

# Suche nach Clustern im Pierre Auger Experiment

Bachelor-Arbeit

zur Erlangung des akademischen Grades

Bachelor of Science

(B.Sc.)

dem Fachbereich Physik der

Universität Siegen

vorgelegt von

Oliver Rosenthal

April 2008



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Kosmische Strahlung . . . . .	1
1.1.1	Energiespektrum . . . . .	1
1.1.2	GZK-Schwelle . . . . .	2
1.2	Ausgedehnte Luftschauer . . . . .	3
1.3	Quellen und Anisotropie der kosmischen Strahlung . . . . .	4
1.4	Thema und Ziel der Arbeit . . . . .	5
<b>2</b>	<b>Das Pierre-Auger-Experiment</b>	<b>6</b>
2.1	Detektorarten . . . . .	7
2.1.1	Boden-Detektoren . . . . .	7
2.1.2	Fluoreszenz-Detektoren . . . . .	8
<b>3</b>	<b>Erstellung von Himmelskarten</b>	<b>10</b>
3.1	HEALPix . . . . .	10
3.1.1	HEALPix Projektion . . . . .	11
3.1.2	Auflösung . . . . .	11
3.2	Koordinatensysteme . . . . .	13
3.2.1	Horizontale Koordinaten . . . . .	13
3.2.2	Äquatorkoordinaten . . . . .	13
<b>4</b>	<b>Significance Map</b>	<b>14</b>
4.1	Coverage Map . . . . .	14
4.2	Event Map . . . . .	15
4.3	Significance Map . . . . .	17
<b>5</b>	<b>Clustersuche</b>	<b>19</b>
5.1	Methode 1 . . . . .	19
5.2	Methode 2 . . . . .	21
5.2.1	Anwendung und Effizienztest . . . . .	24
<b>6</b>	<b>Methoden der Clustersuche im Pierre-Auger-Experiment</b>	<b>29</b>
6.1	Cone-Jet-Algorithmus . . . . .	29
6.2	Maximum-Likelihood-Ratio-Test . . . . .	29

<i>INHALTSVERZEICHNIS</i>	iii
6.3 Die 2-Punkt Korrelationfunktion . . . . .	30
<b>7 Zusammenfassung und Ausblick</b>	<b>32</b>
<b>Literaturverzeichnis</b>	<b>33</b>

# Kapitel 1

## Einleitung

### 1.1 Kosmische Strahlung

Als kosmische Strahlung werden geladene Teilchen bezeichnet, die sich durch das Weltall bewegen und auf die Erde treffen. Diese Strahlung besteht überwiegend aus ca. 98% Atomkernen und zu 2% aus Elektronen, wobei sich die Kerne zu ca. 87% aus Protonen, zu 12% aus Alphateilchen und nur zu etwa 1% aus schweren Elementen zusammensetzen [KK97].

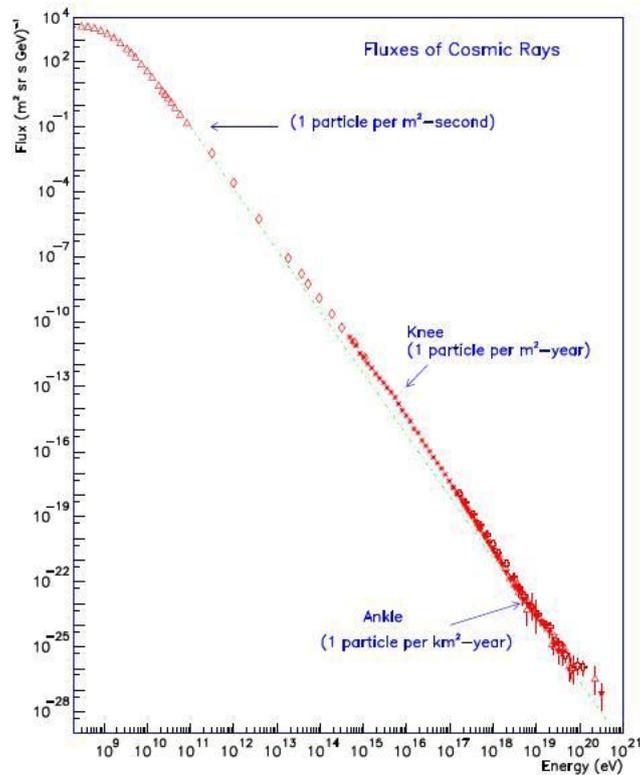
#### 1.1.1 Energiespektrum

Das Energiespektrum der kosmische Strahlung erstreckt sich über einen sehr großen Energiebereich von  $10^9\text{eV}$  bis hin zu ultrahohen Energien von über  $3 \cdot 10^{20}\text{eV}$ . Die Energiespektrum dieser Teilchen folgt dabei folgendem Potenzgesetz:

$$\frac{dN}{dE} \propto E^{-\gamma} \quad (1.1)$$

Der spektrale Index  $\gamma$  hat bis etwa  $3 \cdot 10^{15}\text{eV}$  einen Wert von 2.7 und oberhalb dieser Energieschwelle einen Wert von 3.0. Bei einer Energie von ungefähr  $3 \cdot 10^{18}\text{eV}$  ändert sich dann  $\gamma$  erneut zu einem Wert von 2.7.

In Abbildung 1.1 ist das gemessene Energiespektrum gezeigt. Es sind zwei Knicke erkennbar: Bei etwa  $3 \cdot 10^{15}\text{eV}$ , an der Stelle der ersten Änderung im Index des Potenzgesetzes, das sogenannte Knie, sowie bei etwa  $3 \cdot 10^{18}\text{eV}$ , genannt Knöchel.



**Abbildung 1.1:** Das beobachtete Energiespektrum der kosmischen Strahlung, wobei der Teilchenfluss über die Teilchenenergie aufgetragen ist. Das Knie ist bei einer Energie von  $3 \cdot 10^{15}$  eV und der Knöchel bei etwa  $3 \cdot 10^{18}$  eV zu erkennen [Pie97]

Gründe für das Abknicken des Spektrums sind noch nicht vollständig verstanden und Gegenstand aktueller Forschung.

Der Teilchenfluss bei der Energie des Knies beträgt rund ein Teilchen pro m<sup>2</sup> und Jahr, im Bereich des Knöchels nur noch ungefähr ein Teilchen pro km<sup>2</sup> und Jahr. Damit ist eine direkte Messung der Energie und Teilchenart der kosmischen Strahlung durch Ballon- oder Satellitenexperimente nicht mehr möglich.

Kosmische Teilchen erzeugen allerdings in der Erdatmosphäre einen ausgedehnten Luftschauer mit sehr vielen Sekundärteilchen, die mit Hilfe von großen Detektorflächen auf dem Erdboden nachweisbar sind. Aus den dabei gewonnenen Messdaten können Rückschlüsse auf die primären Teilchen bezüglich ihrer Masse, Energie und Richtung gezogen werden.

### 1.1.2 GZK-Schwelle

Nach einer Theorie von Kenneth Greisen, Georgy Zatsepin und Vadim Kuzmin wird bei einer Energie von  $5 \cdot 10^{19}$  eV das Ende der kosmischen Strahlung vorhergesagt, da alle Teilchen oberhalb der sogenannten GZK-Schwellenenergie mit den Photonen der 2.7K Hintergrundstrahlung wechselwir-

ken [Gre66,Zat66]. Ab dieser Schwelle steht genügend Energie im Schwerpunktsystem zur Verfügung, um die  $\Delta$ -Resonanz anzuregen:

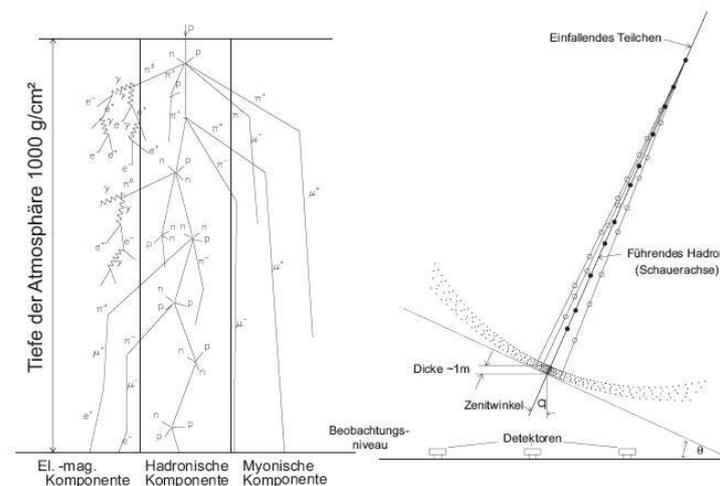
$$p + \gamma_{2.7K} \rightarrow \Delta(1232) \rightarrow p + \pi^0 \tag{1.2}$$

$$p + \gamma_{2.7K} \rightarrow \Delta(1232) \rightarrow n + \pi^+ \tag{1.3}$$

Die Protonen verlieren dadurch soviel Energie, dass sie innerhalb einer zurückgelegten Wegstrecke von 100Mpc die GZK-Schwelle unterschreiten. Detektiert man nun auf der Erde in größerer Zahl Teilchen mit Energien oberhalb der Schwelle, so kann die Quelle demnach nicht weiter als 100Mpc entfernt sein [Blü00].

## 1.2 Ausgedehnte Luftschauer

Kosmische Teilchen wechselwirken in der Atmosphäre mit den Kernen von Luftmolekülen, wodurch Sekundärteilchen entstehen. Diese erzeugen wiederum neue Teilchen, sodass sich kaskadenartige Schauer entwickeln. Ein Proton mit  $10^{15}$ eV erzeugt z.B.  $10^6$  Teilchen [Gru00], zusammen gesetzt aus 80% Photonen, 17% Elektronen, 1.7% Myonen und 0.3% Hadronen [Ast]. Die Sekundärteilchen bilden eine ausgedehnte Scheibe, die sich mit nahezu Lichtgeschwindigkeit entlang der Einfallsrichtung des Primarteilchens durch die Atmosphäre bewegt. Diese Scheibe ist im Zentrum einige Meter dick ist und kann einen Durchmesser von mehreren Kilometern haben. Zur Untersuchung genügt aufgrund der Vielzahl an Teilchen die einzelnen Komponenten stichprobenartig zu messen. In Abbildung 1.2 ist ein solcher Luftschauer schematisch dargestellt.



**Abbildung 1.2:** Schematische Darstellung eines ausgedehnten Luftschauers; Das linke Bild zeigt eine Schauerkaskade, rechts ist ein Luftschauer schematisch im Profil dargestellt [Alk75]

Luftschauer lassen sich in drei verschiedene Komponenten unterteilen: Die hadronische, myonische und elektromagnetische Gruppe.

Die **hadronische Komponente** besteht aus Kernen, Nukleonen sowie Pionen und Kaonen. Bei einer hadronischen Kollision wird der Impuls hauptsächlich in Vorwärtsrichtung übertragen, sodass sich fast alle Hadronen innerhalb von 10 - 20m um die Schauerachse befinden. Die meisten produzierten Hadronen sind Pionen. Die neutralen Pionen zerfallen sehr schnell in zwei Photonen (Lebensdauer  $\tau_{\pi^0} = 8.4 \cdot 10^{-17}\text{s}$ ) und erzeugen so unmittelbar eine elektromagnetische Kaskade. Geladene Pionen und Kaonen können aufgrund ihrer höheren Lebensdauer ( $\tau_{K^\pm} = 2.6 \cdot 10^{-8}\text{s}$ ;  $\tau_{\pi^\pm} = 1.2 \cdot 10^{-8}\text{s}$ ) weiterhin mit der Atmosphäre wechselwirken und zerfallen leptonisch in Myonen und Neutrinos.

So bildet sich die **myonische Komponente** des Schauers aus. Da die energiereichen Myonen durch die relativistische Zeitdilatation eine drastische Verlängerung ihrer Lebensdauer erfahren und durch Ionisation oder andere Prozesse kaum Energie verlieren, können sie mit großer Wahrscheinlichkeit den Erdboden erreichen. Im niederenergetischen Bereich kann es jedoch auch zu Zerfällen von Myonen in Elektronen und Neutrinos kommen.

Die **elektromagnetische Komponente** entsteht durch den Zerfall neutrale Mesonen, die durch hadronische Wechselwirkungen entstanden sind, vor allem neutrale Pionen. Die entstandenen Photonen erzeugen dann durch Paarbildung ein Elektron-Positron-Paar. Da diese durch Bremsstrahlung wieder Photonen emittieren, welche wiederum Elektronen-Positronen-Paare erzeugen, entsteht eine Kaskade von Elektronen, Positronen und Photonen.

