

**TECHNISCHE
UNIVERSITÄT
DRESDEN**

Nachweis hadronisch zerfallender Tau-Leptonen mit multivariaten Methoden im ATLAS-Detektor am LHC

Bachelor-Arbeit
zur Erlangung des Hochschulgrades
Bachelor of Science
im Bachelor-Studiengang Physik

vorgelegt von

Tony Henseleit
geboren am 06.09.1993 in Dessau-Roßlau

Institut für Kern- und Teilchenphysik
Fachrichtung Physik
Fakultät Mathematik und Naturwissenschaften
Technische Universität Dresden
2015

Eingereicht am 11. Juni 2015

1. Gutachter: Prof. Dr. Straessner
2. Gutachter: Prof. Dr. Zuber

Zusammenfassung

Zusammenfassung

Deutsch:

Diese Arbeit befasst sich mit der Verbesserung der Identifikation hadronisch zerfallender τ -Leptonen durch multivariate Methoden aus den rekonstruierten Daten der Messungen von Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 8 TeV, mittels des ATLAS-Detektors am LHC im Jahr 2012. Dabei wird durch Einführung neuer Variablen, die mit dem *CellBased*-Algorithmus berechnet wurden, versucht, die Identifikation mit neuronalen Netzen im Vergleich zu *Boosted Decision Trees* zu verbessern. Weiterhin wird auch der Einfluss dieser Variablen auf die Identifikation mit *Boosted Decision Trees* überprüft. Da bei neuronalen Netzen, neben Art und Anzahl der verwendeten Variablen, auch die Netzwerkarchitektur eine entscheidende Rolle spielt, wird auch diese an den Satz von Variablen angepasst um ein optimales Ergebnis zu erzielen.

Abstract

English:

This thesis deals with the improvement of the identification of hadronically decaying τ -leptons using multivariate methods on reconstructed data, collected from proton-proton-collisions at a center-of-mass energy of 8 TeV by the ATLAS-detector at the LHC in 2012. By introducing new variables, reconstructed based on the *CellBased*-algorithm, an attempt is made to improve the identification with neural networks compared to *Boosted Decision Trees*. Furthermore the influence of these variables on the identification with *Boosted Decision Trees* is checked. Because for neural networks, besides the type of variables and the number of variables, the network architecture plays a crucial role, the architecture is adjusted to the new set of variables in order to achieve best results.

Inhaltsverzeichnis

1	Einleitung	6
2	Das ATLAS-Experiment am <i>Large Hadron Collider</i>	7
3	Theoretische Grundlagen	9
3.1	Entstehung von τ -Leptonen	9
3.2	Standardvariablen	10
3.3	Neuronale Netzwerke	12
3.4	Boosted Decision Trees	16
4	Auswertung	18
4.1	Erweiterte Variablen	18
4.2	Variablenkonfigurationen und ihr Einfluss auf die Identifikation mit MLPs	23
4.2.1	Variation der Variablenkonfiguration	23
4.2.2	Variation der Netzwerkarchitektur	25
4.2.3	Variation der Anfangswerte	27
4.2.4	Fazit	28
4.3	Variablenkonfigurationen und ihr Einfluss auf die Identifikation mit BDTs	30
5	Zusammenfassung der Ergebnisse und Ausblick	34
A	Diagramme	35
A.1	Verteilung der Standardvariablen	35
A.2	Verteilung der erweiterten Variablen	38
B	Literatur	50
	Abbildungsverzeichnis	52
	Tabellenverzeichnis	53

1 Einleitung

Bei τ -Leptonen[1][2] handelt es sich um Teilchen, die Teil des Standardmodells der Teilchenphysik sind, welches in Abb. 1 (S. 6) dargestellt wird. Sie sind der dritten Generation von Fermionen zugeordnet. Damit handelt es sich bei τ -Leptonen um Spin- $\frac{1}{2}$ -Teilchen mit einer einfachen Elementarladung. Aufgrund ihrer im Vergleich zu den anderen Leptonen sehr hohen Masse von $m = 1,777 \text{ GeV}$ sind sie zudem sehr gut geeignet für die Untersuchung physikalischer Phänomene, die in Zusammenhang mit dem Higgs-Boson stehen. Dies ist dadurch begründet, dass das Higgs-Boson über den Higgs-Mechanismus[3][4][5] an die Masse eines Teilchens koppelt. Damit ist die Wahrscheinlichkeit

Drei Generationen der Materie (Fermionen)					
	I	II	III		
Masse	2,3 MeV	1,275 GeV	173,07 GeV	0	125,9 GeV
Ladung	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	0
Spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
Name	u up	c charm	t top	γ Photon	H Higgs Boson
Quarks	4,8 MeV $-\frac{1}{3}$ $\frac{1}{2}$ d down	95 MeV $-\frac{1}{3}$ $\frac{1}{2}$ s strange	4,18 GeV $-\frac{1}{3}$ $\frac{1}{2}$ b bottom	0 0 1 g Gluon	
	<2 eV 0 $\frac{1}{2}$ ν_e Elektron-Neutrino	<0,19 MeV 0 $\frac{1}{2}$ ν_μ Myon-Neutrino	<18,2 MeV 0 $\frac{1}{2}$ ν_τ Tau-Neutrino	91,2 GeV 0 1 Z^0 Z Boson	
	0,511 MeV -1 $\frac{1}{2}$ e Elektron	105,7 MeV -1 $\frac{1}{2}$ μ Myon	1,777 GeV -1 $\frac{1}{2}$ τ Tau	80,4 GeV ± 1 1 W^\pm W Boson	Eichbosonen

Abb. 1: Das Standardmodell der Elementarteilchen¹

des Zerfalls eines Higgs-Bosons in τ -Leptonen um ein Vielfaches höher als in Myonen oder gar Elektronen. Daher ist die Identifikation von τ -Leptonen von essentieller Bedeutung für die Forschung im Bereich des Standardmodells, aber auch darüber hinaus. Ihre kurze mittlere Lebensdauer von nur $\Delta t = 2,906 \cdot 10^{-13} \text{ s}$ erschwert deren Detektion. Sie zerfallen schon vor dem Erreichen aktiver Detektorregionen und sind daher nur über ihre Zerfallsprodukte, das heißt über die von ihnen erzeugten Teilchenjets, identifizierbar. Zu diesem Zweck verwendet man sogenannte multivariate Methoden. Diese ermöglichen die gleichzeitige Betrachtung aller in die Entscheidungsfindung einbezogenen Variablen. Dadurch wird der Identifikationsprozess wesentlich besser, als es bei der aufeinander folgenden Betrachtung der Variablen möglich wäre. Die Standardmethode, welche zu diesem Zweck verwendet wird, sind sogenannte *Boosted Decision Trees*, kurz BDTs. Eine weitere Methode stellen *Multilayer Perceptrons*, kurz MLPs, dar. Dabei handelt es sich um künstliche neuronale Netzwerke. In den Kapiteln 3.3 und 3.4 werden diese Methoden genauer beschrieben.

¹Bild: „Standard model of elementary particles: the 12 fundamental fermions and 5 fundamental bosons.“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 27. August 2010, 16:34 UTC. URL: http://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles-de.svg (Abgerufen: 07. Mai 2015)

2 Das ATLAS-Experiment am *Large Hadron Collider*

Bei dem *Large Hadron Collider*, kurz LHC[6], handelt es sich um den großen Hadronenbeschleuniger des europäischen Kernforschungszentrums *European Organization for Nuclear Research*, kurz CERN, bei Genf. Mit einem Umfang von fast 27 km handelt es sich dabei um den weltgrößten Teilchenbeschleuniger. Einer der verbauten großen Detektoren ist *A Toroidal LHC Apparatus*², kurz ATLAS[7]. Dieser ist speziell auf den Nachweis des Higgs-Bosons ausgelegt und wird in Abb. 2 (S. 7) im Querschnitt dargestellt.

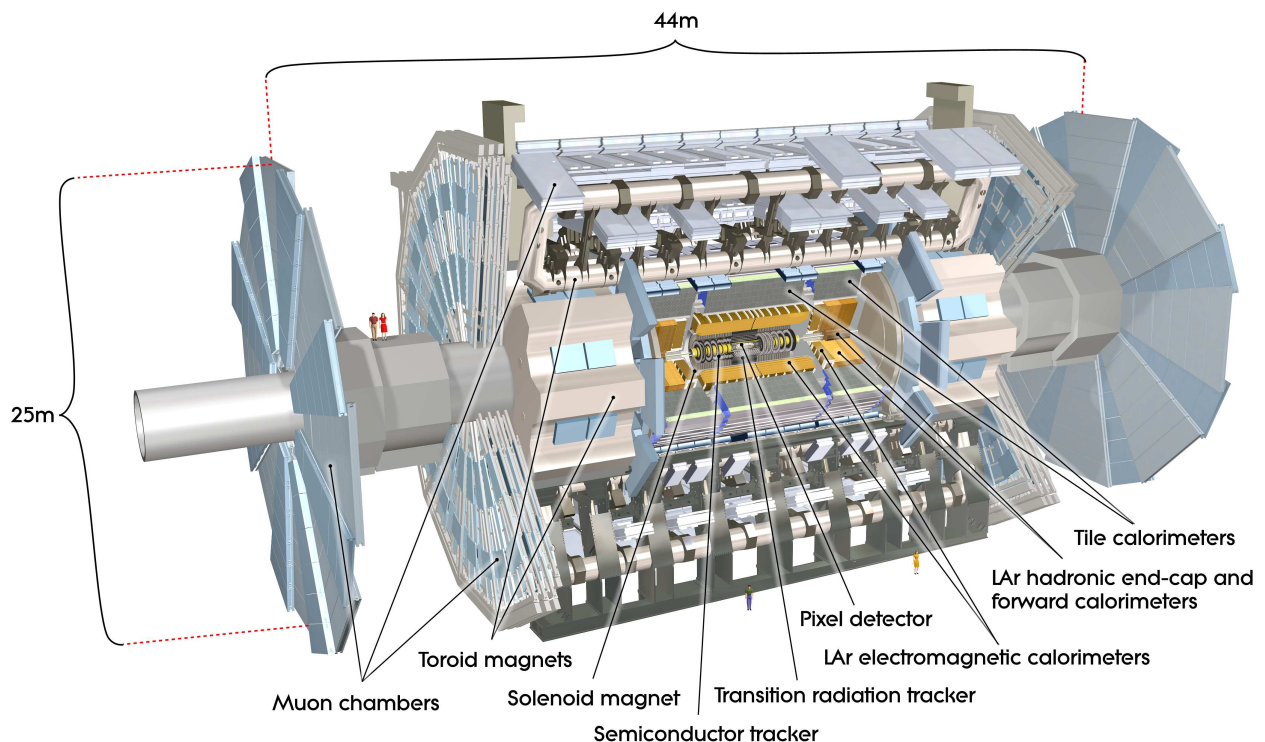


Abb. 2: Querschnitt des ATLAS-Detektors²

Der Detektor besteht aus drei Teilen. Von innen, dem Kollisionspunkt aus, nach außen sind das der innere Detektor, das Kalorimeter und das Myonenspektrometer. Der innere Detektor ist ein Spurdetektor. Durch ein Magnetfeld, erzeugt durch einen Solenoidmagneten, von bis zu zwei Tesla wird die Bahn der geladenen Teilchen gekrümmt. Diese Krümmung ist eine Folge der auf bewegte geladene Teilchen in einem Magnetfeld wirkenden Lorentzkraft. Mit Hilfe des Pixeldetektors, des Silizium-Streifen-Detektors, kurz SCT, und des Übergangsstrahlungsspurdetektors, kurz TRT, werden die Spuren der geladenen Teilchen ermittelt. Diese drei Detektoren sind in der genannten Reihenfolge von innen nach außen angeordnet und haben dementsprechend eine abnehmende Auflösung mit zunehmendem Abstand zum Wechselwirkungspunkt. Während der Pixeldetektor eine Genauigkeit bei der Spurbestimmung von $14 \mu\text{m}$

²Bild: ATLAS Collaboration: „The ATLAS Experiment at the CERN Large Hadron Collider“. JINST 3 (2008) S08003, Abbildung 1.1, S. 4.

erreicht, erreicht der SCT lediglich $30\ \mu\text{m}$ und der TRT 150 mm. Dem inneren Detektor schließen sich nach außen das elektromagnetische, kurz ECAL, und das hadronische Kalorimeter, kurz HCAL, an. Im ECAL, welches aus mit Edelstahl beschichteten Bleiplatten und flüssigem Argon besteht, wird vorrangig die Energie von Elektronen und Photonen deponiert und gemessen. Im anschließenden HCAL wird die Energie der Hadronen deponiert und gemessen, da diese im ECAL nur einen kleinen Teil ihrer Energie deponieren. Um eine entsprechende Wechselwirkung der Hadronen mit dem Absorbermaterial zu gewährleisten, ist dieser Teil des Detektors wesentlich größer als die bisherigen Detektorbestandteile. Er besteht aus abwechselnd angeordneten Eisenabsorbern und Plastiksintillatoren sowie einem Absorber aus flüssigem Argon, Kupfer und Wolfram. Der äußerste und größte Teil des Detektors ist das Myonenspektrometer. Durch ein torodiales Luftspulenmagnetsystem wird darin ein Magnetfeld zwischen drei und acht Tesla erzeugt. Da die Myonen bis zu diesem Punkt kaum Wechselwirkungen eingegangen sind, ist dieser Teil des Detektors in Verbindung mit einem eigenständigen Triggersystem und dem Spurdetektor im inneren Bereich notwendig, um die Myonen nachweisen zu können.

3 Theoretische Grundlagen

3.1 Entstehung von τ -Leptonen

Die τ -Leptonen, welche Gegenstand dieser Arbeit sind, werden indirekt durch die Proton-Proton-Kollision[8] erzeugt, welche am *Large Hadron Collider* bei einer Schwerpunktsenergie von 8 TeV hervorgerufen wird. Die bei der Proton-Proton-Kollision entstehenden W- und Z-Bosonen zerfallen unter anderem unter der Bildung von τ -Leptonen. Diese zerfallen wiederum vor dem Erreichen der aktiven Regionen des ATLAS-Detektors. Dabei werden hadronische und leptonische Zerfälle unterschieden. Im Folgenden werden die wichtigsten Zerfallsmoden[9] der τ -Leptonen dargestellt:

$$\begin{aligned}
 \text{Leptonische Zerfallsmoden: } & \tau^\pm \rightarrow e^\pm + \nu_e/\bar{\nu}_e + \bar{\nu}_\tau/\nu_\tau \\
 & \tau^\pm \rightarrow \mu^\pm + \nu_\mu/\bar{\nu}_\mu + \bar{\nu}_\tau/\nu_\tau \\
 \text{Hadronische Zerfallsmoden: } & \tau^\pm \rightarrow \pi^\pm + \bar{\nu}_\tau/\nu_\tau \\
 & \tau^\pm \rightarrow \pi^\pm + \pi^0 + \bar{\nu}_\tau/\nu_\tau \\
 & \tau^\pm \rightarrow \pi^\pm + 2\pi^0 + \bar{\nu}_\tau/\nu_\tau \\
 & \tau^\pm \rightarrow \pi^\pm + \pi^\pm + \pi^\mp + \bar{\nu}_\tau/\nu_\tau \\
 & \tau^\pm \rightarrow \pi^\pm + \pi^\pm + \pi^\mp + \pi^0 + \bar{\nu}_\tau/\nu_\tau
 \end{aligned}$$

Die leptonischen Zerfälle bilden den geringeren Anteil, wobei jeweils ein leichteres Lepton, das zugehörige Neutrino und ein τ -Neutrino entstehen. In 68% der Fälle zerfällt ein τ -Lepton jedoch hadronisch. Dabei entstehen Quark-Antiquark-Paare und ein τ -Neutrino. Diese Paare bilden wiederum Hadronen. Daher entstehen bei hadronischen τ -Leptonen-Zerfällen meist ein oder drei geladene Pionen und kein, ein oder mehrere neutrale Pionen. Weiterhin sind auch Zerfälle in Kaonen möglich, die aufgrund ihrer geringen Wahrscheinlichkeit im Folgenden vernachlässigt werden. Da diese Vorgänge noch vor dem Erreichen der aktiven Regionen des Detektors ablaufen, bezeichnet man die hierbei entstehenden Teilchen als sichtbare Zerfallsprodukte des τ -Leptons, kurz $\tau_{\text{had-vis}}$. Die Bildung hadronischer Teilchenschauer tritt sowohl bei $\tau_{\text{had-vis}}$ -Teilchen, als auch bei Hadronen aus anderen Quellen auf. Somit besteht die Notwendigkeit, die von τ -Leptonen stammenden Hadronen von jenen, welche durch andere Prozesse entstehen, zu unterscheiden.

Da die Jets der Hadronen, je nach Ursprung, andere Charakteristika besitzen, ist es möglich, mittels bestimmter Algorithmen eine Punktzahl, den sogenannten *Score*, zu ermitteln, der korreliert ist, mit der Wahrscheinlichkeit, ob es sich bei einem bestimmten Kandidaten um ein τ -Lepton handelt oder nicht. Um dies zu ermöglichen, ist es jedoch zunächst notwendig,

aus den Daten des Detektors, die Teilchen-Jets und demzufolge auch die Kandidaten für die Identifikation zu rekonstruieren. Schon bei diesem Prozess werden gewisse Anforderungen an die τ -Kandidaten gestellt[8]. Aus den rekonstruierten Jets werden dann die Identifikationsvariablen ermittelt. In dieser Arbeit wird dabei ein spezielles Augenmerk auf dem erweiterten Variablensatz liegen. Bei diesen Variablen ist es von besonderer Bedeutung, wie sie sich jeweils in festgelegten Bereichen, der sogenannten Zentral-, Kern- und Isolationsregion, verhalten oder verändern.

3.2 Standardvariablen

Die folgende Übersicht führt die Variablen auf, welche standardmäßig für die $\tau_{\text{had-vis}}$ -Identifikation verwendet werden. Da bei der Identifikation die Ereignisse mit einem bzw. drei entstehenden geladenen Pionen getrennt betrachtet werden, im Folgenden als 1-Prong- und 3-Prong-Ereignisse bezeichnet, sind auch einige dieser Variablen prong-spezifisch. Dieser Variablensatz hat sich speziell für die Identifikation mittels BDTs als sehr effizient erwiesen[9].

Übersicht der Identifikationsvariablen:

$$f_{\text{core}}^{\text{cent}} := \frac{\sum_{\Delta R_i < 0.1} E_{\text{T},i}^{\text{EM}}}{\sum_{\Delta R_j < 0.2} E_{\text{T},j}^{\text{EM}}},$$

beschreibt den Anteil der in der Zentralregion deponierten elektromagnetischen Transversalenergie an der im Kerngebiet deponierten elektromagnetischen Transversalenergie.

$$f_{\text{track}} := \frac{p_{\text{T,max}}^{\text{track}}}{E_{\text{T}}^{\tau}} d,$$

beschreibt den Anteil des Transversalimpulses der impulsreichsten Spur an der gesamten Transversalenergie.

$$R_{\text{track}} := \frac{\sum_{\Delta R_i < 0.4} p_{\text{T},i} \cdot \Delta R_i^*}{\sum_{\Delta R_j < 0.4} p_{\text{T},j}},$$

beschreibt den mit dem Transversalimpuls gewichteten Abstand aller Spuren zur τ -Achse des $\tau_{\text{had-vis}}$ -Kandidaten.

* $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ mit $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$ wobei $\eta \hat{=}$ Pseudorapidität, $\phi \hat{=}$ Winkelabstand in Transversalebene, $\theta \hat{=}$ Winkelabstand zur Transversalebene

1-Prong-Variablen:

Die folgenden Variablen werden nur zur Identifikation bei 1-Prong-Ereignissen verwendet.

$$S_{\text{leadtrack}}^{\text{IP}} := \frac{d_0}{\delta d_0},$$

beschreibt den Stoßparameter der impulsreichsten Spur d_0 , gewichtet mit dessen Unsicherheit.

$$N_{\text{track}}^{\text{iso}},$$

beschreibt die Anzahl der Spuren im Bereich von $0.2 < \Delta R < 0.4$.

3-Prong-Variablen:

Die folgenden Variablen werden nur zur Identifikation bei 3-Prong-Ereignissen verwendet.

$$m_{\text{tracks}} := \sqrt{\left(\sum_{i \in \{\text{tracks}\}}^{\Delta R_i < 0.4} E_i \right)^2 - \left(\sum_{i \in \{\text{tracks}\}}^{\Delta R_i < 0.4} \vec{p}_i \right)^2},$$

beschreibt die Gesamtmasse aller Spuren zu einem $\tau_{\text{had-vis}}$ -Kandidaten.

$$S_{\text{T}}^{\text{flight}} := \frac{L_{\text{T}}^{\text{flight}}}{\delta L_{\text{T}}^{\text{flight}}},$$

beschreibt die Zerfallslänge des Sekundärvertex $L_{\text{T}}^{\text{flight}}$ gewichtet mit dessen Unsicherheit. Daher ist diese Größe für 1-Prong-Ereignisse nicht definiert.

$$\Delta R_{\text{max}}^{\text{core}},$$

beschreibt die Spur mit maximalem ΔR in der Kernregion, also für $\Delta R < 0.2$.

