

Lehrstuhlversuch im SS2020

## Search for $t\bar{t}$ resonances with ATLAS data

Fabian Koch

fabian3.koch@tu-dortmund.de

Nils Breer

nils.breer@tu-dortmund.de

Nicole Schulte

nicole.schulte@tu-dortmund.de

Abgabe: xx.xx.2020

TU Dortmund – Fakultät Physik

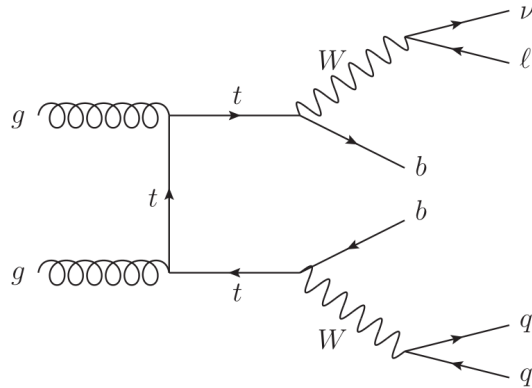
## Inhaltsverzeichnis

<b>1</b>	<b>Theoretische Grundlagen</b>	<b>3</b>
<b>2</b>	<b>Analysestrategie</b>	<b>4</b>
<b>3</b>	<b>Auswertung</b>	<b>5</b>
3.1	Vorselektion . . . . .	5
3.2	Erweiterte Eventselektion . . . . .	6
3.3	Fundamentale Verteilungen . . . . .	6
<b>4</b>	<b>Übereinstimmung Daten und Simulation</b>	<b>9</b>
<b>5</b>	<b>Statistische Analyse der Ergebnisse</b>	<b>13</b>
<b>6</b>	<b>Diskussion</b>	<b>16</b>
<b>7</b>	<b>Anhang</b>	<b>16</b>
7.1	Weitere Plots zu den fundamentalen Verteilungen . . . . .	16
7.2	Weitere Plots zur Überprüfung der Monte Carlo Übereinstimmung mit den Daten . . . . .	22
	<b>Literatur</b>	<b>25</b>

# 1 Theoretische Grundlagen

In dem Lehrstuhlversuch *Search for  $t\bar{t}$  resonance with ATLAS detector* wird ein Datensatz, welcher bei einer Schwerpunktsenergie von  $\sqrt{s} = 8 \text{ TeV}$ , entsprechend einer Luminosität von  $\mathcal{L} = 1 \text{ fb}^{-1}$ , der 2012 am ATLAS Detektor aufgenommen wurde, auf  $Z'$ -Resonanzen untersucht. Suchen nach diesem neuen massiven Teilchen beginnen oft bei Massenskalen von 500 GeV und aufwärts. Die Skala für neue Physik wird in dem meisten Fällen um die 1 TeV gesetzt. Aktuelle Limits auf die  $Z'$  Resonanz schließen Massen kleiner als 3,80 TeV ?? aus.

In dieser Analyse wird der mögliche Zerfall des  $Z'$  in ein Top-Quark und ein Anti-Top-Quark untersucht. Das Top-Quark ist das schwerste bekannte Quark und somit sensitiv auf neue Physik. Am LHC wird es hauptsächlich durch Gluonfusion produziert, wohingegen an Beschleunigern mit geringerer Schwerpunktsenergie wie das Tevatron, hauptsächlich Quark-Antiquark Annihilation für die Produktion verantwortlich ist. Top-Quarks zerfallen fast ausschließlich in ein Bottom-Quark zusammen mit einem  $W$ -Boson. Letztere können bei der Top-Quark Paarproduktion entweder semileptonisch, leptonisch oder hadronisch zerfallen. Der leptonische Zerfall beschreibt den Endzustand mit einem geladenen Lepton und dem zugehörigen Neutrino für beide Eichbosonen. Der hadronische Zerfall beschreibt den Zerfall beider  $W$ -Bosonen in jeweils zwei Quarks. Der semileptonische Zerfall beschreibt dann ein hadronisch zerfallendes und ein leptonisch zerfallendes  $W$ -Boson. Untersuchungen des leptonischen Zerfalls haben den Nachteil, dass durch die Neutrinos ein hoher Anteil fehlender Energie in der Analyse untersucht werden muss, wohingegen die Analyse des hadronischen Zerfalls den Nachteil vieler Jets hervorruft. In diesem Versuch wird demnach der semileptonische Zerfall untersucht. Dieser wird auch *lepton + jets* genannt, da sowohl ein geladenes Lepton und fehlende Energie verlangt wird, als auch mindestens 4 Jets, die von dem hadronischen  $W$ -Zerfall und von den Bottom-Quarks aus dem Top-Quark-Zerfall stammen. Dieser semileptonische Fall ist in Abbildung 1 für die  $t\bar{t}$  Produktion durch die Gluonfusion als Feynman Diagramm dargestellt.



**Abbildung 1:** Feynman Diagramm der semileptonischen Zerfallsmode der  $t\bar{t}$  Produktion durch Gluonfusion.

Die Signaturen der untersuchten Objekte im ATLAS Detektor sind wie folgt. Das Muon interagiert im Detektor zunächst als *Minimal Ionizing Particle*, einem sogenannten MIP. Es hinterlässt somit weder im Spurdetektor noch in den Kalorimetern eine Signatur. Lediglich in den Muonkammern deponiert es Energie. Elektronen werden in den Trackingdetektoren nach ihrer Ladung gekrümmt und deponieren anschließend im elektromagnetischen Kalorimeter ihre Energie. Die Quarks aus dem hadronischen Zerfall hadronisieren und schauen hauptsächlich im hadronischen Kalorimeter. Neutrinos sind nur über fehlende Energie der bereits rekonstruierten Endzustandsteilchen bestimmbar.

## 2 Analysestrategie

Um das Verhältnis von Signal zu Untergrund zu verbessern, muss zunächst eine Eventselektion auf den Datensatz angewendet werden, da dieser eine enorme Datenmenge besitzt. Dafür werden Analysemethoden in C++ verwendet. Die selektierten Events werden dann auf verschiedene Variablen wie die invariante Masse des Systems geprüft um eine *Finale Diskriminate* zu definieren, die zur optimalen Differenzierung zwischen Signal und Untergrund dienen soll. Der Untergrund sollte dabei ein fallendes Spektrum aufweisen, auf dem das Signal optimalerweise ein Peak aufweist. Für die Bestimmung des Untergrundspektrums werden Monte Carlo (MC) Methoden verwendet. Die simulierten Untergründe und die Benennung in der Analyse lauten wie folgt:

- **diboson**: Paarproduktion der  $W$ -/  $Z$ -Eichbosonen; Hierbei ist auch die Kombination  $WZ$  möglich
- **singletop**: Produktion eines einzelnen Top-Quarks
- **wjets**: Produktion eines  $W$ -Bosons im Zusammenhang mit Jets
- **zjets**: Produktion eines  $Z$ -Bosons im Zusammenhang mit Jets
- **ttbar**: Top-Quark Paarproduktion

Weiterhin stehen verschiedene **zprime** MC-samples zur Verfügung, welche für verschiedene  $Z'$  Massen von 400 GeV bis 3000 GeV generiert sind.

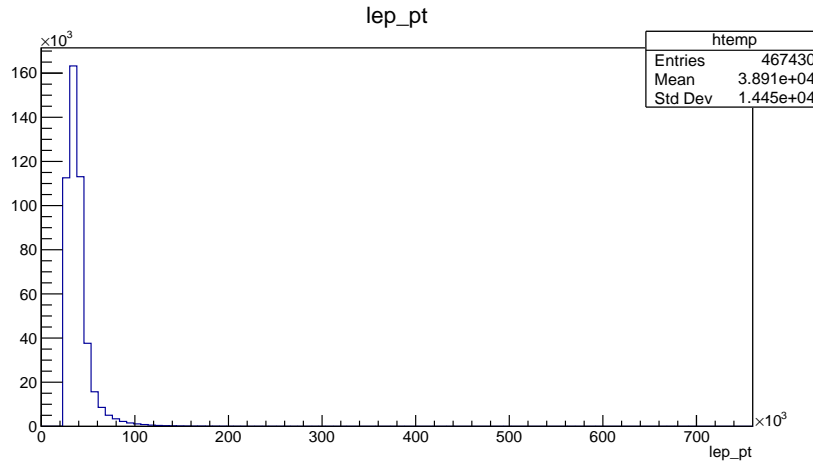
Die Datensamples sind in **ntupeln** im ROOT Format abgespeichert. Diese enthalten **TTrees** in denen verschiedene Informationen, wie beispielsweise die Pseudorapidität der Leptonen, über die rekonstruierten Objekte abgespeichert sind. Die Daten dieser **.root** samples sind bereits vorselektiert worden.

