

Studien zur Entfaltung kinematischer Verteilungen in der elektroschwachen Streuung von W - und Z -Bosonen am LHC

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

vorgelegt von

Max Vincent Stange
geboren am 06.05.1996 in Hoyerswerda

Institut für Kern- und Teilchenphysik
Fachrichtung Physik
Fakultät Mathematik und Naturwissenschaften
Technische Universität Dresden
2018

Eingereicht am 26. Juni 2018

1. Gutachter: Prof. Dr. Michael Kobel
2. Gutachter: Prof. Dr. Kai Zuber

Zusammenfassung

Um die am ATLAS-Detektor gemessenen kinematischen Verteilungen der WZ -Streuung mit anderen Messungen oder bestehenden bzw. zukünftigen Theorien vergleichen zu können, müssen die Einflüsse des Detektors, wie beispielsweise die beschränkte Effizienz und Auflösung, auf die gemessenen Daten beachtet werden. Dies geschieht hier durch die Entfaltung der Daten, bei der ermittelt wird, welche tatsächliche Verteilung dieser Messung wahrscheinlich zugrunde liegt. In dieser Arbeit wird die elektroschwache Streuung von W - und Z -Bosonen betrachtet und dafür exemplarisch Monte Carlo Daten zu der transversalen Masse der Streuung $M_T(WZ)$, dem transversalen Impuls des Z -Bosons $P_T(Z)$ und der Anzahl der gemessenen Jets N_{Jets} bei unterschiedlichen integrierten Luminositäten entfaltet. Dabei werden auch integrierte Luminositäten verwendet, welche im Run 2 des LHC realisierbar sind. Als Vorbereitung für den Entfaltungsprozess wird eine Methode zur Binoptimierung gewählt, welche die Bestimmung einer jeweils geeigneten Binaufteilung ermöglicht. Anschließend werden verschiedene Entfaltungsmethoden angewendet und deren Ergebnis inklusive der Unsicherheit dargestellt, welche sich aus der Propagation der systematischen und statistischen Unsicherheiten ergibt. Daran anknüpfend wird auch die durch den Entfaltungsprozess entstehende Unsicherheit ermittelt. Abschließend folgt die Entfaltung einer statistischen Fluktuation der Monte Carlo Pseudodaten, um einen Eindruck zu erhalten, wie die Entfaltungsergebnisse von tatsächlichen Messdaten aussehen könnten. Das in dieser Arbeit für die Entfaltung beschriebene Vorgehen ist zum Großteil allgemein gehalten und lässt sich so auch auf andere Verteilungen anwenden.

Abstract

For the comparison between kinematic distributions in the WZ -scattering measured by the ATLAS-detector and other measurements or theory predictions, it is necessary to consider the impact of detector effects such as the limited efficiency and resolution. One way to eliminate the detector impact on the data is unfolding. In this bachelor thesis Monte Carlo data of the transverse mass $M_T(WZ)$, the transverse momentum $P_T(Z)$ and the number of jets in the electroweak scattering of W - and Z -bosons are unfolded. The unfolding process is studied at different integrated luminosities which can be partially reached in the Run 2 of the LHC. At first a method for the bin optimization is chosen in preparation of the unfolding. Different unfolding methods are used to calculate the unfolding result and its statistical and systematic uncertainties. Subsequently the uncertainty caused by the unfolding process is determined. Finally statistical fluctuated Monte Carlo pseudo data is unfolded to gain an impression of how the unfolding results of real measured data might look like. The procedure of unfolding, which is used in this thesis, is kept general and could be applied to other data distributions.

Inhaltsverzeichnis

2	Einleitung	3
3	Vorbetrachtung	5
3.1	LHC und ATLAS	5
3.2	Streuung von W - und Z -Bosonen	5
3.2.1	Transversale Masse	6
3.3	Monte Carlo Simulationen	7
3.4	Entfaltung	7
3.4.1	Bin-by-Bin-Entfaltung	9
3.4.2	Entfaltung durch Matrix-Invertierung	10
3.4.3	Bayesische Entfaltung	11
3.5	Statistische und systematische Unsicherheit	12
4	Hauptteil	15
4.1	Binoptimierung	16
4.1.1	Mindestanforderungen und Optimierungskriterien	16
4.1.2	Durchführung der Optimierung	17
4.1.3	Ergebnisse der Binoptimierung	18
4.2	Technical Closure Test	20
4.3	Entfaltungsergebnisse	20
4.4	Entfaltungsunsicherheit	23
4.4.1	Bestimmung der Ungewichtungsfunktion	23
4.4.2	Test der Bestimmung der Ungewichtungsfunktion	25
4.4.3	Berechnung der Entfaltungsunsicherheit	28
4.5	Entfaltung fluktuierter Monte Carlo Daten	30
5	Zusammenfassung	33
6	Anhang	35
7	Literaturverzeichnis	69

2 Einleitung

Eine der erfolgreichsten physikalischen Theorien der modernen Physik stellt das Standardmodell der Teilchenphysik dar, das schon durch zahlreiche Experimente überprüft wurde. Dabei ist nicht nur die Messung der relevanten Daten sondern auch deren Auswertung entscheidend. Die gemessenen Daten können nicht direkt mit den theoretischen Vorhersagen verglichen werden, weil sie durch Detektoreffekte beeinflusst werden. Da der Detektor beispielsweise nur eine begrenzte Auflösung, Effizienz und Kalibrierengenauigkeit besitzt, weichen die gemessenen Daten von der eigentlichen Messverteilung ab. Aus diesem Grund ist es nötig, unter Kenntnis dieser Einflüsse die Detektoreinflüsse auf die gemessenen Daten zu korrigieren, damit sie mit Vorhersagen des Standardmodells oder darüber hinaus gehenden Theorien verglichen werden können.

Dieser als Entfaltung bezeichnete Vorgang wird im Folgenden für kinematische Verteilungen in der elektroschwachen Streuung von W - und Z -Bosonen durchgeführt. Dafür wird exemplarisch die transversale Masse des Streuprozesses $M_T(WZ)$, der transversale Impuls des Z -Bosons $P_T(Z)$ und die Anzahl der gemessenen Jets N_{Jets} entfaltet. Sowohl für die Entfaltung, als auch als Daten werden vom CERN stammende Monte Carlo Simulationen genutzt, welche für den ATLAS-Detektor durchgeführt wurden. Da das allgemeine Vorgehen bei der Entfaltung im Vordergrund steht, wird von der Entfaltung von tatsächlich gemessenen Daten abgesehen. Als Daten werden aus diesem Grund ausschließlich Pseudodaten von Monte Carlo Generatoren verwendet. Begonnen wird mit der Optimierung der Binaufteilung, welche für die zu entfaltenden Verteilungen gewählt wird. Dabei werden unterschiedliche Mindestanforderungen und Optimierungskriterien an die Binaufteilung gestellt, welche sich noch erweitern ließen. Zusätzlich lassen sich anhand der gestellten Mindestanforderungen auch die für eine bestimmte Binanzahl notwendigen integrierten Luminositäten berechnen. Daran anschließend folgt die Entfaltung der Pseudodaten inklusive der Propagation der systematischen und statistischen Unsicherheiten. Zur Entfaltung werden die Bin-by-Bin-Entfaltung, die Entfaltung durch Matrix-Invertierung und die Bayesische Entfaltung angewendet und deren Ergebnisse zur besseren Vergleichbarkeit gemeinsam dargestellt. Der Entfaltung folgend wird die Entfaltungsunsicherheit mittels einer Methode abgeschätzt, welche auf dem Data-driven Closure Test basiert. Dieses Vorgehen bei der Entfaltung wird abschließend noch einmal für eine Fluktuation der Monte Carlo Daten angewendet, um einen Eindruck für die Entfaltung von Daten, welche statistischen Fluktuationen unterliegen, zu gewinnen.

3 Vorbetrachtung

3.1 LHC und ATLAS

Viele Phänomene, welche von dem Standardmodell und darüber hinausführende Theorien vorhergesagt werden, lassen sich erst im Rahmen der Hochenergiephysik untersuchen. Aus diesem Grund bieten sich oftmals Wechselwirkungen von hochenergetischen Teilchen an, um die jeweiligen Vorhersagen zu überprüfen. Dabei spielt der Bau von leistungsfähigen Beschleunigeranlagen eine entscheidende Rolle. Das bisher größte und leistungsstärkste Exemplar stellt dabei der LHC im CERN dar. Er besitzt einen Umfang von etwa 27 km und ermöglicht Hadronenkollisionen mit einer Schwerpunktsenergie in einer Größenordnung von bis zu 14 TeV [12]. Die beschleunigten Hadronen werden dabei bis zu ihrer Kollision von supraleitenden Magneten auf ihrer Kreisbahn gehalten.

Die im LHC stattfindenden Kollisionen werden unter anderen von ATLAS detektiert, welcher einer der am LHC befindlichen Teilchendetektoren ist. Dieser Detektor wurde zylinderförmig um die Strahlachse gebaut und besitzt einen zwiebelschalenförmigen Aufbau. Im Inneren des Detektors sind homogene Magnetfelder mit einer Stärke von bis zu 2 T, um durch die daraus resultierende Ablenkung beispielsweise den Impuls des jeweiligen Teilchens zu bestimmen. Die Teilchen durchqueren von dem Wechselwirkungspunkt aus zuerst den Pixeldetektor, welcher umschlossen ist von einer Spurkammer. Dieser folgt ein elektromagnetisches und anschließend ein hadronisches Kalorimeter. Den äußersten Detektor des ATLAS bildet schließlich der Myonen-Detektor. [2]

Dieser Detektoraufbau ermöglicht nicht nur die Klassifizierung der Teilchen und die Bestimmung ihrer Flugrichtung, sondern lässt sich aus den Messungen auch der Impuls und die Energie des jeweiligen Teilchens ermitteln.

3.2 Streuung von W - und Z -Bosonen

Bei den am CERN durchgeführten Experimenten konnten bisher zahlreiche vom Standardmodell vorhergesagte Vorgänge beobachtet werden. Ein Beispiel stellt die Entdeckung der W - und Z -Bosonen im Jahr 1983 einem $\bar{p}p$ -Collider dar [2, Seite 224].

Zusätzlich zu der Erzeugung von Vektor-Bosonen kann auch deren Streuung untereinander, nachdem sie von den jeweiligen Quarks abgestrahlt wurden, untersucht werden. Exemplarisch

lässt sich die Streuung $WZ \rightarrow WZ$ über die Zerfallsprodukte der beteiligten Vektor-Bosonen beobachten. Zwar besitzen beide Vektor-Bosonen die größte Zerfallsbreite im hadronischen Zerfall [10] [11], jedoch sind diese bei Hadronenkollisionen nur schwer zu untersuchen und man konzentriert deshalb sich auf die leptonischen Zerfallskanäle [2, Seite 224]. Da die W -Bosonen in ein Lepton und das dazugehörige Neutrino zerfallen und das Z -Boson ein Lepton-Antilepton-Paar erzeugt, werden bei der WZ -Streuung vorrangig Endzustände von einem Neutrino, drei Leptonen und mindestens zwei Jets untersucht. Die Jets stammen dabei von den ursprünglichen Quarks, welche die schwachen Eichbosonen abgestrahlt haben. Da diese Quarks zusätzlich auch Gluonen abstrahlen können, sind teilweise auch mehr als nur zwei Jets messbar.

3.2.1 Transversale Masse

Eine wichtige Größe in der Teilchenphysik stellt die invariante Masse m eines Prozesses dar, da sie unabhängig von der Wahl des Bezugssystems und äquivalent zur Schwerpunktenergie ist. Für den Streuprozess $WZ \rightarrow WZ$ ergibt sich die invariante Masse $m(WZ)$ durch die Summe der drei Leptonenergien E^ℓ und der Neutrinoenergie E^ν und der Summe der Leptonenimpulse $\vec{p}^\ell$ und des Neutrinoimpulses $\vec{p}^\nu$.

$$m(WZ) = \sqrt{\left(\sum_{i=1}^3 E^\ell + E^\nu\right)^2 - \left(\sum_{i=1}^3 \vec{p}^\ell + \vec{p}^\nu\right)^2} \quad (3.1)$$

Da sowohl die Energie, als auch der Impuls des Neutrinos nicht vom ATLAS-Detektor direkt gemessen werden können, wird in der Experimentalphysik oft die transversale Masse $M_T(WZ)$ des Vorgangs betrachtet. Um diese zu ermitteln, müssen die Transversalimpulse $\vec{p}_T$ und transversalen Energien $\vec{E}_T$ der Leptonen gemessen werden, welche in der Ebenen senkrecht zur Strahlachse liegen. Wählt man als Strahlachse die z -Achse, ergibt sich die transversale Energie und der Transversalimpuls durch Gleichung (3.2) und (3.3) [9].

$$\vec{p}_T^2 = p_x^2 + p_y^2 \quad (3.2)$$

$$\vec{E}_T^2 = \vec{p}_T^2 + m^2 \quad (3.3)$$

Die übrige Energie der nicht messbaren Teilchen lässt sich anschließend mit $\vec{E}_T^{miss} = -\sum_i \vec{p}_T(i)$ durch Impulserhaltung aus den messbaren Impulsen ermitteln [9]. Nimmt man die Leptonen als masselos an, lässt sich die transversale Masse $M_T(WZ)$ somit allein durch die Impulse der

entstehenden Leptonen definieren [7].

$$M_T(WZ) := \sqrt{\left(\sum_{i=1}^3 p_T^\ell + E_T^{miss}\right)^2 - \left(\sum_{i=1}^3 \vec{p}_T^\ell + \vec{E}_T^{miss}\right)^2} \quad (3.4)$$

3.3 Monte Carlo Simulationen

Sowohl bei der Durchführung, als auch bei der Auswertung von Messprozessen am LHC sind Simulationen unverzichtbar. Dafür werden unter anderem Monte Carlo Generatoren genutzt, welche durch statistische Methoden komplexe Problemstellungen lösen können. Sie werden in vielen Bereichen eingesetzt, in denen keine analytischen Lösungen mehr möglich sind. So können beispielsweise auch Vorhersagen des Standardmodells mittels der Störungstheorie betrachtet werden, was unter anderem eine spezifischere Planung und Auswertung von Experimenten ermöglicht. Zusätzlich zu dem physikalischen Prozess, welcher untersucht werden soll, lässt sich auch die schlussendliche Messung und der Einfluss des Detektors simulieren. Dabei wird eine Kollision mit den dabei stattfindenden Wechselwirkungen, der sich daraus ergebene Teilchenschauer und die anschließende Messung simuliert. Dadurch können durch Monte Carlo Simulationen Pseudo-Experimente durchgeführt werden, was viele Möglichkeiten eröffnet. So können Vorhersagen des Standardmodells oder von Theorien die darüber hinausgehen mit vorhandenen Messwerten verglichen werden und auch Vorgänge, welche noch nicht gemessen werden konnten, in solchen Pseudo-Experimenten berechnet und untersucht werden. [13]

3.4 Entfaltung

Das Ziel der meisten Teilchenphysik-Experimente ist die Messung der Verteilungen von Messgrößen wie zum Beispiel die Impulse, die Energien oder die Streuwinkel der Teilchen bzw. derer Zerfallsprodukte. Diese Messergebnisse müssen anschließend mit den Ergebnissen anderer Experimente oder Theorievorhersagen vergleichbar sein. Jedoch werden die gemessenen Verteilungen auch von dem Detektor selbst beeinflusst, da dieser unter anderem nur eine begrenzte Auflösung und Effizienz besitzt. Somit kommt es, dass die wirkliche Verteilung $f(t)$ der Variablen t als Verteilung $g(s)$ der Messgröße s gemessen wird. Im Folgenden wird $f(t)$ als *truth*-Verteilung und $g(s)$ als *reco*-Verteilung bezeichnet und die Variablen werden als ein-dimensional angenommen. Der Zusammenhang zwischen den beiden Verteilungen lässt sich in Gleichung (3.5) durch ein Fredholm Integral erster Art beschreiben [3, Seite 187]. Dabei stellt $b(s)$ den Messuntergrund der zu messenden Verteilung und $K(s, t)$ die Einflüsse des Messprozesses dar.

$$g(s) = \int K(s, t) f(t) dt + b(s) \quad (3.5)$$

In der Praxis sind die Verteilungen jedoch nicht stetig, sondern liegen in Form von Histogrammen vor, da nur konkrete Einzelereignisse gemessen werden können. Der Übergang von einer stetigen Funktion $h(x)$ zu einer diskretisierten Darstellung η ist dadurch gegeben, dass man die Werte in einem Intervall $[x_{i-1}, x_i]$ in dem Bin η_i zusammenfasst (siehe Gleichung (3.6)). Somit wird ab nun nur noch die diskretisierte *truth*-Verteilung μ bzw. *reco*-Verteilung ν betrachtet. Die Relation in Gleichung (3.5) lässt sich mittels der Response-Matrix R und dem Messuntergrund β in Gleichung (3.7) als Matrixmultiplikation beschreiben.

$$\eta_i := \int_{x_{i-1}}^{x_i} h(x) dx \quad (3.6)$$

$$\nu = R\mu + \beta =: F(\mu) \quad (3.7)$$

Der damit beschriebene Übergang von der wirklichen *truth*-Verteilung μ zu der messbaren *reco*-Verteilung ν wird als Faltung bezeichnet. Die Response-Matrix R und der Messuntergrund β lassen sich dafür durch Monte Carlo Simulationen bestimmen. Dabei werden unter anderem die Detektoreffizienz und die nicht-perfekte Kalibrierung des Detektors, welche zu einer Verschiebung der Verteilung führen kann, betrachtet [3, Seite 188]. Außerdem werden auch statistische Fluktuationen und die Migration zwischen den einzelnen Bins, die durch die beschränkte Detektorauflösung bedingt ist, berücksichtigt.

Die Response-Matrix selbst lässt sich noch in die *fiducial*-Korrektur $fidCorr$, die *efficiency*-Korrektur $effCorr$ und die Migration-Matrix M aufschlüsseln. Die *fiducial*-Korrektur gibt an, welcher Anteil der Einträge in dem jeweiligen Bin der *reco*-Verteilung ν_i auch einen Beitrag in der gesamten *truth*-Verteilung μ liefert, und die *efficiency*-Korrektur beschreibt analog, welcher Anteil von μ_i zu Einträgen in ν führt. Das Migration-Matrixelement M_{ij} gibt schlussendlich den Anteil der Ereignisse des j -ten Bins der *truth*-Verteilung, welche die *efficiency*-Korrektur bereits passiert haben, an, der im i -ten Bin der *reco*-Verteilung einen Beitrag liefert. Die Migration-Matrix beschreibt damit die Migration zwischen den Bins und ist zur besseren Veranschaulichung in Abbildung 3.1 für N_{Jets} dargestellt.

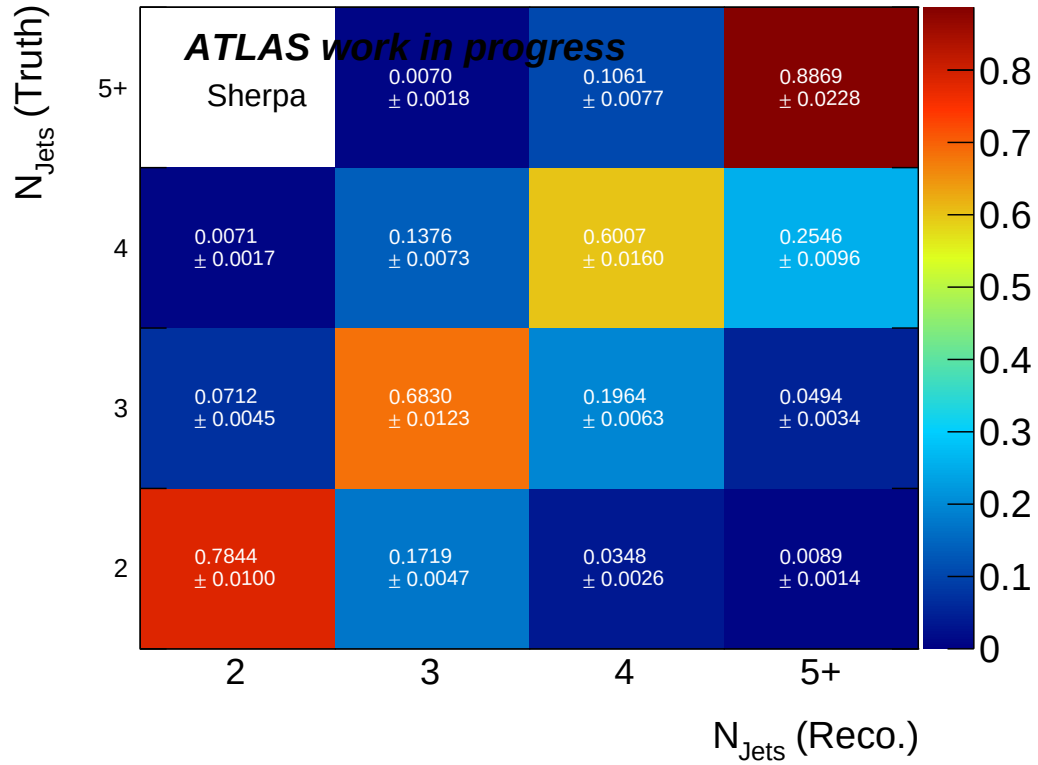


Abbildung 3.1: Migration-Matrix für N_{Jets} wurde von Sherpa ermittelt

Durch die *fiducial*-Korrektur $fidCorr$, die *efficiency*-Korrektur $effCorr$ und die Migration-Matrix M lässt sich die Response-Matrix folgendermaßen darstellen:

$$R_{ij} = \frac{1}{fidCorr_i} \cdot M_{ij} \cdot effCorr_j \quad (3.8)$$

Durch das Falten einer, von der Theorie vorhergesagten, Verteilung lässt sich diese mit den Messwerten vergleichen. Jedoch ist es oft der Fall, dass man die Messwerte mit mehreren Theorien oder auch anderen Messergebnissen vergleichen möchte. Aus diesem Grund nutzt man den umgekehrten Schritt von der *reco*-Verteilung zur *truth*-Verteilung, was als Entfaltung bezeichnet wird. Durch die Entfaltung der Messdaten sind diese nicht mehr detektorabhängig und lassen sich damit besser vergleichen. Zur Realisierung des Entfaltens gibt es verschiedene Methoden, von denen im Folgenden drei vorgestellt werden.

3.4.1 Bin-by-Bin-Entfaltung

Eine einfache Variante Messdaten zu entfalten ist die Bin-by-Bin Methode. Bei dieser Methode muss vorausgesetzt werden, dass die zu ermittelnde *truth*-Verteilung die gleiche Binanzahl wie die *reco*-Verteilung besitzt. Wie in Gleichung (3.9) dargestellt, wird von jedem Bin der *reco*-Verteilung zuerst der Messuntergrund abgezogen und das Ergebnis anschließend mit einem Korrekturfaktor C_i multipliziert [1, Seite 7]. Diese Korrekturfaktoren ergeben sich aus dem

Verhältnis von Einträgen in den jeweiligen Bins der *truth*-Verteilung und der *reco*-Verteilung (ohne Messuntergrund) der Monte Carlo Simulation.

$$\mu_i = C_i(\nu_i - \beta_i) \quad (3.9)$$

$$C_i := \frac{\mu_i^{MC}}{(\nu_i - \beta_i)^{MC}} \quad (3.10)$$

Da bei dieser Methode für jeden Bin μ_i immer nur die Anzahl der Einträge in dem Bin ν_i berücksichtigt werden, muss die Response-Matrix R annähernd diagonal sein. Tritt stärkere Migration zwischen den Bins auf, ist diese Entfaltungsmethode nicht zu empfehlen.

3.4.2 Entfaltung durch Matrix-Invertierung

Eine weitere einfache Methode, welche jedoch die Migration zwischen den Bins berücksichtigt, ist die Matrix-Invertierung. Bildet man das Inverse der Response-Matrix, lässt sich die Gleichung (3.7) nach der *truth*-Verteilung auflösen.

$$\mu = R^{-1}(\nu - \beta) \quad (3.11)$$

Dies kann durch die Methode der Maximum-Likelihood realisiert werden, welche auch für eine unterschiedliche Binaufteilung der *truth*- und *reco*-Verteilung anwendbar ist. Die Likelihood-Funktion $L(x, \theta)$ gibt die Wahrscheinlichkeit dafür an, den Wert x zu messen, in Abhängigkeit von der Wahl des Parameters θ . Wenn für jede Komponente von $x = (x_1, x_2, \dots, x_N)$ die Wahrscheinlichkeitsdichte $f(x, \theta)$ gilt, lässt sich $L(x, \theta)$ durch Gleichung (3.12) ausdrücken [3, Seite 29].

$$L(x, \theta) = \prod_{i=1}^N f(x_i, \theta) \quad (3.12)$$

Nimmt man nun an, dass die gemessene *reco*-Verteilung in jedem einzelnen Bin ν_i um den durch die Theorie vorhergesagten gefalteten Wert $\nu'_i := F(\mu)_i$ Poisson-verteilt ist, lässt sich die Likelihood-Funktion als $L(\nu, \mu)$ formulieren.

$$L(\nu, \mu) = \prod_{i=1}^N \nu_i^{\nu'_i} \frac{e^{-\nu'_i}}{\nu'_i!} \quad (3.13)$$

Da die Likelihood-Funktion für die zur *reco*-Verteilung gehörige *truth*-Verteilung maximal wird, lässt sich μ durch Gleichung (3.14) ermitteln.

$$\frac{\partial \log L(\nu, \mu)}{\partial \mu_i} = 0 \quad \forall i \quad (3.14)$$

Falls die Binanzahl der *truth*- und *reco*-Verteilung übereinstimmt, entspricht die Lösung dieser Gleichung schlussendlich dem Ergebnis der Matrix-Invertierung [1, Seite 6].

3.4.3 Bayesische Entfaltung

Die Bayesische Entfaltung ist eine iterative Entfaltungsmethode. Für diese Methode müssen zuerst die Wahrscheinlichkeit $P(C_i)$ bzw. $P(E_i)$ eingeführt werden, die die Wahrscheinlichkeit darstellen, dass ein Ereignis in dem i -ten Bin der wirklichen bzw. beobachteten Verteilung zu finden ist. Dabei wurden schon die *fiducial*- und *efficiency*-Korrektur mit eingerechnet und der Messuntergrund abgezogen, damit nur Ereignisse betrachtet werden, welche sowohl in der *truth*-Verteilung als auch der *reco*-Verteilung Beiträge liefern. Somit sind die Einträge in den Bins C_i und E_i nur die Ereignisse aus der Schnittmenge der *truth*-Verteilung und *reco*-Verteilung. Die gesamte Ereigniszahl N_{tot} dieser Schnittmenge lässt sich anhand von Gleichung (3.17) ermitteln.

$$P(C_i) := \frac{\mu_i \cdot effCorr_i}{N_{tot}} \quad (3.15)$$

$$P(E_i) := \frac{(\nu_i - \beta_i) \cdot fidCorr_i}{N_{tot}} \quad (3.16)$$

$$N_{tot} := \sum_j \mu_j \cdot effCorr_j = \sum_j (\nu_j - \beta_j) \cdot fidCorr_j \quad (3.17)$$

Die Wahrscheinlichkeit $P(C_i | E_j)$, ein Ereignis im *truth*-Bin C_i anzufinden unter der Bedingung, dass es einen Beitrag im *reco*-Bin E_j liefert, ist durch den Satz von Bayes gegeben.

$$P(C_i | E_j) = \frac{P(E_j | C_i) \cdot P(C_i)}{P(E_j)} \quad (3.18)$$

Die bedingte Wahrscheinlichkeit $P(E_j | C_i)$ entspricht dem Migration-Matrixelement M_{ij} und ist damit bekannt. Die Wahrscheinlichkeiten $P(C_i)$ und $P(E_i)$ sind wiederum durch das Gesetz

der totalen Wahrscheinlichkeit gegeben.

$$P(C_i) = \sum_j P(C_i | E_j) \cdot P(E_j) \quad (3.19)$$

$$P(E_i) = \sum_j P(E_i | C_j) \cdot P(C_j) \quad (3.20)$$

Da die Wahrscheinlichkeit $P(E_j | C_i)$ durch die Migration-Matrix und $P(E_i)$ durch die *reco*-Verteilung gegeben ist, kann man nun $P(C_i)$ verwenden, um die bedingte Wahrscheinlichkeit $P(C_i | E_j)$ zu berechnen. Anschließend kann man mit Hilfe von $P(C_i | E_j)$ auch wieder $P(C_i)$ ermitteln, was das in Gleichung (3.21) dargestellte iterative Vorgehen nahelegt [1, Seite 9]. Dabei wird $P_{Data}(E_i)$ aus der gemessenen *reco*-Verteilung ermittelt und $P_r(E_i)$ ergibt sich anhand von Gleichung (3.20) aus den Werten von $P_r(C_j)$.

$$P_{r+1}(C_i) = \sum_j \frac{P(E_j | C_i) \cdot P_r(C_i)}{P_r(E_j)} \cdot P_{Data}(E_j) \quad (3.21)$$

Somit muss für diese Methode jedoch ein Startwert $P_0(C_i)$ gewählt werden, welcher nach Möglichkeit schon in der Nähe des tatsächlichen Wertes $P(C_i)$ liegen sollte. Dieses iterative Verfahren kann dann abgebrochen werden, wenn das jeweilige Ergebnis relativ konstant bleibt oder ein anderes Abbruchkriterium erfüllt ist.

