CERN-THESIS-2015-165

UNIVERSITÀ DEGLI STUDI DI TRIESTE

DIPARTIMENTO DI FISICA

Corso di Laurea Magistrale in Fisica
Curriculum in Microfisica e Struttura della Materia

Study of the associated production of a Z boson and b jets with the CMS experiment

Studio della produzione associata di un bosone Z e getti adronici da quark b nell'esperimento CMS

Dissertazione di Laurea Magistrale

Relatore:
Prof. Giuseppe Della Ricca
Correlatore:
Dott. Vieri Candelise

Candidato:
Francesco Peri
Matricola: 2300185

Anno Accademico 2014/2015

Sommario

Questo lavoro di tesi è focalizzato sullo studio della produzione associata di un bosone vettore Z e di getti adronici da quark bottom. Questa è una misura importante sia in termini di validazione del Modello Standard sia per quanto riguarda la ricerca di nuova fisica. Lo studio dello stato finale $Z + b$ permette, infatti, di valutare le predizioni della Cromodinamica Quantistica e di testare la parametrizzazione delle funzioni di distribuzione partoniche del protone. Inoltre, il processo è un fondo importante per svariati studi che coinvolgono il bosone di Higgs e per alcuni processi fisici previsti da modelli "Beyond the Standard Model", come il decadimento di quark pesanti di quarta generazione. In questo lavoro viene analizzata la produzione di un bosone Z in associazione con almeno uno o due getti da quark b, confrontando i dati sperimentali e le predizioni teoriche fornite da alcuni generatori Monte Carlo.

La prima parte della tesi è volta a presentare la misura e le tecniche utilizzate nella successiva analisi. Viene fornita una panoramica del Modello Standard, descrivendo i raggiungimenti della teoria e i questi ancora aperti. Sono descritti i generatori Monte Carlo utilizzati nella tesi e i processi fisici legati alla produzione $Z+b$ nel Modello Standard e in alcuni modelli di nuova fisica. Si presentano quindi gli apparati sperimentali utilizzati per la raccolta dati.

Nella seconda parte viene descritta l'analisi utilizzata per la selezione degli eventi e vengono presentati una serie di studi comparativi volti a testare la robustezza dell'analisi stessa. Si prosegue descrivendo quali sono le incertezze più importanti legate ai Monte Carlo utilizzati, stimando le incertezze teoriche derivanti dalla scelta del set di funzioni di densità partoniche e dei fattori di scala e di costante di accoppiamento. In conclusione si presenta il confronto tra i dati sperimentali e le distribuzioni teoriche, valutando quali sono le possibilità di miglioramento e i possibili sviluppi futuri.

Contents

Introduction

Over the past three decades, the Standard Model has proved being a trustworthy and successful theory. It has been able to justify a large number of experimental processes and many predictions of the theory have proved being correct. In recent times, the discovery of the Higgs boson has definitely demonstrated the reliability of the theory and the potential of the Large Hadron Collider and the associated experiments at CERN. Nevertheless, the room for improvement is still significant and the open questions are many. The Standard Model, in fact, does not include the gravitational interaction, it does not explain the asymmetry between matter and antimatter nor it provides explanation to dark matter evidences. The aim of CMS and the other LHC experiments is precisely to find explanation to these questions, as well as to improve the current understanding of the model. The study of the associated production of a Z boson and b-jets is an important measurement both in terms of the Standard Model validation and for the search of new particles and phenomena. It allows to test the predictions of the perturbative Quantum-Chromodynamics and to improve the parametrization of the Parton Distribution Functions, assessing, in particular, the b-content of the proton. Moreover this process is an important background for several Higgs related studies and also in some new physics scenarios. In this thesis the production of a Z boson, together with at least one or two b-jets, is analyzed, comparing the experimental data with the theoretical predictions provided by the NLO calculation of some Monte Carlo generators. In particular the theoretical uncertainties related to the Monte Carlo generators used for the $Z + b$ studies are evaluated, estimating the contribution of the renormalization and factorization scales and of the PDF variations to the total error. The experimental data consist in proton-proton collisions, at a center of mass energy of 8 TeV, collected by the CMS experiment during the 2012. The thesis is divided into six Chapters. In the first Chapter, the Standard Model is presented, providing a short description of the Quantum-Chromodynamics and the Electroweak theory, together with a brief summary of the main issues of the model itself. In the second Chapter the Monte Carlo event generation is reviewed, dwelling on the Monte Carlo generators used in the analysis. The phenomenology of the $Z + b$ final state is described in the third Chapter, while the LHC and the CMS experiment are shortly presented in the fourth Chapter. In the fifth Chapter there is an overview of the strategy used