Im Anschluss an die Eventselektion erfolgt eine Studie, um die Übereinstimmung der Monte Carlo samples mit den Daten zu überprüfen. Dies ist ein wichtiger Schritt um die Qualität der simulierten Daten zu testen. Dann wird eine statistische Analyse vorgenommen, bei der die finale Diskriminate verwendet wird um den Überschuss des Datenpeaks über den kontinuierlichen Untergrund abzuschätzen und, wenn möglich, ein Limit auf die  $Z'$  Masse mit Hilfe eines Hypothesentests zu setzen.

## 3 Auswertung

### 3.1 Vorselektion

Um die enorme Datenmenge zu reduzieren wurde eine Vorselektion auf die Daten angewandt. Diese reduziert die ursprünglich  $66 \cdot 10^6$  Events auf  $6.6 \cdot 10^6$  Events. Um diese Reduktion zu erreichen, werden einige Anforderungen an jedes Event gestellt. Zu Bemerkem ist, dass in manchen Größen mehr Einträge zu finden sind, als Events im Tupeln vorhanden sind. Dies hat den Grund, dass Events mit mehreren Jets oder Leptonen für jeden Jet und jedes Lepton einen separaten Eintrag enthalten. Dies taucht somit auch nur für Größen auf, die die Eigenschaften der Jets oder Leptonen zeigen. Zunächst wird gefordert, dass der entsprechende Myon- oder Elektrontrigger ausgelöst wird. Damit sollen Beiträge von Untergrundleptonen verringert werden. Niederenergetische Leptonen lassen sich nämlich sehr schlecht vom Untergrund unterscheiden, weswegen eine typische Triggerschwelle für den transversalen Impuls bei 25 GeV liegt. Diese Mindestenergie lässt sich in den Daten direkt erkennen, wie in Abbildung 2 für das Datentupel *data.06.el* dargestellt. Da die Tupel nach den jeweiligen Leptonen sortiert sind, wird auch nur der entsprechende Trigger gefordert. Die Anzahl der rekonstruierten Leptonen sollte mindestens eins enthalten. Die Pseudorapidität hängt mit dem transversalen Impuls zusammen. Je geringer der transversale Impuls, desto größer die Pseudorapidität. Die Pseudorapiditätsverteilung spiegelt allerdings auch einige Detektoreigenschaften wieder. So werden in dem Gebiet um  $\eta \approx \pm 1.5$  weniger Events registriert. Dies hat den Grund, dass dort Leitungen oder ähnliche nicht detektierenden Elemente verbaut sind. Die geometrische Akzeptanz des Detektors beträgt also nicht 100 %. Das Maximum der Pseudorapidität liegt bei  $\eta \approx \pm 3$ . Der Azimuthalwinkel  $\Phi$  zeigt keine Anforderungen.



**Abbildung 2:** Histogramm des transversalen Impulses des Elektrons des Datentupels *data.06.el*.

### 3.2 Erweiterte Eventselektion

Selektiert wird auf den **lepton+jets** Kanal. Das bedeutet, es wird genau ein geladenes Lepton aus dem leptonischen Zerfall eines  $W$ -Bosons gefordert. Dieses Lepton, soll einen Transversalimpuls von mindestens  $\text{lep\_pt} > 50 \text{ GeV}$  besitzen. Leptonen mit geringerem Transversalimpuls werden verworfen. In dem leptonischen Verfall, wird zudem ein Neutrino produziert. Daher wird eine fehlende Transversalenergie von  $\text{met\_et} > 40 \text{ GeV}$  gefordert. Da beide Top-Quarks in ein  $W$ -Boson mit einem Bottom-Quark zerfallen, entstehen dadurch bereits zwei Quarks. Aus dem hadronischen Zerfall des zweiten  $W$ -Bosons entstehen zwei weitere Quarks, weswegen in der Selektion Events mit  $\text{jet\_n} < 4$  verworfen werden. Von diesen Jets, werden mindestens zwei Jets verlangt, die b-tagged sein müssen. Das wird dadurch erreicht, indem ein  $MV1$  Wert größer als 0.7892 gefordert wird. Weiterhin soll jeder Jet mindestens einen Transversalimpuls von  $\text{jet\_pt} > 100 \text{ GeV}$  besitzen.

Die Pseudorapidität ist sowohl eine Detektorkomponente als auch eine Teilcheneigenschaft, denn Teilchenmasse und -energie beeinflussen den Abstrahlungswinkel. Der ATLAS Detektor kann innerhalb eines Akzeptanzbereichs von  $|\eta| < 2.4$  messen. Alles was diesen Bereich übersteigt, kann nicht von den Detektorkomponenten erfasst werden. Deswegen wird auf diese Eigenschaft ebenfalls selektiert. Schnitte auf Teilcheneigenschaften wie Transversalimpuls sind notwendig um Untergründe zu reduzieren. Da in diesem Versuch ein Teilchen mit hoher invarianter Masse untersucht wird, ist es sinnvoll hohe Grenzen auf den Transversalimpuls zu setzen.

Die erwarteten Untergründe für diesen Zerfallskanal sind bereits in Kapitel 2 gelistet. Zur Berechnung des Koeffizienten  $\epsilon * A$ , welches für Effizienz mal die Detektorakzeptanz steht, wird das Verhältnis der Anzahl der Events, die einen Cut überstehen, mit der Gesamtanzahl der Events nach der kompletten Selektion gebildet. Dadurch ergeben sich für die Samples die in Tabelle 1 abgebildeten Effizienzen.

Nach der Eventselektion ist Top-Quark Paarproduktion der größte Untergrund, da die Signatur dieses Untergrundes der selektierten Signatur ähnelt.

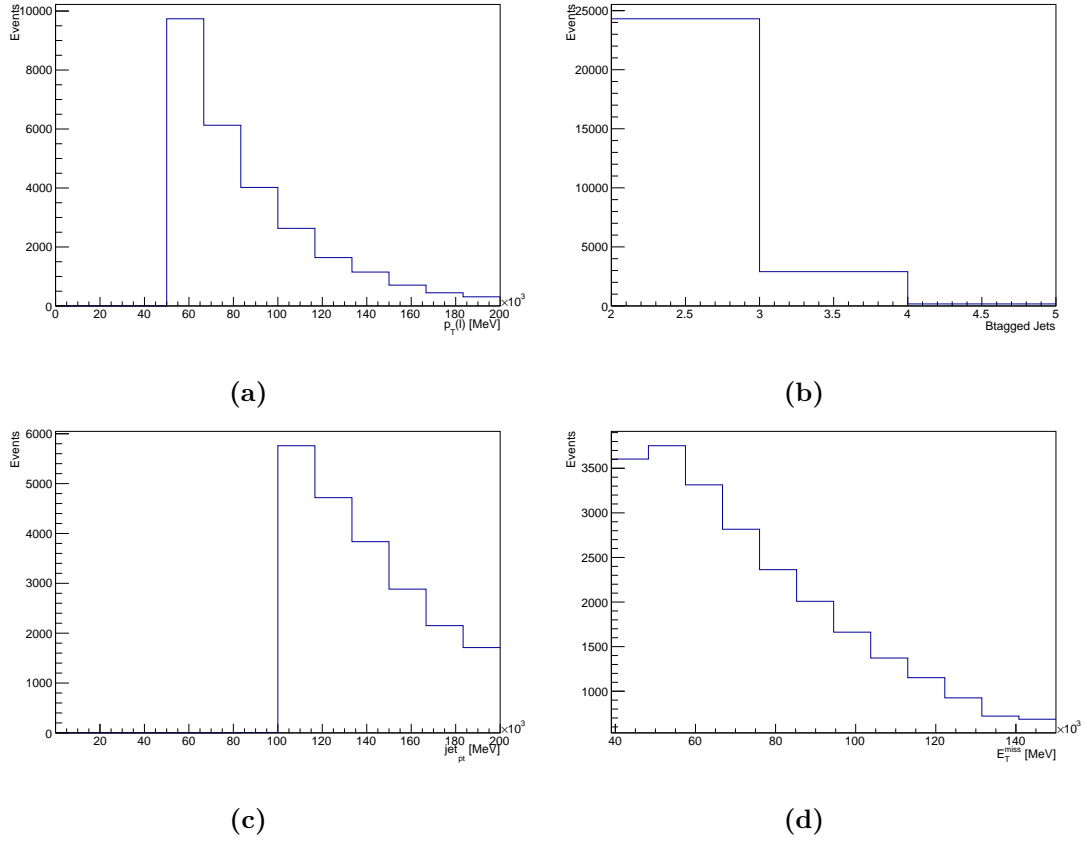
### 3.3 Fundamentale Verteilungen

Die große Untergrundquelle sind nach der vollständigen Selektion die  $t\bar{t}$ -Prozesse. Um ihre Eigenschaften besser verstehen zu können, wurden einige Größen hier aufgetragen. Diese werden dann mit den Erwartungen für  $t\bar{t}$ -Prozesse verglichen. Dabei handelt es sich um Monte-Carlo Simulationen des  $t\bar{t}$ -Prozesses mit Myonen. Einige dieser Größen sind in Abbildung 3 dargestellt. Der Rest ist im Anhang in Kapitel 7.1 in Abbildung 8 - 10 zu finden. Sie stimmen mit den Erwartungen an den  $t\bar{t}$ -Prozess gut überein. So ist zum Beispiel in Abbildung 3b die Anzahl an btagged Jets zu sehen. Btagged Jets sind Jets die mit einer gewissen Wahrscheinlichkeit ( $MV1 > 0.7892$ ) ihren Ursprung in einem b-Quark nahmen. Da in einem  $t\bar{t}$ -Prozess, wie in Abbildung 1 zu sehen, zwei b-Jets erzeugt werden, ist die Verteilung in der Hinsicht wie zu erwarten. Der restliche transversale Impuls wird dann gleichermaßen auf das Myon und das Neutrino aufgeteilt. Diese beiden Verteilungen in Abbildung 3a und 3d weisen einen ähnlichen Peak und Verlauf auf. Allerdings ist die Verteilung durch die fehlende Auflösung für die Neutrinos etwas ausgeschmierter, bzw.