### 1.3 Quellen und Anisotropie der kosmischen Strahlung

Aufgrund des großen Energiebereiches der kosmischen Teilchen geht man davon aus, dass es verschiedene Quellenarten im Kosmos gibt. Man nimmt an, dass Teilchen mit Energien bis zu  $10^{15}\text{eV}$  aus Supernovae und Pulsaren in unserer Galaxie stammen. Der Ursprung ultrahochenergetischer Strahlung wird außerhalb unserer Galaxie angenommen. Mögliche Quellen sind vor allem aktive galaktische Kerne, die zu den energiereichsten bekannten Objekten im Universum gehören.

Diese Teilchen werden nach der Fermi-Theorie [Fer49] durch Magnetfelder in Schockfronten von explodierenden Supernovae und sich ausdehnende magnetische Plasmen beschleunigt. Eine weitere Möglichkeit sind Pulsare, d.h. rotierende Neutronensterne. Durch die große Rotationsenergie kommt es in den magnetischen Polbereichen zu Potentialabfällen, die eine sehr hohe Beschleunigung der Teilchen ermöglicht. Weiterhin werden bei Doppelsternsysteme starke elektromagnetische Felder erzeugt, in denen die geladene Strahlung ebenfalls beschleunigt wird.

Auf ihrem Weg durch das All wird die kosmischen Strahlung aufgrund ihrer elektrischen Ladung von interstellaren magnetischen Feldern beeinflusst. Abhängig von der Stärke und der Struktur des Magnetfeldes, sowie der Energie der Primärteilchen lässt sich der Ursprung der Strahlung meist nicht zurückverfolgen. Aufgrund dieser Ablenkung erwartet man eine gleichmäßige Verteilung der Ankunftsrichtungen auf der Erde, was durch Messungen bestätigt wurde. Für die große Statistik von Teilchen bis  $10^{14}\text{eV}$  wurde eine Abweichung von der Isotropie von nur 0.5% beobachtet. Da jedoch die Ablenkung mit höherer Energie abnimmt, entsprechen die Ankunftsrichtungen höchstenergetischer kosmischer Teilchen oberhalb von  $6 \cdot 10^{19}\text{eV}$  fast genau der Richtungen ihrer Quellen. Diese Teilchen und die dadurch verursachte Anisotropie ermöglicht eine Suche nach ihren Quellen.

## 1.4 Thema und Ziel der Arbeit

Das Thema dieser Arbeit ist die Entwicklung von Algorithmen zur Suche von Clustern in den Daten des Pierre-Auger-Projekts. Die Clustersuche ist eine Methode bei der Anisotropieanalyse zur Entdeckung von Punktquellen aber auch Objekten, die die kosmische Strahlung auf dem Weg Richtung Erde abfangen. Die Arbeit ist folgendermaßen aufgebaut:

- Kapitel 2 enthält eine kurze Beschreibung des Pierre-Auger-Projekts
- In Kapitel 3 wird das Programm HEALPix vorgestellt, mit dessen Hilfe Himmelskarten erstellt werden
- Grundlegend für die Suche nach Clustern sind die Signifikanzen; Wie diese berechnet werden und wie mit diesen Daten eine Himmelskarte konstruiert wird, die sog. Significance Map, wird in Kapitel 4 erklärt
- Zwei Methoden der Clustersuche wurden entwickelt; Kapitel 5 gibt eine ausführliche Beschreibung dieser Algorithmen sowie erste Ergebnisse wieder
- Kapitel 6 gibt einen kurzen Überblick über Methoden der Clustersuche, die im Rahmen des Pierre-Auger-Projektes veröffentlicht wurden
- Das letzte Kapitel fasst noch einmal den Inhalt dieser Arbeit zusammen und gibt einen kurzen Ausblick über mögliche weiterführende Studien

## Kapitel 2

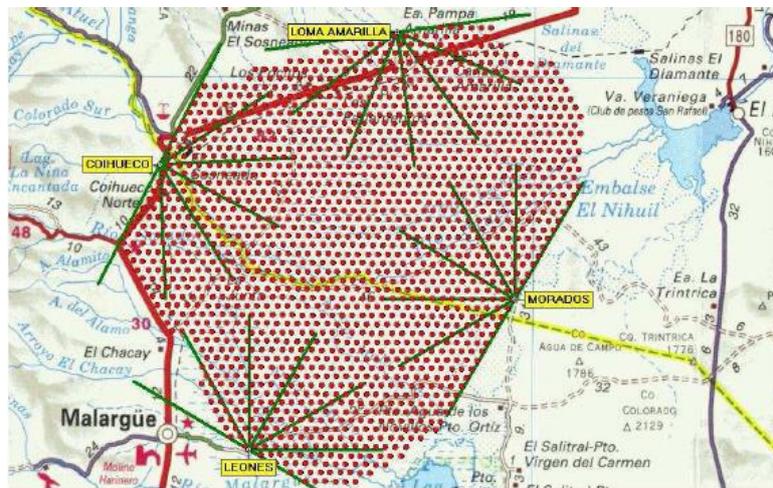
# Das Pierre-Auger-Experiment

Das internationale Pierre-Auger-Projekt hat sich zur Aufgabe gemacht, hochenergetische kosmische Strahlung bezüglich ihrer Herkunft und Teilchenzusammensetzung im Energiebereich zwischen  $10^{17}$ eV und  $10^{21}$ eV zu untersuchen.

Die Frage, ob es bei einer Energie von  $5 \cdot 10^{19}$ eV zum GZK-Cutoff kommt, ist ebenfalls ein wichtiges Thema des Projektes. Experimente wie AGASA oder Fly's Eye konnten bereits Teilchen mit Energien über  $10^{20}$ eV nachweisen, jedoch aufgrund der nicht genügend vorhandenen Statistik nicht eindeutig klären. Sollte das Pierre-Auger-Experiment solche Teilchen nachweisen, wäre ein weiteres Ziel, die Quellen dieser Teilchen zu finden. Da die höchstenergetischen Teilchen auf ihrem Weg nur wenig durch Magnetfelder abgelenkt werden, erzeugen ihre Quellen eine Anisotropie in der Verteilung der Ankunftsrichtung, die mit dem Pierre-Auger-Projekt ebenfalls untersucht wird.

In der Hochebene Pampa Amarilla in der Provinz Mendoza/ Argentinien befindet sich das südliche Pierre-Auger-Observatorium. Bei diesem Projekt werden zwei verschiedene Nachweismethoden kombiniert: 1600 Bodendetektoren, die hexagonal im Abstand von 1.5 km auf einer Gesamtfläche von  $3000\text{km}^2$  verteilt sind, umgeben von vier Stationen mit jeweils 6 Fluoreszenzteleskopen. Das Observatorium ist noch nicht endgültig fertiggestellt, im Mai 2007 waren 75% der Bodendetektoren betriebsbereit [Piea], wobei bereits seit Anfang 2004 Daten mit der immer größer werdenden Anlage aufgenommen werden. Abbildung 2.1 zeigt eine Übersicht des Observatoriums.

Da es das Ziel ist, den gesamten Himmel nach Quellen der höchstenergetischen kosmischen Strahlung abzusuchen, ist ein zweites Observatorium auf der nördlichen Erdhalbkugel in Colorado/USA in Planung.



**Abbildung 2.1:** Das Pierre-Auger-Observatorium in der Pampa Amarilla(Argentinien); das Bild zeigt das ganze Gebiet mit den Bodendetektoren (rote Punkte), umgeben von den vier Teleskopstationen (grün) [Pieb]

## 2.1 Detektorarten

### 2.1.1 Boden-Detektoren

Jeder Bodendetektor ist ein zylinderförmiger Wasser-Cherenkov-Tank mit einem Durchmesser von 3.6 m und einer Höhe von 1.2m, gefüllt mit insgesamt 12000l ultrareinem Wasser. Über eigene Solarpanels werden die Detektoren mit Energie versorgt. Die Erfassung der Ankunftszeit von Teilchen in den Bodendetektoren erfolgt mit Hilfe von GPS (Global Positioning System).

Treten nun hochenergetische, geladene Teilchen in den Tank ein und besitzen im Wasser eine höhere Geschwindigkeit als die Lichtgeschwindigkeit in Wasser, so entsteht sogenannte Cherenkov-Strahlung. Dieses Licht wird von den mit einem speziellen UV-reflektierenden Tyvec-Material ausgekleideten Innenwänden des Tanks reflektiert und kann so von den drei unabhängigen Photomultipliern im Tank registriert werden. Das Wasser muss ultrarein sein, um die Absorption der Photonen im Wasser zu minimieren und damit sich keine Bakterien ausbreiten, um eine Betriebszeit von 20 Jahren bei gleichbleibender Qualität gewährleisten zu können. Abbildung 2.2 zeigt den schematischen Aufbau eines solchen Tanks.

Mit diesen Detektoren soll die Lateralverteilung der myonischen und elektromagnetischen Komponente eines kosmischen Schauers untersucht werden, was Rückschlüsse auf die Energie des Primärteilchens ermöglicht. Zudem lässt sich anhand des Verhältnisses Myon- zu Elektronenzahl auf die Art und Masse des primären Teilchens schließen. Die Richtung eines einfallenden Luftschauers wird aus der Zeitdifferenz der Signale zwischen den einzelnen Detektoren ermittelt.

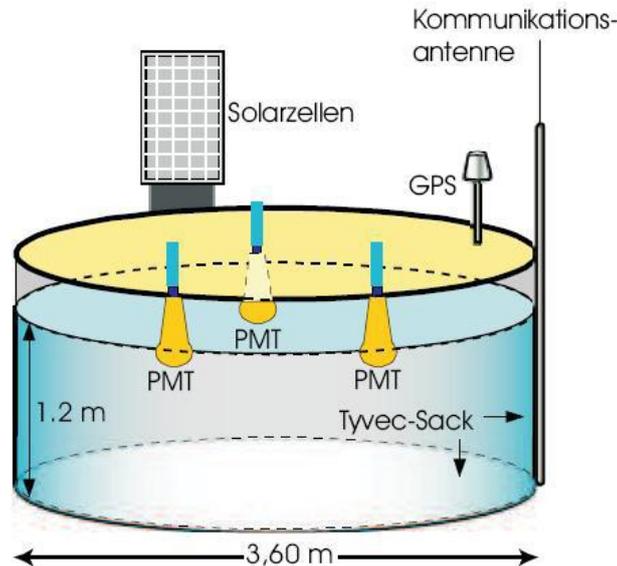


Abbildung 2.2: Schematischer Aufbau eines Bodendetektor-Tanks [Pie97]

### 2.1.2 Fluoreszenz-Detektoren

Jeweils sechs Fluoreszenzteleskope befinden sich innerhalb einer der halbkreisförmigen Stationen, wobei jedes einzelne Teleskop ein Gesichtsfeld von  $30^\circ \times 30^\circ$  und somit eine Teleskopstation ein Gesichtsfeld von  $30^\circ \times 180^\circ$  hat. Jedes dieser Teleskope ist ein Weitwinkel-Schmidt-Teleskop und besteht im Wesentlichen aus einem segmentierten sphärischen Spiegel mit einem Krümmungsradius von 3.4m, einem UV-Filter, einer Korrekturlinse und einer Photomultiplier-Kamera, die aus 440 Photomultipliern zusammengesetzt ist. Die Hauptachse ist  $16^\circ$  gegen den Horizont geneigt, sodass das Teleskop von  $1^\circ$  bis  $31^\circ$  über dem Horizont messen kann. In Abbildung 2.3 ist der Aufbau eines einzelnen Teleskopes dargestellt.

In der Atmosphäre regen geladene Teilchen im Schauer Stickstoffmoleküle der Luft an; Wenn diese in ihren Grundzustand zurückkehren, wird Fluoreszenzlicht im UV-Bereich ausgestrahlt, welches von den Teleskopen beobachtet werden kann. So kann auf diese Weise der Verlauf der Schauer verfolgt sowie anhand der Lichtstärke eine Energiemessung durchgeführt werden. Diese Methode ist erst ab Energien oberhalb von  $10^{17}$ eV und in mondlosen Nächten möglich, da dann der Untergrund an UV-Licht gering genug ist, um die Schauer zu identifizieren.

