Universität Bonn

Physikalisches Institut

Measurement of the WZ cross section in the channel $WZ \to l\nu b\bar{b}$ at $\sqrt{s} = 8\,\mathrm{TeV}$ with the ATLAS detector

Nicklas Christian Denis

Physikalisches Institut der
Universität Bonn
Nussallee 12
D-53115 Bonn

BONN-IB-2015-01
November 2014

Acknowledgements

Foremost, I would like to express my gratitude to Dr. Götz Gaycken who supervised me during my master thesis and provided a lot of advice and guidance.

I would like to thank Prof. Dr. Norbert Wermes for giving me the opportunity to write my thesis in his group and to spend a semester at CERN. Furthermore, I would like to thank Dr. Tatjana Lenz who supervised my thesis in Bonn and also supported me a lot. I thank my colleagues Stephan Hageböck, Elisabeth Schopf, Holger Ritter and Eckhard von Törne for advice, comments and good ideas.

I owe much to Robert, Katrin and Fia who supported me with humor and encouragement.

Contents

CHAPTER 1

Introduction

In this thesis, a measurement of the total cross section of WZ production in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ is presented. The measurement is performed with the ATLAS detector in the decay channel $WZ \to l\nu b\bar{b}$ based on the $20.3\,\text{fb}^{-1}$ dataset from the 2012 run of the LHC.

The process $WZ \to l\nu b\bar{b}$ is one of the important backgrounds in the still-ongoing search for the $H \to b\bar{b}$ decay of the Standard Model Higgs boson in associated WH production. A measurement of the WZ cross section is a suitable cross-check for the WH analysis since WZ production is already a well-explored process of the Standard Model.

Chapter 2 introduces the process $WZ \to l\nu b\bar{b}$ as well as the important backgrounds and gives an overview of the LHC and the ATLAS detector. Chapter 3 discusses the dataset, the simulated event samples and the object reconstruction this analysis is based on. An outline of the analysis strategy is given in chapter 4. The chapters 5 and 6 explain different analysis techniques and their application. Systematic uncertainties on the cross section measurement are discussed in chapter 7. The results are presented in chapter 8 and 9.

CHAPTER 2

Physics and experiment

2.1 WZ as background for WH

In July 2012, a new particle with a mass of 125 GeV was discovered by the ATLAS and CMS experiments [1]. It is consistent with the expected properties of the Standard Model Higgs boson. Up to now, the presumed Higgs boson was observed in the channels $H \to \gamma\gamma$, $H \to WW$, $H \to ZZ$ and $H \to \tau\tau$. Despite a predicted branching ratio of 58%, the particle was not yet observed in the decay $H \to b\bar{b}$ [2]. The search for a Higgs signal in this particular channel is extremely challenging because of the huge amount of background[1]. The largest background is QCD multijet production, i.e. events in which several hadron jets are produced via strong interaction.

A considerable reduction of the multijet background is possible if the search for $H \to b\bar{b}$ is restricted to the associated production of the Higgs boson with a W boson in a process referred to as "Higgs-strahlung". The Feynman graph is shown in figure 2.1: The W boson is produced in the initial parton

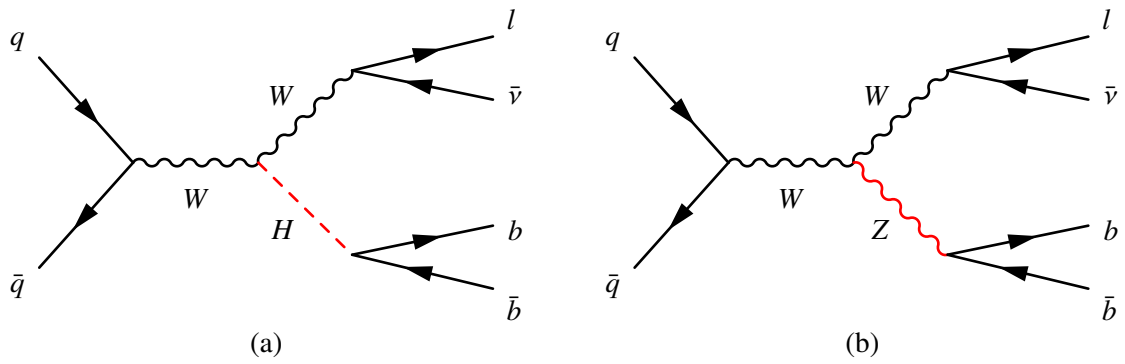

Figure 2.1: The processes (a) $WH \to l\nu b\bar{b}$ and (b) $WZ \to l\nu b\bar{b}$ (dominant production mode)

interaction and then radiates off a Higgs boson. The decay of the W boson to an electron or muon and the corresponding neutrino can be used to trigger the event selection, since isolated high-energetic leptons are rarely produced in multijet events.

Besides multijet, several other background processes play a role. One of these backgrounds is

[1] other processes that create similar detector signatures

WZ production. The dominant production mode, the equivalent of the Higgs-strahlung, is shown in figure 2.1. The goal of this analysis is to measure the total cross section of WZ production in the decay channel $WZ \to l\nu b\bar{b}$.

2.2 Backgrounds for $WZ \to l\nu b\bar{b}$

The detector signature of $WZ \to l\nu b\bar{b}$, visualized in figure 2.2, consists of a high-energetic isolated electron or muon in association with two b-tagged jets. Events with one additional un-tagged jet that may originate from initial or final state gluon radiation are also taken into account. The missing trans-

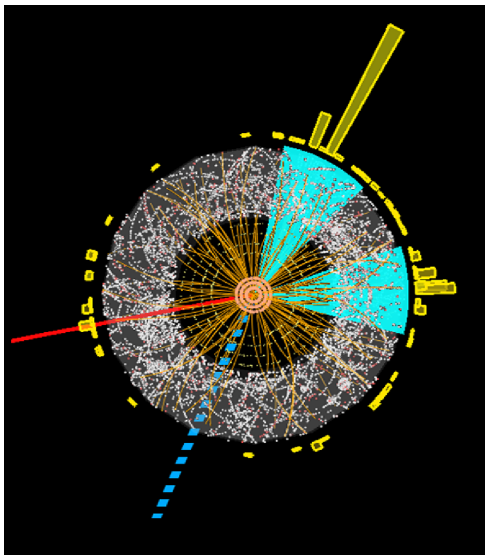

Figure 2.2: ATLAS detector signature of an event with 2 b-jets (blue), a high-energetic muon (red) and a neutrino (dashed blue) [3]

verse energy (see section 3.3) as indicator of a possible neutrino plays a role in the preselection cuts (see section 4.3) and the BDT (see section 5). The required signature can be matched by different processes, which appear as backgrounds in the search for the WZ signal:

Top-antitop pairs The dominant decay mode of the top quark is through the weak interaction producing a bottom quark and a W boson. Hence, $t\bar{t}$-pair production, shown in figure 2.3, can provide the required two b-jets. For the two W bosons, two relevant cases can be distinguished: Either both decay leptonically and one of the leptons is not detected or one decays leptonically and the other hadronically and at least one of the four jets is not detected.

Single top The single top background, shown in figure 2.4, comprises three different processes that all have one top quark in the final state. The required electron or muon can originate from a semileptonic decay of the top quark or, in case of the Wt-channel, from the final-state W boson.

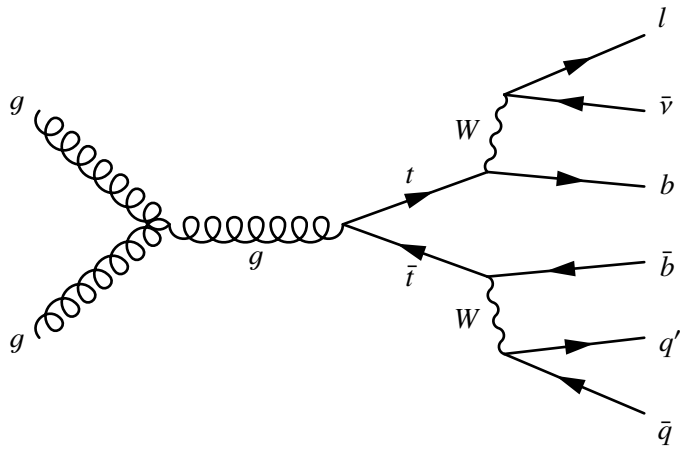

Figure 2.3: The production of a $t\bar{t}$-pair through gluon-gluon fusion with one W boson decaying leptonically and the other hadronically.

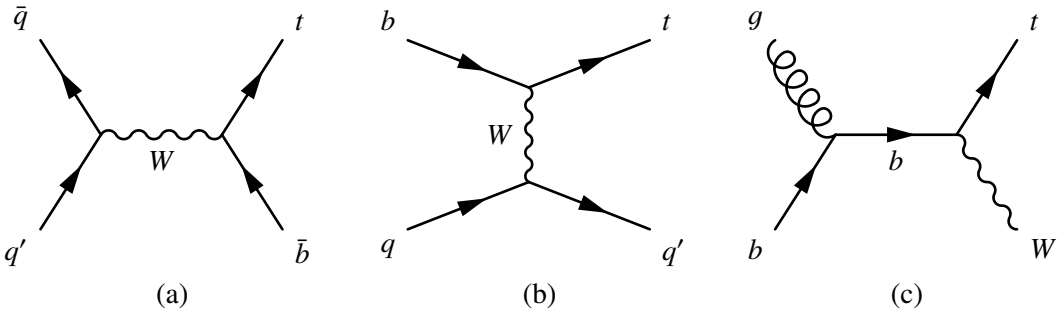

Figure 2.4: Single top background: (a) s-channel (b) t-channel (c) Wt production

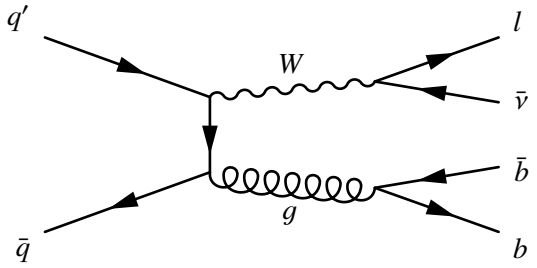

Figure 2.5: The W+bb production as an example of W+jets background.

W+jets The term W+jets refers to processes in which a W boson is produced together with jets from initial or final state radiation. While $W+bb$ production, shown in figure 2.5, has the highest chance of fulfilling the selection requirements, several other processes contribute significantly due to flavor misidentification.

Z+jets The Z+jets background is strongly suppressed by the requirement of exactly one charged lepton.

Diboson While WZ production is considered the signal process in this analysis, WW gives a very small background contribution and ZZ is fully suppressed.

WH/ZH The WH production yields a tiny background contribution whereas ZH is fully suppressed.

QCD multijet Multijet refers to the production of jets via strong interaction, one of the most important backgrounds at hadron colliders. Two main classes of multijet events can be distinguished:

1. semileptonic heavy-flavor decays in jets
2. jets or photon conversions that are misidentified as electrons

Due to the large number of contributing processes and the enormous cross section in combination with a low detector acceptance, no appropriate Monte Carlo simulation is available for multijet background. Instead, shape templates are derived from data by inverting the track isolation (see section 4.3), using the fact that track activity in the vicinity of the presumed lepton is typical for multijet background. This approach is explained in detail in [4]. The estimation of the multijet background by this method bears large uncertainties, which makes multijet suppression a priority of the event selection.

2.3 The ATLAS experiment

2.3.1 Detector systems

ATLAS (A Toroidal LHC Apparatus) is a general-purpose detector at the LHC, built to probe high-energetic proton collisions [5]. The whole detector is shown in figure 2.6. The different subsystems form a series of concentric cylinders around the interaction point. They are briefly discussed in the following, going from the center outwards:

Inner detector The inner detector measures the tracks of charged particles. A solenoid magnet that surrounds the entire inner detector causes the tracks to bend, which allows to determine the momentum of the particle from the curvature. The tracking system in turn consists of three different detectors: A multi-layer silicon pixel detector offers a high tracking precision near the interaction point. It is followed by a silicon-based semiconductor tracker (SCT) similar to the pixel detector but with long strips rather than pixels. The outermost layer of the inner detector, the transition radiation tracker (TRT), uses a combination of ionization-based tracking with gas-filled straw tubes and particle identification (especially electron-pion discrimination) via transition radiation created in a radiator between the straws.

Calorimeter The purpose of the calorimeters, situated outside the solenoid magnet, is to measure the energy of a particle. An incoming particle interacts with high-density material in the calorimeter, creating showers of secondary particles. These are finally stopped and their energy is absorbed

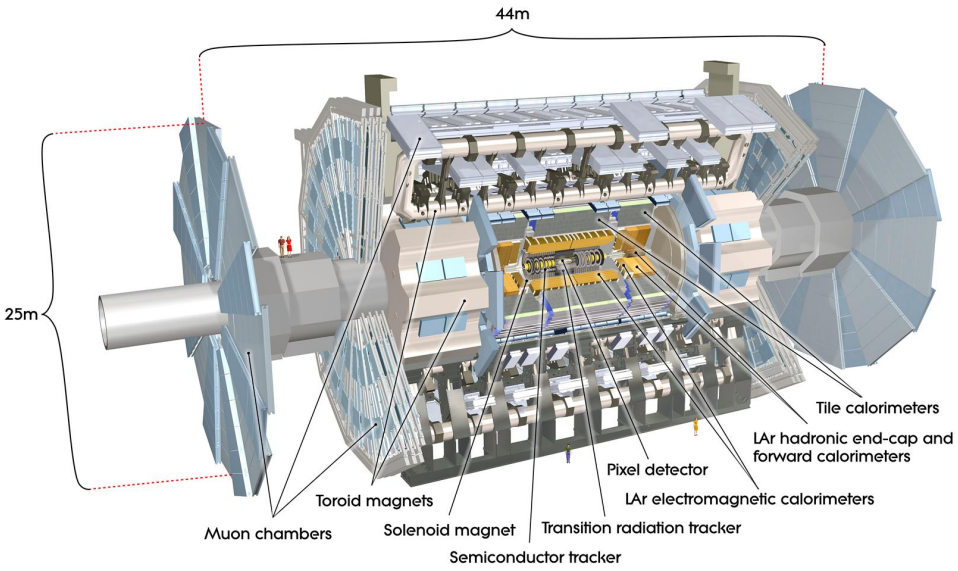

Figure 2.6: The ATLAS detector [3]

in the calorimeter. ATLAS has two calorimeter systems, both of which are designed as sampling calorimeters with alternating passive absorber layers and active sampling layers: The inner electromagnetic calorimeter, which uses lead as absorber and liquid argon (LAr) for sampling, is sensitive to particles that produce electromagnetic showers and plays an important in the identification of electrons and photons. The outer hadronic calorimeter absorbs the energy of strongly interacting particles via hadronic showers. The barrel part of the hadronic calorimeter is a tile calorimeter with steel as absorber and scintillating tiles for sampling.

Muon spectrometer The muon spectrometer is the outermost part of the ATLAS detector. Nearly all particles that are not stopped at the latest in the hadronic calorimeter are muons, which makes muon identification comparatively easy. The muon spectrometer uses drift tube chambers and several other technologies. It also has its own magnet system, which allows to determine the muon momentum from the track curvature.