**Tabelle 1:** Berechte Effizienzen für die einzelnen Cuts der Eventselektion

Sample	lep_n! = 1	lep_pt > 50 GeV	jet_n > 4	jet_pt > 100 GeV	btagged > 2	met_et > 50 GeV	jet_eta <  2.4
ttbar.el	0.9316	0.5071	0.2442	0.1406	0.0605	0.0428	0.0400
ttbar.mu	0.9200	0.4821	0.2328	0.1340	0.0579	0.0409	0.0382
singletop.el	0.9806	0.4323	0.0532	0.0311	0.0103	0.0071	0.0064
singletop.mu	0.9788	0.4104	0.0498	0.0288	0.0097	0.0068	0.0062
diboson.el	0.9044	0.3813	0.0119	0.0048	0.0001	7.97e-05	7.9691e-05
diboson.mu	0.8731	0.3510	0.0091	0.0038	9.99691e-05	7.38902e-05	6.51973e-05
wjets.el	0.9999	0.1709	0.0029	0.0017	1.99407e-05	1.39121e-05	1.28301e-05
wjets.mu	0.9999	0.1589	0.0027	0.0016	2.33554e-05	1.56505e-05	1.49282e-05
zjets.el	0.7167	0.1814	0.0067	0.0032	0.000135906	4.8692e-05	4.59184e-05
zjets.mu	0.5015	0.1256	0.0025	0.0014	7.37231e-05	3.18146e-05	3.00259e-05
zprime400.el	0.9308	0.4464	0.1640	0.0470	0.0192	0.0124	0.0119
zprime400.mu	0.9201	0.4215	0.1597	0.0539	0.0203	0.0137	0.0131
zprime500.el	0.9324	0.5616	0.2539	0.1499	0.0625	0.0417	0.0399
zprime500.mu	0.9224	0.5353	0.2319	0.1357	0.0581	0.0389	0.0378
zprime750.el	0.9237	0.6705	0.3881	0.3503	0.1573	0.1216	0.1155
zprime750.mu	0.9133	0.6451	0.3710	0.3332	0.1526	0.1220	0.1173
zprime1000.el	0.9339	0.7342	0.4599	0.4399	0.2087	0.1789	0.1728
zprime1000.mu	0.9182	0.7055	0.4573	0.4330	0.2021	0.1681	0.1596
zprime1250.el	0.9355	0.7475	0.4811	0.4665	0.2157	0.1898	0.1820
zprime1250.mu	0.9246	0.7319	0.4867	0.4702	0.2217	0.1931	0.1853
zprime1500.el	0.9401	0.7711	0.5014	0.4905	0.2121	0.1905	0.1819
zprime1500.mu	0.9287	0.7483	0.5012	0.4902	0.2125	0.1893	0.1805
zprime1750.el	0.9473	0.7758	0.5140	0.5036	0.2146	0.1932	0.1840
zprime1750.mu	0.9383	0.7571	0.5036	0.4914	0.2091	0.1898	0.1819
zprime2000.el	0.9451	0.7629	0.4988	0.4873	0.1913	0.1747	0.1683
zprime2000.mu	0.9406	0.7611	0.5171	0.5066	0.1968	0.1799	0.1717
zprime2250.el	0.9450	0.7451	0.4948	0.4839	0.1846	0.1688	0.1588
zprime2250.mu	0.9448	0.7639	0.5183	0.5085	0.2012	0.1824	0.1738
zprime2500.el	0.9480	0.7509	0.4958	0.4809	0.1759	0.1598	0.1523
zprime2500.mu	0.9433	0.7595	0.5211	0.5078	0.1933	0.1765	0.1669
zprime3000.el	0.9437	0.7151	0.4510	0.4253	0.1648	0.1458	0.1368
zprime3000.mu	0.9398	0.7314	0.4833	0.4631	0.1781	0.1607	0.1516

flacher.

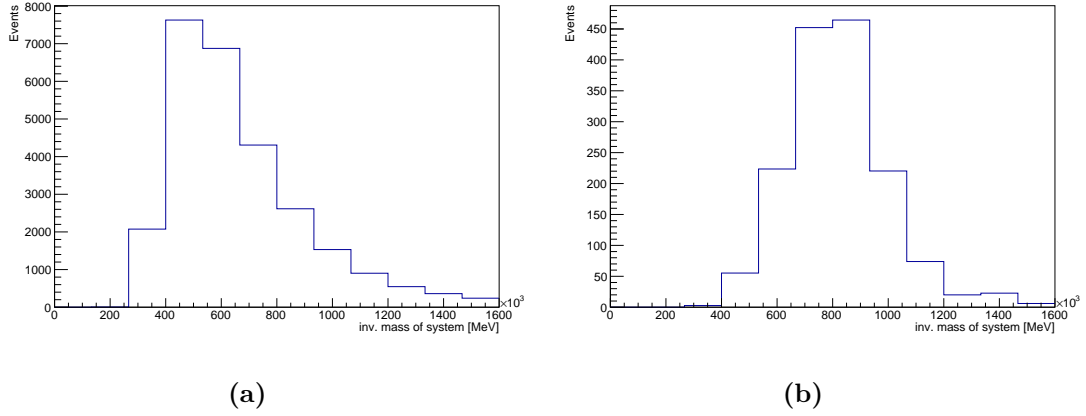


**Abbildung 3:** Darstellung verschiedener Verteilungen der Größen der  $t\bar{t}$  Monte-Carlo Simulation. Zusehen sind der transversale Impuls der Myonen (a), die Anzahl an btagged Jets (b), der transversale Impuls des Jets mit dem höchsten transversalen Impuls des Events (c) und die fehlende transversale Energie (d).

Auch nach der oben durchgeführten Selektion bleiben Untergrundprozesse wie der hier untersuchte  $t\bar{t}$ -Prozess übrig. Um nun eine bessere Unterscheidung zwischen Signalereignissen, also  $Z'$ -Ereignissen, und den Untergründen herzustellen werden weitere Größen konstruiert. Da diese unter Umständen mehr Informationen als einfache Größen enthalten, kann ihre Diskriminierungsstärke deutlich erhöht sein. Untersucht wurden  $\Delta\Phi$ , die Differenz zwischen dem Azimutalwinkel der fehlenden Transversalenergie und der Leptonflugrichtung, die invariante Masse des Systems, die durch die drei Jets mit dem größten  $p_T$  gebildet wird, die invariante Masse und die Pseudorapidität des Systems, die durch die vier Jets mit dem größten  $p_T$ , dem Lepton und dem Neutrino gebildet werden. Für  $\Delta\Phi$  wird darauf geachtet, dass jeweils die kleinere Differenz von  $|\phi_1 - \phi_2|$  und  $|2\pi - |\phi_1 - \phi_2||$  genutzt wird. Diese werden für alle Datentupel und Monte-Carlo Simulationen berechnet. Um die Größe mit der stärksten Diskriminierungsstärke zu



ermitteln, werden die Verteilungen für den größten Untergrundprozess ( $t\bar{t}$ ) mit einer möglichen Signalverteilung ( $Z'(1000)$ ), also einem hypothetischen  $Z'$  mit einer Masse von 1000 GeV) verglichen. Dies ist für die nachher gewählte Diskriminante, der vollständigen invarianten Masse des Systems, hier dargestellt. Die restlichen Vergleiche sind im Anhang in Kapitel 7.1 in Abbildung 11 - 13 zu sehen.



**Abbildung 4:** Vergleich der Verteilung der invarianten Masse des Systems, welches aus den vier Jets mit dem größten  $p_T$ , dem Lepton und dem Neutrino gebildet wird. Dies ist für die  $t\bar{t}$ -Untergrundsimulation a und der Signalsimulation des  $Z'(1000)$  b dargestellt. Diese Größe dient weiter als Diskriminante.