to reconstruct the physical objects and to perform the event selection at particle level. Also a series of comparative studies, performed to test the robustness of the event selection, is reported. In the last Chapter the theoretical uncertainties related to the Monte Carlo generators are presented, together with the collection of techniques used to evaluate them. At the end, the comparison between the experimental data and the theoretical predictions is presented: the $Z+b$ differential cross sections, in function of several kinematic observables, are investigated and some possible improvements to the measurement are proposed.

Chapter 1

The Standard Model

The Standard Model (SM) [1, 2, 3] of particle physics is the quantum field theory which describes the fundamental particles and their interactions. The model has been built thanks to the work of many scientists around the world since the second half of the twentieth century. Currently the theory incorporates three of the four fundamental interactions, namely the strong nuclear force, the weak nuclear force and the electromagnetic interaction. The first is described by the Quantum-Chromodynamics (QCD) while the last two are unified into the Electroweak (EW) Model. Mathematically, the Standard Model is a non-Abelian gauge theory, based on special relativity and renormalizable. The fundamental interactions are described by the unitary group $SU(3) \times SU(2) \times U(1)$; it consists of the product of $SU(3)$,which represents the strong interactions, with $SU(2)_L \times U(1)_Y$ which describes the electromagnetic and weak interactions. The SM has gained immense popularity due to the fact that it is able to describe a large number of experimental processes, but it was also able to correctly provide experimental predictions. For example:

- the tau neutrino, theorized as the third generation neutrino during the 1970s, and which was discovered in 2000 at DONUT [4];
- the top quark, predicted in 1973 and discovered at the Tevatron accelerator in 1995 [5, 6];
- the W, Z bosons, discovered by the UA1 and UA2 Collaborations in 1983 [7, 8];
- the Higgs boson, predicted by Brout, Englert and Higgs in 1964 and then discovered at LHC in 2012 [9, 10].

The SM still leaves open questions. It fails to provide an explanation to some cosmological evidence as the presence of dark matter and the asymmetry between baryons and anti-baryons in the Universe. Also it does not incorporate general relativity and the gravitational theory nor it can explain the neutrinos masses.

1.1 Particle Content

The SM provides for the existence of various elementary particles, represented in Figure 1.1, which can be divided into two broad categories: fermions (particles of spin 1/2, obeying the Pauli exclusion principle) and bosons (integer spin).

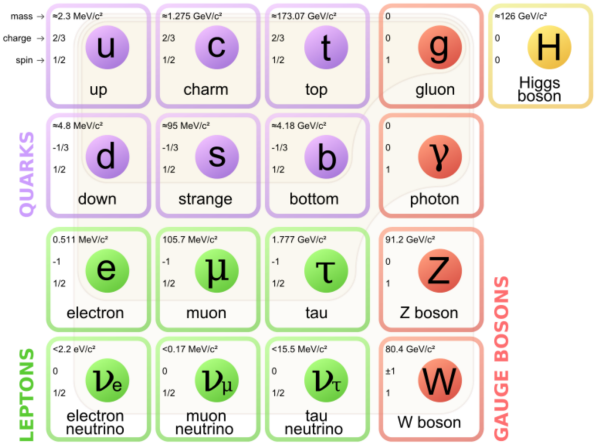

Figure 1.1: The Standard Model elementary particles; matter particles in the first three columns, gauge bosons in the fourth, and the Higgs boson in the fifth. The antiparticles are neglected. In the top left corner of each particle the mass, the electric charge and the spin are reported. Three fermion generations are defined, corresponding to the first three columns [91].