2.3.2 Coordinate system

The coordinate system used in ATLAS defines the nominal interaction point as the origin with the x-axis pointing towards the center of the LHC, the y-axis pointing upwards and z-axis pointing along the beam line towards the LHCb experiment [5]. The azimuthal angle ϕ is measured around the beam axis and the polar angle θ is the angle with respect to the beam axis. Instead of θ, one usually uses the pseudorapidity

$$\eta = -\ln(\tan\theta/2)\,, \tag{2.1}$$

which has the advantage that differences in η are approximately invariant under longitudinal boosts. Distances in the (η, ϕ)-plane are written as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}\,. \tag{2.2}$$

CHAPTER 3

Data, simulation & event reconstruction

3.1 Data

This analysis uses data from proton-proton collisions recorded with the ATLAS detector in the 2012 run at a center-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$. The dataset contains only events that have passed the quality requirements of the ATLAS Good Run List, making sure that the detector systems and triggers were operating correctly. The integrated luminosity of the used dataset amounts to $20.3\,\text{fb}^{-1}$.

3.2 Event simulation

The simulation of collision events plays an important role in the analysis of data at hadron colliders. A typical proton-proton collision is extremely complex and involves the production of many different particles with momenta ranging over several orders of magnitude. Besides the hard parton interaction, soft QCD phenomena such as hadronization play a role. Therefore, the complete event cannot be computed from theory and requires modeling that relies on phenomenological approaches.

The measurement of the WZ cross section presented in this thesis is based on a comparison of data and simulated background (see chapter 4). The signal and the different background processes except multijet

	Process	Generator	Sample size
Signal	WZ	HERWIG	20M
Vector boson + jet	W+jets	SHERPA 1.4.1	168M
	Z+jets	SHERPA 1.4.1	42M
Top-quark	$t\bar{t}$	POWHEG+PYTHIA	75M
	single top t-channel	ACER+PYTHIA	20M
	single top s-channel	POWHEG+PYTHIA	6M
	single top Wt-channel	POWHEG+PYTHIA	9M
Diboson	WW	HERWIG	10M
	ZZ	HERWIG	7.5M
Higgs	WH	PYTHIA8	300000
	ZH	PYTHIA8	300000

Table 3.1: Monte Carlo samples used in this analysis

production were modeled in Monte Carlo simulation [6–8] using the same center-of-mass energy and event pileup[1] as the data. The different background samples and the respective Monte Carlo generators are listed in table 3.1.

3.3 Event reconstruction

The physical event has to be reconstructed from the raw detector data. The reconstruction of objects relevant for this analysis is briefly discussed in the following:

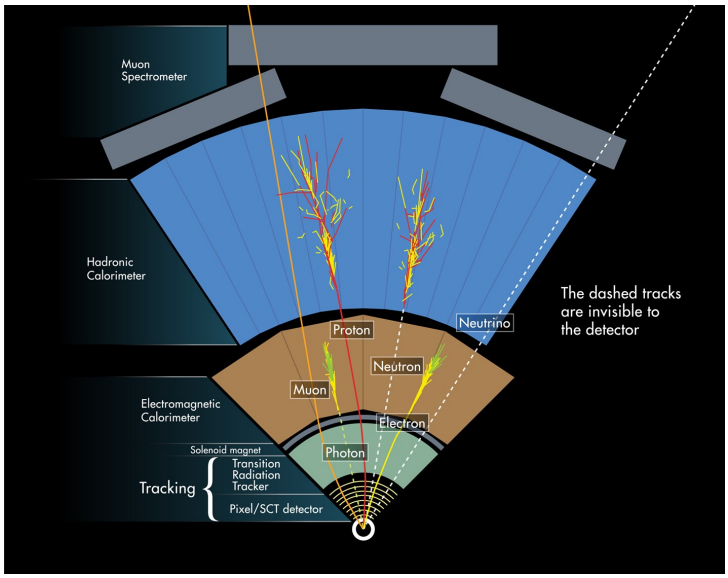

Figure 3.1: Signatures of different particle species in the ATLAS detector [3]

Track & vertex reconstruction Tracks are reconstructed with a clustering algorithm on the basis of combined data from pixel detector, SCT and TRT. A vertex-finder algorithm is then applied to the reconstructed tracks to identify primary and secondary vertices.

Calorimeter clustering Clustering algorithms run on calorimeter entries to identify electromagnetic and hadronic showers.

Electron reconstruction The characteristic signature of an electron consists of a shower in the electromagnetic calorimeter in combination with a track in the inner detector pointing at the shower (see figure 3.1). Photons can create electromagnetic showers as well but do not leave tracks.

Muon reconstruction Due to their higher mass in comparison to the electron, muons do not emit bremsstrahlung and therefore do not create electromagnetic showers. Since they also do not interact strongly, they can pass the hadronic calorimeter and leave a track in the muon spectrometer, which sets them apart from the other particle species.

Jet reconstruction Jets are reconstructed on the basis of calorimeter clusters using the anti-k_T algorithm [9].

[1] additional soft proton-proton interactions in the same bunch crossing as the hard scattering process

Missing transverse energy The neutrino can not be measured directly, as it does not interact with detector matter. However, it can be detected implicitly using a quantity called "missing transverse energy": In the plane transverse to the beam, one can make use of momentum conservation. If there is exactly one neutrino in the event, the negative vectorial sum of all other transverse momenta equals the transverse momentum of the neutrino. The construction of the missing transverse energy is based on calorimeter entries.

***b*-tagging** The identification of b-jets (jets originating from a b-quark), referred to as "b-tagging", is crucial for this analysis. It is a challenging task to discriminate their detector signature against that of light- or c-flavored jets. The properties b-tagging mainly relies on are the large mass of the b-hadrons and their long lifetime, which allows them to travel macroscopic distances before they decay. Thanks to the latter property, there is a chance to resolve the secondary vertex from the primary. This analysis uses the MV1c algortithm for b-tagging. It uses a neural network to exploit all relevant information from the jet. MV1c is an improved version of MV1 [10, 11] trained for better rejection of c-jets. A weight between zero and one as a measure of the "b-likeness" is assigned to each jet. In this analysis, a jet is considered b-tagged if it has an MV1c weight larger than 0.405, which corresponds to an average b-tagging efficiency[2] of 80%.

[2] the percentage of the b-jets that was correctly tagged

CHAPTER 4

WZ analysis strategy

4.1 Signal significance

The cross section measurement is based on a comparison of the data with the simulated background: A possible excess over background is interpreted as the signal strength. It is proportional to the cross section. This method requires to suppress background as far as possible and to maximize the expected signal significance by a suitable event selection, i.e. the expected excess should be clearly larger than fluctuations of the background.

Signal significance is usually expressed in terms of Gaussian standard deviations, stating how large the fluctuation of a normally distributed background would have to be to account for the observed excess [12]. A high significance corresponds to a low probability for such a fluctuation to occur. Based on profile likelihood estimation [13], the following expression for the signal significance can be derived:

$$Z = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right)-s\right)}, \tag{4.1}$$

where s and b are the numbers of signal and background events. For large s and b, this formula can by approximated by

$$Z \approx \frac{s}{\sqrt{s+b}}, \tag{4.2}$$

which, in the case $s \ll b$, reduces to

$$Z \approx \frac{s}{\sqrt{b}}. \tag{4.3}$$

4.2 Analysis overview

A reliable cross section measurement requires an event selection that suppresses the simulated background as far as possible and maximizes the expected signal significance. The same selection criteria are applied to both data and simulation. The selection is divided into two steps:

1. A preselection covers basic requirements such as the existence of a high-energetic charged lepton and two b-tagged jets. Furthermore, kinematic phase space regions where the background description bears large uncertainties, which is e.g. the case for multijet-dominated regions, can be excluded at this point. For details on the preselection see section 4.3.

2. A boosted decision tree (BDT, see section 5) that has been trained on simulated events is applied to optimally separate signal and background. The cut on the BDT classifier distribution is chosen such that the expected significance is maximized.

The normalization factors of chosen backgrounds are allowed to float freely in a fit of the total background to the data. To avoid a bias on the cross section measurement, the fit is performed in signal-depleted "control regions". The so-obtained corrections are then applied to the signal region. The fit is explained in detail in section 6.

The number of signal events is measured by subtracting the number of background events from the number of data events after cut on the BDT classifier. These numbers are determined by integrating the background and data histograms over the remaining range of the classifier. The number of signal events can be written as

$$N_{\text{sig}} = \epsilon \mathcal{L}_{\text{int}} \sigma_{WZ} \,, \tag{4.4}$$

where ϵ is the selection efficiency, $\mathcal{L}_{\text{int}}$ is the integrated luminosity and σ_{WZ} is the fully inclusive cross section of WZ production. Since the Monte Carlo simulation is normalized to the same value of $\epsilon \mathcal{L}_{\text{int}}$ as data, the cross section can be determined by comparing the measured event yield N_{sig} with the number $N_{\text{sig}}^{\text{sim}}$ expected from simulation:

$$\sigma_{WZ} = \frac{N_{\text{sig}}}{N_{\text{sig}}^{\text{sim}}} \sigma_{WZ}^{\text{theo}} \,,$$

where $\sigma_{WZ}^{\text{theo}}$ is the theory prediction of the cross section used to normalize the simulated signal sample.

4.3 Event preselection

The preselection is adopted from the ATLAS $H \to b\bar{b}$ analysis [2] with a few changes explained below. The most important requirements are listed in the following:

- exactly one electron or muon with the following properties:
  - $p_{\text{T}} > 25\,\text{GeV}$
  - $|\eta| < 2.5$ ($|\eta| < 2.47$ for electrons)
  - associated track in the inner detector
  - calorimeter isolation: total energy deposit within a cone of $\Delta R = 0.3$ around the lepton less than 7% of the lepton E_{T}.
  - track isolation: sum of transverse track momenta within a cone of $\Delta R = 0.2$ around the lepton less than 4% of the lepton p_{T}
- jet selection: $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.5$ (the range where b-tagging can be applied)
- two b-tagged jets
- optionally, one additional jet without b-tag

Additional cuts on different event variables are applied to further suppress backgrounds. Here, two changes are made with respect to the ATLAS $H \to b\bar{b}$ analysis:

- $p_{\text{T}} > 35\,\text{GeV}$ (45 GeV in the WH analysis) for one of the b-tagged jets

- $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$ in the high p_{T}^{W} region (see section 4.4)
- $H_{\mathrm{T}} > 160\,\mathrm{GeV}$ (180 GeV in the WH analysis) in the low p_{T}^{W} region (see section 4.4)
- $\Delta R(b,b) > 0.7$ for $p_{\mathrm{T}}^{W} < 200\,\mathrm{GeV}$

The quantity H_{T} is the scalar sum of the transverse momenta of the reconstructed objects. The requirements on the minimum values on H_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ are mainly motivated by multijet reduction while the cut on $\Delta R(b,b)$ is effective against W+jets background. The lowering of the cuts on H_{T} and jet p_{T} with respect to the WH analysis reflects the lower mass of the Z boson in comparison to the Higgs boson.

4.4 Event categories

To improve the sensitivity of the measurement, the events are split in categories. The categorization is done according to the number of jets (two or three) and the transverse momentum of the W boson:

- Low p_{T}^{W} defined as $p_{\mathrm{T}}^{W} < 120\,\mathrm{GeV}$
- High p_{T}^{W} defined as $p_{\mathrm{T}}^{W} \geq 120\,\mathrm{GeV}$

This gives four categories:

- Low p_{T}^{W} 2jet
- Low p_{T}^{W} 3jet
- High p_{T}^{W} 2jet
- High p_{T}^{W} 3jet

The low p_{T}^{W} categories profit from a much larger number of events but suffer from a non-negligible contribution of multijet, whereas for $p_{\mathrm{T}}^{W} \geq 120\,\mathrm{GeV}$, multijet is highly suppressed. The three jet region is dominated by $t\bar{t}$-pair production and has a poor signal-to-background ratio. However, it is very useful to constrain the $t\bar{t}$ normalization in the fit (see chapter 6).

CHAPTER 5

BDT training

5.1 The concept of BDTs

A boosted decision tree (BDT) is a multivariate classification method based on machine learning. Such techniques are used in high-energy physics in order to search for a small signal in a large data set. Event kinematics in a collider experiment can be described by several variables, such as transverse momenta and invariant masses. Each variable follows a certain distribution which in general is different for signal and background. Based on a simulated training sample, the BDT exploits such differences and learns

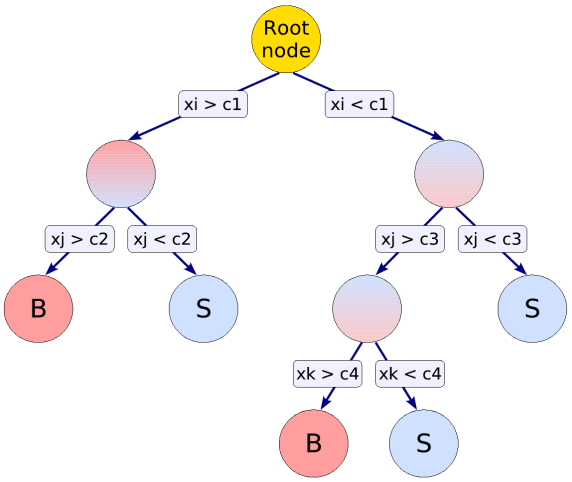

Figure 5.1: Schematic view of a decision tree. The quantities x_i represent variables which are being cut on at values c_j. [14]

to discriminate between signal and background. The result is a classifier that can decide if an event is more signal-like or more background-like.

The working principle of decision trees is sketched in figure 5.1: The root node represents the full training sample. Starting from here, a sequence of binary decisions splits the sample into smaller and smaller sub-samples. The decisions are made by cutting on variables. In each new node of the tree, the variable and cut value that give the best discrimination at this stage are used for this purpose. The

procedure ends when the tree has reached a predefined depth. The leaf nodes at the bottom of the tree are then labeled "signal" or "background", according to the majority of events that end up in the respective node.

"Boosting" is a method to improve and refine the classification output: Instead of a single decision tree, a series of trees (a "forest") is grown based on the same training sample. Events that were misclassified in the first tree are given higher relative weights and thus gain higher importance in the next one, and so on. The most common boosting algorithm, also used in this analysis, is called "AdaBoost" [15] (short for adaptive boosting). Each of the trees puts out -1 for background or $+1$ for signal. All binary classifications are then combined into one classifier by a weighted mean. It has a range from -1 to $+1$. By requiring events to surpass a certain minimum value, the selection is narrowed down to more signal-like events, improving the expected signal significance (see section 4.1).

The training of a BDT is a way of deriving a model from a training sample. A problem that can occur in such techniques is "overtraining": If the model has many degrees of freedom and the size of the training sample is not sufficiently large, it can happen that the training procedure will take statistical fluctuations for characteristic properties. In the extreme case, the model will perfectly mimic every bit of the training sample. For a BDT the number of degrees of freedom depends on the number and depths of the trees. If too many different variables are made available to BDT, this can amplify the effect of overtraining, since each additional variable increases the number of possible decision trees, allowing the model to fit the data even more exactly.

To achieve better stability against small changes in the training sample and to avoid overtraining it is preferable to keep the single tree rather shallow, with only a few consecutive splits, and to go for a higher number of trees instead [4].

5.2 BDT configuration

This analysis uses the TMVA package [16] for the training of the BDT. The configuration is as follows:

- Boost method: AdaBoost (see section 5.1)
- Maximal tree depth: 4 splits
- Number of trees: 400

5.3 Splitting of the training sample

The same simulated events that were used in the BDT training should not be used for a comparison of data and simulation later in the analysis since possible overtraining could lead to a bias. Therefore, the simulated samples are split in two halves, and one BDT classifier is trained independently on each half. When the BDT is applied in the analysis, each event is evaluated by the *respective other* classifier.

