



Lehrstuhl Experimentelle Physik V
Fakultät Physik
Technische Universität Dortmund



Neuartige Boten aus dem All

Messung des extraterrestrischen und atmosphärischen

Myon-Neutrino-Energiespektrums mit IceCube-79

Dissertation

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

vorgelegt von

Dipl.-Phys. Martin Schmitz
geboren in Marl

Dortmund, 20. August 2014

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

2. Gutachter: Prof. Dr. Carsten Westphal

Datum des Einreichens der Arbeit: 20. August 2014

Inhaltsverzeichnis

1. Einleitung	1
2. Astrophysikalische Grundlagen	3
2.1. Der IceCube-Detektor	3
2.2. Astrophysikalische Neutrinos	8
2.3. Primäre kosmische Strahlung	10
2.3.1. Energieverteilung	10
2.3.2. Greisen-Zatsepin-Kuzmin Cutoff	12
2.4. Luftschauber	14
2.4.1. Konventionelle atmosphärische Myonen und Neutrinos	14
2.4.2. Atmosphärische Elektron-Neutrinos	15
2.4.3. Prompte Neutrinos	17
3. Methoden der Datenanalyse	19
3.1. Simulation	19
3.2. Algorithmen zur Rekonstruktion von Ereignissen	21
3.2.1. Line-Fit	21
3.2.2. SPE- und MPE-Fit	22
3.2.3. Energierekonstruktion	24
3.3. Trennung von Signal und Untergrund	25
3.3.1. Begrifflichkeiten	25
3.3.2. Entscheidungsbäume	27
3.3.3. Random Forests	28
3.3.4. Merkmalselektion	29
3.3.5. Validierung	29
3.3.6. Stabilität der Merkmalselektion	33

3.4. Entfaltung von Verteilungen	34
3.4.1. Diskretisierung	35
3.4.2. Unregularisierte Entfaltung	36
3.4.3. Regularisierung	39
3.4.4. Modi in TRUEE	41
3.4.5. Ungewichten von Merkmalen	43
3.4.6. Bestimmung der x-Werte bei breiten Bins	44
4. Analyse der IC 79-Daten	45
4.1. Ereignis Selektion	45
4.1.1. Precuts	48
4.1.2. Merkmalauswahl	51
4.1.3. Selektion mit Hilfe des Random Forests	54
4.1.4. Beschreibung des verbleibenden Untergrunds	60
4.2. Entfaltung	71
4.2.1. Entfaltungsmerkmale und Einstellungen	71
4.2.2. Überprüfung auf Simulationsdaten	76
4.2.3. Modellunabhängigkeit der Entfaltung	77
4.2.4. Untersuchung systematischer Fehler	80
4.2.5. Anwendung auf Daten	83
5. Diskussion	87
5.1. Vergleich mit anderen Analysen	87
5.2. Vergleich zu verschiedenen theoretischen Modellen	91
5.3. Kombination der Ergebnisse der Analysen mit IC 79 und IC 86	95
5.4. Diskussion der astrophysikalischen und prompten Komponente	99
5.4.1. Erklärung durch eine hohe prompte Komponente	99
5.4.2. Vergleich mit der Analyse mit startenden Ereignissen	100
5.5. Bestimmung physikalischer Parameter	101
5.5.1. Methode	102
5.5.2. Ergebnisse	103
6. Zusammenfassung und Ausblick	107
A. Benutzte Ereignisse zur Separation von Ereignissen	117
B. Pullmodusverteilungen für die Entfaltung	121

C. Ergebnisse der Flussmessung	127
D. Ergebnisse der gemittelten Flussmessung	129

1. Einleitung

Seit jeher untersuchen die Menschen das Universum. Naturgemäß war lange Zeit das Photon die einzige Informationsquelle. Während zunächst nur das Auge als Messinstrument gedient hatte, wurden die technischen Messinstrumente im Laufe der Zeit immer besser und der beobachtete Frequenzbereich dehnte sich immer weiter aus. Heutzutage erstreckt sich der Bereich der untersuchten Strahlung vom kosmischen Mikrowellenhintergrund bis hin zu Gammastrahlung im PeV-Regime. Erst im 20. Jahrhundert wurde es möglich andere Teilchen als nur das Photon zur Erforschung des Universums zu nutzen. Die geladene kosmische Strahlung wurde entdeckt und mit verschiedenen Experimenten untersucht. Aufgrund ihrer Ladung werden die Teilchen durch Magnetfelder abgelenkt. Dadurch ist ein Rückschluss auf die Herkunft dieser Teilchen schwierig.

Mit dem Fund hochenergetischer extraterrestrischer Neutrinos im Jahre 2013 [A⁺13a] konnte ein neues Botenteilchen nachgewiesen werden, welches schon seit längerem als solches postuliert wurde [Gre60].

In dieser Arbeit wird das Spektrum von Myon-Neutrinos bestimmt, welche mit Ice-Cube [Hal09] detektiert wurden. Hierbei werden Methoden des maschinellen Lernens [JDM00] genutzt, um den Untergrund durch atmosphärische Neutrinos zu überwinden. Um das Energiespektrum der Myon-Neutrinos zu bestimmen wird die Methode der Entfaltung [SPM02] verwendet. Das Spektrum wird dabei über einen Energiebereich von ~ 100 GeV bis zu einer Energie oberhalb von 3 PeV bestimmt. Im Bereich oberhalb von 100 TeV kann eine wahrscheinlich nicht atmosphärische Komponente nachgewiesen werden. Dies belegt die Existenz dieser Botenteilchen auch im Myon-Kanal. Mit diesem Nachweis besagter Teilchen wird eine neue Ära der Astroteilchenphysik eingeleitet. Nun ist es möglich Astroteilchenphysik auch mit Neutrinos als Botenteilchen zu betreiben.

1. Einleitung

Diese Arbeit gliedert sich folgendermaßen:

Im zweiten Kapitel werden einige physikalische Grundlagen erläutert. Hierzu gehört IceCube, die primäre kosmische Strahlung und auch die Neutrinoastronomie. Ein besonderer Fokus wird auf die bestehenden Untergründe bei der Messung astrophysikalischer Neutrinos gelegt.

Im dritten Kapitel werden die verwendeten Techniken der Datenanalyse erklärt. Der Fokus liegt hier zum einen auf den Methoden des maschinellen Lernens, die notwendig sind um die erwähnten Untergründe zu beseitigen. Zum anderen wird auf die Methode der Entfaltung eingegangen, welche es ermöglicht, das Energiespektrum der Myon-Neutrinos zu bestimmen.

Das vierte Kapitel beschreibt die Analyse der Daten des IC 79-Detektors bis hin zur Bestimmung des Energiespektrums. Im fünften Kapitel werden die Eigenschaften dieses Spektrums untersucht. Hierzu wird das erzeugte Spektrum mit verschiedenen anderen Messungen und theoretischen Modellen verglichen. Zum Schluss folgt eine Zusammenfassung sowie ein Ausblick, welche die zuvor bestimmten Ergebnisse resümiert und zukünftige Verbesserungsmöglichkeiten aufzeigt.

2. Astrophysikalische Grundlagen

In dieser Arbeit werden Daten des Neutrino­detektors IceCube analysiert. Daher wird dieser Detektor im Folgenden zunächst vorgestellt. Im weiteren Verlauf dieses Kapitels wird der Zusammenhang zwischen der Neutrinoastronomie und anderen Bereichen der Astroteilchenphysik erläutert. Auch werden verschiedene Signalmuster und insbesondere der Einfluss der kosmischen Strahlung auf die Messung von Neutrinos mit IceCube diskutiert.

2.1. Der IceCube-Detektor

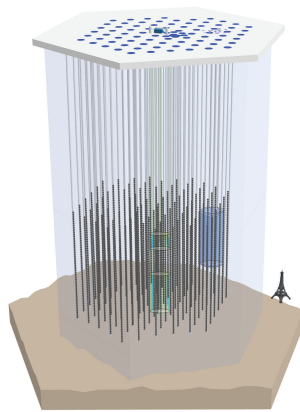


Abbildung 2.1.: Schematische Darstellung des IceCube-Detektors inklusive IceTop. Zum Größenvergleich ist der Eiffelturm abgebildet [KS12].

IceCube ist ein Neutrino­detektor am geographischen Südpol. Mit einem instrumentierten Volumen von einem Kubikkilometer ist er der zur Zeit größte Neutrino-

2. Astrophysikalische Grundlagen

detektor der Welt. IceCube wird häufig auch als Neutrinoobservatorium bezeichnet, da eines der wissenschaftlichen Ziele die Neutrinoastronomie ist [Bot05].

Neutrinos sind Teilchen des Standardmodells, die von Pauli 1930 postuliert wurden und schließlich von Cowan und Reines im Jahre 1956 experimentell nachgewiesen wurden [Gus10, CRH⁺56]. Neutrinos haben im Vergleich zu anderen Teilchen des Standardmodells mit einem Wechselwirkungsquerschnitt von

$$\sigma \sim 10^{-35} \text{ cm}^2 \quad (2.1.1)$$

bei einer Neutrinoenergie von 1 TeV eine geringe Wechselwirkungswahrscheinlichkeit mit Materie. Dies macht sie, zusammen mit der Eigenschaft, dass sie stabil gegen Zerfallsprozesse sind, zu idealen Botenteilchen für astrophysikalische Prozesse [F⁺98].

Durch ihre physikalischen Eigenschaften sind Neutrinos nicht direkt detektierbar. Um sie dennoch zu messen, ist eine indirekte Messung notwendig [KS12]. Hierbei gibt es zwei verschiedene Reaktionsmechanismen. Der dominante Mechanismus ist der Kanal über den geladenen Strom (*englisch: charged current*). In diesem Fall entsteht über

$$\nu_l + N \rightarrow l + X, \quad l = e, \mu, \tau \quad (2.1.2)$$

ein geladenes Lepton, welches im Detektor nachgewiesen werden kann. Dabei ist N ein Nukleus, l ein Lepton gleichen Flavours und X ein hadronischer Endzustand. Neben diesen Wechselwirkungen können Neutrinos auch über den ungeladenen Strom (*englisch: neutral current*) über

$$\nu_l + N \rightarrow \nu_l + X, \quad l = e, \mu, \tau \quad (2.1.3)$$

reagieren. Bei dieser Reaktion entsteht kein geladenes sekundäres Lepton.

Auf Grund des geringen Wechselwirkungsquerschnitts von Neutrinos mit Materie müssen möglichst große Detektoren gebaut werden. Bei der Detektion wird der Tscherenkoweffekt genutzt [KS12].

Geladene Leptonen, welche durch ein Medium propagieren, erzeugen währenddessen Sekundärteilchen. Die Leptonfamilien unterscheiden sich durch ihre Masse und ihre Lebensdauer, deshalb hinterlassen sie im Detektor unterschiedliche Signaturen. Elektronen haben im Vergleich zu Myonen und Tauonen auf Grund ihrer geringen Masse einen sehr großen Energieverlust pro Meter. Daher bilden sie sehr

2. Astrophysikalische Grundlagen

lokalisierte Ereignismuster im Detektor. Tauonen haben eine geringe Lebensdauer von $(290.6 \pm 1.0) \cdot 10^{-15}$ s [B⁺12]. Durch diese Eigenschaft und ihre verschiedenen Zerfallskanäle können Tauonen im Detektor unterschiedliche Signaturen hinterlassen. Detaillierte Informationen finden sich in [C⁺07]. Falls Neutrinos eines beliebigen Flavours über den neutralen Strom interagieren, so bilden sich ähnlich zu Elektron-Neutrinos lokalisierte Muster, die mit IceCube nicht von Interaktionen von Elektron-Neutrinos über den geladenen Strom unterschieden werden können. Diese lokalisierten Muster werden Kaskaden genannt. Da in dieser Arbeit der Fokus auf Myon-Neutrinos liegt, werden im Folgenden einige Eigenschaften für Myon-Wechselwirkungen diskutiert.

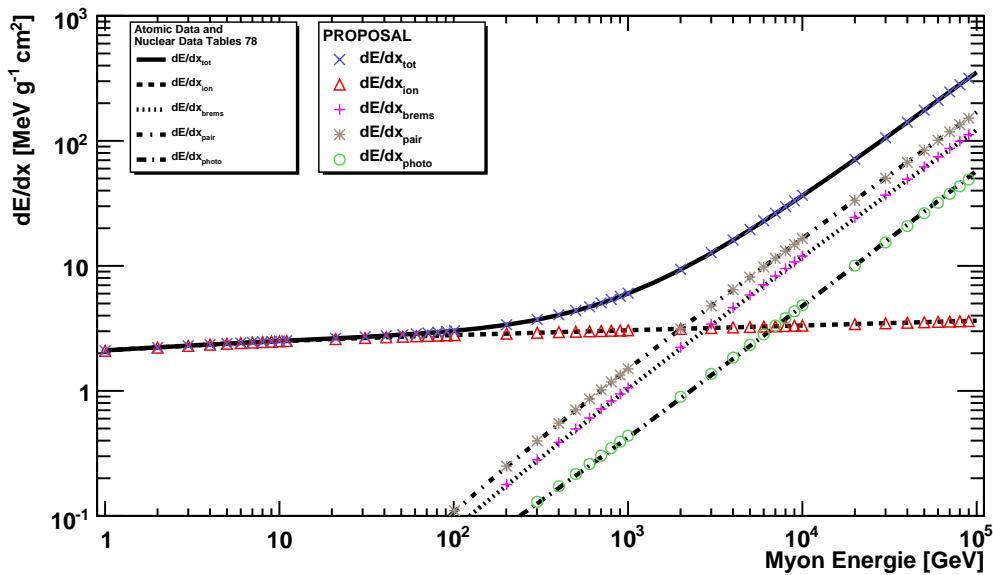


Abbildung 2.2.: Energieverlust von Myonen durch verschiedene Wechselwirkungsprozesse nach [GMS01], sowie nach Rechnung mit PROPOSAL [KFS⁺13]. Für Teilchen unterhalb von 1 TeV ist die Ionisation der dominante Energieverlust. Für Energien über 1 TeV sind Paarproduktion und Bremsstrahlung die dominanten Energieverlustmechanismen [KFS⁺13].

Myonen, welche durch ein Medium propagieren, unterliegen im wesentlichen vier Wechselwirkungen. Dies sind die Ionisation, die Paarproduktion von Elektron-Positron-Paaren, die Bremsstrahlung sowie der photonukleare Effekt. In unter-

2. Astrophysikalische Grundlagen

schiedlichen Energiebereichen sind unterschiedliche Effekte die jeweils dominanten Effekte für den Energieverlust.

Im Energiebereich unter 1 TeV ist der dominante Energieverlust für Myonen die Ionisation. Bei Energien über 1 TeV beginnen Paarproduktion und Bremsstrahlung zu dominieren. Eine Übersicht über den Energieverlust pro Strecke durch verschiedene Prozesse ist in Abbildung 2.2 dargestellt. Obwohl der Energieverlust pro Strecke durch Bremsstrahlungs- und Paarproduktionsprozesse bei Energien über einem TeV ähnlich ist, unterscheiden sich diese Prozesse deutlich in ihren Charakteristiken. Bei der Paarproduktion entstehen sehr viele Elektron-Positron-Paare, welche in Summe einen signifikanten Anteil zum Energieverlust beitragen. Die Bremsstrahlung hingegen ereignet sich deutlich seltener, führt dann jedoch zu einem relativ großen Energieverlust. Im Extremfall, also in dem Fall, dass das Myon einen Großteil der Energie verliert, spricht man von katastrophalen Energieverlusten (*englisch: catastrophic energy loss*). Bei Bremsstrahlungsverlusten entstehen elektromagnetische Kaskaden, welche im Detektor sichtbar sind [KFS⁺ 13].

Das Lepton und die Sekundärteilchen aus den verschiedenen Wechselwirkungen sind in der Lage das umgebende Medium zu polarisieren und so Licht zu erzeugen. Da ihre Geschwindigkeit größer als die Lichtgeschwindigkeit im Medium ist, interferieren die Wellen nicht mehr destruktiv. Dieses Phänomen wird Tscherenkoeffekt genannt und ist der Effekt zur Detektion der geladenen Leptonen [Čer37].

Das entstehende Licht kann mit Photosensoren [A⁺ 10a] detektiert werden. Damit das Licht die Sensoren erreicht, ist ein transparentes Medium notwendig. Um die zuvor erwähnten großen Volumina zu ermöglichen, ist entweder klares Wasser oder Eis notwendig. IceCube nutzt zur Detektion das klare Eis am Südpol. Der ANTARES-Detektor [A⁺ 11c] nutzt hingegen das Wasser des Mittelmeeres als Detektionsmedium.

IceCube befindet sich in einer Tiefe von 1450 m bis 2450 m unter der Oberfläche. Dies ist notwendig, um den Untergrund durch atmosphärische Myonen zu reduzieren. Auf atmosphärische Myonen wird in Kapitel 2.4.1 näher eingegangen.

Der komplette IceCube-Detektor besteht aus 86 Strings. An diesen sind digitale optische Module (kurz DOMs) angebracht, welche einen Photosensor beinhalten. Der volle Ausbau mit 86 Strings beinhaltet auch den Subdetektor DeepCore. Dieser Teil von IceCube besteht aus einigen Strings im tiefen Eis, welche eine höhe-

2. Astrophysikalische Grundlagen

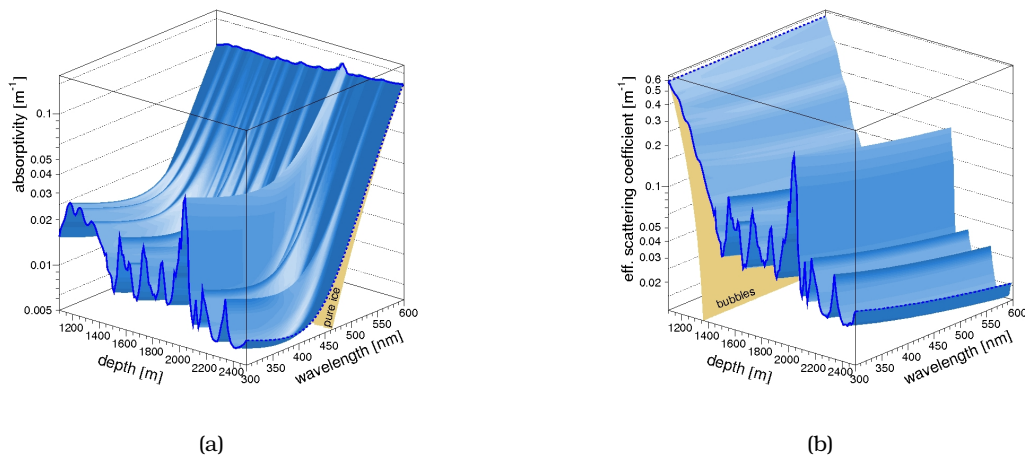


Abbildung 2.3.: (a) Absorptions- und (b) Streukoeffizienten für das WHAM Eismodell in Abhängigkeit von Tiefen und Wellenlänge. Aus diesen Daten können mehrere Bereiche identifiziert werden, in denen das Eis deutlich schlechtere Eigenschaften hat. Der prägnanteste Bereich liegt dabei in einer Tiefe von ca. 2050 m. Dieser auch als dust layer bezeichnete Bereich korrespondiert mit einem Eisalter von ca. 65 000 Jahren [A⁺06b].

re Dichte an DOMs aufweisen, um auch Neutrinos mit Energien unterhalb von 100 GeV messen zu können [W⁺09, A⁺12b].

Wichtig bei der Interpretation der IceCube-Daten ist die Kenntnis der optischen Eigenschaften des Eises. Insbesondere die tiefenabhängigen Streu- und Absorptionskoeffizienten sind von Interesse. Gerade die Streuung hat dabei einen starken Einfluss auf die Detektion. Hinzu kommt, dass es während der Entstehungszeit des Eises immer wieder zu Verschmutzungen durch Staub kam. Daher gibt es innerhalb des Eises verschiedene Staubschichten (*englisch: dust layer*), welche die Ausbreitung der Photonen stark beeinflussen. Um diese Verschmutzungen parametrisieren zu können, beinhalten einige DOMs LED-Flasher. Mit ihrer Hilfe können wohldefinierte Lichtpulse durch das Eis gesendet werden und so die optischen Eigenschaften des Eises vermessen werden. Zusätzlich zu diesen Daten gibt es noch Bohrkerne, welche die Ergebnisse verifizieren. Die Parametrisierung des Eises ist eine der größten Unsicherheiten bei der Messung von Neutrinos mit IceCube [Chi11, A⁺13c].

Der Bau von IceCube begann im antarktischen Sommer 2004/2005. Während in diesem Jahr nur ein einziger String deponiert wurde, wurde die Anzahl im Laufe

2. Astrophysikalische Grundlagen

der Zeit weiter erhöht. Die bis zum jeweiligen Zeitpunkt fertiggestellten Teile des Detektors konnten bereits für Analysen genutzt werden. Die Ausbaustufen des Detektors werden nach der Anzahl funktionsfähiger Strings bezeichnet. Diese waren IC 22, IC 40, IC 59, IC 79 und IC 86. Die fortlaufenden Jahre mit voller Anzahl von Strings werden jeweils mit IC 86-I, IC 86-II usw. bezeichnet.

2.2. Astrophysikalische Neutrinos

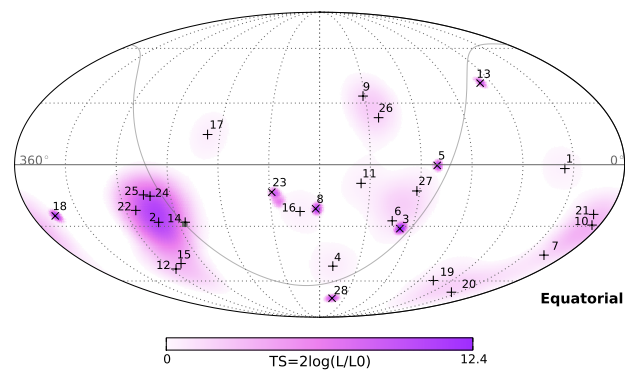


Abbildung 2.4.: Ereignisse aus der Suche nach Hochenergieneutrinos mittels im Detektor startenden Ereignissen mit dem IceCube-Detektor in äquatorialen Koordinaten [A⁺13a]. Obwohl es eine optische Häufung gibt, ist diese nicht statistisch signifikant. Pluszeichen (+) symbolisieren Kaskaden und Kreuze (x) Myonspuren. Auf der Farbachse ist das logarithmierte Verhältnis der Likelihood-Werte für Signal- und Untergrundhypothese aufgetragen.

Im Jahr 2013 ist es IceCube gelungen, Neutrinos, die nicht aus diesem Sonnensystem kommen, mit einem Konfidenzniveau von 4.8σ nachzuweisen [A⁺13a]. Seit dieser Entdeckung gilt es als gesichert, dass es extraterrestrische Neutrinos im Energiebereich oberhalb von 100 TeV gibt. Diese Analyse war der erste Nachweis solcher Neutrinos, seit der Entdeckung der Neutrinos der Supernova 1987A [H⁺87].

Die in [A⁺13a] genutzte Analysetechnik ist komplementär zu der später in dieser Analyse genutzten Technik.

Die Messung dieser Neutrinos ist durch den hohen Untergrund schwierig. Zum einen gibt es einen Untergrund durch atmosphärische Myonen (vgl. Abschnitt

2. Astrophysikalische Grundlagen

2.4.1) und zum anderen einen Untergrund durch atmosphärische Neutrinos (vgl. Abschnitt 2.4.1). Diese dominieren das Signal um viele Größenordnungen. Um den Untergrund zu minimieren, wurden Ereignisse gesucht, die im Detektor starten. Um dies zu realisieren, wurde die äußere Schicht DOMs als Vetoschicht genutzt. Auf diese Weise konnte die Anzahl der atmosphärischen Myonen reduziert werden. Atmosphärische Neutrinos konnten durch ihr Selbstveto minimiert werden. Da diese Neutrinos in Schauern entstehen, werden sie in der Regel von Myonen begleitet [GJKvS14]. Diese Myonen erzeugen ein Signal in der Vetoschicht. Dadurch werden diese Ereignisse verworfen und der Untergrund durch atmosphärische Neutrinos kann weiter reduziert werden.

Diese Methodik in Kombination mit einer Selektion von sehr hochenergetischen Ereignissen ab 30 TeV ermöglichten eine Reduzierung des Untergrunds auf $10.6^{+5.0}_{-3.6}$ Ereignissen. In zwei Jahren, in denen IceCube Daten genommen hat, konnten 28 Ereignisse detektiert werden. Mit Hilfe der Energieverteilung dieser Ereignisse ergibt sich daraus die zuvor erwähnte Signifikanz von 4.8σ . Die Ereignisse sind in Abbildung 2.4 in äquatorialen Koordinaten dargestellt. Obwohl visuell eine Häufung sichtbar ist, ist diese nicht signifikant. Die Verteilung der Ereignisse deutet auf einen isotropen Fluss hin, der eine Energieverteilung proportional zu $E^{-2.2 \pm 0.4}$ besitzt [A⁺13a].

Schon lange waren Neutrinos als Botenteilchen für astrophysikalische Prozesse in der Diskussion. Um das Hochenergieuniversum zu erforschen, gibt es nur wenige Botenteilchen. Protonen verlieren auf Grund ihrer Ladung in aller Regel ihre Richtungsinformation und sind daher nicht geeignet um Quellen direkt zu untersuchen.

Photonen weisen diesen Nachteil nicht auf. Das Universum ist jedoch nicht vollständig durchlässig für Photonen im Energiebereich oberhalb von 1 PeV. Daher sind auch Photonen nicht geeignet, um mit ihnen alle Phänomene des Universums zu untersuchen. Neutrinos haben keinen dieser Nachteile. Werden die Probleme mit dem hohen Untergrund und der kleinen Anzahl an Signalereignissen bei der Detektion dieser Neutrinos überwunden, sind Neutrinos die idealen Botenteilchen, da sie die Richtungsinformation behalten und ungehindert den intergalaktischen Raum durchqueren können.

Für detaillierte Übersichten der Neutrinoastronomie sei auf [KS12] oder [Bec07] verwiesen.

2.3. Primäre kosmische Strahlung

Die Erde wird jederzeit von einer Vielzahl an Teilchen getroffen. Den größten Anteil an dieser Strahlung tragen Protonen und andere Hadronen. Die Energie dieser Teilchen reicht über viele Größenordnungen bis hin zu 10^{22} eV und liegt damit weit jenseits der Energien, die in irdischen Teilchenbeschleunigern erzeugt werden können. Im Folgenden wird auf die primäre kosmische Strahlung eingegangen. Der Fokus liegt vor allem auf dem Energiespektrum und der Komposition, also der chemischen Zusammensetzung. Anschließend werden die Probleme in der theoretischen Beschreibung des Primärflusses beschrieben [YD98].

2.3.1. Energieverteilung der primären geladenen kosmischen Strahlung

Die Energieverteilung der primären geladenen kosmischen Strahlung, welche in Abbildung 2.5 abgebildet ist, erstreckt sich von ca. 10^9 eV bis hin zu 10^{22} eV. Im Energiebereich unterhalb von ungefähr 10^{14} eV kann direkt gemessen werden. Um diesen Energiebereich zu vermessen werden verschiedene Ballon- und Satellitenexperimente genutzt [B⁺12]. Bei höheren Energien ist eine solche Detektion schwierig, da der Fluss stark abfällt und die benötigten Detektionsflächen nicht zur Verfügung stehen. Daher wird in diesen Energiebereichen auf bodengebundene Arrays zurückgegriffen [YD98].

Der Fluss Φ der primären kosmischen Strahlung kann in erster Näherung als stückweise definiertes Potenzgesetz der Form

$$\frac{d\Phi(E)}{dE} = \Phi_0 E^{-\gamma} \quad (2.3.1)$$

parametrisiert werden.

Es gibt einige Besonderheiten, welche sich durch Änderungen des spektralen Indexes γ auszeichnen.

Die erste Besonderheit ist das sogenannte Knie (*englisch: knee*), welches bei einer Energie von $\sim 10^{15}$ eV liegt. Hier wechselt der spektrale Index von 2.7 auf 3.0 [YD98]. Dieser Index bleibt konstant bis zu einer Energie im Bereich von $\sim 10^{17}$ eV bis 10^{18} eV. An dieser Stelle befindet sich das zweite Knie (*englisch: second knee*). An dieser Stelle ändert sich der spektrale Index auf 3.3. Eine weitere Änderung des Index ist bei $\sim 10^{18}$ eV zu beobachten. An diesem, auch Ferse (*englisch: ankle*) genannten Punkt, ändert sich der spektrale Index auf 2.7 [Hör07, YD98].

2. Astrophysikalische Grundlagen

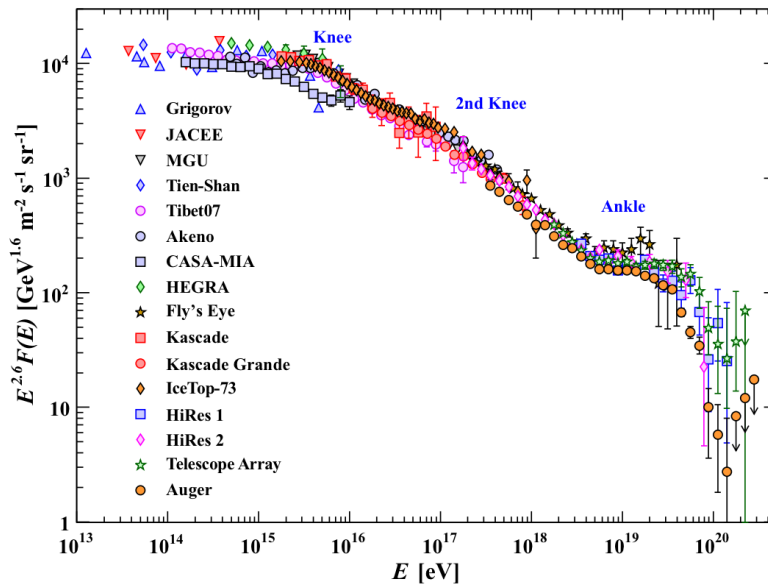


Abbildung 2.5.: Das Energiespektrum der kosmischen Strahlung gemessen mit verschiedenen Experimenten gewichtet mit $E^{-2.6}$ [B⁺ 12].

Bei höchsten Energien ist ein Abbruch zu erwarten (vgl. Abschnitt 2.3.2). Dieser Cutoff ist Gegenstand aktueller Forschung [A⁺ 10c].

Einige experimentelle Daten sind in Abbildung 2.5 dargestellt. Die Messdaten entstammen einer Reihe von Experimenten. Die erwähnten Eigenschaften des Flusses sind grundsätzlich sichtbar. Die Messungen unterscheiden sich zum Teil signifikant. Sehr deutlich wird dies im Bereich des Knies, das in Abbildung 2.6 detaillierter dargestellt ist.

Die unterschiedlichen Messungen machen eine Parametrisierung des Primärflusses schwierig. Bei der Beschreibung des Primärflusses ist neben der Energieverteilung auch die Komposition der Strahlung von Bedeutung. Die Komposition ist experimentell nur schwer zugänglich und daher schlecht vermessen. Üblich genutzte Modelle sind GST [GST13], HillasGaisser [Gai12], polygonato [Hör03] und Zatsepin-Sokolskaya [ZS06]. Diese Modelle sind in Abbildung 2.7 dargestellt. Auch bei diesen Parametrisierungen ist ersichtlich, dass sie sich stark unterscheiden. Dies ist über den gesamten Energiebereich sichtbar. Am deutlichsten wird dies jedoch bei einer Energie oberhalb von 10^7 GeV.

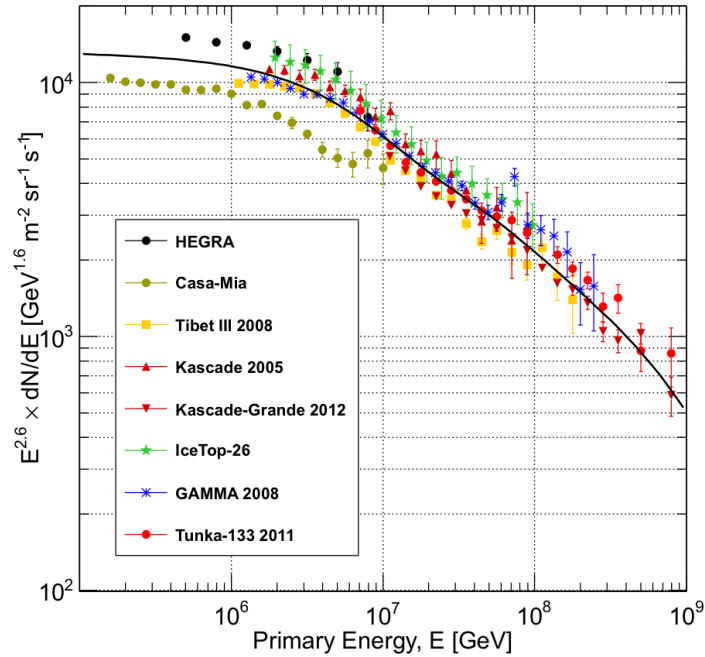


Abbildung 2.6.: Detailliertere Darstellung für die Messpunkte der kosmischen Strahlung im Bereich des Knies.

Die Ergebnisse der verschiedenen Experimente unterscheiden sich deutlich. Vor allem der Bereich des Knies weist stark unterschiedliche Messdaten auf. Durch diese Unsicherheit in den Messungen ist eine theoretische Parametrisierung nur schwer möglich. Eine mögliche Parametrisierung ist die eingezeichnete schwarze Kurve nach [GST13].

2.3.2. Greisen-Zatsepin-Kuzmin Cutoff

Hochenergetische kosmische Protonen können mit dem kosmischen Mikrowellenhintergrund (*englisch: cosmic microwave background, CMB*) über



wechselwirken.

Zum einen ist die Konsequenz dieser Reaktion, dass die Protonen in den höchsten Energien, also bei mehr als $\sim 10^{20}$ eV, vom CMB absorbiert werden und somit

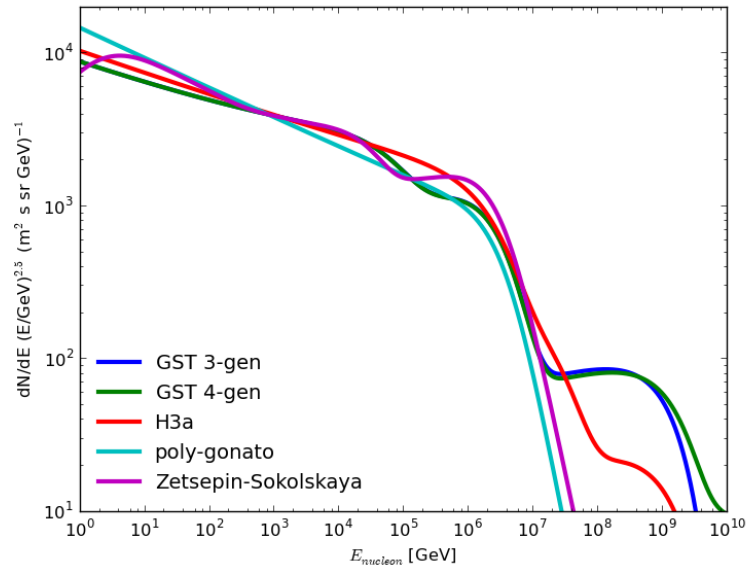


Abbildung 2.7.: Verschiedene Parametrisierungen der primären kosmischen Strahlung [GST13] [Gai12] [Hör03] [ZS06]. Diskrepanzen zwischen den Modellen sind deutlich erkennbar. Vor allem bei Energien oberhalb von $\sim 10^7$ GeV unterscheiden sich die gezeigten Modelle [AEF⁺].

nicht mehr auf der Erde ankommen. Dieser sogenannte Greisen-Zatsepin-Kuzmin Cutoff wurde erstmals in [ZK66] und [Gre66] beschrieben. Zum anderen kann das entstehende geladene Pion über

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \quad (2.3.3)$$

zerfallen und erzeugt Neutrinos. Diese Neutrinos sind, da sie nicht interagieren, auf der Erde als Signal messbar. Dieser Effekt gilt als ein garantierter Effekt für hochenergetische extraterrestrische Neutrinos, wurde jedoch bisher noch nicht beobachtet. Es gibt verschiedene dedizierte Experimente, die solche Neutrinos suchen. Unter anderem sind dies ANITA [M⁺05] und ARA [A⁺12c].

2.4. Luftschauer

Treffen die zuvor erwähnten Hadronen und Photonen der kosmischen Strahlung auf die Atmosphäre, so bilden sich ausgedehnte Luftschauer. Für diese Arbeit sind dabei die Protonen von hoher Bedeutung, da in diesen Schauern Myonen und Neutrinos entstehen. In Schauern, welche von Photonen induziert werden, entstehen in der Regel nur wenige Myonen und Neutrinos, da sie hauptsächlich elektromagnetische Schauer bilden. Im Folgenden geht es um die Myonen und Neutrinos, welche als Sekundärteilchen in den hadronischen Schauern entstehen.

2.4.1. Konventionelle atmosphärische Myonen und Neutrinos

In einem hadronischen Teilchenschauer entstehen zahlreiche Pionen und Kaonen. Beim Zerfall von geladenen Pionen und geladenen Kaonen entstehen über

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \quad \pi^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (2.4.1)$$

$$K^+ \rightarrow \mu^+ + \nu_\mu, \quad K^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (2.4.2)$$

hauptsächlich (Anti-)Myonen und (Anti-)Myon-Neutrinos. Näherungsweise können die Neutrinos aus diesen Zerfällen durch

$$\frac{d^2\Phi}{dE_\nu d\Omega}(E_\nu, \theta) = 0.0285 E_\nu^{-2.7} \left\{ \frac{1}{1 + \frac{6E_\nu \cos \theta^*}{115 \text{ GeV}}} + \frac{0.213}{1 + \frac{1.44E_\nu \cos \theta^*}{850 \text{ GeV}}} \right\} \quad (2.4.3)$$

beschrieben werden [LM00]. Mit Hilfe dieser Gleichung ist es möglich das Verhältnis von Myonen und Neutrinos aus Pion- und Kaon-Zerfällen zu berechnen. Das Verhältnis von Myon-Neutrinos aus Pion- und Kaon-Zerfällen ist in Abbildung 2.8 dargestellt. Ab einer Energie von 100 GeV werden Kaonenzerfälle für Neutrinos dominant. Im Myonkanal bleiben die Pionen dominant.

Da Gleichung (2.4.3) nur näherungsweise gilt, werden normalerweise Rechnungen genutzt, die entweder simulationsgestützte Lösungen oder Hybridlösungen zwischen analytischen und numerischen Lösungen sind.

In der Regel werden zwei Modelle genutzt. Dies sind die Rechnungen von Honda et. al [HKK⁺07]. Die Rechnungen sind ursprünglich für eine Energie von maximal 10⁴ GeV durchgeführt worden. Daher wurden sie extrapoliert, um die volle Energieskala von IceCube abzudecken. Bei Energien ab 100 TeV beginnt das Knie der

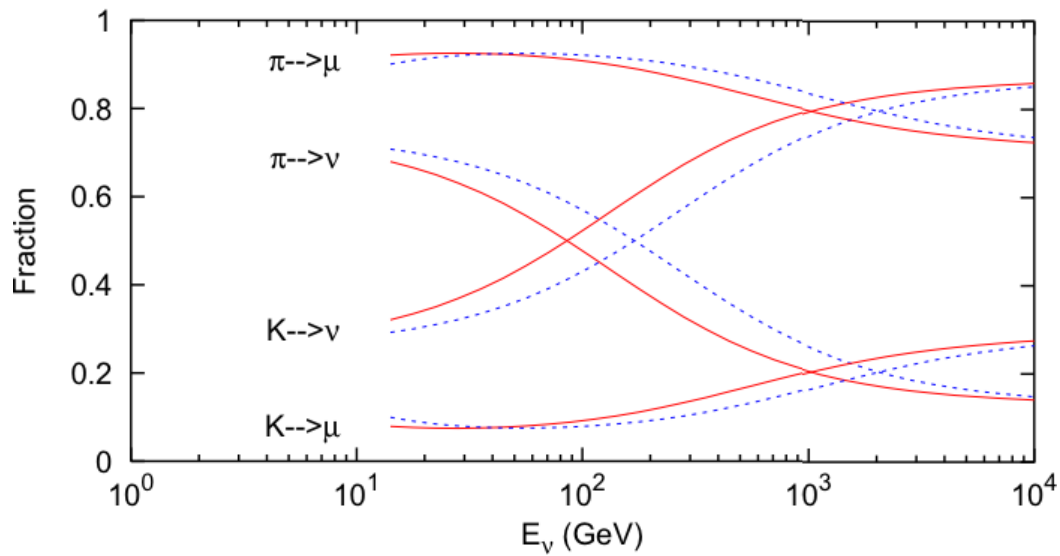


Abbildung 2.8.: Relativer Anteil von Neutrinos und Myonen aus Pion oder Kaonzerfällen am konventionellen Fluss. Die rote Kurve ist der Anteil für vertikale Ereignisse. Die gestrichelte blaue Kurve beschreibt Ereignisse unter einem Zenitwinkel von $\theta = 60^\circ$ [GH02].

kosmischen Strahlung den Neutrinofluss zu beeinflussen. Daher wurden weitere Modifikationen vorgenommen, um diesem Umstand Rechnung zu tragen [A⁺14d]. Die aktuellsten Rechnungen stammen aus [FTD12]. Für die Rechnungen wurden große Mengen an Simulationen genutzt. Diese Simulationen ermöglichen es, den Fluss bis zu einer Energie von 100 TeV zu bestimmen. Bei höheren Energien werden Hybridrechnungen unter Berücksichtigung von analytischen Rechnungen genutzt.

