Università della Calabria

Dipartimento di Fisica

Master's Degree in
Physics

Thesis

Measurements of the production of a Z boson in association with two b-jets in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS experiment

CERN-THESIS-2024-256
10/10/2024

Supervisor
Prof. Evelin Meoni

Candidate
Giuseppina Santoro
ID number 247408

Academic Year 2023/2024

Ogni nuovo mattino,
uscirò per le strade cercando i colori.

(Every new morning,
I will go out into the streets looking for colors.)

– *Cesare Pavese*

Contents

Introduction

The quest to unravel the fundamental structure of the universe has long captivated scientists and philosophers alike. Particle physics is the field that seeks to understand the basic constituents of matter and the forces governing their interactions. The Large Hadron Collider (LHC) at CERN stands as a monumental achievement in this pursuit, capable of colliding protons at unprecedented energies. ATLAS is one of the two general-purpose particle physics experiments at the LHC and it is designed to fully exploit the discovery potential of these collisions, pushing the boundaries of scientific knowledge. Through the ATLAS experiment, it is possible to investigate the fundamental constituents of matter and their interaction.

So far, all experimental results in this field show excellent agreements with the predictions of the Standard Model (SM) theory of particle physics. Despite the SM's remarkable success in explaining a vast array of phenomena, it leaves several pressing questions unanswered, such as those surrounding Dark Matter and the matter-antimatter asymmetry. The development of various theories beyond the Standard Model has occurred over the years, yet validation through experimental evidence remains crucial. This underscores the significance of advancing our understanding in this domain to explore new physics beyond the confines of the SM. The ATLAS experiment aims to validate the SM at new energy scales, investigate the properties of the Higgs boson and search for phenomena that extend beyond the current theoretical framework.

This thesis focuses on the measurement of the production of a Z boson, which decays leptonically, in association with jets originating from b-quarks. This analysis uses proton-proton collision data at $\sqrt{s} = 13$ TeV collected by ATLAS, corresponding to an integrated luminosity of 140 fb^{-1} (referred to as full Run-2 dataset), that is currently the largest and best known dataset available at ATLAS. The process of the production of a Z boson in association with b-jets is mediated by the strong interaction between quarks and gluons, which are the fundamental costituents of the proton.

In the measurement, the Z boson decays leptonically in a lepton pair with opposite charge, either muons ($\mu^+\mu^-$) or two electrons (e^+e^-), while the b-quarks hadronize to form jets. The b-quarks are reconstructed within the detector as b-jets cones of highly collimated hadrons. Jets coming from b-quarks hadronisation are disentangled with jets from other quarks or gluons, exploring measured features related to the properties of the b-quarks (i. e. the lifetime).

Such studies are crucial for probing perturbative Quantum Chromodynamics (pQCD) in high-energy regimes, as well as for improving the understanding of heavy-quark contributions to parton distribution functions (PDFs), which describe how a proton's momentum is distributed among its constituent quarks and gluons. In addition to advancing our theoretical understanding of pQCD, this process, $Z + b$ jets, is also a key background for many analyses at the LHC. A prominent example is the study of Higgs boson production in association with a Z boson (ZH), where the Z boson decays into leptons and the Higgs boson decays into bottom quark pairs ($H \to b\bar{b}$). Although this final state has already been observed at the LHC, improving the precision of measurements requires a better understanding of the $Z + b$ jets background. Delving deeper into Higgs physics and studying processes predicted by the Standard Model is a method of paramount importance to potentially find evidence of new physics, by identifying discrepancies with the Standard Model expectations.

The work of this thesis builds upon the recent findings presented by ATLAS in the measurement of the inclusive and differential cross-sections for a Z boson in association with b- or c- jets with the full Run-2 dataset. Specifically this thesis provides differential cross sections as a function of new observables in the two b-jets final states, not included in the latest ATLAS publication, particularly those relevant to the Higgs boson sector plus few others important to pQCD and PDFs studies.

The differential cross sections as a function of the new variables measured in this thesis are compared with state-of-art MC generators and provide therefore inputs for their improvements, moreover they offer a more neanced understanding of the pQCD in this process.

The thesis begins with a brief overview of the Standard Model (Chapter 1), focusing on the theoretical foundations necessary to understand proton collisions. It also delves into the role of strong interactions, which are fundamental to describe the production of the Z boson in association with b-jets. In this chapter also the state-of-art of the knowledge of the process that is the aim of the thesis is presented.

A technical description of the LHC and the ATLAS detector follows in Chapter 2, with emphasis on the reconstruction and identification tools of the physics objects used in this work. This is crucial for understanding how events involving Z bosons and b-jets are identified and measured experimentally.

The measurement of the Z + b-jets process at the detector level, i.e. before correcting for detector effects, is presented in Chapter 3, detailing the event selection criteria, background estimation techniques and associated uncertainties.

The unfolding technique, used to pass from the detector-level measurement to the cross-section one at particle level, used for the analysis and the results of the analysis are presented in Chapter 4 through differential cross sections, highlighting the dependence on observables not previously considered in the most recent publication on this process. These results provide valuable insights into the contributions of heavy quarks and are compared with theoretical predictions. The systematic uncertainties are discussed in detail, with a particular focus on how they affect the precision of the differential cross section measurements.

Finally, the inclusive and differental cross-section are presented and compared with the state-of-art signal MC predictions.

A delailed description of the unfolding procedure, which is used to correct the data for detector effects, is given in the Appendix. In this work, the Bayesian unfolding method is employed, allowing for a robust transformation from the detector-level measurements to the particle-level truth distributions.

Chapter 1

Theoretical and phenomenological overview

The Standard Model (SM) is currently the theory that best describes the interactions between fundamental particles. Throughout history, numerous experiments have attempted to identify flaws or limitations within this theory. Although there are phenomena that the SM does not fully explain, it remains the most tested and validated theory to date. In this chapter, the SM is discussed in detail in Section 1.1. In Section 1.2 the peculiar theoretical aspects of proton-proton (pp) collisions are addressed and Section 1.3 focuses on Monte Carlo simulations. Section 1.4 delves into the current understanding of the production of the Z boson in association with b-jets in pp collisions.

1.1 The Standard Model of Particle Physics

The fundamental forces in physics are four: strong, electromagnetic, weak, and gravitational. Three of these four fundamental forces, excluding the gravitational interaction, are described by the SM of particle physics [1]. The theory of the SM was developed in the latter half of the 20^{th} century through the work of numerous scientists from around the world. It is a quantum gauge field theory; the gauge symmetry group is $SU(3)$ and $SU(2) \times U(1)$. $SU(3)$ is the symmetry group of the strong interactions, while $SU(2) \times U(1)$ is the symmetry group of the electroweak interactions. Today, a lot of experiments confirm the predictions of the SM. However there are also phenomena that cannot be explained by this theory, such as the Dark Matter of the Universe, the asymmetry between particles and antiparticles, the unification of all forces at very high energy including the gravity as it was at the Big Bang. Due to these limitations, various theories beyond the SM [2] have been proposed, in which the SM maintains the role of theory valid at low energies, though these new theories have yet to be confirmed.

The SM describes the fundamental particles in nature and their interactions. All the fundamental particles of the SM are shown in Fig. 1.1. These particles are divided in two categories: fermions and bosons.

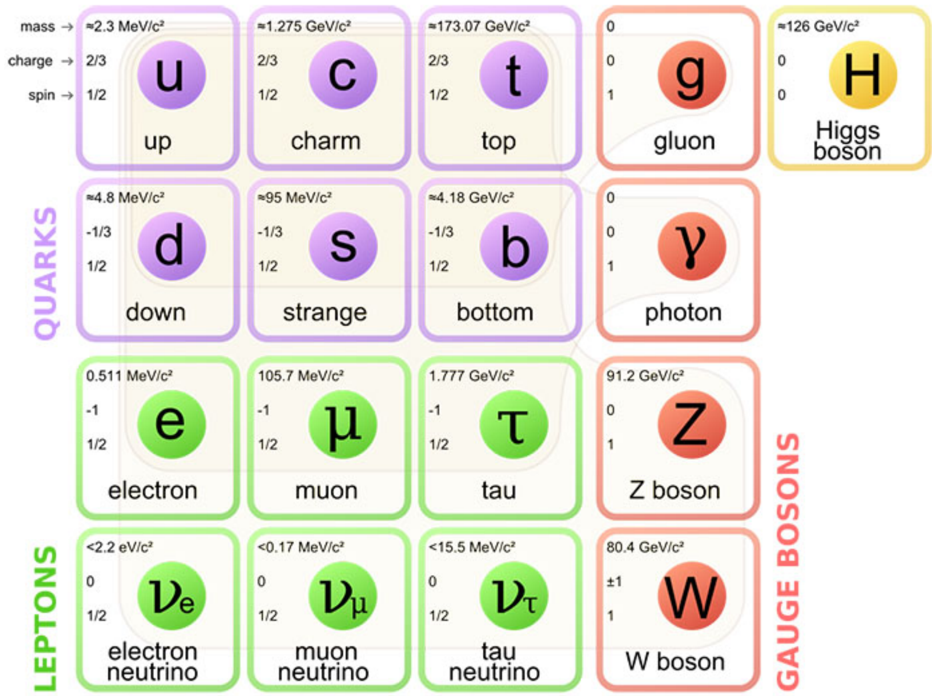

Figure 1.1: Particles predicted by the SM. The quarks are in purple, the leptons are green, the gauge boson in red and the Higgs boson is the yellow one.

The fermions

Fermions are particles with half-integer spin, such as leptons and quarks.

There are three families of leptons, each differing in mass, as illustrated in Fig. 1.1. The first family is composed of the electron (e^-) and the electron neutrino (ν_e). The second family consists of the muon (μ) and the muon neutrino (ν_μ). The third family consists of the tau (τ) and the tau neutrino (ν_τ). Although the SM initially predicted neutrinos to be massless, experimental evidences, such as neutrino oscillations, indicate that they possess a small mass [3]. The masses of other leptons increases moving from the first up to the third family. Electron, muon and tau have the same electric charge e, while neutrino is chargeless. Leptons can be observed as free particles.

Quarks are fundamental particles with fractional electric charges of either

$-\frac{1}{3}$ e or $\frac{2}{3}$ e. There are three families of quarks, each consisting of two quarks as shown in Fig. 1.1.

The first family includes the up quark (u) with a charge of $\frac{2}{3}$ e and the down quark (d) with a charge of $-\frac{1}{3}$ e. The second family includes the charm quark (c) with a charge of $\frac{2}{3}$ e and the strange quark (s) with a charge of $-\frac{1}{3}$ e. The third family is composed of the top quark (t) with a charge of $\frac{2}{3}$ e and the bottom quark (b) with a charge of $-\frac{1}{3}$ e. Regarding the quarks, similarly as for the leptons, those in the first generation are lighter, while those in the last generation are heavier.

They also carry a color charge, which can be in one of three states: red, blue, or green. Due to the color confinement in strong interactions, quarks cannot be observed as free particles but only in bound states known as hadrons. Hadrons can be baryons or mesons. Baryons consist of three quarks bound together, while mesons are composed of a quark-antiquark pair. Protons and neutrons are examples of baryons and are made up of the quark combinations (uud) and (udd), respectively.

Each fermion has an associated antiparticle, known as an antifermion, which have the same mass as their corresponding fermions but opposite charges. For instance, the positron (e^+) is the antiparticle of the electron (e^-) and has a positive electric charge and it has leptonic number -1.

The Bosons

Bosons are particles with integer spin that act as mediators of the fundamental forces.

The SM bosons are: 8 gluons, the photon, the Z boson and the W^+ and W^- bosons and the Higgs boson.

The photon is the mediator of the electromagnetic interaction, which affects charged particles.

The gluon is the mediator of the strong interaction, binding quarks together within protons, neutrons and other hadrons. Gluons carry the color charges and interact by exchanging this charge with quarks, thus they participate in strong interactions themselves.

The Z and W bosons mediate the weak interaction, which is responsible for processes like beta decay in atomic nuclei. The W bosons exist in both positively charged (W^+) and negatively charged (W^-) forms, while the Z boson is neutral. The Z and W bosons are massive, in contrast to the photon and gluons, which are massless.

The Higgs boson is a fundamental particle whose discovery in 2012 confirmed the Higgs mechanism, responsible for giving mass to elementary particles. Its interaction with the Higgs field allows particles to acquire mass.

In Table 1.1 are collected the masses and the charges of all the particles predicted by the SM.

Particle	Type	Charge (e)	Mass (MeV/c^2)
Fermions			
Electron (e^-)	Lepton	-1	0.511
Muon (μ^-)	Lepton	-1	105.66
Tau (τ^-)	Lepton	-1	1776.86
Electron Neutrino (ν_e)	Lepton	0	< 0.0000022
Muon Neutrino (ν_μ)	Lepton	0	< 0.17
Tau Neutrino (ν_τ)	Lepton	0	< 18.2
Up Quark (u)	Quark	$+\frac{2}{3}$	2.16
Down Quark (d)	Quark	$-\frac{1}{3}$	4.67
Strange Quark (s)	Quark	$-\frac{1}{3}$	93
Charm Quark (c)	Quark	$+\frac{2}{3}$	1270
Bottom Quark (b)	Quark	$-\frac{1}{3}$	4180
Top Quark (t)	Quark	$+\frac{2}{3}$	173100
Bosons			
Photon (γ)	Gauge Boson	0	0
Gluon (g)	Gauge Boson	0	0
W Boson ($W^\pm$)	Gauge Boson	± 1	80379
Z Boson (Z^0)	Gauge Boson	0	91188
Higgs Boson (H^0)	Scalar Boson	0	125100

Table 1.1: Fundamental fermions and bosons with their electric charge and mass [4].

1.1.1 The Standard Model Lagrangian

Each interaction of the SM is described by a quantum field theory [5] in a specific group. The SU(3) group is the simmetry group for the color that is used to describe the strong interaction, SU(2) is for the weak isospin, while U(1) is for the hypercharge and these last two are unified in the $SU(2) \times U(1)$ that describes the electroweak theory. A fundamental parameter that determines the strenght of the particle interactions is the coupling constant. The couplings will be described in Section 1.1.2.

The Lagrangian for the Quantum Electrodynamics (QED) for a massless fermion field ψ is

$$\mathcal{L}_{QED} = \overline{\psi} i \gamma^\mu \partial_\mu \psi - e \overline{\psi} \gamma^\mu \psi A_\mu(x) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.1}$$

where A_μ is the electromagnetic vector field and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strenght tensor. The first term in the lagrangian describes the free propagation of fermions without interaction, the second term describes the interaction between the fermion and the gauge field A_μ, while the third term is for the energy and the dynamics of the electromagnetic field.

Imposing the Dirac Lagrangian to be invariant under $SU(2) \times U(1)$ local transformation, the Dirac Lagrangian of the electroweak interaction is

$$\mathcal{L}_{EW} = \sum_{f=q,l} \overline{f} i\gamma^{\mu} D_{\mu} f - \frac{1}{4} W^{i}_{\mu\nu} W^{i,\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.2}$$

where

$$\begin{aligned} W^{i}_{\mu\nu} &= \partial_{\mu} W^{i}_{\nu} - \partial_{\nu} W^{i}_{\mu} + g\epsilon^{ijk} W^{j}_{\mu} W^{k}_{\nu} \\ B_{\mu\nu} &= \partial_{\mu} B_{\nu} - \partial_{\nu} B_{\mu} \end{aligned} \tag{1.3}$$

B_{μ} is the gauge field for U(1) group, while W_{μ} is the gauge field for the SU(2) group and D_{μ} is the covariant derivative defined as

$$D_{\mu} = \partial_{\mu} - ig\vec{T} \cdot \vec{A}_{\mu}(x) \tag{1.4}$$

The vector $\vec{T}$ represents the generators of the SU(2) Lie algebra.

The Quantum Chromo-Dynamics (QCD) Lagrangian for massless fermion field written already after imposing the invariance under SU(3) local transformation is

$$\mathcal{L}_{QCD} = \sum_{f=q} \overline{f} i\gamma^{\mu} D_{\mu} f - \frac{1}{4} G^{a}_{\mu\nu} G^{a,\mu\nu} \tag{1.5}$$

where $G_{\mu\nu}$ is the field strength tensor defined as

$$G_{\mu\nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu} - ig_{s}[A_{\mu}, A_{\nu}] \tag{1.6}$$

with g_s the strong coupling constant and $[A_{\mu}, A_{\nu}]$ the commutator of the A_{μ} and A_{ν} fields. In this last case $\vec{T}$ represents the generators of SU(3) Lie algebra.

The QCD theory describes the interaction between colored particles such as quarks and gluons. These particles cannot be directly observed because of the property of confinement that is related to the running of the strong coupling.

1.1.2 Running couplings

The couplings, both electromagnetic e and strong g_s, are not constant values but they vary with energy [6]. In QED, the behaviour of the coupling depends on the energy of the probe. Indeed, as the energy increases, the coupling strength increases and viceversa. The higher is the coupling, the more intense will be the interaction. Indeed, as Q^2 increases, the denominator decreases and $e(Q^2)$ increases.

In natural units, the electromagnetic coupling e is usually also expressed as $\alpha = \frac{e^2}{4\pi}$, so e and α are proportional.

The running QED coupling is expressed as:

$$\alpha(Q^2) = \frac{\alpha(\mu^2)}{1 - \frac{\alpha(\mu^2)}{3\pi} ln\left(\frac{Q^2}{\mu^2}\right)} \tag{1.7}$$

where Q^2 is the energy of the probe, μ^2 is the reference scale, often taken as $\mu^2 = m_e^2$, where m_e is the mass of the electron. So, here, increasing the energy of the probe makes the interaction more intense.

In QCD, due to possible gluon-gluon interaction, the behaviour is different. The possible tree level Feynman diagrams are shown in Fig. 1.2.

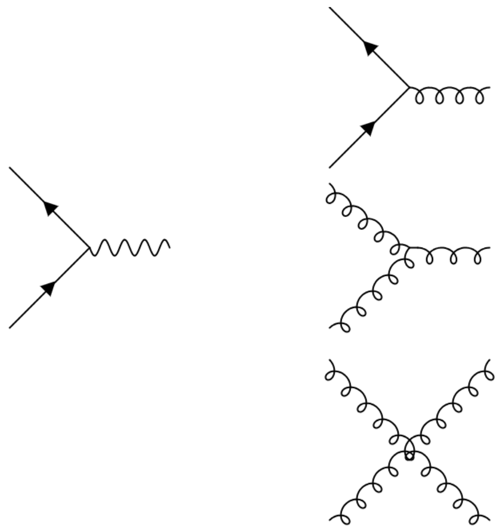

Figure 1.2: The diagram on the left is the only possible tree level Feynman diagram for QED, on the right the possible tree level feynman diagrams in QCD.

The QCD coupling is

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \frac{\alpha_s(\mu^2)}{12\pi}(11n_c - 2n_F) ln\left(\frac{Q^2}{\mu^2}\right)} \tag{1.8}$$

where n_c is the number of colors, n_f is the number of flavors and $\alpha_s = \frac{g_s^2}{4\pi}$.

The different sign in the denominator leads to the opposite behaviour. Here, increasing the energy of the probe, i.e. probing at short distances, the particles behave as free and the phenomenon is known as *asymptotic freedom*. At lower energies and large distances, the coupling becomes stronger, leading to *confinement*, that is the reason why colored particles cannot be observed directly.

In Fig. 1.3 are shown the behaviours of the QED and QCD couplings as a function of the energy of the probe.

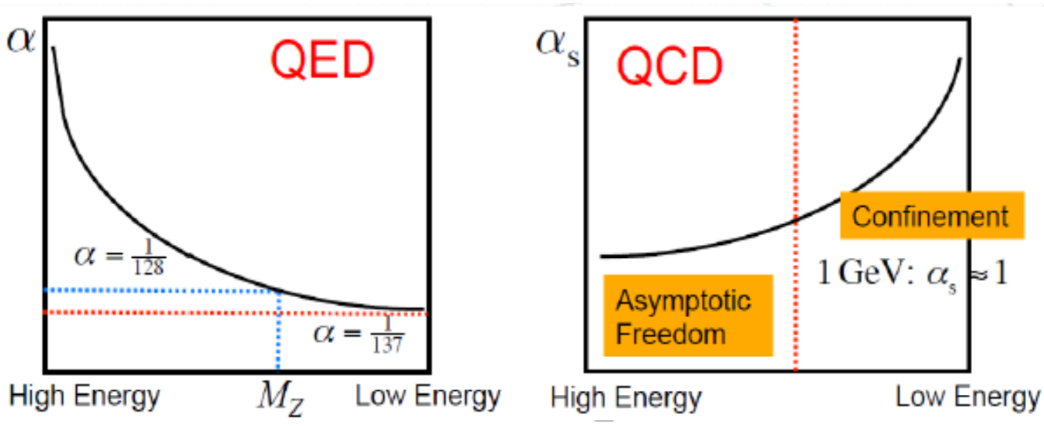

Figure 1.3: QED (on the left) and QCD (on the right) running couplings in function of Q^2.

1.1.3 Higgs mechanism

Another fundamental component of the SM is the Higgs mechanism, crucial for explaining the origin of the mass of the particles.

In order to explain the origin of the mass of elementary particles, in the SM model it is introduced a scalar field, known as the Higgs field. The Higgs field has a vacuum expectation value different from zero which leads to the breaking of the electroweak symmetry. As a consequence of the Higgs mechanism, the W and Z boson acquire mass, the photon remain massless, and it appears a new scalar particle, the Higgs boson. Interacting with the Higgs field also fermions acquire mass.

For example, the Yukawa interaction between the Higgs field and fermion fields is described by the Lagrangian

$$\mathcal{L}_{int} = \frac{gv}{\sqrt{2}}\left(\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L\right) + \frac{g}{\sqrt{2}}\left(\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L\right)h \tag{1.9}$$

The mass of the fermion m_f is given by $m_f = \frac{g_f v}{\sqrt{2}}$ and recognizing that $\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L = \overline{\psi}\psi$, the Lagrangian can be rewritten as:

$$\mathcal{L}_{int} = m_f\overline{\psi}\psi + \frac{m_f}{v}\overline{\psi}\psi h \tag{1.10}$$

where the first term is the mass term for the fermion and the second term shows the interaction between the Higgs boson and the fermion and the coefficient indicates the strenght of this interaction.

Due to this coupling, the probability of interaction between the Higgs boson and heavier particles is higher.

For many years, the Higgs boson was regarded as the "missing piece" of the SM. Its discovery in 2012 marked a significant milestone in particle physics, confirming the Higgs mechanism.

1.2 Proton-proton collision theory

Proton-proton (pp) collisions are crucial experiments in high-energy physics. In these experiments, two protons are accelerated to extremely high velocities and made to collide. When two protons collide, their constituents interact in a variety of ways.

The scattering can be elastic, when the particles in the final state are the same as those in the initial state, or inelastic. In this thesis the focus will be on inelastic scattering.

Most of the inelastic interactions are soft, characterized by low momentum transfer and non-perturbative effects, while in hard interactions, which are described by perturbative Quantum Chromodynamics (pQCD), the momentum transfer is high. Therefore in the study of pp collisions at high energy both perturbative and non-perturbative QCD processes are involved.

The inner constituents of the protons—quarks and gluons—begin to interact directly. This process is referred to as deep-inelastic scattering. In these interactions, pQCD allows for the calculation of cross-sections at fixed order using Parton Distribution Functions (PDFs) and squared matrix elements (MEs).

During the interaction, soft and collinear partons can be emitted from the initial or final states, a process known as parton showering. Parton showers evolve the system from the hard scattering scale down to the hadronization scale by simulating successive emissions, which can be modeled through a series of splittings, resulting in a cascade of lower-energy partons.

Quarks and gluons in the final state recombine to form colorless hadrons through a process called hadronization, which is based on phenomenological models. Hadronization ensures that observable particles are color-neutral, as required by QCD.

In studying pp collisions, additional complexities can arise due to pileup events, which are soft collisions resulting from other parton-parton interactions occurring simultaneously.

In Fig. 1.4, a schematic representation of a collision between two protons is shown. The two objects on the left represent the protons, whose internal structure is described by parton distribution functions (PDFs). During the collision, two partons interact in a hard scattering process. The remaining partons within the protons contribute to what is known as the underlying event. The elementary particles produced in the hard scattering process constitute the parton level. These partons undergo fragmentation, transitioning from the parton level to the particle level, where quarks and gluons are combined into hadrons.

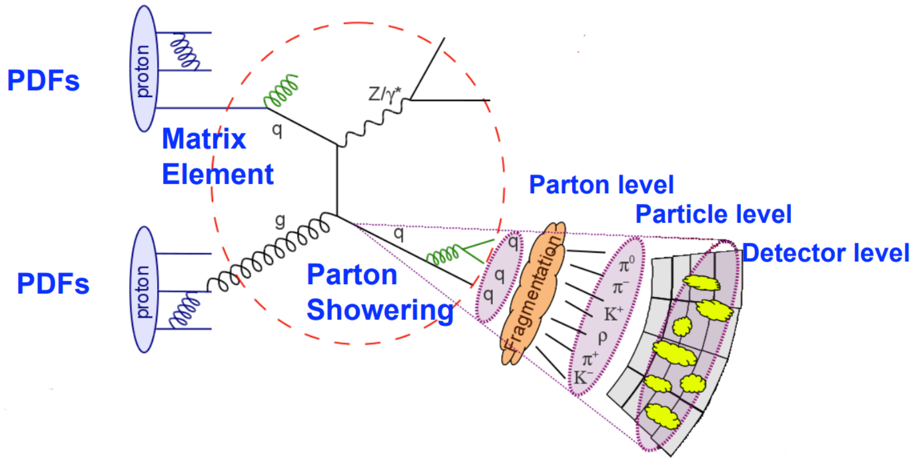

Figure 1.4: Schematization of a pp collision that shows evidently particle level, parton level and detector level.

1.2.1 The internal structure of the proton and the parton distribution functions

The protons are hadrons, so they're not point-like; instead, they are composed of quarks and gluons.

In the Gell-Mann representation, baryons are composed of three quarks, while the mesons are composed of a quark-antiquark pair. This is just a static representation of the baryions that, as the deep inelastic scattering experiments have shown, is not entirely realistic.

Indeed, baryons are dynamic system where quark-antiquark pairs (called sea quarks) are continuosly created and annihilated producing and absorbing gluons.

The hadrons are composed of partons that carry a fraction of the total momentum. The parton distribution functions (PDF) describe the internal structure of hadrons as a function of the fraction x of the momentum of the proton carried by the partons at the scale factor Q^2. The PDFs give the probability to find partons in an hadron.

The kinetic variables can be expressed as

$$Q^2 \equiv -q^2, \qquad x \equiv \frac{Q^2}{2pq} = \frac{Q^2}{2M\nu}, \qquad y \equiv \frac{pq}{qk} \tag{1.11}$$

where Q^2 is the square of the four-momentum transfer of the photon, q^2 is the square of the modulus of the four-momentum transfer vector in the process, x is the Bjorken variable (a measurement of the elasticity of the process), p is the four-momentum of the target particle, q is the four-momentum of the photon, M is the mass of the target particle, ν is the energy transfer, y is the inelasticity parameter and k is the four-momentum of the incident particle.

The PDF are parametrized in the following form:

$$f_i(x, Q^2) = x^{\alpha_i}(1-x)^{\beta_i} g_i(x) \tag{1.12}$$

where g_i is a polynomium in x or in $\sqrt{x}$.

These functions cannot be estimated from first principle; instead, they need to be measured in experiments at a certain scale extracting them from differential cross-sections and then used in the predictions of new measurements at different scales. The DGLAP equations [7] [8] [9] describe the evolution of the partonic distributions in order to change the scale from the initial parametrization to the hard-scattering scale, known as the factorization scale. The integro-differential DGLAP equations are

$$\frac{\partial f_i(x, \mu^2)}{\partial \ln \mu^2} = \sum_{j=q,\bar{q},g} \int_x^1 \frac{dz}{z} P_{ij}\left(\frac{x}{z}, \mu^2\right) f_j(z, \mu^2) \tag{1.13}$$

where μ^2 is the energy scale, associated with the momentum transfer squared, $\frac{\partial f_i(x,\mu^2)}{\partial \ln \mu^2}$ is the rate of change with respect to $\ln \mu^2$. It describes how the PDF evolves as the energy scale μ^2 changes. The sum is made over partons (quarks, antiquarks and gluons). The integral $\int_x^1 \frac{dz}{z}$ accounts for all possible ways the parton can contribute to the observed parton distribution.

Here P_{ij} represents the splitting functions that are calculated as a series of the couplings as:

$$P_{ij} = \frac{\alpha_s}{2\pi} P_{ij}^{(0)} + \left(\frac{\alpha_s}{2\pi}\right)^2 P_{ij}^{(1)} + \left(\frac{\alpha_s}{2\pi}\right)^3 P_{ij}^{(2)} + O(\alpha_s^4) \tag{1.14}$$

where $P_{ij}^{(0)}$, $P_{ij}^{(1)}$ and $P_{ij}^{(2)}$ are the leading-order (LO), next-to-leading-order (NLO), and next-to-next-to-leading-order (NNLO) contributions to the splitting function.

Parton distribution functions (PDFs) are parameterized by various collaborations, including NNPDF [10], CTEQ [11] and MSTW [12].

CTEQ and MSTW follow a traditional methodology that relies on functional parameterizations to describe the PDFs. Specific functional forms are chosen and then adjusted to fit experimental data. CTEQ handles uncertainties using the Hessian approach, while MSTW analyzes the covariance matrix to assess uncertainties.

On the other hand, NNPDF employs a more modern approach based on neural networks. This methodology helps reduce bias associated with the choice of functional forms. It allows a more accurate treatment of statistical uncertainties and involves a larger number of parameters, enhancing the flexibility and precision of the PDF modeling.

In Fig. 1.5 the comparison between these different collaboration is shown.

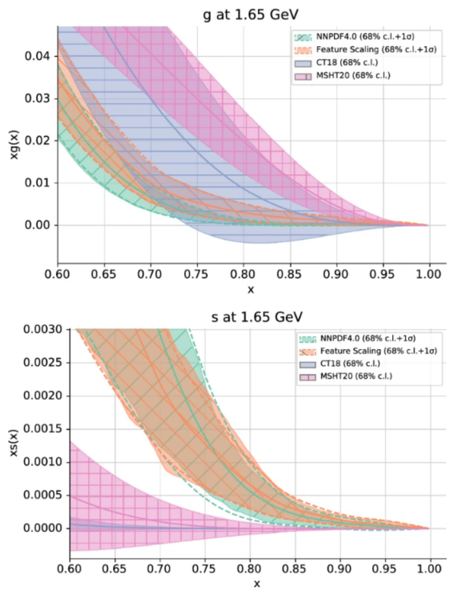

Figure 1.5: Comparison of the large-x extrapolation regions of the gluon (top) and the strange (bottom) PDFs between NNPDF4.0 (green), feature scaling (orange), CT18 (blue), and MSHT20 (pink) [13].

1.2.2 The factorization theorem

The factorization theorem allows to separate the cross-section of pp collisions into a term that depends on the PDFs and a short-distance parton cross-section term.

Considering the collision between two hadrons with four-momenta p_1 and p_2, the cross-section can be written as:

$$\sigma(p_1, p_2) = \sum_{i,j=q,\bar{q},g} \int dx_1 \int dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij} \left(x_1 p_1, x_2 p_2, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right) \tag{1.15}$$

where f_i and f_j are the PDFs, μ_F is the factorization scale and μ_R is the renormalization scale.

$\hat{\sigma_{ij}}$ represents the partonic cross-section for the interaction between the two partons and it can be expanded perturbatively as

$$\hat{\sigma}_{ij} = \hat{\sigma}_{ij}^{LO} + \alpha_s \hat{\sigma}_{ij}^{NLO} + \alpha_s^2 \hat{\sigma}_{ij}^{NNLO} + \alpha_s^3 \hat{\sigma}_{ij}^{N^3LO} + O(\alpha_s^4) \tag{1.16}$$

where the leading-order, LO, is the lowest order in the pQCD calculation, followed by the next-to-leading order, NLO, the next-to-next-to-leading-order, NNLO, and so on up to a certain perturbative order.

1.2.3 Hadronization

After the parton-parton interaction, elementary particles are generated. Due to confinement, quarks and gluons cannot be observed directly in the final state; they form hadrons that cluster in jets [14].

Because of confinement, quarks and gluons cannot exist freely, so they combine to form color-neutral hadrons that retain the memory of the elementary particle from which they were generated. In high-energy collisions, quarks and gluons are produced with very high energies and they can emit additional gluons, which in turn produce quark-antiquark pairs.

The process in which quarks and gluons are bound into colorless objects is known as *hadronization*. The hadrons produced tend to move in collimated directions, which reflect the momentum and direction of the original partons from which the hadronization started.

1.3 Monte Carlo simulations

The simulation of events in particle physics is a complex yet essential task, vital for comparing experimental data with theoretical predictions. While fixed order calculations in pQCD are available only up to a limited order in α_s and just focus on the hard-scattering aspect providing results at parton level, MC simulations are computational models designed to replicate particle collisions as observed by detectors, indeed they aim to the complete simulation of the processes.

A typical hadron-hadron collision can be broken down into three key elements:

1- the partonic distribution function that provide the initial state of the interacting partons;

2- the strong interaction between the two partons;

3- the fragmentation/hadronization of partons to form final state.

Typically, the initial scattering between two partons is described by a matrix element (ME), which calculates the cross-section of the process at a given order in pQCD. In modern MC generators, there are NLO (Next-to-Leading Order) MEs that can handle final states with multiple partons.

However, after the scattering process, the produced partons emit multiple low-energy and collinear gluons, which can subsequently decay into quark-antiquark pairs. This phenomenon, simulated in MC generators through the so-called "Parton Shower" (PS), allows the inclusion of higher-order effects that are not captured in the ME.

Another crucial aspect of MC generators is the description of the hadronization process, through which partons transform into hadrons. This process is modeled using phenomenological approaches.

There are two phenomenological models as shown in Fig. 1.6 that simulate the hadronization:

- the string model, also known as Lund model;
- the cluster model.

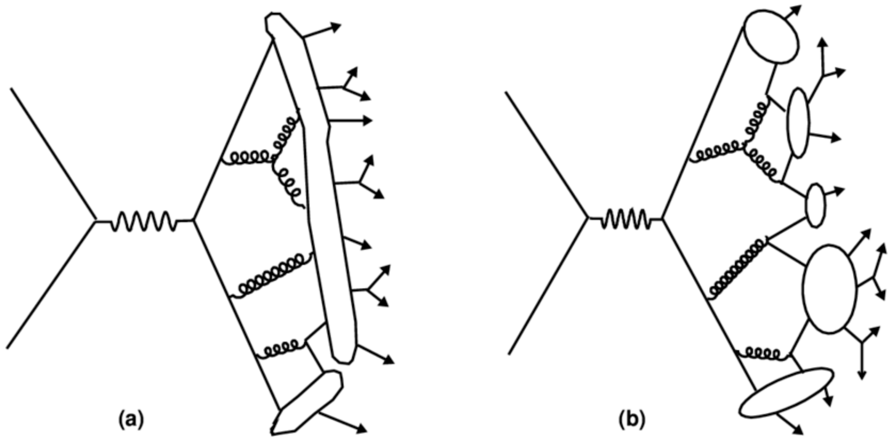

Figure 1.6: Comparison between string (a) and cluster (b) model hadronization.

The string model assumes that as the distance between two partons increases, the potential energy between them also rises. It becomes energetically more favorable to break the stretched "string" and create new segments, which form $q\bar{q}$ pairs. In this case, the gluons produced give rise to kinks on the string.

The cluster model is based on the theory of preconfinement [14], which suggests that colorless clusters of quarks and gluons with finite mass are formed immediately after the parton shower. These clusters are relatively small and localized in space. If a cluster is too heavy, it can split into lighter clusters or it may combine with other clusters to form stable hadrons.

In addition to the description of scattering and hadronization processes, MC generators also account for the underlying event, which includes the softer interactions among the produced particles.

Together, the matrix element, parton shower, hadronization models and the underlying event provide a complete simulation of collision events. Programs like `Pythia`, `Herwig`, `Sherpa`, and `Madgraph` [18] are examples of general-purpose MC generators widely used in LHC analyses, capable of realistically describing the behavior of all final-state particles observed in experiments.

`Phythia, Herwing` and `Sherpa` are general-purpose generators that model everything from the hard scattering to the hadronization phase.

`Phythia` uses a dypole-style parton shower, where gluons and quarks can radiate other partons, leading to jet formation. For the hadronization it uses Lund string model.

`Herwig` places emphasis on angular ordering in the parton shower, which is crucial for maintaining the angular structure of emissions and provides a more accurate description of jet substructure. For hadronization, `Herwig` utilizes the cluster model and is particularly effective for simulating heavy quark decays, making it suitable for studying charm and bottom quarks.

`Sherpa` distinguishes itself from `Pythia` and `Herwig` by integrating matrix element calculations with parton showers, offering a more precise description of complex final states. By combining higher-order matrix elements with parton showers, `Sherpa` avoids double-counting. For the hadronization simulation `Sherpa` uses the cluster model.

