
Verbesserung der Energiregression bei CTA

Lars Möllerherm

September 25, 2018

Experimentelle Physik Vb
Fakultät Physik

Grundlagen

Gammaastronomie
Maschinelles Lernen

Ergebnisse

Rekonstruktion mit RF Regressor
Optimierung durch Mittelwert
Verschachtelte Methode
Energie-Transformation

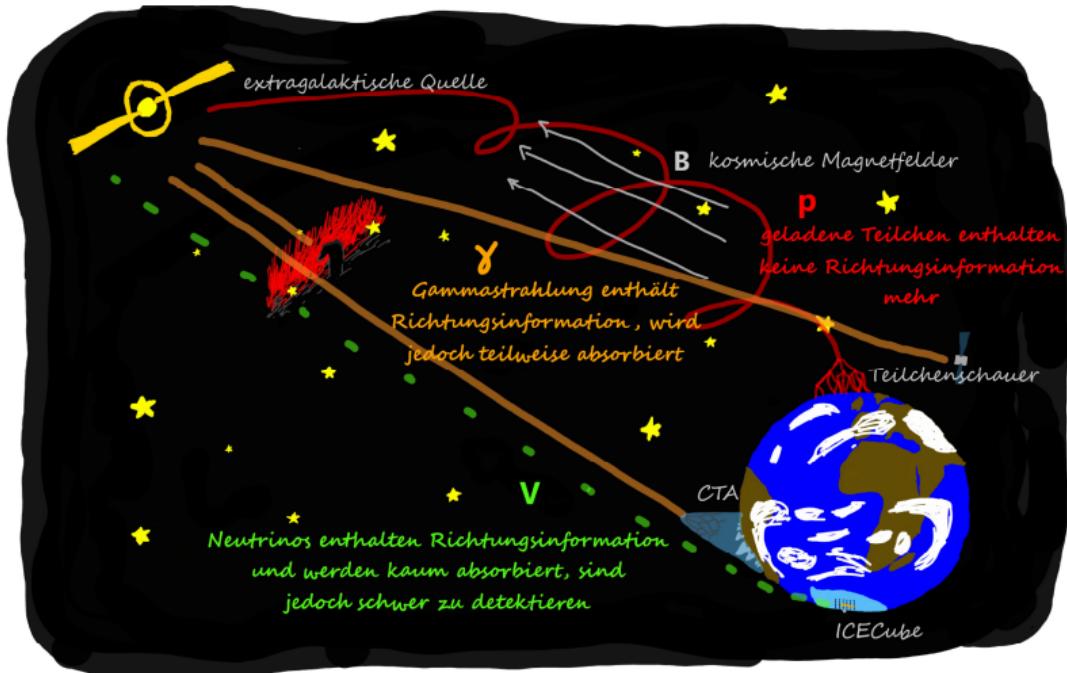
Fazit

Zusammenfassung
Perspektiven

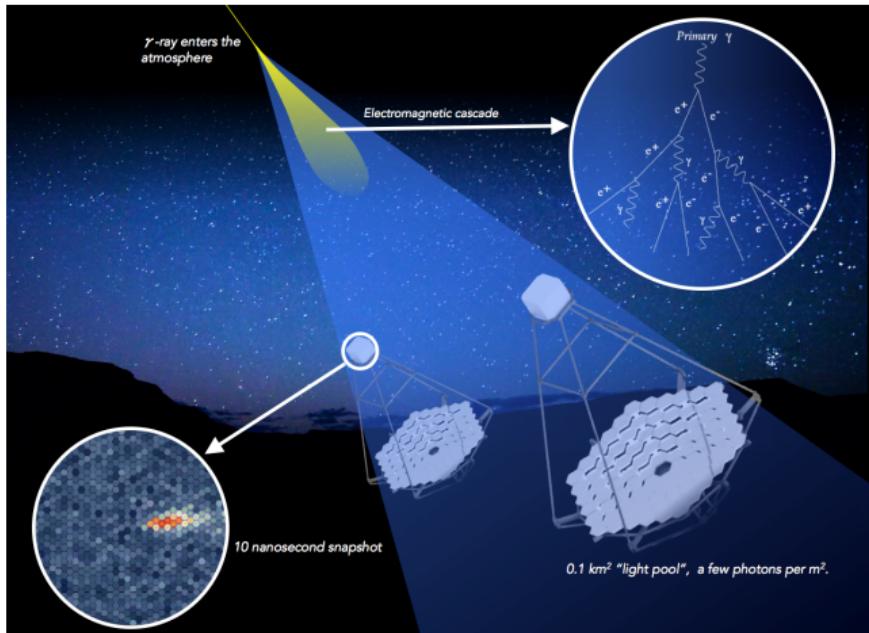


Grundlagen

Botenteilchen



Funktionsweise von CTA

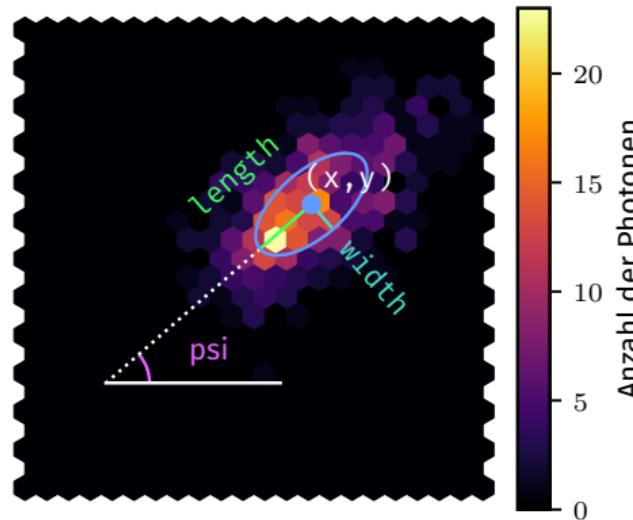


CTA homepage (How CTA works)

- Photon wechselwirkt mit der Atmosphäre: Schauer entsteht.
- Lichtgeschwindigkeit in Atmosphäre 0,03 % langsamer als Vakuum.
- Überlichtschnelle geladene Teilchen senden Kegel aus Cherenkov-Licht aus.
- Durchmesser: 250 m
- Dauer: einige ns.

