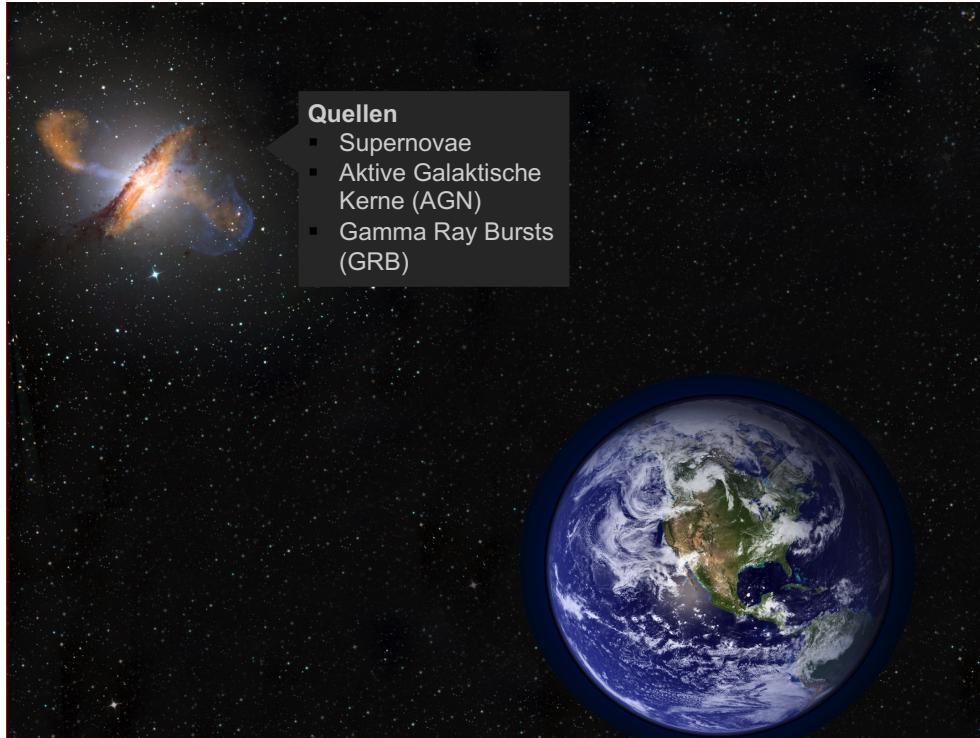


Vorlesung

Statistische Methoden der Datenanalyse

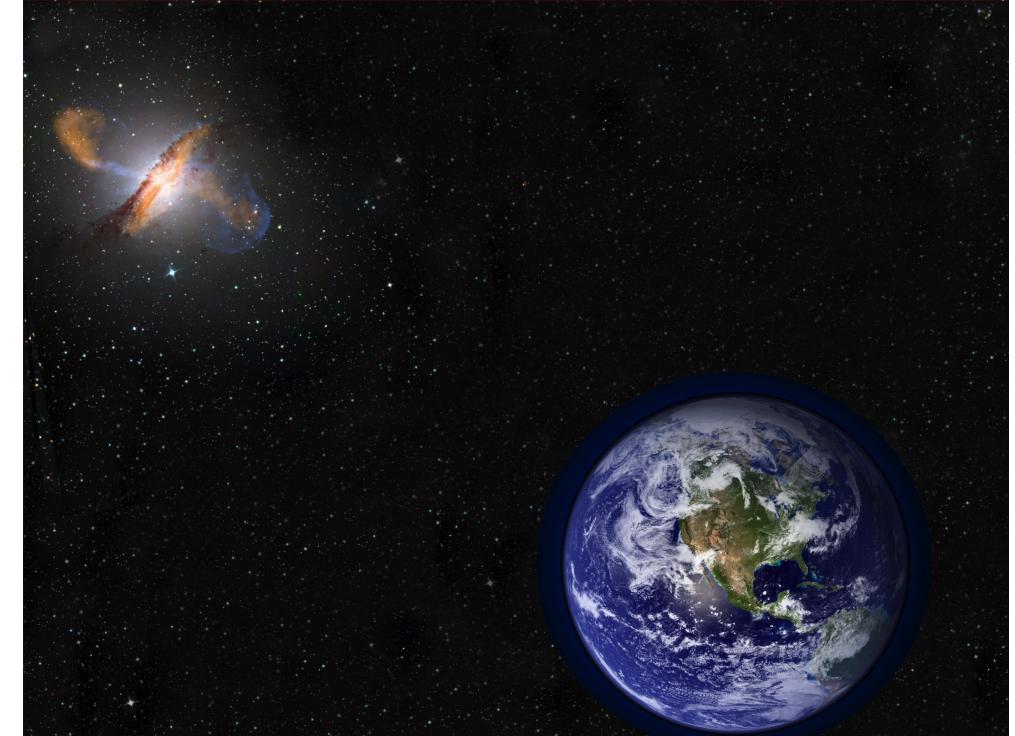
Prof. Dr. Dr. Wolfgang Rhode

Datenanalyse – Ein Beispiel aus der Praxis

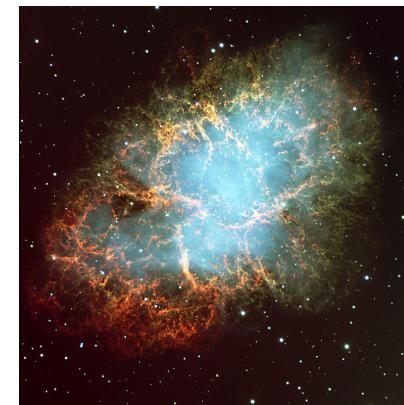


Quellen

- Supernovae
- Aktive Galaktische Kerne (AGN)
- Gamma Ray Bursts (GRB)

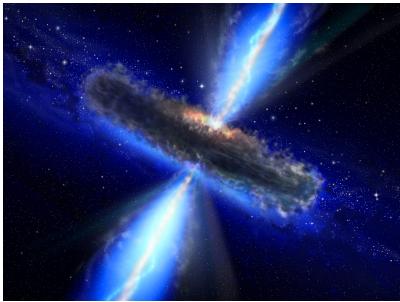


Supernovae



- Supernova
 - Ende Lebenszyklus eines Sternes; Fusionsmaterial *aufgebraucht*
 - Explosion mit hoher Emission von Photonen, geladenen Teilchen und Neutrinos
- Supernovaüberrest (SNR)
 - Expandierende Schockfront
 - Emittieren großen Teil der galaktischen kosmischen Strahlung
 - Beschleunigung durch den sogenannten Fermi Mechanismus

Aktive Galaktische Kerne (AGN)

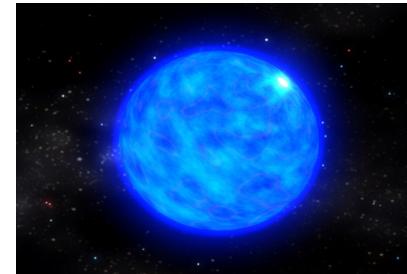


- Bestehen aus:
 - Supermassivem schwarzen Loch im Zentrum
 - Akkretionsscheibe um das Zentrum
 - Relativistischen Jets aus dem Zentrum, senkrecht zur Akkretionsscheibe
- Kategorisierung abhängig unserer Blickrichtung auf die Jets
- Emittieren hochenergetische kosmische Strahlung

Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse

Gamma Ray Bursts (GRB)

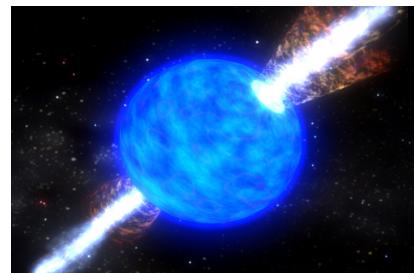


- Extrem hochenergetische Explosionen in weit entfernten Galaxien
- Hohe Strahlungsemision während der Explosion
- Dauer in der Größenordnung von Sekunden

Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse

Gamma Ray Bursts (GRB)

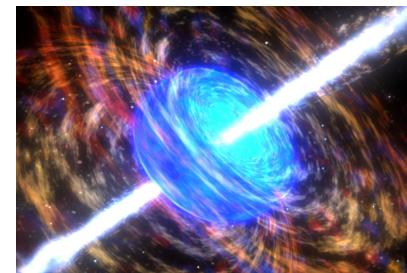


- Extrem hochenergetische Explosionen in weit entfernten Galaxien
- Hohe Strahlungsemision während der Explosion
- Dauer in der Größenordnung von Sekunden

Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse

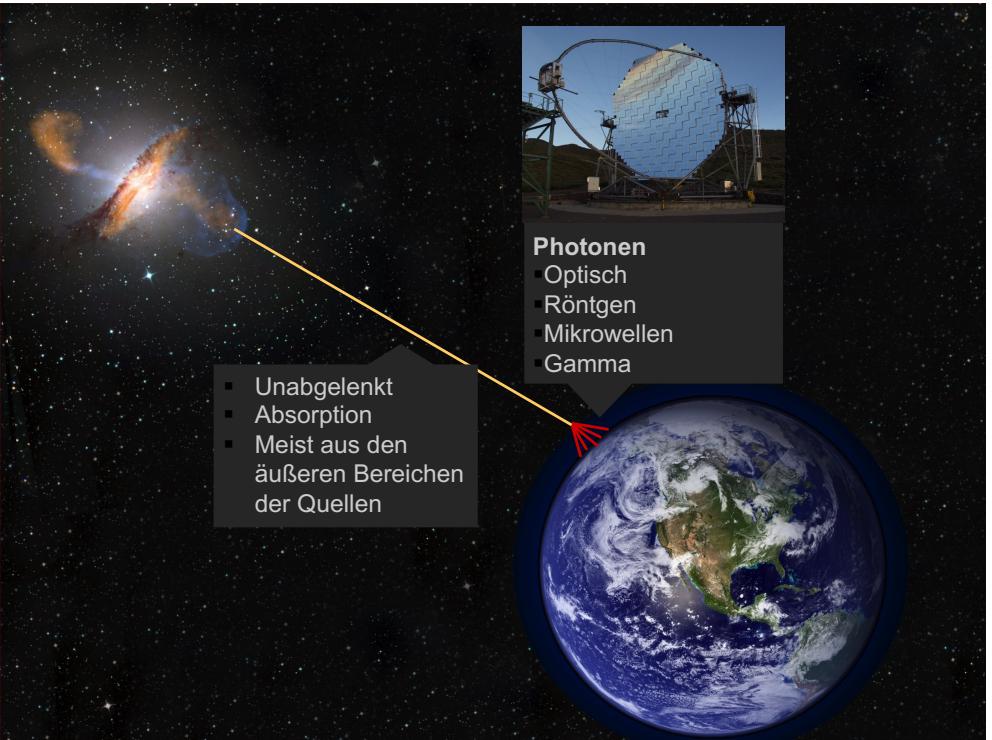
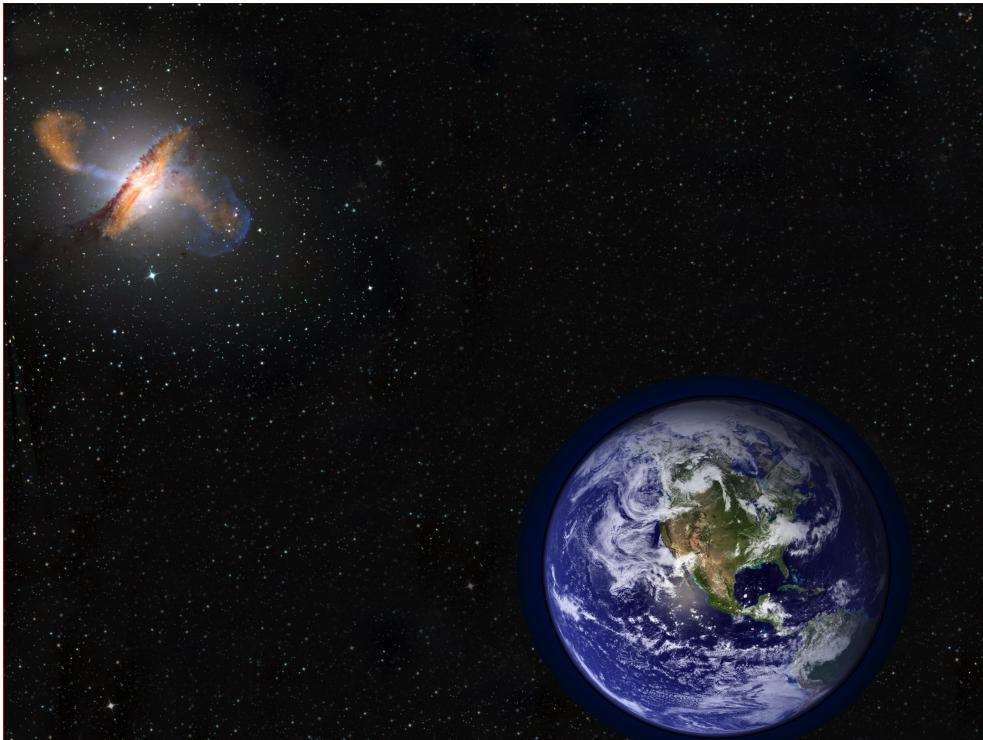
Gamma Ray Bursts (GRB)



- Extrem hochenergetische Explosionen in weit entfernten Galaxien
- Hohe Strahlungsemision während der Explosion
- Dauer in der Größenordnung von Sekunden

Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse



Photonen



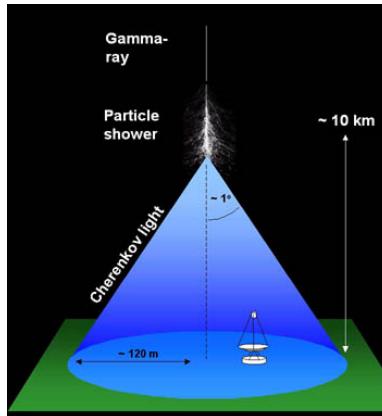
- Direkter Nachweis/Messung der Photonen
 - Sichtbares Licht
 - Radiowellen
 - Röntgen (mit Satelliten)



Tscherenkov Strahlung:

Geladene Teilchen, die sich durch ein Medium bewegen und dabei schneller sind als die Lichtgeschwindigkeit in dem Medium, emittieren Photonen.

Photonen



Prof. Dr. Dr. W. Rhode



- Indirekter Nachweis/Messung hochenergetischer Gamma Strahlung
- Messung der Tscherenkov Photonen, die vom elektromagnetischen Teilchenschauer emittiert werden

Statistische Methoden
der Datenanalyse

Photonen

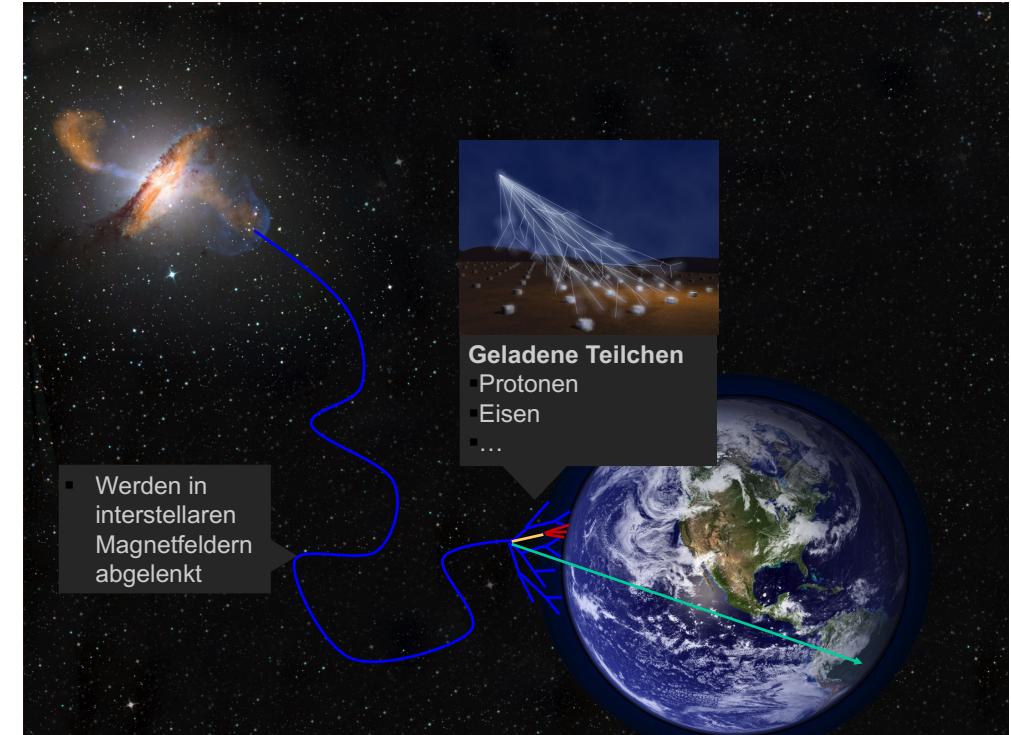
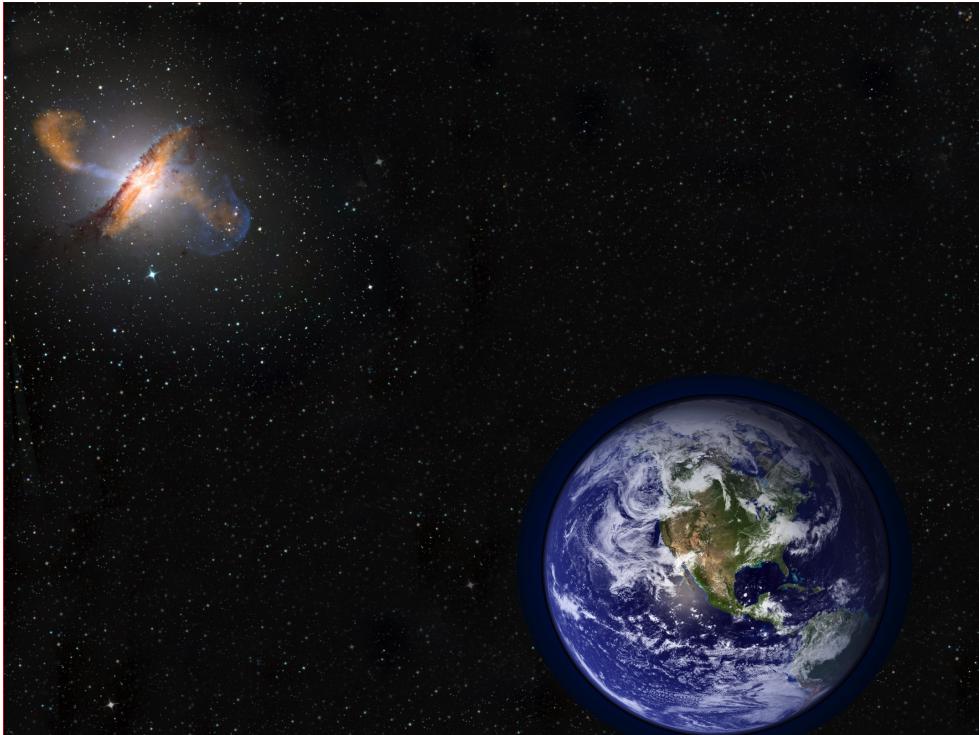


