Technische Universität Dortmund

Fakultät Physik

Lehrstuhl für Experimentelle Physik 5

Unfolding the muon neutrino flux

Leonora Kardum

A document submitted in partial fulfillment of the requirements for the degree of

Doctor rerum naturalium

(Dr. rer. nat.)

at

Faculty of Physics, TU Dortmund University

Project Supervisors

Prof. Dr. Dr. Wolfgang Rhode and Prof. Dr. Carsten Westphal

2023

Abstract

The IceCube Neutrino Observatory, situated at the South Pole within a cubic kilometer of underground ice, is a state-of-the-art experiment for detecting particles of high energies, with a special focus on investigating neutrino physics. The neutrino flux can be divided into three distinct components: astrophysical, originating from extraterrestrial sources; conventional, arising from the decay of pions and kaons in atmospheric cosmic ray cascades; and the prompt component, which has yet to be detected and stems from the decay of charmed hadrons. This study aims to reconstruct the total flux of neutrinos at Earth and places a particular emphasis on examining the predicted angular dependence. Unfolding encompasses a collection of techniques that aim to determine a quantity in a manner independent of specific assumptions, thereby removing the influence of various assumptions made during the process. In this analysis, the energy spectrum of muon neutrinos is unfolded with the employment of an innovative technique for reshaping the observable space to ensure an adequate number of events in the low statistic region at the highest energies.

This work presents the unfolded energy and zenith angle spectrum reconstructed from eleven years of IceCube data in the range from 500 GeV to 4 PeV energies, and compares the findings with both model predictions and previous measurements.

Zusammenfassung

Das IceCube Neutrino Observatory, das sich am Südpol in einem Kubikkilometer unterirdischen Eises befindet, ist ein hochmodernes Experiment zum Nachweis von Teilchen mit hoher Energie, wobei der Schwerpunkt auf der Untersuchung der Neutrinophysik liegt. Der Neutrinofluss kann in drei verschiedene Komponenten unterteilt werden: astrophysikalische Neutrinos aus extraterrestrischen Quellen; konventionelle Neutrinos, aus dem Zerfall von Pionen und Kaonen in atmosphärischen Luftschauern; und prompte Neutrinos, die noch nicht gemessen wurden, aber aus dem Zerfall von Hadronen aus Charm-Quarks erwartet werden. Diese Studie zielt darauf ab, den Gesamtfluss der Neutrinos auf der Erde zu rekonstruieren, und legt einen besonderen Schwerpunkt auf die Untersuchung der vorhergesagten Winkelabhängigkeit. Die Entfaltung umfasst eine Reihe von Techniken, die darauf abzielen, eine physikalische Größe modellunabhängig zu bestimmen. In dieser Analyse wird das Energiespektrum von Myon-Neutrinos mit Hilfe einer innovativen Technik zur Umformung des beobachtbaren Raums entfaltet, um eine angemessene Anzahl von Ereignissen im Bereich der niedrigen Statistik bei den höchsten Energien sicherzustellen.

Diese Arbeit präsentiert das ungefaltete Energie- und Zenitwinkelspektrum, das aus elf Jahren IceCube-Daten im Energiebereich von 500 GeV bis 4 PeV rekonstruiert wurde, und vergleicht die Ergebnisse sowohl mit Modellvorhersagen als auch mit früheren Messungen.

Contents

Chapter 1

Introduction

Neutrinos are among the most intriguing particles known to science, and the study of their properties has been an active area of research for decades. They are elusive and although the most abundant massive particles, very rarely detected. However, despite their challenging nature, they are key to our understanding of the universe and the fundamental laws of physics. The study of neutrinos is a complex and interdisciplinary field, requiring knowledge of particle physics, astrophysics, and cosmology. In this doctoral thesis, we focus on unfolding the muon neutrino flux, which is a crucial parameter in understanding the properties of these peculiar particles and their sources. Therefore, of particular interest is the coverage of flux produced in the Earth's atmosphere and the flux of neutrinos created in galactic origins.

The goal of this doctoral thesis is to unfold the muon neutrino flux observed by the IceCube Neutrino Observatory, a large neutrino detector located at the South Pole. IceCube is designed to detect high-energy neutrinos produced by astrophysical sources and has observed a significant number of muon neutrinos since its completion in 2010. However, the observed flux of muon neutrinos is contaminated by background events from other types of neutrinos and cosmic rays, and is distorted by various detection limitations particular to this detector.

Unfolding is a set of techniques based on the principle of deconvolution, which aims to recover the underlying physical distribution of a source from the observed data usually affected by instrumental and statistical effects. In astrophysics, unfolding is used to reconstruct the energy spectrum of a source, which represents the radiation intensity emitted by the source as a function of the particle energy. Here, we inspect the diffuse neutrino flux describing the intensity arriving to Earth from all directions and unspecified sources. Unfolding is a complex process that requires careful modeling of the source and instrument properties, as well as an accurate estimation of the statistical and systematic uncertainties associated with the data. This work focuses on optimizing each segment by utilizing machine learning methods. A notable feature of unfolding methods is its model independence, enabling the production of results unaffected by researchers' assumptions.

The energy range investigated starts in the region of 500 GeV dominated by neutrinos created in the atmosphere and spans to the astrophysical-dominated energies of 4 PeV. While the atmospheric region has been extensively investigated, the diffuse flux above 2 PeV has not been reconstructed until this work.

The outcome of this work is an unprecedented calculation of the diffuse neutrino flux, with substantially lower uncertainty enabled by a never-used-before amount of measured data. A particular objective is the investigation of the neutrino flux dependence on the angular direction. The observed variance of flux in five given directions is the first calculation of the flux zenith dependence in IceCube.

This work is constructed as follows: The second chapter introduces basic theoretical concepts necessary in this analysis development, with sections covering neutrino physics and the used detection process. The third chapter covers various machine learning methods used in optimizing steps given in chapter six. The following chapter explains the two types of data used in this work. Then, the mathematical basis of unfolding is presented, being the main building block of the work. Following is the chapter describing the complete analysis chain developed, with specifics needed for reproducibility. Finally, the seventh chapter concludes with the interpretation of the results, its relevance in the current state of research, and future outcomes.

Chapter 2

Theoretical background

The universe is a vast expanse of celestial objects emitting energy through various forms. These objects span from stars, galaxies, supernovae, and black holes to less understood bodies like quasars and pulsars, or phenomena like gamma-ray bursts. A meticulous study of these cosmic objects and events can shed light on the fundamental nature of the universe and the processes that govern its evolution.

Multimessenger astroparticle physics is a rapidly evolving field that aims to study the universe through multiple messengers, such as electromagnetic radiation, cosmic rays (CR), gravitational waves, and neutrinos. This approach enables us to gather information from different sources and compare them to form a more comprehensive understanding of astrophysical phenomena, ultimately leading to groundbreaking discoveries and opening new avenues for understanding the universe. Among these messengers, neutrinos have played a crucial role in multimessenger astroparticle physics due to their unique properties, including their ability to travel through dense matter without interacting which enables an undistorted reconstruction of their sources' directions. In this section, we will discuss the historical context of neutrino detection and the important role it plays in multimessenger physics.

2.1 Multimessenger astrophysics

The idea of using multiple messengers to study astrophysical phenomena has been around for several decades. The first observation of cosmic rays being high-energy particles originating from outside the Earth's atmosphere was done in the 1930s by Victor Hess [Hes40]. This triggered the rapid development of cosmic ray detectors including cloud chambers, spark chambers, and emulsions. As a result, muons and pions were discovered, and they will later play a crucial role in astroparticle physics and the study of both atmospheric interactions and galactic sources. In 1960, the physicist John A. Wheeler first proposed the idea of using gravitational waves to study the universe [HTWW65]. He suggested that if two massive objects were to collide, they

would create ripples in the fabric of spacetime that could be detected on Earth. However, it wasn't until 2015 that the Laser Interferometer Gravitational-Wave Observatory (LIGO) made the first detection of gravitational waves, confirming Wheeler's theory and starting a new era of multimessenger astrophysics [A$^+$16b]. The first observation of a gamma-ray burst (GRB) by a gamma-ray experiment was made in 2019 by MAGIC [7419]. In 2020, the HAWC gamma-ray observatory detected a very high-energy GRB that was also detected by the Fermi Gamma-ray Space Telescope.

In 1983, the first neutrino detector was built, the Kamiokande in Japan, succeeded by the Super-Kamiokande and to be succeeded by Hyper-Kamiokande in the following years. In 1987, eleven neutrinos were detected over a time span of thirteen seconds, coming from the direction of supernova SN1987A in the Large Magellanic Cloud, substantially contributing to the understanding of supernovae explosion mechanism [H$^+$88]. This event marked the first time that neutrinos had been observed from a source outside the solar system and has led to the development of dedicated neutrino observatories in the field of astronomy, including but not limited to Antares and KM3NeT in the Mediterranean Sea, the Baikal Deep Underwater Neutrino Telescope (BDUNT) and Baksan Neutrino Observatory (BNO) in Russia, and the Helium And Lead Observatory (HALO) in Canada. Starting construction in 2005, the IceCube Neutrino Observatory (IceCube) became the largest neutrino detector in existence. With this state-of-the-art telescope, many breakthroughs in the field of astroparticle physics were facilitated, and its unprecedented aptitude for data analysis enabled the development of this work. The workings of this detector will be thoroughly covered in the subsequent section.

In 2018, the IceCube neutrino detector detected a high-energy neutrino from outside the solar system, tracing back to a blazar, a type of active galactic nucleus [CFLM$^+$18]. This was the first time that a high-energy neutrino had been associated with a specific source. Not only has this event solidified the importance of astrophysical neutrino detection, but also confirmed the eminence of IceCube in this field.

2.2 Neutrinos

Neutrinos are fundamental particles belonging to the lepton family. They are electrically neutral and interact only weakly with matter, making them extremely difficult to detect. The very properties causing the strain in their detection make them ideal candidates for transferring information over tremendous distances in space. Neutrinos are created in a variety of astrophysical sources, such as the Sun, supernovae, and active galactic nuclei. The detection of neutrinos from these sources can provide valuable information about their properties and the astrophysical phenomena that produce them. In this section, we will discuss the properties of neutrinos and how they can be detected in astrophysics, with a focus on muon neutrinos.

There are three types of neutrinos: electron neutrinos, muon neutrinos, and tau neutrinos, which correspond to the three charged leptons. Although believed to have been massless for a long time, neutrinos have three distinctive mass states, and each neutrino is a combination of mass eigenstates. The hierarchy of the mass has not been determined yet [QV15], and mass states are still calculated only to an upper limit. The neutrino physics was further complicated with the discovery of flavor mixing, where neutrinos also come in combinations of the flavor eigenstates. As a result, neutrinos can be detected in a different flavor than the one created in, referred to as neutrino oscillations. The mixing of neutrino states is conveniently described since 1962 in the Pontecorvo–Maki–Nakagawa–Sakata matrix (PMNS matrix) [IKMY15] parametrized with three mixing angles ($\theta_{12}, \theta_{13}, \theta_{23}$) and the charge-parity violations phase angle δ. In this analysis, muon neutrinos are used in determining the flux, however, the possible oscillation to the muon flavor is not of interest and therefore out of the scope of this work.

2.2.1 The Flux and its components

The intensity of neutrinos arriving at Earth can be examined by looking at either distinct point sources in the sky, corresponding to cataloged astrophysical phenomena [BGH+21], or by investigating the complete flux detected integrated over the observed sky. The latter is referred to as the diffuse flux, and resulting from the observatory's position at the South Pole, the Northern Sky is investigated by measuring neutrinos that previously passed through the inside of our planet.

There are three main components of the neutrino flux: conventional, prompt, and astrophysical, shown in Figure 2.1. The flux is parametrized by a power law in the form

$$\frac{d\Phi}{dEd\Omega} = \Phi^0(\frac{E_\nu}{100TeV})^{-\gamma} \tag{2.1}$$

where Φ^0 is called the normalization constant, and γ the spectral index, determining the slope of the function. Although all are expected to follow the same form, the three different components attain different parameter values due to the variance in their properties.

The conventional neutrino flux is produced by cosmic ray interactions with the Earth's atmosphere. These interactions produce charged particles, such as pions and kaons, which subsequently decay into neutrinos. The main production channels are the pion and kaon decays

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \tag{2.2}$$

$$\pi^- \rightarrow \mu^- + \bar{\nu_\mu} \tag{2.3}$$

$$K^- \rightarrow \mu^- + \bar{\nu_\mu} \tag{2.4}$$

$$K^+ \rightarrow \mu^+ + \nu_\mu \tag{2.5}$$

with pions and kaons both being long-lived particles. The decay mean free path of mesons is given by

$$\lambda_m = \gamma c \tau_m \tag{2.6}$$

with τ_m being the meson's lifetime, while their interaction mean free path is

$$\lambda_m = \frac{A m_m}{\sigma^{air}} \tag{2.7}$$

with σ^{air} as the meson-air cross-section and A as the mean mass number of air nuclei. Combining this gives the critical energy of a meson, at which the interaction starts dominating the decay

$$\epsilon_m = \frac{c \tau_m}{m_m c^2 h_0} \tag{2.8}$$

with h_0 being the scale height. With muon energies E_μ below $\epsilon_\pi = 115 GeV$ and $\epsilon_K = 850 GeV$, decay dominates and the muons follow the creating mesons' spectrum, consequently imitating the spectrum of Cosmic Rays. At energies $E_\mu > \epsilon_\pi, \epsilon_K$, interaction dominates and the muon flux gets suppressed. The large decay mean free path allows mesons to lose energy before creating a muon, and as a consequence create a softer flux. Its spectral index is expected to be around 3.7.

Prompt neutrinos are produced in the same cosmic ray interactions as conventional neutrinos but come from the decay of short-lived particles. These are mainly charmed mesons, and the creation is governed by the decays

$$D^+ \rightarrow \mu^+ + \nu_\mu \tag{2.9}$$

$$D^- \rightarrow \mu^- + \bar{\nu_\mu} \tag{2.10}$$

which is, together with the electron, the favored channel. The tau creation is suppressed compared to the previous two. Although the decay mean free path is as well described with Equation 2.6, the much shorter lifetime of charmed mesons reduces the interaction probability, and the charmed mesons almost always decay before interacting. With no loss of energy in the atmosphere, these particles create neutrinos with a harder energy spectrum. The prompt component starts to substantially contribute to the flux at energies around 100 TeV, however, due to the inability to discriminate neutrinos created from different mesons, both conventional and prompt neutrinos are observed together. Under the common name of atmospheric neutrinos, the two fluxes are often parametrized jointly. In the region of conventional dominance at less than 100 TeV, this component has been observed solely, but nonetheless, the prompt component has not been observed nor parametrized yet [Sch13b].

Last, but not least, is the astrophysical component. It is constituted from neutrinos created in astrophysical sources, independent of processes on Earth and inside its atmosphere. Due to the

range of different sources, setting the expectation for a general spectral index is not straightforward. IceCube has already put restraints on the contribution of GRBs and Active Galactic Nuclei (AGN) to the astrophysical component to at most 20% [A$^+$17a, A$^+$17e]. Starburst galaxies (SBG) are a possible candidate for explaining a substantial amount of astrophysical neutrinos, but the current models limit the energies in this channel of production to the PeV range [PW18]. A neutrino production channel common to several astrophysical sources is the proton-photon interaction

$$p \rightarrow n + \pi^+ \rightarrow n + \mu^+ + \nu_\mu \tag{2.11}$$

provided an adequate column density of target photons. Neutrino production is expected in the proton-proton collisions of Fermi mechanisms, with the spectral index imitating the one of the proton flux [Mé s17]. Despite the previously mentioned features supporting a multi-model parametrization of the astrophysical component, current efforts are to describe it with a single power-law, due to the complexity of the problem and the sparsity of detected neutrinos. At 2.62 by the most recent measures [A$^+$22], its spectral index is determined experimentally with no agreed-upon theoretical model to date.

2.2.2 Angular dependence

The angular dependence of the neutrino flux is a complex matter, spanning from the different flavors of detected neutrinos and the different production mechanisms observed through the three components. The dependency is expected due to the different lengths of paths particles pass depending on their arriving zenith, however several additional factors influence the detected intensity.

As previously described, the conventional component is primarily driven by cosmic rays interacting with the Earth's atmosphere, which results in a higher flux of neutrinos near the horizon where the atmosphere is thicker. Consequently, a strong dependence on the zenith angle arises in conventional-dominated region. Even before the creation of the muon, the mentioned impact of energy loss in the long traversed paths of primary mesons inside the atmosphere will therefore also influence the distribution of resulting neutrinos. When a pion collides with the atmosphere perpendicularly, it will swiftly reach the lower, denser, atmosphere where it interacts with more particles, leading to an increased likelihood of interaction. In contrast, entering at a larger angle, the pion spends more time in the less populated upper atmosphere, which enables it to decay without losing energy to interactions, and surrendering more energy to created neutrinos. This creates an additional energy dependence based on the incoming angle, apart from the expected creation rate. Upon muon creation, which retains the direction due to momentum conservation, the column depth of the atmosphere will impact its decay probability. When the angle of incidence is increased, the distance between the origin of the muons in the

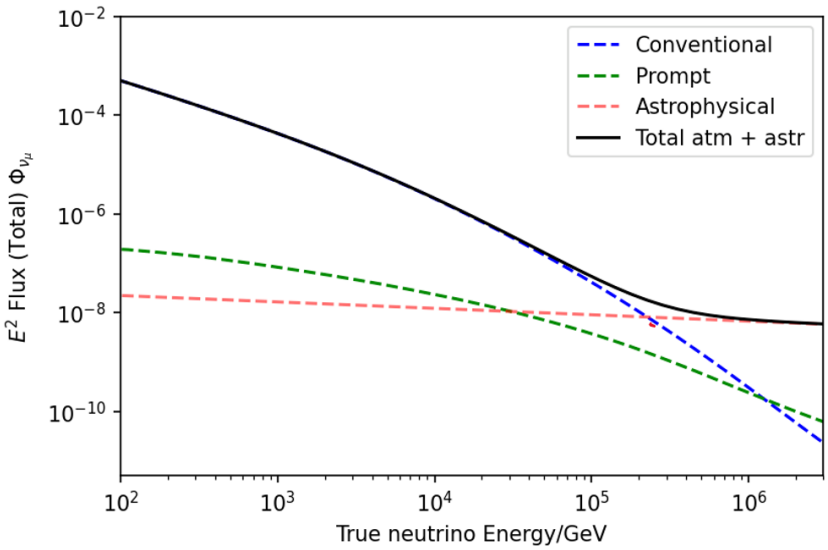

Figure 2.1: The total neutrino flux and its components, weighted with energy squared. Atmospheric components are simulated with MCEq [FEG+15], described in Chapter 4. The dominance of the conventional component is discernible from the flux, with the prompt component being concealed by the conventional at lower energies, and by the astrophysical at higher energies. The flux is dominated by the astrophysical component at energies above 700TeV, conveniently named the signal region. Atmospheric components are constructed from expectancies, while the astrophysical component is an experimental observation taken from [A+16a].

atmosphere and the detector on the earth's surface also increases, reaching its maximum at 180°. At this angle, muons have to travel an average of 400 kilometers to reach the Earth, with an average decay length of only 33 kilometers. Only several high-energy muons can cover these distances, resulting in the lowest rate at this angle. The dependency is shown in Figure 2.2.

Due to the short time spent inside the atmosphere in the fast decays of prompt particles, the column depth does not impact the energy loss, and therefore nor the neutrino rate. The prompt component is expected to be isotropic for this reason, as shown in Figure 2.3. Both conventional and the prompt components might be influenced by additional factors including temperature, season, and similar, but these models have not been developed to include the angular dependence yet [HTWZ19, H+21].

As is the prompt component, the astrophysical component is expected to be isotropic, although its sources are not fully unraveled. The experimental data points to an isotropic nature of the flux, with most high energy neutrinos related to an extragalactic origin, however, recent measures show support for astrophysical neutrino excess in the Galactic plane [C+17].

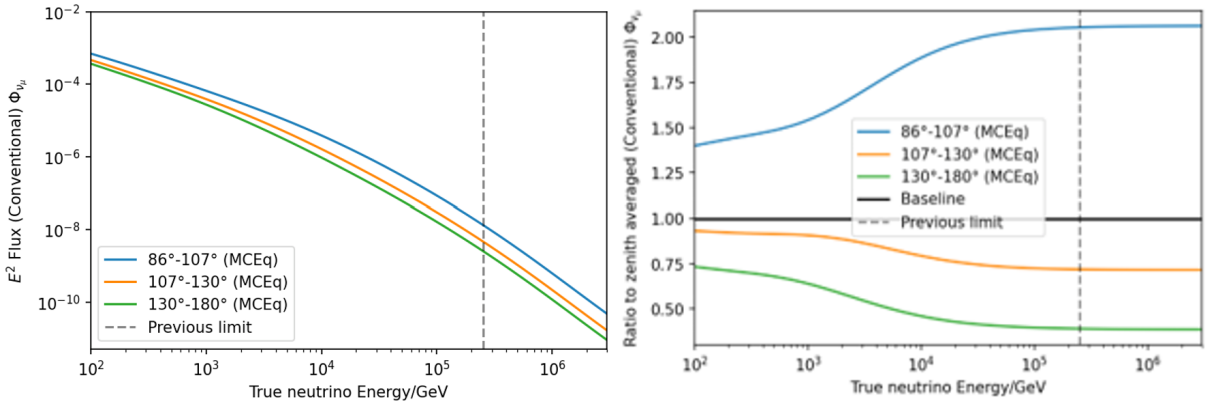

Figure 2.2: Angular dependence of the conventional component. The left figure shows conventional neutrino fluxes weighted with squared energy in three different angle bins. It is imminent to notice the decrease in neutrino rate for larger angles. The right figure shows the ratio between conventional neutrino fluxes in angle bins and the all-sky conventional flux. At high energies, the ratio reaches a constant value. The dashed line shows the highest energy considered in previous analyses.

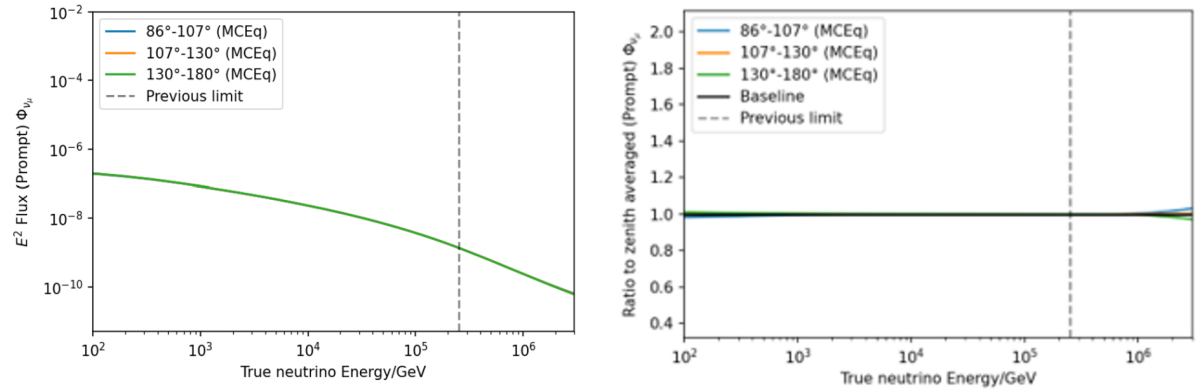

Figure 2.3: Angular dependence of the prompt component. The left figure shows prompt neutrino fluxes weighted with squared energy in three different angle bins. The fluxes exhibit no difference and overlap each other. The right figure shows the ratio between prompt neutrino fluxes in angle bins and the all-sky prompt flux. Due to the isotropic nature of the prompt flux, its ratio remains one in all angle bins. The dashed line shows the highest energy considered in previous analyses.

Considering everything stated above, it concludes the angular dependence of the total neutrino flux arises solely from the anisotropy of the conventional component. The prompt and astrophysical can both be approximated to isotropic. The dominance of the three components

interchanges with energy, with conventional solidly dominating the low energy region, prompt having a rising contribution in middle energy, and astrophysical dominating the high energy range. Consequently, the low-energy region will be characterized by stronger angular dependence, with isotropic behavior rising with energy, as depicted in Figure 2.4.

2.2.3 Why muon neutrinos?

In the large volume of ice, neutrinos interact with matter through charged current (CC) interactions and neutral current (NC) interactions. In CC interactions, the neutrino interacts with a nucleon in the target material and produces a charged lepton of the same flavor as the neutrino. The resulting lepton can then be detected through an intricate series of steps. NC interactions lack the creation of lepton, yet the interaction of a neutrino with the nucleus creates an electromagnetic shower, resulting in a signature inside the detector volume distinctive from a CC signature. Since interactions creating a charged, and therefore detectable, lepton are of interest, NC interactions are cleaned out of the sample.

Electrons are lightweight particles that interact strongly with matter. When an electron loses energy through bremsstrahlung, the emitted photons can create electron and positron pairs, which in turn produce more photons, creating a cascade of particles and light. The energy is emitted in a wide cone shape around the electron's direction as the created shower of particles develops in a tree-like structure. Due to their interactivity with matter, electrons change directions often in collisions and interactions with matter, resulting in a torturous path. The release of electromagnetic radiation on this path results in energy radiated in several directions, that cannot be easily reconstructed to a single course.

Muons, on the other hand, are much heavier than electrons and have a substantially lower cross-section. Their penetrating nature allows them to reach detectors on sea level, with high energy muons even penetrating Earth to reach underground facilities. This sets them apart from other leptons as the favored particle in neutrino detection. Although considered a background in this work, muons created in the atmosphere are a valuable constitution to many research goals of IceCube, including studies on cosmic rays. Particles with lower mean free path than muons have to be detected at the surface of Earth, or even by specialized satellites, leaving muons as the only medium in underground detection with a natural shield against noise.

Neutrinos' low interactivity with matter allows them to surpass large distances inside the rock and ice volume undisturbed. The creation of a muon in the CC interaction of a neutrino most often happens in the bedrock surrounding the detector, and the muon retaining a part of neutrino energy continues to pass the detector volume. The stochastic losses of energy will then leave a track-like signature in the detector around the muon's path, with the losses for the most part being dependent on the muon's energy. However, loss through material ionization is constant for these muon energies. The lack of correlation between the original energy of the

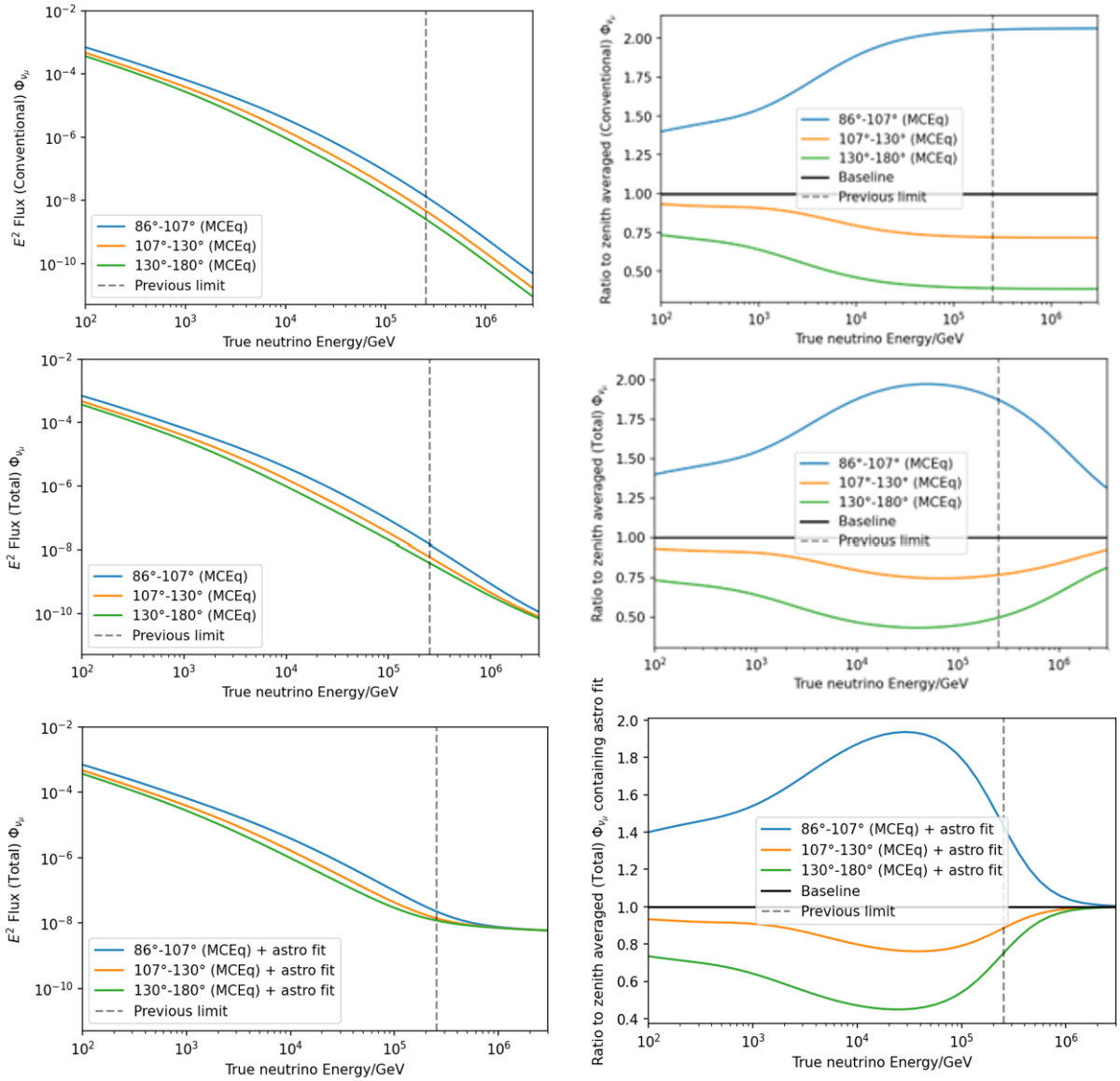

Figure 2.4: Angular dependence of neutrino flux constituted of different components. The left column shows neutrino fluxes weighted with squared energy for the conventional (first row), conventional and prompt (second row), and all three components (third row). The right column shows the angular dependency shown as a ratio to the full-sky flux for the respective fluxes. With rising energy, isotropic components start to dominate the flux, and the ratio gravitates towards one. The lower right plot is the expected outcome of an angular dependence analysis like this work. Atmospheric components are simulated with MCEq [FEG+15], while the astrophysical component is taken from [A+16a].

muon and the energy loss through ionization leads to inability of ionization losses to be used as proxies in energy reconstruction. These losses are governed by the Bethe-Bloch equation

$$-\left\langle \frac{dE}{dx} \right\rangle = K\frac{Z}{A}\frac{z^2}{\beta^2}\left[\ln\left(\frac{2m\gamma^2\beta^2}{I} - \beta^2\right)\right] \tag{2.12}$$

where $K = 4\pi N_A, m_e, r_{2,c^2}$ and Z being the atomic number of the medium, I the ionization potential, N_A Avogadro number, m_e and r_e the mass and radius of the electron, and z the charge of the particle (in this case a muon). The energy loss, shown in Figure 2.5, increases with $\sim \beta^{-2}$ resulting in slower particles being more ionizing. With increasing muon speed, the energy loss reaches a minimum in the $MeV\ g/cm^2$ range in which it remains for energies covered by the IceCube experiment.

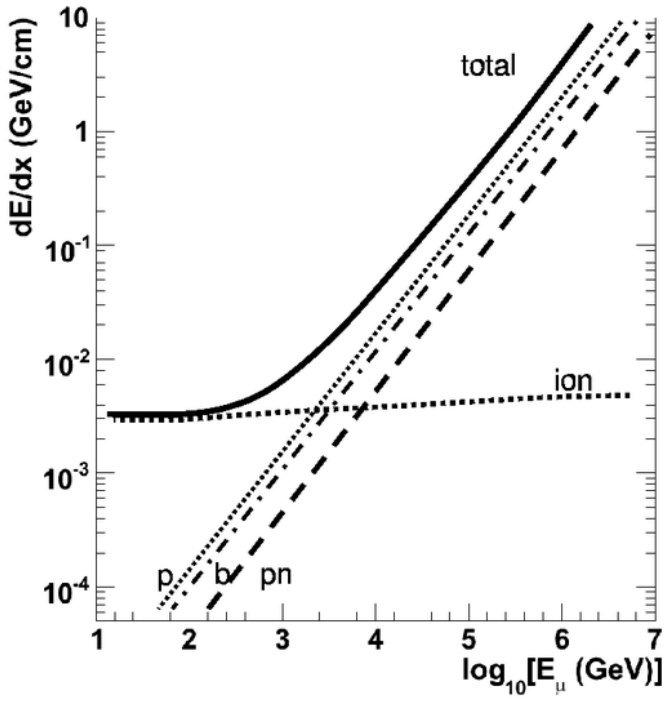

Figure 2.5: Energy loss of muons in water [BVE11]. Labels represent the following energy loss processes: p for pair production, b for bremsstrahlung, pn for photo-nuclear interactions, and ion for ionization. While other interactions show a dependence on muon energy, ionization loss is independent of starting energy and cannot be used as a proxy for its calculation.

2.3 Cherenkov radiation

Cherenkov radiation is a type of electromagnetic radiation that is emitted by a charged particle when it travels through a medium at a velocity higher than the speed of light in that medium.

The radiation is emitted in a cone-shaped pattern with the apex of the cone located at the point where the charged particle interacts with the medium [Kno10]. The angle of the cone is given by the following equation

$$cos\theta_C = (n\beta)^{-1} \tag{2.13}$$

where θ_C is the angle of the cone, n is the refractive index of the medium, and β is the velocity of the charged particle relative to the speed of light in vacuum. From here, the maximal angle of particles approaching the speed of light is $\theta_C max = cos^{-1}(1.31^{-1}) \sim 40.2°$. Furthermore, the intensity of Cherenkov radiation is

$$\frac{dN}{dxd\lambda} = \frac{4\pi^2 z^2 e^2}{hc\lambda^2} sin^2\theta_C \tag{2.14}$$

with N being the number of photons emitted, λ the wavelength, and h the Planck's constant. The intensity peaks at wavelengths between 300 and 400 nanometers, giving Cherenkov radiation a distinctive blue glow visible to human eyes.

Cherenkov radiation can be detected using a variety of methods, including photomultiplier tubes (PMTs) and charged-coupled devices (CCDs), with PMTs being the established method in water and ice detectors. The exact detection procedure can be fully understood with the introduction of the detector operation covered in the following subsection.

2.4 IceCube Neutrino Observatory

The IceCube Neutrino Observatory [A$^+$17b] is a particle detector located at the geographic South Pole and operated on the Amundsen-Scott South Pole station. It is the largest neutrino detector in the world, with an instrumented volume of one cubic kilometer of Antarctic ice. The main goal of IceCube is neutrino detection with the aspiration to infer many different conclusions in the astrophysics area of interest. In this work, we are interested in its detection of muon neutrinos in already mentioned track-like events with the described Cherenkov radiation. The main detection process relies on the mechanism of PMTs.

2.4.1 Main detector units

A photomultiplier tube (PMT) is a highly sensitive device used to detect and amplify weak light signals. The basic principle behind its operation is based on the photoelectric effect, where photons of light are converted into electrons when they strike a photosensitive material. The resulting electrons are then multiplied using a series of electron multiplication stages, or dynodes,

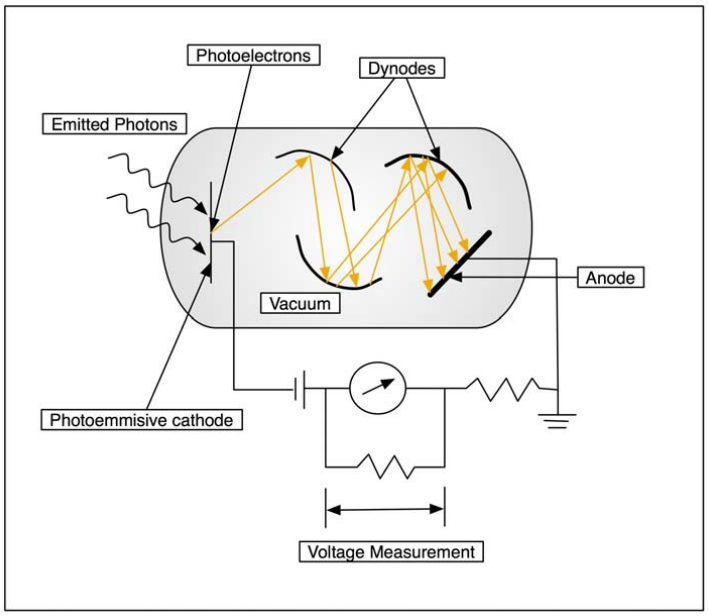

Figure 2.6: The schematic representation of a conventional Photomultiplier tube (PMT), from [BWM01].

resulting in a large output signal. The PMT consists of three main parts: the photocathode, the electron multiplier, and the anode, depicted in Figure 2.6. The photocathode is a photosensitive material that absorbs photons of light and emits electrons through the photoelectric effect. The electron multiplier is a series of dynodes, which are metal electrodes arranged in a chain-like fashion. The anode is a metal electrode that collects the multiplied electrons and produces an output signal. When a photon of light enters the PMT, it strikes the photocathode and ejects an electron. This electron is then accelerated towards the first dynode, where it strikes the surface and causes the emission of several secondary electrons. These secondary electrons are then accelerated towards the next dynode, where they strike the surface and cause the emission of even more secondary electrons. This process is repeated for several dynodes until a large number of electrons are produced and collected at the anode. The key to the PMT's high sensitivity is the ability to multiply electrons at each dynode. This is achieved by using a high voltage applied across the dynodes, which causes the emission of more secondary electrons. The number of electrons produced at each dynode is typically a factor of 10 to 20 times higher than the previous dynode. This multiplication process results in a very high gain, which is the ratio of output electrons to input photons. In addition to its high sensitivity, PMTs also have a fast response time and low noise characteristics.

The amplification of electrons is however a statistical process, and PMTs do not have a well-defined amplification constant with drastic differences between units. Each PMT is described with several variables relating to collected charge per event. An expected charge generated in

some PMT for a single photon is a probability curve known as the single photoelectron (SPE) charge distribution. In optimal conditions, a single photon would generate a response of a single photoelectron, but the stochasticity of the measurement smears the distribution around the value of 1, as shown in Figure 2.7. The collected charge is therefore an additional parameter to be calibrated.