Information über das Schauer

- Hillas Parameter, Wölbung und Krümmung
- Scaled Cuts Technik: $SV = \frac{v - \langle v \rangle}{\sigma_v}$ für $v = w; l$ [1, p. 104]
- Teleskopart
- Anzahl der ausgelösten Teleskope, LST, MST und SST
- gesamte Intensität
- Art des Primärteilchens

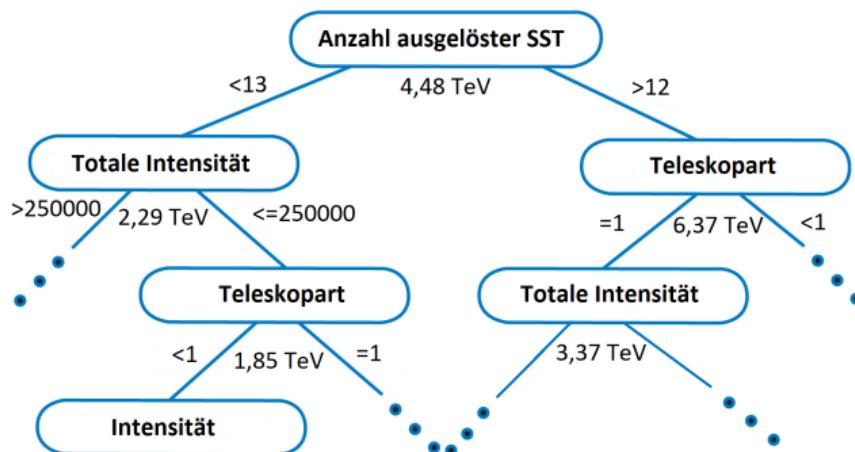


Max Nöthe

Aufgabenspektrum

- Entdecken neuer galaktischer Gammaquellen (SNR, PWN, Doppelsysteme, galaktische Zentren).
- Multiwellenlängen- und Multimessenger-Beobachtungen um hadronische Beschleunigungsprozesse zu untersuchen.
- Entdecken neuer Quellen mit großer Rotverschiebung.
- Detektieren von Zerfällen der Dunklen Materie.
- ...

Entscheidungsbaum



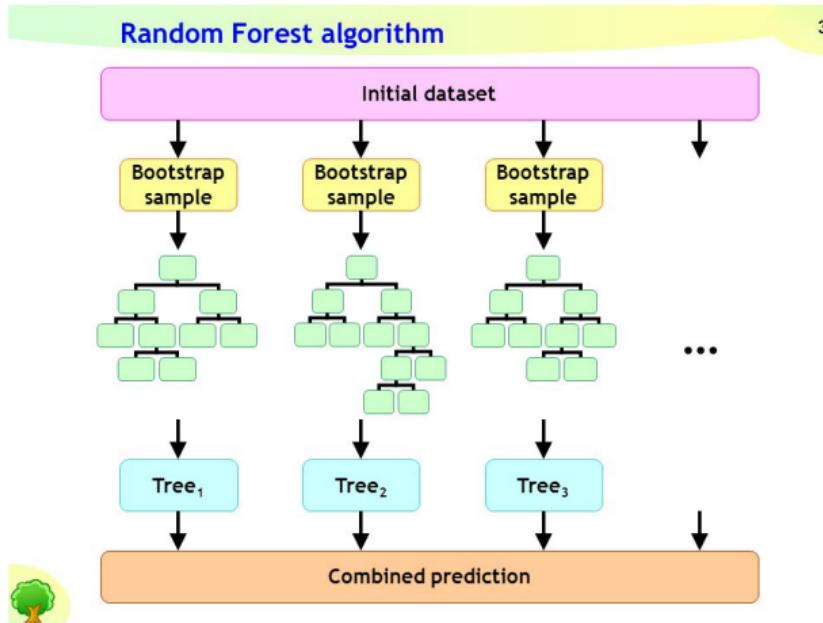
Kriterium

$$H(X_m) = \frac{1}{N_m} \sum_{i \in N_m} (\hat{y}_i - c_m)^2$$

mit

$$c_m = \frac{1}{N_m} \sum_{i \in N_m} \hat{y}_i$$

Random Forest



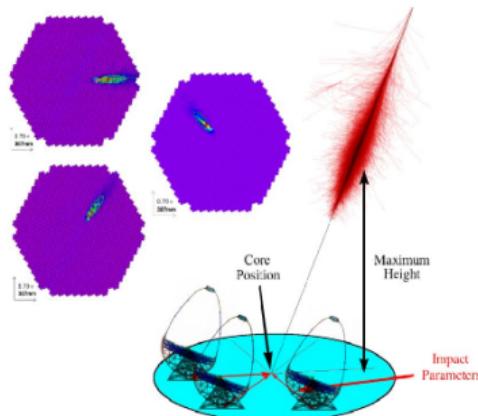
Warum der Random Forest Algorithmus:

- Interpretierbarkeit
- Zeitkomplexität: $\Theta_{\text{pred}}(M \log(N))$
- einfache Datenpräparation
- nicht anfällig gegen unbedeutende Attribute
- stabiles Ergebnis

