

Study of Different Unfolding Methods of Kinematic Distributions of the $WZ \rightarrow WZ$ Scattering with Data and Simulations of the ATLAS Detector at the LHC

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

vorgelegt von

Tim Herrmann
geboren am 22.09.1991 in Dresden

Institut für Kern- und Teilchenphysik
Fachrichtung Physik
Fakultät Mathematik und Naturwissenschaften
Technische Universität Dresden
2017

Eingereicht am 30. März 2017

1. Gutachter: Prof. Dr. Kobel
2. Gutachter: Dr. Siegert

Zusammenfassung

In dieser Arbeit wird für Verteilungen der Vektor-Boson-Streuung im Kanal $WZ \rightarrow WZ$ in leptonischen Endzuständen untersucht, welche Entfaltungsmethoden für die P-value Bestimmung bei statistischen Tests geeignet sind. Die $WZ \rightarrow WZ$ Streuung wird vom derzeit am besten bestätigten Modell der Teilchenphysik, dem Standardmodell der Teilchenphysik, vorausgesagt - wurde aber noch nicht gemessen. Es wird erwartet, 100 fb^{-1} mit dem ATLAS Detektor im Run 2 des LHC aufzunehmen. Mit dieser integrierten Luminosität wird eine Beobachtung des Prozesses, durch eine Messung des Wirkungsquerschnittes, erwartet. Die auf mögliche Abweichungen der $WZ \rightarrow WZ$ Streuung vom Standard Modell sensitiven Verteilungen der transversale Masse vom WZ System $M_T(WZ)$ und des transversalen Impulses des Z-Bosons p_T^Z werden in dieser Arbeit analysiert.

Beim Vergleich von Theorie und Daten müssen Detektoreffekte mit berücksichtigt werden, wofür entweder die Theorievorhersage mit den Detektoreffekten gefaltet oder die Daten entfaltet werden. Im Rahmen der Studien konnten für computergestützte statistische Tests kaum Vorteile des Entfaltungsansatzes gegenüber einer Faltung der Theorievorhersage festgestellt werden. Wenn Datenverteilung und Theorievorhersage unabhängig von Detektoreffekten visuell verglichen werden sollen, dann ist Entfaltung notwendig. Bei visuellen Abschätzungen kann Regularisierung wie in 'Bayesian Iterative Unfolding' nützlich sein, bei computergestützter Berechnung hat die Entfaltungsmethode 'Matrix Inversion' die meisten Vorteile.

Abstract

It is analyzed in this work which unfolding methods are suited for the P-value calculation in statistical tests. It is analyzed for distributions of Vector Boson Scattering in the channel $WZ \rightarrow WZ$ for fully leptonic final states. $WZ \rightarrow WZ$ scattering is predicted by the most successful model of particle physics, the Standard Model of Particle Physics - but was not measured yet. It is scheduled to record 100 fb^{-1} with the ATLAS detector in Run 2 at LHC. With that integrated luminosity an observation of that process, via a cross section measurement, is expected. The distributions of the transverse mass of the WZ system $M_T(WZ)$ and the transverse momentum of the Z boson p_T^Z which are sensitive to deviations of the $WZ \rightarrow WZ$ scattering from the Standard Model are analyzed in this work.

For comparisons between data and theory predictions detector effect have to be considered, for which the theory has to be folded or the data has to be unfolded. In this study, no significant advantages of using the unfolding approach instead of the folding approach have been found for computer based statistical tests. If data and theory distributions should be compared independently from detector effects then unfolding is necessary. For visual comparisons regularization, like in 'Bayesian Iterative Unfolding', can be useful, for computerized calculations the unfolding method 'Matrix Inversion' has the most advantages.

Contents

1	Introduction	1
2	Theoretical Foundation	3
2.1	Standard Model	3
2.1.1	Electroweak Symmetry Breaking	6
2.1.2	Indications for Physics Beyond the Standard Model	7
2.2	Vector Boson Scattering	8
2.2.1	Process Definition and Studied Variables	8
2.3	Effective Field Theories and Anomalous Gauge Coupling	10
3	Experimental Setup	11
3.1	LHC	11
3.2	ATLAS Detector	11
3.2.1	Coordinates	11
3.2.2	Subdetectors	12
3.2.3	Data taking	13
4	Monte Carlo Event Generators	15
5	Statistical Methods	17
5.1	Frequentist vs. Bayesian Interpretation	17
5.2	The P-value	19
5.2.1	The Likelihood and χ^2 Function	20
5.3	Probability in Terms of Standard Deviations and Conventions	21
6	Unfolding	23
6.1	Unfolding Methods	27
6.1.1	Bin-By-Bin Unfolding	28
6.1.2	Matrix Inversion (MI)	28
6.1.3	Regularized Unfolding	29
6.2	Propagation of Uncertainties in the Unfolding Approaches	35

7	Treadment of Statistical Uncertainties in Statistical Tests with Unfolded Distributions	43
7.1	Influence on the P-value	45
7.1.1	Different Test Statistic	53
8	Discussion and Conclusion	59
A	Used Monte Carlo Samples	63
B	Selection Criteria	65
C	Additional Plots	67
	List of Figures	69
	List of Tables	71
D	Bibliography	73

1 Introduction

The scattering of vector bosons is predicted by the Standard Model of Particle Physics but the scattering of W and Z bosons has not been significantly observed yet. During Run 1 data was taken at ATLAS corresponding to an integrated luminosity of 20.3 fb^{-1} at LHC at a center of mass energy of $\sqrt{s} = 8 \text{ TeV}$ in 2012. With that data the cross section of $WZ \rightarrow WZ$ scattering was measured with a significance of 1.8σ [1]. With the planned integrated luminosity of 100 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$ in Run 2 lasting until 2018 [2] it is expected to observe the $WZ \rightarrow WZ$ scattering process. With it the cross section can be measured.

For rising event numbers differential cross sections represented in histograms become interesting in comparisons with different theory predictions. To calculate the significance of an observed data distribution given a theory prediction it is necessary to fold the theory or to unfold the data. In this work different unfolding approaches and methods are compared in order to determine which unfolding approaches are suitable for low event numbers.

Firstly, there is a brief introduction to the theoretical foundations to understand WZ physics in Chapter 2. Afterwards, in Chapter 3 the experimental setup to measure distributions of the WZ scattering is described. Chapter 4 is about Monte Carlo generators needed in order to calculate the theoretical predictions. Chapter 5 introduces statistical methods used to make statements about the validity of theories and their comparability with the observed data. In Chapter 6 several unfolding approaches and methods are introduced and discussed. In Chapter 7 the statistical differences between two unfolding approaches are discussed. Finally, in Chapter 8 the results are summarized.

2 Theoretical Foundation

Theories are developed in order to describe and predict processes in nature. The Standard Model provides a theoretical description of high energy particle physics.

2.1 Standard Model

The Standard Model (SM) is a very successful relativistic quantum field theory (QFT) which has been developed during the 20th century. It predicted several particles successfully. The greatest success so far was maybe the correct prediction of the existence of a Higgs boson in 1964 [3] [4], which has been measured in 2012 [5].

The SM describes the particles and their interactions including three of four known interactions: the electromagnetic, weak and strong interaction. The gravitational force is missing but is expected to have no measurable effect in high energy physics for the current energies of approximately 10^4 GeV. For energies in the order of the Planck energy of 10^{19} GeV quantum gravity is expected to start having a measurable impact. The SM can be fully described by the field content - the particles - and by a Lagrangian function determining the propagation and interaction of these particles.

A more detailed description can be found in [6] and [7].

The particles of the SM can be divided into bosons with integer spins and fermions with half integer spins¹. The bosons of the SM are listed in Table 2.2. All elementary fermions (see Table 2.1) in the SM have a spin of $\frac{1}{2}$. They can be subdivided into quarks and leptons, where quarks interact with the strong force while leptons do not.

The SM Lagrangian follows the Poincaré symmetry of special relativity and three local gauge symmetries, which are:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y. \quad (2.1)$$

$SU(3)_C$ describes the quantum chromodynamics (QCD) describing the strong interaction. While $SU(2)_L \otimes U(1)_Y$ describes the electroweak interaction (ew) which is a unified description of the electromagnetic and weak interaction.

Each local gauge symmetry has an associated charge. For $SU(3)_C$ it is the color charge C ,

¹Natural units, i.e. $c = \hbar = k_B = 1$, are used in this work

	Fermion	Electric Charge Q	Mass m in MeV
1 st Generation	e^- electron	-1	0.5110
	ν_e electron-neutrino	0	$< 2 \cdot 10^{-6}$
	u up	$+\frac{2}{3}$	$2.2^{+0.6}_{-0.4}$
	d down	$-\frac{1}{3}$	$4.7^{+0.5}_{-0.4}$
2 nd Generation	μ^- muon	-1	105.7
	ν_μ muon-neutrino	0	$< 2 \cdot 10^{-6}$
	c charm	$+\frac{2}{3}$	$(1.27 \pm 0.03) \cdot 10^3$
	s strange	$-\frac{1}{3}$	96^{+8}_{-4}
3 rd Generation	τ^- tau	-1	1777
	ν_τ tau-neutrino	0	$< 2 \cdot 10^{-6}$
	t top	$+\frac{2}{3}$	$(1.732 \pm 0.012) \cdot 10^5$
	b bottom	$-\frac{1}{3}$	$(4.18^{+0.04}_{-0.03}) \cdot 10^3$

Table 2.1: Summary of fermions in the Standard Model.[8] The first two fermions of each generation are leptons while the last two are quarks. Each quark exists in three colors. An anti-particle with the same properties but reversed signs of the charges exists for each fermion.

	Boson	Electric Charge Q	Spin	Interaction	Mass m in GeV
γ	photon	0	1	electromagnetic	$< 1 \cdot 10^{-27}$
$W^\pm$	W bosons	0	1	weak	80.385 ± 0.015
Z^0	Z boson	0	1	weak	91.1876 ± 0.0021
g	gluons	0	1	strong	0
H	Higgs	0	0		125.09 ± 0.24

Table 2.2: Summary of bosons in the Standard Model.[8]

for $SU(2)_L$ it is the weak isospin I_W and for $U(1)_Y$ it is the weak hypercharge Y . These charges are the conserved quantities which have to exist in local gauge symmetries according to the Noether-Theorem. Together with the demand for a consistent relativistic QFT it implies that each particle must have an anti-particle with same mass but opposite signed charges.

The local gauge invariance is expressed via generators which are used to define gauge invariant transformations. A $SU(N)$ has $N^2 - 1$ generators and a $U(N)$ has N^2 . In the case of the $SU(2)_L$ the generators are the three Pauli matrices τ_i and in the case of the $SU(3)_C$ these are the eight Gell-Mann matrices T_i . The Pauli matrices act on doublets of particles with weak isospin and the Gell-Mann matrices act on triplets of color charged particles. To fulfill the local gauge symmetries a gauge field has to be introduced for each generator. For the QCD these are the eight gluon fields $G^{1\dots 8}$ and for the EW theory these are the fields B, W^1, W^2 and W^3 .

The position in the weak isospin doublet defines the z-component I_W^3 of the weak isospin. The top element of a doublet has $I_W^3 = \frac{1}{2}$ and the bottom element has $I_W^3 = -\frac{1}{2}$. Only left-handed particles have a weak isospin. There is one lepton doublet L_L for each generation consisting of a charged lepton and its anti-neutrino and one quark doublet Q_L for each generation consisting of the $\frac{2}{3}$ electric charged quark and a $-\frac{1}{3}$ electric charged quark. For each doublet also a second one with its anti-particles in it exists. The lower entry in the quark doublet is a mixture of the 'down'-type quarks in each case.

The SM Lagrangian can be split into:

$$\mathcal{L}_{SM} = \mathcal{L}_{leptonic} + \mathcal{L}_{bosonic} + \mathcal{L}_{Higgs}. \quad (2.2)$$

Here, $\mathcal{L}_{bosonic}$ includes the kinematic terms of the bosons and their interactions, while $\mathcal{L}_{leptonic}$ includes the kinematic terms of the fermions and interactions of fermions with bosons. The term $\mathcal{L}_{higgs}$ adds some extra terms to the $\mathcal{L}_{SM}$ generating among others mass terms of the bosons and fermions.

The leptonic and bosonic Lagrangian are:

$$\begin{aligned} \mathcal{L}_{leptonic} = \sum_{j=1}^3 & i\bar{L}_L^j \gamma^\mu D_\mu L_L^j + i\bar{l}_R^j \gamma^\mu D_\mu l_R^j + \\ & i\bar{Q}_L^j \gamma^\mu D_\mu Q_L^j + i\bar{u}_R^j \gamma^\mu D_\mu u_R^j + i\bar{d}_R^j \gamma^\mu D_\mu d_R^j + h.c. \end{aligned} \quad (2.3)$$

$$\mathcal{L}_{bosonic} = -\frac{1}{4}W_{a,\mu\nu}W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G_{a,\mu\nu}G_a^{\mu\nu} \quad (2.4)$$

with

$$D_\mu = \partial_\mu + ig_W \frac{\tau_a}{2} W_\mu^a + ig_Y Y B_\mu + ig_s T_a G_\mu^a \quad (2.5)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (2.6)$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_W \epsilon^{abc} W_\mu^b W_\nu^c \quad (2.7)$$

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_W f^{abc} G_\mu^b G_\nu^c \quad (2.8)$$

where h.c. stands for the hermitian conjugate of the expression, ϵ^{abc} represents the Levi-Civita symbol and f^{abc} denotes the structure constant of $SU(3)$.

The mass eigenstates of the bosons of the EW interaction photon A , $W^\pm$ -Boson and Z -Boson are a mixture of the gauge fields B, W^1, W^2 and W^3 :

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \cdot \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \quad (2.9)$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad (2.10)$$

2.1.1 Electroweak Symmetry Breaking

In order to have mass terms in the Lagrangian, while keeping it local gauge invariant, the scalar complex $SU(2)_L$ doublet field Φ with an weak hypercharge $Y = \frac{1}{2}$ is part of the SM:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} \quad (2.11)$$

The SM Lagrangian includes that field in the potential $V(\Phi)$, the kinematic term $|D_\mu \Phi|^2$ and the Yukawa coupling terms $\mathcal{L}_{YUKAWA}$:

$$\mathcal{L}_{Higgs} = |D_\mu \Phi|^2 - \underbrace{(\lambda(\Phi^\dagger \Phi)^2 - \mu^2 \Phi^\dagger \Phi)}_{V(\Phi)} + \underbrace{\sum_{j=1}^3 y_l^j \bar{L}_L^j \Phi l_R^j + y_u^j \bar{Q}_L^j i\sigma_2 \Phi u_R^j + y_d^j \bar{Q}_L^j \Phi d_R^j}_{\mathcal{L}_{YUKAWA}} + h.c. \quad (2.12)$$

The potential $V(\Phi)$ introduces two free scalar parameters $\mu, \lambda > 0$. The ground state Φ_0 has to minimize $V(\Phi_0)$. Leading to the requirement:

$$|\Phi_0|^2 = \frac{\mu^2}{2\lambda} =: \frac{v^2}{2} \quad (2.13)$$

with the vacuum expectation value v . There is an infinite number of potential Φ_0 values which fulfill this requirement. In the SM Φ_0 is set to:

$$\langle 0 | \Phi_0 | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (2.14)$$

By choosing a specific value Φ_0 , the $SU(2)_L \otimes U(1)_Y$ gets spontaneous broken, which is called the electroweak symmetry breaking. The ground state Φ_0 remains local gauge invariant under $U(1)_Q$ with the associated electric charge Q , defined as:

$$Q := I_W^3 + \frac{Y}{2}. \quad (2.15)$$

$U(1)_Q$ corresponds to the symmetry of quantum electrodynamics (QED). So the EW-symmetry gets broken into the symmetry of QED:

$$SU(2)_L \otimes U(1)_Y \rightarrow U(1)_Q \quad (2.16)$$

There is a massive excitation of the ground state Φ_0 which leads to the existence of an additional boson, the Higgs boson H . The kinematic term $|D_\mu \Phi|^2$ generates the mass terms of the bosons and the interactions between the Higgs boson and the $W^\pm$ and Z bosons. Through the Yukawa coupling terms $\mathcal{L}_{YUKAWA}$ mass terms for the fermions are introduced, except for the neutrinos which only exist as left-handed particles in the SM. There is one free parameter y_X for each fermion defining its mass. The two new scalar parameters μ and λ can be calculated from the mass ratio of the W and Z -Boson and the mass of the Higgs boson.

2.1.2 Indications for Physics Beyond the Standard Model

Although being a very successful theory the SM can not explain everything. There have to be some physics beyond the Standard Model (BSM) for which several BSM theories exist. Some topics which can only be described by BSM theories are:

- Neutrinos have no mass in the SM. But oscillations between their flavor eigenstates are observed which is only possible with masses.
- The gravitation is missing in the SM.
- The rotation speed of galaxies indicate that there are additional particles, which interact gravitational - called dark matter.
- There is also something which accelerates the expansion of the universe - dark energy - which is also not described by the SM.

2.2 Vector Boson Scattering

The SM allows the scattering of photons, W and Z-bosons, called Vector Boson Scattering (VBS). Triple and Quartic gauge couplings are possible. Triple gauge coupling between the bosons of the electroweak interaction has already been measured at LEP, but quartic gauge couplings has not been. That is one reason why quartic gauge coupling is studied at LHC to verify the SM. Additionally, the decay channel including VBS is sensitive to potential new particles. In order to be a consistent theory the cross section of processes involving quartic gauge couplings has to be finite for rising energies. This is the case with the SM Higgs boson. But there could also be more than one Higgs boson or other particles changing the equilibrium. That is the reason for the study of quartic gauge coupling which could result in signs for particles in addition to the observed Higgs boson.

2.2.1 Process Definition and Studied Variables

For proton collisions like at LHC, VBS can be studied best for the process where one quark of each proton radiates one vector boson, which then scatter.

The processes on leading order which produce two vector bosons from two vector bosons can be seen in Figure 2.1.

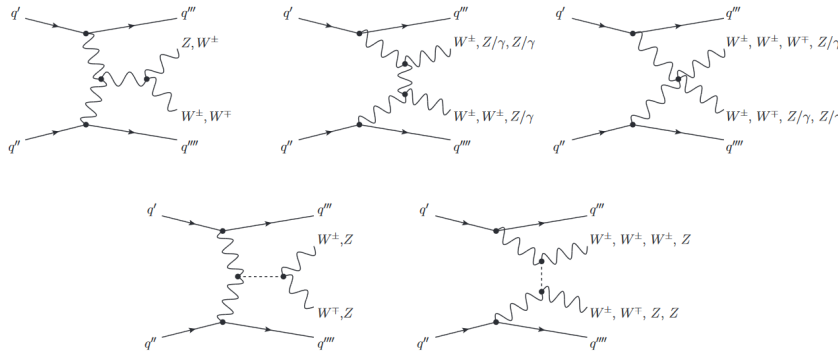


Figure 2.1: "Possible Feynman diagrams for VBS with final states containing massive electroweak gauge bosons. The primes of the quarks indicate possible changes in flavour." [9]

Since in this work the scattering of $WZ \rightarrow WZ$ is studied it is searched for the intermediate state $WZjj$. For a purer signal, only leptonic decays of the W and Z boson are studied, leading to a final state of three charged leptons, one neutrino and two jets. Since also other processes could have the same final state selection criteria are applied on the final state to increase the fraction of signal events. The applied selection criteria on the leptons and the jets can be

found in [10].

If there are additional particles involved in the scattering process they would be identifiable by additional heavy resonances in the spectrum of the invariant mass of the two vector bosons. So, the invariant mass of the sum of W and Z would be an interesting variable, given by:

$$m^{WZ} := \sqrt{\left(\sum_{\ell=1}^3 E^\ell + E^\nu\right)^2 - \left(\sum_{\ell=1}^3 \vec{p}^\ell + \vec{p}^\nu\right)^2} \quad (2.17)$$

where ℓ stands for the three charged leptons and ν denotes the neutrino coming from the W decay. Since the neutrino is not detectable with the ATLAS detector, the momentum $\vec{p}^\nu$ and the energy E^ν remain unknown. The only information which is available is the missing energy in the transverse plane E_T^{miss} . So, the contributions to m^{WZ} in the transverse plane, namely m_T^{WZ} , can be determined. For massless leptons this is [10]:

$$M_T(WZ) := \sqrt{\left(\sum_{\ell=1}^3 p_T^\ell + E_T^{miss}\right)^2 - \left[\left(\sum_{\ell=1}^3 p_x^\ell + E_x^{miss}\right)^2 + \left(\sum_{\ell=1}^3 p_y^\ell + E_y^{miss}\right)^2\right]} \quad (2.18)$$

One has $m_T^{WZ} < m^{WZ}$. [11] The expected distribution is shown in Figure 2.2.

Also the transverse momentum of the Z-boson p_T^Z is sensitive to heavy resonances of the WZ system.

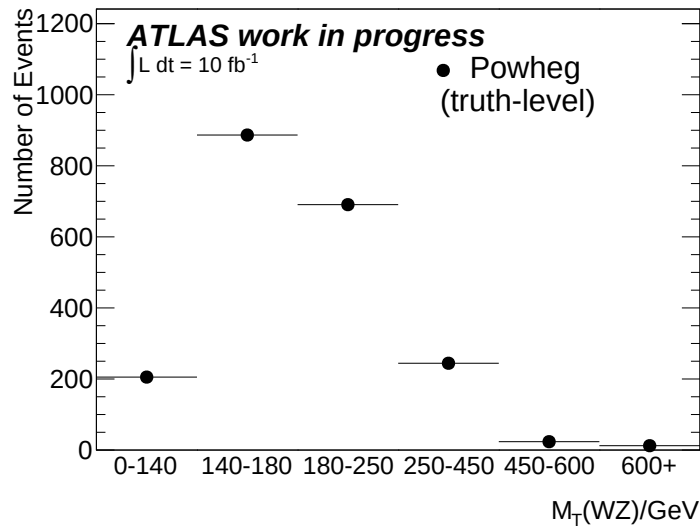


Figure 2.2: Expected $M_T(WZ)$ distribution generated with Powheg.

If the additional resonance is at a high energy e.g. over 1 TeV then the effects which it would

have for the measurable energies can be described by an effective field theory.

2.3 Effective Field Theories and Anomalous Gauge Coupling

Effective field theories (EFTs) introduce additional terms in the SM of the form [12]:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_{d>4} \sum_i \frac{f_i^{(d)}}{\Lambda^{d-4}} \mathcal{O}_i^{(d)} \quad (2.19)$$

where $\mathcal{O}_i^{(d)}$ represents an operator of the order d , Λ is the scale of new physics and is introduced to the power of $-(d-4)$ in order to have the same units as in the SM Lagrangian and $f_i^{(d)}$ is a free parameter which quantifies the deviation from the SM. Additional resonances in the WZ scattering would be of order $d \geq 8$.