`MadGraph`, on the other hand, specializes in generating matrix elements for specific processes, particularly those involving potential new physics scenarios. The generated matrix elements are then passed to `Pythia` or `Herwig` for parton showering and hadronization.

`POWHEG` is a tool used to implement Next-to-Leading Order (NLO) calculations within parton shower simulations, providing high accuracy in incorporating higher-order corrections.

1.3.1 Usage of Monte Carlo simulations in high-energy experiments

MC simulations aid in event selection by refining criteria to isolate rare processes and in background estimation by modeling background processes to identify new physics signals. They help assess the efficiency of experimental components, including detectors and analysis algorithms. Simulations also enable the evaluation of systematic uncertainties by testing the impact of varying conditions. Lastly, as already said, they are crucial for data validation and calibration, ensuring the accuracy and reliability of experimental measurements by comparing simulated and actual detector data.

After event generation, detector simulation is typically carried out. ATLAS experiment, for example, uses software such as Geant4 [15] or AtlFast3 [16]. Geant4 is widely used due to its precision in modeling interactions between particles and detector materials. AtlFast3 offers a faster, albeit less precise, alternative for quick simulations.

Then the potential noise is simulated, including event pileup, to emulate the actual conditions of the detector readout.

In the reconstruction phase, which is quite similar to processing real data, various parameters are calculated and tracks, vertices and clusters are reconstructed to identify objects like particles. For example by analyzing the energy deposits in calorimeters, physicists can reconstruct the jets, which are groups of hadrons produced from the same initial parton, using specialized algorithms (discussed in the following chapter). This phase is critical for extracting meaningful physical information from the simulated data, ensuring that the simulations accurately reflect real experimental scenarios.

1.4 The production of the Z boson in association with b-jets

At the LHC, exploiting the large centre-of-mass energy of pp collisions and the large amount of data, as it is explained in the Chapter 2, it is possible to study in details processes not observed before or studied previously just inclusively. It is the case of the production of a Z boson in association with b-jets that is the main topic of this thesis.

The Feynman diagrams for this process at LO are shown in Fig. 1.7 and 1.8.

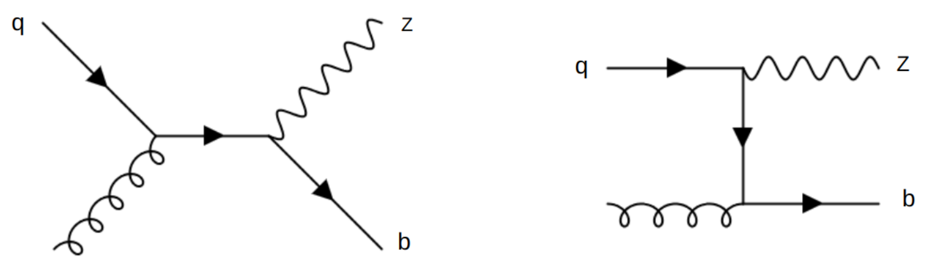

Figure 1.7: Feynman diagrams for the production of the Z boson in association with a b-jet.

The quarks discussed in Section 1.1 have different flavours, each with distinct masses, as shown in Table 1.2.

Due to these varying masses, quarks can be categorized as either light or heavy. Heavy quarks are those for which the strong coupling constant is in the perturbative region at their mass scale, specifically the bottom and top quarks.

Flavour schemes differ based on whether they consider the quark masses as negligible or significant.

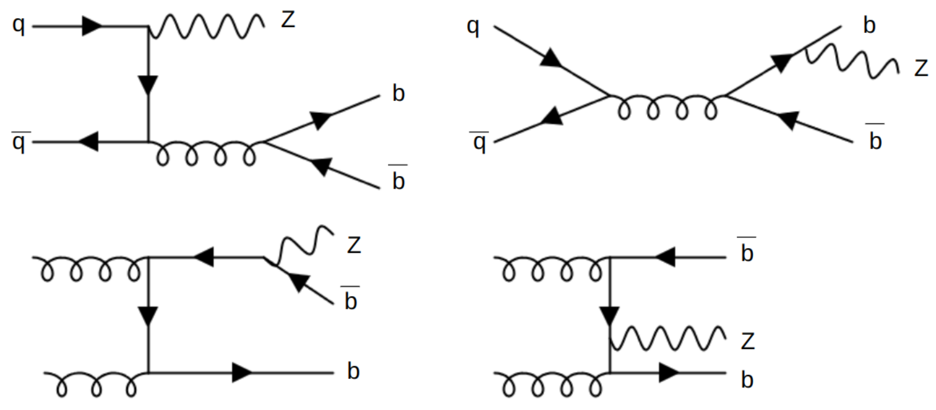

Figure 1.8: Feynman diagrams for the production of the Z boson in association with two b-jets.

Quark	**Simbolo**	**Massa (MeV/c^2)**
Up	u	2.2 - 2.3
Down	d	4.7 - 4.8
Charm	c	1280
Strange	s	96
Top	t	173100
Bottom	b	4180

Table 1.2: Masse approssimative dei quark.

In the four-flavour number scheme (4 FNS), the bottom quark is much heavier than the proton (mp = 938.27 MeV/c^2) , and as a result, it can only be produced in pairs through gluon splitting. Therefore, it is not considered a contributing factor in the determination of the PDFs.

The five-flavour number scheme (5 FNS) is applicable only when the energy scale is higher than the mass of the bottom quark. In this scenario, the bottom quark is treated as massless, allowing it to contribute to the initial state of the proton. Consequently, it is included in the calculations of the PDFs.

The signature measured in this thesis consists of two leptons with opposite charge and same flavour, originating from the decay of the Z boson, and two b-jets resulting from the hadronization of b-quarks.

These kinds of studies provide an important test of pQCD, applicable in the regime of very high energy and short distance interactions.

Moreover, they are crucial for understanding the contribution of heavy quarks to parton distribution functions (PDFs).

The production of Z+b jets is a significant background for many analyses at the LHC, including the study of the Higgs boson produced in association with a Z boson (ZH). In this process, the Z boson decays into leptons and the Higgs boson decays into bottom quark pairs (bb), as shown in the Feynman diagram in Fig.1.9.

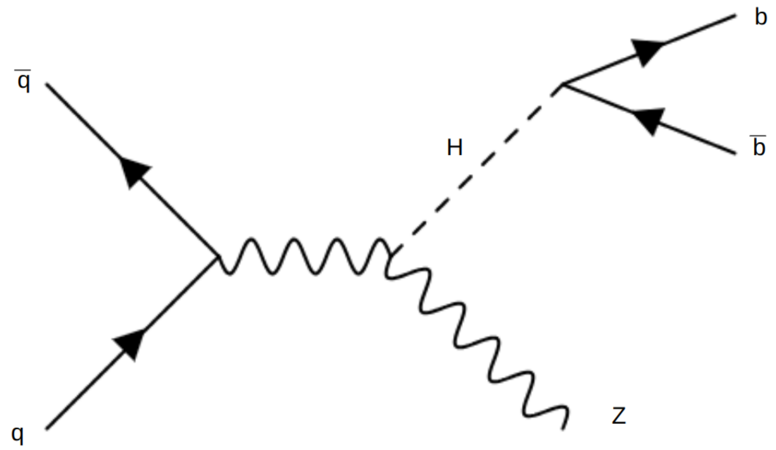

Figure 1.9: Feynman diagram for the production of the Higgs boson decaying in two b-jets in association with a Z boson.

The Higgs boson decay into a pair of bottom quarks is the dominant decay mode predicted by the SM for a Higgs boson with a mass of 125 GeV, which corresponds to the mass of the Higgs boson discovered at the LHC in 2012. This decay channel has a branching ratio (BR) of approximately 58%, as shown in Fig. 1.10. However, due to the large multijet background, this decay can only be studied in associated production at the LHC, where the signal-to-background ratio is more favorable. The downside is a significant reduction in the production rate.

In this context, the production of $W/Z + H(bb)$, where two b-jets are present in the final state along with the leptonic decay of the W ($W \to \ell\nu$) or the Z ($Z \to \ell\ell$), represents one of the leading topologies for studying Higgs physics at the LHC and the Higgs coupling to the b-quark.

This final state has been observed at the LHC, but, to improve the precision of these measurements, a better understanding of the W/Z+b-jets background is required, that indeed represents currently one of the main limit factor of the measurement [19].

The production of Z + b-jets faces several background processes, such as Z + c or light jets, Z boson production in association with an Higgs boson, $t\bar{t}$ events, single top events, diboson production, $Z \to \tau\tau$.

In the production of the Z boson in association with b-jets, the largest background is the production of the Z boson with jets originating from the

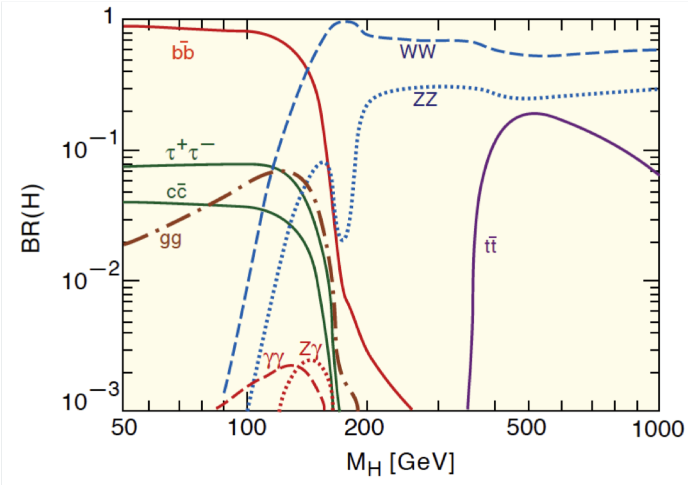

Figure 1.10: Branching ratios of the possible decay channels of the Higgs boson. [17]

hadronization of other quarks, different from the b-quark, such as c quarks or light quarks (up, down, and strange).

Another significant background is $t\bar{t}$ pair events, which are abundantly produced at the LHC. The top quark decays almost exclusively (close to 100%) into a W boson and a b quark. To obtain two leptons and b-quarks in the final state, the relevant decay channel is $t\bar{t} \to W^+bW^-\bar{b} \to l^+\nu bl^-\nu\bar{b}$, which has a branching ratio of 10%.

Single top production is also a background for Z + b-jets production and there are three different channels for this process: t-channel, s-channel and production in association with a W boson.

The production of the Z boson in association with a Higgs boson (ZH events) is another background process, as previously explained. The Z boson decays leptonically and the Higgs boson can decay into two b quarks, producing two b-jets.

Another important background is the diboson final state, where the bosons can be two Z bosons, two W bosons or a W and a Z boson—generally referred to as two vector bosons ($VV = WW, ZZ, WZ$).

1.4.1 State of art of Z + b-jets measurements

The inclusive and differential cross-sections for the production of Z bosons in association with b-jets have been studied in proton-antiproton collisions at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV by the CDF and D0 experiments [20, 21] and in pp collisions at $\sqrt{s} = 7$ TeV at the LHC by the ATLAS and CMS experiments [22, 23]. Additionally, the CMS experiment extended these measurements to $\sqrt{s} = 8$ TeV [24]. Both experiments performed studies at $\sqrt{s} = 13$ TeV using the full Run-2 dataset [25] [26].

The most recent ATLAS result at $\sqrt{s} = 13$ TeV full Run-2 dataset measured both the inclusive and differential cross-sections for Z boson production (decaying into electrons or muons) in association with at least one or at least two b-jets, or at least one c-jet. In this measurement, jets are required to have a transverse momentum of $p_T > 20$ GeV and rapidity $|y| < 2.5$. This phase space closely resembles earlier measurements at $\sqrt{s} = 7$ TeV and those conducted by the CMS experiment, commonly referred to as the resolved regime.

In this measurement only few differential cross-sections have been measured in the final state with at least 2 b-jets. A previous ATLAS measurement [28] was performed at $\sqrt{s} = 13$ TeV, using a partial dataset where a larger set of observables in the final state with at least 2 b-jets were presented. In addition, ATLAS provided measurements of both inclusive and differential cross-sections in the boosted regime using a partial dataset [27]. These measurements are compared with state-of-the-art MC predictions for Z + b-jets production that utilize different flavor and mass schemes. These predictions typically combine next-to-leading-order (NLO) matrix elements with parton showers (PS).

The primary aim of this thesis is to provide differential cross-sections of the Z boson in association with at least two b-jets in the resolved regime. These measurements utilizes pp collision data from the ATLAS experiment at $\sqrt{s} = 13$ TeV, using the complete LHC Run-2 dataset (140 fb^{-1}). Compared to the previous ATLAS measurement at $\sqrt{s} = 13$ TeV in the resolved regime, this work provides measurements as a function of variables not included in the latest ATLAS publication on the full dataset.

Chapter 2

The ATLAS experiment at LHC collider

CERN [29] (European Council for Nuclear Research) is the one of the larger laboratory in the world for particle physics and it is based in Geneva, Swiss. At CERN is installed the Large Hadron Collider (LHC) [30], that is the world's largest and most powerful particle accelerator. The LHC is designed to collide protons at a maximum centre of mass energy $\sqrt{s} = 14\ TeV$. The very high energies at which LHC works are reached through an accelerator chain as shown in the Figure 2.1 .

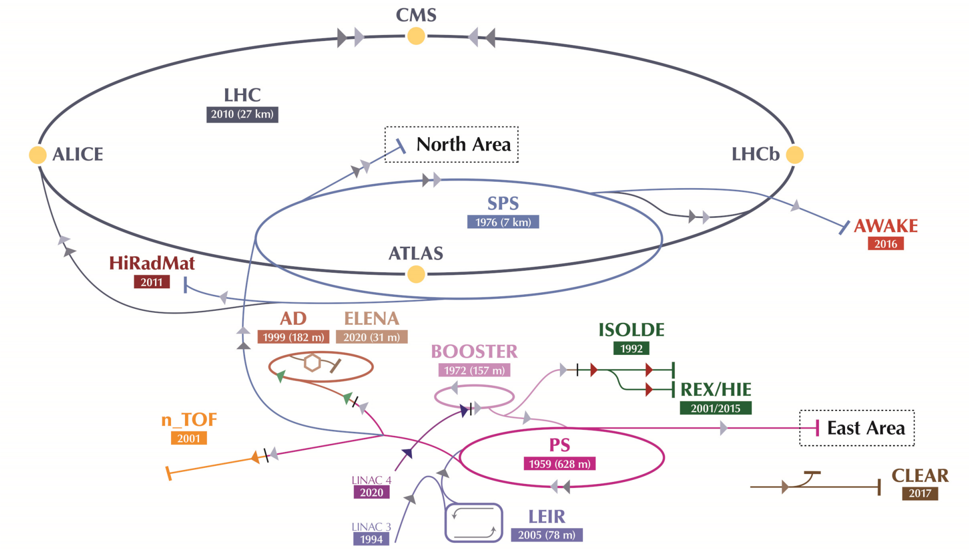

Figure 2.1: CERN accelerator complex.

The accelerator complex at CERN starts from LINAC4 that, since 2020, is the source of the proton beams. Linac4 substituted LINAC2. It accelerates negative hydrogen atoms up to 160 MeV. The ions are than injected in the Proton Synchtotron Boosted where they're divided by their electrons, leaving only protons. Before to be injected into the Proton Synchrotron they're accelerated up to 2 GeV and then up to 26 GeV. In the Super Proton Synchrotron they're accelerated up to 450 GeV. Then the protons are transferred to LHC.

2.1 The Large Hadron Collider

The LHC [32] has a circumference of 27 km and it collides protons or heavy ions, in particular lead nuclei. LHC completed two data-taking campaigns: Run 1 (up to 2012) where the proton collision energy was $\sqrt{s} = 7$ and $8\ TeV$, Run 2 (from 2015 to 2018) where the proton collision energy was $\sqrt{s} = 13\ TeV$. Since 2022, the Run 3 campaign is on-going at $\sqrt{s} = 13.6\ TeV$.

In this work the data collected in Run 2 are used.

In the accelerator, particles circulate in a vacuum tube in order to reduce the probability of collisions with air molecules. The trajectory is manteined by dipole magnets, while quadrupole magnets keep focus the beam. The accelerating cavities are electromagnetic resonators that accelerate particles.

In order to create very high magnetic field to have a circular shape 27 km long, superconducting electromagnets are employed. The dipoles become superconducting below a temperature of 10 K. To have a magnetic field of almost 8.3 T, the LHC operates at 1.9 K. These temperatures are reacheble with a superfluid helium cooling system.

The LHC is made of eight arcs and eight insertions. The arcs contain the dipole bending magnets, while the insertions are long straight sections, each one with two transition regions. The sector is the region between two insertion points and at each sector there are different working units that take care of magnet installation, hardware, powering.

The proton beam at LHC is divided in almost 2808 bunches with 25 ns bunch spacing in Run 2. The number of particles per bunches is of the order of 10^{11}.

LHC accelerates two particles beams in opposite directions before colliding them in four interaction points distributed along its ring. At each interaction point, a detector is housed. ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are the two multipurpose experiments of LHC, while two specific experiments LHCb (Large Hadron Collider beauty) and ALICE (A Large Ion Collider Experiment) are focused respectively on studies of the b-quark and of the heavy ion collisions.

In Table 2.1 the main collider parameters are collected.

Parameter	Value
Tunnel Length	27000 m
Magnet Cooling Temperature	1.8 K
Number of Main Dipole Magnets	1200
Quadrupole Magnet Count	860
Radio Frequency Cavities	10 per beam
Spacing Between Bunches	30 ns (or multiples)
Magnetic Field at Maximum Energy	8.5 T
Target Luminosity	10^{35} cm^{-2}s^{-1}
Beam Energy, Protons	8, 13 TeV (Run 2), 14 TeV (Run 3)
Bunch Collision Frequency	50 MHz
Maximum Bunches Per Beam	2900
Protons Per Bunch	$1.2 \cdot 10^{11}$ (30 ns) or $1.7 \cdot 10^{11}$ (60 ns)
Total Collision Events Per Second	700 million
Crossing Angle at Interaction Points	310 μrad

Table 2.1: Main Parameters of the Collider

2.1.1 The Luminosity at LHC

To assess the performance of the accelerator, a useful quantity is the instantaneous luminosity L, defined as the number of interactions per unit of time and area

$$L = \frac{1}{\sigma}\frac{dN}{dt} \tag{2.1}$$

where $\frac{dN}{dt}$ is the event rate of a given physics process and σ is the cross-section. The instantaneous luminosity is expressed in units of $cm^{-2}s^{-1}$.

Indeed, the integrated luminosity is the number of interactions per unit of area

$$L = \int L \ dt \tag{2.2}$$

In the case of a collider, the instantaneous luminosity is measured as

$$L = f_r \ n_b \ \frac{N_1 N_2}{4\pi\sigma_x\sigma_y} \tag{2.3}$$

where f_r is the beam revolution frequency, n_b is the number of bunches in each beam, N_1 and N_2 are the number of protons in each beam, σ_x and σ_y are the gaussian transverse profile of the beams. This formula is true if the collision is head-on, otherwise, it's necessary to multiply by a factor S that depends on the crossing angle.

During Run 1 LHC delivered to ATLAS 28 fb^{-1} of data, 25 fb^{-1} of the total provided amount of data was usable for physics analyses. During Run

2 the delivered data were 156 fb^{-1} of which 140 fb^{-1} were good for physics analyses. The expected luminosity that LHC will deliver by the end of Run 3 is about 300 fb^{-1}.

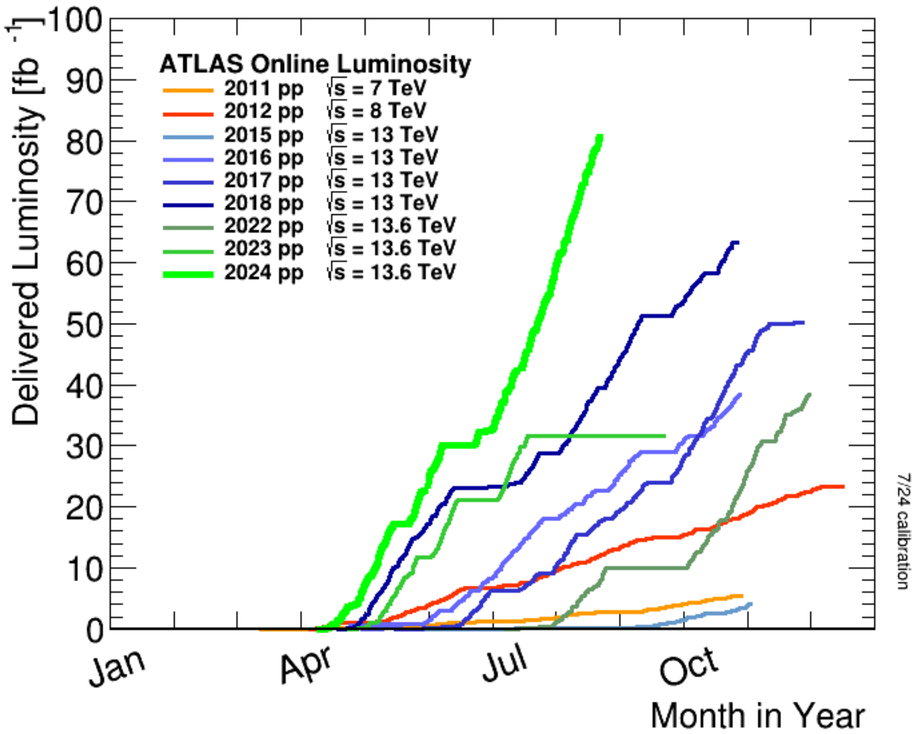

Figure 2.2: Integrated luminosity in function of data taking period.

The complete formula for the luminosity, considering also these parameters, is

$$L = \frac{N_1 N_2 n_b f \gamma}{4\pi\epsilon_n \beta*} F \tag{2.4}$$

where N_1 and N_2 are the number of protons in each beam, n_b is the number of bunches, f is the revolution frequency, γ is a factor that increases with the beam energy, ϵ_n is the normalized emittance, β^* is the beta function at the interaction point, F is the geometric reduction factor.

The luminosity is also influenced by pile-up and underlying events. Pile-up is defined as the mean number of pp interactions per bunch crossing, denoted as $\langle\mu\rangle$. Pile-up poses a challenge for detectors because it is desirable to have a low pile-up in order to collect events originating from a single interaction. However, with increasing luminosity, the pile-up also increases.

In Fig. 2.3, the mean number of interactions per bunch crossing is shown.

During Run 2, the average number of interactions per bunch crossing, $\langle\mu\rangle$ (pile-up), was 33.7.

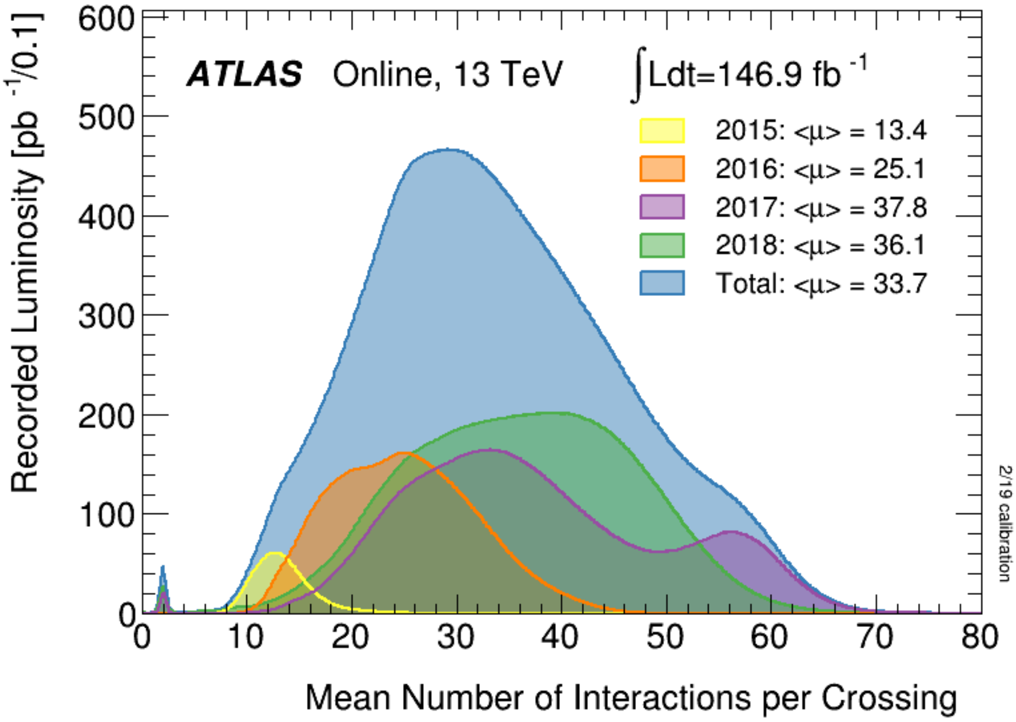

Figure 2.3: Integrated luminosity as a function of the mean number of interactions per crossing.

2.2 The ATLAS detector

The ATLAS detector [31] (in Fig. 2.4) is a general-purpose detector and it has a cylindrical shape. It is placed in one of the four interaction points. ATLAS is an hermetic detector, all around the interaction point and is forward-backward symmetric. It covers a solid angle of nearly 4π steradians. It is 44m long, its diameter is 25 m and it weights 7000 tonnes.

The ATLAS detector is made of subsystems: Inner Detector, Calorimeters, Muon Spectometer and a Magnet System.

The expected resolution of the ATLAS subdetectors are summarised in Table 2.2

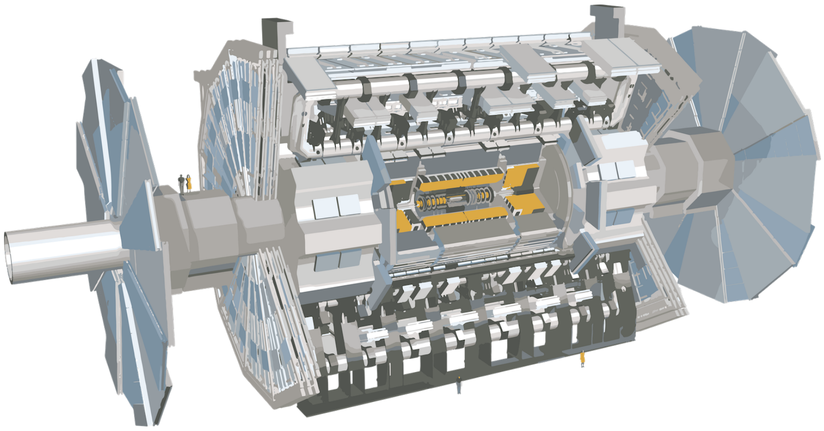

Figure 2.4: Representation of the ATLAS detector.

Detector component	**Required resolution**
Tracking	$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\%$
EM calorimetry	$\frac{\sigma_E}{E} = 10\%/\sqrt{E} \oplus 0.7\%$
Hadronic calorimetry (barrel)	$\frac{\sigma_E}{E} = 50\%/\sqrt{E} \oplus 3\%$
Muon spectrometer	$\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1$ TeV

Table 2.2: Expected resolutions of the ATLAS sub-detectors

2.2.1 The ATLAS reference frame

The ATLAS coordinate system is right handed: the $x-$ axis points to the center of the LHC ring, the $z-$ axis runs along the tunnel while the $y-$ axis points up. The origin of the coordinate system is the geometric center of the detector. Polar coordinates are also used in the (R, ϕ) plane. The angle ϕ is the one in the transverse plane between the x and y axes, and its formula is

$$\phi = \arctan\left(\frac{p_y}{p_x}\right)$$

where p_y and p_x are the momenta along the x and y axes. The angle θ is the angle between the z-axis and the transverse plane. Another fundamental quantity is the rapidity y, which is defined as

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$$

In the relativistic limit, where the particle mass can be neglected, a more convenient quantity called pseudorapidity η is used, which is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$

To parameterize the track, the following parameters are used:
- the polar angles ϕ_0 and θ_0;
- the transverse and longitudinal impact parameter z_0 and d_0;
- the ratio between the charge and the momentum of the particle.

2.2.2 Magnet system

The magnet system [37] is a fundamental component of the ATLAS detector. The ATLAS detector primarily consists of two main magnet systems: a solenoid and a barrel toroid, along with two end-cap toroids. A schematic representation of the ATLAS magnet system is shown in Fig. 2.5.

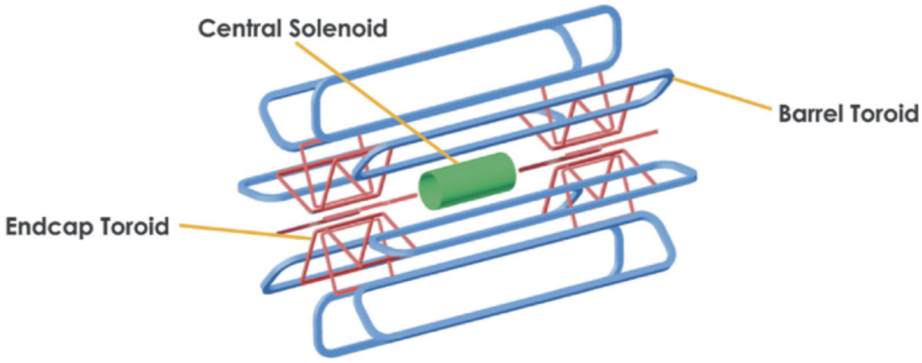

Figure 2.5: ATLAS magnetic system made of solenoid and toroids.

The solenoid provides a 2 T axial magnetic field for the Inner Detector, aligned with the beam axis. It is constructed from 9 km of niobium-titanium superconducting wires embedded in pure aluminum strips and it weighs over 5 tonnes. The solenoid is 5.8 m long, with a diameter of 2.56 m and a thickness of 4.5 cm.

The barrel toroid surrounds the central region of the experiment and is composed of eight coils. It is constructed from 56 km of superconducting wire, measuring 25.3 m in length and 20.1 m in diameter, with a total weight of 830 tonnes. The barrel toroid generates a magnetic field of approximately 0.5 T.

The two end-cap toroids provide a magnetic field of nearly 1 T for the Muon Spectrometer. The conductors in these toroids are made of aluminum, niobium-titanium and copper. Each end-cap toroid is 5.0 m long, with an outer diameter of 10.7 m and weighs 240 tonnes.

2.2.3 Inner Detector

The Inner Detector (ID) [33] (in Fig. 2.6 and 2.7) is the closest detector to the interaction point. It is immersed in a 2 T magnetic field generated by central solenoid. This detector provides pattern recognition, momentum resolution and both primary and secondary vertex measurements for charged tracks. This applies to particles above a specific transverse momentum (p_T) threshold and whithin a pseudorapidity range of $|\eta| < 2.5$. The cylindrical central region provides a full coverage for $|\eta| < 1.5$. The disks in the two endcaps covers a pseudo-rapidity range of $1.5 < |\eta| < 2.5$.

The ID is made of 3 independent sub-detectors:

- Pixel Detector, made of 4 layers in the barrel and 3 layers in each of the two endcaps;
- Semiconductor Tracker, made of 4 barrel strip layers and 9 endcaps;
- Transition Radiation Tracker, made of straw tubes in barrel and endcaps.

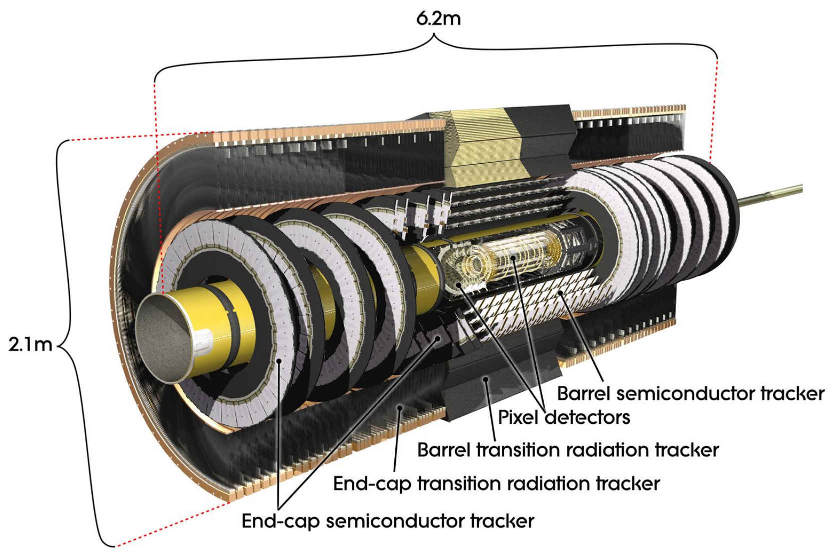

Figure 2.6: Internal structure of the Inner Detector that measures direction, momentum and charge of particles.

The Pixel Detector is the closest sub-detector to the vertex region and typically provides almost four measurements per track in the barrel. Its proximity to the interaction point makes it essential for reconstructing secondary vertices, which are crucial for identifying heavy quarks. The Pixel Detector is based on semiconductor technology that utilizes silicon pixels and strips. When a charged particle passes through the silicon wafer, it generates electron-hole pairs. Under the influence of an applied potential difference, the holes drift towards the p-n junctions, where they are collected. The four layers of the Pixel Detector achieve a resolution of 14 μm $(r-\phi)\times 115\mu m$ (z). The first of the four layers is placed at a distance of 3.3 cm from the LHC

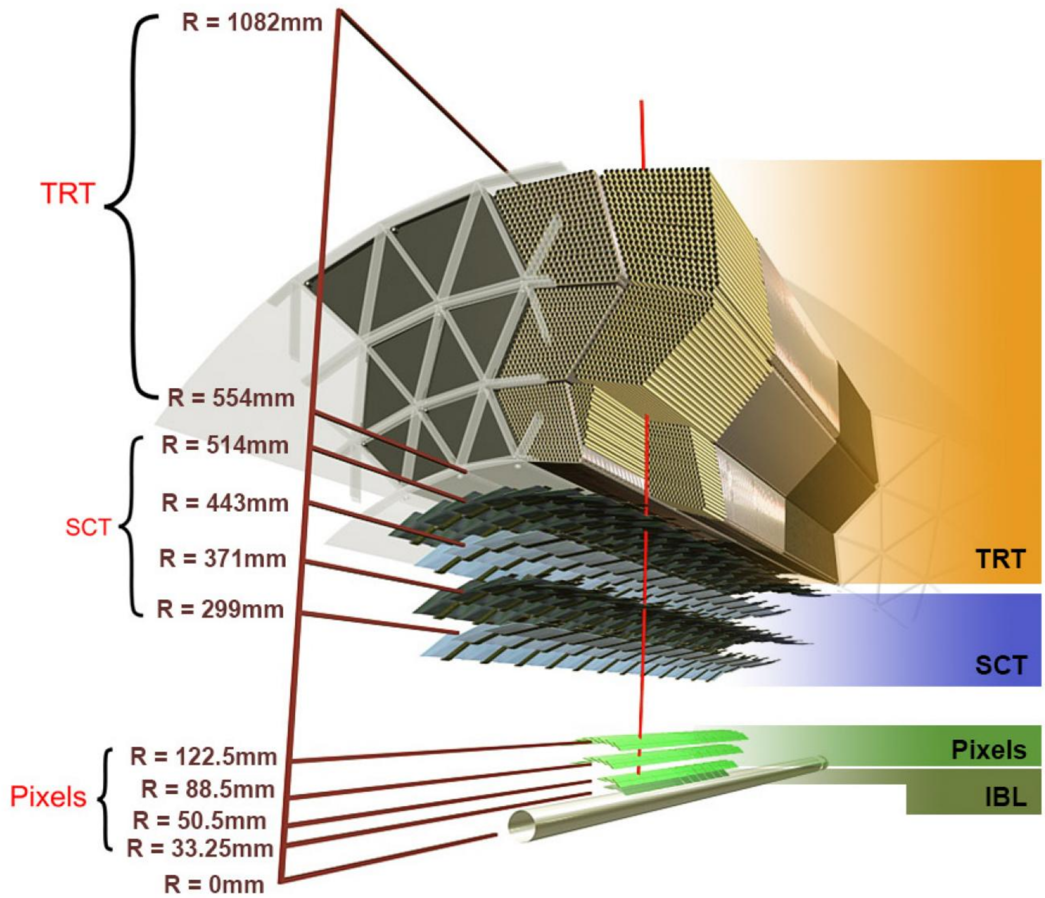

Figure 2.7: Trasversal representation of the Inner Detector.

beam line, while the other three at distances of 50.5, 88.5 and 122.5 mm respectively.

The Semiconductor Tracker (SCT) provides eight measurements per track in the barrel, with strips replacing the pixels used in the Pixel Detector. The silicon strip detector achieves a spatial resolution of $17\mu m$ in the transverse plane and $580\mu m$ along the z-axis.

The Transition Radiation Tracker (TRT) extends track reconstruction capabilities up to $|\eta| = 2.0$ with a resolution of $130\mu m$ in the transverse plane $(R - \phi)$ in the barrel and a resolution of $130\mu m$ along the z-axis in the endcap. The TRT also aids in electron identification by detecting the fraction of hits with an energy deposit above the transition-radiation threshold.

