



LEHRSTUHL EXPERIMENTELLE PHYSIK V  
FAKULTÄT PHYSIK  
TECHNISCHE UNIVERSITÄT DORTMUND

# Data-Mining in der Astroteilchenphysik

Entfaltung des Energiespektrums der von IceCube in der 86 String-Konfiguration  
gemessenen Myon-Neutrinos

## Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Dipl.-Phys. Florian Scheriau

geboren in Herdecke

Dortmund, 26. September 2014



1. Gutachter : Prof. Dr. Dr. Wolfgang Rhode

2. Gutachter : Prof. Dr. Metin Tolan

Datum der Abgabe: 26. September 2014



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Astroteilchenphysik</b>	<b>5</b>
2.1	Die geladene kosmische Strahlung . . . . .	7
2.2	Neutrinoastronomie . . . . .	13
<b>3</b>	<b>IceCube</b>	<b>17</b>
3.1	Der Detektor . . . . .	19
3.2	Simulation . . . . .	24
<b>4</b>	<b>Data-Mining</b>	<b>25</b>
4.1	Definitionen . . . . .	26
4.2	Attributselektion . . . . .	28
4.2.1	Minimum Redundancy Maximum Relevance Algorithmus . . . . .	29
4.2.2	Stabilität . . . . .	31
4.3	Ereignisselektion . . . . .	32
4.3.1	Validierung . . . . .	32
4.3.2	Der Entscheidungsbaum-Algorithmus . . . . .	37
4.3.3	Random Forest . . . . .	39
4.3.4	Boosting . . . . .	41
4.4	Entfaltung . . . . .	42
<b>5</b>	<b>Ereignisselektion</b>	<b>47</b>
5.1	Ereignisvorselektion . . . . .	48
5.2	Attributselektion . . . . .	58
5.3	Ereignisselektion mittels Random Forest . . . . .	61
<b>6</b>	<b>Entfaltung des Energiespektrums</b>	<b>75</b>
6.1	Das entfaltete Energiespektrum . . . . .	75
6.2	Validierung der Entfaltung . . . . .	82
6.3	Bestimmung der systematischen Fehler . . . . .	83
6.4	Entfaltung des Energiespektrums in verschiedenen Zenitbereichen . . . . .	87
6.5	Umgewichtung des Spektrums auf einen beliebigen Zenitbereich . . . . .	92
<b>7</b>	<b>Analyse des entfalteten Energiespektrums</b>	<b>97</b>

7.1	Vergleich des entfalteten Energiespektrums mit anderen Messungen . . .	97
7.2	Vergleich des gemittelten Energiespektrums mit Modellen . . . . .	104
7.3	Bestimmung der Modellparameter . . . . .	108
7.3.1	Projektionen der Likelihood-Funktion . . . . .	112
7.3.2	Signifikanz des beobachteten Exzesses . . . . .	119
7.4	Einschränkung für den Fluss der prompten Neutrinos . . . . .	121
7.5	Bestimmung der Modellparameter mit fester prompter Komponente . .	126
<b>8</b>	<b>Zusammenfassung</b>	<b>129</b>
<b>9</b>	<b>Ausblick</b>	<b>133</b>
<b>A</b>	<b>Anhang</b>	<b>135</b>
A.1	Beschreibung der verwendeten Attribute . . . . .	135
A.2	Verwendete Attribute auf Level 3 . . . . .	138
A.3	Verwendete Attribute auf Level 5 . . . . .	143
A.4	Validierung der Entfaltung . . . . .	147
	<b>Literaturverzeichnis</b>	<b>149</b>
	<b>Abbildungsverzeichnis</b>	<b>158</b>
	<b>Tabellenverzeichnis</b>	<b>161</b>

# 1. Einleitung

*„Problems worthy of attack prove their worth by fighting back.“*

---

*(Piet Hein, **Grooks 1**, 1969)*

Besonders wertvolle Herausforderungen zeichnen sich dadurch aus, dass sie nicht einfach zu lösen sind. Eine solche Herausforderung ist die Neutrinoastronomie. Neutrinos sind ideale Informationsträger für die höchstenergetischen Prozesse im Universum. Sie sind ungeladen und erreichen daher die Erde ohne Ablenkung durch Magnetfelder. Im Gegensatz zu Photonen unterliegen sie keiner energieabhängigen Beschränkung in ihrer freien Weglänge durch das extragalaktische Hintergrundlicht. Eine Quelle, wie beispielsweise ein aktiver Galaxienkern wäre daher durch Neutrinos in einem wesentlich größeren Energiebereich beobachtbar. [Tre00]

Neutrinos haben jedoch, abhängig von ihrer Energie, einen sehr kleinen Wirkungsquerschnitt mit Materie. Der geringe Wirkungsquerschnitt macht ein sehr großes instrumentiertes Volumen zwingend notwendig. Je größer das instrumentierte Volumen ist, desto größer werden natürlich auch die Kosten für den Detektor. [Gru01]

Neutrinos wechselwirken nur unter Vermittlung der schwachen Kraft. Diese Teilchen sind daher nicht direkt nachweisbar. Lediglich durch ihre Interaktionen mit Materie sind sie indirekt nachweisbar. Die von einem Detektor gemessenen hochenergetischen extraterrestrischen Neutrinointeraktionen sind von einem extrem dominanten Untergrund, der aus der geladenen kosmischen Strahlung stammt, überlagert. Signal und Untergrund manifestieren sich sehr ähnlich im Detektor. Der Nachweis von hochenergetischen extraterrestrischen Neutrinos wird dadurch deutlich erschwert. Unterschiede zwischen diesen Klassen sind nur in den räumlichen und energetischen Verteilungen zu finden. [Res09]

Eine Unterscheidung dieser Klassen ist nicht trivial. Es bietet sich zunächst an, durch den Einsatz multivariater Methoden atmosphärische Myonen von den Daten zu se-

parieren. Die deutlich verschiedenen spatialen Verteilungsfunktionen machen diese Unterscheidung vergleichsweise einfach. Auf diese Weise ist es möglich, einen Datensatz zu erzeugen, der nur aus neutrinoinduzierten Ereignissen besteht. Der größte Unterschied zwischen atmosphärischen und hochenergetischen extraterrestrischen Neutrinos ist der jeweilige spektrale Index. Daher bietet sich als nächster Schritt eine Entfaltung des Energiespektrums an, um über die Unterschiede im spektralen Verhalten hochenergetische extraterrestrische Neutrinos nachzuweisen. Bisher ist es nicht gelungen durch diese Methode hochenergetische extraterrestrische Neutrinos nachzuweisen. In [Sch14] wurde mit diesem Verfahren ein Exzess nachgewiesen, der nicht mit konventionellen atmosphärischen Neutrinos erklärt werden kann. Es konnte jedoch nicht unterschieden werden, ob der Exzess durch prompte atmosphärische oder hochenergetische extraterrestrische Neutrinos verursacht wird. [A<sup>+</sup>11b, Sch14]

Für sehr hohe Energien ist es kürzlich gelungen durch eine andere Vorgehensweise einen diffusen hochenergetischen extraterrestrischen Neutrinofluss nachzuweisen. Der auf diese Weise gemessene Fluss ist mit einem in Flavor- und Spatialverteilung isotropen Fluss kompatibel. Auf Grund der sehr geringen Menge an identifizierten Ereignissen und großer systematischer Unsicherheit konnte der genaue spektrale Index des hochenergetischen extraterrestrischen Flusses bisher nicht ermittelt werden. Die überwiegende Zahl der gemessenen Ereignisse stammt, bedingt durch die angewandte Methodik, aus Richtung der Südhalbkugel der Erde. Eine dedizierte Analyse des spektralen Verhaltens von Ereignissen von der Nordhalbkugel ist daher von großem wissenschaftlichen Interesse. [A<sup>+</sup>13b, A<sup>+</sup>14c]

Das Ziel der im Folgenden vorgestellten Analyse ist es, durch eine Einbeziehung der niedrigen und mittleren Energien bei einer Fokussierung auf die Nordhalbkugel die beschriebenen Lücken im Wissen über den Fluss der hochenergetischen extraterrestrischen Neutrinostrahlung zu reduzieren.

Die folgenden Kapitel der vorliegenden Arbeit gliedern sich in Anlehnung an die Entstehungsgeschichte der Analyse. Zu Beginn der Analyse steht das Verständnis der in **Kapitel 2** aufgeführten physikalischen Grundlagen der Astroteilchenphysik. Der Schwerpunkt des Kapitels liegt auf der kosmischen Strahlung und den Produktionsmechanismen von Neutrinos. Nach einem kurzen Überblick über die grundlegenden physikalischen Zusammenhänge wird das IceCube-Experiment im Rahmen von **Kapitel 3** beschrieben. Ein besonderer Fokus liegt dabei auf den Rekonstruktionsmethoden, die im weiteren Verlauf für die Analyse von Bedeutung sind. Das Ziel der vorliegenden Analyse ist die Entfaltung des Energiespektrums der von IceCube gemessenen Myon-Neutrinos. Da diese Ereignisse von einem dominanten Untergrund überlagert werden,

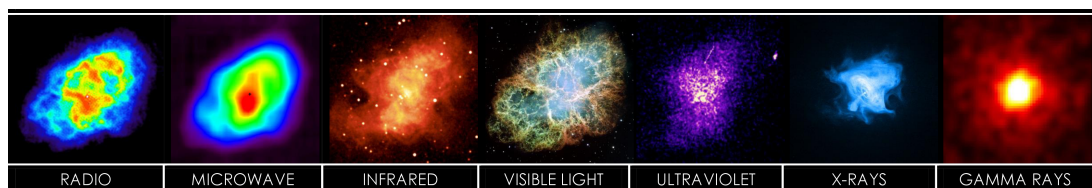


---

müssen die in **Kapitel 4** beschriebenen Methoden genutzt werden, um die gesuchten Neutrinos vom Untergrund zu trennen. Die explizite Umsetzung der Methoden auf die gemessenen Daten wird in **Kapitel 5** beschrieben. Ebenfalls wird in diesem Kapitel der erzeugte Datensatz ausführlich beschrieben. Nachdem im vorherigen Kapitel die Entstehung und die Eigenschaften des erzeugten Datensatzes beschrieben werden, wird in **Kapitel 6** beschrieben, wie das Energiespektrum der von IceCube gemessenen Myon-Neutrinos entfaltet wird. In **Kapitel 7** wird das entfaltete Spektrum mit theoretischen Modellen sowie verwandten Analysen verglichen. Abschließend wird diskutiert, inwieweit es möglich ist, mit dem erzeugten Spektrum Rückschlüsse auf den Fluss der hochenergetischen extraterrestrischen Neutrinos zu ziehen. Die gewonnenen Ergebnisse werden in **Kapitel 8** zusammengefasst und gewertet. Abschließend wird in **Kapitel 9** ein Ausblick auf mögliche Weiterentwicklungen der Analyse gegeben.



## 2. Astroteilchenphysik



**Abbildung 2.1.:**

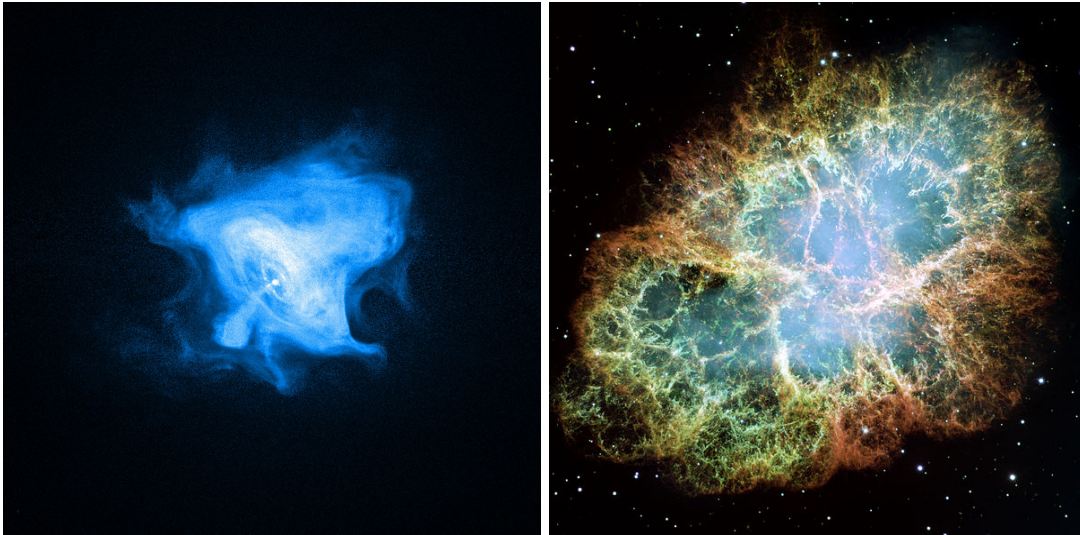
Der Krebsnebel, dargestellt in verschiedenen Wellenlängen. Die Aufnahmen in verschiedenen Wellenlängen unterscheiden sich, da sie unterschiedliche Informationen enthalten. Aus den Bildern in unterschiedlichen Wellenlängenbereichen lassen sich daher jeweils andere Informationen über die zugrundeliegenden physikalischen Abläufe gewinnen. Das Bild wurde [Wik14] entnommen.

Das Ziel der Astroteilchenphysik ist, die Untersuchung und das Verständnis der höchstenergetischen Phänomene des Universums. Da keine Messungen in der Nähe dieser Objekte durchgeführt werden können, muss eine bestehende Informationsquelle genutzt werden. Zu diesem Zweck bietet sich die ungeladene kosmische Strahlung an. Dies wird besonders am Beispiel des Krebsnebels deutlich. Der Krebsnebel ist der Überrest einer Supernova aus dem Jahr 1054 und eines der bekanntesten Objekte am Himmel, das schon seit langem von Menschen beobachtet wird. In Abbildung 2.1 ist der Krebsnebel in verschiedenen Wellenlängenbereichen dargestellt. Von links nach rechts nimmt die mittlere dargestellte Frequenz zu. Die einzelnen Bilder unterscheiden sich sehr deutlich, da jedes der Bilder andere Informationen beinhaltet. Daraus ist erkennbar, dass für ein umfassendes Verständnis der Quelle Informationen über alle Wellenlängenbereiche genutzt werden müssen. Besonders deutlich wird die Wichtigkeit der Beobachtung von mehr als nur einer Informationsquelle, durch den Vergleich von Abbildung 2.2(a) und Abbildung 2.2(b). Die beiden Bilder zeigen den Krebsnebel im optischen Bereich und im Röntgen-Bereich. Im optischen Bereich ist die bekannte Form des Krebsnebels zu sehen. Im Röntgenlicht dagegen ist ein sogenannter Pulsar im Inneren des Krebsnebels zu erkennen. Anhand dieses Beispiels kann geschlossen werden, dass für ein umfassendes Verständnis der höchstenergetischen Phänomene

## 2. Astroteilchenphysik

---

des Universums alle zur Verfügung stehenden Quellen genutzt werden sollten, da sie zusätzliche Informationen enthalten können. [Hes08]



(a) Vergrößerte Darstellung des Röntgen-Bereiches. (b) Vergrößerte Darstellung des optischen Bereiches. Das Bild wurde [NEHL] entnommen. Das Bild wurde [NCSS] entnommen.

**Abbildung 2.2.:** Der Krebsnebel, dargestellt in zwei verschiedenen Wellenlängenbereichen.

In dem Beispiel des Krebsnebels sind Photonen die Informationsträger. Photonen eignen sich besonders als Informationsträger, da sie die Erde erreichen, ohne dabei durch Magnetfelder abgelenkt zu werden. Ebenfalls werden sie, abhängig von ihrer Energie, kaum durch Staub- und Gaswolken absorbiert. Ein Problem für die Beobachtung mittels Photonen ist jedoch ihre begrenzte mittlere freie Weglänge durch das extragalaktische Hintergrundlicht<sup>1</sup>. Diese Interaktion wird als EBL-Absorption bezeichnet. Sie begrenzt die maximale Energie bis zu der hochenergetische Gammastrahlungsquellen beobachtet werden können, da ihre Gammastrahlung ab einer Energie von  $E_\gamma \gtrsim 100 \text{ TeV}$  stark abgeschwächt wird. Es wäre daher wünschenswert, auch oberhalb dieser Energieschwelle Beobachtungen vornehmen zu können. [Tre00, MRMH12]

Neutrinos bieten sich als Ergänzung und Alternative zu Photonen an. Sie sind genau wie Photonen elektrisch ungeladen und erreichen daher ebenfalls die Erde mit ihrer unverfälschten Richtungsinformation. Der Nachteil im Vergleich zu Photonen ist jedoch der geringe Wechselwirkungsquerschnitt von Neutrinos, welcher für die Detektion ein sehr großes instrumentiertes Volumen notwendig macht. Durch das

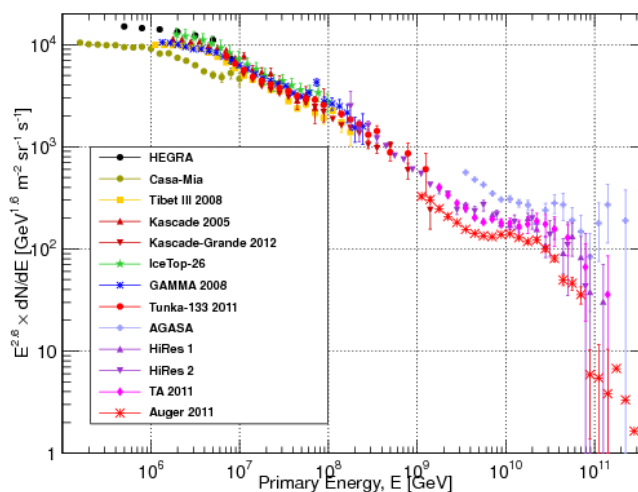
---

<sup>1</sup>Wird im Englischen als Extragalactic Background Light bezeichnet.

Kamiokande-Experiment wurde gezeigt, dass die Observation von Phänomenen außerhalb unseres Sonnensystems mittels Neutrinos trotz dieser Einschränkungen möglich ist, wie sich erstmals durch den Nachweis von Neutrinos aus der Supernova 1987a zeigt. [Gru01, HKK<sup>+</sup>87]

Beobachtungen von astronomischen Objekten werden jedoch unabhängig vom Informationsträger und dem Energiebereich immer vom Untergrund behindert. Es ist daher unerlässlich diesen Untergrund zu verstehen. Eine der wichtigsten Quellen für Untergrund in der Astroteilchenphysik und der Neutrinoastronomie ist die geladene kosmische Strahlung, deren Spektrum in Abbildung 2.3 dargestellt ist. Daher wird diese Strahlung im nachfolgenden Abschnitt 2.1 beschrieben, bevor abschließend in Abschnitt 2.2 auf Neutrinoastronomie eingegangen wird.

## 2.1. Die geladene kosmische Strahlung



**Abbildung 2.3.:**

Spektrum der Energie aller in Luftschauerexperimenten gemessenen Bestandteile der geladenen kosmischen Strahlung. Das Spektrum wird in erster Näherung durch drei verschiedene Potenzgesetze beschrieben. Die wichtigsten Übergangsstellen werden als Knie und Ferse bezeichnet. Die einzelnen Messungen weisen eine hohe Varianz in der Messung des absoluten Flusses der geladenen kosmischen Strahlung auf. Die Abbildung ist [GST13] entnommen.

Die geladene Strahlung verliert auf ihrem Weg zur Erde durch Magnetfelder den größten Teil der Informationen über ihren Ursprung. Auf der Erde kommt sie daher in erster Näherung isotrop verteilt an. Daher lassen sich mit Hilfe der kosmischen Strahlung nur bedingt Rückschlüsse auf spezifische Quellen ziehen. [BCM<sup>+</sup>14]

Das derzeit größte Neutrinoobservatorium ist der IceCube-Detektor mit einem instrumentierten Volumen von einem Kubikkilometer. Auf Grund dessen ist es erstmals möglich, einen diffusen extraterrestrischen Fluss von Neutrinos nachzuweisen. Das größte Hindernis bei der Observation von extraterrestri-

## 2. Astroteilchenphysik

schen Neutrinos ist die geladene kosmische Strahlung, da diese für die Entstehung von atmosphärischen Neutrinos und Myonen verantwortlich ist. Die Teilchen verursachen den extrem dominanten Untergrund bei der Suche nach extraterrestrischen Neutrinos. Daher ist es notwendig, die geladene kosmische Strahlung möglichst genau zu verstehen, da nur dann Rückschlüsse auf extraterrestrische Quellen möglich sind. [HK10, K<sup>+</sup>14, A<sup>+</sup>13b, A<sup>+</sup>14c]

Die folgenden Abschnitte beschreiben die kosmische Strahlung und ihre Interaktionen auf der Erde. Nach einem Überblick über die geladene kosmische Strahlung werden extraterrestrische Neutrinos und ihre möglichen Quellen behandelt.

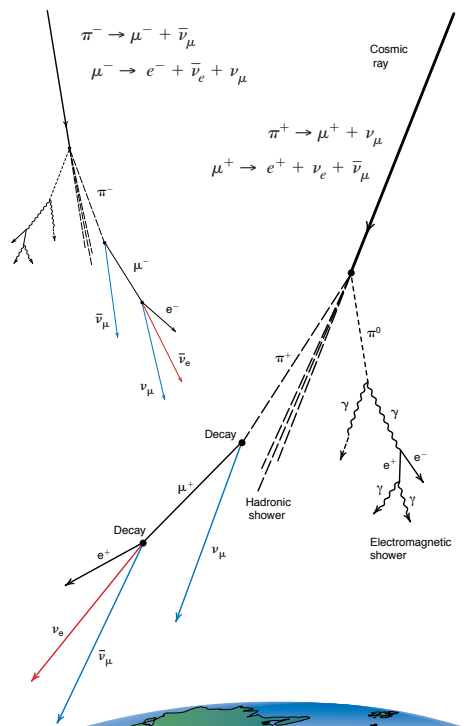
Die geladene Komponente der kosmischen Strahlung besteht zum überwiegenden Teil aus Protonen. Weitere Bestandteile der Strahlung sind beispielsweise Heliumkerne und Elektronen. Das Spektrum der Energie findet sich in Abbildung 2.3. Häufig wird vereinfacht angenommen, dass der Fluss pro Energie der kosmischen Strahlung dem Potenzgesetz

$$\Phi \sim E^{-2,7} \quad (2.1)$$

folgt. Wie an Abbildung 2.3 zu erkennen ist, folgt das Energiespektrum der kosmischen Strahlung jedoch nicht einem ungebrochenen Potenzgesetz. Es lassen sich verschiedene Bereiche erkennen, in denen die kosmische Strahlung jeweils einem unterschiedlichen spektralen Index folgt. Die wichtigsten Übergangspunkte werden als Knie<sup>2</sup> und Ferse<sup>3</sup> bezeichnet. Das Knie liegt bei einer Energie von etwa  $10^{15}$  eV, die Ferse bei einer Energie von etwa  $10^{19}$  eV.

<sup>2</sup>Wird im Englischen als Knee bezeichnet.

<sup>3</sup>Wird im Englischen als Ankle bezeichnet.



**Abbildung 2.4.:**

Hadronische Kaskaden, die durch die Interaktion der geladenen kosmischen Strahlung mit den Atomkernen der Atmosphäre ausgelöst werden. Für die Neutrinoastronomie sind die entstehenden atmosphärischen Neutrinos und Myonen von besonderer Bedeutung, da sie in der Neutrinoastronomie den dominanten Untergrund bilden. Die Abbildung ist [LSG<sup>+</sup>97] entnommen.

Das unterschiedliche spektrale Verhalten wird auf unterschiedliche Ursprünge der Strahlung und damit auf verschiedene Beschleunigungsmechanismen zurückgeführt. Teilchen mit einer Energie unterhalb des Knies werden überwiegend auf einen solaren Ursprung zurückgeführt. Für den Bereich zwischen Knie und Ferse werden galaktische Quellen verantwortlich gemacht. Der Bereich oberhalb der Ferse kann nicht durch galaktische Quellen erklärt werden. Es ist davon auszugehen, dass dieser Energiebereich mehrheitlich von Teilchen mit einem extragalaktischen Ursprung populierte wird. [BE87, GST13]

Ebenfalls zeigt Abbildung 2.3, dass das Spektrum der geladenen kosmischen Strahlung über einen großen Energiebereich von diversen Experimenten vermessen wurde. Die Messungen zeichnen sich jedoch durch eine hohe Varianz in der Bestimmung des absoluten Flusses aus. [A<sup>+</sup>13a, Tos12, GST13, Gai06]

Gelangt ein Teilchen der geladenen kosmischen Strahlung in die Atmosphäre der Erde, kommt es zu einer Interaktion des Primärteilchens mit den Atomkernen der Atmosphäre. Durch diese Reaktion kann eine hadronische Kaskade ausgelöst werden, wie sie in Abbildung 2.4 dargestellt ist. Von besonderer Bedeutung für die Neutrinoastronomie ist dabei, dass innerhalb der hadronischen Kaskade geladene Pionen und Kaonen entstehen. Diese Pionen und Kaonen haben eine geringe Halbwertszeit in der Größenordnung von 10 ns. Sie erreichen nur zu einem sehr geringen Anteil den Erdboden, da sie vorher zu Myonen und Neutrinos zerfallen. Myonen haben eine Lebensdauer von etwa  $2 \mu\text{s}$  und erreichen daher auf Grund ihrer hohen Energie den Erdboden in großer Zahl. Die entsprechenden Zerfälle sind beispielhaft für den Pionkanal in den Gleichungen 2.2 und 2.3 dargestellt. Die Kaonen zerfallen ebenfalls entweder direkt in Myonen und Neutrinos oder indirekt über den Pionkanal. [LMOP47, BCM<sup>+</sup>14, FTD12]



Der Anteil der Myonen und Myon-Neutrinos, die jeweils aus Kaonen und Pionen produziert werden, ist abhängig von der Energie. Die sich daraus ergebenden Verhältnisse für die Kanäle sind in Abbildung 2.5 anhand einer beispielhaften Rechnung dargestellt. Die Grafik zeigt, dass die Abhängigkeit für die Myonenproduktion schwach ausgeprägt ist. Bei der Produktion von Myon-Neutrinos dagegen ist die Abhängigkeit sehr stark

## 2. Astroteilchenphysik

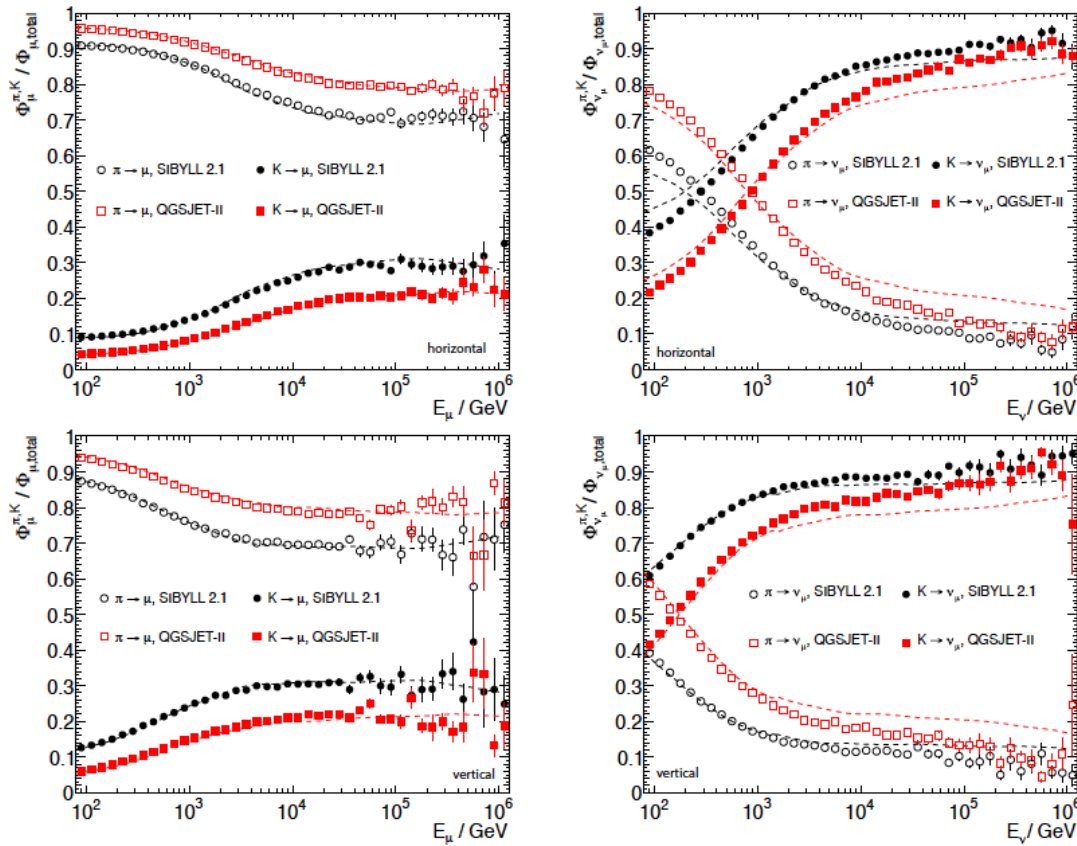


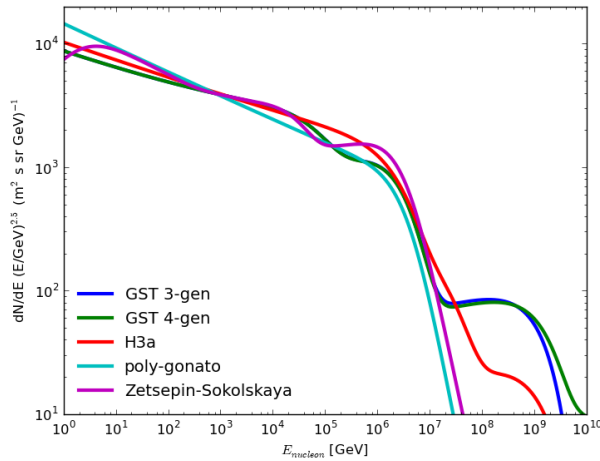
Abbildung 2.5.:

Produktionsverhältnisse für Kaonen und Pionen zeigen eine deutliche Energie- und Winkelabhängigkeit. Die linken Bilder beschreiben die Produktionsverhältnisse von Myonen, die rechten Bilder die Produktionsverhältnisse von Myon-Neutrinos. Die oberen Bilder beschreiben horizontale, die unteren Bilder vertikale Produktionsverhältnisse. Die Abbildung ist [FTD12] entnommen.

ausgeprägt. Da das genaue Verhältnis jedoch unbekannt ist, führt dies zu einer Unsicherheit in der Form des Flusses der atmosphärischen Neutrinos und seiner absoluten Normierung. [FTD12]

Für die Observation von extraterrestrischen Neutrinos stellen diese beiden Komponenten den dominanten Untergrund dar. Der Untergrund der atmosphärischen Myonen ist dabei am einfachsten zu behandeln. Die Erde kann als Abschirmung gegen atmosphärische Myonen genutzt werden. Der Nachteil dieser Vorgehensweise ist jedoch, dass etwa 50% des Winkelbereiches nicht beobachtet werden können. Für den IceCube-Detektor bedeutet dies, dass lediglich der Zenitbereich  $85^\circ \leq \Theta \leq 180^\circ$  beobachtet werden kann. Durch diese Vorgehensweise bleibt jedoch immer noch der Untergrund aus atmosphärischen Neutrinos.




**Abbildung 2.6.:**

Unterschiedliche Primärmodelle der kosmischen Strahlung. Im Bereich der niedrigen Energien beträgt der Abstand zwischen den Modellen wenige Prozent. Die Unterschiede werden erst für große Energien sehr deutlich. Es sind jedoch über den gesamten aufgetragenen Energiebereich Unterschiede in absoluter Normierung und Form der unterschiedlichen Modelle zu erkennen. [AEF<sup>+</sup>]. Die Abbildung ist [Sch14] entnommen.

Zerfall Energieverluste erleiden. Messungen und genauere Rechnungen zeigen jedoch, dass genau wie der Fluss der kosmischen Strahlung, der Fluss der atmosphärischen Neutrinos nicht von einem einfachen, ungebrochenen Potenzgesetz beschrieben werden kann. [BCM<sup>+</sup>14, FTD12]

Eine umfassende theoretische Beschreibung der konventionellen Neutrinos ist auf Grund von hauptsächlich drei Faktoren bisher noch nicht möglich. Wie in Abbildung 2.3 zu sehen ist, gibt es große Unsicherheiten in der Messung des Primärflusses. Diese Unsicherheit spiegelt sich ebenfalls in den unterschiedlichen Modellen wieder, die den Primärfluss beschreiben. Eine Auswahl dieser Modelle findet sich in Abbildung 2.6. Die Abbildung zeigt, dass sich die dargestellten Modelle in absoluter Normierung und Form unterscheiden. Da die konventionellen Neutrinos direkt vom Fluss der geladenen kosmischen Strahlung abhängig sind, übertragen sich die Unsicherheiten des Primärflusses direkt auf die Beschreibung des Energiespektrums der Neutrinos. Die zweite wesentliche Unsicherheit ergibt sich aus den hadronischen Interaktionsmodellen, da für den extrem hohen Energiebereich der Vorwärtsstreuung keine Messungen an Beschleunigern vorliegen und der Wirkungsquerschnitt aus den bisherigen Messungen extrapoliert werden muss. Die dritte wichtige Unsicherheit ist bedingt durch eine

Dieser Untergrund setzt sich aus zwei Komponenten zusammen. Die erste Komponente ist der Fluss konventioneller Neutrinos aus dem Zerfall von Pionen und Kaonen, wie in den Gleichungen 2.2 und 2.3 beschrieben. Eine stark vereinfachte Rechnung ergibt einen Fluss  $\Phi \sim E^{-3,7}$ , der grob einem Potenzgesetz mit einem spektralen Index von  $\gamma = -3,7$  folgt. Der Grund für den im Vergleich zum Primärfluss, um eins verringerten spektralen Index ist, die Lebenszeit der Pionen und Kaonen von mehr als 10 ns. Auf Grund dieser langen Lebenszeit können sie vor dem

zu geringe Grundgesamtheit der zur Verfügung stehenden Simulationen. Dieser Mangel an Simulationen führt gerade für den Bereich der höchsten Energien dazu, dass die Modelle mit ihren bereits beschriebenen Unsicherheiten extrapoliert werden müssen. [FTD12, HKK<sup>+</sup>07]

Die zweite Komponente der atmosphärischen Neutrinos sind die prompten Neutrinos. Liegt die Energie eines Primärteilchens in einer hadronischen Kaskade über der Charmproduktionsschwelle, können extrem kurzlebige charmhaltige Hadronen entstehen. An den Gleichungen 2.4 bis 2.6 ist zu erkennen, dass im Zerfall der Hadronen ein Neutrino entsteht:

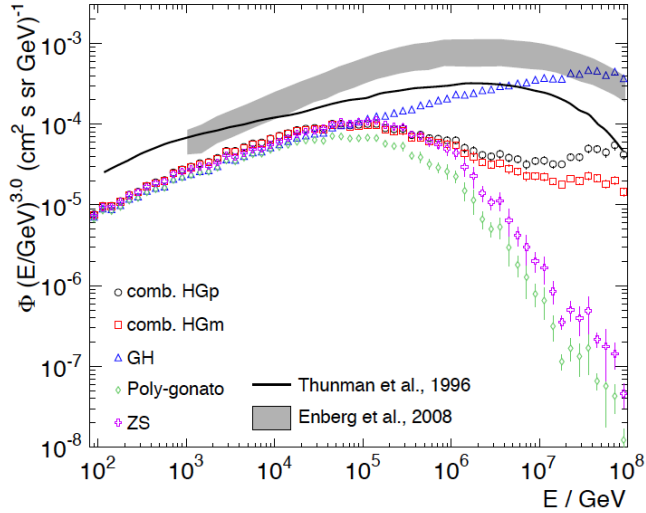
$$D^+ \longrightarrow \bar{K}^0 + l^+ + \nu_l, \quad (2.4)$$

$$D^0 \longrightarrow K^- + l^+ + \nu_l, \quad (2.5)$$

$$\Lambda_c \longrightarrow \Lambda_0 + l^+ + \nu_l. \quad (2.6)$$

Diese charmhaltigen Hadronen haben im Vergleich zu Kaonen und Pionen eine sehr geringe Lebenszeit von etwa einer 1 ps. Die Teilchen können daher nur in einem sehr begrenzten Maß Energie verlieren. Der erwartete spektrale Index sollte daher in erster Näherung dem der geladenen kosmischen Strahlung entsprechen. [ERS08, FTD12]

Wie sich die Unsicherheiten des primären Flusses in den der prompten Leptonen übersetzt ist in Abbildung 2.7 dargestellt. Aus der Unsicherheit des Primärmodells ergibt sich eine große Unsicherheit für die Form und die Normierung des Flusses der prompten Leptonen. Da die prompte Komponente der atmosphärischen Neutrinos bisher nicht nachgewiesen werden konnte, gibt es auch keine Möglichkeit die Unsicherheit



**Abbildung 2.7.:** Auswirkung der Unsicherheiten des Primärflusses auf die prompte Komponente der atmosphärischen Leptonen für das hadronische Interaktionsmodell QGSJet [Ost06]. Der beschriebene Fluss ist gemittelt über alle prompten Leptonen. Zusätzlich werden Modelle der prompten Komponente nach [ERS08] und [TIG96] zum Vergleich dargestellt. Die Abbildung ist [FTD12] entnommen.

ten in der Beschreibung durch Messungen einzuschränken. [ERS08, FTD12, TIG96, Ruh13]

Die großen Unsicherheiten in der theoretischen Beschreibung des prompten Flusses sind von großer Bedeutung bei der Suche nach einem Fluss von hochenergetischen extraterrestrischen Neutrinos. Der spektrale Index der prompten Komponente sollte in etwa  $\gamma = -2,7$  betragen. Erste Messungen des extraterrestrischen Flusses zeigen, dass dieser ein weicherer Spektrum aufweist, als der in der Vergangenheit häufig vermutete spektrale Index von  $\gamma = -2,0$ . Der aktuell beste Fit für den spektralen Index ist  $\gamma = -2,3 \pm 0,3$ . Dies würde bedeuten, dass der Fluss der prompten Neutrinos und der Fluss der hochenergetischen extraterrestrischen Neutrinos einen sehr ähnlichen spektralen Index aufweisen könnten. Aus ähnlichen spektralen Indices folgt, dass die beiden Komponenten das Energiespektrum auf eine sehr ähnliche Weise beeinflussen. Dieser Effekt erschwert die Unterscheidung zwischen den beiden Komponenten. [A<sup>+</sup>14c, A<sup>+</sup>13c, Ruh13]

## 2.2. Neutrinoastronomie

Das Ziel der Neutrinoastronomie ist es, extraterrestrische Phänomene mittels der von ihnen emittierten Neutrinos zu beobachten und zu beschreiben, um das Verständnis der Quellen weiter zu vertiefen. Mit dieser Technik wären dann auch Beobachtungen oberhalb von einigen 100 TeV möglich, da Neutrinos nicht wie Photonen durch Interaktion mit dem extragalaktischen Hintergrundlicht in ihrer freien Weglänge für hohe Energien stark beschränkt sind. Weiter bietet die Beobachtung von Neutrinos aus einer bekannten Quelle viele Möglichkeiten, um das Verständnis über Neutrinos selbst zu vertiefen. So konnten aus den wenigen Neutrinoereignissen, die in Folge der Supernova 1987a gemessen wurden, Rückschlüsse auf die Masse des Neutrinos und die Zahl der Neutrino flavor gemacht werden. [MRMH12, ABKW89]

Für unterschiedliche Quellen liefert der Nachweis oder der Ausschluss von Neutrinoemissionen einen entscheidenden Schritt zu einem besseren Verständnis der Quellen. Ein Beispiel sind aktive galaktische Kerne (AGN), die zu den Objekten im Universum mit dem höchsten Energieausstoß gehören. Diese Objekte konnten bisher nur durch die Emission von Photonen beobachtet werden. Dabei konnte insbesondere hochenergetische Gammastrahlung beobachtet werden. Der genaue Beschleunigungsmechanismus, der die hochenergetische Gammastrahlung produziert, ist dagegen noch nicht bekannt. Grundsätzlich bieten sich hadronische und leptonische Modelle an, um die

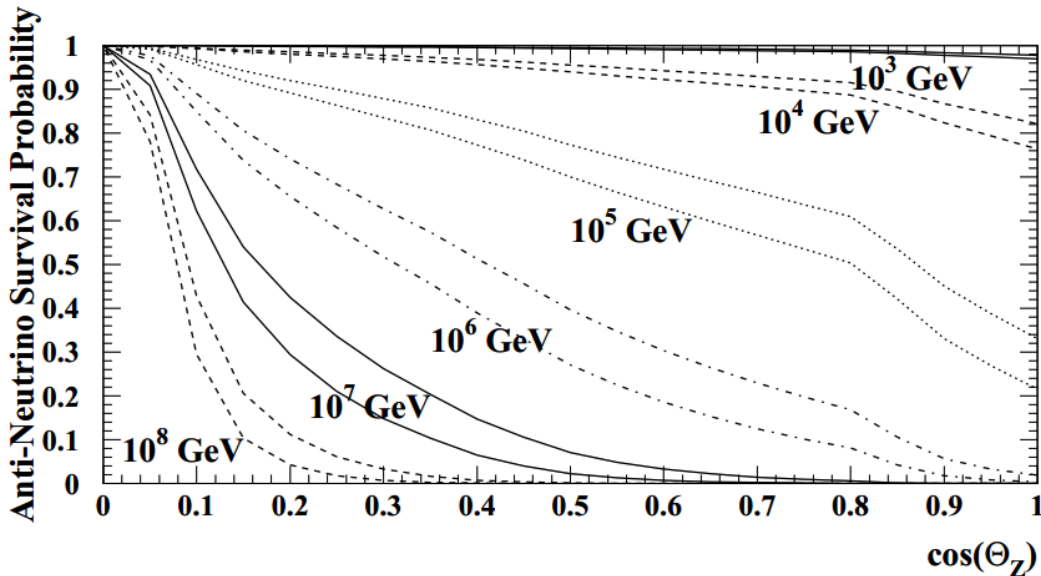


Abbildung 2.8.:

Überlebenswahrscheinlichkeit für Neutrinos beim Durchqueren der Erde für den Winkelbereich  $90^\circ$  bis  $180^\circ$  in Abhängigkeit von der Energie. Ein Winkel von  $180^\circ$  entspricht einem Neutrino, das den IceCube-Detektor vom Nordpol her erreicht. Die Wahrscheinlichkeit, dass ein Neutrino beim Durchqueren der Erde wechselwirkt, steigt mit der Energie und der zurückgelegten Wegstrecke. Die Abbildung ist [ALS02] entnommen.

bisherigen Beobachtungen zu erklären. Neutrinoemissionen bieten die Möglichkeit zwischen hadronischen und leptonicen Beschleunigungsmechanismen in AGNs zu unterscheiden. Beide Beschleunigungsmechanismen erzeugen Photonen, jedoch nur im hadronischen Fall werden geladene Pionen erzeugt, die unter Entstehung von Neutrinos zerfallen. [Bec07, Pro97, KS12, NMB93]

Für eine hadronische Beschleunigung wird ein Produktionsverhältnis der Flavor  $\nu_e : \nu_\mu : \nu_\tau$  an der Quelle von  $1 : 2 : 0$  vorhergesagt. Da Neutrinos jedoch massebehaftet sind, können sie zwischen den verschiedenen Flavoreigenzuständen oszillieren. Die ersten Hinweise auf eine Oszillation der Neutrino flavor ergaben sich aus den Messungen des Homestake Experiments. Der erste Nachweis von Oszillationen gelang jedoch erst später durch das Super-Kamiokande Experiment. Durch die Oszillation zwischen den Flavoren wird auf der Erde ein Verhältnis  $\nu_e : \nu_\mu : \nu_\tau$  von  $1 : 1 : 1$  erwartet. [BCM<sup>+</sup>14, Gri11a, F<sup>+</sup>98, CDDJ<sup>+</sup>98, AJY00]

Wie bereits erwähnt wurde, haben Neutrinos einen kleinen Wirkungsquerschnitt und können daher Materie weitestgehend ohne Wechselwirkung durchqueren. Allgemein ist der Wirkungsquerschnitt jedoch von der Energie abhängig. Dies führt dazu, dass der Fluss von Neutrinos durch die Erde beobachtbar abgeschwächt wird. In Abbildung 2.8

wird die Überlebenswahrscheinlichkeit für Neutrinos bei ihrer Propagation durch die Erde für den Zenitbereich  $90^\circ$  bis  $180^\circ$  bei verschiedenen Energien betrachtet. Es wird deutlich, dass die Wahrscheinlichkeit für eine Wechselwirkung eines Neutrinos beim Durchqueren der Erde mit der Strecke durch die Erde und der Energie des Neutrinos zunimmt. Neutrinos können über zwei Prozesse interagieren. Die auf der Erde ankommenden Neutrinos, können mit den Nukleonen  $N$  über den geladenen Strom interagieren. Infolge dieser Reaktion, beschrieben in Gleichung 2.7, entsteht eine hadronische Kaskade  $X$  und der zum Neutrino passende geladene leptonische Partner  $l$ . [A<sup>+</sup>04, ALS02]



Ebenfalls können die Neutrinos über den neutralen Strom wechselwirken. Bei dieser Wechselwirkung wird das Neutrino gestreut und es entsteht eine hadronische Kaskade. Auf der Grundlage dieser beiden Prozesse lassen sich Neutrinos durch die hadronischen Kaskaden und ihre leptonischen Partner nachweisen.

Für einen Neutrinodetektor ergeben sich bei der Suche nach extraterrestrischen Neutrinos hauptsächlich zwei Formen von Untergrund. Die erste Quelle sind die atmosphärischen Myonen, die in der Atmosphäre über dem Detektor entstehen. Die zweite Quelle sind die atmosphärischen Neutrinos. Für den Nachweis werden daher hocheffiziente Detektoren und hochentwickelte Rekonstruktionsalgorithmen benötigt. In Kapitel 3 werden daher der Aufbau, die Funktionsweise und die Rekonstruktionsalgorithmen des IceCube-Detektors vorgestellt. [A<sup>+</sup>14c]



### 3. IceCube

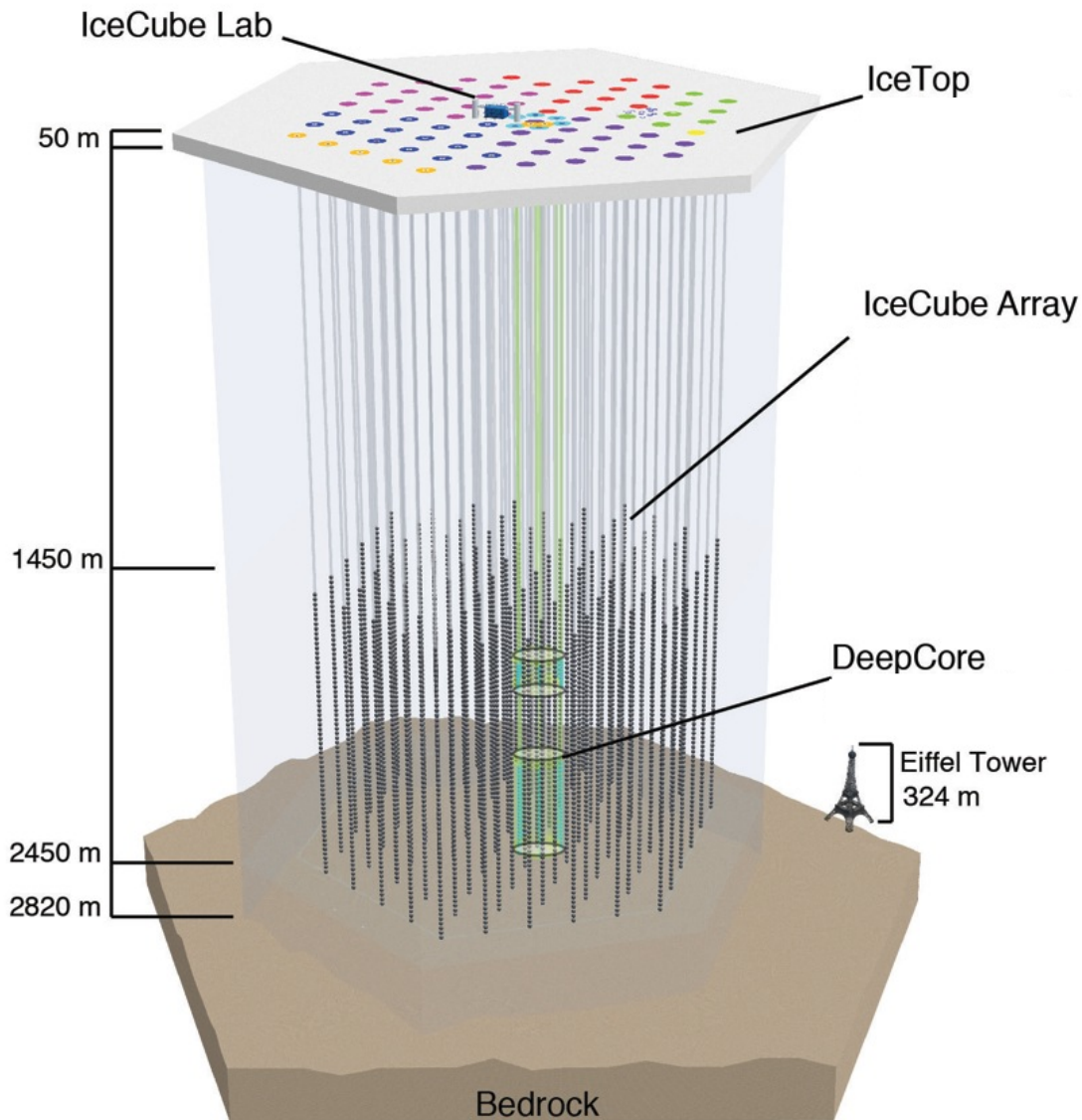
Die Idee Neutrinos als zusätzliche Informationsquelle in der Untersuchung der Phänomene des Universums zu nutzen, ist kurz nach ihrer Entdeckung im Jahr 1954 entstanden. Die erste erfolgreiche Messung von solaren Neutrinos erfolgte im Jahr 1968 durch das Homestake Experiment. [CRH<sup>+</sup>56, RHH68]

Der Fluss hochenergetischer extraterrestrischer Neutrinos ist sehr gering. Die Größe des instrumentierten Volumens ist daher der entscheidende Faktor um Neutrinoastronomie zu betreiben. Daher wurde im Jahr 1960 von Markov vorgeschlagen, Neutrino-teleskope tief unter Wasser zu errichten. In ausreichender Tiefe werden die Detektoren nicht mehr vom Tageslicht erreicht und es ist möglich, die geladenen Interaktionsprodukte der Neutrinos über den Cherenkov-Effekt nachzuweisen. [Mar60, KS12, A<sup>+</sup>13b, A<sup>+</sup>14c, Čer37]

Das derzeit größte Neutrino-teleskop mit einem instrumentierten Volumen von einem Kubikkilometer ist der IceCube-Detektor, dargestellt in Abbildung 3.1. Der Aufbau und die Grundzüge seiner Funktionsweise werden im Folgenden beschrieben. [K<sup>+</sup>14, KS12]

### 3. IceCube

---

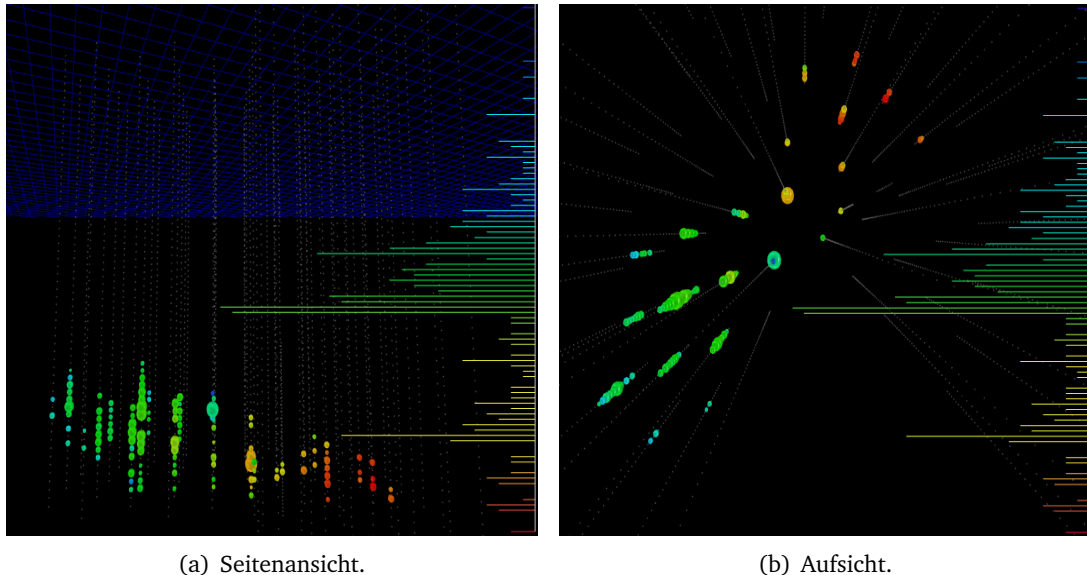


**Abbildung 3.1.:**

Schematische Darstellung des IceCube-Detektors. Auf der Eisschicht über dem Detektor befindet sich die Oberflächenerweiterung IceTop, die sowohl zur Messung der kosmischen Strahlung gedacht ist, als auch eine Vetoschicht für den IceCube-Detektor darstellen kann. Im Eis befinden sich in einer Tiefe von 1,5 km bis 2,5 km die 5160 digitalen optischen Module des IceCube-Detektors. In der Mitte der unteren Hälfte des Detektors befindet sich die Niederenergieerweiterung DeepCore. Diese Erweiterung ist in der Lage, IceCube selbst als Vetoschicht zu nutzen, um damit bis zu einer Energie von ungefähr 10 GeV messen zu können. Für einen Größenvergleich wurde eine maßstabgetreue Zeichnung des Eiffelturms in der Grafik eingezeichnet. Das Bild wurde [The11] entnommen.



### 3.1. Der Detektor



**Abbildung 3.2.:**

Darstellung eines Ereignisses, das vom IceCube-Detektor aufgezeichnet wurde. Die Farbe kodiert den zeitlichen Verlauf des Ereignisses von rot nach grün. Die Größe der einzelnen Kugeln ist proportional zur deponierten Ladung in den digitalen optischen Modulen.

Der IceCube-Detektor, dargestellt in Abbildung 3.1, wird bei der Beantwortung vielfältiger wissenschaftlicher Fragestellungen genutzt. Der Detektor besteht aus 5160 digitalen optischen Modulen (DOMs). Die DOMs sind jeweils mit einem Photomultiplier ausgestattet und daher in der Lage, ein Lichtsignal über eine veränderliche Spannung zu messen. Im IceCube-Detektor sind die DOMs an insgesamt 86 Kabeln, die als Strings bezeichnet werden, befestigt. Die Strings sind in einer Tiefe zwischen 1,5 km und 2,5 km im Eispanzer des Südpols in regelmäßigen Abständen angebracht. Innerhalb der unteren Hälfte des IceCube-Detektors befindet sich die Niederenergieerweiterung DeepCore. Durch einen wesentlich geringeren Abstand der Strings und der einzelnen DOMs ist DeepCore in der Lage die untere Energieschwelle des IceCube-Detektors auf etwa 10 GeV zu senken. Das Eis des Südpols eignet sich auf Grund seiner optischen Transparenz besonders als Detektionsmedium. Auf der Oberfläche über IceCube befindet sich die Oberflächenerweiterung IceTop. Diese bietet die Möglichkeit, als Vetoschicht für den IceCube-Detektor zu fungieren. [A<sup>+</sup>10, A<sup>+</sup>06, Gai10, Tos12, A<sup>+</sup>12, A<sup>+</sup>13f]

Eines der wichtigsten Ziele des Detektors ist der Nachweis und die Vermessung von extraterrestrischen Neutrinos. Für diese sollen unter anderem Richtung, Ankunftszeit und Energie möglichst gut rekonstruiert werden, um mit diesen Informationen Rückschlüsse auf den Ursprung der Neutrinos ziehen zu können. Bei diesen Beobachtungen sind Myon-Neutrinos von besonderem Interesse. Einer der bedeutendsten Gründe für die Betrachtung von Myon-Neutrinos ist die Reichweite der von ihnen induzierten Myonen. Durch die große Reichweite der Myonen ist das effektive Volumen des IceCube-Detektors für Myonen-Neutrinos wesentlich größer als für Elektron-Neutrinos. Die mittlere Weglänge  $x_\mu$  eines Myon-Neutrinos ergibt sich zu:

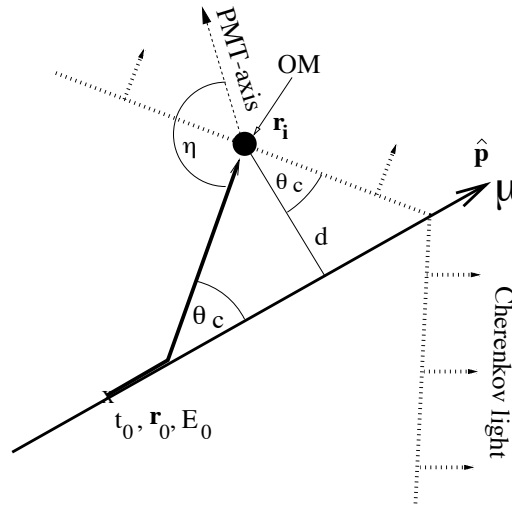
$$x_\mu = \frac{1}{b} \ln \left[ \frac{a + bE_\mu}{a + bE_\mu^{\text{th}}} \right], \quad (3.1)$$

$$a \approx 0,24 \left[ \frac{\text{GeV}}{\text{m}} \right]$$

$$b \approx 3,3 \times 10^{-4} \left[ \frac{1}{\text{m}} \right]$$

Den Myonen ist es daher möglich, abhängig von ihrer Energie, einige Kilometer lange Spuren zu erzeugen. Diese Eigenschaft zeigt sich insbesondere anhand der effektiven Fläche. Die effektive Fläche beschreibt die Akzeptanz des Detektors für Neutrinos verschiedener Energiebereiche und Neutrino flavors. Die effektive Fläche beschreibt wie groß eine Fläche sein muss, die alle auf sie treffenden Neutrinos zu 100 % absorbiert, um die gleiche Effizienz wie der verwendete Detektor zu besitzen. Diese Fläche ist besonders groß für Myon-Neutrinos. Im Folgenden wird daher auf die Messung von Myonen im IceCube-Detektor eingegangen. Eine ausführliche Beschreibung der Berechnung der effektiven Fläche findet sich in [Cle13]. [A<sup>+</sup>06, HK10, Ber09]

Die Detektion und Rekonstruktion von Myonen erfolgt im IceCube-Detektor durch Cherenkov-Licht des Myons und seiner Sekundärteilchen. Ein aufgezeichnetes Ereignis



**Abbildung 3.3.:**

Definition der beschreibenden Parameter eines Myons. Der Parameter  $\mathbf{r}_0$  beschreibt einen beliebigen Punkt auf der Trajektorie des Myons, auf dem es sich zum Zeitpunkt  $t_0$  befindet. Die Richtung, in die sich das Myon bewegt, wird mit  $\hat{\mathbf{p}}$  bezeichnet. Der Parameter  $E_0$  bezeichnet die Energie des Myons. Der Parameter  $\theta_c$  bezeichnet den Winkel, unter dem Cherenkov-Licht emittiert wird. Die Abbildung ist [A<sup>+</sup>04] entnommen.

nis findet sich in Abbildung 3.2. Cherenkov-Licht entsteht, wenn sich ein geladenes Teilchen durch ein dielektrisches Medium bewegt. Die Ladung des Teilchens führt zu einer Polarisation des Mediums, welche eine elektromagnetische Strahlung induziert. Die so erzeugte Strahlung ist auf Grund von destruktiver Interferenz für Teilchen mit einer Gruppengeschwindigkeit unterhalb der Lichtgeschwindigkeit in dem Medium nicht messbar. Erst wenn die Geschwindigkeit des Teilchens größer als die Gruppengeschwindigkeit des Lichts in dem Medium ist, kommt es nicht mehr zu destruktiver Interferenz. Die Cherenkov-Strahlung wird unter dem Winkel  $\theta$ , beschrieben in Gleichung 3.2 und dargestellt in Abbildung 3.3, entlang der Flugbahn des Teilchens emittiert. [Čer37, A<sup>+</sup>04, A<sup>+</sup>14a, Gri11b]

$$\cos(\theta) = \frac{c}{vn} \quad (3.2)$$

$n$ : Brechungsindex des Mediums  
 $c$ : Lichtgeschwindigkeit im Vakuum  
 $v$ : Geschwindigkeit des Teilchens

Damit die DOMs nicht zu einem überwiegenden Teil Rauschen aufzeichnen, wurde für den IceCube-Detektor eine Schwellbedingung entwickelt. Die Schwellbedingung ist das *hard local coincidence* (HLC) Kriterium. Als HLC wird bezeichnet, wenn in einem DOM und einem seiner direkten oder übernächsten Nachbarn innerhalb eines Zeitfensters von  $\pm 1 \mu\text{s}$  Ladung deponiert wird. Ein Ereignis wird aufgezeichnet, wenn mindestens acht DOMs innerhalb  $\pm 5 \mu\text{s}$  angesprochen wurden, die dem HLC Kriterium entsprechen. Darüber hinaus werden anschließend alle DOMs ausgelesen, die in einem Zeitfenster von  $\pm 10 \mu\text{s}$  ein Signal messen. Diese DOMs entsprechen dem *soft local coincidence* (SLC) Kriterium. Die Summe aller durch die DOMs aufgezeichneten Informationen wird in IceCube als ein Ereignis definiert. [A<sup>+</sup>09, A<sup>+</sup>12]

Mit der bekannten Position der einzelnen DOMs und dem zeitlichen Verlauf der deponierten Ladung in den einzelnen DOMs lässt sich die Flugbahn und die Energie des Myons rekonstruieren. Die Geschwindigkeit der Myonen ist von der Energie abhängig, daher liegt die Annahme nah, dass eine Energiemessung durch den Öffnungswinkel des Cherenkovkegels möglich ist. Diese Messmethode ist jedoch unpraktikabel, denn die Auflösung des IceCube-Detektors ist zu gering, als dass eine genaue Messung des Winkels möglich wäre. Darüber hinaus ist zu beachten, dass für hochenergetische Myonen die Geschwindigkeit als  $c$  angenommen werden kann, was zu einem annähernd konstanten Öffnungswinkel führt. [A<sup>+</sup>14a, A<sup>+</sup>14b, A<sup>+</sup>04]

Es ist jedoch möglich, die Energie eines Myons anhand seines Energieverlustes zu schätzen. Die energieabhängigen Energieverluste, die ein Myon erfährt, sind in Abbil-

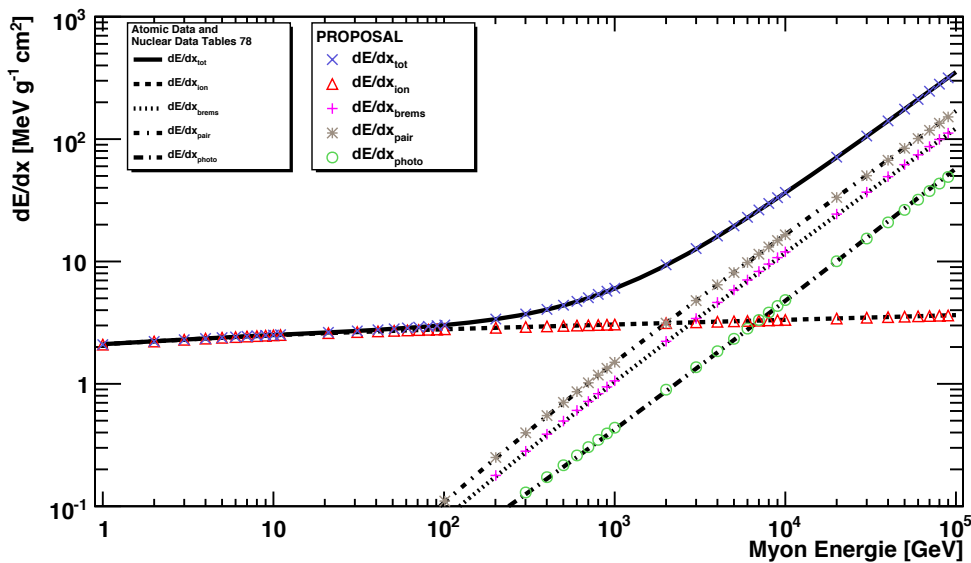


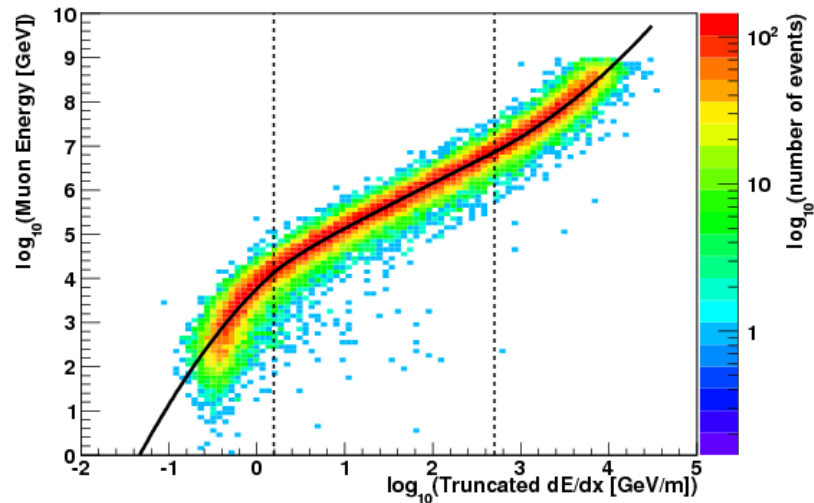
Abbildung 3.4.:

Energieabhängige Energieverluste von Myonen. Bis zu einer Myonenergie von etwa 1 TeV ist die Ionisation der dominante Energieverlust für Myonen. Für höhere Energien ist die Paarproduktion der dominante Energieverlustmechanismus. Die Abbildung ist [KFS<sup>+</sup>13] entnommen.