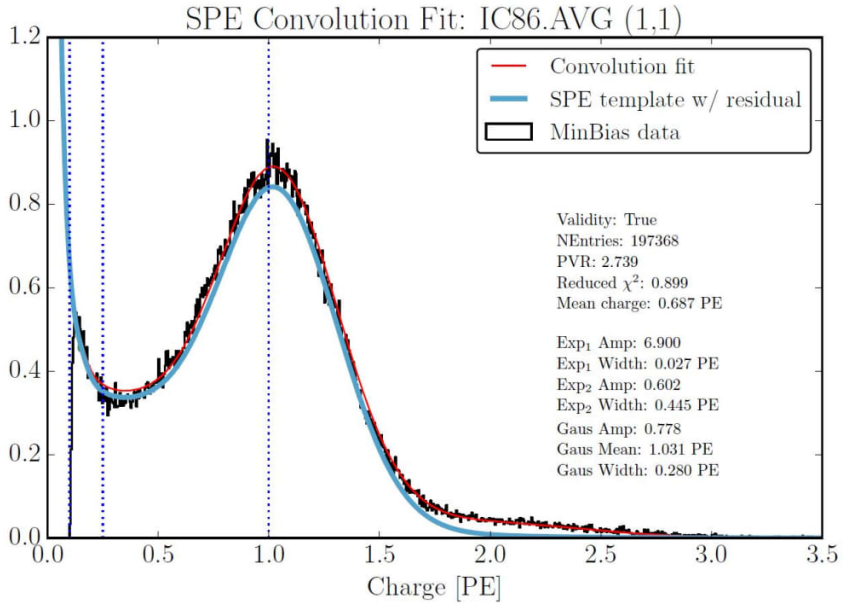

Figure 2.7: Single photoelectron charge distribution in an example DOM, from [BWM01]. The black histogram shows the charge in photoelectrons collected in DOM 1 on String 1 for a single injected photon. An ideal PMT would always generate one photoelectron. The blue and red lines show the fit adjusted to describe the charge collection response of a given PMT, used to calibrate the analyzed data.

The 25-centimeter PMTs are fitted into a 30-centimeter wide Digital Optical Module (DOM) shown in Figure 2.8, together with the mainboard and an LED flasher. To withstand the enormous pressure under over a kilometer of ice, DOMs are built from a pressure-resistant glass able to resist over 400 atmospheres. An incorporated mu-metal wire cage provides magnetic shielding.

2.4.2 IceCube Observatory Architecture

The detector consists of an array of 86 strings, each containing 60 digital optical modules (DOMs) housing a PMT with 17-meter spacing. The strings were deployed vertically into the ice using hot water drills. The first five years of construction were dedicated to the deployment of the strings, which were deployed at depths ranging from 1,450 to 2,450 meters. Strings are

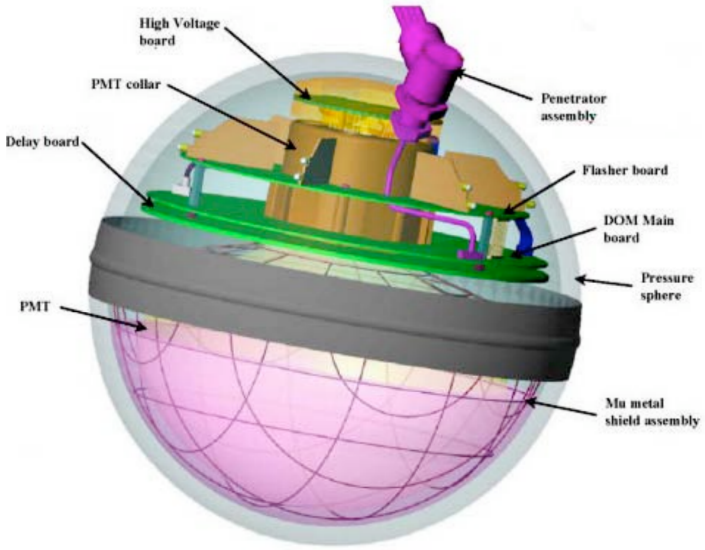

Figure 2.8: The architecture of a single Digital Optical Module in IceCube [Sto05].

spaced at regular intervals of 125 meters.

Since particles with low energy can go undetected in the spacing between strings, a more dense detection volume was deployed at the bottom center. Under the name DeepCore, the volume has strings at 70 meters spacing with DOMs at only 7 meter spacing. This architecture has reduced the lower energy limit of IceCube to 10 GeV. The architecture is shown in Figure 2.9.

The surface contains 81 stations each with two PMT containing water tanks named IceTop. Tanks were designed to detect air-shower particles and serve as a veto primarily in background cleaning.

The DOMs are connected to the surface via a cable, which provides power and data communication. Apart from background events, the proper analysis of IceCube events calls for in-depth knowledge of the surrounding ice and its properties, including the glacial movements and its impurity. In addition, the efficiency and operation of the DOMs at different depths have to be taken into account. We refer to these additional parameters as systematic effects, and they are respectively handled in this and other works by IceCube.

2.4.3 Event detection

The collected charge in each DOM is conveniently represented in 3-dimensional IceCube events, as shown in Figure 2.10. The response of each DOM is visualized with a bubble whose size corresponds to the collected charge, resulting in large bubbles signaling high-energy events. Additionally, bubbles are color-coded based on the detection time, with the color spectrum ranging

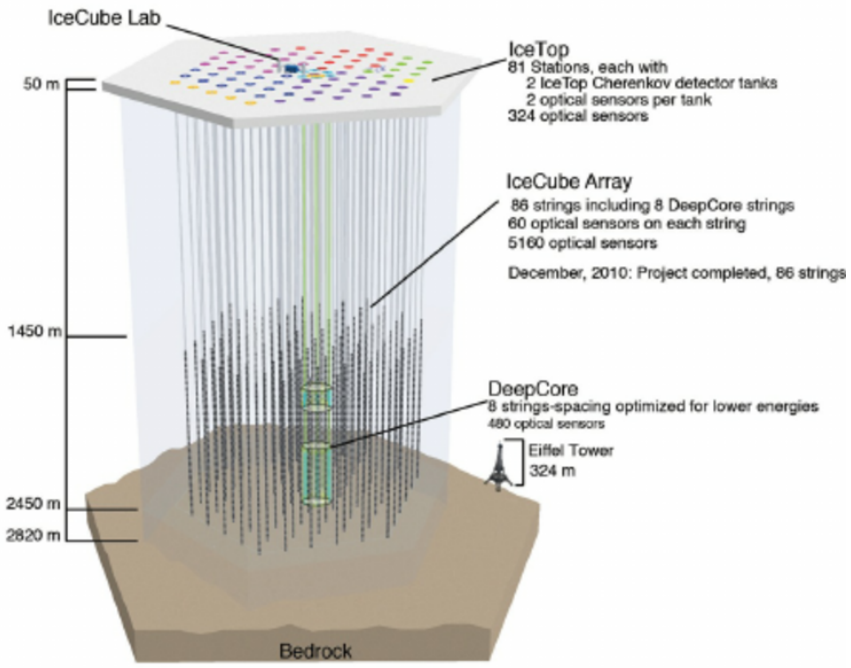

Figure 2.9: The architecture of the IceCube Neutrino Observatory [Pon15]. Distinctive sections of detector volume are marked in the schematic, including specific quantities of detector parts. The Eiffel Tower is for comparison purposes.

from red for early signals to blue for later triggers. The direction of a particle can be simply inferred from the color of the bubbles, with its path pointing from the red DOMs towards the blue range of the spectrum, in track-like events. However, in cascade events, the reconstruction of the initial particle is not straightforward to conclude, and due to the lack of their relevance in this work, the exact mechanism is out of scope.

2.4.4 The Collaboration and its Research aims

Apart from its well-known focus on detecting high-energy neutrinos, the IceCube has a wide range of research objectives, including Dark Matter, Cosmic Ray physics, and Glaciology. It has also developed a real-time notification software with the goal of detecting transient sources, with its findings made publicly available for other scientists to use [A$^+$17c]. Its most outstanding achievement is the constructive and competent collaboration of over 300 scientists spread in 58 institutions around the globe, all working for common scientific goals.

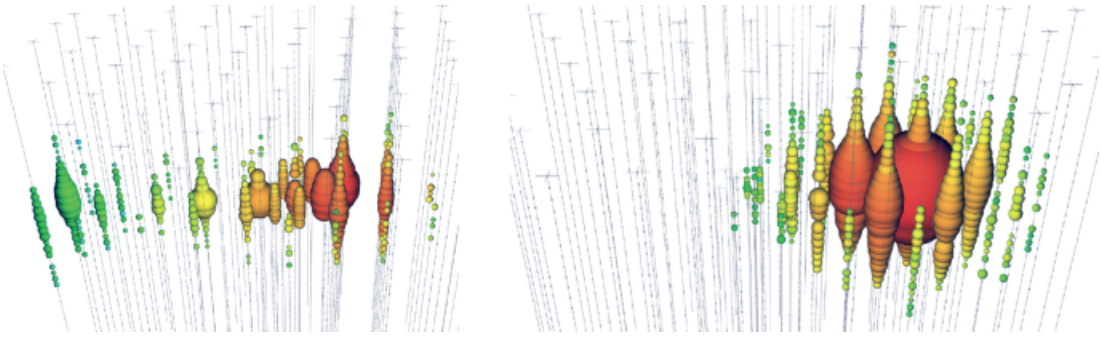

Figure 2.10: Two example detected events in the IceCube Neutrino Observatory [AHPdlH18]. The left figure corresponds to a track-like event, in which the particle's direction is more readily reconstructed. In this example, the particle has moved from center-right to center-left. The right figure represents a high-energy cascade event. The size of the red bubble points to the event being energetic, with the timestamps moving outward in every direction from the center. This is a typical cascade event in which the trajectory of the starting particle is difficult to observe.

2.5 Current results

IceCube has set itself on the pinnacle of neutrino flux calculation. The energy range examined to this date is up to 2 PeV for bin-wise methods, and unbounded for astrophysical fits.

Figure 2.11 shows the results of past efforts in describing the neutrino flux. The atmospheric region corresponds well to previous results created with the preceding AMANDA telescope [A$^+$09], and to results created with previous architectures [A$^+$15a], [A$^+$17d]. The large water telescope ANTARES under the Mediterranean Sea has also produced agreeing results [AM$^+$13], and all are complemented with low energy results from Frejus [DRB$^+$95].

The normalization constant and the spectral index of the astrophysical flux as described in Equation 2.1 has to be determined in separate analyses and with special care. IceCube has produced fits in the low statistic region of high energies by availing data spanning several years. Most recent measurements of the diffuse flux calculated both in a bin-wise and a linear model have produced an estimate of $\frac{d\Phi_{astro}}{dEd\Omega} = 1.44^{+0.25}_{-0.26}(\frac{E_\nu}{100TeV})^{-2.37^{+0.09}_{-0.09}}$ [A$^+$22]. The high energy starting events (HESE) analysis estimated the best fit for the astrophysical flux as $\frac{d\Phi_{astro}}{dEd\Omega} = 6.37^{+1.46}_{-1.62}(\frac{E_\nu}{100TeV})^{-2.87^{+0.20}_{-0.19}}$ [A$^+$21]. Results considered as current best estimate are the astrophysical fit at $\frac{d\Phi_{astro}}{dEd\Omega} = 0.9^{+0.30}_{-0.27}(\frac{E_\nu}{100TeV})^{-2.13\pm0.13}$ [A$^+$16a] and has been used in this analysis in later steps, whenever creating assumptions in the process of optimization.

This work aims to provide bin-wise estimation in the higher energy region visible to IceCube, and enrich the current measurements with results of unparalleled precision.

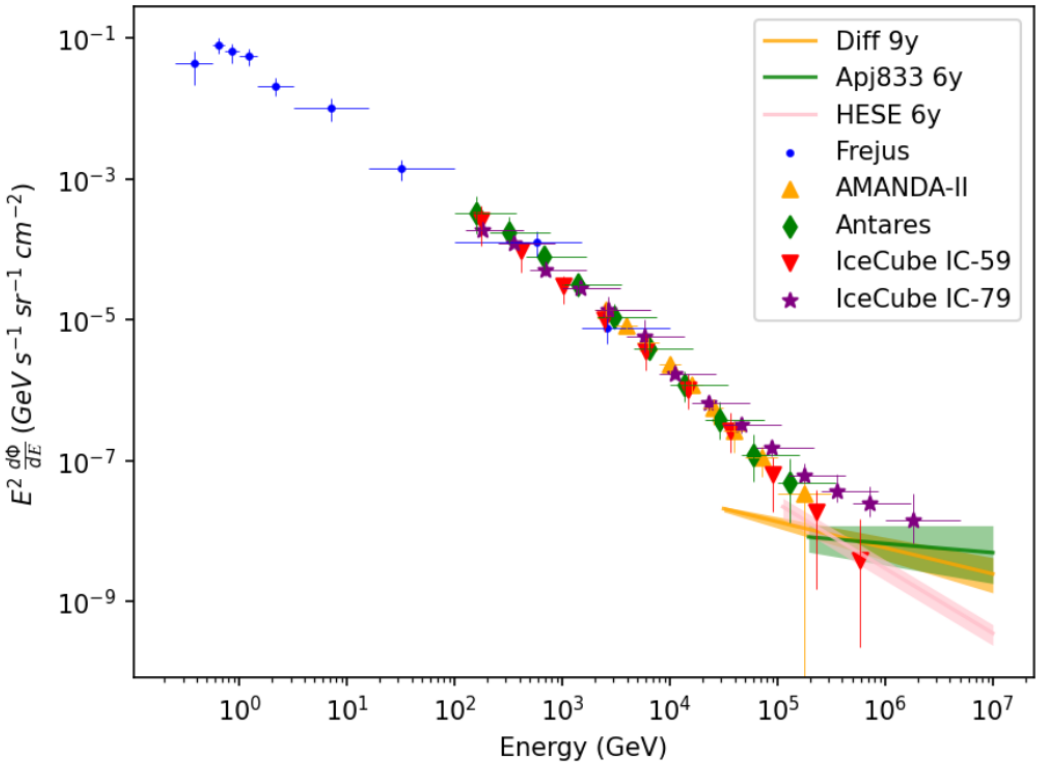

Figure 2.11: Experimental results of neutrino flux measurement to date. The flux is weighted with energy squared for higher prominence of its components. Results include low energy results from the underground Frejus detector [DRB+95], previous architectures of IceCube [A+09], [A+15a], [A+17d], and results from the water detector ANTARES [AM+13]. Astrophysical fits are made in IceCube assuming the flux model to be described by Equation 2.1. HESE analysis [A+21] is depicted in pink, and the results from the diffuse analysis on astonishing 9.5 years of data [A+22] in orange. Experimental fit from six years [A+16a] has been used throughout this work as the common best assumption of the astrophysical flux parametrization.

Chapter 3

Machine Learning methods and Statistical tests

Machine learning is a branch of Artificial Intelligence (AI) that deals with the development of algorithms and models that can learn and make predictions or decisions based on data. Apart from its wide use in industry, it is an unavoidable part of preparation and analysis of data for scientific purposes.

The primary goal of machine learning is to enable computers to learn from data, so they can improve their performance on a particular task over time, or to be utilized for tasks too complex to be analyzed by hand. This process involves several steps, including data collection, data preparation, model training, and model evaluation.

Data collection involves gathering relevant data that can be used to train a machine learning model. Data collection in the context of this work is done by processes described in Section 2.4. The collected data amounts to vast memory, and contains noise and unnecessary information. Data preparation involves cleaning and preprocessing the data, so it can be used for training. The first level of data preparation is done at the South Pole, before being transferred for further investigation. Collected data, in the form used in this thesis, is further described in Chapter 4, while the additional preprocessing steps are given in Chapter 6.

Model training involves feeding the data to a machine learning algorithm, so it can learn from it. The training is highly problem-specific, therefore the algorithm has to be carefully chosen based on the needs of the scientist. Several different methods have been used in this work, however, most steps have included the Random Forest algorithm, and it will be described in more detail in the following section.

Model evaluation involves testing the model's performance on a separate dataset, so we can determine how well it can generalize to new, unseen data, and how well it can perform in future tasks. Although there are general evaluation metrics that can be calculated for any model, the evaluation is also problem-specific, and depending on the goal the metrics have different

relevance.

A general separation of machine learning is to supervised and unsupervised learning. Supervised learning includes algorithms in which they are trained on input data with some known output. Depending on its type, the model adjusts parameters to find the best-describing functions that map the input to the known result. In unfolding, the model is trained on simulated IceCube events combined with the simulated response. In this way, the model learns which events cause the shape of the observables given in simulation, and can generalize this to new data. The algorithms surrounding the unfolding itself (e.g. preprocessing methods) in this work also belong to supervised learning. Unsupervised learning is used in cases where the output is not known or simulated, like pattern recognition, but it is outside the scope of this thesis.

3.1 Random Forest algorithm

A Random Forest is an ensemble learning algorithm that is used for classification and regression tasks. In general, ensemble learning algorithms combine multiple machine learning algorithms to improve their performance. Specifically, a Random Forest is a combination of decision trees.

3.1.1 Decision Tree

Decision trees are a popular and effective machine learning algorithm that can be used for both classification and regression tasks. They are easy to interpret and visualize, making them a popular choice for tasks where understanding the underlying decision-making process is important. A decision tree is a tree-like structure (hence its name) that models decisions and their possible consequences. In a decision tree, each node represents a decision, and each branch represents a possible outcome of that decision, as in Figure 3.1. The decision tree is constructed by recursively splitting the data into subsets based on the values of the input features until a stopping criterion is met. The stopping criterion can be based on various factors, such as the number of data points in a subset, the depth of the tree, or the purity of the subsets. The purity of a subset refers to the degree to which the output values in the subset are the same. The Gini index is a measure of the impurity of a subset and is calculated as the probability of misclassifying a randomly chosen element into that subset. A feature that splits the data into subsets that are more pure with respect to the target variable will have a lower Gini index than a feature that splits the data into subsets with high impurity.

To achieve the optimal set of decisions in a classification task, decision trees seek to reduce entropy through the choice of splits maximizing information gain. Entropy can be understood

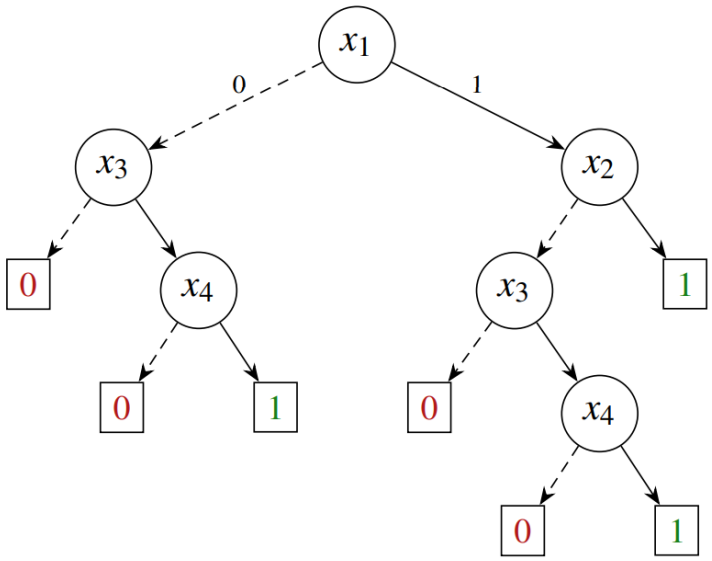

Figure 3.1: Schematic representation of a decision tree, from [IIMS20]. Each decision x is related to a limit, in the case of numerical features like in this work. Any event i with a value of n_n higher than the decision x will attain a number of 1, and move towards the right-hand decision. Dashed and undashed lines show two types of decisions, x are often referred to as nodes, and boxes show leaves at the end of the decision process.

as a measure of uncertainty, or disorder, in the data. It is given with

$$E(Y) = -\sum_{i=1}^{N} -p_i log_2 p_i \tag{3.1}$$

for an element i among N classes, and p_i being the probability of randomly selecting the element from data. An entropy of 0 relates to a pure set of identical elements, therefore higher entropy signals higher impurity.

Information gain is a measure of the reduction in entropy achieved by introducing new information, where this is achieved by splitting the data based on a particular feature in these types of models. Information gain from introducing a new decision is

$$IG(Y|D) = E(Y) - E(Y|D) \tag{3.2}$$

where $E(Y|D)$ is the entropy of Y given the information achieved by decision D, and $E(Y)$ entropy of Y. Additionally, we define a gain ratio as a modification of information gain that takes into account the number of branches created by splitting the data on a particular feature. It prevents the algorithm from favoring features with a large number of possible values. For every

parent (branch), the decision tree calculates the information gain by splitting on some feature and retains decisions that enlarge it. Features in this context are the descriptive observables measured by IceCube and described in Chapter 4.

Once the decision tree is constructed, it can be used to make predictions on new, unseen data by traversing the tree from the root node to a leaf node that corresponds to the predicted output value. The decision tree algorithm can also be used to prune the tree to prevent overfitting, when the tree is too complex and fits the training data too closely, resulting in poor generalization. Decision trees are well known to be prone to this issue and have therefore often been expanded to a more complex model of Random Forest where the building units are decision trees.

3.1.2 Random Forest

A Random Forest is composed of multiple decision trees that are trained on different subsets of the data and with different subsets of the input features. The decision trees in a Random Forest are constructed using a technique called bagging. Bagging involves randomly sampling the data with replacement to create multiple subsets of the data, and then training a decision tree on each subset. The output value of a Random Forest is determined by averaging the output values of the individual decision trees, with the main goal of reducing overfitting or impure decision trees built from imbalanced datasets.

The procedure is as follows. Bagging or bootstrapping is done first, resulting in several subsets being created. Then, random subsets of features, always slightly lower in size than the full set, are created for each node with the aim of diversifying the decisions. Decision trees are constructed, as described in the previous section, using the random subsets of both data and features. Events from the data sample are passed through decisions, and separated into leaves. Finally, the average classification of all trees is taken as the result.

Random Forests have the obvious advantage over decision trees in their reduction of overfitting. By introducing randomness, they are also more robust to outliers. As the algorithm is more complex, it can also handle complex datasets with a tremendous amount of input or variables. However, the disadvantage to decision trees is the reduced ability of interpretation. Due to the simplicity of trees alone, they can be interpreted by looking into the decisions they make along the way. This becomes inconceivable with the rising intricacy of forests.

Forests are built based on several parameters chosen, some of the most important including the number of trees, maximum depth of branches, impurity evaluation, and minimum number of elements in a decision. In general and in this work, the Random Forests used are evaluated with common metrics to determine the optimal set of parameters. Accuracy is the number of correct classifications per all. However, it is not appropriate for imbalanced datasets. Precision is the number of correct classifications per the total events of that class, which effectively scales it to

the size of a given class in that dataset and handles imbalance, and it is given with

$$P = \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i} \tag{3.3}$$

where TP and FP are true positives and false positives. Furthermore, recall is given with

$$R = \sum_{i=1}^{N} \frac{TP_i}{TP_i + FN_i} \tag{3.4}$$

with FN being false negatives, and is also robust to imbalance. Finally, we can define the f-metric

$$F = \frac{2PR}{P + R} \tag{3.5}$$

as the harmonic mean of precision and recall, and use it to evaluate the performance of different forests.

3.2 Goodness of fit

Depending on the problem of interest, the quality of reconstruction can be evaluated with several tests. It is important to use tests appropriate for discrete distributions, such as the unfolded flux resulting from this analysis.

A goodness-of-fit test is a statistical procedure used to assess whether a sample of data fits a particular theoretical probability distribution. The test compares the observed data to the expected frequencies or probabilities predicted by the theoretical distribution. In the final stage of this work, the observed data is compared to the current theoretical model.

However, the goodness-of-fit is often evaluated during the optimization process. Known distributions, simulated as described in Chapter 4, are passed through the unfolding process, and the outcome is compared to true values by a statistical test. This ensures choosing optimal parameters for the analysis chain by seeking the statistically most significant parameter set.

The most commonly used goodness-of-fit test is the Chi-square test. The test calculates the Chi-square statistic by comparing the observed frequencies in different categories with the expected frequencies predicted by the theoretical distribution. The larger the discrepancy between observed and expected values, the larger the Chi-square statistic will be. If the calculated Chi-square value exceeds a critical value determined by the desired level of significance and degrees of freedom, it indicates that the observed data significantly deviates from the expected

distribution. The Chi-square statistic [Wue11] of a set of N observations O is given with

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{\sigma_i^2} \tag{3.6}$$

where E are the expected values, and σ the variance. The square root corresponds to the standard regression error. In a generalized case, the variance is substituted with expectations E. To compare the statistics among multiple observations, degrees of freedom have to be accounted for. Therefore, the reduced Chi-square is introduced as

$$\chi_r^2 = \frac{\chi^2}{\nu} \tag{3.7}$$

where $\nu = (n-1) \cdot (m-1)$ is given from the dimensions of observations. The degrees of freedom for this work are determined by the number of bins to be $\nu = 13$. Dimension of the target is constant, and therefore testing different unfolded fluxes compared to their true value is straightforward using either Chi-square or reduced Chi-square.

Given some null hypothesis being tested against the alternative, the p-value measures the strength of evidence against the null assumption. It is the probability of obtaining the given Chi-square or larger, under this assumption. It is important to note that the p-value does not provide the probability for the observed effect (a flux of a certain shape, in the case of this work), but rather the probability of the hypothesis being true or false. It is often accompanied by a chosen significance level α, which is the limit value of probability to reject the assumption. A p-value smaller than the significance

$$p_\chi < \alpha \tag{3.8}$$

points to strong evidence of the null hypothesis being false. However, in the case of multiple hypotheses, like in the case of testing different parameter sets in this work, it is sufficient to compare p-values of different sets under the same assumptions.

Between a set of observations and expected values, the higher the p-value, and the lower the Chi-square statistic, the agreement between the two is stronger. In the extreme case of observations fully aligning with the theoretical predictions, the p-value would be 1.

There are many common and simple measures to assess the difference between two sets of discrete values, but they rarely consider the statistical variance of observations which is a crucial part of statistical testing. These measures are useful in quick checking of alignments between data, especially histogrammed, and are not restricted to the comparison of observations to expected values or theoretical models but have a vast application in the whole process given here. For a set of size N, we can define the Manhattan distance for evaluating the added closest

distances between two histograms,

$$D_M = \sum_{i=1}^{N} |x_i - y_i| \tag{3.9}$$

for comparing sets x and y. The Chebyshev distance gives the maximum difference in any direction between two vectors

$$D_{Ch} = max_i(|x_i - y_i|). \tag{3.10}$$

and in this work, it has been scaled with the vector x_i to express the distance as a ratio of the expected value.

The lower the evaluated distance among a set of histograms, the better the alignment. The previously given metrics do not take into consideration the population sizes inside bins i and can lead to large distances being driven by small uncertainties in vastly populated bins. A more proficient measure is the Wasserstein distance adjusted to discrete problems, in which case it is referred to as Earth's Mover Distance (EMD). The EMD is a measure of work needed to equalize two distributions [Ngu11] and is given with the minimal amount of work needed to match sets x and y normalized by the value of the lower input

$$D_{EMD} = \frac{min(Work(F, x, y))}{min(x, y)} \tag{3.11}$$

where work is the distance covered by any possible flow F,

$$Work(F, x, y) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} \cdot d_{ij} \tag{3.12}$$

d_{ij} being the starting distance between points x_i and y_j and f_{ij} the weight transferred from larger to lower input during their equalization. Although seemingly complex, EMD is readily available for usage in Python packages and is a strong tool for evaluating the differences in imbalanced datasets.

Chapter 4

Simulation and Observations

A mandatory item in understanding any detector is a proper simulation of events and the response they trigger. In a Monte Carlo simulation, a large number of random samples or iterations are generated to simulate the uncertain elements of a problem. Each sample represents a possible outcome, and the simulation calculates the results based on these samples. By repeating the process many times, the simulation generates a distribution of possible outcomes imitating some given probability distribution of the problem of interest.

4.1 Air Shower modelling

CORSIKA (COsmic Ray SImulations for KAscade) is a Monte Carlo simulation generating air showers initiated by interactions of cosmic rays [HKC+98]. CORSIKA propagates primary particles up to the atmosphere considering the probability distributions of their interactions and decay, described in Section 2.2.1. Simulations consider many possible influences on particles' trajectories and their usage is adjustable, some of which include the impact of Eath's magnetic field, Moliere scattering, Coulomb scattering, and Ionization energy loss in CORSIKA. Due to propagation and energy loss, information on all particles' positions and energies have to be constantly updated. For this work, the propagation of pions and kaons, and consecutively muons, is relevant, while resulting electromagnetic showers are of less importance. However, simulation of light from all interactions is necessary for a successful determination of background.

The hadronic interactions can be described by many different interaction models [DLRF17], including VENUS (Very Energetic NUclear Scattering), QGSJET (Quark Gluon String model with JETs), DPMJET (Dual Parton Model with JETs) and SYBILL. SYBILL [REF+20] has been given an extension with consideration of charm interactions, and therefore includes prompt particles. The newest version is referred to as SYBILL2.3c.

In SYBILL, when two hadrons collide, they break apart into smaller components called quarks and diquarks, that form color triplets or anti-triplets. The differently colored components

of the hadrons combine to form two colored strings, which then undergo fragmentation. In collisions between hadrons and nuclei, the number of target nucleons determines the number of strings, with each target splitting into two components. Secondary particles produced by SIBYLL decay into known particles, but only certain nucleons, antinucleons, charged pions, and all types of kaons can be treated as projectiles. Other particles are tracked but only undergo decay without further interaction. This model is especially relevant as it is considered to be the current most appropriate model in describing interactions leading to neutrinos and has been used to construct the simulation used for training in this work.

The current standard primary flux is modeled based on fits combined with the theoretical Hillas model, called the Hillas-Gaisser primary model with the newest version being referred to as H4a [Gai12]. It was developed separately for protons and neutrons with special attention to muon ratios and pions and kaons production, relevant for the resulting muon neutrino flux. In this work, it was tested alongside other models including the previous version H3a [Gai12], Zatsepin-Sokolskaya model [ZS06], and Gaisser-Stanev-Tilav [GST13], visualized in Figure 4.1.

Solving cascade equations of air showers is conveniently implemented in MCEq (Matrix Cascade Equations) [FEG+15]. This algorithm numerically solves equations for lepton fluxes and gives flux probabilities in discrete energy bins for various scenarios. Aside from the primary model and the hadronic interaction model, it considers the position on Earth and the time of year, due to differences in densities and temperatures of the atmosphere. Most importantly, the lepton flux density is constructed depending on the zenith angle, which is principally important in the making of this work, as described in Section 2.2.2. Different models at different zenith angles are shown in Figure 4.2. The densities can be applied to any simulation of neutrinos propagated in IceCube for the desired model. Throughout this work, a combination of primary model H4a and interaction model SYBILL2.3c at the South pole in different zenith angles has been used to reweight the densities of simulated neutrinos.

After the simulation of the primary particles in the air shower and the construction of their respective fluxes, the resulting neutrinos have to be propagated through ice and bedrock.

4.2 Neutrino propagation

In the same manner as air shower simulations, neutrinos are propagated by randomly sampling known distributions regarding their interactions. Neutrino propagation is clearly experiment-dependent, as neutrinos are propagated through Earth and ice corresponding to the IceCube experiment. These are called *nugen* and can also be tweaked depending on many parameters. After an interaction has been created following everything described in Section 2.2, the response of IceCube DOMs to this particular particle is generated. A reconstruction of the event like the one

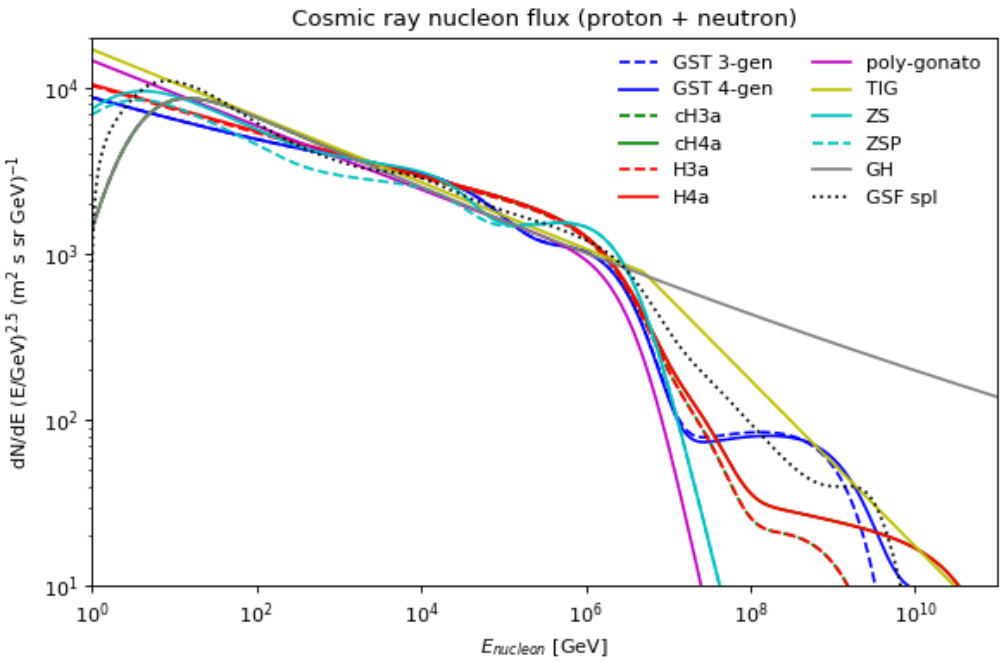

Figure 4.1: Cosmic ray fluxes of nucleons in different primary models. Label are as follows, GST 3-gen and GST 4-gen for 3-generational and 4-generational Gaisser-Stanev-Tilav [GST13], H3a, H4a, cH3a and cH4a for Hillas-Gaisser variations [Gai12], default and pamela for Zatsepin-Sokolskaya [ZS06], GSF for Spline fits [DEF+17], poly-gonato for PolyGonato [Hö03] and GH for Gaisser-Honda combination. Further explanations are available in the cosmic ray documentation [Doc]. The solid red line shows the primary model used in this work.

shown in Figure 2.10 is developed from the properties of the simulated neutrino, as if the event has been truly observed. The events are described in hundreds of features relating to geometry, energy, and direction, and stored with information about preceding events causing its creation, including the primary particles created with air shower simulations. The detector response used in this work is a *nugen* set [GK05] from 2016, constructed with the IC86 architecture, with a total of 9698407 events. Neutrinos are simulated with a E^{-2} spectrum, which does not correspond to the expected flux described in Section 2.2.1. However, the events are reweighted to proper flux using MCEq densities. The angular coverage of this dataset is 0° to 180°, and neutrinos are simulated with energies of 100 GeV up to 100 PeV. The absorption and scattering coefficients are 0, and DOM efficiency is 100% in this dataset, which corresponds to *baseline* values. These parameters are closely described in the following section.

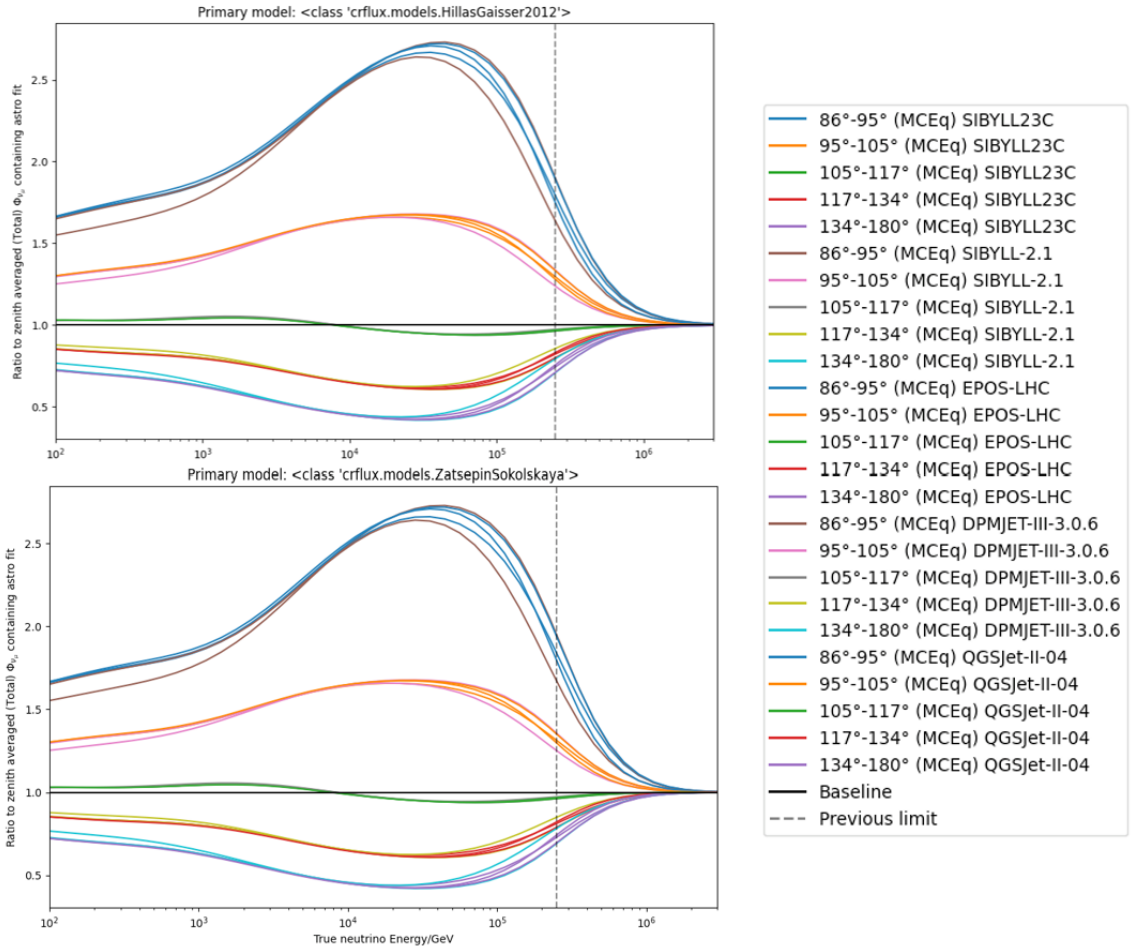

Figure 4.2: Angular dependence of the neutrino flux under different assumptions. The figure shows the same dependency as Figure 2.4 with different primary and hadronic models. The top figure shows the H4a primary model [Gai12] and the bottom the Zatsepin-Sokolskaya [ZS06]. Colored lines show different zenith bands. No significant differences are observed when varying primary models, as their flux is less influential to the resulting angular neutrino distribution. Different interaction models show slight influence in varying zenith bands, but the variations are deemed insignificant due to their size being lower than the statistical uncertainty of flux unfolding.