Fermions are 12 particles (+ 12 antiparticles) which are the matter constituents. There are 6 quarks (up, down, strange, charm, top, bottom) and 6 leptons (electron, muon, tau, electron neutrino, muon neutrino, tau neutrino). Quarks have color charge and they interact via all three of the fundamental interactions included in the Standard Model. They only occur in pairs (mesons) or triplets (baryons) necessarily color-less. The other six particles are part of the group of leptons. They do not have color charge, so they do not interact via the strong force. The neutrinos have no electric charge either, then they only interact through the weak force.

Bosons are the force carriers of fundamental interactions. The Standard Model describes in fact the interaction between fermions through the exchange of other particles, bosons precisely. Bosons are divided into photon, which mediates the electromagnetic interaction, W and Z bosons, carriers of the weak force, and 8 gluons, carriers of the strong interaction. Gluons and photons are massless particles. In addition to these, which are bosons with spin 1, there is the Higgs boson. This is an uncharged massive scalar particle created as an excitation of the Higgs field. This field has a crucial importance in the theory of particle physics

as it allows to understand why some particles, which according to the interaction symmetries should be massless, are instead massive; its presence is justified in the next Section.

1.2 Electroweak interaction

Although at low energies the weak and electromagnetic forces seem quite distinct and have different coupling constants, above the 100 GeV energy threshold they merge together. The electroweak (EW) theory is the unified theoretical description of the weak and electromagnetic forces. As already mentioned the EW interaction is mathematically described by the $SU(2)_L \times U(1)_Y$ gauge group. $SU(2)_L$ is the not abelian weak isospin group which has three generators $T_{1,2,3} = \frac{\tau_{1,2,3}}{2}$ (the τ_i are the Pauli matrices). The subscript "L" in $SU(2)_L$ encodes the fact that the weak charge current interaction acts only on left-handed particles (or right-handed antiparticles), explicitly violating the parity invariance. $U(1)_Y$ is instead the abelian hyper-charge group, which has a single generator, Y/2. The electromagnetic group, generally indicated with $U(1)_{em}$, is a subgroup of $SU(2)_L \times U(1)_Y$. Its generator, the electric charge, is a combination of the hyper-charge and the isospin third component:

$$Q = T_3 + \frac{Y}{2}. \tag{1.1}$$

The Electroweak Lagrangian density is described by the following equation:

$$L = -\frac{1}{4} W_{\mu\nu} W^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \overline{\psi} i \gamma_\mu D^\mu \psi. \tag{1.2}$$

The first two terms are the kinematic terms of the isospin and hyper-charge group associated fields. Defining $\vec{W}_\mu$ and B_μ the $SU(2)_L$ and $U(1)_Y$ fields respectively, the tensors used in the Lagrangian are defined as follows:

$$W^\alpha_{\mu\nu} = \partial_\mu W^\alpha_\nu - \partial_\nu W^\alpha_\mu - g \epsilon^{\alpha\beta\gamma} W^\beta_\mu W^\gamma_\nu, \tag{1.3}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{1.4}$$

It is important to emphasize the fact that the $\vec{W}_\mu$ and B_μ fields so far described are not the fields associated with the vector bosons predicted by the SM, which, as we shall see shortly, are obtained as a linear combination of the gauge field components.