2.2.4 The Calorimeters

In the ATLAS detector, there are two types of calorimeters: electromagnetic and hadronic. The ATLAS calorimeters are positioned around the Inner Detector, providing coverage up to $|\eta| < 4.9$.

Calorimeters are detectors designed to measure the energy of particles passing through them by absorbing their energy and converting it into measurable signals, which are proportional to the energy deposited. In addition to measuring energy, calorimeters can also provide information about the type of particle, thanks to their segmentation and fine granularity, which allows for the tracking of particle trajectories. The measurement process in calorimeters is destructive, as they absorb the entire energy of the particle.

A sampling calorimeter, as the one employed in ATLAS, consists of two main components: the absorber and the active material. When a particle enters the calorimeter, it interacts with the absorber material, producing a cascade of secondary particles, known as a particle shower. This shower is then detected by the active material, which generates the measurable signals.

The Electromagnetic Calorimeter

The Electromagnetic Calorimeter [34] (in Fig. 2.8) is designed to absorb and detect electrons, positrons and photons. It is also called Liquid Argon (LAr) Calorimeter because the detection medium used is Liquid Argon, which is interspersed with lead plates that act as the absorber.

The Electromagnetic Calorimeter consists of three sections: a central barrel section (covering $|\eta| < 1.475$) and two endcaps (covering $1.375 < |\eta| < 3.2$). These sections provide measurements in different regions of pseudorapidity. The central part features a distinctive accordion geometry, which ensures excellent azimuthal coverage over the entire range of the ϕ angle.

The Hadronic Calorimeter

The Hadronic Calorimeter (in Fig. 2.9) measures the energy of showers produced by hadronic jets. It consists of three different sections: the Tile Calorimeter, two Liquid Argon (LAr) Hadronic End-Cap Calorimeters (HEC) and the LAr Forward Calorimeter (FCal), which is composed of three layers.

In the Tile Calorimeter [35], particles interact with steel plates, producing hadronic showers. The scintillating material within the calorimeter generates photons when these showers occur. These photons are then converted into an electric current, which is measured and is proportional to the energy of the particles. The Tile Calorimeter covers the pseudorapidity region $|\eta| < 1.7$.

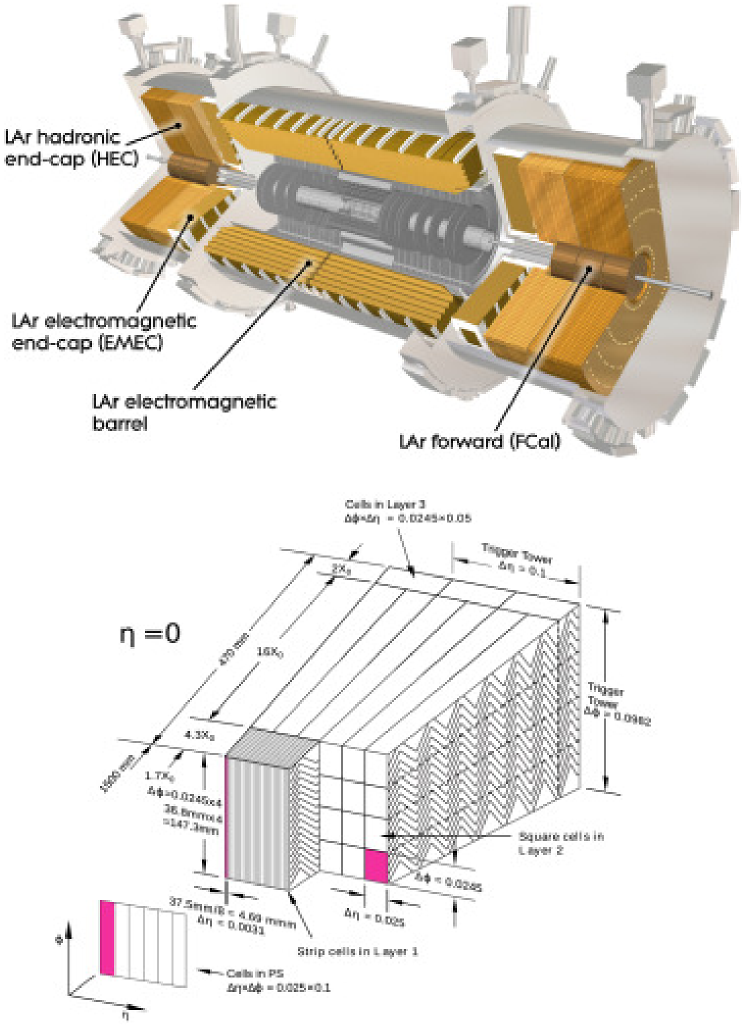

Figure 2.8: Representation of the electromagnetic calorimeter (up), section of the eletromagnetic calorimeter that shows the accordion geometry (down).

The HEC [34] covers the region $1.5 < |\eta| < 3.2$, while the FCal covers $3.1 < |\eta| < 4.9$. Both the HEC and the first layer of the FCal (FCal1) use copper as the absorber material and liquid argon as the detection medium. The second and third layers of the FCal (FCal2 and FCal3) use tungsten as the absorber, also with liquid argon as the detection medium.

2.2.5 The Muon Spectrometer

A specific detector is used to measure the momentum of muons and identify them. This detector is called Muon Spectrometer [36] (in Fig. 2.10). The

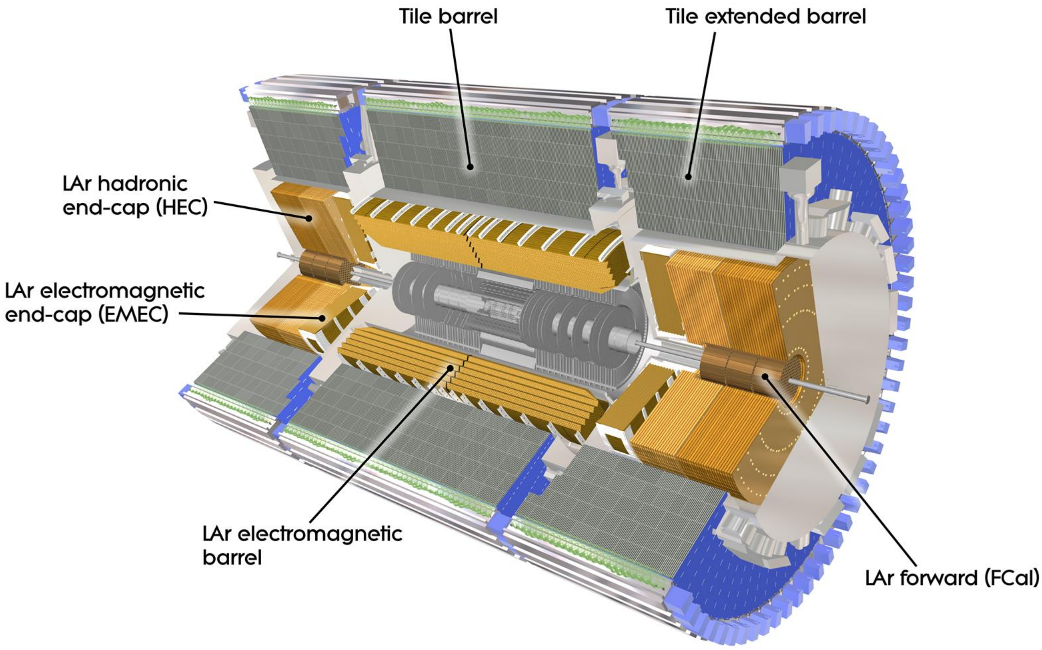

Figure 2.9: Representation of the hadronic calorimeter

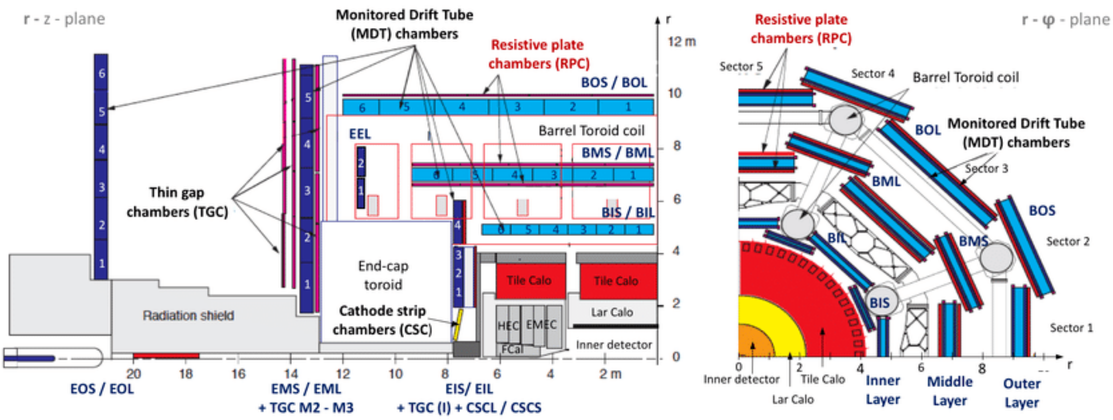

Figure 2.10: Grafical representation of the ATLAS Muon Spectrometer: in the $r - z$ plane (left), in the transverse $x - \phi$ plane (right).

spectrometer is divided into three parts: a barrel and two endcaps. It is composed of several sub-detectors, such as Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs), which provide precise measurements of the muon momentum. Additionally, Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) are employed for fast response in online triggering and have coarser granularity.

The barrel is made of three concentric cylindrical layers (station) at radii of 5 m, 7.5 m and 10 m around the beam axis. In the transition and end-cap regions the chambers are situated at 7.4 m, 10.8 m, 14 m and 21.5 m along the $z-$ axis relative to the geometric center and they're installed vertically.

Superconducting toroidal magnets are used to detect the trajectories of the particles. The large barrel toroid covers the region with pseudorapidity $|\eta| < 1.4$, while the two endcaps cover the regions $1.6 < |\eta| < 2.7$.

The muon spectrometer aims for a momentum resolution $\Delta p_T/p_T < 1 \times 10^{-4} \times p/\text{GeV}$ for $p_T > 300$ GeV. At lower momenta, the resolution is limited by multiple scattering and energy loss fluctuations. To achieve the required resolution, a three-point measurement must have each point measured with an accuracy better than $50\,\mu$m, setting the requirements for the muon chambers' design and stability.

Monitored Drift Tube (MDT) chambers are used in most regions, offering a resolution of approximately $80\,\mu$m. In regions with high pseudorapidity ($|\eta| > 2$), Cathode Strip Chambers (CSCs) are deployed to handle the high rates and background conditions. The Thin Gap Chambers, instead, covers a pseudorapidity region of $1.05 < |\eta| < 2.4$ while the Resistive Plate Chambers (RPC) have a psudorapidity range of $|\nu| < 1.05|$.

The detection elements of the MDT chambers are aluminium tubes of 30 mm diameter and 400 μm wall thickness, with a 50 μ m diameter central wire. These tubes operate with a non-flammable Ar-CH_4-N_2 gas mixture at a pressure of 3 bar. Each tube provides a single-wire resolution of ~ 80 μm. In the inner station, tubes are arranged in 2×4 monolayers and in the middle/outer stations in 2×3 monolayers, to improve resolution. The tubes are mounted on spacer frames that ensure mechanical stability and alignment.

The muon momentum is measured by the deflection caused by the magnetic field, with sampling occurring in the three layers of the Monitored Drift Tubes (MDTs). When a muon passes through the tube, the electrons produced by ionization drift toward the wire.

The Cathode Strip Chambers (CSCs) complement the Monitored Drift Tubes in the regions $2.0 < |\eta| < 2.7$ and are located in the endcaps. These are multi-wire proportional chambers with cathode planes segmented into strips. In these strips the distance between anode and cathode matches the anode wire pitch. The precision coordinate is determined measuring the charge induced on the segmented cathode. With a wire pitch of 2.54 mm and cathode readout pitch of 5.08 mm, CSCs achieve spatial resolutions better than 60 μm.

The Thin Gap Chambers (TGCs) and the Resistive Plate Chambers

(RPC) are responsible for triggering in both the endcaps and the barrel region. They complements the measurement of the MDTs in the radial direction. The TGCs consist of two layers: one for low-p_T triggering and the other for high-p_T triggering. The RPCs are detectors composed of resistive plates with 2 mm gas-filled gaps, where an electric field of 4.9 kV/mm is applied between the plates. They offer excellent time resolution of 1.5 ns, making them ideal for trigger systems. This configuration allows avalanches to develop along ionizing particle tracks towards the anode plates. On the other hand, the TGCs use an anode wire at 2900V, placed between resistive cathode plates in a gas-filled chamber. TGCs also provide good timing resolution for particle detection.

2.2.6 Trigger and Data Acquisition System

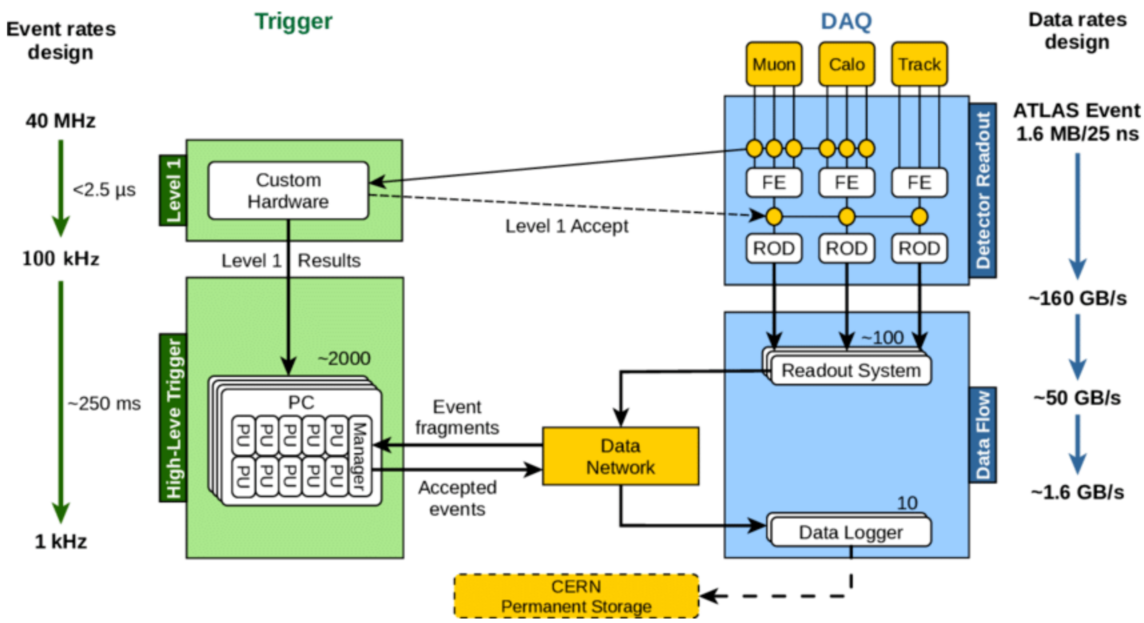

Figure 2.11: Schematization of the ATLAS Trigger and Data Acquisition system.

At ATLAS, a huge amount of data is collected, so the Trigger and Data Acquisition (TDAQ) system (in Fig. 2.11) [38] is essential for selecting only the data useful for physics analysis. This system reduces the event rate from 40 MHz to 1 kHz by discarding non-interesting events as quickly as possible. The TDAQ system consists of two trigger levels: the Low-Level Trigger (LLT) and the High-Level Trigger (HLT).

The **Low-Level Trigger** collects data from the calorimeters and the muon spectrometer to identify regions of interest (RoIs) in η and ϕ. It reduces the event rate from 40 MHz to 100 kHz within 2.5 μs. This is a hardware-based trigger.

The **High-Level Trigger** uses the RoIs identified by the LLT as input and processes the event data for full reconstruction. It is a software-based trigger that employs online selection algorithms to further reduce the event rate from 100 kHz to 1 kHz with a processing time of $\simeq$ 200 ms and an event size of about 2 Mbyte.

Events that pass through the trigger system are sent to CERN Tier 0 for offline reconstruction and saved as Analysis Object Data (AODs). These datasets are processed to produce different derived AOD (DAOD) formats, which contain a subset of events and reduced reconstruction information tailored for specific physics analysis and performance groups. These DAOD types are utilized by analysts who create condensed ntuples for further processing or final physics results. Additionally, simulated data are produced in AOD format and undergo further reductions, ultimately resulting in a ntuple format that is usable for physics analyses.

Data quality is assessed both online and offline to reject data affected by potential hardware or software issues. In the full Run-2 dataset, almost 94 % of the recorded events were deemed good for physics analysis [59].

2.3 Object Reconstruction and Identification in ATLAS

When particles pass through the detector and leave a signal, the next step is to reconstruct the full kinematics of the particle and perform its identification. The output of pp collisions consists of a series of electronic signals that must be reconstructed into physical objects. This section illustrates how the reconstruction of the physics objects needed to perform the measurements described in this thesis at ATLAS, namely the reconstruction of electrons, muons, jets , b-jets and missing transverse energy, used to identify the neutrinos.

2.3.1 Electrons

Electrons lose part of their energy due to bremsstrahlung when passing through materials, causing emitted photons that at high energy generate electron-positron pairs, leading to an electromagnetic shower. The reconstruction of electrons [39] in the ATLAS detector involves several key steps:

Energy Deposit in the Electromagnetic Calorimeter (EMCal): Electrons deposit energy in the EMCal, which is measured by summing the energy from each calorimeter layer. A sliding-window algorithm is used to identify localized energy deposits, forming clusters. This algorithm searches for nearby clusters to determine which should be kept based on energy criteria.

Tracks in the Inner Detector (ID): Charged particles, including electrons, leave tracks in the silicon pixel detector and the SCT detector, which offers three-dimensional measurements. Track reconstruction in the ATLAS ID begins by forming clusters of hits, which are then converted into three-dimensional space-points using pattern recognition and fitting algorithms. For tracks not meeting initial criteria, a Gaussian-sum filter (GSF) is applied. The GSF better accounts for energy loss due to bremsstrahlung, improving the accuracy of critical track parameters, such as the transverse impact parameter significance. This method provides more precise reconstruction for electron tracks compared to standard fitting techniques.

Track-Cluster Matching: The final electron reconstruction process involves matching track candidates to calorimeter seed clusters with strict criteria. The primary electron track is selected based on factors such as the distance in η and ϕ between the tracks and the cluster barycenters in the second layer of the calorimeter, the number of hits in the silicon detectors and the number of hits in the innermost silicon layer. Reconstructed clusters are formed by expanding around the seed cluster, using different window sizes for the barrel (3×7) and endcap (5×5) regions. The cluster energy is then calibrated using multivariate techniques after the selection of electron candidates to ensure accurate representation of the original electron energy.

Electron identification: Electrons are identified using a likelihood (LH) method. The LH identification method distinguishes between signal (prompt electrons) and background (jets mimicking electrons, electrons from photon conversions and non-prompt electrons) by calculating likelihoods for both signal (L_S) and background (L_B). These likelihoods are computed as the product of probability density functions (pdfs) and various x quantities

$$L_{S(B)}(x) = \prod_{i=1}^{n} P_{S(B),i}(x_i) \tag{2.5}$$

where $P_{S,i}(x_i)$ is the value of the signal PDF for the quantity i at value x_i and $P_{B,i}(x_i)$ is the corresponding PDF for the background.

The discriminant d_L is calculated as:

$$d_L = \frac{L_S}{L_S + L_B} \tag{2.6}$$

The discriminant d_L peaks sharply at one for signals and at zero for backgrounds. To facilitate selection, d_L is transformed using an inverse sigmoid function into $d_L^{'}$, creating a more convenient range of values. Indeed, the sigmoid function transforms continuos real numbers into a range of $(0, 1)$. Electrons with $d_L^{'}$ above a specific threshold are identified as signal. The LH-based identification is advantageous over cut-based methods because it integrates information from multiple discriminating variables, enhancing overall identification efficiency even if individual criteria are not met.

There are three operating points (OPs) that are commonly used in physics analyses and are defined based on the minimum value of the LH discriminant. These points are known as LooseLH, MediumLH and TightLH and are ordered in decreasing identification efficiency with an inversely proportional increase in electron signal purity.

In this analysis, electrons must meet the Tight likelihood requirement and satisfy the conditions of $p_T > 27GeV$ and $|\eta| > 2.47$. The isolation based on calorimeter measurements is characterized by the variable $E_{T,\text{ cone }20}/p_T$, where $E_{T,\text{ cone }20}$ represents the sum of transverse energy from energy clusters contained within a cone of $\Delta R = 0.2$ centered around the electron cluster that matches the track. Conversely, the track-based isolation is represented by the variable $p_{T,\text{ varcone }20}/p_T$, where $p_{T,\text{ varcone }20}$ is the scalar sum of the transverse momenta of tracks located within a cone that varies in size based on p_T and is defined as $\Delta R = \min(0.2, 10\text{ GeV}/E_T)$.

2.3.2 Muons

Muons are charged particles that leave tracks in the inner detector (ID), muon spectrometer (MS) and occasionally small energy deposits in the calorimeters. The muon reconstruction process [40] in the Muon Spectrometer begins with searching for hit patterns in each muon chamber to form segments.

In MDT (Monitored Drift Tube) chambers, a Hough transform is used to identify hits aligned along a trajectory and straight-line fits are applied to these hits. In CSC (Cathode Strip Chamber) detectors, separate searches are conducted in the η and ϕ planes to form segments.

Muon track candidates are constructed by combining hits from segments in different layers of the MS, starting with segments in the middle layers (where more trigger hits are available) and then extending to outer and inner layers. Segments are selected based on hit multiplicity and fit quality, with a minimum requirement of two matching segments to form a track.

Four reconstructed muon types are defined based on the algorithms and subdetectors used to reconstruct the muon tracks and they are described below:

- Combined (CB) Muons: Tracks are independently reconstructed in both the Inner Detector (ID) and Muon Spectrometer (MS), then combined using a global refit. Most CB muons follow an outside-in approach, starting in the MS and matching to ID tracks.

- Segment-tagged (ST) Muons: ID tracks are classified as muons if they match at least one MS track segment, useful for low-p_T muons or regions with reduced MS coverage.

-Calorimeter-tagged (CT) Muons: ID tracks are identified as muons if they match an energy deposit in the calorimeter with a minimum ionizing particle, this type improves acceptance where MS is less instrumented;

- Extrapolated (ME) Muons: Only MS tracks are used, extrapolated to the interaction point, useful for the high-$|\eta|$ region, $2.5 < |\eta| < 2.7$, where there is no ID coverage.

Similar to electron identification, muons are identified using certain discriminating variables with four identification working points:

- Loose, where the CB and ME muons are considered while the CT and ST are restricted to the $|\eta| < 0.1$ region. This identification criterium maximise the reconstruction efficiency while providing good-quality muon tracks.

-Medium, where only the CB and ME tracks are considered. This selection minimises the systematic uncertainties associated with muon reconstruction and calibration.

- Tight, that considers only CB muons with hits in at least two stations of the MS and satisfying the Medium selection criteria. This criterium is useful to maximise the purity of muons at the cost of some efficiency.

- High-p_T, that maximises the momentum resolution for tracks with transverse momentum above 100 GeV. The muons that pass the Medium selection and with at least three hits in three MS stations are selected.

These points vary based on the requirements for signal purity and background rejection.

In this analysis, muons must satisfy the *Medium* identification criterion, where only CB and ME muons are considered and they are required to have $p_T > 27\ GeV$ and $|\eta| > 2.5$.

2.3.3 Jets

Quarks and gluons are produced during proton-proton collisions and are captured in the hadronic calorimeter. The reconstruction and identification of jets, which are collimated bunches of energetic particles, are crucial and challenging aspects of ATLAS data analysis.

Jets are reconstructed from electromagnetic topo-clusters (EMTopo jets) using algorithms that ensure the process is infrared and collinear safe (Fig. 2.12), meaning that it is stable against soft (low-energy) emissions and collinear splittings, respectively.

This approach is also designed to be independent of the detector, fully specified with clear kinematic definitions and consistent regardless of the level at which the algorithm is applied.

At ATLAS, the anti-k_t algorithm [43] is used for jet reconstruction. This algorithm produces jets with a conical shape, making it different from other clustering algorithms that may result in more irregular shapes as shown in Fig. 2.13

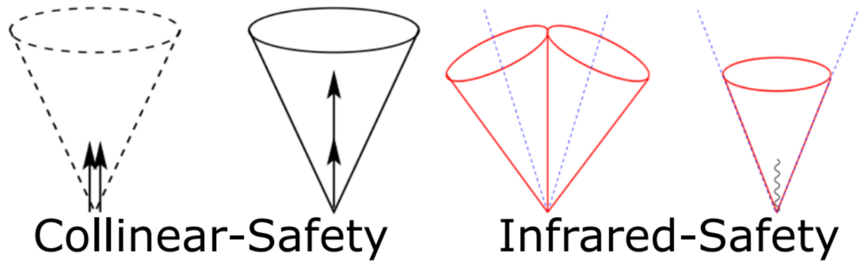

Figure 2.12: Schematic representation of the problem which is avoid requiring the jets to be Infrared and Collinear safe.

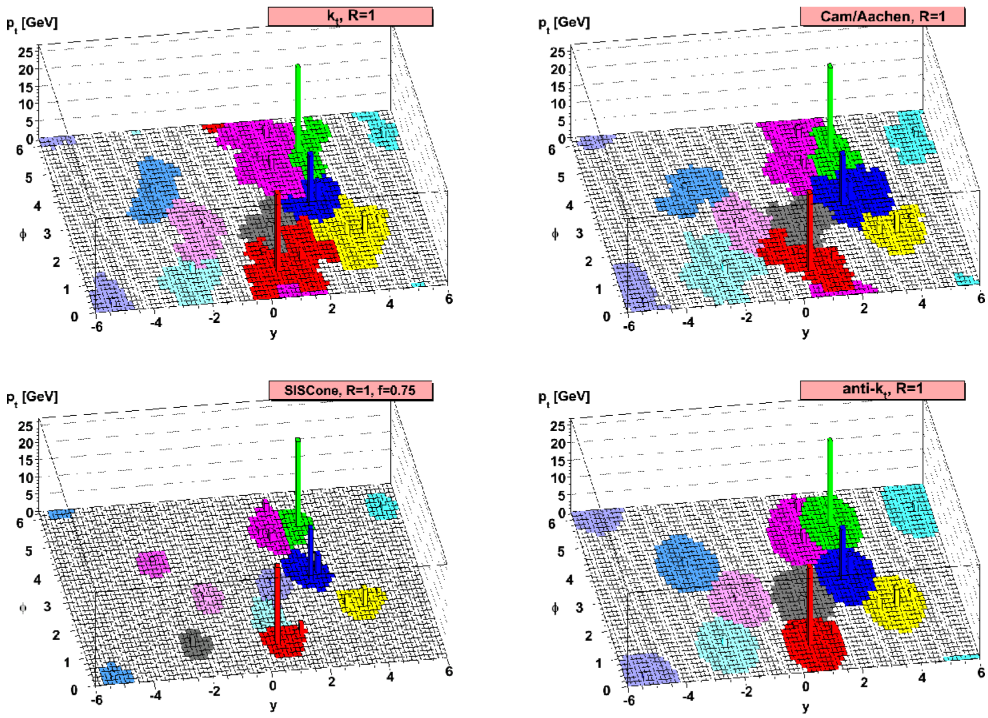

Figure 2.13: Comparison between different cone algorithm in p_T, ϕ, y coordinate system.

The anti-k_t algorithm calculates the distance between two objects (i and j) as:

$$d_{ij} = min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{2.7}$$

while the distance between the object i and the beam is

$$d_i = \frac{1}{p_{T,i}^2} \tag{2.8}$$

where p_T is the transverse momentum, R is the jet parameter and ΔR_{ij} is defined as

$$\Delta R_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{2.9}$$

with y_i the rapidity, ϕ_i is the azimuth angle.

If $d_{ij} < d_i$ the two objects are combined in a new one, while if $d_i < d_{ij}$ the object i is declared as a jet and it is removed from the particle list.

The procedure continues until no more particles remain to be examined.

The positive aspects of the anti-kt algorithm are that soft particles ($p_T \rightarrow 0$) result in $d \rightarrow \infty$, so they are clustered last reducing the impact of low-energy noise. Collinear particles have $d \rightarrow 0$, ensuring they are clustered first, maintaining the integrity of high-energy jets.

The jet radius parameter R determines the size of the jet cone and affects the jet reconstruction process. In this analysis:

- A radius of R=0.4 is used for the resolved region, where the aim is to resolve closely separated jets.

- A radius of R=1.0 is used for the boosted region, to capture broader jets.

The EMTopo jets are built from the energy deposited by charged particles in the calorimeter. In this analysis a new approach, called *particle flow*, combine the topo-clusters of charged particles, in addition to the topo-clusters of neutral particles, with the momenta of the tracks. These jets are called *PFlow jets* and they offers a better resolution for low-p_T jet, since for those jets the momentum resolution of the tracker is better than the energy resolution, and a better accuracy in jet reconstruction, indeed it allows to associate the tracks to vertices determing whether particles originate from pile-up vertices and eventually reject them.

There are some detector effects, such as the calorimeter's lack of compensation for hadronic energy deposits, dead materials and the escape of particles beyond the calorimeters, that effects the jet energy measurements. To consider them a calibration procedure, called Jet Energy Scale (JES) calibration, is applied and it alignes the energies of reconstructed jets with the energies of their stable particle constituents. [42]

This procedure is made of the following steps:

- **Origin Correction:** Each jet's four-vector is adjusted to point towards the primary vertex of the event rather than the detector's center. This adjustment enhances the angular resolution of the jet without significantly affecting its p_T.

- **Pileup Correction:** In densely populated environments, jet reconstruction may be influenced by pileup interactions, which can alter the area and energy of the reconstructed jets. To mitigate this effect, the p_T of redundant contributions is subtracted from the reconstructed jet p_T. These corrections depend on the number of reconstructed primary vertices, the jet's pseudorapidity and the bunch spacing.

- **Absolute Calibration:** The absolute JES and η calibration is obtained to adjust the discrepancy between the reconstructed jet energy and the particle-level jet energy. This calibration is derived from MC simulated samples of di-jet events.

- **Global Sequential Calibration:** The global sequential calibration (GSC) is applied in addition to the previous calibrations to address residual dependencies on various other jet characteristics. For instance, a significant dependence on JES is noted based on the jet's origin, as quark-initiated and gluon-initiated jets can exhibit variations in shape and particle composition.

- **In-Situ Calibration:** A final calibration is performed using real data in-situ to rectify differences between the MC simulated samples and actual data. Corrections are derived from dijet events, $Z + jets$, photon+jets and multijet events.

To differentiate between jets originating from the main primary vertex and those from pileup interactions, a multivariate analysis (MVA) discriminant known as the Jet Vertex Tagger (JVT) is utilized. The output of the JVT algorithm indicates the probability that a jet comes from the main primary vertex and this probability is employed for selection in physics analyses.

The selected jet criteria are $p_T > 20GeV$ and $|y| < 2.5$.

2.3.4 b-Jets

Flavor tagging is a crucial procedure applied after jet reconstruction and identification to determine the type of parton from which the jet originated. Based on this identification, jets can be classified as b-jets (originating from b-quarks), c-jets (from c-quarks), or light jets (from u, d, or s quarks).

To differentiate between these types of jets, specific properties are analyzed. A key indication of the jet's origin is the presence of heavy B and D mesons, which contain b and c quarks, respectively. In contrast, jets from u and d quarks are generally lighter and have longer lifetimes, leading to their decay outside the detector. Therefore, secondary vertices are typically not detected for light quarks.

For b and c quarks, however, the secondary vertices often decay within the detector's volume. The distance between the primary and secondary vertices is crucial for flavor tagging, as shown in Figure 2.14. This distance is measured using the longitudinal and transverse impact parameters, denoted as z_0 and d_0, respectively.

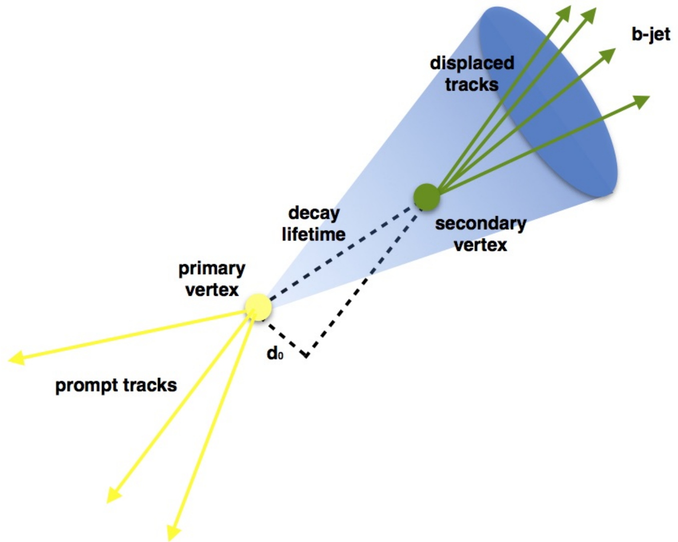

Figure 2.14: Representation of the longitudinal and transverse impact parameter (z_0 and d_0) that represent the distances between the primary and secondary vertices.

These impact parameters, along with the lifetime and other track information, are used as inputs for low-level algorithms. These algorithms generate preliminary tagging results. High-level algorithms then use the outputs from these low-level algorithms to calculate the probability that a jet is originated from a specific type of quark—whether light, b, or c.[43]

In this analysis the high-level algorithm used is the DL1r algorithm that uses a recurrent deep neural network approach. It is a more recent version that, with respect to the previous one, uses a recurrent neural network to estimate the impact parameter. The output scores from DL1r enable the establishment of various operational working points by applying different cuts on the scores. Each cut, and consequently each operational working point, correlates with a specific b-jet tagging efficiency. In order to determine the cuts a simulated $t\bar{t}$ sample is used. The defined operational working points cater to b-jet tagging efficiencies of 60%, 70%, 77%, and 85%, with the 60% efficiency representing the most stringent criteria and the 85% efficiency being the most relaxed. These operational working points can serve as bin edges, allowing the output of the DL1r b-tagging discriminant to be categorized into four intervals that correspond to defined ranges of b-tagging efficiencies: 85-77% (bin 1), 77-70% (bin 2), 70-60% (bin 3), and 60% (bin

4). For this analysis, the operational working point chosen is the one corresponding to a 70% efficiency.

The efficiency for each flavor is defined as

$$\epsilon^f = \frac{N^f_{pass}}{N^f_{all}} \tag{2.10}$$

where f represents the flavour, N^f_{pass} is the number of jets of flavour f tagged by the b-tagging algorithm and N^f_{all} is the number of all jets of flavour f.

To account for differences in tagging efficiency between data and simulations, a scale factor (SF) is determined by comparing the efficiencies:

$$SF^f = \frac{\epsilon^f_{data}}{\epsilon^f_{MC}} \tag{2.11}$$

This scale factor is then applied to the simulation results to bring them in line with real data, ensuring accurate jet flavor identification across the analysis.

Appropriate calibrations with their associated uncertainties are also established for mistag efficiencies. The calibration for b-jet events is obtained by comparing real data with $t\bar{t}$ MC samples that necessitate the presence of two leptons of opposite charge along with b-jets in the final state.

2.4 Missing Transverse Energy

In proton proton collisions the fraction of the proton energy that the partons possess during the collision is unknown. Because of this, the momentum conservation cannot be applied.

In the transverse plane the total initial momentum sum is zero, so, using the concept of missing transverse energy $E_{\text{miss},T}$, it is possible in this plane to apply momentum conservation laws.

The calculation of $E_{\text{miss},T}$ is given by the formula:

$$E_{\text{miss},T} = \sqrt{(E_{\text{miss},x})^2 + (E_{\text{miss},y})^2}$$

where:

$$E_{\text{miss},x} = -\sum_{i=1}^{N_{\text{cell}}} E_i \sin\theta_i \cos\phi_i$$

$$E_{\text{miss},y} = -\sum_{i=1}^{N_{\text{cell}}} E_i \sin\theta_i \sin\phi_i$$

In these equations, E_i represents the energy in the calorimeter cell, θ_i denotes the polar angle and ϕ_i refers to the azimuthal angle.

The missing transverse energy is so defined as the negative vector sum of the transverse momentum contributions from all identified hard physics objects.

Chapter 3

Z + b-jets measurements at detector level

One of the primary objectives of this work is to present detailed measurements of Z+b-jets at the detector level; this is indeed the first mandatory step for the cross-section measurement. Section 3.1 presents the strategy of the measurement at detector level. Section 3.2 presents the selected observables measured in this thesis. Section 3.3 describes the conditions under which the data were collected and the simulated samples used in this analysis. The object definitions and the event selection applied to data and the simulated samples are presented in Section 3.4. Section 3.5 is dedicated to the estimation of the backgrounds. Section 3.6 presents the estimation of the uncertainties at detector level. The kinematic distributions of the measured observables after the event selection at detector level are shown in Section 3.7, data are compared with simulations that comes from MC and data-driven techniques.

