

Arbeit zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
(Dr. rer. nat.)

Charmante Myonen im Eis

**Messung des hochenergetischen atmosphärischen
Myon-Energiespektrums mit IceCube
in der Detektorkonfiguration IC86-I**

Tomasz Fuchs
geboren in Ruda Śląska

2016

Lehrstuhl für Experimentelle Physik V
Fakultät Physik
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Dr. Wolfgang Rhode
Zweitgutachterin: Prof. Dr. Julia Tjus
Abgabedatum: 30. Juni 2016

Kurzfassung

In dieser Dissertation wird der Fluss hochenergetischer Myonen in einem Energiebereich von 10 TeV bis 1 PeV mit IceCube-Daten vom 13.05.2011 bis zum 15.05.2012 rekonstruiert und analysiert. Für diesen Zweck werden aus einem Datensatz mit Myon-Bündel-Ereignissen diejenigen Ereignisse separiert, bei denen ein Myon die anderen energetisch dominiert. Zur Separation wird ein Random-Forest-Modell verwendet, welches einen Datensatz hochenergetischer Myonen mit einer Effizienz von $(40,8 \pm 0,6) \%$ und einer Reinheit von $(93,1 \pm 0,4) \%$ erzeugt. Die Attribute für diese Separation werden durch den mRMR-Algorithmus bestimmt. Aus diesem Datensatz wird durch eine regularisierte Entfaltung mit der Software TRUEE das Energiespektrum der Myonen rekonstruiert. Die Hypothese einer prompten und konventionellen Komponente atmosphärischer Myonen ergibt Flussnormalisierungen von $N_{\text{Konv.}} = 1,03 \pm 0,06$ und $N_{\text{Prompt}} = 1,59 \pm 1,57$. Aufgrund der hohen Unsicherheit in der prompten Flussnormalisierung kann ein rein konventioneller Fluss nicht ausgeschlossen werden. Wird der Anteil an flavorfreien und charmhaltigen Beiträgen zur prompten Komponente als korrekt angenommen, so ergibt sich ein p-Wert von 0,045 für eine Überfluktuation der atmosphärischen Komponente oberhalb von 60 TeV des gemessenen IceCube-Neutrino fluxes. Dieser p-Wert impliziert eine astrophysikalische Komponente.

Abstract

In this thesis the flux of high-energy muons in the energy regime from 10 TeV to 1 PeV is reconstructed and analyzed using data collected with the IceCube detector in the time span 13.05.2011 to 15.05.2012. From a data set containing muon bundles only those events are selected which contain a muon that is energetically dominating the others in the bundle. For the separation a Random Forest model is applied, resulting in a data set of high-energy muons with an efficiency of $(40,8 \pm 0,6) \%$ and a purity of $(93,1 \pm 0,4) \%$. Attributes considered in the separation are selected by the mRMR algorithm. The energy spectrum of muons is reconstructed with a regularized unfolding using the software TRUEE. The hypothesis of a prompt and a conventional component of atmospheric muons results in flux normalizations of $N_{\text{conv.}} = 1,03 \pm 0,06$ and $N_{\text{prompt}} = 1,59 \pm 1,57$. Due to the large uncertainty of the prompt component, the hypothesis of a pure conventional flux cannot be excluded. Using these normalizations, it is possible to determine if the measured high-energy neutrino flux above 60 TeV is of atmospheric origin. The p-value for this hypothesis is found to be 0,045, which indicates the need of an astrophysical component to explain the excess at high energies.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung | 1 |
| 2 | Kosmische Strahlung | 3 |
| 2.1 | Ursprung der kosmischen Strahlung | 3 |
| 2.2 | Primäre kosmische Strahlung | 5 |
| 2.3 | Sekundäre kosmische Strahlung | 10 |
| 2.4 | Prompte Myonen | 11 |
| 2.5 | Wechselwirkungen von Myonen | 13 |
| 3 | IceCube | 17 |
| 3.1 | Allgemein | 17 |
| 3.2 | Detektionsprinzip | 18 |
| 3.3 | Ziele | 21 |
| 4 | Signalklassifikation und -rekonstruktion | 23 |
| 4.1 | Signalklassifikation | 23 |
| 4.2 | Signalrekonstruktion | 36 |
| 5 | Verwendete Datensätze und Ereignisseigenschaften | 41 |
| 5.1 | Verwendete Datensätze | 41 |
| 5.2 | Ereignisklassen | 44 |
| 5.3 | Algorithmen zur Ereignisrekonstruktion | 48 |
| 5.4 | DDDDR-Algorithmus | 50 |
| 6 | Rekonstruktion des Myonspektrums | 53 |
| 6.1 | Qualitätsschnitte | 54 |
| 6.2 | Attributauswahl | 56 |
| 6.3 | Separation | 63 |
| 6.4 | Analyse des finalen Datensatzes | 72 |
| 6.5 | Neutrinountergrund | 77 |
| 6.6 | Rekonstruktion des Myonspektrums | 78 |
| 6.7 | Resultate im Kontext der Neutrinoastronomie | 89 |
| 7 | Zusammenfassung und Ausblick | 91 |

Inhaltsverzeichnis

| | |
|-------------------|------------|
| Literatur | 95 |
| Danksagung | 101 |

1 Einleitung

Die Astronomie begleitet die Menschheit bereits seit Jahrtausenden. Lange wurden Antworten auf physikalische Fragen gesucht, indem Objekte am Himmel mit optischen Teleskopen untersucht wurden. Dieses Gebiet erfuhr eine Erweiterung mit der Entdeckung der atmosphärischen Höhenstrahlung im Jahr 1912 von Viktor Hess[1]. Durch diese Entdeckung wurde der Grundstein für die Astroteilchenphysik gelegt. Dieses Teilgebiet der Physik beschäftigt sich mit Teilchen, welche im Universum erzeugt worden sind. Heutzutage werden Photonen im Bereich von niederenergetischer Radiostrahlung bis zu hochenergetischer Gammastrahlung mit modernen Teleskopen untersucht. Auch die jüngst detektierten Gravitationswellen[2] durch das Experiment LIGO am 14. September 2015 liefern grundlegende Erkenntnisse über das Universum[3]. Ein weiterer junger Zweig der Astronomie ist die Untersuchung von Neutrinos, welche im Jahr 2013 den Beweis lieferte, dass es Quellen hochenergetischer Neutrinos außerhalb unseres Sonnensystems geben muss. Die astrophysikalischen hochenergetischen Neutrinos werden von atmosphärischen Neutrinos überlagert und dominieren diese ab einer Energie von etwa 100 TeV. Der Fluss von atmosphärischen Myonenneutrinos besteht aus einer konventionellen Komponente, welche durch den Zerfall von Pionen und Kaonen entsteht, und einer unbekannt prompten Komponente von hochenergetischen Neutrinos aus dem Zerfall kurzlebiger Teilchen. Die Neutrinos aus der prompten Komponente beginnen, wie auch die astrophysikalischen, bei ca. 100 TeV relevante Beiträge zum atmosphärischen Neutrinofluss zu liefern. Daher ist das Verständnis der Entstehungsprozesse dieser prompten Neutrinos von großer Bedeutung, damit die Beiträge aus astrophysikalischen Neutrinos genauer gemessen werden können. Da bei der Produktion von Neutrinos auch geladene Leptonen erzeugt werden, können auch die Myonen untersucht werden und dadurch Rückschlüsse auf die Neutrinoflüsse gezogen werden.

Die vorliegende Arbeit hat das Ziel das atmosphärische Myonspektrum zu rekonstruieren. Zu diesem Zweck wird der IceCube-Detektor verwendet und die erste Datennahme-Periode der vollständigen Konfiguration des Detektors des Jahres 2011 untersucht. Mit Hilfe von aktuellen statistischen Methoden der Datenanalyse sollen hochenergetische Myonen selektiert und ihr Energiespektrum entfaltet werden. Auf diese Weise sollen Normalisierungen für die prompte und konventionelle Komponente des atmosphärischen Myonenflusses bestimmt werden. Die Übertragung dieser

Ergebnisse auf die Neutrinoastronomie soll zeigen, dass der Überschuss an hochenergetischen Neutrinos astrophysikalisch ist und nicht durch die prompte Komponente des atmosphärischen Flusses erklärt werden kann.

Die Inhalte dieser Arbeit sind wie folgt strukturiert:

In [Kapitel 2](#) wird die kosmische Strahlung erläutert. In diesem Zuge werden zum einen die Quellen der Strahlung und zum anderen die Teilchenarten betrachtet, welche die Erde erreichen. Abschließend findet eine Erläuterung der Myonwechselwirkungen statt.

Die Grundlagen des IceCube-Detektors werden in [Kapitel 3](#) erläutert. Sowohl allgemeine Eigenschaften, das Detektionsprinzip als auch die Ziele des Detektors werden in diesem Kapitel beleuchtet.

Statistische Methoden der Datenanalyse werden in [Kapitel 4](#) beschrieben. Die verwendeten Methoden in dieser Arbeit umfassen eine Attributauswahl, eine Separation und eine Spektrumsrekonstruktion.

Die dieser Arbeit zugrundeliegenden Datensätze werden in [Kapitel 5](#) diskutiert. Ereigniseigenschaften, vorhandene Simulationsdatensätze und zur Verfügung stehende Rekonstruktionsalgorithmen werden in diesem Kapitel erläutert.

Die Ergebnisse dieser Arbeit und die daraus resultierenden Schlussfolgerungen sind in [Kapitel 6](#) aufgeführt. In diesem Kapitel werden die Normalisierungen für den prompten und konventionellen Myonfluss bestimmt und in Hinblick auf vorhandene Neutrinomessungen betrachtet.

Abschließend wird in [Kapitel 7](#) eine Zusammenfassung und ein Ausblick gegeben. Im Ausblick wird auf die Auswirkung bei der Analyse von Daten aus nachfolgenden Datennahme-Perioden und mit möglichen Detektorerweiterungen des IceCube-Detektors eingegangen.

2 Kosmische Strahlung

Für die Analyse hochenergetischer Myonen und generell in der Astroteilchenphysik ist das Verständnis der kosmischen Strahlung von großer Bedeutung. Wenn die kosmische Strahlung die Erdatmosphäre erreicht, kann diese in eine primäre und eine sekundäre Komponente unterteilt werden. Die primäre Komponente beschreibt alle Teilchen, welche die Erdatmosphäre erreichen und einen Luftschauer erzeugen können. Als sekundäre Komponente gelten alle Teilchen, welche von dem Primärteilchen während der Propagation zur Erdoberfläche produziert werden. Diese sekundären Teilchen können mit Teleskopen auf der Erde detektiert werden. Im Speziellen wird auf die produzierten Myonen in solchen Luftschauern eingegangen. Weiterhin werden die Energieverluste von Myonen beschrieben, da diese für die vorliegende Arbeit relevant sind.

2.1 Ursprung der kosmischen Strahlung

Die kosmische Strahlung wird durch verschiedene Quellen im Universum produziert. Hierbei können die hochenergetischen Teilchen, welche die Erde erreichen, in eine galaktische und eine extragalaktische Komponente aufgeteilt werden. Zunächst wird die Erzeugung geladener Teilchen und anschließend die Produktion von hochenergetischen Photonen und Neutrinos erläutert.

Niederenergetische kosmische Strahlung ($E < 50 \text{ PeV}$) kann in Supernovaüberresten (SNR¹) produziert werden[4, 5]. In einer Supernova kommt es zu einer Schockbeschleunigung, wenn die Schockfront auf das interstellare Medium trifft. Innerhalb der Schockfront und des interstellaren Mediums sind turbulente Magnetfelder vorhanden, wodurch die Teilchen abgelenkt werden können. Die hochenergetischen Teilchen der Schockfront können durch die turbulenten Magnetfelder mehrmals die Ebene zwischen Schockfront und dem interstellaren Medium passieren. Durch diesen stochastischen Prozess entsteht ein Energiespektrum, welches einem Potenzgesetz folgt und auch durch Beobachtungen von Quellen bestätigt werden konnte. Dieser Prozess der Beschleunigung wird als Fermi-Beschleunigung erster Ordnung bezeichnet, da

¹Supernova remnant

Teilchen, welche den Schock passieren, bei jedem Durchgang einen Energiezuwachs $\Delta E \propto (v/c_0)^1$ erfahren. Hierbei ist v die Geschwindigkeit eines Teilchens und c_0 gibt die Vakuumlichtgeschwindigkeit an.

Für extrem hochenergetische Teilchen ($E > 50 \text{ EeV}$) kann eine Beschleunigung nicht mehr in der Galaxis stattfinden. Der Grund hierfür ist, dass der Lamor-Radius der geladenen Teilchen der Größenordnung der Galaxis entspricht, sodass diese die Galaxis verlassen würden bevor sie zu noch höheren Energien beschleunigt werden können. Kandidaten für die Produktion hochenergetischer kosmischer Strahlung sind aktive galaktische Kerne[6], welche auch AGN² genannt werden. Diese aktiven galaktischen Kerne haben ein massives schwarzes Loch in ihrem Zentrum und sind von einem Staubtorus umgeben. Durch die Akkretion von Masse durch das schwarze Loch entsteht ein Jet, welcher senkrecht zur Torusebene des Staubes orientiert ist. In diesem Jet sind diffuse Magnetfelder vorhanden, die durch Fermi-Beschleunigung erster Ordnung Teilchen zu höheren Energien beschleunigen können. Eine weitere Möglichkeit hochenergetische kosmische Strahlung zu produzieren ist die Fermi-Beschleunigung zweiter Ordnung, jedoch ist dieser Prozess ineffizient, da sie nur einen Energiezuwachs proportional zu $(v/c_0)^2$ liefert[7].

Durch die Beschleunigung von Elektronen und die Ablenkung in magnetischen Feldern können hochenergetische Photonen durch Synchrotronstrahlung entstehen. Dieser Prozess produziert Photonen in der Quelle mit Energien von einigen GeV bis TeV. Höherenergetische Photonen können mit den anwesenden Elektronen über den inversen Compton Prozess zusätzliche Energie aufnehmen. Dadurch ist es möglich die Photonen bis zu einer Energie von $E \approx 100 \text{ TeV}$ zu beschleunigen. Weiterhin können hochenergetische Photonen aus Zerfällen des π^0 -Mesons in Proton-Proton- oder Proton-Photon-Wechselwirkungen erzeugt werden.

Treten hadronische Wechselwirkungen in der Quelle auf, können auch Neutrinos erzeugt werden. Bei diesen Wechselwirkungen können geladene Mesonen entstehen, die wiederum in Neutrinos zerfallen können. Die geladenen Mesonen mit dem größten Beitrag zu diesen Neutrinos liefern π^\pm und K^\pm . In Gleichung (2.1.1) wird der Prozess der Neutrinoproduktion für π^- aufgeführt. Dieser Prozess verläuft für die weiteren geladenen Mesonen analog und lässt sich durch die Zerfallsgleichung

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \rightarrow e^- + \bar{\nu}_e + \nu_\mu + \bar{\nu}_\mu \quad (2.1.1)$$

beschreiben. Aufgrund der Distanzen zwischen der Erde und einer Quelle (z.B. mehrere hundert Lichtjahre im Falle von AGNs[8]) lässt sich wegen der Neutrinooszillation nicht auf das ursprünglich in der Quelle produzierte Flavor-Verhältnis schließen. Von Quellen hochenergetischer Neutrinos wird somit ein Flavor-Verhältnis

²Active Galactic Nucleus

$\nu_\tau : \nu_\mu : \nu_e$ von 1 : 1 : 1 erwartet. Ein Hinweis auf dieses Flavor-Verhältnis wurde mit dem IceCube-Detektor gemessen[9].

2.2 Primäre kosmische Strahlung

Die kosmische Strahlung, welche die Erde erreicht, lässt sich in Photonen, Neutrinos und die geladene kosmische Strahlung unterteilen. Diese drei Strahlungsarten zeigen beim Eintritt in die Atmosphäre ein unterschiedliches Verhalten und auch die Detektionsmethoden sind verschieden. Mit Hilfe des IceCube-Detektors lassen sich prinzipiell alle Teilchenarten der kosmischen Strahlung detektieren, aufgrund dessen wird in diesem Abschnitt auf die primäre kosmische Strahlung eingegangen.

2.2.1 Photonen

Das wohl am längsten bekannte Teilchen der primären kosmischen Strahlung ist das Photon. Photonen im optischen Bereich werden kaum von der Atmosphäre absorbiert und erreichen somit den Erdboden. Aus diesem Grund können diese direkt mit optischen erdgebundenen Teleskopen detektiert werden. Im zwanzigsten Jahrhundert wurde begonnen einen Großteil des Spektrums der Photonen zu untersuchen. In der aktuellen Forschung wird das Spektrum von der Radiostrahlung bis hin zur hochenergetischen Gammastrahlung vermessen.

Radioastronomie wird primär mit bodengestützten Teleskopen durchgeführt, da die Absorption in der Atmosphäre für diese Wellenlängen gering ist. Die Untersuchung von Photonen im optischen Bereich geschieht mit Hilfe von erdgebundenen Teleskopen und auch Satelliten werden zur Analyse dieses Frequenzbereiches genutzt. Zur Detektion von hochenergetischen Photonen werden Satelliten-Experimente verwendet. Diese Experimente sind allerdings aufgrund ihrer geringen Detektionsfläche von einigen Quadratmetern[10] nicht im Stande den Fluss der höchstenergetischen Photonen signifikant zu detektieren, da dieser Fluss vergleichsweise gering ist. Eine andere Methode der Detektion in diesem Energiebereich ist die Verwendung von erdgebundenen Cherenkov-Teleskopen. Hochenergetische Gammastrahlung erreicht nicht den Erdboden, sondern wechselwirkt mit den Atomen der Atmosphäre. Dadurch werden Teilchenschauer ausgelöst, welche wiederum Cherenkov-Licht produzieren, das anschließend von den Teleskopen detektiert wird. Die Detektion der höchstenergetischen Photonen geschieht somit indirekt.

Hochenergetische Gammastrahlung wechselwirkt allerdings mit dem extragalaktischen Hintergrundlicht, wodurch eine Absorption bereits ab Energien von

$E \approx 100 \text{ GeV}$ auftritt[11]. Bei der Untersuchung hochenergetischer Photonquellen, wie aktiver galaktischer Kerne, muss aus diesem Grund beachtet werden, dass eine Abschwächung des gemessenen Flusses aufgrund der Propagation durch den interstellaren Raum auftritt.

2.2.2 Neutrinos

Die Neutrinoastronomie ist das jüngste Forschungsgebiet bei der Untersuchung der primären kosmischen Strahlung. Neutrinos wechselwirken nur schwach mit Materie, wodurch große Detektionsvolumen nötig sind[12]. Die erste Detektion extraterrestrischer Neutrinos fand bei der Untersuchung von solaren Neutrinos statt, welche Energien von einigen MeV aufweisen[13]. Mit der Detektion solarer Neutrinos wurde entdeckt, dass ausgesendete Neutrino flavor in andere Neutrino flavor übergehen können, da ihre Masseneigenzustände und Flavoreigenzustände nicht identisch sind[14].

In den vergangenen Jahrzehnten ist der Fluss atmosphärischer Neutrinos extensiv untersucht worden[15–18]. Diese Untersuchungen dienten dazu Quellen hochenergetischer astrophysikalischer Neutrinos zu identifizieren. Der erste Nachweis von hochenergetischen astrophysikalischen Neutrinos ist im Jahr 2013 von der IceCube-Kollaboration geliefert worden[9]. Die Anzahl der nicht-atmosphärischen Ereignisse ist aktuell allerdings zu gering, um Quellen oder Quelltypen dieser Neutrinos eindeutig bestimmen zu können.

2.2.3 Geladene kosmische Strahlung

Die geladene kosmische Strahlung besteht aus Atomkernen, die durch den interstellaren Raum propagiert sind. Aufgrund dieser Propagation und der großen Propagationsstrecke erreichen meist nur stabile Kerne die Erde. Instabile Kerne können durch Spallationsprozesse in geringer Entfernung zur Erde erzeugt werden und so die Erdatmosphäre erreichen[19]. Das Spektrum der geladenen kosmischen Strahlung ist in Abbildung 2.1 für Primärteilchenenergien oberhalb von 10 TeV dargestellt. Für geringe Energien wird das Spektrum direkt über Satelliten- oder Ballon-Experimente untersucht. Für höhere Energien $E > 100 \text{ TeV}$ ist der Fluss der geladenen kosmischen Strahlung zu gering, sodass eine indirekte Detektion über Luftschauer geschieht[20–23].

Das Spektrum der geladenen kosmischen Strahlung folgt einem gebrochenen Potenzgesetz, welches drei markante Stellen aufweist. Die erste Veränderung des spektralen Index findet bei einer Energie von $E \approx 3 \text{ PeV}$ statt und wird als Knie bezeichnet.

Das Knie markiert den Punkt, bei dem die galaktische Komponente der Protonen abnimmt, da galaktische Quellen nicht zu einer Beschleunigung zu höheren Energien im Stande sind[24].

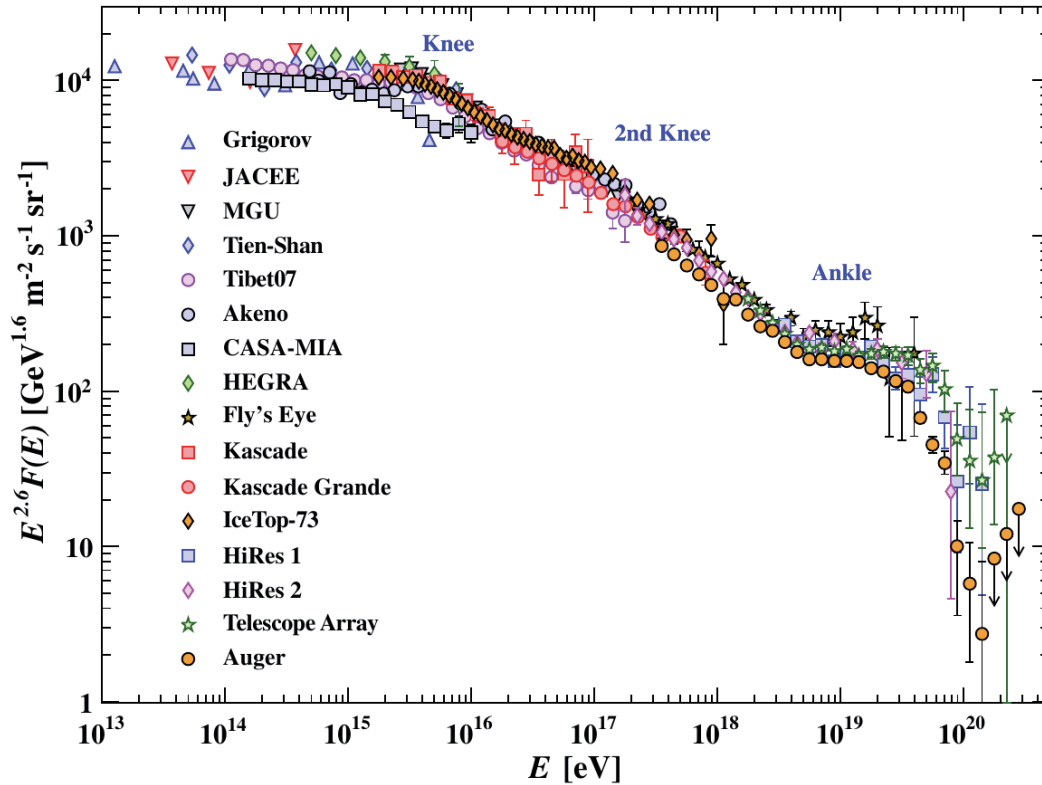


Abbildung 2.1: Differenzieller Fluss der kosmischen Strahlung in Abhängigkeit der Energie gemessen durch verschiedene Experimente[25]. Dieses Spektrum ist zur besseren Darstellung mit $E^{2,6}$ multipliziert. Dadurch sind das Knie (≈ 3 PeV), das zweite Knie (≈ 80 PeV) und die Ferse (≈ 50 EeV) des Spektrums der kosmischen Strahlung leichter zu identifizieren.

Eine weitere Änderung des spektralen Index findet bei einer Energie von $E \approx 80$ PeV statt und wird als zweites Knie bezeichnet. Beim zweiten Knie wird angenommen, dass die Beschleunigung der Ionen von der Ladung Z des betrachteten Ions abhängt. Je mehr Ladung ein Ion trägt, desto stärker kann es von einer Quelle beschleunigt werden. Somit werden alle Elemente in einer Quelle bis zu verschiedenen Energiemaxima beschleunigt. Die schwersten relevanten Ionen sind Eisenkerne. Der Bereich zwischen dem ersten Knie und dem zweiten Knie beschreibt die Beschleunigungsgrenzen der einzelnen Elemente, bis an dem Punkt des zweiten Knies die Beschleunigungsgrenze der Eisenkerne erreicht ist.

Der letzte markante Punkt ist bei $E \approx 50 \text{ EeV}$ zu finden und wird Ferse genannt. Bei der Ferse wird angenommen, dass die dort produzierten hochenergetischen Ionen nicht mehr aus unserer Galaxie stammen können, sondern nur von extragalaktischen Quellen erzeugt werden. Mögliche Quellen dieser hochenergetischen Strahlung wurden in Abschnitt 2.1 aufgeführt. Bei der Detektion dieser Ionen ist zu beachten, dass diese nicht beliebig hohe Energien besitzen können. Protonen über einer Energie von $E \approx 50 \text{ EeV}$ können mit dem Mikrowellenhintergrund über

$$p + \gamma \rightarrow \Delta^+ \rightarrow \begin{cases} p + \pi^0 \\ n + \pi^+ \end{cases} \quad (2.2.1)$$

wechselwirken. Diese Wechselwirkung wird als GZK-Unterdrückung[26, 27] bezeichnet und Hinweise auf diese Wechselwirkung lassen sich für hohe Energien in dem gezeigten Spektrum erkennen. Die Unterdrückung durch die GZK-Wechselwirkung ist abhängig von dem Verhältnis E/A , wobei A die Massenzahl des betrachteten Ions angibt.

Aufgrund der vielen massen- bzw. ladungsabhängigen Effekte im Spektrum der geladenen kosmischen Strahlung ist es sinnvoll dieses für verschiedene Elemente zu betrachten. Bei der Entwicklung von Flussmodellen der geladenen kosmischen Strahlung ist es von Vorteil nicht alle Elemente einzeln zu beschreiben, sondern diese in Gruppen einzuteilen. Solch eine Gruppeneinteilung findet sich in [28] und ist im Folgenden aufgeführt:

- H ($Z = 1$)
- He ($Z = 2$)
- CNO-Gruppe ($Z = 3 - 9$)
- Mg-Si-Gruppe ($Z = 10 - 24$)
- Mn-Fe-Gruppe ($Z = 25 - 27$)

Elemente mit einer Ladungszahl von $Z > 27$ treten zwar auf, doch ist ihr Beitrag zum totalen Fluss der kosmischen Strahlung gering[29]. In der aktuellen Forschung gibt es eine Vielzahl an Modellen für den Fluss der kosmischen Strahlung. Einige der am häufigsten verwendeten Modelle sind das Polygonato-, das TIG- und das Gaisser-Modell, auf welche im Nachfolgenden eingegangen wird[24, 30, 31]. In dieser Arbeit wird zusätzlich noch das GST-Modell untersucht, welches eine auf aktuellsten Messungen basierende modifizierte Version des Gaisser-Modells darstellt[29].

Das Polygonato-Modell wurde von Jörg R. Hörandel entwickelt und ist ein phänomenologisches Modell, welches anhand von Daten verschiedener Experimente konstruiert wird. Der Name Polygonato gibt bereits Aufschluss darüber, welcher Ansatz hinter

diesem Modell steht. Der Begriff Polygonato stammt aus dem Griechischen und bedeutet „viele Knie“. Das Polygonato-Modell erzeugt für die Massengruppen eigene Positionen des Knies der kosmischen Strahlung. Diese Abschwächung bei der Position des Knies der einzelnen Elementgruppen ist proportional zu ihrer Ladung, beginnend mit Protonen bei $E \approx 3 \text{ PeV}$ bis hin zu Eisen bei Energien von $E \approx 80 \text{ PeV}$. Die Abschwächung nach dem jeweiligen Knie wird als exponentielle Unterdrückung parametrisiert. Das Polygonato-Modell beschreibt den Fluss bis zum zweiten Knie hinreichend genau, allerdings ist die Flussvorhersage für Energien oberhalb des zweiten Knies zu gering.

Eine weitere Parametrisierung des Flusses der primären Strahlung wurde von Thunman, Ingelman und Gondolo berechnet und wird in der Literatur als TIG-Modell bezeichnet[30]. Der Fluss für geladene Nukleonen wird dabei gemäß

$$\Phi_N(E) \left[\frac{\text{Nukleonen}}{\text{cm}^2 \text{ s sr GeV}/A} \right] = \begin{cases} 1,7 E^{-2,7} & \text{für } E < 5 \cdot 10^6 \text{ GeV} \\ 174 E^{-3} & \text{für } E > 5 \cdot 10^6 \text{ GeV} \end{cases} \quad (2.2.2)$$

parametrisiert und berücksichtigt das Knie bei einer Energie von $E = 5 \cdot 10^6 \text{ GeV}$. Die simple Form des Flusses liegt darin begründet, dass sie zur Berechnung von Myon- und Myonneutrinoströmen verwendet werden kann. Dies wurde durchgeführt, um die prompte Komponente der atmosphärischen Strahlung zu beschreiben. Diese Komponente wird nochmals in Abschnitt 2.4 erläutert.

In dieser Arbeit wird das Primärmodell nach Gaisser[31] verwendet. Dieses Primärmodell betrachtet drei Populationen der kosmischen Strahlung separat. Zusätzlich werden die Beiträge der einzelnen Elementgruppen für jeden Energiebereich eigens betrachtet. Die drei Populationen sind die Energiebereiche vor dem Knie (Population 1), zwischen Knie und Ferse (Population 2) und nach der Ferse (Population 3). Für jeden dieser Energiebereiche wird eine exponentielle Unterdrückung eingefügt, welche mit einer Schockbeschleunigung als Ursprung dieser Populationen verträglich ist. Die Flüsse der Elementgruppen lassen sich mit

$$\Phi_i(E) = \sum_{j=1}^3 a_{i,j} E^{-\gamma_{i,j}} \times \exp\left(-\frac{E}{Z_i R_{c,j}}\right) \quad (2.2.3)$$

berechnen, wobei i die Elementgruppe und j die Population im Energiebereich bezeichnet. Die charakteristischen Rigiditäten sind durch $R_{c,j}$ und die Ladungen mit Z_i für die jeweiligen der Elementgruppen beschrieben. Das massenaufgelöste Spektrum ist in diesem Modell durch eine direkte Messung mit einem Ballon-Experiment für die erste Population durchgeführt worden[32]. Dieses direkt gemessene Massenspektrum wird für die weiteren Populationen extrapoliert.

Für die dritte Population, dessen Teilchen auf die Produktion innerhalb extragalaktischer Quellen zurückzuführen sind, werden zwei mögliche Modelle angegeben. Falls diese Population nur aus Protonen besteht, wird dieses Modell als H4a oder auch H3p bezeichnet. Ist diese letzte Population allerdings von Eisen dominiert, so wird dieses Modell H3a genannt. Das H3a-Modell liefert bessere Übereinstimmungen mit gemessenen Spektren und wird aus diesem Grund in dieser Arbeit verwendet. Allerdings liefern Primärteilchen mit Energien oberhalb von $E > 50 \text{ EeV}$ kaum einen Beitrag bei der Analyse von Myonen mit Energien bis zu $E = 1 \text{ PeV}$.

2.3 Sekundäre kosmische Strahlung

Bei der Propagation von hochenergetischen geladenen Teilchen oder Photonen in der Atmosphäre werden sekundäre Teilchen erzeugt. Diese sekundären Teilchen können hochenergetische Leptonen, Hadronen oder Mesonen sein, welche wiederum andere Teilchen erzeugen. Die sekundäre kosmische Strahlung kann in eine hadronische, myonische und elektromagnetische Komponente unterteilt werden.

Einige der am häufigsten produzierten Teilchen sind Pionen und Kaonen, wobei im Folgenden nur auf Pionen eingegangen wird, da für Kaonen Analoges gilt. Pionen können geladen als π^\pm oder ungeladen als π^0 innerhalb der Atmosphäre produziert werden. Aufgrund der Lebensdauer geladener Pionen von $\tau \approx 260 \mu\text{s}$ propagieren diese durch die Atmosphäre und verlieren einen Teil ihrer Energie bevor sie zerfallen[25]. Die dominanten Zerfallsprozesse für das negativ geladene und das neutrale Pion ergeben sich durch die folgenden Zerfallsgleichungen:

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (2.3.1)$$

$$\pi^0 \rightarrow \gamma\gamma \quad (2.3.2)$$

Analog zu Gleichung (2.3.1) entsteht für ein positiv geladenes Pion ein positiv geladenes Myon und ein Myonneutrino. In hochenergetischen Proton-Proton-Kollisionen können kurzlebige Teilchen entstehen, die als prompte Komponente bezeichnet werden (vgl. Abschnitt 2.4). Der Zerfall der Pionen in Myonen stellt die myonische Komponente dar, während der Zerfall des ungeladenen Pions die elektromagnetische Komponente darstellt. Die Photonen der elektromagnetischen Komponente produzieren auf dem Weg zur Erdoberfläche Elektron-Positron-Paare, welche bis zur Erde propagieren und Cherenkov-Strahlung erzeugen.

In der hadronischen Komponente der sekundären kosmischen Strahlung werden Protonen, Neutronen, Pionen und andere Hadronen produziert. Diese können durch weitere Wechselwirkungen wieder Baryonen oder Mesonen produzieren. Die Mesonen

können wie zuvor beschrieben über Zerfallsprozesse in die anderen Komponenten der sekundären kosmischen Strahlung übergehen.

Geladene Teilchen, welche sich mit einer Geschwindigkeit oberhalb der Lichtgeschwindigkeit im Medium bewegen, erzeugen niederenergetische Photonen, auch Cherenkov-Strahlung genannt. Diese Photonen können durch optische Elemente, welche im UV-Bereich sensitiv sind, detektiert werden. Der Winkel der abgestrahlten Photonen bei der Erzeugung ist primär abhängig von dem Brechungsindex des Mediums, in dem sich das Teilchen bewegt. Für Luft beträgt der Winkel $\sim 3^\circ$, wohingegen er bei Eis $\sim 40^\circ$ beträgt.

2.4 Prompte Myonen

Bei der Kollision von hochenergetischen Protonen mit Teilchen der Atmosphäre können kurzlebige Sekundärteilchen produziert werden. Diese Teilchen werden als prompte Komponente der kosmischen Strahlung bezeichnet, da diese sehr schnell zerfallen und somit kaum Energie bei der Propagation durch die Atmosphäre verlieren. Die Zerfallsdauer eines D-Mesons ist mit $\tau \approx 1$ ps sehr viel kleiner als die des geladenen Pions mit $\tau \approx 260 \mu\text{s}$. An dieser Stelle sei erwähnt, dass die prompte Komponente in der Atmosphäre nicht mit der charmhaltigen Komponente gleichzusetzen ist. Anhand von Abbildung 2.2 ist zu erkennen, dass für höhere Energien der Anteil der flavorfreen Elternteilchen einen nicht zu vernachlässigenden Beitrag zum Fluss der atmosphärischen Myonen leistet.

Die charmhaltigen Hadronen, welche zum atmosphärischen Myonenfluss beitragen können, sind die D-Mesonen und das Λ_c -Baryon. Diese charmhaltigen Hadronen zerfallen direkt aufgrund ihrer kurzen Lebensdauer und produzieren dabei Neutrinos und Myonen. Die dominanten Zerfallskanäle für die Myonproduktion werden durch

$$D^0 = c\bar{u} \quad \rightarrow \quad K^- + \mu^+ \nu_\mu \quad (2.4.1)$$

$$D^+ = c\bar{d} \quad \rightarrow \quad \bar{K}^0 + \mu^+ \nu_\mu \quad (2.4.2)$$

$$D_s^+ = c\bar{s} \quad \rightarrow \quad \mu^+ \nu_\mu \quad (2.4.3)$$

$$\Lambda_c = udc \quad \rightarrow \quad \Lambda + \mu^+ \nu_\mu \quad (2.4.4)$$

beschrieben. Ein Nachweis der prompten Komponente der atmosphärischen Myonen ist von besonderem Interesse, da die Produktion dieser Teilchen für kleine Winkel nicht von Teilchenbeschleunigern gemessen werden kann. Durch diese Messung könnte der $c\bar{c}$ -Wirkungsquerschnitt bei einer Proton-Proton-Kollision durch eine Analyse von atmosphärischen Myonen berechnet werden[33]. Da die charmhaltige Komponente auch Neutrinos enthält, kann sie ebenso über den Neutrinofluss gemessen werden.

Dies ist jedoch problematisch, da der Fluss prompter Myonneutrinos im gleichen Energiebereich liegt wie der Fluss hochenergetischer astrophysikalischer Neutrinos und von diesen dominiert wird[9].

