

# Studien zur Entfaltung der $\Delta R(b\bar{b})$ -Verteilung des $t\bar{t} + b\bar{b}$ -Prozesses mit TUnfold

Bachelorarbeit

Kerstin Märkle

An der Fakultät für Physik  
Institut für Experimentelle Teilchenphysik

Erstgutachter:	Prof. Dr. Ulrich Husemann
Zweitgutachter:	Dr. Matthias Schröder
Betreuender Mitarbeiter:	Jan van der Linden

Karlsruhe, 14. 01. 2021



---

Diese Arbeit wurde vom Erstgutachter der Bachelorarbeit akzeptiert.

**Karlsruhe, den 14.01.2021**

.....  
(Prof. Dr. Ulrich Husemann)



---

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

**Karlsruhe, den 14.01.2021**

.....  
(Kerstin Märkle)



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Physikalischer Hintergrund</b>	<b>3</b>
2.1	Standardmodell der Teilchenphysik . . . . .	3
2.2	Compact-Muon-Solenoid-Experiment . . . . .	4
2.3	Objektrekonstruktion . . . . .	7
2.4	Ereignissimulation . . . . .	7
2.5	Der $t\bar{t} + b\bar{b}$ -Prozess . . . . .	8
<b>3</b>	<b>Entfaltungsmethode</b>	<b>11</b>
3.1	Allgemeine Beschreibung der Entfaltung . . . . .	11
3.2	Entfaltung mit Hilfe von TUnfold . . . . .	13
3.3	Bedingungen für die Migrationsmatrix $\mathbf{M}$ . . . . .	17
<b>4</b>	<b>Studium der Entfaltung</b>	<b>19</b>
4.1	Binning . . . . .	19
4.2	Erzeugung der Pseudodaten für die Entfaltung . . . . .	20
4.3	Studium zur Variation der Migrationsmatrix durch verschiedene Bedingungen an Reinheit und Stabilität . . . . .	22
4.4	Studien zur Entfaltung durch Variation der Pseudodaten . . . . .	28
4.5	Untersuchung der systematischen Unsicherheiten . . . . .	32
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>39</b>
	<b>Literatur</b>	<b>41</b>



# 1 Einleitung

Die Teilchenphysik beschäftigt sich mit der Erforschung der Elementarteilchen und ihrer Wechselwirkungen. Eine der wichtigsten Theorien ist dabei das Standardmodell der Teilchenphysik (SM). Viele Vorhersagen aus diesem Modell ließen sich in den letzten Jahrzehnten experimentell bestätigen. Beispielsweise ermöglichte das Standardmodell im Jahr 2012 die Entdeckung des Higgs-Teilchens [1]. Das SM gilt daher als die bisher erfolgreichste Theorie, die einen Einblick in die Eigenschaften der Elementarteilchen sowie deren Wechselwirkungen ermöglicht.

Zu Überprüfung physikalischer Modelle werden hochenergetische Streuprozesse durchgeführt. Um eine theoretische Vorhersage der Streuprozesse zu erhalten, werden Ereignisse mit Hilfe der Monte-Carlo-Methode erzeugt. Jedoch lassen sich die gemessenen Daten nicht direkt mit der theoretischen Vorhersage vergleichen. Durch eine begrenzte Akzeptanz, Effizienz und Genauigkeit des Detektors weichen die gemessenen Daten von den simulierten Daten ab. Diese Detektoreffekte lassen sich durch eine Entfaltung der gemessenen Daten korrigieren. Die Entfaltung ermöglicht damit einen Vergleich der simulierten Ereignisse mit den gemessenen Daten ohne Verschmierungen des Detektors.

In dieser Arbeit wird die Entfaltung mit dem TUnfold-Algorithmus für die  $\Delta R(b\bar{b})$ -Verteilung des  $t\bar{t} + b\bar{b}$ -Prozesses untersucht. Dieser Prozess ist einer der wichtigsten Untergrundprozesse des  $t\bar{t}H(b\bar{b})$ -Prozesses. Die Untersuchung der Entfaltungsmethode steht dabei im Vordergrund. Dazu werden simulierte Daten des Compact-Muon-Solenoid-Experiments aus dem Jahr 2018 mit einer integrierten Luminosität von  $L_{\text{int}} = 59.70 \text{ fb}^{-1}$  verwendet. Statt gemessenen Daten werden poissonverschmierte Pseudodaten aus den simulierten Ereignissen erzeugt und entfaltet. Mit diesen Pseudodaten werden Studien zur Optimierung des Binnings der zu entfaltenden Verteilungen durchgeführt. Anschließend wird die integrierte Regularisierungsmethode des TUnfold-Algorithmus untersucht. Mit Hilfe von vergrößerten statistischen Fluktuationen der Pseudodaten sowie einer Skalierung der Pseudodaten wird die Stabilität der Entfaltung getestet. Zum Abschluss der Studien werden durch eine Variation der simulierten Verteilungen mit unterschiedlichen Gewichtungen der Ereignisse systematische Unsicherheiten auf die entfalteten Ergebnisse betrachtet.

Das Kapitel 2 liefert einen kurzen Überblick über die physikalischen Grundlagen. Anschließend wird in Kapitel 3 der Entfaltungsprozess und das Werkzeug TUnfold vorgestellt, welches in dieser Arbeit zur Entfaltung verwendet wird. Die Studien zur Entfaltung mit TUnfold werden in Kapitel 4 präsentiert. Dabei liegt der Schwerpunkt dieser Arbeit auf der

Methode der Entfaltung. Abschließend werden in Kapitel 5 die Ergebnisse dieser Arbeit zusammengefasst.

## 2 Physikalischer Hintergrund

In der Teilchenphysik werden zur Erforschung der Teilchen und ihrer Wechselwirkungen hochenergetische Streuexperimente durchgeführt. Diese Streuprozesse werden zusätzlich mit Hilfe der Monte-Carlo-Methode simuliert. Anschließend lassen sich die simulierten Ereignisse mit den gemessenen Daten vergleichen. Die Simulation basiert auf einer Theorie, die den Teilchenprozess möglichst genau vorhersagen soll. Das Standardmodell der Teilchenphysik fasst hierfür die wesentlichen Erkenntnisse der Teilchenphysik in einem mathematischen Modell zusammen.

Dieses Kapitel gibt in Abschnitt 2.1 einen Überblick über das Standardmodell der Teilchenphysik, welches die physikalische Grundlage dieser Arbeit bildet.

Für die Erzeugung und Detektion von hochenergetischen Streuprozessen werden große Teilchenbeschleuniger und Detektoren benötigt. Der derzeit größte Teilchenbeschleuniger der Welt ist der Large Hadron Collider (LHC) am CERN. Dieser wird in Abschnitt 2.2 zusammen mit dem Compact-Muon-Solenoid-Detektor (CMS-Detektor) beschrieben.

Durch die kurzen Zerfallszeiten lassen sich bei Partonen nicht die ursprünglichen Teilchen, welche durch die Streuprozesse erzeugt werden, detektieren. Jedoch können mit Hilfe von Objektrekonstruktion, welche in 2.3 beschrieben wird, Rückschlüsse auf die physikalischen Informationen der ursprünglichen Teilchen gezogen werden.

In Abschnitt 2.4 wird auf die Simulation von Ereignissen eingegangen, welche zum Vergleich der gemessenen Daten und der theoretischen Vorhersage erzeugt werden.

Abschließend wird in Abschnitt 2.5 der  $t\bar{t} + b\bar{b}$ -Prozess, dessen kinematische Verteilung von  $\Delta R(b\bar{b})$  in dieser Arbeit untersucht wird, beschrieben. Dabei wird der semileptonische Zerfallskanal des  $t\bar{t} + b\bar{b}$ -Prozesses betrachtet. Anhand der Zerfallsprodukte lassen sich für die Daten Selektionskriterien definieren, die zur Reduzierung des Untergrunds verwendet werden.

### 2.1 Standardmodell der Teilchenphysik

Mit Hilfe des Standardmodells der Teilchenphysik können die fundamentalen Bestandteile der Physik beschrieben werden. Mit dieser Theorie werden drei der elementaren Grundkräfte der Natur erklärt: die elektromagnetische, die schwache und die starke Wechselwirkung. Im Folgenden wird basierend auf [2] ein Überblick über die Elementarteilchen und deren Wechselwirkungen gegeben.

Die Materie wird im Standardmodell in drei verschiedene Typen unterteilt.

Die Leptonen sind Fermionen und besitzen damit einen halbzahligen Spin. Sie lassen sich in drei Generationen einteilen, wobei jede Generation von einem einfach negativ geladenen Teilchen sowie einem neutralen Neutrino vertreten wird. Die negativ geladenen Teilchen werden als Elektronen, Myonen und Tau-Leptonen bezeichnet. Für die Zuordnung der Generation sind die Neutrinos nach ihren jeweiligen geladenen Partnern benannt. Zusammen mit den sechs Antiteilchen, dem Positron, dem Anti-Myon, dem Anti-Tau und den drei Anti-Neutrinos gibt es insgesamt zwölf Leptonen.

Neben den leptonischen Fermionen gibt es die Quarks, welche im Gegensatz zu den Leptonen eine nicht ganzzahlige elektrische Ladung tragen. Sie besitzen jeweils eine Flavour-Quantenzahl, mit der sie sich klassifizieren lassen. Es wird zwischen den Flavours Up, Down, Strange, Charm, Beauty und Truth unterschieden. Diese dienen ebenso als Bezeichnungen der Quarks. Zusätzlich zu ihrer elektrischen Ladung tragen sie eine von insgesamt drei Farbladungen. Durch das Confinement der starken Wechselwirkung können farbgeladene Teilchen nicht als freie Teilchen vorkommen. Dies führt zur Hadronisierung, wodurch es zu Verbindungen aus zwei oder drei Quarks bzw. Antiquarks kommt. Diese Verbindungen werden Mesonen und Baryonen genannt. Dabei besteht die Materie im Universum fast nur aus den zwei bekanntesten Baryonen: dem Proton und dem Neutron.

Den dritten Teilchentyp bilden die Bosonen, welche im Gegensatz zu den Fermionen einen ganzzahligen Spin besitzen. Im Standardmodell gibt es vier Eichbosonen: Das neutrale Photon und Z-Boson, die elektrisch geladenen W-Bosonen und farbgeladene Gluonen. Die ersten drei genannten Eichbosonen beschreiben die elektroschwache Wechselwirkung, welche bei hohen Energien die schwache Wechselwirkung und die elektromagnetische Wechselwirkung vereint. Das Photon, welches die elektromagnetische Wechselwirkung beschreibt, koppelt nur an elektrisch geladene Teilchen. Die schwache Wechselwirkung hingegen koppelt sowohl an die elektrisch geladenen Leptonen und Quarks als auch an die Neutrinos. Damit koppelt die elektroschwache Wechselwirkung ebenfalls an alle Leptonen und an die Quarks.

Zusätzlich gibt es noch die starke Wechselwirkung. Sie wird über den Austausch von Gluonen beschrieben. Diese wechselwirken nur mit den farbgeladenen Quarks oder mit anderen Gluonen. Durch ihre doppelte Farbladung bewirken sie einen Austausch der Quark-Farbe.

Das Higgs-Boson nimmt im Standardmodell eine gesonderte Stellung ein. Es wird als gequantelte Anregung des Higgsfeldes beschrieben. Dabei verleiht das Higgsfeld allen Elementarteilchen ihre Masse.

Mit Hilfe des Standardmodells der Teilchenphysik lassen sich eine Vielzahl der beobachteten physikalischen Teilchenprozesse erklären. Jedoch ist es unvollständig, da es die Gravitation, welche eine Anziehung zwischen massiven Objekten bewirkt, nicht beinhaltet. Ebenfalls liefert das Standardmodell keine Erklärung für die Masse der Neutrinos, welche von Null verschieden ist. Dies deutet darauf hin, dass das Standardmodell nur ein Teil einer noch allumfassenderen Theorie sein könnte, welche die Gesetze des Universums beschreibt.

## 2.2 Compact-Muon-Solenoid-Experiment

Am Compact-Muon-Solenoid-Experiment werden Streuexperimente mit hochenergetischen Teilchen durchgeführt. Diese dienen zur Überprüfung des Standardmodells der Teilchenphysik sowie zur Überprüfung von Theorien, welche über das Standardmodell hinausgehen.

Der Large Hadron Collider (LHC), welcher als Teilchenbeschleuniger des CMS-Experiments dient, wird in Abschnitt 2.2.1 beschrieben. Anschließend wird in Abschnitt 2.2.2 auf den CMS-Teilchendetektor eingegangen und in Abschnitt 2.2.3 dessen Kinematik beschrieben.

### 2.2.1 Large Hadron Collider

Der Large Hadron Collider (LHC) am Europäischen Kernforschungszentrum CERN bei Genf ist der bislang größte und hochenergetischste Teilchenbeschleuniger. Dieser wurde gebaut um sowohl Protonen als auch Blei-Ionen auf nahezu Lichtgeschwindigkeit zu beschleunigen und anschließend zur Kollision zu bringen. Damit die Teilchen miteinander kollidieren können, werden zwei Strahlen in entgegengesetzter Richtung erzeugt. Diese werden mit Hilfe von supraleitenden Magneten in zwei voneinander getrennten Strahlrohren auf ihren Bahnen gehalten. Seit 2016 wird der LHC durchgehend mit einer Schwerpunktsenergie von  $\sqrt{s} = 13 \text{ TeV}$  betrieben. Dabei wurde zwischen 2016 und 2018 eine integrierte Luminosität von  $L_{\text{int}} = 137 \text{ fb}^{-1}$  aufgenommen. Die integrierte Luminosität ist dabei ein Maß für die Größe des aufgenommenen Datensatzes.

In den vier großen Teilchendetektoren ALICE [3], ATLAS [4], CMS [5] und LHCb [6] werden die Teilchenstrahlen des LHCs zur Kollision gebracht und die Signale der Zerfallsprodukte aufgenommen. Das ATLAS- und das CMS-Experiment verfolgen dabei die gleichen physikalischen Ziele, jedoch unter anderen technischen Voraussetzungen. Damit lassen sich gemessene Daten des CMS-Experiments mit den Daten aus dem ATLAS-Experiment und umgekehrt überprüfen.

### 2.2.2 Der Compact-Muon-Solenoid-Detektor

Der Compact-Muon-Solenoid-Detektor (CMS-Detektor) wurde nach seiner kompakten Bauweise, seiner guten Auflösung von Myonenspuren sowie seinem starken Solenoid-Magneten benannt. Er besteht aus mehreren Schichten, welche in Abbildung 2.1 dargestellt sind. Sie dienen jeweils zur Detektion von unterschiedlichen Teilchen.

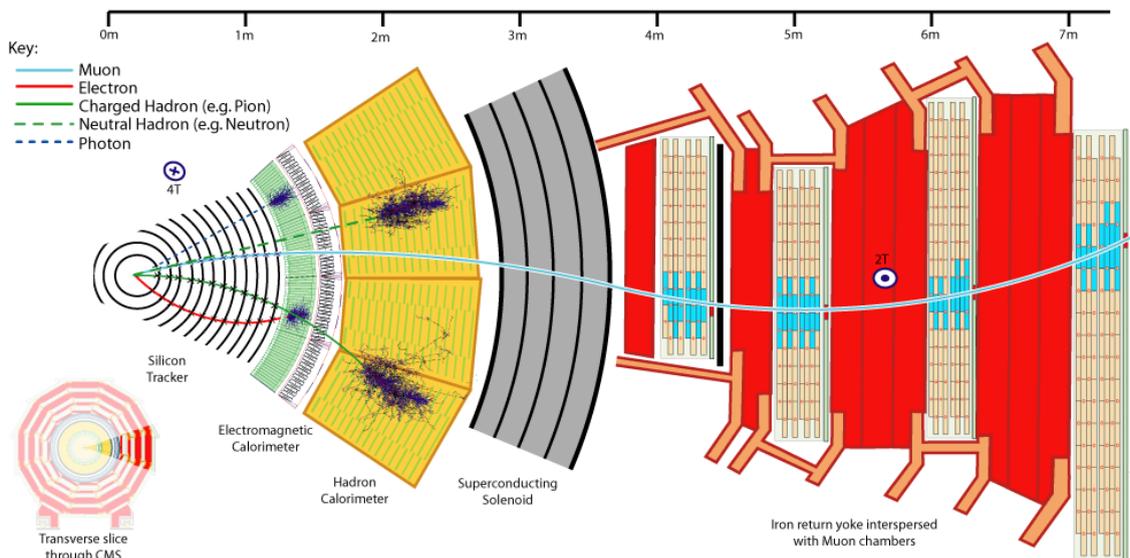


Abbildung 2.1: In diesem Schnitt des CMS-Detektors sind die einzelnen Schichten zu sehen. Diese Schichten besitzen jeweils eine unterschiedliche Funktionsweise und werden zur Detektion von verschiedenen Teilchen genutzt [7].

Die innerste Schicht des CMS-Detektors besteht aus einem Silizium-Spurdetektor [8], welcher die Spuren von geladenen Teilchen misst. Sowohl die geladenen Leptonen als auch

die geladenen Hadronen werden detektiert. Das Magnetfeld im Inneren des Detektors bewirkt eine Krümmung der Spur, wodurch der Impuls und die Ladung der gemessenen Teilchen bestimmt werden können.