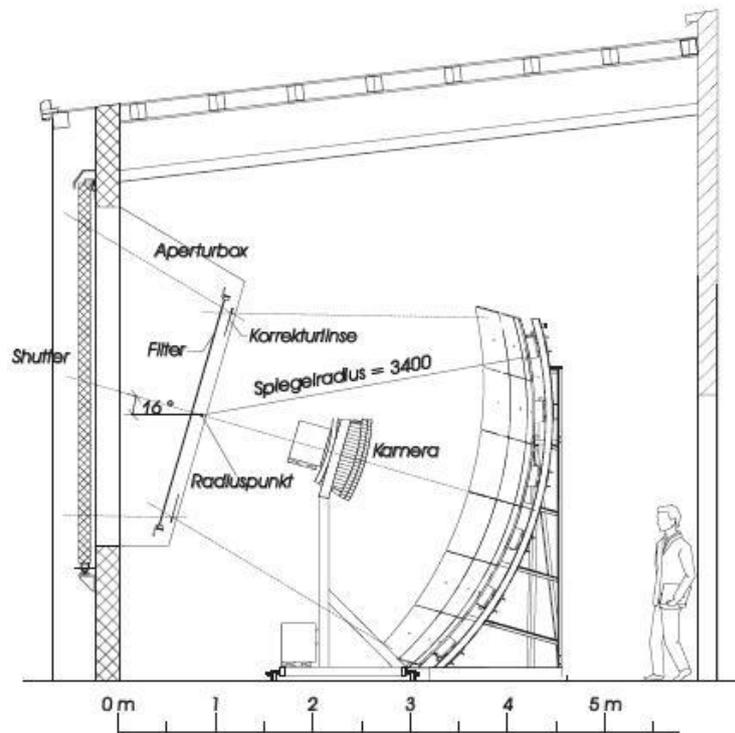


Abbildung 2.3: Schematischer Aufbau eines einzelnen Fluoreszenzteleskop [Bec05]

## Kapitel 3

# Erstellung von Himmelskarten

Die Anisotropieanalyse macht es notwendig, digitalisierte Himmelskarten der vom Pierre-Auger- Observatorium registrierten Schauer zu erstellen. Dies bedeutet, dass die kontinuierliche Verteilung der Einfaltsrichtungen diskretisiert werden muss, der Himmel wird also in kleine Flächen unterteilt. So werden bei der Analyse nicht mehr die exakten Richtungen betrachtet, sondern jeweils diese Flächen, die als Pixel bezeichnet werden, sowie die Schauer, die in einem bestimmten Zeitraum durch diese Pixel die Erde erreicht haben. Mit welcher Auflösung der Himmel pixelisiert werden muss, richtet sich in der Regel nach dem Auflösungsvermögen der Detektoren. Diese Himmelskarten machen es nun möglich, Zusammenhänge zwischen benachbarten Richtungen herzustellen und so Cluster zu bilden.

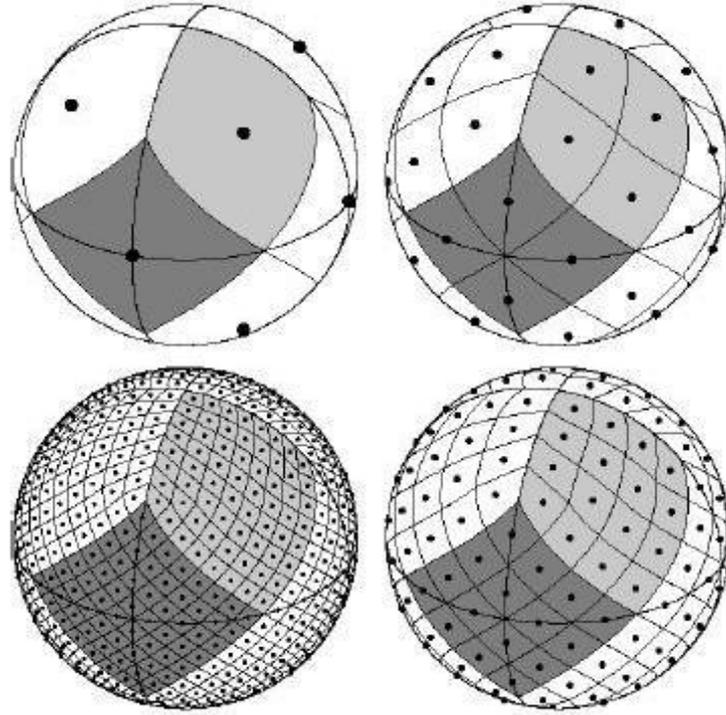
### 3.1 HEALPix

Das Hierarchical Equal Area iso-Latitude Pixelization (HEALPix) ist ein Paket aus Programmen und Routinen zur Pixelierung von Daten auf einer Sphäre. Mathematische Algorithmen verhelfen bei der Untersuchung und Verarbeitung von astronomischen Daten und eine Graphiksoftware zur Erstellung von diskretisierten, sphärischen Karten.

Das HEALPix-Pixelierungsschema basiert auf 3 Eigenschaften, die den Einsatz effizienter Routinen ermöglichen [Gor05]:

- Die Auflösung der Pixelisierung ist durch den Parameter  $N_{\text{side}}$  gegeben, der die Werte  $2^k$  für  $k = 0, 1, 2, \dots$  annehmen kann. Die Anzahl der Pixel  $N_{\text{pix}}$  ist über  $N_{\text{pix}} = 12 \times N_{\text{side}}^2$  von  $N_{\text{side}}$  abhängig, kann also nicht jeden beliebigen Wert annehmen. Somit enthält die geringste Auflösung 12 Pixel bei  $N_{\text{side}} = 1$ . Eine höhere Auflösung erhält man, indem man jeden Pixel in 4 weitere aufteilt.
- Die Mittelpunkte der Pixel liegen alle auf einem der  $N_{\text{ring}} = 4 \times N_{\text{side}} - 1$  Ringe mit konstanter Rektaszension.
- Die Pixel haben innerhalb einer Auflösung alle dieselbe Flächengröße, sodass keine Richtung auf der Sphäre durch die Pixelgröße ausgezeichnet wird.

Beispiele für pixelisierte Sphären nach dem HEALPix-Schema sind in Abbildung 3.1 zu sehen.



**Abbildung 3.1:** Pixelisierte Sphären nach dem HEALPix-Schema mit den ersten 4 Auflösungen: Links oben besitzt die Sphäre  $N_{\text{pix}} = 12$  Pixel ( $N_{\text{side}} = 1$ ). Die Anzahl der Pixel erhöht sich im Uhrzeigersinn auf  $N_{\text{pix}} = 48, 192$  und  $768$  ( $N_{\text{side}} = 2, 4$  und  $8$ ) [Gor05]

### 3.1.1 HEALPix Projektion

Sämtliche in dieser Arbeit bearbeiteten Daten und daraus erstellten Karten werden in der Mollweide-Projektion einer Sphäre in Äquatorkoordinaten (siehe Kapitel 3.2) erzeugt, wie im nächsten Kapitel, z.B. Abbildung 4.1 zu sehen ist. Die Deklination ist vertikal aufgetragen, mit seinem Nullpunkt auf dem Äquator bis zu  $+90^\circ$  im Norden bzw.  $-90^\circ$  im Süden. Die Rektaszension wird horizontal gemessen, startend bei  $0^\circ$  im Mittelpunkt der Projektion bis zu  $360^\circ$  Richtung Osten. Die Farben geben die Einträge jedes Pixels an, vom Minimum (dunkelblau) bis zum Maximum (dunkelrot).

### 3.1.2 Auflösung

Die Auflösung der Karten, d.h. die Anzahl der Pixel, sollte in etwa der Auflösung der Bodendetektoren von ca.  $1^\circ$  entsprechen. Dazu muss die Pixelanzahl berechnet werden: Der Raumwinkel  $\Omega$  ist definiert als ein Segment der Kugeloberfläche, welches sich unter einem Öffnungswinkel  $2\alpha$  auf einer Kugel ergibt. Der Größe des Raumwinkels ist gleich der Größe des Flächenstücks  $S$  im Verhältnis zum Quadrat des Radius  $R$  der Kugel:

$$\Omega = \frac{S}{R^2} \quad (3.1)$$

Daraus lässt sich eine Gleichung für den Raumwinkel bezogen auf den Öffnungswinkel und damit der Auflösung herleiten [Chi]. Die zur Berechnung notwendigen Größen sind in Abbildung 3.2 dargestellt:

$$\begin{aligned}
 S &= 2\pi R h & R &= h + H & \alpha &= \frac{a}{R} & H &= R \cos \alpha \\
 \Rightarrow h &= R - H = R - R \cos \alpha = R(1 - \cos \alpha) \\
 \Omega &= \frac{2\pi R h}{R^2} = \frac{2\pi R \cdot R(1 - \cos \alpha)}{R^2} \\
 \Omega &= 2\pi(1 - \cos \alpha)
 \end{aligned} \tag{3.2}$$

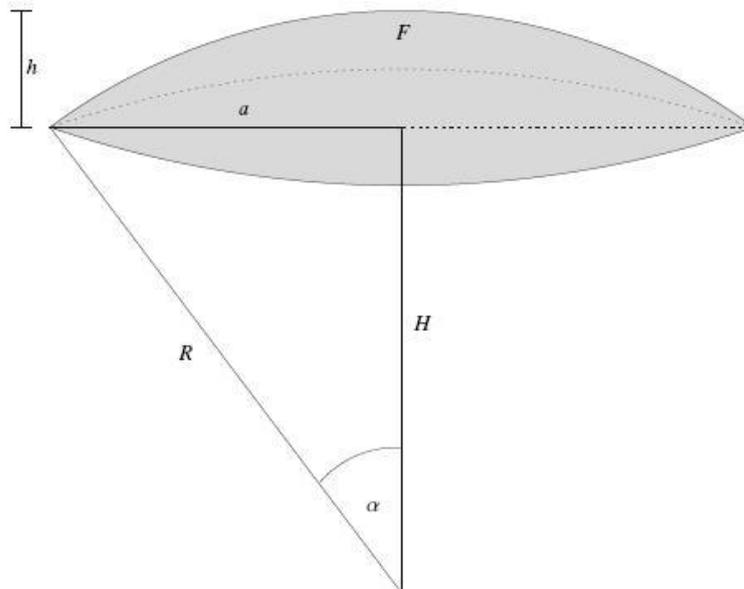
Das Flächenstück  $S$  ist, bezogen auf die Sphäre der Erde, die Größe eines Pixels. Somit erhält man die Pixelanzahl  $P$  daraus, wie oft die Fläche  $S$  auf eine komplette Kugeloberfläche passt:

$$\begin{aligned}
 P &= \frac{A_K}{S} = \frac{\Omega_K \cdot R^2}{\Omega \cdot R^2} = \frac{4\pi}{2\pi(1 - \cos \alpha)} \\
 P &= \frac{2}{1 - \cos \alpha}
 \end{aligned} \tag{3.3}$$

Eine Auflösung von  $1^\circ$  entspricht  $\alpha = 0.5^\circ$ , sodass sich für die Pixelanzahl ergibt:

$$P = \frac{2}{1 - \cos 0.5^\circ} \approx 52525$$

Da sich die Anzahl der Pixel in HEALPix nach  $N_{\text{pix}} = 12 \times N_{\text{side}}^2$  richtet, können keine Karten mit der berechneten Zahl erstellt werden. Daher enthalten sämtliche Karten in dieser Arbeit 49152 Pixel ( $N_{\text{side}} = 64$ ).



**Abbildung 3.2:** Darstellung des Raumwinkels  $\alpha$  sowie allen Relationen zur Kugeloberfläche, mit denen eine Gleichung für  $\alpha$  hergeleitet werden kann [Chi]

## 3.2 Koordinatensysteme

Himmelsrichtungen können mithilfe von Koordinaten genau beschrieben werden. Dazu ist es jedoch wichtig zu wissen, aus welcher Perspektive und damit in welchem Koordinatensystem die Richtungen bestimmt werden. Bestimmt man die Koordinaten aus der Sicht des Observatoriums, so geht man von den sogenannten horizontalen Koordinaten aus. Diese sind sowohl von der Position des Beobachters, als auch von der Zeit abhängig. Beim Äquatorkoordinatensystem handelt es sich um einen festen Bezugsrahmen, der in dieser Arbeit für die Beschreibungen der HEALPix Maps verwendet wird.