3.1 Strategy of the measurements

The events that contains an electron or a muon pair with leptons of opposite charge in association with at least two b-jets are selected at detector level.

After the selection, the events that contain a Z boson in association with b-jets, c-jets or light-jets are separated with a fit to flavour-sensitive observables. All other backgrounds are estimated via MC simulations or data-driven technique.

Once all the backgrounds have been estimated, a comparison between data and simulations is performed at three stages:

1- before the flavour fit, when both the signal and the background are taken directly from the simulation;

2- after scaling both signal and background in order to test the quality of the fit;

3- after scaling only the background while leaving the signal unscaled, allowing for a first evaluation of the impact of potential signal mismodelling.

In the second part of the analysis, described in the next Chapter, the data distributions are background subtracted and corrected for experimental effects, such as efficiencies and resolution, in order to provide the measured cross-section that are compared with the predictions of signal MC generators.

3.2 List of measured observables

A recent paper on the inclusive and differential cross-section measurements for the production of a Z boson in association with b- and c- jets has been published by the ATLAS Collaboration [26]. This paper presents results from the ATLAS experiment, specifically focusing on the Z + $\geq$ 1 b-jet, Z + $\geq$ 2 b-jets, and Z +$\geq$ 1 c-jet processes. However, the paper omits certain variables that are crucial for Higgs physics and the main goal of this thesis work is the measurement of these observables. The Higgs-boson decay into two b-quarks is the dominant decay for a Higgs boson of 125 GeV of mass, namely the mass of the Higgs discovered at the LHC. Due to the large multijet background this decay is predominantly studied in association with a W or a Z boson. In the case of association with a Z boson, events where a Z boson is produced is association with b-jets, leading to a signature similar to the Higgs boson one, represents a key background.

In addition, this thesis includes variables that are essential for improving the understanding of perturbative QCD and the parton distribution functions (PDFs) of the proton. This thesis focus on the Z + $\geq$ 2 b-jets final state and the measured observables are:

- The transverse momentum of the Z boson (p_T^Z), that is reconstructed from its decay products. It is a fundamental observable for Higgs boson studies, that is an analysis performed in different p_T regions of the Z boson.
- The rapidity of the Z boson, $|Y_Z|$. It is sensitive to the PDFs. Moreover both the $|Y_Z|$ and the p_{TZ} are essential for gaining deeper insights into QCD.
- The transverse momentum of the leading b-jet, the one with higher p_T, p_T^b.
- The rapidity of the lead b-jet, $|Y_b|$. Additionally, rapidity is critical for better understanding the PDFs of the proton.
- The azimutal angle between the two leading b-jets, $\Delta\phi$.

- The difference in rapidity between two leading b-jets,ΔY_{bb}, measured as

$$\Delta Y = |Y_1 - Y_2| \tag{3.1}$$

where Y_1 and Y_2 are the rapidities of the two leading b-jets. ΔY is related to the kinematic separation of the jets in the longitudinal direction and plays a key role in characterizing the event topology.

- The angular distance between two leading b-jets, ΔR, measured in the $\eta \times \phi$ space as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.2}$$

ΔR is related to gluon splitting, making it an important tool for studying this process. It is also relevant for Higgs studies, as the composition of the backgrounds varies with ΔR_{bb}.

- The invariant mass of the system composed of the two jets (M_{bb}) and it is defined as:

$$M_{bb} = \sqrt{(E_1 + E_2)^2 - |\vec{p_1} + \vec{p_2}|^2} \tag{3.3}$$

Both $\Delta\phi$ and M_{bb} have been previously presented in the ATLAS paper [26] and will be used in this thesis for cross-checking the analysis.

3.3 Data and simulated event samples

The data analyzed in this study were collected using the ATLAS detector during the years 2015 to 2018, achieving a total integrated luminosity of 140 fb^{-1} with an uncertainty of 0.83 % as detailed in [28].

Candidate events were selected using either a single-electron or single-muon trigger, which were subject to minimum transverse momentum (p_T) thresholds as well as quality and isolation requirements. Initially, the p_T threshold was set at 24 GeV for electrons and 20 GeV for muons. Due to the increased instantaneous luminosity over the course of data collection, these thresholds were subsequently raised to 26 GeV for both electrons and muons.

As outlined in Section 1.3, MC simulations play a crucial role not only in comparing experimental data with theoretical models but also in transitioning the data from detector level to particle level through the unfolding process, as well as in estimating systematic uncertainties.

The specific MC simulations employed for both signal and background processes in this study are summarized in Table 3.1.

For the simulation of the detector response, the GEANT 4 [15] toolkit was utilized. This ensures a realistic representation of how particles interact with the various components of the ATLAS detector. The pile-up is simulated with PYTHIA using A3 tune [45] and the NNPDF 2.3 LO PDF set [46].

Process	Generator	Order of pQCD in ME (FS)
$Z \to \ell\ell$	MGAMC+PY8 FxFx	0-3p NLO (5FS)
	SHERPA 2.2.11	0-2p NLO, 3-5p LO (5FS)
$t\bar{t}$	NNLO+NNLL	NLO
	POWHEG+PY8	NLO
single top ($s/t/Wt$-channel)		NLO
	POWHEG+PY8	NLO
$qg/q\bar{q} \to VV \to \ell\ell/\nu\nu + q\bar{q}$	SHERPA 2.2.1	1p NLO, 2-3p LO
$qq \to ZH \to \ell\ell/\nu\nu + b\bar{b}$	POWHEG+PY8	NLO
$gg \to ZH \to \ell\ell/\nu\nu + b\bar{b}$	POWHEG+PY8	NLO

Table 3.1: List of processes, generators, and the corresponding order of perturbative QCD in matrix element calculations (final state).

To simulate the signal $Z \to ll$, event generators such as MadGraph (MGAMC+PY8 FXFX) combined with Pythia, that is the generator used for the unfolding, as detailed in chapter 4, and SHERPA 2.2 11 were used [47].

MADGRAPH5 generates matrix elements for up to three partons at next-to-leading order (NLO) accuracy. The showering and the hadronization has been performed using Pythia8, using NNPDF 2.3 LO PDF set. The different jet multiplicities are merged using the FxFx prescription [48]. The PDF set used for event generation is NNPDF3.1luxQED with $\alpha_s = 0.018$. This sample is referred to as MGAMC+PY8 FXFX.

SHERPA 2.2.11 generates matrix elements at NLO accuracy for up to two partons and at leading-order (LO) accuracy for up to five partons. The PDF sets used are the Hessian NNPDF 3.0 NNLO [49], calculated at next-to-next-to-leading order in QCD with $\alpha_s = 0.018$. The different jet multiplicities are merged using MEPS@NLO prescription [50] [51] [52].

Both generators employ the five-flavor scheme (5FNS), where the b and c quarks are treated as massless and included in the proton's initial state.

The MGaMC+Py8 FxFx and the Sherpa signal samples as the background samples described in the next are produced using the full detector simulation described above. To test the 4FNS predictions, a sample of $Z+b\bar{b}$ events generated by $MadGraph5_aMC@NLO$ (v3.4.1) is produced using the 4FNS with massive b-quarks at NLO with two partons in the matrix element. It uses the 4-flavour NNPDF 3.1 NLO PDF set with $\alpha_s = 0.118$. $MadGraph5_aMC@NLO$ is combined with Pythia (v8.307) for the parton shower and hadronization. This sample is referred to as MG+Py8 FxFx

(4FNS). It isn't provided with full detector simulation there, for it is used only in the comparison between the measured unfolded cross-sections and the predictions in Section 4.3 and 4.4, as shown there, are currently available only for a subset of the measured observables.

Signal predictions from Sherpa, MG+Py8 FxFx and MG+Py8 FxFx (4FNS) are affected by theoretical uncertainties. The two main uncertainties affecting the predictions are the QCD scale uncertainties and the PDF uncertainties. QCD scale uncertainties have been evaluated using 7-point variations of the renormalization (μ_R) and factorisation (μ_F) scales coherently in the matrix elements and in the parton shower. The scales are varied independently by factors of 0.5 and 2 while avoiding cases where the factorization and renormalization scales move in opposite directions ($0.5 < \frac{\mu_R}{\mu_F} < 2$). PDFs uncertainties are evaluated using the Hessian eigenvector set for each PDFs set adopted in the predictions.

For the simulation of $t\bar{t}$ processes, PowhegBox v2 [54] [55] [56] generator was employed at NLO accuracy, using the NNPDF3.0 [49] PDF set. This sample was then interfaced with Pythia 8.230 [57], which uses the NNPDF 2.3 [46] LO PDF set.

For the single top quark production in association with W bosons, as well as in the s- and t- channels, the event simulation is done using PowhegBox v2 at NLO using the NNPDF3.0 PDF set interfaced with Pythia 8.230 using the NNPDF 2.3 LO PDF set.

Semileptonic diboson (VV) final states were simulated using Sherpa v.2.2.11 which provides NLO matrix elements for up to one parton emission and LO matrix elements for up to three parton emissions, calculated with the NNPDF 3.0 NNLO PDF set.

Processes involving the production of a Z boson in association with a Higgs boson (ZH) were modeled using Powheg interfaced with Pythia 8.230, with simulations carried out at NLO accuracy. NNPDF 3.0 NNLO PDF set has been used.

3.4 Object definition and Event selection

As described in Section 2.2.6, the Trigger and Data Acquisition system plays a crucial role in selecting only the datasets suitable for physics analysis. The list of runs where all parts of the detector functioned properly are included in the Good Run List (GRL) and only the runs of the GRL are used in the analysis.

Muon candidates are required to meet the *Medium* quality selection and pass the *FCTight* isolation criteria, as outlined in Section 2.3.2. These requirements help reduce background contamination from muons produced in hadronic decays. The tracks of selected muons must be associated with the primary vertex, with specific constraints on both the transverse and longi-

Object definition	**Electron channel**	**Muon channel**
Leptons	Single electron trigger Tight Isolated $d_0/\sigma_{d_0} < 5, \lvert z_0 \sin(\theta)\rvert < 0.5\,\text{mm}$ $p_T > 27\,\text{GeV}$ $\lvert\eta\rvert < 1.37\,\text{or}\,1.52 < \lvert\eta\rvert < 2.47$	Single muon trigger Medium Isolated $d_0/\sigma_{d_0} < 3, \lvert z_0 \sin(\theta)\rvert < 0.5\,\text{mm}$ $p_T > 27\,\text{GeV}$ $\lvert\eta\rvert < 2.5$
Jets	$p_T > 20\,\text{GeV}\,\text{and}\,\lvert y\rvert < 2.5$ $\Delta R(\text{jet}, \ell) > 0.4$	$p_T > 20\,\text{GeV}\,\text{and}\,\lvert y\rvert < 2.5$ $\Delta R(\text{jet}, \ell) > 0.4$
Flavour-tagged jets	$p_T > 20\,\text{GeV}\,\text{and}\,\lvert y\rvert < 2.5$ DL1r@85%	$p_T > 20\,\text{GeV}\,\text{and}\,\lvert y\rvert < 2.5$ DL1r@85%

Table 3.2: Object definition for electron and muon channels

tudinal impact parameters. The transverse impact parameter must satisfy $\frac{|d_0|}{\sigma(d_0)} < 3.0$, and the longitudinal impact parameter must meet $|z_0 \sin\theta| < 0.5\,\text{mm}$, where θ is the polar angle. Additionally, muons must have a transverse momentum (p_T) of at least 27 GeV and lie within the pseudorapidity range $|\eta| < 2.5$.

Electron candidates must meet the *Tight* quality standards and satisfy the *Tight VarRad* isolation criteria, as outlined in Section 2.3.1. This is implemented to minimize background electron contamination from hadronic decays. Electron tracks must be linked to the primary vertex, with constraints on both the transverse and longitudinal impact parameters. The transverse impact parameter must satisfy $\frac{|d_0|}{\sigma(d_0)} < 5.0$, while the longitudinal impact parameter is constrained by $|z_0 \sin\theta| < 0.5$ mm, where θ is the polar angle. Additionally, electrons must have a minimum transverse momentum (p_T) of 27 GeV and be located within the ranges $1.52 < |\eta| < 2.47$ or $|\eta| < 1.37$.

Jets are reconstructed using the anti-k_T algorithm with a radius parameter $R = 0.4$, as implemented in the *FastJet* package [53]. The p_T of the jet must exceed 20 GeV and fall within $|y| < 2.5$. To reduce pile-up, it is essential to apply the *Jet Vertex Tagger* algorithm with the *Tight* working point.

For b-tagging, the jet must satisfy the 85% efficiency working point, meaning the algorithm selects b-jets with 85% efficiency. In this case as well, the jet is required to have $p_T > 20$ GeV and $|y| < 2.5$.

Table 3.2 summarizes the selection criteria for leptons and jets.

Events containing exactly one Z boson candidate are required to have precisely two leptons of the same flavor (i.e., ee or $\mu\mu$) and opposite charge, with an invariant mass in the range $76\,\text{GeV} < m_{\ell\ell} < 106\,\text{GeV}$, centered around the mass of the Z boson (m_Z). Additionally, one of the two leptons must match the lepton that triggered the event.

To distinguish these events from the significant $t\bar{t}$ background, events with the transverse momentum of the two leptons smaller than 150 GeV ($p_T^{\ell\ell} < 150\,\text{GeV}$) are required to have the missing transverse energy smaller than 60 GeV ($E_T^{\text{miss}} < 60\,\text{GeV}$).

For the MC samples, the flavour of the hadron that produced the jet is categorized as either a b-, c-, or light jet using cone-based criteria. Specifically, a jet is labeled as a b-jet if it lies within $\Delta R = 0.3$ of at least one b-hadron with $p_T > 5$ GeV. In cases where a b-hadron matches two jets, only the jet closest in ΔR is tagged as a b-jet. If no b-jets are identified, a jet is classified as a c-jet if it is within $\Delta R = 0.3$ of a c-hadron with $p_T > 5$ GeV. Jets that do not satisfy either of these conditions are classified as light jets.

After events pass the selection criteria, they are categorized into two distinct regions: the *1-tag region*, where at least one jet has been flavor-tagged, and the *2-tag region*, where at least two jets have been flavor-tagged. In this analysis only the 2-tag region, centered on $Z+ \geq 2\,b$-jets events, is taken into account.

The event selection and signal region requirements are summed in Table 3.3.

Event selection	
Leptons	Exactly 2, same-flavour, opposite-charge
$m_{\ell\ell}$	$76\,\text{GeV} < m_{\ell\ell} < 106\,\text{GeV}$
E_T^{miss}	$E_T^{\text{miss}} < 60\,\text{GeV}$ if $p_T^{\ell\ell} < 150\,\text{GeV}$
Flavour-tagged jets	≥ 1 or ≥ 2 jets, DL1r@85%
Signal regions	
1-tag	≥ 1 flavour-tagged jets
2-tag	≥ 2 flavour-tagged jets

Table 3.3: Event selection and signal regions

3.5 Background estimation

The backgrounds in the production of a Z boson in association with b-jets are processes where the Z boson is produced in association with jets but the jet flavour is different from the one targeted in the measurement, dileptonic $t\bar{t}$ events, dibosons, ZH, single top quark production, $Z \to \tau\tau$ and multi-jet events and in the case of the final state in association with at least 2 b-jets, also the Z+1 b-jet process is considered as background.

Minor backgrounds, namely single top-quark, diboson, ZH and $Z \to \tau\tau$ are estimated using the MC samples detailed in Section 3.3, while major backgrounds such as the ttbar and the Z+jets process with jets originated by light or c-jets or Z + 1 b-jet are estimated via data-driven technique, as well as the multijet background that cannot be estimated with MC with high precision.

In this analysis, a significant amount of missing transverse energy is observed, which is typically attributed to the presence of neutrinos. However, in the case of Z boson production with b-jets, neutrinos are not expected in the final state, suggesting that the observed E_T^{miss} primarily arises from the $t\bar{t}$ background. The cut presented in Table 3.3 is done in order to reduce the $t\bar{t}$ contribution.

The contribution of $t\bar{t}$ events is estimated using data-driven technique. This involves applying the same selection as those in the signal region, but selecting events with opposite-flavor leptons ($e^+\mu^-$ or $e^-\mu^+$) and requiring a larger invariant mass range of $71\,\text{GeV} < m_{\ell\ell} < 111\,\text{GeV}$. This control region (CR) is dominated by $t\bar{t}$ events with a small contribution from other processes (i.e. diboson, Z+jets and single top) . The contribution of non-$t\bar{t}$ processes in the CR is subtracted from the data using MC predictions. Finally for each measured observable the $t\bar{t}$ contribution in the signal region is estimated as the distribution of the data in CR after the non-ttbar processes subtraction scaled by a Transfer Factors (TFs). The TF is derived using a $t\bar{t}$ MC simulated sample and it is obtained as the ratio of the distribution of the given observable in the signal region(ee or $\mu\mu$) to the distribution of the same observable in the CR ($e\mu$).

Also multi-jet events constitute a background in the analysis of Z boson production in association with b-jets and their contribution is evaluated using data-driven method. Detailed studies have been done in the recent ATLAS publication with the same dataset and it has been concluded that in the phase space of the measurement this background is negligible and therefore it has been neglected in the publication and in the current thesis work . For completeness the method adopted is reported here. In this case, the control region definitions are similar to those of the signal regions, but with some modifications. In the muon channel, the isolation requirement for muons is removed and both muon candidates are required to have the same charge. In the electron channel, the identification, impact parameter cuts and isolation criteria are removed and the electrons must have the same charge. Contributions from non-multi-jet sources are estimated using simulations and subtracted from the observed data. The contribution to the background from the production of Z bosons in association with jets of flavors other than b is estimated using a technique known as flavor fit. This method involves a maximum-likelihood fit based on flavor-sensitive distributions. A fit of the $m_{\ell\ell}$ distribution to data in the signal and multijet control region is performed in the $40\,\text{GeV} < m_{\ell\ell} < 160\,\text{GeV}$ window. The normalisations of the Z+jets processes and of the multijet background templates are free to float in the fit, while the normalisation of the other processes is fixed. Various alternative approaches are tested by varying the fit strategy (i.e. mass range) and the control region definitions.

The DL1r b-tagging discriminant output of the leading flavor-tagged jet and the ones of the sub-leading flavor-tagged jet are used. This output is divided into four distinct bins, each corresponding to specific b-tagging efficiency ranges: 85–77% (bin 1), 77–70% (bin 2), 70–60% (bin 3), and below 60% (bin 4). The joint distribution of the DL1r discriminant outputs for both the leading and sub-leading flavor-tagged jets results in a ten-bin distribution. Separate templates are generated using MC simulation for events containing $Z+\geq 2$ b-jets, $Z+\geq 1$ b-jet, $Z+\geq 1$ c-jet, and $Z+$ light jets and a single combined template is constructed for all other non-$Z+$ jets backgrounds. In the fit the normalization of each Z+jets component is free to float, while normalization of the non-Z+jets backgrounds is fixed to their MC estimate. The fit is done simultaneously in the electron and muon channels. The scale factors extracted by the fit for MG FxFx and Sherpa are reported in Table 3.5.

The results on the flavour fit in this thesis have been taken from the ATLAS publication performed with the same dataset. In particular these are the values obtained for the measurement of the inclusive cross-section. In the ATLAS publication for the background estimation in the differential cross-sections, the flavour fit has been performed in different bins of the measured observables, but this goes beyond the scopes of this thesis.

Process (Sherpa)	**electron**	**muon**
$Z+\geq 2b$	1.10 ± 0.00	1.16 ± 0.00
$Z+1b$	0.87 ± 0.02	0.81 ± 0.02
$Z+c(l)$	1.06 ± 0.01	1.11 ± 0.01
Process (MadGraph)	**electron**	**muon**
$Z+\geq 2b$	0.93 ± 0.00	0.89 ± 0.00
$Z+1b$	1.06 ± 0.00	1.19 ± 0.02
$Z+c(l)$	1.12 ± 0.01	1.12 ± 0.00

Table 3.4: Sherpa and MadGraph scale factors for the Z + 2 b-jets measurement for the electron channel, the muon channel and their combination.

3.6 Uncertainties in the detector-level measurements

Uncertainties in detector-level measurements arise from systematic effects associated with the objects definition and the detector-level selection process and from the determination of backgrounds. These corrections are inherently linked to the imperfect understanding of the detector. All systematic shifts are also applied to the MC predictions of both signal samples and backgrounds.

The uncertainties related to the detector-level selection of physics objects originate from systematic sources that propagate throughout the analysis.

The contribution of lepton-related uncertainties arises from factors such as the trigger, reconstruction, identification and isolation processes. To match the MC simulations with the data, Scale Factors are applied to account for these uncertainties. Additionally, corrections for the electron energy scale and resolution as those related to the muon momentum scale, the inner-detector and the muon spectrometer resolution are considered. These uncertainties are treated as uncorrelated between the two lepton channels.

Uncertainties in the reconstruction of jets are addressed through adjustments in the Jet Energy Scale (JES) and Jet Energy Resolution (JER). These are estimated by scaling and smearing the jet four-momentum in the simulation based on the calibration procedure.

The largest source of uncertainty arises from flavour tagging. The uncertainties for $Z+\geq 2$ b-jets stem from the calibration of the DL1r tagger. As previously discussed in Section 2.3.4, the discrepancies in flavour tagging between data and MC simulations are used to correct the flavour tagging efficiency in the MC. For b-jets, the correction factors are close to unity and the associated uncertainties are described by 45 independent parameters. In the case of c-jets and light jets, the correction factors range between 1 and 1.3, with uncertainties characterized by 20 independent parameters.

Another minor source of uncertainty is the imperfect modeling of pile-up effect. Instead, the uncertainty in the integrated luminosity for the Run 2 data is 0.83%.

An additional systematic uncertainty is derived by taking the difference between the post-fit Z+jets background evaluated using MG FxFx and Sherpa simulated samples.

Another uncertainty is introduced to evaluate possible bias from the extrapolation of $t\bar{t}$ contribution to the signal region using an alternative validation region.

The uncertainty for the diboson background is determined by varying μ_F and μ_R and taking the envelope of the predicted variations as the error estimate. Other uncertainties, such as those from single-top, ZH and $Z \to \tau\tau$ processes, are due to normalization of QCD scale, PDF and α_s variations.

3.7 Data and Monte Carlo comparison

In this section, comparisons between the predictions and the data are presented to ensure there are no significant discrepancies between them. Signal events are simulated using MC samples, while background processes are modeled either with MC simulations or data-driven technique, as detailed in the previous sections. The comparison is evaluated in three scenarios: before the flavour fit, after the flavour fit with both signal and background scaled to assess the goodness of the fit, after the flavour fit but with only the backgrounds scaled to evaluate the impact of the signal mismodeling.

In the following figures, the comparison between the data and the predictions is performed using for the signal MC simulation alternatively `Sherpa` and `MG FxFx`. Each plot presents on the top the event distributions as a function of a given observable (the complete list of the observables is reported in Sect. 3.2), while the lower section of each plot shows the ratio between the predictions and the data. This ratio allows for a clearer assessment of whether the predictions are compatible with the data within the range of uncertainties. The uncertainty bands include the uncertainties on the statistics of the simulated sample, on the event-selection (namely the uncertainties of physics objets) and on the background. The comparison is shown for the muon channel, the electron channel and their combination. The combination is meant the sum of electron and muon channel results.

In Fig. 3.1 and 3.2 the distribution of the events in function of the transverse momentum p_T and the rapidity $|Y|$ of the Z boson are shown. In Fig. 3.7 and 3.3 the distributions in function of the p_T and of the $|Y|$ of the lead b-jet are shown. In Fig. 3.4, 3.5, 3.6 the distribution of the events as a function of $\Delta\Phi$, ΔY, ΔR and in Fig. 3.7 as a function of the invariant mass of the system of the two b-jets are presented. $\Delta\Phi$ and M_{bb} have already been published in the paper [26].

The event yields in both data and predictions are summarized for all three scenarios, across the muon, electron and combined electron+muon channels respectively on Table 3.5, 3.6 and 3.7 before and after the flavour fit. In particular, it is reported the total number of predicted events and events observed in data. The expected amount of signal and various background is expressed as a fraction of the total number of predicted events. In the table, the signal and Z+jets background predictions are from the $MG + FxFx$. The uncertainty in the total predicted number of events is statistical only.

For each of the shown distributions, it is observed, as expected, an improved data-to-MC agreement when signal and Z+jets background are scaled using the results of the flavour fit, with respect to the pre-fit results. Also, as expected, a quite large reduction of the uncertainties is observed after applying the flavour-fit procedure, this is due to the fact that the flavour fit is applied to each of the shifted distributions obtained applying each systematic, the effect of the fit is a constraint the predictions to the data and this results in a reduction of the systematics. When only the Z+jets backgrounds are scaled, first comments on the agreement between the signal expectation and the data can be done. In the following part of the section I am commenting mostly this last set of results focusing on the combined final state (Z→ll) being the ones less affected by statistical fluctuation (plot of each figure).

Fig.3.1, showing the distribution of the Z_{p_T}, demonstrates good agreement between predictions and data at low p_T, while the discrepancies be-

Process	**Not Scaled**	**All Scaled**	**Background Scaled**
Total predicted	**194000 $\pm$ 600**	**192000 $\pm$ 600**	**204000 $\pm$ 600**
Z+ $\geq$ 2 b-jets	48%	43%	46%
Z+ $\geq$ 1 b-jets	10%	11%	11%
Z+c-jets	19%	24%	23%
Z+light-jets	8%	8%	7%
Z $\rightarrow \tau\tau$	<1%	<1%	<1%
ZW, ZZ, ZH (ZX)	2%	2%	2%
tt (MC)	12%	12%	11%
Single-top	<1%	<1%	<1%
Data		192531	

Table 3.5: Number of events for the muon channel in the three scenarios: before flavour fit, after flavour fit with signal and Z+jets background scaled, after flavour fit scaling only the Z+jets background. The expected amount of signal and various background is expressed as a fration of te total number of predicted events. The signal predictions are from `MG+FxFx`.

Process	**Not Scaled**	**All Scaled**	**Background Scaled**
Total predicted	**116000 $\pm$ 400**	**116000 $\pm$ 400**	**121000 $\pm$ 500**
Z+ $\geq$ 2 b-jets	48%	44%	46%
Z+ $\geq$ 1 b-jets	10%	11%	11%
Z+c-jets	20%	23%	23%
Z+light-jets	7%	7%	7%
Z $\rightarrow \tau\tau$	<1%	<1%	<1%
ZW, ZZ, ZH (ZX)	2%	2%	2%
tt (MC)	13%	13%	12%
Single-top	<1%	<1%	<1%
Data		116668	

Table 3.6: Number of events for the electron channel in the three scenarios: before flavour fit, after flavour fit with signal and Z+jets background scaled, after flavour fit scaling only the Z+jets background. The expected amount of signal and various background is expressed as a fration of te total number of predicted events. The signal predictions are from `MG+FxFx`.

Process	Not Scaled	All Scaled	Background Scaled
Total predicted	**310000 ± 700**	**309000 ± 700**	**325000 ± 800**
Z+ ≥ 2 b-jets	48%	43%	46%
Z+ ≥ 1 b-jets	10%	11%	11%
Z+c-jets	20%	24%	23%
Z+light-jets	8%	7%	7%
Z $\rightarrow \tau\tau$	<1%	<1%	<1%
ZW, ZZ, ZH (ZX)	2%	2%	2%
tt (MC)	12%	12%	12%
Single-top	<1%	<1%	<1%
Data		309199	

Table 3.7: Number of events for the lepton channel in the three scenarios: before flavour fit, after flavour fit with signal and Z+jets background scaled, after flavour fit scaling only the Z+jets background. The expected amount of signal and various background is expressed as a fration of te total number of predicted events. The signal predictions are from `MG+FxFx`.

come more pronounced at high p_T. It is evident that the best prediction is provided by the MG FxFx simulation. A similar trend is observed looking at the p_T of the leading b-jet (Fig.3.7), in this last case MG FxFx is still compatible within the uncertainties with the data, while Sherpa largely underestimates the data in the high p_T region. These observation triggered the decision of using MG FxFx for the unfolding procedure discussed in the following Chapter.

Fig. 3.2 and 3.3 show the distribution of the $|Y|(Z)$ and the leading b-jet$|Y|$, a very good agreement between predictions and data is observed.

For what concerns the di-bjets angular distributions, shown in Fig.3.4, Fig.3.5, Fig. 3.6 and Fig. 3.7, a good agreement between predictions and data is observed for $\Delta\Phi_{bb}$, ΔY_{bb} and ΔR_{bb}. While for M_{bb} (Fig. 3.7), crucial for Higgs-boson studies and searches of new physics, a large mismodeling is observed in the high mass range, mainly for Sherpa.

Distribution of events passing event selection as a function of p_T of the leading b-jet between the before flavour fit (left), after flavour fit scaling signal (Z+ b-jets) and Z backgrounds (center) and after flavour fit scaling only Z backgrounds (right).In the first line is shown the muon channel, in the second the electron channel, in the third their combination.

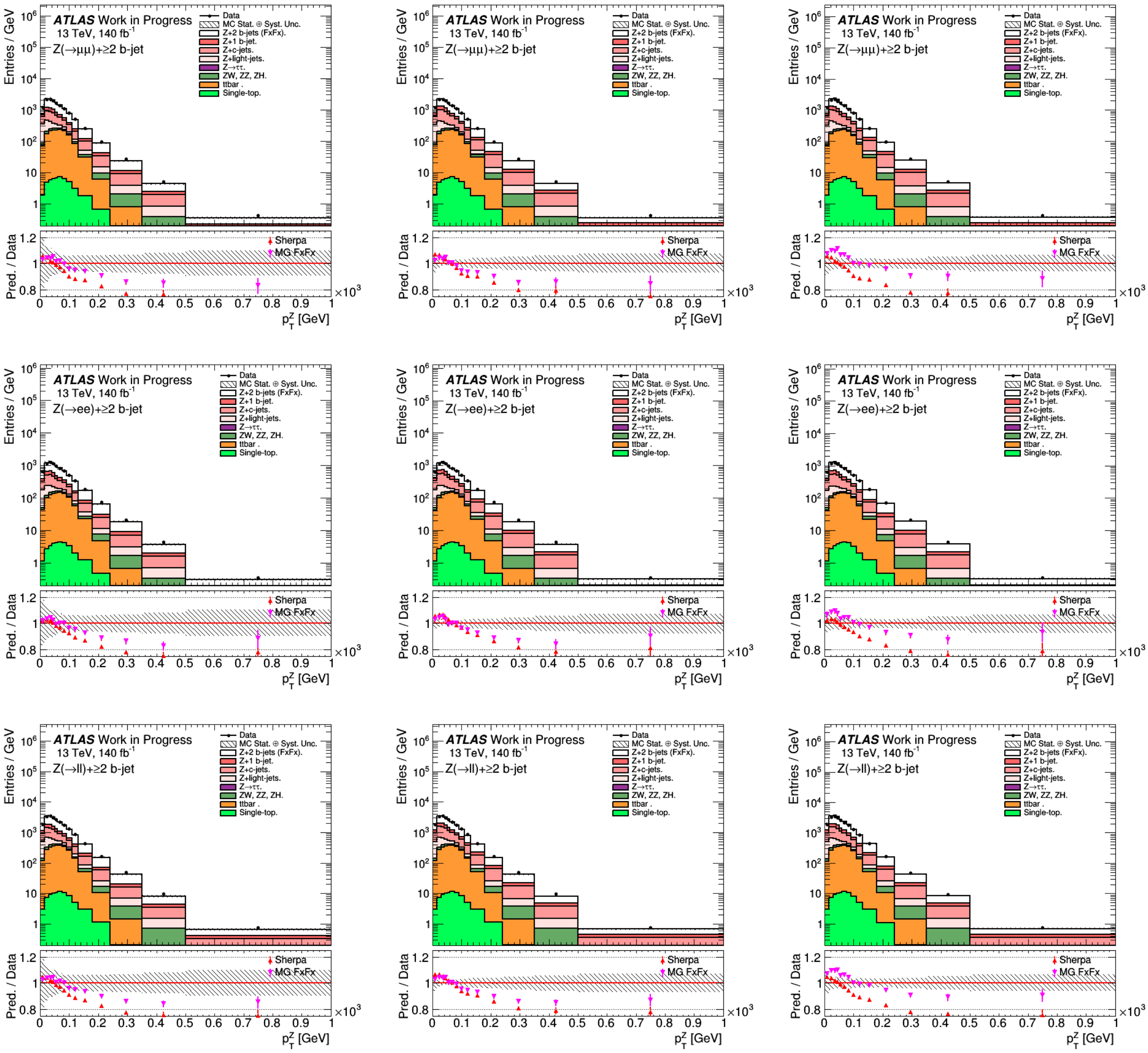

Figure 3.1: Distribution of events passing event selection as a function of the transverse momentum of the Z boson before flavour fit (left), after flavour fit scaling signal (Z+ b-jets) and Z backgrounds (center) and after flavour fit scaling only Z backgrounds (right).In the first line is shown the muon channel, in the second the electron channel, in the third their combination.

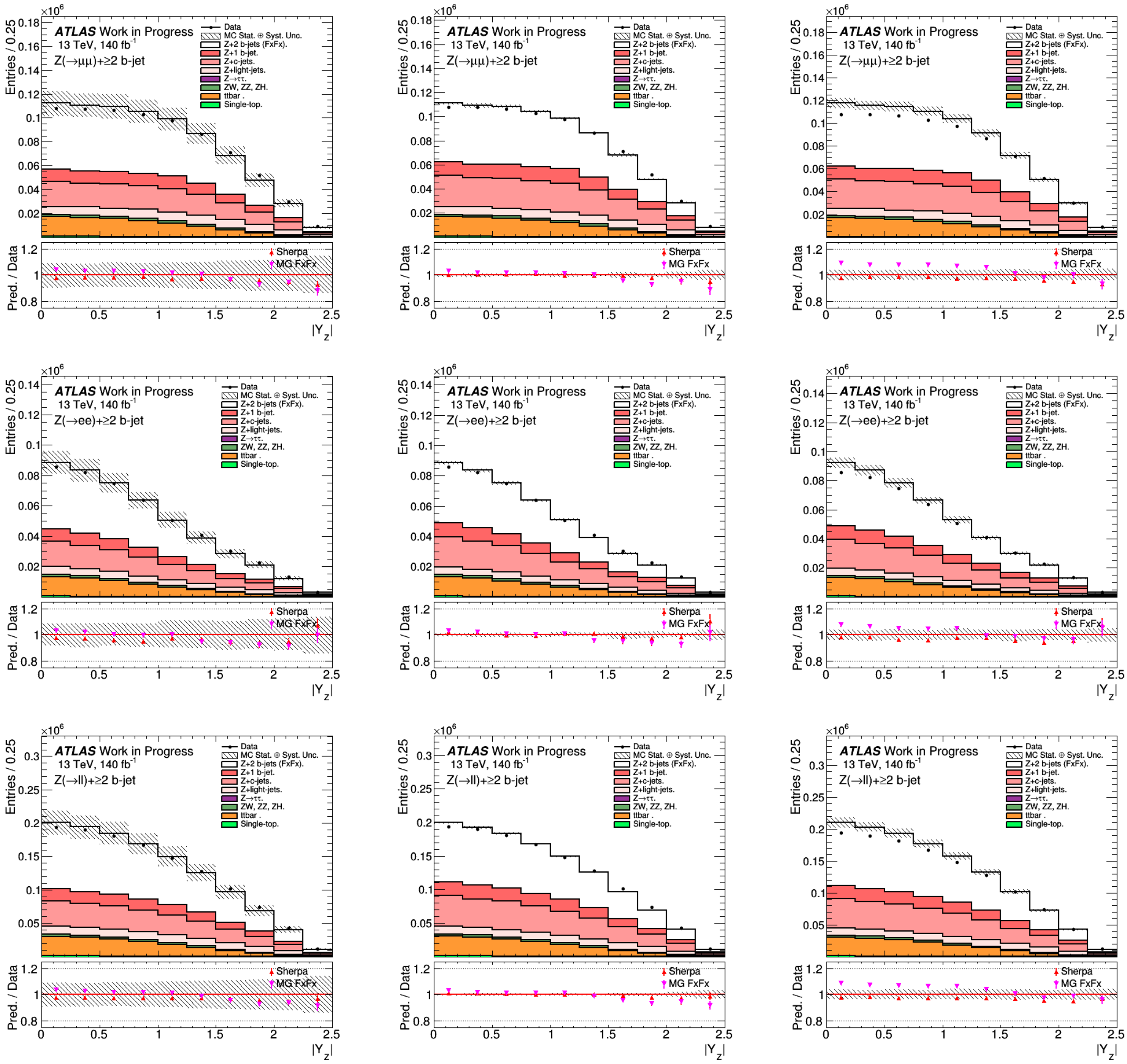

Figure 3.2: Distribution of events passing event selection as a function of $|Y|$ of the Z boson before flavour fit (left), after flavour fit scaling signal (Z+ b-jets) and Z backgrounds (center) and after flavour fit scaling only Z backgrounds (right).In the first line is shown the muon channel, in the second the electron channel, in the third their combination.