Im elektromagnetischen Kalorimeter [9] (ECAL) erzeugen Elektronen und Photonen elektromagnetische Schauer. Diese werden hauptsächlich durch Bremsstrahlung und Elektron-Positron-Paarbildung verursacht. Die Detektion der Schauerteilchen ermöglicht es, die Energie der ursprünglichen Teilchen zu rekonstruieren.

Die starke Wechselwirkung verursacht im hadronische Kalorimeter [10] (HCAL) ebenfalls eine Schauerentwicklung. Damit lässt sich sowohl die Energie von geladenen als auch die von nicht geladenen Hadronen messen.

Beim CMS-Experiment befinden sich die Kalorimeter innerhalb des Solenoidmagneten. Dadurch kann die Energie von Teilchen einerseits genauer vermessen werden, da die Wechselwirkung von Teilchen durch den Magneten nicht relevant ist, andererseits ist jedoch die Länge der Kalorimeter durch die Größe des Magneten beschränkt.

Zur Detektion von Myonen wird ein weiterer Spurdetektor, bestehend aus Gasionisationskammern [11], verwendet. Photonen und Elektronen geben bereits durch Paarbildung und Bremsstrahlung ihre Energie vollständig im ECAL ab. Auch die Hadronen werden im HCAL zum Großteil absorbiert. Da die Myonen durch ihre geringere spezifische Ladung im Magnetfeld weniger stark abgelenkt werden als Elektronen, erzeugen sie weniger Bremsstrahlung und können das ECAL und das HCAL durchdringen. Damit lässt sich die Spur, welche in dem zusätzlichen Spurdetektor gemessen wird, einem Myon zuordnen.

Die verschiedenen Schichten mit ihren unterschiedlichen Funktionsweisen ermöglichen es, mit Hilfe des CMS-Detektors sehr viele physikalische Prozesse, welche durch Teilchenkollisionen entstehen, nachzuvollziehen.

Jedoch besitzt auch der CMS-Detektor nur eine begrenzte Auslösung. So können Akzeptanz- und Effizienzeffekte auftreten, welche zu einer Verschmierung der Messdaten führen.

### 2.2.3 Kinematik im CMS-Detektor

Auf Grund der zylindrischen Form des Detektors werden Polarkoordinaten zur Berechnung der kinematischen Größen verwendet. Das Koordinatensystem wird so gelegt, dass die  $z$ -Achse entlang der Strahlrichtung verläuft. Der Polarwinkel  $\theta$  bezeichnet den Winkel zur  $z$ -Achse. Der Azimutwinkel in der  $x$ - $y$ -Ebene wird  $\varphi$  genannt.

Da der Impuls in Richtung der  $z$ -Achse nicht null ist, bewegt sich das Ruhesystem der Teilchen entlang der  $z$ -Achse. Daher ist es sinnvoll, Größen zu verwenden, welche gegenüber einem Boost parallel zur  $z$ -Achse invariant sind. Diese Invarianz gilt z.B. für den Transversalimpuls

$$\mathbf{p}_T = \sqrt{\mathbf{p}_x^2 + \mathbf{p}_y^2}. \quad (2.1)$$

Es wird angenommen, dass die Teilchen vor der Kollision einen verschwindenden Transversalimpuls besitzen. Durch die Impulserhaltung lässt sich damit die fehlende Transversalenergie  $\cancel{E}_T$  über die Summe der gemessenen Transversalimpulse  $\mathbf{p}_{T,i}$  durch

$$\cancel{E}_T = \left| \sum_i^N \mathbf{p}_{T,i} \right| \quad (2.2)$$

berechnen. Dabei ist  $N$  die Anzahl der detektierten Teilchen. Der fehlende Transversalimpuls lässt Rückschlüsse auf möglicherweise nicht detektierte Teilchen wie beispielsweise Neutrinos ziehen.

Die Rapidität, welche als Maß für die relativistische Geschwindigkeit gilt, wird als

$$y = \frac{1}{2} \ln \left( \frac{E_T + cp_z}{E - cp_z} \right) \quad (2.3)$$

definiert. Dabei ist  $E$  die Energie des Teilchens und  $p_z$  der Impuls entlang der Strahlachse. Durch die hohen Geschwindigkeiten der Teilchen gilt:  $pc \gg mc^2$ . Damit lässt sich die Rapidität näherungsweise als

$$y \approx \frac{1}{2} \ln \left( \frac{p(1 + \cos(\theta))}{p(1 - \cos(\theta))} \right) = -\ln \tan \left( \frac{\theta}{2} \right) =: \eta, \quad (2.4)$$

schreiben, wobei  $\eta$  die Pseudorapidität beschreibt. Letztere kann direkt aus der Messung bestimmt werden.

Die Winkelöffnung in der  $\eta$ - $\varphi$ -Ebene von zwei Teilchen ist definiert durch

$$\Delta R = \sqrt{(\Delta\varphi_{ij})^2 + (\Delta\eta_{ij})^2}. \quad (2.5)$$

Dabei beschreibt  $\Delta\varphi_{ij}$  die Differenz der Winkel und  $\Delta\eta_{ij}$  die Differenz der Pseudorapidität der Teilchen.

## 2.3 Objektrekonstruktion

Durch die Detektion von Zerfallsprodukten aus Teilchenkollisionen entstehen große Datenmengen. Diese werden mit verschiedenen Analysemethoden untersucht. Damit lassen sich die physikalischen Informationen der ursprünglichen Teilchen, welche durch die Kollision erzeugt werden, rekonstruieren. Erst durch diese Objektrekonstruktion lassen sich Aussagen über den eigentlichen Teilchenprozess treffen.

Zur Objektrekonstruktion aus den gemessenen Signalen werden verschiedene Algorithmen verwendet. Dadurch lassen sich die physikalischen Informationen aus den Datensätzen gewinnen. Für die grundlegende Objektrekonstruktion wird der Particle-Flow-Algorithmus [12] verwendet. Mit diesem lassen sich Teilchen identifizieren und Informationen wie den Impuls oder die Energie der Teilchen bestimmen.

Anschließend wird der Anti- $k_T$ -Jet-Clustering-Algorithmus [13] verwendet. Dieser fasst einzelne Teilchen, die mit dem Particle-Flow-Algorithmus identifiziert wurden zu Teilchenjets zusammen. Dabei wird ein Radiusparameter von  $R = 0.4$  verwendet. Nun können die Energien und die Impulse der ursprünglichen Teilchen aus den Jets rekonstruiert werden.

Mit Hilfe des Deep-Jet-Algorithmus [14] lassen sich Informationen zum Ursprung der Jets gewinnen. Der Algorithmus berechnet beispielsweise die Wahrscheinlichkeit, dass ein Jet aus einem Bottom-Quark stammt. Dies wird auch als b-tag-Information bezeichnet.

Damit können Rückschlüsse auf den Teilchenprozess gezogen werden, aus dem die Jets stammen.

## 2.4 Ereignissimulation

Zum Vergleich der gemessenen Daten mit der theoretischen Vorhersage werden simulierte Ereignisse erzeugt. Dazu werden Ereignisgeneratoren verwendet, welche auf der Monte-Carlo-Methode basieren. Die Datensätze, die in dieser Arbeit verwendet werden,

stammen aus Simulationen mit dem Powheg-Matrixelement-Generator [15] und dem Pythia-8-Ereignisgenerator [16]. Der harte hochenergetische Prozess der Kollision wird durch den Matrixelement-Generator beschrieben, während mit dem Ereignisgenerator die Hadronisierung und der Zerfall der Teilchen simuliert wird.

Jedoch lassen sich die gemessenen Daten nicht direkt mit den simulierten Daten vergleichen. Der Detektor besitzt eine limitierte Akzeptanz, wodurch Verluste in der Selektion der Daten auftreten. Zusätzlich erzeugt eine begrenzte Auflösung des Detektors Verschmierungen der Daten. Um diese Detektoreffekte auf die simulierten Daten anzuwenden, wird eine Simulation des CMS-Detektors unter Verwendung des Software-Pakets Geant4 [17] verwendet.

Ziel der Entfaltungsmethode, welche in dieser Arbeit untersucht wird, ist es diese Detektoreffekte der gemessenen Daten mit Hilfe der simulierten Ereignisse zu korrigieren. Damit lassen sich die simulierten Ereignisse mit den gemessenen Daten ohne Verschmierung vergleichen.

## 2.5 Der $t\bar{t} + b\bar{b}$ -Prozess

In dieser Arbeit wird der  $t\bar{t} + b\bar{b}$ -Prozess im semileptonischen Zerfallskanal betrachtet (siehe Abb. 2.2).

Aufgrund der zwei zusätzlichen Bottom-Quarks, welche hauptsächlich aus einem abgestrahlten Gluon beim Zerfall der Top-Quarks entstehen, ist dieser Prozess besonders schwer von dem  $t\bar{t}H(b\bar{b})$ -Prozess (siehe Abb. 2.2) zu unterscheiden. Damit ist der  $t\bar{t} + b\bar{b}$ -Prozess einer der wichtigsten Untergrundprozesse des  $t\bar{t}H(b\bar{b})$ -Prozesses. Für die Untersuchung des  $t\bar{t}H(b\bar{b})$ -Prozesses ist es daher von Nutzen, den  $t\bar{t} + b\bar{b}$  Prozess genau zu verstehen, um ihn als Untergrundprozess möglichst gut von dem eigentlichen Prozess zu unterscheiden. In dieser Arbeit wird daher die  $\Delta R(b\bar{b})$ -Verteilung des  $t\bar{t} + b\bar{b}$ -Prozesses, d. h. der Abstand der beiden  $b$ -Jets in der  $\eta$ - $\varphi$ -Ebene, untersucht.

### 2.5.1 Beschreibung des $t\bar{t} + b\bar{b}$ -Prozesses im semileptonischen Zerfallskanal

Meist entstehen Top-Quark-Antiquark-Paare ( $t\bar{t}$ -Paare) durch Gluon-Fusion oder durch Quark-Antiquark-Vernichtung. Durch die starke Wechselwirkung kann ein  $b\bar{b}$ -Paar erzeugt werden. Tritt dies in Assoziation mit der Produktion eines Top-Quark-Antiquark-Paares auf, so spricht man vom  $t\bar{t} + b\bar{b}$ -Prozess. Die große Masse des Top-Quarks bewirkt einen Zerfall über die schwache Wechselwirkung vor einer möglichen Hadronisierung. Daher können nur die Zerfallsprodukte der Top-Quarks gemessen werden.

Das Top-Quark zerfällt fast ausschließlich in ein Bottom-Quark und ein W-Boson. Im semileptonischen Zerfallskanal des  $t\bar{t}$ -Paares wird der Zerfall eines W-Bosons in ein geladenes Lepton und sein dazugehöriges Neutrino und des anderen W-Bosons in ein Quark-Antiquark-Paar betrachtet.

Neben dem semileptonischen Zerfallskanal gibt es die hadronischen und leptonischen Zerfallskanäle des  $t\bar{t}$ -Systems. Wie durch die Bezeichnung bereits angedeutet wird, zerfallen beide W-Bosonen in diesen Zerfallskanälen jeweils hadronisch beziehungsweise leptonisch.

### 2.5.2 Selektionen

Um möglichst viele Ereignisse zu selektieren, die dem  $t\bar{t} + b\bar{b}$  Prozess im semileptonischen Zerfallskanal entsprechen (und gleichzeitig möglichst wenig Untergrundprozesse) werden Selektionen an den Daten vorgenommen, die den Zerfallsprodukten des zu untersuchenden Prozesses entsprechen. Ausgewählt werden daher Ereignisse mit mindestens sechs Jets, wobei mindestens vier Jets aus einem Bottom-Quark stammen. Des Weiteren werden nur Ereignisse selektiert, die genau ein Elektron oder Myon besitzen.

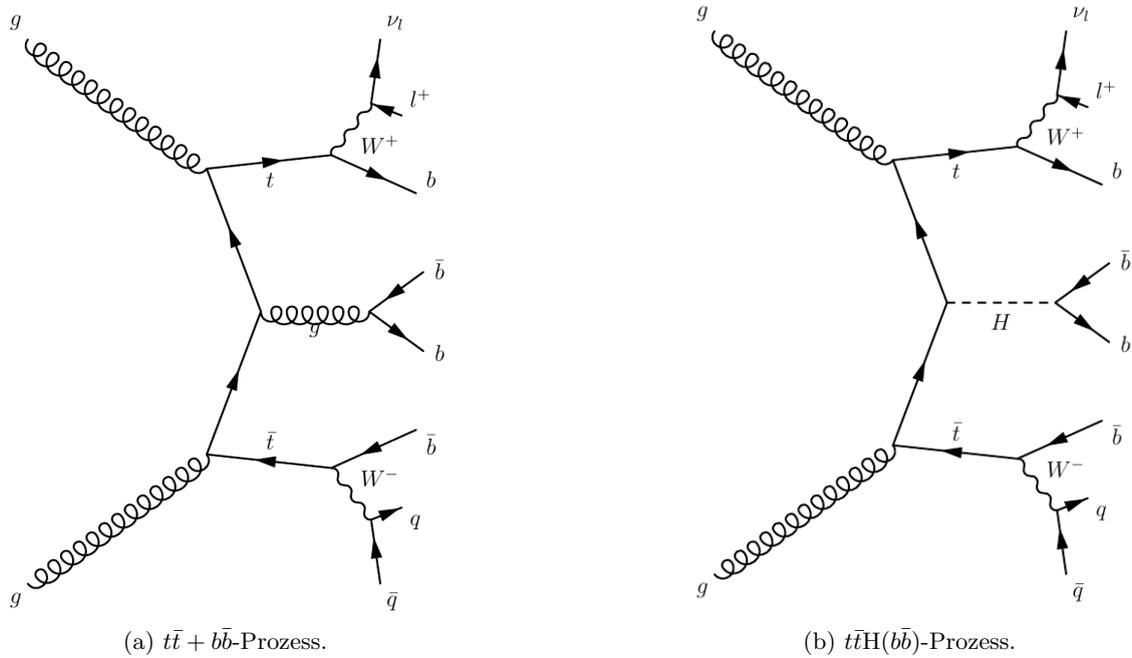


Abbildung 2.2: Beispielhafte Feynmandiagramme für den  $t\bar{t} + b\bar{b}$ -Prozess 2.2a und den  $t\bar{t}H(b\bar{b})$ -Prozess 2.2b im semileptonischen Zerfallskanal.



## 3 Entfaltungsmethode

Um eine Theorie, wie die des Standardmodells der Teilchenphysik, anhand von Daten, welche am LHC mit dem CMS-Detektor gemessen werden, zu überprüfen, ist ein Vergleich der theoretischen Vorhersage mit den gemessenen Daten notwendig.

Dazu werden Ereignisse mit Hilfe der Monte-Carlo-Methode simuliert. Da im CMS-Detektor Akzeptanz- und Effizienz-Effekte auftreten, können die Ereignisse jedoch nicht direkt mit den gemessenen Daten eines Experiments verglichen werden. Um die Daten dennoch vergleichen zu können, wird eine detaillierte Detektorsimulation auf die simulierten Ereignisse der theoretischen Vorhersage angewandt. In der Simulation wird dabei die Verschmierung der gemessenen Daten durch die Detektoreffekte nachempfunden. Damit lässt sich nun die theoretische Vorhersage mit den experimentell gemessenen Daten vergleichen. Man spricht auch von einer Faltung der simulierten Daten mit Detektoreffekten.

Interessant ist es jedoch auch, die Daten ohne die Detektoreffekte zu vergleichen. Dazu muss die Verschmierung der gemessenen Daten durch Detektoreffekte korrigiert werden. Eine Möglichkeit dies zu erreichen, bietet die Entfaltungsmethode. Das Ergebnis hängt nach dem Entfalten nicht mehr von einer detaillierten Detektorsimulation ab und die gemessenen Daten können mit der theoretischen Vorhersagen der simulierten Daten verglichen werden.

In Abschnitt 3.1 wird zunächst die Entfaltung im Allgemeinen vorgestellt. Anschließend wird in Abschnitt 3.2 die Methode des TUnfold-Algorithmus und deren Komponenten beschrieben. Im letzten Abschnitt wird auf die Bedingungen an die Migrationsmatrix, welche für die Entfaltung verwendet wird, eingegangen.

### 3.1 Allgemeine Beschreibung der Entfaltung

Eine Faltung wird im Allgemeinen über einen mathematischen Operator beschrieben. Dieser lässt sich als Produkt der Funktionen  $f$  und  $g$  schreiben, siehe z.B. [18, S. 241]. Dabei ergibt sich die gefaltete Funktion zu

$$h(x) = (f * g)(x). \quad (3.1)$$

Für die Untersuchung der Verteilungen von kinetischen Größen kann die Funktion  $f$  als ursprüngliche Verteilung gedeutet werden. Die Funktion  $g$  beschreibt dabei die Verschmierung durch Detektoreffekte und die Funktion  $h$  die gemessene Verteilung nach der Verschmierung.

Die Entfaltung beschreibt nun einen Prozess, bei dem aus der gemessenen Verteilung  $h$  die ursprüngliche Verteilung  $f$  gewonnen wird.

Dies lässt sich jedoch in der Hochenergiephysik vereinfacht darstellen, da die Experimente üblicherweise als diskretisierte Zählexperimente durchgeführt werden. Ereignisse werden bestimmten Regionen des Phasenraums zugeordnet, welche auch Bins genannt werden. Die Verteilungen sind diskret und können somit als Vektoren dargestellt werden.