## 4 Übereinstimmung Daten und Simulation

Grundlage für Messungen neuer Physik ist immer eine sinnvolle Übereinstimmung der Daten mit der Simulation, vor allem in einem Phasenraum, in dem der Untergrund dominiert. Dafür werden zunächst die erwartete Anzahl der Events berechnet, welche zu einer integrierten Luminosität von  $\mathcal{L} = 1 \text{ fb}^{-1}$  korrespondieren. Mit Hilfe der Formel

$$N = \mathcal{L} \sigma A \epsilon \quad (1)$$

kann diese Anzahl berechnet werden. Dabei entspricht der Koeffizient  $\epsilon \cdot A$  den in Kapitel 3.2 berechneten Werten. Der Wirkungsquerschnitt der Prozesse ist zusammen mit den erwarteten Events in Tabelle 2 aufgelistet. Die erwarteten Events sind für das ausgewählte Sample `data.mu.2.root` aufgelistet. Insgesamt werden insgesamt 66.8647 Untergrundevents erwartet. Gemessene Events nach der Selektion in diesem Datenset sind 94 Events.

**Tabelle 2:** Anzahl erwarteter Ereignisse für die einzelnen Prozesse im sample `data.mu.2.root`. Angegeben ist der zugehörige Wirkungsquerschnitt der für die Berechnung der einzelnen Werte nach Formel (1) benötigt wird.

Prozess	$N_{\text{expected}}$	$\sigma$ / pb
ttbar	3.0282	252.82
singletop	3.3576	52.47
diboson	2.9967	29.41
wjets	6.3202	2516.20
zjets	51.1618	36214
zprime400	103.4000	1.1e2
zprime500	77.0800	8.2e1
zprime750	18.8000	2.0e1
zprime1000	51.7000	5.5
zprime1250	17.8600	1.9
zprime1500	7.8020	8.3e-1
zprime1750	2.8200	3.0e-1
zprime2000	1.3160	1.4e-1
zprime2250	0.0630	6.7e-2
zprime2500	0.0330	3.5e-2
zprime3000	0.0113	1.2e-2

Die selektierten Events beziehen sich noch auf das gegebene Sample, mit individuellen Wirkungsquerschnitten und Sample-Größen, je nach Prozess. Deswegen müssen noch Gewichte eingeführt werden, die die MC-Samples auf das Datensample normieren. Diese Gewichte werden mit Hilfe der Formel

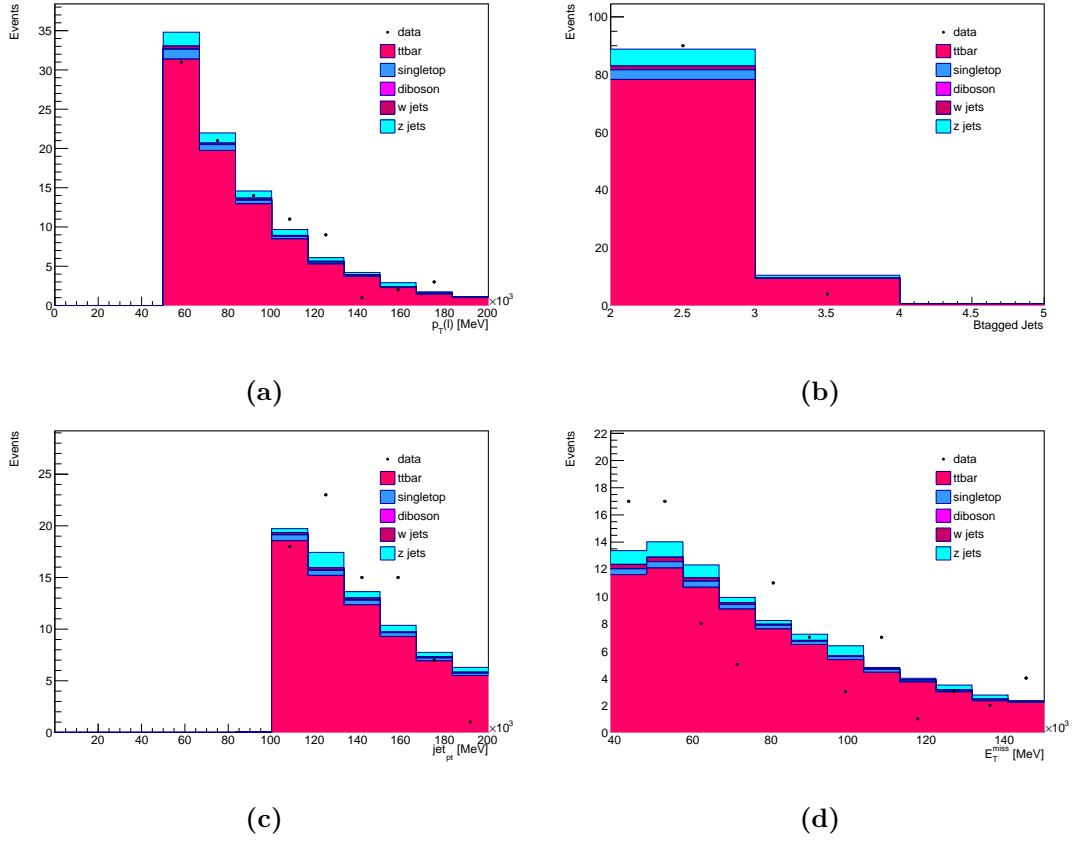
$$w = \frac{\mathcal{L}\sigma}{N_{\text{MC}}} \quad (2)$$

berechnet. Dabei beschreibt  $N_{\text{MC}}$  die Anzahl der MC events vor der Vorselektion. Die Gewichte mit den Samplegrößen sind in Tabelle 3 aufgelistet. Der Wirkungsquerschnitt ist in Tabelle 2 zu finden.

**Tabelle 3:** Berechnete Gewichte für die einzelnen Prozesse. Der für die Berechnung notwendige Wirkungsquerschnitt ist in Tabelle 2 aufgelistet. Die integrierte Luminosität beträgt  $\mathcal{L} = 1 \text{ fb}^{-1}$ . Die Gewichte werden mit Hilfe von Formel (2) berechnet.

Prozess	$N_{\text{MC}}$	Gewichtung $w$
ttbar	7847944	0.03221
singletop	1468942	0.03572
diboson	922521	0.03188
wjets	66536222	0.54428
zjets	37422926	0.06724
zprime400	100000	1.10000
zprime500	100000	0.82000
zprime750	100000	0.20000
zprime1000	100000	0.55000
zprime1250	100000	0.19000
zprime1500	100000	0.08300
zprime1750	100000	0.03000
zprime2000	100000	0.01400
zprime2250	100000	0.00067
zprime2500	100000	0.00035
zprime3000	100000	0.00012

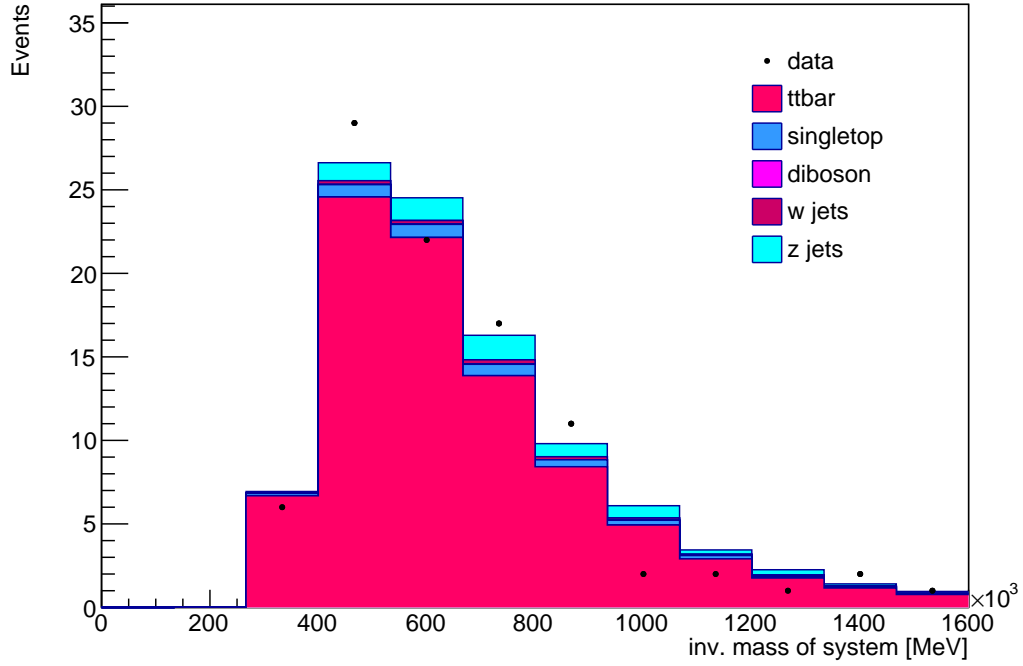
Diese Gewichte werden an die entsprechenden Samples multipliziert. In sogenannten Stacked Plots werden die einzelnen Untergründe aufeinander gestapelt und die Datenpunkte als overlay eingefügt um die Übereinstimmung von Daten und Simulation zu überprüfen. Exemplarisch werden in Abbildung 5 vier Stacked Plots auf fehlerhafte Modellierung diskutiert, die restlichen werden im Anhang 7.2 aufgelistet.