 π^0 -Variablen:

$$f_{\text{vis-}p_{\text{T}}} := \frac{\sum_{i \in \{\text{tracks}, \pi^0\}}^{\Delta R_i < 0.2} p_{\text{T},i}}{p_{\text{T}}^{\tau}},$$

beschreibt den Anteil des Transversalimpulses aller π^0 -Teilchen am Gesamtimpuls des $\tau_{\text{had-vis}}$ -Kandidaten.

N_{π^0} ,

beschreibt die Anzahl der zugehörigen π^0 -Teilchen eines $\tau_{\text{had-vis}}$ -Kandidaten.

$$m_{\text{T}}^{\text{vis}} := \sqrt{\left(\sum_{i \in \{\text{tracks}, \pi^0\}}^{\Delta R_i < 0.2} E_i \right)^2 - \left(\sum_{i \in \{\text{tracks}, \pi^0\}}^{\Delta R_i < 0.2} \vec{p}_i \right)^2},$$

beschreibt die Gesamtmasse aller π^0 -Teilchen eines $\tau_{\text{had-vis}}$ -Kandidaten.

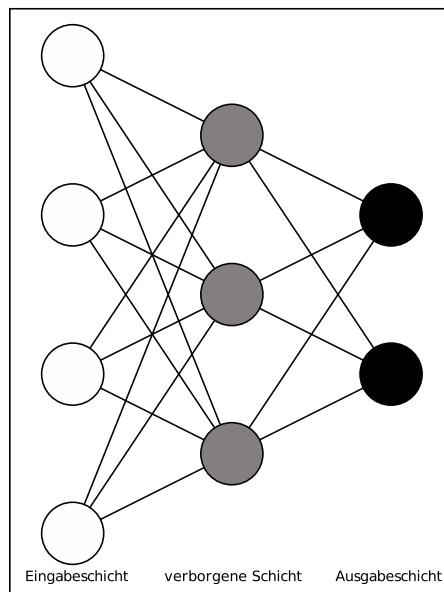
Die Verteilungen dieser Variablen für Signal und Untergrund, sowohl für 1-Prong- als auch für 3-Prong-Ereignisse, werden in Abb. 25-27 (S. 35-37) im Anhang A.1 dargestellt. Anhand dieser Histogramme kann die Separationskraft der jeweiligen Variable abgeschätzt werden.

3.3 Neuronale Netzwerke

Bei einem künstlichen neuronalen Netz[10] handelt es sich um eine Methode zur multivariaten Analyse von Daten. Es unterscheidet sich von anderen mathematischen Verfahren, wie der multiplen Regression dadurch, dass als biologisches Vorbild ursprünglich das menschliche Gehirn diente, wobei dieser Aspekt bei modernen Netzwerken nicht mehr von Bedeutung ist.

Der Aufbau neuronaler Netzwerke ist in Abb. 3 (S. 12) dargestellt. Sie bestehen aus einzelnen Neuronen, die in Schichten angeordnet sind.

Die Neuronen der ersten Schicht werden als Eingabeneuronen bezeichnet. An diese werden Informationen, wie die auszuwertenden Daten, übergeben. Die Neuronen der letzten Schicht, auch Ausgabeneuronen genannt, geben das gewünschte Ergebnis aus, welches anhand der Daten ermittelt wird. Zwischen diesen Schichten befindet sich eine beliebige Anzahl von Schichten aus verborgenen Neuronen. Diese sind von außen nicht sichtbar und ermöglichen eine nicht-lineare Transformation der Eingabedaten. Dabei gilt das *Universal Approximation Theorem*[11], wonach ein Netzwerk, dessen Informationsfluss lediglich in einer Richtung verläuft, ein sogenanntes *Feed-Forward-Network*, auch mit nur einer verborgenen Schicht mit



endlicher Neuronenanzahl jede kontinuierliche Funktion auf einem kompakten Unterraum des

³Erstellt mit LaTeXDraw

Abb. 3: Aufbau eines neuronalen Netzes³

\mathbb{R}^N approximieren kann, wenn die Aktivierungsfunktionen stetig, beschränkt und monotonsteigend sind. Ein weiterer grundlegender Bestandteil eines solchen Netzes sind die Verbindungsgewichte zwischen den einzelnen Neuronen. Diese beeinflussen die Stärke eines Signals zwischen zwei Neuronen entweder hemmend oder verstärkend. Besteht keine Verbindung, ist das Gewicht zwischen den jeweiligen Neuronen null. Die Gewichte eines neuronalen Netzes werden dabei durch ein Training mittels Testdaten mit bekanntem Ergebnis erzeugt.

Die Datenverarbeitung eines Neurons erfolgt in mehreren Schritten.[12] Aus den eingehenden Signalen wird, entsprechend einer Propagierungs- oder Kombinationsfunktion, die Eingabe ermittelt. Meist geschieht das in Form einer gewichteten Summe. Aus dieser wird, anhand der Aktivierungsfunktion, der Aktivitätszustand des Neurons bestimmt. Es gibt viele mögliche Aktivierungsfunktionen, darunter sind häufig Schwellenwertfunktionen wie logistische Funktionen oder die *Heaviside*-Funktion. Deren Schwellenwert kann dabei, im Zuge des Trainingsverfahrens durch Änderung der Gewichtung von Bias-Neuronen, Neuronen die stets ein konstantes Signal erzeugen, beliebig verschoben werden. Die Ausgabe des Neurons entspricht dem Ergebnis der Aktivierungsfunktion. Entsprechend der Wichtung der Verbindungen entstehen daraus die Signale, welche an folgende Neuronen übergeben werden.

Das Training neuronaler Netze ist somit unabdingbar für den Einsatz eines neuronalen Netzes. Dabei werden Datensätze verwendet, bei denen das gewünschte Ergebnis für das neuronale Netz bereits bekannt ist. Im Fall der $\tau_{\text{had-vis}}$ -Identifikation handelt es sich dabei um Monto-Carlo-Simulationen von τ -Leptonen-Zerfällen sowie Daten eines Durchlaufs von 2012 vom LHC, welche als Untergrund verwendet werden. Bei diesen experimentellen Daten handelt es sich um Daten von *Dijet*-Ereignissen, bei denen der Anteil an hochenergetischen τ -Kandidaten sehr gering ist. Für das Training stehen verschiedene Methoden zur Verfügung, wobei das Prinzip immer ähnlich ist:

Durch Vergleich des Ergebnisses des neuronalen Netzes mit dem bekannten Ergebnis wird die Abweichung ermittelt und die Gewichte werden angepasst, sodass es beim nächsten Durchlauf dem gewünschten Ergebnis näher kommt. Diese Modifikation kann einerseits *offline* bzw. stapelweise erfolgen, wobei mehrere Trainingsbeispiele durchlaufen werden bevor die Anpassung durchgeführt wird. Dies wird auch als *Batch*-Trainingsverfahren bezeichnet. Andererseits ist diese Veränderung auch *online* möglich, wobei nach jedem Beispiel direkt eine Veränderung vorgenommen wird. Die Modifikation der Gewichte stellt dabei eine komplizierte Aufgabe dar, für deren Lösung mehrere Algorithmen mit unterschiedlichen Vor- und Nachteilen existieren. Eine Möglichkeit stellt das Gradientenabstiegsverfahren dar. Ein solches Verfahren ermittelt, durch Gradientenbildung und schrittweises Bewegen in Richtung des größten negativen Anstiegs, lokale oder globale Minima n -dimensionaler Felder. Der absolute Fehlerterm des Feldes, häufig die summierte quadratische Abweichung der Ergebnisse des Netzwerks von den Vergleichswerten, in Abhängigkeit von den Gewichten des Netzes, stellt ein solches n -dimensionales

Feld dar, wobei n die Anzahl der Gewichte des Netzwerkes ist. Die Veränderung der Gewichte erfolgt durch Bestimmung der partiellen Ableitungen der Fehlerfunktion nach den Gewichten des neuronalen Netzes. Für die Gewichte der letzten Schicht ist die problemlos möglich, wird jedoch für die Gewichte der vorherigen Schichten erschwert, da deren Einfluss nicht direkt sichtbar ist. Die Ermittlung der partiellen Ableitungen erfolgt daher mit dem *Backpropagation of Error*-Algorithmus[13], welcher rekursiv die partiellen Ableitungen einer Schicht aus denen der vorherigen bestimmt.

Durch Anwendung des Gradientenabstiegsverfahrens auf eine solches Feld, können also die Gewichte für ein Netzwerk mit minimalen Abweichungen vom gewünschten Ergebnis ermittelt werden. Ein weiteres Verfahren, welches auch auf dem *Backpropagation of Error*-Algorithmus beruht, ist der *Broyden-Fletcher-Goldfarb-Shanno*-Algorithmus[14][15][16][17], kurz: BFGS, welcher zusätzlich die zweiten Ableitungen der Fehlerfunktion in die Berechnung der neuen Gewichte mit einbezieht. Dieser wird hier jedoch nicht weiter vertieft.

Entscheidende Nachteile dieser Methoden sind in Abb. 4 (S. 14) dargestellt. Ein grundlegendes Problem ist, dass häufig lokale, an Stelle globaler Minima ermittelt werden (I). Weiterhin kann das Verfahren bei ausgedehnten Ebenen mit geringem Anstieg einen langen Zeitraum arbeiten, ohne ein Minimum zu erreichen (II). Auch das Überspringen guter Minima, aufgrund der Proportionalität der Schrittweite

zum Betrag des Gradienten, ist ein häufiges Problem (III). In manchen Fällen erreicht das Verfahren, da es in eine Schleife gerät, auch nie ein Minimum (IV & V). Zuletzt ist noch zu Erwähnen, dass das Verfahren nicht zwischen Sattelpunkten und Minima unterscheidet (VI), weshalb es in solchen Fällen ein schlechtes Ergebnis liefern kann. Einen von vielen Lösungsansätzen stellt die Einbeziehung vorheriger Gradienten in die Bestimmung der Schrittweite dar. Dies wird auch als Gradientenabstiegsverfahren mit Trägheitsterm bezeichnet. Somit würde die „Nachwirkung“ des vorherigen Gradienten zu einer schnelleren Arbeitsweise im Bereich flacher Ebenen, zur Vermeidung von Schleifen und teilweise zur Vermeidung lo-

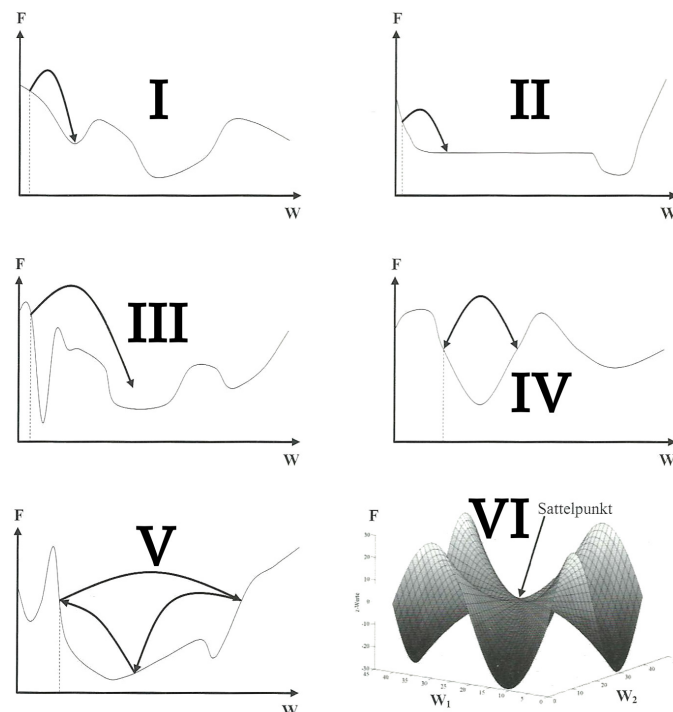


Abb. 4: Probleme des Gradientenverfahrens⁴

im Bereich flacher Ebenen, zur Vermeidung von Schleifen und teilweise zur Vermeidung lo-

⁴VGL.: Rey, Günther Daniel/Wender Karl F.: „Neuronale Netze“. S.45

kaler Minima und Sattelpunkte beitragen. Allerdings ist der Einfluss des Trägheitsterm nicht immer ausreichend um die Probleme des Gradientenabstiegsverfahrens auszugleichen.

Der Backpropagation-Algorithmus ist, wie schon zuvor erwähnt, eine der Möglichkeiten die Gewichte eines neuronalen Netzes anzupassen und wurde für die Netze in dieser Arbeit verwendet. Im Folgenden werden die Berechnungsgrundlagen an einem einfachen Beispiel erläutert.[18] Dabei handelt es sich um ein Netz mit nur einer verborgenen Schicht, dessen Neuronen die Aktivierungsfunktion $f(x) = \tanh(x)$ haben, und einer Ausgabeschicht mit linearer Aktivierungsfunktion. Damit ergibt sich die Ausgabe y_{ANN} des Netzes aus den Gewichten w_{ij} der Verbindungen und der Ausgabe der Eingabeneuronen x_i wie folgt:

$$y_{\text{ANN}} = \sum_{i=1}^{n_h} y_i w_{i1} = \sum_{i=1}^{n_h} \tanh \left(\sum_{j=1}^{n_{\text{var}}} x_j w_{ij} \right) w_{i1},$$

wobei n_{var} und n_h die Anzahl der Neuronen in der Ausgabeschicht bzw. in der verborgenen Schicht sind. Der Index $j = 1$ markiert dabei die Ausgabeschicht. Das neuronale Netz wird nun mit N Trainingsereignissen $x_a = x(x^1, x^2, \dots, x^{n_{\text{var}}})_a$ mit $a \in 1, \dots, N$ trainiert, deren gewünschte Ausgabe y_a bekannt ist. Damit definiert sich, mit der jeweiligen Netzausgabe $y_{\text{ANN},a}$ und dem aktuellen Satz von Gewichten w , die Fehlerfunktion E des aktuellen Netzes wie folgt:

$$E(x_1, \dots, x_N | w) = \frac{1}{N} \sum_{a=1}^N E(x_a | w) = \frac{1}{N} \sum_{a=1}^N \frac{1}{2} (y_{\text{ANN},a} - y_a)^2.$$

Anhand des Gradienten der Fehlerfunktion im Gewichtsraum $-\nabla_w E$ und der positiven Lernrate η wird, mittels Gradientenabstiegsverfahren, aus dem Satz von Gewichten $w^{(p)}$ der nächstfolgende Satz berechnet:

$$w^{(p+1)} = w^{(p)} - \eta \nabla_w E.$$

Explizit erfolgt dann die Veränderung der Gewichte folgendermaßen:

$$\begin{aligned} \Delta w_{i1} &= -\eta \sum_{a=1}^N \frac{\partial E_a}{\partial w_{i1}} = -\eta \sum_{a=1}^N (y_{\text{ANN},a} - y_a) y_{i,a} \\ \Delta w_{ij} &= -\eta \sum_{a=1}^N \frac{\partial E_a}{\partial w_{ij}} = -\eta \sum_{a=1}^N (y_{\text{ANN},a} - y_a) y_{i,a} (1 - y_{i,a}) w_{i1} x_{j,a}. \end{aligned}$$

Dabei wurde verwendet, dass $\tanh'(x) = \tanh(x)(1 - \tanh(x))$ ist. Dieses Verfahren wird wiederholt, bis der Wert der Fehlerfunktion annähernd gegen einen festen Wert konvergiert, das heißt es finden kaum noch Gewichtsänderungen statt. Dann ist das Training abgeschlossen.

Da in diesem Fall die Veränderung der Gewichte nach dem Training durch N Datensätze erfolgt ist, handelt es sich hierbei um das *stapelweise* Lernen. Für sogenanntes *online learning* müsste $N = 1$ gewählt werden. Dementsprechend würden die Gewichte nach jedem Ereignis modifiziert werden.

3.4 Boosted Decision Trees

Bei *Boosted Decision Trees* handelt es sich um *Decision Trees*[19], zu deutsch Entscheidungsbäume, die mittels *Boosting* stark verbessert werden. Für die $\tau_{\text{had-vis}}$ -Identifikation ist es derzeit die übliche Methode, da sie die besten Ergebnisse liefert.

Der Aufbau von *Decision Trees* ist relativ simpel. Jeder Knoten des Baumes repräsentiert einen Entscheidungspunkt, bei dem anhand von Schwellenwerten für das zu betrachtende Ereignis, über die nächsttiefere Stufe des Baumes entschieden wird. Dies führt entweder zum nächsten Knoten und damit zur nächsten Entscheidung oder zu einem *Leaf*, zu deutsch Blatt, welches das jeweilige Ereignis als Signal oder Untergrund identifiziert. Dieser Aufbau kann in Abb. 5 (S. 16) an einem sehr einfachen Beispiel nachvollzogen werden. Dabei soll anhand grundlegender Eigenschaften zwischen Protonen, Neutronen, Myonen, Elektronen und Positronen unterschieden werden. Es handelt sich bei diesem Beispiel nicht um die Grundlage eines Experiments, sondern um eine Verdeutlichung der ablaufenden Prozesse eines Entscheidungsbau- mes. Analog zur $\tau_{\text{had-vis}}$ -Identifikation sind die Unterscheidungen nicht immer eindeutig, das heißt es werden stets sowohl

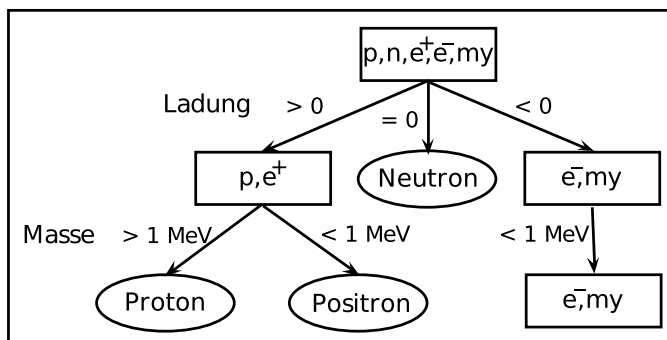


Abb. 5: Aufbau von Entscheidungsbäumen⁵

Untergrund- als auch Signalereignisse herausgefiltert. In der Abbildung wird weiterhin ersichtlich, dass manche Entscheidungen, wenn sie nicht den richtigen Schwellenwert haben, kein Ergebnis bzw. kein gutes Ergebnis liefern. Daher ist es notwendig auch *Decision Trees* vor ihrem Einsatz an die Problemstellung anzupassen. Im Beispiel müsste der Wert für die Masse so angepasst werden, dass auch eine Unterscheidung zwischen Elektronen und Myonen möglich ist, um die Identifikation zu verbessern. Für die praktische Anwendung ist das Ergebnis einfacher Entscheidungsbäume zu instabil unter der Veränderung des verwendeten Datensatzes.

Durch Boosting wird dieses Problem behoben.[20] Die Idee dabei ist, dass eine große Anzahl von Bäumen trainiert wird, deren gewichtetes Mittel gegenüber diesen Fluktuationen unempfindlich ist. Dazu wird für jedes Signalereignis, welches falsch klassifiziert wird, ein neuer Baum

⁵Erstellt mit LaTeXDraw

erstellt, bei dem besagtes Ereignis stärker gewichtet bzw. „geboostet“ wird. Dieser Prozess wird beliebig oft wiederholt. Wird das Ereignis im neuen Baum als Signal identifiziert, erhält der Baum das Gewicht $+1$, ansonsten -1 . Danach wird ein gewichtetes Mittel über alle Bäume gebildet, welches als Anhaltspunkt für die Güte des so entstandenen *Boosted Decision Tree* gilt. Für den Prozess des *Boosting* gibt es verschiedene Verfahren[21], darunter *Adaptive Boosting* oder auch *Gradient Boosting*, die allerdings im Folgenden nicht weiter erläutert werden.

4 Auswertung

Das Training der MLPs und BDTs, welche in dieser Arbeit ausgewertet werden, ist mit Hilfe des TMVA-Pakets[18] aus dem Analyse-Framework ROOT[22] erfolgt. Die Softwareumgebung und Daten wurden zur Verfügung gestellt.[23]

4.1 Erweiterte Variablen

Der erweiterte Variablensatz beinhaltet zumeist Abwandlungen der Standardvariablen, die sich durch die betrachteten Öffnungswinkel um die τ -Achse herum voneinander unterscheiden. Durch diese zusätzlichen Detailinformationen soll bei der Identifikation ein besseres Ergebnis erzielt werden. Es gibt dadurch auch Überschneidungen zwischen diesen und den Standardvariablen. Im Folgenden werden die bisher verwendeten Standardvariablen als „alte“ Variablen und die erweiterten Variablen als „neue“ Variablen bezeichnet.

Anzahl der Zerfallsprodukte:

$$N_{\text{chrg/neut/all}}^{\text{cent/core/iso}} \quad \text{mit} \quad \begin{cases} \text{cent} & \hat{=} \Delta R < 0.1 \\ \text{core} & \hat{=} \Delta R < 0.2 \\ \text{iso} & \hat{=} \Delta R < 0.4 \end{cases}$$

$$\quad \text{und} \quad \begin{cases} \text{chrg} & \hat{=} \text{nur geladene Zerfallsprodukte} \\ \text{neut} & \hat{=} \text{nur ungeladene Zerfallsprodukte} \\ \text{all} & \hat{=} \text{alle Zerfallsprodukte} \end{cases}$$

Dabei sind $N_{\text{all}}^{\text{iso}}$ und $N_{\text{track}}^{\text{iso}}$ sowie $N_{\text{neut}}^{\text{core}}$ und N_{π^0} gleich definiert. Die neu gewonnenen Detailinformationen sind am Beispiel von $N_{\text{chrg}}^{\text{cent}}$ und $N_{\text{chrg}}^{\text{core}}$ in Abb. 6 (S. 18) sowie Abb. 7 (S. 19) deutlich zu erkennen, auch wenn sie in diesem Fall für die Identifikation nicht hilfreich sind.

