

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG



Timo Steinsberger

Datenbasierte Bestimmung des $t\bar{t}$ -Untergrundes für die Messung der
Produktion von W-Bosonen in Assoziation mit Jets am
ATLAS-Experiment bei $\sqrt{s} = 8 \text{ TeV}$

Bachelorarbeit

HD-KIP-15-78

KIRCHHOFF-INSTITUT FÜR PHYSIK

**Fakultät für Physik und Astronomie
Ruprecht-Karls-Universität Heidelberg**

Bachelorarbeit im Studiengang Physik
vorgelegt von

Timo Pascal Steinsberger

geboren in Esslingen am Neckar

2015

Datenbasierte Bestimmung des $t\bar{t}$ -Untergrundes für die
Messung der Produktion von W-Bosonen in Assoziation mit
Jets am ATLAS-Experiment bei $\sqrt{s} = 8 \text{ TeV}$

Diese Bachelorarbeit wurde erstellt von
Timo Pascal Steinsberger
am Kirchhoff-Institut für Physik in Heidelberg
unter der Betreuung von
Frau Dr. Monica Dunford

Abstract

The study of the production of W-bosons in association with jets, so called W+jets events, is an important test of perturbative Quantum Chromodynamics and makes it possible to examine the parton distribution function of the proton. Furthermore these events are an important background in studies on Higgs-Bosons and in the search for physics beyond the Standard Model. A crucial part of W+jets studies is the accurate estimation of the background processes. For events with four or more jets, the production of pairs of top and antitop quarks, so called $t\bar{t}$ -events, form the major background. In this thesis, the $t\bar{t}$ -background in W+jets candidate events measured at the ATLAS experiment at a center of mass energy of $\sqrt{s}=8\text{ TeV}$ in 2012 is determined with a data-driven method. Two fit procedures have been tested: a template likelihood fit and a χ^2 -fit with functional forms. This is done separately for the leptonic W decays $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ for events with three, four and five jets exclusively and for events with more than six jets inclusively. The functional form fit turns out to be less susceptible to systematic uncertainties and therefore allows to reduce the uncertainties on the W+jets cross section.

Zusammenfassung

Die Untersuchung der Produktion von W-Bosonen in Verbindung mit Jets, sogenannte W+Jets Ereignisse, ermöglicht einen wichtigen Test der perturbativen Quantenchromodynamik und eine Untersuchung der Substruktur von Protonen. Außerdem bilden W+Jets Ereignisse einen wichtigen Untergrund in Studien mit Higgs-Bosonen und bei der Suche nach neuer Physik jenseits des Standardmodells. Die Grundlage für Studien von W+Jets Ereignissen ist die präzise Abschätzung des Untergrundes. In W+Jets Kandidaten mit vier oder mehr Jets stellt dabei die Bildung von Top-Antitopquark-Paaren, sogenannte $t\bar{t}$ -Ereignisse, den größten Untergrund dar. In dieser Arbeit wird der $t\bar{t}$ -Untergrund in W+Jets Kandidaten bestimmt, die 2012 am ATLAS-Experiment bei einer Schwerpunktsenergie von $\sqrt{s}=8\text{ TeV}$ aufgezeichnet wurden. Hierzu wird eine datenbasierte Methode verwendet. Die Untergrundbestimmung wird für Ereignisse mit drei, vier, fünf und mit mehr als sechs Jets sowohl für den Zerfallskanal $W \rightarrow e\nu$ als auch für den Zerfallskanal $W \rightarrow \mu\nu$ durchgeführt. Hierzu wird einmal ein Likelihood-Fit und einmal ein χ^2 -Fit mit Funktionen benutzt. Dabei erweist sich der Fit mit Funktionen als weniger anfällig gegenüber systematischen Unsicherheiten und ermöglicht es somit, die Unsicherheiten auf den W+Jets Wirkungsquerschnitt zu reduzieren.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Das Standardmodell der Teilchenphysik	1
1.2	QCD	2
1.3	Das ATLAS-Experiment	3
2	W+Jets und Untergrundprozesse	7
2.1	W+Jets Ereignisse	7
2.2	Untergrundprozesse	8
2.3	$t\bar{t}$ -Untergrund	8
2.4	Multijet-Untergrund	9
2.5	Weitere Untergrundprozesse	9
2.6	Definition der Signalregion	9
3	Multijet-Untergrundabschätzung	12
3.1	Motivation	12
3.2	Multijet-Templates	12
3.3	Fit zur Bestimmung des Multijet-Anteils	13
3.4	Ergebnisse	14
4	$t\bar{t}$-Untergrundabschätzung	15
4.1	Motivation	15
4.2	Transformierte Aplanarität	15
4.3	Gewinnung datenbasierter Templates	16
4.4	Fit mit TFractionFitter	18
4.5	Fit mit Funktionen	22
4.6	Fitbezogene Fehlerquellen	29
4.7	Systematische Unsicherheiten	30
4.8	Gesamtunsicherheit des $t\bar{t}$ -Anteils in der Signalregion	33
4.9	Messwerte	33
4.10	2-b-Jet-Kriterium in der $t\bar{t}$ -Kontrollregion	37
4.11	Konsequenzen für den W+Jets Wirkungsquerschnitt	39
5	Zusammenfassung	41
A	Anhang	43

Kapitel 1

Einleitung

1.1 Das Standardmodell der Teilchenphysik

Bis zu Beginn des 20. Jahrhunderts galten Atome (griech. atomos = unteilbar) als die elementaren Bestandteile unserer Welt. Experimente mit kosmischer Strahlung und Beschleunigern zeigten jedoch, dass auch diese eine Substruktur besitzen und aus noch kleineren Bausteinen bestehen, den sogenannten Elementarteilchen. Drei der vier elementaren Kräfte konnten mittels Quantenfeldtheorien mit korrespondierenden Teilchen, sogenannten Eichbosonen verknüpft werden. Die heute bekannten Teilchen werden im Standardmodell der Teilchenphysik (SM, siehe Abbildung 1.1) zusammengefasst.

Das Standardmodell umfasst zwei Gruppen von Teilchen, die Fermionen mit Spin $1/2$ und die Eichbosonen mit ganzzahligem Spin.

Die Grundbausteine der Materie sind Fermionen. Diese lassen sich in Quarks und Leptonen einteilen. Quarks treten immer in gebundenen Zuständen, sogenannten *Hadronen*, auf. Sie besitzen Ladungen von $+2/3$ und $-1/3$ und kommen in drei Generationen vor. Dabei unterscheiden sich die einander entsprechenden Quarks unterschiedlicher Generationen in ihrer Masse. Die einzigen beiden stabilen Hadronen sind Neutronen und Protonen. Dies bestehen aus den beiden Quarks der leichtesten Generation, dem Up- (u) und dem Down-Quark (d). Die Quarks der zweiten Generation sind das Charm- (c) und das Strangequark (s) und die der dritten Generation das Top- (t) und das Bottom-Quark (b).

Auch Leptonen kommen in drei Generationen vor. Hier gibt es zum einen die drei geladenen Leptonen Elektron (e), Myon (μ) und das schwere Tauon (τ). Sie tragen eine elektrische Ladung von -1 . Zu jedem geladenen Lepton korrespondiert ein ungeladenes und sehr leichtes Neutrino (Elektron-neutrino ν_e , Myonneutrino ν_μ und Tauneutrino ν_τ). Diese wechselwirken nur schwach mit anderen Teilchen und kommen daher nicht in gebundenen Zuständen vor.

Zu jedem Fermion (f) korrespondiert auch ein Antiteilchen (\bar{f}). Teilchen und Antiteilchen unterscheiden sich durch das Vorzeichen ihrer elektrischen Ladung und können in Paaren gebildet werden.

Die Eichbosonen übertragen elementare Kräfte. Das masselose Photon koppelt an die elektrische Ladung und korrespondiert zur elektromagnetischen Wechselwirkung, die eine unendliche Reichweite besitzt. Dagegen besitzen die schweren W^\pm - und Z^0 -Bosonen aufgrund ihrer Masse nur eine

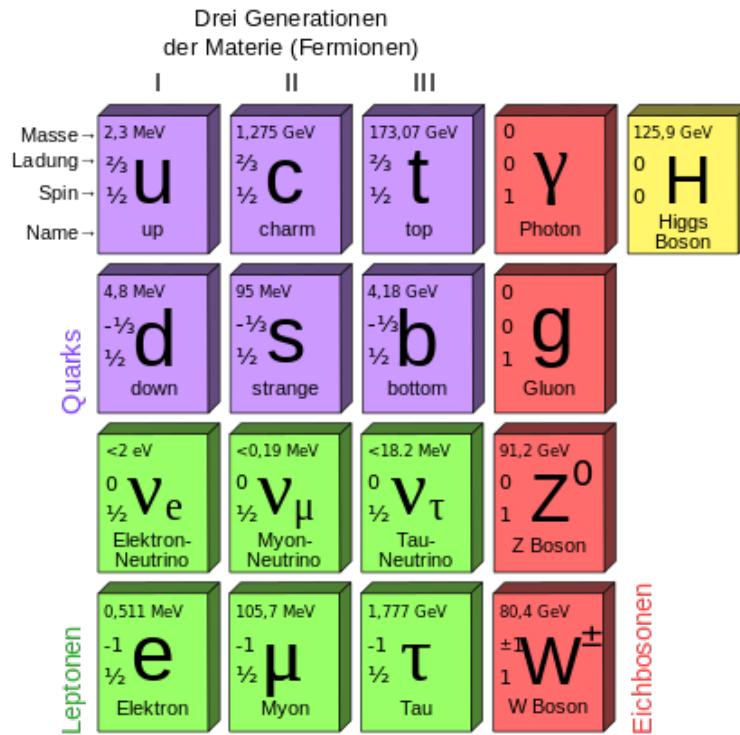


Abbildung 1.1: Das Standardmodell der Teilchenphysik. Die Grafik ist aus [24] entnommen.

sehr geringe Reichweite. Sie übertragen die schwache Kraft, die etwa für den radioaktiven Zerfall verantwortlich ist. Sie koppeln an die sogenannte *Hyperladung*. Elektromagnetische und schwache Kraft wurden in der elektroschwachen Theorie zusammengefasst. Die masselosen Gluonen (g) korrespondieren zur starken Kraft, die Quarks und die aus ihnen gebildeten Baryonen in Atomkernen zusammenhält. Auch ihre Reichweite ist begrenzt. Die starke Wechselwirkung wird von der Quantenchromodynamik (QCD) beschrieben, in der Quarks eine Farbladung tragen. Das Higgs-Boson (H) wurde als letztes Elementarteilchen nachgewiesen. Es verleiht den anderen Elementarteilchen ihre Masse.

Das Standardmodell ist zwar sehr erfolgreich in der Beschreibung der meisten Prozesse, allerdings offenbar unvollständig. So ist in ihm keine Quantenfeldtheorie für die Gravitation enthalten. Auch die in der Astronomie beobachtete dunkle Materie und dunkle Energie werden nicht durch das Standardmodell beschrieben. Daher wird an theoretischen Erweiterungen gearbeitet und experimentell nach neuen Teilchen gesucht. Als besonders aussichtsreich gelten dabei supersymmetrische Theorien [23, S. 501ff].

1.2 QCD

Die Quantenchromodynamik beschreibt die Wechselwirkung zwischen Quarks. Diese tragen als einzige Fermionen eine *Farbladung*, an die die sogenannten *Gluonen* koppeln. Das besondere an der starken Wechselwirkung ist, dass das Potential näherungsweise linear [23, S. 271ff.] mit zunehmendem Abstand der gebundenen Teilchen steigt. Wird die Entfernung gesteigert, ist zu einem

bestimmten Zeitpunkt genug potentielle Energie vorhanden, um ein Quark-Antiquark-Paar zu bilden. Jedes der beiden ursprünglichen Quarks wird dann von einem der beiden neuen begleitet. Daher treten Quarks immer in farbneutralen, gebundenen Zuständen, sogenannten Hadronen, auf. Dieses Phänomen wird *Confinement* genannt. Wird ein solcher Zustand durch einen Zerfall oder eine Kollision aufgebrochen, so entstehen zahlreiche neue Quarks, die sich wieder zu farbneutralen Hadronen zusammenfinden. Dieser Prozess wird als *Hadronisierung* bezeichnet. Die Menge der so entstandenen Teilchen fliegt in eine Richtung und bildet einen sogenannten *Jet*.

Die elektromagnetische und die schwache Kraft koppeln nur schwach an Fermionen. Für niedrige Energieskalen wird die schwache Wechselwirkung zusätzlich durch die große Masse von W- und Z-Bosonen unterdrückt. Daher konvergiert bei störungstheoretischen Berechnung der Wirkungsquerschnitten bereits nach Prozessen mit ein oder zwei elektroschwachen Wechselwirkungen. Die Kopplungskonstante der starken Wechselwirkung ist dagegen für niedrige Energieskalen $\alpha_s \approx 1$. Allerdings sinkt sie für große Schwerpunktsenergien so weit ab, dass wieder störungstheoretische Berechnungen durchgeführt werden können. Man spricht dann von perturbativer Quantenchromodynamik (pQCD). Diese kann mit Monte Carlo Generatoren simuliert werden. Dabei unterscheidet man zwischen leading-order-Generatoren, die nur Prozesse erster Ordnung berücksichtigen, und next-to-leading-order-Generatoren, die auch Prozesse zweiter Ordnung berechnen. Um diese Generatoren zu verbessern, ist es wichtig, sie regelmäßig mit Daten zu vergleichen. Zuverlässige Monte Carlo Simulationen sind auch deshalb wichtig, weil die QCD aufgrund ihrer starken Kopplung in den meisten Zerfallsprozessen involviert ist. So treten immer Jets als Zerfalls- oder Nebenprodukte bei der Kollision von Hadronen auf. In Analysen müssen die Jets aus den untersuchten Prozessen von Jets aus Untergrundprozessen unterschieden werden. Dies trifft insbesondere auf Protonenbeschleuniger wie den Large Hadron Collider am CERN zu, da hier Quarks und Gluonen mit Farbladung interagieren.

Eine bestimmte Art von QCD Prozessen sind die *W+Jets Ereignisse*, in denen sowohl ein W-Boson aus der schwachen Wechselwirkung, als auch Jets, auftreten. Sie ermöglichen einen Test der theoretischen Vorhersagen. Um derartige Analysen zu ermöglichen, wird in dieser Arbeit ein großer Untergrund zu W+Jets Ereignissen, der sogenannte *$t\bar{t}$ -Untergrund* abgeschätzt.

1.3 Das ATLAS-Experiment

1.3.1 Der Large Hadron Collider LHC

Der Large Hadron Collider (LHC) am europäischen Kernforschungszentrum CERN in Genf ist der momentan größte Teilchenbeschleuniger der Welt. In ihm werden Protonen in einem Ring in Paketen, sogenannten *Bunches*, beschleunigt und zur Kollision gebracht. Die Kollisionen finden im Innern von vier großen Detektoren statt: ATLAS, CMS, ALICE und LHCb. Die in dieser Arbeit untersuchten Daten wurden 2012 bei einer Schwerpunktennergie von $\sqrt{s} = 8 \text{ TeV}$ mit dem ATLAS Detektor aufgenommen.

1.3.2 Der ATLAS-Detektor

Der ATLAS-Detektor (siehe Abbildung 1.2, Daten und Aufbau siehe [8]) ist auf die Beobachtung möglichst vieler verschiedener Prozesse optimiert. Er besitzt einen zylindrischen Aufbau mit einer

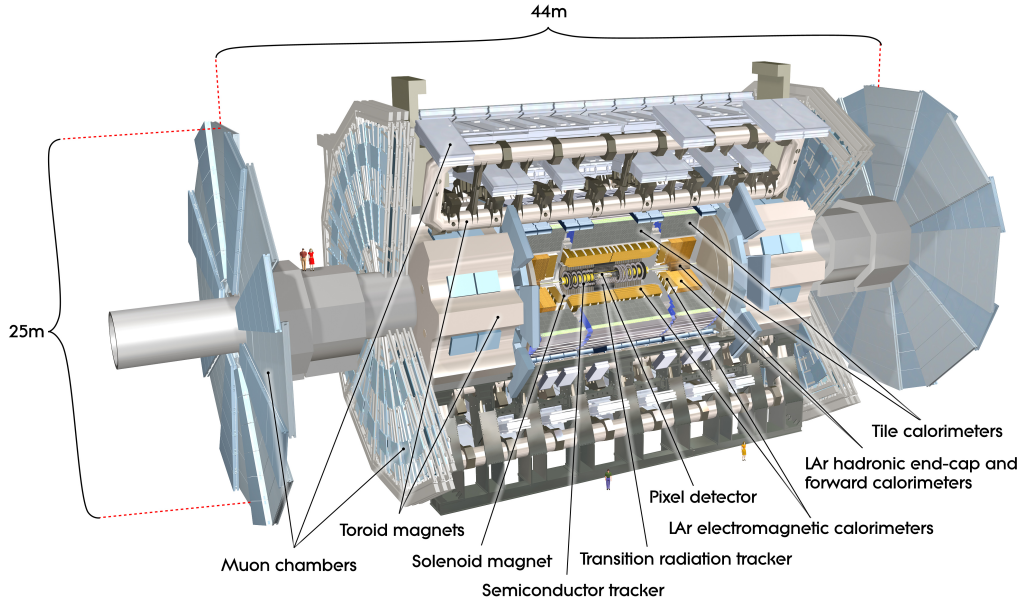


Abbildung 1.2: Der ATLAS Detektor [19]

Länge von 44 m und einem Durchmesser von 22 m. Er gliedert sich in zwei *Endkappen* und den Mantel, der *Barrel* genannt wird. Für die weitere Beschreibung wird folgendes rechtshändiges Koordinatensystem verwendet: Die x-Achse zeigt vom Interaktionspunkt zum Mittelpunkt des LHCs, die y-Achse zeigt nach oben und die z-Achse verläuft entlang der Strahlrichtung. x- und y-Achse definieren die transversale Ebene, in der der Azimutwinkel ϕ gemessen wird. Der Winkel zur z-Achse wird mit θ bezeichnet. Zur Beschreibung physikalischer Prozesse ist jedoch die Pseudorapidität

$$\eta = -\ln \tan \frac{\theta}{2} \quad (1.1)$$

als Maß für den Winkel üblicher. Dabei bezeichnet $\eta = 0$ einen Winkel senkrecht zur Strahlrichtung. Für vernachlässigbare Massen nähert sich η der Rapidität

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (1.2)$$

an. Um den Winkelabstand zweier Trajektorien im Detektor anzugeben wird der Abstand

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (1.3)$$

benutzt.

Die Hauptbestandteile des Detektors sind der innere Spurdetektor, die Kalorimeter und die Myonenkammern. Zudem erzeugen ein Solenoid ein Magnetfeld mit einer Flussdichte von 2 Tesla im Spurdetektor und Toroidmagneten ein Magnetfeld von 0,5-1 Tesla in den Myonenkammern.

Der innere Spurdetektor besteht aus einem Silizium-Pixel-Detektor, einem Silizium-Streifen-Detektor und einem Übergangsstrahlungsdetektor (TRT). Die Aufgabe des Spurdetektors ist die präzise Vermessung der Trajektorien der entstandenen Teilchen um den Kollisionspunkt zu rekonstruieren,

geladene Teilchen zu identifizieren und deren Impuls mit Hilfe der Krümmung in einem Magnetfeld zu vermessen.

Der innere Spurdetektor ist von einem elektromagnetischen Blei-Flüssigargon(LAr)-Kalorimeter umgeben. In diesem geben Elektronen und Photonen den Großteil ihrer Energie ab. Die Projektion der Energie in die transversale Ebene $E_T = E \cdot \sin \theta$ kann gemessen werden. Hadronen geben hingegen nur einen vergleichsweise geringen Teil ihrer Energie ab und erreichen das hadronische Kachelkalorimeter. Dieses verwendet Eisen als Absorbermaterial und misst die Energie der Hadronen. Die beiden Kalorimeter kommen sowohl im Barrel als auch in den Endkappen vor.

Myonen durchdringen als einzige detektierbare Teilchen beide Kalorimeter. Sie werden in den Myonenkammern über Ionisation von Gas nachgewiesen. Das Magnetfeld zwingt sie auf eine gekrümmte Bahn, sodass ihr Impuls vermessen werden kann. Die Myonenkammern bestehen dabei aus vier Arten von Teildetektoren: Driftröhren (*Monitored Drift Tubes* MDTs) decken den größten Bereich von $|\eta| < 2,7$ ab. In den Endkappen erhöhen Kathoden-Streifenkammern (*Cathode Strip Chambers* CSCs) die Auflösung im Bereich $2,0 < |\eta| < 2,7$. Hinzu kommen Widerstandsplattenkammern (*Resistive Plate Chambers* RPCs) im Barrel bei $|\eta| < 1,05$ und *Thin Gap Chambers* (TGCs) im Bereich $1,05 < |\eta| < 2,7$.

Neutrinos können nicht direkt detektiert werden. Da die in Proton-Proton-Kollisionen interagierenden Partonen einen vernachlässigbaren Impuls senkrecht zur Strahlrichtung tragen, müssen sich sämtliche Impulse in der transversalen Ebene vektoriell zu null addieren. Die von Neutrinos getragene transversale Energie E_T^{miss} ist dann der fehlende Beitrag zur in den Kalorimetern und in den Myonenkammern gemessenen Energie.

1.3.3 Das ATLAS Triggersystem

Das ATLAS Experiment benutzt ein dreistufiges Triggersystem, um interessante Daten zu selektieren und die Datenmengen auf ein verarbeitbares Maß zu reduzieren. Der Level-1 Trigger (L1) basiert auf programmierbarer Hardware und trifft innerhalb von $2,5 \mu\text{s}$ eine Vorauswahl. Hierzu nutzt er nur einen Teil der Daten und definiert sogenannte *regions of interest* (RoI), in denen sich detektierte Teilchen befinden. Die Ereignisrate wird so auf 75 kHz reduziert. Anschließend nutzt der Level-2-Trigger (L2) die vollen Informationen aus den RoIs, um die Daten weiter zu selektieren. So wird die Rate auf 3,5 kHz verringert. Abschließend rekonstruiert der Event Filter (EF) ganze Ereignisse und trifft dann eine Endauswahl mit einer Rate von nur noch ca. 200 Hz. L2 und EF werden zum High-Level Trigger (HLT) zusammengefasst. Die verbleibenden Daten werden vollständig aufgezeichnet und stehen für Analysen zur Verfügung.

Der Trigger ordnet den gemessenen Signalen sogenannten *Objekte* zu. Dies sind Leptonen, Photonen, Hadronen oder Jets. Hierzu rekonstruiert er die Trajektorien der gemessenen Teilchen. Orte, an denen ein Teilchen zerfällt und sich die Zerfallsprodukte trennen, werden als *Vertices* bezeichnet. Die Rekonstruktion der Jets wird mit dem Anti- k_t -Algorithmus [7] durchgeführt. Im Kollisionspunkt der beiden Protonen stellt den Ursprung der Zerfallsketten dar und wird daher als *primärer Vertex* bezeichnet. Der Jet mit dem größten transversalen Impuls wird *führender Jet* genannt. Treffen zwei Bunches aufeinander, so kann es zu mehreren Proton-Proton-Kollisionen kommen, die sich überlagern. Hierbei spricht man von *Pile-up*.