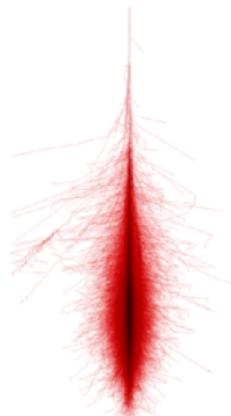
Pavel Polishchuk

Ergebnisse

Energierekonstruktion mit dem Random Forest



$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$



Teleskopspezifische
Vorhersage

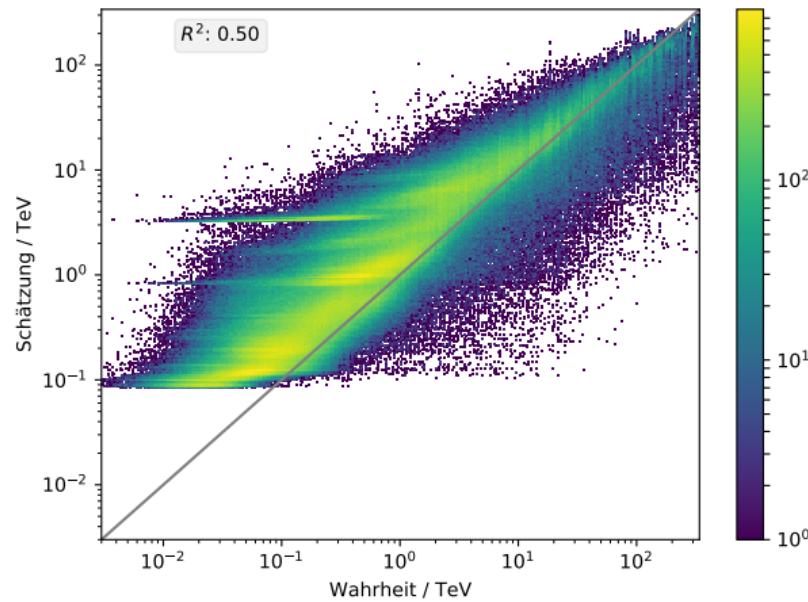
Mittelwertbildung

Eventspezifische
Vorhersage

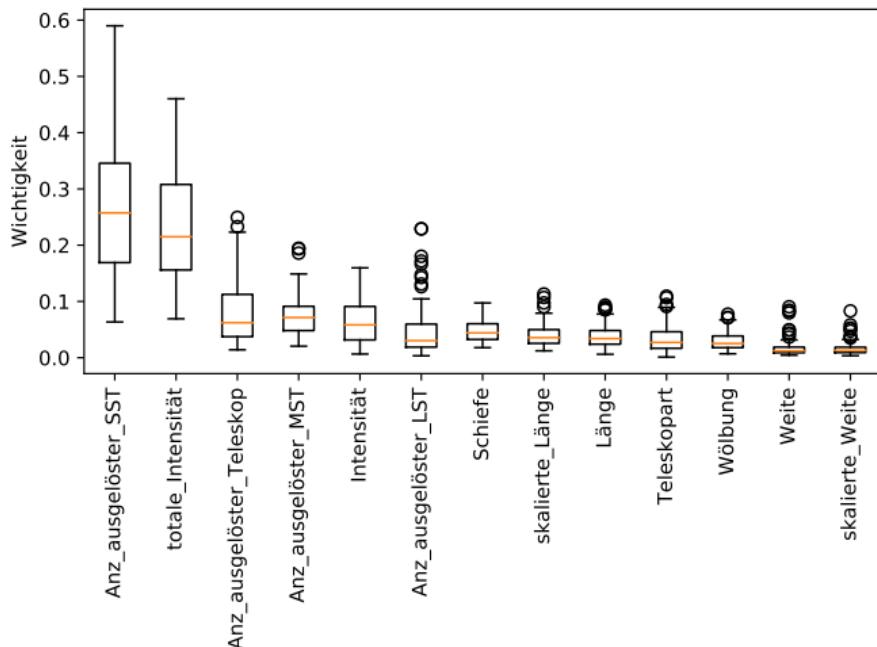
Schätzung jedes Teleskop-Ereignis

R^2 -Score:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$



Rangliste der Attribute

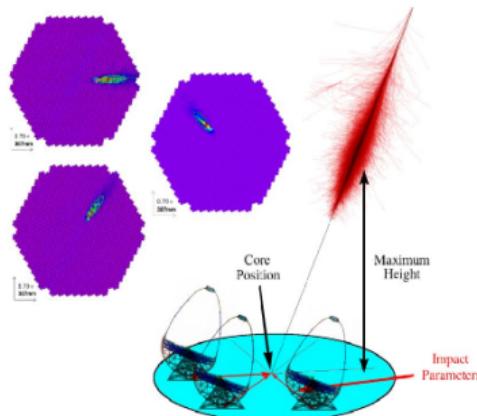


Kriterium

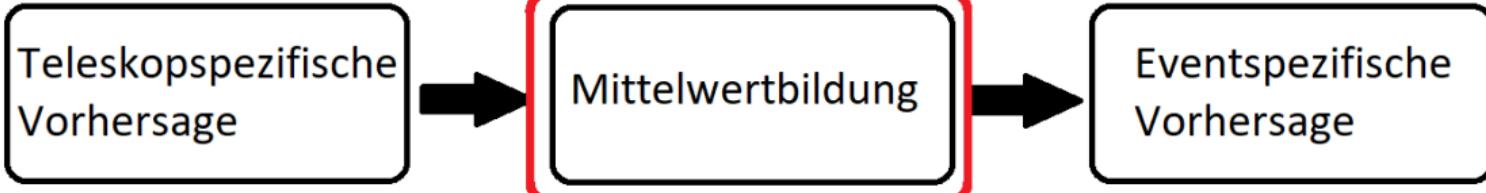
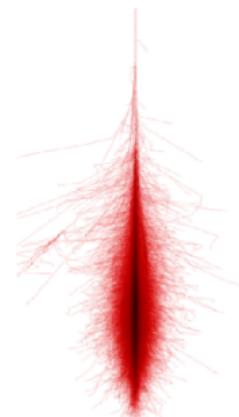
Rangordnung der Attribute in jedem Entscheidungsbaum

- Aufgrund von Bootstrapping und dem CART-Algorithmus kann eine Verzerrung auftreten.
- Attribute mit einer kleinen Anzahl an Kategorien werden unterschätzt.

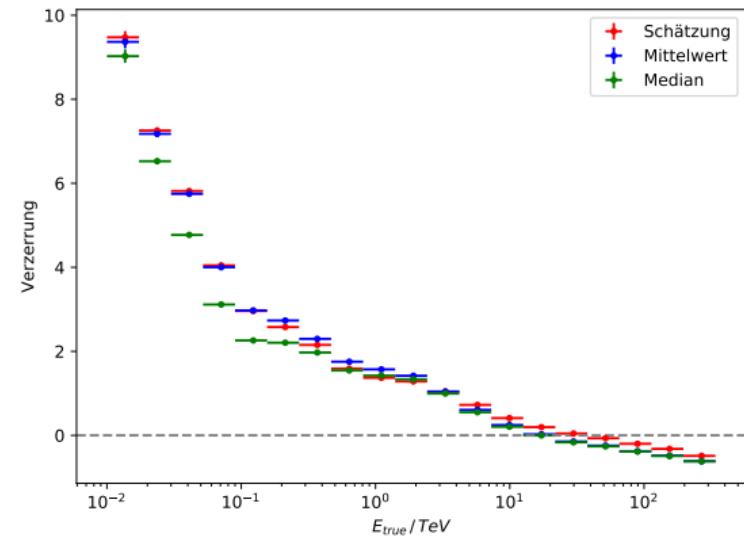
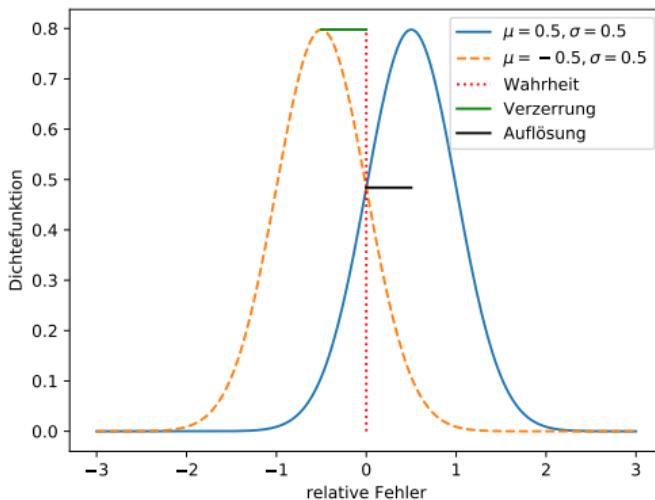
Energierekonstruktion mit dem Random Forest



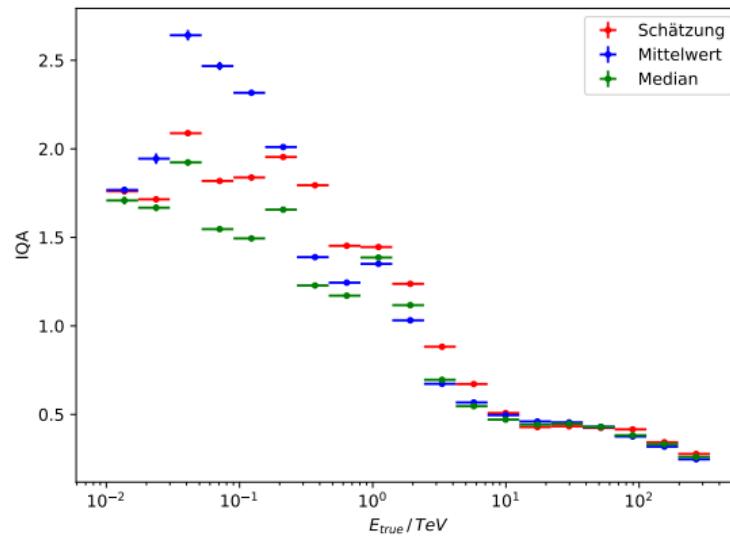
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$