Die Komponente der prompten Myonen besteht allerdings auch aus flavorfreien Elternteilchen. Auch diese Mesonen haben eine kurze Lebensdauer, jedoch ist der myonische Zerfallskanal meist subdominant. Diese niedrige Zerfallswahrscheinlichkeit wird allerdings durch die absolute Häufigkeit dieser Teilchen in Luftschauern ausgeglichen. Die beitragenden Mesonen und deren relevante Zerfallskanäle sind nachfolgend aufgeführt:

$$\pi^0 = \frac{1}{\sqrt{2}} (u\bar{u} - d\bar{d}) \quad \rightarrow \quad \mu^- + \mu^+ + \gamma \quad (2.4.5)$$

$$\eta \approx \frac{1}{\sqrt{6}} (u\bar{u} + d\bar{d} - 2s\bar{s}) \quad \rightarrow \quad \mu^- + \mu^+ + \gamma \quad (2.4.6)$$

$$\eta' \approx \frac{1}{\sqrt{3}} (u\bar{u} + d\bar{d} + s\bar{s}) \quad \rightarrow \quad \mu^- + \mu^+ + \gamma \quad (2.4.7)$$

$$\rho^0 = \frac{1}{\sqrt{2}} (u\bar{u} - d\bar{d}) \quad \rightarrow \quad \mu^- + \mu^+ \quad (2.4.8)$$

$$\omega = \frac{1}{\sqrt{2}} (u\bar{u} + d\bar{d}) \quad \rightarrow \quad \mu^- + \mu^+ + \pi^0 \quad (2.4.9)$$

$$\phi = s\bar{s} \quad \rightarrow \quad \mu^- + \mu^+ \quad (2.4.10)$$

Es ist zu erkennen, dass beim Zerfall dieser Mesonen ein $\mu^- \mu^+$ -Paar auftritt. Ein Nachweis zweier Myonen in IceCube ist nur möglich, wenn die Myonen einen hohen Transversalimpuls ($p_T \geq 10$ GeV) besitzen[34]. Aufgrund dieses hohen Transversalimpulses ist dann ein Abstand der beiden Spuren von ca. 150 m gegeben, jedoch ist die Häufigkeit dieser Ereignisse gering.

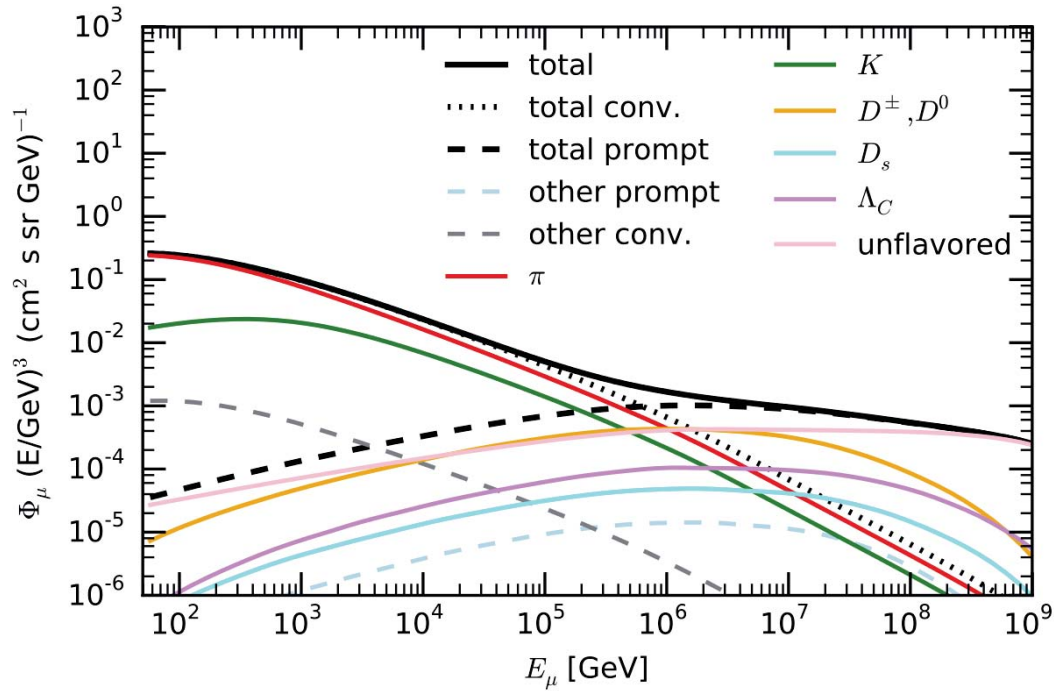


Abbildung 2.2: Atmosphärischer Myonfluss für das Spektrum der kosmischen Strahlung berechnet mit SIBYLL-2.3 RC1[33] und dem TIG-Modell. Dargestellt sind die Beiträge von verschiedenen Elternteilchen zum Myonfluss. Zusätzlich sind die Beiträge für die prompte und konventionelle Komponente nochmals separat aufgeführt.

2.5 Wechselwirkungen von Myonen

Für die Rekonstruktion des hochenergetischen Myonflusses sind die Wechselwirkungen, die ein Myon während der Propagation durch ein Medium durchführen kann, von großer Bedeutung. Dieser Abschnitt gibt eine Übersicht über die möglichen Wechselwirkungen eines Myons. Eine detailliertere Betrachtung der Energieverluste geladener Leptonen und der verwendeten Software zur Simulation derselben findet sich in [35].

Bei der Betrachtung der Wechselwirkungen wird in diesem Abschnitt der Zerfall des Myons als Wechselwirkung betrachtet, da auch dieser ein hochenergetisches Sekundärteilchen erzeugt. Die möglichen Wechselwirkungen lassen sich wie folgt zusammenfassen:

- Ionisation
- Elektron-Positron-Paarproduktion
- Bremsstrahlung
- Photonukleare Wechselwirkung
- Zerfall

Diese Wechselwirkungen unterscheiden sich in ihrer Häufigkeit des Auftretens, der Größe des möglichen Energieverlustes und der dabei produzierten Sekundärteilchen. Für diese Arbeit sind die Wechselwirkungen von Myonen von großer Bedeutung, aufgrund dessen wird im Nachfolgenden detailliert auf die einzelnen Wechselwirkungen eingegangen.

Die Ionisation von Materie bei der Propagation eines hochenergetischen Myons ist die am häufigsten stattfindende Wechselwirkung und in Gleichung (2.5.1) dargestellt.



Durch die Ionisation der Materie werden Elektronen aus der Materie gelöst und erzeugen Cherenkov-Strahlung, da sie eine relativistische Geschwindigkeit besitzen. Zwar ist diese Wechselwirkung die am häufigsten Auftretende, doch ist ihr Beitrag zum Gesamtenergieverlust ab einer Myonenergie von ≈ 1 TeV subdominant.

Den größten Anteil am Gesamtenergieverlust der Myonen oberhalb dieser Energie trägt die Produktion von Elektron-Positron-Paaren bei, welche sich mit Gleichung (2.5.2) beschreiben lässt.



Die Elektron-Positron-Paarproduktion dominiert ab einer Energie von ≈ 1 TeV den Gesamtverlust, da sie hochenergetische Elektronen und Positronen erzeugen kann. Im Folgenden wird nur auf die Wechselwirkung der Elektronen eingegangen, allerdings gilt das Verhalten analog für die Positronen. Die Elektronen bewegen sich durch die Materie und wechselwirken wiederum mit der Materie und können hochenergetische Bremsstrahlungsphotonen erzeugen. Bremsstrahlung tritt auf, wenn geladene Teilchen in die Nähe eines Atomkerns gelangen und mit dem elektromagnetischen Feld dieses Atoms wechselwirken. Diese Wechselwirkung ist der dominante Energieverlust für Elektronen. Die Energie der erzeugten Elektronen kann fast vollständig der

Energie des erzeugenden Myons entsprechen, allerdings ist es wahrscheinlicher, dass sie weitaus geringer ist. Diese Energieabhängigkeit ist in Abbildung 2.3 dargestellt.

Der nächstgrößere Beitrag von Energieverlusten eines Myons wird durch Bremsstrahlungsverluste, beschrieben in Gleichung (2.5.3), erzeugt.

$$\mu A \rightarrow \mu A \gamma \quad (2.5.3)$$

Der Energieverlust der Bremsstrahlung ist gleichverteilt im Logarithmus der Energieverluste und der Energie des erzeugenden Myons. Somit können Bremsstrahlungsinteraktionen extrem hohe Energieverluste erzeugen, welche allerdings weitaus seltener sind als die Ionisation oder die Elektron-Positron-Paarproduktion. Bei hohen Energien werden geringe Bremsstrahlungsverluste durch den Landau-Pomeranchuk-Migdal(LPM)- und Ter-Mikaelian(TM)-Effekt unterdrückt. Beiden Effekten liegt zugrunde, dass das geladene Lepton das Photon nicht punktförmig abstrahlt, sondern der Abstrahlungsprozess auf einer bestimmten Strecke geschieht. Während dieser Strecke sind die Wellenfunktionen des geladenen Leptons und des Photons kohärent. Eine Störung dieser Kohärenz beeinflusst den Wirkungsquerschnitt der Bremsstrahlung und der Paarproduktion für hohe Energien. Für Myonen trägt dieser Effekt erst bei extremen Energien von $E > 10^{22}$ eV signifikant zum Gesamtenergieverlust bei, wohingegen für Elektronen bereits ab einer Energie von $E \approx 100$ TeV ein signifikanter Beitrag erkennbar ist. Der LPM-Effekt beschreibt die Veränderung des Wirkungsquerschnitts, wenn das geladene Lepton während des Abstrahlungsprozesses eine Compton-Streuung mit dem Atomkern durchführt. Der TM-Effekt beschreibt die Wechselwirkung des Photons mit den Hüllenelektronen während des Abstrahlprozesses.

Weiterhin kann das Myon noch inelastisch mit einem Atom wechselwirken, welches als photonukleare Wechselwirkung bezeichnet wird und in Gleichung (2.5.4) dargestellt ist.

$$\mu A \rightarrow \mu A' X \quad (2.5.4)$$

Der Atomkern wird bei diesem Prozess zerstört und eine hadronische Kaskade wird erzeugt. Dieser Energieverlust hat den geringsten Beitrag zum Gesamtenergieverlust des Myons, kann aber ähnlich hohe Einzelverluste erzeugen wie die Bremsstrahlung. Da dieser Prozess nicht mit der perturbativen QCD berechnet werden kann, werden hier phänomenologische Modelle wie das Regge-Modell verwendet. Parametrisierungen dieser Wechselwirkungen sind in [36–42] zu finden.

Der dominante Zerfallsprozess des Myons ist in Gleichung (2.5.5) dargestellt.

$$\mu^- \rightarrow \nu_\mu e^- \bar{\nu}_e \quad (2.5.5)$$

Der Zerfall des Myons ist allerdings als Energieverlust für IceCube zu vernachlässigen, da er erst bei Energien unterhalb der Detektionsschwelle auftritt. Der Zerfall des geladenen Leptons ist allerdings im Falle des Taus der dominante Energieverlust und muss aus diesem Grund dort beachtet werden.

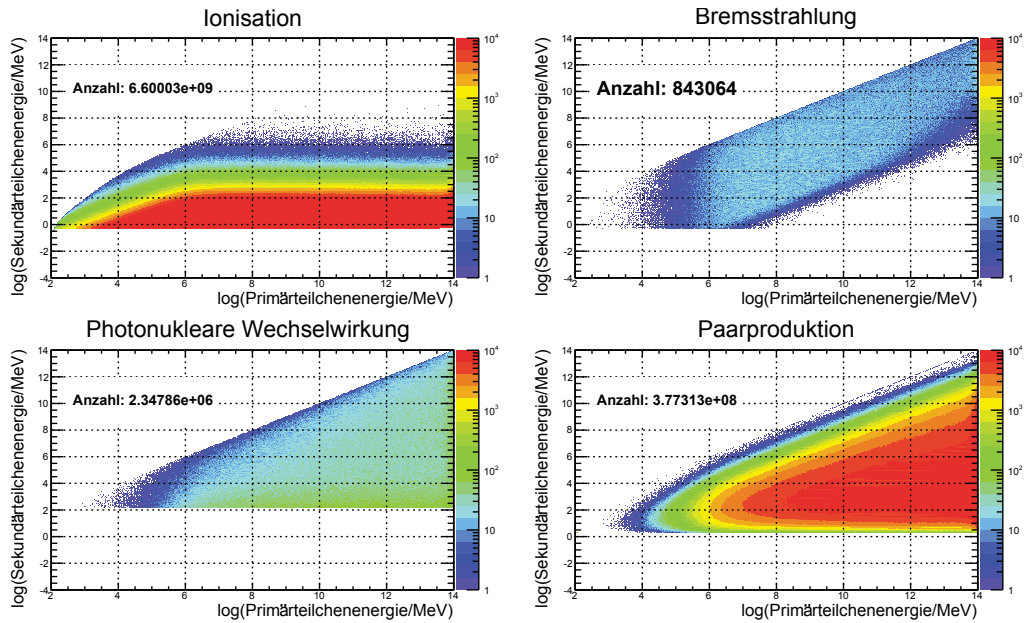


Abbildung 2.3: Stochastische Energieverluste von Myonen für verschiedene Wechselwirkungen abhängig von der Energie des Myons. Die Farbe gibt die Anzahl der aufgetretenen Wechselwirkungen an. Zur Erstellung dieser Grafiken wurden 10^4 Myonen mit einer Startenergie von 10^{20} eV durch Südpoleis propagiert[35].

3 IceCube

Der IceCube-Detektor[12] befindet sich am geografischen Südpol und besitzt ein instrumentiertes Volumen von einem Kubikkilometer. IceCube dient zum Nachweis hochenergetischer Neutrinos und kann Neutrinos von 10 GeV bis zu einigen PeV detektieren. Der Bau des Detektors begann im Dezember 2004 und endete im Dezember 2010. Das gesamte Experiment besteht aus einem InIce-, einem DeepCore- und einem IceTop-Array, auf welche im Nachfolgenden detaillierter eingegangen wird. Weiterhin befindet sich am Südpol das IceCube-Lab, welches zur Datennahme und der ersten Analyse der aufgenommenen Daten genutzt wird.

3.1 Allgemein

In der zur Zeit vorliegenden und finalen Form besteht der InIce-Detektor aus 86 Sensorreihen, auch Strings genannt, an welchen jeweils 60 digitale optische Module, kurz DOMs, angebracht sind. Die DOMs befinden sich in einer Tiefe von 1450 m bis 2450 m und haben an einem String einen Abstand von 17 m zueinander. Die Strings sind hexagonal in einem Abstand von 125 m zueinander angeordnet. Die zuvor beschriebenen DOMs werden aufgrund ihrer Position als InIce-Detektor bezeichnet. Mit dem InIce-Detektor können eine Vielzahl von Teilchen detektiert werden und darüber hinaus kann er als Veto für hochenergetische Photonen genutzt werden. Der Detektionsbereich des InIce-Detektors für Neutrinos liegt zwischen 100 GeV bis einigen PeV.

An der Oberfläche der jeweiligen Strings befinden sich Stationen, welche aus jeweils zwei DOMs in einem Wassertank bestehen. Dieses Array an Oberflächenstationen wird auch IceTop genannt. Das Einfrieren des Tankwassers geschah in einem kontrollierten Gefrierprozess, damit das Wasser so wenige Lufteinschlüsse wie möglich entwickelt[43]. Das IceTop-Array dient als Instrument zur Analyse der geladenen kosmischen Strahlung und kann zusammen mit dem InIce-Detektor als Veto für Neutrinos verwendet werden.

Im Zentrum des IceCube-Detektors befindet sich der DeepCore-Detektor, welcher sowohl einen geringeren Abstand der Strings als auch der DOMs zueinander aufweist.

Die DOMs sind in einen Abstand von 7 – 12 m angebracht und die Strings haben einen Abstand von 60 m zueinander. Die Verringerung der Abstände der DOMs und Strings dient dazu die untere Energieschwelle des IceCube-Detektors auf 10 GeV zu senken. Der Vorteil von DeepCore ist die Tatsache, dass DeepCore sich im Zentrum des InIce-Detektors befindet. Dadurch kann der InIce-Detektor als Veto für das DeepCore-Array verwendet werden, um niederenergetische Neutrinos nachzuweisen. DeepCore kann Neutrinos mit Energien von 10 GeV bis einige 100 GeV detektieren.

Die Konstruktion des gesamten Detektors konnte nur im Sommer erfolgen und somit wurde der Detektor sukzessiv fertiggestellt. Von Beginn des Baus bis zu der Fertigstellung des Detektors fand bereits eine Datennahme mit den partiellen Konfigurationen statt. Die Bezeichnungen für die Jahre während der Konstruktion werden an der Anzahl der aktiven InIce-Strings festgemacht, wodurch Analysen mit der IC1-, IC22-, IC40-, IC59-, IC79- und IC86-Konfiguration durchgeführt wurden. Da die Konstruktion nun abgeschlossen ist, lassen sich Analysen mit Daten über mehrere Jahre durchführen, da ein homogener und symmetrischer Detektor vorliegt.

3.2 Detektionsprinzip

Das Grundprinzip bei der Detektion von hochenergetischen Teilchen in IceCube ist der Cherenkov-Effekt[45]. Der Cherenkov-Effekt tritt auf, wenn geladene Teilchen sich mit einer Geschwindigkeit v durch ein Medium bewegen, welche über der Lichtgeschwindigkeit in dem Medium c_m liegt. Das abgestrahlte Licht breitet sich innerhalb eines Kegels aus, welcher einen Öffnungswinkel θ besitzt, der von dem Brechungsindex n in dem Medium und der Energie des Teilchens abhängt. Dieser Winkel berechnet sich über

$$\cos(\theta) = \frac{1}{n\beta},$$

wobei in der Variablen $\beta = \frac{v}{c_0}$ die Energie des Teilchens eingeht und c_0 die Vakuumlichtgeschwindigkeit ist. Für hochenergetische Teilchen, wie sie bei IceCube detektiert werden, kann $\beta = 1$ angenommen werden. Dadurch ergibt sich ein charakteristischer Winkel für ein bestimmtes Medium. Für hochenergetische Teilchen in IceCube ergibt sich mit $n = 1,31$ dieser Winkel zu $\theta = 40,2^\circ$.

Die Anzahl an Photonen, welche durch ein hochenergetisches Teilchen erzeugt werden, sind allerdings nur ein Bruchteil der Photonen, welche durch die DOMs detektiert werden. Hauptsächlich wechselwirken die Teilchen mit dem Eis und erzeugen so weitere Teilchen, welche wiederum eine höhere Geschwindigkeit haben

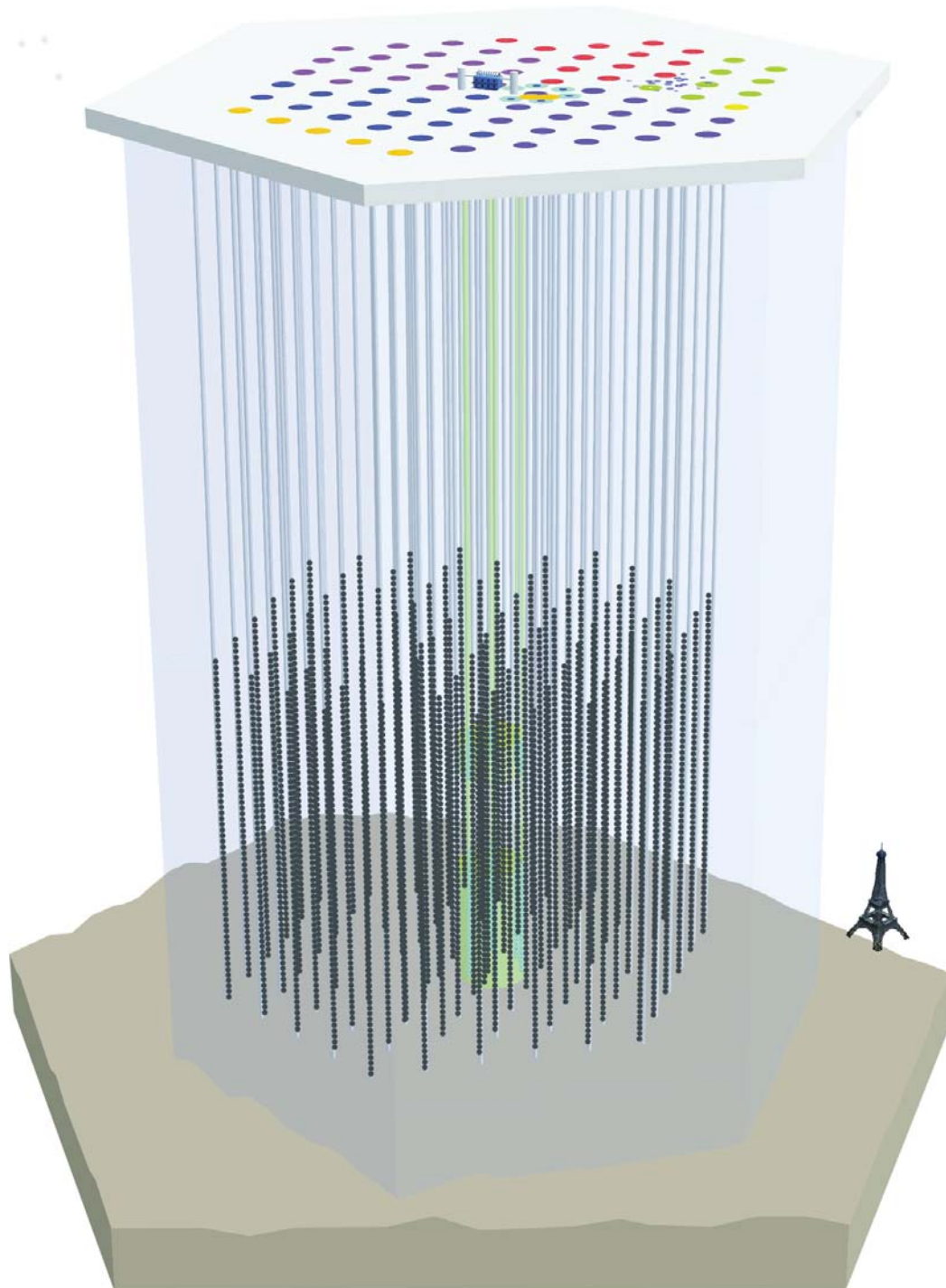


Abbildung 3.1: Schematische Darstellung des IceCube-Detektors mit dem DeepCore- und dem IceTop-Array nach [44]. Als Größenvergleich ist zusätzlich der Eiffelturm dargestellt.

als die Lichtgeschwindigkeit im Medium (vgl. Abschnitt 2.5). Dadurch produzieren diese Teilchen weitere Cherenkov-Strahlung, welche aufgrund der hohen Anzahl an Wechselwirkungen den Großteil der detektierten Photonen bildet.

Der Nachteil des IceCube-Detektors oder vielmehr seiner Position besteht darin, dass ein natürliches Medium als Detektionsmedium verwendet wird. Einerseits ist die mittlere Absorptionslänge des Südpoleises mit $a \approx 100$ m hoch, andererseits ist die mittlere Streulänge der Photon mit $b \approx 20$ m gering, wenn die DOM- und String-Abstände als Referenz betrachtet werden. Aus diesem Grund werden meistens die DOMs von bereits gestreute Photonen erreicht, sodass eine Richtungsrekonstruktion erschwert wird. Mit Hilfe von sogenannten Flasher-Runs müssen Messungen der Eiseigenschaften durchgeführt werden. Aus diesen Messungen lassen sich die zuvor erwähnten Größen bestimmen.

Eine weitere Schwierigkeit stellt die Inhomogenität des Eises dar, da sich die Eisschicht über Millionen von Jahren gebildet hat. Dadurch sind die zuvor erwähnten Streu- und Absorptionslängen nur für bestimmte Regionen im Eis anwendbar. Bei dem für IceCube relevanten Bereich gibt es mehrere Schichten, welche eine geringere Absorptionslänge aufweisen. Die prominenteste dieser Schichten befindet sich in einer Tiefe von ungefähr 1900 m – 2100 m und ist eine Staubschicht, welche die Parameter stark beeinflusst. Die Abhängigkeiten der charakteristischen Größen für diese Schichten und für den prominenten Bereich, der auch „dustlayer“ genannt wird, sind in der Abbildung 3.2 dargestellt.

Ein weiterer Effekt, der für die Lichtausbreitung berücksichtigt werden muss, ist die Bewegung des Eises. Das Südpoleis bewegt sich gleichbleibend, allerdings tiefenabhängig, in eine Richtung. Obwohl der Detektor eine nahezu rotationssymmetrische Struktur besitzt, kommt es durch die Eisbewegung zu einer richtungsabhängigen Veränderung der Absorptions- und Streulängen.

Der letzte entscheidende Effekt wird nicht durch das Südpoleis hervorgerufen, sondern von dem Eis, welches sich in den Bohrlöchern befindet. Dieses Eis unterscheidet sich von dem Südpoleis, da es geschmolzen und wieder gefroren wurde. Durch dieses Schmelzen und Einfrieren befinden sich in diesem Eis winzige Luftporen. Diese Luftporen nehmen Einfluss auf die Propagation des Lichts an den jeweiligen DOMs, sodass die Photonen in der Nähe der DOMs stärker gestreut werden.

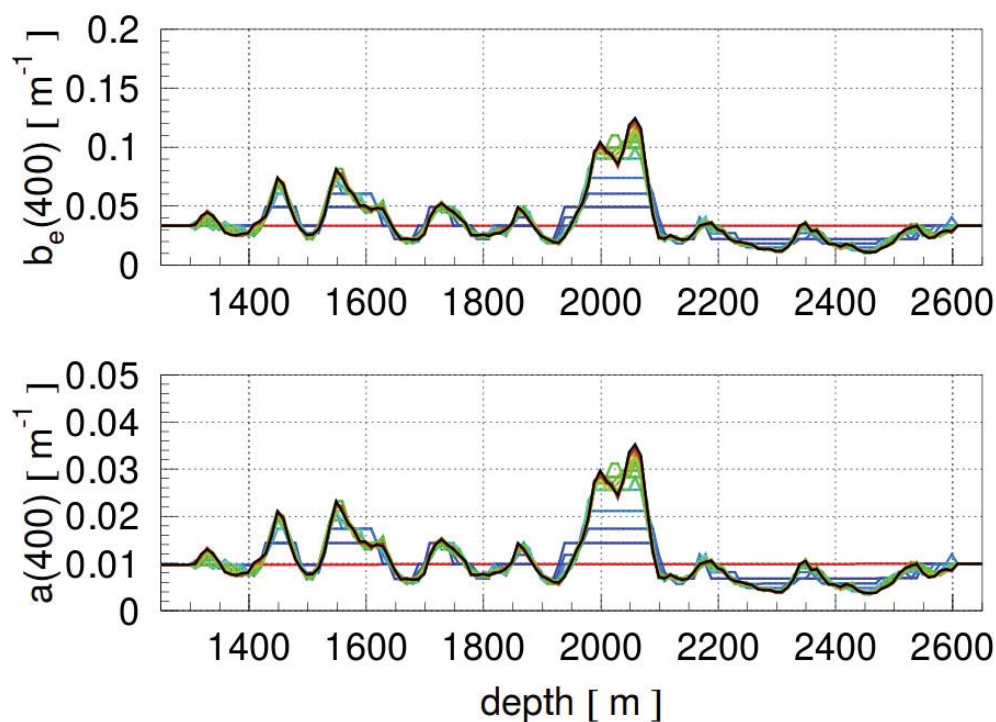


Abbildung 3.2: Absorptionslänge a und die Streulänge b des am Südpol befindlichen Eises in Abhängigkeit der Tiefe[46].

3.3 Ziele

Das primäre Ziel des IceCube-Detektors besteht in dem Nachweis von hochenergetischen astrophysikalischen Neutrinos. Diese Neutrinos können in astrophysikalischen Quellen wie aktiven galaktischen Kernen oder Supernovaüberresten produziert werden (vgl. Abschnitt 2.1). Diese Kategorie der Neutrinos wurde im Jahr 2014 mit der IC86-Detektorkonfiguration nachgewiesen[9]. Aufgrund dieses Ergebnisses ist das weitere Ziel die Quelle bzw. Quell-Klasse dieser hochenergetischen astrophysikalischen Neutrinos zu finden. Damit ließen sich Modelle für die Beschleunigung von hochenergetischen Teilchen in diesen Quellen besser verstehen und andere ausschließen. Zusätzlich können Neutrinos aus Supernovae detektiert werden. Diese Detektion geschieht nicht direkt, sondern äußert sich über eine gestiegene Rauschrate aller DOMs. Dadurch kann ein Alarm für eine Supernova an weitere Experimente gesendet werden.

Des Weiteren kann IceCube dazu verwendet werden exotische Teilchen zu detektieren. Diese exotischen Teilchen sind dunkle Materie und magnetische Monopole. Um dunkle Materie nachzuweisen ist es nötig Gravitationszentren zu untersuchen, in denen Neutrinos über die Wechselwirkung

$$\chi\chi \rightarrow \nu\bar{\nu} \quad (3.3.1)$$

entstehen können[47]. Als Quellen zur Detektion dieser Annihilation bieten sich massive oder nahe Objekte an. Als nahe Objekte zur Detektion von dunkler Materie können die Erde und die Sonne genutzt werden, welche im Zentrum eine Ansammlung von dunkler Materie enthalten könnten[48]. Bei der Analyse von massiven Objekten bieten sich nahe Galaxien und Galaxie-Cluster[49] oder der galaktische Halo unserer Milchstraße an[50]. Magnetische Monopole werden von IceCube durch die Cherenkov-Photonen von δ -Elektronen[51] für kleine Werte von $\beta = v/c$ oder über direkte Photonen bei Geschwindigkeiten höher als $\beta \approx 0,75$ detektiert[52]. Diese Signal ist sehr deutlich, da bereits bei den δ -Elektronen mehr als 1000 Elektronen pro cm entstehen[51] bzw. der magnetische Monopol ≈ 8300 -mal mehr Cherenkov-Photonen produziert als ein Myon mit identischer Energie[52].

Eine weiteres Ziel von IceCube ist die Analyse der geladenen kosmischen Strahlung. Mit Hilfe von IceTop kann sowohl die Komposition als auch das Spektrum der kosmischen Strahlung bestimmt werden[53]. Auch der Fluss von ungeladener kosmischer Strahlung in Form von hochenergetischen Neutronen kann mit IceTop untersucht werden[54]. Bei der Untersuchung der Neutronen kann lediglich eine obere Schranke für den Fluss aufgestellt werden. Weiterhin kann das Spektrum der atmosphärischen Myonen bestimmt werden. Dies kann, wie in dieser Arbeit, dazu genutzt werden den Fluss der konventionellen und prompten Komponente zu bestimmen.

4 Signalklassifikation und -rekonstruktion

Relevante Informationen sind bei der Analyse von Daten meist durch unwichtige Informationen verunreinigt. Diese Verunreinigung wird typischerweise als Untergrund bezeichnet, während die relevanten Daten als Signal bezeichnet werden. Das Verhältnis von Signal- zu Untergrundeignissen kann stark variieren und unbalanciert sein (für IceCube bis zu $1 : 10^6$ [55]). Eine hohe Verunreinigung kann statistische Aussagen erschweren oder unmöglich machen, daher ist es notwendig den Untergrund soweit wie möglich zu reduzieren. Zur Auswahl der Signalereignisse sind bestimmte Schritte notwendig, damit stabile und aussagekräftige Ergebnisse erreicht werden. Aus selektierten Signalereignissen kann auf eine Signalverteilung zurückgeschlossen werden. Auf die Schritte zur Signalklassifikation und -rekonstruktion wird im Folgenden eingegangen, wobei die allgemeinen Methoden vorgestellt werden, jedoch mit dem Ziel diese auf IceCube-Daten anzuwenden.

4.1 Signalklassifikation

Die Klassifikation von Daten hat in den vergangenen Jahren immer mehr an Bedeutung gewonnen. Algorithmen zur Klassifikation werden in vielen Teilbereichen unseres Alltages verwendet, um beispielsweise Gefahrstoffe zu entdecken oder Werbung zu personalisieren. Auch in der Physik findet dieses Gebiet immer mehr Beachtung, da die Menge an Daten stetig steigt und die Komplexität der Problemstellungen zunimmt. Im Allgemeinen lassen sich Schritte für eine Klassifikation in eine Attributauswahl, eine Signalklassifikation und eine Validierung unterteilen, welche in diesem Abschnitt erläutert werden.

4.1.1 Attributauswahl

Bei der Betrachtung von zu analysierenden Daten werden diese oft durch eine Vielzahl an Attributen beschrieben. Bei dem Beispiel einer Signalklassifikation stellt sich die Frage, welche dieser Attribute eine hohe Korrelation zu der Signal- bzw. Untergrundklasse aufweisen, anhand derer zuverlässig Signal und Untergrund

unterschieden werden können. Diese Attribute sind von Bedeutung, da sie maschinellen Lernalgorithmen zur Verfügung stehen und auf dieser Basis ein Modell zur Signalklassifikation erstellt wird. Beispiele für Lernalgorithmen sind der Entscheidungsbaum und der Random Forest, welche in den nachfolgenden Abschnitten erläutert werden. Um eine möglichst optimale Auswahl dieser Attribute zu treffen, ist es sinnvoll, unabhängig und mit fortschrittlichen Algorithmen eine solche Auswahl an Attributen vorzunehmen. Algorithmen, die dies als Ziel haben, sind beispielsweise die Forward Selection[56][57], die Backward Elimination[56][58] und der mRMR¹-Algorithmus[59][60]. Der mRMR-Algorithmus wird in diesem Abschnitt im Detail erläutert, da er in der vorliegenden Analyse verwendet wird.

Das Ziel der Anwendung des mRMR-Algorithmus ist eine Auswahl an Attributen zu finden, welche die höchste Relevanz bezüglich der Signalklasse aufweist, gleichzeitig aber die geringste Redundanz der Attribute untereinander besitzt. Zur Berechnung dieses Zusammenhangs wird die Transinformation I nach Gleichung (4.1.1) berechnet.

$$I(x, y) = \iint p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy \quad (4.1.1)$$

Nehmen die Attribute x und y diskrete Werte an, so kann die Transinformation mit Hilfe einer einfachen Summation berechnet werden. Ein Problem, welches bei diesem Verfahren auftritt, ist, dass die Wahrscheinlichkeitsdichten $p(x)$, $p(y)$ und $p(x, y)$ der einzelnen Attribute nicht analytisch bekannt sind, sondern nur Schätzer für diese Dichten. Für diskrete Attribute kann die Wahrscheinlichkeitsdichte durch die empirische Wahrscheinlichkeitsdichte geschätzt werden. Für kontinuierliche Attribute können die Wahrscheinlichkeitsdichten mit Hilfe einer Kerndichtenschätzung bestimmt werden[61][62].

Bei einer Klassifikation ist es relevant, ob ein Attribut x eine Abhängigkeit zu der Signalklasse c aufweist. Unter diesem Gesichtspunkt lassen sich mit Gleichung (4.1.1) die Relevanz D und Redundanz R mit den Gleichungen (4.1.2) und (4.1.3) berechnen.

$$D(S, c) = \frac{1}{|S|} \sum_{x \in S} I(x, c) \quad (4.1.2)$$

$$R(S) = \frac{1}{|S|^2} \sum_{x, y \in S} I(x, y) \quad (4.1.3)$$

¹minimum Redundancy Maximum Relevance

Hierbei gibt S in diesem Abschnitt die Menge von potentiellen Attributen an, welche die Daten am besten beschreiben sollen. Der durch eine Attributauswahl zu minimierende Ausdruck Φ ergibt sich anschließend zu

$$\Phi(D, R) = D - R. \quad (4.1.4)$$

Unter Verwendung der Relevanz und Redundanz können alternative Ausdrücke erstellt werden, welche minimiert werden können. Als Beispiel ist hier der Quotient aus der Relevanz und Redundanz $\Phi^*(D, R) = D/R$ genannt. Für diese Analyse wird hingegen die Berechnung des zu optimierenden Ausdruckes nach Gleichung (4.1.4) durchgeführt.