Damit lässt sich das Problem der Faltung als Matrixmultiplikation

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad (3.2)$$

darstellen [18, S. 246]. Der Vektor  $\mathbf{x}$  gibt dabei die wahre Verteilung an, welche nicht durch Detektoreffekte verschmiert ist. Der Vektor  $\mathbf{y}$  gibt die gemessene Verteilung der Ereignisse nach der Verschmierung an.

Die Verschmierung des Detektors, welche durch eine begrenzte Auflösung und Effizienzeffekte verursacht wird, sorgt dafür, dass Ereignisse teils dem falschen Bin zugeordnet werden. Für die diskreten Verteilungen lässt sich diese Verschmierung als Matrix  $\mathbf{A}$  darstellen.

Im Folgenden werden Verteilungen, welche durch eine Detektorsimulation oder durch den tatsächlichen Detektor verschmiert sind, als Daten auf dem Detektorlevel bezeichnet. Die ursprünglichen simulierten Daten ohne Verschmierung beziehungsweise die entfalteten Daten aus dem Detektorlevel werden als Daten auf dem Generatorlevel bezeichnet.

Das Problem der Entfaltung, bei der die Verschmierung des Detektors auf die Daten aus dem Detektorlevel korrigiert werden soll, lässt sich nun mathematisch als Inversionsproblem darstellen. Jedoch ist dies ein mathematisch schlecht gestelltes Problem, wodurch die Möglichkeit besteht, dass keine Lösung für das Problem existiert oder die Lösung nicht eindeutig ist.

Zusätzlich muss beachtet werden, dass die Lösung für dieses Inversionsproblem von dem Vektor  $\mathbf{y}$  abhängt, welcher statistischen Fluktuationen unterliegt. Durch eine einfache Matrixinversion ohne zusätzliche Bedingungen könnten diese statistischen Fluktuationen vergrößert werden und negative Korrelationen zwischen benachbarten Bins entstehen [18, S. 242].

Statt einer einfachen Inversion der Matrix lässt sich das Problem zu einer  $\chi^2$ -Minimierung umformulieren. Sie dient zur Anpassung einer Funktion  $g(x, \lambda)$  an eine gemessene Verteilung  $y_i$ . Die Funktion  $g(x, \lambda)$  beschreibt die Faltung der wahren Daten  $\mathbf{x}$  und hängt von den variierbaren Parametern  $\lambda$  ab. Für deren Bestimmung wird die quadratische Abweichung

$$\chi^2(\lambda) = \sum_{i=1}^N \left( \frac{y_i - g(x_i, \lambda)}{\sigma_{y_i}} \right)^2 \quad (3.3)$$

der gemessenen Werte  $y_i$  von der Funktion  $g(x, \lambda)$  berechnet. Die quadratischen Abweichungen werden zusätzlich über die Unsicherheiten  $\sigma_{y_i}$  der gemessenen Werte gewichtet. Somit fließen Werte, welche eine geringere Unsicherheit besitzen, stärker in die Funktion  $\chi^2$  ein. Anschließend wird das Minimum der Funktion  $\chi^2(\lambda)$  bestimmt. Die Parameter  $\lambda$  der Funktion  $g(x, \lambda)$ , werden auf die Werte festgelegt, bei der die Funktion ihr Minimum annimmt.

Durch eine zusätzliche Glättungsbedingung, welche auch als Regularisierung bezeichnet wird, können die statistischen Fluktuationen des Vektors  $\mathbf{y}$  gedämpft werden.

Eine solche  $\chi^2$ -Minimierung mit einer zusätzlichen Glättungsbedingung ist auch in dem Werkzeug TUnfold [19], welches in dieser Arbeit für die Entfaltung der Daten verwendet wird, implementiert.

Das Werkzeug TUnfold kann für Entfaltungen von mehrdimensionalen Verteilungen in der Hochenergiephysik verwendet werden. Der Algorithmus unterstützt die Korrektur von Untergrund- und Verschmierungseffekten. Die zusätzliche Glättung, welche in dem Werkzeug verwendet wird, basiert auf der Tikhonov-Regularisierung, welche in Referenz [20] genauer beschrieben ist. In TUnfold sind verschiedene Methoden integriert, um eine optimale Glättung zu erreichen, ohne eine Überanpassung an die simulierten Daten zu verursachen.

## 3.2 Entfaltung mit Hilfe von TUnfold

Im Folgenden wird der TUnfold-Algorithmus und dessen Methoden basierend auf [19] beschrieben.

### 3.2.1 Der TUnfold-Algorithmus

Der implementierte Entfaltungs-Algorithmus in TUnfold basiert auf der Bestimmung des stationären Punktes der Lagrangefunktion

$$\mathcal{L}(x, \lambda) = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3, \text{ wobei} \quad (3.4)$$

$$\mathcal{L}_1 = (Ax - y)^T V_{yy}^{-1} (Ax - y), \quad (3.5)$$

$$\mathcal{L}_2 = \tau^2 x^T L^T L x \text{ und} \quad (3.6)$$

$$\mathcal{L}_3 = \lambda \sum_i (y - Ax). \quad (3.7)$$

Wie schon im vorherigen Abschnitt beschrieben, geben die Vektoren  $\mathbf{x}$  und  $\mathbf{y}$  die Verteilungen der Daten auf Generator- und Detektorlevel an. Die Matrix  $\mathbf{A}$  stellt die Verschmierung des Detektors dar und enthält die Wahrscheinlichkeiten von Ereignissen in Bins auf dem Generatorlevel zu Ereignissen in Bins auf dem Detektorlevel zu migrieren. Die Dimension der Matrix  $\mathbf{A}$  ist  $n \times m$ . Dabei ist  $n$  die Anzahl der Bins auf Detektorlevel und  $m$  die Anzahl der Bins auf Generatorlevel.

$\mathcal{L}_1$  enthält die Kovarianzmatrix  $\mathbf{V}_{\mathbf{y}\mathbf{y}}$  des Vektors  $\mathbf{y}$ , welche die Quadrate der statistischen Unsicherheiten enthält. Sie ist somit diagonal. Damit entspricht der erste Term  $\mathcal{L}_1$  der schon erwähnten  $\chi^2$ -Minimierung.

Durch eine einfache Matrixinversion können statistische Fluktuationen der gemessenen Daten verstärkt werden [18, S. 242]. Nach der Entfaltung ergeben sich dadurch starke Fluktuationen auf die entfaltete Verteilung  $\mathbf{x}$  sowie negative Korrelationen der benachbarten Bins. Um dies zu verhindern, wird die zusätzliche Glättungsbedingung  $\mathcal{L}_2$ , welche auf der Tikhonov-Regularisierung basiert, im TUnfold-Algorithmus berücksichtigt.

In TUnfold ist eine Regularisierung der zweiten Ableitung der entfalteten Verteilung implementiert. Dabei wird die zweite Ableitung durch die Matrix  $\mathbf{L}$  approximiert. Die Matrix  $\mathbf{L}$  besitzt  $m$  Spalten und  $n - 2$  Reihen, wobei alle Elemente einer Reihe bis auf

$$L_{i,i} = 1, \quad (3.8)$$

$$L_{i,i+1} = -2, \quad (3.9)$$

$$L_{i,i-1} = 1 \quad (3.10)$$

den Wert null haben. TUnfold enthält zusätzlich noch weitere Möglichkeiten für die Regularisierung von mehrdimensionalen Verteilungen. Diese werden bei den Studien dieser Arbeit jedoch nicht betrachtet.

Wie stark diese Glättung ist, wird durch den Regularisierungsparameter  $\tau$  bestimmt. Dieser lässt sich durch verschiedene Methoden, auf die in den folgenden Unterkapiteln noch näher eingegangen wird, bestimmen.

Der dritte Term  $\mathcal{L}_3$  basiert auf einer Flächenbeschränkung mit einem Lagrange-Parameter  $\lambda$ . Da die Anzahl der Ereignisse in einem Bin einer Poissonverteilung statt einer Normalverteilung folgt, kann eine Verzerrung der Normierung der entfalteten Verteilung  $\mathbf{x}$  auftreten. Durch die zusätzliche Flächenbeschränkung wird die Normierung des Ergebnisses  $\mathbf{x}$  korrigiert.

Zur Bestimmung der Verteilung  $\mathbf{x}$  auf Generatorlevel wird der stationäre Punkt der Lagrangefunktion  $\mathcal{L}(\mathbf{x}, \lambda)$  berechnet. Dabei entspricht der stationäre Punkt dem Minimum der Lagrangefunktion. Die Verteilung  $\mathbf{x}$  an diesem Punkt wird als entfaltete Verteilung und damit als Lösung des Entfaltungs-Algorithmus bezeichnet. Als Eingabe in den TUnfold-Algorithmus werden die Migrationsmatrix  $\mathbf{A}$ , die Verteilung  $\mathbf{y}$  auf Detektorlevel und eine passend gewählte Regularisierungsstärke  $\tau$  benötigt. Auf diese Eingabeparameter wird in den folgenden Unterkapiteln näher eingegangen.

### 3.2.2 Wahl des Regularisierungsparameters $\tau$ in TUnfold

Wie bereits in dem vorherigen Abschnitt beschrieben wird in dem TUnfold-Algorithmus die Tikhonov-Regularisierung für die Glättung der entfalteten Verteilung  $\mathbf{x}$  verwendet.

Die Stärke der Glättung wird über den Regularisierungsparameter  $\tau$  bestimmt. Dieser spielt eine wichtige Rolle für ein sinnvolles Ergebnis der Entfaltung. Sowohl ein zu großer als auch ein zu kleiner Wert für  $\tau$  kann das Ergebnis der Entfaltung negativ beeinflussen. In den Studien zur Entfaltung wird genauer untersucht, welche Auswirkungen die Wahl des Regularisierungsparameters hat. Dabei werden die Folgen einer schlechten Wahl des Parameters deutlich.

Ein zu kleiner Wert kann große statistische Fluktuationen der entfalteten Daten sowie negative Korrelationen der benachbarten Bins verursachen. Ein zu großer Wert kann jedoch eine zu starke Anpassung an die theoretisch vorhergesagte Verteilung  $\mathbf{x}$  zur Folge haben.

Um eine geeignete Regularisierungsstärke für die Entfaltung zu wählen, gibt es verschiedene Methoden. Der TUnfold-Algorithmus beinhaltet eine einfache Version der L-Curve-Methode [21] sowie die Methode der Minimierung der globalen Korrelation [19]. Jedoch liefert die L-Curve-Methode bei den Anwendungsversuchen keine zuverlässigen Ergebnisse. Für die Untersuchung des TUnfold-Algorithmus an der  $\Delta R(bb)$ -Verteilung des  $t\bar{t} + b\bar{b}$ -Prozesses wird daher die zweite Methode verwendet, welche im Folgenden näher beschrieben wird. Weitere Methoden, welche in TUnfold implementiert sind, können in Referenz [20] nachgelesen werden.

#### Minimierung der globalen Korrelation

Die durchschnittliche globale Korrelation der entfalteten Verteilung ist ein Maß für die Korrelation aller Bins untereinander. Diese wird zur Untersuchung der Regularisierungsstärke verwendet. Sie ist definiert durch

$$\bar{\rho} = \sum_i \frac{\rho_i}{n}, \text{ wobei} \quad (3.11)$$

$$\rho_i = \sqrt{1 - \frac{1}{(\mathbf{V}_{\mathbf{xx}})^{-1}(\mathbf{V}_{\mathbf{xx}})_{ii}}}, \quad (3.12)$$

$n$  die Anzahl der Bins und  $\mathbf{V}_{\mathbf{xx}}$  die Kovarianzmatrix des Vektors  $\mathbf{x}$  ist. Zur Untersuchung der Regulierungsstärke  $\tau$  wird für verschiedene Werte von  $\tau$  die Entfaltung durchgeführt und jeweils die durchschnittliche globale Korrelation berechnet. Für die Regulierungsstärke tau wird der Wert verwendet, welcher die durchschnittlichen globalen Korrelation minimiert.

### 3.2.3 Migrationsmatrix als Eingabe für den Entfaltungs-Algorithmus

Für den Entfaltungs-Algorithmus werden die Migrationsmatrix  $\mathbf{A}$  und die zu entfaltende Verteilung  $\mathbf{y}$  benötigt. Die Migrationsmatrix  $\mathbf{A}$  wird aus simulierten Verteilungen mit Hilfe der Monte-Carlo-Methode bestimmt. Dazu werden Daten auf Generatorlevel erzeugt. Durch eine präzise Detektorsimulation werden die dazugehörigen verschmierten Daten auf Detektorlevel berechnet. Diese Daten werden anschließend in eine Matrix  $\mathbf{M}$  eingelesen. Aus dieser Matrix lassen sich die Übergangswahrscheinlichkeit, dass Ereignisse eines bestimmten Bins aus dem Generatorlevel in einem Bin aus dem Detektorlevel gemessen werden, berechnen. Diese Matrix  $\mathbf{M}$ , welche die Anzahl der migrierenden Ereignisse enthält, wird zur Initialisierung des TUnfold-Algorithmus verwendet.

Durch die Detektoreffekte kann es vorkommen, dass Ereignisse die Detektorselektion erfüllen, jedoch nicht die Generatorselektion. Diese Ereignisse werden „Fakes“ genannt, da sie nur durch die Detektoreffekte fälschlicherweise als Ereignisse auf Detektorlevel selektiert werden. Es kann jedoch ebenso vorkommen, dass Ereignisse die Generatorselektion erfüllen, jedoch nicht die Detektorselektion. Diese Ereignisse werden „Misses“ genannt und fehlen somit in der Verteilung auf Detektorlevel. Sowohl die Fakes als auch die Misses werden in der Matrix  $\mathbf{M}$  berücksichtigt. Dazu wird eine zusätzliche Spalte und eine zusätzliche Zeile mit Index  $i, j = 0$  benötigt.

Im Matricelement  $\mathbf{M}_{ij}$  werden die Ereignisse gezählt, welche in dem Bin  $j > 0$  auf Generatorlevel erzeugt werden und auf Detektorlevel in dem Bin  $i > 0$  detektiert werden. Die Fakes, welche die Generatorselektion nicht erfüllen und zu den Bins  $i > 0$  auf Detektorlevel gehören, werden in die dafür dedizierten Einträge  $\mathbf{M}_{i0}$  gezählt. Die Misses, welche in den Bins  $j > 0$  auf Generatorlevel erzeugt werden, werden in die Einträge  $\mathbf{M}_{0j}$  gezählt.

Die Matrix  $\mathbf{M}$  besitzt damit die Dimension  $(n + 1 \times m + 1)$ . Dabei ist  $m$  die Anzahl der Bins der wahren Verteilung auf Generatorlevel und  $n$  die Anzahl der Bins auf Detektorlevel. Damit der TUnfold-Algorithmus ein gutes Ergebnis liefert, darf die Anzahl der Bins auf Generatorlevel nicht größer sein als die der Bins auf Detektorlevel.

Für den Entfaltungsalgorithmus werden die Migrationsmatrix  $\mathbf{A}$  und die Vektoren  $\mathbf{x}_0$  und  $\mathbf{y}_0$ , welche die erwarteten Verteilungen der Ereignisse auf dem Generatorlevel und Detektorlevel beschreiben, aus der Eingabematrix  $\mathbf{M}$  mit Ereignissen aus der Simulation durch

$$\mathbf{A}_{ij} = \frac{\mathbf{M}_{ij}}{\mathbf{g}_j}, \text{ für } i, j > 0 \text{ und} \quad (3.13)$$

$$\mathbf{g}_j = \sum_{i=0}^m \mathbf{M}_{ij}, \quad (3.14)$$

$$(3.15)$$

berechnet. Die Misses werden somit über die Normierung der Migrationsmatrix berücksichtigt, indem die Einträge  $\mathbf{M}_{i0}$  der Migrationsmatrix bei der Berechnung des Gewichtungsfaktors  $g_j$  mit dazu gezählt wird.

Die Fakes werden für die Matrix  $\mathbf{A}$  nicht beachtet. Oft ist es sinnvoll diese als Untergrund bereits vor dem Einlesen in den TUnfold-Algorithmus von der Verteilung auf Detektorlevel abzuziehen. Der Algorithmus ermöglicht jedoch ebenfalls, die Untergrundereignisse von den Daten auf Detektorlevel vor der eigentlichen Entfaltung abzuziehen. Die Verteilung

$$\mathbf{y} = \mathbf{y}_0 - f^b \mathbf{b} \text{ mit der Kovarianzmatrix} \quad (3.16)$$

$$(\mathbf{V}_{\mathbf{y}\mathbf{y}})_{ij} = (\mathbf{V}_{\mathbf{y}\mathbf{y}})_{ij}^0 + \delta_{ij}(f^b(\delta \mathbf{b}_i))^2 + (\delta f^b)^2 \mathbf{b}_i \mathbf{b}_j \quad (3.17)$$

wird anschließend für den Entfaltungsalgorithmus verwendet. Dabei beschreibt der Vektor  $\mathbf{y}_0$  die Verteilung auf Generatorlevel vor dem Abzug der Untergrundereignisse und  $\mathbf{V}_{\mathbf{y}\mathbf{y}}^0$  die Kovarianzmatrix des Vektors. Der Vektor  $\mathbf{b}$  mit der Unsicherheit  $\delta \mathbf{b}$  gibt zusammen mit dem Normierungsfaktor  $f^b$  und der Unsicherheit  $\delta f^b$  die Verteilung des Untergrunds an.