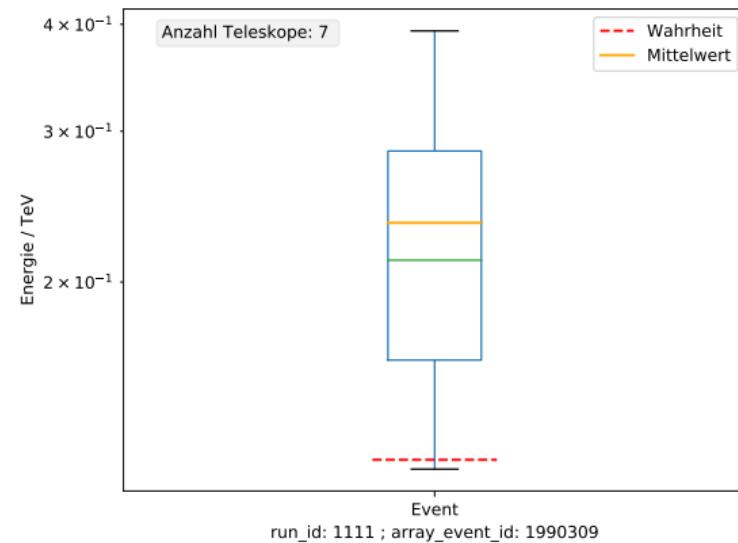
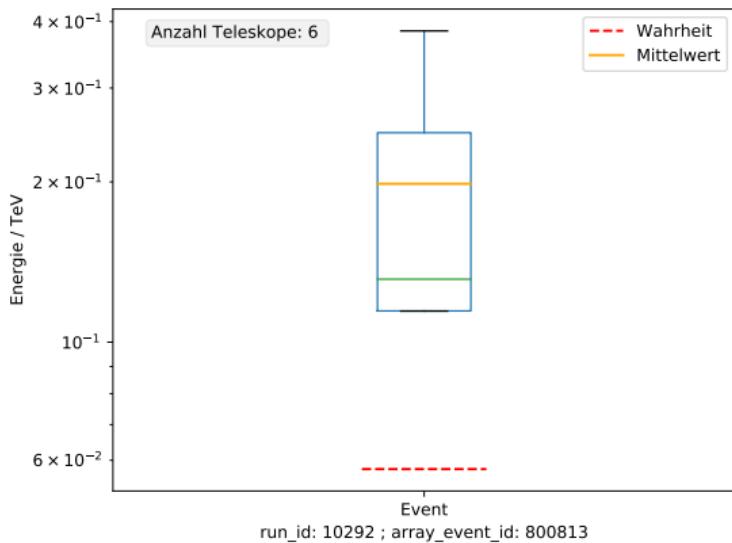
Arithmetisches Mittel und Median



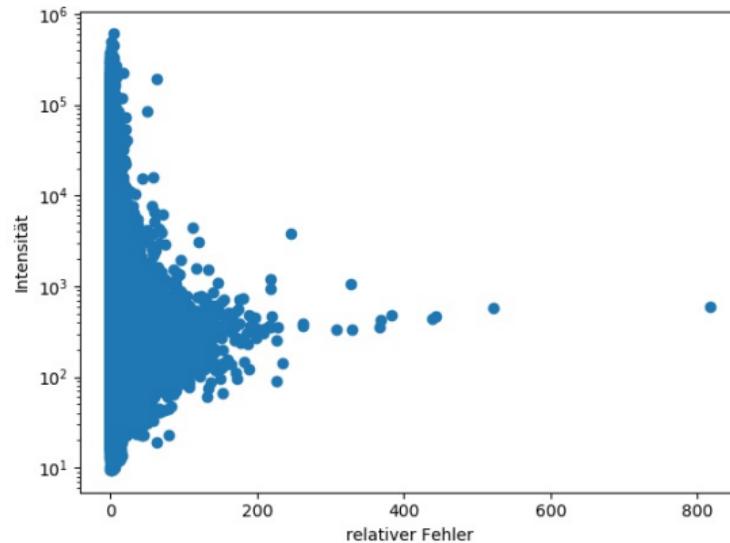
Mittelwert des relativen Fehlers für verschiedene Energieräume.



Die Hälfte des IQA von 68,26 % des relativen Fehlers für verschiedene Energiebereiche.



Intensität als Gewicht



Sensitivität als Gewicht

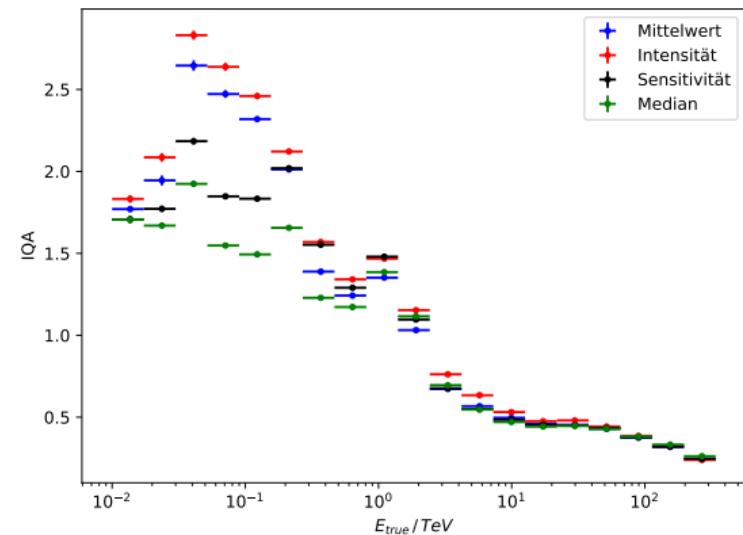
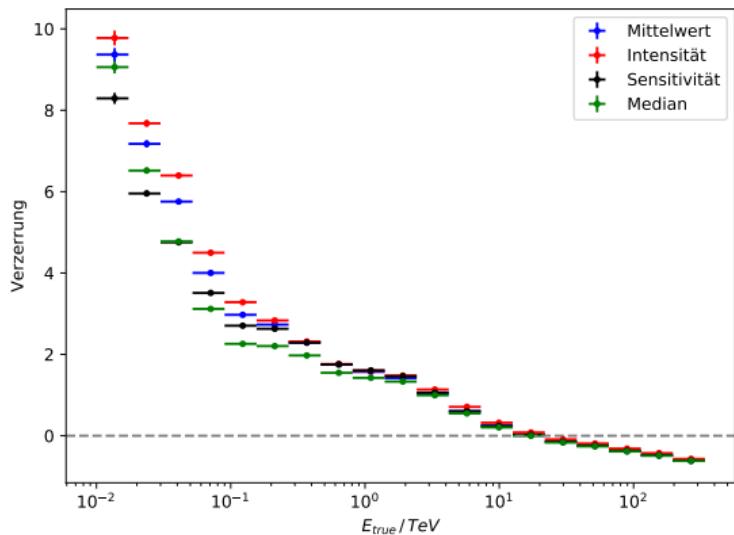
	Large-Sized Telescope (LST)	Medium-Sized Telescope (MST)			Small-Sized Telescope (SST)		
		FlashCam	NectarCam	SCT	ASTRI	GCT	SST-1M
Required energy range	20 GeV – 3 TeV	80 GeV – 50 TeV			1 TeV – 300 TeV		
Energy range (in which subsystem provides full system sensitivity)	20 GeV – 150 GeV	150 GeV – 5 TeV			5 TeV – 300 TeV		
Number of telescopes	4 (South) 4 (North)	25 (South) 15 (North)			70 (South) 0 (North)		
Optical design	Parabolic	Modified Davies-Cotton	Schwarzschild-Couder	Schwarzschild-Couder		Davies-Cotton	
Primary reflector diameter	23.0 m	11.5 m	9.7 m	4.3 m	4.0 m	4.0 m	
Secondary reflector diameter	--	--	5.4 m	1.8 m	2.0 m	--	