Dieses Vorgehen kann bereits als Attributauswahl verwendet werden, allerdings ist dieses Verfahren für eine große Anzahl von Attributen nicht praktikabel. Bei einer gewünschten Anzahl S an Attributen wird Gleichung (4.1.4) $\frac{A!}{(A-S)!}$ -mal berechnet, wenn A die Gesamtanzahl der Attribute angibt. Diese Anzahl der Berechnungen beträgt bereits bei der Auswahl von $S = 5$ und einer Gesamtzahl $A = 20$ an Attributen rund zwei Millionen, da alle Kombinationen von 5 Attributen betrachtet werden müssen. Aus diesem Grund wird $\Phi(D, R)$ nicht für alle Kombinationen berechnet, sondern die Menge an Attributen in S iterativ erweitert. Somit wird zunächst das Attribut hinzugefügt, welches die höchste Relevanz zu der Zielklasse aufweist. Im zweiten Iterationsschritt wird das Attribut hinzugefügt, welches die höchste Relevanz zu der Zielklasse aufweist, allerdings die geringste Redundanz zu dem bereits gewählten Attribut besitzt. Dieses Vorgehen wird nach Gleichung (4.1.5)

$$\max_{y \notin S_{m-1}} \left(I(y, c) - \frac{1}{m-1} \sum_{x \in S_{m-1}} I(x, y) \right) \quad (4.1.5)$$

durchgeführt, bis in der Menge S die Anzahl an gewünschten Attributen erreicht ist. Diese Art der iterativen Berechnung reduziert die Anzahl der Auswertungen auf $\sum_{i=1}^S (A - i)$, wodurch bei dem vorherigen Beispiel lediglich 90 Auswertungen benötigt würden. Das beschriebene Vorgehen weicht von der in dieser Arbeit benutzten Implementierung für kontinuierliche Attribute geringfügig ab. Eine detaillierte Beschreibung des verwendeten Verfahrens findet sich in [63].

Da die Daten, welche zur Auswahl der Attribute verwendet werden, statistische Fluktuationen aufweisen, muss die Stabilität der finalen Auswahl an Attributen untersucht werden. Zu diesem Zweck kann eine Attributauswahl k -mal auf verschiedenen Teilmengen der Daten ausgeführt und anschließend ein Maß für die Stabilität berechnet werden. Zwei übliche Maße sind die Stabilitäten nach Jac-

card[64] und Kuncheva[65]. Die Stabilität nach Jaccard ist lediglich auf zwei Mengen mit Gleichung

$$J(S_a, S_b) = \frac{|S_a \cap S_b|}{|S_a \cup S_b|} \quad (4.1.6)$$

definiert, kann aber mit Gleichung (4.1.7) auf mehr als zwei Mengen erweitert werden[63]. Die Stabilität nach Jaccard J und nach Kuncheva K lassen sich für k Anwendungen einer Attributauswahl auf verschiedene Teilmengen mit den Gleichungen (4.1.7) und (4.1.8) berechnen.

$$J(S_1 \dots S_k) = \frac{2}{k(k-1)} \sum_{i=1}^k \sum_{j=i+1}^k J(S_i, S_j) \quad (4.1.7)$$

$$K(S_1 \dots S_k) = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^k I_C(S_i, S_j) \quad (4.1.8)$$

Die Funktion I_C in Gleichung (4.1.8) ist ein Konsistenzmaß definiert auf dem Bereich $[0, 1]$ und ist detailliert in [65] beschrieben. Die so erhaltenen Maße für die Stabilität liegen in dem Bereich $[0, 1]$. Ein Wert von 0 bedeutet, dass in allen k Attributmengen keine Übereinstimmung von Attributen vorhanden ist, wohingegen ein Wert von 1 aussagt, dass in allen k Attributmengen dieselben Attribute ausgewählt werden.

4.1.2 Entscheidungsbaum

Zur Klassifikation von Ereignissen bei Zwei- oder Mehrklassenproblemen können verschiedene Algorithmen verwendet werden. Viele der heutzutage verwendeten Algorithmen sind sogenannte baumbasierte Algorithmen, welche auf dem Algorithmus des Entscheidungsbaumes basieren und auch als CART² bezeichnet werden[66]. In diesem Abschnitt wird die Funktionsweise eines Entscheidungsbaumes erläutert, da er für das Verständnis des verwendeten Random-Forest-Algorithmus unerlässlich ist (vgl. Abschnitt 4.1.3).

Ein Entscheidungsbaum beginnt an seiner Wurzel mit allen zur Verfügung stehenden Attributen A und der jeweiligen Klassenzugehörigkeit für jedes Ereignis. Das Ziel bei der Verwendung des Entscheidungsbaumes ist, eine Trennung der jeweiligen Klassen. Zu diesem Zweck ist der Algorithmus in folgende Schritte unterteilt:

1. Betrachtung eines Attributs a_i der Menge A .

²Classification and Regression Trees

2. Auswahl von Schnitten in dem Wertebereich des Attributs a_i .
3. Berechnung von Maßen, welche die Trennungsstärken nach der Anwendung der gewählten Schnitte beschreiben.
4. Auswahl des Schnitts mit der maximalen Trennstärke für die verschiedenen Klassen. Die maximale Trennung wird anhand des gewählten Maßes in Schritt 3 bestimmt.

Dieses Verfahren wird für alle Attribute a_i der Attributmenge A durchgeführt. Abschließend wird das Attribut zum Trennen verwendet, welches die maximale Trennstärke besitzt. Nachdem diese Trennung durchgeführt wurde, entstehen ausgehend vom Ausgangsknoten zwei neue Äste, welche die getrennten Ereignisse enthalten.

Das Kernproblem dieses Algorithmus besteht darin eine maximale Trennung zwischen den verschiedenen Klassen zu erreichen und die Trennungsstärke zu quantifizieren. Zur Lösung dieses Problems gibt es verschiedene Ansätze, welche in dieser Arbeit nicht vollständig erläutert werden können. Die bekanntesten Algorithmen zur Quantifizierung einer Trennung sind der GINI-Index[66], die Marshall-Korrektur[67] und der Informationsgewinn³[68]. Welcher Algorithmus sich für die Trennung am besten eignet, hängt von dem zu untersuchenden Datensatz ab, allerdings stellt sich bei einer Untersuchung der Quantifizierungsalgorithmen heraus, dass alle im allgemeinen Fall eine genügend hochwertige Trennung gewährleisten[69].

Der Informationsgewinn I wird hier im Detail besprochen, da dieser von dem verwendeten Random Forest durchgeführt wird. Die Basis des Algorithmus ist die Minimierung der Entropie H der einzelnen Klassen in den jeweiligen Knoten. Zu diesem Zweck wird die Entropie vor der Trennung und die erwartete Entropie EH nach der Trennung für alle Äste k berechnet. Für ein Zwei-Klassen-Problem (p und n) lassen sich die dafür benötigten Gleichungen wie folgt darstellen:

$$H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n} \log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n} \log_2\left(\frac{n}{p+n}\right) \quad (4.1.9)$$

$$EH(a) = \sum_{i=1}^k \frac{p_i + n_i}{p+n} H\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right) \quad (4.1.10)$$

$$\Rightarrow I(a) = H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - EH(a) \quad (4.1.11)$$

Zunächst wird die Entropie H mit Hilfe von Gleichung (4.1.9) für den aktuellen Knoten berechnet, an dem eine Trennung durchgeführt werden soll. Anschließend wird gemäß Gleichung (4.1.10) für ein Attribut a und einen gewählten Schnitt die

³engl.: Information Gain

erwartete Entropie EH berechnet. Diese Berechnung bildet die Summe über die Entropien an jedem entstehenden Knoten k . Damit finale Knoten mit wenigen Ereignissen kein starkes Gewicht erhalten, werden die einzelnen Entropien für jeden Knoten mit der Anzahl der darin auftretenden Ereignisse gewichtet. Abschließend wird mit Gleichung (4.1.11) der Informationsgewinn durch diesen Schnitt berechnet. Aus diesen berechneten Informationsgewinnen $I_j(a)$ für alle Schnitte j eines Attributes wird anschließend derjenige gewählt, welcher den maximalen Informationsgewinn aufweist.

Die Anforderungen an die Attribute, welche der Entscheidungsbaum verarbeiten kann, sind sehr gering und nachfolgend aufgeführt:

- Attribute müssen definiert sein und müssen somit immer einen Eintrag besitzen.
- Die Definitionsbereiche der Attribute müssen begrenzt sein.
- Attribute müssen einen eindeutigen Typ besitzen. Somit dürfen sich ihre Bedeutungen nicht für verschiedene Ereignisse ändern.

Aufgrund dessen sind für die Verwendung eines Entscheidungsbaumes nominale, ordinale und kardinale Attribute erlaubt. Für nominale Attribute sind Schnitte zur Trennung nicht klar definiert, da die auftretenden Werte nicht geordnet sind. Es gibt zwei Wege diese Attribute zu behandeln. Diese beiden Verfahren werden zur Veranschaulichung am Beispiel des nominalen Attributes *Farbe* mit den Ausprägungen *Rot*, *Blau* und *Grün* erläutert. Zunächst kann der Entscheidungsbaum alle Ausprägungen dieses Attributes als Trennung wählen. Nach der Trennung würden somit drei Knoten entstehen, welche jeweils nur Ereignisse mit *Rot*, *Blau* oder *Grün* enthalten. Dies ist eine sehr komplizierte Art einen Entscheidungsbaum zu erstellen, da für jeden Knoten mehrere Folgeknoten erlaubt sein müssen. Weiterhin wäre die Einteilung der Ereignisse in die richtigen Knoten zeitintensiv, wenn viele Ausprägungen des nominalen Attributes existieren würden.

Eine weitere Möglichkeit zur Verwendung nominaler Attribute ist die Nutzung der Eigenschaft, dass sich nominale Attribute zumindest vergleichen lassen. Mit Hilfe dieser Eigenschaft kann an einem Knoten abgefragt werden, ob ein bestimmtes Ereignis die Ausprägung *Rot* besitzt. Die Trennung der Ereignisse findet anschließend in die Knoten *Rot* und *Nicht-Rot* statt. Dies ist eine einfache und schnelle Einteilung, welche für viele nominale Attribute performant ist. Ein weiterer Vorteil dieser Verarbeitung von nominalen Attributen ist, dass die Struktur zum Aufbauen eines Baumes dieselbe ist, wie für nicht-nominale Attribute. Dadurch gibt es an jedem Knoten nur zwei Verzweigungen zu Folgeknoten. Ein so aufgebauter Baum wird auch binärer Entscheidungsbaum genannt.

Der zuvor erwähnte Separationsprozess wird auch für die entstehenden Äste durchgeführt bis ein Abbruchkriterium erreicht wird. Übliche Abbruchkriterien lassen sich wie folgt zusammenfassen:

- Der zu trennende Knoten enthält nur Ereignisse einer Klasse und ist somit ein reiner Knoten.
- Eine im Vorhinein definierte maximale Baumtiefe T ist erreicht.
- Weniger als X Ereignisse in dem Knoten gehören einer bestimmten Klasse an.
- Bei einer optimalen Trennung ist die Anzahl der Ereignisse in den Ästen geringer als ein Minimum Y .
- Die optimale Trennung besitzt einen kleineren Informationsgewinn als Z .

Welches Abbruchkriterium gewählt wird, hängt von dem gestellten Problem ab. Meist werden mehrere Abbruchkriterien für das Erstellen eines Entscheidungsbaumes gewählt. Ist der Entscheidungsbaum vollständig erstellt, so muss jedem finalen Knoten des Entscheidungsbaumes, welche auch Blätter genannt werden, jeweils eine Klasse zugeteilt werden. Die zugeteilte Klasse wird aus der größten Anzahl der darin befindlichen Klassenereignisse bestimmt. Beim Anwenden des Entscheidungsbaumes auf Daten erhält dann jedes Ereignis, welches sich abschließend in diesem Blatt befindet, dessen Klassenzugehörigkeit. In der hier vorliegenden Arbeit wird die Trennung so lange durchgeführt bis alle entstehenden Blätter nur Ereignisse einer Klasse enthalten, sodass nur reine Knoten entstehen. Eine schematische Darstellung eines so aufgebauten Entscheidungsbaumes findet sich in Abbildung 4.1.

Der Entscheidungsbaum ist einer der simpelsten Algorithmen zur Klassifikation von Daten. Bei wohldefinierten Klassen werden stabile Ergebnisse erreicht, welche die Ansprüche an eine Separation bereits erfüllen können. Sind die Klassen allerdings sehr ähnlich und erst in höheren Dimensionen unterscheidbar, so ist die Wahl eines Entscheidungsbaumes zur Klassifikation nicht robust.

Eines der großen Probleme bei der Nutzung von Entscheidungsbäumen, wie bei vielen wenig komplexen maschinellen Lernalgorithmen, besteht in dem Risiko einer zu starken Anpassung an den Trainingsdatensatz, welcher zum Erstellen des Baumes verwendet wurde. Somit findet eine Überanpassung des Entscheidungsbaumes an den Datensatz und nicht an die jeweiligen Klassen statt. Bei der Anwendung auf einen anderen Datensatz werden bei einer Überanpassung stark unterschiedliche Ergebnisse erzielt. Ein Verfahren zur Überprüfung von Überanpassung wird in Abschnitt 4.1.5 beschrieben.

Ein weiteres Problem besteht bei der resultierenden Klassenzugehörigkeit. Ein voll ausgebauter Entscheidungsbaum trifft am Ende des Prozesses eine Entscheidung, ob

ein Ereignis ein Signal- oder Untergrundeignis ist. Aus dieser Information kann allerdings nicht entnommen werden, wie sicher sich der Entscheidungsbaum ist, dass dieses Ereignis tatsächlich ein Signalereignis darstellt. Bei sehr ähnlichen Untergrund- und Signalereignissen kann es von dem Trainingsdatensatz abhängen, wie das jeweilige Ereignis klassifiziert wird. Die Information, dass bestimmte Ereignisstypen unklassifizierbar oder ununterscheidbar voneinander sind, geht hierbei verloren und solche Ereignisse werden genau so klassifiziert wie eindeutige Signal- oder Untergrundeignisse. Einen fortgeschrittenen Ansatz bietet der Random-Forest-Algorithmus, welcher eine Konfidenz berechnet wie signalähnlich einzelne Ereignisse sind (vgl. Abschnitt 4.1.3).

Zur Vollständigkeit dieses Algorithmus sei noch erwähnt, dass der Entscheidungsbaum auch in der Lage ist eine kontinuierliche Klasse zu schätzen und somit eine Regression durchzuführen. In dieser Anwendung kann die Erstellung des Baumes nicht mit der Berechnung des Informationsgewinns erfolgen. Im kontinuierlichen Fall werden bei jedem Schnitt die Varianzen der gesuchten Variable in den entstehenden Ästen berechnet und mit der Varianzen des ursprünglichen Knotens verglichen. Durch dieses Verfahren wird die Varianz der abschließenden Blätter minimiert und der Wert der kontinuierlichen Variable wird durch eine Mittelung aller darin befindlichen Trainingsereignisse berechnet.

4.1.3 Random Forest

Aufgrund der zuvor erwähnten Probleme des Entscheidungsbaumes bietet es sich an weiterentwickelte maschinelle Lernalgorithmen zu verwenden. Die Gruppe der Ensemble-Methoden ist eine einfache Erweiterung von simplen Lernalgorithmen. Bei diesem Ansatz werden mehrere wenig komplexe Lernalgorithmen verwendet, um eine stabilere Klassifikation zu gewährleisten. Solch ein simpler Lernalgorithmus kann beispielsweise der Entscheidungsbaum sein, welcher im vorherigen Abschnitt beschrieben wurde. Zwei gängige Ensemble-Methoden sind der Boosted Decision Tree[70] und der Random Forest[71].

Der Boosted Decision Tree führt mehrere Klassifikationen durch und hält jeweils einen Anteil α des Trainingsdatensatzes zum Testen zurück. Nach dem Aufbau des ersten Entscheidungsbaumes wird der zurückgehaltene Anteil klassifiziert. Falsch klassifizierte Ereignisse aus diesem Anteil erhalten bei der Erstellung des nächsten Entscheidungsbaumes ein höheres Gewicht. Somit wird die Falschklassifizierungsrate iterativ minimiert. Zusätzlich erhält jeder trainierte Baum ein Gewicht, das in die Berechnung der Konfidenz eingeht.

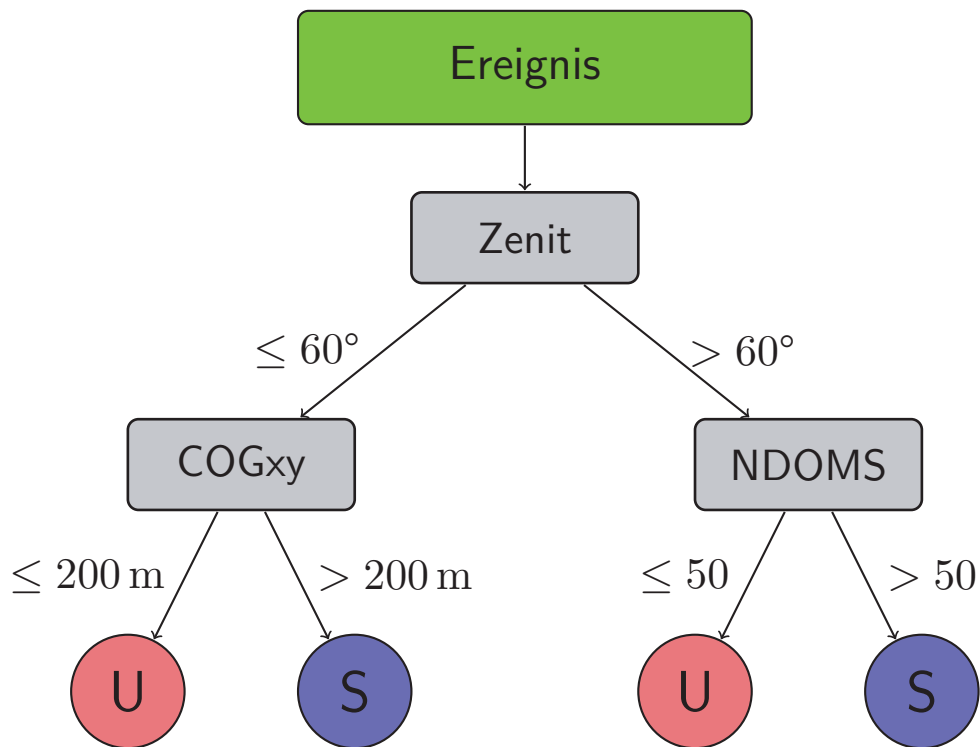


Abbildung 4.1: Schematische Darstellung eines Entscheidungsbaumes mit einer Tiefe von zwei zur Trennung der Untergrundklasse U und der Signalklasse S . Die gewählten Attribute sind in diesem Beispiel der Zenitwinkel $Zenit$, die Anzahl der DOMs $NDOMS$ und die radiale Position des Schwerpunktes COG_{xy} des Ereignisses.

In dieser Arbeit wird der Algorithmus des Random Forest verwendet, welcher M unabhängige Entscheidungsbäume erstellt. Dadurch entsteht ein Entscheidungswald, welcher mehrere Vorteile gegenüber dem einfachen Entscheidungsbaum besitzt. Der Algorithmus des Entscheidungsbaumes ist völlig deterministisch, dadurch muss der Algorithmus angepasst werden, damit er M unabhängige Entscheidungsbäume erzeugen und unterschiedliche Ergebnisse liefern kann. Dies wird beim Random Forest dadurch erreicht, dass zufällig ein Anteil α des Datensatzes zum Testen zurückgehalten wird und pro Knoten jedes Entscheidungsbaumes nicht jedes Attribut zur Verfügung steht. Die Einstellungsmöglichkeiten für den in dieser Arbeit verwendeten Random Forest lassen sich wie folgt zusammenfassen:

- Anzahl der Entscheidungsbäume (M)
- Anzahl der Attribute zum Trainieren (A)
- Anzahl der Attribute, welche an jedem Knoten zur Verfügung stehen (a)
- Maximale Tiefe der Bäume

Der Anteil an zurückgehaltenen Daten kann für die genutzte Implementierung des Random Forest nicht definiert werden, sondern wird vom Algorithmus mit $\alpha = 10\%$ ausgeführt[72]. Durch diese zufällige Variation des Trainingsdatensatzes und der Attribute, die an jedem Knoten jeweils zur Verfügung stehen, entsteht ein robuster Lernalgorithmus. Eine Überanpassung ist durch die zufälligen Variationen selten, da weder der Datensatz noch die zur Verfügung stehenden Attribute für jeden Entscheidungsbaum identisch sind.

Wenn die Einstellungen für den Random Forest gewählt sind, kann die Erstellung der M Bäume stattfinden. Einer der Vorteile des Random Forest ist, dass dieses Training parallel geschehen kann. Dieser Vorteil ist besonders in der heutigen Zeit relevant, da Mehrkern- und Grafikprozessoren immer mehr an Bedeutung gewinnen[73]. Dadurch können die Erstellung des Random-Forest-Modells beschleunigt und mehr Entscheidungsbäume falls nötig zum Trainieren gewählt werden. Die Anwendung eines Random-Forest-Modells auf Daten erzeugt einen klassifizierten Datensatz, welcher eine Klassifikation in Signal oder Untergrund sowie einen Konfidenzwert enthält. Dieser kontinuierliche Konfidenzwert beschreibt wie sicher die Klassifikation eines bestimmten Ereignisses ist. Bei einem Zwei-Klassen-Problem gibt er an, wie viele Entscheidungsbäume dieses Ereignis als Signal klassifiziert haben, dividiert durch die Anzahl aller Entscheidungsbäume (siehe Gleichung (4.1.12)). Für verschiedene Konfidenzen lassen sich zusätzlich Qualitätsparameter berechnen, welche in Abschnitt 4.1.4 genauer betrachtet werden.

Um festzustellen wie wichtig jedes Attribut bei der Klassifikation ist, kann untersucht werden, wie stark sich die Falschklassifizierungsrate ändert, falls eines

der Attribute aus einem Entscheidungsbaum entfernt wird. Dies geschieht über die zurückgehaltenen Daten für jeden Entscheidungsbaum. Diese werden durch jeden Entscheidungsbaum klassifiziert und es wird eine Falschklassifizierungsrate berechnet. Eine Entfernung eines Attributes würde eine zufällige Entscheidung an jedem Knoten bedeuten, an welchem das Attribut verwendet wird. Dies ist schwierig zu realisieren, da zu diesem Zweck der Algorithmus eines jeden Baumes geändert werden müsste. In der Praxis werden die Werte des zu untersuchenden Attributes im zurückgehaltenen Datensatz zufällig vertauscht und damit jegliche Korrelation zur Signal- und Untergrundklasse aufgelöst. Im Anschluss wird das entstandene Modell auf den modifizierten zurückgehaltene Datensatz angewandt, um eine weitere Falschklassifizierungsrate zu berechnen. Der relative Anstieg dieser Rate kann als Wichtigkeit für das jeweilige Attribut angesehen werden. War dieses Attribut wichtig zur Klassifikation, so steigt die Falschklassifizierungsrate stark ($\approx 10\%$) an, wohingegen sie bei einem unwichtigen Attribut nur leicht ansteigt ($\approx 1\%$). Dies kann dazu genutzt werden, um unwichtige Attribute zu entfernen. Weiterhin kann die Falschklassifizierungsrate zur Konsistenzüberprüfung einer vorherigen Attributauswahl, wie in Abschnitt 4.1.1 beschrieben, verwendet werden.

Zur Vollständigkeit sei noch zu erwähnen, dass mit Hilfe des Random-Forest-Algorithmus auch eine Regression einer zu schätzenden Variable durchgeführt werden kann. Um eine kontinuierliche Variable zu schätzen, wird die Regression an jedem Knoten wie bei einem Entscheidungsbaum (vgl. Abschnitt 4.1.2) durchgeführt. Die zufällige Wahl der Attribute an jedem Knoten und der zurückgehaltene Datensatz werden beibehalten. Für jeden Entscheidungsbaum liegt am Ende einer Regression ein Wert vor, der aus den gemittelten Ereignissen in dem finalen Knoten besteht. Um den abschließenden Wert des Ereignisses für den Random Forest zu erhalten, wird über alle Entscheidungsbäume M gemittelt. Dieser Algorithmus wird als Random-Forest-Regressor bezeichnet.

4.1.4 Qualitätsparameter

Nachdem das trainierte Random-Forest-Modell auf einen Testdatensatz angewandt wurde, lassen sich verschiedene Qualitätsparameter für den klassifizierten Datensatz berechnen. Diese Qualitätsparameter sind im Nachfolgenden beschrieben.

$$K = \frac{M_+}{M} \quad (4.1.12)$$

Die Gleichung (4.1.12) beschreibt die Konfidenz des Algorithmus für ein Ereignis, wobei M_+ die Anzahl der Entscheidungsbäume ist, welche dieses Ereignis als Signal klassifiziert haben, und M die Gesamtanzahl der erstellten Entscheidungsbäume. Die

Konfidenz ist somit auf die Anzahl der verwendeten Bäume normiert und nimmt nur Werte im Bereich von 0 bis 1 an. Eine Klassifikation des Random Forest findet bei einem Konfidenzwert von $K^* = 0.5$ statt, sodass alle Ereignisse mit einer höheren Konfidenz zu der einen Klasse und die Ereignisse darunter zu der anderen Klasse zugeordnet werden. Oftmals ist diese Wahl der Grenze nicht sinnvoll, da noch immer viel Untergrund im Datensatz vorhanden ist oder zu viel Signal verworfen wird. Eine Festlegung der Klasse kann ebenso für einen selbst festgelegten Wert der Konfidenz geschehen, welcher meist mittels der Betrachtung der Reinheit R und der Effizienz E gewählt wird. Diese beiden Qualitätsparameter hängen vom Schnitt K^* in der Konfidenz K ab, welche als Klassifikationsgrenze genutzt wird. In vielen Anwendungen der Astroteilchenphysik wird meist ein Konfidenzschnitt von $K^* = 1$ oder $K^* = 0.98$ verwendet, falls ein hoch reiner Datensatz benötigt wird [55, 74]. Dieser Wert kann gesenkt werden, wenn nicht genügend Ereignisse nach dem Schnitt vorhanden sind und eine höhere Verunreinigung keine gravierenden Auswirkungen auf die weitere Analyse hat.

$$R(K^*) = \frac{S(K^*)}{S(K^*) + U(K^*)} \quad (4.1.13)$$

Die Reinheit R beschreibt das Verhältnis der richtig klassifizierten Ereignisse einer Klasse zu der gesamten Anzahl aller Ereignisse. In Gleichung (4.1.13) werden die beiden Klassen Signal S und Untergrund U als Beispiel herangezogen, um die Reinheit der Signalklasse bei einem bestimmten Konfidenzschnitt K^* zu ermitteln. Die auf Simulationsdaten bestimmte Reinheit sagt aus, wie viel Untergrund zu erwarten ist, falls bei einem bestimmten Konfidenzschnitt K^* eine Klassifikation vorgenommen wird. Die Optimierung für einen möglichst reinen Datensatz ist meist nicht wünschenswert, da bei diesem Vorgehen viele Signalereignisse verworfen werden. Aus diesem Grund wird der Konfidenzwert K^* meist unter Berücksichtigung der Effizienz E gewählt.

$$E(K^*) = \frac{S(K^*)}{S} \quad (4.1.14)$$

Die Effizienz E nach Gleichung (4.1.14) beschreibt das Verhältnis der verbleibenden Ereignisse einer Klasse nach einem Konfidenzschnitt K^* und der Gesamtanzahl der Ereignisse dieser Klasse. Dies ist für die Signalklasse dargestellt, da die Effizienz bezüglich dieser Klasse relevant ist. Sie ist im Wertebereich von 0 bis 1 definiert, wobei beim Wert 0 kein Signalereignis selektiert wird und bei einem Wert von 1 alle Signalereignisse selektiert werden. Bei einer Klassifikation sollte die Effizienz möglichst hoch sein, um für weitere Verarbeitungsschritte möglichst viele Signalereignisse zur Verfügung zu haben. Die Effizienz E kann auch für die Untergrundklasse bestimmt werden, allerdings ist dieser Wert in der Praxis oft nicht relevant.

4.1.5 Kreuzvalidierung

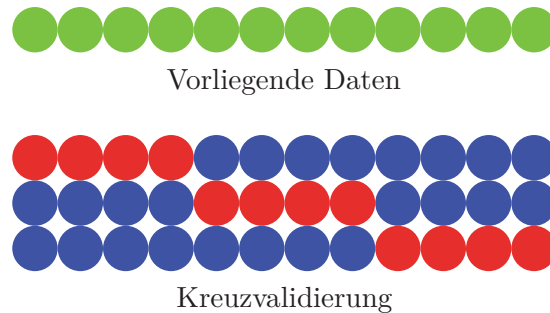


Abbildung 4.2: Schematische Darstellung der Durchführung einer dreifachen Kreuzvalidierung zur Bestimmung der Varianz von Klassifikationsalgorithmen nach [75]. Die Modelle werden mithilfe der blau markierten Datenpunkte trainiert und auf die rot markierten Ereignisse angewandt. Aus klassifizierten rot markierten Ereignissen werden anhand dessen Qualitätsparameter und deren Varianzen bestimmt.

Bei einer Untersuchung der Qualitätsparameter ist es sinnvoll das Verhalten der Random-Forest-Klassifikation auf statistische Fluktuationen der Trainingsdaten zu untersuchen. Theoretisch kann es beim Random Forest, wie bei jedem Lernalgorithmus, zu einer Überanpassung an den Trainingsdatensatz kommen. Durch eine Überanpassung würde das Random-Forest-Modell den Anschein einer besseren Klassifikation erwecken, als praktisch möglich ist. Dies kann dadurch geschehen, dass der Lernalgorithmus statistische Fluktuationen einzelner Ereignisse als Klasseneigenschaft erkennt und darauf optimiert. Da dies ein unerwünschter Effekt ist, muss überprüft werden, ob solch eine Überanpassung vorliegt. Eine einfache Methode hierzu bietet beispielsweise die Kreuzvalidierung.

Bei der Methode der Kreuzvalidierung wird der Trainingsdatensatz in n Teile aufgeteilt. Das Random-Forest-Modell wird anschließend auf $(n - 1)$ Teilen des Trainingsdatensatzes erstellt. Auf dem verbleibenden Datensatz wird anschließend das trainierte Random-Forest-Modell angewandt. Diese Prozedur wird wiederholt bis alle Teilmengen ein Mal auf ein trainiertes Random-Forest-Modell angewendet wurden. Eine dreifache Kreuzvalidierung ist grafisch in Abbildung 4.2 dargestellt. Die Anzahl der Kreuzvalidierungsschritte ist problemabhängig, jedoch haben sich Werte von $n = 5$ und $n = 10$ in der Praxis als sinnvoll erwiesen[71]. Für jede Klassifikation werden die Qualitätsparameter anhand der Gleichungen (4.1.12) bis (4.1.14) bestimmt. Aus den einzelnen Werten der Kreuzvalidierungsschritte lässt sich ein empirischer Mittelwert und eine empirische Varianz berechnen.

Liegt eine Überanpassung des Random-Forest-Modells vor, so werden die Varianzen der Qualitätsparameter so groß, dass sie nicht die statistische Fluktuation der Daten beschreiben, sondern die Überanpassung des Random-Forest-Modells. Liegt keine Überanpassung vor, so nehmen die Varianzen der Qualitätsparameter Werte an, die den erwarteten statistischen Fluktuationen der zu testenden Daten entsprechen. Ein großer Vorteil dieser Methode ist nicht nur, dass sie eine Überprüfung der Überanpassung liefert, sondern auch eine Fehlerabschätzung für die bestimmten Qualitätsparameter. Eine typische Anwendung ist eine Abschätzung des verbleibenden Untergrundes der gemessenen Ereignisse bei einem bestimmten Konfidenzschnitt K^* .

4.2 Signalrekonstruktion

In der Physik sind nur wenige Größen direkt messbar, da beispielsweise bei einer Waage die Stauchung einer Feder gemessen und daraus ein Gewicht ermittelt werden kann. Beim speziellen Fall von IceCube wird eine elektrische Ladung gemessen, welche einer Lichtmenge entspricht, die wiederum proportional zu einem Energieverlust eines Teilchens im Detektor ist. Findet ein Energieverlust eines Teilchens statt, so soll von der gemessenen Detektorantwort der wahre Energieverlust oder die Energie des primären Teilchens rekonstruiert werden. Weiterhin ist bei vielen Fragestellungen in der Physik nicht die Information über ein einzelnes Ereignis von Interesse, sondern eine zugrundeliegende Verteilung vieler gemessener Ereignisse. In der Astroteilchenphysik ist vor allem die Messung von Teilchenflüssen von Bedeutung, allerdings können die Flüsse nicht direkt gemessen werden, sondern nur indirekt über sekundäre Größen. Eine Technik diese Flüsse zu rekonstruieren ist die Methode der Entfaltung. Im Folgenden wird die Entfaltung für den Fall einer Flussrekonstruktion von hochenergetischen Myonen in IceCube betrachtet.

Die Grundlage der Entfaltung ist die in Gleichung (4.2.1) aufgeführte Fredholmsche Integralgleichung erster Art[76].

$$g(y) = \int_a^b A(y, x) f(x) dx + b(y) + \epsilon(y) \quad (4.2.1)$$

Die Funktion $f(x)$ ist die wahre Verteilung der gesuchten Größe, welche im Beispiel dieser Arbeit die primäre Energie der gemessenen Myonen ist. Diese gesuchte Energie wird mit der Antwort des Detektors $A(y, x)$ gefaltet. Dadurch wird eine bestimmte Energie x in eine gemessene Variable y überführt. Der zweite Term $b(y)$ in Gleichung (4.2.1) beschreibt den vorhandenen Untergrund für eine gemessene Observable y und der dritte Term $\epsilon(y)$ einen statistischen Fehler bei der Messung

dieser Observablen. Die Summe aus den zuvor genannten Termen beschreibt die vom Detektor gemessene Verteilung einer Observablen $g(y)$.

Bei realen Messungen werden Stichproben der Observablen y aus der Verteilung $g(y)$ gemessen. Zur Schätzung der Verteilung $g(y)$ bietet es sich an anhand der Stichproben ein Histogramm zu betrachten, in welchem die Häufigkeiten von Ereignissen in den Intervallen $[y_i, y_{i+1}]$ aufgeführt sind. Die Fredholmsche Integralgleichung erster Art lässt sich in eine diskretisierte Form überführen, welche in Gleichung (4.2.2) dargestellt ist.

$$\vec{g} = \underline{A} \vec{f} + \vec{b} + \vec{\epsilon} \quad (4.2.2)$$

Eine begrenzte Akzeptanz und ein begrenztes Auflösungsvermögen lassen sich anhand der Antwortmatrix \underline{A} illustrieren. Für den Fall eines idealen Detektors wäre die Antwortmatrix eine Einheitsmatrix. Eine Einheitsmatrix als Antwortmatrix würde bedeuten, dass jedes erzeugte Ereignis detektiert wird und die Energiezuordnung immer dem wahren Energiebin entspricht. Im Falle einer begrenzten Akzeptanz gilt für die Diagonalelemente A_{ii} der Antwortmatrix $A_{ii} < 1$, da weniger Ereignisse gemessen werden als generiert wurden. In diesem speziellen Fall dieser Arbeit sagt der Begriff Akzeptanz aus, wie hoch der Anteil der gemessenen Myonen an der Anzahl der gesamten Myonen ist, welche in der Atmosphäre generiert wurden. Diese Akzeptanz ist im Allgemeinen abhängig von verschiedenen Größen und im Falle von Myonen unter anderem von der Energie. Das begrenzte Auflösungsvermögen spiegelt sich dadurch wieder, dass die Antwortmatrix nicht mehr die Form einer Diagonalmatrix hat. Die Struktur der Antwortmatrix \underline{A} ist die einer Banddiagonalmatrix, da Ereignisse einer bestimmten Energie unter Umständen unter- oder überschätzt werden. Bei der vorliegenden Arbeit wird das Auflösungsvermögen dadurch begrenzt, dass die Signatur eines Teilchens mit einer bestimmten Energie im Detektor nicht immer identisch ist. Dies liegt daran, dass Energieverluste eines Myons stochastische Prozesse sind (vgl. Abschnitt 2.5). Dadurch können Myonen mit identischer Energie unterschiedlich hohe Energieverluste im Detektor erfahren, wodurch die Algorithmen zur Energierekonstruktion und somit der Detektor nur eine begrenzte Auflösung besitzen können.