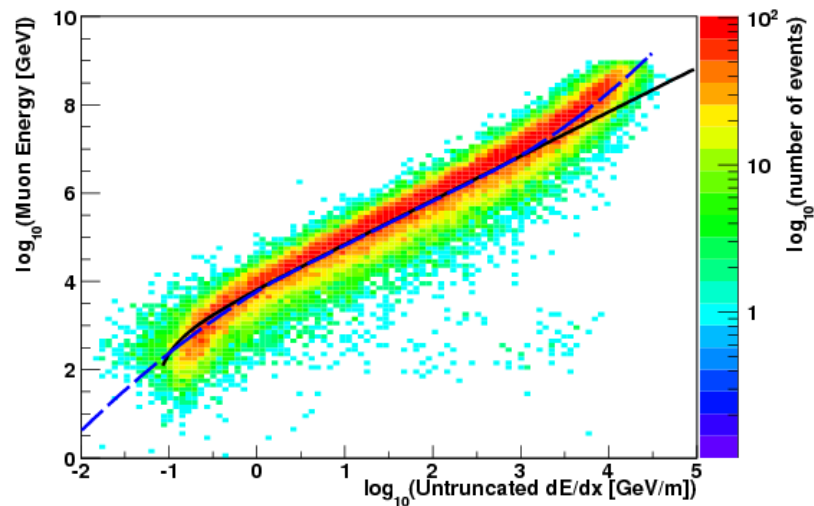
Abbildung 3.4 zusammengefasst. Für niedrige Energien dominiert die Ionisation. Für höhere Energien dominiert die Paarproduktion. Energieverluste durch Paarproduktion und Bremsstrahlung sind von der Myonenergie abhängig. Jedoch stehen einzelne Energieverluste durch Bremsstrahlung in keiner für die Energierekonstruktion nutzbaren Beziehung zu der Myonenergie. Im Vergleich von Abbildung 3.5(a) und Abbildung 3.5(b) ist der negative Einfluss der Bremsstrahlungsverluste für die Energieschätzung zu erkennen. Für die Berechnung der Myonenergie bleibt daher nur die Paarproduktion. Werden nur die nicht durch Bremsstrahlung verursachten Energieverluste pro Meter gegen die Myonenergie, wie in Abbildung 3.5(a), aufgetragen, ist eine deutliche Abhängigkeit mit niedriger Varianz zu erkennen. [A<sup>+</sup>13e]

Eines der wichtigsten Ziele des IceCube-Detektors ist die Suche nach Punktquellen von Neutrinos. Bei dieser Suche bieten sich ebenfalls Myon-Neutrinos an. Die Trajektorie des Myons stimmt energieabhängig mit der des zugehörigen Neutrinos bis auf unter  $1^\circ$  überein. Ebenfalls lassen Myonen eine sehr gute Winkelrekonstruktion zu, die bis auf unter  $1^\circ$  genau ist. Diese Winkelauflösung für Myon-Neutrinos ist im Vergleich zu Elektron-Neutrinos sehr gut, da Elektron-Neutrinos nur eine Winkelauflösung von etwa  $10^\circ$  zulassen. Trotz dieser sehr guten Voraussetzungen und sehr ausgefeilten Tech-

niken ist es bisher noch nicht gelungen eine weitere extraterrestrische Quelle von Neutrinos neben der Sonne und der Supernova 1987a zu identifizieren. [A<sup>+</sup>13d, A<sup>+</sup>04]



(a) Energieverlust pro Meter bereinigt von Bremsstrahlungsenergieverlusten. Die Abbildung wurden [A<sup>+</sup>13e] entnommen.



(b) Energieverlust pro Meter nicht bereinigt von Bremsstrahlungsenergieverlusten. Die Abbildung wurden [A<sup>+</sup>13e] entnommen.

**Abbildung 3.5.:**

Energieverlust von Myonen pro Meter abhängig von ihrer Energie. Im Vergleich der beiden Abbildungen ist zu erkennen, wie die Bremsstrahlung die Abhängigkeit zwischen Energieverlust pro Meter und Myonenergie verzerrt.

## 3.2. Simulation

Das Ziel der Simulation eines Detektors in der Teilchenphysik ist es, den gesamten Detektor und alle möglichen Typen zu messender Ereignisse genau zu beschreiben. Für die vorliegende Analyse ergibt sich damit die Aufteilung der Simulation in zwei Klassen: Die Simulation von Myon-Neutrinos, die das Signal beschreiben und die Simulation von atmosphärischen Myonen, die den Untergrund beschreiben. Die erzeugten Monte-Carlo-Simulationen (MC) dieser beiden Klassen bilden das Fundament für alle weiteren verwendeten Methoden.

Die Simulation der Klassen in IceCube erfolgt im wesentlichen durch zwei Generatoren. Neutrinos werden durch den Neutrinogenerator NuGen und Myonen durch den Generator Corsika erzeugt. Das grundlegende Prinzip für beide Generatoren ist gleich. Auf Grundlage von Pseudozufallszahlen werden Ereignisse erzeugt, die einer vorher definierten Winkel- und Energieverteilung folgen. Lediglich die simulierten Prozesse unterscheiden sich. NuGen erzeugt Neutrinos, die in der Nähe des IceCube-Detektors zur Interaktion gezwungen werden. Das Ziel dieser erzwungenen Interaktion ist, dass die Interaktionsprodukte in IceCube gemessen werden können. Corsika erzeugt Luftschauer, bei denen die erzeugten Myonen eine Chance haben, den Detektor zu erreichen. [HKC<sup>+</sup>98, GK05]

Erreichen Leptonen, die in der Simulation von Corsika und NuGen erzeugt wurden, einen definierten Bereich um den Detektor, werden diese durch die Software PROPOSAL weiter propagiert. Die während der Propagation erzeugten Photonen werden anschließend zu den DOMs propagiert. Dadurch kann die Antwort des Detektors auf die Photonen simuliert werden. [KFS<sup>+</sup>13, Chi13, LMW<sup>+</sup>07]

Anschließend durchlaufen die simulierten Ereignisse die gleiche Analyseketten wie die vom Detektor real aufgezeichneten Ereignisse. Wenn alle physikalischen Gegebenheiten richtig simuliert werden, finden sich in den Attributen für Daten und MC kein Unterschied. Dies ist jedoch für IceCube-Daten und MC nicht immer gegeben. Daher müssen solche Unterschiede untersucht werden und ihr Einfluss auf die gegebene Analyse abgeschätzt werden. Es kann sich dabei anbieten, auf Untermengen der Daten in der Analyse zu verzichten, wenn diese keine gute Übereinstimmung zwischen Daten und MC aufweisen.

Mit den so erzeugten MC und den gemessenen Daten kann begonnen werden, das gewünschte Signal von dem unerwünschten Untergrund zu trennen. Die für diesen Zweck geeigneten Methoden werden im folgenden Kapitel beschrieben.

## 4. Data-Mining

Data-Mining kann wörtlich aus dem Englischen übersetzt werden mit Datenbergbau. Im Unterschied zum klassischen Bergbau sind die zu fördernden Ressourcen allerdings nicht Gold, Kohle oder Salz, sondern Informationen. Im klassischen Bergbau sind die besonders wertvollen Ressourcen nur zu einem sehr geringen Anteil in der Erdkruste vorhanden und müssen daher mit erheblichem Aufwand gefördert werden. Der Goldanteil in der Erdkruste liegt bei etwa  $1 : 10^9$ . Diese Analogie ist insbesondere in der Astroteilchenphysik sehr passend. Die für Physiker wichtigen Informationen sind zumeist in einer enormen Menge an irrelevanten Daten verborgen. In der Astroteilchenphysik beispielsweise liegt das Verhältnis von Signal zu Untergrund für atmosphärische Neutrinos bei etwa  $1 : 10^8$ . Für hochenergetische extraterrestrische Neutrinos ist das Verhältnis von Signal zu Untergrund noch ungünstiger. [Hay12, A<sup>+</sup>14c, Ruh13]

Der Prozess diese gewünschten Informationen zu fördern wird als Data-Mining bezeichnet. Im Vergleich zum Bergbau haben die gewonnenen Informationen keinen materiellen Wert. Eine einfache Überlegung kann aber eine Abschätzung für ihren Wert bringen und damit die Wichtigkeit der im weiteren Verlauf dieses Kapitels beschriebenen Methoden liefern.

Die Experimente in der Astroteilchenphysik kosten in ihrem Aufbau und Betrieb viele Millionen Euro. Dabei haben sie eine begrenzte Lebenszeit, in der sie Daten nehmen können. Wenn durch den optimalen Einsatz der Methoden des Data-Minings die Effizienz um einen Faktor von 20 % gesteigert werden kann, ist dies direkt mit einer um 20 % längeren Laufzeit vergleichbar. Die Kosten des Experiments bleiben jedoch konstant. Es bietet sich also eine einfache Möglichkeit die Wirtschaftlichkeit von Experimenten bei einer gleichzeitigen Erhöhung des wissenschaftlichen Nutzens direkt zu steigern. Der Aufbau des IceCube-Detektors wurde mit mehreren hundert Millionen Euro gefördert. Bei einer Steigerung der Effizienz um 20 % würde das einem Mehrwert von mehreren zehn Millionen Euro entsprechen. [Jam12]

In den folgenden Abschnitten werden die im Rahmen dieser Arbeit verwendeten und teilweise neu entwickelten Methoden vorgestellt. Die Reihenfolge der Vorstellung der Methoden erfolgt dabei analog zu der Reihenfolge, wie sie in Kapitel 5 und Kapitel 6 verwendet werden.

### 4.1. Definitionen

Grundlage für jede Form von Data-Mining sind Daten. Diese Daten setzen sich aus der Menge aller Ereignisse zusammen. Ein Ereignis ist beispielsweise ein von IceCube gemessenes Myon. Jedes Ereignis wird dabei von der Menge seiner Attribute, häufig auch als Parameter oder Observablen bezeichnet, und deren Ausprägungen vollständig beschrieben. Die Attribute sind beispielsweise ein

rekonstruierter Winkel oder die im Detektor deponierte Energie. Die Attribute werden abhängig vom Datentyp unterschiedlich bezeichnet. Eine kurze Zusammenfassung findet sich in Tabelle 4.1.

Innerhalb eines Data-Mining Prozesses kann jedem Attribut eine Rolle zugewiesen werden. Abhängig von ihrer Rolle erfüllen die Attribute verschiedene Aufgaben. Die wichtigsten Rollen sind Regulär, ID und Label. Reguläre Attribute sind alle Attribute ohne eine besondere Rolle. Im Attribut ID wird eine eindeutige Identifikation für jedes Ereignis codiert. Das mit Abstand wichtigste Attribut ist das Label. Im Label wird die Klassenzugehörigkeit eines Ereignisses beschrieben. Die Klasse definiert dabei zum Beispiel, ob es sich bei dem Ereignis um Signal oder Untergrund handelt. [WEH11, Cow98]

Es gibt zwei Wege die Information zu erhalten. Für MC ist die Information über ihre Klassenzugehörigkeit bekannt. Für gemessene Daten kann die Information durch einen Experten erfolgen, der anhand der Attributverteilung eine Zuordnung vornehmen kann. Daten, die kein Label enthalten, werden als ungelabelte Daten bezeichnet.

**Tabelle 4.1.:**

Die Tabelle gibt eine Übersicht über die verschiedenen Typen von Attributen, die im Bereich des Data-Mining verwendet werden.

Bezeichnung	Datentyp	Beispiele
numerisch	double	-1; 1 ; 2,32 ; 123
integer	integer	1; 2; -8
(poly)nominal	beliebig	grün; Haus; -2; 11
binominal	boolean	True; False
ordinal	text	kalt > mild > warm



Ziel bei ungelabelten Daten ist, ihre Klassenzugehörigkeit möglichst genau zu schätzen. Dieser Prozess wird als Klassifikation bezeichnet. [WEH11, Cow98]

**Tabelle 4.2.:**

Die Konfusionsmatrix für eine Separation eines Datensatzes mit zwei Klassen. Ereignisse der Klasse positiv, die als solche erkannt werden, werden als  $r_p$  bezeichnet. Ereignisse der Klasse positiv, die nicht als solche erkannt werden, werden als falsch negativ  $f_n$  bezeichnet. Entsprechend ergeben sich für die Klasse negativ die Fälle richtig negativ  $r_n$  und falsch positiv  $f_p$ .

	echt positiv	echt negativ
vorhergesagt positiv	richtig positiv $r_p$	falsch positiv $f_p$
vorhergesagt negativ	falsch negativ $f_n$	richtig negativ $r_n$

Um die Qualität einer Klassifikation zu überprüfen, wird eine Konfusionsmatrix erstellt. Ein Beispiel für eine Konfusionsmatrix für eine Separation eines Datensatzes mit den zwei Klassen positiv und negativ ist in Tabelle 4.2 dargestellt. Ereignisse der Klasse positiv, die als positiv klassifiziert wurden, werden als richtig positiv  $r_p$  bezeichnet. Ereignisse der Klasse positiv, die

der Klasse negativ zugeordnet wurden, werden als falsch negativ  $f_n$  bezeichnet. Entsprechend ergibt sich für die negative Klasse die Bezeichnung falsch positiv  $f_p$  und richtig negativ  $r_n$ . [WEH11, Cow98]

Aus diesen vier Grundbausteinen lassen sich weitere Kenngrößen ableiten, die Aussagen über die Qualität einer Klassifikation liefern können. Die wichtigsten Kenngrößen für den Fall von zwei Klassen sind die Reinheit<sup>1</sup>

$$p = \frac{r_p}{r_p + f_p}, \quad (4.1)$$

die Effizienz<sup>2</sup>

$$r = \frac{r_p}{r_p + f_n}, \quad (4.2)$$

und die Accuracy

$$a = \frac{r_p + r_n}{r_p + r_n + f_p + f_n}. \quad (4.3)$$

Für ein Klassifikationsproblem mit mehr als einer Klasse sind die entsprechenden Klassen so zu erweitern, dass Größen jeweils pro Klasse angegeben werden. [WEH11, Cow98]

<sup>1</sup>Die Reinheit wird auch als Precision bezeichnet.

<sup>2</sup>Die Effizienz wird auch als  $t_p$ -Rate und als Recall bezeichnet.

### 4.2. Attributselektion

Die von IceCube gemessenen Daten lassen sich für jedes Ereignis auf drei Grundbausteine zurückführen. Die bekannten Positionen der DOMs, die Zeit zu denen sie angesprochen wurden, und die Ladungen, die in ihnen deponiert wurden. Auf Grundlage dieser Informationen werden eine Fülle von weiteren Attributen berechnet, wie Trajektorien, Längen und Energien. Insgesamt werden für jedes Ereignis mehr als 1 000 Attribute erstellt.

Viele dieser Attribute weisen dabei eine sehr hohe Korrelation untereinander auf. Ein Beispiel für sehr hoch korrelierte Attribute sind die mehr als 10 Zenitwinkelschätzer in IceCube. Alle diese Schätzer sind untereinander korreliert und weisen Korrelationskoeffizienten größer als 0,95 auf. Es lässt sich daher feststellen, dass die über 1 000 Attribute ein hohes Maß an Redundanz aufweisen.

Die Daten des IceCube-Detektors werden für viele verschiedene physikalische Fragestellungen aufbereitet. Daher haben viele der berechneten Attribute keine oder nur eine sehr geringe Aussagekraft für eine spezielle Fragestellung. Für eine Analyse von Myon-Neutrinos machen viele in der Suche nach magnetischen Monopolen generierten Attribute keinen Sinn, wobei a priori nicht ausgeschlossen werden kann, dass gerade diese Attribute eine besonders hohe Trennkraft aufweisen. Es ergibt sich jedoch insgesamt eine hohe Anzahl nicht relevanter Attribute.

Nicht relevante und redundante Attribute bringen für eine spätere Separation der Daten keinen Nutzen und können im Extremfall die Qualität der Trennung zwischen Signal und Untergrund negativ beeinflussen. Durch die vielen irrelevanten Informationen kann es passieren, dass ein zur Trennung der Daten verwendeter Algorithmus nicht mehr vollständig in der Lage ist, die relevanten Informationen zu finden. Darüber hinaus bedeuten mehr Attribute, dass mehr Speicherplatz und Rechenleistung in allen weiteren Schritten benötigt werden. Dies kann dazu führen, dass auf Grund von eingeschränkten Ressourcen einige Algorithmen nicht verwendet werden können, da zum Beispiel nicht mehr ausreichend Arbeitsspeicher vorhanden ist. Das Ziel ist also, eine optimale Repräsentation der Daten bei einer möglichst geringen Anzahl an Attributen zu finden.

Die einfachste Möglichkeit dieses Problem anzugehen ist, mit physikalischem Sachverstand alle Attribute zu betrachten und jene Attribute zu wählen, die auf Grund von Wissen und Erfahrung die Daten am besten beschreiben. Diese Vorgehensweise ist nicht optimal. Es ist mit Wissen und Erfahrung sehr schwer eine optimale Abwägung

aller Faktoren zu treffen. Insbesondere die Frage, ab wann ein Attribut zu redundant ist oder ab wann es relevant genug ist, lässt sich nicht einfach beantworten. Die Situation wird noch komplexer, wenn Wechselwirkungen zwischen Attributen betrachtet werden. Solche Wechselwirkungen sind, dass zwei Attribute alleine sehr irrelevant sind, jedoch in Kombination mit dem jeweils Anderen eine hohe Relevanz aufweisen. Ein einfaches Beispiel sind die Azimutrekonstruktionen in IceCube. Alleine haben diese Rekonstruktionen für die Unterscheidung zwischen Signal und Untergrund keine Aussagekraft, zusammen haben sie jedoch eine hohe Aussagekraft. Denn ihre Differenz gibt einen Aufschluss über die Qualität der Rekonstruktion des Ereignisses. Die Qualität der Rekonstruktion ist wiederum sehr stark von der Topologie des Ereignisses abhängig. Die Topologie kann für die Trennung sehr relevant sein. Auch wenn durch diese Vorgehensweise eine gute Repräsentation der Daten gefunden wird, bleibt ein weiteres Problem. Es ist davon auszugehen, dass jeder Experimentator, der dieser Vorgehensweise folgt, zu einem anderen Satz an Attributen kommen wird. Das bedeutet, dass die so erzeugten Repräsentationen der Daten nicht stabil sind. Genau dies ist jedoch ein wichtiges Ziel. Der Ausgang eines Experiments darf nicht von Zufällen, wie der Meinung des Experimentators über bestimmte Attribute, abhängen. Eine mögliche Lösung dieses Problems ist die Wahl eines geeigneten Algorithmus, der nicht von zufälligen Schwankungen abhängt.

Ein solcher Algorithmus sucht anhand der zur Verfügung stehenden Simulationen nach der besten Repräsentation für ein gegebenes Problem, wie die Unterscheidung zwischen Signal und Untergrund. Es ergibt sich jedoch zwangsläufig, dass die Simulationen statistischen Schwankungen unterliegen. Diese Fluktuationen sollten jedoch keinen oder nur einen möglichst kleinen Einfluss auf die Wahl der Attribute haben. Wie stark ein Algorithmus von statistischen Schwankungen abhängig ist, kann also als Qualitätsmerkmal des Algorithmus gewertet werden.

Im Folgenden wird beispielhaft auf den *Minimum Redundancy Maximum Relevance* Algorithmus eingegangen und dabei seine Funktionsweise erläutert. Abschließend werden zwei Verfahren vorgestellt, die in der Lage sind, die Stabilität einer Attributselektion zu messen.

### 4.2.1. Minimum Redundancy Maximum Relevance Algorithmus

Physikalische Daten werden durch einen  $N$ -dimensionalen Vektor beschrieben. Die Anzahl der Dimensionen ist dabei durch die Anzahl der Attribute gegeben. Das Ziel

der Attributselektion ist es, stabil für eine gegebene Fragestellung die optimale Repräsentation der Daten in einer möglichst kleinen Zahl von Dimensionen zu finden.

Eine Methode, die diese Zielsetzung erfüllt, ist der *Minimum Redundancy Maximum Relevance* Algorithmus (MRMR). Der Algorithmus ist so aufgebaut, dass die gewählten Attribute eine möglichst hohe Korrelation zum Zielattribut aufweisen unter der Bedingung einer minimalen Korrelation untereinander. Dabei ist zu beachten, dass die richtige Wahl des Korrelationsmaßes von großer Bedeutung ist. Für die Korrelation eines Attributes zu einem binominalen Zielattribut, erschließt sich leicht, dass hier die Wahl der Pearson Korrelation unvorteilhaft ist und daher ein geeignetes Maß verwendet werden sollte. [DP05, Sti89]

Der MRMR Algorithmus verwendet, um die Korrelation zum Zielattribut und die Korrelationen der Attribute untereinander zu messen, Mutual Information. Die Korrelation von  $j$  Attributen  $\vec{x}$  zum Zielattribut  $y$  kann geschrieben werden als  $R(\vec{x}, y)$ . Auf ähnliche Weise kann die Korrelation von zwei Attributen  $x$  und  $x'$  geschrieben werden als  $D(x, x')$ . Damit ergibt sich für die Wahl einer Untermenge an Attributen die zu maximierende Größe

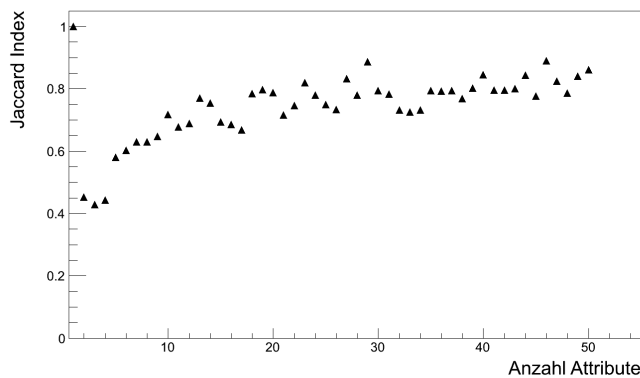
$$Q = R(\vec{x}, y) - \frac{1}{j} \sum_{x' \in F_j} D(x', x). \quad (4.4)$$

Im ersten Schritt wird das Attribut gewählt, welches die höchste Korrelation zum Zielattribut aufweist. Da bisher keine weiteren Attribute gewählt wurden, gilt das  $Q = R(\vec{x}, y)$ . Daher ist die Wahl des ersten Attributs trivial. Im zweiten Schritt steht somit das erste Attribut bereits fest. Das zweite Attribut wird nun so gewählt, dass  $Q$  maximal wird. Daher wird ein Attribut gewählt, das zwar eine hohe Korrelation zum Zielattribut aufweist, jedoch eine geringe Korrelation zu dem bereits gewählten Attribut. Diese Vorgehensweise wird solange wiederholt, bis alle Attribute ausgewählt wurden. Die Reihenfolge, in der die Attribute gewählt werden, entspricht einer Liste ihrer Relevanz zum Zielattribut unter der Bedingung einer minimalen Redundanz zu bisher gewählten Attributen. [DP05, Sti89]

Dies beantwortet jedoch nicht die Frage, welche die optimale Anzahl an Attributen für ein gegebenes Problem ist. Diese Frage ist ein weiteres Optimierungsproblem. Ein erster Hinweis auf die optimale Anzahl an Attributen ergibt sich aus der Stabilität der Attributselektion. Die minimale Anzahl an Attributen, die gewählt werden sollte, ergibt sich dabei dadurch, dass der Algorithmus stabile Ergebnisse liefern soll. Dies

ist jedoch häufig erst ab einer Mindestzahl von Attributen gegeben. Daher wird im Folgenden auf die Berechnung der Stabilität einer Attributselektion eingegangen.

#### 4.2.2. Stabilität



**Abbildung 4.1.:** Der Jaccard-Index in Abhängigkeit von der Anzahl der gewählten Attribute am Beispiel der Analyse von [Lot12].

Da eine endliche Stichprobe immer eine nicht verschwindende Varianz aufweist, haben Fluktuationen einen Einfluss auf die Ergebnisse des zur Attributselektion genutzten Algorithmus. Im Idealfall ist es wünschenswert, dass solche Fluktuationen einen möglichst kleinen Einfluss auf die Attributselektion haben. Um diesen Einfluss quantifizieren zu können bietet sich der Jaccard-Index

$$J = \frac{|A \cap B|}{|A \cup B|} \quad (4.5)$$

an. Im einfachsten Fall wird der Datensatz in zwei disjunkte Mengen geteilt. Auf beiden Mengen wird eine beliebige Attributselektion durchgeführt, wodurch sich die beiden Attributmengen  $A$  und  $B$  ergeben. Für diese beiden Mengen ist  $|A \cap B|$  die Anzahl der Attribute, die in beiden Mengen vorhanden sind. Der Term  $|A \cup B|$  beschreibt den Betrag der Vereinigung der beiden Mengen, also die Anzahl aller Attribute, die in beiden Durchläufen gewählt werden. Der Nenner aus Gleichung 4.5 kann als Normierung verstanden werden, da durch ihn  $J$  auf das Intervall  $[0,1]$  normiert wird. [SM11, Jac12]

Da bei einem Stichprobenumfang von zwei eine große Varianz von  $J$  zu erwarten ist, bietet sich eine Verallgemeinerung von Gleichung 4.5 auf beliebig viele Mengen an. Diese ergibt sich zu

$$J(A_1, A_2, \dots, A_n) = \frac{|A_1 \cap A_2 \cap \dots \cap A_n|}{|A_1 \cup A_2 \cup \dots \cup A_n|}. \quad (4.6)$$

Der Jaccard-Index für eine Attributselektion zeigt häufig ein in Abbildung 4.1 dargestelltes Verhalten. Für sehr kleine Anzahlen an Attributen ist eine geringe Stabilität zu erkennen. Der Grund dafür liegt darin, dass einige wenige trennstarke Attribute etwa gleich häufig gewählt werden. Welches Attribut gewählt wird, ist dann stark von den statistischen Fluktuationen des gegebenen Datensatzes abhängig. Genau diese Abhängigkeit ist jedoch in einer Attributselektion unerwünscht. Für eine steigende Anzahl an gewählten Attributen ist dann eine langsam sättigende Steigerung der Stabilität zu erkennen. Hieraus lässt sich ein Kriterium für eine minimale Anzahl an zu wählenden Attributen ableiten. Befindet sich die Stabilität in Sättigung, so ist davon auszugehen, dass der Einfluss von statistischen Fluktuationen auf die Wahl der Attribute minimal ist. Es sollte daher mindestens eine Zahl an Attributen gewählt werden, bei der eine Sättigung in der Stabilität der Attributselektion beobachtet wird. [SM11, Jac12]

Die maximale Anzahl an Attributen kann jedoch auf diese Weise nicht festgelegt werden, da jedes zusätzlich gewählte Attribut mehr Informationen trägt. Die maximale Anzahl an Attributen kann daher über Optimierung der Effizienz der Attributselektion für verschiedene Anzahlen an Attributen ermittelt werden. Eine maximale Anzahl an Attributen ergibt sich anschließend aus der Optimierung der Effizienz der Separation für verschiedene Anzahlen an Attributen.

### 4.3. Ereignisselektion

In Abschnitt 4.2 wurden Methoden beschrieben, um für ein gegebenes Problem eine optimale Repräsentation der Daten zu finden. In diesem Abschnitt werden Methoden beschrieben, die ausgehend von einer Repräsentation der Daten eine optimale Trennung zwischen verschiedenen Klassen erlauben. Zunächst werden im Unterabschnitt 4.3.1 Methoden beschrieben, die es ermöglichen, beliebige Verfahren des maschinellen Lernens zu validieren. Abschließend werden in den Unterabschnitten 4.3.2 bis 4.3.4 Methoden des maschinellen Lernens beschrieben, die in der Lage sind, eine Trennung zwischen verschiedenen Klassen vorzunehmen.

#### 4.3.1. Validierung

Validierung im Kontext des Data-Minings bezeichnet unter Nutzung eines Algorithmus die statistischen und systematischen Fehler eines Modells zu berechnen. Die Validierung ist die einzige Möglichkeit aussagekräftige Fehler auf die Kenngrößen, wie

	$N_1$	$N_2$	$N_3$	$N_4$	$N_5$
1. Schritt					
2. Schritt					
3. Schritt					
4. Schritt					
5. Schritt					

Abbildung 4.2.:

Rot markiert die Teilmengen, die zum Testen genutzt werden. Grün markiert die Teilmengen, die zum Trainieren genutzt werden. In der fünffachen Kreuzvalidierung wird die Gesamtmenge aller Ereignisse in fünf disjunkte Teile  $N_i$  getrennt. In jedem Schritt werden vier Teile für das Trainieren genutzt. Die verbleibende Menge wird zum Testen des Modells benutzt. Dieser Vorgang wird fünfmal wiederholt, so dass jede Menge genau einmal zum Testen des Modells genutzt wurde. Anschließend kann über die Ergebnisse der fünf Tests gemittelt und deren Varianz bestimmt werden.

Reinheit und Effizienz, eines Modells zu berechnen. Diese Fehler sind sehr wichtig, da ohne eine Angabe des Fehlers jede Angabe einer Kenngröße nur bedingt sinnvoll ist. Darüber hinaus ist es durch die Fehler möglich, zwei der wichtigsten Fehlerquellen in einem Experiment zu erkennen und abzuschätzen. Ein großer Fehler der Validierung kann ein deutliches Anzeichen für Überanpassung des Modells<sup>3</sup> an die MC sein. Liegt der in Daten gemessene Wert signifikant außerhalb der Fehler der auf MC bestimmten Kenngröße, so kann dies ein Anzeichen für einen Unterschied zwischen den Daten und den MC sein. Ein solcher Unterschied kann zwei Ursachen haben. Die erste Ursache ist eine fehlerhafte Simulation des Detektors. Die zweite Ursache ist eine nicht bekannte und daher nicht simulierte Komponente in den Daten.

Im Folgenden werden die zwei wichtigsten Validierungsmethoden Kreuzvalidierung (CV) und Bootstrap-Validierung (BV) vorgestellt. Weiterführend wird eine erweiterte Form der BV vorgestellt, die die Verteilungsfunktion jeder Klasse in allen Attributen schätzen kann. [K<sup>+</sup>95, LM68, Efr79]

### Kreuzvalidierung

Um ein gegebenes Modell zu validieren, ist die Kreuzvalidierung eine sehr weit verbreitete Methode. In dieser werden die Daten in  $N$  disjunkte Teile aufgespalten. Im ersten Schritt werden  $N - 1$  Teile für das Lernen eines Modells  $M_i$  genutzt. Die nicht

<sup>3</sup>Überanpassung des Modells an einen Datensatz wird auch als Overtraining bezeichnet.

gewählte Teilmenge  $N_i$  wird anschließend genutzt, um  $M_i$  darauf anzuwenden. Dabei wird für jede gesuchte Kenngröße der Wert für diesen Datensatz berechnet. Im nächsten Schritt wird eine andere Menge für die Berechnung der Kenngröße gewählt. Dieses Vorgehen wird so lange wiederholt bis alle Mengen genau einmal zum Testen verwendet wurden. Es ergeben sich daher  $N$  verschiedene Messungen einer Kenngröße. Für diese kann der Mittelwert und die zugehörige Standardabweichung berechnet werden. [HTF01]

Die Anzahl der disjunkten Teilmengen definieren dabei die Bezeichnung der Methode. Für  $N = 2$  wird die CV als Splitvalidierung bezeichnet. Werden so viele Teilmengen gebildet, dass jedes Ereignis eine Teilmenge bildet, wird die CV als *Leave One Out* bezeichnet. Alle anderen Fälle werden als  $N$ -fache CV bezeichnet. Eine schematische Darstellung einer fünffachen CV findet sich in Abbildung 4.2. [HTF01]

#### **Bootstrap-Validierung**

Der grundlegende Unterschied der BV zur CV ist der Übergang von disjunkten zu nicht disjunkten Teilmengen. Für jeden Durchgang der BV werden eine feste Anzahl von Ereignissen aus der Grundmenge per Bootstrapping<sup>4</sup> gezogen, die für das Trainieren des Modells verwendet werden. Die nicht gezogenen Ereignisse werden anschließend auf das gebildete Modell angewandt. Auf Grundlage der angewandten Ereignisse werden die gewünschten Kenngrößen berechnet. Dieser Vorgang wird  $N$ -mal wiederholt. Abhängig von der Anzahl der Wiederholungen wird von einer  $N$ -fachen BV gesprochen. [Efr79, HTF01]

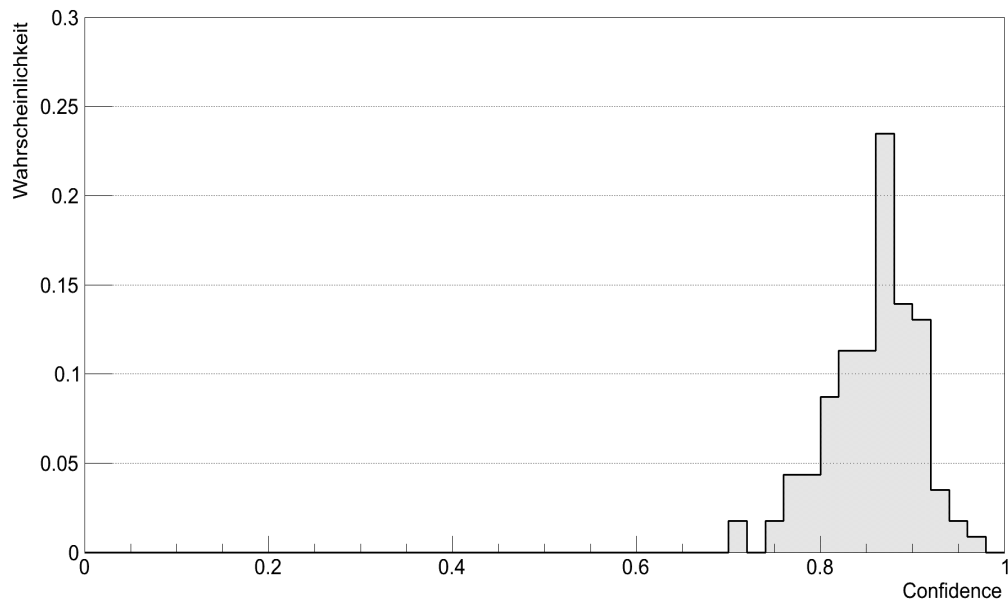
#### **Erweiterung der Bootstrap-Validierung**

Im idealen Fall steht für ein gegebenes Klassifikationsproblem eine ausreichend große Menge an MC zur Verfügung. In der Realität ist die Anzahl der MC stark beschränkt. Dies führt dazu, dass einige Bereiche des durch die Attribute aufgespannten Phasenraums nur mit einer sehr geringen Anzahl an Ereignissen besetzt sind. Aussagen über diese Bereiche sind daher nur auf Grundlage einer sehr kleinen Menge an Ereignissen möglich. Genau in diesem Bereich kann das Bootstrapping eine Möglichkeit bieten, valide Aussagen über spärlich besetzte Bereiche des Phasenraums zu machen. Die Bedingung für die Anwenden dieser Methode ist, dass der Klassifikationsalgorithmus

---

<sup>4</sup>Bootstrapping bezeichnet Ziehen mit Zurücklegen, einzelne Ereignisse können also mehrfach gezogen werden.





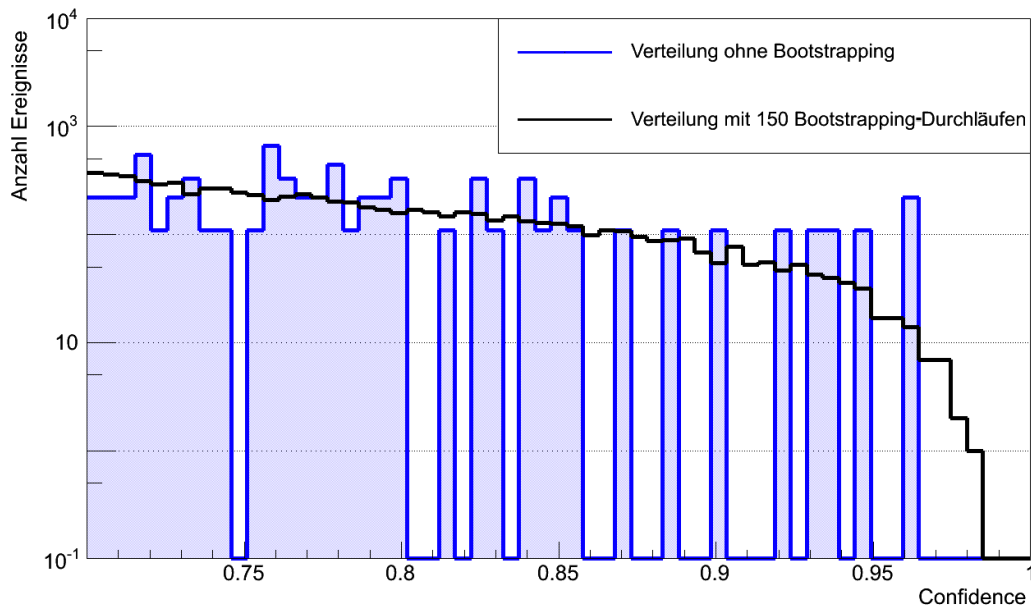
**Abbildung 4.3.:**

Wahrscheinlichkeitsverteilung für ein Ereignis, bestimmt durch 100 Bootstrapping-Durchläufe. Das Ereignis wird abhängig von den verschiedenen Modellen unterschiedlichen Confidencen zugeordnet.

dem Ereignis eine als Wahrscheinlichkeit interpretierbare Größe zuordnet, mit der er ein Ereignis einer Klasse zuordnet. Im Folgenden wird diese Größe als Confidence bezeichnet.

In jedem Durchgang des Bootstrapping werden andere Ereignisse zum Lernen der Methode genutzt. Daher sind die entstehenden, von den statistischen Fluktuationen abhängigen Modelle unterschiedlich. Jedes Modell ist jedoch gleichwertig. Daher kann über die Modelle gemittelt werden. Dadurch ist es im Bootstrapping möglich, eine Schätzung der Kenngrößen vorzunehmen. Es ist jedoch ebenfalls möglich, über einzelne Ereignisse zu mitteln. Dadurch kann die in Abbildung 4.3 dargestellte Wahrscheinlichkeitsdichte eines einzelnen Ereignisses in der Confidence geschätzt werden.

Alle weiteren Operationen werden dann nicht mehr auf einzelnen Ereignissen ausgeführt, sondern auf Wahrscheinlichkeitsdichten. In Abbildung 4.4 wird die durch Bootstrapping erzeugte Confidence-Verteilung mit der Verteilung der gleichen Daten ohne diese Methode verglichen. Beide Verteilungen sind Schätzungen der wahren Verteilung. Die Verteilung, die durch Bootstrapping gewonnen wurde, besitzt jedoch die wesentlich geringere Varianz. Für eine Analyse bedeutet dies, dass, wenn in einem Datensatz zwei Klassen voneinander getrennt werden sollen, nicht ab einem



**Abbildung 4.4.:**

Verteilung der Confidence nach Bootstrapping. Die beiden Verteilungen sind Schätzungen für die wahre Verteilung. Die Verteilungsfunktion auf Grundlage von Bootstrapping ist jedoch wesentlich genauer.

Schwellwert alle Ereignisse verworfen werden müssen, sondern nur noch die Teile der Wahrscheinlichkeitsdichte, die unterhalb des Schwellwerts liegen. Gerade die Ausläufer der Wahrscheinlichkeitsdichte geben dann einen Beitrag zu den Bereichen des Phasenraums, die bei einmaliger Anwendung einer Methode auf MC nur sehr spärlich besetzt sind. Da die Information über einzelne Ereignisse vorhanden ist, kann dadurch die Verteilungsfunktion nach einem Schnitt in allen Attributen besser geschätzt werden.

Die effektive Anzahl an MC kann durch dieses neue Verfahren um einen Faktor, der proportional zur Anzahl der Bootstrappingdurchläufe ist, erhöht werden. In den meisten Experimenten ist die zur Verfügung stehende Anzahl an MC durch die Computing-Ressourcen begrenzt. Dieses Verfahren benötigt jedoch keine zusätzlichen MC, sondern maximiert nur deren Nutzung bei einem vergleichsweise geringen Einsatz von Computing-Ressourcen.

Normalerweise wird zur Trennung zwischen Signal und Untergrund ein Schnitt bei einem festen Schwellwert durchgeführt, so dass beispielsweise alle Ereignisse, die eine Confidence größer als 0,5 aufweisen, als Signal klassifiziert werden. Durch das im bisherigen Verlauf dieses Abschnittes beschriebene Verfahren steht nun jedoch nicht

mehr nur eine Schätzung der Confidence für ein Ereignis zur Verfügung, es ist vielmehr die Wahrscheinlichkeitsdichte für jedes Ereignis vorhanden. Diese Wahrscheinlichkeitsdichte bietet mehr Möglichkeiten, eine Trennung zwischen den Klassen herzustellen, da nun auch die Momente und andere Größen die nur für Verteilungen definiert sind, für Schnitte genutzt werden können.

#### 4.3.2. Der Entscheidungsbaum-Algorithmus

Ein Entscheidungsbaum ist ein Algorithmus zur Klassifikation von Daten. Dabei ist zu beachten, dass es verschiedene Implementierungen gibt, so dass nur bedingt von *dem* Entscheidungsbaum-Algorithmus gesprochen werden kann. Der grundlegende Aufbau ist jedoch immer ähnlich. Im Folgenden wird daher beispielhaft die Funktionsweise eines gierigen<sup>5</sup> Entscheidungsbaumes anhand eines Zweiklassenproblems erklärt. [Qui86, BFO<sup>+</sup>83]

Das Ziel ist es, mittels MC einen Entscheidungsbaum aufzubauen, der in der Lage ist, nicht gelabelte Daten zu klassifizieren. Um dies zu erreichen, muss der Algorithmus zunächst das Attribut identifizieren, das eine möglichst gute Trennung des Datensatzes erlaubt. Das Attribut sollte also einen maximalen Informationsgehalt bezüglich des Labels aufweisen. Durch die Nutzung eines für das Problem geeigneten Korrelationsmaßes kann dieses Attribut leicht identifiziert werden. Für ein Label mit zwei Ausprägungen bietet sich hier die punktbiserielle Korrelation an. Die punktbiserielle Korrelation  $r_{pb}$  eines Attributes  $X$  bezüglich des Labels  $L$  mit den zwei Klassen 1 und 0 ist gegeben durch Gleichung 4.7.

$$r_{pb} = \frac{(\bar{X}_0 - \bar{X}_1) \sqrt{\frac{n_0 n_1}{n_0 + n_1}}}{\sigma_X} \quad (4.7)$$

$\bar{X}_i$ : Mittelwert des Attributs  $X$  unter der Bedingung der Ausprägung  $i$   
 $n_i$ : Anzahl der Ereignisse der Ausprägung  $i$

Für dieses Attribut sucht der Algorithmus nach einem Schnitt, der in den beiden daraus resultierenden Untermengen eine maximale Ungleichverteilung der Klassen erzeugt. Ein Maß für die Ungleichverteilung ist der in Gleichung 4.8 beschriebene Gini-Koeffizient  $G$ . [Bis07, Gin12, T<sup>+</sup>54, BFO<sup>+</sup>83]

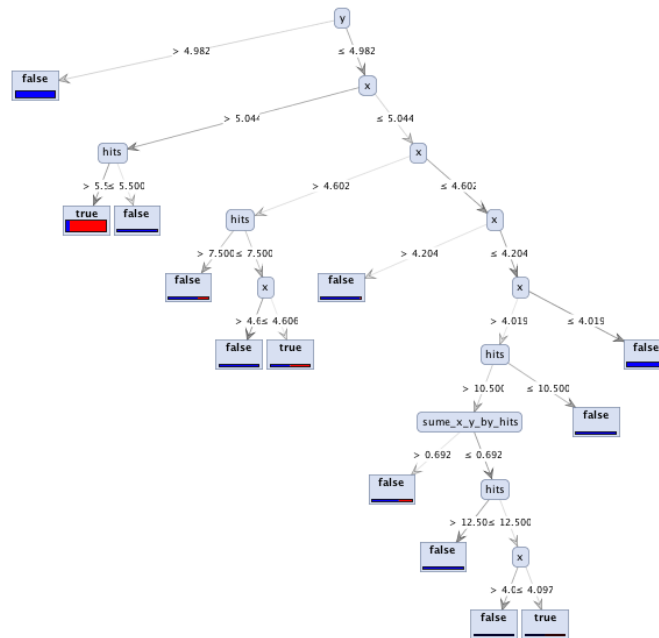
$$G = \sum_i p_i^2 \quad (4.8)$$

$p_i$ : Prozentualer Anteil der Ereignisse der Klasse  $i$  an der Gesamtmenge

<sup>5</sup>Wird im Englischen als greedy bezeichnet.

Daraus ergibt sich als Optimierungskriterium für den Schnitt im Attribut  $X$  ein maximaler Zuwachs des Gini-Koeffizienten in den zwei entstehenden Untermengen. Der erste auf diese Weise entstehende Schnitt wird als Wurzelknoten des Entscheidungsbaums bezeichnet. Die für den Wurzelknoten beschriebene Vorgehensweise wird auf den sich durch den Schnitt ergebenden Untermengen solange wiederholt, bis in den entstehenden Untermengen nur noch Ereignisse einer Klasse enthalten sind. Dieser letzte Knoten wird als Blatt bezeichnet. Die Summe aller Knoten und Blätter wird als Entscheidungsbaum bezeichnet, ein Beispiel findet sich in Abbildung 4.5.

Das Maximierungskriterium einer maximalen Ungleichverteilung der Klassen, also von möglichst reinen Blättern, für einen Entscheidungsbaum kann leicht dazu führen, dass dieser übertrainiert wird. Das bedeutet, dass die gewählten Schnitte so diffizil sind, dass sie nur noch auf Grundlage von statistischen Fluktuationen gewählt werden. Die auf Grundlage von Rauschen gewählten Schnitte zeigen in der Übertragung auf andere Daten jedoch keine Effizienz. Dieses Verhalten ist daher nicht wünschenswert. In einer Kreuzvalidierung zeigt sich Überanpassung durch einen vergrößerten Fehler deutlich. Für Entscheidungsbäume gibt es zwei einfache Methoden, mit denen Überanpassung



**Abbildung 4.5.:** Entscheidungsbaum zur Klassifikation von Daten in die zwei Klassen true und false. Das Verhältnis der Klassen sowie die absolute Anzahl der Ereignisse wird in jedem Blatt dargestellt. Die Höhe des Balken gibt die absolute Anzahl der Ereignisse an. Das Verhältnis der Klassen wird durch die Farben Rot (true) und Blau (false) gekennzeichnet.

vermieden werden kann. Die erste Möglichkeit ist, die maximale Anzahl von Knoten zu beschränken. Der Baum kann sich dadurch nicht mehr so weit verästeln, die angewandten Schritte basieren mit einer hohen Wahrscheinlichkeit auf einer ausreichend großen Anzahl an Ereignissen, so dass es nicht zu einer Überanpassung kommt. Durch diese Vorgehensweise kann aber nicht ausgeschlossen werden,

dass es zu Überanpassung kommt. Es besteht zusätzlich die Gefahr, dass durch die beschränkte Tiefe des Baums keine optimale Trennung der Klassen erreicht wird. [JDM00, KM98, OH95, BFO<sup>+</sup>83]

Eine weitere Möglichkeit, Überanpassung entgegenzuwirken ohne die Klassifizierungsstärke einzuschränken, ist das Zurückschneiden<sup>6</sup> von Entscheidungsbäumen. Dabei werden Knoten entfernt, die nur einen sehr geringen Informationsgewinn liefern. Dadurch wird der Entscheidungsbaum auf die wesentlichen Schnitte reduziert, womit das Risiko von Überanpassung verringert wird. Die einfachste Form des Zurückschneidens ist das *Reduced Error Pruning*. Bei dieser Form des Zurückschneidens wird bei den Blättern des Entscheidungsbaums beginnend jeder Knoten durch ein Blatt der häufigsten Klasse in diesem Knoten ersetzt, wenn sich dadurch die Accuracy des Entscheidungsbaumes nicht ändert. Diese Vorgehensweise ist einfach und kann daher sehr schnell umgesetzt werden, sie führt trotzdem zu guten Ergebnissen. Jedoch kann auch mit komplexeren Verfahren die Gefahr von Overtraining nicht ausgeschlossen werden. [JDM00, KM98, OH95]

### 4.3.3. Random Forest

Eine Möglichkeit, die Nachteile des Entscheidungsbaums in Bezug auf Trennkraft und Robustheit auszugleichen ist der Random Forest. Der Random Forest ist ein Ensemble von  $N$  unterschiedlichen Entscheidungsbäumen. Durch Mittlung über die  $N$  Bäume ergibt sich für jedes klassifizierte Ereignis die Confidence  $S_x$ , beschrieben in Gleichung 4.9, welche als Wahrscheinlichkeit interpretiert werden kann, mit der ein Ereignis der Klasse  $x$  zugeordnet wird.

$$S_x = \frac{N_x}{N} \quad (4.9)$$

$N_x$ : Anzahl der Entscheidungsbäume, die das Ereignis der Klasse  $x$  zuordnen  
 $N$ : Gesamte Anzahl der Entscheidungsbäume

Für den Fall eines Zweiklassenproblems ergeben sich zwei Confidence-Werte. Der erste Wert ist Confidence(Signal), der zweite ist Confidence(Untergrund). Speziell für ein Zweiklassenproblem gilt

$$\text{Confidence(Signal)} = 1 - \text{Confidence(Untergrund)}. \quad (4.10)$$

<sup>6</sup>Wird im Englischen als pruning bezeichnet.

Daher bietet es sich an, für ein Zweiklassenproblem jeweils nur eine der beiden Kenngrößen zu nennen. Für ein Mehrklassenproblem ergibt sich entsprechend für jede Klasse eine Confidence.

Die einzelnen Entscheidungsbäume eines Random Forest werden sehr ähnlich zu den in Unterabschnitt 4.3.2 beschriebenen Vorgehensweisen aufgebaut. Aus der Gesamtheit der  $L$  für das Training zur Verfügung stehenden Ereignisse wird zufällig eine Anzahl von Ereignissen per Bootstrapping gezogen. Auf der Grundlage dieser  $l$  Ereignisse wird ein Entscheidungsbaum bis zur maximalen Verästelung aufgebaut und nicht zurückgeschnitten. An jedem Knoten stehen dabei jedoch nur eine Anzahl  $m$  der insgesamt  $M$  Attribute zur Verfügung. Die Attribute werden dabei genau wie die Ereignisse zufällig mit Zurücklegen gezogen. Durch diese Vorgehensweise entstehen  $N$  unterschiedliche und stark übertrainierte Entscheidungsbäume, die als einzelne Klassifikatoren schlechte Ergebnisse liefern. In der Klassifikation wird jedes Ereignis von allen Entscheidungsbäumen im Random Forest klassifiziert. Dabei wird jedes Ereignis von jedem Baum immer einer Klasse zugeordnet. Durch die Mittlung über die  $N$  Entscheidungsbäume ergibt sich dann die nicht übertrainierte und robuste Klassifikation des Random Forest. [Bre01]

Eine sehr einfache Verbesserung des Random Forest gegenüber Entscheidungsbäumen ist der Übergang von den einfachen Mehrheitsentscheidungen in den Blättern der Entscheidungsbäume zu einer mittleren Wahrscheinlichkeit für jedes Ereignis. Bei der Mehrheitsentscheidung wird ein Ereignis der im Blatt häufigsten Klasse zugeordnet. Der Anteil anderer Klassen im Blatt wird für die weitere Entscheidung nicht beachtet. Die Verhältnisse der Klassen lassen sich jedoch als Wahrscheinlichkeit  $p_x$  für die Zugehörigkeit zu einer Klasse interpretieren. Über diese einzelnen Wahrscheinlichkeiten kann dann in Anlehnung an Gleichung 4.9 summiert werden. Daraus ergibt sich die in Gleichung 4.11 beschriebene mittlere Wahrscheinlichkeit  $\bar{p}_x$ , die beschreibt, mit welcher Wahrscheinlichkeit der Random Forest ein Ereignis einer Klasse  $x$  zuordnet.

$$\bar{p}_x = \frac{1}{N} \sum_i p_{i,x} \quad (4.11)$$

$p_{i,x}$ : Die Wahrscheinlichkeit, mit der der  $i$ -te Entscheidungsbaum das Ereignis der Klasse  $x$  zuordnet  
 $N$ : Anzahl der Entscheidungsbäume

Es lässt sich zeigen, dass durch diese Vorgehensweise eine Verbesserung der Klassifikation erreicht werden kann. [BZM07]

Um eine Klassifikation vorzunehmen, muss für die  $S_x$  bzw.  $p_{i,x}$  ein Schwellwert festgelegt werden, ab dem ein Ereignis einer Klasse zugewiesen wird. Für die Ermittlung

dieses Schwellwerts gibt es keine allgemein gültige Vorgehensweise. Generell ist der Schwellwert stark von dem gegebenen Problem abhängig und sollte entsprechend angepasst werden.

Eine wichtige Größe, die dabei helfen kann, die Bedeutung der einzelnen genutzten Attribute für einen Random Forest zu beurteilen, ist die Tree Importance. In jedem Entscheidungsbaum eines Random Forest werden unterschiedliche Schnitte auf unterschiedlichen Attributen durchgeführt. Wird der Zuwachs des Gini-Index summiert, den jedes Attribut im Mittel für einen Baum bringt, ergibt sich aus diesem der mittlere Zuwachs des Gini-Index. Dieser mittlere Zuwachs wird als Tree Importance bezeichnet. [MKM<sup>+</sup>09, Bre01]

#### 4.3.4. Boosting

Die dem Boosting zugrunde liegende Idee ist, aus mehreren schwachen Lernalgorithmen einen zusammengesetzten starken Lernalgorithmus zu generieren. Es zeigt sich also eine Ähnlichkeit zum Random Forest, beschrieben in Unterabschnitt 4.3.3. Boosting ist im Gegensatz dazu nicht an einen Algorithmus gekoppelt, sondern kann auf jeden Algorithmus angewandt werden. Zu Beginn wird allen Ereignissen im Trainingsset ein Gewicht von 1 zugewiesen. Im nächsten Schritt wird der Algorithmus auf dem Trainingsset trainiert. Anschließend werden die Gewichte von falsch klassifizierten Ereignissen um einen Faktor  $g$  erhöht. Alle anderen Gewichte bleiben unverändert. Durch die Veränderung der Gewichte vor jedem neuen Durchlauf des Algorithmus entstehen jeweils andere Modelle, die auf unterschiedlichen Trainingssets gelernt wurden. Die einzelnen Modelle unterscheiden sich daher in ihrer Klassifikation. Diese Vorgehensweise wird solange wiederholt, bis eine vorher festgelegte Anzahl an Durchläufen erreicht wurde. Die Klassifikation erfolgt nach Ausbildung aller Algorithmen in der Mittlung über ihre Entscheidungen, genau wie beim Random Forest. [FS97]

## 4.4. Entfaltung

Myon-Neutrinos können außerhalb des IceCube-Detektors interagieren. In dieser Interaktion verteilt sich die Energie des Neutrinos auf eine hadronische Kaskade und ein Myon. Die genaue Aufteilung der Energie zwischen Myonen und Kaskaden ist unbekannt. Das entstandene Myon propagiert anschließend bis es den IceCube-Detektor erreicht. Auf der zurückgelegten Strecke kann das Myon Energieverluste erleiden, beispielsweise durch Bremsstrahlung, Paarproduktion und Ionisation. Da diese Energieverluste außerhalb des Detektors stattfinden, sind sie für den Detektor nicht messbar. Erreicht das Myon den Detektor können energieabhängige Attribute, wie der Energieverlust pro Meter  $\frac{dE}{dx}$ , die deponierte Ladung in den DOMs und die Anzahl der angesprochenen DOMs rekonstruiert beziehungsweise ausgelesen werden. Die Ausprägung dieser Attribute hängt von der Energie des Myons ab. Diese wiederum hängt von der Energie des Myon-Neutrinos ab. Die Abhängigkeit ist jedoch stark verschmiert und teilweise sehr schwach, auf Grund der oben beschriebenen unbekannteren Energieverluste. Für ein einzelnes Myon-Neutrino ist es daher unmöglich seine genaue Energie aus der Beobachtung im Detektor zu rekonstruieren. Ein solches Problem wird daher als schlecht gestelltes inverses Problem bezeichnet. Es ist jedoch möglich, das ebenfalls schlecht gestellte inverse Problem der Rekonstruktion des Energiespektrums der Myon-Neutrinos zu lösen. Im Folgenden wird daher anhand der Rekonstruktion des Energiespektrums der Myon-Neutrinos ein Lösungsweg für die gegebene Problemstellung skizziert. [A<sup>+</sup>14a, A<sup>+</sup>13e, KFS<sup>+</sup>13, MDK<sup>+</sup>13, BL98]

Ein schlecht gestelltes inverses Problem kann durch ein Fredholm Integral 1. Ordnung beschrieben werden:

$$g(y) = \int_a^b A(y, x) f(x) dx + b(y). \quad (4.12)$$

$g(y)$ : Gemessene Verteilungsfunktion  
 $f(x)$ : Gesuchte Verteilungsfunktion  
 $A(y, x)$ : Kern des Integrals  
 $b(y)$ : Verteilungsfunktion des Untergrunds

Durch die Technik der Entfaltung ist es möglich, aus den energieabhängigen Attributen das wahre Energiespektrum zu rekonstruieren. Das Ziel ist es, die Faltung der gesuchten Verteilungsfunktion mit dem Kern des Integrals aufzulösen. Durch eine ausrei-



chend gute Separation des Untergrundes vereinfacht sich das Integral zu: [MDK<sup>+</sup>13, BL98]

$$g(y) = \int_a^b A(y, x) f(x) dx. \quad (4.13)$$

Das Fredholm Integral kann durch eine Diskretisierung der Gleichung 4.13 in Gleichung 4.14 überführt werden.

$$\vec{g} = A\vec{a} \quad (4.14)$$

*A*: Antwortmatrix  
*g*: Histogramm der gemessenen Verteilung  
*a*: Histogramm der gesuchten Verteilung

Die einfachste Möglichkeit Gleichung 4.14 zu lösen, ist die in Gleichung 4.15 beschriebene Umstellung nach  $\vec{a}$  unter Verwendung der inversen Matrix von  $A$ :

$$\vec{a} = A^{-1}\vec{g}. \quad (4.15)$$

Um diese Gleichung lösen zu können, muss die Matrix  $A$  bekannt sein. Diese kann im Fall von IceCube durch die Verwendung von MC berechnet werden, da für MC sowohl die Attribute, als auch die wahre Energie für jedes Teilchen bekannt ist. Aus den Eigenschaften der Matrix ergeben sich für nicht triviale  $A$  Probleme, welche konzeptionell bei jedem Versuch der Lösung eines schlecht gestellten inversen Problems auftreten. Diese werden im Folgenden besprochen. [BL98, MDK<sup>+</sup>13, Mil12]

Für einen idealen Detektor, der alle Energieverluste genau messen kann, ist die Invertierung von  $A$  trivial, da in diesem Fall  $A$ , wie in Gleichung 4.16 beschrieben, der Einheitsmatrix entspricht, welche durch Invertieren in sich selbst übergeht. [BL98, MDK<sup>+</sup>13, Mil12]

$$A = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{bmatrix} \Rightarrow A^{-1} = A \quad (4.16)$$

*A*: Matrix für einen idealen Detektor

Ebenfalls trivial ist eine begrenzte Akzeptanz. Dies bedeutet, dass der Detektor abhängig von der Energie Ereignisse nur mit einer unterschiedlichen Wahrscheinlichkeit detektieren kann. Wie in Gleichung 4.17 zu erkennen ist, sind die Hauptdiagonalele-

#### 4. Data-Mining

---

mente kleiner als Eins, was der begrenzten Akzeptanz für unterschiedliche Energien entspricht. Solange die Akzeptanz bekannt ist und nicht gegen Null geht, ergeben sich dabei keine Probleme. Physikalisch betrachtet, können Myon-Neutrinos unterhalb einer minimalen Energie kaum noch detektiert werden. Wird die Messung innerhalb sinnvoller energetischer Grenzen durchgeführt, ergeben sich durch die Akzeptanz keine Probleme. [BL98, MDK<sup>+</sup>13, A<sup>+</sup>09]

$$A = \begin{bmatrix} 0,7 & 0 & \dots & 0 \\ 0 & 0,9 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 0,97 \end{bmatrix} \Rightarrow A^{-1} = \begin{bmatrix} 1,43 & 0 & \dots & 0 \\ 0 & 1,11 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 1,03 \end{bmatrix} \quad (4.17)$$

A: Matrix für einen idealen Detektor  
mit begrenzter Akzeptanz

Auf Grund der begrenzten Auflösung und der Verschmierung bei der Messung der Energie ergeben sich wie in Gleichung 4.18 beschrieben von Null verschiedene, sehr kleine Nebendiagonalelemente. Diese Nebendiagonalelemente besitzen auf Grund der begrenzten Menge an MC, die zum Bestimmen der Matrix zur Verfügung stehen, einen sehr großen relativen Fehler. Wesentlich wichtiger ist jedoch, dass die Matrix singular oder zumindest annähernd singular ist. Dieser Umstand bedingt, dass die Gleichung nicht oder nur bedingt invertierbar ist. Daraus folgt, dass Gleichung 4.15 keine eindeutige Lösung besitzt. Neben der Lösung, die zum gesuchten Spektrum führt, ergeben sich Lösungen, die von unphysikalischen Oszillationen überlagert sind. Diese unphysikalischen Lösungen treten konzeptionell bei jedem Versuch auf ein schlecht gestelltes inverses Problem zu lösen. Es ist daher notwendig, eine Möglichkeit zu finden, das Problem der unerwünschten Oszillationen zu umgehen. [BL98, MDK<sup>+</sup>13]

$$A = \begin{bmatrix} 0,4 & 0,1 & \dots & 0,002 \\ 0,2 & 0,8 & \dots & 0,01 \\ \vdots & 0,1 & \ddots & \vdots \\ 0,001 & \dots & 0,1 & 0,92 \end{bmatrix} \quad (4.18)$$

A Matrix für einen Detektor  
mit begrenzter Akzeptanz  
und verschmierter Energiemessung

Eine Möglichkeit, die unerwünschten Oszillationen zu unterdrücken ist, Krümmungen in den Lösungen zu unterdrücken. Eine Möglichkeit dies zu erreichen, bietet die Singulärwertzerlegung der Matrix  $A$ :

$$A = UEV^*. \quad (4.19)$$

Durch die Singulärwertzerlegung wird  $A$  dargestellt, durch die unitäre  $m \times m$  Matrix  $U$  und die adjungierte der  $n \times n$  Matrix  $V$ , sowie die reelle  $m \times n$  Matrix:

$$E = \left[ \begin{array}{ccc|ccc} \epsilon_1 & & & \vdots & & \\ & \ddots & & \dots & 0 & \dots \\ & & \epsilon_i & \vdots & & \\ \hline & \vdots & & \vdots & & \\ \dots & 0 & \dots & \dots & 0 & \dots \\ & \vdots & & \vdots & & \end{array} \right]. \quad (4.20)$$

Die  $\epsilon_i$  sind die Singulärwerte der Matrix  $A$ . Sie sind von ihrer Größe her aufsteigend, so dass gilt:

$$\epsilon_1 \geq \epsilon_2 \geq \dots \geq \epsilon_i. \quad (4.21)$$

Die Singulärwerte beschreiben die Eigenschaften der Matrix  $A$ . Je kleiner der Betrag der Singulärwerte, desto stärker tragen sie zu den unerwünschten unphysikalischen Oszillationen bei. Der einfachste Ansatz wäre, diese ab einem bestimmten Schwellwert abzuschneiden, wodurch sich die Matrix  $A'$  ergibt. Die Eigenschaften von  $A$  und  $A'$  sind sehr ähnlich, jedoch lässt sich  $A'$  auf Grund der verworfenen Singulärwerte invertieren ohne die oben beschriebenen Oszillationen der Lösung. Diese Vorgehensweise wird als Regularisierung bezeichnet. Durch die Anzahl der nicht verworfenen Singulärwerte kann die Stärke der Regularisierung eingestellt werden. Durch jede Form der Regularisierung gehen Informationen verloren. Sie ist jedoch nötig, um ein schlecht gestelltes inverses Problem zu lösen. Es ist daher zu beachten, dass die Stärke der Regularisierung an das gegebene Problem angepasst werden muss, da sonst Eigenschaften des zu entfaltenden Spektrums durch diese Regularisierung unterdrückt werden können. Wie in [A<sup>+</sup>11a] gezeigt wird, ist es mit dem beschriebenen Verfahren möglich, das Energiespektrum der von IceCube gemessenen Myon-Neutrinos zu bestimmen. [A<sup>+</sup>11a, BMM08]

#### 4. Data-Mining

---

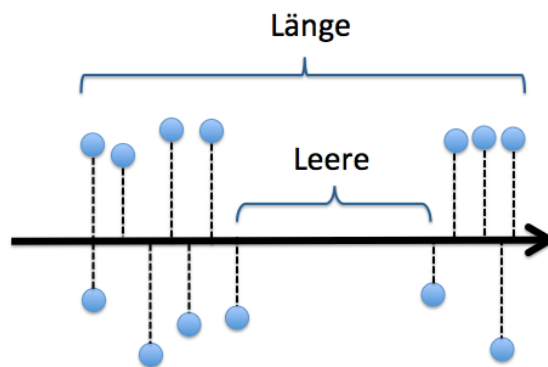
Eine Software, die solche schlecht gestellten inversen Probleme lösen kann ist TRUEE. Die Software basiert auf dem in [Blo96] beschriebenen RUN Algorithmus. Zur Entfaltung wird entgegen der oben beschriebenen Singulärwertzerlegung eine Maximum Likelihood-Methode genutzt. Die Regularisierung erfolgt nach dem in [Tik63] beschriebenen Verfahren, über die zweite Ableitung der gesuchten Lösung. Diese Vorgehensweise ist sehr ähnlich zur Einführung von Straftermen für kleine Singulärwerte. [MDK<sup>+</sup>13, Sch14, Blo96, Moo20, Tik63]

Die Software wird seit vielen Jahren im Bereich der Teilchenphysik eingesetzt. Eine genaue Beschreibung findet sich in [MDK<sup>+</sup>13] und [Sch14].