**Abbildung 5:** Gestapelte Untergründe im Vergleich mit den selektierten Datensample. Zusehen sind der transversale Impuls der Myonen (a), die Anzahl an btagged Jets (b), der transversale Impuls des Jets mit dem höchsten transversalen Impuls des Events (c) und die fehlende transversale Energie (d).

Die größte Diskrepanz ist in der Verteilung der fehlenden Transversalenergie zu sehen. In den Bins, sind die Datenpunkte entweder deutlich über den Untergrunddaten oder deutlich unter ihnen abgebildet. Ein erhöhter Datenpunkt könnte prinzipiell einen Hinweis auf ein Signal liefern, da aber ebenso viele Punkte deutlich unter dem Untergrundsignal sind, liegt der Ursprung vermutlich eher in einem systematischen Fehler. Da diese Grafiken bereits deutlich auf Monte Carlo Fehlmodellierungen hinweisen, verlieren die Ergebnisse deutlich an Signifikanz.

Als nächstes wird die Diskriminante untersucht.



**Abbildung 6:** Verteilung der invarianten Masse des Systems, welches aus den vier Jets mit dem größten  $p_T$ , dem Lepton und dem Neutrino gebildet wird.

Im zweiten und fünften Bin sind Exzesse der Daten über dem Untergrund feststellbar. Jedoch auf Grund der vorher bemerkten Fehlmodellierung der Untergründe kann hier nicht von einer Entdeckung oder einem Hinweis auf ein  $Z'$  Signal gesprochen werden.

## 5 Statistische Analyse der Ergebnisse

Im vorherigen Kapitel wurde die Übereinstimmung der Daten mit den Simulationen bereits qualitativ durchgeführt. Um nun die Übereinstimmung quantitativ zu bestimmen, wird ein  $X^2$ -Test durchgeführt. Dieser  $X^2$ -Test ist ein statistisches Maß um die Abweichung zwischen einer beobachteten und erwarteten Verteilung zu bestimmen. Die erwartete Verteilung entspricht dabei den Monte-Carlo Simulationen, also der Untergrundverteilungen plus einer spezifischen Massenhypothese des  $Z'$ . Die beobachtete Verteilung entspricht den gemessenen Daten. Die Prüfgröße  $X^2$  berechnet sich dann durch

$$X^2 = \sum_{j=1}^m \frac{(N_j - n_{0j})^2}{n_{0j}} \quad (3)$$

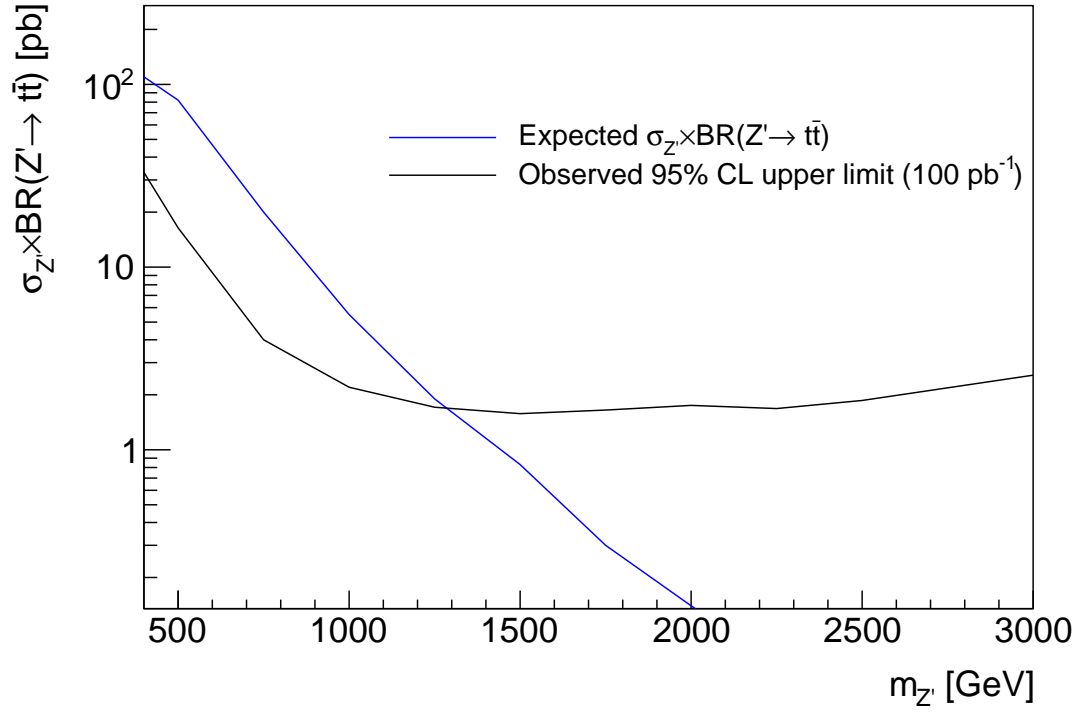
und gibt die Größe der Abweichung in jedem Bin der Verteilung an. Dabei ist  $N_j$  die beobachtete und  $n_{0j}$  die erwartete Häufigkeit. Dies wird auf dem Datensample

`data.mu.2.root` für die Diskriminante, also der invarianten Masse des Systems durchgeführt. Ist die Nullhypothese, dass es ein  $Z'$  gibt, wahr, so muss der Unterschied zwischen  $N_j$  und  $n_{0j}$  sehr gering sein. Das Konfidenzlevel berechnet sich aus  $p = 1 - \alpha$ , wobei  $\alpha$  die Signifikanz des Tests ist. Damit eine Hypothese angenommen wird, muss  $X^2 < X^2_{1-\alpha}$  gelten. Das Konfidenzlevel wird durch das Paket `TMath` von `root` berechnet und ist von der Anzahl an Freiheitsgraden, bzw. Bins abhängig. Da keine Hinweise auf ein mögliches  $Z'$  gefunden wurden, kann in einem nächsten Schritt eine obere Ausschlussgrenze im 95 % Konfidenzintervall für die verschiedenen  $Z'$ -Massenhypothesen gefunden werden. Um die Gewichte der einzelnen Hypothesen zu berücksichtigen wird ein scaler genutzt, der die Gewichte entsprechend skaliert bis das 95 % Konfidenzintervall erreicht wurde. Diese Ergebnisse sind in Tabelle 4 dargestellt. Wie zuvor festgestellt, kann keine der Massenhypothesen bestätigt werden. Sie alle erfüllen die geforderte Signifikanz nicht.

**Tabelle 4:** Die berechneten Parameter des  $X^2$ -Tests für eine reine Untergrundverteilung und den verschiedenen  $Z'$ -Massenhypothesen. Für die Massenhypothesen wurde außerdem die erwartete obere Ausschlussgrenze des Wirkungsquerschnittes im 95 % Konfidenzintervall berechnet.

Process	Freiheitsgrade	Konfidenzlevel	$X^2$	scaler	$\sigma/\text{pb}$
Reiner Untergrund	11	0.9162	5.29	-	-
$Z'(400)$	11	0.0209	-	0.3	33
$Z'(500)$	11	0.0008	-	0.2	16.4
$Z'(750)$	11	0.012	-	0.2	4
$Z'(1000)$	11	0.0461	-	0.4	2.2
$Z'(1250)$	11	0.0257	-	0.9	1.71
$Z'(1500)$	11	0.0429	-	1.9	1.58
$Z'(1750)$	11	0.0470	-	5.5	1.65
$Z'(2000)$	11	0.0483	-	12.5	1.75
$Z'(2250)$	11	0.0493	-	25.1	1.68
$Z'(2500)$	11	0.0497	-	53.2	1.86
$Z'(3000)$	11	0.489	-	213.4	2.56

Abschließend werden die so berechneten erwarteten oberen Ausschlussgrenzen der Wirkungsquerschnitte im 95 % Konfidenzintervall der unterschiedlichen  $Z'$ -Massenhypothesen gegen ihre Masse aufgetragen. Dies wird mit den erwarteten Wirkungsquerschnitten verglichen. Das ist in Abbildung 7 dargestellt. Aus diesen Plots ist erkennbar, dass nur Massen bis 1250 GeV erlaubt sind. Allerdings wurden in den in Kapitel 4 angefertigten stacked Plots keine entsprechenden Hinweise gefunden. Die höheren Massen können ausgeschlossen werden, da die Daten ansonsten nicht zu dem simulierten Untergrund passen würden.