The last term of the Lagrangian is the interaction term between fermions and bosons. This interaction manifests through the gauge covariant derivative D_μ. The γ_μ are the Dirac matrices and ψ is the Dirac spinor associated to leptons and quarks.

$$D_\mu \psi = \left[\partial_\mu - igW_\mu \frac{\tau}{2} - i\frac{g'}{2} B_\mu \right] \psi, \tag{1.5}$$

$$\psi = \begin{pmatrix} \nu_l \\ l \end{pmatrix}_{l=e,\mu,\tau}, \begin{pmatrix} u_i \\ d_i \end{pmatrix}^{u_i=u,c,t}_{d'_i=\sum_j V_{i,j} d_j; d_j=d,s,b}. \tag{1.6}$$

In the covariant derivative formulation, g and g' are the $SU(2)_L$ and $U(1)_Y$ gauge coupling constants. $V_{i,j}$ is the Cabibbo-Kobayashi-Maskawa mixing matrix [12] which connects the weak quark eigenstates d'_i with the mass eingenstates d_j (see Section 1.3).

The theory described so far involves massless particles. The gauge symmetry, in fact, forbids to write mass (bilinear) terms for bosons and fermions. This obviously contradicts the experimental observations, which confirm the fact that many particles are irrevocably massive. The solution to the problem is provided by the mechanism of spontaneous symmetry breaking of $SU(2)_L$x$U(1)_Y$, proposed by Higgs, Englert and Brout [13, 14], which consists in the introduction of an additional scalar field in the Electroweak Lagrangian. It is necessary to provide mass for the W^+, W^-, Z gauge bosons, but the photon must stay massless. In this model a complex scalar doublet $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ is added to the Lagrangian through a scalar field invariant term:

$$L_s = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda^2 (\phi^\dagger \phi)^2. \tag{1.7}$$

For a negative value of μ^2, ϕ develops a vacuum expectation value (vev):

$$< \phi >_0 = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \tag{1.8}$$

where $v = \sqrt{-\mu^2/\lambda}$. The vev must be in the neutral part of the doublet to avoid the breaking of the $U(1)_{em}$ explicit symmetry. After a gauge transformation, the scalar field can be rewritten as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \tag{1.9}$$

where H is the Higgs scalar field. An expansion of the $|D_\mu \phi|^2$ term of the scalar Lagrangian brings to:

$$|D_\mu \phi|^2 = \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8} g^2 (v+H)^2 |W^1_\mu + iW^2_\mu|^2 + \frac{1}{8}(v+H)^2 |gW^3_\mu - ig' B_\mu|^2. \tag{1.10}$$

New $W^\pm_\mu$, Z_μ,A_μ fields can be defined as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}}, A_\mu = \frac{gW_\mu^3 + g'B_\mu}{\sqrt{g^2+g'^2}}. \tag{1.11}$$

With this formulation, quadratic terms appear in the Lagrangian:

$$M_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}M_Z^2 Z_\mu Z^\mu + \frac{1}{2}M_A^2 A_\mu A^\mu, \tag{1.12}$$

where $M_W = \frac{1}{2}vg$, $M_Z = \frac{1}{2}v\sqrt{g^2+g'^2}$, $M_A = 0$. The mixing between A_μ and Z_μ is often expressed in function of the Weinberg's angle $\theta_W = \arctan\frac{g'}{g}$:

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W, \tag{1.13}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W. \tag{1.14}$$

After this process the $U(1)_{em}$ symmetry is still unbroken and massive W, Z bosons have been generated. Adding other invariant terms to the Lagrangian, it is possible to generate both the fermion and the quark masses with the same mechanism. A more detailed discussion of the Higgs mechanism can be found in [15].

1.3 The Cabibbo Kobaiashi Maskawa matrix

The Cabibbo Kobaiashi Maskawa matrix (CKM) is a unitary matrix which allows to describe in a simple way the flavor changing phenomena which involve weak interactions between quarks. It is a 3×3 matrix:

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

,

where every element $V_{qq'}$ corresponds to a transition amplitude between a quark q to a quark q'. The quark mixing described by the CKM matrix is theoretically justified by the fact that the mass eingestates for left handed quark doublets do not coincide with the weak ones which in fact generate as a linear combination of the mass eingestates. The mixing can be parametrized by using at least three real angles and one CP-violating complex phase. In this terms the explicit expression of the matrix is:

$$\begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & -c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}$$

,

where $s_{ij} = \sin\theta_{ij}$, $c_{ij} = \cos\theta_{ij}$ and δ is the CP violating phase. Also the Wolfenstein parametrization for the CKM matrix is often used [16]:

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{us}|^2 + |V_{ud}|^2}},\ s_{23} = A\lambda^2 = \frac{|V_{cb}|}{|V_{us}|},\ s_{13}e^{i\delta} = A\lambda^3(\rho + i\eta) = V_{ub}^*. \tag{1.15}$$

The CKM formalism incorporated the Cabibbo theory [17] which guaranteed the weak interaction universality before the discovery of heavy quarks.