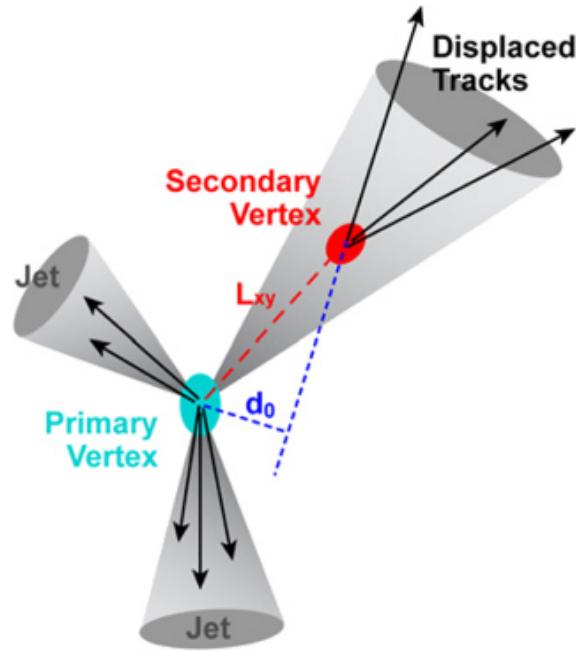


Abbildung 1.3: Identifikation eines sekundären Vertex aus dem Zerfall eines b-Hadrons. Die Grafik wurde aus [13] entnommen.

1.3.4 b-Tagging

Eine für diese Arbeit relevante Methode ist die Identifikation von sogenannten b-Jets. Dies sind Jets, die durch ein b-Quark initiiert werden. b-Hadronen besitzen eine Lebensdauer von typischer Weise etwa 1,5 ps, in denen sie in der transversalen Ebene bei einem typischen Impuls von $p_T = 70 \text{ GeV}/c$ eine Strecke von $L_{xy} \approx 6,4 \text{ mm}$ vom primären Vertex aus zurücklegen [9]. Beim Zerfall des b-Hadrons entstehen neue Teilchen. Deren Trajektorie wird nun rekonstruiert und ihr geringster Abstand zum primären Vertex, der *impact parameter* d_0 (siehe Abbildung 1.3), berechnet. Ist dieser signifikant von Null verschieden, so kann der sekundäre Vertex identifiziert und einem b-Jet zugeordnet werden. Auch die invariante Masse der Zerfallsprodukte von typischer Weise 5-6 GeV sind eine Signatur eines b-Hadrons. Zudem können auch Leptonen mit ungewöhnlich niedrigem Impuls, sogenannte *soft leptons* entstehen. Der verwendete b-Tagging Algorithmus MV1 nutzt alle drei Signaturen. Seine Akzeptanz lässt sich so einstellen, dass ein bestimmter Anteil tatsächlicher b-Jets als solche erkannt wird. Bei zunehmender Akzeptanz steigt allerdings auch der Anteil falscher Identifikationen. Für Effizienzen von 60, 70, 80 und 85% sind kalibrierte Schwellwerte vorhanden. Sie werden als *working points* bezeichnet.

Kapitel 2

W+Jets und Untergrundprozesse

2.1 W+Jets Ereignisse

W+Jets Ereignisse sind die Produktion von W-Bosonen am primären Vertex in Assoziation mit Jets. W-Bosonen zerfallen zu zwei Dritteln hadronisch. Dies lässt sich allerdings kaum von reinen QCD-Prozessen unterscheiden. Der interessantere Fall ist daher die zweite Möglichkeit: ein Zerfall in ein geladenes Lepton und das dazu gehörige Antineutrino. Die entsprechende Signatur ist ein isoliertes Lepton, fehlende transversale Energie E_T^{miss} aufgrund des Neutrinos und Jets. Tauonen können aufgrund ihrer kurzen Lebensdauer von nur $3,4 \cdot 10^{-13}$ s nur über ihre Zerfallsprodukte detektiert werden. Die stabilen Elektronen und die langlebigen Myonen erreichen hingegen den Detektor. Sie können dazu genutzt werden, W+Jets Ereignisse zu identifizieren. In dieser Arbeit wird sowohl der elektronische Zerfallskanal $W \rightarrow e\nu$, als auch der myonische Zerfallskanal $W \rightarrow \mu\nu$ untersucht. In Abbildung 2.1 sind beispielhaft einige Produktionsprozesse von W+Jets Ereignissen dargestellt. Dabei steuert häufig ein Valenzquark den nötigen Impuls bei. Die elektrische Ladung des W-Bosons und somit des Leptons ist in den beiden linken Prozessen abhängig von der Ladung des Quarks, weshalb durch den Vergleich von W^+ und W^- Rückschlüsse auf die Partonenverteilung im Protonen gezogen werden können.

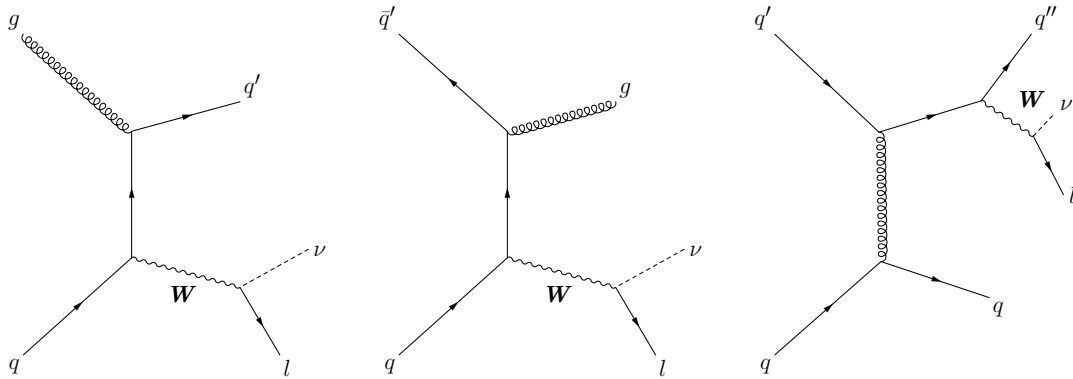


Abbildung 2.1: Produktion von W+Jets Ereignissen mit leptonischem W-Zerfall durch Wechselwirkung von Quarks und Gluonen (links), Quarks und Antiquarks (Mitte) und zwischen zwei Quarks durch den Austausch eines Gluons (rechts). Die entstandenen Quarks und Gluonen bilden Jets.

Untersuchungen von W+Jets Ereignissen ermöglichen einen Test der Vorhersagen von Mon-

te Carlo Simulationen, welche sich perturbativer Quantenchromodynamik (pQCD) bedienen. Mit den in dieser Arbeit verwendeten Generatoren ALPGEN [17] für die harte Wechselwirkung und PYTHIA6 für die Partonenschauer sind zwar Simulationen vorhanden, die die Daten vorheriger Experimente im Rahmen der Messgenauigkeit beschreiben. Allerdings treten Unterschiede zu anderen Generatoren wie SHERPA [14] auf. Zudem sind W+Jets Ereignisse ein wichtiger Untergrund zu einigen interessanten Prozessen, wie zum Beispiel der Bildung von einzelnen Topquarks, Higgs-Bosonen oder postulierten supersymmetrischen Teilchen. Daher ist das Ziel, eine möglichst genaue Messung der W+Jets Produktion durchzuführen. Hierfür muss jedoch der Untergrund präzise bestimmt werden. Diese Arbeit widmet sich dabei dem sogenannten $t\bar{t}$ -Untergrund.

Im Laufe der Untergrundbestimmung wird eine Monte Carlo Simulation von W+Jets Ereignissen als Schätzung benötigt. Hierzu werden die Generatoren ALPGEN und PYTHIA6 verwendet.

2.2 Untergrundprozesse

Um W+Jets Ereignisse möglichst genau untersuchen zu können, muss die Kontamination der W+Jets-Selektion mit Untergrundprozessen abgeschätzt werden. Diese Arbeit konzentriert sich dabei auf den $t\bar{t}$ -Untergrund (siehe Kapitel 4), der in Ereignissen mit mehr als fünf Jets den größten Untergrund darstellt. Hierfür muss zuvor auch der Multijet-Untergrund bestimmt werden (siehe Kapitel 3). Für diesen wird jedoch keine separate Diskussion der systematischen Unsicherheiten durchgeführt, da diese mit denen des $t\bar{t}$ -Untergrundes korreliert sind und deswegen dort eingehen (siehe Kapitel 4.7). Hinzu kommen noch einige weitere Prozesse mit geringerer Häufigkeit. Im Folgenden wird ein Überblick über alle berücksichtigten Untergrundprozesse gegeben.

2.3 $t\bar{t}$ -Untergrund

Ein wichtiger Untergrund zu W+Jets Ereignissen ist die Produktionen von Top-Antitop-Paaren. Drei wichtige Produktionsprozesse sind in Abbildung 2.2 dargestellt. Aufgrund ihrer extrem kurzen

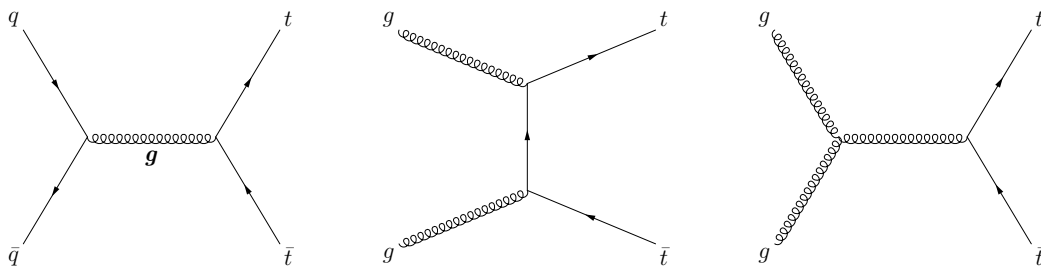


Abbildung 2.2: Produktionsprozesse von $t\bar{t}$ -Paaren bei der Kollision zweier Protonen.

Lebensdauer von nur $6 \cdot 10^{-25}$ s zerfallen t- und \bar{t} -Quark, bevor sie Hadronen bilden können. Das (Anti-)Topquark zerfällt dabei zu $91 \pm 4\%$ [15] in ein (Anti-)b-Quark und ein W-Boson. Daher treten $t\bar{t}$ -Ereignisse besonders bei höheren Jetmultiplizitäten auf. Die typische Signatur eines $t\bar{t}$ -Ereignisses sind genau wie für W+Jets Ereignisse Jets, Leptonen und fehlende transversale Energie. Ab einer Jetmultiplizität von vier treten mehr $t\bar{t}$ -Ereignisse als W+Jets Ereignisse auf.

Die Tatsache, dass mit dem Zerfall von Topquarks meist b-Jets verbunden sind, kann ausgenutzt werden, um diesen Untergrund durch ein b -Veto in der Signalregion zu reduzieren. Damit wird der

$t\bar{t}$ -Untergrund erst ab fünf Jets der größte Untergrund und ist stets geringer als W+Jets. Um die Unsicherheiten bei der Abschätzung des $t\bar{t}$ -Untergrundes zu minimieren, wird in dieser Arbeit eine datenbasierte Untergrundabschätzung durchgeführt. Als Kontrollregion dienen die vom b-Veto aus der Signalselektion ausgeschlossenen Ereignisse. Im Laufe des Verfahrens (siehe Kapitel 4), ebenso wie für die Abschätzung des Multijet-Untergrundes, werden auch simulierte $t\bar{t}$ -Ereignisse benötigt. Als Monte Carlo Generator wird hierzu auf POWHEG [12] mit dem PERUGIA 2011 Tune[21] für die harte Wechselwirkung und PYTHIA6[22] für den Partonenschauer zurückgegriffen. Der hadron-Parameter wurde hierfür auf die Topquarkmasse gesetzt. Zur Berücksichtigung theoretischer Unsicherheiten wird im Myonenkanal auch SHERPA [14] benutzt.

2.4 Multijet-Untergrund

Multijet-Ereignisse sind Prozesse der starken Wechselwirkung (QCD). Im Laufe der Hadronisierung können sich auch schwere Quarks oder instabile Hadronen bilden, die schwach zerfallen und dabei ein Lepton emittieren. Dieses bleibt meist im Inneren des Jets. Man bezeichnet es daher als *nicht isoliert*. Um diesen Hintergrund zu reduzieren wird daher in der Signalselektion ein Isolationskriterium benutzt. Allerdings passieren dennoch einige Multijet-Ereignisse die Signalselektion. Dies geschieht sowohl in der W+Jets Signalregion (siehe Kapitel 2.6) als auch in der für die Bestimmung des $t\bar{t}$ -Untergrundes verwendeten $t\bar{t}$ -Kontrollregion (siehe 4.3.1). Aufgrund der Komplexität der QCD-Prozesse muss hier auf Multijet-Templates aus Daten zurückgegriffen werden. In Kapitel 3 wird eine datenbasierte Bestimmung des Multijet-Untergrundes durchgeführt.

2.5 Weitere Untergrundprozesse

Neben den beiden großen Untergründen passieren auch noch weitere Prozesse in geringer Zahl die Signalselektion. Diese können durch Monte Carlo Generatoren simuliert werden. Im konkreten sind das Diboson-Ereignisse mit zwei W-, zwei Z- oder W- und Z- Bosonen (HERWIG[10]), die Produktion einzelner Top-Quarks, im Folgenden Single Top-Ereignisse genannt (JIMMY[6] PYTHIA, AcerMC[16]), Tauonen, die fälschlicherweise für Myonen oder Elektronen gehalten werden (ALPGEN PYTHIA, SHERPA), der Zerfall $Z^0 \rightarrow e^+e^-$ (SHERPA, nur Elektronenkanal), sowie der Zerfall $Z^0 \rightarrow \tau^+\tau^-$ (ALPGEN JIMMY). Für die Bestimmung von Multijet- und $t\bar{t}$ -Untergrund werden in dieser Arbeit die Simulationen von W+Jets Ereignissen und den hier erwähnten weiteren Untergrundprozessen in einem Signal-Template zusammengefasst.

2.6 Definition der Signalregion

W+Jets Ereignisse sind nur ein Teil der vom ATLAS Detektor aufgezeichneten Ereignisse. Daher ist es erforderlich, Kriterien zu definieren, welche insbesondere W+Jets Ereignisse selektieren und somit die Signalregion festlegen. Die Signalregion sollte somit so gewählt werden, dass zum einen viele W+Jets Ereignisse darin liegen, um eine gute Statistik zu erhalten. Zum anderen sollten darin möglichst wenige Ereignisse aus Untergrundprozessen vorhanden sein. Für die Signalregion wurden im Elektronenkanal die Trigger EF_e24vhi_medium1 und EF_e60_medium1 und im Myonenkanal die Trigger EF_mu24i_tight und EF_mu36_tight mit einer integrierten Luminosität von $20,3 \text{ fb}^{-1}$

verwendet. Die wichtigsten auf die aufgezeichneten Daten angewendeten Selektionskriterien sind:

1. Für das gesamte Ereignis muss gelten:

- (a) funktionierender Detektor
- (b) kein ungewöhnlich starkes Rauschen im Detektor
- (c) Erfolgreiche Rekonstruktion der Jets und Tajektorien
- (d) mindestens ein primärer Vertex und drei damit verbundene Trajektorien
- (e) $E_T^{miss} > 25 \text{ GeV}$ aufgrund des Neutrinos
- (f) $m_T > 40 \text{ GeV}/c^2$
- (g) $p_T < 20 \text{ GeV}/c$ für ein zweites Elektron, das die Elektronenidentifikation medium++ (siehe [11]) beziehungsweise keine zwei Myonen (Kriterien siehe Punkt 3.) mit $p_T > 20 \text{ GeV}/c$, $|\eta| < 2,47$ und nicht $1,37 < |\eta| < 1,52$ erfüllt, um den Untergrund $Z^0 \rightarrow e^+e^-$ zu verringern.

2. Das Signalelektron im Elektronenkanal muss folgendes erfüllen:

- (a) $p_T^e > 25 \text{ GeV}/c$
- (b) $|\eta| < 2,47$ und nicht $1,37 < |\eta| < 1,52$
- (c) Das Elektron muss die Elektronenidentifikation tight++ (siehe [11]) passieren.
- (d) Das Elektron muss isoliert sein. Hierfür wird verlangt, dass die Summe der transversalen Impulse von anderen Tajektorien in einem Gebiet um das Elektron herum (ohne Elektron) p_T^{cone} beziehungsweise die Summe der transversalen Energien in den Kalorimetern E_T^{cone30} deutlich kleiner ist als der transversale Impuls des Elektrons p_T^e selbst. Dabei bezeichnet "cone30" einen Kreis mit Radius $\Delta R = 0,3$.

Im Detail gilt:

- i. $p_T^{cone30}/p_T^e < 0,07$
- ii. $E_T^{cone30}/p_T^e < 0,14$
- (e) $|z_0 \cdot \sin \theta| < 0,5 \text{ mm}$
- (f) $|d_0|/\sigma_{d_0} < 5$

3. Für das Signalmyon im Myonenkanal gilt:

- (a) $p_T^\mu > 25 \text{ GeV}/c$
- (b) $|\eta| < 2,4$ und nicht $1,37 < |\eta| < 1,52$
- (c) Das Myon muss die Myonenidentifikation medium plus (siehe [18]) passieren
- (d) Als Isolationskriterium: $p_T^{cone20}/p_T^\mu < 0,1$
- (e) Zusätzlich gilt für das Signalmyon und andere Myonen je nach Impuls
 - i. $p_T^\mu \leq 15 \text{ GeV}/c$: $p_T^{cone40}/p_T^\mu < 0,06$ und $E_T^{cone30}/p_T^\mu < 0,06$
 - ii. $15 \text{ GeV}/c < p_T^\mu \leq 20 \text{ GeV}/c$: $p_T^{cone30}/p_T^\mu < 0,08$ und $E_T^{cone30}/p_T^\mu < 0,12$
 - iii. $20 \text{ GeV}/c < p_T^\mu \leq 25 \text{ GeV}/c$: $p_T^{cone30}/p_T^\mu < 0,12$ und $E_T^{cone30}/p_T^\mu < 0,18$

- iv. $25 \text{ GeV}/c < p_T^\mu$: $p_T^{cone30}/p_T^\mu < 0,12$ und $E_T^{cone30}/p_T^\mu < 0,30$
- (f) $|z_0 \cdot \sin \theta| < 1 \text{ mm}$
- 4. Der Anti- k_t (LCTopo) Algorithmus mit $R = 0,4$ fasst Teilchen in Jets zusammen. Jets werden nur dann gezählt, wenn:
 - (a) Der Jet erfüllt die Qualitätskriterien Not badLooseMinus
 - (b) $p_T^{jet} > 30 \text{ GeV}/c$
 - (c) $|\eta| < 4,4$ wegen der Geometrie des Detektors
 - (d) Jet und Lepton müssen getrennt sein: $\Delta R > 0.5$
 - (e) die meisten Trajektorien im Jet müssen zum primären Vertex zeigen, um Pile-up zu reduzieren
- 5. Um den $t\bar{t}$ -Untergrund zu senken, werden Ereignisse mit b-Jet aus der Signalregion ausgeschlossen (b-Veto). b-Jets werden mit dem MV1-Algorithmus und einer Effizienz von 60% identifiziert (siehe 1.3.4). Dabei wird von einem b-Jet verlangt, dass
 - (a) gute Qualität der Daten
 - (b) $p_T > 20 \text{ GeV}/c$
 - (c) $|\eta| < 2,5$, da b-Tagging nur hier mit dem Trackingdetektor möglich ist.

In dieser Arbeit werden Ereignisse mit unterschiedlicher *Jetmultiplizität* untersucht. Dies ist die Anzahl der in diesem Ereignis vorkommenden Jets im Sinne von Kriterium 4. Dabei werden die Jetmultiplizitäten drei, vier und fünf exklusiv, also einzeln, untersucht. Ereignisse mit 6 oder mehr Jets wurden dagegen aufgrund der geringen Statistik zusammengefasst.

Kapitel 3

Multijet-Untergrundabschätzung

3.1 Motivation

Multijet-Ereignisse sind neben $t\bar{t}$ -Ereignissen der zweite große Untergrund. Als Prozess der starken Wechselwirkung können sie jedoch nur sehr schwer simuliert werden. Daher wird eine datenbasierte Abschätzung des Multijet-Untergrundes mit Hilfe eines Likelihood-Fits durchgeführt. Als Fitvariable ist die fehlende transversale Energie E_T^{miss} geeignet.

Der Fit wird mit zwei Templates durchgeführt: einem Multijet-Template aus Daten und einem Template, dass die Monte Carlo Vorhersage für W+Jets Ereignisse, $t\bar{t}$ und die weiteren Untergrundprozesse (siehe 2.5) enthält. Der Fit wird in der Signalregion (SR) mit angewandtem b-Veto durchgeführt. Außerdem wird der Fit auch in der $t\bar{t}$ -Kontrollregion (CR) wiederholt. Dies ist notwendig, um bei der datenbasierten Abschätzung des $t\bar{t}$ -Untergrundes (siehe Kapitel 4) die Kontamination des $t\bar{t}$ -Templates abzuschätzen.

Als Grundlage für den Fit müssen hierzu möglichst reine Multijet-Templates gewonnen werden. Diese Arbeit konzentriert sich auf die Bestimmung des $t\bar{t}$ -Untergrundes (siehe Kapitel 4), für die die Abschätzung des Multijet-Untergrundes ein notwendiger Zwischenschritt ist. Entsprechend sind die systematischen Fehlerquellen für beide Untergründe korreliert. Es wird hier auf eine Diskussion dieser Fehlerquellen verzichtet und stattdessen in Kapitel 4.7 direkt die Auswirkung auf den $t\bar{t}$ -Untergrund bestimmt.

3.2 Multijet-Templates

Um Multijet-Templates zu erhalten, müssen drei Schritte durchgeführt werden. Zuerst müssen zwei Multijet-Kontrollregionen definiert werden: eine, die die Multijet-Verteilung in der Signalregion und eine zweite, die die Multijet-Verteilung in der $t\bar{t}$ -Kontrollregion widerspiegelt. Im letzten Schritt muss die Kontamination der Daten durch W+Jets und $t\bar{t}$ -Ereignisse, sowie durch die anderen Untergrundprozesse abgeschätzt und subtrahiert werden.

3.2.1 Definition der Multijet-Kontrollregionen

Der Multijet-Untergrund wird von den Multijet-Ereignissen gebildet, bei denen ein Jet oder ein Teil davon vom Detektor fälschlicher Weise als isoliertes Signal-Lepton identifiziert wurde. Um ein

mit Multijet-Ereignissen angereichertes Template zu erhalten, werden daher nun die Isolationskriterien für Leptonen (siehe Kapitel 2.6) invertiert. Es wird verlangt, dass das vermeintliche Lepton von einem Jet umgeben und somit nicht isoliert ist und dass sich sein Impuls nicht übermäßig von dem anderer Komponenten des umgebenden Jets unterscheidet. Die genaue Vorgabe ist dabei

1. im Elektronenkanal: $p_T^{cone30}/p_T^e > 0,1$
2. im Myonenkanal: $0,1 < p_T^{cone20}/p_T^\mu < 0,5$.