### 3.2.1 Horizontale Koordinaten

Schauerrichtungen, die vom Observatorium registriert und gespeichert werden, werden in lokalen Horizontalkoordinaten gemessen. Die Referenzebene ist die Ebene tangential zur Erdkugel durch den Standpunkt des Beobachters und durchschneidet die Himmelskugel am Horizont. Man geht von zwei Koordinaten aus: der Azimuth gibt den Winkelabstand zwischen dem Fußpunkt des durch die Einfallsrichtung gehenden Höhenkreises und dem Südpunkt, dem sogenannten Meridian, an. Der Zenitwinkel ist der Winkelabstand zwischen dem Punkt senkrecht über dem Beobachter, dem Zenit, und der Einfallsrichtung. Aufgrund der Erdrotation sind die Koordinaten zeitabhängig, d.h. ein bestimmter Punkt ist erst nach einem siderischen Tag, also einer vollen Erdumdrehung, wieder in der exakt identischen Richtung sichtbar. Daher ist bei der Analyse von Schauerdaten die Zeit immer von größter Bedeutung.

### 3.2.2 Äquatorkoordinaten

Das Äquatorkoordinatensystem nimmt die Erde als Ausgangspunkt und den Äquator als Bezugsebene, Nulllinie. Jede Himmelsrichtung wird jetzt in Beziehung zu Erde und Himmelsäquator beschrieben, in direkter Analogie zu den bekannten geographischen Koordinaten auf der Erdoberfläche. Die Deklination  $\delta$  gibt den Winkelabstand zwischen Einfallsrichtung und Äquator Richtung Norden bzw. Süden an, sodass der Nordpol bei  $\delta = +90^\circ$ , der Südpol bei  $\delta = -90^\circ$  liegt. Wie bei der geographischen Länge wird die zweite Koordinate auf dem Äquator gemessen. Der gewählte Nullpunkt wird Frühlingspunkt genannt, da die Sonne genau zu Frühlingsbeginn in diesem Punkt steht. Von dort aus zählt man bis zum Schnittpunkt von Äquator und Deklinationskreis, der senkrecht zum Äquator durch die zu bestimmende Richtung verläuft.

# Kapitel 4

## Significance Map

Grundlegend für die Suche und Konstruktion von Clustern ist die Kenntnis der Signifikanzen in alle möglichen Himmelsrichtungen. Als Signifikanz wird die statistische Abweichung der gemessenen von der erwarteten Anzahl an Events aus einer bestimmten Richtung bezeichnet. Somit ist der erste Schritt die Erstellung einer sogenannten Significance Map aus einer Healpix Map, die für jeden Pixel die dazugehörige Signifikanz liefert.

### 4.1 Coverage Map

Eine Coverage Map ist eine Karte, die die erwartete Anzahl an Ereignissen anzeigt, die aus allen Himmelsrichtungen von den Detektoren registriert werden. Es gibt zwei grundlegende Methoden, solch eine Karte zu erstellen: Shuffling and das semianalytische Verfahren.

Beim Shuffling werden die jeweils drei Schauerparameter realer Ereignisse in Listen übertragen und durch zufällige Auswahl eines Zenitwinkels, eines Azimutwinkels und der Ankunftszeit falsche Ereignisse erstellt und in die Map gefüllt.

Beim semianalytischen Verfahren wird ein Fit für die Verteilung des Zenitwinkels bestimmt und so mit Werten dieses Fits wie beim Shuffling falsche Ereignisse konstruiert.

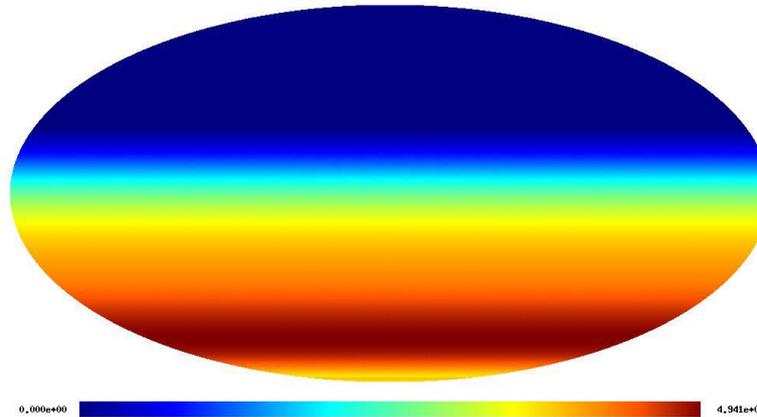
Beide Methoden führen zu vergleichbaren Ergebnissen.

In Abbildung 4.1 ist die in dieser Arbeit verwendete Coverage Map zu sehen, die von M. Grigat im Rahmen seiner Diplomarbeit erstellt wurde. Der verwendete Datensatz zur Berechnung der Coverage Map enthält alle 445815 Schauer, die von den Bodendetektoren von Januar 2004 bis Dezember 2006 registriert wurden und die T5 Trigger Bedingung erfüllt haben. Für die genaue Vorgehensweise zur Erstellung der Coverage Map wird auf die Diplomarbeit von M. Grigat verwiesen [Gri07].

Das Pierre-Auger-Observatorium befindet sich auf der Karte bei einer Deklination von  $\delta = -35.25^\circ$ ; die eingeschränkte Sicht der Bodendetektoren kann einfach durch Addieren der Zenitwinkel-Obergrenze berücksichtigt werden. Da für die Ereignis Selektion Zenitwinkel mit Werten von maximal  $60^\circ$  benutzt worden sind, ist der Himmel bis zu einer Deklination von  $\delta = -35.25^\circ + 60^\circ = 24.75^\circ$  sichtbar. In der Karte ist zu erkennen, dass der Himmel oberhalb dieser Grenze nicht berücksichtigt wurde. Daher besitzen sämtliche Pixel in diesem Bereich keine Einträge und bilden in der Karte eine große dunkelblaue Fläche. Im Süden entsteht keine Grenze durch den Zenitwinkel, da nach  $\delta = -35.25^\circ - 60^\circ = -95.25^\circ$  der ganze Himmel südlich des Observatoriums abgedeckt wird; ein Kreis mit einem Winkelradius von  $5.25^\circ$  um

den Südpol ist jederzeit vom Observatorium sichtbar.

Desweiteren erkennt man in der Coverage Map die konstanten Einträge entlang der Rektaszension mit einem Maximum von 49410 Schauern pro Pixel im Deklinationsbereich von  $-60^\circ$  bis  $-75^\circ$ , sowie einem kontinuierlichen Abfall der Einträge Richtung Norden bzw. Süden vom Maximum.



**Abbildung 4.1:** Die in dieser Arbeit verwendete Coverage Map; sie besitzt eine konstante Anzahl an Einträgen entlang der Rektaszension mit einem Maximum von 49410, sowie keinen Einträgen oberhalb einer Deklination von  $\delta=24.75^\circ$

## 4.2 Event Map

Als Event Map wird eine Karte bezeichnet, in der unveränderte Schauerdaten importiert wurden, d.h. nur tatsächliche vom Observatorium gemessene Schauer. Im folgenden wird anhand der Coverage Map solche Karten simuliert.

Der Unterschied zwischen einer Event Map und einer Coverage Map besteht darin, daß auch statistische Fluktuationen beachtet werden, d.h. die Anzahl der Ereignisse pro Pixel höher oder niedriger sein kann als erwartet. Dies erreicht man in den simulierten Karten dadurch, indem die Anzahl der Einträge in den Pixeln der Coverage Map verschmiert werden, d.h. diese durch zufällige Zahlen aus Poisson- oder Gaußverteilungen ersetzt werden.

Die Verschmierung der Anzahl der in der Coverage Map pro Pixel gegebenen Anzahl von Schauern wurde auf folgende Weise durchgeführt:

0 Einträge  $\rightarrow$  0 Einträge

bis zu 100 Einträge  $\rightarrow$  Poissonverteilung

über 100 Einträge  $\rightarrow$  Gaußverteilung

Anhand der gegebenen Coverage Map wurde durch diese Methode eine Event Map erstellt, wobei darauf geachtet wurde, dass sämtliche Pixel oberhalb der Deklination von  $24.75^\circ$ , also alle Richtungen, die aufgrund des eingeschränkten Zenitwinkels nicht erfasst wurden, von der Verschmierung ausgeschlossen werden. Dies wurde dadurch erreicht, dass sämtliche Pixel mit Eintrag 0 nicht durch Zufallszahlen

gefüllt wurden. In diesem Fall war die Vorgehensweise möglich, da die Bodendetektoren laut der Coverage Map aus allen möglichen Richtungen Schauer registriert haben sollten, also alle Pixel südlich der Deklination von  $24.75^\circ$  Einträge besitzen. Andernfalls hätte unterschieden werden müssen, wo sich die leeren Pixel befinden, da bei denen, die im vom Observatorium sichtbaren Bereich liegen, ebenfalls statistisch Fluktuationen angenommen werden müssten.

Jeder Pixel mit  $N \leq 100$  Einträgen wird durch eine Zahl  $K$  neu besetzt, die sich zufällig aus einer Poisson-Verteilung

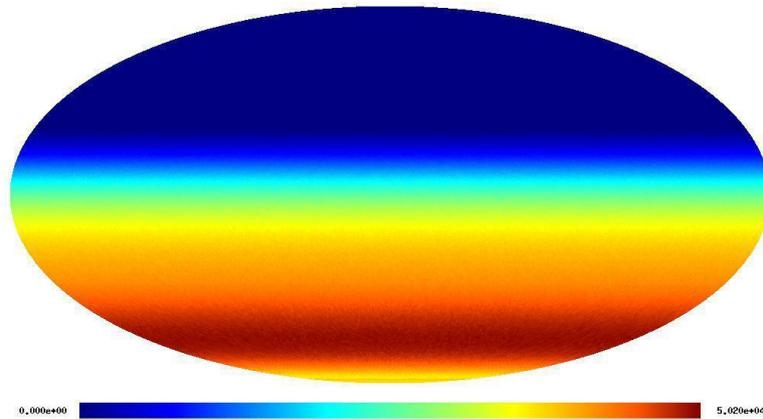
$$P(N) = \frac{e^{-N}}{K!} \cdot N^K \quad (4.1)$$

ergibt.

Alle Pixel, die in der Coverage Map mehr als 100 Einträge besitzen, werden in der Event Map mit einer Zufallszahl  $K$  aus einer Gaußverteilung gefüllt. Hat ein Pixel  $N$  Einträge, so wird zufällig ein Wert aus einer Gaußverteilung mit einem Erwartungswert  $N$  und einer Standardabweichung  $\sigma = \sqrt{N}$  gewählt:

$$P(N) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{1}{2} \frac{(K-N)^2}{N}} \quad (4.2)$$

In Abbildung 4.2 ist eine mögliche Event Map dargestellt. Anders als bei der Coverage Map sind keine scharfen Grenzen zwischen den horizontalen Farbstreifen mehr zu erkennen; stattdessen verschwimmen diese, was zeigt, dass die konstante Anzahl an Einträgen entlang der Rektaszension nicht mehr vorhanden ist. Dennoch bleibt die grobe Struktur erhalten: die maximale Anzahl im Deklinationsbereich von  $-60^\circ$  bis  $-75^\circ$  sowie der kontinuierliche Abfall der Einträge Richtung Norden bzw. Süden vom Maximum.