5.4 Choice of input variables

The BDT has to be supplied with a sinstead uitable set of input variables. The most important requirement is a good separation of signal and background. The set of variables should fully exploit the available information and, in view of overtraining, not contain more variables than necessary. Furthermore, it makes sense to discard a variable if one knows it to be badly modeled. After the preselection (see section 4.3), the BDT classifier distribution is used to select signal-like events. The cut value is

Baseline	Baseline WZ	Lorentz inv.
$m(bb)$	$m(bb)$	$p^{b_1} \cdot p^{b_2}$
$\Delta R(b,b)$	$\Delta R(b,b)$	$p^{b_1} \cdot p^{l}$
$\lvert\Delta\eta(b,b)\rvert$	$\lvert\Delta\eta(b,b)\rvert$	$p^{b_2} \cdot p^{l}$
$p_{\mathrm{T}}^{b_1}$, $p_{\mathrm{T}}^{b_2}$	$p_{\mathrm{T}}^{b_1}$, $p_{\mathrm{T}}^{b_2}$	$p^{b_1} \cdot p^{\nu}$
p_{T}^{l}	–	$p^{b_2} \cdot p^{\nu}$
$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss}}$	$p^{l} \cdot p^{\nu}$
p_{T}^{W}	p_{T}^{W}	θ_1
H_{T}	H_{T}	θ_2
m_{T}^{W}	m_{T}^{W}	γ^z
$\lvert\Delta\phi(W,bb)\rvert$	$\lvert\Delta\phi(W,bb)\rvert$	
$\lvert\Delta\phi(l,E_{\mathrm{T}}^{\mathrm{miss}})\rvert$	$\lvert\Delta\phi(l,E_{\mathrm{T}}^{\mathrm{miss}})\rvert$	
$\min[\Delta\phi(l,b)]$	$\min[\Delta\phi(l,b)]$	
–	$\Delta\eta(b_1,l)$	
MV1c(b_1)	MV1c(b_1)	MV1c(b_1)
MV1c(b_2)	MV1c(b_2)	MV1c(b_2)

Table 5.1: The different sets of discriminating variables used as BDT input.

chosen such that the expected significance as defined in equation 4.2 is maximized. Different sets of input variables were tested, comparing their performance as measured by the maximum significance.

Three chosen sets of variables are shown in table 5.1. The so-called Baseline variables are the standard choice in the ATLAS $H \to b\bar{b}$ analysis. They include the transverse momenta $p_{\mathrm{T}}^{b_1}$,$p_{\mathrm{T}}^{b_2}$, p_{T}^{l} and p_{T}^{W} of the b-tagged jets (p_{T}-ordered), the lepton and the W boson, the absolute value $E_{\mathrm{T}}^{\mathrm{miss}}$ of the missing transverse energy and angular distances between different objects[1]. Furthermore, the invariant mass $m(bb)$ of the dijet system, and the transverse mass of the vector boson m_{T}^{W} are included. The quantity H_{T} is the sum of the transverse momenta of the reconstructed objects. The most powerful variable in this analysis is $m(bb)$ as it shows a resonance at the mass of the Z boson for the signal sample.

The "Lorentz-invariant" variables [4] are designed to describe the physical system in a way which is correlation-free by construction and independent of the transverse boost. They are products of the four-momenta p^{b_1}, p^{b_2}, p^{l} and p^{ν} of the b-tagged jets, the charged lepton and the neutrino. In addition to the Lorentz-invariants, this set of variables includes the angle θ_1 between the $b\bar{b}$-system and the beam and the angle θ_2 between the plane defined by the $b\bar{b}$-system and the beam-line on the one hand and the charged lepton on the other hand. Furthermore, the longitudinal boost γ^z of the system is included.

Figure 5.2 shows the BDT peformance for the different sets of variables. As can be read off from the plot, the Lorentz-invariant variables (green) with a maximum significance of 3.98 perform slightly worse than the Baseline variables (red) with a maximum significance of 4.18. The loss of separation power is presumably due to the non-consideration of the transverse boost. The Lorentz-invariant variables were therefore not further considered in this analysis[2]

[1] $\min[\Delta\phi(l,b)]$ means the distance in ϕ between the lepton and the closest b-tagged jet

[2] However, in the light of systematic uncertainties there are still reasons to work with the Lorentz-invariant variables [4].

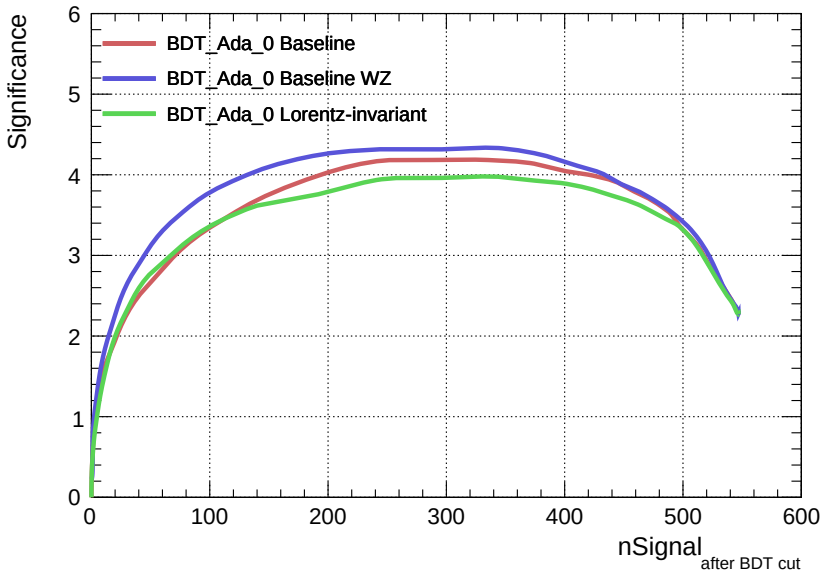

Figure 5.2: The statistical significance as measured by formula 4.1 plotted against the number of signal events after cut on the classifier output while varying the cut value, comparing different sets of discriminating variables (80% b-tagging working point)

The training was redone several times testing different variables as addition to the Baseline variables to investigate whether the expected significance could be enhanced any further. Likewise, it was tested whether one or more variables could be omitted without losing performance. The result of this study is a slightly modified set of variables, called Baseline WZ: The difference in pseudorapidity $\Delta\eta(b_1, l)$ between the first b-tagged jet and the lepton is added; in exchange, the transverse momentum p_{T}^{l} of the lepton is removed from the selection of variables. The linear correlation matrix of the Baseline WZ variables is shown in figure A.1 and A.2. Note that $\Delta\eta(b_1, l)$ appears to be uncorrelated to any other variable, which proves that it indeed delivers additional information. The separation of signal and background for the Baseline WZ variables is shown in figure 5.4. The separation power of the additional variable $\Delta\eta(b_1, l)$ can be nicely seen. Figure 5.3 illustrates that, by contrast, the variable p_{T}^{l} gives practically no separation of signal and background. A comparison of data and simulation for all variables is shown in figure A.3, A.4, A.5 and A.6. The BDT trained on the basis of the Baseline WZ variables achieves a maximum significance of 4.33.

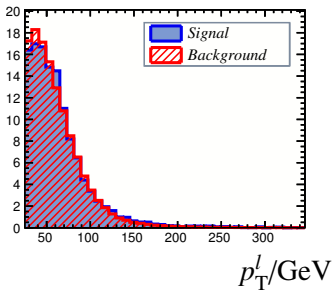

Figure 5.3: Signal-background comparison for the transverse momentum p_{T}^{l} of the lepton.

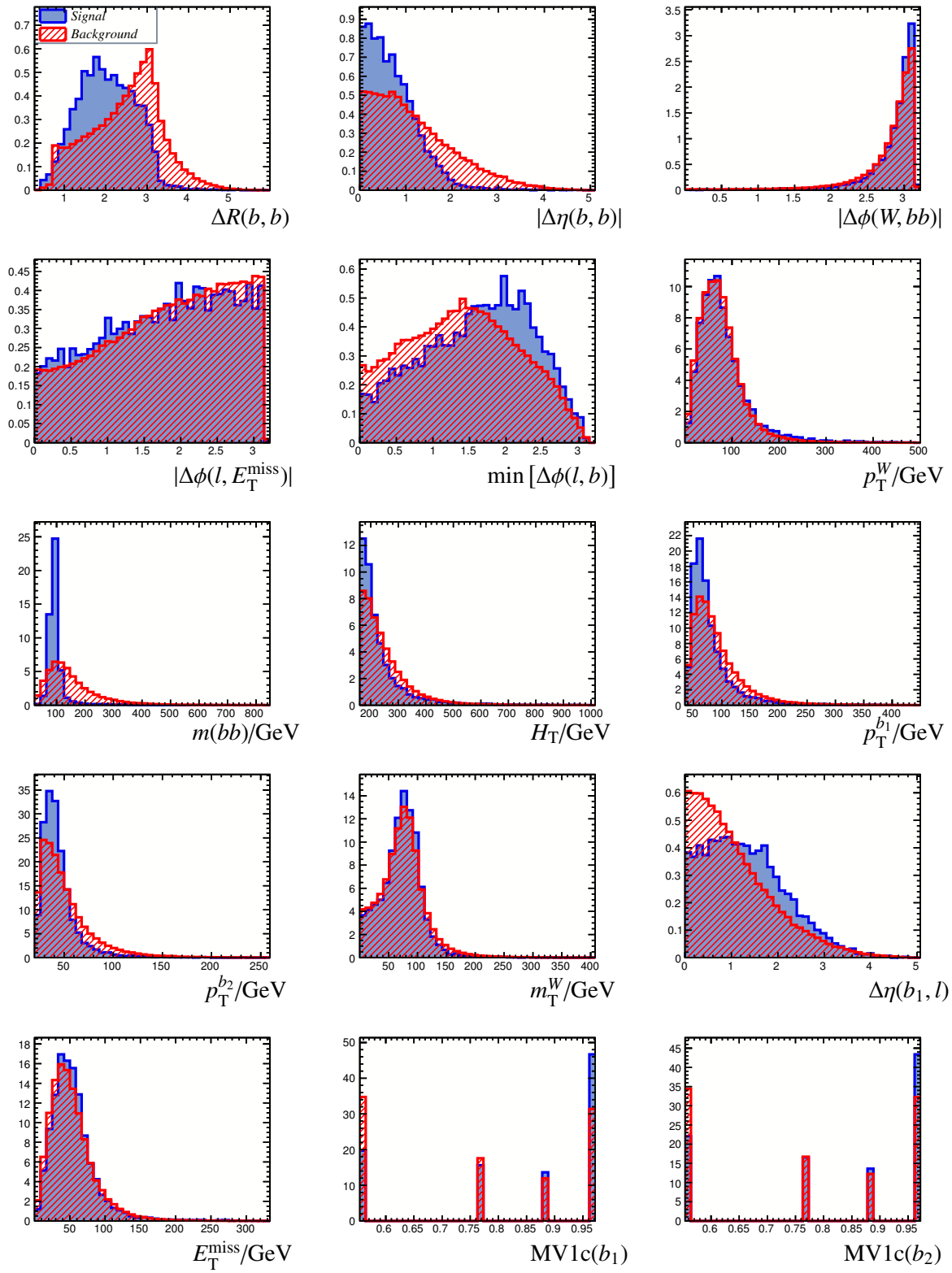

Figure 5.4: Signal-background comparison for the Baseline *WZ* variables with signal and background normalized to the same arbitrary value.

5.5 Comparison of *b*-tagging working points

Two different working points for *b*-tagging (see section 3.3) were tested using the *WZ* Baseline variables:

1. MV1c> 0.405, equivalent to 80% efficiency (expected event yield: $S = 547$, $B = 57850$)
2. MV1c> 0.703, equivalent to 70% efficiency (expected event yield: $S = 343$, $B = 24624$)

The lower MV1c cut gives better *b*-tagging efficiency but leads to a higher percentage of misidentified *c*-jets and light jets. To compensate this disadvantage, the MV1c weights MV1c(b_1) and MV1c(b_2) for both *b*-tagged jets are used in the BDT training in addition to the other discriminating variables at the 80% working point. The MV1c distribution is split into four bins between 0.405 and 1.000 (see figure 5.4).

The comparison of the significance curves clearly shows that the 80% working point with the use of the MV1c weights in the BDT training gives the better result. Therefore, the 80% working point was used in the further analysis.

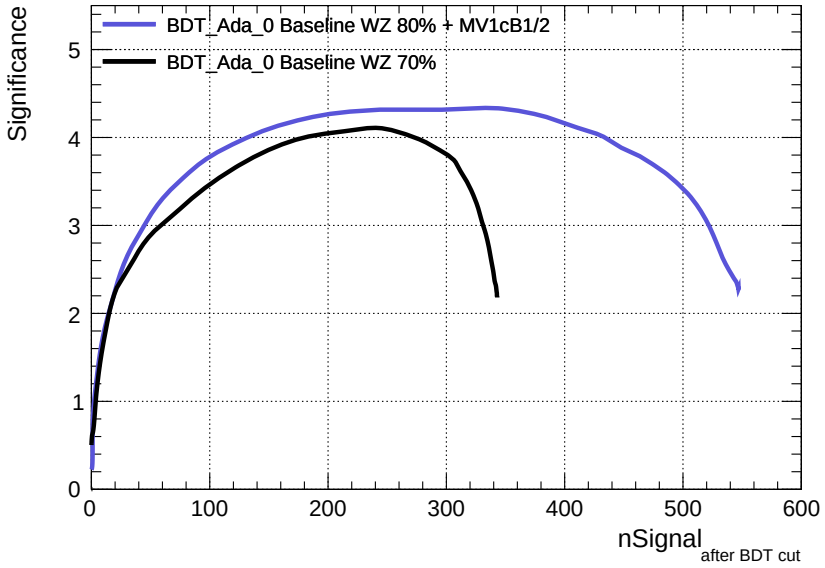

Figure 5.5: The statistical significance for two different *b*-tagging working points

5.6 BDT classifier distribution

The Baseline *WZ* variables are used to produce the final classifier for the cross section measurement. The distribution is shown in figure 5.6 with signal and background normalized to the same arbitrary value. The optimal cut value is found to be 0.13. The diagram also compares training and test sample: The training sample shows a slightly stronger separation of signal and background than the test sample, which indicates overtraining. However, the splitting of the training sample described in section 5.3 prevents a bias from overtraining.

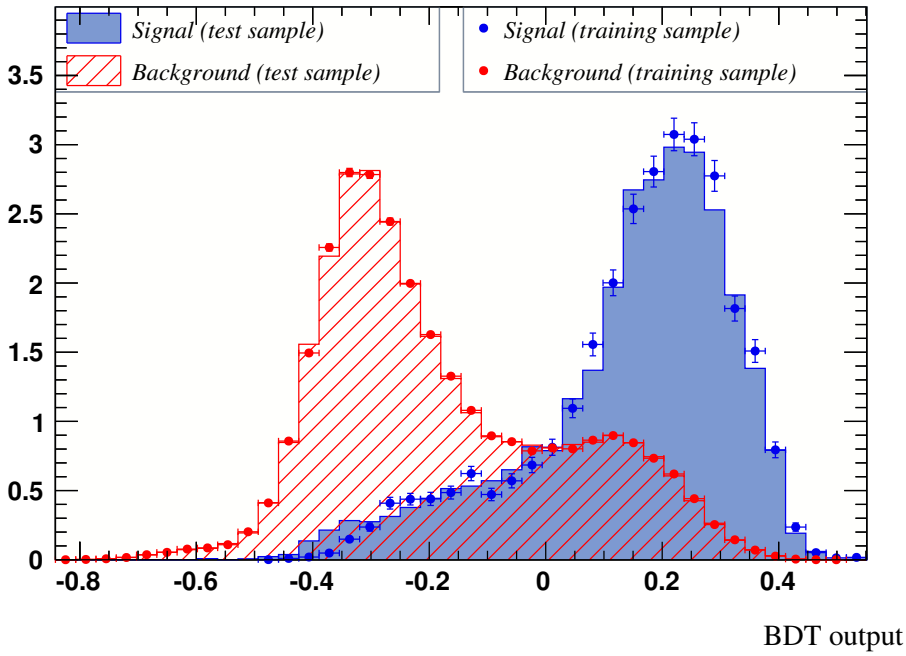

Figure 5.6: The BDT classifier distribution shown separately for signal and background simulation (normalized to the same value) and for training and test sample.