2.4.2. Atmosphärische Elektron-Neutrinos

Die zuvor erwähnten Pionen und Kaonen in Luftschauern zerfallen vorwiegend in ein Myon und ein Myon-Neutrino, da der Zerfall in ein Elektron und ein Elektron-Neutrino helizitätsunterdrückt ist. Das Verzweigungsverhältnis für den Zerfall

$$\pi^+ \rightarrow e^+ + \nu_e \quad (2.4.4)$$

beträgt nur $(1.230 \pm 0.004) \cdot 10^{-4}$ [B⁺12]. Der Zerfall des Myons aus Gleichung (2.4.1) erzeugt nur niederenergetische Elektron-Neutrinos, da die Myonen, bevor

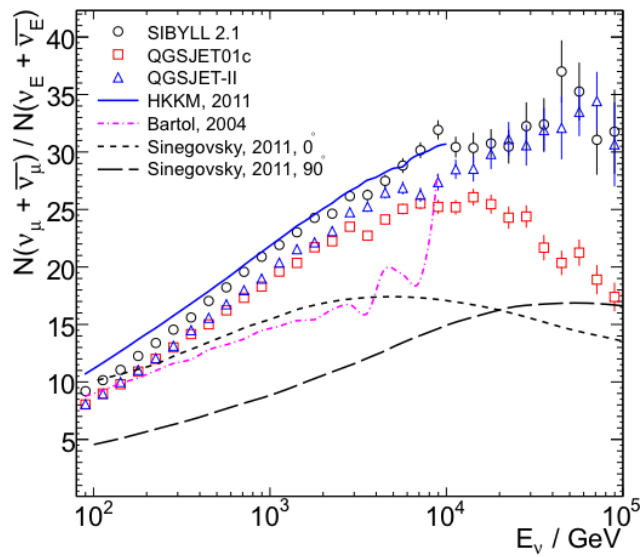


Abbildung 2.9.: Verhältnis von konventionellen atmosphärischen Elektron- und Myon-Neutrinos aus verschiedenen Rechnungen [FTD12]. Über den ganzen Energiebereich sind Myon-Neutrinos dominant; bei höheren Energien steigt dieser Anteil weiter an.

sie zerfallen, bereits viel Energie verloren haben. Die signifikanten Beiträge zum Elektron-Neutrinofluss kommen aus Dreikörper-Kaon-Zerfällen. Für die geladenen Kaonen ist dies der Zerfall über

$$K^+ \rightarrow \pi^0 e^+ \nu_e \quad K^- \rightarrow \pi^0 e^- \bar{\nu}_e. \quad (2.4.5)$$

Ein zweiter Beitrag kommt aus dem Zerfall von K_L^0 über

$$K_L^0 \rightarrow \pi^+ e^- \bar{\nu}_e \quad K_L^0 \rightarrow \pi^- e^+ \nu_e. \quad (2.4.6)$$

analog zu Gleichung (2.4.3) ist die analytische Näherung zur Beschreibung für diese atmosphärischen Elektron-Neutrinos durch

$$\frac{d^2\Phi}{dE_\nu d\Omega}(E_\nu, \theta) = 0.0024 E_\nu^{-2.7} \left\{ \frac{0.05}{1 + \frac{1.5E_\nu \cos \theta^*}{850 \text{ GeV}}} + \frac{0.185}{1 + \frac{1.5E_\nu \cos \theta^*}{194 \text{ GeV}}} + \frac{11.4/E_\nu}{1 + \frac{1.21E_\nu \cos \theta^*}{115 \text{ GeV}}} \right\} \quad (2.4.7)$$

gegeben [LM00]. Mit diesen Gleichungen ist es möglich, das Verhältnis aus Elektron- und Myon-Neutrinos $\frac{\nu_\mu + \bar{\nu}_\mu}{\nu_e + \bar{\nu}_e}$ zu berechnen. Dieses Verhältnis ist energieabhängig

2. Astrophysikalische Grundlagen

und in Abbildung 2.9 zu sehen. In dem gesamten Energiebereich sind die Myon-Neutrinos dominant.

Für astrophysikalische Neutrinoquellen ist das Flavourverhältnis auf Grund von Oszillationen $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$ [AJY00]. Der relative Untergrund durch atmosphärische Elektron-Neutrinos ist daher in diesem Detektionskanal geringer als im Myon-Kanal. Aktuelle Messungen des atmosphärischen Elektron-Neutrino-flusses finden sich in [A⁺13b].

2.4.3. Prompte Neutrinos

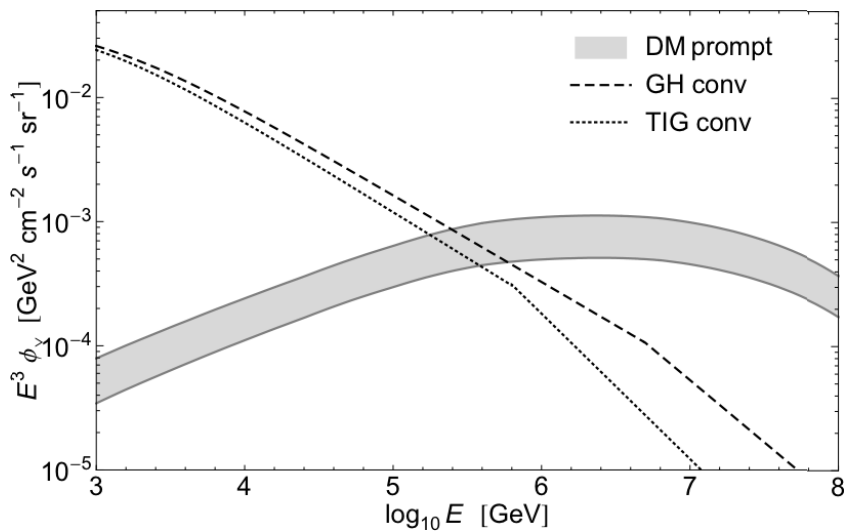


Abbildung 2.10.: Der prompte Neutrinofluss gewichtet mit E^{-3} mit Fehlerband im Vergleich zum konventionellen Neutrinofluss mit zwei verschiedenen Modellen für die primäre kosmische Strahlung [ERS08]. GH ist das GaisserHonda-Modell [GH02] und TIG ist das Modell nach [TIG96]. Die prompte Komponente wird ab ca. 1 PeV dominant.

Neben Pionen und Kaonen können in der Atmosphäre auch Hadronen mit schwereren Quarks (hauptsächlich Charm-Quarks) entstehen. Diese Teilchen zerfallen schneller und daher werden die entstehenden Neutrinos prompte atmosphärische Neutrinos genannt. Die Neutrinos aus Pion- und Kaon-Zerfällen werden hingegen konventionelle atmosphärische Neutrinos genannt. Da die Lebensdauer der Teilchen mit schweren Quarks kurz ist, verlieren sie kaum Energie, bevor sie zerfallen.

2. Astrophysikalische Grundlagen

Daher erben die entstehenden prompten Neutrinos die Energieverteilung des primären Spektrums. Das Spektrum der prompten atmosphärischen Neutrinos ist dadurch näherungsweise durch

$$\Phi(E) \propto E^{-2.7} \tag{2.4.8}$$

gegeben. Genauere theoretische Berechnungen sind in [ERS08] zu finden. Die theoretische Vorhersage für diesen Fluss gestaltet sich jedoch schwierig. Der Grund dafür ist, dass die Produktion dieser Teilchen in Vorwärtsrichtung nur schwer berechenbar ist. In Beschleunigerexperimenten ist die Vermessung der benötigten Größen schwierig, da das Strahlrohr die Messung behindert. Mit einer Messung des prompten Flusses durch atmosphärische Myonen oder aber durch atmosphärische Neutrinos wäre somit ein großer Beitrag der Astroteilchenphysik zur klassischen Teilchenphysik möglich. Für eine nähere Betrachtung der Unsicherheiten für prompten Neutrinos siehe [Enb14].

3. Methoden der Datenanalyse

In der modernen Teilchenphysik tritt in aller Regel das Problem auf, dass die Datenmengen immer größer werden und gleichzeitig das gesuchte Signal immer kleiner wird. Um dieses Problem zu lösen, werden Methoden benötigt, um große Datenmengen effizient untersuchen zu können. Im folgenden Kapitel werden die in der Analyse genutzten Methoden vorgestellt. Dazu wird zunächst in Abschnitt 3.1 der Fokus auf die Erzeugung von Simulationsdaten gelegt. Diese sind für viele sogenannte überwachte Lernalgorithmen (englisch *supervised learning algorithms*) notwendig. Anschließend werden die Methoden der beiden großen Teilbereiche der Analyse vorgestellt. Dies ist zum einen die in Abschnitt 3.3 vorgestellte Trennung von Signal und Untergrund mittels multivariaten Methoden. Hierbei geht es vor allem um die Trennung von atmosphärischen Myonen und neutrinoinduzierten Myonen. Zum anderen ist dies die in Abschnitt 3.4 vorgestellte Methodik der Entfaltung, welche es ermöglicht, das Energiespektrum der zuvor selektierten Neutrinos aus gemessenen energieabhängigen Merkmalen zu bestimmen.

3.1. Simulation

Für viele Datenanalyse-Verfahren ist eine Kenntnis von verschiedenen Arten an Ereignissen im Detektor unabdingbar. Aus diesem Grunde wird der komplette Messprozess simuliert. Diese Simulation ist sehr umfangreich und umschließt zahlreiche einzelne Komponenten. Im Folgenden werden die Komponenten erklärt, welche für die in der Analyse genutzten Simulationen wichtig sind. Da exotische Teilchen wie WIMPs, Monopole oder dunkle Materie in dieser Analyse nicht betrachtet werden, werden die entsprechenden Module hier nicht erläutert.

Für diese Arbeit relevante Teilchengeneratoren sind CORSIKA [HKC⁺98] und NuGen [GK05]. CORSIKA ist ein Simulationsprogramm für Luftschauer. Es wurde ur-

3. Methoden der Datenanalyse

sprünglich für das KASCADE-Experiment [A⁺03] entwickelt, ist nun aber in vielen Experimenten der Astroteilchenphysik im Einsatz. Vor allem um die hadronischen Komponenten der Luftschauber zu simulieren, nutzt CORSIKA verschiedene andere Bibliotheken. Für die hadronischen Komponenten sind dies vor allem QGSJet [Ost06] und SIBYLL [AEG⁺09]. Neutrinos mit Energien größer als ~ 100 GeV werden in IceCube mit NuGen erzeugt. Für Neutrinos mit niedrigeren Energien wird GENIE [A⁺10d] genutzt.

Die Myonen aus den Teilchenschauern, sowie die Leptonen aus den Neutrino-Nukleon-Reaktionen müssen durch verschiedene Medien propagiert werden. Im Fall von Myonen ist diese Wegstrecke zum Teil mehrere Kilometer lang. Vor allem im Detektor ist eine genaue Beschreibung der Energieverluste wichtig. Die Stochastizität der Bremsstrahlung kann verschiedene Muster im Detektor hinterlassen. Zur Propagation von Leptonen wird PROPOSAL bzw. MMC genutzt, dieser Propagator legt seinen Fokus auf eine genaue Beschreibung der Physik bei hoher Rechengeschwindigkeit [KFS⁺13].

Beim Propagieren des Leptons entstehen Tscherenkowphotonen. Diese Photonen müssen ebenfalls durch das Medium propagiert werden. Dies ist eine rechenintensive Aufgabe, da im Idealfall jedes Photon einzeln behandelt werden sollte. Um dies zu gewährleisten, nutzt IceCube zwei verschiedene Photonpropagatoren, ppc und clsim [Chi13]. Beide sind für die Nutzung von Grafikkarten optimiert. Durch die hohe parallele Rechenkraft von Grafikkarten ist eine effiziente Propagation möglich. Abschließend wird die gesamte Hardware des Detektors simuliert. Hierzu gehört insbesondere das Rauschverhalten der Elektronik [Lar13]. Um die enormen Datenmengen produzieren zu können, werden weltweit verteilte Datenzentren zusammengeschlossen. Das benutzte IceProd Framework ist in [DV14] beschrieben.

3.2. Algorithmen zur Rekonstruktion von Ereignissen

Um eine effiziente Datenverarbeitung zu gewährleisten, ist es wichtig die aufgenommenen Ereignisse möglichst genau zu beschreiben. Neben einfachen Observablen wie N_{Ch} , also der Anzahl der getroffenen DOMs, gibt es auch komplexe Rekonstruktionsalgorithmen, welche in IceCube verwendet werden. Im Folgenden werden die wichtigsten von ihnen vorgestellt. Physikalisch relevante Merkmale sind vor allem die Energie sowie die Richtungsinformationen. Je nach Flavour des Neutrinos ist die Auflösung von IceCube sehr unterschiedlich.

Für Myon-Neutrinos ist die Winkelauflösung auf Grund der Signatur mit unter einem Grad sehr gut [A⁺14c]. Da ein Myon-Neutrino jedoch einige Kilometer vor dem Detektor interagieren kann und nur ein kurzer Teil der Spur im Detektor sichtbar ist, ist die Energieauflösung bei Myonen-Neutrinos schlechter als bei Elektron-Neutrinos. Elektron-Neutrino-Signaturen sind häufig vollkommen im Detektor enthalten. Damit kann IceCube wie ein Kalorimeter genutzt werden und hat eine deutlich bessere Energieauflösung als bei Myon-Neutrinos. Die sphärische Struktur der Signaturen macht jedoch die Winkelauflösung für Elektron-Neutrinos problematisch. Die Winkelauflösung ist bei diesen Ereignissen schlechter als 10° [M⁺09].

Für die Separation von Signal und Untergrund können unter Umständen physikalisch uninteressante Merkmale von großer Bedeutung sein. In der Separation ist es gewünscht, dem Algorithmus möglichst viele nicht redundante Informationen über das Ereignis zur Verfügung zu stellen. Im Folgenden werden einige Rekonstruktionsalgorithmen kurz vorgestellt, welche für die weitere Analyse von Bedeutung sind.

3.2.1. Line-Fit

Der Line-Fit ist ein einfacher Algorithmus, um die Eigenschaften eines Ereignisses zu beurteilen. Es wird angenommen, dass ein Ereignis eine Spur hinterlässt. Sowohl die Geometrie des Tscherenkowkegels als auch die Eigenschaften des Eises werden ignoriert. Dadurch lässt sich die χ^2 -Funktion

$$\chi^2 = \sum_{i=0}^{N_{\text{Fit}}} (\mathbf{r}_i - \mathbf{r} - \mathbf{v} \cdot t_i)^2 \quad (3.2.1)$$

definieren. Die Anzahl aller Treffer wird mit N_{Hit} bezeichnet. Die χ^2 -Funktion ist abhängig von dem Punkt des Vertex \mathbf{r} sowie dem Geschwindigkeitsvektor \mathbf{v} und dem individuellen Ortsvektor eines DOMs \mathbf{r}_i . Nach der Minimierung lässt sich aus dem Geschwindigkeitsvektor die Richtung $\mathbf{e} = \frac{\mathbf{v}}{|\mathbf{v}|}$ und daraus der Zenitwinkel $\Theta = -\arccos(v_z/|\mathbf{v}|)$ bestimmen.

Ein weiterer wichtiger, aus dieser Minimierung ableitbarer Parameter ist $v_{LF} = |\mathbf{v}|$. Dieser Wert gibt die mittlere Geschwindigkeit der Lichtpropagation an. Dieser Wert ist wichtig, da er ein Diskriminanzkriterium zwischen Spur- und Kaskadenereignissen ist. Kaskadenereignisse neigen zu $v_{LF} \sim 0$, da bei ihnen eine sphärische Lichtausbreitung vorliegt.

Ein großer Vorteil dieser Methode ist die analytische Lösbarkeit. Die Minimierung kann daher sehr effizient geschehen. Diese Minimierung wird zum einen als frühes Diskriminanzkriterium genutzt und zum anderen dienen die bestimmten Parameter auch als Startpunkt für komplexere Algorithmen [A⁺04].

3.2.2. SPE- und MPE-Fit

Eine genauere Berechnung der Parameter des Ereignisses ist über den SPE- (**S**ingle **P**hoto**E**lectron) bzw. den MPE-Algorithmus (**M**ulti-**P**hoto**E**lectron) möglich. Dieser benutzt eine Likelihoodbeschreibung des Problems. Die Likelihoodfunktion des Problems lautet

$$\mathcal{L}(\mathbf{x}|\mathbf{a}) = \prod_{x_i} p(x_i|\mathbf{a}). \quad (3.2.2)$$

$\mathbf{a} = (\mathbf{r}_0, t_0, \mathbf{p}, E_0)$ ist die Menge an physikalischen Parametern, welche das Ereignis definieren. Diese Parameter sind zur Verdeutlichung in Abbildung 3.1 dargestellt. Die Wahrscheinlichkeitsdichtefunktion $p(x_i|\mathbf{a})$ ist die Wahrscheinlichkeitsdichtefunktion bei einer Messung x_i die Spurparameter \mathbf{a} zu realisieren. Für einen perfekten Detektor wäre diese Funktion auf Grund der Zeitauflösung eine Deltafunktion. Durch die Verschmierung des Detektors ist dies jedoch in der Regel nicht gegeben. Zu dieser Verschmierung tragen zum einen die Eiseigenschaften, aber auch die stochastischen Verluste durch Bremsstrahlung und Rauscheffekte bei. Es ist daher üblich für eine weitere Beschreibung nur den Zeitunterschied zwischen der geometrischen und der wahren Ankunftszeit

$$t_{res} = t_{geo} - t_{hit} \quad (3.2.3)$$

3. Methoden der Datenanalyse

zu betrachten. Die Zeit t_{geo} ist hierbei die Zeit, die das Photon braucht, wenn es ungestreut zum DOM propagiert. Der Wert t_{hit} ist hingegen die Zeit, die es unter Berücksichtigung aller Effekte benötigt. Die Wahrscheinlichkeitsdichtefunktion $p(t_{res})$ kann aus der Simulation einzelner Photonen bestimmt werden. Im SPE-Algorithmus bildet dann das Produkt der $p(t_{res})$ -Verteilungen die zu maximierende Likelihood-Funktion. Als Zeitpunkt wird die steigende Flanke des ersten Lichtpulses genutzt. Dies ist jedoch nur korrekt, wenn jeder DOM nur ein Photon detektiert hat. Diesen Umstand berücksichtigt der MPE-Algorithmus. Hierbei wird die kumulative Dichtefunktion für die Ankunftszeiten genutzt. Eine detaillierte Beschreibung der Wahrscheinlichkeitsdichtefunktion ist in [A⁺04] zu finden. Mit Hilfe des MPE-Algorithmus ist für hochenergetische Spuren eine Winkelauflösung von weniger als 1° möglich [A⁺14c].

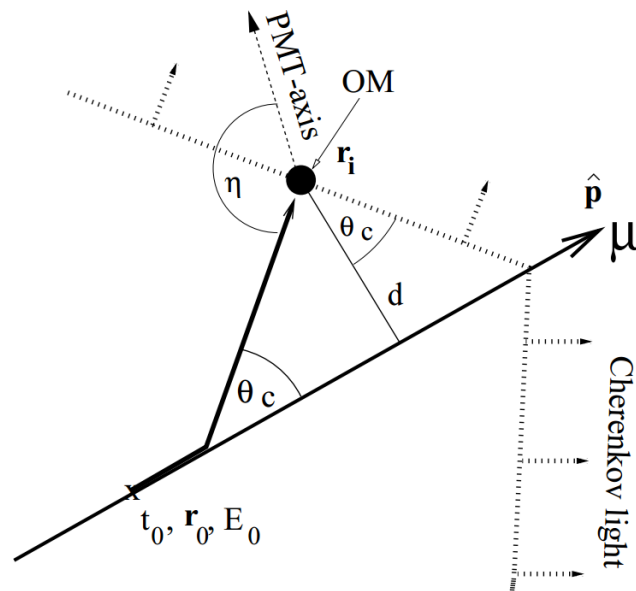


Abbildung 3.1.: Skizze zur Veranschaulichung der Parameter im SPE- bzw. MPE-Fit. \mathbf{p} ist der Richtungsvektor der Spur. \mathbf{r}_0 ist ein beliebiger Ort auf der Spur. E_0 und t_0 sind die Energie und der Zeitpunkt beim Durchgang des Teilchens durch \mathbf{r}_0 . θ_c ist der Tscherenkowwinkel [A⁺04].

Aus beiden Maximierungen kann als Gütekriterium der Wert der negativen Log-Likelihood-Funktion $\log \mathcal{L}$ gewonnen werden. Dieser Wert ist jedoch unnormiert und daher schwer zu interpretieren. Eine bessere Interpretation kann durch ei-

ne Normierung auf die Anzahl der Messkanäle gewonnen werden. Es ist dabei Konvention den Wert der reduzierten Likelihoodfunktion $rlogl$ als

$$rlogl = \frac{-\log \mathcal{L}}{N_{Ch} - 5} \quad (3.2.4)$$

zu definieren [Sch13].

3.2.3. Energierekonstruktion

Die Energie ist eines der wichtigsten zu rekonstruierenden Merkmale. Im Folgenden wird nur auf die Energierekonstruktion von Myon-Neutrinos eingegangen und nicht auf die Rekonstruktion von Elektron-Neutrinos. Details zur Energierekonstruktion von Elektron-Neutrinos sind unter anderem in [M⁺09] zu finden.

Da die Energie des primären Neutrinos in der Regel für Myonen nicht zugänglich ist, schätzen alle Energieschätzer in IceCube die Energie des Myons. Aus Konvention wird die Energie geschätzt, welche das Myon am Punkt der Spur hatte, die dem Detektorzentrum am nächsten liegt.

Die Grundlage der Energieschätzer basiert darauf, dass der Energieverlust pro Strecke $\frac{dE}{dX}$ näherungsweise eine lineare Abhängigkeit zur Energie besitzt und durch

$$\frac{dE}{dX} = a + bE \quad (3.2.5)$$

gegeben ist. Die Konstanten a und b sind aus numerischen Rechnungen bestimmbar [KFS⁺13].

Es gibt zwei verschiedene Algorithmen zur Rekonstruktion der Energie. Der MuEX genannte Algorithmus benutzt eine diffuse Approximation für das räumliche und zeitliche Verhalten der Tscherenkowphotonen. Hierbei werden die optischen Eigenschaften des Eises berücksichtigt [Sch13] [A⁺14a].

Ein weiterer Ansatz ist es, die Anzahl der gemessenen Photoelektronen in jedem DOM zu nutzen. Diese Anzahl wird mit der Vorhersage unter der Bedingung einer Energie und einer Spurrekonstruktion verglichen. Die wahrscheinlichste Energie wird als Rekonstruktion genutzt. Eine solcher Ansatz ist in TruncatedEnergy realisiert. Da stochastische Energieverluste die Energieauflösung verringern, wurden jene bei der Berechnung von TruncatedEnergy ausgenommen. [A⁺14a, A⁺13d]

3.3. Trennung von Signal und Untergrund

Da in der (Astro)Teilchenphysik meist immer seltenere Teilchen und Zerfälle untersucht werden, wird eine effektive Trennung von Signal und Untergrund wichtiger. Während in der Vergangenheit einfache Schnitte zur Unterdrückung des Untergrundes genutzt wurden, werden die Methoden heutzutage deutlich differenzierter. Methoden des maschinellen Lernens treten mehr in den Fokus, da diese eine effiziente Klassifizierung von Daten in hochdimensionalen Räumen ermöglichen.

Viele Begrifflichkeiten, welche benutzt werden, stammen aus der Informatik. Diese Begriffe werden zunächst eingeführt. Bei einer Selektion soll einem Ereignis anhand seiner Merkmale eine Klasse zugeordnet werden. Um diese Art von Problem zu lösen, gibt es eine breite Auswahl an Klassifikationsalgorithmen.

Eine große Gruppe von Lernalgorithmen sind die sogenannten *baumartigen* Algorithmen, die auf Entscheidungsbäumen (englisch *decision trees*) basieren. Daher wird in Abschnitt 3.3.2 zunächst der Entscheidungsbaum vorgestellt, um anschließend in Abschnitt 3.3.3 den in der Analyse genutzten Random Forest zu erläutern. Im Speziellen wird in Kapitel 3.3.5 auf die Validierung eingegangen, die es ermöglicht, die Varianz des genutzten Modells zu schätzen.

3.3.1. Begrifflichkeiten

In der modernen Datenanalyse werden viele Begriffe benutzt. Diese werden im Folgenden kurz erläutert.

Merkmal

Ein Ereignis, welches im Detektor registriert wird, besteht zunächst aus 5160 Zeitreihen, in denen die Spannung, die der Photomultiplier detektiert, gegen die Zeit aufgetragen wird. Diese Zeitreihen werden bereits im Eis digitalisiert und zum Datenzentrum weitergeleitet. Anschließend werden diese Zeitreihen in Ankunftszeiten und Photoelektronen im DOM umgerechnet. Aus diesen Werten werden daraufhin viele weitere Werte berechnet. Ein Wertetyp, also z.B. die Summe der detektierten Photonen wird in dieser Arbeit als Merkmal bezeichnet. Andere übliche Bezeichnungen sind Observable, Attribut oder auch Feature.

3. Methoden der Datenanalyse

Klasse

Bei einer Separation sollen mehrere Klassen von Ereignissen voneinander getrennt werden. Eine Klasse bezeichnet einen Typ von Ereignissen. In der hier dargestellten Analyse wird eine binomiale Klassifikation durchgeführt. Das heißt, dass es zwei Klassen, ν_μ und μ , gibt. Diese werden als Signal und Untergrund bezeichnet.

Überwachtes Lernen

Überwachtes Lernen (*englisch: supervised learning*) ist eines der Teilgebiete des maschinellen Lernens. Dem Klassifikationsalgorithmus steht beim überwachten Lernen im Trainingsdatensatz die Wahrheit zur Verfügung. Diese Wahrheit kann zum Beispiel aus Expertenwissen oder aus Simulationen stammen. In dem vorliegenden Fall sind die Trainingsdaten Simulationsdaten.

Modell

Wird in einer Analyse ein überwachtes Lernverfahren benutzt, so generiert der Algorithmus eine Handlungsvorschrift, nach der unbekannte Ereignisse behandelt werden. Diese Handlungsvorschrift wird auch Modell genannt.

Confidence

Viele Lernverfahren bestimmen nicht nur die Zugehörigkeit zu einer Klasse, sondern geben dazu auch eine Konfidenz für die Zugehörigkeit an. Dieser Wert ist üblicherweise zwischen 0 und 1 und wird entweder als confidence oder im binominalen Fall auch als signalness bzw. backgroundness bezeichnet.

Reinheit

Die Reinheit ist eine Größe zur Bestimmung der Güte einer Klassifikation. Sie ergibt sich aus

$$R = \frac{\text{Anzahl Signalereignisse im resultierenden Datensatz}}{\text{Anzahl aller Ereignisse im resultierenden Datensatz}} \quad (3.3.1)$$

und ist damit ein Maß für die Verunreinigung des Datensatzes durch Untergrund.

Effizienz

Die Effizienz ist ein weiteres Merkmal zur Bestimmung der Güte eines Klassifikators. Die Effizienz ist über

$$E = \frac{\text{Anzahl gefundener Signalereignisse im Datensatz}}{\text{Anzahl aller Signalereignisse im Datensatz}} \quad (3.3.2)$$

definiert.

3.3.2. Entscheidungsbäume

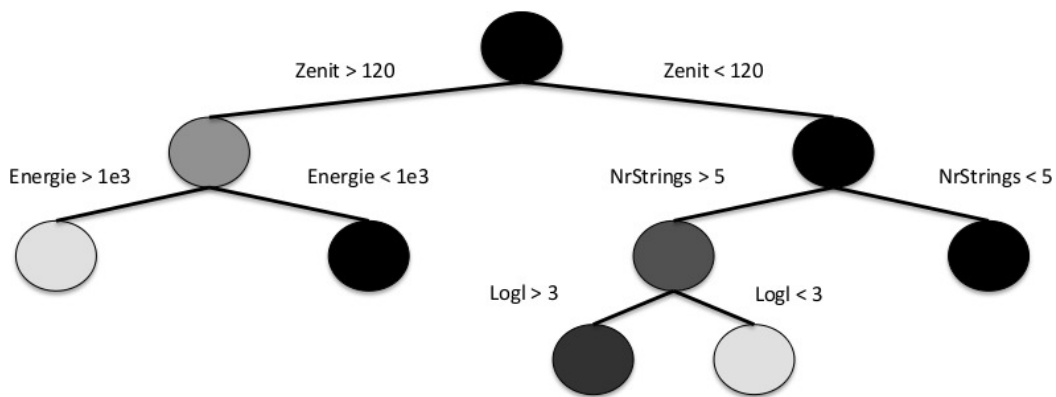


Abbildung 3.2.: Beispiel eines Entscheidungsbaumes, der auf Daten des IceCube-Detektors entstehen könnte. Die Farbe kodiert die Reinheit an einer Schnittstelle.

Zu Beginn einer Klassifikation wird das Merkmal und der Wert, bei dem die Trennung optimal ist, gesucht. Um die Güte der Trennung beurteilen zu können, können verschiedene Maße benutzt werden. Ein häufig genutztes Maß ist der Gini-Index [Gin12] oder die Kreuzentropie [BP96]. Durch Ausführen des so gefundenen optimalen Schnitts entstehen zwei Unterräume, die ebenfalls nach obigem Vorgehen separiert werden. Dieses Vorgehen wird rekursiv angewendet, bis die Daten nicht mehr trennbar sind. Die Enden werden auch Blätter (englisch: *leaves*) genannt. Die Punkte, an denen geschnitten wird, werden Knoten (englisch: *knot*) genannt.

Bekannte Einschränkungen von Entscheidungsbäumen

Entscheidungsbäume haben bekannterweise das Problem, dass sie übertrainiert werden können. Das bedeutet, dass sich der Lerner zu stark an die Trainingsdaten anpasst. Der Lerner passt sich an die Spezifika z.B. das zufällige Rauschen von Merkmalen an und ist somit nicht auf jedem Testdatensatz gleich gut. Um diesen Effekt zu unterdrücken, wird oftmals *pruning* genutzt. Bei dieser Technik werden die Bäume soweit zurückgeschnitten, sodass sie nicht mehr bis zum Ende ausgebaut werden. Hierdurch wird der Effekt reduziert, es entsteht jedoch eine gewisse Unschärfe [JDM00].

Gegenüber anderen Lernalgorithmen besitzen Entscheidungsbäume die Einschränkung, dass sie in jedem Knoten nur in einem Merkmal schneiden. Dadurch schneiden sie im Raum der Merkmale nur parallel zu einer Achse. So sind bestimmte Muster schwerer zu erkennen als mit anderen Algorithmen. Eine geeignete Datenaufbereitung kann dieses Problem reduzieren.

3.3.3. Random Forests

Ein Random Forest ist ein Ensemble von Entscheidungsbäumen. Um den oben genannten Effekt des Übertrainierens entgegen zu wirken, wird eine Menge i von Entscheidungsbäumen, die sich in ihrem Aufbau unterscheiden, genutzt. Es gibt verschiedene Möglichkeiten diese Unterscheidungen zu realisieren. Die am häufigsten genutzte ist die Realisierung nach [Bre01]. Jeder Baum innerhalb des Random Forests wird nur auf einer mit dem Bootstrap-Verfahren (vgl. Kapitel 3.3.5) ausgewählten Teilmenge der gesamten Ereignisse erstellt. Zusätzlich wird an jedem Knoten nur der beste Schnitt in k zufällig ausgesuchten Merkmalen gesucht. Mit Hilfe dieser Methode unterscheiden sich die Bäume in ihrem Aufbau. Nach Mittelung über die Entscheidungen der einzelnen Bäume ergibt sich die sogenannte confidence (bzw. signalness), die über

$$c = \frac{1}{N} \sum_{i=0}^N P_i, \quad P_i = \{0, 1\} \quad (3.3.3)$$

definiert ist. Hierbei ist P_i die Entscheidung über die Zugehörigkeit zu einer Klasse und N die Gesamtzahl der genutzten Bäume. In vielen Anwendungen wird ein zu klassifizierendes Ereignis als einer Klasse zugehörig klassifiziert, wenn die confidence größer als 0.5 ist, also die Mehrheit der Bäume für ein Signal spricht. Da

in dem Anwendungsfall der Neutrinoastronomie jedoch sehr ungleich populierte Klassen vorliegen und die Ansprüche an die Reinheit des Signals hoch sind, wird dieser Wert üblicherweise verändert.

3.3.4. Merkmalselektion

Die typische Anzahlen von Merkmalen in einem Datensatz der Astroteilchenphysik sind größer als 100. Allerdings enthält der Datensatz auch Merkmale, die keinerlei Information darüber tragen, zu welcher Klasse das Ereignis gehört. Ein gutes Beispiel ist die Detektionszeit des Teilchens. Andererseits gibt es Merkmale, die sich so sehr ähneln, dass die Information auch in nur einer Zahl ausgedrückt werden kann. Ein Beispiel sind verschiedene Schätzer eines Winkels. Im Hinblick auf Rechenzeit und die Effizienz des Lernalgorithmus ist es daher erforderlich die Anzahl der Merkmale zu reduzieren.

Für diese Aufgabe gibt es dedizierte Algorithmen. Der in dieser Analyse genutzte Algorithmus ist der *minimum redundancy maximum relevance* (kurz MRMR) [DP05, SM10] Algorithmus. In diesem Algorithmus wird iterativ ein Merkmal x zu den gewählten Merkmalen hinzugefügt. Das Merkmal wird so gewählt, dass

$$Q_j(x) = I(x, y) - \frac{1}{j} \sum_{x' \in F_j} R(x', x) \quad (3.3.4)$$

maximal ist. Hierbei ist $I(x, y)$ die Relevanz zum Label und $R(x', x)$ die Redundanz zu einem bereits gewählten Merkmal in der Menge der gewählten Merkmale F_j . Die Anzahl der bereits gewählten Merkmale wird mit j bezeichnet, während das Zielmerkmal (Signal oder Untergrund) mit y bezeichnet wird. $I(x, y)$ und $R(x', x)$ können mit der Transinformation [Bat94] (englisch: *mutual information*) berechnet werden. Diese ist definiert über

$$MI = \sum_{i,j} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i)P(y_j)}. \quad (3.3.5)$$

Die Wahrscheinlichkeitsdichtefunktionen werden mit P bezeichnet.

3.3.5. Validierung

Jedes gelernte Modell hat einen inhärenten Fehler auf seine Qualitätswerte wie Reinheit und Effizienz. Um eine Einschätzung dieser Fehler zu erlangen, ist es

3. Methoden der Datenanalyse

notwendig diese durch eine Validierung zu bestimmen. Im Fall von Übertrainieren, also dem übermäßigen Anpassen an den Trainingsdatensatz, wird dieser Fehler sehr groß. Übertraining ist ein unerwünschtes Verhalten, da sich das Modell an das zufällige Rauschen des Trainingsdatensatzes anpasst und es dadurch instabil wird. Um den Fehler des genutzten Modells zu bestimmen, gibt es im Allgemeinen zwei verschiedene Techniken. Beide werden im Folgenden erklärt.

Kreuzvalidierung

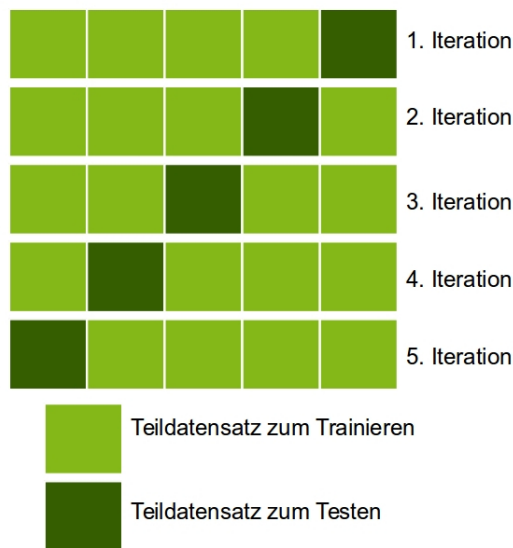


Abbildung 3.3.: Schematische Darstellung einer fünffachen Kreuzvalidierung. Der Datensatz wird in fünf disjunkte Teile geteilt. In jeder Iteration werden vier Teile zum Lernen genutzt und ein Teil zum Testen. Auf diese Weise werden die Qualitätsparameter jeweils fünfmal bestimmt. Aus diesen kann eine Abschätzung der Varianz erfolgen.

Bei einer Kreuzvalidierung werden die zur Verfügung stehenden Trainingsdaten zunächst in k disjunkte Teile aufgeteilt. Im Allgemeinen wird in diesem Fall einer k -fachen Kreuzvalidierung gesprochen. Der Extremfall $k = 2$ wird als *Splitvalidierung* und der Extremfall $k = n - 1$, wobei n die Anzahl der vorhandenen Ereignisse ist, wird als *Leave One Out Validierung* bezeichnet.

Um nun ein Modell zu validieren, wird die Methode auf $k - 1$ der k Teile trainiert. Der k -te Teil wird zum Testen des Modells genutzt. Anschließend wird diese Vor-

3. Methoden der Datenanalyse

gehensweise mit jedem Teil als Testdatensatz wiederholt werden. Auf diese Weise werden Qualitätsparameter wie Reinheit und Effizienz k -mal bestimmt. So ist eine Schätzung für den Mittelwert und die Varianz der Qualitätsparameter des Modells möglich [LM68].

Bootstrap-Validierung

Die Validierung mittels Bootstrapping unterscheidet sich von einer Kreuzvalidierung vor allem darin, dass die Datensätze nicht in jeder Iteration disjunkt sind. Stattdessen werden in jeder Iteration der Validierung k Ereignisse aus der Grundmenge von n Ereignissen gezogen. Es wird Ziehen mit Zurücklegen genutzt, so dass Ereignisse auch mehrfach in diesem Datensatz vorkommen können. Der so erzeugte Teildatensatz wird dann zum Trainieren eines Lernalgorithmus genutzt. Die nicht verwendeten Ereignisse werden zum Testen genutzt. Dieses Vorgehen kann mit unterschiedlichen Zufallszahlen für die Auswahl der Ereignisse wiederholt werden. Für jede Iteration des Algorithmus werden die Qualitätsparameter wie Reinheit oder Effizienz bestimmt. Aus der Verteilung der Qualitätsparameter lässt sich der Mittelwert und die Varianz bestimmen [Efr79].

Nutzung der Wahrscheinlichkeitsdichtefunktion zur Schätzung von Untergrundverteilungen

In der Praxis tritt in der Astroteilchenphysik häufig das Problem von begrenzten Simulationsdaten auf. Insbesondere die Untergrundsimulationen entsprechen einer Detektorlebenszeit von weniger als einem Monat. Der Grund ist der enorme Rechenaufwand bei der Simulation von Luftschauern sowie bei der Propagation der Photonen. Einzelne verbleibende Untergrundereignisse tragen dadurch ein hohes statistisches Gewicht. Dies führt zu zwei Problemen, welche gelöst werden müssen. Zum einem ist die relative Varianz der Qualitätsparameter sehr groß, da kleine Schwankungen zu großen relativen Varianzen führen. Zum anderen können die Eigenschaften des verbleibenden Untergrunds nur unzureichend durch die verbleibenden Untergrundereignisse beschrieben werden. Insbesondere in Merkmalen wie der Energie ist es wichtig nicht nur die Anzahl, sondern auch die Verteilung der verbleibenden Ereignisse schätzen zu können.

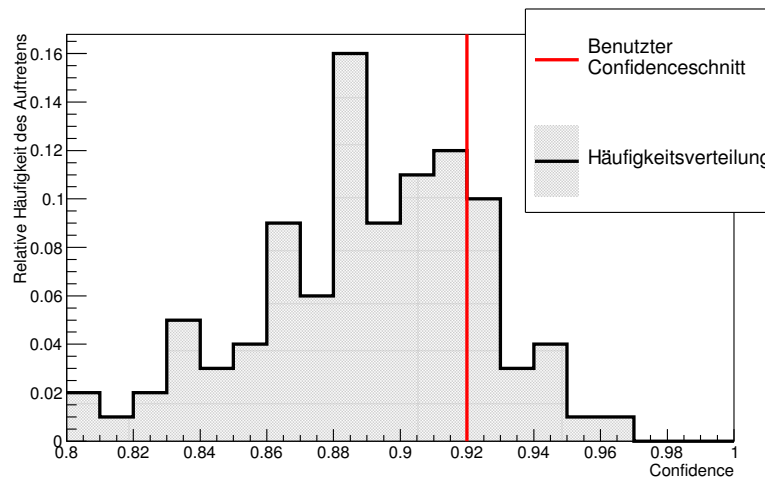


Abbildung 3.4.: Die Confidenceverteilung für ein Untergrundereignis, welches nur mit geringer Wahrscheinlichkeit als Signalereignis klassifiziert wird. Bei einer Kreuzvalidierung wäre dieses Ereignis mit einer Wahrscheinlichkeit von 19% als Signalereignis klassifiziert worden. In der Pull-Validierung wird dieses Ereignis mit einem Gewicht von 0.19 genutzt.

Mit Hilfe der Pull-Validierung ist dies möglich. Hierzu wird ausgenutzt, dass in einer Pull-Validierung jedes Ereignis mehrfach zum Testen genutzt wird. Jedem Ereignis kann auf diese Art und Weise an Stelle eines Confidencewerts eine Confidenceverteilung zugewiesen werden. Diese Verteilung kann als unverzerrter Schätzer der wahren Wahrscheinlichkeitsdichtefunktion des Ereignisses in der confidence interpretiert werden. Ein Beispiel für eine solche normierte Verteilung ist in Abbildung 3.4 dargestellt. Häufig ist die Wahrscheinlichkeit, dass ein Untergrundereignis nicht durch den Confidenceschnitt als Signalereignis klassifiziert wird, klein. In diesem Fall beträgt diese Wahrscheinlichkeit 19%. In einer Kreuzvalidierung wäre dieses Ereignis nur mit eben jener Wahrscheinlichkeit als Signalereignis klassifiziert worden. In der hier vorgestellten Methode wird das Ereignis als Untergrundereignis genutzt und ihm wird das Gewicht 0.19 zugewiesen. So wird mit jedem Ereignis verfahren. Die Anzahl der Ereignisse, die auf diese Art und Weise zur Untergrundverteilung im Signalbereich beitragen, kann durch diese Methode beträchtlich erhöht werden.