Eine alternative Form der Diskretisierung von Gleichung (4.2.1) kann mit Hilfe von Basisfunktionen durchgeführt werden. Dieses Vorgehen wird hier erläutert, da es in der verwendeten Entfaltungsoftware TRUEE[77] genutzt wird. Die gesuchte Verteilung $f(x)$ kann mit Hilfe einer Summe aus m Basisfunktionen p_j und Amplituden a_j nach Gleichung (4.2.3) beschrieben werden.

$$f(x) = \sum_{j=1}^m a_j p_j(x) \quad (4.2.3)$$

Unter Verwendung dieser Basistransformation ergibt sich für die bereits beschriebene Gleichung (4.2.1) die folgende Form:

$$g(y) = \sum_{j=1}^m \left(a_j \int_a^b A(y, x) p_j(x) dx \right) + b(y) + \epsilon(y) \quad (4.2.4)$$

Die Diskretisierung der Verteilungen geschieht anschließend durch die Integrationen

$$g_i = \int_{y_{i-1}}^{y_i} g(y) dy, \quad b_i = \int_{y_{i-1}}^{y_i} b(y) dy \quad \text{und} \quad \underline{\underline{A}}_{ij} = \int_{y_{i-1}}^{y_i} A_j(y) dy \quad (4.2.5)$$

in dem Attributsintervall $[y_i, y_{i+1}]$. Da der statistische Fehler bei der Messung der Observablen klein gegenüber der Variation der Ereignisse ist, kann $\epsilon(y) = 0$ angenommen werden. Sollte die Verteilung $\epsilon(y)$ nicht vernachlässigbar sein, so könnte sie mit dem verbleibenden Untergrund $b(y)$ durch $b^*(y) = b(y) + \epsilon(y)$ zusammengefasst und die nachfolgende Berechnung analog durchgeführt werden. Dadurch ergibt sich für Gleichung (4.2.2)

$$\vec{g} = \underline{\underline{A}} \vec{a} + \vec{b}. \quad (4.2.6)$$

Der Vektor \vec{a} enthält die Koeffizienten der Basisfunktionen. Die gesamte Fredholmsche Integralgleichung erster Art ist in dieser Darstellung nur noch von den Amplituden der Basisfunktionen abhängig. Als Basisfunktionen bieten sich kubische B-Splines an, welche die Eigenschaft besitzen, dass über den gesamten Definitionsbereich $\sum_j p_j(x) = 1$ gilt. Die Antwortmatrix $\underline{\underline{A}}$ in Gleichung (4.2.6) kann anhand eines Testsignals für den wahren Detektor bestimmt werden. Dies ist in der Praxis meist nicht möglich, da oft kein solches Signal erzeugt werden kann. Für den Fall des IceCube-Detektors kann kein solches Testsignal erzeugt werden, da die detektierten Teilchen zu hochenergetisch sind um sie in Beschleunigern zu produzieren. Aus diesem Grund müssen zur Erzeugung von Testsignalen Monte-Carlo-Simulationen verwendet werden. Bei diesen Simulationen ist zu beachten, dass Simulation und gemessene Daten übereinstimmen sollten, wenn sie in der Entfaltung genutzt werden. Um dies sicher zu stellen, werden beide Verteilungen miteinander verglichen und der kumulative Abstand der Verteilungsfunktion als Maß für die Übereinstimmung betrachtet (vgl. Abschnitt 6.2).

Eine direkte Lösung der Gleichung (4.2.2) ist nach Gleichung (4.2.7) gegeben.

$$\vec{a} = \underline{\underline{A}}^{-1} (\vec{g} - \vec{b}) \quad (4.2.7)$$

Solch eine Lösung lässt sich jedoch nur für quadratische und invertierbare Matrizen $\underline{\underline{A}}$ eindeutig angeben. Alternativ kann die Bestimmung der Amplituden a_j

iterativ erfolgen, indem eine Anpassung an die gemessene Verteilung g durchgeführt wird. Da die gemessenen Ereignisse unabhängig voneinander sind, kann angenommen werden, dass die Anzahl der gemessenen Ereignisse g_i^m in einem Intervall einer Poisson-Verteilungen mit Mittelwert g_i folgen. Mit Hilfe einer Minimierung nach der Log-Likelihood-Methode[78] ergibt sich zur Minimierung die negative Log-Likelihood-Funktion

$$S(\vec{a}) = \sum_i (g_i(\vec{a}) - g_i^m \ln(g_i(\vec{a}))). \quad (4.2.8)$$

Diese direkte Lösung des Problems nach Gleichung (4.2.7) und die iterative Lösung nach Gleichung (4.2.8) können jedoch zu unphysikalischen Lösungen führen. Dies ist darin begründet, dass eine Entfaltung ein schlecht gestelltes Problem ist, wodurch die Lösung dieser zu einer stark oszillierenden Lösung führen kann. Aufgrund der Invertierung der Matrix werden kleine Eigenwerte der Matrix \underline{A} verstärkt, sodass geringe statistische Variationen einen signifikanten Einfluss auf das Ergebnis haben.

Um diese unphysikalischen Oszillationen zu unterdrücken wird bei der Entfaltung eine Regularisierung durchgeführt. Durch die Regularisierung wird die entfaltete Verteilung geglättet. Die Stärke der Regularisierung sollte bei der Anwendung möglichst gering sein, da ansonsten physikalische Merkmale in der entfalteten Verteilung unterdrückt werden könnten, jedoch genügend hoch um unphysikalische Oszillationen zu unterdrücken. Bei der Lösung des Problems mit Hilfe einer Likelihood Minimierung bietet es sich an einen Regularisierungsterm hinzuzufügen, welcher Oszillationen unterdrückt. Ein üblicher Regularisierungsterm r wird nach Tikonov berechnet und besteht für diesen Fall aus dem Quadrat der zweiten Ableitung der resultierenden Lösung[79].

$$r = \int (f''(x))^2 dx \quad (4.2.9)$$

Durch die Diskretisierung mit kubischen B-Splines kann Gleichung (4.2.9) in eine Matrixmultiplikation mit den Splinekoeffizienten umgeschrieben werden. Der resultierende Ausdruck wird somit durch Gleichung (4.2.10) beschrieben.

$$r(\vec{a}) = \int (f''(x))^2 dx = \vec{a}^T \underline{\underline{C}} \vec{a} \quad (4.2.10)$$

Bei der Matrix C handelt es sich um eine Banddiagonalmatrix, welche analytisch aus den Splinefunktionen bestimmt werden kann. Zur Kontrolle der Regularisierung ist es sinnvoll deren Einfluss einstellen zu können. Zu diesem Zweck wird der Parameter τ als Faktor zum Regularisierungsterm (4.2.10) hinzugefügt. Das sich so ergebene Minimierungsproblem kann als

$$S(\vec{a}) + \frac{1}{2}\tau \cdot r(\vec{a}) \rightarrow \min \quad (4.2.11)$$

geschrieben werden, wobei der Faktor $\frac{1}{2}$ zur leichteren formalen Berechnung gewählt wird. Im Anschluss an die Minimierung lässt sich aus den kubischen B-Splines ein Schätzer für die wahre Verteilung rekonstruieren. Die Bestimmung der geschätzten Verteilung aus den ermittelten Parametern der Basisfunktion geschieht durch eine binweise Integration mit

$$f_k = \frac{\left(\sum_{j=1}^m a_j \int_{x_{k-1}}^{x_k} p_j(x) dx \right)}{x_k - x_{k-1}}. \quad (4.2.12)$$

Dieses gesamte Verfahren zur Bestimmung der gesuchten Verteilung ist in [80] erläutert und wurde in der Software TRUEE[77] in C++ implementiert. Die Software TRUEE ist eine weiterentwickelte Version der Software RUN[81], welche in FORTRAN77 implementiert ist. In der hier vorliegenden Analyse wird TRUEE verwendet, da diese Software bereits in mehreren Bereichen in der Astroteilchenphysik erfolgreiche Ergebnisse lieferte[55, 74, 77, 82–85].

5 Verwendete Datensätze und Ereigniseigenschaften

Als Grundlage für jede Analyse in der Physik dienen die in dem Experiment gemessenen Daten. Da alle relevanten Schlüsse und Ergebnisse auf Basis dieser Daten geschehen, werden in diesem Kapitel die Datensätze beschrieben, welche dieser Arbeit zugrunde liegen. Zunächst wird eine Übersicht über die vorliegenden Daten und die verschiedenen Monte-Carlo-Datensätze gegeben. Weiterhin werden die enthaltenen Ereignisklassen beschrieben und eine Klassifizierung dieser wird vorgenommen. Abschließend werden die verwendeten Rekonstruktionsalgorithmen beschrieben, im Speziellen der DDDDR-Algorithmus.

5.1 Verwendete Datensätze

In der vorliegenden Analyse werden IceCube-Daten in der Datennahmeperiode vom 13.05.2011 bis zum 15.05.2012 analysiert. Diese Datennahmeperiode wird als IC86-I bezeichnet und beinhaltet alle aufgezeichneten Daten des InIce-, IceTop- und DeepCore-Detektors. Um eine künstliche Erzeugung von Signalen durch spezifisch gewählte Schnitte zu vermeiden, stehen während der Erstellung der Analyse nur 10 % der gesamten Daten zur Verfügung. Dieser Datensatz wird auch als Burn-Datensatz bezeichnet, da anhand dieser Daten die Analyse optimiert wird und aus diesem Grund keine physikalischen Ergebnisse aus dem Datensatz gezogen werden dürfen. Dieser Datensatz besteht aus jedem zehnten Datennahmeintervall, welches im Fall von IceCube in der Regel acht Stunden beträgt. Dadurch ist der Burn-Datensatz über das Jahr verteilt, sodass mit ihm auch saisonabhängige Analysen durchgeführt werden können. Ist die Analysemethode optimiert und liefert verlässliche Ergebnisse, so werden die verbleibenden 90 % der Daten freigegeben und die Analyse darf nicht mehr verändert werden. Aufgrund dieses Vorgehens werden die nötigen Analyse-schritte anhand dieses Burn-Datensatzes erläutert und das abschließend entstehende Spektrum wird mit den verbleibenden Daten gezeigt (vgl. Kapitel 6).

Als Monte-Carlo-Simulationen werden zwei separate Datensätze genutzt, welche verschiedene Energiebereiche der Primärteilchen abdecken. Beide Datensätze wurden

mit der Primärkomposition des Polygonato-Modells (vgl. Abschnitt 2.2.3) und der Software CORSIKA[86] produziert. Der Winkelbereich, in dem Ereignisse produziert wurden, ist für beide Datensätze 0° bis 90° im Zenitwinkel Θ und 0° bis 360° im Azimutwinkel ϕ . Die gesamten zur Verfügung stehenden Simulationsdaten bestehen aus einem Datensatz mit hochenergetischen Ereignissen mit der IceCube internen Nummer 10309[87] und einem Datensatz mit niederenergetischen Ereignissen der IceCube internen Nummer 9622[88]. Für den Niederenergie Datensatz wurden Primärteilchen im Energiebereich von 600 GeV bis 10^5 GeV und einem spektralen Index von $-2,7$ erzeugt. Der Hochenergie Datensatz besteht aus Primärteilchen im Energiebereich von 10^5 GeV bis 10^{11} GeV und einem spektralen Index von -2 . Der spektrale Index des 10309-Datensatzes entspricht nicht dem tatsächlichen spektralen Index der kosmischen Strahlung, jedoch wurde dieser gewählt damit eine höhere Anzahl an hochenergetischen Ereignissen erzeugt wird. Eine Übersicht der verwendeten Monte-Carlo-Simulationen ist in Tabelle 5.1 zu finden.

Tabelle 5.1: Beschreibung der verwendeten Datensätze. Aufgeführt sind die verwendeten Monte-Carlo-Datensätze und ihre internen Nummern. Zudem sind die jeweiligen Energiebereiche, in denen Primärteilchen produziert wurden, und die spektralen Indizes der Datensätze aufgeführt.

| Datensatz Nr. | E_{\min} / GeV | E_{\max} / GeV | γ |
|---------------|------------------|------------------|----------|
| 9622 | 600 | 10^5 | $-2,7$ |
| 10309 | 10^5 | 10^{11} | $-2,0$ |

Die Häufigkeiten der generierten Elemente in Abhängigkeit der Energie lässt sich, in den produzierten Monte-Carlo-Simulationen, auf Flüsse nach verschiedenen Primärmodellen umgewichten. Allgemein lassen sich differentielle Teilchenflüsse als

$$\frac{d\Phi}{dE} = \frac{dN}{dt dA d\Omega dE} \quad (5.1.1)$$

schreiben. Die Variablen in der Gleichung sind der Winkelbereich Ω , der Energiebereich E , die Messzeit t und die effektive Fläche A des Experiments. Der Fluss $d\Phi$ ist jeweils für einen Primärteilchentyp definiert, da im Falle der geladenen kosmischen Strahlung mehrere Elemente einen Teilchenschauer verursachen können. Die Berechnung von Gewichten w der Ereignisse findet über die Gleichung

$$w = \frac{dN_{\text{erwartet}}/dt dA d\Omega dE}{dN_{\text{generiert}}/dA d\Omega dE} \quad (5.1.2)$$

statt. Dadurch besitzt das Gewicht w die Einheit einer Rate und kann somit für eine beliebige Detektionszeit verwendet werden. In dieser Arbeit wird eine Gewichtung

gemäß des GaisserH3a-Primärmodells vorgenommen, welches auch die höchste Übereinstimmung mit den gemessenen Daten liefert.

Die gemessenen Daten und simulierten Datensätze sind bereits von der IceCube Kollaboration aufbereitet, um die Datenrate zu reduzieren und die Qualität der gesamten Daten zu erhöhen[89]. Diese Aufbereitung dient speziell der Analyse von Myon-Ereignissen und wird auch als Level 3 bezeichnet. Die vorherigen Anpassungen sind Schnitte auf verschiedene Variablen, welche nachfolgend aufgeführt sind.

1. Der mittlere Abstand der aktiven DOMs zur rekonstruierten Spur ist kleiner als 90 m **oder** die gesamte Ladung des Ereignisses ist größer als 100 pe^1 .
2. Die relative Likelihood des Fits, welcher die Ereignisrichtung rekonstruiert, muss kleiner sein als 8,3 **oder** die Anzahl der DOMs mit direkten Photonen muss größer sein als 6 und gleichzeitig muss die Spurlänge einen bestimmten Wert überschreiten, der wiederum von der Anzahl der DOMs mit direkten Photonen abhängt.
3. Ein zenitabhängiger Schnitt der gesamten Ladung Q_{tot} in dem Ereignis. Der Schnitt ist ein Polynom vierten Grades im Raum $\cos(\Theta)$ und $\log_{10}(Q_{\text{tot}})$.

Der erste durchgeführte Schnitt betrifft diese Analyse kaum, da hochenergetische Ereignisse eine höhere detektierte Gesamtladung als 100 pe besitzen. Der zweite Schnitt beschreibt die Qualität der rekonstruierten Teilchenspur. Dies ist für diese Analyse ein sinnvoller Schnitt, da aus dem rekonstruierten Winkel die Strecke berechnet wird, welche das Myon von der Oberfläche bis zum Detektor zurücklegt. Der verbleibende Schnitt ist nur für Ereignisse mit hohen Zenitwinkeln relevant und beginnt erst bei $\cos(\Theta) < 0,3$ Ereignisse mit mehr Ladung als 1000 pe zu verwerfen. Allerdings werden bei $\cos(\Theta) = 0$ lediglich Ereignisse mit weniger Ladung als $\approx 1350 \text{ pe}$ verworfen. Diese Grenze liegt nur geringfügig oberhalb des gewählten Qualitätsschnittes in Abschnitt 6.1.

¹pe: Photoelektronen

5.2 Ereignisklassen

Bei der Detektion von Myonen in IceCube treten prinzipiell drei verschiedene Ereignistypen auf. Diese Ereignisse werden in dieser Arbeit als hochenergetische Myonen (im folgendenen HE-Myon-Ereignisse), Ballon- und Myon-Bündel-Ereignisse bezeichnet. Alle detektierten Ereignisse sind Myon-Bündel-Ereignisse, allerdings ergibt sich die Einteilung in verschiedene Klassen durch die Energieverteilung der Myonen in dem Bündel oder anhand der Topologie der detektierten Ereignisse. In der vorliegenden Arbeit werden die Ereignistypen anhand der Quotienten

$$E_{\mu_{\text{rel}}} = \frac{E_{\mu_{\text{max}}}}{E_{\text{tot}}} \quad (5.2.1)$$

$$Q_{\text{rel}} = \frac{Q_{\text{max}}}{Q_{\text{tot}}} \quad (5.2.2)$$

festgelegt. Die Variable $E_{\mu_{\text{max}}}$ gibt die Energie des höchstenergetischsten Myons an, wohingegen E_{tot} die Gesamtenergie des Myon-Bündel-Ereignisses angibt. Im zweiten Quotienten Q_{rel} gibt die Variable Q_{max} die maximale Ladung eines DOMs im IceCube Ereignis an und Q_{tot} die in dem Ereignis detektierte Gesamtladung in allen DOMs. Auf die detaillierte Beschreibung der HE-Myon-, Ballon- und Myon-Bündel-Ereignisse wird im Folgenden näher eingegangen.

HE-Myon-Ereignisse erreichen den Detektor und beinhalten nicht nur das höchstenergetischste Myon, sondern meist auch niederenergetische Myonen. Bei HE-Myon-Ereignissen besitzt das Myon mit der höchsten Energie allerdings über 50% der gesamten Myon-Bündel-Energie. Diese Ereignisse zeichnen sich oftmals durch einen hohen stochastischen Energieverlust aus. Diesen Energieverlust kann nur ein hochenergetisches Myon erzeugen und dieser kann nicht durch eine Vielzahl an niederenergetischen Myonen vorgetäuscht werden. Hohe stochastische Energieverluste bilden eine kugelförmige Signatur im Detektor, welche von mehreren DOMs detektiert wird. Solch ein Ereignis ist in Abbildung 5.1 dargestellt.

Eine Signatur ähnlich zu der der HE-Myon-Ereignisse wird von Ballon-Ereignissen erzeugt. Bei Ballon-Ereignissen erweckt das Ereignis den Anschein einen hohen stochastischen Energieverlust erzeugt zu haben. Die aufgenommene Lichtmenge dieses Energieverlustes ist jedoch in einem einzelnen DOM konzentriert. Diese Signatur ist kein hochenergetischer Energieverlust, sondern niederenergetisch und nahe eines DOMs erzeugt worden. Somit ist diese Ereignisklasse eine rein räumliche Signatur eines Ereignisses, welche durch einen kleinen stochastischen Energieverlust an einem bestimmten Ort zu einer bestimmten Zeit erzeugt wird. Ein solches Ballon-Ereignis ist in Abbildung 5.2b dargestellt.

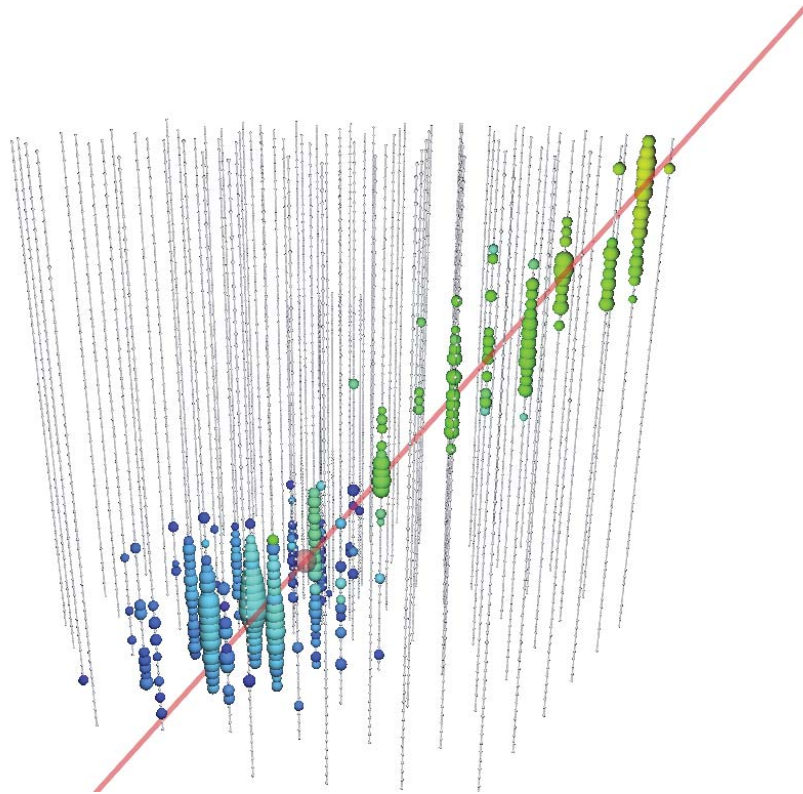


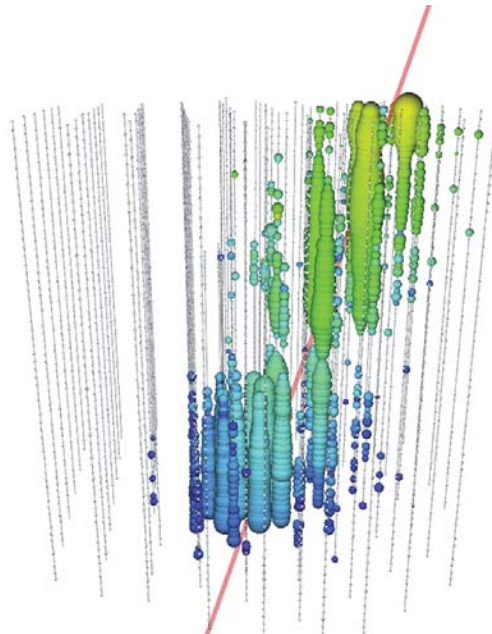
Abbildung 5.1: Signatur eines typischen HE-Myon-Ereignisses. Dargestellt ist ein hochenergetisches Myon im IceCube-Detektor. Eine animierte Version dieses Ereignisses ist in [90] zu finden. Die Größe der Kugeln gibt die gesamte detektierte Ladung in den jeweiligen DOMs an und die eingezeichnete Gerade die rekonstruierte Spur. Die Zeit ist in der Farbe der DOMs kodiert von früh(rot) bis spät(blau). Zu erkennen ist ein hochenergetischer, stochastischer Energieverlust des Myons.

Die verbleibende Ereignisklasse umfasst die Myon-Bündel-Ereignisse und deckt den verbleibenden Parameterraum der Myonereignisse ab. Diese Ereignisse enthalten kein hochenergetisches Myon und meist auch keinen hochenergetischen stochastischen Verlust oder eine Signatur, die diesen vortäuschen könnte. Diese Ereignisklasse bildet den Untergrund bei der Selektion der HE-Myon-Ereignisse. Solch ein Myon-Bündel-Ereignis ist in Abbildung 5.2a grafisch dargestellt.

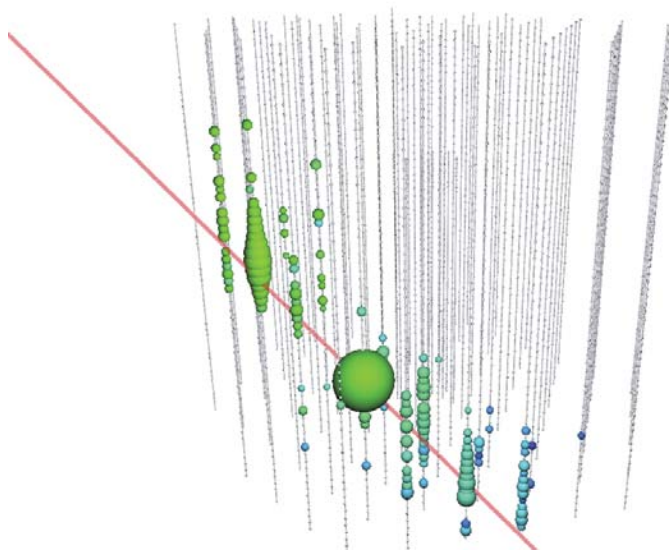
Tabelle 5.2: Beschreibung der Klasseneinteilung von Myonereignissen. Aufgeführt sind die Ereignisklassen mit den Parameterbereichen der Quotienten, welche diese Ereignistypen definieren.

| Ereignisklasse | Bedeutung | $E_{\mu_{\text{rel}}}$ | Q_{rel} |
|----------------|------------|------------------------|------------------|
| HE-Myon | Signal | ≥ 0.5 | < 0.2 |
| Ballon | Untergrund | ≥ 0.5 | ≥ 0.2 |
| Myon-Bündel | Untergrund | < 0.5 | – |

Diese Einteilung wird gewählt, da eine Energierekonstruktion von einzelnen Myonen nur bei den HE-Myon-Ereignissen möglich ist, denn dort wird ein hoher Anteil der detektierten Lichtmenge durch ein einzelnes Myon erzeugt. Alle Ereignisklassen mit den Definitionen im Parameterraum anhand der Gleichungen (5.2.1) und (5.2.2) sind in Tabelle 5.2 aufgeführt. Da diese Ereignisklassen Ausprägungen von atmosphärischen hadronischen Schauern sind, ist die Trennung dieser Klassen weitaus schwieriger als beispielsweise die Trennung von Neutrinos und atmosphärischen Myonen, da es sich hierbei um zwei verschiedene Teilchensorten handelt. Aus diesem Grund ist eine Verunreinigung bei der Trennung kaum zu vermeiden, da diese Klassen kontinuierlich ineinander übergehen.



(a) Myon-Bündel-Ereignis



(b) Ballon-Ereignis

Abbildung 5.2: Beispiele für Myon-Bündel- und Ballon-Ereignisse. Dargestellt sind ein simuliertes Myon-Bündel-Ereignis und ein simuliertes Ballon-Ereignis in IceCube. Animierte Version dieser Ereignisse sind in [91] und [92] zu finden. Die Größe der Kugeln gibt die gesamte detektierte Ladung in den jeweiligen DOMs an und die eingezeichnete Gerade die rekonstruierte Spur. Die Zeit ist in der Farbe der DOMs kodiert von früh(rot) bis spät(blau). Das Myon-Bündel-Ereignis hat einen nahezu gleichbleibenden Energieverlust pro Strecke während das Ballon-Ereignis einen hochenergetischen Energieverlust vortäuscht. 47

5.3 Algorithmen zur Ereignisrekonstruktion

Zur Separation der HE-Myon-Ereignisse von den Untergrundereignissen stehen für die aufbereiteten Datensätze auf Level 3 bereits einige Ergebnisse von Rekonstruktionsalgorithmen zur Verfügung. Diese Algorithmen dienen der Ereignisbeschreibung allgemein, jedoch wurden bereits Rekonstruktionen durchgeführt, die speziell für Myon-Ereignisse entwickelt wurden. Da einige der Attribute aus diesen Algorithmen zur Separation und abschließenden Entfaltung der Daten verwendet werden, wird hier ein Überblick über die Rekonstruktionsalgorithmen gegeben.

DDDDR Der **DDDDR**²-Algorithmus dient zur Rekonstruktion von Energieverlusten entlang einer Teilchenspur und kann somit stochastische Energieverluste rekonstruieren. Die Ergebnisse dieses Algorithmus sind aufgrund dessen von besonderer Bedeutung für diese Analyse und aus diesem Grund wird dieser Algorithmus in Abschnitt 5.4 im Detail erläutert.

HitStatisticsValues Bei den **HitStatisticsValues** handelt es sich um eine Sammlung von Kenngrößen, welche mit anderen Algorithmen berechnet werden und bereits viel Aufschluss über das detektierte Ereignis geben. Die Wichtigsten sind der Ereignisschwerpunkt, die größte deponierte Ladung in einem DOM und die gesamte detektierte Ladung des Ereignisses.

HitMultiplicityValues Der Algorithmus **HitMultiplicityValues** enthält die grundlegenden, zählbaren Informationen über ein Ereignis. Diese sind die Anzahl an aktiven DOMs, DOMs mit nur einem Puls und die aktiven Strings eines Ereignisses.

LineFit Bei **LineFit** handelt es sich um ein einfaches Richtungsbestimmungsverfahren, das mit Hilfe einer Minimierung nach der Methode der kleinsten Quadrate eine Spurrekonstruktion durchführt. Dieser Algorithmus ist sehr robust, liefert allerdings keine sehr genaue Berechnung der tatsächlichen Spur.

MPEFit Der **MPEFit**-Algorithmus nutzt alle detektierten Photonen in den jeweiligen DOMs und rekonstruiert mit einer Likelihood-Minimierung eine Teilchenspur. Die **MPE**³-Algorithmen konkurrieren mit den **SPE**⁴-Algorithmen, welche nur die ersten detektierten Photonen zur Rekonstruktion nutzen. Algorithmen, welche nur die ersten Photonen zur Rekonstruktion nutzen, werden in dieser Analyse nicht genutzt.

²DDDDR: **D**ata-**D**erived **D**eterministic **D**ifferential **D**eposition **R**econstruction

³MPE: **M**ulti-**P**hoto**E**lektron

⁴SPE: **S**ingle-**P**hoto**E**lektron

PoleMuonLlhFit PoleMuonLlhFit ist eine erste Spurrekonstruktion, welche am Südpol durchgeführt wird und eine Likelihood-Minimierung zur Ermittlung der Spur nutzt.

SplineMPE Richtungsrekonstruktionen in IceCube unterscheiden sich meist nur durch den Algorithmus, der zur Bestimmung der Teilchenspur aus den DOM-Signalen genutzt wurde. Der SplineMPE-Algorithmus ist der einzige, der die Eigenschaften des Südpoleises bei der Rekonstruktion verwendet. Die Streuwahrscheinlichkeiten für jede Position eines Photons im Detektor werden dabei in einem Raster berechnet und durch Splines innerhalb dieses Raumes interpoliert. Durch die Berücksichtigung der Eiseigenschaften weist der SplineMPE-Algorithmus die kleinsten Abweichungen zur tatsächlichen Richtung des Teilchens auf.

Truncated_Energy_BIN Dieser Algorithmus rekonstruiert die Energie des Myons anhand des durchschnittlichen Energieverlustes des Ereignisses entlang der Spur. Zu diesem Zweck wird die Strecke im Detektor in Teilstücke eingeteilt und 40 % der Teilstücke mit den höchsten Energiedepositionen werden nicht berücksichtigt. Diese Methode benötigt eine rekonstruierte Spur, welche in dieser Analyse der SplineMPE-Algorithmus liefert.

Einige dieser Rekonstruktionsalgorithmen werden auch auf verschiedene Zeitfenster des Ereignisses angewendet. Dies geschieht zum einen um einen zeitlichen Verlauf von einigen Variablen zu erhalten und zum anderen können Koinzidenzen von weiteren Ereignissen oder das Rauschen von zeitlich späteren DOMs vermindert werden. Der Nullpunkt dieser Zeitfenster ist das Triggersignal, welches eine Datennahme für dieses Ereignis veranlasst hat. Zur Verfügung stehen die Zeitfenster $A - E$, welche in Tabelle 5.3 aufgeführt sind.

Tabelle 5.3: Aufgeführt sind die Zeitfenster relativ zum Triggersignal in denen Rekonstruktionsalgorithmen ausgeführt werden können.

| Zeitfenster | $t_{\text{Start}} / \text{ns}$ | $t_{\text{Stop}} / \text{ns}$ |
|-------------|--------------------------------|-------------------------------|
| A | -15 | 15 |
| B | -15 | 25 |
| C | -15 | 75 |
| D | -15 | 125 |
| E | -15 | 175 |

5.4 DDDDR-Algorithmus

Der Algorithmus **D**ata-**D**erived **D**eterministic **D**ifferential **D**eposition **R**econstruction (im Folgenden als DDDDR bezeichnet) dient zur Rekonstruktion der Energieverluste, welche ein Myon bei der Propagation durch den IceCube-Detektor erfährt. Dieser Algorithmus ist von besonderer Bedeutung, da HE-Myon-Ereignisse oft einen hohen stochastischen Verlust aufweisen, der nicht durch viele niederenergetische Myonen erzeugt werden kann. Im Folgenden wird das Prinzip dieses Algorithmus beschrieben und die in dieser Arbeit verwendeten Einstellungen aufgeführt. Eine detailliertere Beschreibung des DDDDR-Algorithmus ist in [93] zu finden.

Da die entstandenen Photonen eine bestimmte Strecke im Eis propagieren, können diese vom Eis gestreut oder absorbiert werden. Diese Effekte bewirken einen exponentiellen Abfall der Intensität I abhängig von der Distanz d des DOMs zur Spur und der Absorptionslänge $\lambda_{\text{Ab.}}$ der Form:

$$I \propto \exp(-d/\lambda_{\text{Ab.}}(z)) \quad (5.4.1)$$

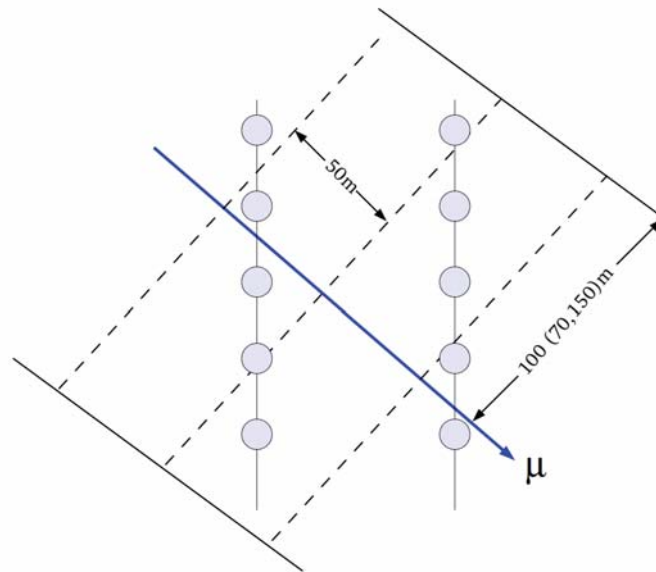
Die Absorptionslänge $\lambda_{\text{Ab.}}$ ist von der Tiefe des Eises z abhängig und wird innerhalb des DDDDR-Algorithmus berücksichtigt. Die detektierte Lichtintensität in den DOMs wird über Bins mit einer definierten Länge gemittelt, damit eine robuste Rekonstruktion der Energieverluste stattfindet. Weiterhin werden nicht alle DOMs mit ihren Pulsen betrachtet, sondern lediglich solche, welche sich in einem Zylinder mit einem definierten Radius um die rekonstruierte Spur befinden. Diese beiden Einstellungsparameter sind in Abbildung 5.3a grafisch dargestellt.

Das Resultat des DDDDR-Algorithmus ist ein Energieverlustprofil wie es in Abbildung 5.3b dargestellt ist. Die propagierte Tiefe bezeichnet die Strecke, die das Myon von der Oberfläche bis zum Energieverlust zurückgelegt hat. Um auf tatsächliche Energieverluste schließen zu können, wurden bei der Entwicklung des Algorithmus MC-Simulationen untersucht und aus den wahren Energieverlusten ein Skalierungsfaktor berechnet. Im Beispielprofil in Abbildung 5.3b sind zwei stochastische Verluste zu erkennen. Aus diesem Profil lassen sich weitere Attribute wie Median, Mittelwert oder Maximum berechnen. Weiterhin kann aus dem Profil eine Rekonstruktion der signifikanten stochastischen Energieverluste berechnet werden.

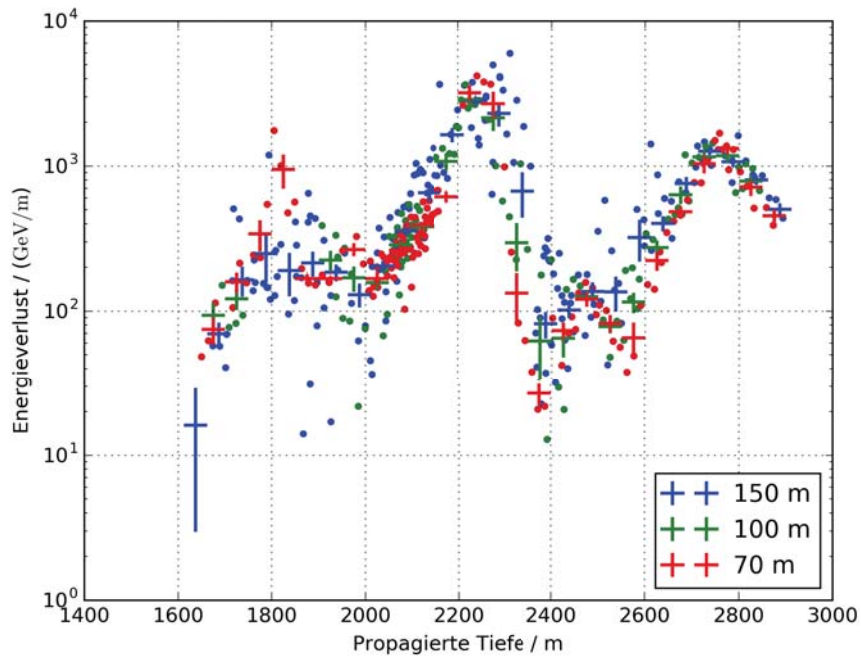
In der hier vorliegenden Analyse werden folgende Einstellungen für den DDDDR-Algorithmus gewählt:

- Spurrekonstruktion: `SplineMPE`
- Binbreite: 50 m
- Maximaler Spurabstand: 100 m
- Pulse: `SRTOfflinePulses`

Da die aufgenommene Ladung an jedem DOM aus den Pulsen rekonstruiert wird, muss für den Algorithmus eine Liste von Pulsen angegeben werden. Die verwendeten Pulse in dieser Arbeit sind `SRTOfflinePulses`. Dies sind rekonstruierte Pulse nach der Datennahme, wobei die Rekonstruktion dieser Pulse nicht am Südpol geschehen ist. Die Abkürzung `SRT` steht für Seeded Radius Time. Dadurch werden Pulse, die zeitlich 1000 ns nach dem Ereignis liegen, und Pulse, die weiter als 150 m von der rekonstruierten Spur entfernt sind, nicht betrachtet.



(a)



(b)

52

Abbildung 5.3: Einstellungsmöglichkeiten des DDDR-Algorithmus und dessen Resultat. (a) Schematische Darstellung der Einstellungsmöglichkeiten, welche bei dem DDDR-Algorithmus durchgeführt werden können[93]. Hierbei kann die Breite der Bins zur Rekonstruktion der Energieverluste und der zu berücksichtigende Abstand von der Spur eingestellt werden. (b) Dargestellt ist der rekonstruierte Energieverlust eines Ereignisses für verschiedene Spurbstände in Abhängigkeit der propagierten Tiefe.