CHAPTER 6

Normalization of the backgrounds

6.1 Motivation

This chapter discusses how the normalizations of the different backgrounds are corrected. This is of particular importance for multijet production. The shape of the multijet distributions is determined from data (see section 2.2) but the normalization is a priori not known. Separate multijet templates are available for the electron and muon channel, which must be normalized independently. The other background processes are modeled using Monte Carlo generators, but the cross sections bear large uncertainties in the particular phase space of this analysis. This is further discussed in section 7.2.

The normalizations of the most important backgrounds are corrected based on data. This is achieved with a global fit that combines information from multiple control regions (see section 6.2).

6.2 Description of the global fit

The fit is performed in control regions, i.e. regions of the phase space that are signal-depleted while kinematically similar to the signal region. Ideally, each control region is dominated by a different kind of background. If the latter requirement is fulfilled, the respective background can be normalized with better accuracy, because in that case it is nearly equivalent to the total background in that region and can be directly compared with data. In this analysis, however, it is difficult to single out individual backgrounds and one will have to live with several different backgrounds in some of the control regions.

The input of the fit is a selection of statistically uncorrelated[1] histograms. These are created from kinematic variables selected among those that were used in the BDT (see section 5.4). The range in each variable is chosen such that the signal contribution is negligible. The set of input histograms should be chosen such that it gives the best agreement between data and simulation. The choice of input histograms is further discussed in section 6.4. The fit minimizes a global χ^2 that is built from all bins of all histograms. The normalizations of several backgrounds are free parameters in the fit. Backgrounds that only contribute little are hardly constraint by the fit and their normalizations are fixed to the theoretical value. This is explained in section 6.3.

To take account of kinematic differences between low p_{T}^{W} and high p_{T}^{W}, the scale factors a determined independently for each of the two regions.

[1] containing disjoint sets of events

6.3 Backgrounds to be rescaled

The least significant backgrounds, namely WH/ZH, diboson and Z+jets, are fixed to their theoretical values, since inaccuracies in their normalizations will affect the cross section measurement only insignificantly. This leaves single top, $t\bar{t}$, W+jets and multijet. The W+cc part of the W+jets is also small and is therefore fixed to the theoretical value. The multijet normalization must in any case be allowed to float in the fit (see section 6.1).

As it turns out, the fit is not stable with this many free parameters, in the sense that small changes in the input histograms lead to a largely different result or that one of the scale factors becomes unreasonably large ($s.f. \gg 1$) or small ($s.f. \approx 0$). The $t\bar{t}$ background can be normalized accurately since it dominates the 3jet control regions (see figure 6.1 and 6.2), but the relative scaling of the different W+jets and single top backgrounds is problematic. As a compromise, the single top samples are also fixed to their theoretical value, leaving the following six backgrounds:

- multijet electron channel
- multijet muon channel
- W+light
- W+cl
- W+b
- $t\bar{t}$

Since the contribution of the multijet muon channel in the high p_{T}^{W} region is tiny, the scale factor for this background is adopted from low p_{T}^{W} and kept fixed in the fit. With this smaller selection of backgrounds, a satisfying level of stability is achieved. The uncertainty on the cross section due to normalization errors is discussed in section 7.2.

6.4 Fit input

6.4.1 Control regions

A problem that arises with a control-region-based fit is how to extrapolate the obtained scale factors to the signal region. Since the kinematics of signal and control region are never perfectly equivalent, applying the same factors to the signal region, which is the method followed in this analysis, induces systematic uncertainties. However, this type of uncertainty is not discussed in the scope of this thesis. The strategy is therefore to, if possible, restrict the choice of input histograms to control regions which are kinematically largely similar to the signal region so that the presumed extrapolation error is small.

A simple way of obtaining possible control regions which are statistically uncorrelated by construction is to consider events with no or only one b-tagged jet instead of two while keeping the other preselection cuts (see section 4.3). The categories will be referred to as 0tag, 1tag and 2tag. A signal-depleted subcategory of the 2tag region is a good control region because the background composition depends heavily on the number of b-tags. Such a subcategory can be defined by setting an upper limit on the BDT classifier output. The value BDT output = 0.00 is used for this purpose. Since the 1tag2jet region has also a non-negligible signal contribution, the same upper limit is applied here. This defines six control regions which are listed after b-tag and jet multiplicity in table 6.1.

N(b-tags)	N(jets)	Description	Used in fit
0tag	2jet	Largely dominated by W+light.	No
	3jet	Largely dominated by W+light.	No
1tag	2jet	(BDT<0.00) Largest contributions from W+light and W+cl.	Yes
	3jet	Largest contributions from $t\bar{t}$, W+light and W+cl.	No
2tag	2jet	(BDT<0.00) Largest contributions from $t\bar{t}$, W+b	Yes
	3jet	(BDT<0.00) Largely dominated by $t\bar{t}$.	Yes

Table 6.1: The available control regions, sorted by b-tag and jet multiplicity.

The uncertainty on the total background is not significantly reduced by the 0tag regions. Since the background composition is very different and dominated by components which hardly matter in the signal region, this control region is not used in the fit. Since the jet multiplicity in W+jets has a large uncertainty and this background is an important component in the 1tag3jet region, this region is also not considered in the fit.

6.4.2 Input histograms

In order to reliably normalize the electron and muon samples of the multijet background, input histograms need to be split by electron and muon channel.

The variable that is used for the fit input histograms should ideally show differently shaped distributions for each kind of background in order to constrain the scale factors in an unambiguous way. There is, however, no single variable that is clearly superior than others in this respect. Furthermore, the variable should not feature significant mismodeling. This requirement makes the BDT classifier output a good choice since it combines well and not so well modeled variables. The resulting level of mismodeling is reduced compared to the mismodeling seen in some of the input variables.

While the fits works well with the BDT output as the only variable in the high p_{T}^{W} region, an additional refinement is made for low p_{T}^{W}: The control regions are subdivided by a cut on the BDT output at -0.25, creating the subcategories BDT[-0.60,-0.25] and BDT[-0.25,0.00]. Using the BDT output as fit variable in the first category and $m(bb)$ in the latter further improves the fit stability. The final choice of input histograms is listed in table 6.2; the corresponding plots are shown in figure 6.1, 6.2, 6.3, 6.4 and 6.5.

control region	low p_{T}^{W}	high p_{T}^{W}
1tag2jet (1T2J)	BDT output [-0.70,0.00] mu.	BDT output [-0.70,0.00] el. BDT output [-0.70,0.00] mu
2tag2jet (2T2J)	BDT output [-0.6,-0.25] el. BDT output [-0.6,-0.25] mu. $m(bb)$ for BDT[-0.25,0.00] el. $m(bb)$ for BDT[-0.25,0.00] mu.	BDT output [-0.5,0.00] el. BDT output [-0.5,0.00] mu.
2tag3jet (2T3J)	BDT output [-0.6,-0.25] el. BDT output [-0.6,-0.25] mu. $m(bb)$ for BDT[-0.25,0.00] el. $m(bb)$ for BDT[-0.25,0.00] mu.	BDT output [-0.5,0.00] el. BDT output [-0.5,0.00] mu.

Table 6.2: Overview of the fit input histograms, “mu.” and “el.” standing for muon and electron channel, respectively. The corresponding plots are shown in figure 6.1, 6.2, 6.3, 6.4 and 6.5.

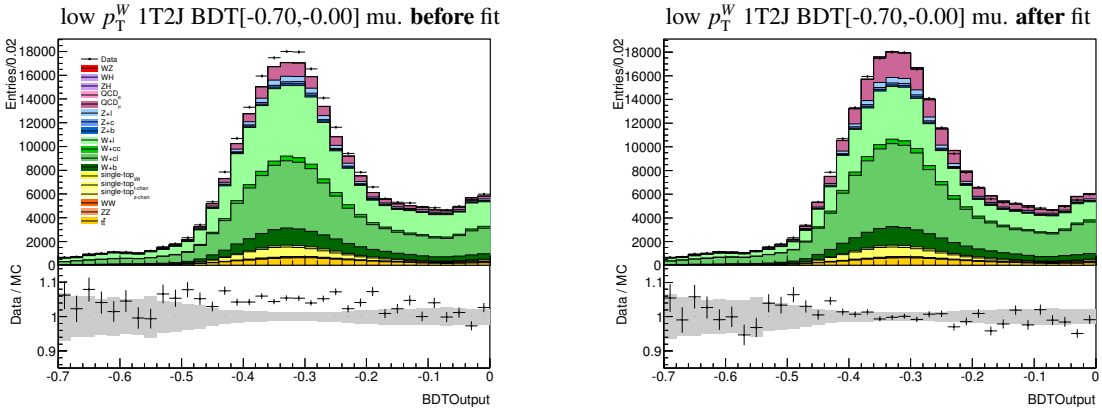

Figure 6.1: Fit input histograms for low p_{T}^{W} (statistical error on background displayed in grey), part 1

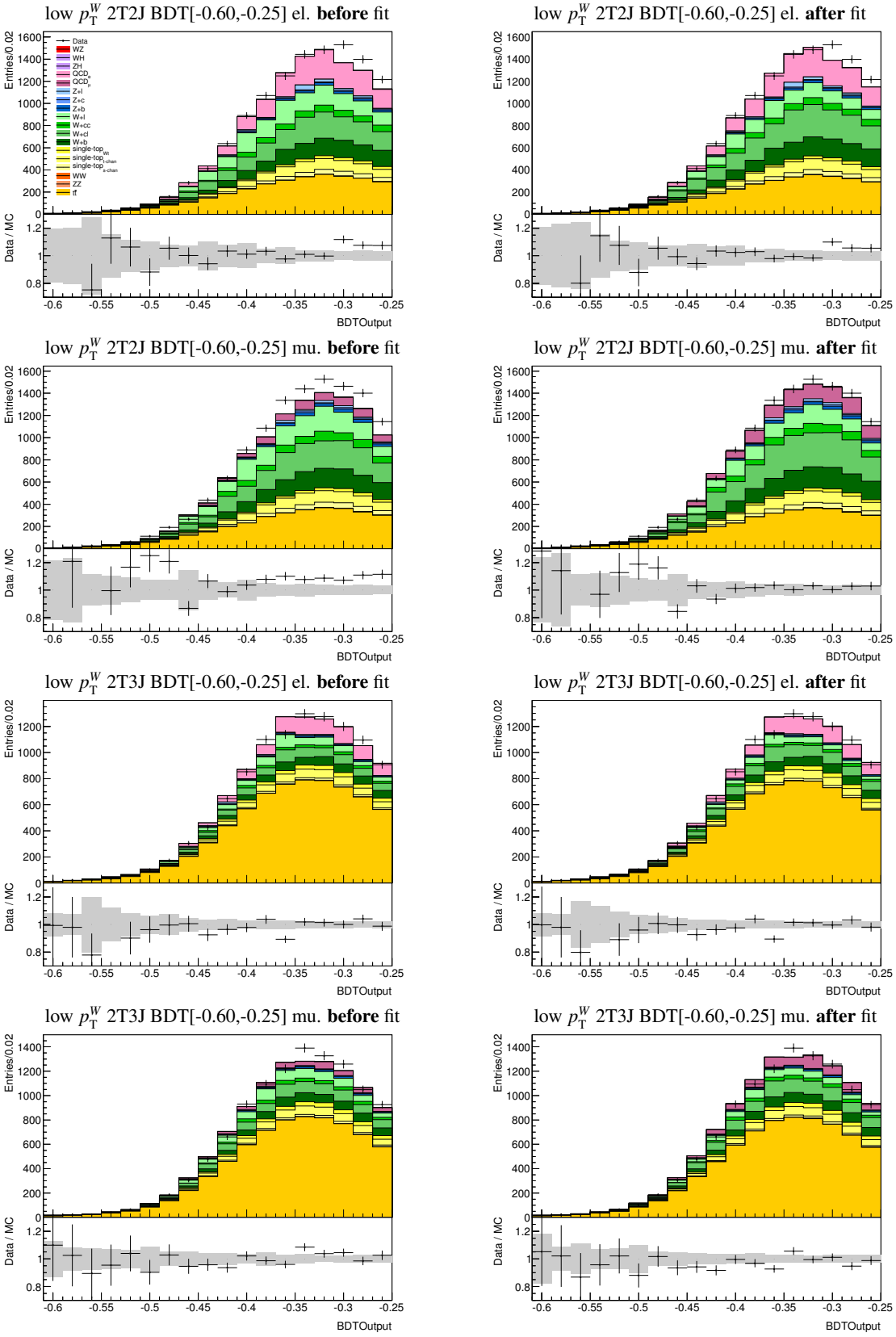

Figure 6.2: Fit input histograms for low p_T^W (statistical error on background displayed in grey), part 2

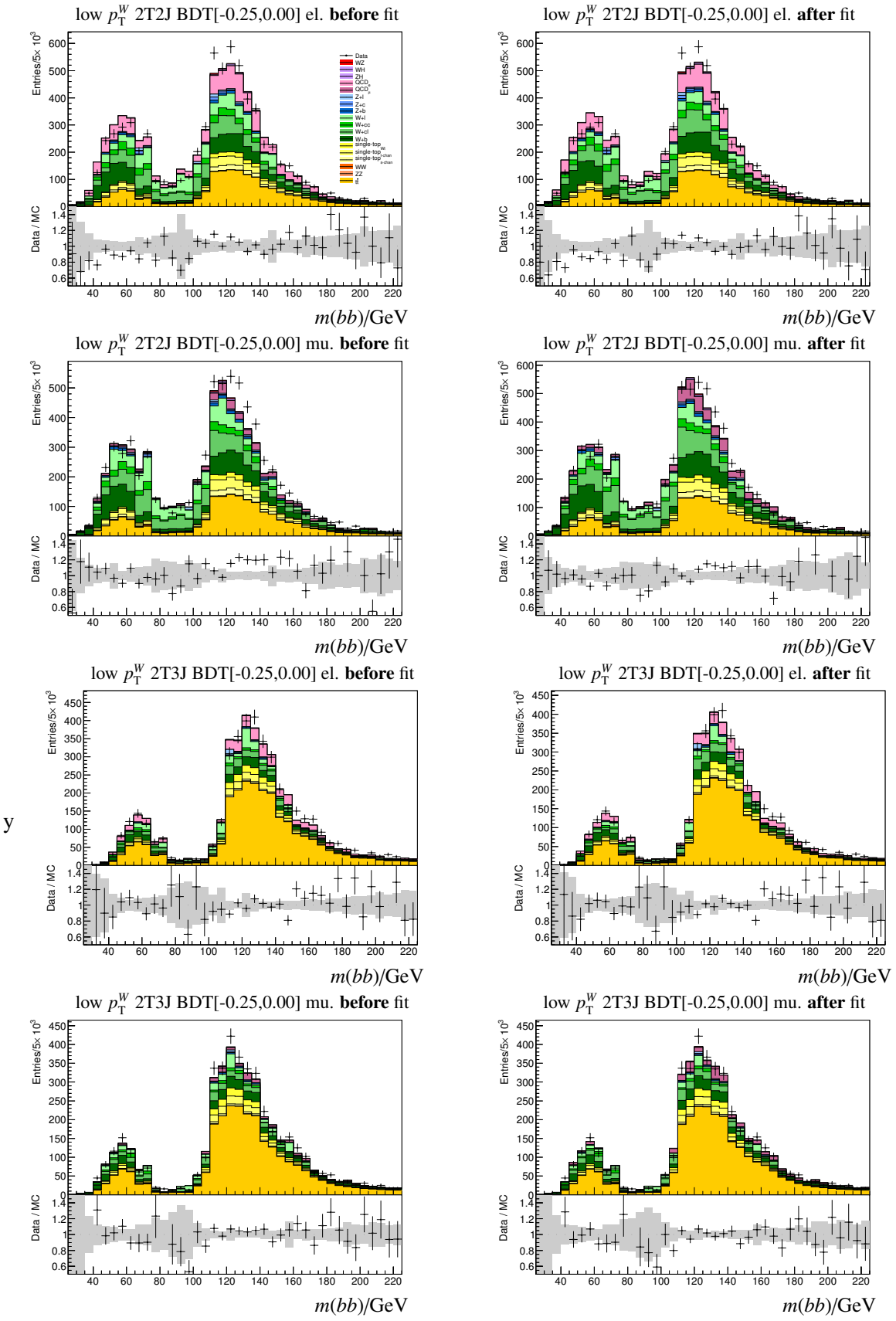

Figure 6.3: Fit input histograms for low p_T^W (statistical error on background displayed in grey), part 3