CTA homepage (CTA technology)

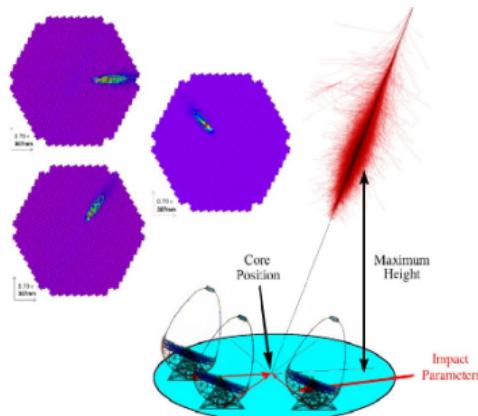
Vollsensitiver Bereich: 2

Teilsensitiver Bereich: 1

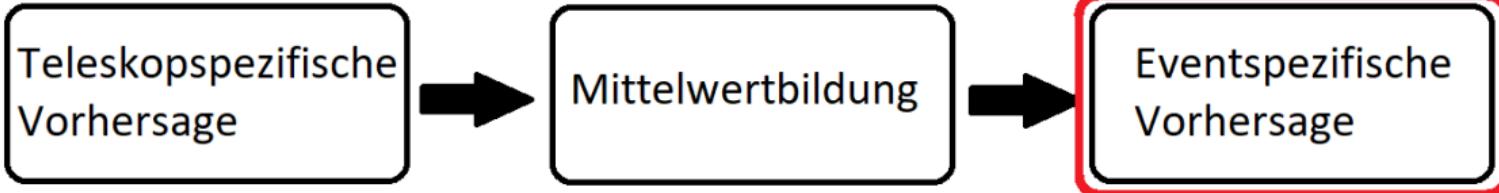
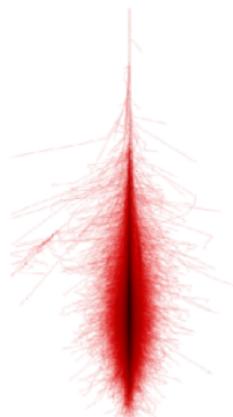
Außerhalb der Bereiche: 0,10



Energierekonstruktion mit dem Random Forest



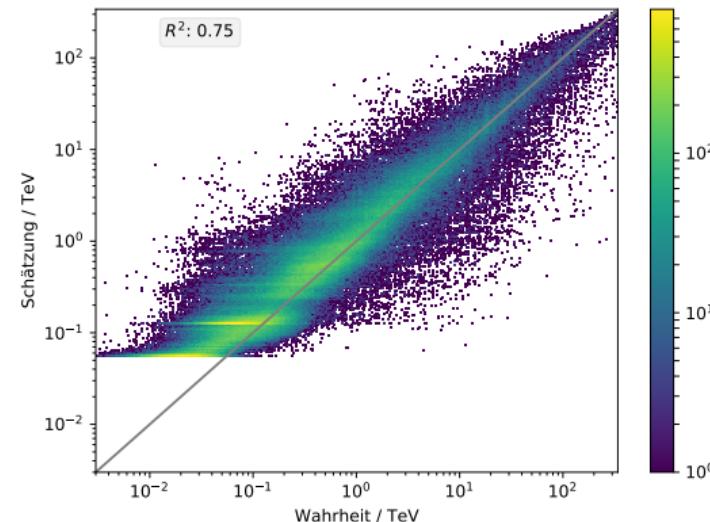
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$



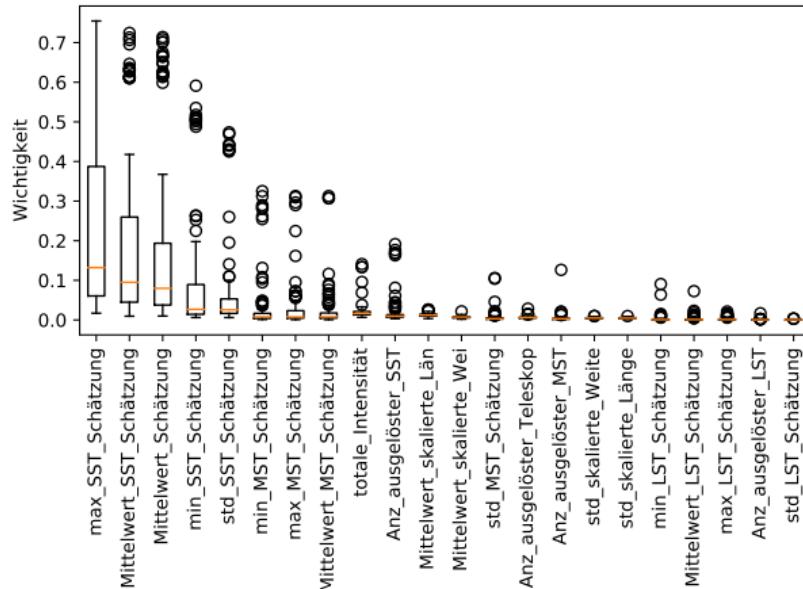
Eventspezifische Vorhersage

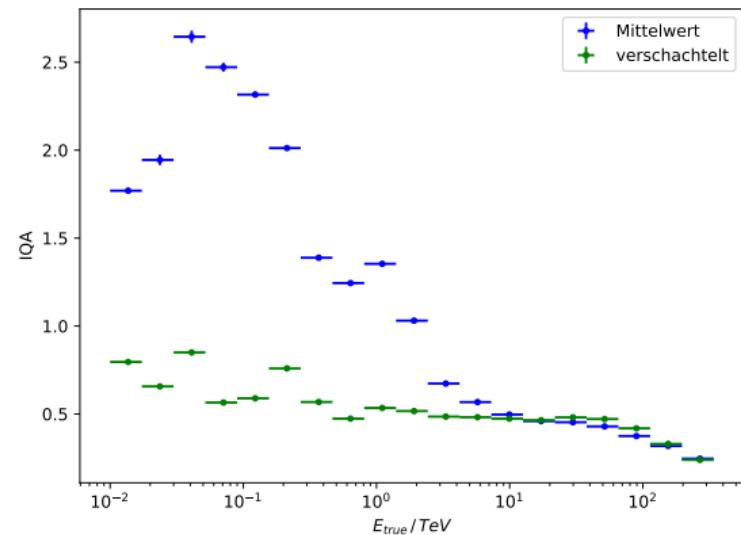
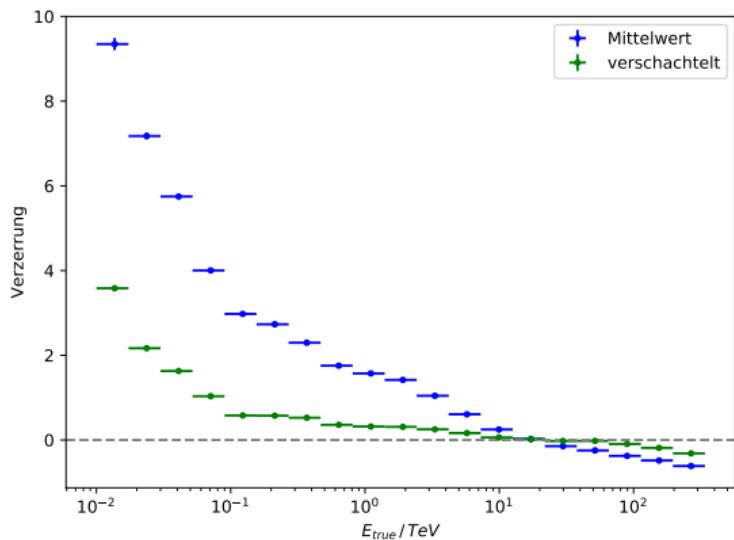
■ Attribute:

- totale Intensität, Anzahl ausgelöster Teleskope, SST, MST, LST
- Mittelwert und Standardabweichung der skalierten Größen
- Mittelwert der vorherigen Schätzungen
- Mittelwert, Standardabweichung, Maximum und Minimum der Schätzungen der SST, MST und LST

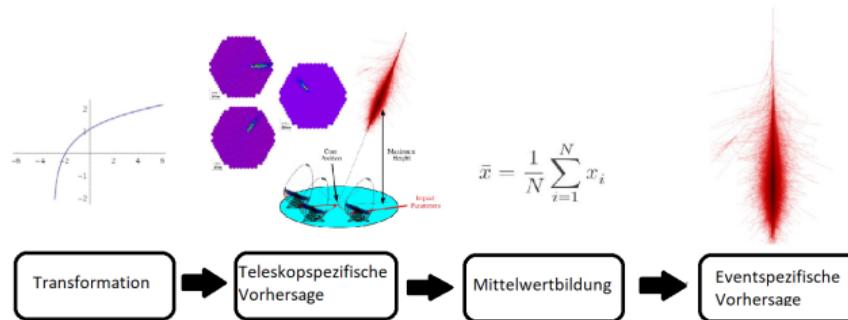


Rangsliste der Attribute des zweiten Random Forest





Transformation der Energien

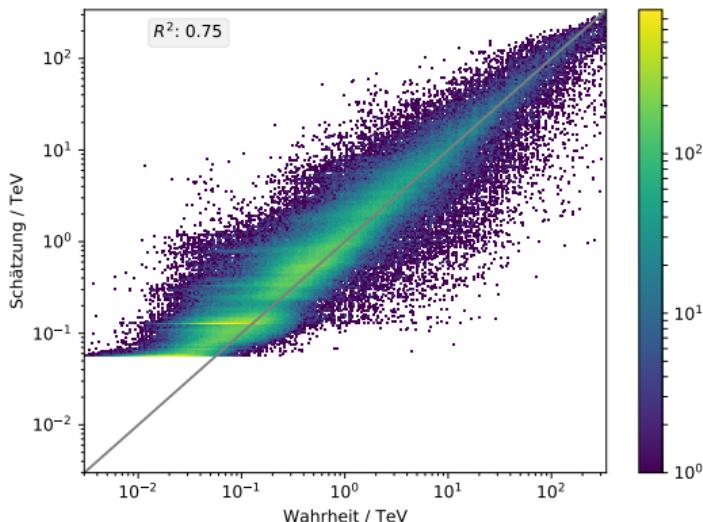


Bijektive Transformation auf \mathbb{R}^+ :

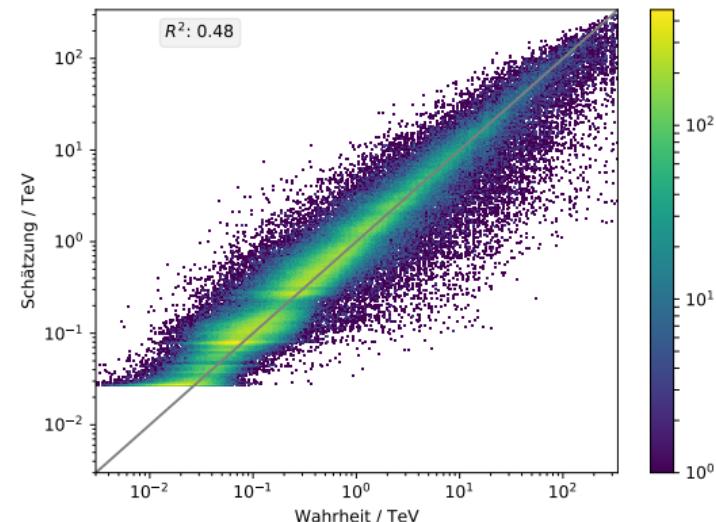
$$E_{\text{trafo}} = \ln(E_\gamma + 3)$$

Rücktransformation:

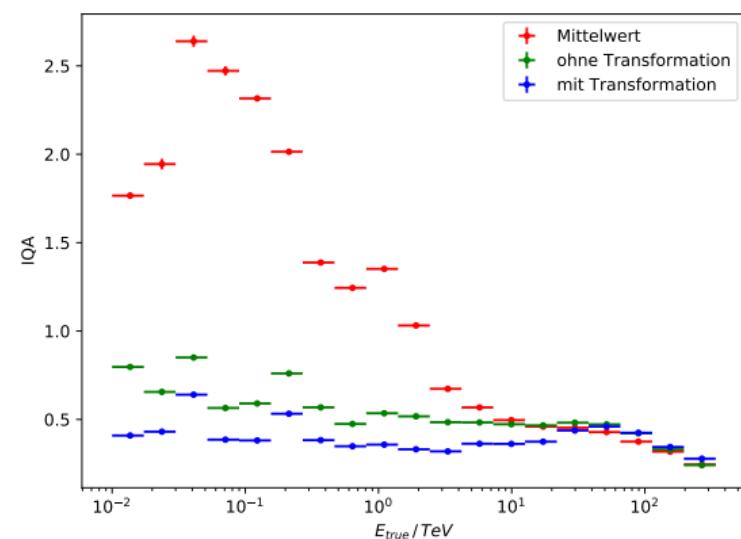
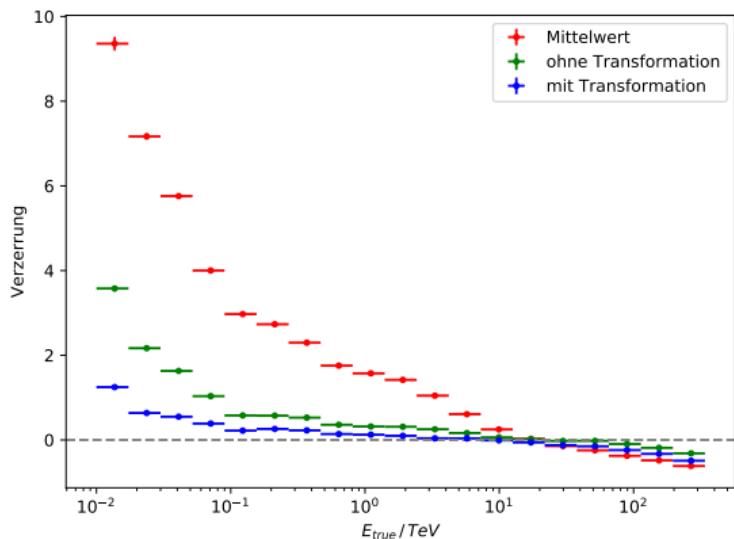
$$E_\gamma = \exp(E_{\text{trafo}}) - 3$$