3.5 Statistische und systematische Unsicherheit

Neben dem Entfaltungsergebnis ist auch die durch die Entfaltung propagierten Unsicherheiten von Interesse, damit die entfalteten Daten im Rahmen dieser Unsicherheit ausgewertet werden können. Die statistischen Unsicherheiten lassen sich durch die Entfaltung von Pseudo-Experimenten bestimmen. In jedem dieser Pseudo-Experimente werden die zu entfaltenden Daten im Rahmen der statistischen Unsicherheit fluktuiert und anschließend entfaltet. Dadurch lässt sich die Unsicherheit des Entfaltungsergebnisses bestimmen, welche durch die statistische Unsicherheit der Daten bedingt ist. [14]

Zusätzlich besitzt die Response-Matrix eine statistische Unsicherheit, da die Monte Carlo Generatoren nur eine endliche Anzahl von Ereignissen simulieren können. Die Response-Matrix wird im Rahmen dieser Unsicherheit variiert und jeweils zur Entfaltung der Daten genutzt [14]. Da die sich daraus ergebende Unsicherheit lediglich durch die beschränkte Anzahl der mittels Monte Carlo simulierten Ereignisse bedingt ist, wird diese zu den systematischen Unsicherheiten gezählt.

Um die restlichen systematischen Unsicherheiten zu bestimmen, benötigt man zwei Gruppen

von Monte Carlo Daten. Eine enthält Daten des nominellen Monte Carlo Generators und die andere besteht aus Datensätzen von jeweils systematisch variierten Monte Carlo Generatoren, welche als Systematiken bezeichnet werden. Die Systematiken ergeben sich durch die Variation der nominellen Simulation innerhalb ihrer systematischen Unsicherheiten. Die Entfaltung wird anschließend sowohl anhand der Response-Matrix des nominellen Monte Carlo Generators als auch mittels der Response-Matrizen, welche aus den Systematiken stammen, durchgeführt [14]. Die einzelnen systematischen Unsicherheiten ergeben sich aus der Abweichung des mittels der jeweiligen Systematik bestimmten Entfaltungsergebnisses zu dem nominellen Entfaltungsergebnis. Schlussendlich lässt sich daraus die gesamte systematische Unsicherheit bestimmen.

4 Hauptteil

In den folgenden Kapiteln wird die Entfaltung der Verteilung der transversalen Masse des Streuprozesses $M_T(WZ)$, des transversalen Impulses des Z -Bosons $P_T(Z)$ und der Jetanzahl N_{Jets} bei der elektroschwachen Streuung von W - und Z -Bosonen durchgeführt. Die verwendeten Entfaltungsmethoden sind dabei Bin-by-Bin-Entfaltung, Matrix-Invertierung und Bayesische Entfaltung und sind bereits in ROOUnfold [6] und EWUnfolding [5] enthalten. Als Prior wird jeweils die *truth*-Verteilung der Monte Carlo Simulation genutzt, welche auch zur Entfaltung verwendet wird. Zur Berechnung der statistischen und systematischen Unsicherheiten wird EWUnfolding und 100 Systematiken (siehe Tabelle 6.7) verwendet. Für den verwendeten Messuntergrund wird außerdem keine zusätzliche Unsicherheit angenommen. Die verwendeten Monte Carlo Generatoren sind dabei Madgraph¹ und Sherpa², wobei aber der gleiche Messuntergrund verwendet wird.

Um den Messuntergrund der zu entfaltenden Daten zu unterdrücken, wurden an die simulierten Ereignisse Selektionskriterien gestellt. Im Endzustand werden mindestens zwei Jets gefordert, welche einen transversalen Impuls von mindestens 40 GeV, eine Pseudorapidität von $|\eta| < 4,5$ und eine invariante Masse von über 500 GeV besitzen. Zusätzlich müssen diese beiden auch in unterschiedliche Hemisphären gerichtet sein und es darf kein Jet auftreten, welcher mit hoher Wahrscheinlichkeit aus einem *bottom*-Quark stammt. Außerdem müssen drei Leptonen (Elektron oder Myon) mit $|\eta| < 2,5$, mittlerer Qualität und einem transversalen Impuls von über 15 GeV gemessen werden, wovon ein Lepton einen transversalen Impuls von mindestens 25 GeV besitzt. Zwei der drei Leptonen müssen ein Lepton-Antilepton-Paar bilden mit $|m_{\ell\ell} - m_Z| < 10$ GeV. Das dritte Lepton muss eine hohe Qualität und einen transversalen Impuls von über 20 GeV besitzen. Das W -Boson, von welchem dieses Lepton stammt, muss eine transversale Masse von mindestens 30 GeV aufweisen. Abschließend darf kein viertes Lepton mit $p_T > 5$ GeV, niedriger Qualität und $|\eta| < 2,5$ auftreten.

Um eine bessere Vorstellung von den verwendeten Pseudodaten in dem dadurch definierten Phasenraum zu bekommen, sind in Abbildung 6.1 exemplarisch die von Sherpa simulierten *reco*-Verteilungen für eine Luminosität von 1200 fb^{-1} dargestellt.

¹MGPY8EG_NNP30LO_A14NNPDF23LO_WZlvlljjEWK, LO

²Sherpa_222_NNP30NNLO_llvjj_EW6, NLO für bis zu ein Parton

4.1 Binoptimierung

Bevor man mit der Auswertung und damit auch mit der Entfaltung der Messdaten beginnen kann, muss man die Diskretisierung der gemessenen Verteilung festlegen. Für die Messung von eindimensionalen Messgrößen, entspricht dies der Binaufteilung in dem jeweiligen Histogramm. Da dabei die Messereignisse zusammengefasst und nicht mehr einzeln ausgewertet werden, hat die Wahl der Binaufteilung großen Einfluss auf die spätere Auswertung und muss abhängig von der zu erwartenden Verteilung getroffen werden.

Die Wahl der optimalen Binaufteilung wird im Folgenden immer anhand von Monte Carlo Verteilungen und allgemeinen Kriterien an die Verteilung getroffen. Außerdem wird sowohl für die *truth*- als auch für die *reco*-Verteilung die gleiche Binaufteilung verwendet.

4.1.1 Mindestanforderungen und Optimierungskriterien

Um die Anzahl der wählbaren Bins von Anfang an einzuschränken, bietet es sich an, Mindestanforderungen an die zu messende Verteilung zu stellen. Diese bilden einen Rahmen für die folgende Optimierung und stellen eine obere Grenze für die Binanzahl dar. Dabei bieten sich Kriterien an, welche entscheidend für die weitere Auswertung der Daten sind.

Damit noch möglichst viel Information in den gemessenen Verteilungen vorhanden ist, muss die Anzahl der Bins, in welche sie aufgeteilt ist, entsprechend groß sein. Aus diesem Grund wird im Folgendem immer die maximale Binanzahl gewählt, welche die Mindestanforderungen noch erfüllt. Somit beschränkt sich die endgültige Optimierung nur noch auf die Lage der Bin-Grenzen.

Eine Mindestanforderung, welche für die folgende Binoptimierung getroffen wurde, bezieht sich auf die Mindestanzahl von Ereignissen in jedem Bin. Es sollten in jedem Bin mindestens so viele Ereignisse vorhanden sein, dass man das zu messende Signal bei statistischen Fluktuationen der gemessenen Daten noch von dem Messuntergrund unterscheiden kann. Nimmt man an, dass die gemessene Gesamtereigniszahl in einem Bin poissonverteilt ist, lässt sich diese Bedingung durch Gleichung (4.1) formulieren.

$$signal - Z \cdot \sqrt{signal + background} > 0 \quad (4.1)$$

Dabei wird gefordert, dass das Signal größer ist als die Z -fache statistische Unsicherheit der Messdaten. Die Wahl eines minimalen Z führt somit zu der jeweiligen statistischen Sicherheit des gemessenen Signals in jedem Bin. Da dieses Kriterium von der Gesamtanzahl der Messereignisse abhängt, hängt es auch direkt von der über die Zeit integrierten Luminosität der Messung ab.

Die zweite Mindestanforderung wird in Hinblick auf die folgende Entfaltung getroffen, indem eine minimale *stability* in jedem Bin gefordert wird. Die *stability* ist der Anteil der Ereignisse in

einem Bin der *truth*-Verteilung, welcher dem gleichen Bin der *reco*-Verteilung zugeordnet wird. Da die *stability* ein Maß für die Migration zwischen den Bins darstellen soll, wird vorher schon die *fiducial*- und die *efficiency*-Korrektur mit eingerechnet. Somit kommen die Ereignisse, für die die *stability* bestimmt werden soll, nur aus der Schnittmenge der Ereignisse in der *truth*- und *reco*-Verteilung.

Je größer die geforderte *stability* in jedem Bin ist, desto mehr gleicht die Migration-Matrix M einer Diagonalmatrix. Eine höhere *stability* and damit eine verminderte Migration zwischen den Bins erleichtert die Entfaltung und verbessert deren Ergebnis.

Diese beiden Mindestanforderungen werden im Rahmen der folgenden Optimierung als ausreichend angesehen, da sie sowohl Bezug zur statistischen Sicherheit der Daten, als auch zur folgenden Entfaltung besitzen.

Damit muss nur noch die optimale Lage der Bins definiert werden. Im Allgemeinen bietet es sich an, eine Funktion in Abhängigkeit von der Binaufteilung zu definieren, welche einen Wert dafür liefert, wie gut diese Aufteilung im Rahmen der Mindestanforderungen getroffen wurde. Diese Funktion kann dann aus verschiedenen Kriterien zusammengesetzt sein und muss nur noch optimiert werden. Für die folgende Optimierung wurde als alleiniges Kriterium angesetzt, dass die Summe der *stability* aller Bins maximal werden soll. So werden die Binpositionen in Hinblick auf die Entfaltung optimiert.

4.1.2 Durchführung der Optimierung

Um die jeweilige Verteilung optimal in Bins aufzuteilen, benötigt man jedoch schon Kenntnisse davon, wie die Ereignisanzahlen verteilt sind und welche Form die Response-Matrix aufweist. Dafür eignet sich eine möglichst feine Anfangsaufteilung in Bins. Die folgende Binoptimierung muss sich anschließend nur noch mit dem optimalen Zusammenfassen der Bins auseinandersetzen.

Zu Beginn muss das Intervall so groß gewählt werden, dass sowohl in der *truth*- als auch *reco*-Verteilung voraussichtlich alle Messereignisse innerhalb dieses Intervalls anzutreffen sind. Für die Anzahl der Jets N_{Jets} reicht ein Intervall von 2 bis 12 Jets aus. Für die transversale Masse $M_T(WZ)$ wird das Intervall von 0 GeV bis 3000 GeV und für den transversalen Impuls des Z -Bosons $P_T(Z)$ das Intervall von 0 GeV bis 1200 GeV gewählt. Die Aufteilung des Intervalls für N_{Jets} ist dadurch limitiert, dass nur eine Aufteilung in ganze Zahlen sinnvoll ist. Die Verteilungen von $M_T(WZ)$ und $P_T(Z)$ lassen sich hingegen beliebig fein auflösen. Die feine anfängliche Binaufteilung ließe sich selbst wiederum optimieren, da in Bereichen von zum Beispiel hoher Ereigniszahlen eine feinere Aufteilung nötig ist als in Bereichen niedriger Ereigniszahlen. Im Folgenden soll es jedoch für die Verteilungen von $M_T(WZ)$ und $P_T(Z)$ ausreichen, das erste Drittel der Intervalle in 100 gleichgroße Bins aufzugliedern und die restlichen Zweidrittel in 50 ebenfalls gleichgroße Bins zu unterteilen. Alle folgenden Ergebnisse der Optimierung beruhen demzufolge auf der Annahme, dass dies die feinst mögliche Aufteilung sei. Eine noch feine-

re und der jeweiligen Verteilung besser angepasste Anfangsaufteilung wäre damit noch eine zusätzliche Verbesserung des Optimierungsverfahrens.

Die Optimierung selbst beginnt mit dem Abschätzen der maximalen Binanzahl und der ungefähren Binpositionen, anhand der durch das jeweilige Z gewählten statistischen Mindestanforderung. Dafür können im Intervall aufsteigend die einzelnen Bins zu jeweils einem Bin zusammengefasst werden und sobald die statistische Mindestanforderung erfüllt ist, beginnt der nächste Bin. Dies macht man solange bis am Ende des Intervalls nur noch ein Bereich übrig ist, der die Mindestanforderung nicht erfüllt, und der Rest des Intervalls in Bins aufgeteilt ist, welche die Mindestanforderung gerade so noch erfüllen.

Dies lässt sich wiederholen, indem man nun anfängt das Intervall beginnend mit den hohen Werten zu Bins zusammenzufassen. Dabei verbleibt nun am Anfang des Intervalls ein Bereich, welcher nicht genug Werte besitzt, um die Bedingung zu erfüllen. Dadurch kann man nicht nur die maximale Binanzahl ermitteln, für die das Signal außerhalb der Z -fachen statistischen Unsicherheitsgrenzen liegt, sondern auch welchen Spielraum die Bingrenzen besitzen. Die Bingrenzen können nicht niedriger gewählt werden als in der Binaufteilung, welche bei den niedrigsten Werten begonnen hat, und nicht höher als in der Binaufteilung, die mit den höchsten Werten begonnen hat. Jegliche Bingrenzen können nur innerhalb dieses Rahmens gewählt werden.

Die Mindestanforderung durch die minimale *stability* lässt sich hingegen nicht im Vorfeld schon so leicht betrachten, da nicht nur Bereiche der eindimensionalen *truth*- und *reco*-Verteilung sondern auch Teile der Response-Matrix zusammengefasst werden. Somit wird diese Bedingung in die Optimierung der Binpositionen übernommen. Diese Optimierung wird nun für die maximale Binanzahl durchgeführt, bei denen das Signal in jedem Bin außerhalb der statistischen Z Unsicherheitsgrenzen liegt. Dabei werden alle Binaufteilungen durchprobiert, welche die beiden Mindestanforderungen erfüllen, und die Aufteilung mit der größten Summe der *stability* der Bins wird als optimale Binaufteilung verwendet. Falls keine Variante gefunden wurde, welche beide Mindestanforderungen erfüllt, wird die Binanzahl solange verringert bis eine optimale Aufteilung gefunden werden kann, die diese erfüllt.

4.1.3 Ergebnisse der Binoptimierung

Da das beschriebene Verfahren erlaubt, für eine gegebene über die Zeit integrierte Luminosität und geforderte Mindestanforderungen die Binanzahl zu berechnen, lässt sich dadurch auch die minimale integrierte Luminosität ermitteln, welche man für eine bestimmte Binanzahl und Mindestanforderungen benötigen würde. Dies ist in den Tabellen 6.1 bis 6.6 für N_{Jets} , $M_T(WZ)$ und $P_T(Z)$ für einige Werte exemplarisch dargestellt. Die minimale integrierte Luminosität wurde dabei bis zu einer oberen Grenze von 5000 fb^{-1} ermittelt. Liegt die integrierte Luminosität über dieser Grenze, kann dies bedeuten, dass entweder die benötigte integrierte Luminosität einfach nur sehr hoch ist oder aber auch, dass eine solche Aufteilungen nicht

möglich ist.

In den Tabellen 6.1 bis 6.6 fällt auf, dass bei kleinen Binanzahlen hauptsächlich die Mindestanforderung durch das Z die minimale integrierte Luminosität bestimmt. Mit steigender Binanzahl gewinnt jedoch die minimale *stability* an Bedeutung und das umso stärker, je weniger die jeweilige anfängliche Migrationsmatrix einer Diagonalmatrix ähnelt.

Mit Hilfe der Binoptimierung können nun die Binaufteilungen für die Monte Carlo Daten, welche im Folgenden entfaltet werden sollen, gewählt werden. Dabei werden exemplarisch die Verteilungen mit der jeweiligen minimalen integrierten Luminosität entfaltet, welche für drei Bins, $Z = 2,0$ und *stability* $> 0,3$ nötig ist. Zusätzlich sollen die Verteilungen auch bei einer integrierten Luminosität von 1200 fb^{-1} entfaltet werden, da bei dieser hohen integrierten Luminosität für alle Verteilungen eine größere Binanzahl gewählt werden kann und die Form der jeweiligen Verteilung dadurch besser ersichtlich ist. Als Mindestanforderungen werden für N_{Jets} $Z = 3,0$ und *stability* $> 0,5$ und für $M_T(WZ)$ und $P_T(Z)$ $Z = 3,5$ und *stability* $> 0,6$ verwendet. Die Ergebnisse der Binoptimierung sind in Tabelle 4.1 dargestellt.

Verteilung	$\int L dt$ in fb^{-1}	Bingrenzen
Sherpa		
$M_T(WZ)$	133	[0; 200; 280; 3000] GeV
$P_T(Z)$	133	[0; 80; 160; 1200] GeV
N_{Jets}	184	[1,5; 2,5; 3,5; 12,5]
Madgraph		
$M_T(WZ)$	97	[0; 200; 280; 3000] GeV
$P_T(Z)$	97	[0; 80; 164; 1200] TeV
N_{Jets}	162	[1,5; 2,5; 3,5; 12,5]
Sherpa		
$M_T(WZ)$	1200	[0; 170; 220; 270; 320; 430; 3000] GeV
$P_T(Z)$	1200	[0; 40; 64; 84; 112; 152; 196; 280; 1200] GeV
N_{Jets}	1200	[1,5; 2,5; 3,5; 4,5; 12,5]
Madgraph		
$M_T(WZ)$	1200	[0; 160; 200; 250; 310; 370; 480; 3000] GeV
$P_T(Z)$	1200	[0; 36; 56; 72; 88; 108; 132; 156; 196; 244; 324; 1200] GeV
N_{Jets}	1200	[1,5; 2,5; 3,5; 4,5; 12,5]

Tabelle 4.1: Die dargestellten Bingrenzen sind das Ergebnis der Binoptimierung für unterschiedliche integrierten Luminositäten und Mindestanforderungen. Bei den Verteilungen mit einer integrierten Luminosität von 1200 fb^{-1} wurde für N_{Jets} $Z = 3,0$ und *stability* $> 0,5$ und für $M_T(WZ)$ und $P_T(Z)$ $Z = 3,5$ und *stability* $> 0,6$ verwendet. Bei den Binoptimierungen mit kleinerer integrierter Luminosität gilt $Z = 2,0$ und *stability* $> 0,3$.

4.2 Technical Closure Test

Bevor man mit der Entfaltung von Daten beginnt, ist es im Allgemeinen erforderlich die verwendete Entfaltungsmethode zu testen. Einer der wichtigsten Tests stellt dabei der Technical Closure Test dar, welcher fundamentale Probleme und Bugs aufzeigen kann. Bei diesem Test werden die gleichen Monte Carlo Daten entfaltet, welche auch zur Bestimmung der Response-Matrix verwendet wurden [1, Seite 17]. Wichtig ist, dass bei der Entfaltung keine zusätzliche Regularisierung der Daten vorgenommen wird. Das Entfaltungsergebnis muss mit der Monte Carlo Vorhersage der *truth*-Verteilung übereinstimmen, damit der Test als bestanden angesehen werden kann. Signifikante Abweichungen zwischen dem Entfaltungsergebnis und der Monte Carlo *truth*-Verteilung bedeuten, dass bei der Entfaltung ein grundlegendes Problem aufgetreten ist. Dies könnte beispielsweise ein technischer Bug im benutzten Entfaltungsprogramm sein, aber auch ein Problem bei der Wahl der Binaufteilung sein. Wenn zum Beispiel das gewählte Intervall für eine Verteilung zu klein ist, können Ereignisse aus den *Overflow*-Bins in die betrachteten Bins migrieren, was das Entfaltungsergebnis verfälscht.

Dieser Test wurde auf die in Tabelle 4.1 aufgeführten Verteilungen durchgeführt. Die jeweilige relative Abweichung des Entfaltungsergebnisses von der Monte Carlo Vorhersage ist in den Abbildungen 6.2, 6.3 und 6.4 für Sherpa und in den Abbildungen 6.5, 6.6 und 6.7 für Madgraph zusehen. Für alle Verteilungen liegt die maximale relative Abweichung im Bereich von 10^{-6} und ist damit verschwindend gering. Diese Abweichungen könnten unter anderem durch Rundungsfehler aufgetreten sein, da z.B. vor der Entfaltung die zu entfaltende Verteilung von einem ROOT-Histogramm in ein python numpy-array und anschließend wieder in ein ROOT-Histogramm umgewandelt wurde.

Da die Abweichungen jedoch nur sehr gering sind, können sie als unbedenklich und der Technical Closure Test damit als bestanden angesehen werden.

4.3 Entfaltungsergebnisse

Da der Technical Closure Test bestanden wurde, kann nun die Entfaltung von Daten folgen. Als Daten werden dabei jeweils die *reco*-Verteilung des jeweils anderen Monte Carlo Generator zuzüglich des Messuntergrundes verwendet. Exemplarisch ist in Abbildung 4.1 das Entfaltungsergebnis der Madgraph *reco*-Verteilung der transversalen Masse $M_T(WZ)$ bei einer integrierten Luminosität von 1200 fb^{-1} abgebildet. Zur Entfaltung wurden dabei die Ergebnisse der Sherpa Simulation genutzt. Der verwendete Prior ist die Sherpa *truth*-Verteilung, welcher zusammen mit der Madgraph *truth*-Verteilung ebenfalls in Abbildung 4.1 dargestellt ist. Zusätzlich zu den Entfaltungsergebnissen der einzelnen Entfaltungsmethoden sind auch die jeweiligen Unsicherheiten eingezeichnet, welche sich aus der statistischen und systematischen Unsicherheit ergeben.

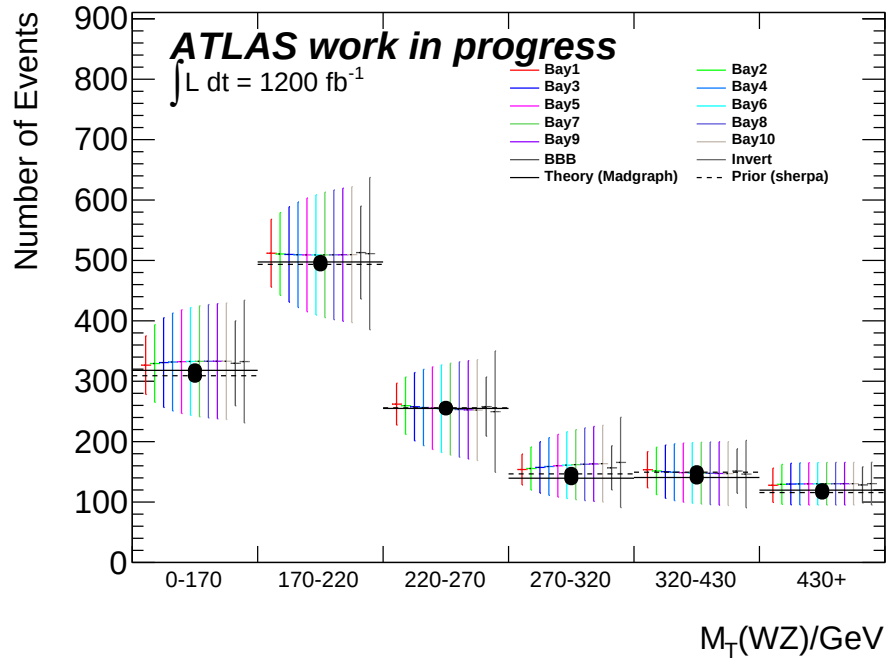


Abbildung 4.1: Dargestellt ist das Entfaltungsergebnis der Madgraph *reco*-Verteilung für $M_T(WZ)$ bei einer integrierten Luminosität von 1200 fb^{-1} . Entfaltet wurde dabei mittels Sherpa und die eingezeichneten Unsicherheiten umfassen sowohl die statistische als auch systematische Unsicherheit.

Welchen Beitrag die statistische und systematische Unsicherheit jeweils für die Gesamtunsicherheit liefern, ist in Abbildung 4.2 für die Entfaltung der Madgraph *reco*-Verteilung bei 1200 fb^{-1} erkennbar. Die dargestellte Entfaltungsmethode ist die Matrix-Invertierung mittels der Response-Matrix von Sherpa.

Die von EWUnfolding berechnete Gesamtunsicherheit wird zum Großteil durch die statistische Unsicherheit dominiert. Eine größere Anzahl an verwendeten Systematiken würde somit zwar die systematische Unsicherheit noch leicht erhöhen, hätte aber wohl nur einen sehr geringen Einfluss auf die Gesamtunsicherheit. Dies gilt auch für die Unsicherheiten der anderen Entfaltungsmethoden unabhängig davon ob Sherpa oder Madgraph als Daten bzw. für die Bestimmung der Response-Matrix genutzt wurde und für geringere integrierte Luminositäten steigt der Beitrag der statistischen Unsicherheit noch weiter an.

Die restlichen Entfaltungsergebnisse inklusive ihrer Gesamtunsicherheiten sind in den Abbildungen 6.8, 6.9 und 6.10 für die Entfaltung der Madgraph *reco*-Verteilungen mittels Sherpa und in den Abbildungen 6.11, 6.12 und 6.13 für die Entfaltung der Sherpa *reco*-Verteilungen mittels Madgraph zu finden. Vergleicht man die Bayesische Entfaltungsmethode mit der Entfaltung durch Matrix-Invertierung fällt auf, dass sowohl die Entfaltungsergebnisse der Bayesischen Methode als auch deren Unsicherheiten gegen die Ergebnisse bzw. Unsicherheiten der Matrix-Invertierung streben. Dies war zu erwarten, da im Allgemeinen die Lösung durch das Bayesische Verfahren mit steigender Iterationsanzahl gegen die Lösung durch Matrix-Invertierung kon-

vergiert [8]. Nach schon drei oder vier Iterationen stimmt für die betrachteten Verteilungen das Entfaltungsergebnis der Bayesischen Entfaltungsmethode sehr gut mit dem Entfaltungsergebnis der Matrix-Invertierung überein.

Bezüglich der Gesamtunsicherheit ist außerdem auch ersichtlich, dass die Bin-by-Bin-Entfaltung und die Bayesische Entfaltung mit nur einer Iteration die geringste Propagation statistischer und systematischen Unsicherheiten besitzen. Dass dies bei der Wahl der gewünschten Entfaltungsmethode jedoch nicht das einzige Kriterium sein sollte, wird im folgenden Kapitel deutlich.

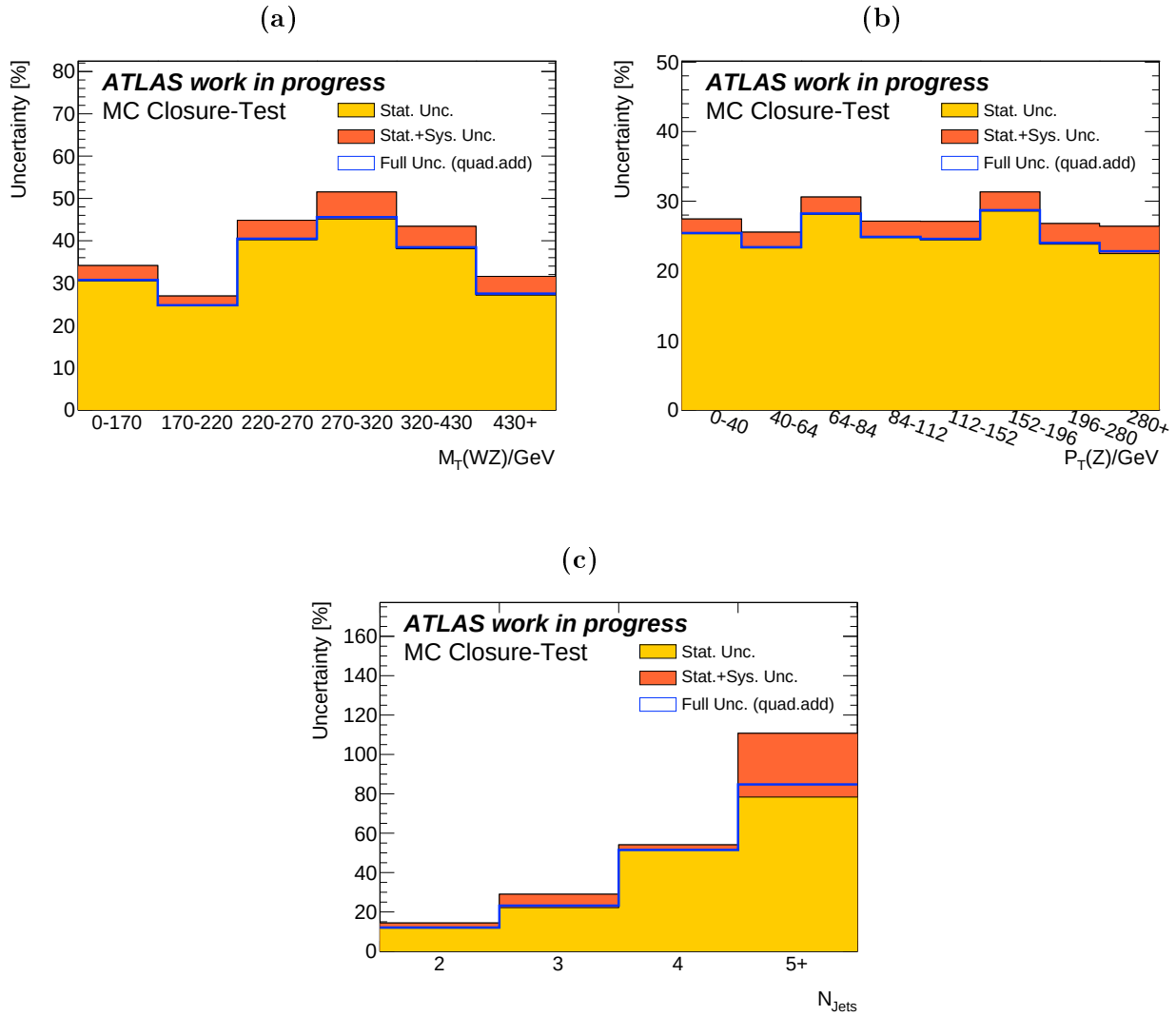


Abbildung 4.2: Zu sehen ist die ermittelte Unsicherheit der Entfaltungsergebnisse der Madgraph *reco*-Verteilungen für $M_T(WZ)$ (a), $P_T(Z)$ (b) und N_{Jets} (c) bei einer integrierten Luminosität von 1200 fb^{-1} . Entfaltet wurde dabei durch Matrix-Invertierung mittels der Response-Matrix von Sherpa.