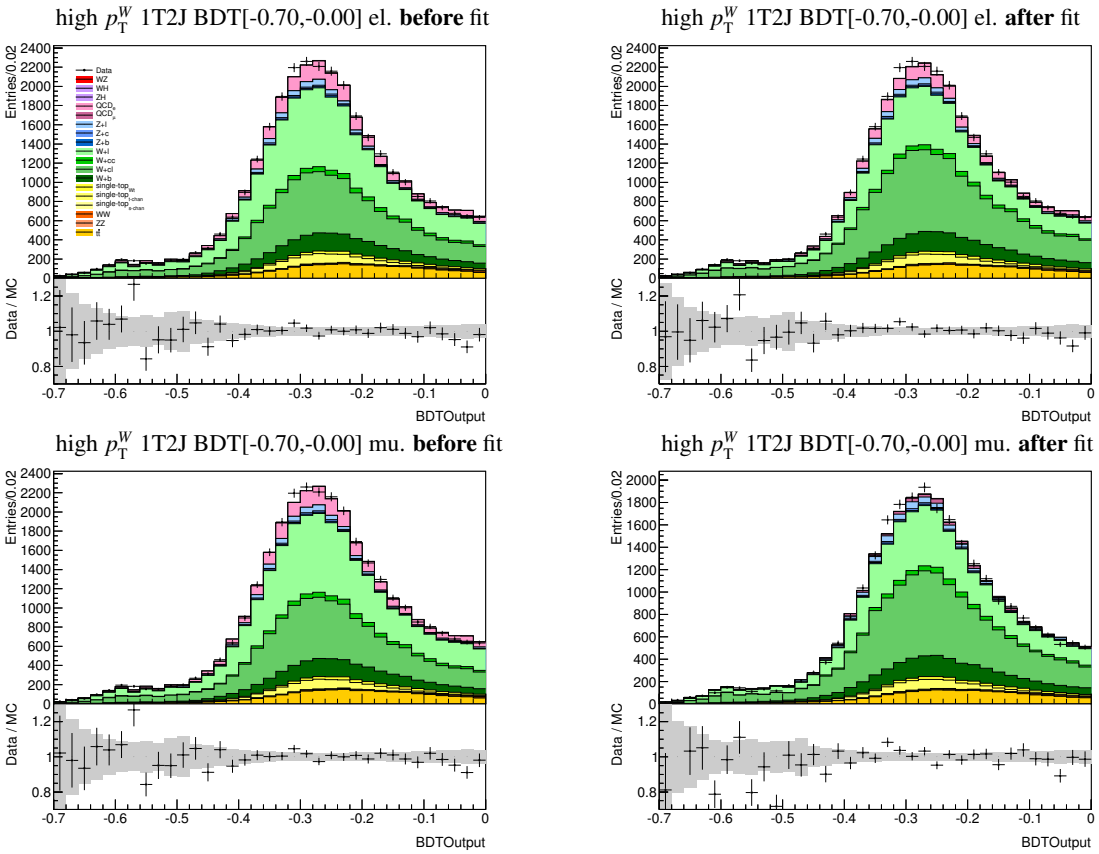

Figure 6.4: Fit input histograms for high p_T^W (statistical error on background displayed in grey), part 1

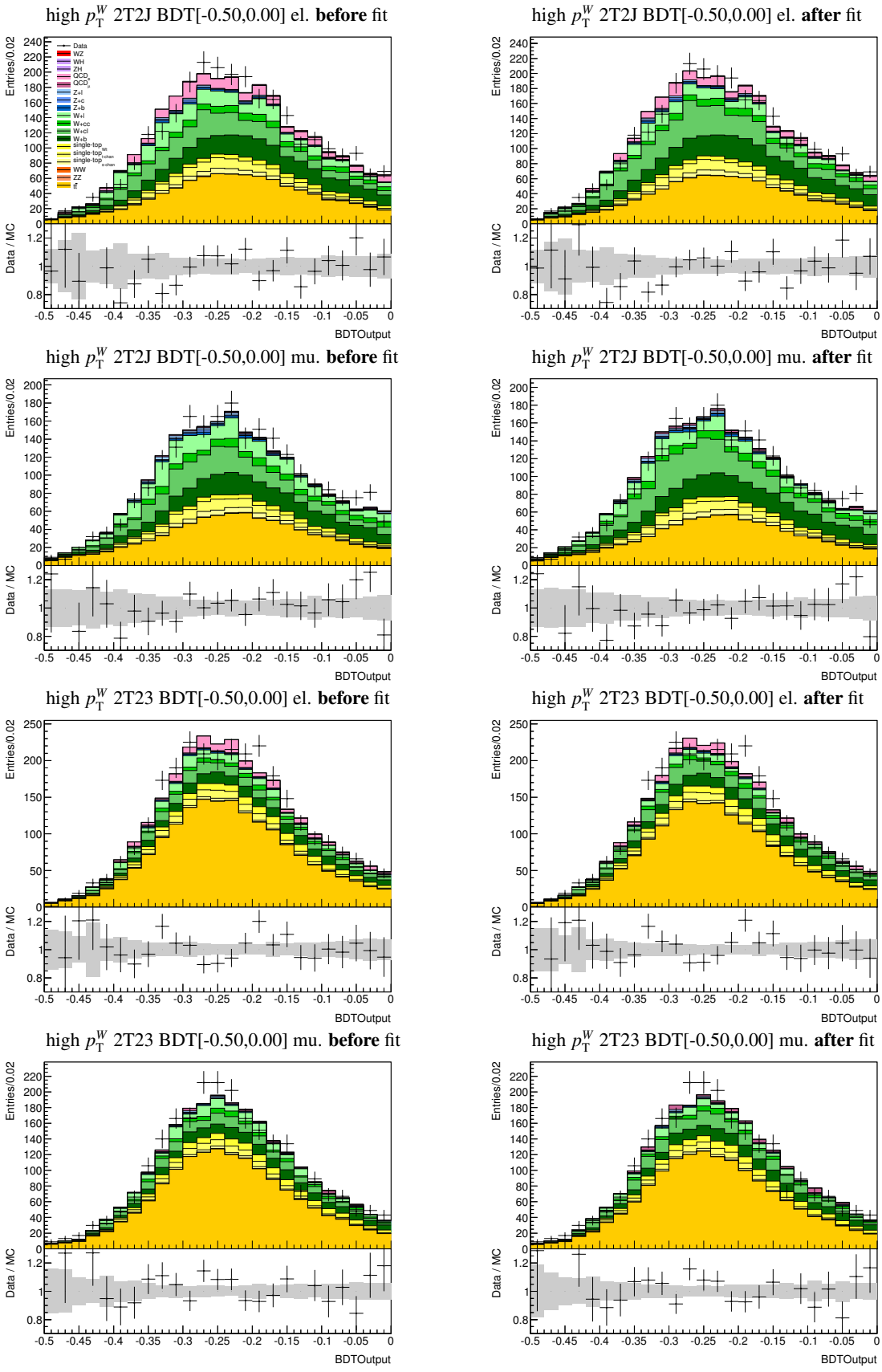

Figure 6.5: Fit input histograms for high p_T^W (statistical error on background displayed in grey), part 2

6.5 Scale factor results

The scale factor results for low p_{T}^{W} and high p_{T}^{W} are listed in table 6.3. The scale factor for the multijet muon channel in the high p_{T}^{W} region was adopted from low p_{T}^{W}. The scale factors of both regions are in very good agreement within their uncertainties, with the exception of the multijet electron channel. The scale factors for those backgrounds that are modeled with Monte Carlo simulation do not deviate very much from one, in agreement with the expectation.

Background sample	low p_{T}^{W}	high p_{T}^{W}
Multijet el.	1.00 ± 0.07	0.79 ± 0.03
Multijet mu.	1.89 ± 0.16	from low p_{T}^{W}
W+light	0.74 ± 0.03	0.74 ± 0.04
W+cl	1.24 ± 0.03	1.34 ± 0.07
W+b	1.09 ± 0.07	1.10 ± 0.10
$t\bar{t}$	0.994 ± 0.004	0.978 ± 0.014

Table 6.3: The reference scale factors. The derivation of the errors is explained in section 7.2

.

CHAPTER 7

Systematic uncertainties

7.1 Modeling of variables

A major source of systematic uncertainty on the cross section measurement is the imperfection of the Monte Carlo simulation, which may become visible in form of deviations between data and simulation in certain variables. However, if the shape of the simulated background in a control region does not fit the data, this may also be caused by the incorrect normalization of background samples. The investigation of variable mismodeling should therefore be done *after* the normalization fit (see chapter 6), assuming that remaining deviations are then mainly caused by incorrect shapes of the simulated samples. The impact of mismodeling on the cross section measurement is estimated by the following method:

1. Find variables that show significant deviations between data and simulated background. Consider only the control region defined by BDT output < 0.00 to avoid a bias on the cross section measurement. Choose a mostly uncorrelated subset of the mismodeled variables that covers the worst modeling problems.

2. Derive a systematic uncertainty for each variable by reweighting the simulated events so that the deviations in the control region are compensated. This is done by the following steps:
   a) Smooth out statistical fluctuations[1] in both the background and data histogram of the respective variable.
   b) Obtain the appropriate weight for each bin i from the control region as $N_i^{\text{data}}/N_i^{\text{sim}}$ where N_i^{data} and N_i^{sim} are the numbers of data and simulated background events in the respective bin.
   c) Apply the weights to both control region and signal region.
   d) Redo the normalization fit (see chapter 6).
   e) Calculate the difference of the reweighted and un-reweighted distribution bin-by-bin for any desired variable.

3. Sum the contributions from the different reweightings in quadrature, assuming that the reweighted variables are uncorrelated.

[1] The smoothing is done by replacing each bin content with a value obtained from averaging over the respective bin and the neighboring bins, this procedure being repeated once.

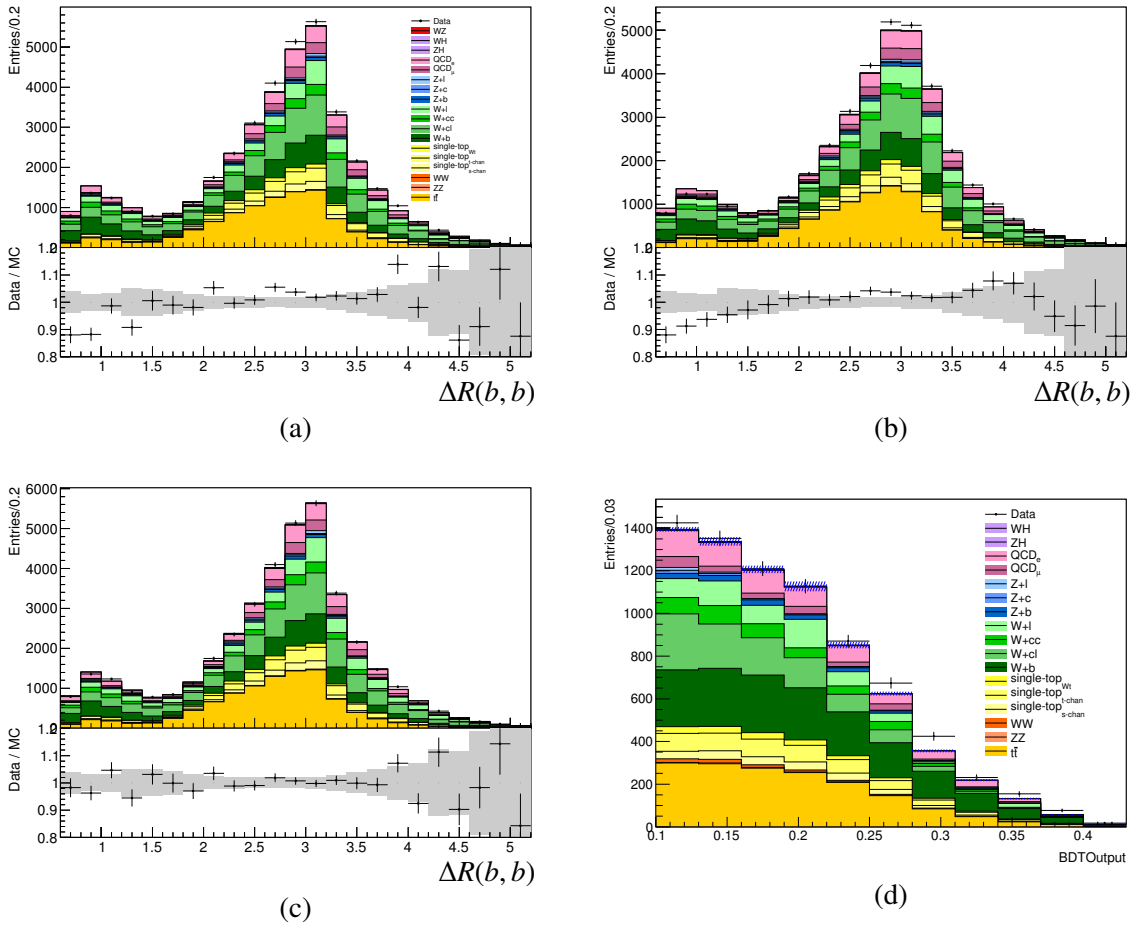

Figure 7.1: The derivation of systematic uncertainties through event reweighting using the example of $\Delta R(b,b)$ in the low p_{T}^{W} 2jet region (statistical error on background displayed in grey): The diagrams show the $\Delta R(b,b)$ distribution for BDT output < 0.00 **(a) un-reweighted** with noticeable mismodeling, **(b)** with smoothing applied to data and total background for the reduction of statistical fluctuations, **(c) reweighted** with weights obtained from the smoothed histograms and **(d)** the BDT output > 0.10 with the systematic uncertainty (hatched blue) from this particular reweighting on top of the total background (signal simulation switched off).

The reweighting procedure is illustrated in figure 7.1 using the example of $\Delta R(b,b)$. Note that the reweighting is only used for error estimation; the cross section measurement is based on the original, un-reweighted samples.

Significant deviations are mostly to be found in $\Delta R(b,b)$, m_{T}^{W} and $p_{\mathrm{T}}^{b_2}$ (see figure 7.1 and 7.2 for the low p_{T}^{W} 2jet distributions). The linear correlation (see figure A.1 and A.2) for any pair out of these variables is smaller than 30%; hence, the overestimation of the total uncertainty due to the assumption of uncorrelated variables will be not too large. Modeling uncertainties are derived in all four signal regions, 2jet/3jet and low/high p_{T}^{W}. The strongest impact of mismodeling is observed in low p_{T}^{W} 2jet. For high p_{T}^{W}, statistical fluctuations due to the smaller sample size overshadow systematic uncertainties. Nevertheless a reasonable systematic uncertainty can be derived after application of the smoothing procedure.