6 Rekonstruktion des Myonspektrums



In diesem Kapitel werden die Analyseschritte, welcher zur Untersuchung der IceCube-Daten aus dem Jahr 2011 durchgeführt werden, detailliert beschrieben. Das Ziel dieser Analyse ist die Rekonstruktion des Spektrums hochenergetischer Myonen. Die Verarbeitung der Daten wird in einem mehrstufigen Prozess durchgeführt. Zunächst werden gerade Schnitte in bestimmten Attributen durchgeführt und ein Vergleich zwischen verwendeten Daten und Simulation gezeigt. Daraufhin wird die Attributauswahl zur Separation der Signalereignisse vom Untergrund erläutert. Abschließend wird das Spektrum der hochenergetischen Myonen mit Hilfe einer regularisierten Entfaltung rekonstruiert. Diese Einzelschritte sind zur Übersicht in der folgenden Aufzählung aufgeführt:

1. Vorselektion der Daten mit Hilfe von Qualitätsschnitten.
2. Attributselektion zur Trennung zwischen HE-Myon- und Untergrundereignissen.
3. Erstellung und Anwendung eines Modells zur Separation der HE-Myon- von Untergrundereignissen.
4. Rekonstruktion des Myonspektrums mit Hilfe einer Entfaltung.

Mit dem rekonstruierten Spektrum hochenergetischer Myonen können Flussnormalisierungen bestimmt und weitere Hypothesen für den Fluss hochenergetischer Myonen überprüft werden. Auf Grundlage dieser bestimmten Flussnormalisierungen werden Rückschlüsse auf den Fluss atmosphärischer Neutrinos gezogen.

6.1 Qualitätsschnitte

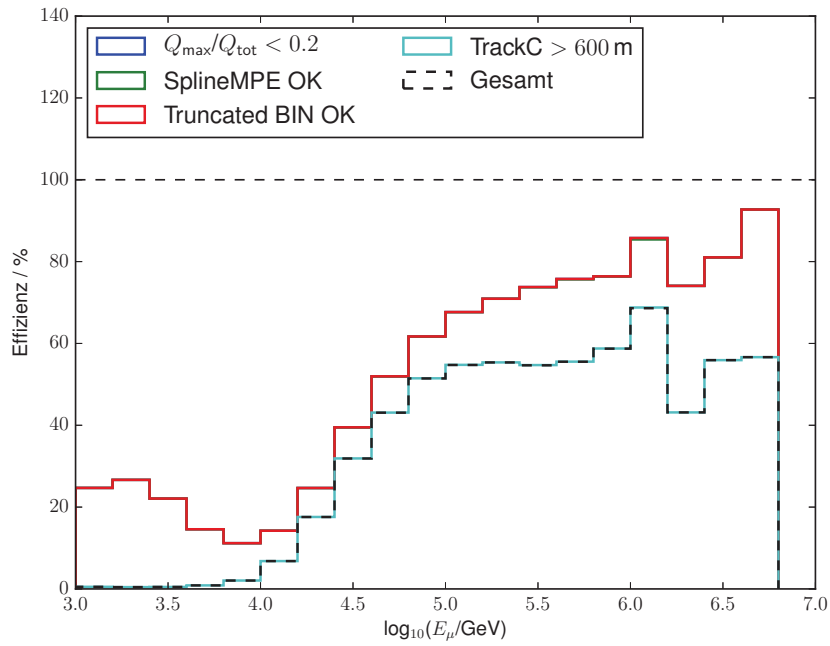
Bevor die verwendeten Daten mit fortgeschrittene Separationsalgorithmen getrennt werden, bietet es sich an Qualitätsschnitte durchzuführen. Diese Schnitte dienen dazu die Qualität der aufgenommenen Daten zu erhöhen und das Verhältnis zwischen HE-Myon-Ereignissen und Untergrundereignissen zu verbessern. Die folgenden Qualitätsschnitte werden in dieser Analyse durchgeführt und im Weiteren erläutert:

- $\text{TrackC} \geq 600 \text{ m}$
- $Q_{\text{tot}} \geq 1000 \text{ pe}$
- `SplineMPE` nicht fehlgeschlagen
- `SplineMPE_Truncated_Energy_BIN` nicht fehlgeschlagen
- $Q_{\text{max}}/Q_{\text{tot}} < 0.2$
- $\cos(\text{SplineMPE}_{\text{Zenit}}) > 0.1$

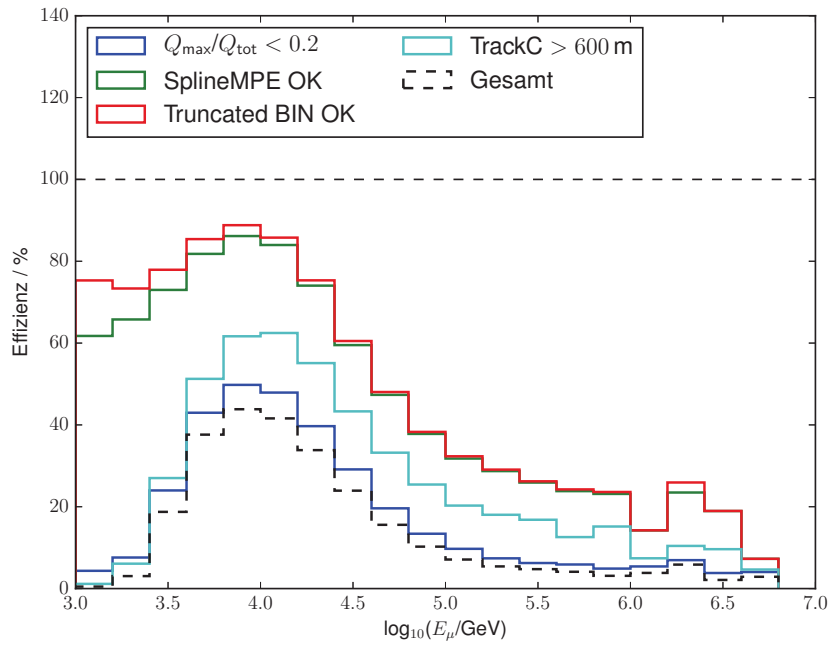
Der erste Qualitätsschnitt fordert, dass im Zeitfenster C (siehe Tabelle 5.3) mindestens eine Spurlänge von 600 m vorhanden ist. Durch diesen Schnitt werden Randereignisse verworfen, welche den Detektor nur streifen. Diese Ereignisse besitzen durch die geringe Strecke eine schlechtere Winkel- und Energieauflösung.

Der Schnitt in der gesamten detektierten Ladung Q_{tot} stellt sicher, dass niederenergetische Ereignisse verworfen werden. Diese Ereignisse sind für diese Analyse nicht relevant, stellen allerdings einen großen Teil der Signalereignisse in IceCube dar, da sie häufig von einzelnen niederenergetischen Myonen verursacht werden. Die Forderung hoher Ladung in Ereignissen bevorzugt allerdings auch die Selektion von schweren Elementen, welche jedoch deutlich seltener HE-Myon-Ereignisse erzeugen. Das Verhalten der Massenkomposition aufgrund der Selektionsschritte wird detailliert in Abschnitt 6.4 diskutiert.

Da der `SplineMPE`-Rekonstruktionsalgorithmus im Allgemeinen die geringste Abweichung zu der wahren Richtung von Ereignissen aufweist und in dieser Analyse verwendet wird, sollte dieser nicht fehlgeschlagen. Dies kann geschehen, falls der Fit für diesen Algorithmus nicht konvergiert. Eine genaue Richtungsrekonstruktion ist von großer Bedeutung, da die gemessene Energie im Detektor nicht der Oberflächenenergie entspricht. Um Rückschlüsse auf die Oberflächenenergie zu ziehen, muss die Strecke bekannt sein, die das Teilchen durch das Eis zurückgelegte und dort Energie verlor bis es den Detektor erreichte. Für die Berechnung dieser Strecke ist eine genaue Richtungsrekonstruktion von hoher Bedeutung.



(a) HE-Myon-Ereignisse



(b) Myon-Bündel- und Ballon-Ereignisse

Abbildung 6.1: Selektionseffizienz der vorhandenen Ereignisklassen. Dargestellt sind die Effizienzen der Qualitätsschnitte auf die Selektion der HE-Myon-Ereignisse (a) und der Untergrundereignisse (b).

In dem Attribut `SplineMPE_Truncated_Energy_BIN` wird ein weiterer Schnitt durchgeführt. Somit wird sichergestellt, dass der Algorithmus nicht fehlgeschlagen ist. Dieses Attribut ist nicht relevant für die Separation der HE-Myon-Ereignisse von den Untergrundereignissen, sondern wird durchgeführt, da dieses Attribut zur Energierekonstruktion verwendet wird (vgl. Abschnitt 6.6).

Der Qualitätsschnitt in dem Attribut Q_{\max}/Q_{tot} bewirkt, dass die zuvor genannten Ballon-Ereignisse verworfen werden. Dieser Schnitt wirkt sich nur auf die Untergrundereignisse aus, da HE-Myon-Ereignisse per Definition in Abschnitt 5.2 dieser Forderung genügen. Mit diesem Schnitt wird die Ereignisrate am Stärksten reduziert, im Vergleich zu allen von Level 3 auf Level 4 getätigten Schnitten.

Die Auswirkungen der einzelnen Schnitte auf HE-Myon-, Myon-Bündel- und Ballon-Ereignisse sind in den Abbildungen 6.1a und 6.1b dargestellt. Die Reduktion der HE-Myon-Ereignisse geschieht nur durch die Forderung, dass der `SplineMPE`-Algorithmus nicht fehlschlägt und eine Mindestspurlänge im Detektor gegeben sein muss. Ist der `SplineMPE`-Algorithmus gescheitert, so gilt `SplineMPE_Truncated_Energy_BIN` als Energie-Rekonstruktions-Algorithmus automatisch als gescheitert. Durch die hohe Qualität der Spur ist dies jedoch selten der Fall, wenn die Richtungsrekonstruktion erfolgreich ist, weswegen der Schnitt in Abbildung 6.1a einen identischen Verlauf hat und diese sich gegenseitig verdecken. Selbiges gilt für die geforderte Spurlänge und die Summe aller Schnitte, da dieser Schnitt bereits eine erfolgreiche Spur benötigt und die Spurlänge ansonsten den Wert Null erhält. Aufgrund dieser Schnitte werden für Energien $E \geq 10^5$ GeV HE-Myon-Ereignisse mit einer Effizienz von $\approx 50\%$ selektiert, während die Untergrundereignisse nur eine Selektionseffizienz von $\approx 5 - 10\%$ aufweisen. Zusammenfassend lässt sich feststellen, dass durch die gewählten Schnitte das Verhältnis zwischen Signal- und Untergrundereignissen für die weitere Analyse verbessert wird, ohne die Signalereignisse stark zu reduzieren.

6.2 Attributauswahl

Die Auswahl an Attributen, welche sich zur Trennung von HE-Myon- und Untergrundereignissen eignen, ist aufgrund der hohen Dimensionalität und der Komplexität des Problems nicht offensichtlich. Aus diesem Grund wird für die Attributauswahl der mRMR-Algorithmus verwendet (vgl. Abschnitt 4.1.1). Dieser Algorithmus selektiert die Attribute, welche die höchste Korrelation zur Signalklasse aufweisen, jedoch die geringste Korrelation zueinander besitzen. Bevor die so gewählten Attribute genutzt werden können, muss jedoch sichergestellt werden, dass sie in einer ausreichenden Qualität simuliert sind. Aus diesem Grund wird eine Vorselektion von

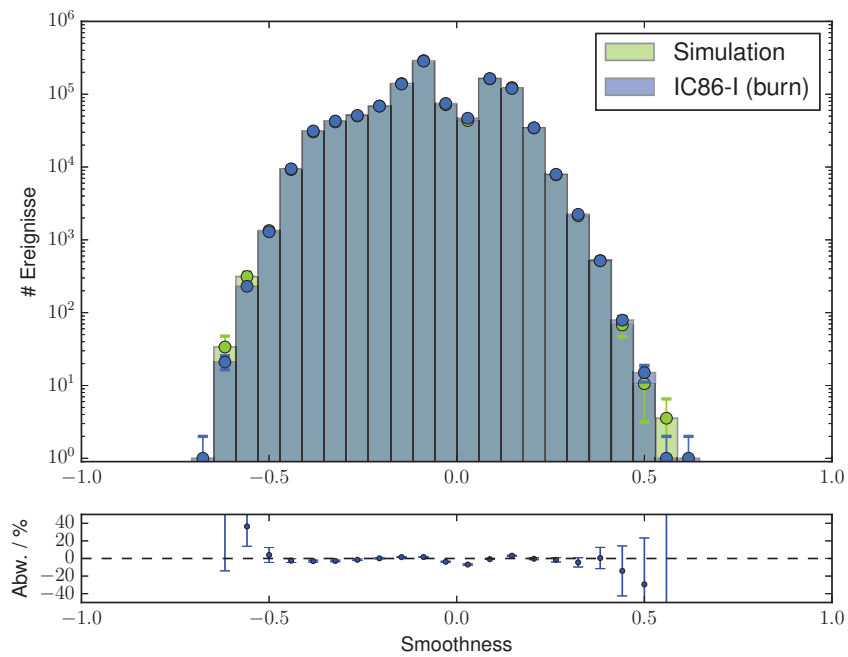
Attributen vorgenommen, bei denen die Simulation und die aufgenommenen Daten eine hinreichend hohe Übereinstimmung aufweisen.

Bei dieser Vorselektion spielt die absolute Normierung der simulierten Ereignisse keine Rolle. Ein Algorithmus, der eine quantitative Aussage über die Qualität der Übereinstimmung zweier Verteilungen liefert, ist die Methode des ungebinnten Kolmogorov-Smirnoff-Tests[94]. Dieser Test betrachtet die kumulierten Verteilungen zweier Mengen und berechnet aus dem maximalen Abstand dieser einen p-Wert, welcher eine Aussage macht, ob die verglichenen Verteilungen aus der gleichen Grundmenge stammen. Da die Anzahl der detektierten Daten- und simulierten MC-Ereignisse hoch ist und nicht alle Detektoreigenschaften in der Simulation berücksichtigt werden können, hat der berechnete p-Wert dieses Tests keine Aussage. Es kommt hinzu, dass die simulierten Ereignisse gewichtet erzeugt wurden, der Kolmogorov-Smirnoff-Test jedoch nur ungewichtete Verteilungen betrachtet. Stattdessen wird die maximale Distanz d_{\max} der kumulierten Verteilungen $F_{\text{Daten}}(x)$ und F_{MC} genutzt, um eine quantitative und sortierbare Güte der Übereinstimmung zu erhalten. Dieses Vorgehen lässt sich durch die folgende Gleichung beschreiben:

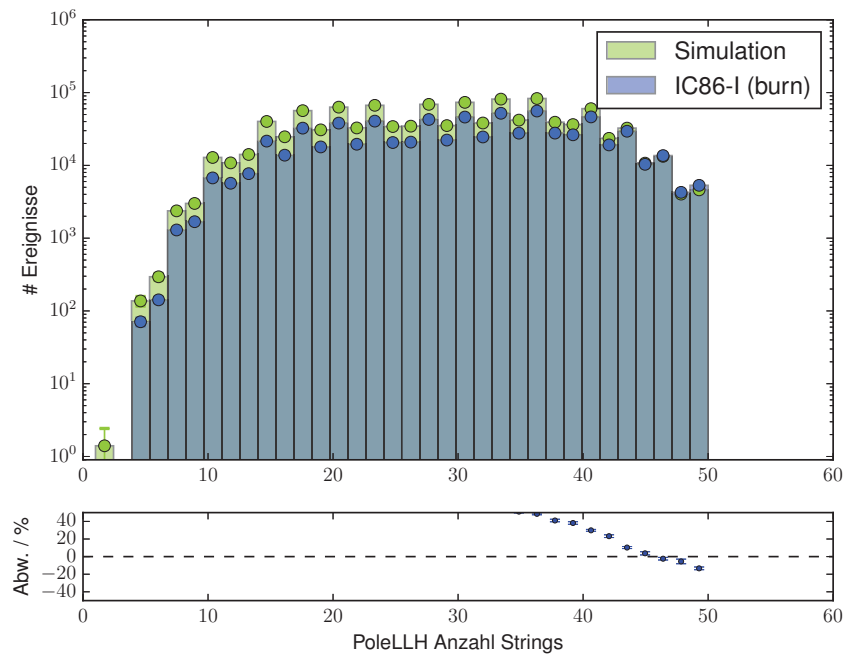
$$d_{\max} = \sup_x |F_{\text{Daten}}(x) - F_{\text{MC}}(x)| \quad (6.2.1)$$

Die Güte der Übereinstimmung der vorhandenen Attribute wird anschließend anhand der Distanzen d_{\max} aus Gleichung (6.2.1) bestimmt. Bei dem Vergleich der Distanzwerte des KS-Tests ist ein rapider Anstieg in der Verteilungen von d_{\max} der möglichen Attribute zu finden. Alle Attribute bis zu diesem rapiden Anstieg der Distanzwerte werden für die weitere Analyse verwendet.

Beispielhafte Verteilungen für eine hohe und eine niedrige Übereinstimmung sind in Abbildung 6.2 dargestellt. Eine hohe Übereinstimmung liefert das Attribut **Smoothness** (siehe Abbildung 6.2a), welches die Homogenität der Energieverluste auf der rekonstruierten Spur angibt. Der Wert ist auf die Spurlänge normiert und gibt den Abstand von Spurmittelpunkt im Wertebereich $[-1, 1]$ an. Ein positiver Wert bedeutet, dass der Schwerpunkt des Ereignisses nach dem Spurmittelpunkt ist. Ein Schwerpunkt vor dem Spurmittelpunkt ist favorisiert, da die Energie des Ereignisses mit der propagierten Länge der Teilchen abnimmt. Das Attribut **PoleLLH Anzahl Strings** hingegen weist eine schlechte Übereinstimmung zu den gemessenen Daten auf (siehe Abbildung 6.2b). Dies kann dadurch erklärt werden, dass es sich um einen schnellen Algorithmus handelt, welcher erste Schätzer für Ereigniseigenschaften berechnet. Diese Berechnung muss schnell geschehen, da sie am Ort der Datennahme geschieht und dort die Rechenkapazität gering ist.



(a)



(b)

Abbildung 6.2: Attributvergleich von Simulationen und Daten zur Bestimmung der Güte der Übereinstimmung. Dargestellt sind die Attribute *Smoothness* (a) und *PoleLLH Anzahl Strings* (b) aus der Monte-Carlo-Simulation im Vergleich mit dem Burn-Datensatz, sowie die prozentuale Abweichung. Das Attribut *Smoothness* weist eine hohe Übereinstimmung von Simulation und Daten auf, wohingegen das Attribut *PoleLLH Anzahl Strings* eine geringe Übereinstimmung zeigt.

Bei der Berechnung der Distanzen spielt auch die Annahme über das vorliegende Spektrum und die Komposition der kosmischen Strahlung eine Rolle. Werden die simulierten Daten mit einem anderen Primärmodell gewichtet, so kann es zu Abweichungen der aufgenommenen Daten und der Simulation kommen. Dies ist beispielsweise anhand von Abbildung 6.3 und des Attributs Q_{tot} dargestellt, welche die gesamte aufgenommene Ladung in einem Ereignis beschreibt. Die in Abbildung 6.3a dargestellte Parametrisierung der kosmischen Strahlung wird in dieser Arbeit verwendet, da sie das aktuellste und am häufigsten verwendete Modell beschreibt.

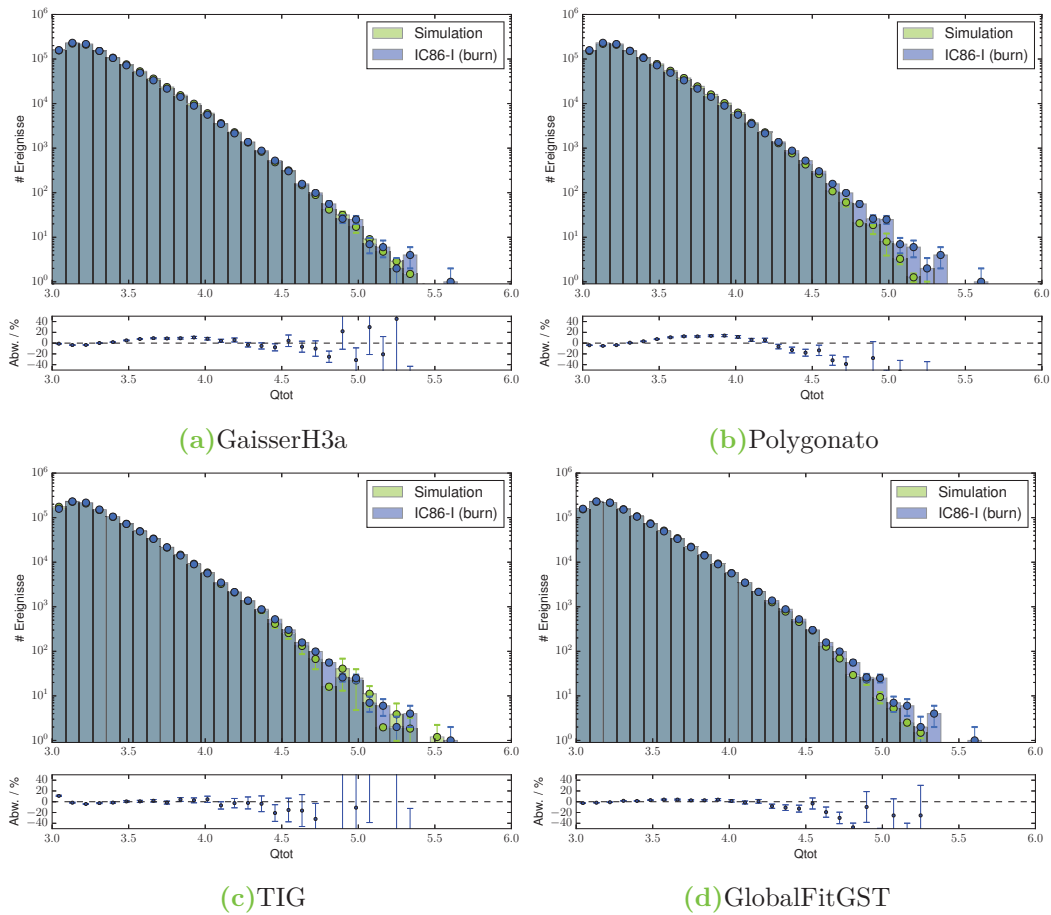


Abbildung 6.3: Vergleich der Gewichtungen der verfügbaren Flussmodelle der kosmischen Strahlung für die totale aufgenommene Ladung der Ereignisse. Dargestellt sind die Gewichtungen nach GaisserH3a, Polygonato, TIG und GlobalFitGST, welche in Kapitel 2.2.3 näher beschrieben sind.

Durch die Betrachtung der maximalen Abstände der kumulativen Verteilungen reduziert sich die Anzahl der verfügbaren Attribute von 400 auf 142 Attribute. Anschließend werden weitere Attribute erzeugt, welche das Potential haben die Separation zu verbessern. Diese selbst erzeugten Attribute sind Kombinationen aus bereits vorhandenen Attributen und erhalten das Präfix „U_“, um sie einfacher kenntlich zu machen. Durch die Hinzunahme weiterer Attribute erhöht sich die Gesamtanzahl an Attributen von 142 auf 165.

Mit diesen Attributen wäre es bereits möglich einen Attributsatz für die spätere Separation mit Hilfe des mRMR-Algorithmus zu erzeugen. Die Laufzeit dieses Algorithmus skaliert jedoch quadratisch mit der Anzahl der verfügbaren Attribute, sodass eine vorherige Reduktion der Attribute sinnvoll ist. Zu diesem Zweck werden die vorhandenen Attribute auf eine Korrelation nach Pearson[95] zueinander überprüft. Dies ist im Fall von IceCube besonders sinnvoll, da viele Attribute einen Schätzer für die gleiche gesuchte Größe liefern. Ein einfaches Beispiel sind die verschiedenen Zeitfenster, in denen die exakt gleiche Rekonstruktion durchgeführt wird (siehe Tabelle 5.3). Attribute, welche eine Korrelation von 0.9 oder höher haben, werden anschließend entfernt. Dies ist eine schwache Forderung, jedoch ist der exakte Wert nicht von Bedeutung, da stark korrelierte Attribute von dem mRMR-Algorithmus zusätzlich verworfen werden. Dadurch reduziert sich die Anzahl der zu betrachtenden Attribute von 165 zu 91.

Mit diesen 91 Attributen kann mit Hilfe des mRMR-Algorithmus ein Satz an Attributen erstellt werden, welcher sich zur Separation eignet. Jedoch stellt sich im späteren Verlauf der Analyse heraus, dass eine Favorisierung von niederenergetischen Ereignissen vorliegt, wenn zur Energie korrelierte Attribute verwendet werden. Dies ist für diese Analyse problematisch, da besonders der hochenergetische Anteil der Myonen von Interesse ist, da dort die prompte Komponente erwartet wird. Der Grund liegt darin, dass für die maschinellen Lernalgorithmen (vgl. Abschnitt 6.3) ein niederenergetisches Myon genau so wichtig ist wie ein hochenergetisches. Da niederenergetische Ereignisse eine höhere Wahrscheinlichkeit haben ein einzelnes hochenergetisches Myon im Vergleich zu den anderen Myonen im Bündel zu haben, werden Energieschätzer durch den Lernalgorithmus favorisiert. Aus diesem Grund ist es sinnvoll Attribute, die mit der Energie korreliert sind, nicht in den mRMR-Algorithmus und somit in der Separation zu verwenden. Dadurch werden zusätzlich 17 Attribute entfernt, sodass von den bisherigen 91 Attributen 74 zur Selektion bereitgestellt werden.

Der mRMR-Algorithmus wird auf diesen vorselektierten Attributen angewandt, wobei die Durchführung mehrmals auf verschiedenen zufällig gezogenen Teilmengen der Daten stattfindet. Dieses Vorgehen ist nötig, da das Ergebnis mit begrenzten

Monte-Carlo-Daten schwanken kann und es sinnvoll ist die Stabilität der Attributselektion für verschiedene Anzahlen an Attributen zu betrachten. Dazu werden die Monte-Carlo-Daten in fünf Teile unterteilt und anschließend die Stabilitäten nach Kuncheva und Jaccard berechnet (vgl. Abschnitt 4.1.1). An dieser Stelle sei nochmals erwähnt, dass eine Stabilität von 1 der Wert für die stabilste Attributauswahl darstellt. Die Stabilitäten sind grafisch in Abbildung 6.4 dargestellt. Anhand dieser Abbildung, ist zu erkennen, dass die ersten sechs Attribute immer gleich ausgewählt werden. Für sieben oder mehr Attribute wird weiterhin eine stabile Selektion nahe der maximalen Stabilität erreicht. Diese verringerte Stabilität kann auf die noch vorhandenen ähnlichen Attribute zurückgeführt werden. In dieser Arbeit wird eine Anzahl an 30 Attributen gewählt, da dies eine hohe Dimensionalität besitzt, die Anzahl der Attribute allerdings weiterhin so gering ist, dass die Erstellung des Random-Forest-Modells in einer moderaten Zeit geschieht. Die gesamte Attributselektion lässt sich zusammenfassend wie folgt darstellen:

$$400 \xrightarrow{-\text{MC/Daten}} 142 \xrightarrow{+\text{Eigene}} 165 \xrightarrow{-\text{Korrelierte}} 91 \xrightarrow{-\text{Energie}} 74 \xrightarrow{-\text{mRMR}} 30$$

Die gewählten Attribute sind in Tabelle 6.1 inklusive ihrer Reihenfolge, in der sie durch den mRMR-Algorithmus ausgewählt werden, aufgeführt. Weiterhin ist die in Abschnitt 4.1.3 beschriebene Attributwichtigkeit aufgeführt, welche durch den Random-Forest-Algorithmus berechnet wird (vgl. Abschnitt 4.1.3). Es ist festzustellen, dass eine hohe Übereinstimmung der Auswahl des mRMR-Algorithmus und der Random-Forest-Attributwichtigkeit vorliegt. Die ersten zehn Attribute des mRMR-Algorithmus werden auch vom Random Forest als wichtig erachtet, jedoch kommt es dort auch zu Abweichungen. Ein Grund für diese Abweichung könnte sein, dass der Random Forest zu Beginn der Separation bereits den vollen Datensatz verwendet und dadurch mehr Möglichkeiten zur Optimierung besitzt als der iterative mRMR-Algorithmus. Durch die vielfachen Schnitte, welche ein Baumalgorithmus ausführt, wird die Separation immer wieder auf Teildatensätzen durchgeführt. Dadurch kann eine stärkere Separationskraft und Korrelation in einem bestimmten Subparameterraum vorliegen, welche von dem mRMR-Algorithmus nicht erkannt wird. Zusammenfassend lässt sich feststellen, dass die Attributauswahl des mRMR-Algorithmus und der Vergleich zu der Attributwichtigkeit ein zufriedenstellendes Ergebnis liefern.

Tabelle 6.1: Auflistung der ausgewählten Attribute des mRMR-Algorithmus und die Wichtigkeit der jeweiligen Attribute bei der Erstellung des Random-Forest-Modells. In der Spalte RF sind die Wichtigkeiten für das Random-Forest-Modell aufgetragen, wohingegen in der Spalte σ_+ der Anstieg der Falschklassifizierungsrate für die jeweiligen Attribute aufgeführt ist (vgl. Abschnitt 4.1.3).

| mRMR | Attribut | RF | σ_+ |
|------|--|----|------------|
| 1 | U_DDDDR_PEAKVSMED | 1 | 8.27 |
| 2 | HitStatisticsValues.z_max | 8 | 2.65 |
| 3 | U_DDDDR_MEDVSMEAN | 2 | 7.88 |
| 4 | SplineMPE.empty_hits_track_length | 19 | 1.31 |
| 5 | U_DDDDR_MEANVSPEAK | 6 | 2.80 |
| 6 | DDDDR.nDOMsCascade | 24 | 1.03 |
| 7 | HitStatisticsValues.cog_z_sigma | 4 | 3.69 |
| 8 | U_SmoothnessE_ABS | 15 | 1.64 |
| 9 | LineFit_TTPParams.lf_vel_z | 5 | 2.91 |
| 10 | PoleMuonLlhFitCutsFirstPulseCuts.n_dir | 18 | 1.47 |
| 11 | HitStatisticsValues.cog_z | 20 | 1.29 |
| 12 | LineFit_TTPParams.n_hits | 3 | 3.85 |
| 13 | SplineMPE.track_hits_separation_length | 9 | 2.37 |
| 14 | SplineMPE.zenith | 7 | 2.78 |
| 15 | SplineMPEFitParams.nmini | 14 | 1.70 |
| 16 | SplineMPEDirectHitsA.n_early_doms | 30 | 0.52 |
| 17 | SplineMPEMuEXDifferential_r.value | 11 | 2.00 |
| 19 | SplineMPE.avg_dom_dist_q_tot_dom | 13 | 1.84 |
| 19 | LineFit_TTPParams.lf_vel | 16 | 1.50 |
| 20 | U_QmaxOverQtot | 12 | 1.84 |
| 21 | SplineMPEDirectHitsA.n_early_strings | 29 | 0.55 |
| 23 | U_SmoothnessD_ABS | 17 | 1.47 |
| 23 | U_CogRxy | 21 | 1.21 |
| 24 | U_DeltaZen | 22 | 1.18 |
| 26 | U_DDDDR_MEANVSCASC | 10 | 2.34 |
| 26 | SplineMPEDirectHitsE.n_dir_strings | 25 | 1.01 |
| 27 | SplineMPEDirectHitsC.n_dir_strings | 27 | 0.82 |
| 29 | MPEFitHighNoiseFitParams.nmini | 26 | 0.85 |
| 29 | PoleMuonLlhFitCutsFirstPulseCuts.s_dir | 28 | 0.73 |
| 30 | SplineMPE.track_hits_distribution_smoothness | 23 | 1.17 |

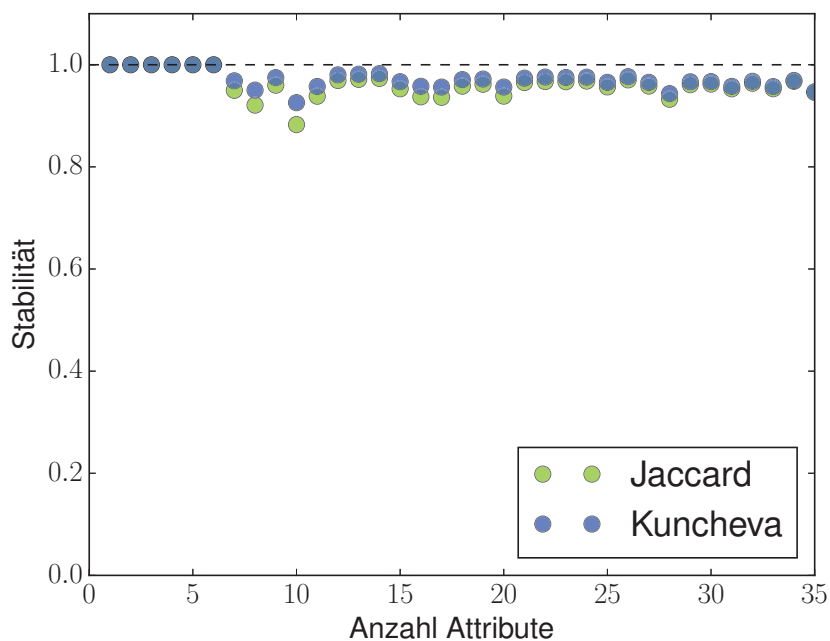


Abbildung 6.4: Stabilität der Attributauswahl mit Hilfe des mRMR-Algorithmus. In grün ist die Stabilität nach Jaccard und in blau nach Kuncheva dargestellt. Zur besseren Visualisierung ist in schwarz die maximale Stabilität mit dem Wert 1 eingezeichnet.

6.3 Separation

Nachdem die Attributauswahl durchgeführt wurde, wird in diesem Abschnitt die Separation der HE-Myonen von dem Untergrund beschrieben. Um diese Separation durchzuführen, wird der Random-Forest-Algorithmus (vgl. Abschnitt 4.1.3) verwendet. Zunächst werden die vorhandenen Monte-Carlo-Daten in einen Trainings- und einen Testdatensatz aufgeteilt. Dies dient zur Überprüfung einer Überanpassung des Random-Forest-Modells, da der Algorithmus sich an den Datensatz anpassen könnte und nicht an die Topologie der Signal- und Untergründereignisse. Für den Trainings- und Testdatensatz werden 50% der HE-Myon- und Myon-Bündel-Ereignisse für die jeweiligen Datensätze zufällig gezogen. Dies geschieht ohne Zurücklegen, damit zwei disjunkte Datensätze entstehen. Der Trainingsdatensatz wird zur Optimierung von Parametern des Modells verwendet, während der Testdatensatz in der weiteren Analyse, im Speziellen bei der Entfaltung, verwendet wird.

Für das Training des Random-Forest-Modells werden zur Erstellung die folgenden Einstellungen gewählt:

- 200 Entscheidungsbäume
- 30 Attribute
- 5 Attribute pro Knoten
- Vollständig ausgebaute Bäume

Die Auswahl von 30 Attributen wurde im vorherigen Abschnitt durchgeführt. Für einen Random Forest empfiehlt es sich $\log_2(A)$ oder \sqrt{A} Attribute pro Knoten zu wählen, wenn A die Gesamtanzahl an Attributen ist. Beide zuvor erwähnten Gleichungen liefern den Wert 5 als empfohlene Anzahl Attribute pro Knoten, welcher auch in dieser Arbeit verwendet wird. Im Weiteren wird ein Trainingsverhältnis von HE-Myon- zu Untergrundereignisse von 20000 : 60000 verwendet. Dieses Verhältnis spiegelt nicht das reale Verhältnis wider, jedoch liefert es hinreichend gute Ergebnisse bei geringer Laufzeit. Das wahre Verhältnis ist in Tabelle 6.2 dargestellt und beträgt ungefähr $10^5 : 10^6$, sodass der dominierende Anteil der Ereignisse aus Untergrund besteht. Dieses wahre Verhältnis wird für den Testdatensatz beibehalten. Zur Validierung der Ergebnisse wird eine fünffache Kreuzvalidierung auf dem zuvor erstellten Trainingsdatensatz durchgeführt, wie sie in Abschnitt 4.1.5 beschrieben ist.

Die Anzahl der gewählten Bäume ergibt sich in dieser Analyse durch die Anzahl von Bäumen, die keine signifikante Veränderung in der Konfidenzverteilung liefern. Dabei wird versucht so viele Bäume wie nötig und so wenig wie möglich zu wählen. Eine zu große Anzahl an Bäumen wirkt sich negativ auf die Rechenzeit aus, jedoch nicht auf die Qualität der Ergebnisse. Dies liegt daran, dass bei einer zu großen Anzahl an Bäumen sehr ähnliche Bäume trainiert werden, wodurch sich die Klassifikation nicht verbessert.