Auch die Qualitätsparameter werden auf diese Weise besser geschätzt als es mit einer Kreuzvalidierung möglich ist. In jeder Iteration werden $k < n$ der Ereignisse

zum Testen genutzt. Hierbei ist n die Gesamtzahl der Ereignisse. In jeder Iteration wird der gewünschte Qualitätsparameter bestimmt. Da dies häufig geschieht, ist somit eine genaue Schätzung des Mittelwertes und der Varianz möglich [Sch14].

3.3.6. Stabilität der Merkmalselektion

Genau wie die Qualitätsparameter der Separation muss auch eine Merkmalselektion auf ihre Stabilität hin geprüft werden. Es ist wichtig, dass die Auswahl der Merkmale nicht vom genutzten Datensatz abhängt. Der Datensatz wird in k disjunkte Teile aufgeteilt. Ähnlich zur Kreuzvalidierung werden die Merkmale auf jedem dieser k Teildatensätze bestimmt. Es entstehen somit k Listen von gewählten Merkmalen. Im Idealfall sind diese Listen identisch. In der Realität sind sie es häufig nicht. Maße für die Stabilität, also die Güte der Merkmalselektion ist z.B. der Jaccard-Index [Jac12]. Der Jaccard-Index ist über

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3.3.6)$$

definiert, wobei A und B die Mengen der in zwei Iterationen gewählten Merkmale sind. Der Wert ist auf den Wertebereich $[0, 1]$ definiert. Der Wert von 1 bezeichnet dabei eine maximal stabile Selektion.

3.4. Entfaltung von Verteilungen

Bei Messproblemen besteht häufig die Aufgabe, dass von der Wirkung auf die Ursache geschlossen werden soll. Dieses Problem wird als *inverses Problem* bezeichnet. Ein solches Problem liegt bei der Bestimmung von Energiespektren mit IceCube vor. In diesem konkreten Fall werden Merkmale gemessen, welche von der Energie abhängig sind, jedoch nicht die Energie des Teilchen selbst. Beispielsweise sind der rekonstruierte Energieverlust pro Meter, die Anzahl der getroffenen optischen Module oder auch die Länge der Spur im Detektor von der Energie des Primärteilchens abhängig.

Diese gemessenen Größen unterliegen einem gewissen Messfehler durch die begrenzte Auflösung des Detektors. Hinzu kommt, dass z.B. die Bremsstrahlungsverluste eines Myons eine hohe Stochastizität aufweisen. Dies führt zu weiteren Verschmierungen in den Observablen. Aus diesen Gründen ist eine exakte Rekonstruktion der Neutrinoenergie für jedes individuelle Teilchen nicht möglich. Um dennoch ein Spektrum der gesuchten Teilcheneigenschaft mit hinreichender Genauigkeit zu bestimmen, werden Entfaltungsalgorithmen genutzt. Diese Algorithmen bestimmen das Spektrum in dem gesuchten Merkmal. Bei dieser Technik tritt das einzelne Ereignis in den Hintergrund.

Um das oben skizzierte inverse Problem zu lösen, gibt es in der Astroteilchenphysik zwei populäre Ansätze. Diese Methoden der *Entfaltung* und der *Vorwärtsfaltung* unterscheiden sich inhärent und die Aussagen, die aus den Ergebnissen getroffen werden können, sind unterschiedlich.

Der Ansatz der Vorwärtsfaltung ist, dass die verschiedenen Komponenten einer Messung, abhängig von wenigen Parametern, modelliert und simuliert werden. Anschließend werden die Parameter jeder Komponente an die Daten angepasst. Bei einer Bestimmung des astrophysikalischen Myon-Neutrino-Flusses mittels Vorwärtsfaltung wären die Komponenten der konventionelle atmosphärische Myon-Neutrino-Fluss, der prompte atmosphärische Myon-Neutrino-Fluss, der astrophysikalische Myon-Neutrino-Fluss sowie gegebenenfalls der verbleibende Fluss durch atmosphärische Myonen.

Dadurch, dass die Theorie durch wenige Parameter modelliert wird, ist die Vorwärtsfaltung modellabhängig. Die Aussage einer Vorwärtsfaltung ist, dass die ermittelten Parameter im Rahmen der Modelle die Daten am besten beschreiben.

Ist das in der Natur realisierte Modell nicht in der Vorwärtsfaltung enthalten, so sind auch die ermittelten Parameter der anderen Komponenten beeinflusst und unter Umständen falsch. Die Aussage über einen Parameter ist somit nur eine bedingte Aussage. Sie bedingt nämlich, dass die Modelle die Natur hinreichend gut beschreiben.

Die Entfaltung hingegen ist modellunabhängig. Wie in diesem Abschnitt ersichtlich werden wird, ist für eine erfolgreiche Entfaltung nur die korrekt simulierte Abhängigkeit zwischen Observable und gesuchter Variable entscheidend.

In dieser Analyse wird das Entfaltungsproblem mit Hilfe der Entfaltungssoftware TRUEE [MDK⁺13] gelöst. TRUEE basiert auf dem RUN [Blo96] Algorithmus. Mit Hilfe der eingebaute Methoden zur Validierung, ist es möglich die gewählten Entfaltungseinstellungen auf ihre Stabilität und Güte zu überprüfen. Außerdem unterstützt TRUEE die Benutzung von bis zu drei Observablen, während andere Entfaltungsalgorithmen oft nur eine Observable unterstützen [Ady11].

Im Folgenden werden zunächst der Algorithmus, welcher in TRUEE genutzt wird, vorgestellt und die Handhabung von Regularisierungsproblemen dargestellt. Anschließend werden die Validierungsmodi erläutert.

3.4.1. Diskretisierung

Das Problem der Entfaltung lässt sich mathematisch durch eine Fredholmsche Integralgleichung erster Art beschreiben:

$$g(y) = \int_c^d A(y, x)f(x) dx + b(y). \quad (3.4.1)$$

Hier ist x die gesuchte und y die gemessene Größe. $A(x, y)$ ist die sogenannte Antwortfunktion des Detektors und $b(y)$ ist eine mögliche Untergrundverteilung der gemessenen Größe. Da diese Integralgleichung in der Regel nicht analytisch lösbar ist, ist es von Vorteil diese zu diskretisieren [BL98]. Die gesuchte Funktion $f(x)$ wird durch eine Linearkombination kubischer B-Splines [dB94] parametrisiert:

$$f(x) = \sum_{j=1}^n a_j p_j(x) \quad (3.4.2)$$

$p_j(x)$ sind die Basisfunktionen der B-Splines und a_j die jeweiligen Koeffizienten. Diese Parametrisierung erlaubt die Vereinfachung von Gleichung (3.4.1) über

$$\int_c^d A(y, x) f(x) dx = \sum_{j=1}^n a_j \int_c^d A(y, x) p_j(x) dx = \sum_{j=1}^n a_j A_j(y), \quad (3.4.3)$$

wobei $A_j(y)$ über

$$A_j(y) = \int_c^d A(y, x) p_j(x) dx \quad (3.4.4)$$

definiert ist.

Die Untergrundverteilung $b(y)$ und die Observablenverteilung $g(y)$ können durch Histogrammieren diskretisiert werden. Somit entstehen ein Observablen- und ein Untergrundvektor, welche über

$$g_i = \int_{y_{i-1}}^{y_i} g(y) dy \quad \text{bzw.} \quad b_i = \int_{y_{i-1}}^{y_i} b(y) dy \quad (3.4.5)$$

definiert sind.

Dies führt auf die Matrixgleichung

$$\mathbf{g} = \underline{\mathbf{A}} \mathbf{a} + \mathbf{b}. \quad (3.4.6)$$

In dieser Gleichung ist $\underline{\mathbf{A}}$ die sogenannte Antwortmatrix bzw. *response matrix* und \mathbf{a} ist der Vektor, der die Koeffizienten der B-Splines beinhaltet.

Im Folgenden wird der Untergrund \mathbf{b} vernachlässigt, da er in der vorliegenden Analyse durch die Signal-Untergrund-Trennung ausreichend unterdrückt wurde. Wenn dies nicht möglich ist, kann der Untergrund entweder über Simulationen oder über Kalibrationsmessungen bestimmt werden und anschließend kann

$$\mathbf{g}' = \underline{\mathbf{A}} \mathbf{a} \quad \text{mit} \quad \mathbf{g}' = \mathbf{g} - \mathbf{b} \quad (3.4.7)$$

mit denselben Methoden, wie im Fall ohne Untergrund, gelöst werden.

3.4.2. Unregularisierte Entfaltung

Übliche Lösungsmethoden für Gleichung (3.4.6) basieren entweder auf einem Kleinste-Quadrate- oder einem Log-Likelihood-Ansatz. Für den Log-Likelihood-Ansatz lautet die zu minimierende Funktion:

$$S(\mathbf{a}) = \sum_i (g_i(\mathbf{a}) - g_i^m \ln g_i(\mathbf{a})). \quad (3.4.8)$$

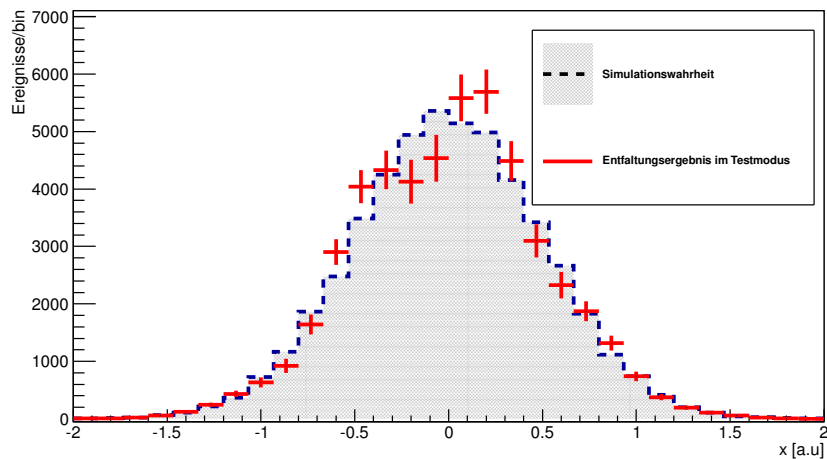


Abbildung 3.5.: Beispiel für eine minimal regularisierte Entfaltung mit TRUEE. Als Simulationswahrheit wurde eine Gaußverteilung mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = 0.5$ genutzt. Diese Simulationswahrheit ist in grau unterlegt. Die entfaltete Verteilung unterliegt einem stark schwankendem Verhalten.

Der Wert der gemessenen Observablen wird mit g_i^m bezeichnet und g_i ist der aus Antwortmatrix und B-Splinekoeffizienten berechnete Wert für die gemessene Observable. Bei einer solchen Lösungsmethode (und auch bei ähnlichen Ansätzen wie z.B. über die Invertierung der Antwortmatrix) ergeben sich oft unphysikalische Lösungen. Um dies zu verstehen, ist es hilfreich die Eigenschaften der Kovarianzmatrix der Lösung zu untersuchen.

Bestimmung der Kovarianzmatrix

Um für die gegebene Log-Likelihood-Funktion die Kovarianzmatrix zu bestimmen, kann die Funktion in zweiter Ordnung um das Minimum der Funktion entwickelt werden. Da um das Minimum \hat{a} entwickelt wird, ist die erste Ableitung der Funktion 0 und somit ergibt sich

$$F(\mathbf{a}) = F(\mathbf{a}_0) + \frac{1}{2} \sum_{i,k} G_{ik}(a_i - \hat{a}_i)(a_k - \hat{a}_k), \quad (3.4.9)$$

wobei G_{ik} gegeben ist durch

$$G_{ik} = \frac{\partial F}{\partial a_i \partial a_k} . \quad (3.4.10)$$

Wird die Log-Likelihood-Funktion betrachtet, ergibt sich in dieser Näherung eine Normalverteilung. Die Matrix $\underline{\mathbf{G}}$ ist die Hessematrix. Es zeigt sich, dass die Kovarianzmatrix $\underline{\mathbf{V}}$ gegeben ist durch

$$\underline{\mathbf{V}} = \underline{\mathbf{H}}^{-1} . \quad (3.4.11)$$

Zerlegung der Hessematrix

Die Hessematrix ist eine symmetrische, positiv semidefinite Matrix. Diese kann in

$$\underline{\mathbf{D}} = \underline{\mathbf{U}}^T \underline{\mathbf{H}} \underline{\mathbf{U}} \quad (3.4.12)$$

zerlegt werden. Hier sind die Diagonalelemente von $\underline{\mathbf{D}}$ die (reellen) Eigenwerte von $\underline{\mathbf{H}}$ und $\underline{\mathbf{U}}$ eine orthogonale Matrix mit der Eigenschaft

$$\underline{\mathbf{U}}^T \underline{\mathbf{U}} = \underline{\mathbf{I}} . \quad (3.4.13)$$

Wird diese Transformation genutzt und auf die Entwicklung zweiter Ordnung der Log-Likelihood-Funktion angewandt, folgt:

$$S(\mathbf{b}) = S_0 + \mathbf{b}^T \mathbf{h}_b + \frac{1}{2} \mathbf{b} \mathbf{b}^T , \quad (3.4.14)$$

wobei \mathbf{b} der transformierte Vektor zu \mathbf{a} gemäß

$$\mathbf{a} = \underline{\mathbf{U}} \underline{\mathbf{D}}^{\frac{1}{2}} \mathbf{b} \quad (3.4.15)$$

ist und \mathbf{h}_b über

$$\mathbf{h}_b = \underline{\mathbf{D}}^{\frac{1}{2}} \underline{\mathbf{U}}^T (-\underline{\mathbf{H}} \mathbf{a} + \mathbf{h}) \quad (3.4.16)$$

definiert ist.

In diesem Koordinatensystem ist die Lösung leicht ersichtlich. Das gesuchte Minimum \mathbf{b} ist gegeben durch

$$\mathbf{b} = -\mathbf{h}_b \quad (3.4.17)$$

und die Kovarianzmatrix zu \mathbf{b} ist die Einheitsmatrix. Wird nun \mathbf{b} geschätzt und zurücktransformiert, so folgt:

$$\hat{\mathbf{a}} = \underline{\mathbf{U}}\underline{\mathbf{D}}^{-\frac{1}{2}}\mathbf{b} = \sum_j \left(\frac{1}{D_{jj}}\right)^{\frac{1}{2}} \mathbf{b}_j \mathbf{u}_j. \quad (3.4.18)$$

Die gesuchte Funktion $f(x)$ kann nun alternativ auch durch die Koeffizienten \mathbf{b} und den Basisfunktionen $p'_j(x)$ ausgedrückt werden. Analog zu Gleichung (3.4.2) führt dies zu

$$f(x) = \sum_{j=1}^n b_j p'_j(x). \quad (3.4.19)$$

Wird diese Funktion für eine Entfaltung analysiert, so zeigt sich, dass die Schwankung für höhere Indizes j steigt. Beispiele für solche Basisfunktionen sind in Abbildung 3.6 zu sehen. Durch dieses Verhalten ist nun auch die nicht zufriedenstellende Lösung des inversen Problems durch einfache Invertierung erklärt. Die insignifikanten Beiträge, welche durch Eigenwerte mit hohem Index entstehen, führen zu unphysikalischen Oszillationen. Die einfache Lösung eines harten Abschneidens von Eigenwerten ab einem bestimmten Index j ist nicht zu empfehlen, da sonst das Gibbssche Phänomen auftritt [HH79]. Eine geschicktere Wahl der Unterdrückung, welche auch Regularisierung genannt wird, ist im nächsten Teil im Folgenden beschrieben.

3.4.3. Regularisierung

Um die erwähnten Oszillationseffekte zu kontrollieren, wird eine Regularisierungsmethode benutzt. Für die Regularisierung gibt es verschiedene Ansätze. Häufig wird in die Log-Likelihood-Funktion ein weiterer Term eingefügt, der fluktuierende Lösungen unterdrückt. Die in TRUEE genutzte Regularisierung ist die Tikhonov-Regularisierung [Tik43] mit der zweiten Ableitung. Dieser kann formuliert werden als

$$r(\mathbf{a}) = \frac{\tau}{2} \int (f''(x))^2 dx. \quad (3.4.20)$$

Der Parameter τ legt die Stärke der Regularisierung fest. Bei Benutzung von kubischen B-Splines mit äquidistanten Stützstellen vereinfacht sich dieser Term zu

$$r(\mathbf{a}) = \frac{\tau}{2} \underline{\mathbf{a}} \underline{\mathbf{C}} \mathbf{a}^T, \quad (3.4.21)$$

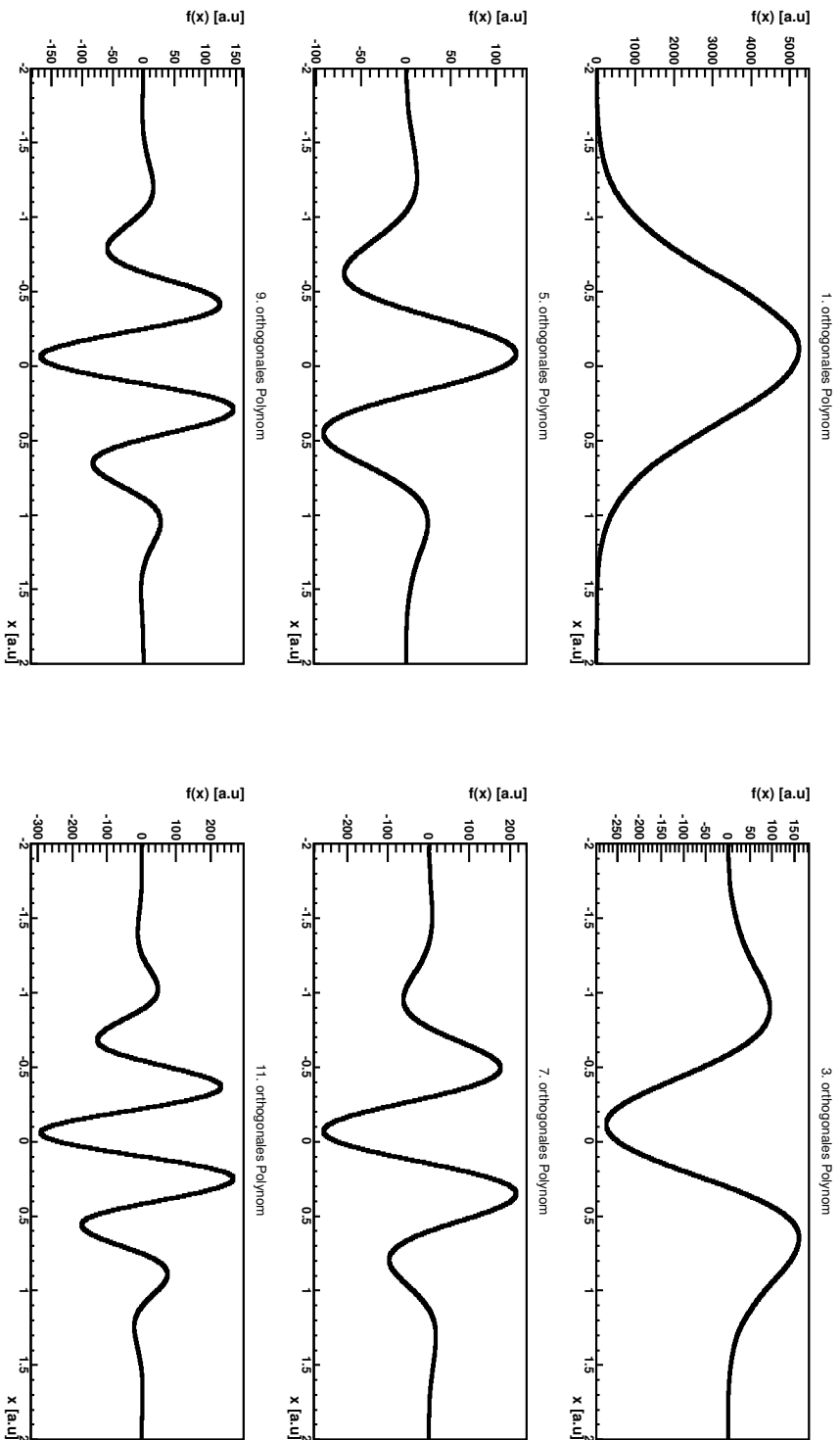


Abbildung 3.6.: Sechs verschiedene Basisfunktionen $p_j'(x)$ für die in Abbildung 3.5 gezeigte Entfaltung. Die kurvigen Beiträge zur Lösung sind die Moden mit hohem Index und daher kleinem Eigenwert. In der regularisierten Lösung werden die Moden mit hohem Index immer stärker unterdrückt.

wobei $\underline{\mathbf{C}}$ eine konstante, symmetrische, positivsemidefinite Matrix ist.

Die nun zu minimierende Log-Likelihood-Funktion lautet somit in parabolischer Näherung:

$$R(\mathbf{a}) = S(\tilde{\mathbf{a}}) + (\mathbf{a} - \tilde{\mathbf{a}})^T \mathbf{h} + \frac{1}{2} (\mathbf{a} - \tilde{\mathbf{a}})^T \underline{\mathbf{H}} (\mathbf{a} - \tilde{\mathbf{a}}) + \frac{\tau}{2} \mathbf{a} \underline{\mathbf{C}} \mathbf{a}^T \quad (3.4.22)$$

Um nun die Kovarianzmatrix zu analysieren werden $\underline{\mathbf{C}}$ und $\underline{\mathbf{H}}$ gleichzeitig diagonalisiert. Nach kurzer Rechnung [BL98] folgt für den Lösungsvektor der regularisierten Lösung \mathbf{c}

$$\mathbf{c} = (\underline{\mathbf{I}} + \tau \underline{\mathbf{S}})^{-1} \mathbf{b}, \quad (3.4.23)$$

wobei $\underline{\mathbf{S}}$ eine Diagonalmatrix ist, auf deren Diagonale die Eigenwerte von $\underline{\mathbf{C}}$ stehen und \mathbf{b} der unregularisierte Lösungsvektor ist. Für einen einzelnen Koeffizienten der Lösung folgt

$$c_j = \frac{1}{1 + \tau S_{jj}} b_j. \quad (3.4.24)$$

Somit erzeugt die Regularisierung ein allmähliches Abschneiden der Terme, die zu einer hohen Krümmung führen. Die Kovarianzmatrix der regularisierten Lösung ist nach einfacher Fehlerfortpflanzung gegeben durch

$$\underline{\mathbf{V}}[\mathbf{c}] = (\underline{\mathbf{I}} + \tau \underline{\mathbf{S}})^{-2}. \quad (3.4.25)$$

3.4.4. Modi in TRUEE

Bei einer Entfaltung müssen verschiedene Parameter optimiert werden. Dies sind

1. Auswahl der mit dem Zielmerkmal korrelierten Merkmale
2. Binning der korrelierten Merkmale
3. Binning des Zielmerkmals
4. Anzahl der Knoten des genutzten B-Splines
5. Stärke der Regularisierung.

Die optimale Wahl dieser Parameter ist a priori unklar. Gewünscht ist eine Entfaltung mit möglichst gutem Auflösungsvermögen und hoher Stabilität. Um dies zu gewährleisten, wird über verschiedene Möglichkeiten iteriert, um die gewünschte Einstellung zu finden. TRUEE bietet dazu zwei Möglichkeiten, welche im Folgenden erklärt werden.

Test-Mode

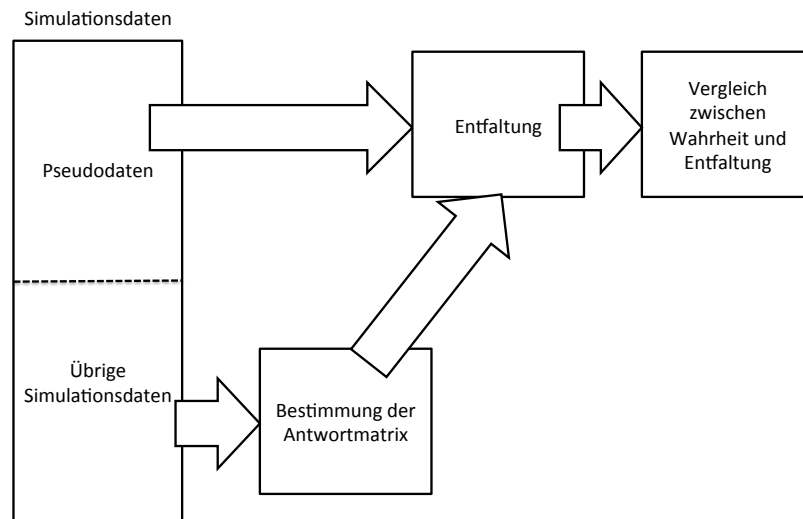


Abbildung 3.7.: Schematische Darstellung des Testmodes. Damit die Güte einer Entfaltungseinstellung überprüft werden kann, werden die vorhandenen Simulationsdaten geteilt. Dann wird der eine Teil als reguläre Simulationsdaten benutzt. Der andere hingegen dient als Pseudodaten und wird entfaltet. Anschließend kann das Ergebnis mit der Simulationswahrheit verglichen werden.

Um die Güte der Entfaltung für verschiedene Einstellungen beurteilen zu können, steht dem Nutzer der Test-Mode zur Verfügung, welcher in Abbildung 3.7 schematisch dargestellt ist.

Die vorhandenen Simulationsdaten werden aufgeteilt. Der eine Teil wird zur Bestimmung der Antwortmatrix genutzt. Der andere Teil wird als Pseudodaten behandelt und kann so entfaltet werden. Das bedeutet, dass im Unterschied zu echten Daten, für diese Pseudodaten die simulierten Wahrheiten über die gesuchte Verteilung bekannt sind. So kann die entfaltete Verteilung mit der wahren Verteilung mittels eines χ -Quadrat-Tests [Gag06], Kolmogorov-Smirnoff-Tests [MJ51] oder über ein Histogramm verglichen werden.

Pull-Mode

Der Pull-Mode ist die Erweiterung des Test-Modus, welcher die Stabilität der gewählten Einstellungen überprüft. Um dies zu erreichen, werden statt einer Entfaltung mehrere Entfaltungen (sogenannte *pulls*) durchgeführt. Dies führt zu N Entfaltungsergebnissen, die zur Analyse nutzbar sind. Dazu wird der Testwert

$$\frac{\text{Entfaltungsergebnis} - \text{Wahrheit}}{\text{Fehler}} \quad (3.4.26)$$

für jedes Bin in jeder Iteration definiert.

Dieser Wert kann für jedes Bin in ein Histogramm gefüllt werden. Nach N Wiederholungen entstehen genau so viele Histogramme wie Bins mit jeweils N Einträgen. Ist die Entfaltung erfolgreich verlaufen, so weist jedes dieser Histogramme eine Normalverteilung auf, da für große N der zentrale Grenzwertsatz gilt. In der Regel können jedoch leichte Abweichungen im Mittelwert dieser Verteilung festgestellt werden. Dies ist bedingt durch den Regularisierungsterm in Gleichung (3.4.20). Dieser Term verzerrt den Schätzer und führt daher zu systematischen Abweichungen. Diese Abweichungen sollten aber klein gegen den statistischen Fehler, und somit klein gegen 1, sein.

3.4.5. Umgewichten von Merkmalen

Eine herausragende Eigenschaft von TRUEE ist es, dass die Qualität der Entfaltung auch auf Daten überprüft werden kann. Nach der Entfaltung ist sowohl das Energiespektrum der Simulation als auch das Energiespektrum der Daten bekannt.

Mit dieser Information können nun die Simulationsdaten entsprechend der Datenverteilung umgewichtet werden. Anschließend kann mit Hilfe dieser Gewichte die Verteilung von anderen, an der Entfaltung nicht beteiligten, Merkmalen verglichen werden. Diese umgewichteten Verteilungen sollten, falls die Simulationsdaten die Eigenschaften des Detektors beschreiben, den Verteilungen der echten Daten entsprechen.

3.4.6. Bestimmung der x-Werte bei breiten Bins

Das Ergebnis einer Entfaltung mit TRUEE ist ein Histogramm mit Ereigniszahlen in definierten Bins. Sollen diese gebinnnten Werte mit einer kontinuierlichen Theorie verglichen werden, so ist es notwendig für jedes Bin einen entsprechenden x-Wert zu bestimmen, bei dem der Funktionswert verglichen werden soll.

Ein ähnliches Problem tritt auf, wenn das Ergebnis einer Entfaltung, wie in der Astroteilchenphysik üblich, nicht als Fluss, $\frac{d\Phi}{dE}$, sondern skaliert mit einer Potenz der Energie $E^y \frac{d\Phi}{dE}$, angegeben werden soll. Dadurch ist es notwendig den Energiewert genau zu bestimmen.

Nach [LW95] ist der x-Wert eines breiten Bins gegeben durch

$$f(x_{lw}) = \frac{1}{\Delta x} \int_{x_1}^{x_2} f(x) dx, \quad (3.4.27)$$

wobei f die Funktion ist, welche gebinnt wurde und $\Delta x = x_2 - x_1$ die Binbreite ist. Da aber f im Allgemeinen unbekannt ist, ist dieses Problem nicht direkt lösbar. In [LW95] wird vorgeschlagen eine Theoriefunktion zu nutzen, um damit die Funktion f zu nähern. In TRUEE gibt es die Möglichkeit die Spline-Funktionen als Approximation der wahren Funktion zu benutzen, um damit den x-Wert zu bestimmen. Durch die Lösung von

$$\hat{f}(x_{lw}) = \frac{1}{\Delta x} \int_{x_1}^{x_2} \hat{f}(x) dx \quad \text{mit} \quad \hat{f}(x) = \sum_j a_j p_j(x) \quad (3.4.28)$$

kann der x-Wert berechnet werden. Dieses Vorgehen hat im Vergleich zum Vorgehen nach [LW95] den Vorteil, dass es unabhängig von einer Theorieannahme ist.

Eine Bestimmung über eine Theoriefunktion würde zu systematischen Fehlern durch die Wahl der Theorie führen. Dies entfällt bei dem hier genutzten Verfahren.

4. Analyse der IC 79-Daten

In diesem Kapitel wird die Analyse des IC 79-Datensatzes diskutiert. Ziel dieser Analyse ist die Bestimmung des Energiespektrums von Myon-Neutrinos.

Das Kapitel gliedert sich in drei Teile. Im ersten Teil wird die Ereignis Selektion beschrieben, die einen hochreinen, aber trotzdem umfangreichen Datensatz erzeugt.

Die genutzten Methoden der Datenanalyse wurden im vorherigen Kapitel beschrieben. Anschließend wird die Entfaltung des Energiespektrums mit TRUÉE erläutert. Im dritten Teil dieses Kapitel wird das so vermessene Spektrum mit anderen Messdaten und theoretischen Vorhersagen verglichen.

4.1. Ereignis Selektion

Es gibt zwei verschiedene Techniken, um einen Großteil des myonischen Untergrundes zu reduzieren. Die eine Technik basiert darauf, dass nur im Detektor startende Ereignisse genutzt werden. Hierzu wird die äußere Schicht an DOMs als Veto genutzt. Dieser Schnitt wird auch *Containmentschnitt* genannt. Die andere basiert darauf, die Erde als Schild für Myonen zu nutzen. Diese Methode ist eher traditionell und wird im Folgenden als Schildtechnik bezeichnet.

Die Technik, welche startende Ereignisse benutzt, wurde in [A⁺13a] mit großem Erfolg genutzt. Mit dieser Methode ist es möglich IceCube als 4π Neutrinodetektor zu betreiben und in alle Richtungen sensitiv zu sein. Durch das sogenannte atmosphärische Selbstveto kann sowohl der konventionelle als auch der prompte Anteil von atmosphärischen Neutrinos weiter reduziert werden [GJKvS14]. Beim Selbstveto wird ausgenutzt, dass bei der Produktion der atmosphärischen Neutrinos auch Myonen entstehen. Diese können entweder direkt in dem Zerfall des Hadrons oder aber bei einem anderen Zerfall im selben Luftschauer entstehen.

4. Analyse der IC 79-Daten

Diese Myonen hinterlassen mit einer gewissen Wahrscheinlichkeit im Detektor ein Signal, welches zeitlich koinzident zum Neutrino-Signal ist. Somit schneidet der Containmentschnitt diese atmosphärischen Ereignisse ebenfalls aus dem Datensatz.

Um die astrophysikalische Komponente zu untersuchen, ist eine effektive Untergrundunterdrückung notwendig. Um dies zu erreichen, benötigt diese Analyse zusätzlich noch einen Schnitt in der deponierten Energie. Auf Grund dieses Schnittes und des Containmentschnittes ist das effektive Volumen dieser Analyse sehr gering. Entsprechend gering sind die Zählraten. In der neuesten Veröffentlichung mit drei Jahren von IceCube-Daten konnten 37 Ereignisse bei einer Schätzung von 8.4 ± 4.2 Myon- und $6.6^{+5.9}_{-1.6}$ atmosphärischen Neutrino-Untergrundereignissen detektiert werden [A⁺14b]. Dies führt zwar zu einer hohen Signifikanz, da aber die Ereignisse meist Kaskaden sind, ist eine Entdeckung von Quellen bisher nicht möglich [A⁺13a, A⁺14b].

Die Schildtechnik macht sich die Eigenschaft zu nutze, dass nur Neutrinos die Erde ungehindert durchqueren können. Detektierbare Teilchen, welche durch die Erde propagieren, sind also, wenn die Richtungsrekonstruktion fehlerfrei wäre, alles Neutrinos. Da die Rekonstruktion mit einem Fehler behaftet ist, tritt trotzdem ein myonischer Untergrund auf, der weiter unterdrückt werden muss.

Die Schildtechnik führt zu einer weniger starken Einschränkung des effektiven Volumens. Da keine Containmentschnitte gesetzt werden, können auch Myon-Neutrinos, welche mehrere Kilometer vom Detektor entfernt interagieren, detektiert werden. Durch den Schnitt im Zenitwinkel ist jedoch der sichtbare Bereich auf ca. 2π eingeschränkt. Trotzdem konnten bereits in der 59-String Konfiguration von IceCube ca. 28 000 Ereignisse gefunden werden [Ruh13]. Diese Daten liegen in einem Energiebereich von ~ 100 GeV bis hin zu 1 PeV. Daher ist mit dieser Methode neben einer Untersuchung der astrophysikalischen auch eine Untersuchung der atmosphärischen Komponente möglich.

Da, wie in Abbildung 4.1 ersichtlich, die Erde im Energiebereich > 100 TeV nicht mehr vollständig durchlässig für Neutrinos ist, ist bei diesen sehr hohen Energien der Winkelbereich mit der Schildtechnik kleiner als 2π . Die Schildtechnik alleine unterdrückt nicht den gesamten Untergrund aus atmosphärischen Myonen. Durch das Signal-Untergrund-Verhältnis sind die falsch rekonstruierten Myonen immer noch dominant gegenüber den Myon-Neutrinos (vgl. Abbildung 4.3(b)).

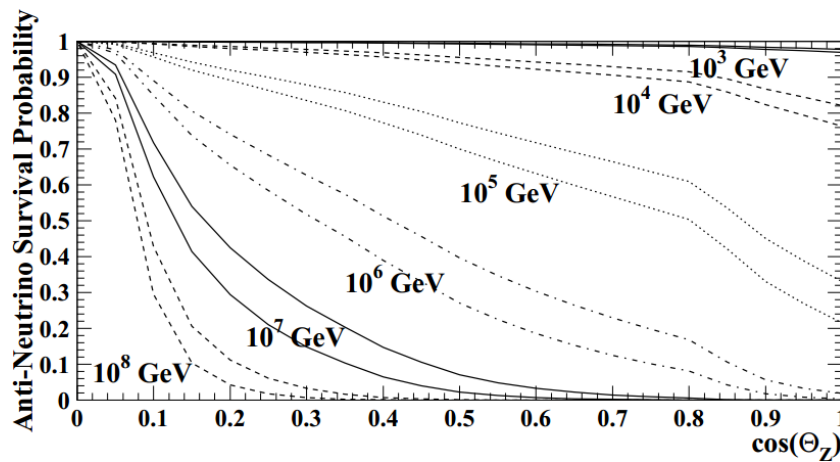


Abbildung 4.1.: Überlebenswahrscheinlichkeit für Anti-Neutrinos, die durch die Erde propagieren, mit verschiedenen Energien in Abhängigkeit des Zenitwinkels. Für hohe Winkel und somit lange Strecken ist die Erde nicht mehr durchlässig für Neutrinos. Die oberen Kurven sind jeweils nur für Interaktionen über den geladenen Strom (CC-Interaktionen); die unteren Kurven sind für Interaktionen über den geladenen als auch den ungeladenen Strom (CC- sowie NC-Interaktionen) [ALS02].

In dieser Arbeit wird die Schildtechnik mit Myon-Neutrinos genutzt. Viele der genutzten Methoden kam bereits in der IC 59-Analyse von Tim Ruhe [Ruh13] zum Einsatz. Die verwendete Ereignisselektion wurde in der Analyse von Florian Scheiriau [Sch14] für die Analyse der IC 86 Daten entwickelt und in den Details an den IC 79 Datensatz angepasst.

Die Ereignisselektion gliedert sich grob in drei Phasen. Zunächst werden sogenannte Precuts, also Schnitte zur Vorselektion, genutzt, um grobe Verunreinigungen durch Myonen und Elektron-Neutrinos zu beseitigen. Der resultierende Datensatz wird dann in der zweiten Phase mit Hilfe des MRMR-Algorithmus untersucht, um die besten Diskriminanzmerkmale zu finden. Diese werden dann in der dritten Phase genutzt, um mit Hilfe des Random Forests Signal und Untergrund zu trennen. Diese drei Phasen werden in den folgenden Abschnitten näher beschrieben.

4. Analyse der IC 79-Daten

Detektor	Winkelbereich
AMANDA	$100^\circ - 180^\circ$
IC 40	$97^\circ - 180^\circ$
ANTARES	$90^\circ - 180^\circ$
IC 59	$90^\circ - 180^\circ$
IC 79	$86^\circ - 180^\circ$
IC 86	$86^\circ - 180^\circ$

Tabelle 4.1.: Winkelbereiche für die Analysen von AMANDA [A⁺10b], IC 40 [A⁺11a], ANTARES [AM⁺12], IC 59 [Ruh13], IC 86 [Sch14] sowie in IC 79.

4.1.1. Precuts

Teil dieser Analyse sind zwei Precuts. Der erste Precut ist der bereits erwähnte Schnitt auf den rekonstruierten Zenitwinkel. Als Zenitwinkelrekonstruktion wird die Rekonstruktion aus dem MPE-Algorithmus genutzt, da dieser die beste Auflösung besitzt [A⁺14c]. Als Schnittwert wird 86° gewählt. Im Vergleich zu früheren Analysen, deren Winkelbereiche in Tabelle 4.1 aufgeführt sind, ist das ein sehr großer Bereich. Dieser Winkelbereich muss natürlich in einem Vergleich der Analysen berücksichtigt werden, da der Myon-Neutrino-Fluss, wie in Abschnitt 2.4.1 bereits erwähnt, winkelabhängig ist.

Der zweite genutzte Precut ist der Schnitt auf ν_{LF} , welcher in Abschnitt 3.2 beschrieben ist. Der gewählte Schnitt von > 0.1 , welcher in 4.3(a) dargestellt ist, schneidet einen Großteil der Elektron-Neutrinos aus dem Datensatz. Mit diesem Schnitt werden Ereignisse selektiert, die nicht kaskadenartig, sondern spurartig sind.

Der Recall für atmosphärische Myonen durch diese Schnitte beträgt

$$\text{Recall}(\mu) = (19.3 \pm 0.1) \% .$$

Der Recall für atmosphärische Myon-Neutrinos beträgt

$$\text{Recall}(\nu_\mu) = (71.4 \pm 0.1) \% .$$

Die Fehler wurden mittels der Bootstrapvalidierung bestimmt.