### 3.2.4 Statistische und systematischen Unsicherheiten der Migrationsmatrix $\mathbf{A}$

Die Migrationsmatrix  $\mathbf{A}$  enthält üblicherweise Unsicherheiten, die durch verschiedene Quellen verursacht werden. Dabei enthält sie sowohl statistische Unsicherheiten, als auch systematische Unsicherheiten.

Die statistischen Unsicherheiten werden durch die Monte-Carlo-Methode verursacht. Die simulierten Ereignisse werden in die Matrix  $\mathbf{M}$  gezählt. Dadurch ergeben sich statistische Unsicherheiten  $\delta \mathbf{M}_{ij}$  auf die einzelnen Bins, welche voneinander unabhängig sind. Durch den Entfaltungs-Algorithmus resultieren diese statistischen Unsicherheiten in der Kovarianzmatrix  $\mathbf{V}_{\mathbf{x}\mathbf{x}}^{\mathbf{M},\text{stat}}$  der entfalteten Verteilung  $\mathbf{x}$ .

Zusätzlich lassen sich durch eine Variation der Matrix  $\mathbf{M} \rightarrow \mathbf{M} + \delta \mathbf{M}$  systematische Unsicherheiten untersuchen. Diese systematische Variation  $\delta \mathbf{M}$  wird durch Variation von experimentellen Bedingungen erzeugt. Sie erzeugt dabei eine systematische Verschiebung der entfalteten Verteilung  $\delta \mathbf{x}$ . Die dazugehörige Kovarianzmatrix lässt sich durch

$$\mathbf{V}_{\mathbf{x}\mathbf{x}}^{\mathbf{M},\text{sys}} = \delta \mathbf{x}(\delta \mathbf{x})^T \quad (3.18)$$

berechnen.

### 3.3 Bedingungen für die Migrationsmatrix $M$

Für ein gutes Ergebnis des Entfaltungsprozesses spielt es eine wichtige Rolle, wie fein die Einteilung der Bins („Binning“) der Verteilungen auf Generator- und Detektorlevel gewählt wird. Wird dieses Binning zu fein gewählt, kann dies große statistische Fluktuationen auf die entfalteten Daten sowie hohe Korrelation von benachbarten Bins verursachen. Daher ist es wichtig, geeignete Bedingungen für das Binning zu finden.

Für die statistische Aussagekraft ist es sinnvoll, ein Minimum an Ereignissen innerhalb eines Bins festzulegen. Zusätzlich werden Bedingungen an die Reinheit und an die Stabilität der Verteilung gestellt.

#### 3.3.1 Reinheit und Stabilität

Die Reinheit  $\mathbf{p}$  und die Stabilität  $\mathbf{s}$  der Daten wird durch die Diagonaleinträge der Migrationsmatrix und den Verteilungen auf Generator- und Detektorlevel über

$$\mathbf{p}_i = \frac{M_{ii}}{y_i^0} \quad (3.19)$$

$$\mathbf{s}_j = \frac{M_{jj}}{x_j^0} \quad (3.20)$$

berechnet. Die Reinheit  $\mathbf{p}$  gibt den Anteil der Ereignisse eines Bins auf Detektorlevel an, der von dem selben Bin auf Generatorlevel stammt. Die Stabilität  $\mathbf{s}$  gibt die Wahrscheinlichkeit an, dass ein Ereignis, welches in einem Bin auf Generatorlevel erzeugt wird auch in diesen Bin auf Detektorlevel zugeordnet wird. Dabei werden die Verteilungen ohne die Fakes und Misses verwendet.

Eine hohe Anforderung an Reinheit und Stabilität haben zur Folge, dass die Auflösung der Verteilungen geringer wird. Eine zu geringe Anforderung jedoch bewirkt eine große Fluktuation der entfalteten Daten. Für optimale Ergebnisse der Entfaltung sollte daher eine passende Anforderung an Reinheit und Stabilität gefunden werden. In Kapitel 4.1 werden unterschiedliche Anforderungen getestet und untersucht, wie sich die Entfaltung auswirkt.



## 4 Studium der Entfaltung

In diesem Kapitel werden die Studien zur Entfaltung der  $\Delta R(b\bar{b})$ -Verteilung des  $t\bar{t} + b\bar{b}$ -Prozesses mit TUnfold beschrieben.

Als Eingangsparameter benötigt der TUnfold-Algorithmus eine Matrix  $\mathbf{M}$ , welche die simulierten Daten auf Generator- und Detektorlevel enthält. Dazu werden die Daten in verschiedene Bereiche unterteilt. Diese Einteilung wird auch Binning genannt. In Abschnitt 4.1 wird das Binning beschrieben, welches durch Bedingungen an die Anzahl der Ereignisse pro Bin und an die Werte der Reinheit und der Stabilität der Daten (siehe Abschnitt 3.3.1) erstellt wird.

Zusätzlich zur Eingangsmatrix  $\mathbf{M}$  werden Pseudodaten benötigt, welche durch den TUnfold-Algorithmus entfaltet werden. Für die Studien in dieser Arbeit werden poissonverschmierte Pseudodaten verwendet. Die Erzeugung der poissonverschmierten Pseudodaten wird in Abschnitt 4.2 beschrieben.

Anschließend werden verschiedene Studien zur Entfaltung mit dem TUnfold-Algorithmus durchgeführt. In Abschnitt 4.3 wird untersucht, wie sich die Entfaltung auf verschiedene Anforderungen an die Reinheit und an die Stabilität auswirkt.

Für die Studien zur Auswirkung der Variation der Pseudodaten auf die Entfaltung spielt die Regularisierung ebenfalls eine wichtige Rolle. Dazu wird in Abschnitt 4.4 erforscht, ob der TUnfold-Algorithmus eine Vergrößerung der statistischen Fluktuation ausgleichen kann.

Die Auswirkung des TUnfold-Algorithmus auf die skalierten Pseudodaten wird ebenfalls in Abschnitt 4.4 untersucht.

Wie schon in Abschnitt 3.2.4 beschrieben, lassen sich auch systematische Unsicherheiten mit Hilfe des TUnfold-Algorithmus untersuchen. In Abschnitt 4.5 werden die Studien zur Entfaltung mit systematischen Unsicherheiten diskutiert.

### 4.1 Binning

Der Entfaltungs-Algorithmus von TUnfold benötigt eine Eingangsmatrix, welche die simulierten Ereignisse auf Generator- und Detektorlevel enthält. Für den Entfaltungs-Algorithmus wird daraus, wie in Abschnitt 3.2.3 beschrieben, die Migrationsmatrix berechnet. Die Breite der einzelnen Bins auf Generator- und Detektorlevel kann auf unterschiedliche Weise festgelegt werden. In dieser Arbeit werden verschiedene Bedingungen an die Reinheit und an die Stabilität (siehe Abschnitt 3.3.1) der Daten getestet. Diese erzeugen

jeweils ein unterschiedliches Binning der Daten und damit auch eine unterschiedliche Eingangsmatrix  $\mathbf{M}$  für die Entfaltung.

Wie schon in Abschnitt 3.2.3 erwähnt, wird auch eine Anforderung an die Anzahl der Ereignisse pro Bin sowohl auf Generator- als auch auf Detektorlevel gestellt. Dabei werden die Ereignisse aus der Monte-Carlo-Simulation mit Gewichtungen verwendet, welche sowohl die Selektionskriterien auf Generator- sowie auf Detektorlevel erfüllen. Dies sorgt für statistisch aussagekräftige Ergebnisse. Für die Studien in dieser Arbeit wird eine Mindestanzahl an 30 Ereignissen pro Bin gefordert. Diese Anforderung gilt für alle Matrizen, welche für die Entfaltungen verwendet werden.

Um diese Anforderung zu erfüllen, werden zu Beginn die simulierten Ereignisse auf Generator- und Detektorlevel mit einem feinen Binning im Bereich  $\Delta R$  in  $[0; 4,0]$  mit gleichgroßen Bins der Breite 0,1 histogrammiert. Anschließend wird gefordert, dass mindestens 30 Ereignisse in einem Bin sowohl auf Generatorlevel als auch auf Detektorlevel enthalten sind. Dabei werden die gewichteten Ereignisse für die Berechnung der Anzahl der Ereignisse verwendet<sup>1</sup>. Aufgrund der höheren Anzahl der Ereignisse für kleinere Werte für  $\Delta R$  wird zu Beginn der kleinste  $\Delta R$ -Werte betrachtet. Ist die Anzahl der Ereignisse in diesem Bin kleiner als 30, wird der nächste Bin mit diesem zusammengefügt. Dadurch steigt die Anzahl der Ereignisse. Es werden solange Bins zusammengefügt, bis die Anforderung an die Anzahl der Bins erfüllt ist. Danach wird der Bin mit dem nächstgrößeren  $\Delta R$ -Wertebereich betrachtet, welcher noch nicht zu dem letzten Bin hinzugefügt wurde. Das Prozedere wird wiederholt bis in allen Bins die Anforderung erfüllt ist.

Für die Anforderung an die Reinheit und an die Stabilität wird ein ähnliches Schema verwendet. Jedoch muss hier zusätzlich eine Matrix  $\mathbf{M}$  mit den simulierten Ereignissen gefüllt werden. Die  $\Delta R$ -Werte auf Generatorlevel werden auf der  $y$ -Achse gegen die  $\Delta R$ -Werte auf Detektorlevel auf der  $x$ -Achse aufgetragen. Das Binning der beiden Verteilungen ist dabei symmetrisch. Durch die Verschmierung des Detektors kann es vorkommen, dass Ereignisse nur die Detektorselektion oder nur die Generatorselektion erfüllen. Wie schon in Kapitel 3.2.3 beschrieben, werden diese Ereignisse ebenfalls in die dafür dedizierten Einträge der Matrix  $\mathbf{M}$  gezählt.

Nun werden, wie bei der Anforderung an die Anzahl der Ereignisse, die Bins zusammengefasst, bis die Bedingung an die Reinheit und an die Stabilität der Daten erfüllt ist. Statt das Binning im zweidimensionalen zu betrachten, wird dabei das Binning der eindimensionalen Verteilung verwendet. Nach jedem Zusammenfassen der eindimensionalen Bins muss hierfür eine neue Matrix  $\mathbf{M}$  zur Berechnung der Reinheit und der Stabilität (siehe Abschnitt 3.3.1) generiert werden.

Für den letzten Bin besteht jedoch nicht die Möglichkeit, noch weitere Bins zusammenzufassen. Dies hat zur Folge, dass der Wert für die Reinheit und für die Stabilität im letzten Bin die Anforderung nicht notwendigerweise erfüllt. Bei starken Abweichungen an die geforderten Werte wird der letzte Bin mit dem vorletzten Bin nachträglich zusammengefasst. Die Matrix, welche nun die Anforderungen erfüllt, wird als Eingangsmatrix  $\mathbf{M}$  für den Entfaltungs-Algorithmus verwendet.

## 4.2 Erzeugung der Pseudodaten für die Entfaltung

Zur Untersuchung des TUnfold-Algorithmus werden unterschiedliche Pseudodaten verwendet, welche ebenfalls als Eingabe für den Entfaltung-Algorithmus dienen. Die Entfaltung

<sup>1</sup>In der Arbeit waren alle Gewichte  $< 1$ . Für den Fall, dass es Ereignisse mit Gewichten  $> 1$  gibt, müsste diese Prozedur gegebenenfalls angepasst werden.

wird für verschiedene Anforderung an Reinheit und Stabilität getestet. Hierfür werden poissonverschmierte Zufallsdaten als Pseudodaten für die Entfaltung verwendet. Zusätzlich wird untersucht, wie sich TUnfold auf vergrößerte statistische Schwankungen auswirkt. Daher werden zusätzlich Pseudodaten erzeugt, welche eine vergrößerte statistische Schwankung besitzen. Ebenfalls wird getestet, wie sich ein Faktor, mit dem die Pseudodaten skaliert werden, auf das Ergebnis der Entfaltung auswirkt. Dies dient der Untersuchung eines größer gemessenen Wirkungsquerschnitts mit Hilfe einer Umgewichtung der Ereignisse.

#### 4.2.1 Erzeugung von poissonverschmierten Pseudozahlen

Zählexperimente besitzen üblicherweise eine Binomialverteilung

$$B(k|p, n) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (4.1)$$

Wird die Anzahl der Ereignisse sehr groß und die Wahrscheinlichkeit  $p$  klein, geht die Verteilung in eine Poissonverteilung

$$P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (4.2)$$

über. Dabei ist  $\lambda$  der Erwartungswert der Verteilung.

Eine solche Poissonstatistik wird auch für die Erzeugung von Pseudodaten zur Untersuchung der Entfaltung auf die  $\Delta R$ -Verteilung des  $t\bar{t} + b\bar{b}$ -Prozesses verwendet. Dazu wird die Verteilung auf dem Detektorlevel  $\mathbf{y}$  verwendet. Die Anzahl der Ereignisse in einem Bin  $\mathbf{y}_i$  wird bei der Erzeugung einer poissonverschmierten Zufallszahl als Erwartungswert  $\lambda_i = \mathbf{y}_i$  verwendet. Für die Unsicherheit der Pseudodaten ergibt sich aus der Standardabweichung der Poissonverteilung

$$\sigma_{\mathbf{y}_i} = \sqrt{\lambda} = \sqrt{\mathbf{y}_i}. \quad (4.3)$$

Die Pseudodaten werden für alle Entfaltungen zur Studie des Binnings der Migrationsmatrizen einzeln erzeugt, da sich für jedes Binning eine andere Verteilung von  $\Delta R$  auf dem Detektorlevel ergibt.

#### 4.2.2 Erzeugung von Pseudodaten mit vergrößerten statistischen Schwankungen

Zur Studie der Regularisierung des TUnfold-Algorithmus werden zusätzlich poissonverschmierte Pseudozahlen erzeugt, welche vergrößerte statistische Schwankungen besitzen. Dies dient zur Untersuchung, ob die Regularisierung diese statistischen Schwankungen ausgleichen kann. Dazu werden, wie in Abschnitt 4.2.1 beschrieben, poissonverschmierte Zufallsdaten  $\mathbf{z}_i$  aus der theoretischen Verteilung auf dem Detektorlevel  $\mathbf{y}_i$  erzeugt.

Für die Vergrößerung der statistischen Unsicherheit wird anschließend die Differenz der erzeugten Zufallszahl zur Anzahl der Ereignisse in einem Bin um einen Faktoren  $f$  vergrößert. Damit ergeben sich die Pseudodaten

$$\mathbf{p} = (\mathbf{z} - \mathbf{y}) \cdot f + \mathbf{y}. \quad (4.4)$$

Für die statistische Unsicherheit, welche ebenfalls vergrößert wird, gilt anschließend

$$\sigma_{\mathbf{y}_i} = \sqrt{\lambda_i} = \sqrt{\mathbf{y}_i} \cdot f. \quad (4.5)$$

### 4.2.3 Erzeugung von skalierten Pseudodaten

In dieser Arbeit wird die Auswirkung eines größer gemessenen Wirkungsquerschnitts des  $t\bar{t} + b\bar{b}$ -Prozesses, als durch die Simulation vorhergesagt wird, auf die Entfaltung mit TUnfold untersucht. Dazu werden skalierte Pseudodaten verwendet. Diese werden erzeugt, indem statt der theoretischen Verteilung auf Generatorlevel  $\mathbf{y}$  eine um einen Faktor  $g$  skalierte Verteilung

$$\mathbf{y}' = \mathbf{y} \cdot g \quad (4.6)$$

für den Erwartungswert  $\lambda$  verwendet wird. Die statistische Unsicherheit ergibt sich ebenfalls aus der vergrößerten Verteilung zu

$$\sigma_{\mathbf{y}_i} = \sqrt{\lambda_i} = \sqrt{\mathbf{y}_i \cdot g}. \quad (4.7)$$

## 4.3 Studium zur Variation der Migrationsmatrix durch verschiedene Bedingungen an Reinheit und Stabilität

Für ein aussagekräftiges Ergebnis der Entfaltung spielt das Binning eine wichtige Rolle. Dabei werden die Ereignisse auf Generator- und Detektorlevel, wie in Abschnitt 4.1 beschrieben, in Bereiche unterteilt. Diese Bereiche werden durch Bedingungen an die Reinheit und an die Stabilität der Daten bestimmt.

In dieser Arbeit wird die Auswirkung der Anforderungen an die Reinheit und an die Stabilität auf die Entfaltung untersucht. Dazu werden unterschiedliche Werte als Anforderung getestet und die Entfaltung mit den daraus resultierenden Eingangsmatrizen durchgeführt.

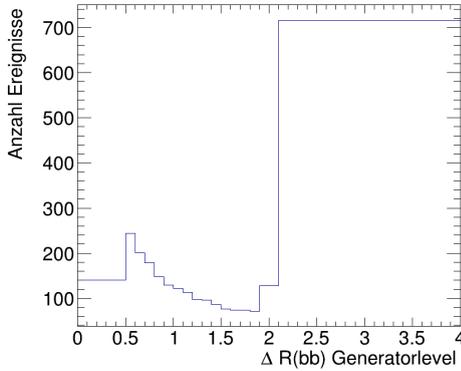
Wie bereits in Abschnitt 4.1 erwähnt, wird auch eine Anforderung von mindestens 30 Ereignissen pro Bin gestellt. Die Verteilung, welche sich daraus ergibt, wird in Abschnitt 4.3.1 diskutiert.