4.4 Entfaltungsunsicherheit

Wie in Kapitel 3.5 beschrieben, lassen sich die statistischen und systematischen Unsicherheiten durch den Entfaltungsprozess propagieren. Jedoch führt auch der Entfaltungsprozess an sich zu einer zusätzlichen Unsicherheit, welche durch den Data-driven Closure Test bestimmt werden kann. Bei diesem Test wird die Monte Carlo *truth*-Verteilung so umgewichtet, dass sie nach der Faltung zu den Daten passt. Die Umgewichtung der einzelnen Bineinträge findet mittels einer Umgewichtungsfunktion statt, welche möglichst glatt sein soll. Diese Forderung an die Umgewichtungsfunktion hat zur Folge, dass das anschließende Faltungsergebnis nach der Umgewichtung nicht exakt mit den Daten übereinstimmen muss, sondern lediglich eine signifikant verbesserte Übereinstimmung gefordert wird [1, Seite 19]. Die Glattheit der Umgewichtungsfunktion wird gefordert, um den Einfluss der statistischen Fluktuation in den Daten zu verringern. Anschließend wird die umgewichtete gefaltete Verteilung wieder mit der jeweiligen Entfaltungsmethode entfaltet. Für diese Entfaltung wird die Response-Matrix verwendet, welche aus der Monte Carlo Vorhersage ohne Umgewichtung stammt. Abschließend lässt sich aus der Differenz zwischen dem Entfaltungsergebnis und der ursprünglichen umgewichteten *truth*-Verteilung die Entfaltungsunsicherheit abschätzen [1, Seite 20].

4.4.1 Bestimmung der Umgewichtungsfunktion

Der erste Schritt bei der Umgewichtung der *truth*-Verteilung ist es, die Forderungen an die Umgewichtungsfunktion genauer festzulegen. Zum Vergleich der gefalteten umgewichteten *truth*-Verteilung $F(Rew(\mu))$ (inklusive des Messuntergrundes) mit den Daten wird ein Chi-Quadrat-Test zwischen $F(Rew(\mu))$ und den Daten durchgeführt, wobei $F(Rew(\mu))$ als Erwartungswert betrachtet wird. Da der Chi-Quadrat-Test eine Gaußverteilung um den Erwartungswert in jedem einzelnen Bin annimmt, wird damit die statistische Unsicherheit der Daten berücksichtigt. Das Umgewichtete $Rew(\mu)$ wird dadurch realisiert, dass jeder Bineintrag der *truth*-Verteilung μ_i mit einem Umgewichtungsfaktor rew_i multipliziert wird. Die gesamte Verteilung lässt sich somit durch das Produkt $S_{rew}\mu$ umgewichten, wobei S_{rew} eine Diagonalmatrix mit den Elementen der Umgewichtungsfaktoren rew ist.

Die Glattheit der Umgewichtungsfunktion lässt sich aus der numerischen zweiten Ableitung der Umgewichtungsfaktoren rew bestimmen. Die numerische erste Ableitung ist dabei durch Gleichung (4.2) definiert und die numerische zweite Ableitung ergibt sich rekursiv daraus. Als Maß für die Glattheit wird wie in Gleichung (4.3) die Wurzel der Betragsquadrate der numerischen zweiten Ableitungen genommen.

$$(d^1rew)_i := rew_{i+1} - rew_i \quad (4.2)$$

$$|d^2rew| := \sqrt{\sum_i (d^2rew)_i^2} \quad (4.3)$$

Somit lässt sich nun sowohl ein Wert für die Übereinstimmung der gefalteten umgewichteten *truth*-Verteilung $F(Rew(\mu))$ und den Daten als auch für die Glattheit der Ungewichtungsfunktion bestimmen. Diese beiden Eigenschaften lassen sich nun in der Optimierungsfunktion $K_{opt}(S_{rew})$ zusammenführen, welche einen Wert dafür liefert, wie gut die Ungewichtungsfaktoren die gestellten Forderungen erfüllt. Der Glättungsparameter λ bestimmt dabei wie hoch die Priorität der Glattheit ist.

$$K_{opt}(S_{rew}) = \chi^2(data, F(S_{rew}\mu)) + \lambda \cdot |d^2rew| \quad (4.4)$$

Je besser die Forderungen an die Ungewichtungsfunktion erfüllt sind, desto kleiner wird der Wert, welcher von der Optimierungsfunktion $K_{opt}(S_{rew})$ ausgegeben wird. Somit lassen sich die Ungewichtungsfaktoren durch eine Minimierung von $K_{opt}(S_{rew})$ bestimmen.

Bevor man jedoch mit dem Ungewichten beginnen kann, muss man eine Wahl für den Glättungsparameter λ treffen. Eine allgemeine Festlegung auf einen Wert für λ wäre dabei nicht empfehlenswert, da die Minimierung der Optimierungsfunktion stark von der jeweilig betrachteten Verteilung und integrierten Luminosität abhängt. Besser wäre es Kriterien festzulegen, mit denen sich ein optimaler Glättungsparameter λ definieren lässt, und diesen dann für die jeweilige Verteilung zu ermitteln.

Da die Ähnlichkeit zu den Daten und die Glattheit der Ungewichtungsfunktion die Hauptforderungen an die Ungewichtungsfunktion sind, bieten sie sich auch als Kriterien für den optimalen Glättungsparameter an. Dieser soll im Folgenden so gewählt werden, dass die Priorität der Glattheit solange erhöht wird, wie der p-Wert von $\chi^2(data, F(S_{rew}\mu))$ nach der Minimierung von $K_{opt}(S_{rew})$ noch oberhalb eines Minimalwertes liegt. Dadurch wird gewährleistet, dass die *truth*-Verteilung durch eine möglichst glatte Funktion umgewichtet wird.

Um zu testen, wie stark der optimale Glättungsparameter durch statistische Fluktuationen beeinflusst wird, wurde dieser exemplarisch für die $M_T(WZ)$ -Verteilung bei einer integrierten Luminosität von 1000 fb^{-1} und 1500 fb^{-1} berechnet. Als Daten wurden jeweils Verteilungen verwendet, welche in jedem Bin um die Madgraph *reco*-Verteilung im Rahmen einer Poissonstatistik fluktuiert wurden. Die Response-Matrix und die umzugewichtende *truth*-Verteilung wurden mittels Sherpa bestimmt und die verwendete Binaufteilung entspricht der aus Tabelle 4.1 für Sherpa und einer integrierten Luminosität von 1200 fb^{-1} . Die Verteilung der bestimmten optimalen Glättungsparameter sind für 10000 Fluktuationen in den Abbildungen 4.3a und 4.3b zu sehen.

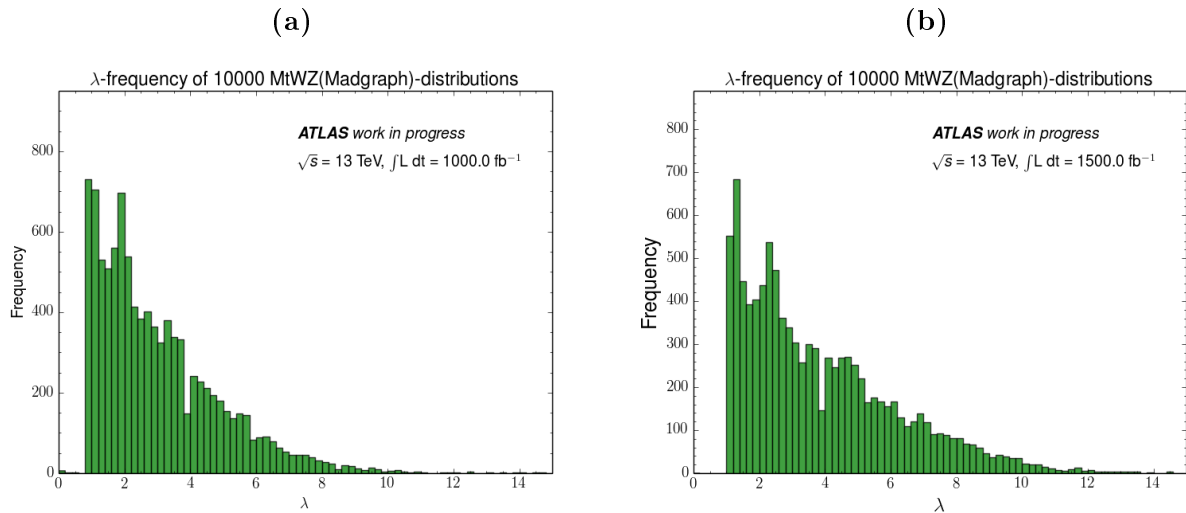


Abbildung 4.3: Dargestellt ist die absolute Häufigkeit des ermittelten optimalen Glättungsparameters λ für 10000 Fluktuationen der $M_T(WZ)$ Verteilung bei einer integrierten Luminosität von 1000 fb^{-1} in (a) und 1500 fb^{-1} in (b).

Aus den Abbildungen 4.3a und 4.3b ist ersichtlich, dass der optimale Glättungsparameter λ trotz Fluktuationen innerhalb der gleichen Größenordnung bleibt. Die Priorität der Glattheit ist somit bei der Ungewichtung verschiedener Daten zum gleichen Erwartungswert recht ähnlich. Um die Ungewichtungsfunktion für eine *truth*-Verteilung und einen Satz von Daten zu bestimmen, kann man so einen Parameter λ ermitteln, der für diese Verteilungen passend wäre. Ein möglicher Nachteil dabei ist jedoch, dass die Priorität der Glattheit zu hoch gewählt wird, falls die Daten schon sehr gut mit der Monte Carlo *reco*-Verteilung übereinstimmen. Dadurch kann es passieren, dass die Ungewichtung die Übereinstimmung zu den Daten nur noch geringfügig verbessert. Da in einem solchen Fall jedoch keine signifikante Ungewichtung mehr nötig ist und ein sehr großer Wert für λ sehr unwahrscheinlich ist (siehe Abbildung 4.3a und 4.3b), sollte dieser Spezialfall einer zu großen Priorität der Glattheit kein Problem darstellen.

4.4.2 Test der Bestimmung der Ungewichtungsfunktion

Die Ungewichtung soll die Monte Carlo *truth*-Verteilung möglichst gut an die Daten anpassen und dabei zusätzlich den Einfluss der statistischen Fluktuation der Daten minimieren. Um dies zu testen, werden Sherpa Daten für die *truth*-Verteilung und die Ermittlung der Response-Matrix verwendet und als Daten wird eine Fluktuation um eine leicht veränderte Madgraph *reco*-Verteilung betrachtet. So soll die Bestimmung der Ungewichtungsfunktion für einen Fall getestet werden, bei dem die theoretische Vorhersage signifikant von der den Daten zugrunde liegenden Theorie abweicht.

Durchgeführt wird dies für $M_T(WZ)$ bei einer integrierten Luminosität von 1200 fb^{-1} und der dazu passenden Binaufteilung in Tabelle 4.1 für Sherpa. An die Bineinträge der Madgraph

reco-Verteilung werden nun die Faktoren $[1,1; 0,9; 0,8; 1,0; 1,2; 1,4]$ multipliziert, um eine von der Sherpa *reco*-Verteilung deutlicher unterscheidbare Verteilung zu erhalten. Die poisson-fluktuierten Werte stellen damit Daten zu einer Theorie dar, welche leicht abweichend zu der für Sherpa verwendeten Theorie ist. Durch die Umgewichtungsfunktion soll die Sherpa *truth*-Verteilung so umgewichtet werden, dass sie nach der Faltung zu der veränderten Madgraph *reco*-Verteilung passt und dabei der statistischen Fluktuation, welche auf die Verteilung angewendet wurde, möglichst wenig nachgeht.

Um zu testen, wie gut die Umgewichtung die statistische Fluktuation der Daten vernachlässigt hat, wird die gefaltete umgewichtete *truth*-Verteilung ($F(Rew(\mu))$) mittels Chi-Quadrat-Test mit weiteren statistischen Fluktuationen um die veränderte Madgraph *reco*-Verteilung verglichen. Würde die Umgewichtung die Fluktuation der Daten vollständig vernachlässigen, entspräche ($F(Rew(\mu))$) der veränderten Madgraph *reco*-Verteilung und die p-Werte der Chi-Quadrat-Tests zu den anderen Fluktuationen ergäbe dann eine Gleichverteilung.

Da es aber unmöglich ist anhand einer Fluktuation die exakte ursprüngliche Verteilung zu ermitteln, wird für ($F(Rew(\mu))$) im Folgenden lediglich eine möglichst gleichmäßigere Verteilung dieser p-Werte als bei einer Umgewichtung ohne zusätzliche Glattheit der Umgewichtungsfunktion gefordert. Um dies zu testen, wurde die Umgewichtung mittels der Optimierungsfunktion aus Gleichung (4.4) mit verschiedenen Werten für λ inklusive des als optimal definierten λ durchgeführt. Dadurch lässt sich auch überprüfen, wie stark der Einfluss der Glattheit der Umgewichtungsfunktion ist. Dies wurde für insgesamt 20 Fluktuationen um die veränderte *reco*-Verteilung wiederholt. Ein Beispiel ist in Abbildung 4.4 dargestellt, um zu vergleichen ob das Umgewichten mittels des berechneten optimalen Glättungsparameter λ eine signifikant bessere Übereinstimmung zu der eigentlichen *reco*-Verteilung besitzt als die exakte Umgewichtung ohne zusätzliche Glattheit.

Die Verteilung der p-Werte bei der Umgewichtung mit Glättungsterm in Abbildung 4.4b besitzt einen eindeutig ebeneren Verlauf als ohne Glättungsterm in Abbildung 4.4a. Dies war bei fast allen der betrachteten 20 Verteilungen der Fall, von denen in Abbildung 6.14 noch drei weitere exemplarisch dargestellt sind. Aus diesen Graphiken lässt sich schlussfolgern, dass die gewählte Umgewichtungsmethode einen gewissen Anteil der statistischen Fluktuationen vernachlässigt und so einen klaren Vorteil gegenüber einer exakten Umgewichtung besitzt.

Der Einfluss unterschiedlicher Glättungsparametern ist in den Abbildungen 6.15 bis 6.18 erkennbar. Sie wurden aus den 20 Fluktuationen so ausgewählt, dass jeweils ein anderer Einfluss des Glättungsparameters gezeigt wird, und sind somit nicht repräsentativ für alle 20 Fluktuationen.

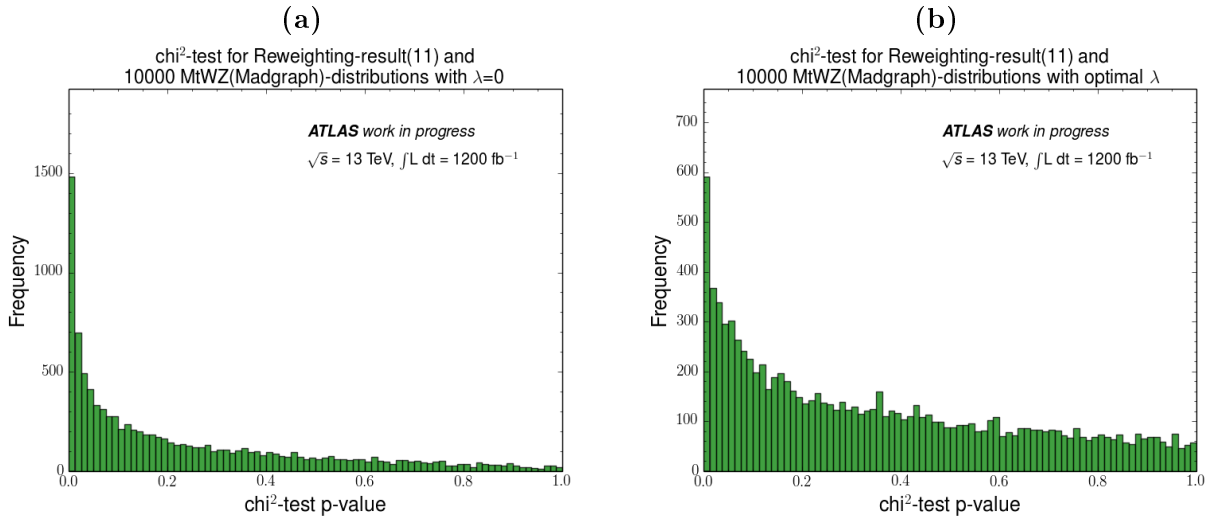


Abbildung 4.4: Abgebildet ist die p-Wert-Verteilung des Chi-Quadrat-Tests zwischen der gefalteten umgewichteten Sherpa *truth*-Verteilung von $M_T(WZ)$ und 10000 Fluktuationen um eine leicht veränderte Madgraph *reco*-Verteilung. Verglichen wird dabei eine Ungewichtung ohne die Berücksichtigung der Glattheit (a) und mit der Wahl des optimalen Glättungsparameters (b). Die Ungewichtungsfunktion wurde dabei für eine Fluktuation um die veränderte Madgraph *reco*-Verteilung bestimmt.

In Abbildung 6.15 ist zum Beispiel erkennbar, dass es auch Fluktuationen geben kann, bei denen eine zusätzliche Glattheit der Ungewichtungsfunktion keine Verbesserung darstellt. In diesem Beispiel lag die Fluktuation in allen Bins oberhalb der eigentlichen *reco*-Verteilung, was somit nicht durch eine zusätzliche Glattheit ausgeglichen werden kann.

Die Fluktuation, die in Abbildung 6.16 untersucht wird, ließe sich am besten mit einem Glättungsparameter von etwa $\lambda = 8$ ausgleichen. Bei dieser Fluktuation führt eine gewisse Glattheit zur deutlich besseren Übereinstimmung mit der *reco*-Verteilung, wohingegen eine zu hohe Glattheit diese Übereinstimmung wieder verfälscht.

In Abbildung 6.17 stimmte die Fluktuation von Anfang an schon sehr gut mit der eigentlichen *reco*-Verteilung überein und eine zusätzliche Glattheit verschlechtert so nur das Ergebnis.

Im Kontrast dazu nimmt in Abbildung 6.18 die Übereinstimmung zwischen der gefalteten umgewichteten *truth*-Verteilung und der *reco*-Verteilung mit steigendem Glättungsparameter immer weiter zu.

Der Großteil der 20 Fluktuationen wies das in Abbildung 6.18 bzw. Abbildung 6.16 dargestellte Verhalten auf, bei dem die Übereinstimmung zwischen dem gefalteten Ungewichtungsergebnis und der veränderten Madgraph *reco*-Verteilung bis $\lambda = 16$ immer weiter zu nimmt bzw. sich schon wieder verringert.

In welcher Größenordnung der Glättungsparameter liegt, der zu der höchsten Übereinstimmung mit der *reco*-Verteilung führt, ist damit recht stark von der jeweiligen statistischen Fluktuation abhängig. Zusätzlich kommt voraussichtlich auch noch ein großer Einfluss davon,

wie gut die simulierte *reco*-Verteilung mit der tatsächlich zu messenden Verteilung übereinstimmt. Die Wahl eines festen Wertes für λ wäre somit nur sehr schwer abzuschätzen und wahrscheinlich nicht zielführend. Da die Umgewichtung mit dem als optimal definierten Glättungsparameters zu einer signifikanten Vernachlässigung der statistischen Fluktuationen führt, liefert im Allgemeinen die Minimierung der Optimierungsfunktion K_{opt} mittels dieses optimalen Glättungsparameters einen sinnvollen Wert für die Umgewichtungsfunktion. Ob es noch bessere Methoden der Umgewichtung gibt, steht noch aus und müsste somit noch geprüft werden, jedoch soll diese Methode für die folgende Berechnung der Entfaltungsunsicherheit als ausreichend angesehen werden.

4.4.3 Berechnung der Entfaltungsunsicherheit

Mittels der gewählten Umgewichtungsmethode lässt sich die Entfaltungsunsicherheit durch das bloße Anwenden der Entfaltungsmethode ermitteln. Zuvor wird die *truth*-Verteilung der Simulation, welche auch für den Entfaltungsprozess genutzt wird, so umgewichtet, dass sie zu den Daten passt. Als Daten werden dabei die Pseudodaten (*reco*-Verteilung) des jeweils anderen Monte Carlo Generators verwendet. Anschließend wird die umgewichtete *truth*-Verteilung mittels der zur Entfaltung verwendeten Response-Matrix gefaltet und danach mit der jeweiligen Entfaltungsmethode wieder entfaltet. Die Abweichung des Entfaltungsergebnisses zur umgewichteten *truth*-Verteilung ist dann ein Maß für die Entfaltungsunsicherheit.

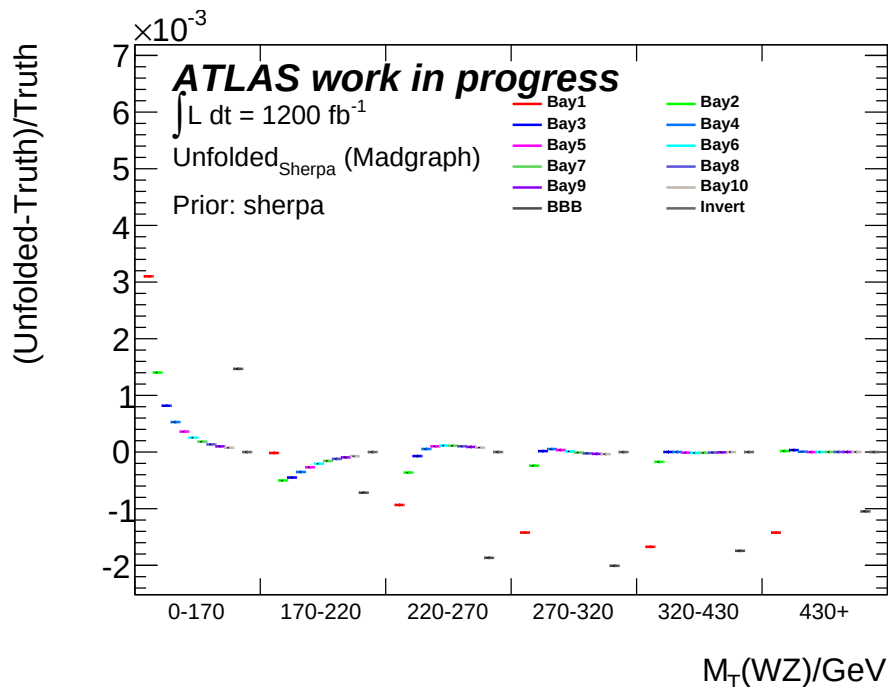


Abbildung 4.5: Zu sehen ist die relative Abweichung des Entfaltungsergebnisses von $M_T(WZ)$ von der *truth*-Verteilung. Die *truth*-Verteilung entspricht der Sherpa *truth*-Verteilung, welche im Bezug zur Madgraph *reco*-Verteilung umgewichtet wurde. Entfaltet wurde die entsprechende Faltung der umgewichteten *truth*-Verteilung.

In Abbildung 4.5 ist exemplarisch die Entfaltungsunsicherheit für die Entfaltung der $M_T(WZ)$ Madgraph *reco*-Verteilung mittels Sherpa abgebildet.

Die restlichen Entfaltungsergebnisse sind in den Abbildungen 6.19, 6.20 und 6.21 für die Entfaltung der Madgraph *reco*-Verteilungen und in den Abbildungen 6.22, 6.23 und 6.24 für die Entfaltung der Sherpa *reco*-Verteilungen dargestellt. Es fällt auf, dass die ermittelten Entfaltungsunsicherheiten größtenteils deutlich geringer sind als die bisher ermittelte Gesamtunsicherheit. Im Allgemeinen liegt die Entfaltungsunsicherheit unterhalb von wenigen Prozent oder ist noch kleiner. Jedoch ist beispielsweise aus Abbildung 6.21b ersichtlich, dass der Entfaltungsprozess auch zu größeren Unsicherheiten führen kann, welche signifikant gegenüber den systematischen und statistischen Unsicherheiten sind.

Vergleicht man die Unsicherheiten der einzelnen Entfaltungsmethoden untereinander, wird wieder deutlich, dass die Entfaltungsunsicherheit der Bayesischen Methode gegen die Entfaltungsunsicherheit der Matrix-Invertierung konvergiert. Die Entfaltungsunsicherheit der Matrix-Invertierung ist dabei zwangsweise gleich null, da sie lediglich die Umkehroperation zur Faltung mittels Response-Matrix darstellt. Die Bin-by-Bin-Entfaltung und die Bayesische Entfaltung mit wenigen Iterationen führen zu den größten Entfaltungsunsicherheiten, wobei jedoch schon nach etwa zwei oder drei Iterationen die Entfaltungsunsicherheit der Bayesischen Methode stark abgefallen ist.

Legt man das Hauptaugenmerk auf die Entfaltungsunsicherheiten der Bayesischen Methode mit mindestens zwei Iterationen, besitzt im Allgemeinen die Entfaltung von $M_T(WZ)$ eine größere Entfaltungsunsicherheit als bei $P_T(Z)$ und die Entfaltung von N_{Jets} besitzt die höchsten Unsicherheiten. Dies resultiert daraus, dass bei $P_T(Z)$ bzw. auch $M_T(WZ)$ das Entfaltungsergebnis der Bayesischen Methode deutlich stärker gegen die Lösung der Matrix-Invertierung konvergiert. Begründet werden kann dies dadurch, dass die Response-Matrix von $P_T(Z)$ die höchste *stability* aufweist, wohingegen die *stability* bei N_{Jets} am geringsten ist. Eine höhere *stability* verringert damit die auftretende Entfaltungsunsicherheit.

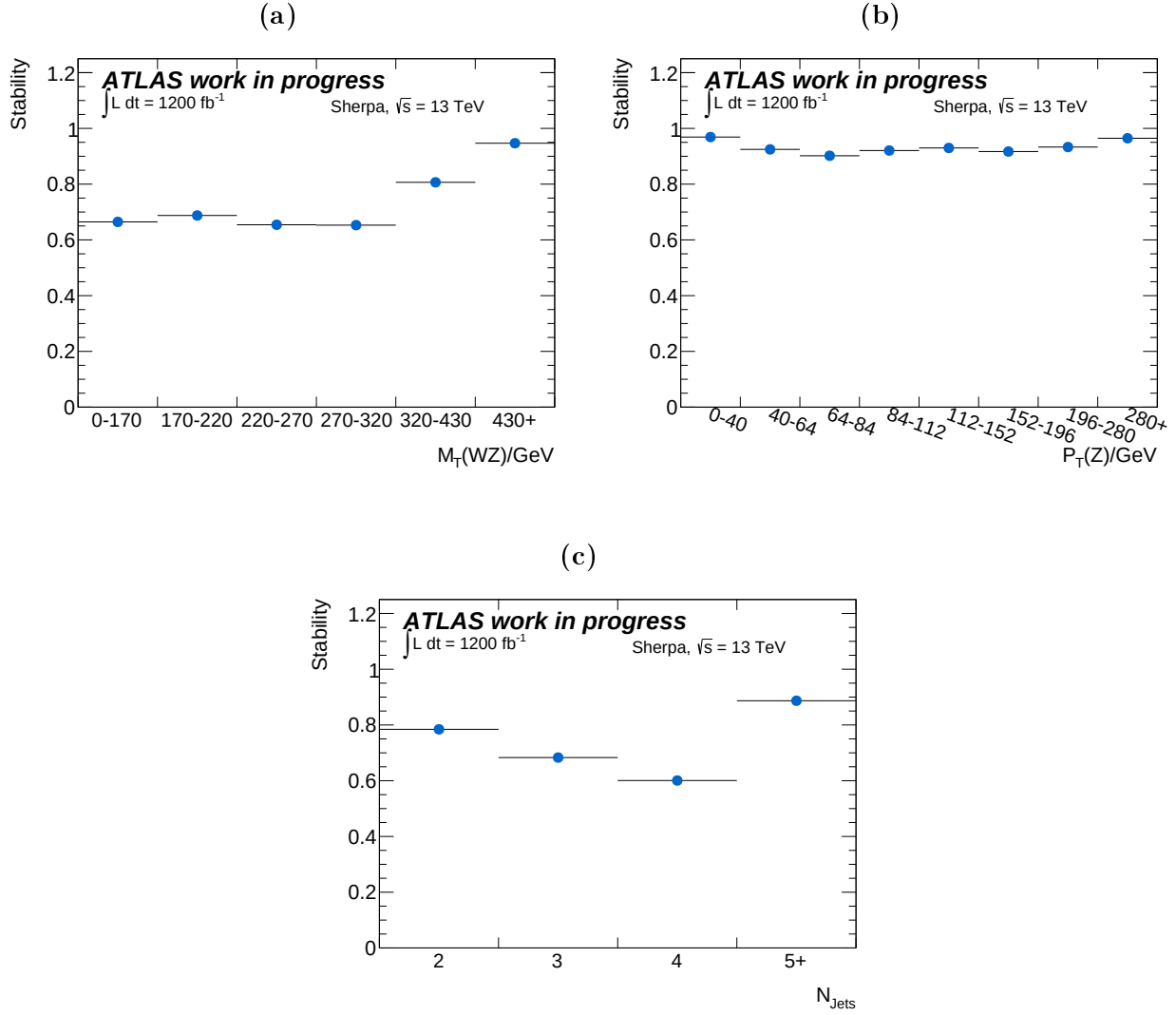


Abbildung 4.6: Die dargestellte *stability* ergibt sich aus der jeweiligen Response-Matrix der Sherpa Simulation für $M_T(WZ)$ (a), $P_T(Z)$ (b) und N_{Jets} (c).