Ohne Transformation



Mit Transformation



Absoluter Fehler

Mittlere quadratische Fehler:

$$\text{mse} = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2$$

	mse
Teleskopspezifische Schätzung	102,94 TeV ²
Arithmetisches Mittel	65,31 TeV ²
Median	66,61 TeV ²
Gewichtung mit Intensität	58,61 TeV ²
Gewichtung mit Sensitivität	64,72 TeV ²
Eventspezifische Schätzung	37,18 TeV ²

Transformation:

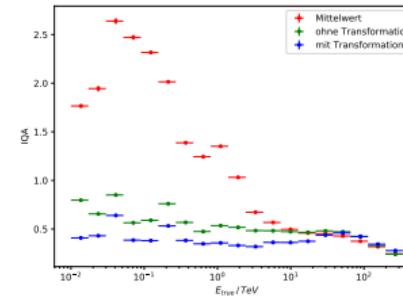
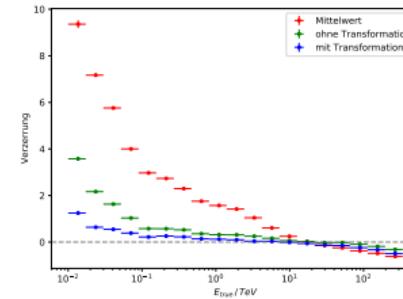
Teleskopspezifische Schätzung	149,48 TeV ²
Arithmetisches Mittel	94,60 TeV ²
Eventspezifische Schätzung	53,65 TeV ²

Ergebnisse: Energie-Transformation

Fazit

Zusammenfassung der Ergebnisse

- Es wurden keine Cuts angewendet.
- Anpassung der Methode auf die Aufgabenspezifische Analyse.
- Beste Auflösung und geringste Verzerrung des relativen Fehlers bei der verschachtelten Methode mit Transformation. Es könnte Verzerrung und Auflösung in weiten Bereichen um ca. 75 % gesenkt werden.
- Geringster mittlerer quadratischer Fehler bei der verschachtelten Methode ohne Transformation. Der absolute Fehler wurde um ca. 66 % gesenkt.



Mögliche Perspektiven

- Abstand von Teleskop und Schauer und Höhe des Schauers als Attribut hinzunehmen
- aufgabenspezifische Wahl der Transformation
- Kriterium für Entscheidungsbaum entwickeln, welches den absoluten Fehler nutzt





Mathieu de Naurois. "Very High Energy astronomy from H.E.S.S. to CTA. Opening of a new astronomical window on the non-thermal Universe". PhD thesis. Ecole Polytechnique, 2012. URL: <http://tel.archives-ouvertes.fr/tel-00687872>.

■ Datensatz:

- 3 322 938 Teleskopevents
- 996 653 Arrayevents
- 78 % punktgerichtet und 22 % diffuse
- 33 % Trainings- und 66 % Testdatensatz

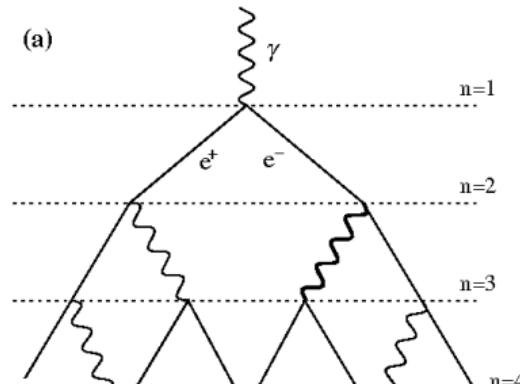
■ Random Forest:

- maximale Tiefe = 10
- Größe des Waldes = 100
- Anzahl der Attribute = \sqrt{N}
- minimale Blattgröße = 1

Schauer Modell

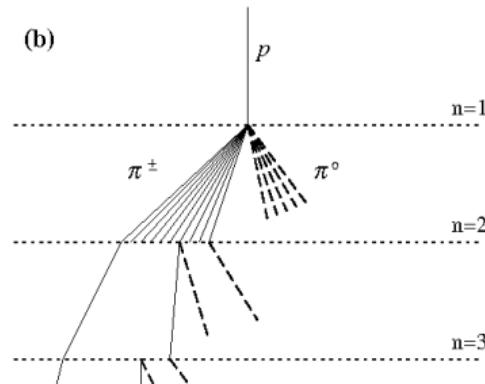
Gamma induziertes Schauer:

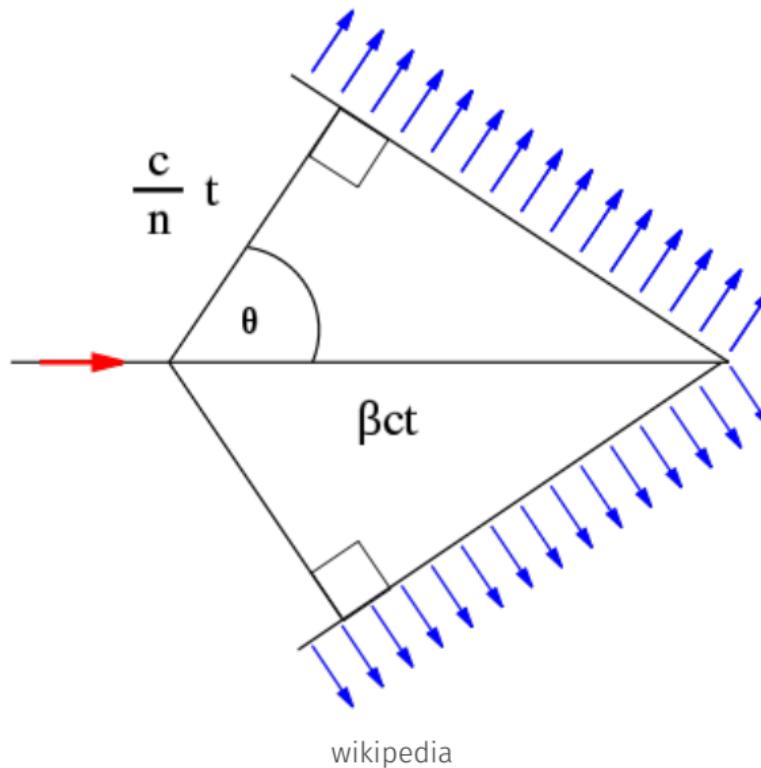
- γ
- $e^+; e^-$



Proton induziertes Schauer

- Hadronische Komponente: Kernfragmente, π , Kaon, ...
- elektromagnetische Komponente: γ von π^0 zerfall
- Muonen vom Zerfall der geladenen Mesonen
- Neutrios vom Zerfall der geladenen Mesonen und der Muonen





