

Gamma/Hadron – Separation

Momentane Situation bei MAGIC

Möglichkeiten durch ein externes Programm

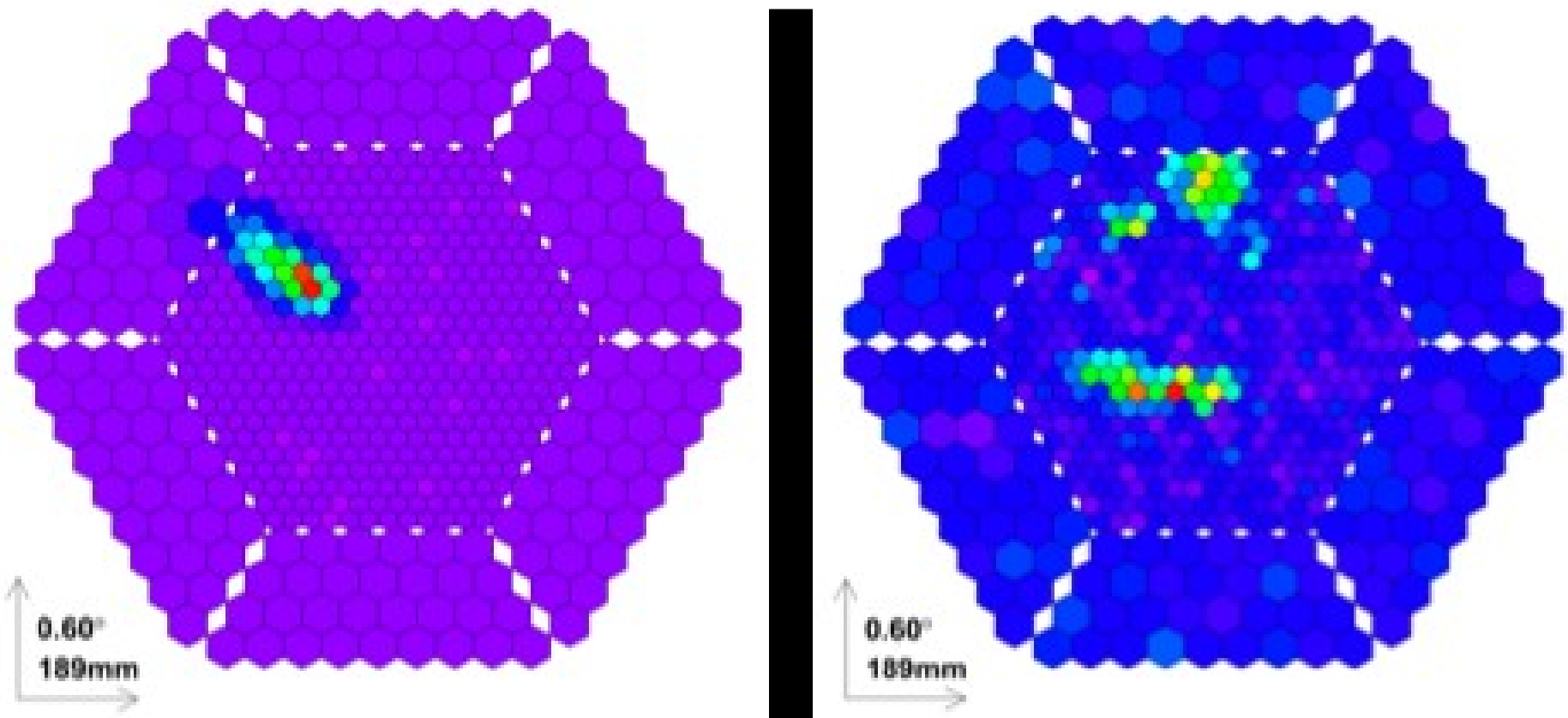
Martin Schulze

Technische Universität Dortmund

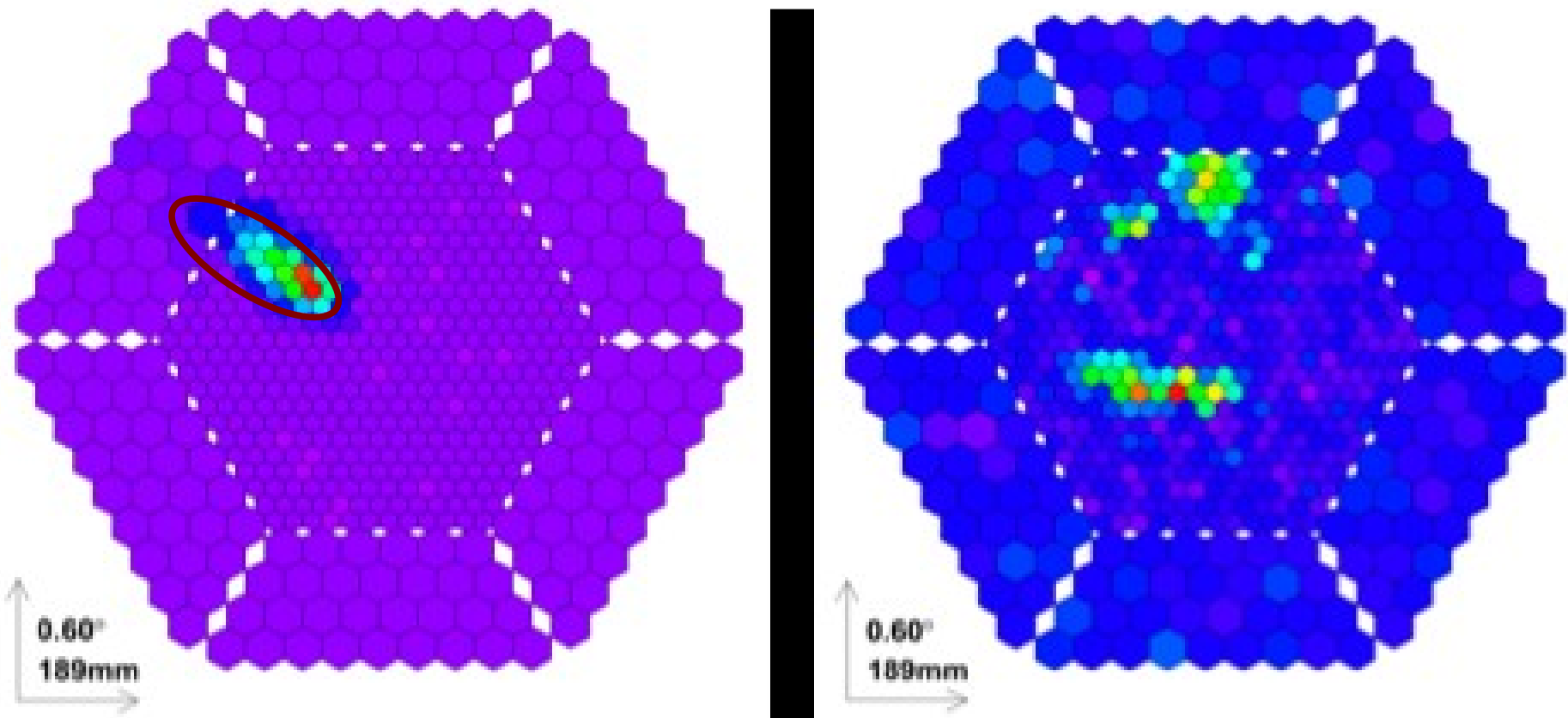
Lehrstuhl EVb



Darstellung eines Gammas (links) und Hadronereignisses (rechts) in der Kamera



Darstellung eines Gammas (links) und Hadronereignisses (rechts) in der Kamera



Von dieser Ellipse wird, unter anderem, die Breite (width), Länge (length), Anzahl der Photoelektronen (Size) berechnet. Zusätzlich hat man Informationen über die Zeitverteilung des Schauers.

Analyse Kette bei MAGIC: MARS (Barcelona – Version)

CALLISTO – Kalibration der Rohdaten

Analyse Kette bei MAGIC: MARS (Barcelona – Version)

CALLISTO – Kalibration der Rohdaten



STAR – Berechnung der (Hillas-)Parameter und Imagecleaning

Analyse Kette bei MAGIC: MARS (Barcelona – Version)

CALLISTO – Kalibration der Rohdaten



STAR – Berechnung der (Hillas-)Parameter und Imagecleaning



OSTERIA – Training der Random Forests zur Separation und Energie Berechnung

Analyse Kette bei MAGIC: MARS (Barcelona – Version)

CALLISTO – Kalibration der Rohdaten



STAR – Berechnung der (Hillas-)Parameter und Imagecleaning



OSTERIA – Training der Random Forests zur Separation und Energie Berechnung



MELIBEA – Separation der Gammas von den Hadronen

Analyse Kette bei MAGIC: MARS (Barcelona – Version)