The additional terms coming from additional resonances in the WZ scattering can be interpreted as anomalous quartic gauge couplings (aQGC). By setting limits on these aQGC parameters, limits are implicitly set on additional resonances in the WZ scattering. Unfortunately, there was no simulation of these anomalous couplings available at the time scale of this work. Fortunately, statistical tests for parameters of quartic and triple gauge coupling are very similar. Since for triple gauge couplings there are already enough events the studies are performed on a aTGC parameter.

The term of aTGC studied in this work as example for a BSM theory is :

$$\mathcal{L}_{\lambda_Z aTGC} = -ig_{WWZ} \frac{\lambda_Z}{M_W^2} W_\mu^{\nu+} W_\nu^{-\rho} Z_\rho^\mu \quad (2.20)$$

where λ_Z is the strength of the additional couplings.[12]

3 Experimental Setup

The studies performed in this work are based on simulated data of the ATLAS detector. Therefore, the experimental setup of the ATLAS detector at the Large Hadron Collider is shortly described here.

3.1 LHC

Near Geneva 45 - 170 m under the surface lies the circular accelerator Large Hadron Collider (LHC)[13] with a circumference of 26.7 km. It is designed to accelerate and collide two circulating beams of heavy ions or protons. There are periods when protons are collided and other periods for heavy ion collisions. The proton beams consist of up to 2808 proton bunches which get accelerated from 450 GeV to 6.5 TeV. The protons get pre-accelerated in a series of linear and circular accelerators.

At four points detectors are situated where the beams can be crossed. Two of the four detectors are designed to detect as many high energetic decay products as possible of proton-proton collisions: CMS and ATLAS. They act as two independent experiments to perform high precision measurements of the Standard Model and to search for new physics beyond the Standard Model. The biggest success so far was the observation of the Higgs Boson, which was discovered in both the detectors in 2012 [5].

3.2 ATLAS Detector

The ATLAS detector [14] is a general purpose detector which is designed to perform precise measurements of the SM and search for BSM. Therefore it is necessary that it measures particles in all directions with a good energy, momentum and position precision. It should identify as many particles coming from the collision point as possible.

3.2.1 Coordinates

To describe coordinates in the ATLAS-detector Cartesian coordinates are used. The origin of ordinates is the nominal collision point of the particles. The x-axis points to the middle of the ring accelerator, the y-axis points upwards and the z-axis points in the beam direction such

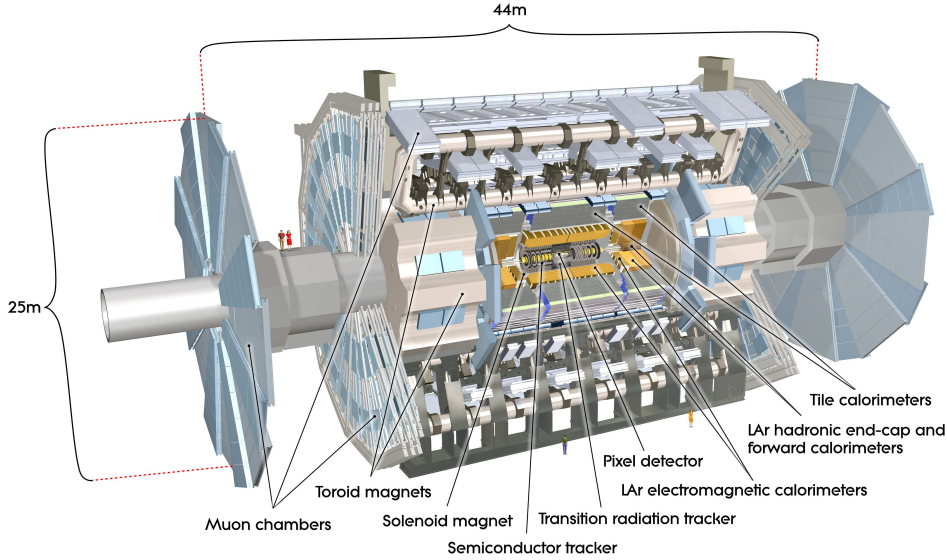


Figure 3.1: Schematic representation of the ATLAS detector [14].

that it is a right handed coordinate system.

Spherical coordinates are used to denote direction in space starting from the collision point with the polar angle ϕ and the azimuthal angle θ . The polar angle ϕ is the angle with respect to the x-direction in the plane orthogonal to the beam direction. The azimuthal angle θ is the angle with respect to the beam direction. As measure of the angle θ the pseudorapidity η is defined as

$$\eta := -\ln(\tan(\theta/2)). \quad (3.1)$$

The pseudorapidity is defined such that $\Delta\eta$ is independent from boosts along the beam direction for massless particles.

3.2.2 Subdetectors

Only particles with a long enough life time reach the detector. They have to fly from the beam center to the border of the beam tube, which are approximately 35 mm. Particles expected to reach the detector are photons, electrons, muons, neutrinos, and light hadrons like pions, protons and neutrons. To distinguish between them four subdetectors are present: the inner detector, the electromagnetic calorimeter, the hadronic calorimeter and the muon system.

The subdetector nearest to the beam pipe is the inner detector, surrounded by a solenoid magnet producing a magnetic field of 2 Tesla. It covers the central region, defined as $|\eta| < 2.5$, and detects charged particles with silicon stripes, pixels and drift tubes filled with xenon and argon. Due to the magnetic field charged particles describe a curved path from which their

transverse momentum can be calculated.

The electromagnetic calorimeter surrounds the inner detector. It consists of a barrel covering the area till $|\eta| < 1.475$. At the end of the barrel there are two end-caps, covering the space till $|\eta| = 3.2$. The electromagnetic calorimeter is built to measure the energy of light electromagnetic interacting particles, like photons, electrons or positrons. There are alternating layers of lead and liquid argon. Lead acts as absorber and let the electromagnetic interacting particles decay into showers. In the liquid argon layers the number of particles is measured. From the measured properties of the shower the energy of the initial particle is recalculated.

The electromagnetic calorimeter is surrounded by the hadronic calorimeter, designed to measure the energy of hadrons. It also consists of layers of absorbers and active medium. Here, steal is used as absorber. It consists of a tile barrel, two end-caps (HEC) and a liquid argon forward calorimeter (FCal). The barrel covers $|\eta| < 1.7$, HEC covers $1.5 < |\eta| < 3.2$ and FCal covers $3.1 < |\eta| < 4.9$ the barrel uses scintillating tiles as active material, whereas HEC and FCal uses liquid argon. The hadronic calorimeter works similar to the electromagnetic by producing showers in the absorbers which are measured by the active material.

The outermost subdetector is the muon system. It measures muons which are the only detectable particles, which are expected to escape the hadronic calorimeter. The muon system is embedded in a magnetic field coming from a toroid magnet. Through this field the muons get redirected and similar to the inner detector the momentum can be calculated from this. The muon trajectories are measured with drift tubes and Cathode Strip Chambers.

3.2.3 Data taking

The measurements of the detector are clustered into events containing all information associated with one inelastic p-p scattering process. Due to technical and financial constrains not all events can be stored. There is a three level trigger system which selects only the interesting events. The information read out from the subdetectors is kept in a pipe line in order to minimize the dead time of the detector.

The hardware based level-one trigger (L1) uses only a few information from the detector to decide within $2.5 \mu s$ whether to keep an event. If it is kept L1 defines regions of interest (RoI) where something interesting happened.

The level-two trigger (L2) uses the RoIs and all measurements within these regions to decide within 40 ms if the event should be discarded.

The last trigger is the event filter which uses all detector information to make a decision within

4 s and reduces the event rate to 200 Hz. L2 and event filter together form the High-Level Trigger whose accepted events are permanently stored.

4 Monte Carlo Event Generators

In order to compare theories with data, theory predictions are needed. Since the calculation of the theory prediction is very complex Monte Carlo methods are used. Which means that events are randomly generated according to their theoretical probabilities. A more detailed description can be found in [15].

The theoretical probabilities can be determined from the Lagrangian by applying the time evolution operator on the initial state to calculate the final state. For high energetic particles (above 1 GeV) only a perturbative calculation is known. In general the calculation in the perturbation theory is stopped after a few orders. Omitting the higher order terms leads to a limited precision. For low energetic QCD-processes (less than 1 GeV) this has to be handled using non-perturbative models because the above sequence does not converge anymore.

In the p-p collisions with a center of mass energy in the order of 13 TeV the protons have to be described as build of partons because of the high energies. The collision occurs between these partons. For the simulation, the momentum of the colliding partons needs to be known. These momentums are distributed according to the Parton Distribution Function (PDF). In the MC simulation, the momenta of the colliding particles are randomly chosen according to the PDF. It is possible that more than two partons interact during a proton-proton collision. This is included in the simulation as 'underlying event'.

The simulation of the inelastic interaction of the two partons is divided into three steps. The first step simulates the high energetic processes where perturbation theory works. In the second step low energy processes where perturbation calculations break down are simulated. The last step is the consideration of the experimental setup which affects the measurement.

The first step involves the hard scattering process which is calculated till a given order in perturbation theory. The time evolution of the incoming partons are simulated according to these probabilities. Since single partons are possible in this intermediate final state this stage of event simulation is called parton level.

The second step performs the showering and hadronization of the partons into hadrons. In

this step, also the decay of particles with short live time not reaching the detector is simulated. At the end, only particles being able to reach the detector are left. These particles are called to be on hadron level.

The last step is the simulation of the detector effects where it is simulated what is actually measured. This includes the interactions of the incoming particles with the material distribution in the detector which leads to electronic signals in the subdetectors. From the measured information tracks and calorimeter clusters are reconstructed. Which are used to reconstruct particles like electrons, muons or jets. The same algorithms are used for the reconstruction of real data.

In this step, the occurrence of 10-40 inelastic proton-proton scatterings per bunch crossing [16] is also included in the simulation. This leads to additional particles in the detector - mostly low energy jets. Events passing also this last step are called on detector level.

A crucial point in the simulation is the matching of partons originating from the hard scattering process and those radiated in the hadronization process. This makes the distribution of the number of jets on hadron level interesting for developers of Monte Carlo generators to check how good this matching works.

For the generation of events three different MC generators are used in this work. Their individual characteristics are shortly outlined below. The Powheg and Sherpa samples from [10] are used in this work.

Sherpa Sherpa [17] is a multi-purpose event generator for the simulation of user defined processes with NLO QCD accuracy up to hadron level.

Powheg Powheg [18] is an event generator for predefined processes simulated also with NLO QCD accuracy. Powheg has been interfaced with Pythia for the shower simulation for the generation of the samples used in this work.

MC@NLO MC@NLO [19] also uses predefined processes and simulates events with NLO QCD accuracy. Several BSM models are available via an interface to FeynRules. MC@NLO has been interfaced with Herwig for the shower simulation for the generation of the samples used in this work.

5 Statistical Methods

The aim of experiments in fundamental research is to look how well the world is described by theories and to improve the theories if the measurement can not be fully described by current theories. In high energy particle physics, such as conducted at LHC, that means, that the Standard Model is tested and it is checked whether the data can be described by the Standard Model or by a BSM model. Since it is not possible to measure quantum field processes directly, but rather the decaying products of such processes it is not possible to determine the processes which were involved, but rather to make statements about probabilities of measurements and theories. Therefore, statistical tests are performed to decide whether to keep or discard a theory.

At first two different interpretations of the term 'probability' are discussed. Afterwards, two statistical tests are presented. At last, there is a short overview over conventions under which circumstances theories are kept or discarded.

5.1 Frequentist vs. Bayesian Interpretation

In the frequentist interpretation, the probability p of an outcome m is defined as the fraction of cases with outcome m if a test is repeated an infinite amount of times $N \rightarrow \infty$ under identical circumstances.

$$p(m) = \lim_{N \rightarrow \infty} \frac{\text{times outcome } m \text{ occurred}}{N} \quad (5.1)$$

The test does not need to be possible, as for instance the direct measurement whether a theory is true is not possible. A longer discussion about Frequentist vs. Bayesian interpretation can be found for instance in [20].

In this definition, the probability of a specific theory is one for the theory realized in nature and zero for all others, because multiple direct measurements of a theory have always the same result. If the in nature realized theory is not known and not directly measurable this quantity is not useful in the frequentist interpretation. Instead, the conditional probability to measure a certain dataset given a fixed theory $p(\text{data}|\text{theory})$ is a meaningful probability because it is possible to measure different datasets in repeated experiments.

The other interpretation of 'probability' is the Bayesian interpretation, where 'probability'

is interpreted as 'degree of belief'. Thus, the probability that a specific $theory_i$ is true can be anything. The probability-vector $P(theory_i)$ is called prior. In an experiment, the prior is given a priori and reflects the 'degree of believe' that a certain theory is true. From the following equations:

$$P(data \cup theory) = P(data|theory) \cdot P(theory) = P(theory|data) \cdot P(data) \quad (5.2)$$

$$P(data) = \sum_i P(data|theory_i) \cdot P(theory_i) \quad (5.3)$$

follows the Bayes Theorem with which the posterior $P(theory|data)$ can be calculated:

$$P(theory|data) = \frac{P(data|theory) \cdot P(theory)}{\sum_i P(data|theory_i) \cdot P(theory_i)}. \quad (5.4)$$

The posterior represents the degree of belief of theories updated by the measurement.

The critical point of the Bayes Theorem is, that it needs a prior. The prior can be chosen subjective, what would also make the posterior subjective. This makes it difficult to share a result between multiple researchers with different priors and the prior might also change over time.

As alternative, the prior can be taken from the posterior of a previous measurement. As a consequence, due to the use of the prior all the results of the measurements depend on each other. Thus, if these results are combined it is difficult to ensure that no information included in priors is double counted. Also, a mistake in a previous measurement would result in aftereffects in all the following measurements which uses some prior information from the suspicious one. Additionally, the first ever made measurement has still to assume some prior which is not motivated from a previous measurement, but rather from subjective considerations. This alternative is better than a completely subjective prior in so far, that the influence of the initial prior gets smaller for every applied measurement.

It can be summarized that Bayesian statements have several drawbacks because of the used priors. That is why it is tried to make statements independent from priors, e.g. frequentist statements.

Due to the general definition of the Bayesian interpretation as 'degree of belief', the frequentist interpretation is the subset of the Bayesian, where 'degree of believe' equals the percentage of times one would measure something if the experiment were repeated infinite times. To better distinguish frequentist and Bayesian interpretation, a statement is only called Bayesian if it is not frequentist in this work - for example if some prior information is used.

To make statements, whether to keep or discard a theory different statistical methods are used, from which one is analyzed in this work.

The P-value is analyzed which describes how probable it is to measure the measured distribution or one which yields more extrem values in the considered variables given a theory prediction. It only needs the theory prediction, one measurement and the uncertainties.

An introduction to other statistical topics can be found in [21].

5.2 The P-value

To calculate the P-value a measurement (data) and a theory prediction including its uncertainties are needed. The P-value is the probability to measure the observed data or something more extreme given the theory. To quantify how extreme a measurement n is a test statistic $Q(n)$ is used. (e.g. the negated probability to measure n could be used as test statistic) Let $Q_{obs} := Q(n_{obs})$ be the observed value of the test statistic. The P-value is defined as:

$$\text{P-value} := P(Q \geq Q_{obs}) = \int_{Q_{obs}}^{\infty} \frac{dP}{dQ} dQ \quad (5.5)$$

One can estimate the P-value by throwing N toys around the theory prediction according to its uncertainties. For each toy_n the value of the test statistic is calculated $Q(toy_n)$. At the end, one data-distribution is observed with its corresponding test statistic value Q_{obs} . The estimated P-value is then given by:

$$\hat{P}(Q \geq Q_{obs}) = \frac{\text{number of toys with } Q(toy_n) \geq Q_{obs}}{\text{total number of toys}} \quad (5.6)$$

Since only a limited number of toys can be generated, the estimated P-value varies around the true P-value. The toys are binomial distributed into values above and below Q_{obs} . The variance of the number of toys with $Q \geq Q_{obs}$ is then given by:

$$v = N \cdot \text{P-value} \cdot (1 - \text{P-value}). \quad (5.7)$$

The standard deviation of the estimated P-value is consequently:

$$\sigma_{\text{P-value}} = \frac{\sqrt{v}}{N} = \sqrt{\frac{\text{P-value} \cdot (1 - \text{P-value})}{N}}. \quad (5.8)$$

The determined P-value is used to reject or to keep a certain theory. The more strict statement is to discard a theory. That is why it is valid to overestimate the P-value which could result in a kept theory which would have been rejected in a exact P-value calculation. This is called conservative estimate. Underestimating a P-value could result in a falsely rejected theory what is avoided. Thus, if the P-value can not be calculated exactly a conservative estimate is used.

5.2.1 The Likelihood and χ^2 Function

The probability to measure a dataset n given a theory prediction ν with uncertainty σ is described by the likelihood function $\mathcal{L}(n; \nu, \sigma)$. For a binned distribution, where the theory predicts ν_i events in bin i and with Gaussian distributed uncertainties σ_i the likelihood function becomes:

$$\mathcal{L}(n; \nu, \sigma) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \frac{-(n_i - \nu_i)^2}{2\sigma_i^2} \quad (5.9)$$

This likelihood could be used as test statistic in P-value calculations. Since a natural choice for more extreme is less probable, $-\mathcal{L}$ is considered as test statistic and not $\mathcal{L}$. The likelihood function from Equation 5.9 can be split into:

$$\mathcal{L}(n; \nu, \sigma) = e^{-\frac{\chi^2(n; \nu, \sigma)}{2}} \cdot C(\sigma) \quad (5.10)$$

with

$$C(\sigma) := \frac{1}{(2\pi)^{\frac{N}{2}}} \cdot \prod_i \frac{1}{\sqrt{\sigma_i^2}} \quad (5.11)$$

$$\chi^2(n; \nu, \sigma) := \sum_i \frac{(n_i - \nu_i)^2}{\sigma_i^2} \quad (5.12)$$

In P-value calculations with only Gaussian distributed variables one gets the same results if one uses the negated likelihood as test statistic as if the chi-squared is used. $\mathcal{L}$ is strictly monotonically decreasing with rising χ^2 (see Equation 5.10) and therefore $-\mathcal{L}$ and χ^2 have the same ordering relation between different measurements n . Since for the P-value calculation only the ordering is important, χ^2 can be used as test statistic instead of $-\mathcal{L}$. So,

$$P(-\mathcal{L} \geq -\mathcal{L}_{obs}) = P(\chi^2 \geq \chi_{obs}^2) \quad (5.13)$$

Since a natural choice for more extreme is less probable $-\mathcal{L}$ is considered as test statistic and not $\mathcal{L}$.

In [22] the following generalization is presented which allows the implementation of correlations between the bins through the covariance matrix V of the measurement n .

$$\chi_{corr}^2 = (n - \nu) \cdot V^{-1} \cdot (n - \nu) \quad (5.14)$$

One can show, that the test statistic χ_{corr}^2 has still the same ordering relation as the negation of the corresponding likelihood function (see e.g. [21]).

For the P-value calculation the expected distribution of the χ_{corr}^2 -values is needed. One can show that (see e.g. [21]):

Theorem 5.1:

Let n be a histogram of random variables with b bins whose entries are Gaussian distributed with covariance V and mean ν . Then, the distribution of χ_{corr}^2 -values depends only on b , and can be analytical described by the χ^2 -distribution for b degrees of freedom.

This enables the possibility to calculate the P-value without computational time expensive toy-study if n is Gaussian distributed with Covariance V and mean ν .

If only Gaussian distributed statistical uncertainties (i.e. more than 5 events per bin and $\sigma_i^2 = \nu_i$) are considered, then the chi-squared function $\chi^2(n; \nu, \sigma)$ is called Pearson's chi-squared $\chi_P^2(n; \nu)$:

$$\chi_P^2(n; \nu) := \chi^2(n; \nu, \sqrt{\nu}) = \sum_i \frac{(n_i - \nu_i)^2}{\nu_i}. \quad (5.15)$$

$$(5.16)$$

One can also state in this case, that n is distributed according to the covariance matrix $V_{P,ij} := \delta_{ij}\nu_i$. So, Theorem 5.1 is applicable saying that $\chi_P^2(n; \nu)$ is χ^2 -distributed.

In some cases, the variance of n as covariance matrix V is not used for the χ^2 calculation, but rather an estimator $\hat{V}$. The χ^2 where the statistical uncertainties estimated from data are used is called Neyman chi-squared. It is defined as:

$$\chi_N^2(n; \nu) := \chi^2(n; \nu, \sqrt{n}) = \sum_i \frac{(n_i - \nu_i)^2}{n_i} \quad (5.17)$$

Since n is no longer distributed according to the used Covariance matrix $\hat{V}$ in the χ_{corr}^2 calculation Theorem 5.1 does not apply any more. So, $\chi_N^2(n; \nu)$ does not have to be χ^2 -distributed.

5.3 Probability in Terms of Standard Deviations and Conventions

The intermediate result of a statistical test is a probability. These probabilities can get very small. To have more suitable numbers the probability is given in terms of standard deviations a measurement would be apart from the mean of a Gaussian distributed variable for the given P-value. This value is called significance. In this work both sides of the Gaussian are included in the calculation of the P-value for the conversion. In Figure 5.1 the correspondence of probability to significance in terms of standard deviation σ is shown.

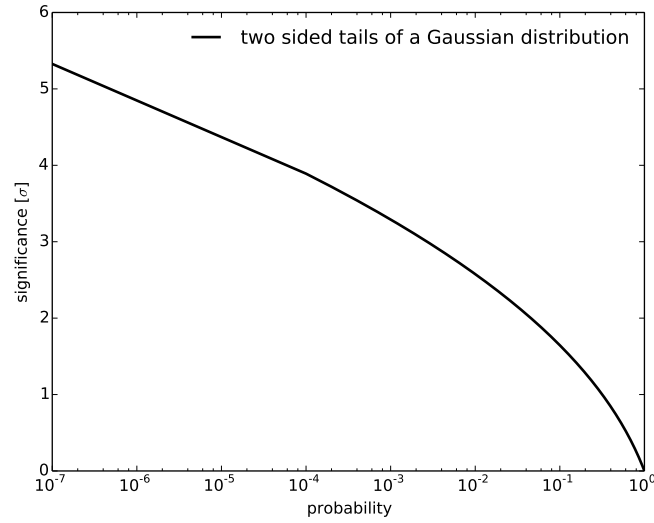


Figure 5.1: Correspondence of the significances to the probabilities for the two tails of a Gaussian distribution.