## 5. Ereignisselektion

Die vorliegende Analyse gliedert sich in zwei Teile. Der erste Teil ist die Generation eines qualitativ hochwertigen Datensatzes von Neutrinoereignissen. Diese Arbeiten werden innerhalb der RapidMiner Softwareumgebung durchgeführt. Der Schwerpunkt der methodischen Arbeiten entstammt dem ersten Bereich. Dies zeigt sich besonders an der für diese Analyse weiterentwickelten Bootstrapvalidierung. Diese ermöglicht es die Verteilung des Untergrundes bei sehr kleinen Ereigniszahlen besser zu beschreiben.

Insbesondere ist es durch diese Methode möglich, die Untergrundverteilung in den für die Entfaltung genutzten Attributen genauer beschreiben zu können. Durch diese genaue Beschreibung ist es möglich, wesentlich mehr Ereignisse in der Separation zuzulassen, ohne dabei die Reinheit zu verringern. Der zweite Teil der Analyse ist die Entfaltung des Energiespektrums der von IceCube gemessenen Myon-Neutrinos. Die in diesem Teil der Analyse verwendeten Methoden beruhen dabei auf der Arbeit von [Sch14]. [MWK<sup>+</sup>06]



**Abbildung 5.1.:**

Der schwarze Pfeil stellt die rekonstruierte Trajektorie dar. Die blauen Kreise sind stilisierte DOMs, deren Position auf die Trajektorie projiziert wird. Der Abstand zwischen der ersten und der letzten Projektion entlang der Trajektorie ist die Länge des Ereignisses. Der maximale Abstand zwischen zwei Projektionen wird als Leere bezeichnet.

## 5. Ereignisselektion

---

Die verwendeten Daten wurden im Zeitraum 13.05.2011 bis zum 16.05.2012 vom IceCube-Neutrinoobservatorium aufgezeichnet. Die Daten durchlaufen einige standardisierte Vorverarbeitungsschritte für den gewünschten Beobachtungskanal, bei denen die Daten verschiedenen Rekonstruktionsverfahren unterzogen werden. Dieses Aufbereitinungsniveau der Daten wird als Level3 bezeichnet und bildet den Ausgangspunkt der vorliegenden Arbeit.

Jedes Attribut wird in allen in Tabelle 5.1 beschriebenen Zeitfenstern berechnet. Das verwendete Zeitfenster wird durch die Verwendung eines Suffix gekennzeichnet. Ist kein Suffix angegeben, so wurde das Attribut für das Zeitfenster C berechnet.

Eine vollständige Auflistung und Beschreibung aller genutzten Attribute findet sich im Anhang in Abschnitt A.1. Ein Histogramm aller verwendeten Attribute auf Level3 findet sich im Anhang in Abschnitt A.2.

**Tabelle 5.1.:**

Bezeichnung der Zeitfenster, wie sie in der Ereignisrekonstruktion von IceCube genutzt werden. Je kleiner das Zeitfenster, desto eher erreichen die Photonen den DOM ohne auf ihrem Weg gestreut worden zu sein.

Bezeichnung	Zeitintervall
A	$[-15 \text{ ns}, +15 \text{ ns}]$
B	$[-15 \text{ ns}, +25 \text{ ns}]$
C	$[-15 \text{ ns}, +75 \text{ ns}]$
D	$[-15 \text{ ns}, +150 \text{ ns}]$
E	$[-15 \text{ ns}, +250 \text{ ns}]$

### 5.1. Ereignisvorselektion

Vor der Selektion der Ereignisse mit dem Random Forest bietet es sich an, Teile der Daten zu verwerfen, die für die Analyse nicht genutzt werden können. Grund für das Verwerfen kann dabei sein, dass dieser Teil der Daten von einem nicht abtrennbaren Untergrund verunreinigt ist, oder dass die Qualität der Ereignisse nicht ausreichend für eine weitere Analyse ist. In Anlehnung an das Level3 wird das so erzeugte Verarbeitungsniveau als Level4 bezeichnet. Ein weiterer Vorteil, der sich aus dem Verwerfen von nicht nutzbaren Ereignissen ergibt, ist eine deutliche Reduktion der Datenmenge, was die Weiterverarbeitung der Daten erleichtert.

Wie in Abschnitt 2.2 beschrieben, ist die Erde opak für atmosphärische Myonen. Die Erde kann daher als Abschirmung gegen diesen unerwünschten Untergrund eingesetzt werden. In Abbildung 5.2 ist  $\text{SplineMPE}_\ominus$  für Daten und MC auf Level3 dargestellt. Für Zenitwinkel kleiner  $90^\circ$  ist der Untergrund der atmosphärischen Myonen um fünf Größenordnungen häufiger vertreten als das gesuchte Signal der atmosphärischen und hochenergetischen extraterrestrischen Neutrinos. Für Zenitwinkel größer  $90^\circ$  ist der

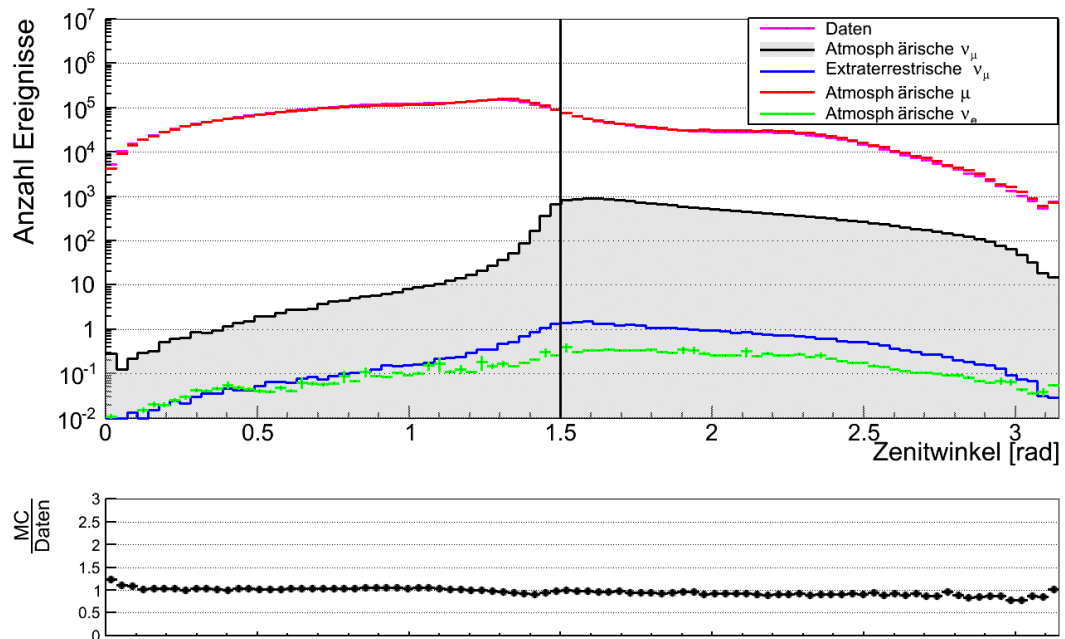


Abbildung 5.2.:

Verteilung des Attributs  $\text{SplineMPE}_\theta$  der Ereignisse auf Level3. Wie zu erkennen ist, sind für Winkel kleiner  $90^\circ$  die atmosphärischen Myonen auf Level3 extrem dominant.

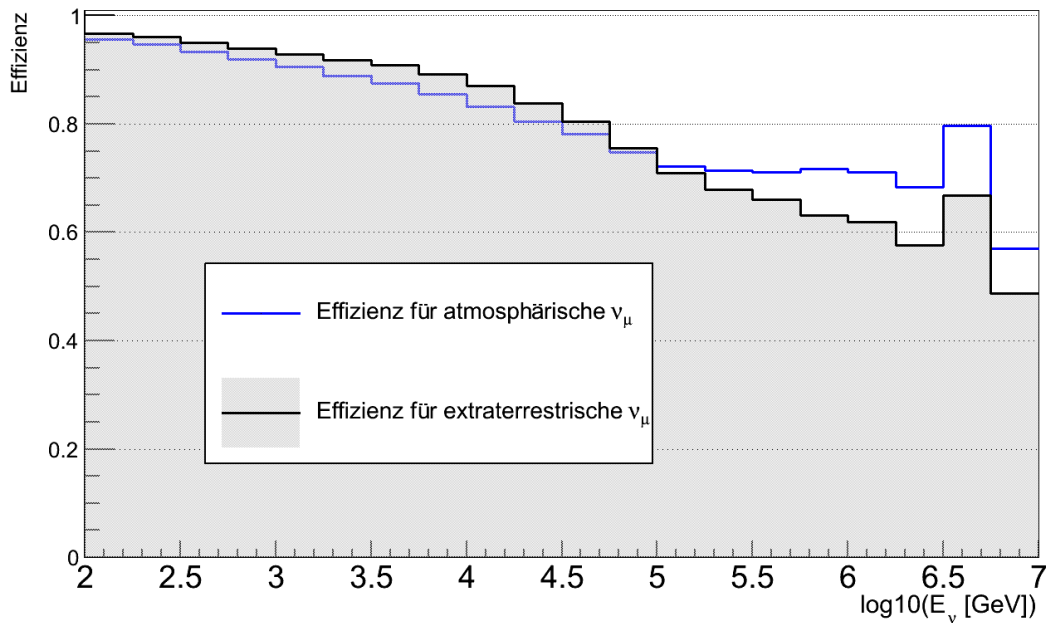
Untergrund ebenfalls dominant. Dies liegt daran, dass ein verhältnismäßig kleiner Anteil der atmosphärischen Myonen falsch rekonstruiert wird. In der Optimierung aller Vorselektionsschnitte ergibt sich ein Schnitt bei

$$\text{SplineMPE}_\theta \geq 86^\circ \quad (5.1)$$

als optimal. Dieser Wert ist ebenfalls vorteilhaft, da [Sch14] den gleichen Wert verwendet. Da der Fluss der atmosphärischen Myon-Neutrinos abhängig vom mittleren Zenitwinkel aller Ereignisse ist, lassen sich durch diese Wahl die Analysen leichter vergleichen.

Das Verwerfen von Ereignissen vom Südhimmel führt zu einem deutlichen Verlust von hochenergetischen Ereignissen. Dieser Verlust ist in Abbildung 5.3 zu erkennen. Der Effizienzverlust im Bereich der Hochenergie durch diesen Schnitt ist größer als alle weiteren Verluste addiert. Da die Erde als Abschirmung gegen atmosphärische Myonen genutzt wird, ist es jedoch nicht möglich auf diesen Schnitt zu verzichten und den vollen Winkelbereich zu betrachten.

## 5. Ereignisselektion



**Abbildung 5.3.:**

Die energieabhängige Effizienz des Schnittes in  $\text{SplineMPE}_\Theta$  für den Energiebereich von 10 GeV bis 10 PeV.

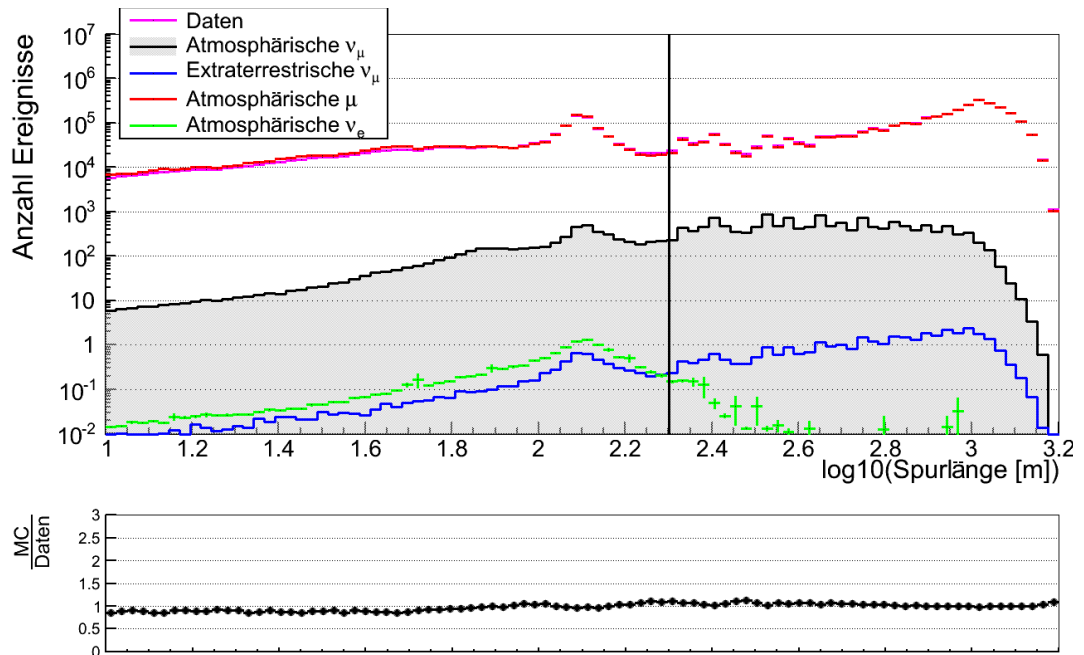
Die rekonstruierte Länge der Ereignisse, schematisch erklärt in Abbildung 5.1, ist in Abbildung 5.4 dargestellt. Der Modus der Verteilung der Elektron-Neutrinos und ein großer Teil der Verteilung der atmosphärischen Myonen liegt unter 200 m. Beides sind Typen von Ereignissen, die in der Entfaltung unerwünscht sind. Der durchschnittliche Abstand der IceCube-Strings ist 125 m. Ereignisse mit einer rekonstruierten Länge von unter 200 m können also maximal zwei Strings angesprochen haben. In den meisten Fällen handelt es sich um sehr niederenergetische Myonen. Diese Ereignisse weisen häufig eine geringe Qualität auf und lassen sich daher nur schwer rekonstruieren. Es bietet sich daher an, diesen Bereich der Daten zu verwerfen und nur Ereignisse zuzulassen, die

$$\text{Länge} \geq 200 \text{ m} \tag{5.2}$$

erfüllen. Durch diesen Schnitt werden neben dem unerwünschten Untergrund auch große Mengen an atmosphärischen Myon-Neutrinos verworfen.

Die verworfenen Ereignisse lassen sich in drei Klassen unterteilen. Die erste Klasse besteht aus sehr niederenergetischen Neutrinos. In Abbildung 5.5 liegt der maximale





**Abbildung 5.4.:**

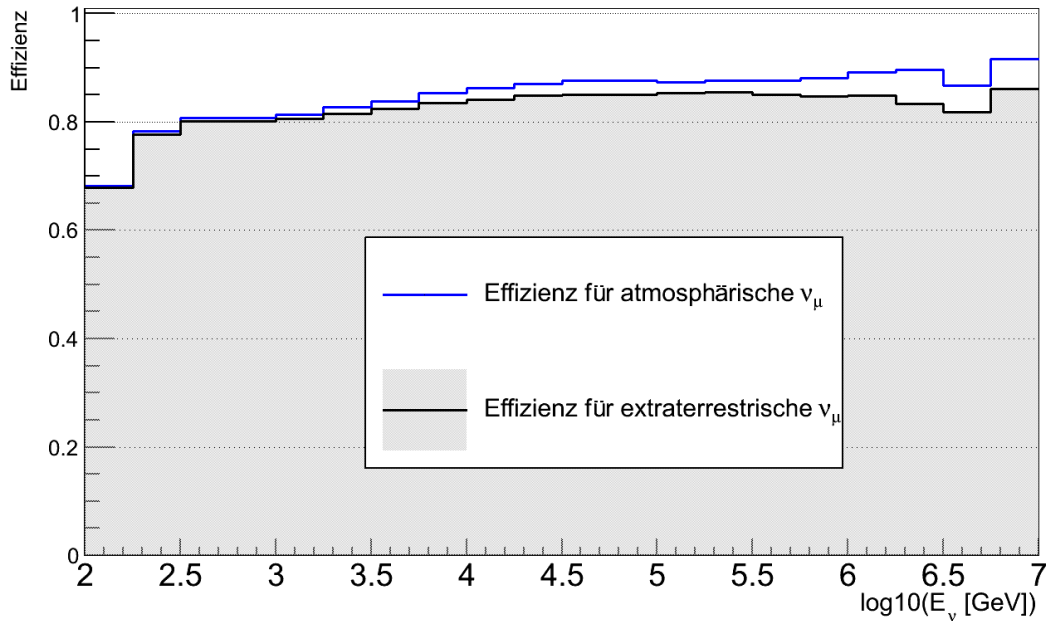
Verteilung des Attributs LängeC auf Level3. Wie erkennbar ist, befindet sich der Modus für Elektron-Neutrinos bei Längen kleiner als 200 m, was dem gewählten Schnitt entspricht.

Effizienzverlust durch diesen Schnitt bei 100 GeV. Da diese Ereignisse in großer Zahl vorhanden sind, ist ihr Verlust für die Entfaltung des Spektrums zu vernachlässigen. Die zweite Klasse von Ereignissen sind Neutrinos, die weit außerhalb des Detektors interagiert haben. Das resultierende Myon hat auf dem Weg bis zum IceCube-Detektor den überwiegenden Teil seiner Energie bereits verloren. Diese Ereignisse lassen sich konzeptionell nicht von niederenergetischen Neutrinos unterscheiden. In der Entfaltung tragen diese Ereignisse ebenfalls zu den höchstenergetischen Bins bei, jedoch in einem sehr geringen Maß. Daher ist das Verwerfen dieser Ereignisse ebenfalls zu vernachlässigen. Die dritte Klasse besteht aus Myon-Neutrinos, die über den neutralen Strom interagiert haben. Die in diesem Fall entstehende Kaskade ist nicht von Kaskaden die durch Elektron-Neutrinos induziert werden zu unterscheiden. Da die Ereignismuster von Elektronen und Myonen im IceCube-Detektor unterschiedlich sind, bietet es sich daher an, diese für eine Entfaltung nicht zu vermischen.

Die Schnitte in  $\text{SplineMPE}_\Theta$  und LängeC dienen dazu, schwierig zu separierenden Untergrund vor der eigentlichen Selektion durch einen Random Forest abzutrennen.

In einem weiteren Schritt werden Ereignisse mit einer geringen Qualität entfernt. Eine Möglichkeit dies zu erreichen ist, nur Ereignisse, bei denen die Myonenergieschätzung

## 5. Ereignisselektion



**Abbildung 5.5.:**

Effizienz des Schnittes in LängeC für den Energiebereich von 10 GeV bis 10 PeV. Bei den Effizienzverlusten im hochenergetischen Bereich handelt es sich zu einem überwiegenden Teil um Myon-Neutrinos, welche über den neutralen Strom Wechselwirkungen ausgeführt haben. Diese Wechselwirkungen erzeugen Kaskaden, welche eine wesentlich andere Signatur und Energieabhängigkeit als Myonen haben. Es ist daher wünschenswert, diese aus dem finalen Datensatz zu entfernen.

durch  $\text{TEBINS}_E$  und  $\text{TEDOMS}_E$  erfolgreich war, zu nutzen. Die beiden Energieschätzer betrachten nur DOMs in einem festen Radius  $r_{\text{max}}$  um die rekonstruierte Trajektorie. Die rekonstruierte Trajektorie wird in eine feste Anzahl an Bins unterteilt und die enthaltenen DOMs gezählt. Wird ein minimaler Anteil der Bins nicht mit mindestens einem DOM gefüllt, so gilt die Rekonstruktion als gescheitert. Die beiden Algorithmen unterscheiden sich durch die Anzahl der verwendeten Bins. Der Parameter  $\text{TEBINS}_E$  verwendet eine Anzahl von Bins abhängig von der Länge des Tracks, wohingegen in  $\text{TEDOMS}_E$  jeder DOM einem Bin entspricht. Eine Veranschaulichung der Vorgehensweise findet sich in Abbildung 5.8. Für Myonen ist zu erwarten, dass sie bedingt durch Paarproduktion entlang ihrer Trajektorie gleichmäßig Ladung in den DOMs deponieren. Ist die Trajektorie jedoch falsch rekonstruiert kommt es nicht zu einer gleichmäßigen Verteilung der Ladung entlang der rekonstruierten Flugbahn. Insbesondere im Fall von koinzidenten Ereignissen ist davon auszugehen, dass die Trajektorie falsch rekonstruiert ist. Da koinzidente Ereignisse ein vollkommen anderes Verhalten aufweisen als einzelne Myonen, bietet es sich an, solche Ereignisse vor der Entfaltung zu verwer-

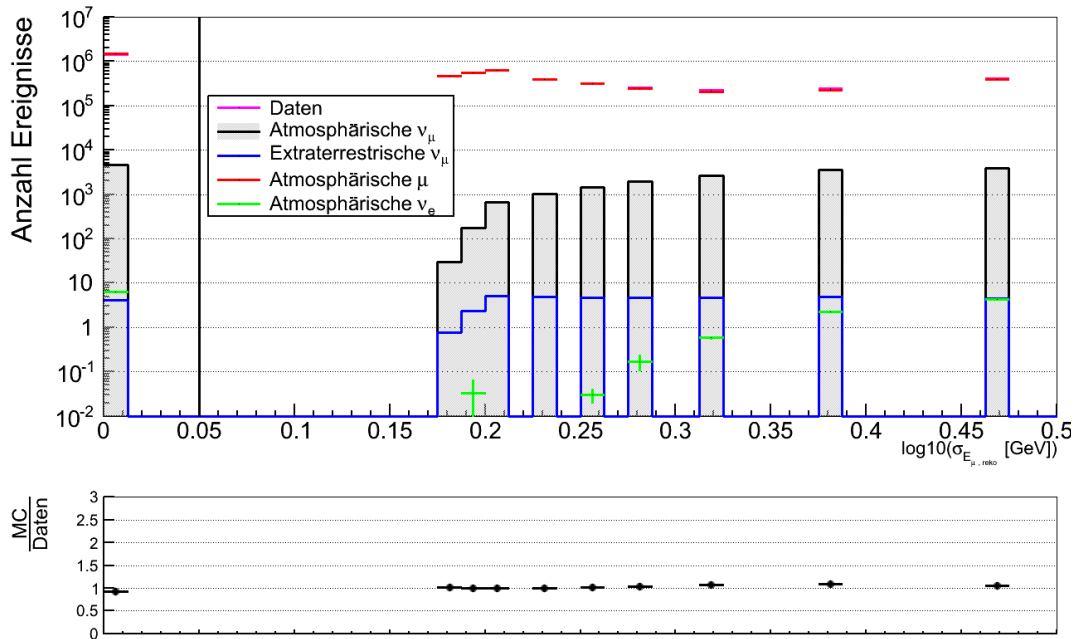


Abbildung 5.6.:

Energieauflösung im Zehnerlogarithmus der Myonenergie. Ereignisse, bei denen die Energierekonstruktion erfolgreich war, weisen eine Energieauflösung  $> 0$  auf. Der angewandte Schnitt ist ebenfalls in der Abbildung eingezeichnet. Durch den Schnitt wird der Modus der Untergrundverteilung verworfen.

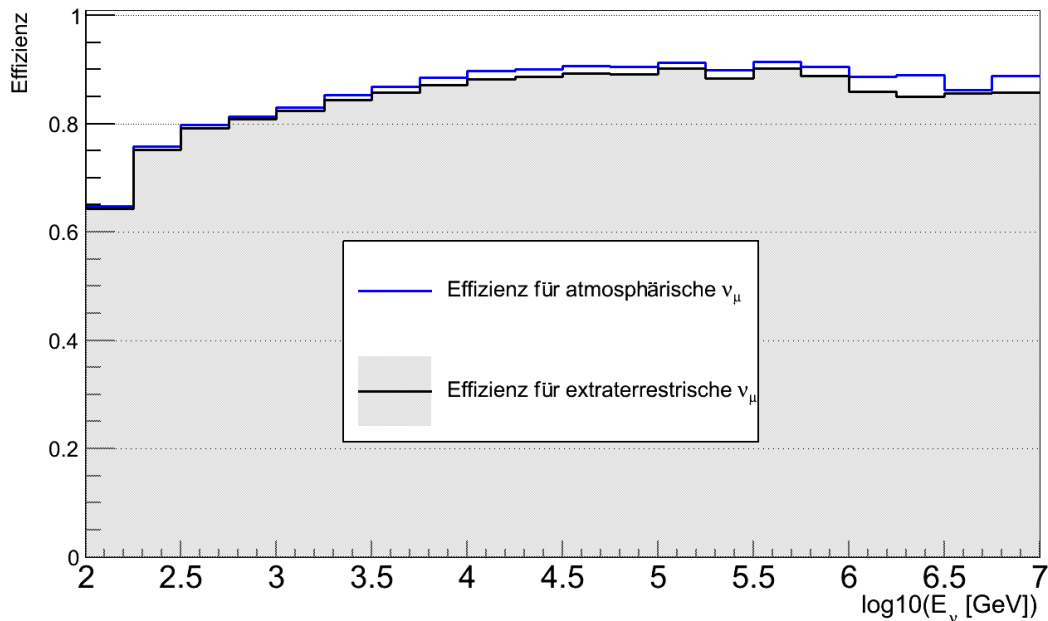
fen. Ist die Energieschätzung gescheitert, so zeigt sich dies durch den Wert Null für die Energieauflösung. Es werden daher nur Ereignisse zugelassen, welche die Bedingung

$$\text{TEBINS}_\sigma \neq 0 \wedge \text{TEDOMS}_\sigma \neq 0 \quad (5.3)$$

erfüllen. In Abbildung 5.6 ist  $\text{TEBINS}_\sigma$  dargestellt. Wie in Abbildung 5.7 erkennbar ist, werden durch diesen Qualitätsschnitt auch hochenergetische Ereignisse verworfen. Diese müssen jedoch, um zu einem Scheitern des Algorithmus zu führen, eine geringe Qualität aufweisen. Daher ist es vertretbar diese für die weitere Analyse nicht zu berücksichtigen.

Der zweite Schnittwert zur Sicherstellung der Ereignisqualität ist die maximale Länge von Lücken in der Trajektorie, genannt Leere. Eine schematische Erklärung des Attributs Leere findet sich in Abbildung 5.1. Die Verteilung des Attributs wird in der Abbildung 5.9(a) dargestellt. Das Attribut wird für einen Schnitt genutzt, da Ereignisse mit sehr großen Löchern entlang ihrer rekonstruierten Trajektorie durch koinzidente Myonen oder falsch rekonstruierte Ereignisse erzeugt werden können. Koinzidente Er-

## 5. Ereignisselektion



**Abbildung 5.7.:**

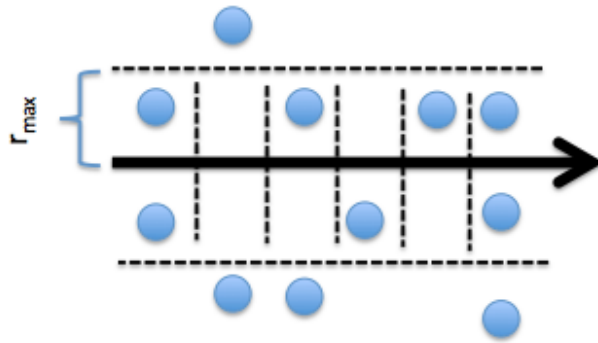
Die Effizienz des Schnitts auf eine erfolgreiche Energierekonstruktion für den Energiebereich von 10 GeV bis 10 PeV. Durch den Schnitt werden hauptsächlich niederenergetische Ereignisse entfernt.

eignisse sind ein unerwünschter Untergrund, der vor der Entfaltung entfernt werden sollte. Der Schnitt wurde so gewählt, dass nur Ereignisse zugelassen werden, die eine

$$\text{Leere} \leq 400 \text{ m} \quad (5.4)$$

aufweisen. Der Einfluss dieses Schnitts auf den erzeugten Datensatz ist gering. In Abbildung 5.9(b) wird deutlich, dass durch den Schnitt im Attribut Leere nur ein geringer Anteil der Ereignisse verworfen wird.

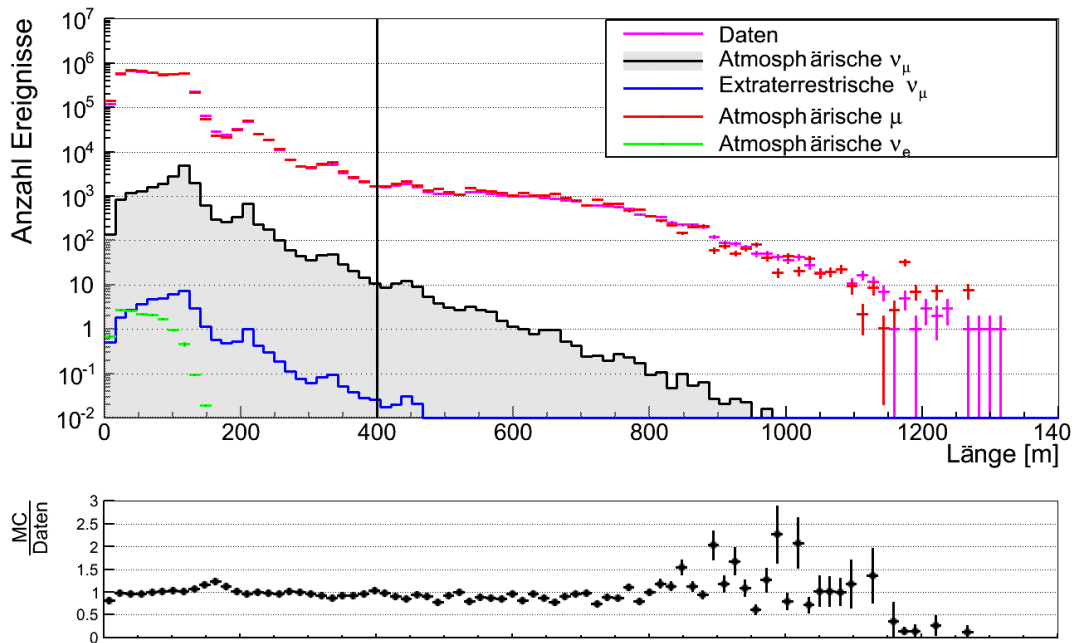
Es ist von besonderem Interesse, wie sich die Kombination von allen Schnitten auf die energieabhängige Effizienz, dargestellt in Abbildung 5.10(a), auswirkt. Die Abbildung zeigt, dass über den gesamten betrachteten Energiebereich ein deutlicher Effizienzverlust durch die Schnitte entsteht. Im Vergleich von Abbildung 5.10(a) mit Abbildung 5.10(b), ist zu erkennen, dass der größte Teil der Verluste im Bereich sehr hoher Energien aus dem Zenitschnitt stammt. Die Effizienzverluste im Bereich der niedrigen Energien sind zwar sehr groß, in diesem Bereich sind sehr viele Ereignisse vorhanden, so dass hier ein Verlust von 50 % der Ereignisse keinen Einfluss auf die Rekonstruktion des Spektrums hat.



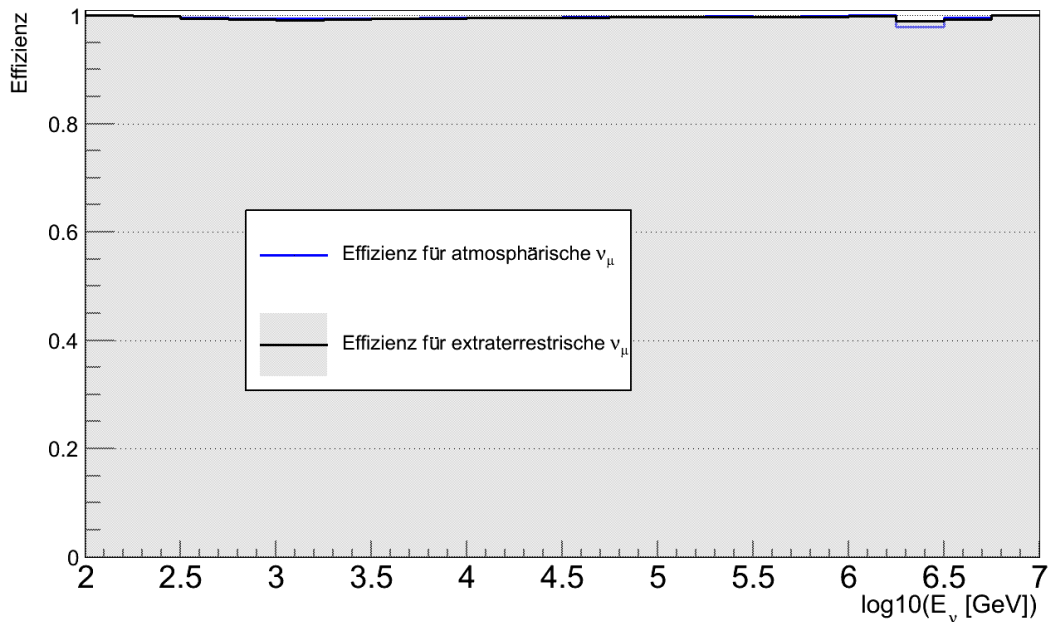
**Abbildung 5.8.:**

Die Parameter  $TEBINS_E$  und  $TEDOMS_E$  verwenden nur DOMs, die maximal eine Entfernung von  $r_{\max}$  zur rekonstruierten Trajektorie aufweisen. Die gestrichelten Linien skizzieren in vereinfachter Form das Unterteilen in Bins entlang der Trajektorie.

## 5. Ereignisselektion

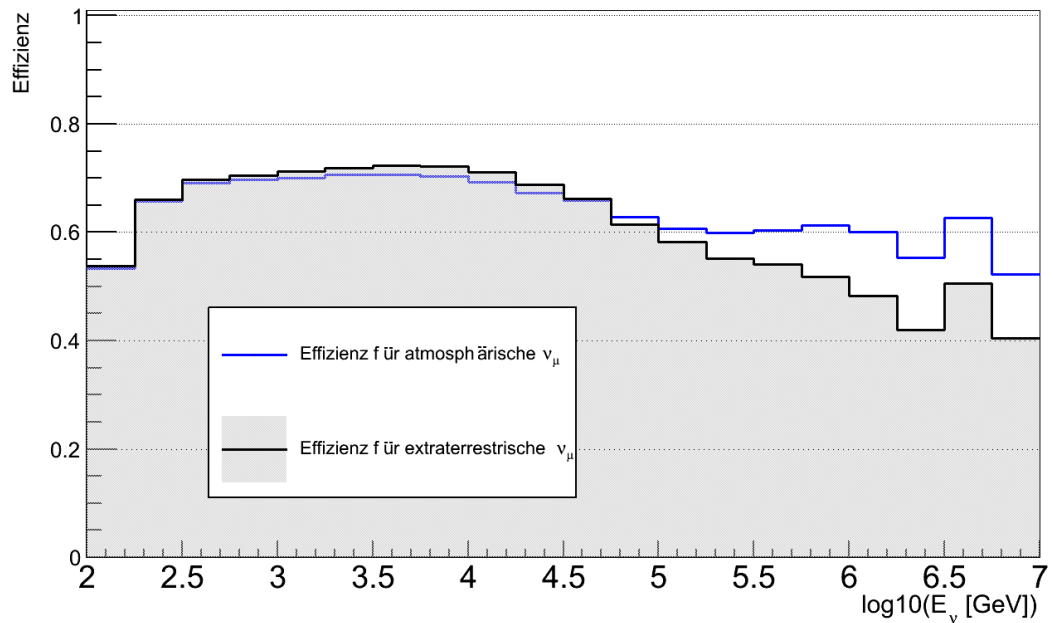


(a) Verteilung der IceCube-Daten und MC für das Attribut Leere. Darüber hinaus wurde der Schnitt bei einer Leere von 400 m eingezeichnet. In der Abbildung ist zu erkennen, dass das Attribut hinreichend gut simuliert wird, was eine Grundbedingung ist, um die Auswirkungen eines Schnittes zu verstehen.

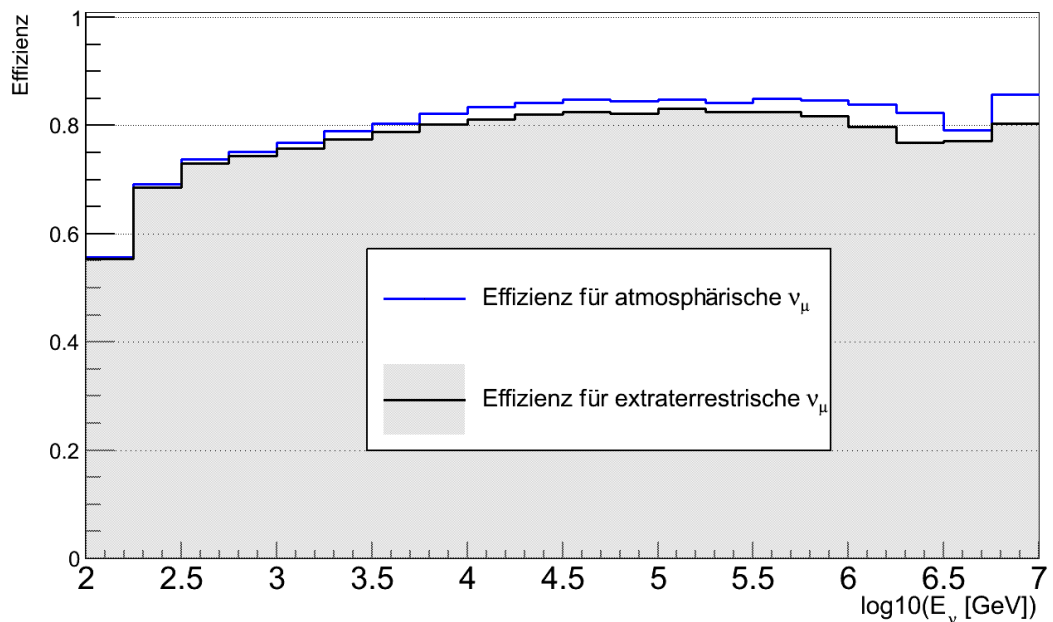


(b) Effizienz des Schnittes auf Leere für den Energiebereich von 10 GeV bis 10 PeV. Wie erkennbar ist, wirkt sich der Schnitt nur wenig auf die Effizienz der Analyse aus.

**Abbildung 5.9.:** Das Attribut Leere als Vorselektionsschnitt.



(a) Effizienz für die Summe aller Vorselektionsschnitte für den Energiebereich von 10 GeV bis 10 PeV. Es ist erkennbar, dass die Effizienz für ein Signal hochenergetischer extraterrestrischer Neutrinos im Vergleich zu atmosphärischen Myon-Neutrinos höher ist.



(b) Die Effizienz für die Summe aller Vorselektionsschnitte bis auf den Zenitschnitt für den Energiebereich von 10 GeV bis 10 PeV. Alle übrigen Schnitte resultieren in einen Effizienzverlust im Bereich der Hochenergie von etwa 20%.

**Abbildung 5.10.:**

Effizienz aller Vorselektionsschnitte für den Energiebereich von 10 GeV bis 10 PeV, sowie der Einfluss des Zenitschnittes.

### 5.2. Attributselektion

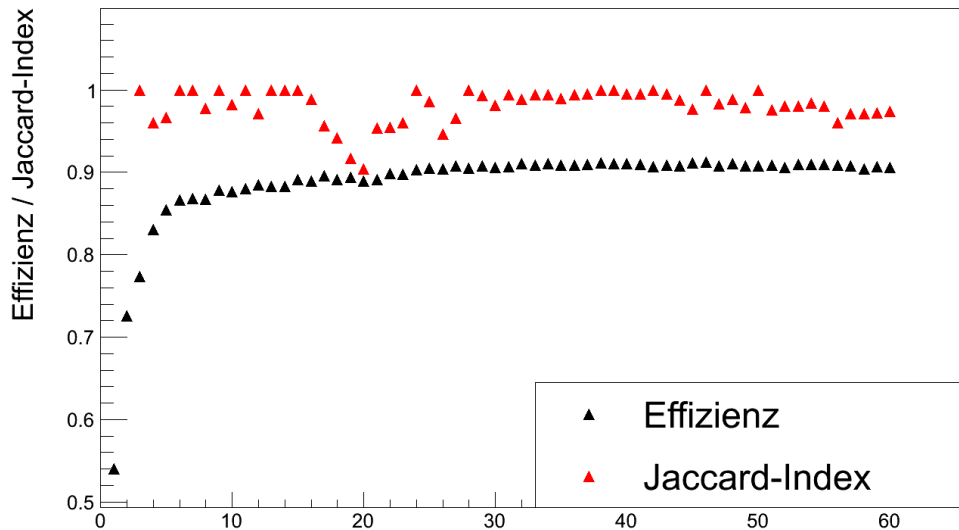
Die Selektion der Ereignisse für die weitere Nutzung wird mittels des in Unterabschnitt 4.2.1 beschriebenen MRMR Algorithmus durchgeführt. Da der Algorithmus nicht dafür ausgelegt ist schlecht simulierte Attribute zu erkennen, werden alle Attribute einer Prüfung unterzogen. Bei dieser wird auf eine gute Übereinstimmung zwischen Daten und MC getestet. Da schlecht simulierte Attribute sich stark auf die Qualität der Analyse auswirken können, ist dieser Schritt von besonderer Wichtigkeit. Bei dieser Überprüfung werden ebenfalls konstante Attribute und Attribute mit mehr als 30% fehlenden Werten entfernt. Bei allen anderen Attributen werden fehlende Werte durch 0 ersetzt. Durch diese Vorgehensweise reduziert sich die Anzahl der Attribute von über 1 000 auf 100.

Um aus den 100 Attributen die beste Repräsentation der Daten für die weitere Analyse zu wählen, wird zunächst eine MRMR Attributselektion für alle Anzahlen an zu wählenden Attributen im Intervall  $[1,60]$  durchgeführt. Das Ziel dieser Vorgehensweise ist eine Mindestanzahl an Attributen zu finden, ab der eine Attributmenge stabil ausgewählt werden kann. Die obere Grenze des Intervalls wird auf 60 festgelegt, da in den Analysen [Ruh13] und [Sch14] gezeigt wurde, dass etwa 20 bis 40 Attribute für die Beschreibung der IceCube-Daten ausreichend sind. Daher wird darauf verzichtet das Intervall unnötig zu erweitern.

Wie in Abbildung 5.11 zu erkennen ist, ergibt sich für alle Attributanzahlen ein Jaccard Index von  $\gtrsim 0,9$ . Daraus kann geschlossen werden, dass die Attributselektion bereits ab dem ersten gewählten Attribut stabil ist. In der Abbildung sind kleine Einbrüche in der Stabilität zu erkennen. Diese Einbrüche lassen sich durch Gruppen von Attributen mit ähnlicher Abhängigkeit zum Zielattribut erklären. Welches Attribut einer solchen Gruppe gewählt wird, kann leicht von kleinen Fluktuation abhängen. Da die Einbrüche der Stabilität gering sind, kann geschlossen werden, dass der Einfluss von Fluktuationen in den Daten auf die gewählten Attribute ebenfalls gering ist. Es ist daher anzunehmen, dass sich aus der Stabilität der Attributselektion kein direkter Hinweis auf die Anzahl der zu verwendenden Attribute ergibt.

Es ergibt sich daher keine, wie in Unterabschnitt 4.2.1 beschrieben, durch die Stabilität der Attributselektion gegebene untere Grenze an Attributen. Ein solches Verhalten zeigte sich deutlich in bisherigen IceCube-Analysen, wie in [Ruh13] und [Lot12] beschrieben, es ist jedoch nicht zwingend zu erwarten. Es gibt zwei Hauptfaktoren, die dieses Verhalten erklären können. Erstens ist davon auszugehen, dass auf Grund des sehr restriktiven Verwerfens von Attributen vor der eigentlichen Attributselektion viele





**Abbildung 5.11.:**

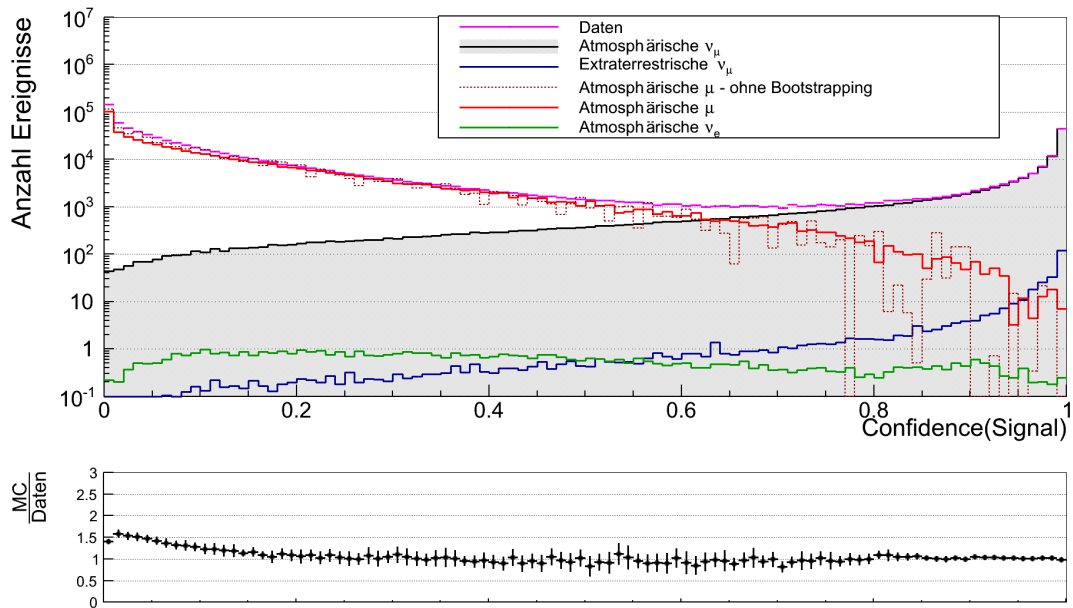
Stabilität und Effizienz der Attributselektion in Abhängigkeit von der Anzahl der gewählten Attribute. Die Stabilität zeigt für jede Anzahl an gewählten Attributen ein hohes Maß an Stabilität. Die Effizienz für etwa 30 Attribute eine Sättigung.

Attribute mit einer ähnlichen Abhängigkeit zum Zielattribut bereits verworfen wurden. Ebenfalls hat sich die Qualität der Daten durch einen homogeneren Detektor und eine bessere Beschreibung des Eises deutlich verbessert, was zu einer Verringerung von Uneindeutigkeiten in der Rangfolge der Abhängigkeiten zum Zielattribut geführt hat. Es ist wahrscheinlich, dass eine Mischung aus beiden Gründen das in Gleichung 4.5 beschriebene Stabilitätsverhalten erklärt.

Eine der Hauptforderungen an die Attributselektion ist eine maximale Stabilität, daher kann das beobachtete Verhalten als vorteilhaft angesehen werden. Für die Wahl einer Anzahl an Attributen muss jedoch ein anderes Kriterium herangezogen werden.

In Unterabschnitt 4.2.1 wird beschrieben wie über die Effizienz der Attributselektion eine obere Grenze für die Anzahl der zu wählenden Attribute bestimmt werden kann. Daher wird mit jeweils 10 000 Signal und Untergrund MC für alle durch den MRMR Algorithmus gewählten Merkmalsmengen die Effizienz für einen Random Forest aus zehn Bäumen berechnet. Die Anzahl der in jedem Knoten zufällig gezogenen Attribute ergibt sich nach  $\log_2(\text{Anzahl Attribute})$ . Die Ergebnisse finden sich in Abbildung 5.11.

## 5. Ereignisselektion



**Abbildung 5.12.:**

Random Forest Confidence Verteilung für Daten und MC. Für Confidence Werte  $< 0,5$  werden die Daten hinreichend gut durch Atmosphärische Myonen beschrieben. Für Confidence Werte  $> 0,8$  werden die Daten hinreichend gut durch atmosphärische Myon-Neutrinos beschrieben. Die gute Übereinstimmung zwischen Daten und MC zeigt sich ebenfalls im Verhältnis in dem  $\sum_{\text{Daten}}^{\text{MC}}$  aufgetragen wurde.

Abbildung 5.11 zeigt, dass die Effizienz für mehr als etwa 30 Attribute eine Sättigung erreicht. Für die weitere Verwendung wird daher

$$M_{\text{Attribute}} = 30 \quad (5.5)$$

gewählt. Die 30 durch den MRMR Algorithmus gewählten Attribute finden sich in Tabelle 5.2. Eine Beschreibung aller Attribute findet sich im Anhang.

Der genutzte Random Forest für die Bestimmung der Attributanzahl unterscheidet sich vom dem für die anschließende Trennung genutzten Random Forest. Der Rechenaufwand für die Erstellung der Modelle ist jedoch sehr gering. Ebenfalls sollte das Verhalten der Effizienz in Abhängigkeit von der Anzahl der Attribute sehr ähnlich sein. Daher ist die Verwendung dieses Verfahrens gerechtfertigt.

### 5.3. Ereignisselektion mittels Random Forest

Die nach dem in Abschnitt 5.2 beschriebenen Verfahren gewählten Attribute bilden die Repräsentation der Daten, die für die weitere Separation zwischen Signal und Untergrund genutzt wird. Um eine Separation basierend auf den 30 Attributen zu erreichen, wird ein, in Unterabschnitt 4.3.3 beschriebener Random Forest Algorithmus genutzt. Die für die Generierung des Modells gewählten Parameter des Random Forests wurden durch eine manuelle Optimierung gewählt. Es ist dabei zu beachten, dass eine Optimierung nach einem fixen Kriterium wie  $\frac{\text{Signal}}{\sqrt{\text{Untergrund}}}$  nicht zielführend ist. Der Grund dafür ist, dass das Ziel nicht die Separation als solche ist, sondern die auf dem separierten Datensatz aufbauende Entfaltung. Das Ziel der Separation ist einen Datensatz zu erzeugen, mit dem eine optimale Entfaltung möglich ist. Da jedoch eine Steigerung der Effizienz der Separation sich nicht zwangsläufig in eine bessere Entfaltung übersetzt, muss für jede getestete Einstellung immer die Auswirkung auf die Entfaltung geprüft werden. Daher ist eine geschlossene automatische Optimierung der Parameter des Random Forests nicht möglich.

In der Optimierung ergeben sich für den Random Forest folgende Einstellungen als optimal: Die Anzahl der genutzten Entscheidungsbäume im Random Forest ergibt sich zu

$$N_{\text{Bäume}} = 200. \quad (5.6)$$

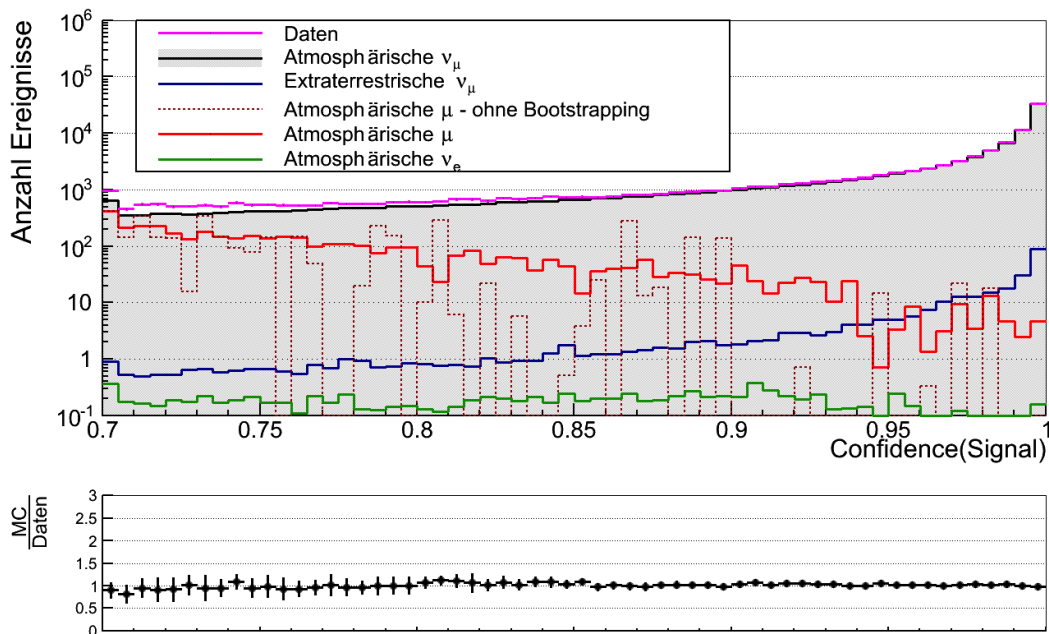
Die Anzahl der zufällig gezogenen Attribute aus den maximal 30 zur Verfügung stehenden Attributen ergibt sich zu

$$m_{\text{Attribute}} = 5. \quad (5.7)$$

Dies passt zu der in [Bre01] beschriebenen Regel, nach der die Anzahl der zufällig zu ziehenden Attribute entweder dem Logarithmus Dualis oder der Wurzel der Anzahl der zur Verfügung stehenden Attribute entsprechen sollte. Für 30 Attribute ergeben sich sowohl für den Logarithmus Dualis als auch für die Wurzel jeweils gerundet fünf als Wert. In Tabelle 5.2 wird deutlich, dass ab einer Zahl von etwa 25 Attributen die Bedeutung der Attribute für die Trennung abnimmt. Dieses Ergebnis ist zu erwarten, da für mehr als 30 Attribute keine Steigerungen in der Effizienz festzustellen ist.

Der letzte Parameter ist die Tiefe der einzelnen Entscheidungsbäume, welcher nicht beschränkt wird. Es ist jedoch festzuhalten, dass die drei beschriebenen Parameter des

## 5. Ereignisselektion



**Abbildung 5.13.:**

Die Abbildung zeigt im oberen Histogramm die Random Forest Confidence-Verteilung für Daten und MC in einer Ausschnittsvergrößerung für den Confidence Bereich  $> 0,7$ . Die Daten werden hinreichend durch die MC beschrieben. Es ist ebenfalls zu erkennen, dass der Anteil von Untergrund für hohe Confidence-Werte extrem abnimmt.

Algorithmen keinen starken Einfluss auf die Trennkraft des Random Forests haben und die Maxima in der Optimierung der Parameter sehr breit sind.

Eine weitere Optimierung zeigt, dass die absolute Anzahl an MC, sowie das Verhältnis von Signal und Untergrund MC beim Trainieren des Random Forests einen nicht zu vernachlässigenden Einfluss auf das Ergebnis hat. Ein Optimum ergibt sich für

$$L_{\text{Signal}} = 100\,000 \quad (5.8)$$

Signal MC Ereignisse und

$$L_{\text{Untergrund}} = 10\,000 \quad (5.9)$$

Untergrund MC Ereignisse. Mit diesen zwei Werten sind alle Parameter des Random Forests bestimmt. Eine Zusammenfassung der Random Forest Einstellungen findet sich in Tabelle 5.3.

**Tabelle 5.2.:**

Auflistung der Tree Importance, beschrieben in Unterabschnitt 4.3.3, der 30 durch den MRMR Algorithmus gewählten Attribute. Die Tree Importance aller Attribute wird auf das Attribut  $\text{Bayesian}_\Theta$  normiert, da dieses die höchste Trennkraft aufweist. Von besonderer Bedeutung für die Trennung sind Attribute, welche eine Aussage über die Qualität der Rekonstruktion machen.