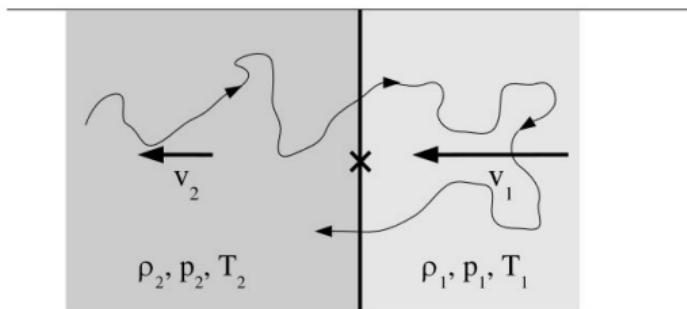
wikipedia

Fermi-Beschleunigung

erster Ordnung

$$\left\langle \frac{\delta E}{E} \right\rangle \approx \frac{2}{3} \frac{\delta v}{c},$$

δv Geschwindigkeitsunterschied hinter und vor der Druckwelle

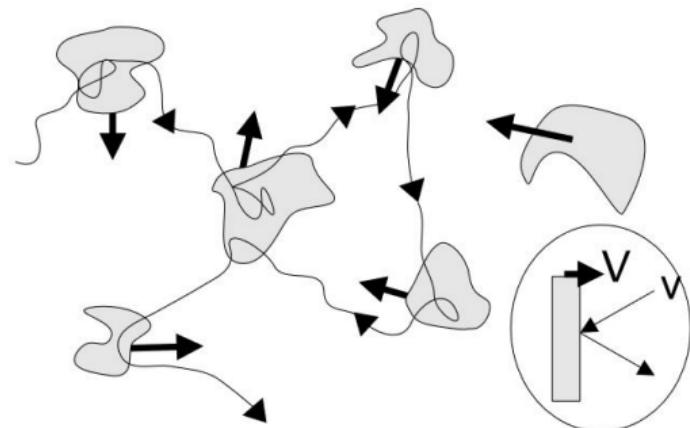


Mathieu de Naurois

zweiter Ordnung

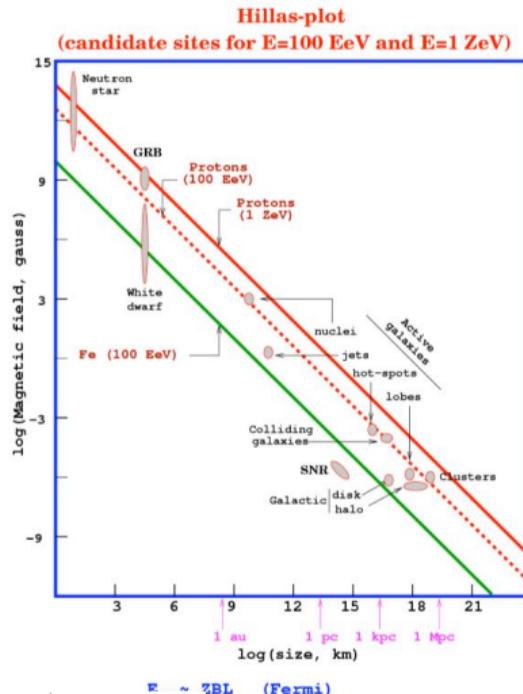
$$\left\langle \frac{\delta E}{E} \right\rangle = \frac{8}{2} \left(\frac{v}{c} \right)^2.$$

v Druckwellengeschwindigkeit



Mathieu de Naurois

Gammaquellen



Process	Energy Losses	Particle Energy	Radiated Spectrum
Synchrotron	$\frac{dE}{dt} \propto E^2$	$E_\gamma \approx \frac{3eB}{4\pi m_e^3 c^4} E_e^2$ up to $\sim \frac{3eB}{4\pi m_e^3 c^4} E_C^2$	$I(E_\gamma) \propto E_\gamma^{(1-s)/2}$
Inverse Compton (Thomson)	$\frac{dE}{dt} \propto E^2$	$E_\gamma \approx \frac{E_e^2 E_0}{m_e^2 c^4}$ up to $\gamma_C^2 E_0$	$I(E_\gamma) \propto E_\gamma^{(1-s)/2}$
Inverse Compton (Klein-Nishina)	$\frac{dE}{dt} \propto E^2$	$E_\gamma \approx \frac{E_e^2 E_0}{m_e^2 c^4}$ up to $\gamma_C m_e c^2$	$I(E_\gamma) \propto E_\gamma^{(1-s)/2}$
Bremsstrahlung	$\frac{dE}{dt} \propto E$	$E_\gamma \approx 0.5 E_e$	$I(E_\gamma) \propto E_\gamma^{-s}$
Hadronic Interaction	$\frac{dE}{dt} \propto E$	$E_\gamma \approx 0.1 E_p$	$I(E_\gamma) \propto E_\gamma^{-s}$ above 67.5 MeV

Mathieu de Naurois

Unfolding

Aufgrund der endlichen Auflösung ist das Signal

$$g(x) = \int_Y R(x|y) \cdot f(y) dy \text{ oder } y = A \cdot x. \quad (1)$$

Um das wirkliche Energiespektrum f zu erhalten, muss entfaltet werden. Methoden zum Entfalten:

- mit maximum likelihood fit
- oder mit Bayes' Theorem

Bayes' Theorem:

$$P(C|E, \Lambda, I) = \frac{P(X|C, \Lambda, I) \cdot P(C|I)}{\sum_C P(E|C, \Lambda, I) \cdot P(C|I)} \quad (2)$$

I = informationsstand unter der die Analyse gemacht wurde und

$$\lambda_{ij} = P(E_j|C_i, I) \quad (3)$$

Resolutionmatrix. Mit

$$\theta_{ij} = \frac{\lambda_{ij} \cdot P(C_i|I)}{\sum_i \lambda_{ij} \cdot P(C_i|I)} = P(C_i|E_j, I) \quad (4)$$

lässt sich dann C berechnen

Konvergenz des RF

Wenn die Anzahl der Entscheidungsbäume in einem RF erhöht wird, konvergiert der generalisierte Fehler

$$PE = P_{X,y}(mg(X,y) < 0) \quad (5)$$

gegen

$$P_{X,y}(P_\theta(h(X,\theta) = y) - \max_{j \neq y} P_\theta(h(X,\theta) = j) < 0) \quad (6)$$

und es kann durch eine Vergrößerung des Waldes nicht zum Übertraining kommen.

$$mg(X,y) = av_k I(h_k(X) = y) - \max_{j \neq y} av_k I(h_k(X) = j) \quad (7)$$

die Gewinn-Funktion, $h_k(X)$ = Vorhersage des k -ten Entscheidungsbaumes; $I(\cdot)$ die charakteristische Funktion; $mg(X,y) > 0 \rightarrow$ richtiges Ergebnis des RF; θ = für den Entscheidungsbaum gewählte Attribute.