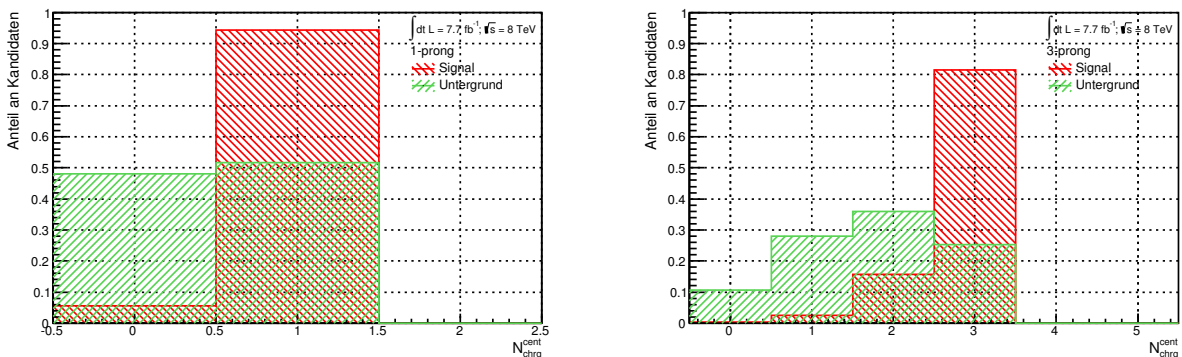


Abb. 6: Darstellung der Verteilungshistogramme der Variablen $N_{\text{chrg}}^{\text{cent}}$ von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts)

Während innerhalb der Zentralregion ein eindeutiger Unterschied zwischen Signal und Un-

tergrund sowohl für 1-, als auch für 3-Prong-Ereignisse besteht, ist die Anzahl der geladenen Teilchen innerhalb des Kerngebietes identisch. Einige der Variablen sind allerdings stark korreliert und weisen untereinander eine ähnliche Verteilung auf, weshalb in solchen Fällen die Fülle der Detailinformationen eher gering ist.

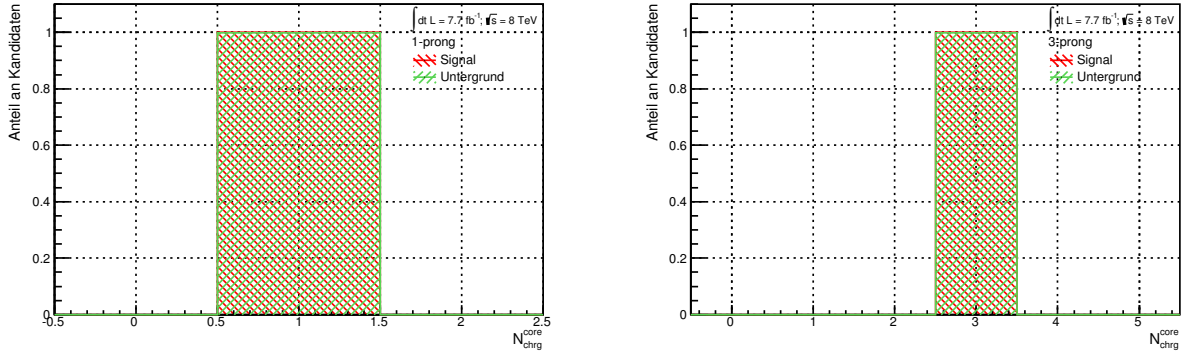


Abb. 7: Darstellung der Verteilungshistogramme der Variablen $N_{\text{chrg}}^{\text{core}}$ von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts)

Maximales ΔR zwischen geladenen Pionen und τ -Achse:

$$\Delta R_{\text{max, chrg}}^{\text{core/iso}}$$

Abstand zwischen der impulsreichsten und der impulsärmsten Spur:

$$\Delta R_{p_{\text{T}}, \text{minmax}}^{\text{iso}}$$

p_{T} -gewichteter Abstand der Spuren zur τ -Achse:

$$R_{\text{cal, chrg/ neut/ all}}^{\text{core/iso}} = \frac{\sum_{i \in \{\pi^{\pm}\}/\{\pi^0\}/\{\pi^{\pm}, \pi^0\}}^{\Delta R_i < 0.2/0.4} p_{\text{T}, i} \cdot \Delta R_i}{\sum_{j \in \{\pi^{\pm}\}/\{\pi^0\}/\{\pi^{\pm}, \pi^0\}}^{\Delta R_j < 0.2/0.4} p_{\text{T}, j}}$$

Dabei sind $R_{\text{cal, chrg}}^{\text{iso}}$ und R_{track} gleich definiert.

Massen der Zerfallsprodukte:

$$m_{\text{chrg/ neut/ all}}^{\text{cent/core/iso}} = \sqrt{\left(\sum_{i \in \{\pi^{\pm}\}/\{\pi^0\}/\{\pi^{\pm}, \pi^0\}}^{\Delta R_i < 0.1/0.2/0.4} E_i \right)^2 - \left(\sum_{i \in \{\pi^{\pm}\}/\{\pi^0\}/\{\pi^{\pm}, \pi^0\}}^{\Delta R_i < 0.1/0.2/0.4} \vec{p}_i \right)^2}$$

Dabei sind $m_{\text{chrg}}^{\text{iso}}$ und m_{track} sowie $m_{\text{all}}^{\text{core}}$ und m_{τ}^{vis} gleich definiert. Die Übereinstimmungen

zwischen den alten und neuen Variablen, trotz unterschiedlicher Algorithmen zur Ermittlung dieser, sind am Beispiel der Masse aller geladenen Zerfallsprodukte in der Isolationsregion bei Betrachtung von Abb. 8 (S. 20) und Abb. 9 (S. 20) ersichtlich.

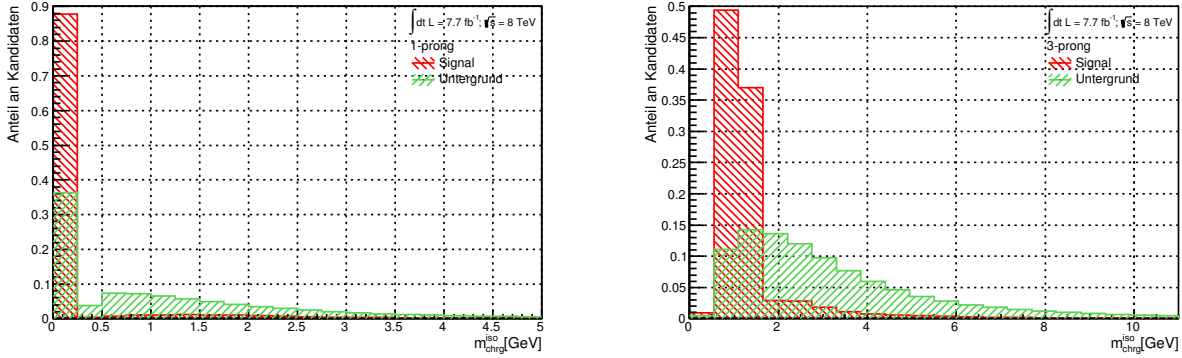


Abb. 8: Darstellung der Verteilungshistogramme der Variablen $m_{\text{chrg}}^{\text{iso}}$ von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts)

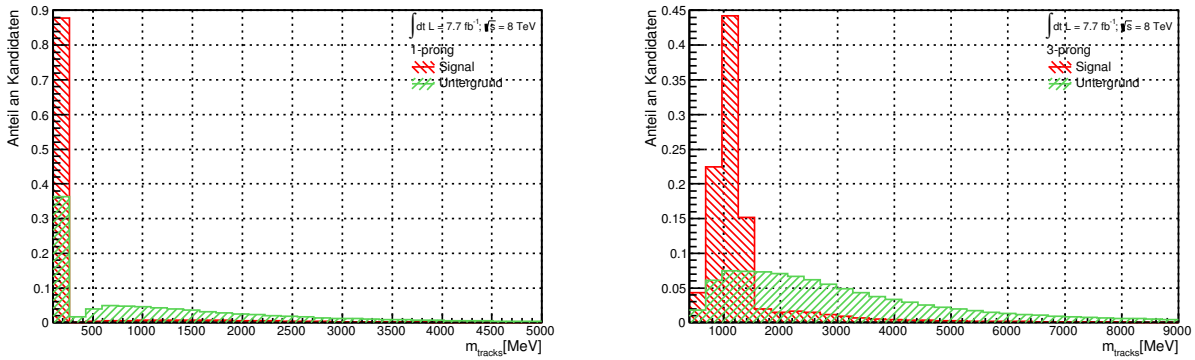


Abb. 9: Darstellung der Verteilungshistogramme der Variablen m_{tracks} von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts)

Energieanteile der Zerfallsprodukte:

$$f_{E_{T,\text{chrg/neut/all}}^{\text{cent-core/core-iso/cent-iso}}} = \frac{\sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm,\pi^0\}} \Delta R_i < 0.1/0.2/0.1 E_{T,i}}{\sum_{j \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm,\pi^0\}} \Delta R_j < 0.2/0.4/0.4 E_{T,j}}$$

Anteil des Transversalimpulses des führenden geladenen Pions an Transversalenergie:

$$f_{\text{lead,chrg}}^{\text{cent/core/iso}} = \frac{P_{\text{lead,chrg/neut/all}}^{\Delta R_i < 0.1/0.2/0.4}}{\sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm,\pi^0\}} \Delta R_i < 0.1/0.2/0.4 E_{T,i}}$$

Anteil des Transversalimpulses der Zerfallsprodukte am Gesamtimpuls des Kandidaten:

$$f_{p_{T,\text{chrg/neut/all}}^{\text{cent/core/iso}}} = \frac{\sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}}^{\Delta R_i < 0.1/0.2/0.4} p_{T,i}}{p_T^\tau}$$

Dabei sind $f_{p_{T,\text{all}}}^{\text{core}}$ und $f_{\text{vis-}p_T}$ gleich definiert. Allerdings kann man durch Vergleich der Abb. 10 (S. 21) mit Abb. 11 (S. 21) deutlich erkennen, dass der neue Algorithmus zu einer leicht anderen Verteilung der Variable führt. Daher ist der Informationsgehalt der neuen Variablen, auch bei gleicher Bedeutung wie zuvor, ein anderer und muss untersucht werden.

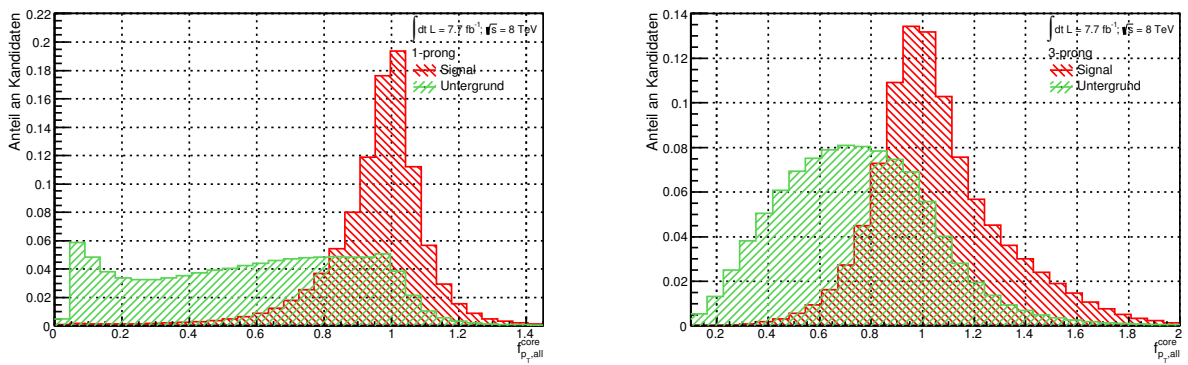


Abb. 10: Darstellung der Verteilungshistogramme der Variablen $f_{p_{T,\text{all}}}^{\text{core}}$ von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts)

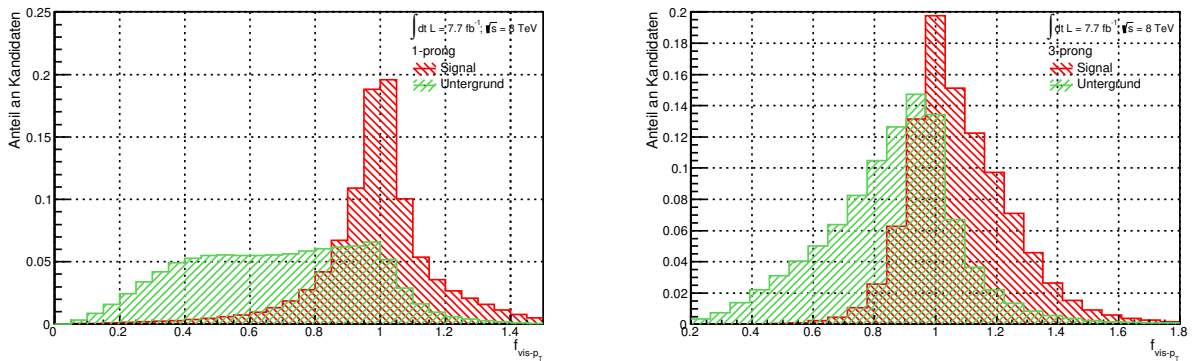


Abb. 11: Darstellung der Verteilungshistogramme der Variablen $f_{\text{vis-}p_T}$ von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts)

Im oberen Teil von Abb. 12 (S. 22) wird die Receiver Operating Characteristic-Kurve, kurz ROC-Kurve, eines neuronalen Netzes mit einer verborgenen Schicht, dessen Neuronenzahl der Variablenanzahl entspricht, unter Verwendung der Standardvariablen, bezeichnet als „*defaultMLP*“, verglichen mit der eines ähnlichen Netzes, bei dem die Variable $f_{p_{T,\text{all}}}^{\text{core}}$ hinzugefügt wird, bezeichnet als „*MLP069*“. Bei einer solchen Kurven ist die Untergrundunterdrückung, im Diagramm als *Inverse Background Efficiency* bezeichnet, über der Signalidenti-

fikationswahrscheinlichkeit, im Diagramm als *Signal Efficiency* bezeichnet, aufgetragen. Dabei gibt die Signalidentifikationswahrscheinlichkeit den prozentualen Anteil des echten Signals an, welcher auch als solches identifiziert wird, während die Untergrundunterdrückung angibt, jedes wievielte Ereignis aus dem Untergrund als Signalereignis identifiziert wird. Zum Beispiel würden an dem Punkt (0,405 | 10) im linken Diagramm 40,5% des Signals korrekt erkannt werden, wobei jedes zehnte Untergrundereignis auch dem Signal zugeordnet werden würde, dies entspricht 10%. Dementsprechend ist ein MLP oder BDT mit höherer Untergrundunterdrückung bei gleicher Signalidentifikationswahrscheinlichkeit besser für die Identifikation geeignet, weshalb dies als Kriterium für den Vergleich herangezogen wird. Damit ist die ROC-Kurve ein Maß für die Güte einer Identifikation. Der untere Teil der Abbildung zeigt immer das Verhältnis aus allen eingetragenen ROC-Kurven zu der in der Legende als erstes erwähnten. Offensichtlich scheint die neue Variable gegenüber der Alten keine neuen Informationen zu liefern, obwohl die Verteilungen leicht voneinander abweichen. Die leichte Verschlechterung ist wahrscheinlich eine Folge der Korrelation der alten mit der neuen Variable und der daraus resultierenden neuen Freiheitsgrade im Gradientenabstiegsverfahren für die Ermittlung des Minimums der Fehlerfunktion.

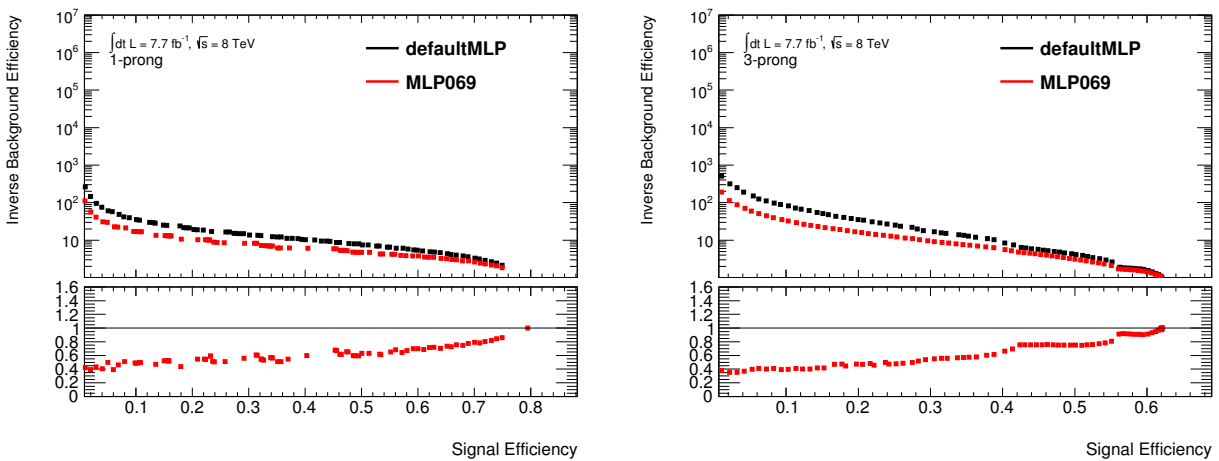


Abb. 12: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) eines MLP's mit Standardvariablen und $f_{pT,all}^{core}$ sowie Darstellung des zugehörigen Verhältnisses zum Vergleichswert (unten)

Die Verteilungshistogramme aller übrigen Variablen aus dem erweiterten Variablensatz, die bereits gezeigten eingeschlossen, sind in Anhang A.2 in Abb. 28-39 (S. 38-49) zu finden. Auf die Variablen mit dem positivsten Einfluss auf die Identifikation wird im Folgenden spezieller eingegangen.

4.2 Variablenkonfigurationen und ihr Einfluss auf die Identifikation mit MLPs

4.2.1 Variation der Variablenkonfiguration

Zunächst werden die Ergebnisse für den Vergleich der Identifikation mit neuronalen Netzen mit verschiedenen Variablenkonfigurationen diskutiert. Dazu wird für jede neue Variable ein Netz mit den Standardvariablen, einschließlich der jeweiligen neuen Variable, mit je einer verborgenen Schicht trainiert. Dessen Neuronenanzahl entspricht der Variablenanzahl. Dabei hat sich herausgestellt, dass neuronale Netze eine bemerkenswerte Instabilität in Bezug auf eine Veränderung der Konfiguration der Variablen aufweisen. In Tab. 1 (S. 23) sind die Variablen aufgeführt, welche in den Netzen in Abb. 13 (S. 23) und Abb. 15 (S. 25) verwendet werden.

Tab. 1: Auflistung der Kennzahlen der trainierten MLPs und der jeweils zu den Standardvariablen hinzugefügten neuen Variable

Neuronales Netz	MLP040	MLP042	MLP044	MLP077
Zusätzliche Variable	$f_{E_T,all}^{core-iso}$	$f_{E_T,all}^{cent-iso}$	$f_{E_T,all}^{cent-core}$	N_{neut}^{cent}

Während für 1-Prong-Ereignisse die Güte der Identifikation durch die MLPs mit $f_{E_T,all}^{cent-iso}$ und $f_{E_T,all}^{cent-core}$ eher gering voneinander abweicht, ist der Unterschied zu dem MLP mit $f_{E_T,all}^{core-iso}$ deutlich zu erkennen. Bei 3-Prong-Ereignissen sind zwischen allen MLPs die Abweichungen groß, wobei die Untergrundunterdrückung für das MLP mit $f_{E_T,all}^{core-iso}$ nahezu konstant eins ist. Dementsprechend kann hier fast keine Trennung zwischen Untergrund und Signal erfolgen.

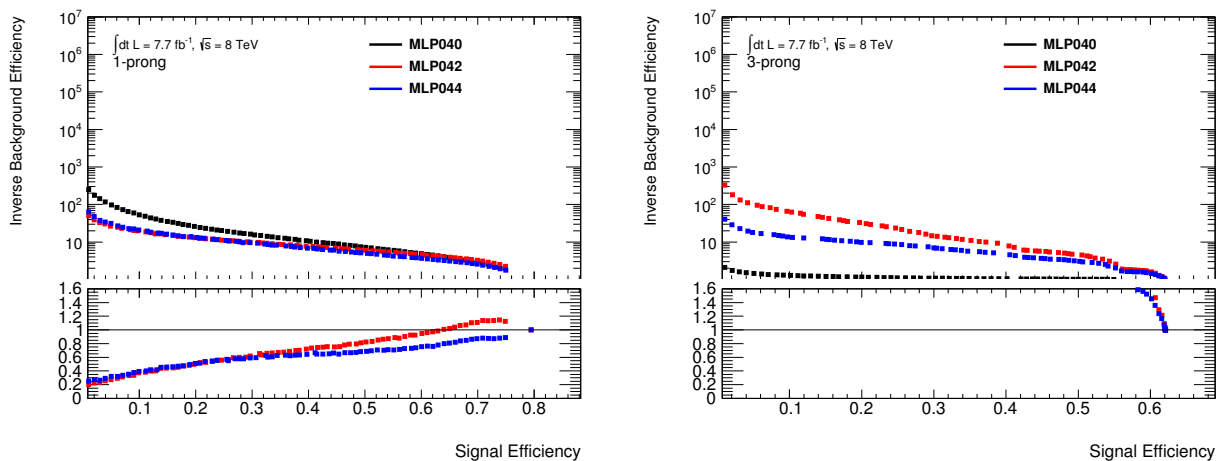


Abb. 13: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) eines MLP's mit Standardvariablen und $f_{E_T,all}^{cent-core}/core-iso/cent-iso$ sowie Darstellung des zugehörigen Verhältnisses zum Vergleichswert (unten)

Dieses Ergebnis ist fragwürdig, da die Verteilungen von $f_{E_T,all}^{core-iso}$ und $f_{E_T,all}^{cent-iso}$, wie in Abb. 14 (S.