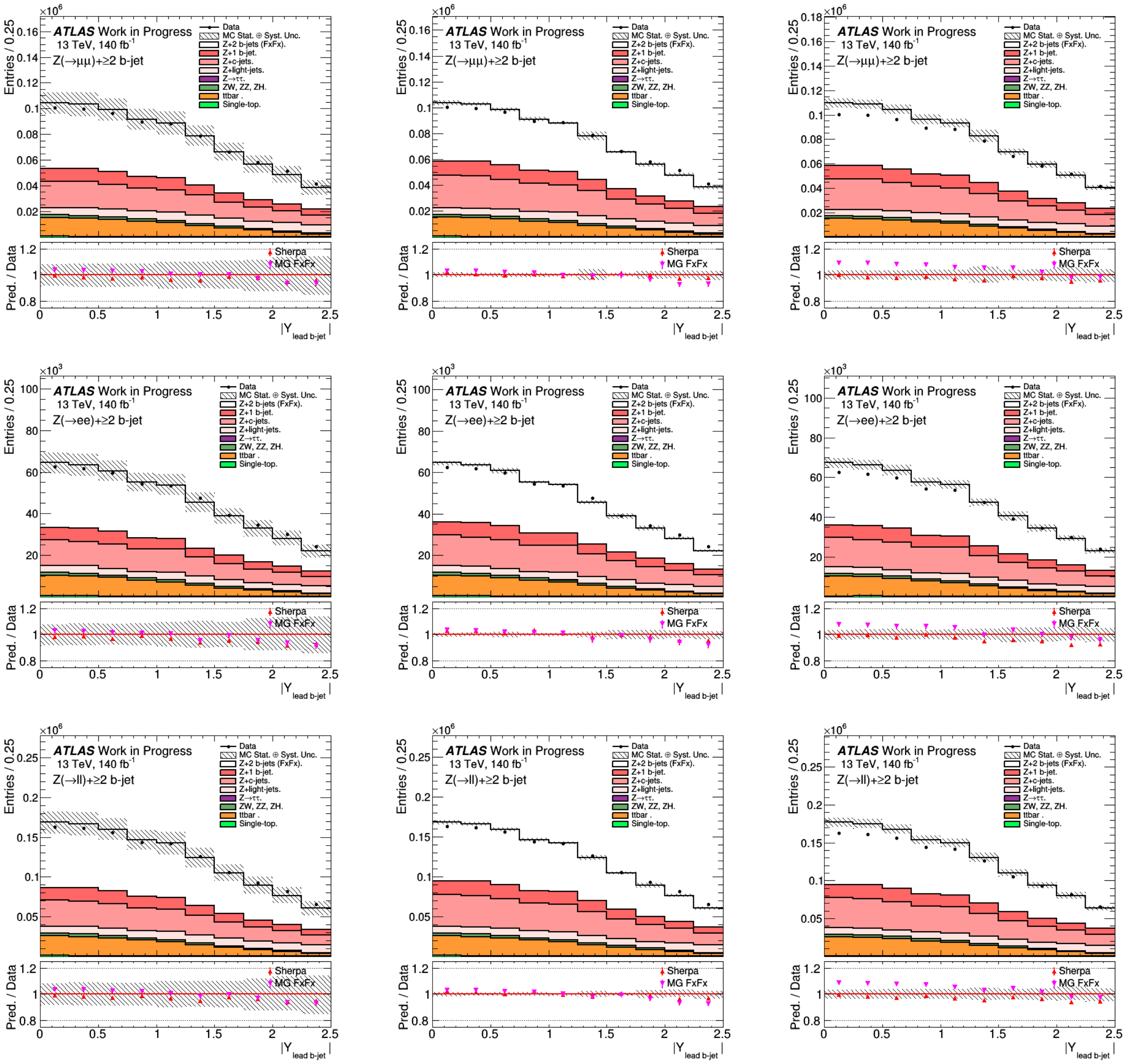

Figure 3.3: Distribution of events passing event selection as a function of Y of the leading b-jet before flavour fit (left), after flavour fit scaling signal (Z+ b-jets) and Z backgrounds (center) and after flavour fit scaling only Z backgrounds (right).In the first line is shown the muon channel, in the second the electron channel, in the third their combination.

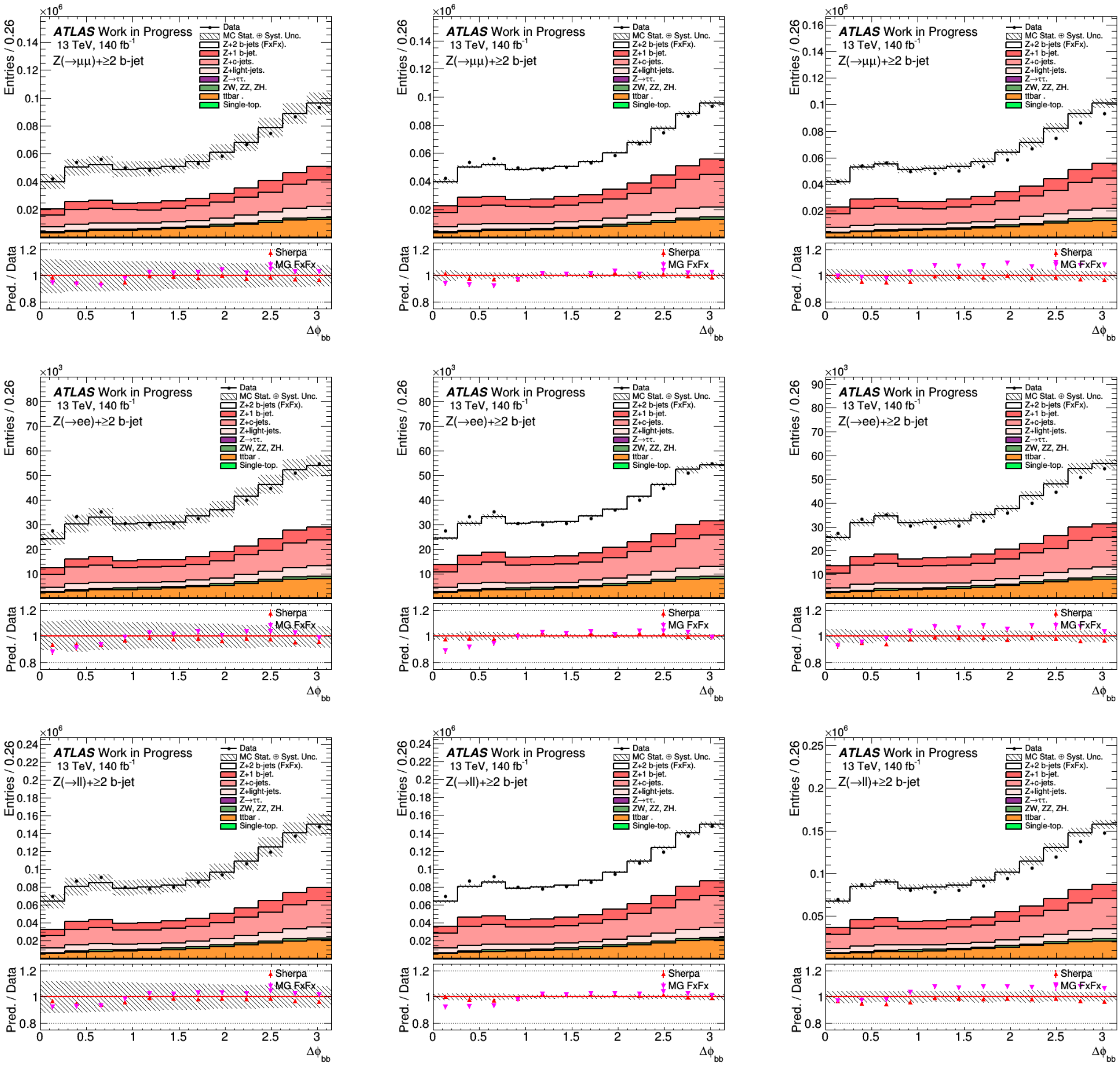

Figure 3.4: Distribution of events passing event selection as a function of $\Delta\Phi$ before flavour fit (left), after flavour fit scaling signal (Z+ b-jets) and Z backgrounds (center) and after flavour fit scaling only Z backgrounds (right). In the first line is shown the muon channel, in the second the electron channel, in the third their combination.

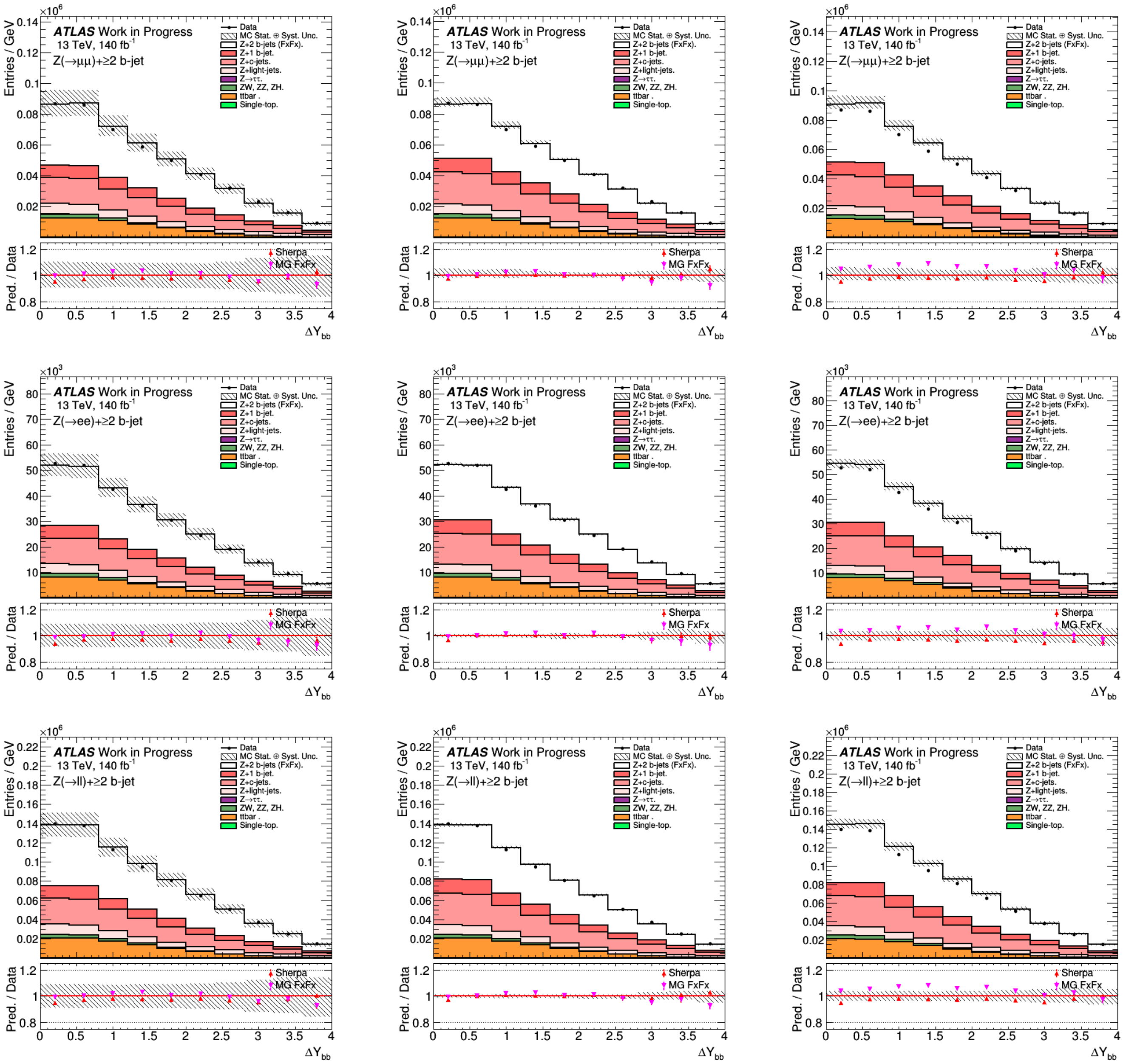

Figure 3.5: Distribution of events passing event selection as a function of ΔY before flavour fit (left), after flavour fit scaling signal (Z+ b-jets) and Z backgrounds (center) and after flavour fit scaling only Z backgrounds (right).In the first line is shown the muon channel, in the second the electron channel, in the third their combination.

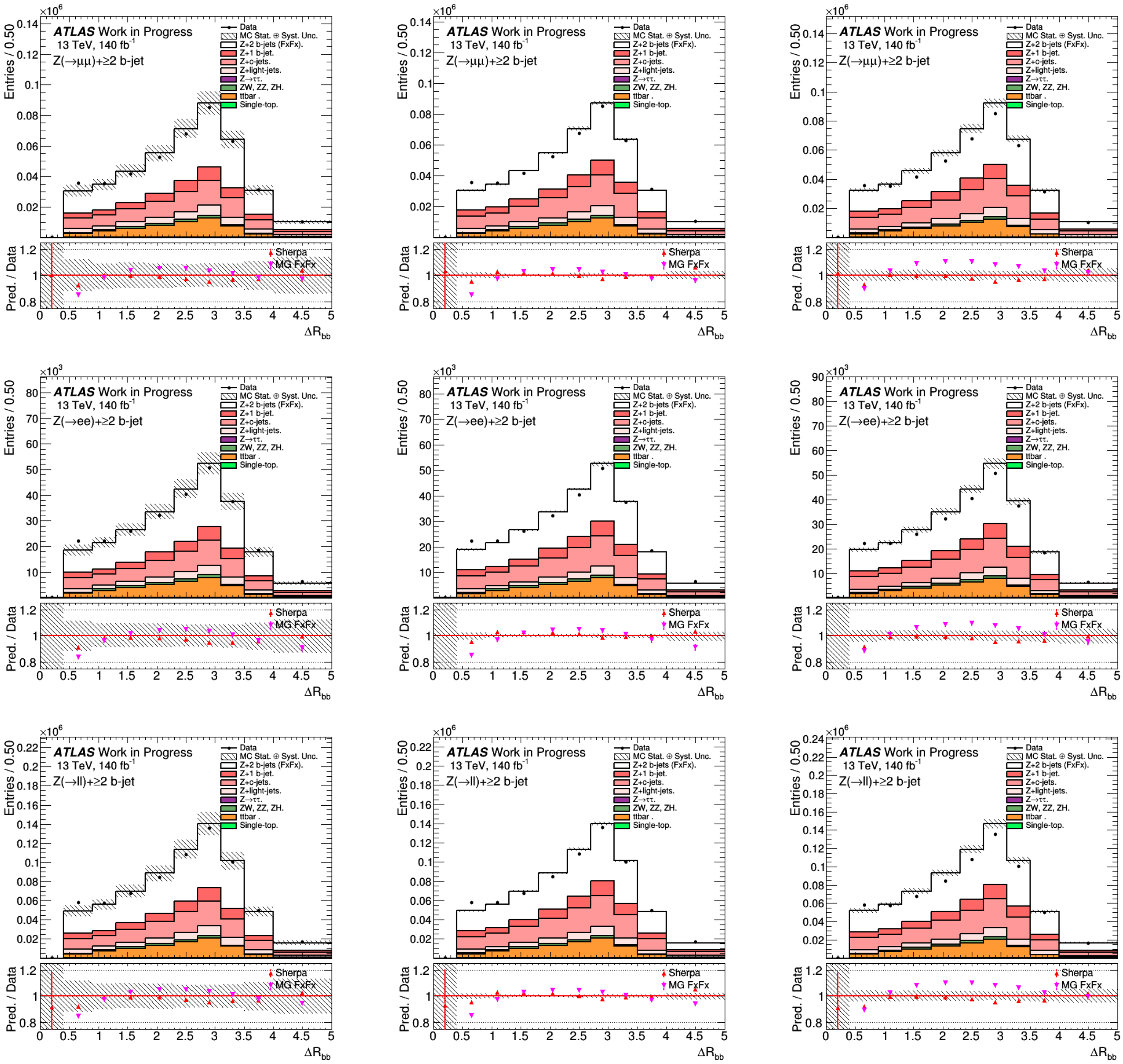

Figure 3.6: Distribution of events passing event selection as a function of ΔR before flavour fit (left), after flavour fit scaling signal (Z+ b-jets) and Z backgrounds (center) and after flavour fit scaling only Z backgrounds (right).In the first line is shown the muon channel, in the second the electron channel, in the third their combination.

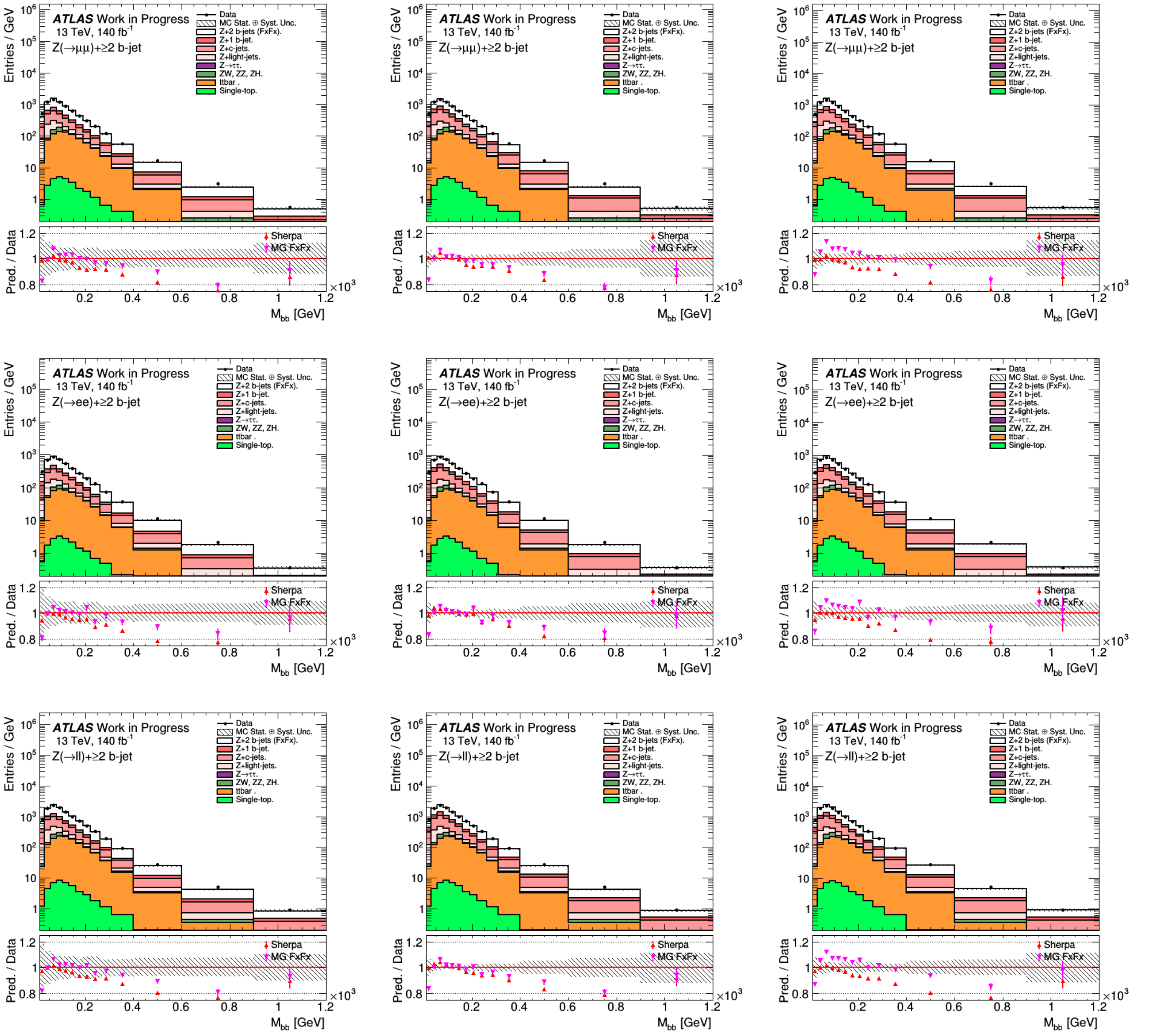

Figure 3.7: Distribution of events passing event selection as a function of M_{bb} before flavour fit (left), after flavour fit scaling signal (Z+ b-jets) and Z backgrounds (center) and after flavour fit scaling only Z backgrounds (right).In the first line is shown the muon channel, in the second the electron channel, in the third their combination.

Chapter 4

Z + b-jets cross-section measurements

In experimental physics, the measured observables often differ from the true ones due to detector effects. To enable meaningful comparisons between experimental data and simulations, it is crucial to correct for these distortions. This correction process is known as Unfolding procedure.

This chapter presents the unfolding technique used in this thesis and the final cross-section results after applying the procedure, often referred to as unfolded results. In particular, Section 4.1 delves into the unfolding technique that has been employed in the analysis. Section 4.2 presents the systematic uncertainties associated with the unfolded results. Finally, in Section 4.3 and in Section 4.4, the inclusive differential cross sections and the differential ones of the Z boson production in association with at least 2 b-jets (Z + $\geq$ 2 b-jets) are presented.

4.1 Unfolding in Z+ $\geq$ 2 b-jets measurements

The measured distributions are affected by detector effects, such as resolution, non-linearities, efficiencies and acceptance. The role of the unfolding technique is to transform the measured distribution at the detector level into the particle-level distribution, also referred to as the truth distribution. Several techniques are commonly used for unfolding, including the bin-by-bin method and the Bayesian unfolding, detailed in Appendix A. Each method has its advantages and is chosen based on the complexity of the measurement and the level of regularization required to mitigate statistical fluctuations and detector smearing. Unfolding ensures that the reconstructed data can be compared directly with theoretical predictions, either fixed-order pQCD calculations or MC simulations at the truth level.

The distributions derived from the signal event yields — after subtracting the estimated background contributions from the data — are then corrected for detector-level effects. The unfolded data are detector-independent and can finally be compared with theoretical predictions.

The phase-space at particle level corresponds to the fiducial phase-space of the measurement. In this analysis the fiducial phase-space is very close to the detector-level one; this implies that the unfolding procedure relies at minimum on MC-based acceptance corrections and therefore the modeling uncertainties related to this corrections are small.

In Tab. 4.1 the acceptance cuts that identify the fiducial phase space of the cross-section measurements are specified.

Object Selection	**Acceptance cuts**
$Z \to \mu\mu$	2 OS μ with $p_T > 27\,\text{GeV}$, $\|\eta\| < 2.5$, $m_{\mu\mu} = 91\,\text{GeV} \pm 15\,\text{GeV}$
$Z \to ee$	2 OS e with $p_T > 27\,\text{GeV}$, $\|\eta\| < 2.5$, $m_{ee} = 91\,\text{GeV} \pm 15\,\text{GeV}$
bjet	$p_T > 20\,\text{GeV}$ and $\|y\| < 2.5$, $\Delta R(\text{bjet}, \ell) > 0.4$

Table 4.1: Acceptance cuts of the fiducial phase space.

The leptons at the particle level are dressed leptons and they're constructed by taking the final-state leptons originating from the decay of the Z boson and by adding any photons radiated within a cone of $\Delta R = 0.1$ around their direction. Particle-level jets are identified using the anti-k_t algorithm with a radius parameter $R = 0.4$, applied to all final-state particles that have a lifetime longer than 30 ps, excluding the decay products of the dressed Z boson. The particles are classified as b-jets, c-jets or light jets following the same procedure as in Chapter 2.

The measured integrated cross section of the Z-boson production in association with b-jets, more precisely with at least 2 b-jets (Z + $\geq$ 2 b-jets), in a fiducial phase space is given by:

$$\sigma_{\text{fid}}(Z \to \ell^+\ell^- + \geq 2\,b\text{-jets}) = \frac{N^{\text{obs}}_{\text{data}} - B}{C_Z L} \tag{4.1}$$

where $N^{\text{obs}}_{\text{data}}$ is the number of observed data events passing the detector-level selection, B is the number of background events passing the same event selection (namely the sum of background predictions of Table 3.6), L is the integrated luminosity (140 fb^{-1}), and C_Z is defined as

$$C_Z = \frac{N^{\text{MC pass all cuts}}_{\text{selected}}}{N^{\text{MC fiducial volume}}_{\text{truth}}} \tag{4.2}$$

where $N^{\text{MC pass all cuts}}_{\text{selected}}$ is the number of reconstructed signal events passing the detector-level selection, and $N^{\text{MC fiducial volume}}_{\text{truth}}$ is the number of truth-level events in the fiducial volume. This term acts as a correction factor.

The method applied for the measurement of the inclusive cross-section does not account for migrations between bins (i.e. an observable with a given value at particle level, falling in a given bin, that is reconstructed with a different value at reconstruction level and falls in another bin). Therefore this method can be used to perform differential cross section measurements only if the bin-to-bin migrations are small compared to the bin width of the distribution. If this condition is not valid more complex methods has to be used.

In this thesis the differential distributions are corrected to the particle level using an iterative Bayesian unfolding method implemented via the `RooUnfold` package [58].

The differential cross section can be measured as:

$$\frac{d\sigma_i}{dX_i} = \frac{M_{ij}^{-1}(N_j^{\mathrm{obs}} - B_j)f_j}{\epsilon_i L \Delta X_i} \tag{4.3}$$

where N_j^{obs} is the number of events in bin j at the reconstruction level, B_j is the number of background events estimated in bin j at the reconstruction level and M_{ij} is the response matrix (or migration matrix) that describes the migration effects from bin i at the particle level to bin j at the detector level for events passing both the detector-level and particle-level selections. M_{ij}^{-1} is often called unfolding matrix. In this equation, ΔX_i represents the width of bin i, while f_j is a factor that corrects for unmatched events in bin j, i.e., events passing the detector-level selection but not the particle-level selection, L is the integrated luminosity, ϵ_i is the reconstruction efficiency in the i-th bin.

In the Bayesian unfolding method one treats the response matrix as a description of the probability of observing a reconstructed distribution in data given the true distribution at particle level. In general the relation between the distribution of an observable α at particle level ($T(\alpha)$) and the detector level distribution of this observable ($R(\alpha)$) is the following:

$$R(\alpha_i) = \sum_{j=1}^{N_T} M_{ij} T(\alpha_j) = \sum_{j=1}^{N_T} M(R(\alpha_i)|T(\alpha_j))T(\alpha_j) \tag{4.4}$$

where $T(\alpha_j)$ represents the number of events at the particle level in bin j, $R(\alpha_i)$ represents the number of events in detector bin i for the observable α and $M_{ij} = M(R(\alpha_i)|T(\alpha_j))$ is the response matrix. This matrix gives the conditional probability that the detector-level outcome $R(\alpha_i)$ results from the particle-level input $T(\alpha_j)$. N_T is the total number of bins in the particle-level distribution. The aim of the unfolding procedure is to compute the probability $M(T(\alpha_j)|R(\alpha_i))$, which is the unfolding matrix. This can be obtained via Bayes' theorem:

$$M(T(\alpha_j)|R(\alpha_i)) = \frac{M(R(\alpha_i)|T(\alpha_j)) \cdot P_0(T(\alpha_j))}{\sum_{l=1}^{N_T} M(R(\alpha_l)|T(\alpha_j)) \cdot P_0(T(\alpha_l))} \tag{4.5}$$

where $P_0(T(\alpha_j))$ denotes the prior probability set of the hypothesis $T(\alpha_j)$. The prior is provided by the signal MC simulation. An estimator for the unfolded distribution is then computed as:

$$\hat{U}(\alpha_j) = \frac{1}{\epsilon_j} \sum_{i=1}^{N_R} M(T(\alpha_j)|R(\alpha_i)) \cdot D(\alpha_i) \tag{4.6}$$

where $D(\alpha_i)$ is the number of measured background-subtracted data in detector bin i and ϵ_j is the reconstruction efficiency namely the probability that the cause j produces an effect. The efficiency is defined as:

$$\epsilon_j = \sum_{k=1}^{k=N_R} P(R(\alpha_k)|T(\alpha_j)) \tag{4.8}$$

In the iterative Bayesian unfolding adopted in this thesis, the procedure is applied iteratively. For the first iteration, the prior is taken from the particle-level prediction, while for the second iteration, the prior is based on the unfolded distribution from the previous iteration and the unfolding matrix. The greater the number of iterations, the lower the dependence on the MC truth distribution. However, while this dependence decreases, the statistical uncertainty increases.

Thus, the number of iterations must be chosen in order to balance the two effects. In this thesis two iterations are found to be a good compromise between modeling and statistical uncertainty and therefore are used for all the measured observables. The nominal MC used to obtain the migration matrices, efficiencies and correction factors is MG+FxFx.

4.1.1 Migration Matrix

The migration matrix expresses the conditional probability that a given effect is produced by a given cause (i.e. the probabitily that an observable with a value at reconstructed level falling in a given bin has been generated at particle level with a value that falls in a bin that might be different).

The MC used to obtain response matrices, efficiencies and correction factors is MG FxFx.

In Fig.s 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 the migration matrices for the observables measured in this thesis are shown in muon channel, electron channel and in their combination (electron+muon channel). Namely the figures show the migration matrix of: the p_T of the Z boson, the $|Y|$ of the Z boson, the p_T of the leading b-jet, the $|Y|$ of the leading b-jet, $\Delta\Phi$, ΔY, ΔR and M_{bb}. In order to fill these matrices the MC signal events that pass truth-level and reconstruction-level selections are used.

In this representation, the truth values are on the horizontal axis, while the reconstructed ones are on the vertical axis. The entries in each column are normalized to the total number of truth events that fulfill the event selection in that bin, so that the sum of all the entries in a given column represents the reconstruction efficiency to the corresponding truth bin.

The distribution of values in the migration matrix is similar across the different channels, indicating that the reconstruction and identification efficiencies do not vary significantly between them.

In the case of the Z boson's p_T and rapidity Y, the concentration is well-localized along the diagonal of the migration matrix, whereas in other cases, it is spread across the entire matrix, though the largest concentration still remains on the diagonal. This indicates that, for p_T and Y of the Z boson, migration effects are nearly zero. For the other variables, these effects are still very small.

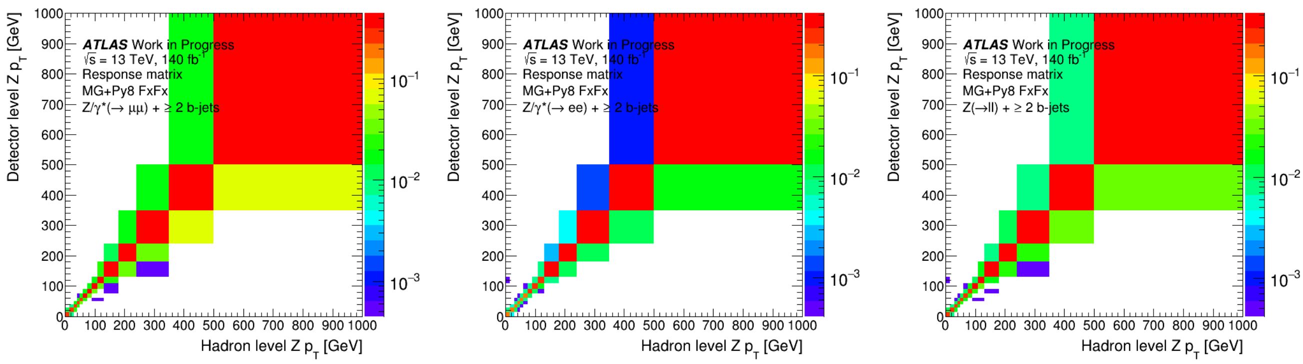

Figure 4.1: Migration Matrices for the p_T of the Z boson in electron channel, muon channel and in their combination (electron+muon channel).

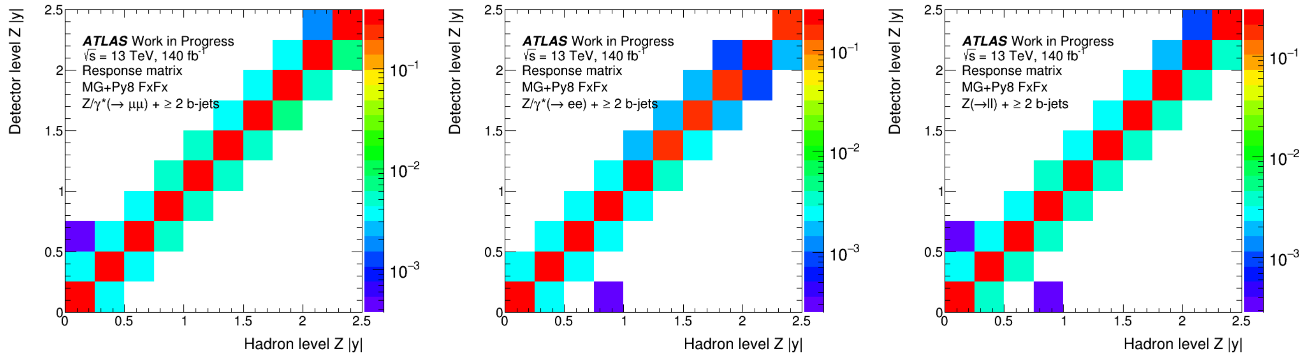

Figure 4.2: Migration Matrices for the $|Y|$ of the Z boson in electron channel, muon channel and in their combination (electron+muon channel).

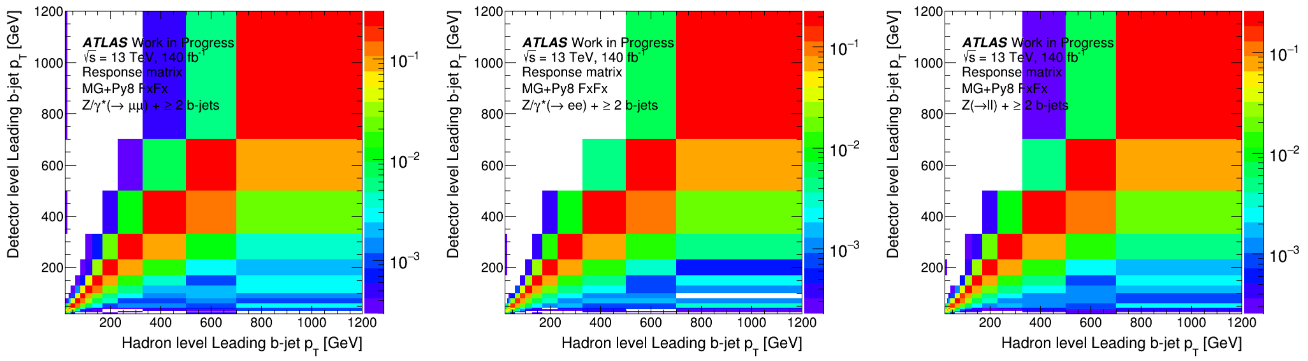

Figure 4.3: Migration Matrices for the p_T of the lead b-jet in electron channel, muon channel and in their combination (electron+muon channel).

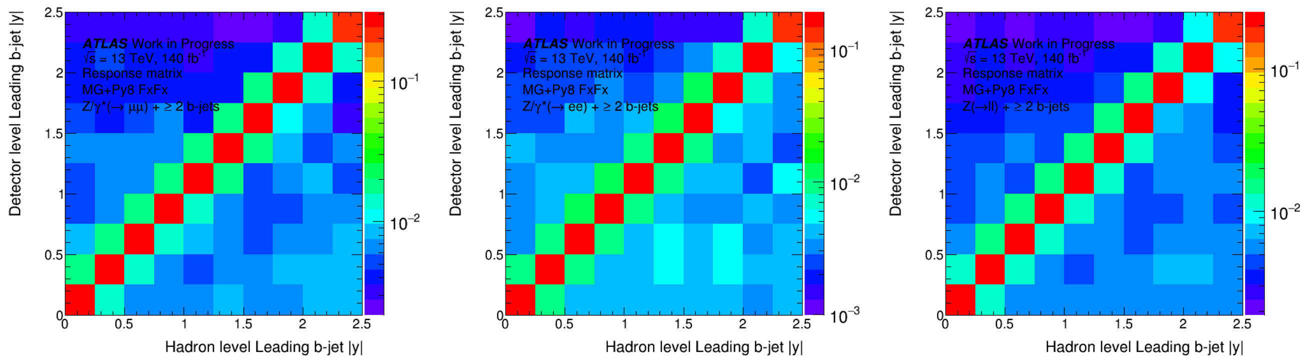

Figure 4.4: Migration Matrices for the $|Y|$ of the lead b-jet in electron channel, muon channel and in their combination (electron+muon channel).