Als Trigger wurde im Elektronenkanal der Trigger EF_e24vh_medium1 und im Myonenkanal der Trigger EF_mu24_tight mit einer integrierten Luminosität von jeweils $2,7 \text{ fb}^{-1}$ verwendet. Mit diesen Selektionskriterien erhält man Templates, die zu großen Teilen aus Multijet-Ereignissen bestehen. Auf die Daten wird nun einmal ein b-Veto angewendet, um eine Verteilung wie in der Signalregion zu erreichen. Wird auf die Daten hingegen ein b-Tagging angewendet, so erhält man die Verteilung in der $t\bar{t}$ -Kontrollregion. Das b-Tagging wird in Kapitel 1.3.4 genauer erläutert.

3.2.2 Kontamination

Da allerdings auch Ereignisse der anderen Untergrundprozesse (siehe 2.5), W+Jets und $t\bar{t}$ -Ereignisse diese Selektion zu geringen Teilen passieren, müssen diese subtrahiert werden. Dies geschieht auf Basis einer jeweiligen Monte Carlo Simulation.

Das so erhaltene Template ist nun geeignet, um über einen Fit den Multijet-Anteil in der Signalregion und der $t\bar{t}$ -Kontrollregion abzuschätzen.

3.3 Fit zur Bestimmung des Multijet-Anteils

In der Signalregion werden das datenbasierte Multijet-Template und das Monte Carlo basierte Template mit den übrigen Prozessen an die Daten gefittet. Diese enthalten sämtliche Cuts wie die eigentliche Signalregion, mit Ausnahme der Forderung von $E_T^{miss} > 25 \text{ GeV}$. Es wird ein Fitbereich von 15-75 GeV gewählt, da hier ausreichend Statistik vorhanden ist und die detektorbasierten Unsicherheiten der Monte Carlos (siehe 4.7) noch begrenzt sind. Auch der Profilunterschied zwischen Multijet-Ereignissen und dem Template mit den verbleibenden Prozessen ist hier am deutlichsten. Der Fit wird mit der in ROOT [5] implementierten Klasse TFractionFitter [3] durchgeführt. Diese skaliert zum einen die gefitteten Templates, zum anderen variiert sie aber auch die einzelnen Bins im Rahmen eines statistischen Fehlers, der immer als die Wurzel des Bineintrages angenommen wird. Als Ergebnis wird der Anteil der eingegebenen Templates an den Daten innerhalb des Fitbereichs inklusive Fehler angegeben. Eine ausführliche Diskussion der Vor- und Nachteile dieser Technik findet sich in 4.4.4. Nach dem Fit wird aus dem angegebenen Multijet-Anteil ein Skalierungsfaktor dieses Templates errechnet, mit dem wiederum der Anteil an Multijet-Ereignissen bezüglich der Daten in der Signalregion mit $E_T^{miss} > 25 \text{ GeV}$ berechnet wird. Ein Beispiel eines solchen Fits ist in Abbildung 3.1a zu sehen.

In der $t\bar{t}$ -Kontrollregion wird analog vorgegangen. Ein Beispiel hierfür ist in Abbildung 3.1b zu sehen.

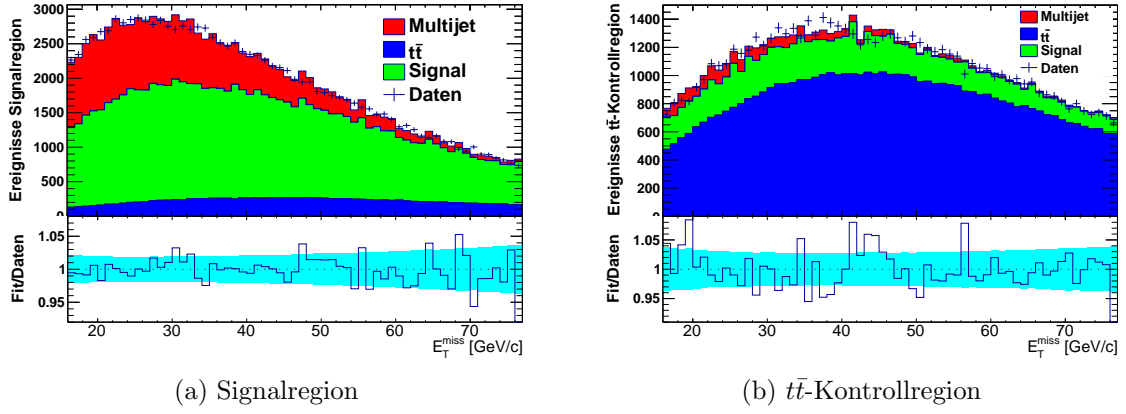


Abbildung 3.1: Fits zur Bestimmung des Multijet-Anteils, (a) in der Signalregion und (b) in der $t\bar{t}$ -Kontrollregion. Das Beispiel stammt jeweils aus dem Elektronenkanal für Ereignisse mit vier Jets. Das Signal-Template enthält sowohl W+Jets Ereignisse als auch die kleineren berücksichtigten Untergrundprozesse (siehe Kapitel 2.5). Durch die Variation der Bins durch TFractionFitter liegt das Fitergebnis fast immer innerhalb des unten eingezeichneten statistischen Fehlerbands der Daten.

3.4 Ergebnisse

Die Ergebnisse der Fits sind für den Elektronenkanal in Tabelle 3.1a und für den Myonenkanal in Tabelle 3.1b zusammengefasst. Es ist insbesondere zu bemerken, dass der Anteil im Elektronenkanal mit 17 – 18% für Ereignisse mit drei bis fünf Jets nahezu konstant ist und starken Einfluss hat. Im Myonenkanal liegt der Multijet-Anteil mit 12-15% etwas niedriger. In der $t\bar{t}$ -Kontrollregion spielen Multijet-Ereignisse dagegen nur eine geringe Rolle, da sie nur selten das b-Jet Kriterium erfüllen.

(a) Multijet-Anteil im Elektronenkanal

$\#Jets$	SR [%]	$t\bar{t}$ CR [%]
2	$15,1 \pm 0,1$	$9,6 \pm 0,3$
3	$17,1 \pm 0,3$	$4,8 \pm 0,3$
4	$17,5 \pm 0,5$	$2,0 \pm 0,4$
5	$17,9 \pm 1,1$	$0,7 \pm 0,5$
≥ 6	$15,9 \pm 2,1$	$2,2 \pm 0,4$

(b) Multijet-Anteil im Myonenkanal

$\#Jets$	SR [%]	$t\bar{t}$ CR [%]
2	$13,4 \pm 0,1$	$21,6 \pm 0,3$
3	$15,1 \pm 0,2$	$10,7 \pm 0,3$
4	$14,3 \pm 0,4$	$4,5 \pm 0,3$
5	$12,7 \pm 0,8$	$3,7 \pm 0,5$
≥ 6	$12,3 \pm 1,5$	$2,0 \pm 0,9$

Tabelle 3.1: Anteil von Multijet-Ereignissen an den Daten in der Signalregion (SR) und in der $t\bar{t}$ -Kontrollregion ($t\bar{t}$ CR) bei einer b-Taggingeffizienz von 60% (a) im Elektronenkanal und (b) im Myonenkanal in Prozent.

Kapitel 4

$t\bar{t}$ -Untergrundabschätzung

4.1 Motivation

$t\bar{t}$ -Ereignisse sind neben Multijet-Ereignissen der zweite große Untergrundprozess in der W+Jets Signalregion. Es steht zwar mit POWHEG [12] eine Monte Carlo Simulation zur Verfügung, allerdings wird hier eine datenbasierte Untergrundabschätzung durchgeführt, um die vorhandenen theoretischen und detektorbasierten Unsicherheiten zu reduzieren. Dies ist wegen der Größe des $t\bar{t}$ -Untergrundes besonders wichtig. Dabei wird ein aus Daten gewonnenes $t\bar{t}$ -Template, gemeinsam mit einem datenbasierten Multijet-Template und einem Monte Carlo basierten Signal-Template, bestehend aus W+Jets Ereignissen und den verbleibenden kleineren Untergrundprozessen (siehe Kapitel 2.5), an die Messdaten gefittet, um die Normierung des $t\bar{t}$ -Templates zu erhalten. Dies geschieht in der Signalregion. Hierfür muss zuerst eine geeignete Fitvariable gefunden werden. Als solche wurde die transformierte Aplanarität gewählt. Anschließend müssen datenbasierte Templates für Multijet- und $t\bar{t}$ -Ereignisse gewonnen werden. Mit diesen wird dann der oben erwähnte Fit durchgeführt. Hierfür wird auf zwei Methoden zurückgegriffen: die in Root integrierte Klasse TFractionFitter [3], sowie ein χ^2 -Fit mit Funktionen, der in Kapitel 4.5 entwickelt wird. Die anschließende Abschätzung statistischer und systematischer Fehlerquellen wird zeigen, dass der Fit mit Funktionen in der Lage ist, bei gleichen statistischen Unsicherheiten die systematischen Fehler im Vergleich zu TFractionFitter zu reduzieren.

4.2 Transformierte Aplanarität

Als Variable für den Fit zur Bestimmung des $t\bar{t}$ -Anteils wird wie in der W+Jets Analyse der Daten aus dem Jahr 2011 [2] die transformierte Aplanarität $A' = \exp(-8A)$ genutzt. Dabei ist die Aplanarität A 1,5 mal der kleinste Eigenwert λ_3 des normierten Impulstensors [2]

$$M_{xyz} = \sum_i \begin{pmatrix} P_{x,i}^2 & P_{x,i}P_{y,i} & P_{x,i}P_{z,i} \\ P_{y,i}P_{x,i} & P_{y,i}^2 & P_{y,i}P_{z,i} \\ P_{z,i}P_{x,i} & P_{z,i}P_{y,i} & P_{z,i}^2 \end{pmatrix} \quad (4.1)$$

mit den Eigenwerten

$$\lambda_1 > \lambda_2 > \lambda_3 \quad (4.2)$$

und

$$\sum_i \lambda_i = 1, \quad (4.3)$$

wobei die Summe über alle detektierten Jets und Leptonen gebildet wurde[1]. Die transformierte Aplanarität quantifiziert, wie viel Impuls aus der durch die beiden führenden Jets gebildeten Ebene herauszeigt. Planare Ereignisse besitzen den Wert $A' \approx 1$ während sphärische Ereignisse den Wert $A' \approx 0$ annehmen. Aufgrund ihrer großen Masse bewegen sich Topquarks vergleichsweise langsam. Entsprechend ist der Lorentzboost ihrer Zerfallsprodukte im Detektorsystem schwächer. Daher neigen $t\bar{t}$ -Ereignissen zu einem sphärischeren Erscheinungsbild im Detektor als W+Jets Ereignisse. Dies resultiert in einem für den $t\bar{t}$ -Fit ausreichenden Profilunterschied der beiden Verteilungen bezüglich der transformierten Aplanarität.

4.3 Gewinnung datenbasierter Templates

Als Multijet-Template wird ein datenbasiertes Template genutzt. Dies wird genau wie in Kapitel 3 beschrieben gewonnen. Um ein datenbasiertes $t\bar{t}$ -Template zu gewinnen, sind drei Schritte notwendig: Zuerst muss eine $t\bar{t}$ -Kontrollregion definiert werden. Anschließend muss die Kontamination dieser Daten durch andere Prozesse abgeschätzt und subtrahiert werden. Abschließend muss eine Korrektur durchgeführt werden, um das Profil in der Signalregion zu erhalten.

4.3.1 Definition einer $t\bar{t}$ -Kontrollregion

Die produzierten Top- und Antitopquarks zerfallen zu $91 \pm 4\%$ [15] in Bottom- und Antibottomquarks und W-Bosonen. Wird nun auf die W+Jets Kandidaten in der Signalregion ein b-Veto angewandt, kann der Hintergrund an $t\bar{t}$ -Ereignissen deutlich reduziert werden. Wird als $t\bar{t}$ -Kontrollregion dagegen ein b-Tagging durchgeführt, erhält man ein Sample, das von $t\bar{t}$ -Ereignissen dominiert wird. Dieses kann genutzt werden, um das Profil der $t\bar{t}$ -Verteilung in der Signalregion abzuschätzen.

Für das b-Tagging (siehe Kapitel 1.3.4) können unterschiedliche b-Taggingeffizienzen benutzt werden. Diese beruhen auf unterschiedlich harten Identifikationskriterien für die b-Jets. Es sind vier unterschiedliche Effizienzen verfügbar: 60%, 70%, 80% und 85%. Dabei bedeutet 60%, dass 60% aller b-Jets diese Selektion passieren. Für die anderen working points werden zwar mehr b-Jets richtig erkannt, allerdings kommt es auch häufiger zu Fehlidentifikationen anderer Objekte, wodurch die Kontamination zunimmt. Dem gegenüber steht der Vorteil einer größeren Statistik.

Zusätzlich kann gewählt werden, ob mindestens zwei b-Jets identifiziert wurden oder ob mindestens einer ausreicht. Das gleichzeitige Auftreten zweier b-Jets ist ein starker Hinweis auf ein $t\bar{t}$ -Ereignis, allerdings wird durch diese Forderung die Statistik deutlich reduziert.

Es gilt also, eine Auswahl zu treffen, die einerseits eine ausreichend hohe Statistik zur Verfügung stellt und andererseits möglichst rein ist, um den Fehler durch die Kontamination des Templates zu reduzieren. Für das b-Veto in der Signalregion wurde eine Effizienz von 60% bei einem identifizierten b-Jet (1.60) gewählt. Ein b-Jet erlaubt, dass $t\bar{t}$ -Ereignisse auch bei der Identifikation von nur einem der beiden b-Jets aussortiert werden, andererseits sorgen die strengen Kriterien der b-Taggingeffizienz von 60% dafür, dass möglichst wenige W+Jets Ereignisse fälschlicherweise verloren gehen. Konsistenter Weise ist 1.60 auch die Standardeinstellung für die $t\bar{t}$ -Kontrollregion und wur-

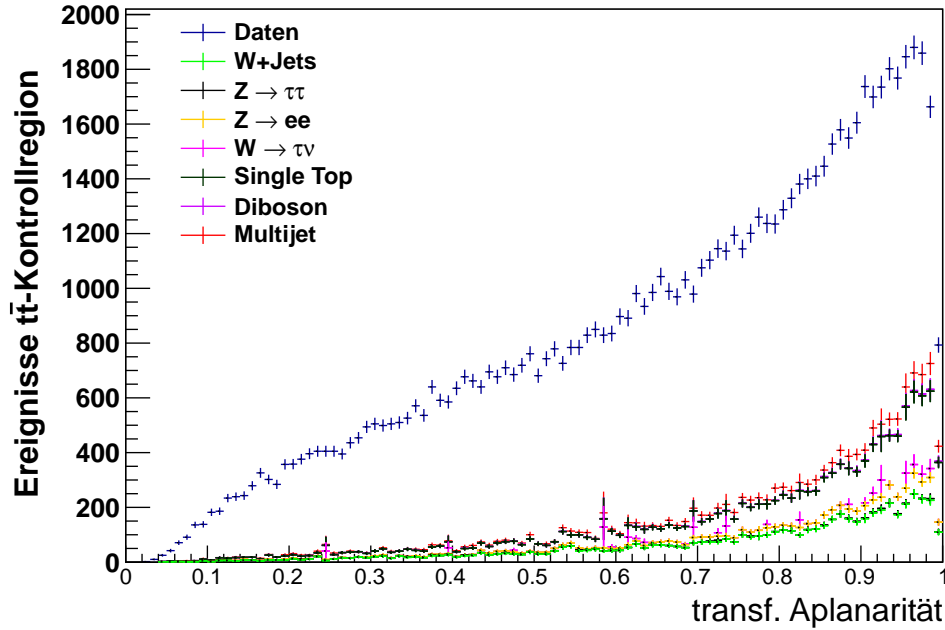


Abbildung 4.1: Daten-Template aus der $t\bar{t}$ -Kontrollregion mit gestapelter Kontamination bei vier Jets im Elektronenkanal und der Selektion 1_60. Hier liefern W+Jets Ereignisse und Ereignisse mit einem einzelnen Topquark ("Single Top") die größten Beiträge.

de für sämtliche Studien im Elektronenkanal genutzt. So sind Signalregion und $t\bar{t}$ -Kontrollregion orthogonal, weshalb keine statistischen Korrelationen zwischen den beiden Regionen bestehen. Im Myonenkanal wurde zusätzlich eine Studie durchgeführt, inwiefern die Auswahl einer anderen b-Taggingeffizienz oder eines 2 b-Jets Kriteriums in der $t\bar{t}$ -Kontrollregion zu besseren Ergebnissen führen kann (siehe 4.10). Das Ergebnis bestätigte allerdings die Wahl der $t\bar{t}$ -Kontrollregion 1_60.

4.3.2 Kontamination

Für den Fit ist es wichtig, ein möglichst reines $t\bar{t}$ -Template zu nutzen. Im Daten-Template aus der $t\bar{t}$ -Kontrollregion ist jedoch eine Kontamination durch andere Prozesse wie W+Jets und Multijet-Ereignisse, sowie den in Kapitel 2.5 erwähnten kleineren Untergrundprozessen vorhanden, die nun abgezogen werden muss. Für Multijet-Ereignisse wurde der Anteil in Kapitel 3 bestimmt. Die Ergebnisse finden sich in Kapitel 3.4. Ein Template wird wie in Kapitel 3.2 beschrieben aus Daten gewonnen und entsprechend dem Multijet-Anteil normiert. Für alle anderen Kontaminationen werden Monte Carlo Simulationen genutzt. In Abbildung 4.1 ist das Daten-Template aus der Kontrollregion für Ereignisse mit vier Jets im Elektronenkanal in der Selektion 1_60 mit gestapelten Kontaminationsprozessen abgebildet. Die Anteile für alle Untergrundprozesse und Jetmultiplizitäten sind für den Elektronenkanal in Tabelle A.1 und für den Myonenkanal in Tabelle A.2 nachzulesen.

Die Kontamination wird von den Daten subtrahiert. Somit erhält man ein möglichst reines $t\bar{t}$ -Template in der Kontrollregion. Im nächsten Schritt muss nun die Verzerrung des Profils durch das b-Tagging in der $t\bar{t}$ -Kontrollregion und das b-Veto in der Signalregion ausgeglichen werden.

4.3.3 Korrektur der Verzerrung durch das b-Tagging

Die Anwendung eines b-Taggings in der $t\bar{t}$ -Kontrollregion führt zu einer Verzerrung der Verteilung von $t\bar{t}$ -Ereignisse im Vergleich zur Signalregion mit angewendetem b-Veto. Neben der Normierung unterscheidet sich insbesondere auch das für den Fit relevante Profil. Um nun aus dem datenbasierten $t\bar{t}$ -Template in der $t\bar{t}$ -Kontrollregion (CR) eines in der Signalregion (SR) zu erhalten, muss eine Korrektur angewandt werden. Hierfür wird auf Monte Carlo Simulationen zurückgegriffen. Die Korrektur sieht wie folgt aus:

$$N(A')_{SR}^{Daten} = c \cdot \frac{N(A')_{SR}^{MC}}{N(A')_{CR}^{MC}} \cdot N(A')_{CR}^{Daten} \quad (4.4)$$

$$= C(A') \cdot N(A')_{CR}^{Daten} \quad (4.5)$$

Dabei bezeichnet $N(A')_{SR}^{MC}$ die Monte Carlo Vorhersage für die Anzahl der $t\bar{t}$ -Ereignisse in der Signalregion und $N(A')_{CR}^{MC}$ die Vorhersage für die Anzahl der Ereignisse in der $t\bar{t}$ -Kontrollregion im Bin bei A' . $N(A')_{CR}^{Daten}$ ist der Bineintrag des datenbasierten $t\bar{t}$ -Templates in der $t\bar{t}$ -Kontrollregion nach Abzug der Kontamination. Die Normierungskonstante c wurde so gewählt, dass die Korrektur die Normierung des datenbasierten $t\bar{t}$ -Templates aus der Kontrollregion nicht ändert. Für die Bestimmung des Wertes von $C(A')$ wurde POWHEG verwendet. Im Myonenkanal war zudem SHERPA verfügbar. Dieser wurde genutzt um die Unsicherheit auf die Korrektur abzuschätzen. Diese ist der statistische Fehler von POWHEG oder, falls größer, die Differenz der Korrekturhistogramme von POWHEG und SHERPA. In Abbildung 4.2 ist als Beispiel das Korrekturhistogramm für Ereignisse mit vier Jets im Myonenkanal zu sehen. Am ansteigenden Verlauf ist zu erkennen, dass sich das Profil in Signal- und Kontrollregion deutlich unterscheidet. Vermutlich lässt es sich dadurch erklären, dass in sphärischen Ereignissen (kleines A') die Zerfallsprodukte des identifizierten b-Hadrons den Jet deutlicher verlassen und somit einen signifikanteren Impact parameter d_0 (siehe Kapitel 1.3.4) besitzen als bei planaren Ereignissen. Folglich ist hier die b-Tagging Effizienz höher und beim Transfer in die Signalregion muss nach unten korrigiert werden.

Die Verzerrung durch das b-Tagging fällt für unterschiedliche Variablen verschieden stark aus. So ist insbesondere die Verteilung bezüglich der Rapidität der Jets $Jet \eta$ besonders stark betroffen, da die b-Taggingeffizienz stark von der Detektorregion abhängt (siehe Abbildung A.1). Für den transversalen Impuls der Jets treten deutlichere Unterschiede zwischen POWHEG und SHERPA auf (siehe Abbildung A.2). Dagegen sind die Unsicherheiten für die transformierte Aplanarität vergleichsweise gering, was ein weiteres Argument für die Wahl dieser Fitvariablen ist.

Nun sind alle Templates vorhanden, um mit einem Fit den $t\bar{t}$ -Untergrund in der Signalregion abzuschätzen.

4.4 Fit mit TFractionFitter

Der Fit zur Bestimmung des $t\bar{t}$ -Untergrunds in der Signalregion wird zuerst mit der in Root implementierten Klasse TFractionFitter durchgeführt. Im Folgenden wird zuerst deren Funktionsweise geschildert. Anschließend wird der Fit zur Bestimmung des $t\bar{t}$ -Untergrundes beschrieben bevor auf die Probleme dieser Methode eingegangen wird.