CALLISTO – Kalibration der Rohdaten



STAR – Berechnung der (Hillas-)Parameter und Imagecleaning



OSTERIA – Training der Random Forests zur Separation und Energie Berechnung



MELIBEA – Separation der Gammas von den Hadronen



FLUXLC – Berechnung des Energieflusses und des Spektrums

Analyse Kette bei MAGIC: MARS (Barcelona – Version)

CALLISTO – Kalibration der Rohdaten



STAR – Berechnung der (Hillas-)Parameter und Imagecleaning



OSTERIA – Training der Random Forests zur Separation und Energie Berechnung



MELIBEA – Separation der Gammas von den Hadronen



FLUXLC – Berechnung des Energieflusses und des Spektrums



UNFOLDING – Entfaltung des Spektrums

Analyse Kette bei MAGIC: MARS (Barcelona – Version)

CALLISTO – Kalibration der Rohdaten



STAR – Berechnung der (Hillas-)Parameter und Imagecleaning



OSTERIA – Training der Random Forests zur Separation und Background



MELIBEA – Separation der Gammas von den H



FLUXLC – Berechnung des Energieflusses und des Spektrums



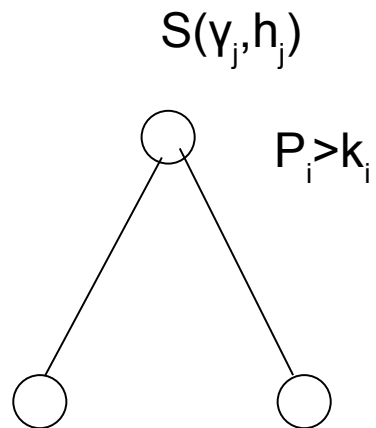
UNFOLDING – Entfaltung des Spektrums

Wird durchgeführt
am PIC in
Barcelona

Prinzipieller Aufbau eines Random Forrest

Benötigt werden zwei Trainingssets von Daten:

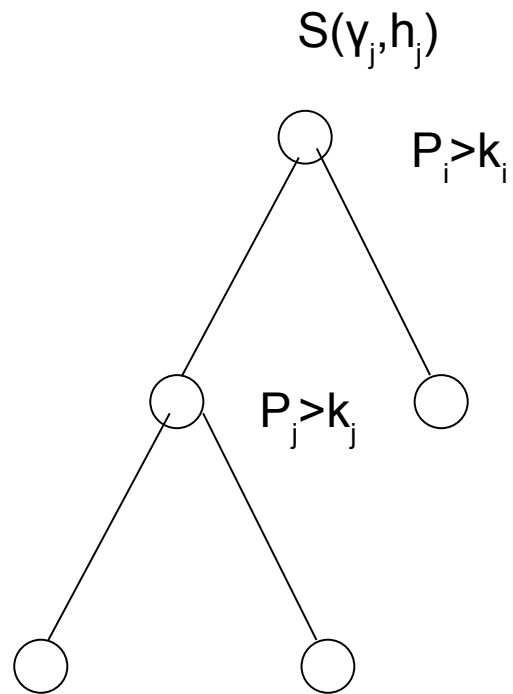
- a) Gammaevents (γ) (aus MC-Simulationen) und
- b) Hadronen (h) (Off-Data)



Prinzipieller Aufbau eines Random Forrest

Benötigt werden zwei Trainingssets von Daten:

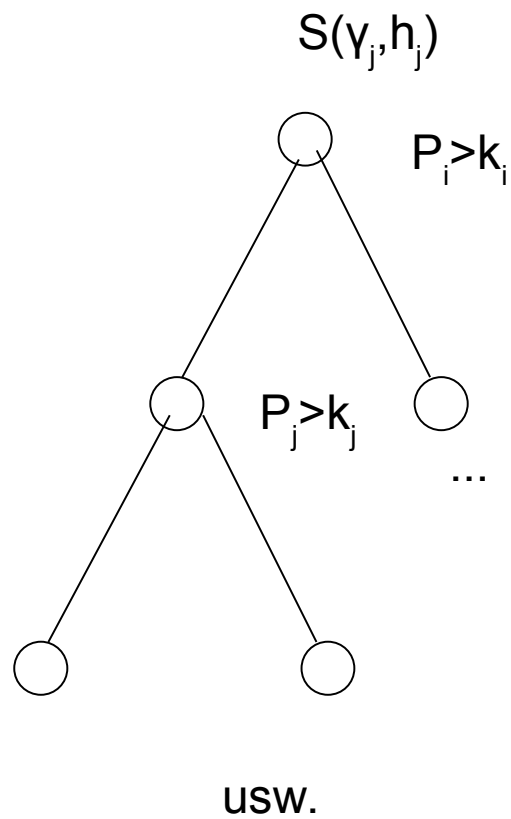
- a) Gammaevents (γ) (aus MC-Simulationen) und
- b) Hadronen (h) (Off-Data)



Prinzipieller Aufbau eines Random Forrest

Benötigt werden zwei Trainingssets von Daten:

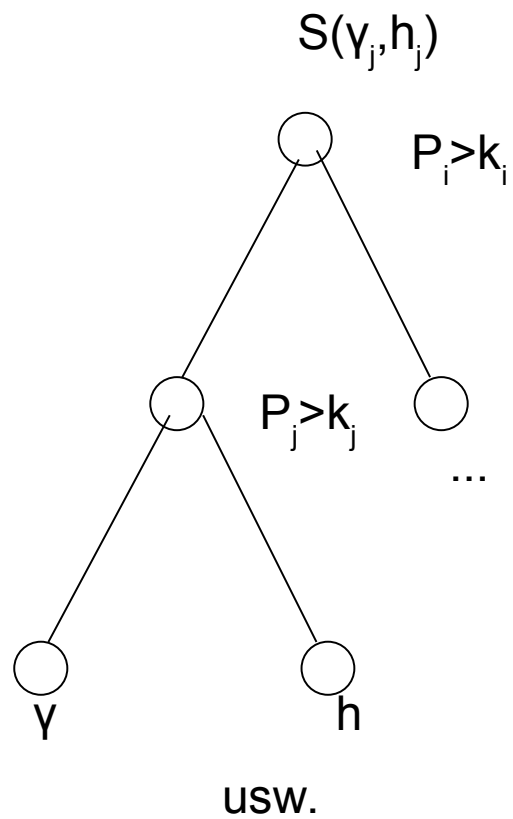
- a) Gammaevents (γ) (aus MC-Simulationen) und
- b) Hadronen (h) (Off-Data)



Prinzipieller Aufbau eines Random Forrest

Benötigt werden zwei Trainingssets von Daten:

- a) Gammaevents (γ) (aus MC-Simulationen) und
- b) Hadronen (h) (Off-Data)

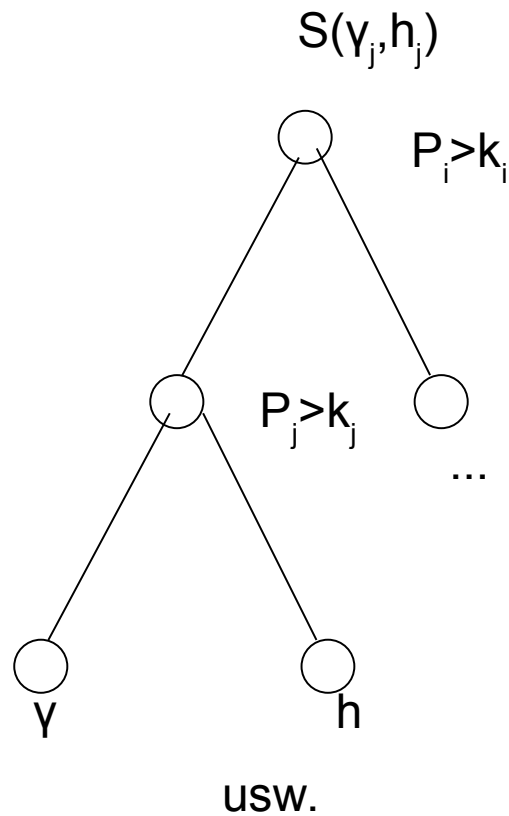


Prinzipieller Aufbau eines Random Forrest

Benötigt werden zwei Trainingssets von Daten:

- a) Gammaevents (γ) (aus MC-Simulationen) und
- b) Hadronen (h) (Off-Data)

Jedes Event wird klassifiziert durch einen Vektor $v(P)$ dessen Komponenten die (Hillas-) Parameter sind. Zusätzlich wird jedem Trainingsevent eine Zahl (Hadroness) zugeordnet (0 oder 1)

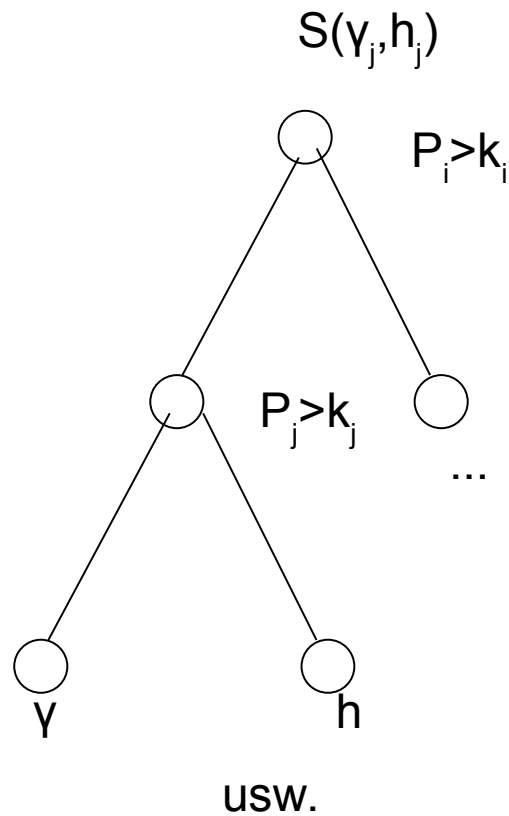


Prinzipieller Aufbau eines Random Forrest

Benötigt werden zwei Trainingssets von Daten:

- a) Gammaevents (γ) (aus MC-Simulationen) und
- b) Hadronen (h) (Off-Data)

Jedes Event wird klassifiziert durch einen Vektor $v(P)$ dessen Komponenten die (Hillas-) Parameter sind. Zusätzlich wird jedem Trainingsevent eine Zahl (Hadroness) zugeordnet (0 oder 1)



Hadroness:

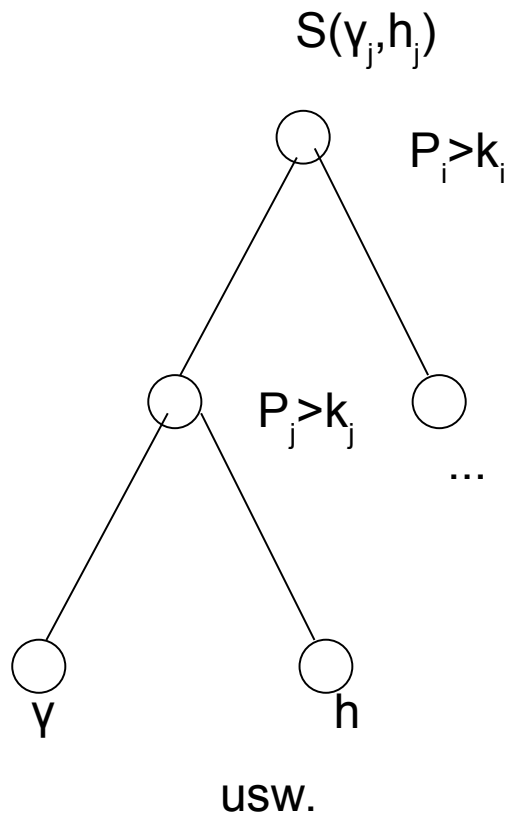
$$h(v) = \frac{\sum_{i=1}^{n_{trees}} l_i(v)}{n_{trees}}$$

Prinzipieller Aufbau eines Random Forrest

Benötigt werden zwei Trainingssets von Daten:

- a) Gammaevents (γ) (aus MC-Simulationen) und
- b) Hadronen (h) (Off-Data)

Jedes Event wird klassifiziert durch einen Vektor $v(P)$ dessen Komponenten die (Hillas-) Parameter sind. Zusätzlich wird jedem Trainingsevent eine Zahl (Hadroness) zugeordnet (0 oder 1)



Hadroness:

$$h(v) = \frac{\sum_{i=1}^{n_{trees}} l_i(v)}{n_{trees}}$$

Gini - Index:

$$Q_{Gini} = 2 \cdot \left(\frac{N_{glinks} \cdot N_{hlinks}}{N_{links} \cdot N_{links}} + \frac{N_{grechts} \cdot N_{hrechts}}{N_{rechts} \cdot N_{rechts}} \right)$$

Ein genauerer Einblick in OSTERIA:

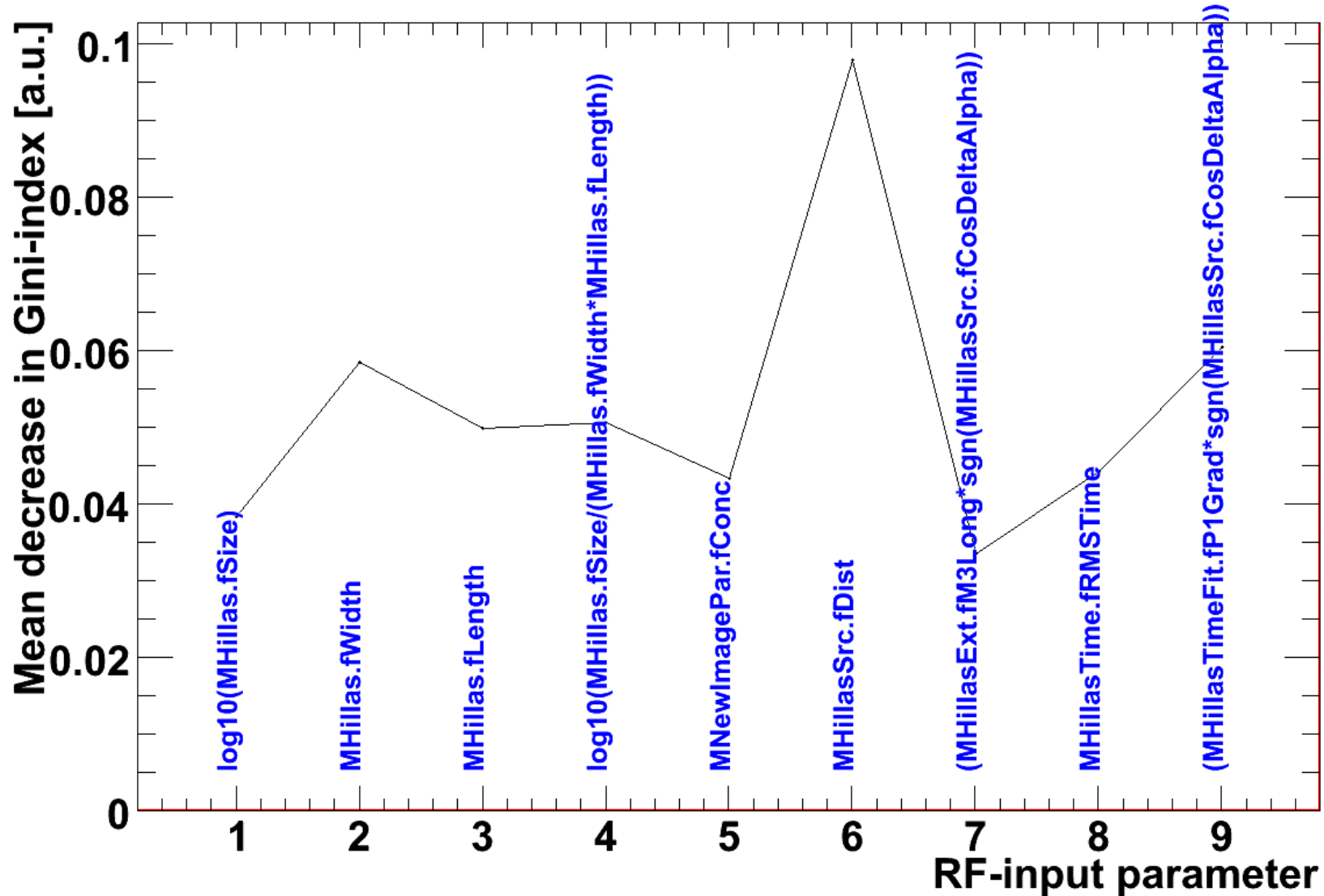
Einstellmöglichkeiten:

- Anzahl der Bäume für den Random Forest
- Unphysikalische Parameter ausschließen, z.B. : $Width < 0$
- Einfache Filter cuts, z.B.: $NumIslands > 3$
- Verhältnis von Hadronen zu Gammas der Trainingsevents ändern
(Standardeinstellung: 1)
- Die Parameter zur Separation einstellen:

→ Size	→ Dist
→ Width	→ $M3Long * \text{sgn}(\text{CosDeltaAlpha})$
→ Length	→ RMSTime
→ $Size / (Width * Length)$	→ $P1Grad * \text{sgn}(f\text{CosDeltaAlpha})$

Ergebnisse von OSTERIA: z.B.: In Form des Plots für den mean Gini decrease

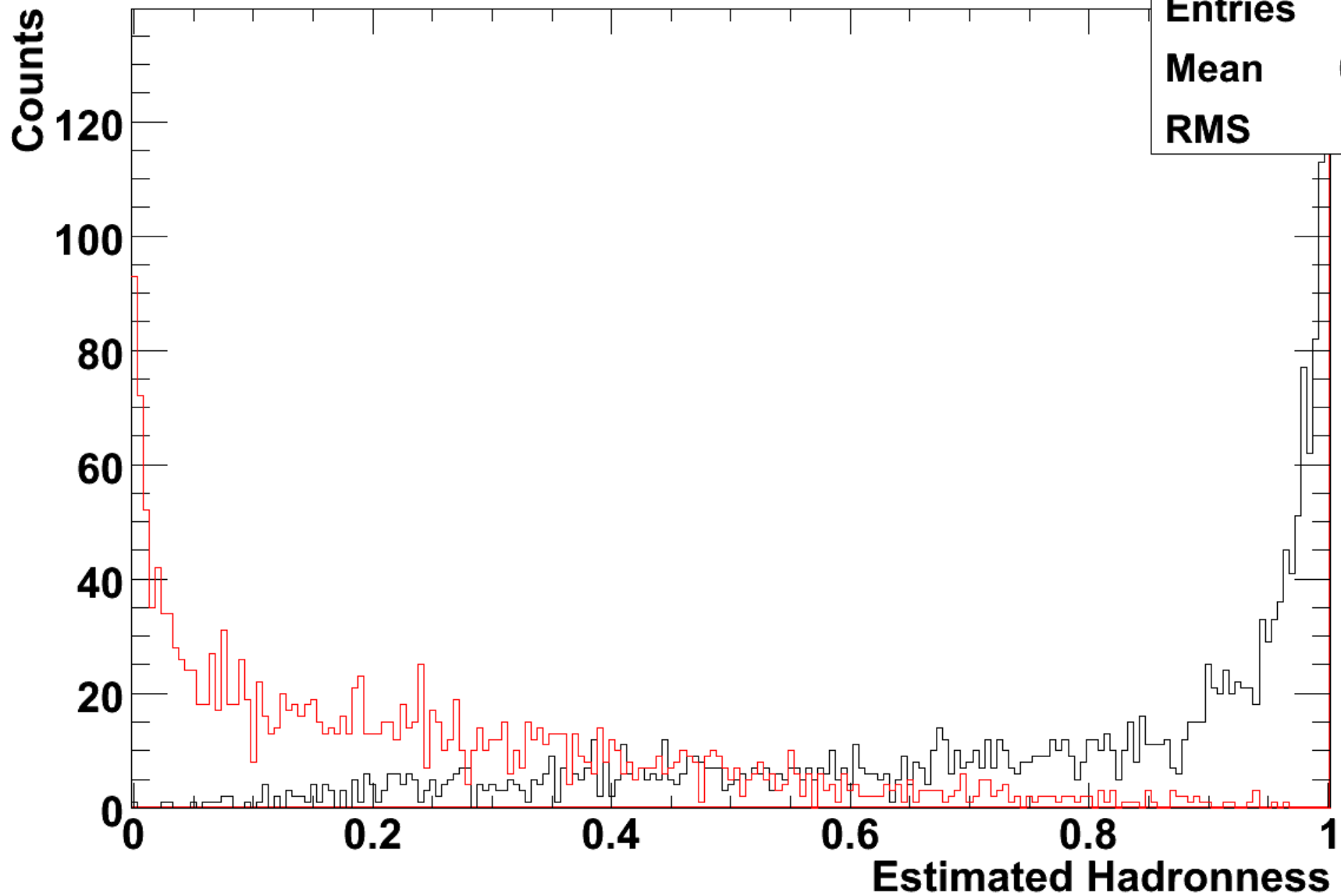
Importance of RF-input parameters measured by mean Gini decrease