- Direkter Nachweis/Messung der Photonen
 - Sichtbares Licht
 - Radiowellen
 - Röntgen (mit Satelliten)

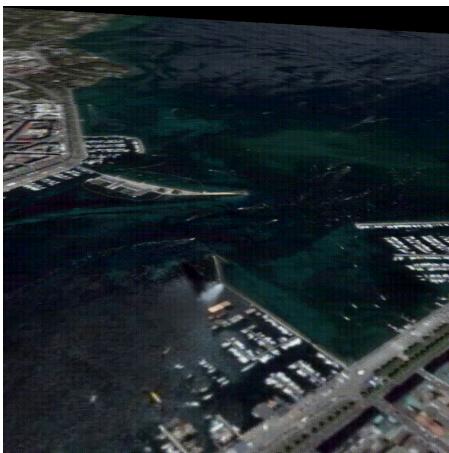
Prof. Dr. Dr. W. Rhode

- Indirekter Nachweis/Messung hochenergetischer Gamma Strahlung
- Messung der Tscherenkov Photonen, die vom elektromagnetischen Teilchenschauer emittiert werden

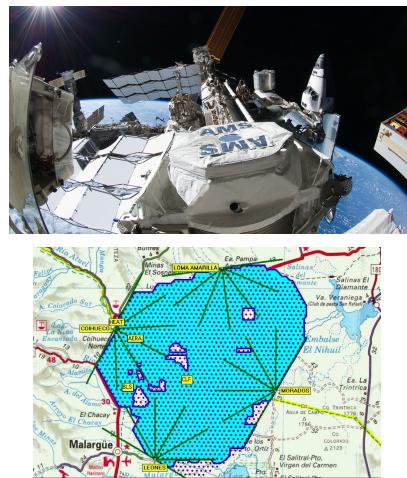
Statistische Methoden
der Datenanalyse



Geladene Strahlung

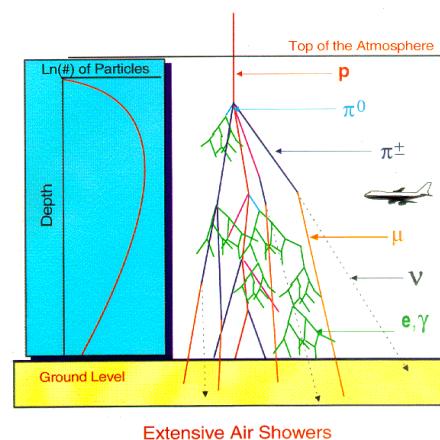


Prof. Dr. Dr. W. Rhode



Statistische Methoden
der Datenanalyse

Geladene Strahlung

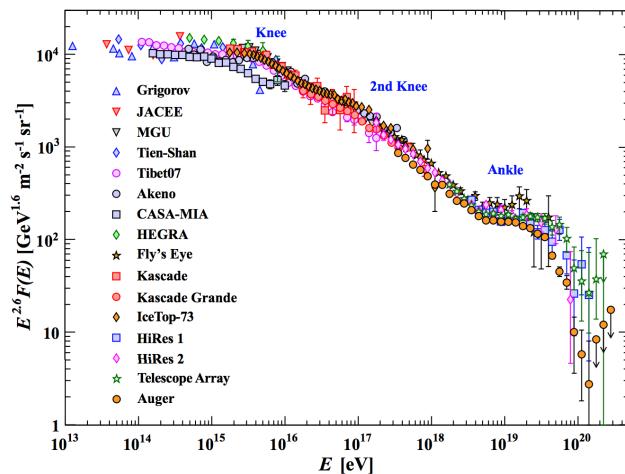


Prof. Dr. Dr. W. Rhode

- Trifft ein hochenergetisches Teilchen der geladenen kosmischen Strahlung auf der Atmosphäre entsteht ein Luftschauder
 - Teilchen, die in diesen Schauern entstehen, können direkt gemessen werden
 - Indirekte Messung:
 - Tscherenkov Strahlung
 - Fluoreszenzlicht von in der Luft angeregten Stickstoffmolekülen