There are some predefined significances. If a theory deviates more than 2σ , then it is usually rejected. One speaks of an evidence if 3σ are reached for the rejection of the background only hypothesis and of a discovery if 5σ are reached.

6 Unfolding

In high energy physics (HEP) binned distributions like $M_T(WZ)$ shown in Figure 2.2 on page 9 are analyzed to study differential cross sections. With the histogram, the phase space is divided into several phase space regions represented by bins.

The detector is not perfect. It has only limited coverage, resolution, reconstruction and identification power. Because of these imperfections, the measured distribution is not distributed like the unknown underlying true distribution, but is distorted (see Figure 6.1). Consequently, it is distinguished between the underlying truth distribution which a perfect detector would measure and the measured reconstructed distribution. Distributions and events which one would get with a perfect detector are called on 'truth level'. In terms of the MC generator semantic this corresponds to on 'hadron level'. Distributions and events which include detector effects are called on 'reco level'. In terms of the MC generator semantic this corresponds to on 'detector level'.

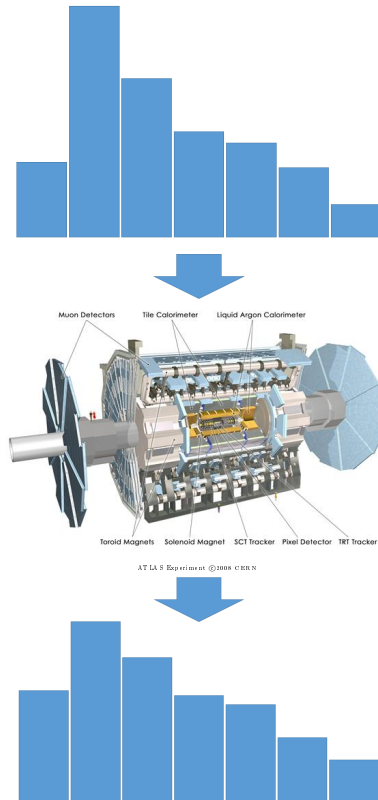


Figure 6.1: Schematic overview of the detector effects. The distributions are getting distorted due to the detector effects.

The detector effects can be modeled as a folding of the truth distribution μ . The general formula including background is:

$$\nu = \int_{-\infty}^{\infty} R(\nu, \mu) \mu \, d\mu + \beta. \quad (6.1)$$

With binned distributions, this becomes:

$$\nu_i = \sum_j R_{ij} \mu_j + \beta_i \quad (6.2)$$

where ν_i is the i -th bin of the theory prediction on reco level, μ_j is the j -th bin of the theory prediction on truth level, R_{ij} is the (i, j) -element of the response matrix and β_i is the i -th bin of the background.

The background term is included because one is in general interested in a special production process, called the signal process, which is identifiable by its final state. Since there are also other processes which produce the same final state the expected number of events in a certain phase space is increased.

The response matrix defines how the bin entries on truth level are connected with the ones on reco level. A perfect detector would have a unity-matrix as response matrix whereas a realistic one differs because of the detector effects. The response matrix depends on the variable being unfolded and the binning.

In a concrete measurement, the distribution n is observed on reco level. By undoing the folding it is possible to estimate the distribution $\hat{\mu}$ on truth level from the observed data:

$$n_j = \sum_j R_{ij} \hat{\mu}_j + \beta_i. \quad (6.3)$$

Only events which pass the selection criteria on reco level are part of the distribution n on reco level. For the comparison of unfolded distributions with theory predictions also well defined selection criteria on truth level are needed. These selection criteria on truth level are in general as close as possible to those on reco level in order to not loose information and to not extrapolate something into the data in the unfolding process. Not every event which passes the selection on reco level also passes the corresponding selection on truth level and vice versa. This is also corrected for in the response matrix.

In order to disentangle the different effects included in the response matrix it is expanded as a product of efficiency correction *effCorr*, inverse fiducial correction *fidCorr* and migration matrix M (as it is done in the EWUnfolding framework¹ [23]):

¹In the EWUnfolding framework the migration matrix is also called response matrix. This convention is not used here because it makes the semantic differentiation with the total response matrix difficult.

$$R_{ij} = \frac{1}{fidCorr_i} \cdot M_{ij} \cdot effCorr_j \quad (6.4)$$

where $effCorr_j$ is the probability that an event which passes the selection on truth level also passes it on reco level, M_{ij} is the migration between the i -th bin on reco level and the j -th bin on truth level and $fidCorr_i$ is the probability that a signal event which passes the selection on reco level also passes it on truth level.² A visual representation is shown in Figure 6.2.

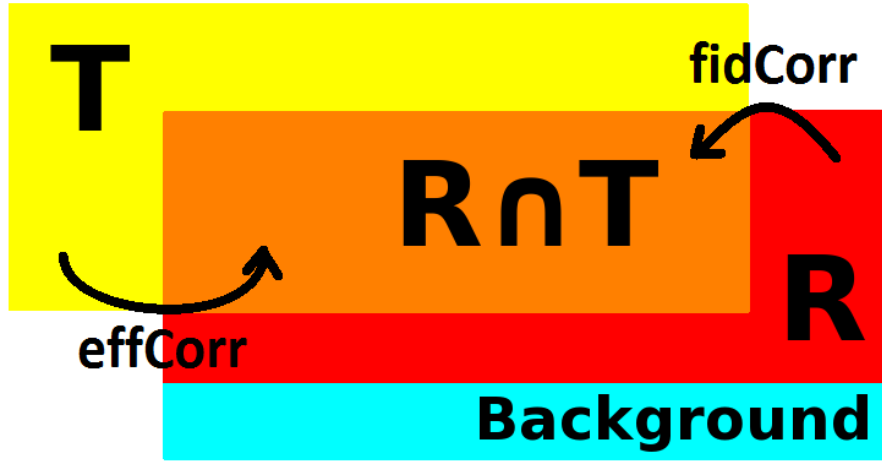


Figure 6.2: Different selection levels of events. **R** means an event passes the selection on reco level (reco-selection), **T** means an event passes the selection on truth level (truth-selection). If an event passes both selection criteria it lies in the joint phase space $\mathbf{R} \cap \mathbf{T}$. The background accounts for the events which also pass the selection on reco level, but come from another than the signal process.

Simulated signal (s. s.) and background events are used to estimate the constituents of the response matrix and the background on reco level. The events are simulated on truth level according to the Standard Model prediction. Then a full detector simulation is applied from which each event gets its properties on reco level.

Then the fiducial correction can be estimated by:

$$fidCorr_i = \frac{\text{number of s. s. events in reco-bin } i \text{ passing the reco- and truth-selection}}{\text{number of s. s. events in reco-bin } i \text{ passing the reco-selection}}. \quad (6.5)$$

The efficiency correction can be estimated by:

$$effCorr_j = \frac{\text{number of s. s. events in truth-bin } j \text{ passing the reco- and truth-selection}}{\text{number of s. s. events in truth-bin } j \text{ passing the truth-selection}} \quad (6.6)$$

²In some other unfolding setups the effect of the fiducial correction is included in the background term β .

and the migration matrix can be estimated by:

$$M_{ij} = \frac{\text{number of s. s. events in both reco-bin } i \text{ and truth-bin } j}{\text{number of s. s. events in truth-bin } j \text{ passing the reco- and truth-selection}}. \quad (6.7)$$

In Figure 6.3 the values are shown for the example of the M_T^{WZ} distribution simulated with Powheg.

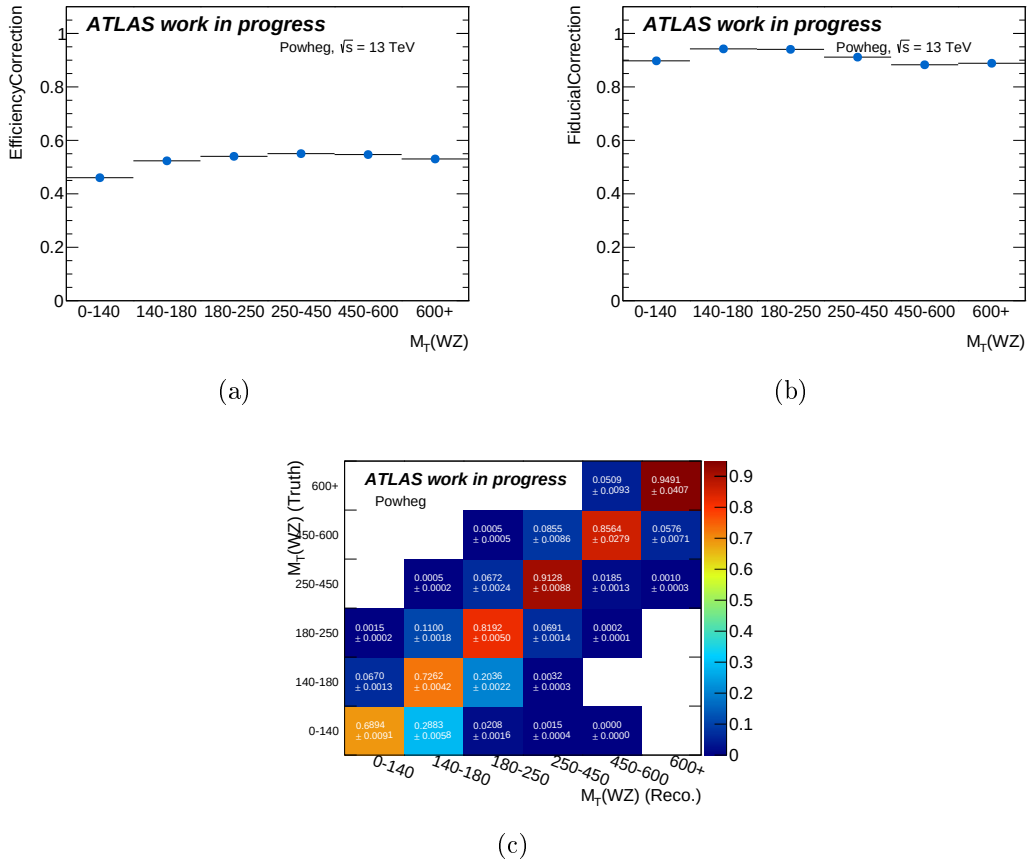


Figure 6.3: Input for unfolding $M_T(WZ)$ generated with Powheg. The efficiency correction (a), fiducial correction (b) and migration matrix (c) are shown.

At the end one wants to compare theory predictions with the experimental data or experiments with experiments (for instance results from CMS and ATLAS). In the case of the comparison between experiments one has to unfold both measured distributions and can only compare on truth level because the events are shifted with different detector effects to reco level. In the case of the theory - experiment comparison there are two possible approaches:

1. One option is to fold the theory prediction and compare 'folded theory prediction' and 'measurement' on reco level. There are three different approaches how to perform the folding of the theory prediction.

- a) For the folding the most exact approach is to use a full detector simulation called **the folding approach with full detector simulation**.
 - b) However, it is also possible to use a simplified detector simulation, for instance [24], called **folding approach with simplified detector simulation**.
 - c) Also applying the response matrix derived from the SM simulation is possible called **the folding approach with response matrix**.
2. Another option is to unfold the measurement and compare the 'theory prediction' and the 'unfolded measurement' on truth level. This approach can be further subdivided by the information used for statistical tests with unfolded distributions.
- a) If only the unfolded distribution, a theory prediction and their uncertainties are used for statistical tests it is called **the fast unfolding approach**.
 - b) If in the statistical test the response matrix is also used, the method is called **the extended unfolding approach**.

In all approaches, there are typically two person groups involved in comparing the predictions with experimental results. On one side, there are the experimentalists retrieving the data and preparing it for statistical analyses. On the other side, there are the theorists who want to compare their theories with experimental data. Typically, one dataset is compared to many different theories. So, it is worth the effort to prepare the data as far as possible to minimize the effort of the theorists. This is the motivation for the fast unfolding approach where theorists do not need any knowledge about the detector and can compare fast on truth level. However, there are some drawbacks regarding the calculation of the statistical uncertainties, which are discussed in Section 6.2. Furthermore, all approaches using the response matrix derived from the SM have additional systematic uncertainties due to the assumption of the SM. This is discussed in Section 6.2 as well.

In the following work, the unfolding approaches are studied. In discussions, they are compared to folding approaches.

6.1 Unfolding Methods

In this chapter, different unfolding methods are presented and discussed. A good unfolding method makes the statistical tests on truth level easy for theorists. That means that they can perform it fast without losing significance compared to other methods, e.g. through conservative estimated uncertainties. In order to make meaningful statistical statements possible the unfolding results should have a similar ordering relations on truth level and on reco level with respect to test statistics like the Likelihood.

It is often stated, that the unfolding method should also provide a good estimate of the theory

distribution $\hat{\mu}$ on truth level. A good estimate has a small variance and is unbiased. If an estimate has the smallest possible variance for its bias it is called efficient. The bias b is defined as

$$b = E(\hat{\mu}) - \mu \quad (6.8)$$

where $E(\hat{\mu})$ is the expectation value:

$$E(\hat{\mu}) = \int \hat{\mu} \cdot \mathcal{L}(\hat{\mu}; \mu) d\hat{\mu}. \quad (6.9)$$

Since the bias depends on the theory μ it is either 0 or has to be estimated or calculated with the knowledge of the theory.

In the following different unfolding methods are presented and discussed.

6.1.1 Bin-By-Bin Unfolding

One simple way of unfolding is the Bin-By-Bin unfolding. It does not correct for migrations between bins and just scales the bins according to the Monte Carlo simulation from reco- to truth level. So, it corrects only for lost or additional events per bin.

The calculation is very simple, by just multiplying a correction factor C to each bin.

$$\hat{\mu}_k = C_k \cdot (n_k - \beta_k) \quad (6.10)$$

with

$$C_k = \frac{\text{number of simulated signal events in reco-bin } k \text{ passing the reco-selection}}{\text{number of simulated signal events in truth-bin } k \text{ passing the truth-selection}}. \quad (6.11)$$

This method only yields promising results if the migration matrix is nearly diagonal. Since the method does not correct for migrations it is in general very biased if migrations are present.

6.1.2 Matrix Inversion (MI)

The mathematically exact way to unfold would be to resolve the folding equation 6.3 for μ . One gets:

$$n = R\hat{\mu} + \beta \quad (6.12)$$

$$\iff \hat{\mu} = R^{-1}(n - \beta) \quad (6.13)$$

If the inverse of the response matrix exists and is unique then the unfolding method Matrix Inversion is also the solution from Maximum Likelihood (ML) estimate which is in general, but especially in this case, efficient and unbiased[21]. Since MI is the exact mathematical solution it has some nice properties in the uncertainty propagation which is discussed in Section 6.2.

Matrix Inversion is only applicable if the inverse of the response matrix exists and is unique. The response matrix can be chosen quadratic by construction. Then the question for a unique inverse boils down to the question if the migration matrix is invertible. One can show [25] that a matrix is always invertible if it is strictly diagonally dominant. In the current case this means, that if the main diagonal of the migration matrix is larger than 50% then the matrix is invertible. Since a main diagonal of the migration matrix larger than 50% is also for other reasons (see Section 6.2) preferable and is chosen accordingly, it should be in general possible to use Matrix Inversion.

Nevertheless, there are reasons why Matrix Inversion is avoided in particle physics and other applications like image reconstruction. In general, the response matrix smears the distribution on truth level to a smoother distribution on reco level. This way, large bin-to-bin fluctuations on truth level lead to smaller bin-to-bin fluctuations on reco level.

For a response matrix with large smearing (see Figure 6.4 (b)), the information in the data on reco level (see Figure 6.4 (a)) only tells approximately which bin it comes from on truth level. This becomes problematic when the unfolding is done with Matrix Inversion. In this case, small fluctuations on reco level get blown up to larger fluctuations on truth level. In Figure 6.4 (c) and (d) this effect is shown for a generated toy for the calculation of the statistical uncertainties on truth level. This way, large bin-to-bin fluctuations are quite probable on truth level and can not be declined by the data. The unfolded result is shown in Figure 6.4 (e) with its large variance and anti correlations shown in (f). The correlation matrix C can be calculated from the covariance matrix V and is given by: $C_{ij} = \frac{V_{ij}}{\sqrt{V_{ii}V_{jj}}}$.

If one wants to publish a best guess how the distribution on truth level may look like or if one wants to compare by eye large bin-to-bin fluctuations are hindering. That is why one tries to suppress large fluctuations and make the unfolded distribution smoother with dedicated estimators which make large bin-to-bin fluctuations unlikely. The information of smoothness is not in the data and one has to introduce it ad hoc. This process is called regularization. However, from a frequentist statistical point of view there can not be a better estimator than the ML estimator [26]. The ML solution has the smallest possible variance and zero bias.

6.1.3 Regularized Unfolding

The regularization can be done in multiple ways. At the moment, the method most used in particle physics is Bayesian iterative unfolding. There, unsmooth distributions are avoided by a recursive algorithm which starts with a prior distribution and converges towards the MI solution. It is stopped after a fixed number of iterations. Unsmooth MI solutions need

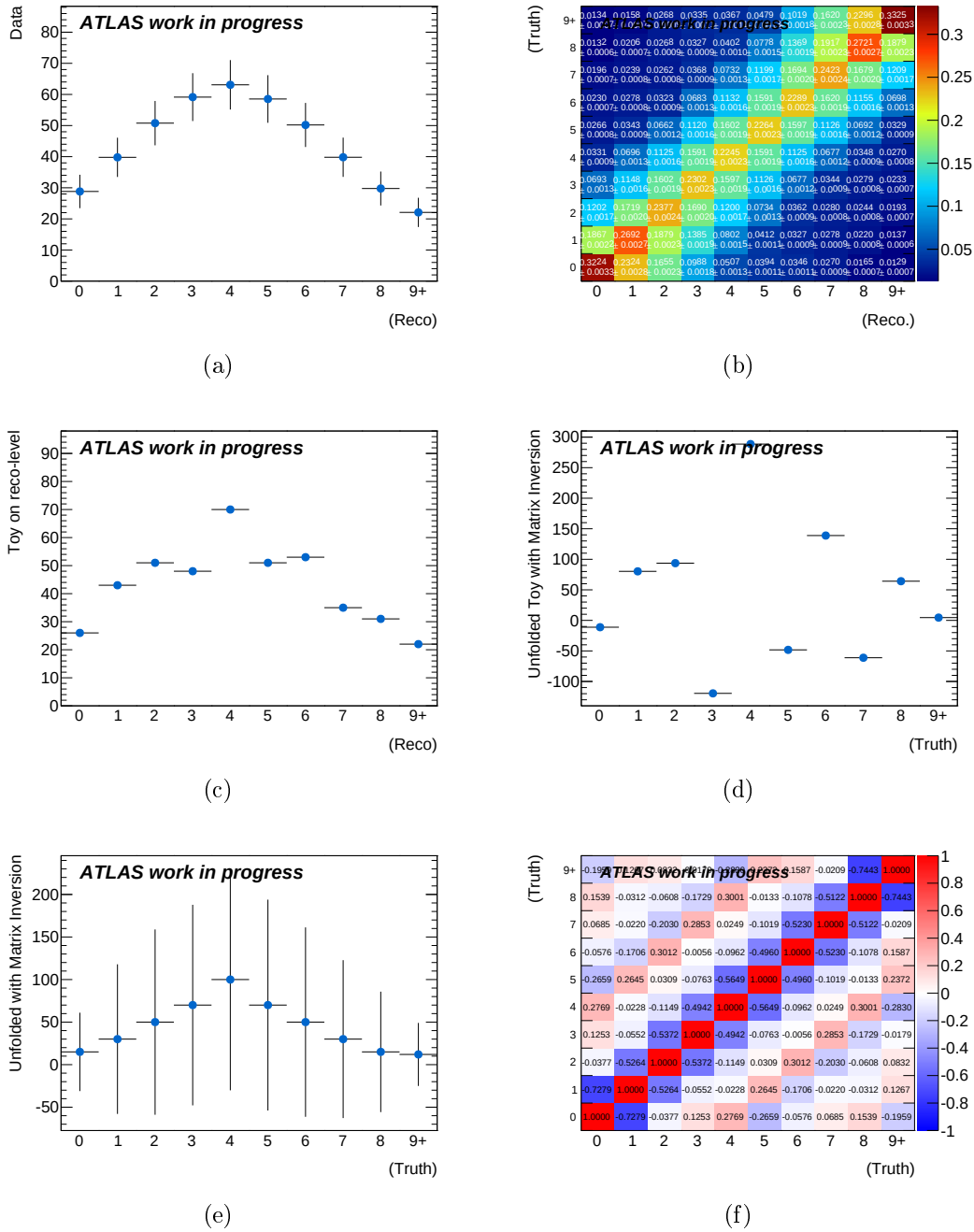


Figure 6.4: Example of large anti correlations in the unfolding result derived with Matrix Inversion. (a) the used Data distribution on reco level, (b) the used response matrix, (c) a toy on reco level with some bin-to-bin fluctuations, (d) unfolding result of the same toy, (e) unfolded data distribution, (f) correlation matrix of the unfolded data distribution

longer to converge than smooth MI solutions. Thus, since it is stopped after a fixed number of iterations these unfolded distributions are smoothed compared to MI solutions. The number of iterations and the initial prior are parameters to adjust the strength of regularization.

Also for the ML approach a Likelihood term can be inserted to make smooth distributions more likely (implemented e.g. in TUnfold [27]). Inside that term is a parameter to tune the

strength of the regularization. Such a likelihood term is also included in other regularized unfolding methods as for instance SVD [28] or IDS [29]. In the SVD method smooth functions are not only considered as more likely but information about bin-to-bin fluctuations are also explicitly ignored in order to get a smoother result.

There are multiple possible ways to interpret regularization.

One could interpret regularization in a Bayesian way. In this case, regularization would mean to include some prior knowledge into the unfolded result. The unfolded result would be interpreted as best guess how the distribution may look like on truth level.

It has the known drawbacks of Bayesian methods: It makes combinations difficult and leads to cascading problems if mistakes are found in an old measurement. Additionally, the choice of the prior is subjective. It has to be noted, that in a Bayesian approach there is no perfect degree of regularization, because it just represents the 'degree of belief' how smooth the unfolded distribution has to be.

In this thesis, it is interpreted in a frequentist way. This is possible by considering the regularized unfolded result as a biased estimate used to perform statistical tests. Since the ML estimator is already efficient it is only possible to make the variance smaller by including a bias into the estimate. One tries to choose the bias such, that it is small for theories one thinks are more likely - namely smooth distributions. The regularization does not increase the significance but just makes data - theory comparisons by eye easier. In a computational setup the correlations can be handled by the covariance matrix.