4.2.1 Systematic uncertainties

There are underlying differences in the shape and clarity of ice throughout the kilometer cube volume in IceCube. Additionally, the thousands of DOMs cannot be expected to perform all at the same level. To mitigate the effects of these differences, an additional error is added to the unfolded flux (or any other measurement done). These are jointly referred to as systematic uncertainties.

When exposed to photons, DOMs exhibit a linear accumulation of charge until they reach saturation. Therefore, the efficiency can be modeled with a linear function up to a certain energy. Simulations are used to calibrate their efficiency and are made so that the response of DOMs in simulations matches the one from real measured events. Although DOM efficiency closely relates to the quantum efficiency of the encompassed PMT, it is not the same, and the efficiency is an IceCube-related measure for scaling Monte Carlo to resemble the environment in ice. Simulations are made with varied DOM efficiency, and there are sets available of 90%, 95%, 100%, 105%, and 110%. Due to PMTs having a lower limit of detected photons for triggering a new event, different efficiencies can impact the number of events that are measured in the volume.

Another set of important systematic uncertainties is the ice absorption and ice scattering coefficients. As in the case of DOM efficiency, ice absorption doesn't have a one-to-one correspondence to the known absorption coefficient, but it is rather a scaling value for simulations. Different ice coefficients are especially relevant since different layers of ice exhibit different properties, some with lower clarity due to dustiness or freezing speed. Both coefficients are varied in values of 95%, 100%, and 105%.

The last varied parameter is the hole ice, which describes the difference between the old glacial ice and the refrozen ice made of water resulting from IceCube drilling.

To quantify the effects of these systematics, Monte Carlo simulations with different values are generated. Then, each of the sets is used as a training sample in unfolding. Resulting unfolded fluxes will imminently have slight differences in their shape due to the algorithm learning to unfold particles from a different environment in ice. When fluxes are finally unfolded by using different systematic sets, they are compared, and the absolute relative difference in flux per bin is taken as the systematic error. The final flux is unfolded using the already mentioned *baseline* values, but an additional error bar in the size of the attained relative difference is added, which covers the potential variance that can arise from ice and DOM properties. Details on the specific values and the resulting influence are shown in Appendix A.5.

4.3 Observations

As described in Section 2.2.3, a dataset consisting of through-going muon tracks is constructed. For reasons already stated, the main component of observations in IceCube, atmospheric muons, has to be cleared out.

Track-like events exhibit very poor energy resolution, as it is not known how much energy has a muon already deposited before reaching IceCube. However, the directional resolution is great, and it is important both for sample cleaning in this step, and later for angular studies. This is mainly due to the very small angle between the primary neutrino and the observed muon,

resulting from the relativistic boost of high-energy neutrinos [Sch13a].

In the sample, dominating events are atmospheric muons. However, at angles of over 85° in zenith, the muons have to cross kilometers of rock or ice and are therefore mainly absorbed. Simply looking at high enough zenith angles clears the dataset of the dominating background. This is, however, not sufficient since several down-going muon tracks can be falsely reconstructed as up-going in the process of event reconstruction. Several proficient steps of event selection need to be done to properly separate the background and enable background-free unfolding.

Event selection is a tedious task and is highly analysis-dependent. Events are selected in a series of steps called *levels*. For collected light to be registered as an event, at least eight DOMs have to be simultaneously triggered in a time window of less than five microseconds. If this limit is passed, hits from four microseconds before and six after are stored with the same event. This constitutes a level 1 event that is further processed as described in the following manner.

4.3.1 Level 2 processing

The hits are translated into waveforms, and the first angular reconstructions start already at the South Pole, to lower the amount of data to be transferred by a limited bandwidth. This processing is low-level and mostly focuses on the preparation of information transfer.

4.3.2 Level 3 processing

In this step, the sky is divided into two large regions. The Northern Sky is everything above the zenith angle of 85° degrees, and the Southern Sky is everything below. Analyses often carry the name corresponding to the sky part of interest. As stated, atmospheric muons cannot traverse the Earth at these angles, but there still exists a substantial background due to these events being misreconstructed into the region of the Northern Sky. Therefore, the level 3 processing focuses on handling faulty directional reconstruction.

Since reconstruction algorithms assume a single primary particle causing the track-like signatures (or cascade-like in other analyses), a filter by the name *HiveSplitter* is applied and separates events cleaning out ones with a higher possibility of multiple primaries. The boolean logic

$$(((\frac{l_{dir}}{180m})^2 + (\frac{n_{dir}}{10})^2) \;\; AND \;\; n_{dir} > 6) \;\; OR \;\; \frac{\alpha_r}{n_{dof}} < 9 \;\; OR \;\; \alpha_r \cdot n_{dof} > 7.5 \tag{4.1}$$

corresponds to level 3. Here, it is mandated that the length of track from first to last triggered DOM l_{dir} or the number of DOMs triggered by at least one photon n_{dir} is high, or that the reduced likelihood α_r associated with some degrees of freedom n_{dof} is larger than a certain limit. This cut solely reduced the domination of atmospheric muons to neutrino-created muons to three orders of magnitude from the starting ten-fold, or higher, ratio.

Some misreconstructed events remain in the sample, and a higher-level cleaning process is needed.

4.3.3 Level 4 processing

Selections done in level 4 are always highly adapted to the goal of the analysis, and the steps presented here are the current approach to event selection for analyses looking at the diffuse sky, especially in the field of flux reconstruction [Rä17, A$^+$22].

To further separate muon events originating from astrophysical neutrinos with high precision, the decision tree algorithm, described in Section 3.1.1, is employed. Events that have been simulated with CORSIKA are labeled as background, and simulated neutrino events from *nugen* are labeled as signal. The signal includes exclusively charged current events with a reconstruction error lower than five degrees, to further mitigate the misreconstruction effects, and are weighted to an E^{-2} spectrum to avoid bias of the denser populated low energy region. The sample is split into ten mutually exclusive parts, and are used for training and testing in rotating sets.

With labeled data, the decision tree is trained to produce the probability of events belonging to the signal based on several features, of which the most important are the already introduced l_{dir}, n_{dir}, α_r, and additional topological features. The resulting score distribution is shown in Figure 4.3.

Only events with a signal probability of higher than 90% are kept. Additional hard cuts are introduced, keeping specifically events with

$$n_{DOM} > 12,\ n_{dir} > 6,\ l_{dir} > 200m,\ l_{empty} < 400m,\ cos\theta^2_{geo} < 0.2 \tag{4.2}$$

where, in addition to already introduced features, l_{empty} is the largest distance between hit positions at the track, and θ_{geo} is the angle between the track and a plane that geometrically separates the track into parts with equal n_{dir}.

A possibility of misclassifying cascade-like events as high-energy neutrino-induced events remains, due to high energy deposited and their poor angular resolution, and another decision tree is trained with the goal of rejecting neutral current events.

The tree is trained in a similar manner to the one previously described, however, the background sample from CORSIKA now consists of cascade events induced by electron neutrinos interacting with nucleons. This model performs substantially worse, and the tree is often not able to distinguish cascade events, as seen in Figure 4.4. However, the cascade rejection is a useful additional cleaning step and has been shown to boost the performance of the overall selection process. All events with a cascade score higher than 50% are removed from the sample.

To determine the competence of a neutrino sample in reconstructing the flux, the most rele-

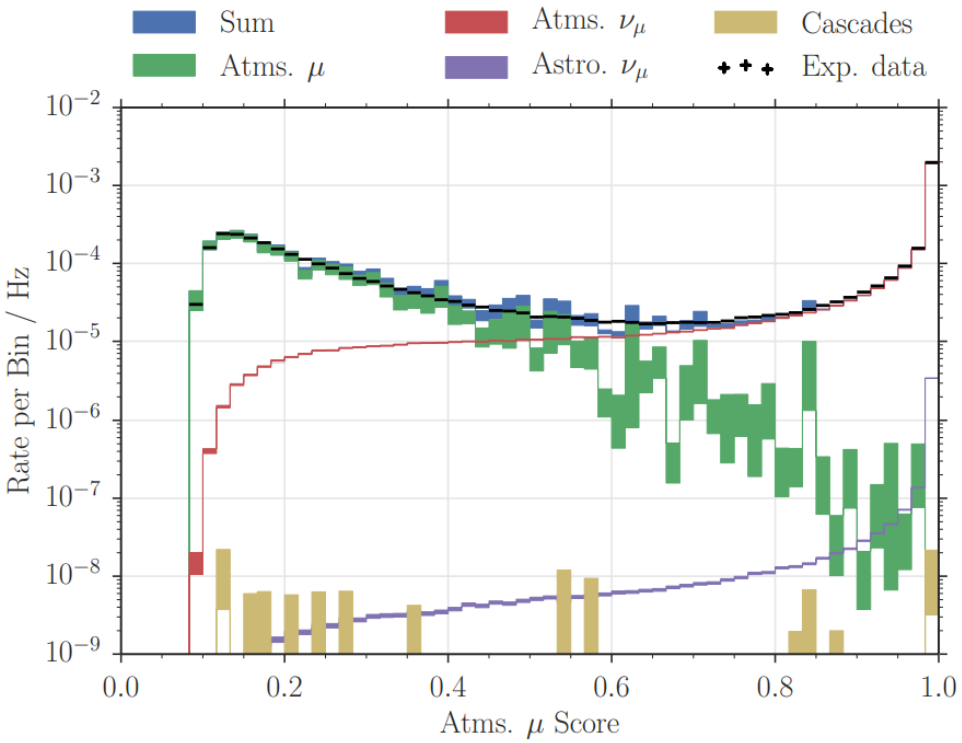

Figure 4.3: Score distribution of the decision tree trained to separate astrophysical neutrino induced events from muon neutrinos [Rä17]. The score corresponds to probability of the decision tree classifying a single event as the signal. The atmospheric muons depicted in green have a visibly descending distribution towards high signal scores. Events of interest are depicted in purple, and have a steadily rising distribution of scores between values of 16% and 100%.

vant measure is the purity of the sample

$$purity = \frac{n_\nu}{n_\nu + n_\mu} \tag{4.3}$$

which gives an approximate expectation of the contribution of neutrino-induced events to the full set of events used in the analysis. Processing event samples created in IceCube up to level 4 by previously described methods results in a purity of 99.87%, meaning that any set will have at most a 0.13% contribution of atmospheric muons, and is a milestone value in event selection. Due to their nature, the atmospheric muon background would be most severe at low energies and around the horizon.

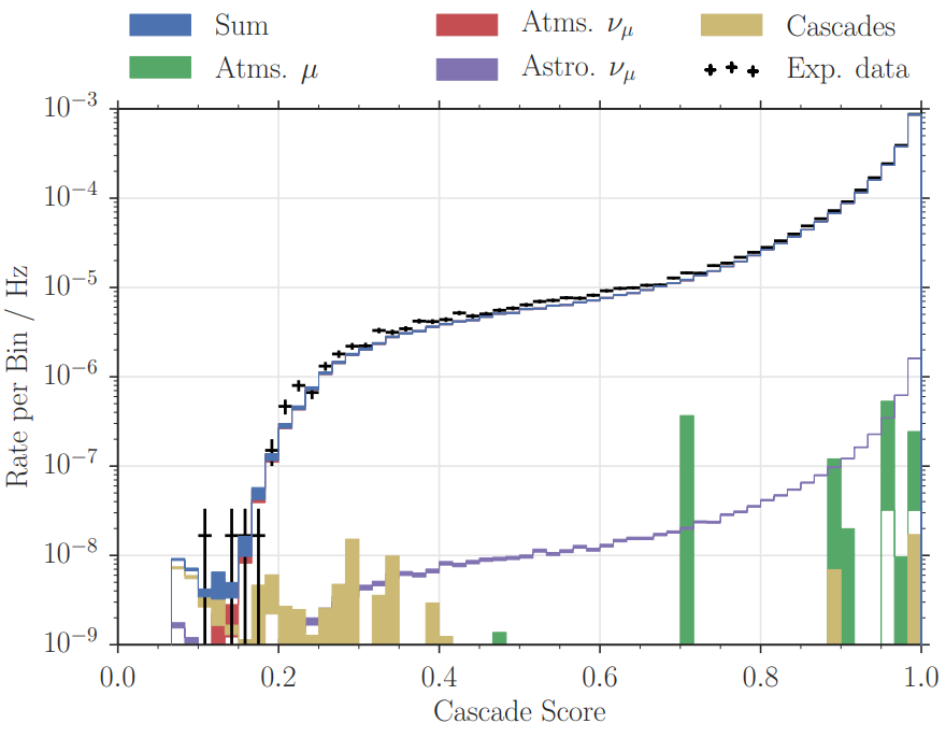

Figure 4.4: Score distribution of the decision tree trained to separate cascade-like from track-like neutrino induced events [Rä17]. The score corresponds to the probability of the decision tree classifying a single event as a signal. The cascades depicted in yellow accumulate at low scores, pointing to a deficient performance of classification.

4.3.4 Level 5 processing

Level 5 is the final level selection, and is adjusted to the specific analysis done, like the one presented in this work. Level 5 can be considered the further selection of events in the energy range of interest, which is from 500 GeV up to 4 PeV. A slightly higher than conventional low energy limit is used to avoid the problematic region of low energy events on the horizon.

Furthermore, the sample is used in full, and also split into angular regions, to explore the dependence described in Section 2.2.2. In this work, a total of eleven years of data is used, starting in 2011 and ending in 2021, and processed to the fifth level. Data was taken with the IC86 configuration, meaning all of the 86 strings holding DOMs in IceCube were active. The total lifetime is 341132384.3 seconds or 10 years, 298 days, and 7 hours. The total number of events expected in this time frame is approximately 850000, and a subset of 10% of data called *Burnsample* has been used to develop several tests. The event rate after full selection is 1.5 mHz.

The energy resolution σ_E of reconstructed events is estimated at 0.3 for 100 TeV energies.

The angular resolution of given energy is 0.25° degrees, a rather remarkable result for an event set spanning a total of 95° degrees. The active detection time in IceCube is separated into time windows called *Runs* usually lasting from several minutes to eight hours, each of which contains separate *Events*.

Chapter 5

Unfolding

In IceCube and other experiments, the neutrino energy is not measured directly but has to be determined from related measures. The many different features of the reconstructed event are related to the produced muon energy, and consecutively to the primary neutrino energy.

Unfolding, or deconvolution, is the process of determining the function in inverse problems [Spa13]. The sought-after quantity x, in this case the neutrino energy, is smeared with the detector response matrix A, which represents all the stochastic processes happening during the detection. The measured observables are given with y.

$$g(y) = \int A(x, y) \cdot f(x)dx + b(y) \tag{5.1}$$

When these functions are discretized, the problem translates to a linear model of the form

$$\vec{g} = A_{m,n} \cdot \vec{f} \tag{5.2}$$

in the special cases where the background $b(y)$ can be neglected. Here, the background in the dataset has been handled in a multi-step process described in Chapter 4.

The migration function A, or detector response, is a matrix of size m, n where m is the dimension of g and n is the number of bins in the target space f. It indicates the conditional probability to measure that a given true quantity x is measured with y. A has to map the complete detection process, from the propagation and interactions of primary neutrinos and emerged leptons to the whole detector response described in measured features.

The condition number of a matrix is a measure of how sensitive the solution of a linear system of equations is to small changes in the coefficients of the system [Din17]. In other words, the condition number of a matrix A measures how much the solution to a general case $Ax = y$ changes when the entries of A and y are perturbed by small amounts. It is defined as the

product of the norm of the matrix and the norm of its inverse:

$$cond(A) = \|A\| \left\|A^{-1}\right\| \tag{5.3}$$

where $\|.\|$ is any matrix norm, such as the Euclidean norm or the maximum absolute row sum norm. We refer to problems containing a matrix with a high condition number as ill-conditioned. This leads to numerical instability and inaccurate results when solving the system of equations using numerical methods. Conversely, a matrix with a low condition number is well-conditioned and small perturbations in the input will not significantly affect the output. Therefore, the condition of the matrix is proportional to the uncertainties arising in the solution.

In unfolding, the matrix is ordinarily ill-conditioned due to the size and complexity of the input. The problem cannot be solved by simply looking for the inverse matrix, as this usually results in values with large errors. The migration matrix can be determined empirically from the simulation (Monte Carlo) set and subsequently divided by the energy density function of the training dataset to make it independent of the training energy distribution. The searched spectrum is then estimated with the Maximum likelihood estimation where all bins are assumed Poissonian. The likelihood for each bin for some expected value λ is given with

$$\alpha(\vec{g}|\vec{f}) = \prod_{u=1}^{m} \frac{\lambda_u^{g_u}}{g_u!} \cdot exp(-\lambda_u) \tag{5.4}$$

where we defined the expectancy for λ in Equation 5.2, leading to

$$\alpha(\vec{g}|\vec{f}) = \prod_{u=1}^{m} \frac{(A\vec{f})^{g_u}}{g_u!} \cdot exp(-(A\vec{f})_u) \tag{5.5}$$

For simplicity, the logarithm of the likelihood function is searched for. This leads to an alternative form of the sought likelihood

$$\alpha(\vec{g}|\vec{f}) = \sum_{u=1}^{m} (g_u ln(A\vec{f})_u - (A\vec{f})_u) - ln(g_u!)) \tag{5.6}$$

in which the constant term $ln(g_u!)$ can be neglected. Due to the mentioned ill-conditioning of the matrix, the possibility of high variance has to be handled, therefore a regularization factor $R(\vec{f})$ has to be added to the expression.

5.1 Regularization

Regularization helps to stabilize the solution by adding a penalty term to the objective function, which prevents the solution from being too complex or overfitting to the data. Tikhonov regular-

ization [Kre89] is a specific form of regularization that is commonly used in inverse problems, where the goal is to reconstruct an unknown function from noisy or incomplete data. Tikhonov regularization, also known as Ridge Regression or L2 regularization, was developed by Andrey Tikhonov in the 1940s as a way to solve ill-posed problems in geophysics, however, its use has widened and in this work, it is adapted to the flux reconstruction problem. The basic idea behind Tikhonov regularization is to add a penalty term to the objective function that measures the complexity of the solution. The penalty term is a function of the parameters of the solution and is designed to favor solutions that are smooth or have small magnitudes. This encourages the solution to be as simple as possible, while still fitting the data reasonably well. Tikhonov regularization has several advantages over other regularization methods. First, it is simple and easy to implement. Second, it has a closed-form solution that can be computed efficiently using standard linear algebra routines. Third, it has a well-defined regularization parameter that can be tuned using cross-validation or other methods.

In the particular case of flux unfolding, regularization can be applied with the assumption of a smooth solution since the neutrino flux is expected to be a power-law. Although the power-law has expected slope changes, as explained in Section 2.2.1, the flux it describes should not exhibit any prominent breaks. This is a solid assumption that can be introduced to handling the solution and is the only limit imposed to control the bias that researchers' assumptions create in the results.

In smooth functions, a small second derivative is expected, therefore, a normal distribution centered around 0 is introduced. This way, the likelihood becomes a prior with expectation zero for the second derivative of the logarithm of the solution which favors power laws. For the covariance of this expression, a diagonal matrix with some regularization strength τ is introduced. This regularization strength can be adjusted and must be chosen so that unphysical solutions are suppressed without significant distortion of the results. Adding the normal distribution, Equation 5.6 expands to

$$\alpha(\vec{g}|\vec{f}) = \sum_{u=1}^{m}(g_u ln(A\vec{f})_u - (A\vec{f})_u)) - \frac{1}{2}\left[(C\vec{f})^T Diag(\mathbb{I}\cdot\tau)(C\vec{f})\right] \tag{5.7}$$

with C as the Thikonov matrix.

Certain ranges, and therefore bins, of the target distribution f can be empty. This would lead to a logarithm of zero in the given expression, which is not defined. For this reason, a small offset d is added to the event spectrum. Both the regularization strength and the offset are optimized through an iterative method explained in Chapter 6. Adding the offset, we arrive at

the full expression for the likelihood

$$\alpha(\vec{g}|\vec{f}) = \sum_{u=1}^{m}(g_u ln(A\vec{f})_u - (A\vec{f})_u)) - \frac{1}{2}\left[(C\vec{f} + \vec{d})^T Diag(\mathbb{I} \cdot \tau)(C\vec{f} + \vec{d})\right] \tag{5.8}$$

that is to be maximized.

No analytical solution can be found to maximize the likelihood for $\vec{f}$. To find the event spectrum $\vec{f}$ with a maximum probability to be causing the measured observables $\vec{g}$ in IceCube, sampling methods are deployed. Due to the complexity of the problem, a detailed sampling method has to be found.

5.2 MCMC sampling

Monte Carlo Markov Chain (MCMC) sampling is a computational technique used to generate samples from complex probability distributions that are difficult to directly sample from [vRCB16]. MCMC works by constructing a Markov chain whose stationary distribution is the target distribution that we want to sample from. The chain is constructed in such a way that it satisfies the detailed balance condition, which ensures that the chain converges to the stationary distribution regardless of the starting point.

The basic idea is to generate a sequence of random samples from the distribution of interest. At each step, the current state of the Markov chain is used to probabilistically determine the next state. This is done by proposing a new state based on the current state, and then accepting or rejecting the proposal based on a predefined probability. Here, this measure is the likelihood α defined with Equation 5.8 multiplied with the function's prior, as taken from Bayesian probability:

$$p_n^{MCMC}(\vec{f}|\vec{g}) \propto \alpha_n(\vec{g}|\vec{f})p(\vec{f}) \tag{5.9}$$

Each walker compares the likelihood of the proposed state to the likelihood of the current state and uses a transition probability to balance the trade-off between exploring the state space and converging to the target distribution. The resulting sequence of states generated by the Markov chain converges to the target distribution as the number of samples approaches infinity. The convergence rate depends on the mixing rate of the chain, which is determined by the proposal distribution used to generate new states, the number of samples, and the number of walkers employed.

Flux is expressed in terms of energy per unit area, per unit time, and per unit solid angle. Finally, obtaining a flux from some event spectrum f is by

$$\Phi_{\nu_{mu}} = \frac{f}{\Delta E \cdot 2\pi\Delta cos\theta \cdot t_l \cdot A_{eff}} \tag{5.10}$$

for considering the angular band $\Delta cos\theta = cos\theta_{min} - cos\theta_{max}$ and the energy range ΔE in some lifetime t_l. The flux is commonly measured in $GeV^{-1} \cdot s^{-1} \cdot sr^{-1} \cdot cm^{-2}$. Oftentimes, the flux is weighted with energy squared to visually amplify the differences in the slope of the components, and is in this case measured in $GeV \cdot s^{-1} \cdot sr^{-1} \cdot cm^{-2}$.

The effective area can be understood as the size of the detector that has the perfect efficiency in detecting neutrinos. It can be simply inferred from the ratio of detected neutrinos to the true number and is done also from the neutrino propagation simulations. Effective area is

$$A_{eff} = \frac{n_{obs}}{n_{sim}} \cdot (\rho N_A) \cdot \sigma(E_\nu) \tag{5.11}$$

with $\sigma(E_\nu)$ as the neutrino cross-section and N_A the nucleon density, and is energy and angle-dependent. n_{obs} is the number of neutrinos detected in the total of n_{sim}, not to be confused with n_{exp} that measures the amount of experimental events seen in real observations.

5.3 Unfolding in the context of flux reconstruction

One of the most desired features of any method aiming at reconstructing the neutrino flux is being unbiased. The main advantage of this unfolding method is that it uses training data to build the response matrix, but it does not depend on the weighting of the input. This means that the method is less sensitive to the uncertainties in the input data, which can lead to biases in other methods. Furthermore, being independent of flux assumption is preferred with the goal of getting the true reconstruction, even in the case of the neutrino flux being modeled or assumed wrong.

This unfolding method also does not depend on any assumed shape of the model, such as power-law or broken power-law, since the results are bin-wise points not described by a single function. This makes the method more robust and applicable in all cases. Methods that do assume a shape, can ever only fit functions to that particular shape, while in this case the results are self-governing.

However, unfolding methods also have some limitations. One of the main challenges is the need for a well-defined detector response. This is achieved by enabling a tremendous amount of data for the method to learn the behavior of the stochastic processes. Another limitation is the fine-tuning of regularization parameters, but nonetheless, given enough testing on varied samples, this limitation is surpassed.

5.4 Comparison to other methods

The early predecessor of this technique is the *RUN (Regularized UNfolding)* program, widely used since its development with the aim of unfolding neutrino cross-sections [Blo02]. *RUN*

relies on the same concept of discretizing the response from Equation 5.1 to a linear model from Equation 5.2, and optimizing the strength of introduced Thikonov regularization through a single parameter τ. It is also a model-independent bin-wise method, but with a limit of three for the number of observables used. Based on its workings, an advanced program was developed and published as *TRUEE (Time-dependent Regularized Unfolding for Economics and Engineering problems)*, introducing automated observable selection, parameter optimization, and the determination of input bias [Mil12]. It was used in IceCube for model-independent reconstructions producing quality results [A$^+$15b, Saa14], up to the introduction of here used, advanced, method.

DSEA+ (Dortmund Spectrum Estimation Algorithm) is an unfolding method [RVW$^+$19], unlike ones introduced before. It works by translating the task into a multinomial classification problem, in which it assigns a probability to each event of the sample. This enables *DSEA+* to retain information event-wise, a feature unavailable to unfolding methods building a Hessian matrix. The probability of specific events can be constructed using common machine learning libraries, giving this algorithm a special edge due to its simplicity of use. It can be considered an expansion to the familiar Bayesian Unfolding [D'A10], with several limitations resolved. During the development of this work, *DSEA+* was extensively tested and compared to the unfolding method used here. Due to a different formulation of the problem solution, it can be effectively used in problems with well-known models, or in studying variations [H$^+$21].

DSEA+ is mathematically similar to bin-wise likelihood unfolding [Bun18], often used in IceCube for flux reconstruction. Most commonly, IceCube has used model-dependent likelihood methods, in which the power-law assumption is fitted to the observed data. As mentioned, these methods have a compelling limitation from the introduced bias.

However, using likelihood, different flux models can be accounted for and compared. Utilizing a power-law assumed likelihood, a flux measurement with outstanding precision was recently published [A$^+$22].

In addition to the mentioned advantages of here used method in unfolding, other aspects of the analysis chain have been improved, with almost all preparatory and surrounding steps optimized and evaluated using machine learning methods described in Chapter 3.

Chapter 6

Analysis chain

This chapter presents all the steps done in the analysis and the intermediate results achieved. First, all the data has to be properly prepared. Following is the transformation of data, including feature selection and mandatory preprocessing. Then, bias tests are presented along with optimization of parameters entering unfolding. Finally, the algorithm is applied and results are shown for several different cases studied here.

6.1 Data preparation

To construct the training sample, the Monte Carlo simulation described in Chapter 4.2 is collected. As a result of neutrino propagation, they contain track-like events from CC interactions. The so-called $frames$ of data in IceCube files contain 1346 columns called features which describe each event. Some reconstruction algorithms repeatedly fail for specific simulations, therefore the first step is always handling missing information. Cleaning the frame from empty features reduces its size to 1263. Likewise to empty features, features with constant values through all events do not carry any information and their removal results in a frame of size 965.

When two or more features are highly correlated, they often provide redundant or repetitive information to the learning algorithm, which can lead to several issues. This can increase the computational complexity of the model and slow down training and inference processes, leading to longer processing times and higher resource requirements. Highly correlated features can introduce instability in the following manner; if the algorithm learns the strong linear relationship among these features, a slight change or noise in one of them can lead to the model's assumption of the variance in other features, too. Similarly, this can lead to undesired overfitting, by overtraining the model to the same information through several features seemingly of the same importance. The migration matrix A described in Chapter 5 is preferred to be of smaller dimensions, to reduce its condition number, which is additionally handled by regularization parameters described in Section 5.1. Therefore, to achieve stability and reduce dimensionality while saving

relevant information, highly correlated features are removed from the set. However, handling correlation is a tedious task and it will be done in two different ways. At this early stage of data preparation, a simple Pearson cut [FPP07] is applied, to reduce the dataset to a practical size. From a pair of any two variables x and y, the Pearson correlation coefficient is

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{6.1}$$

and attains a value between -1 and 1, depending on the strength of the relationship. In any pair with $|r_{xy}| > 0.95$, one of the features is removed. A cut of 0.95 is rather conservative, however, correlation will be dealt with in depth later. Applying this cut to the full frame reduces its size to only 272, which is not surprising knowing that many reconstruction algorithms aim to calculate the same features of an event.

Simulated events do not have any timestamps, as these events do not actually hold place. There are several features related to simulation production, which do not exist in frames of real observations. All the events of one simulation are also marked as one *Run*, information completely redundant for unfolding. Every feature of this nature is removed, and the final size of the dataset in this step is 247 features with 9698407 events.

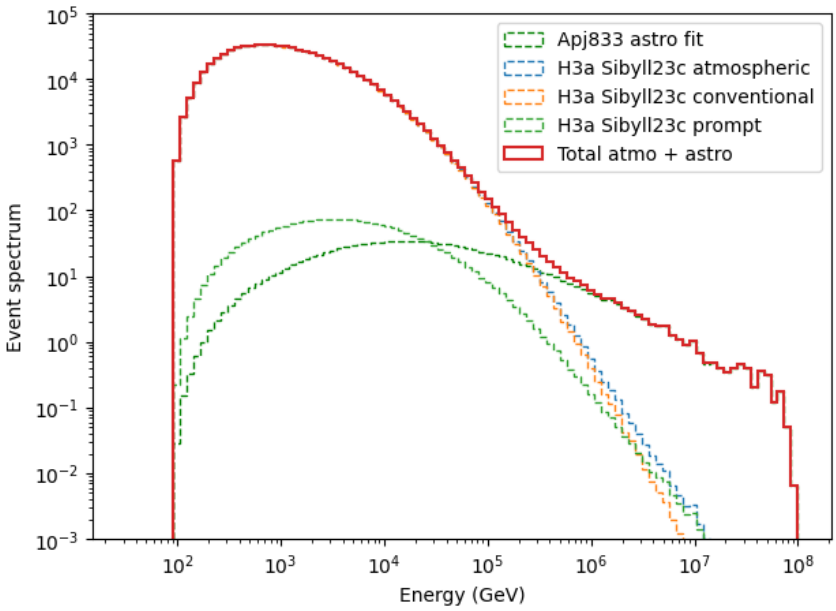

Figure 6.1: Expected event spectrum and its components. The conventional and prompt components are simulated with MCEq [FEG+15], while the astrophysical fit is taken from experimental results [A+16a]. Reweighting is applied to the whole Monte Carlo simulation, spanning from 100 GeV to 100 PeV in energy.

As mentioned, the simulations do not follow the shape of a realistic flux but are rather de-

Model	Type	Expected number of events
SYBILL2.3c	Prompt	1912.37
SYBILL2.3c	Conventional	653743.99
SYBILL2.3c	Atmospheric	655660.77
HESE	Astrophysical	8475.36
Apj833	Astrophysical	1147.25
SYBILL2.3c + Apj833	Total	656808.02

Table 6.1: Expected total event spectrum for the lifetime used in this analysis. Atmospheric components are simulated with MCEq, assuming a primary model H4a [Gai12] and interaction model SYBILL2.3c [REF+20]. The astrophysical event spectrum is taken from experimental results in IceCube, including the High Energy Starting Events analysis [A+21] and the six-year astrophysical fit [A+16a].

scribed by a single power law. Reweighting assigns a rate in seconds for each type of event so that they cumulatively give a flux described with density profiles given with MCEq. Here, the software is applied to produce a flux from a SYBILL interaction model with Hillas Gaisser primaries, more closely explained in Section 4.1.

With weights calculated, it is simple to arrive at an expected number of events during some time, since weights represent rates. From

$$E(n_{exp}) = \sum_{i=1}^{n_{sim}} w_i t_l \tag{6.2}$$

the expected number of events in experimental data n_{exp} during lifetime t_l of almost eleven years is 656808.02, of which 655660.77 is of atmospheric origin and 1147.25 of astrophysical, as seen in Table 6.1 and visualized in Figure 6.1. IceCube has in fact observed a total of over 850000 events pointing to an underestimation of current atmospheric models, but nonetheless, a ratio of atmospheric to astrophysical events can be expected around 571.

Poisson sampling is a method used to generate a sample of data based on event rates that follow a Poisson distribution. The Poisson distribution is commonly used to model the occurrence of rare events over a fixed interval of time. This is utilized to create *pseudosamples*, random sets that imitate the nature of the underlying distribution. Having varying sets from the same expected flux allows us to test consistencies of different assumptions while preserving the sta-

tistical properties of the underlying Poisson distribution, such as the average event rate and the random occurrence of events. Here, the weight represents the average number of events occurring during the given lifetime. A random number generator is used to determine the occurrence or non-occurrence of each event within the sampling interval. Pseudosamples with different random seeds following the same underlying distributions are visualized in Figure 6.2. As per the definition, *pseudosamples* from the same underlying distributions should create differences in the unfolded flux at the highest to the level of statistical error. The tests in this work were mainly performed using a pseudosample, which is the best possible approximation of what we should see in later unfolding, to avail the possibility of biasing oneself by inspecting the true measurement before developing the whole analysis chain.

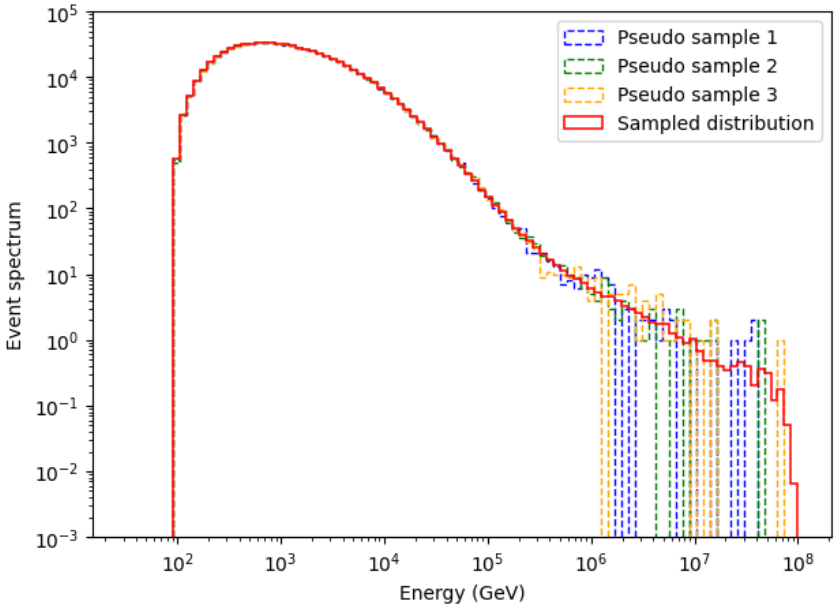

Figure 6.2: Creation of three pseudosamples to be used in testing and optimization during the analysis. The pseudosamples are based on the underlying distribution in red given by the atmospheric and astrophysical assumptions described in this and Section 4.1, with each event sampled in a randomized process. Due to higher populated low energy region, pseudosamples highly imitate the underlying distribution, however, the low statistic high energy region produces less or none events in pseudosamples, as expected.

Based on event rates, the sky is divided into five regions to attain approximately the same amount of statistics in each angular bin. Each angular bin will contain approximately a fifth of all data. The most populated horizon bin will therefore be shortest in size. The five bins are later used for unfolding the flux, and were used to simulate expectancies shown in Figures 2.2, 2.3, 2.4, and 4.2. Figure 6.3 depicts the resulting angular bin edges which are 1.5, 1.65, 1.83, 2.04,

2.33, and 3.14 in radians, translating to approximately 86°, 95°, 105°, 117°, 134°, and 180° in degrees.

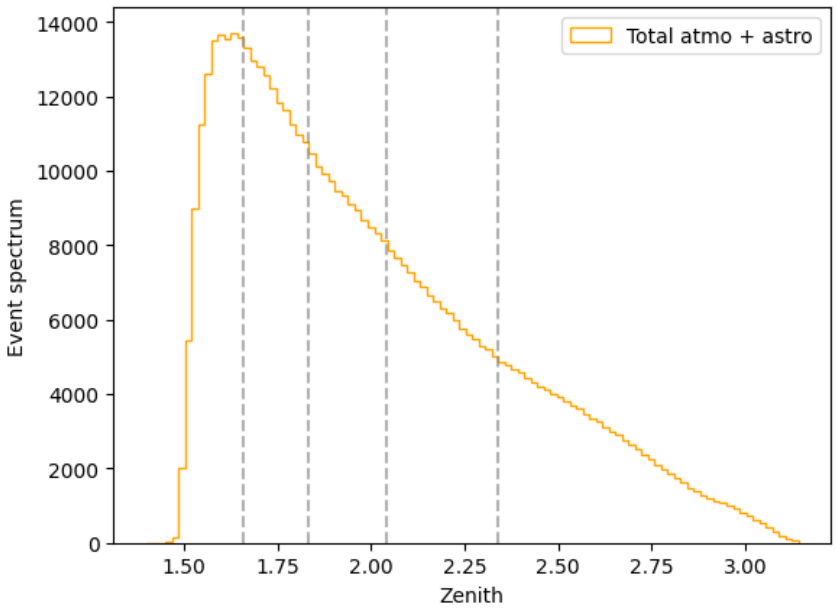

Figure 6.3: Simulated distribution of events per each angular bin, assuming a primary model H4a [Gai12], interaction model SYBILL2.3c [REF+20] and the six-year astrophysical fit [A+16a]. Each of the five regions between dashed gray lines contains approximately a fifth of the total 850000 events. The zenith is given in radians.

6.2 Effective area

Effective area, given in equation 5.11, is calculated solely from simulation data, and can therefore be retrieved at this early stage of the chain.