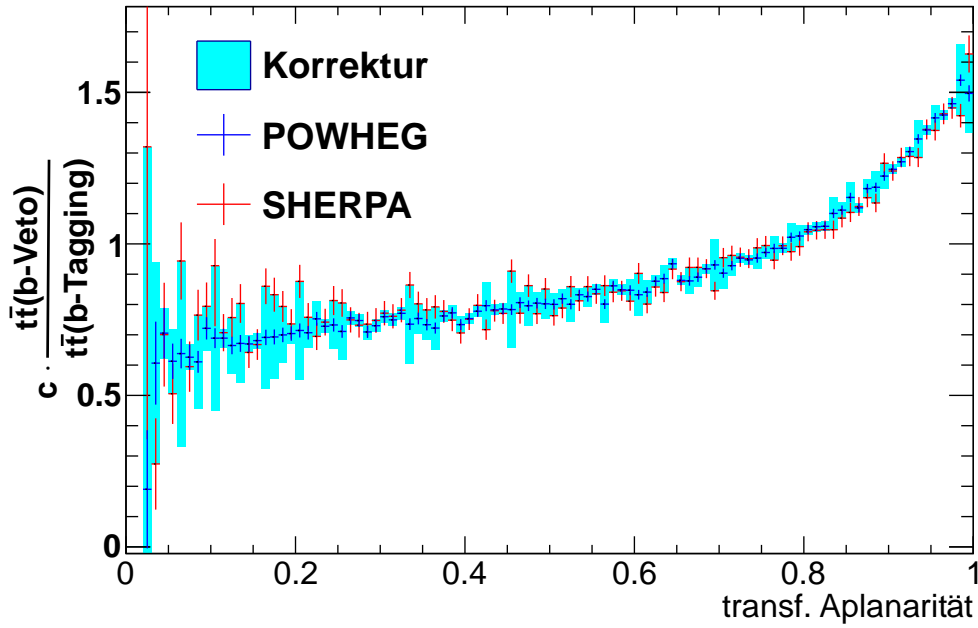


Abbildung 4.2: Korrekturhistogramm, um im datenbasierten $t\bar{t}$ -Template aus der Kontrollregion (CR) Verzerrungen durch das b-Veto in der Signalregion (SR) auszugleichen. Dieses Beispiel ist für vier Jets im Myonenkanal. Als Wert der Korrektur dient der Wert von POWHEG, SHERPA dient zur Abschätzung der Unsicherheit.

4.4.1 Funktionsweise

TFractionFitter ist eine Klasse, um Histogramme an Histogramme zu fitten und bedient sich dabei der in [3] beschriebenen Fitroutine HMCMLL, in der die Likelihood-Funktion maximiert wird. Es werden ein Daten-Template und mehrere Templates für den Daten zugrundeliegende Prozesse definiert. Diese werden dann so normiert, dass sie möglichst gut an die Daten passen. TFractionFitter besitzt dabei die Fähigkeit, neben der Normierung des gesamten Histogramms auch einzelne Bins zu variieren. Als Rahmen dafür wird ein Poissonfehler des Bins j von

$$\sigma_j = \sqrt{N_j} \quad (4.6)$$

angenommen, wobei N_j der Binninhalt des Bins j ist. TFractionFitter kann somit auch Unsicherheiten auf das Profil der verwendeten Templates ausgleichen. Dies ist insbesondere bei datenbasierten Templates von Vorteil, in denen eine begrenzte Statistik zu großen relativen Unsicherheiten führt. Als Ergebnis liefert TFractionFitter den Anteil f_i der jeweiligen Templates i an den Daten innerhalb des Fitbereichs inklusive Fehler. Dieser kann in einen Skalierungsfaktor s_i mittels

$$s_i = f_i \cdot \frac{N_{\text{Daten}}^{\text{Fitbereich}}}{N_i^{\text{Fitbereich}}} \quad (4.7)$$

umgerechnet werden. Multipliziert man diesen Skalierungsfaktor mit der Anzahl der Ereignisse des gesamten, unskalierten Templates N_i^{total} , so erhält man die Gesamtzahl der entsprechenden

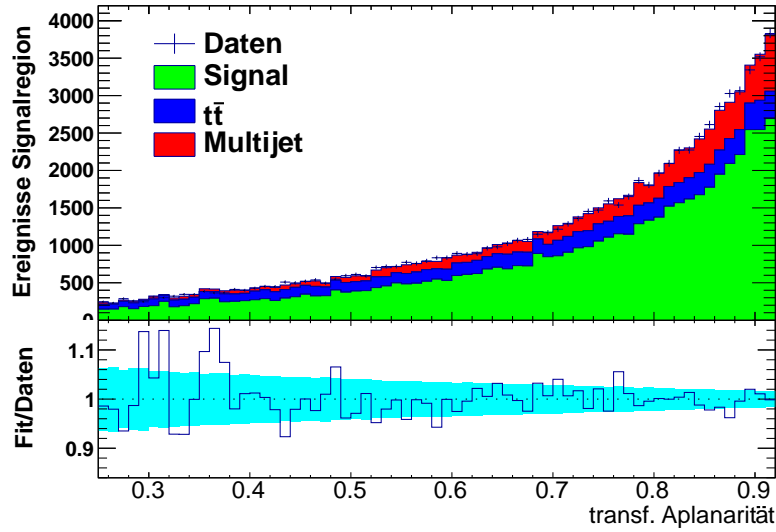


Abbildung 4.3: Fit zur Bestimmung des $t\bar{t}$ -Untergrundes mit TFractionFitter für 4 Jets im Elektronenkanal. Das Signal-Template enthält die Monte Carlo Simulationen für W+Jets Ereignissen, sowie die in Kapitel 2.5 beschriebenen kleinen Hintergrundprozesse. Das Fitergebnis liegt innerhalb des statistischen Fehlerbands der Daten (unten).

Ereignisse in der Signalregion.

$$N_i^{SR} = s_i \cdot N_i^{total} \quad (4.8)$$

4.4.2 Fit

Das Daten-Template besteht aus Daten, die die Definitivskriterien der Signalregion erfüllen. Daran gefittet werden ein Signal-Template, das Monte Carlo Simulationen von W+Jets Ereignissen und den in Kapitel 2.5 beschriebenen kleinen Hintergrundprozessen enthält. Als zweites Template wird das in 4.3 gewonnene datenbasierte $t\bar{t}$ -Template verwendet. Hinzu kommt ein Multijet-Template, dass wie in Kapitel 3 beschrieben gewonnen wurde. Dessen Anteil an den Daten wird an das Ergebnis der Abschätzung des Multijet-Untergrunds in der Signalregion aus Kapitel 3 fixiert. Die genauen Werte sind in den Tabellen 3.1a und 3.1b nachzulesen. Dies ist notwendig, da sich das Profil von W+Jets Ereignissen und Multijet-Ereignissen bezüglich der transformierten Aplanarität nur geringfügig unterscheiden. Das $t\bar{t}$ -Template und das Signal-Template können dagegen frei variiert werden. Beim Signal-Template wird dabei ein Skalierungsfaktor nahe eins erwartet, da die Monte Carlo Simulationen bereits auf die theoretische Erwartung normiert sind. Da diese jedoch mit einer Unsicherheit belastet sind, sind geringe Variationen durchaus zu erwarten.

Als Fitbereich wird der maximale Bereich gewählt, in dem der Fit stabil funktioniert und in dem systematische Unsicherheiten keinen übermäßigen Einfluss haben, um die maximale Menge an Information zu verwenden. Eine genauere Diskussion dieses Problems findet in 4.7.1 statt. Im Elektronenkanal ergab sich ein Fitbereich von $A' = 0,25$ bis $A' = 0,91$ für Ereignisse mit drei, vier und fünf Jets, für Ereignisse mit sechs und mehr Jets reicht der Fitbereich von $A' = 0,31$ bis $A' = 0,91$. Außerdem wurde wegen der geringen Statistik für 5 und ≥ 6 Jets ein Rebinning um den Faktor 3 durchgeführt.

Als Beispiel ist in Abbildung 4.3 der Fit für Ereignisse mit vier Jets im Elektronenkanal dargestellt.

4.4.3 Ergebnisse

Die bislang erzielten Fehler sind lediglich die Unsicherheit aufgrund des Fits. Weitere Fehlerquellen werden in den Kapiteln 4.6 und 4.7 diskutiert und abgeschätzt. Die endgültigen Ergebnisse finden sich in den Tabellen 4.3 und 4.5. Die Werte inklusive des Fehlers durch den Fit an sich können aber bereits hier festgehalten werden. Der Anteil an $t\bar{t}$ -Ereignisse in der Signalregion inklusive dieses Fehlers σ_{fit} sind in Tabelle 4.1 zusammengefasst.

(a) Elektronenkanal			(b) Myonenkanal		
#Jets	$t\bar{t}$ -Anteil [%]	σ_{fit} [%]	#Jets	$t\bar{t}$ -Anteil [%]	σ_{fit} [%]
3	3,46	0,83	3	4,06	1,16
4	12,96	1,08	4	15,57	1,50
5	27,85	2,85	5	27,86	3,35
≥ 6	35,53	7,80	≥ 6	46,16	19,21

Tabelle 4.1: $t\bar{t}$ -Anteil in der Signalregion, links im Elektronen- und rechts im Myonenkanal bestimmt mit TFractionFitter. Der Fehler σ_{fit} beinhaltet lediglich den statistischen Fehler des Fits an sich. Eine Zusammenfassung mit allen Fehlern findet sich in den Tabellen 4.3 und 4.5.

4.4.4 Probleme

Im Laufe des Fitprozesse zeigten sich einige Probleme, die überwiegend mit der Variation der Bins zusammenhängen.

Zum einen trägt die Abschätzung des statistischen Fehlers der eingegebenen Templates nach Gleichung 4.6 als Wurzel des Bineintrages bei Monte Carlo Simulationen nicht der Tatsache Rechnung, dass die Simulation meist bei einer weit höheren Statistik durchgeführt und anschließend auf die Luminosität der Daten herab skaliert wurde. Wird das Template durch den Fit wieder nach oben skaliert, so überschätzt TFractionFitter den Fehler deutlich.

Zum zweiten wird bei sehr niedriger Statistik der relative Fehler auf Bins sehr groß. Dies kommt im Bereich sehr kleiner Werte für A' vor, insbesondere im Multijet-Template. Dieses wurde bei geringer Luminosität aufgenommen und muss deshalb im $t\bar{t}$ -Fit um einen Faktor von ungefähr drei im Elektronenkanal und ungefähr fünf im Myonenkanal nach oben skaliert werden, weshalb die Fehler sehr groß werden. Dies führt dazu, dass TFractionFitter jegliches Profil der Daten durch geringe relative Variationen eines einzelnen eingegebenen Templates ausgleichen kann, ohne dass die anderen zur Geltung kommen. Dadurch wird das Programm instabil und läuft in Unendlichkeitsschleifen. Somit muss auf einen großen Fitbereich verzichtet werden.

Zum dritten kann TFractionFitter keine offensichtlichen Fluktuationen als solche erkennen. Diese treten insbesondere für die Monte Carlo Simulationen bei einzelnen Bins in Form deutlich von den benachbarten Bins abweichender Werte auf. Da diese aber im Bereich geringer Statistik den einzig nennenswerten Unterschied im Profil der Templates darstellen, gehen sie sehr stark in den Fit ein und verzerren das Ergebnis.

Das vierte Problem liegt in der Motivation der Variation einzelner Bins. Für TFractionFitter ist

diese rein mathematisch durch die Maximierung der zur Quantifizierung der Güte des Fits benutzten Likelihood-Funktion begründet. Dabei bleibt jedoch außen vor, dass physikalisch betrachtet kontinuierliche Verteilungen bezüglich der kinetischen Variablen A' sinnvoll sind. Daher sollten Fluktuationen zwischen benachbarten Bins ausgeglichen werden. Für TFractionFitter ist allerdings auch eine Verstärkung von Ausreißern möglich.

Die vier geschilderten Probleme werden auch noch in der Diskussion der systematischen Fehler in Kapitel 4.7 eine Rolle spielen. Deshalb erscheint es sinnvoll, eine alternative Fitmethode zu verwenden, die insbesondere im Bereich geringer Statistik robuster gegen Fluktuationen in einzelnen Bins ist. Eine solche Fitmethode wird im Folgenden entwickelt.

4.5 Fit mit Funktionen

4.5.1 Motivation

Für einen Fit, bei dem Modelle an Messdaten skaliert werden, gibt es zwei mögliche Fehlerquellen. Zum einen kann vorkommen, dass das angenommene Modell die tatsächliche Physik nicht im Rahmen ihrer Unsicherheit beschreibt. Für Multijet- und $t\bar{t}$ -Ereignisse wird hier versucht, dieses Problem mit datenbasierten Templates zu lösen, während für W+Jets Ereignisse und die anderen Prozesse im Signal-Template auf Monte Carlo Simulation vertraut werden muss. Die zweite Fehlerquelle ist eine unzureichende Statistik. Die daraus resultierenden statistischen Fluktuationen verzerren die tatsächlich zugrundeliegende Verteilung. Dies trifft sowohl auf das Daten-Template zu, an das gefittet wird, als auch auf die Templates der angefitteten Prozesse.

Neben dem vorhandenen Wissen, wie viele Einträge sich in einem Bin befinden und wie groß die Unsicherheit darauf ist, können durch die folgende Annahme weitere Informationen gewonnen werden:

Annahme:

Der statistischen Verteilung der Messwerte bezüglich der kinetischen Variable A' liegt eine kontinuierliche Wahrscheinlichkeitsverteilung zugrunde.

Im folgenden wird gezeigt werden, dass diese Annahme im Rahmen der Messgenauigkeit mit den Messungen verträglich ist. Außerdem wird aus den Konsequenzen dieser Annahme Profit geschlagen. Diese sind:

1. Die Messwerte streuen zufällig um den durch die Wahrscheinlichkeitsverteilung gegebenen Erwartungswert.
2. Die Wahrscheinlichkeitsverteilung lässt sich durch differenzierbare Funktionen approximieren.

Daher ist es möglich, Funktionen an die Daten zu fitten, die die statistischen Fluktuationen um die Erwartungswerte ausgleichen und Ausreißer abschwächen. Das Anpassen von Funktionen ist zudem für geringe Statistik stabiler als die zuvor genutzte Klasse TFractionFitter. Hierdurch ist eine Erweiterung des möglichen Fitbereichs hin zu geringeren transformierten Aplanaritäten realisierbar. Dies ist insbesondere deshalb relevant, weil, wie in Kapitel 4.7.1 gezeigt werden wird, Bereiche hoher transformierter Aplanarität besonders stark von systematischen Fehlerquellen betroffen sind.

Um einen geeigneten Fit mit Hilfe von Funktionen zu realisieren wird im folgenden eine Fitmethode inklusive Fehlerabschätzung entwickelt. Deren Ergebnisse werden anschließend mit denen von `TFractionFitter` verglichen.

4.5.2 Fitmethode

Aufgrund der einfacheren Handhabung wird die Streuung um den Mittelwert durch eine Gaußverteilung angenähert, sodass ein χ^2 -Fit durchgeführt werden kann.

Das Vorgehen ist das Folgende:

Zuerst wird jeweils eine Funktion an die drei Templates Signal, $t\bar{t}$ und Multijet gefittet. Hierfür wird die Rootfunktion `TH1::Fit` verwendet. Als geeignet erwies sich jedes mal das Produkt aus einem Polynom 4. Grades und einer Exponentialfunktion eines Polynoms 3. Grades, also

$$F_i^0(A') = [p_1 + p_2 \cdot A' + p_3 \cdot (A')^2 + p_4 \cdot (A')^3 + p_5 \cdot (A')^4] \cdot \exp[p_6 \cdot A' + p_7 \cdot (A')^2 + p_8 \cdot (A')^3] \quad (4.9)$$

Anschließend wird die Summe dieser drei Funktionen mit jeweils einem Skalierungsfaktor pro Funktion als freien Parameter an die Daten gefittet. Der Skalierungsfaktor für die Multijet-Funktion wird jedoch auf das Ergebnis aus Kapitel 3 fixiert, sodass sich lediglich die Normierung von $t\bar{t}$ und Signal ändern kann.

$$F_{Daten}^0(A') = s_{Signal} \cdot F_{Signal}^0 + s_{t\bar{t}} \cdot F_{t\bar{t}}^0 + s_{Multijet} \cdot F_{Multijet}^0 \quad (4.10)$$

Als Beispiel sind die vier Fits für Ereignisse mit vier Jets im Elektronenkanal abgebildet (siehe Abbildungen 4.4, 4.5, 4.6 und 4.7). Sie modellieren die drei dem Fit zu Grunde liegenden Templates innerhalb des Fitbereichs von 0,00-0,91 gut. Die $\chi_{red}^2 = \chi^2/n.d.f.$ Werte sind $\chi_{red}^2 \approx 1$. Sie sind für den Elektronenkanal in Tabelle A.3 und für den Myonenkanal in Tabelle A.4 zusammengefasst. Im vom Fit gelieferten Fehler auf den Skalierungsfaktor ist noch nicht berücksichtigt, dass auch Fehler durch die Unsicherheit auf die Fitkurven der einzelnen Templates auftreten. Diese werden im Folgenden diskutiert und abgeschätzt.

4.5.3 Fehler durch Unsicherheit der Profile

Der abschließende Fit an die Daten liefert als Ergebnis Skalierungsfaktoren für die Funktionen $F_{Signal}(A')$ und $F_{t\bar{t}}(A')$ (siehe Gleichung 4.10) inklusive eines Fehlers σ_{norm} . Wird der Skalierungsfaktor um diesen Fehler geändert, so steigt der χ^2 -Wert um eins an.

Allerdings hängt das Ergebnis auch von den Profilen der drei Fitfunktionen F_i aus den vorausgehenden Fits an die einzelnen Templates ab. Diese sind allerdings auch nur innerhalb eines Fehlers bekannt. Es werden also zunächst Fehlerfunktionen benötigt, die angeben, wie diese Fitfunktionen im Rahmen der Messgenauigkeit auch aussehen könnten. Anschließend kann über die Spanne der Skalierungsfaktoren, je nach Wahl der für den Fit an die Daten gewählten Fehlerfunktion, der Fehler σ_{shape} aufgrund der Unsicherheit des Profils abgeschätzt werden.

Als ersten Ansatz für geeignete Fehlerfunktionen wird je einer der acht Parameter des Fits (siehe Gleichung 4.9) im Rahmen seines durch den Fit gelieferten Fehlers variiert. Dies führt jedoch zu Funktionen, die die Templates nicht einmal ansatzweise beschreiben. Der Grund hierfür ist eine starke Korrelation zwischen einzelnen Parametern: Wird ein Parameter im Rahmen des angege-

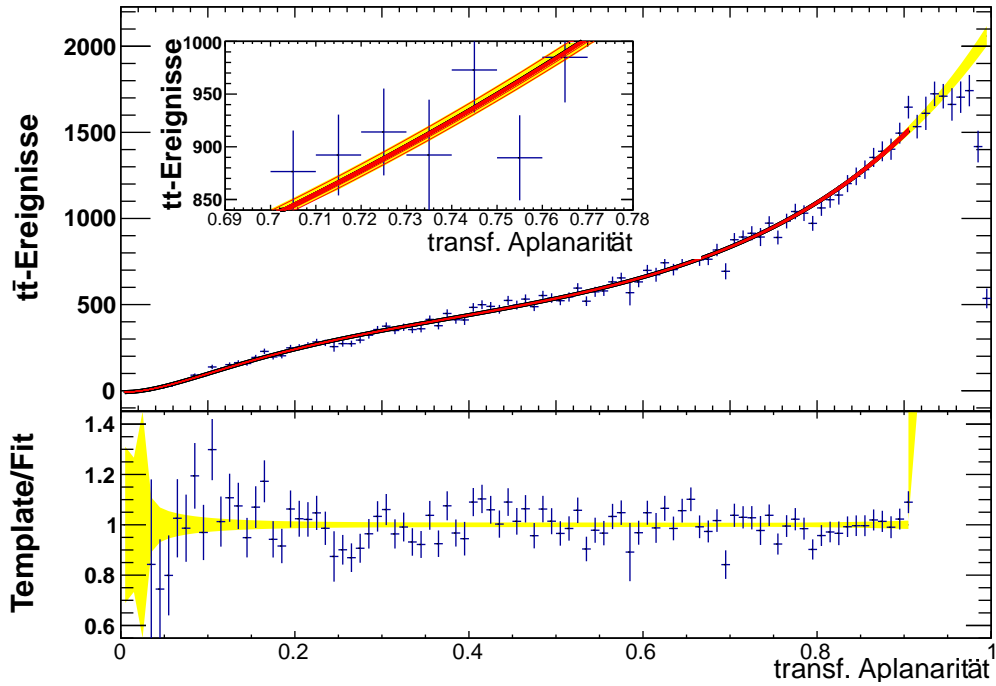


Abbildung 4.4: Fit einer Funktion an das $t\bar{t}$ -Template für Ereignisse mit vier Jets im Elektronenkanal. Wie vergrößert dargestellt decken die roten Fehlerfunktionen das gesamte 1- σ -Konfidenzband (gelb) ab. Der Fitbereich reicht von 0,00 bis 0,91.

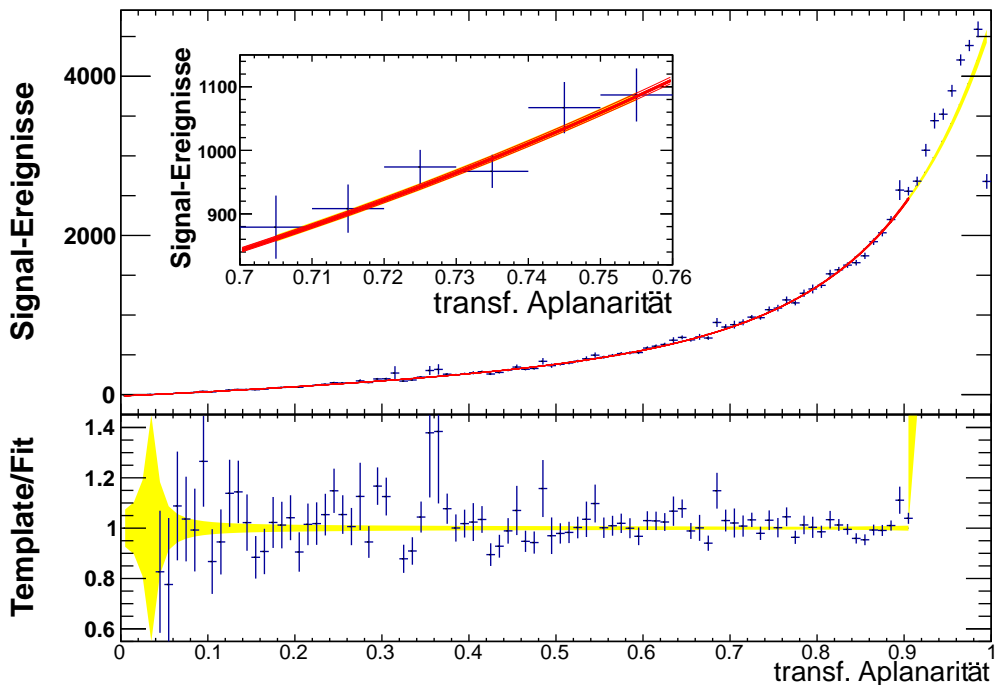


Abbildung 4.5: Fit einer Funktion an das Signal-Template für Ereignisse mit vier Jets im Elektronenkanal. Die kleinen Fehlerbalken der Monte Carlo Simulation führen zu einem schmalen 1- σ -Konfidenzband (gelb). Dies schlägt sich auch in den roten Fehlerfunktionen nieder (siehe Vergrößerung).