It is sometimes stated that one trades off statistical uncertainties against the bias by the regularization strength. This is not meant the way that one somehow has to minimize a total uncertainty. Instead one has to weigh how much statistical information should be discarded³ in order to get a better visual comparable distribution⁴.

A use case of estimators including regularization is the validation of MC generators by eye. In general, it is expected that the Monte Carlo generators predict smooth distributions. Therefore, in order to easily see differences, it is useful to smoothen the measured distribution. By doing so, the visibility of differences over several bins is enhanced.

Also in the frequentist approach there is no well defined optimal strength of regularization. The optimal strength depends on the choice of the experimentalist how smooth compatible theories have to be. So, a variety of possible regularization strength can be 'optimal' in some sense (see some optimization algorithms in [21]).

One remaining issue is the calculation of the bias. In the extended unfolding approach it is

³Statistical information decreases with rising conservative estimated bias.

⁴Visual comparability gets better with less variance and less correlations.

possible to calculate this bias for symmetric distributed uncertainties simply by folding and regularized unfolding the theory prediction. The bias is then given as the difference between the regularized unfolded theory and the initial theory prediction on truth level.

6.1.3.1 Iterative Bayesian Unfolding

The Iterative Bayesian unfolding [30] [31] method is based on Bayes Theorem. It introduces cause and effect cells in order to handle the migrations with the Bayes Theorem. The bins on truth level are interpreted as cause cells C_j and the bins on reco level as effect cells E_i . The variable $n(C_j)$ resp. $n(E_i)$ denotes the number of events in that bin on truth resp. reco level. The number of events $n(E_i)$ entering the migration matrix on reco level is given by the data distribution on reco level n_i subtracted by the background β_i multiplied by the fiducial correction

$$n(E_i) = (n_i - \beta_i) \cdot fidCorr_i. \quad (6.14)$$

The number of events $n(C_j)$ coming from the migration matrix on truth level corresponds to the estimated truth distribution $\hat{\mu}_j$ multiplied by the efficiency correction

$$n(C_j) = \hat{\mu}_j \cdot effCorr_j. \quad (6.15)$$

The aim of the unfolding process is to determine $n(C_j)$ with which $\hat{\mu}_j$ can be calculated. All these calculations happen in the joint phase space where all events have passed the selection on reco and truth level.

In the following these number of events are converted into probabilities. Therefore one defines:

$$N_{tot} := \sum_i n(E_i) = \sum_j n(C_j) \quad (6.16)$$

$$p(E_i) := \frac{n(E_i)}{N_{tot}} \quad (6.17)$$

$$p(C_j) := \frac{n(C_j)}{N_{tot}} \quad (6.18)$$

where $p(E_i)$ is the probability to find an event in the i -th effect cell, $p(C_j)$ is the probability to find an event in the j -th cause cell and N_{tot} is the total number of events in the effect cells. Since all regarded events lie in the joint phase space, N_{tot} is also the total number of events in the cause cells.

One wants to determine $p(C_j)$ with the knowledge of $p(E_i)$. The Bayes Theorem (see Equation 5.4) is used to calculate $p(C_j)$ in an iterative approach:

$$p_n(C_j|E_i) = \frac{p(E_i|C_j) \cdot p_{n-1}(C_j)}{\sum_k p(E_i|C_k) \cdot p_{n-1}(C_k)}, \quad (6.19)$$

where $p(E_i|C_j)$ is the probability to find an event in the i -th effect cell if it is in the j -th truth cell. The probability can be calculated from simulations and equals M_{ij} . The prior distribution for the first iteration $p_0(C_j)$ has to be defined by the experimentalists. A common choice is the distribution predicted from the MC sample

$$p_0(C_j) = \frac{\text{number of s. s. events in truth-bin } j \text{ passing the reco- and truth-selection}}{\text{number of s. s. events passing the reco- and truth-selection}}. \quad (6.20)$$

For the following iterations the outcome from the previous iteration is used as new prior $p_n(C_j) := \sum_i p_n(C_j|E_i) \cdot p(E_i)$. This way the prior gets updates every iteration and the dependence on the initial prior gets smaller. For infinite number of iterations the prior converges towards the MI solution.

6.1.3.2 Prior Dependence in Bayesian Unfolding

The regularization strength and thus also the bias of the estimator depends on the number of iterations and the initial prior. In order to study the dependence of the bias from the number of iterations for a given prior two different Monte Carlo generators (Powheg and Sherpa) are used. The advantage of MC data is the availability of the truth-information. In this study it is assumed that the statistical uncertainty is the only uncertainty and that it is Gaussian distributed. Since it is symmetric the bias is given by the difference of unfolded result and prediction on truth level.

In a first test the response matrix from Powheg and a prior from Sherpa are used to unfold the distribution on reco level predicted by Powheg. It is shown in Figure 6.5.

There is a small deviation between the predictions of Sherpa (dashed black line) and Powheg (solid black line) on truth level. With an increasing number of iteration, the unfolding result converges towards the Powheg prediction on truth level. Already after 3 iterations it is near the Powheg prediction, which is also the result of Matrix Inversion within the statistical uncertainties. The statistical uncertainties are also shown generated with 10.000 toys thrown around the theory prediction on reco level. The statistical uncertainties increase with a rising number of iterations because each toy converges towards its Matrix Inversion solution starting from the prior. So, for a low number of iterations all toys are relatively near to the initial prior and with rising number of iterations they diverge until they reach the spreading of the Matrix Inversion solution. In the shown example, the statistical uncertainties are already nearly spread to its maximum after 4-5 iterations.

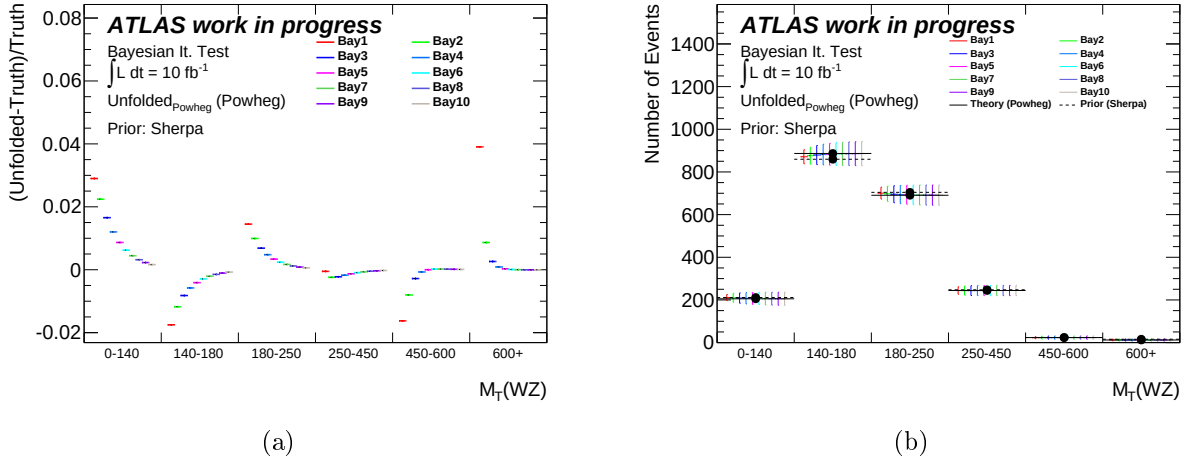


Figure 6.5: The Powheg prediction on reco level is unfolded with the response matrix derived with Powheg and the prior from Sherpa. In Subfigure (a) the bias is plotted and in Subfigure (b) distributions on truth level. Per bin there are multiple data points. The dashed black line indicates the prior, the solid black line the theory prediction from Powheg and the colored points denote the unfolded result with Bayesian unfolding after 1 to 10 iterations. The error bars represent the statistical uncertainties derived from toy studies.

A second test is performed with a flat prior instead of the Sherpa prior (see Figure 6.6), in order to see the dependences from the prior.

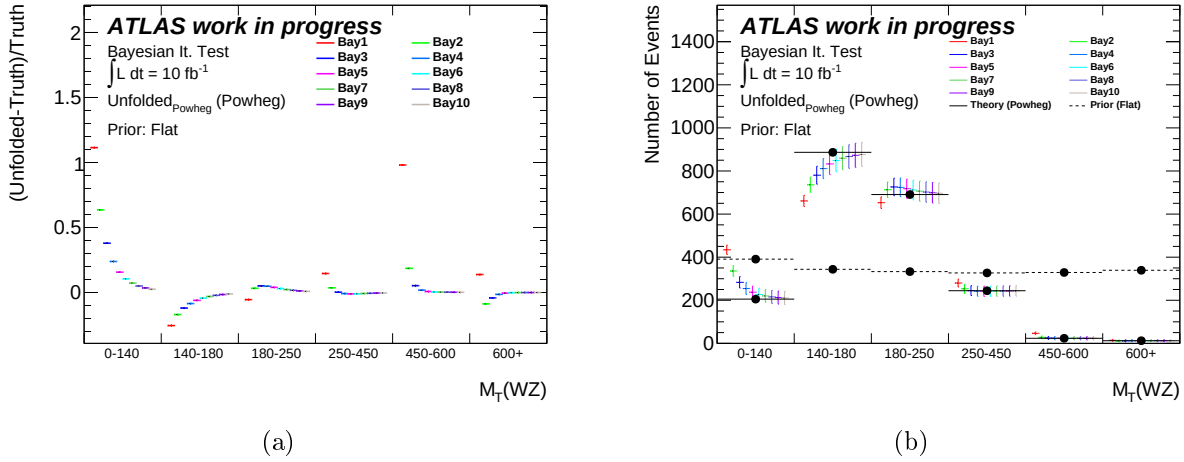


Figure 6.6: The Powheg prediction on reco level is unfolded with the response matrix derived with Powheg and a flat prior. In subfigure (a) the bias is plotted and in subfigure (b) distributions on truth level. Per bin there are multiple data points. The dashed black line indicates the prior, the solid black line the theory prediction from Powheg and the colored points denote the unfolded result with Bayesian unfolding after 1 to 10 iterations. The error bars represent the statistical uncertainties derived from toy studies.

The least informed prior in the joint phase space of reco and truth selection is the flat prior. In the phase space of the truth level selection it is slightly deformed due to the application of

the efficiency correction. The prior in the phase space of the truth level selection is shown as dashed points in Figure 6.6. Here, the bias is relatively large and 7 to 8 iterations are needed in order to make the bias small compared to the statistical uncertainties.

It can be summarized from this tests with different priors, that the optimal number of iterations depends strongly on the prior or more exactly depends on the similarity of prior and the theory one wants to compare with. This is a special problem of Bayesian iterative unfolding. In other unfolding methods one only declares how strong the regularization should be and no prior is needed. Because of this better definition, the resulting bias is just due to decreased bin-to-bin fluctuations and not due to a mismodelling of a prior. So, for comparisons by eye in a frequentist interpretation another unfolding method could give broader usable results.

In a last test, the so called technical closure test (see Figure 6.7) is performed.

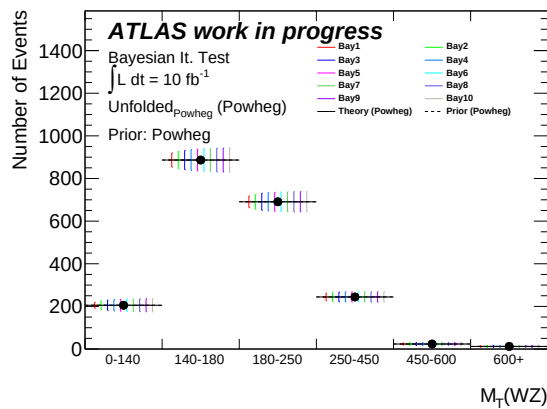


Figure 6.7: Closure test. The Powheg prediction on reco level is unfolded with the response matrix derived with Powheg and the prior from Powheg. Shown are the distributions on truth level. Per bin there are multiple data points. The dashed black line indicates the prior, the solid black line the theory prediction from Powheg and the colored points denote the unfolded result with Bayesian unfolding after 1 to 10 iterations. The error bars represent the statistical uncertainties derived from toy studies.

There, everything is used from Powheg. One expects that the unfolded data und the known truth data agree perfectly. This is used to check whether the implementation is correct. One can see perfect agreement.

6.2 Propagation of Uncertainties in the Unfolding Approaches

To calculate P-values one has to know how the unfolded data is distributed, namely its uncertainties.

In Figure 6.8 the handling of uncertainties is shown for the extended unfolding approach. The

theory prediction gets folded and unfolded in order to propagate uncertainties correctly. In the fast unfolding approach (see Figure 6.9) only the data gets unfolded that is why some uncertainties can only be estimated. In the following, all occurring uncertainties are presented and it is described how they can be handled in the different approaches. The uncertainties are propagated to the unfolded result through toy studies. Each item gets varied within its uncertainties and its effect gets propagated to truth level through unfolding. The uncertainties are usually provided as covariance matrices U on truth level.⁵ When a comparison is done by theorists they only provide their theory prediction including its uncertainties.

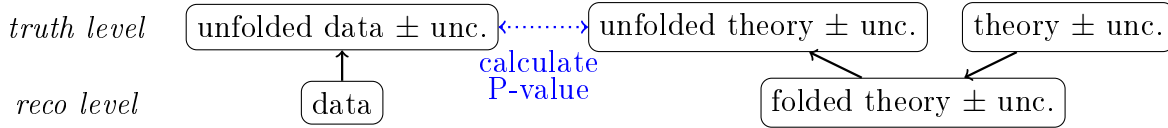


Figure 6.8: Extended unfolding approach where the theory is folded and unfolded with the help of the response matrix in order to propagate uncertainties correctly.

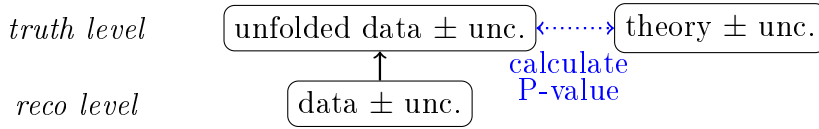


Figure 6.9: Fast unfolding approach where only data gets unfolded.

Uncertainties of the theory At first there are uncertainties of the theory prediction. This usually includes PDF and scale uncertainties. The scale uncertainties denote the uncertainties due to the limited order to which the theory is calculated in perturbation theory. If the theory prediction is generated with a MC generator, then the theory uncertainties also involve statistical uncertainties due to limited MC events.

The theory predicts to which probability one will measure an event in a certain bin on truth level. These probabilities have the above mentioned theory uncertainties. If one multiplies the response matrix followed by an unfolding with Matrix Inversion then one gets again the same initial probabilities plus uncertainties on truth level for the unfolded distributions (since $R \cdot R^{-1} = I$). So, in statistical tests on truth level one can directly apply the theory uncertainties on truth level without folding and unfolding them if Matrix Inversion is used.

If regularized unfolding is performed then the folding and unfolding does not give the same uncertainties anymore. Since the unfolding avoids large bin-to-bin fluctuations the theory uncertainties are expected to get smaller through the folding and unfolding. So, if one uses the theory uncertainties directly from truth level in the fast unfolding approach one estimates them conservatively.

⁵The covariance matrix can handle only symmetric uncertainties. There are generalizations of the covariance matrix [32], which can also include asymmetric uncertainties.

Statistical uncertainties The expected data distribution on reco level is Poisson distributed around the theory prediction on reco level due to statistical fluctuations. To include them in statistical tests on truth level they have to be propagated from reco to truth level. This can be done by throwing toys on reco level according to the statistical uncertainties and unfold them. From the toys on truth level the variance and correlations bundled in the covariance matrix can be calculated. This is only possible in the extended unfolding approach. In the fast unfolding approach, there is no information available how the theory prediction on reco level looks like. The theory prediction on reco level can only be estimated by the data distribution on reco level. The resulting differences are analyzed in Chapter 7.

Statistical and systematic uncertainties of the response matrix In the unfolding approaches, MC events are used to estimate the response matrix. These MC events have statistical uncertainties due to the limited number of generated events. Additionally, the response matrix has some systematic uncertainties like the uncertainty on the reconstruction efficiency. These uncertainties can be propagated to the unfolded result by a toy study. One could generate multiple toy response matrices according to the uncertainties and unfold the data with each of them. Finally, the uncertainties of the MC sample can be expressed by the covariance matrix of the unfolded distributions. In both unfolding approaches this uncertainty has to be calculated for the unfolded data distribution.

In the extended unfolding approach one could also calculate this for the unfolded theory prediction. For each toy one would expect that the toy generated response matrix is the correct one and would therefor fold and unfold the theory with the same varied response matrix. The additional uncertainty is zero if Matrix Inversion is used, since also for the varied response matrix $R + \Delta R$ yields: $(R + \Delta R) \cdot (R + \Delta R)^{-1} = I$. Only for regularized unfolding there is a difference. Further studies have to show how large that effect is.

Background uncertainties There are also statistical and systematic uncertainties on the background estimation. They can also be propagated by varying them in the unfolding process to the data distribution on truth-level.

The theory on truth level gets no uncertainty due to uncertainties of the background estimate for all unfolding approaches and methods. By adding and subtracting the background β to the folded theory prediction nothing changes for varied background estimates since $+(\beta + \Delta\beta) - (\beta + \Delta\beta) = 0$.

Unfolding uncertainties due to regularization If one uses biased estimators, like one does if one uses regularization, then the bias is an additional systematic uncertainty in the fast unfolding approach.

Since the bias depends on both the theory prediction and the unfolding ingredients like the response matrix it is not possible to calculate the bias in the fast unfolding approach. Usually

it is estimated by a smooth toy-theory near the data. The bias is then quoted as the bias this concrete theory would have. This method is sometimes conservative, but could also underestimate the actual bias which leads to wrong significances. That is why one uses estimators with very small expected biases, so that an underestimation would be mostly covered by other conservatively estimated uncertainties or the influence gets very small compared to other correctly approximated uncertainties - typically the statistical uncertainty.

Systematic uncertainty from estimating the response matrix with a SM simulation

Finally, all unfolding methods have one additional uncertainty coming from the fact that they use a response matrix to model the detector effects. The reason for the additional uncertainty is that the bin entries of migration matrix, efficiency correction and fiducial correction also depend on the distribution of other variables than the one under investigation, called hidden variables. E.g. if one unfolds the transverse momentum of jets the response matrix depends on the distribution of the pseudorapidity of the jets because the detector is less efficient in the forward region. In this example the pseudorapidity of the jets is a hidden variable. One has to assume a distribution for these hidden variables and usually uses the prediction of the Standard Model.

There is a bias if one compares with models deviating from the Standard Model. To calculate these uncertainties correctly one would need to apply a full detector simulation to the theory and construct a response matrix from it. The difference between the unfolded result using the response matrix of the theory and using the response matrix of the SM is the systematic uncertainty, that one is aiming to quantify. The problem is, that one wants to avoid the full detector simulation with the unfolding approaches, so this is not an option. What is usually done to estimate this uncertainty, is to create unfolding matrices with two different Monte Carlo generators. Since they perform the hadronization etc. differently there are some differences present in hidden variables. From these differences, the systematic uncertainties can be estimated.

This is a very rough estimate and an underestimation of that uncertainty is quite possible. That is why a full detector simulation or at least a simplified detector simulation is performed whenever possible. In that case, unfolding is not needed.

The expected effect of hidden variables depends on the form of the migrations and corrections. If the response matrix is a unit matrix, then also the differential response matrix of any value of a hidden variable has to be a unit matrix. Thus, if the corrections are nearly one and there are very few migrations between the bins the effect of hidden variables is expected to be small. This happens because the response matrix has only little freedom to vary depending on a hidden variable. Therefore, the binning is chosen such that the migrations are below a certain value - typical 40-20 %. The migrations get smaller by merging neighboring bins.

As an example, the effect of hidden variables was estimated by a Sherpa vs. Powheg comparison. The migration matrices of the $M_T(WZ)$ variable can be seen in Figure 6.10. The highest amount of migrations can be found in the first bin with 33%. Since the matrices are generated with a limited number of MC events, they have statistical uncertainties shown in Figure 6.10, too. The only differences between these two matrices can stem from statistical uncertainties or from hidden variables since the same detector simulation was used for their generation.

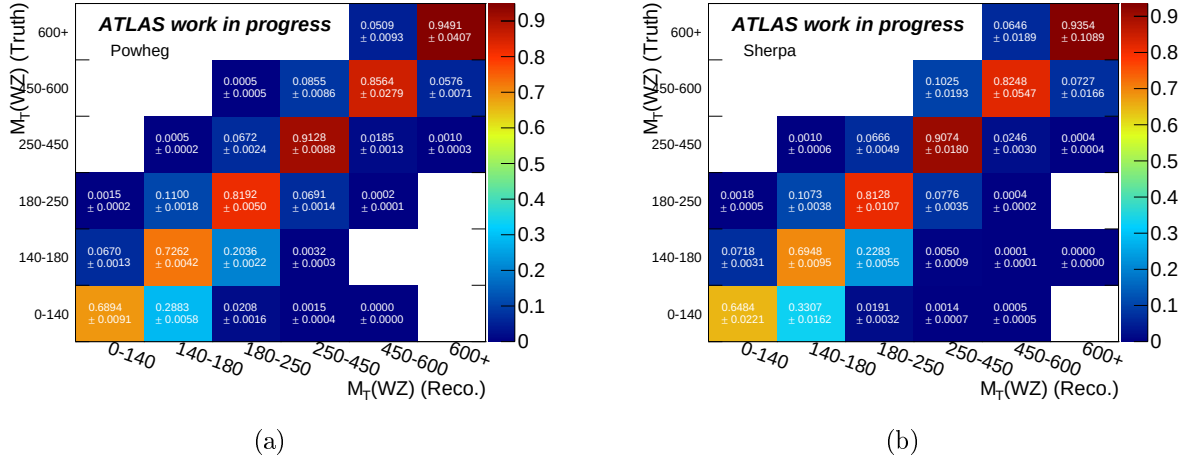


Figure 6.10: Migration Matrix of Powheg (a) and Sherpa (b) of the $M_T(WZ)$ variable.

In Figure 6.11 it is tested whether the observable differences of the two matrices are due to statistical fluctuations. Therefor, in Figure 6.11 two plots are shown that allow for a better comparison of the two matrices. Figure 6.11 (a) shows the relative difference and (b) the pull between them. The pull is defined as:

$$pull(Powheg/Sherpa) = (Powheg/Sherpa - 1)/\Delta(Powheg/Sherpa). \quad (6.21)$$

It is the probability in terms of standard deviations for such a large difference of Sherpa and Powheg to occur just because of statistical fluctuations.