4.5 Entfaltung fluktuierter Monte Carlo Daten

Da die bisherige Entfaltung sich ausschließlich auf das Entfalten von Monte Carlo Daten beschränkte, wurde eine gute Vergleichbarkeit des Entfaltungsergebnisses mit der theoretischen Verteilung ermöglicht. Jedoch wurde dadurch immer nur eine Monte Carlo *reco*-Verteilung entfaltet, welche lediglich den simulierten Erwartungswerte der Messverteilung darstellt. Um abzuschätzen wie im Vergleich dazu die Entfaltungsergebnisse und Entfaltungsunsicherheiten bei echten gemessenen Daten sein könnten, wird nun eine statistische Fluktuation auf die Monte Carlo Daten angewendet und die daraus resultierenden Verteilungen als Pseudodaten verwendet.

Entfaltet werden die $M_T(WZ)$ -, $P_T(Z)$ - und N_{Jets} -Verteilung bei 1200 fb^{-1} mittels der durch

Sherpa bestimmten Response-Matrix. Die zu entfaltenden Daten sind Monte Carlo Madgraph *reco*-Verteilungen, welche in jedem Bin im Rahmen einer Poissonstatistik fluktuiert wurden. Als Binaufteilung wurden die entsprechenden Ergebnisse der Binoptimierung aus Tabelle 4.1 verwendet. In Abbildung 4.7 ist exemplarisch das Entfaltungsergebnis von $M_T(WZ)$ dargestellt.

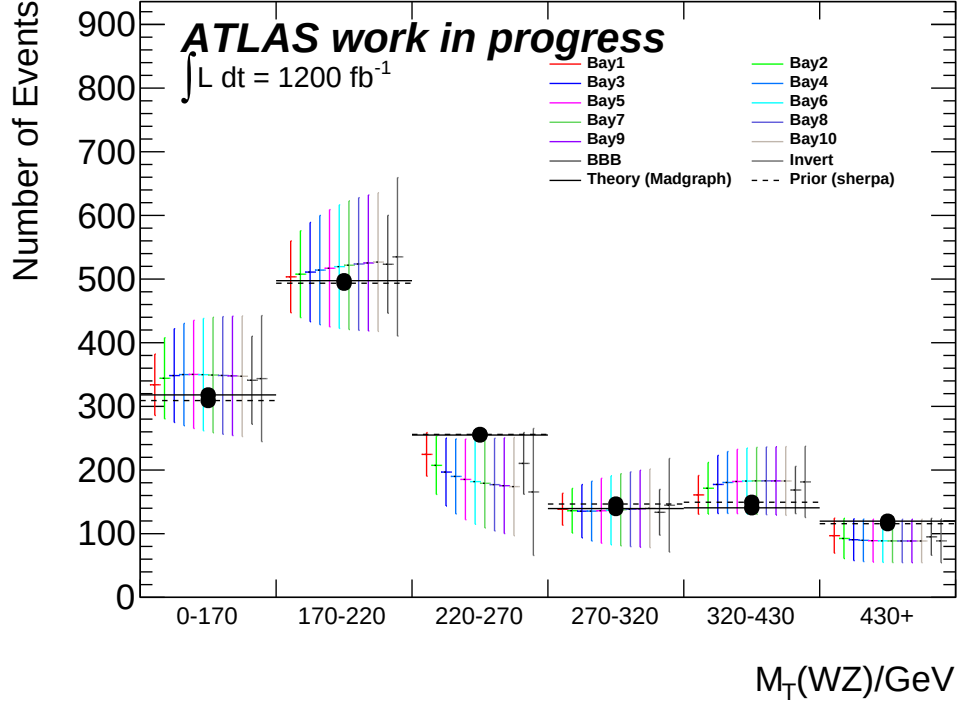


Abbildung 4.7: Dargestellt ist das mittels Sherpa durch verschiedene Entfaltungsmethoden ermittelte Entfaltungsergebnis inklusive der statistischen und systematischen Unsicherheiten. Als Daten wurden Fluktuationen der Madgraph *reco*-Verteilung von $M_T(WZ)$ bei 1200 fb^{-1} verwendet.

Aufgrund der statistischen Fluktuation weicht das Entfaltungsergebnis deutlicher von der theoretischen Vorhersage ab, als es in Abbildung 6.8b der Fall war. Jedoch bleibt diese Abweichung im Rahmen der dargestellten Unsicherheiten. Auch für $P_T(Z)$ und N_{Jets} treten in der Abbildung 6.25 diese statistischen Schwankungen auf. Die theoretischen Werte liegen bei den Großteil der Bins innerhalb der Unsicherheitsgrenzen und weichen nur in zwei Bins von Abbildung 6.25a etwas stärker von dem theoretischen Wert ab.

Die Entfaltungsunsicherheit für $M_T(WZ)$ ist in Abbildung 4.8 dargestellt und die Entfaltungsunsicherheiten für $P_T(Z)$ und N_{Jets} sind in Abbildung 6.26 zu finden.

Die Entfaltungsunsicherheit liefert für $M_T(WZ)$ und $P_T(z)$ einen höheren Beitrag als für die Entfaltung der unfluktuierten Monte Carlo Daten, welcher aber noch deutlich unter der statistischen Unsicherheit liegt. Die Entfaltungsunsicherheit von N_{Jets} blieb in der gleichen Größenordnung. Der Anstieg der Entfaltungsunsicherheit ist mit der größeren Formabweichung zwischen den zu entfaltenden Daten und der *reco*-Verteilung, des für die Bestimmung der

Response-Matrix genutzten Monte Carlo Generators, zu begründen.

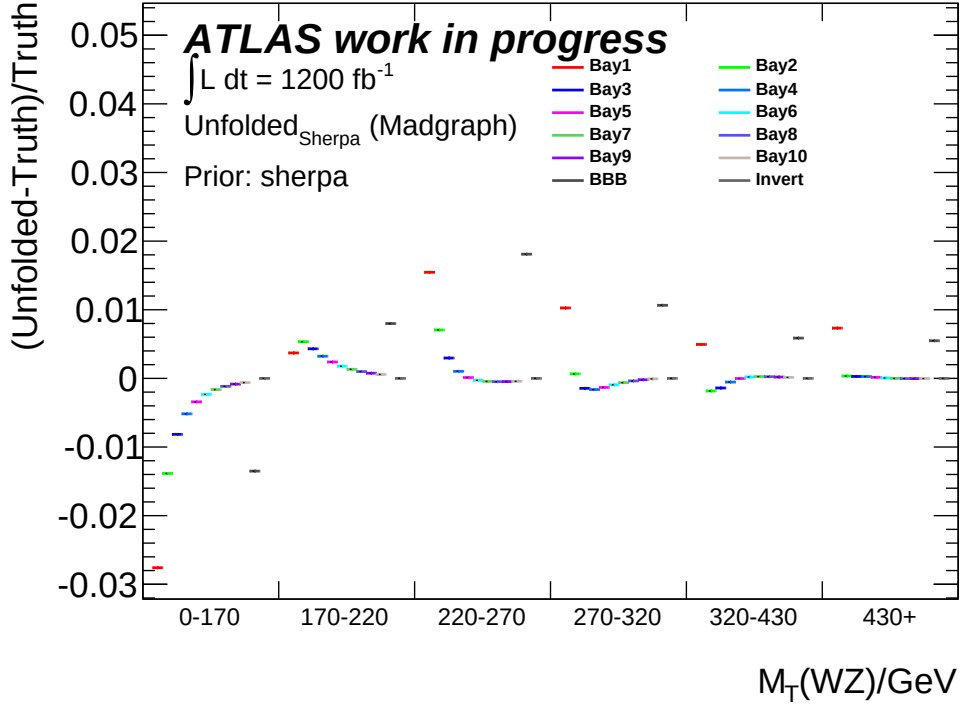


Abbildung 4.8: Die dargestellte Entfaltungsunsicherheit wurde für die Entfaltung mittels Sherpa bestimmt. Als Daten wurden Fluktuationen der Madgraph *reco*-Verteilung von $M_T(WZ)$ bei 1200 fb^{-1} verwendet.

Im Allgemeinen kann man jedoch schlussfolgern, dass im Falle von $M_T(WZ)$, $P_T(Z)$ und N_{Jets} in der elektroschwachen Streuung von W - und Z -Bosonen bei einer integrierten Luminosität von 1200 fb^{-1} voraussichtlich die Entfaltungsunsicherheiten kleiner als die statistischen Unsicherheiten sind, aber schon einen signifikanten Beitrag zur Gesamtunsicherheit liefern können. Für kleinere Luminositäten ist jedoch wieder zu erwarten, dass die statistischen Unsicherheiten deutlicher überwiegen.

5 Zusammenfassung

Ziel dieser Arbeit war es die Entfaltung für kinematische Verteilungen in der elektroschwachen Streuung von W - und Z -Bosonen durchzuführen, wobei der Schwerpunkt auf der Binoptimierung und der Bestimmung der Entfaltungsunsicherheit lag. Dazu wurden Monte Carlo Pseudodaten von Sherpa und Madgraph der transversalen Masse der Streuung $M_T(WZ)$, des transversalen Impulses des Z -Bosons $P_T(Z)$ und der Anzahl der Jets N_{Jets} bei unterschiedlichen integrierten Luminositäten entfaltet.

Begonnen wurde mit der Binoptimierung, bei welcher eine minimale statistische Sicherheit der Daten und eine minimale *stability* in jedem Bin gefordert und dann die dafür maximal möglich Binanzahl gewählt wurde. Die Binpositionen wurden so gewählt, dass die Summe der *stability* aus allen Bins maximal wird. Da diese Kriterien recht allgemein gehalten sind, wären sie auch für andere Verteilungen anwendbar. Zusätzlich ließe sich die beschriebene Binoptimierung auch um weitere Kriterien erweitern. Falls die Binoptimierung jedoch beispielsweise in Hinblick auf die Gesamtunsicherheit des Entfaltungsergebnisses inklusive der Entfaltungsunsicherheit optimiert werden soll, ist das beschriebene Vorgehen nicht mehr anwendbar, da dann die Entfaltung selbst Teil der Binoptimierung werden würde.

Außerdem ließ sich aus den Mindestanforderungen an die Binaufteilung die für die jeweilige Verteilung minimal notwendige integrierte Luminosität abschätzen, welche für die untersuchten Verteilungen im Allgemeinen in einem sehr hohen Bereich liegt. Jedoch ist zu betonen, dass die entfalteten Verteilungen von $M_T(WZ)$ und $P_T(Z)$ in der elektroschwachen Streuung von W - und Z -Bosonen mit drei Bins, $Z = 2$ und *stability* $> 0,3$ mit der integrierten Luminosität von 150 fb^{-1} im Run 2 des LHC realisiert werden können [4].

Zur Entfaltung selbst wurde die Bin-by-Bin-Entfaltung, die Entfaltung durch Matrix-Invertierung und die Bayesische Entfaltung mit unterschiedlicher Iterationsanzahl verwendet. Aus den Entfaltungsergebnissen ist ersichtlich wie stark die statistischen und systematischen Unsicherheiten von den jeweiligen Entfaltungsmethoden propagiert werden. Außerdem wird durch die verwendeten Beispiele auch verdeutlicht, dass die Entfaltungsergebnisse der Bayesischen Entfaltung mit steigender Iterationszahl gegen die Entfaltungsergebnisse der Matrix-Invertierung konvergieren.

Bei der Berechnung der Entfaltungsunsicherheit wurde eine Methode zur Bestimmung der Ungewichtungsfunktion dargestellt und anschließend an einer Verteilung hinsichtlich ihres Verhaltens gegenüber statistischen Fluktuationen der Daten getestet. Weiterführend müsste

noch geprüft werden, wie sich diese Methode für andere Verteilungen und Theorien verhält, und auch der Vergleich zu anderen Vorgehensweisen für die Ungewichtung wäre von Interesse. Die ermittelten Entfaltungsunsicherheiten zeigten eine deutliche Abhängigkeit von der jeweiligen Entfaltungsmethode und lagen für die untersuchten Verteilungen im Allgemeinen deutlich unter dem Anteil der statistischen Unsicherheit.

Die in dieser Arbeit durchgeführte Studie zur Entfaltung kinematischer Verteilungen lässt sich über die exemplarisch untersuchten Verteilungen hinaus anwenden. Das Vorgehen bei der Bin-optimierung, der Entfaltung, dem Technical Closure Test und dem Data-driven Closure Test wurden allgemein gehalten und sind damit voraussichtlich auch für die Entfaltung anderer Verteilungen anwendbar.

6 Anhang

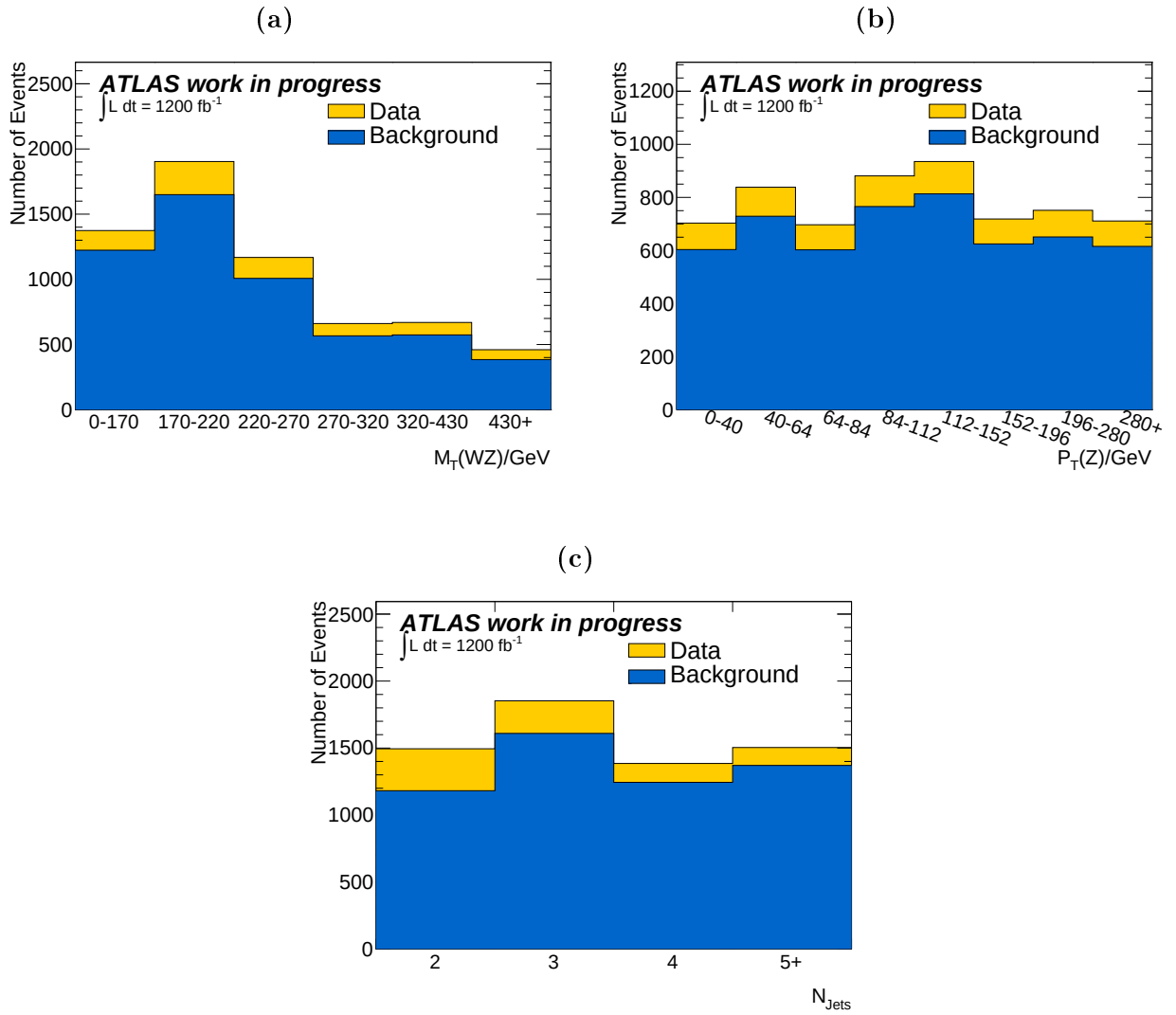


Abbildung 6.1: Die Pseudodaten mit Messuntergrund entsprechen den mittels Sherpa simulierten *reco*-Verteilungen von $M_T(WZ)$ (a), $P_T(Z)$ (b) und N_{Jets} (c) für eine integrierte Luminosität von 1200 fb^{-1} .

$M_T(WZ)$ Sherpa								
$\int Ldt$ in fb^{-1}		<i>stability</i>						
Binanzahl	Z	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	1,5	75	75	75	75	75	96	247
	2,0	133	133	133	133	133	170	440
	2,5	208	208	208	208	208	266	687
	3,0	299	299	299	299	299	384	990
	3,5	408	408	408	408	408	522	1347
	4,0	532	532	532	532	532	682	1759
4	1,5	108	108	108	108	120	197	1177
	2,0	192	192	192	192	213	351	2092
	2,5	300	300	300	300	333	549	3268
	3,0	432	432	432	432	479	790	4706
	3,5	588	588	588	588	652	1075	>5000
	4,0	768	768	768	768	852	1404	>5000
5	1,5	129	129	129	139	197	412	3243
	2,0	229	229	229	247	350	733	>5000
	2,5	358	358	358	385	548	1145	>5000
	3,0	515	515	515	555	788	1649	>5000
	3,5	701	701	701	755	1073	2244	>5000
	4,0	916	916	916	986	1402	2931	>5000
6	1,5	162	162	162	197	301	823	>5000
	2,0	289	289	289	350	535	1463	>5000
	2,5	451	451	451	548	835	2286	>5000
	3,0	650	650	650	788	1203	3292	>5000
	3,5	884	884	884	1073	1637	4481	>5000
	4,0	1154	1154	1154	1402	2139	>5000	>5000
7	1,5	188	188	215	260	500	1786	>5000
	2,0	334	334	383	462	889	3175	>5000
	2,5	521	521	598	722	1389	4961	>5000
	3,0	751	751	861	1040	2000	>5000	>5000
	3,5	1022	1022	1172	1415	2722	>5000	>5000
	4,0	1334	1334	1530	1848	3555	>5000	>5000

Tabelle 6.1: In dieser Tabelle sind die minimalen integrierten Luminositäten $\int Ldt$ in fb^{-1} für $M_T(WZ)$ der Sherpa Verteilung aufgeführt. Berechnet wurden sie für verschiedene Mindestanforderungen und Binanzahlen.

$P_T(Z)$ Sherpa								
$\int Ldt$ in fb^{-1}		<i>stability</i>						
Binanzahl	Z	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	1,5	75	75	75	75	75	75	75
	2,0	133	133	133	133	133	133	133
	2,5	208	208	208	208	208	208	208
	3,0	300	300	300	300	300	300	300
	3,5	408	408	408	408	408	408	408
	4,0	533	533	533	533	533	533	533
4	1,5	100	100	100	100	100	100	100
	2,0	177	177	177	177	177	177	177
	2,5	277	277	277	277	277	277	277
	3,0	399	399	399	399	399	399	399
	3,5	543	543	543	543	543	543	543
	4,0	710	710	710	710	710	710	710
5	1,5	128	128	128	128	128	128	128
	2,0	227	227	227	227	227	227	227
	2,5	354	354	354	354	354	354	354
	3,0	510	510	510	510	510	510	510
	3,5	694	694	694	694	694	694	694
	4,0	907	907	907	907	907	907	907
6	1,5	151	151	151	151	151	151	151
	2,0	268	268	268	268	268	268	268
	2,5	419	419	419	419	419	419	419
	3,0	603	603	603	603	603	603	603
	3,5	821	821	821	821	821	821	821
	4,0	1072	1072	1072	1072	1072	1072	1072
7	1,5	182	182	182	182	182	182	182
	2,0	323	323	323	323	323	323	323
	2,5	505	505	505	505	505	505	505
	3,0	727	727	727	727	727	727	727
	3,5	990	990	990	990	990	990	990
	4,0	1292	1292	1292	1292	1292	1292	1292

Tabelle 6.2: In dieser Tabelle sind die minimalen integrierten Luminositäten $\int Ldt$ in fb^{-1} für $P_T(Z)$ der Sherpa Verteilung aufgeführt. Berechnet wurden sie für verschiedene Mindestanforderungen und Binanzahlen.

N_{Jets} Sherpa								
$\int Ldt$ in fb^{-1}		<i>stability</i>						
Binanzahl	Z	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	1,5	103	103	103	103	227	1274	>5000
	2,0	184	184	184	184	404	2265	>5000
	2,5	287	287	287	287	631	3539	>5000
	3,0	413	413	413	413	908	>5000	>5000
	3,5	563	563	563	563	1236	>5000	>5000
	4,0	735	735	735	735	1615	>5000	>5000
4	1,5	227	227	227	227	1274	>5000	>5000
	2,0	404	404	404	404	2265	>5000	>5000
	2,5	631	631	631	631	3539	>5000	>5000
	3,0	908	908	908	908	>5000	>5000	>5000
	3,5	1236	1236	1236	1236	>5000	>5000	>5000
	4,0	1615	1615	1615	1615	>5000	>5000	>5000
5	1,5	502	502	502	1274	>5000	>5000	>5000
	2,0	892	892	892	2265	>5000	>5000	>5000
	2,5	1393	1393	1393	3539	>5000	>5000	>5000
	3,0	2006	2006	2006	>5000	>5000	>5000	>5000
	3,5	2730	2730	2730	>5000	>5000	>5000	>5000
	4,0	3566	3566	3566	>5000	>5000	>5000	>5000
6	1,5	1274	1274	1274	>5000	>5000	>5000	>5000
	2,0	2265	2265	2265	>5000	>5000	>5000	>5000
	2,5	3539	3539	3539	>5000	>5000	>5000	>5000
	3,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	3,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	4,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
7	1,5	3189	3189	3189	>5000	>5000	>5000	>5000
	2,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	2,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	3,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	3,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	4,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000

Tabelle 6.3: In dieser Tabelle sind die minimalen integrierten Luminositäten $\int Ldt$ in fb^{-1} für N_{Jets} der Sherpa Verteilung aufgeführt. Berechnet wurden sie für verschiedene Mindestanforderungen und Binanzahlen.

$M_T(WZ)$ Madgraph								
$\int Ldt$ in fb^{-1}		<i>stability</i>						
Binanzahl	Z	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	1,5	54	54	54	54	54	71	214
	2,0	97	97	97	97	97	127	380
	2,5	151	151	151	151	151	198	594
	3,0	217	217	217	217	217	285	855
	3,5	296	296	296	296	296	388	1163
	4,0	387	387	387	387	387	507	1519
4	1,5	79	79	79	79	85	142	859
	2,0	141	141	141	141	151	253	1527
	2,5	220	220	220	220	236	395	2385
	3,0	317	317	317	317	340	569	3435
	3,5	431	431	431	431	462	775	4675
	4,0	563	563	563	563	603	1012	>5000
5	1,5	95	95	95	99	142	277	1778
	2,0	169	169	169	176	253	492	3160
	2,5	264	264	264	275	395	769	4938
	3,0	381	381	381	397	569	1108	>5000
	3,5	518	518	518	540	775	1508	>5000
	4,0	677	677	677	705	1012	1970	>5000
6	1,5	115	115	115	142	239	602	4595
	2,0	205	205	205	253	425	1071	>5000
	2,5	321	321	321	395	664	1673	>5000
	3,0	462	462	462	569	957	2409	>5000
	3,5	629	629	629	775	1302	3279	>5000
	4,0	821	821	821	1012	1700	4283	>5000
7	1,5	142	152	153	198	392	1277	>5000
	2,0	253	270	272	352	697	2271	>5000
	2,5	395	422	424	550	1089	3548	>5000
	3,0	569	608	611	792	1568	>5000	>5000
	3,5	775	828	831	1078	2134	>5000	>5000
	4,0	1012	1081	1086	1408	2788	>5000	>5000

Tabelle 6.4: In dieser Tabelle sind die minimalen integrierten Luminositäten $\int Ldt$ in fb^{-1} für $M_T(WZ)$ der Madgraph Verteilung aufgeführt. Berechnet wurden sie für verschiedene Mindestanforderungen und Binanzahlen.

$P_T(Z)$ Madgraph								
$\int Ldt$ in fb ⁻¹		<i>stability</i>						
Binanzahl	Z	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	1,5	55	55	55	55	55	55	55
	2,0	97	97	97	97	97	97	97
	2,5	152	152	152	152	152	152	152
	3,0	218	218	218	218	218	218	218
	3,5	297	297	297	297	297	297	297
	4,0	388	388	388	388	388	388	388
4	1,5	73	73	73	73	73	73	73
	2,0	129	129	129	129	129	129	129
	2,5	202	202	202	202	202	202	202
	3,0	291	291	291	291	291	291	291
	3,5	397	397	397	397	397	397	397
	4,0	518	518	518	518	518	518	518
5	1,5	92	92	92	92	92	92	92
	2,0	164	164	164	164	164	164	164
	2,5	256	256	256	256	256	256	256
	3,0	368	368	368	368	368	368	368
	3,5	501	501	501	501	501	501	501
	4,0	655	655	655	655	655	655	655
6	1,5	112	112	112	112	112	112	112
	2,0	198	198	198	198	198	198	198
	2,5	310	310	310	310	310	310	310
	3,0	447	447	447	447	447	447	447
	3,5	608	608	608	608	608	608	608
	4,0	794	794	794	794	794	794	794
7	1,5	130	130	130	130	130	130	130
	2,0	232	232	232	232	232	232	232
	2,5	362	362	362	362	362	362	362
	3,0	521	521	521	521	521	521	521
	3,5	709	709	709	709	709	709	709
	4,0	927	927	927	927	927	927	927

Tabelle 6.5: In dieser Tabelle sind die minimalen integrierten Luminositäten $\int Ldt$ in fb⁻¹ für $P_T(Z)$ der Madgraph Verteilung aufgeführt. Berechnet wurden sie für verschiedene Mindestanforderungen und Binanzahlen.

N_{Jets} Madgraph								
$\int Ldt$ in fb^{-1}		<i>stability</i>						
Binanzahl	Z	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	1,5	91	91	91	91	293	>5000	>5000
	2,0	162	162	162	162	520	>5000	>5000
	2,5	253	253	253	253	813	>5000	>5000
	3,0	364	364	364	364	1170	>5000	>5000
	3,5	496	496	496	496	1592	>5000	>5000
	4,0	648	648	648	648	2080	>5000	>5000
4	1,5	293	293	293	1006	>5000	>5000	>5000
	2,0	520	520	520	1788	>5000	>5000	>5000
	2,5	813	813	813	2793	>5000	>5000	>5000
	3,0	1170	1170	1170	4022	>5000	>5000	>5000
	3,5	1592	1592	1592	>5000	>5000	>5000	>5000
	4,0	2080	2080	2080	>5000	>5000	>5000	>5000
5	1,5	1006	1006	1006	>5000	>5000	>5000	>5000
	2,0	1788	1788	1788	>5000	>5000	>5000	>5000
	2,5	2793	2793	2793	>5000	>5000	>5000	>5000
	3,0	4022	4022	4022	>5000	>5000	>5000	>5000
	3,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	4,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
6	1,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	2,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	2,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	3,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	3,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	4,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
7	1,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	2,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	2,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	3,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	3,5	>5000	>5000	>5000	>5000	>5000	>5000	>5000
	4,0	>5000	>5000	>5000	>5000	>5000	>5000	>5000

Tabelle 6.6: In dieser Tabelle sind die minimalen integrierten Luminositäten $\int Ldt$ in fb^{-1} für N_{Jets} der Madgraph Verteilung aufgeführt. Berechnet wurden sie für verschiedene Mindestanforderungen und Binanzahlen.