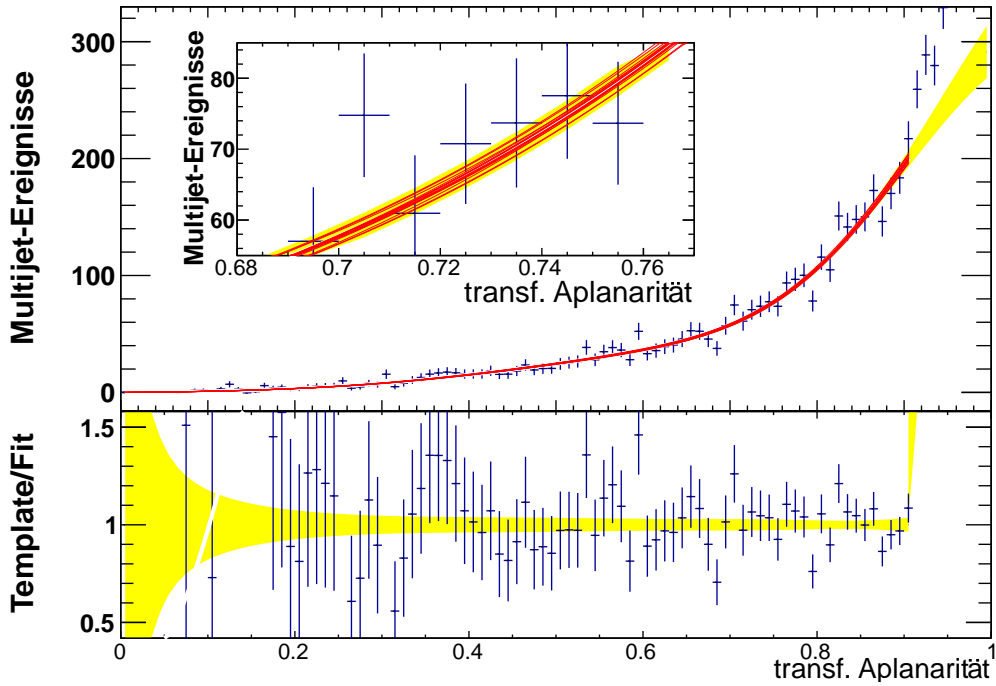


Abbildung 4.6: Fit einer Funktion an das Multijet-Template für Ereignisse mit vier Jets im Elektronenkanal. Wegen der geringsten Statistik ist hier das gelbe 1- σ -Konfidenzband besonders breit. Die roten Fehlerfunktionen tragen dem Rechnung.

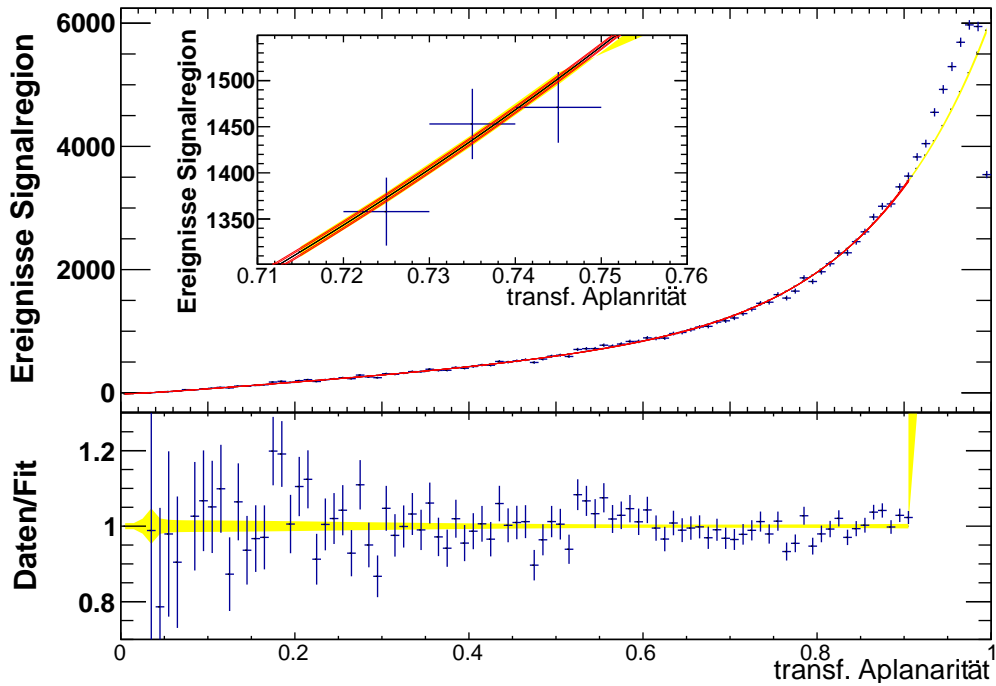


Abbildung 4.7: Der Fit der Summe aus den Fitfunktionen für die einzelnen Templates $t\bar{t}$, Signal und Multijet an die gemessenen Ereignisse in der Signalregion, hier für Ereignisse mit vier Jets im Elektronenkanal. In rot sind diejenigen Fehlerfunktionen eingezeichnet, die die jeweils größten bzw. kleinsten Skalierungsfaktoren für Signal und $t\bar{t}$ -Untergrund liefern.

benen Fehlers geändert, so müssten auch die anderen Parameter geändert werden, um wieder die Daten zu beschreiben.

Man muss also zu einem neuen Satz an unkorrelierten Parametern (p'_1, \dots, p'_8) wechseln. Dies ist mit Hilfe der Kovarianzmatrix möglich.

Die Kovarianz ist ein Maß dafür, wie stark zwei Werte X und Y mit gemeinsamer Wahrscheinlichkeitsverteilung miteinander korreliert sind, also wie häufig hohe Werte von X mit hohen bzw. niedrigen Werten von Y einhergehen. Sie ist folgendermaßen definiert[4, Seite 845]:

Seien X und Y zwei reelle Zufallsvariablen mit den Erwartungswerten $E(X)$, $E(Y)$ und der Erwartungswert $E(X \cdot Y)$ existiert, dann heißt

$$Cov(X, Y) := E[(X - E(X)) \cdot (Y - E(Y))] \quad (4.11)$$

die Kovarianz von X und Y .

Dies lässt sich umformen zu

$$Cov(X, Y) = E[X \cdot Y - E(X) \cdot Y - X \cdot E(Y) + E(X) \cdot E(Y)] \quad (4.12)$$

$$= E(X \cdot Y) - E(X) \cdot E(Y) - E(X) \cdot E(Y) + E(X) \cdot E(Y) \quad (4.13)$$

$$= E(X \cdot Y) - E(X) \cdot E(Y) \quad (4.14)$$

Insbesondere folgt daraus, dass die Kovarianz einer Variablen mit sich selbst gleich der Varianz dieser Variablen ist.

$$Cov(X, X) = \sigma_X^2 \quad (4.15)$$

Außerdem ist die Kovarianz symmetrisch unter der Vertauschung der Variablen.

$$Cov(X, Y) = Cov(Y, X) \quad (4.16)$$

Sind die beiden Variablen X und Y unabhängig voneinander, d. h.

$$E(X \cdot Y) = E(X) \cdot E(Y) \quad (4.17)$$

so folgt

$$Cov(X, Y) = 0 \quad (4.18)$$

Im Laufe des Fits wird die Kovarianzmatrix V automatisch erstellt. Diese enthält als Einträge die Kovarianzen der acht Parameter.

$$V_{j,k} = Cov(p_j, p_k) \quad 1 \leq j, k \leq 8 \quad (4.19)$$

Hierfür nutzt man, dass

$$V_{j,k}^{-1} = \frac{1}{2} \cdot \frac{\partial^2 \chi^2(p_1, \dots, p_8)}{\partial p_j \partial p_k} \Big|_{\hat{p}_j, \hat{p}_k} \quad (4.20)$$

wobei \hat{p}_j den bestimmten Messwert bezeichnet (siehe [20]). Die Kovarianzmatrix ist wegen Gleichung 4.16 symmetrisch.

Die zu dieser Darstellung gehörige Basis wird von den acht achtdimensionalen kanonischen Einheitsvektoren

$$\vec{e}_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j - te \text{ Komponente} \quad (4.21)$$

gebildet.

Das zu lösende Problem ist, dass diese Matrix im alten Koordinatensystem $S = (\vec{e}_1, \dots, \vec{e}_8)$ auch auf den Nichtdiagonalelementen Einträge ungleich null besitzt, d. h. die Parameter sind miteinander korreliert. Dies wird nun behoben, indem die Kovarianzmatrix diagonalisiert wird. Es wird in ein System $S' = (\vec{e}'_1, \dots, \vec{e}'_8)$ transformiert. Dabei sind die acht Basisvektoren \vec{e}'_j gerade die normierten Eigenvektoren der Kovarianzmatrix. Aus deren Symmetrie folgt die Orthogonalität der Eigenvektoren.

In der neuen Basis sieht die Kovarianzmatrix wie folgt aus:

$$V' = \begin{pmatrix} \sigma_1'^2 & 0 & \dots & 0 \\ 0 & \sigma_2'^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & \sigma_8'^2 \end{pmatrix} \quad (4.22)$$

mit den zu den Eigenvektoren \vec{e}'_j gehörigen Eigenwerten $\lambda_j = \sigma_j'^2$ als Diagonaleinträgen.

Auf diese Weise werden acht unabhängige Fitparameter und die dazugehörigen Fehler gefunden. Dies wird nun genutzt, um 16 Fehlerfunktionen pro Template zu erstellen. Als Startpunkt wählt man die acht Parameter der Fitfunktion $F_i^0(A')$ (siehe Gleichung 4.10), also der Funktion, die der Fit als Ergebnis liefert. In der Vektorschreibweise ist dieser Parametersatz gegeben als

$$\vec{p}_0 = \sum_{j=1}^8 p_j \cdot \vec{e}_j \quad (4.23)$$

Dieser wird nun für jeden unabhängigen Parameter p_j einmal um dessen Fehler σ'_j nach oben bzw. nach unten variiert.

$$p_{j,\pm}^{\vec{}} = \vec{p}_0 \pm \sigma'_j \cdot \vec{e}'_j \quad (4.24)$$

Geometrisch kann das Vorgehen folgender Maßen interpretiert werden: Jede Parameterkonfiguration entspricht einem Punkt im achtdimensionalen Parameterraum. Die Konfiguration \vec{p}_0 des besten Fits F_i^0 ist umgeben von einem Ellipsoiden, in dessen Innern sich 68% aller möglichen Parameterkonfigurationen befinden. Zur Fehlerabschätzung werden die Punkte an den Enden der Halbachsen betrachtet. Letztere sind gegeben durch die Eigenvektoren der Kovarianzmatrix und ihre Länge ist gleich der Wurzel der zugehörigen Eigenwerte.

Die 16 Fehlerfunktionen sind in den Abbildungen 4.4, 4.5 und 4.6 rot eingezeichnet. Sie liegen

alle nahe beim besten Fit und beschreiben so noch die Daten im Rahmen der Messgenauigkeit. Gleichzeitig füllen sie das gelb eingezeichnete $1 - \sigma$ -Konfidenzband mit einer Häufung in der Mitte und abfallender Dichte zu den Rändern hin aus. Dies zeigt, dass die Wahl der Fehlerfunktionen sinnvoll ist.

Mit den so gefundenen Fehlerfunktionen wird der Fehler σ_{shape} abgeschätzt. Hierfür wird ein konservativer Ansatz gewählt: Der finale Fit an die Daten mit Gleichung 4.10 wird für alle $17^3 = 4913$ möglichen Kombinationen aus besten Fits und Fehlerfunktionen für die drei Templates $t\bar{t}$, Signal und Multijet wiederholt. Der Ansatz ist deshalb konservativ, weil zu erwarten ist, dass nicht alle drei tatsächlichen Verteilungen auf dem Rand des Fehlerbereichs liegen, sondern eher in dessen Innern. Als Fehler wird nun die größte Abweichung der erhaltenen Skalierungsfaktoren s_{Signal} und $s_{t\bar{t}}$ vom Ergebnis des Fits F_{Daten}^0 gewählt. Diese ist nach oben und nach unten sehr symmetrisch und ist in etwa noch einmal so groß wie der Fehler σ_{fit} .

Um die Fitmethode auf Instabilitäten zu prüfen wurden in Abbildung 4.7 die vier Fitergebnisse mit dem jeweils größten bzw. kleinsten Skalierungsfaktor für das Signal und das $t\bar{t}$ -Template eingezeichnet. Auch diese liegen sehr nahe beim besten Fit und stimmen somit mit den Daten überein. Der Algorithmus erscheint somit stabil.

Um die Güte des Fits bewerten zu können wird ebenfalls für sämtliche Fehlerkombinationen der χ^2 -Wert des Fits bestimmt und der niedrigste gespeichert. Der erhalten Wert $\chi_{red}^{2,Daten,best}$ ist für den Elektronenkanal in Tabelle A.3 und für den Myonenkanal in Tabelle A.4 festgehalten und liegt etwas niedriger als beim Fit mit den jeweils besten Fitfunktionen aus den Fits an die drei Templates $t\bar{t}$, Signal und Multijet. Da es sich hier nur um die χ_{red}^2 -Werte für einzelne Punkte im durch die Fits erlaubten Parameterraum handelt, ist dieser Wert eine Obergrenze für den im Rahmen der Messgenauigkeit geringsten χ_{red}^2 -Wert. Dies zeigt, dass die Fitmethode Ergebnisse liefert, die im Rahmen ihrer Messgenauigkeit sehr gut mit den Daten vereinbar sind. Dies bestätigt die in Kapitel 4.5.1 gemachte Annahme kontinuierlicher Verteilungen.

Nun kann der Gesamtfehler des Fits σ_{fit} bestimmt werden. Hierzu werden die beiden Fehler σ_{norm} und σ_{shape} als unkorreliert angenommen. Dies ist insofern gerechtfertigt, als dass sie auf unterschiedliche Weisen bestimmt wurden. Außerdem liefert der Fit für jede Kombination von Fehlerfunktionen zur Bestimmung von σ_{shape} einen Fehler σ_{norm} , der immer in etwa gleich groß ist. Somit ergibt sich für den Gesamtfehler

$$\sigma_{fit} = \sqrt{\sigma_{norm}^2 + \sigma_{shape}^2} \quad (4.25)$$

4.5.4 Ergebnisse

Die Messwerte inklusive des Fehlers σ_{fit} durch den Fit an sich mit seinen beiden Teilkomponenten σ_{norm} und σ_{shape} wird für den Elektronenkanal in Tabelle 4.2a und für den Myonenkanal in Tabelle 4.2b angegeben. Die beiden Teilfehler σ_{shape} und σ_{norm} sind in etwa gleich groß. Die Fehler σ_{fit} sind in etwa gleich groß wie beim Fit mit TFractionFitter (siehe Tabelle 4.1). Diese Konsistenz bestätigt die Sinnhaftigkeit der Fehlerabschätzung in Kapitel 4.5.3. Die gemessenen $t\bar{t}$ -Anteile sind für die beiden Methoden ebenfalls sehr ähnlich.

(a) $t\bar{t}$ -Anteil im Elektronenkanal					(b) $t\bar{t}$ -Anteil im Myonenkanal				
#Jets	$t\bar{t}$ -Anteil [%]	σ_{norm}	σ_{shape}	σ_{fit}	#Jets	$t\bar{t}$ -Anteil [%]	σ_{norm}	σ_{shape}	σ_{fit}
3	3,94	0,42	0,59	0,72	3	4,41	0,49	0,68	0,84
4	14,99	0,74	0,96	1,21	4	15,34	0,82	0,97	1,27
5	26,19	1,89	2,17	2,88	5	29,29	2,19	3,38	4,02
≥ 6	37,32	4,98	5,97	7,78	≥ 6	43,69	5,40	5,92	8,01

Tabelle 4.2: $t\bar{t}$ -Anteil in der Signalregion (a) im Elektronenkanal und (b) im Myonenkanal bestimmt mit einem Fit mit Funktionen mit statistischen Fehlern durch den Fit. Der Fehler σ_{norm} ist der von TH1::Fit direkt gelieferte statistische Fehler auf die Skalierung der $t\bar{t}$ -Funktion im Fit an die Daten in der Signalregion, σ_{shape} ist der Fehler durch die Unsicherheit bei den Fitfunktionen und σ_{fit} ist als deren quadratische Summe die Gesamtunsicherheit aufgrund des Fits.

4.6 Fitbezogene Fehlerquellen

Die Unsicherheiten des $t\bar{t}$ -Fits in beiden Methoden (TFractionFitter und Fit mit Funktionen) sind statistisch begründet. Würde man mehr Daten nehmen, so würden die für den Fit verwendeten Templates gegen die Verteilung in der Signalregion konvergieren und die Unsicherheiten würden sich verringern. Gleiches gilt für die Normierung des Multijet-Templates beim $t\bar{t}$ -Fit in der Signalregion, sowie in der $t\bar{t}$ -Kontrollregion beim Subtrahieren der Kontamination des datenbasierten $t\bar{t}$ -Templates (siehe Kapitel 4.3). Beide Werte wurden in Kapitel 3 mit Hilfe eines Fits mit TFractionFitter bestimmt und besitzen einen statistischen Fehler. Dessen Auswirkung auf die Abschätzung des $t\bar{t}$ -Untergrundes wird im Folgenden untersucht.

4.6.1 Normierung des Multijet-Templates in der Signalregion

Die Normierung des Multijet-Templates wurde in Kapitel 3 innerhalb einer statistischen Unsicherheit bestimmt. Die Werte und Unsicherheiten sind für den Elektronenkanal in Tabelle 3.1a und für den Myonenkanal in Tabelle 3.1b zusammengefasst. Um die hieraus verursachte Unsicherheit auf den $t\bar{t}$ -Untergrund zu ermitteln, wird die fixierte Normierung des Multijet-Templates einmal um ihren Fehler nach oben und einmal nach unten variiert und der Fit wiederholt. Anschließend wird die größere der resultierenden Abweichungen des $t\bar{t}$ -Anteils in der Signalregion als Fehler $\sigma_{QCD,SR}$ angenommen. Dies wird für beide Fitmethoden durchgeführt. Die Ergebnisse sind für den Elektronenkanal in den Tabellen 4.3 und 4.4 und für den Myonenkanal in den Tabellen 4.5 und 4.6 zusammengefasst. Es zeigt sich, dass die Normierung des Multijet-Templates im Vergleich zur Unsicherheit σ_{fit} nur einen geringen Einfluss auf das Fitergebnis hat.

4.6.2 Multijet-Kontamination in der $t\bar{t}$ -Kontrollregion

In der $t\bar{t}$ -Kontrollregion wird der Multijet-Anteil ebenfalls mit einem Fit bestimmt und ist daher nur innerhalb der statistischen Unsicherheit des Fits bekannt (siehe für den Elektronenkanal Tabelle 3.1a und für den Myonenkanal Tabelle 3.1b). In der $t\bar{t}$ -Kontrollregion wurde er gebraucht, um das datenbasierte $t\bar{t}$ -Template vom Untergrund zu bereinigen. Für die Abschätzung der dar-

aus resultierenden Unsicherheit auf den $t\bar{t}$ -Untergrund wird der bestimmte Multijet-Anteil jeweils einmal plus und minus seine Unsicherheit variiert und der $t\bar{t}$ -Fit in der Signalregion mit den neuen Templates wiederholt. Auch diesmal wird die größere der beiden resultierenden Abweichungen des Ergebnisses für den $t\bar{t}$ -Anteil als Fehler $\sigma_{QCD,CR}$ angenommen. Auch dies wird für beide Fitmethoden durchgeführt. Die Ergebnisse sind ebenfalls für den Elektronenkanal in den Tabellen 4.3 und 4.4 und für den Myonenkanal in den Tabellen 4.5 und 4.6 zu finden. Auch dieser Fehler spielt verglichen mit der Unsicherheit des Fits σ_{fit} eine geringe Rolle.

4.6.3 Der statistische Fehler

Der statistische Fehler σ_{stat} setzt sich nun aus den drei beschriebenen Komponenten σ_{fit} (Fehler des Fits), $\sigma_{QCD,SR}$ (Fehler durch Unsicherheit auf Normierung des Multijet-Templates in der Signalregion) und $\sigma_{QCD,CR}$ (Fehler durch Unsicherheit auf Normierung der Multijet-Templates in der $t\bar{t}$ -Kontrollregion) zusammen. Diese sind als statistische Fehler aufgrund dreier separat durchgeführter Fits unabhängig voneinander. Daher ist der statistische Fehler

$$\sigma_{stat} = \sqrt{\sigma_{fit}^2 + \sigma_{QCD,SR}^2 + \sigma_{QCD,CR}^2} \quad (4.26)$$

Auch er findet sich in den Tabellen 4.3, 4.4, 4.5 und 4.6.

4.7 Systematische Unsicherheiten

Neben den statistischen Fehlern gibt es noch mehrere systematische Fehlerquellen. Diese haben für unterschiedliche Wertebereiche der transformierten Aplanarität verschieden starken Einfluss. Daher wird zuerst untersucht, welche Bereiche besonders stark betroffen sind und wie durch eine geschickte Wahl des Fitbereichs der systematische Fehler reduziert werden kann. Anschließend wird der verbleibende Einfluss systematischer Fehlerquellen auf den $t\bar{t}$ -Untergrund abgeschätzt.

4.7.1 Bestimmung des optimalen Fitbereichs

Für eine sinnvolle Wahl des Fitbereichs wurden zwei Fehlerquellen untersucht. Zum einen war dies die Jetenergieskala (JES), zum anderen Pile-up, der nur im Myonenkanal untersucht wurde. Die Ergebnisse sind dabei auf den Elektronenkanal übertragbar, da dort genauso gleichzeitige Kollisionen voneinander separiert werden müssen.

Die Jetenergieskala betrifft die Umrechnung von Detektorsignalen in die Energie des detektierten Jets. Die Kalibration ist jedoch nur innerhalb einer Unsicherheit korrekt. Um die Auswirkung dieser Unsicherheit auf den Messwert abzuschätzen lässt man die Kalibration der Daten unverändert und variiert die Monte Carlo Simulationen folgendermaßen: Man geht davon aus, dass eine Umrechnung am Rand des Fehlerbereichs die richtige wäre (JESup/down: höhere/niedrigere Jetenergie) und simuliert dann, wie die Ereignisse in der davon abweichenden Umrechnung des Detektors aussehen würden.