Due to the unitarity of the matrix, the following relations are valid: $\sum_i V_{ij}V_{ik}^* = \delta_{jk}$ and $\sum_j V_{ij}V_{kj}^* = \delta_{jk}$. Since there are six different quark flavors, there are six possible combinations. These can be represented as triangles in the complex plane. The triangles have the same area, $\frac{1}{2}J$, where J is the Jarlskog invariant [18], a measurable variable related to the CP violation.

1.4 Quantum Chromodynamics

The QCD is the non-abelian gauge field theory which describes the strong interaction. Mathematically it is based on the $SU(3)$ component of the $SU(3) \times SU(2) \times U(1)$ Standard Model symmetry. The QCD Lagrangian is given by:

$$L = -\frac{1}{4}F_{\mu\nu}^A F^{A\mu\nu} + \sum_q \overline{\psi}_{q,a}(i\gamma^\mu\partial_{ab} - g_s\gamma^\mu t_{ab}^C A_\mu^C - m_q\partial_{ab})\psi_{q,b}. \tag{1.16}$$

The field tensor is:

$$F_{\mu\nu}^A = \partial_\mu A_\nu^A - \partial_\nu A_\mu^A - g_s f^{ABC} A_\mu^B A_\nu^C, \tag{1.17}$$

where the f^{ABC} are the structure constant of the group. The first term of the Lagrangian is therefore the kinetic term associated to the gluon field A_μ^C. Gluons are the gauge bosons of the theory. They are color-charged-states which carry a combination of color and anticolor. Being the gluon number equal to eight the C index goes from 1 to 8. Again the γ_μ are the Dirac matrices while g_s is the strong coupling constant (often $\alpha_s = \frac{g_s^2}{4\pi}$ is used instead of g_s). The t_{ab}^C are the eight generator of the symmetry derived from the 3×3 Gell-Mann matrices. The $\psi_{q,b}$ is the spinor of a quark field for a quark q, with a color index a ($a = 1, 2, 3$).

A characteristic of the non-abelian gauge theories, where the gauge bosons carry the interaction charge, is the fact that they can couple with themselves. In particular 3 and 4 gluon vertexes generate from the field strength tensor together with the free field term.

The QCD has two distinctive features. The first is the confinement. This is the phenomenon by which quarks cannot be detected as free particles. As already mentioned, in fact, hadrons are color-less groups of quarks and gluons. This is due to the fact that the force between two quarks does not diminish as the quarks move away from each other. In particular, the color-field potential grows in direct proportion to the distance between the quarks. For this reason, when they reach a certain distance, the creation of quark-antiquark pairs becomes possible. A simple way to imagine the mechanism is through the Lund Model. In this case an imaginary color string connects a pair of quarks. At some point as they move away, it becomes energetically favorable for the string to break in two (see Figure 1.2). In this process, two smaller strings have been created and the two new extremes represent the new pair of quarks.

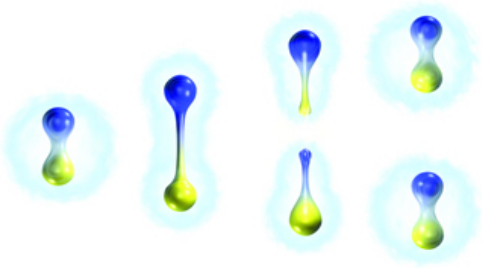

Figure 1.2: A schematic representation of a quark-antiquark production [92]
.