23) zu sehen ist, sehr ähnlich sind. Ein so stark abweichendes Ergebnis bei der Identifikation war nicht zu erwarten. Im Vergleich dazu weicht die Verteilung von $f_{E_T, \text{all}}^{\text{cent-core}}$ von den anderen etwas stärker ab. Dennoch liefern die MLPs mit $f_{E_T, \text{all}}^{\text{cent-core}}$ und $f_{E_T, \text{all}}^{\text{cent-iso}}$ ein eher ähnliches Ergebnis. Dieses Verhalten wird auch bei weiteren Tests beobachtet. Allerdings kann kein erkennbares Muster anhand der Verteilungen oder Korrelationen der Variablen untereinander festgestellt werden.

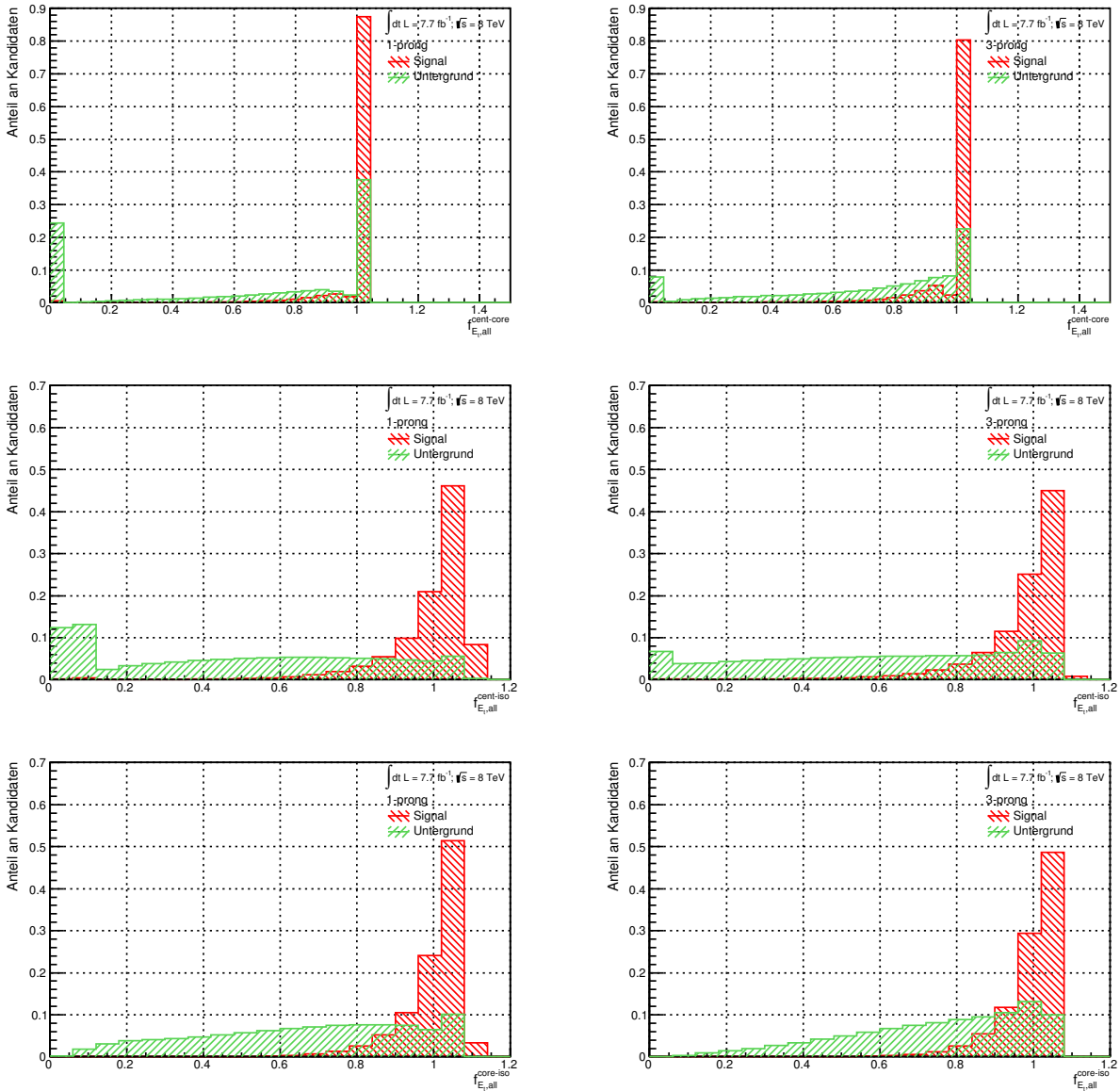


Abb. 14: Darstellung der Verteilungshistogramme der Variablen, verwendet in den neuronalen Netzen aus Abb. 13 (S. 23), von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts)

Darüber hinaus ist in Abb. 15 (S. 25) ersichtlich, dass für ein MLP mit $N_{\text{neut}}^{\text{cent}}$ eine signifikante Verschlechterung der Identifikation eintritt, obwohl die Variable sowohl für 1-Prong-, als auch für 3-Prong-Ereignisse bei Signal und Untergrund in ihrer Verteilung identisch ist. Daher sollte

sie keinen Einfluss auf die Identifikation haben. Diese Verschlechterung lässt sich teilweise mit der zunehmenden Korrelation durch die zusätzliche Variable erklären, ist aber wahrscheinlich auch auf die Instabilität der neuronalen Netze zurückzuführen. Die gleiche Verschlechterung wird beobachtet, wenn eine Standardvariable für die Identifikation zwei mal verwendet wird.

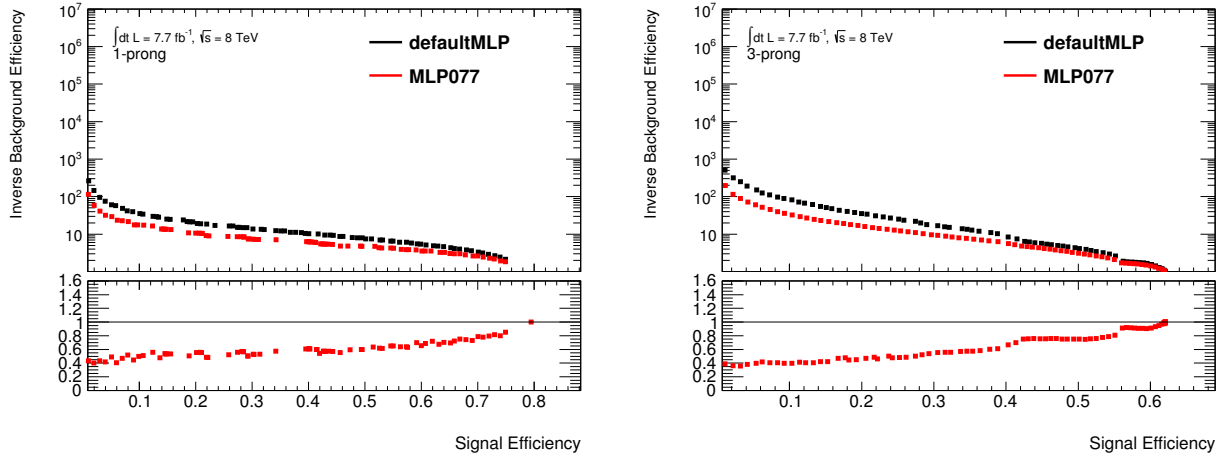


Abb. 15: Darstellung der ROC-Kurve für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) eines MLP's mit Standardvariablen und $N_{\text{neut}}^{\text{cent}}$ sowie Darstellung des zugehörigen Verhältnisses zum Vergleichswert (unten)

4.2.2 Variation der Netzwerkarchitektur

Um zu überprüfen, ob die beobachtete Instabilität die Folge einer starken Abhängigkeit der verwendeten Variablenkonfiguration von der Netzwerkarchitektur ist, wird der Einfluss dieser auf die Identifikation überprüft. Die verwendeten Konfigurationen sind in Tab. 2 (S. 25) dargestellt. Dabei ist N die Anzahl an Eingabevariablen.

Tab. 2: Auflistung der Kennzahlen der trainierten MLPs und der jeweils verwendeten Netzwerkarchitektur durch Angabe der Neuronenanzahl pro verborgener Schicht von der ersten Schicht nach der Eingabeschicht (links) zur letzten Schicht vor der Ausgabeschicht (rechts)

Neuronales Netz	MLP080	MLP081	MLP082	MLP111	MLP121
Netzwerkarchitektur	N	N,N	N,N,N	$5,5,5,5$	$15,15,15,15$
Neuronales Netz	MLP113	MLP114	MLP115	MLP116	MLP117
Netzwerkarchitektur	10	10,10	10,10,10	10,10,10,10	10,10,10,10,10

Die Zahlen in den Tabellen sind von Links nach Rechts so angeordnet, dass die erste Zahl die Anzahl der Neuronen in der ersten verborgenen Schicht nach der Eingabeschicht angibt und die letzte die Anzahl der Neuronen in der letzten verborgenen Schicht vor dem Ausgabeneuron. Damit entspricht die Anzahl der Zahlen der Anzahl an verborgenen Schichten. In Abb. 16 (S.

26), Abb. 17 (S. 26) und Abb. 18 (S. 27) wird deutlich, dass auch die Netzwerkarchitektur einen signifikanten Einfluss auf die Identifikation mit neuronalen Netzen hat. Allerdings wird auch hier deutlich, dass es einen weiteren Effekt geben muss, der das Ergebnis stark beeinflusst, denn es zeichnen sich keine klaren Tendenzen ab. Sowohl die Variation der Anzahl an verborgenen Schichten, als auch der Anzahl an Neuronen pro verborgener Schicht, führen auf teilweise sehr unterschiedliche Ergebnisse, die weder eine Tendenz zu einer großen, noch zu einer kleinen Anzahl von Schichten bzw. Neuronen aufweisen. Um eine Aussage darüber treffen zu können, welche Konfiguration sich am besten für die Identifikation eignet, ist es also zunächst notwendig herauszufinden, was die Ursache für die beobachtete Instabilität ist.

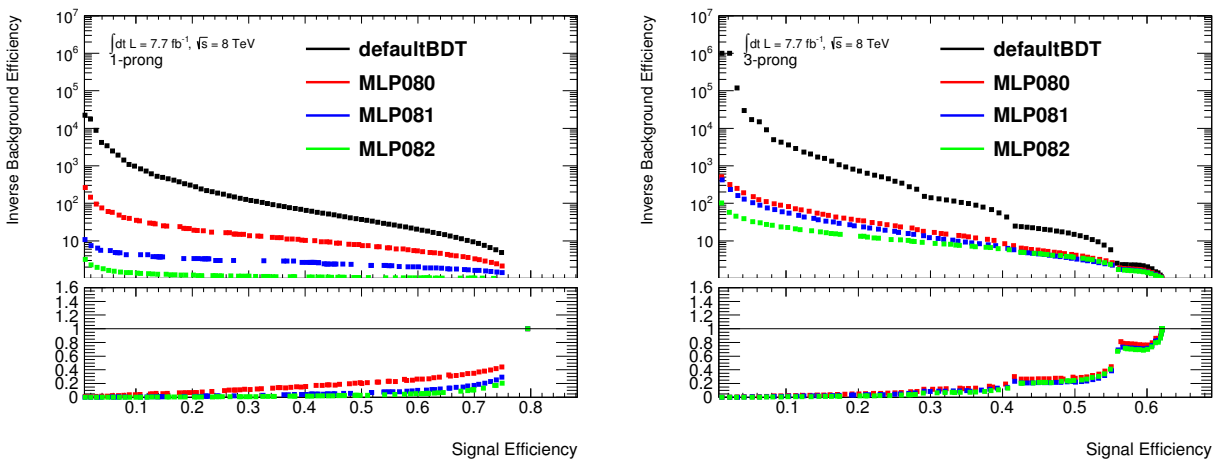


Abb. 16: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) für MLPs bei Variation der Anzahl der verborgenen Schichten mit einer Neuronenanzahl pro Schicht gleich der Variablenanzahl sowie Darstellung der zugehörigen Verhältnisse zum Vergleichswert (unten)

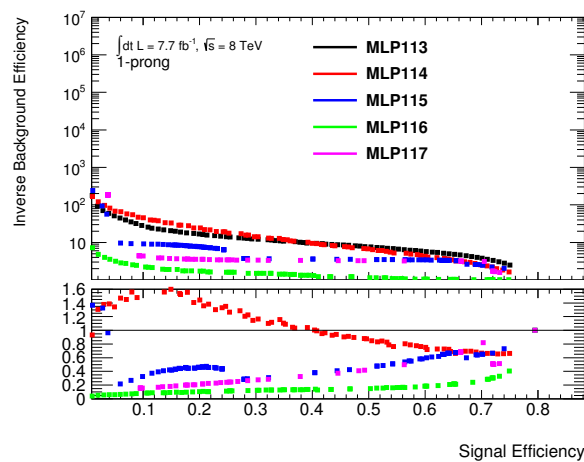


Abb. 17: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen für MLPs bei Variation der Anzahl der verborgenen Schichten mit 10 Neuronen pro Schicht sowie Darstellung der zugehörigen Verhältnisse zum Vergleichswert (unten)

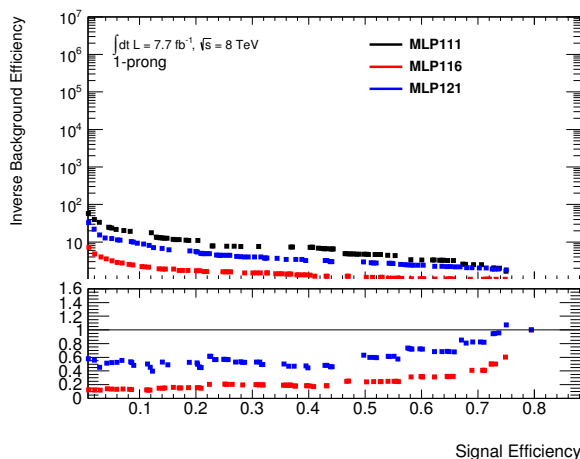


Abb. 18: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen für MLPs bei Variation der Neuronenanzahl pro verborgener Schicht sowie Darstellung der zugehörigen Verhältnisse zum Vergleichswert (unten)

4.2.3 Variation der Anfangswerte

Bei der Ermittlung des Minimums der Fehlerfunktion im Gewichtsraum des neuronalen Netzes wird ein numerisches Verfahren angewandt. Da es sich hierbei um keine analytische Lösung handelt, erhält man kein exaktes Ergebnis. Die Güte der Identifikation mit dem Netz ist also stets davon abhängig, wie „gut“ das gefundene Minimum ist. Bei einem guten Algorithmus für das Näherungsverfahren, sollte das Ergebnis weitestgehend unabhängig von den jeweiligen Anfangswerten der Gewichte sein. Um also die Stabilität des Trainingsalgorithmus zu testen, werden baugleiche neuronale Netze mit der gleichen Variablenkonfiguration trainiert und lediglich der *Seed* des Pseudozufallszahlen-Generators für die Anfangswerte wird verändert. In Abb. 19 (S. 27) sowie in weiteren Messungen wird deutlich, dass der verwendete Algorithmus eine hohe Instabilität in Bezug auf die verwendeten Anfangswerte der Gewichte aufweist.

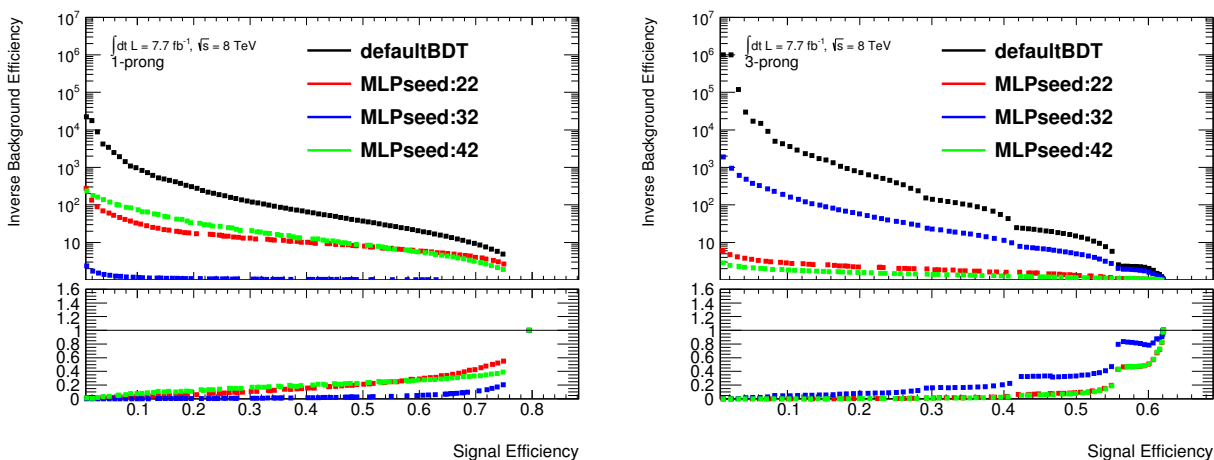


Abb. 19: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) für MLPs mit unterschiedlichen Anfangswerten sowie Darstellung der zugehörigen Verhältnisse zum Vergleichswert (unten)

Hier ist dies am Beispiel für die *Seeds* 22, 32 und 42 im Vergleich zum *defaultBDT* aufgeführt.

4.2.4 Fazit

In Anbetracht der Tatsache, dass der verwendete Algorithmus sehr instabil ist und sich daher nur schlecht für das Training neuronaler Netze zur Identifikation von τ -Leptonen eignet, ist es nicht möglich eindeutig zu ermitteln, welche Variablenkonfiguration in Zusammenhang mit welcher Netzwerkarchitektur das beste Ergebnis liefert, da man keinen klaren Trend erkennen kann. Weiterhin ist es nicht möglich, bestimmte Variablen oder Variablenkonfigurationen für die Identifikation als ungeeignet zu befinden, da man nie ausschließen kann, dass ein bestimmter Satz von Anfangswerten zu einem signifikant besseren Ergebnis führt, als die überprüften. Allerdings ist es möglich, durch Prüfung verschiedener Netze mit unterschiedlichen Variablenkonfigurationen und unterschiedlicher Infrastruktur bei verschiedenen *Seeds*, einige leichte Tendenzen festzustellen. Die folgenden Ergebnisse werden ermitteln, indem von je drei verwendeten *Seeds*, der mit dem besten Ergebnis für den Vergleich herangezogen wird. Als Netzwerkarchitektur hat sich dabei eine Kombination aus vier bis fünf verborgenen Schichten mit je 13 bis 14 Neuronen pro Schicht bewährt. Die besten Variablen als Ergänzung zu den Standardvariablen sind in Tab. 3 (S. 28) aufgeführt. Die besten Variablen für den Austausch sind in Tab. 4 (S. 29) dargestellt.

Tab. 3: Auflistung der Kennzahlen der besten trainierten MLPs, vom besten (links) zum schlechtesten (rechts), und der jeweils zu den Standardvariablen hinzugefügten neuen Variable für die Identifikation von 1-Prong-Ereignissen (oben) und 3-Prong-Ereignissen (unten)

Ranking	1	2	3	4	5
1-Prong-Variable	$f_{E_T, \text{chrg}}^{\text{core-iso}}$	$R_{\text{cal,all}}^{\text{iso}}$	$R_{\text{cal,all}}^{\text{core}}$	$f_{\text{lead, chrg}}^{\text{cent}}$	$f_{E_T, \text{neut}}^{\text{core-iso}}$
Neuronales Netz	MLP045	MLP059	MLP039	MLP049	MLP047
Ranking	1	2	3	4	5
3-Prong-Variable	$\Delta R_{p_T, \text{minmax}}^{\text{iso}}$	$m_{\text{all}}^{\text{iso}}$	$m_{\text{neut}}^{\text{cent}}$	$m_{\text{neut}}^{\text{core}}$	$m_{\text{chrg}}^{\text{iso}}$
Neuronales Netz	MLP041	MLP037	MLP056	MLP057	MLP055

Die ROC-Kurven, für die in Tab. 3 (S. 28) und in Tab. 4 (S. 29) aufgeführten MLPs, werden jeweils in Abb. 20 (S. 29) und Abb. 21 (S. 29) dargestellt. Es ist deutlich zu erkennen, dass für 1-Prong-Ereignisse starke Verbesserungen erreicht werden, während dies für 3-Prong-Ereignisse nicht möglich ist. Die Kurve, welche als „*defaultBDT*“ bezeichnet wird, stellt das Ergebnis des besten bekannten BDT's dar. Mit „*MLP035_default*“ wird das MLP mit Standardvariablen und gleicher Netzwerkarchitektur wie die anderen MLPs bezeichnet.