Die Variation der Jetenergieskala führt in den Monte Carlo Simulationen zu zwei Veränderungen, die in Abbildung 4.8 zu sehen sind. Hier ist das Signal-Template für vier Jets im Elektronenkanal dargestellt. Zum einen ändert sich die Anzahl der Ereignisse für die einzelnen Jetmultiplizitäten, zum anderen ändert sich auch das Profil der Verteilung. Die Änderung der Normierung

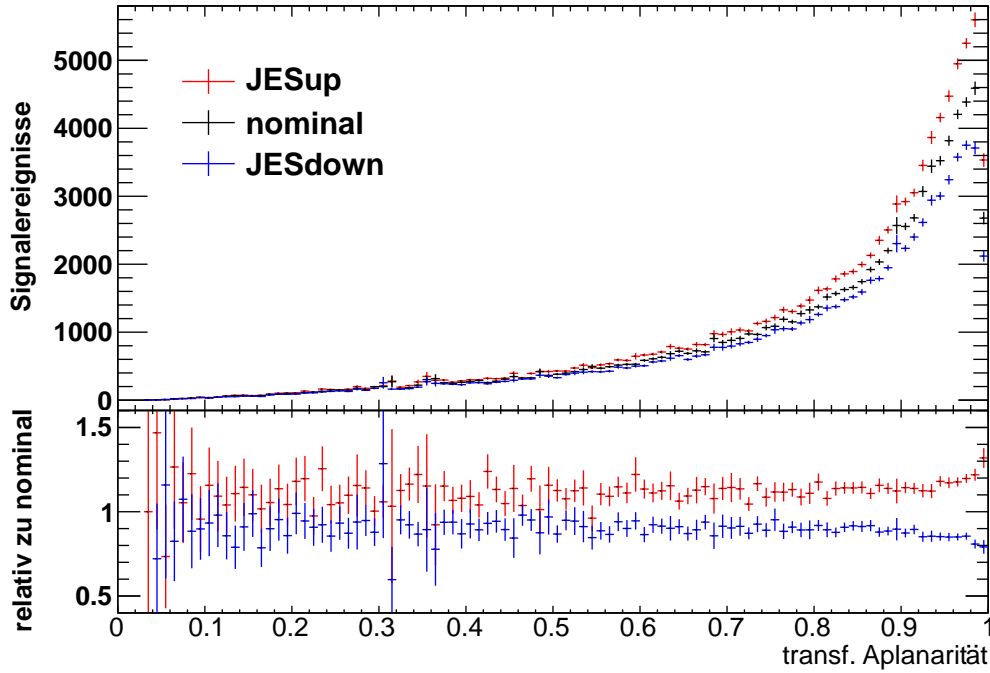


Abbildung 4.8: Signal-Templates mit variiertener Jetenergieskala im Vergleich und relativ zum nominalen Template, das zur Messwertbestimmung verwendet wurde, für Ereignisse mit vier Jets im Elektronenkanal. Bis $A' = 0,9$ unterscheiden sich die Templates überwiegend in ihrer Normierung. Erst danach laufen sie auseinander. Dieser Bereich wird daher aus dem $t\bar{t}$ -Fit ausgeschlossen.

ist darauf zurückzuführen, dass nun je nach Korrektur mehr oder weniger Jets die Schwelle von $p_T = 30 \text{ GeV}/c$ überschreiten und folglich als Signaljets klassifiziert werden. Es kommt somit zu einer Migration zwischen den Jetmultiplizitäten. Dies ist an der Verschiebung der Templates JESup und JESdown nach oben beziehungsweise nach unten zu erkennen. Da der Fit die Normierung der Templates ändert, führt dies lediglich zu einer Änderung des Skalierungsfaktors, aber nicht unbedingt zu einer Änderung des $t\bar{t}$ -Anteils. Für diesen ist überwiegend das Profil relevant, dass sich für hohe transformierte Aplanaritäten ändert, was am Anstieg beziehungsweise am Abfall im relativen Histogramm im Bereich von $A' > 0,9$ zu erkennen ist. Am stärksten ist hiervon das Signal-Template aus W+Jets Ereignissen und den kleinen Untergrundprozessen aus Kapitel 2.5 betroffen, da es vollständig auf Monte Carlo Simulationen beruht, während das Multijet- und das $t\bar{t}$ -Template aus Messdaten stammen¹. Wird dieser Bereich aus dem Fit ausgeschlossen, so kann diese Unsicherheit deutlich reduziert werden.

Als zweites wurde im Myonenkanal der Pile-up untersucht. Je mehr Kollisionen dem Ereignis zugeordnet werden, desto größer ist die Gefahr einer Fehlzuordnung und damit eines falsch eingeordneten Signals. Dies kann zu Verzerrungen in den Templates führen, die möglichst aus dem Fitbereich ausgeschlossen werden sollten. Um kritische Bereiche zu erkennen wurden die Templates in vier

¹Auch hier spielen Monte Carlo Simulationen beim Abziehen der Kontamination eine Rolle. Für das Multijet-Template beträgt die Kontamination unter 4% und hat daher geringen Einfluss. Für das $t\bar{t}$ -Template beträgt die Kontamination durch simulierte Prozesse 6-20% (siehe Tabellen A.1 und A.2). Hinzu kommt, dass für die Korrektur der Verzerrung durch das b-Veto in Kapitel 4.3.3 Monte Carlo Simulationen benötigt werden. Ob die so verursachten Unsicherheiten durch eine andere Wahl einer alternativen $t\bar{t}$ -Kontrollregion reduziert werden kann, wird in Kapitel 4.10 untersucht.

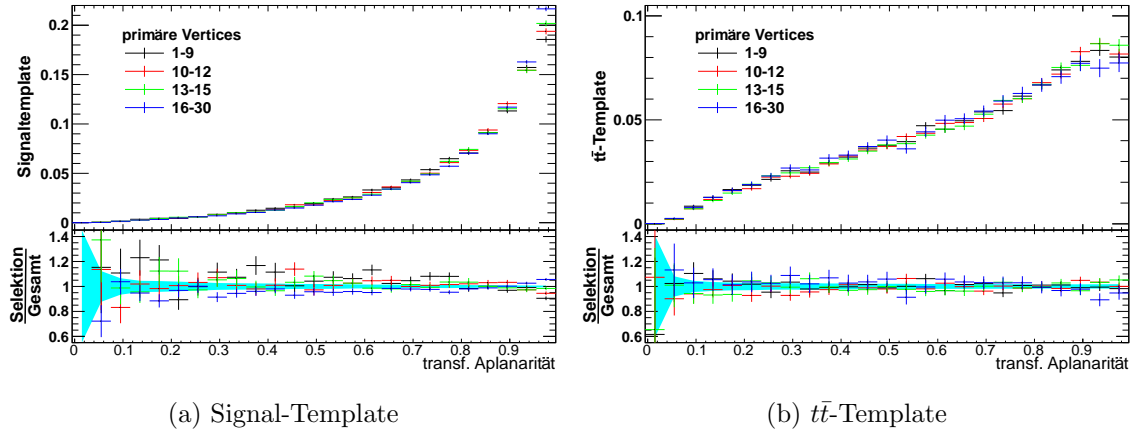


Abbildung 4.9: Das Signal-Template bestehend aus W+Jet Ereignissen und kleinen Untergrundprozessen in der Signalregion und das datenbasierte $t\bar{t}$ -Template nach Abzug der Kontamination in der $t\bar{t}$ -Kontrollregion aufgeteilt nach der Anzahl primärer Vertices für vier Jets im Myonenkanal. Oben: Normiert im direkten Vergleich untereinander. Unten: Normiert relativ zur Summe aller vier Teilselektionen, die in Cyan mit ihrem statistischen Fehlerband eingezeichnet ist.

Subtemplates mit unterschiedlicher Anzahl an primären Vertices aufgeteilt. Die Bereiche wurden dabei so gewählt, dass die in ihnen enthaltene Statistik in etwa gleich groß ist. Die gewählten Subtemplates enthalten Ereignisse mit 1-9, 10-12, 13-15 und 16-30 Vertices. Diese Templates wurden dann zur besseren Vergleichbarkeit auf eins normiert und gemeinsam bezüglich der transformierten Aplanarität dargestellt.

Wie in Abbildung 4.9 zu sehen ist, treten sowohl für das Monte Carlo basierte Signal-Template (a) als auch für das datenbasierte $t\bar{t}$ -Template (b) keine signifikanten Unterschiede zwischen den einzelnen Teilselektionen auf². Daher kann auf eine weitere Verringerung des Fitbereichs verzichtet werden.

4.7.2 Detektorbasierte Unsicherheiten

Zur Abschätzung der systematischen Unsicherheit wurden die vier dominanten Fehlerquellen aus der Analyse der Daten aus dem Jahr 2011 untersucht [2]. Diese sind die Jetenergieskala (JES), deren Auflösung (JER), sowie die MET Soft-Term Skala (MET) mit deren Auflösung (METres). Die MET Soft-Term Skala dient zur Berechnung der Energie von Detektorsignalen, die keinem Jet oder Lepton zugeordnet werden können. Dies ist insbesondere wichtig, um die fehlende transversale Energie E_t^{miss} zu bestimmen. Dies wirkt sich daher indirekt über die Bestimmung des Multijet-Anteils (Kapitel 3), aber auch direkt in den Templates für den abschließenden $t\bar{t}$ -Fit über der transformierten Aplanarität aus. Für beide Skalen ist die Auflösung JER und METres nur innerhalb eines Fehlers bekannt. Dies bedeutet, dass die in Monte Carlo Simulationen bekannten Energien verschmiert werden müssen, wobei die Breite der Verteilungen nur mit einer Unsicherheit bekannt ist. Die resultierenden Unsicherheiten auf den $t\bar{t}$ -Untergrund werden im Folgenden abgeschätzt.

Für die Jetenergieskala und die MET Soft-Term Skala werden für alle Monte Carlo Simulationen

²In der Monte Carlo Vorhersage durch POWHEG und auch durch SHERPA treten hingegen Unterschiede bei Ereignissen mit 16-30 primären Vertices in der $t\bar{t}$ -Kontrollregion auf. Dies ist in Abbildung A.3 zu sehen.

Templates verwendet, die den detektierten Signalen höhere (JESup und METup) beziehungsweise niedrigere Energien (JESdown und METdown) zuordnen. Es ändern sich somit das Signal-Template, das Multijet-Template durch die Subtraktion des Untergrunds und das $t\bar{t}$ -Template sowohl durch den Abzug der Kontamination, als auch durch die Korrektur der Verzerrung durch das b-Veto. Außerdem ändert sich der Multijet-Anteil, da sich auch hier das Ergebnis des Fits (siehe Kapitel 3) ändert. Mit den neuen Templates wird der $t\bar{t}$ -Fit wiederholt. Hierdurch erhält man neue Werte für den $t\bar{t}$ -Anteil. Nun wird für jede Skala die größere der beiden Abweichungen vom ursprünglichen Messwert als Fehler angenommen. Auf diese Weise erhält man die beiden Fehler σ_{JES} und σ_{MET} .

Für die Abschätzung der Unsicherheit aufgrund der Auflösungen JER und METres ist insbesondere der Fall einer schlechteren Auflösung als angenommen interessant. Es werden nun die gleichen Schritte wie bei der Abschätzung der Fehler σ_{JES} und σ_{MET} durchgeführt, allerdings mit nur einem Template mit einer im Rahmen der Unsicherheit schlechteren Auflösung. Auf diese Weise erhält man die beiden systematischen Fehler σ_{JER} für die Auflösung der Jetenergieskala und σ_{METres} für die Auflösung der MET Soft-Term Skala. Für den systematischen Fehler insgesamt gilt damit

$$\sigma_{sys} = \sqrt{\sigma_{JES}^2 + \sigma_{MET}^2 + \sigma_{JER}^2 + \sigma_{METres}^2} \quad (4.27)$$

Die Werte aller Fehler finden sich für den Elektronenkanal je nach Fitmethode in den Tabellen 4.3 und 4.4 und für den Myonenkanal in den Tabellen 4.5 und 4.6.

4.8 Gesamtunsicherheit des $t\bar{t}$ -Anteils in der Signalregion

Aus dem statistischen Fehler σ_{stat} und dem systematischen Fehler σ_{sys} folgt nun der Gesamtfehler

$$\sigma_{tot} = \sqrt{\sigma_{stat}^2 + \sigma_{sys}^2} \quad (4.28)$$

Auch dieser ist im Elektronenkanal für den Fit mit TFractionFitter in Tabelle 4.3 und für den Fit mit Funktionen in Tabelle 4.4, sowie für den Myonenkanal für den Fit mit TFractionFitter in Tabelle 4.5 und für den Fit mit Funktionen in Tabelle 4.6 zusammengefasst. Es fällt auf, dass der Gesamtfehler meist durch die systematischen Unsicherheiten dominiert wird. Die vier diskutierten systematischen Unsicherheiten sind für den Fit mit Funktionen zumeist kleiner als mit TFractionFitter, wodurch auch der Gesamtfehler kleiner ist. Für fünf Jets im Elektronenkanal ist der Gesamtfehler σ_{tot} für beide Methoden in etwa gleich. Dies liegt an einem auffällig kleinen Wert von σ_{JER} bei der Verwendung von TFractionFitter. Durch den Fit mit Funktionen konnte außerdem die systematische Unsicherheit für Ereignisse mit fünf und mit mehr als sechs Jets im Myonenkanal unter die statistischen Unsicherheiten reduziert werden.

4.9 Messwerte

Die Messwerte für den $t\bar{t}$ -Anteil in der Signalregion $f_{t\bar{t}}$ und alle berücksichtigten Unsicherheiten sind für den Elektronenkanal für den Fit mit TFractionFitter in Tabelle 4.3 und für den Fit mit Funktionen in Tabelle 4.4 zusammengefasst. Die Messwerte für den Myonenkanal finden sich in den Tabellen 4.5 und 4.6.

$t\bar{t}$ -Anteil im Elektronenkanal mit TFractionFitter [%]											
	Wert	statistische Fehler			systematische Fehler				gesamt		
#Jets	$f_{t\bar{t}}$	σ_{fit}	$\sigma_{QCD,SR}$	$\sigma_{QCD,CR}$	σ_{JES}	σ_{MET}	σ_{JER}	σ_{METres}	σ_{stat}	σ_{sys}	σ_{tot}
3	3,46	0,83	0,07	0,02	1,73	0,46	0,87	0,77	0,83	2,14	2,29
4	12,96	1,08	0,12	0,09	1,74	1,23	3,51	1,78	1,09	4,47	4,61
5	27,85	2,85	0,22	0,19	1,13	0,78	0,89	2,06	2,87	2,63	3,89
≥ 6	35,53	7,80	0,22	0,12	8,64	1,97	10,44	1,76	7,81	13,81	15,86

Tabelle 4.3: $t\bar{t}$ -Anteil im Elektronenkanal bestimmt mit TFractionFitter in Prozent

$t\bar{t}$ -Anteil im Elektronenkanal beim Fit mit Funktionen [%]											
	Wert	statistische Fehler			systematische Fehler				gesamt		
#Jets	$f_{t\bar{t}}$	σ_{fit}	$\sigma_{QCD,SR}$	$\sigma_{QCD,CR}$	σ_{JES}	σ_{MET}	σ_{JER}	σ_{METres}	σ_{stat}	σ_{sys}	σ_{tot}
3	3,94	0,72	0,06	0,03	0,65	0,52	0,53	0,02	0,73	0,99	1,23
4	14,99	1,21	0,12	0,09	1,49	1,00	2,38	0,56	1,22	3,03	3,27
5	26,19	2,88	0,30	0,20	2,06	0,05	1,96	0,20	2,90	2,85	4,07
≥ 6	37,32	7,78	0,35	0,93	3,07	0,87	7,42	1,66	7,84	8,25	11,38

Tabelle 4.4: $t\bar{t}$ -Anteil im Elektronenkanal bestimmt mit Funktionen in Prozent

$t\bar{t}$ -Anteil im Myonenkanal mit TFractionFitter [%]											
	Wert	statistische Fehler			systematische Fehler				gesamt		
#Jets	$f_{t\bar{t}}$	σ_{fit}	$\sigma_{QCD,SR}$	$\sigma_{QCD,CR}$	σ_{JES}	σ_{MET}	σ_{JER}	σ_{METres}	σ_{stat}	σ_{sys}	σ_{tot}
3	5,26	1,16	0,02	0,51	1,74	0,48	1,45	0,31	1,27	2,34	2,66
4	15,48	1,50	0,04	0,72	3,05	0,78	1,82	0,72	1,67	3,71	4,06
5	27,79	3,35	0,46	0,44	1,74	0,62	4,48	0,11	3,41	4,84	5,93
≥ 6	37,01	19,21	0,50	0,95	25,61	1,10	4,46	0,21	19,24	26,11	32,43

Tabelle 4.5: $t\bar{t}$ -Anteil im Myonenkanal bestimmt mit TFractionFitter in Prozent

$t\bar{t}$ -Anteil im Myonenkanal beim Fit mit Funktionen [%]											
	Wert	statistische Fehler			systematische Fehler				gesamt		
#Jets	$f_{t\bar{t}}$	σ_{fit}	$\sigma_{QCD,SR}$	$\sigma_{QCD,CR}$	σ_{JES}	σ_{MET}	σ_{JER}	σ_{METres}	σ_{stat}	σ_{sys}	σ_{tot}
3	4,30	0,78	0,01	0,04	0,94	0,03	0,74	0,12	0,78	1,20	1,43
4	15,69	1,23	0,03	0,10	1,34	0,11	1,14	0,21	1,30	1,78	2,20
5	29,63	3,99	0,36	0,63	0,45	0,51	2,33	0,82	4,05	2,56	4,79
≥ 6	43,35	7,74	0,59	0,80	3,27	2,25	3,90	3,57	7,80	6,61	10,23

Tabelle 4.6: $t\bar{t}$ -Anteil im Myonenkanal bestimmt mit Funktionen in Prozent

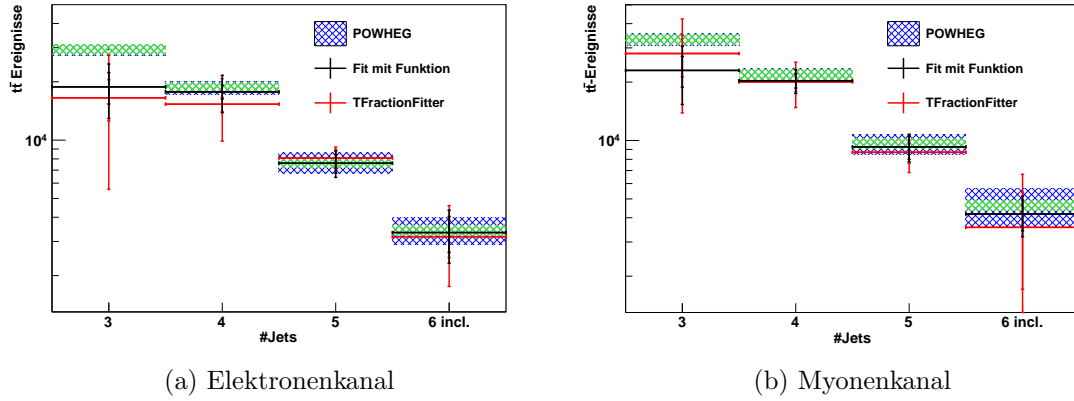


Abbildung 4.10: Anzahl der $t\bar{t}$ -Ereignisse in der Signalregion (a) im Elektronenkanal und (b) im Myonenkanal. Die inneren Fehlerbalken sind die statistischen Fehler. Auf die Vorhersage der POWHEG Simulation wurde eine theoretische Unsicherheit auf den Wirkungsquerschnitt von 6% angenommen (grün). Dazu kommen auch hier die vier diskutierten detektorbasierten systematischen Unsicherheiten (blau).

Die gemessenen $t\bar{t}$ -Anteile $f_{t\bar{t}}$ beider Methoden sind bereits im Rahmen der statistischen Unsicherheit sehr gut miteinander vereinbar.

Der $t\bar{t}$ -Anteil lässt sich mittels

$$N_{t\bar{t}}^{SR} = f_{t\bar{t}} \cdot N_{Daten}^{SR} \quad (4.29)$$

in die Zahl der $t\bar{t}$ -Ereignisse in der Signalregion $N_{t\bar{t}}^{SR}$ umrechnen. Die Ergebnisse sind in Tabelle A.5 zusammengefasst und für den Elektronenkanal in Abbildung 4.10a und für den Myonenkanal in Abbildung 4.10b, gemeinsam mit der Vorhersage von POWHEG, graphisch dargestellt.

Die datenbasierten Ergebnisse sind nicht nur miteinander, sondern, zumindest bezüglich ihrer Normalisierung, auch mit der theoretischen Vorhersage von POWHEG konsistent. Lediglich für drei Jets gibt es Abweichungen, die in beiden Zerfallskanälen auftreten. Es fällt auf, dass POWHEG mit den berücksichtigten Unsicherheiten Vorhersagen macht, die für Ereignisse mit drei beziehungsweise vier Jets deutlich kleinere Fehlerbalken aufweisen als die datenbasierten Methoden. Um die Zuverlässigkeit dieser Vorhersagen bewerten zu können, bleibt zu untersuchen, wie gut sie die gemessenen Verteilungen des $t\bar{t}$ -Untergrundes beschreiben. Dies wird in Kapitel 4.9.2 mit den Variablen H_T und 1-Jet P_T getan.

4.9.1 Vergleich der Fitmethoden

Wie in den Tabellen 4.3, 4.4, 4.5 und 4.6 zu sehen ist, führen der Fit mit TFractionFitter und der Fit mit Funktionen im Rahmen ihrer statistischen Unsicherheiten σ_{fit} zu den selben Ergebnissen. In dieser Hinsicht sind sie äquivalent. Allerdings ist die Unsicherheit, insbesondere die systematische Unsicherheit, für den Fit mit Funktionen deutlich geringer als beim Fit mit TFractionFitter. Dies hat damit zu tun, welche Fitbereiche für den Fit relevant sind. Der Fit mit Funktionen ist durch die Berücksichtigung des Verlaufs über benachbarte und deshalb korrelierte Bins in der Lage, auch aus Regionen mit geringer Statistik Informationen zu gewinnen. Konkret sind dies hier die Regionen mit geringer transformierter Aplanarität ($A' < 0,25$), also sphärische Ereignisse.

Diese streuen auf große Detektorbereiche, wodurch sich systematische Fehlerquellen teils wegmiteln. Planare Ereignisse sind dagegen anfälliger für systematische Fehlerquellen. TFractionFitter berücksichtigt dagegen keine Korrelation zwischen benachbarten Bins. Im Bereich geringer Statistik sind zudem die relativen Fehler sehr groß. Daher kann TFractionFitter eine falsche Normierung durch Variation der einzelnen Bins ausgleichen und gewinnt so kaum Information. Dies führt zu Instabilitäten, wegen denen der ganze Bereich $A' < 0,25$ ausgeschlossen werden musste. Im Bereich hoher Statistik hingegen sind die relativen Fehler gering und zwingen TFractionFitter zu einer Normierung. Effektiv bezieht diese Fitklasse also den Großteil ihrer Informationen aus diesem Bereich. Da dieser jedoch stark von systematischen Fehlern betroffen ist, wirkt sich dies zwangsläufig auf das Ergebnis aus. Der Fit mit Funktionen dagegen nutzt auch die Bereiche geringer Statistik und ist daher weniger anfällig für systematische Unsicherheiten.

Wäre im systematisch stabilen Bereich (geringes A') dagegen auch eine hohe Statistik vorhanden, so wäre die Fitklasse TFractionFitter im Vorteil, weil sie komplexe Strukturen mit der Auflösung einzelner Bins für den Fit verwenden kann, ohne dabei durch eine womöglich falsche Funktion limitiert zu werden. Durch die geringe Statistik wird dieser Vorteil allerdings zunichte gemacht. Dies kann nur durch das völlige Ausschließen der Bereiche mit geringer Statistik behoben werden. Der Fit mit Funktionen läuft bei der Wahl geeigneter Startwerte dagegen in allen Fitregionen stabil. Zusammenfassend lässt sich also feststellen, dass der Fit mit Funktionen von den beiden Fitmethoden mit konsistenten Ergebnissen die wesentlich weniger fehleranfällig und stabilere ist. In den weiteren Vergleichen zwischen Monte Carlo Simulation und datenbasierten Ergebnissen wird daher nur noch auf das Ergebnis des Fits mit Funktionen eingegangen.