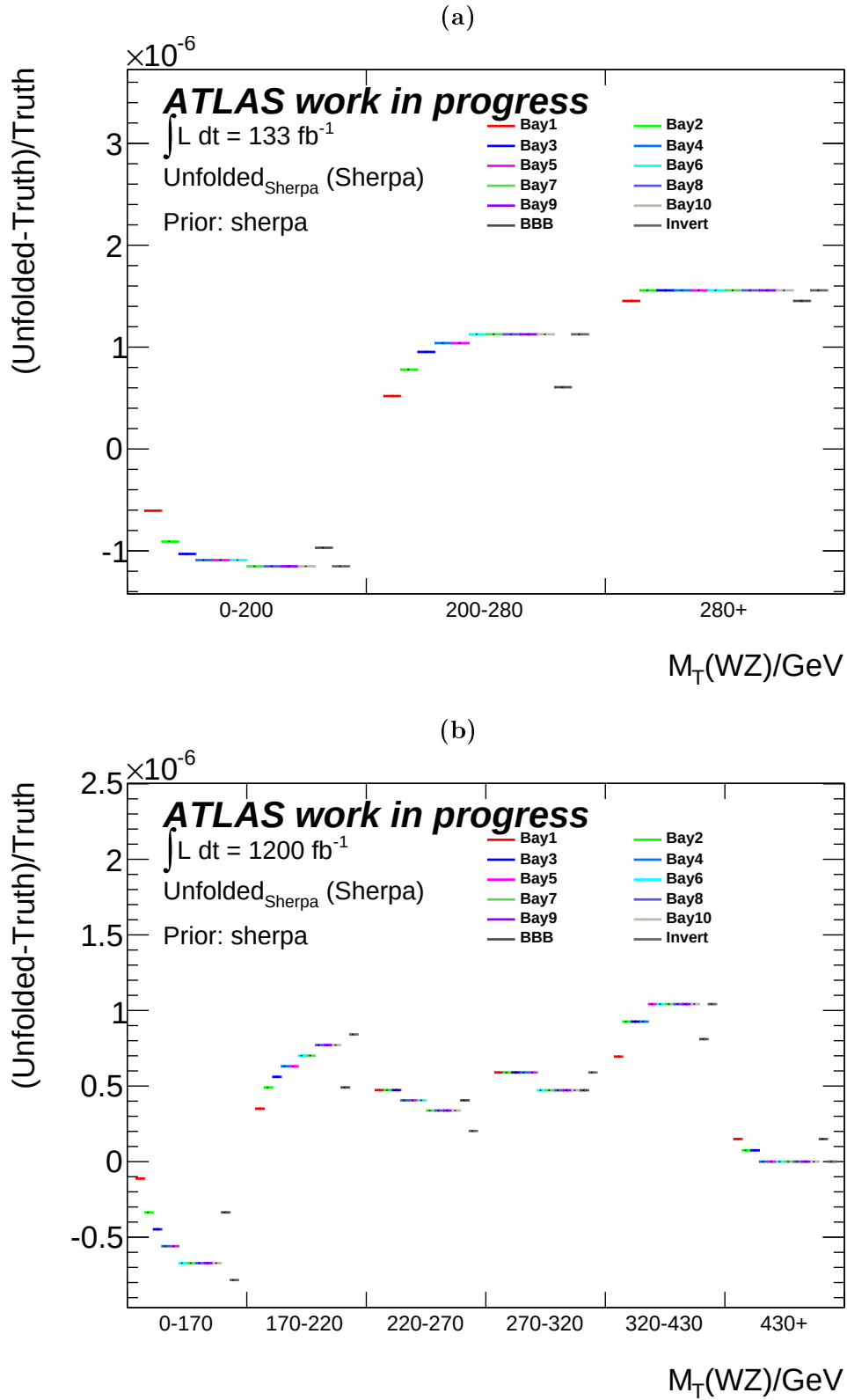


Abbildung 6.2: Dargestellt sind die Ergebnisse des Technical Closure Tests für die Entfaltung von $M_T(WZ)$ mittels Sherpa bei einer integrierten Luminosität von 133 fb^{-1} (a) bzw. 1200 fb^{-1} (b).

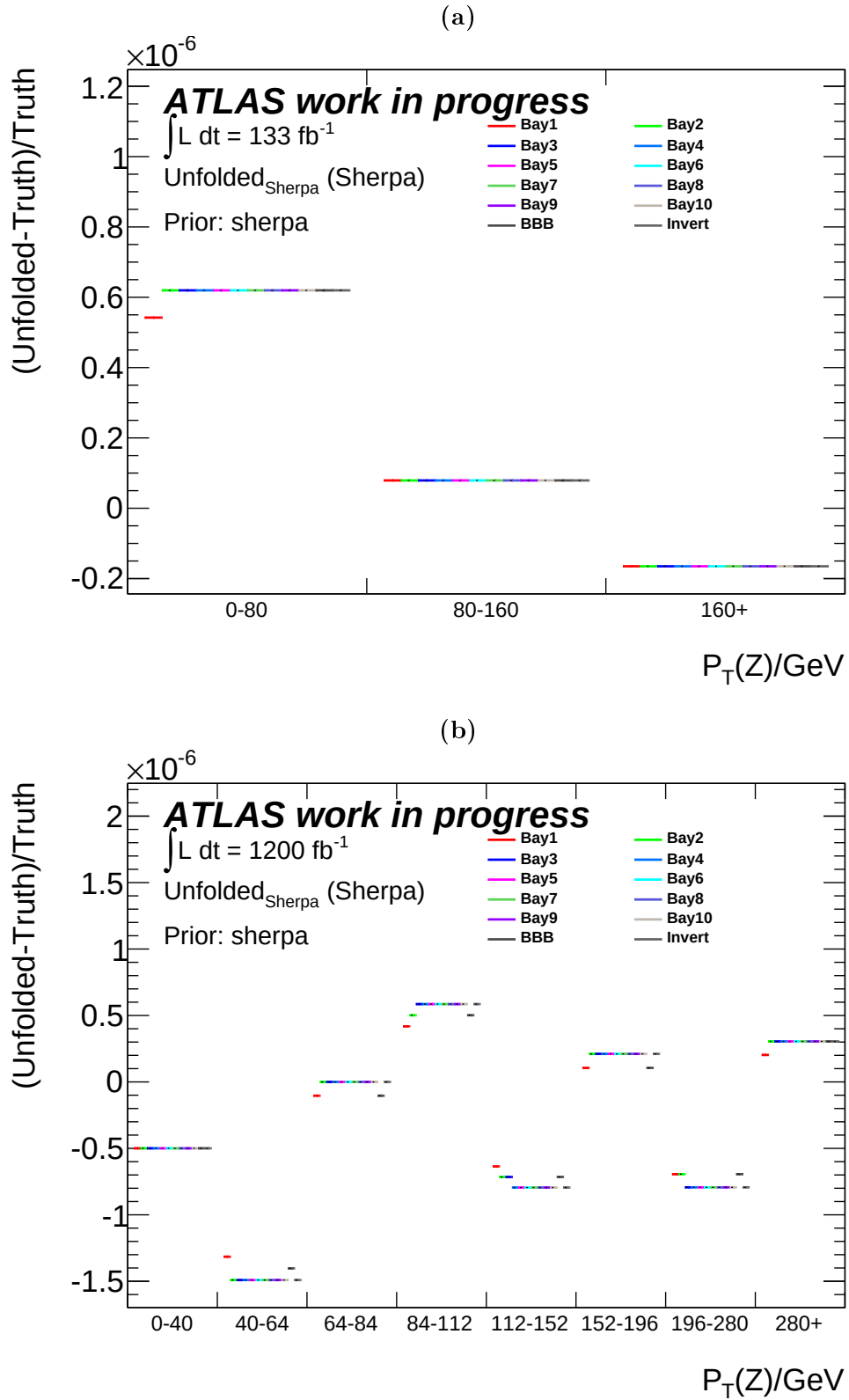


Abbildung 6.3: Dargestellt sind die Ergebnisse des Technical Closure Tests für die Entfaltung von $P_T(Z)$ mittels Sherpa bei einer integrierten Luminosität von 133 fb^{-1} (a) bzw. 1200 fb^{-1} (b).

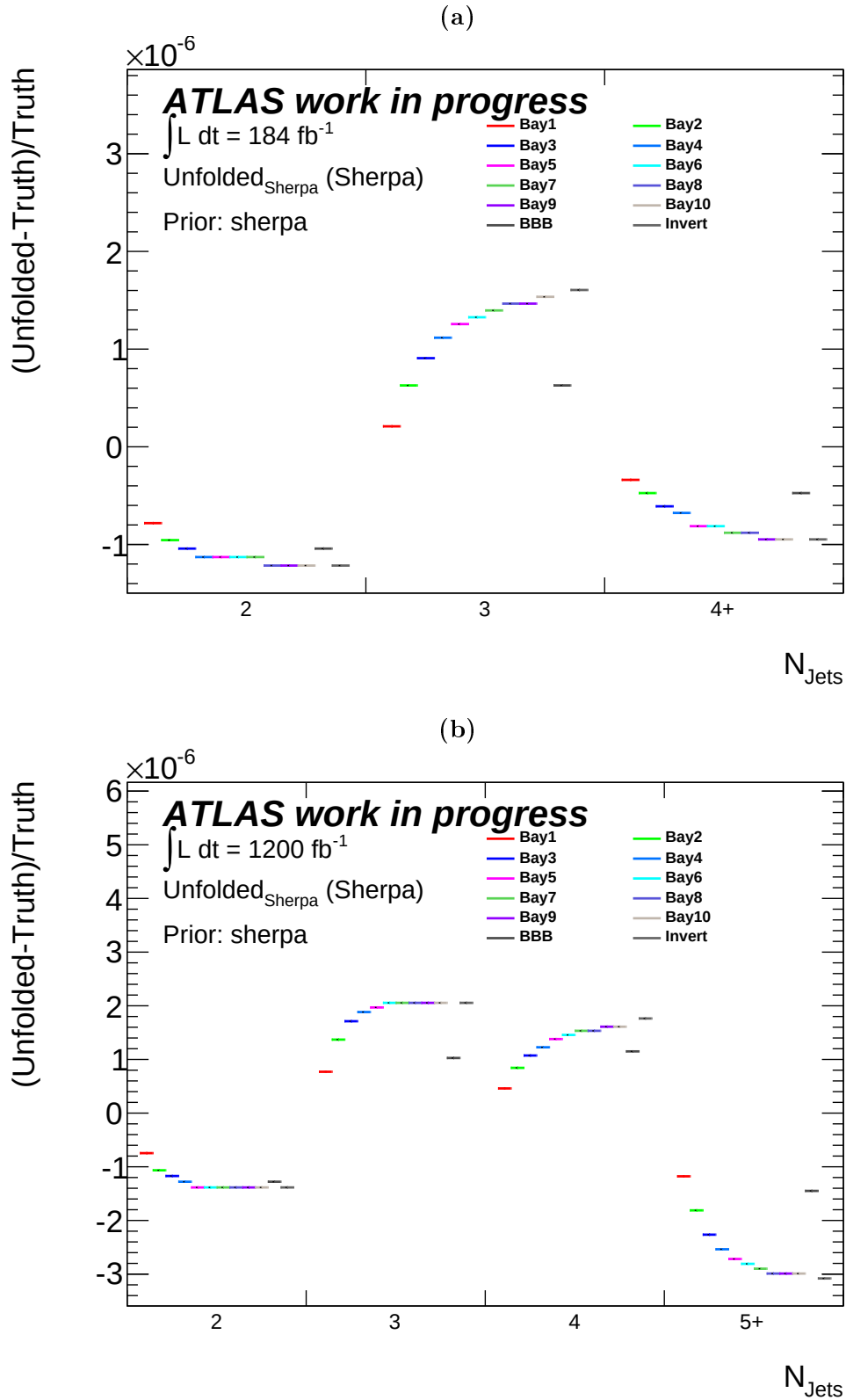


Abbildung 6.4: Dargestellt sind die Ergebnisse des Technical Closure Tests für die Entfaltung von N_{Jets} mittels Sherpa bei einer integrierten Luminosität von 184 fb^{-1} (a) bzw. 1200 fb^{-1} (b).

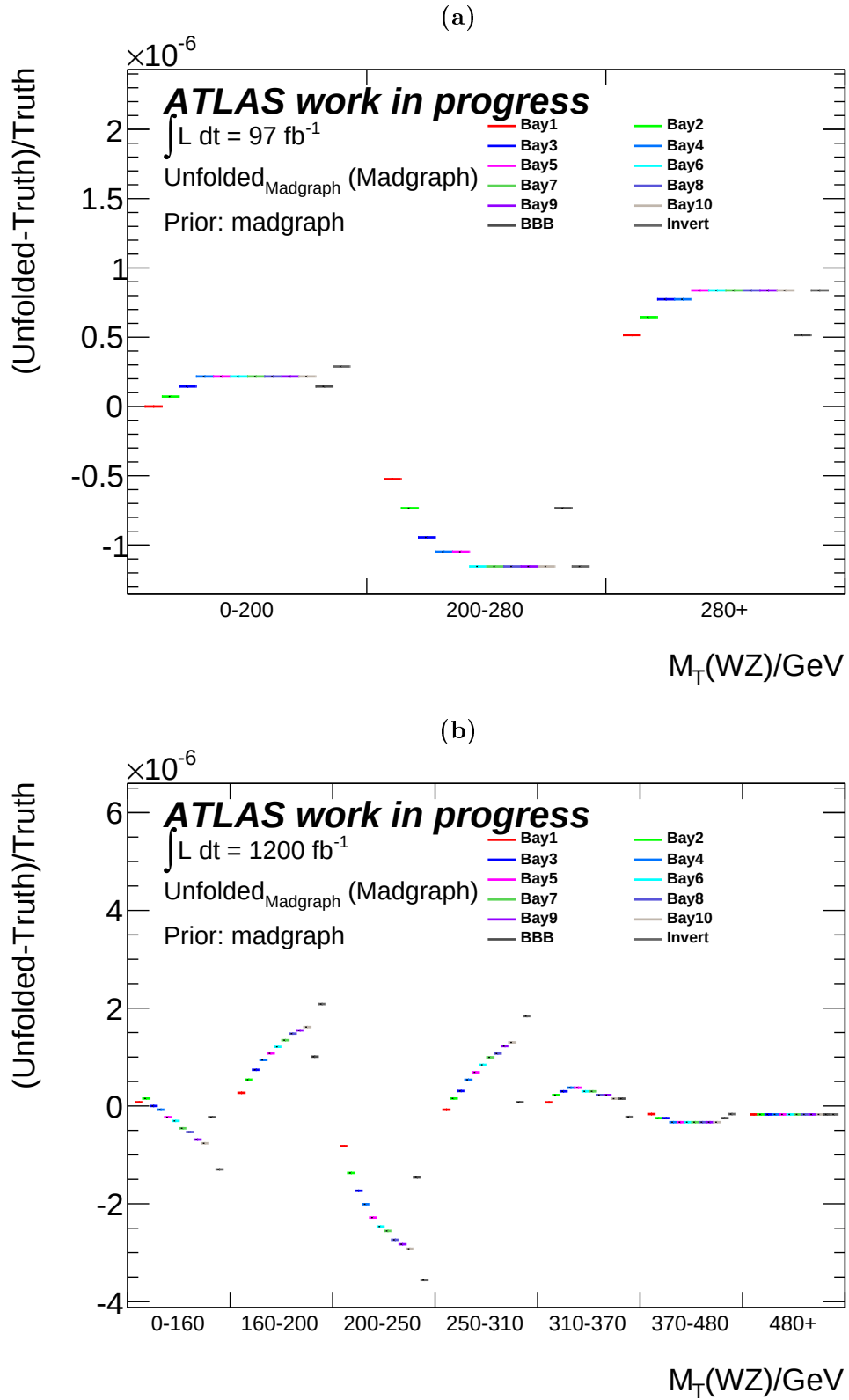


Abbildung 6.5: Dargestellt sind die Ergebnisse des Technical Closure Tests für die Entfaltung von $M_T(WZ)$ mittels Madgraph bei einer integrierten Luminosität von 97 fb^{-1} (a) bzw. 1200 fb^{-1} (b).

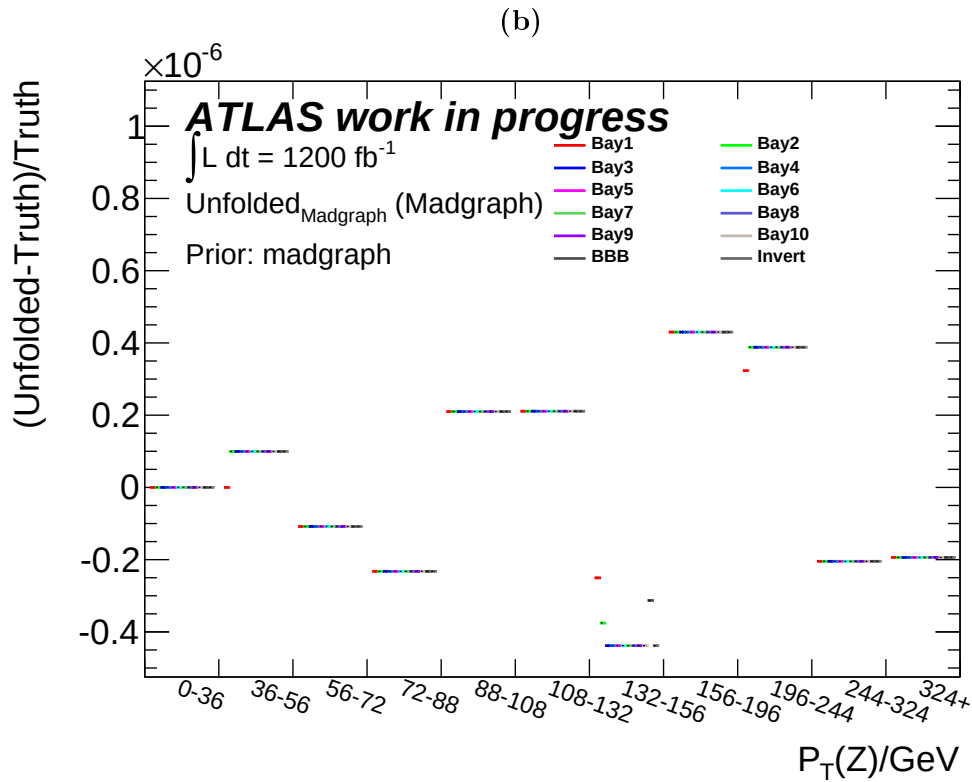
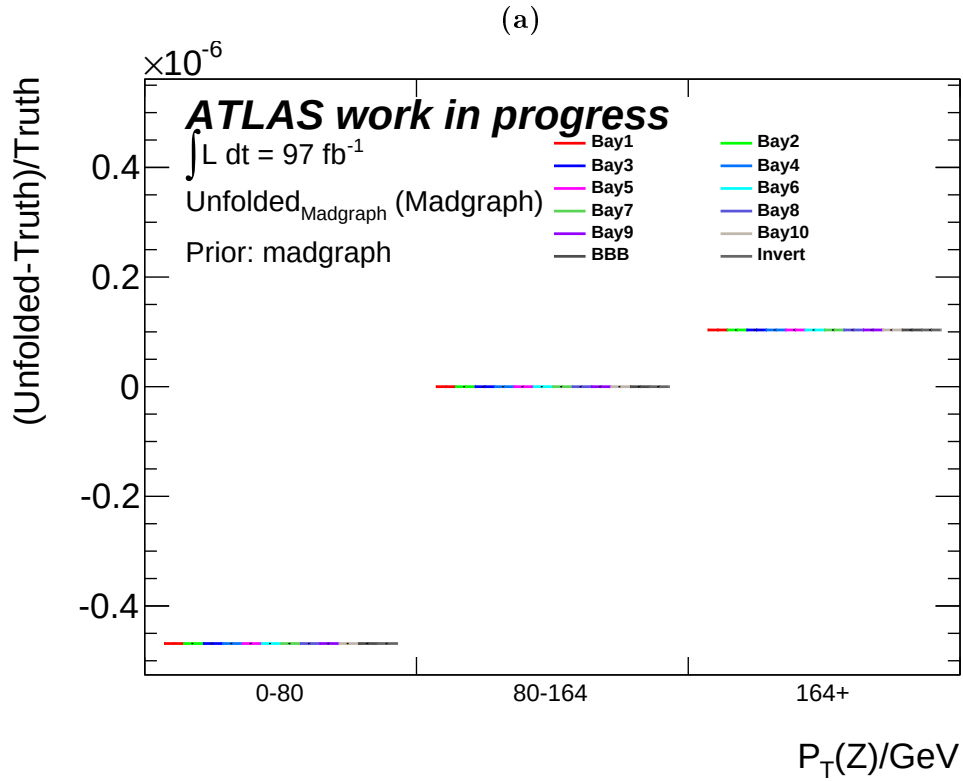


Abbildung 6.6: Dargestellt sind die Ergebnisse des Technical Closure Tests für die Entfaltung von $P_T(Z)$ mittels Madgraph bei einer integrierten Luminosität von 97 fb^{-1} (a) bzw. 1200 fb^{-1} (b).

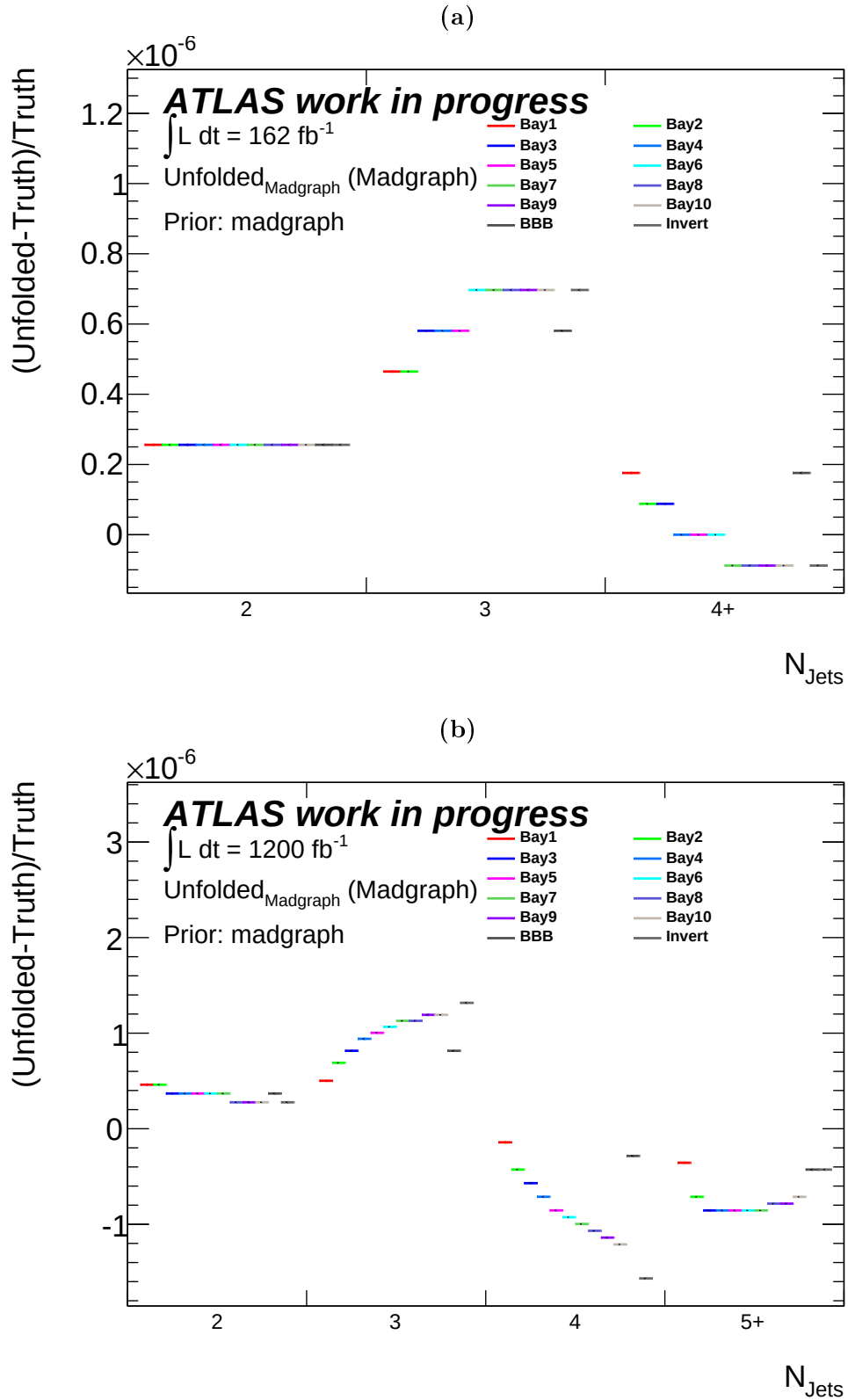


Abbildung 6.7: Dargestellt sind die Ergebnisse des Technical Closure Tests für die Entfaltung von N_{Jets} mittels Madgraph bei einer integrierten Luminosität von 162 fb^{-1} (a) bzw. 1200 fb^{-1} (b).

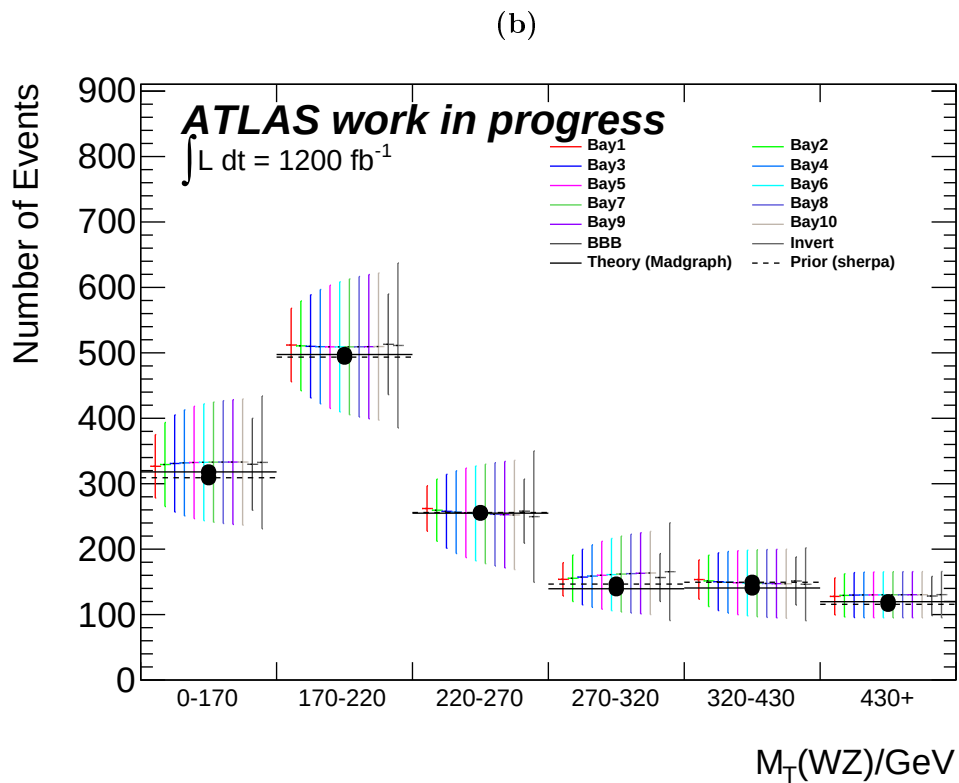
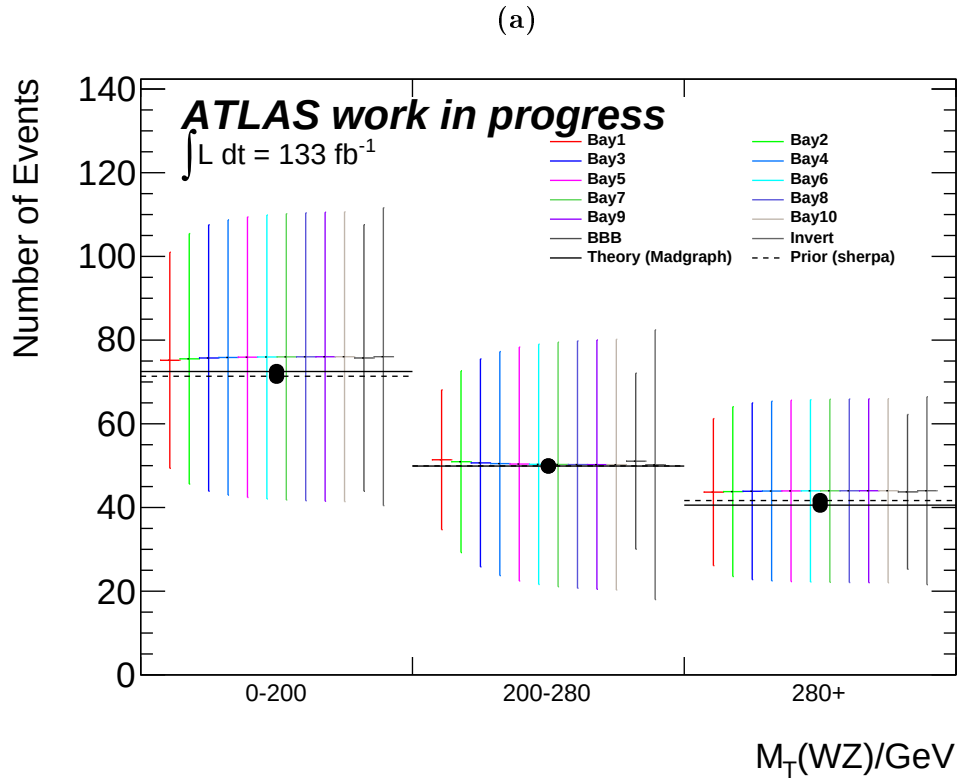


Abbildung 6.8: Dargestellt sind die mittels Sherpa durch verschiedene Entfaltungsmethoden ermittelten Entfaltungsergebnisse der Madgraph *reco*-Verteilung für $M_T(WZ)$ bei 133 fb^{-1} (a) bzw. 1200 fb^{-1} (b). Zusätzlich sind auch die Madgraph *truth*-Verteilung, der verwendete Prior und die statistischen und systematischen Gesamtunsicherheiten eingezeichnet.