**Abbildung 7:** Berechnete erwarteten oberen Ausschlussgrenzen der Wirkungsquerschnitte im 95 % Konfidenzintervall der unterschiedlichen  $Z'$ -Massenhypothesen gegen ihre Masse. Zum Vergleich ist ebenfalls die erwartete Verteilung der Wirkungsquerschnitte dargestellt.

## 6 Diskussion

Die Ergebnisse dieses Versuchs zeigen einen Hinweis auf einen systematischen Fehler entweder in der Monte Carlo Simulation selber oder in der Selektion hin. Die Datenpunkte in den Stacked Plots, wie beispielsweise in Abbildung 5 zeigen deutliche Streuung von den gestapelten Untergründen. Auch in der Finalen Diskriminanten sind diese Abweichungen zu finden, weswegen alle späteren Ergebnisse eine Fortpflanzung dieses systematischen Fehlers beinhalten und somit an Relevanz verlieren.

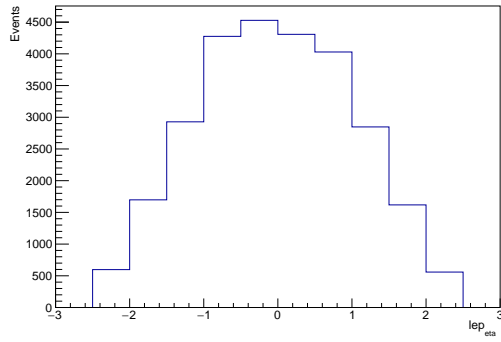
Zur Bewertung und Beurteilung des Lehrstuhlversuchs ist folgendes Anzumerken. Zu einem wurden eine Methode der Datenanalyse aufgezeigt, was zu einem sehr interessant war, aber zum anderen auch sehr anspruchsvoll. Der Versuch ist nur mit weiterführenden Kenntnissen in die Programmiersprachen C++ und Root zu bewältigen und fällt deutlich schwerer, wenn diese nicht in der Tiefe vorhanden sind. Des Weiteren ist der Umfang des Versuchs unserer Meinung nach etwas zu hoch. Auf der einen Seite, ist es uns verständlich, dass nur die volle Analyse das komplette Verständnis für diesen Fachbereich liefert, allerdings ist es im Hinblick darauf, dass noch weitere Versuche durchgeführt werden müssen, zeitlich nur schwer machbar in diesem Umfang. Hinzu kommt die aktuelle Situation, die es einem nur ermöglicht remote miteinander zu arbeiten oder Fragen zu stellen. Dies hat die Zusammenarbeit und das Verständnis manchmal sehr erschwert.

## 7 Anhang

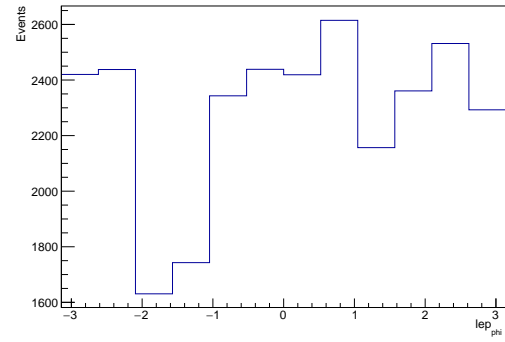
### 7.1 Weitere Plots zu den fundamentalen Verteilungen

Die restlichen Plots des  $t\bar{t}$ -Prozesses dargestellt sind in den Abbildungen 8 - 10 zusammengefasst. Die restlichen Vergleiche der konstruierten Variablen zwischen der  $t\bar{t}$  Untergrund- und der  $Z'(1000)$  Signal- Simulation sind in den Abbildungen 11 - 13 aufgetragen.

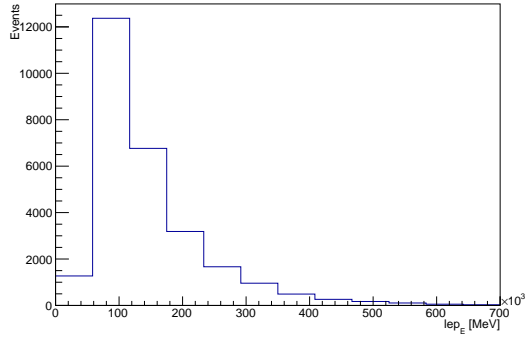




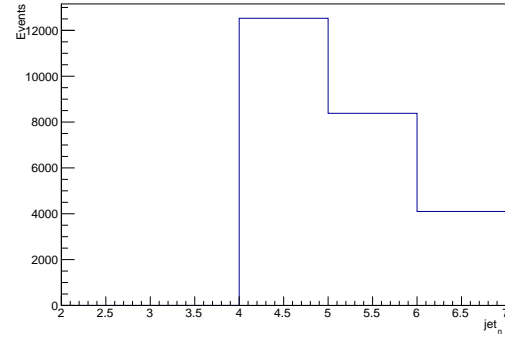
(a)



(b)

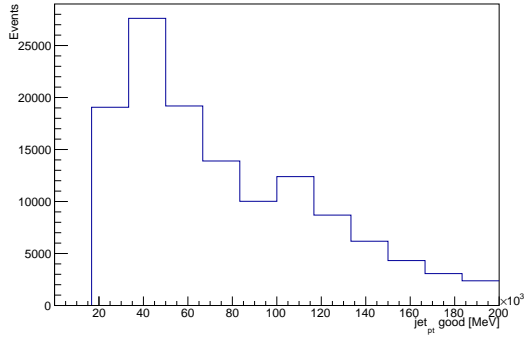


(c)

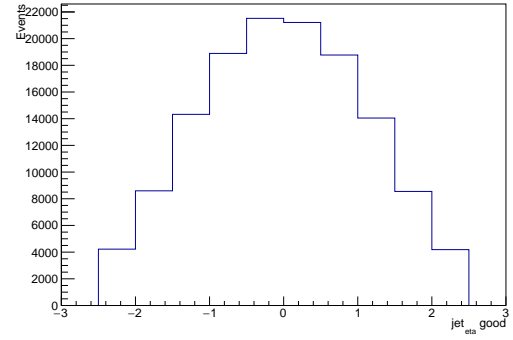


(d)

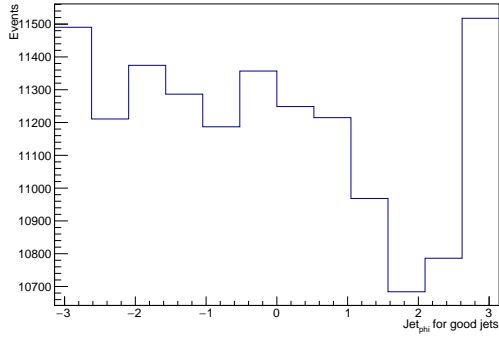
**Abbildung 8:** Darstellung verschiedener Verteilungen der Größen der  $t\bar{t}$  Monte-Carlo Simulation. Zusehen sind die Pseudorapidity der Myonen (a), der Azimuthalwinkel der Leptonen (b), der Energie der Myonen (c) und die Anzahl an nutzbaren Jets (d).



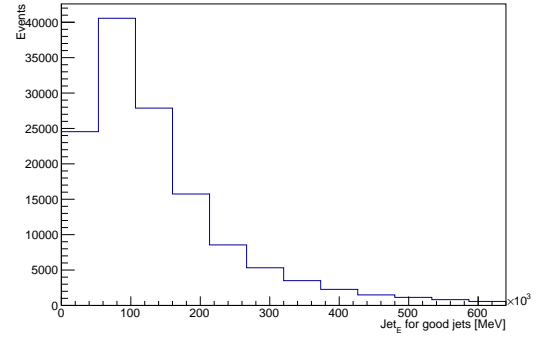
(a)



(b)

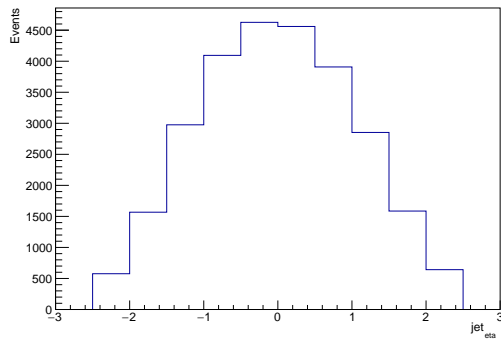


(c)

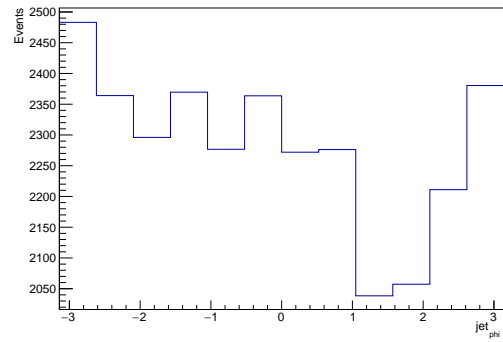


(d)

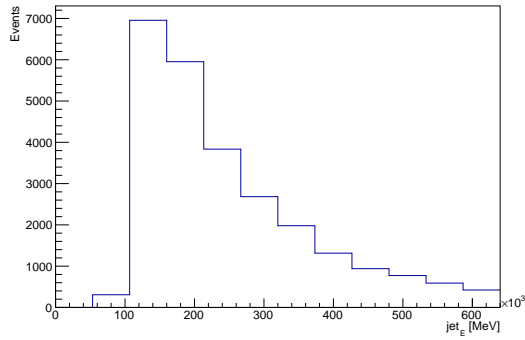
**Abbildung 9:** Darstellung verschiedener Verteilungen der Größen der  $t\bar{t}$  Monte-Carlo Simulation. Zusehen sind der transversale Impuls der nutzbaren Jets (a), die Pseudorapidität der nutzbaren Jets (b), der Azimuthalwinkel der nutzbaren Jets (c) und die Energie der nutzbaren Jets (d).