**Abbildung 4.2:** Eine berechnete Event Map, auf die sich die Ergebnisse in dieser Arbeit beziehen; aufgrund der Neubesetzung der Einträge durch poisson- oder gaußverteilte Zufallszahlen ist die konstante Anzahl entlang der Rektaszension aufgehoben, mit maximal 50200 Einträgen pro Pixel

### 4.3 Significance Map

Die Signifikanz ist nichts anderes als die Differenz der Einträge zwischen der Event Map und der Coverage Map, ausgedrückt in Einheiten der Standardabweichung. Somit muss für jeden einzelnen Pixel die Signifikanz auf folgende Weise berechnet werden:

$K \rightarrow$  Einträge der Event Map

$N \rightarrow$  Einträge der Coverage Map

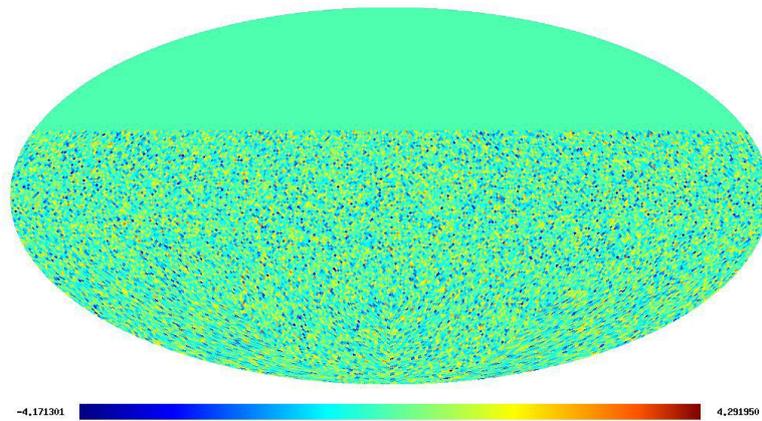
$$\text{Signifikanz} = \frac{K - N}{\sqrt{N}} \quad (4.3)$$

In Abbildung 4.3 ist die Significance Map zu sehen, die sich aus der Coverage Map sowie der Event Map aus Abbildung 4.2 ergeben hat. Es ist keinerlei Struktur mehr zu erkennen; Es ergaben sich Signifikanzen von  $-4.17\sigma$  bis  $4.29\sigma$ , die, wie zu erwarten war, willkürlich auf der Karte verteilt sind. Ob die Karte korrekt erstellt wurde, lässt sich anhand der Eigenschaft ihrer Signifikanzverteilung überprüfen: die Einträge der Event Map ergeben sich fast ausschließlich aus Zufallszahlen einer Gaußverteilung (siehe Kapitel 3.2) mit einem Erfahrungswert  $N$  und einer Standardabweichung  $\sigma = \sqrt{N}$  ( $N, K$  haben die gleiche Bedeutung wie oben). Die Differenz  $K-N$  lässt sich so interpretieren, dass die jeweilige Gaußverteilung auf der x-Achse um  $N$  Einheit nach links verschoben wird und somit einen Erfahrungswert von Null besitzt. Die Division durch  $\sqrt{N}$  normiert letztendlich sämtliche Erwartungswerte auf 1. Somit sind die Signifikanzen der Significance Map nichts anderes als Zufallszahlen aus einer Normalverteilung um Null und  $\sigma = 1$ , so dass die Signifikanzverteilung der Karte die Form dieser Gauß-Funktion annehmen sollte.

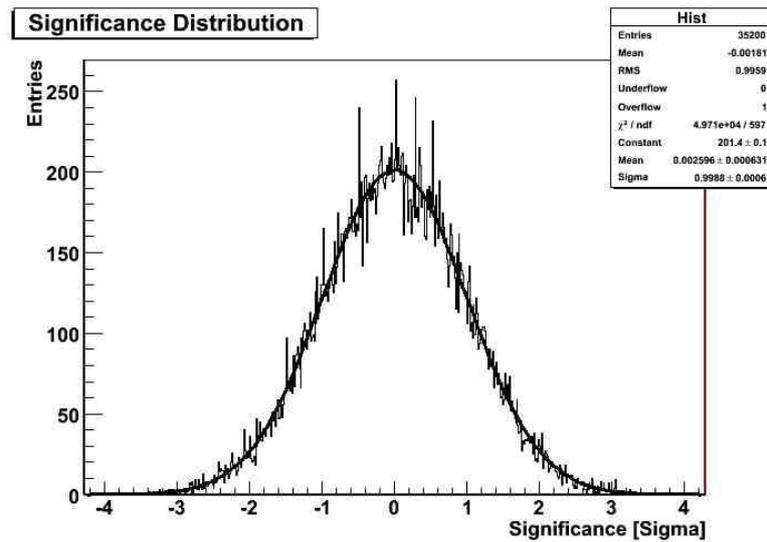
In Abbildung 4.4 ist die Verteilung für die gegebene Karte aufgetragen; ein Gauß-Fit gibt mit dem Parameter  $\sigma = 0.9988$  und dem Erwartungswert  $0.0026$  ein erwartetes Ergebnis. Im Histogramm wurden nur die verschmierten Pixel einbezogen, weshalb dieses nur 35200 Einträge besitzt, während in der graphischen Darstellung der Karte (Abbildung 4.3) für alle Pixel im nördlichen Bereich, die nicht vom Observatorium erfasst werden können,  $\sigma = 0$  gesetzt wurde.

Durch die Verteilung ist auch zu erklären, warum die Karte auf den ersten Blick aus hauptsächlich aus gelben und blauen Punkten auf hellgrünem Grund besteht: Aufgrund der besonderen Eigenschaft der Gaußverteilung liegen 68.27% der Werte im Bereich von  $-\sigma$  bis  $\sigma$ , hier also etwa 24000 Pixel zwischen  $-1$  und  $1$  (grün). 27.18% liegen im Bereich von  $-2\sigma$  bis  $-\sigma$  bzw.  $\sigma$  bis  $2\sigma$ , die in der Karte die gelben/-gelbgrünen sowie hellblauen/blauen Pixel ausmachen. Orangene bzw. rote sowie dunkelblaue Pixel, die in der Karte kaum zu finden sind, werden im weiteren bei der Clustersuche von großer Bedeutung sein.

Die im nachfolgenden Kapitel beschriebenen Clusteralgorithmen und angegebenen Ergebnisse beziehen sich jeweils auf diese Significance Map.



**Abbildung 4.3:** Die Significance Map, berechnet mit der Coverage Map in Abbildung 4.1 und der Event Map in Abbildung 4.2; die verschiedenen Signifikanzen sind willkürlich verteilt, mit Extrema von  $-4.17\sigma$  und  $4.29\sigma$



**Abbildung 4.4:** Die Signifikanzverteilung der Significance Map in Abbildung 4.3; ein Gauß-Fit gibt ein erwartetes Ergebnis mit einem Erwartungswert  $\text{mean} \approx 0$  und der Standardabweichung  $\sigma \approx 1$

# Kapitel 5

## Clustersuche

Es wurden zwei aufeinander aufbauende Algorithmen entwickelt. Beide basieren darauf, dass um die höchstsignifikanten Pixel in der Significance Map mittels der Suche nach den nächsten Nachbarn Cluster erstellt werden.

### 5.1 Methode 1

Der erste Algorithmus besteht aus einer einfachen Methode, Cluster aus Pixeln mit den höchsten positiven oder negativen Signifikanzen sowie seinen nächsten Nachbarn zu erstellen.

Die Clustersuche erfolgt nach folgenden 4 Schritten:

1. Suche den Pixel mit der größten positiven Signifikanz oberhalb von  $3\sigma$  sowie seine nächsten Nachbarn

Die Nachbarsuche erfolgt mittels eines  $3 \times 3$ -Rasters, sodaß insgesamt 8 Nachbarn um die größte Signifikanz gefunden werden.

2. Diese Pixel werden zu einem Cluster zusammengefaßt und für diesen die Gesamtsignifikanz berechnet

Die Gesamtsignifikanz wird in der Weise berechnet, dass die Signifikanzen aller Pixel im Cluster quadratisch addiert werden, wobei das Vorzeichen erhalten bleibt:

Signifikanz eines Pixels:  $s_i$  und Gesamtsignifikanz:  $S$

$$S^2 = \sum \text{sign}(s_i) \cdot s_i^2$$

$$S = \text{sign}(S^2) \sqrt{S^2} \quad (5.1)$$

3. Die Schritte 1 und 2 werden jeweils für die nächstkleineren Signifikanzen oberhalb  $3\sigma$  wiederholt

Somit wird die Clustersuche mit dem Pixel mit der zweitgrößten Signifikanz fortgesetzt, dann mit der drittgrößten usw., bis alle Punkte oberhalb von  $3\sigma$  berücksichtigt worden sind. Hierbei wird die Bedingung beachtet, daß sich keine Cluster überschneiden, d.h. dass jeder Pixel höchstens zu

einem Cluster gehören darf. Ist also ein Pixel der nächste Nachbar von zwei hochsignifikanten Pixeln oberhalb  $3\sigma$ , so wird dieser der höheren Signifikanz zugeordnet. Im Fall, dass ein Pixel eine statistische Abweichung oberhalb  $3\sigma$  besitzt, jedoch der nächste Nachbar einer höheren Signifikanz und damit bereits in einem Cluster enthalten ist, so wird für diesen kein neuer Cluster erzeugt.

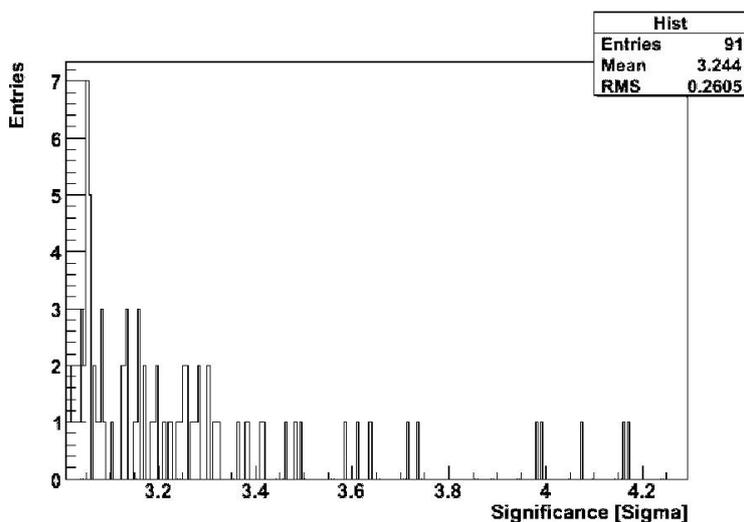
4. Der Algorithmus wird für die größten negativen Signifikanzen unterhalb  $-3\sigma$  wiederholt

In diesem Fall werden die Cluster in Reihenfolge der größten negativen Signifikanzen gesucht. In der gegebenen Significance Map (Abbildung 4.3) beginnt somit der Algorithmus mit dem Pixel mit  $-4.17\sigma$ . Da die Signifikanzverteilung einer Normalverteilung mit  $\sigma \approx 1$  entspricht, besitzen nur 0.27% aller Pixel eine statistische Abweichung von mehr als  $3\sigma$  bzw.  $-3\sigma$ , was bei einer Map mit insgesamt 35200 signifikanten Pixeln 95 Punkten entspricht. Dies sollte in etwa auch der Anzahl der gefundenen Cluster wiedergeben.

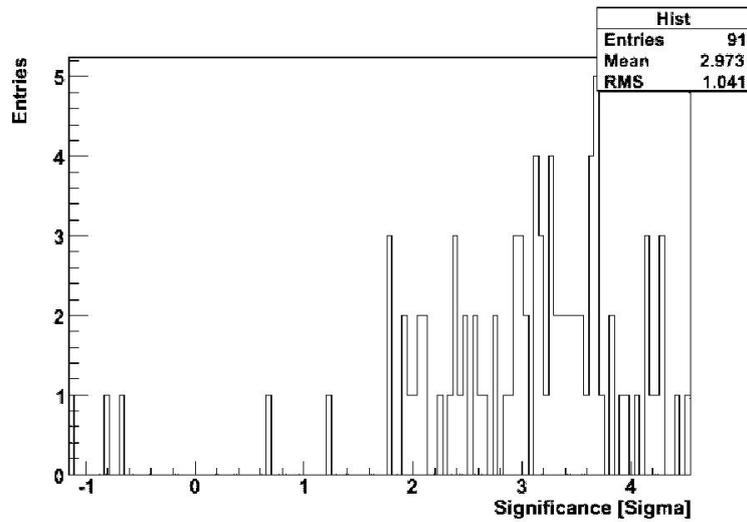
In der gegebenen Significance Map (Abbildung 4.3) wurden mit dem Algorithmus 91 Cluster gefunden. In den Abbildungen 5.1 sowie 5.2 sind die Signifikanzverteilungen, sowohl für den einzelnen zentralen Pixel als auch für den ganzen Cluster dargestellt. Hierbei wurden alle Cluster beachtet; die Signifikanzen, die sich aus Clustern mit zentralen Pixeln negativer Abweichung ergeben haben, wurden jeweils mit  $-1$  multipliziert, das heißt an der y-Achse gespiegelt. Da sich die positiven und negativen Signifikanzen aufgrund der Symmetrie der Gaußfunktion gleich verteilen, wird das Ergebnis nicht verfälscht, sondern durch diese Spiegelung eine mögliche Symmetrie aufgrund doppelt so vieler Werte besser erkennbar.

Die Verteilung der Signifikanzen für den zentralen Pixel entspricht dem Ende der Signifikanzverteilung der Significance Map, also der Normalverteilung oberhalb  $3\sigma$ .

Die Clustersignifikanzen verteilen sich um einen Wert von  $3\sigma$ , wie es auch zu erwarten war: Der zentrale Pixel besitzt eine Signifikanz von etwa  $3\sigma$  (zwischen  $3\sigma$  und  $4.29\sigma$ ), dessen Nachbarn als Zufallszahlen einer Gaußverteilung (siehe Kapitel 4.3) im Schnitt eine Gesamtsignifikanz von 0.