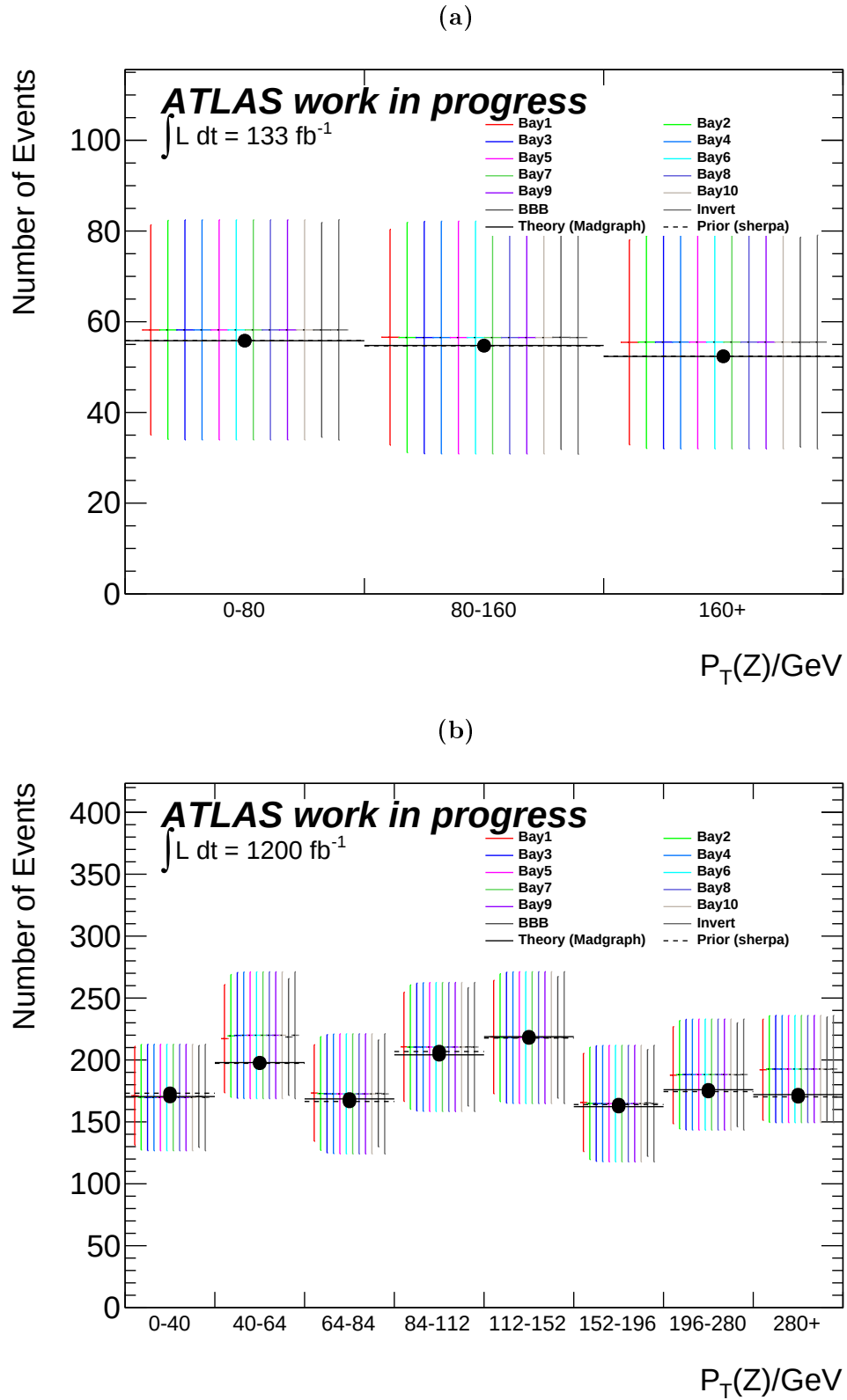


Abbildung 6.9: Dargestellt sind die mittels Sherpa durch verschiedene Entfaltungsmethoden ermittelten Entfaltungsergebnisse der Madgraph *reco*-Verteilung für $P_T(Z)$ bei 133 fb^{-1} (a) bzw. 1200 fb^{-1} (b). Zusätzlich sind auch die Madgraph *truth*-Verteilung, der verwendete Prior und die statistischen und systematischen Gesamtunsicherheiten eingezeichnet.

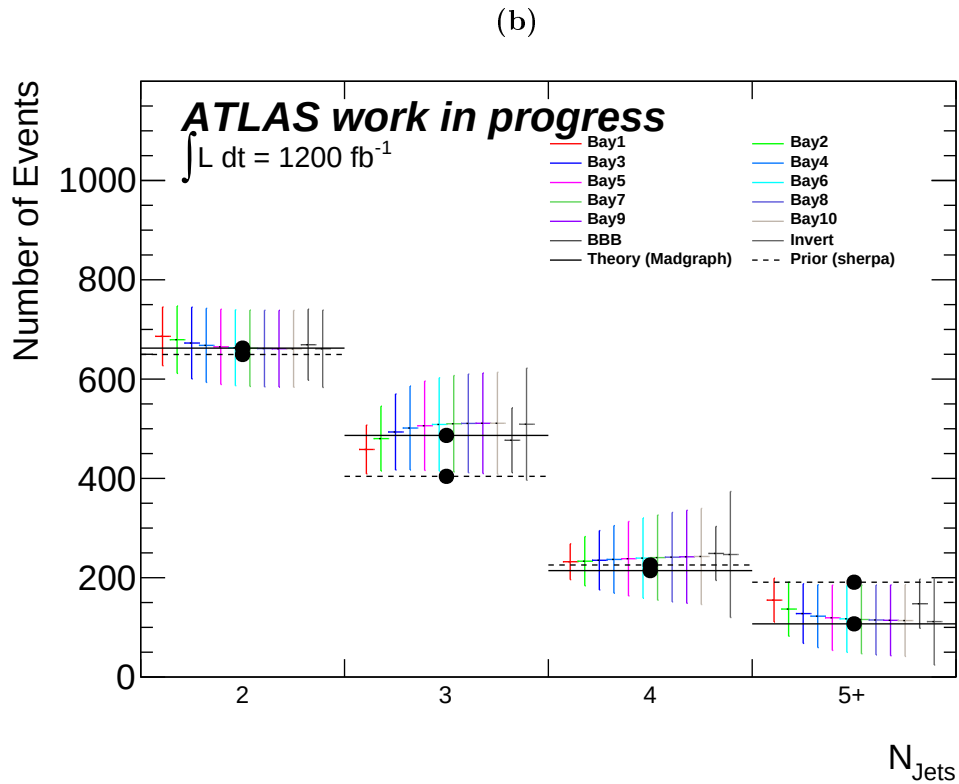
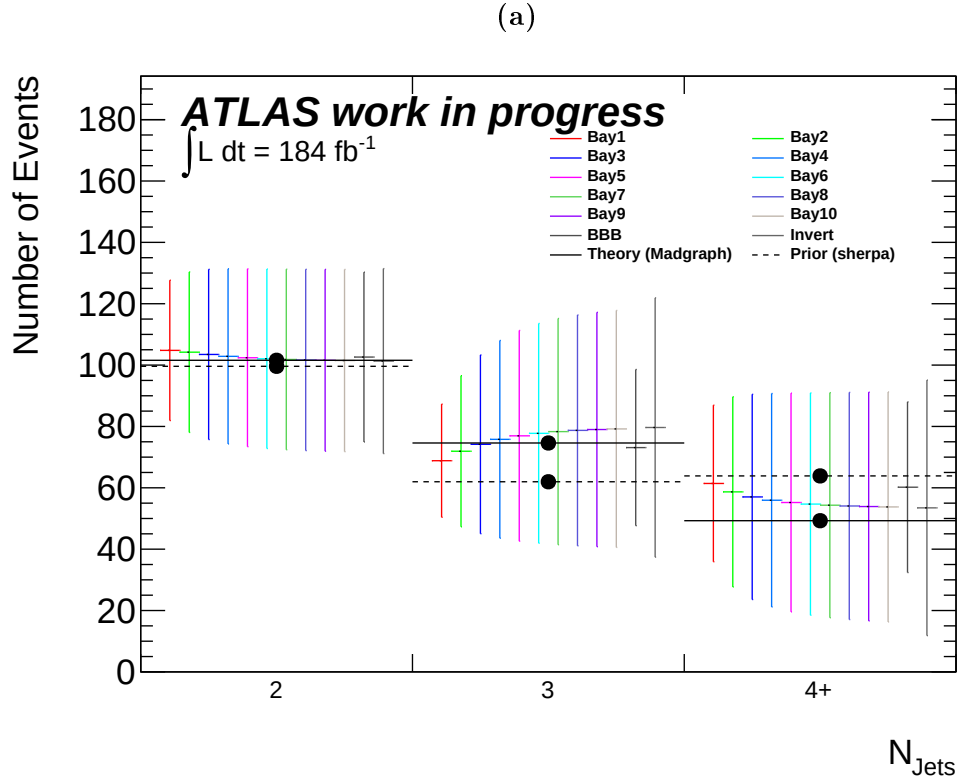


Abbildung 6.10: Dargestellt sind die mittels Sherpa durch verschiedene Entfaltungsmethoden ermittelten Entfaltungsergebnisse der Madgraph *reco*-Verteilung für N_{Jets} bei 184 fb^{-1} (a) bzw. 1200 fb^{-1} (b). Zusätzlich sind auch die Madgraph *truth*-Verteilung, der verwendete Prior und die statistischen und systematischen Gesamtunsicherheiten eingezeichnet.

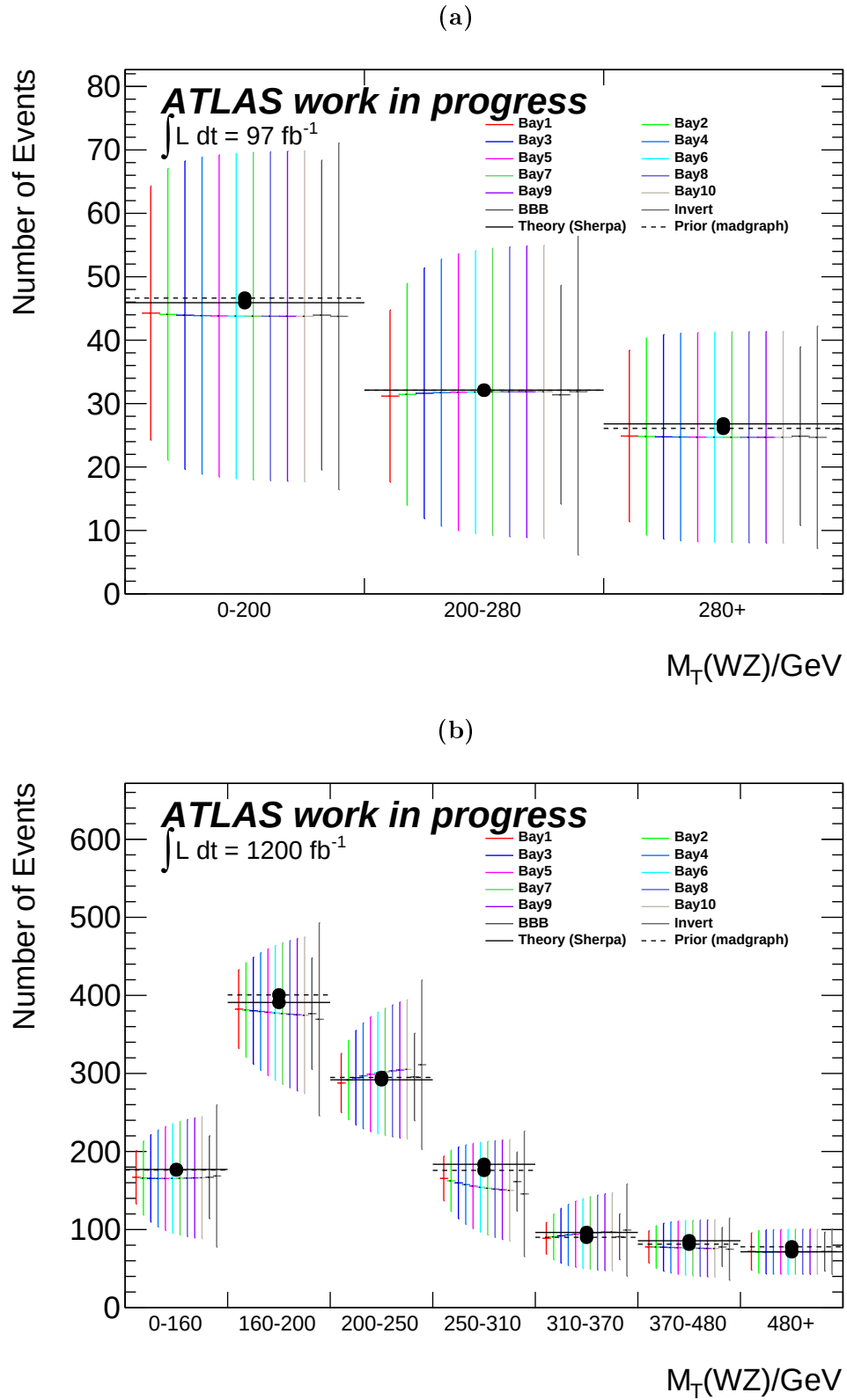


Abbildung 6.11: Dargestellt sind die mittels Madgraph durch verschiedene Entfaltungsmethoden ermittelten Entfaltungsergebnisse der Sherpa *reco*-Verteilung für $M_T(WZ)$ bei 97 fb^{-1} (a) bzw. 1200 fb^{-1} (b). Zusätzlich sind auch die Sherpa *truth*-Verteilung, der verwendete Prior und die statistischen und systematischen Gesamtunsicherheiten eingezeichnet.

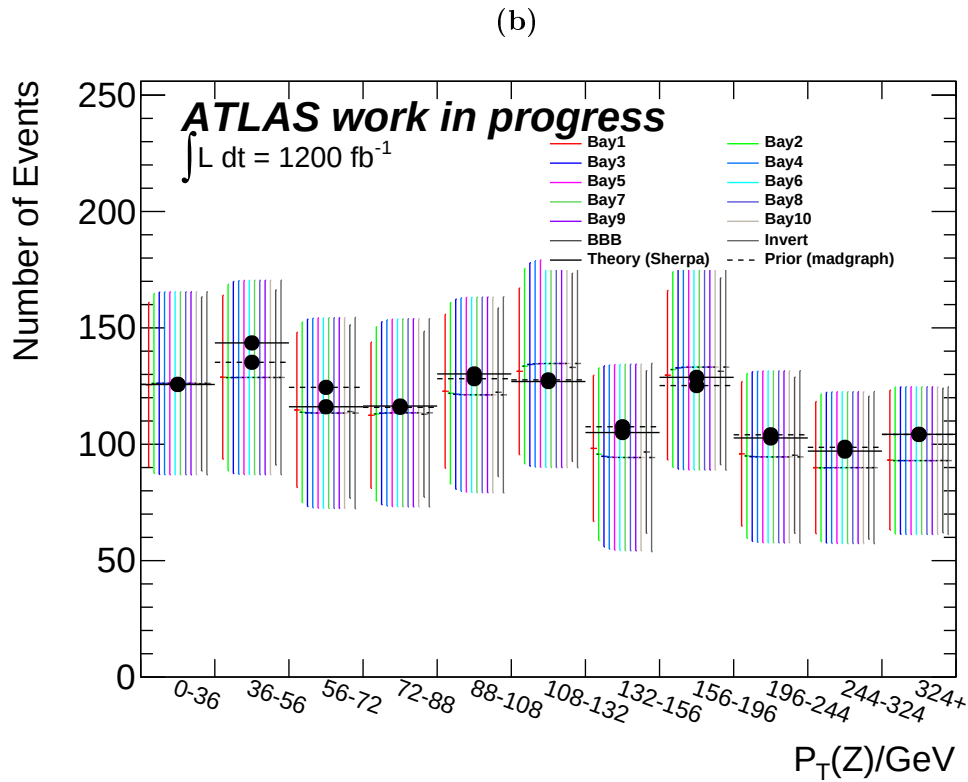
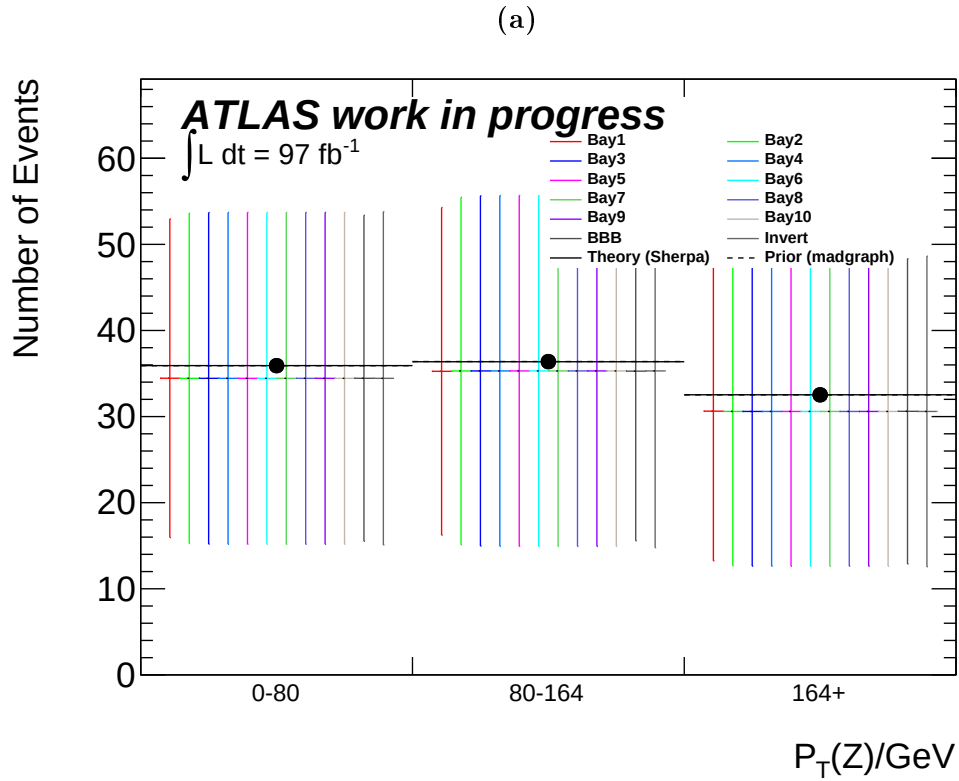


Abbildung 6.12: Dargestellt sind die mittels Madgraph durch verschiedene Entfaltungsmethoden ermittelten Entfaltungsergebnisse der Sherpa *reco*-Verteilung für $P_T(Z)$ bei 97 fb^{-1} (a) bzw. 1200 fb^{-1} (b). Zusätzlich sind auch die Sherpa *truth*-Verteilung, der verwendete Prior und die statistischen und systematischen Gesamtunsicherheiten eingezeichnet.

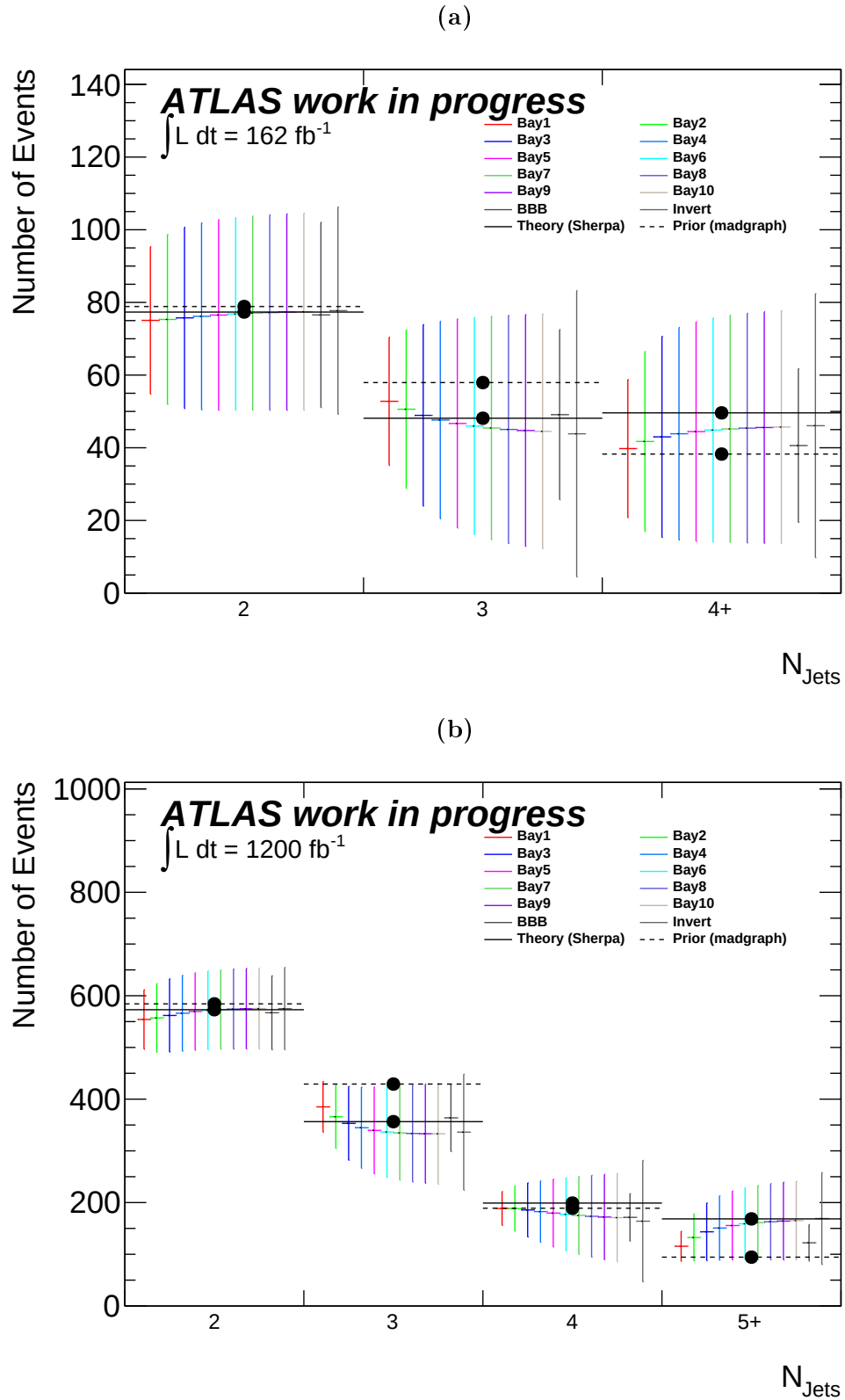


Abbildung 6.13: Dargestellt sind die mittels Madgraph durch verschiedene Entfaltungsmethoden ermittelten Entfaltungsergebnisse der Sherpa *reco*-Verteilung für N_{Jets} bei 162 fb^{-1} (a) bzw. 1200 fb^{-1} (b). Zusätzlich sind auch die Sherpa *truth*-Verteilung, der verwendete Prior und die statistischen und systematischen Gesamtunsicherheiten eingezeichnet.

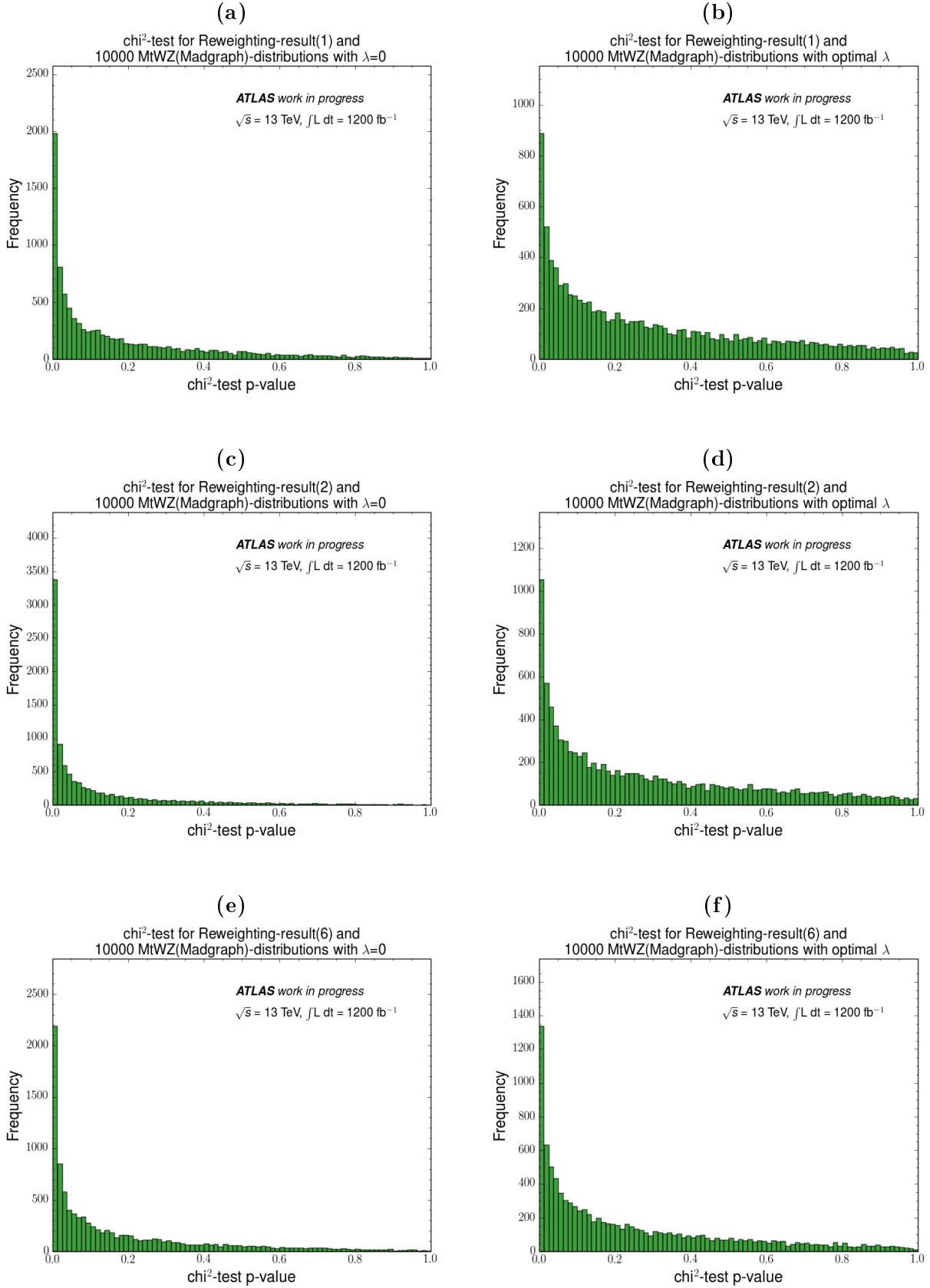


Abbildung 6.14: Verglichen werden die p-Wert-Verteilungen des in Kapitel 4.4.2 beschriebenen Chi-Quadrat-Tests für eine Umgewichtung ohne der Berücksichtigung der Glattheit (a), (c) und (e) und mit der Wahl des optimalen Glättungsparameters (b), (d) und (f) für verschiedene Fluktuationen von $M_T(WZ)$.

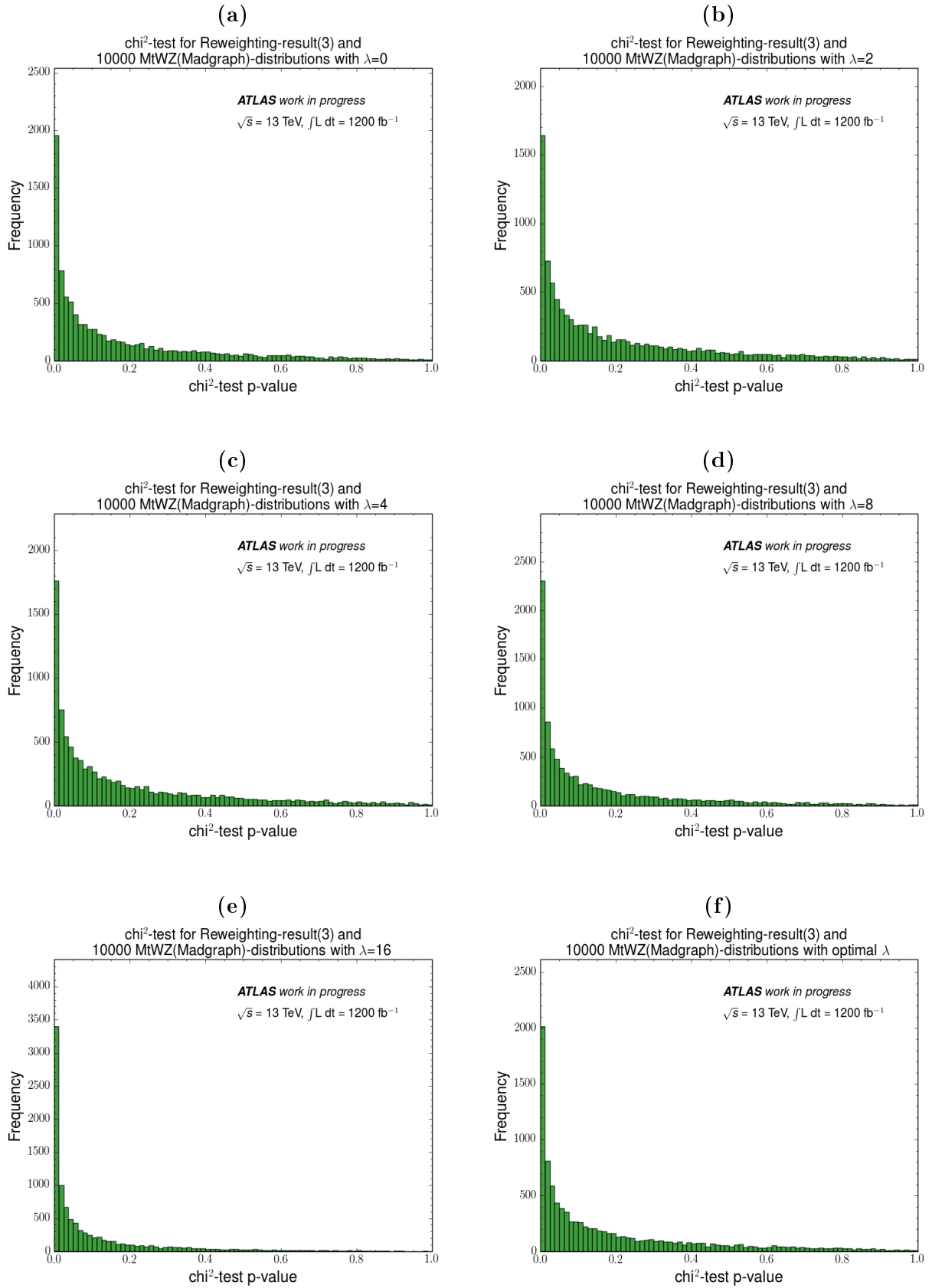


Abbildung 6.15: Verglichen werden die p-Wert-Verteilungen des in Kapitel 4.4.2 beschriebenen Chi-Quadrat-Tests die Umgewichtung mit unterschiedlichen Glättungsparametern für eine Fluktuationen um die veränderte Madgraph *reco*-Verteilung von $M_T(WZ)$.