(a)

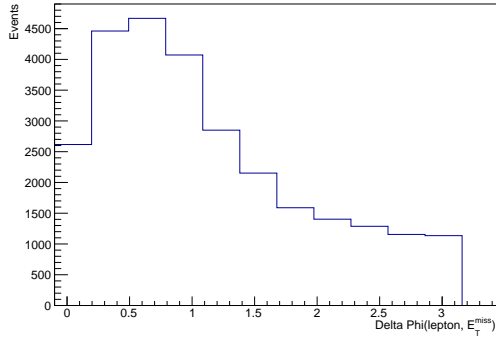


(b)

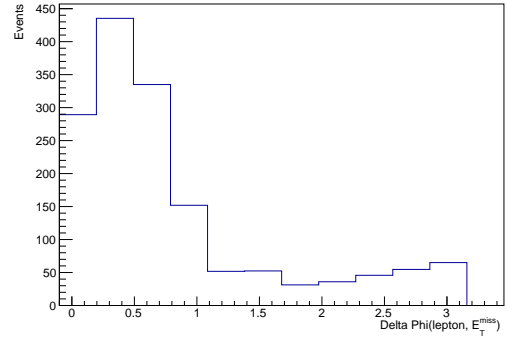


(c)

**Abbildung 10:** Darstellung verschiedener Verteilungen der Größen der  $t\bar{t}$  Monte-Carlo Simulation. Zusehen sind die Pseudorapidity des Jets mit dem höchsten  $p_T$  (a), der Azimuthalwinkel des Jets mit dem höchsten  $p_T$  (b) und die Energie des Jets mit dem höchsten  $p_T$  (c).

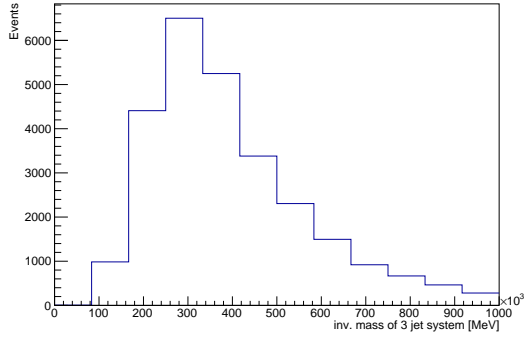


(a)

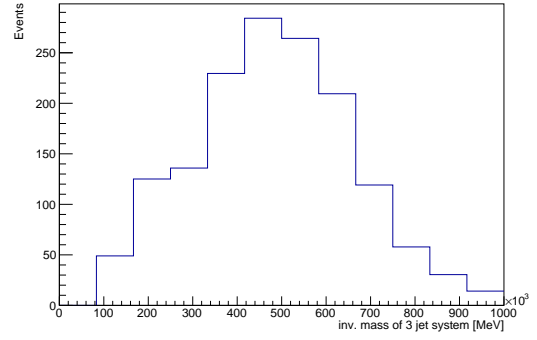


(b)

**Abbildung 11:** Vergleich der Verteilung der Differenz zwischen dem Azimutalwinkel der fehlenden Transversalenergie und der Leptonflugrichtung. Dies ist für die  $t\bar{t}$ -Untergrundsimulation a und der Signalsimulation des  $Z'(1000)$  b dargestellt.

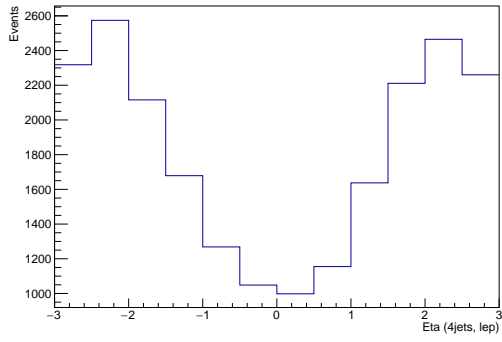


(a)

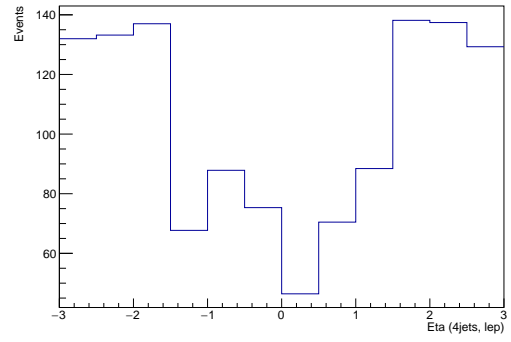


(b)

**Abbildung 12:** Vergleich der Verteilung der invarianten Masse des Systems, die durch die drei Jets mit dem größten  $p_T$  gebildet wird. Dies ist für die  $t\bar{t}$ -Untergrundsimulation a und der Signalsimulation des  $Z'(1000)$  b dargestellt.



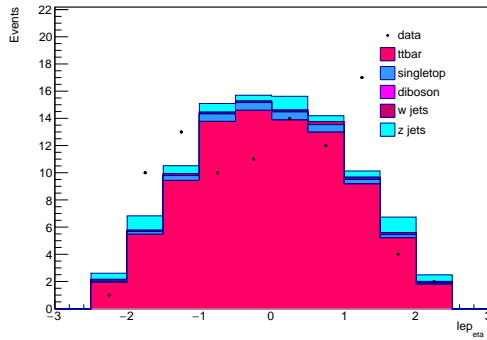
(a)



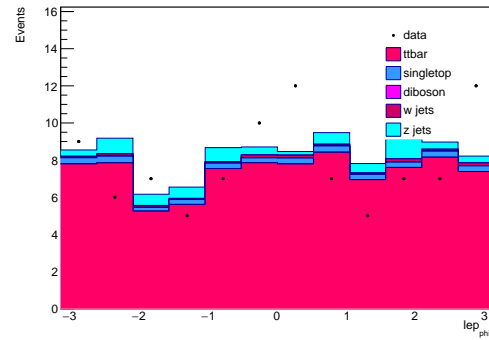
(b)

**Abbildung 13:** Vergleich der Verteilung der Pseudorapidity des Systems, die durch die vier Jets mit dem größten  $p_T$ , dem Lepton und dem Neutrino gebildet wird. Dies ist für die  $t\bar{t}$ -Untergrundsimulation a und der Signalsimulation des  $Z'(1000)$  b dargestellt.

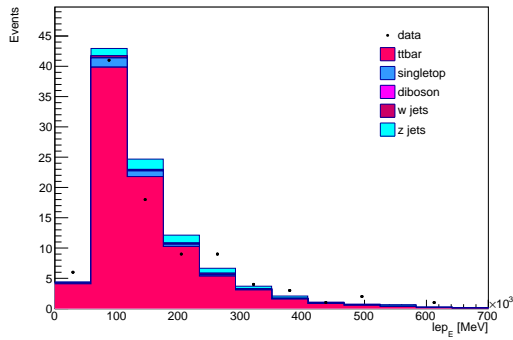
## 7.2 Weitere Plots zur Überprüfung der Monte Carlo Übereinstimmung mit den Daten



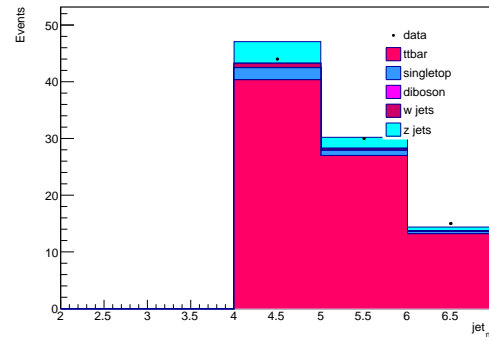
(a)



(b)