Attribut	Importance	Attribut	Importance
1 $\text{Bayesian}_\Theta$	1,00	16 $r_{\text{weight}}$	0,25
2 $\text{Paraboloid}_{\text{err1}}$	0,95	17 $Z_{\text{min}}$	0,25
3 $\text{MuEXAngular}4_\sigma$	0,83	18 $\sigma_{\text{COG}, Z}$	0,24
4 $\text{StochastizitätE}$	0,74	19 $ \text{StochastizitätC} $	0,21
5 $\text{MPEFitHighNoise}_{r\log l}$	0,68	20 $r_{\text{COG}}$	0,19
6 $\text{LineFit}_\Theta$	0,64	21 $\text{StochastizitätC}$	0,17
7 $\text{TEBINS}_E$	0,61	22 $\text{MuEX}_E$	0,13
8 $\text{TEDOMS}_E$	0,58	23 $\text{StringsC}$	0,10
9 $d_{\text{COG}}$	0,52	24 $\text{TEBINS}_\sigma$	0,10
10 $\text{Paraboloid}_{\text{err2}}$	0,49	25 $\text{Bayesian}_z$	0,07
11 $Z_{\text{max}}$	0,41	26 $\text{DOMsA}$	0,06
12 $\text{LängeC}$	0,39	27 $\text{StringsA}$	0,05
13 $ \text{SplieMPE}_\Theta - \text{LineFit}_\Theta $	0,38	28 $\text{DOMsC}$	0,04
14 $\text{LineFit}_c$	0,33	29 $\text{TEDOMS}_\sigma$	0,01
15 $\text{LängeA}$	0,26	30 $\text{SPE}_{\text{status}}$	0,01

**Tabelle 5.3.:**

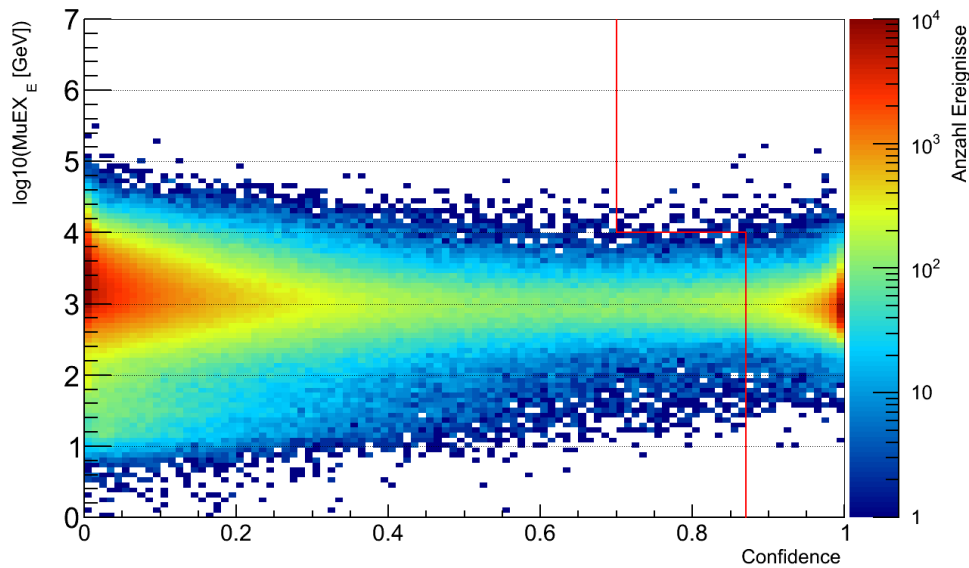
Zusammenfassung der genutzten Random Forest Parameter

Parameter	Wert
$N_{\text{Bäume}}$	200
$M_{\text{Attribute}}$	30
$m_{\text{Attribute}}$	5
$L_{\text{Signal}}$	100 000
$L_{\text{Untergrund}}$	10 000

Durch ein Training des Random Forests mit den beschriebenen Einstellungen ergibt sich für MC und Daten die Confidence(Signal)-Verteilung (im Folgenden Confidence), dargestellt in Abbildung 5.12. Wie an der Darstellung zu erkennen ist, werden die Daten von den MC hinreichend gut beschrieben. Die Übereinstimmung zwischen Daten und MC ist die Grundvoraussetzung dafür, die Ergebnisse des erzeugten Modells für die weitere Analyse zu nutzen. Der entscheidende Punkt, damit die Ergebnisse des Modells einen Nutzen für die weitere Analyse bringen, ist, dass das Signal ein scharfes Maximum für den Wert Confidence von eins aufweist, wohingegen

der dominante Untergrund ein scharfes Maximum für eine Confidence von null aufweist. Durch die scharfen Maxima von Signal und Untergrund ist es möglich, diese beiden Bestandteile der Daten voneinander zu trennen. Um die mögliche Trennkraft

## 5. Ereignisselektion



**Abbildung 5.14.:**

Das zweidimensionale Histogramm zeigt den von Confidence und  $\text{MuEX}_E$  aufgespannten Phasenraum. Die zwei Populationen von Signal und Untergrund sind deutlich zu erkennen. In rot ist der genutzte zweidimensionale Schnitt eingezeichnet.

eines Schnittes in der Confidence besser einschätzen zu können, ist in Abbildung 5.13 eine Ausschnittsvergrößerung des Confidence-Bereiches größer 0,7 dargestellt. In Abbildung 5.13 ist deutlich zu sehen, dass der Untergrund der atmosphärischen Myonen ab einem Confidence-Wert von etwa 0,8 ungefähr eine Größenordnung unter dem gesuchten Signal der atmosphärischen Myon-Neutrinos liegt.

Das Signal der hochenergetischen extraterrestrischen Neutrinos ist jedoch, wie zu erwarten, entartet mit den atmosphärischen Neutrinos. Dies zeigt sich daran, dass beide Ereignistypen ein scharfes Maximum in der Confidence-Verteilung bei eins aufweisen. Das Ziel ist es, diese Komponenten durch ihre Unterschiede im spektralen Verhalten zu unterscheiden. Da der Fluss der atmosphärischen und hochenergetischen extraterrestrischen Neutrinos einem unterschiedlichen spektralen Index folgen sollte, ist zu erwarten, dass der Unterschied besonders prominent im Bereich der Hochenergie ist, da in diesem Bereich der Fluss der extraterrestrischen Neutrinos über dem Fluss der atmosphärischen Neutrinos liegen sollte. Daher ist es von besonderer Bedeutung, eine möglichst hohe Effizienz in diesem Energiebereich zu erzielen. Die Confidence hängt, wie in Abbildung 5.14 zu erkennen ist, nur schwach von der Myonenenergie ab. Die Separation in allen Energiebereichen ist daher ähnlich effizient. Da ein besonderes Interesse darin besteht, die Effizienz im hochenergetischen Bereich zu steigern, bietet es

sich an, den finalen Datensatz nicht über einen einfachen Schnitt in der Confidence zu erzeugen.

Vielmehr bietet es sich an, einen Schnitt im durch Confidence und Energieschätzer aufgespannten Phasenraum durchzuführen, da durch dieses Vorgehen die Effizienz energieabhängig gesteigert werden kann.

**Tabelle 5.4.:**

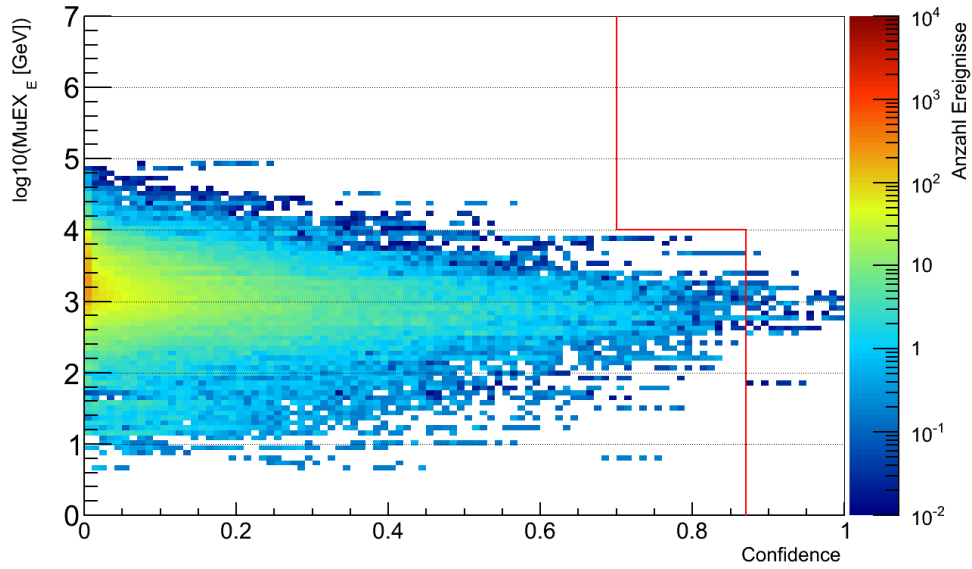
Durch das Anwenden des in Gleichung 5.10 beschriebenen Schnittes auf Daten und MC ergeben sich die beschriebenen Kenngrößen des generierten Datensatzes.

Kennzahl	Wert	Fehler
Ereignisse	93 280	
Ereignisse $E_\mu > 1 \text{ TeV}$	39 088	
Ereignisse $E_\mu > 10 \text{ TeV}$	233	
Atm. $\nu_\mu$	93 266	$\pm 390$
Atm. $\nu_\mu E_\mu > 1 \text{ TeV}$	37 231	$\pm 360$
Atm. $\nu_\mu E_\mu > 10 \text{ TeV}$	210	$\pm 6$
Atm. $\mu$	491	$\pm 263$
Atm. $\mu E_\mu > 1 \text{ TeV}$	169	$\pm 140$
Atm. $\mu E_\mu > 10 \text{ TeV}$	0,01	$\pm 0,01$
Reinheit	0,996	$\pm 0,002$
Reinheit $E_\mu > 1 \text{ TeV}$	0,997	$\pm 0,003$
Reinheit $E_\mu > 10 \text{ TeV}$	0,99996	$\pm 0,00004$

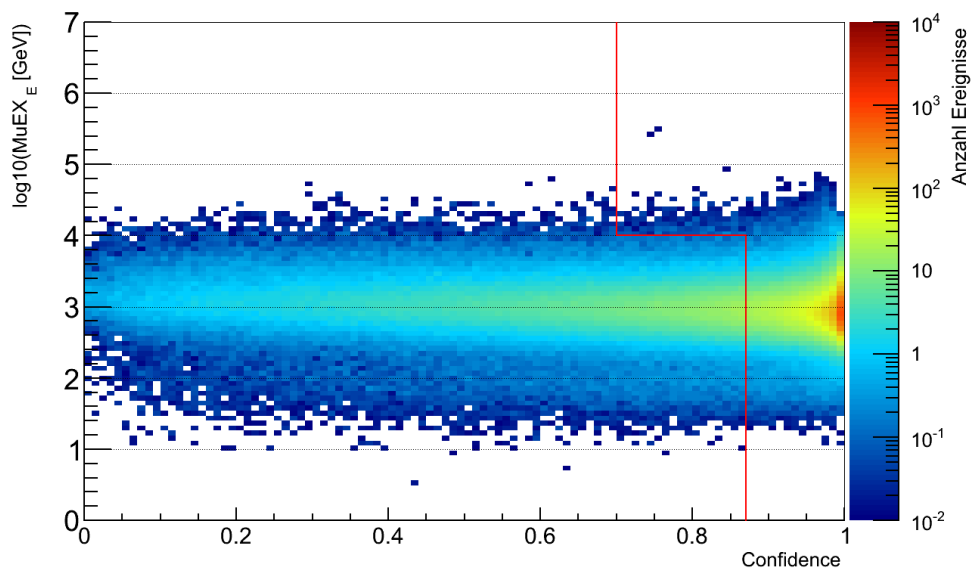
In Abbildung 5.14 sind deutlich die zwei Populationen in den Daten erkennbar. Ein Vergleich der Abbildung mit den MC-Vorhersagen für Signal und Untergrund in den Abbildungen 5.15(a) und 5.15(b) ergibt, dass die beiden Populationen leicht durch Signal und Untergrund erklärbar sind. Ebenfalls ist erkennbar, dass es für den Phasenraumbereich hoher Energien und hoher Confidence einen Bereich gibt, in dem es keine Vorhersage von Untergrundereignissen, jedoch eine nicht verschwindende Vorhersage von Signalereignissen gibt.

Bei dieser Vorgehensweise ist zu beachten, dass in diesem Phasenraumbereich die Anzahl der simulierten Untergrundereignisse gering ist. Daher ist ein Schnitt in diesem Bereich ohne eine zusätzliche Absicherung mit dem Risiko verbunden den finalen Datensatz mit nicht beschriebenem Untergrund zu verunreinigen. Um diesem Risiko entgegenzuwirken, wurde die in Abschnitt 4.3.1 beschriebene Erweiterung des Bootstrappings angewandt, um die zur Verfügung stehenden MC optimal auszunutzen. Durch dieses Vorgehen kann nahezu ausgeschlossen werden, dass durch einen Schnitt

## 5. Ereignisselektion



(a) Zweidimensionaler Phasenraum für Untergrund MC.

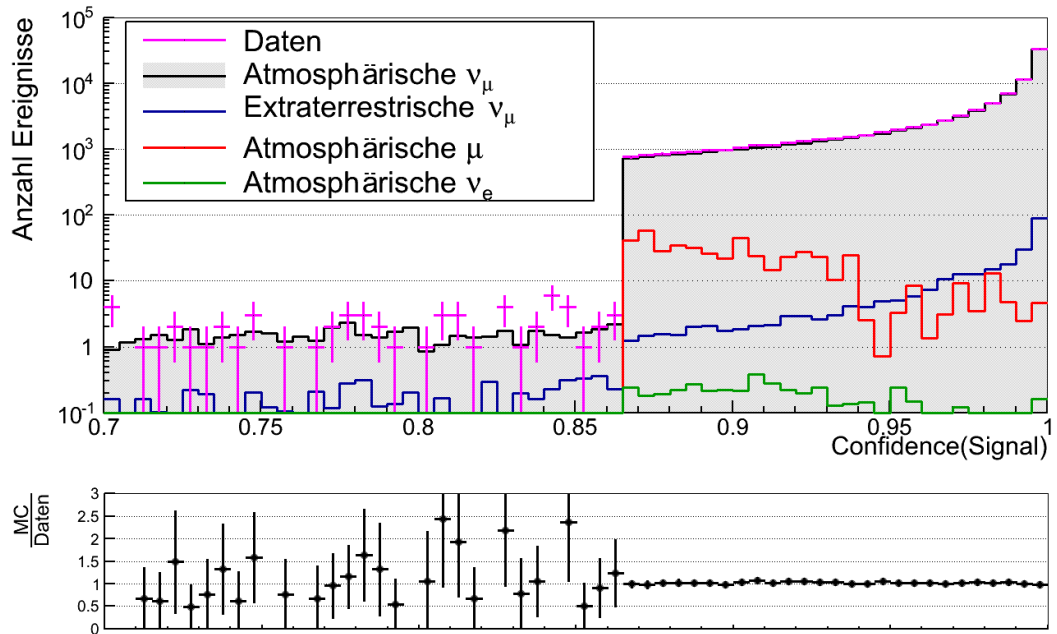


(b) Zweidimensionaler Phasenraum für Signal MC.

### Abbildung 5.15.:

Die zweidimensionalen Histogramme (a) und (b) zeigen die Verteilung von Signal und Untergrund im von Confidence und Energierekonstruktion aufgespannten Phasenraum. In rot ist der genutzte zweidimensionale Schnitt eingezeichnet.




**Abbildung 5.16.:**

Die Abbildung zeigt im oberen Histogramm die Random Forest Confidence-Verteilung für Daten und MC in einer Ausschnittsvergrößerung für den Confidence-Bereich  $> 0,7$ . Die Daten werden hinreichend durch die MC beschrieben. Es ist ebenfalls deutlich zu erkennen, dass der Anteil von Untergrund für hohe Confidence-Werte extrem abnimmt.

in diesem Bereich des Phasenraums unerwünschter Untergrund in den finalen Datensatz gelangt.

Im Rahmen einer Optimierung ergibt sich der zweidimensionale Schnitt F als optimal:

$$F = \begin{cases} \text{Confidence} \geq 0,87; & \text{falls } \text{MuEX}_E < 10 \text{ TeV} \\ \text{Confidence} \geq 0,70; & \text{sonst} \end{cases} \quad (5.10)$$

Durch Anwendung des in Gleichung 5.10 beschriebenen Schnitts auf Daten und MC ergeben sich für Daten und MC die in Tabelle 5.4 beschriebenen Werte. Wie zu erkennen ist, ist der gewählte Schnitt in der Lage einen hochreinen Datensatz von Myon-Neutrinoereignissen zu generieren. Der Datensatz zeichnet sich insbesondere durch seine sehr hohe Effizienz und Reinheit für Energien größer 10 TeV aus. Dies ist von besonderer Bedeutung, da eines der vorrangigen Ziele dieser Analyse die genaue Vermessung des Energiespektrums der von IceCube gemessenen Myon-Neutrinos im Bereich über 10 TeV ist.

Die Confidence-Verteilung verändert sich durch den in Gleichung 5.10 beschriebenen Schnitt zu Abbildung 5.16. Durch den zweidimensionalen Schnitt werden hochenergetische Ereignisse selektiert, die ansonsten nicht berücksichtigt werden. Der Vergleich der Effizienzen nach der Selektion durch den Random Forest in Abbildung 5.17(a), mit der Effizienz nach den Vorselektionsschnitten in Abbildung 5.10(a) zeigt, dass der Effizienzverlust durch den Random Forest im Bereich der Hochenergie bei etwa 10 % liegt. Die Gesamteffizienz im Bereich der Hochenergie liegt bei etwa 50 %. Eine Betrachtung der Effizienz ohne Berücksichtigung des Zenitschnittes findet sich in Abbildung 5.17(b). Der Random Forest erreicht in diesem Fall eine Effizienz von etwa 70 % im Bereich der Hochenergie. Die gesonderte Betrachtung ohne die Auswirkungen des Zenitschnittes ergibt sich daraus, dass für jede vergleichbare Analyse ein solcher Schnitt zwingend erforderlich ist.

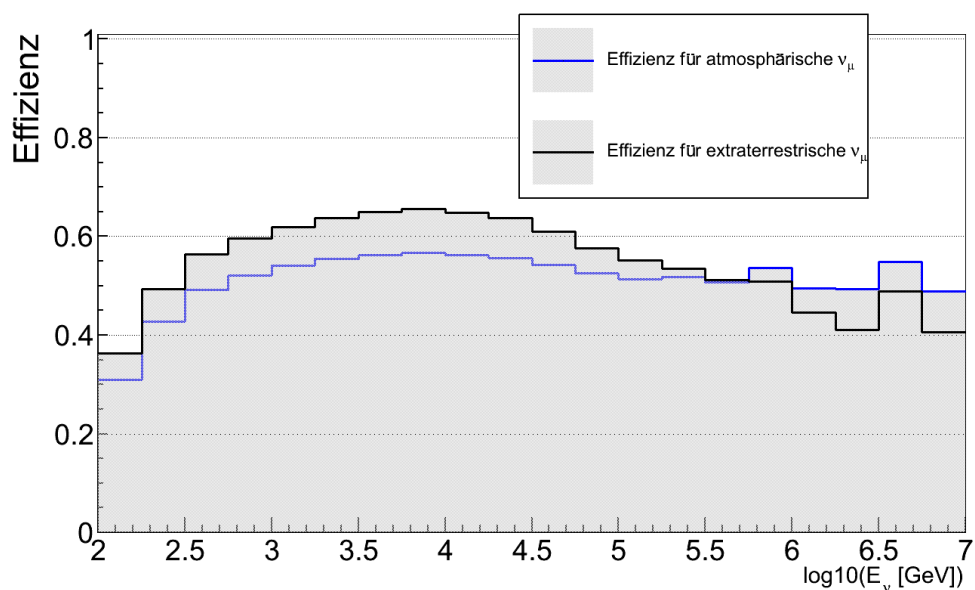
Um die Effizienz der Ereignisselektion besser beurteilen zu können, bietet es sich an, die sich ergebende effektive Fläche, beschrieben in Kapitel 3, mit bereits bestehenden Analysen zu vergleichen. Ein Vergleich der Analyse mit den in Abbildung 5.18 beschriebenen Analysen lässt eine deutliche Effizienzsteigerung feststellen. Über den gesamten Energiebereich liegt die effektive Fläche der vorliegenden Analyse über der der Referenzanalysen. Ab einer Energie von etwa einem PeV ist für alle Analysen die Anzahl der zur Verfügung stehenden MC begrenzt, was sich in den in Abbildung 5.18 erkennbaren Fluktuationen der effektiven Fläche äußert.

Wie in Abbildung 5.18 zu erkennen ist, konnte die Effizienz der vorliegenden Analyse deutlich gegenüber vergleichbaren Analysen gesteigert werden.

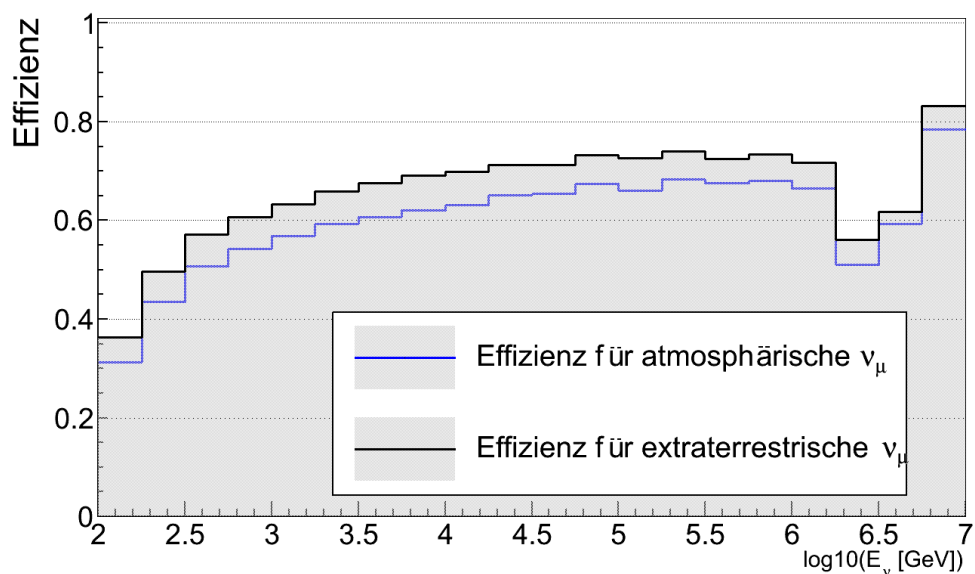
Um sicherzustellen, dass alle verwendeten Attribute auch nach einer Anwendung des in Gleichung 5.10 beschriebenen Schnittes ein gutes Daten-MC-Verhältnis aufweisen, werden diese auf ein hinreichendes Verhältnis geprüft. Keines der verwendeten Attribute zeigt eine Auffälligkeit. Es kann daher davon ausgegangen werden, dass die Separation hinreichend gute Ergebnisse liefert. Die zugehörigen Abbildungen finden sich im Anhang.

Da das Ziel dieser Analyse die Entfaltung des Energiespektrums der von IceCube gemessenen Myon-Neutrinos ist, wird im Folgenden die im finalen Datensatz gemessene Myonenergie genauer betrachtet. Zu diesem Zweck wird der Myon-Energieschätzer  $\text{MuEX}_E$  genutzt. Wie in Abbildung 5.19(a) zu sehen ist, ist das Attribut hinreichend gut simuliert und weist in allen Bins einen subdominanten Untergrund auf.

Das Attribut ist von besonderem Interesse, da ab einer Energie von etwa 10 TeV rekonstruierter Myonenergie ein Überschuss über der reinen atmosphärischen Vorhersage



(a) Effizienz der Ereignisselektion durch den Random Forest bezogen auf das Level3.

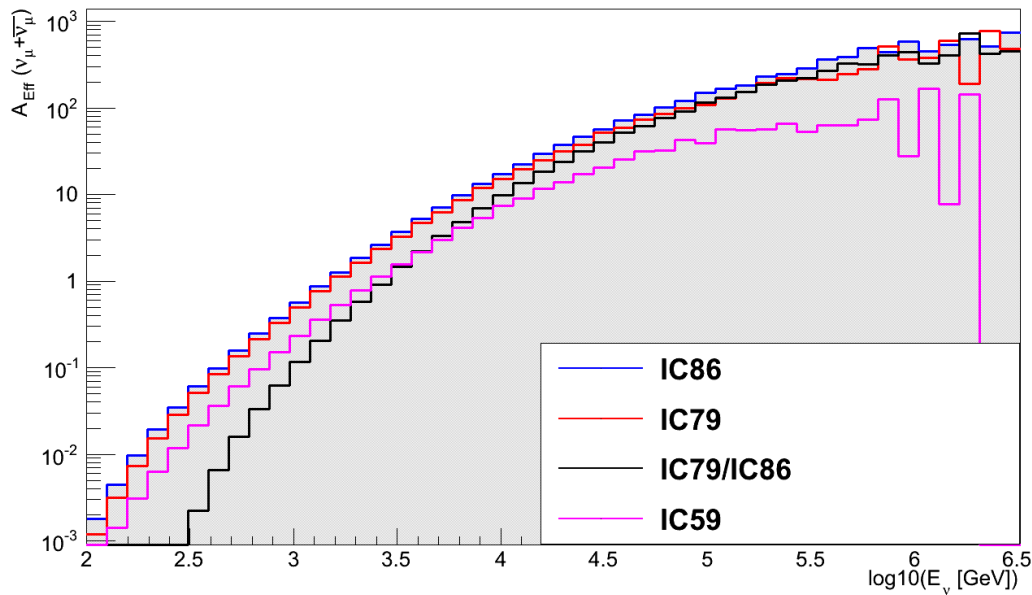


(b) Effizienz der Ereignisselektion durch den Random Forest bezogen auf das Level3 mit einem Zenitschnitt bei  $86^\circ$ .

**Abbildung 5.17.:**

Die Effizienz der Ereignisselektion im Energiebereich von 10 GeV bis 10 PeV. Abbildung (a) zeigt die Effizienz der Ereignisselektion unter Berücksichtigung aller Vorselektionsschnitte. Abbildung (b) zeigt den Einfluss des Schnittes auf den rekonstruierten Zenitwinkel. Es zeigt sich, dass ein Großteil des Effizienzverlusts auf den Zenitschnitt zurückzuführen ist.

## 5. Ereignisselektion

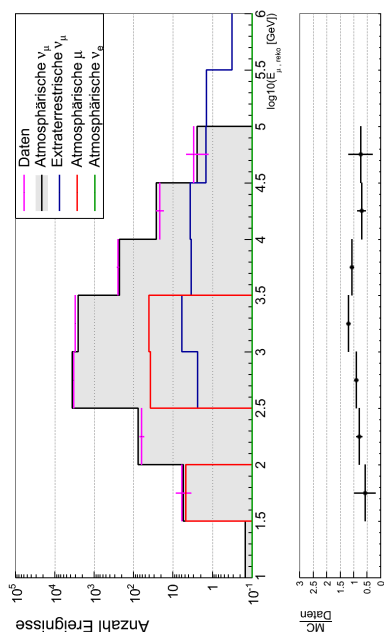


**Abbildung 5.18.:**

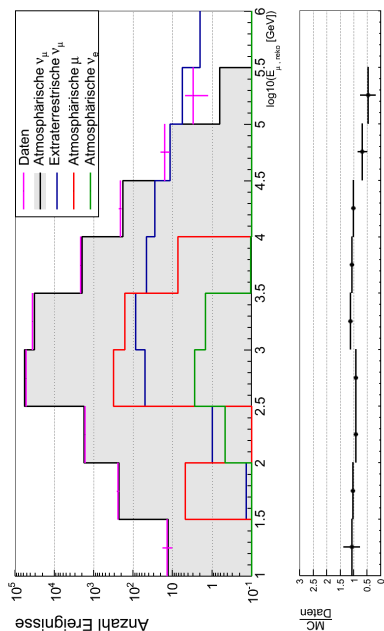
Vergleich der effektiven Fläche der vorliegenden IC86 Analyse mit drei vergleichbaren Ereignisselektionen. Blau: Ereignisselektion mittels Random Forest auf Basis der IceCube 79 String-Konfiguration [Sch14]. Schwarz: Ereignisselektion mittels manueller Schnitte auf Basis der IceCube 79 und 86 String-Konfiguration [A<sup>+</sup>ep]. Rot: Ereignisselektion mittels Random Forest auf Basis der IceCube 59 String-Konfiguration [Ruh13]. Die vorliegende Analyse ist in allen Energiebereichen deutlich effizienter als die Referenzen. Erst ab einer Energie von etwa einem PeV kann auf Grund einer mangelnden Anzahl an MC keine klare Aussage getroffen werden.

zu beobachten ist. Der zu beobachtende Überschuss ähnelt dem in Abbildung 5.19(a) eingezeichneten isotropen Fluss hochenergetischer extraterrestrischer Myon-Neutrinos, welcher mit den Ergebnissen von [A<sup>+</sup>14c] kompatibel ist.

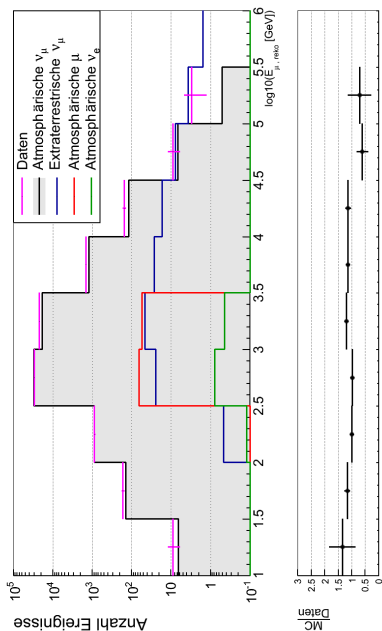
Ferner muss ausgeschlossen werden, dass der beobachtete Überschuss an hochenergetischen Ereignissen nicht durch falsch rekonstruierte atmosphärische Myonen entsteht. Der Verdacht ist auszuräumen, indem die MuEX<sub>E</sub>-Verteilung in verschiedenen Zenitwinkelbereichen betrachtet wird. Sollte der Exzess durch atmosphärische Myonen verursacht werden, wäre ein Exzess im Winkelbereich  $86^\circ \leq \Theta \leq 90^\circ$  zu erwarten. Wie in Abbildung 5.19(b) jedoch zu erkennen ist, findet sich in diesem Zenitbereich keine Überfluktuation. Ebenfalls ist davon auszugehen, dass hochenergetische Neutrinos den Detektor nicht durch die Erde erreichen, da die Erde, wie in Kapitel 2 beschrieben, für hochenergetische Neutrinos undurchlässig ist. Daher ist ein Exzess im Zenitbereich  $120^\circ \leq \Theta \leq 180^\circ$  ebenfalls nicht zu erwarten. In Abbildung 5.19(d) findet sich in diesem Bereich keine auffällige Überfluktuation. Unter der Annahme, dass es sich bei der Überfluktuation um extraterrestrische Neutrinos handelt, ist zu erwar-



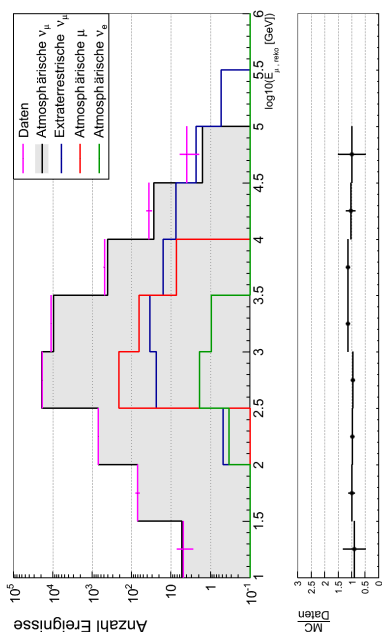
(a) MuEX<sub>E</sub> Verteilung für den Winkelbereich 86° bis 180°.



(b) MuEX<sub>E</sub> Verteilung für den Winkelbereich 86° bis 90°.



(c) MuEX<sub>E</sub> Verteilung für den Winkelbereich 90° bis 120°.



(d) MuEX<sub>E</sub> Verteilung für den Winkelbereich 120° bis 180°.

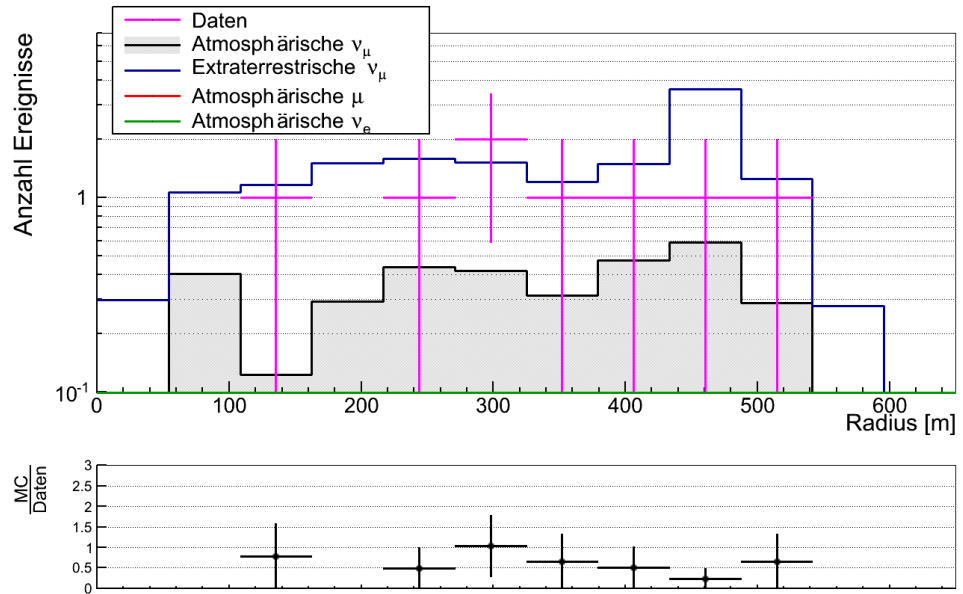
**Abbildung 5.19.:**

MuEX<sub>E</sub>-Verteilung für verschiedene Winkelbereiche. Abbildung (a) zeigt einen Exzess der Daten oberhalb der Vorhersage für atmosphärische Neutrinos. Der Exzess ist kompatibel mit der Vorhersage für einen Fluss von extraterrestrischen Neutrinos, wie er in [A<sup>+</sup>14c] gemessen wurde. Ein deutlicher Exzess der gemessenen Daten oberhalb der Vorhersage im Winkelbereich 86° bis 90° würde auf den Einfluss von atmosphärischen Myonen hindeuten. Wie in Abbildung (b) zu erkennen ist, findet sich in diesem Bereich kein Exzess. Der Exzess ist vielmehr auch in den Winkelbereichen 90° bis 120° und 120° bis 180° kompatibel mit der Vorhersage für atmosphärische Neutrinos.

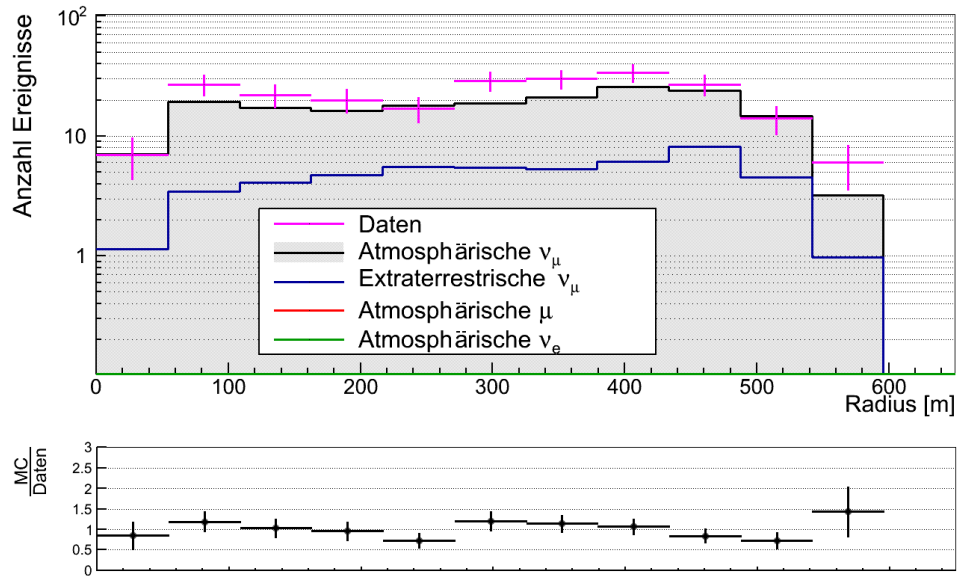
teten einen ausgeprägten Exzess über der Vorhersage der atmosphärischen Neutrinos im Winkelbereich  $90^\circ \leq \Theta \leq 120^\circ$  zu sehen. Dieser Bereich wird in Abbildung 5.19(c) dargestellt. Der erwartete Exzess ist zu erkennen. Aus der energieabhängigen Zenitverteilung der Ereignisse lassen sich also keine Hinweise auf den Einfluss von atmosphärischen Myonen ziehen.

Eine weitere Möglichkeit zu überprüfen, ob der Exzess aus atmosphärischen Myonen stammt, ist die  $r_{\text{COG}}$ -Verteilung der hochenergetischen Ereignisse. Sollten Myonen für den Exzess verantwortlich sein, könnte sich dies darin manifestieren, dass eine erhöhte Anzahl an Ereignissen am Rand des Detektors auftritt. Diese Ereignisse bergen das Risiko, dass ihre Trajektorie falsch rekonstruiert wird. Anhand der Abbildungen 5.20(a) und 5.20(b) ist zu erkennen, dass in der  $r_{\text{COG}}$ -Verteilung kein auffälliger Exzess zu beobachten ist. Insgesamt gibt es keine Hinweise darauf, dass der beobachtete Exzess durch falsch rekonstruierte atmosphärische Myonen zustande kommt.

Der Datensatz wird durch die beschriebenen Tests gegen den Einfluss von atmosphärischen Myonen abgesichert. Die in [Sch14] und [A<sup>+</sup>13b] beschriebenen Analysen zeigen einen ähnlichen Exzess. Es ist daher davon auszugehen, dass es sich bei diesen Ereignissen nicht um atmosphärische Neutrinos handelt. Es stellt sich daher die Frage, wie sich dieser Exzess auf das Spektrum auswirkt.



(a)  $r_{COE}$  für Ereignisse mit  $MuEX_E > 50$  TeV.



(b)  $r_{COE}$  für Ereignisse mit  $MuEX_E > 10$  TeV.

**Abbildung 5.20.:**

$r_{COE}$  für zwei Schnitte in  $MuEX_E$ . Sollte der in Abbildung 5.19 aufgezeigte Exzess durch Myonen am Rand des Detektors verursacht werden, würde sich dies durch einen Exzess für hohe Werte von  $r_{COE}$  zeigen. Dieser Exzess sollte durch eine Beschränkung der Betrachtung auf die hochenergetischen Ereignisse deutlicher werden.





## 6. Entfaltung des Energiespektrums

In diesem Kapitel wird die Entfaltung des Energiespektrums der von IceCube gemessenen Myon-Neutrinos beschrieben. Dabei wird die Software TRUÉE für die Entfaltung eingesetzt. Die Software wurde im Rahmen der Arbeit von [Mil12] entwickelt und in [MDK<sup>+</sup>13] ausführlich beschrieben. Die Weiterentwicklung und Anpassung der Software an IceCube werden in [Sch14] und [Ruh13] beschrieben. Die im weiteren Verlauf beschriebene Entfaltung der IceCube-Daten beruht in weiten Teilen auf den Vorarbeiten dieser Analysen. Insbesondere wurden die in [Sch14] entwickelten Validierungsverfahren eingesetzt, um die Qualität des erzeugten Spektrums sicherzustellen.

Neben einer zuverlässigen und qualitativ hochwertigen Entfaltungsoftware ist für ein optimales Entfaltungsergebnis der genutzte Datensatz von großer Bedeutung. Die Eigenschaften und die Entstehung des für die Entfaltung genutzten Datensatzes wurden in Kapitel 5 beschrieben. Die für die Entfaltung verwendeten systematischen Datensätze werden in Tabelle 6.3 aufgelistet. Wie aus den Datensätzen das Energiespektrum gewonnen wird, wird im Folgenden beschrieben.

### 6.1. Das entfaltete Energiespektrum

Nachdem bereits in [A<sup>+</sup>14c] ein Fluss von hochenergetischen extraterrestrischen Neutrinos nachgewiesen und in [Sch14] ein Abflachen des Spektrums nachgewiesen wurde, ist das Ziel der vorliegenden Entfaltung insbesondere eine Verbesserung der Auflösung. Da der genutzte Datensatz im Vergleich zu anderen Analysen über den gesamten Energiebereich eine deutlich gesteigerte Effizienz aufweist, sollte es möglich sein, die Unsicherheiten der Entfaltung weiter zu verringern. Daher können möglicherweise Rückschlüsse auf die prompte und hochenergetische extraterrestrische Komponente gezogen werden. Eine ausführliche Diskussion findet sich in Kapitel 7.

## 6. Entfaltung des Energiespektrums

**Tabelle 6.1.:**

In der ersten Spalte ist die Binnummer angegeben. In der zweiten Spalte werden Binbreite und Grenzen beschrieben. Die dritte Spalte zeigt die Anzahl der entfaltenen Ereignisse. In der vierten Spalte befindet sich der entfaltete Fluss, gewichtet mit dem Quadrat der Energie des  $x$ -Wertes. In der fünften Spalte wird die Unsicherheit des Flusses in Prozent angegeben. Die nächste Spalte gibt den  $x$ -Wert der Entfaltung wieder.

Bin	Binbreite [ $\log_{10}(\frac{E}{\text{GeV}})$ ]	Entfalt. Ereignisse	$E^2 \times \text{Fluss}$ [ $\frac{\text{GeV}}{\text{sr s cm}^2}$ ]	Unsicherheit [%]	$x$ -Wert [GeV]
1	[2,1; 2,4]	12698,2	$2,67 \times 10^{-4}$	+11,3 -14,6	$1,88 \times 10^2$
2	[2,4; 2,7]	20386,6	$1,10 \times 10^{-4}$	+19,3 -16,4	$3,36 \times 10^2$
3	[2,7; 3,0]	19811,9	$6,56 \times 10^{-5}$	+41,3 -21,7	$7,25 \times 10^2$
4	[3,0; 3,3]	15757,8	$3,28 \times 10^{-5}$	+13,3 -20,4	$1,42 \times 10^3$
5	[3,3; 3,6]	10304,7	$1,50 \times 10^{-5}$	+32,8 -50,1	$2,71 \times 10^3$
6	[3,6; 3,9]	5366,2	$7,78 \times 10^{-6}$	+31,5 -42,0	$6,04 \times 10^3$
7	[3,9; 4,2]	2537,7	$2,39 \times 10^{-6}$	+13,8 -11,8	$1,13 \times 10^4$
8	[4,2; 4,5]	1001,8	$6,68 \times 10^{-7}$	+20,3 -12,8	$2,17 \times 10^4$
9	[4,5; 4,8]	312,7	$2,77 \times 10^{-7}$	+29,3 -37,0	$4,80 \times 10^4$
10	[4,8; 5,1]	93,4	$8,99 \times 10^{-8}$	+40,0 -63,4	$8,84 \times 10^4$
11	[5,1; 5,4]	32,0	$4,36 \times 10^{-8}$	+46,6 -74,1	$1,85 \times 10^5$
12	[5,4; 5,7]	13,8	$2,05 \times 10^{-8}$	+60,6 -89,6	$3,63 \times 10^5$
13	[5,7; 6,0]	6,6	$1,09 \times 10^{-8}$	+93,1 -86,2	$7,11 \times 10^5$
14	[6,0; 6,5]	4,8	$7,33 \times 10^{-9}$	+151,8 -104,5	$1,76 \times 10^6$

Das Entfaltungsproblem wird durch Gleichung 4.12 beschrieben. Da der verwendete Datensatz eine hohe Reinheit aufweist, kann der Einfluss des Untergrundes  $b(y)$  vernachlässigt werden. Das Entfaltungsproblem wird zu Gleichung 6.1 vereinfacht.

$$g(y) = \int_a^b A(y, x) f(x) dx + \cancel{b(y)} \quad (6.1)$$

$g(y)$ : Gemessene Verteilungsfunktion  
 $f(x)$ : Gesuchte Verteilungsfunktion  
 $A(y, x)$ : Der Kern des Integrals  
 $b(y)$ : Verteilungsfunktion des Untergrunds

Da die Entfaltung unter Nutzung der Software TRUEE ausgeführt wird, können drei Attribute gleichzeitig genutzt werden. Tests mit verschiedenen Kombinationen von Attributen zeigen, dass die Kombination der Attribute LängeC, DOMsC und MuEX<sub>E</sub> in der Entfaltung ein optimales Ergebnis liefert. Diese Kombination von Attributen ergab bereits in den in [RM<sup>+</sup>12] und [Sch14] beschriebenen Entfaltungsanalysen sehr gute Ergebnisse. Die gewählten Attribute werden im Folgenden nun genauer vorgestellt.

Das erste gewählte Attribut, LängeC, beschreibt die Länge eines Ereignisses. Das Attribut eignet sich insbesondere als Energieschätzer im Bereich der niedrigen Energien, da

Myonen mit hoher Energie den Detektor komplett durchqueren können und so immer eine hohe LängeC aufweisen. Das Attribut LängeC wird in Abbildung 6.1(a) dargestellt. Wie sich erkennen lässt, ergibt sich für das Attribut eine hinreichend gute Übereinstimmung zwischen gemessenen Daten und Simulation. Es wird ebenfalls deutlich, dass der noch vorhandene Untergrund nicht im Bereich der hohen Energien liegt. In Abbildung 6.1(b) wird die Korrelation des Attributs mit der gesuchten Neutrinoenergie dargestellt. Wie zu erkennen ist, zeigt sich über den gesamten Energiebereich eine Korrelation mit einer hohen Varianz.

**Tabelle 6.2.:**  
Einstellungen, mit denen die Entfaltungsssoftware TRUEE betrieben wird.

Parameter	Wert
Knoten	20
Freiheitsgrade	7
Binning LängeC	10
Binning DOMsC	10
Binning MuEX <sub>E</sub>	10

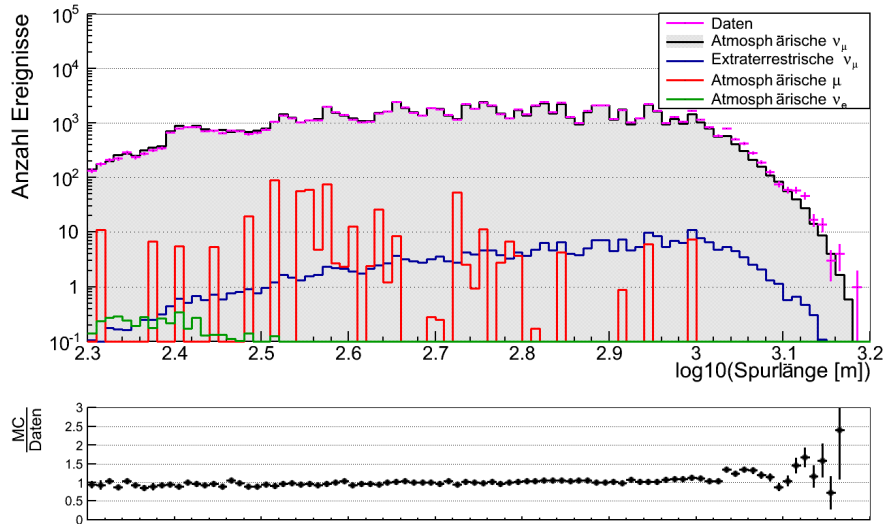
Das zweite gewählte Attribut ist DOMsC. Das Attribut gibt die Anzahl der in einem Ereignis angesprochene DOMs wieder. Da die Anzahl der angesprochenen DOMs während eines Ereignisses stark mit dem Energieverlust des Myons korreliert ist, ist das Attribut ein guter Schätzer für den Energieverlust der Myonen. Eine histogrammierte Darstellung des Attributes findet sich in Abbildung 6.2(a). Das Attribut wird genau wie LängeC hinreichend gut von den MC beschrieben und die vorhandenen Untergrundergebnisse befinden sich im Bereich der niedrigen Energien. Die Korrelation des Attributs mit der Neutrinoenergie findet sich in Abbildung 6.2(b). Wie zu erkennen ist, ist die Abhängigkeit für mittlere und hohe Energien stark ausgeprägt, allerdings zeigt sich ebenfalls eine recht große Varianz.

Die hohe Varianz der ersten beiden vorgestellten Attribute erklärt sich daraus, dass in ihnen keine Unterscheidung von verschiedenen Ereignistopologien vorgenommen wird. Für DOMsC bedeutet dies, dass ein Ereignis, dessen Trajektorie eine Staubschicht im Detektor durchläuft, konzeptionell eine geringere Anzahl an angesprochenen DOMs aufweist als ein Ereignis, dessen Trajektorie nicht durch diese Staubschicht läuft. Eine kleine Varianz ist jedoch vorteilhaft für die Rekonstruktion des Energiespektrums.

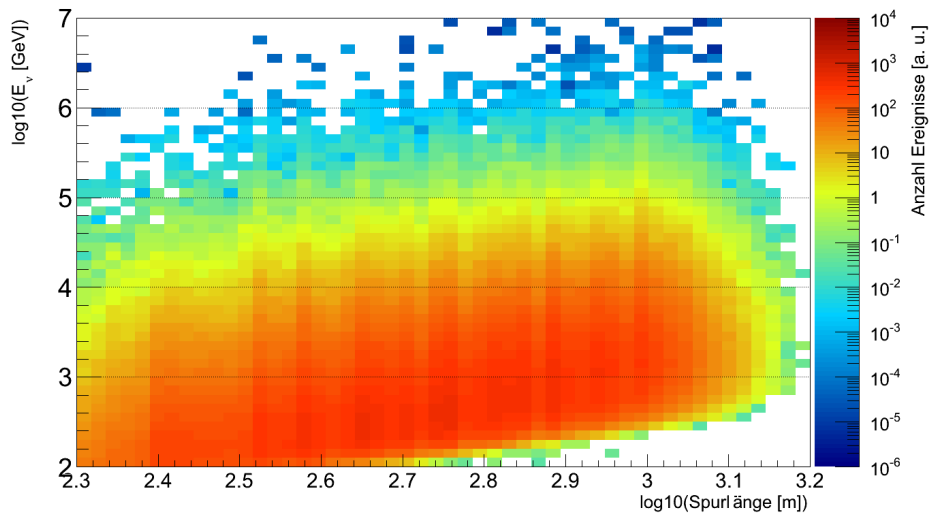
Das in Abbildung 6.3 beschriebene Attribut MuEX<sub>E</sub> basiert auf einer Likelihoodschätzung der Myonenergie. Bei dieser Schätzung wird die gesamte Ereignistopologie mit einbezogen. Daher eignet es sich besonders gut als Energieschätzer für Myonen.

Eine Anpassung der Einstellungen der Software TRUEE an den verwendeten Datensatz liefert die in Tabelle 6.2 zusammengefassten Einstellungen. Mit den drei gewählten Attributen wird eine Entfaltung des Energiespektrums durchgeführt. Eine Zusam-

## 6. Entfaltung des Energiespektrums

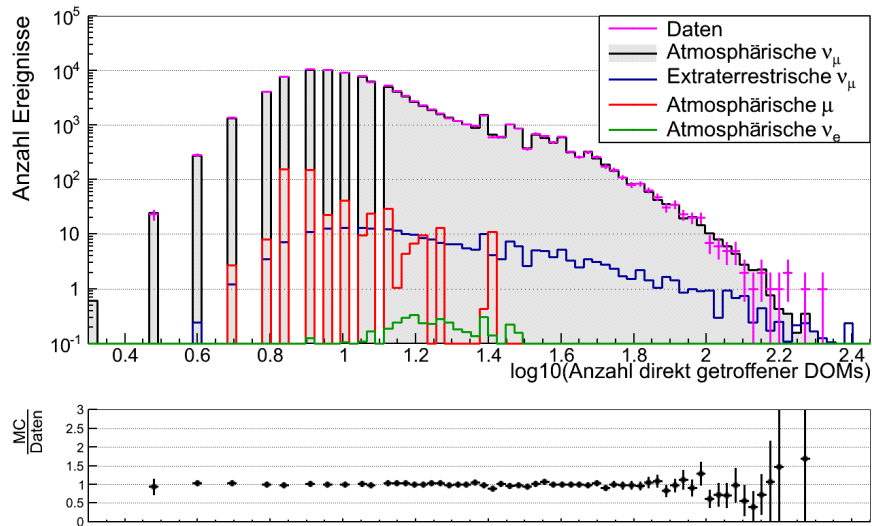


(a) Die Verteilung der Daten wird durch die MC hinreichend gut beschrieben. Es wird ebenfalls deutlich, dass der vorhandene Untergrund erstens nicht im Bereich der Hochenergie liegt, und zweitens nur Bins populiert, in denen er gegenüber dem Signal der Myon-Neutrinos stark subdominant ist. Die wellige Struktur des Attributs ergibt sich aus dem diskreten Abstand der Strings von etwa 125 m.

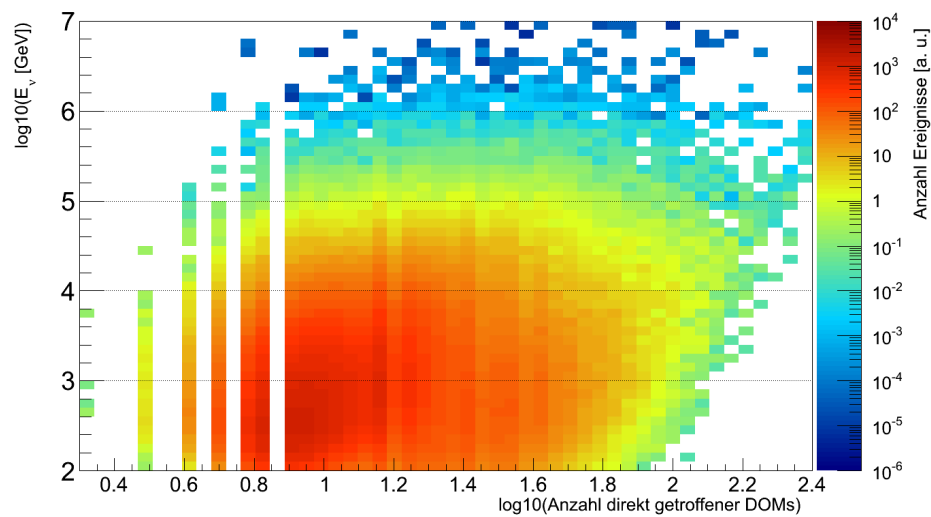


(b) LängeC korreliert stark mit der Energie der Myon-Neutrinos. Es ist ebenfalls erkennbar, dass das Attribut eine geringe Auflösung besitzt. Das Attribut ist gerade für Ereignisse mit kurzen Trajektorien ein verlässlicher Energieschätzer.

**Abbildung 6.1.:** LängeC als Entfaltungsobservable.



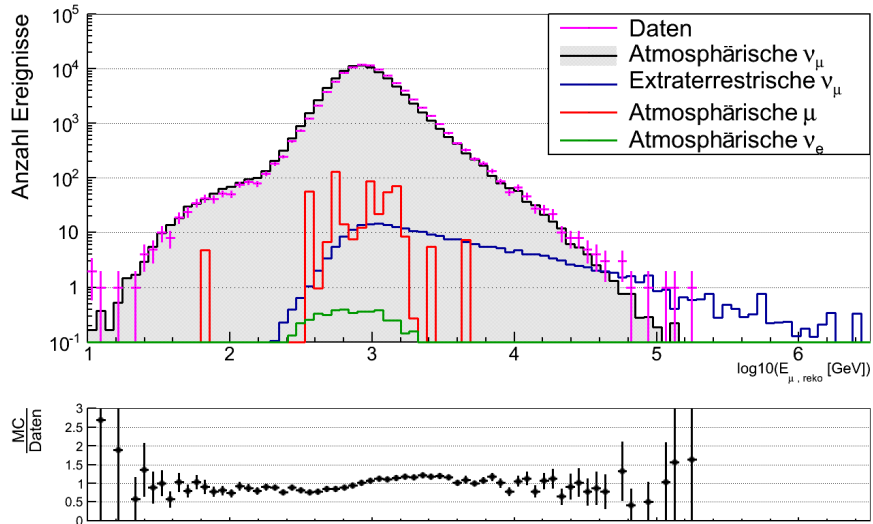
(a) Die Verteilung der Daten werden hinreichend gut von den MC beschrieben. Es wird ebenfalls deutlich, dass der vorhandene Untergrund erstens nicht im Bereich der Hochenergie liegt und zweitens nur Bins populiert, in denen er gegenüber dem Signal der Myon-Neutrinos stark subdominant ist. Die Lücken in der Verteilung des Attributs ergeben sich daraus, dass hier der Logarithmus einer ganzen Zahl gebildet wird.



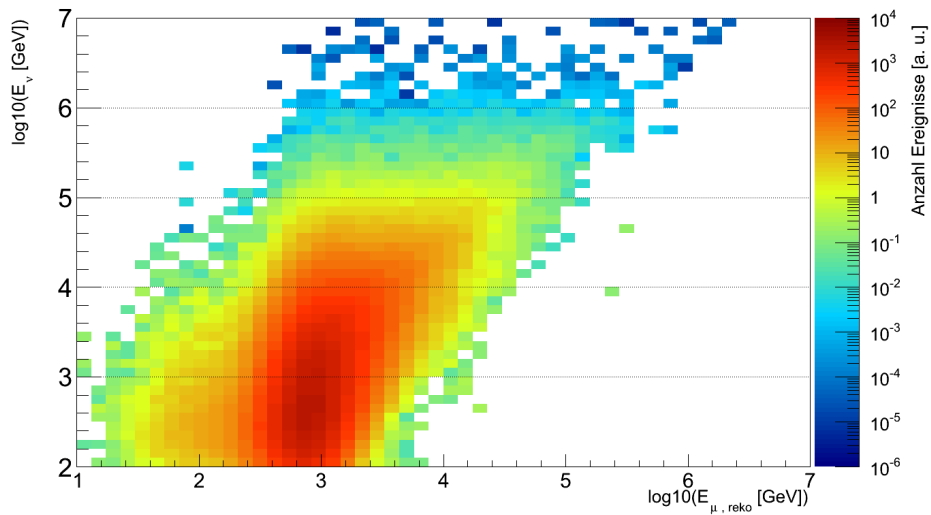
(b) DOMsC korreliert stark mit der Energie der Myon-Neutrinos. Es ist ebenfalls erkennbar, dass das Attribut eine geringe Auflösung besitzt.

**Abbildung 6.2.:** DOMsC als Entfaltungsobservable.

## 6. Entfaltung des Energiespektrums

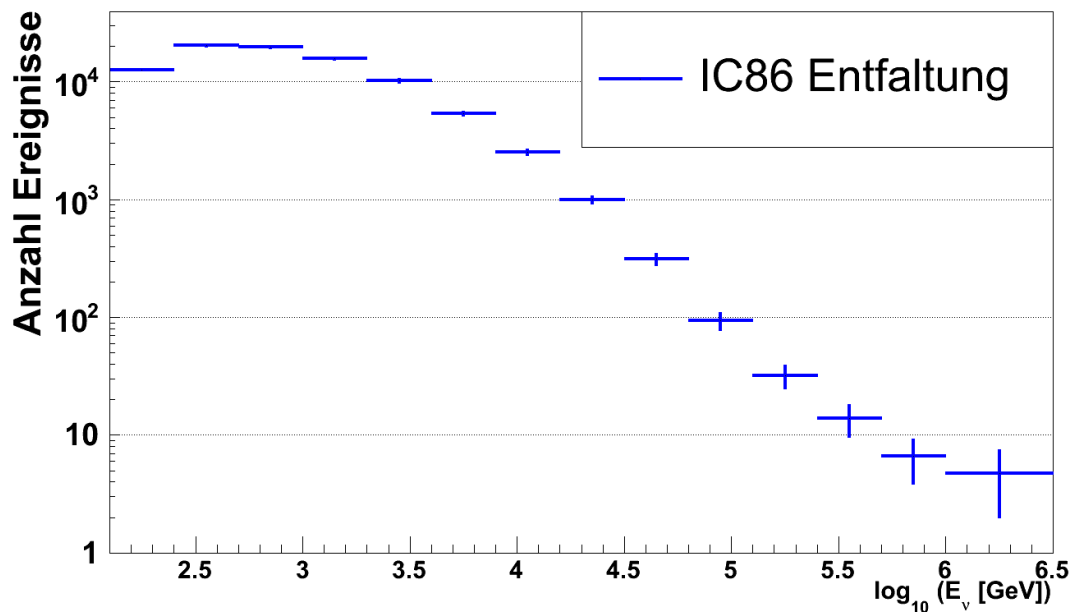


(a) Die Verteilung der Daten wird hinreichend gut von den MC beschrieben. Es ist eine leichte Verschiebung zwischen Daten und MC zu erkennen. Dieser Effekt wird durch eine nicht optimal eingestellte DOM-Effizienz verursacht. Die daraus resultierende Unsicherheit wird in der Entfaltung berücksichtigt.



(b)  $\text{MuEX}_E$  korreliert stark mit der Energie der Myon-Neutrinos. Im Vergleich zu den anderen Entfaltungsobservablen ist die Varianz gerade im Bereich der höchsten Energien gering.

**Abbildung 6.3.:**  $\text{MuEX}_E$  als Entfaltungsobservable.



**Abbildung 6.4.:** Entfaltetes Energiespektrum ohne systematische Fehler und Akzeptanzkorrektur. Die Abbildung zeigt, dass alle Entfaltungsbins populiert sind. Dies ist insbesondere für die letzten Bins von Bedeutung, da durch die große Anzahl an entfaltenen Ereignissen eine valide Entfaltung möglich ist.

menfassung der Ergebnisse findet sich in Tabelle 6.1. Die grafische Darstellung des Ergebnisses findet sich in Abbildung 6.4. Das Ergebnis ist die Anzahl der Ereignisse, die vom Algorithmus in jedes Bin entfaltet wird. Es ist dabei von großer Bedeutung, dass alle Bins ausreichend populiert sind, da die Entfaltung für Bineinträge  $< 1$  einer besonderen Verifikation bedarf. Dies ist in dieser Analyse allerdings nicht der Fall, da nach der Entfaltung jedes Bin mehr als ein Ereignis enthält, wie in Tabelle 6.1 deutlich wird.

Die bisherigen Ergebnisse beschreiben das von IceCube gemessene Energiespektrum. Die Ergebnisse sind jedoch noch mit der Akzeptanz des Detektors gefaltet. Ein direkter Vergleich mit Vorhersagen für den Fluss ist daher nicht ohne weiteres möglich. Es bietet sich daher an, das Energiespektrum um die Akzeptanz des Detektors mittels MC zu bereinigen. Durch Ausführen der Akzeptanzkorrektur ergibt sich der in Abbildung 6.5 gezeigte Fluss. Eine ausführliche Beschreibung der Vorgehensweise in der Akzeptanzkorrektur findet sich in [Ruh13]. Es ist notwendig zu testen, wie stabil die erzielten Ergebnisse sind. Zu diesem Zweck wird die Entfaltung validiert. Innerhalb der Validierung werden ebenfalls die systematischen Fehler abgeschätzt. Diese Vorgehensweise wird im folgenden Abschnitt beschrieben.