One can see in Figure 6.11 (b) that bins next to the main diagonal have in general large pulls. So, even with two Monte Carlo generators describing the SM one may see effects of hidden variables in the response matrix for migrations in the order of 30%. In contrast, in Figure 6.12 and Figure 6.13 the effect is studied for the $p_T(Z)$ distribution which has migrations of less than 10%. No significant differences between Powheg and Sherpa can be found there.

So, in order to decrease the influence of other hidden variables it is better to have less bins. But with too few bins the unfolded variable becomes a hidden variable inside each bins, because inside each bin the SM distribution of that variable is assumed. Again, like for the regularization, there is no general optimal binning because it depends on the theory to which the data gets compared to. If the response matrix depends strongly on other hidden variables

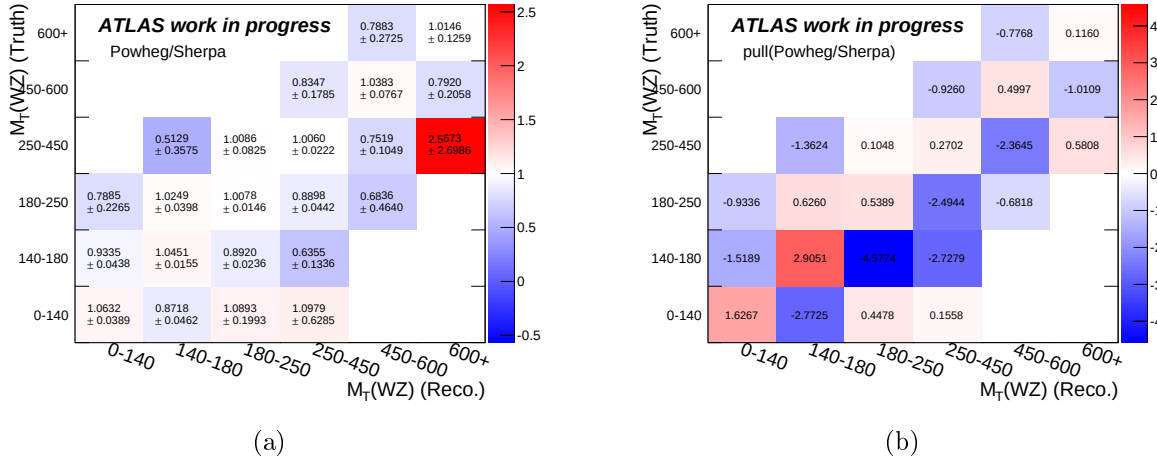


Figure 6.11: Differences between the migration matrices of Sherpa and Powheg. Subfigure (a) shows the relative difference including statistical uncertainty and subfigure (b) the pull for the statistical uncertainty.

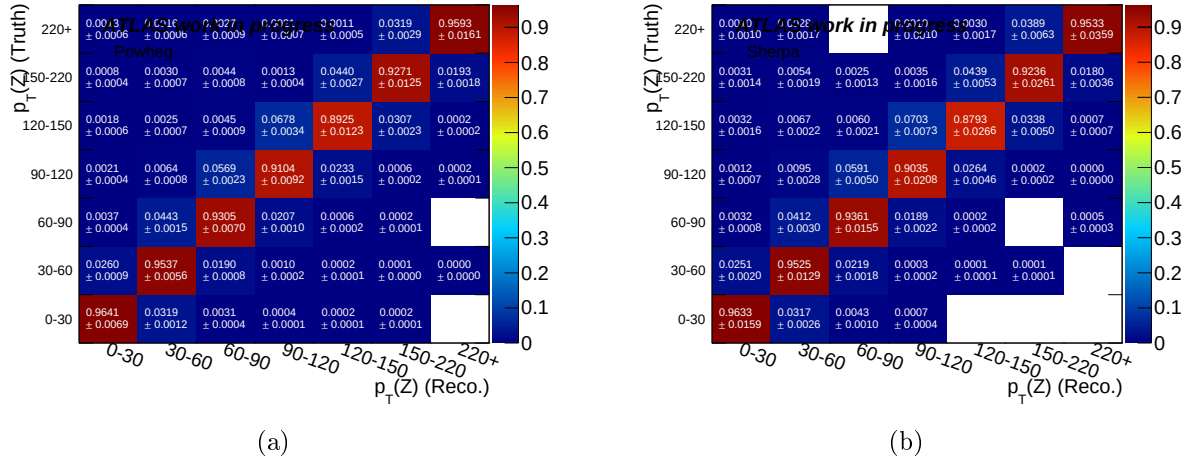
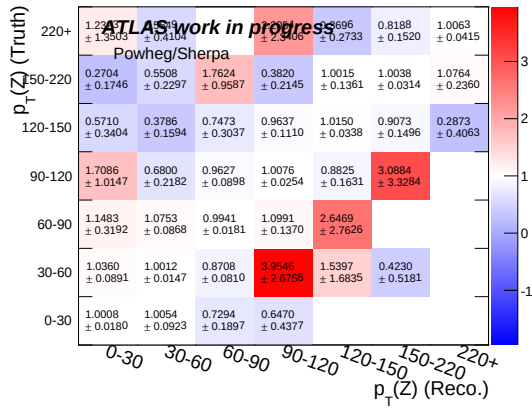


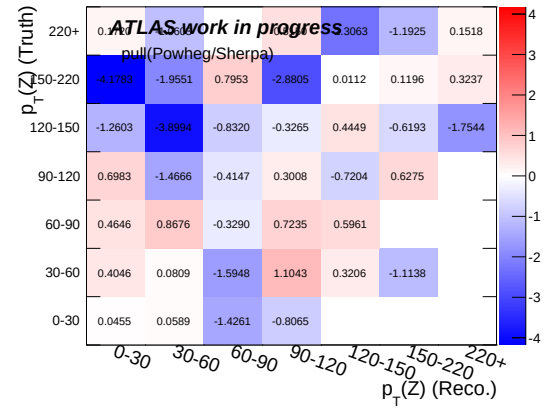
Figure 6.12: Migration Matrix of Powheg (a) and Sherpa (b) of the p_T^Z variable

it is better to have few bins, if the response matrix depends strongly on the variable under investigation it is better to have more bins.

If the response matrix depends strongly on one hidden variable multidimensional unfolding might be an option.



(a)



(b)

Figure 6.13: Differences between the migration matrices of Sherpa and Powheg. Subfigure (a) shows the relative difference including statistical uncertainty and subfigure (b) the pull for the statistical uncertainty.

7 Treatment of Statistical Uncertainties in Statistical Tests with Unfolded Distributions

In Figure 6.9 in section 6.2 it is shown how the uncertainties are handled in the fast unfolding approach. As mentioned in Chapter 6 in a fast unfolding setup the statistical uncertainties are only available as estimated from data. In this chapter the influence of using statistical uncertainties estimated from data instead of statistical uncertainties predicted from theory in comparisons of a theory with data is analyzed.

Therefor, the two ways of handling statistical uncertainties are compared. In a first approach, which corresponds to the fast unfolding approach, statistical uncertainties are estimated from data (see Figure 7.1).

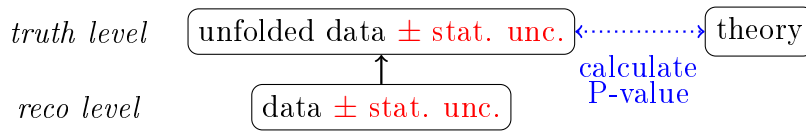


Figure 7.1: Handling of statistical uncertainties in the fast unfolding approach.

In a second approach, which corresponds to the extended unfolding approach, the statistical uncertainties are handled correctly (see Figure 7.2). The statistical uncertainties are calculated from the theory distribution which gets folded to reco level, where it gets Poisson fluctuated. By unfolding these the correct statistical uncertainties are received.

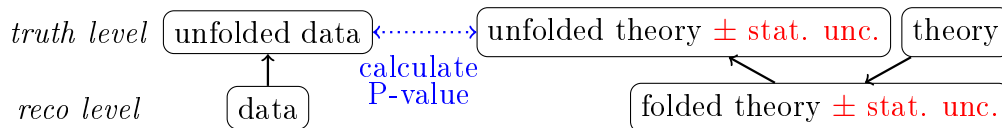


Figure 7.2: Handling of statistical uncertainties in the extended unfolding approach.

The effect of the estimated statistical uncertainties is studied on the $M_T(WZ)$ distribution, used to calculate how well the observe data can be described by different aTGC points. For the studies of the P-value one aTGC point is studied which are close to the limits set in [10].

The distribution is shown in Figure 7.3 (d) ¹.

A MC@NLO sample is used which is generated according to the SM prediction but can be reweighted to different aTGC parameters. The efficiency correction, fiducial correction and migration matrix generated according to the SM prediction are shown in Figure 7.3 (a)-(c). The fiducial correction is nearly one indicating that the data is not extrapolated into new phase spaces during the unfolding. The efficiency correction is about 50% representing the detector inefficiencies. The migration matrix is strictly diagonally dominant enabling Matrix Inversion as unfolding method.

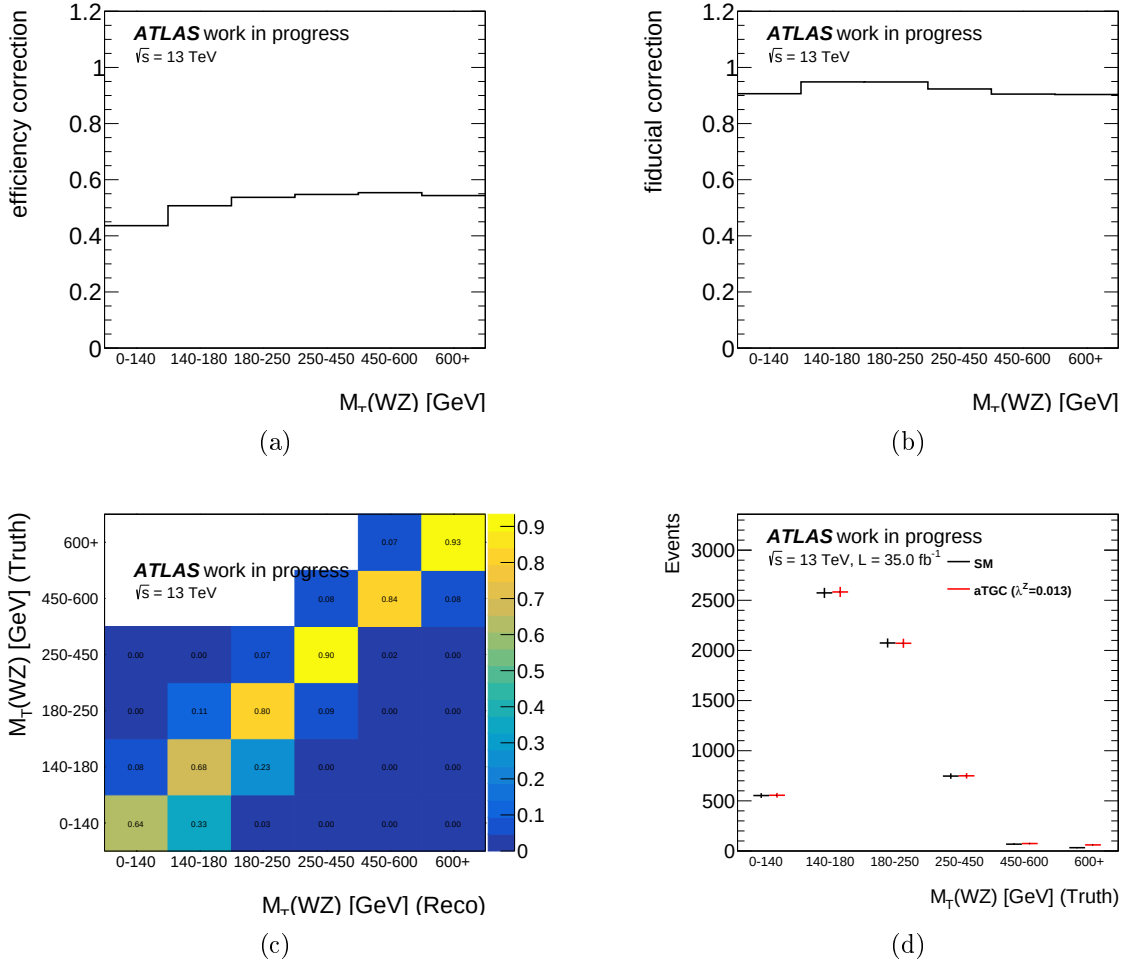


Figure 7.3: Input for unfolding $M_T(WZ)$ generated with MC@NLO. The efficiency correction (a), fiducial correction (b), migration matrix (c) and expected truth distribution for the aTGC parameter $\lambda^Z = 0.013$ in comparison to the SM prediction (d) is shown.

By this example the influence on the P-value of the estimated statistical uncertainties is studied in the following.

¹There is a normalization issue in the plots. There are less events than in [10]. But since only statistical studies are performed in this work the correct assignment to a luminosity is not critical.

7.1 Influence on the P-value

Matrix Inversion is used as unfolding method. In order to perform the P-value calculation some values need to be defined. The generalized χ_{corr}^2 function is used as test statistic. Only statistical uncertainties are applied. The theory prediction on truth level is μ and on reco level ν . The covariance matrix of statistical uncertainties on truth level is U and on reco level V . It is assumed, that the bin entries are Poisson distributed, by using

$$V_{ij} = \delta_{ij}\nu_i. \quad (7.1)$$

The estimator $\hat{V}$ is

$$\hat{V}_{ij} = \delta_{ij}n_i. \quad (7.2)$$

One test statistic, denoted with the index P, uses the covariance matrix U of the statistical uncertainties on truth level predicted by the theory.² It represents the extended unfolding approach

$$\chi_{P,corr}^2 := (\hat{\mu} - \mu) \cdot U^{-1} \cdot (\hat{\mu} - \mu). \quad (7.3)$$

The other test statistic, denoted with the index N uses the covariance matrix $\hat{U}$ of the statistical uncertainties on truth level estimated from the data distribution. It represents the fast unfolding approach

$$\chi_{N,corr}^2 := (\hat{\mu} - \mu) \cdot \hat{U}^{-1} \cdot (\hat{\mu} - \mu). \quad (7.4)$$

In the following it is shown that the $\chi_{P,corr}^2$ calculation on truth level gives the same value as a Pearson's chi-squared calculation on reco level. The same yields for $\chi_{N,corr}^2$ and the Neyman chi-squared calculation[33]. Let U be the covariance on truth level and V be the covariance matrix on reco level. It is proven in [21] for Matrix Inversion, that

$$U = R^{-1} \cdot V \cdot (R^{-1})^T. \quad (7.5)$$

This also yields for the estimate:

$$\hat{U} = R^{-1} \cdot \hat{V} \cdot (R^{-1})^T \quad (7.6)$$

Putting first Equation 7.5 and 6.2, secondly Equation 7.1 and at last Equation 5.16 in Equation

²It is later shown that it is equivalent to χ_P^2

7.3 one gets:

$$\chi_{P,corr}^2 = (\hat{\mu} - \mu) \cdot U^{-1} \cdot (\hat{\mu} - \mu) = (n - \nu) \cdot V^{-1} \cdot (n - \nu) = \sum_i \frac{(n_i - \nu_i)^2}{\nu_i} = \chi_P^2(n; \nu) \quad (7.7)$$

Analogous this is possible for the Neyman chi-squared. Putting first Equation 7.6 and 6.13, secondly Equation 7.2 and at last Equation 5.17 in Equation 7.4 one gets:

$$\chi_{N,corr}^2 = (\hat{\mu} - \mu) \cdot \hat{U}^{-1} \cdot (\hat{\mu} - \mu) = (n - \nu) \cdot \hat{V}^{-1} \cdot (n - \nu) = \sum_i \frac{(n_i - \nu_i)^2}{n_i} = \chi_N^2(n; \nu) \quad (7.8)$$

As shown in Section 5.2.1 with Theorem 5.1 χ_P^2 resp. $\chi_{P,corr}^2$ is expected to be chi-squared distributed.

For the calculation of the distribution of the $\chi_{N,corr}^2$ values theorists would need to know the statistical uncertainties V predicted by the theory. Theorists who perform the P-value calculation without the knowledge of the response matrix have to assume that the data is distributed according to some covariance. The straight forward way is to assume that it is distributed according to $\hat{V}$. In the next subsection a modified covariance is tested. If it is assumed, that the $\chi_{N,corr}^2$ values are distributed according to $\hat{V}$, then Theorem 5.1 is applicable and a chi-squared distribution is expected. But, theorists would make a mistake by assuming this because in reality the $\chi_{N,corr}^2$ values are distributed according to V . Among others the impact of that mistake is studied in this chapter.

The studies are performed on the $M_T(WZ)$ distribution for two different integrated luminosities. The two different integrated luminosities 13 fb^{-1} and 35 fb^{-1} are chosen to analyze the dependence on the number of events which are shown in Table 7.1. The 13 fb^{-1} is the integrated luminosity available at the latest publication of aTGC limits in WZ production at ATLAS [10] and the 35 fb^{-1} is the estimated full integrated luminosity of year 2016 [16] which is available at the creation time of this work. The P-value is calculated for the theory prediction with an anomalous triple gauge coupling of $\lambda^Z = 0.013$.

In Figure 7.4 the expected distributions are shown.

At first the correlation between the $\chi_{N,corr}^2$ values and the $\chi_{P,corr}^2$ values is investigated. In order to study it 25.000 toys are thrown Poisson distributed on reco level around the aTGC hypothesis ($\lambda^Z = 0.013$) according to the statistical uncertainties.³ Each toy is unfolded to

³For the correlation study it is not important that they are distributed according to the statistical uncertain-

(a)						
bin [GeV]	0-140	140-180	180-250	250-450	450-600	600+
SM ($\lambda^Z = 0$)	284	1160	1296	516	46	22
aTGC ($\lambda^Z = 0.013$)	286	1163	1296	517	51	38

(b)						
bin [GeV]	0-140	140-180	180-250	250-450	450-600	600+
SM ($\lambda^Z = 0$)	106	431	481	192	17	8
aTGC ($\lambda^Z = 0.013$)	106	432	481	192	19	14

Table 7.1: Prediction of the SM ($\lambda^Z = 0$) and an aTGC point ($\lambda^Z = 0.013$) for the number of events in the bins of the $M_T(WZ)$ distribution on reco level for 35 fb^{-1} (a) and 13 fb^{-1} (b) generated with MC@NLO.

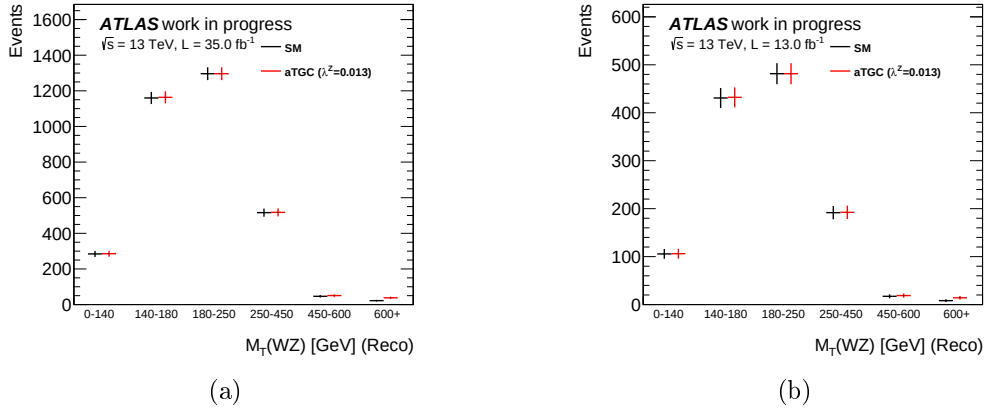


Figure 7.4: Prediction of the SM ($\lambda^Z = 0$) and an aTGC point ($\lambda^Z = 0.013$) with statistical uncertainties of the $M_T(WZ)$ distribution on reco level for 35 fb^{-1} (a) and 13 fb^{-1} (b) generated with MC@NLO.

truth level with Matrix Inversion. For the calculation of $\chi_{N,corr}^2$ the needed covariance matrix $\hat{U}$ is calculated with Formula 7.6, where $\hat{V}$ is a diagonal matrix with the number of events in each bin of the toy distribution on the main diagonal. For the calculation of $\chi_{P,corr}^2$ the needed covariance matrix U is also calculated from Formula 7.5, but with V which has the number of events of the theory prediction on its main diagonal. The resulting scattering is shown in Figure 7.5.

In Figure 7.5 it can be seen, that the toys are relatively widely spread around the main diagonal where $\chi_{P,corr}^2$ would correspond to $\chi_{N,corr}^2$. For the lower luminosity the toys are more widely spread and stripes get visible at higher $\chi_{N,corr}^2$ values. The scattering around the main diagonal is also more asymmetric for the lower luminosity.

The reason for the scattering around the main diagonal is the different definition of the test statistics. As discussed earlier they are equivalent to the Pearson's resp. Neyman chi-squared

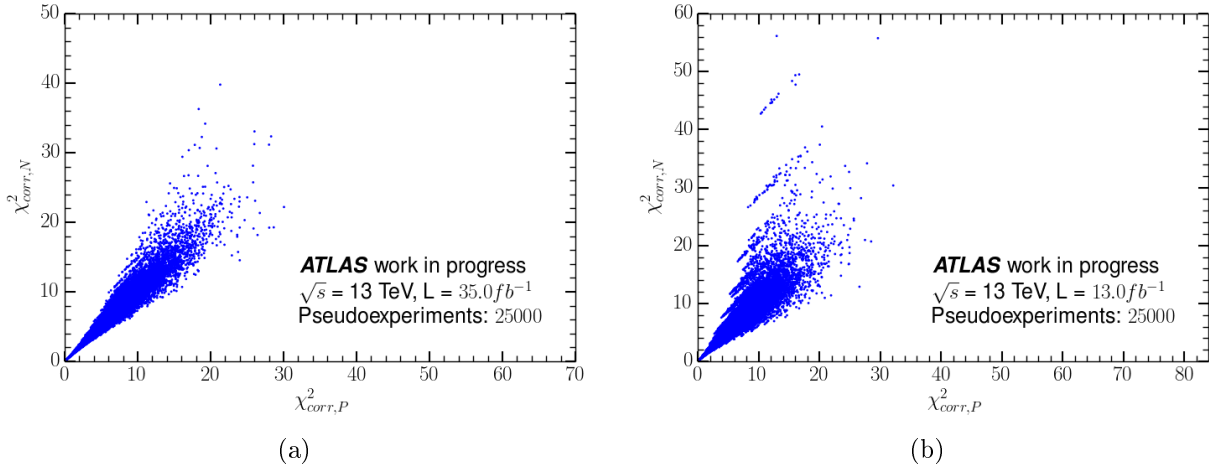


Figure 7.5: Comparison of $\chi_{P,corr}^2$ and $\chi_{N,corr}^2$ for $M_T(WZ)$ at an integrated luminosities of $13 fb^{-1}$ (a) and $35 fb^{-1}$ (b).

on reco level. If there is an under-fluctuation in one bin on reco level then this leads in the Neyman chi-squared to a smaller estimated uncertainty in that bin than in Pearson's. This leads to $\chi_N^2 > \chi_P^2$ since one summand of χ^2 is inversely proportional to σ_i . So, if under-fluctuations are dominant on reco level for a toy then this toy lies in the scattering plot above the main diagonal. Consequently, a toy with mainly over-fluctuations on reco level lies below the main diagonal.