Neben der Vermessung des konventionellen atmosphärischen Myon-Neutrino-Flusses ist auch die Vermessung des prompten atmosphärischen Myon-Neutrino-Flusses

4. Analyse der IC 79-Daten

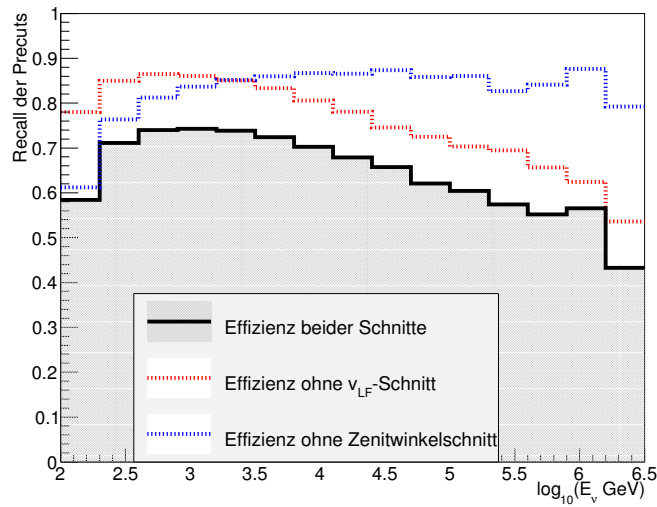
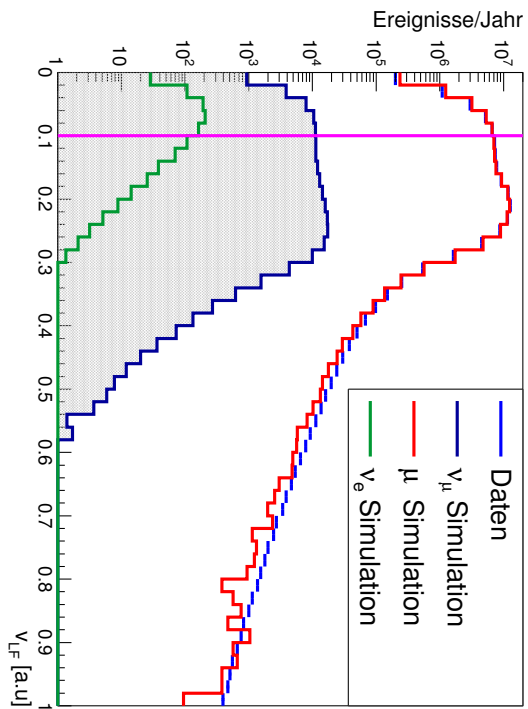
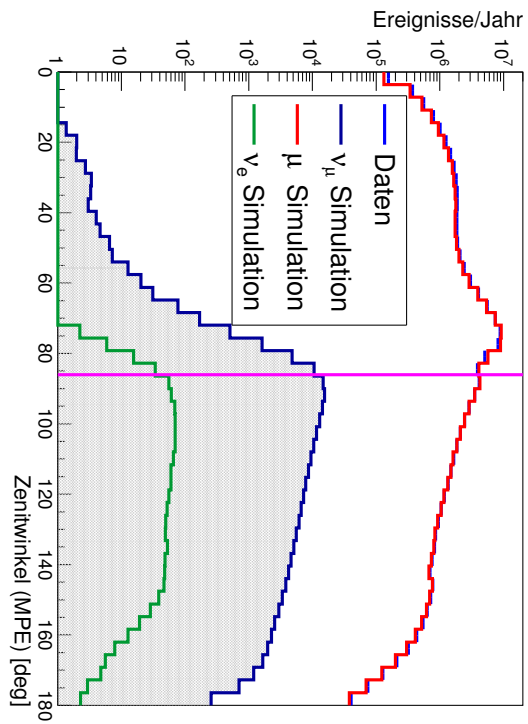


Abbildung 4.2.: Recall der kombinierten und einzelnen Precuts. Ein Großteil des Verlustes an Signalereignissen ist durch den Zenitcut zu erkennen. Dieser ist jedoch notwendig in einer solchen Analyse.

und des hochenergetischen astrophysikalischen Myon-Neutrino-Flusses Ziel dieser Arbeit. Entscheidend für eine Analyse der astrophysikalischen Komponente ist eine sehr hohe Signaleffizienz oberhalb von 100 TeV. Der energieaufgelöste Recall ist in Abbildung 4.2 zu sehen. Der Großteil an Signalverlust im Hochenergiebereich wird durch den Zenitschnitt hervorgerufen. Dieser ist jedoch in einer solchen Analyse unumgänglich. Der ν_{LF} -Schnitt schneidet nur wenige Signalereignisse aus dem Hochenergiebereich aus dem Datensatz. Die verbleibenden 50% Signalereignisse für die höchsten Energien sind daher als hoch anzusehen.



(a)



(b)

Abbildung 4.3.: Darstellung der gewählten Precuts. (a) Schnitt auf $v_{\mu F}$. Mit Hilfe dieses Schnittes wird der Untergrund durch Elektron-Neutrinos verringert. (b) Schnitt auf den vom MPE-Algorithmus rekonstruierten Zenitwinkel. Mit diesem Schnitt ist eine starke Reduktion des Untergrundes durch atmosphärische Myonen möglich. Die jeweils genutzten Schnitte sind in Violet eingzeichnet. Die Ereignisse die einen Wert besitzen, der kleiner als der gewählte Schrittwert ist, werden verworfen.

4.1.2. Merkmalauswahl

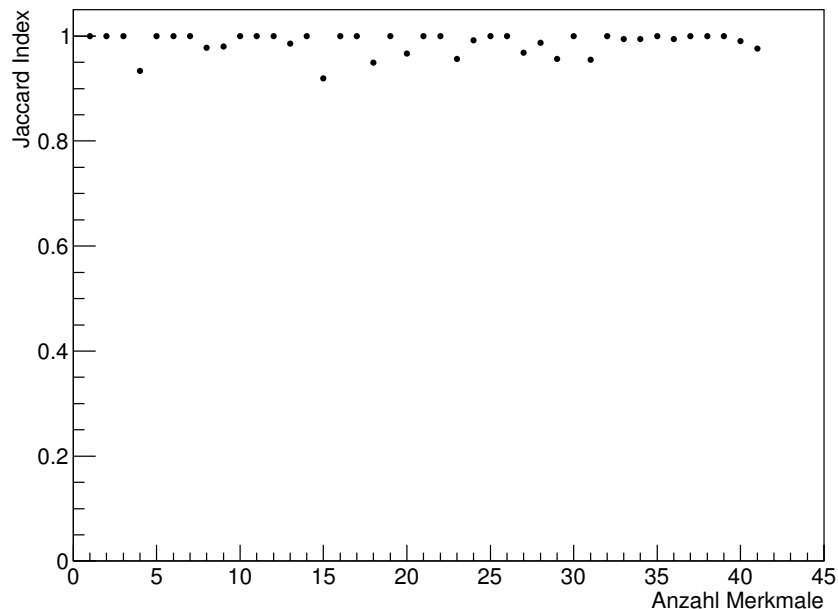


Abbildung 4.4.: Stabilität der Merkmalsselektion für verschiedene Anzahlen von gewünschten Merkmalen. Die Stabilität wird mit dem Jaccard-Index gemessen.

Um die geeigneten Merkmale zu finden, wird die MRMR-Methode [DP05], wie in Kapitel 3.3.4 beschrieben, genutzt. Zunächst werden verschiedene Anzahlen von gesuchten Merkmalen getestet und auf ihre Stabilität geprüft. Das Ergebnis dieser Überprüfung ist in Abbildung 4.4 zu sehen. Die Stabilität der Merkmalsselektion ist hoch. Es werden grundsätzlich also auch auf statistisch unabhängigen Datensätzen dieselben Merkmale gewählt. Um die genaue Anzahl der Merkmale zu wählen, wird die Wichtigkeit der Merkmale analysiert. Die Merkmals-Wichtigkeit ergibt sich aus dem Vergleich zwischen der Fehlklassifikation mit einem Datensatz mit dem unveränderten Merkmal und einem Datensatz, in dem das Merkmal verwechselt ist. Das Verwechseln wird hierbei mittels zufälliger Permutation innerhalb des Datensatzes erreicht [Bre01].

4. Analyse der IC 79-Daten

Merkmal	Merkmals-Wichtigkeit
L3Cut_PLogL	1.0
MPE_NDirC	0.57
Ztravel · MPE_NDirC	0.57
TruncatedMean_Bins_Res	0.48
MPE_NDirC · TruncatedMean_Doms_Res	0.47
MPE_ClosestAppCoG	0.47
MPE_LDirC	0.46
MPE_Zd – LineFit_Zd	0.30
Ztravel	0.17
MPE_ClosestApp	0.17
COGZ	0.16
NCh	0.12
TruncatedMean_Doms_Res	0.06
MPE_Zd	0.06
MPE_NDirB – MPE_NDirC	0.06
NString · cos(MPE_Zd)	0.05
Zsigma	0.05
radius	0.05
MPE_Sall	0.04
DeltaAzimuthMPELineFit	0.04
NString	0.04
SPE4_Bay_rlogl	0.04
SPE4_Bay_Zd	0.04
TruncatedMean_Bins_Res – TruncatedMean_Doms_Res	0.01
SPE4_Bay_Status	0.01

Tabelle 4.2.: Merkmals-Wichtigkeit für die gewählten Merkmale. Eine Merkmalsbeschreibung ist in Anhang A zu finden.

4. Analyse der IC 79-Daten

Die Merkmals-Wichtigkeit ist für 25 Merkmale in Tabelle 4.2 aufgeführt. Die Merkmals-Wichtigkeit ist auf 1 normiert. Mehr als 25 Merkmale sind nicht notwendig, da die Wichtigkeit dort auf Werte nahe 0 abfällt. Daher werden in dieser Separation 25 Merkmale genutzt.

Die gewählten Merkmale sind in Anhang A aufgeführt. Obwohl die Auswahl von einem Algorithmus getroffen wurde, sind die gewählten Merkmale nachvollziehbar. Im Allgemeinen lassen sie sich in vier Gruppen aufteilen.

Winkelschätzer

Winkelschätzer haben einen hohen Informationsgehalt. Umso näher ein Ereignis in der Nähe des Horizonts rekonstruiert wird, desto wahrscheinlicher ist es ein fehlrekonstruiertes Myon.

Energieabhängige Merkmale

Je mehr Energie ein Myon besitzt, desto eindeutiger ist das Signal im Detektor, da mehr Informationen vorliegen. Wenn ein Ereignis nur wenige DOMs getroffen hat, so ist der Informationsgehalt über das Ereignis gering. Aus diesem Grund werden energieabhängige Merkmale aufgenommen.

Gütekriterien der Rekonstruktion

Offensichtliche Qualitätskriterien der Rekonstruktion sind die Werte der reduzierten Log-Likelihood-Funktion, welche die Güte der Rekonstruktion beschreibt. Aber auch die Unterschiede in der Rekonstruktion durch unterschiedliche Algorithmen sind im Endeffekt ein Gütekriterium, da dieser Unterschied darauf hinweist, dass die Schätzer eine hohe Varianz beinhalten.

Ortsbeschreibungen

Der Untergrund dieser Messung – fehlrekonstruierte atmosphärische Myonen – besteht aus Ereignissen, welche *down-going* sind, also vom Detektorzentrum in Richtung Erdmittelpunkt propagieren. Daher neigen diese Ereignisse dazu im oberen Bereich des Detektors ihren Schwerpunkt zu haben. Auch Ereignisse, die am

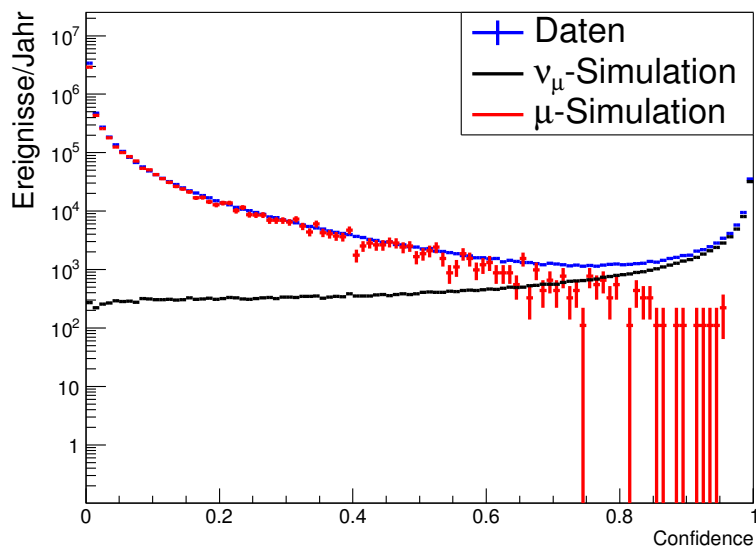
Rand des Detektors ihren Schwerpunkt besitzen, können myonischen Ursprungs sein. Diese sogenannten *Cornerclipper* sind schwerer zu rekonstruieren und haben daher eine größere Varianz in ihrer Winkelschätzung.

4.1.3. Selektion mit Hilfe des Random Forests

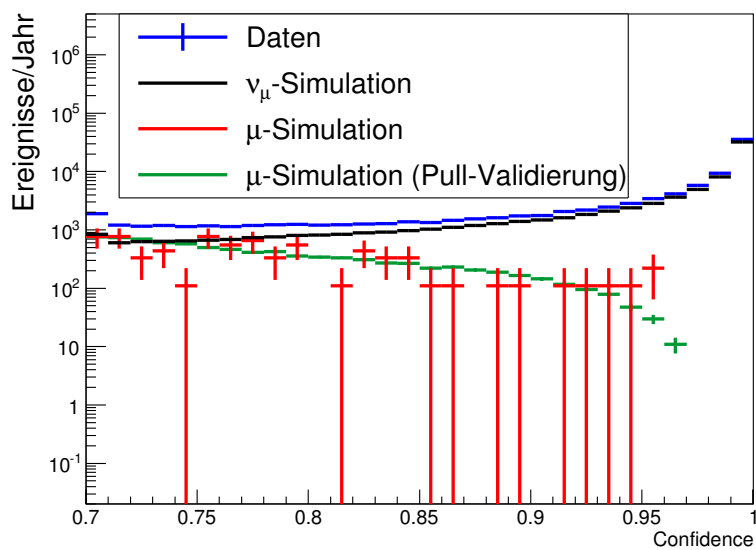
Der resultierende Datensatz wird mit Hilfe eines Random Forests, wie er in Kapitel 3.3 beschrieben ist, weiter verarbeitet. Es werden 250 Bäume und 5 zufällige Merkmale pro Knoten gewählt. Der Random Forest weist jedem Ereignis einen Wert zu, die sogenannte Confidence. Mit Hilfe dieser Methode wurde die Information aus den 25 Merkmalen auf eine Größe reduziert. Die Verteilung der Confidence ist in Abbildung 4.5 zu sehen. In Abbildung 4.5(a) ist die Confidence für den gesamten definierten Bereich zwischen 0 und 1 zu sehen. Dem Untergrund werden wie erwartet niedrige Werte zugewiesen. Der Modus der Verteilung liegt bei 0. Dem definierten Signal (Myon-Neutrinos), werden hohe Werte zugewiesen. Der Modus der Signalverteilung ist bei 1. Im ganzen definierten Bereich stimmen Simulationen und Daten überein.

In Abbildung 4.5(b) ist die Confidence für den Bereich zwischen 0.7 und 1 dargestellt. Dieser Bereich ist wichtig, da in diesem Bereich ein Schnitt angebracht wird, um zwischen Signal und Untergrund zu trennen. Auf Grund von limitierter Simulationsstatistik wird die Vorhersage von Myonen bei sehr hohen Confidence-Werten, wie in der Abbildung ersichtlich, schwierig, da die statistischen Schwankungen groß werden. Um eine bessere Vorhersage für den Untergrund zu treffen, wurde die Pull-Validierung genutzt. Diese Verteilung ist weniger von Schwankungen betroffen und liefert eine bessere Untergrundbeschreibung. Bei Schnitten oberhalb von 0.9 können Datensätze erreicht werden, die eine Reinheit höher als 99% besitzen. Diese Datensätze können als näherungsweise rein betrachtet werden.

4. Analyse der IC 79-Daten



(a)



(b)

Abbildung 4.5.: Confidence-Verteilung, welche der Random Forest erzeugt. a) Verteilung für den gesamten definierten Bereich der Confidence. b) Vergrößerung auf den Bereich [0.7, 1] mit Hinzunahme der Ergebnisse der Pull-Validierung.

4. Analyse der IC 79-Daten

Für diese Analyse wurde ein Confidenceschnitt von 0.92 gewählt. Nach Anwendung dieses Schnittes wird eine Verteilung in der rekonstruierten Spurlänge (LDirC) erreicht, wie sie in Abbildung 4.6 zu sehen ist. Wie dort ersichtlich ist, ist bei Spurlängen von unter 200 m eine schlechte Übereinstimmung von Simulationen und Daten zu beobachten. Diese Längen sind für IceCube sehr klein, da der Abstand zwischen den Strings 125 m beträgt. Bei solch kleinen Längen können Effekte wie das simulierte Rauschen eine große Rolle spielen.

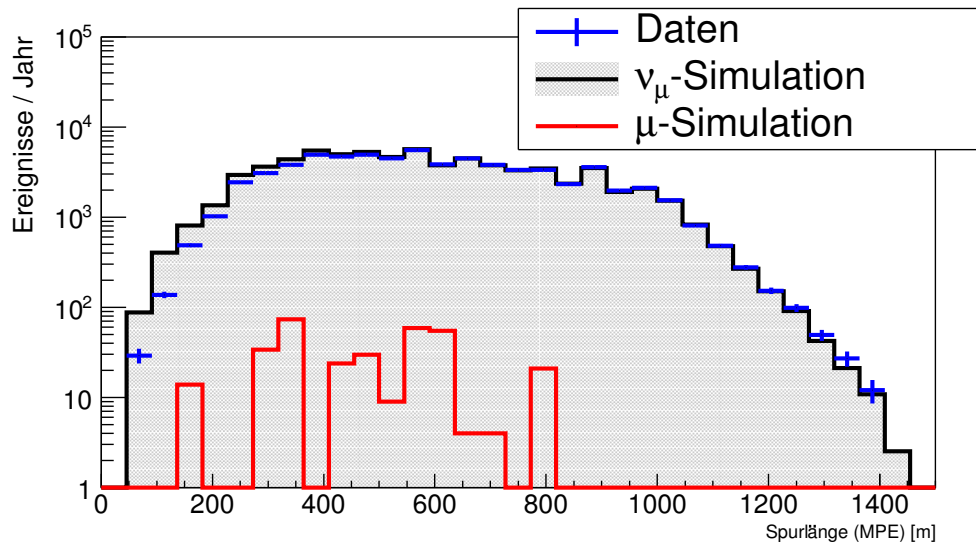


Abbildung 4.6.: Rekonstruierte Spurlänge nach einem Schnitt auf eine Confidence von 0.92. Bei Spurlängen von unter 200 m ist eine Unstimmigkeit zwischen Daten und Simulation zu sehen. Der myonische Untergrund wird mit Hilfe der Pull-Validierung abgeschätzt.

Um sicher zu gehen, dass dieser Unterschied zwischen Simulationen und Daten in der späteren Entfaltung keine Rolle spielt, wird ein zusätzlicher Schnitt auf die rekonstruierte Spurlänge von 200 m angewendet. Die Reinheiten und Ereigniszahlen für den kombinierten Schnitt sind in Tabelle 4.3 zu sehen.

In Abbildung 4.7 ist der Recall für Myonneutrinos für verschiedene Confidence-Schnitte unter Berücksichtigung des Spurlängenschnitts in Abhängigkeit von der Energie aufgetragen. Ein Confidence-Schnitt nahe bei eins reduziert die Anzahl der Ereignisse drastisch. Je lockerer der Schnitt, desto mehr Untergrundereig-

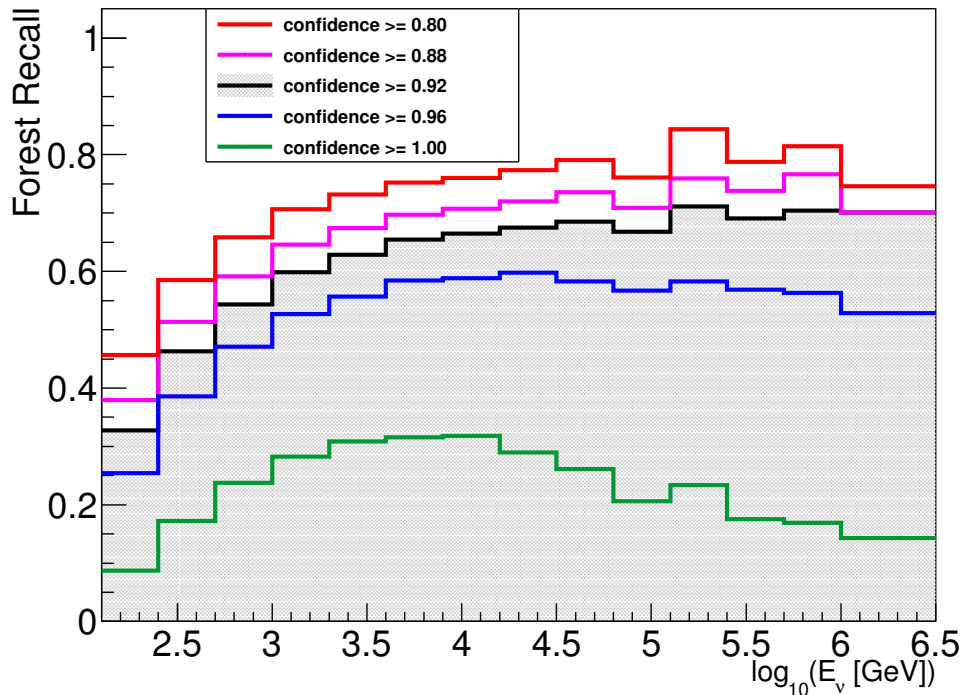


Abbildung 4.7.: Recall des Random Forests für Myon-Neutrinos für verschiedene Confidence-Schnitte. Für Schnitte nahe 1 ist der Recall gering. Der gewählte Schnitt von 0.92 ist grau hinterlegt. Dieser Schnitt ist ein guter Kompromiss zwischen Reinheit und Effizienz.

nisse werden zugelassen. Für diese Analyse wurde ein Confidence-Schnitt von 0.92 gewählt, da dieser einen guten Kompromiss zwischen Reinheit und Effizienz erreicht. Bei diesem Schnitt können 66 885 Ereignisse erfasst werden.

Wie in Tabelle 4.3 zu sehen, hat die DOM-Sensitivität einen starken Einfluss auf die Normierung des Flusses. Die DOM-Sensitivität ist die relative Photon-Detektionseffizienz der Photomultiplier innerhalb der Simulation.

Die korrekte Wahl der DOM-Sensitivität ist nicht durch die absolute vorhergesagte Anzahl an Neutrinos gegeben. Die Normierung hängt unter anderem stark von der Normierung des primären Flusses ab, welche einen großen Fehler hat. Viel wichtiger ist die Übereinstimmung in den Verteilungen von energieabhängigen Merkmalen (vgl. z.B. Abbildung 4.8). Hier ist eine bessere Übereinstimmung

4. Analyse der IC 79-Daten

für den Datensatz mit 110% zu beobachten. Aus diesem Grunde wird der 110%-Datensatz als Datensatz zur Bestimmung der Antwortmatrix in der Entfaltung genutzt. Der Datensatz mit 100% sowie ein Datensatz mit 120% DOM-Sensitivität wird als Referenzdatensatz zur Bestimmung von systematischen Fehlern genutzt. Die Untersuchung von systematischen Fehlern ist in Abschnitt 4.2.4 zu finden.

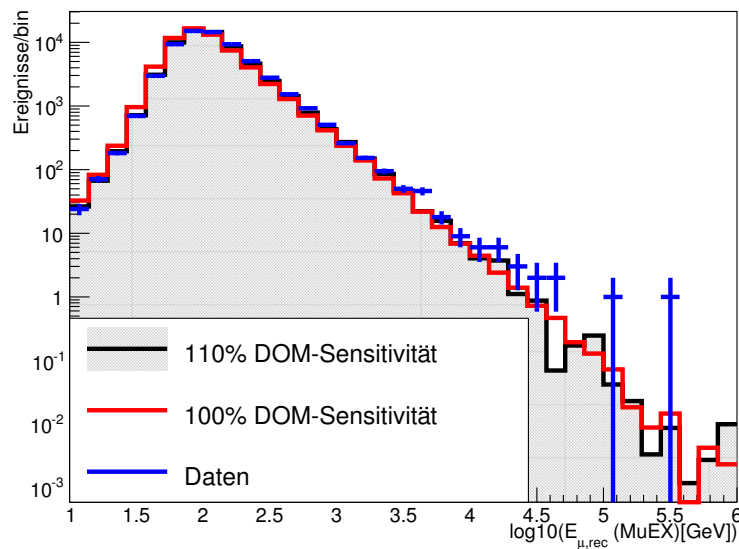


Abbildung 4.8.: Energieschätzung des Myons mit dem MuEX-Algorithmus für zwei verschiedene Einstellungen der DOM-Sensitivität. Es ist ein klarer Versatz zu sehen. Dies ist bedingt dadurch, dass die Anzahl der detektierten Photonen durch eine Steigerung der DOM-Sensitivität steigt und somit sich die zuzuordnende Energie ändert.

Confidence- schnitt	Detektierte Ereignisse	Vorhergesagte		Atmosphärische Myonen		Reinheit
		Ereignisse (100%)	Ereignisse (110%)	Ereignisse (110%)	Myonen	
1.00	26896	23630 ± 865	33230 ± 1213	0 ⁺²⁸ 0 ⁻⁰	1.000 ^{+0.000} -0.001	
0.99	41076	36860 ± 686	50306 ± 987	0 ⁺²⁸ 0 ⁻⁰	1.000 ^{+0.000} -0.001	
0.98	48071	43611 ± 668	58787 ± 893	0 ⁺²⁸ 0 ⁻⁰	1.000 ^{+0.000} -0.001	
0.97	52970	48360 ± 611	64749 ± 809	9 ⁺²⁸ -9	1.000 ^{+0.000} -0.001	
0.96	56787	52111 ± 583	69375 ± 771	19 ⁺⁵⁶ -19	1.000 ^{+0.000} -0.001	
0.95	59854	55214 ± 545	73136 ± 715	56 ⁺⁸⁵ -75	0.999 ^{+0.0001} -0.002	
0.94	61204	56597 ± 536	76437 ± 715	94 ⁺¹⁰³ -94	0.999 ± 0.002	
0.93	64763	60254 ± 508	79295 ± 658	216 ± 150	0.997 ± 0.002	
0.92	66885	62370 ± 480	81900 ± 630	310 ± 188	0.995 ± 0.003	
0.91	68824	64279 ± 470	84194 ± 611	442 ± 216	0.930 ± 0.003	
0.90	70584	65172 ± 461	86385 ± 386	592 ± 244	0.991 ± 0.004	

Tabelle 4.3.: Vorhergesagte und detektierte Ereignisse für verschiedene Confidenceschnitte sowie die resultierende Reinheit. Die Vorher-
sage für Myon-Neutrinos im Datensatz ist für zwei verschiedene DOM-Sensitivitäten angegeben. Die angegebenen Fehler
ergeben sich aus einer 150-fachen Pull-Validierung.

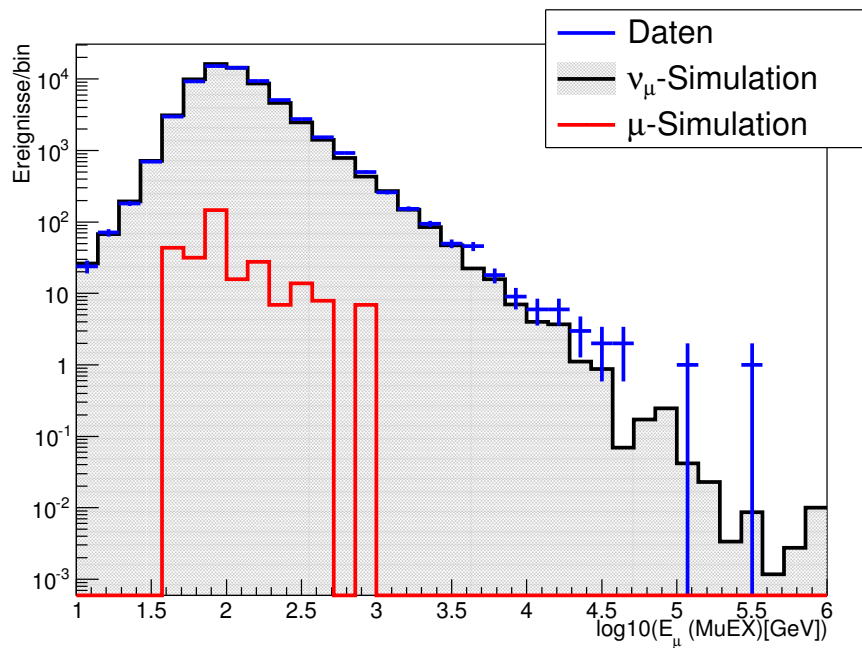
4.1.4. Beschreibung des verbleibenden Untergrunds

Mit Hilfe der Pull-Validierung (vgl. Abschnitt 3.3.5) ist es möglich die Untergrundverteilungen trotz geringen Stichprobenumfangs der Simulationsdaten genauer als mit herkömmlichen Methoden, wie zum Beispiel einer Kreuzvalidierung, zu beschreiben. Wichtige Merkmale sind die energieabhängigen Merkmale, welche auch in der Entfaltung genutzt werden. Diese und der Zenitwinkel aus dem MPE-Algorithmus sind in Abbildung 4.9 sowie in Abbildung 4.10 zu sehen. Zum einen ist eine sehr gute Übereinstimmung zwischen den Daten und der Simulation zu beobachten. Zum anderen liegt der verbleibende myonische Untergrund nur bei niedrigen Energien. Daher wird er die Entfaltung nur vernachlässigbar beeinflussen.

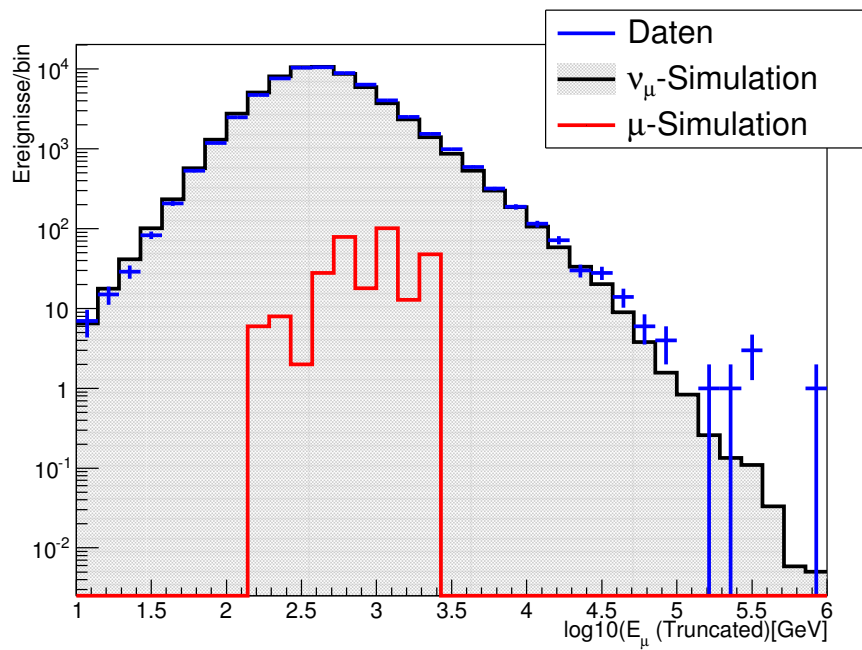
Vor allem in Abbildung 4.9(a) ist ein deutlicher Überschuss an Myonneutrinos bei sehr hohen Energien zu beobachten. Dieser ist mit großer Wahrscheinlichkeit durch eine astrophysikalische Komponente zu erklären. Diese Komponente wird in Kapitel 5 ausführlich untersucht.

Im Folgenden werden einige weitere Eigenschaften des Datensatzes untersucht. Hierbei steht insbesondere die Frage im Mittelpunkt, ob der Überschuss in der Hochenergie durch Untergrund erklärt werden kann.

4. Analyse der IC 79-Daten



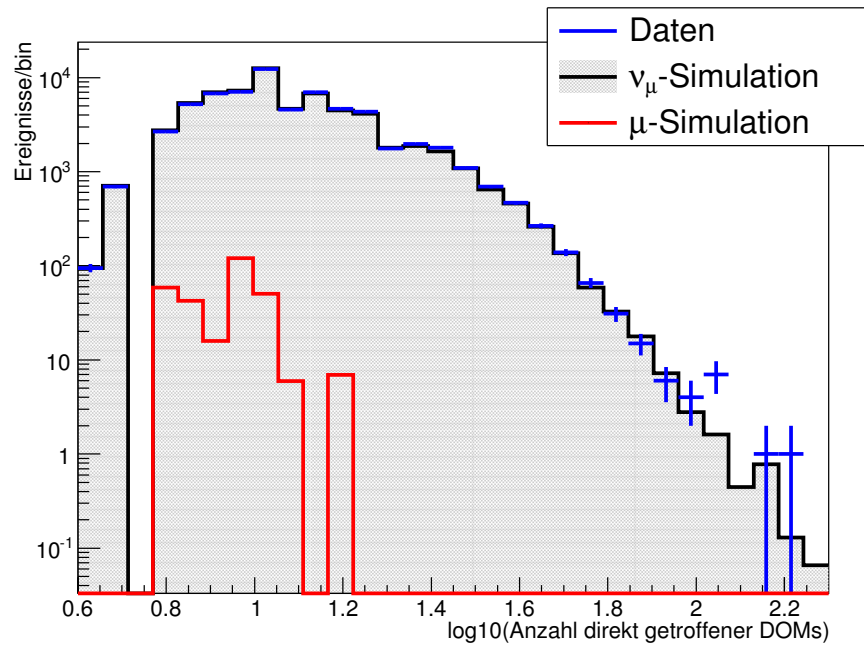
(a)



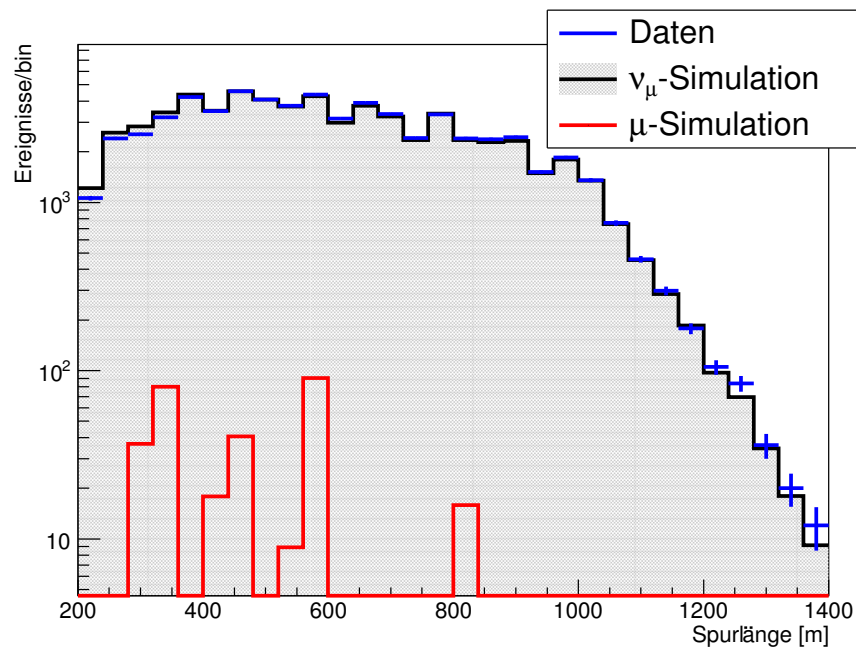
(b)

Abbildung 4.9.: Signal- und Untergrundverteilungen für verschiedene Merkmale. a) Energieschätzung mit MuEX b) Energieschätzung mit TruncatedMean. In diesen Merkmalen kann eine Übereinstimmung zwischen Daten und Simulationen beobachtet werden.

4. Analyse der IC 79-Daten



(a)



(b)

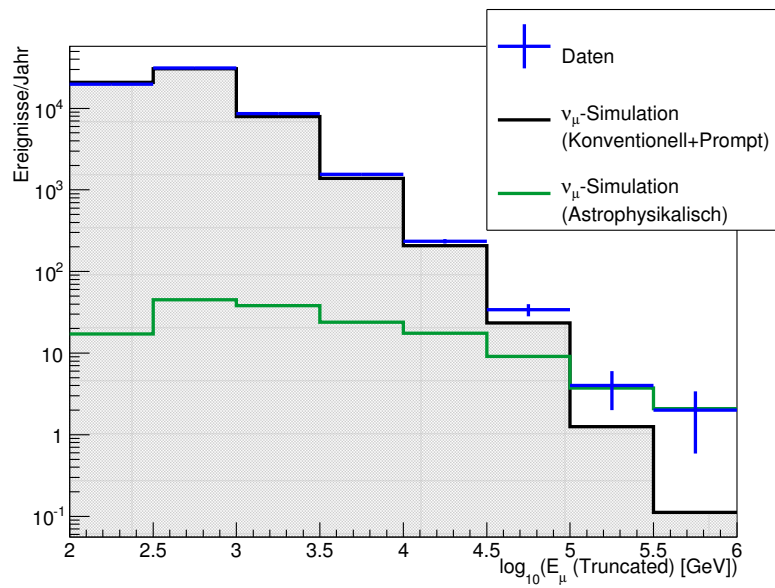
Abbildung 4.10.: Signal und Untergrundverteilungen für verschiedene Merkmale. a) Anzahl direkt getroffener DOMs b) Spurlänge. In diesen Merkmalen kann eine Übereinstimmung zwischen Daten und Simulationen beobachtet werden.

Energieschätzer in unterschiedlichen Winkelbereichen

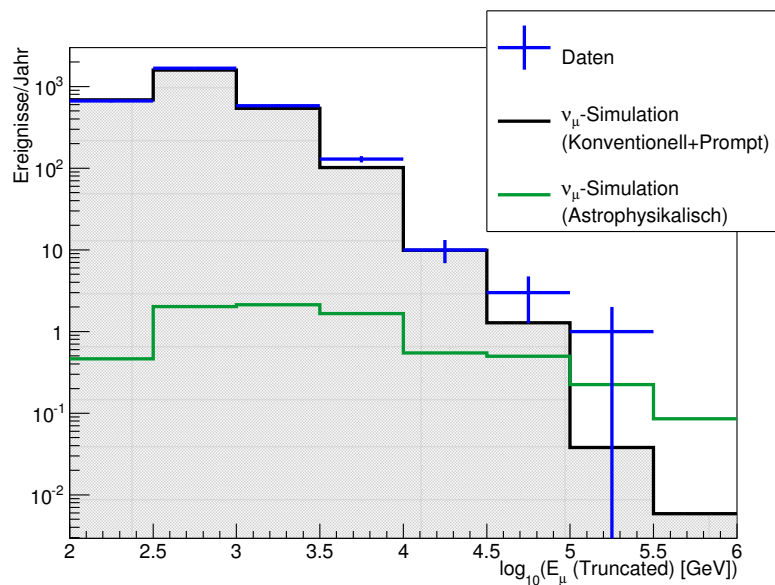
Je eher ein Signal vom Südpol kommt, desto wahrscheinlicher ist es ein Untergrundereignis, welches die Eis- und Gesteinsschichten überwunden hat. Daher ist es wichtig zu überprüfen, in welchem Winkelbereich der Überschuss am deutlichsten ist. Um dies zu überprüfen ist in den Abbildungen 4.11 und 4.12 die mit dem TruncatedMean-Algorithmus geschätzte Energie in verschiedenen Winkelbereichen abgebildet. Der Überschuss an Ereignissen oberhalb von $10^{4.5}$ GeV, welcher auch hier zu sehen ist, hat seinen Ursprung im Winkelbereich von $90-120^\circ$. Weder unterhalb des Horizontes ($86-90^\circ$) noch für den Winkelbereich zwischen 120° und 180° ist eine derartige Häufung zu sehen. Dies ist dadurch zu erklären, dass die Erde in diesem Winkelbereich nicht mehr durchlässig für Neutrinos ist (vgl. Abbildung 4.1).

Neben der atmosphärischen Komponente ist in den Abbildungen 4.11 und 4.12 auch eine mögliche astrophysikalische Komponente eingezeichnet. Diese ist proportional zu E^{-2} und hat eine Normierung von $10^{-8} \frac{\text{GeV}}{\text{sr s m}^2}$. Dies entspricht dem in [A⁺13a] gemessenen Fluss.

4. Analyse der IC 79-Daten



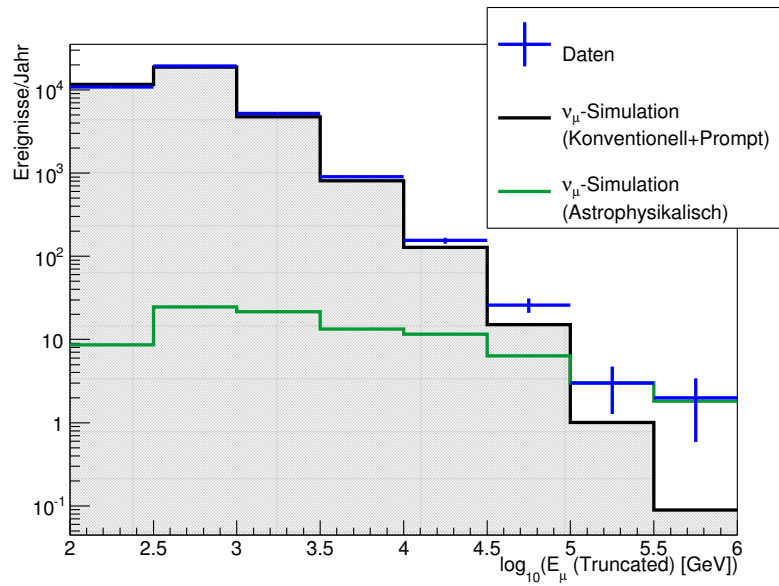
(a)



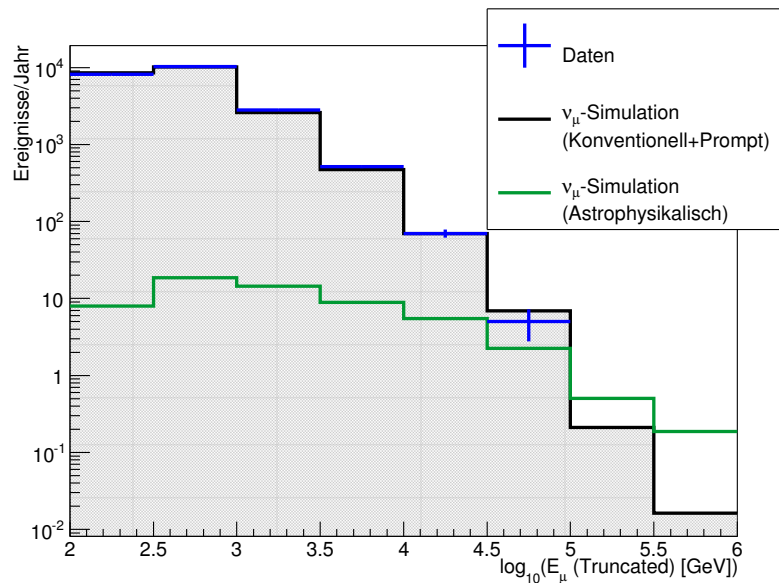
(b)

Abbildung 4.11.: Verteilung der rekonstruierten Myonenergie a) im gesamten Winkelbereich, b) im Winkelbereich zwischen 86° und 90° . Der Überschuss, der bei hohen rekonstruierten Energien zu sehen ist, ist nicht auf einen Überschuss am Horizont zurückzuführen. Für die astrophysikalische Komponente wird ein E^{-2} -Fluss mit einer Normierung von $10^{-8} \frac{\text{GeV}}{\text{sr s m}^2}$ angenommen.