## 6. Entfaltung des Energiespektrums

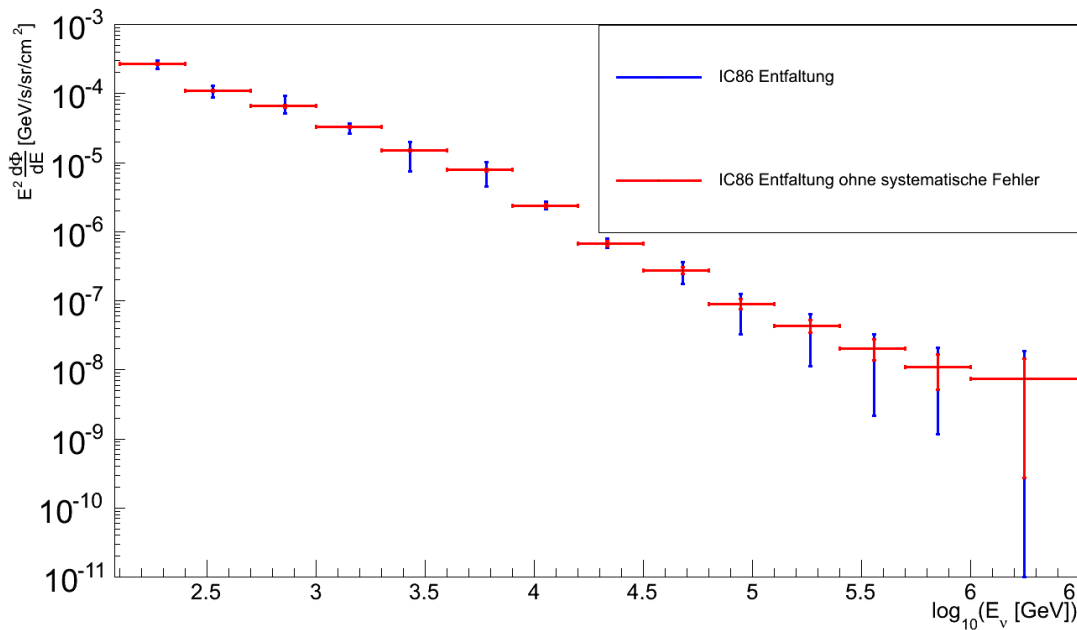


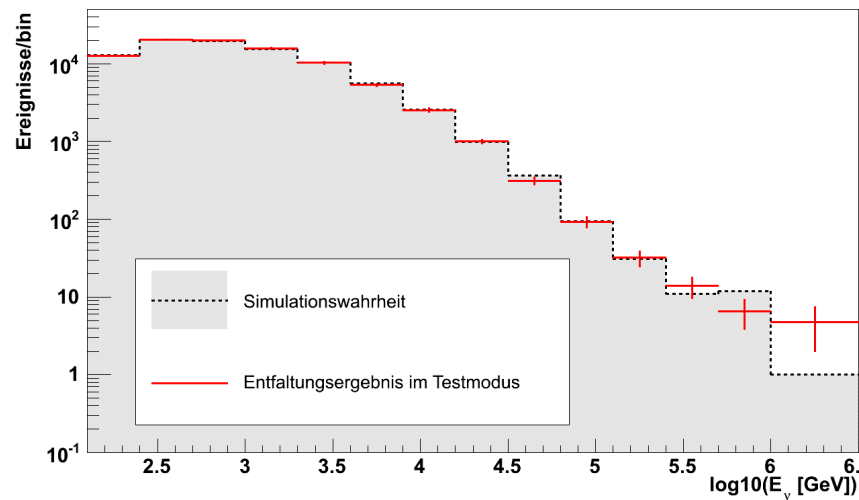
Abbildung 6.5.: Entfaltetes Energiespektrum mit und ohne systematische Fehler.

### 6.2. Validierung der Entfaltung

Damit sichergestellt werden kann, dass die Entfaltung einwandfrei funktioniert hat, kann der Testmode von TRUEE genutzt werden. In diesem Modus wird statt echter Daten ein MC Datensatz entfaltet, um die Abweichung des Entfaltungsergebnisses von der Wahrheit einschätzen zu können. Ein zufällig gewähltes Ergebnis wird in Abbildung 6.6 dargestellt. Das Entfaltungsergebnis stimmt in den meisten Bins gut mit der Wahrheit überein. In den letzten Bins sind Unterschiede zu erkennen. Es stellt sich daher die Frage, ob diese Unterschiede rein auf Grundlage von Fluktuationen in diesem speziellen Datensatz entstanden sind oder ob die Entfaltung systematisch falsche Ergebnisse liefert.

Um diese Frage beantworten zu können, wird der Pullmode von TRUEE genutzt. Insgesamt werden 1 000 unterschiedliche MC Datensätze entfaltet. Für jedes Energiebin wird der Abstand der Entfaltungsergebnisse zur Wahrheit in Einheiten des Fehlers in ein Histogramm eingetragen. Wenn die Entfaltung einwandfrei funktioniert, ergibt sich für jedes Bin eine Normalverteilung um Null mit RMS von 1. In Abbildung 6.7 findet sich ein beispielhaft ausgewähltes Ergebnis für ein Bin, eine vollständige Übersicht findet sich im Anhang in den Abbildungen A.9 und A.10. Die überwiegende Anzahl der Ergebnisse entsprechen einer Normalverteilung um 0 mit einem RMS von 1 und die





**Abbildung 6.6.:**

Im Testmode von TRUÉE werden MC entfaltet. Für den genutzten MC Datensatz ist die wahre Verteilung in den Bins bekannt, daher kann der Unterschied zum Entfaltungsergebnis verglichen werden.

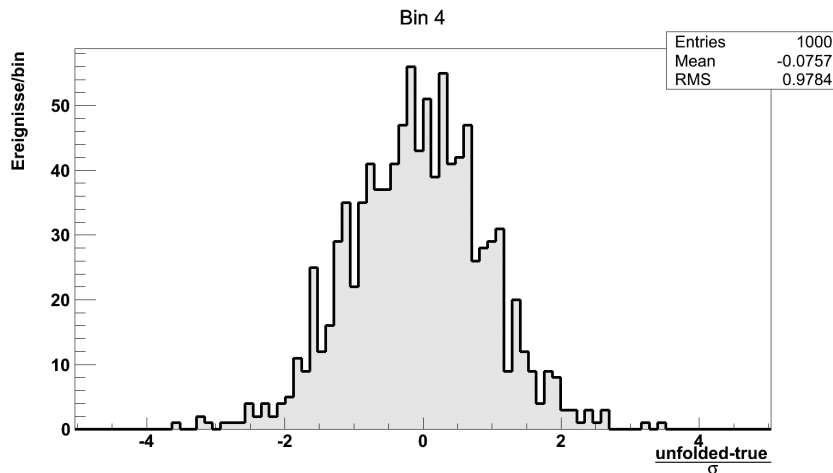
beobachteten Abweichungen sind klein. Es ist daher davon auszugehen, dass die Entfaltung valide Ergebnisse liefert. [MDK<sup>+</sup>13, Sch14]

Einige Werte weisen eine geringe Verschiebung der Schwerpunkte zu Werten kleiner null auf. Dieses Verhalten deutet darauf hin, dass die Entfaltung den Wert systematisch leicht unterschätzt. Die Abweichungen sind im Verhältnis zu dem systematischen Fehler klein. Der Einfluss der Abweichungen auf das Entfaltungsergebnis kann daher ignoriert werden. Insgesamt ergibt sich, dass die Entfaltung wie gewünscht funktioniert.

### 6.3. Bestimmung der systematischen Fehler

Mit dem Pullmode ergibt sich über die Validierung der Entfaltung ebenfalls eine ideale Möglichkeit die systematischen Fehler der Entfaltung zu berechnen. Zu diesem Zweck werden die in Tabelle 6.3 beschriebenen systematischen Signal MC genutzt. Diese Datensätze unterscheiden sich in jeweils einer Eigenschaft von den für die eigentliche Entfaltung verwendeten MC, sie eignen sich daher dafür den Einfluss dieser Eigenschaft auf das Entfaltungsergebnis zu quantifizieren. Für die Berechnung der systematischen Fehler werden die Datensätze aus Tabelle 6.3 in der Entfaltung wie Daten behandelt. Die Entfaltung selber bleibt vollkommen unverändert.

## 6. Entfaltung des Energiespektrums



**Abbildung 6.7.:**

Für 1 000 MC Entfaltungen wird der Abstand zwischen der wahren Anzahl von Ereignissen in einem Bin und die Anzahl an Ereignissen, die in das Bin entfaltet wurden, in Einheiten des Fehlers aufgetragen. Für eine ideale Entfaltung ergibt sich eine Normalverteilung um null mit einer Breite von eins. Ist die Verteilung zu negativen Werten verschoben, bedeutet dies eine systematische Unterschätzung des Wertes. Ist die Verteilung zu positiven Werten verschoben, bedeutet dies eine systematische Überschätzung.

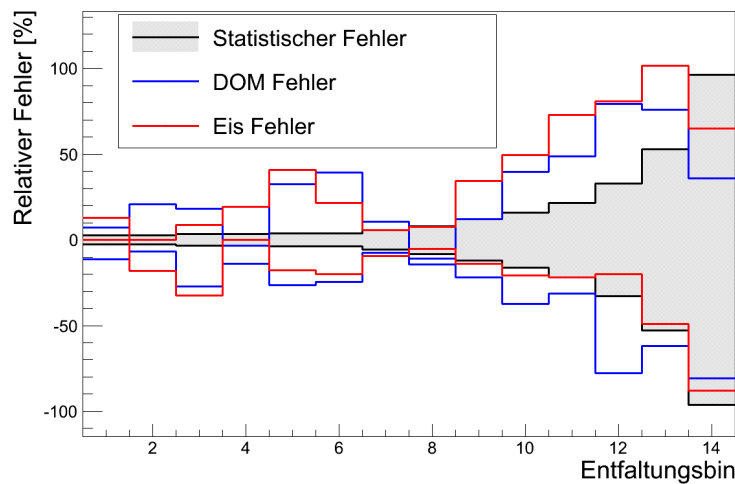
Die Wahl der Datensätze zur Ermittlung des systematischen Fehlers ist identisch zu [Sch14]. Dies hat den Vorteil, dass die beiden Spektren durch gleiche Fehlerrechnungen besser vergleichbar sind. Die Datensätze lassen sich in zwei Klassen zusammenfassen.

Die erste Klasse beschreibt eine veränderte Effizienz, der in den DOMs verbauten Photoelektronenvervielfacher. Die Klasse umfasst die Datensätze Dom1 mit einer um 10% erhöhten Effizienz und Dom2 mit einer um 10% verringerten Effizienz. Die Effizienz der Photoelektronenvervielfacher verändert effektiv die Anzahl von Photoelektronen, die während eines Ereignisses detektiert werden. Der gleiche Effekt entsteht durch eine Veränderung des Myon-Nukleon Wirkungsquerschnitts, daher lassen sich durch diese Datensätze beide Effekte zusammen behandeln.

**Tabelle 6.3.:**

Die verwendeten systematischen Datensätze jeweils mit einer kurzen Beschreibung des Unterschieds zu den Standard MC.

Bezeichnung	Besonderheit
Dom1	DOM Effizienz +10%
Dom2	DOM Effizienz -10%
Eis1	Photon Streuung +10%
Eis2	Photon Absorption +10%
Eis3	Photon Absorption und Streuung -7,1%



**Abbildung 6.8.:**

Systematische und statistische Fehler für jedes der vierzehn Entfaltungsbins. Die großen systematischen Fehler der Bins fünf und sechs lassen sich durch die maximale Steigung des Zusammenhangs zwischen Myonenergie und Energieverlust pro Meter, beschrieben in Abbildung 3.5(b), erklären.

Die zweite Klasse beschreibt eine Veränderung der Eiseigenschaften, was sich in einer veränderten Absorption und Streuung von Photoelektronen ausdrückt. Zu dieser Klasse gehören die Datensätze Eis1 mit einer um 10% erhöhten Streuung der Photoelektronen, Eis2 mit einer um 10% erhöhten Absorption der Photoelektronen und Eis3 mit einer um 7,1% verringerten Streuung und Absorption der Photoelektronen. Diese Datensätze werden genutzt, um zu berücksichtigen, dass das von IceCube genutzte Detektionsmedium ein nicht vollkommen homogenes natürliches Medium ist. Die Absorptions- und Streulänge des Eises wurde umfassend untersucht, die Eigenschaften des Eises unterliegen jedoch Schwankungen. Um die aus diesen Schwankungen resultierenden Unsicherheiten zu berücksichtigen, werden die beschriebenen Datensätze genutzt. [LMW<sup>+</sup>07, A<sup>+</sup>06]

Im Pullmode kann durch die Verschiebung des Schwerpunktes der Pullverteilung, siehe Abbildung 6.7, quantifiziert werden, wie stark der Einfluss eines der systematischen Datensätze ist. Für jeden der verwendeten Datensätze ergibt sich daraus eine Abweichung nach oben oder nach unten. In der Summe ergeben sich daher asymmetrische Fehler.

Eine Auflistung aller Fehler findet sich in Tabelle 6.4. Die resultierenden Unsicherheiten lassen sich anhand ihres Ursprungs zusammenfassen. Dabei können sowohl die negativen als auch die positiven Fehler jeweils quadratisch addiert werden. Die sich daraus ergebende Verteilung der Unsicherheiten findet sich in Abbildung 6.8. Durch

## 6. Entfaltung des Energiespektrums

**Tabelle 6.4.:**

In der ersten Spalte ist die Binnummer angegeben. In der zweiten Spalte findet sich der statistische Fehler. Die folgenden Spalten beschreiben die, aus den in Tabelle 6.3 beschriebenen Datensätzen, resultierenden systematischen Fehler. Die letzten beiden Spalten geben die Summe der positiven und negativen Unsicherheiten in jedem Bin der Entfaltung an.

Bin	Stat. Fehler [%]	Dom1 [%]	Dom2 [%]	Eis1 [%]	Eis2 [%]	Eis3 [%]	$\sum$ Positiv [%]	$\sum$ Negativ [%]
1	2,6	-11,4	7,1	4,9	3,9	11,0	14,6	11,3
2	2,6	20,9	-6,9	-2,8	-7,1	-16,4	19,2	19,3
3	3,5	18,3	-27,2	8,8	-3,1	-32,3	21,7	41,3
4	3,5	-14,1	-3,5	9,9	11,5	11,8	20,4	13,3
5	3,8	-26,4	32,6	-17,6	3,9	40,1	50,1	32,8
6	3,8	-24,6	39,3	-18,1	-7,9	21,5	42,0	31,5
7	5,6	-7,4	10,5	-7,9	-5,2	5,8	11,8	13,8
8	8,1	-14,5	-10,9	-5,2	6,3	4,3	12,0	20,3
9	12,0	-22,0	12,2	-14,0	10,2	32,0	37,0	29,3
10	16,1	-37,4	39,6	-20,9	3,9	49,2	63,4	40,0
11	21,7	-31,5	48,8	-21,7	3,7	72,6	74,1	46,6
12	32,8	-77,8	79,2	-19,6	-2,9	80,9	89,6	60,6
13	52,9	-62,0	75,8	-26,5	-41,2	101,5	89,2	93,1
14	96,2	-80,1	35,8	-64,4	-60,0	65,2	104,5	151,8

die hohe Anzahl an Ereignissen im Bereich der niedrigen bis mittleren Energien ergibt sich ein geringer statistischer Fehler, der mit abnehmender Anzahl von Ereignissen stark zunimmt. Generell ist festzustellen, dass jedes Bin einen dominanten systematischen Fehler aufweist.

Wie in Abbildung 6.8 zu erkennen ist, ergeben sich für die Bins fünf und sechs verhältnismäßig große systematische Fehler. Ein ähnliches Verhalten wurde bereits in [Sch14] beschrieben. Die Fehler lassen sich sehr wahrscheinlich auf den in Abbildung 3.5(b) beschriebenen Zusammenhang zwischen Myonenergie und Energieverlust pro Meter zurückführen. Der Zusammenhang zwischen den beiden Größen ist nicht linear. Es zeigt sich, dass Ereignisse aus dem Bereich der größten Steigung für die großen systematischen Fehler verantwortlich sind. Dies erklärt sich daher, dass in diesem Bereich eine leichte Veränderung des Energieverlustes pro Meter in einer starken Änderung der rekonstruierten Myonenergie niederschlägt. Genau eine solche Veränderung des Energieverlustes ist in den MC modelliert, welche zur Bestimmung des systematischen Fehlers genutzt wurden. Daher ist die in Abbildung 6.8 dargestellte Struktur der Entfaltung zu erwarten.

## 6.4. Entfaltung des Energiespektrums in verschiedenen Zenitbereichen

**Tabelle 6.5.:**  
Definition der drei Zenitbereiche für die Entfaltung der drei Winkelbereiche.

	Zenitbereich
1	$86^\circ \leq \theta \leq 100^\circ$
2	$100^\circ \leq \theta \leq 120^\circ$
3	$120^\circ \leq \theta \leq 180^\circ$

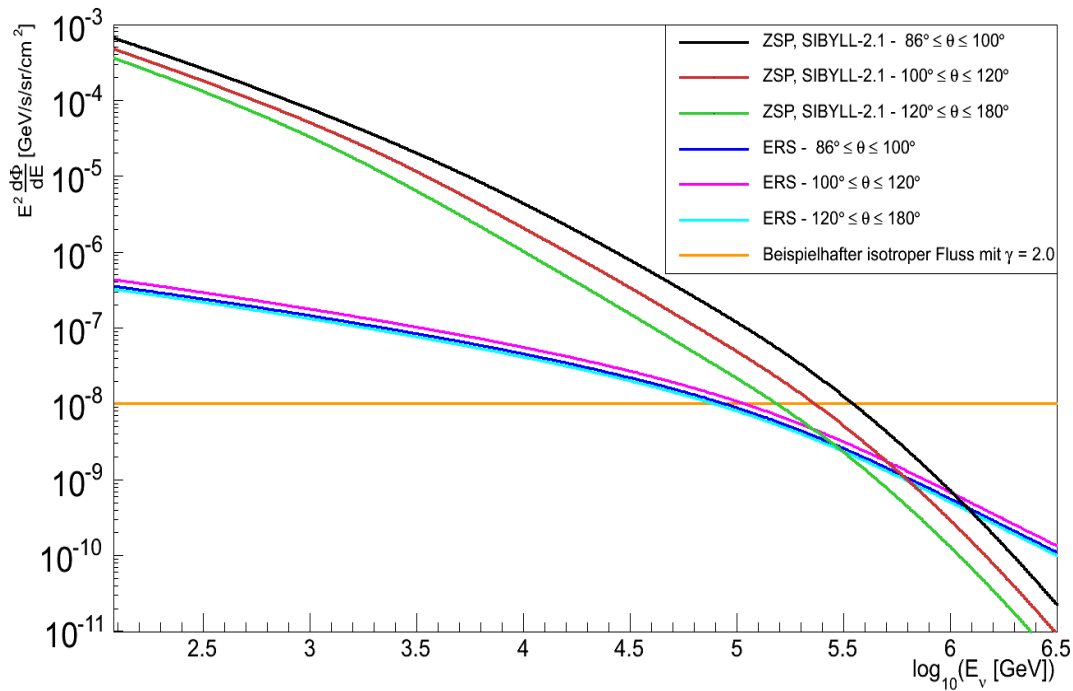
Die Abhängigkeit des Flusses von konventionellen, prompten Neutrinos und extraterrestrischen Neutrinos ist unterschiedlich. Der Fluss der konventionellen Neutrinos beispielsweise weist eine deutliche Abhängigkeit vom Kosinus des betrachteten Zenitwinkels auf, wohingegen ein isotroper Fluss von hochenergetischen extraterrestrischen Neutrinos unabhängig vom beobachteten Zenitwinkel ist. Daher ist das Verhalten des gemessenen Flusses in verschiedenen Zenitbereichen von besonderem Interesse. In Abbildung 6.9 wird die Abhängigkeit der unterschiedlichen Flüsse vom betrachteten Winkelbereich

illustriert. [Gai90, ERS08]

Die Entfaltung der unterschiedlichen Zenitbereiche erfolgt analog zur Entfaltung des Spektrums im vollen Zenitbereich. Durch die Eingrenzung des betrachteten Zenitbereiches verringert sich die Anzahl, der zur Verfügung stehenden Ereignisse. Das Fehlen von Ereignissen zeigt sich insbesondere in gering populierte Energiebins. In Abbildung 6.4 zeigt sich, dass gerade die Bins im Bereich der Hochenergie nur eine geringe Anzahl an entfaltenen Ereignissen aufweisen. Daher kann die Entfaltung der Zenitbereiche nur bis zu einer geringeren maximalen Energie durchgeführt werden. Insbesondere ergibt sich aus der in Abbildung 2.8 dargestellten Absorptionswahrscheinlichkeit für Neutrinos beim Durchqueren der Erde, dass nicht aus allen Zenitbereichen gleich viele hochenergetische Ereignisse erwartet werden.

Die gewählten Zenitbereiche wurden so gewählt, dass in jedem der Bereiche jeweils etwa 30 000 Ereignisse enthalten sind. Die Bereiche werden in Tabelle 6.5 zusammengefasst. Mit den definierten Winkelbereichen können die drei sich ergebenden Datensätze entfaltet werden. Dabei ist jedoch zu beachten, dass sich die Datensätze unterscheiden und die Entfaltungseinstellungen von TRUEE für jeden Datensatz einzeln optimiert werden müssen. Die genutzten Einstellungen sind in Tabelle 6.6 aufgeführt. Vergleicht man diese Einstellungen mit den Einstellungen für die Entfaltung im vollen Winkelbereich aus Tabelle 6.2, ist zu erkennen, dass sich die Einstellungen nur wenig unterscheiden. Die Unterschiede liegen in der Anzahl der Bins und in der Anzahl der Freiheitsgrade.

## 6. Entfaltung des Energiespektrums



**Abbildung 6.9.:**

Der dargestellte Fluss der konventionellen Neutrinos basiert auf dem hadronischen Interaktionsmodell SYBILL-2.1 [FTD12] und dem Modell des Primärflusses ZSP [A<sup>+</sup>11c]. Der Fluss zeigt eine deutliche Abhängigkeit vom betrachteten Winkelbereich. Der Fluss der prompten Neutrinos basiert auf dem ERS Modell [ERS08]. Für dieses Modell zeigt sich ebenfalls eine Abhängigkeit vom betrachteten Winkelbereich, die jedoch unterschiedlich und geringer ausgeprägt ist, als die des Flusses der konventionellen Neutrinos. Ebenfalls ist ein beispielhafter isotroper Fluss eingezeichnet, welcher keine Abhängigkeit vom betrachteten Winkelbereich aufweist. Die Flüsse wurden in den in Tabelle 6.5 beschriebenen Winkelbereichen dargestellt.

Mit den beschriebenen Einstellungen kann die Entfaltung ausgeführt werden. Die sich daraus ergebenden Ergebnisse werden in den Tabellen 6.7 - 6.9 dargestellt. Eine Darstellung der Entfaltungen ohne systematische Fehler findet sich in Abbildung 6.10, die entsprechende Abbildung mit systematischen Fehlern ist Abbildung 6.11. In den Abbildungen ist zu erkennen, dass sich für den Winkelbereich  $100^\circ \leq \theta \leq 120^\circ$  für die beiden letzten Bins mit null verträgliche Fehler ergeben.

## 6.4. Entfaltung des Energiespektrums in verschiedenen Zenitbereichen

**Tabelle 6.6.:**

Einstellungen für die drei Zenitbereiche, mit denen die Entfaltungssoftware TRUEE betrieben wird. Die Unterschiede zu den Einstellungen im vollen Zenitbereich aus Tabelle 6.2 liegen hauptsächlich in einer geringeren maximalen Energie.

Parameter	Wert	Wert	Wert
	$86^\circ \leq \theta \leq 100^\circ$	$100^\circ \leq \theta \leq 120^\circ$	$120^\circ \leq \theta \leq 180^\circ$
Knoten	20	20	20
Freiheitsgrade	7	7	6
Binning LängeC	10	10	10
Binning DOMsC	10	10	10
Binning MuEX <sub>E</sub>	10	10	10
Anzahl Bins	11	11	10
E <sub>max</sub> [log(E <sub>ν</sub> )]	5,4	5,4	5,1

**Tabelle 6.7.:**

Entfaltungsergebnisse für den Winkelbereich  $86^\circ \leq \theta \leq 100^\circ$ . In der ersten Spalte ist die Binnummer angegeben. In der zweiten Spalte werden Binbreite und Grenzen beschrieben. In der dritten Spalte befindet sich der entfaltete Fluss, gewichtet mit dem Quadrat der Energie des  $x$ -Wertes. In der vierten Spalte wird die Unsicherheit des Flusses in Prozent angegeben. Die nächste Spalte gibt den  $x$ -Wert der Entfaltung wieder.

Bin	Binbreite [log <sub>10</sub> ( $\frac{E}{\text{GeV}}$ )]	E <sup>2</sup> × Fluss [ $\frac{\text{GeV}}{\text{sr s cm}^2}$ ]	Unsicherheit [%]	$x$ -Wert [GeV]
1	[2,10; 2,40]	$4,32 \times 10^{-4}$	+15,94 -9,22	$1,89 \times 10^{-2}$
2	[2,40; 2,70]	$1,93 \times 10^{-4}$	+22,25 -14,00	$3,62 \times 10^{-2}$
3	[2,70; 3,00]	$9,40 \times 10^{-5}$	+39,77 -48,18	$7,09 \times 10^{-2}$
4	[3,00; 3,30]	$5,54 \times 10^{-5}$	+8,75 -7,35	$1,41 \times 10^{-3}$
5	[3,30; 3,60]	$3,02 \times 10^{-5}$	+36,93 -13,46	$2,79 \times 10^{-3}$
6	[3,60; 3,90]	$1,23 \times 10^{-5}$	+27,48 -39,51	$5,57 \times 10^{-3}$
7	[3,90; 4,20]	$4,37 \times 10^{-6}$	+11,04 -26,86	$1,11 \times 10^{-4}$
8	[4,20; 4,50]	$1,34 \times 10^{-6}$	+17,49 -22,93	$2,21 \times 10^{-4}$
9	[4,50; 4,80]	$4,20 \times 10^{-7}$	+31,77 -40,81	$4,37 \times 10^{-4}$
10	[4,80; 5,10]	$1,60 \times 10^{-7}$	+44,86 -48,12	$8,71 \times 10^{-4}$
11	[5,10; 5,40]	$6,36 \times 10^{-8}$	+40,56 -75,71	$1,77 \times 10^{-5}$

## 6. Entfaltung des Energiespektrums

**Tabelle 6.8.:**

Entfaltungsergebnisse für den Winkelbereich  $100^\circ \leq \theta \leq 120^\circ$ . In der ersten Spalte ist die Binnummer angegeben. In der zweiten Spalte werden Binbreite und Grenzen beschrieben. In der dritten Spalte befindet sich der entfaltete Fluss, gewichtet mit dem Quadrat der Energie des  $x$ -Wertes. In der vierten Spalte wird die Unsicherheit des Flusses in Prozent angegeben. Die nächste Spalte gibt den  $x$ -Wert der Entfaltung wieder.

Bin	Binbreite [ $\log_{10}(\frac{E}{\text{GeV}})$ ]	$E^2 \times \text{Fluss}$ [ $\frac{\text{GeV}}{\text{sr s cm}^2}$ ]	Unsicherheit [%]	$x$ -Wert [GeV]
1	[2,10; 2,40]	$2,76 \times 10^{-4}$	+5,23 -7,70	$1,87 \times 10^{-2}$
2	[2,40; 2,70]	$1,28 \times 10^{-4}$	+27,23 -27,02	$3,61 \times 10^{-2}$
3	[2,70; 3,00]	$6,72 \times 10^{-5}$	+44,23 -32,44	$7,06 \times 10^{-2}$
4	[3,00; 3,30]	$3,61 \times 10^{-5}$	+20,70 -16,27	$1,40 \times 10^{-3}$
5	[3,30; 3,60]	$1,77 \times 10^{-5}$	+44,38 -59,76	$2,79 \times 10^{-3}$
6	[3,60; 3,90]	$6,51 \times 10^{-6}$	+39,09 -52,70	$5,55 \times 10^{-3}$
7	[3,90; 4,20]	$2,00 \times 10^{-6}$	+30,48 -12,08	$1,10 \times 10^{-4}$
8	[4,20; 4,50]	$6,46 \times 10^{-7}$	+30,29 -10,88	$2,20 \times 10^{-4}$
9	[4,50; 4,80]	$2,14 \times 10^{-7}$	+18,12 -37,30	$4,38 \times 10^{-4}$
10	[4,80; 5,10]	$6,71 \times 10^{-8}$	+40,12 -104,18	$8,77 \times 10^{-4}$
11	[5,10; 5,40]	$1,95 \times 10^{-8}$	+125,42 -162,15	$1,75 \times 10^{-5}$

**Tabelle 6.9.:**

Entfaltungsergebnisse für den Winkelbereich  $120^\circ \leq \theta \leq 180^\circ$ . In der ersten Spalte ist die Binnummer angegeben. In der zweiten Spalte werden Binbreite und Grenzen beschrieben. In der dritten Spalte befindet sich der entfaltete Fluss, gewichtet mit dem Quadrat der Energie des  $x$ -Wertes. In der vierten Spalte wird die Unsicherheit des Flusses in Prozent angegeben. Die nächste Spalte gibt den  $x$ -Wert der Entfaltung wieder.

Bin	Binbreite [ $\log_{10}(\frac{E}{\text{GeV}})$ ]	$E^2 \times \text{Fluss}$ [ $\frac{\text{GeV}}{\text{sr s cm}^2}$ ]	Unsicherheit [%]	$x$ -Wert [GeV]
1	[2,10; 2,40]	$1,84 \times 10^{-4}$	+4,05 -13,51	$1,88 \times 10^{-2}$
2	[2,40; 2,70]	$1,02 \times 10^{-4}$	+13,27 -14,43	$3,60 \times 10^{-2}$
3	[2,70; 3,00]	$4,71 \times 10^{-5}$	+29,41 -16,02	$7,07 \times 10^{-2}$
4	[3,00; 3,30]	$2,01 \times 10^{-5}$	+11,50 -13,09	$1,40 \times 10^{-3}$
5	[3,30; 3,60]	$9,23 \times 10^{-6}$	+45,86 -46,36	$2,78 \times 10^{-3}$
6	[3,60; 3,90]	$3,61 \times 10^{-6}$	+42,09 -56,53	$5,54 \times 10^{-3}$
7	[3,90; 4,20]	$1,02 \times 10^{-6}$	+29,08 -49,32	$1,10 \times 10^{-4}$
8	[4,20; 4,50]	$2,98 \times 10^{-7}$	+37,85 -20,02	$2,19 \times 10^{-4}$
9	[4,50; 4,80]	$1,13 \times 10^{-7}$	+55,28 -30,99	$4,33 \times 10^{-4}$
10	[4,80; 5,10]	$4,28 \times 10^{-8}$	+116,51 -51,26	$8,73 \times 10^{-4}$



## 6.4. Entfaltung des Energiespektrums in verschiedenen Zenitbereichen

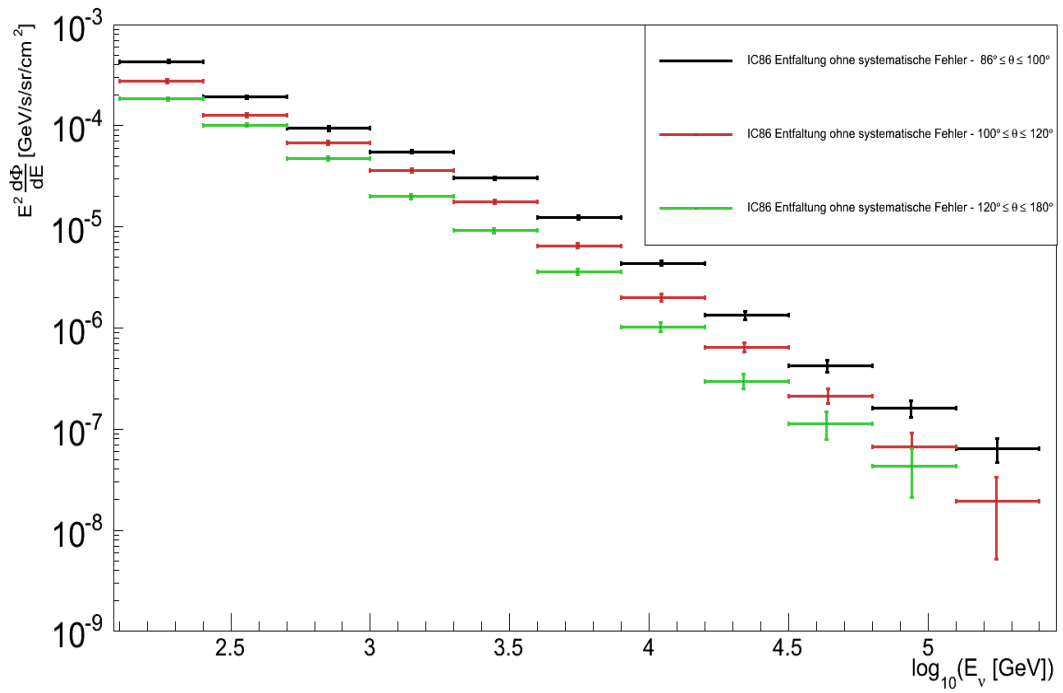


Abbildung 6.10.: Entfaltete Energiespektren in drei Zenitbereichen ohne systematische Fehler.

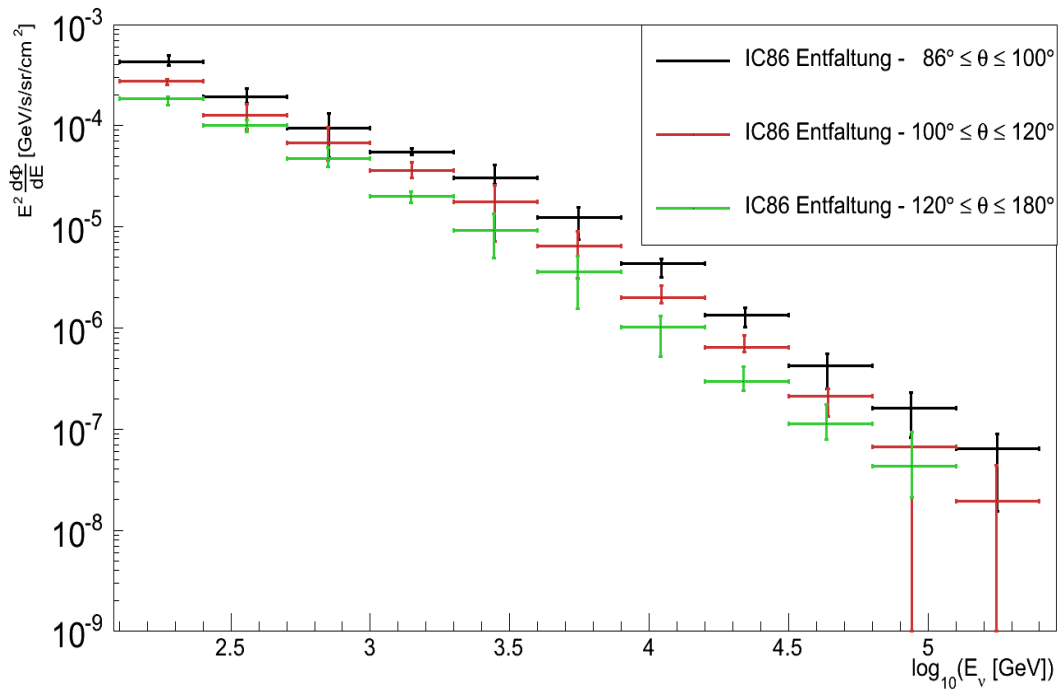


Abbildung 6.11.: Entfaltete Energiespektren in drei Zenitbereichen mit systematischen Fehlern.

## 6.5. Umgewichtung des Spektrums auf einen beliebigen Zenitbereich

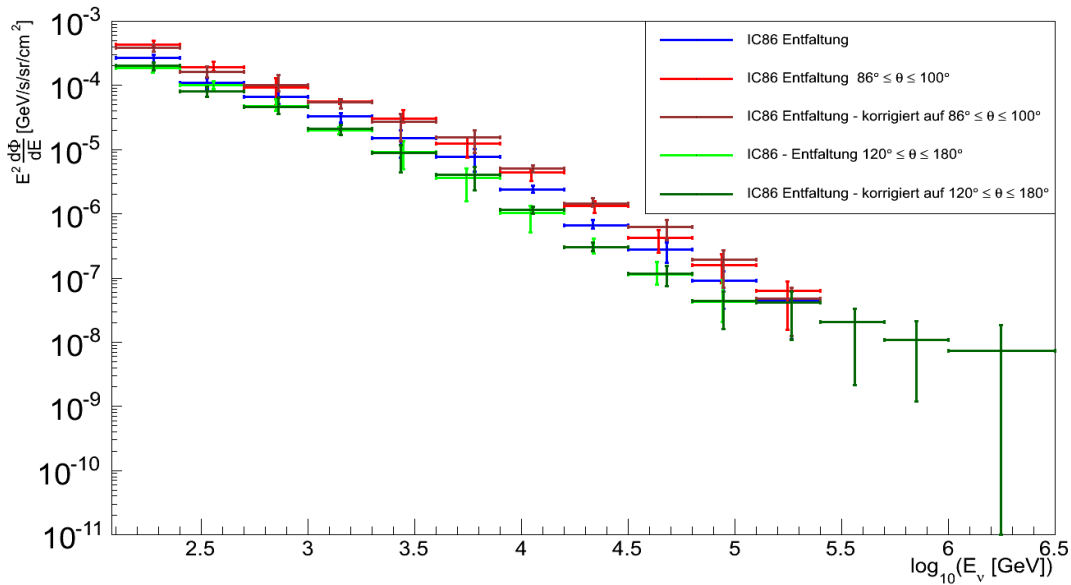


Abbildung 6.12.:

Umgewichtung des Spektrums im vollen Winkelbereich auf die Spektren in verschiedenen Winkelbereichen.

Wie im vorangegangenen Abschnitt beschrieben wird, zeigen die Flüsse, die zum entfalteten Spektrum beitragen, unterschiedliche Abhängigkeiten vom betrachteten Winkelbereich. Dies bedeutet, dass Messungen, die einen unterschiedlichen Winkelbereich beschreiben, nicht direkt vergleichbar sind. Es bietet sich daher an, diese Unterschiede zu korrigieren. Im Folgenden wird daher eine Methode beschrieben, welche die Möglichkeit bietet, den Einfluss des betrachteten Winkelbereichs modellabhängig zu korrigieren.

Für Modellvorhersagen des Flusses der konventionellen, prompten und hochenergetischen extraterrestrischen Neutrinos ist die Abhängigkeit vom betrachteten Winkelbereich  $\Delta\Theta_{\text{Messung}}$  bekannt. Es ist daher möglich, eine Kombination der Flüsse zu wählen, welche das Entfaltungsergebnis im betrachteten Winkelbereich der Messung hinreichend beschreibt. Der Fluss, der durch die Modelle vorhergesagt wird, sollte in diesem Fall die Datenpunkte der Entfaltung hinreichend beschreiben. Daraus ergibt sich der vorhergesagte Fluss  $\Phi_{\text{Messung}, i}$  für jeden der  $i$  Datenpunkte. Ebenfalls ist es möglich, für die gleichen Modelle eine Flussvorhersage für einen beliebigen anderen Winkel-

bereich  $\Delta\Theta_{\text{Ziel}}$  zu treffen und diese an den  $x$ -Werten der Datenpunkte auszuwerten, wodurch der Fluss für jeden Datenpunkt  $\Phi_{\text{Ziel},i}$  definiert wird. Aus dem Quotienten der beiden Flüsse ergibt sich ein energieabhängiger Korrekturfaktor  $\beta_i$  für jeden Datenpunkt der Entfaltung.

Mit dieser Methodik ist es möglich, das entfaltete Energiespektrum für den vollen Winkelbereich auf einen beliebigen anderen Messbereich zu korrigieren. Um zu zeigen, dass die Methode funktioniert, wird das Spektrum auf die entfalteten Spektren im Winkelbereich  $86^\circ \leq \theta \leq 100^\circ$  und  $120^\circ \leq \theta \leq 180^\circ$  korrigiert. Die Ergebnisse der Winkelkorrektur für das Spektrum des vollen Winkelbereiches auf die beiden anderen Spektren finden sich in Abbildung 6.12. Das verwendete Modell für den Fluss der konventionellen Neutrinos ist ZSP [A<sup>+</sup>11c] mit dem hadronischen Interaktionsmodell SYBILL-2.1 [AEG<sup>+</sup>09]. Für den prompten Fluss wird ERS [ERS08] verwendet. Der isotrope Fluss hochenergetischer Neutrinos entspricht dem besten Fit an die Entfaltungsergebnisse aus Abschnitt 7.3. Alle Datenpunkte der korrigierten Spektren sind mit denen der Entfaltung im entsprechenden Winkelbereich kompatibel. Aus diesem Ergebnis lässt sich schließen, dass die Methode in der Anwendung auf Daten funktioniert.

Um einen zukünftigen Vergleich einer Modellvorhersage zum Spektrum im vollen Winkelbereich zu erleichtern, wird das Spektrum für neun verschiedene Winkelbereiche  $90^\circ \leq \theta \leq \theta_{\text{max}}$  korrigiert. Die sich für die einzelnen Bins ergebenden Flüsse, finden sich in Tabelle 6.10. Eine Darstellung einer Auswahl von korrigierten Flüssen ist in Abbildung 6.13 dargestellt. Da sich durch die verwendete Methode die Zentralwerte und die relativen statistischen und systematischen Fehler des Flusses nicht ändern, ergeben sich diese Werte aus Tabelle 6.1.

## 6. Entfaltung des Energiespektrums

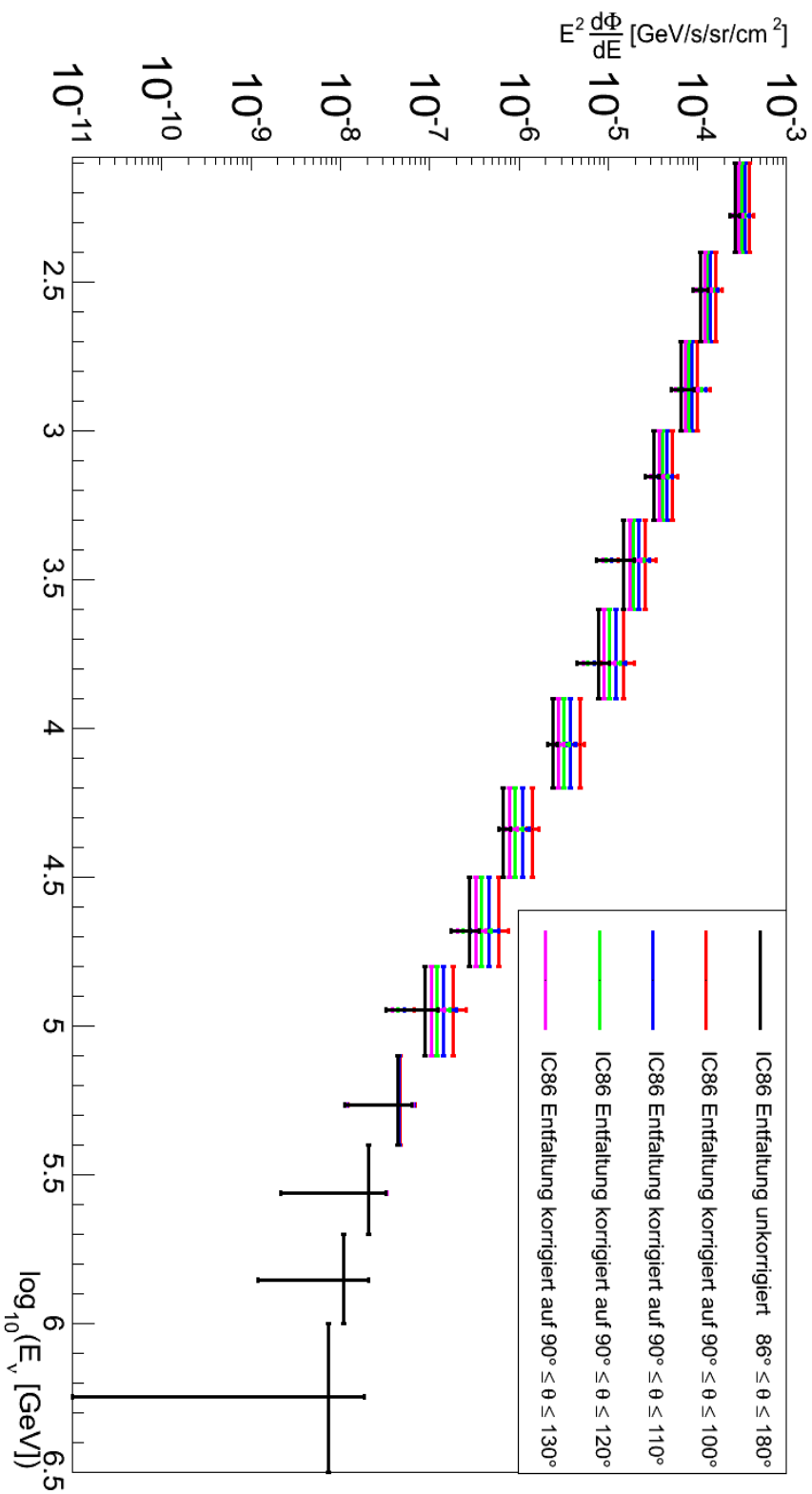


Abbildung 6.13.: Umgewichtung des Spektrums im vollen Winkelbereich auf die Spektren in vier Winkelbereichen.

**Tabelle 6.10.:** Umgewichteter Fluss in verschiedenen Zenitbereichen. Der beschriebene Fluss bezieht sich auf den Fluss von  $90^\circ \leq \theta \leq \theta_{\max}$ . Durch die verwendete Methode wird der Fluss auf einen Winkelbereich korrigiert, daher ergeben sich alle anderen Werte wie die relativen Unsicherheiten und Binbreiten aus den Ergebnissen des vollen Winkelbereiches in Tabelle 6.1.

Bin	$\theta_{\max} = 100^\circ$	$\theta_{\max} = 110^\circ$	$\theta_{\max} = 120^\circ$	$\theta_{\max} = 130^\circ$	$\theta_{\max} = 140^\circ$	$\theta_{\max} = 150^\circ$	$\theta_{\max} = 160^\circ$	$\theta_{\max} = 170^\circ$	$\theta_{\max} = 180^\circ$
	$E^2 \times \text{Fluss} \left[ \frac{\text{GeV}}{\text{sr s cm}^2} \right]$								
1	$3,80 \times 10^{-4}$	$3,39 \times 10^{-4}$	$3,11 \times 10^{-4}$	$2,91 \times 10^{-4}$	$2,78 \times 10^{-4}$	$2,68 \times 10^{-4}$	$2,62 \times 10^{-4}$	$2,58 \times 10^{-4}$	$2,57 \times 10^{-4}$
2	$1,60 \times 10^{-4}$	$1,41 \times 10^{-4}$	$1,29 \times 10^{-4}$	$1,20 \times 10^{-4}$	$1,14 \times 10^{-4}$	$1,10 \times 10^{-4}$	$1,07 \times 10^{-4}$	$1,05 \times 10^{-4}$	$1,05 \times 10^{-4}$
3	$9,98 \times 10^{-5}$	$8,77 \times 10^{-5}$	$7,93 \times 10^{-5}$	$7,34 \times 10^{-5}$	$6,91 \times 10^{-5}$	$6,61 \times 10^{-5}$	$6,42 \times 10^{-5}$	$6,31 \times 10^{-5}$	$6,27 \times 10^{-5}$
4	$5,29 \times 10^{-5}$	$4,59 \times 10^{-5}$	$4,09 \times 10^{-5}$	$3,74 \times 10^{-5}$	$3,49 \times 10^{-5}$	$3,31 \times 10^{-5}$	$3,20 \times 10^{-5}$	$3,13 \times 10^{-5}$	$3,11 \times 10^{-5}$
5	$2,61 \times 10^{-5}$	$2,21 \times 10^{-5}$	$1,93 \times 10^{-5}$	$1,74 \times 10^{-5}$	$1,61 \times 10^{-5}$	$1,51 \times 10^{-5}$	$1,45 \times 10^{-5}$	$1,42 \times 10^{-5}$	$1,41 \times 10^{-5}$
6	$1,49 \times 10^{-5}$	$1,21 \times 10^{-5}$	$1,03 \times 10^{-5}$	$9,09 \times 10^{-6}$	$8,29 \times 10^{-6}$	$7,76 \times 10^{-6}$	$7,41 \times 10^{-6}$	$7,22 \times 10^{-6}$	$7,15 \times 10^{-6}$
7	$4,82 \times 10^{-6}$	$3,83 \times 10^{-6}$	$3,21 \times 10^{-6}$	$2,81 \times 10^{-6}$	$2,54 \times 10^{-6}$	$2,37 \times 10^{-6}$	$2,26 \times 10^{-6}$	$2,19 \times 10^{-6}$	$2,17 \times 10^{-6}$
8	$1,40 \times 10^{-6}$	$1,09 \times 10^{-6}$	$9,08 \times 10^{-7}$	$7,90 \times 10^{-7}$	$7,12 \times 10^{-7}$	$6,60 \times 10^{-7}$	$6,27 \times 10^{-7}$	$6,09 \times 10^{-7}$	$6,03 \times 10^{-7}$
9	$5,95 \times 10^{-7}$	$4,60 \times 10^{-7}$	$3,79 \times 10^{-7}$	$3,28 \times 10^{-7}$	$2,95 \times 10^{-7}$	$2,73 \times 10^{-7}$	$2,59 \times 10^{-7}$	$2,51 \times 10^{-7}$	$2,48 \times 10^{-7}$
10	$1,83 \times 10^{-7}$	$1,43 \times 10^{-7}$	$1,19 \times 10^{-7}$	$1,05 \times 10^{-7}$	$9,50 \times 10^{-8}$	$8,86 \times 10^{-8}$	$8,45 \times 10^{-8}$	$8,22 \times 10^{-8}$	$8,15 \times 10^{-8}$
11	$4,72 \times 10^{-8}$	$4,57 \times 10^{-8}$	$4,47 \times 10^{-8}$	$4,42 \times 10^{-8}$	$4,38 \times 10^{-8}$	$4,35 \times 10^{-8}$	$4,34 \times 10^{-8}$	$4,33 \times 10^{-8}$	$4,33 \times 10^{-8}$
12	$2,05 \times 10^{-8}$	$2,05 \times 10^{-8}$	$2,05 \times 10^{-8}$	$2,05 \times 10^{-8}$	$2,05 \times 10^{-8}$	$2,05 \times 10^{-8}$	$2,05 \times 10^{-8}$	$2,05 \times 10^{-8}$	$2,05 \times 10^{-8}$
13	$1,09 \times 10^{-8}$	$1,09 \times 10^{-8}$	$1,09 \times 10^{-8}$	$1,09 \times 10^{-8}$	$1,09 \times 10^{-8}$	$1,09 \times 10^{-8}$	$1,09 \times 10^{-8}$	$1,09 \times 10^{-8}$	$1,09 \times 10^{-8}$
14	$7,33 \times 10^{-9}$	$7,33 \times 10^{-9}$	$7,33 \times 10^{-9}$	$7,33 \times 10^{-9}$	$7,33 \times 10^{-9}$	$7,33 \times 10^{-9}$	$7,33 \times 10^{-9}$	$7,33 \times 10^{-9}$	$7,33 \times 10^{-9}$



## 7. Analyse des entfalteteten Energiespektrums

Im vorherigen Kapitel wurde das Ergebnis und die Entstehung des entfalteteten Energiespektrums der von IceCube gemessenen Myon-Neutrinos vorgestellt. Das entfaltetete und akzeptanzkorrigierte Spektrum inklusive systematischer Fehler findet sich in Abbildung 6.5. Dieses Spektrum bildet den Kern der physikalischen Aussage der vorliegenden Arbeit.

Um das entfaltetete Spektrum besser beurteilen zu können, wird im folgenden Abschnitt das Spektrum mit einer Auswahl an relevanten Messergebnissen verglichen. Daran anschließend wird das Spektrum mit einer Auswahl an Modellen verglichen. Abschließend wird die Signifikanz des beobachteten Exzesses diskutiert.

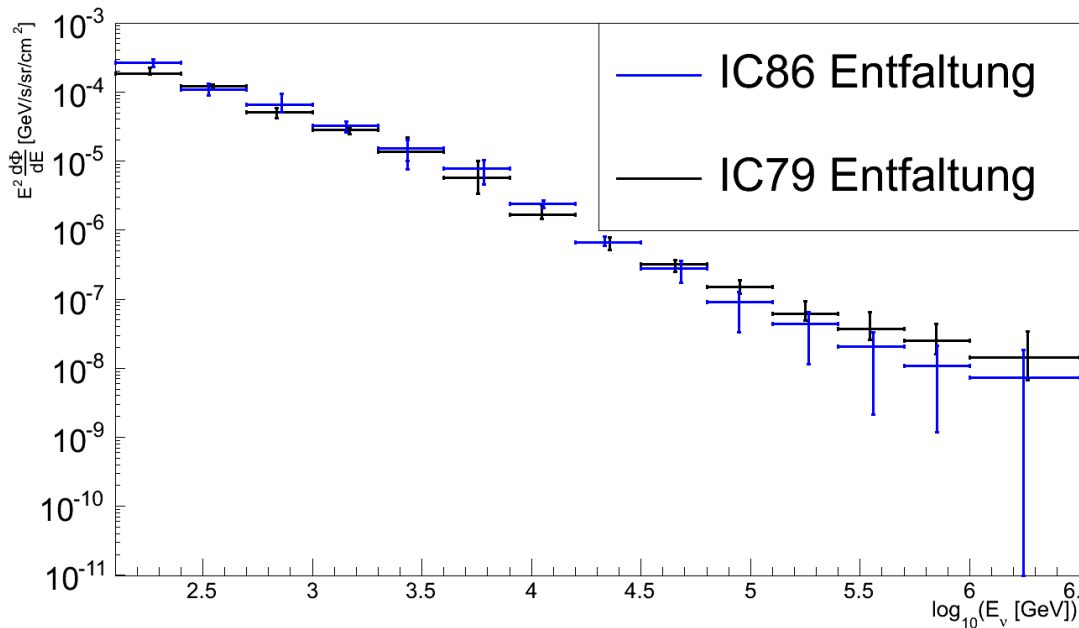
### 7.1. Vergleich des entfalteteten Energiespektrums mit anderen Messungen

Den Ergebnissen der IC79 Analyse von [Sch14] kommt eine besondere Bedeutung zu, da beide Analysen sich methodisch stark ähneln. Die Analysen wurden auf unterschiedlichen Daten und auf Grundlage verschiedener MC durchgeführt. Der entscheidende Unterschied ist, dass die vorliegende Analyse etwa 30 000 Neutrino-Ereignisse mehr nachweisen konnte und insbesondere in der Hochenergie eine deutlich gesteigerte Effizienz aufweist. Der Vergleich bietet daher eine Möglichkeit, die Validität der Analysen zu testen.

Die beiden Energiespektren sind in Abbildung 7.1 dargestellt. Die Flüsse in den einzelnen Bins weisen mindestens überschneidende Fehlerbalken auf. Eine Vielzahl der Datenpunkte ist im Rahmen der Fehler verträglich.

Es ist auffällig, dass die IC79 Analyse für Energien über 50 TeV einen konstant höheren Fluss misst. Der Unterschied ist durch die großen Fehler nicht signifikant. Diese

## 7. Analyse des entfalteten Energiespektrums



**Abbildung 7.1.:**

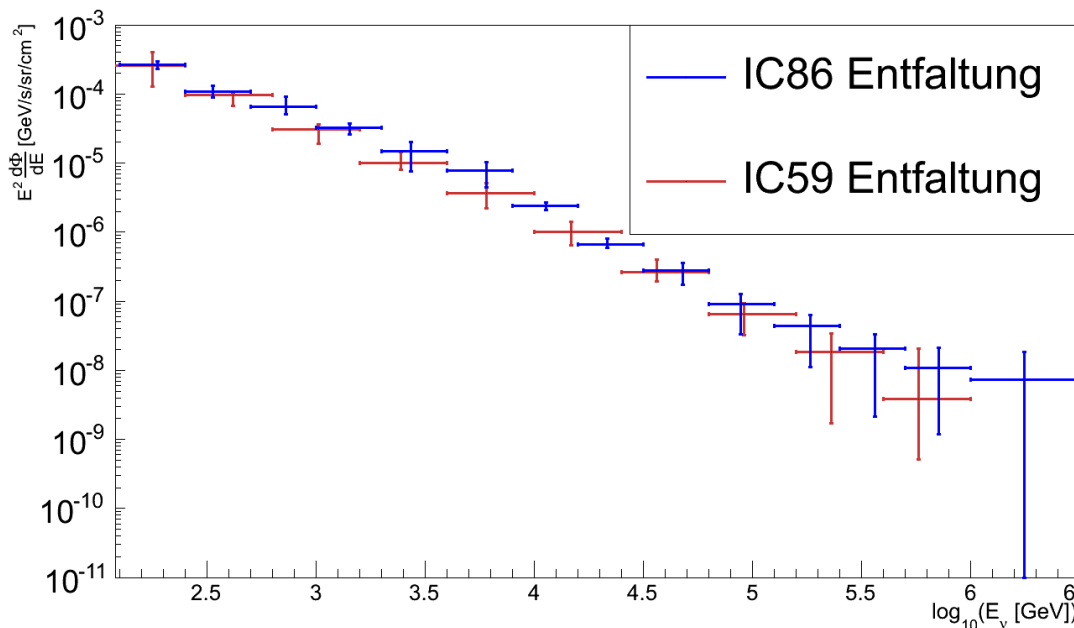
Vergleich der IC86 Entfaltung mit der IC79 Entfaltung [Sch14]. Alle Entfaltungspunkte weisen mindestens überlappende Fehlerbalken auf. Für hohe Energien größer 50 TeV liegt, der von IC79 gemessene Fluss konstant über dem von IC86 gemessenen Fluss.

Unterschiede zeigen sich auch in den, im Vergleich zur IC79 Analyse, wesentlich größeren Fehlern. Dieser Unterschied lässt sich durch die sehr geringen Ereigniszahlen im Bereich der Hochenergie erklären. In Abbildung 6.4 ist zu erkennen, dass die letzten Bins von nur wenigen Ereignissen populiert werden. Leichte Schwankungen im Rahmen der Fehler sind daher zu erwarten. Ein steter Wechsel von Über- und Unterfluktuationen ist durch die Regularisierung und ihrer Forderung nach Glätte jedoch unwahrscheinlich. Daher ist der Unterschied in Form einer konstanten Über- oder Unterfluktuation zu erwarten. Da die Datenpunkte im Bereich der Hochenergie jedoch im Rahmen der Fehler kompatibel sind, ist davon auszugehen, dass beide Analysen den gleichen Fluss im Bereich der Hochenergie messen, die Messung jedoch auf Grund des geringen Flusses eine große Varianz aufweist. Es ist daher zu schließen, dass die beiden Messungen kompatibel sind. Für weitere Tests bietet es sich an, die Analyse mit weiteren Entfaltungsergebnissen zu vergleichen.

Die methodisch ähnliche IC59 Entfaltung von [Ruh13] basiert auf einer wesentlich kleineren Detektorkonfiguration. Die kleinere Detektorkonfiguration wird besonders deutlich an der effektiven Fläche, dargestellt in Abbildung 5.18. Die Entfaltung war trotzdem in der Lage, das Energiespektrum der von IceCube gemessenen



## 7.1. Vergleich des entfalteten Energiespektrums mit anderen Messungen



**Abbildung 7.2.:**

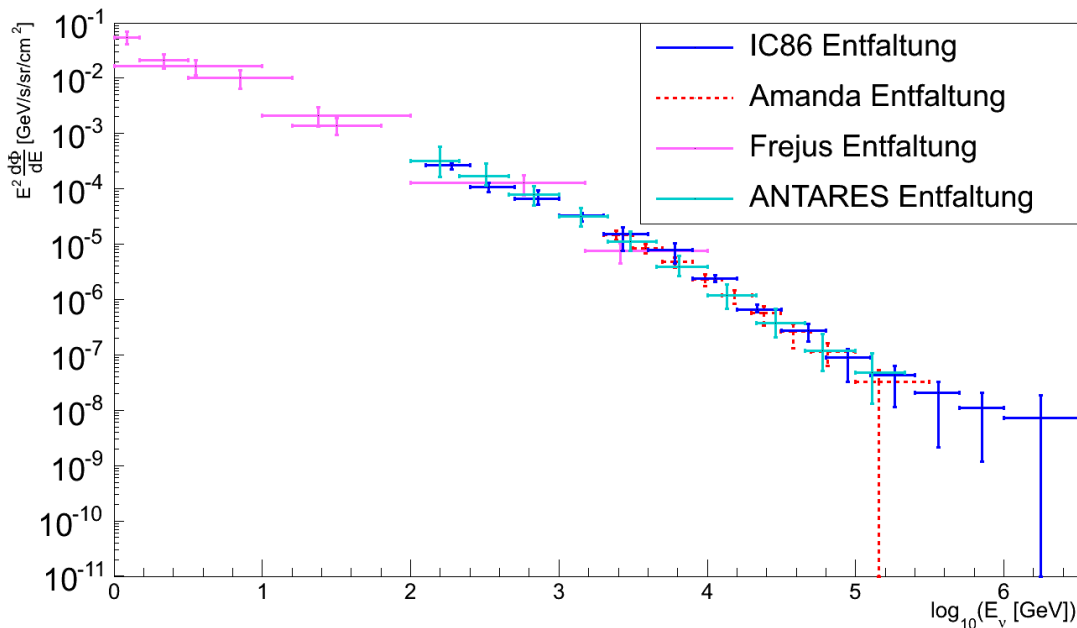
Vergleich der IC86 Entfaltung mit der IC59 Entfaltung [Ruh13]. Ein direkter Vergleich der einzelnen Messpunkte ist auf Grund des unterschiedlichen Binnings nicht direkt möglich. Es ist deutlich sichtbar, dass die beiden Ergebnisse kompatibel sind.

Myon-Neutrinos bis zu einer Energie von einem PeV zu rekonstruieren. Ein Vergleich der Analysen findet sich in Abbildung 7.2. Die beiden Analysen wurden in einem unterschiedlichen Binning entfaltet, was einen direkten Vergleich der Entfaltungspunkte verhindert. An den übereinander gezeichneten Datenpunkten in der Abbildung ist jedoch zu erkennen, dass die Entfaltungsergebnisse kompatibel sind.

In Abbildung 7.3 werden noch drei weitere Entfaltungsanalysen mit der vorliegenden Arbeit verglichen. Die Entfaltung von [DRB<sup>+</sup>95] wurde mit Daten des Fréjus-Detektors durchgeführt. Die Entfaltung deckt den Bereich sehr niedriger Energien bis zu einer Energie von 10 TeV ab. Die Überschneidung im Energiebereich ist gering und die Auflösung ist vergleichsweise klein. Die Analyse ist die einzige Entfaltung, die diesen Energiebereich abdeckt. Daher bietet es sich an, anhand dieser Analyse die Entfaltung im Bereich geringer Energien zu verifizieren. Im Vergleich ist zu erkennen, dass die beiden Entfaltungen bündig ineinander übergehen. Es ist daher davon auszugehen, dass die Ergebnisse konsistent sind.

Die zweite Entfaltung von [Mün07] wurde mit Daten des AMANDA-Detektors durchgeführt. Die dritte Entfaltung [AM<sup>+</sup>13] nutzt Daten des ANTARES-Detektors. Beide Messungen decken einen mit der IC86 Entfaltung vergleichbaren Energiebereich ab.