The other important property of QCD is the asymptotic freedom. This phenomenon causes bonds between particles to become weaker and weaker as they come closer and the energy increases. This property is the basis of the fact that, at high energies, the perturbation theory can be applied onto hard scattering of quantum-chromodynamics. This approach is called perturbative QCD (pQCD). In the pQCD framework the predictions are described in terms of the $\alpha_s(\mu_r^2)$ strong coupling constant at a specific energy scale μ_r (the renormalization scale). The strong coupling follows the Renormalization Group Equation:

$$\mu_r^2 \frac{\mathrm{d}\alpha_s}{\mathrm{d}\mu_r^2} = -(b_0\alpha_s^2 + b_1\alpha_s^3 + ...) = \beta(\alpha_s), \tag{1.18}$$

where the b_i are the beta function coefficients. The Asymptotic Freedom originates from the minus sign in equation above. In fact, it involves the decrease of the coupling constant α_s while the transferred momentum in the process becomes

larger (see Figure 1.3). An exact solution of equation 1.18 is possible when only the b_0 coefficient is taken into account:

$$\alpha_s(\mu_r^2) = \frac{1}{b_0 \log(\frac{\mu_r^2}{\Lambda_{QCD}^2})}. \tag{1.19}$$

The term Λ_{QCD} is the renormalization integral's cut-off which defines the phase-space beyond which the integrals would not converge. It corresponds to the energy scale at which non-perturbative effects become important. Its value is experimentally determined to be $\Lambda_{QCD} \simeq 200$ MeV.

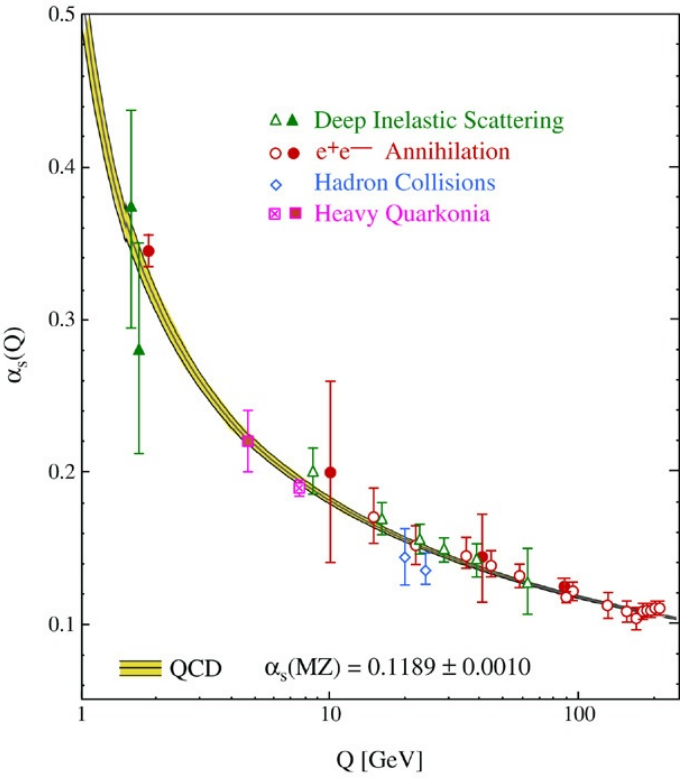

Figure 1.3: The asymptotic behavior of the strong coupling constant [93].

1.5 Cross section and Feynman diagrams

Even if the topic is not strictly related to the Standard Model description, the concepts of cross section and Feynman diagram are of particular importance in particle physics and they are therefore reported here.