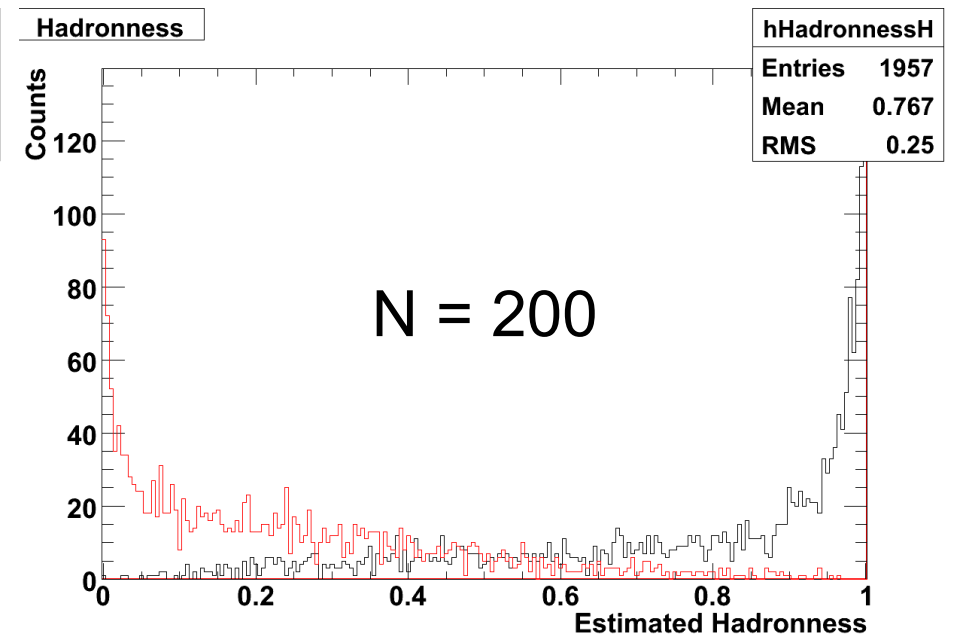
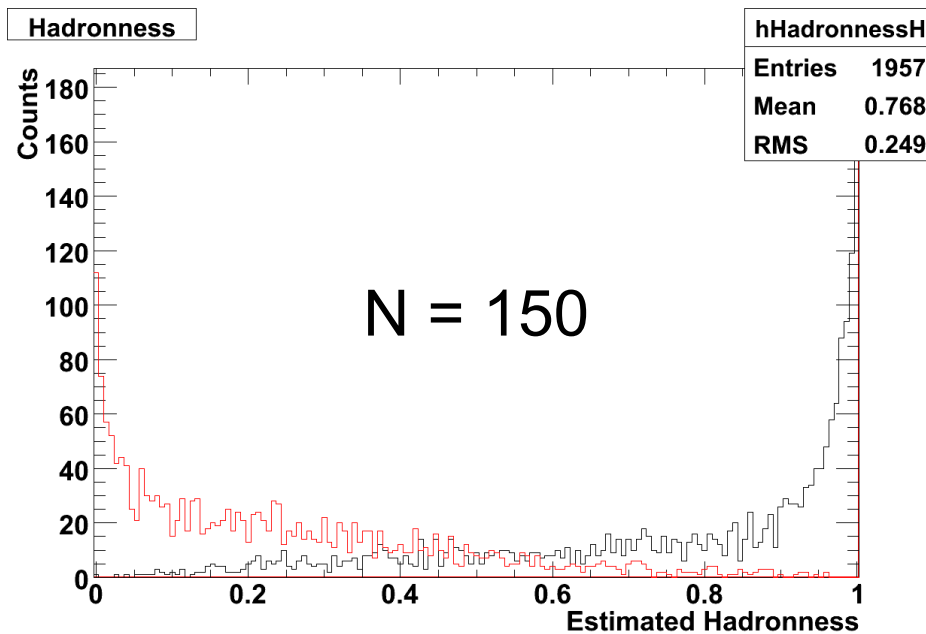
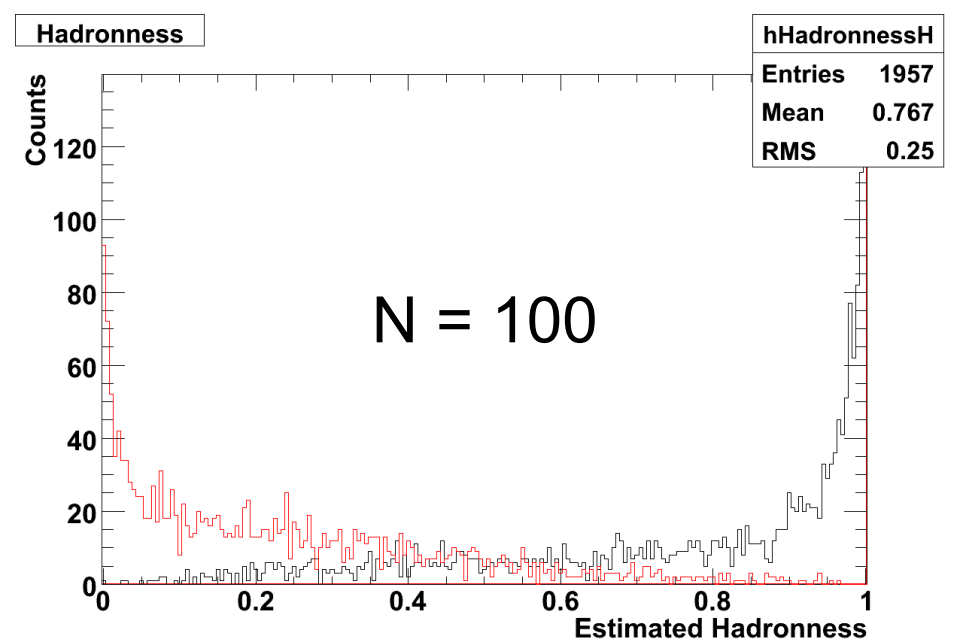
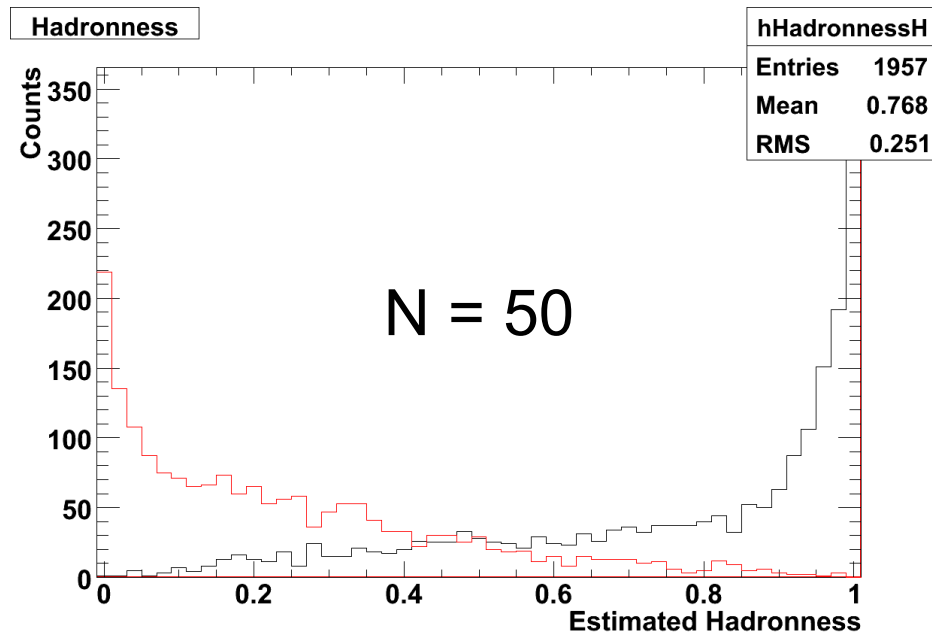
bzw..