Für die Reinheit und Stabilität werden im Folgenden eine Mindestgröße von 0.1, 0.3, 0.4 und 0.5 getestet und in Abschnitt 4.3.2 – 4.3.5 ausgewertet. Wie schon in Abschnitt 4.2 beschrieben, werden poissonverschmierte Pseudodaten für die Entfaltung bei dieser Studie verwendet. Zur Bestimmung des Regularisierungsparameters wird die Minimierung der globalen Korrelation (siehe Abschnitt 3.2.2) verwendet.

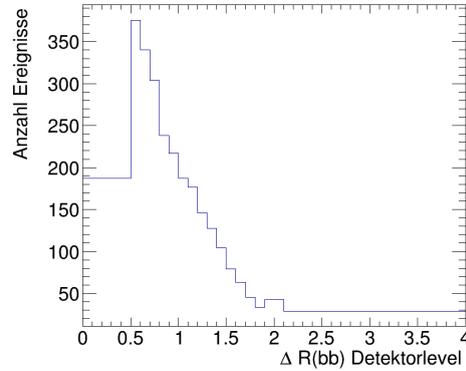
### 4.3.1 Binning mit mindestens 30 Ereignissen pro Bin

Für ein statistisch aussagekräftiges Ergebnis wird zu Beginn ein Binning erstellt, welches die Anforderung von 30 Ereignissen pro Bin erfüllt. Die resultierenden Verteilungen von  $\Delta R$  des  $t\bar{t} + b\bar{b}$ -Prozesses sind in Abbildung 4.1 zu sehen. Dabei werden in dieser Verteilung nur Ereignisse betrachtet, welche sowohl die Generator- und Detektorselektion erfüllen.

Diese Verteilungen weichen besonders für den letzten Bin stark voneinander ab. Auf dem Detektorlevel wird das kleinste  $\Delta R$  der Jets, welche aus dem Zerfall von Bottom-Quarks stammen, aufgetragen. Es kann vorkommen, dass fälschlicherweise Jets aus dem Zerfall der Top-Quarks zur Berechnung von  $\Delta R$  verwendet werden. Dadurch wird für viele Ereignissen mit hohen Werten für  $\Delta R$  auf dem Generatorlevel ein zu kleiner Wert für  $\Delta R$  auf dem Detektorlevel zugeordnet. Durch die falsche Zuweisung der Ereignisse auf dem Detektorlevel für große Werte für  $\Delta R$  auf dem Generatorlevel, erhöht sich die Anzahl der Ereignisse in den restlichen Bins.



(a) Verteilung auf Generatorlevel.



(b) Verteilung auf Detektorlevel.

Abbildung 4.1: Die Verteilungen der simulierten Ereignisse auf Generator- (4.1a) und Detektorlevel (4.1b) sind über  $\Delta R$  aufgetragen. Das Binning wird mit der Anforderung an 30 Ereignissen pro Bin erstellt. Dazu werden Ereignisse verwendet, welche die Selektionskriterien für das Generator- sowie für das Detektorlevel erfüllen.

### 4.3.2 Reinheit und Stabilität $\geq 0.1$

Für das Binning wird nun, zusätzlich zur Anforderung an die Anzahl der Ereignisse pro Bin, eine Bedingung an die Reinheit und an die Stabilität gestellt. In diesem Abschnitt wird ein Wert von  $\geq 0.1$  gefordert.

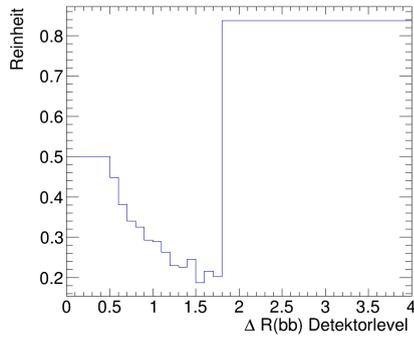
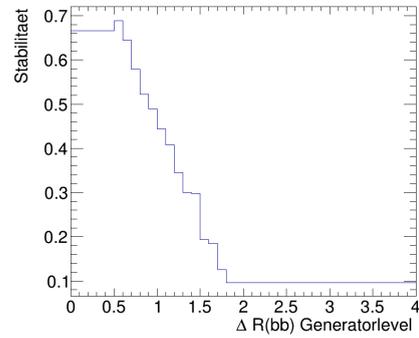
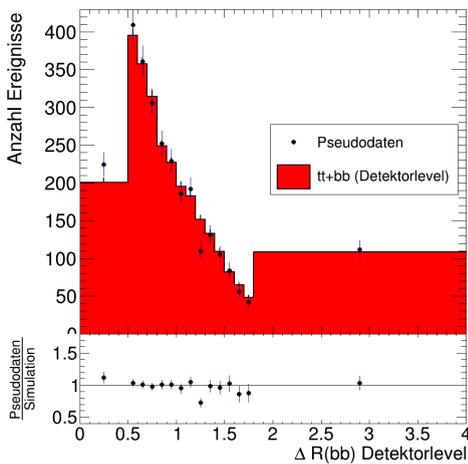
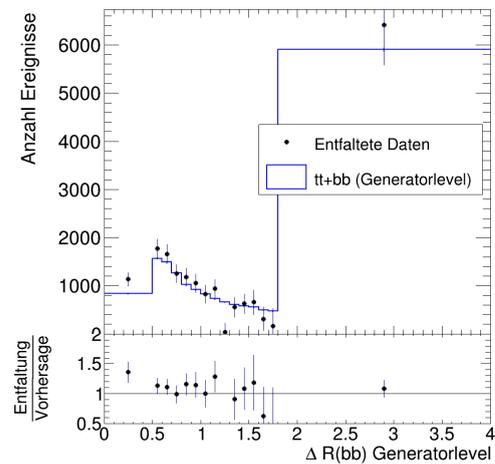
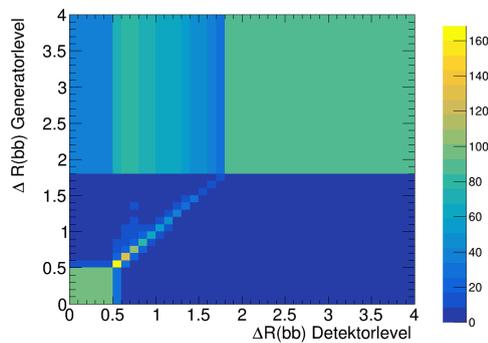
Die Detektorverschmierung für die großen Werte von  $\Delta R$  auf Generatorlevel bewirkt eine kleine Stabilität der Werte in diesem Bereich. Für die Anforderung an die Stabilität werden daher in diesem Bereich einige Bins zusammengefasst.

Das Binning ist in den mittleren Bereichen von  $\Delta R$  jedoch immer noch sehr fein. Für kleine Werte von  $\Delta R$  ist die Stabilität und die Reinheit relativ groß (siehe Abb. 4.2a und 4.2b). Die Stabilität nimmt für größere Werte stetig ab. Die Reinheit wird für größerer Werte ebenfalls kleiner. Da nur sehr wenige Ereignisse mit hohen  $\Delta R$ -Werten auf Detektorlevel niedrige  $\Delta R$ -Werte auf Generatorlevel besitzen, ist die Reinheit des letzten Bins sehr groß.

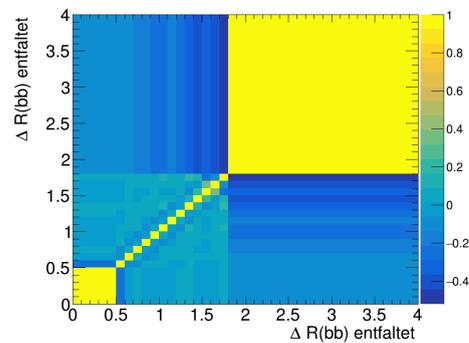
Die Matrix, welche in Abbildung 4.2e dargestellt ist, zeigt die Verschmierung des Detektors deutlich. Für den letzte Bin auf dem Generatorlevel führt die Verschmierung zu großen Werten auf den Nebendiagonalelementen. Dies sorgt für eine geringe Stabilität der Daten für große Werte von  $\Delta R$ .

Die erzeugten poissonverschmierte Pseudodaten sind in Abbildung 4.2c zusammen mit der Verteilung von  $\Delta R$  auf Detektorlevel aufgetragen. In Abbildung 4.2d können die entfalteten Daten mit der Verteilung auf Generatorlevel verglichen werden. Durch die Minimierung der globalen Korrelation ergibt sich für den Regularisierungsparameters ein Wert von  $\tau = 1 \cdot 10^{-5}$ .

Es ist ein deutlicher Zusammenhang zwischen der statistischen Unsicherheiten und dem Wert für die Reinheit der Daten zu erkennen. Bei Abnahme der Reinheit für größer werdende Werte für  $\Delta R$  nimmt die Korrelation der entfalteten Daten zu. Dies ist anhand der negative Korrelation der übernächsten Nachbarn der entfalteten Daten (siehe Abb. 4.2f) zu sehen. Besonders der letzte Bin weist eine starke negative Korrelation zu den anderen Bins auf. Die größer werdende Korrelation verursacht eine Vergrößerung der statistischen Unsicherheit.

(a) Reinheit der Daten über  $\Delta R$  auf Detektorlevel aufgetragen.(b) Stabilität der Daten über  $\Delta R$  auf Generatorlevel aufgetragen.(c) Simulierte Verteilung von  $\Delta R$  auf dem Detektorlevel zusammen mit den poissonverschmierten Pseudodaten über  $\Delta R$  aufgetragen.(d) Entfaltete Daten zusammen mit der simulierten Verteilung von  $\Delta R$  auf dem Generatorlevel über  $\Delta R$  aufgetragen.

(e) Eingangsmatrix für die Entfaltung.



(f) Korrelationsmatrix der entfalteten Daten.

Abbildung 4.2: In dieser Abbildung sind die Verteilungen, welche mit einem Binning durch die Anforderung von 0.1 an die Reinheit und an die Stabilität der Daten erzeugt wurden, zusammen mit den entfalteten Ergebnissen zu sehen.

### 4.3.3 Reinheit und Stabilität $\geq 0.3$

Nun wird ein Mindestwert von 0.3 für die Reinheit und Stabilität gefordert. Wie in Abbildung 4.3a und 4.3b zu sehen ist, werden dadurch für höhere Werte von  $\Delta R$  weitere Bins zusammengefasst. Die Migrationsmatrix dieses Binnings ist in Abbildung 4.3e zu sehen.

Durch die Minimierung der globalen Korrelation ergibt sich ein Wert von  $\tau = 3 \cdot 10^{-6}$  für den Regulierungsparameter. Die Pseudodaten und die entfalteten Daten sind in der Abbildung 4.2c und 4.3d zusammen mit den simulierten Verteilungen auf Generator- und Detektorlevel aufgetragen. Dabei ist im Vergleich zu der Abbildung 4.2d zu sehen, dass die statistische Unsicherheit der Daten für eine höhere Anforderung an die Reinheit und an die Stabilität der Daten abnimmt. Auch die Fluktuation der entfalteten Daten nimmt im Vergleich zu dem Ergebnis der vorherigen Entfaltung ab. Die Bins korrelieren jedoch weiterhin negativ mit den übernächsten Nachbarn (siehe Abb. 4.3f). Auch der letzte Bin weist weiterhin eine starke negative Korrelation mit seinen näheren Nachbarn auf.

### 4.3.4 Reinheit und Stabilität $\geq 0.4$

Die Anforderung an die Reinheit und an die Stabilität wird auf einen Wert von  $\geq 0.4$  gesetzt. Dadurch nimmt die Anzahl der Bins weiter deutlich ab (siehe Abb. 4.4a und 4.4b). Die Minimierung der Globalen Korrelation liefert für dieses Binning einen Wert von  $\tau = 8 \cdot 10^{-6}$  für den Regularisierungsparameter.

Die Pseudodaten und die entfalteten Daten sind in Abbildung 4.4c und 4.4d mit den simulierten Verteilungen aufgetragen. Auch hier nimmt die statistische Unsicherheit der Daten weiter ab. Die entfalteten Daten liegen hier sehr nah an der theoretischen Verteilung. Der letzte Bin weist dennoch eine deutliche negative Korrelation zu seinen Nachbarbins auf (siehe Abb. 4.4f). Jedoch besitzen die mittleren Bins keine starke negative Korrelation zu ihren übernächsten Nachbarn mehr.

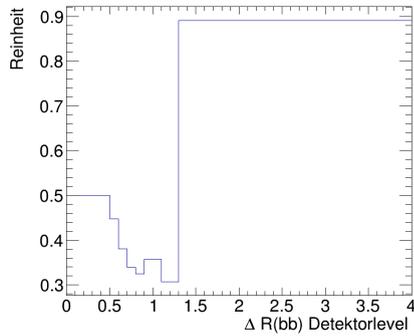
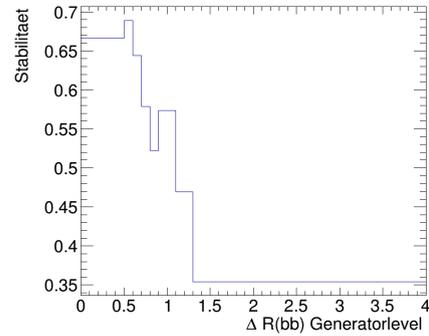
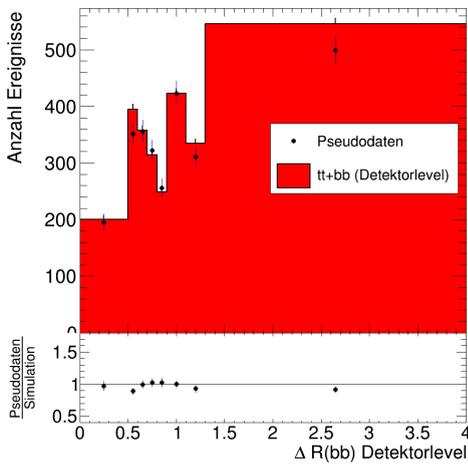
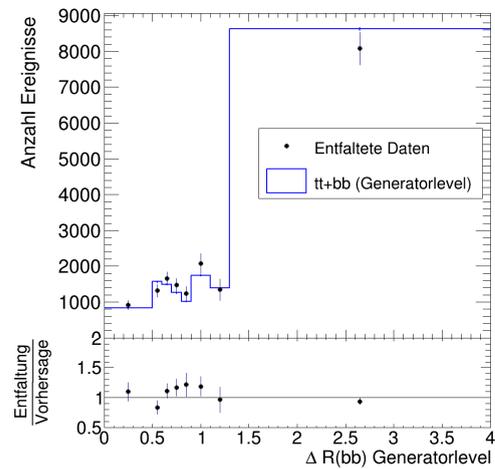
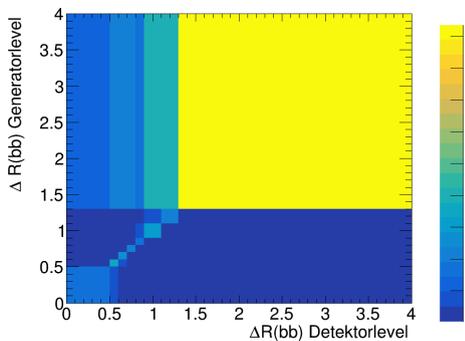
### 4.3.5 Reinheit und Stabilität $\geq 0.5$

Wie schon für die vorherigen Anforderungen an die Reinheit und an die Stabilität nimmt die Anzahl der Bins mit steigenden Werten für die Reinheit und Stabilität ab. Für den geforderten Wert von  $\geq 0.5$  bleiben lediglich zwei Bins übrig (siehe Abb. 4.5). Diese reichen jedoch nicht aus um ein aussagekräftiges Ergebnis durch die Entfaltung zu erhalten, weshalb hierfür auch keine Entfaltung durchgeführt wird.

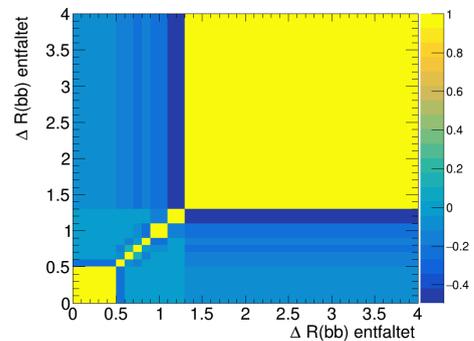
### 4.3.6 Fazit

Für eine höhere Anforderung an die Werte der Reinheit und Stabilität der Daten nimmt die Anzahl der Bins deutlich ab. Jedoch nimmt die statistische Unsicherheit der entfalteten Daten ebenfalls ab. Es gilt daher immer abzuwägen, ob eine feine Auflösung der Verteilungen oder ein höherer Wert an Reinheit und Stabilität für die Entfaltung verwendet werden soll. Die Auflösung sollte groß genug sein, um eine physikalische Aussage über das Ergebnis treffen zu können.

Bei dem Vergleich der verschiedenen Anforderungen an die Werte der Reinheit und Stabilität hat sich der Wert 0.4 als bestes Ergebnis herausgestellt. Eine höhere Anforderung führt zu einer zu geringen Anzahl an Bins, wodurch eine Entfaltung kein aussagekräftiges Ergebnis liefert. Eine kleinerer Wert hat zur Folge, dass die statistischen Schwankungen auf die entfaltete Verteilung  $x$  zunehmen.