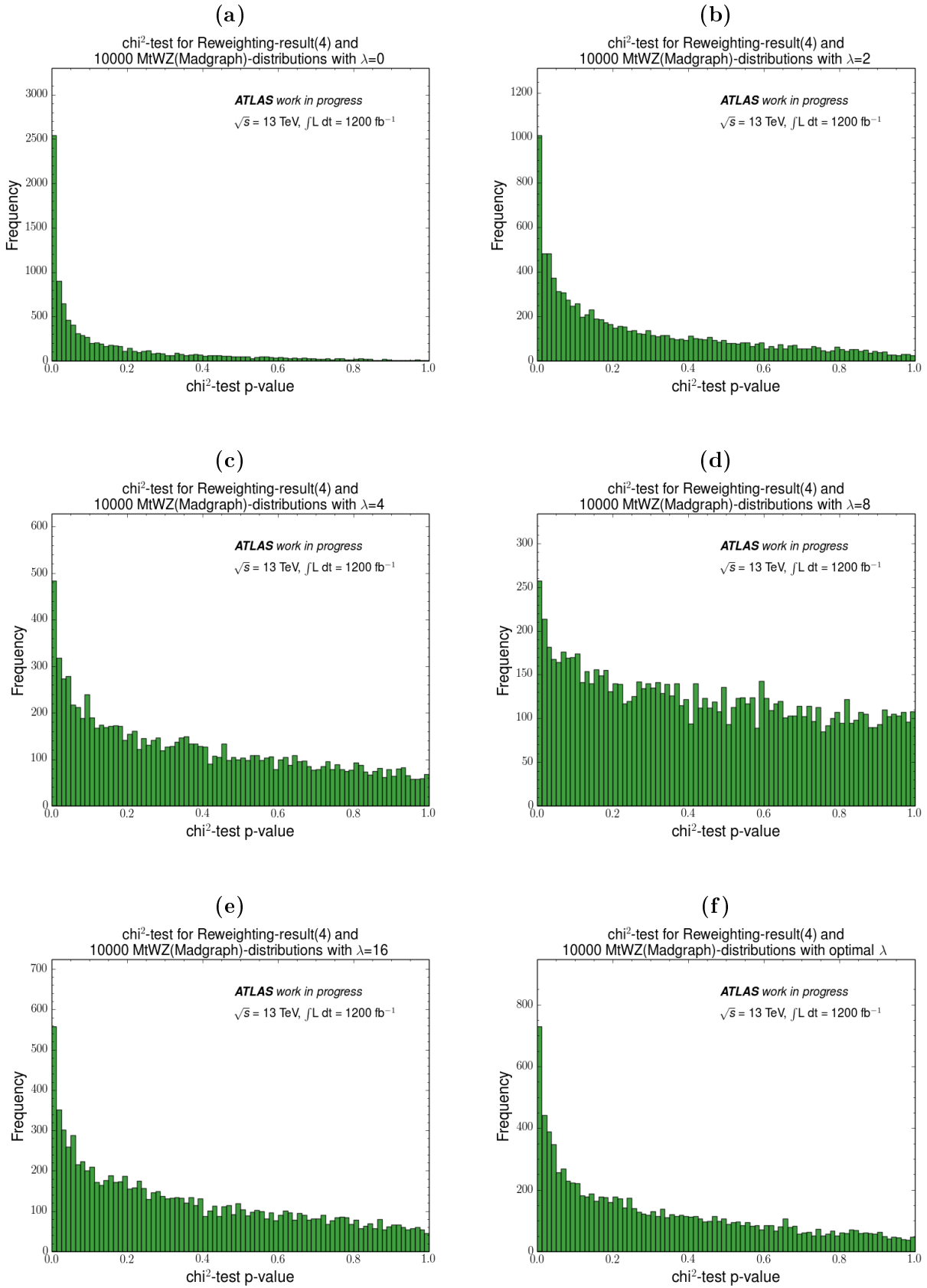


Abbildung 6.16: Verglichen werden die p-Wert-Verteilungen des in Kapitel 4.4.2 beschriebenen Chi-Quadrat-Tests die Umgewichtung mit unterschiedlichen Glättungsparametern für eine Fluktuationen um die veränderte Madgraph *reco*-Verteilung von $M_T(WZ)$.

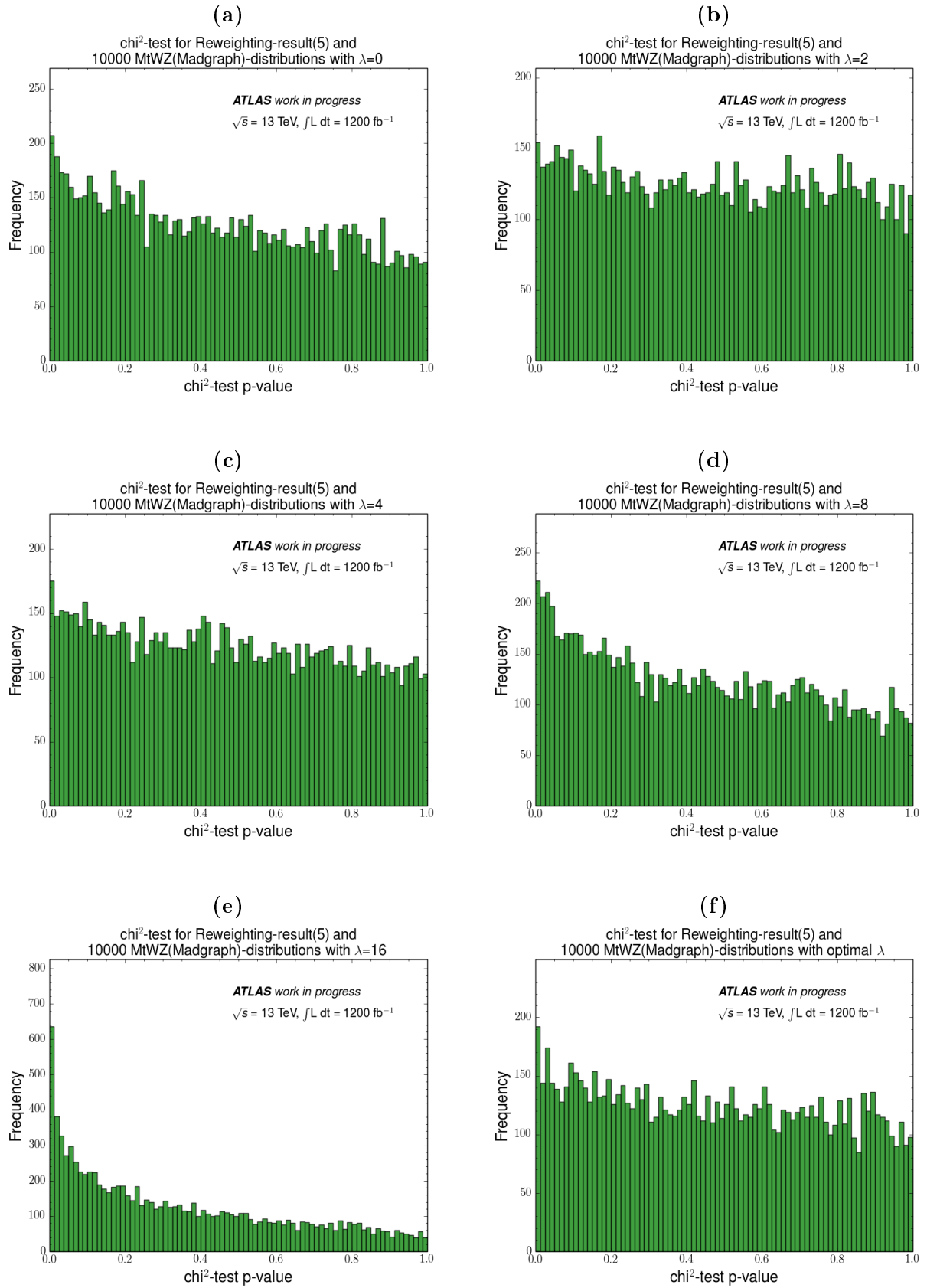


Abbildung 6.17: Verglichen werden die p-Wert-Verteilungen des in Kapitel 4.4.2 beschriebenen Chi-Quadrat-Tests die Umgewichtung mit unterschiedlichen Glättungsparametern für eine Fluktuationen um die veränderte Madgraph *reco*-Verteilung von $M_T(WZ)$.

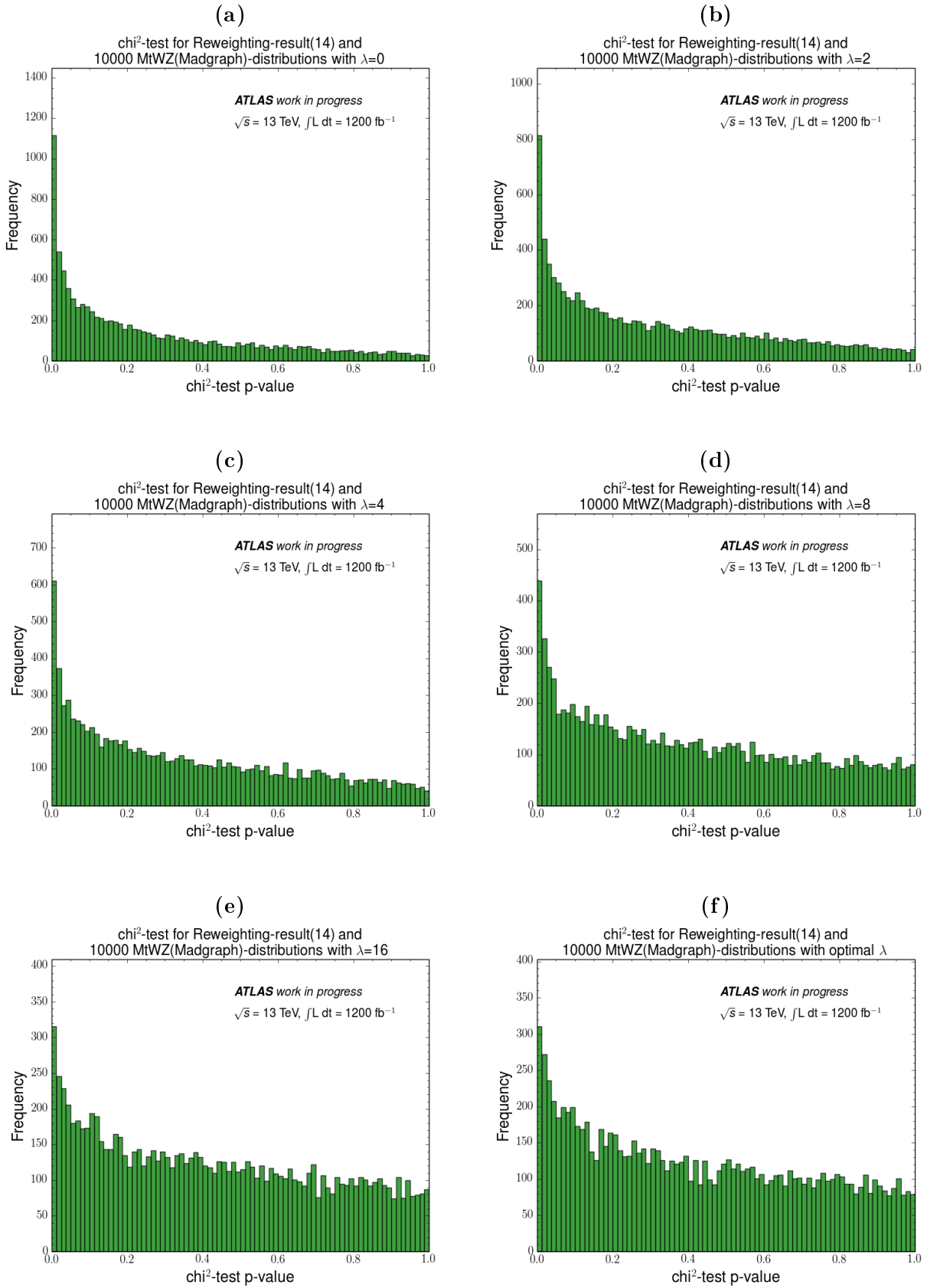


Abbildung 6.18: Verglichen werden die p-Wert-Verteilungen des in Kapitel 4.4.2 beschriebenen Chi-Quadrat-Tests die Umgewichtung mit unterschiedlichen Glättungsparametern für eine Fluktuationen um die veränderte Madgraph *reco*-Verteilung von $M_T(WZ)$.

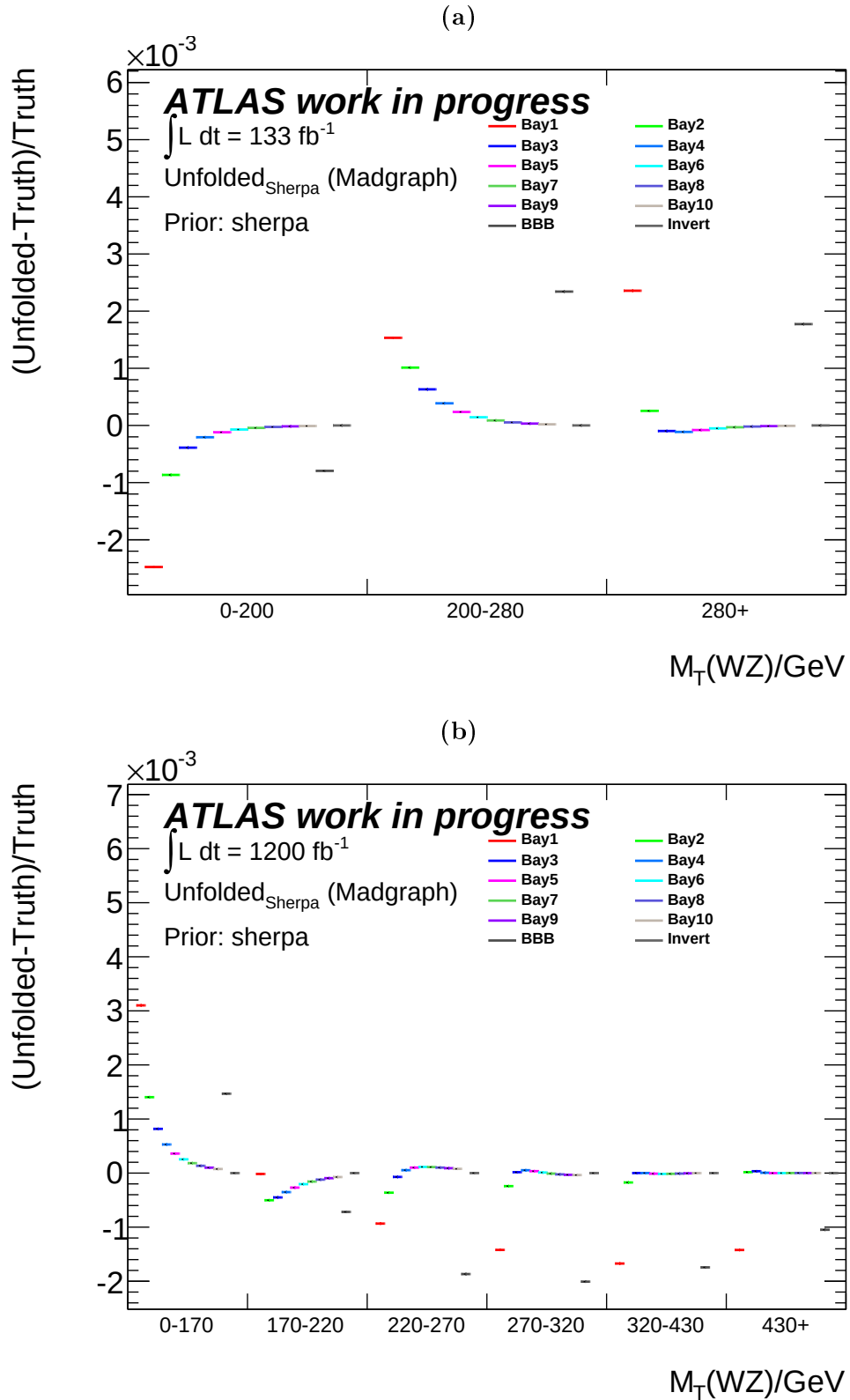


Abbildung 6.19: Die dargestellte relative Abweichung des Entfaltungsergebnisses von der *truth*-Verteilung lässt sich als ein Maß für die jeweilige Entfaltungsunsicherheit auffassen. Die *truth*-Verteilung entspricht der Sherpa *truth*-Verteilung von $M_T(WZ)$ bei 133 fb^{-1} (a) bzw. 1200 fb^{-1} (b), welche im Bezug zur Madgraph *reco*-Verteilung umgewichtet wurde. Entfaltet wurde die entsprechende Faltung der umgewichteten *truth*-Verteilung mittels Sherpa.

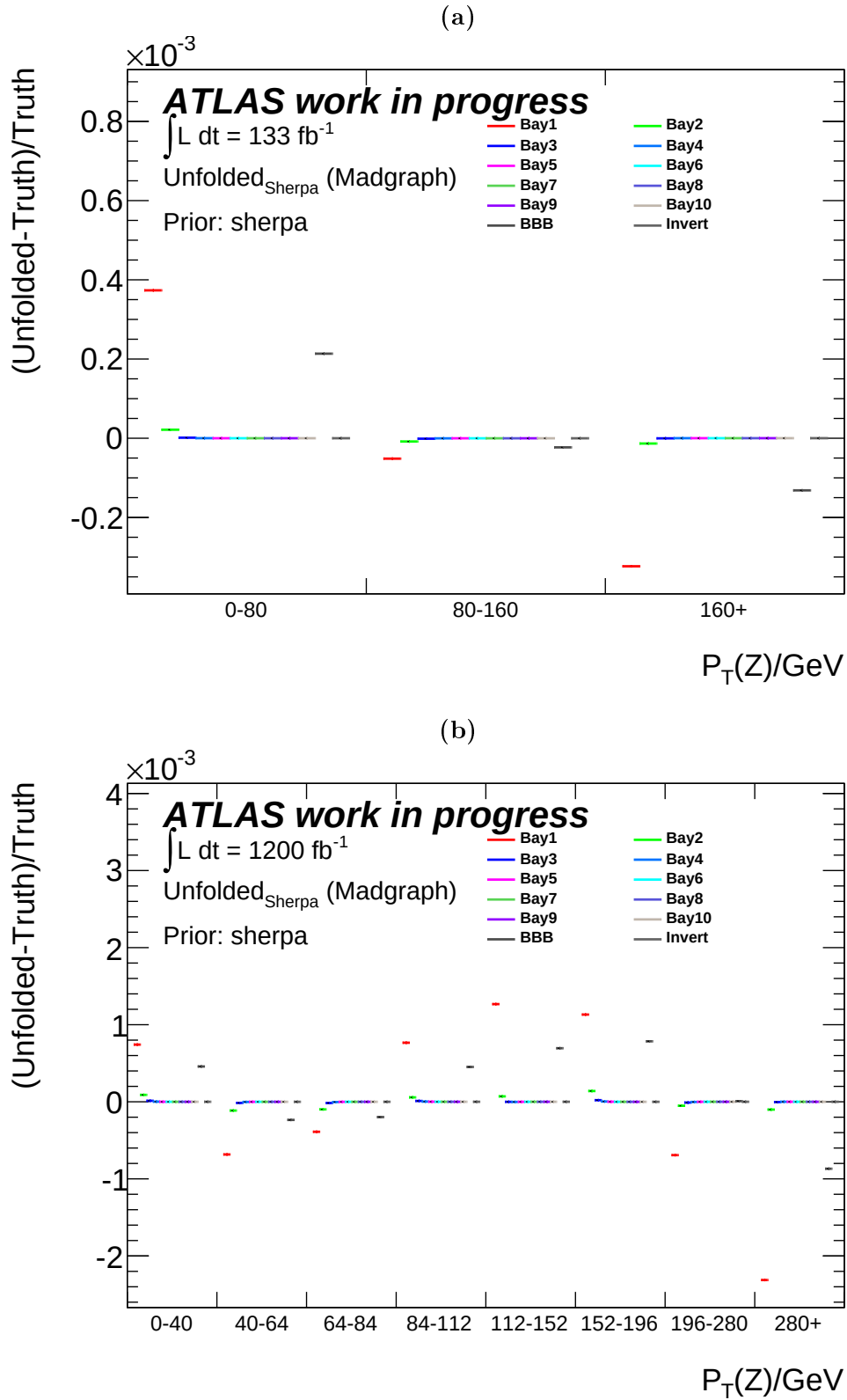
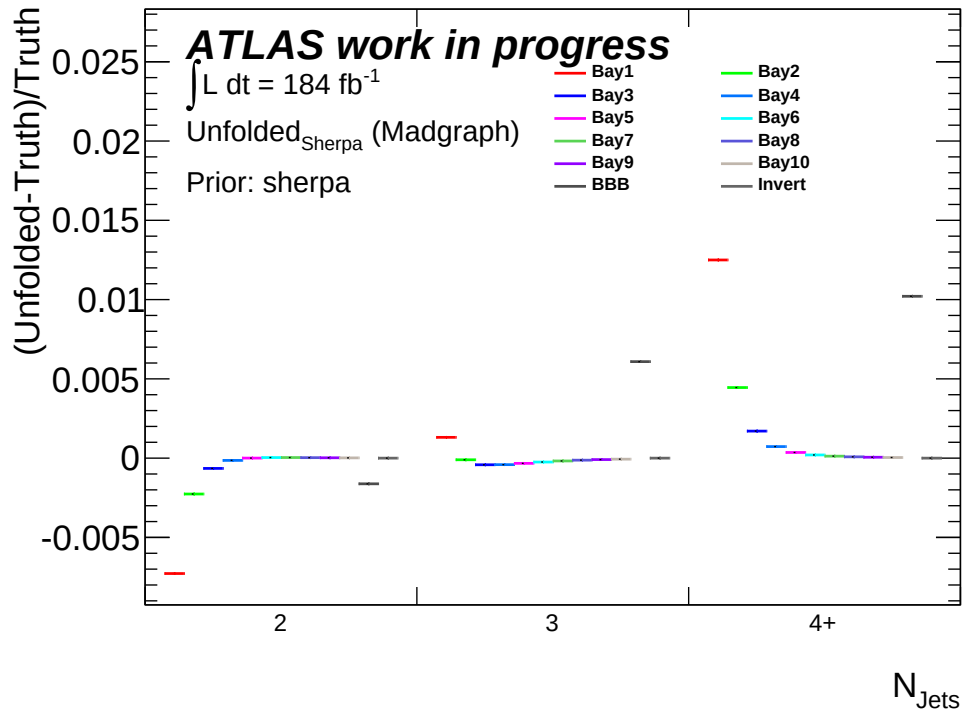


Abbildung 6.20: Die dargestellte relative Abweichung des Entfaltungsergebnisses von der *truth*-Verteilung lässt sich als ein Maß für die jeweilige Entfaltungsunsicherheit auffassen. Die *truth*-Verteilung entspricht der Sherpa *truth*-Verteilung von $P_T(Z)$ bei 133 fb^{-1} (a) bzw. 1200 fb^{-1} (b), welche im Bezug zur Madgraph *reco*-Verteilung umgewichtet wurde. Entfaltet wurde die entsprechende Faltung der umgewichteten *truth*-Verteilung mittels Sherpa.

(a)



(b)

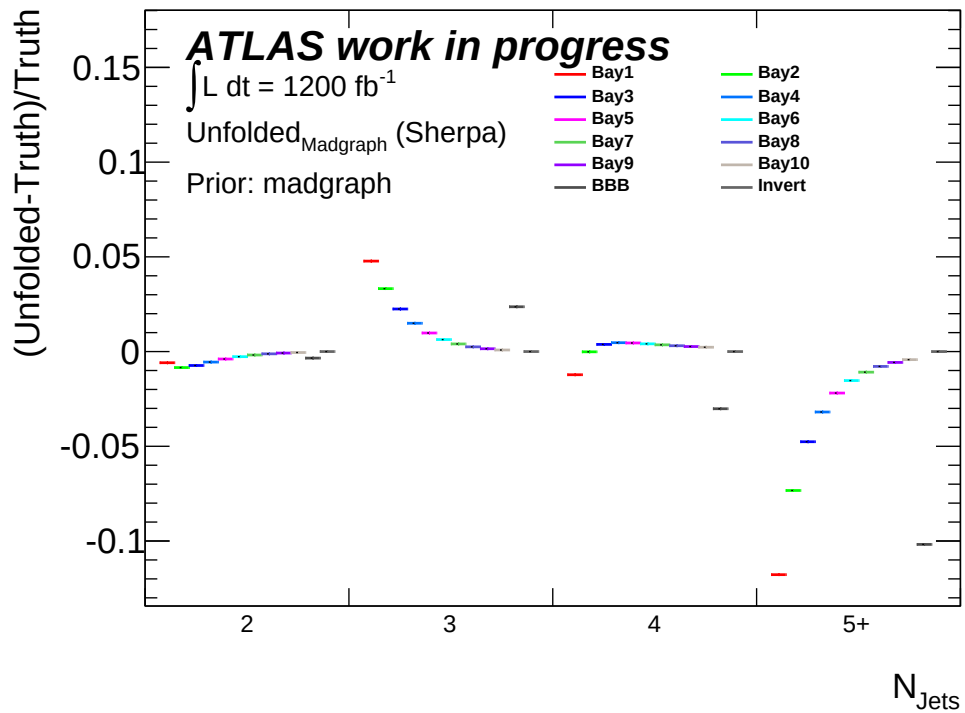


Abbildung 6.21: Die dargestellte relative Abweichung des Entfaltungsergebnisses von der *truth*-Verteilung lässt sich als ein Maß für die jeweilige Entfaltungsunsicherheit auffassen. Die *truth*-Verteilung entspricht der Sherpa *truth*-Verteilung von N_{Jets} bei 184 fb^{-1} (a) bzw. 1200 fb^{-1} (b), welche im Bezug zur Madgraph *reco*-Verteilung umgewichtet wurde. Entfaltet wurde die entsprechende Faltung der umgewichteten *truth*-Verteilung mittels Sherpa.