The cross section is an effective area which measures the quantum mechanical probability for a interaction to occur. The conventional unit is the barn b, where $\mathrm{b} = 10^{-28}\,\mathrm{m}^2$, but smaller submultiples are also used (pb, fb...). Often it is the differential cross section of a process the result to be obtained from a scattering experiment. This is a description of the variation of the cross section as a function of a specific kinematic variable. In a simple case, where only two initial state particles are involved in a scattering, the cross section of the process will be:

$$\mathrm{d}\sigma = \frac{1}{2s\beta_i}|M_{f,i}|^2\mathrm{d}\phi_n \tag{1.20}$$

where $s = (p_1 + p_2)^2$ is the center-of-mass energy, $\mathrm{d}\phi_n$ is the Lorentz invariant phase space of the final state and β_i is:

$$\beta_i = \sqrt{1 - 2(x_1 + x_2) + (x_1 - x_2)^2} \tag{1.21}$$

with $x_1 = p_1^2/s$, $x_2 = p_2^2/s$. $M_{f,i}$ is the scattering matrix element which describes the evolution of the initial states into the final states. If H' is the perturbative part of the system Hamiltonian, M is defined as:

$$M_{f,i} = < \psi_f|H'|\psi_i > \tag{1.22}$$

where $\psi_{i,f}$ are the initial (final) states. The matrix element for a specific scattering event can be conveniently calculated by the use of so-called Feynman diagrams. The Feynman diagrams are graphical representation of rules derived from the Standard Model scattering theory. An example of a Feynman diagram is represented in Figure 1.4.

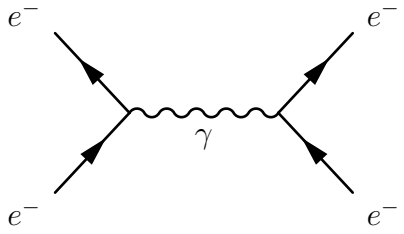

Figure 1.4: Feynman diagram showing electromagnetic interaction between a pair of electrons.

External lines correspond to incoming/outgoing particles, inner lines represent quantum field theory propagators while the points where three or more particles

meet is a vertex where the interactions occur. The style used for each line depends on the kind of particle that is represented: a bosonic propagator is usually represented with wiggly lines, while fermions are drawn using solid lines. A Feynman diagram is also used to represent the space-time evolution of the process: in general one axis is used to show the space coordinate of the system (top-bottom), while the other corresponds to the time coordinate (left-right). It is possible that closed loops of virtual particles appear in the diagrams. Integrals involving loop diagrams can diverge therefore a renormalization procedure of the theory is needed to restore the integral finiteness. A complete description of Feynman rules can be found at [21].

When dealing with scattering processes occurring in an accelerator, the cross section σ is often connected to the accelerator luminosity L:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \sigma L. \tag{1.23}$$

where L is the time-dependent luminosity. In a collider where n bunches with a transverse-area A, constituted by N particles, collide with a frequency f, the luminosity can be calculated as:

$$L = \frac{nN^2 f}{A}. \tag{1.24}$$

Also the time-integrated luminosity L_{int} is often used:

$$L_{int} = \int L \mathrm{d}t. \tag{1.25}$$

1.6 The Drell-Yan process

In this Section there is a brief description of the Drell-Yan process, which is crucial for this analysis. The Drell-Yan process consists in the production of lepton-antilepton pairs from the electroweak decay of γ^* or Z bosons created in a hadron-hadron collision, where a quark and an antiquark annihilate each other. A simple representation of the process, mediated by a virtual photon, is depicted in Figure 1.5.

The invariant mass spectrum of the leptons is a falling distribution which, for small values of the dilepton mass, is dominated by uncorrelated leptonic decays. As the mass becomes larger, the background decreases and the Z production becomes the dominant process. The result is that a resonance peak manifests around the Z-mass value.

The Drell-Yan cross section measurement allows to investigate the internal structure of hadrons and mesons. In particular it is a direct probe of the antiquark content of hadrons: in proton-proton collisions, in fact, the initial-state quark can

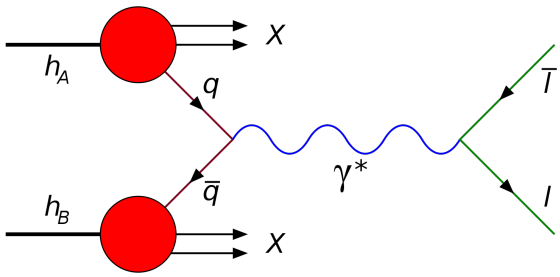

Figure 1.5: Drell-Yan process [94].

be a sea or valence quark, while the antiquark must necessarily come from the sea. This process is also an important background in many scenarios of BSM physics.