Hadronness

hHadronnessH	
Entries	1957
Mean	0.767
RMS	0.25



Was bewirkt jetzt eine Änderung in, z.B., der Anzahl der Bäume?



Aber...

- Was ist denn die „beste“ Einstellung?
Gilt diese Einstellung für jede Quelle?

Aber...

- Was ist denn die „beste“ Einstellung?
Gilt diese Einstellung für jede Quelle?
- Ist ein Random Forest überhaupt am besten geeignet, um die Gamma/Hadronen Separation durchzuführen

Aber...

- Was ist denn die „beste“ Einstellung?
Gilt diese Einstellung für jede Quelle?
- Ist ein Random Forest überhaupt am besten geeignet, um die Gamma/Hadronen Separation durchzuführen
- Welche Parameter sind am besten geeignet zur Separation?
Oder besser: Welche Kombination der Parameter liefert das beste Ergebnis?

Aber...

- Was ist denn die „beste“ Einstellung?
Gilt diese Einstellung für jede Quelle?
- Ist ein Random Forest überhaupt am besten geeignet, um die Gamma/Hadronen Separation durchzuführen?
- Welche Parameter sind am besten geeignet zur Separation?
Oder besser: Welche Kombination der Parameter liefert das beste Ergebnis?



Welche neuen Erkenntnisse können sich daraus ergeben?

Aber...

- Was ist denn die „beste“ Einstellung?
Gilt diese Einstellung für jede Quelle?
- Ist ein Random Forest überhaupt am besten geeignet, um die Gamma/Hadronen Separation durchzuführen
- Welche Parameter sind am besten geeignet zur Separation?
Oder besser: Welche Kombination der Parameter liefert das beste Ergebnis?



Welche neuen Erkenntnisse können sich daraus ergeben?

Problem: Die Einstellungen können nur durch ausprobieren oder erhöhten Programmieraufwand gefunden werden.

Lösung: Ein externes Programm, das alle gewünschten
Eigenschaften schon hat



Lösung: Ein externes Programm, das alle gewünschten Eigenschaften schon hat



Ab 2001 wurde YALE am Lehrstuhl für künstliche Intelligenz unter der Führung von Prof. K. Morik im Fachbereich Informatik der TU – Dortmund entwickelt.

Seit 2007 wird die Software von der Firma Rapid I unter dem Namen RapidMiner vertrieben.

Es existiert eine Community Edition (free) und eine Enterprise Edition für den kommerziellen Gebrauch.

Lösung: Ein externes Programm, das alle gewünschten Eigenschaften schon hat



Ab 2001 wurde YALE am Lehrstuhl für künstliche Intelligenz unter der Führung von Prof. K. Morik im Fachbereich Informatik der TU – Dortmund entwickelt.

Seit 2007 wird die Software von der Firma Rapid I unter dem Namen RapidMiner vertrieben.

Es existiert eine Community Edition (free) und eine Enterprise Edition für den kommerziellen Gebrauch.

Das Programm ist Operator gesteuert, d.h. recht intuitiv zu begreifen.

Geschrieben wurde es in Java, d.h. es ist Plattformunabhängig

Es existieren schon die benötigten Verfahren (Random Forest, neuronale Netze, SVM etc.)

CALLISTO



STAR



OSTERIA



MELIBEA



FLUXLC



UNFOLDING

CALLISTO



STAR



Übersetzungsskript
ROOT in XML
(M. Schwacke)