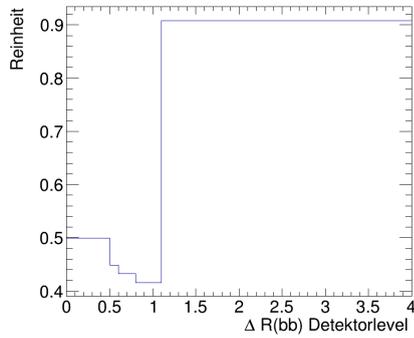
(a) Reinheit der Daten über  $\Delta R$  auf Detektorlevel aufgetragen.(b) Stabilität der Daten über  $\Delta R$  auf Generatorlevel aufgetragen.(c) Simulierte Verteilung von  $\Delta R$  auf dem Detektorlevel zusammen mit den poissonverschmierten Pseudodaten über  $\Delta R$  aufgetragen.(d) Entfaltete Daten zusammen mit der simulierten Verteilung von  $\Delta R$  auf dem Generatorlevel über  $\Delta R$  aufgetragen.

(e) Eingangsmatrix für die Entfaltung.

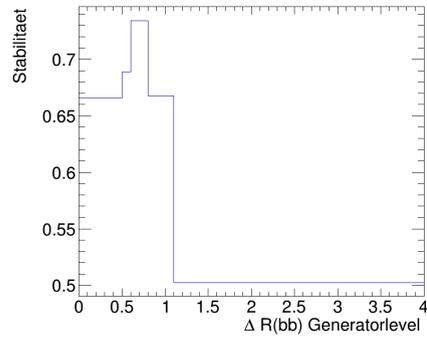


(f) Korrelationsmatrix der entfaltenen Daten.

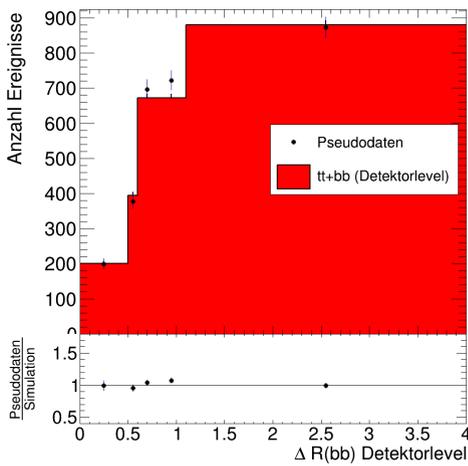
Abbildung 4.3: In dieser Abbildung sind die Verteilungen, welche mit einem Binning durch die Anforderung von 0.3 an die Reinheit und an die Stabilität der Daten erzeugt wurden, zusammen mit den entfaltenen Ergebnissen zu sehen.



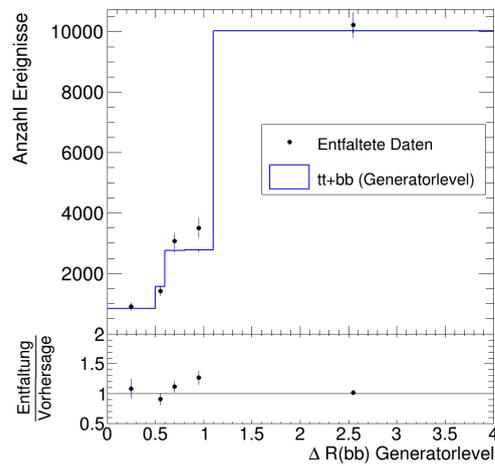
(a) Reinheit der Daten über  $\Delta R$  auf Detektorlevel aufgetragen.



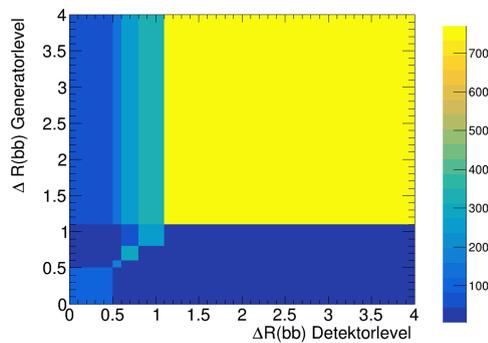
(b) Stabilität der Daten über  $\Delta R$  auf Generatorlevel aufgetragen.



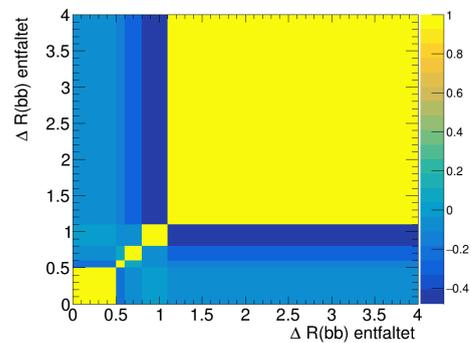
(c) Simulierte Verteilung von  $\Delta R$  auf dem Detektorlevel zusammen mit den poissonverschmierten Pseudodaten über  $\Delta R$  aufgetragen.



(d) Entfaltete Daten zusammen mit der simulierten Verteilung von  $\Delta R$  auf dem Generatorlevel über  $\Delta R$  aufgetragen.



(e) Eingangsmatrix für die Entfaltung.



(f) Korrelationsmatrix der entfalteten Daten.

Abbildung 4.4: In dieser Abbildung sind die Verteilungen, welche mit einem Binning durch die Anforderung von 0.4 an die Reinheit und an die Stabilität der Daten erzeugt wurden, zusammen mit den entfalteten Ergebnissen zu sehen.

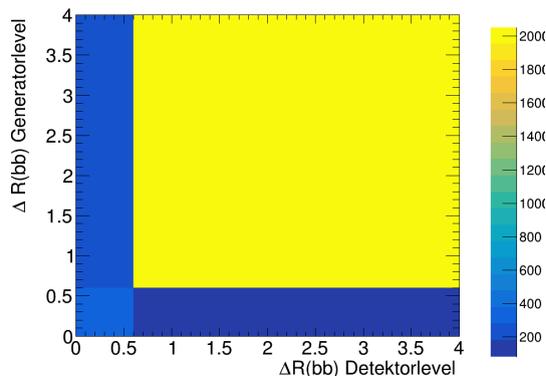


Abbildung 4.5: Eingangsmatrix für die Entfaltung mit einem Binning, welches mit einer Mindestanforderung von 0.5 an die Reinheit und an die Stabilität erstellt wird.

## 4.4 Studien zur Entfaltung durch Variation der Pseudodaten

Die Regularisierung dient zum Ausgleich von statistischen Fluktuationen der zu entfaltenden Daten. Dabei lässt sich die Regularisierungsstärke über den Regularisierungsparameter  $\tau$  variieren. In Abschnitt 4.4.1 wird die Auswirkung der Wahl einer Regularisierungsstärke auf die Entfaltung beobachtet. Anschließend wird die Stabilität der Entfaltung getestet, indem die statistische Fluktuation der Pseudodaten künstlich vergrößert wird. Dabei wird untersucht, ob die Regularisierung bei der Entfaltung diese statistische Fluktuation ausgleichen kann.

In Abschnitt 4.4.3 werden für die Entfaltung skalierte Pseudodaten verwendet. Diese Skalierung stellt die Simulation eines höheren Wirkungsquerschnitts in echten Daten dar.

### 4.4.1 Variation der Regularisierungsstärke

Zur Untersuchung der Entfaltung werden wieder poissonverteilte Pseudodaten verwendet. Dabei wird jedoch, wie in Abschnitt 4.2.2 erläutert, die statistische Fluktuation um einen Faktor von  $f = 5$  vergrößert. Dies dient einem deutlicheren Vergleich zwischen entfalteten Daten mit und ohne Regularisierung. Aus der vorherigen Studie zur Untersuchung des Binnings hat sich die Anforderung von  $\geq 0.4$  für die Reinheit und für die Stabilität der Daten als gut erwiesen. Daher wird im Folgenden die Eingangsmatrix aus dem Binning, welches mit dieser Anforderung erstellt wird, verwendet. Die Pseudodaten sind zusammen mit der theoretischen Verteilung auf Detektorlevel in 4.7a dargestellt.

Zur Untersuchung der Regularisierung wird eine Entfaltung ohne Regularisierung durchgeführt. Dazu wird der Regularisierungsparameter  $\tau$  auf Null gesetzt. Anschließend wird die Minimierung der globalen Korrelation zur Bestimmung einer optimalen Regularisierungsstärke verwendet und die Entfaltung durchgeführt. Dabei wird ein Wert von  $\tau = 2 \cdot 10^{-6}$  ermittelt (siehe Abb. 4.6). Zur Untersuchung einer Überregularisierung wird der Regularisierungsparameter auf  $\tau = 0.1$  gesetzt und eine weitere Entfaltung durchgeführt. In Abbildung 4.7b – 4.7d sind die entfalteten Daten mit den verschiedenen Regularisierungsbedingungen aufgetragen.

Durch den Vergleich der entfalteten Verteilungen mit und ohne Regularisierung (siehe Abb. 4.7b und 4.7c) ist zu sehen, dass die Wahl des Regularisierungsparameters aus der Minimierung der globalen Korrelation die statistischen Fluktuationen und Unsicherheiten der entfalteten Daten verringert. Dieser Effekt ist jedoch nicht sehr stark. Ein zu großer

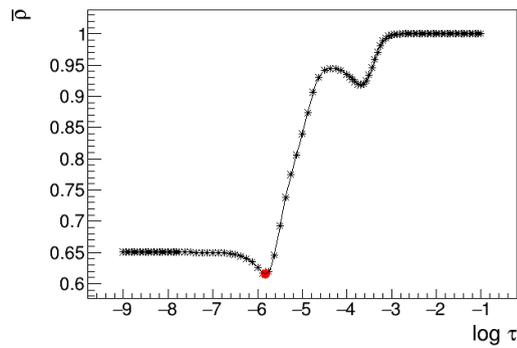


Abbildung 4.6: Die globale Korrelation  $\bar{\rho}$  ist über  $\log \tau$  aufgetragen. Diese wird aus der Entfaltung mit den Pseudodaten ermittelt, welche eine um den Faktor  $f = 3$  vergrößerte statistische Fluktuation besitzen. Das Minimum der globalen Korrelation ist rot markiert und liegt bei  $\tau = 2 \cdot 10^{-6}$ .

Tabelle 4.1: Ermittelte Regularisierungsstärken für die verschiedenen Faktoren der Vergrößerung der statistischen Fluktuation.

Vergrößerungsfaktor $f$	Regularisierungsstärke $\tau$
3	$3 \cdot 10^{-6}$
5	$2 \cdot 10^{-6}$
10	$1 \cdot 10^{-6}$

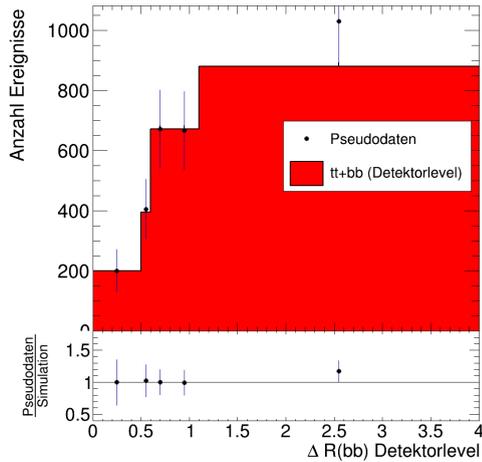
Wert für den Regularisierungsparameter (siehe Abb. 4.7d) führt jedoch zu einer zu starken Glättung, wodurch die entfaltete Verteilung von der theoretischen Vorhersage stark abweicht. In der Abbildung 4.8 sind die Korrelationsmatrizen der entfalteten Verteilungen mit den unterschiedlichen Regularisierungsstärken abgebildet. Dabei ist zu sehen, dass die Korrelation der benachbarten Bins der entfalteten Verteilung mit der Regularisierungsstärke aus der Minimierung der globalen Korrelation geringer ist als bei der entfalteten Verteilung ohne Regularisierung. Die Korrelation der Bins für  $\tau = 0.1$  ist dagegen sehr stark. Es gilt daher eine Überregularisierung zu vermeiden.

Die Regularisierung mit der Regularisierungsstärke aus der Minimierung der globalen Korrelation liefert das beste Ergebnis. Daher wird in den restlichen Studien dieser Arbeit diese Methode zur Bestimmung der Regularisierungsstärke verwendet.

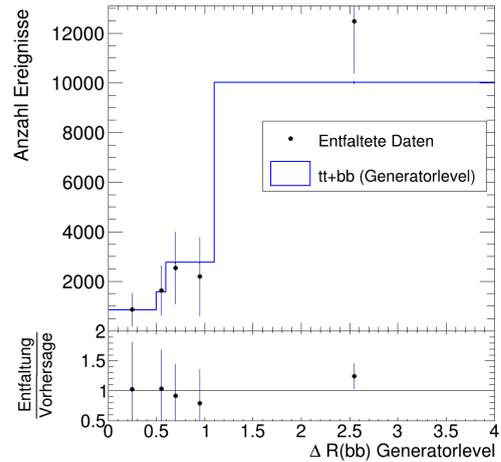
#### 4.4.2 Vergrößerung der statistischen Fluktuation der Pseudodaten

Für die Untersuchung, ob eine Vergrößerung der statistischen Unsicherheit durch die Regularisierung ausgeglichen werden kann, werden Pseudodaten mit unterschiedlich vergrößerten statistischen Fluktuationen erstellt. Es werden Faktoren von  $f = 3, 5$  und  $10$  zur Vergrößerung der statistischen Fluktuation in dieser Studie verwendet. Die Pseudodaten sind mit ihrer statistischen Unsicherheit in Abbildung 4.9 zusammen mit der Verteilung von  $\Delta R$  auf Detektorlevel aufgetragen.

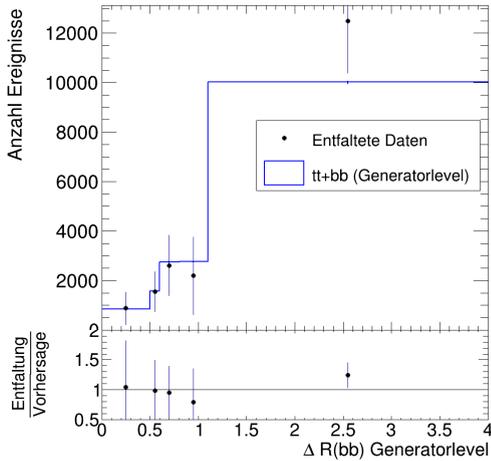
Für eine optimale Regularisierung wird für alle Pseudodaten jeweils die Regularisierungsstärke durch die Minimierung der globalen Korrelation bestimmt. Es ergeben sich die Werte in Tabelle 4.1.



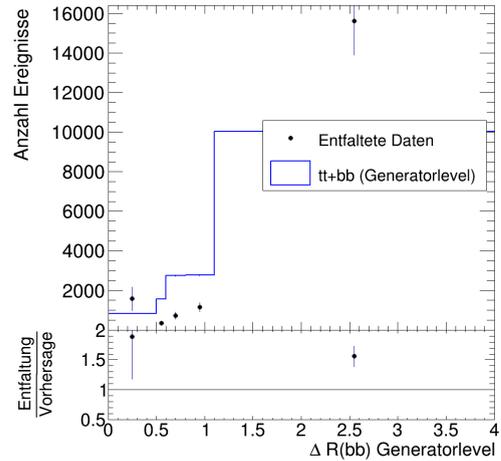
(a) Die poissonverschmierten Pseudodaten sowie die Verteilung auf Detektorlevel ist über  $\Delta R$  aufgetragen.



(b) Die entfaltete Daten ohne Regularisierung sind zusammen mit der Verteilung auf Generatorlevel aufgetragen.

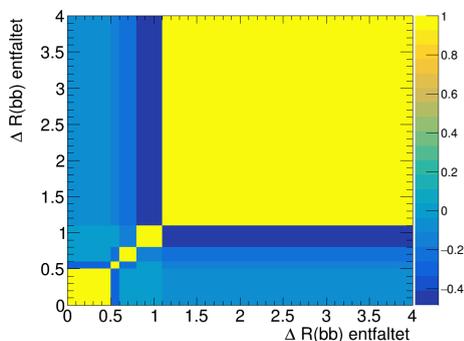


(c) Die entfaltenen Daten, welche mit der Regularisierungsstärke aus der Minimierung der globalen Korrelation erzeugt werden, sind zusammen mit der Verteilung auf dem Generatorlevel aufgetragen.

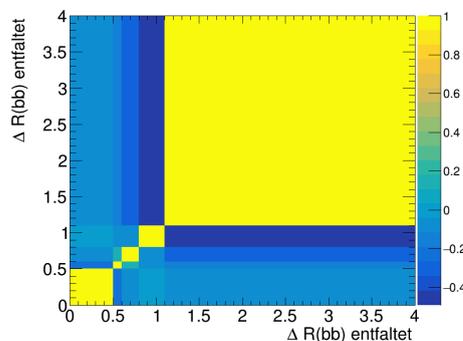


(d) Die entfaltenen Daten, welche mit einer Regularisierungsstärke von  $\tau = 0.1$  erzeugt werden sind zusammen mit der Verteilung auf Generatorlevel aufgetragen.