Each event results from a certain type of interaction and is simulated based on the probability of this interaction W_{int}. To simplify further steps, a measure referred to as *OneWeight* is assigned to each event

$$w_{one} = \left(\frac{W_{int}}{E^{-\gamma}}\right) \int^{\Delta E} E^{-\gamma} dE \cdot Area \cdot \Omega \tag{6.3}$$

where the simulation is generated in Area for energies E in the solid angle Ω. γ is the spectral index from Equation 2.1 for the flux used in simulation and described in Section 4.2. The weight of each event i following some parametrized flux Φ_ν is then

$$w_i = \frac{w_{one}}{n_{sim}} \cdot \Phi_\nu \tag{6.4}$$

Note that equation 6.4 can be used to simply reweight any known simulation, however only in case of a power-law flux Φ_ν as defined in Equation 2.1. This can be used to reweight various experimental astrophysical fits parametrized with a normalization constant and a spectral index. In contrast, atmospheric components are described with complex functions far from a simple power-law and have to be simulated with MCEq as presented earlier. Furthermore, the effective area is then calculated for each energy as

$$A_{eff} = \frac{w_{one}}{E_\nu \cdot \frac{10 \cdot log(10)}{m_{eff}}} \tag{6.5}$$

where m_{eff} is the number of bins in which the effective area is calculated. For the whole energy range, which will later be separated into 13 energy bins, the effective area is calculated in 52 energy bins equidistant in logarithmic space. Therefore, for each energy bin, an average from 4 values is taken. The effective area is calculated separately for different sets given in Section 4.2 corresponding to angular ranges and is given in Table 6.2 and in Figure 6.4. As expected, the effective area at high energies is best for horizontal events and is in comparison deteriorated for very high angles.

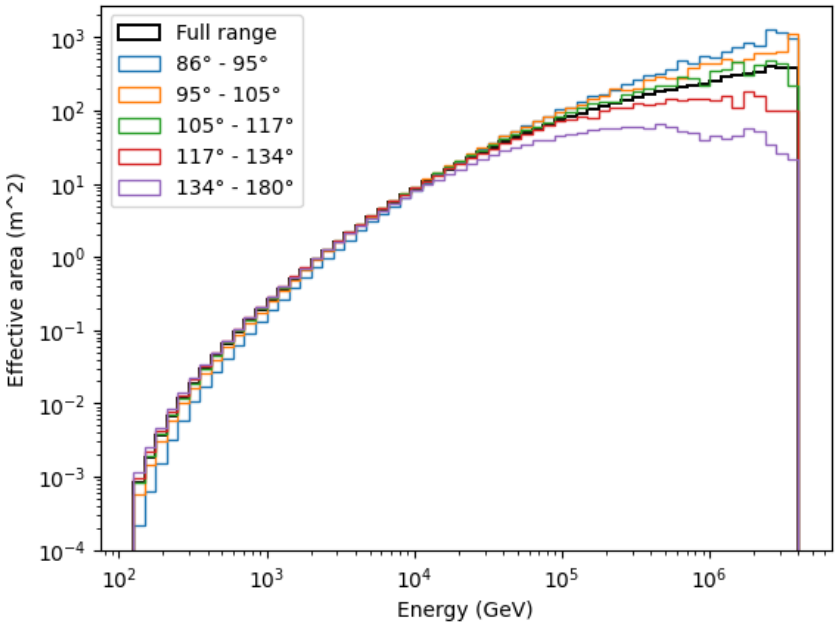

Figure 6.4: Effective area in squared meters for the inspected energy range. The colored lines show the effective area for the five angular bins used in unfolding.

After simulation reading, it is necessary to read the measurements. As discussed in Section 4.3.4, the studies are first run on only 10% of the full observations called the *Burnsample* data, and later redone on the full set. From subsets of observations called *Runs*, IceCube generates the

Energy range [GeV]	A_{eff} [m^2] ($86° - 180°$)	A_{eff} [m^2] ($86° - 95°$)	A_{eff} [m^2] ($95° - 105°$)	A_{eff} [m^2] ($105° - 117°$)	A_{eff} [m^2] ($117° - 134°$)	A_{eff} [m^2] ($134° - 180°$)
$10^{2.7} - 10^{3.0}$	$1.167 \cdot 10^{-1}$	$7.943 \cdot 10^{-2}$	$1.091 \cdot 10^{-1}$	$1.220 \cdot 10^{-1}$	$1.284 \cdot 10^{-1}$	$1.275 \cdot 10^{-1}$
$10^{3.0} - 10^{3.3}$	$4.381 \cdot 10^{-1}$	$3.260 \cdot 10^{-1}$	$4.246 \cdot 10^{-1}$	$4.624 \cdot 10^{-1}$	$4.747 \cdot 10^{-1}$	$4.589 \cdot 10^{-1}$
$10^{3.3} - 10^{3.6}$	$1.411 \cdot 10^{0}$	$1.155 \cdot 10^{0}$	$1.425 \cdot 10^{0}$	$1.492 \cdot 10^{0}$	$1.495 \cdot 10^{0}$	$1.416 \cdot 10^{0}$
$10^{3.6} - 10^{3.9}$	$3.918 \cdot 10^{0}$	$3.488 \cdot 10^{0}$	$4.145 \cdot 10^{0}$	$4.165 \cdot 10^{0}$	$4.070 \cdot 10^{0}$	$3.735 \cdot 10^{0}$
$10^{3.9} - 10^{4.2}$	$9.524 \cdot 10^{0}$	$9.247 \cdot 10^{0}$	$1.028 \cdot 10^{1}$	$1.017 \cdot 10^{1}$	$9.764 \cdot 10^{0}$	$8.632 \cdot 10^{0}$
$10^{4.2} - 10^{4.5}$	$2.046 \cdot 10^{1}$	$2.208 \cdot 10^{1}$	$2.311 \cdot 10^{1}$	$2.214 \cdot 10^{1}$	$2.048 \cdot 10^{1}$	$1.706 \cdot 10^{1}$
$10^{4.5} - 10^{4.8}$	$3.905 \cdot 10^{1}$	$4.618 \cdot 10^{1}$	$4.671 \cdot 10^{1}$	$4.289 \cdot 10^{1}$	$3.847 \cdot 10^{1}$	$2.907 \cdot 10^{1}$
$10^{4.8} - 10^{5.1}$	$6.726 \cdot 10^{1}$	$9.002 \cdot 10^{1}$	$8.577 \cdot 10^{1}$	$7.460 \cdot 10^{1}$	$6.460 \cdot 10^{1}$	$4.254 \cdot 10^{1}$
$10^{5.1} - 10^{5.4}$	$1.070 \cdot 10^{2}$	$1.694 \cdot 10^{2}$	$1.451 \cdot 10^{2}$	$1.233 \cdot 10^{2}$	$9.486 \cdot 10^{1}$	$5.262 \cdot 10^{1}$
$10^{5.4} - 10^{5.7}$	$1.576 \cdot 10^{2}$	$2.866 \cdot 10^{2}$	$2.365 \cdot 10^{2}$	$1.808 \cdot 10^{2}$	$1.223 \cdot 10^{2}$	$6.011 \cdot 10^{1}$
$10^{5.7} - 10^{6.0}$	$2.156 \cdot 10^{2}$	$4.505 \cdot 10^{2}$	$3.539 \cdot 10^{2}$	$2.346 \cdot 10^{2}$	$1.508 \cdot 10^{2}$	$5.599 \cdot 10^{1}$
$10^{6.0} - 10^{6.3}$	$2.805 \cdot 10^{2}$	$7.083 \cdot 10^{2}$	$4.915 \cdot 10^{2}$	$2.813 \cdot 10^{2}$	$1.472 \cdot 10^{2}$	$4.684 \cdot 10^{1}$
$10^{6.3} - 10^{6.6}$	$3.516 \cdot 10^{2}$	$1.005 \cdot 10^{3}$	$6.555 \cdot 10^{2}$	$3.389 \cdot 10^{2}$	$1.204 \cdot 10^{2}$	$3.546 \cdot 10^{1}$

Table 6.2: Effective areas per energy and in different angular regions. The area is calculated in four slices of each energy bin, with their average entering the final flux calculation from Equation 5.10.

GoodRun list, which contains all *Runs* in a given lifetime that satisfy the minimum requirements for its events to be used in further analyses. The *GoodRun* selection includes the lower limit of active DOMs at any point in time, a minimum amount of detector running with no known issues, and similar. After cleaning the sample from bad runs, the level 5 observations are read and ready for following studies.

6.3 Feature selection

There are several properties a feature set should satisfy to be used in unfolding. From the currently available 247 features, only a few will be used to limit the size and instability of the migration matrix.

Conditions for the feature set can be roughly divided into three categories. First, it is necessary the set contains features with the best available reconstruction of true values. Several algorithms reconstruct the same observables in a different manner and can be compared based on their ability to reconstruct a known simulated distribution. Secondly, features in simulation should resemble the same features observed in real data, as it would otherwise point to a misunderstanding or misconstruction of the physical processes involved. Lastly, out of all the available features, it is desired to retain those containing the most information needed for energy estimation, as energy estimation is the basic component of flux reconstruction. In conclusion, features should have three properties: proper reconstruction of values, alignment to experimental data, and information gain in unfolding.

6.3.1 Reconstruction algorithms

Unfolding the energy flux necessarily needs an energy estimator, and in addition, the study of angular dependence calls for a proper zenith estimator. Values from which the target is inferred are often called *proxies*. Although many observables in IceCube are correlated with primary energy, including the length of track, number of hits, affected DOMS, and similar, a variable estimating the starting energy is desired.

Based on the underlying equations and fits, or the boolean logic they use, many different algorithms exist. Example directional reconstructors are *LineFit*, *BestTrack*, and *MPE*. *BestTrackDirectHits* and *LineFitGeoSplit* focus on counting responses in DOMs, for example. For energy estimation, all types of *TruncatedEnergy* reconstructors are commonly used, along with *SplineMPEMuExDifferential*. Algorithms belonging to the same groups aim to reconstruct the same set of values. Therefore, all directional algorithms will have a zenith estimation available.

In this step of feature selection, the power of simulations is yet again used. For some simulated set of neutrinos, the true zenith distribution is known. Neutrinos are propagated as in

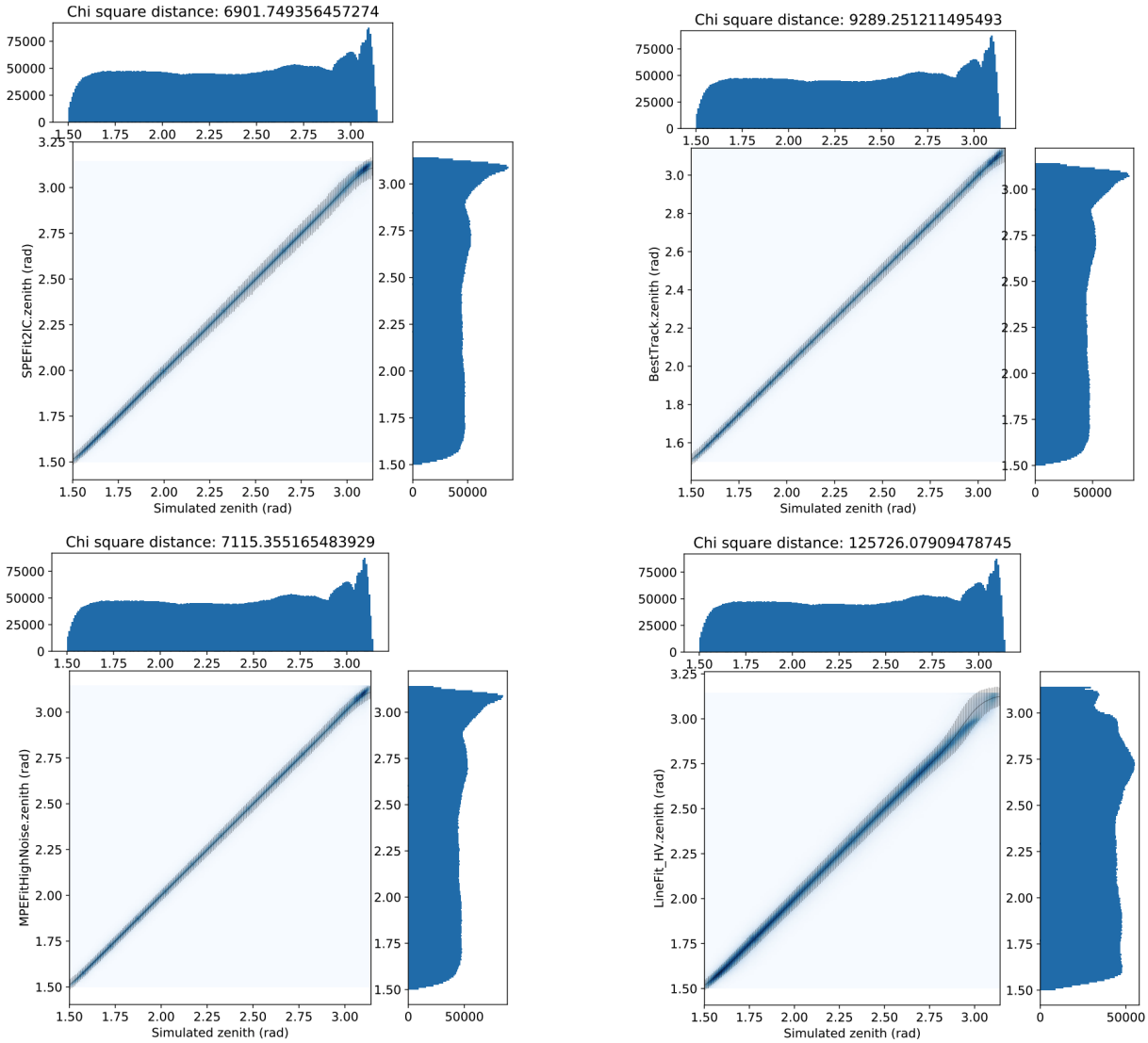

Figure 6.5: Testing of directional reconstruction algorithms. The top histogram on each plot shows the true simulated zenith distribution. On the right, the reconstructed value of each algorithm is depicted. The middle plots show the alignment of the reconstructed to the true value, with gray lines showing the spread of agreement. Here, tests of four reconstructors are showcased, which attained the following Chi-square statistics: SPEFit2IC with $\chi^2 = 7020.6$ (upper left), BestTrack with $\chi^2 = 9428.7$ (upper right), MPEFitHighNoise with $\chi^2 = 7222.29$ (lower left), and LineFit_HV with $\chi^2 = 127758.0$ (lower right). The SPEFit2IC exhibited the best performance among all available direction reconstructors.

Section 4.2, and the constructed response is compared to true value. To quantify the quality of agreement between reconstructed and real values, the Chi-square statistic is calculated for each

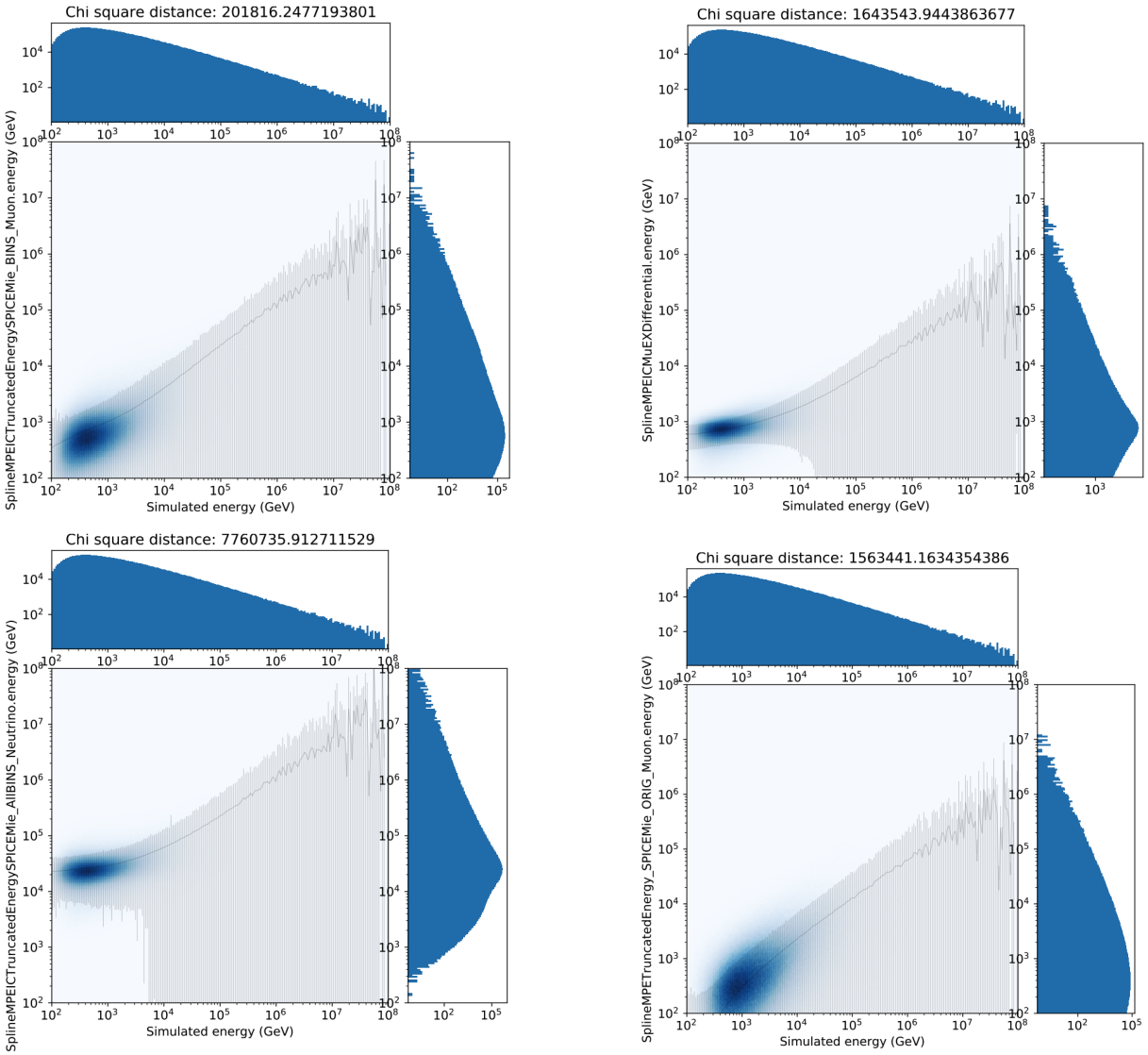

Figure 6.6: Testing of energy reconstruction algorithms. The top histogram on each plot shows the true simulated energy distribution. On the right, the reconstructed value of each algorithm is depicted. The middle plots show the alignment of the reconstructed to the true value, with gray lines showing the spread of agreement. Here, tests of four reconstructors are showcased, which attained the following Chi-square statistics: SplineMPEICTruncatedEnergySPICEMie_BINS_Muon with χ^2 = 204975.1 *(upper left), SplineMPEICMuEXDifferential with* χ^2 = 1669570.6 *(upper right), SplineMPEICTruncatedEnergySPICEMie AllBINSNeutrino with* χ^2 = 7885030.7 *(lower left), and SplineMPETruncatedEnergy_SPICEMie _ORIG_Muon with* χ^2 = 1558268.6 *(lower right). The SplineMPEICTruncatedEnergySPICEMie_BINS_Muon exhibited the best performance among all available energy reconstructors.*

available algorithm. A modified form of the Equation 3.6 is used, in which the variance equals the expected value. Here, E is the simulated, and O the reconstructed distribution. Comparisons are shown in Figure 6.5.

An equivalent approach is applied to every available energy reconstruction algorithm, with results shown in Figure 6.6. As explained in Section 4.3, track-like events have poor energy resolution compared to angular resolution. Therefore, Chi-test results are strongly inferior in this group of reconstructors. Nonetheless, a type of *TruncatedEnergy* showed upper performance compared to others. Finally, based on statistical agreement to true values, *SPEFit2IC.zenith* and *SplineMPEICTruncatedEnergySPICEMie_BINS_Muon.energy* are added to the feature set.

6.3.2 Data to Monte Carlo agreement

If the features of the simulation weighted to some assumed models are distributing similarly to distribution inside real data, it is considered the simulation is properly working, and the underlying assumptions are true. Customarily in IceCube, data to Monte Carlo agreement is done by binning distributions into histograms and visually inspecting if any substantial deviations exist. As part of the standard process, these comparisons will be done later, but first a more elaborate approach is done.

Instead of merely checking the agreement between the overall distributions, a general Random Forest algorithm described in Section 3.1 is employed. This necessity comes from the fact that features may look similar in histograms, although following different underlying distributions or correlating to different variables. This is especially true if the choice of binning is random, and not properly adjusted to reveal distributions' characteristics. This can be easily understood with a simple example; if the uniform function is defined as

$$f(x) = \begin{cases} 1, & |x| < 1; \\ 0 & \text{otherwise.} \end{cases} \tag{6.6}$$

and the triangular functions is defined as

$$f(x) = \begin{cases} 1 - |x - c|, & |x| < 1, |c| < 1; \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

with an arbitrary constant c, the rather dissimilar functions will look the same binned to a histogram of too narrow binning. If only visual inspection of histograms would be applied to variables with these underlying functions, their equivalence could be falsely claimed. The example is explained in Figure 6.7. This is especially relevant since many features represent $flags$, boolean values separating events into types based on some criteria. For example, if there exists

a *flag* separating events into low energy and high energy based on the energy compared to some critical value, a balanced dataset would result in opposite values per each event but identical histograms. This is more detailed in Appendix Figure A.13.

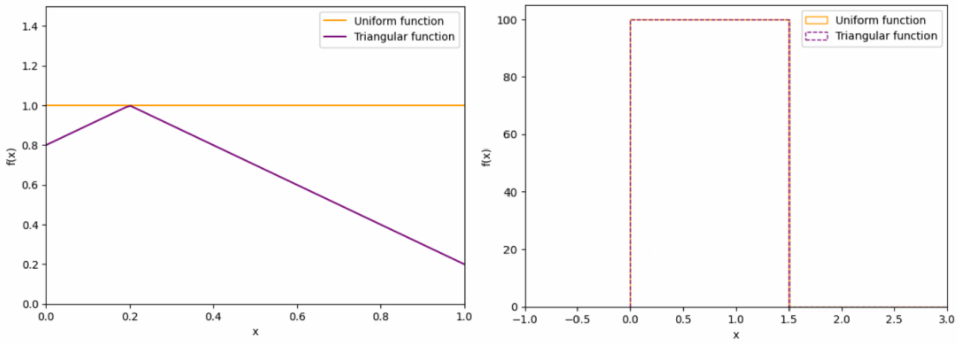

Figure 6.7: Comparison of plots of uniform and triangular functions (left) and their respective histograms (right). The triangular function is defined with a constant $c = 0.2$ in this example. Functions are shown for a hundred x divided linearly between 0 and 1. On the right, both histograms of values resulting from two functions are shown, and they fully overlap. Since the binning of the histogram is too narrow, the differences between resulting values are engulfed by the width of one bin, and the distributions seem identical.

An additional complication comes from features depending on other features, whose relationship is not well understood. To mitigate the presented problem, the alignment is tested by using a Random Forest, because of its ability to construct decisions based on underlying functions not attainable to analysts. Both real observations and simulations are used in this step. Events from the *Burnsample* are labeled with value 1 (for being of type Signal) and Monte Carlo events get assigned a value of 0. The Random Forest model is built with the goal of separating events into these two categories, with the target value to be determined being this label, and not the energy, as in the main part of the analysis. Being able to separate events into data and Monte Carlo is not desirable, but the model is built this way to expose properties that contribute to their separation.

The main premise is that there exists an underlying difference in features if the model can differentiate among two sets when this feature is present. The model is specifically trained and tuned to separate data from simulation with the best possible performance, and then the effects of features in this discrimination are interpreted. To quantify an effect a feature has on some model's decisions, the change of purity described in Section 3.1.1 is investigated. As noted, purity is calculated for each decision tree constituting the Random Forest. When a feature is added to the set, the purity of each decision tree is calculated and averaged out over the whole forest, and the process is repeated for each inspected feature. The averaged purities of all features

are normalized to one so that it holds

$$\sum_{j=1}^{m} I_j = 1 \tag{6.8}$$

where I_j is referred to as *feature importance* in a set of tested features of size m. It follows that in a set of features approximately equally contributing to information gain, their importance would center around

$$I_j \approx \frac{1}{m}. \tag{6.9}$$

After the model is trained and the feature importances are evaluated, the Random Forest is applied to an unseen testing subset of data with known labels. Evaluating the probability assigns a value between 0 and 1 for each event. If the underlying differences between two sets (data and simulation) truly exist, the Random Forest will be able to assign probability scores of over 0.5 for events with a known label of 1, and vice versa. If no differences occur, the model will randomly guess each event belonging to any class due to the absence of proper information needed for discrimination, and the probability scores should center around 0.5.

For an imperfect feature set in which discrimination is possible, a feature contributes substantially if its importance is much higher than the mean, as given in Equation 6.9. Calculated importances are binned into a histogram, and any feature from a set I with a deviation from the median importance higher than a limit so it holds

$$I_j > Median(I) + 2 \cdot MAD \tag{6.10}$$

is removed. The MAD (Median Absolute Deviation) is defined similarly to the standard deviation

$$MAD = Median(|I_j - Median(I)|) \tag{6.11}$$

with the Median being the value of which exactly half the set is larger.

After all importances are evaluated and every feature crossing the critical value from Equation 6.10 is removed, the process is repeated for the newly attained, smaller, feature set. This process is repeated until a set is constructed so that no feature importance satisfies Equation 6.10. As a result of retraining the whole model with a new feature set in each epoch, the probability scores change and the limit applied gets progressively stricter. Migration of the probability score towards the desired value of 0.5 and of importances to Equation 6.9 is expected with the advance of the iterations.

With each iteration, the frequency of importances is varied, therefore features that may have not satisfied the limit in early iterations would be removed in a later set. The more iterations a feature passes without crossing the critical value, the less it contributes to discrimination. Therefore, a feature's alignment between data and simulation is correlated with the iteration in

which it was removed. From this, it follows that features can be scored based on this iterative procedure.

The procedure applied to the cleaned set described in Section 6.1 ran for 8 iterations, until no more features were marked for removal. On the final set, the model is applied once more with the goal of predicting the label, and it attains a precision of only 0.513, near the ideal value of 0.5. Prediction scores of all events center tightly around 0.5 as well. The procedure is visualized in Figures 6.8, 6.9, 6.10, 6.11, with the resulting iteration and score shown in Figures 6.12 and 6.13.

The algorithm has been tested on various weightings of atmospheric models. Slight differences exist in early iterations, however, the final chosen set is independent of weighting. Although weighting is seemingly important in reshaping the Monte Carlo to best imitate real data, this result stems from the fact that the best-aligned features are always of a geometrical nature. Straightforward measures like position, sphericity, length, and similar, are always better reconstructed in IceCube than inferred values relating to energy and zenith. Therefore, this algorithm prefers geometrical observations. The set resulting from this procedure is not the final feature set used in the analysis, but rather the scores of data to Monte Carlo alignment for each feature are stored, based on the last iteration it was included in. These scores are used in combination with the third, and final, feature selection algorithm described in the following section.

6.3.3 Recursive Feature selection

The final selection procedure relates to the choice of features most appropriate for energy reconstruction. Different types of algorithms either recursively add or eliminate features from given sets, and compare them based on some optimization metric.

RFECV (Recursive Feature Elimination with Cross-Validation) is a predeveloped algorithm from the *scikit-learn* package [PVG+11]. For the algorithm to evaluate the performance of a certain set, an underlying classifier has to determine energy. For this, a Random Forest from Section 3.1 is trained, however, in this case with a target of energy reconstruction, and not data type discrimination like in Section 6.3.2. Data is discretized in the same manner as it will be for the final unfolding, and the events are classified based on the outcome probability score into the same energy bins in the range of 500 Gev to 4 PeV, described in Section 4.3.4. For every size of set being tested, weighted precision is calculated, a modified version of metric given in Equation 3.3. In this case, precisions are multiplied by the size of their respective category, corresponding to the 13 energy bins in this model. This handles the imbalance of the dataset and is relevant due to lower statistics toward the higher end of energies.

Information gain imminently rises with each feature added, however, it is desired to also preserve low-dimensionality, as discussed in Section 6.1. The benefit of a small information gain from a less relevant feature is often outweighed by the deterioration coming from dimension

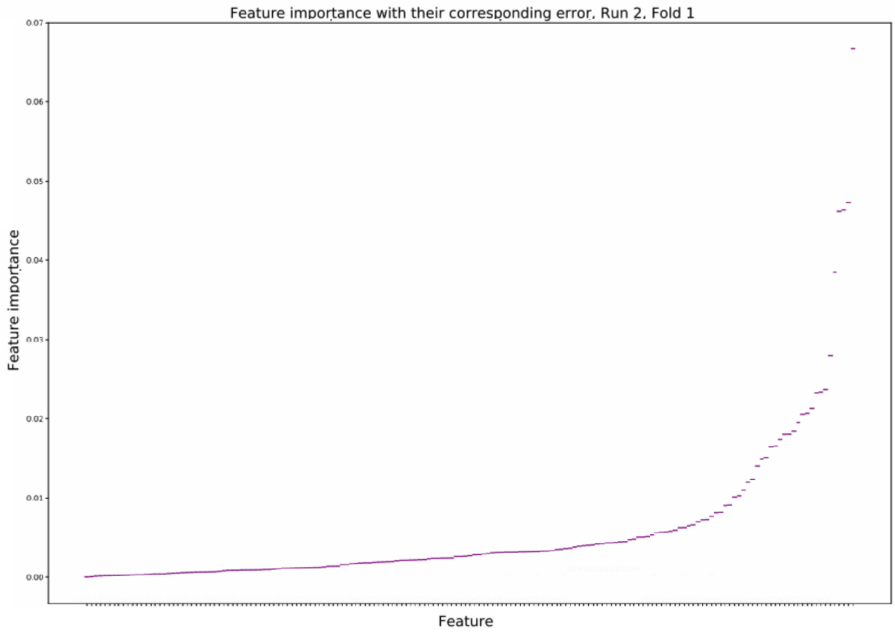

Figure 6.8: First step of the iterative feature selection, on the example of the second iteration with a feature set of size 169. After the Random Forest is fitted and importance is assigned to each feature, they are sorted and shown on the x-axis with their corresponding importance on the y-axis.

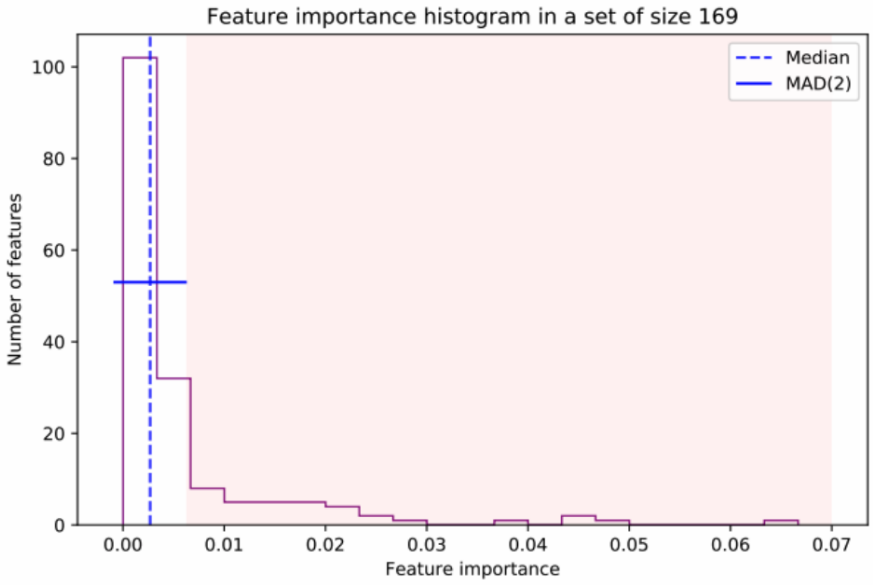

Figure 6.9: Second step of the iterative feature selection. The importance is binned into a histogram and the median value is calculated. The solid blue line shows a spread of two MADs around the Median. All features related to values larger than the limit, in the red-shaded area, ought to be removed.

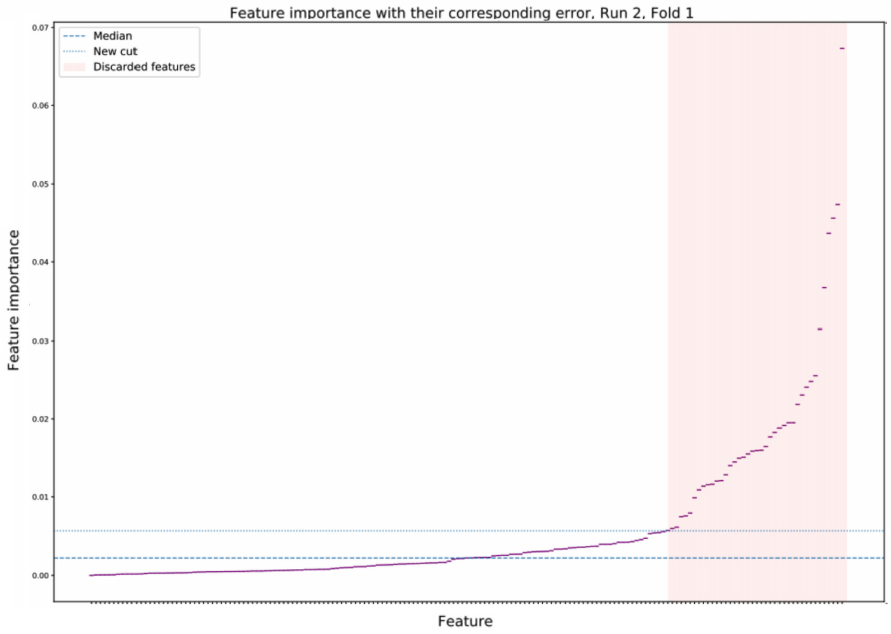

Figure 6.10: On features from Figure 6.8, a cut depicted in Figure 6.9 is applied. All features whose importance is higher than the dotted line are discarded.

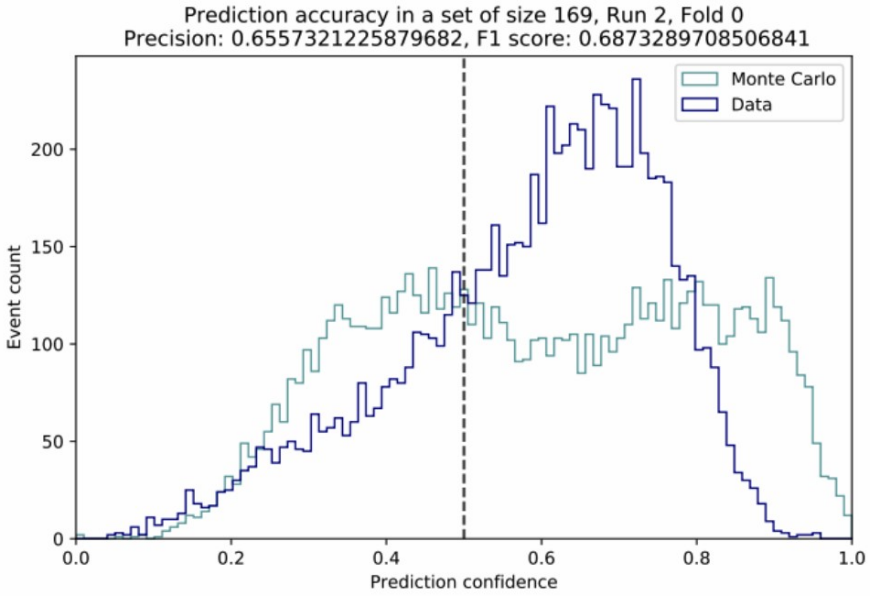

Figure 6.11: Probability scores assigned to events of each category. Although the separation is not optimal, the probabilities of signal events (dark blue) gravitate towards one. In the case of two indistinctive sets, all probabilities should center around the middle.

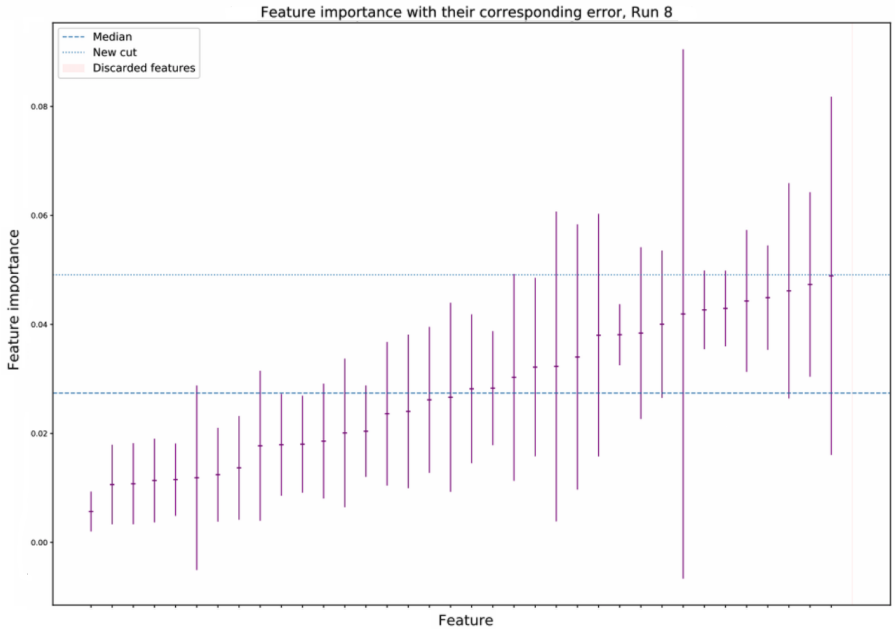

Figure 6.12: In the final iteration, no importance is crossing the limit. Values distribute more evenly around the Median.

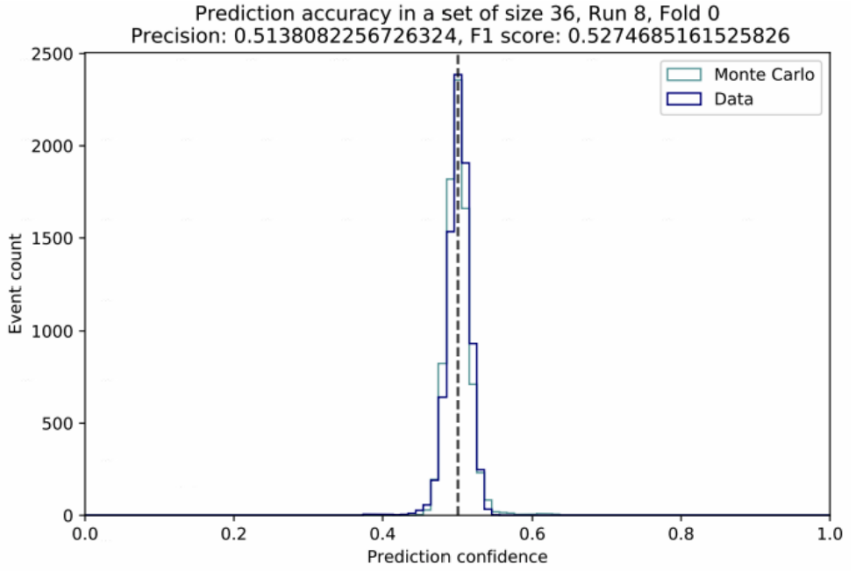

Figure 6.13: The final iteration results in 36 features that are indistinctive to the Random Forest, visible from events having probabilities with equivalent distributions in both categories. Scores centering around 0.5 are often referred to as good as guessing.

addition. The size of the final feature set is iteratively enlarged until sufficient information is reached that cannot be considerably improved with more data. This exhibits a saturation, or visually a plateau, in performance metrics per dimension added.