## 7. Analyse des entfalteten Energiespektrums



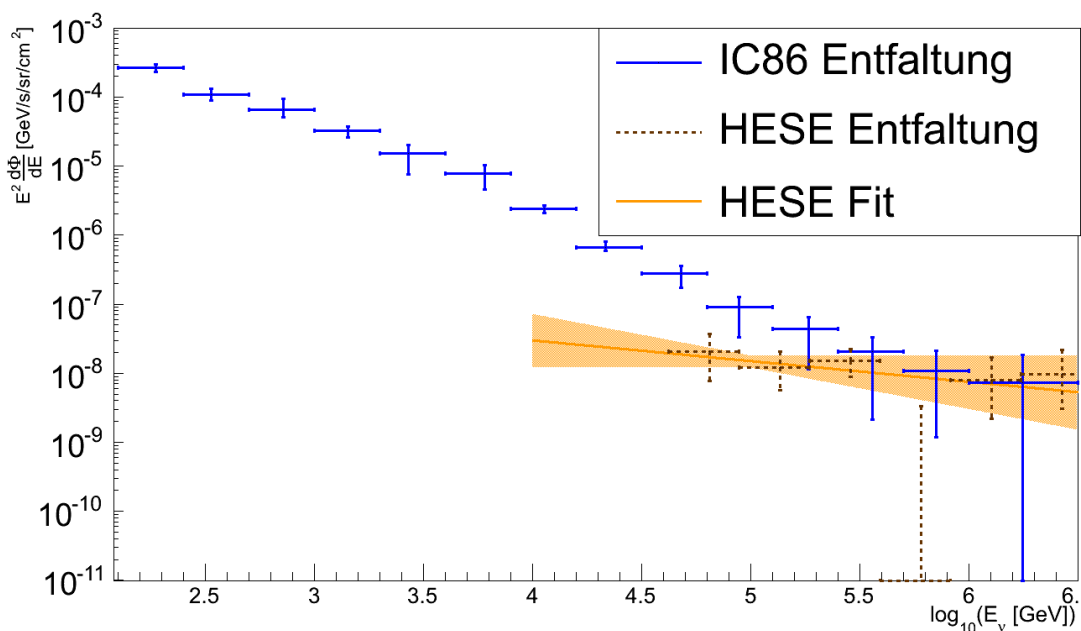
**Abbildung 7.3.:**

Vergleich mit der AMANDA [Mün07], ANTARES [AM<sup>+</sup>13] und Fréjus [DRB<sup>+</sup>95] Entfaltung. Die Analysen weisen alle ein unterschiedliches Binning der Entfaltung auf, was einen direkten Vergleich der einzelnen Bins verhindert. Für den Bereich 10 GeV bis 100 TeV weisen die Ergebnisse von IC86, ANTARES und AMANDA eine gute Übereinstimmung auf. Auch die Fréjus Datenpunkte stimmen im Rahmen des sich überschneidenden Energiebereiches mit den anderen Datenpunkten überein.

Sie sind jedoch nicht in der Lage Aussagen über den Bereich höchster Energien zu treffen. Der Grund für die begrenzte Auflösung bei hohen Energien liegt in dem kleinen Volumen der genutzten Experimente. Für beide Analysen gilt, dass sie ein anderes Binning nutzen. Die übereinander gezeichneten Ergebnisse in Abbildung 7.3 zeigen, dass alle dargestellten Analysen konsistent sind.

Die *High Energy Starting Events* Analyse (HESE), beschrieben in [A<sup>+</sup>14c], war in der Lage, eine Messung des Flusses der hochenergetischen extraterrestrischen Neutrinos durchzuführen. Die Rekonstruktion des Spektrums ist im Vergleich zur IC86 Analyse in Abbildung 7.4 dargestellt. Das rekonstruierte Spektrum von [A<sup>+</sup>14c] passt im hochenergetischen Bereich gut zur vorliegenden IC86 Entfaltung. Die beiden ersten Punkte des Spektrums liegen deutlich unterhalb der IC86 Entfaltung. Dies erklärt sich damit, dass die IC86 Entfaltung eine Messung der Summe aller vom IceCube-Detektor gemessenen Flüsse ist, wohingegen [A<sup>+</sup>14c] nur den Fluss der hochenergetischen extraterrestrischen Neutrinos misst. In Abbildung 7.4 ist ebenfalls ein Fit an die HESE-

## 7.1. Vergleich des entfalteten Energiespektrums mit anderen Messungen



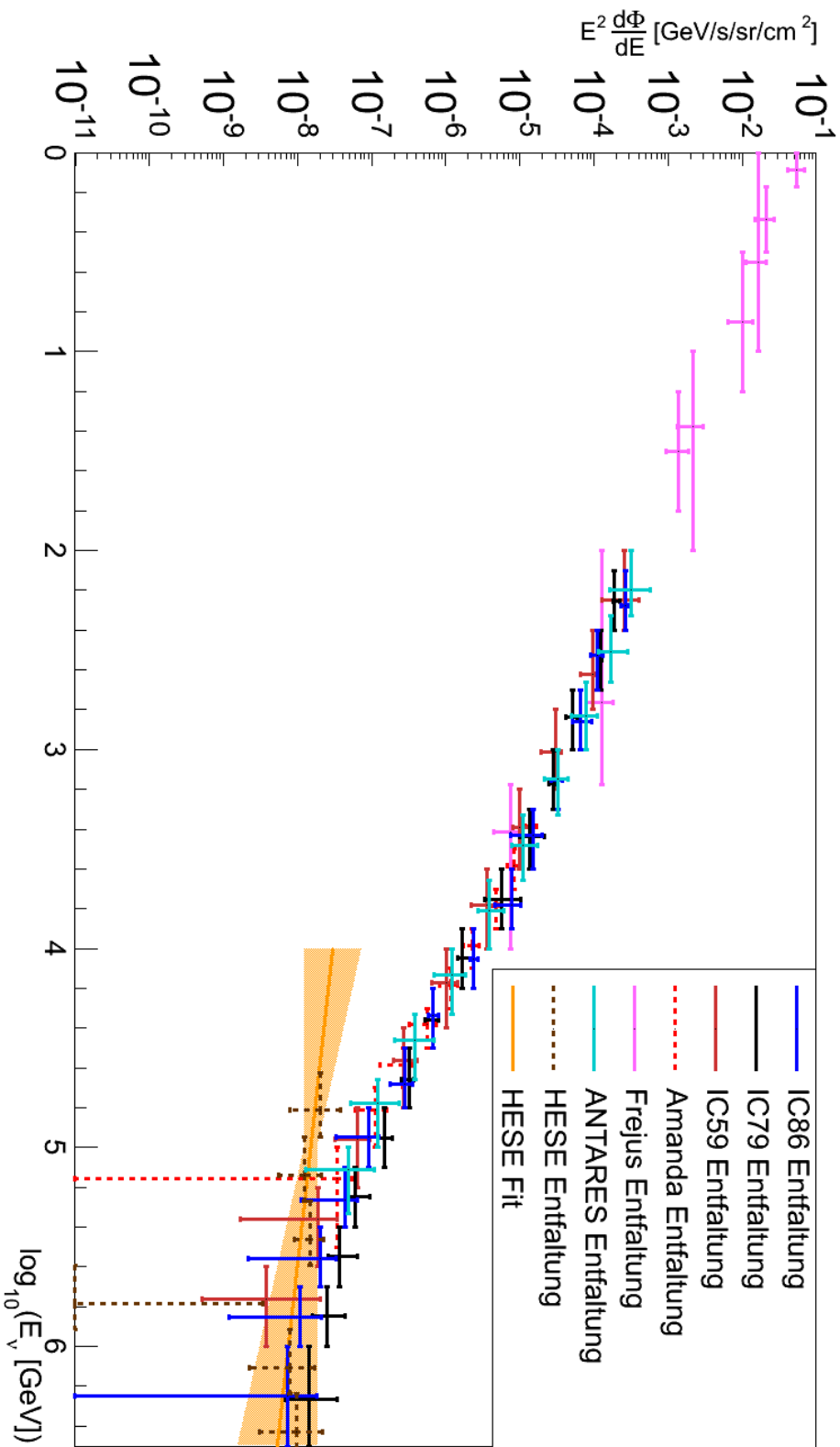
**Abbildung 7.4.:**

Vergleich der IC86 Entfaltung mit der HESE Entfaltung [A<sup>+</sup>14c]. Die Analyse ist in der Lage, durch eine starke Beschränkung auf die Hochenergie und die Nutzung einer speziellen Vetotechnik nur den hochenergetischen extraterrestrischen Neutrinofluss zu rekonstruieren. Der eingezeichnete Fehler der HESE Ergebnisse ergibt sich aus der Unsicherheit des Fits des spektralen Index von  $\gamma = -2,3^{+0,3}_{-0,3}$ . Trotz dieser recht unterschiedlichen Voraussetzungen sind die gemessenen Flüsse kompatibel. [A<sup>+</sup>14c]

Ergebnisse eingezeichnet. Die Übereinstimmung des Fits mit den IC86 Datenpunkten ist ebenfalls ein deutliches Indiz dafür, dass beide Messungen konsistent sind.

In Abbildung 7.5 sind alle bisher beschriebenen Ergebnisse zusammengefasst. Es wird deutlich, dass die Spektren miteinander kompatibel sind. Eine Ausnahme stellt die HESE Analyse dar, da keine Entfaltung aller Flüsse vorgenommen wird, sondern nur des hochenergetischen extraterrestrischen Flusses. Auf Grund der unterschiedlichen Messverfahren wird die HESE Analyse erst in dem Bereich mit den anderen Analysen vergleichbar, in dem die atmosphärische Komponente subdominant wird und alle Analysen nur noch eine Messung der extraterrestrischen hochenergetischen Neutrinos vornehmen.

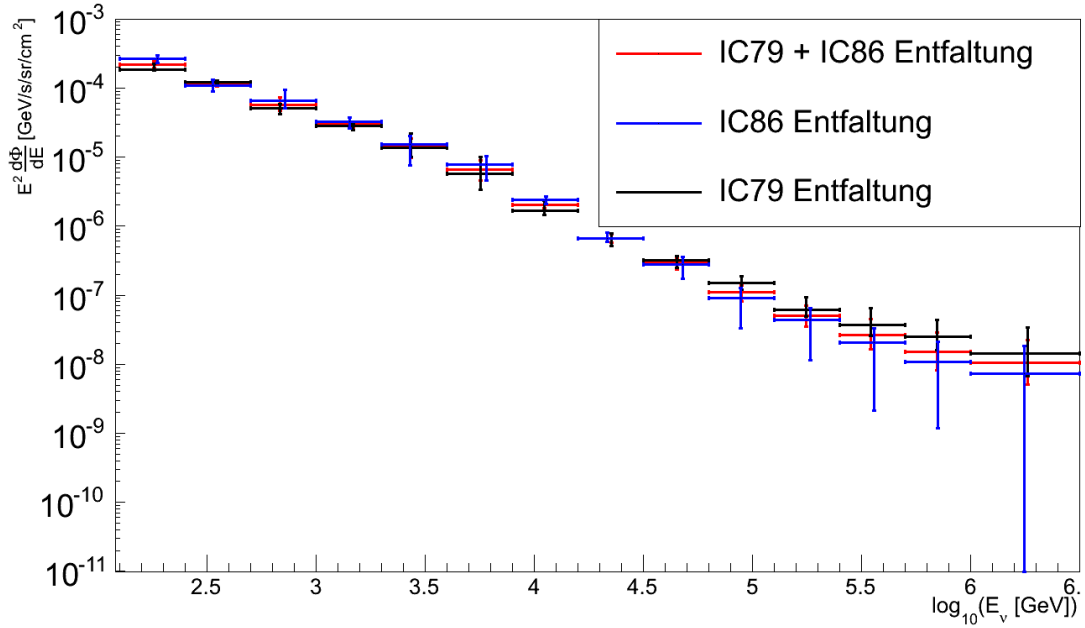
## 7. Analyse des entfalteten Energiespektrums



**Abbildung 7.5.:**

Vergleich der IC86 Entfaltung mit den AMANDA [Mün07], ANTARES [AM<sup>+</sup>13], Fréjus [DRB<sup>+</sup>95], HESE [A<sup>+</sup>14c], IC59 [Ruh13] und IC79 [Sch14] Entfaltungen. Zwischen 100 GeV und etwa 100 TeV liegen die AMANDA, ANTARES, Fréjus, HESE, IC59, IC79 und IC86 übereinander, sodass die einzelnen Messpunkte kaum noch zu erkennen sind. Dies ist ein starkes Indiz dafür, dass alle Messungen miteinander kompatibel sind. Für den Bereich der Neutrinoenergien größer als 100 TeV lässt sich feststellen, dass die vier Messungen mit den Datenpunkten in diesem Bereich ebenfalls konsistente Ergebnisse liefern. Der eingezeichnete Fehler der HESE Ergebnisse ergibt sich aus der Unsicherheit des Fits des spektralen Index von  $\gamma = -2.3_{-0.3}^{+0.3}$ .

## 7.1. Vergleich des entfalteten Energiespektrums mit anderen Messungen



**Abbildung 7.6.:**

Die IC79 [Sch14] und IC86 Entfaltungsergebnisse sind zu einer Messung verbunden. Die gemessenen Flüsse in jedem Bin werden nach Gleichung 7.1 gemittelt. Das Vorgehen verdoppelt nahezu die Anzahl, der für die Entfaltung zur Verfügung stehenden Ereignisse, was die Varianz des Ergebnisses verringert.

Im Laufe dieses Abschnittes wurde gezeigt, dass das gewonnene Spektrum kompatibel zu anderen Entfaltungsergebnissen des Myon-Neutrino fluxes ist. Da die IC79 und IC86 Analysen das gleiche Binning aufweisen und die Daten aus dem gleichen Winkelbereich stammen, können die Ergebnisse zu einem gemittelten Spektrum zusammengefasst werden.

Dazu wird das gewichtete Mittel

$$\langle \Phi \rangle = \frac{\sum (w_i \Phi_i)}{\sum w_i} \quad (7.1)$$

des Fluxes  $\Phi$  in jedem Bin  $i$  berechnet. Das Gewicht

$$w_i = \frac{1}{\sigma_{i,\text{stat}}^2} \quad (7.2)$$

so gewählt, dass es dem Quadrat des inversen statistischen Fehlers entspricht. Da eine gemeinsame systematische Fehlerbetrachtung für die beiden Analysen nicht vorliegt, werden die systematischen Fehler der beiden Analysen nach Gauß fortgepflanzt. Diese Vorgehensweise führt zu einer leichten Überschätzung der Fehler. Eine tabel-

## 7. Analyse des entfalteten Energiespektrums

---

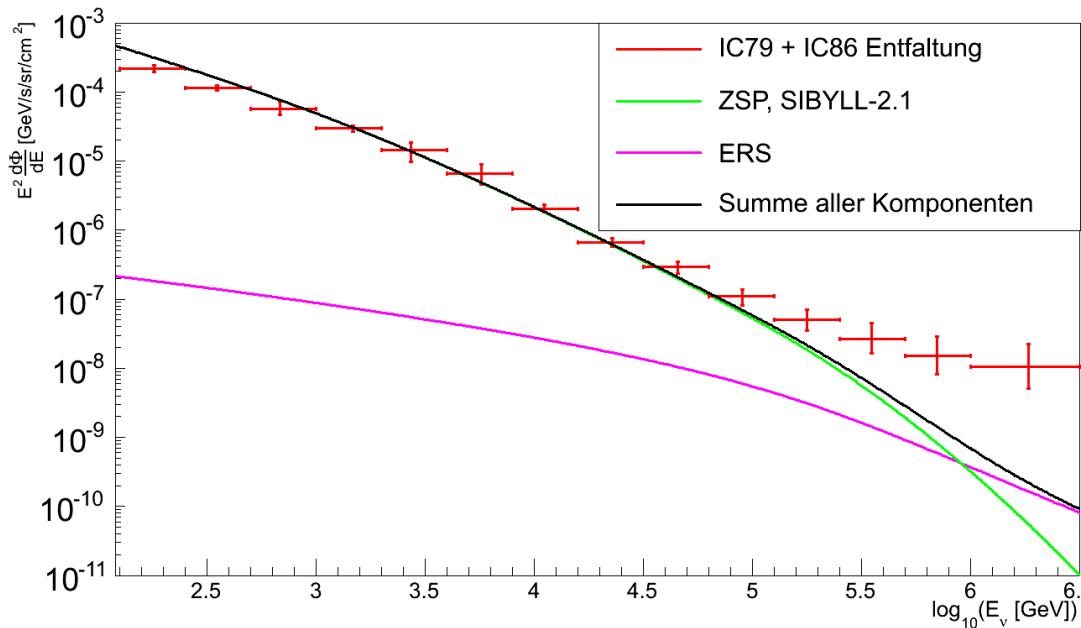
larische Auflistung der Flüsse in den Bins findet sich in Tabelle 7.1. Eine grafische Darstellung des Ergebnisses im Vergleich zur IC79 und zur IC86 Analyse findet sich in Abbildung 7.6. Das so gewonnene Spektrum weist auf Grund der größeren Anzahl von Ereignissen, auf denen es beruht, eine kleinere Unsicherheit auf. Diese Kombination wird im folgenden Abschnitt genutzt, um die Entfaltung mit den unterschiedlichen Modellrechnungen zu vergleichen.

**Tabelle 7.1.:** Fluss und Unsicherheit für die gemittelte IC79 und IC86 Entfaltung.

Bin	$E^2 \times \text{Fluss}$ $\left[ \frac{\text{GeV}}{\text{sr s cm}^2} \right]$	Unsicherheit [%]
1	$2,18 \times 10^{-4}$	+13 -7
2	$1,15 \times 10^{-4}$	+10 -10
3	$5,63 \times 10^{-5}$	+20 -13
4	$3,03 \times 10^{-5}$	+7 -12
5	$1,43 \times 10^{-5}$	+35 -27
6	$6,61 \times 10^{-6}$	+41 -29
7	$2,01 \times 10^{-6}$	+19 -9
8	$6,69 \times 10^{-7}$	+14 -12
9	$2,98 \times 10^{-7}$	+16 -21
10	$1,11 \times 10^{-7}$	+24 -35
11	$5,10 \times 10^{-8}$	+35 -38
12	$2,62 \times 10^{-8}$	+48 -48
13	$1,51 \times 10^{-8}$	+60 -48
14	$1,06 \times 10^{-8}$	+102 -51

### 7.2. Vergleich des gemittelten Energiespektrums mit Modellen

Wie in Kapitel 2 bereits dargestellt, gibt es in der Beschreibung des Flusses der atmosphärischen Neutrinos große Unsicherheiten auf Grund der geladenen kosmischen Strahlung und des hadronischen Interaktionsmodells. Wie bereits ausgeführt, gibt es für die Flussvorhersage der atmosphärischen Neutrinos keine geschlossene Fehlerbetrachtung. Es ist jedoch sicher, dass es eine große energieabhängige Unsicherheit gibt. Da diese nicht bekannt ist, wird im Folgenden darauf verzichtet einen Fehler, auf die Modellvorhersagen anzugeben. Diese Darstellung soll in keiner Weise suggerieren, dass die eingezeichneten Modellvorhersagen fehlerlos sind, vielmehr ist gerade für hohe Energien mit sehr großen Unsicherheiten zu rechnen.



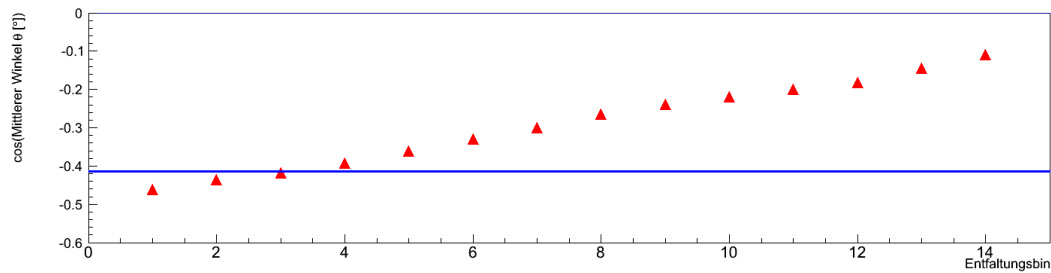
**Abbildung 7.7.:**

Vergleich mit konventioneller und prompter Vorhersage. Die Rekonstruktion wird bis zu einer Neutrinoenergie von etwa 50 TeV hinreichend von durch ZSP [A<sup>+</sup>11c] und SYBILL-2.1 [AEG<sup>+</sup>09] beschrieben. Die Vorhersage aus [ERS08] ist nicht in der Lage den deutlichen Exzess der Messung über der atmosphärischen Vorhersage zu erklären.

Ebenfalls wird im Folgenden darauf verzichtet, die Unterschiede zu betrachten, die sich aus der Benutzung verschiedener Primärmodelle ergeben. Der Grund für diese Vorgehensweise ist, dass durch die unbekanntens Unsicherheiten in den Modellen ein direkter Vergleich schwierig ist und daher eine Genauigkeit suggeriert werden würde, die nicht vorhanden ist.

Das erste Modell mit dem das gewonnene Spektrum verglichen werden soll, ist ein Modell für konventionelle atmosphärische Myon-Neutrinos basierend auf [A<sup>+</sup>11c]. Es wird in Abbildung 7.7 im Vergleich zum gemittelten Spektrum dargestellt. Der Vergleich des Spektrums zum Modell der konventionellen atmosphärischen Neutrinos zeigt, dass die ersten drei Datenpunkte unterhalb der Flussvorhersage des Modells liegen. Eine mögliche Erklärung dieses Verhaltens liegt im mittleren Winkel der einzelnen Entfaltungsbins. Wie in Abschnitt 6.5 dargestellt wird, ist der Fluss der konventionellen atmosphärischen Neutrinos vom betrachteten Kosinus des Zenitwinkels abhängig. Der in der Abbildung dargestellte Fluss ist der integrierte Fluss über den Winkelbereich  $86^\circ \leq \theta \leq 180^\circ$ . Dies ist der Winkelbereich in dem in der IC79 und IC86 Analyse Ereignisse selektiert wurden. Eine Betrachtung des mittleren Winkels

## 7. Analyse des entfalteten Energiespektrums



**Abbildung 7.8.:**

Mittlerer Winkel in den Entfaltungsbins. In blau ist der mittlere Zenitwinkel für alle Entfaltungsbins von  $114,5^\circ$  eingezeichnet. In rot ist der mittlere Zenitwinkel in jedem der 14 Entfaltungsbins eingezeichnet. Der mittlere Winkel sinkt mit steigendem Schwerpunkt der Entfaltungsbins von maximal  $115,9^\circ$  im ersten Entfaltungsbins bis auf minimal  $96,2^\circ$  im letzten Entfaltungsbins.

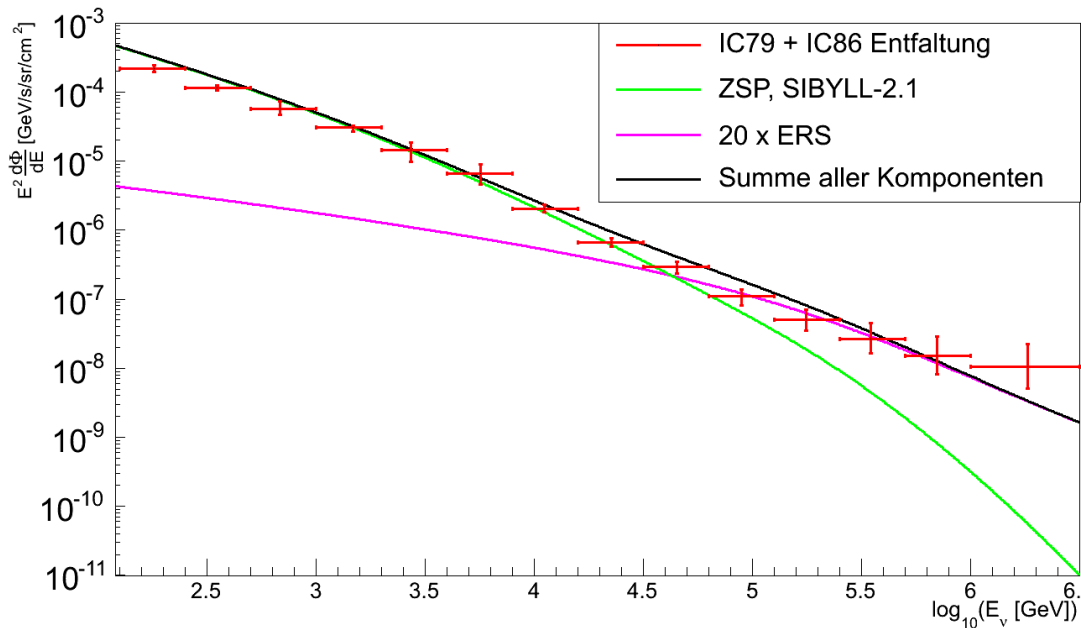
in jedem der Entfaltungsbins in Abbildung 7.8 zeigt, dass der mittlere Winkel in den Entfaltungsbins nicht konstant ist, sondern mit steigender Energie abnimmt. Daher ist in jedem Entfaltungsbins eine kleine vom mittleren Winkel abhängige Verschiebung des Entfaltungsergebnisses gegenüber der Modellvorhersage zu erwarten. Dieser Effekt kann die leichte Überschätzung der ersten drei Punkte erklären. Daher kann geschlossen werden, dass das Modell konsistent mit den gemessenen Daten bis zu einer Energie von 50 TeV ist. Oberhalb dieser Energie liegt die Vorhersage deutlich unterhalb der Messung.

Es stellt sich daher die Frage, ob die beobachtete Diskrepanz im Spektrum durch einen Fluss von prompten atmosphärischen Neutrinos beschrieben werden kann. In Abbildung 7.7 ist zu erkennen, dass die auf [ERS08] beruhende Vorhersage nicht zu dem rekonstruierten Spektrum passt.

Wird der prompte Fluss beispielhaft um einen Faktor zwanzig gesteigert, wie in Abbildung 7.9 aufgezeichnet, ist der vorhergesagte Fluss in der gleichen Größenordnung wie die entfalteten Daten. Die Form des Spektrums deckt sich nicht gut mit der Vorhersage. Die Entfaltung scheint im Bereich der Hochenergie einem härteren Spektrum zu folgen. Die Unterschiede sind jedoch nicht so prominent, dass die Hypothese ausgeschlossen werden kann.

Der Exzess der Messung über der atmosphärischen Vorhersage von konventionellen Neutrinos aus Pionen und Kaonen lässt sich mit einem Fluss hochenergetischer extraterrestrischer Neutrinos erklären. Um dies deutlich zu machen, wird in Abbildung 7.10 neben den atmosphärischen Flüssen ebenfalls ein Fluss von hochenergetischen extraterrestrischen Neutrinos eingezeichnet. Die eingetragene Komponente basiert auf den Ergebnissen von [A<sup>+</sup>14c]. Die Abbildung zeigt, dass die Summe der drei Flüsse die





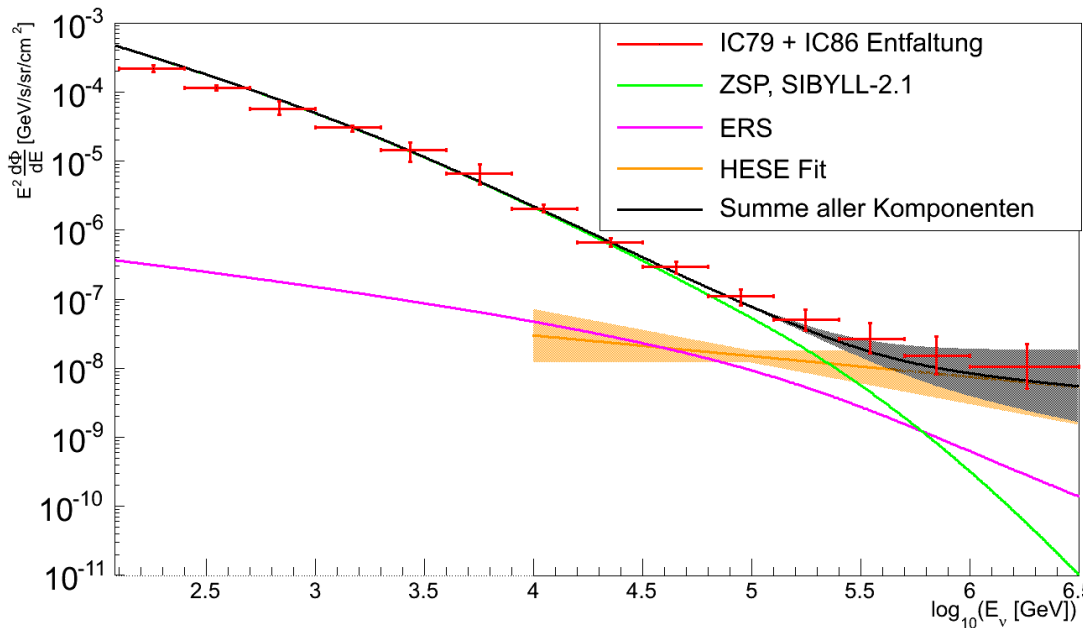
**Abbildung 7.9.:**

Vergleich mit einer konventionellen Komponente, beschrieben durch ZSP [A<sup>+</sup>11c] und SYBILL-2.1 [AEG<sup>+</sup>09], und einer um einen Faktor zwanzig gesteigerten prompten Komponente, beschrieben durch ERS [ERS08]. Eine deutlich gesteigerte prompte Komponente kommt in der Summe den gemessenen Punkten sehr nah, die Form des Exzesses scheint aber einen härteren Fluss zu bevorzugen. Ein Fit der einzelnen Komponenten findet sich in Abschnitt 7.3. Eine gesonderte Betrachtung der maximal erlaubten Normierung der prompten Komponente findet sich in Abschnitt 7.4.

gemittelte Entfaltung beschreibt. Lediglich der Bereich zwischen 50 und 300 TeV zeigt einen leichten Exzess der Entfaltung über der Summe der Vorhersagen. Es ist jedoch zu beachten, dass jede der eingezeichneten Komponenten eine unbekannte Unsicherheit aufweist.

Im Rahmen dieses Abschnittes wurde das Verhältnis der gemittelten Entfaltung mit einer Reihe von Flussvorhersagen verglichen. Es wird dabei aufgezeigt, dass der rekonstruierte Fluss für Energien größer 100 TeV deutlich über einer rein konventionellen Vorhersage liegt. Aus den bisherigen Ergebnissen lässt sich keine eindeutige Aussage treffen, welches Modell am besten die Daten beschreibt. Es bietet sich daher an, diese Frage mittels einer Likelihood-Methode genauer zu untersuchen. Diese entscheidende Untersuchung wird im folgenden Abschnitt durchgeführt.

## 7. Analyse des entfalteteten Energiespektrums



**Abbildung 7.10.:**

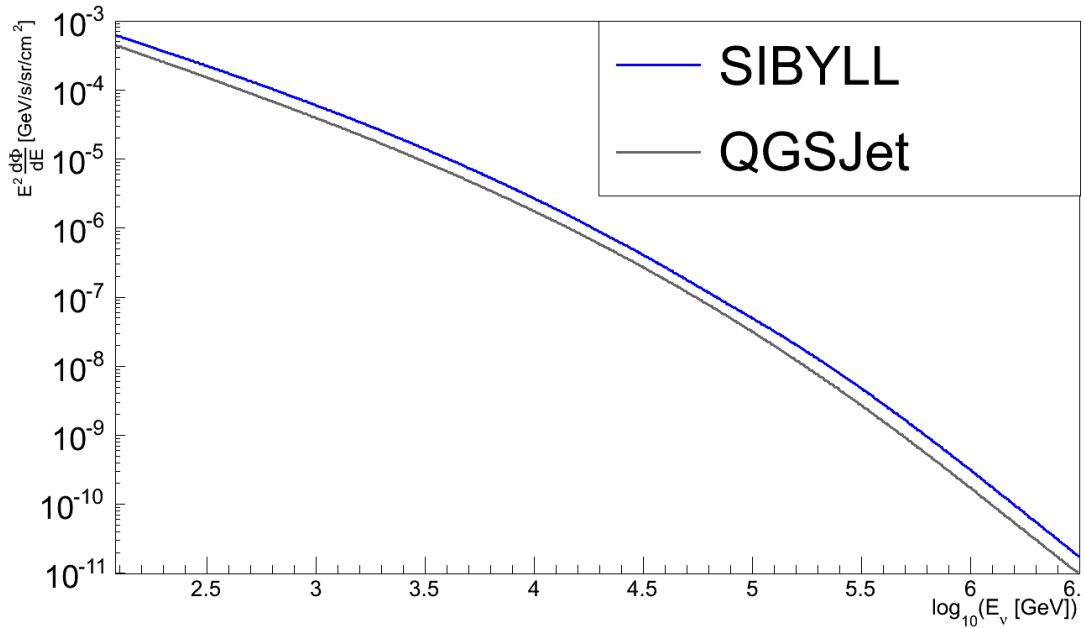
Vergleich mit der konventionellen atmosphärischen Komponente beschrieben durch ZSP [A<sup>+</sup>11c] und SYBILL-2.1 [AEG<sup>+</sup>09], der prompten Komponente ERS [ERS08] und den HESE Ergebnissen aus [A<sup>+</sup>14c]. Die Messung wird durch die Summe der drei Flüsse gut beschrieben. Nur im Energiebereich zwischen 50 und 300 TeV zeigt sich eine Differenz zwischen Vorhersage und Messung. Der eingezeichnete Fehler der HESE Ergebnisse ergibt sich aus der Unsicherheit des Fits des spektralen Index von  $\gamma = -2,3^{+0,3}_{-0,3}$ .

### 7.3. Bestimmung der Modellparameter

Die Frage nach der Kombination von Modellen, die am besten zu den gemessenen Daten passen, ist nicht einfach zu beantworten.

Es bietet sich daher an, alle Möglichkeiten, die sich aus der absoluten Normierung der verschiedenen Flüsse, sowie dem spektralen Index des Flusses der extraterrestrischen Neutrinos ergeben, zu testen. Da das wichtigste Ziel dieser Untersuchung der Ursprung des Exzesses über der konventionellen Vorhersage ist, werden nur Datenpunkte mit einer Energie größer als 10 TeV betrachtet. Dies hat den Vorteil, dass in diesem Bereich der Einfluss des konventionellen Flusses stetig abnimmt und so die Unsicherheiten aus dem hadronischen Interaktions- und Primärmodell mit steigender Energie unbedeutender werden.

Das hadronische Interaktionsmodell wird in der nachfolgenden Untersuchung nicht betrachtet. Die Unterschiede, die sich aus den verschiedenen hadronischen Interaktionsmodellen ergeben, sind hauptsächlich in der absoluten Normierung zu finden,



**Abbildung 7.11.:** Vergleich zwischen den hadronischen Interaktionsmodellen QGSJet [Ost06] und SIBYLL-2.1 [AEG<sup>+</sup>09].

wie in Abbildung 7.11 zu erkennen ist. Die absolute Normierung wird jedoch gefittet, daher kann auf eine Erweiterung des Fits auf verschiedene Interaktionsmodelle verzichtet werden.

Ebenfalls kann auf einen Fit des Primärmodells für die konventionelle Komponente verzichtet werden. Der Energiebereich, in dem sich die Modelle deutlich unterscheiden, liegt bei Energien größer als 10 TeV. In diesem Bereich wird die Messung des konventionellen Flusses zunehmend von der zusätzlichen Komponente überlagert. Es besteht also wenig Aussicht, eine klare Aussage über das Primärmodell treffen zu können. Auf eine Betrachtung dieser Parameter wird daher verzichtet.

Die zu bestimmenden Parameter ergeben sich daher zu der Normierung des konventionellen Flusses  $N_{\text{Atmo}}$ , der Normierung des prompten Flusses  $N_{\text{Prompt}}$ , der Normierung  $N_{\text{Astro}}$  und dem spektralen Index  $\gamma$  des extraterrestrischen Flusses. Der Bereich in dem die Parameter gefittet werden, wird in Tabelle 7.2 aufgelistet. Eine ausführliche Beschreibung der einzelnen Parameter findet sich in Unterabschnitt 7.3.1.

Für den Test der verschiedenen Modelle werden die Entfaltungsergebnisse der IC59 [Ruh13], IC79 [Sch14], IC86, AMANDA [Mün07] und ANTARES [AM<sup>+</sup>13] Entfaltungen genutzt. Der Grund für diese Auswahl ist, dass es sich bei allen Datensätzen um Entfaltungen handelt, deren Ergebnisse bereits systematische Fehler enthalten. Daher

können die Ergebnisse aus verschiedenen Jahren und unterschiedlichen Detektoren einfach kombiniert und für einen globalen Fit genutzt werden.

Ein weiterer Grund für die Wahl ist, dass die gewählten Entfaltungen einen ähnlichen mittleren Zenitwinkel aufweisen und in einem ähnlichen Energiebereich Messungen durchgeführt haben. Von den ausgewählten Analysen zeigen nur die IC79 Analyse und die IC86 Analyse einen deutlichen Exzess über der Vorhersage für den konventionellen Fluss. Die anderen Analysen werden gewählt, um den Fit besser im Energiebereich zwischen 10 TeV und 100 TeV einzuschränken. In diesem Energiebereich würde sich, wie in Abbildung 7.9 zu sehen ist, der Einfluss einer ausgeprägten prompten Komponente zeigen.

**Tabelle 7.2.:**

Das globale Minimum der Likelihood-Funktion ergibt sich für die, in der Tabelle gegebenen Parameterwerte.  $N_{\text{Atmo}}$ : Normierung der konventionellen Neutrinos. Die Normierung bezieht sich dabei auf das Primärmodell ZSP [A<sup>+</sup>11c], und das hadronische Interaktionsmodell SIBYLL-2.1 [AEG<sup>+</sup>09].  $N_{\text{Prompt}}$ : Normierung der prompten Neutrinos in Einheiten des in [ERS08] beschriebenen Flusses.  $N_{\text{Astro}}$ : Normierung der hochenergetischen extraterrestrischen Neutrinos. Die Normierung bezieht sich auf ein Vielfaches eines Flusses von  $1 \times 10^{-8} \left[ \frac{\text{GeV}}{\text{sr s cm}^2} \right]$  bei einer Energie von einem PeV.  $\gamma$ : Spektraler Index des Flusses der hochenergetischen extraterrestrischen Neutrinos.

Parameter	Bester Fit	Fit Bereich
$N_{\text{Atmo}}$	$1,02^{+0,09}_{-0,06}$	[0,7; 1,3]
$N_{\text{Prompt}}$	$0,20^{+5,90}_{-0,20}$	[0; 20]
$N_{\text{Astro}}$	$1,47^{+0,44}_{-0,38}$	[0; 2]
$\gamma$	$-2,39^{+0,20}_{-0,45}$	[-1; -3,5]

Für den Fit wird eine Likelihood-Methode genutzt. Daher wird für jede Kombination der oben beschriebenen Parameter eine Flussvorhersage  $\mathcal{F}_j$  gebildet. Anschließend wird der Abstand  $r_{i,\sigma}$  jedes der  $N$  Messpunkte  $x_i$  in Einheiten des Fehlers zu der Flussvorhersage berechnet. Aus  $r_{i,\sigma}$  kann die Wahrscheinlichkeit

$$p_i(x_i|\mathcal{F}_j) = \int_{-\infty}^{r_{i,\sigma}} \mathcal{N}(1,0) dx \quad (7.3)$$

durch Integration über die Standardnormalverteilung  $\mathcal{N}(1,0)$  berechnet werden. Diese beschreibt, wie wahrscheinlich der Datenpunkt unter der Bedingung der gegebenen Flussvorhersage  $\mathcal{F}$  ist. Das Produkt aller  $p_i(x_i|\mathcal{F}_j)$  für eine Flussvorhersage  $\mathcal{F}_j$  wird als Likelihood-Funktion  $L$  bezeichnet.

$$L = p(\mathcal{F}_j|x_i) = \prod_i^N p_i(x_i|\mathcal{F}_j) \quad (7.4)$$

Die Gleichung 7.4 beschreibt die Wahrscheinlichkeit einer Flussvorhersage unter der Bedingung der Datenpunkte  $x_i$ . Das Maximum der Likelihood-Funktion ergibt sich für die Flussvorhersage  $\mathcal{F}_j$ , welche die Summe aller Datenpunkte am besten beschreibt. Es ist Konvention nicht das Maximum von Gleichung 7.4 zu berechnen, sondern das Minimum von

$$-\log(L) = -\sum_{i=1}^N \log(p_i(x_i|\mathcal{F}_j)). \quad (7.5)$$

zu bestimmen. Auf diese Weise kann die Kombination von Parametern gefunden werden, welche die Daten am besten beschreiben. Alle betrachteten Komponenten, sowie die zugehörigen Wertebereiche und der sich abschließend ergebene beste Fit, finden sich in Tabelle 7.2. [BL98]

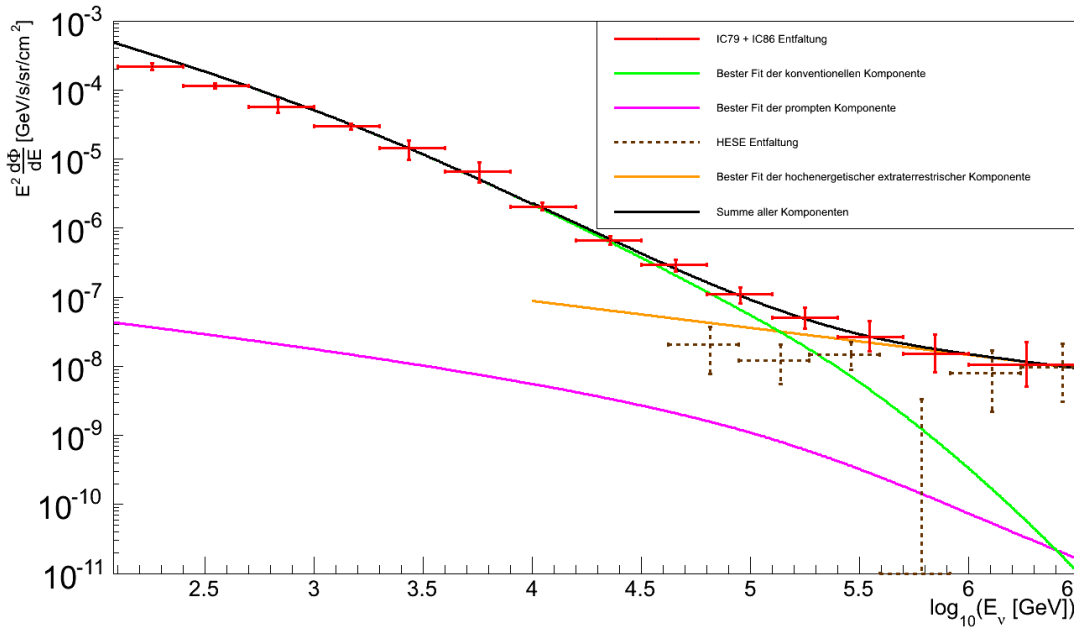
Werden alle Ergebnisse der einzelnen Komponenten kombiniert, kann das Ergebnis mit dem in Abbildung 7.6 beschriebenen gemittelten Spektrum verglichen werden. Das Ergebnis findet sich in Abbildung 7.12, es ist zu erkennen, dass der Fit die gemessenen Datenpunkte beschreibt. Ebenfalls ist der in Abbildung 7.10 sichtbare Exzess der Daten über der Vorhersage im Bereich zwischen 50 TeV und 300 TeV nicht mehr vorhanden.

In Abbildung 7.12 sind ebenfalls die HESE Ergebnisse eingezeichnet. Der beste Fit dieser Analyse für die Normierung des Flusses der hochenergetischen extraterrestrischen Neutrinos bei einer Energie von 1 PeV ist  $N_{\text{Astro}}^{\text{HESE}} = 0,75 \times 10^{-8} \left[ \frac{\text{GeV}}{\text{sr s cm}^2} \right]^1$ . Im Vergleich zu den Ergebnissen aus Tabelle 7.2 ist zu erkennen, dass der beschriebene Fit eine höhere Normierung des Flusses der hochenergetischen extraterrestrischen Komponente präferiert. Der beste Fit der HESE Ergebnisse für den spektralen Index ist  $\gamma^{\text{HESE}} = 2,3 \pm 0,3$ . Dieses Ergebnis ist kompatibel mit den Ergebnissen aus Tabelle 7.2. In Anbetracht der gegebenen bekannten Unsicherheiten ist daher zu schließen, dass die Ergebnisse des Fits kompatibel mit den HESE Ergebnissen sind. [A<sup>+</sup>14c]

Im Folgenden werden die Unsicherheiten, die sich für die einzelnen Komponenten ergeben, beschrieben. Zu diesem Zweck wird die eindimensionale Projektion der Likelihood-Funktion für die einzelnen Komponenten betrachtet. Die Frage nach der Signifikanz des Exzesses über der Vorhersage eines konventionellen Flusses wird in Unterabschnitt 7.3.2 behandelt.

<sup>1</sup>In [A<sup>+</sup>14c] wird keine Unsicherheit für dieses Ergebnis angegeben.

## 7. Analyse des entfalteten Energiespektrums



**Abbildung 7.12.:**

Das Ergebnis des besten Fits aus Tabelle 7.2 im Vergleich zum gemittelten Spektrum aus der IC79 und IC86 Analyse. Die dargestellte hochenergetische extraterrestrische Komponente ist kompatibel mit der in [A<sup>+</sup>14c] beschriebenen HESE Entfaltung.

### 7.3.1. Projektionen der Likelihood-Funktion

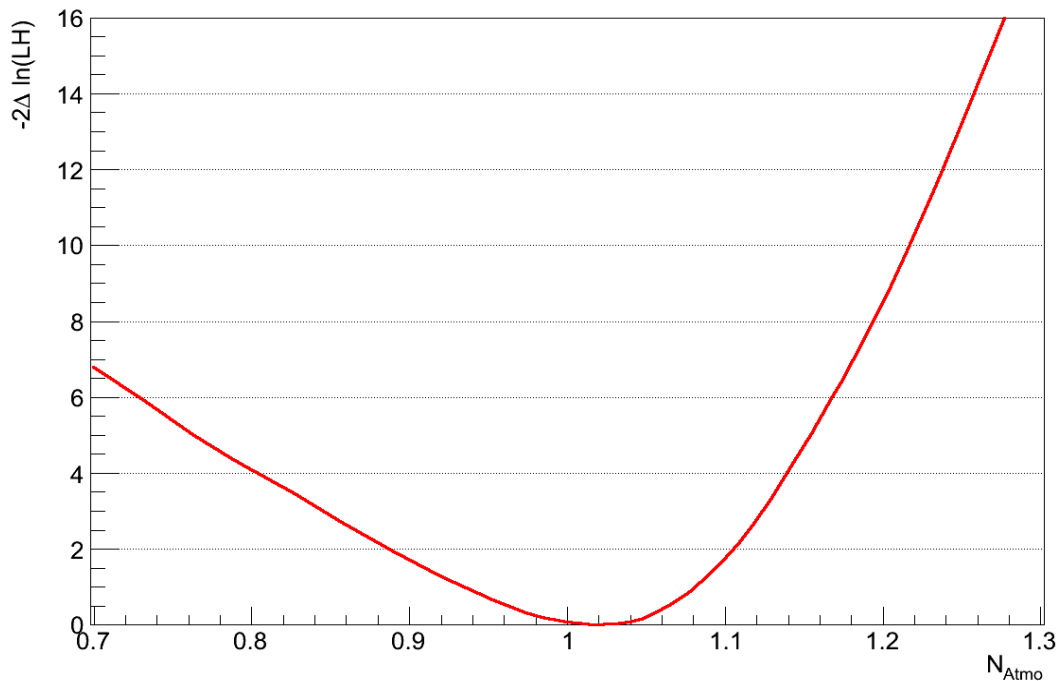
Der erste betrachtete Parameter ist die Normierung des konventionellen Flusses. Der Standard bezieht sich dabei auf das Primärmodell ZSP [A<sup>+</sup>11c], und das hadronische Interaktionsmodell SIBYLL-2.1 [AEG<sup>+</sup>09]. Der Fit der konventionellen Komponente wird im Wertebereich zwischen 0,7 und 1,3 der Standardnormierung durchgeführt.

Die Likelihood-Funktion in Abhängigkeit der Normierung der konventionellen Neutrinos findet sich in Abbildung 7.13. Das Minimum ergibt sich zu  $1,02^{+0,09}_{-0,06}$ . Die  $n\sigma$ -Umgebung einer Likelihood-Funktion wird durch den Anstieg der Likelihood-Funktion  $L$  bezogen auf das Minimum  $L_{\text{Min}}$  auf

$$L_n = L_{\text{Min}} \frac{n^2}{2} \quad (7.6)$$

definiert. Da die Funktion im Fall der konventionellen Normierung asymmetrisch ist, ergeben sich auch asymmetrische Fehler.

Das Ergebnis von  $1,02^{+0,09}_{-0,06}$  ist konsistent mit der Beobachtung, dass ein einfacher konventioneller Fluss bereits, wie in Abbildung 7.7 aufgeführt, die gemessenen Daten-

**Abbildung 7.13.:**

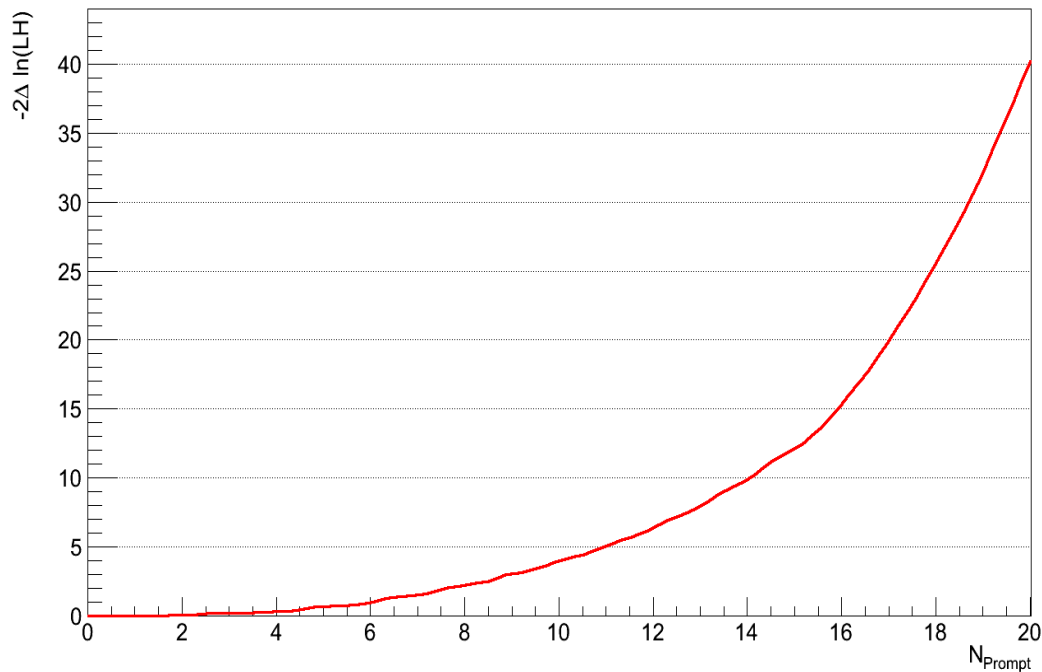
Eindimensionale Projektion der Likelihood-Funktion für die Normierung der konventionellen Neutrinos.

punkte beschreiben kann. Hier zeigt sich, dass die Normierung des Flusses der konventionellen atmosphärischen Neutrinos hinreichend von dem Modell beschrieben wird. Daher war ein Wert in der Nähe von 1,0 zu erwarten.

Die Projektion der Likelihood-Funktion für die prompt Komponente findet sich in Abbildung 7.14. Die Normierung bezieht sich auf ein Vielfaches der in [ERS08] beschriebenen prompten Komponente. Der Fit wird im Intervall  $[0; 20]$  durchgeführt. Das Minimum der dargestellten Likelihood-Funktion liegt bei  $0,2^{+5,90}_{-0,20}$ . Es ist ein sehr breites Minimum für die Normierung der prompten Komponente erkennbar. An den Fehlern des Fits ist zu erkennen, dass die  $1\sigma$ -Umgebung erst bei einem sechsfachen der prompten Komponente endet. Die  $2\sigma$ -Umgebung endet erst bei einem zehnfachen der Normierung. In Abbildung 7.9 wird gezeigt, dass der Exzess über den konventionellen Neutrinos durch eine zwanzigfache prompte Komponente erklärbar ist. Wie sich an dem Fit zeigt, wird eine solche starke prompte Komponente von den Ergebnissen des Fits nicht präferiert.

Die nächste Komponente ist die Normierung einer hochenergetischen extraterrestrischen Komponente. Die Normierung bezieht sich auf ein Vielfaches eines Flusses von  $1 \times 10^{-8} \left[ \frac{\text{GeV}}{\text{sr s cm}^2} \right]$  bei einer Energie von einem PeV. Der Fitbereich dieser Komponente

## 7. Analyse des entfalteten Energiespektrums



**Abbildung 7.14.:**

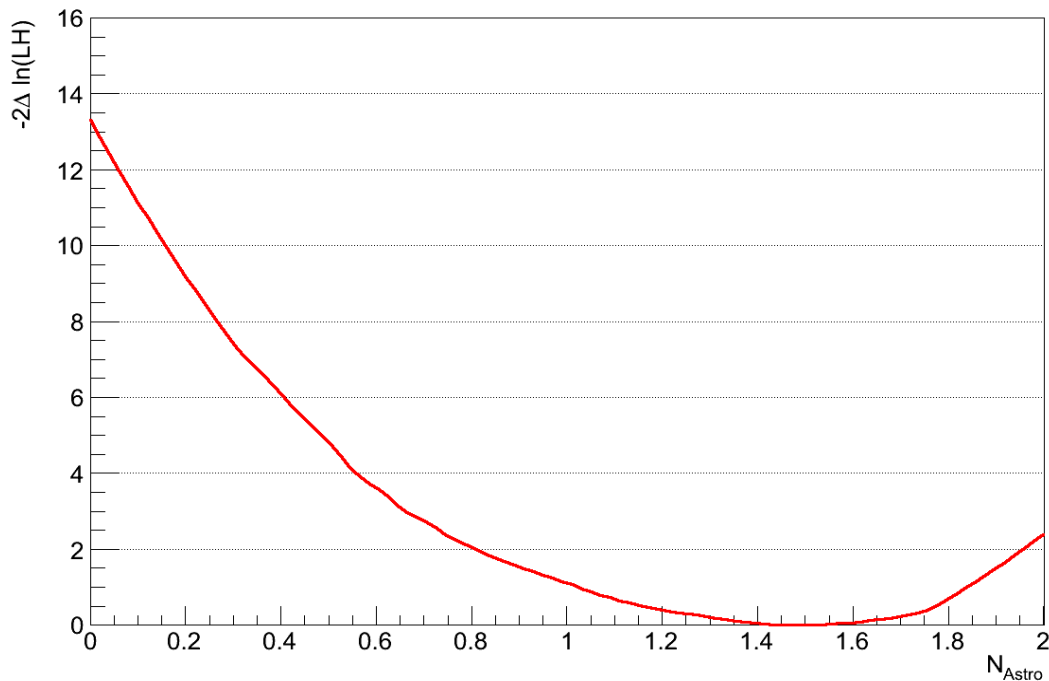
Eindimensionale Projektion der Likelihood-Funktion für die Normierung der prompten Neutrinos.

liegt im Intervall  $[0; 2]$ . Die zugehörige Abbildung der Likelihood-Funktion findet sich in Abbildung 7.15. Das Minimum der Likelihood-Funktion ergibt sich zu  $1,47^{+0,44}_{-0,38}$ . Das Ergebnis des Fits ist kompatibel mit der in  $[A^{+14c}]$  beschriebenen Messung, es wird jedoch eine höhere absolute Normierung präferiert. Von besonderer Bedeutung ist der Wert der Likelihood-Funktion bei einem Wert von null. In der Abbildung ist zu erkennen, dass der Achsenabschnitt bei etwa  $-2\Delta \ln(L) = 13,5$  liegt. Dies bedeutet, nach Gleichung 7.6 einen Abstand von mehr als einer dreifachen Standardabweichung zum Minimum der Likelihood-Funktion. Dieser Wert ist nicht die Signifikanz des Ergebnisses, jedoch zeigt er, dass der Fit Ergebnisse mit einer nicht verschwindenden hochenergetischen extraterrestrischen Komponente präferiert.

Der spektrale Index ist von zentraler Bedeutung, da dieser zusammen mit der Normierung den Fluss eindeutig charakterisiert. Er wird im Intervall  $[-1; -3,5]$  gefittet. Das Minimum ergibt sich zu  $-2,39^{+0,20}_{-0,45}$ . Das Ergebnis des Fits ist, wie das Ergebnis für die Normierung der Komponente, kompatibel mit der Messung in  $[A^{+14c}]$ . Eine Darstellung der Likelihood-Funktion findet sich in Abbildung 7.16.

Die eindimensionalen Projektionen der Likelihood-Funktion geben jeweils nur eine Aussage über einen der gefitteten Parameter. Die Wechselwirkungen der Parameter





**Abbildung 7.15.:**

Eindimensionale Projektion der Likelihood-Funktion für die Normierung der hochenergetischen extraterrestrischen Komponente.

untereinander werden durch diese Darstellungsform nicht wiedergegeben. Daher bietet es sich an, zusätzlich zweidimensionale Projektionen zu betrachten.

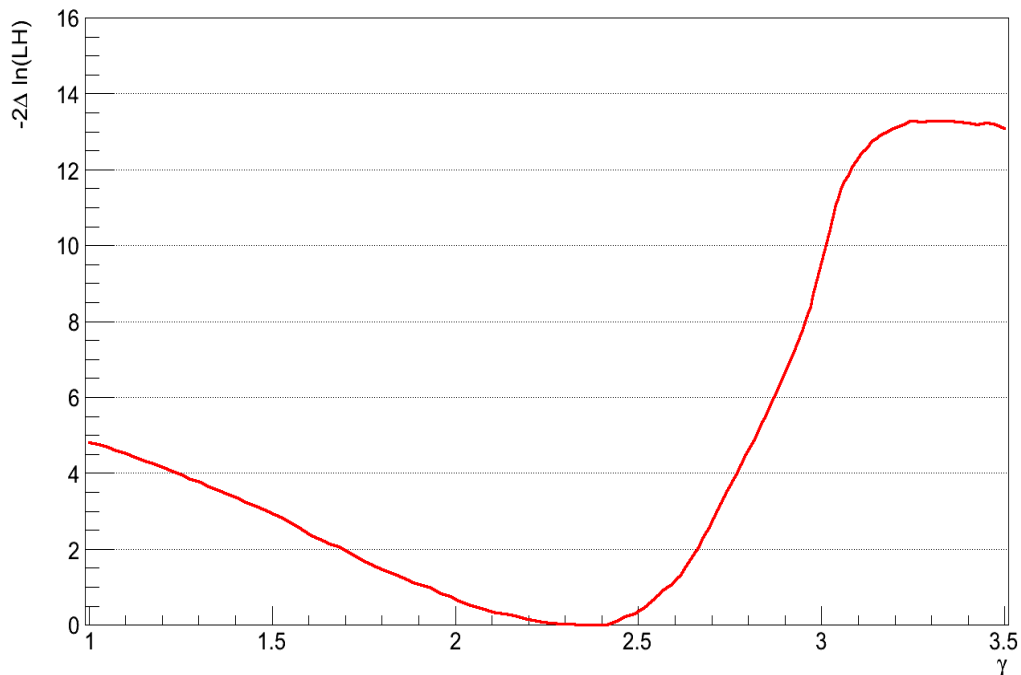
Bei den nachfolgenden zweidimensionalen Projektionen wird die  $z$ -Achse, nach Gleichung 7.6, so gewählt, dass sie dem Vielfachen einer Standardabweichung entspricht. Durch diese Wahl lässt sich leichter erkennen, welche Kombinationen den besten Fit liefert.

Die erste Projektion wird in Abbildung 7.17 dargestellt. Sie zeigt das Zusammenspiel vom spektralem Index und der Normierung der prompten Komponente. Durch die Projektion zeigt sich eine deutliche Korrelation zwischen der prompten Komponente und dem spektralen Index des hochenergetischen extraterrestrischen Flusses. Generell deuten die Ergebnisse des Fits auf ein sehr weiches Spektrum mit einem spektralen Index kleiner  $-2,0$  hin.

Die zweite Projektion, dargestellt in Abbildung 7.18, beschreibt die Wechselwirkung der Normierung der prompten Komponente mit der Normierung des hochenergetischen extraterrestrischen Flusses. Genau wie für die vorherige Projektion zeigt sich eine starke Korrelation zwischen den beiden Parametern. Für kleine Werte der promp-

## 7. Analyse des entfalteten Energiespektrums

---



**Abbildung 7.16.:**

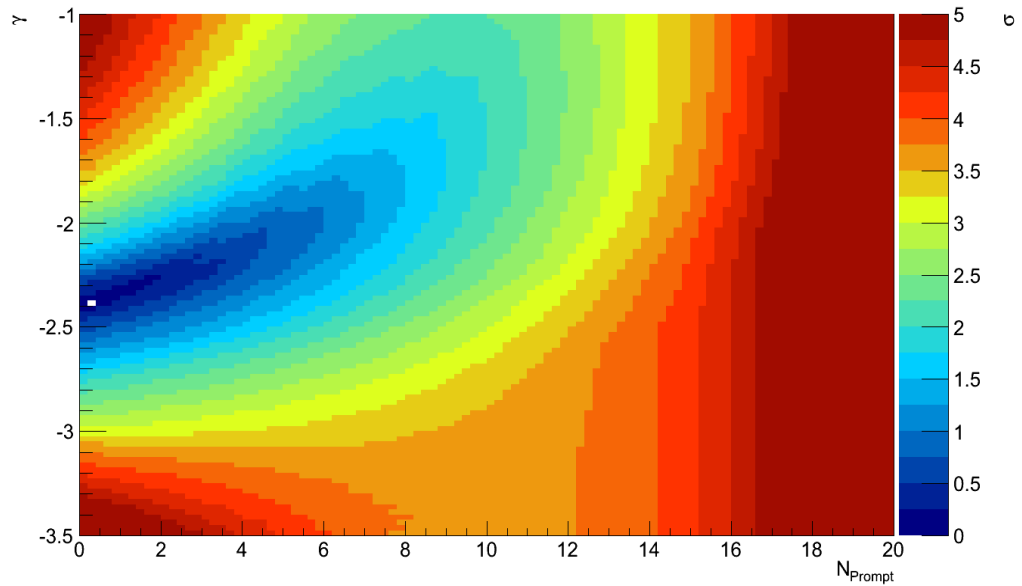
Eindimensionale Projektion der Likelihood-Funktion für den spektralen Index der hochenergetischen extraterrestrischen Komponente.

ten Komponente ergibt sich ein sehr deutlicher hochenergetischer extraterrestrischer Fluss.