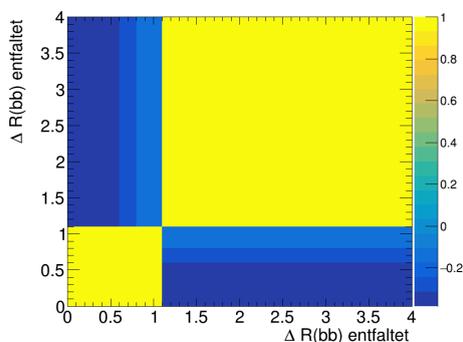
Abbildung 4.7: Die poissonverschmierten Pseudodaten in 4.7a besitzen eine statistische Fluktuation, welche um den Faktor  $f = 5$  vergrößert wurde. In den Abbildungen 4.7b – 4.7d sind die Ergebnisse der Entfaltungen mit unterschiedlicher Regularisierungsstärke zu sehen.



(a) Korrelationsmatrix der entfalteteten Daten ohne Regularisierung.



(b) Korrelationsmatrix der entfalteteten Daten, welche mit der Regularisierungsstärke aus der Minimierung der globalen Korrelation erzeugt werden.



(c) Korrelationsmatrix der entfalteteten Daten, welche mit einer Regularisierungsstärke von  $\tau = 0.1$  erzeugt werden.

Abbildung 4.8: In den Abbildungen 4.8a – 4.8c sind die Korrelationsmatrizen der entfalteteten Daten, welche in Abbildung 4.7 zu aufgetragen sind, zu sehen. Dabei variiert die Regularisierungsstärke, welche für die Entfaltungen verwendet wurde. Daraus ergeben sich die deutlichen Unterschiede der Korrelationsmatrizen.

Für größere statistische Unsicherheiten nimmt der Wert für die Optimierung der Regularisierungsstärke  $\tau$  ab. Jedoch sind die Abweichungen nur minimal. Die unterschiedliche Regularisierungsstärke hat damit keine größeren Auswirkungen auf die entfalteten Daten.

Die Entfaltung wird nun für die unterschiedlichen Pseudodaten mit der jeweiligen optimierten Regularisierungsstärke durchgeführt. Der Vergleich der entfalteten Verteilung von  $\Delta R$  mit der theoretischen Vorhersage wird in Abbildung 4.9 gezeigt.

Durch die Entfaltung ist eine Vergrößerung der statistischen Fluktuation der Daten zu sehen. Die Regularisierung kann die statistische Fluktuation somit nicht ausgleichen. Größere statistische Unsicherheiten von gemessenen Daten lassen sich somit nicht mit Hilfe des TUnfold-Algorithmus dämpfen.

#### 4.4.3 Entfaltung mit skalierten Pseudodaten

Der gemessene Wirkungsquerschnitt für die  $t\bar{t} + b\bar{b}$ -Produktion ist höher als der Wirkungsquerschnitt, welcher durch Monte-Carlo-Generatoren vorhergesagt wird. Dies gilt sowohl für den hadronischen Zerfallskanal [22] als auch für den leptonenischen Zerfallskanal [23] des  $t\bar{t}$ -Paares. Um den Effekt der Abweichung des Wirkungsquerschnitts in gemessenen Daten relativ zu den simulierten Vorhersagen zu untersuchen, werden hierzu Pseudodaten erzeugt, welche anschließend mit einem Faktor von  $g = 1.3$  skaliert werden (siehe Abschnitt 4.2.3).

In Abbildung 4.10 sind die Pseudodaten zusammen mit der Verteilung von  $\Delta R$  auf Detektorlevel sowie die entfalteten Daten zusammen mit der Verteilung auf Generatorlevel aufgetragen. Aus der Minimierung der globalen Korrelation ergibt sich für den Regularisierungsparameter ein Wert von  $\tau = 7 \cdot 10^{-6}$ .

Die statistische Fluktuation der entfalteten Daten ist größer als die Fluktuation von entfaltenen Daten ohne Skalierungsfaktor. Jedoch sind auch die entfaltenen Daten ebenfalls nach oben skaliert. Dies ermöglicht der Term der Flächenbeschränkung, welcher in Abschnitt 3.2.1 beschrieben wird. Somit ist auch eine aussagekräftige Entfaltung für eine nicht-normierte Verteilung möglich.

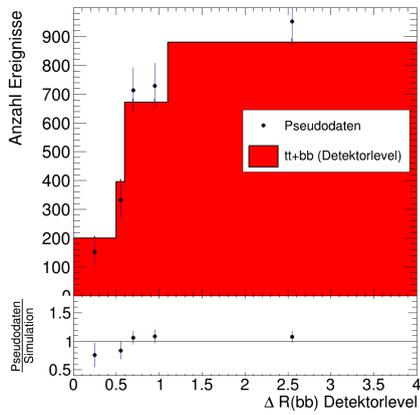
### 4.5 Untersuchung der systematischen Unsicherheiten

Wie schon im vorherigen Kapitel in Abschnitt 3.2.4 beschrieben, können zusätzlich zu den statistischen Unsicherheiten auch systematische Unsicherheiten auftreten. Diese lassen sich durch eine Variation der Eingangsmatrix  $\Delta \mathbf{M}$  untersuchen. Damit können Aussagen über die Stabilität der Entfaltung getroffen werden.

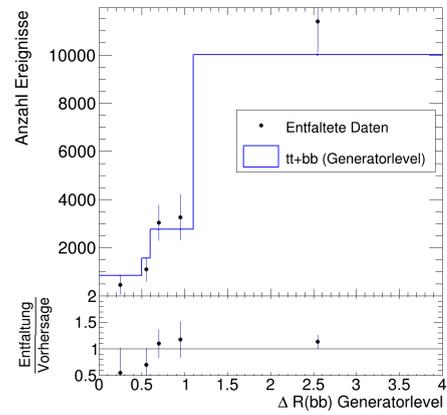
In dieser Arbeit wird eine systematische Unsicherheit der  $\Delta R$ -Verteilung des  $t\bar{t} + b\bar{b}$ -Prozesses verwendet, welche bei der Ableitung der b-tagging-Skalenfaktoren entsteht. Diese wird verwendet um die Unterschiede des b-tagging-Algorithmus in simulierten und echten Daten zu korrigieren. Durch diese Korrektur gibt es eine Gewichtung nach oben und nach unten, welche auf die simulierten Daten zusätzlich angewendet wird (siehe Abb. 4.11).

Für die Untersuchung systematischer Unsicherheiten der entfaltenen Daten werden durch die Verteilungen mit zusätzlichen Gewichtungen der Ereignisse auf Detektorlevel verschiedene Matrizen als Eingabe für die Entfaltung erzeugt.

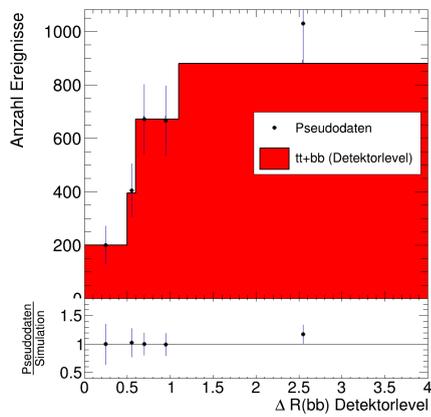
Das Binning wird für alle Matrizen gleich gewählt. Dies ermöglicht einen Vergleich der entfaltenen Daten. Für diese Studie der systematischen Unsicherheit wird das Binning der Matrix verwendet, welches durch simulierte Ereignisse ohne zusätzliche Gewichtung erzeugt wird. Dabei wird auch hier gefordert, dass die Stabilität und die Reinheit in jedem



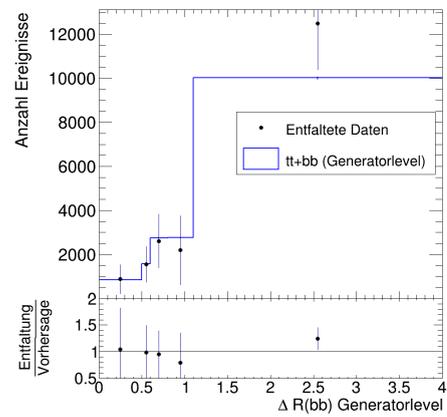
(a) Pseudodaten mit einer um den Faktor  $f = 3$  vergrößerten statistischen Fluktuation.



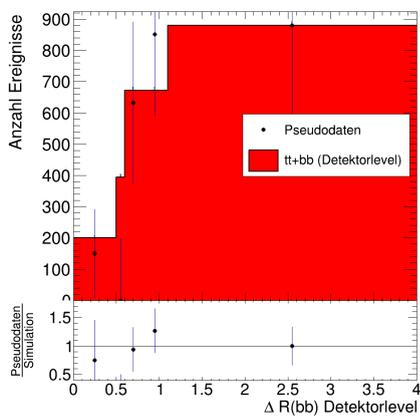
(b) Entfaltete Pseudodaten mit einer um den Faktor  $f = 3$  vergrößerten statistischen Fluktuation.



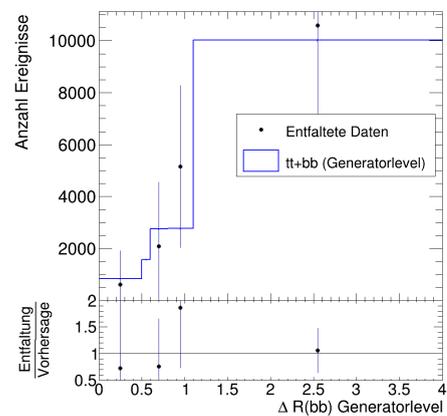
(c) Pseudodaten mit einer um den Faktor  $f = 5$  vergrößerten statistischen Fluktuation.



(d) Entfaltete Pseudodaten mit einer um den Faktor  $f = 5$  vergrößerten statistischen Fluktuation.

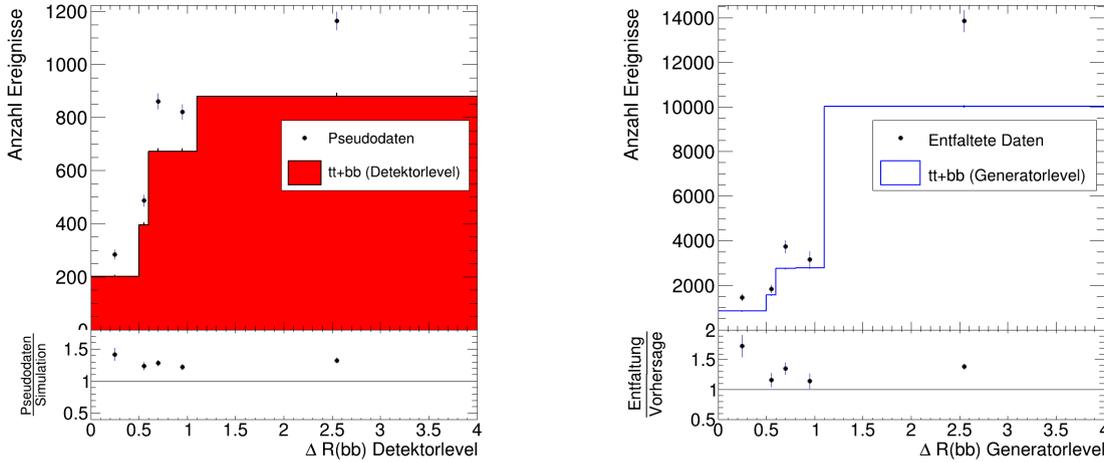


(e) Pseudodaten mit einer um den Faktor  $f = 10$  vergrößerten statistischen Fluktuation.



(f) Entfaltete Pseudodaten mit einer um den Faktor  $f = 10$  vergrößerten statistischen Fluktuation.

Abbildung 4.9: In den Abbildungen 4.9a, 4.9c und 4.9e sind Pseudodaten über  $\Delta R$  aufgetragen, deren statistische Fluktuation vergrößert ist. In den Abbildungen 4.9b, 4.9d und 4.9f sind die dazugehörigen entfalteten Daten aufgetragen.



(a) Die skalierten Pseudodaten sind zusammen mit der unskalierten Verteilung auf Detektorlevel aufgetragen.

(b) Die entfalteten Daten sind zusammen mit der unskalierten Verteilung auf Generatorlevel aufgetragen.

Abbildung 4.10: Die Pseudodaten sind mit einem Faktor von  $g = 1.3$  skaliert (siehe Abb. 4.10a). Die daraus entfalteten Daten sind in Abbildung 4.10b zu sehen. Das Binning erfüllt die Anforderung von  $\geq 0.4$ .

Bin einen Wert von  $\geq 0.4$  besitzen, da dies, wie in Abschnitt 4.3 beschrieben, die besten Ergebnisse liefert.

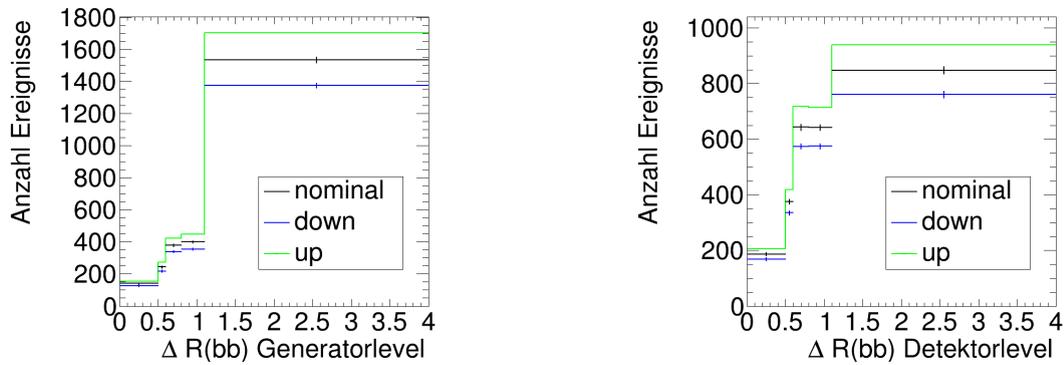
Mit den zwei zusätzlichen Gewichtungen ergeben sich drei verschiedene Matrizen für die Entfaltung (siehe Abb. 4.12). Dabei wird die Matrix, welche mit Ereignissen ohne zusätzliche Gewichtung erzeugt wird, als  $\mathbf{M}_0$  bezeichnet. Die beiden anderen Matrizen werden nach ihrer zusätzlichen Gewichtung als  $\mathbf{M}_{\text{up}}$  und  $\mathbf{M}_{\text{down}}$  bezeichnet.

Auch die Pseudodaten, mit denen die systematische Unsicherheit untersucht wird, werden aus den simulierten Ereignissen ohne zusätzliche Gewichtung erzeugt (siehe Abb. 4.13). Dafür werden wie gehabt poissonverschmierte Zufallszahlen auf Basis der Verteilung auf Detektorlevel erzeugt. Für alle drei Entfaltungen mit den verschiedenen Eingangsmatrizen werden dieselben Pseudodaten verwendet.

Für den Vergleich der entfalteten Verteilungen wird auch die Regularisierungsstärke für alle Entfaltungen gleich gewählt. Dazu wird die Methode der Minimierung der globalen Korrelation auf die Entfaltung mit der Eingangsmatrix  $\mathbf{M}_0$  angewendet. Dabei ergibt sich für die Regularisierungsstärke, wie schon im vorherigen Teilkapitel, ein Wert von  $\tau = 7 \cdot 10^{-6}$ . Da die Verteilungen und die statistischen Unsicherheiten der Verteilungen mit unterschiedlichen Gewichtungen nicht stark voneinander abweichen, eignet sich die gewählte Regularisierungsstärke auch für die Entfaltungen mit den beiden anderen Matrizen  $\mathbf{M}_{\text{up}}$  und  $\mathbf{M}_{\text{down}}$ .

Für alle drei Matrizen wird eine separate Entfaltung durch TUnfold mit den gleichen Pseudodaten und dem gleichen Wert für die Regularisierungsstärke durchgeführt. Es wird der differentielle Wirkungsquerschnitt ermittelt. Dieser wird über

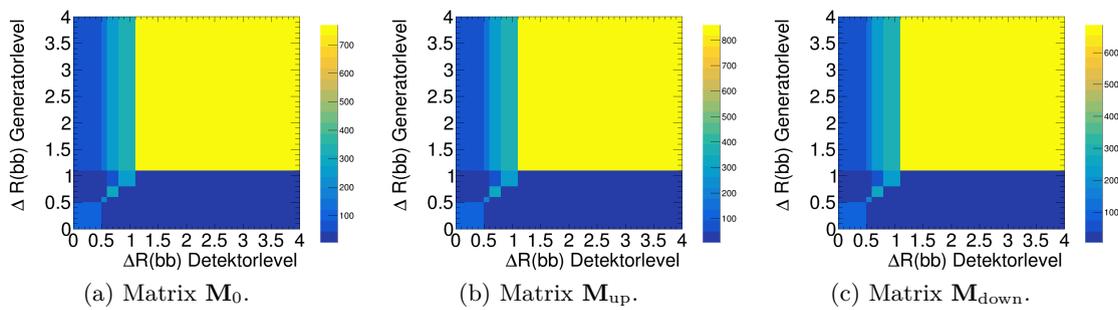
$$\frac{d\sigma}{d\Delta R(\text{bb})} = \frac{1}{L_{\text{int}}} \cdot \frac{dN^{\text{Signal}}}{d\Delta R(\text{bb})} \quad (4.8)$$



(a) Verteilung auf Generatorlevel über  $\Delta R$  aufgetragen.

(b) Verteilung auf Detektorlevel über  $\Delta R$  aufgetragen.

Abbildung 4.11: Die Verteilungen der Ereignisse, die sowohl die Generator- als auch die Detektorselektion erfüllen, ist sowohl mit als auch ohne zusätzliche Gewichungen in den Abbildungen 4.11a und 4.11b aufgetragen. Die Verteilungen mit den zusätzlichen Gewichungen geben dabei die systematische Unsicherheit an.