4. Analyse der IC 79-Daten



(a)



(b)

Abbildung 4.12.: Verteilung der rekonstruierten Myonenergie mit dem Truncated-Algorithmus in verschiedenen Winkelbereichen. a) Der Winkelbereich zwischen 90° - 120° . b) Der Winkelbereich zwischen 120° - 180° . Der Überschuss, welcher im gesamten Zenitwinkelbereich zu sehen ist, ist hauptsächlich auf einen Überschuss im Winkelbereich zwischen 90° - 120° zurückzuführen. Dies deutet darauf hin, dass diese Ereignisse Myon-Neutrinos sind. Für die astrophysikalische Komponente wird ein E^{-2} -Fluss mit einer Normierung von $10^{-8} \frac{\text{GeV}}{\text{sr s m}^2}$ angenommen.

Zenitwinkelverteilungen

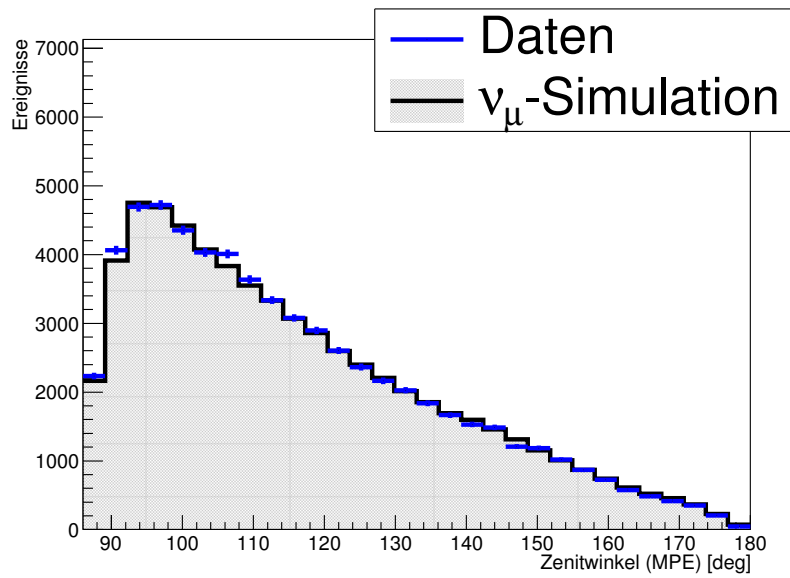
Um weiter zu überprüfen, ob die detektierten Ereignisse Untergrundereignisse sind, kann die Zenitverteilung analysiert werden. Wie auch schon im Test zuvor wäre eine Häufung der Ereignisse am Horizont ein Indiz für einen Untergrundbeitrag. In Abbildung 4.13 ist die Zenitwinkelrekonstruktion aus dem MPE-Algorithmus für die Simulation und den Datensatz aufgetragen. In Abbildung 4.13(a) ist der gesamte Datensatz dargestellt und in Abbildung 4.13(b) der Teil des Datensatzes, der eine rekonstruierte Myonenergie (Truncated-Algorithmus) von mehr als $10^{4.5}$ GeV hat. Für beide Verteilungen ist zum einen eine Übereinstimmung zwischen Daten und Simulationen zu erkennen. Zum anderen tritt keine signifikante Häufung am Horizont auf. Dies deutet erneut darauf hin, dass die detektierten hochenergetischen Ereignisse keine Untergrundereignisse aus atmosphärischen Myonen sind.

Schwerpunktverteilungen

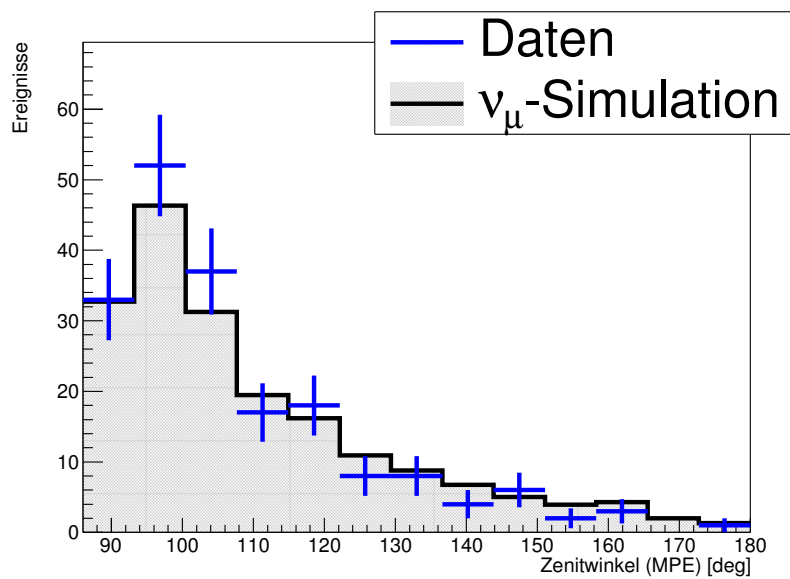
Eine weitere Eigenschaft des Untergrundes ist, dass er von oben in den Detektor eintritt. Aus diesem Grund würden die Schwerpunkte der Ereignisse zu Werten am Rand des Detektors neigen. Um dies zu überprüfen, ist in Abbildung 4.14 die Schwerpunktverteilung zu sehen. Es ist sowohl für den ganzen Energiebereich als auch für hohe Energien eine Übereinstimmung zwischen Daten und Simulationsdaten zu beobachten.

Für die Ereignisse, deren mit dem Truncated-Algorithmus geschätzte Myon-Energie größer als $10^{4.5}$ GeV ist, gibt es keine Überfluktuationen im oberen Bereich des Detektors. Dies ist ein weiteres Anzeichen dafür, dass diese Ereignisse keine atmosphärischen Myonen sind.

4. Analyse der IC 79-Daten

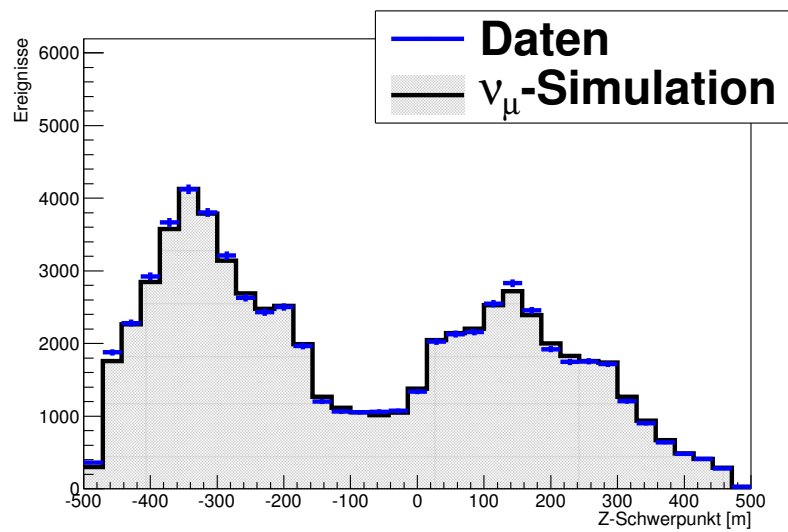


(a)

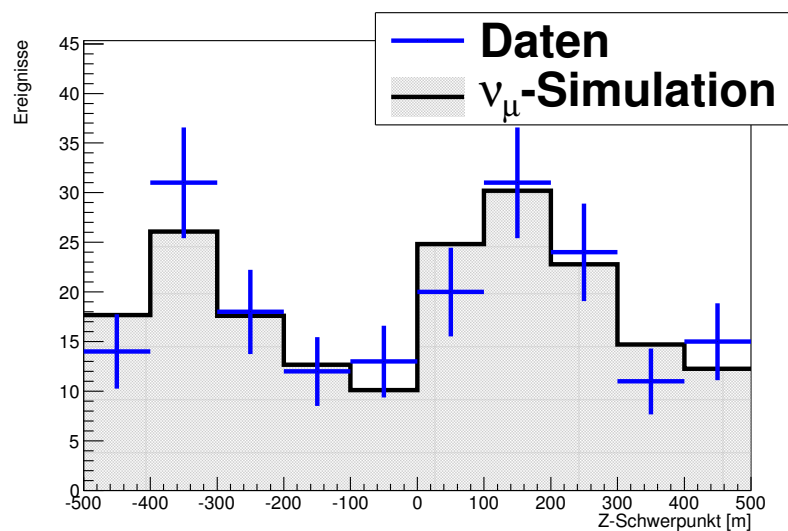


(b)

Abbildung 4.13.: Verteilungen der Zenitwinkel im Datensatz verglichen mit Simulationsdaten. a) Zenitwinkelverteilung für den gesamten Datensatz. b) Zenitwinkelverteilung für Ereignisse, welche eine geschätzte Energie von mehr als $10^{4.5}$ GeV (Truncated-Algorithmus) besitzen.



(a)



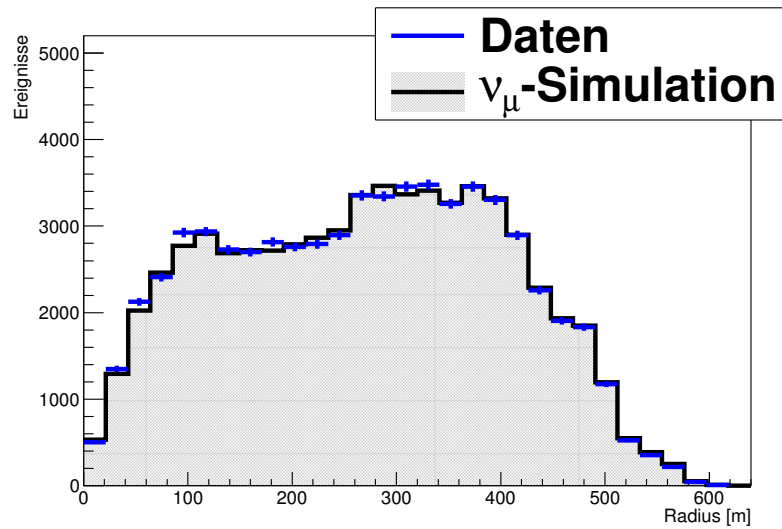
(b)

Abbildung 4.14.: Verteilungen der Schwerpunkte im Datensatz verglichen mit Simulationsdaten. Das Koordinatensystem ist so gewählt, dass 0 den Mittelpunkt des Detektors bezeichnet. a) Schwerpunktsverteilung für den gesamten Datensatz. b) Schwerpunktsverteilung für Ereignisse, welche eine geschätzte Energie von mehr als $10^{4.5}$ GeV (Truncated-Algorithmus) besitzen.

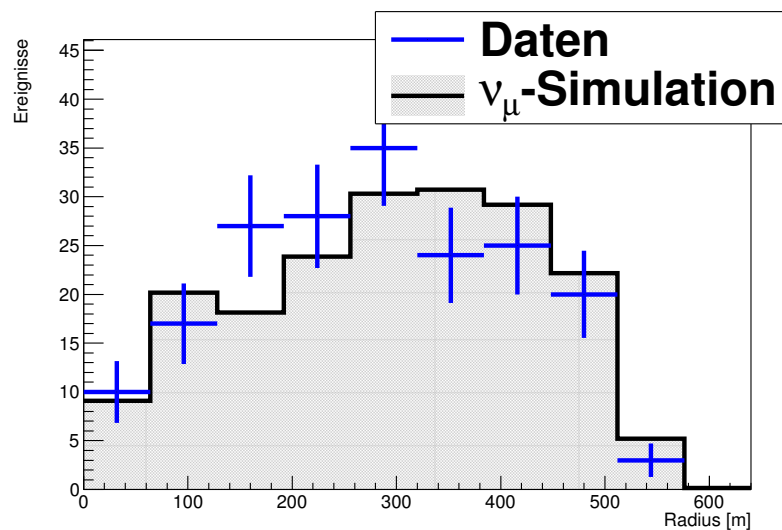
Radiusverteilungen

Ein ähnlicher Test ist auch mit den Radien möglich. Hierzu wird der Radius, welcher sich aus X - und Y -Schwerpunkt ergibt, genutzt. Ein hoher Radius bedeutet, dass die Ereignisse am Rand des Detektors liegen. Solche *Corner Clipper* genannte Ereignisse besitzen eine größere Unsicherheit in ihrer Rekonstruktion und sind daher mit höherer Wahrscheinlichkeit atmosphärische Myonen.

Die Verteilungen für den gesamten Datensatz und für Ereignisse, welche eine rekonstruierte Energie mit dem Truncated-Algorithmus von mehr als $10^{4.5}$ GeV haben, ist in Abbildung 4.15 zu sehen. Sowohl für den gesamten Energiebereich als auch für den hochenergetischen Teil des Datensatzes ist eine Übereinstimmung zwischen Daten und Simulationsdaten zu beobachten. Insbesondere ist zu sehen, dass die Ereignisse nicht dazu neigen, ihren Schwerpunkt am Rand des Detektors zu besitzen. Dies ist wiederum ein Anzeichen dafür, dass der Überschuss nicht aus atmosphärischen Myonen besteht.



(a)



(b)

Abbildung 4.15.: Verteilungen der Radien im Datensatz verglichen mit Simulationsdaten. Das Koordinatensystem ist so gewählt, dass 0 der Mittelpunkt des Detektors bezeichnet. a) Radienverteilung für den gesamten Datensatz. b) Radienverteilung für Ereignisse, welche eine geschätzte Energie von mehr als $10^{4.5}$ GeV (Truncated-Algorithmus) haben.

4.2. Entfaltung

Bisher wurden in dieser Arbeit ausschließlich Ergebnisse der Selektion gezeigt, bei denen die rekonstruierte Myon-Energie oder ein anderes gemessenes Merkmal Gegenstand der Diskussion war. Die wichtige Information ist jedoch das Energiespektrum der Neutrinos. In Abschnitt 3.4 wurde die Technik der Entfaltung vorgestellt, um dieses Problem zu lösen.

Der folgende Abschnitt gliedert sich in drei Teile. Zunächst werden die gewählten Entfaltungsmerkmale vorgestellt und motiviert. Anschließend werden verschiedene Tests der Entfaltung auf Simulationsdaten vorgestellt. Dies umschließt unter anderem den Test- und Pull-Modus. Zum Schluss wird das entfaltete Spektrum auf Daten vorgestellt und die Tests mittels Umgewichten der Merkmale gezeigt.

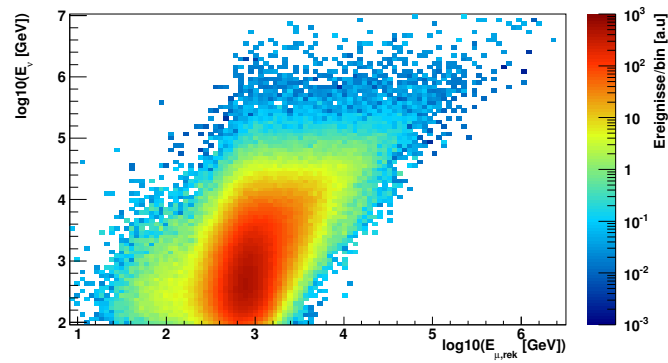
4.2.1. Entfaltungsmerkmale und Einstellungen

Die Merkmale, die für die Entfaltung genutzt wurden, sind der Energieschätzer MuEX, die rekonstruierte Spurlänge aus dem MPE-Algorithmus MPE_LDirC, sowie die Anzahl der direkt getroffenen DOMs aus dem MPE-Algorithmus MPE_NDirC.

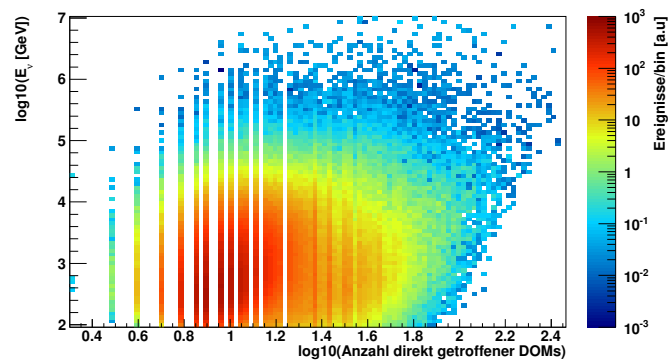
Die Wahl des Energieschätzers von MuEX ist dadurch begründet, dass es die beste Korrelation mit der gesuchten Energie besitzt und gleichzeitig eine geringe Korrelation mit den anderen Entfaltungsobservablen aufweist.

Eine Hinzunahme des anderen Energieschätzers TruncatedMean bringt keine Verbesserung, da die Korrelation zwischen diesen beiden Schätzern sehr hoch ist. Die Observablen LDirC und NDirC wurden in dieser Kombination bereits in der IC 59 Analyse [Ruh13] genutzt, wo es sich gezeigt hat, dass die Kombination dieser beiden Merkmale sich gut für eine Entfaltung mit TRUEE eignet. Die Abhängigkeit der Merkmale zur Energie ist in Abbildung 4.16 zu sehen.

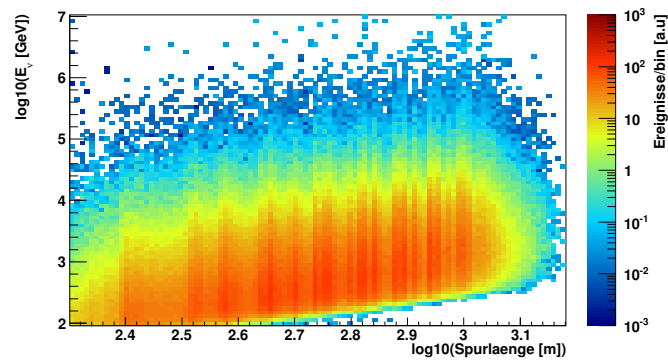
4. Analyse der IC 79-Daten



(a)



(b)



(c)

Abbildung 4.16.: Abhängigkeitsplots zwischen der simulierten wahren Neutrino-Energie und den genutzten Entfaltungsmerkmalen. (a) Energierekonstruktion des Myons mit dem MuEX-Algorithmus. (b) Anzahl direkt getroffener DOMs. c) Rekonstruierte Spurlänge.

4. Analyse der IC 79-Daten

Um die Energieauflösung dieser Merkmalskombination einschätzen zu können, werden monoenergetische Ereignisse entfaltet. Um dies zu erreichen werden aus dem Datensatz Ereignisse herausgesucht, welche die Bedingung

$$4.4 < \log(E_\nu[\text{GeV}]) < 4.6$$

erfüllen. Die Breite ist dabei so gewählt, dass genügend Ereignisse im Datensatz verbleiben, damit die Entfaltung nicht von statistischen Fluktuationen beeinflusst ist. Die Entfaltung solcher Ereignisse ist in Abbildung 4.17 zu sehen. Das Anpassen einer Gaußverteilung an diese Werte führt zu einer Standardabweichung von $\sigma = 0.26$.

Aus diesem Grund wird für die Entfaltung eine Binbreite von 0.3 in $\log(E_\nu[\text{GeV}])$ gewählt. Das letzte Bin, welches sich oberhalb von 1 PeV befindet, bildet eine Ausnahme. Da die zu erwartenden Ereigniszahlen im letzten Bin sehr gering sind, wird in diesem Bin eine Breite von 0.5 in $\log(E_\nu[\text{GeV}])$ gewählt. Da auf diese Weise die Ereigniszahlen in diesem Bin erhöht werden, wird auch der relative statistische Fehler dort verringert.

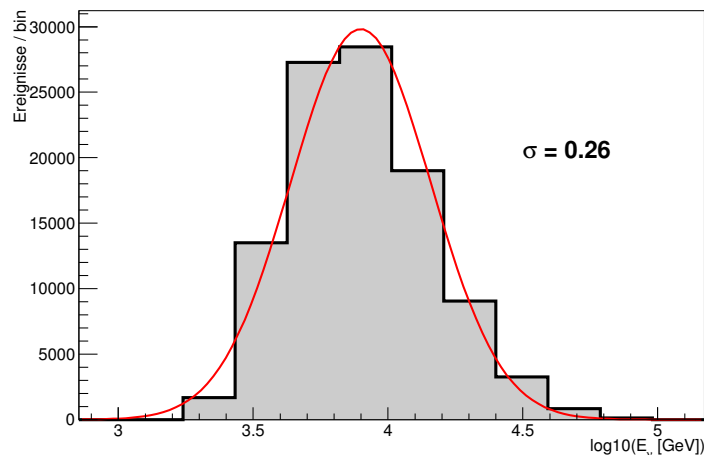


Abbildung 4.17.: Entfaltung von Ereignissen, welche näherungsweise monoenergetisch sind, zur Bestimmung der Energieauflösung. In diesem Fall sind diese Ereignisse aus dem Intervall $4.4 < \log(E_\nu[\text{GeV}]) < 4.6$. Das Ergebnis ist eine Verteilung, welche näherungsweise gaußverteilt ist. Die angepasste Gaußverteilung hat eine Standardabweichung von 0.26. Dies begründet die Energieauflösung von 0.3.

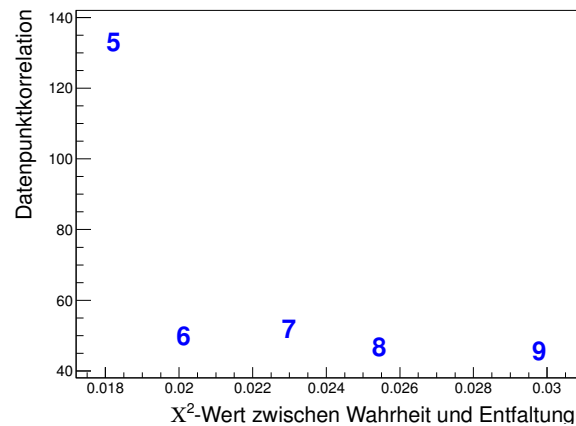


Abbildung 4.18.: L-Kurve für verschiedene Werte für die Anzahl der Freiheitsgrade (blau). Wenige Freiheitsgrade bedeuten eine starke Regularisierung. Aufgetragen ist das Maß für die Korrelation zwischen den Bins gegen den χ^2 -Wert zwischen Wahrheit und Entfaltung.

Die Stärke der Regularisierung ist eine wichtige Einstellung bei der Entfaltung. Je stärker regularisiert wird, desto stärker wird der Log-Likelihoodschätzer verzerrt und somit nicht mehr verlässlich. Auf Grund der stärkeren Regularisierung wird auch die Bin-zu-Bin-Korrelation größer und die Varianz der Messung geringer (vgl. auch Gleichung (3.4.25)).

Bei zu geringer Regularisierung sind die nicht physikalischen Effekte aus der singulären Antwortmatrix im Spektrum zu finden. Eine Abwägung zwischen Erwartungstreue und Varianz wird auch *Bias-Variance-Tradeoff* genannt. In TRUEE wird die sogenannte Datenpunktkorrelation (*englisch: data point correlation (DPC)*) berechnet. Diese gibt an, wie stark die Bins miteinander korreliert sind. Details hierzu sind in [Mil12] zu finden.

Dieser Wert kann genutzt werden, um den Einfluss der Regularisierung zu analysieren. Üblicherweise wird der Wert der DPC in Abhängigkeit des χ^2 -Werts zwischen Simulationswahrheit und Entfaltungsergebnis aufgetragen. Dieser L-Kurve genannte Plot ist in Abbildung 4.18 zu sehen. Es ist ersichtlich, dass kleine Freiheitsgrade, also hohe Regularisierungen, zu sehr hohen Bias führen. Durch diesen Plot wird der erlaubte Bereich auf 6-8 eingegrenzt. Bei höheren Werten sind unphysikalische Effekte zu sehen.

4. Analyse der IC 79-Daten

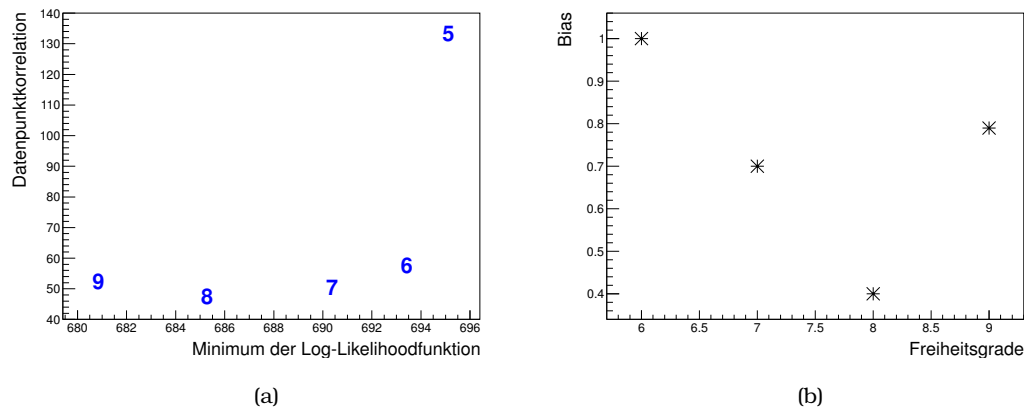


Abbildung 4.19.: Weitere Tests zur Wahl der Regularisierungsstärke. (a) L-Kurve auf Daten. Anstelle des χ^2 -Werts wird der minimale Wert der Log-Likelihood-Funktion benutzt. Ein kleinerer Wert bedeutet eine bessere Übereinstimmung und einen kleineren Einfluss der Regularisierung. (b) Maximimale Verschiebungen des Mittelwertes der Pull-Verteilungen für verschiedene Freiheitsgrade. Der Wert von 8 wird anhand beider Plots klar präferiert.

Im Laufe dieser Arbeit wurde TRUEE so erweitert, dass die L-Kurve auch auf Daten nutzbar ist. Dazu wird nicht der χ^2 -Wert genutzt, sondern der Wert der Log-Likelihood-Funktion ohne Berücksichtigung des Regularisierungsterms. Je kleiner dieser Wert ist, desto besser ist die Übereinstimmung. Da dieser Wert auf Daten zur Verfügung steht, kann dieser Test auf Daten durchgeführt werden. Das Ergebnis ist in Abbildung 4.19 zu sehen. Generell ist ein ähnlicher Trend wie in Abbildung 4.19(a) zu sehen. Jedoch ist auch zu sehen, dass der Wert mit 8 Freiheitsgraden besser als die Ergebnisse mit 6 oder 7 Freiheitsgraden ist.

Dies kann mit einem weiteren Test überprüft werden, bei dem der Pull-Modus genutzt wird. Es wird für jede in Frage kommende Einstellung der Pull-Modus ausgeführt. Anschließend kann für jedes Bin der Bias, aus der Verschiebung der Verteilung gegen 0 abgelesen werden. Als Maß des Bias wird nun das Maximum dieser Werte genutzt. Eine grafische Darstellung dieser Werte ist in Abbildung 4.19(b) zu sehen. Durch diesen Test wird deutlich ersichtlich, dass 8 Freiheitsgrade die gewünschte Regularisierungsstärke ist. Eine Übersicht über alle Entfaltungseinstellungen ist in Tabelle 4.4 gegeben.

4. Analyse der IC 79-Daten

Einstellung	Wert	Merkmal	Anzahl Bins
Freiheitsgrade	8	MuEX_Energy	15
Knoten	30	MPE_NDirC	10
		MPE_LDirC	10

Tabelle 4.4.: Einstellungen für die durchgeführte Entfaltung.

4.2.2. Überprüfung auf Simulationsdaten

Um die Güte der gewählten Einstellung zu überprüfen, werden verschiedene Tests durchgeführt. Der erste von diesen Tests ist die Ausführung des Test-Modes auf Simulationsdaten. Dies ist in Abbildung 4.20 zu sehen. Der Vergleich zwischen Simulationswahrheit und Entfaltungsergebnis zeigt eine gute Verträglichkeit innerhalb der Fehler, so, dass davon ausgegangen werden kann, dass diese Entfaltung zuverlässige Resultate ergibt. Um dies weiter zu verifizieren, wird der Pull-Mode genutzt.

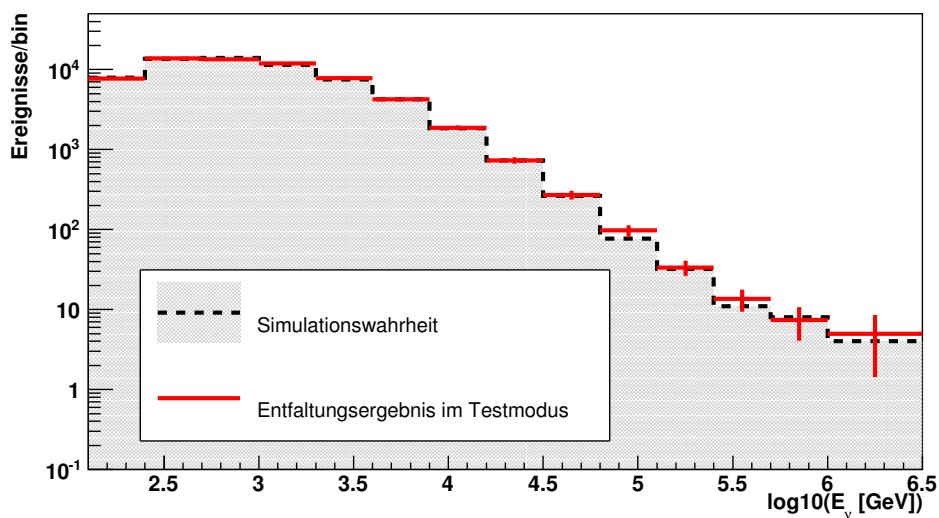


Abbildung 4.20.: Entfaltung eines Testdatensatzes im Test-Modus von TRUEE. Die Simulationswahrheit ist grau hinterlegt, das Ergebnis der Entfaltung ist in rot dargestellt. Die Entfaltung zeigt eine sehr gute Übereinstimmung mit der Simulationswahrheit.

4. Analyse der IC 79-Daten

Die Pull-Verteilungen für jedes Bin sind in den Abbildungen B.3 bis B.5 im Anhang abgebildet. Der Wert

$$\frac{\text{Wahrheit} - \text{Entfaltung}}{\sigma}$$

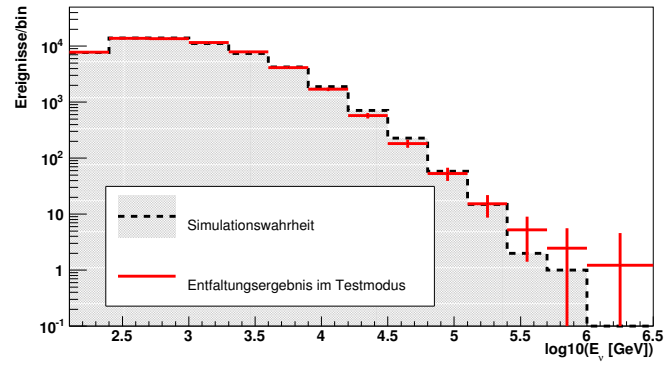
wird für jedes bin in jeder Iteration bestimmt. Die Verteilungen sind alle geringer im Vergleich zur vorherigen Analyse [Ruh13]. Insbesondere unter Berücksichtigung der systematischen Fehler (vgl. Abschnitt 4.2.4) ist der Bias durch den Entfaltungsalgorithmus vernachlässigbar gering.

4.2.3. Modellunabhängigkeit der Entfaltung

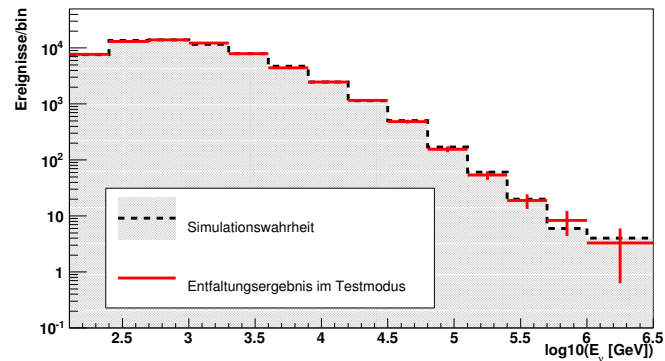
Um zu zeigen, dass die Entfaltung ohne Einfluss hinsichtlich der Simulationsannahme arbeitet, werden weitere Tests durchgeführt. Um dies zu erreichen, wird die Simulationsannahme, welche genutzt wird um die Antwortmatrix zu bestimmen, nicht verändert. Die Daten, welche jedoch benutzt werden, um die Entfaltung zu testen, werden hinsichtlich ihres zugrunde liegenden theoretischen Modells verändert.

Bei diesem Test werden insgesamt vier verschiedene Hypothesen für den Fluss getestet. Zwei von ihnen beinhalten keine astrophysikalische Komponente. Diese sind in Abbildung 4.21 dargestellt. Zum einen wird ein atmosphärisches Spektrum nach [HKK⁺07] mit einer prompten atmosphärischen Komponente nach [ERS08] genutzt, zum anderen wird die prompte Komponente in ihrer Normierung um den Faktor 30 erhöht. Diese Annahme ist durch andere Analysen [Ruh13, Sch13] bereits ausgeschlossen, wird aber genutzt, damit die Stärke der Entfaltung demonstriert wird.

4. Analyse der IC 79-Daten



(a)



(b)

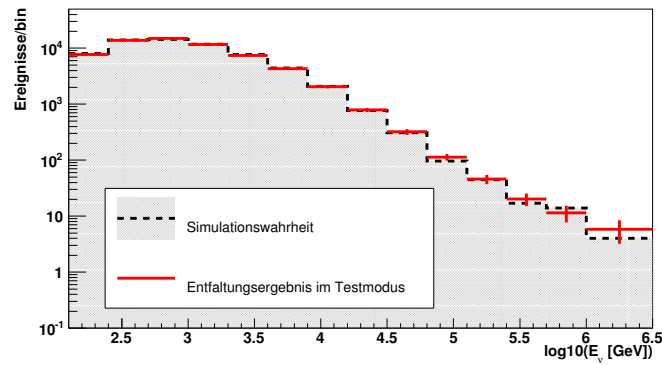
Abbildung 4.21.: Entfaltungen mit verschiedenen Annahmen für die Testdaten. (a) Atmosphärisches Spektrum nach [HKK⁺07] mit der Vorhersage für prompte Neutrinos nach [ERS08]. (b) Konventionelles atmosphärisches Spektrum nach [HKK⁺07] mit einer 30-fachen prompten Komponente nach [ERS08].

Die beiden in Abbildung 4.22 gezeigten Entfaltungen beinhalten eine konventionelle und prompte Komponente. Als astrophysikalische Komponente wird jeweils ein isotroper Fluss angenommen. Der Fluss ist somit gegeben durch

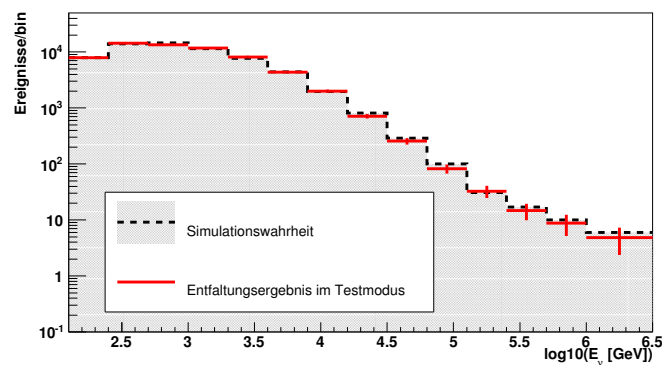
$$\Phi(E) = N_0^{-\gamma} \text{ mit } \gamma = 1.7 \text{ bzw. } 2.4. \quad (4.2.1)$$

Die Annahme von $\gamma = 1.7$ ist eine Annahme, die durch die Analyse von startenden Spuren basierend auf drei Jahren IceCube-Daten [A⁺14b] nicht mehr favorisiert wird. Die Normierungen der astrophysikalischen Flüsse werden so gewählt, dass

4. Analyse der IC 79-Daten



(a)



(b)

Abbildung 4.22.: Entfaltungen mit verschiedenen Annahmen für die Testdaten. Beide Testdatensätze beinhalten ein atmosphärisches Spektrum nach [HKK⁺07] und eine prompt Komponente nach [ERS08] sowie eine astrophysikalische Komponente. (a) Die astrophysikalische Komponente ist $\sim E^{-1.7}$. Die Normierung ist so gewählt, dass der Fluss bei 1 PeV $10^{-8} \frac{\text{GeV}}{\text{sr s m}^2}$ beträgt. (b) Die astrophysikalische Komponente ist $\sim E^{-2.4}$. Die Normierung ist so gewählt, dass der Fluss bei 1 PeV $10^{-8} \frac{\text{GeV}}{\text{sr s m}^2}$ beträgt.

der Fluss bei 1 PeV $10^{-8} \frac{\text{GeV}}{\text{sr s m}^2}$ beträgt. Dies entspricht dem entfalteten Fluss bei 1 PeV aus [A⁺13a].

Die Abbildungen 4.21 und 4.22 zeigen, dass die Entfaltung modellunabhängig arbeitet. Die Verteilungen können für alle genutzten Hypothesen rekonstruiert

werden. Auch Annahmen, welche eine starke Änderung gegenüber der Simulationsannahme aufweisen, sind gut entfaltbar.

4.2.4. Untersuchung systematischer Fehler

Wie im vorherigen Abschnitt gezeigt, ist die Methodik der Entfaltung unabhängig von dem genutzten theoretischen Modell. Für die Entfaltung ist jedoch die korrekte Simulation zwischen gemessenem Merkmal y und Zielmerkmal x wichtig. Dieser Zusammenhang wird genutzt, um die Antwortmatrix zu bilden. In der Simulation für IceCube gibt es jedoch Parameter, mit einer gewissen Varianz. Dies sind vor allem die Eigenschaften des Eises, die DOM-Sensitivität sowie die Unsicherheiten in den Wirkungsquerschnitten der Energieverluste von Myonen.

Diese Effekte müssen berücksichtigt werden und in die systematischen Fehler eingehen. Modellannahmen wie die Normierung der kosmischen Strahlung, der spektrale Index der kosmischen Strahlung oder auch das gewählte Modell für hadronische Wechselwirkungen müssen nicht berücksichtigt werden, da die Entfaltung modellunabhängig ist.

Bei genauer Betrachtung ist der Effekt der ungenauen Beschreibung der Myon-Nukleon Wirkungsquerschnitte und der Unsicherheit in der DOM-Sensitivität ähnlich. Beide Effekte führen dazu, dass weniger Licht erzeugt wird. Die reduzierten Wirkungsquerschnitte führen zu diesem Effekt, da weniger Paarproduktionsverluste auftreten und somit weniger Tscherenkowlicht entsteht. Die DOM-Sensitivität führt ebenso dazu, dass weniger der erzeugten Photonen detektiert werden. Daher wird diese Fehlerquelle im Folgenden zu einem Fehler zusammengefasst.

Um das Eis in zwei Kilometern Tiefe unterhalb des Südpol zu modellieren, wurden an einigen DOMs LEDs angebracht. Diese können Licht aussenden, welches benachbarte DOMs detektieren. Anhand des gemessenen Signals lassen sich anschließend die physikalischen Eigenschaften wie Absorptionslänge und Streulänge mit Hilfe einer Minimierung bestimmen. Diese Minimierung hat jedoch eine gewisse Unsicherheit. Um diese Unsicherheiten in die Entfaltung mit aufzunehmen, werden Simulationen erzeugt, in denen die entsprechenden Parameter verändert wurden.

Um den Einfluss der Unsicherheiten bestimmen zu können wird der Pull-Modus von TRUEE genutzt. Es wird die Option genutzt, die Antwortmatrix auf anderen Si-

4. Analyse der IC 79-Daten

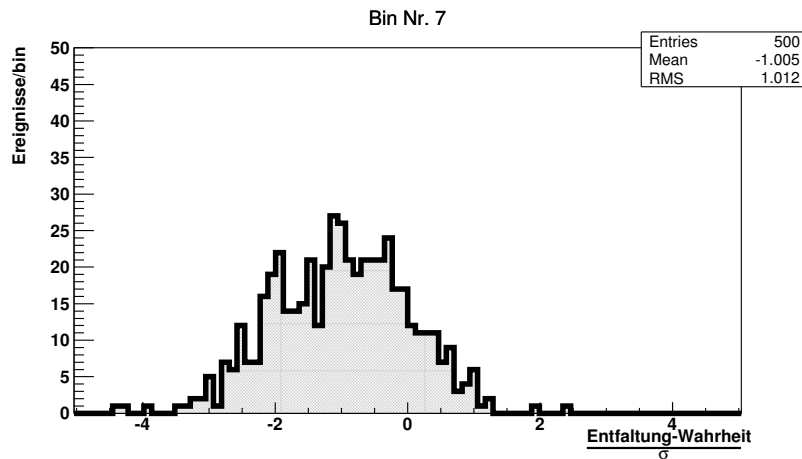


Abbildung 4.23.: Pull-Verteilung für das siebte Bin der Entfaltung unter der Annahme, dass die DOM-Sensitivität um 10% geringer ist als angenommen. Die Entfaltung unterschätzt den wahren Fluss in den Simulationsdaten um ca. 1σ . Dieser Effekt wird in den systematischen Fehlern berücksichtigt.

mulationsdaten zu bestimmen, als denen, die zum Testen genutzt werden. TRUEE erzeugt eine Pull-Verteilung für jedes Bin. Ein Beispiel für einen solchen Plot ist in Abbildung 4.23 zu sehen. Bei diesem Pull-Mode wurde die DOM-Sensitivität im Testdatensatz um 10% verringert. Dadurch werden weniger Photonen detektiert und die Abhängigkeit zwischen rekonstruierter Energie und wahrer Energie verändert sich. Wie in Abbildung 4.23 zu sehen, unterschätzt die Entfaltung im siebten Bin den wahren Bineintrag um ca. 1σ . Dieser Wert wird für alle systematischen Datensätze bestimmt. Der systematische Fehler ergibt sich aus der Addition der Varianzen aus den einzelnen Fehlerquellen gemäß

$$\sigma_{sys,i}^2 = \sum_{\text{Fehlerquellen}} \sigma_{sys,i,k}^2 \quad (4.2.2)$$

Der Wert $\sigma_{sys,i,k}^2$ ist die mittlere Verschiebung durch die k -te Fehlerquelle im i -ten Bin. Bei dieser Rechnung wird berücksichtigt, dass die Fehlerquellen immer nur zu einer Über- oder Unterschätzung führen. Daher ergeben sich asymmetrische Fehler.