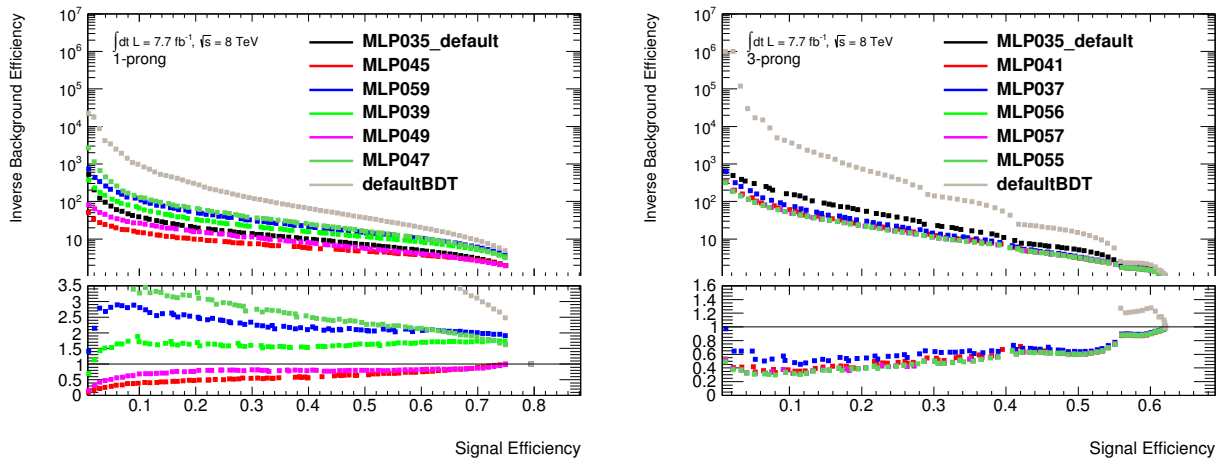


Abb. 20: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) für die besten 5 MLPs bei Hinzufügen einer Variable sowie Darstellung der zugehörigen Verhältnisse zum Vergleichswert (unten)

Tab. 4: Auflistung der Kennzahlen der besten trainierten MLPs, vom besten (links) zum schlechtesten (rechts), und der jeweils ausgetauschten Standardvariable mit der verwendeten neuen Variable für die Identifikation von 1-Prong-Ereignissen (links) und 3-Prong-Ereignissen (rechts)

Ranking	1	2	3	1
Standardvariable	f_{vis-p_T}	R_{track}	$f_{\text{core}}^{\text{cent}}$	$\Delta R_{\text{max}}^{\text{core}}$
Neue Variable	$f_{p_T, \text{chrg}}^{\text{core}}$	$R_{\text{cal, all}}^{\text{iso}}$	$f_{E_T, \text{chrg}}^{\text{core-iso}}$	$\Delta R_{\text{max, chrg}}^{\text{iso}}$
Neuronales Netz	MLP137	MLP142	MLP131	MLP156

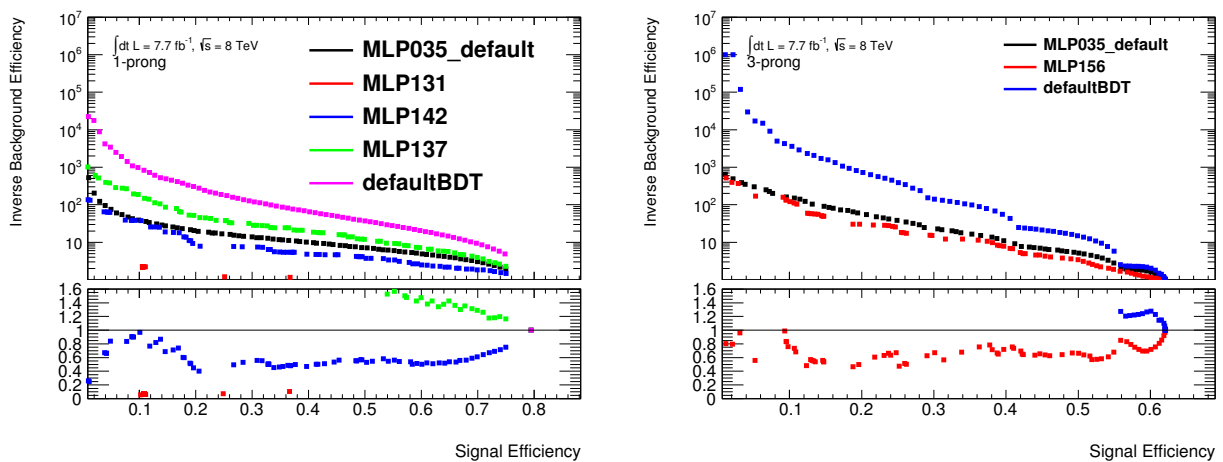


Abb. 21: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) für die besten MLPs bei Austausch einer alten durch eine neue Variable sowie Darstellung der zugehörigen Verhältnisse zum Vergleichswert (unten)

4.3 Variablenkonfigurationen und ihr Einfluss auf die Identifikation mit BDTs

Weiterhin wird der Einfluss der neuen Variablen auf den BDT überprüft. Analog zu der Untersuchung der MLPs, wird zu diesem Zweck für jede neue Variable in Kombination mit den Standardvariablen die Identifikation mit dem BDT mit dem besten bekannten Ergebnis, im Folgenden als „*defaultBDT*“ bezeichnet, und mit dem selbst erstellten BDT mit Standardvariablen, welcher als „*BDT080_default*“ bezeichnet wird, verglichen.

Die fünf besten Variablen sind sowohl für 1-Prong-, als auch für 3-Prong-Ereignisse, in Tab. 5 (S. 30) aufgeführt. Die Güte der Identifikation mit diesen Variablensätzen wird in Abb. 22 (S. 31) dargestellt. Daraus wird ersichtlich, dass es für 1-Prong-Ereignisse keine Verbesserung gegenüber dem „*defaultBDT*“ gibt. Bei 3-Prong-Ereignissen ist hingegen eine deutliche Verbesserung im Bereich unterhalb einer Signalidentifikationswahrscheinlichkeit von 30% erkennbar. Allerdings sind für die $\tau_{\text{had-vis}}$ -Identifikation nur die höheren Bereiche relevant, in denen nur geringe bis keine Verbesserungen vorhanden sind. Es handelt sich dabei für die Identifikation von 1-Prong-Ereignissen um Signalidentifikationswahrscheinlichkeiten von 40%, 60% oder 70%, je nachdem ob es eine *tight*, *medium* oder *loose identification* ist. Für die Identifikation von 3-Prong-Ereignissen liegen diese Werte bei 35%, 55% und 65%. Welche Identifikation gewählt wird, hängt davon ab, welche Menge an „Fehlidentifikationen“ man tolerieren kann.

Da die verwendeten Einstellungen des „*defaultBDT*“ nicht bekannt sind, können sie nicht für diese Untersuchungen verwendet werden. Daher ist es von Bedeutung, dass die aufgeführten Variablenkonfigurationen im Vergleich zu den Standardvariablen bei gleichen Einstellungen für den BDT eine signifikante Verbesserung bewirken. Ob und inwiefern sich die neuen Variablensätze auf die Identifikation auswirken, wenn die Einstellungen des „*defaultBDT*“ verwendet würden, lässt sich an dieser Stelle nicht sagen. Es ist aber zu erwarten, dass die bisherigen Ergebnisse mit der Konfiguration des *defaultBDT* qualitativ reproduziert werden können.

Tab. 5: Auflistung der Kennzahlen der besten trainierten BDTs, vom besten (links) zum schlechtesten (rechts), und der jeweils zu den Standardvariablen hinzugefügten neuen Variable für die Identifikation von 1-Prong-Ereignissen (oben) und 3-Prong-Ereignissen (unten)

Ranking	1	2	3	4	5
1-Prong-Variable	$m_{\text{neut}}^{\text{iso}}$	$f_{\text{lead,chg}}^{\text{iso}}$	$f_{p_{\text{T}},\text{chg}}^{\text{iso}}$	$f_{p_{\text{T}},\text{neut}}^{\text{cent}}$	$f_{p_{\text{T}},\text{all}}^{\text{cent}}$
BDT	BDT175	BDT174	BDT076	BDT071	BDT068
Ranking	1	2	3	4	5
3-Prong-Variable	$f_{p_{\text{T}},\text{chg}}^{\text{core}}$	$f_{p_{\text{T}},\text{chg}}^{\text{cent}}$	$f_{p_{\text{T}},\text{all}}^{\text{iso}}$	$m_{\text{neut}}^{\text{iso}}$	$m_{\text{all}}^{\text{iso}}$
BDT	BDT075	BDT074	BDT070	BDT175	BDT168

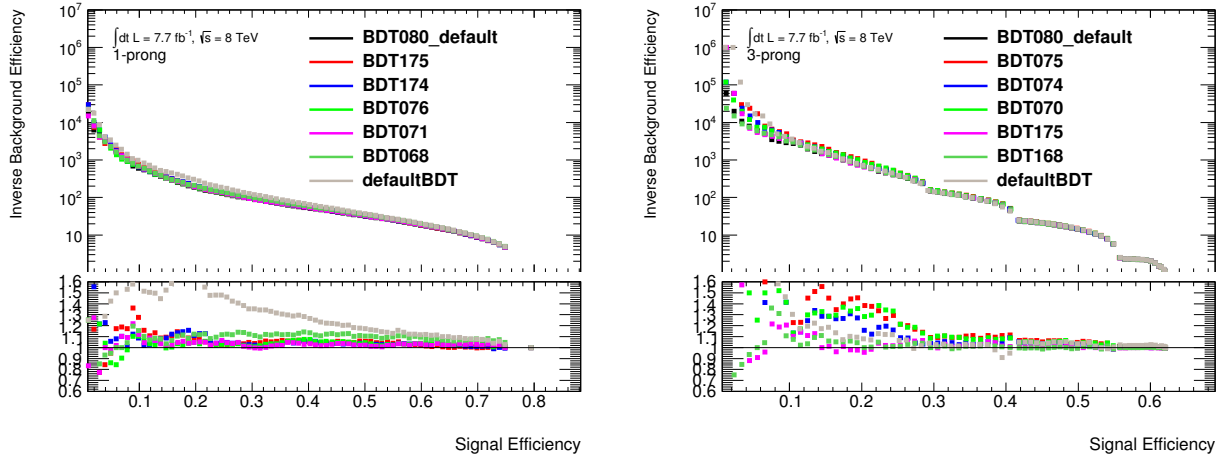


Abb. 22: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) für die besten 5 BDTs bei Hinzufügen einer Variable sowie Darstellung der zugehörigen Verhältnisse zum Vergleichswert (unten)

In Abb. 23 (S. 32) wird die Güte der Identifikation mit einer Kombinationen aus den drei besten Variablen dargestellt. Diese werden in Tab. 6 (S. 31) nach ihrer Eignung für die Identifikation geordnet. Für 1-Prong-Ereignisse gibt es für alle möglichen Kombinationen keine Verbesserung der Identifikation gegenüber dem „*defaultBDT*“, analog zu den Ergebnissen für die einzelnen Variablen. Für 3-Prong-Ereignisse gibt es hingegen eine signifikante Verbesserung um eine bis zu 15% höhere Untergrundunterdrückung im relevanten Bereich zwischen 30% und 40% Signalidentifikationswahrscheinlichkeit. Die Verbesserung im Bereich geringerer Signalidentifikationswahrscheinlichkeiten ist für die $\tau_{\text{had-vis}}$ -Identifikation uninteressant. Im Vergleich zum selbst erstellten BDT mit Standardvariablen, „*BDT080_default*“, gibt es hingegen in jedem Fall eine signifikante Verbesserung im relevanten Bereich mit einer bis zu 15% höheren Untergrundunterdrückung. Die zu beobachtenden Verbesserung sind dabei kaum größer als jene, die auch bei Hinzufügen einzelner Variablen aufgetreten sind.

Tab. 6: Auflistung der Kennzahlen der besten trainierten BDTs, vom besten (links) zum schlechtesten (rechts), und der jeweils zu den Standardvariablen hinzugefügten Kombination neuer Variablen für die Identifikation von 1-Prong-Ereignissen (oben) und 3-Prong-Ereignissen (unten)

Ranking	1	2	3	4
1-Prong-Var.	$m_{\text{neut}}^{\text{iso}} / f_{\text{lead,chg}}^{\text{iso}} / f_{p_{\text{T},\text{chg}}}^{\text{iso}}$	$m_{\text{neut}}^{\text{iso}} / f_{p_{\text{T},\text{chg}}}^{\text{iso}}$	$f_{\text{lead,chg}}^{\text{iso}} / f_{p_{\text{T},\text{chg}}}^{\text{iso}}$	$m_{\text{neut}}^{\text{iso}} / f_{\text{lead,chg}}^{\text{iso}}$
BDT	BDT181	BDT179	BDT180	BDT178

Ranking	1	2	3	4
3-Prong-Var.	$f_{p_{\text{T},\text{chg}}}^{\text{core}} / f_{p_{\text{T},\text{all}}}^{\text{iso}}$	$f_{p_{\text{T},\text{chg}}}^{\text{core}} / f_{p_{\text{T},\text{chg}}}^{\text{cent}} / f_{p_{\text{T},\text{all}}}^{\text{iso}}$	$f_{p_{\text{T},\text{chg}}}^{\text{cent}} / f_{p_{\text{T},\text{all}}}^{\text{iso}}$	$f_{p_{\text{T},\text{chg}}}^{\text{core}} / f_{p_{\text{T},\text{chg}}}^{\text{cent}}$
BDT	BDT179	BDT181	BDT180	BDT178

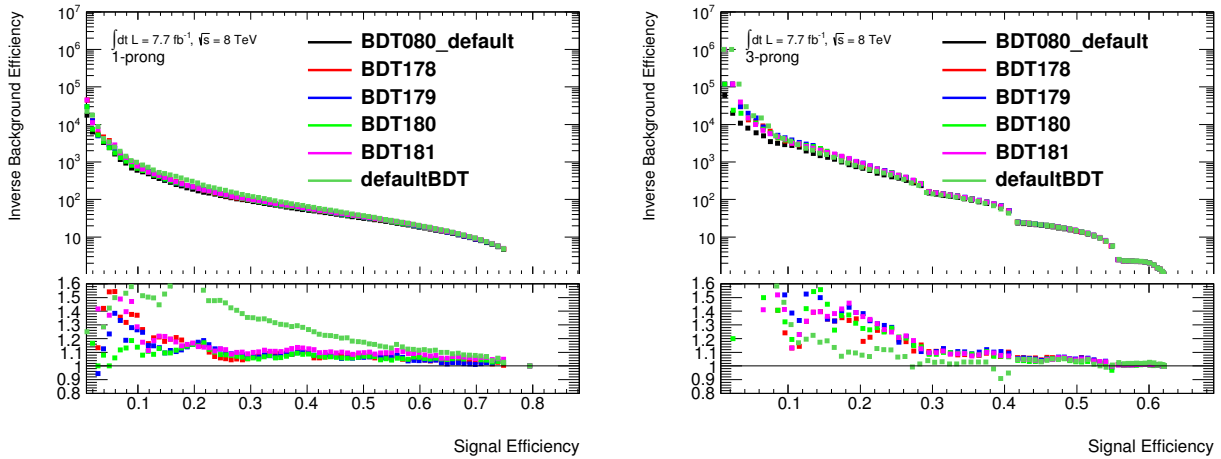


Abb. 23: Darstellung der ROC-Kurven für die Identifikation von 1-Prong-Ereignissen (links-oben) und 3-Prong-Ereignissen (rechts-oben) für die besten BDTs bei Hinzufügen einer Kombination von Variablen sowie Darstellung der zugehörigen Verhältnisse zum Vergleichswert (unten)

Allgemein haben sich für die Identifikation speziell die Massen der Zerfallsprodukte und die Anteile der Transversalimpulse der Zerfallsprodukte am Gesamttransversalimpuls des Ereignisses als sehr gut geeignete Variablen für die Verbesserung der Identifikation erwiesen. Durch die Verteilungen jener Variablen, die sich als separationskräftig erwiesen haben, in Anhang A.2, wird dieses Ergebnis überwiegend bestätigt. Insbesondere die Variablen mit Bezug auf die Isolationsregion haben sich, gegenüber ihren Gegenstücken für die Zentral- und Kernregion, als geeigneter erwiesen.

Die schlechteren Identifikationen von 1-Prong- im Vergleich zu 3-Prong-Ereignissen, sind nicht durch die lineare Korrelation der Variablen mit den Standardvariablen oder untereinander begründbar, da diese, wie in Abb. 24 (S. 33) dargestellt wird, bei 3-Prong-Ereignissen wesentlich stärker ausgeprägt ist, als bei 1-Prong-Ereignissen. Das trifft gleichermaßen auf die Signal- und die Untergrunddaten zu. Allerdings gibt diese Darstellung keine Informationen über die Korrelation in höherer Ordnung, was möglicherweise ein Grund für die schlechtere Identifikation sein könnte. Wahrscheinlicher ist es, dass die Separationskraft der Variablen an sich die Hauptursache ist. Eine günstige Kombination führt, wie bei der Identifikation der 3-Prong-Ereignisse ersichtlich, zu einer signifikanten Verbesserung des BDT's.

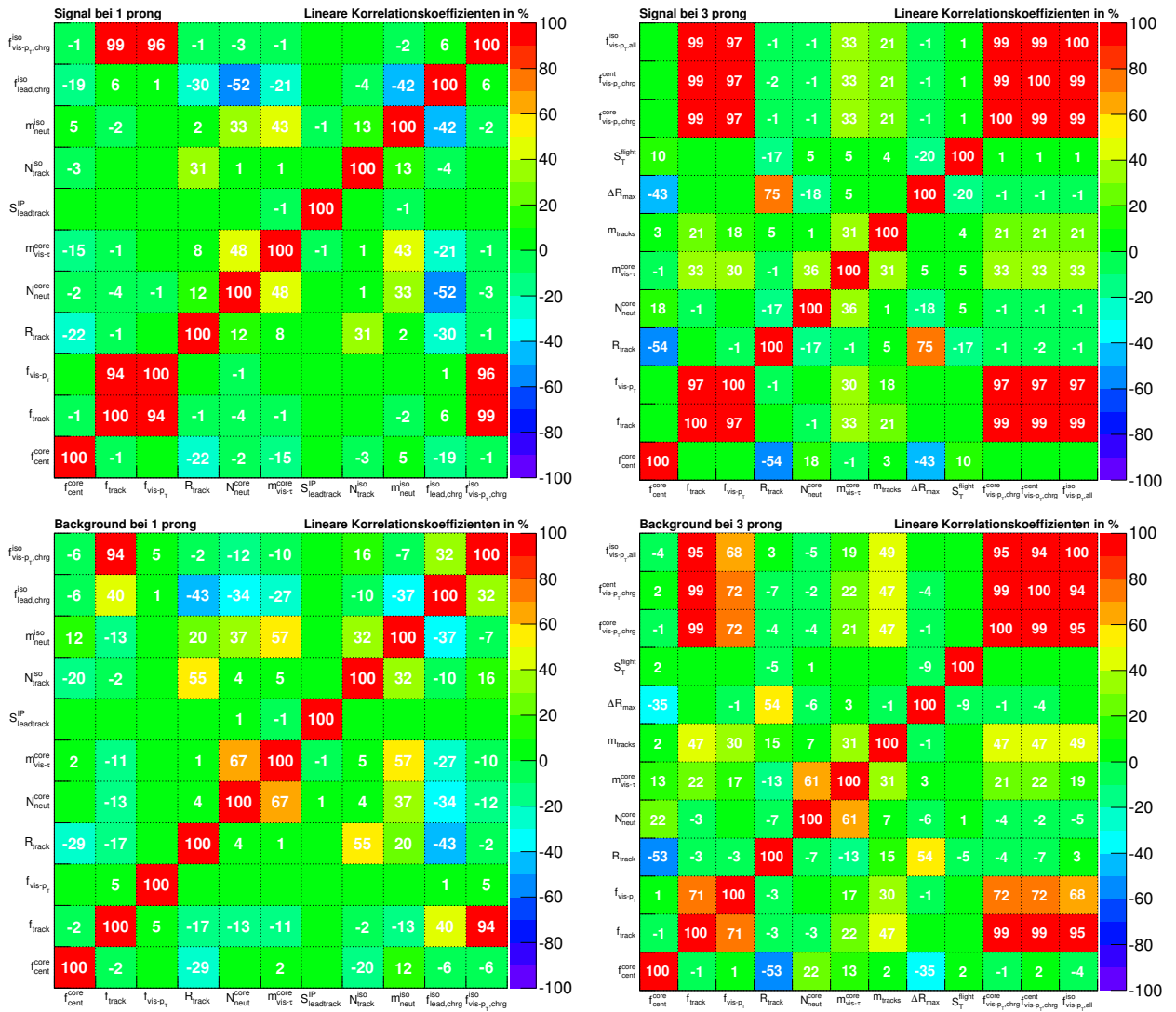


Abb. 24: Darstellung der linearen Korrelation in % zwischen jeweils zwei der Variablen aus der besten Variablen-Kombination für BDTs von Signaldaten (oben) und Untergrunddaten (unten) von 1-Prong-Ereignissen (links) und 3-Prong-Ereignissen (rechts)

5 Zusammenfassung der Ergebnisse und Ausblick

Abschließend ergibt sich, dass künstliche neuronale Netze für die Identifikation von $\tau_{\text{had-vis}}$ -Kandidaten nicht geeignet sind, wenn sie mittels des hier verwendeten Gradientenabstiegsverfahrens trainiert werden. Dieser hat sich als zu instabil erwiesen. Darüber hinaus konnte aufgrund dieses Umstandes keine Variablen- und Netzwerkkonfiguration gefunden werden, die den aktuell führenden *Boosted Decision Tree* in der Güte seiner Identifikation erreicht oder gar übersteigt. Dennoch bleibt es notwendig, die Untersuchungen mittels stabilerer Algorithmen für die Ermittlung des globalen Minimums fortzuführen um das volle Potential der neuronalen Netze abschätzen zu können, da sich trotz der Instabilität teilweise vielversprechende Tendenzen gezeigt haben.