**Abbildung 5.1:** Die Signifikanzverteilung der zentralen Pixel aller 91 gefundenen Cluster; sie entspricht der Signifikanzverteilung der gegebenen Significance Map oberhalb  $3\sigma$



**Abbildung 5.2:** Die Signifikanzverteilung der 91 gefundenen Cluster; alle Signifikanzen verteilen sich wie erwartet um  $3\sigma$

## 5.2 Methode 2

Die 2.Methode zur Clustersuche baut auf dem 1.Algorithmus auf, konstruiert also ebenfalls Cluster anhand der Nächsten-Nachbar-Suche, die bei einem hochsignifikanten Pixel startet.

Da in der 1.Methode bei einem Schnitt von  $3\sigma$  sehr viele Cluster entstehen und man nur die höchstsignifikanten Stellen finden möchte, wurde nun der Schnitt für den zentralen Pixel auf  $4\sigma$  erhöht.

Die Suche der Cluster folgte folgenden 6 Schritten:

1. Suche den Pixel mit der größten positiven Signifikanz oberhalb von  $4\sigma$  sowie seine nächsten Nachbarn mit Signifikanzen von mindestens  $1\sigma$

Die Suche erfolgt wie in der ersten Methode, wobei nun eine Bedingung für die Nachbarn zu erfüllen ist, um in den Cluster aufgenommen zu werden.

2. Suche für dessen nächste Nachbarn wiederum die Nachbarn mit Signifikanzen oberhalb von  $1\sigma$
3. Wiederhole den 2.Schritt für die jeweiligen neuen Pixel solange, bis keine Punkte oberhalb des Cuts mehr gefunden werden

Damit ist die Clustergröße im Gegensatz zur 1.Methode variabel; so werden während der Kalkulation für jeden neuen Pixel im Cluster die nächsten Nachbarn gesucht und zum Cluster hinzugefügt, bis die Suche keinen Pixel oberhalb  $1\sigma$  mehr findet.

4. Berechne die Gesamtsignifikanz des Clusters

Die Berechnung erfolgt wie zuvor:

Signifikanz eines Pixels:  $s_i$  und Gesamtsignifikanz:  $S$

$$S^2 = \sum \text{sign}(s_i) \cdot s_i^2$$

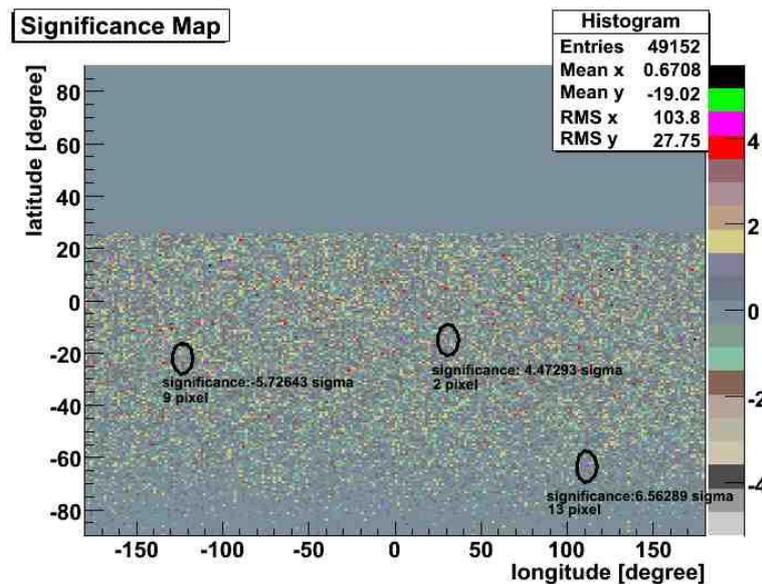
$$S = \text{sign}(S^2) \sqrt{S^2} \quad (5.2)$$

- Setze die Clustersuche fort, indem die Schritte 1 bis 4 jeweils für die Pixel mit den nächstkleineren Signifikanzen oberhalb  $4\sigma$  wiederholt werden

Somit wird die Clustersuche mit den Pixel mit den nächstgrößeren Signifikanzen fortgesetzt, mit der Bedingung, dass jeder Pixel höchstens zu einem Cluster gehören darf.

- Führe für die größten negativen Signifikanzen unterhalb  $-4\sigma$  den Algorithmus (Schritt 1 bis 5) noch einmal durch

In der gegebenen Significance Map (Abbildung 4.3) wurde mit diesem Algorithmus drei Cluster gefunden, wie in Abbildung 5.3 zu sehen ist, zwei mit positiver und einer mit negativer Signifikanz: ein Cluster bestehend aus 13 Pixeln mit einer Gesamtsignifikanz von  $6.56\sigma$ , ein Cluster mit zwei Pixeln und der Signifikanz  $4.47\sigma$  und einen 9-Pixel-Cluster mit  $-5.73\sigma$ .



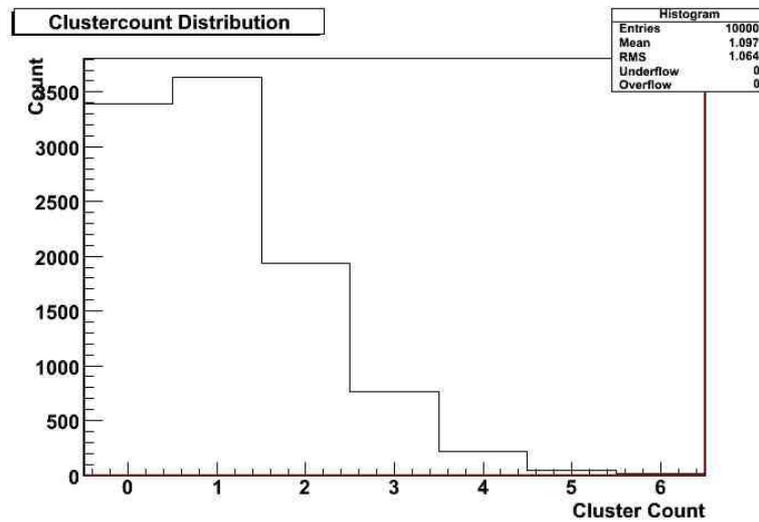
**Abbildung 5.3:** die Significance Map wie in Abbildung 4.3, in der die drei gefundenen Cluster nach der 2.Methode eingezeichnet sind

Welche Ergebnisse jeweils mit diesem Algorithmus zu erwarten sind, wurde anhand einer Langzeitstudie untersucht. Dazu wurden Cluster, in diesem Fall nur mit zentralem positiver Signifikanz, in 10.000 verschiedenen Event Maps gesucht und die Ergebnisse gesammelt.

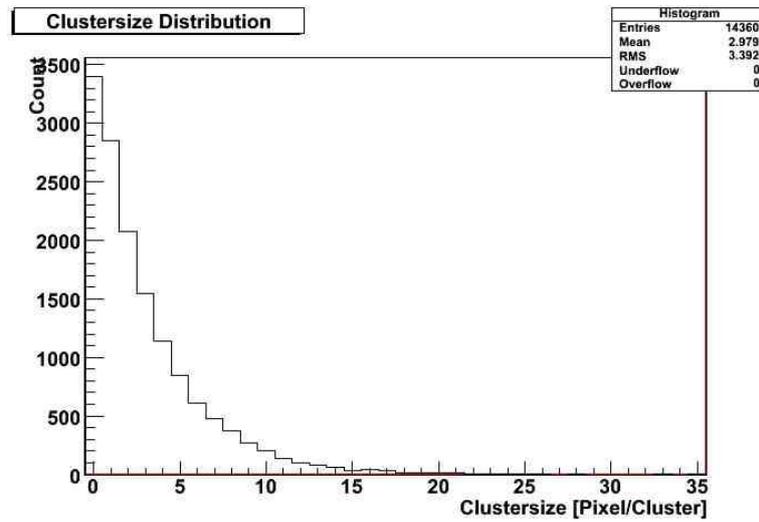
Abbildung 5.4 zeigt zunächst, wieviele Cluster jeweils in diesen 10.000 Maps gefunden wurden. Man sieht, dass sich in den meisten Fällen kein oder nur ein Cluster, und mit dem Algorithmus nie mehr als sechs Cluster ergeben haben. Dieses ist einfach dadurch zu erklären, dass mit den festgelegten Schnitt von  $4\sigma$  kaum noch Cluster gefunden werden können.

Die jeweilige Größe, also die Anzahl der Pixel pro Cluster, ist in Abbildung 5.5 aufgetragen. Es ist klar ein exponentieller Abfall zu erkennen mit einem Maximum bei Eins. In dieser Abbildung zu beachten, dass in allen Fällen, in denen der Algorithmus keine Treffer lieferte, eine Clustergröße von Null gezählt wurde. Der Verlauf der Verteilung kommt dadurch zu Stande, dass nur Pixel mit einer Signifikanz von mindestens  $1\sigma$  in Frage kommen, und damit aufgrund der gaußverteilten Signifikanzen nur 16% aller möglichen Punkte. Da sich diese Pixel zufällig auf der Karte verteilen, fällt die Wahrscheinlichkeit, dass zwei oder mehr Punkte oberhalb  $1\sigma$  direkt nebeneinander liegen, nach Gesetzen der Stochastik exponentiell ab.

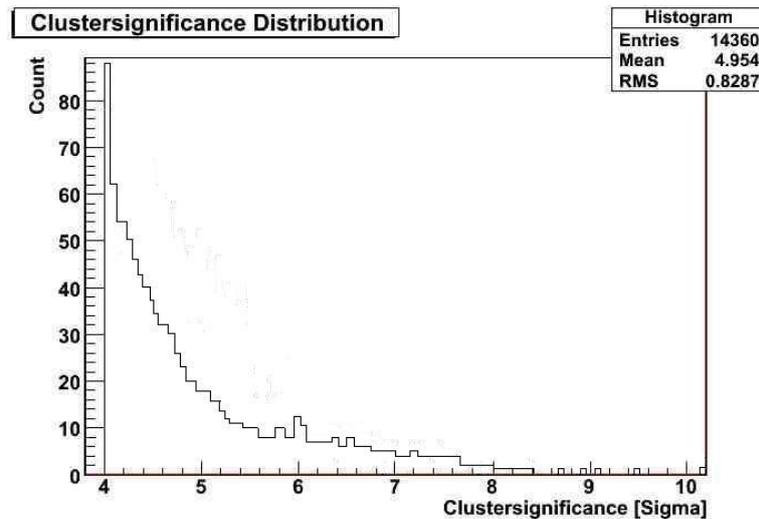
In Abbildung 5.6 ist die Verteilung aller Clustersignifikanzen dargestellt. Man sieht auch hier einen exponentiellen Abfall, wobei die Gesamtsignifikanz immer größer als  $4\sigma$  ist. Der Peak bei etwa  $4\sigma$  entsteht dadurch, dass, wie im vorherigen Histogramm zu sehen ist, die meisten Cluster nur aus einem Pixel bestehen.