Das Verhalten der Konfidenzverteilung ist in den Abbildungen 6.8 und 6.9 für Random-Forest-Modelle mit 10, 50, 100, 150, 250 und 300 Bäumen im Vergleich zum verwendeten Random-Forest-Modell mit 200 Bäumen dargestellt. Die Vergleichsverteilungen sind auf die Anzahl der verwendeten Bäume renormiert, damit ein besserer Vergleich der Verteilungen durchgeführt werden kann. Für eine geringe Anzahl an Bäumen ist eine Verbesserung der Signalklassifikationen zu erkennen, wenn mehr Bäume hinzugenommen werden. Bei 150 Bäumen ist das Einsetzen einer Sättigung zu beobachten, welche die Konfidenzverteilungen bis zu der Maximalanzahl von 300 Bäumen nur marginal verändert. Der Fokus bei der Wahl der Anzahl an Bäumen liegt hier in der Verbesserung der Signaldetektion, welche bei einem Konfidenzwert von 1 liegt und dort ab einer Anzahl von 150 Bäumen nicht mehr

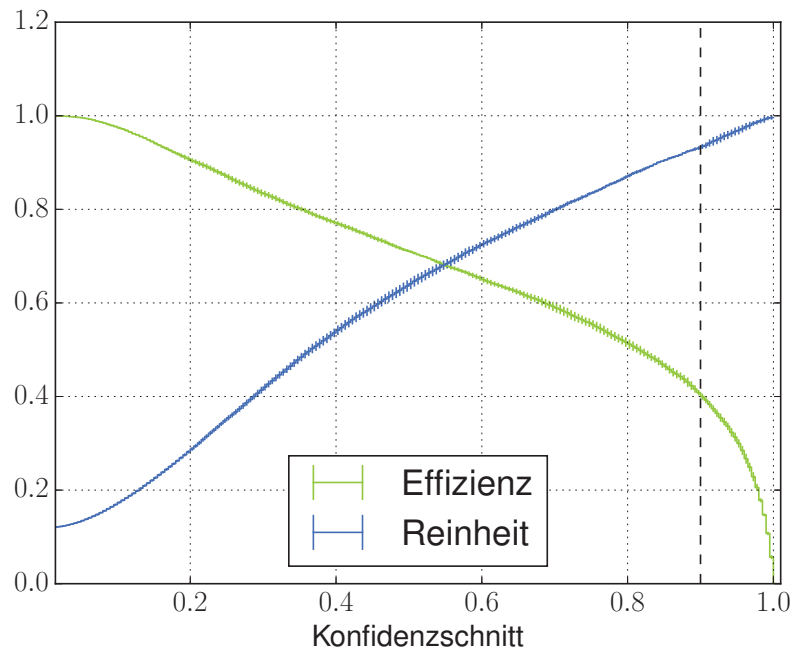
ansteigt. Um sicherzustellen, dass eine Anzahl an Bäumen im Sättigungsbereich gewählt wird, wird in dieser Analyse der nächst höhere Wert von 200 Bäumen gewählt.

Mit diesen 200 Entscheidungsbäumen des Random Forest wird das Modell erstellt. Damit eine Abschätzung der Ereigniszahlen und Fehler auf die Qualität der Separation bestimmt werden kann, wird eine fünffache Kreuzvalidierung durchgeführt um eine Überanpassung zu überprüfen. Die Methode der Kreuzvalidierung wurde bereits in Abschnitt 4.1.5 detailliert erläutert. Bei jedem Schritt der Kreuzvalidierung wird das zu Beginn in diesem Abschnitt verwendete Signal-Untergrund-Verhältnis von 20000 : 60000 zum Trainieren des Random-Forest-Modells verwendet.

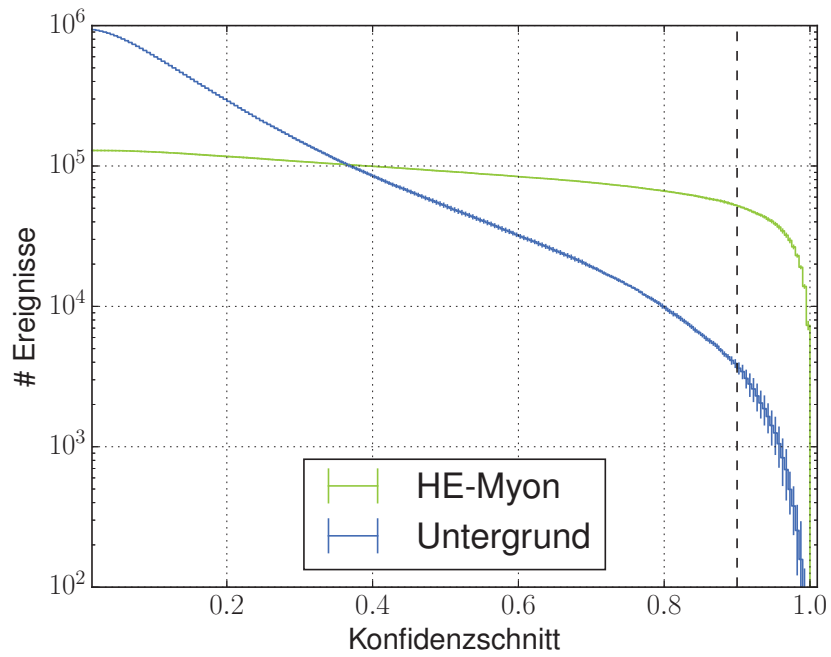
Das Resultat der Kreuzvalidierung sind fünf Modelle, welche auf fünf disjunkten Teilmengen des Trainingsdatensatzes angewendet werden. Dadurch werden alle Qualitätsparameter, welche in Abschnitt 4.1.4 beschrieben sind, für jede getestete Teilmenge des Trainingsdatensatzes bestimmt. Die gemittelten Qualitätsparameter und deren Standardabweichung sind in Abbildung 6.5a dargestellt.

Die berechneten Reinheiten und Effizienzen beziehen sich auf die reinen Ereignisse ohne die Energieabhängigkeit zu betrachten. Reinheit und Effizienz sind zwar von großer Bedeutung, allerdings im Hinblick auf die anschließende Entfaltung des hochenergetischen atmosphärischen Spektrums nicht immer optimal. Hilfreich bei der Wahl eines Konfidenzschnitts ist die Korrelation von Konfidenz und der Oberflächenenergie der hochenergetischen Myonen wie sie in Abbildung 6.6 dargestellt ist. Insgesamt ist ein leichter Trend zu erkennen, der niedrigerenergetische Ereignisse $E < 10^{4,5}$ GeV bei höheren Konfidenzen favorisiert. Dieser Trend ist schwach, jedoch ist weitaus wichtiger, dass höher energetische Ereignisse annähernd unkorreliert zu ihrer Energie klassifiziert werden. Für hohe Konfidenzen ist eine Verteilung an eindeutig zu klassifizierenden Ereignissen zu erkennen, welche das Maximum bei ungefähr $K = 1.0$ und $E = 10^{4,5}$ GeV besitzt und Ausläufer bis $K \approx 0.90$ hat. Aufgrund dessen wird ein Konfidenzschnitt $K^* = 0.9$ gewählt und Ereignisse mit $K > K^*$ ausgewählt, wodurch weiterhin eine hohe Anzahl an energetisch relevanten Ereignissen eingeschlossen wird. Dieser gewählte Konfidenzwert ist zusätzlich in Abbildung 6.5 dargestellt.

In Abbildung 6.5a ist die Reinheit und Effizienz der verwendeten Testdatensätze gezeigt. Der Verlauf beider Kurven ist wie er bei einer funktionierenden Separation zu erwarten ist, da die Effizienz sinkt, während die Reinheit steigt. Bei dem gewählten Konfidenzschnitt $K^* = 0.9$ wird eine Effizienz von $(40,8 \pm 0,6) \%$ erreicht und eine Reinheit von $(93,1 \pm 0,4) \%$. Diese Reinheit ist eine signifikante Verbesserung des Signal zu Untergrundverhältnisses von $\approx 1 : 10$ vor der Separation zu $\approx 10 : 1$ nach dem Konfidenzschnitt.



(a)



(b)

Abbildung 6.5: Dargestellt ist die Reinheit und Effizienz in (a) und die Anzahl der verbleibenden Ereignisse für die HE-Myon- und Untergrundereignisse in (b). Die X-Achse gibt den Schnittwert K^* für die Konfidenz K an, wenn nur Ereignisse mit einer Konfidenz mit der Bedingung $K > K^*$ gewählt werden.

Die Reinheit und Effizienz sind normierte Zahlen, welche sich nur bedingt auf weitere Analyseschritte übertragen lassen. Aus diesem Grund ist es sinnvoll die Anzahl der verbleibenden Signal- und Untergründereignisse in Abhängigkeit eines Konfidenzschnitts K^* darzustellen. Diese absoluten Ereigniszahlen im Burn-Datensatz für die jeweiligen Ereignisklassen und der 0.90 Konfidenzschnitt sind in Abbildung 6.5b grafisch dargestellt. Bei dem gewählten Konfidenzschnitt sind immer noch mehrere Zehntausend Signalereignisse vorhanden, allerdings auch mehrere tausend Untergründereignisse. Eine hohe Anzahl an verbleibenden Signalereignissen ist von Bedeutung, da das erwartete Spektrum der Myonen einen spektralen Index von $-3,7$ besitzt. Die exakten Zahlen werden in Abschnitt 6.4 diskutiert und sind in Tabelle 6.2 dargestellt.

Mit dem zuvor definierten Trainingsdatensatz wird nun das finale Random-Forest-Modell gebildet und auf den zuvor zurückgehaltenen Testdatensatz angewendet. Die erwarteten Reinheiten und Effizienzen und deren Varianzen können dadurch auf die Resultate des Testdatensatzes übertragen werden, da sie der selben Grundmenge entsprechen. Die Konfidenzverteilungen des Testdatensatzes separat aufgeteilt in die Ereignisklassen und die Konfidenzverteilung des Testdatensatzes und der detektierten Ereignisse sind in Abbildung 6.7 dargestellt. Anhand von Abbildung 6.7a ist zu erkennen, dass der Algorithmus die Klassifikation sinnvoll vornimmt. Es sind zwei klare Maxima zu erkennen, welche jeweils die Signal- und Untergrundklasse beschreiben. Während sich das Maximum der Signalklasse bei $K = 1$ befindet, ist das Maximum der Untergrundklasse nicht bei $K = 0$, sondern bei $K \approx 0.07$. Ein Effekt für diese Verschiebung könnte die Heterogenität der Untergrundklasse sein, da diese aus einer Vielzahl an chemischen Elementen besteht. Die HE-Myon-Ereignisse hingegen werden überwiegend von leichten Elementen produziert. Dies wird detaillierter in Abschnitt 6.4 diskutiert.

In Abbildung 6.7b ist die Konfidenzverteilung des Testdatensatzes und der detektierten Ereignisse dargestellt. Die Form der Konfidenzverteilungen weichen leicht voneinander ab, jedoch ist der prinzipielle Verlauf ähnlich. Für die Abweichung dieser Verteilung kann es mehrere Gründe geben, welche hier kurz erläutert werden. Abweichungen in diesen Verteilungen sind nicht direkt vergleichbar mit den gezeigten Daten und Monte-Carlo Vergleichen, welche in Abschnitt 6.2 beschrieben sind. Ein Grund für die mangelnde Vergleichbarkeit ist, dass die Konfidenzverteilung eine Projektion eines dreißigdimensionalen Raumes darstellt, da dreißig Attribute zur Separation genutzt werden. Eine Ursache für die Abweichung in dieser Projektion könnten nicht simulierte Detektoreffekte sein, welche durch unvollständige Eis-, DOM- oder die Staubschichtmodelle verursacht werden. Die zweite Ursache für die Abweichung könnte durch unvollständige Schauersimulation entstehen. In der Simulation der Schauer bei den hier relevanten Energien ist sowohl die Myonanzahl in Schauern, als auch das Pion-Kaon-Verhältnis nicht genau bekannt. Bei diesem

Effekt spielt auch die Komposition der kosmischen Strahlung eine Rolle, welche besonders im Bereich des Knies nicht exakt bekannt ist. Zusammenfassend ist festzustellen, dass trotz einer Vielzahl an unsicheren Parameter in der Simulation eine hohe Übereinstimmung der Konfidenzverteilung vorliegt und die Abweichung im Bereich $\leq 10\%$ liegen.

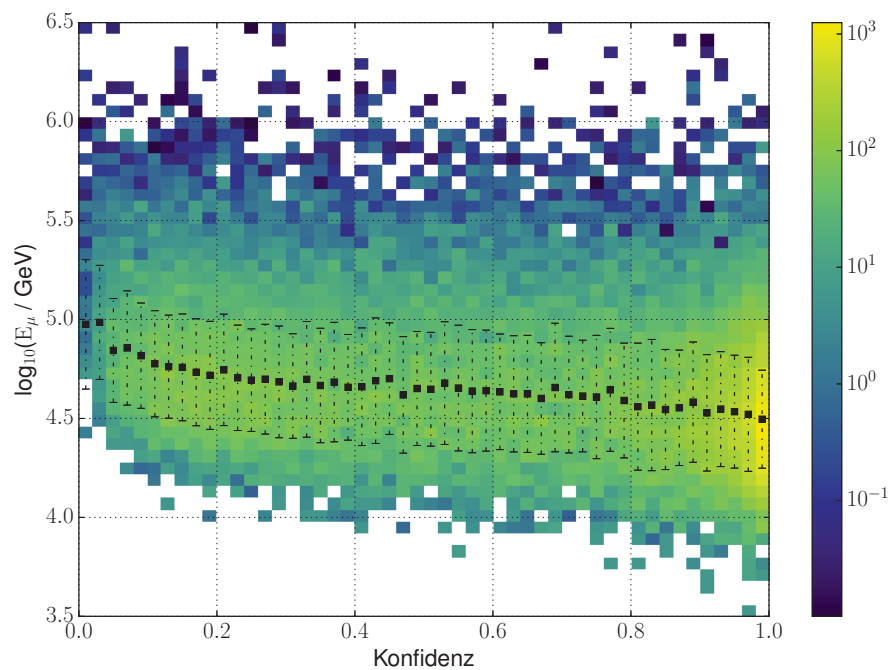
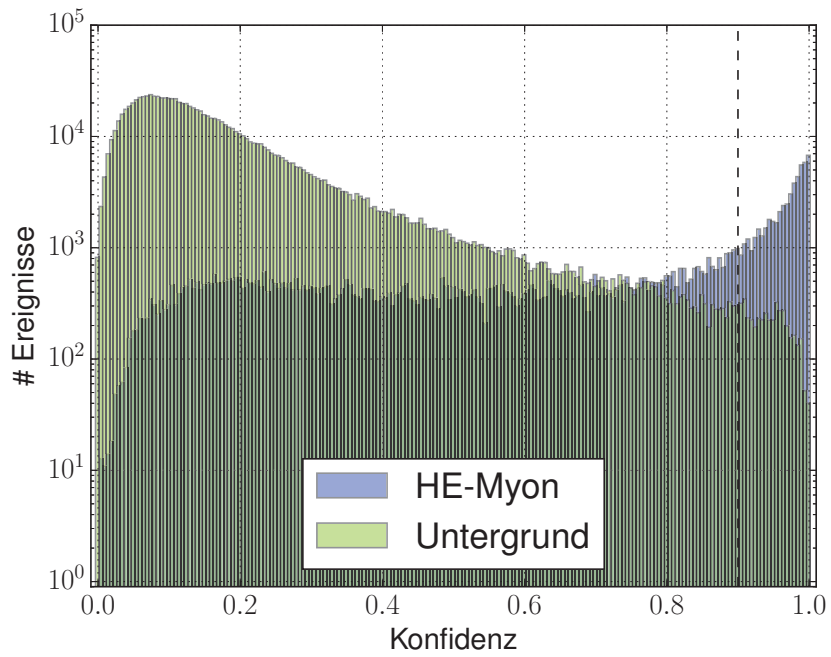
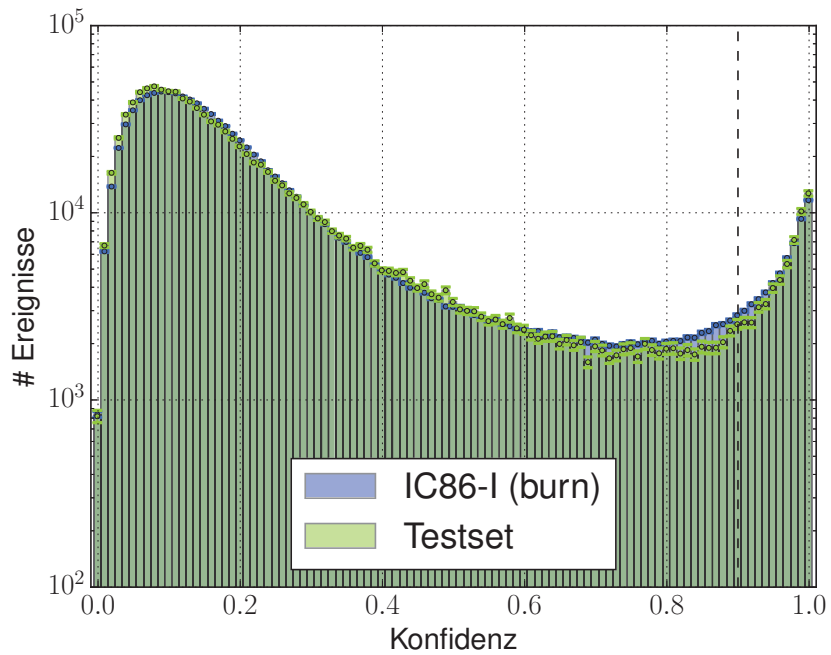


Abbildung 6.6: Dargestellt ist die Konfidenz aufgetragen gegen die Oberflächenenergie der HE-Myon-Ereignisse. Die Intensität in dieser Darstellung gibt die Anzahl an Ereignissen an, welche im Burn-Datensatz erwartet werden. Zusätzlich sind für die Konfidenzbins Mittelwerte und Varianzen gezeigt.



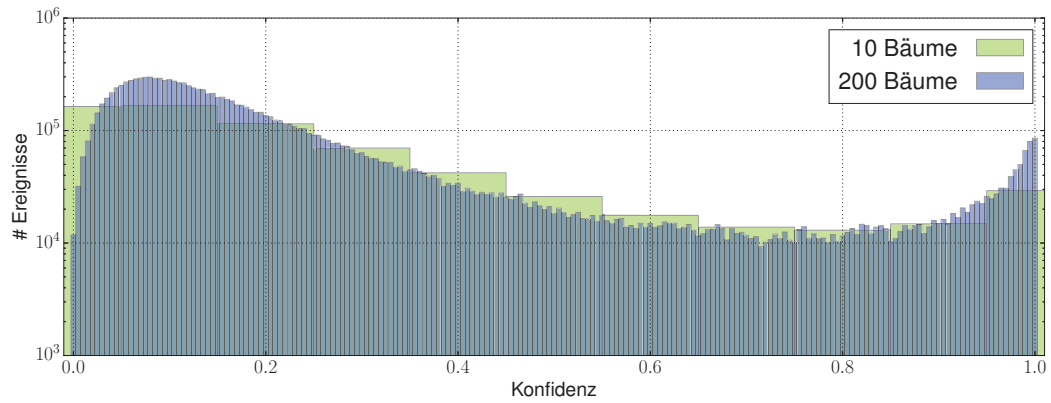
(a)



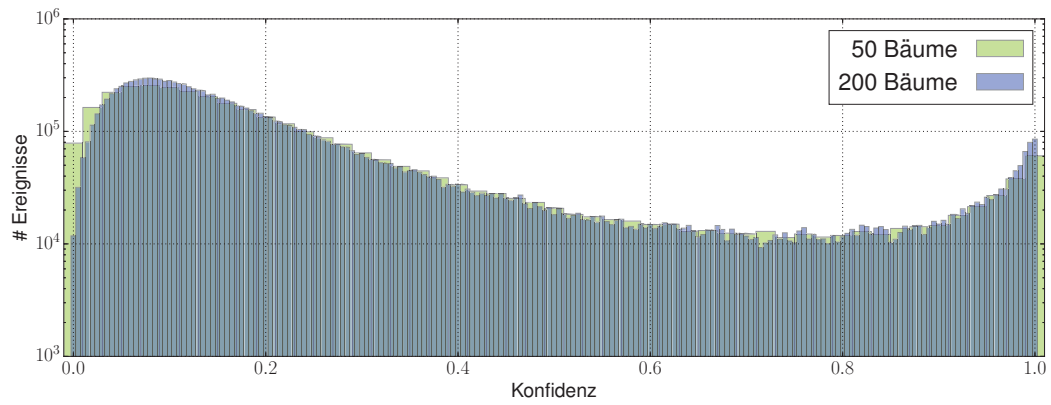
(b)

Abbildung 6.7: Ereignisanzahlen für die verschiedenen Ereignistypen (a) und der Vergleich des Testdatensatzes mit dem Burn-Datensatz (b). Weiterhin ist der Konfidenzschnitt von $K^* = 0.9$ eingezeichnet, welcher in der nachfolgenden Rekonstruktion des atmosphärischen Myonspektrums gewählt wird.

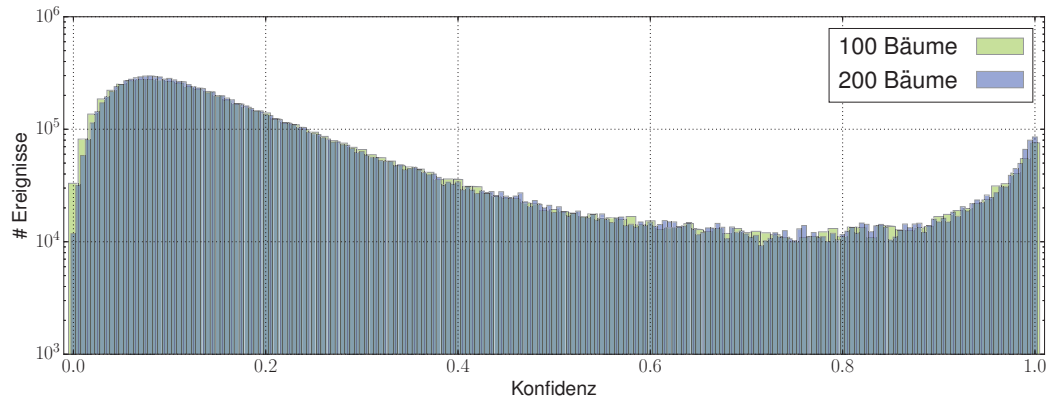
6 Rekonstruktion des Myonspektrums



(a)

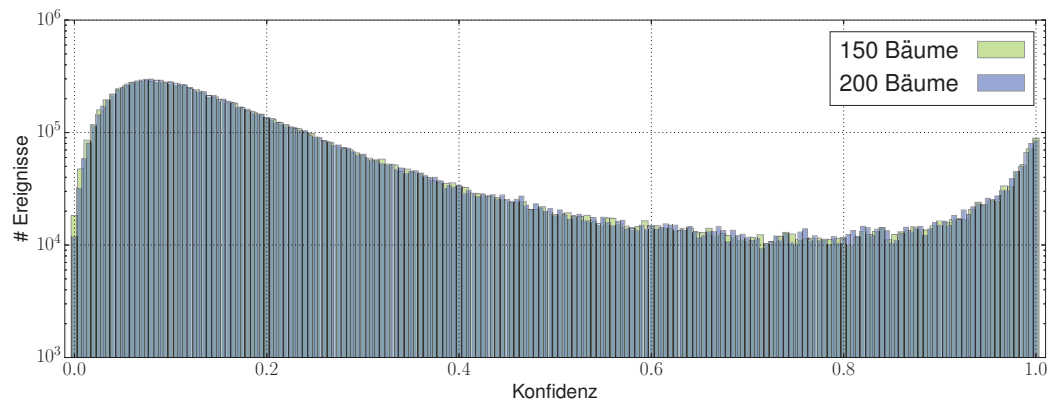


(b)

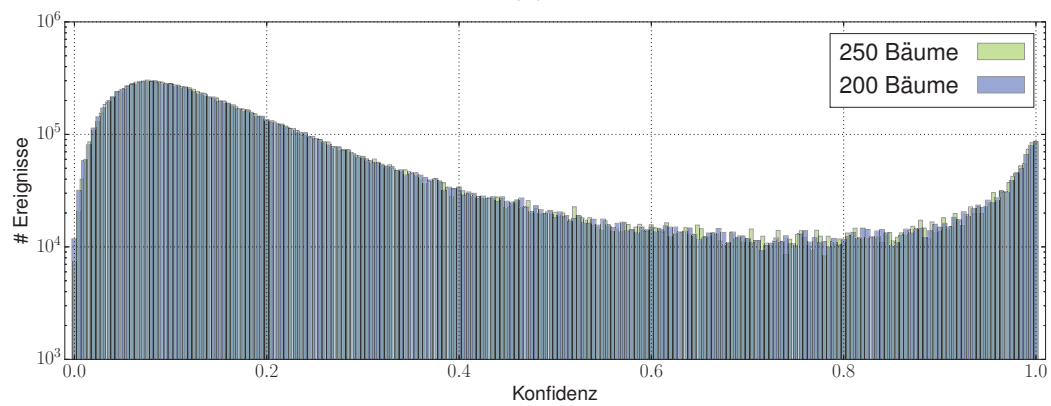


(c)

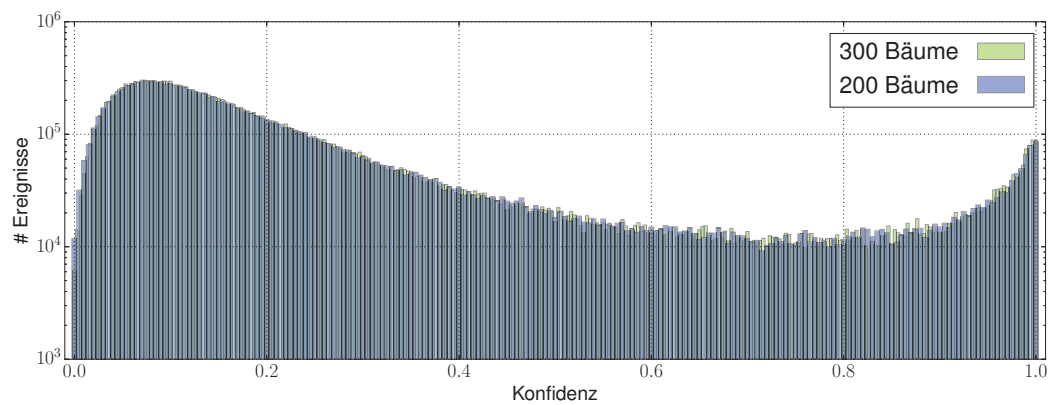
Abbildung 6.8: Dargestellt sind die Konfidenzverteilungen für 10, 50 und 100 Bäume im Vergleich zur gewählten Anzahl von 200 Bäumen für die Random-Forest-Modelle.



(a)



(b)



(c)

Abbildung 6.9: Dargestellt sind die Konfidenzverteilungen für 150, 250 und 300 Bäume im Vergleich zur gewählten Anzahl von 200 Bäumen für die Random-Forest-Modelle.

6.4 Analyse des finalen Datensatzes

Nachdem die Separation der HE-Myon-Ereignisse im vorherigen Abschnitt durchgeführt wurde, wird in diesem Abschnitt auf die Eigenschaften des daraus entstandenen finalen Datensatzes eingegangen. Dabei bietet es sich an, den finalen Datensatz in den verschiedenen Verarbeitungsschritten zu betrachten. Zunächst wird auf die Anzahl erwarteter Ereignisse auf verschiedenen Leveln der Analyse eingegangen. Diese Ereigniszahlen sind für den Hochenergie- (HE), Niederenergie Datensatz (LE) und den Burn-Datensatz in Tabelle 6.2 aufgeführt. Weiterhin sind die Ereignisanzahlen in die verschiedenen Ereignisklassen, welche in Abschnitt 5.2 beschrieben wurden, eingeteilt. Wie zuvor werden auch in Tabelle 6.2 die Gewichte entsprechend des GaisserH3a-Modells verwendet. Die Anzahl an einzigartigen ungewichteten Ereignissen ist in Tabelle 6.3 aufgeführt. Für Level 4 und 5 existieren keine Ballon-Ereignisse mehr, da diese anhand von rekonstruierten Größen klassifiziert werden. Die durchgeführten Qualitätsschnitte entfernen diese Ereignisklasse dadurch vollständig. Das Wegfallen dieser Ereignisse ist besonders bei niedrigen Primärteilchenenergien festzustellen, da dort wenig Ladung deponiert wird und dadurch auch einfacher ein großer Teil der Ladung in nur einem DOM konzentriert werden kann. Im Niederenergie Datensatz ist zusätzlich eine drastische Abnahme an Ereignissen von Level 3 auf Level 4 zu beobachten, welche durch den Qualitätsschnitt in der mindestens detektierten Ladung von 1000 pe erzeugt wird. Anhand der Tabelle wird zusätzlich verdeutlicht, dass die absolute Anzahl von Ereignissen in den simulierten Datensätzen mit dem IC86-I Burn-Datensatz verträglich ist.

Weiterhin ist die Selektionseffizienz der Untergrund- und HE-Myon-Ereignisse von Level 3 auf Level 5 zu betrachten. Diese Selektionseffizienz ist als Funktion der Energie für die beiden Ereignisklassen in Abbildung 6.10 dargestellt. Ab einer Energie von $10^{4,6}$ GeV (≈ 40 TeV) werden die hochenergetischen Myonen mit einer Effizienz von $\approx 10\%$ selektiert. Die Untergrundunterdrückung bei der gesamten Selektionskette liegt bei unter 1% und wird für hohe Energien auf Werte von unter 10^{-4} reduziert. Oberhalb einer Myonoberflächenenergie von $10^{5,8}$ GeV (≈ 600 TeV) werden keine Untergrundeignisse in dieser Arbeit selektiert. Dies ist eine vielversprechende Eigenschaft für ähnliche Analysen auf den nachfolgenden IceCube-Messungen der Jahren ab 2012, da dadurch der Energiebereich erweitert werden kann und der zu erwartende Untergrund gering ist. Zwar ist die Selektionseffizienz für höhere Energien leicht abnehmend, jedoch lässt sich dies voraussichtlich mit zusätzlich simulierten HE-Myon-Ereignissen in der Separation mit dem Random-Forest-Algorithmus korrigieren.

Ein weiterer interessanter Aspekt der Datenselektion ist das Verhalten der Massenzusammensetzung bei den verschiedenen Verarbeitungsleveln. Bei dieser Betrachtung ist zu

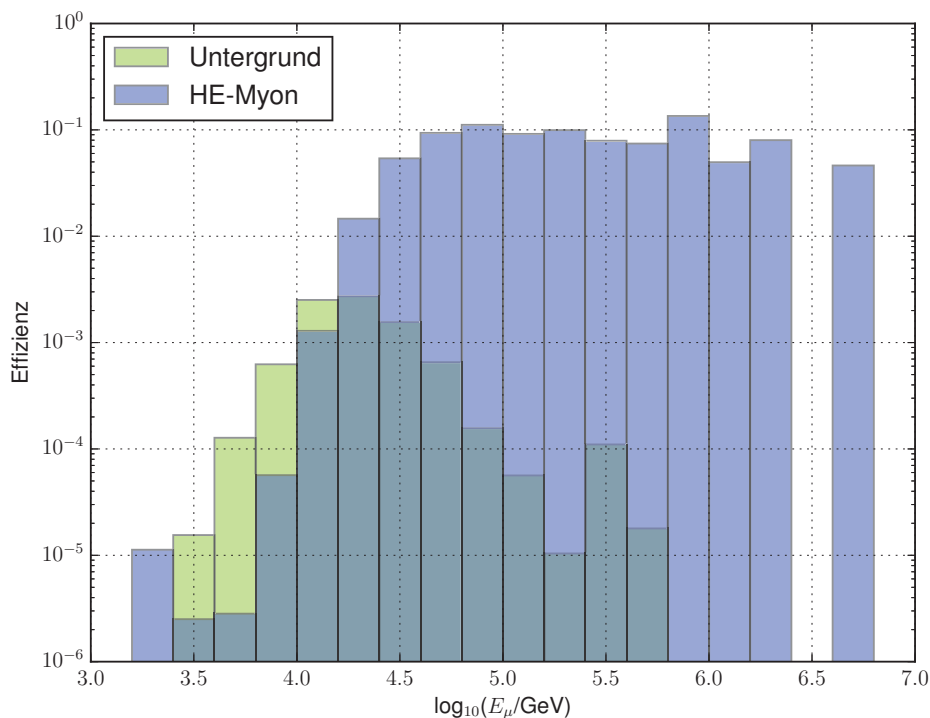


Abbildung 6.10: Dargestellt ist die Selektionseffizienz der Untergrund- und HE-Myon-Ereignisse nach Qualitätsschnitten und einem Konfidenzschnitt von $K^* = 0.9$ in Abhängigkeit der Myonenergie.

beachten, dass in der Simulation als Primärteilchen H , He , N , Al und Fe verwendet werden. Koinzidente Ereignisse werden allerdings anhand des Polygonato-Modells simuliert, sodass auch weitere Elemente in den Daten detektiert werden können. Nicht immer ist das simulierte Primärteilchen für das detektierte Ereignis verantwortlich, da koinzidente Ereignisse höher energetisch sein können und den Detektor zentraler treffen. Aus diesem Grund ist es sinnvoll die Elemente zu betrachten, welche von den Rekonstruktionsalgorithmen als Ereignis betrachtet werden.

In Abbildung 6.11 sind die Massenkompositionen für Level 3-5 gewichtet auf das GaisserH3a-Modell und der Detektionszeit des Burn-Datensatzes dargestellt. Für alle betrachteten Level ist zu erkennen, dass die rekonstruierten Elemente am Häufigsten auch die simulierten Primärteilchen sind. Dies ist zu erwarten, da nicht jedes Primärteilchen ein koinzidentes Ereignis besitzt und die Wahrscheinlichkeit bei hochenergetischen Primärteilchen ein weiteres hochenergetisches koinzidentes Ereignis zu erhalten gering ist. Bei dem Vergleich der Massenverteilung von Level 3 auf Level 4 wird deutlich, dass bei der Anwendung der Qualitätsschnitte schwere

Elemente favorisiert werden. Diese Favorisierung von schwereren Elementen wird durch den Qualitätsschnitt auf die Ladung von mindestens 1000 pe erzeugt. Ereignisse aus schweren Elemente besitzen eine höhere Anzahl an Myonen als reine Proton- oder He -Ereignisse, da mehr Nukleonen zur π - und K -Erzeugung zur Verfügung stehen. Diese π - und K -Mesonen zerfallen, wodurch mehr Myonen zum Ereignis beitragen. Ein Großteil der detektierten Lichtmenge wird durch Elektronen erzeugt, welche durch Paarproduktion von in Schauern enthaltenen Myonen erzeugt werden. Dadurch ist die detektierte Ladung proportional zu der Anzahl an Myonen, welche sich im Schauer befinden.

Die Änderung der Massenverteilung von Level 4 auf Level 5 bevorzugt wiederum leichtere Elemente, was auch dem erwarteten Ergebnis entspricht. Schwere Elemente verteilen ihre Gesamtenergie auf eine hohe Anzahl an Nukleonen, weswegen die Wahrscheinlichkeit sinkt ein hochenergetisches Myon im Schauer zu erzeugen. Bei Proton- oder He -Ereignissen kann in einer frühen Wechselwirkung ein hochenergetisches Myon erzeugt werden, welches einen hohen Anteil der Gesamtenergie des Primärteilchens trägt. Auch die Erzeugung prompter Hadronen findet in frühen Interaktionen des Primärteilchens statt, da diese massereich sind und eine hohe Schwerpunktsenergie bei der Wechselwirkung benötigen.