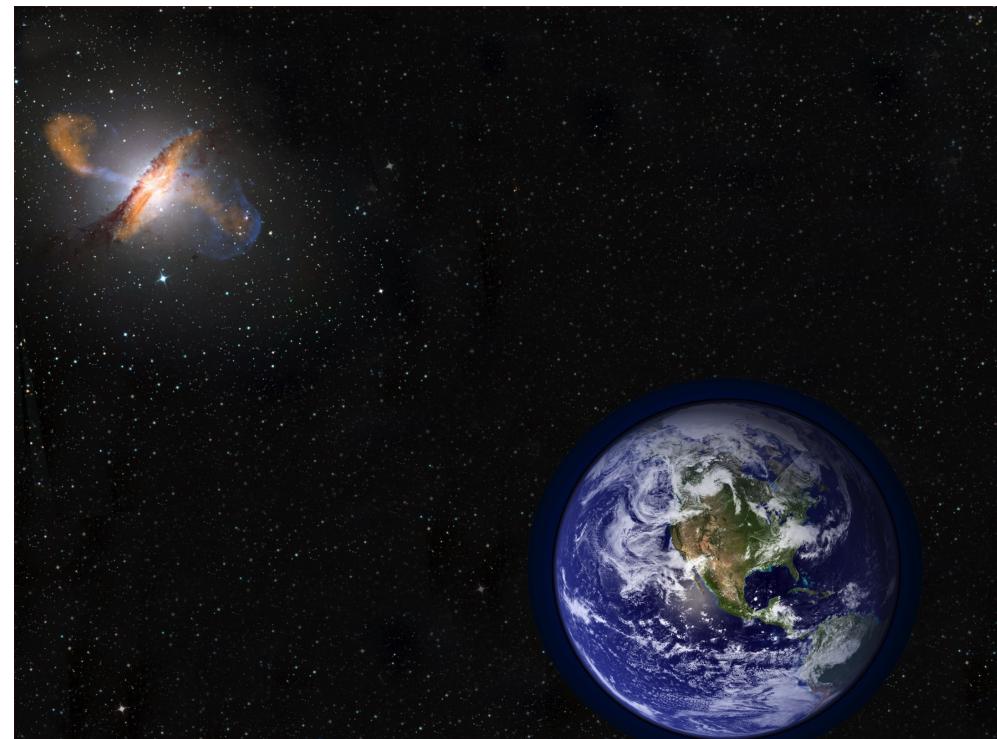
Statistische Methoden
der Datenanalyse

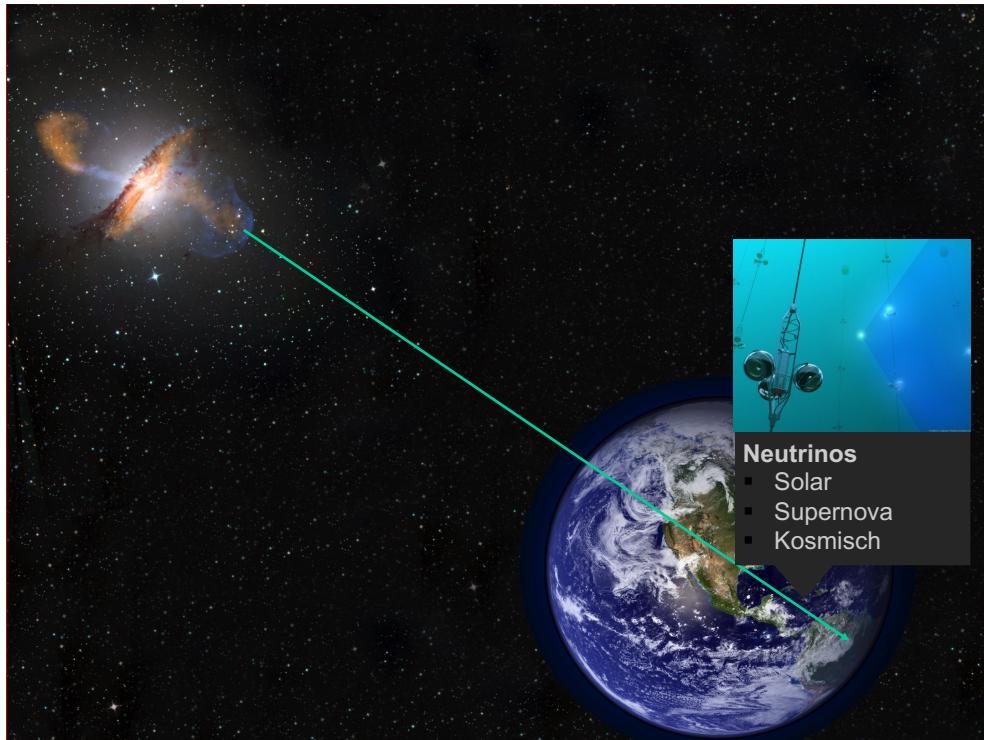
Geladene Strahlung



Prof. Dr. Dr. W. Rhode

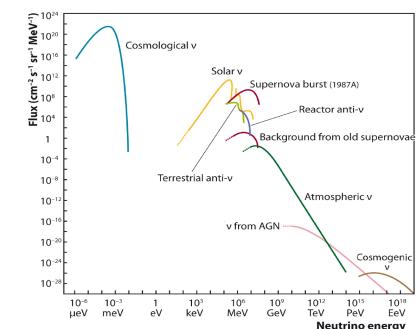
Statistische Methoden
der Datenanalyse





Neutrinos

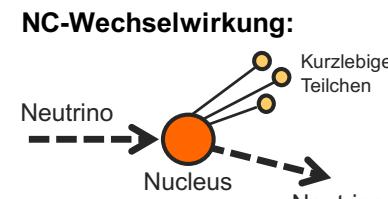
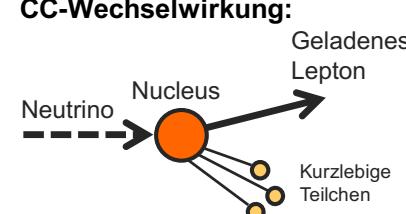
mass →	$>2.3 \text{ GeV}/c^2$	$=1.275 \text{ GeV}/c^2$	$=173.07 \text{ GeV}/c^2$	0	$=16.26 \text{ GeV}/c^2$
charge →	$+2/3$	$+1/3$	$-1/3$	0	0
spin →	$1/2$	$1/2$	$1/2$	1	0
	up	charm	top	gluon	Higgs boson
QUARKS					
	$\sim 4.8 \text{ MeV}/c^2$	$\sim 65 \text{ MeV}/c^2$	$\sim 14.18 \text{ GeV}/c^2$	0	
	$-1/3$	$-1/3$	$-1/3$	0	
	down	strange	bottom	photon	
LEPTONS					
	$0.811 \text{ MeV}/c^2$	$108.7 \text{ MeV}/c^2$	$1.777 \text{ GeV}/c^2$	9.12 GeV/c^2	
	-1	-1	-1	0	
	$1/2$	$1/2$	$1/2$	1	
	electron	muon	tau	Z boson	
GAUGE BOSONS					
	$<2.2 \text{ eV}/c^2$	$<0.17 \text{ MeV}/c^2$	$<15.5 \text{ MeV}/c^2$	$<6.0 \text{ GeV}/c^2$	
	$1/2$	0	0	± 1	
	electron neutrino	muon neutrino	tau neutrino	W boson	



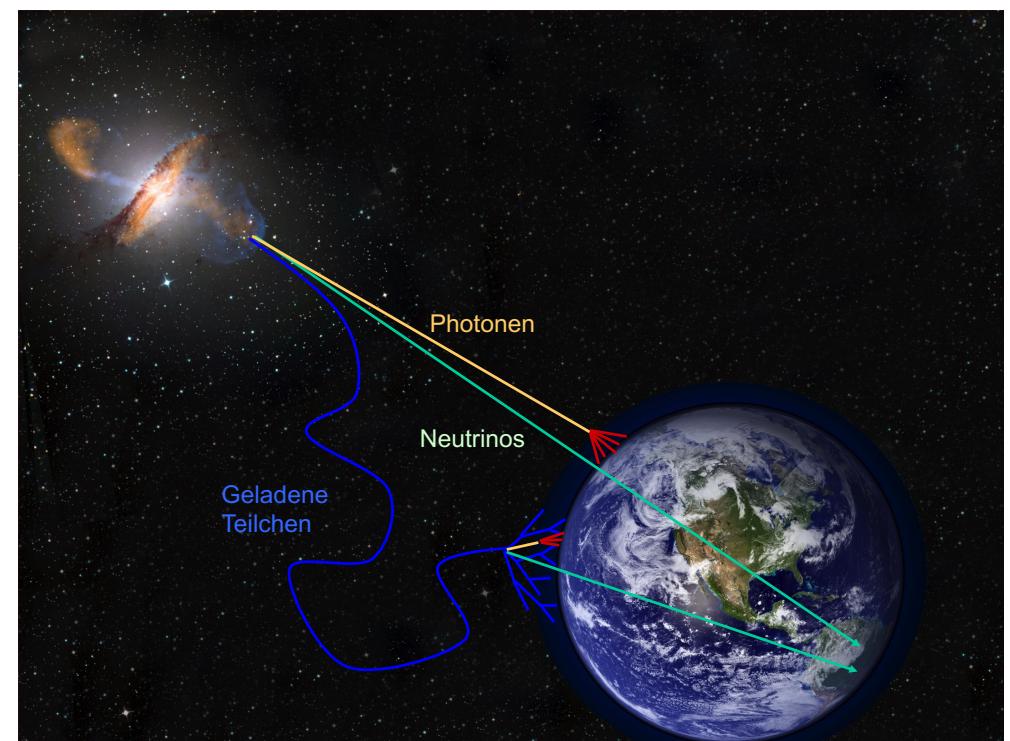
Statistische Methoden
der Datenanalyse

Neutrinos

- Da Neutrinos nur Teil der schwachen Wechselwirkung sind, ist eine direkte Messung nicht möglich
 - Messung der Tscherenkov Photonen, die von Teilchen, die nach der Neutrinowechselwirkung entstehen, emittiert werden
 - Quellen für hochenergetische kosmische Neutrinos ($> 100 \text{ TeV}$) unbekannt
 - Experimenteller Nachweis ihrer Existenz erst Ende 2013 durch IceCube

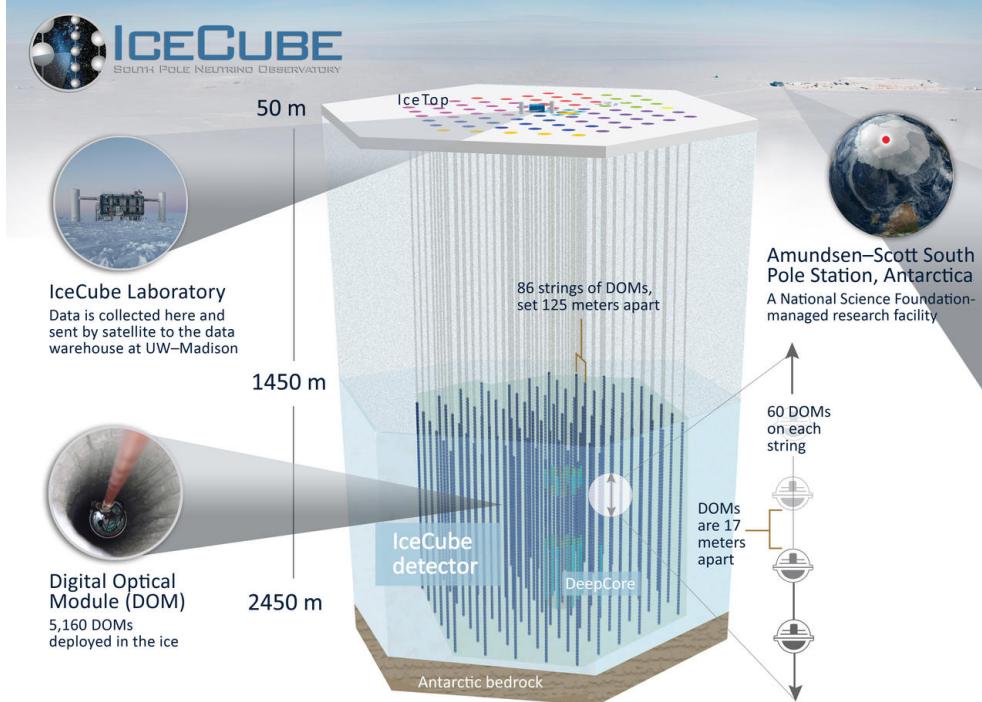
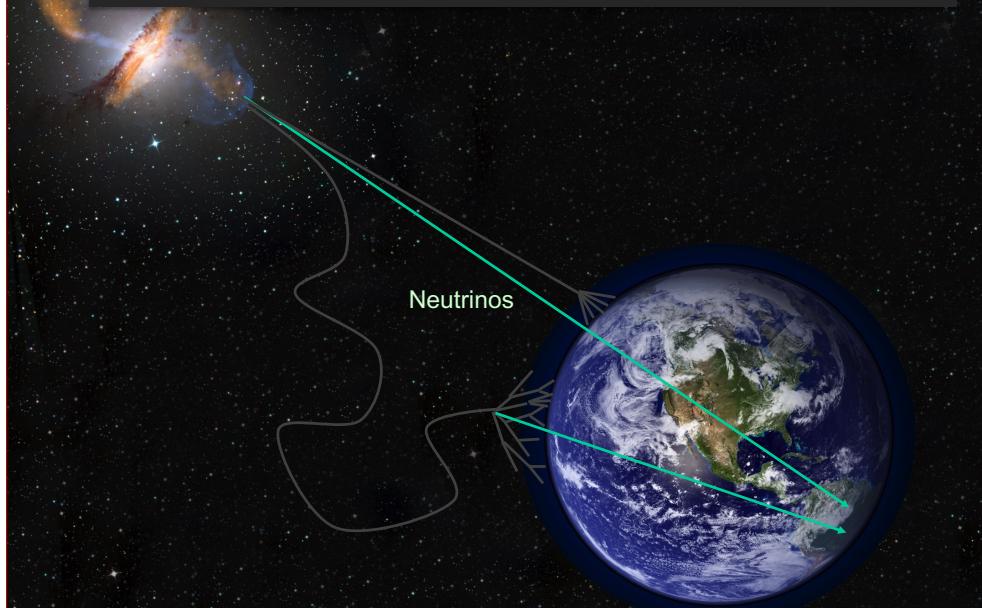