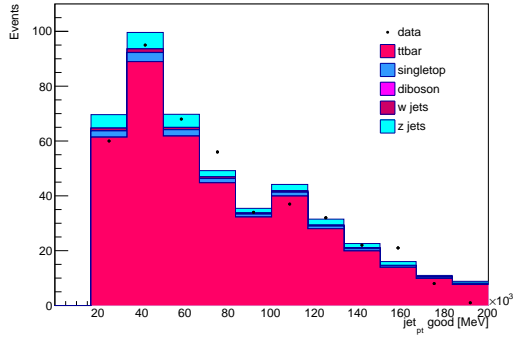


(c)

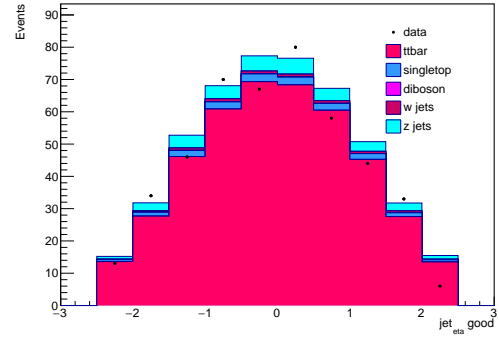


(d)

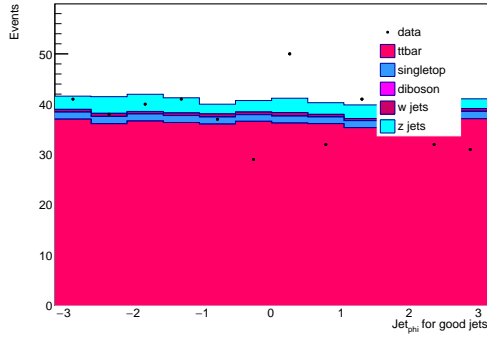
**Abbildung 14:** Darstellung verschiedener Verteilungen der Größen der  $t\bar{t}$  Monte-Carlo Simulation. Zusehen sind die Pseudorapidität der Myonen (a), der Azimuthalwinkel der Leptonen (b), der Energie der Myonen (c) und die Anzahl an nutzbaren Jets (d).



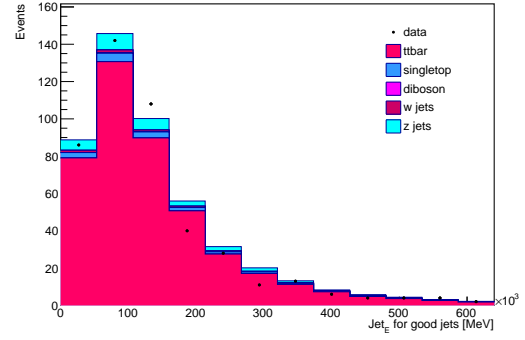
(a)



(b)

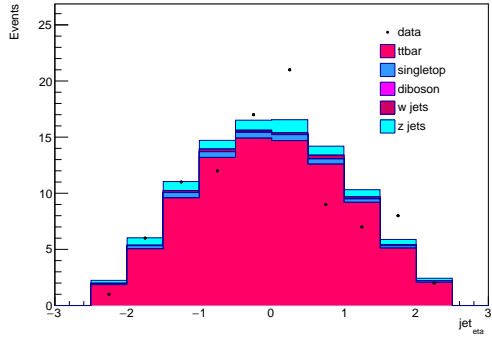


(c)

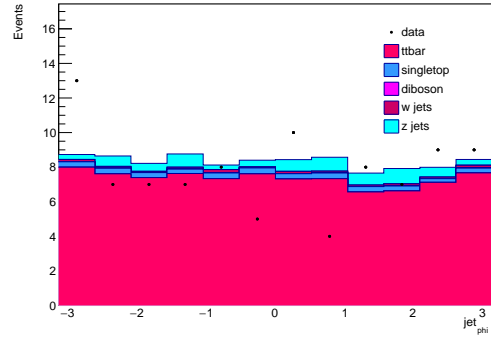


(d)

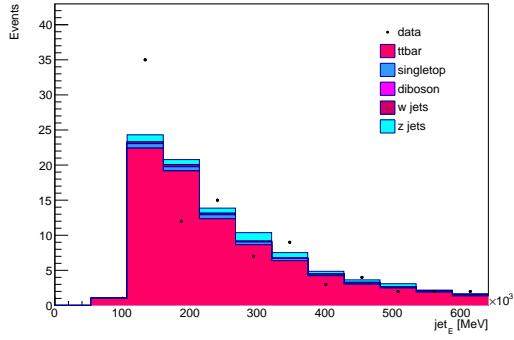
**Abbildung 15:** Darstellung verschiedener Verteilungen der Größen der  $t\bar{t}$  Monte-Carlo Simulation. Zusehen sind der transversale Impuls der nutzbaren Jets (a), die Pseudorapidität der nutzbaren Jets (b), der Azimuthalwinkel der nutzbaren Jets (c) und die Energie der nutzbaren Jets (d).



(a)



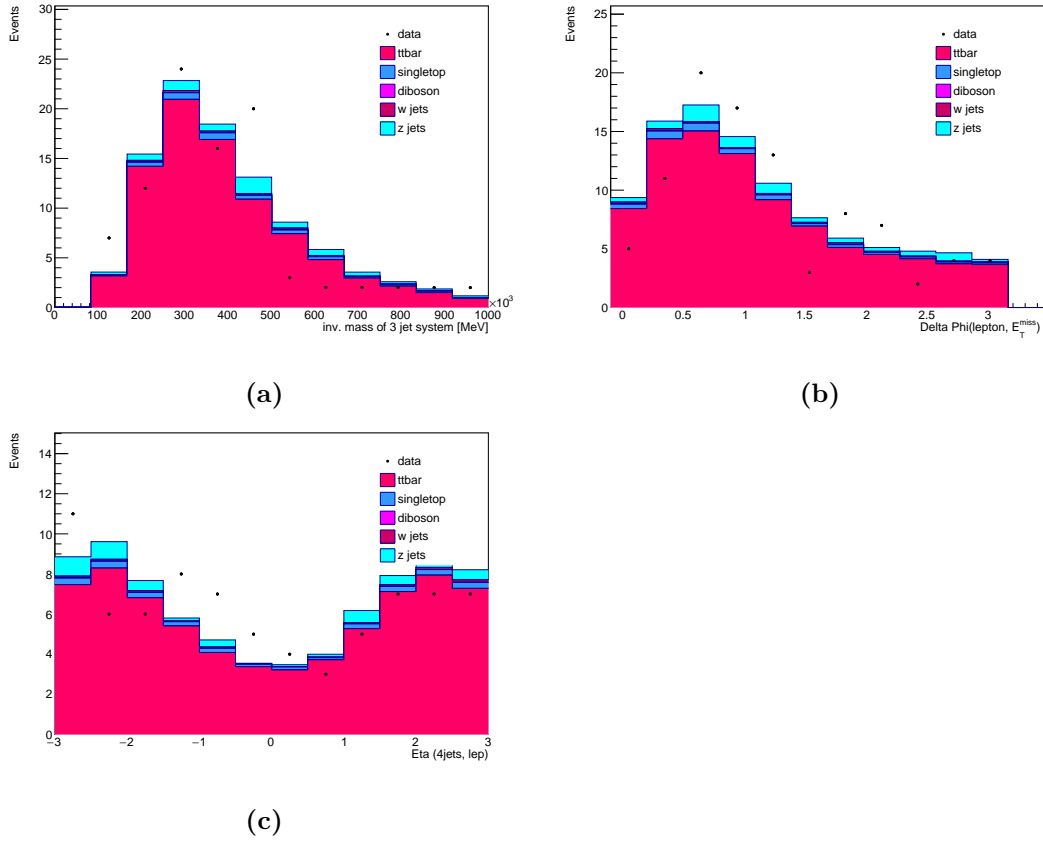
(b)



(c)

**Abbildung 16:** Darstellung verschiedener Stacked Plots als Überprüfung der Übereinstimmung von Simulation und Daten. Zusehen sind die Pseudorapidität des Jets mit dem höchsten  $p_T$  (a), der Azimuthalwinkel des Jets mit dem höchsten  $p_T$  (b) und die Energie des Jets mit dem höchsten  $p_T$  (c).





**Abbildung 17:** Darstellung verschiedener Stacked Plots als Überprüfung der Übereinstimmung von Simulation und Daten. Zusehen ist die invariante des Dreijetsystems aus den drei Jets mit dem höchsten  $p_T$  (a), der Azimuthwinkel zwischen der fehlenden Transversalenergie und dem geladenen Lepton (b) und die Pseudorapidität des gesamten Systems (geformt aus den vier Jets mit dem größten  $p_T$ , dem Lepton und dem Neutrino) (c).

## Literatur

- [1] Atlas Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation, , JINST 3 (2008) S08003
- [2] CMS Collaboration, *Search for  $t\bar{t}$  production in proton-proton collisions at 13 TeV*, Journal of High Energy Physics, , JHEP 04 (2019) 031