Tabelle 6.2: Gewichtete Ereignisse für verschiedene Level der Verarbeitung. Weiterhin sind die Ereigniszahlen in die verschiedenen Ereignisklassen und Datensätze aufgeschlüsselt.

| Datensatz | Klasse | Level 3 | Level 4 | Level 5 |
|------------|-------------|-----------|-----------|---------|
| LE: 9622 | HE-Myon | 506 282 | 19 478 | 10 601 |
| | Myon-Bündel | 338 562 | 129 459 | 673 |
| | Ballon | 1 045 735 | 0 | 0 |
| HE: 10309 | HE-Myon | 531 109 | 116 710 | 41 419 |
| | Myon-Bündel | 1 599 604 | 884 423 | 3 858 |
| | Ballon | 607 886 | 0 | 0 |
| IC-86 Burn | | 4 621 075 | 1 077 756 | 57 433 |

Tabelle 6.3: Einzigartige Ereignisse der Datensätze für die verschiedenen Analyselevel.

| Datensatz | Klasse | L3 | L4 | L5 |
|------------|-------------|-----------|-----------|--------|
| LE: 9622 | HE-Myon | 74 347 | 2 878 | 803 |
| | Myon-Bündel | 47 493 | 18 413 | 48 |
| | Ballon | 153 589 | 0 | 0 |
| HE: 10309 | HE-Myon | 222 553 | 63 899 | 6 434 |
| | Myon-Bündel | 2 073 225 | 1 262 145 | 1 245 |
| | Ballon | 139 761 | 0 | 0 |
| IC-86 Burn | | 4 621 075 | 1 077 756 | 57 433 |

6 Rekonstruktion des Myonspektrums

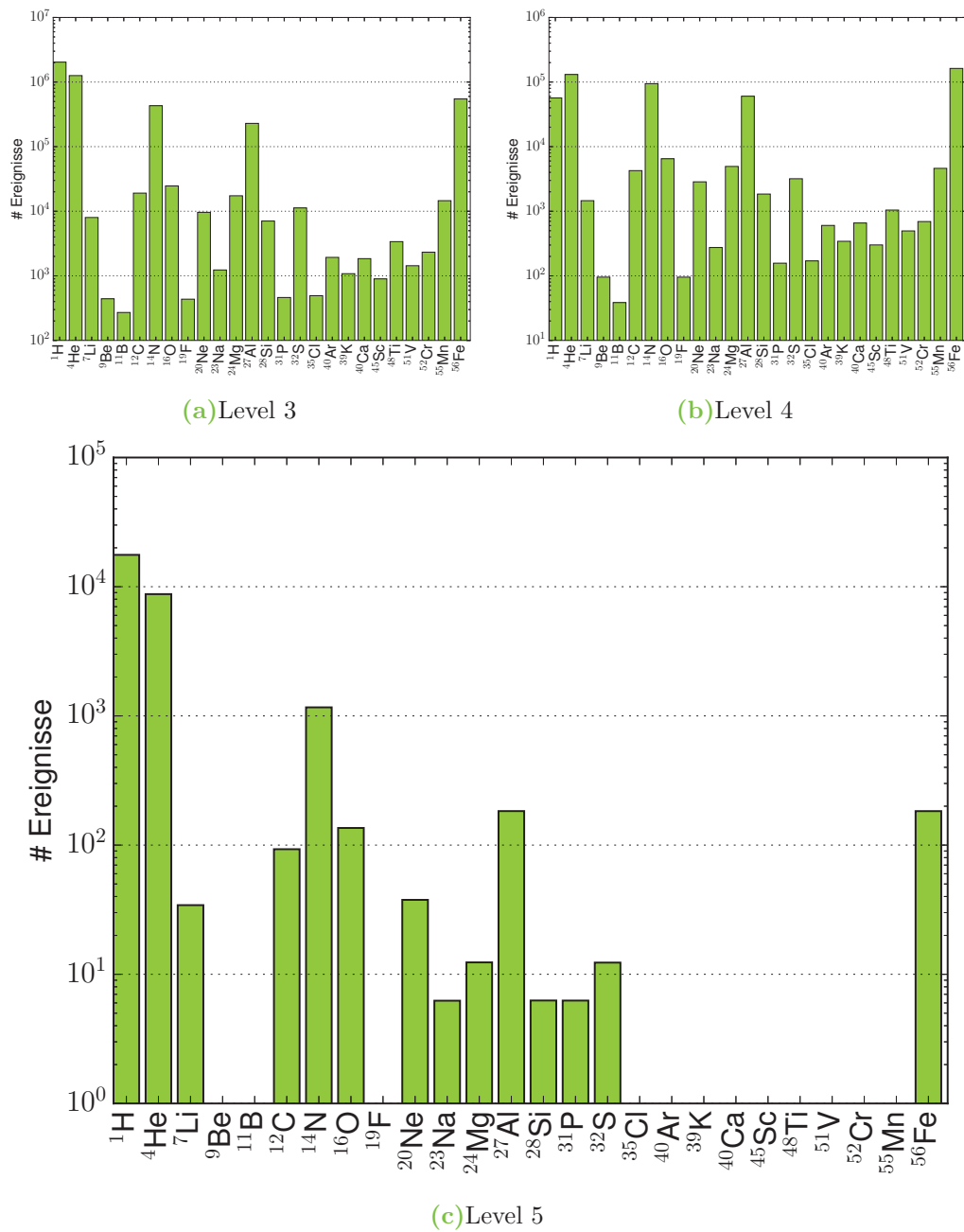


Abbildung 6.11: Massenkomposition für die verschiedenen Level, welche die Daten durchlaufen. a) Ursprüngliche Massenkomposition. b) Massenkomposition nach der Anwendung von Qualitätsschnitten. c) Massenkomposition nach dem Schnitt auf die Random-Forest-Konfidenz von $K \geq 0.90$.

6.5 Neutrinoenergie

In dem zuvor beschriebenen finalen Datensatz wurde auf die verbleibende Anzahl an Untergrundereignissen hingewiesen. Diese Untergrundereignisse sind nur Ereignisse der kosmischen Strahlung, welche keine HE-Myon-Ereignisse sind. Da die Rate an Ereignissen für hohe Zenitwinkel stark sinkt, bleibt zu überprüfen, ob neutrinoinduzierte Ereignisse einen signifikanten Untergrund bilden. Zu diesem Zweck wird ein Myonneutrinoenergie mit der internen Nummer 10602[96] betrachtet, welcher folgende Eigenschaften besitzt:

- Teilchenart: Myonneutrinos
- Energiebereich: $10 \text{ GeV} - 10^9 \text{ GeV}$
- Zenitwinkel: $0^\circ - 180^\circ$
- Energiespektrum: E^{-2}

Der Fluss der Myonneutrinos, welcher von IceCube detektiert werden kann, lässt sich in eine konventionelle, prompte und astrophysikalische Komponente unterteilen. Wie für die kosmische Strahlung gibt es verschiedene theoretische Modelle, die den Fluss der einzelnen Komponenten beschreiben. Für diese Analyse wird der Fluss der konventionellen Myonneutrinos nach Honda[97] und der prompte Fluss nach Enberg[98] gewichtet. Der Nachweis astrophysikalischer Neutrinos gelang erst vor einigen Jahren, weswegen der Fluss mit großen Unsicherheiten aufgrund der geringen Ereignisrate belegt ist[9]. Diese Unsicherheit ist im Rahmen dieser Analyse gravierender, da der Fluss nur in der Leptonenfamilie der Myonen benötigt wird. Der verwendete astrophysikalische Fluss in dieser Arbeit ist nach [99] berechnet worden.

Nachdem die Gewichtung der Myonneutrinos berechnet wurde, werden die Qualitätschnitte aus Abschnitt 6.1 auf den Datensatz angewendet. Die erwartete Anzahl an Ereignissen, welche diese Qualitätschnitte passieren, beträgt $\mathcal{O}(10)$ für den verwendeten Datensatz und das Flussmodell. Solche Ereigniszahlen sind zu erwarten, da Ereignisse nur aus Richtung Süden ($\cos(\Theta) > 0.1$) betrachtet werden. Dadurch reduziert sich der totale Fluss der Myonneutrinos in diesem Winkelbereich auf $\approx 1\%$ im Vergleich zur gesamten Erdatmosphäre, die bei Neutrinoanalysen verwendet wird. Zusätzlich zum geringen Fluss erweist sich die niedrige Propagationsstrecke im Interaktionsmedium Eis als kritisch, da nur ein geringer Anteil an Neutrinos im Detektor wechselwirkt. Meist interagieren die Neutrinos im Eis vor dem Detektor und die in dieser Wechselwirkung entstandenen Sekundärteilchen propagieren zum Detektor. Aufgrund der verhältnismäßig geringen Tiefe von IceCube ist dies ein weiterer

reduzierender Faktor der Neutrinorate. Die sich ergebende Anzahl von $\mathcal{O}(10)$ Ereignissen wird jedoch überschätzt, da in dem verwendeten Datensatz koinzidente Untergrundereignisse simuliert werden. Diese koinzidenten Untergrundereignisse sind Ereignisse der kosmischen Strahlung, welche die nötige Ladungsdeposition produzieren können und damit Ereignisse, die die Qualitätsschnitte passieren. Dadurch passieren Myonneutrinos mit einigen 10 GeV die Qualitätsschnitte, welche ein hohes Gewicht enthalten. Werden nur Myonen mit einer Energie im relevanten Energiebereich $E_\mu \geq 100$ TeV betrachtet, so reduziert sich die Anzahl an erwarteten Ereignissen auf 1,6 Ereignisse im Burn-Datensatz. Aufgrund dieser geringen Anzahl an Ereignissen bereits nach Qualitätsschnitten kann der Untergrund an erwarteten Myonneutrinos als vernachlässigbar betrachtet werden, da im finalen Datensatz in diesem Energiebereich $\mathcal{O}(1000)$ HE-Myon-Ereignisse selektiert werden.

6.6 Rekonstruktion des Myonspektrums

Mit dem finalen Datensatz wird in diesem Abschnitt das Spektrum hochenergetischer Myonen rekonstruiert. Dazu wird die Methode der Entfaltung genutzt, welche in Abschnitt 4.2 detailliert beschrieben wurde. Dieser Algorithmus ist in dem Softwarepaket TRUÉE[77] implementiert und bietet eine Reihe an Einstellungsmöglichkeiten zur Entfaltung. Zwei offensichtliche Einstellungsmöglichkeiten sind die Anzahl der zu entfaltenden Bins und die zur Entfaltung genutzten Attribute. Innerhalb von TRUÉE wird die zu entfaltende Verteilung mit Hilfe von Splines dargestellt. Die Anzahl dieser Splines ist festgelegt durch die Anzahl der Knoten. Die Anzahl der tatsächlich verwendeten Knoten ist jedoch um 6 höher, da für die Ränder jeweils 3 zusätzliche Knoten benötigt werden, um die Anzahl der Splines zu realisieren. Die Anzahl der Freiheitsgrade beschreibt, ab welchen Eigenwerten der Entfaltung eine Regularisierung stattfindet. Je höher die Anzahl der Freiheitsgrade, desto schwächer ist die Regularisierung. Die gewählten Einstellungen für diese Entfaltung sind im Folgenden aufgeführt:

- Entfaltete Bins: 10
- Freiheitsgrade: 9
- Knoten: 16
- Verwendete Attribute: $\log_{10}(dE/dx_{\text{SplineMPE BINS}})$, Tiefe

Bei den verwendeten Attributen wird ein Binning von 25 gewählt, welches in Abbildung 6.12 für beide Attribute dargestellt ist. In dieser Analyse werden nur zwei Attribute verwendet, da die Statistik an simulierten Ereignissen gering ist.

Würde der Parameterraum auf ein zusätzliches Attribut erweitert, so kommt es zu starken Fluktuation des Ergebnisses, da zu viele Bins leer oder nur schwach besetzt sind. In den Abbildungen 6.12a und 6.12b sind die verwendeten Attribute in Abhängigkeit der Oberflächenenergie der HE-Myonen aufgetragen. Mittelwerte und Varianzen für beide Attribute und jeden Wert dieser Attribute sind jeweils in diesen Abbildungen dargestellt. Während bei dem Energieschätzer eine gute Korrelation zur Oberflächenenergie des Myons vorliegt, ist kaum ein Zusammenhang bei der Tiefe erkennbar. Der Energieschätzer $dE/dx_{\text{SplineMPEBINS}}$ schätzt die Energie bzw. den Energieverlust des HE-Myon-Ereignisses im Detektor. Um Rückschlüsse auf die Oberflächenenergie zu ziehen, muss etwas über den Energieverlust des Ereignisses von der Oberfläche zum Detektor bekannt sein. Die Größe dieser Energieverluste hängt direkt von der Tiefe des Ereignisses ab, da während dieser Propagationslänge Myonen in stochastischen Prozessen Energieverluste erfahren (vgl. Abschnitt 2.5). Aufgrund dessen ist es sinnvoll die Tiefe in die Entfaltung einzubeziehen, da die Oberflächenenergie davon abhängt.

In Abbildung 6.12b ist zusätzlich die kritische Energie E_C dargestellt. Ihr Wert sagt aus, wie viel Energie ein Myon im Durchschnitt besitzen muss, damit es den Detektor erreicht. Die Berechnung dieser Energie beruht auf der Gleichung nach Barrett[100], die im Nachfolgenden dargestellt ist:

$$\frac{dE}{dx} = a + b \times E. \quad (6.6.1)$$

Die Gleichung (6.6.1) beschreibt den mittleren Energieverlust dE/dx eines Myons während der Propagation der Strecke x durch ein Medium. Alle stochastischen Energieverluste sind gemittelt angenommen und die Konstanten a und b sind mediumsabhängig. Durch Lösung dieser Gleichung lässt sich die kritische Energie ausdrücken als

$$E_C = \frac{a}{b} (\exp(b \cdot x \cdot \rho) - 1). \quad (6.6.2)$$

Die Konstanten a und b sind in der Einheit von Metern Wasseräquivalent angegeben, wodurch die Dichte ρ des Mediums berücksichtigt werden muss. Für den vorliegenden Fall von IceCube findet die Propagation der Myonen im Südpoleis statt, weswegen die folgenden Konstanten verwendet werden[35]:

$$\begin{aligned} a &= 0,259 \text{ mwe}^{-1} \\ b &= 0,364 \cdot 10^{-3} \text{ GeV mwe}^{-1} \\ \rho &= 0,917 \text{ g cm}^{-3} \end{aligned}$$

Einige Ereignisse in Abbildung 6.12b liegen unterhalb dieser kritischen Energie. Die kritische Energie ist jedoch nur der gemittelte Fall, weswegen ein Myon auch weniger

Energie durch stochastische Prozesse verlieren kann und somit eine längere Strecke propagieren kann. Zusammenfassend ist allerdings zu sagen, dass der Großteil der verbleibenden Ereignisse der Erwartung der kritischen Energie entsprechen und oberhalb dieser liegen.

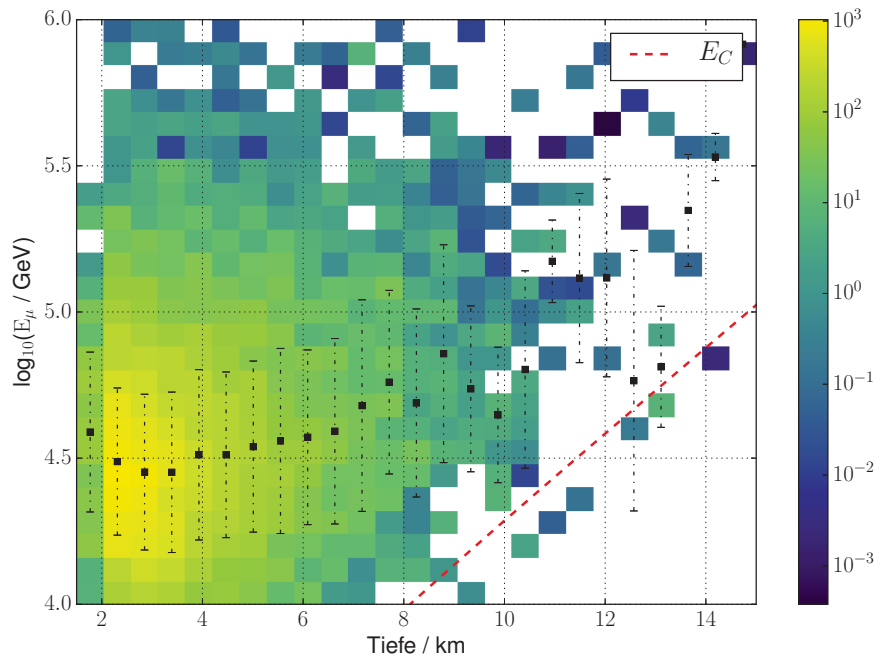
Das Ergebnis dieser Entfaltung mit den gewählten Einstellungen ist ein Energiespektrum, wie es in Abbildung 6.13 für den vollständigen Datensatz dargestellt ist. Die rekonstruierten Anzahlen des Datensatzes sind in der Tabelle 6.4 aufgeführt. Um mit der Anzahl der Ereignisse auf einen physikalischen Fluss schließen zu können, wird eine Entfaltung der Simulationsdaten durchgeführt. Anschließend wird anhand der Ereignisse und der Information über den simulierten Fluss in den jeweiligen Bins auf den physikalischen Fluss zurückgeschlossen. Somit wird das rekonstruierte Ereignisspektrum mit der Gleichung

$$\Phi_i = \frac{N_i}{E_{i+1} - E_i} \times \frac{\int_{E_i}^{E_{i+1}} \Phi_{\text{Theo.}}(E) dE}{N_{\text{MC},i}} \quad (6.6.3)$$

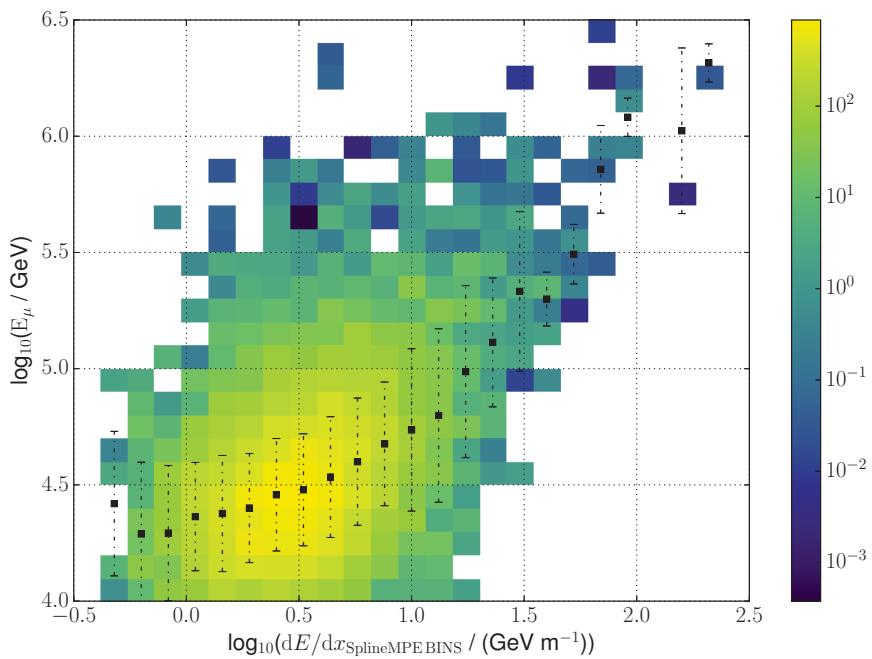
in einen differentiellen Fluss überführt. Als theoretischer konventioneller Fluss wird in dieser Arbeit der bestimmte Myonfluss in [101] verwendet, da dieser den simulierten Monte-Carlo-Ereignissen zu Grunde liegt. Da dieser Fluss mit SIBYLL-2.1 berechnet wurde, wird der konventionelle Fluss in dieser Arbeit als SIBYLL-2.1 bezeichnet. Der prompte Fluss wird nach der Berechnung von Enberg[98] bestimmt, jedoch findet eine Berücksichtigung des GaisserH3a-Modells nach [93] statt.

Das rekonstruierte und akzeptanzkorrigierte Spektrum für den Burn-Datensatz und die gewählten Einstellungen sind in Abbildung 6.14 zusammen mit dem erwarteten konventionellen und prompten Fluss dargestellt. Das Spektrum in Abbildung 6.14a ist mit E^3 multipliziert, respektiv mit $E^{3,7}$ in Abbildung 6.14b. In beiden Darstellungen des entfalteten Spektrums ist zu erkennen, dass das Spektrum mit der Vorhersage für den konventionellen Myonfluss übereinstimmt. Eine Aussage über den prompten Fluss an dieser Stelle zu treffen ist schwierig, da dieser bei den betrachteten Energien im letzten Bin lediglich ca. 30 % des Gesamtflusses beträgt.

Die Fehlerbalken des rekonstruierten Spektrums sind in zwei Komponenten aufgeteilt, eine statistische und eine durch Monte-Carlo-Daten hervorgerufene Komponente. Die statistische Komponente bezeichnet die Fluktuation der vorhandenen Daten aus dem Experiment. Diese Fluktuation ist gering, da die Rate an hochenergetischen Myonen hoch ist, wenn diese mit den in IceCube üblichen Neutrinoraten verglichen wird. Aus den vorhandenen Simulationsdaten ergeben sich jedoch höhere Unsicherheiten, da diese für das Jahr 2011 nur begrenzt zur Verfügung stehen. Um den Fehler aus den begrenzten Monte-Carlo-Daten abzuschätzen, werden aus dem vorhandenen Datensatz weitere Datensätze mit Ziehen und Zurücklegen erzeugt. Anschließend



(a)



(b)

Abbildung 6.12: Abhängigkeit zwischen den zur Entfaltung verwendeten Attributen und der gesuchten Myonenergie in GeV an der Oberfläche mit Mittelwerten und Varianzen. a) Korrelation der Myonenergie zwischen der Eindringtiefe des Ereignisses und die kritische Energie E_C . b) Energieverlust pro Meter für den SplineMPE-Rekonstruktionsalgorithmus in Abhängigkeit der Myonenergie.

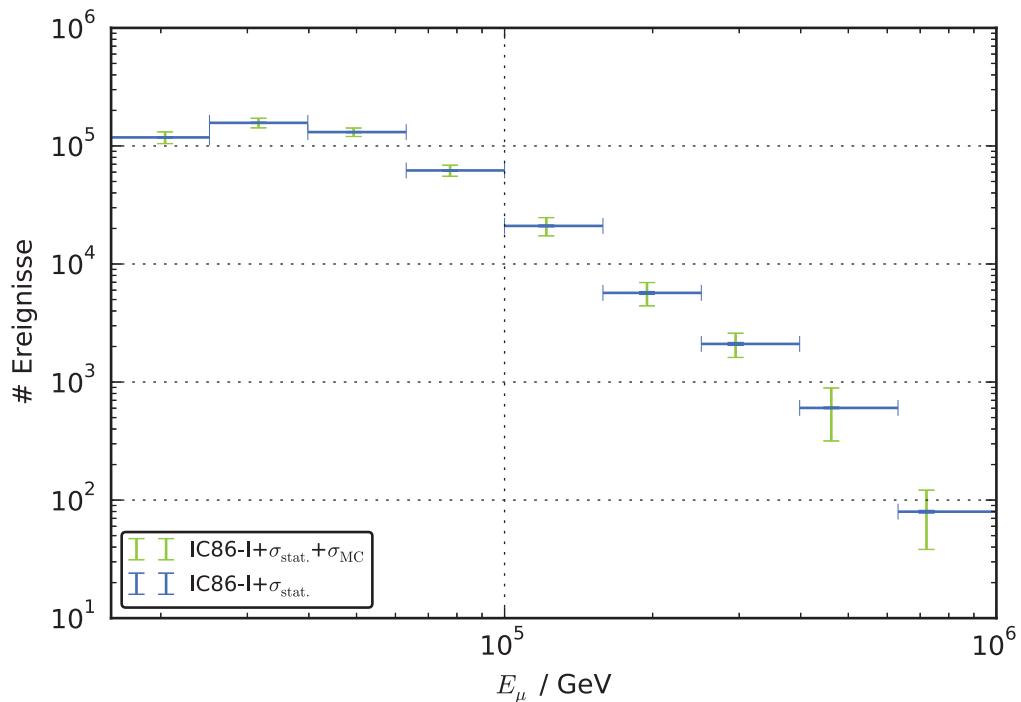


Abbildung 6.13: Dargestellt ist das entfaltete Ereignisspektrum hochenergetischer Myonen in IceCube für das Jahr 2011.

wird die Entfaltung mit jedem dieser Datensätzen durchgeführt und die Varianz der Datenpunkte berechnet. Anhand der Ergebnisse dieses Vorgehens ist zu erkennen, dass die begrenzte Anzahl an Simulationsdaten besonders für höhere Energien der dominierende Faktor ist, da dort nur wenige Ereignisse vorhanden sind.

Nachdem nun die Analyse vollständig durchgeführt wurde und gezeigt wurde, dass sie stabile und verlässliche Aussagen macht, kann die Analyse auf dem Datensatz des kompletten Jahres 2011 angewendet werden. Bei der Verwendung der Daten des gesamten Jahres darf aufgrund von Kollaborationsrichtlinien die Analyse nicht mehr verändert werden. Dadurch werden die Einstellungen, die in den vorherigen Abschnitten gewählt wurden, unverändert übernommen und auch die Simulationsdaten bleiben dieselben. Das rekonstruierte Spektrum hochenergetischer Myonen in IceCube für das Jahr 2011 findet sich in Abbildung 6.15. Aufgrund der selben verwendeten Monte-Carlo-Simulationen sind die Unsicherheiten aus diesen konstant geblieben, lediglich die Unsicherheit aus der statistischen Fluktuation der gemessenen Daten ist gesunken. Ein Vergleich der Abbildungen 6.14 und 6.15 zeigt, dass die rekonstruierten Spektren innerhalb ihrer Fehlerbalken konsistente Ergebnisse liefern. Die Datenpunkte und die Unsicherheiten des Spektrums multipliziert mit

Tabelle 6.4: Aufgeführt sind die entfalteten Ereignisanzahlen an hochenergetischer Myonen in IceCube für das Jahr 2011.

| Bins in $\log_{10}(E/\text{GeV})$ | # Ereignisse | σ_{stat} | σ_{tot} | $\sigma_{\text{rel}} / \%$ |
|-----------------------------------|--------------|------------------------|-----------------------|----------------------------|
| 4,0 - 4,2 | 49447, 1 | 454, 5 | 10253, 4 | 20, 7 |
| 4,2 - 4,4 | 118133, 0 | 1130, 3 | 13445, 5 | 11, 4 |
| 4,4 - 4,6 | 156935, 2 | 1621, 8 | 14617, 2 | 9, 3 |
| 4,6 - 4,8 | 131062, 4 | 1417, 0 | 10783, 1 | 8, 2 |
| 4,8 - 5,0 | 62001, 0 | 880, 8 | 6680, 2 | 10, 8 |
| 5,0 - 5,2 | 21021, 5 | 339, 8 | 3698, 1 | 17, 6 |
| 5,2 - 5,4 | 5691, 9 | 146, 3 | 1272, 0 | 22, 3 |
| 5,4 - 5,6 | 2103, 5 | 58, 6 | 488, 2 | 23, 2 |
| 5,6 - 5,8 | 603, 9 | 8, 5 | 287, 0 | 47, 5 |
| 5,8 - 6,0 | 79, 8 | 1, 9 | 41, 6 | 52, 1 |

$E^{3,7}$ sind in Tabelle 6.5 aufgeführt. Für Energien $E > 10^{5,6}$ GeV ist eine Überfluktuation oberhalb des konventionellen Flusses zu beobachten. Diese Fluktuation ist nicht signifikant und kann durch den begrenzten Simulationsdatensatz verursacht werden.

Im Folgenden wird das rekonstruierte Spektrum untersucht und verschiedene Hypothesen für das rekonstruierte Spektrum werden angenommen. Die offensichtlichste Hypothese ist, dass das Spektrum aus einer konventionellen und einer prompten Komponente mit bekannter Form besteht. Dadurch kann eine Normalisierung für diese beiden Komponenten bestimmt werden, um deren Beiträge zum Gesamtfluss zu ermitteln. Der angenommene Fluss lässt sich somit ausdrücken als:

$$\Phi(E) = N_{\text{Konv.}} \cdot \Phi_{\text{Konv.}}(E) + N_{\text{Prompt}} \cdot \Phi_{\text{Prompt}}(E) \quad (6.6.4)$$

Weiterhin kann auch eine Nullhypothese angenommen werden, dass keine prompte Komponente vorliegt und nur die konventionelle Komponente vorhanden ist. Da die prompte Komponente der atmosphärischen Myonen noch nicht gemessen wurde, ist dies eine sinnvolle Hypothese, die es zu testen gilt.

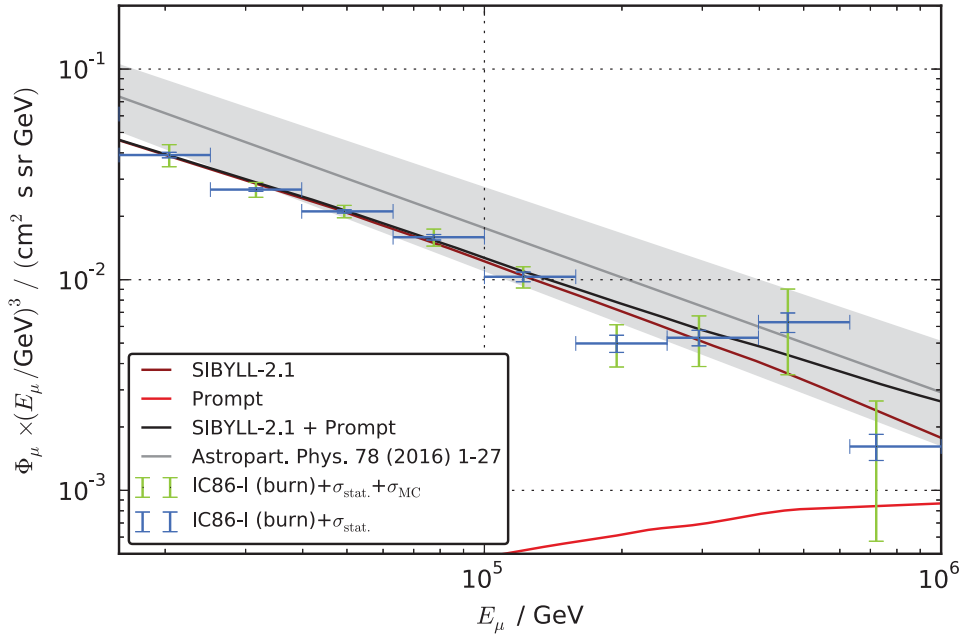
$$\Phi(E) = N_{\text{Konv.}}^* \cdot \Phi_{\text{Konv.}}(E) \quad (6.6.5)$$

Abschließend wird noch die naive Hypothese getestet, dass nur eine konventionelle Komponente vorliegt und diese einem Potenzgesetz folgt. Die Annahme eines Potenzgesetzes für die konventionelle Komponente ist gerechtfertigt, da dies in erster Näherung der Fall ist.

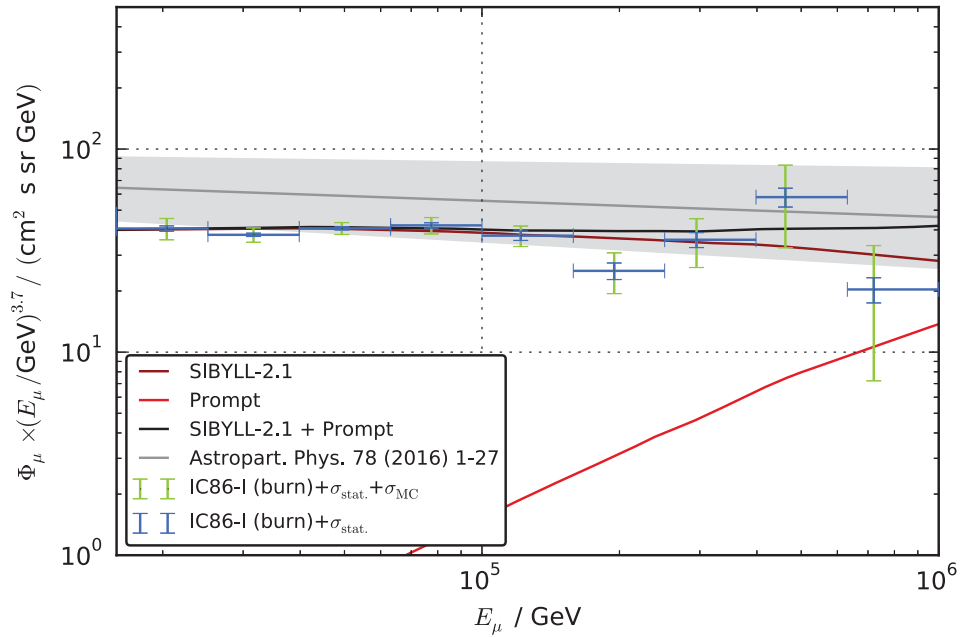
$$\Phi(E) = N \cdot E^\gamma \quad (6.6.6)$$

6 Rekonstruktion des Myonspektrums

Die zu testenden Hypothesen mit ihren Ergebnissen sind in der Tabelle 6.6 aufgeführt und eine grafische Darstellung dieser Hypothesen findet sich in Abbildung 6.17. Anhand der berechneten Werte für χ^2/ndf , welche in Klammern angegeben sind, lässt sich schließen, dass eine hohe Übereinstimmung aller Hypothesen mit dem rekonstruierten Spektrum vorliegt.

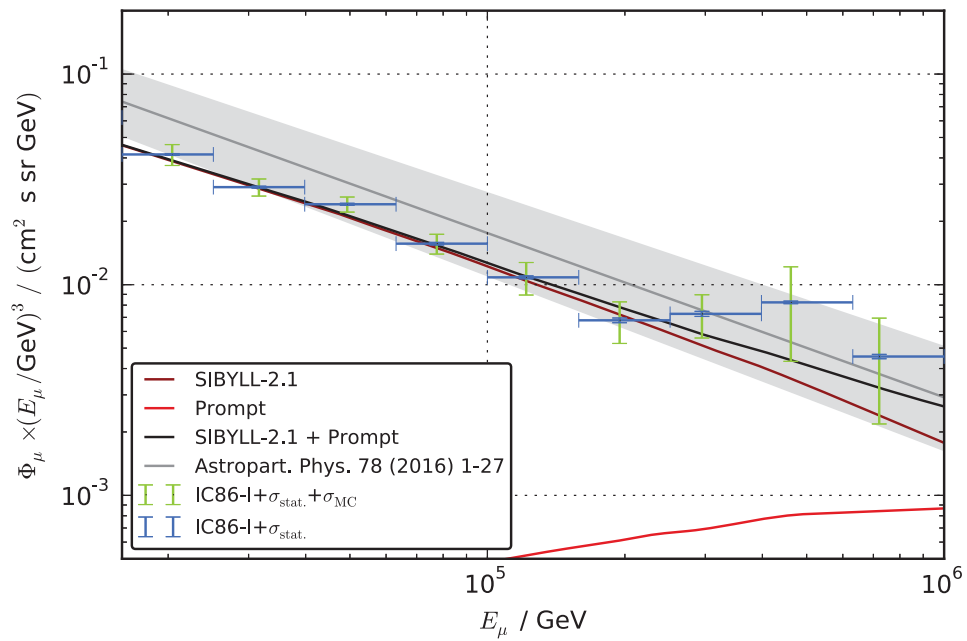


(a)

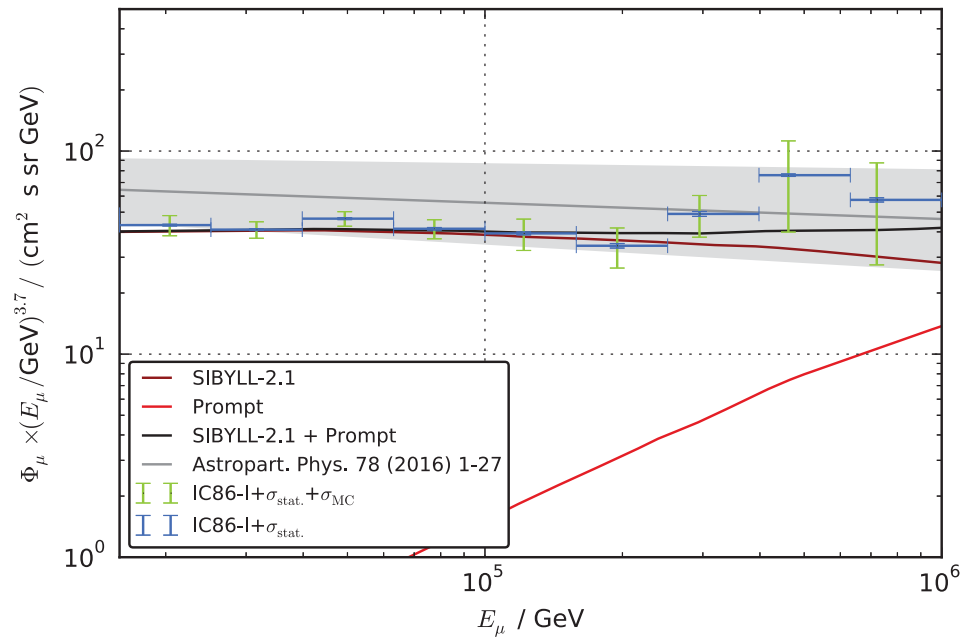


(b)

Abbildung 6.14: Dargestellt sind die entfaltenen HE-Myon-Flüsse multipliziert mit $(E_\mu/\text{GeV})^3$ (a) und $(E_\mu/\text{GeV})^{3,7}$ (b) für den Burn-Datensatz. Der zu erwartende prompte und konventionelle Fluss und deren Summe sind zusätzlich zu den entfaltenen Daten eingezeichnet.