**Abbildung 5.4:** Die Anzahl der Cluster pro Karte für 10000 verschiedene Event Maps zeigt in etwa  $\frac{2}{3}$  der Fälle Null oder 1, und nie mehr als 6 gefundene Cluster



**Abbildung 5.5:** Die Verteilung der Größen aller Cluster, die in den 10000 Event Maps gefunden wurden, zeigt einen exponentiellen Abfall von Null bis zu einem Maximum von 36 Pixel pro Cluster



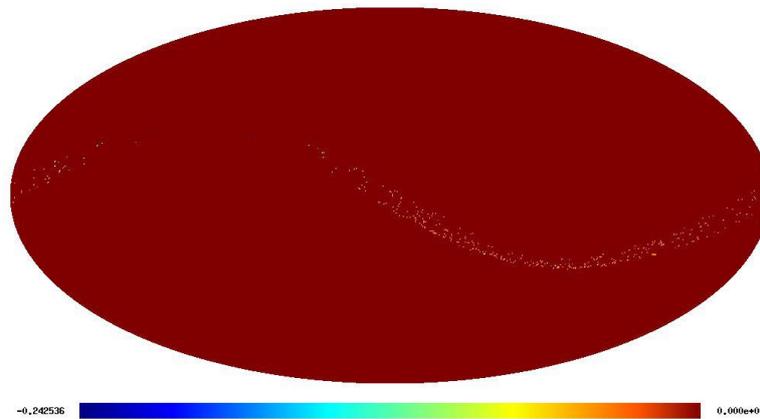
**Abbildung 5.6:** Die Verteilung der Signifikanzen aller Cluster, die in den 10000 Event Maps gefunden wurden, fällt von Null exponentiell ab bis zu einem Maximum von  $11\sigma$

### 5.2.1 Anwendung und Effizienztest

Um den zweiten Algorithmus auf seine Effizienz hin zu testen, wurde dieser an einem Beispiel angewendet. Dazu wurde eine Coverage Map und eine Event Map erstellt, in der der Verlauf des Mondes implementiert wurde. Der Mond fungiert als Störquelle, da er Schauer auf dem Weg zur Erde abfängt, und somit vom Observatorium weniger Teilchen registriert werden als erwartet; dies wurde in der

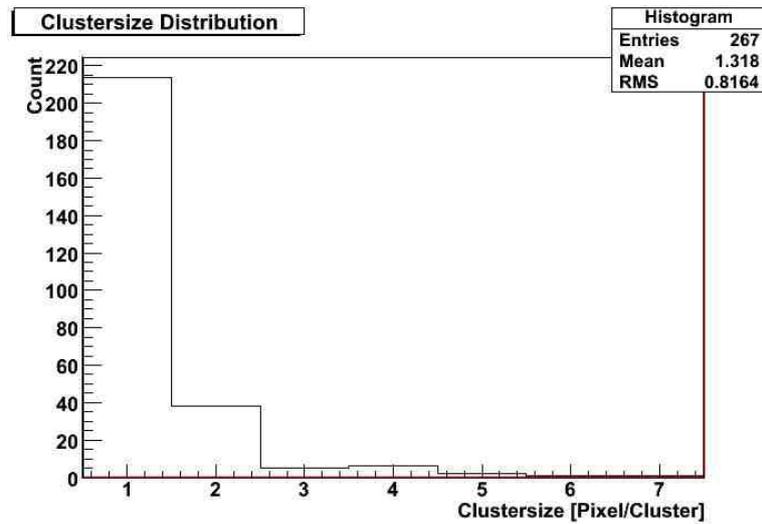
Event Map berücksichtigt, indem die Pixel im Bereich des Mondschattens weniger Einträge besitzen als in der Coverage Map. Anders als in Kapitel 4.2 wurden die statistischen Fluktuationen vernachlässigt, so dass die Einträge der Event Map außerhalb der Mondspur gegenüber der Coverage Map unverändert blieben. Somit wurde allein der idealisierte Mondschatten betrachtet.

Der erste Schritt ist es, ein Significance Map zu erstellen, wie in Kapitel 4.3 geschildert wurde. In Abbildung 5.7 ist diese Karte zu sehen: Anhand der negativen Signifikanzen ist die Spur des Mondes klar zu erkennen, in allen anderen Bereichen ist die Signifikanz Null. Der Einfluss des Mondes auf die Anzahl der einfallenden Schauer ist nur sehr gering, wie die sehr kleinen Signifikanzen zeigen (Maximum bei etwa  $-0.24\sigma$ ). Somit würde sein Einfluss vollkommen untergehen, wenn auch die statistischen Fluktuationen mit beachtet würden. Die Mondspur wäre also nicht mehr erkennbar.



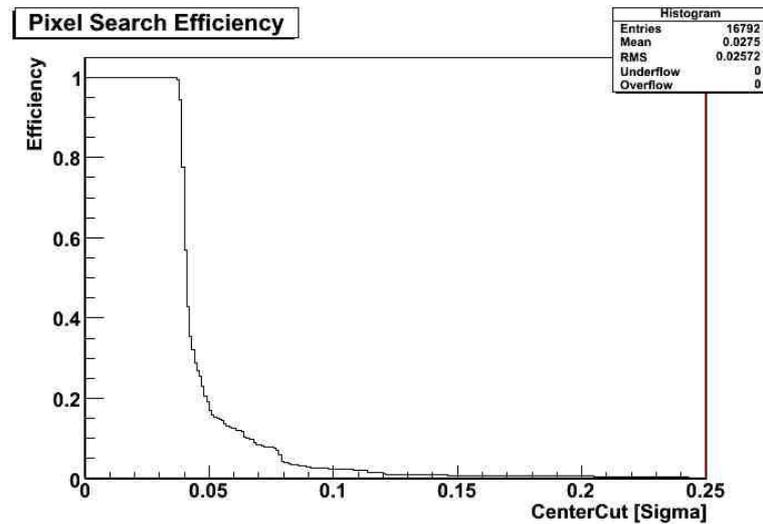
**Abbildung 5.7:** Die Significance Map, erstellt anhand der Coverage Map und der Event Map, in der der Mondverlauf implementiert wurde; entlang der Mondspur sind die Signifikanzen ungleich Null zu erkennen, mit einem Maximum von  $-0.24\sigma$

Zunächst wird getestet, ob der Algorithmus sämtliche 352 Pixel mit Signifikanzen ungleich null findet. Dazu wurde der Schnitt für den zentralen Pixel und die Nachbarpixel auf Null gesetzt und auf die Significance Map angewendet. Wie Abbildung 5.8 zeigt, wurden 267 Cluster unterschiedlicher Größe gefunden, wobei diese sämtliche 352 signifikanten Pixel enthielten.



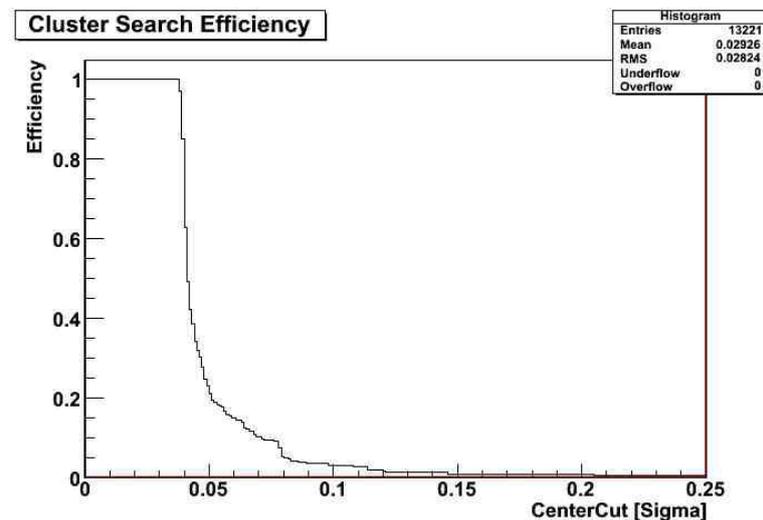
**Abbildung 5.8:** Die Verteilung der Clustergrößen bei Signifikanzschwellen des zentralen Pixels und der Nachbarpixel von  $0\sigma$ ; Addiert man alle Clustergrößen, so ist zu erkennen, dass alle 352 signifikanten Pixel in den 267 Cluster enthalten sind

Nun wurde für den weiteren Effizienztest die Signifikanzschwelle für den zentralen Pixel eines Clusters schrittweise in negativer Richtung erhöht, bis der Algorithmus keine Ergebnisse mehr lieferte. In Abbildung 5.9 ist die Effizienzkurve dargestellt, wobei auf der x-Achse der Betrag des Signifikanzschnitts und auf der y-Achse die Effizienz aufgetragen ist. Die Effizienz 1 entspricht der Tatsache, dass alle 352 Pixel vom Algorithmus gefunden und aus ihnen Cluster erstellt werden. Man sieht, dass bis zu einem Cut von  $0.038\sigma$  alle Pixel gefunden werden, was mit den Daten der Significance Map übereinstimmt; sämtliche Pixel, die aufgrund des Mondschattens weniger Einträge zählen als erwartet, weisen eine Signifikanz von mindestens  $-0.038\sigma$  auf. Ab diesem Punkt fällt die Kurve stark ab, so dass einem Schnitt von  $0.05\sigma$  nur noch 15% und bei  $0.1\sigma$  5% der signifikanten Pixeln gefunden werden. Oberhalb eines Cuts von  $0.24\sigma$  werden keine Pixel mehr gefunden, was mit der Karte in Abbildung 5.8 korrespondiert, in der die größte negative Signifikanz mit  $-0.24\sigma$  angegeben ist.



**Abbildung 5.9:** Die Effizienzkurve bzgl. der vom Algorithmus gefundenen Pixel; die Effizienz bleibt bis  $0.18\sigma$  konstant 1, was 352 Pixel entspricht, und fällt ab dieser Schwelle bis  $0.24\sigma$  rapide ab

Im nächsten Histogramm (Abbildung 5.10) ist die Effizienz bezüglich der gegebenen Cluster aufgetragen. Insgesamt ist diese Effizienzkurve im Vergleich zur vorherigen Kurve sehr ähnlich, was dadurch zu erklären ist, dass die meisten Cluster aus nur einem Pixel bestehen (Abbildung 5.8), und so bei immer höheren Signifikanzschwellen Pixel nicht mehr vom Algorithmus berücksichtigt werden und damit keine 1-Pixel-Cluster mehr bilden können.



**Abbildung 5.10:** Die Effizienzkurve bzgl. der vom Algorithmus erstellten Cluster; die Verteilung ist der Kurve in Abbildung 5.9 sehr ähnlich, da die meisten Cluster aus nur einem Pixel bestehen (siehe Abbildung 5.8)

Bei beiden Effizienzbetrachtungen wurde darauf verzichtet, auch die Signifikanzschwelle der Nachbarpixel zu variieren. Eine Veränderung würde sich zwar auf die Größe, jedoch nicht auf die Anzahl der Cluster auswirken, da ein Cluster nur entstehen kann, sobald ein Pixel den Cut für zentrale Pixel überschreitet, unabhängig davon, wie viele nächste Nachbarn für den Cluster infrage kommen. Bei der Untersuchung, wieviele Pixel vom Programm erfasst werden, spielt die Anzahl der nächsten Nachbarn eine nur untergeordnete Rolle, da, wie bereits gezeigt wurde, die meisten Pixel in der Significance Map isoliert von den anderen signifikanten Pixel liegen und jeweils alleine einen Clustern bilden. Insgesamt wurde am Verlauf des Mondes und seines Einflusses als Störquelle gezeigt, dass der Algorithmus seine Aufgaben erfüllt und erwartungsgemäß funktioniert.

## Kapitel 6

# Methoden der Clustersuche im Pierre-Auger-Experiment

Die Anisotropieanalyse ist einer der Schwerpunkte des Pierre-Auger-Projektes. So wurden seit Beginn der Messungen auch eine Reihe von Untersuchungen veröffentlicht, die versucht haben, mithilfe von Clustern mögliche Punktquellen ausfindig zu machen. Dieser Abschnitt gibt einen kleinen Überblick, welche Methoden der Clustersuche bisher im Rahmen dieses Projektes angewendet wurden.

### 6.1 Cone-Jet-Algorithmus

In einem kürzlich erschienenen Forschungsbericht suchte Th. Hebbeker Cluster mithilfe eines Cone-Jet-Algorithmus [Heb07]. Allgemein werden benachbarte Einfallrichtungen in Cluster verpackt, deren Abstand

$$r = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (6.1)$$

mit dem Azimutwinkel  $\phi$  und der Pseudorapidität  $\eta$  kleiner dem Abstand  $r$  ist. Dieser Algorithmus wurde modifiziert, so dass

$$r^2 = \Delta\phi_{12}^2 \cdot \cos\theta_1 \cdot \cos\theta_2 + \Delta\theta_{12}^2 \quad (6.2)$$

mit dem Polarwinkel  $\theta$  und dem Azimutwinkel  $\phi$ . Sämtliche Richtungen, die  $r < 0.1$  voneinander entfernt lagen, was einem Raumwinkel von  $5.7^\circ$  entspricht, wurden zu einem Cluster zusammengefasst. Anhand der 28 Ereignisse mit den höchsten Energien oberhalb von 56EeV, die bis Juli 2007 von den Bodendetektoren für einen Zenitwinkel bis  $60^\circ$  gemessen wurden, konnte der Algorithmus insgesamt vier Cluster finden, einer bestehend aus 3 und drei aus 2 Schauern.