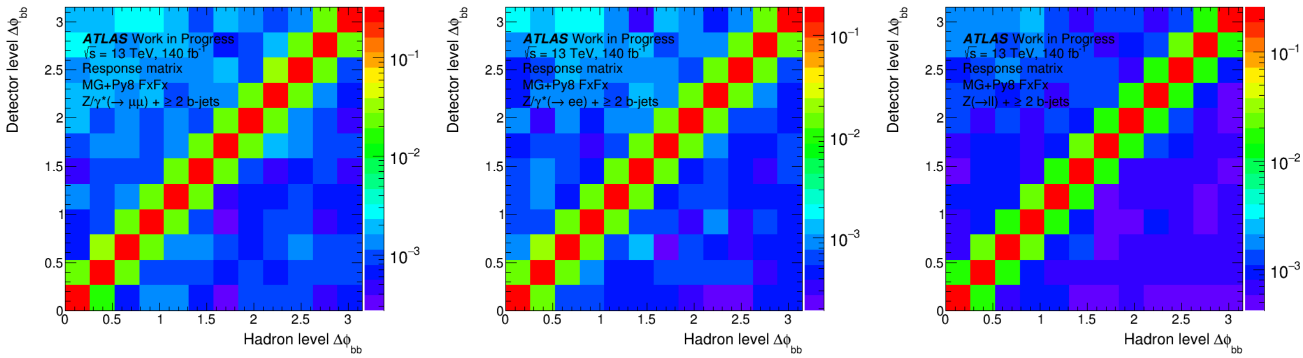

Figure 4.5: Migration Matrices for $\Delta\Phi$ in electron channel, muon channel and in their combination (electron+muon channel).

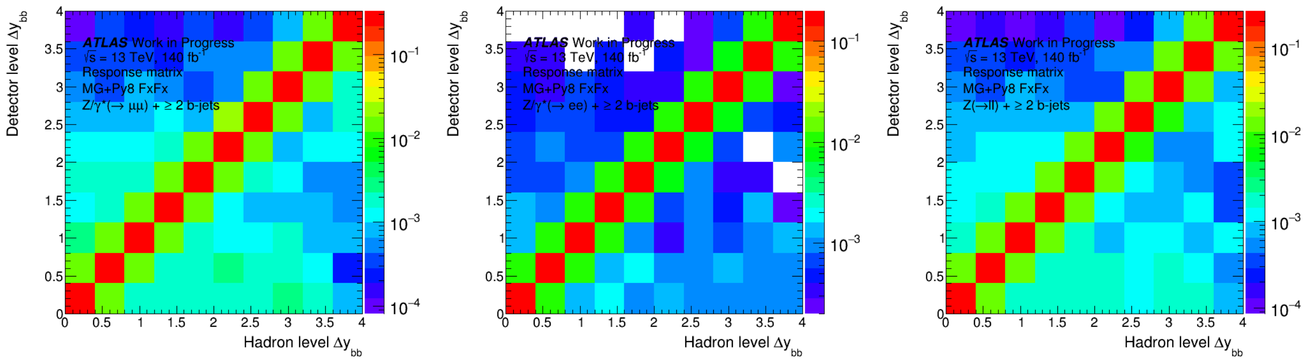

Figure 4.6: Migration Matrices for ΔY in electron channel, muon channel and their combination (electron+muon channel).

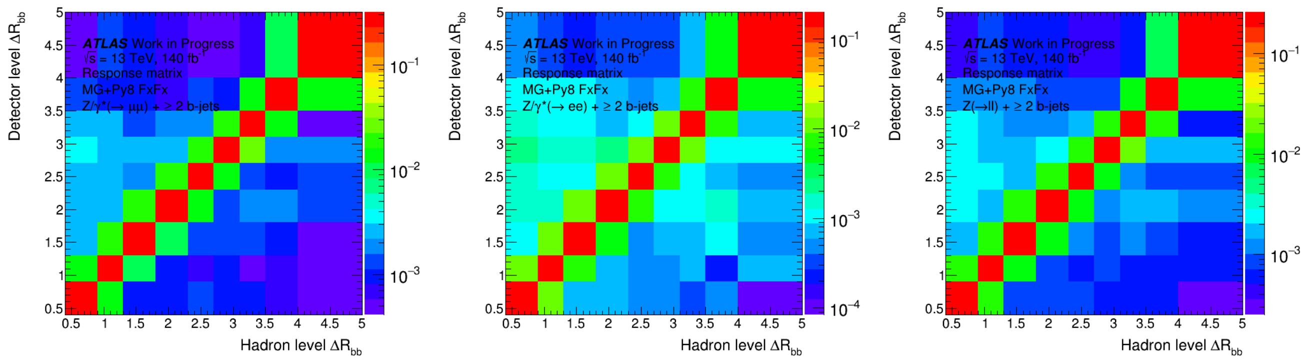

Figure 4.7: Migration Matrices for ΔR in electron channel, muon channel and in their combination (electron+muon channel).

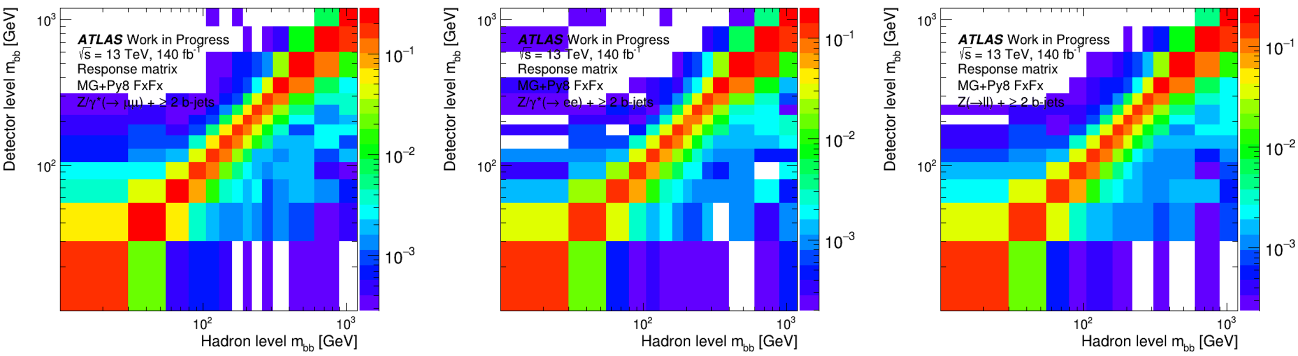

Figure 4.8: Migration Matrices for M_{bb} in electron channel, muon channel and in their combination (electron+muon channel).

4.1.2 Reconstruction Efficiencies

The reconstruction efficiencies are shown in Fig.s 4.9 4.10 for the p_T and $|Y|$ of the Z boson, Fig.s 4.11 and 4.12 for the p_T and $|Y|$ of the lead b-jet, Fig. 4.13 for $\Delta\Phi$, Fig. 4.14 for ΔY_{bb}, Fig.4.15 for ΔR_{bb} and Fig. 4.16 for M_{bb}.

The reconstruction efficiency is defined as the ratio of the number of reconstructed events that are matched to a truth event (i.e. events that pass both the reconstruction and truth-level selections) to the total number of truth events. It is derived from the projection of the migration matrices along the axis representing the truth variables (reconstructed matched events) and then divided by the truth distribution. The reconstruction efficiencies are obtained using MG5+Py8 FxFx, which is the MC employed in the unfolding process, compared with Sherpa 2.2.11.

The figures show an excellent comparison between the two MC simulations used, with differences remaining well below 10%. ΔR is the only variable that, at a few points, exhibits a difference between the MC simulations exceeding 10%. Thus, the overall behavior demonstrates an excellent agreement.

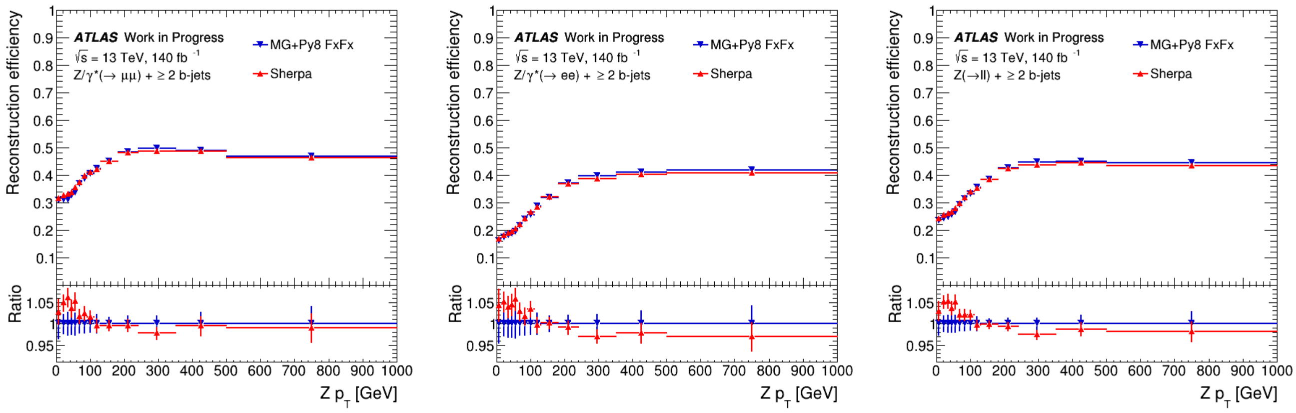

Figure 4.9: Reconstruction efficiency as a function of the p_t of the Z boson in muon channel (left), electron channel (center) and in their combination (electron+muon channel) (right).

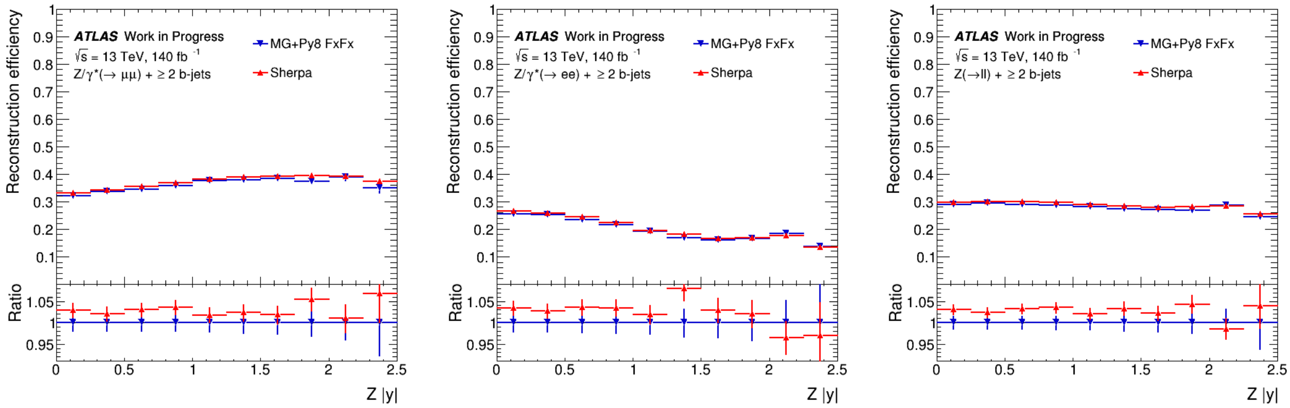

Figure 4.10: Reconstruction efficiency as a function of the $|Y|$ of the Z boson in muon channel (left), electron channel (center) and in their combination (electron+muon channel) (right).

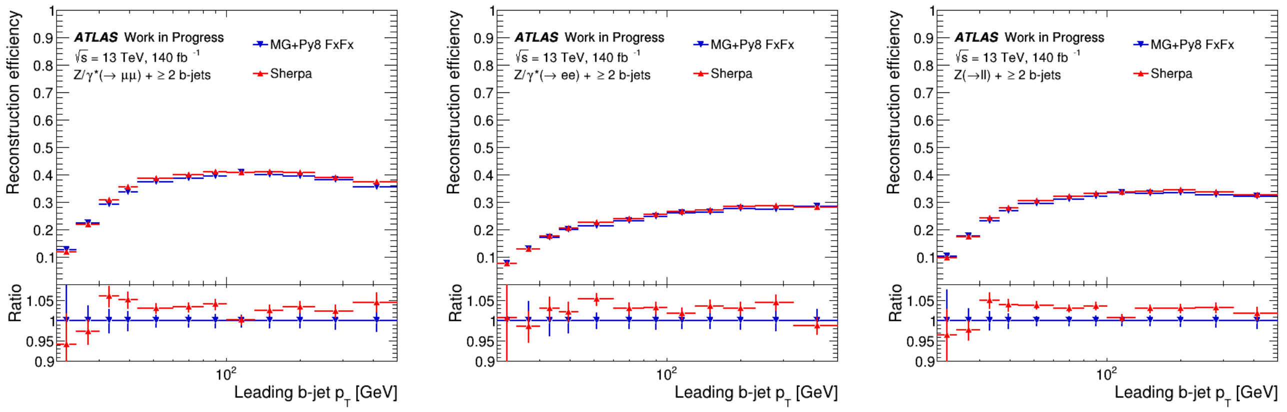

Figure 4.11: Reconstruction efficiency as a function of the p_t of the lead b-jet in muon channel (left), electron channel (center) and in their combination (electron+muon channel) (right).

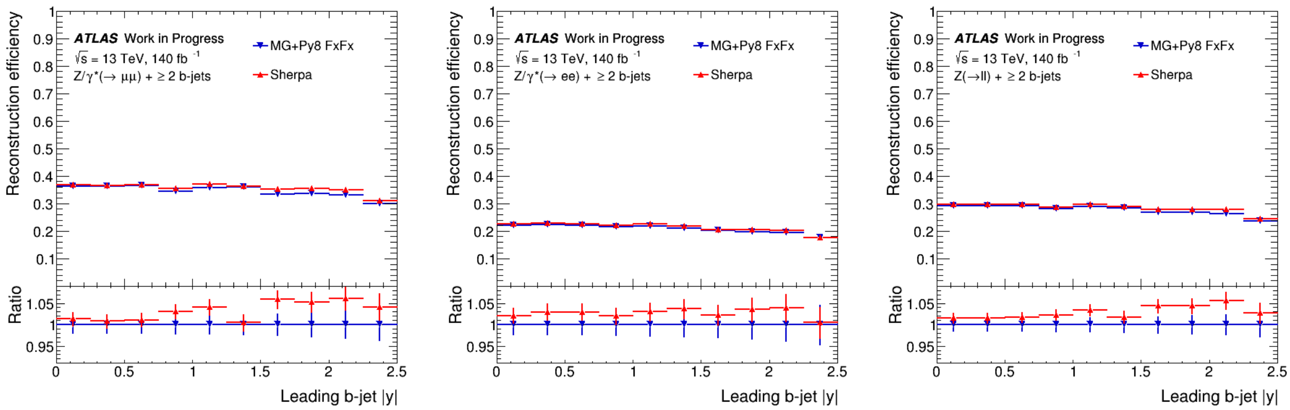

Figure 4.12: Reconstruction efficiency as a function of the $|Y|$ of the lead b-jet in muon channel (left), electron channel (center) and in their combination (electron+muon channel) (right).

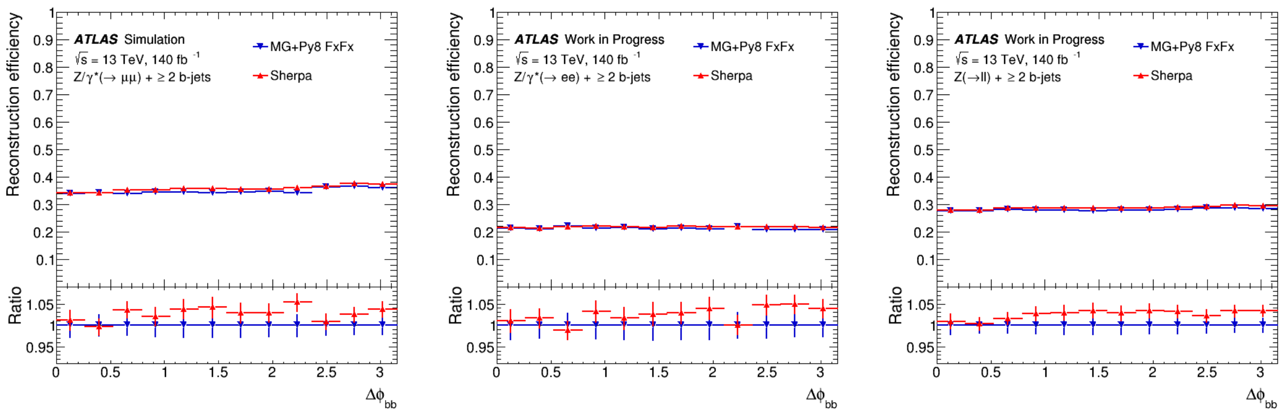

Figure 4.13: Reconstruction efficiency as a function of $\Delta\Phi$ in muon channel (left), electron channel (center) and in their combination (electron+muon channel) (right).

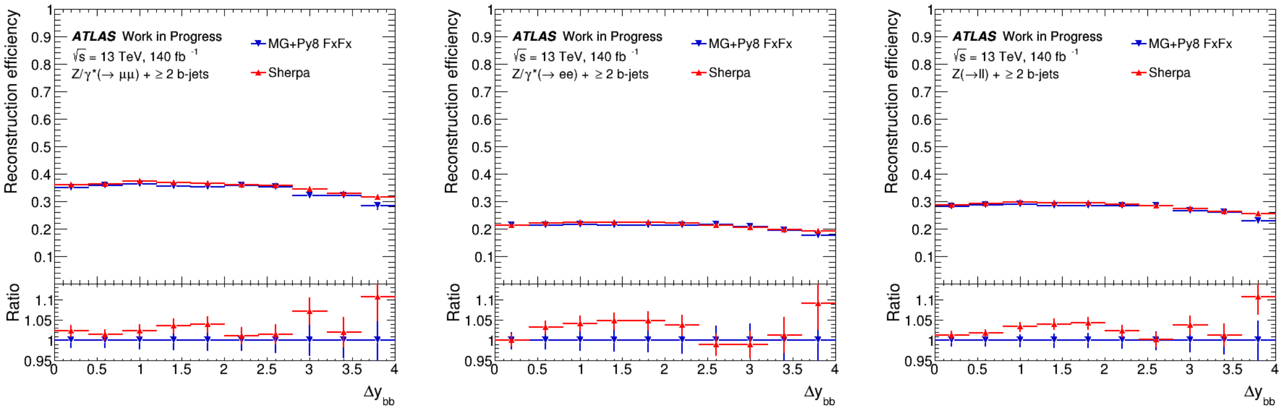

Figure 4.14: Reconstruction efficiency as a function of ΔY in muon channel (left), electron channel (center) and in their combination (electron+muon channel) (right).

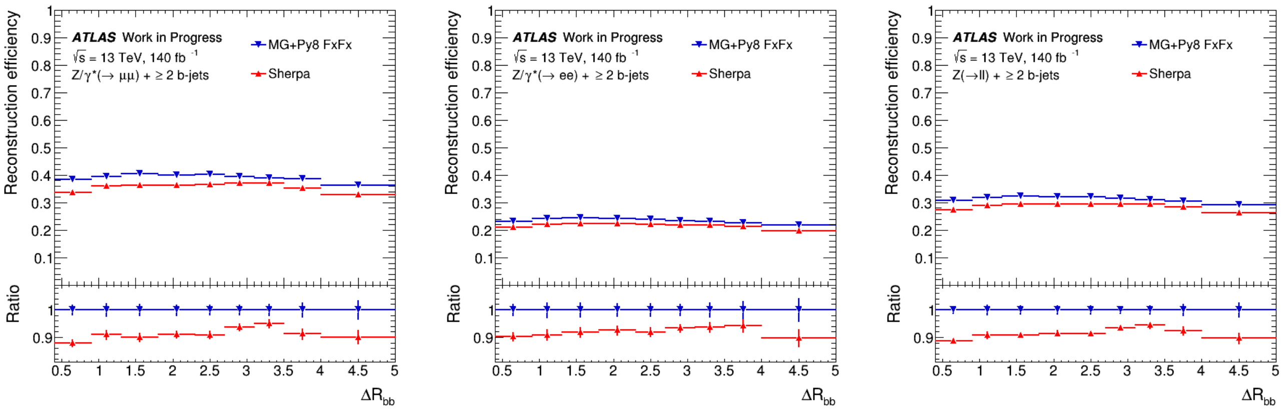

Figure 4.15: Reconstruction efficiency as a function of ΔR in muon channel (left), electron channel (center) and in their combination (electron+muon channel) (right).

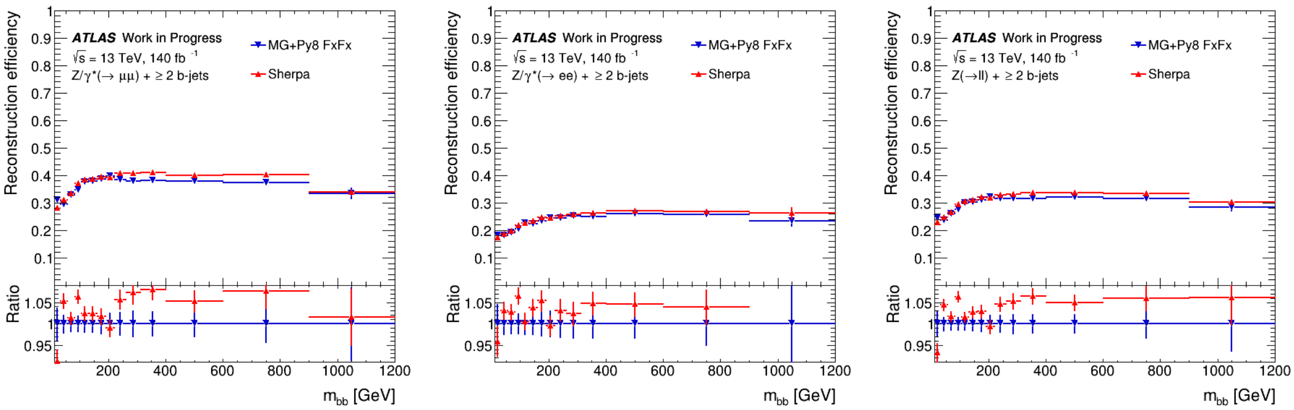

Figure 4.16: Reconstruction efficiency as a function of M_{bb} in muon channel (left), electron channel (center) and in their combination (electron+muon channel) (right).

4.1.3 Fake corrections

Corrections are also applied for "fake" events, i.e., events that fall outside the considered phase space at the particle level but migrate into the studied region at the detector level. These corrections account for migrations across the boundaries of the fiducial volume.

The fake correction factor is defined as the ratio of the number of reconstructed events that are matched to a truth event (i.e., events that pass both the reconstruction and truth-level selections) to the total number of reconstructed events (i.e. events that pass the reconstruction regardless the truth-level selection). It is applied, before the unfolding procedure, as a multiplicative factor both to the background-subtracted data and to the reconstructed MC events.

The fake factors, obtained using MGAMC+PY8 FXFX and SHERPA 2.2.11, are shown in Figs. 4.17 and 4.18 as a function of the p_T and $|Y|$ of the Z boson, Fig.s 4.19 4.20 as a function of the p_T and $|Y|$ of the leading b-jet, Fig. 4.21 as a function of $\Delta\Phi$, 4.22 as a function of ΔY_{bb}, 4.23 as a function of ΔR and Fig. 4.24 for the invariant mass of the two b-jets.

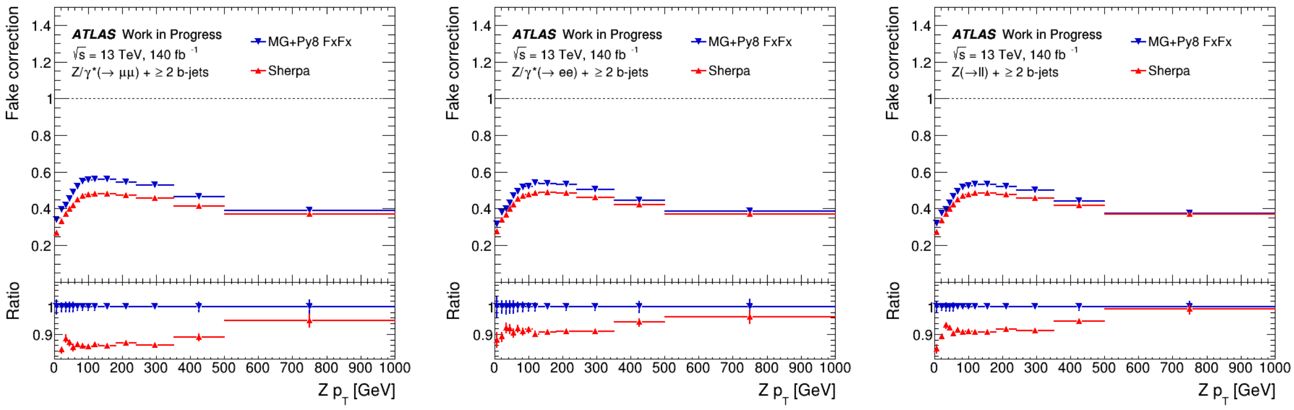

Figure 4.17: Fake corrections as a function of the p_T of the Z boson. The fake corrections are shown in the muon channel (on the left), in the electron channel (in the center) and in their combination (muon+electron channel) (on the right).

In general, a good agreement can be observed between the two MC simulations, with difference within 10% for $\Delta\phi$, M_{bb} and $|Y|$ of the leading b-jets. However, the difference increase to 15% for ΔR, the rapidity Y of the Z boson and the transverse momentum p_T of the Z boson.

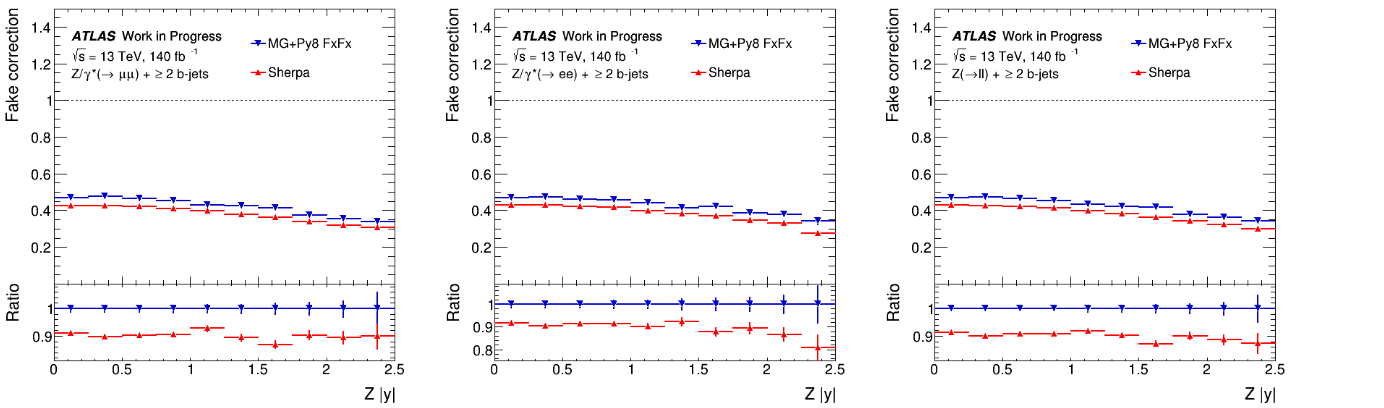

Figure 4.18: Fake corrections as a function of the $|Y|$ of the Z boson. The fake corrections are shown in the muon channel (on the left), in the electron channel (in the center) and in their combination (muon+electron channel) (on the right).

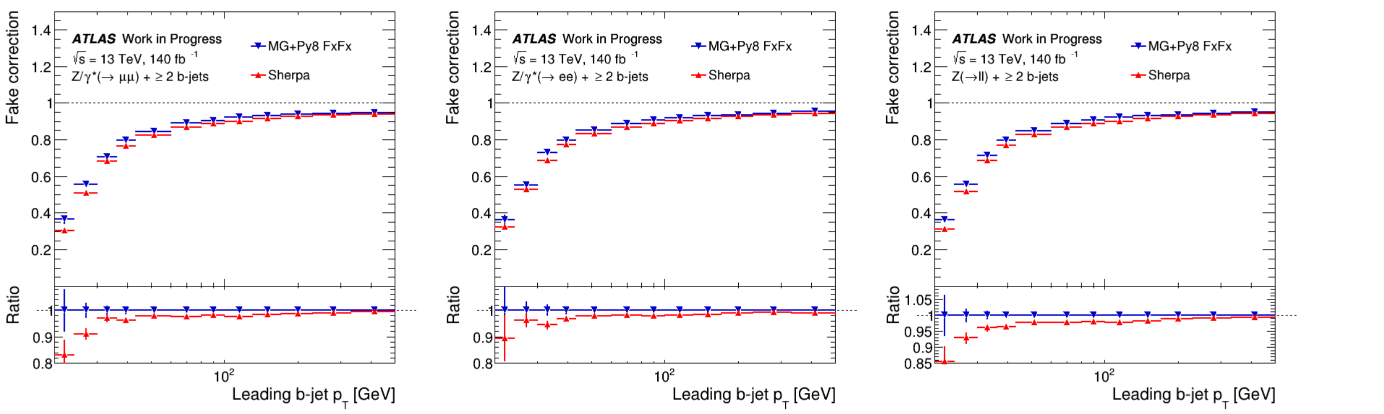

Figure 4.19: Fake corrections as a function of the p_T of the lead b-jet. The fake corrections are shown in the muon channel (on the left), in the electron channel (in the center) and in their combination (muon+electron channel) (on the right).

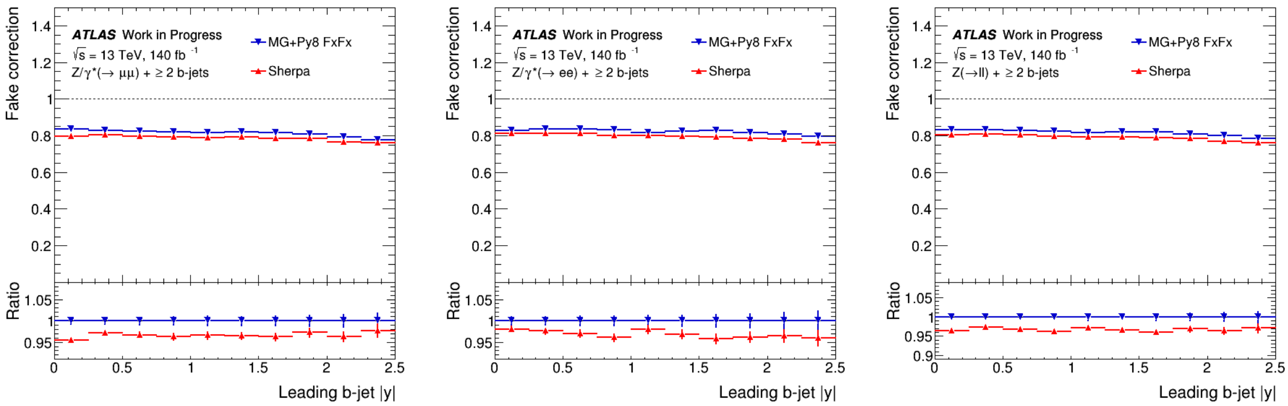

Figure 4.20: Fake corrections as a function $|Y|$ of the lead b-jet. The fake corrections are shown in the muon channel (on the left), in the electron channel (in the center) and in their combination (muon+electron channel) (on the right).

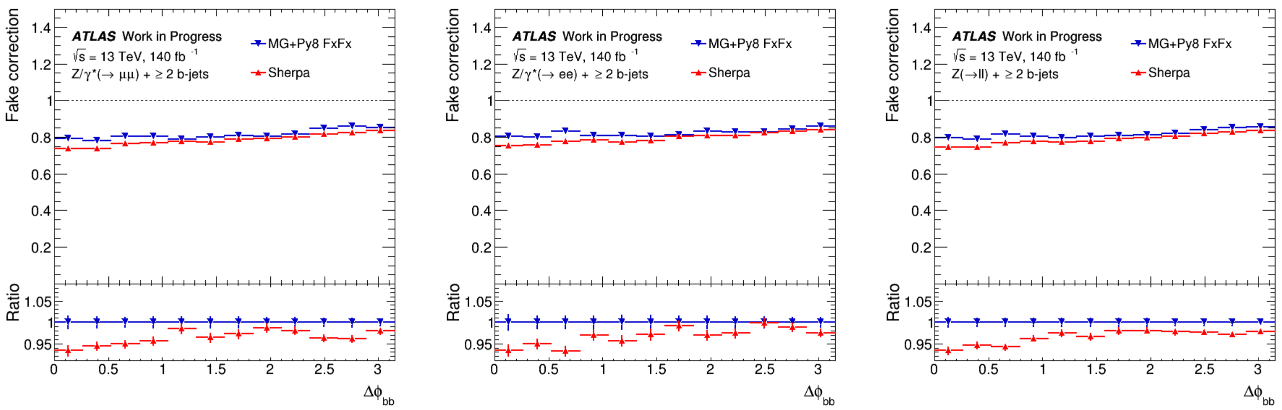

Figure 4.21: Fake corrections as a function the $\Delta\Phi$. The fake corrections are shown in the muon channel (on the left), in the electron channel (in the center) and in their combination (muon+electron channel) (on the right).

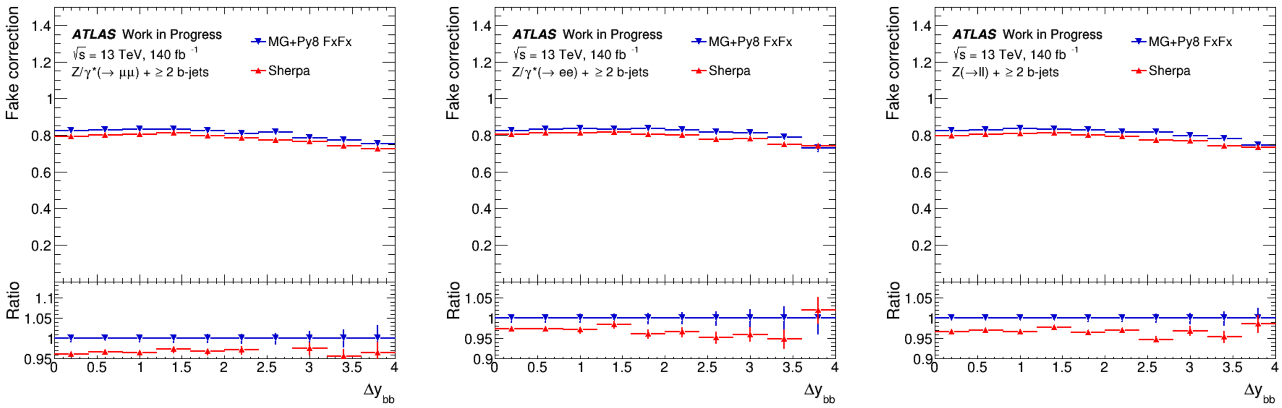

Figure 4.22: Fake corrections as a function of the ΔY. The fake corrections are shown in the muon channel (on the left), in the electron channel (in the center) and in their combination (muon+electron channel) (on the right).

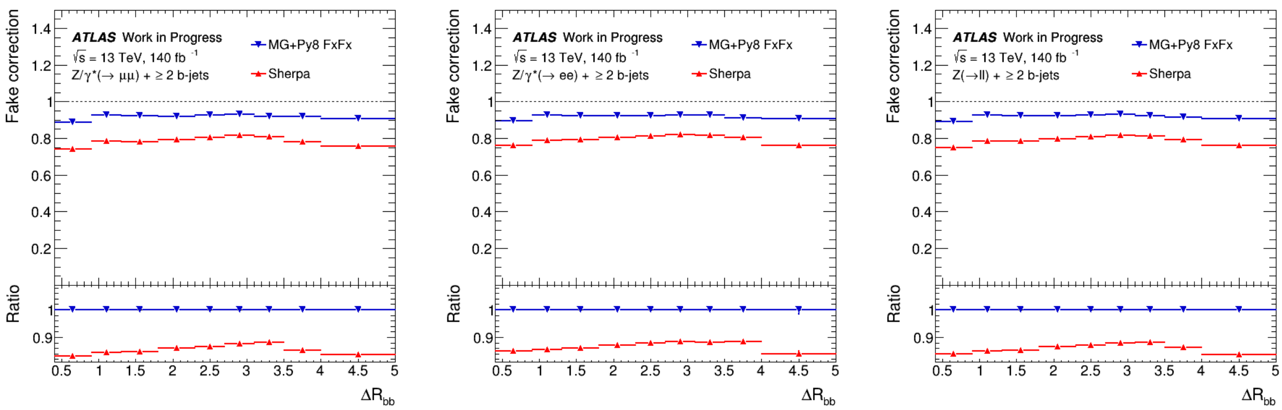

Figure 4.23: Fake corrections as a function of the ΔR. The fake corrections are shown in the muon channel (on the left), in the electron channel (in the center) and in their combination (muon+electron channel) (on the right).