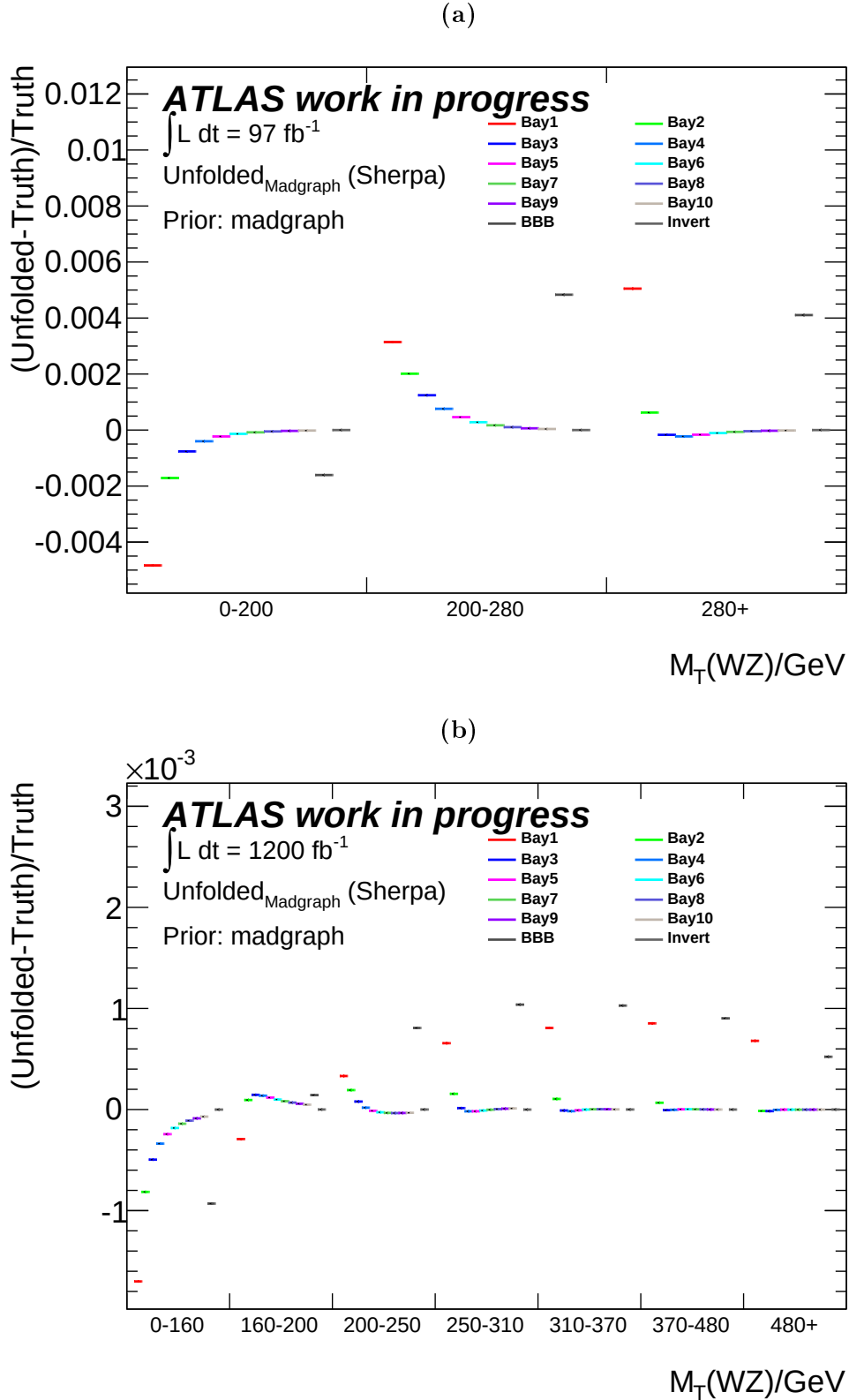


Abbildung 6.22: Die dargestellte relative Abweichung des Entfaltungsergebnisses von der *truth*-Verteilung lässt sich als ein Maß für die jeweilige Entfaltungsunsicherheit auffassen. Die *truth*-Verteilung entspricht der Madgraph *truth*-Verteilung von $M_T(WZ)$ bei 97 fb^{-1} (a) bzw. 1200 fb^{-1} (b), welche im Bezug zur Sherpa *reco*-Verteilung umgewichtet wurde. Entfaltet wurde die entsprechende Faltung der umgewichteten *truth*-Verteilung mittels Madgraph.

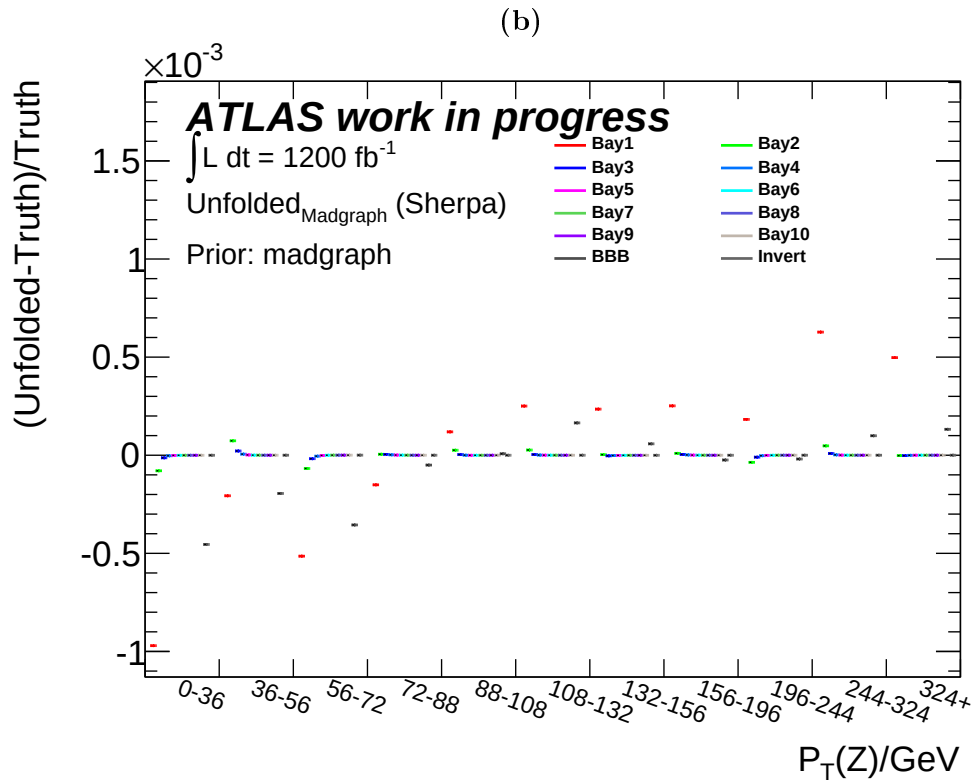
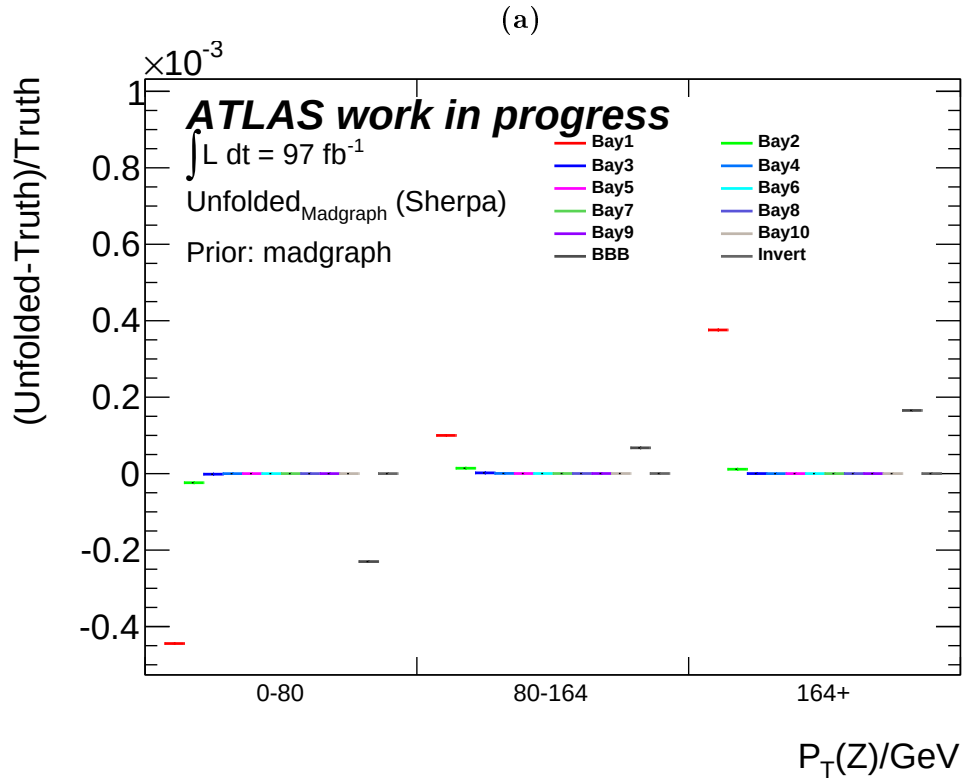


Abbildung 6.23: Die dargestellte relative Abweichung des Entfaltungsergebnisses von der *truth*-Verteilung lässt sich als ein Maß für die jeweilige Entfaltungsunsicherheit auffassen. Die *truth*-Verteilung entspricht der Madgraph *truth*-Verteilung von $P_T(Z)$ bei 97 fb^{-1} (a) bzw. 1200 fb^{-1} (b), welche im Bezug zur Sherpa *reco*-Verteilung umgewichtet wurde. Entfaltet wurde die entsprechende Faltung der umgewichteten *truth*-Verteilung mittels Madgraph.

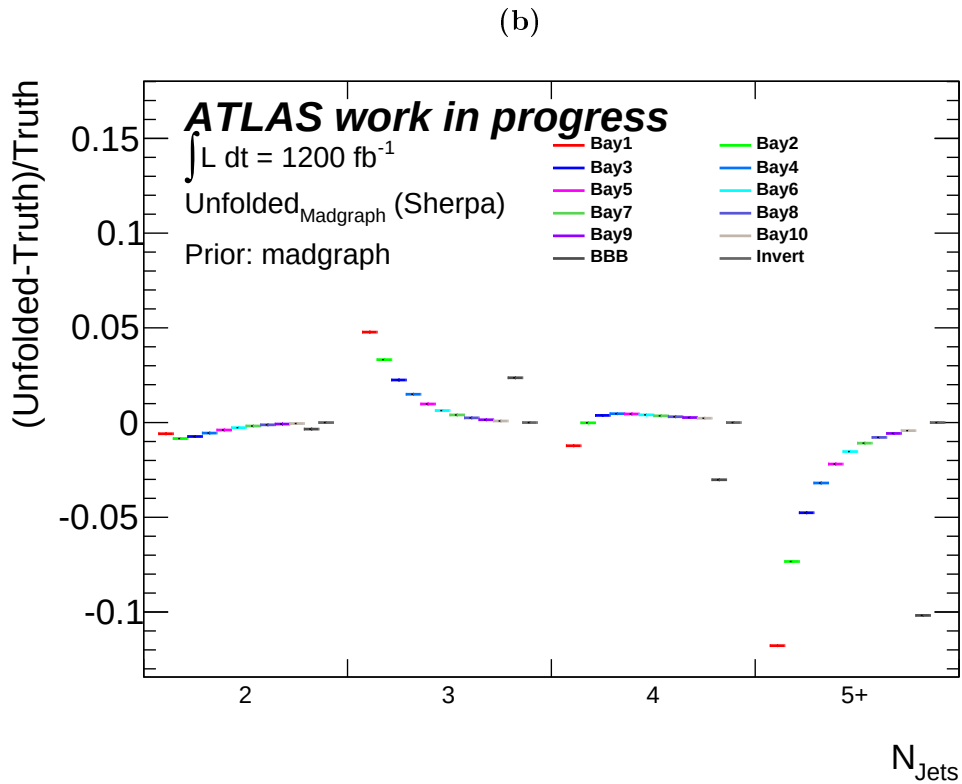
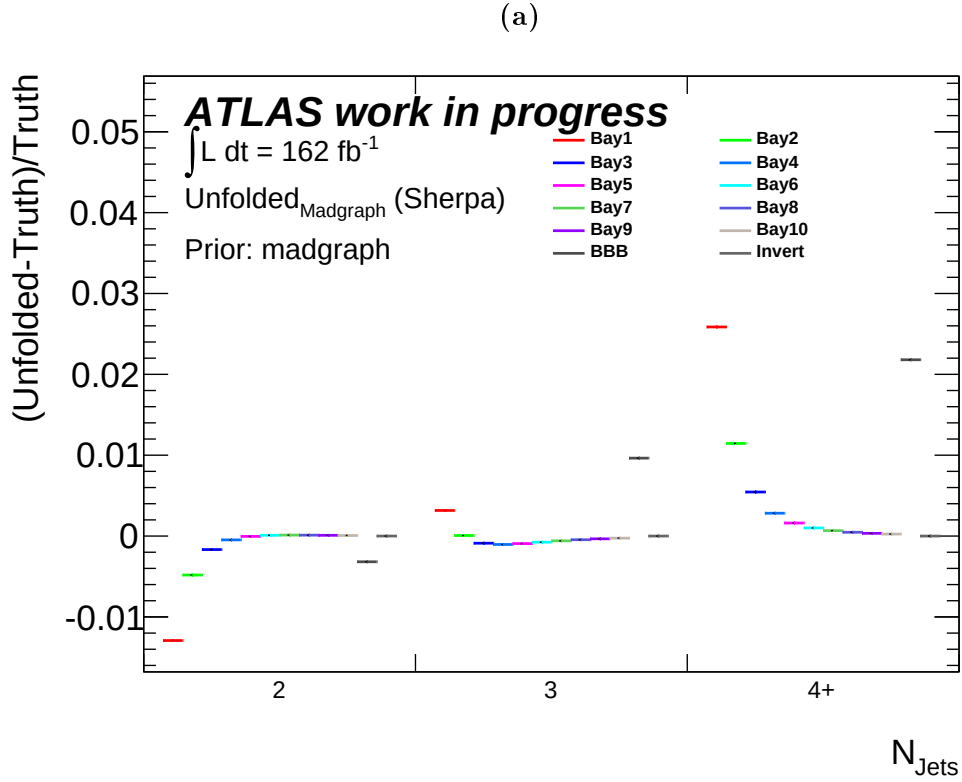


Abbildung 6.24: Die dargestellte relative Abweichung des Entfaltungsergebnisses von der *truth*-Verteilung lässt sich als ein Maß für die jeweilige Entfaltungsunsicherheit auffassen. Die *truth*-Verteilung entspricht der Madgraph *truth*-Verteilung von N_{Jets} bei 162 fb^{-1} (a) bzw. 1200 fb^{-1} (b), welche im Bezug zur Sherpa *reco*-Verteilung umgewichtet wurde. Entfaltet wurde die entsprechende Faltung der umgewichteten *truth*-Verteilung mittels Madgraph.

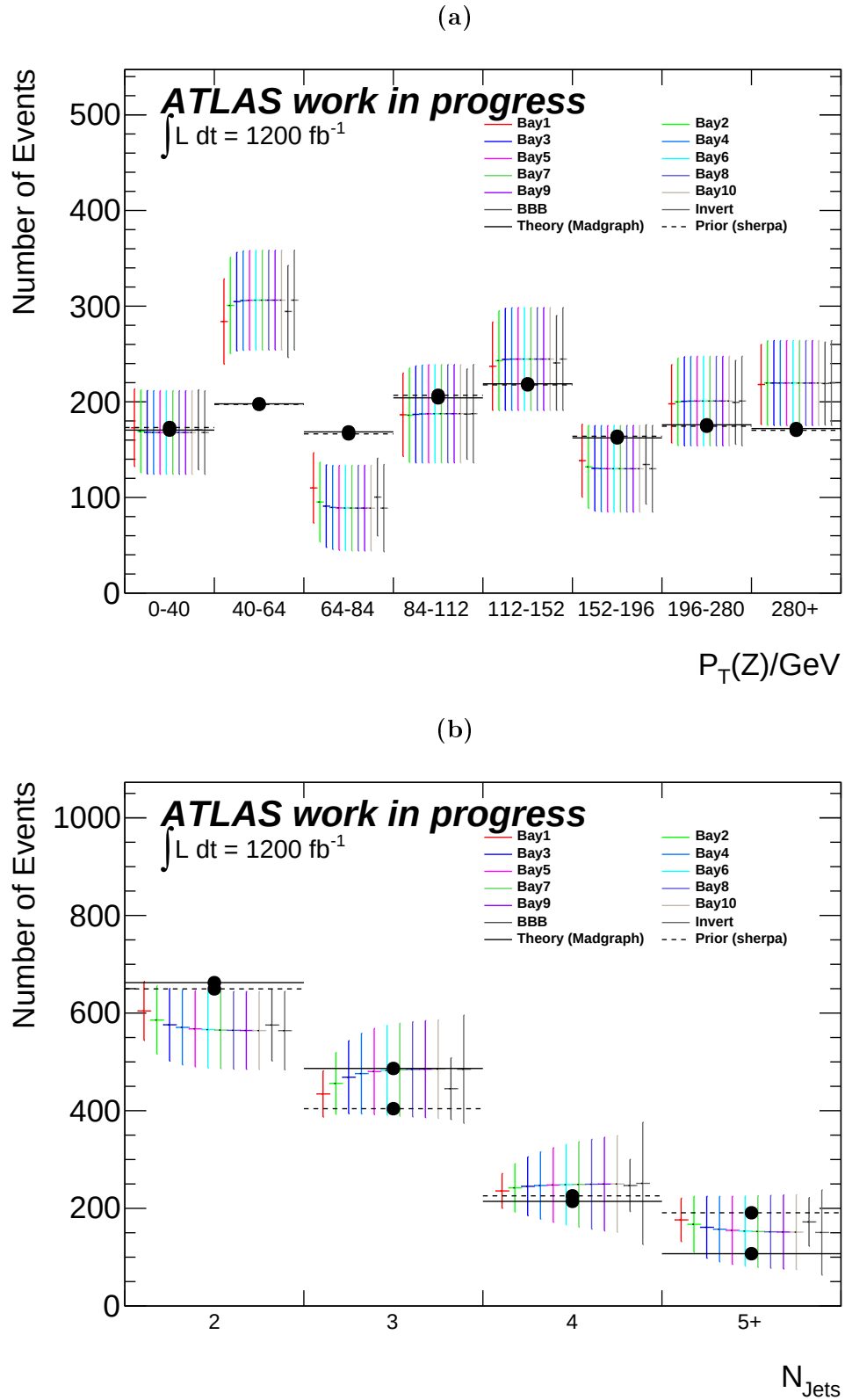
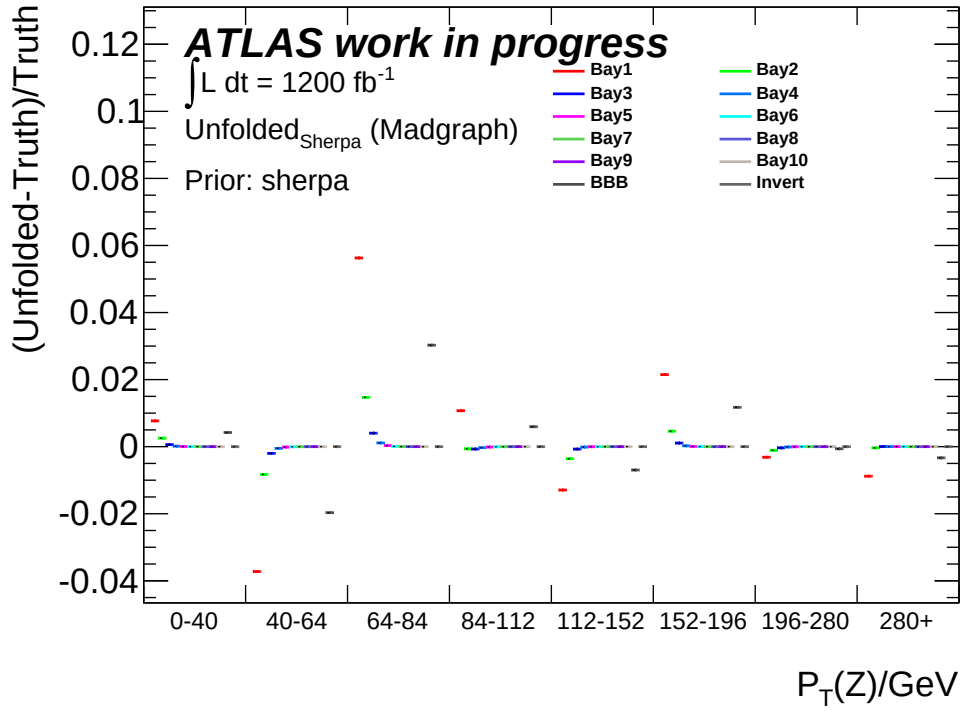


Abbildung 6.25: Dargestellt sind die mittels Sherpa durch verschiedene Entfaltungsmethoden ermittelten Entfaltungsergebnisse inklusive ihrer statistischen und systematischen Unsicherheiten. Als Daten wurden Fluktuationen der Madgraph *reco*-Verteilung von $P_T(Z)$ (a) bzw. von N_{Jets} (b) bei 1200 fb^{-1} verwendet.

(a)



(b)

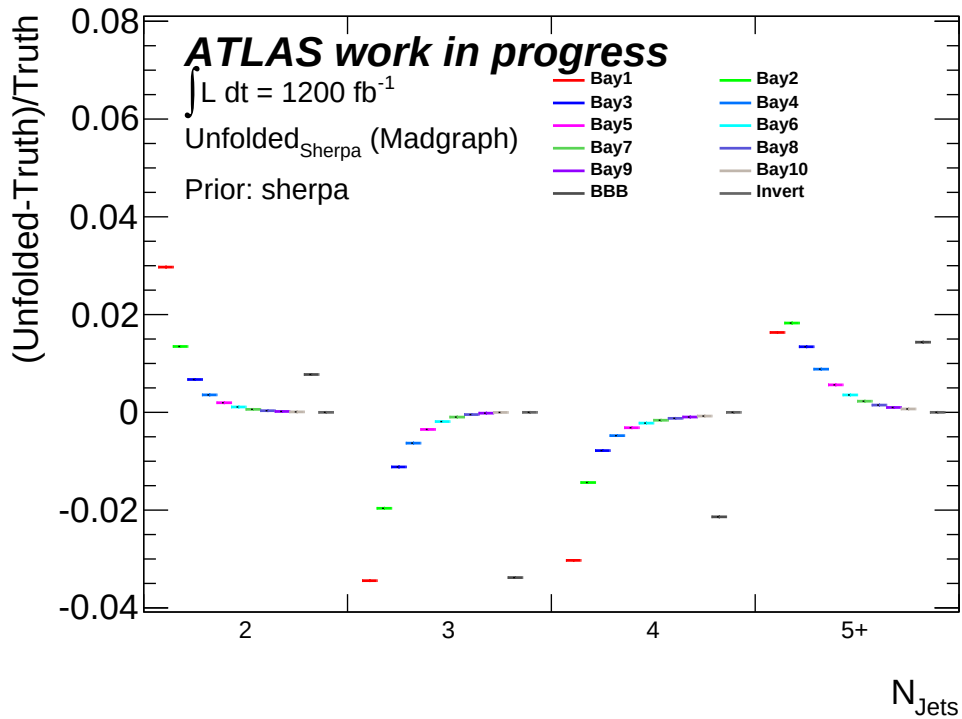


Abbildung 6.26: Die dargestellten Entfaltungsunsicherheiten wurden für die Entfaltung mittels Sherpa bestimmt. Als Daten wurden Fluktuationen der Madgraph *reco*-Verteilung von $P_T(Z)$ (a) bzw. von N_{Jets} (b) bei 1200 fb^{-1} verwendet.

Tabelle 6.7: Aufgezählt sind die zur Berechnung der systematischen Unsicherheit verwendeten Systematiken

MET_SoftTrk_ResoPara	MET_SoftTrk_ResoPerp
MET_SoftTrk_ScaleDown	MET_SoftTrk_ScaleUp
MUON_EFF_STAT_LOWPT__1down	MUON_EFF_STAT_LOWPT__1up
MUON_EFF_STAT__1down	MUON_EFF_STAT__1up
MUON_EFF_SYS_LOWPT__1down	MUON_EFF_SYS_LOWPT__1up
MUON_EFF_SYS__1down	MUON_EFF_SYS__1up
MUON_ID__1down	MUON_ID__1up
MUON_ISO_STAT__1down	MUON_ISO_STAT__1up
MUON_ISO_SYS__1down	MUON_ISO_SYS__1up
MUON_MS__1down	MUON_MS__1up
MUON_SAGITTA_RESBIAS__1down	MUON_SAGITTA_RESBIAS__1up
MUON_SAGITTA_RHO__1down	MUON_SAGITTA_RHO__1up
MUON_SCALE__1down	MUON_SCALE__1up
MUON_TTVA_STAT__1down	MUON_TTVA_STAT__1up
MUON_TTVA_SYS__1down	MUON_TTVA_SYS__1up
EG_RESOLUTION_MATERIALCALO__1down	EG_RESOLUTION_MATERIALCALO__1up
EG_RESOLUTION_MATERIALCRYO__1down	EG_RESOLUTION_MATERIALCRYO__1up
EG_RESOLUTION_MATERIALGAP__1down	EG_RESOLUTION_MATERIALGAP__1up
EG_RESOLUTION_MATERIALIBL__1down	EG_RESOLUTION_MATERIALIBL__1up
EG_RESOLUTION_MATERIALID__1down	EG_RESOLUTION_MATERIALID__1up
EG_RESOLUTION_MATERIALPP0__1down	EG_RESOLUTION_MATERIALPP0__1up
EG_RESOLUTION_PILEUP__1down	EG_RESOLUTION_PILEUP__1up
EG_RESOLUTION_SAMPLINGTERM__1down	EG_RESOLUTION_SAMPLINGTERM__1up
EG_RESOLUTION_ZSMEARING__1down	EG_RESOLUTION_ZSMEARING__1up
EG_SCALE_E4SCINTILLATOR__1down	EG_SCALE_E4SCINTILLATOR__1up
EG_SCALE_G4__1down	EG_SCALE_G4__1up
EG_SCALE_L1GAIN__1down	EG_SCALE_L1GAIN__1up
EG_SCALE_L2GAIN__1down	EG_SCALE_L2GAIN__1up
EG_SCALE_LARCALIB_EXTRA2015PRE__1down	EG_SCALE_LARCALIB_EXTRA2015PRE__1up
EG_SCALE_LARCALIB__1down	EG_SCALE_LARCALIB__1up
EG_SCALE_LARELECCALIB__1down	EG_SCALE_LARELECCALIB__1up
EG_SCALE_LARELECUNCONV__1down	EG_SCALE_LARELECUNCONV__1up
EG_SCALE_LARUNCONVCALIB__1down	EG_SCALE_LARUNCONVCALIB__1up
EG_SCALE_MATCALO__1down	EG_SCALE_MATCALO__1up
EG_SCALE_MATCRYO__1down	EG_SCALE_MATCRYO__1up
EG_SCALE_MATID__1down	EG_SCALE_MATID__1up
EG_SCALE_MATPP0__1down	EG_SCALE_MATPP0__1up
EG_SCALE_PEDESTAL__1down	EG_SCALE_PEDESTAL__1up
EG_SCALE_PS__1down	EG_SCALE_PS__1up
EG_SCALE_S12EXTRALASTETABINRUN2__1down	EG_SCALE_S12EXTRALASTETABINRUN2__1up
EG_SCALE_S12__1down	EG_SCALE_S12__1up
EG_SCALE_WTOTS1__1down	EG_SCALE_WTOTS1__1up
EG_SCALE_ZEESTAT__1down	EG_SCALE_ZEESTAT__1up
EG_SCALE_ZEESYST__1down	EG_SCALE_ZEESYST__1up
EL_EFF_ID_TOTAL_INPCOR_PLUS_UNCOR__1down	EL_EFF_ID_TOTAL_INPCOR_PLUS_UNCOR__1up
EL_EFF_Iso_TOTAL_INPCOR_PLUS_UNCOR__1down	EL_EFF_Iso_TOTAL_INPCOR_PLUS_UNCOR__1up
EL_EFF_Reco_TOTAL_INPCOR_PLUS_UNCOR__1down	EL_EFF_Reco_TOTAL_INPCOR_PLUS_UNCOR__1up
EL_EFF_TriggerEff_TOTAL_INPCOR_PLUS_UNCOR__1down	EL_EFF_TriggerEff_TOTAL_INPCOR_PLUS_UNCOR__1up
EL_EFF_Trigger_TOTAL_INPCOR_PLUS_UNCOR__1down	EL_EFF_Trigger_TOTAL_INPCOR_PLUS_UNCOR__1up
FT_EFF_B_systematics__1down	FT_EFF_B_systematics__1up

Erklärung

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

Max Vincent Stange
Dresden, Juni 2018

7 Literaturverzeichnis

- [1] A. Armbruster, K. Kroeninger, B. Malaescu, and F. Spano. Practical considerations for unfolding. Technical Report ATL-COM-PHYS-2014-277, CERN, Geneva, Apr 2014.
- [2] Giles Barr, Robin Devenish, Roman Walczak, and Tony Weidberg. *Particle Physics in the LHC Era*. Oxford University Press, New York, 2016.
- [3] Olaf Behnke, Kevin Kröninger, Grégory Schott, and Thomas Schörner-Sadenius. *Data Analysis in High Energy Physics - A Practical Guide to Statistical Methods*. John Wiley & Sons, New York, 2013.
- [4] Frédéric Bordry and Cristina Agrigoroae. Preparing for the last year of LHC Run 2. *CERN Document Server*, Feb 2018.
- [5] CERN. *EWUnfolding Framework*, . <https://svnweb.cern.ch/trac/atlasphys-sm/browser/Physics/StandardModel/ElectroWeak/Analyses/EWUnfolding/branches/EWUnfolding-00-00-01-branch?order=author>.
- [6] CERN Tim Adye, Richard Claridge, Kerstin Tackmann, and Fergus Wilson. *ROOUnfold Framework*, Geneva, 2011.
- [7] The ATLAS collaboration. Measurement of $W^\pm Z$ boson pair-production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector and confidence intervals for anomalous triple gauge boson couplings. *ATLAS-CONF-2016-043*, 2016.
- [8] G. D’Agostini. A multidimensional unfolding method based on bayes’ theorem. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 362(2):487 – 498, 1995.
- [9] M. Tanabashi et al. (Particle Data Group). 2018 review of particle physics, kinematics. <http://pdg.lbl.gov/2018/reviews/rpp2018-rev-kinematics.pdf>. Accessed: 18.06.2018.
- [10] M. Tanabashi et al. (Particle Data Group). 2018 review of particle physics, w boson. <http://pdg.lbl.gov/2018/listings/rpp2018-list-w-boson.pdf>. Accessed: 18.06.2018.
- [11] M. Tanabashi et al. (Particle Data Group). 2018 review of particle physics, z boson. <http://pdg.lbl.gov/2018/listings/rpp2018-list-z-boson.pdf>. Accessed: 18.06.2018.

- [12] Lyndon R Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001. 164 p, 2008. Seite 1.
- [13] David P. Landau and Kurt Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, Cambridge, 2009.
- [14] Christopher Meyer. Standardmodelunfoldingnew 2018-04-24. <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/StandardModelUnfoldingNew>. Accessed: 18.06.2018.