Weiterhin haben sich die neuen Variablen als hilfreich für die Verbesserung der Identifikation erwiesen. Es traten sowohl starke als auch geringere Verbesserungen auf, weshalb es nötig ist, in dieser Richtung weitere Nachforschungen anzustellen, was aufgrund der beschränkten Arbeitszeit im Zuge dieser Arbeit nicht möglich war. Einerseits wäre durch eine Optimierung der Einstellungen des verwendeten *Boosted Decision Tree's* ein aussagekräftigerer Vergleich zum aktuellen Spitzenwert möglich. Andererseits ist auch die Überprüfung weiterer Kombinationen von Variablen empfehlenswert, da die Anzahl der möglichen Konfigurationen aus alten und neuen Variablen den Umfang dieser Arbeit um ein Vielfaches überschreitet. Allerdings bietet dieser Einblick eine gute Orientierung für spätere Untersuchungen. Danach bleibt der *Boosted Decision Tree* in näherer Zukunft die bevorzugte Methode zur Identifikation von $\tau_{\text{had-vis}}$ -Kandidaten.

A Diagramme

A.1 Verteilung der Standardvariablen

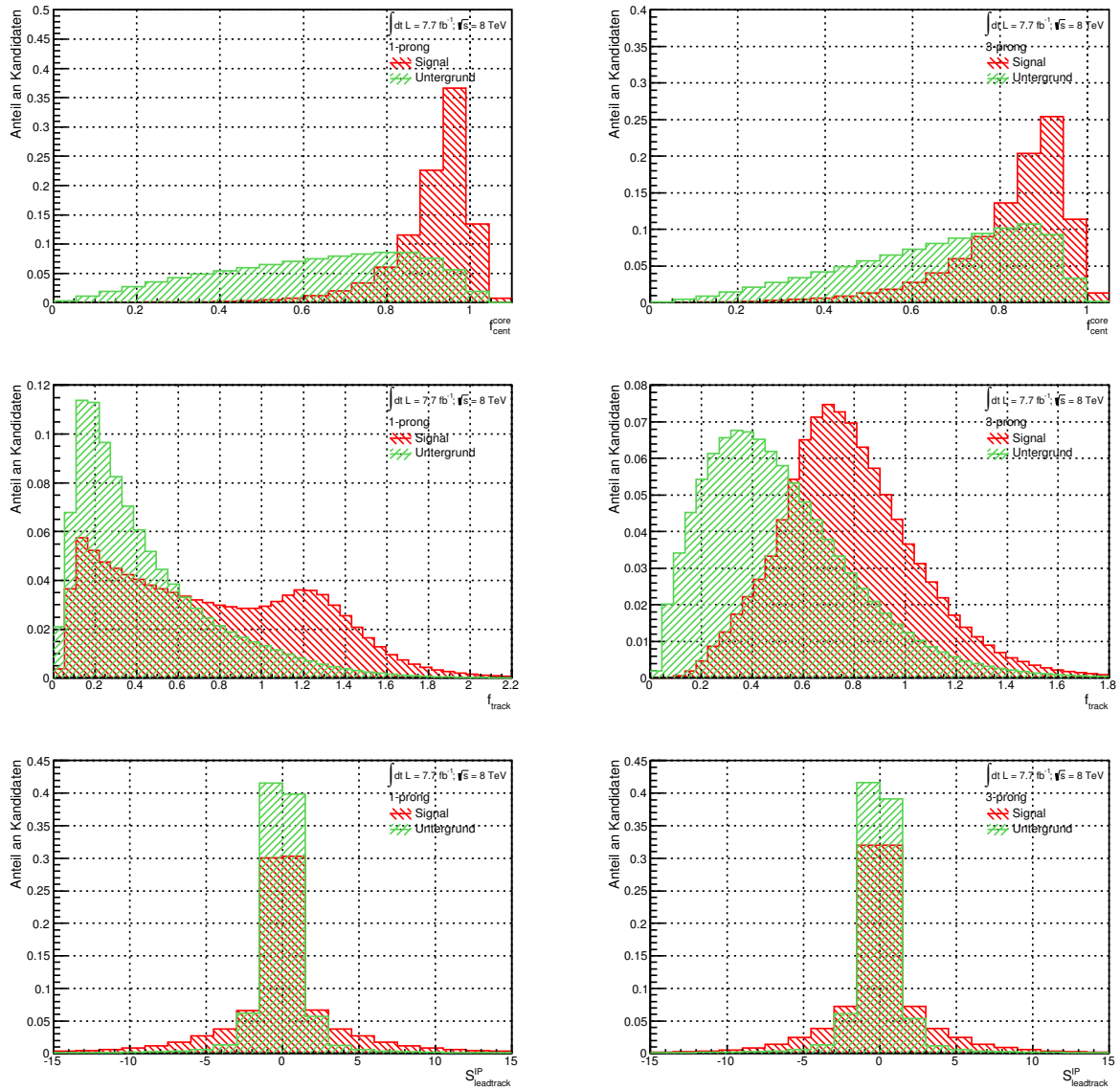


Abb. 25: Verteilungshistogramme der Standardvariablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil I

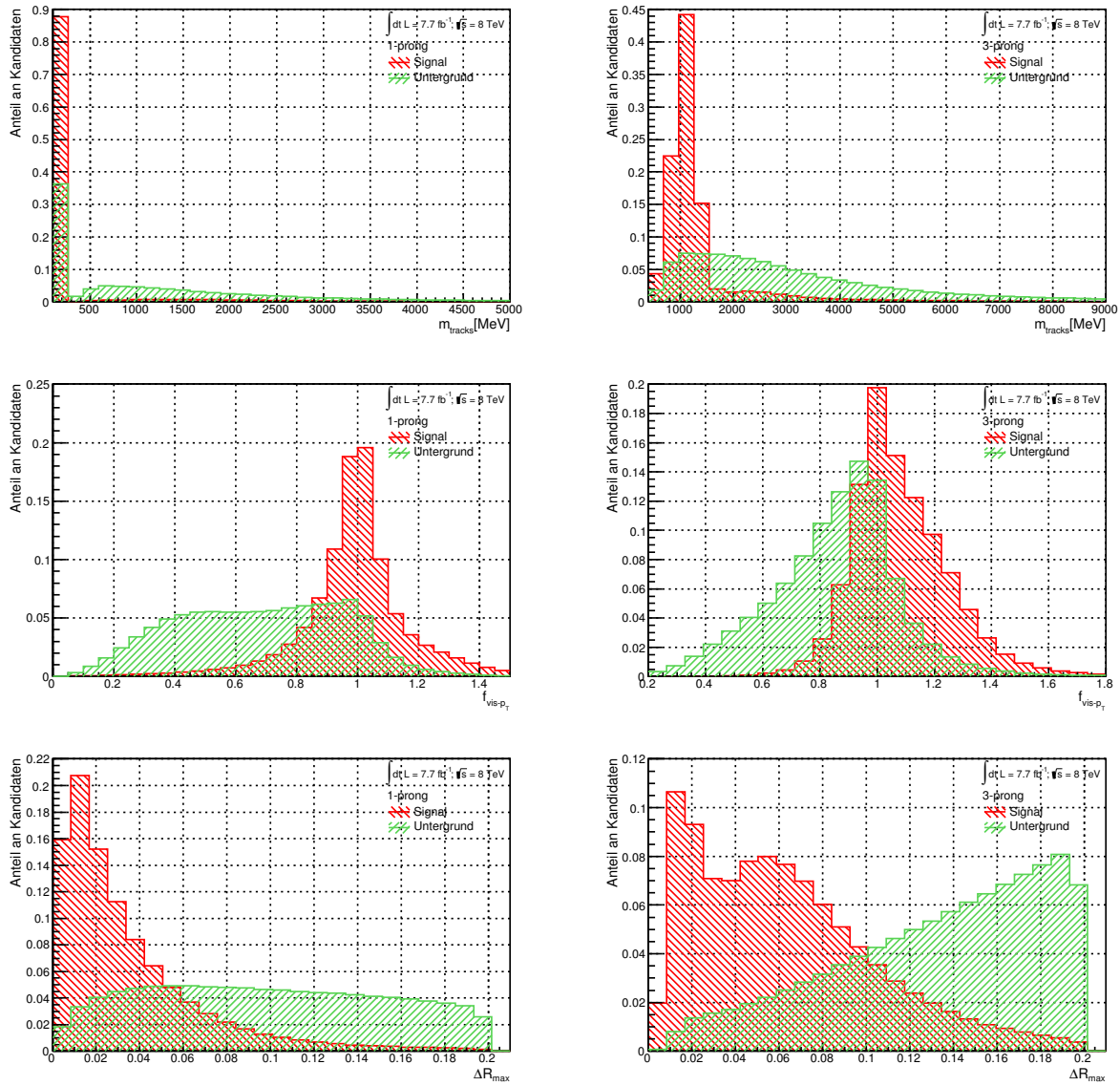


Abb. 26: Verteilungshistogramme der Standardvariablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil II

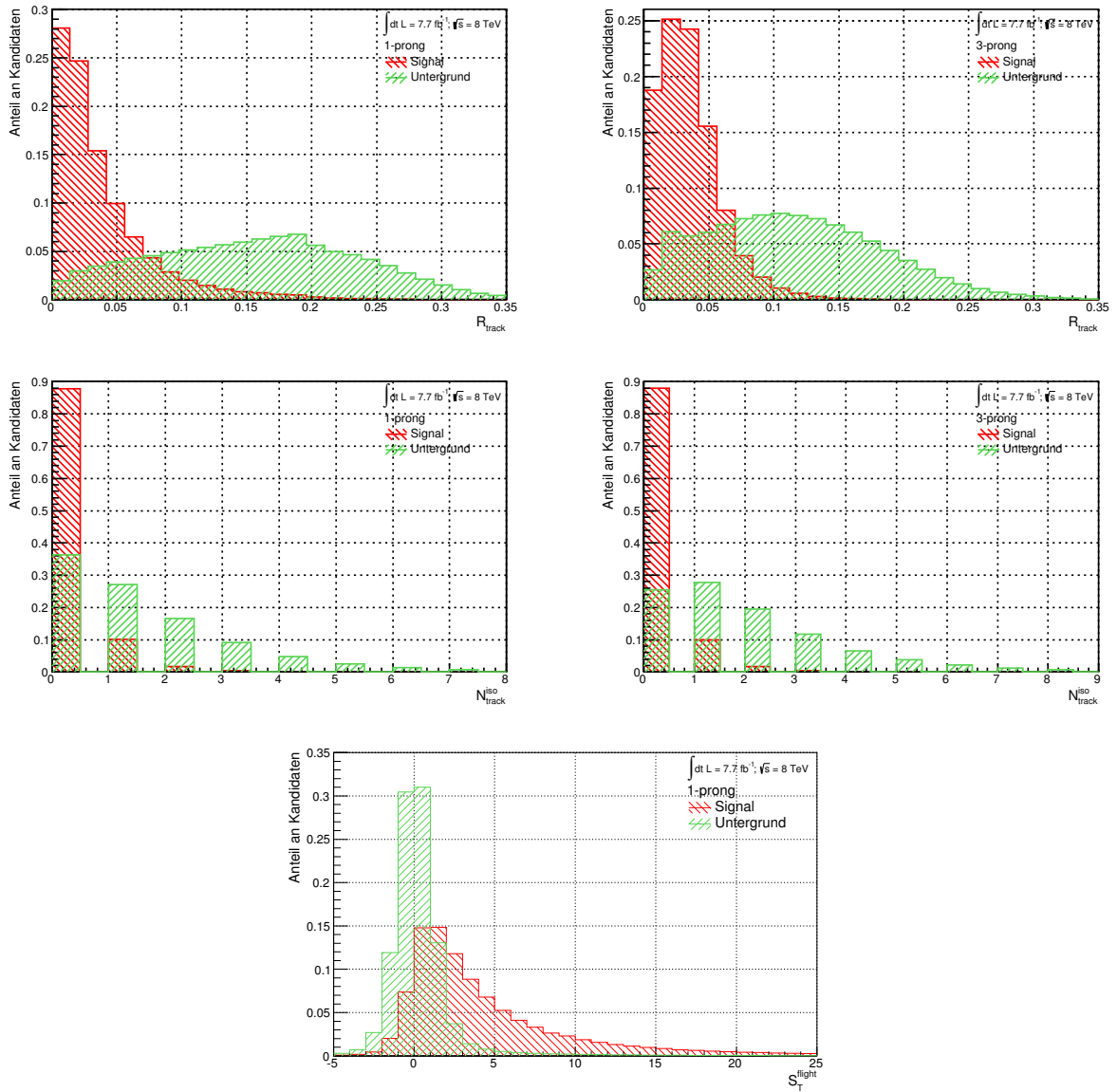


Abb. 27: Verteilungshistogramme der Standardvariablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil III