The other observations can be understood by investigating the influence of one bin j on the difference between χ_N^2 and χ_P^2 . Let the measured data be an over- or under-fluctuation of $1\sigma_j$. Then the measured data is given as:

$$n_j = \nu_j \pm \sigma_j = \nu_j \pm \sqrt{\nu_j} \quad (7.9)$$

and the contribution of that bin to the difference between χ_N^2 and χ_P^2 can be calculated to:

$$\frac{(n_j - \nu_j)^2}{n_j} - \frac{(n_j - \nu_j)^2}{\nu_j} \quad (7.10)$$

$$= \frac{(\nu_j \pm \sqrt{\nu_j} - \nu_j)^2}{\nu_j \pm \sqrt{\nu_j}} - \frac{(\nu_j \pm \sqrt{\nu_j} - \nu_j)^2}{\nu_j} \quad (7.11)$$

$$= \frac{\nu_j}{\nu_j \pm \sqrt{\nu_j}} - 1 \quad (7.12)$$

$$= \frac{1}{\mp \sqrt{\nu_j} - 1} \quad (7.13)$$

For a high number of expected events ν_j this is approximately $\mp \nu_j^{-\frac{1}{2}}$. Thus, the scattering is expected to be symmetric around the main diagonal in that limit. The smaller the bin entry ν_j gets the larger becomes the difference between χ_N^2 and χ_P^2 resulting in a broader scattering.

This explains the broader scattering for the lower luminosity.

For a low number of expected events ν_j the '-1' in Formula 7.13 is not negligible anymore. This introduces asymmetries between over- and under-fluctuations. The difference between χ_N^2 and χ_P^2 is larger for under-fluctuations (e.g. for $\nu_j = 9$ the over-fluctuation results in a difference of $\frac{1}{-\sqrt{9-1}} = -\frac{1}{4}$ and the under-fluctuation results in a difference of $\frac{1}{+\sqrt{9-1}} = \frac{1}{2}$). Thus, a shift of the scattering towards higher χ_N^2 values is expected the lower the bin entries ν_j get. This can be seen in the plots by comparing the asymmetries of Figure 7.5 (a) and (b).

Since in the toys only integer number of events are allowed, stripes coming from the low statistic bins above the main diagonal are expected, which can be seen in Figure 7.5 (b). Thus, the stripes are due to under-fluctuations in the two low statistic bins.

To simulate the statistical uncertainties the toys are Poisson distributed on reco level. The chi-squared distribution assumes Gaussian uncertainties. Thus, these toys can be used to verify if the $\chi_{P,corr}^2$ values are really chi-squared distributed. This is done in Figure 7.6. As can be seen in the ratio plot the $\chi_{P,corr}^2$ distribution is well described by the chi-squared distribution. As there are more than 5 events in each bin this is expected.

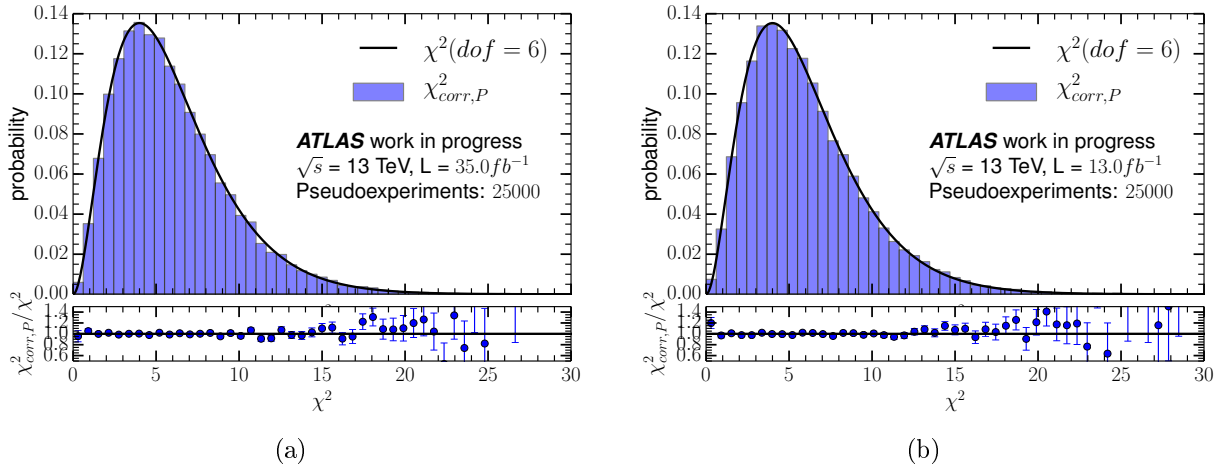


Figure 7.6: Distribution of $\chi_{P,corr}^2$ values for toys of the $M_T(WZ)$ distribution.

Also the above mentioned mistake the theorists would make can be calculated, if they would assume that toys are distributed according to $\hat{V}$. The theorists would assume, that the $\chi_{N,corr}^2$ values are chi-squared distributed, but in reality they are distributed as shown in Figure 7.7. The reason for the broader distribution of the $\chi_{N,corr}^2$ values can be seen in the scatter plot in Figure 7.5. The $\chi_{P,corr}^2$ distribution corresponds to a projection onto the x-axis. This is chi-squared distributed. The projection onto the y-axis corresponds to the shown $\chi_{N,corr}^2$ distribution. As discussed for the scatter plot the $\chi_{N,corr}^2$ values are asymmetric distributed around the main diagonal towards larger $\chi_{N,corr}^2$ values. This leads in the accumulated $\chi_{N,corr}^2$ to a broader distribution.

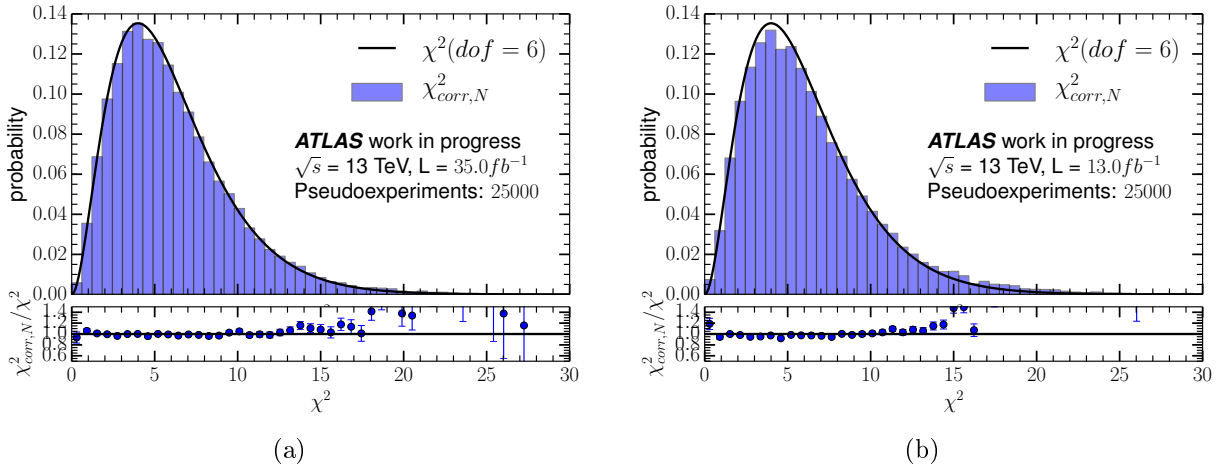


Figure 7.7: Distribution of $\chi^2_{N,corr}$ values for toys of the $M_T(WZ)$ distribution.

The P-value corresponds to the right tail of the $\chi^2_{N,corr}$ distribution. Therefore, theorists would underestimate the P-value which could lead to a false rejection of theories.

In the following it is studied how large the differences are between the P-value calculated from chi-squared distribution and calculated from the actual $\chi^2_{N,corr}$ distribution. The P-value is given in terms of standard deviations a gaussian distribution would have for that P-value, called significance.

Figure 7.8 shows the significances depending on the measured χ^2 value. It can be seen that significances calculated with the $\chi^2_{P,corr}$ distribution give the same result as the ones calculated with a chi-squared distribution within the uncertainties. The significances calculated from the $\chi^2_{N,corr}$ distribution are deviating. The deviation gets larger for higher significances.

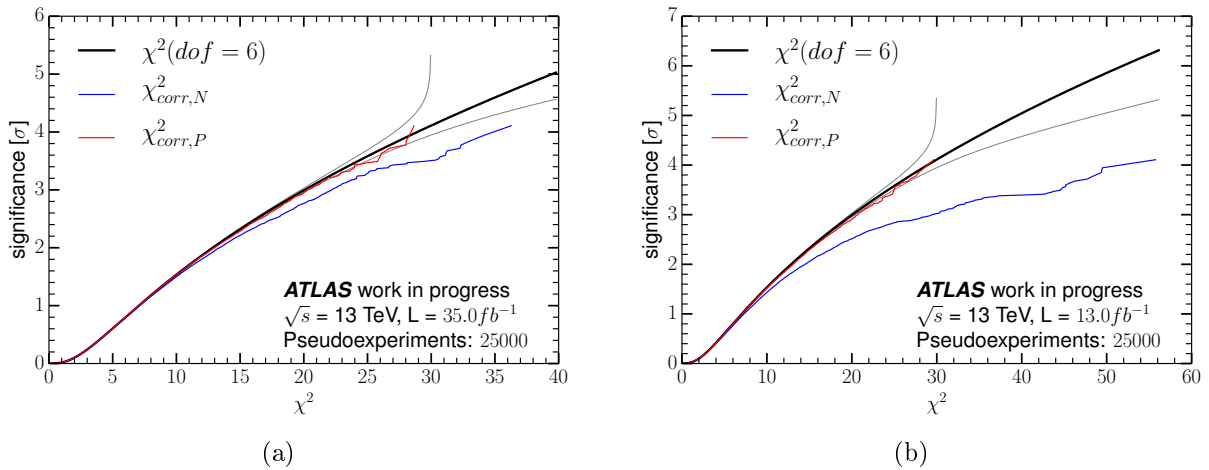


Figure 7.8: Dependence of the significance on the measured χ^2 value. The black line shows the dependence for the chi-squared distribution, the red line for the actual $\chi^2_{N,corr}$ distributed and blue line for the actual $\chi^2_{P,corr}$ distribution. The grey lines represent the expected 1σ up and down variation of the chi-squared distribution due to the limited number of toys.

As discussed earlier the $\chi_{N,corr}^2$ distribution differs through low statistic bins from a chi-squared distribution. High statistic bins are chi-squared distributed and cover partly the differences coming from the low statistic bins. For the rejection of theories the 2σ region of the significance is of interest. The theorists would discard a theory if it has a significance of 2σ or more. That is why the actual significance from the $\chi_{N,corr}^2$ distribution is especially interesting for the point where the estimated significance from chi-squared distribution reaches 2σ . In Figure 7.8 (a) this actual significance is 1.93σ and in (b) it is 1.83σ . These values are calculated by first defining the χ_0^2 value for which the chi-squared distribution reaches 2σ (for the 6 d.o.f. represented by the 6 bins of the $M_T(WZ)$ distribution: $\chi_0^2 = 12.85$). Then the actual significance is calculated with this χ_0^2 value from the actual $\chi_{N,corr}^2$ distribution.

$N = 25\,000$ toys are used to calculate the actual significance. Thus, the P-value of 0.05 corresponding to the 2σ limit can be estimated with an relative uncertainty of 3% ($=\sqrt{\frac{1-P\text{-value}}{P\text{-value}\cdot N}}$ according to Equation 5.8). This results in a relative uncertainty of 2% of the significance near 2σ .

To quantify the difference in a more general way its dependence is studied with different test distributions on reco level.

The test distributions have high and low statistic bins. The high statistic bins contain 1 000 000 events. The number of events in the low statistic bins, the number of low statistic bins and the number of high statistic bins are varied. Since the $\chi_{N,corr}^2$ distribution on truth level is identical to the χ_N^2 distribution on reco level the actual significance is directly determined on reco level.

In Figure 7.9 the actual significance is shown for zero high statistic bins depending on the number of events in the low statistic bins and on the number of low statistic bins. The actual significance is for all combinations significantly below 2σ . In Figure 7.10 also the number of high statistic bins get varied while the number of low statistic bins is fixed to one.

In Figure 7.11 the actual significance is shown for 15 events in the low statistic bins and varied number of low and high statistic bins.

There are no results for distributions with less than 3 bins, because only a few χ_N^2 -values are possible if only a few events are in the low statistic bins. Since multiple toys generate the same χ_N^2 value if only a few are possible, there are steps in the significance plot.

It can be summarized that the handling of the statistical uncertainties in the fast unfolding approach leads to two problems. Firstly, the different test statistic $\chi_{N,corr}^2$ compared to $\chi_{P,corr}^2$ leads to a different order relation and therefore to different P-values for the same observed distribution. Secondly, it is not possible to estimate the P-value resulting from the $\chi_{N,corr}^2$ distribution correctly in the fast unfolding approach. The actual significance depends on the number of events of theory on reco level. This information is needed in order to calculate the actual significance or to correct the estimated significance. Both is not possible in the fast unfolding approach since the theory prediction on reco level remains unknown.

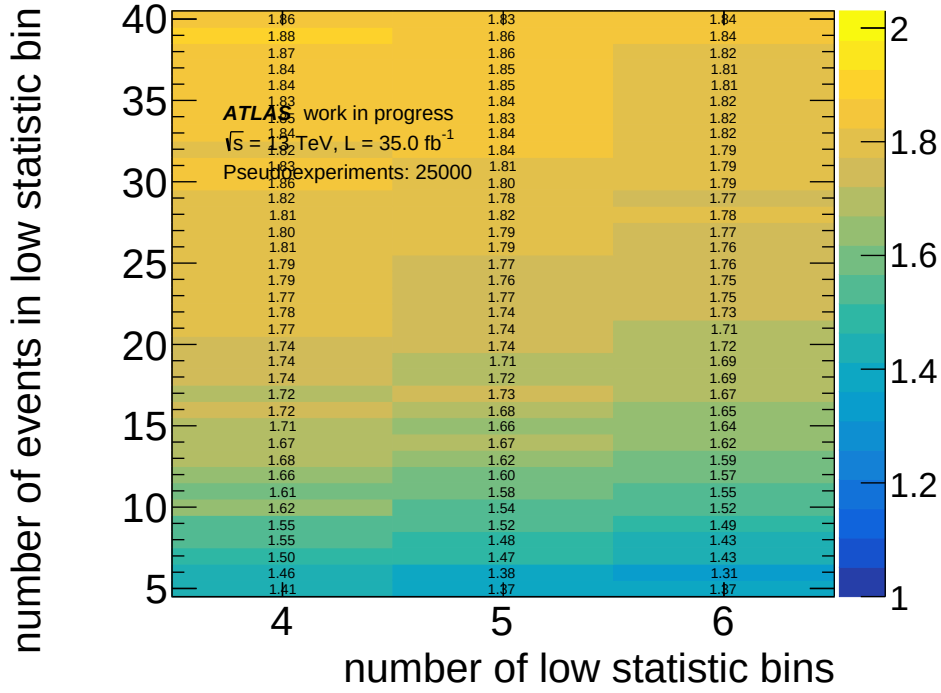


Figure 7.9: Actual significance of estimated 2σ -limit with zero high statistic bins

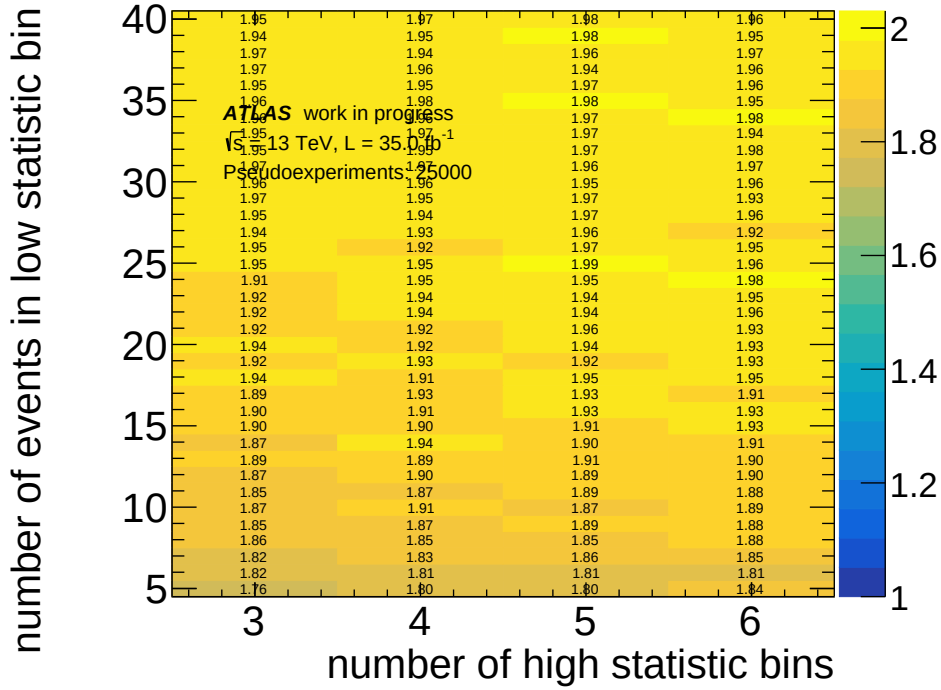


Figure 7.10: Actual significance of estimated 2σ -limit with one low statistic bin

To overcome this problem a modified test statistic $\chi_{N2,corr}^2$ is tested in the next subsection.

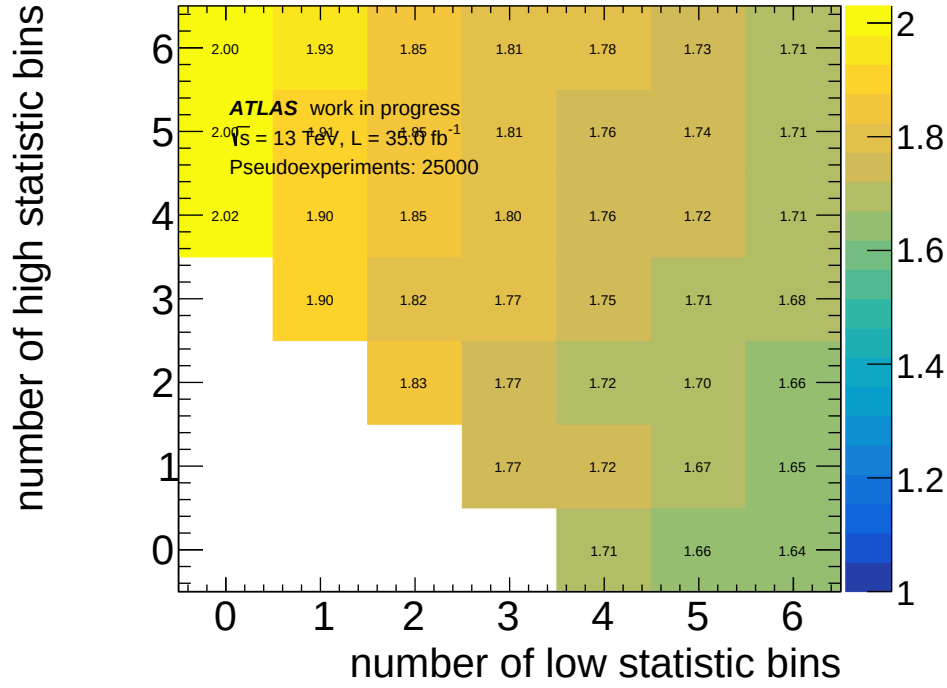


Figure 7.11: Actual significance of estimated 2 σ -limit with 15 events in low statistic bins

7.1.1 Different Test Statistic

To minimize the difference between the estimated covariance $\hat{U}_{ij}$ and the true variance U_{ij} a modified estimate $\hat{U}_{m,ij} = \sqrt{\frac{\mu_i}{\hat{\mu}_i}} \cdot \hat{U}_{ij} \cdot \sqrt{\frac{\mu_j}{\hat{\mu}_j}}$ is proposed. It is motivated from the case with a unit matrix as migration matrix. In this case the modified estimated covariance $\hat{U}_{m,ij}$ equals U_{ij} , since:

$$\hat{U}_{m,ij} = \sqrt{\frac{\mu_i}{\hat{\mu}_i}} \cdot \hat{U}_{ij} \cdot \sqrt{\frac{\mu_j}{\hat{\mu}_j}} \quad (7.14)$$

$$= \sqrt{\frac{\nu_i}{n_i}} \cdot \hat{U}_{ij} \cdot \sqrt{\frac{\nu_j}{n_j}} \quad (7.15)$$

$$= \sqrt{\frac{\nu_i}{n_i}} \cdot \frac{fidCorr_i}{effCorr_i} \hat{V}_{ij} \frac{fidCorr_j}{effCorr_j} \cdot \sqrt{\frac{\nu_j}{n_j}} \quad (7.16)$$

$$= \sqrt{\frac{\nu_i}{n_i}} \cdot \frac{fidCorr_i}{effCorr_i} n_{ij} \delta_{ij} \frac{fidCorr_i}{effCorr_i} \cdot \sqrt{\frac{\nu_i}{n_i}} \quad (7.17)$$

$$= \frac{fidCorr_i}{effCorr_i} \nu_{ij} \delta_{ij} \frac{fidCorr_i}{effCorr_i} \quad (7.18)$$

$$= \frac{fidCorr_i}{effCorr_i} V_{ij} \frac{fidCorr_j}{effCorr_j} \quad (7.19)$$

$$= U_{ij} \quad (7.20)$$

The modified estimate is chosen symmetric in j and i in order to get a symmetric covariance in the non diagonal case of the migration matrix, too. The corresponding test statistic is denoted:

$$\chi_{N2,corr}^2 := (\hat{\mu} - \mu) \cdot \hat{U}_m^{-1} \cdot (\hat{\mu} - \mu) \quad (7.21)$$

Since theorists would assume a fixed covariance matrix, they would expect a chi-squared distribution according to Theorem 5.1. The correspondence of the $\chi_{N2,corr}^2$ values to the $\chi_{P,corr}^2$ values for the investigated $M_T(WZ)$ distributions is shown in Figure 7.12.