As shown in Figure 6.14, weighted precision rises up to a set size of 8. With size, the included features are optimized and are comprised of the best information providers in energy estimation.

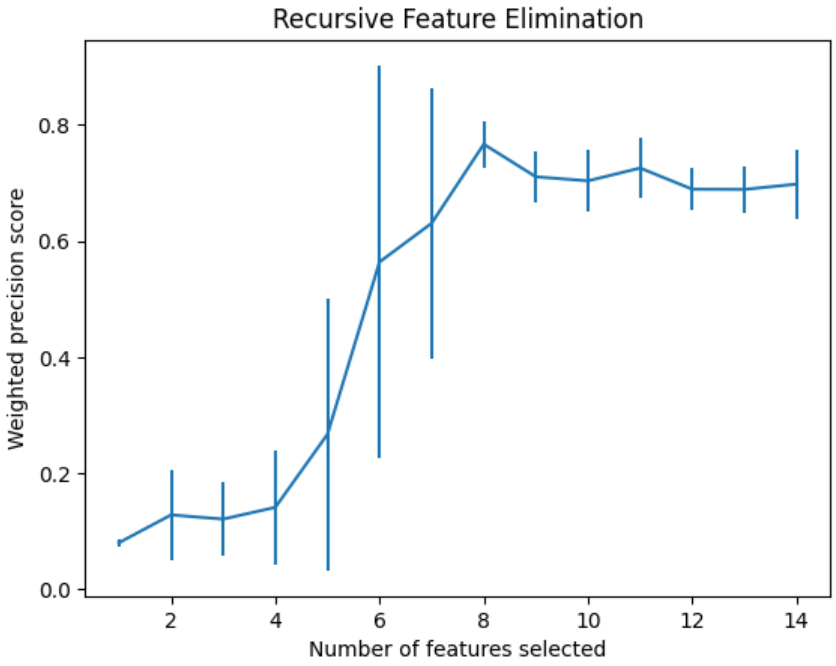

Figure 6.14: Weighted precision for different sizes of tested feature sets. Uncertainty comes from the standard deviation of precision on various sets examined. At eight features, the information gain is sufficient for energy reconstruction and precision reaches a maximum.

Eight features chosen are *energy*, *n_late_doms*, *n_dir_doms*, *avg_dom_dist_q_tot_dom*, *value*, *rlogl*, *dir_track_length* and *dir_track_hit_distribution_smoothness*, described in Table 6.3. These are reconstructed in various algorithms, and the choice of reconstructors is based on results from Section 6.3.2. From the given eight, features with the highest scores from the previous selection algorithm are retained. A zenith reconstruction is necessary and has been chosen together with energy reconstruction, in the first selection from Section 6.3.1. The features are described in Table 6.3 and their agreement to measured data is visualized in Appendix Figures A.16 and A.17.

Feature	**Description**
SPEFit2IC. zenith	Reconstructed zenith.
SplineMPETruncatedEnergy _SPICEMie _BINS _Muon. energy	Reconstructed neutrino energy in GeV.
BestTrackDirectHitsA. n _late _doms	The total number of DOMs, which have at least one late pulse. A late pulse has a time residual, that is after the direct hits time window.
SplineMPEDirectHitsICC. n _dir _doms	The total number of DOMs, which have at least one direct pulse. A direct rpulse has a time residual, that is inside the given direct hits time window.
BestTrackCharacteristics. avg _dom _dist _q _tot _dom	The average DOM distance from the track.
SplineMPEMuEXDifferential _r. value	Value derived from the specific reconstruction algorithm.
SPEFit2GeoSplit1FitParams. rlogl	Reduced negative log likelihood for this reconstruction algorithm.
SplineMPEDirectHitsB. dir _track _length	The length of the track in meters, which is defined as the distance along the track from the first hit DOM to the last hit DOM perpendicular to the track direction.
SplineMPEDirectHitsC. dir _track _hit _distribution _smoothness	The smoothness value, based on the direct hit DOMs: How uniformly distributed are the direct hit DOM projections onto the track.

Table 6.3: Final set of features used in the analysis, chosen in the presented three-way selection algorithm.

6.4 Rebinning the observable space

Classifiers and regressors make predictions for a certain quantity based on its relation to the magnitudes of different observables relevant to their reconstruction. In general, both classifiers and regressors split the observables in a certain number of bins spanning its range. The multi-dimensional observable space is usually divided into equidistant bins in linear space. This can distort the analysis for two reasons. First, with rising number of observables, the number of bins grows exponentially. Since the amount of data acquired is fixed, the statistics in the expansive number of bins must reduce. Secondly, bin edges are not appropriately chosen this way. They can either be too narrow (e.g. it is nonsensical to use bins narrower than the sensitivity of the detector), or they can be too wide and deteriorate the information gain. Additionally, the observable space is binned with no regard to the question in the analysis.

The target space in this analysis consists of 13 energy bins divided in logarithmic space. If similar features that span a large range of values are binned in linear space, the resulting bins towards the high-energy region will be wide and necessarily result in sparse statistics.

Furthermore, some features attain only certain values. For example, if a boolean $flag$ can attain values 0 and 1, a two-bin linear space is enough to describe it. If a classifier divides the maximum range into 11 equidistant bins, the bins spanning values from 0.1 to 0.9 will inevitably be empty.

For reasons stated above, an additional Decision Tree classifier is introduced with the goal of rebinning the observable space used in unfolding which takes into consideration the various features' distributions. We train a Decision Tree with the selected features with the target being the neutrino energy, however, it is not used to later actually estimate energy from some set. Decision trees optimize cuts in observable space when making a decision for classification. The cuts presented in Section 3.1.1 are then used as bin edges of the observable space. In the construction of the Decision Tree, the size of leaves can be mandated therefore enabling the choice of number of samples in each observed bin with the purpose of facilitating enough statistics in each. The leaf size can be either set up as a fraction of the data or as an absolute number of events.

The tree is set up with a lower limit of 20 events in each constructed bin and an upper limit on the number of bins at 2500.

As a result, events are populating observable bins taking into consideration the minimum amount of events. In the multidimensional observable space, each bin will contain at minimum the number mandated with the leaf size. This enables the classifier to attain enough information about an event while preventing the encounter of empty bins.

The produced bin cuts can implicate some properties of the distributions and can reveal physical relationships of observables to the energy. For example, the dominant processes for

energy loss of muons at high energies are *bremsstrahlung* and pair production. Since the lost energy is dependent on the starting energy, algorithms that reconstruct these losses, like $MuEXDifferential$, are more informative as proxies for primary energy estimation. In contrast, at low energies the main energy loss process is ionization. Ionization processes are independent of starting energy, therefore the energy loss is not relevant for reconstruction. Other features correlating to primary energy, like track length, become relevant in this range. To encompass these properties, it is desired to create more bins in regions in which features are more informative. For this example, more cuts are made in the track length observable at low energies. Moving with rising energy towards the *bremsstrahlung* dominated region, fewer cuts are made until track length is considered only binary. For these events, information is mainly retrieved from other features. For simple interpretation, shorter bins imply more information gain. The given example is shown in Figure 6.15.

6.5 Migration Matrix and the MCMC chain

All the necessary information for building the migration matrix A is ready at this stage of the analysis. Using the Monte Carlo simulated events, reweighted as in Section 6.1, the observable space $\vec{g_{sim}}$ is populated into 2500 bins defined with cuts from the Decision Tree built in previous Section 6.4. Finally, the event spectrum $\vec{f_{sim}}$ associated with $\vec{g_{sim}}$ is also calculated from the simulations. The simulated spectrum and its accompanying simulated response define the intermediate detection process described by A. For the training dataset, all components from Equation 5.2 are known. Additionally, the migration matrix is divided with the assumed spectrum to erase the dependency of the process on the chosen weighting.

The resulting matrix has a size of $2500 \cdot 15$, for 2500 observable bins and the 13 target bins with additional overflow and underflow bins. The condition equals $2.7336 \cdot 10^2$. This can be roughly interpreted as the ratio of output error to input error if the matrix was inversed. For example, if the assumptions made in the event spectrum and the observed space would have uncertainties of order 10^{-10}, the inversion would produce a solution with uncertainties approximately two orders of magnitude stronger, at 10^{-8}.

The measured events from real data are then also populated into the same observable space and construct the *real observable space* $\vec{g}$. Now, migration matrix A and $\vec{g}$ are known in the likelihood expression from Equation A.3 and it remains to find the optimal event spectrum $\vec{f}$ maximizing this likelihood.

To get the best estimate for $\vec{f}$, Markov Chain walkers are employed as described in Section 5.2. 1000 walkers are created each starting at a random position and with each step they calculate the likelihood from Equation 5.8 for some combination of regularization parameters τ and d. The difference between the next step and the current step is randomly sampled from a normal

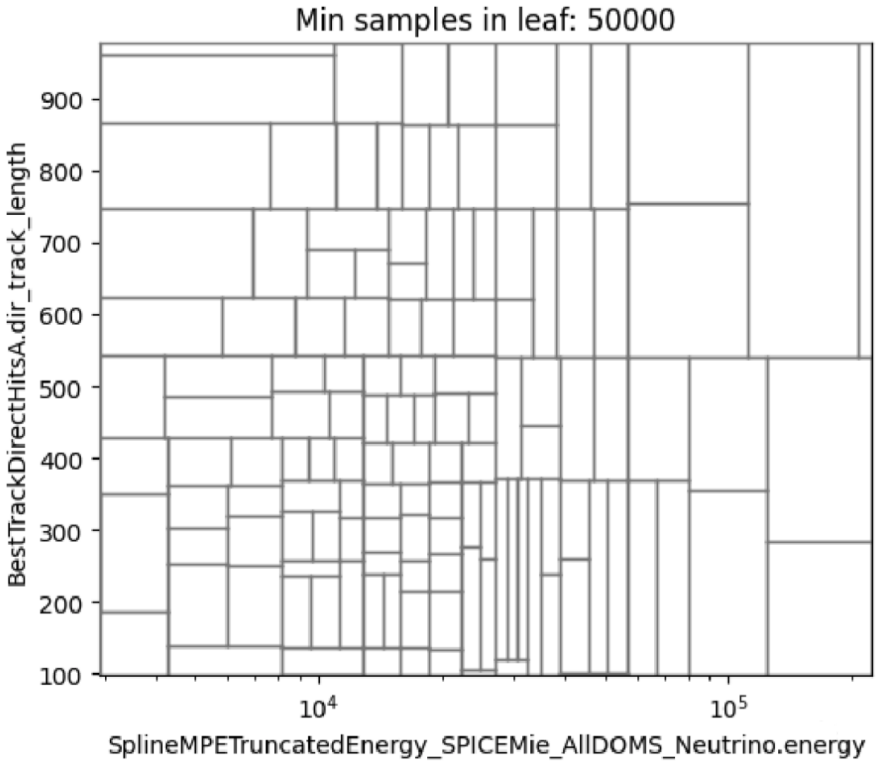

Figure 6.15: Rebinning of the observable space on a track length example. The minimum leaf size is greatly exaggerated to better visualize binning. Towards the lower end of the energy reconstruction values shown on the x-axis, the track length provides significant information in energy estimation. This region is split into more bins to enlarge the number of different values the track length can attain when used in a machine learning algorithm. Toward the higher end, length becomes irrelevant and the number of bins is reduced to control dimensionality. The relevance of the feature is highest in the regions with the densest cuts.

distribution. After 3000 jumps have been made, the chain is discarded and the result is used as the first estimate in a future chain of 10000 steps. The discarded chain is often referred to as *burn−in*, and it is used only to ensure the first random assumption is not too far from the optimal solution. The 10000 steps are all stored and represent the distributions of estimates in each bin. For each energy bin, 1000 walkers create 10000 estimates, therefore the resulting MCMC chain is a matrix of size $100000 \cdot 13$. The walkers converge to the best estimate in each energy bin, constructing $\vec{f}$ as the medians of each respective distribution.

Statistically more supported energy bins converge to the solution faster, as expected. The performance of the MCMC chain can therefore be deduced from the spread of the estimates' distributions. Low energy region produces distributions with well-defined peaks, while the high-energy region tends to be spread out around the best estimate, due to the higher complexity of

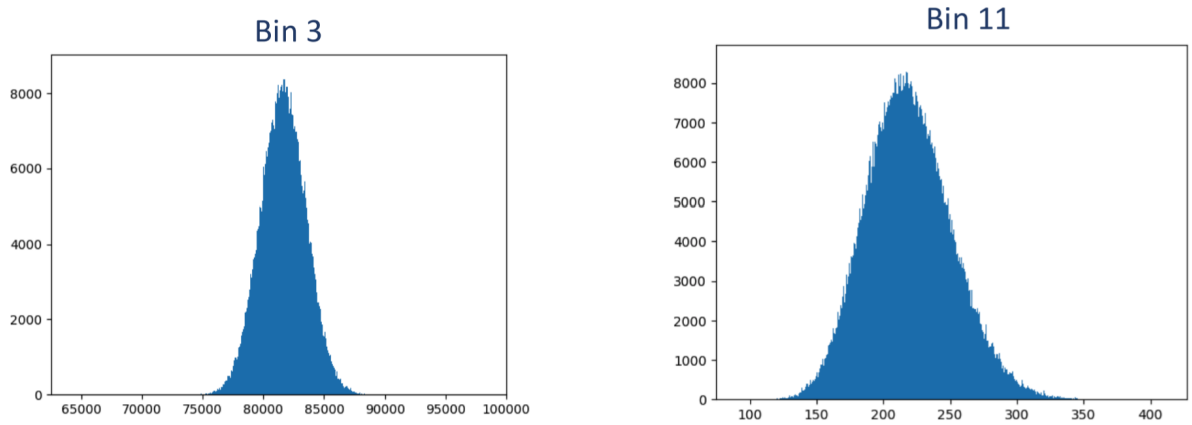

Figure 6.16: Histograms of MCMC estimates for two example bins. Bin 3 shows the energy range from 2 TeV to 4 TeV, and Bin 11 from 500 TeV to 1 PeV. The x-axis is the sampled space of $\vec{f}$, and the distributions peak at the best estimate. The distribution has in total 100000 entries. The low-energy and higher statistic region (left) produces a narrower distribution with a better defined Median. The high-energy region (right) exhibits a larger spread and slower convergence.

arriving at the solution. An example is shown in Figure 6.16, and the full range is presented in the Appendix A.15.

6.6 Optimization of Regularization parameters

Unfolding is iteratively repeated for various values of regularization parameters while keeping the analysis settings from Section 6.5 constant. As mentioned, regularization strength τ has a greater influence on the resulting spectrum than the offset d. Both parameters are sampled from a range and the full unfolding algorithm is applied to the pseudo data.

To evaluate the performance of a regularization set, the pseudosample of known distribution is unfolded. The resulting flux is compared to the true flux via the Chi-square statistics from Equation 3.6 in which for each bin E is the true flux, O is the unfolded flux, and the variance is the statistical error of unfolding. Values for both parameters are sampled more densely from regions exhibiting low Chi-square statistics and large p-values. Each unfolding is additionally tested with Manhattan, Chebyshev, and Earth Mover's distances from Section 3.2.

An example of statistical tests on sampled values is shown in Appendix Table A.1, while Figure 6.17 shows the distribution of p-values for different regularization strengths τ.

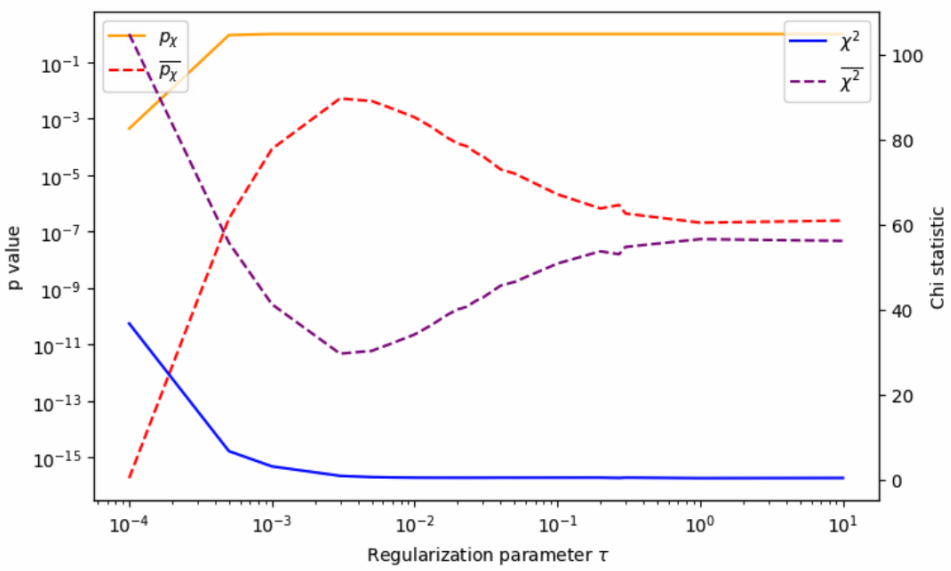

Figure 6.17: Statistical test on the regularization strength parameter τ. Chi-square statistic χ^2 and p-values are defined as in Equation 3.6, while the modified $\overline{\chi^2}$ and $\overline{p_\chi}$ are divided with expectancy instead of variance. The modified values better test the agreement between true and reconstructed flux, while unmodified take into consideration the coverage of uncertainty. Optimal strength is at the maximum of p-values, where the Chi-square statistic showing the deviation from the true value is at the lowest.

6.7 Additional tests

After all parameters have been optimized, additional tests are run to confirm the assumptions made. Primarily, the impartialness of the method on the weighting used to construct the migration matrix is checked. Analysis settings are fixed and the same pseudosample is unfolded varying the assumed flux in the simulated training set. A more proficient statistical test on many theoretical fluxes has been previously done in another work [Bö18] and showed no substantial bias even in assumptions of non-physical fluxes. The flux is still retrievable even in cases of training on only one source of events (one component of flux).

As described in Section 4.3.3, the data may include uncleaned background contamination of up to 0.13% comprised of atmospheric muons. To simulate the effect this could have on unfolding, 0.13% of events are sampled based on a parametrized muon flux from [Sol19]. Unfolding is run with usual analysis settings assuming optimal conditions and no background. The resulting flux has no apparent change in shape compared to a pure sample, with nonsignificant changes in p-values. The test is shown in the Appendix Figure A.14.

6.8 Results

A similar analysis chain using the same unfolding algorithm has been previously done and applied to three years of data, therefore this analysis can be considered an extension, or upgrade, to said analysis. Previous results are discussed in the Appendix A.3

The optimized settings are first applied to a simulated pseudosample. The simulated flux is correctly reconstructed inside the statistical error range per each energy bin. The statistical uncertainty is substantially smaller than in previous results, e.g. the results made with 3 years of data shown in Appendix Figure A.4, due to considerably higher amounts of data and proficiency of each step in the analysis. The event spectrum is shown in Figure 6.18.

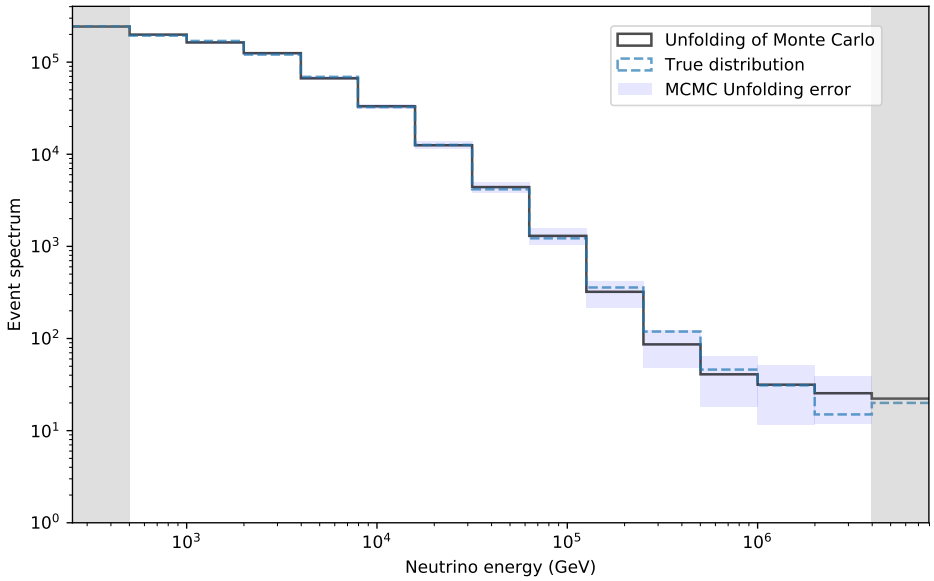

Figure 6.18: Event counts of the simulated pseudosample. The black solid line shows the unfolding in this work, and the dashed blue line the true event spectrum of the simulated set. The shaded area around bin counts shows the statistical uncertainty of unfolding. Event spectrum contains two additional bins: underflow starting from zero energy, and overflow reaching infinity. Additional bins are shown in gray.

Using Equation 5.10, the event spectrum is translated into flux, using the real effective area and the lifetime of the measured sample to be unfolded. The resulting flux is weighted with energy squared, as it is commonly reported in this form. The unfolded weighted flux corresponds well to true simulated values, with the statistical uncertainty always encompassing the true flux, as desired. The statistical uncertainty is less than 10% in the majority of the considered energy range, with an average of 18.13% and a 53.17% relative error in the highest energy bin. The offset of the estimate to true value is mainly below 5%, with the offset being at worst at 10.8% in

the highest bin and at best at 0.63% around 10 TeV. Additionally, statistical tests show excellent agreement. The results are shown in Figure 6.19 and are an example of what to expect from the application on real data.

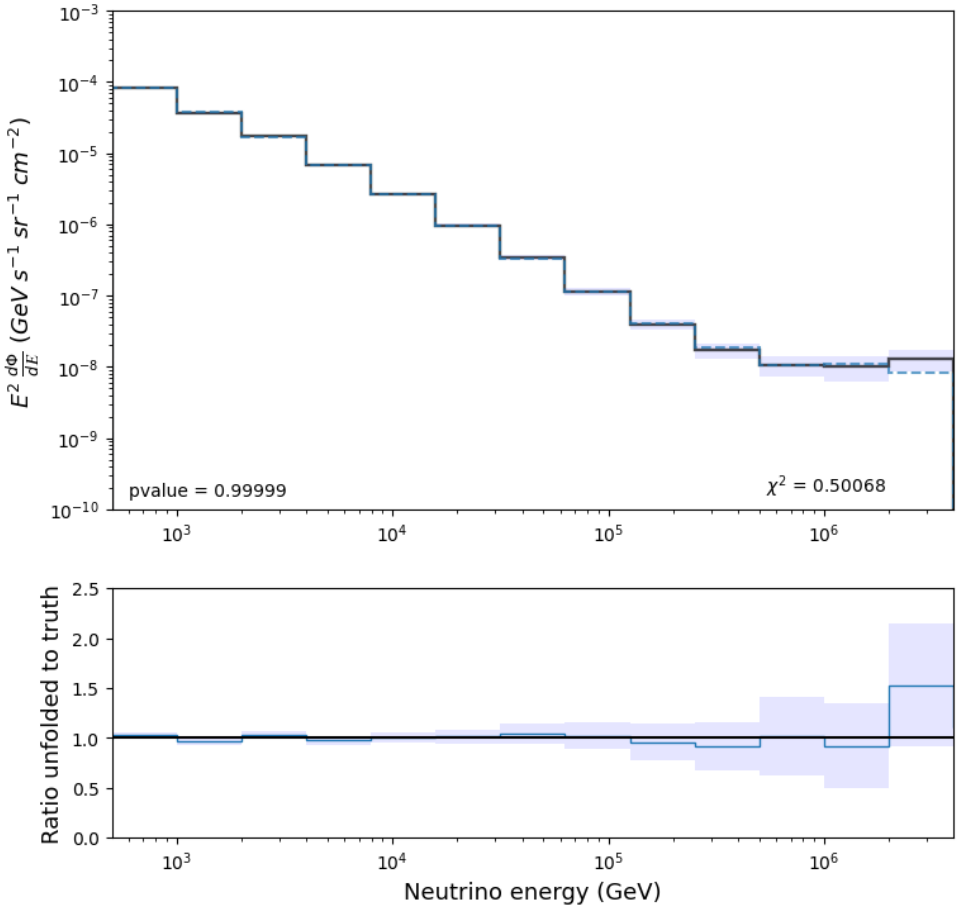

Figure 6.19: Weighted flux of the simulated pseudosample. The black solid line shows the unfolding in this work, and the dashed blue line the true flux of the simulated set. The shaded area around the unfolded values shows the statistical uncertainty of unfolding. The lower plot shows the ratio of the unfolded and the true simulated flux, with the black line representing a ratio of one and perfect agreement. The shading represents the scaled statistical uncertainty. Values in the bottom corners of the upper plot show the Chi-square statistics of the unfolding agreement from Equation 3.6.

All results spanning from unfolding true data measured with IceCube are shown with crosses as opposed to histograms made with simulated data, per the convention in the collaboration. Additionally, uncertainties of unfolding real data include errors spanning from systematic differences, described in more detail in Section A.5. Preliminary results are reached by unfolding *Burnsample* data, described in Section 4.3.4, with adjustments where needed. In comparison

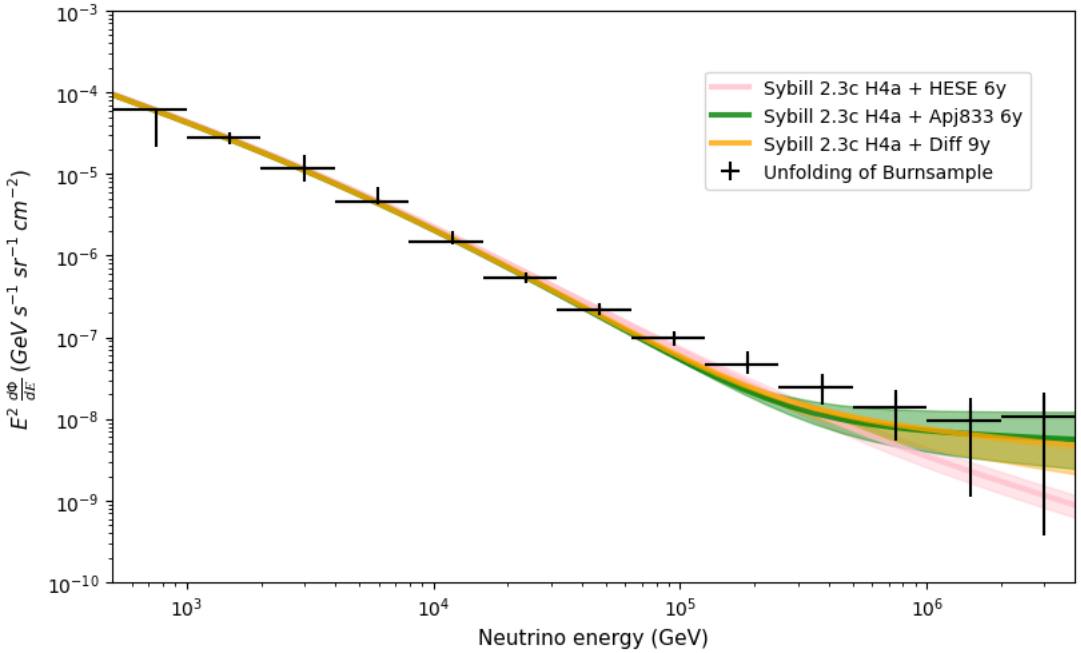

Figure 6.20: Weighted flux of the Burnsample data. The black solid lines show the unfolding in this work and the evaluated total uncertainty. Recent results from IceCube for the astrophysical components are shown in solid lines considering their corresponding energy ranges, with the shading showing uncertainty. The assumed astrophysical fit used in the total flux estimation [A$^+$16a] is in green, HESE analysis [A$^+$21] is shown in pink, and 9.5 years diffuse analysis [A$^+$22] in orange. All astrophysical fits are combined with the atmospheric model prediction given by SYBILL 2.3c [DLRF17].

to the full dataset and the pseudosample, *Burnsample* has an approximately ten times shorter lifetime but equivalent effective area. Since the true distribution is not known, and is the goal of this analysis, *Burnsample* unfolding can only be compared to current theoretical models, and cannot be assessed with statistical tests as in the case of simulated samples. Figure 6.20 shows the unfolded *Burnsample* weighted flux, compared to the total neutrino flux predictions made with different astrophysical models. Appendix Figure A.18 compares results to separated atmospheric components simulated with different cosmic ray models and experimental fits on the astrophysical component. Results show good alignment with both theoretical and experimental results, with a slight disagreement between 2000 and 4000 TeV. The disagreement might be accredited to the yet unknown precise spectral index of the prompt component, and the exact shape of the atmospheric flux at these energies where the two components are expected to exchange dominance. All experimental astrophysical fits, including this work, agree inside the uncertainty ranges. The amplified statistical uncertainty in the high energy region comes from the limited

statistics of the *Burnsample* and will diminish in application to all data.

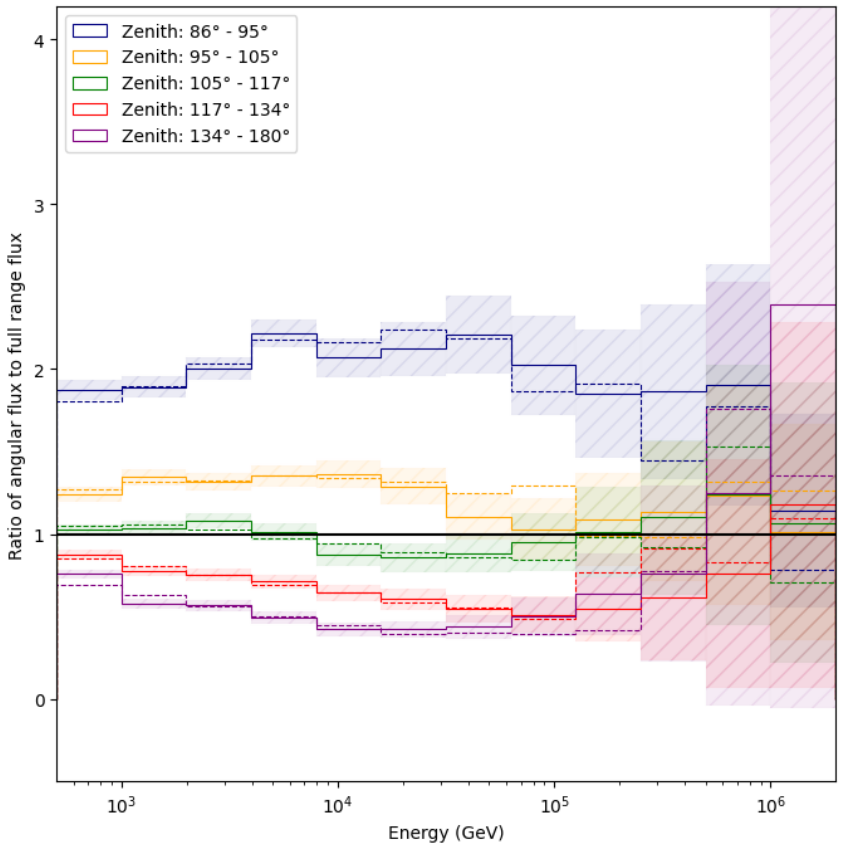

Figure 6.21: Ratio of fluxes in angular bins to the full range flux. The dashed lines show the true ratio pseudosampled from simulations, and the solid lines show the unfolded ratio with the statistical uncertainty shown as the shaded area around the best estimate. The statistical uncertainty does not overlap at energies lower than 1 PeV enabling discrimination among angular fluxes.

In the same manner, with the aim of first testing the algorithm, angular datasets are pseudosampled from the given simulation. The zenith bins are defined as described in Section 6.1 and visualized in Figure 6.3. The unfolded angular fluxes are calculated using effective areas given in Table 6.2 and examined in comparison to the full range flux ranging from 86 to 180 degrees.

Pseudosampling from a flux following a negative power law leads to higher statistical uncertainty in the regions of lower populations, which in this case corresponds to higher energy. In the case of unfolding a simulation following a known distribution, it is more sensical to compare it to the true simulated values instead of theoretical models, especially in the case of testing with proof-of-concept aims. For bins in which the statistical uncertainties overlap in the angular unfolding, it is not possible to discriminate fluxes. However, if bins exist with no overlapping

errors we conclude the fluxes are separable. In the example of one simulated and divided pseudosample, the results of angular flux ratios to the full range are given in Figure 6.21.

A clear separation is possible up to energies of 1 PeV in the five given zenith bins. At higher energies, the statistical uncertainty encompasses different bins and the bins become indistinguishable. The separation works really well in the atmospheric-dominated region for five zenith bins, which is a step further from the previously examined three bins.

Therefore, in application on real data, the atmospheric-dominated region can be tested against theoretical models, with behavior as described and given in Figure 2.4.

6.8.1 Unfolding eleven years of data

Finally, the algorithm is applied to the full dataset. Data consists of almost eleven years and is processed as described in Section 4.3.4. Regularization is chosen as shown in Section A.2, and the effective area is described in Section 6.2. The final solution is sampled with MCMC walkers from Section 6.5.

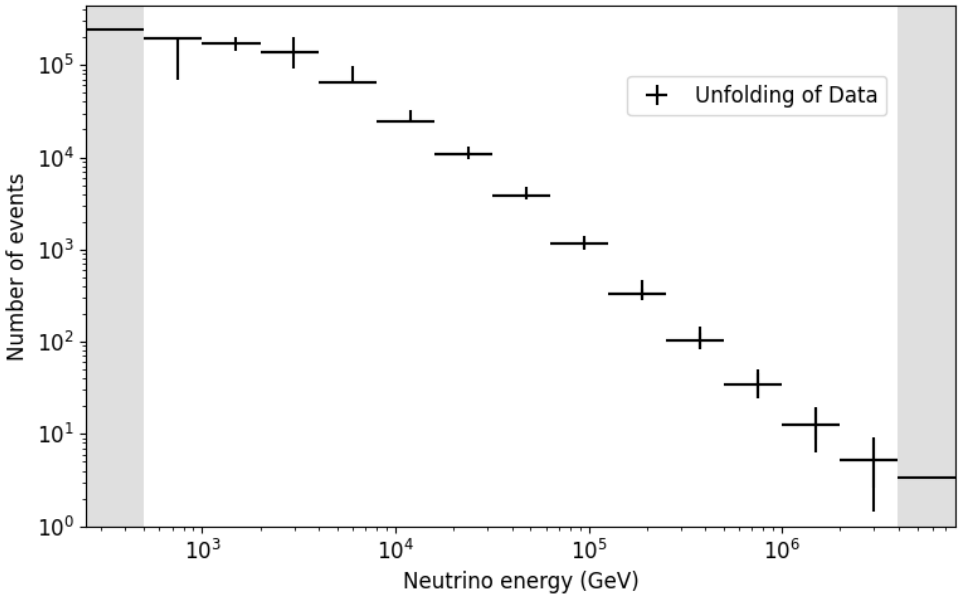

Figure 6.22: Event counts of the unfolded dataset. The black solid line shows the unfolding in this work. Vertical error bars show the total uncertainty of unfolding. Event spectrum contains two additional bins: underflow starting from zero energy, and overflow reaching infinity, for which the effective area and statistical uncertainty are not defined. Additional bins are shown in gray.

The solution, presented in Figures 6.22, 6.23, and 6.24, shows very low uncertainty, as expected through simulation studies. Exact values for the investigated energy range are given in Table 6.4. Results compared to separated components are given in Figure A.19. The atmospheric-

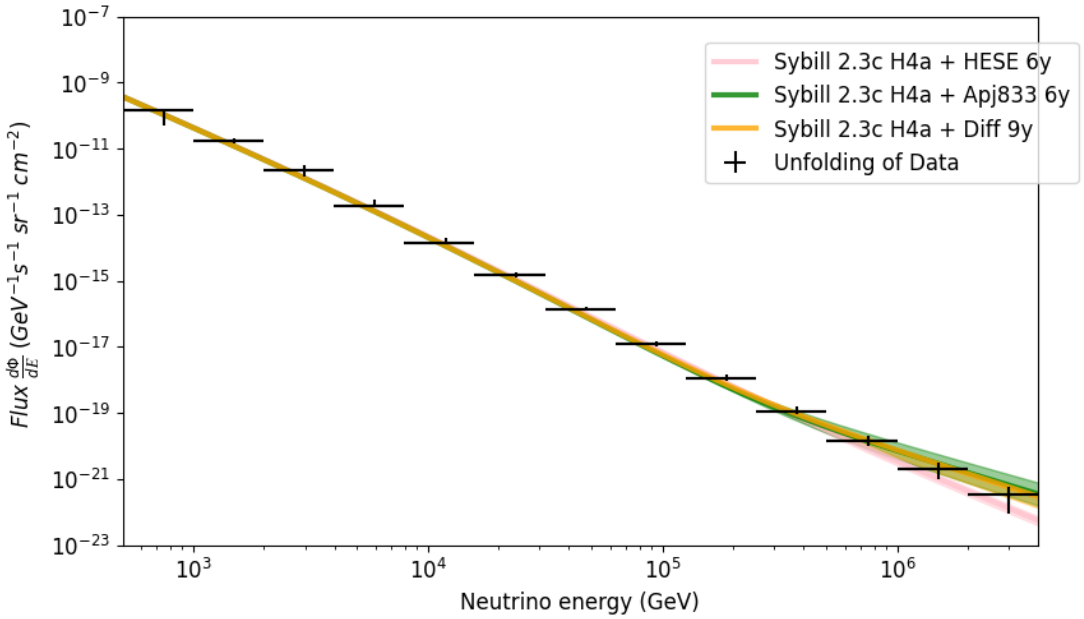

Figure 6.23: Flux of the unfolded data. The black solid line shows the unfolding in this work. Recent results from IceCube for the astrophysical components are shown in solid lines considering their corresponding energy ranges, with the shading showing uncertainty. The assumed astrophysical fit used in the total flux estimation [A$^+$16a] is in green, HESE analysis [A$^+$21] is shown in pink, and 9.5 years diffuse analysis [A$^+$22] in orange. All astrophysical fits are combined with the atmospheric model prediction given by SYBILL 2.3c [DLRF17]. The results are given in Table 6.4.

dominated region, spanning lower energies up to 100 TeV, is in good agreement with the atmospheric models. The astrophysical region also shows great alignment with previous results from IceCube. However, there is an excess of events in the region where conventional and prompt components are expected to interchange dominance, more prominent on visualizations weighted with energy, as given in Figure 6.24. The cross-over energy is not well known and is only assumed to be around 300 TeV. Different prompt models give different results for the exact cross-over energy, and the atmospheric models used here for comparison have already been shown to not properly simulate our prompt component expectancy [ERF$^+$17].