### 6.2 Maximum-Likelihood-Ratio-Test

S. BenZvi und S. Westerhoff verwendeten auf der Suche nach Punktquellen die Maximum-Likelihood-Ratio-Test [Ben05]. Aus einer gegebenen Himmelsrichtung  $x$  erreicht ein Schauer als Ergebnis der kosmischen Hintergrundstrahlung die Erde mit der Wahrscheinlichkeit  $R_i$ . Fällt ein Schauer, der von einer Quelle erzeugt wurde, in der Richtung  $x_i$  ein, so ist die Wahrscheinlichkeit, dieses Ereignis in

einer anderen Position  $x$  finden,  $Q_i(x, x_s)$ . Für die Wahrscheinlichkeitsverteilung eines Ereignisses  $i$  gilt:

$$P_i(x, x_s) = \frac{n_s}{N} Q_i(x, x_s) + \frac{N - n_s}{N} R_i(x, x_s) \quad (6.3)$$

$n_s$  ist die Anzahl der Schauer, die durch Punktquellen erzeugt wurden,  $N$  die Gesamtanzahl der Schauer. Die Likelihood-Funktion ergibt sich aus dem Produkt aller möglicher Wahrscheinlichkeiten  $P_i$ :

$$L(n_s, x_s) = \prod P_i(x_i, x_s) \quad (6.4)$$

Die Ratio bildet sich nun aus zwei Voraussetzungen: Entweder findet man für eine bestimmte Himmelsrichtung sowohl Hintergrund-, als auch Quelleneignisse, oder es sind keine Schauer zu erkennen, die von Quellenteilchen erzeugt werden ( $n_s = 0$ ). Daraus erhält man schließlich:

$$\ln R(n_s, x_s) = \ln \frac{L(n_s, x_s)}{L(0, x_s)} = \sum \ln \left[ \frac{n_s}{N} \left( \frac{Q_i(x, x_s)}{R_i(x_i)} - 1 \right) + 1 \right] \quad (6.5)$$

Wie in dieser Arbeit erfolgt die Suche mittels einer Himmelskarte; für jeden Pixel werden alle möglichen  $R_i$  und  $Q_i$  berechnet, wobei sich  $Q_i$  aus der Winkelungenauigkeit für jedes Ereignis  $i$  und  $R_i$  aus erwarteten Messungen bei einer isotropischen Schauerverteilung ergibt. Somit ergibt sich  $R_i$  aus einer mit den Parametern der  $N$  Ereignissen erstellten Coverage Map. Nach der Berechnung aller Koeffizienten kann  $\ln R$  für jeden Punkt in Abhängigkeit von  $n_s$  maximiert werden. Erstellt man nun mit  $\ln R$  eine Himmelskarte, so erhält man lokale Maxima, in denen sich  $n_s$  Schauer befinden. Diese Richtungen werden zu Clustern verpackt.

Verwendet wurden die 639 Ereignisse oberhalb 10EeV, die im Zeitraum von Januar bis Dezember 2004 von den Bodendetektoren bei einem Zenitwinkel von maximal  $75^\circ$  gemessen wurden. Es haben sich einige signifikante lokale Maxima gebildet, in denen sich meist Doublets befinden. Das globale Maximum befindet sich bei den Äquatorkoordinaten mit einem Cluster von drei Ereignissen.

### 6.3 Die 2-Punkt Korrelationsfunktion

Eine verbreitete Methode ist die Clustersuche mittels einer 2-Punkt Korrelationsfunktion. Die Funktion zählt Paare von Ereignissen oberhalb einer Energieschwelle  $E$ , deren Schauerrichtungen weniger als ein definierter Winkel  $\Theta$  auseinanderliegen. Diese Zahl wird in Zusammenhang gesetzt zu der Anzahl an Paaren bei einer isotropen Verteilung, die durch Monte-Carlo Simulationen generiert wird. Anhand eines Autokorrelations-Scans werden die Energie und der Winkel  $\Theta$  bestimmt, bei der das Clustering-Signal am Stärksten ist. Dazu wird die bei einem Datensatz von  $N$  Ereignissen die Anzahl an Paaren  $n_p$  für jedes  $E$  und  $\Theta$  berechnet und die Wahrscheinlichkeit bestimmt, dass  $n_p$  oder mehr Paare gefunden werden:

$$P_{\text{data}}(N, \Theta) = \sum_{n=n_p}^{\infty} P_{\text{mc}}(N, \Theta, n) \quad (6.6)$$

$P_{\text{mc}}$  ist dabei die Wahrscheinlichkeit für die simulierten Daten. Für eine bestimmte Kombination der Energie und des Winkels wird  $P_{\text{data}}$  minimal, was das gesuchte stärkste Signal anzeigt. Um die wahre Signifikanz des Signals zu berechnen, wird der gleiche Scan über  $n_{\text{mc}}$  Monte-Carlo Datensätze

durchgeführt und so für jede Berechnung  $P_{\min}^i$  bestimmt. Die Zufallswahrscheinlichkeit ergibt sich aus:

$$P = \frac{n_{mc}^*}{n_{mc}} \quad (6.7)$$

wobei  $n_{mc}^*$  die Anzahl der Datensätze ist, bei der  $P_{\min}^i \leq P_{\min}$ .

Diese Variante der Clustersuche wurde von S.Mollerach durchgeführt, bei der sie mit Pierre-Auger-Daten von Januar 2004 bis zum 15. März 2007 oberhalb von 10EeV (1672 Ereignisse) das Signal überprüfte, welches bei  $2.5^\circ$  für Energien oberhalb 40EeV von AGASA gemeldet wurde [Mol07]. Auch D. Semikoz bediente sich der Autokorrelation-Funktionen und unterschied zwischen klein- ( $\theta \leq 5^\circ$ ) oder großräumiger Anisotropie ( $20^\circ \leq \theta \leq 30^\circ$ ) [Sem07a].

Auf ähnliche Weise werden nicht nur Korrelationen zwischen einfallenden Schauern, sondern auch zwischen Schauerrichtungen und möglichen Quellen, vor allem aktive galaktische Kerne, hergestellt. Beispiele dafür sind die Ergebnisse von D.Semikoz [Sem07b] oder B.Siffert [Sif07].

## Kapitel 7

# Zusammenfassung und Ausblick

Das Ziel dieser Arbeit war es, anhand der Entwicklung von einfachen Clusteralgorithmen einen Einblick in die Anisotropieanalysen des Pierre-Auger-Projektes zu gewinnen. So wurden in Kapitel 5 zwei Methoden der Clustersuche beschrieben, die im Rahmen dieser Arbeit entwickelt und auf ihre Fähigkeiten und Ergebnisse hin getestet wurden:

Methode 1 sucht in einer gegebenen Significance Map (Kapitel 3 und 4) nach den Pixeln mit den höchsten positiven bzw. negativen Signifikanzen (die Mindestsignifikanz ist variabel einstellbar, bei der Entwicklung wurde diese Grenze bei  $3\sigma$  bzw.  $-3\sigma$  festgesetzt) und bildet aus jeweils einem dieser Pixel sowie seinen nächsten Nachbarn Cluster. Da die Bedingung festgelegt wurde, dass jeder Pixel in dieser Karte zur höchsten einem Cluster gehören darf, bestehen alle Cluster aus mindestens einem und maximal 9 Pixeln.

Der zweite Algorithmus baut auf der ersten Methode auf, sucht also zunächst nach den Pixeln mit den höchsten positiven bzw. negativen Signifikanzen. Um die Anzahl der Cluster zu reduzieren, wurde die Signifikanzschwelle auf  $4\sigma$  bzw.  $-4\sigma$  erhöht. Es wurden zunächst die nächsten Nachbarn gesucht, wobei diesmal die Einschränkung eingeführt wurde, dass die Nachbarn eine Signifikanz von mindestens  $1\sigma$  bzw.  $-1\sigma$  aufweisen müssen. Auch für diese Pixel wurden die nächsten Nachbarn ermittelt und die Suche für alle neuen Pixel in Cluster solange fortgesetzt, bis keine neuen Punkte mit der oben genannten Mindestsignifikanz gefunden wurden. Damit ist die Clustergröße variabel, wobei durch die Eigenschaft der Signifikanzverteilung (Kapitel 3.3) und durch die Anzahl der Pixel in der Karte ein natürliche Grenze festgesetzt ist.

Es gibt zahlreiche weitere Möglichkeiten der Clustersuche, wie bereits in Kapitel 7 zu sehen ist, sei es durch die Entwicklung neuer Algorithmen oder der Modifizierung und Anwendung bereits bestehender Verfahren, wie den Jet-Algorithmus.

Die Methoden dieser Arbeit basieren darauf, dass sich die Cluster um jeweils einen Pixel mit besonders hoher Signifikanz bilden. Im Gegensatz dazu könnte z.B. nicht nach einzelnen hochsignifikanten Pixeln gesucht werden, sondern anhand eines  $3 \times 3$ -Rasters nach den 9 direktbenachbarten Pixeln (angeordnet: ein zentraler Pixel und seine 8 nächsten Nachbarn) mit der höchsten Gesamtsignifikanz, um so aus diesen einen Cluster zu bilden.

Somit gibt es eine ganze Reihe von Möglichkeiten, diese Arbeit vorzuführen.

# Literaturverzeichnis

- [Alk75] ALKHOFER, O. C.: *Introduction to Cosmic Radiation*. Verlag Carl Thiemig, 1975.
- [Ast] ASTROTEILCHENPHYSIK IN DEUTSCHLAND. <http://www.astroteilchenphysik.de/topics/cr/grnd.htm>, Abruf: 14.10.2007.
- [Bec05] BECKER, J. R.: *Untersuchung der optischen Eigenschaften der Fluoreszenzteleskope des Pierre Auger Observatoriums*. Diplomarbeit, Universität Karlsruhe, 2005.
- [Ben05] BENZVI, S. UND WESTERHOFF, S.: *Search for Point Sources in Auger SD Data Using a Maximum Likelihood Method*. GAP-Note 2005-058, 2005.
- [Blü00] BLÜMER, J. UND KAMPERT, K.-H.: *Die Suche nach den Quellen der kosmischen Strahlung*. Physikalische Blätter 56, Seiten 39–45, 2000.
- [Chi] CHINI, R.: *Einführung in die Astronomie – Skript*. <http://www.astro.ruhr-uni-bochum.de/astro/topics/vorlesungen/chini/raumwinkel.pdf>, Abruf: 14.10.2007.
- [Fer49] FERMI, E.: *On the Origin of the Cosmic Radiation*. Physical Review 75, Seiten 1169–1174, 1949.
- [Gor05] GORSKI, K. M. ET AL: *HEALPix – a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere*. The Astrophysical Journal 622, Seite 759, 2005.
- [Gre66] GREISEN, K.: *End to the Cosmic-Ray Spectrum?* Physical Review Letters 16, Seiten 748–750, 1966.
- [Gri07] GRIGAT, M.: *Sky Coverage Computation for the Surface Detector of the Pierre Auger Observatory*. Diplomarbeit, Universität Siegen, 2007.
- [Gru00] GRUPEN, C.: *Astroteilchenphysik – Das Universum im Licht der kosmischen Strahlung*. Friedr. Vieweg & Sohn Verlagsgesellschaft mbH. Braunschweig/Wiesbaden, 2000.
- [Heb07] HEBBEKER, TH.: *Search for Clustering of Shower Directions with a Cone Jet Algorithm*. GAP-Note 2007-084, 2007.
- [KK97] KLAPDOR-KLEINGROTHAUS, H. V. UND ZUBER, K.: *Teilchenastrophysik*. B. G. Teubner, 1997.

- [Mol07] MOLLERACH, S.: *Studies of clustering in the arrival directions of cosmic rays detected at the Pierre Auger Observatory above 10 EeV*. ICRC 2007, 2007.
- [Piea] THE PIERRE AUGER COLLABORATION. <http://www.auger.org/news/>, Abruf: 14.10.2007.
- [Pieb] THE PIERRE AUGER COLLABORATION. <http://www.auger.org.ar/survey/Ozi/sd-array-full-1673-SD-20041116.jpg/>, Abruf: 14.10.2007.
- [Pie97] THE PIERRE AUGER COLLABORATION: *The Pierre Auger Observatory Design Report, Second Edition*, 1997.
- [Sem07a] SEMIKOZ, D.: *Small and medium scale clusters in the arrival directions of cosmic rays in the Pierre Auger Observatory data*. GAP-Note 2007-010, 2007.
- [Sem07b] SEMIKOZ, D.: *Update on correlations with AGNs*. GAP-Note 2007-086, 2007.
- [Sif07] SIFFERT, B. ET AL: *Could other UHECR sources result in the observed correlation with AGN*. GAP-Note 2007-074, 2007.
- [Zat66] ZATSEPIN, G. T. UND KUZMIN, V. A.: *Upper Limit of the Spectrum of Cosmic Rays*. Journal of Experimental and Theoretical Physics 4, Seiten 78–80, 1966.

**Erklärung**

Hiermit erkläre ich, dass ich die vorliegende Bachelor-Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate und Ergebnisse Anderer kenntlich gemacht habe.

(Ort)

(Datum)

(Unterschrift)