(a) Matrix  $M_0$ .

(b) Matrix  $M_{up}$ .

(c) Matrix  $M_{down}$ .

Abbildung 4.12: Erzeugte Matrizen für die Entfaltung mit zusätzlichen Gewichungen der Ereignisse.

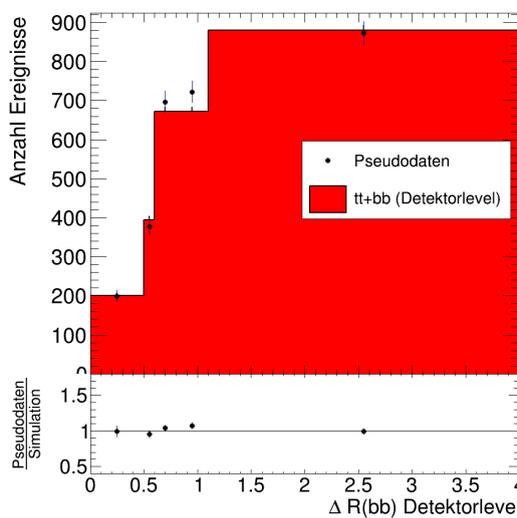


Abbildung 4.13: Die Verteilung der Pseudodaten, welche zur Untersuchung der systematischen Unsicherheit verwendet wird, wird zusammen mit der simulierten Verteilung aufgetragen ohne zusätzliche Gewichtung.

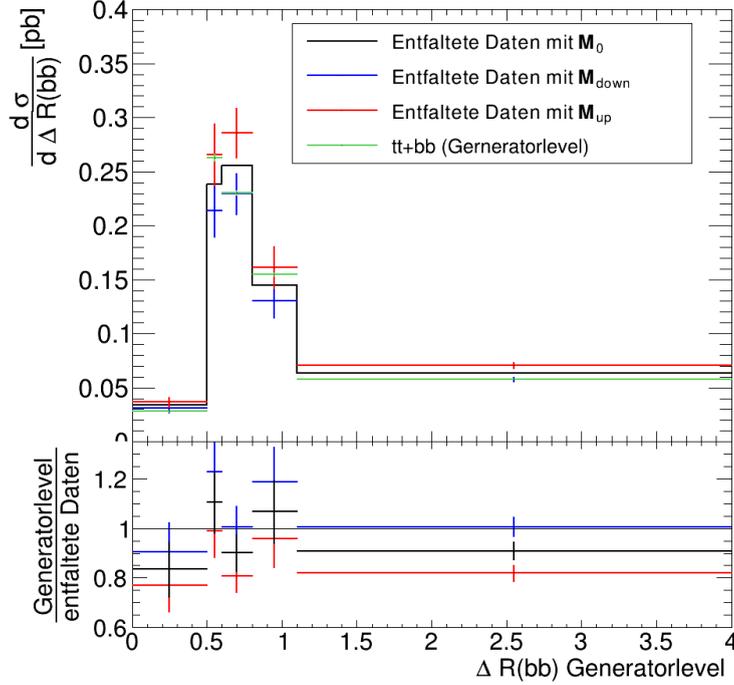


Abbildung 4.14: Die entfalteten Daten sind über  $\Delta R$  auf dem Generatorlevel zusammen mit der theoretischen Vorhersage aufgetragen. Für die Entfaltung wird die Eingangsmatrix  $\mathbf{M}$  verwendet, welche eine Mindestanforderung an Reinheit und Stabilität der Daten von 0.4 erfüllt. Zusätzlich werden Entfaltungen mit Matrizen durchgeführt, die mit zusätzlichen Gewichtungen auf die Ereignissen erstellt werden. Dies dient zur Untersuchung von systematischen Unsicherheiten, welche durch die Abweichungen der entfalteten Daten zu sehen ist.

bestimmt. Für die diskretisierte Verteilung lässt sich dieser Ausdruck vereinfachen zu

$$\frac{d\sigma}{d\Delta R(\text{bb})_i} = \frac{N_i^{\text{Signal}}}{L_{\text{int}} \cdot \Delta_{\text{Bin},i}}. \quad (4.9)$$

Dabei ist  $N_i^{\text{Signal}}$  die Anzahl der entfalteten Ereignisse in dem Bin  $i$ , welche von dem Signalprozess stammen,  $L_{\text{int}}$  die integrierte Luminosität, welche von dem Detektor gemessen wird, und  $\Delta_{\text{Bin},i}$  die Breite des  $i$ -ten Bins. Die verwendete integrierte Luminosität der simulierten Daten beträgt  $L_{\text{int}} = 59.70 \text{ fb}^{-1}$ .

Die entfalteten Verteilungen sind in Abbildung 4.14 aufgetragen. Die Abweichung der entfalteten Daten der gewichteten Verteilungen können nun als systematische Unsicherheit beschrieben werden. Damit ergeben sich zu den statistischen Unsicherheiten der entfalteten Daten, welche durch die Kovarianzmatrix 4.4f beschrieben werden, die systematischen Unsicherheiten  $\Delta \mathbf{x}$ , welche in Abbildung 4.14 zu sehen sind.

Die theoretische Vorhersage liegt damit im Bereich der systematischen Unsicherheit. Lediglich der Wert der theoretischen Vorhersage im ersten Bin liegt außerhalb der unteren Abweichung der entfalteten Daten.

Die Entfaltung mit TUnfold liefert ein stabiles Ergebnis für die entfalteten Pseudodaten. Es lässt sich darauf schließen, dass die Entfaltung mit TUnfold für echte gemessene Daten

vergleichbare stabile Ergebnisse liefert. Für echte gemessene Daten müssen zusätzlich noch weitere systematische Unsicherheiten berücksichtigt werden. Diese lassen sich wie die Unsicherheit behandeln, welche in dieser Arbeit verwendet worden ist. Damit lässt sich der Prozess der Entfaltung, wie er in dieser Arbeit auf Pseudodaten angewendet wurde, auf echte gemessene Daten anwenden.



## 5 Zusammenfassung und Ausblick

Als häufigster Untergrundprozesse des  $t\bar{t}H(b\bar{b})$ -Prozesses spielt der  $t\bar{t} + b\bar{b}$ -Prozess eine wichtige Rolle. Dessen kinematische Verteilungen können durch hochenergetische Streuexperimente am LHC mit Hilfe des CMS-Detektors gemessen werden. Dabei erzeugen die Detektoreffekte eine Verschmierung der Verteilung, welche sich durch eine Entfaltung der Daten korrigieren lässt.

In dieser Arbeit wurde das Werkzeug TUnfold für die Entfaltung der  $\Delta R(b\bar{b})$ -Verteilung des  $t\bar{t} + b\bar{b}$ -Prozesses verwendet. Ziel dieser Arbeit war die Untersuchung der Stabilität der entfalteten Ergebnisse. Dafür wurde eine Optimierung des Binnings durchgeführt, die Regularisierung der Entfaltung studiert sowie systematische Unsicherheiten der entfalteten Ergebnisse untersucht. Da die Methode der Entfaltung im Vordergrund stand, wurden statt gemessenen Daten poissonverschmierte Pseudodaten für die Entfaltung verwendet.

Für das Binning wurde zu Beginn eine Mindestanforderung an 30 Ereignisse pro Bin gestellt. Dies ermöglicht ein statistisch aussagekräftiges Ergebnis. Es wurden geeignete Mindestanforderungen an die Reinheit und die Stabilität der Daten untersucht. Die Pseudodaten wurden mit den Binnings, welche sich aus verschiedenen Werten für die Mindestanforderung für die Reinheit und die Stabilität ergaben, entfaltet.

Aus dem Vergleich der Ergebnisse der Entfaltungen lässt sich eine optimale Anforderung an die Stabilität und an die Reinheit von  $\geq 0.4$  bestimmen. Für höhere Anforderungen nimmt die Anzahl der Bins zu stark ab und für geringere Anforderungen nimmt die statistische Fluktuation der entfalteten Daten sowie deren statistische Unsicherheit zu.

Anschließend wurde die Regularisierung des TUnfold-Algorithmus untersucht. Zur Bestimmung einer geeigneten Regularisierungsstärke wird die Minimierung der globalen Korrelation getestet. Um die Auswirkungen der Regularisierung zu beobachten, werden zusätzlich eine Entfaltung ohne Regularisierung und eine Entfaltung mit einer sehr starken Regularisierung durchgeführt. Die Entfaltung mit der Regularisierungsstärke aus der Minimierung der globalen Korrelation ermöglicht eine Verringerung der statistischen Fluktuation der entfalteten Ergebnisse. Dadurch nimmt die Korrelation der benachbarten Bins ebenfalls ab. Eine zu starken Regularisierungsstärke verursacht große Korrelationen zwischen den Bins und führt zu einer starken Abweichung der entfalteten Verteilung von der erwarteten Verteilung. Die Regularisierungsstärke aus der Minimierung der globalen Korrelation hat sich damit als sinnvoll erwiesen.

Zur Untersuchung der Stabilität der Entfaltung werden poissonverschmierte Pseudodaten erzeugt, deren statistische Fluktuationen künstlich um einen Faktor vergrößert werden. Diese statistischen Fluktuationen werden durch die Entfaltung verstärkt. Der TUnfold-Algorithmus ermöglicht es somit nicht, große statistische Fluktuationen zu dämpfen. Zusätzlich wurde eine Entfaltung mit skalierten Pseudodaten durchgeführt. Die Skalierung bewirkt eine Vergrößerung der statistischen Fluktuation der entfalteten Daten. Der Term der Flächenbegrenzung im TUnfold-Algorithmus ermöglicht jedoch trotz vergrößerter statistischer Unsicherheiten ein aussagekräftiges Ergebnis der Entfaltung.

Für die Untersuchung von systematischen Unsicherheiten werden simulierte Verteilungen mit unterschiedlichen Gewichtungen verwendet. Damit lassen sich die systematischen Unsicherheiten der entfalteten Daten abschätzen. In dieser Arbeit wurde beispielhaft eine Unsicherheit bezüglich des b-tagging-Algorithmus verwendet. Die Abweichungen der entfalteten Pseudodaten von der ursprünglichen simulierten Verteilung liegen im Bereich der systematischen Unsicherheit. Damit liefert die Entfaltung mit TUnfold ein stabiles Ergebnis.

Für weitere Untersuchungen empfiehlt es sich, den TUnfold-Algorithmus auf echte gemessene Daten anzuwenden. Dazu müssten weitere systematische Unsicherheiten berücksichtigt werden. Die Studien dieser Arbeit lassen vermuten, dass diese Methode der Entfaltung für gemessene Daten ebenfalls ein vergleichbar stabiles Ergebnis liefern würde. Zur Untersuchung der Allgemeingültigkeit der Methoden dieser Arbeit lassen sich diese an weiteren Entfaltungen kinematischer Verteilungen erproben.

# Literatur

- [1] S. Chatrchyan u. a. „Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC“. In: *Physics Letters B* 716.1 (2012), S. 30–61. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.021>. URL: <http://www.sciencedirect.com/science/article/pii/S0370269312008581>.
- [2] D. J. Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. New York, NY: Wiley, 2008. URL: <https://cds.cern.ch/record/111880>.
- [3] ALICE Collaboration. „The ALICE experiment at the CERN LHC“. In: *Journal of Instrumentation* 3.08 (2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002. URL: <https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08002>.
- [4] ATLAS Collaboration. „The ATLAS Experiment at the CERN Large Hadron Collider“. In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08003. 437 p. DOI: 10.1088/1748-0221/3/08/S08003. URL: <https://cds.cern.ch/record/1129811>.
- [5] CMS Collaboration. „The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment“. In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08004. 361 p. DOI: 10.1088/1748-0221/3/08/S08004. URL: <https://cds.cern.ch/record/1129810>.
- [6] LHCb Collaboration. „The LHCb Detector at the LHC“. In: *JINST* 3.LHCb-DP-2008-001. CERN-LHCb-DP-2008-001 (2008). Also published by CERN Geneva in 2010, S08005. DOI: 10.1088/1748-0221/3/08/S08005. URL: <https://cds.cern.ch/record/1129809>.
- [7] S. R. Davis. „Interactive Slice of the CMS detector“. In: (Aug. 2016). URL: <https://cds.cern.ch/record/2205172>.
- [8] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/368412>.
- [9] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/349375>.
- [10] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/357153>.
- [11] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/343814>.
- [12] CMS Collaboration. „Particle-flow reconstruction and global event description with the CMS detector“. In: *Journal of Instrumentation* 12.10 (Okt. 2017), P10003–P10003. DOI: 10.1088/1748-0221/12/10/p10003. URL: <https://doi.org/10.1088%2F1748-0221%2F12%2F10%2Fp10003>.

- [13] M. Cacciari, G. P. Salam und G. Soyez. „The anti-ktjet clustering algorithm“. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), S. 063–063. DOI: 10.1088/1126-6708/2008/04/063. URL: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [14] CMS Collaboration. „Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector“. In: (Nov. 2018). URL: <https://cds.cern.ch/record/2646773>.
- [15] S. Alioli u. a. „A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX“. In: *Journal of High Energy Physics* 2010.6 (Juni 2010). ISSN: 1029-8479. DOI: 10.1007/jhep06(2010)043. URL: [http://dx.doi.org/10.1007/JHEP06\(2010\)043](http://dx.doi.org/10.1007/JHEP06(2010)043).
- [16] T. Sjöstrand u. a. „An introduction to PYTHIA 8.2“. English. In: *Computer Physics Communications* 191 (2015), S. 159–177. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2015.01.024.
- [17] S. Agostinelli u. a. „Geant4—a simulation toolkit“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), S. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [18] V. Blobel und E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Teubner Studienbücher Physik. Stuttgart: Teubner, 1998. URL: <https://cds.cern.ch/record/437773>.
- [19] S. Schmitt. „TUnfold, an algorithm for correcting migration effects in high energy physics“. In: *Journal of Instrumentation* 7.10 (2012), T10003–T10003. DOI: 10.1088/1748-0221/7/10/t10003. URL: <https://doi.org/10.1088/1748-0221/7/10/t10003>.
- [20] A. Tikhonov. „Solution of Incorrectly Formulated Problems and the Regularization Method“. In: *Soviet Math. Dokl.* 5 (1963), S. 1035/1038. URL: <https://ci.nii.ac.jp/naid/10004315593/en/>.
- [21] P. C. Hansen. „The L-Curve and its Use in the Numerical Treatment of Inverse Problems“. In: *in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering*. WIT Press, 2000, S. 119–142.
- [22] CMS Collaboration. „Measurement of the  $t\bar{t}b\bar{b}$  production cross section in the all-jet final state in pp collisions at  $\sqrt{s}=13$  TeV“. In: *Physics Letters B* 803 (2020), S. 135285. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2020.135285>. URL: <http://www.sciencedirect.com/science/article/pii/S0370269320300897>.
- [23] CMS Collaboration. „Measurement of the cross section for  $t\bar{t}$  production with additional jets and b jets in pp collisions at  $\sqrt{s} = 13$  TeV“. In: *Journal of High Energy Physics* 2020.7 (Juli 2020). ISSN: 1029-8479. DOI: 10.1007/jhep07(2020)125. URL: [http://dx.doi.org/10.1007/JHEP07\(2020\)125](http://dx.doi.org/10.1007/JHEP07(2020)125).

## Abbildungsverzeichnis

2.1	Schnitt des CMS-Detektors. . . . .	5
2.2	Feynmandiagramme des $t\bar{t} + b\bar{b}$ -Prozesses und des $t\bar{t}H(b\bar{b})$ -Prozesses. . . . .	9
4.1	Simulierte Ereignisse auf Generatorlevel mit 30 Ereignissen pro Bin. . . . .	23
4.2	Verteilungen des Binnings mit Reinheit und Stabilität $\geq 0.1$ . . . . .	24
4.3	Verteilungen des Binnings mit Reinheit und Stabilität $\geq 0.3$ . . . . .	26
4.4	Verteilungen des Binnings mit Reinheit und Stabilität $\geq 0.4$ . . . . .	27
4.5	Verteilungen des Binnings mit Reinheit und Stabilität $\geq 0.5$ . . . . .	28
4.6	Globale Korrelation $\rho$ über der Regularisierungsstärke $\tau$ aufgetragen. . . . .	29
4.7	Entfaltungen mit unterschiedlicher Regularisierungsstärke. . . . .	30
4.8	Korrelationsmatrizen der entfalteten Daten mit unterschiedlicher Regularisierungsstärke. . . . .	31
4.9	Pseudodaten mit vergrößerter statistischer Fluktuation und deren entfaltete Ergebnisse. . . . .	33
4.10	Entfaltung skalierten Pseudodaten. . . . .	34
4.11	Verteilungen auf Generatorlevel und Detektorlevel mit unterschiedlichen Gewichtungen. . . . .	35
4.12	Erzeugte Matrizen für die Entfaltung mit zusätzlichen Gewichtungen der Ereignisse. . . . .	35
4.13	Verteilung der Pseudodaten für die Untersuchung der systematischen Unsicherheit. . . . .	35
4.14	Entfaltete Verteilungen mit unterschiedlichen Gewichtungen. . . . .	36

## Tabellenverzeichnis

4.1	Ermittelte Regularisierungsstärken für die verschiedenen Faktoren der Vergrößerung der statistischen Fluktuation. . . . .	29
-----	---	----