(a)



(b)

Abbildung 6.15: Dargestellt sind die entfalteten HE-Myon-Flüsse multipliziert mit $(E_\mu/\text{GeV})^3$ (a) und $(E_\mu/\text{GeV})^{3,7}$ (b) für das Jahr 2011. Der zu erwartende prompte und konventionelle Fluss und deren Summe sind zusätzlich zu den entfalteten Daten eingezeichnet.

Tabelle 6.5: Aufgeföhrt ist der entfaltetete Fluss hoehenergetischer Myonen in IceCube für das Jahr 2011 multipliziert mit $(E_\mu/\text{GeV})^{3,7}$.

| Bins in $\log_{10}(E/\text{GeV})$ | Fluss $\times (E_\mu/\text{GeV})^{3,7} / (\text{cm}^2 \text{ s sr GeV})$ | σ_{stat} | σ_{tot} | $\sigma_{\text{rel}} / \%$ |
|-----------------------------------|--|------------------------|-----------------------|----------------------------|
| 4,0 - 4,2 | 48,6 | 0,446 | 10,1 | 20,7 |
| 4,2 - 4,4 | 43,2 | 0,413 | 4,92 | 11,4 |
| 4,4 - 4,6 | 41 | 0,424 | 3,82 | 9,31 |
| 4,6 - 4,8 | 46,5 | 0,502 | 3,82 | 8,23 |
| 4,8 - 5,0 | 41,4 | 0,589 | 4,46 | 10,8 |
| 5,0 - 5,2 | 39,3 | 0,636 | 6,92 | 17,6 |
| 5,2 - 5,4 | 34,2 | 0,878 | 7,64 | 22,3 |
| 5,4 - 5,6 | 49 | 1,37 | 11,4 | 23,2 |
| 5,6 - 5,8 | 76,1 | 1,08 | 36,1 | 47,5 |
| 5,8 - 6,0 | 57,5 | 1,37 | 30 | 52,1 |

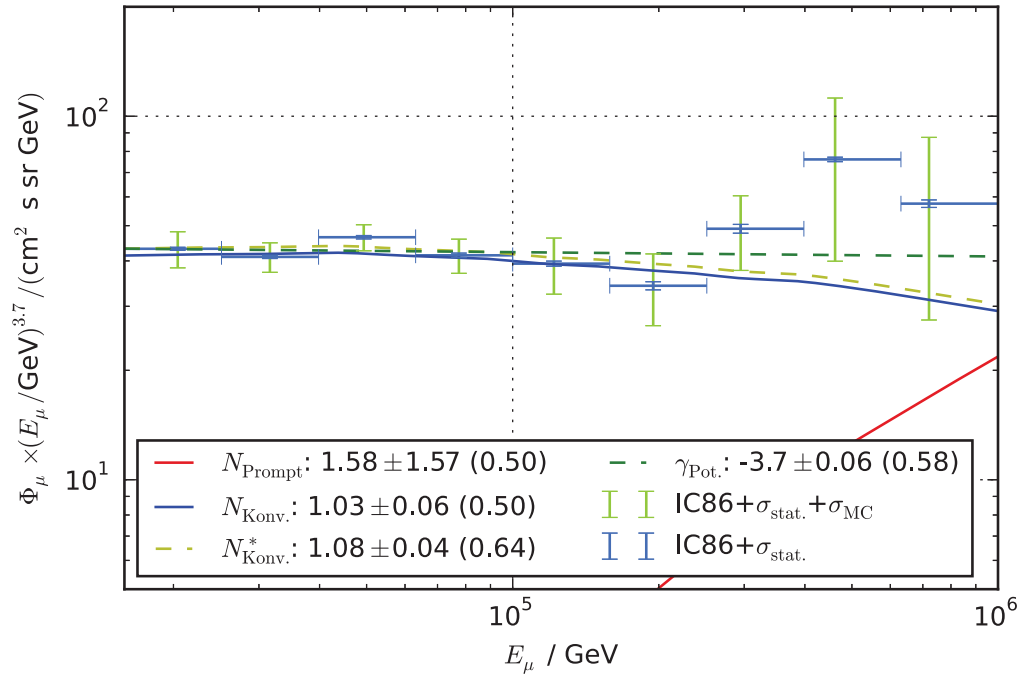


Abbildung 6.16: Rekonstruiertes Myonspektrum mit den zu testenden Hypothesen. Für jede Hypothese ist zusätzlich der Wert von χ^2/ndf angegeben.

Tabelle 6.6: Aufgeführt sind die Fitresultate für die Hypothesen des gemessenen Spektrum der hochenergetischen Myonen.

| Fitbezeichnung | Fitwert | Unsicherheit |
|------------------------|---------|--------------|
| $N_{\text{Konv.}}$ | 1,03 | 0,06 |
| N_{Prompt} | 1,59 | 1,57 |
| $N_{\text{Konv.}}^*$ | 1,08 | 0,05 |
| $\gamma_{\text{Pot.}}$ | -3,71 | 0,06 |

6.7 Resultate im Kontext der Neutrinoastronomie

In den Abschnitten 2.3 und 2.4 wurde die sekundäre kosmische Strahlung, welche in der Atmosphäre entsteht, beschrieben. Dabei konzentrierte sich die Diskussion auf die entstehenden Myonen, da diese der Fokus der vorliegenden Arbeit sind. Allerdings können bei der Erzeugung von hochenergetischen Myonen auch Neutrinos entstehen, welche die Erde erreichen und in IceCube detektiert werden können. Die entstehenden Neutrinos bestehen, wie die Myonen, aus einer konventionellen und einer prompten Komponente. Zusätzlich zu diesen beiden Komponenten aus der sekundären kosmischen Strahlung, tritt dort auch eine astrophysikalische Komponente auf. Diese astrophysikalische Komponente beginnt im selben Energiebereich ($E > 100 \text{ TeV}$) dominant zu werden wie die prompte, allerdings ist der astrophysikalische Fluss um einen Faktor ≈ 10 höher. Dadurch ist die Vermessung der prompten Komponente mit Neutrinos aufgrund der ähnlichen spektralen Indizes und dem hohen Untergrund der astrophysikalischen Neutrinos problematisch.

In diesem Abschnitt werden die ermittelten Normalisierungen für die Flüsse aus der vorliegenden Analyse im Kontext der Neutrinomessungen gezeigt. Die ermittelten Normalisierungen aus der Hypothese nach Gleichung 6.6.4 und deren Unsicherheit sind in Abbildung 6.17 dargestellt. Zusätzlich zu den Vorhersagen durch diese Analyse ist auch der von IceCube gemessene Neutrinofluss nach [74] aufgetragen. Anhand des Fehlerbandes des Fits und der Fehlerbalken der gemessenen Neutrinovertelung ist zu erkennen, dass der gemessene Fluss hochenergetischer Neutrinos nicht durch eine rein atmosphärische Komponente erklärt werden kann. Um diese Tatsache zu quantifizieren wird ein χ^2 -Test durchgeführt, jedoch ist es nicht sinnvoll dies für alle bestimmten Messwerte zu durchzuführen. Der p-Wert aus dem χ^2 -Test für die gemessenen Datenpunkte und die ermittelten Normalisierungen der prompten und konventionellen Komponente in den letzten vier Bins beträgt lediglich 0,045. Somit ist es unwahrscheinlich, dass es sich um eine statistische Fluktuation der relevanten Messpunkte handelt.

Das Ergebnis aus den Fitwerten der Myonanalyse ist jedoch nur begrenzt auf die Neutrinoflüsse anwendbar. Der Grund für diese Einschränkung liegt darin, dass eine Myonanalyse die Summe der prompten Komponente, bestehend aus dem Zerfall charmhaltiger und flavorfreier Elternteilchen, misst (vgl. Abschnitt 2.4). Myonneutrinos werden allerdings nur durch die charmhaltige Komponente erzeugt. Da diese beiden Komponenten für Myonen in dem analysierten Energiebereich annähernd gleiche Beiträge liefern, ist die Aussage für Myonneutrinos nur gültig für die Annahme, dass die zugrunde liegenden Myonflussmodelle korrekt sind.

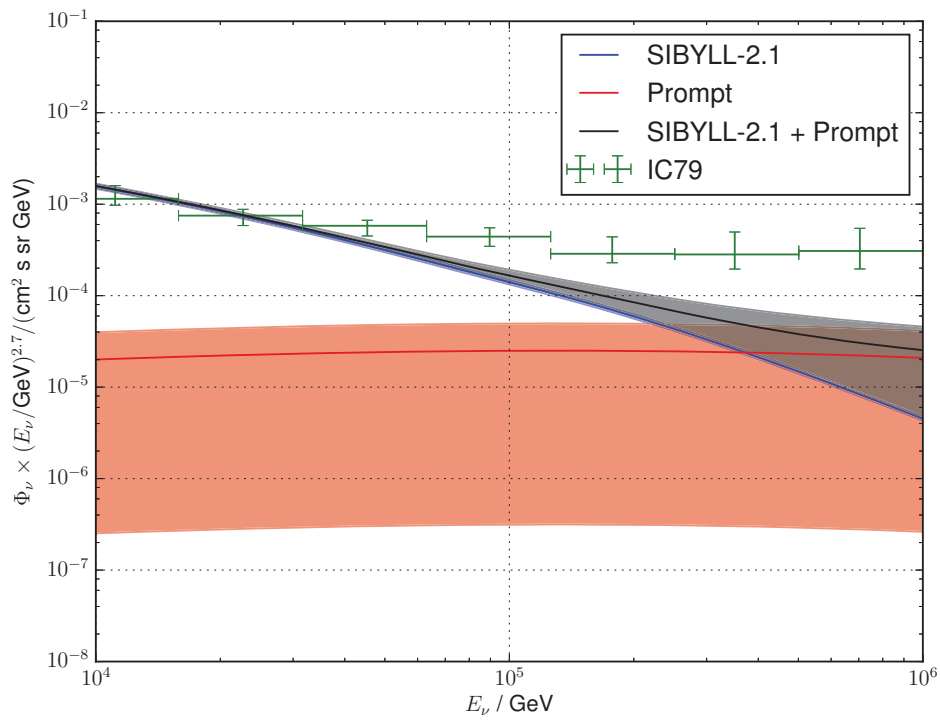


Abbildung 6.17: Dargestellt sind die ermittelten Flussnormierungen für Myonneutrinos und der gemessene Fluss der Myonneutrinos multipliziert mit $(E_\mu/\text{GeV})^{2.7}$ und der gemessene Neutrinofluss nach [74].

7 Zusammenfassung und Ausblick

Das Ziel dieser Arbeit war die Messung des hochenergetischen atmosphärischen Myon-Energiespektrums mit IceCube in der Detektorkonfiguration IC86-I, um daraus resultierende Implikationen bezüglich der konventionellen und prompten Komponente atmosphärischer Myonen und astrophysikalischer Neutrinos zu treffen. Dazu wurden Myon-Bündel-Ereignisse in verschiedene Klassen eingeteilt. Eine dieser für diese Arbeit relevante Klasse umfasst Myon-Bündel aus mehreren Myonen, bei denen ein Myon die anderen energetisch dominiert. Diese Ereignisse wurden mit fortgeschrittenen Algorithmen des Data-Minings von den verbleibenden Bündel-Ereignissen getrennt. Für die Selektion von hochenergetischen Myonen wurde eine Attributauswahl durchgeführt, welche in mehrere Schritte unterteilt ist. Zunächst wurden simulierte Ereignisse mit den detektierten Ereignissen verglichen und Attribute, welche eine große Abweichung aufwiesen, verworfen. Attribute mit einer hohen Korrelation zueinander wurden ebenfalls entfernt. Um eine Favorisierung von niederenergetischen Ereignissen zu vermeiden, wurden energiekorrelierte Attribute entfernt. Aus diesen vorselektierten Attributen wurden 30 Attribute von dem mRMR-Algorithmus für die nachfolgende Separation ausgewählt.

Die Selektion von Ereignissen mit hochenergetischen Myonen aus den Myon-Bündel-Ereignissen wurde anhand des maschinellen Lernalgorithmus Random Forest durchgeführt. Das Random-Forest-Modell wurde zu diesem Zweck mit einem Trainingsdatensatz erstellt, welcher aus 50 % der vorhandenen simulierten Daten besteht. Das verwendete Random-Forest-Modell besteht aus 200 Bäumen, da dies ausreichend genaue Ergebnisse liefert bei moderater Rechenzeit. Die Anwendung des Modells auf den Testdatensatz erreicht eine Signaleffizienz von $(40,8 \pm 0,6) \%$ und eine Reinheit von $(93,1 \pm 0,4) \%$ bei einem Konfidenzschnitt von 0,9.

Nachdem die Separation durchgeführt wurde, wurden die Eigenschaften des daraus entstandenen finalen Datensatzes beschrieben. Es zeigte sich, dass sich durch die Separation das Verhältnis von Signal- zu Untergrundereignissen von ungefähr 1 : 10 vor der Separation zu ungefähr 10 : 1 nach der Separation verbesserte. Die totale Selektionseffizienz von Signalereignissen durch die gesamten angewendeten Analyseschritte beträgt ca. 10 %, während Untergrundereignisse unter 0,1 % liegen. Die Massenkomposition des finalen Datensatzes weist überwiegend leichte Elemente auf, da diese einen höheren Anteil der Energie auf einzelne Nukleonen verteilen. Abschließend wurde noch der verbleibende Neutrino-Untergrund für diese Analyse

untersucht mit dem Resultat, dass die Anzahl der Ereignisse im relevanten Energiebereich weniger als 1 % der Gesamttereignisse beträgt.

Anhand des final entstandenen Datensatzes wurde ein atmosphärisches Myon-Energiespektrum mit der Software TRUUE entfaltet. Die dominierende Unsicherheit des Spektrums wurde durch die begrenzten Simulationsdatensätze verursacht. Mit Hilfe dieser Ergebnisse konnten mehrere Hypothesen bezüglich atmosphärischer Myonen überprüft und Normalisierung für den konventionellen und prompten Fluss bestimmt werden. Die Normierungen zeigten eine Übereinstimmung mit den Vorhersagen für die konventionelle Komponente, jedoch ist die Normierung der prompten Komponente nahezu mit Null verträglich. Allerdings liefert die Hypothese eines rein konventionellen Flusses ähnliche Ergebnisse wie die Annahme einer konventionellen und prompten Komponente.

Abschließend wurden die berechneten Normalisierung des Myonflusses im Hinblick auf die Neutrinoastronomie betrachtet. Zu diesem Zweck wurden die bestimmten Normalisierungen auf die Myonneutrinos übertragen. Der konventionelle Fluss für niederenergetische Myonneutrinos wird durch diese Analyse bestätigt, jedoch kann der Fluss der hochenergetischen Myonneutrinos nicht durch die bestimmte prompte Komponente beschrieben werden. Dies ist ein weiterer Hinweis darauf, dass die gemessenen hochenergetischen Neutrinos aus Quellen hochenergetischer kosmischer Strahlung stammen und nicht in der Atmosphäre produziert werden.

Die erzielten Ergebnisse dieser Arbeit sind vielversprechend und motivieren weitere darauf aufbauende Analysen. Verbesserungen bezüglich der vorhandenen Daten, der Analysemethoden und der möglichen Detektorerweiterungen können in zukünftigen Analysen berücksichtigt werden.

Zwischen der Datennahme-Saison IC86-I und IC86-II wurde eine Verbesserung der Datennahme-Filter durchgeführt. Die Datennahme-Filter wurden in der Saison IC86-II bis IC86-V nicht verändert, wodurch der gesamte Zeitraum in einer zusammenhängenden Analyse betrachtet werden kann. Der weitaus wichtigere Vorteil für diese Analyse besteht jedoch darin, dass siebenmal mehr simulierte Ereignisse zur Verfügung stehen. Mit der zusätzlichen Menge an simulierten Daten kann sowohl die Unsicherheit im relevanten Bereich oberhalb von 100 TeV verkleinert, als auch der obere Energiebereich erweitert werden. Dadurch wären Aussagen mit geringeren Unsicherheiten über den prompten Fluss möglich. Weiterhin liegen für den Datennahme-Zeitraum systematische Datensätze bereit, welche bei der Abschätzung systematischer Unsicherheiten genutzt werden können.

Weitere Verbesserungen der Ergebnisse können erzielt werden, indem die Analyse verändert wird. Diese Veränderung könnte schon bei der Klasseneinteilung der Myon-Bündel-Ereignisse beginnen. Die bisherige Forderung, dass ein Myon 50 % der Gesamtenergie des Myon-Bündel-Ereignisses besitzen muss, um als Signal zu gelten, wurde gewählt damit die Energie dieses hochenergetischen Myons rekon-

struiert werden kann. Alternativ wäre es auch möglich, den relativen Abstand des höchstenergetischen Myons zum Myon mit nächst niedrigerer Energie als Kriterium zur Klasseneinteilung zu wählen. Dies könnte noch erweitert werden und die Myonen mit der zweithöchsten Energie als Signal betrachtet werden, damit Zerfälle in zwei Myonen aus dem prompten Fluss betrachtet werden können. Eine Alternative zur Klasseneinteilung wäre der Verzicht auf eine Einteilung. Eine gewählte Klasseneinteilung hat als Zweck nur Ereignisse zu betrachten, welche wahrscheinlich durch ein Myon dominiert werden, sodass die rekonstruierte Energie des Ereignisses mit der des höchstenergetischen Myons korreliert. Stattdessen könnte die Energie des hochenergetischen Myons mit Hilfe eines Random-Forest-Regressors oder ähnlichen Algorithmen für jedes Ereignis geschätzt werden. Ereignisse, welche eine geringe Varianz der Schätzung aufweisen, können anschließend zur Entfaltung verwendet werden.

Zusätzlich kann an der Analyse noch das Verhältnis des Trainings- zu Testdatensatzes verändert werden. In der vorliegenden Analyse wurde aufgrund der begrenzten Simulationsdaten ein Verhältnis von 1 : 1 verwendet. Dadurch standen sowohl genügend Ereignisse zur Erstellung eines Random-Forest-Modells zur Verfügung, also auch zur Entfaltung des Myonspektrums. Aufgrund der Verfügbarkeit von mehr Monte-Carlo-Ereignissen in den Folgejahren kann dieses Trainingsverhältnis angepasst werden. Falls das Random-Forest-Modell in der gleichen Qualität trainiert werden würde, dann könnte eine dreizehnfache Anzahl an Ereignissen zur Entfaltung genutzt werden. Als direkte Konsequenz würden die Unsicherheiten der Entfaltung drastisch sinken und genauere Aussagen über das Spektrum und die prompte Komponente könnten getroffen werden.

Weitere Möglichkeiten der Verbesserung betreffen geplante Detektorerweiterungen. Eine dieser Erweiterungen von IceCube wird als Gen2 bezeichnet und ist in mehrere Teile unterteilt. Die kostengünstigste Erweiterungsmöglichkeit besteht darin, das Oberflächenarray IceTop zu erweitern und so ungefähr 100 km^2 an Oberfläche abzudecken. Bei identischem Detektor innerhalb des Eises wäre es möglich zusätzlich die Masse des Primärteilchens zu bestimmen, welches ein hochenergetisches Myon erzeugt hat. Da es wahrscheinlicher ist, dass hochenergetische Myonen von leichten Elementen erzeugt werden, kann dies zur Verbesserung der Separation verwendet werden. Eine andere Erweiterungsmöglichkeit besteht darin, den Detektor innerhalb des Eises auf ein instrumentiertes Volumen von 10 km^3 zu vergrößern. Um dieses Volumen zu instrumentieren, muss allerdings der Abstand zwischen den einzelnen Strings erhöht werden. Durch die Erhöhung der Stringabstände wird die Detektion von stochastischen Verlusten innerhalb eines Myon-Bündels erschwert. Die Erweiterung kann jedoch trotz alledem genutzt werden, um den Qualitätsschnitt von mindestens 600 m innerhalb des Detektors auf geringere Werte zu reduzieren. Dazu könnte weiterhin die hohe Auflösung des IceCube-Detektors genutzt werden, um stochastische Verluste zu detektieren und zu rekonstruieren.

Literatur

- [1] V. F. Hess. „Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten“. In: *Physikalische Zeitschrift* 13 (1912).
- [2] A. Einstein. „Über Gravitationswellen“. In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (1918).
- [3] B. P. Abbott et al. „Observation of Gravitational Waves from a Binary Black Hole Merger“. In: *Phys. Rev. Lett.* 116 (2016).
- [4] R. Blandford und D. Eichler. „Particle acceleration at astrophysical shocks: A theory of cosmic ray origin“. In: *Physics Reports* 154.1 (1987).
- [5] E. Berezhko und H. Völk. „Spectrum of cosmic rays produced in supernova remnants“. In: *Astrophysical Journal* 661.2 II (2007).
- [6] J. P. Rachen, T. Stanev und P. L. Biermann. „Extragalactic ultra-high energy cosmic rays. II. Comparison with experimental data“. In: *Astronomy and Astrophysics* 273 (1993).
- [7] E. Fermi. „On the Origin of the Cosmic Radiation“. In: *Phys. Rev.* 75 (1949).
- [8] F. L. Franca et al. „A New Cosmological Distance Measure Using Active Galactic Nucleus X-Ray Variability“. In: *The Astrophysical Journal Letters* 787.1 (2014).
- [9] The IceCube Collaboration. „Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector“. In: *Science* 342.6161 (2013).
- [10] W. B. Atwood et al. „The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission“. In: *The Astrophysical Journal* 697.2 (2009).
- [11] D. Mazin et al. „Potential of EBL and cosmology studies with the Cherenkov Telescope Array“. In: *Astroparticle Physics* 43 (2013).
- [12] A. Achterberg et al. „First year performance of the IceCube neutrino telescope“. In: *Astroparticle Physics* 26.3 (2006).
- [13] R. Davis, D. S. Harmer und K. C. Hoffman. „Search for Neutrinos from the Sun“. In: *Phys. Rev. Lett.* 20 (1968).
- [14] V. Gribov und B. Pontecorvo. „Neutrino astronomy and lepton charge“. In: *Phys. Lett. B* 28.7 (1969).
- [15] K. Daum et al. „Determination of the atmospheric neutrino spectra with the Fréjus detector“. In: *Zeitschrift für Physik C Particles and Fields* 66 (1995).

- [16] E. Richard et al. „Measurements of the atmospheric neutrino flux by Super-Kamiokande: energy spectra, geomagnetic effects, and solar modulation“. In: *arXiv:1510.08127* (2015).
- [17] M. G. Aartsen et al. „Measurement of the Atmospheric ν_e Spectrum with IceCube“. In: *Phys. Rev. D* 91 (2015).
- [18] M. G. Aartsen et al. „Atmospheric and astrophysical neutrinos above 1 TeV interacting in IceCube“. In: *Phys. Rev. D* 91 (2015).
- [19] C. J. Cesarsky. „Cosmic-Ray Confinement in the Galaxy“. In: *Annual Review of Astronomy and Astrophysics* 18.1 (1980).
- [20] W. Apel et al. „The spectrum of high-energy cosmic rays measured with KASCADE-Grande“. In: *Astroparticle Physics* 36.1 (2012).
- [21] W. Apel et al. „KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays“. In: *Astroparticle Physics* 47 (2013).
- [22] J. Abraham et al. „Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory“. In: *Phys. Lett. B* 685.4–5 (2010).
- [23] The Pierre Auger collaboration. „Measurement of the cosmic ray spectrum above 4×10^{18} eV using inclined events detected with the Pierre Auger Observatory“. In: *Journal of Cosmology and Astroparticle Physics* 2015.08 (2015).
- [24] J. R. Hörandel. „On the knee in the energy spectrum of cosmic rays“. In: *Astroparticle Physics* 19.2 (2003).
- [25] K. A. Olive et al. „Review of Particle Physics“. In: *Chin. Phys.* C38 (2014).
- [26] K. Greisen. „End to the Cosmic-Ray Spectrum?“ In: *Phys. Rev. Lett.* 16 (1966).
- [27] G. T. Zatsepin und V. A. Kuz'min. „Upper Limit of the Spectrum of Cosmic Rays“. In: *Soviet Journal of Experimental and Theoretical Physics Letters* 4 (1966).
- [28] T. Stanev, P. L. Biermann und T. K. Gaisser. „Cosmic rays. IV. The spectrum and chemical composition above 10 GeV“. In: *Astroparticle Physics* 274 (1993).
- [29] T. K. Gaisser, T. Stanev und S. Tilav. „Cosmic ray energy spectrum from measurements of air showers“. In: *Frontiers of Physics* 8.6 (2013).
- [30] M. Thunman, G. Ingelman und P. Gondolo. „Charm production and high energy atmospheric muon and neutrino fluxes“. In: *Astroparticle Physics* 5.3–4 (1996).
- [31] T. K. Gaisser. „Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio“. In: *Astroparticle Physics* 35.12 (2012).
- [32] H. S. Ahn et al. „Measurements of the Relative Abundances of High-energy Cosmic-ray Nuclei in the TeV/Nucleon Region“. In: *The Astrophysical Journal* 715.2 (2010).

-
- [33] A. Fedynitch et al. „A new version of the event generator Sibyll“. In: *PoS(ICRC2015)* 558 (2015).
- [34] D. Soldin. „Composition from high pT muons in IceCube“. In: *EPJ Web of Conferences* 99 (2015).
- [35] J.-H. Koehne et al. „PROPOSAL: A tool for propagation of charged leptons“. In: *Computer Physics Communications* 184.9 (2013).
- [36] V. Butkevich und S. P. Mikheyev. „Cross section of the muon-nuclear inelastic interaction“. In: *Soviet Journal of Experimental and Theoretical Physics* 95 (2002).
- [37] H. Abramowicz und A. Levy. „The ALLM parameterization of $\sigma_{tot}(\gamma^*p)$ - an update“. In: *arXiv:hep-ph/9712415* (1997).
- [38] H. Abramowicz et al. „A parametrization of $\sigma_T(\gamma^*p)$ above the resonance region for $Q^2 \geq 0$ “. In: *Phys. Lett. B* 269 (1991).
- [39] W. Rhode. „Measurements of the muon-flux with the Fréjus-detector“. In: *Nuclear Physics B Proceedings Supplements* 35 (1994).
- [40] E. V. Bugaev und Y. V. Shlepin. „Photonuclear interaction of high energy muons and tau leptons“. In: *Nuclear Physics B Proceedings Supplements* 122 (2003).
- [41] L. B. Bezrukov und E. V. Bugaev. „Nucleon shadowing effects in photonuclear interactions“. In: *Soviet Journal of Nuclear Physics* 33 (1981).
- [42] M. Derrick et al. „Measurement of total and partial photon proton cross-sections at 180-GeV center-of-mass energy“. In: *Z.Phys.* C63 (1994).
- [43] R. Abbasi et al. „IceTop: The surface component of IceCube“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 700 (2013).
- [44] F. Halzen und S. R. Klein. „IceCube: An instrument for neutrino astronomy“. In: *Rev. Sci. Instrum.* 81.8 (2010).
- [45] P. A. Cherenkov. „Visible emission of clean liquids by action of gamma radiation“. In: *Doklady Akademii Nauk SSSR* 2 (1934).
- [46] M. Aartsen et al. „Measurement of South Pole ice transparency with the IceCube LED calibration system“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 711 (2013).
- [47] D. J. Koskinen. „IceCube-DeepCore-Pingu: Fundamental neutrino and dark matter physics at the south pole“. In: *Modern Physics Letters A* 26.39 (2011).
- [48] M. G. Aartsen et al. „Search for Dark Matter Annihilations in the Sun with the 79-String IceCube Detector“. In: *Phys. Rev. Lett.* 110 (2013).
- [49] M. G. Aartsen et al. „IceCube search for dark matter annihilation in nearby galaxies and galaxy clusters“. In: *Phys. Rev. D* 88 (2013).

- [50] M. Aartsen et al. „Multipole analysis of IceCube data to search for dark matter accumulated in the Galactic halo“. In: *The European Physical Journal C* 75.1, 20 (2015).
- [51] S. P. Ahlen. „Monopole-track characteristics in plastic detectors“. In: *Phys. Rev. D* 14 (1976).
- [52] D. R. Tompkins. „Total Energy Loss and Čerenkov Emission from Monopoles“. In: *Phys. Rev.* 138 (1965).
- [53] M. G. Aartsen et al. „Measurement of the cosmic ray energy spectrum with IceTop-73“. In: *Phys. Rev. D* 88 (2013).
- [54] M. S. Sutherland. „Search for High Energy Neutron Point Sources in IceTop“. In: *PoS(ICRC2015)* 250 (2015).
- [55] M. Aartsen et al. „Development of a general analysis and unfolding scheme and its application to measure the energy spectrum of atmospheric neutrinos with IceCube“. In: *The European Physical Journal C* 75.3, 116 (2015).
- [56] M. A. Efronymson. *Mathematical methods for digital computers*. New York: Wiley, 1960.
- [57] S. Weisberg. *Applied Linear Regression 2nd ed.* Wiley, 1985.
- [58] N. Jaspens, N. Draper und H. Smith. „Applied Regression Analysis“. In: *Mathematics of Computation* 22 (1968).
- [59] C. Ding und H. Peng. „Minimum Redundancy Feature Selection from Microarray Gene Expression Data“. In: *J Bioinform Comput Biol.* 2003.
- [60] C. Ding, H. Peng und F. Long. „Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (2005).
- [61] E. Parzen. „On Estimation of a Probability Density Function and Mode“. In: *The Annals of Mathematical Statistics* 33.3 (1962).
- [62] B. W. Silverman. „Choosing the Window Width when Estimating a Density“. In: *Biometrika* 65.1 (1978).
- [63] B. Schöwe und K. Morik. „Fast-Ensembles of Minimum Redundancy Feature Selection“. In: *Ensembles in Machine Learning Applications*. Bd. 373. Studies in Computational Intelligence. Springer Berlin Heidelberg, 2011.
- [64] P. Jaccard. „The Distribution of the Flora in the Alpine Zone“. In: *New Phytologist* 11.2 (1912).
- [65] L. I. Kuncheva. „A stability index for feature selection“. In: *Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications* (2007).
- [66] L. Breiman et al. *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 1984.
- [67] J. Mingers. „An empirical comparison of selection measures for decision-tree induction“. In: *Machine Learning* 3.4 (1989).
- [68] J. R. Quinlan. „Induction of Decision Trees“. In: *Machine Learning* 1.1 (1986).

-
- [69] W. Buntine und T. Niblett. „A Further Comparison of Splitting Rules for Decision-Tree Induction“. In: *Machine Learning* 8.1 (1992).
- [70] J. H. Friedman. „Greedy function approximation: A gradient boosting machine.“ In: *Ann. Statist.* 29.5 (2001).
- [71] L. Breiman. „Random Forests“. In: *Machine Learning* 45.1 (2001).
- [72] O. Ritthoff et al. *Yale: Yet Another Learning Environment*. 2001.
- [73] T. Sharp. „Computer Vision – ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV“. In: Springer Berlin Heidelberg, 2008.
- [74] F. Scheriau. „Data-Mining in der Astroteilchenphysik. Entfaltung des Energiespektrums der von IceCube in der 86 String-Konfiguration gemessenen Myon-Neutrinos“. Dissertation. TU Dortmund, 2014.
- [75] R. Cruz. *Brackets around tikz picture*. 2016. URL: <http://tex.stackexchange.com/questions/268321/brackets-around-tikz-picture>.
- [76] I. Fredholm. „Sur une classe d'équations fonctionnelles“. In: *Acta Mathematica* 27.1 (1903).
- [77] N. Milke et al. „Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 697 (2013).
- [78] A. W. F. Edwards. *Likelihood*. Johns Hopkins University Press, 1992.
- [79] A. N. Tikhonov. „On the solution of ill-posed problems and the method of regularization“. In: *Dokl. Akad. Nauk SSSR* 151 (1963).
- [80] V. Blobel und E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Teubner-Studienbücher : Physik. Vieweg+Teubner Verlag, 1998.
- [81] V. Blobel. „The RUN Manual: Regularized Unfolding for High-energy Physics Experiments“. In: *OPAL Technical Note TN361* (1996).
- [82] F. Clevermann. „Neutrinoenergieentfaltung von Punktquellkatalogen mit IceCube. Lösung inverser Probleme bei kleinen Statistiken“. Dissertation. TU Dortmund, 2013.
- [83] M. Doert. „The talkative AGN next door: Broad-band spectral variability of the TeV blazar Markarian 501“. Dissertation. TU Dortmund, 2013.
- [84] J.-H. Köhne. „Der Leptonpropagator PROPOSAL“. Dissertation. TU Dortmund, 2014.
- [85] M. Schmitz. „Neuartige Boten aus dem All. Messung des extraterrestrischen und atmosphärischen Myon-Neutrino-Energiespektrum mit IceCube-79“. Dissertation. TU Dortmund, 2014.
- [86] D. Heck et al. „CORSIKA: a Monte Carlo code to simulate extensive air showers“. In: *Wissenschaftliche Berichte* 6019 (1998).

- [87] The IceCube Collaboration. *Hochenergie Datensatz der kosmischen Strahlung*. 2016. URL: <http://internal.icecube.wisc.edu/simulation/dataset/10309>.
- [88] The IceCube Collaboration. *Niederenergie Datensatz der kosmischen Strahlung*. 2016. URL: <http://internal.icecube.wisc.edu/simulation/dataset/9622>.
- [89] The IceCube Collaboration. *IC86 Muon Level3*. 2016. URL: https://wiki.icecube.wisc.edu/index.php/IC86_Muon_Level3.
- [90] T. Fuchs. *HE-Myon-Ereignis*. 2016. URL: <https://www.youtube.com/watch?v=ealHnZ6Jk0s>.
- [91] T. Fuchs. *Myon-Bündel-Ereignis*. 2016. URL: <https://www.youtube.com/watch?v=QZWKp60Yrtc>.
- [92] T. Fuchs. *Ballon-Ereignis*. 2016. URL: <https://www.youtube.com/watch?v=6KT8d8Kw2bw>.
- [93] M. Aartsen et al. „Characterization of the atmospheric muon flux in IceCube“. In: *Astroparticle Physics* 78 (2016).
- [94] A. N. Kolmogorov. „Sulla Determinazione Empirica di una Legge di Distribuzione“. In: *Giornale dell'Istituto Italiano degli Attuari* 4 (1933).
- [95] K. Pearson. „Notes on regression and inheritance in the case of two parents“. In: *Proceedings of the Royal Society of London* 58 (1895).
- [96] The IceCube Collaboration. *Neutrinodatensatz*. 2016. URL: <http://internal.icecube.wisc.edu/simulation/dataset/10602>.
- [97] M. Honda et al. „New calculation of the atmospheric neutrino flux in a three-dimensional scheme“. In: *Phys. Rev. D* 70 (2004).
- [98] R. Enberg, M. H. Reno und I. Sarcevic. „Prompt neutrino fluxes from atmospheric charm“. In: *Phys. Rev. D* 78 (2008).
- [99] M. Aartsen et al. „Evidence for Astrophysical Muon Neutrinos from the Northern Sky with IceCube“. In: *Phys. Rev. Lett.* 115 (2015).
- [100] P. H. Barrett et al. „Interpretation of Cosmic-Ray Measurements Far Underground“. In: *Rev. Mod. Phys.* 24 (1952).
- [101] A. Fedynitch, J. Becker Tjus und P. Desiati. „Influence of hadronic interaction models and the cosmic-ray spectrum on the high-energy atmospheric muon and neutrino flux“. In: *Phys. Rev. D* 86.7 (2012).

Danksagung

Ich danke Prof. Dr. Dr. Wolfgang Rhode für die Betreuung meiner Arbeit und die Unterstützung bei Fragen oder Problemen. Darüber hinaus bedanke ich mich für die Möglichkeit mein Wissen auf Konferenzen zu teilen und mich mit anderen Physikern auf diesem Gebiet auszutauschen.

Weiterhin möchte ich meiner Zweitgutachterin Prof. Dr. Julia Tjus für den Austausch zu meiner Arbeit danken und die Bereitschaft die zweite Gutachterin dieser zu sein.

Ein Dank gilt außerdem der IceCube-Kollaboration, welche immer das Maximum von mir erwartet und mich so zu einem besseren Wissenschaftler gemacht hat. Insbesondere möchte ich Patrick Berghaus danken, der mir mit seinem umfangreichen Wissen über Myonen in IceCube stets zur Seite stand.

Meinem Lehrstuhl Experimentelle Physik V möchte ich danken für das schöne Arbeitsklima und die Bereitschaft Probleme und Fragestellungen gemeinsam lösen zu können. Ich danke auch Maximilian Meier, Thorben Menne, Tim Ruhe, Alexander Sandrock, Sandra Schellhase, Kai Schennetten, Philipp Schlunder, Johannes Werthebach, Mirco Hünnefeld für das Korrekturlesen meiner Arbeit. Ganz besonders danke ich Sabrina Einecke für die ausführliche und intensive Korrektur dieser Arbeit.

Abschließend möchte ich mich besonders bei meiner Familie bedanken, die während meines Weges immer an mich geglaubt hat. Der Zuspruch und die Unterstützung durch euch haben mir Halt gegeben während des gesamten Weges.

Vielen Dank euch allen!