The weighting of the resulting flux with $E^{3.7}$ as presented in Figure 6.25 shows the break in the spectrum at higher energies. This visualization presents in a simplified way the retrieved different slopes of the spectral index describing the flux. Points moving away from a horizontal line point to a change of spectral index, with higher change associated with higher deviation from horizontal alignment. Another weighted flux is given in Appendix Figure A.20 visualized to highlight the spectral break.

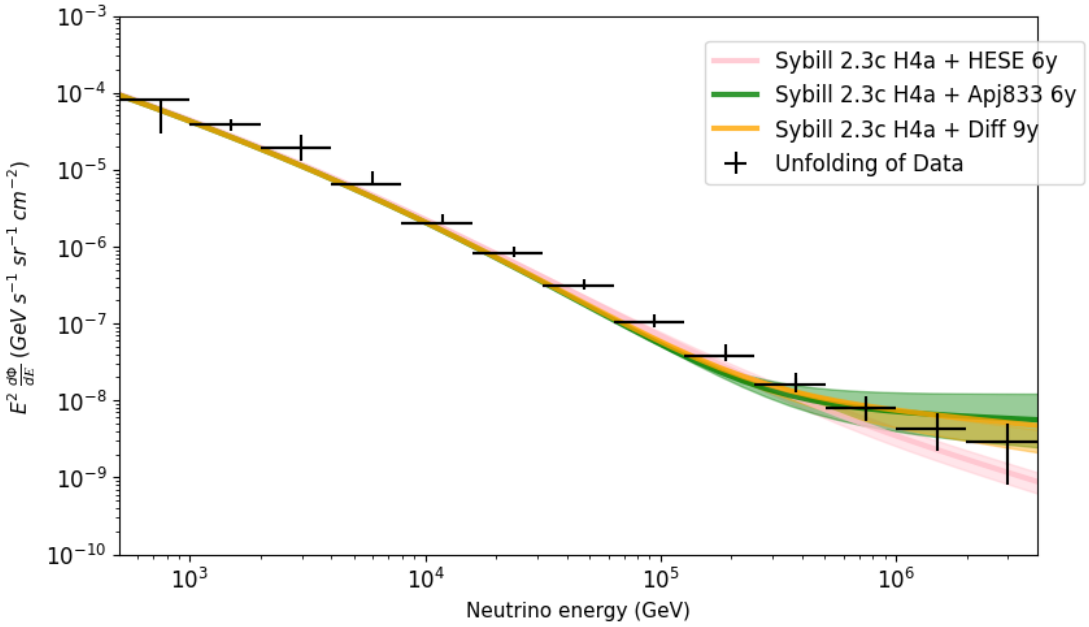

Figure 6.24: Weighted flux of the unfolded data. The black solid line shows the unfolding in this work. Recent results from IceCube for the astrophysical components are shown in solid lines considering their corresponding energy ranges, with the shading showing uncertainty. The assumed astrophysical fit used in the total flux estimation [A$^+$16a] is in green, HESE analysis [A$^+$21] is shown in pink, and 9.5 years diffuse analysis [A$^+$22] in orange. All astrophysical fits are combined with the atmospheric model prediction given by SYBILL 2.3c [DLRF17]. The unfolded flux shows good agreement in the atmospheric-dominated region with theoretical models, and astrophysical region with previous fits. Similarly, as with Burnsample unfolding, there is an excess in the cross-over region of conventional and prompt components.

Additionally, these results are compared to previous results developed on less data, and described in Appendix A.3

The two works show excellent agreement in the atmospheric-dominated region, where best estimates overlap. This work's best estimate is included in the uncertainty region of the previous analysis, which implies agreement between the two results. However, due to limited data, the previous results yield very high statistical uncertainty, therefore no better conclusions can be drawn from the comparison. The comparison is shown in Figure A.9.

6.8.2 Unfolding eleven years of data in angular bins

The process following the one described in the previous Section 6.8.1 is reapplied to subsets of data separated into zenith bins as defined in Section 6.1. The process is comparable, with

Energy range [GeV]	A_{eff} [m^2]	$\frac{d\Phi}{dE}$ [$GeV\ s^{-1}\ sr^{-1}\ cm^{-2}$]	$\sigma_{stat.}$ [$GeV\ s^{-1}\ sr^{-1}\ cm^{-2}$]	$\sigma^{+}_{syst.}$ [$GeV\ s^{-1}\ sr^{-1}\ cm^{-2}$]	$\sigma^{-}_{syst.}$ [$GeV\ s^{-1}\ sr^{-1}\ cm^{-2}$]
$10^{2.7} - 10^{3.0}$	$1.167 \cdot 10^{-1}$	$1.445 \cdot 10^{-10}$	$\pm 3.151 \cdot 10^{-12}$	$0.000 \cdot 10^{0}$	$-9.279 \cdot 10^{-11}$
$10^{3.0} - 10^{3.3}$	$4.381 \cdot 10^{-1}$	$1.715 \cdot 10^{-11}$	$\pm 5.159 \cdot 10^{-13}$	$2.673 \cdot 10^{-12}$	$-2.792 \cdot 10^{-12}$
$10^{3.3} - 10^{3.6}$	$1.411 \cdot 10^{0}$	$2.134 \cdot 10^{-12}$	$\pm 5.757 \cdot 10^{-14}$	$1.002 \cdot 10^{-12}$	$-6.986 \cdot 10^{-13}$
$10^{3.6} - 10^{3.9}$	$3.918 \cdot 10^{0}$	$1.847 \cdot 10^{-13}$	$\pm 6.242 \cdot 10^{-15}$	$8.869 \cdot 10^{-14}$	$0.000 \cdot 10^{0}$
$10^{3.9} - 10^{4.2}$	$9.524 \cdot 10^{0}$	$1.431 \cdot 10^{-14}$	$\pm 7.228 \cdot 10^{-16}$	$4.673 \cdot 10^{-15}$	$0.000 \cdot 10^{0}$
$10^{4.2} - 10^{4.5}$	$2.046 \cdot 10^{1}$	$1.483 \cdot 10^{-15}$	$\pm 8.422 \cdot 10^{-17}$	$2.943 \cdot 10^{-16}$	$-1.807 \cdot 10^{-16}$
$10^{4.5} - 10^{4.8}$	$3.905 \cdot 10^{1}$	$1.374 \cdot 10^{-16}$	$\pm 1.034 \cdot 10^{-17}$	$2.983 \cdot 10^{-17}$	$-1.507 \cdot 10^{-17}$
$10^{4.8} - 10^{5.1}$	$6.726 \cdot 10^{1}$	$1.183 \cdot 10^{-17}$	$\pm 1.184 \cdot 10^{-18}$	$2.866 \cdot 10^{-18}$	$-1.670 \cdot 10^{-18}$
$10^{5.1} - 10^{5.4}$	$1.070 \cdot 10^{2}$	$1.079 \cdot 10^{-18}$	$\pm 1.529 \cdot 10^{-19}$	$4.412 \cdot 10^{-19}$	$-1.322 \cdot 10^{-19}$
$10^{5.4} - 10^{5.7}$	$1.576 \cdot 10^{2}$	$1.135 \cdot 10^{-19}$	$\pm 2.246 \cdot 10^{-20}$	$4.985 \cdot 10^{-20}$	$-1.754 \cdot 10^{-20}$
$10^{5.7} - 10^{6.0}$	$2.156 \cdot 10^{2}$	$1.403 \cdot 10^{-20}$	$\pm 4.289 \cdot 10^{-21}$	$6.477 \cdot 10^{-21}$	$-4.132 \cdot 10^{-21}$
$10^{6.0} - 10^{6.3}$	$2.805 \cdot 10^{2}$	$1.945 \cdot 10^{-21}$	$\pm 9.693 \cdot 10^{-22}$	$1.087 \cdot 10^{-21}$	$-5.942 \cdot 10^{-22}$
$10^{6.3} - 10^{6.6}$	$3.516 \cdot 10^{2}$	$3.263 \cdot 10^{-22}$	$\pm 2.350 \cdot 10^{-22}$	$2.459 \cdot 10^{-22}$	$-1.624 \cdot 10^{-22}$

Table 6.4: The unfolded flux from 11 years of data, shown in Figure 6.23. The table shows the unfolded flux in each considered energy bin with their corresponding effective area, statistical error, and upper and lower systematic error.

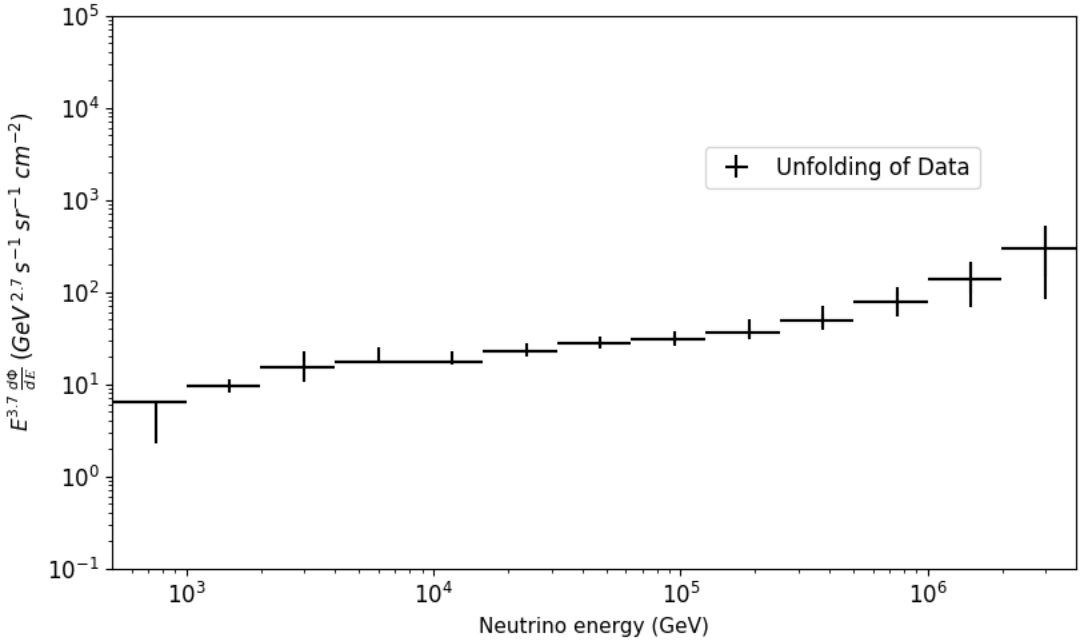

Figure 6.25: Flux of the unfolded data weighted with $E^{3.7}$. The black solid line shows the unfolding in this work with vertical lines showing the total uncertainty. The deviation from the horizontal alignment at higher energies implies a break in the spectrum, predicted by the overtaking dominance of the astrophysical component.

datasets having appropriately recalculated effective areas and systematic errors for their respective coverage.

The unfolded fluxes are first visualized and inspected without weighting with energy, shown in Figure 6.26. The five resulting sets of points are separable in several bins, mostly positioned in the lower energy atmospheric-dominated region. The uncertainties start to overlap at approximately 125 TeV, or the 9th energy bin.

As shown in Section 2.2.2, the particularities of the angular fluxes are best shown when considered in ratio to the all-range flux. Therefore, each of the resulting angular fluxes is divided with the flux shown in Figure 6.24. Figure 6.27 shows the resulting calculated ratios, and the appropriate total error (considering the varying systematic error at different zenith angles, given in Section A.5). As already implied in Figure 6.26, the bins are separable up to the astrophysical-dominated region. The non-symmetrical uncertainties span from the same nature of the angular systematic errors.

The unfolded angular ratio can be compared to the various atmospheric models and astrophysical fits. As already described, the differences in the models are more prominent when inspected through angular ratios. The unfolded data points are tested against the theoretical models described in Section 4.1 with the procedure given in Section 3.2.

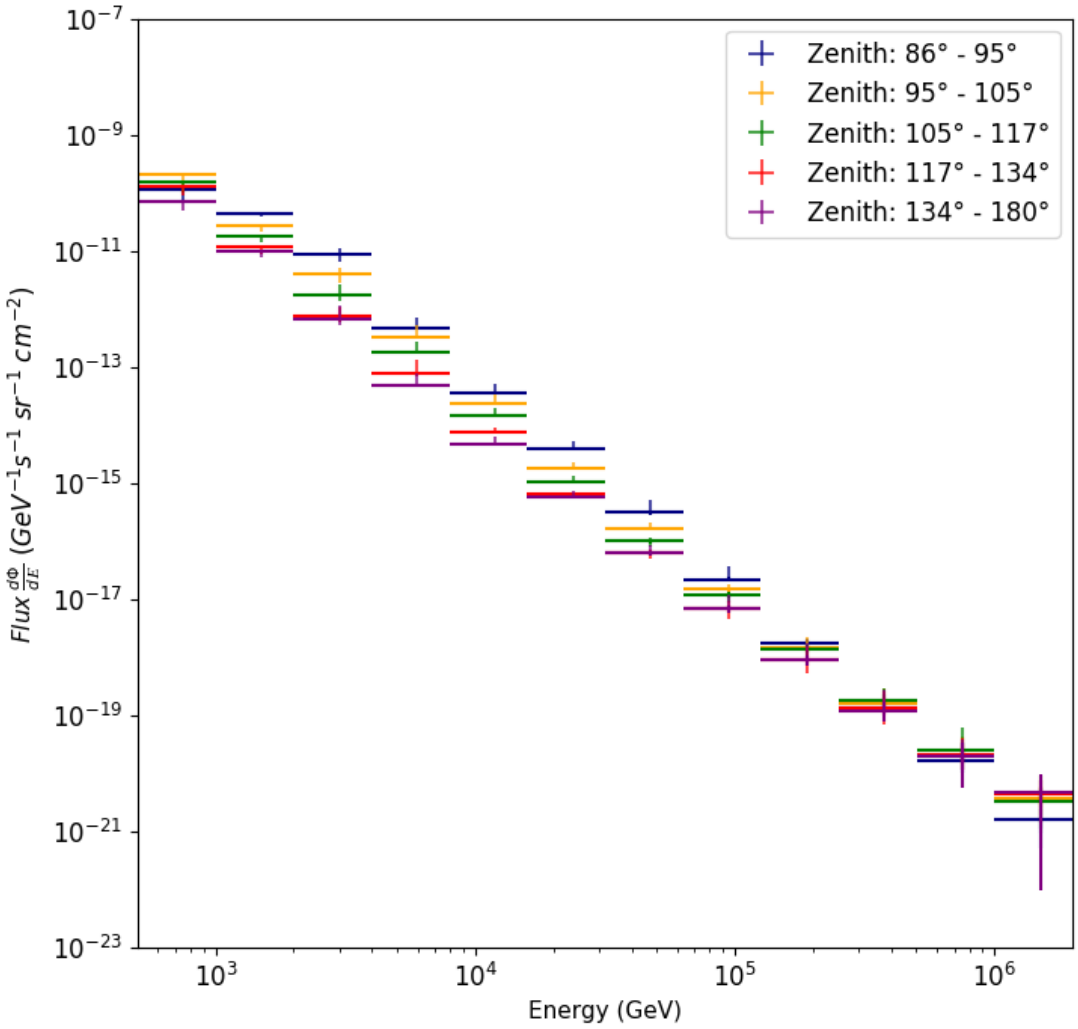

Figure 6.26: Unfolded flux of the five zenith bins considered in this work. The vertical lines show the total uncertainty of unfolding. Bins in the atmospheric-dominated region are separable, while the astrophysical region exhibits overlapping error bars.

The lowest Chi-square statistic is given for the QGSJet interaction model, the H3a Hillas primary model [Gai12], and the diffuse 9.5-year astrophysical fit [A+22]. The Chi-square statistics of the agreement between the estimate and the various atmospheric models combined with the diffuse 9.5 year astrophysical fit are given in Table 6.5. The statistics for the two other astrophysical fits are shown in Appendix Table A.2 and Appendix Table A.3.

All tested models containing an astrophysical contribution have a p-value higher than 0.99, confirming that the unfolding indeed follows the predicted angular behavior. A comparison of the unfolded ratio and the described model is visualized in Figure 6.28.

Additionally, atmospheric models with no astrophysical component were tested and all were shown to underperform in comparison to models with an astrophysical contribution. For refer-

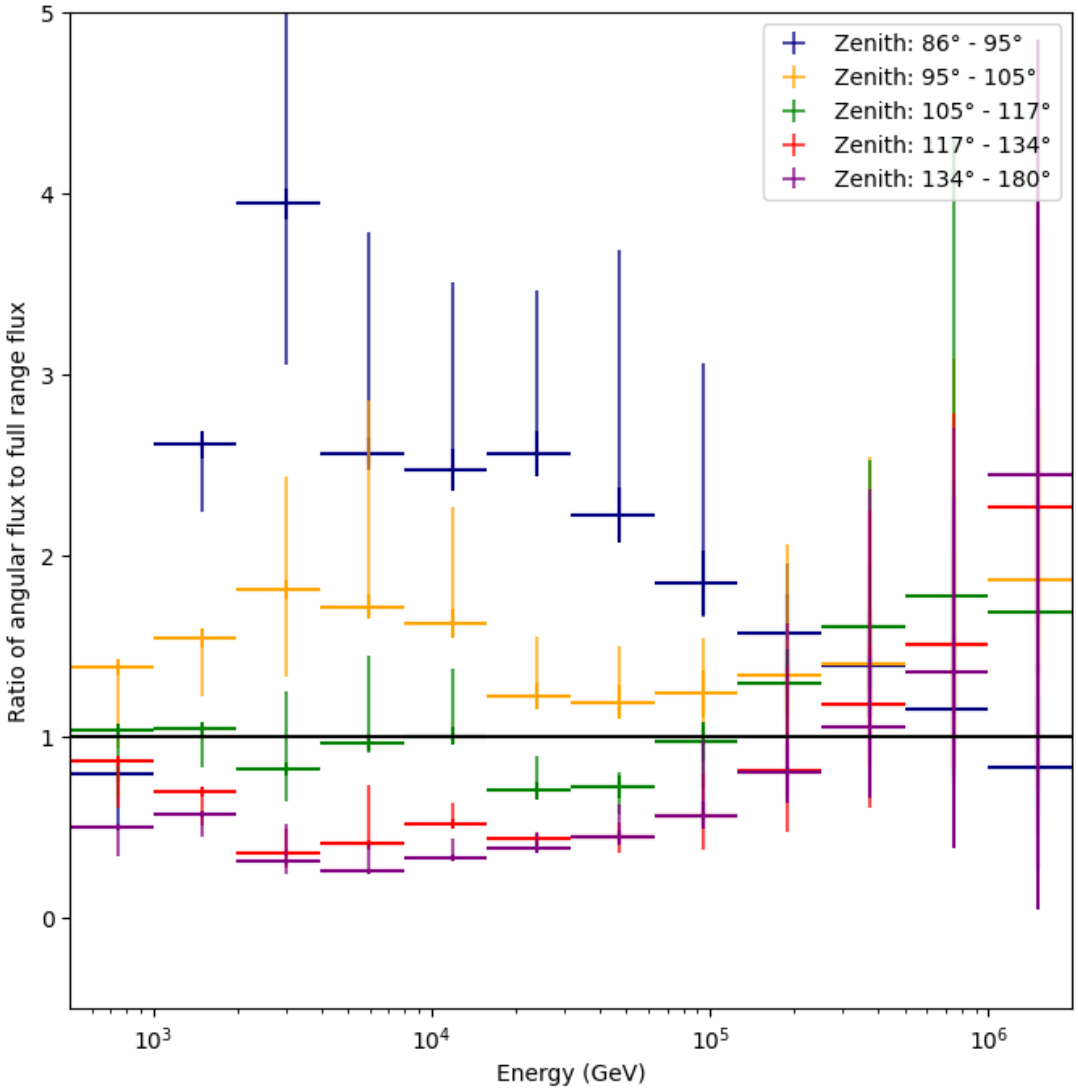

Figure 6.27: Unfolded ratios of the five zenith bins to the all range flux. The vertical lines show the total uncertainty of unfolding. Bins in the atmospheric-dominated region are separable, while the astrophysical region exhibits overlapping error bars.

ence, the Chi-statistics of purely atmospheric models are given in Appendix Table A.4, and an example is visualized in Appendix Figures A.21, A.22.

Diff 9y	**H4a**	**4-gen**	**H3a**	**default**
SIBYLL23C	10.9931	10.9484	10.9649	11.0692
SIBYLL-2.1	11.2653	11.179	11.2263	11.3669
EPOS-LHC	11.1335	11.0439	11.0931	11.2236
DPMJET-III-3.0.6	10.8505	10.8757	10.8348	10.9073
QGSJet-II-04	10.861	10.8335	10.8327	10.9281

Table 6.5: Chi-square statistic of agreement between the unfolded best estimates and the expectation made with the diffuse 9.5 year astrophysical fit [$A^{+}22$] and different atmospheric models. Columns show different interaction models, while rows represent primary models. Tested models are described in Section 4.1.

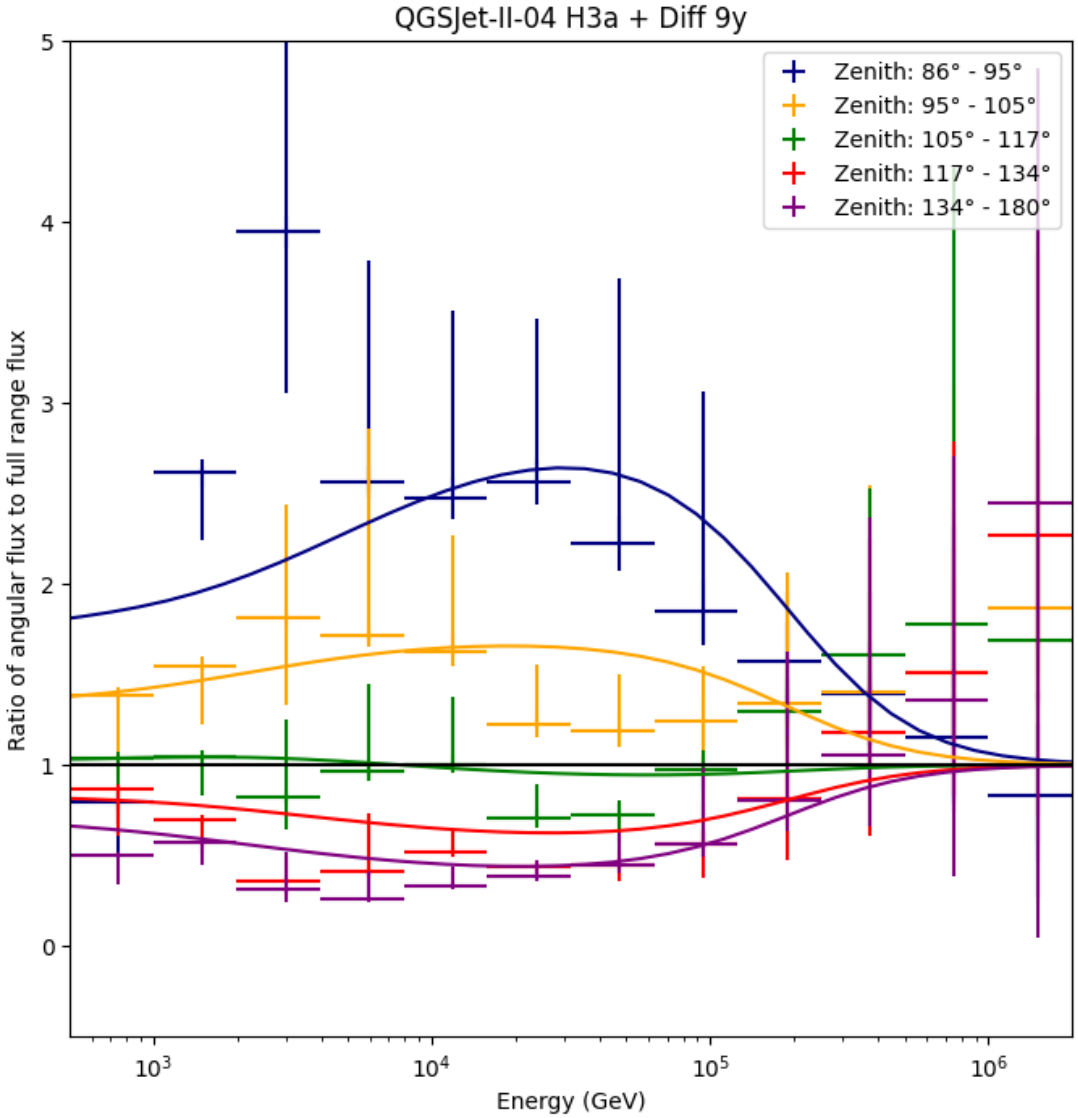

Figure 6.28: Unfolded ratios of the five zenith bins to the all range flux. The vertical lines show the total uncertainty of unfolding. The prediction is shown in solid lines, color-coded to correspond to the respective angular ranges. The simulation was made with MCEq [FEG+15] following the QGSJet interaction model, the H3a Hillas primary model [Gai12], and the diffuse 9.5 year astrophysical fit [A+22].

Chapter 7

Discussion and future aims

In this concluding chapter, the research efforts made in this work with the aim of unfolding the muon neutrino flux are brought together. The primary objective has been to estimate the neutrino energy spectrum using IceCube data in range spanning from 500 GeV to 4 PeV with unprecedented precision. The flux has been successfully unfolded and presented in this work, with additional interesting conclusions made during the analysis.

A contributing factor to the great quality of the reconstruction, inferred from the low uncertainty surrounding the best estimates, is the unparalleled amount of data available in this analysis. Previous studies have utilized a maximum of 9.5 years of data, and used different methods, resulting in less sensitivity. The dataset also demonstrates excellent purity, at 99.87%, allowing a background-less unfolding. The purity is a result of a meticulous cleaning process that encompasses both the energy and angular dependence of the events in IceCube.

One of the defining features of this work is the incorporation of machine learning techniques at various stages of the analysis. From data preprocessing to feature selection, intelligent models were utilized with the aim of both improving the results and reducing resource usage. Machine learning has enabled the optimal usage of a tremendous amount of data, and enhanced the overall precision of the results.

This work presented a three-step feature selection process, ensuring the observables used for unfolding are not only relevant but also yield precise results and encompass our knowledge about the detector. The features are selected considering their reconstruction power, their agreement to data, and their relevance in unfolding. Additionally, a novel rebinning technique was applied to the observable dataset, establishing proper usage of the available statistics and the underlying information about the interconnection of observables not available to researchers.

Finally, these results have demonstrated that unfolding is a state-of-the-art technique for extracting valuable information from the complex nature of IceCube data. The independence on assumptions made in the process has been repeatedly demonstrated through various tests. The lower uncertainty range shows the power of the method, while its good agreement with previous

results and theoretical models further grounds its applicability.

A special aim of this work was the investigation of ratios of angular fluxes, not before evaluated nor tested in IceCube. Inspection of ratios allows seeing the peculiarities of the total neutrino flux, coming from the nature of the three distinct components, in a superior manner compared to the investigation of fluxes alone. This work also demonstrated the applicability of the method for these studies and paves the way for more detailed studies of the angular dependence of both atmospheric and astrophysical neutrinos.

The results confirm both the expected shape of the total neutrino flux and the predicted anisotropy at the atmospheric-dominated energies.

Results can be further used in studies of the various neutrino models, and in probing the nature of the composition of the neutrino flux depending on the number of constituting parts. The angular studies can be and are encouraged to be used in future studies of the atmosphere, as the results encompass the interactions neutrinos partake in while traveling through air.

As the statistical uncertainty is already impressively low, due to the amount of data used, attaining more data is not expected to significantly improve the results in the following years of data taking in IceCube. However, the results would benefit from turning to better detector studies that would lead to a lower systematic impact, which holds true for many other studies aside from flux reconstruction.

In conclusion, the combination of the presented unfolding technique with high-quality data and the power of machine learning enabled unrivaled results of the muon neutrino flux measurement. I hope these results inspire future researchers and help them in analyzing and drawing conclusions about the curious Universe surrounding us and the nature of our atmosphere, through the lens of neutrinos.

Appendix A

A.1 Unfolding

The expanded derivations of expressions used in this work and described in Chapter 5 are given here. The Poissonian likelihood

$$\alpha(\vec{g}|\vec{f}) = \prod_{u=1}^{m} \frac{\lambda_u^{g_u}}{g_u!} \cdot exp(-\lambda_u) \tag{A.1}$$

has the expectancy λ defined with Equation 5.2

$$\alpha(\vec{g}|\vec{f}) = \prod_{u=1}^{m} \frac{(A\vec{f})^{g_u}}{g_u!} \cdot exp(-(A\vec{f})_u) \cdot exp(R) \tag{A.2}$$

with an added regularization parameter R. Applying a logarithm to the whole expression

$$ln(\alpha(\vec{g}|\vec{f})) = ln(\prod_{u=1}^{m} \frac{(A\vec{f})^{g_u}}{g_u!} \cdot exp(-(A\vec{f})_u) \cdot exp(R)) \tag{A.3}$$

results in

$$l(\vec{g}|\vec{f}) = \sum_{u=1}^{m} (g_u ln(A\vec{f})_u - (A\vec{f})_u) - ln(g_u!)) + R \tag{A.4}$$

from which the independent term is dropped producing

$$l(\vec{g}|\vec{f}) = \sum_{u=1}^{m} (g_u ln(A\vec{f})_u - (A\vec{f})_u)) + R \tag{A.5}$$

Regularization is applied in the form of a Thikonov matrix

$$l(\vec{g}|\vec{f}) = \sum_{u=1}^{m} (g_u ln(A\vec{f})_u - (A\vec{f})_u)) - \frac{1}{2}\left[(C\vec{f})^T (C\vec{f})\right] \tag{A.6}$$

which is then multiplied with the parameter τ to enable strength of regularization adjustment

$$l(\vec{g}|\vec{f}) = \sum_{u=1}^{m}(g_u ln(A\vec{f})_u - (A\vec{f})_u)) - \frac{1}{2}\left[(C\vec{f})^T Diag(\mathbb{I}\cdot\tau)(C\vec{f})\right] \tag{A.7}$$

This describes the final form of likelihood to be sampled by the random walkers. C is the Tikhonov matrix

$$C = \begin{bmatrix} 1 & -2 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & ... & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & ... & 0 & 1 & -2 & 1 \end{bmatrix} \tag{A.8}$$

sometimes referred to as the Ridge Regression matrix. Tikhonov regularization penalizes non-smooth solutions based on the strength of regularization τ which is optimized for a given problem.

A.2 Regularization optimization

The starting estimate for the regularization parameters is sought by a trial-and-error approach. After approximate orders of magnitude are established, regularization strength and offset are sampled from a wide range of combinations. Each set is tested against the known distribution, with measures from Section 3.2 evaluated.

Effects of unregularized unfolding can be seen in Figure A.1.

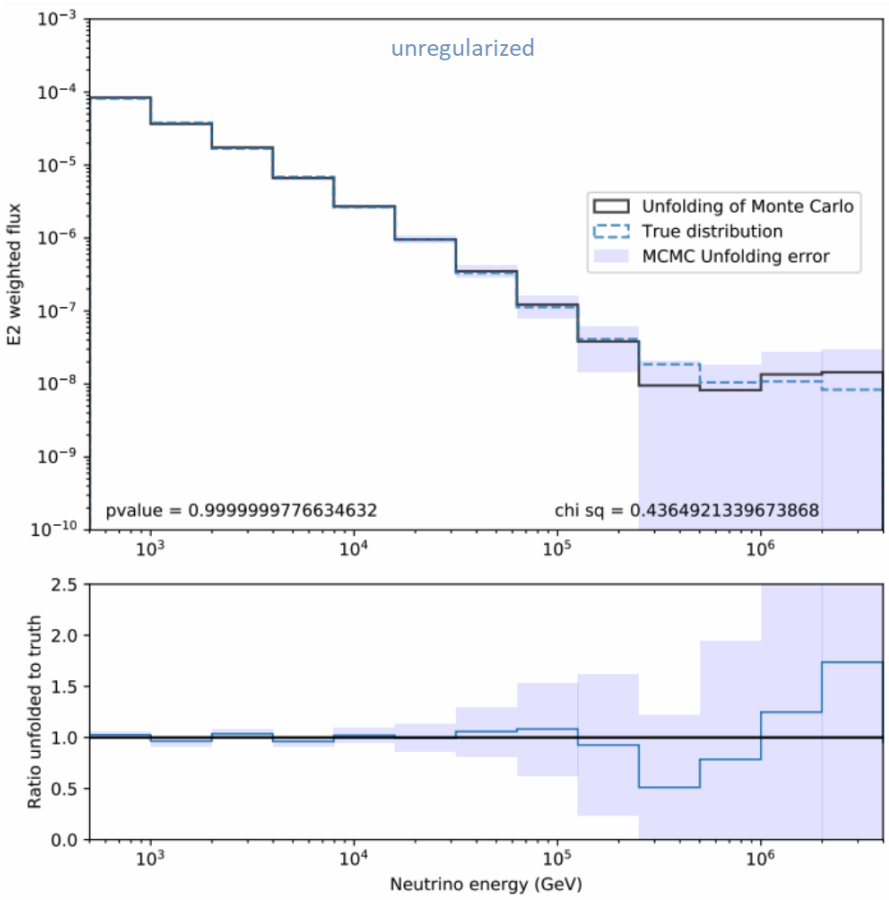

Figure A.1: Unregularized unfolding of a pseudosample. The estimation is substantially deteriorated with high uncertainties due to the complexity of the problem.

	d = 14.0	d = 20.0
$\tau = 0.017$	$\chi^2 = 0.5235357081622899$ $p_\chi = 0.9999999298541741$ $\overline{\chi^2} = 39.14657058541254$ $\overline{p_\chi} = 0.0001894514072518325$ $D_M = 16272.909784071877$ $D_{Ch} = 0.5035670298077569$ $D_{EDM} = 1251.762291082452$	$\chi^2 = 0.5011805386997424$ $p_\chi = 0.9999999466661433$ $\overline{\chi^2} = 38.86891403134682$ $\overline{p_\chi} = 0.00020980654771567952$ $D_M = 16221.306851060957$ $D_{Ch} = 0.489050930037561$ $D_{EDM} = 1247.7928346969966$
$\tau = 0.02$	$\chi^2 = 0.5183509468671078$ $p_\chi = 0.9999999341011512$ $\overline{\chi^2} = 40.12986328007975$ $\overline{p_\chi} = 0.00013173679422423244$ $D_M = 16495.915866311097$ $D_{Ch} = 0.5166573541549274$ $D_{EDM} = 1268.9166051008538$	$\chi^2 = 0.5001238587562805$ $p_\chi = 0.9999999473688342$ $\overline{\chi^2} = 40.105845685319515$ $\overline{p_\chi} = 0.00013291573598958012$ $D_M = 16502.97752278177$ $D_{Ch} = 0.5037914315913393$ $D_{EDM} = 1269.459809444752$
$\tau = 0.023$	$\chi^2 = 0.5226548192756095$ $p_\chi = 0.9999999305914219$ $\overline{\chi^2} = 41.47324765217568$ $\overline{p_\chi} = 7.982088049592654e-05$ $D_M = 16783.149865543128$ $D_{Ch} = 0.531696383751279$ $D_{EDM} = 1291.0115281187022$	$\chi^2 = 0.49514285761256244$ $p_\chi = 0.9999999505779448$ $\overline{\chi^2} = 40.681762242894756$ $\overline{p_\chi} = 0.00010729762054787487$ $D_M = 16621.21027907897$ $D_{Ch} = 0.517751333788659$ $D_{EDM} = 1278.5546368522284$
$\tau = 0.027$	$\chi^2 = 0.5243726855408345$ $p_\chi = 0.9999999291476315$ $\overline{\chi^2} = 42.64604224197798$ $\overline{p_\chi} = 5.132759622194328e-05$ $D_M = 17015.955213002337$ $D_{Ch} = 0.5443434943998559$ $D_{EDM} = 1308.9196317694107$	$\chi^2 = 0.5006840872965537$ $p_\chi = 0.9999999469972527$ $\overline{\chi^2} = 42.16514668423613$ $\overline{p_\chi} = 6.15432629730593e-05$ $D_M = 16937.51592031455$ $D_{Ch} = 0.5299642180430068$ $D_{EDM} = 1302.8858400241957$

Table A.1: Example of tested combinations of regularization parameters on an unfolded pseudosample against the known true distribution. The columns show two different offsets d, and rows show four strengths τ. Values correspond to various statistical tests with definitions given in Section 3.2.

A.3 Unfolding three years of data

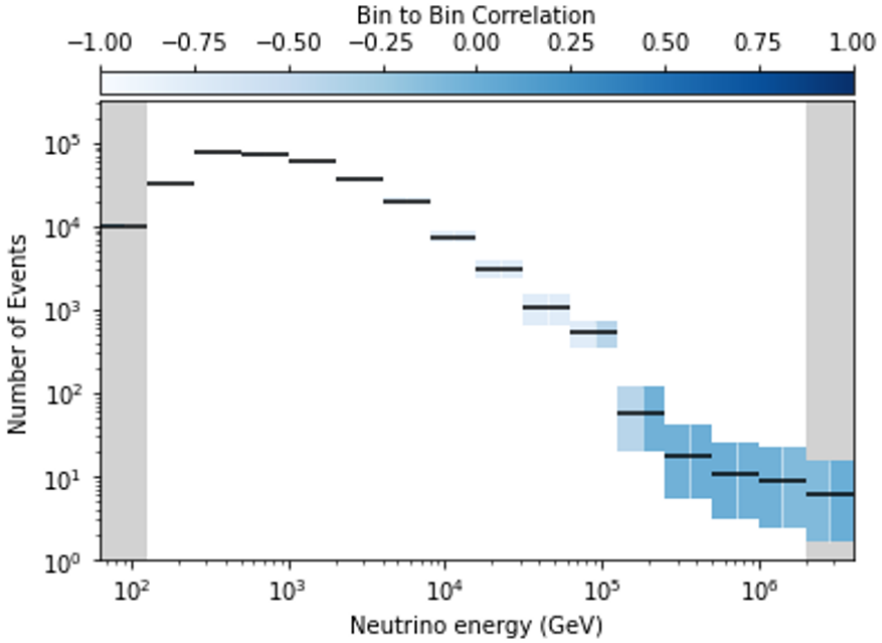

Figure A.2: Event counts of the unfolded three-year sample. The black solid line shows the unfolding in this work, with the shaded area around bin counts showing the statistical uncertainty of unfolding. The intensity of shading shows the Pearson correlation between estimate distributions of neighboring bins. The event spectrum contains additional two bins: underflow starting from zero energy, and overflow reaching infinity. Additional bins are shown in gray.