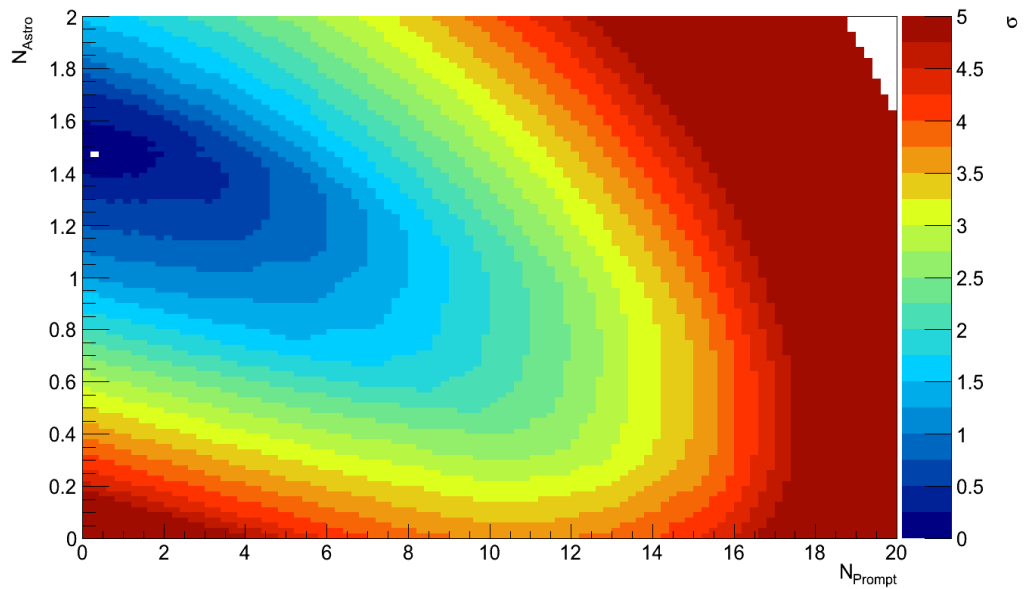
An dieser Stelle zeigt sich besonders deutlich, dass die beiden Komponenten durch ihre hohe Korrelation untereinander entartet sind. Eine klare Aussage über die hochenergetische extraterrestrische Komponente wird ohne weitere Messungen, welche die Normierung der prompten Komponente einschränken, deutlich erschwert.

Die zuletzt zu betrachtende Projektion zeigt die Wechselwirkung der beiden Komponenten der hochenergetischen extraterrestrischen Neutrinos. Die Projektion findet sich in Abbildung 7.19. Der Bereich des besten Fitergebnisses für die beiden Komponenten ist deutlich zu erkennen. Aus der Projektion kann geschlossen werden, dass hohe Normierungen des extraterrestrischen Flusses nur für einen geringen spektralen Index wahrscheinlich sind.

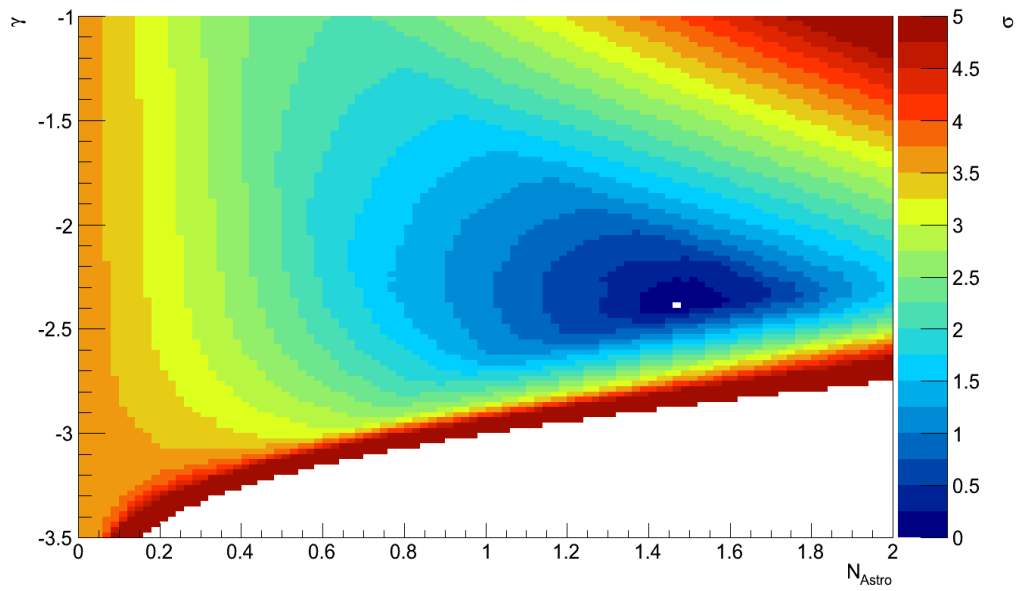
Die Projektionen der Likelihood-Funktion beschreiben jeweils die Fehler eines Fits. Sie geben nicht die Signifikanz des Ergebnisses wieder. Daher wird im folgenden Unterabschnitt 7.3.2 die Signifikanz des Ergebnisses berechnet.



**Abbildung 7.17.:**  
Zweidimensionale Projektion der Likelihood-Funktion für den spektralen Index und die prompt Komponente. Das weiße Bin entspricht dem wahrscheinlichsten Fit.



**Abbildung 7.18.:**  
Zweidimensionale Projektion der Likelihood-Funktion für die Normierung des hochenergetischen extraterrestrischen Flusses und der prompten Komponente. Das weiße Bin entspricht dem wahrscheinlichsten Fit.



**Abbildung 7.19.:**

Zweidimensionale Projektion der Likelihood-Funktion für die Normierung des hochenergetischen extraterrestrischen Flusses und des spektralen Index. Das weiße Bin entspricht dem wahrscheinlichsten Fit.

### 7.3.2. Signifikanz des beobachteten Exzesses

Wie in Abbildung 7.9 und Abbildung 7.10 zu sehen ist, gibt es zwei grundsätzliche Möglichkeiten die gemessenen Daten zu erklären. Die Nullhypothese  $H_0$  geht davon aus, dass es keinen hochenergetischen extraterrestrischen Fluss von Neutrinos gibt und der Exzess nur auf Grund von konventionellen und prompten Neutrinos entsteht. In der zweiten Hypothese  $H_1$  wird davon ausgegangen, dass es zu den Komponenten von  $H_0$  noch einen weiteren Fluss von hochenergetischen extraterrestrischen Neutrinos gibt.

Für beide Hypothesen kann eine Likelihood-Funktion, wie in Gleichung 7.4 beschrieben, aufgestellt werden. Sind für beide Hypothesen Likelihood-Funktionen definiert, kann mit Hilfe eines Likelihood-Quotienten-Tests

$$\Lambda = \frac{\sup \left( \prod_{i=1}^N p_{i,0}(x_i) \right)}{\sup \left( \prod_{i=1}^N p_{i,1}(x_i) \right)} \quad (7.7)$$

die Signifikanz bestimmt werden. Bei diesem Test wird der Wert der Likelihood-Funktion für die wahrscheinlichste Kombination der Parameter von  $H_0$  durch den Wert für die wahrscheinlichste Kombination der Parameter von  $H_1$  geteilt. Um aus  $\Lambda$  die Signifikanz zu berechnen, kann  $\Lambda$  als

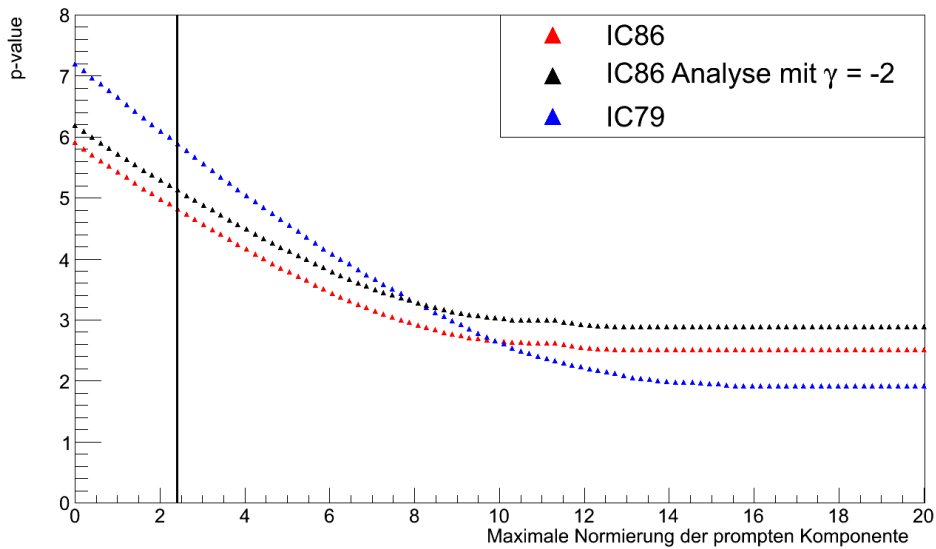
$$-2\Lambda = -2 \log \sup \left( \prod_{i=1}^N p_{i,0}(x_i) \right) + 2 \log \sup \left( \prod_{i=1}^N p_{i,1}(x_i) \right) \quad (7.8)$$

ausgedrückt werden. Diese Funktion folgt einer  $\chi^2$ -Verteilung, daher kann aus ihrem Funktionswert eine Signifikanz bestimmt werden. [Wil38, BL98]

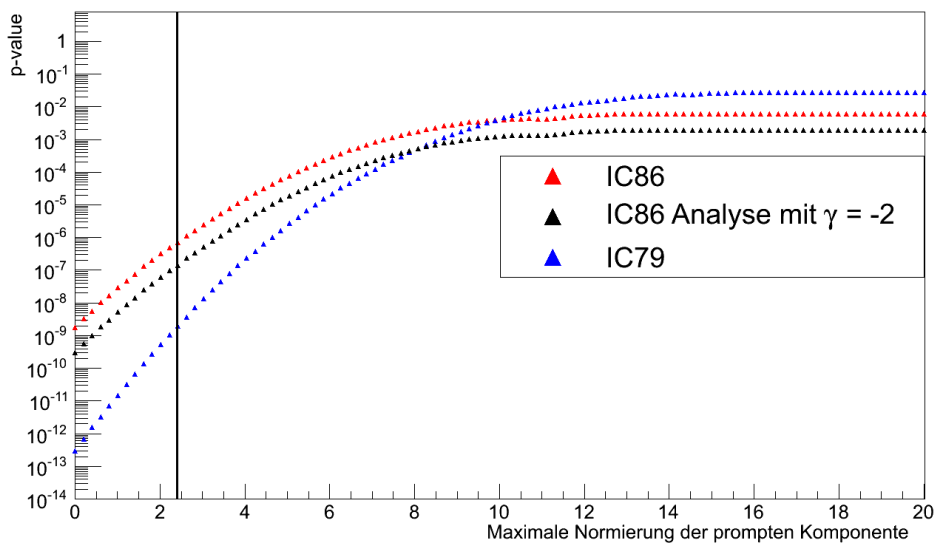
Die sich auf Grundlage von Gleichung 7.8 ergebene Signifikanz ist  $2,5 \sigma$  für  $H_1$ . Die so erhaltene Signifikanz liegt über, der in [Sch14] beschriebenen Signifikanz von etwa  $1,9 \sigma$ . Der Unterschied stammt aus der Nutzung der zusätzlichen Messpunkte der AMANDA und ANTARES Entfaltung. Durch die größere Anzahl an Messpunkten ergibt sich eine geringere maximal erlaubte Normierung der prompten Komponente im Fit. Es jedoch ist nicht möglich, durch die zusätzlichen Messpunkte die Entartung zwischen der prompten Komponente und dem hochenergetischen extraterrestrischen Fluss vollständig aufzulösen. Es ist jedoch möglich, die Signifikanz zu steigern.

Wie in Unterabschnitt 7.3.1 gezeigt wird, weist die prompte Komponente eine starke Entartung mit der hochenergetischen extraterrestrischen Komponente auf. Wird  $H_0$  so verändert, dass die prompte Komponente gleich null gesetzt wird, ergibt sich ei-

## 7. Analyse des entfalteten Energiespektrums



(a) Signifikanz in Abhängigkeit von der Normierung der prompten Komponente. In rot sind die Ergebnisse, des in dieser Analyse beschriebenen globalen Fits dargestellt.



(b) Signifikanz ausgedrückt als p-value, dass der gemessene Exzess durch eine zufällige Fluktuation in den Daten zu erklären ist, in Abhängigkeit von der Normierung der prompten Komponente.

### Abbildung 7.20.:

Signifikanz des beobachteten Exzesses in Abhängigkeit der maximal erlaubten Normierung der prompten Komponente in  $H_0$ . In rot sind die Ergebnisse dieser Analyse aufgetragen. In schwarz aufgetragen sind die Ergebnisse dieser Analyse, jedoch wurde der spektrale Index der extraterrestrischen Neutrinos auf  $\gamma = -2,0$  fest eingestellt. In blau sind die Ergebnisse der IC79 Analyse [Sch14] eingezeichnet. Der aktuelle beste Fit der Normierung der prompten Komponente, beschrieben in  $[A^+ 14e]$ , ist  $0,00^{+2,41}_{-0,00}$ . Die schwarze Linie entspricht der  $1\sigma$ -Umgebung dieses Fits.

ne Signifikanz von  $5,91 \sigma$ . Es zeigt sich, dass der im Spektrum beobachtete Exzess nicht mit konventionellen atmosphärischen Neutrinos erklärt werden kann. Der Zusammenhang zwischen Signifikanz und Normierung der prompten Komponente wird in Abbildung 7.20(a) dargestellt. In der Abbildung wird die Signifikanz der Analyse für die maximal im Fit von  $H_0$  erlaubte Normierung der prompten Komponente aufgezeigt. Es zeigt sich, dass sich die maximale Signifikanz für eine minimale Normierung der prompten Komponente ergibt.

Die Signifikanz in Einheiten von  $\sigma$  lässt sich leicht über die Standardnormalverteilung in einen p-value umrechnen. Der p-value beschreibt die Wahrscheinlichkeit, mit der ein gegebenes Ergebnis durch  $H_0$  zu erklären ist. Eine Darstellung der Ergebnisse findet sich in Abbildung 7.20(b).

Für eine steigende prompte Normierung fällt die Signifikanz annähernd linear ab, um bei einem Wert von etwa zehn zu sättigen. Zum Vergleich wird in Abbildung 7.20(a) ebenfalls das Ergebnis von [Sch14] aufgetragen. Die Analyse zeigt ein ähnliches Verhalten. Für minimale prompte Normierungen ergibt sich eine leicht höhere Signifikanz, dafür erreicht sie für eine quasi freie prompte Normierung eine geringere Signifikanz. An diesem Verhalten ist zu erkennen, dass die Daten nicht ohne eine Komponente zusätzlich zu den konventionellen atmosphärischen Neutrinos erklärt werden können. Zusammen mit den Ergebnissen von [Sch14] und [A<sup>+</sup>14c] ist die wahrscheinlichste Erklärung für das beobachtete Verhalten ein Fluss von hochenergetischen extraterrestrischen Neutrinos.

Wird der spektrale Index im Fit fest auf  $\gamma = -2,0$  gesetzt, werden die Freiheitsgrade des Fits um eins reduziert. Daher ergibt sich aus Gleichung 7.8 eine höhere Signifikanz des Ergebnisses. Die Ergebnisse sind ebenfalls in Abbildung 7.20 aufgetragen. Die globale Signifikanz steigt durch die Fixierung des spektralen Index auf  $2,9 \sigma$ . Für eine nicht vorhandene prompte Komponente steigt die Signifikanz auf  $6,1 \sigma$ .

## 7.4. Einschränkung für den Fluss der prompten Neutrinos

Im vorherigen Abschnitt zeigt sich, dass auf Grund des beschränkten Wissens über die prompte Komponente der atmosphärischen Neutrinos nur bedingt Aussagen über den Fluss der extraterrestrischen Neutrinos möglich sind. Insbesondere zeigt sich in Abbildung 7.20, dass die Signifikanz für extraterrestrische Neutrinos stark von der Normierung des Flusses der prompten Neutrinos abhängt. Es ist daher von besonderem Interesse, den Fluss der prompten Neutrinos besser bestimmen zu können.

## 7. Analyse des entfalteten Energiespektrums

---

Ein erster Ansatz ist einen Fit durchzuführen, bei dem der in den Daten gemessene hochenergetische Exzess nur durch den prompten Fluss erklärt wird. Die Likelihood-Funktion für die Normierung der prompten Komponente findet sich in Abbildung 7.21. Das Minimum der Likelihood-Funktion ergibt sich zu  $N_{\text{Prompt}} = 10,7^{+2,2}_{-2,0}$ . Da ein Fluss von extraterrestrischen Neutrinos nachgewiesen wurde, kann ausgeschlossen werden, dass der gemessene Exzess vollständig durch den Fluss der prompten Neutrinos erklärt wird. Der Fit kann daher als ein Maximum für den Fluss der prompten Neutrinos angesehen werden.

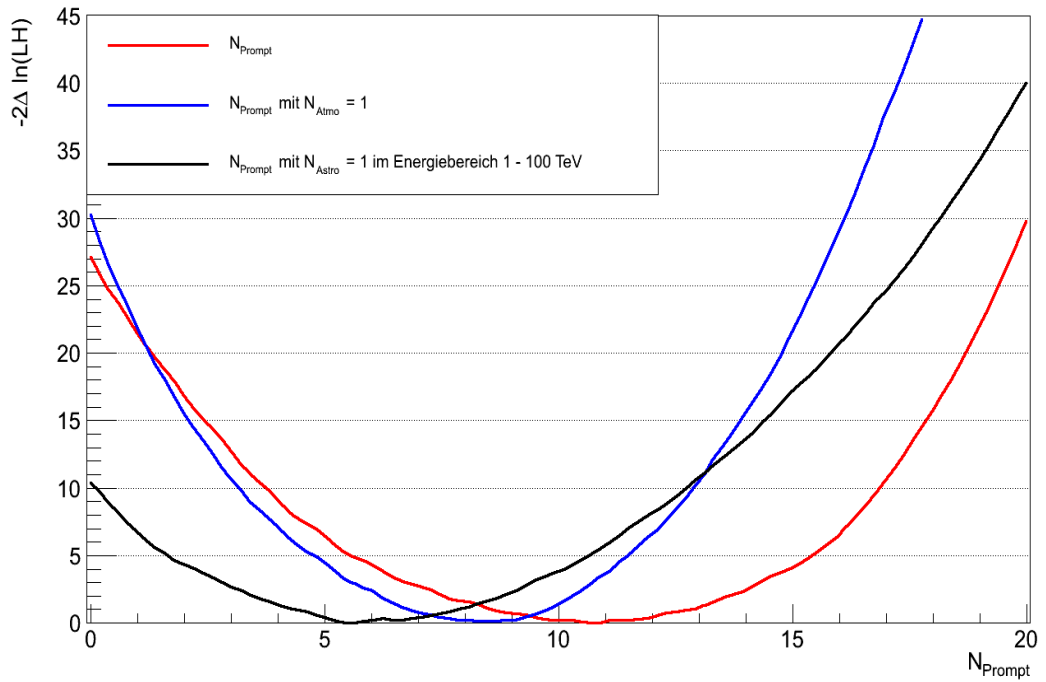
Betrachtet man die Wechselwirkung zwischen der Normierung des prompten und konventionellen Flusses für diesen Fit, so ist eine deutliche Korrelation zwischen den beiden Komponenten zu erkennen. Wie jedoch aus den Vergleichen zwischen Modellvorhersagen und Daten deutlich wurde, werden die Datenpunkte bereits sehr gut durch die Vorhersage für den konventionellen Fluss beschrieben. Es ist daher vertretbar, den Fluss bei  $1 \times N_{\text{Atmo}}$  zu fixieren. Durch diese Einschränkung ergibt sich das Minimum, der in Abbildung 7.21 dargestellten Likelihood-Funktion zu  $N_{\text{Prompt}} = 8,5^{+1,4}_{-1,6}$ . Im Vergleich zum vorherigen Fit verringert sich der maximal vom Fit erlaubte prompte Fluss deutlich.

Für eine weitere Einschränkung des prompten Flusses bietet es sich an, ausschließlich den Energiebereich  $[1; 100]$  TeV im Fit zu betrachten. Für höhere Energien wird der Einfluss der extraterrestrischen Komponente zunehmend dominant, was eine Bestimmung der maximal erlaubten prompten Komponente erschwert. Energien kleiner einem TeV werden ebenfalls nicht berücksichtigt, um den Fit nicht durch mögliche Unsicherheiten der theoretischen Beschreibung in diesem Bereich zu beeinflussen. Alle anderen Fitparameter bleiben im Vergleich zu den bisherigen Untersuchungen jedoch unverändert. Der beste Fit der prompten Komponente ergibt sich damit zu  $N_{\text{Prompt}} = 5,5^{+2,4}_{-1,0}$ .

Wie der Vergleich zu den bisherigen Fits in Abbildung 7.21 zeigt, ergibt sich, wie erwartet ein wesentlich geringerer Wert für die Normierung der prompten Komponente. Jedoch vergrößert sich der Fehler, so dass insgesamt nur eine bedingt bessere Eingrenzung möglich ist. Für diesen Fit gibt es ebenfalls nicht die oben beschriebene starke Abhängigkeit zwischen den Normierungen des prompten und atmosphärischen Flusses. Dieser Umstand erklärt sich durch die wesentlich höhere Anzahl an Messwerten im Bereich eines dominanten atmosphärischen Flusses. Der beste Fit der atmosphärischen Normierung ergibt sich zu  $N_{\text{Atmo}} = 0,98^{+0,06}_{-0,03}$ . Das Ergebnis ist konsistent mit dem in Tabelle 7.2 beschriebenen besten Fit der atmosphärischen Komponenten.



#### 7.4. Einschränkung für den Fluss der prompten Neutrinos



**Abbildung 7.21.:**

In rot ist die Likelihood-Funktion für die Normierung der prompten Neutrinos unter der Annahme, dass es keine extraterrestrischen Neutrinos gibt, dargestellt. Die in blau eingezeichnete Likelihood-Funktion ergibt sich durch die zusätzliche Bedingung einer Normierung der konventionellen Neutrinos. In schwarz ist die Likelihood-Funktion für die Normierung der prompten Neutrinos für den Fit im Energiebereich [1; 100] TeV eingezeichnet.

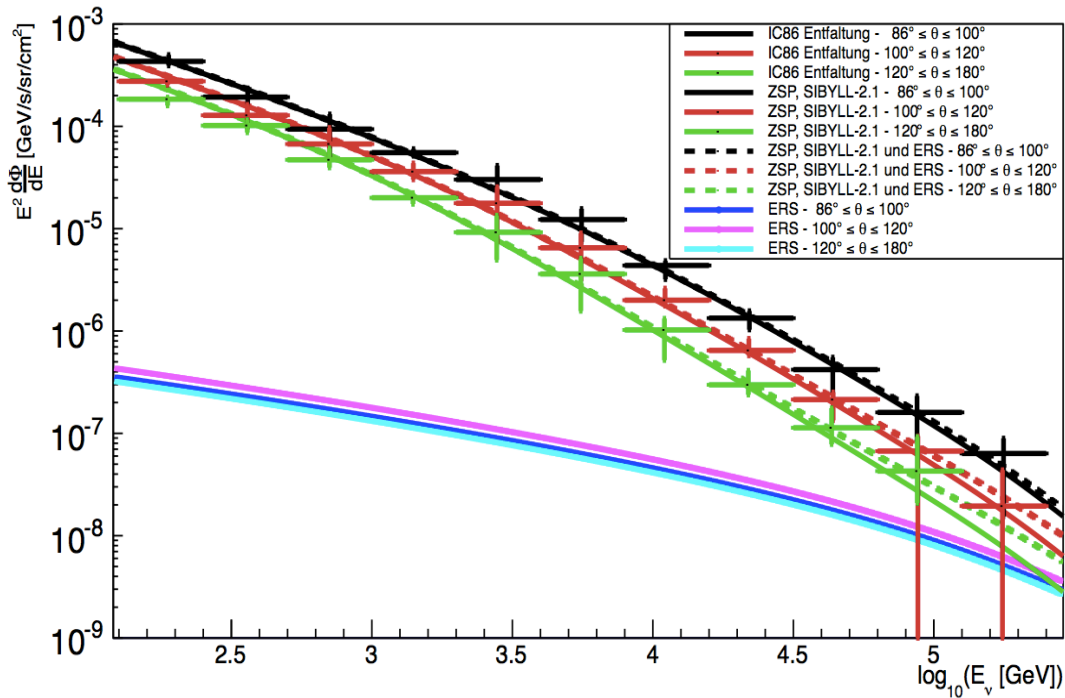
Der Fluss der prompten und konventionellen Komponente der atmosphärischen Neutrinos weist eine unterschiedliche Abhängigkeit vom Zenit  $\theta$  auf. Der Fluss der prompten Neutrinos ist in erster Näherung unabhängig vom Zenit. Der Fluss

$$\Phi_{\text{konv}} \propto \frac{1}{\cos(\theta)} \quad (7.9)$$

der konventionellen Neutrinos weist jedoch eine deutliche Abhängigkeit vom Zenit auf.

In Abschnitt 6.4 wird die Entfaltung des Energiespektrums in drei Zenitbereichen beschrieben. Diese Ergebnisse werden in Abbildung 7.22 mit der Flussvorhersage für den konventionellen Fluss nach [A<sup>+</sup>11c] und dem prompten Fluss nach [ERS08] verglichen. In der Abbildung ist deutlich zu erkennen, dass der konventionelle Fluss für die unterschiedlichen Zenitbereiche unterschiedlich ist. Dieses Verhalten spiegelt die, in Gleichung 7.9 beschriebene, Abhängigkeit des Flusses vom Zenit wieder. Es wird

## 7. Analyse des entfalteteten Energiespektrums



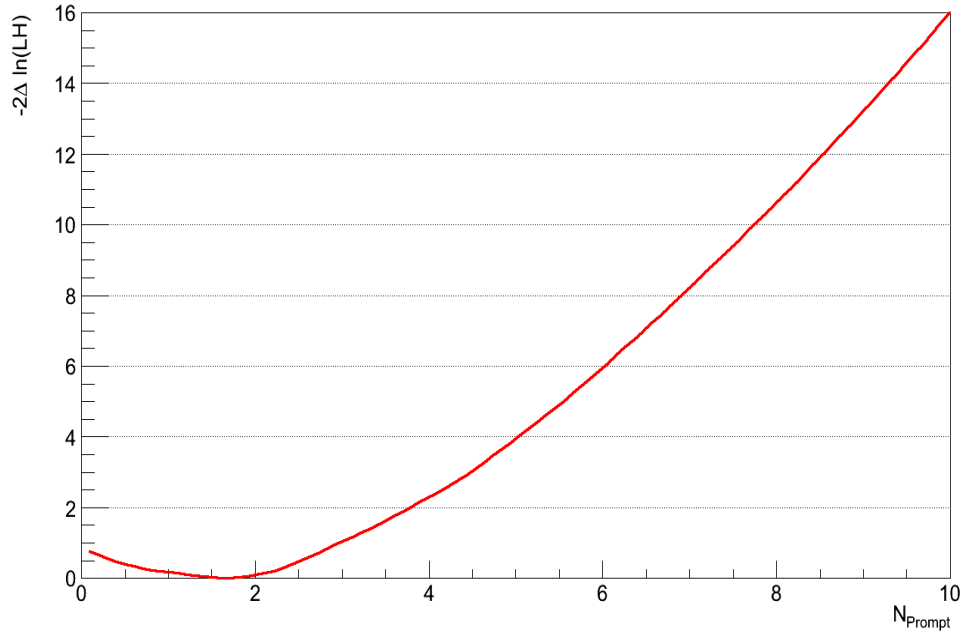
**Abbildung 7.22.:**

Entfaltete Energiespektren in drei Zenitbereichen im Vergleich zur Vorhersage. Der konventionelle Fluss basiert auf ZSP [A<sup>+</sup>11c] und SYBILL-2.1 [AEG<sup>+</sup>09], der prompte Fluss entspricht ERS [ERS08].

ebenfalls deutlich, dass die Abhängigkeit des prompten Flusses wesentlich geringer und im Vergleich zum konventionellen Fluss anders ist. Es bietet sich daher an, dieses unterschiedliche Verhalten zu nutzen, um eine bessere Eingrenzung der prompten Komponente zu erreichen. [FTD12, Gai90]

Da es in den Winkelbereichen der drei Spektren keine Überschneidung gibt, können die drei Entfaltungen als statistisch unabhängig betrachtet werden. Ebenfalls ist in dem Energiebereich der Spektren der extraterrestrische hochenergetische Fluss nicht dominant. Daher kann analog zur bisherigen Vorgehensweise die prompte Komponente durch einen Fit bestimmt werden ohne diese zusätzliche Komponente zu betrachten. Es bietet sich an, die drei Spektren der Zenitbereiche für einen Fit ebenfalls um IC59, AMANDA und ANTARES Ergebnisse zu ergänzen. Führt man den Fit aus, ergibt sich die in Abbildung 7.23 beschriebene Likelihood-Funktion für die Normierung der prompten Komponente. Der sich daraus ergebende beste Fit ergibt sich zu  $N_{\text{Prompt}} = 1,7^{+1,3}_{-1,7}$ . Der beste Fit der konventionellen Komponente ergibt sich zu  $N_{\text{Atmo}} = 1,03^{+0,04}_{-0,06}$ .

Vergleicht man die Ergebnisse des Fits mit den Ergebnissen des Fits unter Berücksichtigung der Entfaltung des Spektrums im vollen Winkelbereich aus Tabelle 7.2 ,ergibt



**Abbildung 7.23.:**

In rot ist die Likelihood-Funktion für die Normierung der prompten Neutrinos für den Fit der drei Zenitbereichentfaltungen dargestellt. Der Fit erfolgte unter Vernachlässigung der extraterrestrischen Neutrinos, da diese im Energiebereich nicht dominant sind.

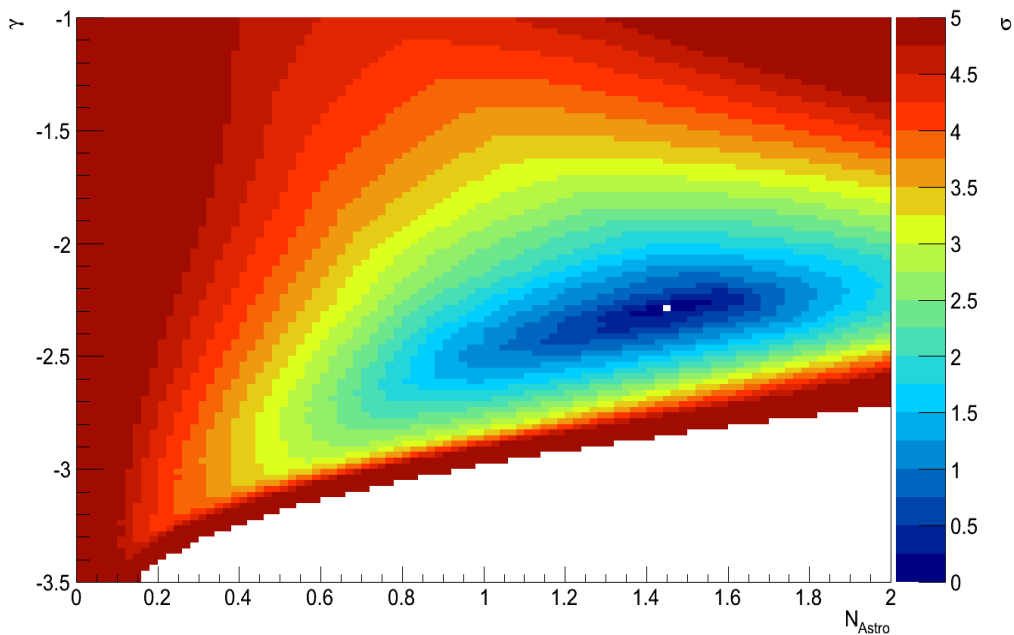
sich, dass die Ergebnisse für die Normierung der konventionellen Neutrinos im Fehler übereinstimmen. Es ist daher davon auszugehen, dass das genutzte Modell die Daten gut beschreibt. Für den Fluss der prompten Komponente ergibt sich ebenfalls ein konsistentes Bild, da die beiden Ergebnisse ebenfalls im Fehler übereinstimmen. Der Fit in den unterschiedlichen Zenitbereichen ist jedoch in der Lage, den Fluss wesentlich besser einzugrenzen. Dies gilt ebenfalls im Vergleich zu den Fits mit Einschränkungen auf dem vollen Zenitbereich.

Die Berechnung der Signifikanz, in Abhängigkeit von der prompten Normierung des Fits, an das Spektrum im vollen Zenitbereich aus Abbildung 7.20, ergibt für einen Wert von  $N_{\text{Prompt}} = 1,7^{+1,3}_{-1,7}$  eine Signifikanz von  $5,2^{+0,7}_{-0,6} \sigma$ . Für den Fit mit  $\gamma = 2,0$  ergibt sich eine Signifikanz von  $5,5^{+0,7}_{-0,6} \sigma$ .

Der bisher beste Fit der prompten Komponente wird in [A<sup>+</sup>14e] beschrieben. Der beste Fit ergibt sich zu  $N_{\text{Prompt}}^{\text{IceCube}} = 0,00^{+2,41}_{-0,00}$ . Die Ergebnisse beziehen sich ebenfalls auf das prompte Modell nach [ERS08]. Vergleicht man den in diesem Abschnitt berechneten besten Fit der Normierung der prompten Komponente von  $N_{\text{Prompt}} = 1,7^{+1,3}_{-1,7}$  mit

diesem Ergebnis, ist zu erkennen, dass beide Ergebnisse konsistent sind. Es ist jedoch ebenfalls deutlich, dass die Unsicherheit im Fit verringert werden konnte.

### 7.5. Bestimmung der Modellparameter mit fester prompter Komponente



**Abbildung 7.24.:**

Zweidimensionale Projektion der Likelihood-Funktion für die Normierung des hochenergetischen extraterrestrischen Flusses und des spektralen Index unter der Bedingung  $N_{\text{Prompt}} = 1,7$ . Das weiße Bin entspricht dem wahrscheinlichsten Fit.

Die Ausführungen in diesem Kapitel zeigen, dass es möglich ist, die prompte Komponente durch die Zuhilfenahme von zusätzlichen Informationen weiter einzuschränken. Unter der Bedingung dieser Einschränkung ergibt sich eine wesentlich höhere Signifikanz des Ergebnisses. Wertet man die Ergebnisse des Likelihood-Fits aus Abschnitt 7.3, unter der Bedingung einer festen prompten Normierung von  $N_{\text{Prompt}} = 1,7$  aus, ergibt sich die in Abbildung 7.25 aufgetragene Likelihood-Funktion für den spektralen Index und die Normierung des hochenergetischen astrophysikalischen Flusses.

Durch die Festlegung der prompten Komponente verändert sich die Likelihood-Funktion aus Abbildung 7.18 deutlich zu der in Abbildung 7.24 dargestellten Likelihood-Funktion.

## 7.5. Bestimmung der Modellparameter mit fester prompter Komponente

---

Der Fit zeigt eine noch deutlichere Präferenz für weiche Spektren der hochenergetischen extraterrestrischen Neutrinos. Der beste Fit für diese Komponente ergibt  $\gamma = -2,3 \pm 0,2$  und  $N_{\text{Astro}} = 1,5^{+0,3}_{-0,4}$ . Der beste Fit der atmosphärischen Normierung ergibt sich zu  $N_{\text{Atmo}} = 1,02 \pm 0,06$ . Das Ergebnis ist somit konsistent im Vergleich zum Fit ohne Einschränkung der prompten Komponente.

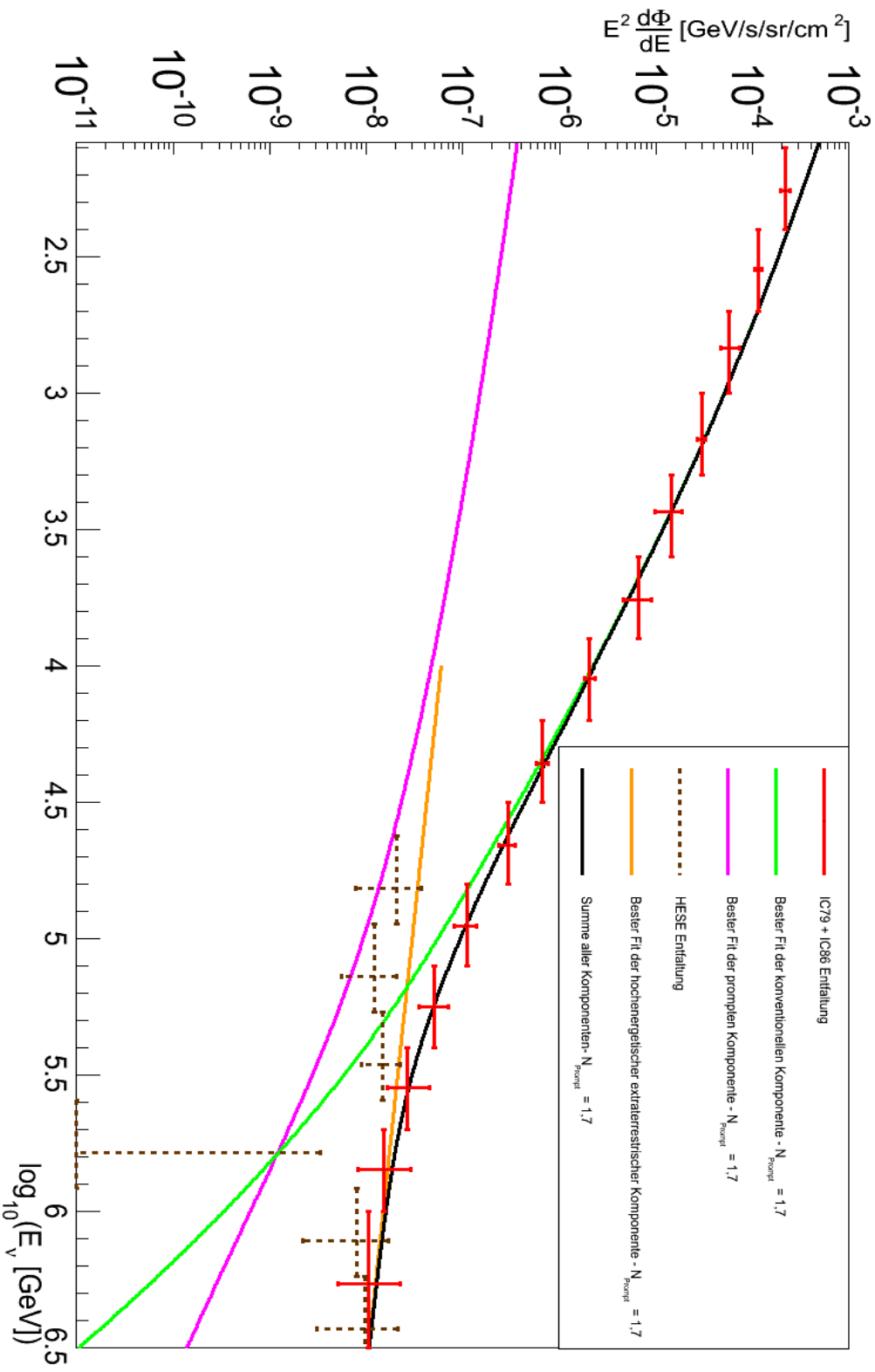
Eine Darstellung des besten Fits im Vergleich zum gemittelten Spektrum und der HESE Entfaltung findet sich in Abbildung 7.25. Wie zu erkennen ist, beschreibt der beste Fit die Datenpunkte der beiden Entfaltungen hinreichend. Der beste Fit der HESE Analyse für die Normierung des Flusses der hochenergetischen extraterrestrischen Neutrinos bei einer Energie von 1 PeV ist  $N_{\text{Astro}}^{\text{HESE}} = 0,75 \times 10^{-8} \left[ \frac{\text{GeV}}{\text{sr s cm}^2} \right]^2$ . Es ist daher erkennbar, dass im Vergleich zur HESE Entfaltung eine höhere Normierung des Flusses präferiert wird. Der beste Fit der HESE Entfaltung für den spektralen Index ist  $\gamma^{\text{HESE}} = -2,3 \pm 0,3$ . Dieses Ergebnis passt sehr gut zu den Ergebnissen des beschriebenen Fits, es weist jedoch im Vergleich eine größere Unsicherheit auf. Es kann also geschlossen werden, dass die beiden Analysen konsistent sind. [A<sup>+</sup>14c]

Eine abschließende Wertung und Zusammenfassung aller Ergebnisse dieser Arbeit wird im folgenden Kapitel vorgenommen.

---

<sup>2</sup>In [A<sup>+</sup>14c] wird keine Unsicherheit für dieses Ergebnis angegeben.

## 7. Analyse des entfalteten Energiespektrums



**Abbildung 7.25.:**  
 Das Ergebnis des besten Fits unter der Bedingung  $N_{\text{prompt}} = 1,7$  im Vergleich zum gemittelten Spektrum aus der 1079 [Sch14] und 1086 Analyse und der HESE Entfaltung [A<sup>+</sup>14c]. Die dargestellte hochenergetische extraterrestrische Komponente ist kompatibel mit der HESE Entfaltung.

## 8. Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurde eine Methodik zur Separation eines hochreinen und effizienten Datensatzes von Myon-Neutrinos anhand der Daten des IceCube-Detektors in der 86-String Konfiguration entwickelt. Die Analysetechniken basieren auf den in [Ruh13] beschriebenen Methoden. Im Vergleich wurden die Vorselektionschnitte sowie der Confidenceschnitt deutlich verbessert. Insbesondere der zweidimensionale Schnitt in Confidence und einem Energieschätzer ist eine Neuerung, die für die Erzeugung von Neutrinospektren bisher noch nicht genutzt wurde. Durch diese Verbesserungen ergibt sich eine deutlich erhöhte Effizienz im Bereich der Hochenergie. Die entscheidendste Verbesserung ist jedoch, die Weiterentwicklung des Bootstrapping. Mit dieser Methode ist eine wesentlich bessere Schätzung der Randverteilungen einer beliebigen Verteilung möglich. Diese Methode ist die bedeutendste methodische Entwicklung im Rahmen dieser Arbeit. Diese Neuentwicklung kann in jeder Analyse, die eine multivariante Methode nutzt, einen Beitrag zu einem besseren Verständnis der Daten liefern, da durch diese Methode die bestehenden MC und Daten wesentlich besser genutzt werden können. Hiermit kann die effektive Menge an MC um einen Faktor 10 bis 100 gesteigert werden. Die dazu erforderliche Menge an Computing-Ressourcen beträgt jedoch nur wenige CPU-Stunden.

Unter Verwendung der neuen und optimierten Methoden wurde ein Datensatz von mehr als 93 000 Neutrinoereignissen erzeugt. Dieses Ergebnis liegt deutlich über den 28 000 in [Ruh13] und 63 000 in [Sch14]. Die Reinheit des Datensatzes liegt bei  $(99,6 \pm 0,2)\%$  und ist damit höher als in den beiden genannten Analysen. Die Effizienz des Datensatzes liegt im Bereich der Hochenergie bei über 60 %, was ebenfalls eine deutliche Verbesserung gegenüber vergleichbaren Analysen darstellt. Der in Abbildung 5.18 dargestellte Vergleich der Effizienz verschiedener IceCube-Datensätze zeigt, dass der vorliegende Datensatz die aktuell effizienteste Trennung zwischen Signal und Untergrund aufweist.

Eine Betrachtung der hochgradig mit der Neutrinoenergie korrelierten Attribute, wie die Rekonstruktion der Myonenergie, beschrieben in Abbildung 5.19, gibt bereits die

ersten Hinweise darauf, dass die Daten nicht ausschließlich mit dem Fluss der konventionellen atmosphärischen Neutrinos erklärt werden können.

Die Software TRUUEE wurde im nächsten Schritt verwendet, um das Energiespektrum der von IceCube gemessenen Myon-Neutrinos zu bestimmen. Das Spektrum wurde zwischen 200 GeV und 3,3 PeV entfaltet. Die entsprechenden Ergebnisse inklusive aller Unsicherheiten sind in Tabelle 6.1 aufgeführt und in Abbildung 6.5 dargestellt. Das Spektrum zeigt, dass der gemessene Fluss ab etwa 50 TeV über der Vorhersage für konventionelle Neutrinos liegt. Ebenfalls war es auf Grund der Größe des Datensatzes erstmals möglich, den Fluss der atmosphärischen Neutrinos in drei Zenitbereichen zu entfalten.

Trotz deutlich verbesserter Effizienz der Separation im Bereich der Hochenergie zeigt sich der gefundene Exzess oberhalb der Vorhersage für konventionelle atmosphärische Neutrinos nicht so signifikant wie im IC79 Spektrum. [Sch14].

Da der Exzess über dem Fluss der konventionellen atmosphärischen Neutrinos gering ist, zeigt sich der Einfluss erst ab einer Neutrinoenergie von etwa 50 TeV. Um eine genauere Bestimmung zuzulassen, wurden weitere Datenpunkte der IC59, IC79, AMANDA und ANTARES Entfaltung herangezogen. Mittels einer Likelihood-Methode wurde versucht festzustellen, wie der beobachtete Exzess im Spektrum über dem Fluss der konventionellen atmosphärischen Neutrinos erklärt werden kann. Durch diese Vorgehensweise war es möglich, die Signifikanz für eine Erklärung des Exzesses durch hochenergetische extraterrestrische Neutrinos im Vergleich zu [Sch14] von  $1,9 \sigma$  auf  $2,5 \sigma$  zu steigern. Es zeigt sich, dass auf diese Weise keine eindeutige Antwort auf den Ursprung des Exzesses auf Grundlage der verwendeten Daten gefunden werden kann, wenn es nicht gelingt die prompte Komponente der atmosphärischen Neutrinos besser einzugrenzen. Zu diesem Zweck wurden die drei Entfaltungen der unterschiedlichen Zenitbereiche sowie die Daten der IC59, AMANDA und ANTARES Entfaltung für einen weiteren Likelihood-Fit genutzt mit dem Ziel, die Normierung der prompten Komponente der atmosphärischen Neutrinos zu bestimmen. Im Rahmen dieser Untersuchung ergibt sich ein mit [A<sup>+</sup>14e] kompatibles Ergebnis, die Unsicherheiten konnten jedoch im Vergleich verringert werden. Unter der Bedingung des Fitergebnisses für die prompte Komponente ergibt sich eine Signifikanz von  $5,2 \sigma$ .

Durch die Bestimmung der prompten Komponente war es ebenfalls möglich, den hochenergetischen extraterrestrischen Fluss besser zu bestimmen. Das beste Fitergebnis des spektralen Index ist identisch mit dem in [A<sup>+</sup>14c], jedoch mit geringeren Unsicherheiten. Daraus kann geschlossen werden, dass die Ergebnisse von [A<sup>+</sup>14c] unabhängig



---

bestätigt werden können. Dies ist von besonderer Bedeutung, da [A<sup>+</sup>14c] hauptsächlich Ereignisse von der Südhalbkugel der Erde nutzt, wohingegen die vorliegende Analyse hauptsächlich Ereignisse von der Nordhalbkugel nutzt. Somit ergänzen sich die Ergebnisse zu einem in sich stimmigen Bild eines Flusses isotroper hochenergetischer extraterrestrischer Neutrinos.

Im Rahmen dieser Arbeit war es möglich, einen isotropen Fluss hochenergetischer extraterrestrischer Neutrinos nachzuweisen. Der Fluss ist damit jedoch nicht vollständig vermessen und verstanden. Daher folgt nach der Zusammenfassung und Wertung der Ergebnisse ein Ausblick auf mögliche Analyseansätze, die über die beschriebene Analyse hinausgehen.



## 9. Ausblick

*„Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!“*

---

*(Lewis Carroll, Through the Looking-Glass, and What Alice Found There, 1871)*

Mit dem obigen Zitat von Carroll wird die Red-Queen-Hypothese der Evolution beschrieben. Die Hypothese besagt, dass egal wie gut ein Organismus an ein Ökosystem angepasst ist, er trotzdem aussterben wird, wenn er sich nicht schnell genug an neue Veränderungen anpassen kann und daher von anderen Organismen verdrängt wird. Dies lässt sich sehr gut auf Analysen in der Physik übertragen. Die gerade durchgeführte Analyse ist Vergangenheit und die Methoden müssen angepasst und verbessert werden, damit die nächste Analyse ebenfalls konkurrenzfähig ist. [VV73]

Die deutlichsten Verbesserungen lassen sich sehr wahrscheinlich durch einen verstärkten Einsatz der Erweiterung des Bootstrapping erreichen. Durch Einsetzen dieser Methode wird für jedes Ereignis die Wahrscheinlichkeitsverteilung der Confidence geschätzt. Anstatt nun nur einen Schnitt in dieser Verteilung vorzunehmen, um eine Separation zwischen Signal und Untergrund zu erreichen, kann auf Basis der Wahrscheinlichkeitsverteilung eine Separation durchgeführt werden. Insbesondere bietet es sich an, auf Grundlage der Momente der Verteilung eine Klassifikation vorzunehmen.

Die Klassifikation auf Grundlage der Momente kann darüber hinaus mit zusätzlichen Informationen wie Energie und Winkel kombiniert werden. Spätestens an dieser Stelle sollte die Separation jedoch nicht mehr einfach mit Schwellwerten durchgeführt werden, sondern mit Hilfe einer multivariaten Methode.

Der Fluss der verschiedenen Komponenten der von IceCube gemessenen Neutrinos zeigt ein unterschiedliches Verhalten im Zenitwinkel. Dieser Unterschied wurde im

Rahmen dieser Arbeit erfolgreich zur Bestimmung der prompten Komponente der atmosphärischen Neutrinos genutzt. Durch eine deutliche Steigerung der Anzahl der nutzbaren Ereignisse ist eine Aufteilung in mehr als drei Zenitbereiche möglich, was eine bessere Bestimmung der prompten und damit der hochenergetischen extraterrestrischen Komponente möglich machen würde. Um die Grundgesamtheit der Datensätze weiter zu erhöhen, bietet es sich an, das Schema der Datensätze auf Grundlage der Daten eines Jahres aufzugeben und die Daten mehrerer Jahre und Detektorkonfigurationen zu verbinden.

# A. Anhang

## A.1. Beschreibung der verwendeten Attribute

Alle in der vorliegenden Analyse verwendeten Attribute werden im folgenden aufgeführt. Dabei wird jedes Attribut kurz beschrieben. Eine Darstellung aller Attribute vor und nach der Separation findet sich in Abschnitt A.2 und A.3.

Die Attribute lassen sich in Gruppen von ähnlichen Attributen gliedern. Die Ähnlichkeit ergibt sich dabei entweder daraus, dass die Attribute vom gleichen Algorithmus oder, dass die Attribute von verschiedenen Algorithmen, die versuchen dasselbe zu rekonstruieren, generiert werden.

Jedes Attribut wird für die in Tabelle 5.1 aufgeführten Zeitfenster berechnet. Aus der Endung des Attributs ergibt sich das verwendete Zeitfenster. Wird das Attribut ohne Kürzel für das Zeitfenster angegeben, bezieht sich die Angabe auf das Zeitfenster C.

Die erste Gruppe von Attributen ergibt sich direkt aus der räumlichen und zeitlichen Verteilung, der während eines Ereignisses angesprochen DOMs. Diese sich so ergebenden Attribute sind:

**Z<sub>max</sub>** Maximale  $z$ -Position eines DOMs mit deponierter Ladung während eines Ereignisses, wobei das Zentrum des Detektors den Ursprung des Koordinatensystems bildet. [A<sup>+</sup>04]

**Z<sub>min</sub>** Minimale  $z$ -Position eines DOMs mit deponierter Ladung während eines Ereignisses. [A<sup>+</sup>04]

**r<sub>weight</sub>** Durch die Gewichtung des Abstands  $r_i$ , der in einem Ereignis angesprochenen DOMs mit der Menge der deponierten Ladung  $q_i$ , kann der mittlere gewichtete Abstand

$$r_{\text{weight}} = \frac{\sum_{i=1} q_i r_i}{\sum_{i=1} q_i} \quad (\text{A.1})$$

aller an einem Ereignis beteiligten DOMs zur rekonstruierten Trajektorie bestimmt werden.

**$\sigma_{\text{COG}, z}$**  Wird die deponierte Ladung  $q_i$  in den DOMs an der Position  $r_i$  während eines Ereignisses als ein Gewicht behandelt, kann der Schwerpunkt der Ladungsverteilung berechnet werden. Das Zentrum des Koordinatensystems ist durch den Mittelpunkt des Detektors gegeben. Das Attribut beschreibt die Standardabweichung in der Bestimmung der  $z$ -Position der Ladungsverteilung. [A<sup>+</sup>04]

**r<sub>COG</sub>** Wird die deponierte Ladung  $q_i$  wie ein Gewicht behandelt, kann der Schwerpunkt der Ladungsverteilung berechnet werden. Das Zentrum des Koordinatensystems ist durch den Mittelpunkt des Detektors gegeben. Das Attribut beschreibt den Radius des Ladungsschwerpunkts von der Mittelachse des Detektors. [A<sup>+</sup>04]

**d<sub>COG</sub>** Abstand des Ladungsschwerpunktes berechnet für Photonen, die im ersten und vierten zeitlichen Viertel des Ereignisses berechnet wurden. Für eine Kaskade sollte sich ein Wert von etwa 0 einstellen. Für Myonen sind abhängig von der Energie deutlich größere Werte zu erwarten. [A<sup>+</sup>04]

**Strings** Anzahl der Strings mit angesprochenen DOMs.

**DOMs** Anzahl der angesprochenen DOMs.

**Länge** Werden alle in einem Zeitfenster angesprochenen DOMs auf die rekonstruierte Trajektorie projiziert, ergibt der maximale Abstand von zwei Projektionen die Länge des Ereignisses.

**Leere** Werden alle angesprochenen DOMs eines Ereignisses auf die rekonstruierte Trajektorie projiziert, beschreibt die Leere den maximalen Abstand zwischen zwei aufeinander folgenden Projektionen.

**Stochastizität** Die Stochastizität beschreibt, wie gleichmäßig Ladung in den DOMs entlang der Trajektorie deponiert wurde. [A<sup>+</sup>11a]

Die zweite Gruppe von Attributen ergibt sich aus der Rekonstruktion der Trajektorien der Ereignisse, durch verschiedene Rekonstruktionsalgorithmen. Die unterschiedlichen Algorithmen verfolgen dabei jedoch verschiedene Ansätze. Daher kann sich aus der Benutzung von Attributen aus unterschiedlichen Algorithmen ein Vorteil für die Selektion ergeben.

**LineFit<sub>θ</sub>** Rekonstruktion der Trajektorie unter der Annahme von ebenen Wellen und der Vernachlässigung von Streuung. Das Attribut beschreibt den rekonstruierten Zenitwinkel. [A<sup>+</sup>14b]

**LineFit<sub>c</sub>** Der LineFit-Algorithmus berechnet die über  $x$ -,  $y$ -,  $z$ -Komponente gemittelte Lichtausbreitungsgeschwindigkeit des Ereignisses. Für Kaskaden ist ein Wert von etwa 0 zu erwarten. Für Myonen sind abhängig von der Energie deutlich größere Werte zu erwarten. [A<sup>+</sup>14b]

**MPEFitHighNoise<sub>r<sub>log</sub></sub>** Rekonstruktion der Trajektorie mittels Likelihood-Ansatz. Das Attribut ist der Wert der Likelihood-Funktion der Winkelrekonstruktion geteilt durch die um fünf verringerte Anzahl an angesprochenen DOMs.

**Paraboloid<sub>err1</sub>** Rekonstruktion der Trajektorie mittels Likelihood-Ansatz. Die Likelihood-Funktion wird in diesem Verfahren in eine quadratische Form gebracht. Das Attribut beschreibt die  $1\sigma$ -Umgebung für die Zenitrekonstruktion. [Cow98]

**Paraboloid<sub>err2</sub>** Das Attribut beschreibt die  $1\sigma$ -Umgebung für die Azimutrekonstruktion. [Cow98]

**MuEXAngular $4_\sigma$**  Rekonstruktion der Trajektorie mittels Likelihood-Ansatz. Das Attribut beschreibt die Unsicherheit in der Winkelrekonstruktion durch die Breite des Likelihoodminimums. [Cow98, A<sup>+</sup>14d]

**Bayesian $_z$**  Rekonstruktion der Trajektorie unter der Annahme, dass es sich bei dem Ereignis um ein atmosphärisches Myon handelt. Das Attribut beschreibt die rekonstruierte  $z$ -Position des Vertex. [A<sup>+</sup>14d]

**Bayesian $_\Theta$**  Das Attribut beschreibt den rekonstruierten Zenitwinkel. [A<sup>+</sup>14d]

**SPE $_{\text{status}}$**  Das Attribut beschreibt, ob die Rekonstruktion durch den SPEFit erfolgreich war. [A<sup>+</sup>14d]

**|SplineMPE $_\Theta$  – LineFit $_\Theta$ |** Differenz zweier rekonstruierter Zenitwinkel. Der Abstand ist ein Maß für die Qualität eines Ereignisses.

**SplineMPE $_\Theta$**  Das Attribut beschreibt den rekonstruierten Zenitwinkel. [A<sup>+</sup>14d]

Die dritte Gruppe von Attributen ergibt sich aus der Rekonstruktion der Myonenergie am Zentrum des Detektors. Eine ausführliche Beschreibung der verwendeten Attribute findet sich in [A<sup>+</sup>13e] und [A<sup>+</sup>14a].