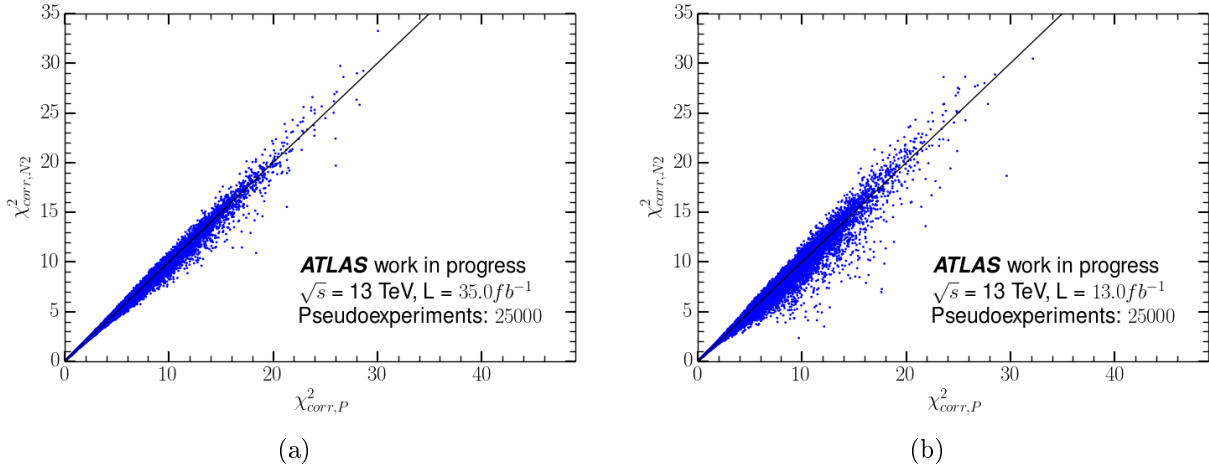


Figure 7.12: Comparison of $\chi_{P,corr}^2$ and $\chi_{N2,corr}^2$ for $M_T(WZ)$.

One can see, that this distribution is less scattered than for $\chi_{N,corr}^2$ shown in Figure 7.5. Also in the cumulated distribution of the $\chi_{N2,corr}^2$ values in Figure 7.13 and the plot of the actual significance (see Figure 7.14) look better. Nevertheless there are still some differences between $\chi_{P,corr}^2$ and $\chi_{N2,corr}^2$.

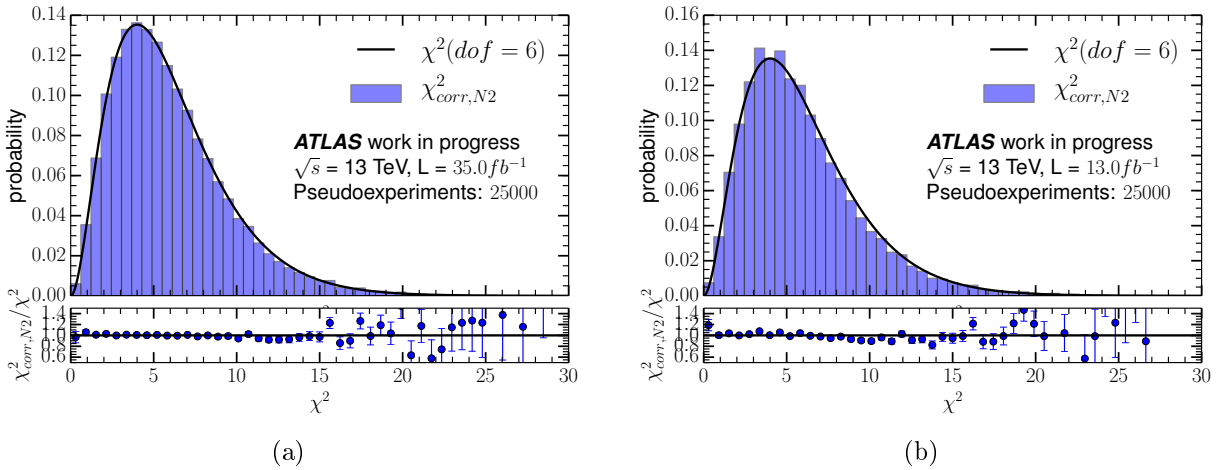


Figure 7.13: Distribution of $\chi_{N2,corr}^2$ values for toys of the $M_T(WZ)$ distribution.

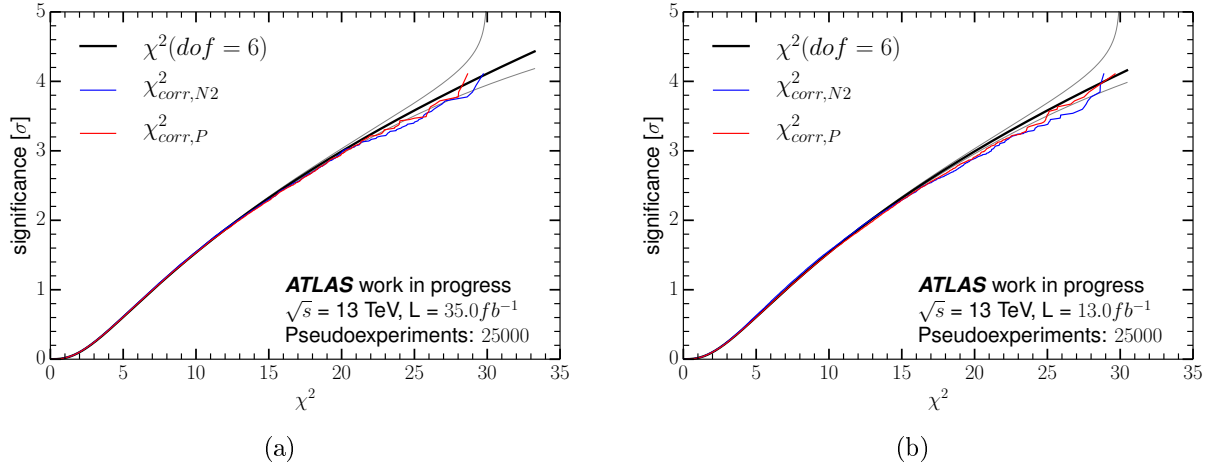


Figure 7.14: Comparison of the significances of $\chi^2_{P,corr}$ and $\chi^2_{N2,corr}$ values for toys of the $M_T(WZ)$ distribution.

This is further studied on a less diagonal migration matrix of the number of jets $NJet$ distribution shown in Figure 7.15. The used pseudo data is scaled to 175 fb^{-1} to have enough events in the last bin.

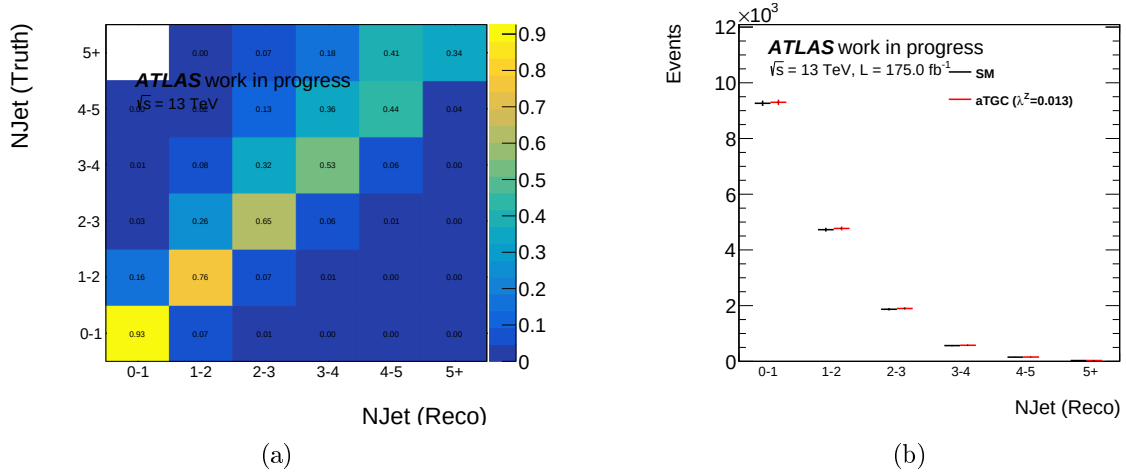


Figure 7.15: Migration matrix (a) and distribution on truth level (b) of the number of jets ($NJet$) distribution.

Here, the $\chi^2_{N2,corr}$ test statistic does not perform better than $\chi^2_{N,corr}$. It is even worse. This can be seen in Figure 7.16, 7.17 and Figure 7.18.

So, if the migration matrix gets significant different from a diagonal matrix then the new test statistic $\chi^2_{N2,corr}$ performs even worse then the unmodified one. Further studies have to show under which circumstances $\chi^2_{N2,corr}$ performs better than $\chi^2_{N,corr}$. Nevertheless, the deviations depend on the response matrix, which is not known in the fast unfolding approach.

This modified test statistic is maybe helpful if the response matrix got lost and in the descrip-

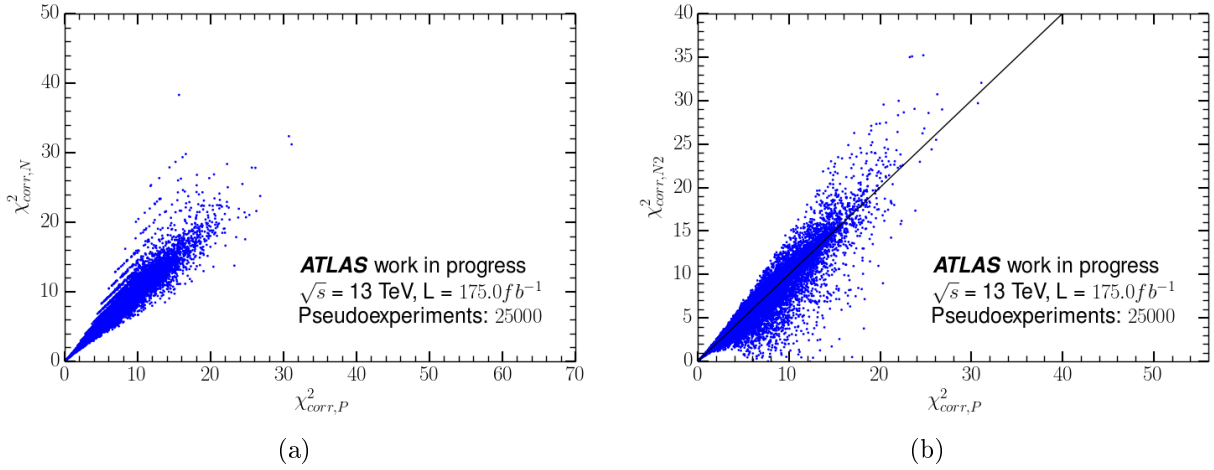


Figure 7.16: Comparison of $\chi^2_{P,corr}$, $\chi^2_{N,corr}$ (a) and $\chi^2_{N2,corr}$ (b) values for toys of the $NJet$ distribution.

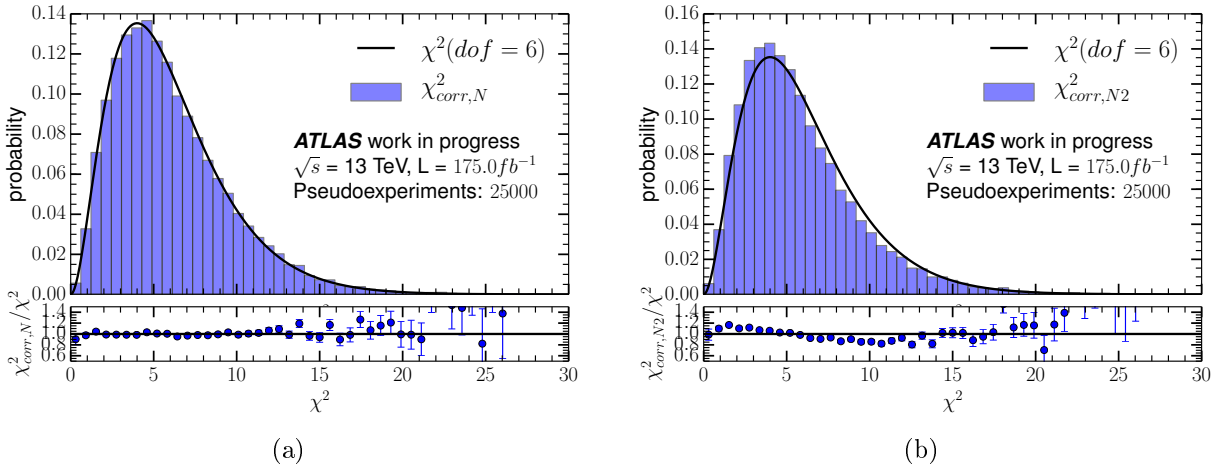


Figure 7.17: Distribution of $\chi^2_{N,corr}$ (a) and $\chi^2_{N2,corr}$ (b) values for toys of the $NJet$ distribution.

tion of the study it is stated that the response matrix had no strong migrations. But it would be better to publish and use the response matrix.

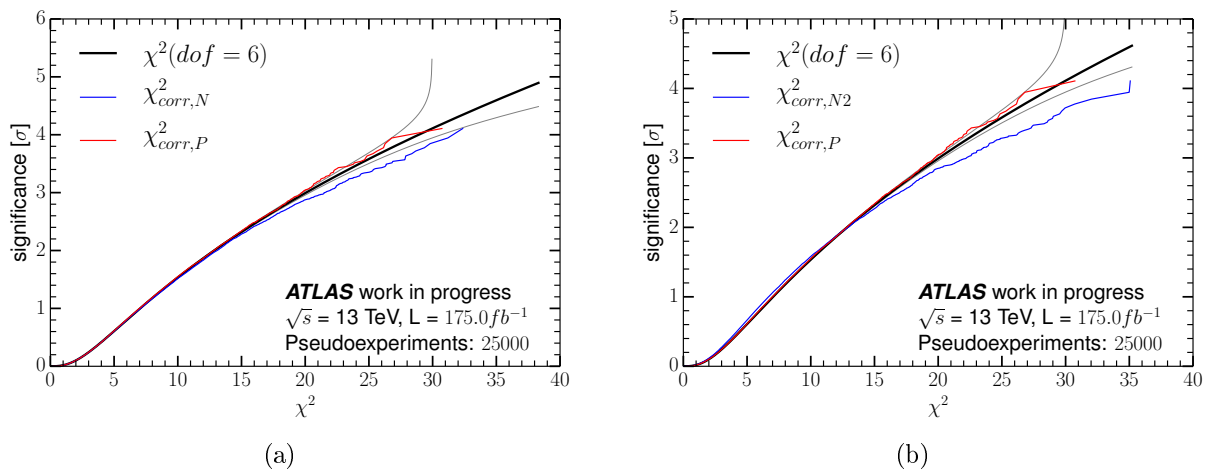


Figure 7.18: Comparison of the significances of $\chi^2_{P,corr}$, $\chi^2_{N,corr}$ (a) and $\chi^2_{N2,corr}$ (b) values for toys of the $NJet$ distribution.

8 Discussion and Conclusion

Using the example of distributions sensitive to $WZ \rightarrow WZ$ scattering it was analyzed which unfolding methods are best suitable for the comparison of data with theory.

The process of comparing theories with experimental data is usually divided into two steps, the preparation of the data and the comparison of the prepared data with theories. The first one is usually done by experimentalists and the latter by theorists.

Since one prepared dataset is potentially compared to many theories it is tried to provide a dataset with which the comparison with theories can be done fastly with a compact set of information. Furthermore, as few uncertainties as possible should be overestimated in the comparison to not lose significance. For potential comparisons by eye, correlated uncertainties between bins are hindering and are avoided.

Because of detector effects the true underlying (truth) distribution gets distorted to the reconstructed (reco) distribution. The comparison of theory with data can be done on reco or on truth level.

If the comparison is performed on truth level, detector effects have to be undone from the data using unfolding. For this purpose a response matrix derived from a Standard Model (SM) simulation is used. It includes the probabilities that an event coming from a certain bin j on truth level will be in a certain bin i on reco level. The mathematically exact way of unfolding is 'Matrix Inversion'. In the unfolding process regularization can be used to avoid bin-to-bin fluctuations in the unfolded data distribution. Four different unfolding approaches are compared:

the fast unfolding approach with Matrix Inversion/with regularization where only the unfolded data distribution including its uncertainties are provided by the experimentalists and

the extended unfolding approach with Matrix Inversion/with regularization where the response matrix is additionally used in the statistical test.

If the comparison is performed on reco level the theory has to be folded with the detector effects. This can be done in a **folding approach with full detector simulation**, in a **folding approach with simplified detector simulation**, such as discussed in [24], or with

a precalculated response matrix from the SM in a **folding approach with response matrix**. These seven approaches are compared in Table 8.1.

It is assumed that the uncertainties are bundled in covariance matrices in the unfolding approaches and that these are handled as Gaussian distributed uncertainties in the statistical tests. Therefore, five events per bin are needed in these approaches in order that the assumption of Gaussian distributed statistical uncertainties holds.

In the fast unfolding approaches, statistical uncertainties are estimated from the data distribution. In order not to underestimate this uncertainty, an additional systematic uncertainty would have to be included which reduces the significance. The estimate of the statistical uncertainties is better for higher number of events per bin.

The highest computational time for theorists occurs for the folding approach with full detector simulation, whereas the fast unfolding approaches are the fastest. The response matrix is usually calculated from a SM simulation. Due to effects of hidden variables the response matrix depends on the model it was generated with. This has to be accounted for with an additional systematic uncertainty reducing the significance.

All *unfolding* approaches need a similar amount of computational time. Therefore, the unfolding approach yielding the highest significance can be chosen. For computational comparisons Matrix Inversion in the extended unfolding approach has the most advantages because it has the highest expected significance.

If a response matrix is used then there are not many differences between folding the theory or unfolding the data in the extended approach with it. The advantage of folding the theory is that theorists can easily switch to a simplified or full detector simulation and can thereby use the full potential of the data.

As soon as clear structures are visible in the data distribution it can be useful to unfold with regularization for fast comparisons by eye. It is possible to see structures if the total uncertainty in each bin is smaller than the content of each bin. To ensure that the estimate of the statistical uncertainty is not too rough in a 2σ test, there should be more than 20-40 events in each bin on reco level.

The strength of regularisation and the binning can not be optimized in a general way, since the optimum depends on the theory it is compared to. One aim of published data is that it is useful for comparisons with as many theories as possible. To do so, it should be published unregularized on reco level with a fine binning together with a response matrix.

Approach and method	Needed infos by theorist (including unc.)		Computational time for theorists	Additional correlations between bins	Additional sources of unc.	Significance loss	Needed events per bin on reco level
folding approach with full detector simulation	full detector reco level data	sim.,	high	no		no	0
folding approach with simplified detector simulation	simplified sim., reco level data	detector	medium	no	not all detector effects covered	low	0
folding approach with response matrix	response matrix, reco level data		low	no	hidden variables	medium	0
extended unfolding approach with Matrix Inversion	response matrix, unfolded data	un-	low	medium	hidden variables	medium	5
extended unfolding approach with regularization	response matrix, unfolded data	un-	medium-low	medium-low	hidden variables, regularization	medium-high	5
fast unfolding approach with Matrix Inversion	unfolded data		very low	medium	hidden variables, estimated Δ_{stat} .	medium-high	$\gg 20$
fast unfolding approach with regularization	unfolded data		very low	medium-low	hidden variables, regularization, estimated Δ_{stat} .	high	$\gg 20$

Table 8.1: Comparison of different unfolding approaches and methods.

A Used Monte Carlo Samples

Monte Carlo samples from Powheg, Sherpa and MC@NLO are used. For each generator they include processes, where two partons go in and a W and a Z boson radiate with at least one strong coupling. Only leptonic decays of the W and Z are simulated.

Powheg QCD-Processes:

mc15_13TeV.361601.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_WZlvll_mll4.merge.DAOD_STDM5.e4475_s2726_r7772_r7676_p2669

Sherpa QCD-Processes:

mc15_13TeV.361064.Sherpa_CT10_llvSFMinus.merge.DAOD_STDM5.e3836_s2608_s2183_r7725_r7676_p2669

mc15_13TeV.361065.Sherpa_CT10_llvOFMinus.merge.DAOD_STDM5.e3836_s2608_s2183_r7725_r7676_p2669

mc15_13TeV.361066.Sherpa_CT10_llvSFPlus.merge.DAOD_STDM5.e3836_s2608_s2183_r7725_r7676_p2669

mc15_13TeV.361067.Sherpa_CT10_llvOFPlus.merge.DAOD_STDM5.e3836_s2608_s2183_r7725_r7676_p2669

MC@NLO QCD-Processes:

mc15_13TeV.361570.McAtNloHerwigEvtGen_AUET2CT10_WmZ_tau_tau_0_0_0.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361571.McAtNloHerwigEvtGen_AUET2CT10_WpZ_e_e_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361572.McAtNloHerwigEvtGen_AUET2CT10_WmZ_e_e_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361573.McAtNloHerwigEvtGen_AUET2CT10_WpZ_e_mu_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361574.McAtNloHerwigEvtGen_AUET2CT10_WmZ_e_mu_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361575.McAtNloHerwigEvtGen_AUET2CT10_WpZ_e_tau_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361576.McAtNloHerwigEvtGen_AUET2CT10_WmZ_e_tau_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361577.McAtNloHerwigEvtGen_AUET2CT10_WpZ_mu_e_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361578.McAtNloHerwigEvtGen_AUET2CT10_WmZ_mu_e_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361579.McAtNloHerwigEvtGen_AUET2CT10_WpZ_mu_mu_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361580.McAtNloHerwigEvtGen_AUET2CT10_WmZ_mu_mu_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361581.McAtNloHerwigEvtGen_AUET2CT10_WpZ_mu_tau_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361582.McAtNloHerwigEvtGen_AUET2CT10_WmZ_mu_tau_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361583.McAtNloHerwigEvtGen_AUET2CT10_WpZ_tau_e_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361584.McAtNloHerwigEvtGen_AUET2CT10_WmZ_tau_e_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361585.McAtNloHerwigEvtGen_AUET2CT10_WpZ_tau_mu_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361586.McAtNloHerwigEvtGen_AUET2CT10_WmZ_tau_mu_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361587.McAtNloHerwigEvtGen_AUET2CT10_WpZ_tau_tau_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

mc15_13TeV.361588.McAtNloHerwigEvtGen_AUET2CT10_WmZ_tau_tau_0_0_0p13.merge.DAOD_STDM5.e4439_s2608_r7772_r7676_p2722

B Selection Criteria

Objects reconstructed as AntiKt4EMTopoJets, having $p_T > 25$ GeV and $|\eta| < 4.5$ are selected as Jet objects. An overlap removal with identified leptons is applied. Jets that can be assigned to pile-up effects are rejected.