Die resultierenden systematischen Fehler sind in Abbildung 4.24 dargestellt. Es ist ersichtlich, dass über einen weiten Energiebereich die systematischen Fehler dominieren. Nur bei höchsten Energien ist der statistische untere Fehler domi-

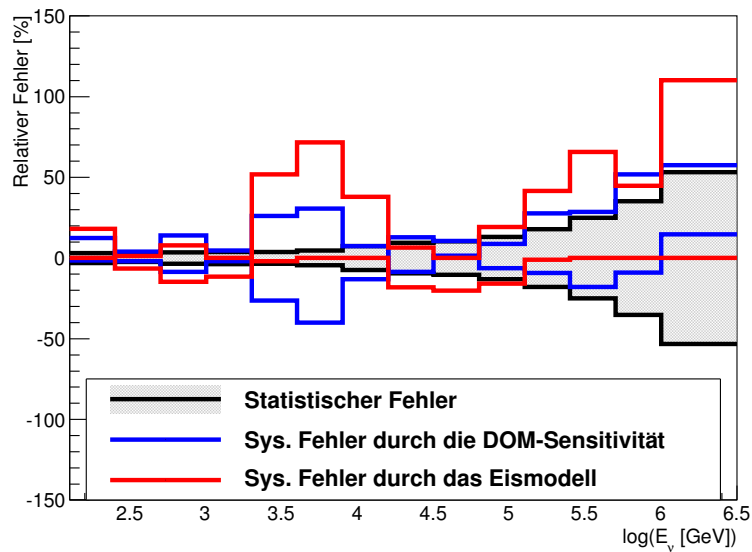


Abbildung 4.24.: Übersicht über die systematischen Fehler. Zum Vergleich sind die statistischen Fehler ebenfalls eingezeichnet.

nant. Die großen systematischen Fehler im Energiebereich zwischen 1 TeV und 10 TeV sind auf den Zusammenhang zwischen $\frac{dE}{dx}$ -Wert und Myonenergie zurückzuführen. Wie in Abbildung 4.25 zu sehen ist, ist die Steigung in diesem Bereich deutlich höher als bei größeren Energien. Daher resultiert hier durch eine Veränderung der Detektionseffizienz durch die DOM-Sensitivität oder das Eismodell ein größerer systematischer Fehler.

Für eine genauere Messung des atmosphärischen Neutrino-Energiespektrums hat eine Vergrößerung des Datensatzes keinen großen Einfluss auf die Fehler. Um ein besseres Verständnis zu erlangen, kann zum einen die Systematik der Messung verbessert werden. Die andere Möglichkeit ist, statt eines einfachen Energiespektrums ein zweidimensionales Spektrum in Zenit und Energie zu entfalten. Dies ist entweder durch Zenitschnitte oder durch eine echte zweidimensionale Entfaltung möglich.

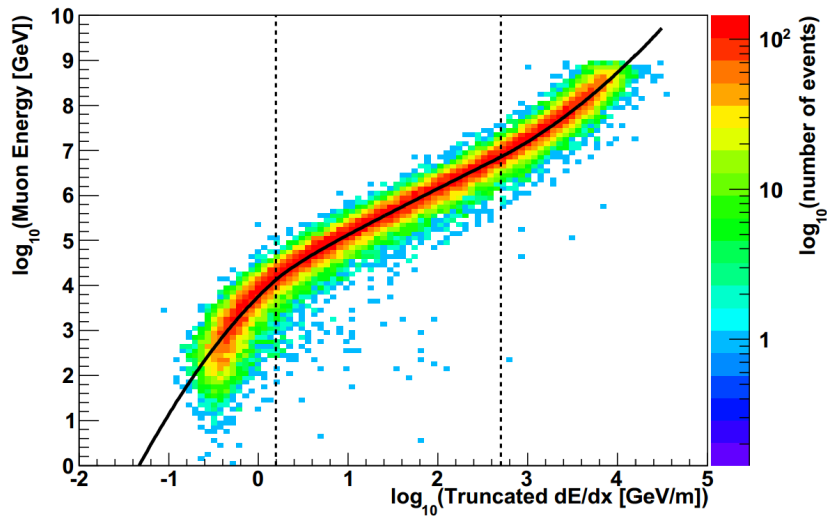


Abbildung 4.25.: Abhängigkeit zwischen berechnetem $\frac{dE}{dx}$ aus dem Truncated-Algorithmus und der wahren Myon-Energie. Die Steigung im Bereich von 1 TeV ist deutlich höher. Dies begründet den großen systematischen Fehler im Energiebereich der Neutrinos zwischen 1 TeV und 10 TeV Neutrinoenergie [A⁺13d].

4.2.5. Anwendung auf Daten

Die Anwendung der Entfaltung auf Daten führt zu Ergebnissen, welche in Tabelle C.1 und Abbildung 4.26 dargestellt sind. Es ist eine deutliche Überfluktuation im Hochenergiebereich zu erkennen. Diese Überfluktuation ist nicht durch konventionelle atmosphärische Neutrino zu erklären, sie muss also entweder prompten oder astrophysikalischen Ursprungs sein. Die Eigenschaften dieses Spektrums werden in Kapitel 5 ausführlich diskutiert.

Ungewichtete Merkmale

Um die Qualität der Entfaltung weiter beurteilen zu können, können an der Entfaltung unbeteiligte Merkmale genutzt und mit der entfalteten Energieverteilung neugewichtet werden. (vgl Kapitel 3.4). Diese Verteilung muss mit der Verteilung auf Daten übereinstimmen, da nun dasselbe zugrunde liegende Spektrum gezeigt wird. Dieser Test wird für den durch den MPE-Algorithmus geschätzten

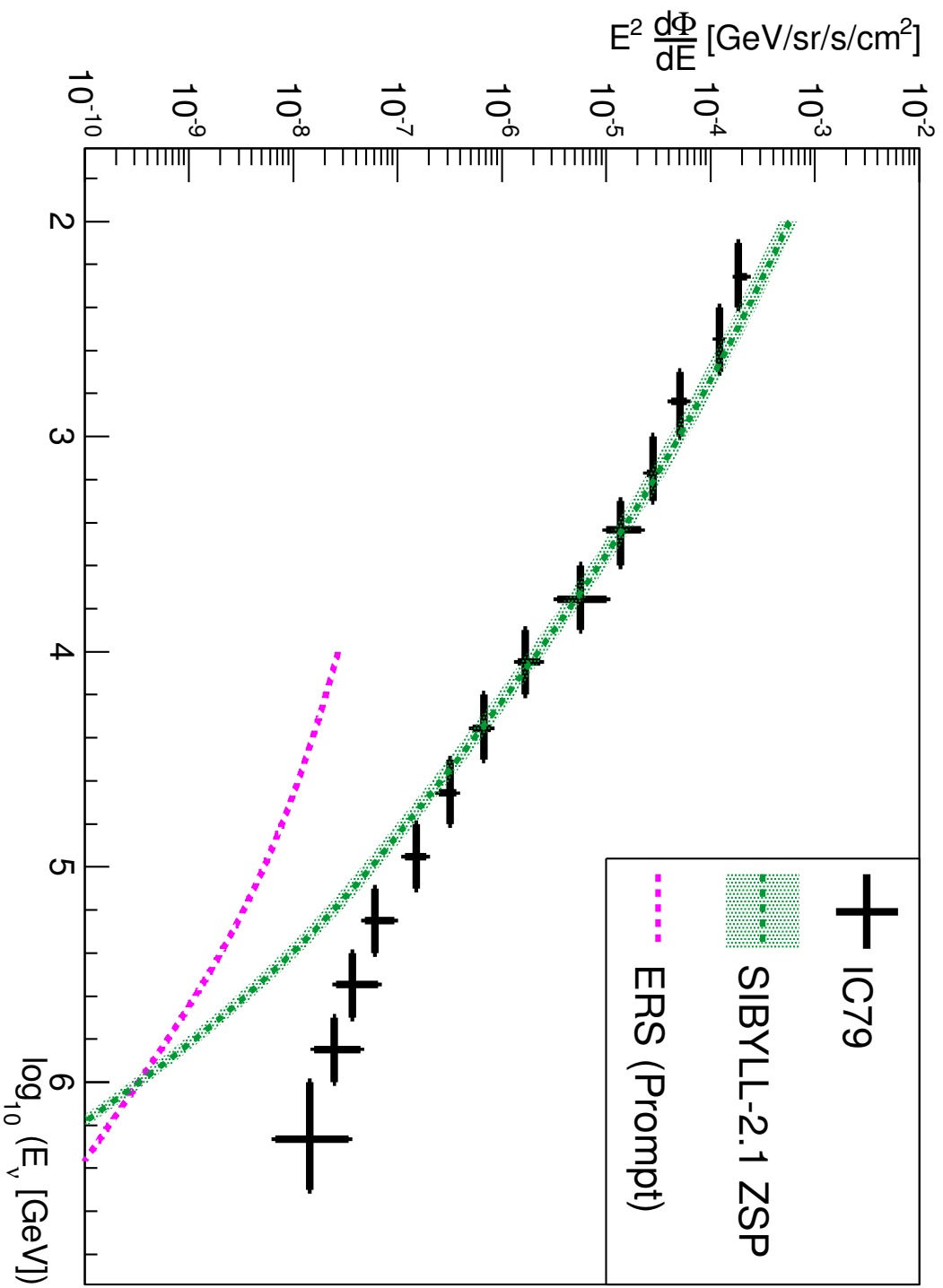


Abbildung 4.26: Gemessener und mit E^{-2} skaliertes Fluss aus der Analyse der IC79 Daten im Vergleich zu Rechnungen aus [FTD12] mit SIBYLL-2.1 als hadronisches Interaktionsmodell und ZSP [$A^+ 1 \text{ Ib}$] als primäres Flussmodell für die konventionelle Komponente und [ERS08] für die prompte Komponente. Es ist eine weitere Komponente oberhalb von 100 TeV sichtbar.

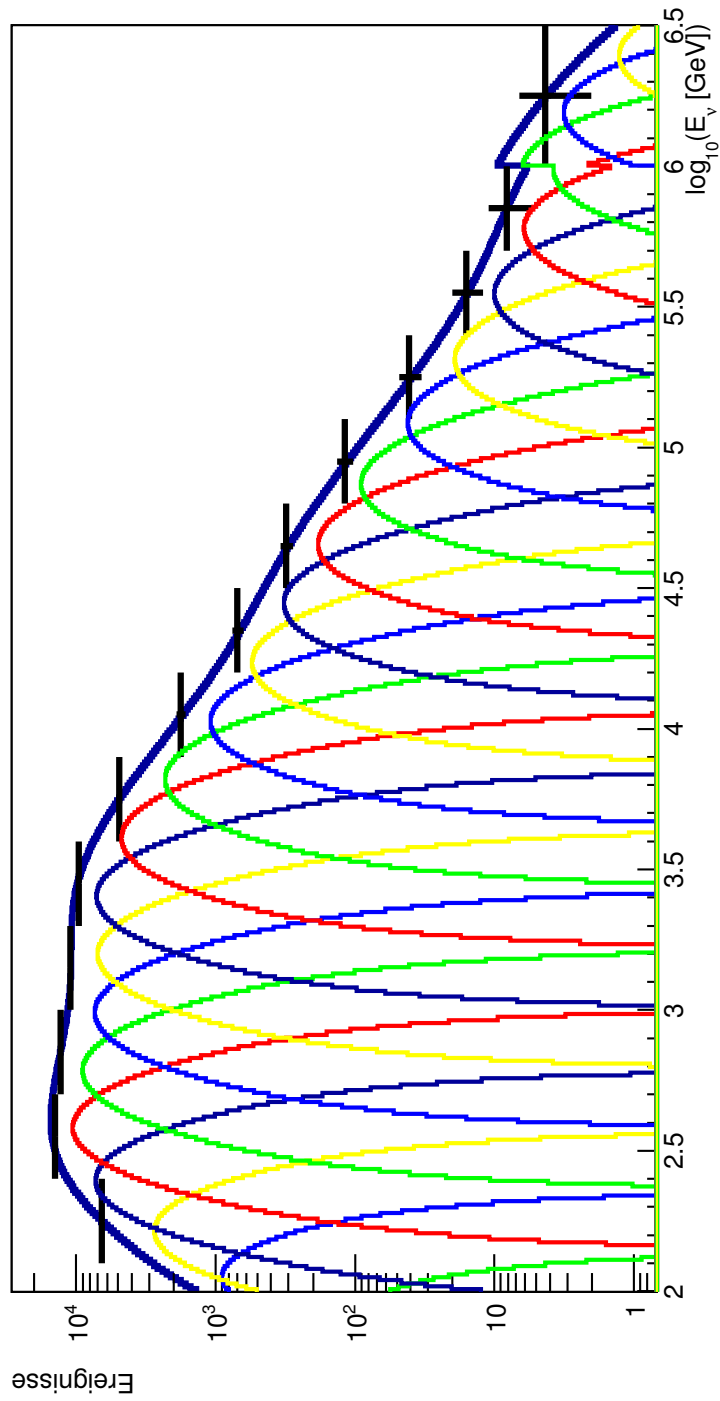
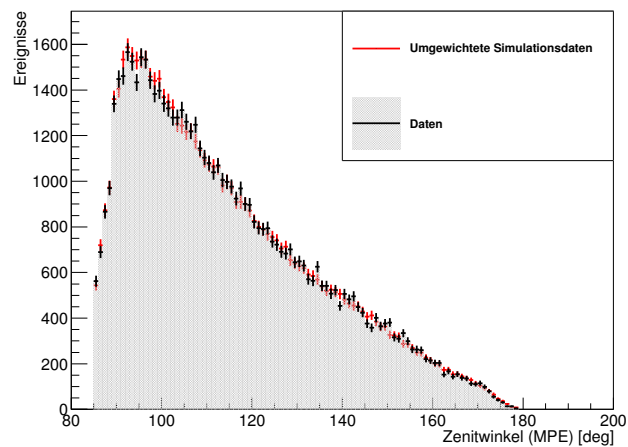
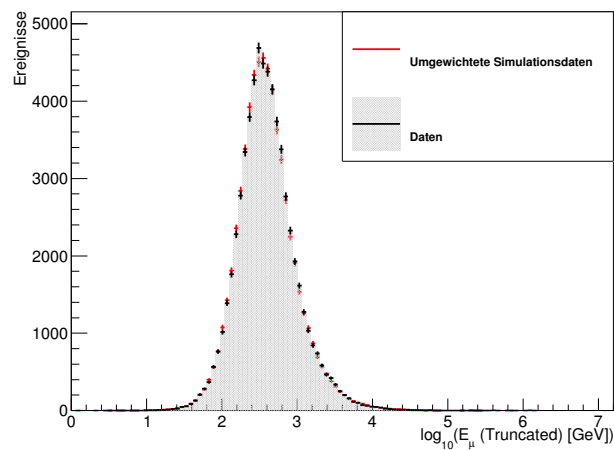


Abbildung 4.27.: Entfaltete Ereignisse des IC 79-Datensatzes mit Spline und zugrunde liegenden B-Splines. Die Sprungstelle bei 10^6 GeV ist auf die Umrechnung von differentiellen Raten auf Ereignisse zurückzuführen.

4. Analyse der IC 79-Daten



(a)



(b)

Abbildung 4.28.: Umgewichtete Verteilungen. a) Der Zenitwinkel geschätzt durch den MPE-Algorithmus. b) Die Myonenergie geschätzt durch den Truncated-Algorithmus. Die Simulationsdaten werden hierbei dem Entfaltungsergebnis entsprechend umgewichtet. Die sehr gute Übereinstimmung lässt auf eine gelungene Entfaltung schließen.

Zenit-Winkel und die durch den Truncated-Algorithmus geschätzte Myon-Energie durchgeführt. Das Ergebnis ist in 4.28 zu sehen. Die umgewichteten Simulationsdaten sind kompatibel mit den Daten. Es kann von einer gelungenen Entfaltung ausgegangen werden.

5. Diskussion

Nachdem im vorherigen Kapitel die Analyse der Daten erläutert und das Spektrum bestimmt wurde, wird in diesem Kapitel der Fluss mit verschiedenen Messungen und Rechnungen verglichen. Dieses Kapitel gliedert sich in mehrere Teile. Zunächst wird der gemessene Fluss der Myon-Neutrinos mit gemessenen Flüssen aus anderen IceCube-Analysen sowie Analysen anderer Experimente verglichen. Anschließend wird der atmosphärische Teil des Spektrums mit verschiedenen Modellen für die atmosphärische Komponente verglichen. Es wird auch ein Fokus auf das Modell der primären kosmischen Strahlung gelegt.

Der nächste Teil dieses Kapitels ist der astrophysikalischen Komponente dieser Analyse gewidmet. Zu diesem Zweck wird die Analyse mit der Analyse mit starrenden Ereignissen aus [A⁺14b] als Vergleich genutzt. Um die Eigenschaften des astrophysikalischen Spektrums besser bestimmen zu können, wird anschließend ein Hypothesentest durchgeführt.

5.1. Vergleich mit anderen Analysen

Zunächst wird die vorgestellte Analyse mit anderen Entfaltungen verglichen. Dazu ist zunächst in Abbildung 5.1 der Vergleich zwischen der hier präsentierten Analyse und zwei weiteren IceCube-Analysen aus [Ruh13] und [Sch14] zu sehen. Im Allgemeinen ist eine Kompatibilität zwischen den Analysen zu sehen. Der erste Datenpunkt ist in der hier präsentierten Analyse geringer als in den anderen beiden Analysen. Dies ist kein signifikanter Effekt und wahrscheinlich statistischer Natur.

Es ist anzumerken, dass zum einen die Energieauflösung und die Fehler im Vergleich zur IC 59-Analyse deutlich verbessert werden konnten. Zum anderen konnte

5. Diskussion

auch der Energiebereich, in dem entfaltet wurde, durch den größeren Stichprobenumfang vergrößert werden.

Im Bereich oberhalb von 100 TeV sind die Punkte der hier vorgestellten IC 79-Analyse oberhalb der beiden anderen Analysen. Dies ist jedoch nicht signifikant. Im IC 79-Datensatz ist entweder eine stärkere astrophysikalische Komponente oder eine statistische Überfluktuation der prompten bzw. astrophysikalischen Komponente im Datensatz enthalten.

In Abbildung 5.2 ist der in dieser Arbeit gemessene Fluss mit Messungen von anderen Experimenten verglichen. In dieser Abbildung wurden Daten von AMANDA [A⁺ 10b], Frejus [D⁺95] sowie ANTARES [AM⁺ 12] genutzt. Die überlappenden Bins aus der Frejus-Analyse sind auf zwei unterschiedliche Entfaltungen mit vollständig und nicht vollständig enthaltenen Ereignissen zurückzuführen.

Im Allgemeinen sind die Datensätze ebenfalls kompatibel. Die Flusspunkte im Nieder-Energiebereich des Spektrums liegen leicht unter denen von ANTARES, sind jedoch mit ihnen kompatibel. Im Energiebereich von 1 TeV bis 10 TeV sind sich alle Messungen sehr ähnlich.

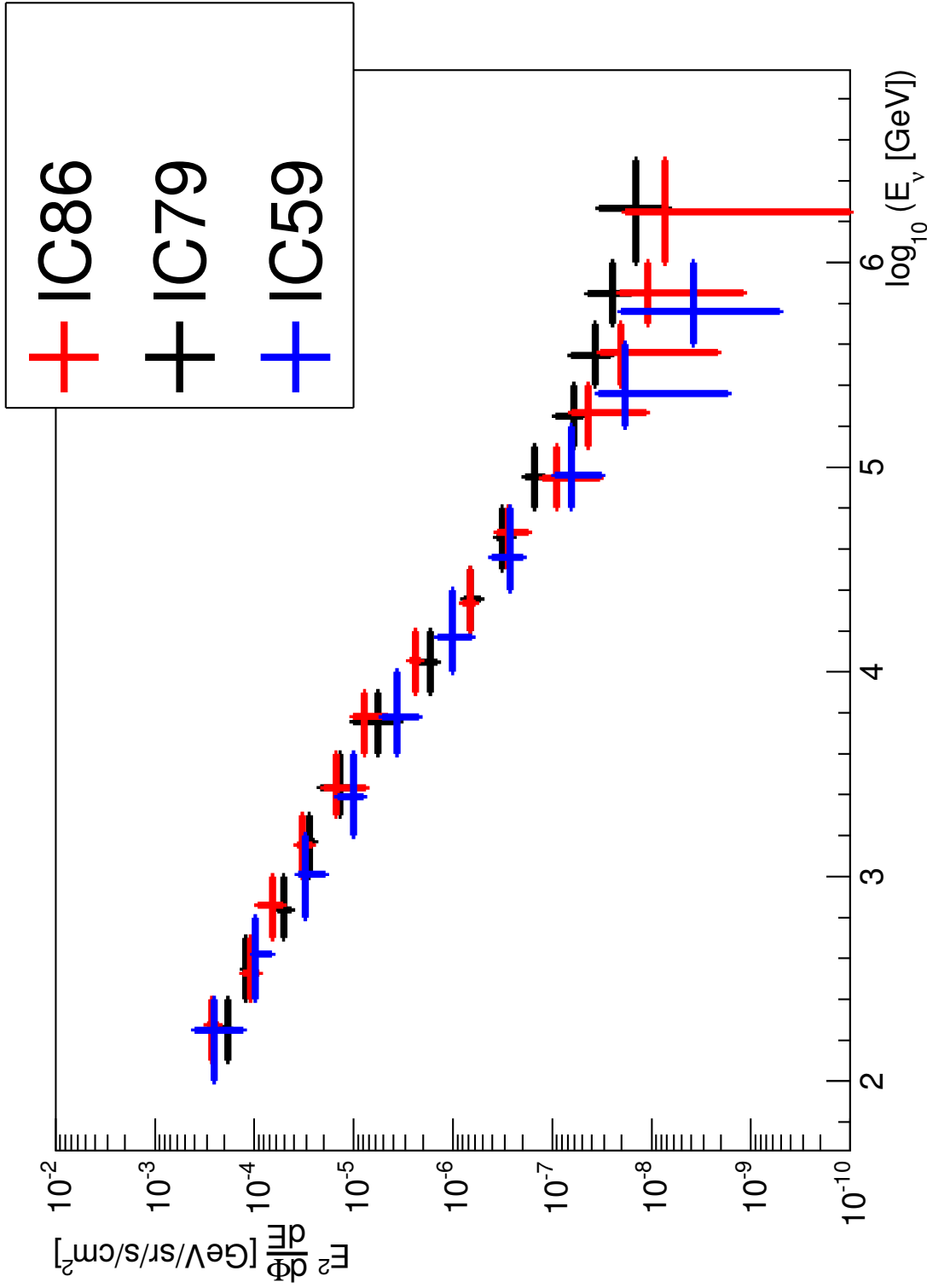


Abbildung 5.1.: Gemessener und mit E^2 skaliertes Fluss dieser Analyse im Vergleich zu anderen Messungen des Myon-Neutrino-Spektrums mit IceCube. Die verwendeten Daten stammen aus [Ruh13] und [Sch14]. Es ist eine Übereinstimmung zwischen den Datensätzen zu beobachten. Im Hochenergiebereich ist die Entfaltung der IC 79-Daten leicht höher als die der IC 86-Daten.

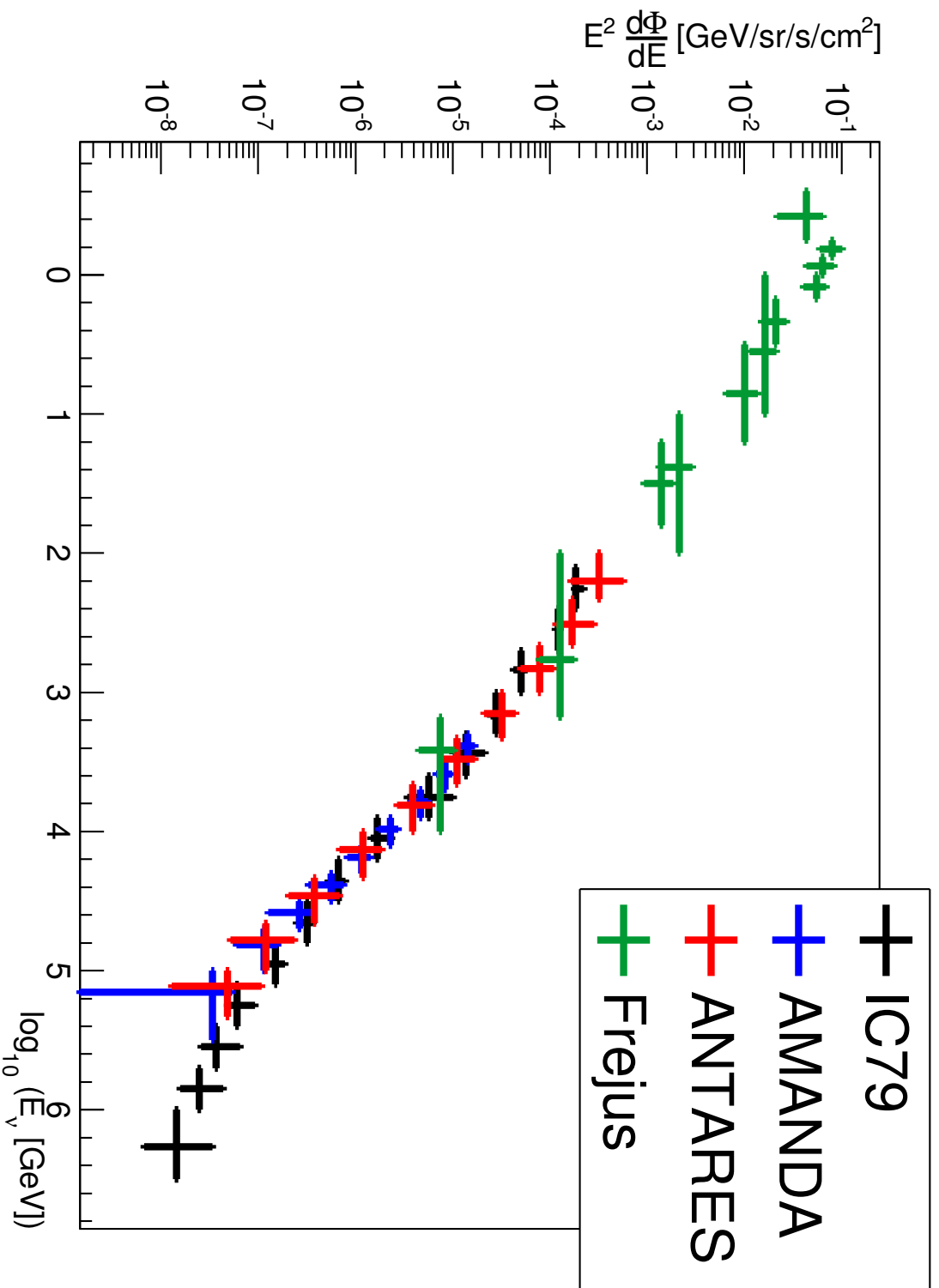


Abbildung 5.2.: Gemessener und mit E^2 skaliertes Fluss dieser Analyse im Vergleich zu anderen Messungen des Myon-Neutrino-Spektrums anderer Experimente. Die Daten kommen hier aus [A⁺ 10b], [D⁺ 95] sowie [AM⁺ 12].

5.2. Vergleich zu verschiedenen theoretischen Modellen

Neben den Messdaten stehen auch theoretische Modelle zur Verfügung. In diesem Abschnitt wird dabei der Fokus auf die konventionelle atmosphärische Komponente gelegt. Die prompte Komponente wird im nächsten Abschnitt im Detail analysiert. In den Abbildungen 5.3, 5.4 und 5.5 sind die Messdaten von IceCube aus dieser Analyse sowie aus [Ruh13] und [Sch14] abgebildet. In Abbildung 5.3 wurde SIBYLL-2.1 [AEG⁺09] als hadronisches Interaktionsmodell gewählt. Für das primäre Flussmodell sind verschiedene Kurven nach [Hör03], [GST13], [A⁺11b] sowie [FTD12] abgebildet. Die Raten, die von diesem hadronischen Wechselwirkungsmodell vorhergesagt werden, sind höher als bei anderen Modellen. Vor allem im Energiebereich unter 1 TeV macht sich dies bemerkbar. Die Kurven liegen in diesem Energiebereich mit allen primären Modellen oberhalb der Messwerte.

Wird QGSJet-II [Ost06] als Interaktionsmodell genutzt, sind die vorhergesagten Flüsse geringer. Die entsprechenden Kurven sind für verschiedene primäre Wechselwirkungsmodelle in Abbildung 5.4 zu sehen. Im Energiebereich unter 1 TeV passen die theoretischen Rechnungen besser mit den Flusspunkten überein. Einzige Ausnahme ist wieder der erste Messpunkt aus der IC 79-Analyse. Dieser liegt leicht unterhalb der Kurven. Dies ist jedoch kein signifikanter Effekt und wahrscheinlich statistischer Natur.

In Abbildung 5.5 sind die Modelle aus [HKK⁺07] dargestellt. Da die veröffentlichten Datenpunkte nur bis 10 TeV reichen, wird der Fluss für den IceCube-Energiebereich extrapoliert und für verschiedene primäre Modelle umgewichtet [Sch13]. Die Varianz in den Primärmodellen aus dieser Rechnung ist geringer als bei den Rechnungen aus [FTD12]. Der Energiebereich unterhalb von 1 TeV ist ähnlich wie bei den Rechnungen mit QGSJet-II und somit kompatibel mit den Messdaten.

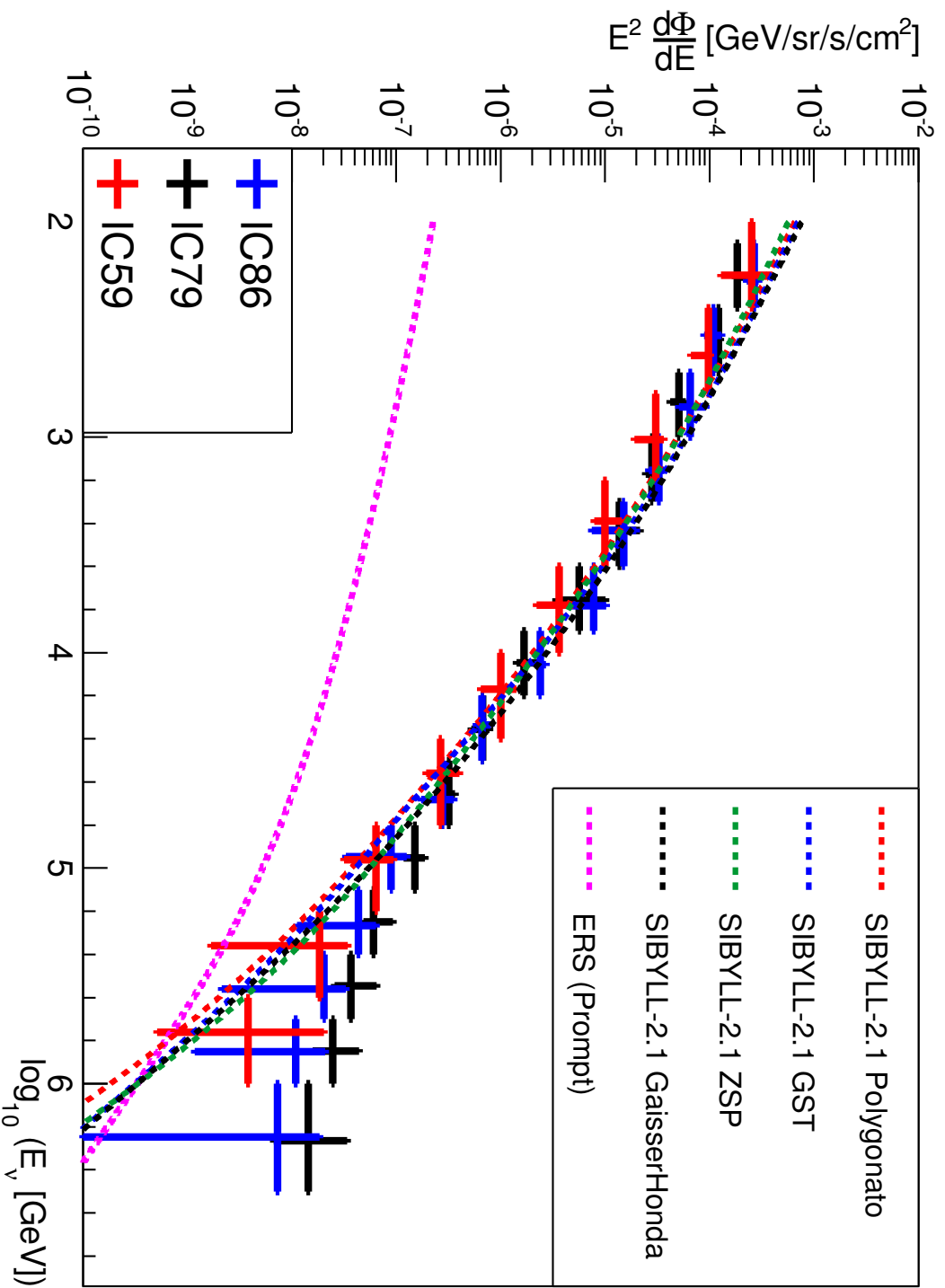


Abbildung 5.3.: Vergleich zwischen Rechnungen aus [FTD12] mit SIBYLL-2.1 als hadronisches Interaktionsmodell und aktuellen IceCube-Messungen [Ruh13][Sch14]. Für das Modell der primären kosmischen Strahlung wurden verschiedene Modelle aus [Hör03], [GST13], [A⁺11b] sowie [FTD12] genutzt. Für die prompte Komponente ist die Rechnung aus [ERS08] eingezeichnet.

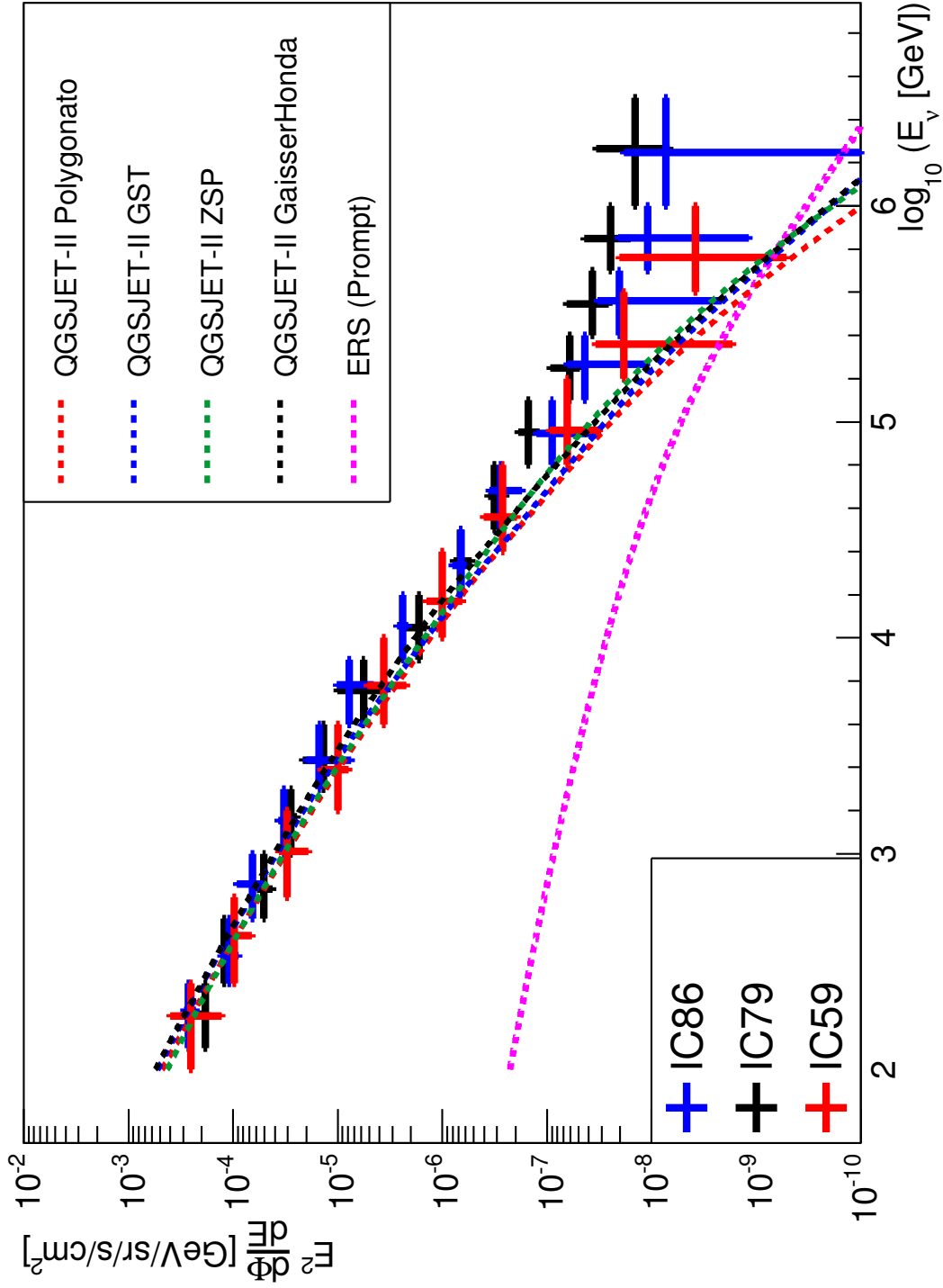


Abbildung 5.4.: Vergleich zwischen Rechnungen aus [FTD12] mit QGSJet-II [Ost06] als hadronisches Interaktionsmodell und aktuellen IceCube-Messungen [Ruh13][Sch14]. Für das Modell der primären kosmischen Strahlung wurden verschiedene Modelle aus [Hör03], [GST13], [A⁺11b] genutzt. Für die prompte Komponente ist die Rechnung aus [ERS08] eingezeichnet.

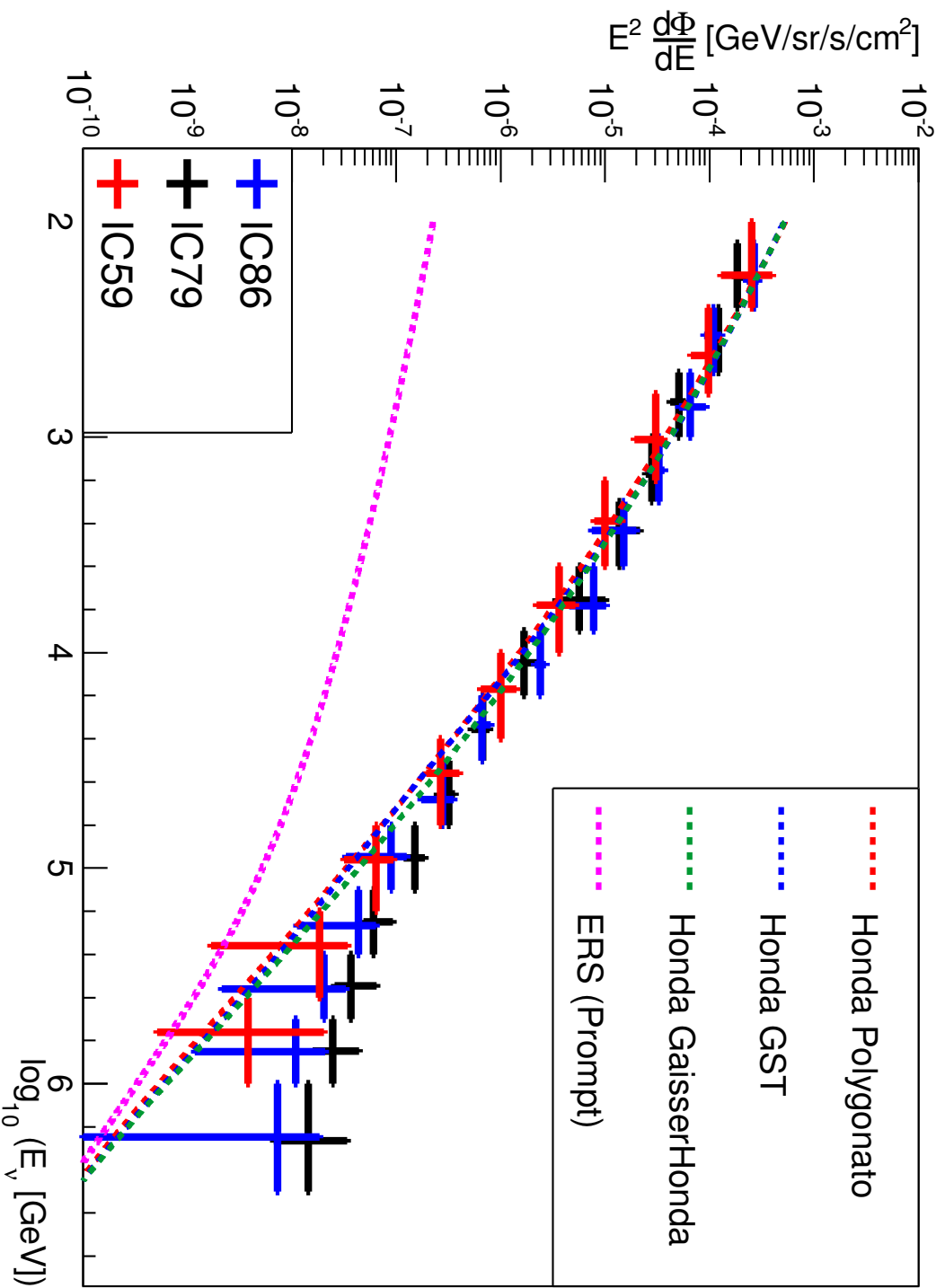


Abbildung 5.5.: Vergleich zwischen Rechnungen aus [HKK⁺07] und aktuellen IceCube-Messungen [Ruh13][Sch14]. Da die Ursprungsrechnungen nur für einen Energiebereich bis 10 TeV geeignet sind, werden diese extrapoliert [Sch13]. Für das Modell der primären kosmischen Strahlung wurden verschiedene Modelle aus [Hör03], [GST13] sowie [FTD12] genutzt. Für die prompte Komponente ist die Rechnung aus [ERS08] eingezeichnet.