**MuEX $_E$**  Rekonstruktion der Myonenergie mittels Likelihood-Ansatz. [A<sup>+</sup>14a]

**TEDOMS $_\sigma$**  Energieauflösung der zugehörigen Energieschätzung. [A<sup>+</sup>13e, A<sup>+</sup>14a]

**TEDOMS $_E$**  Rekonstruktion der Myonenergie. [A<sup>+</sup>13e, A<sup>+</sup>14a]

**TEBINS $_\sigma$**  Energieauflösung der zugehörigen Energieschätzung. [A<sup>+</sup>13e, A<sup>+</sup>14a]

**TEBINS $_E$**  Rekonstruktion der Myonenergie. [A<sup>+</sup>13e, A<sup>+</sup>14a]

## A.2. Verwendete Attribute auf Level 3

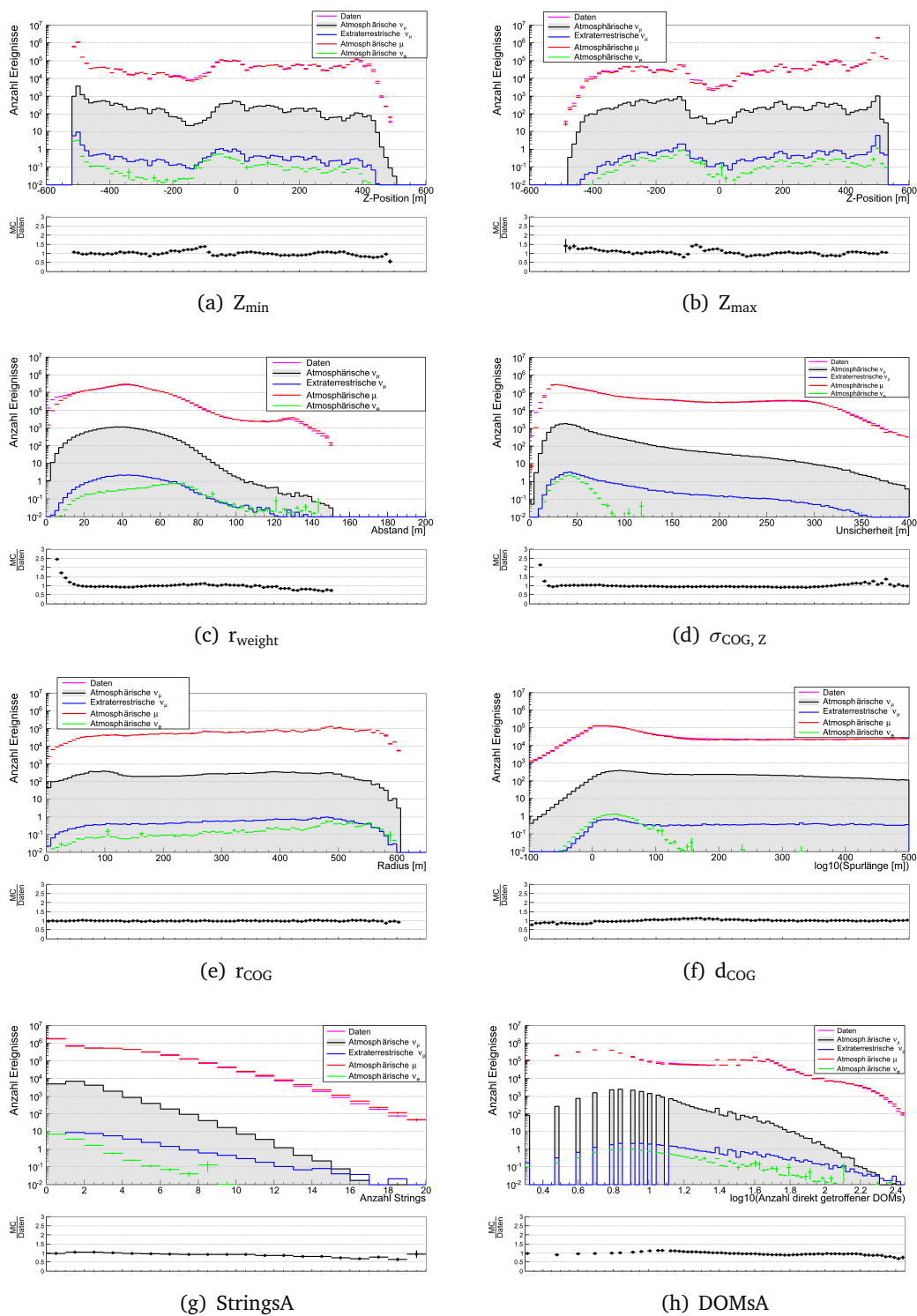


Abbildung A.1.: Verwendete Attribute auf Level 3



## A.2. Verwendete Attribute auf Level 3

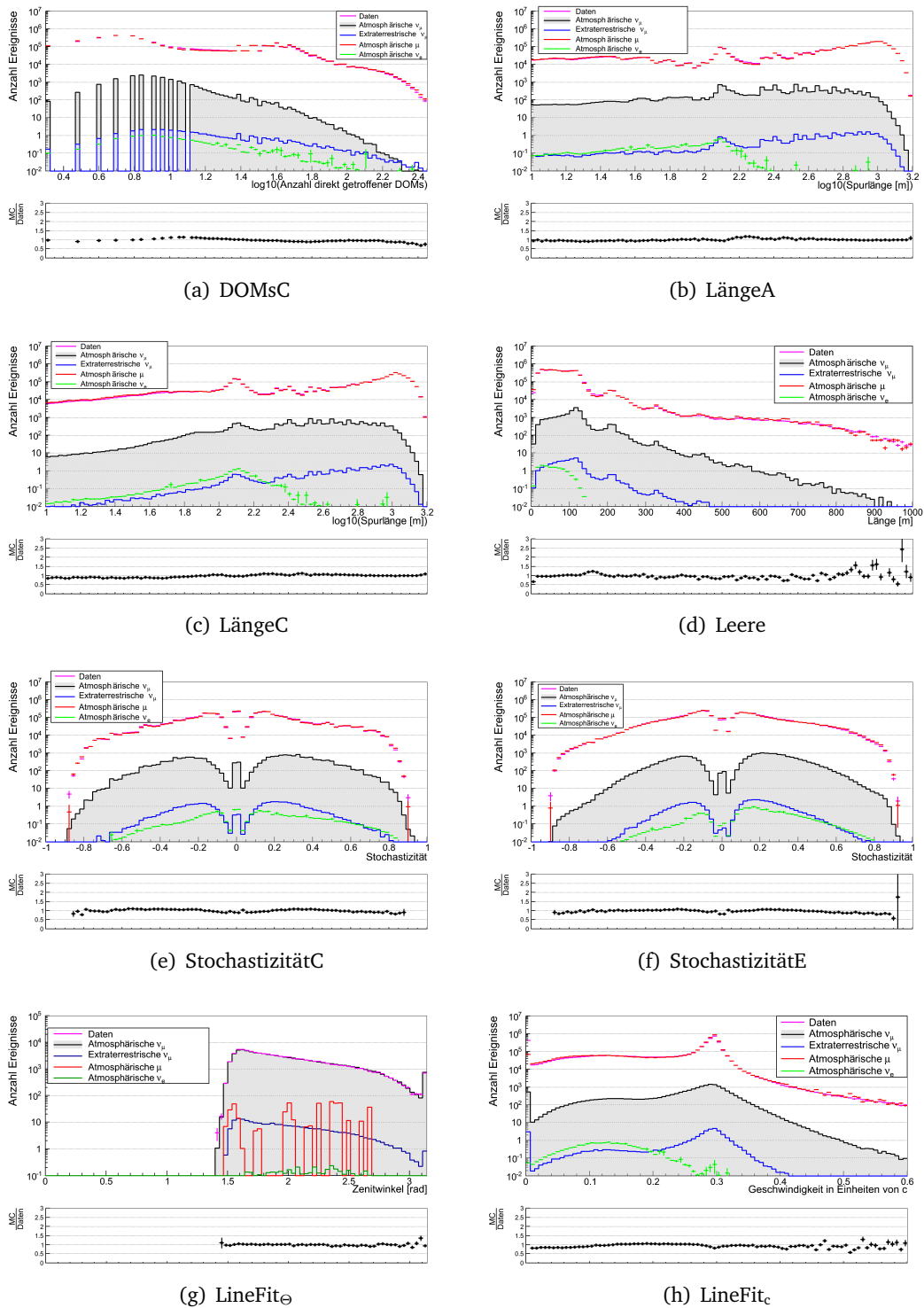
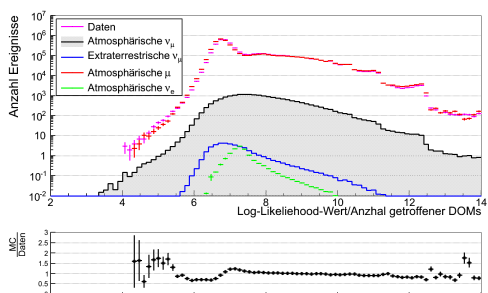
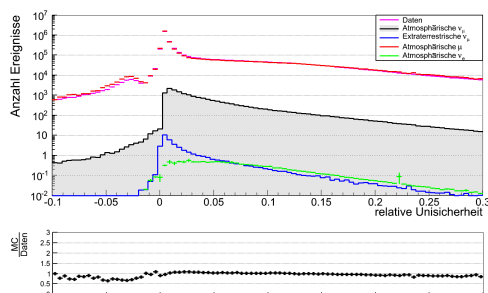


Abbildung A.2.: Verwendete Attribute auf Level 3

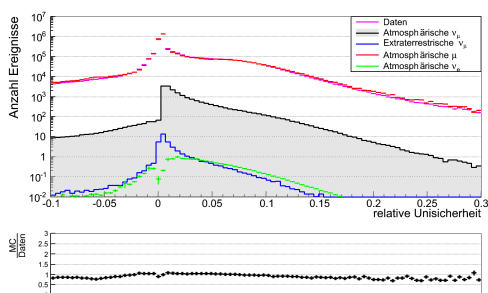
# A. Anhang



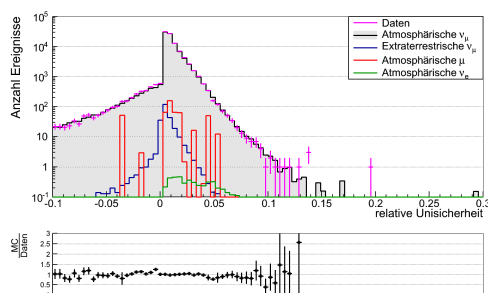
(a) MPEFitHighNoise<sub>err1</sub>



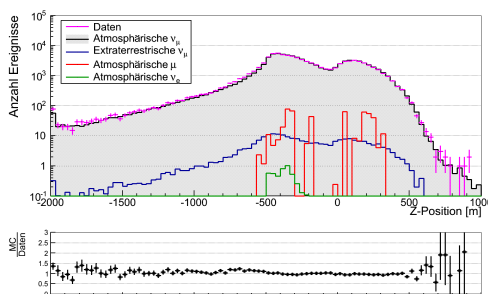
(b) Paraboloid<sub>err1</sub>



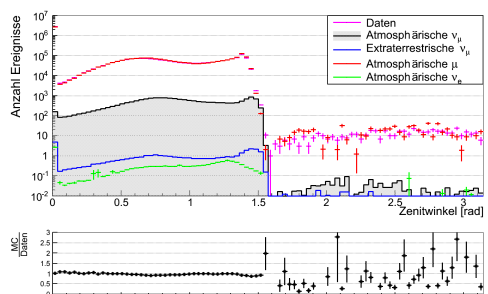
(c) Paraboloid<sub>err2</sub>



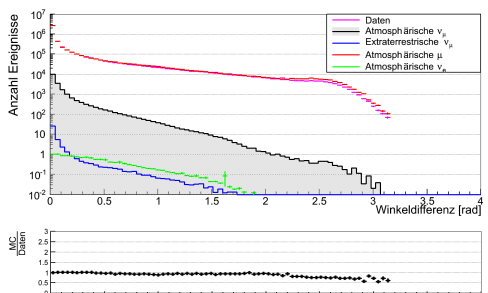
(d) MuEXAngular<sub>4σ</sub>



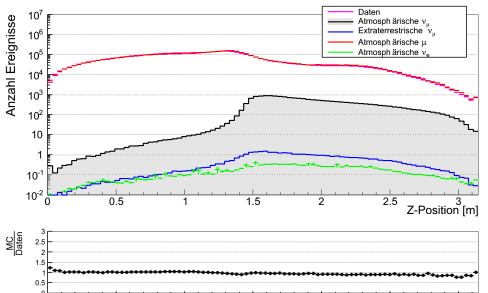
(e) Bayesian<sub>z</sub>



(f) Bayesian<sub>θ</sub>



(g) abs(SplieMPE<sub>θ</sub> - LineFit<sub>θ</sub>)



(h) SplieMPE<sub>θ</sub>

Abbildung A.3.: Verwendete Attribute auf Level 3

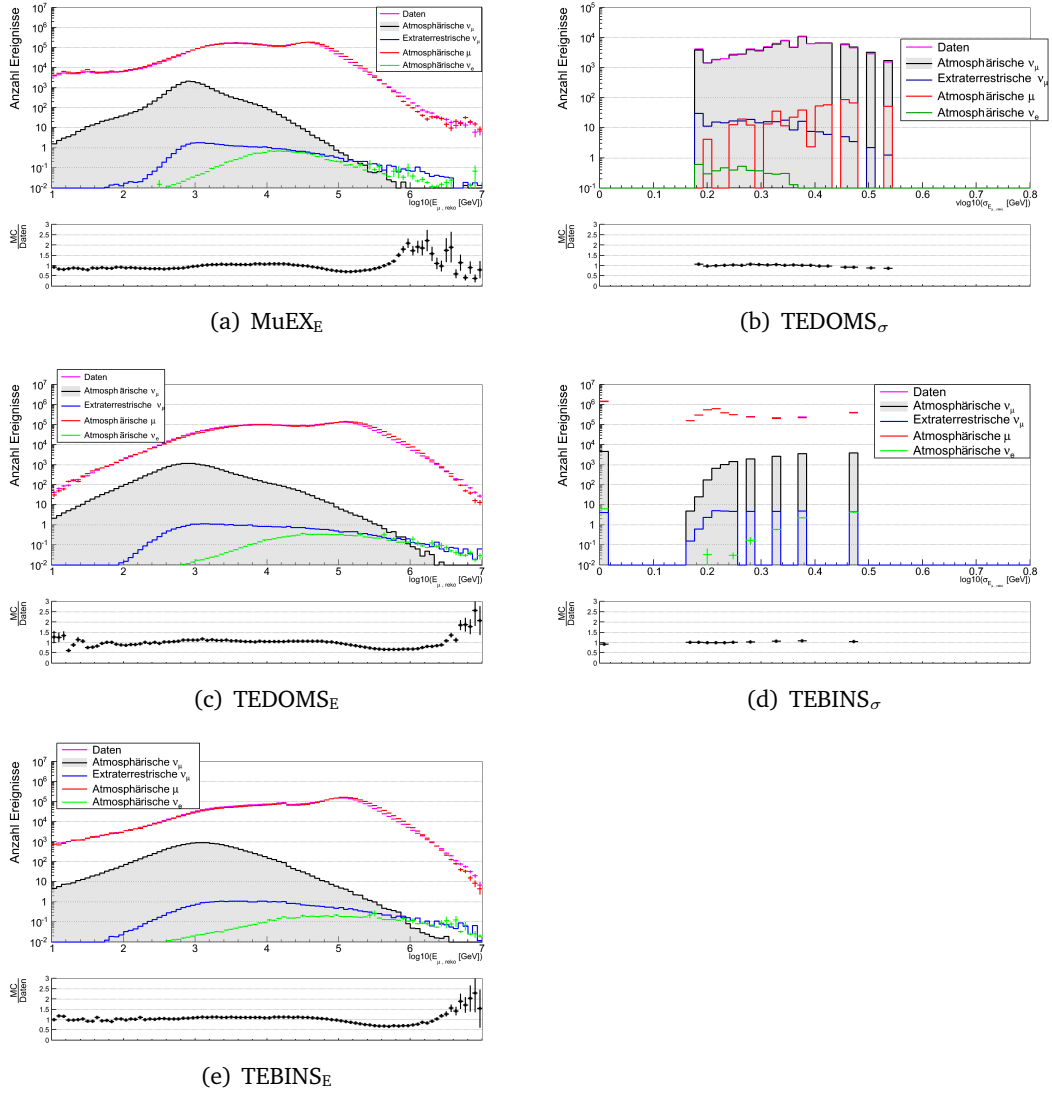


Abbildung A.4.: Verwendete Attribute auf Level 3



### A.3. Verwendete Attribute auf Level 5

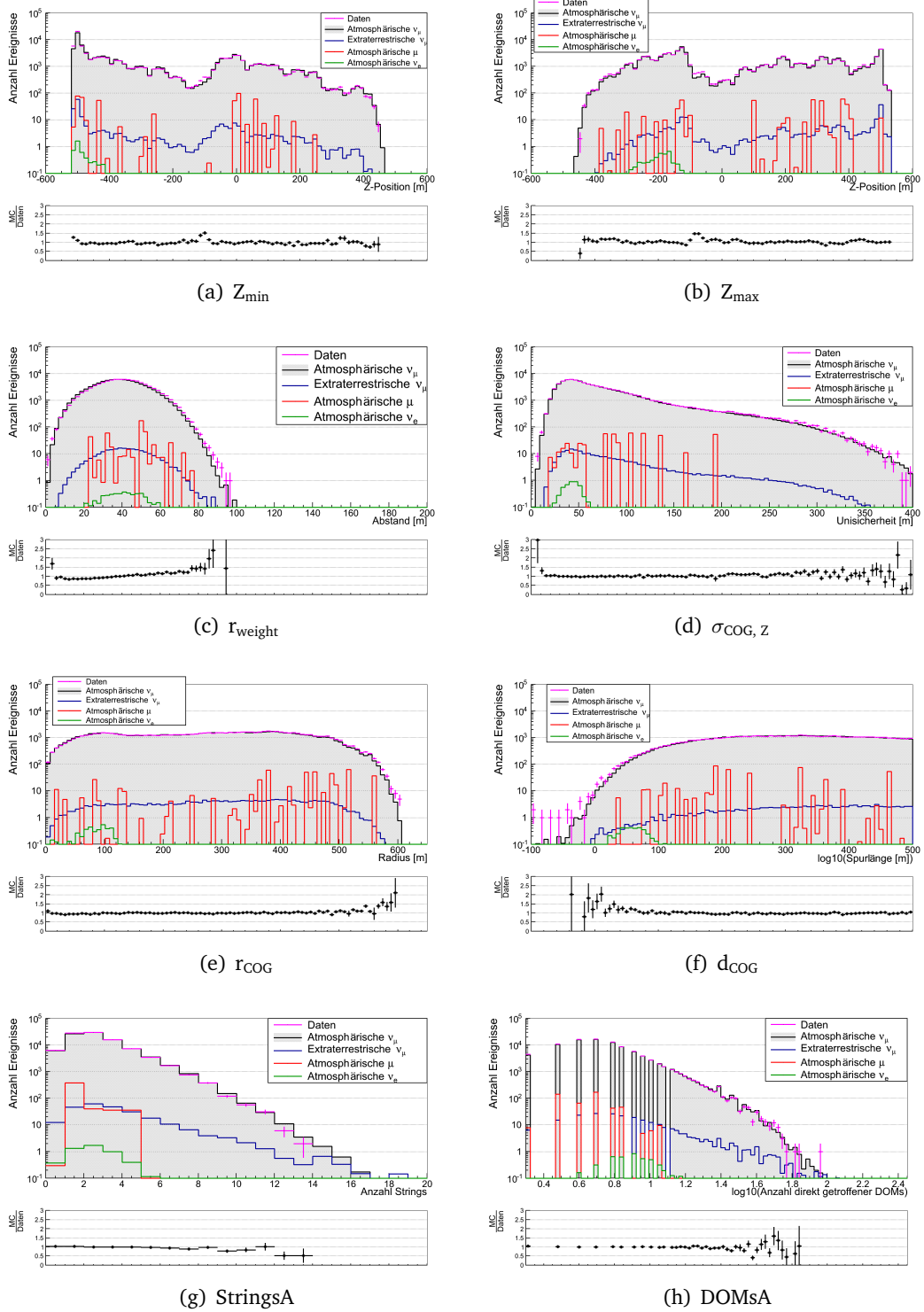


Abbildung A.5.: Verwendete Attribute auf Level 5

# A. Anhang

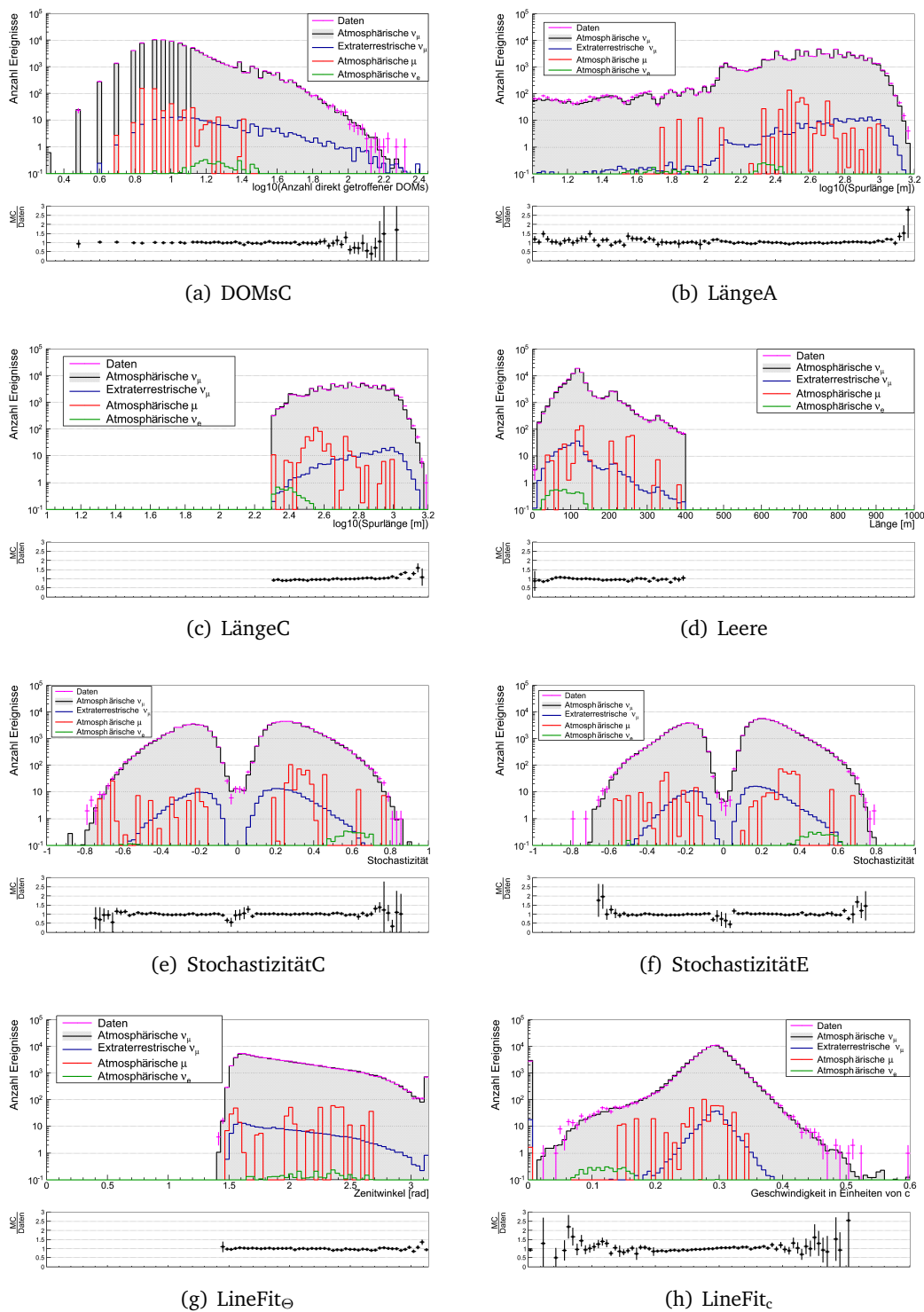


Abbildung A.6.: Verwendete Attribute auf Level 5

### A.3. Verwendete Attribute auf Level 5

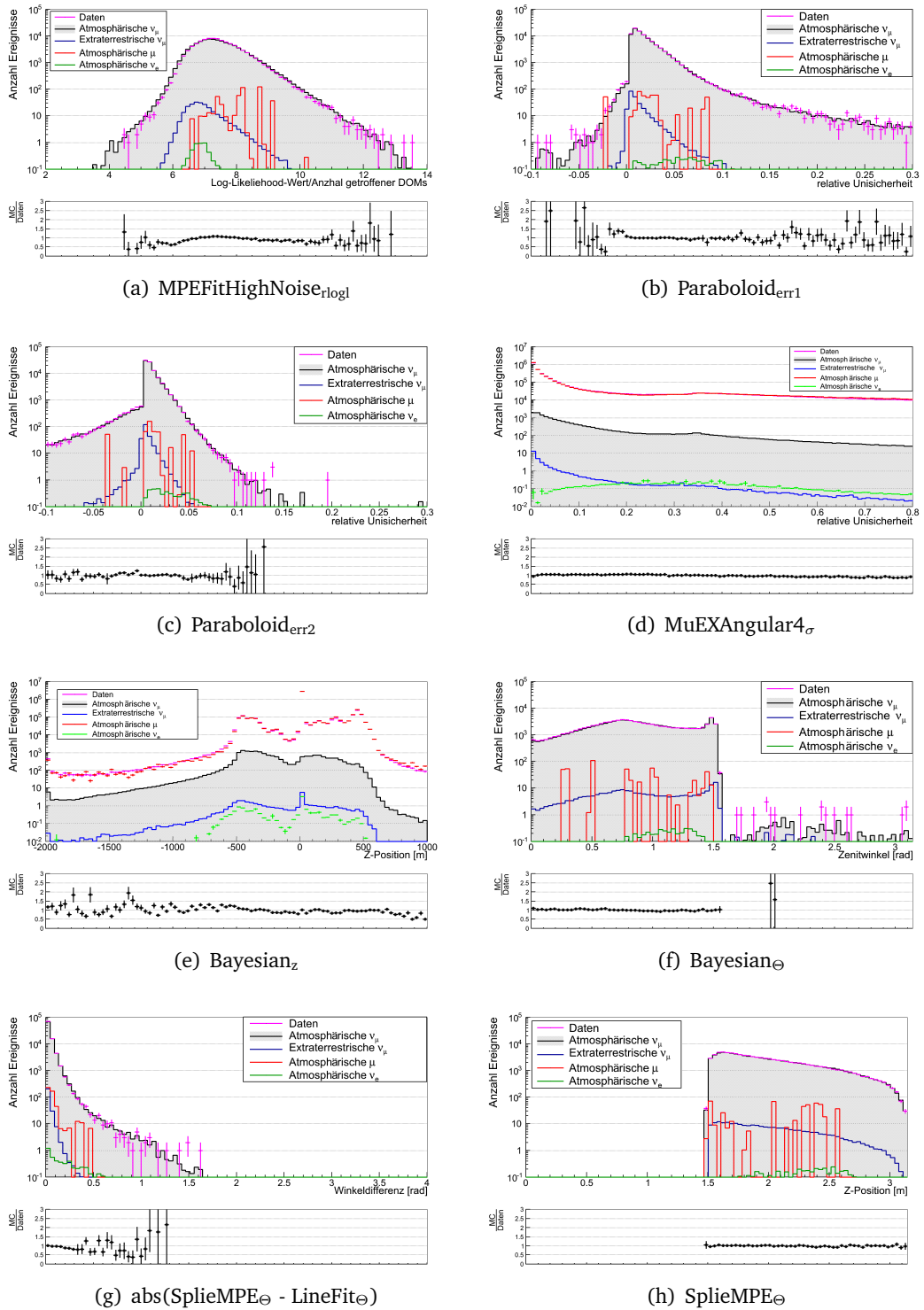
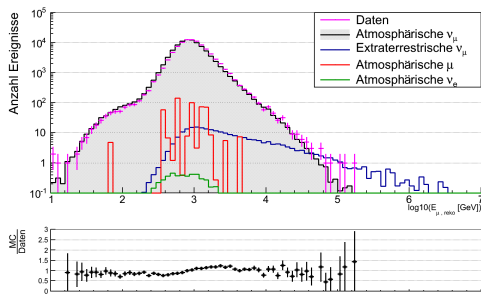
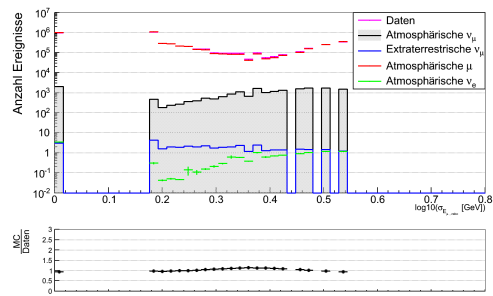


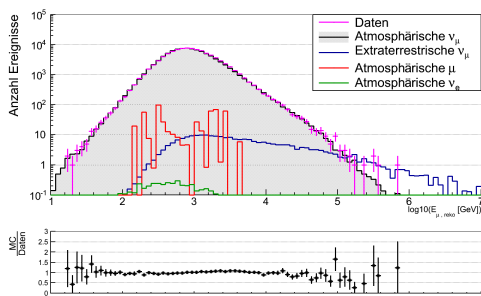
Abbildung A.7.: Verwendete Attribute auf Level 5



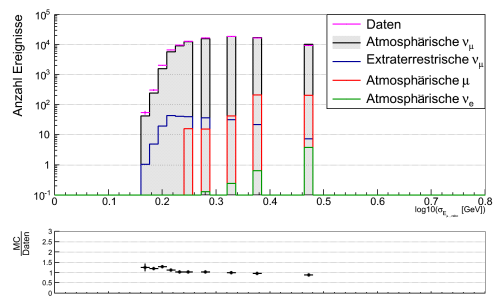
(a) MuEX<sub>E</sub>



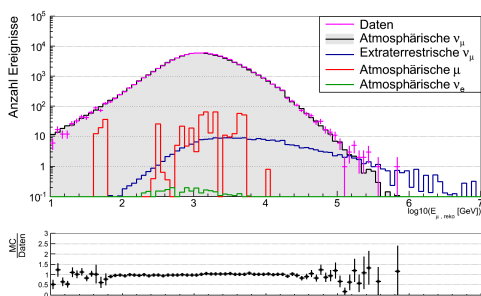
(b) TEDOMS<sub>σ</sub>



(c) TEDOMS<sub>E</sub>



(d) TEBINS<sub>σ</sub>



(e) TEBINS<sub>E</sub>

Abbildung A.8.: Verwendete Attribute auf Level 5



## A.4. Validierung der Entfaltung

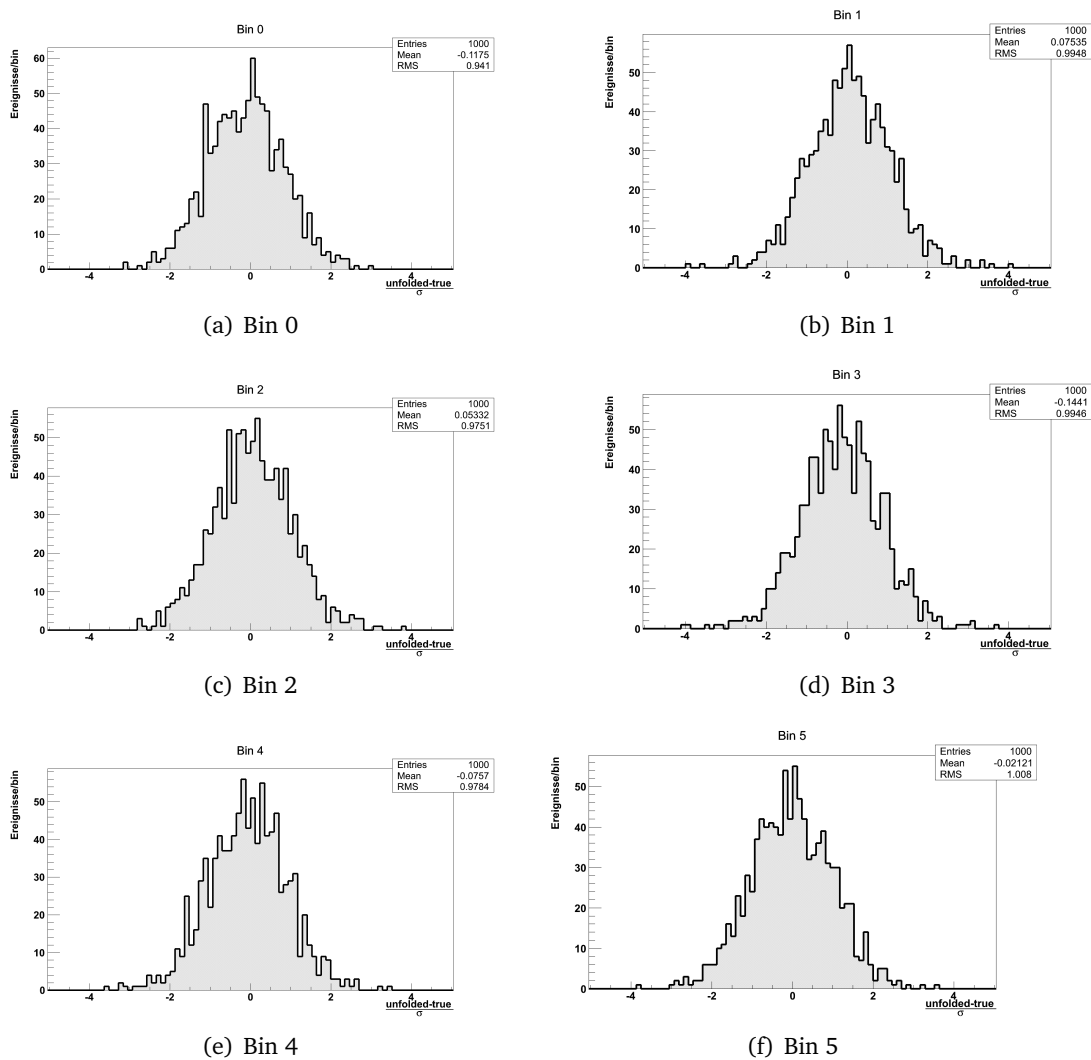
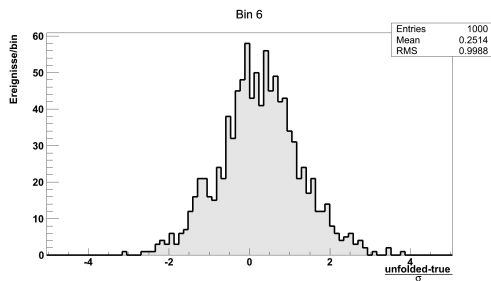
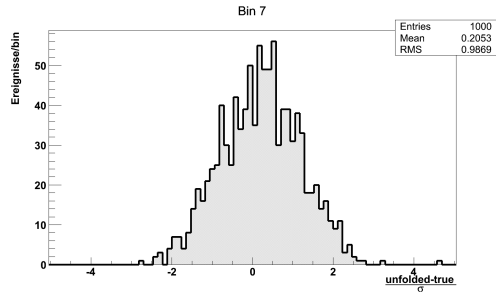


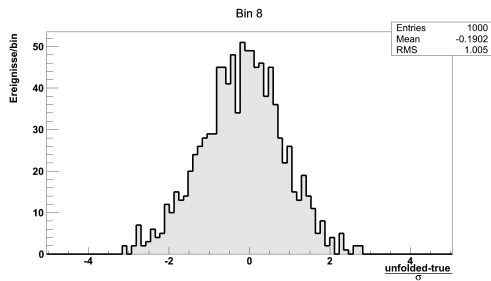
Abbildung A.9.: Pullmode - Bins 0 bis 5



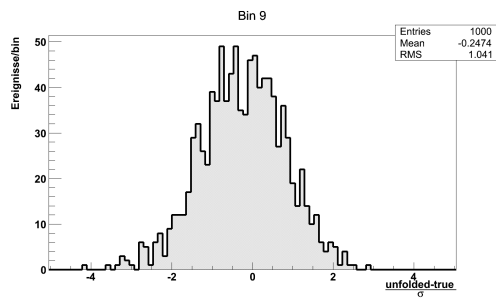
(a) Bin 6



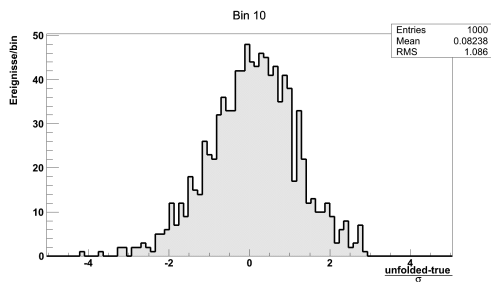
(b) Bin 7



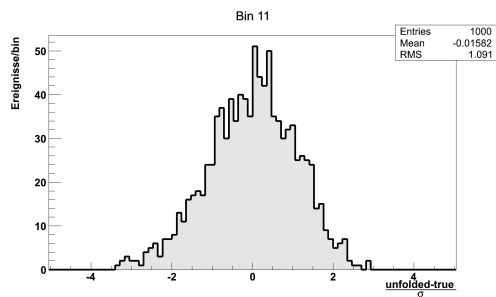
(c) Bin 8



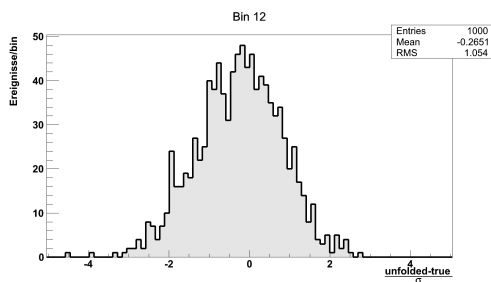
(d) Bin 9



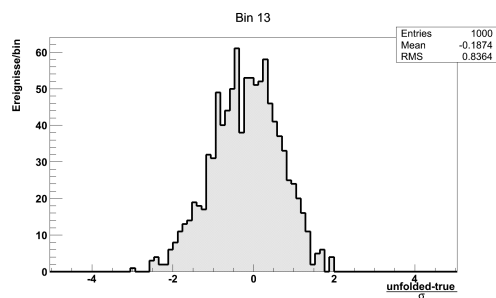
(e) Bin 10



(f) Bin 11



(g) Bin 12



(h) Bin 13

Abbildung A.10.: Pullmode - Bins 6 bis 13

# Literaturverzeichnis

- [A<sup>+</sup>04] J. Ahrens et al. Muon track reconstruction and data selection techniques in AMANDA. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 524(1):169–194, 2004.
- [A<sup>+</sup>06] M. Ackermann et al. Optical properties of deep glacial ice at the South Pole. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 111(D13), 2006.
- [A<sup>+</sup>09] R. Abbasi et al. The IceCube data acquisition system: Signal capture, digitization, and timestamping. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 601(3):294–316, 2009.
- [A<sup>+</sup>10] R. Abbasi et al. Calibration and characterization of the IceCube photomultiplier tube. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 618(1):139–152, 2010.
- [A<sup>+</sup>11a] R. Abbasi et al. Measurement of the atmospheric neutrino energy spectrum from 100 GeV to 400 TeV with IceCube. *Physical Review D*, 83(1):012001, 2011.
- [A<sup>+</sup>11b] R. Abbasi et al. Search for a diffuse flux of astrophysical muon neutrinos with the IceCube 40-string detector. *Physical Review D*, 84(8):082001, 2011.
- [A<sup>+</sup>11c] O. Adriani et al. PAMELA measurements of cosmic-ray proton and helium spectra. *arXiv preprint arXiv:1103.4055*, 2011.
- [A<sup>+</sup>12] R. Abbasi et al. The design and performance of IceCube DeepCore. *Astroparticle physics*, 35(10):615–624, 2012.
- [A<sup>+</sup>13a] M. Aartsen et al. Measurement of the cosmic ray energy spectrum with IceTop-73. *Physical Review D*, 88(4):042004, 2013.
- [A<sup>+</sup>13b] M.G. Aartsen et al. Evidence for high-energy extraterrestrial neutrinos at the IceCube detector. *Science*, 342(6161):1242856, 2013.

- [A<sup>+</sup>13c] M.G. Aartsen et al. Search for a diffuse flux of astrophysical muon neutrinos with the IceCube 59-string configuration. *arXiv preprint arXiv:1311.7048*, 2013.
- [A<sup>+</sup>13d] M.G. Aartsen et al. The IceCube Neutrino Observatory Part I: Point Source Searches. 2013.
- [A<sup>+</sup>13e] R. Abbasi et al. An improved method for measuring muon energy using the truncated mean of  $dE/dx$ . *Nucl.Instrum.Meth.*, A703:190–198, 2013.
- [A<sup>+</sup>13f] R. Abbasi et al. IceTop: The surface component of IceCube. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 700:188–220, 2013.
- [A<sup>+</sup>14a] M.G. Aartsen et al. Energy Reconstruction Methods in the IceCube Neutrino Telescope. *JINST*, 9:P03009, 2014.
- [A<sup>+</sup>14b] M.G. Aartsen et al. Improvement in Fast Particle Track Reconstruction with Robust Statistics. *Nucl.Instrum.Meth.*, A736:143–149, 2014.
- [A<sup>+</sup>14c] M.G. Aartsen et al. Observation of High-Energy Astrophysical Neutrinos in Three Years of IceCube Data. *arXiv preprint arXiv:1405.5303*, 2014.
- [A<sup>+</sup>14d] M.G. Aartsen et al. Observation of the cosmic-ray shadow of the Moon with IceCube. *Phys.Rev.*, D89:102004, 2014.
- [A<sup>+</sup>14e] M.G. Aartsen et al. Search for a diffuse flux of astrophysical muon neutrinos with the IceCube 59-string configuration. *Physical Review D*, 89(6):062007, 2014.
- [ABKW89] W. Arnett, J. Bahcall, R. Kirshner, and S. Woosley. Supernova 1987A. *Annual review of Astronomy and Astrophysics*, 27:629–700, 1989.
- [AEF<sup>+</sup>] E. AHN, R. ENGEL, A. FEDYNITCH, T. GAISSER, F. RIEHN, and T. STANEV. Atmospheric neutrinos at high energy. *atmosphere*, 3:8.
- [AEG<sup>+</sup>09] E.-J. Ahn, R. Engel, T. Gaisser, P. Lipari, and T. Stanev. Cosmic ray interaction event generator SIBYLL 2.1. *Physical Review D*, 80(9):094003, 2009.
- [AJY00] H. Athar, M. Jeżabek, and O. Yasuda. Effects of neutrino mixing on high-energy cosmic neutrino flux. *Physical Review D*, 62(10):103007, 2000.
- [ALS02] I. Albuquerque, J. Lamoureux, and G. Smoot. Astrophysical neutrino event rates and sensitivity for neutrino telescopes. *Astrophys.J.Suppl.*, 141:195–209, 2002.
- [AM<sup>+</sup>13] S. Adrián-Martínez et al. Measurement of the atmospheric  $\nu_\mu$  energy spectrum from 100 GeV to 200 TeV with the ANTARES telescope. *EUROPEAN PHYSICAL JOURNAL C*, 73(10), 2013.

- [A<sup>+</sup>ep] M.G. Aartsen et al. Evidence for Astrophysical Muon Neutrinos from the Northern Sky. *Physical Review Letters*, in prep.
- [BCM<sup>+</sup>14] J. Beringer, W. Collins, J. Martin, et al. The Review of Particle Physics. *Phys. Rev.*, D86, 2014.
- [BE87] R. Blandford and D. Eichler. Particle acceleration at astrophysical shocks: A theory of cosmic ray origin. *Physics Reports*, 154(1):1–75, 1987.
- [Bec07] J. Becker. *Neutrinos on the rocks: On the phenomenology of potential astrophysical neutrino sources*. PhD thesis, TU Dortmund, 2007.
- [Ber09] P. Berghaus. Muons in IceCube. *Nuclear Physics B-Proceedings Supplements*, 196:261–266, 2009.
- [BFO<sup>+</sup>83] L. Breiman, J. Friedman, R. Olshen, C. Stone, D. Steinberg, and P. Colla. CART: Classification and regression trees. *Wadsworth: Belmont, CA*, 156, 1983.
- [Bis07] C. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2007.
- [BL98] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Springer, 1998.
- [Blo96] V. Blobel. The run manual-regularized unfolding for high-energy physics experiments. *Technical Note TN*, 361, 1996.
- [BMM08] I.N. Bronštejn, G. Musiol, and H. Mühlig. *Taschenbuch der Mathematik*. Deutsch, 2008.
- [Bre01] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
- [BZM07] A. Bosch, A. Zisserman, and X. Munoz. Image Classification using Random Forests and Ferns. In *ICCV*, pages 1–8. IEEE, 2007.
- [CDDJ<sup>+</sup>98] B. Cleveland, T. Daily, R. Davis Jr, J. Distel, K. Lande, C. Lee, P. Wildenhain, and J. Ullman. Measurement of the solar electron neutrino flux with the Homestake chlorine detector. *The Astrophysical Journal*, 496(1):505, 1998.
- [Čer37] P. Čerenkov. Visible radiation produced by electrons moving in a medium with velocities exceeding that of light. *Physical Review*, 52(4):378, 1937.
- [Chi13] D. Chirkin. Photon tracking with GPUs in IceCube. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 725:141–143, 2013.

- [Cle13] F. Clevermann. *Neutronoenergieentfaltung von Punktquellkatalogen mit IceCube - Lösung inverser Probleme bei kleinen Statistiken*. PhD thesis, Technische Universität Dortmund, 2013.
- [Cow98] G. Cowan. *Statistical data analysis*. Oxford University Press, 1998.
- [CRH<sup>+</sup>56] C. Cowan, F. Reines, F. Harrison, H. Kruse, and A. McGuire. Detection of the free neutrino: A confirmation. *Science*, 125:103, 1956.
- [DP05] C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205, 2005.
- [DRB<sup>+</sup>95] K. Daum, W. Rhode, P. Bareyre, R. Barloutaud, G. Chardin, B. Degrange, J. Ernwein, B. Kuznik, H. Meyer, L. Mosca, et al. Determination of the atmospheric neutrino spectra with the Frejus detector. *Zeitschrift für Physik C Particles and Fields*, 66(3):417–428, 1995.
- [Efr79] B. Efron. Bootstrap methods: another look at the jackknife. *The annals of Statistics*, pages 1–26, 1979.
- [ERS08] R. Enberg, M. Reno, and I. Sarcevic. Prompt neutrino fluxes from atmospheric charm. *Physical Review D*, 78(4):043005, 2008.
- [F<sup>+</sup>98] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Physical Review Letters*, 81(8):1562, 1998.
- [FS97] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting, 1997.
- [FTD12] A. Fedynitch, J. Tjus, and P. Desiati. Influence of hadronic interaction models and the cosmic ray spectrum on the high energy atmospheric muon and neutrino flux. *Physical Review D*, 86(11):114024, 2012.
- [Gai90] T.K. Gaisser. *Cosmic Rays and Particle Physics*. Cambridge University Press, 1990.
- [Gai06] T. Gaisser. The Cosmic-ray Spectrum: from the knee to the ankle. *Journal of Physics: Conference Series*, 47(1):15, 2006.
- [Gai10] T. Gaisser. Cosmic-ray physics with IceCube. In *38th COSPAR Scientific Assembly*, volume 38, page 2674, 2010.
- [Gin12] C. Gini. *Variabilità e mutabilità. Contributi allo studio delle relazioni e delle distribuzioni statistiche*. 1912.
- [GK05] A. Gazizov and M. Kowalski. ANIS: High energy neutrino generator for neutrino telescopes. *Computer Physics Communications*, 172(3):203–213, 2005.

- [Gri11a] P. Grieder. *Extensive Air Showers: High Energy Phenomena and Astrophysical Aspects - A Tutorial, Reference Manual and Data Book Volumen I (Astrophysics and Space Science Library)*. Springer, 2011.
- [Gri11b] P. Grieder. *Extensive Air Showers: High Energy Phenomena and Astrophysical Aspects - A Tutorial, Reference Manual and Data Book Volumen II (Astrophysics and Space Science Library)*. Springer, 2011.
- [Gru01] C. Grupen. *Astroteilchenphysik: Das Universum im Licht der kosmischen Strahlung (German Edition)*. Springer, 2001.
- [GST13] T. Gaisser, T. Stanev, and S. Tilav. Cosmic ray energy spectrum from measurements of air showers. *Frontiers of Physics*, 8(6):748–758, 2013.
- [Hay12] W. Haynes. *CRC handbook of chemistry and physics*. CRC press, 2012.
- [Hes08] J. Hester. The crab nebula: an astrophysical chimera. *Annu. Rev. Astron. Astrophys.*, 46:127–155, 2008.
- [HK10] F. Halzen and S. Klein. Invited review article: IceCube: An instrument for neutrino astronomy. *Review of Scientific Instruments*, 81(8):081101, 2010.
- [HKC<sup>+</sup>98] D. Heck, J. Knapp, J. Capdevielle, G. Schatz, T. Thouw, et al. *CORSIKA: A Monte Carlo code to simulate extensive air showers*, volume 6019. FZKA, 1998.
- [HKK<sup>+</sup>87] K. Hirata, T. Kajita, M. Koshiba, M. Nakahata, and Y. Oyama. Observation of a neutrino burst from the supernova SN1987A. *Physical Review Letters*, 58:1490–1493, 1987.
- [HKK<sup>+</sup>07] M. Honda, T. Kajita, K. Kasahara, S. Midorikawa, and T. Sanuki. Calculation of atmospheric neutrino flux using the interaction model calibrated with atmospheric muon data. *Phys. Rev. D*, 75:043006, Feb 2007.
- [HTF01] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.
- [Jac12] P. Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.
- [Jam12] H. James. Written Testimony of Mr. James H. Yeck, IceCube Project Director before the UNITED STATES HOUSE OF REPRESENTATIVES, 2012.
- [JDM00] A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37, 2000.

- [K<sup>+</sup>95] R. Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, volume 14, pages 1137–1145, 1995.
- [K<sup>+</sup>14] A. Karle et al. IceCube-status and recent results. *arXiv preprint arXiv:1401.4496*, 2014.
- [KFS<sup>+</sup>13] J.-H. Koehne, K. Frantzen, M. Schmitz, T. Fuchsz, W. Rhode, D. Chirkin, and J. Tjus. PROPOSAL: A tool for propagation of charged leptons. *Computer Physics Communications*, 184(9):2070–2090, 2013.
- [KM98] M. Kearns and Y. Mansour. A Fast, Bottom-Up Decision Tree Pruning Algorithm with Near-Optimal Generalization. In *ICML*, volume 98, pages 269–277, 1998.
- [KS12] U. Katz and C. Spiering. High-energy neutrino astrophysics: Status and perspectives. *Progress in Particle and Nuclear Physics*, 67(3):651–704, 2012.
- [LM68] P. Lachenbruch and M. Mickey. Estimation of error rates in discriminant analysis. *Technometrics*, 10(1):1–11, 1968.
- [LMOP47] C. Lattes, H. Muirhead, G. Occhialini, and C. Powell. Processes involving charged mesons. *Nature*, 159(4047):694–697, 1947.
- [LMW<sup>+</sup>07] J. Lundberg, P. Miočinović, K. Woschnagg, T. Burgess, J. Adams, S. Hundertmark, P. Desiati, and P. Niessen. Light tracking through ice and water — Scattering and absorption in heterogeneous media with PHOTONICS. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 581(3):619–631, 2007.
- [Lot12] M. Lotze. Precut Improvements using Multivariate Methods on IceCube-59 Monte Carlo Simulations, 2012.
- [LSG<sup>+</sup>97] B. Louis, V. Sandberg, G. Garvey, H. White, G. Mills, et al. The evidence for oscillations. *Los Alamos Sci.*, 25:116–127, 1997.
- [Mar60] MA Markov. On high energy neutrino physics. In *Proceedings of the 1960 Annual International Conference on HEP at Rochester*, pages 578–581, 1960.
- [MDK<sup>+</sup>13] N. Milke, M. Doert, S. Klepser, D. Mazin, V. Blobel, and W. Rhode. Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 697:133–147, 2013.



- [Mil12] N. Milke. *Unfolding of the atmospheric neutrino flux spectrum with the new program TRUEE and IceCube*. PhD thesis, Technische Universität Dortmund, 2012.
- [MKM<sup>+</sup>09] B. Menze, B. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. Hamprecht. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10(1):213, 2009.
- [Moo20] E. Moore. On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, 26:394–395, 1920.
- [MRMH12] M. Meyer, M. Raue, D. Mazin, and D. Horns. Limits on the extragalactic background light in the Fermi era. *Astronomy & Astrophysics/Astronomie et Astrophysique*, 542(1), 2012.
- [Mün07] K. Münich. Messung des atmosphärischen Neutrinospektrums mit dem AMANDA-II-Detektor, 2007.
- [MWK<sup>+</sup>06] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler. Yale: Rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 935–940. ACM, 2006.
- [NCSS] NASA, CXC, SAO, and F. Seward. <http://chandra.harvard.edu/photo/2008/crab/>.
- [NEHL] NASA, ESA, J Hester, and A. Loll. <http://hubblesite.org/newscenter/archive/releases/2005/37/image/a/>.
- [NMB93] L. Nellen, K. Mannheim, and P. Biermann. Neutrino production through hadronic cascades in AGN accretion disks. *Physical Review D*, 47(12):5270, 1993.
- [OH95] J. Oliver and J. Hand. On Pruning and Averaging Decision Trees. In *ICML*, pages 430–437, 1995.
- [Ost06] S. Ostapchenko. QGSJET-II: towards reliable description of very high energy hadronic interactions. *Nuclear Physics B-Proceedings Supplements*, 151(1):143–146, 2006.
- [Pro97] R. Protheroe. High Energy Neutrinos from Blazars. In D. T. Wickramasinghe, G. V. Bicknell, & L. Ferrario, editor, *IAU Colloq. 163: Accretion Phenomena and Related Outflows*, volume 121 of *Astronomical Society of the Pacific Conference Series*, pages 585–+, 1997.
- [Qui86] J. Quinlan. Induction of Decision Trees. *Mach. Learn.*, 1(1):81–106, March 1986.

- [Res09] E. Resconi. Status and prospects of the IceCube neutrino telescope. *Nucl.Instrum.Meth.*, A602:7–13, 2009.
- [RHH68] D. Raymond, D. Harmer, and K. Hoffman. Search for neutrinos from the sun. *Physical Review Letters*, 20(21):1205, 1968.
- [RM<sup>+</sup>12] T. Ruhe, K. Morik, et al. Data Mining IceCube, 2012.
- [Ruh13] T. Ruhe. *Data-Mining on the Rocks: A measurement of the atmospheric muon neutrino flux using IceCube in the 59-string configuration and a novel data mining based approach to unfolding*. PhD thesis, Technische Universität Dortmund, 2013.
- [Sch14] M. Schmitz. *Neuartige Boten aus dem All - Messung des extraterrestrischen und atmosphärischen Myon-Neutrino-Energiespektrums mit IceCube-79*. PhD thesis, TU Dortmund, 2014.
- [SM11] B. Schowe and K. Morik. Fast-ensembles of minimum redundancy feature selection. In *Ensembles in Machine Learning Applications*, pages 75–95. Springer, 2011.
- [Sti89] S. Stigler. Francis Galton’s account of the invention of correlation. *Statistical Science*, pages 73–79, 1989.
- [T<sup>+</sup>54] R. Tate et al. Correlation between a discrete and a continuous variable. Point-biserial correlation. *The Annals of mathematical statistics*, 25(3):603–607, 1954.
- [The11] The IceCube Collaboration. <http://icecube.wisc.edu/gallery>, 2011.
- [TIG96] M. Thunman, G. Ingelman, and P. Gondolo. Charm production and high energy atmospheric muon and neutrino fluxes. *Astroparticle Physics*, 5(3):309–332, 1996.
- [Tik63] A. Tikhonof. On the solution of ill-posed problems and the method of regularization. In *Dokl Akad Nauk SSSR*, volume 3, pages 501–504, 1963.
- [Tos12] S. Toscano. Observation of anisotropy in the arrival direction distribution of cosmic rays above TeV energies with IceCube. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 692:165–169, 2012.
- [Tre00] M. Treichel. *Teilchenphysik und Kosmologie: Eine Einführung in Grundlagen und Zusammenhänge (German Edition)*. Springer, 2000.
- [VV73] L. Van Valen. A new evolutionary law. *Evolutionary theory*, 1:1–30, 1973.
- [WEH11] I. Witten, F Eibe, and M. Hall. *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann, 2011.

- [Wik14] Wikimedia Commons. Eingestellt von: Torres997. Lizenz des Bildes: CC BY-SA 3.0. Crab Nebula in multiwavelength. [https://en.wikipedia.org/wiki/File:Crab\\_Nebula\\_in\\_multiwavelength.png](https://en.wikipedia.org/wiki/File:Crab_Nebula_in_multiwavelength.png), 2014.
- [Wil38] S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

# Abbildungsverzeichnis

2.1	Der Krebsnebel in verschiedenen Wellenlängen . . . . .	5
2.2	Vergrößerte Darstellung des Krebsnebels in verschiedenen Wellenlängen	6
2.3	Spektrum der geladenen kosmischen Strahlung . . . . .	7
2.4	Hadronische Kaskaden in der Atmosphäre . . . . .	8
2.5	Produktionsverhältnisse für Kaonen und Pionen . . . . .	10
2.6	Unterschiedliche Primärmodelle der kosmischen Strahlung . . . . .	11
2.7	Die prompte Komponente der atmosphärischen Leptonen . . . . .	12
2.8	Überlebenswahrscheinlichkeit für Neutrinos beim Durchqueren der Erde	14
3.1	Schematische Darstellung des IceCube-Detektors . . . . .	18
3.2	Ein Ereignis im IceCube-Detektor . . . . .	19
3.3	Parameter zur Beschreibung eines Myons mittels Cherenkov Licht . . . .	20
3.4	Energieverluste von Myonen . . . . .	22
3.5	Energieverlust von Myonen pro Meter . . . . .	23
4.1	Der Jaccard-Index in Abhängigkeit von der Anzahl der gewählten Attribute . . . . .	31
4.2	Schematische Darstellung der Kreuzvalidierung . . . . .	33
4.3	Wahrscheinlichkeitsverteilung für ein Ereignis . . . . .	35
4.4	Verteilung der Confidence nach Bootstrapping . . . . .	36
4.5	Schematische Darstellung eines Entscheidungsbaums . . . . .	38
5.1	Schematische Darstellung der Attribute Länge und Leere. . . . .	47
5.2	Das Attribut SplineMPE $_{\Theta}$ . . . . .	49
5.3	Energieabhängige Effizienz des Schnittes in SplineMPE $_{\Theta}$ . . . . .	50
5.4	Das Attribut LängeC . . . . .	51
5.5	Energieabhängige Effizienz des Schnittes in LängeC . . . . .	52
5.6	Das Attribut TEBINS $_{\sigma}$ . . . . .	53
5.7	Energieabhängige Effizienz des Schnittes in TEBINS $_{\sigma}$ . . . . .	54
5.8	Radius $r_{\max}$ des TEBINS $_E$ und TEDOMS $_E$ Algorithmus. . . . .	55
5.9	Das Attribut Leere als Vorselektionsschnitt Parameter. . . . .	56
5.10	Effizienz aller Vorselektionsschnitte . . . . .	57
5.11	Stabilität und Effizienz der Attributselektion . . . . .	59
5.12	Random Forest Confidence Verteilung . . . . .	60
5.13	Confidence Bereich $> 0,7$ . . . . .	62

5.14	Zweidimensionaler Phasenraum für Daten . . . . .	64
5.15	Verteilung von MC im Phasenraum, aufgespannt von Confidence und Energierestruktion . . . . .	66
5.16	Confidence-Bereich $> 0,7$ nach dem finalen Schnitt . . . . .	67
5.17	Effizienz der Ereignisselektion . . . . .	69
5.18	Effektive Flächen . . . . .	70
5.19	MuEX <sub>E</sub> -Verteilung für verschiedene Winkelbereiche . . . . .	71
5.20	r <sub>COG</sub> -Verteilung für verschiedene Schnitte in MuEX <sub>E</sub> . . . . .	73
6.1	LängeC als Entfaltungsobservable . . . . .	78
6.2	DOMsC als Entfaltungsobservable . . . . .	79
6.3	MuEX <sub>E</sub> als Entfaltungsobservable . . . . .	80
6.4	Entfaltetes Energiespektrum ohne systematische Fehler und Akzeptanz- korrektur . . . . .	81
6.5	Entfaltetes Energiespektrum mit und ohne systematische Fehler . . . . .	82
6.6	Testmode Entfaltung von TRUEE . . . . .	83
6.7	Beispielhaftes Pullmodeergebnis . . . . .	84
6.8	Fehler der Entfaltung . . . . .	85
6.9	Flüsse in verschiedenen Winkelbereichen . . . . .	88
6.10	Entfaltete Energiespektren in drei Zenitbereichen ohne systematische Fehler . . . . .	91
6.11	Entfaltete Energiespektren in drei Zenitbereichen mit systematischen Fehlern . . . . .	91
6.12	Umgewichtung des Spektrums im vollen Winkelbereich auf die Spek- tren in verschiedenen Winkelbereichen . . . . .	92
6.13	Umgewichtung des Spektrums im vollen Winkelbereich auf die Spek- tren in vier Winkelbereichen . . . . .	94
7.1	Vergleich mit der IC79 Entfaltung . . . . .	98
7.2	Vergleich mit der IC59 Entfaltung . . . . .	99
7.3	Vergleich mit der AMANDA, ANTARES und Fréjus Entfaltung . . . . .	100
7.4	Vergleich mit der HESE Analyse . . . . .	101
7.5	Vergleich mit den AMANDA, ANTARES, Fréjus, HESE, IC59 und IC79 Entfaltungen . . . . .	102
7.6	Mittelung von IC79 und IC86 . . . . .	103
7.7	Vergleich mit atmosphärischer Vorhersage . . . . .	105
7.8	Mittlerer Winkel in den Entfaltungsbins . . . . .	106
7.9	Vergleich mit einer, um einen Faktor 20 gesteigerten prompten Kompo- nente . . . . .	107
7.10	Vergleich mit atmosphärischem und hochenergetischem extraterrestri- schem Fluss . . . . .	108
7.11	Vergleich zwischen den hadronischen Interaktionsmodellen QGSJet und SIBYLL . . . . .	109
7.12	Bestes Fitergebnis im Vergleich zum gemittelten Spektrum . . . . .	112

7.13	Eindimensionale Projektion der Likelihood-Funktion für die Normierung des Flusses der konventionellen Neutrinos . . . . .	113
7.14	Eindimensionale Projektion der Likelihood-Funktion für die Normierung der prompten Neutrinos . . . . .	114
7.15	Eindimensionale Projektion der Likelihood-Funktion für die Normierung der hochenergetischen extraterrestrischen Komponente . . . . .	115
7.16	Eindimensionale Projektion der Likelihood-Funktion für den spektralen Index der hochenergetischen extraterrestrischen Komponente . . . . .	116
7.17	Zweidimensionale Projektion der Likelihood-Funktion für den spektralen Index und die prompte Komponente . . . . .	117
7.18	Zweidimensionale Projektion der Likelihood-Funktion für die Normierung des hochenergetischen extraterrestrischen Flusses und die prompte Komponente . . . . .	117
7.19	Zweidimensionale Projektion der Likelihood-Funktion für die Normierung des hochenergetischen extraterrestrischen Flusses und des spektralen Index . . . . .	118
7.20	Signifikanz des beobachteten Exzesses in Abhängigkeit der maximal erlaubten Normierung der prompten Komponente in $H_0$ . . . . .	120
7.21	Eindimensionale Projektion der Likelihood-Funktion für die Normierung der prompten Neutrinos ohne extraterrestrische Neutrinos . . . . .	123
7.22	Entfaltete Energiespektren in drei Zenitbereichen im Vergleich zur Vorhersage . . . . .	124
7.23	Eindimensionale Projektion der Likelihood-Funktion für die Normierung der prompten Neutrinos ohne extraterrestrische Neutrinos . . . . .	125
7.24	Zweidimensionale Projektion der Likelihood-Funktion für die Normierung des hochenergetischen extraterrestrischen Flusses und des spektralen Index unter der Bedingung $N_{\text{prompt}} = 1,7$ . . . . .	126
7.25	Bestes Fit Ergebnis im Vergleich zum gemittelten Spektrum auf Basis einer festen prompten Komponente . . . . .	128
A.1	Verwendete Attribute auf Level 3, Teil 1 . . . . .	138
A.2	Verwendete Attribute auf Level 3, Teil 2 . . . . .	139
A.3	Verwendete Attribute auf Level 3, Teil 3 . . . . .	140
A.4	Verwendete Attribute auf Level 3, Teil 4 . . . . .	141
A.5	Verwendete Attribute auf Level 5, Teil 1 . . . . .	143
A.6	Verwendete Attribute auf Level 5, Teil 2 . . . . .	144
A.7	Verwendete Attribute auf Level 5, Teil 3 . . . . .	145
A.8	Verwendete Attribute auf Level 5, Teil 4 . . . . .	146
A.9	Pullmode - Bins 0 bis 5 . . . . .	147
A.10	Pullmode - Bins 6 bis 13 . . . . .	148

# Tabellenverzeichnis

4.1	Übersicht über verschiedene Attributtypen . . . . .	26
4.2	Die Konfusionsmatrix . . . . .	27
5.1	Zeitfenster in IceCube . . . . .	48
5.2	Tree Importance . . . . .	63
5.3	Random Forest Parameter . . . . .	63
5.4	Kenngrößen des erzeugten Datensatzes . . . . .	65
6.1	Entfaltungsergebnisse . . . . .	76
6.2	Einstellungen der Entfaltung . . . . .	77
6.3	Verwendete systematische Datensätze . . . . .	84
6.4	Unsicherheiten der Entfaltungsergebnisse . . . . .	86
6.5	Definition der Zenitbereiche für die Entfaltung der Winkelbereiche . . . . .	87
6.6	Einstellungen der Entfaltungen für die drei Zenitbereiche . . . . .	89
6.7	Entfaltungsergebnisse für den Winkelbereich $86^\circ \leq \theta \leq 100^\circ$ . . . . .	89
6.8	Entfaltungsergebnisse für den Winkelbereich $100^\circ \leq \theta \leq 120^\circ$ . . . . .	90
6.9	Entfaltungsergebnisse für den Winkelbereich $120^\circ \leq \theta \leq 180^\circ$ . . . . .	90
6.10	Umgewichteter Fluss in verschiedenen Zenitbereichen . . . . .	95
7.1	Gemittelte Entfaltung . . . . .	104
7.2	Globales Minimum der Likelihood-Funktion . . . . .	110





# Danksagung

Zum Abschluss dieser Arbeit ist es mir ein Anliegen, vielen Personen zu danken. Ohne Hilfe und Beistand dieser Personen bei verschiedensten kleinen und auch großen Problemen wäre die Anfertigung der Arbeit nicht möglich gewesen. Mir ist dabei jedoch wichtig klarzustellen, dass mir jede Hilfe gleich viel wert ist.

Beginnen möchte ich mit meiner Frau Andrea, die mir während der Anfertigung dieser Arbeit, im Laufe von Studium und Promotion jederzeit mit Hilfe und Verständnis beiseite stand.

Meiner Tochter Marie möchte ich danken, dass sie so viele schöne Momente in mein Leben gebracht hat und mir somit ermöglicht hat, die Arbeit mit einer positiven Einstellung verfassen zu können.

Mein Dank gilt ebenfalls Prof. Dr. Dr. Wolfgang Rhode für die Betreuung während meiner Promotion und den Möglichkeiten, die er mir an seinem Lehrstuhl geboten hat.

Für die Zweitkorrektur dieser Arbeit danke ich Prof. Dr. Metin Tolan.

Martin Schmitz danke ich für die kollegiale Zusammenarbeit am Lehrstuhl und der IceCube Kollaboration, sowie für unzählige Diskussionen über Data-Mining, Entfaltung und Astroteilchenphysik.

Für die Unterstützung in verschiedenster Form danke ich ebenfalls allen Mitarbeitern und Studierenden des Dortmunder Lehrstuhls für Astroteilchenphysik. Danke für viele Diskussionen, Anregungen und Ratschläge!!!

Daher an alle:

# DANKE!