OSTERIA



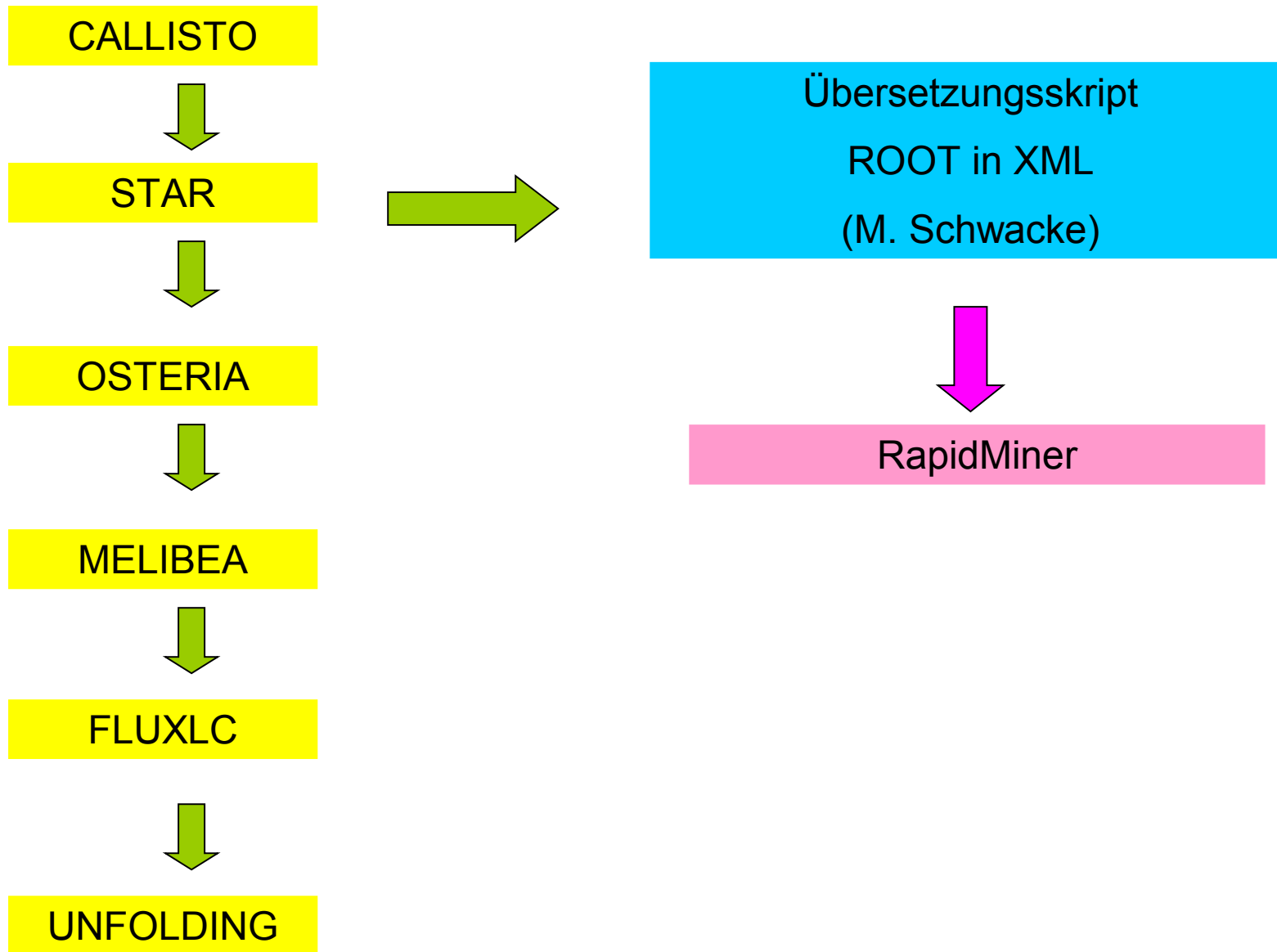
MELIBEA

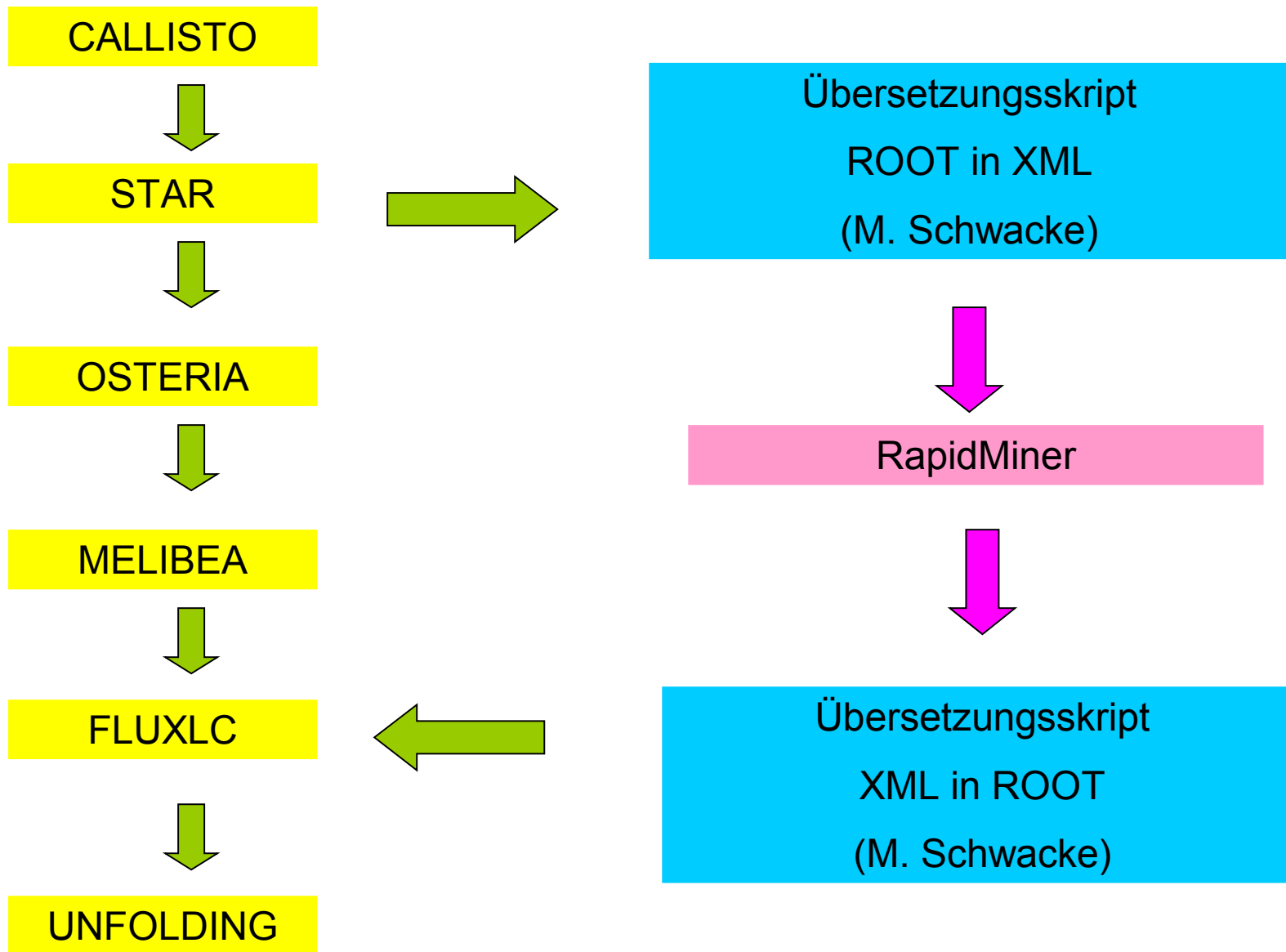


FLUXLC



UNFOLDING





Ziele:

Ziele:

- Verbesserung der Separation, d.h.: reinere Gamma – Daten und damit eine bessere Energiebestimmung
- Ein Vergleich der verschiedenen statistischen Verfahren (Benchmarking)
- Auffinden von neuen Parameterkombinationen zur Separation durch den Einsatz von Evolutionären Algorithmen

Ziele:

- Verbesserung der Separation, d.h.: reinere Gamma – Daten und damit eine bessere Energiebestimmung
- Ein Vergleich der verschiedenen statistischen Verfahren (Benchmarking)
- Auffinden von neuen Parameterkombinationen zur Separation durch den Einsatz von Evolutionären Algorithmen

Weiter: Die Methode ist natürlich nicht nur auf MAGIC beschränkt!

In Arbeit ist der Einsatz bei IceCube und LHCb



Was macht man im Stereoskopischen Modus?



Was macht man im Stereoskopischen Modus?

Dasselbe!



MAGIC:

<http://wwwmagic.mppmu.mpg.de>

Allgemeines zum
Random Forest

<http://oz.berkeley.edu/users/breiman/>

RapidMiner

<http://rapid-i.com/>