Prior to this work, the algorithm was applied to three years of data collected in IceCube from 2012 to 2015. lifetime of the sample is 91709733 seconds with 318224 signal events. The cleaning procedure is different from the one presented in Section 4.3.3, but samples have comparable purities. The analysis chain, including preprocessing and feature selection, has slight differences compared to the new analysis which has been optimized for current data and the state of prowess.

The energy range considered in the three-year analysis spans slightly lower energies from 125.9 GeV up to 1.995 PeV split into 14 bins, and additional underflow and overflow. The zenith bins divide the sky into only three regions, due to statistical constraints, with bins from 86° to 107°, 107° to 111 °, and 111° to 180°.

Correspondingly to the analysis with extended data, the energy range is slightly reduced when considering the flux in zenith bins, to avoid the already low statistic region of high energy distorting results after splitting the data into thirds. Therefore, the reduced energy range is from

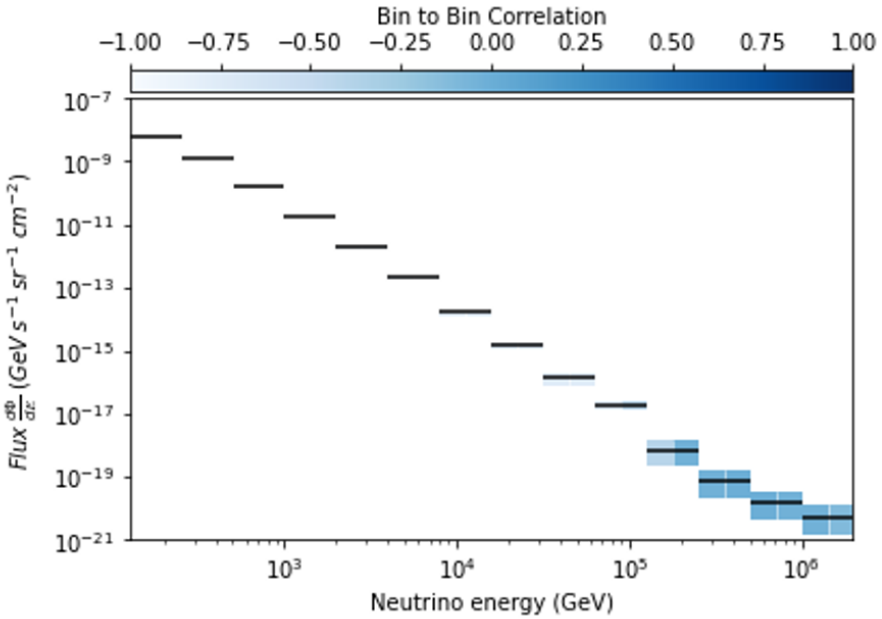

Figure A.3: Flux of the unfolded three-year sample. The black solid line shows the unfolding in this work, with the shaded area around bin counts showing the statistical uncertainty of unfolding. The intensity of shading shows the Pearson correlation between estimate distributions of neighboring bins. Flux is calculated from Equation 5.10 for the energy range from $10^{2.1}$ *to* $10^{6.1}$ *GeV.*

125.9 GeV to 251.2 TeV.

Figures **??**, A.3, and A.4 show the unfolded event spectrum, flux, and the weighted flux of the three-year sample. Figures A.5, A.6, and A.7 show the unfolded flux in the three considered angular ranges.

The best estimate of angular unfolding agrees well with the assumptions made with MCEq simulating the atmospheric components, as shown in Figure A.8. Contrary to the extended analysis, the astrophysical fit is not added to the MCEq simulation as its dominance starts at higher energies.

Both the full sky and angular unfoldings exhibit much larger uncertainties in the high energy region, due to limited statistics of the sample. Uncertainties of this scale do not allow of testing any models.

After the development of this work, the new results were compared to previous ones and shown in Figure A.9.

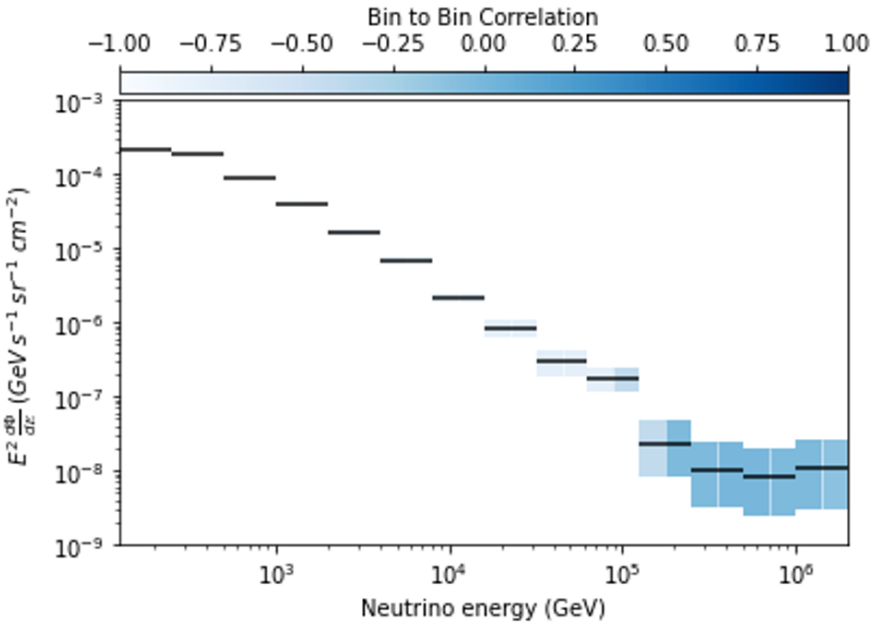

Figure A.4: Weighted lux of the unfolded three-year sample. Black solid line shows the unfolding in this work, with the shaded area around bin counts showing the statistical uncertainty of unfolding. The intensity of shading shows the Pearson correlation between estimate distributions of neighboring bins. Flux is calculated from Equation 5.10 and is multiplied with the energy squared for the energy range from $10^{2.1}$ *to* $10^{6.1}$ *GeV.*

A.4 Effective lifetime

Any simulated dataset like *CORSIKA* or *nugen*, per definition, does not have a lifetime. It is however convenient to define an *effective lifetime* of a simulation with the aim of quantifying the statistical error. An effective lifetime is used to define a minimum amount of data that has a comparable Poissonian error to the sample that is to be unfolded.

A simulation set with i weighted events x_i with accompanying weights w_i has a total number of weighted events

$$N = \sum_{i=1}^{n} w_i \tag{A.9}$$

with variance

$$\sigma_N^2 = \sum_{i=1}^{n} w_i^2 \tag{A.10}$$

The number of *effective events* n_{eff} is defined as the minimum number of events in the simulation

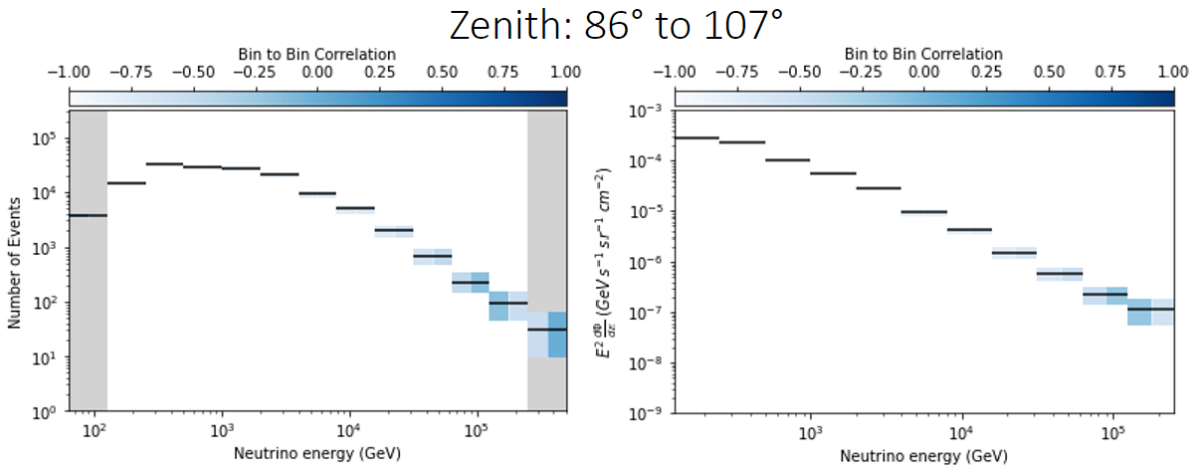

Figure A.5: Event spectrum (left) and the weighted flux (right) of the unfolded three-year sample in the 86° to 111° zenith band, also known as the horizon bin. The black solid line shows the unfolding in this work, with the shaded area around bin counts showing the statistical uncertainty of unfolding. The intensity of shading shows the Pearson correlation between neighboring bins. Flux is calculated from Equation 5.10 and is multiplied with the energy squared for the energy range from $10^{2.1}$ *to* $10^{5.4}$ *GeV.*

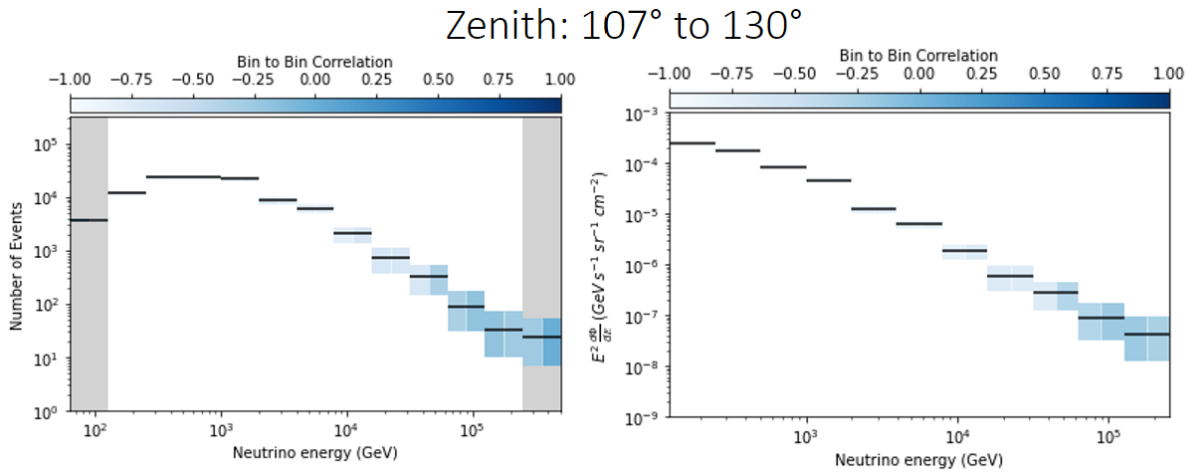

Figure A.6: Event spectrum (left) and the weighted flux (right) of the unfolded three-year sample in the 107° to 130° zenith band, with markings as in Figure A.5.

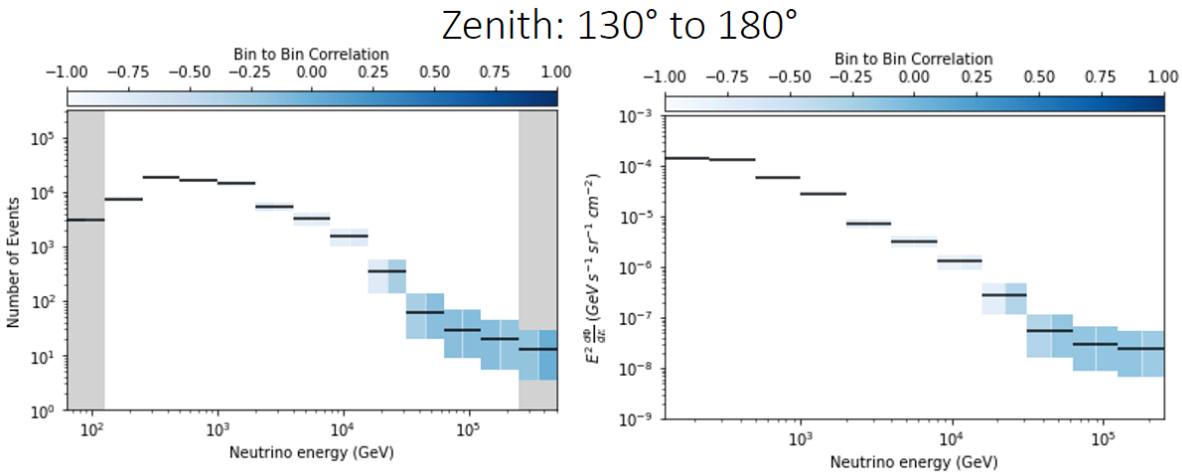

Figure A.7: Event spectrum (left) and the weighted flux (right) of the unfolded three-year sample in the 130° to 180° zenith band, with markings as in Figure A.5.

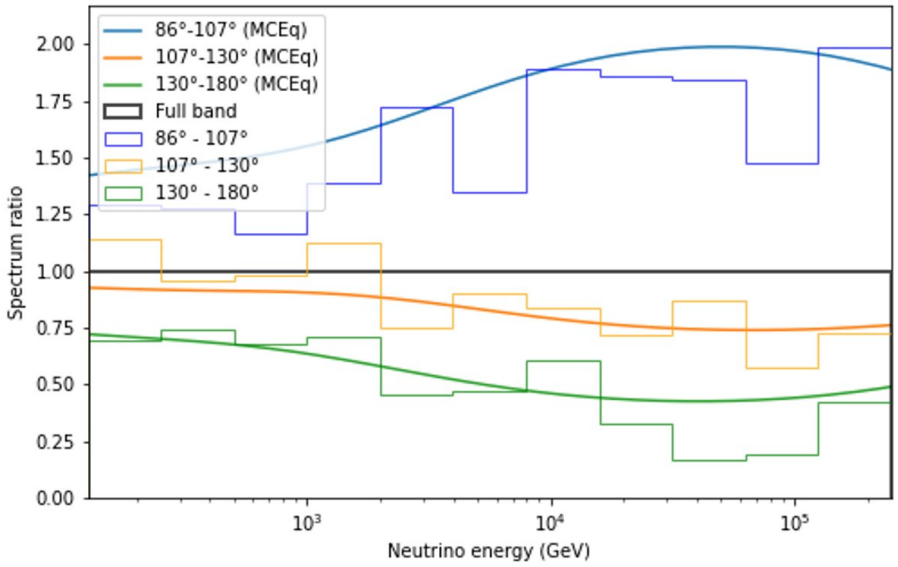

Figure A.8: Ratio of angular flux to the full sky flux from the three-year sample. The unfolded spectra from Figures A.5, A.6, A.7 are divided with the full range unfolding from Appendix Figure A.4 per energy bin. As explained in Section 2.2.2, the low energy region is dominated by the anisotropic conventional component producing the dependancy in the ratio. As the energy crosses into the prompt-dominated region, the ratio decreases. The MCEq [FEG+15] estimates for each band are simulated with the H4a primary model and the SYBILL.3c interaction model, with disregarded astrophysical contribution.

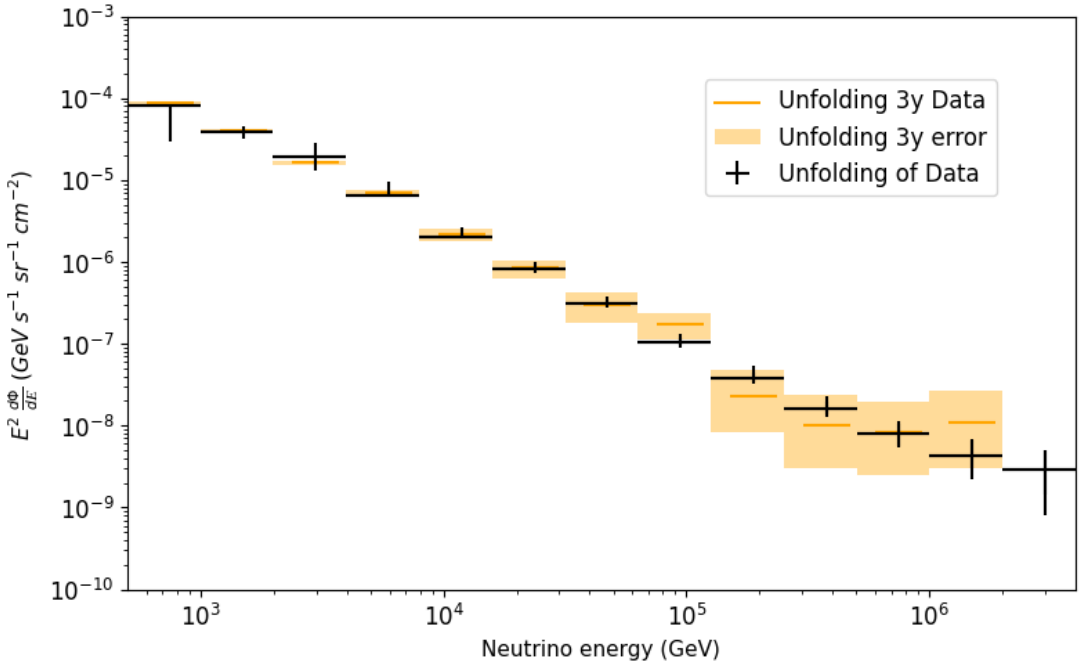

Figure A.9: Weighted flux of the unfolded 11 years data (this work) compared to unfolding of 3 years of data, shown in Appendix Figure A.4. The black solid line shows the unfolding in this work while the orange marks show the previous analysis. The last bin is not comparable as the previous analysis encompassed a shorter energy range. The two fluxes are in agreement inside the previous analysis uncertainty range, shown in the orange shaded area.

to create the same Poissonian error as in the sample, and is given with

$$n_{eff} = \frac{N^2}{\sigma_N^2} = \frac{(\sum_{i=1}^{n} w_i)^2}{\sum_{i=1}^{n} w_i^2} \tag{A.11}$$

The ratio of the weighted simulated events and the effective events is the *effective weight w*

$$w = \frac{T}{n_{eff}} \tag{A.12}$$

which gives the ratio of available events to the minimum events needed to produce comparable errors. Effective weight below one implies statistics better than Poissonian in the given bin. From this, it follows to define the minimum effective lifetime as

$$lifetime_{eff} = \frac{lifetime_{sample}}{w} \tag{A.13}$$

A given simulation can be used to work with a sample observed over a period of time up to the

effective lifetime under the requirement of per bin statistics being Poissonian or better. Therefore, a higher effective lifetime is associated with a superior simulation.

For the full energy range used in this work, with the used simulation and weightings as described in Section 4.2, the effective lifetime and effective weights per considered energy bin are shown in Figure A.10

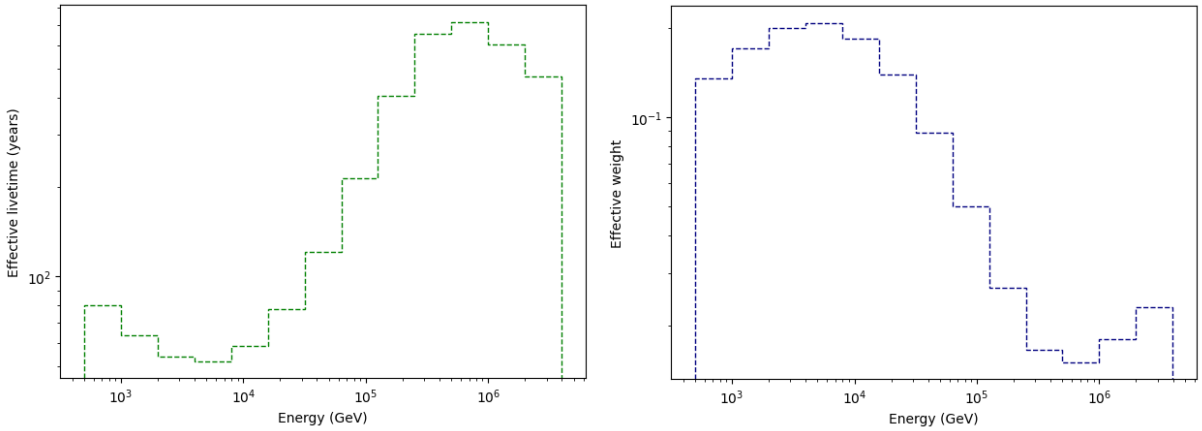

Figure A.10: Effective lifetime (left) and the effective weights (right) for the used energy bins. The values are reciprocal and related by Equation A.13. The simulation set is weighted in the same manner as it is for unfolding in this work.

A.5 Systematic uncertanties

Systematic uncertainties are inherent to scientific experiments, and represent the uncertainties stemming from imperfect knowledge of all the factors influencing a measurement. Unlike statistical uncertainties, which arise from random fluctuations and can be reduced with increased data, systematic uncertainties are associated with known or unknown biases, majorly constituted of limitations in instrumentation, or assumptions made about the detector. In IceCube, several systematic influences are known but not described and cannot be considered straightforward.

Hole ice, the refrozen column of ice in which the DOM strings are embedded, introduces systematic uncertainties related to optical properties. Hole ice is believed to contain residual air bubbles, which can alter the optical properties of the ice, making it more scattering. This introduces uncertainties in the simulation since the scattering properties directly affect photon propagation. The scattering in hole ice furthermore affects the angular sensitivity of the DOMs, as it increases the probability of downgoing light scattering into the PMT. This differs from the behavior of the so-called bulk ice, the glacial ice surrounding the detector components and the refrozen columns. Different models of hole ice $p0$ and $p1$ are considered in the unfolding.

The DOM efficiency represents the number of detected photons relating to the energy deposited in the detector, involving uncertainties both in photon production and detection. The quantum efficiency of the photomultipliers affects the probability of photon detection. Any uncertainties in the fits of quantum efficiency will propagate to DOM efficiency. Furthermore, the efficiency of photon transmission through the glass sphere of the DOM can be affected by factors like surface properties and optical characteristics. Additionally, the presence of detector components can shadow certain areas of the DOMs, reducing their effective light collection and lead to a systematic error, in combination with the uncertainty coming from the descriptions of their positions and geometries. To consider variations of DOM efficiency, values including 90%, 95%, 100%, 105%, and 110% are considered, both in relation to energy and zenith.

Finally, systematic uncertainties include the depth-dependent values of ice scattering and absorption. These values are also varied in 95%, 100%, and 105%, and considered both depending on energy and zenith.

The energy-dependent systematic error per each of the considered factors is given in Appendix Figure A.11. The systematic errors show a bias in the first energy bin, manifesting as an underestimation of the flux for each factor considered. In the remaining energy range, no bias is present and the errors distribute as expected.

Systematic uncertainties are evaluated in the same manner, but in angular regions defined in Section 6.1. Due to the angular-dependent nature of the considered systematic factors, different errors result for the whole range unfolding and the angular studies.

The angular systematic errors are substantially higher than the general systematic error for

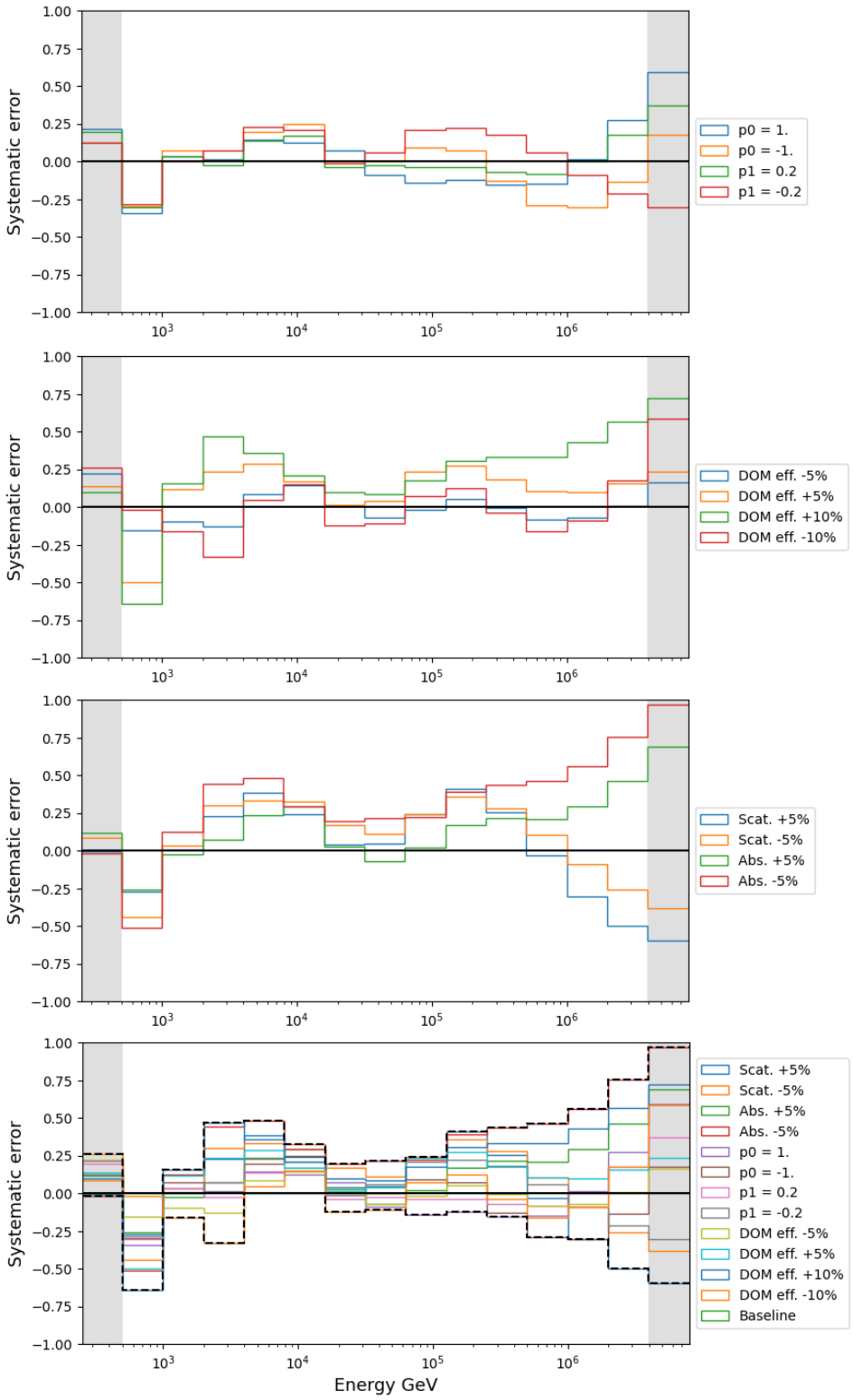

Figure A.11: The systematic error spanning from hole ice (first), DOM efficiency (second), and ice properties (third). All systematic uncertainties are shown (bottom) with the black dashed lines showing their maximum value per energy bin.

the full zenith range considered. The difference between the maximal systematic error of zenith bands and the maximal systematic error of the full range, shown in Appendix Figure A.12, implies an overestimation bias, as the errors do not distribute symmetrically. At the highest energies, the systematic error is dominated by the influence of ice properties. The higher energy region has a dominating upper systematic error, with the middle zenith band exhibiting the highest uncertainty.

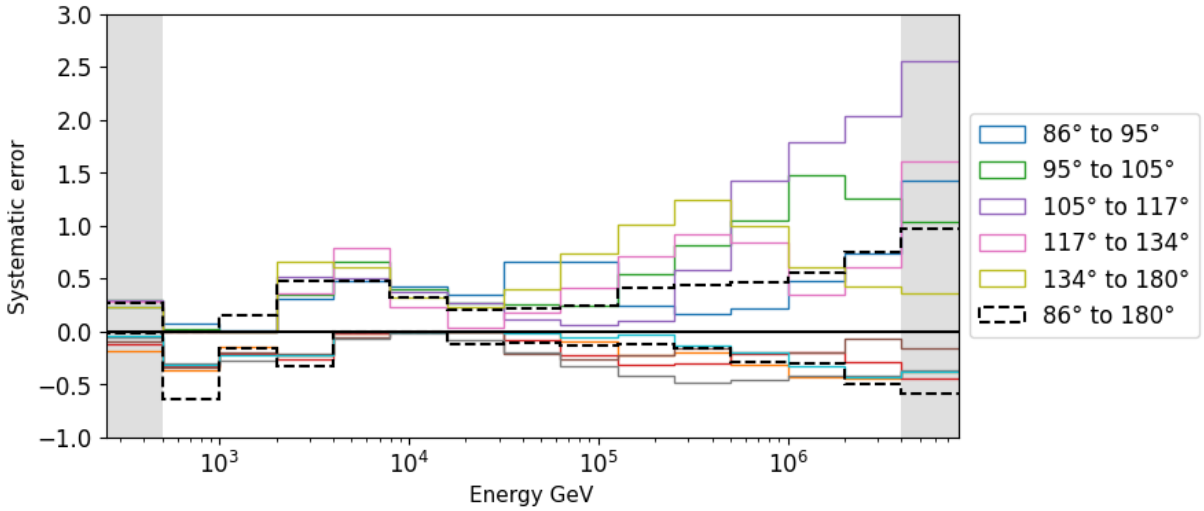

Figure A.12: The maximal systematic error spanning from hole ice, DOM efficiency, and ice properties, in considered angular regions. The maximal systematic uncertainty for the full angular range per energy bin is shown in black dashed lines, as in Appendix Figure A.11.

A.6 Supporting material

The remaining supporting figures and tables relating to the contents of any Section of the presented work are given in this Section with their respective descriptions in captions.

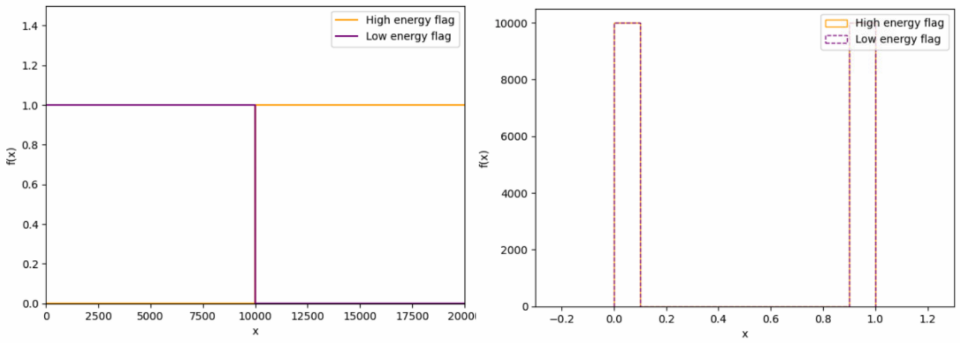

Figure A.13: Comparison of plots of energy flags (left) and their respective histograms (right). The lowest energy for an event to be considered high energy in this example is 10000, and the flags are applied to a generated set of x with balanced values between 0 and 20000. On the right, both histograms of values resulting from two functions are shown, and they fully overlap. This is a result of flags acting as counters of events, therefore producing similar counts for approximately evenly distributed input. Inspecting histograms one might falsely deduce the same underlying logic, rather than the balance of the set. To interpret the similarity of histograms, a proper understanding of features is needed, which is not always feasible for the hundreds of available ones.

Apj833 6y	H4a	4-gen	H3a	default
SIBYLL23C	11.0971	11.0332	11.0683	11.1795
SIBYLL-2.1	11.439	11.3364	11.3984	11.5534
EPOS-LHC	11.3007	11.1914	11.2586	11.4028
DPMJET-III-3.0.6	10.9074	10.9098	10.8895	10.9704
QGSJet-II-04	11.0044	10.9463	10.9721	11.0839

Table A.2: Chi-square statistic of agreement between the unfolded best estimates and the six-year astrophysical fit [A$^+$16a] and different atmospheric models. Columns show different interaction models, while rows represent primary models. Tested models are described in Section 4.1.

HESE 6y	H4a	4-gen	H3a	default
SIBYLL23C	11.356	11.3973	11.324	11.4714
SIBYLL-2.1	11.5097	11.4783	11.4603	11.6174
EPOS-LHC	11.4529	11.436	11.4024	11.5578
DPMJET-III-3.0.6	11.2624	11.419	11.253	11.3563
QGSJet-II-04	11.1897	11.3003	11.1627	11.2763

Table A.3: Chi-square statistic of agreement between the unfolded best estimates and the HESE fit [A$^+$21] and different atmospheric models. Columns show different interaction models, while rows represent primary models. Tested models are described in Section 4.1.

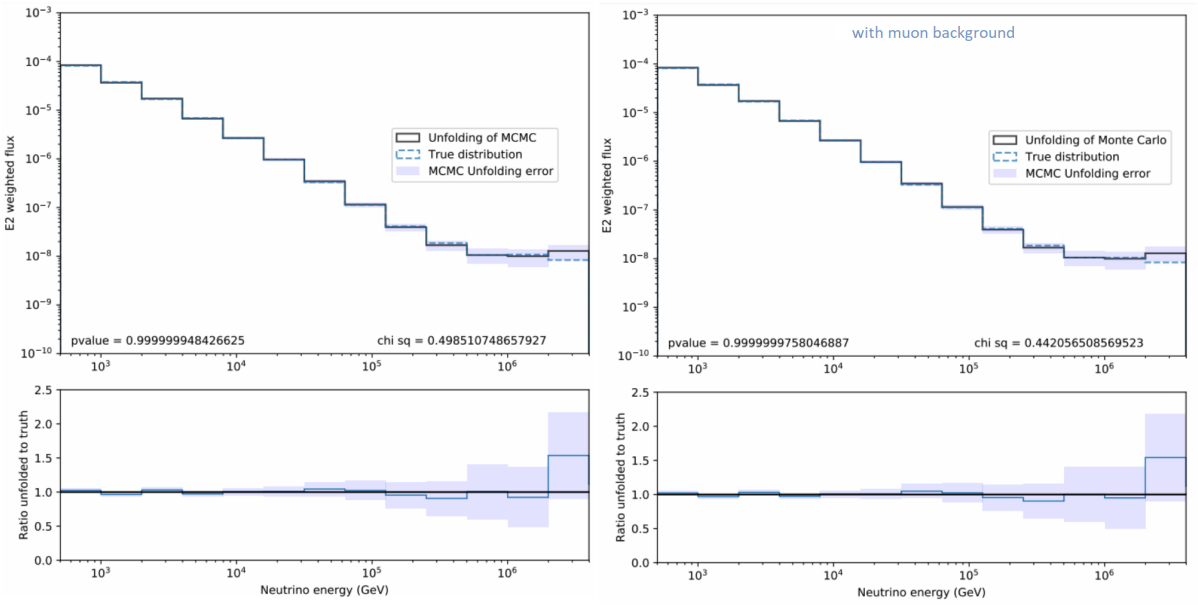

Figure A.14: Weighted flux of the pseudosample (left) and pseudosample with added contamination (right). Events are added following a muon at surface energy distribution at 0.13% to imitate the impurity of the sample described in Section 4.3.3 used in the analysis. There is a slight difference in absolute bin counts, but the unfolded fluxes are comparable inside error bars.

Only atmo.	**H4a**	**4-gen**	**H3a**	**default**
SIBYLL23C	18.2808	17.7452	18.5804	20.7619
SIBYLL-2.1	33.8936	34.0061	33.8781	33.976
EPOS-LHC	33.649	33.7194	33.6479	33.7701
DPMJET-III-3.0.6	14.0361	14.0889	14.0186	13.9873
QGSJet-II-04	33.6899	33.7854	33.6854	33.7213

Table A.4: Chi-square statistic of agreement between the unfolded best estimates and different atmospheric models, with a missing astrophysical contribution. Columns show different interaction models, while rows represent primary models. Tested models are described in Section 4.1.

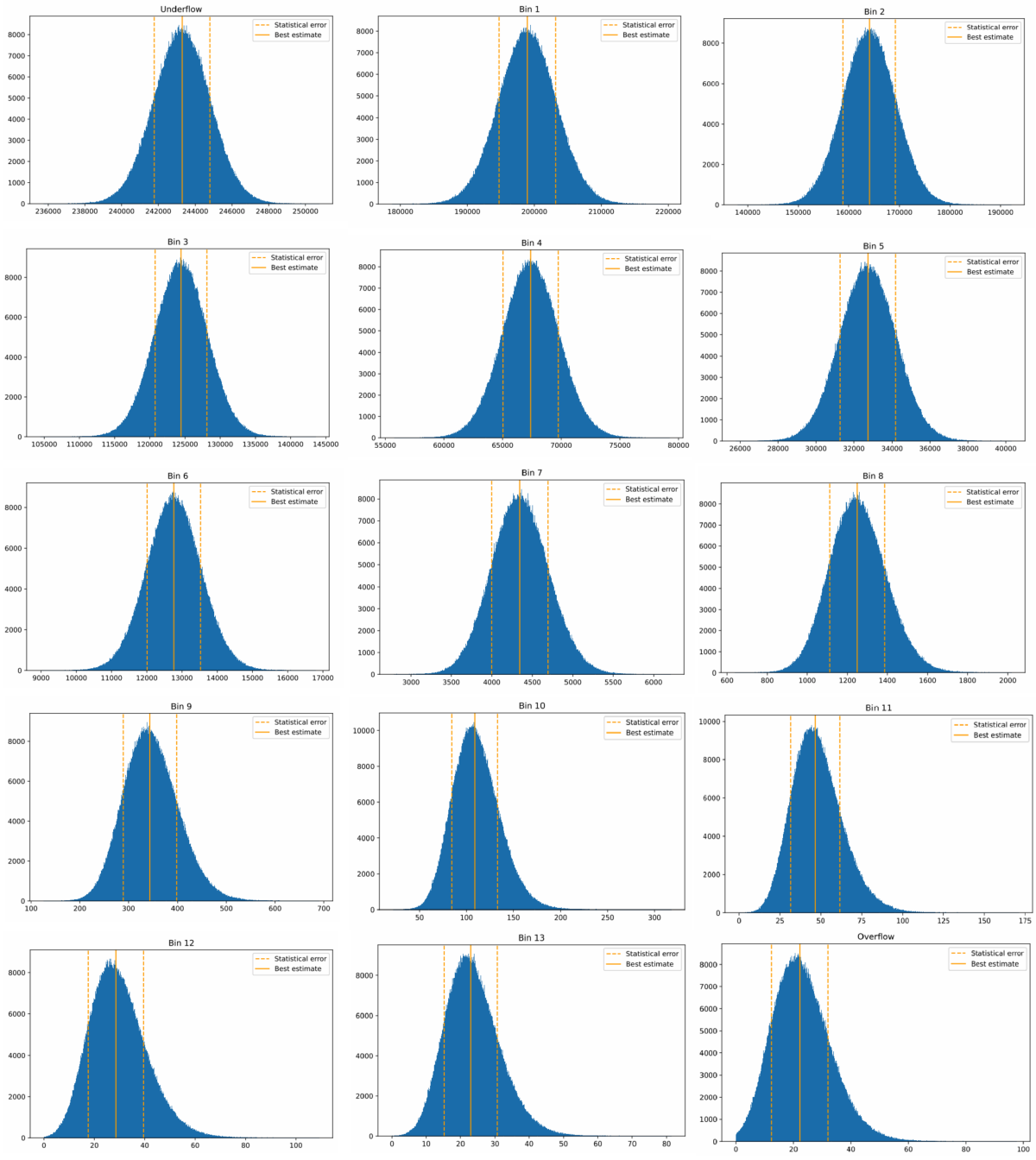

Figure A.15: Distribution of Markov Chain Monte Carlo estimates for all target space bins, including underflow and overflow. Each distribution has a total of 1000000 entries. The Y-axis shows the bin counts in the space of estimates shown on the x-axis. The orange line shows the median of the distribution with the spread in dashed lines.