A.2 Verteilung der erweiterten Variablen

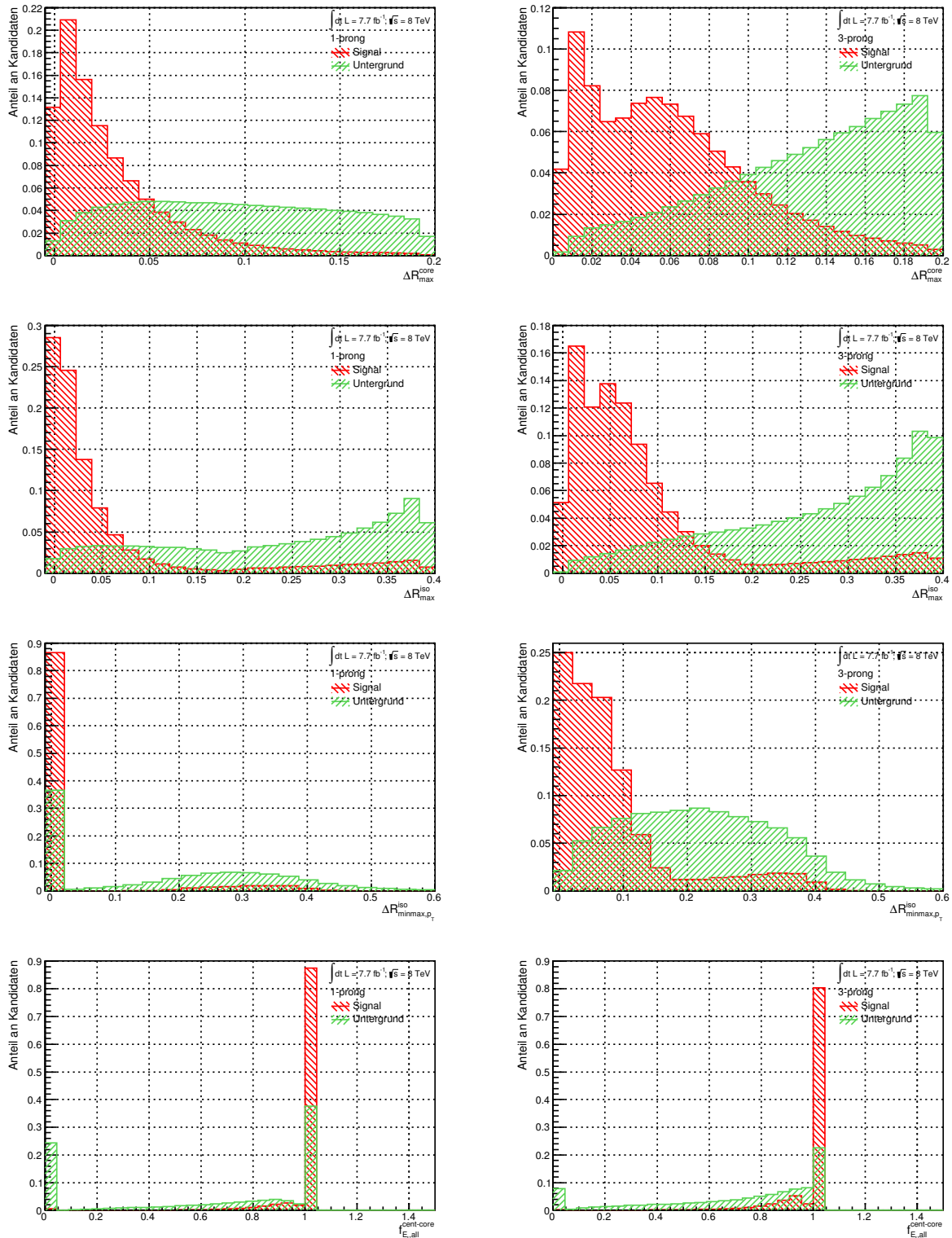


Abb. 28: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil I

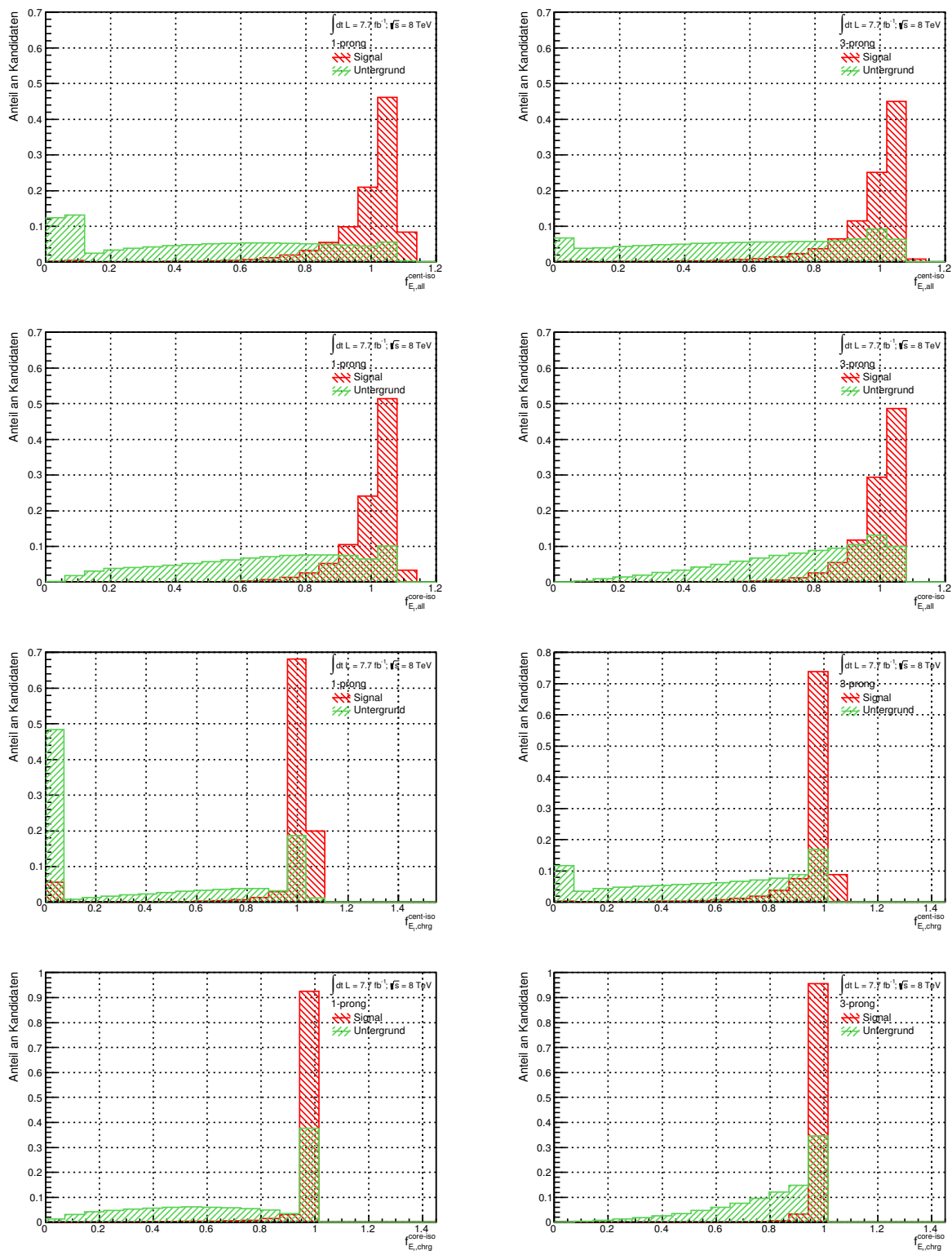


Abb. 29: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil II

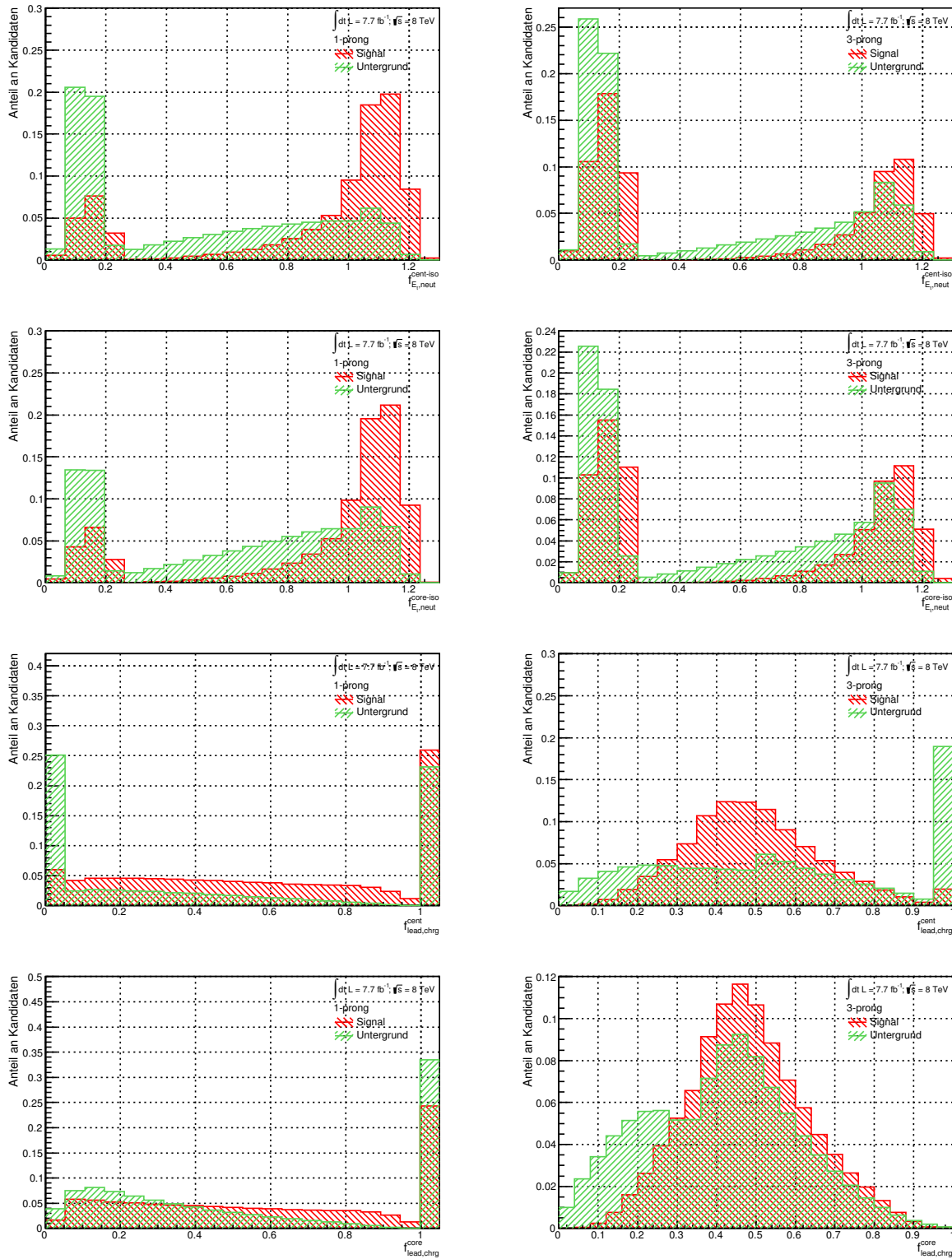


Abb. 30: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil III

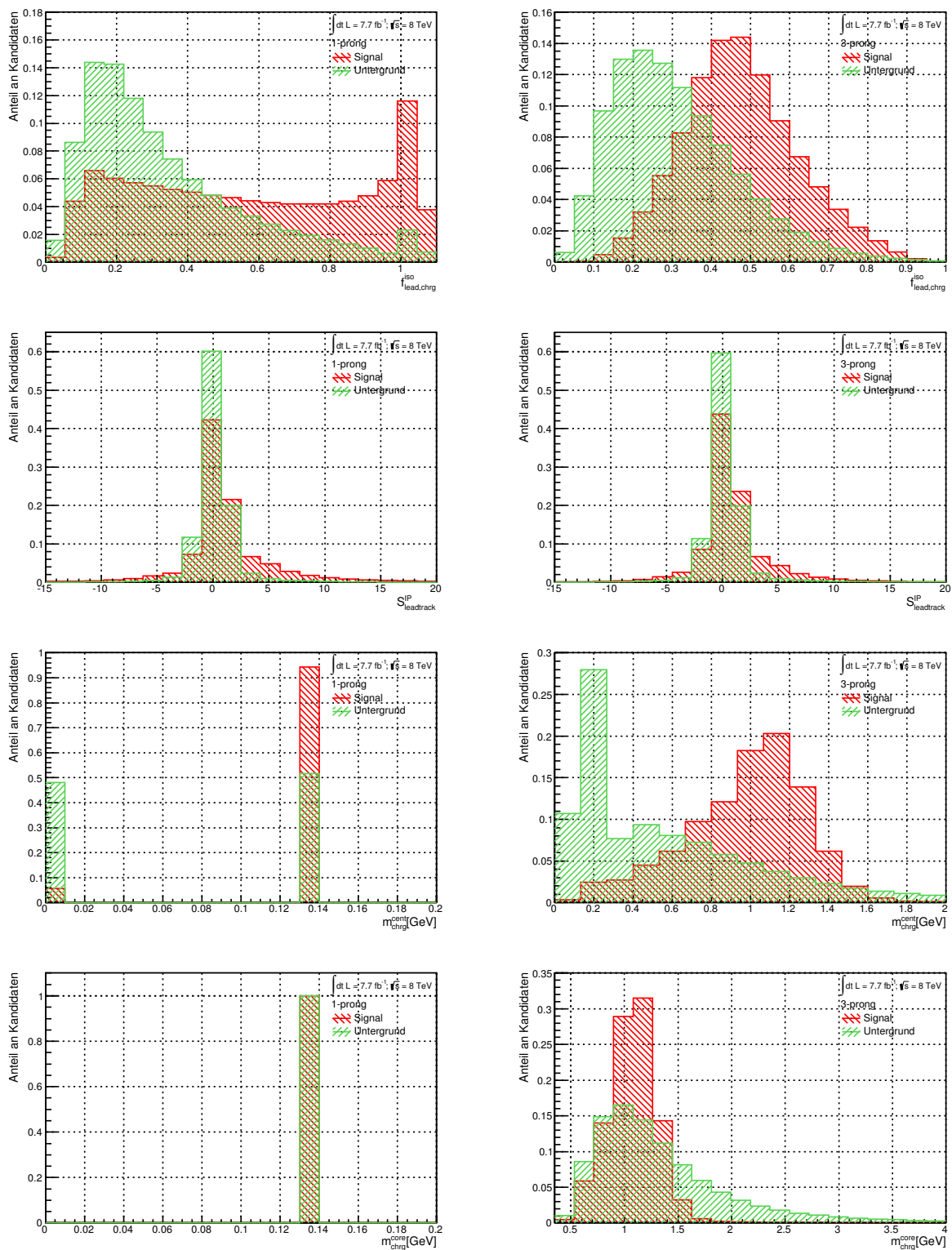


Abb. 31: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil IV

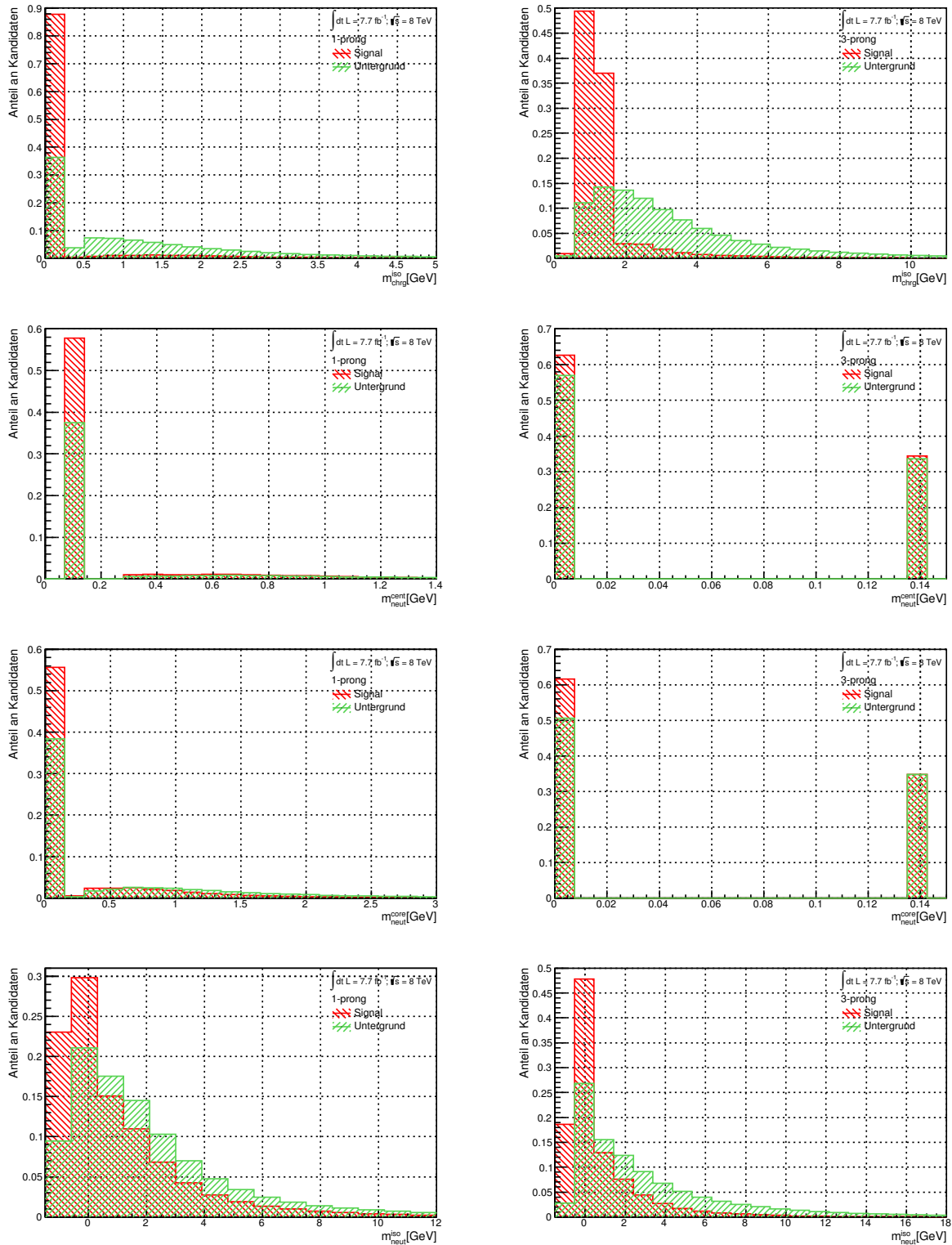


Abb. 32: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil V

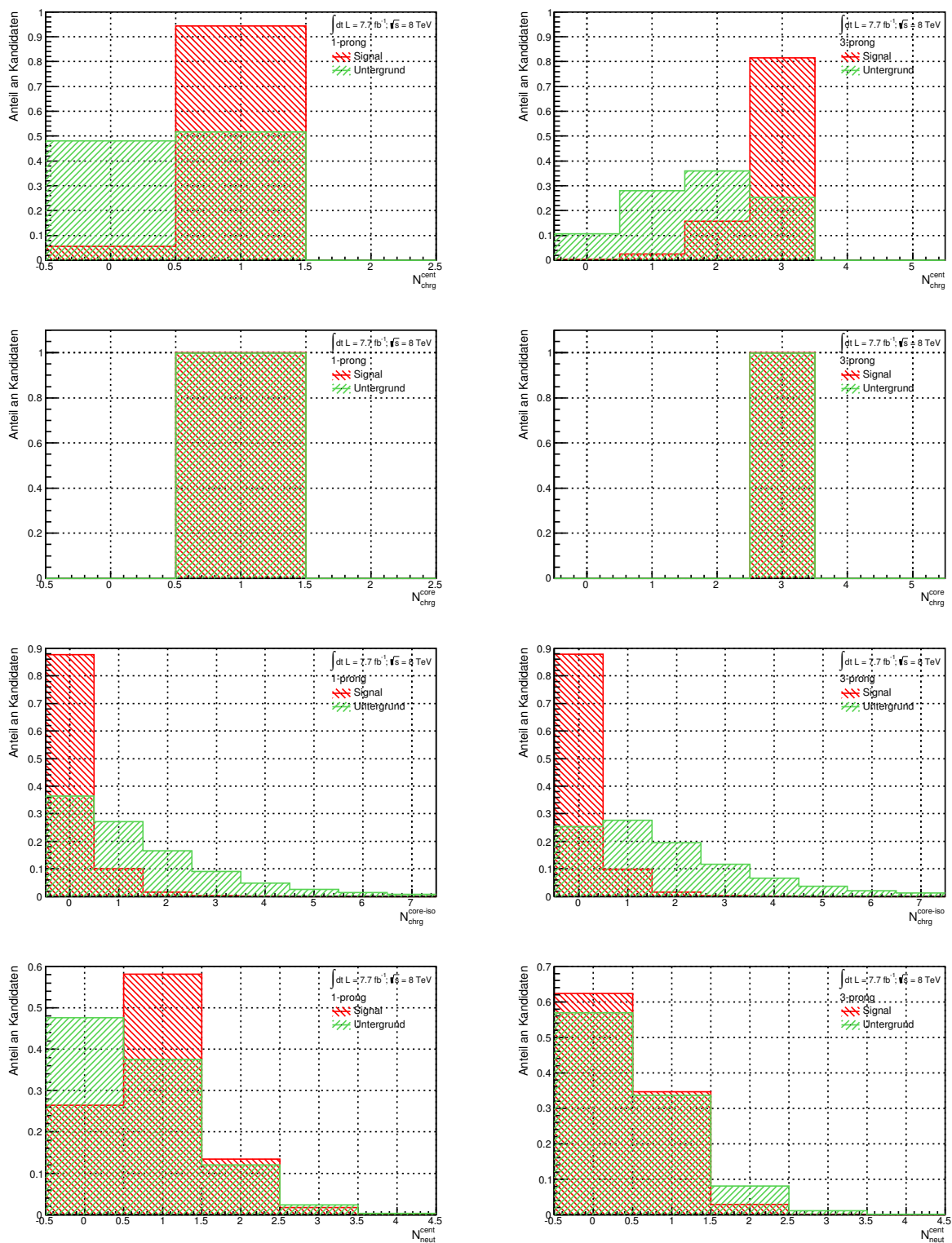


Abb. 33: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil VI

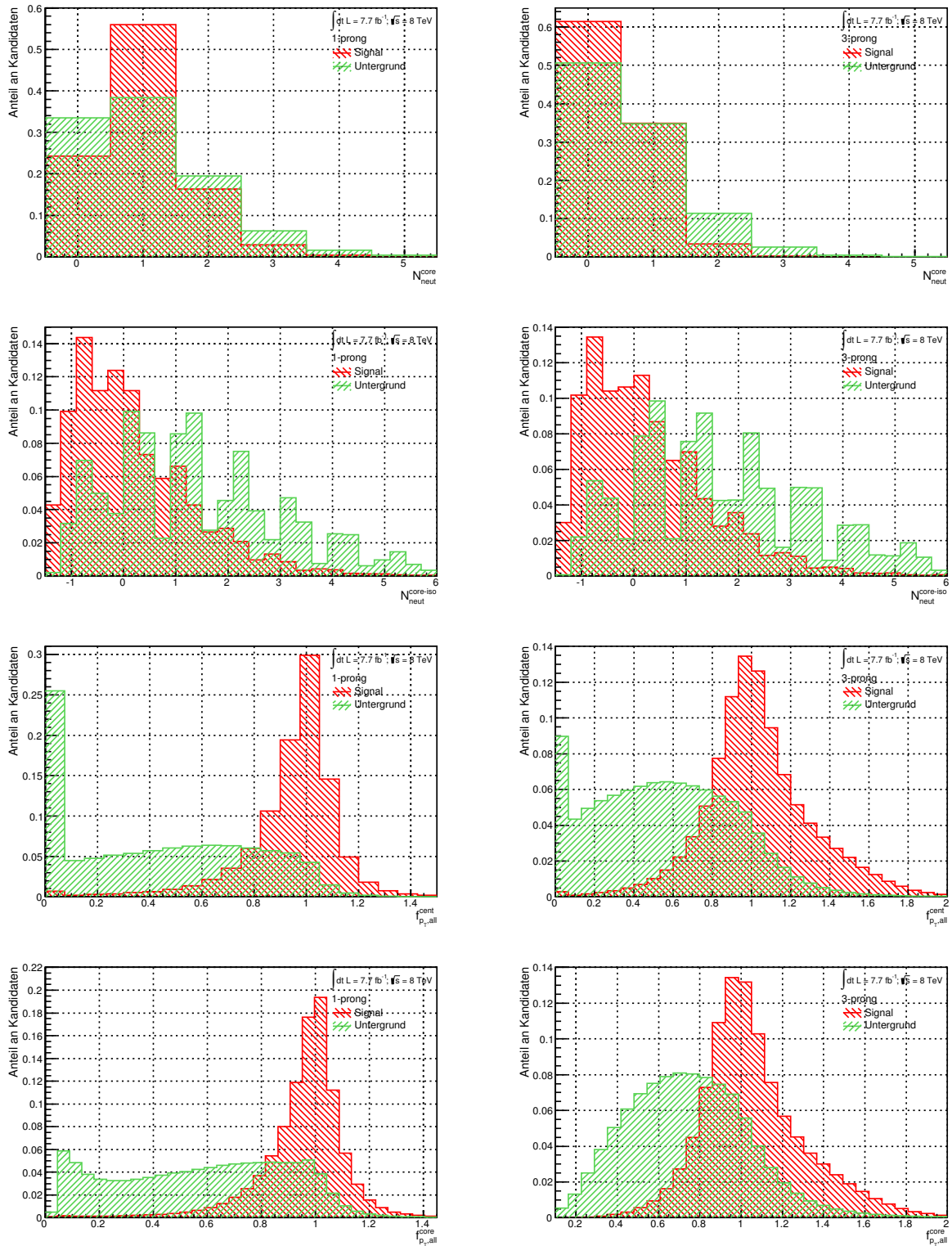


Abb. 34: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil VII

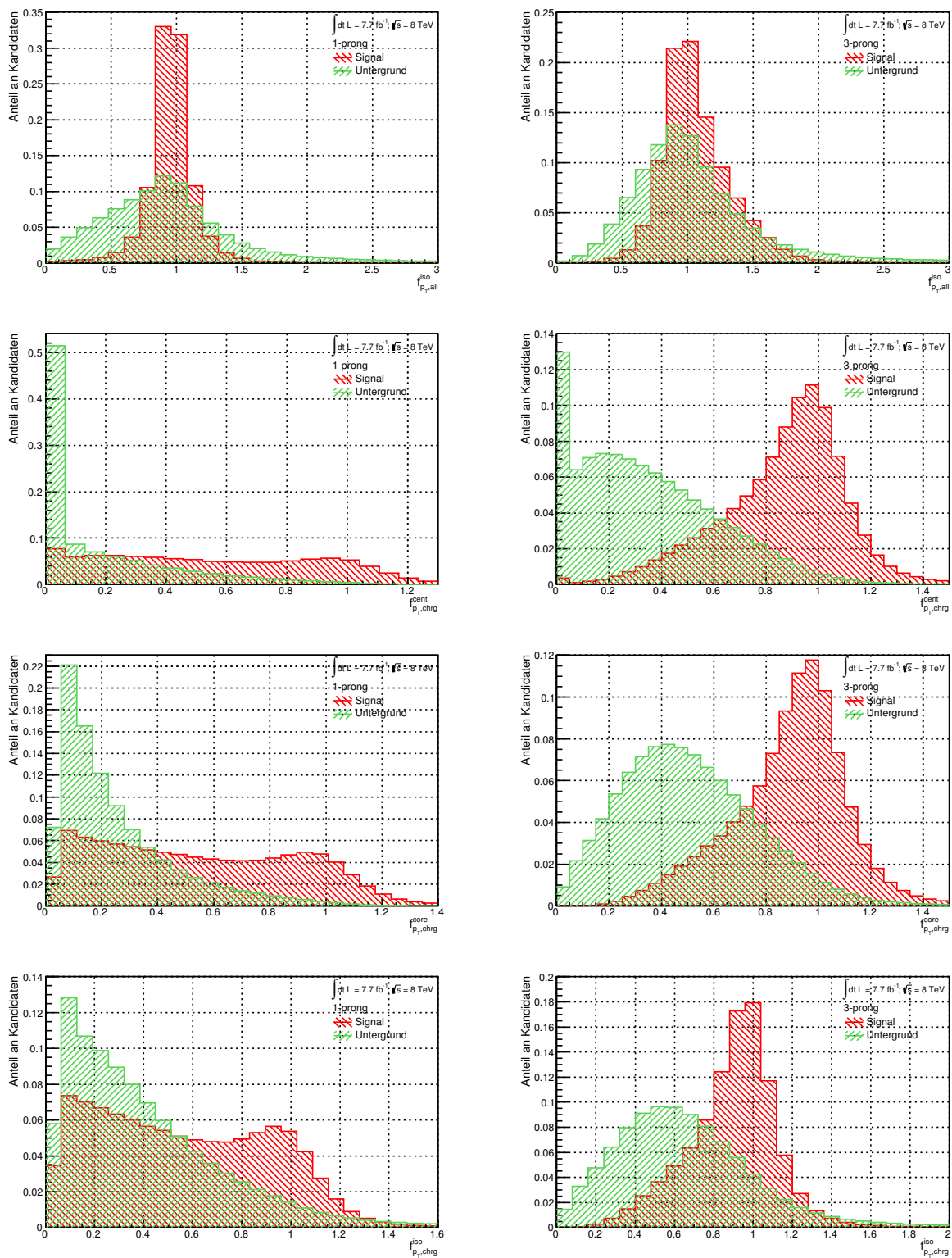


Abb. 35: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil VIII

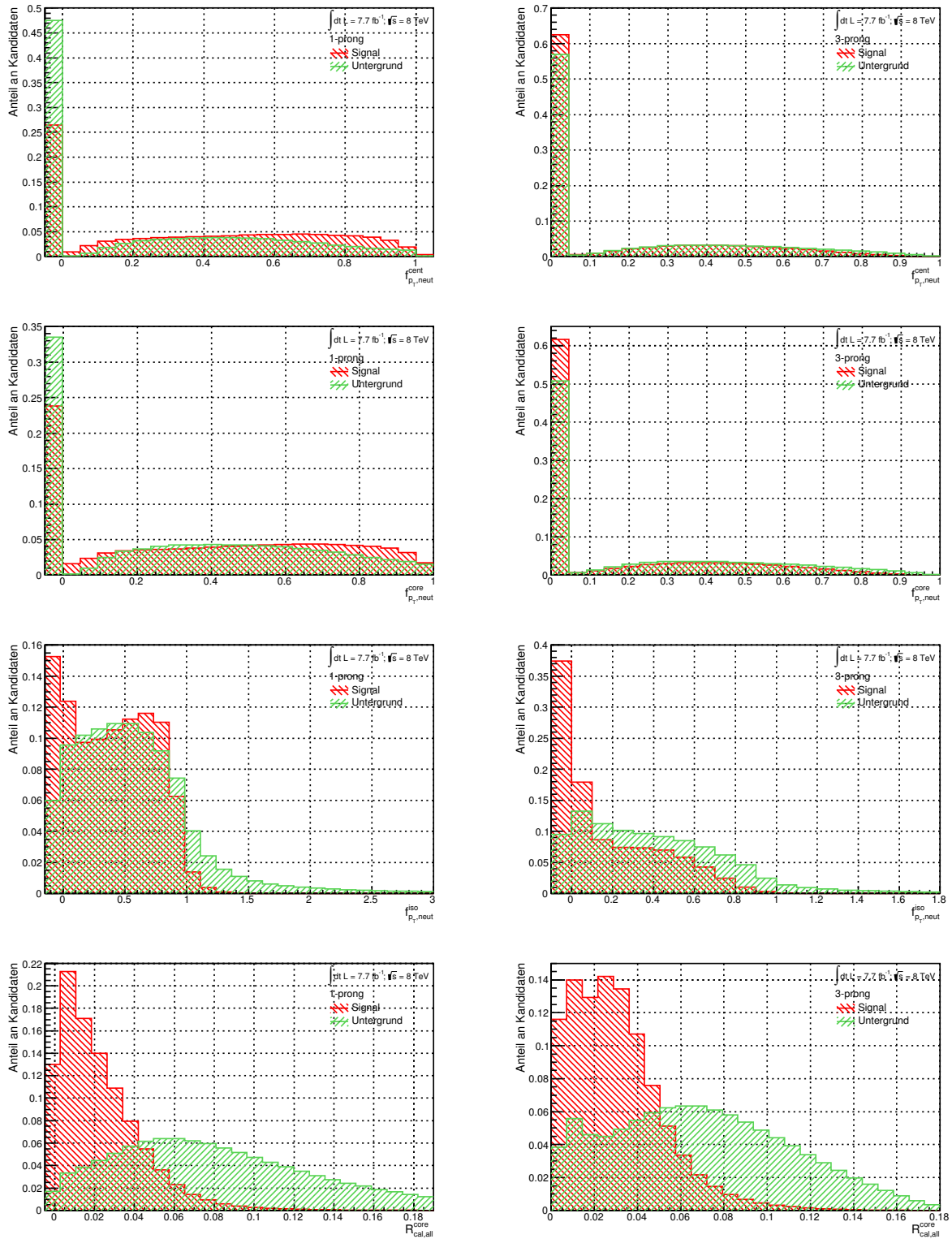


Abb. 36: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil IX

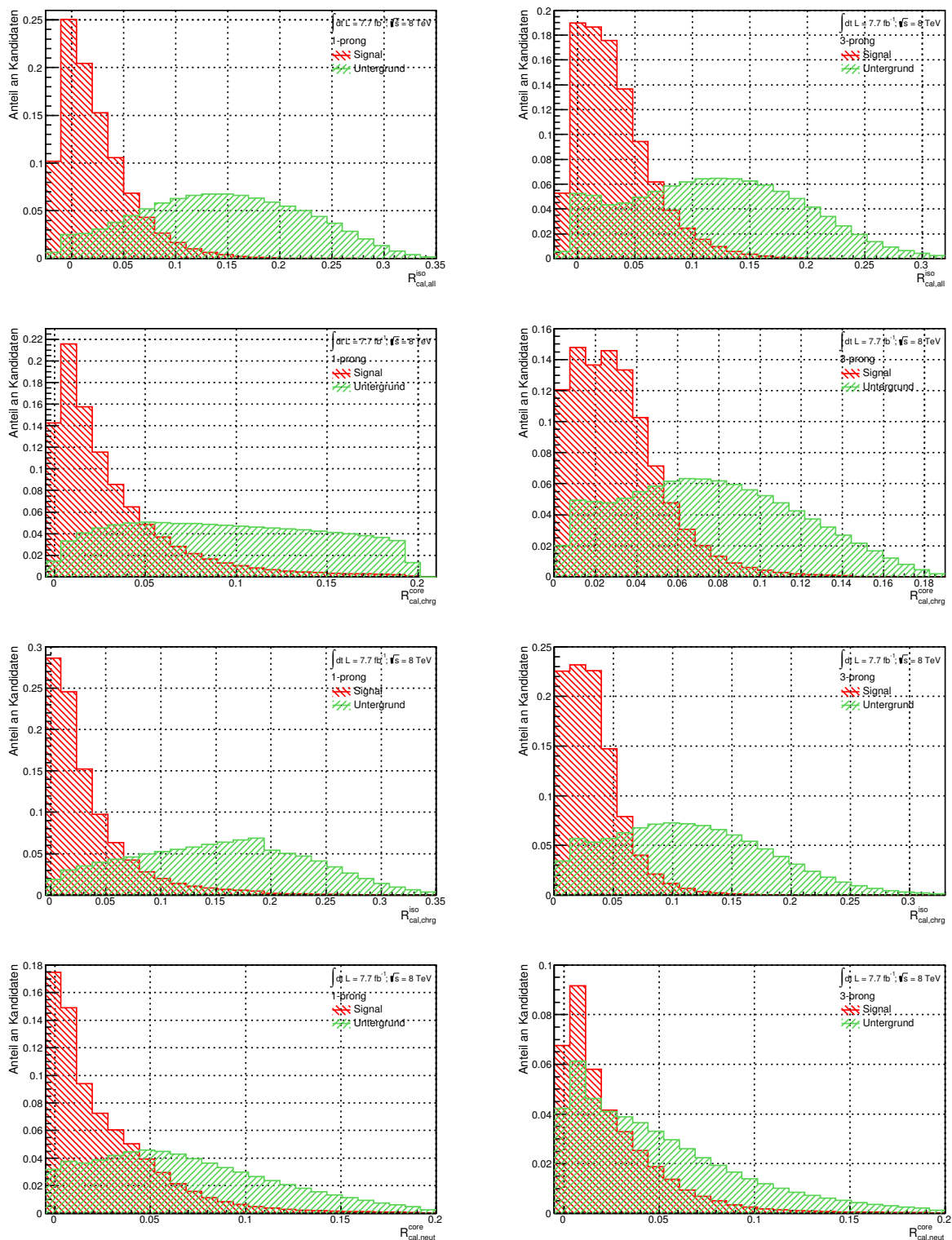


Abb. 37: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil X

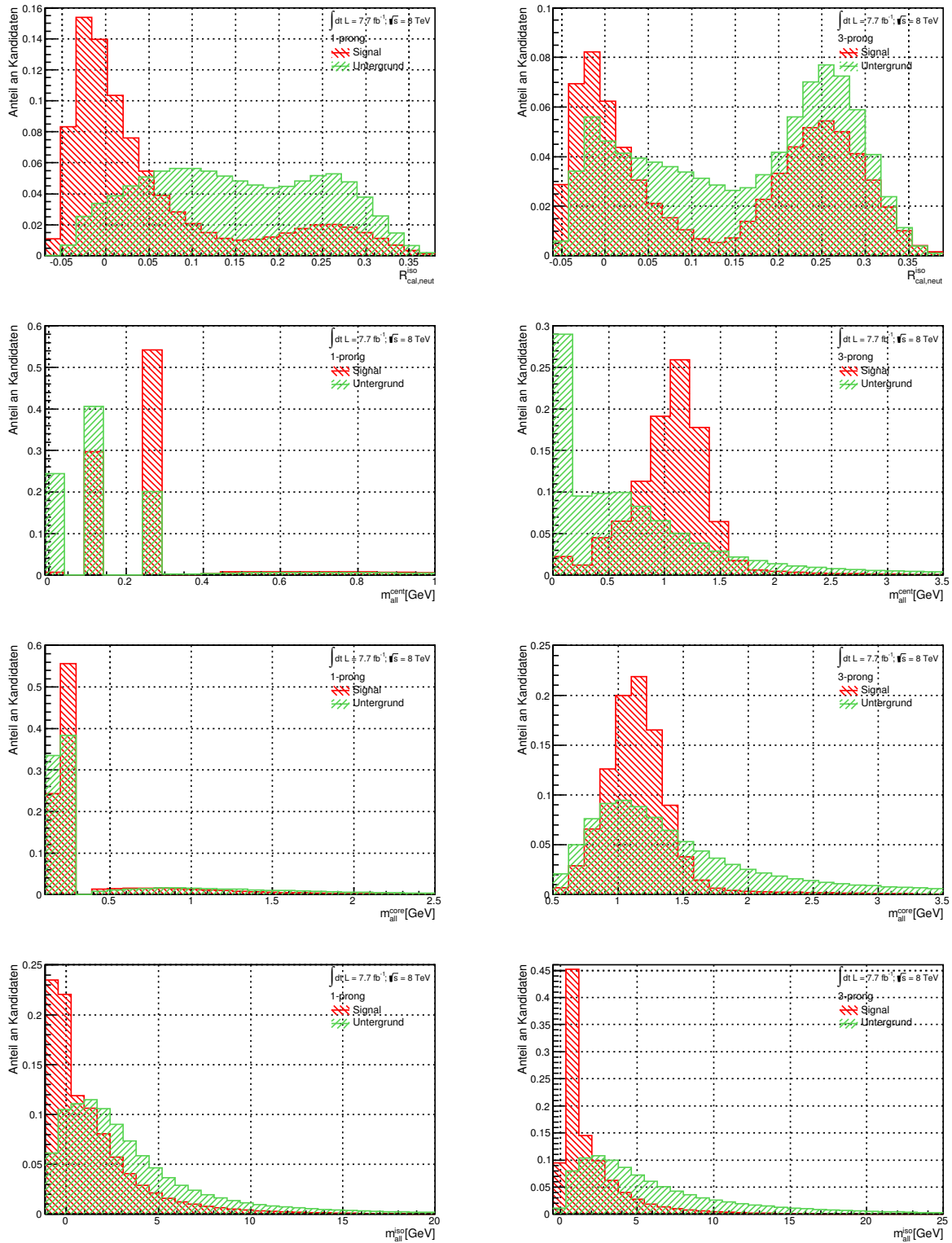


Abb. 38: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil XI

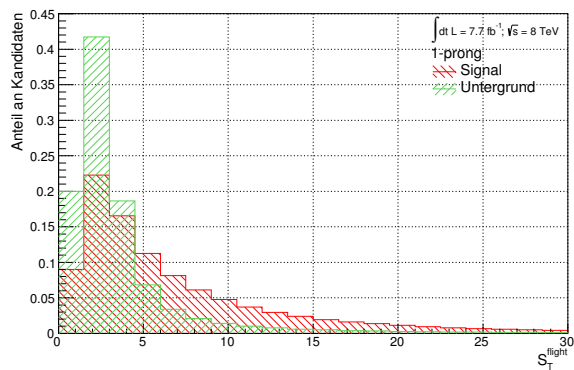


Abb. 39: Verteilungshistogramme der erweiterten Variablen von Signal (rot) und Untergrund (grün) für 1-Prong-Ereignisse (links) und 3-Prong-Ereignisse (rechts), Teil XII

B Literatur

- [1] Olive, K. A./et al. (Particle Data Group): „The Review of Particle Physics“. 2014-2015, Chin. Phys. C38 (2014) 090001.
- [2] Perl, M. L./et al.: „Evidence for Anomalous Lepton Production in e^+e^- Annihilation“. In: Physical Review Letters. 35, Nr. 22, 1975, S. 1489-1492.
- [3] Higgs, Peter: „Broken symmetries and the masses of gauge bosons“. In: Physical Review Letters. Band 13, 1964, S. 508–509.
- [4] Englert, Brout: „Broken Symmetry and the Mass of Gauge Vector Mesons“. In: Physical Review Letters. Band 13, 1964, S. 321–323.
- [5] Kibble, T. W. B.: „Symmetry breaking in non-Abelian gauge theories“. In: Physical Review Letters. 155, 1967, S. 1554.
- [6] Evans, L. R./Bryant, P.: „LHC Machine“. JINST 3 (2008) S08001.
- [7] ATLAS Collaboration: „The ATLAS Experiment at the CERN Large Hadron Collider“. JINST 3 (2008) S08003.
- [8] The ATLAS Collaboration: „Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV“. URL: <http://arxiv.org/pdf/1412.7086.pdf> (Abgerufen: 5.April 2015)
- [9] Hanisch, Stefanie: „Optimisation of the Hadronic Tau Identification Based on the Classification of Tau Decay Modes with the ATLAS Detector“. URL: <http://cds.cern.ch/record/2004897/files/CERN-THESIS-2015-026.pdf?version=1> (Abgerufen: 3.Mai 2015)
- [10] Rey, Günther Daniel/Wender, Karl F.: „Neuronale Netze. Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung“. 2.Auflage. Bern: Huber, 2011.
- [11] Cybenko., G.: „Approximations by superpositions of sigmoidal functions“, Mathematics of Control, Signals, and Systems, 2 (4), 303-314, 1989.
- [12] Kriesel, David: „Ein kleiner Überblick über Neuronale Netze“. URL: http://www.dkriesel.com/_media/science/neuronalenetze-de-zeta2-2col-dkrieselcom.pdf (Abgerufen: 21. April 2015)
- [13] Kinnebrock, Werner: „Neuronale Netze: Grundlagen, Anwendungen, Beispiele“. R. Oldenbourg Verlag, München 1994.
- [14] Broyden, Charles G.: „The convergence of a class of double-rank minimization algorithms“. In: Journal of the Institute of Mathematics and Its Applications. 6, 1970, S. 76–90.

-
- [15] Fletcher, Roger: „A New Approach to Variable Metric Algorithms“. In: *Computer Journal*. 13, Nr. 3, 1970, S. 317–322.
- [16] Goldfarb, Donald: „A Family of Variable Metric Updates Derived by Variational Means“. In: *Mathematics of Computation*. 24, Nr. 109, 1970, S. 23–26.
- [17] Shanno, David F.: „Conditioning of quasi-Newton methods for function minimization“. In: *Mathematics of Computation*. 24, Nr. 111, Juli 1970, S. 647–656.
- [18] Hoecker, A./Speckmayer, P./Stelzer, J./Therhaag, J./von Toerne, E./Voss, H.: „TMVA 4. Users Guide“. URL: <http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf> (Abgerufen: 3.Mai 2015)