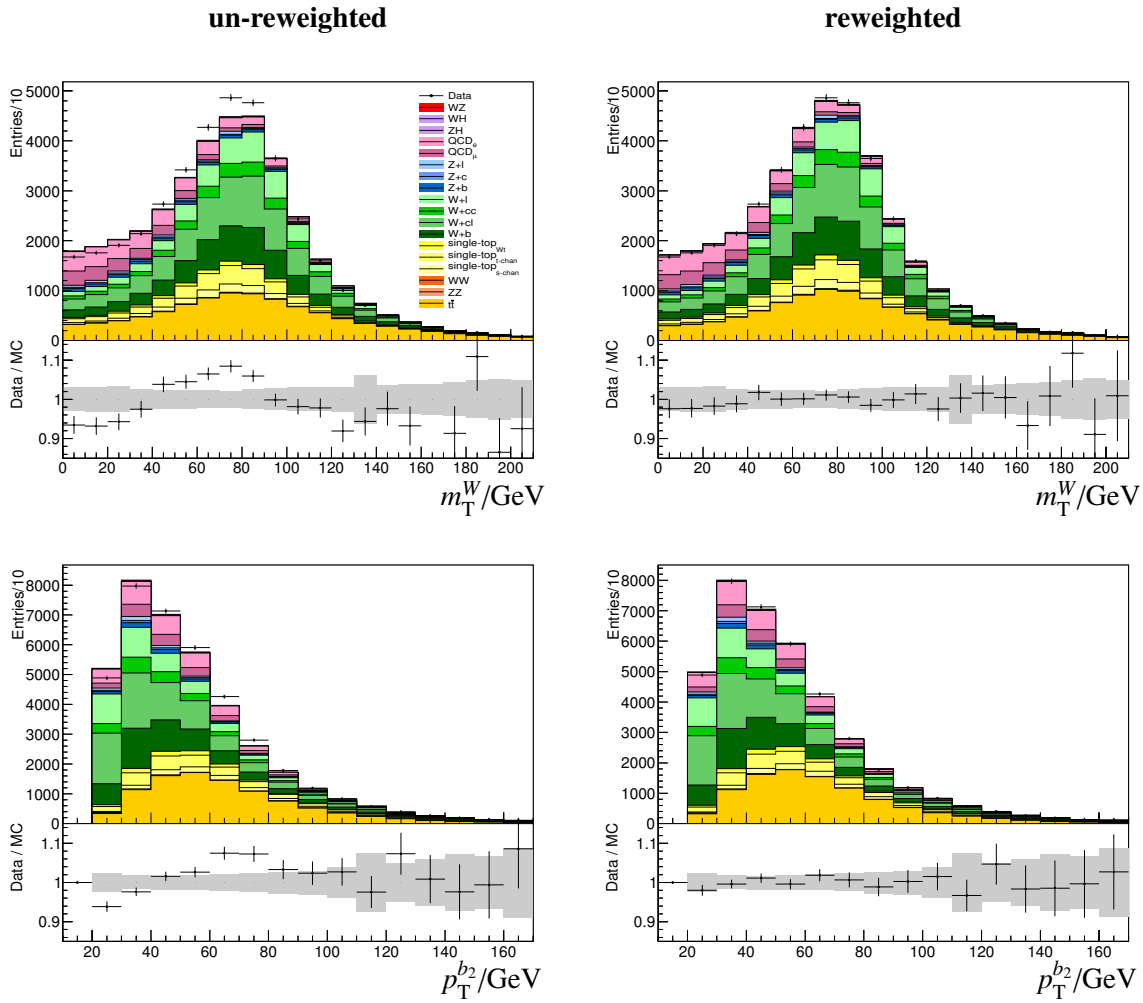

Figure 7.2: Un-reweighted and reweighted distributions for m_T^W and $p_T^{b_2}$ in the low p_T^W 2jet region where the mismodeling is most clearly visible (statistical error on background displayed in grey).

The mismodeling of the jet-related variables $\Delta R(b,b)$ and $p_T^{b_2}$ can mainly be attributed to the W+jets background samples, since the jets are radiated off via strong interaction (see figure 2.5), causing similar theoretical difficulties as QCD multijet. However, this does not explain the mismodeling of m_T^W, because this variable is not directly related to the jets. While it seems tempting to attribute the m_T^W mismodeling mainly to the imperfect reconstruction of E_T^{miss}, this is contradicted by the observation that E_T^{miss} itself is better described (see A.6). The modeling problems of m_T^W remain subject to further investigation.

7.2 Background normalization

The normalizations of the different backgrounds that were corrected in the global fit (see chapter 6) bear uncertainties. The exact outcome of the scale factors depends to a degree on which variables are used in the fit. This is partly due to statistical fluctuations, which are unique for any variable. Mismodeling of variables may also play a role. The systematic uncertainty due to mismodeling is, however, already covered by the procedure described in section 7.1; therefore a certain degree of correlation between both error components cannot be excluded.

To estimate uncertainties on the scale factors, the fit is repeated several times using different input variables. Based on the uncertainties of the scale factors, an uncertainty on the total background in the signal region is derived. The exact procedure is described in the following:

1. Define alternative sets of input histograms by choosing different variables.

2. Perform the fit for each set of input histograms. Calculate the standard deviation of each scale factor. Based on the reference values (see table 6.3), an upward variation and a downward variation is obtained for each scale factor by adding or subtracting the standard deviation.

3. For each scale factor, do the following:

   a) Fix it to the value which corresponds to the upward variation.

   b) Refit the other backgrounds and obtain the new scale factors.

   c) Do the same for the downward variation.

   d) For both upward and downward variation, observe the effect of the new scale factors on the total background in the signal region. Obtain the systematic uncertainty for any desired variable by calculating the difference of the re-fitted and the reference total background bin-by-bin and then averaging the absolute values of upward and downward variation.

4. Sum the different contributions in quadrature.

To create alternative sets of input histograms, one uses the same splitting in low p_{T}^{W} and high p_{T}^{W} and the same control regions as in the reference fit but different variables. The control regions are subdivided by a BDT cut at -0.25 so that a different variable can be used for each subcategory in the same fit. Table 7.1 shows all used combinations of input variables, the resulting scale factors and the respective standard deviations.

	Input variables	multijet el.	multijet mu.	W+light	W+cl	W+b	$t\bar{t}$
Low p_{T}^{W}	ref	1.00	1.89	0.74	1.24	1.09	0.994
	BDT & $\Delta R(b,b)$	1.01	1.77	0.73	1.29	1.04	0.988
	m_{T}^{W} & $m(bb)$	0.94	1.57	0.77	1.23	1.20	0.985
	m_{T}^{W} & $\Delta R(b,b)$	0.94	1.51	0.76	1.27	1.16	0.980
	$E_{\mathrm{T}}^{\mathrm{miss}}$ & $m(bb)$	1.08	1.86	0.80	1.20	1.05	0.988
	$E_{\mathrm{T}}^{\mathrm{miss}}$ & $\Delta R(b,b)$	1.09	1.79	0.79	1.23	1.00	0.985
	std. dev.	0.07	0.16	0.03	0.03	0.07	0.004
High p_{T}^{W}	ref	0.79	—	0.74	1.34	1.10	0.978
	BDT & $\Delta R(b,b)$	0.80	—	0.70	1.46	0.80	1.018
	m_{T}^{W} & $m(bb)$	0.76	—	0.69	1.47	0.88	1.014
	m_{T}^{W} & $\Delta R(b,b)$	0.77	—	0.66	1.53	0.79	1.018
	$E_{\mathrm{T}}^{\mathrm{miss}}$ & $m(bb)$	0.82	—	0.76	1.35	0.93	1.011
	$E_{\mathrm{T}}^{\mathrm{miss}}$ & $\Delta R(b,b)$	0.83	—	0.73	1.41	0.84	1.016
	std. dev.	0.03	—	0.04	0.07	0.10	0.014

Table 7.1: Scale factors for different combinations of fit input variables and the respective standard deviations. The notation m_{T}^{W} & $m(bb)$ means that m_{T}^{W} was used for BDT output < -0.25 and $m(bb)$ for BDT output > -0.25.

7.3 *b*-tagging efficiency

Uncertainties on the b-tagging efficiency and on the rejection efficiency for c-jets and light jets are taken into account. Systematic variations are performed by a reweighting of events in the different MV1c bins, depending on the jet p_{T}. From this, 10 significant uncertainty components are derived for b-jets, 15 for c-jets and 10 for light jets, with an upward and a downward variation for each. These uncertainties are adopted from the ATLAS $H \to b\bar{b}$ analysis [2].

Depending on the reweighting, more or less jets of the different types will be b-tagged, which corresponds to a higher or lower b-tagging efficiency, and the total background in the signal region will change accordingly. The uncertainty on the total background is derived for each component by calculating the difference of the reweighted and the unreweighted distribution bin-by-bin and then averaging the absolute values of upward and downward variation.

CHAPTER 8

Cross section measurement

The cross section is measured based on the BDT classifier distribution above a cut value of 0.13 by the method described in section 4.2. This is done separately in each of the four event categories (see section 4.4). The distributions are shown in figure 8.1 and 8.2. The results are presented in table 8.1. The uncertainty on the cross section is divided into five components:

1. the statistical error on data
2. the statistical error on the simulated background (acting as a systematic uncertainty)
3. the systematic uncertainty from the modeling of variables (see section 7.1)
4. the systematic uncertainty from the background normalizations (see section 7.2)
5. the systematic uncertainty from b-tagging efficiency (see section 7.3)

The statistical error on the simulated background is to be considered a systematic uncertainty, since it is an uncertainty of the theoretical model. It can be reduced by generating more events.

In the most sensitive category, low p_{T}^{W} 2jet, the measured signal strength amounts to about half the Standard Model expectation. In the second most sensitive category, high p_{T}^{W} 2jet, no clear excess over background is observed. The 3jet categories are not reliable for a standalone measurement since a possible excess over background is fully overshadowed by statistical fluctuations; they nevertheless contribute to the overall measurement precision when the different categories are combined in a weighted mean.

A comparison of the different systematic uncertainties shows that the statistical error on the simulated background (2.) considerably deteriorates the measurement precision. Therefore, more Monte Carlo events should be produced. The b-tagging uncertainty (5.) has the second-largest effect, reflecting the difficulty of a reliable jet-flavor-identification. The modeling (3.) uncertainty, which concerns the shape of the simulated samples, has also a non-negligible influence. The main problem here is presumably the insufficient description of W+jets background. The influence of the normalization uncertainty (4.) is comparatively small. This can partly be explained by the fact that the different backgrounds are largely similar in shape and can compensate each other to a degree without significantly changing the total background.

The results of the four categories are combined into one final value by a weighted mean. The different error components are treated as uncorrelated, neglecting that a certain degree of correlation is to be assumed between modeling and normalization uncertainty (see section 7.2). Correlations of specific

Event category	$\left(\frac{S}{\sqrt{B}}\right)_{\text{sim}}$	σ_{WZ}/pb					
Low p_{T}^{W} 2jet	4.5	13	$\pm(5)_{\text{stat}}$	$\pm(6)_{\text{stat,bg}}$	$\pm(3)_{\text{mod}}$	$\pm(2)_{\text{norm}}$	$\pm(4)_{b\text{-tag}}$
Low p_{T}^{W} 3jet	1.2	11	$\pm(18)_{\text{stat}}$	$\pm(11)_{\text{stat,bg}}$	$\pm(3)_{\text{mod}}$		
High p_{T}^{W} 2jet	2.5	2	$\pm(9)_{\text{stat}}$	$\pm(4)_{\text{stat,bg}}$	$\pm(2)_{\text{mod}}$	$\pm(1)_{\text{norm}}$	
High p_{T}^{W} 3jet	0.9	57	$\pm(25)_{\text{stat}}$	$\pm(8)_{\text{stat,bg}}$	$\pm(3)_{\text{mod}}$		
Combined		11	$\pm(5)_{\text{stat}}$	$\pm(3)_{\text{stat,bg}}$	$\pm(1)_{\text{mod}}$	$\pm(1)_{\text{norm}}$	$\pm(4)_{b\text{-tag}}$

Table 8.1: The cross section measured in four different event categories with five error components: the statistical error on data, the statistical error on the simulated background, the systematic uncertainty from the modeling of variables, the systematic uncertainty from background normalization and the systematic uncertainty from the b-tagging efficiency. The second column shows the significance one expects from simulation.

error components across the four categories are taken into account: The normalization uncertainty is correlated between 2jet and 3jet because the scale factors for both regions are determined in the same fit. The b-tagging uncertainty is correlated between all four regions since the event reweighting from which this uncertainty is derived applies globally across the categories. The weighted mean is therefore determined in two steps: First, separate means are calculated for low and high p_{T}^{W} using error components 1-3 (inverse sum of squares) as weights, then the mean of both values is calculated using error components 1-4.

The combined result for the WZ cross section is

$$\sigma_{WZ} = \left[11 \pm 5(\text{stat}) \pm 3(\text{stat, bg}) \pm 4(\text{syst})\right] \text{pb}\,.$$

The statistical error on the simulated background is not merged with the other systematic uncertainties because it can be reduced by generating more simulated events. Given the large uncertainties, the agreement with the theory prediction [17, 18]

$$\sigma_{WZ_{\text{theory}}} = (20.3 \pm 0.8)\,\text{pb}$$

is better than 2σ. However, the measurement is equally well compatible with the background-only model.