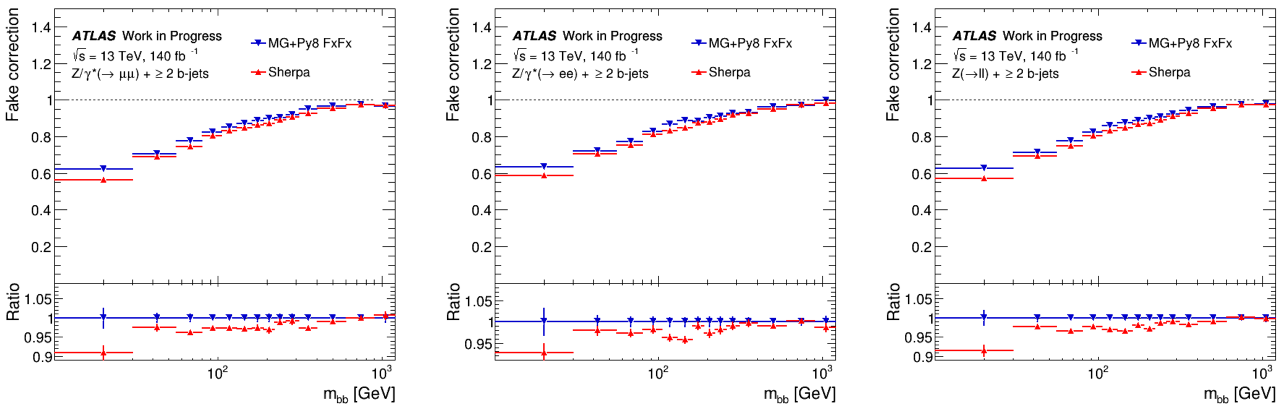

Figure 4.24: Fake corrections as a function of the M_{bb}. The fake corrections are shown in the muon channel (on the left), in the electron channel (in the center) and in their combination (muon+electron channel) (on the right).

4.1.4 Systematic uncertainties of the unfolded procedure

Uncertainties in the unfolding procedure arise from both the modeling of the observables in the MC simulation used for the unfolding and the statistical fluctuations within the MC samples:

- MC Statistics: The uncertainty from the limited statistics of the MG FxFx MC samples used for the unfolding is propagated into the systematic uncertainties using toy MC samples. The unfolding inputs are fluctuated independently according to Gaussian distributions. The procedure is performed 100 times and the RMS of the results is taken as systematic uncertainty.
- MC Modeling: the prior, the migration matrix, the efficiencies and the fake corrections used in the unfolding procedure of each observable are taken by the MG FxFx MC generator, these inputs may have an impact on the unfolded results. Sherpa generator is used to evaluate this uncertainty. In Sherpa indeed different hadronisation and parton-shower models are used for the signal simulation. For the estimation of this systematics, data are unfolded with Sherpa signal samples (i.e. using the prior, the migration matrix, the efficiencies and the fake corrections of Sherpa) and the difference between the unfolded data obtained with Sherpa and the nominal ones, obtained with MG FxFx, is taken as uncertainty.

4.2 Systematic Uncertainties of the Unfolded Results

Systematic uncertainties in the cross-section measurements stem from effects related to detector-level selection, background estimation, the unfolding method and the statistical uncertainty.

Uncertainties related to detector-level selection and background estimation are discussed in Section 3.6. These uncertainties are evaluated for each observable by propagating systematic variations from their respective sources to both the unfolding inputs derived from the signal MC samples and the subtracted background contributions. The dominant source of uncertainty is the b-tagging. The uncertainty for each background component is estimated by applying shifts to the subtracted background contributions. Uncertainties for Z+1 b-jet, Z+c-jets and Z+light-jets are determined through the flavor fit. Uncertainties related to the $t\bar{t}$ background are evaluated using the dedicated data-driven method.

The uncertainties associated with the unfolding procedure are detailead in Section 4.1.4.

In Figs. 4.25 4.26 the relative contribution of each systematic source as a function of the p_T and the $|Y|$ of the Z boson is illustrated. In Figs. 4.27 and 4.28 the relative contribution of each systematic source as a function of the p_T and the $|Y|$ of the lead b-jet is shown. In Figs. 4.29 4.30 4.31 4.32 it is shown respectively for the $\Delta\Phi$, ΔY_{bb}, ΔR and M_{bb}.

The uncertainties for each source have different colors as in the legend; the line in black represents the total systematic band.

The main experimental uncertainties are the ones related to jets, namely the combination of JES and JER, and the one related to b-tagging. They contribute about at the same level in each bin of each observable, with jets systematics being slighly larger in some bin, mainly as expected in low and high p_T regions where JES is determined with minor precision with respect to the central part of the spectrum. The other larger uncertainty is the one related to the unfolding procedure, where the main component impacting the cross-section measurement is the modeling uncertainty [1] Depending on the observable, the relative contribution of the main uncertainty changes.

[1] The approach used in this thesis is slightly conservtive. In the recent ATLAS paper on this analysis a more sofistated technique is used to estimate these systematics, but the approach used in the paper there goes beyond the scopes of this thesis. The range of unfolding uncertainty on $\Delta\Phi_{bb}$ moves from 5-8% (in this thesis) to 2-7% (in the paper).

For example, in the measurement of the pT and the $|Y|$ of the Z boson the unfolding term is the dominant one, while for the other observables the contribution of JES, JER and unfolding is on average similar.

Focusing on the results of the combination, where statistical fluctuations get mitigated, one observes than the total uncertainty ranges between 10% and 15% in the measurement of Z's $|Y|$, leading b-jet $|Y|$, $\Delta\Phi_{bb}$, ΔY_{bb}.

The total uncertainty on ΔR_{bb} is slighly larger, due to an issue with the electron channel, that at this stage hasn't been yet further investigated. It ranges between 10% and 20% in the Z's p_T, where the measurement is done up to 1 TeV. It ranges between 10% and 30% in the leading b-jets p_T, where the measurement is done up to 500 GeV, and in the M_{bb}, measured up to 1.2 TeV.

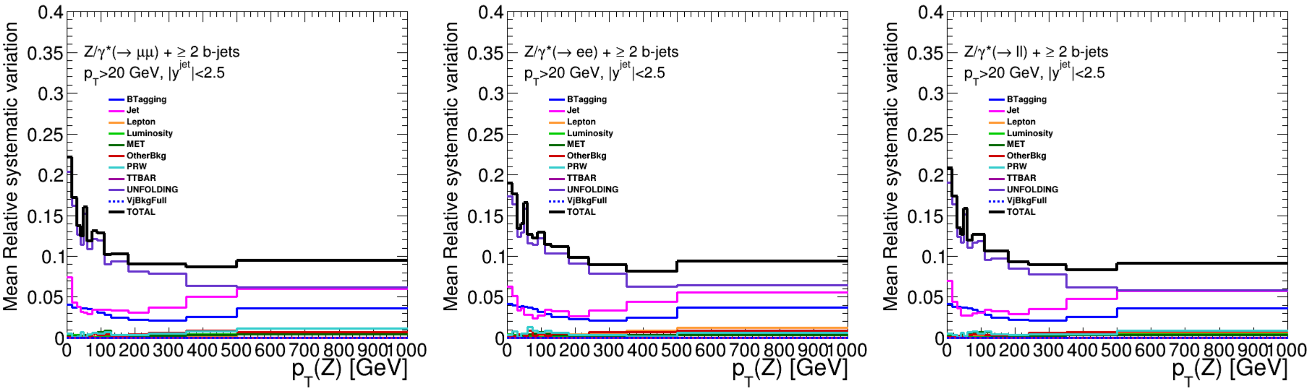

Figure 4.25: Relative sistematic uncertainties on the unfolded results of the p_T of the Z boson in muon (left), electron (center) and electron + muon (right) channels.

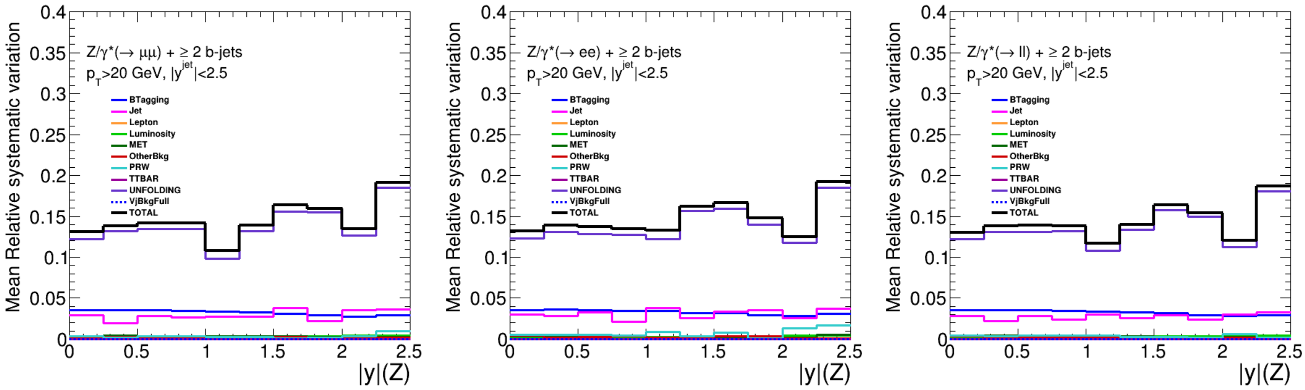

Figure 4.26: Relative sistematic uncertainties on the unfolded results of the $|Y|$ of the Z boson in muon (left), electron (center) and electron + muon (right) channels.

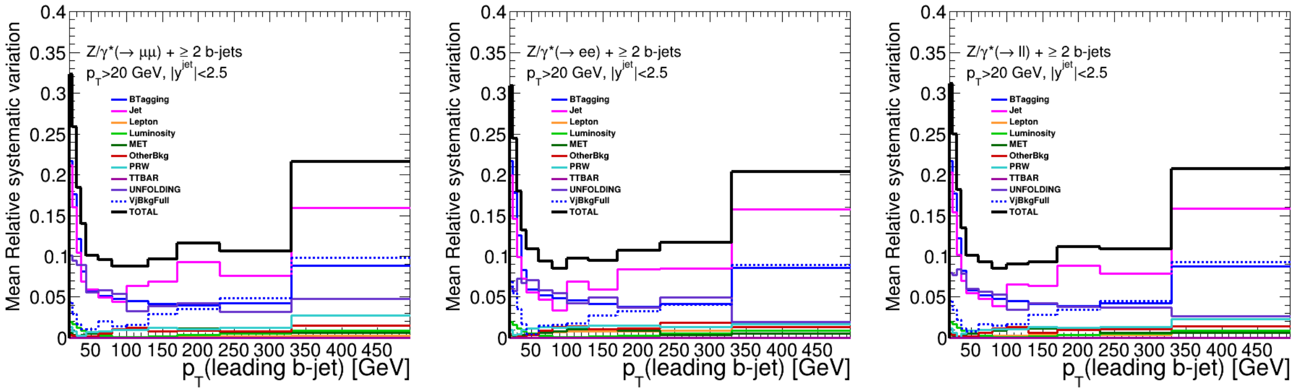

Figure 4.27: Relative sistematic uncertainties on the unfolded results of the p_T of the lead b-jet in muon (left), electron (center) and electron + muon (right) channels.

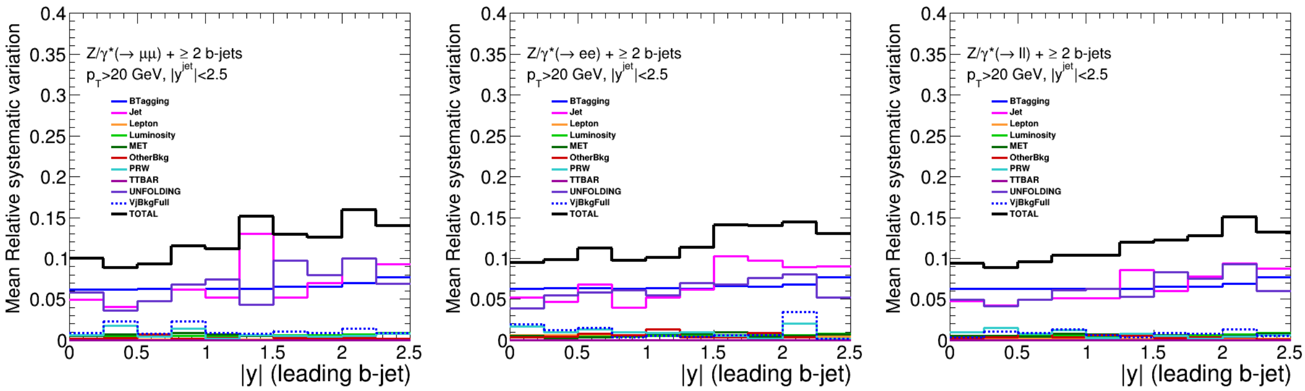

Figure 4.28: Relative sistematic uncertainties on the unfolded results of the $|Y|$ of the lead b-jet in muon (left), electron (center) and electron + muon (right) channels.

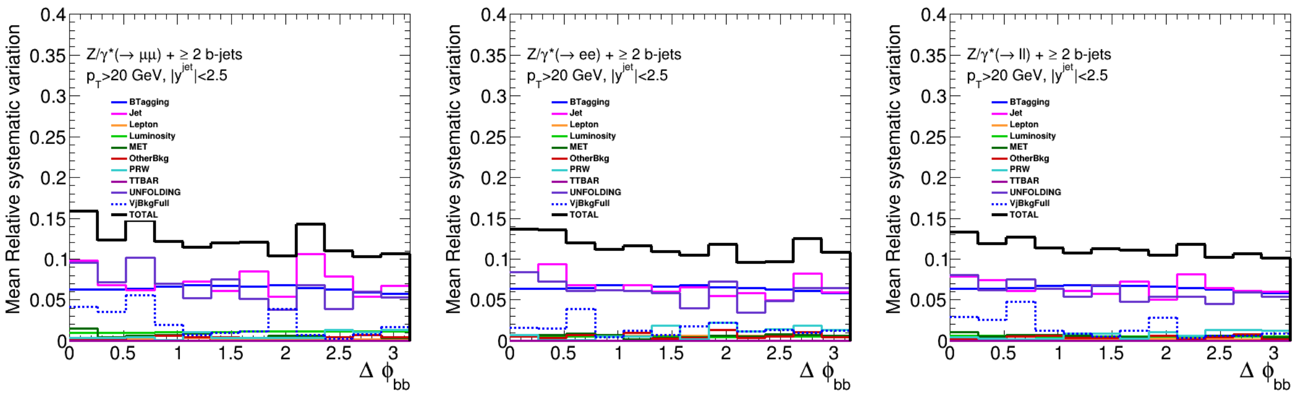

Figure 4.29: Relative sistematic uncertainties on the unfolded results of the $\Delta\Phi$ between the two b-jets in muon (left), electron (center) and electron + muon (right) channels.

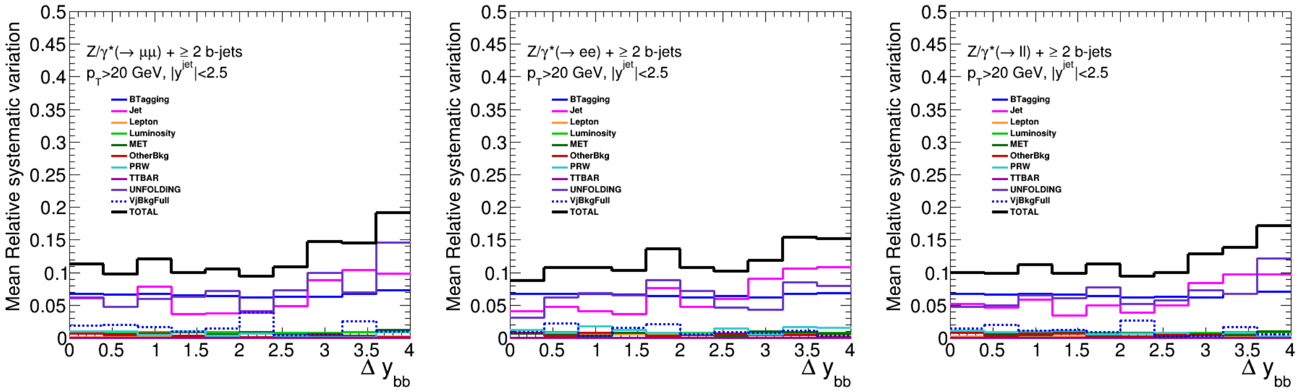

Figure 4.30: Relative sistematic uncertainties on the unfolded results of the ΔY between the two b-jets in muon (left), electron (center) and electron + muon (right) channels.

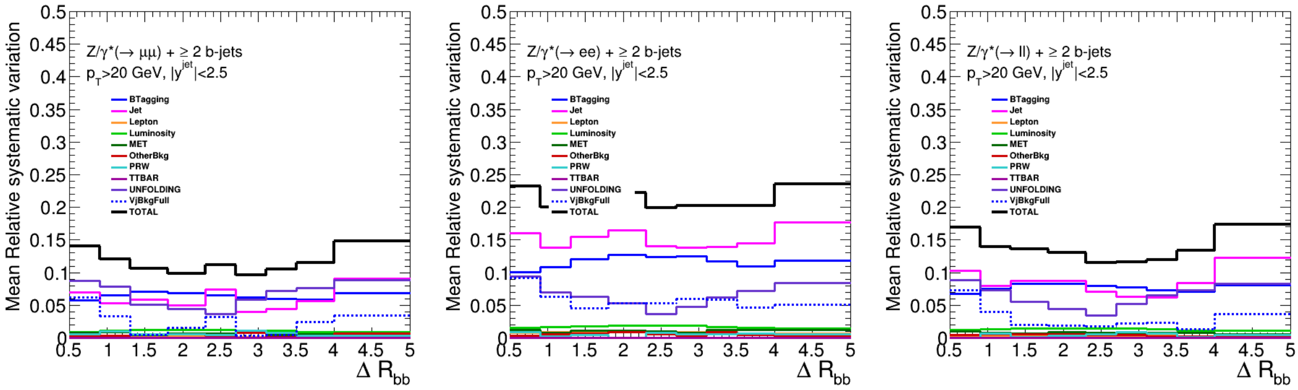

Figure 4.31: Relative sistematic uncertainties on the unfolded results of the ΔR between the two b-jets in muon (left), electron (center) and electron + muon (right) channels.

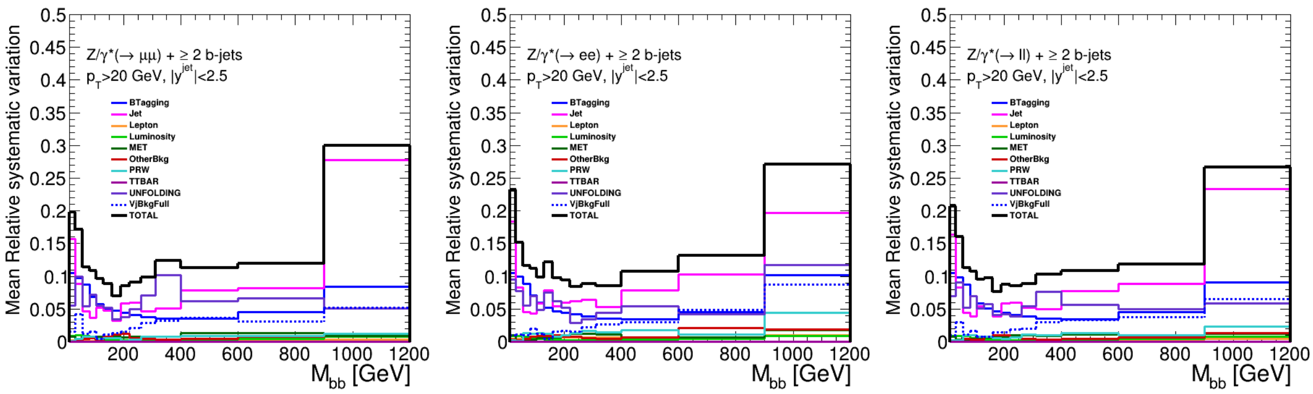

Figure 4.32: Relative sistematic uncertainties on the unfolded results of the invariant mass of the two b-jets M_{bb} in muon (left), electron (center) and electron + muon (right) channels.

4.3 Inclusive Cross-Section

The inclusive cross-section has been measured as described in Section 4.1: the systematic uncertainties are propagated through the unfolding via the unfolding factor (C_Z)and via the variation of the subtracted backgrounds. The measure is compared with standard 5FNS multi-leg MG+Py8 FxFx and Sherpa and with the 4FNS prediction as provided by MG+Py8 FxFx (4FNS). The results obtained as expected are identical to the ones published by ATLAS in the recent paper [26]. They are reported in Table 4.2.

2*$Z \to \ell^+\ell^- + \geq 2$ b-jet cross-section	2*$\sigma \pm$ stat. $\pm$ syst. [pb]
Data	$1.39 \pm 0.01 \pm 0.13$
MG5+Py8 FxFx 5FS (NLO)	$1.56 \pm 0.00\,(\text{stat})\,^{+0.14}_{-0.14}\,(\text{scale + PDF})$
Sherpa 5FS (NLO)	$1.25 \pm 0.00\,(\text{stat})\,^{+0.49}_{-0.21}\,(\text{scale + PDF})$
MGaMC+Py8 Zbb 4FS (NLO)	$1.27 \pm 0.01\,(\text{stat})\,^{+0.19}_{-0.16}\,(\text{scale + PDF})$

Table 4.2: Measured inclusive cross-sections for $Z+ \geq 2$ b-jet combining electron and muon channels.

Either 5FN and 4FN predictions agree with data for the inclusive cross-section and this is line also with previous ATLAS publication with partial Run-2 dataset [28].

4.4 Measured Differential Cross-Sections

The differential cross-sections are displayed in Figs. 4.33 4.34 4.35 4.36 4.37 4.38 4.39 4.40. For each measured observable it is shown the cross-section in the electron, muon and combined (electron+muon) channels. All measurements are compared with the predictions from the 5FNS multi-leg generators MG+Py8 FxFx and Sherpa, while $\Delta\Phi_{bb}$ and M_{bb}, Fig.s 4.35 and 4.36, are also compared with the 4FNS predictions provided by MG+Py8 FxFx (4FNS). [2]

In each figure the nominal measured cross-section is shown as black points, the black bars correspond to statistical uncertainties (they are not visible, being negligible, thanks to the large dataset used), while the dashed bands are the total uncertainties (fully dominated by systematic uncertainties). Also the predictions are shown with their uncertainties (fully dominated by scale uncertainties).

[2]For the other observables predictions from MG+Py8 FxFx(4FNS) are not currently available).

The distributions of the transverse momentum of the Z boson and of the b-jets probe pQCD over a wide range of scales and provide important input to the background prediction for other SM processes and searches beyond the SM. The differential cross-section as a function of $p_T(Z)$ is shown in Figure 4.33, while the differential cross-section as a function of the leading b-jet is shown in Figure 4.35. In general, in both cases the measured spectrum tends to be slightly harder than all predicted spectra. Overall, the prediction from MG+Py8 FxFx demonstrates the best agreement with data, while the Sherpa prediction still describes the data within larger theory uncertainties.

The distributions of the Z-boson rapidity and of the leading b-jet rapidity are sensitive to PDFs in addition to probing pQCD. The differential cross-section as a function of $|Y(Z)|$ is shown in Fig. 4.34, while the differential cross-section as a function of the leading b-jet $|Y|$ is shown in Fig. 4.36. In general, a good agreement is observed between data and predictions within the uncertainties. Sherpa (simulated with NNNPDF 3.0) describes very well the shape of the data, while MG+Py8 FxFx (simulated with NNNPDF3.1luxQED) tends to present a narrow distributions than data.

Fig.s 4.37, 4.38, 4.39 and 4.40 show the results on di-b-jets angular observables. The measurement of the angular separation between the two leading b-jets allows characterisation of the hard radiation at large angles and the soft radiation for collinear emissions. Fig. 4.37 shows the differential cross-section as a function of $\Delta\Phi_{bb}$. Both MG+Py8 FxFx and Sherpa describe the data shape well with some modulation, MG+Py8 FxFx (4FS) tends to slightly underestimate small and large $\Delta\Phi_{bb}$ values corresponding to collinear and back-to-back b-jets. Fig. 4.38 shows the differential cross-section as a function of the rapidity separation of the 2 leading b-jets ΔY_{bb}. This observable has some sensitivity to PDFs, but that is below the experimental uncertainties. A good modeling of MG+Py8 FxFx and Sherpa is observed. Fig. 4.39 shows the differential cross-section as a function of ΔR_{bb}. The ΔR_{bb} observable is sensitive to the various production mechanisms of the Z + $b\,\bar{b}$ final state. The region at low ΔR_{bb} is dominated by the production of two b-jets from gluon splitting. Probing this region requires two b-jets in the final state, so it is not sensitive to very small angles of the splitting. The interplay of the modelling of $\Delta\Phi_{bb}$ and ΔY_{bb} influences the prediction of the ΔR_{bb}. Sherpa predictions describe the shape of this observable quite well, featuring a substantial improvement at low ΔR_{bb} relative to the LO version reported by ATLAS using s = 7 TeV [22]. MG+Py8 FxFx shows some differences in shape with respect to data, but it still agree with data within the uncertainty.

Finally Fig. 4.40 presents the differential cross-section as a function of M_{bb}. The invariant mass of the two leading b-jets is an important observable in the measurement of associated ZH production with Higgs boson decays into $b\ \bar{b}$ and in searches for physics beyond the SM in the same final state. In general, all calculations predict steeper growth below 80 GeV and steeper decrease for higher values. Overall none of the predictions is able to describe the shape of the data distribution, despite they are still in agreement with data within the uncertainties in most of the bins.

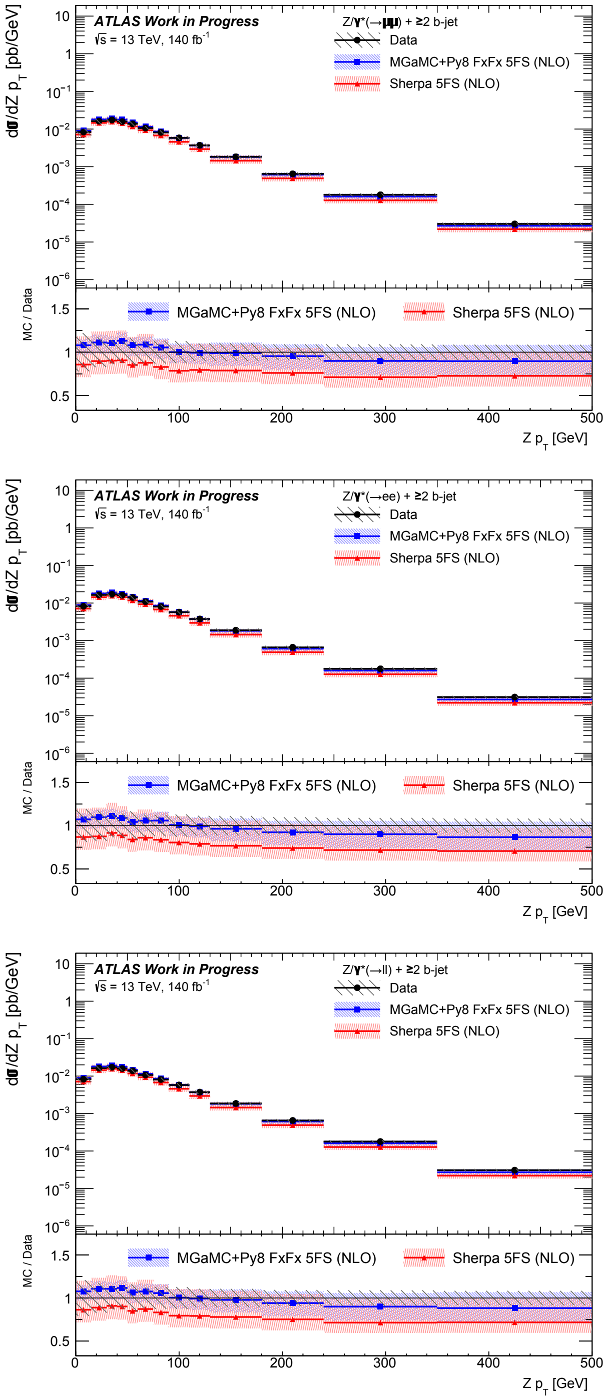

Figure 4.33: Differential cross-section for the production of a Z boson in association with at least 2 b-jets as a function of the p_T of the Z boson in electron, muon and electron+muon channels.

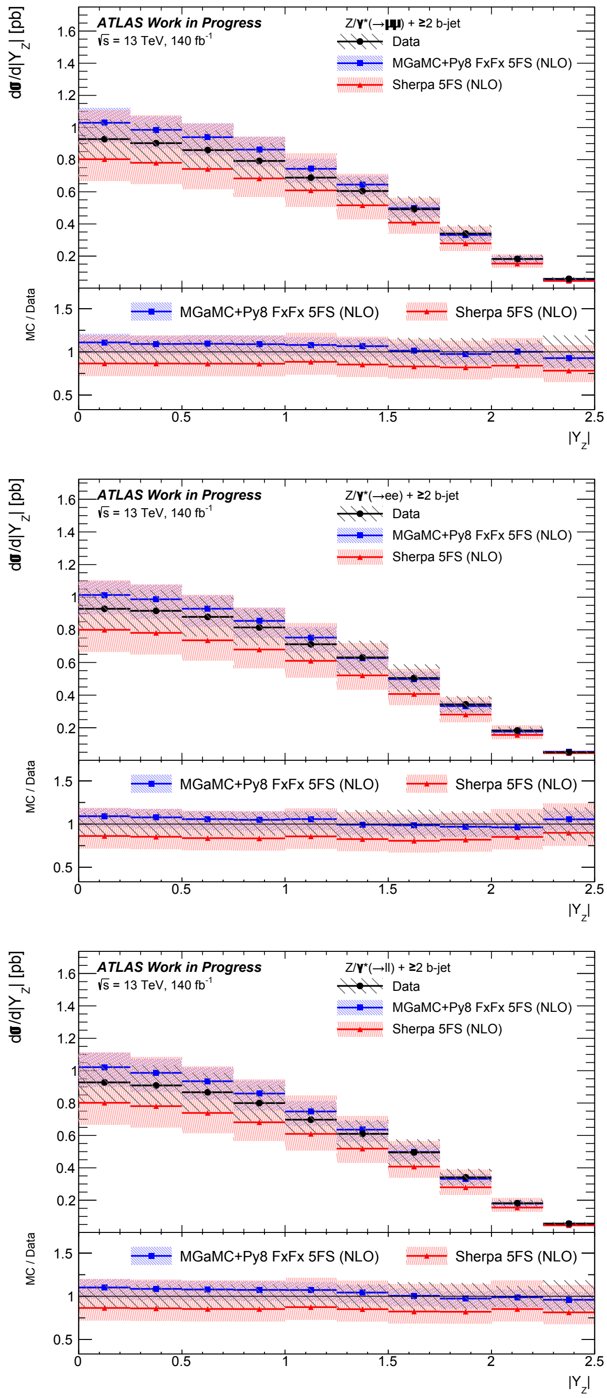

Figure 4.34: Differential cross-section for the production of a Z boson in association with at least 2 b-jets as a function of the $|Y|$ of the Z boson in electron, muon and electron+muon channels.

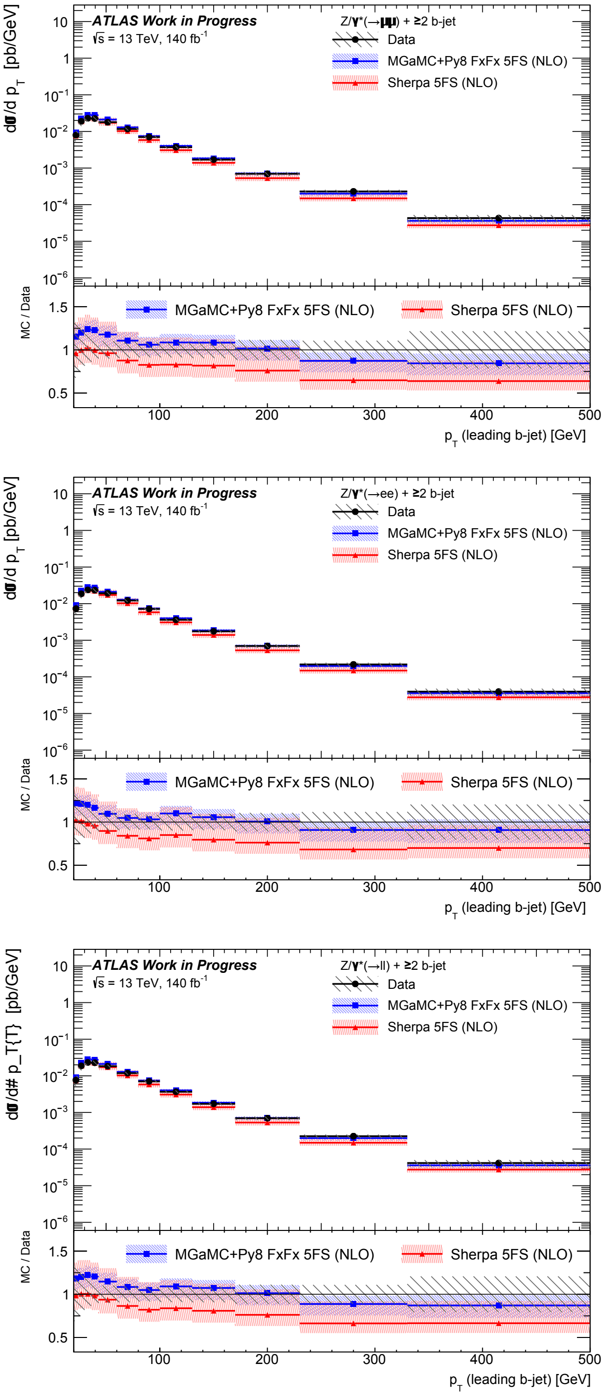

Figure 4.35: Differential cross-section for the production of a Z boson in association with at least 2 b-jets as a function of the p_T of the lead b-jet in electron, muon and electron+muon channels.

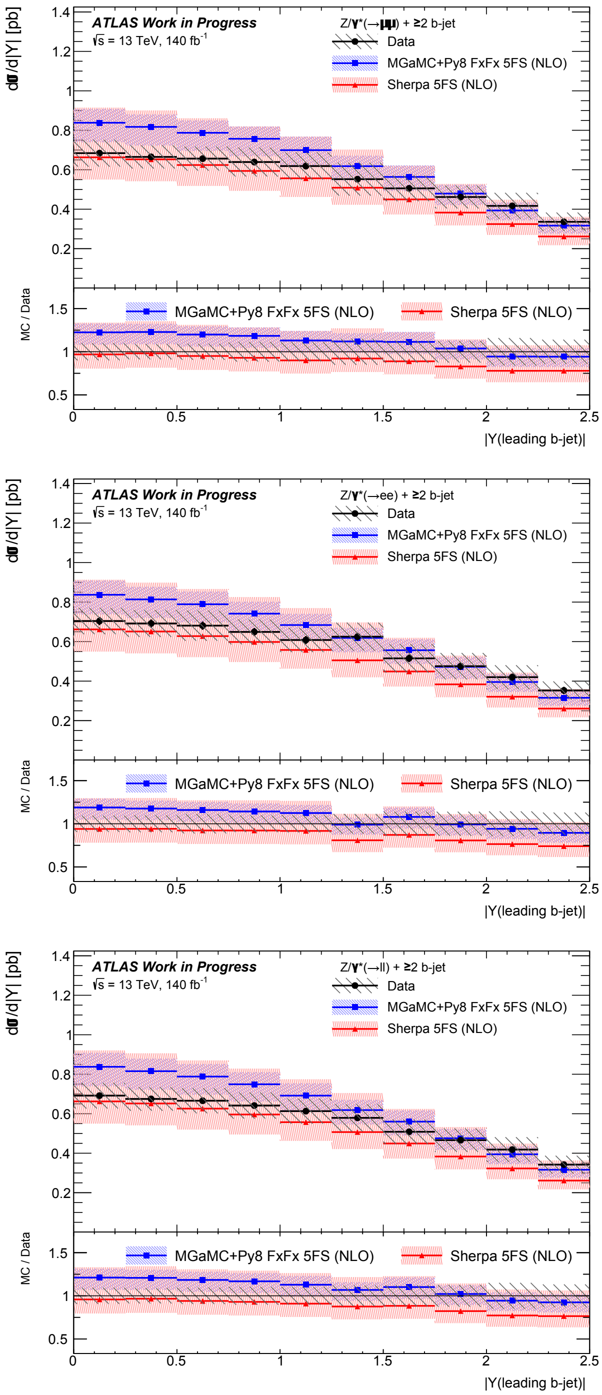

Figure 4.36: Differential cross-section for the production of a Z boson in association with at least 2 b-jets as a function of the $|Y|$ of the lead b-jet in electron, muon and electron+muon channels.