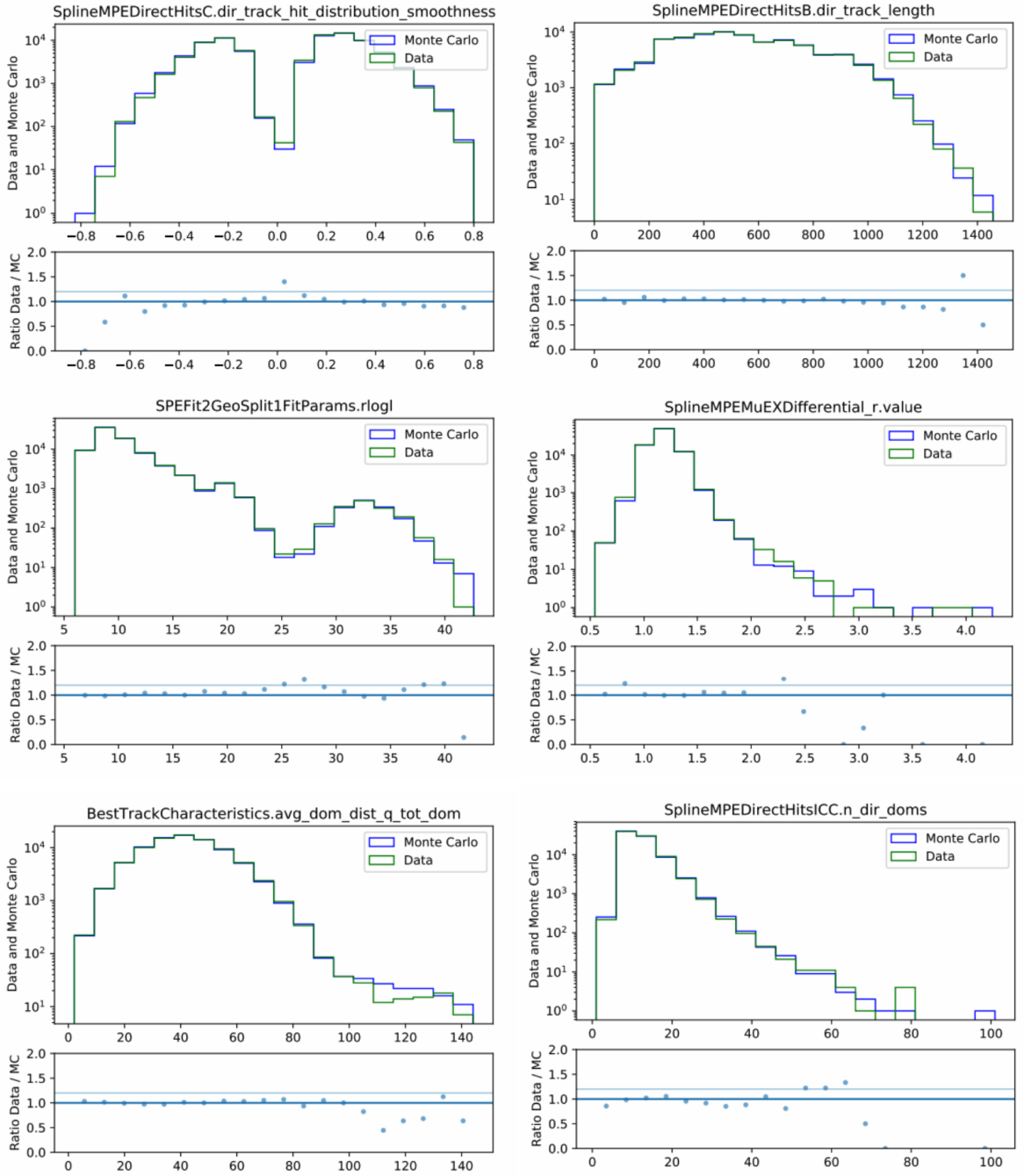

Figure A.16: Data to Monte Carlo agreement of selected features (1). The Monte Carlo simulation is scaled to the size of measurements and weighted with the predictions of the flux. The features were chosen in a three-step selection procedure presented in Section 6.3

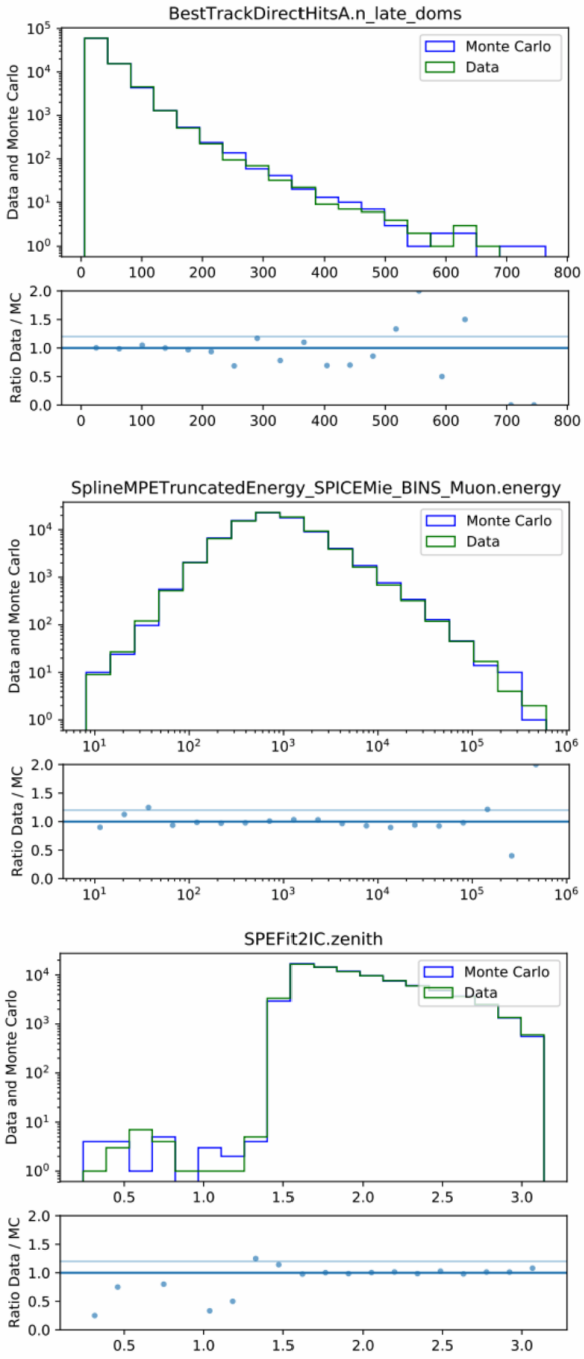

Figure A.17: Data to Monte Carlo agreement of selected features (2). The Monte Carlo simulation is scaled to the size of measurements and weighted with the predictions of the flux. The features were chosen in a three-step selection procedure presented in Section 6.3

.

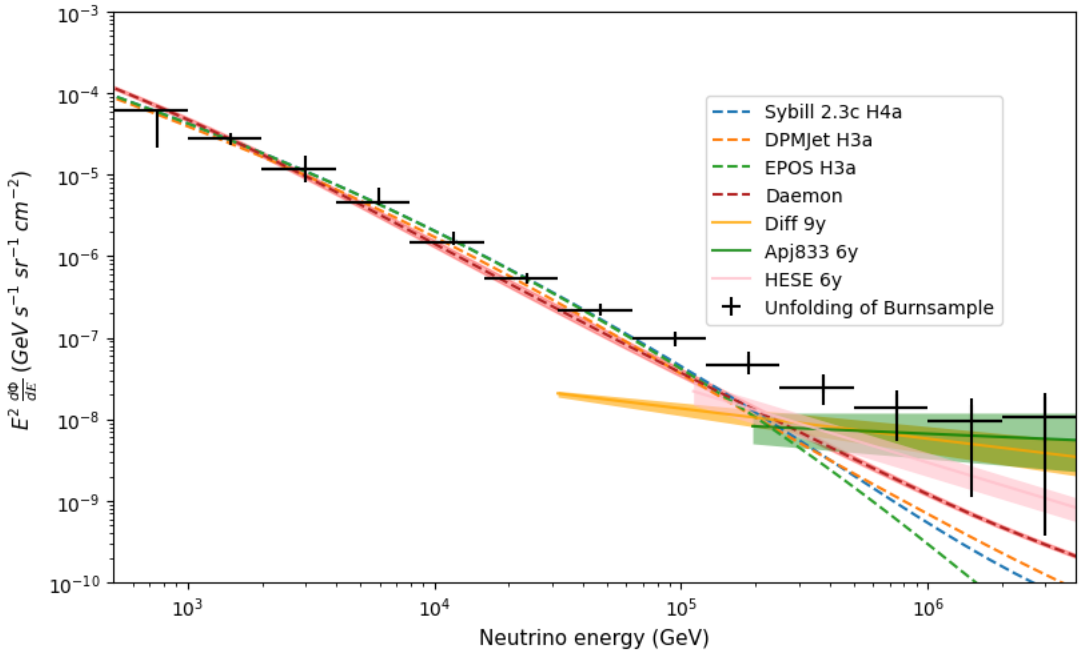

Figure A.18: Weighted flux of the Burnsample data. The black solid lines show the unfolding in this work and the evaluated total uncertainty. The dashed lines show different atmospheric models, with varying interaction models described in Section 4.1. Recent results from IceCube for the astrophysical components are shown in solid lines in their corresponding energy ranges, with the shading showing uncertainty. The assumed astrophysical fit used in the total flux estimation [A$^+$16a] is in green, HESE analysis [A$^+$21] is shown in pink, and 9.5 years diffuse analysis [A$^+$22] in orange.

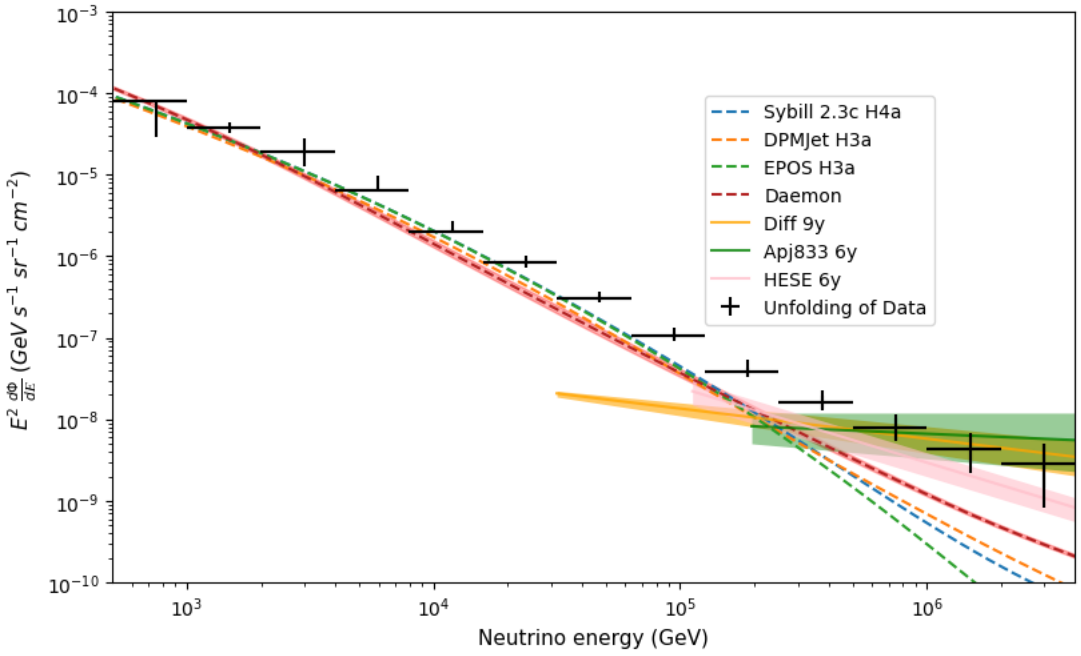

Figure A.19: Weighted flux of the unfolded data. The black solid lines show the unfolding in this work and the evaluated total uncertainty. The dashed lines show different atmospheric models, with varying interaction models described in Section 4.1. Recent results from IceCube for the astrophysical components are shown in solid lines in their corresponding energy ranges, with the shading showing uncertainty. The assumed astrophysical fit used in the total flux estimation [A$^+$16a] is in green, HESE analysis [A$^+$21] is shown in pink, and 9.5 years diffuse analysis [A$^+$22] in orange.

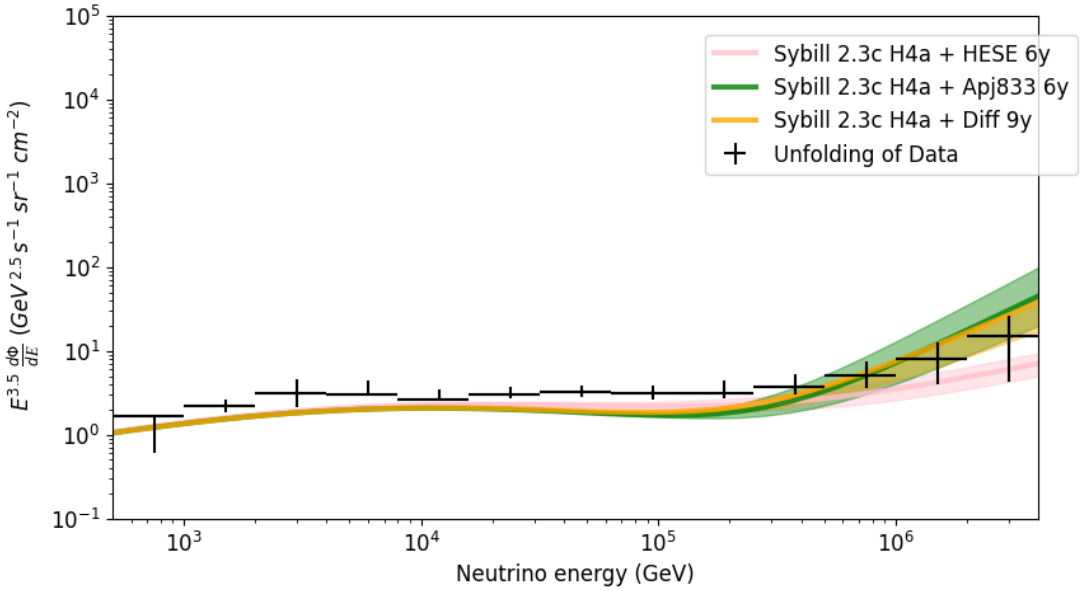

Figure A.20: Flux of the unfolded data weighted with $E^{3.5}$. The black solid line shows the unfolding in this work with vertical lines showing the total uncertainty. The deviation from the horizontal alignment at higher energies implies a break in the spectrum, predicted by the overtaking dominance of the astrophysical component. The assumed astrophysical fit used in the total flux estimation [A$^+$16a] is in green, HESE analysis [A$^+$21] is shown in pink, and 9.5 years diffuse analysis [A$^+$22] in orange. All astrophysical fits are combined with the atmospheric model prediction given by SYBILL 2.3c [DLRF17].

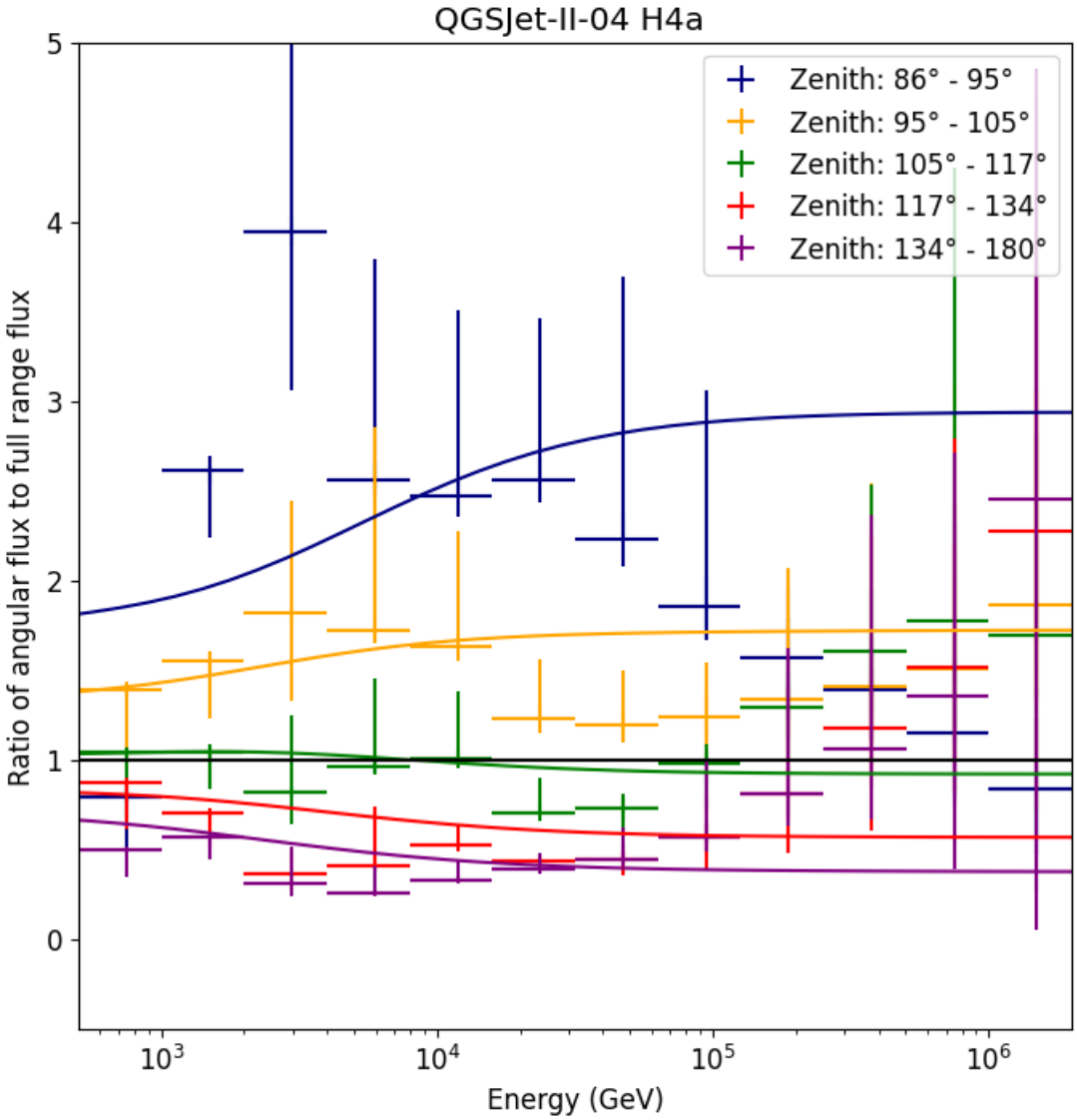

Figure A.21: Unfolded ratios of the five zenith bins to the all range flux. The vertical lines show the total uncertainty of unfolding. The prediction is shown in solid lines, color-coded to correspond to the respective angular ranges. The simulation was made with MCEq [FEG+15] containing only the conventional component following the QGSJet interaction model and the H4a Hillas primary model [Gai12], with no astrophysical contribution. The misalignment between a purely conventional assumption and the angular results is evident.

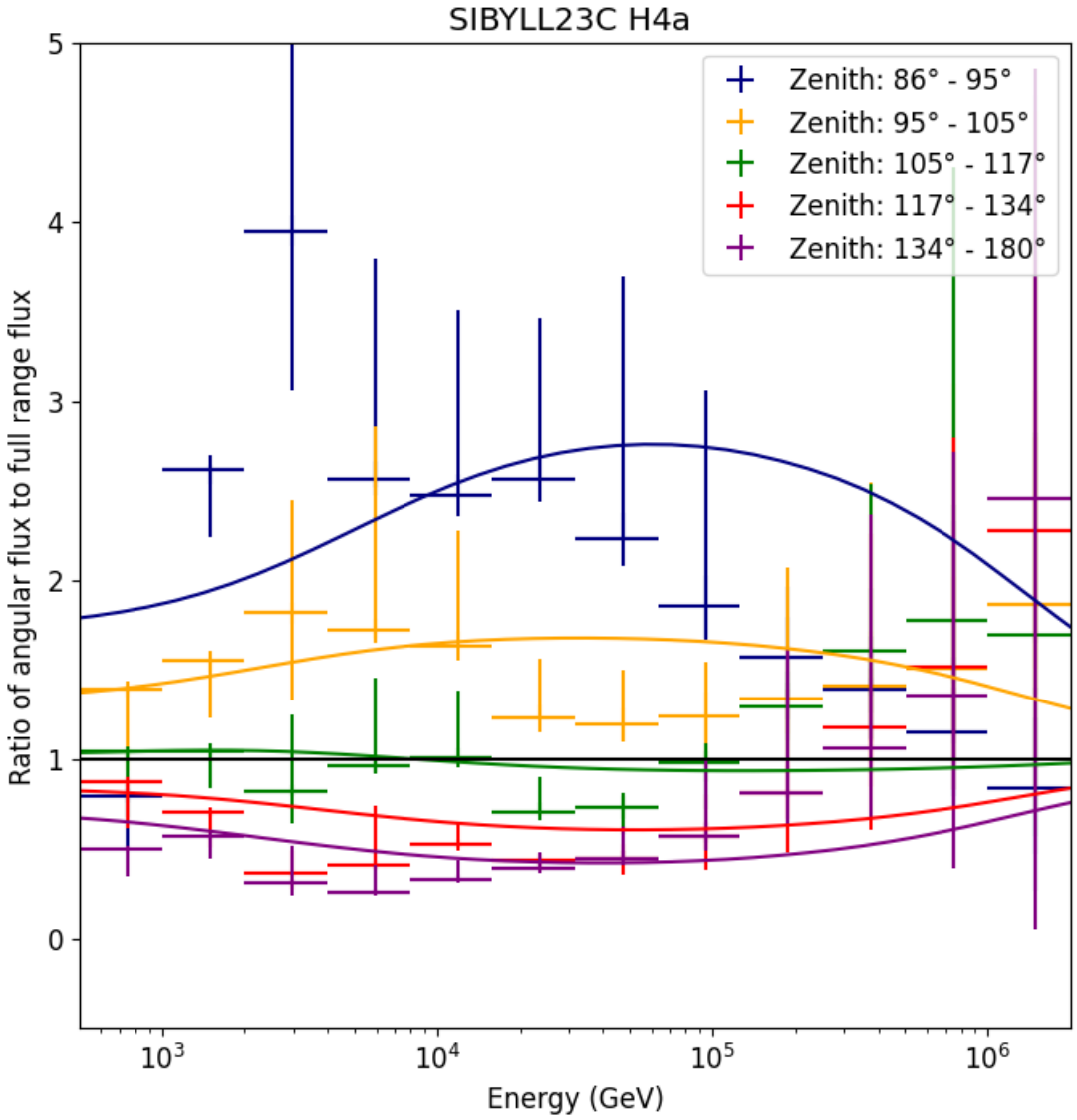

Figure A.22: Unfolded ratios of the five zenith bins to the all range flux. The vertical lines show the total uncertainty of unfolding. The prediction is shown in solid lines, color-coded to correspond to the respective angular ranges. The simulation was made with MCEq [FEG+15] containing only the atmospheric components following the SYBILL 2.3c [DLRF17] interaction model and the H4a Hillas primary model [Gai12], with no astrophysical contribution.

Bibliography

[7419] Teraelectronvolt emission from the -ray burst grb 190114c. *Nature*, 575(7783):455–458, November 2019.

[A+09] R. Abbasi et al. Determination of the atmospheric neutrino flux and searches for new physics with AMANDA-II. *Physical Review D*, 79(10), may 2009.

[A+15a] M. G. Aartsen et al. Development of a general analysis and unfolding scheme and its application to measure the energy spectrum of atmospheric neutrinos with IceCube. *The European Physical Journal C*, 75(3), mar 2015.

[A+15b] M. G. Aartsen et al. Development of a general analysis and unfolding scheme and its application to measure the energy spectrum of atmospheric neutrinos with IceCube. *The European Physical Journal C*, 75(3), mar 2015.

[A+16a] M. G. Aartsen et al. Observation and characterization of a cosmic muon neutrino flux from the northern hemisphere using six years of icecube data. *The Astrophysical Journal*, 833(1):3, dec 2016.

[A+16b] B. P. Abbott et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016.

[A+17a] M. G. Aartsen et al. Constraints on galactic neutrino emission with seven years of icecube data. *The Astrophysical Journal*, 849(1):67, oct 2017.

[A+17b] M. G. Aartsen et al. The IceCube neutrino observatory: instrumentation and online systems. *Journal of Instrumentation*, 12(03):P03012–P03012, mar 2017.

[A+17c] M. G. Aartsen et al. The IceCube realtime alert system. *Astroparticle Physics*, 92:30–41, jun 2017.

[A+17d] M. G. Aartsen et al. Measurement of the ν_μ energy spectrum with IceCube-79. *The European Physical Journal C*, 77(10), oct 2017.

[A+17e] M. G. Aartsen et al. The contribution of fermi-2lac blazars to diffuse tev–pev neutrino flux. *The Astrophysical Journal*, 835(1):45, jan 2017.

[A+21] R. Abbasi et al. IceCube high-energy starting event sample: Description and flux characterization with 7.5 years of data. *Physical Review D*, 104(2), jul 2021.

[A+22] R. Abbasi et al. Improved characterization of the astrophysical muon–neutrino flux with 9.5 years of icecube data. *The Astrophysical Journal*, 928(1):50, mar 2022.

[AHPdlH18] M. Ahlers, K. Helbing, and C. Pérez de los Heros. Probing particle physics with IceCube. *The European Physical Journal C*, 78(11), nov 2018.

[AM+13] S. Adrián-Martínez et al. Measurement of the atmospheric energy spectrum from 100 GeV to 200 TeV with the ANTARES telescope. *The European Physical Journal C*, 73(10), oct 2013.

[BGH+21] C. Bellenghi, T. Glauch, C. Haack, T. Kontrimas, H. Niederhausen, R. Reimann, M. Wolf, and on behalf of the IceCube collaboration. A new and improved icecube pointsource analysis. *Journal of Instrumentation*, 16(11):C11002, nov 2021.

[Blo02] V. Blobel. An unfolding method for high energy physics experiments. 09 2002.

[Bun18] M. Bunse. Dsea rock-solid, regularization and comparison with other deconvolution algorithms, 2018.

[BVE11] B. Baret and Veronique Van Elewyck. High-energy neutrino astronomy: Detection methods and first achievements. *Reports on Progress in Physics*, 74:046902, 03 2011.

[BWM01] L. Boddy, M. Wilkins, and C. Morris. Pattern recognition in flow cytometry. *Cytometry*, 44:195–209, 08 2001.

[Bö18] M. Börner. *Bestimmung des Energiespektrums von atmosphärischen Myonneutrinos mit 3 Jahren Daten des IceCube-Detektors*. PhD thesis, TU Dortmund, 2018.

[C+17] IceCube Collaboration et al. The icecube neutrino observatory - contributions to icrc 2017 part ii: Properties of the atmospheric and astrophysical neutrino flux, 2017.

[CFLM+18] The IceCube Collaboration, Fermi-LAT, MAGIC, AGILE, ASAS-SN, HAWC, H.E.S.S., INTEGRAL, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, Swift/NuSTAR, VERITAS, VLA/17B-403 teams, et al. Multimessenger observations of a flaring blazar coincident with high-energy neutrino icecube-170922a. *Science*, 361(6398):eaat1378, 2018.

[D'A10] G. D'Agostini. Improved iterative bayesian unfolding, 2010.

[DEF+17] H. Dembinski, R. Engel, A. Fedynitch, T. Gaisser, F. Riehn, and T. Stanev. Data-driven model of the cosmic-ray flux and mass composition from 10 gev to 10^{11} gev, 2017.

[Din17] F. B. Diniz. Condition number and matrices, 2017.

[DLRF17] L. Dedenko, A. Lukyashin, T. Roganova, and G. Fedorova. Testing of the venus 4.12, dpmjet 2.55, qgsjet ii-03 and sibyll 2.3 hadronic interaction models via help of the atmospheric vertical muon spectra. *EPJ Web of Conferences*, 158:06006, 01 2017.

[Doc] CRFlux Documentation. Cosmic ray nucleon flux model. Taken from https://crfluxmodels.readthedocs.io/en/latest/module-models.

[DRB+95] K. Daum, W. Rhode, P. Bareyre, R. Barloutaud, G. Chardin, B. Degrange, J. Ernwein, B. Kuznik, H. Meyer, L. Mosca, L. Moscoso, O. Perdereau, M. Schubnell, S. Tisserant, and Y. Wei. Determination of the atmospheric neutrino spectra with the frjus detector. *Zeitschrift für Physik C*, 66:417–428, 09 1995.

[ERF+17] R. Engel, F. Riehn, A. Fedynitch, T. K. Gaisser, and T. Stanev. The hadronic interaction model Sibyll – past, present and future. *EPJ Web Conf.*, 145:08001, 2017.

[FEG+15] A. Fedynitch, R. Engel, T. K. Gaisser, F. Riehn, and T. Stanev. Calculation of conventional and prompt lepton fluxes at very high energy, 2015.

[FPP07] D. Freedman, R. Pisani, and R. Purves. *Statistics: Fourth International Student Edition*. Emersion: Emergent Village Resources for Communities of Faith Series. W.W. Norton & Company, 2007.

[Gai12] T. K. Gaisser. Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio. *Astroparticle Physics*, 35(12):801–806, jul 2012.

[GK05] A. Gazizov and M. Kowalski. ANIS: High energy neutrino generator for neutrino telescopes. *Computer Physics Communications*, 172(3):203–213, nov 2005.

[GST13] T. K. Gaisser, T. Stanev, and S. Tilav. Cosmic ray energy spectrum from measurements of air showers, 2013.

[H+88] K. S. Hirata et al. Observation in the Kamiokande-II Detector of the Neutrino Burst from Supernova SN 1987a. *Phys. Rev. D*, 38:448–458, 1988.

[H+21] K. Hymon et al. Seasonal Variations of the Unfolded Atmospheric Neutrino Spectrum with IceCube. *PoS*, ICRC2021:1159, 2021.

[Hes40] V. F. Hess. The discovery of cosmic radiation. *Thought: Fordham University Quarterly*, 15(2):225–236, 1940.

[HKC+98] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. CORSIKA: A Monte Carlo code to simulate extensive air showers. 2 1998.

[HTWW65] B. K. Harrison, K. S. Thorne, M. Wakano, and J. A. Wheeler. *Gravitation Theory and Gravitational Collapse*. University of Chicago Press, 1965.

[HTWZ19] P. Heix, S. Tilav, C. Wiebusch, and M. Zöcklein. Seasonal variation of atmospheric neutrinos in icecube, 2019.

[Hö03] J. R. Hörandel. On the knee in the energy spectrum of cosmic rays. *Astroparticle Physics*, 19(2):193–220, May 2003.

[IIMS20] Y. Izza, A. Ignatiev, and J. Marques-Silva. On explaining decision trees, 2020.

[IKMY15] J. Iizuka, T. Kitabayashi, Y. Minagawa, and M. Yasuè . Parametrization of pontecorvo–maki–nakagawa–sakata mixing matrix based on CP-violating bipair neutrino mixing. *Modern Physics Letters A*, 30(05):1550019, feb 2015.

[Kno10] G. F. Knoll. *Radiation Detection and Measurement*. Wiley, 2010.

[Kre89] R. Kress. *Tikhonov Regularization*, pages 243–258. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989.

[Mé s17] P. Mé száros. Astrophysical sources of high-energy neutrinos in the IceCube era. *Annual Review of Nuclear and Particle Science*, 67(1):45–67, oct 2017.

[Mil12] N. Milke. *Unfolding of the atmospheric neutrino flux spectrum with the new program TRUEE*. PhD thesis, TU Dortmund, 2012.

[Ngu11] X. Nguyen. Wasserstein distances for discrete measures and convergence in nonparametric mixture models. 01 2011.

[Pon15] W. Van De Pontseele. Studies supporting pmt characterization for the icecube collaboration: Precise photodiode calibration, 2015.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[PW18] A. Palladino and W. Winter. A multi-component model for observed astrophysical neutrinos. *Astronomy &amp Astrophysics*, 615:A168, jul 2018.

[QV15] X. Qian and P. Vogel. Neutrino mass hierarchy. *Progress in Particle and Nuclear Physics*, 83:1–30, 2015.

[REF+20] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev. Hadronic interaction model sibyll 2.3d and extensive air showers. *Physical Review D*, 102(6), 9 2020.

[RVW+19] T. Ruhe, T. Voigt, M. Wornowizki, M. Borner, W. Rhode, and K. Morik. Mining for Spectra - The Dortmund Spectrum Estimation Algorithm. In Marco Molinaro, Keith Shortridge, and Fabio Pasian, editors, *Astronomical Data Analysis Software and Systems XXVI*, volume 521 of *Astronomical Society of the Pacific Conference Series*, page 394, October 2019.

[Rä17] L. Rädel. *Measurement of High-Energy Muon Neutrinos with the IceCube Neutrino Observatory*. Dissertation, RWTH Aachen University, Aachen, 2017. Veröffentlicht auf dem Publikationsserver der RWTH Aachen University; Dissertation, RWTH Aachen University, 2017.

[Saa14] A. Saa, editor. *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, volume 44 of *The Astroparticle Physics Conference*, 2014.

[Sch13a] A. Schukraft. *Search for a diffuse flux of extragalactic neutrinos with the IceCube neutrino observatory*. PhD thesis, Aachen, 2013. Ausgezeichnet mit dem Hertha-Sponer-Preis 2014 der Deutschen Physikalischen Gesellschaft.; Aachen, Techn. Hochsch., Diss., 2013.

[Sch13b] A. Schukraft. A view of prompt atmospheric neutrinos with IceCube. *Nuclear Physics B - Proceedings Supplements*, 237-238:266–268, apr 2013.

[Sol19] D. Soldin. Atmospheric muons measured with icecube. *EPJ Web Conf.*, 208:08007, 2019.

[Spa13] F. Spanò. Unfolding in particle physics: A window on solving inverse problems. *EPJ Web of Conferences*, 55:03002–, 07 2013.

[Sto05] R. Stokstad. Design and performance of the IceCube electronics. In *11th Workshop on Electronics for LHC and Future Experiments (LECC 2005)*, page 4, 9 2005.

[vRCB16] D. van Ravenzwaaij, P. Cassey, and S. Brown. A simple introduction to markov chain monte–carlo sampling. *Psychonomic Bulletin Review*, 25, 03 2016.

[Wue11] K. L. Wuensch. *Chi-Square Tests*, pages 252–253. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[ZS06] V. I. Zatsepin and N. V. Sokolskaya. Three component model of cosmic ray spectra from 10 GeV to 100 PeV. *Astronomy &amp Astrophysics*, 458(1):1–5, oct 2006.

Acknowledgments

My great gratitude goes to my supervisor Prof. Dr. Dr. Wolfgang Rhode for the tremendous support throughout my time as a PhD student in Dortmund. I appreciate the interest in my research, the encouragement to progress, and the creation of opportunities for me to travel and develop as a scientist.

I am grateful to Dr. Dominik Elsasser for always being helpful, having kind words and great advice.

I would like to thank the whole E5b group for making me feel welcome and for engaging in great scientific discussions. I appreciate every conversation and group meeting exchanged in English. I am thankful to the IceCube group for helping me develop this work, and teaching me all the peculiarities of running the analysis in our collaboration, especially at the start of my journey.

My gratitude goes to the IceCube collaboration, for enabling everything needed for my PhD dissertation and the supervision of my work. This could not be possible without the considerate cooperation of all the members. I would like to thank the whole IceCube group in Madison, Wisconsin for hosting my research stay. During my time in the States, I have learned so much and improved both this work and my understanding of astrophysics in general.

I would like to especially thank Paolo Desiati. I am grateful for exciting scientific conversations and new ideas, the support of my work during and after my research stay, and for joining me on the quest of enjoying only the best espressos Madison has to offer.

My research has been supported by Collaborative Research Center 876 ”Providing Information by Resource-Constrained Data Analysis” and Collaborative Research Center 1491 ”Cosmic interacting matters”. Special thanks to The Lamarr Institute for Machine Learning and Artificial Intelligence for funding this work and various travels to scientific conferences.

I would like to thank my friends and family for teaching me invaluable things in life. I would like to thank my friends in Dortmund for making my stay ever so more enjoyable, and my friends back home for love and the understanding of my sometimes hectic life schedule.

To Antea, for showing me what a selfless friendship is, and for every late-night conversation filled with either laughter or tears.

To Mario, for showing me what love is. I truly believe there is no other person who can complete me so impeccably and make every moment shared together nothing short of perfect.

To my sister Lara and to my Lorena, Karlo, and Zoe, without whom this thesis would have been finished way earlier, but nonetheless the journey would be much less joyful. I hope I can one day inspire the kids to make something of their own.

Finally, my greatest appreciation is for my mother Aleksandra. Without her support, love, and kindness, I would not be the person and the scientist I am today. I hope this work and everything I do in the future will continue to make her proud.