4.9.2 H_T und P_T des führenden Jets

Bis hierhin wurden datenbasierte Templates lediglich dazu verwendet, die Normierung des $t\bar{t}$ -Untergrundes zu ermitteln. Anhand von Daten lassen sich allerdings auch die Vorhersagen differentieller Wirkungsquerschnitte überprüfen. Dies wurde im Elektronenkanal für den transversalen Impuls des führenden Jets ($1\text{-Jet } p_T$) und die skalare Summe der transversalen Impulse H_T getan. Hierzu wurde analog zur A' -Verteilung ein $t\bar{t}$ -Template in der $t\bar{t}$ -Kontrollregion aufgenommen, von Untergrund und W+Jets Ereignissen bereinigt und korrigiert (siehe Kapitel 4.3). Abschließend wurde es auf das Ergebnis aus Tabelle 4.4 normiert und mit der Vorhersage von POWHEG verglichen. In den Abbildungen 4.11a und 4.11b sind die Ergebnisse für Ereignisse mit drei Jets zu sehen.

Es zeigt sich, dass genau wie bei der Normierung auch hier für drei Jets Unterschiede auftreten. Sowohl bei H_T als auch beim 1-Jet p_T gibt es hier einen Bereich niedriger Werte, in dem die POWHEG Simulation stärker ansteigt als die Daten, um dann außerhalb des Fehlerbereichs des relativen Histogramms konstant zu bleiben, was für diesen Bereich einen Unterschied in der Normierung, nicht aber im Profil, bedeutet. Für H_T lässt sich der selbe Anstieg auch für Ereignisse mit vier beziehungsweise fünf Jets beobachten, bleibt dort aber innerhalb der Unsicherheit der Messdaten. Dies ist auf die Übereinstimmung in der Normalisierung aus Messung und Simulation zurückzuführen. Insgesamt zeigt sich für diese Jetmultiplizitäten eine große Übereinstimmung zwischen theoretischer Vorhersage und Experiment.

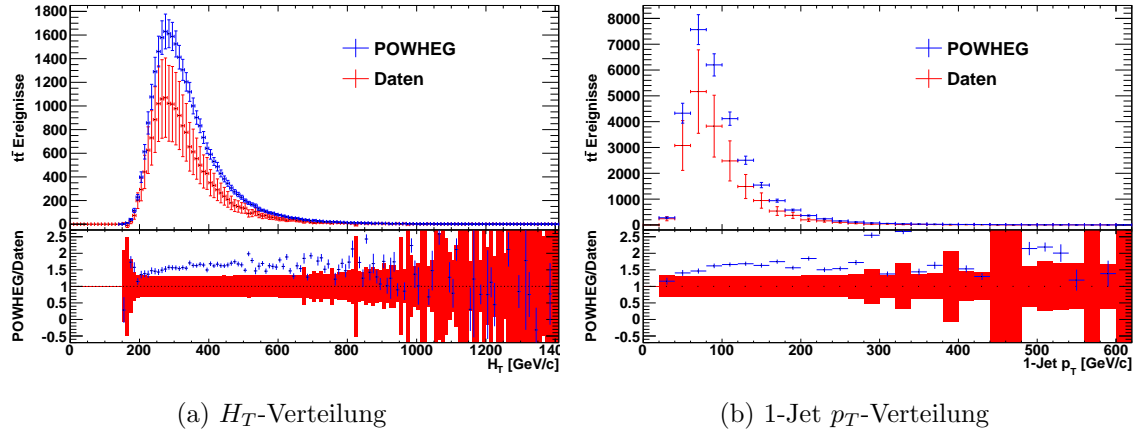


Abbildung 4.11: $t\bar{t}$ -Verteilung von H_T (a) und dem transversalen Impuls des führenden Jets 1-Jet p_T (b) für Ereignisse mit drei Jets im Elektronenkanal. In beiden Fällen besteht ein deutlicher Unterschied in der Normierung zwischen dem datenbasierten Ergebnis und der Vorhersage durch POWHEG. Zudem besteht ein unterschiedlicher Verlauf für niedrige Werte von H_T beziehungsweise 1-Jet p_T . Die großen Fehlerbalken kommen durch die Unsicherheit der Normierung zu Stande. In den Fehlerbalken der POWHEG Vorhersage sind eine theoretische Unsicherheit von 6%, sowie die in Kapitel 4.7 beschriebenen detektorbasierten Unsicherheiten berücksichtigt.

4.10 2-b-Jet-Kriterium in der $t\bar{t}$ -Kontrollregion

Die systematischen Unsicherheiten der Monte Carlo Simulationen gehen an vier Stellen in die Unsicherheit des $t\bar{t}$ -Untergrunds ein:

1. das Profil des Signal-Templates, bestehend aus Simulationen für W+Jets Ereignisse und die kleinen Untergrundprozesse (siehe 2.5, sowohl direkt im $t\bar{t}$ -Fit als auch indirekt über die Bestimmung des Multijet-Untergrundes.)
2. die Kontamination des Multijet-Templates
3. die Kontamination des $t\bar{t}$ -Templates
4. die Korrektur der Verzerrung zwischen Signal- und $t\bar{t}$ -Kontrollregion durch das b-Tagging

Die Punkte drei und vier können durch die Wahl einer alternativen $t\bar{t}$ -Kontrollregion verändert werden. Hier wird untersucht, ob die Forderung eines zweiten identifizierten b-Jets die Unsicherheiten senken kann. Von einer Änderung der b-Taggingeffizienz für die Kontrollregion wird abgesehen, da dies nur hin zu höheren Effizienzen (70%, 80% und 85%) verfügbar wäre. Dies führt insbesondere für b-Taggingeffizienzen von 80% und 85% zu einem massiven Anstieg der Kontamination. Dies wurde für die mit Monte Carlo Simulationen abgeschätzten Untergrundprozesse und W+Jets Ereignisse untersucht. Die Ergebnisse sind in Tabelle A.6 festgehalten. Eine $t\bar{t}$ -Kontrollregion mit einer b-Taggingeffizienz von 70% wird deshalb nicht verwendet, weil dann einige Ereignisse sowohl in der Signalregion mit b-Taggingeffizienz 60% als auch in der $t\bar{t}$ -Kontrollregion vorkommen würden. Die Konsequenz wäre eine Korrelation von statistischen Fehlern in beiden Regionen. Es bleibt somit zu prüfen, ob eine $t\bar{t}$ -Kontrollregion mit mindestens zwei erforderlichen b-Jets und einer b-Taggingeffizienz von 60% zu geringeren Fehlern führt.

Wie in Tabelle A.2 zu sehen ist, kann die Kontamination in der $t\bar{t}$ -Kontrollregion durch die Bedingung eines zweiten identifizierten b-Jets um einen Faktor 3 reduziert werden. Gleichzeitig sinkt auch die Anzahl der Ereignisse um den selben Faktor (siehe Tabelle A.7). Zudem muss eine stärkere Korrektur durchgeführt werden, um das $t\bar{t}$ -Template in der Signalregion zu erhalten (siehe Abbildung A.4).

Mit dem neuen $t\bar{t}$ -Template mit zwei erforderlichen b-Jets wird der $t\bar{t}$ -Fit mit Funktionen (siehe Kapitel 4.5) und die Abschätzung statistischer (siehe Kapitel 4.6) und detektorbasierter (siehe Kapitel 4.7) Unsicherheiten wiederholt.

Die Messwerte für den $t\bar{t}$ -Anteil in der Signalregion sind im Rahmen der statistischen Fehler konsistent mit denen mit nur einem erforderlichen b-Jet. Trotz der deutlich geringeren Zahl an Ereignissen in der $t\bar{t}$ -Kontrollregion steigt der statistische Fehler im Vergleich zu nur einem erforderlichen b-Jet (siehe Tabelle 4.6) nicht oder nur geringfügig an. Ebenso bleibt das χ^2 nahezu unverändert (siehe Tabelle A.4). Die Unsicherheit durch die MET Soft-Term Skala und deren Auflösung bleiben gleich oder nehmen ab. Dies entspricht der Erwartung, da sie sich auch indirekt über die Normierung der Multijet-Kontamination in der $t\bar{t}$ -Kontrollregion auf den $t\bar{t}$ -Fit auswirken. In der Kontrollregion mit mindestens zwei erforderlichen b-Jets ist die Kontamination mit Multijet-Ereignissen jedoch wesentlich kleiner und hat somit geringeren Einfluss. Die Unsicherheiten aufgrund der Jetenergieskala und deren Auflösung nehmen dagegen so stark zu, dass der Gesamtfehler größer als bei Verwendung der Kontrollregion mit mindestens einem erforderlichen b-Jet. Die Ergebnisse sind in Tabelle 4.7 zusammengefasst.

	Wert	statistische Fehler			systematische Fehler				gesamt		
#Jets	$f_{t\bar{t}}$	σ_{fit}	$\sigma_{QCD,SR}$	$\sigma_{QCD,CR}$	σ_{JES}	σ_{MET}	σ_{JER}	σ_{METres}	σ_{stat}	σ_{sys}	σ_{tot}
3	3,90	0,71	0,01	0,04	0,99	0,23	0,83	0,00	0,71	1,31	1,49
4	15,30	1,31	0,03	0,18	1,30	0,09	1,34	0,25	1,32	1,89	2,30
5	29,82	4,41	0,38	0,78	2,36	0,49	4,33	0,67	4,50	5,00	6,73
≥ 6	46,37	9,10	0,65	0,00	5,39	2,56	3,05	1,08	9,12	6,79	11,37

Tabelle 4.7: $t\bar{t}$ -Anteil im Myonenkanal bestimmt mit Funktionen mit zwei erforderlichen b-Jets und einer b-Taggingeffizienz von 60% in der $t\bar{t}$ -Kontrollregion in Prozent.

Die Erklärung für die größeren systematischen Unsicherheiten liegt nicht in Korrektur. Diese ist lediglich ein Maß dafür, wie groß die Wahrscheinlichkeit ist, die b-Jets aus einem $t\bar{t}$ -Ereignis bei einer bestimmten transformierten Aplanarität zu identifizieren. Diese Wahrscheinlichkeit hängt allerdings nicht von der Jetenergieskala ab. Dies ist beispielhaft für vier Jets bei mindestens zwei erforderlichen b-Jets in der $t\bar{t}$ -Kontrollregion in Abbildung A.5 dargestellt.

Eine mögliche Erklärung liegt hingegen in einer Antikorrelation systematischer Unsicherheiten in Signal- und $t\bar{t}$ -Template. Führt die Unsicherheit der Jetenergieskala zur Bildung einer Struktur im Signal-Template relativ zum nominalen Signal-Template, so entsteht im $t\bar{t}$ -Template durch die Subtraktion der Kontamination der Prozesse aus dem Signal-Template eine ähnliche, aber negative Struktur. Die Summe aus Signal-Template und $t\bar{t}$ -Template bleibt also nahezu unverändert. Dies führt zu ähnlichen Skalierungsfaktoren für Signal-Template und $t\bar{t}$ -Untergrund und somit zu geringen systematischen Fehlern. Diese Antikorrelation ist umso stärker, je ähnlicher sich Signal-Template und Kontamination im $t\bar{t}$ -Template sind. Das härtere 2-b-Jet-Kriterium in der

$t\bar{t}$ -Kontrollregion führt jedoch zu größeren Unterschieden zwischen diesen beiden Templates als das b-Tagging mit nur mindestens einem erforderlichen b-Jet. Folglich gleichen sich Unsicherheiten in Signal- und $t\bar{t}$ -Template weniger gut aus und die resultierenden systematischen Unsicherheiten sind größer.

Es wurde also gezeigt, dass ein Kriterium von nur mindestens einem erforderlichen b-Jet in der Signalregion die bessere Wahl ist und deshalb beibehalten werden sollte.

4.11 Konsequenzen für den W+Jets Wirkungsquerschnitt

Die Unsicherheit auf die Anzahl der $t\bar{t}$ -Ereignisse hat Auswirkungen auf die Bestimmung des Wirkungsquerschnitt für W+Jets Ereignisse. Dieser wird durch

$$\sigma_{W+Jets}^{\#Jets} = \frac{N_{Daten}^{\#Jets} - N_{Untergrund}^{\#Jets}}{L \cdot \epsilon}, \quad (4.30)$$

bestimmt, wobei $N_{Daten}^{\#Jets}$ die Zahl der Ereignisse in der Signalregion mit einer bestimmten Anzahl an Jets und $N_{Untergrund}^{\#Jets}$ die Summe über die Ereignisse aller Untergrundprozesse bezeichnet. L ist die Luminosität und ϵ die Nachweiseffizienz. In diese geht unter anderem auch ein, wie viele W+Jets Ereignisse durch die Definition der Signalregion verloren gegangen sind. Dies wird in späteren Analysen erfolgen. Die relative Unsicherheit aufgrund der Unsicherheit des $t\bar{t}$ -Untergrundes kann aber bereits hier abgeschätzt werden. Diese ist

$$\left(\frac{\Delta \sigma_{W+Jets}^{\#Jets}}{\sigma_{W+Jets}^{\#Jets}} \right)_{t\bar{t}} = \frac{\Delta N_{t\bar{t}}^{\#Jets}}{N_{Daten}^{\#Jets} - N_{Untergrund}^{\#Jets}} \quad (4.31)$$

$$\approx \frac{\Delta N_{t\bar{t}}^{\#Jets}}{N_{W+Jets, MC}^{\#Jets}}, \quad (4.32)$$

wobei im zweiten Schritt angenommen wurde, dass die Anzahl der W+Jets Ereignisse aus Monte Carlo Simulationen näherungsweise bekannt ist. Hier wurden wieder ALPGEN und PYTHIA verwendet. Die Ergebnisse für den in dieser Arbeit entwickelten Fit mit Funktionen sind in Tabelle 4.8a und für den herkömmlichen Fit mit TFractionFitter in Tabelle 4.8b zusammengefasst. Hier zeigt sich, dass insbesondere bei sechs und mehr Jets der Vorteil durch den Fit mit Funktionen. Außerdem sind die Ergebnisse ein enormer Fortschritt gegenüber der Analyse der Daten aus dem Jahr 2011[2], als diese Unsicherheit deutlich höher lag und so bei Ereignissen mit sechs und mehr Jets im Elektronenkanal 100% betrug. Dies ist zum einen auf die Anwendung des b-Vetos in der Signalregion zurückzuführen, was an den Ergebnissen aus dem Template-Fit mit TFractionFitter abzulesen ist. Zum anderen konnte die Unsicherheit aufgrund des $t\bar{t}$ -Untergrundes durch die Verwendung des in dieser Arbeit entwickelten Fits mit Funktionen noch weiter reduziert werden, sodass sie im angesprochenen Beispiel von Ereignissen mit sechs und mehr Jets nur noch 26,5% beträgt. Die Unsicherheit durch den $t\bar{t}$ -Untergrund liegt nun auch deutlich unter den Unsicherheiten aufgrund von Jets aus dem Jahr 2011, ist also nicht mehr die dominante Fehlerquelle.

(a) Fit mit Funktionen			(b) TFractionFitter		
#Jets	Elektronenkanal	Myonenkanal	#Jets	Elektronenkanal	Myonenkanal
3	1,8%	2,0%	3	3,3%	3,6%
4	5,5%	3,4%	4	7,8%	6,4%
5	8,1%	8,6%	5	7,8%	10,7%
≥ 6	26,5%	15,6%	≥ 6	36,9%	46,6%

Tabelle 4.8: Unsicherheit des W+Jets Wirkungsquerschnitts durch Unsicherheiten auf den $t\bar{t}$ -Untergrund (a) bestimmt mit dem Fit mit Funktionen und (b) bestimmt mit TFractionFitter.

Kapitel 5

Zusammenfassung

In dieser Arbeit wurde der $t\bar{t}$ -Untergrund und der Multijet-Untergrund in W+Jets Kandidaten mit einer datenbasierten Methode abgeschätzt. Für den Multijet-Untergrund wurde keine Diskussion systematischer Unsicherheiten durchgeführt, da es sich lediglich um eine Vorbereitung zur Bestimmung des $t\bar{t}$ -Untergrundes handelte. Ausführlichere Analysen sollten hier noch durchgeführt werden, da Multijet-Ereignisse für geringe Jetmultiplizitäten den größten Untergrund zu W+Jets Ereignissen darstellen.

Die Bildung von $t\bar{t}$ -Paaren ist im Elektronenkanal ab Ereignissen mit fünf Jets und im Myonenkanal ab Ereignissen mit vier Jets der dominante Untergrundprozess. Daher ermöglicht erst eine präzise Kenntnis dieses Untergrunds eine genaue Analyse von W+Jets Ereignissen. Durch die Anwendung eines b-Vetos in der Signalregion und die Verwendung eines χ^2 -Fits mit Funktionen ist es gelungen, die untersuchten systematischen Unsicherheiten durch die Jetenergieskala, die MET Soft-Term Skala und durch die Auflösung der beiden Skalen je nach untersuchter Jetmultiplizität um einen Faktor 1,5-3 zu reduzieren (siehe Tabellen 4.3, 4.4, 4.5 und 4.6). Hierdurch konnte die resultierende Unsicherheit auf den W+Jets Wirkungsquerschnitt deutlich verringert werden. Unter der Annahme, im Vergleich zur Analyse der Daten aus dem Jahr 2011 [2] gleich bleibender anderer Unsicherheiten, stellt der $t\bar{t}$ -Untergrund nun auch bei Ereignissen mit fünf und mehr Jets nicht mehr die dominierende Fehlerquelle dar. Dies ermöglicht weiterführende Analysen, etwa zur Partonenverteilungsfunktion (PDF) des Protons und zum Test der perturbativen Quantenchromodynamik. Des Weiteren wurden die theoretischen Vorhersagen der Monte Carlo Simulationen mit POWHEG und PYTHIA weitgehend bestätigt. Merkbare Differenzen zwischen Vorhersage und Messwert für die Normierung des $t\bar{t}$ -Untergrund treten lediglich für W+Jets Kandidaten mit drei Jets auf. Hier stellen $t\bar{t}$ -Ereignisse mit lediglich vier bis fünf Prozent der Ereignisse in der Signalregion jedoch nur einen kleinen Untergrund dar, wohingegen Multijet-Ereignisse den größten Untergrund bilden. Dies rechtfertigt es, zukünftig in vorläufigen Analysen zur Untergrundabschätzung vereinfachend auf die Monte Carlo Simulation mit POWHEG und PYTHIA zurückzugreifen. Um die Ergebnisse zu validieren und zu verfeinern sollte allerdings die beschriebene datenbasierte Methode verwendet werden, da vereinzelt Unterschiede zwischen Daten und Simulation auftreten, wie etwa in Verteilungen bezüglich H_T oder 1-Jet p_T und beim Pile-up festgestellt wurde. Teils liefern auch unterschiedliche Monte Carlo Generatoren deutlich verschiedene Vorhersagen, wie etwa POWHEG und SHERPA für die transversalen Impulse von Jets. Für Ereignisse mit fünf beziehungsweise mit

mehr als sechs Jets sind die Unsicherheiten auf den Wirkungsquerschnitt aus der datenbasierten Methode in etwa gleich groß wie die Unsicherheiten aus der Vorhersage von POWHEG (siehe Abbildungen 4.10a und 4.10b). Daher lohnt es sich insbesondere hier, weiterhin datenbasierte $t\bar{t}$ -Untergrundabschätzungen in der Analyse von W+Jets Ereignissen einzusetzen.

Anhang A

Prozess	3 Jets	4 Jets	5Jets	≥ 6 Jets
W+Jets	$14,47 \pm 0,13$	$7,29 \pm 0,11$	$4,91 \pm 0,14$	$3,61 \pm 0,15$
$W \rightarrow \tau\nu$	$0,87 \pm 0,13$	$0,73 \pm 0,18$	$0,25 \pm 0,11$	$1,23 \pm 0,60$
$Z \rightarrow \tau^+\tau^-$	$0,18 \pm 0,02$	$0,12 \pm 0,01$	$0,10 \pm 0,02$	$0,12 \pm 0,03$
$Z \rightarrow e^+e^-$	$2,48 \pm 0,04$	$1,65 \pm 0,04$	$1,38 \pm 0,05$	$1,28 \pm 0,07$
Single top	$10,78 \pm 0,10$	$7,29 \pm 0,10$	$5,06 \pm 0,13$	$3,49 \pm 0,15$
Diboson	$0,35 \pm 0,01$	$0,14 \pm 0,01$	$0,07 \pm 0,01$	$0,01 \pm 0,00$
Multijet	$4,76 \pm 0,30$	$1,96 \pm 0,35$	$0,67 \pm 0,46$	$2,15 \pm 0,41$
Gesamt	$33,88 \pm 0,38$	$19,17 \pm 0,43$	$12,44 \pm 0,52$	$11,89 \pm 0,77$

Tabelle A.1: Kontamination der $t\bar{t}$ -Kontrollregion im Elektronenkanal bei einer b-Taggingeffizienz von 60% in Prozent.

(a) ein b-Jet				
Prozess	3 Jets	4 Jets	5 Jets	≥ 6 Jets
W+Jets	$14,62 \pm 0,10$	$7,72 \pm 0,09$	$5,22 \pm 0,11$	$4,07 \pm 0,14$
$W \rightarrow \tau\nu$	$0,04 \pm 0,01$	$0,02 \pm 0,01$	$0,02 \pm 0,01$	$0,00 \pm 0,00$
$Z \rightarrow \tau^+\tau^-$	$0,04 \pm 0,01$	$0,01 \pm 0,01$	$0,01 \pm 0,01$	$0,00 \pm 0,00$
Single top	$4,24 \pm 0,03$	$3,45 \pm 0,04$	$2,98 \pm 0,06$	$2,45 \pm 0,07$
Diboson	$0,15 \pm 0,00$	$0,06 \pm 0,03$	$0,02 \pm 0,00$	$0,01 \pm 0,00$
Multijet	$10,68 \pm 0,27$	$4,51 \pm 0,28$	$3,71 \pm 0,52$	$2,03 \pm 0,87$
Gesamt	$29,77 \pm 0,29$	$15,78 \pm 0,30$	$11,96 \pm 0,54$	$8,57 \pm 0,88$

(b) zwei b-Jets				
Prozess	3 Jets	4 Jets	5 Jets	≥ 6 Jets
W+Jets	$4,26 \pm 0,12$	$2,46 \pm 0,11$	$1,82 \pm 0,10$	$1,82 \pm 0,13$
$W \rightarrow \tau\nu$	$0,00 \pm 0,00$	$0,01 \pm 0,01$	$0,00 \pm 0,00$	$0,00 \pm 0,00$
$Z \rightarrow \tau^+\tau^-$	$0,00 \pm 0,00$	$0,00 \pm 0,00$	$0,00 \pm 0,00$	$0,00 \pm 0,00$
Single top	$2,72 \pm 0,05$	$2,37 \pm 0,06$	$2,08 \pm 0,08$	$2,03 \pm 0,10$
Diboson	$0,13 \pm 0,01$	$0,04 \pm 0,00$	$0,01 \pm 0,00$	$0,00 \pm 0,00$
Multijet	$3,07 \pm 0,42$	$1,71 \pm 0,39$	$0,75 \pm 0,46$	$0,00 \pm 0,87$
Gesamt	$10,17 \pm 0,44$	$6,59 \pm 0,41$	$4,66 \pm 0,48$	$3,85 \pm 0,89$

Tabelle A.2: Kontamination der $t\bar{t}$ -Kontrollregion im Myonenkanal bei einer b-Taggingeffizienz von 60% (a) für mindestens einen und (b) für mindestens zwei erforderliche b-Jets in Prozent.