Statistische Methoden
der Datenanalyse



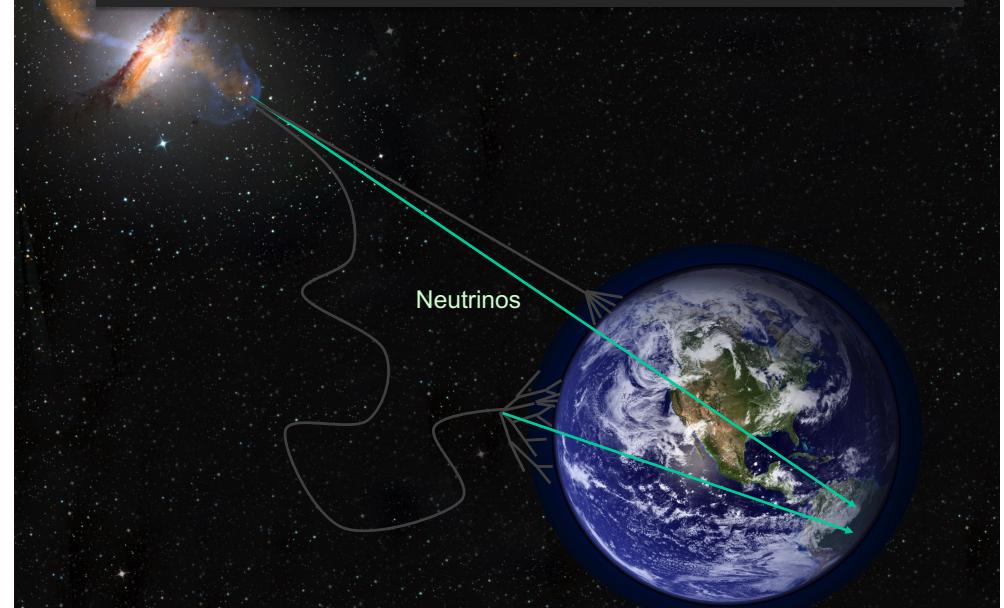
Datenanalyse am Beispiel:

Messung des Energiespektrums von Myon-Neutrinos in IceCube



Datenanalyse am Beispiel:

Messung des Energiespektrums von Myon-Neutrinos in IceCube

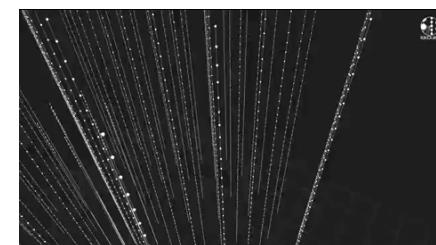


tu technische universität
dortmund

Experimentelle Physik Vb
Astroteilchenphysik

IceCube – Event Signature

Kaskaden

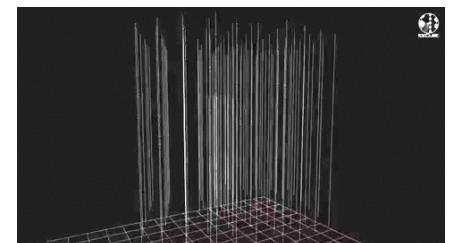


CC-Wechselwirkung
 e/τ -Neutrino



NC-Wechselwirkung
 $e/\mu/\tau$ -Neutrino

Spurartige Ereignisse (Tracks)



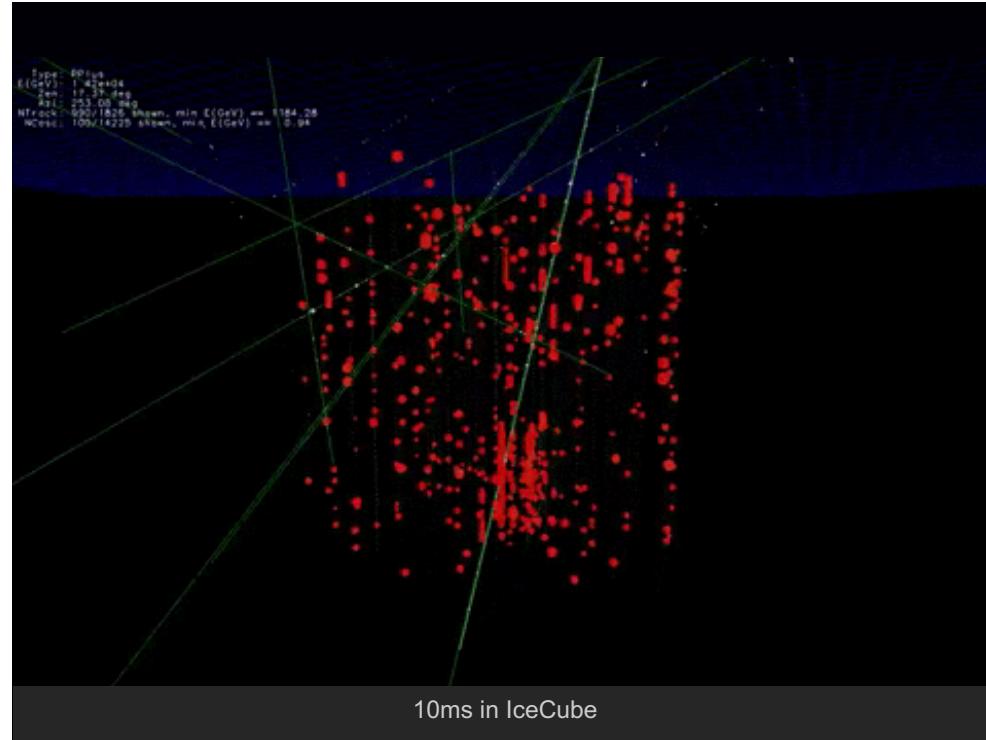
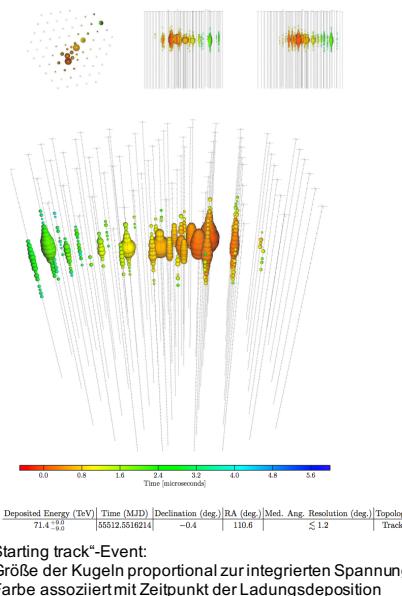
CC-Wechselwirkung
 μ -Neutrino

atmosphärisches μ
Statistische Methoden
der Datenanalyse

Prof. Dr. Dr. W. Rhode

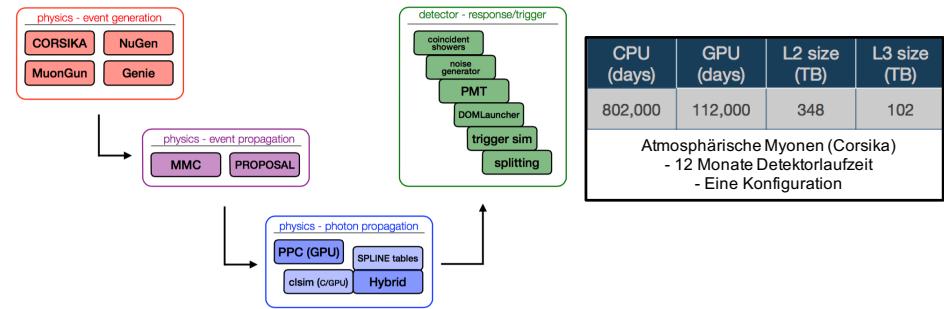
IceCube - Rekonstruktionen

- Die tatsächliche Messung eines Ereignisses besteht aus über fünftausend 190ns langen Spannungszeitreihen (für jeden DOM eine)
- Rekonstruktionen von physikalischen Observablen nötig



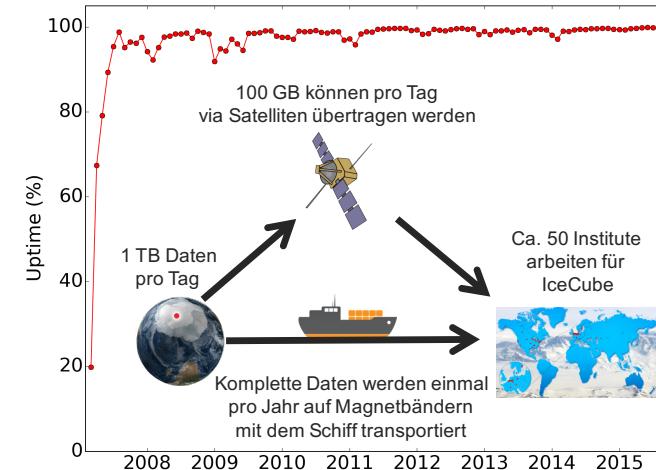
IceCube - Simulationen

- Kalibration mit *Test-Neutrinos* unmöglich und viele moderne Analyseverfahren, die auf Simulationen basieren, im Einsatz
- Extrem aufwändige und detaillierte Simulationen startend bei der Teilchenproduktion bis zum Ende der Detektion nötig