The applied selection criteria on the leptons and the jets can be found in [10].

Muon object selection			
Selection	Baseline selection	Z selection	W selection
$p_T > 7$ GeV	x	x	x
$ \eta < 2.5$	x	x	x
Loose quality	x	x	x
$ d_0^{\text{BL}}/\sigma(d_0^{\text{BL}}) < 3$	x	x	x
$ \Delta z_0^{\text{BL}} \sin \theta < 0.5$ mm	x	x	x
LooseTrackOnly isolation	x	x	x
μ -jet Overlap Removal		x	x
$p_T > 15$ GeV		x	x
Medium quality		x	x
Gradient Loose isolation		x	x
$p_T > 20$ GeV			x

Table B.1: Overview of muon object selections used in this work. [34]

Electron object selection			
Selection	Baseline selection	Z selection	W selection
$p_T > 7$ GeV	x	x	x
Electron object quality	x	x	x
$ \eta^{\text{cluster}} < 2.47, \eta < 2.5$	x	x	x
LooseLH+BLayer identification	x	x	x
$ d_0^{\text{BL}}/\sigma(d_0^{\text{BL}}) < 5$	x	x	x
$ \Delta z_0^{\text{BL}} \sin \theta < 0.5$ mm	x	x	x
LooseTrackOnly isolation	x	x	x
e -to- μ and e -to- e overlap removal	x	x	x
e -to-jets overlap removal		x	x
$p_T > 15$ GeV		x	x
Exclude $1.37 < \eta^{\text{cluster}} < 1.52$		x	x
MediumLH identification		x	x
Gradient Loose isolation		x	x
$p_T > 20$ GeV			x
TightLH identification			x
Gradient isolation			x

Table B.2: Overview of electron object selections used in this work. [34]

Event selection	
Event cleaning	Reject LAr, Tile and SCT corrupted events and incomplete events
Primary vertex	Hard scattering vertex with at least two tracks
MC trigger (2015)	HLT_e24_lhmedium_L1EM18VH HLT_e60_lhmedium HLT_e120_lhloose HLT_mu20_iloose_L1MU15 HLT_mu50
Data trigger (2015)	HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium HLT_e120_lhloose HLT_mu20_iloose_L1MU15 HLT_mu50
Trigger (2016)	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0 HLT_mu26_ivarmedium HLT_mu50
ZZ veto	Less than 4 baseline leptons
N leptons	Exactly three leptons passing the Z lepton selection
Leading lepton p_T	$p_T^{\text{lead}} > 25$ GeV (in 2015) or $p_T^{\text{lead}} > 27$ GeV (in 2016)
Z leptons	Two same flavor oppositely charged leptons passing Z lepton selection
Mass window	$ M_{\ell\ell} - M_Z < 10$ GeV
W lepton	W lepton passes W selection
W transverse mass	$m_T^W > 30$ GeV

Table B.3: Overview of analysis event selection used in this work. [34]

C Additional Plots

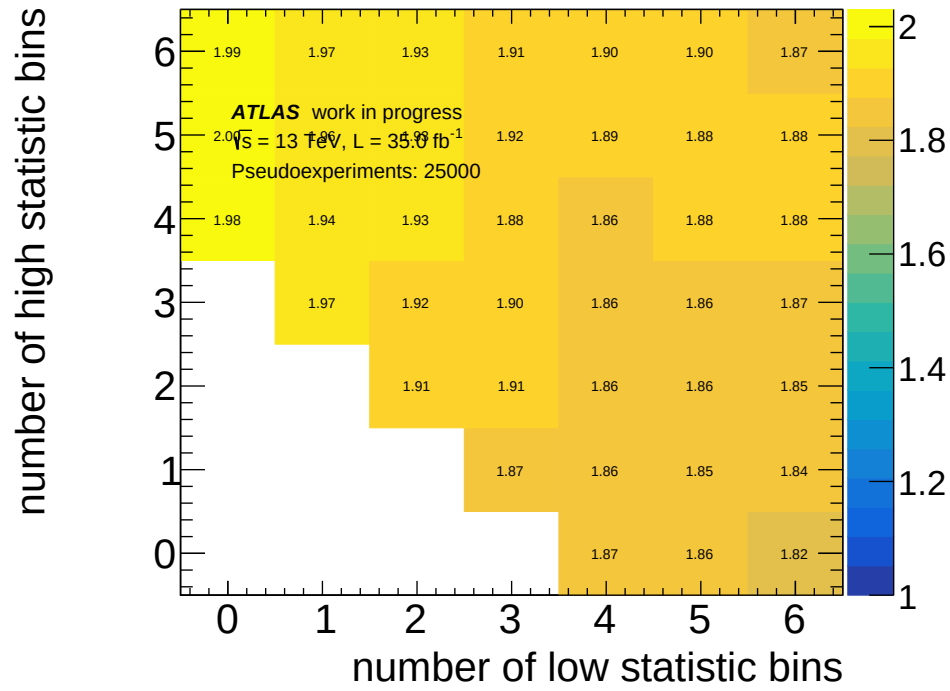


Figure C.1: Actual significance of estimated 2σ -limit with 38 events in low statistic bins

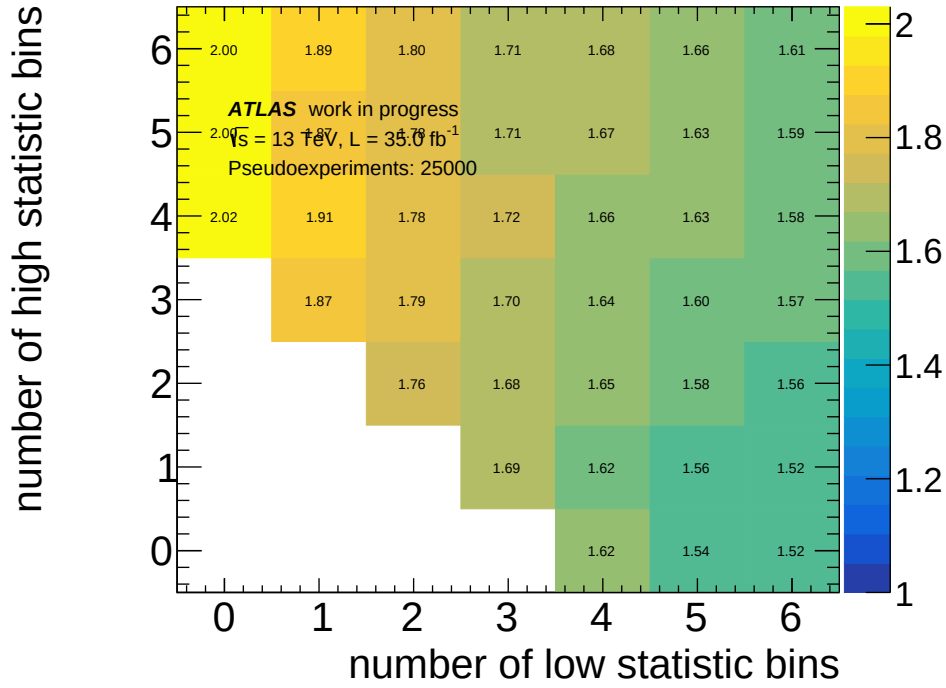


Figure C.2: Actual significance of estimated 2σ -limit with 10 events in low statistic bins

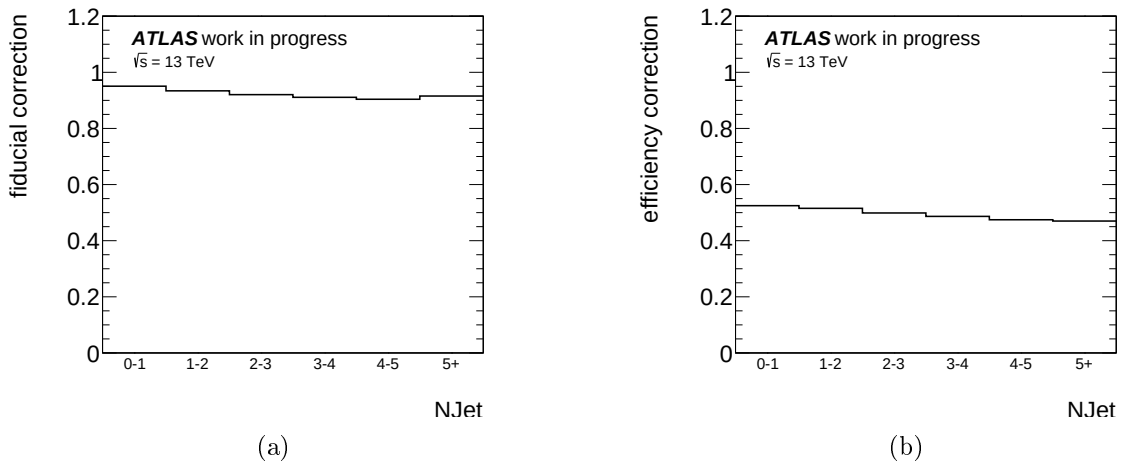


Figure C.3: Fiducial correction (a) and efficiency correction (b) of the number of jets distribution.

List of Figures

2.1	Possible Feynman diagrams for VBS with final states containing massive electroweak gauge bosons.[9]	8
2.2	Expected $M_T(WZ)$ distribution generated with Powheg.	9
3.1	Schematic representation of the ATLAS detector [14].	12
5.1	Correspondence of the significances to the probabilities for the two tails of a Gaussian distribution.	22
6.1	Schematic overview of the detector effects.	23
6.2	Different selection levels of events in unfolding.	25
6.3	Input for unfolding $M_T(WZ)$ generated with Powheg.	26
6.4	Example of large anti correlations in the unfolding result derived with Matrix Inversion.	30
6.5	Powheg prediction on reco level of $M_T(WZ)$ unfolded with the response matrix derived with Powheg and the prior from Sherpa for different numbers of iterations of the Bayesian Iterative algorithm.	34
6.6	Powheg prediction on reco level of $M_T(WZ)$ unfolded with the response matrix derived with Powheg and a flat prior for different numbers of iterations of the Bayesian Iterative algorithm.	34
6.7	Powheg prediction on reco level of $M_T(WZ)$ unfolded with the response matrix derived with Powheg and the prior from Powheg for different numbers of iterations of the Bayesian Iterative algorithm.	35
6.8	Visualization of the extended unfolding approach.	36
6.9	Visualization of the fast unfolding approach.	36
6.10	Migration Matrix of Powheg and Sherpa of the $M_T(WZ)$ variable.	39
6.11	Differences between the migration matrices of Sherpa and Powheg of the $M_T(WZ)$ variable.	40
6.12	Migration Matrix of Powheg and Sherpa of the p_T^Z variable	40
6.13	Differences between the migration matrices of Sherpa and Powheg of the p_T^Z variable.	41
7.1	Handling of statistical uncertainties in the fast unfolding approach.	43
7.2	Handling of statistical uncertainties in the extended unfolding approach.	43

7.3	Input for unfolding $M_T(WZ)$ generated with MC@NLO.	44
7.4	Prediction of the SM ($\lambda^Z = 0$) and an aTGC point ($\lambda^Z = 0.013$) of the $M_T(WZ)$ distribution on reco level generated with MC@NLO.	47
7.5	Comparison of $\chi^2_{P,corr}$ and $\chi^2_{N,corr}$ for $M_T(WZ)$	48
7.6	Distribution of $\chi^2_{P,corr}$ values for toys of the $M_T(WZ)$ distribution.	49
7.7	Distribution of $\chi^2_{N,corr}$ values for toys of the $M_T(WZ)$ distribution.	50
7.8	Dependence of the significance on the measured χ^2 value for $M_T(WZ)$	50
7.9	Actual significance of estimated 2σ -limit with zero high statistic bins	52
7.10	Actual significance of estimated 2σ -limit with one low statistic bin	52
7.11	Actual significance of estimated 2σ -limit with 15 events in low statistic bins	53
7.12	Comparison of $\chi^2_{P,corr}$ and $\chi^2_{N2,corr}$ for $M_T(WZ)$	54
7.13	Distribution of $\chi^2_{N2,corr}$ values for toys of the $M_T(WZ)$ distribution.	54
7.14	Comparison of the significances of $\chi^2_{P,corr}$ and $\chi^2_{N2,corr}$ values for toys of the $M_T(WZ)$ distribution.	55
7.15	Migration matrix and distribution on truth level of the number of jets ($NJet$) distribution.	55
7.16	Comparison of $\chi^2_{P,corr}$, $\chi^2_{N,corr}$ and $\chi^2_{N2,corr}$ values for toys of the $NJet$ distribution.	56
7.17	Distribution of $\chi^2_{N,corr}$ and $\chi^2_{N2,corr}$ values for toys of the $NJet$ distribution.	56
7.18	Comparison of the significances of $\chi^2_{P,corr}$, $\chi^2_{N,corr}$ and $\chi^2_{N2,corr}$ values for toys of the $NJet$ distribution.	57
C.1	Actual significance of estimated 2σ -limit with 38 events in low statistic bins	67
C.2	Actual significance of estimated 2σ -limit with 10 events in low statistic bins	68
C.3	Fiducial correction and efficiency correction of the number of jets distribution.	68

List of Tables

2.1	Summary of fermions in the Standard Model.[8]	4
2.2	Summary of bosons in the Standard Model.[8]	4
7.1	Prediction of the SM ($\lambda^Z = 0$) and an aTGC point ($\lambda^Z = 0.013$) of the $M_T(WZ)$ distribution on reco level for generated with MC@NLO.	47
8.1	Comparison of different unfolding approaches and methods.	61
B.1	Overview of muon object selections used in this work. [34]	65
B.2	Overview of electron object selections used in this work. [34]	66
B.3	Overview of analysis event selection used in this work. [34]	66

D Bibliography

- [1] G. Aad et al.(ATLAS Collaboration), **Measurements of $W^\pm Z$ production cross sections in pp collisions at $\sqrt{s} = 8\text{TeV}$ with the ATLAS detector and limits on anomalous gauge boson self-couplings**, 10.1088/1126-6708/2009/02/007, 2016
- [2] **Longer term LHC schedule**, <https://lhc-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-long-term.htm>, Accessed: 27.03.2017
- [3] P. W. Higgs, **Broken Symmetries and the Masses of Gauge Bosons**, Phys. Rev. Lett. 13, 508, 1964
- [4] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, **Global Conservation Laws and Massless Particles**, Phys. Rev. Lett. 13, 585, 1964
- [5] ATLAS Collaboration, **Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC**, arXiv:1207.7214 [hep-ex], 31.8.2012
- [6] D. J. Griffiths, **Introduction to elementary particles**, Second Revised edition, Physics textbook, WILEY-VCH, 2008, ISBN: 978-3-527-40601-2
- [7] M. Thomson, **Modern Particle Physics**, Cambridge university press, 2016, ISBN: 978-1-107-03426-6
- [8] C. Patrignani et al. (Particle Data Group), **Review of Particle Physics**, Regents of the University of California, 2016
- [9] F. Socher, **Study of Electroweak Gauge Boson Scattering in the WZ Channel with the ATLAS Detector at the Large Hadron Collider**, PhD thesis, Dresden, TU Dresden, Jul 2016
- [10] ATLAS Collaboration, **Measurement of $W^\pm Z$ boson pair-production in pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS Detector and confidence intervals for anomalous triple gauge boson couplings**, ATLAS-CONF-2016-043, 2016
- [11] J. Beringer; et al. (2012). 10.1103/PhysRevD.86.010001 section 43.6.1 (page 425), 2012

-
- [12] M. Rauch, **Vector-Boson Fusion and Vector-Boson Scattering**, arXiv:1610.08420 [hep-ph], 2016
- [13] Lyndon Evans and Philip Bryant, **LHC Machine**, JINST 3 S08001, 2008
- [14] The ATLAS Collaboration et al, **The ATLAS Experiment at the CERN Large Hadron Collider**, JINST 3 S08003, 2008
- [15] A. Buckley et al., **General-purpose event generators for LHC physics**, Physics Reports 504 no. 5, (2011) 145–233, 2011
- [16] **LuminosityPublicResults**,
<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>, Accessed: 25.03.2017
- [17] T. Gleisberg, S. Hoeche, F. Krauss, M. Schoenherr, S. Schumann, F. Siegert, J. Winter, **Event generation with SHERPA 1.1**, arXiv:0811.4622 [hep-ph], 2008
- [18] P. Nason, **A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms**, arXiv:hep-ph/0409146, 2004
- [19] S. Frixione, B.R. Webber, **Matching NLO QCD computations and parton shower simulations**, arXiv:hep-ph/0204244, 2002
- [20] O. Behnke, K. Kröninger, G.Schott, T. Schörner-Sadenius, **Data Analysis in High Energy Physics**, WILEY-VCH, 2013, ISBN: 978-3-527-41058-3
- [21] G. Cowan, **Statistical Data Analysis**, Oxford Science Publications, 1998, ISBN: 0-19-850156-0
- [22] K.A. Oliver et al. (Particle Data Group), Chin. Phys. C 38 090001 (2014) and 2015 update
- [23] EWUnfolding Source Code, <svn+ssh://svn.cern.ch/repos/atlasphys-sm/Physics/StandardModel/ElectroWeak/Analyses/EWUnfolding/branches/EWUnfolding-00-00-01-branch/Code>, Revision: 257250, Accessed: 25.04.2016
- [24] S. Oryn, X. Rouby, V. Lemaitre, **Delphes, a framework for fast simulation of a generic collider experiment**, arXiv:0903.2225 [hep-ph], 2009
- [25] Gene H. Golub, James M. Ortega **Scientific Computing and Differential Equations**, 1992, ISBN: 0122892550
- [26] G. Cowan, **Comments on unfolding**, ATLAS Searches Meeting CERN, 24 January, 2011

-
- [27] S. Schmitt, **TUnfold, an algorithm for correcting migration effects in high energy physics**, Journal of Instrumentation, vol 7, 10 T10003, 2012
- [28] Andreas Höcker and Vakhtang Kartvelishvili, **SVD approach to data unfolding**, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol 372, 3 469-481, 1996
- [29] B. Malaescu, **An Iterative, Dynamically Stabilized(IDS) Method of Data Unfolding**, arXiv:1106.3107 [physics.data-an], 2011
- [30] G. D'Agostini, **A multidimensional unfolding method based on Bayes' theorem**, Meth. in Phys. Res. A362 (1995) 487-498, 1995
- [31] G. D'Agostini, **Improved iterative Bayesian unfolding**, arXiv:1010.0632 [physics.data-an], 2010
- [32] ATLAS Collaboration **Measurement of dijet cross sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector** arXiv:1312.3524v3 [hep-ex], page 16, 2014
- [33] Robert D. Cousins, Samuel J. May, Yipeng Sun **Should unfolded histograms be used to test hypotheses?** arXiv:1607.07038v1 [physics.data-an], 2016
- [34] ATLAS Collaboration, **Measurement of the $W^\pm Z$ boson pair-production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector and limits on aTGCs**, ATL-COM-PHYS-2016-714, 2016
- [35] G. Choudalakis, on behalf of the ATLAS Collaboration, **Unfolding in ATLAS**, arXiv:1104.2962 [hep-ex], 2011
- [36] G. Cowan, **A survey of unfolding methods for particle physics**, in Advanced Statistical Techniques in Particle Physics. Proceedings, Conference, Durham, UK, pg. 248, 2002
- [37] L. Lyons **Unfolding: Introduction**, in Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, H.B. Prosper and L. Lyons eds., Geneva Switzerland , pg. 225, 2011
- [38] V. M. Panaretos **A Statistician's View on Deconvolution and Unfolding**, in Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, H.B. Prosper and L. Lyons eds., Geneva Switzerland , pg. 229, 2011

- [39] V. Blobel, **Unfolding Methods in Particle Physics**, in Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, H.B. Prosper and L. Lynons eds., Geneva Switzerland , pg. 240, 2011
- [40] G. Zech, **Regularization and error assignment to unfolded distributions**, in Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, H.B. Prosper and L. Lynons eds., Geneva Switzerland , pg. 252, 2011
- [41] Analysis framework **ROOT** root.cern.ch

Danksagung

An dieser Stelle möchte ich allen danken die zum Gelingen dieser Arbeit und meines Studiums beigetragen haben.

Im Besonderen möchte ich mich bei Prof. Michael Kobel bedanken für die Möglichkeit, dieses interessante Thema in seiner Arbeitsgruppe bearbeiten zu können, die zielführenden Diskussionen und die herzliche Unterstützung.

Des Weiteren möchte ich mich bei der gesamten VBS-Arbeitsgruppe für die stets sehr angenehme und produktive Arbeitsatmosphäre bedanken. Carsten Bittrich, Felix Socher, Franziska Iltzsche, Stefanie Todt und Sarah Krebs halfen stets bei Fragen aller Art, ob technischer oder physikalischer Natur. Besonderer Dank gilt Carsten Bittrich und Felix Socher für die zahlreichen produktiven Diskussionen inklusive Denkanstößen. Außerdem möchte ich mich bei meinen Korrekturlesern für die vielen hilfreichen Kommentare bedanken.

In addition, I would like to thank Dr. Valerio Dao, who helped me in the first month of this work and got me a lot of valuable information about my topic. I would like to thank Bogdan Malaescu for his discussions and suggestions during and before the P&P week in November and Dimitris Iliadis for providing the Powheg and Sherpa samples.

Meinen Freunden, insbesondere denen die mich im Studium begleitet haben, möchte ich danken für eine schöne Studienzeit.

Mein besonderer Dank gilt meiner Familie, insbesondere meiner Eltern, für die emotionale und finanzielle Unterstützung während meines gesamten Studiums.

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.

Tim Herrmann

Dresden, März 2017