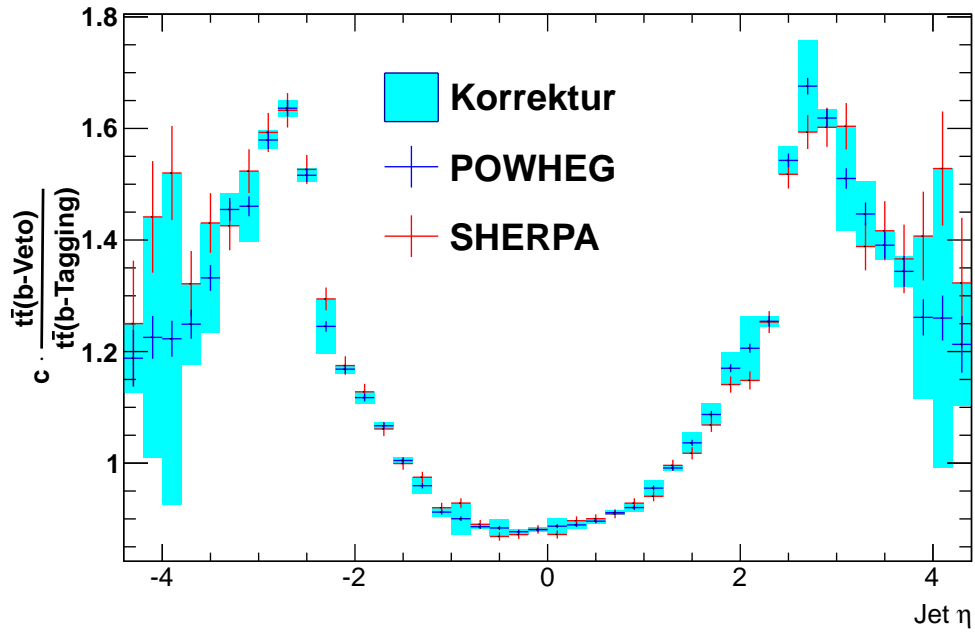


Abbildung A.1: Korrekturhistogramm für die Pseudorapidität von Jets für vier Jets im Myonenkanal. Im Bereich von $|\eta| < 2$ sind die Korrekturen gering und genau bekannt. Außerhalb dieses Bereichs nehmen sowohl die Korrekturen als auch die Unsicherheiten stark zu.

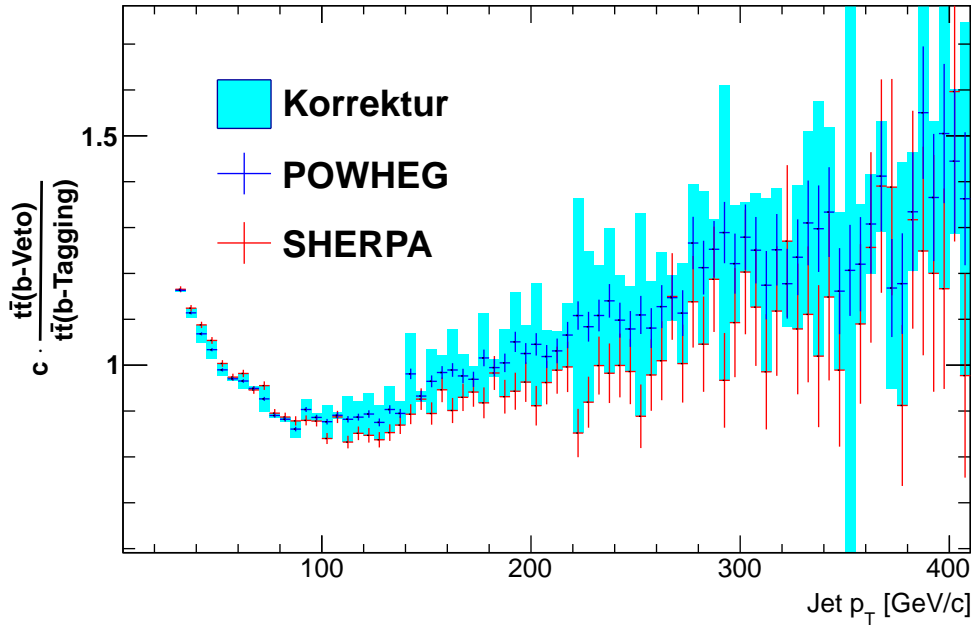


Abbildung A.2: Korrekturhistogramm für den transversalen Impuls der Jets für vier Jets im Myonenkanal. Hier treten deutliche Unterschiede zwischen den Vorhersagen von POWHEG und SHERPA auf.

#Jets	Signal	Multijet	$t\bar{t}$	Daten	Daten, best
3	1,14	1,17	1,02	1,32	1,19
4	1,26	0,99	1,30	1,29	1,23
5	1,28	0,91	1,06	1,20	1,17
≥ 6	0,98	1,26	0,92	1,02	0,89

Tabelle A.3: χ_{red}^2 -Werte für den Fit mit Funktionen im Elektronenkanal für den Fit an das $t\bar{t}$ -Template, an das Signal-Template, an das Multijet-Template und abschließend für den Fit der Summe der drei Fitfunktionen an die Daten in der Signalregion. "Daten, best" ist der niedrigste χ_{red}^2 -Wert des Fits an die Daten mit einer Kombination aus Fehlerfunktionen (siehe Kapitel 4.5).

1 und 2 b-Jets			1 b-Jet			2 b-Jets		
#Jets	Signal	Multijet	$t\bar{t}$	Daten	Daten, best	$t\bar{t}$	Daten	Daten, best
3	1,83	1,06	0,70	1,85	1,77	0,55	1,82	1,77
4	1,77	1,09	0,68	0,84	0,80	0,59	0,84	0,81
5	1,06	0,99	0,62	0,99	0,97	0,68	1,00	0,97
6	0,92	0,94	0,57	1,50	1,32	0,84	1,37	1,17

Tabelle A.4: χ_{red}^2 -Werte für den Fit mit Funktionen im Myonenkanal für den Fit an das $t\bar{t}$ -Template, an das Signal-Template, an das Multijet-Template und abschließend für den Fit der Summe der drei Fitfunktionen an die Daten in der Signalregion sowohl für einen als auch für zwei erforderliche b-Jets in der $t\bar{t}$ -Kontrollregion. "Daten, best" ist der niedrigste χ_{red}^2 -Wert des Fits an die Daten mit einer Kombination aus Fehlerfunktionen (siehe Kapitel 4.5).

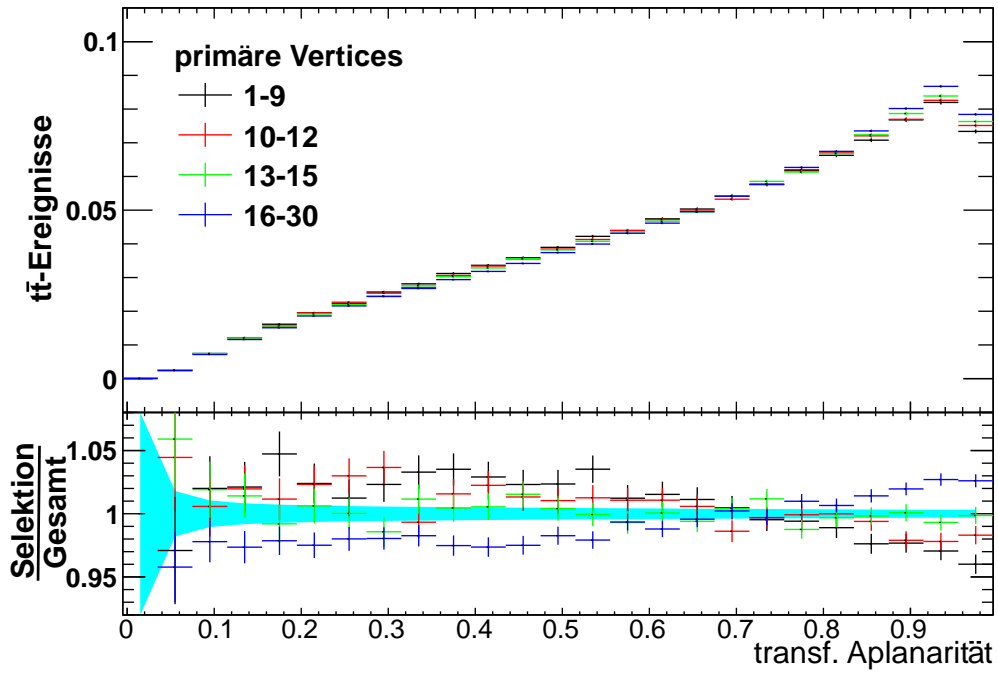


Abbildung A.3: Vorhersage von POWHEG für die Verteilung $t\bar{t}$ -Ereignissen in der $t\bar{t}$ -Kontrollregion. Oben: normiert im direkten Vergleich. Unten: normiert relativ zur Summe der vier Teilselektionen. Es sind deutliche Unterschiede zwischen Ereignissen mit 16-30 primären Vertices im Vergleich zu den anderen abgebildeten Teilselektionen. Diese Unterschiede treten im datenbasierten $t\bar{t}$ -Template nicht auf (siehe Abbildung 4.9b).

#Jets	Elektronenkanal		Myonenkanal	
	TFractionFitter	Funktionen	TFraktionFitter	Funktionen
3	16530 ± 10958	18825 ± 5886	28068 ± 14198	22966 ± 7637
4	15324 ± 5447	17731 ± 3863	20049 ± 5263	20326 ± 2854
5	8083 ± 1130	7601 ± 1181	8694 ± 1854	9270 ± 1500
≥ 6	3175 ± 1417	3335 ± 1017	3569 ± 3128	4181 ± 986

Tabelle A.5: Anzahl von $t\bar{t}$ -Ereignissen in der Signalregion bestimmt mit TFractionFitter und mit einem Fit mit Funktionen.

Selektion	3Jets	4Jets	5Jets	6Jets*
1.60	$19,65 \pm 0,11$	$11,59 \pm 0,10$	$8,57 \pm 0,12$	$6,90 \pm 0,18$
1.70	$23,41 \pm 0,10$	$13,76 \pm 0,10$	$9,94 \pm 0,12$	$8,33 \pm 0,19$
1.80	$33,96 \pm 0,11$	$20,71 \pm 0,10$	$14,70 \pm 0,13$	$12,08 \pm 0,21$
1.85	$42,67 \pm 0,24$	$27,01 \pm 0,11$	$19,19 \pm 0,14$	$15,15 \pm 0,2$
2.60	$7,27 \pm 0,13$	$5,04 \pm 0,11$	$4,12 \pm 0,13$	$4,13 \pm 0,18$
2.70	$8,16 \pm 0,12$	$5,56 \pm 0,10$	$4,60 \pm 0,12$	$4,27 \pm 0,16$
2.80	$10,22 \pm 0,11$	$6,82 \pm 0,10$	$5,50 \pm 0,12$	$5,05 \pm 0,16$
2.85	$12,27 \pm 0,11$	$7,88 \pm 0,10$	$6,27 \pm 0,11$	$5,52 \pm 0,15$

Tabelle A.6: Kontamination der $t\bar{t}$ -Kontrollregion im Myonenkanal ohne Multijet-Ereignisse für mindestens einen und für mindestens zwei erforderliche b-Jets für b-Tagging-Effizienzen von 60%, 70%, 80% und 85%. Die Angaben sind in Prozent. *Die Ergebnisse sind bei Ereignissen mit sechs Jets für die Selektionen mit einem Jet exklusive und für die Selektionen mit mindestens zwei erforderlichen b-Jets inklusive höherer Jetmultiplizitäten.

Selektion	3 Jets	4 Jets	5 Jets	≥ 6 Jets
1 b-Jet	$103\,099 \pm 503$	$74\,072 \pm 363$	$33\,117 \pm 264$	$16\,136 \pm 197$
2 b-Jets	$31\,991 \pm 232$	$26\,737 \pm 197$	$13\,316 \pm 131$	$6\,888 \pm 103$

Tabelle A.7: Anzahl der Ereignisse in der $t\bar{t}$ -Kontrollregion nach Abzug der Kontamination für ein und zwei erforderliche b-Jets im Myonenkanal. Das 2-b-Jet-Kriterium führt zu einer deutlichen Reduktion der Anzahl der Ereignisse in der $t\bar{t}$ -Kontrollregion.

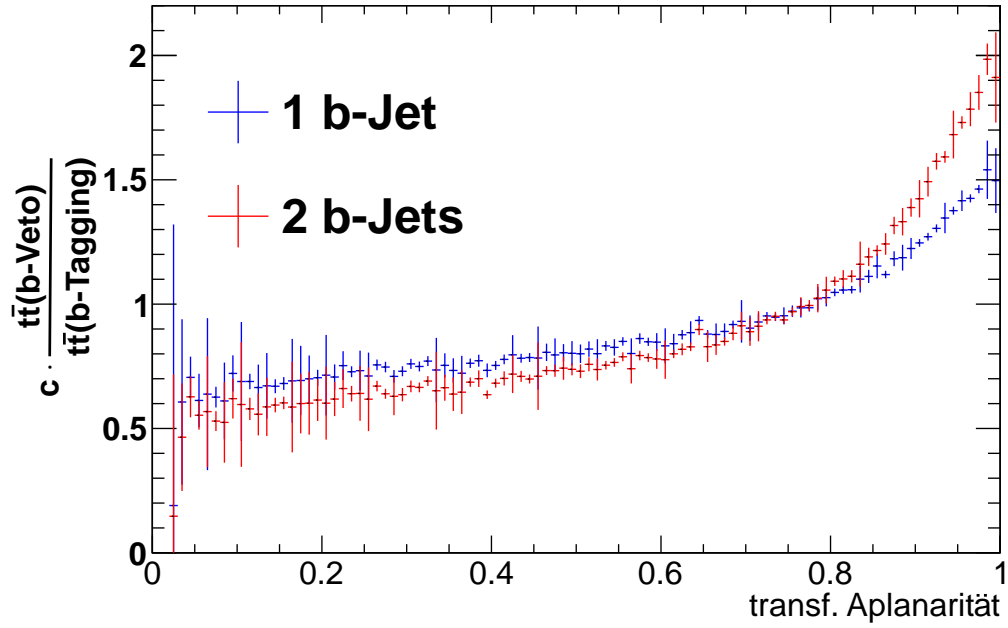


Abbildung A.4: Korrekturhistogramme für einen und für zwei erforderliche b-Jets im Myonenkanal für Ereignisse mit vier Jets und einer b-Taggingeffizienz von 60%.

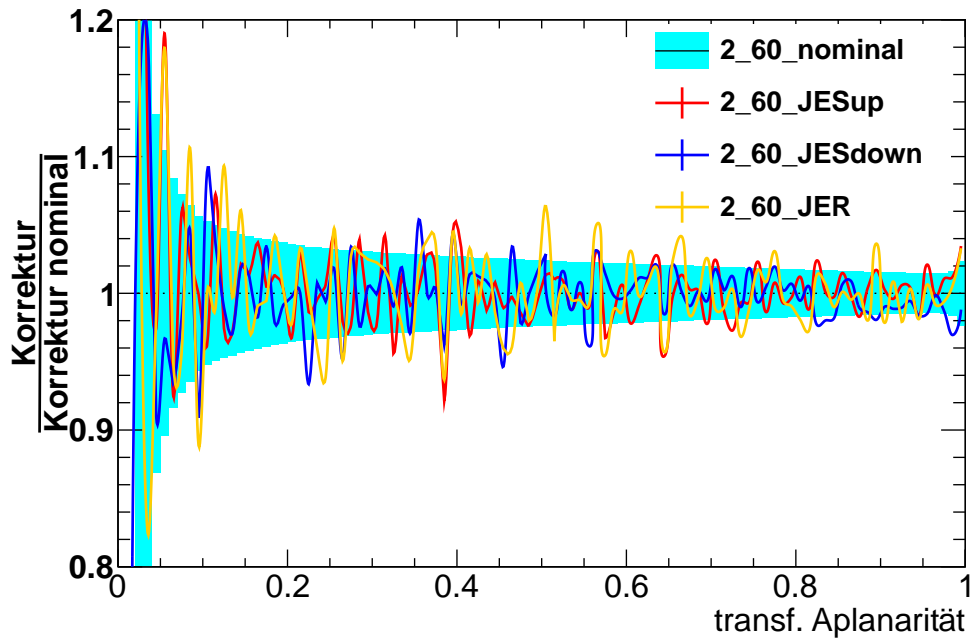


Abbildung A.5: Korrekturhistogramme für Ereignisse mit vier Jets im Myonenkanal bei mindestens zwei erforderlichen b-Jets in der $t\bar{t}$ -Kontrollregion für die systematischen Unsicherheiten durch die Jet Energie Skala relativ zur nominalen Korrektur. Die Fehlerbalken für die Korrekturhistogramme sind in etwa gleich groß wie das eingezeichnete Fehlerband für das nominale Korrekturhistogramm. Es sind lediglich statistische Fluktuationen zu erkennen.

Literatur

- [1] G. Aad u. a. „Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at $\sqrt{s}=7$ TeV“. In: *European Physical Journal C* 72, 2211 (Nov. 2012), S. 2211. DOI: 10.1140/epjc/s10052-012-2211-y. arXiv: 1206.2135 [hep-ex].
- [2] G. Aad u. a. „Measurements of the production cross sections in association with jets with the ATLAS detector“. In: *European Physical Journal C* 75 (Feb. 2015), S. 82. DOI: 10.1140/epjc/s10052-015-3262-7. arXiv: 1409.8639 [hep-ex].
- [3] Roger Barlow und Christine Beeston. „Fitting using finite Monte Carlo samples“. In: *Computer Physics Communications* 77.2 (1993), S. 219–228.
- [4] Bronstein u. a. *Taschenbuch der Mathematik, 7., überarbeitete und erweiterte Auflage*. 7. Aufl. Verlag Hari Deutsch, Frankfurt am Main, Thun, 2008. ISBN: 978-3-8171-2007-9.
- [5] R Brun u. a. *ROOT-An Object-Oriented Data Analysis Framework*. Online; Stand: 29. September 2015. Juli 2004. URL: <https://root.cern.ch/>.
- [6] J. M. Butterworth, J. R. Forshaw und M. H. Seymour. „Multiparton interactions in photoproduction at HERA“. In: *Zeitschrift für Physik C Particles and Fields* 72 (Dez. 1996), S. 637–646. DOI: 10.1007/BF02909195. eprint: hep-ph/9601371.
- [7] Matteo Cacciari, Gavin P. Salam und Gregory Soyez. „The anti- k_t jet clustering algorithm“. In: *Journal of High Energy Physics* 2008.04 (2008), S. 063. URL: <http://stacks.iop.org/1126-6708/2008/i=04/a=063>.
- [8] ATLAS Collaboration. „The ATLAS Experiment at the CERN Large Hadron Collider“. In: *Journal of Instrumentation* 3.08 (2008), S08003. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08003>.
- [9] The ATLAS Collaboration. „b-tagging in dense environments“. 19. Aug. 2014. URL: <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-014/ATL-PHYS-PUB-2014-014.pdf>.
- [10] Gennaro Corcella u. a. „HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)“. In: *Journal of High Energy Physics* 2001.01 (2001), S. 010. URL: <http://stacks.iop.org/1126-6708/2001/i=01/a=010>.
- [11] *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*. Techn. Ber. ATLAS-CONF-2014-032. Geneva: CERN, Juni 2014. URL: <http://cds.cern.ch/record/1706245>.

- [12] Stefano Frixione, Paolo Nason und Carlo Oleari. „Matching NLO QCD computations with parton shower simulations: the POWHEG method“. In: *Journal of High Energy Physics* 2007.11 (2007), S. 070. URL: <http://stacks.iop.org/1126-6708/2007/i=11/a=070>.
- [13] Angela Barbaro Galtieri, Fabrizio Margaroli und Igor Volobouev. „Precision measurements of the top quark mass from the Tevatron in the pre-LHC era“. In: *Reports on Progress in Physics* 75.5 (2012), S. 056201. URL: <http://stacks.iop.org/0034-4885/75/i=5/a=056201>.
- [14] T. Gleisberg u. a. „Event generation with SHERPA 1.1“. In: *Journal of High Energy Physics* 2009.02 (2009), S. 007. URL: <http://stacks.iop.org/1126-6708/2009/i=02/a=007>.
- [15] Particle Data Group. *t-quark*. 21. Aug. 2014. URL: <http://pdg.lbl.gov/2014/listings/rpp2014-list-t-quark.pdf>.
- [16] B.P. Kersevan und E. Richter-Was. „The Monte Carlo Event Generator AcerMC version 1.0 with interfaces to {PYTHIA} 6.2 and {HERWIG} 6.3“. In: *Computer Physics Communications* 149.3 (2003), S. 142–194. ISSN: 0010-4655. DOI: [http://dx.doi.org/10.1016/S0010-4655\(02\)00592-1](http://dx.doi.org/10.1016/S0010-4655(02)00592-1). URL: <http://www.sciencedirect.com/science/article/pii/S0010465502005921>.
- [17] Michelangelo L. Mangano u. a. „ALPGEN, a generator for hard multiparton processes in hadronic collisions“. In: *Journal of High Energy Physics* 2003.07 (2003), S. 001. URL: <http://stacks.iop.org/1126-6708/2003/i=07/a=001>.
- [18] E. Moyse u. a. *REPORT OF THE MUON RECONSTRUCTION SELECTION PANEL*. Techn. Ber. ATL-COM-MUON-2011-040. Geneva: CERN, Dez. 2011. URL: <https://cds.cern.ch/record/1407916>.
- [19] Joao Pequeno. „Computer generated image of the whole ATLAS detector“. März 2008. URL: <http://cds.cern.ch/record/1095924>.
- [20] Prof. Dr. Günter Quast. *Funktionsanpassung mit der χ^2 -Methode*. 26. Feb. 2015. URL: <http://www-ekp.physik.uni-karlsruhe.de/~quast/Skripte/Chi2Method.pdf>.
- [21] Torbjörn Sjöstrand, Stephen Mrenna und Peter Skands. „PYTHIA 6.4 physics and manual“. In: *Journal of High Energy Physics* 2006.05 (2006), S. 026. URL: <http://stacks.iop.org/1126-6708/2006/i=05/a=026>.
- [22] Torbjörn Sjöstrand, Stephen Mrenna und Peter Skands. „PYTHIA 6.4 physics and manual“. In: *Journal of High Energy Physics* 2006.05 (2006), S. 026. URL: <http://stacks.iop.org/1126-6708/2006/i=05/a=026>.
- [23] Mark Thomson. *Modern Particle Physics*. 2. Aufl. Cambridge, United Kingdom: Cambridge University Press, 2013. ISBN: 978-1-107-03426-6.
- [24] Wikipedia. *Standardmodell* — *Wikipedia, Die freie Enzyklopädie*. [Online; Stand 2. September 2015]. 2015. URL: <https://de.wikipedia.org/w/index.php?title=Standardmodell&oldid=139557287>.

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 10. Oktober 2015,