Prof. Dr. Dr. W. Rhode

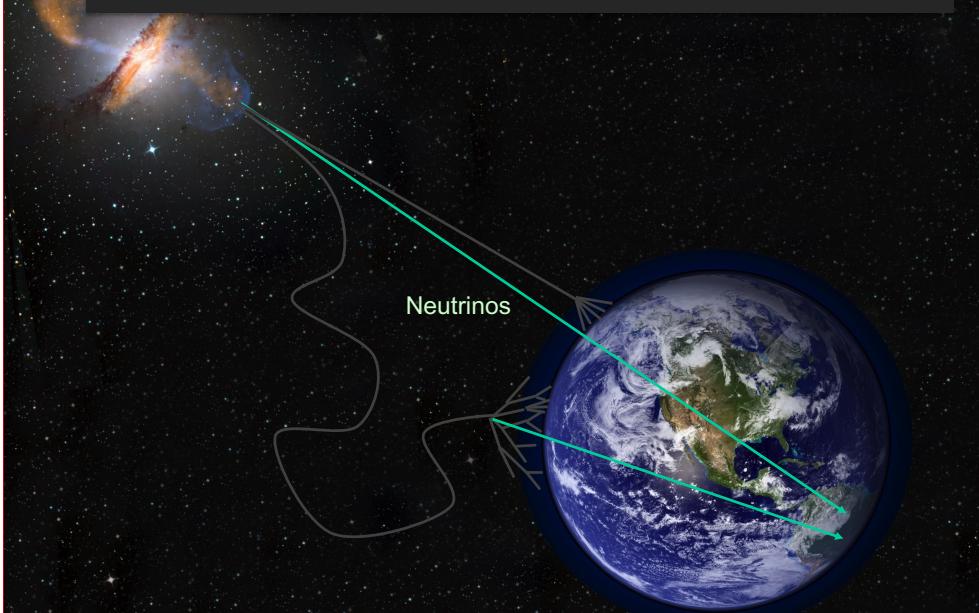
IceCube – Big Data



Prof. Dr. Dr. W. Rhode

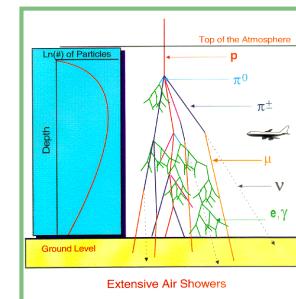
Datenanalyse am Beispiel:

Messung des Energiespektrums von Myon-Neutrinos in IceCube

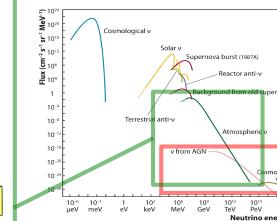


Myon-Neutrinos in IceCube

- Für Neutrino Energien > 100GeV muss zwischen zwei verschiedenen Ursprüngen der Neutrinos unterschieden werden
 - Atmosphärischen Ursprungs: konventionell und prompt
 - Kosmologischen Ursprungs: Genaue Quellen bisher unklar



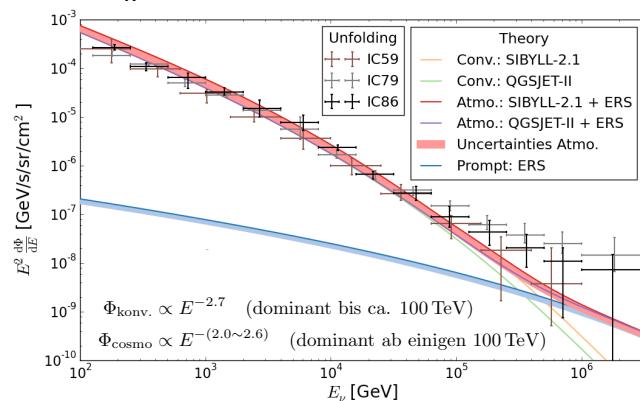
Prof. Dr. Dr. W. Rhode



Statistische Methoden
der Datenanalyse

Messung des Energiespektrums

- Gesucht wird das Energiespektrum aller Myon-Neutrinos in IceCube ab einer Neutrinoenergie von > 100 GeV



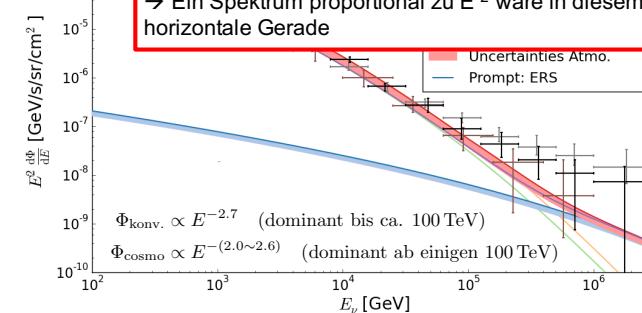
Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse

Das Energiespektrum wird als energieabhängiger Fluss angegeben:

$$\frac{\text{Teilchenzahl}}{\text{GeV s sr cm}^2} : \frac{\text{Teilchenzahl pro Zeiteinheit pro Fläche am Himmel pro Detektionsfläche differential in der Energie}}$$

Zu Darstellungszwecken wird dieser Fluss häufig mit E^2 gewichtet
→ Ein Spektrum proportional zu E^2 wäre in diesem Plot eine horizontale Gerade

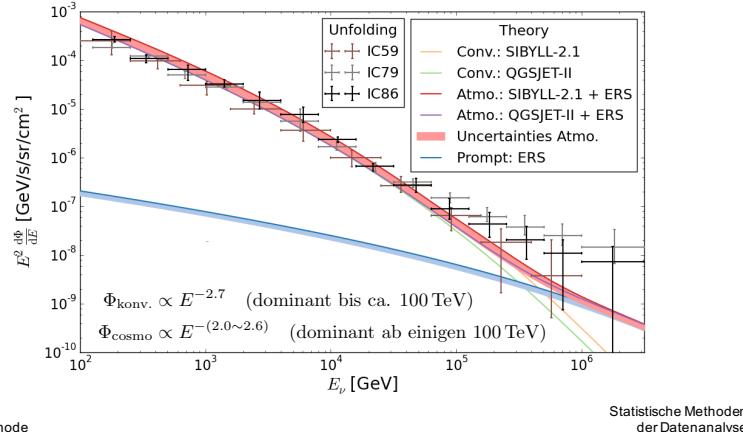


Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse

Messung des Energiespektrums

- Gesucht wird das Energiespektrum aller Myon-Neutrinos ins IceCube ab einer Neutrinoenergie von > 100 GeV



Prof. Dr. Dr. W. Rhode

Messung des Energiespektrums

- Gesucht wird das Energiespektrum aller Myon-Neutrinos ins IceCube ab einer Neutrinoenergie von > 100 GeV



Herausforderungen:

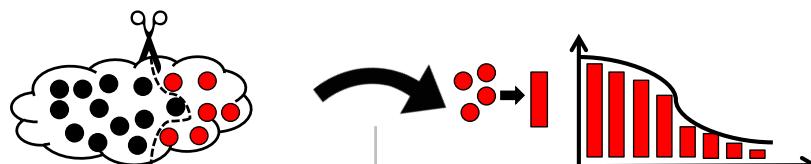
- Der indirekte Nachweis der Myon-Neutrinos durch Myonen (CC-Wechselwirkung) im Detektor
 → Dem Myon im Detektor ist der Ursprung nicht anzusehen. Somit ist eine Unterscheidung zwischen neutrino-induzierten Myonen und atmosphärischen Myonen nicht trivial
- Es kann nur die Energie des Myons im Detektor gemessen werden
 → Physik und Detektoreffekte verschmieren die Energie und machen eine direkte Umrechnung von Myon- zu Neutrinoenergie nicht möglich

Statistische Methoden
der Datenanalyse

Prof. Dr. Dr. W. Rhode

Messung des Energiespektrums

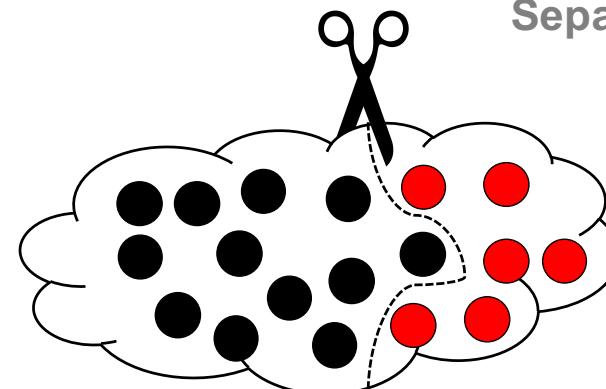
- Die Analyse wird in zwei Schritte unterteilt, in denen jeweils eine der beiden Herausforderungen angegangen wird