- [19] Rokach, Lior/Maimon, O.: „Data mining with decision trees: theory and applications“. World Scientific Pub Co Inc., 2008.
- [20] Roe, Byron/Yang, Hai-Jun/Zhu, Ji: „Boosted Decision Trees, A Powerful Event Classifier“. URL: <http://gallatin.physics.lsa.umich.edu/hyang/papers/phystat05-proc.pdf> (Abgerufen: 09. Mai 2015)
- [21] Autermann, Christian: „Boosted Decision Trees. A modern method of data analysis“. URL: http://wwwiexp.desy.de/users/auterman/talks/20070706_hh_bdt.pdf (Abgerufen: 09. Mai 2015)
- [22] The ROOT team: „ROOT Users Guide 5.26“. Dezember 2009, URL: https://root.cern.ch/download/doc/Users_Guide_5_26.pdf (Abgerufen: 3.Mai 2015)
- [23] Madysa, Nico: „Identification of Hadronic Tau Lepton Decays at the ATLAS Detector Using Artificial Neural Networks“. (In Vorbereitung)

Abbildungsverzeichnis

1	Das Standardmodell der Elementarteilchen	6
2	Querschnitt des ATLAS-Detektors	7
3	Aufbau eines neuronalen Netzes	12
4	Probleme des Gradientenverfahrens	14
5	Aufbau von Entscheidungsbäumen	16
6	Verteilungshistogramme der Variablen $N_{\text{chrg}}^{\text{cent}}$	18
7	Verteilungshistogramme der Variablen $N_{\text{chrg}}^{\text{core}}$	19
8	Verteilungshistogramme der Variablen $m_{\text{chrg}}^{\text{iso}}$	20
9	Verteilungshistogramme der Variablen m_{tracks}	20
10	Verteilungshistogramme der Variablen $f_{p_{\text{T}},\text{all}}^{\text{core}}$	21
11	Verteilungshistogramme der Variablen $f_{\text{vis-}p_{\text{T}}}$	21
12	ROC-Kurve für MLP mit Standardvariablen und $f_{p_{\text{T}},\text{all}}^{\text{core}}$	22
13	ROC-Kurve für MLP mit Standardvariablen und $f_{E_{\text{T}},\text{all}}^{\text{cent-core/core-iso/cent-iso}}$	23
14	Verteilungshistogramme der Variablen in Abb. 13 (S. 23)	24
15	ROC-Kurve für MLP mit Standardvariablen und $N_{\text{neut}}^{\text{cent}}$	25
16	ROC-Kurven für MLPs bei Variation der verborgenen Schichten, Teil I	26
17	ROC-Kurven für MLPs bei Variation der verborgenen Schichten, Teil II	26
18	ROC-Kurven für MLPs bei Variation der verborgenen Schichten, Teil III	27
19	ROC-Kurven für MLPs mit unterschiedlichen Anfangswerten	27
20	ROC-Kurven für beste 5 MLPs bei Hinzufügen einer Variablen	29
21	ROC-Kurven für beste MLPs bei Austausch einer alten durch eine neue Variable	29
22	ROC-Kurven für beste 5 BDTs bei Hinzufügen einer Variablen	31
23	ROC-Kurven für beste BDTs bei Hinzufügen einer Kombination von Variablen	32
24	Linearen Korrelation der besten Variablenkombination für BDTs	33
25	Verteilungshistogramme der Standardvariablen, Teil I	35
26	Verteilungshistogramme der Standardvariablen, Teil II	36
27	Verteilungshistogramme der Standardvariablen, Teil III	37
28	Verteilungshistogramme der erweiterten Variablen, Teil I	38
29	Verteilungshistogramme der erweiterten Variablen, Teil II	39
30	Verteilungshistogramme der erweiterten Variablen, Teil III	40
31	Verteilungshistogramme der erweiterten Variablen, Teil IV	41
32	Verteilungshistogramme der erweiterten Variablen, Teil V	42
33	Verteilungshistogramme der erweiterten Variablen, Teil VI	43
34	Verteilungshistogramme der erweiterten Variablen, Teil VII	44
35	Verteilungshistogramme der erweiterten Variablen, Teil VIII	45
36	Verteilungshistogramme der erweiterten Variablen, Teil IX	46
37	Verteilungshistogramme der erweiterten Variablen, Teil X	47

38	Verteilungshistogramme der erweiterten Variablen, Teil XI	48
39	Verteilungshistogramme der erweiterten Variablen, Teil XII	49

Tabellenverzeichnis

1	Zuordnung der MLPs zur für Identifikation verwendeten Variablenkonfiguration . . .	23
2	Zuordnung der MLPs zur für Identifikation genutzten Netzwerkarchitektur	25
3	Variablenranking der MLPs bei Hinzufügen einer Variablen	28
4	Variablenranking der MLPs bei Austausch einer Variablen	29
5	Variablenranking der BDTs bei Hinzufügen einer Variablen	30
6	Variablenranking der BDTs bei Hinzufügen einer Kombination von Variablen	31

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit im Rahmen der Betreuung am Institut für Kern- und Teilchenphysik ohne unzulässige Hilfe Dritter verfasst und alle Quellen als solche gekennzeichnet habe.

Tony Henseleit
Dresden, 06. Juni 2015