5.3. Kombination der Ergebnisse der Analysen mit IC 79 und IC 86

Da die Analysen der Datensätze von IC 79 und IC 86 im selben Winkelbereich und im selben Energiebinning vorliegen, können diese kombiniert werden, um so ein Spektrum mit kleineren Fehlern zu erlangen. Hierzu wird über

$$\Phi_i = \frac{1}{w_{i,IC79} + w_{i,IC86}} (w_{i,IC79} \Phi_{i,IC79} + w_{i,IC86} \Phi_{i,IC86}) \quad (5.3.1)$$

der gewichtete Mittelwert zwischen den Entfaltungsergebnissen gebildet. Hierbei ist Φ_i der gemittelte Fluss im i -ten Bin, $\Phi_{i,X}$ der gemessene Fluss der jeweiligen Messung im i -ten Bin und

$$w_i = \frac{1}{\sigma_{i,X,stat}^2} \quad (5.3.2)$$

das Quadrat des zugehörigen inversen statistischen Fehlers.

Für die Fehlerrechnung wird der komplette Fehler, also inklusive systematischen Fehler genutzt, und mittels Gaußscher Fehlerfortpflanzung fortgepflanzt. Für den Fehler ergibt sich somit:

$$\sigma_{\Phi_i} = \frac{1}{w_{i,IC79} + w_{i,IC86}} \sqrt{w_{i,IC79}^2 \sigma_{i,IC79}^2 + w_{i,IC86}^2 \sigma_{i,IC86}^2}. \quad (5.3.3)$$

Hierbei bezeichnet $\sigma_{i,X}$ den Fehler inklusive systematischen Fehler der jeweiligen Messung im i -ten Bin.

Die Ergebnisse sind in den Abbildungen 5.6, 5.7 und 5.8 gegen Rechnungen aus [FTD12] und [HKK⁺07] für verschiedene Primärmodelle aufgetragen. Die Werte sind in Anhang D.1 zu sehen. Die kleineren Fehler machen sich im ganzen Energiebereich bemerkbar. Auch hier passt der Energiebereich unter 1 TeV eher zu dem Modell mit QGSJet-II. Für die Ergebnisse oberhalb dieser Grenze passt das Modell mit SIBYLL-2.1 besser.

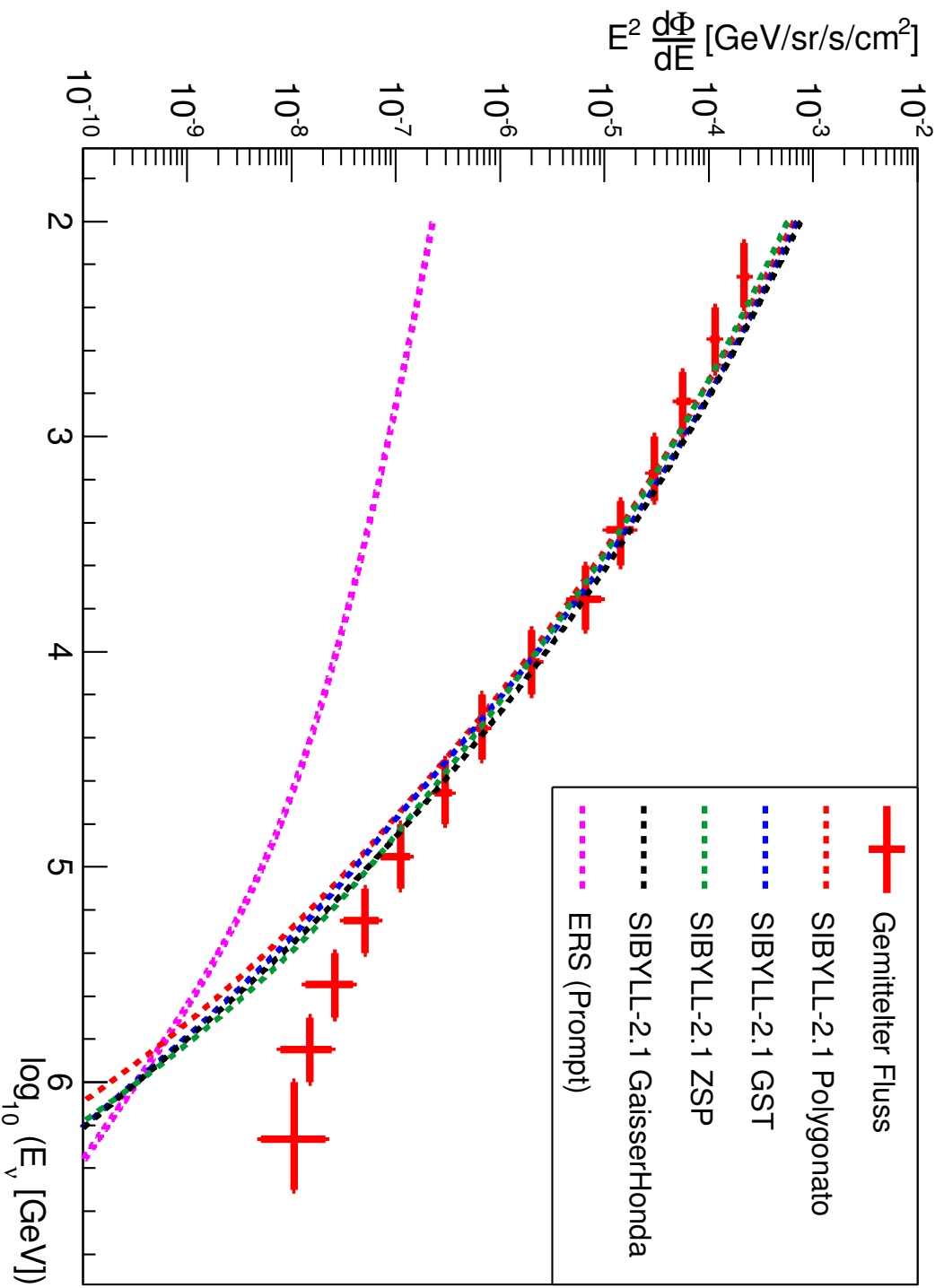


Abbildung 5.6.: Vergleich zwischen Rechnungen aus [FTD12] mit SIBYLL-2.1 als hadronisches Interaktionsmodell und dem gemittelten Fluss aus den IC 79- und IC 86-Analysen. Für das Modell der primären kosmischen Strahlung wurden verschiedene Modelle aus [Hör03], [GST13], [A⁺11b] sowie [FTD12] genutzt. Für die prompte Komponente ist die Rechnung aus [ERS08] eingezeichnet.

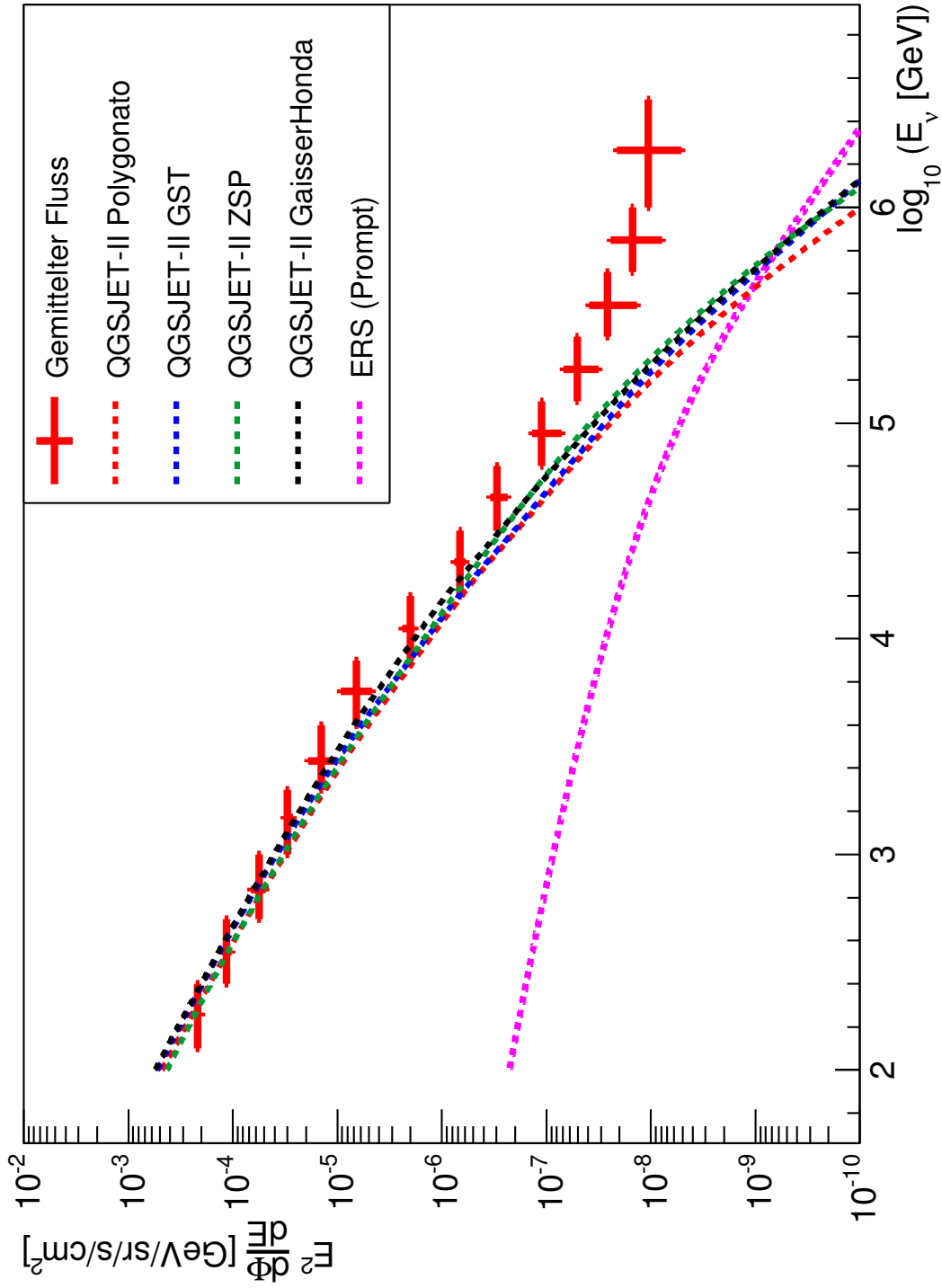


Abbildung 5.7.: Vergleich zwischen Rechnungen aus [FTD12] mit QGSJet-II [Ost06] als hadronisches Interaktionsmodell und dem gemittelten Fluss aus den IC 79- und IC 86-Analysen. Für das Modell der primären kosmischen Strahlung wurden verschiedene Modelle aus [Hör03], [GST13], [A⁺11b] sowie [FTD12] genutzt. Für die prompte Komponente ist die Rechnung aus [ERS08] eingezeichnet.

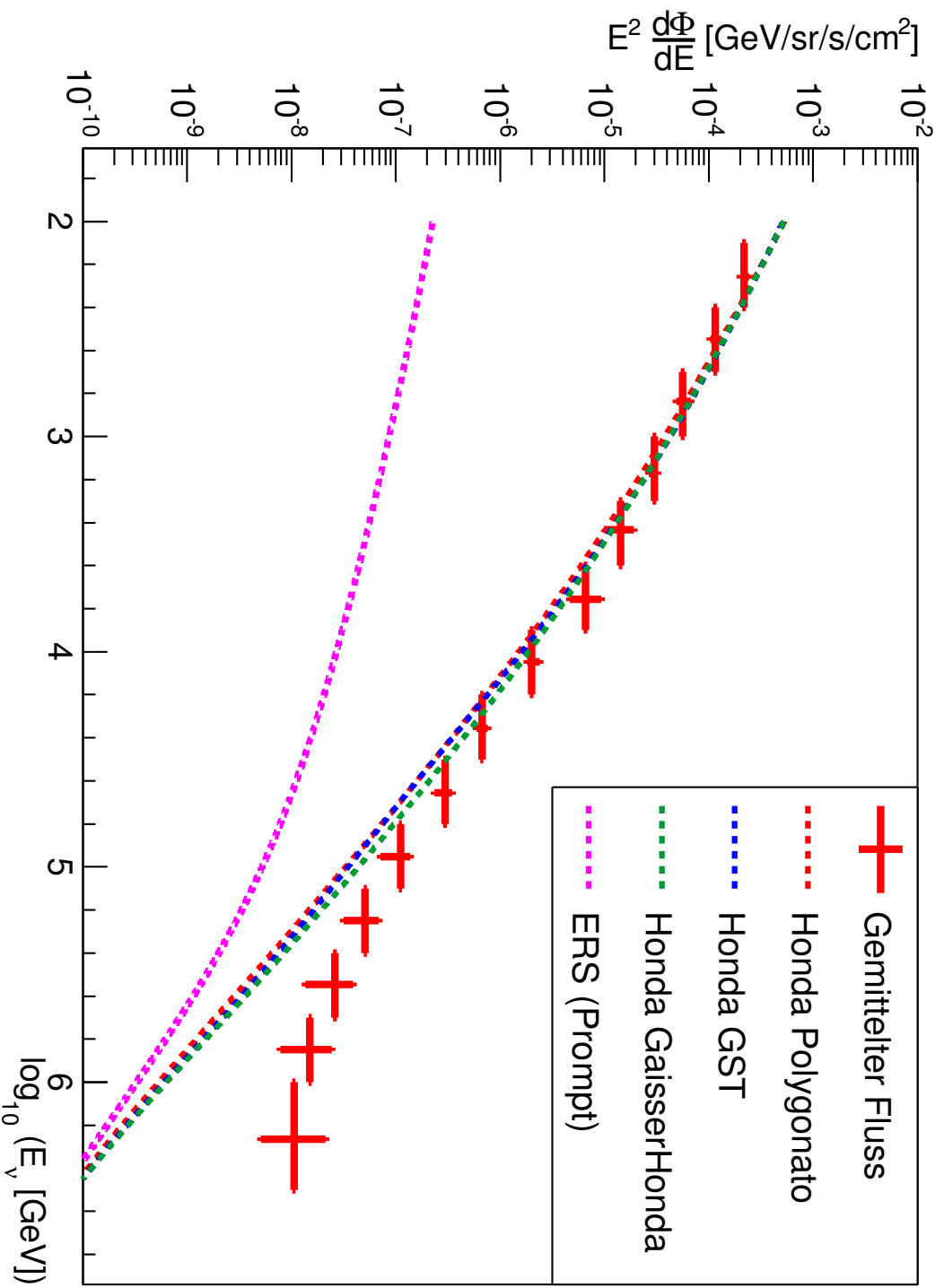


Abbildung 5.8: Vergleich zwischen Rechnungen aus [HKK⁺07] und dem gemittelten Fluss aus den IC 79- und IC 86-Analysen. Da die Ursprungsrechnungen nur für einen Energiebereich bis 10 TeV geeignet sind, werden diese extrapoliert [Sch 13]. Für das Modell der primären kosmischen Strahlung wurden verschiedene Modelle aus [Hör03], [GST13] sowie [FTD12] genutzt. Für die prompte Komponente ist die Rechnung aus [ERS08] eingezeichnet.

5.4. Diskussion der astrophysikalischen und prompten Komponente

Im Bereich oberhalb von 100 TeV ist eine weitere Komponente zu sehen. Im Folgenden wird diese diskutiert. Zunächst wird der Frage nachgegangen, ob diese Komponente prompten und somit irdischen Ursprungs sein kann. Anschließend werden die Entfaltungsergebnisse noch mit den aktuellen Ergebnissen aus [A⁺ 14b] verglichen.

5.4.1. Erklärung durch eine hohe prompte Komponente

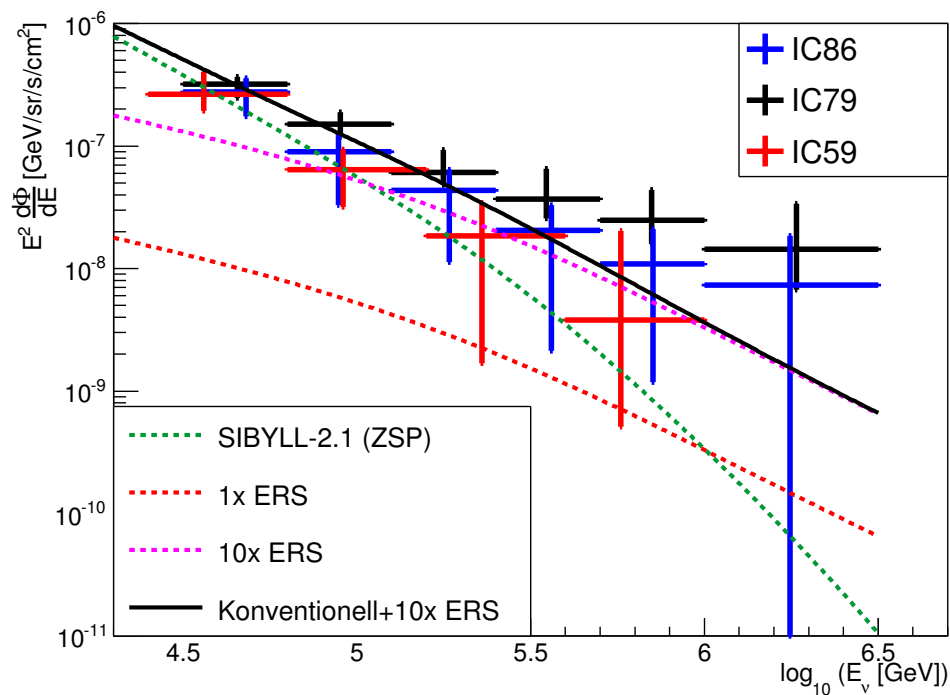


Abbildung 5.9.: Vergleich zwischen aktuellen IceCube-Messungen dieser Analyse und aus [Ruh13] und [Sch14] mit prompten Flüssen. Der prompte Fluss besitzt eine grosse Unsicherheit, daher wurde auch der 10-fache Standardfluss nach [ERS08] gezeichnet. Die konventionelle Komponente ist aus [FTD12] mit SIBYLL-2.1 als hadronisches Interaktionsmodell. Als primäres Flussmodell wird ZSP [A⁺ 11b] genutzt.

In Abbildung 5.9 sind aktuelle IceCube-Messdaten im Vergleich zu den Rechnungen aus [ERS08] dargestellt. Es wurde, da der Fluss eine große Unsicherheit besitzt, auch der theoretische prompte Fluss mit einer 10-fachen Normierung eingezeichnet. Es ist ersichtlich, dass die konventionelle Komponente bei einer Energie über 1 PeV keinen signifikanten Beitrag mehr zum Fluss liefert. Um den gemessenen Fluss ohne einen astrophysikalischen Fluss zu erklären, ist es also notwendig, einen hohen prompten Fluss anzunehmen. Auch bei einer 10-fachen prompten Komponente können die Daten aus der Analyse des IC 79-Datensatzes nicht vollständig erklärt werden. Bei höheren Normierungen als diese werden Defekte im Bereich um 100 TeV sichtbar. Daher ist eine Normierung, welche höher als die 10-fache Standardnormierung aus [ERS08] ist, unwahrscheinlich. Vor allem die beiden Messpunkte im PeV-Bereich aus der IC 79-Messung machen eine Erklärung des Flusses durch eine reine prompte Komponente unwahrscheinlich.

5.4.2. Vergleich mit der Analyse mit startenden Ereignissen

Ein weiterer Vergleich ist der Vergleich mit der Messung aus [A⁺14b], also der Analyse von startenden Ereignissen in 3 Jahren IceCube-Daten. Jedes Bin in dieser Entfaltung ist ein Potenzgesetz der Form

$$\Phi_i = N_i E^{-2}, \quad (5.4.1)$$

wobei N_i die Normierung des Potenzgesetzes im i -ten Bin ist. Die Ergebnisse sind im Vergleich zu den Messungen aus dieser Analyse sowie der Analyse aus [Sch14] in Abbildung 5.10 zu sehen. Bei der Interpretation dieser Daten ist es wichtig, das Selbstveto [GJKvS14] zu beachten. Mit der zuvor bereits erwähnten Methode ist in der Analyse von startenden Ereignissen die atmosphärische Komponente stark unterdrückt. Dies ist in Abbildung 5.10 auch deutlich zu sehen. Die Analyse von startenden Ereignissen liegt unterhalb von ~ 300 TeV deutlich unter den Ergebnissen von Flussrechnungen für den atmosphärischen Neutrinofluss und auch den Messergebnissen aus Entfaltungen des Myon-Neutrinos-Spektrums in diesem Energiebereich.

Im Bereich, in dem die atmosphärische Komponente jedoch vernachlässigbar ist, also dem Bereich oberhalb von 1 PeV, sind die Messergebnisse aus den drei Analysen äquivalent.

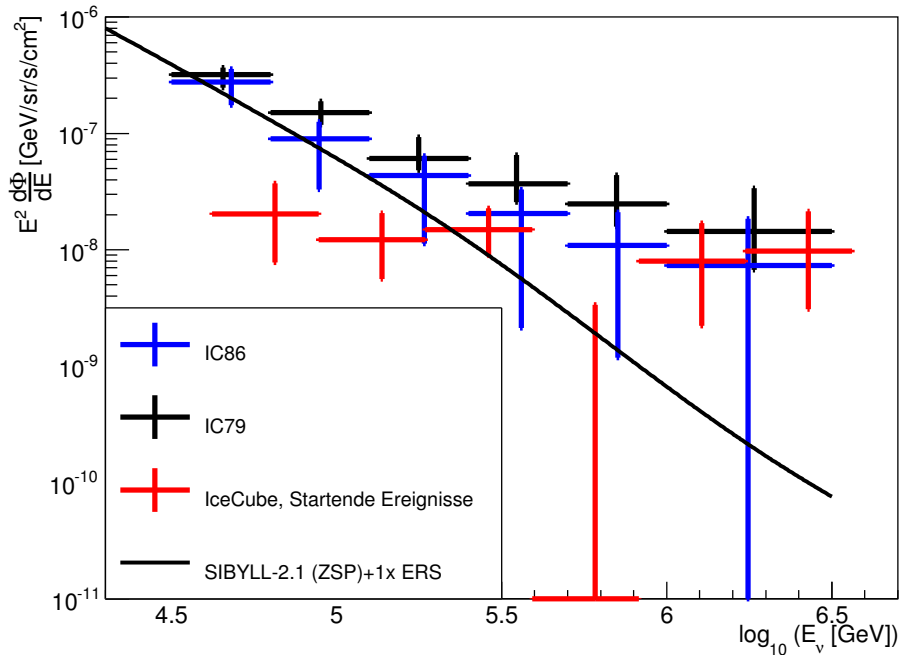


Abbildung 5.10.: Vergleich zwischen aktuellen IceCube-Messungen aus dieser Analyse und aus [Sch14] sowie den entfaltetten Ergebnissen aus [A⁺14b]. Als Vergleich ist die Summe aus konventioneller Komponente aus [FTD12] mit SIBYLL-2.1 als hadronisches Interaktionsmodell und ZSP [A⁺11b] als primäres Flussmodell sowie einer einfachen prompten Komponente nach [ERS08] eingezeichnet.

5.5. Bestimmung physikalischer Parameter

Um die astrophysikalische Komponente weiter zu charakterisieren, wird eine Minimierung genutzt, um die optimale Kombination freier Parameter zu finden. Die freien Parameter dieser Minimierung sind dabei die Normierungen der astrophysikalischen Komponente N_{Astro} , der spektrale Index der astrophysikalischen Komponente γ , die Normierung der prompten Komponente N_{prompt} , die Normierung der atmosphärischen Komponente N_{Atmo} sowie das primäre Flussmodell und das hadronische Wechselwirkungsmodell. Die Gesamtheit dieser Parameter wird als ϑ bezeichnet. Für die Minimierung werden die Daten aus IC 59, IC 79 und IC 86 genutzt.

5.5.1. Methode

Zunächst wird eine Wahrscheinlichkeit für eine Hypothese H berechnet. Hierzu wird der Abstand zwischen den Datenpunkten und der Summe der Flüsse unter dieser Hypothese gebildet.

$$\Delta\Phi_{\sigma,i} = \frac{\Phi_i - \tilde{\Phi}_{\partial|H}}{\sigma_i}. \quad (5.5.1)$$

Hierbei ist Φ_i ein Messungswert, $\tilde{\Phi}_{\partial|H}$ die Summe der theoretischen Flüsse unter der Hypothese H und σ_i der jeweilige Fehler des gemessenen Flusses. Die Wahrscheinlichkeit, dass der Datenpunkt mit der Theorie kompatibel ist, ergibt sich somit aus

$$p_i = 1 - \int_{-\infty}^{\Delta\Phi_{\sigma,i}} \mathcal{N}(1, 0) dx. \quad (5.5.2)$$

Die Standardnormalverteilung wird als $\mathcal{N}(1, 0)$ bezeichnet. Mit Hilfe dieser Wahrscheinlichkeiten lässt sich eine Likelihood-Funktion über

$$\mathcal{L} = \prod_i p_i(\partial|H) \quad (5.5.3)$$

definieren. Das weitere Vorgehen ist das Übliche für den Vergleich von Likelihoods (siehe zum Beispiel [BL98]). Es wird zunächst das Verhältnis der Likelihoods für die Signalthypothese H und der Nullhypothese H_0 gebildet:

$$\Lambda = \frac{\mathcal{L}(\partial|H_0)}{\mathcal{L}(\partial|H)}. \quad (5.5.4)$$

Im vorliegenden Fall steht die Nullhypothese für eine Erklärung des Flusses ohne eine astrophysikalische Komponente.

Mit Hilfe des Theorems von Wilk [Wil38] lässt sich aus diesem Verhältnis eine Signifikanz für die Hypothese H berechnen. Wilks Theorem besagt, dass

$$f(-2 \ln \Lambda) \sim \chi^2_{L-M} \quad (5.5.5)$$

gilt, wobei L die Freiheitsgrade der Signalthypothese und M die Freiheitsgrade der Nullhypothese sind und f die Verteilungsfunktion bezeichnet.

5. Diskussion

Aus dieser Näherung ergibt sich auch, dass die 1σ -Umgebung an dem Punkt x gegeben ist, für den gilt:

$$\mathcal{L}_{max} - \mathcal{L}_x = \frac{1}{2}. \quad (5.5.6)$$

Für das hadronische Wechselwirkungsmodell wurden die Modelle SIBYLL-2.1 und QGSJet-II genutzt. Das Primärmodell kann entweder Polygonato, GaisserHonda, GST oder aber ZSP sein. Die atmosphärische Normierung kann über den Bereich zwischen 0.7 und 1.3 variieren, während die prompte Komponente eine bis zu 20-fache Standardnormierung annehmen kann. Für die astrophysikalische Komponente wird ein Potenzgesetz der Form

$$E^2\Phi = N_{\text{Astro}} \left(\frac{E}{1 \text{ PeV}} \right)^{2+\gamma} \quad (5.5.7)$$

angenommen. Der Index γ kann zwischen -1.0 und -3.5 variieren, während die Normierung, N_{Astro} , maximal bei $2 \cdot 10^{-8} \frac{\text{GeV}}{\text{sr s m}^2}$ liegt.

5.5.2. Ergebnisse

Parameter	Wert der besten Optimierung	1σ -Bereich
γ	-2.11	[-2.74 , -1.45]
N_{Astro}	1.13	[0.66 , 1.70]
N_{Prompt}	8.28	[4.65 , 11.31]
N_{Atmo}	0.89	[0.86 , 1.05]
Wechselwirkungsmodell	ZSP	
Interaktionsmodell	SIBYLL-2.1	

Tabelle 5.1.: Ergebnisse der Minimierung für die verschiedenen Parameter.

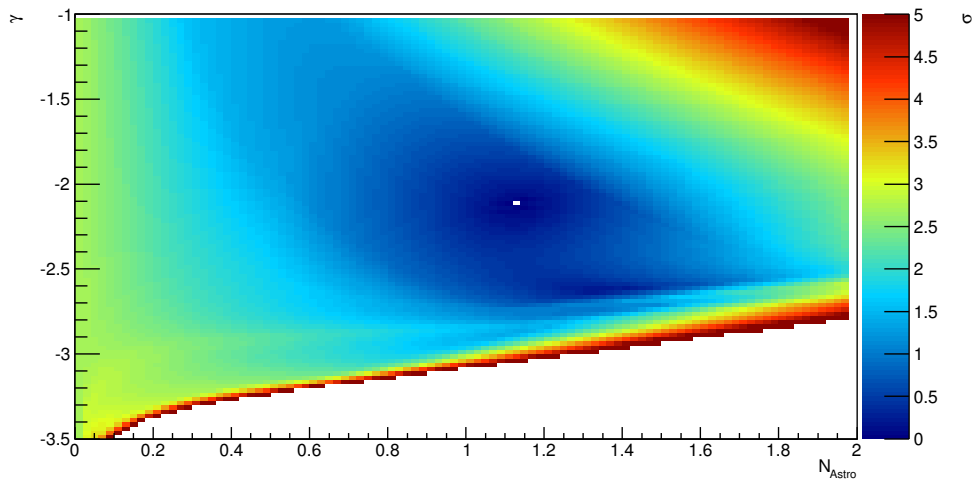
Die Ergebnisse der Minimierung sind in Tabelle 5.1 mit Angabe des 1σ -Bereichs angegeben. Es ist natürlich zu beachten, dass die Ergebnisse korreliert sind. Aus diesem Grund ist es wichtig nicht nur die eindimensionale Projektion zu betrachten, sondern auch zweidimensionale Projektionen der Log-Likelihood-Landschaft. Hierbei wird für jeden Punkt in den beiden Parametern der jeweils beste Wert in den anderen Parametern aufgetragen.

Die Projektionen für γ gegen N_{Astro} , sowie γ gegen N_{Prompt} sind in Abbildung 5.11 dargestellt. In Abbildung 5.12 ist die Projektion in die $N_{\text{Prompt}}-N_{\text{Astro}}$ -Ebene zu

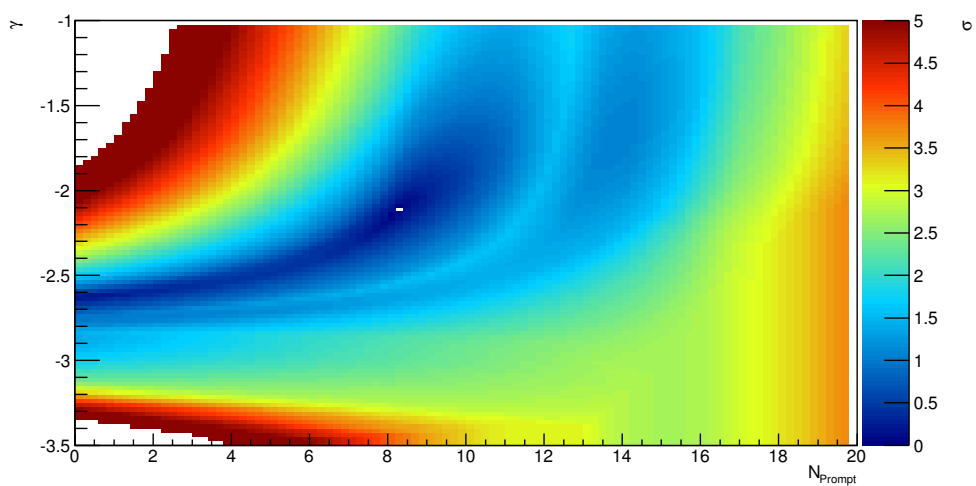
sehen. Auf den Farbachsen steht jeweils die Signifikanz der Hypothese im Vergleich zum optimalen Fit. Vor allem in Abbildung 5.11(b) wird die Schwierigkeit deutlich, die auftritt, wenn die astrophysikalischen Parameter bestimmt werden sollen. Es ist eine sehr starke Abhängigkeit zwischen dem gewählten spektralen Index und der Normierung der prompten Komponente zu sehen. Es gibt einen weiten Bereich an erlaubten prompten Normierungen in Abhängigkeit der astrophysikalischen Komponente. Dies ist auch in den 1σ -Bereichen in Tabelle 5.1 ersichtlich. Um die astrophysikalische Komponente weiter zu charakterisieren, ist eine genauere Kenntnis der prompten Komponente von entscheidender Bedeutung.

Die maximal mögliche prompte Komponente ist auch für die Signifikanz gegenüber der Nullhypothese wichtig. Für eine 20-fache prompte Normierung ergibt sich eine Signifikanz von 1.93σ . Für eine prompte Komponente von maximal 5-facher Standardnormierung ergeben sich 4.60σ und für eine reine konventionelle Komponente ergeben sich 7.22σ . Eine Übersicht für die Signifikanz in Abhängigkeit der maximalen prompten Normierung ist in Abbildung 5.13 zu sehen. Für prompte Normierungen bis hin zu 8.7-facher Standardnormierung werden Signifikanzen oberhalb von 3σ erreicht. Damit das Ergebnis für die astrophysikalische Komponente nicht signifikant ist, müsste die prompte Komponente deutlich höher sein, als sie in [ERS08] berechnet wurde. Eine solche sehr starke prompte Komponente wäre ebenfalls eine wichtige Entdeckung.

5. Diskussion



(a)



(b)

Abbildung 5.11.: Projektionen auf die a) γ - N_{Astro} -Ebene, sowie auf die b) γ - N_{Prompt} -Ebene der Log-Likelihood-Landschaft. Auf der Z-Achse ist die Signifikanz in Bezug auf den besten Fit zu sehen. Der weiße Punkt repräsentiert den Punkt der besten Optimierung.

5. Diskussion

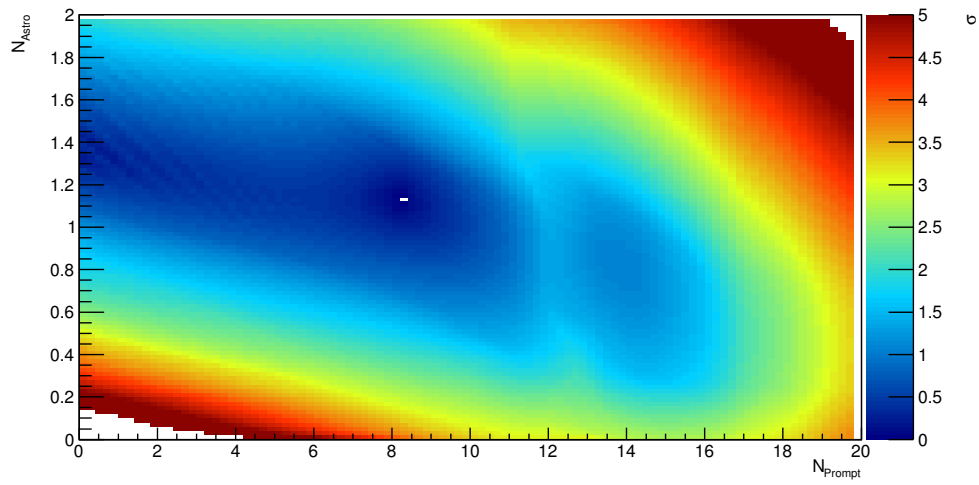


Abbildung 5.12.: Projektion auf die N_{Prompt} - N_{Astro} -Ebene der Log-Likelihood-Landschaft. Auf der Z-Achse ist die Signifikanz in Bezug auf den besten Fit zu sehen. Der weiße Punkt repräsentiert den Punkt der besten Optimierung.

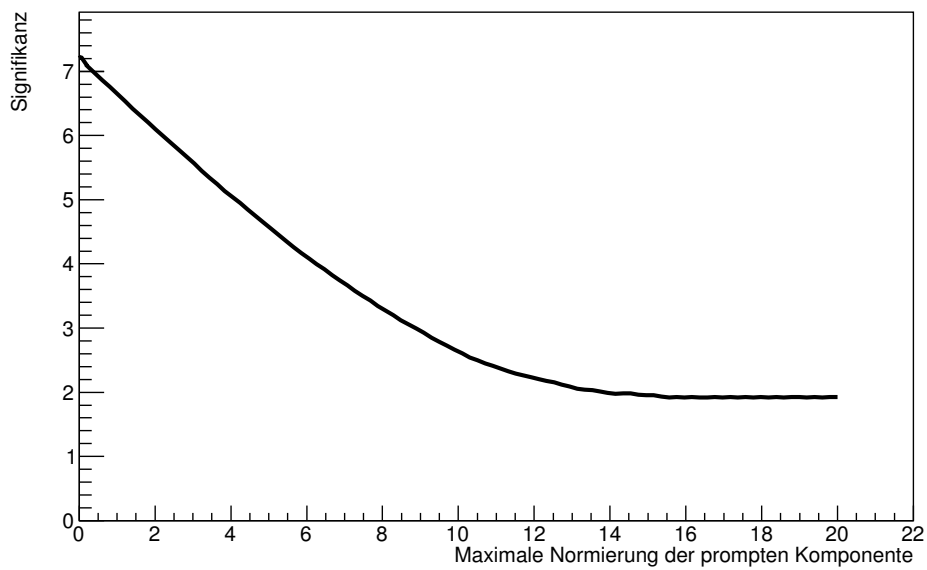


Abbildung 5.13.: Signifikanz in Abhängigkeit der Nullhypothese für verschiedene maximale prompte Normierungen der Nullhypothese.

6. Zusammenfassung und Ausblick

Zu Beginn dieser Arbeit wurde die Bedeutung der Neutrinoastronomie und die Verbindung zu anderen Gebieten der Astroteilchenphysik verdeutlicht. Der IceCube-Detektor, mit dem diese Arbeit ermöglicht wurde, wurde diskutiert. Auch wurde das Detektionsproblem erläutert, welches auf den niedrigen Wirkungsquerschnitt zwischen Neutrinos und Materie und dem hohen Untergrund zurückzuführen ist.

Es wurden weiterhin Methoden des maschinellen Lernens vorgestellt, welche in der Lage sind das Problem des hohen Untergrunds zu überwinden. Diese Methoden wurden auf die Daten des IC 79-Datensatzes angewandt und es konnte ein hochreiner Datensatz an Myon-Neutrinos erzeugt werden. Mit Hilfe neu entwickelter Validierungsmethoden wurde der verbleibende Untergrund untersucht. Es konnte gezeigt werden, dass der verbleibende Untergrund niederenergetisch ist. Bei der Analyse des Datensatzes zeigt sich ein deutlicher Überschuss an Neutrinos bei sehr hohen rekonstruierten Myonenergien. Bei näherer Untersuchung konnte gezeigt werden, dass dieser Untergrund keine typischen Untergrund-Eigenschaft aufweist. Er liegt weder am Rande des Detektors, noch kommt er aus Richtung des Horizonts. Dadurch ist eine Erklärung des Überschusses ausgeschlossen.

Mit der Methode der Entfaltung wurde eine Möglichkeit aufgezeigt, aus diesem Datensatz ein Spektrum in der Neutrinoenergie zu berechnen, obwohl die Neutrinoenergie nicht direkt messbar ist. Diese Methode wurde auf den zuvor erstellten Datensatz angewandt und es zeigt sich, dass auch in der Neutrinoenergie ein deutlicher Überschuss oberhalb von 100 TeV zu beobachten ist.

Die erzielten Ergebnisse wurden mit anderen Messungen verglichen. Die Messungen im Bereich der atmosphärischen Neutrinoenergie sind kompatibel mit vorherigen Messungen. Die Ergebnisse oberhalb von 100 TeV sind nicht durch konventionelle atmosphärische Neutrinos zu erklären. Die Anpassung und der statistische Test über einen Likelihood-Quotienten-Test ergibt, dass es höchst-

wahrscheinlich eine astrophysikalische Komponente gibt. Die Signifikanz dieser Aussage ist abhängig davon, wie hoch die Normierung der prompten Komponente ist. Mit der Annahme aus [ERS08] ergibt sich eine Signifikanz von 6.65σ , dass der gemessene Datensatz nicht irdischen Ursprungs ist.

Eine abschließende Entscheidung wie stark die prompte Komponente ist, kann mit den vorliegenden Messdaten nicht getroffen werden. Es werden weitere Messdaten und Analysen notwendig sein, um diese beiden Komponenten zu unterscheiden. Hierzu können zunächst einmal die Daten weiterer Jahre analysiert werden. Die Datenpunkte in dem Bereich, in dem die astrophysikalische Komponente dominant wird, ist bisher mit großen statistischen Fehlern belegt. Durch eine längere Messdauer werden diese Fehler natürlich kleiner und damit wird eine besser Charakterisierung des Flusses möglich.

Die Daten für ein Jahr IceCube umfassen ca. 100 000 Ereignisse. Mit dieser Menge an Daten ist es möglich, das Spektrum der atmosphärischen Neutrinos auch zweidimensional, also in Energie und Zenitwinkel, zu untersuchen. Auf diese Art kann der atmosphärische Untergrund besser definiert werden, wodurch der Einfluss der Unsicherheiten im Bereich der astrophysikalischen Komponente geringer wird.

Um die prompte Komponente weiter einzuschränken, ist es möglich, diese im Myon-Kanal zu untersuchen. In den Zerfällen der charmhaltigen Teilchen, bei denen Myon-Neutrinos entstehen, werden gleichzeitig auch Myonen mit ähnlichen charakteristischen Eigenschaften erzeugt. Eine Untersuchung dieser Teilchen kann dazu führen, dass die prompten Neutrinos besser verstanden werden und sich somit die Unsicherheiten reduzieren. Diese Verbesserung zieht einen deutlichen Fortschritt in der Vermessung des diffusen astrophysikalischen Neutrinoflusses nach sich.