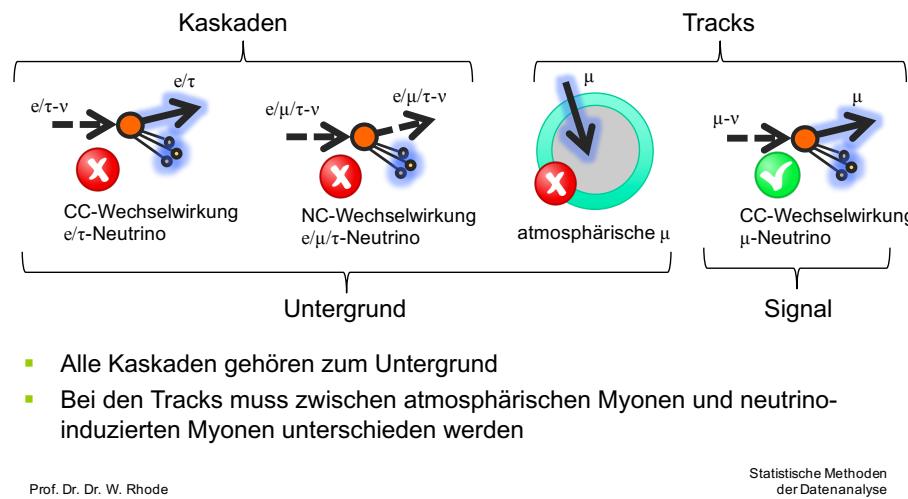
Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse

Schritt 1: Separation



Trennung von Signal und Untergrund

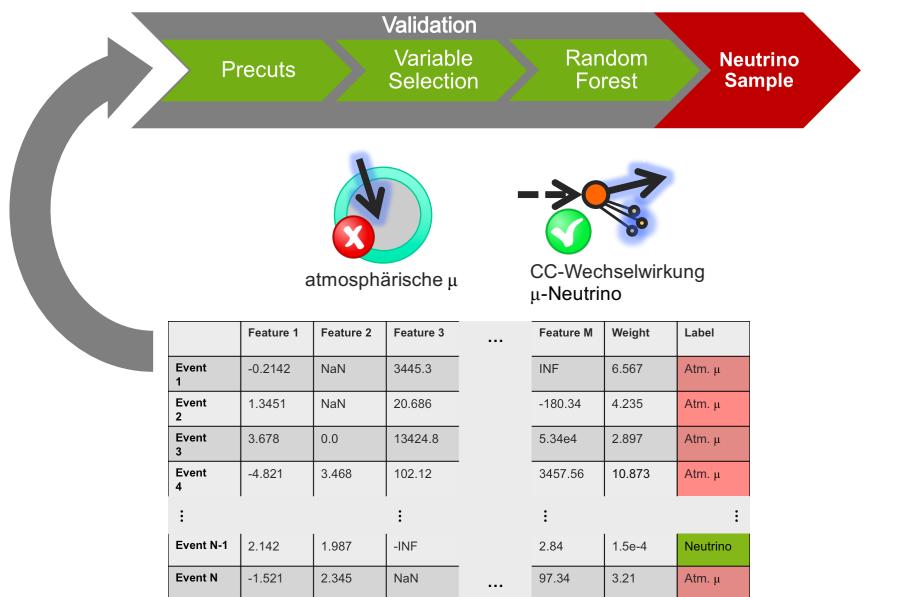


Prof. Dr. Dr. W. Rhode

Trennung von Signal und Untergrund – Die Daten

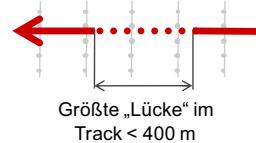
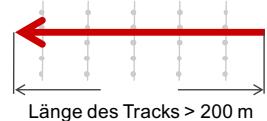
- Die Daten liegen in Form von (großen) Tabellen vor und bringen eine Herausforderung mit sich

hochdimensional $M \sim 2000$			Sig $\sim \frac{1}{1000000}$	Bkg	Simulation (noch) nicht perfekt	Steiles Spektrum $\Phi \propto E^{-2.7}$	
	Feature 1	Feature 2	Feature 3	...	Feature M	Weight	Label
Event 1	-0.2142	NaN	3445.3		INF	6.567	Atm. μ
Event 2	1.3451	NaN	20.686		-180.34	4.235	Atm. μ
Event 3	3.678	0.0	13424.8		5.34e4	2.897	Atm. μ
Event 4	-4.821	3.468	102.12		3457.56	10.873	Atm. μ
:				⋮	⋮	⋮	⋮
Event N-1	2.142	1.987	-INF		2.84	1.5e-4	Neutrino
Event N	-1.521	2.345	NaN	...	97.34	3.21	Atm. μ



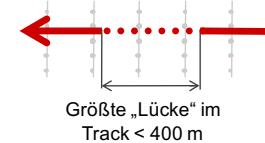
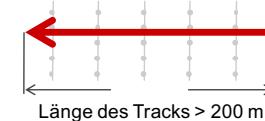
Trennung von Signal und Untergrund

- Schnitte, um Kaskaden und Events schlechter Qualität zu entfernen

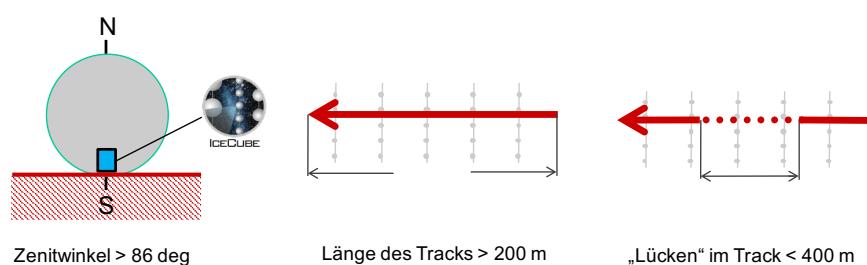
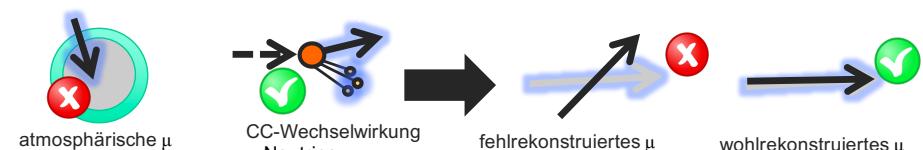
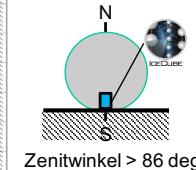
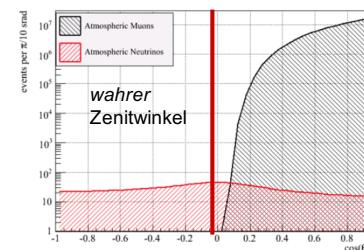


Trennung von Signal und Untergrund

- Schnitte, um Kaskaden und Events schlechter Qualität zu entfernen

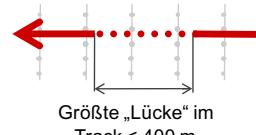
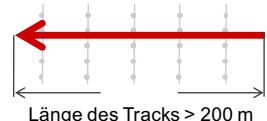


- Die Erde als Abschirmung gegen atmosphärische Myonen
 - Eindringtiefe von Myonen ins Eis: ca. 1km pro TeV