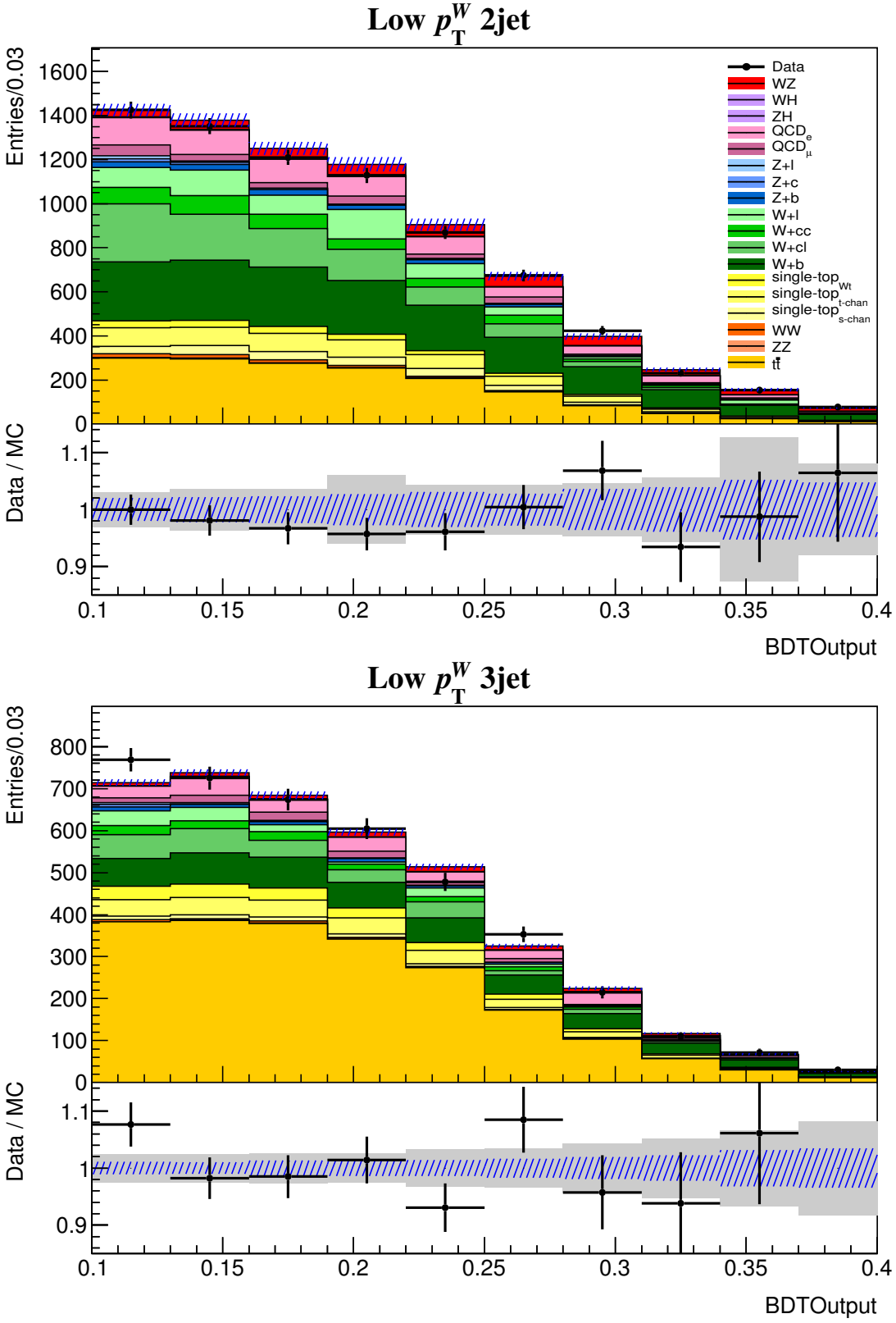

Figure 8.1: BDT classifier range for the cross section measurement in the low p_T^W region (hatched blue: sytematic uncertainty on background, grey: total error on background)

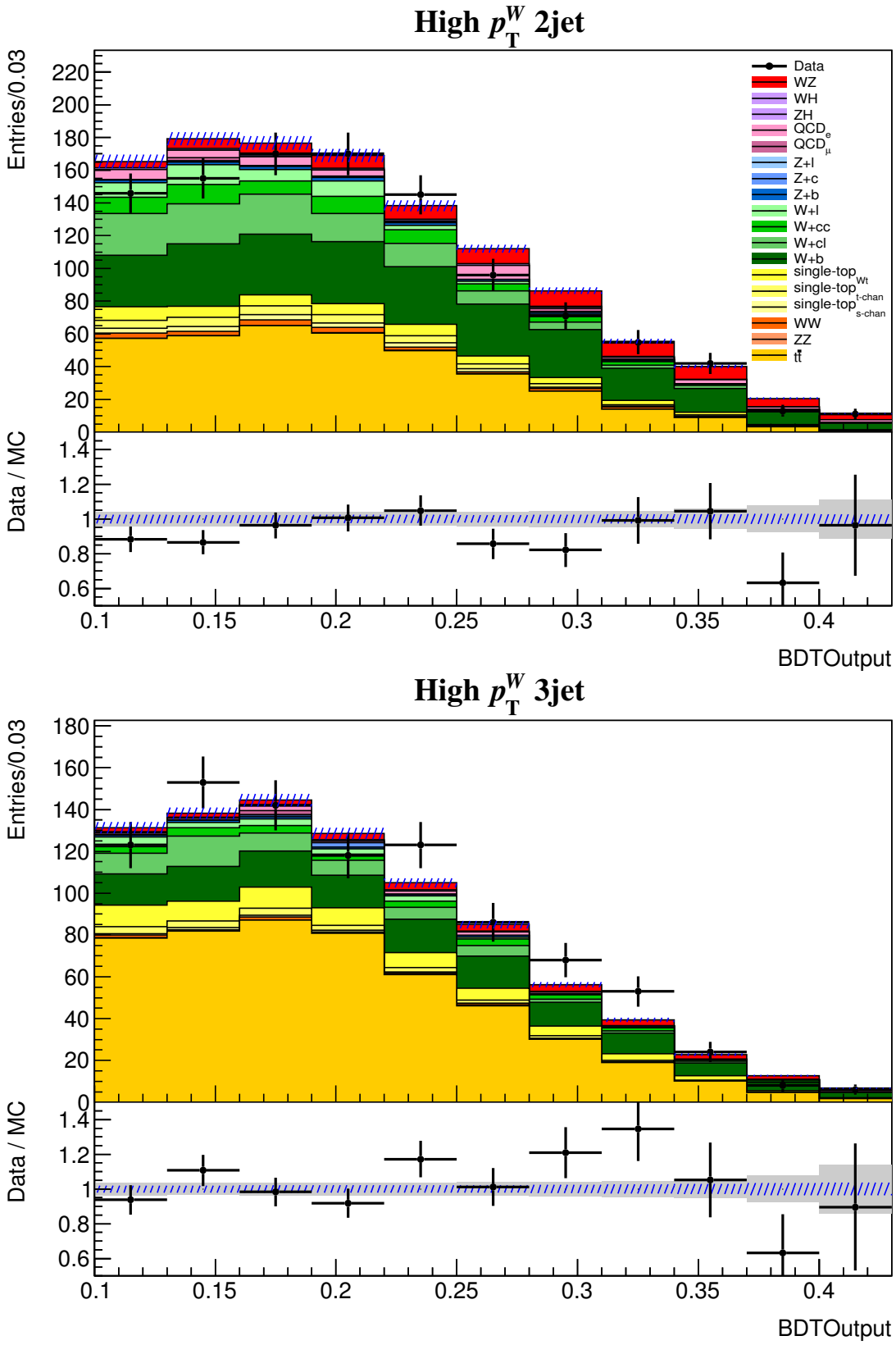

Figure 8.2: BDT classifier range for the cross section measurement in the high p_T^W region (hatched blue: sytematic uncertainty on background, grey: total error on background)

CHAPTER 9

Conclusions

The total cross section of WZ production in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ has been measured in the $WZ \to lvbb$ channel with ATLAS, based on the $20.3\,\text{fb}^{-1}$ dataset from the 2012 run.

The cross section measurement is based on a comparison of the total event count of data and simulated background in a signal-enriched event selection. An excess over background is interpreted as signal. Both data and simulated events undergo the same selection criteria, based on a high p_T lepton trigger. A boosted decision tree trained on simulated events is used to further suppress background.

The cross section measurement was performed separately in different event categories, divided according to jet multiplicity and transverse momentum of the reconstructed W boson. The results were then combined into one final value. Systematic uncertainties were derived for the modeling of kinematic properties, the normalization of different backgrounds and the identification efficiency for b-flavored jets. The cross section was measured to be $\sigma_{WZ} = [11 \pm 5(\text{stat}) \pm 3(\text{stat, bg}) \pm 4(\text{syst})]\,\text{pb}$. The agreement with the theory prediction $\sigma_{WZ_\text{theory}} = (20.3 \pm 0.8)\,\text{pb}$ is better than 2σ. The result is also compatible with the background-only model.

APPENDIX A

Appendix

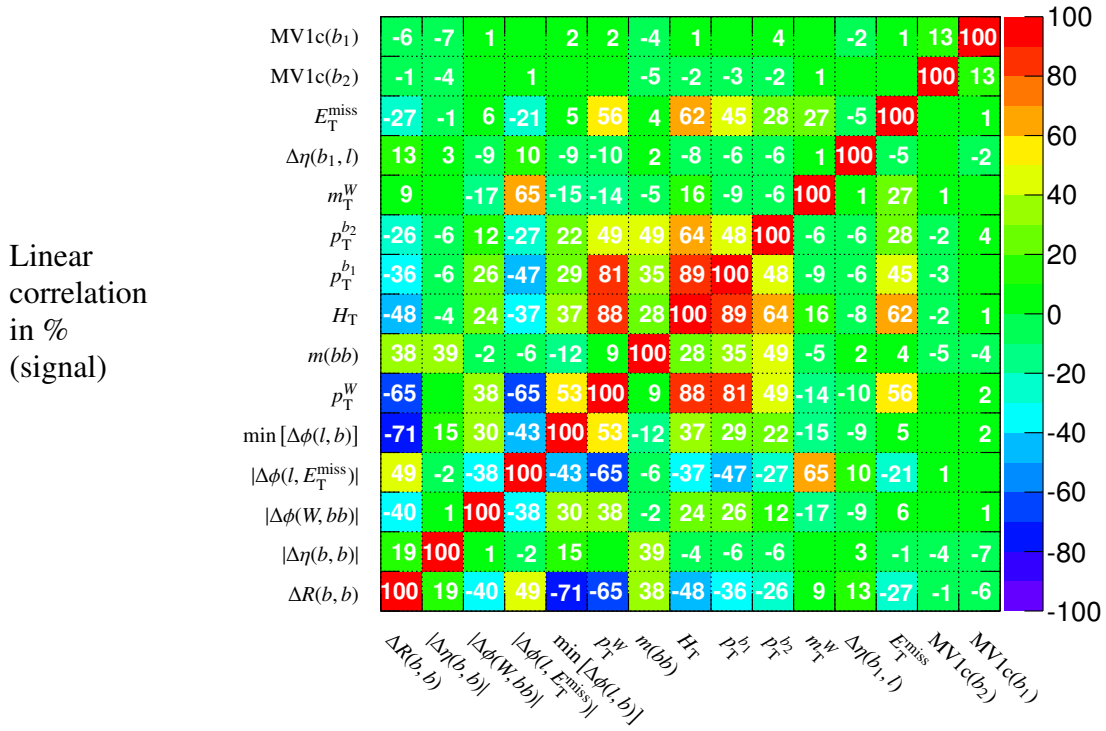

Figure A.1: Signal linear correlation coefficients in % for the Baseline WZ variables. Empty cells mean zero correlation.

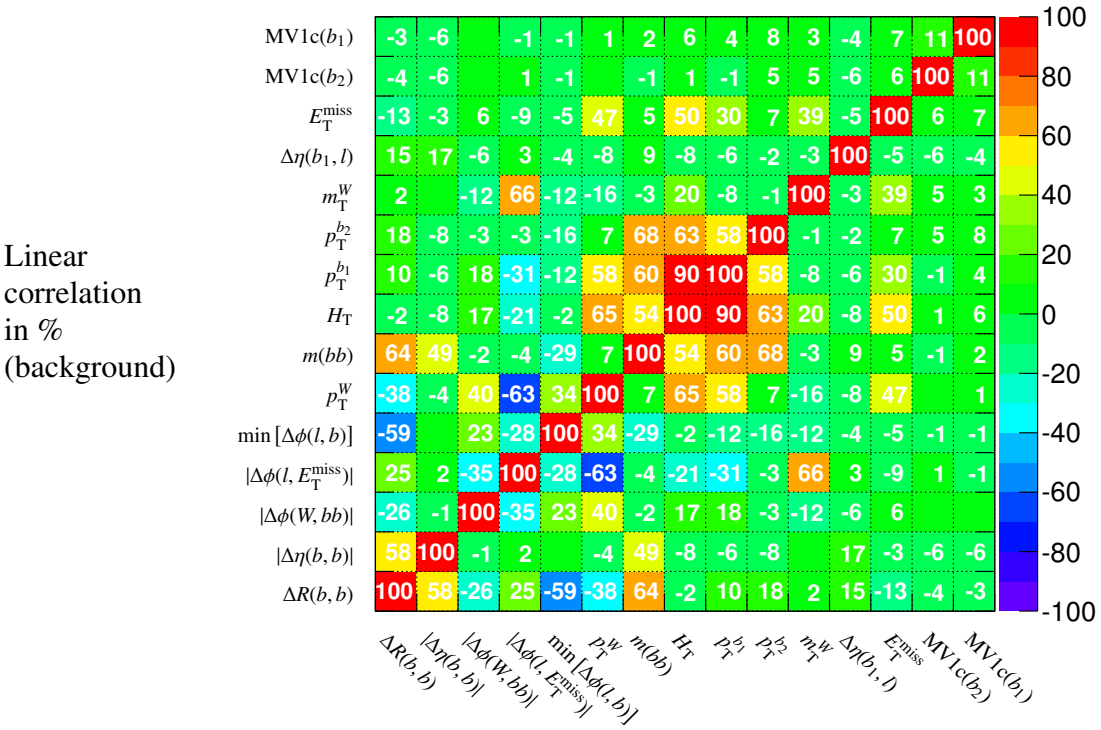

Figure A.2: Background linear correlation coefficients in % for the Baseline WZ variables. Empty cells mean zero correlation.

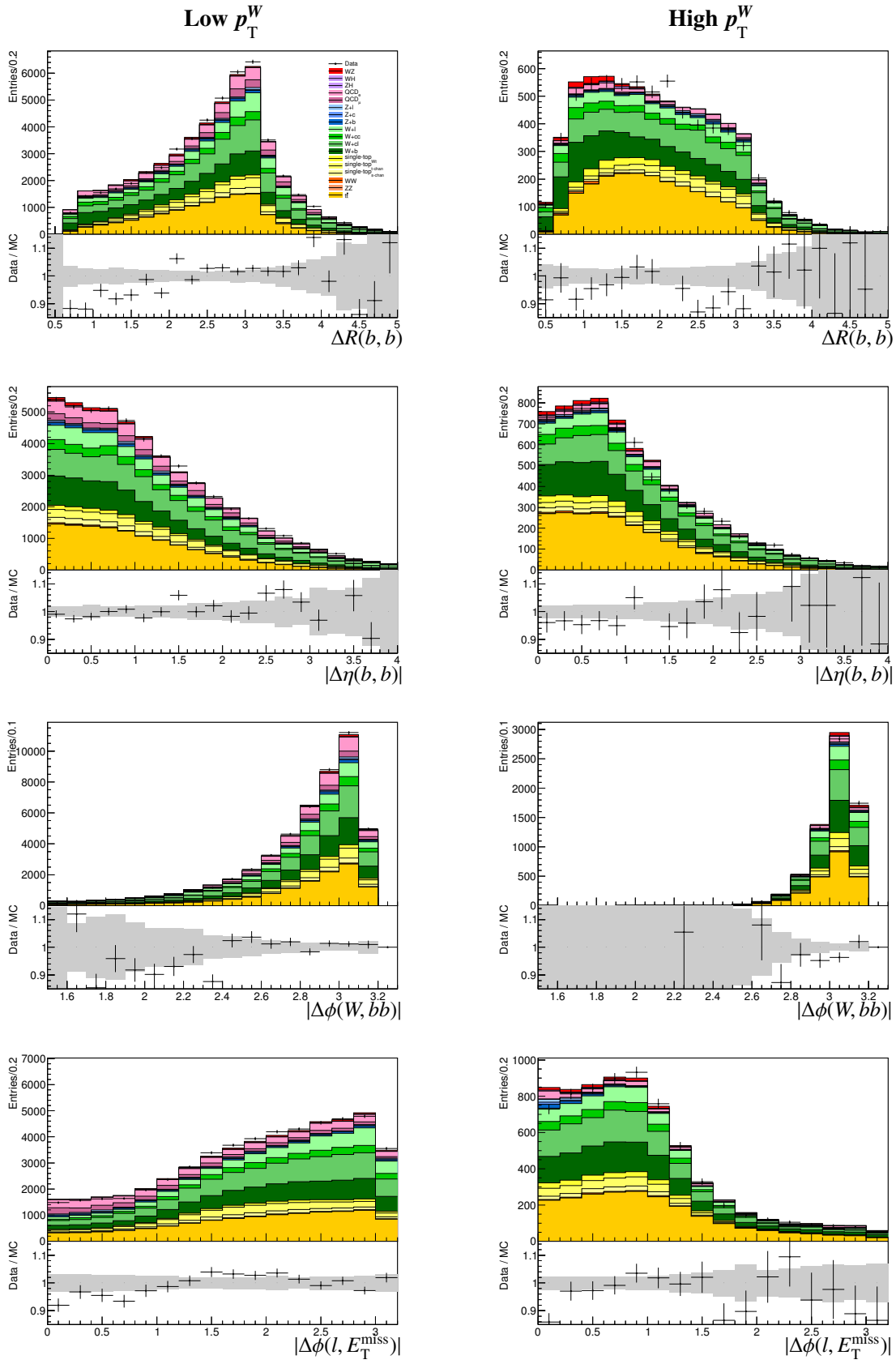

Figure A.3: Comparison of data and simulation for the Baseline WZ variables in low p_{T}^{W} 2jet (after fit, statistical error on background displayed in grey), part 1