Weiter ist natürlich nicht nur der diffuse Fluss von Interesse, sondern auch die Quellen dieser Strahlung. Mit Hilfe der Methode des Stackings, also dem gleichzeitigen Betrachten mehrerer Quell-Kandidaten, ist es möglich, den Untergrund weiter zu reduzieren. Werden die Kandidaten für eine solche Suche optimal gewählt, ist somit eine bessere Sensitivität erreichbar, als bei einer einfachen Punkt-Quellen-Suche. Für eine gute Wahl dieser Quellen stehen verschiedene Kataloge zur Verfügung, welche zumeist auf theoretischen Annahmen basieren [A⁺06a]. Auch eine Suche nach zeitlicher Koinzidenz zu Gamma Ray Bursts ist nach dem

Beweis der Existenz von astrophysikalischen Neutrinos von großem Interesse [A⁺12a].

Für weitere Untersuchungen sind natürlich auch weitere Ausbaustufen von IceCube möglich. Generell sind hier zwei Ideen interessant. Die eine ist eine Vergrößerung des Volumens von IceCube durch weitere Strings. Durch diese Maßnahme können mehr Neutrinos detektiert werden und somit wird das Entdeckungspotential erhöht. Alternativ ist es möglich, ein Myon-Veto zu bauen. Dazu wird an der Oberfläche ein Array ähnlich zu IceTop [G⁺03] errichtet. Bei ausreichender Größe ermöglicht dieses Array ein Veto auf atmosphärische Myonen. Mit dieser Hilfe können Neutrinos aus dem Südhimmel deutlich effizienter selektiert werden. Alternativ zu einer Erweiterung von IceCube ist auch eine Erweiterung des ANTARES-Teleskops eine Möglichkeit. Für das KM3NET genannte Projekt wurden bereits Anfang 2014 die ersten optischen Module versenkt [AM⁺13].

Die Ergebnisse dieser Arbeit zeigen auf, dass es astrophysikalische Neutrinos im Myon-Kanal gibt. Die Möglichkeit die Winkelauflösung von Myonen zu nutzen wird es in Zukunft ermöglichen, mit Neutrinos Astronomie zu betreiben. Wie in dieser Arbeit gezeigt wurde, ist das Problem des hohen Myon-Untergrunds und die Rekonstruktion des Energiespektrums überwunden. Mit IceCube-Daten aus mehreren Jahren und weiteren größeren Detektoren wird damit ein neues Zeitalter der Astronomie eingeleitet. Die Entdeckung von Neutrino-Quellen und damit die Entdeckung der Quellen der geladenen kosmischen Strahlung liegt in greifbarer Nähe.

Danksagung

Während ich diese Arbeit verfasst habe, wurde mir die Hilfe vieler Menschen zu Teil. Bei diesen Menschen möchte ich mich für das Gelingen dieser Arbeit bedanken.

Zu aller erst möchte ich mich bei Prof. Dr. Dr. Wolfgang Rhode bedanken. Die exzellente Betreuung während meiner Arbeit am Lehrstuhl für Astroteilchenphysik war maßgeblich für den Erfolg dieser Arbeit.

Außerdem möchte ich mich bei meinem Zweiprüfer Prof. Dr. Carsten Westphal bedanken, der sich die Zeit nahm diese Arbeit zu begutachten.

Weiter möchte ich mich bei Florian Scheriau bedanken, der mir während dieser Arbeit ein konstruktiver und kontroverser Diskussionspartner war und ohne den diese Arbeit nicht so wäre, wie sie ist.

Ein besonderer Dank geht an die Dortmunder Astroteilchenphysikgruppe, die mir sowohl während meiner Promotion als auch während meiner Diplomarbeit immer wieder wertvolle Hilfe leistete. In dieser Arbeitsgruppe habe ich den Feinschliff meiner Ausbildung erhalten.

Weiter möchte ich mich bei den vielen Menschen der IceCube-Kollaboration bedanken, die in unzähligen Meetings und Telefonkonferenzen wichtige Impulse für diese Arbeit gegeben haben. Ohne die erfolgreiche internationale Kooperation wäre diese Arbeit nicht möglich gewesen.

Der wichtigste Dank jedoch geht an meine Familie, ohne die ich niemals an diesen Punkt gelangt wäre. Vielen Dank für die jahrelange Unterstützung.

Abbildungsverzeichnis

2.1. Schematische Darstellung des IceCube-Detektors inklusive IceTop.	3
2.2. Energieverlust von Myonen durch verschiedene Wechselwirkungsprozesse.	5
2.3. Absorptions- und Streukoeffizienten für das WHAM Eismodell in Abhängigkeit von Tiefen und Wellenlänge.	7
2.4. Ereignisse aus der Suche nach Hochenergieneutrinos mittels im Detektor startenden Ereignissen mit dem IceCube-Detektor	8
2.5. Das Energiespektrum der kosmischen Strahlung gemessen mit verschiedenen Experimenten.	11
2.6. Detailliertere Darstellung für die Messpunkte der kosmischen Strahlung im Bereich des Knies.	12
2.7. Verschiedene Parametrisierungen der primären kosmischen Strahlung	13
2.8. Relativer Anteil von Neutrinos und Myonen aus Pion oder Kaonzerfällen am konventionellen Fluss.	15
2.9. Verhältnis von konventionellen atmosphärischen Elektron- und Myon-Neutrinos aus verschiedenen Rechnungen.	16
2.10 Der prompte Neutrinofluss gewichtet mit E^{-3} mit Fehlerband im Vergleich zum konventionellen Neutrinofluss.	17
3.1. Skizze zur Veranschaulichung der Parameter im SPE- bzw. MPE-Fit.	23
3.2. Beispiel eines Entscheidungsbaumes, der auf Daten des IceCube-Detektors entstehen könnte.	27
3.3. Schematische Darstellung einer fünffachen Kreuzvalidierung.	30
3.4. Confidenceverteilung für ein Untergrundereignis, welches nur mit geringer Wahrscheinlichkeit als Signalereignis klassifiziert wird.	32
3.5. Beispiel für eine minimal regularisierte Entfaltung mit TRUEE.	37

3.6. Sechs verschiedene Basisfunktionen $p'_j(x)$ für die in Abbildung 3.5 gezeigte Entfaltung.	40
3.7. Schematische Darstellung des Testmodes.	42
4.1. Überlebenswahrscheinlichkeit für Anti-Neutrinos, die durch die Erde propagieren, mit verschiedenen Energien in Abhängigkeit des Zenitwinkels.	47
4.2. Recall der kombinierten und einzelnen Precuts.	49
4.3. Darstellung der gewählten Precuts.	50
4.4. Stabilität der Merkmalssektion für verschiedene Anzahlen von gewünschten Merkmalen.	51
4.5. Confidence-Verteilung, welche der Random Forest erzeugt.	55
4.6. Rekonstruierte Spurlänge nach einem Schnitt auf eine Confidence von 0.92.	56
4.7. Recall des Random Forests für Myon-Neutrinos für verschiedene Confidence-Schnitte.	57
4.8. Energieschätzung des Myons mit dem MuEX-Algorithmus für zwei verschiedene Einstellungen der DOM-sensitivität.	58
4.9. Signal- und Untergrundverteilungen für MuEX und TruncatedMean	61
4.11. Verteilung der rekonstruierten Myonenergie.	64
4.12. Verteilung der rekonstruierten Myonenergie mit dem Truncated-Algorithmus in verschiedenen Winkelbereichen.	65
4.13. Verteilungen der Zenitwinkel im Datensatz verglichen mit Simulationsdaten.	67
4.14. Verteilungen der Schwerpunkte im Datensatz verglichen mit Simulationsdaten.	68
4.15. Verteilungen der Radien im Datensatz verglichen mit Simulationsdaten.	70
4.16. Abhängigkeitsplots zwischen der simulierten wahren Neutrino-Energie und den genutzten Entfaltungsmerkmalen.	72
4.17. Entfaltung von Ereignissen, welche näherungsweise monoenergetisch sind, zur Bestimmung der Energieauflösung.	73
4.18. L-Kurve für verschiedene Werte für die Anzahl der Freiheitsgrade.	74
4.19. L-Kurve auf Daten sowie Maximale Verschiebungen des Mittelwertes der Pull-Verteilungen für verschiedene Freiheitsgrade.	75
4.20. Entfaltung eines Testdatensatzes im Test-Modus von TRUEE.	76

4.21	Entfaltungen mit verschiedenen Annahmen für die Testdaten.	78
4.22	Entfaltungen mit verschiedenen Annahmen für die Testdaten unter Berücksichtigung einer astrophysikalischen Komponente.	79
4.23	Pull-Verteilung für das siebte Bin der Entfaltung unter der Annahme, dass die DOM-Sensitivität um 10% geringer ist als angenommen. . .	81
4.24	Übersicht über die systematischen Fehler.	82
4.25	Abhängigkeit zwischen berechnetem $\frac{dE}{dx}$ aus dem Truncated-Algorithmus und der wahren Myon-Energie.	83
4.26	Gemessener und mit E^{-2} skaliertes Fluss aus der Analyse der IC 79 Daten im Vergleich zu Rechnungen aus [FTD12].	84
4.27	Entfaltete Ereignisse des IC 79-Datensatzes mit Spline und zugrunde liegenden B-Splines.	85
4.28	Umgewichtete Verteilungen.	86
5.1.	Gemessener und mit E^2 skaliertes Fluss dieser Analyse im Vergleich zu anderen Messungen des Myon-Neutrino-Spektrums mit IceCube.	89
5.2.	Gemessener und mit E^2 skaliertes Fluss dieser Analyse im Vergleich zu anderen Messungen des Myon-Neutrino-Spektrums anderer Experimente.	90
5.3.	Vergleich zwischen Rechnungen aus [FTD12] mit SIBYLL-2.1 als hadronisches Interaktionsmodell und aktuellen IceCube-Messungen.	92
5.4.	Vergleich zwischen Rechnungen aus [FTD12] mit QGSJet-II [Ost06] als hadronisches Interaktionsmodell und aktuellen IceCube-Messungen.	93
5.5.	Vergleich zwischen Rechnungen aus [HKK ⁺ 07] und aktuellen IceCube-Messungen.	94
5.6.	Vergleich zwischen Rechnungen aus [FTD12] mit SIBYLL-2.1 als hadronisches Interaktionsmodell und dem gemittelten Fluss aus den IC 79- und IC 86-Analysen.	96
5.7.	Vergleich zwischen Rechnungen aus [FTD12] mit QGSJet-II [Ost06] als hadronisches Interaktionsmodell und dem gemittelten Fluss aus den IC 79- und IC 86-Analysen.	97
5.8.	Vergleich zwischen Rechnungen aus [HKK ⁺ 07] und dem gemittelten Fluss aus den IC 79- und IC 86-Analysen.	98
5.9.	Vergleich zwischen aktuellen IceCube-Messungen dieser Analyse und aus [Ruh13] und [Sch14] mit prompten Flüssen.	99

5.10. Vergleich zwischen aktuellen IceCube-Messungen aus dieser Analyse und aus [Sch14] sowie den entfalteten Ergebnissen aus [A ⁺ 14b]. . .	101
5.11. Projektionen auf die a) γ - N_{Astro} -Ebene, sowie auf die b) γ - N_{Prompt} -Ebene der Log-Likelihood-Landschaft.	105
5.12. Projektion auf die N_{Prompt} - N_{Astro} -Ebene der Log-Likelihood-Landschaft.	106
5.13. Signifikanz in Abhängigkeit der Nullhypothese für verschiedene maximale prompte Normierungen der Nullhypothese.	106

A. Benutzte Ereignisse zur Separation von Ereignissen

COGZ Z-Koordinate des Schwerpunktes aller getroffenen DOMs, die jeweils mit der Anzahl der Photoelektronen gewichtet werden.

Plogl Wert der reduzierten Log-Likelihood-Funktion. In diesem Fall ist sie durch

$$\text{plogl} = \frac{\ln \mathcal{L}}{N_{Ch-3.5}}$$

definiert. Der Wert der Log-Likelihood-Funktion \mathcal{L} ist der Wert für den MPE-Algorithmus.

MPE_ClosestApp Kürzester Abstand zwischen dem Detektormittelpunkt und der Spurrekonstruktion aus dem MPE-Algorithmus.

MPE_ClosestAppCoG Abstand zwischen dem Schwerpunkt des Ereignisses und dem Detektormittelpunkt.

MPE_LDirC Länge der Projektion aller getroffenen DOMs auf die Spurrekonstruktion des MPE-Algorithmus.

MPE_NDirC Anzahl der DOMs, die mindestens einen direkten Treffer hatten.

MPE_Sall Stochastizität des Ereignisses rekonstruiert mit dem MPE-Algorithmus.

MPE_Zd Rekonstruierter Zenitwinkel aus dem MPE-Algorithmus.

A. Benutzte Ereignisse zur Separation von Ereignissen

NCh Anzahl getroffener DOMs.

NString Anzahl getroffener Strings.

SPE4_rlogl Wert der reduzierten Log-Likelihood-Funktion aus dem SPE-Algorithmus mit vier Iterationen. In diesem Fall wurde dieser Wert über

$$\text{plogl} = \frac{\ln \mathcal{L}}{N_{Ch-5}}.$$

definiert.

SPE4_Status Statusobservable, ob der SPE-Algorithmus in vier Iterationen funktioniert hat.

SPE4_Bay_Zd Zenitwinkel aus dem SPE-Algorithmus mit vier Iterationen und bayesischer Annahme, dass es sich um ein atmosphärisches Myon handelt.

TruncatedMean_Bins_Res, TruncatedMean_Doms_Res Energieauflösung des TruncatedMean-Algorithmus in der DOMs- und der Binsvariante.

Zsigma Standardabweichung aller Z-Koordinaten der getroffenen DOMs.

Ztravel Vertikal zurückgelegte Strecke in Metern.

Radius Radius bezogen auf den X- und Y-Wert des Schwerpunktes der DOMs

IMPE_Zd - LineFit_Zd Absoluter Unterschied zwischen den rekonstruierten Zenitwinkeln des LineFit- und des MPE-Algorithmus.

IMPE_Ad - LineFit_Ad Absoluter Unterschied zwischen den rekonstruierten Azimutwinkeln des LineFit- und des MPE-Algorithmus.

INString * cos(MPE_Zd) Anzahl getroffener Strings gewichtet mit dem Kosinus des Winkels.

A. Benutzte Ereignisse zur Separation von Ereignissen

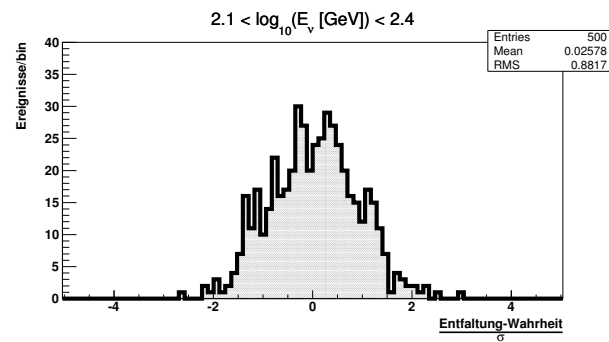
$| \text{TruncatedMean_Bins_Res} - \text{TruncatedMean_Doms_Res} |$ Absoluter Unterschied zwischen den Energieauflösungen der beiden Modi des TruncatedMean-Algorithmus.

$Z_{\text{travel}} * \text{MPE_NDirC}$ Vertikal zurückgelegte Strecke kombiniert mit der Anzahl direkt getroffener DOMs mit mindestens einem Treffer.

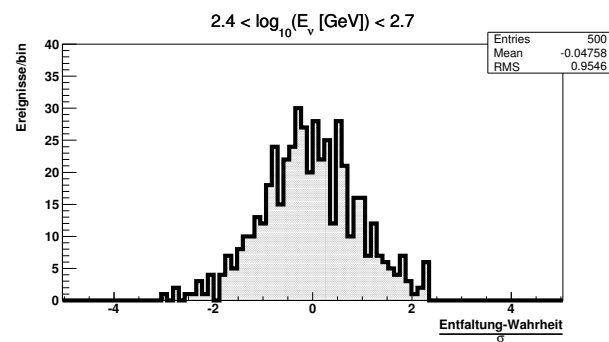
$\text{MPE_NDirB} - \text{MPE_NDirC}$ Unterschied zwischen der Anzahl direkt getroffener DOMs in verschiedenen Zeitfenstern.

$\text{MPE_NDirC} * \text{TruncatedMean_Doms_Res}$ Anzahl direkt getroffener DOMs gewichtet mit der Energieauflösung.

B. Pullmodusverteilungen für die Entfaltung



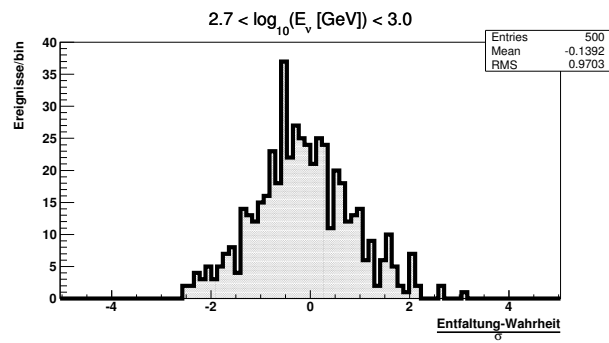
(a)



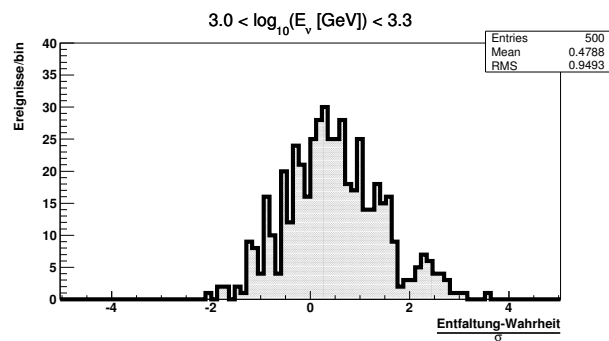
(b)

Abbildung B.1.: Übersicht über die Pull-Verteilungen für die ersten beiden Bins. Die Pull-Verteilungen zeigen keine auffälligen Abweichungen von 0 als Mittelwert.

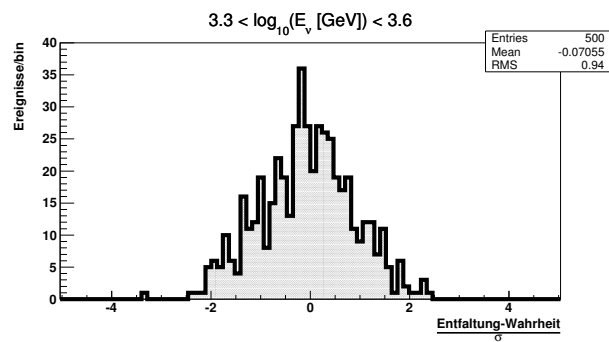
B. Pullmodusverteilungen für die Entfaltung



(a)



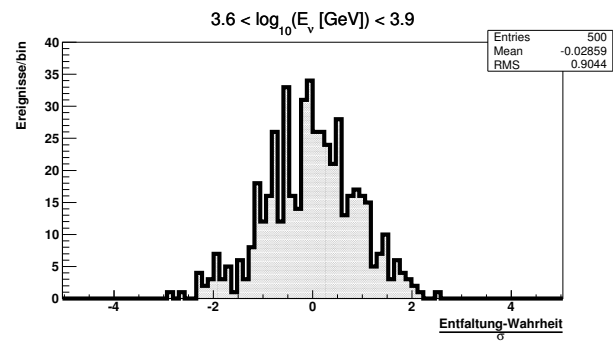
(b)



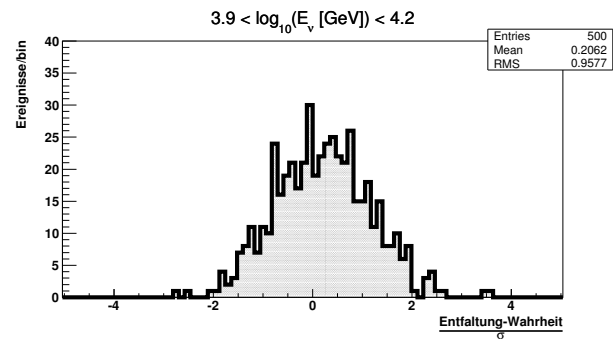
(c)

Abbildung B.2.: Übersicht über die Pull-Verteilungen für die Bins 3-5. Alle Pull-Verteilungen zeigen keine auffälligen Abweichungen von 0 als Mittelwert.

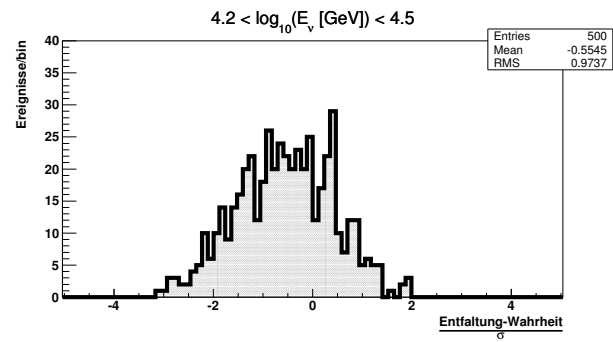
B. Pullmodusverteilungen für die Entfaltung



(a)



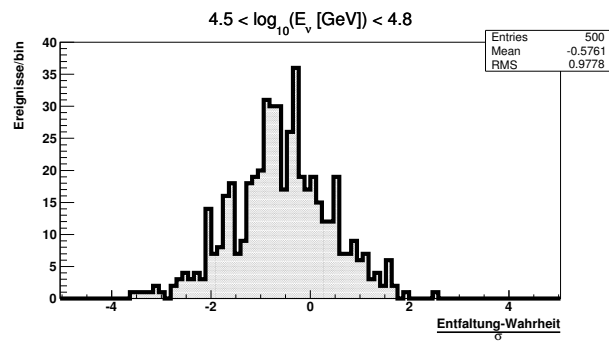
(b)



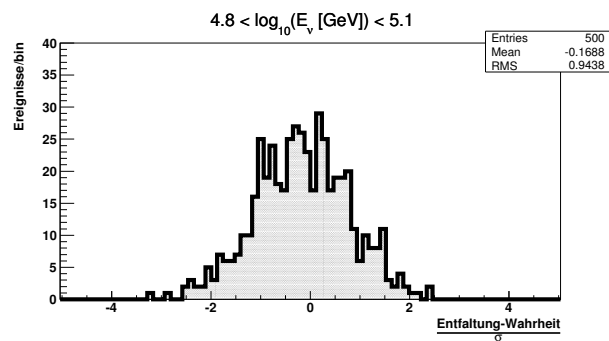
(c)

Abbildung B.3.: Übersicht über die Pull-Verteilungen für die Bins 6-8. Alle Pull-Verteilungen zeigen keine auffälligen Abweichungen von 0 als Mittelwert.

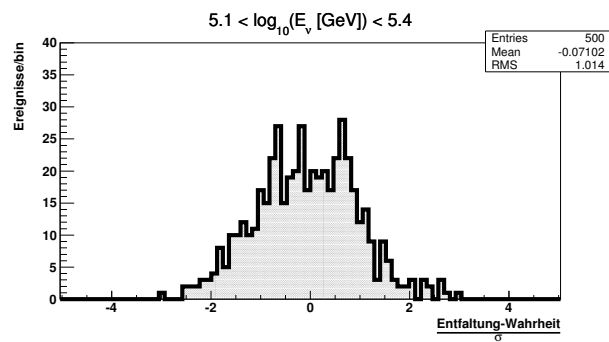
B. Pullmodusverteilungen für die Entfaltung



(a)



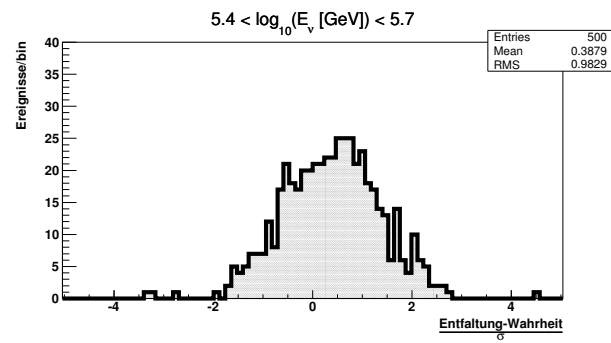
(b)



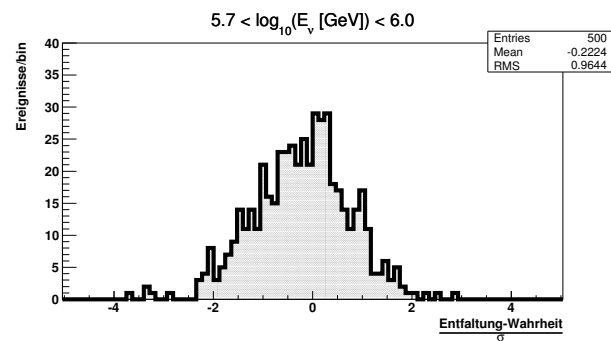
(c)

Abbildung B.4.: Übersicht über die Pull-Verteilungen für die Bins 9-11. Alle Pull-Verteilungen zeigen keine auffälligen Abweichungen von 0 als Mittelwert.

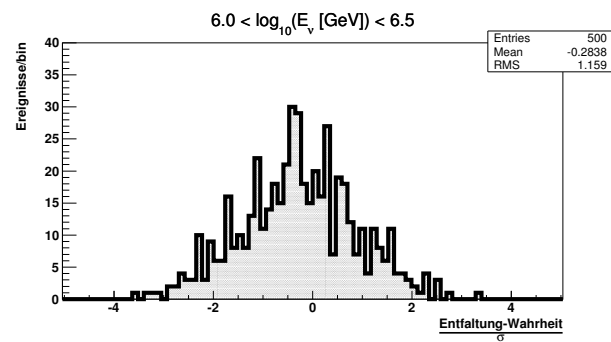
B. Pullmodusverteilungen für die Entfaltung



(a)



(b)



(c)

Abbildung B.5.: Übersicht über die Pull-Verteilungen für die Bins 12-14. Alle Pull-Verteilungen zeigen keine auffälligen Abweichungen von 0 als Mittelwert.

C. Ergebnisse der Flussmessung

Energiebereich $\log_{10}(E[\text{GeV}])$	Zentrale Energie $\log_{10}(E[\text{GeV}])$	Fluss $E^2 \frac{d\Phi}{dE} [\frac{\text{GeV}}{\text{sr s m}^2}]$	Unsicherheit [%]
2.10 - 2.40	2.26	1.84e-04	-0.03, 0.22
2.40 - 2.70	2.55	1.22e-04	-0.07, 0.05
2.70 - 3.00	2.84	5.07e-05	-0.17, 0.16
3.00 - 3.30	3.17	2.80e-05	-0.12, 0.06
3.30 - 3.60	3.43	1.37e-05	-0.27, 0.58
3.60 - 3.90	3.76	5.69e-06	-0.40, 0.78
3.90 - 4.20	4.05	1.68e-06	-0.15, 0.39
4.20 - 4.50	4.36	6.69e-07	-0.22, 0.17
4.50 - 4.80	4.66	3.20e-07	-0.23, 0.15
4.80 - 5.10	4.95	1.51e-07	-0.22, 0.25
5.10 - 5.40	5.25	6.08e-08	-0.20, 0.53
5.40 - 5.70	5.55	3.71e-08	-0.31, 0.76
5.70 - 6.00	5.85	2.48e-08	-0.36, 0.77
6.00 - 6.50	6.26	1.44e-08	-0.53, 1.36

Tabelle C.1.: Entfaltungsergebnis nach der Entfaltung mit TRUEE. Der Wert der zentralen Energie ist mit der Methode, welche in Kapitel 3.4.6 erläutert ist, bestimmt. Die angegebenen Fehler schliessen systematische Fehler ein.

D. Ergebnisse der gemittelten Flussmessung

Energiebereich $\log_{10}(E[\text{GeV}])$	Fluss $E^2 \frac{d\Phi}{dE} [\frac{\text{GeV}}{\text{sr s m}^2}]$	Unsicherheit [%]
2.1 - 2.4	2.18e-04	- 0.07, 0.13
2.4 - 2.7	1.15e-04	- 0.10, 0.10
2.7 - 3.0	5.63e-05	- 0.14, 0.20
3.0 - 3.3	3.03e-05	- 0.12, 0.07
3.3 - 3.6	1.43e-05	- 0.27, 0.35
3.6 - 3.9	6.61e-06	- 0.29, 0.41
3.9 - 4.2	2.01e-06	- 0.09, 0.19
4.2 - 4.5	6.69e-07	- 0.12, 0.14
4.5 - 4.8	2.98e-07	- 0.21, 0.16
4.8 - 5.1	1.11e-07	- 0.35, 0.24
5.1 - 5.4	5.10e-08	- 0.38, 0.35
5.4 - 5.7	2.62e-08	- 0.48, 0.48
5.7 - 6.0	1.51e-08	- 0.48, 0.60
6.0 - 6.5	1.06e-08	- 0.51, 1.02

Tabelle D.1.: Ergebnisse der gemittelten Flussmessung für die Messungen des IC 79 und IC 86 Datensatzes aus Kapitel 5.3

Literaturverzeichnis

- [A⁺03] T. Antoni et al. The cosmic-ray experiment KASCADE. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 513(3):490–510, 2003.
- [A⁺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⁺06a] A. Achterberg et al. On the selection of AGN neutrino source candidates for a source stacking analysis with neutrino telescopes. *Astroparticle physics*, 26(4):282–300, 2006.
- [A⁺06b] 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⁺10a] 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⁺10b] R. Abbasi et al. The energy spectrum of atmospheric neutrinos between 2 and 200 TeV with the AMANDA-II detector. *Astroparticle physics*, 34(1):48–58, 2010.
- [A⁺10c] J. Abraham et al. Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory. *Physics Letters B*, 685(4):239–246, 2010.

- [A⁺10d] C. Andreopoulos et al. The GENIE neutrino monte carlo generator. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 614(1):87–104, 2010.
- [A⁺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⁺11b] O. Adriani et al. PAMELA Measurements of Cosmic-Ray Proton and Helium Spectra. *Science*, 332:69–, 2011.
- [A⁺11c] M. Ageron et al. ANTARES: the first undersea neutrino telescope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 656(1):11–38, 2011.
- [A⁺12a] R. Abbasi et al. An absence of neutrinos associated with cosmic-ray acceleration in γ -ray bursts. *Nature*, 484:351–353, 2012.
- [A⁺12b] R. Abbasi et al. The design and performance of IceCube DeepCore. *Astroparticle physics*, 35(10):615–624, 2012.
- [A⁺12c] P. Allison et al. Design and initial performance of the Askaryan Radio Array prototype EeV neutrino detector at the South Pole. *Astroparticle physics*, 35(7):457–477, 2012.
- [A⁺13a] M.G. Aartsen et al. Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector. *Science*, 342(6161):1242856, 2013.
- [A⁺13b] M.G. Aartsen et al. Measurement of the Atmospheric ν_e Flux in IceCube. *Physical Review Letters*, 110(15):151105, 2013.
- [A⁺13c] M.G. Aartsen et al. The IceCube Neutrino Observatory Part VI: Ice Properties, Reconstruction and Future Developments. *arXiv preprint arXiv:1309.7010*, 2013.
- [A⁺13d] R. Abbasi et al. An improved method for measuring muon energy using the truncated mean of dE/dx . *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 703:190–198, 2013.

- [A⁺14a] M.G. Aartsen et al. Energy reconstruction methods in the IceCube neutrino telescope. *Journal of Instrumentation*, 9(03):P03009, 2014.
- [A⁺14b] M.G. Aartsen et al. Observation of High-Energy Astrophysical Neutrinos in Three Years of IceCube Data. 2014.
- [A⁺14c] M.G. Aartsen et al. Observation of the cosmic-ray shadow of the Moon with IceCube. *Physical Review D*, 89(10):102004, 2014.
- [A⁺14d] 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.
- [Ady11] T. Adye. Unfolding algorithms and tests using RooUnfold. *arXiv preprint arXiv:1105.1160*, 2011.
- [AEF⁺] E. Ahn, R. Engel, A. Fedynitch, T. Gaisser, F. Riehn, and T. Stanev. Atmospheric neutrinos at high energy. *atmosphere*, 3:8.
- [AEG⁺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. *The Astrophysical Journal Supplement Series*, 141(1):195, 2002.
- [AM⁺12] S. Adrián-Martínez et al. Measurement of atmospheric neutrino oscillations with the ANTARES neutrino telescope. *Physics Letters B*, 714(2):224–230, 2012.
- [AM⁺13] S. Adrián-Martínez et al. Detection potential of the KM3NeT detector for high-energy neutrinos from the fermi bubbles. *Astroparticle Physics*, 42:7–14, 2013.
- [B⁺12] J. Beringer et al. Review of particle physics. *Physical Review D*, 86(1), 2012.
- [Bat94] Roberto Battiti. Using mutual information for selecting features in supervised neural net learning. *Neural Networks, IEEE Transactions on*,

- 5(4):537–550, 1994.
- [Bec07] Julia K. Becker. *Neutrinos on the Rocks: On the Phenomenology of Potential Astrophysical Neutrino Sources*. PhD thesis, TU Dortmund, 2007.
- [BL98] Volker Blobel and Erich Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Springer DE, 1998.
- [Blo96] Volker Blobel. The RUN Manual-Regularized Unfolding for High-Energy Physics experiments. *Technical Note TN*, 361, 1996.
- [Bot05] O. Botner. The IceCube neutrino observatory. *Nuclear Physics B-Proceedings Supplements*, 143:367–370, 2005.
- [BP96] A.D. Brink and N.E. Pendock. Minimum cross-entropy threshold selection. *Pattern Recognition*, 29(1):179–188, 1996.
- [Bre01] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [C⁺07] D.F. Cowen et al. Tau neutrinos in IceCube. In *Journal of Physics: Conference Series*, volume 60, page 227. IOP Publishing, 2007.
- [Čer37] PA Čerenkov. Visible radiation produced by electrons moving in a medium with velocities exceeding that of light. *Physical Review*, 52(4):378, 1937.
- [Chi11] D. Chirkin. Study of South Pole ice transparency with IceCube flashers. In *International Cosmic Ray Conference*, volume 4, page 161, 2011.
- [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.
- [CRH⁺56] C. L. Cowan, Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the Free Neutrino: A Confirmation. *Science*, 124:103–104, 1956.
- [D⁺95] K Daum et al. Determination of the atmospheric neutrino spectra with the Frejus detector. *Zeitschrift für Physik C Particles and Fields*, 66(3):417–428, 1995.
- [dB94] Carl de Boor. *A Practical Guide to Splines (Applied Mathematical Sciences)*. Springer, 8 1994.

- [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.
- [DV14] JC Díaz-Vélez. The IceProd (IceCube Production) Framework. In *Journal of Physics: Conference Series*, volume 513, page 032026. IOP Publishing, 2014.
- [Efr79] Bradley Efron. Bootstrap methods: another look at the jackknife. *The annals of Statistics*, pages 1–26, 1979.
- [Enb14] R. Enberg. Neutrinos from charm production in the atmosphere. *arXiv preprint arXiv:1402.0880*, 2014.
- [ERS08] R. Enberg, M. H. Reno, and I. Sarcevic. Prompt neutrino fluxes from atmospheric charm. *Physical Review D*, 78(4):043005, 2008.
- [F⁺98] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Physical Review Letters*, 81(8):1562, 1998.
- [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.
- [G⁺03] Thomas K Gaisser et al. IceTop: The surface component of IceCube. In *International Cosmic Ray Conference*, volume 2, page 1117, 2003.
- [Gag06] ND Gagunashvili. Comparison of weighted and unweighted histograms. *arXiv preprint physics/0605123*, 2006.
- [Gai12] T. Gaisser. Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio. *Astroparticle Physics*, 35(12):801–806, 2012.
- [GH02] T.K. Gaisser and M. Honda. Flux of atmospheric neutrinos. *Annual Review of Nuclear and Particle Science*, 52:153–199, 2002.
- [Gin12] C. Gini. *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*. Number Teil 1. Tipogr. di P. Cuppini, 1912.
- [GJKvS14] T. Gaisser, K. Jero, A. Karle, and J. van Santen. A generalized self-veto probability for atmospheric neutrinos. *arXiv preprint arXiv:1405.0525*, 2014.

- [GK05] A. Gazizov and M. Kowalski. ANIS: High energy neutrino generator for neutrino telescopes. *Computer Physics Communications*, 172:203–213, 2005.
- [GMS01] D. Groom, N. Mokhov, and S. Striganov. Muon stopping power and range tables 10 MeV–100 TeV. *Atomic Data and Nuclear Data Tables*, 78(2):183–356, 2001.
- [Gre60] K. Greisen. Cosmic ray showers. *Annual Review of Nuclear Science*, 10(1):63–108, 1960.
- [Gre66] K. Greisen. End to the Cosmic-Ray Spectrum? *Physical Review Letters*, 16:748–750, April 1966.
- [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.
- [Gus10] J. R. Gustafson. Wolfgang Pauli 1900 to 1930: His Early Physics in Jungian Perspective. *ArXiv e-prints*, March 2010.
- [H⁺87] K Hirata et al. Observation of a neutrino burst from the supernova SN1987A. *Physical Review Letters*, 58(14):1490, 1987.
- [Hal09] F. Halzen. IceCube science. In *Journal of Physics: Conference Series*, volume 171, page 012014. IOP Publishing, 2009.
- [HH79] E. Hewitt and R. Hewitt. The Gibbs-Wilbraham phenomenon: an episode in Fourier analysis. *Archive for history of Exact Sciences*, 21(2):129–160, 1979.
- [HKC⁺98] D. Heck, J. Knapp, JN Capdevielle, G Schatz, T Thouw, et al. CORSIKA: A Monte Carlo code to simulate extensive air showers. 6019, 1998.
- [HKK⁺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. *Physical Review D*, 75(4):043006, 2007.
- [Hör03] J. Hörandel. On the knee in the energy spectrum of cosmic rays. *Astroparticle Physics*, 19(2):193–220, 2003.
- [Hör07] J. Hörandel. Cosmic rays from the knee to the second knee: 10^4 to 10^{18} eV. *Modern Physics Letters A*, 22(21):1533–1551, 2007.

- [Jac12] P. Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.
- [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.
- [KFS⁺13] J.-H. Koehne, K. Frantzen, M. Schmitz, T. Fuchs, W. Rhode, D. Chirkin, and J. Becker Tjus. PROPOSAL: A tool for propagation of charged leptons. *Computer Physics Communications*, 184:2070–2090, 2013.
- [KS12] U Katz and C Spiering. High-energy neutrino astrophysics: Status and perspectives. *Progress in Particle and Nuclear Physics*, 67(3):651–704, 2012.
- [Lar13] Michael James Larson. Simulation and identification of non-poissonian noise triggers in the IceCube neutrino detector. Master’s thesis, University of Alabama, 2013.
- [LM68] P. Lachenbruch and M. Mickey. Estimation of error rates in discriminant analysis. *Technometrics*, 10(1):1–11, 1968.
- [LM00] J. Learned and K. Mannheim. High-energy neutrino astrophysics. *Annual Review of Nuclear and Particle Science*, 50(1):679–749, 2000.
- [LW95] G.D. Lafferty and T.R. Wyatt. Where to stick your data points: The treatment of measurements within wide bins. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 355(2):541–547, 1995.
- [M⁺05] P Miocinovic et al. Tuning into UHE Neutrinos in Antarctica-The ANITA Experiment. *arXiv preprint astro-ph/0503304*, 2005.
- [M⁺09] E. Middell et al. Improved Reconstruction of Cascade-like Events in IceCube. 2009.
- [MDK⁺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] Natalie Milke. *Unfolding of the atmospheric neutrino flux spectrum with the new program TRUEE and IceCube*. PhD thesis, TU Dortmund, 2012.
- [MJ51] F. Massey Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.
- [Ost06] S. Ostapchenko. QGSJET-II: towards reliable description of very high energy hadronic interactions. *Nuclear Physics B-Proceedings Supplements*, 151(1):143–146, 2006.
- [Ruh13] Tim 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, TU Dortmund, 2013.
- [Sch13] Anne Schukraft. *Search for a diffuse flux of extragalactic neutrinos with the IceCube Neutrino Observatory*. PhD thesis, RWTH Aachen University, 2013.
- [Sch14] Florian Scheriau. *Data-Mining in der Astroteilchenphysik: Entfaltung des Energiespektrums der von IceCube gemessenen Myoneneutrinos in der 86-String Konfiguration*. PhD thesis, TU Dortmund, 2014.
- [SM10] B. Schowe and K. Morik. Fast-ensembles of minimum redundancy feature selection. In *Proceedings of LWA2010 - Workshop-Woche: Lernen, Wissen & Adaptivitaet*, Kassel, Germany, 2010.
- [SPM02] J. L. Starck, E. Pantin, and F. Murtagh. Deconvolution in Astronomy: A Review. *Publications of the Astronomical Society of the Pacific*, 114:1051–1069, 2002.
- [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.
- [Tik43] A.N Tikhonov. On the stability of inverse problems. *Dokl. Akad. Nauk SSSR*, 39:95 – 198, 1943.
- [W⁺09] C. Wiebusch et al. Physics capabilities of the IceCube DeepCore detector. *arXiv preprint arXiv:0907.2263*, 2009.
- [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.

- [YD98] S. Yoshida and H. Dai. The extremely high energy cosmic rays. *Journal of Physics G: Nuclear and Particle Physics*, 24(5):905, 1998.
- [ZK66] G. Zatsepin and V. Kuzmin. Upper limit of the spectrum of cosmic rays. *JETP Lett.*, 4:114–116, 1966.
- [ZS06] V.I. Zatsepin and Natalia V. Sokolskaya. Three component model of cosmic ray spectra from 100-gev up to 100-pev. *Astronomy and Astrophysics*, 458:1–5, 2006.