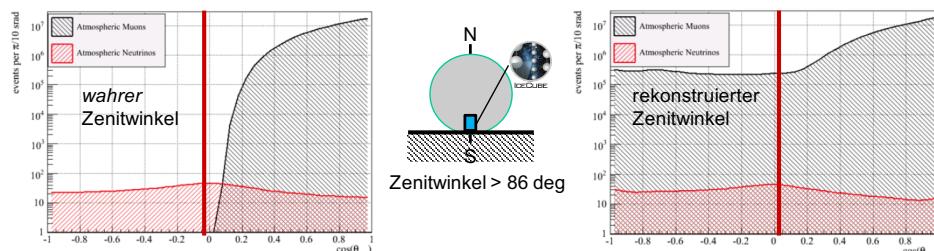


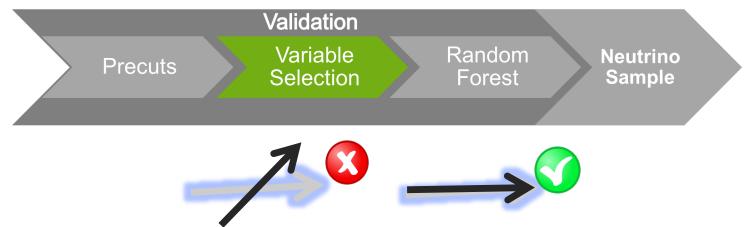
Trennung von Signal und Untergrund

- Schnitte, um Kaskaden und Events schlechter Qualität zu entfernen

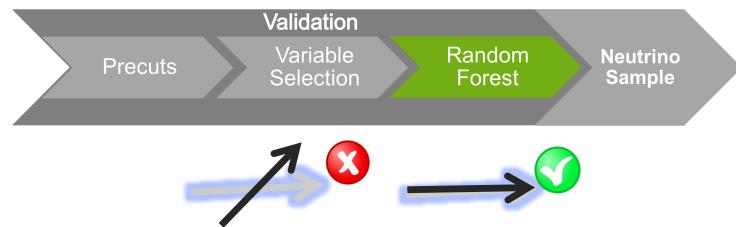
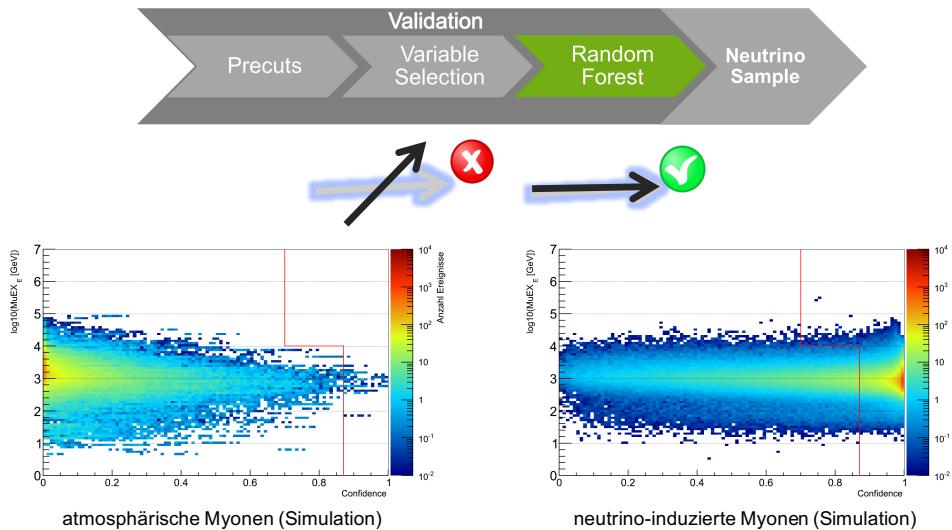


- Die Erde als Abschirmung gegen atmosphärische Myonen
 - Eindringtiefe von Myonen ins Eis: ca. 1km pro TeV

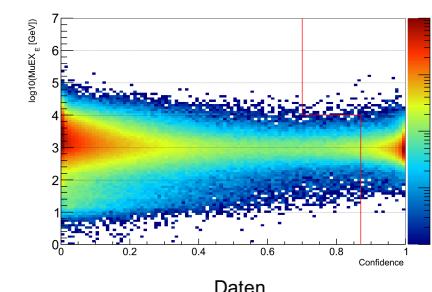




- Hochdimensionaler Datensatz mit vielen redundanten oder sogar unnützen Informationen für die Fragestellung
- Teilweise passen die Simulationen und die Messung in Variablen nicht zu einander
- Viele Dimensionen machen den Datensatz unhandlich und verhindern eine effektive Trennung
 - Auswahl von wenigen Variablen, die das Event möglichst gut beschreiben und viele Informationen über die Art des Events enthalten

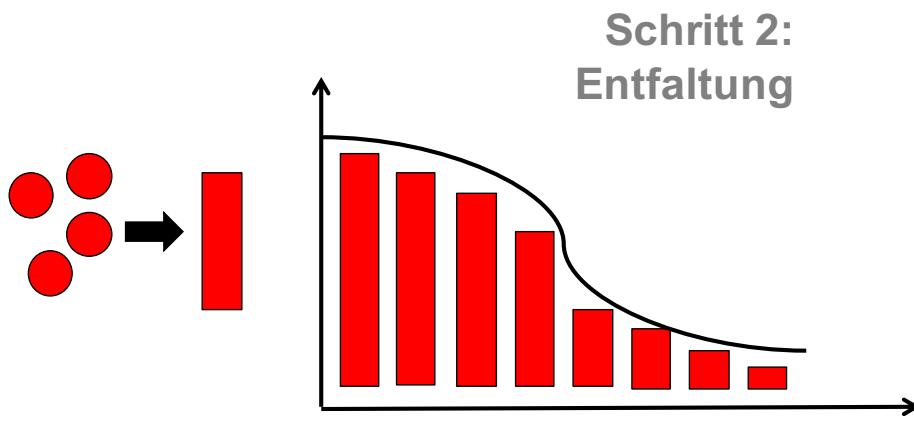


- Gesucht wird ein Modell, dass aus den Strukturen und Zusammenhängen im Datensatz Vorhersagen ableitet, ob es sich beim Event um ein Signal- oder Untergrund-Event handelt
 - Algorithmen des maschinellen Lernens (sogenannt Lernalgorithmen) können auf *gelabelten* Events (z.B. aus Simulationen) trainiert werden und treffen Vorhersagen für Events unbekannter Zugehörigkeit
 - Hier wird ein Random Forest genutzt, dessen Vorhersage (*Confidence*) zwischen 0 und 1 liegt und die Sicherheit angibt mit der der Random Forest das Event als Signal klassifiziert.

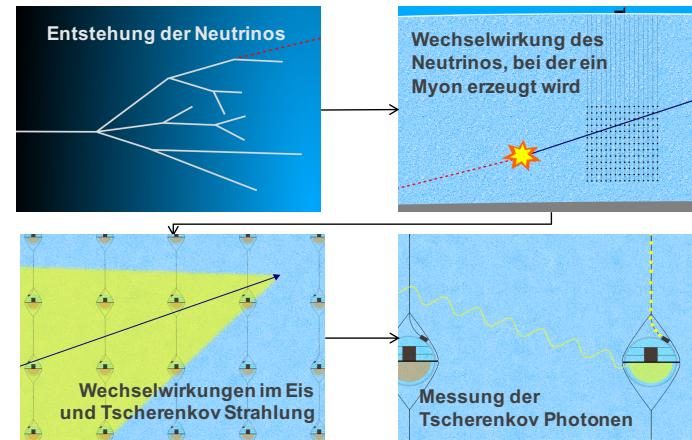


Energie	Atmos. μ	Neutrino μ
Gesamter Bereich	491 ± 263	93280
$E_\nu > 1 \text{ TeV}$	169 ± 140	39088
$E_\nu > 10 \text{ TeV}$	0.01 ± 0.01	233

- $99.6 \pm 0.2\%$ sind tatsächlich Myon-Neutrino Events
- 99.9999% des Untergrunds wurde abgelehnt



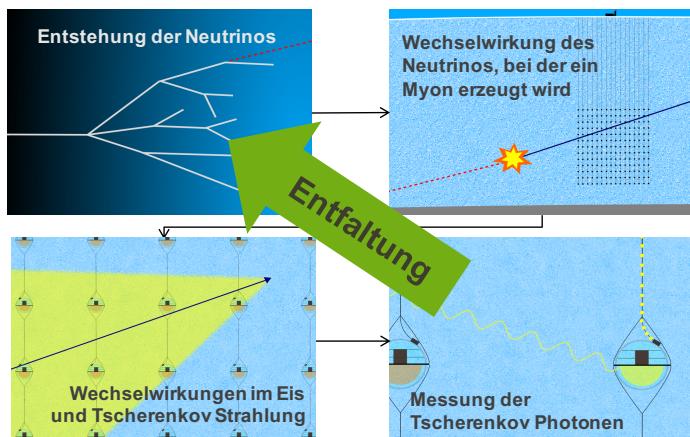
Entfaltung



Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse

Entfaltung

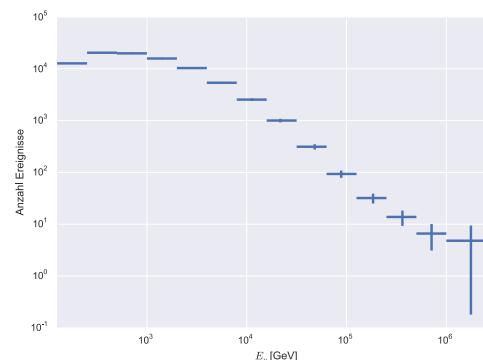


Prof. Dr. Dr. W. Rhode

$$g(y) = \int A(x, y) f(x) dx$$

Statistische Methoden
der Datenanalyse

Entfaltung des Myon-Neutrino Energiespektrums

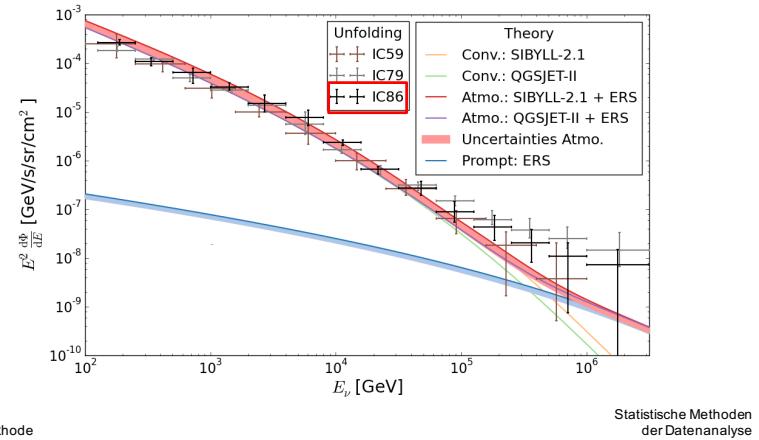


Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse

- Das Ergebnis der Entfaltung ist das Spektrum in Ereigniszahlen
- Mit Hilfe der Simulationen wird eine Akzeptanz-korrektur bestimmt
- Division mit dem beobachteten Himmelsbereich und der Messzeit verwandelt die Ereigniszahlen in einen Fluss
- Durch Vielfaches Entfalten von Simulationen, die systematische Variationen enthalten, werden die systematische Unsicherheiten abgeschätzt

Entfaltung des Myon-Neutrino Energiespektrums



Prof. Dr. Dr. W. Rhode

Statistische Methoden
der Datenanalyse