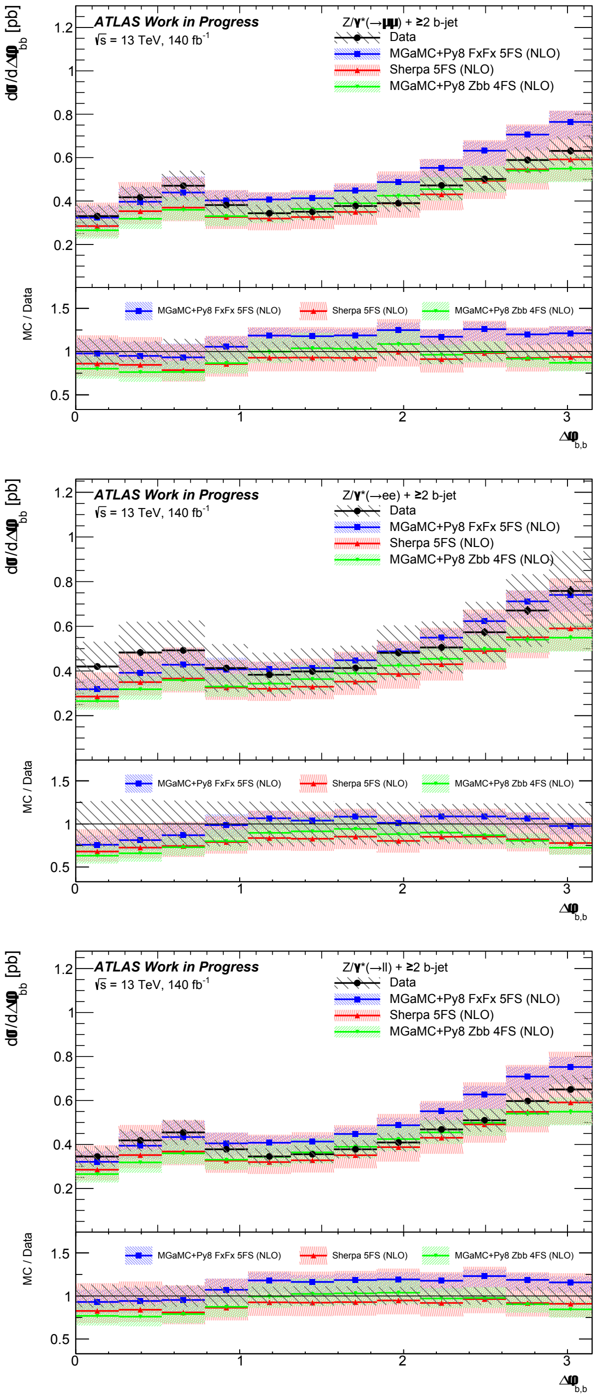

Figure 4.37: Differential cross-section for the production of a Z boson in association with at least 2 b-jets as a function of $\Delta\Phi$ in electron, muon and electron+muon channels.

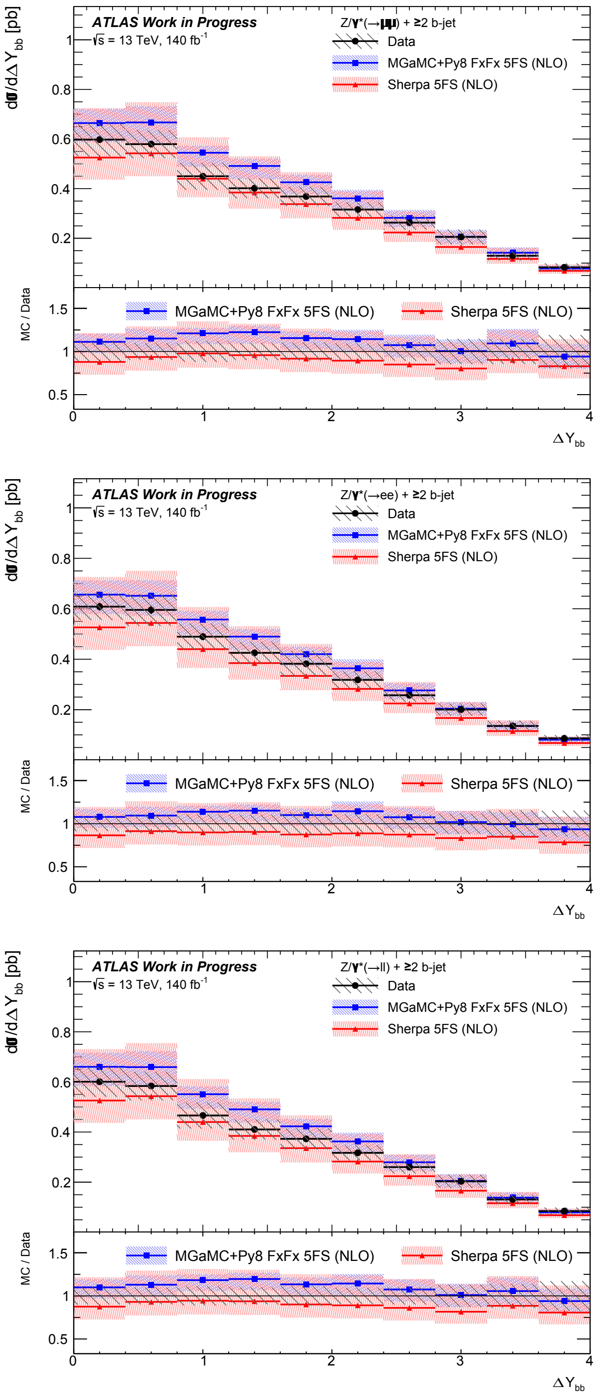

Figure 4.38: Differential cross-section for the production of a Z boson in association with at least 2 b-jets as a function of ΔY in electron, muon and electron+muon channels.

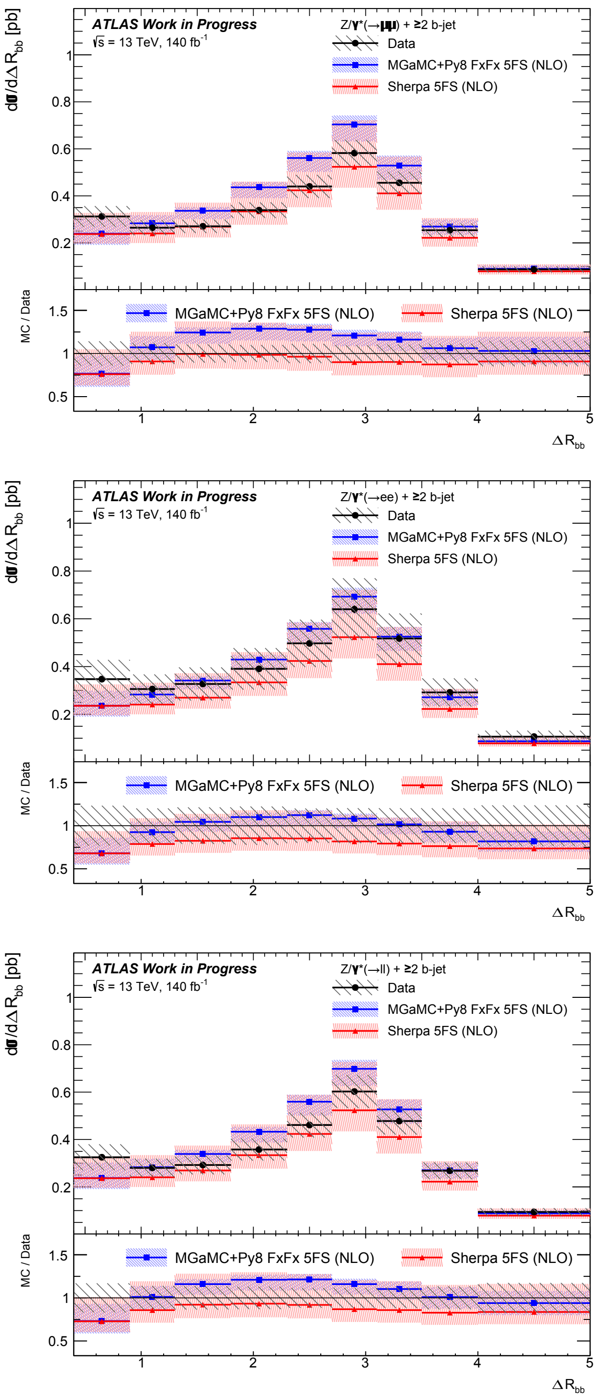

Figure 4.39: Differential cross-section for the production of a Z boson in association with at least 2 b-jets as a function of ΔR in electron, muon and electron+muon channels.

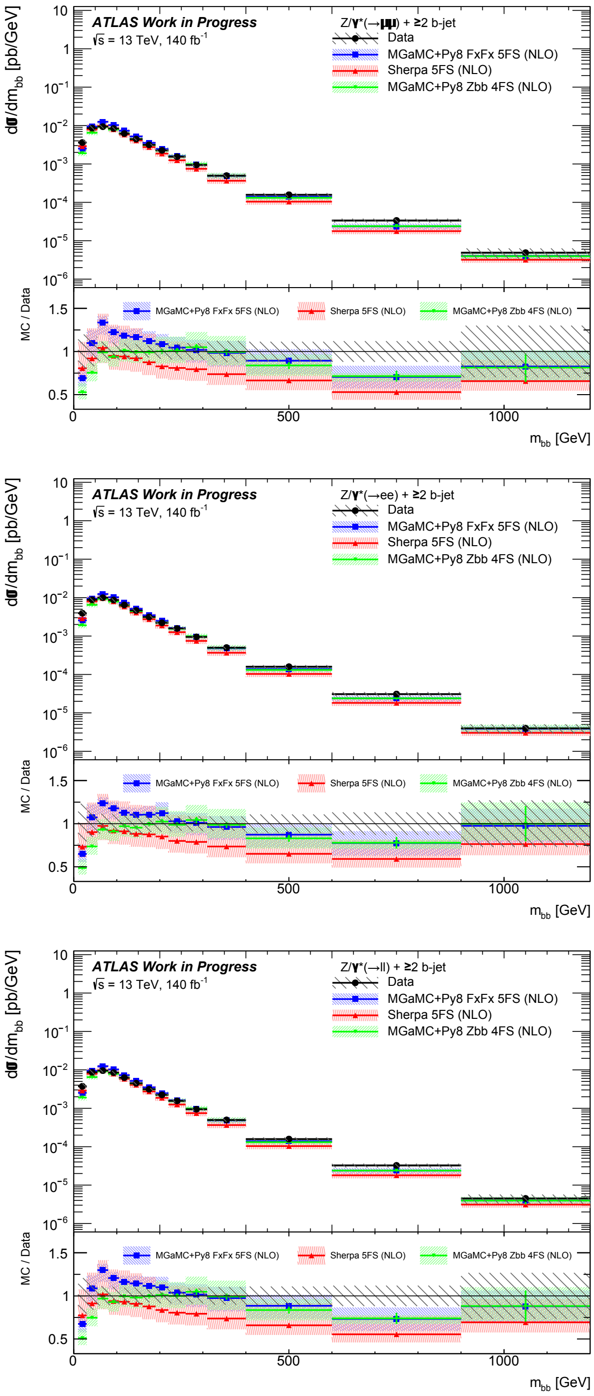

Figure 4.40: Differential cross-section for the production of a Z boson in association with at least 2 b-jets as a function of M_{bb} in electron, muon and electron+muon channels.

Conclusion

This thesis presents the cross-section measurements of the Z boson production in association with at least two b-jets (Z + $\geq$ 2 b-jets) in proton-proton collisions at a center-of-mass energy of $\sqrt{s}$ = 13 TeV collected by the ATLAS experiment at the LHC at CERN. The measurement has been performed using the full Run 2 dataset corresponding to an integrated luminosity of 140 fb^{-1}.

In this analysis, the Z boson is identified through its leptonic decay ($e^+ + e^-$ or $\mu^+ + \mu^-$) and it is associated to the b-jets, namely jets coming from the hadronisation of b-quarks.

The inclusive and differential cross-section measurements are provided in a fiducial space space very close to the to the detector acceptance to minimize as much as possible modeling uncertainties and it is performed separately in the electron channel and the muon channel and in their combination. The differential cross-sections were measured as a function of the following observables: the transverse momentum (p_T) and rapidity ($|Y|$) of the Z boson, the transverse momentum (p_T) and rapidity ($|Y|$) of the leading b-jet, $\Delta\Phi_{bb}$, ΔY_{bb} and ΔR_{bb} between the leading and sub-leading b-jets and the invariant mass of the two b-jets M_{bb}. The inclusive cross-section measurements and the differential ones as a function of $\Delta\Phi_{bb}$ and M_{bb} have been recently published in a paper [26] and have been used in this thesis for validating the analysis described in this document. The measurements of differential cross-section as a function of the p_T and the $|Y|$ of the Z boson and of the leading b-jet and as a function of ΔY_{bb} and ΔR_{bb} represent the novelty of this thesis. Indeed they haven't been never performed before with the full Run 2 dataset.

The backgrounds have been estimated partially with MC simulations and partially with data-driven techniques. In particular the background from $t\bar{t}$ pairs and from multijets has been estimated with data-driven techniques. The signal (Z+ $\geq$2 b-jets) has been separated by Z+jets background, namely Z+light-jets, Z+c-jets and Z+1b-jet with a fit to data using a flavour sensitive observable.

After subtracting the background from the data, the data at detector level have been unfolded to particle level using the Bayesian unfolding method. The systematic uncertainties at the detector level has been propagated through the unfolding procedure and the unfolding has been performed separately for each systematic variation in a correlated manner.

The unfolded data distributions, namely the cross-section measurements, are compared with the available signal predictions from state-of-the-art MC generators. In particular, the measurements are compared with standard 5FS multi-leg MGaMC+Py8 FxFx and Sherpa 2.2.11 predictions, both have NLO matrix-element supplemented by PS for the higher orders, and with a 4FS prediction, as implemented in MGaMC+Py8 4FS, that also have a NLO matrix-element supplemented by PS. In general, the 5FS MG+Py8 FxFx and Sherpa generators describe the Z + $\geq$ 2 b-jets data within the experimental and theory uncertainties. Sherpa predictions show larger uncertainties (up to 40%) with respect to MG+Py8 FxFx (that are up to 20-30%).

The generator that describes better the inclusive data is MG+Py8 FxFx, while Sherpa tends to underestimate the measured inclusive cross-section. The best description of the shape of angular variables namely ($|Y_Z|$, $|Y|$ of the leading b-jet,$\Delta\Phi_{bb}$, ΔY_{bb}, ΔR_{bb}) is provided by Sherpa. It is interesting to mention the large improvement of new MC generations in describing the small ΔR_{bb} region with respect to previous MCs used in Run 1. The spectra of p_T distributions are slightly softer than the data, eithever using MG+Py8 FxFx and Sherpa, but MG+Py8 FxFx shows the best agreement with data. For two observables, namely $\Delta\Phi_{bb}$ and M_{bb}, MGaMC+Py8 4FS is also compared to data. Also in this case the predictions are consistent with the data within the uncertainties with a tendency to underestimate the measured inclusive cross-section and to underpredict collinear and back-to-back b-jet emissions. Finally none of the three MCs describes the shape of data for M_{bb}.

In comparison with the recent ATLAS publication on this final state with the same dataset, this thesis explores additional observables that are important test of pQCD prediction, have sensitivity to PDFs and last but not least are important inputs for the tuning of the MC generators used to describe Z+b-jets processes in analysis where they are background, i.e. Higgs measurements or searches of new physics.

In comparison with the previous ATLAS measurement with partial Run-2 dataset where some of these variables were measured, more extreme regions of the phase space are explored. The higher integrated luminosity contributes to increase the number of selected events for

rare final states with two b-jets. Advances in reconstruction and identification tools, mainly the b-tagging one, allow a reduction of the uncertainties and thus increase the sensitivity to the signal processes. Therefore this measurement provides important inputs for the improvement of theoretical predictions and MC generators of boson production in association with b-jets, allowing a better quantitative understanding of perturbative QCD.
The uncertainties in these measurements range between 10% and 20% and are fully dominated by systematics. This result can be still improved enhancing the performances of the flavour-tagging techniques and improving unfolding methods. Moreover this thesis only explores the resolved region, where the two b-quarks are reconstructed as two separated jets, significant work remains to be done in the boosted region, where the jets have higher transverse momentum and are reconstructed as a single, large-radius jet. This region will allow for studying the internal structure of jets and color flow effects. These opportunities will be further enhanced by Run 3, which is operating at an increased energy of $\sqrt{s} = 13.6$ TeV and is expected to provide a luminosity of almost 300 fb^{-1} (two times the current dataset). To conclude, the upcoming studies of Z+b-jets at ATLAS promise to be extremely exciting in the incoming future. The experiment is going to explores new territories, probing new final states and new observables supported by improved reconstruction tools, these could lead to substantial improvement in our understanding of pQCD.

Unfolding

Several techniques are commonly used for unfolding, including the bin-by-bin method, regularized unfolding and the Bayesian approach. Each method has its advantages and is chosen based on the complexity of the measurement and the level of regularization required to mitigate statistical fluctuations and detector smearing.

Unfolding ensures that the reconstructed data can be compared directly with theoretical predictions or simulations, which are typically expressed at the truth level.

Bin-by-bin

The bin-by-bin method is the simplest unfolding technique. The unfolded distribution is computed as:

$$\hat{x}_i = (y_i - b_i)\frac{N_i^{gen}}{N_i^{rec}} \tag{7}$$

where N_i^{gen} is the total number of generated MC events in bin i, N_i^{rec} is the total number of reconstructed MC events in bin i, $\hat{x}_i$ is the unfolded distribution for bin i.

This method calculates the ratio between the number of reconstructed events and the number of true events in each bin of the distribution.

To apply this method two conditions must be satisfied:

1. The true and experimental variables must have the same binning;

2. The method works well only if there is little event migration between bins. In other words, the detector smearing (the uncertainty in reconstructing the true observable) should be smaller than the bin size.

An advanced version of this method which allows for bin migration by constructing a response matrix that links the number of generated events in one bin to the number of observed events in all bins.

Once this matrix is built, it is inverted and applied to the observed data to recover the true distribution.

In this case, the unfolded result is given by:

$$\hat{x} = A^{-1}(y - b) \tag{8}$$

where A is the response matrix and y and b are the observed data and background, respectively.

While this approach can handle bin migration, it introduces new challenges. Indeed, the inverse of the response matrix does not always exist. Even when it does, the inversion can amplify statistical fluctuations, particularly in cases with limited data, leading to unstable and unreliable results.

Another drawback can be that a requirement of this method is that the true and experimental variables still require identical binning.

Therefore, despite its simplicity, the bin-by-bin method has significant limitations. The strict requirements on binning and stability, combined with its inability to handle large event migration or statistical noise, make it less suitable for many practical applications.

Regularized unfolding

To address the shortcomings of simpler unfolding methods, a more sophisticated approach is **regularized unfolding**. This method works by suppressing spurious oscillations in the solution, which can arise due to statistical fluctuations in the data. The key idea of regularization is to incorporate prior knowledge about the smoothness of the true distribution. By doing so, it aims to stabilize the unfolding process and prevent overfitting to noise in the observed data.

The regularization is achieved by imposing a constraint based on prior assumptions about the degree of smoothness in the true distribution. However, it is crucial to determine the appropriate weight of this prior information using statistical techniques because it can introduce bias into the result.

While regularization improves stability, the unfolded solution is limited in terms of the number of data points. This limitation leads to a loss of resolution and a decrease in statistical accuracy. As a result, the method is best suited for cases where stability and smoothness are more important than maximizing precision.

Despite its advantages in stabilizing the results, regularized unfolding has not been widely adopted due to technical complexities.

Bayes method

The Bayesian method is undoubtedly the best unfolding technique. Unlike other methods, it can be applied to multidimensional problems, whereas many traditional approaches are limited to one-dimensional cases. It allows for the use of cells of different sizes in the distribution of both the true and experimental values. Additionally, the Bayesian method can handle event migration between bins, which is a major limitation in simpler unfolding techniques.

One of its key advantages is that it does not require matrix inversion, avoiding the risk of amplifying statistical fluctuations and ensuring a more stable solution. The method also naturally provides the correlation matrix of the unfolded results, which is crucial for evaluating uncertainties and correlations between bins. Moreover, the Bayesian unfolding is relatively simple and fast to implement, making it a practical choice for many complex analyses.

Bayes theorem

In the Bayes theorem one has to identify an effect E and several independent causes C_i. Supposing to know the initial probability of the $i-th$ cause $P(C_i)$ and the conditional probability of the $i-th$ cause to produce the effect $P(E|C_i)$, the Bayes formula is

$$P(C_i|E) = \frac{P(E|C_i)P(C_i)}{\sum_{l=1}^{n_c} P(E|C_l)P(C_l)} \tag{9}$$

This means that the probability that a specific event E is caused by an $i-th$ cause is proportional to the probability that the $i-th$ cause produced the event E and the initial probability of the cause $P(C_i)$.

A powerful feature of this theorem is that, increasing the number of the observations, also the knowledge of $P(C_i)$ increases.

In contrast, $P(E|C_i)$,which describes the likelihood of the event given the cause, remains constant and can be estimated using MC methods.

If there are multiple events, denoted as $n(E)$, the expected number of events assignable to each cause is

$$\hat{n}(C_i) = n(E)P(C_i|E) \tag{10}$$

and if the effects are more than one (E_j), it is possible to evaluate, similarly to Equ. 9, $P(C_i|E_j)$ as

$$P(C_i|E_j) = \frac{P(E_j|C_i)P_0(C_i)}{\sum_{l=1}^{n_c} P(E_j|C_l)P_0(C_l)} \tag{11}$$

where $P_0(C_i)$ represents the initial probability of the causes.

For the above formula to hold, the following conditions must be met:

- $\sum_{i=1}^{n_c} P_0(C_i) = 1$, ensuring that the total probability across all causes equals 1;

- $\sum_{i=1}^{n_c} P(C_i|E_j) = 1$, meaning each effect can come from one or more causes, including background sources;

- it is not required that each cause produces at least one effect.

Given a number of observations for each event, an estimator for the unfolded distribution, i.e. the expected number of events assigned to each cause due to the observed events is, using Equ. 10,

$$\hat{n}(C_i)|_{obs} = \sum_{j=1}^{n_E} n(E_j) P(C_i|E_j) \tag{12}$$

and, considering the estimation of the efficiency, the true number of events is

$$\hat{n}(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_E} n(E_j) P(C_i|E_j) \tag{13}$$

The true total number of events is then equal to

$$\hat{N}_{true} = \sum_{i=1}^{n_C} \hat{n}(C_i) \tag{14}$$

The efficiency is estimated as

$$\hat{\epsilon} = \frac{N_{obs}}{\hat{N}_{true}} \tag{15}$$

and it differs from the a priori overall efficiency $\epsilon_0 = \frac{N_{rec}}{N_{gen}}$.

As the initial distribution $P_0(C)$ approaches the true one, the agreement improves, meaning that the estimated $\hat{P}(C)$ will lie between the initial guess $P_0(C)$ and the true distribution.

It is possible to perform the unfolding iteratively:

1. The initial distribution $P_0(C)$ and the initial expected number of events $n_0(C_i)$ are chosen;

2. $\hat{n}(C)$ and $\hat{P}(C)$ are calculated;

3. χ^2 comparison between $\hat{n}(C)$ and $n_0(C)$ is done;

4. $P_0(C)$ is replaced by $\hat{P}(C)$ and $n_0(C)$ is replaced by $\hat{n}(C)$;

5. if the χ^2 is smaller than a certain chosen threshold the iterations are stopped, otherwise these steps are repeated.

As previously mentioned, this method allows for an efficient treatment of the background, providing flexibility in the choice of the initial probabilities for the causes.

In fact, this approach can also be applied to study different physics processes that share similar signatures.

Estimation of the uncertainties

In order to determine the uncertainties in this procedure, it is possible to rewrite $n(C)$ as

$$\hat{n}(C_i) = \sum_{j=1}^{n_E} M_{ij} n(E_j) \tag{16}$$

where M_{ij} is the unfolding matrix.

It is crucial to identify the sources of uncertainty for each term that contributes to the covariance matrix of $\hat{n}(C_i)$:

- $P_0(C_i)$ is only affected by systematic uncertainties, which are evaluated by studying the stability of the results when varying the initial assumptions.

- The contribution from $n(E_j)$ is evaluated as a function of the unfolding matrix and the true number of events. If the relative frequency in each cell is sufficiently low, either due to a large number of bins or low efficiency, the numbers in each of the cells can be approximated by an independent Poisson distribution.

- The contributions from $P(E_j|C_i)$ are estimated using MC simulations and account for both statistical (arising from the limited number of simulated events) and systematic (due to assumptions in the simulation) errors. For this contribution, correlations must also be considered. In fact, the total number of events generated for each C_i cell is distributed among the E_j effect-cells. For this reason, it is not always possible to neglect the covariance between $P(E_{j1}|C_i)$ and $P(E_{j2}|C_i)$ due to migration effects. However, the covariance between terms related to different cause-cells can generally be neglected.

Acknowledgments

Il cammino verso la realizzazione di un tanto desiderato traguardo è un viaggio condiviso, in cui ogni passo è segnato dall'aiuto e dal sostegno di chi ci circonda. Desidero quindi esprimere il mio ringraziamento a tutti coloro che hanno fatto parte di questo capitolo della mia vita.

Attraverso questo percorso ho avuto la possibilità di esplorare il mondo della fisica, per il quale ho sempre avuto e continuo ad avere una profonda curiosità. L'amore e la dedizione per questa disciplina sono state anche supportate dall'esempio dei docenti, ognuno dei quali ha segnato il mio percorso di studi e ha contribuito ad ampliare le mie conoscenze. Tra tutti un profondo ringraziamento va alla Professoressa Evelin Meoni che ha seguito con dedizione, professionalità e piena disponibilità il mio percorso di tesi magistrale, per avermi introdotto nel mondo delle particelle e per avermi permesso di dare il mio contributo nell'esperimento ATLAS. La ringrazio anche per l'incoraggiamento e il supporto che mi ha riservato riguardo al futuro della mia carriera.

Ringrazio tutti i miei amici, i colleghi universitari e i colleghi del conservatorio per il sostegno e l'amicizia che hanno reso speciale questo percorso.

Una cosa che in questi 5 anni non è mai venuta a mancare è stato l'appoggio della mia famiglia, dei miei zii, dei miei cugini, di Rita e dei miei nonni. Ognuno di loro sa il ruolo che ha nella mia vita. Ringrazio particolarmente, però, chi è costantemente presente e chi c'è stato nei momenti di difficoltà sacrificando sempre un po' di sé e della propria famiglia.

Colei che veramente ha condiviso tutto con me in questi 5 anni è Teresa. Senza di lei probabilmente non sarei qui. Ogni esame, ogni giornata di studio, anche a distanza, ogni problema o gioia personale l'ho condivisa con lei. Se di questi anni universitari dovessi trovare la cosa che più mi sta a cuore, è sicuramente l'amicizia che abbiamo creato.

Un ringraziamento speciale ad Ivo, per avermi insegnato che amarsi significa anche e soprattutto sostenersi in ogni campo della propria

vita. Lo ringrazio per esserci sempre stato negli ultimi due anni, per aver tifato per me con tutte le sue forze per ogni sfida che ho affrontato, per avermi profondamente amato, nonostante le difficoltà che la distanza porta con sé.

Ringrazio mio padre per essere, come lui sa, un pilastro fondamentale della mia vita, per avermi insegnato che la costanza, la perseveranza e l'amore per ciò che si fa prima o poi verranno premiati. Lo ringrazio anche per tutto ciò che ogni giorno fa per me perché senza di lui non sarei quella che sono.

A mia mamma, di cui sento la presenza e il sostegno continuo, perché nonostante il destino ci abbia diviso, so che sarebbe fiera che io abbia realizzato quello che lei ha sempre saputo, fin da bambina, essere il mio sogno.

A mio fratello Pietro, per essere sempre al mio fianco, compagno di vita e di avventure. La sua presenza costante, il suo sostegno e la sua capacità di farmi sorridere anche nei momenti più difficili hanno reso questo percorso più leggero. A lui, a cui sono indissolubilmente legata, dedico con amore questo traguardo.

Bibliography

[1] David Griffiths. *Introduction to Elementary Particles.* Second, revised Edition, WILEY-VHC.

[2] J. D. Lykken. *Beyond the Standard Model.* CERN Yellow Report. arXiv:1005.1676.

[3] G.Bellini, L.Ludhova, G.Ranucci, F.L.Villante. *Neutrino oscillations.* arXiv:1310.7858v1.

[4] Quantum Diaries url: www.quantumdiaries.org

[5] F.Mandl, G.Shaw. *Quantum Field Theory.* Wiley and Sons, 2010.

[6] F.Halzen, A.D.Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics.* Wiley and Sons, 1984.

[7] G.Altarelli, G.Parisi. *Asymptotic freedom in parton language.* Nuclear Physics B, 126 (2): 298-318,1977.

[8] Y. L. Dokshitzer. *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+ e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.* Sov. Phys. JETP, 46: 641–653, 1977.

[9] V. N. Gribov and L. N. Lipatov. *Deep inelastic ep scattering in perturbation theory.* Sov. J. Nucl. Phys., 15: 438–450, 1972.

[10] NNPDF Collaboration. url:nnpdf.mi.infn.it

[11] CTEQ Collaboration. url:cteq.gitlab.io

[12] MSTW Collaboration. *Parton distributions for LHC.* Eur.Phys.J.C 63 (2009).

[13] S.Carrazza, J. Cruz-Martinez, R. Stegeman. *A data-based parametrization of parton distribution functions.* Eur. Phys. J. C (2022) 82:163. https://doi.org/10.1140/epjc/s10052-022-10136-z

[14] B.R. Webber. *Fragmentation and Hadronization.* CERN-TH/99-387, Cavendish-HEP-99/16 arXiv:hep-ph/9912292.

[15] S. Agostinelli et al. *GEANT4 - a simulation toolkit.* Nucl. Instrum. Meth. A 506 (2003) 250.

[16] ATLAS Collaboration *AtlFast3: the next generation of fast simulation in ATLAS*. CERN-EP-2021-174, 6 September 2021. arXiv:2109.02551.

[17] R.L. Workman et al. (Particle Data Group). *QUARKS*. Prog.Theor.Exp.Phys. 2022, 083C01 (2022).

[18] ATLAS Collaboration. https://opendata.atlas.cern

[19] ATLAS Collaboration. *Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13TeV with the ATLAS detector*. CERN-EP-2020-087, 08 March 2021. arXiv:2007.02873v2

[20] CDF Collaboration. *Measurement of cross sections for b jet production in events with a Z boson in $p\overline{p}$ collisions at $\sqrt{s} = 1.96TeV$*. Phys. Rev. D 79 (2009) 052008. arXiv: 0812.4458[hep-ex].

[21] D0 Collaboration. *Measurement of the $p\overline{p} \to W+b+X$ production cross section at $\sqrt{s} = 1.96TeV$*. Phys. Lett. B 718 (2013) 1314. arXiv: 1210.0627[hep-ex].

[22] ATLAS Collaboration. *Measurement of differential production cross-sections for a Z boson in association with b-jets in 7 TeV proton–proton collisions with the ATLAS detector*. JHEP 10 (2014) 141. arXiv: 1407.3643[hep-ex].

[23] CMS Collaboration. *Measurement of the $Z/\gamma * +bjet$ cross section in pp collisions at $\sqrt{s} = 7TeV$*. JHEP 06 (2012) 126. arXiv:1204.1643[hep-ex].

[24] V. Khachatryan, A. M. Sirunyan, A. Tumasyan et al. (CMS Collaboration). *Measurements of the associated production of a Z boson and b jets in pp collisions at $\sqrt{s} = 8\ TeV$*. Eur. Phys. J. C 77, 11 (2017) pp.751, https://doi.org/10.1140/epjc/s10052-017-5140-y.

[25] A. Tumasyan et al. (CMS Collaboration). *Measurement of the production cross section for Z+b-jet in proton-proton collisions at sqrts = $13TeV$*. Phys. Rev. D 105 (2022) 092014. https://doi.org/10.1103/PhysRevD.105.092014.

[26] ATLAS Collaboration. *Measurements of the production cross-section for a Z boson in association with b- or c-jets in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. CERN-EP-2024-081. arXiv:2403.15093v1.

[27] ATLAS Collaboration. *Cross-section measurements for the production of a boson in association with high-transverse-momentum jets in pp collisions at $\sqrt{s} = 13\ TeV$ with the ATLAS detector*. arXiv: 2205.02597 [hep-ex].

[28] ATLAS Collaboration. *Measurements of the production cross-section for a Z boson in association with b-jets in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. CERN-EP-2020-022. arXiv:2003.11960v2

[29] CERN. https://home.cern/science/accelerators/

[30] CERN. *LHC, the guide.2017*. https://cds.cern.ch/record/2255762/files/CERN-Brochure-2017-002-Eng.pdf

[31] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*. 2008 JINST 3 S08003, ISSN:1748-0221

[32] CERN. *LHC Machine*. 2008 JINST 3 S08003, ISSN:1748-0221

[33] ATLAS Collaboration. *The ATLAS Inner Detector commissioning and calibration*. Eur. Phys. J. C 70 (2010) 787, issn: 1434-6052.

[34] ATLAS Collaboration. *ATLAS Liquid-Argon Calorimeter: Technical Design Report*. CERN,1996. URL: http://cds.cern.ch/record/331062

[35] ATLAS Collaboration. *ATLAS Tile Calorimeter: Technical Design Report*. CERN,1996. URL: http://cds.cern.ch/record/331062.

[36] ATLAS Collaboration. *ATLAS Muon Spectrometer: Technical Design Report*. CERN,1997. https://cds.cern.ch/record/331068.

[37] ATLAS Collaboration. *Magnet system and magnetic field*. 2008 JINST 3 S08003.

[38] ATLAS Collaboration. *Operation of the ATLAS trigger system in Run 2*. CERN,2020. JINST 15 (2020) P10004

[39] ATLAS Collaboration. *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at s = 13 TeV*. CERN-EP-2018-273. arXiv:1902.04655

[40] ATLAS Collaboration. *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at s=13 TeV*. CERN-EP-2016-033. arxiv:1603.05598.

[41] M.Cacciari, G.P.Salam, G.Soyez. *The anti-kt jet clustering algorithm*. LPTHE-07-03 arxiv:0802.1189.

[42] ATLAS Collaboration. *New techniques for jet calibration with the ATLAS detector*. CERN-EP-2023-028. arXiv:2303.17312.

[43] ATLAS Collaboration. *ATLAS flavour-tagging algorithms for the LHC Run 2 collision dataset*. CERN-EP-2022-226. arxiv:2211.16345.

[44] ATLAS Collaboration. *Calibration of the light-flavour jet mistagging efficiency of the -tagging algorithms with +jets events using 139 fb^{-1} of ATLAS proton–proton collision data at $\sqrt{s}$= 13 TeV.* CERN-EP-2022-211. arxiv:2301.06319.

[45] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model.* ATL-PHYS-PUB-2016-017, 2016. url: https://cds.cern.ch/record/2206965

[46] R. D. Ball et al. *Parton distributions with LHC data.* Nucl. Phys. B 867 (2013) 244. arXiv: 1207.1303 [hep-ph].

[47] ATLAS Collaboration. *Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment.* JHEP 08 (2022)089. arXiv: 2112.09588 [hep-ex].

[48] R. Frederix and S. Frixione. *Merging meets matching in MC@NLO.* JHEP 12 (2012) 061. arXiv: 1209.6215 [hep-ph].

[49] NNPDF Collaboration, R.D. Ball et al. *Parton distributions for the LHC Run II.* JHEP 04 (2015) 040. arXiv: 1410.8849 [hep-ph].

[50] S. Höche, F. Krauss, M. Schönherr and F. Siegert. *A critical appraisal of NLO+PS matching methods.* JHEP 09 (2012) 049. arXiv: 1111.1220 [hep-ph].

[51] S. Höche, F. Krauss, M. Schönherr and F. Siegert. *QCD matrix elements + parton showers. The NLO case.* JHEP 04 (2013) 027. arXiv: 1207.5030 [hep-ph].

[52] S. Catani, F. Krauss, B. R. Webber and R. Kuhn. *QCD Matrix Elements + Parton Showers.* JHEP 05 (2009) 053. arXiv: 0903.1219 [hep-ph].

[53] M. Cacciari, G. P. Salam and G. Soyez. *FastJet User Manual.* Eur. Phys. J. C 72 (2012) 1896. arXiv: 1111.6097 [hep-ph].

[54] S. Frixione, G. Ridolfi and P. Nason. *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadron production.* JHEP 09 (2007) 126. arXiv: 0707.3088 [hep-ph].

[55] S. Frixione, P. Nason and C. Oleari. *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method.* JHEP 11 (2007) 070. arXiv: 0709.2092 [hep-ph].

[56] S. Alioli, P. Nason, C. Oleari and E. Re. *A general framework for implementing NLO calculations inshower Monte Carlo programs: the POWHEG BOX.* JHEP 06 (2010) 043. arXiv: 1002.2581 [hep-ph].

[57] T. Sjöstrand et al. *An introduction to PYTHIA 8.2.* Comput. Phys. Commun. 191 (2015) 159. arXiv: 1410.3012 [hep-ph].

[58] ROOT Unfolding Framework url: hep-unx.rl.ac.uk/˜adye/software/unfold/RooUnfold.html

[59] ATLAS Collaboration. url:twiki.cern.ch