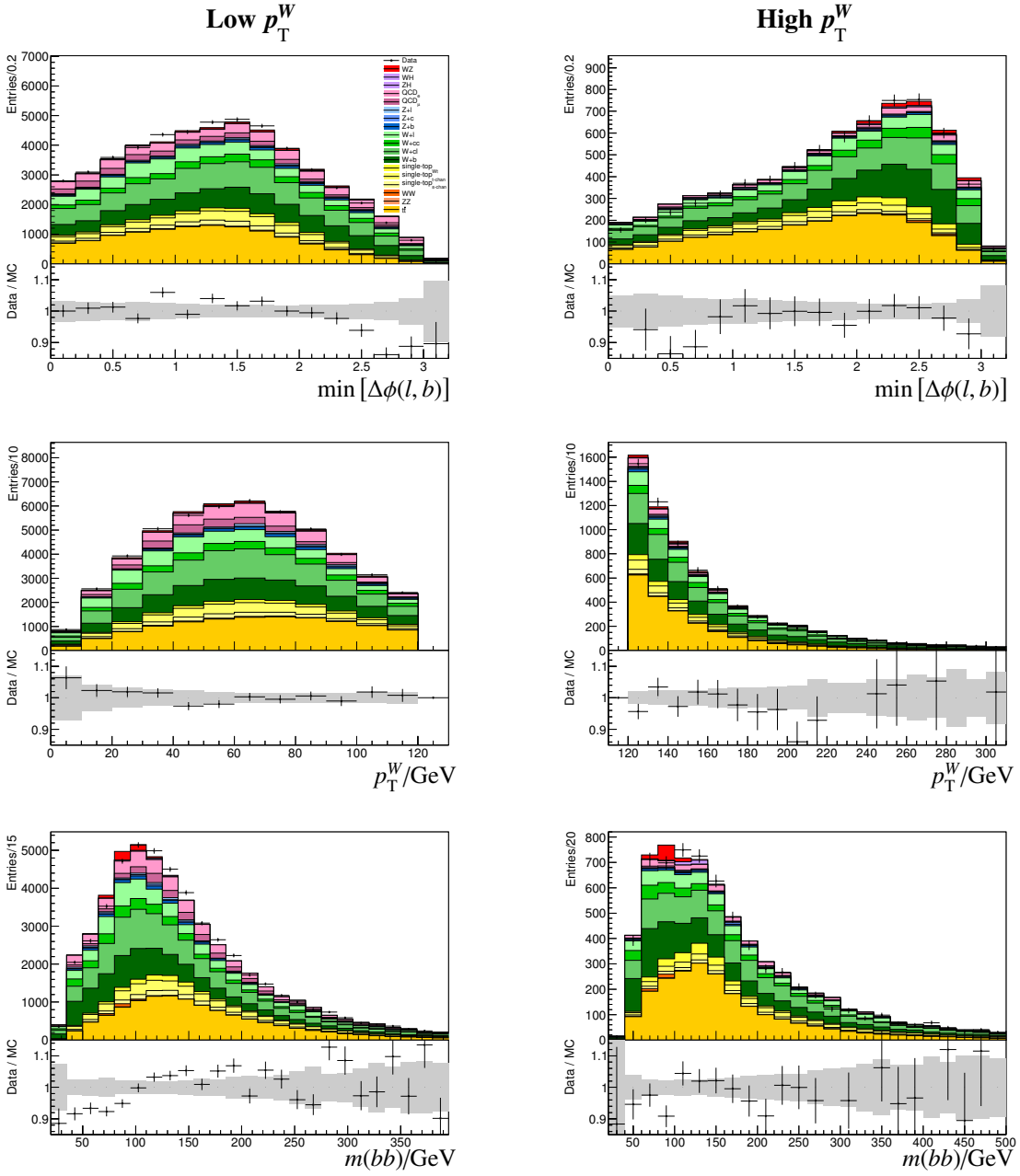

Figure A.4: Comparison of data and simulation for the Baseline WZ variables in low p_{T}^{W} 2jet (after fit, statistical error on background displayed in grey), part 2

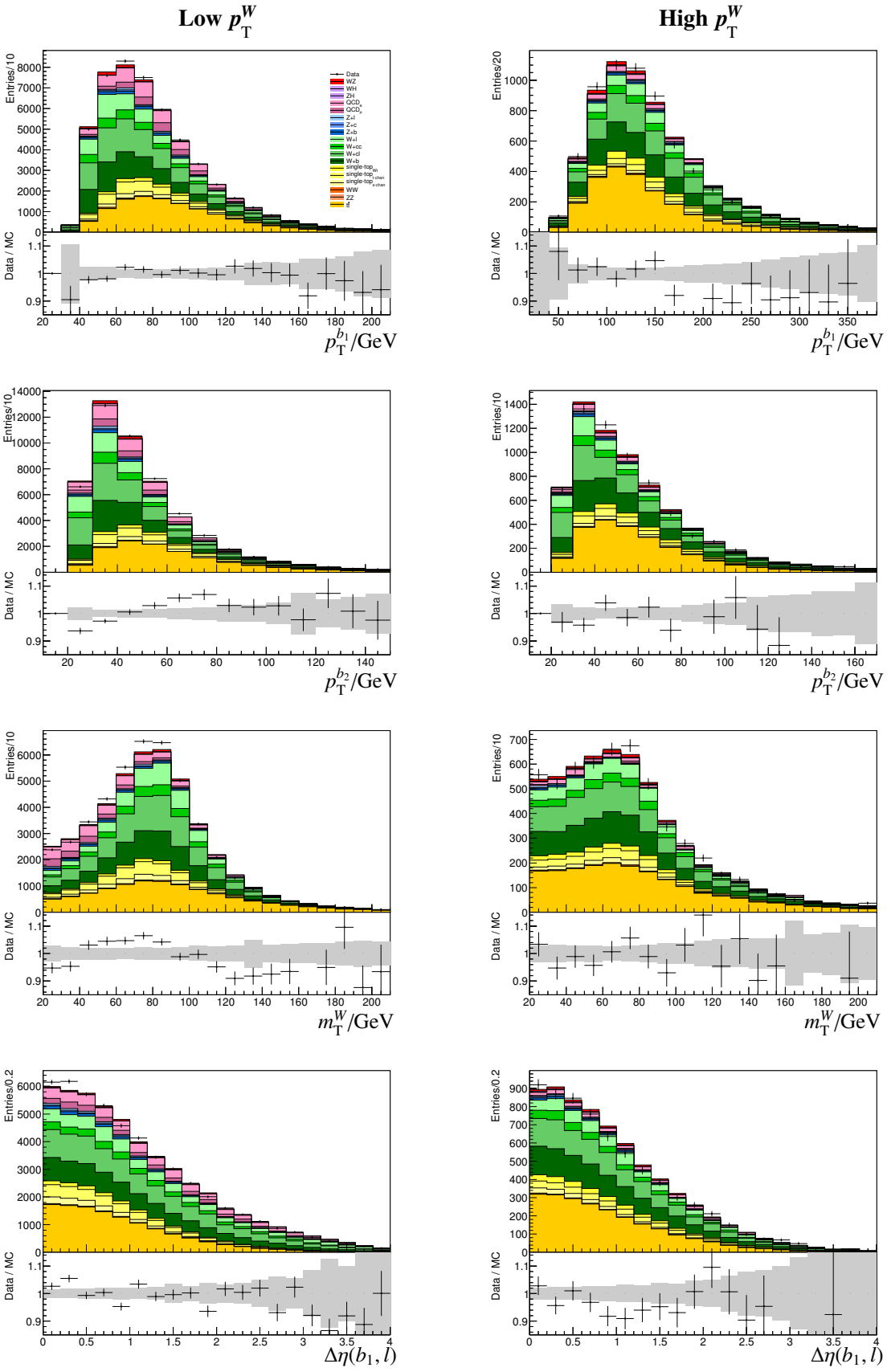

Figure A.5: Comparison of data and simulation for the Baseline WZ variables in low p_{T}^{W} 2jet (after fit, statistical error on background displayed in grey), part 3

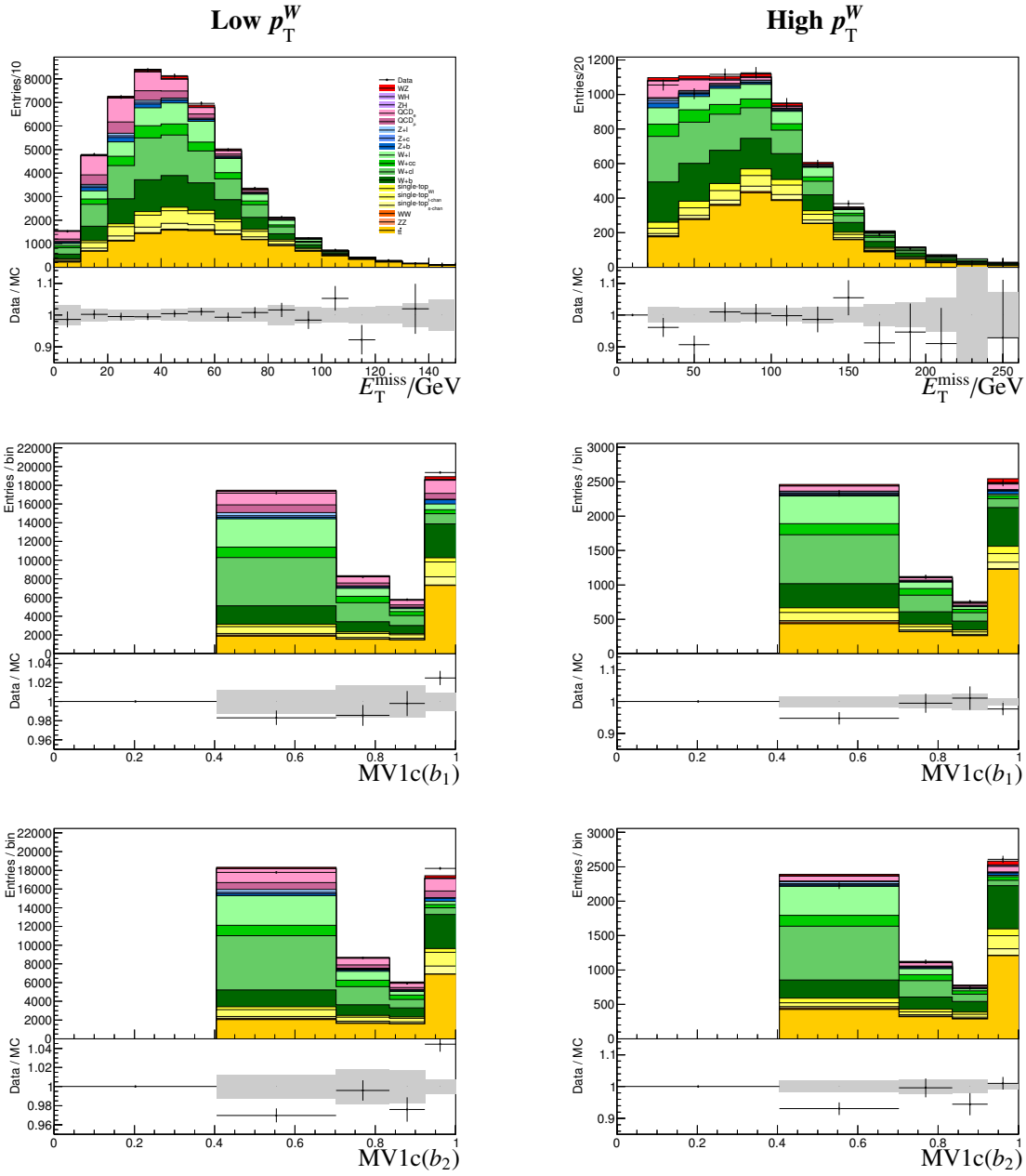

Figure A.6: Comparison of data and simulation for the Baseline *WZ* variables in low p_{T}^{W} 2jet (after fit, statistical error on background displayed in grey), part 4

Bibliography

[1] G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys.Lett.* B716 (208) 1–29, DOI: 10.1016/j.physletb.2012.08.020, arXiv:1207.7214 [hep-ex].

[2] G. Aad et al., "Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector" (2014), arXiv:1409.6212 [hep-ex].

[3] *ATLAS Photos*, URL: http://www.atlas.ch/photos/.

[4] J. Therhaag, "Search for the $H \to b\bar{b}$ decay mode of the Standard Model Higgs boson in associated production with a vector boson in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS experiment", PhD Thesis: University of Bonn, 2014, URL: http://nbn-resolving.de/urn:nbn:de:hbz:5n-36319.

[5] G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* 3 (2008) S08003, DOI: 10.1088/1748-0221/3/08/S08003.

[6] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys.Rept.* 504 (2011) 145–233, DOI: 10.1016/j.physrep.2011.03.005, arXiv:1101.2599 [hep-ph].

[7] T. Sjostrand, "Monte Carlo Generators" (2006) 51–74, arXiv:hep-ph/0611247 [hep-ph].

[8] F. James, "Monte Carlo theory and practice", *Reports on Progress in Physics* 43.9 (1980) 1145, URL: http://stacks.iop.org/0034-4885/43/i=9/a=002.

[9] M. Cacciari, G. P. Salam and G. Soyez, "The anti- k t jet clustering algorithm", *Journal of High Energy Physics* 2008.04 (2008) 063, URL: http://stacks.iop.org/1126-6708/2008/i=04/a=063.

[10] "Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment", tech. rep. ATLAS-CONF-2014-004, CERN, Feb. 2014.

[11] "Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data", tech. rep. ATLAS-CONF-2014-046, CERN, July 2014.

[12] R. J. Barlow, *Statistics, A Guide to the Use of Statistical Methods in the Physical Sciences*, 1st ed., John Wiley & Sons Ltd., 1997.

[13] G. Cowan and E. Gross, "Discovery significance with statistical uncertainty in the background estimate" (2013).

[14] A. Hoecker et al., *TMVA Users Guide*, URL: http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf.

[15] Y. Freund and R. E. Schapire,
"A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting",
Journal of Computer and System Sciences 55.1 (1997) 119 –139, ISSN: 0022-0000,
DOI: http://dx.doi.org/10.1006/jcss.1997.1504,
URL: http://www.sciencedirect.com/science/article/pii/S002200009791504X.

[16] A. Hoecker et al., *TMVA 4.2*, URL: http://tmva.sourceforge.net.

[17] V. Lombardo, "Diboson production cross section at LHC" (2013) 339–342,
arXiv:1305.3773 [hep-ex].

[18] V. Lombardo, *Diboson production cross section at LHC, Moriond talk*, 2013,
URL: http://moriond.in2p3.fr/QCD/2013/SundayAfternoon/Lombardo.pdf.

List of Figures

List of Tables