1.7 Beyond the Standard Model

Despite the great success of the Standard Model, it can not be the ultimate theory of the physics of fundamental interactions as it still has significant open questions. In this Section the main issues of the model are summarized, focusing on the experimental observations that can not be explained within the current SM formulation.

- **The Gravitational interaction.** The SM includes only three of the 4 fundamental forces. The gravity is neglected by the actual formulation of the model. Different hypothesis have been made to formulate an inclusive theory. One of them consists in thinking at the gravity as to an exchange of bosons, similarly to the other interactions. The hypothetical boson associated to the gravitational interaction is the graviton, a massless spin-2 particle.

- **The Hierarchy Problem.** The problem is basically linked to the significant discrepancy between the coupling constants of the gravitational and the electroweak forces (the gravity is approximately 10^{32} times weaker than the weak force). The same question can be formulated taking into account the particle masses. In this case the problem is why the mass of the Higgs boson is so small compared to the Planck mass. To understand the problem from the mathematical point of view, consider the first order radiative

corrections to the tree-level propagator of the Higgs boson:

$$\partial M_H^2 \simeq -\frac{g_f^2}{8\pi^2}\Lambda^2 + ..., \tag{1.26}$$

where Λ is the renormalization cut-off and g_f is the coupling with a fermion f which creates the single loop correction (see Figure 1.6). If the cut-off is chosen at the Plank energy scale, the term above is about 32 order of magnitude larger than the tree-level mass, therefore the Higgs mass should be much larger than 125 GeV. To correct the mass and solve the hierarchy problem, there are many solutions which are currently discussed. For example, in the supersymmetric extension of the SM, a bosonic superpartner gives a positive contribution to the mass correction, canceling the divergent term. During the first run of the LHC data taking no evidence of supersymmetric particles has been found. Other theories involve multiple dimensions or assume that the Higgs is a composite particle.

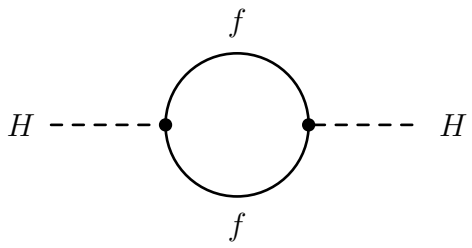

Figure 1.6: Feynman diagram showing the one loop correction to the Higgs boson propagator.

- **Neutrino masses.** In the Standard Model, the Higgs boson does not couple with neutrinos and they are treated as massless particles. Nevertheless Super-Kamiokande [22], Opera [23] and other experiments have definitely proven that they are massive since they can change their flavor with an oscillation frequency that is proportional to the mass square difference of the neutrino states.

- **Dark Matter.** The modern astrophysics does not explain the fact that our Universe contains only a 5% of ordinary matter, while the 27% is dark matter and the rest is dark energy. The presence of dark matter has been inferred from the gravitational dynamics of the galaxies and in particular from the rotational velocity of their halo, which is not compatible with the hypothesis that they are composed only by ordinary matter. At the moment no experiment has been able to detect an eventual dark matter component of the cosmic rays, nor it was possible to produce dark matter particles using particle accelerators.

- **Matter-Antimatter asymmetry.** The Big-Bang models predict that our Universe was created with an equal amount of matter and antimatter; nevertheless nowadays it is mainly composed of matter particles. This asymmetry can not be explained by the actual formulation of the Standard Model and it becomes experimentally evident when studying the CP violation of the weak interaction, as the decay of heavy-flavored B-mesons.

Chapter 2

Monte Carlo Event Generators

This Chapter is focused on the Monte Carlo (MC) event generation. MC generators are software libraries which can simulate high energy interactions between particles generating random events to reproduce the collisions which occur in particle accelerators. The theoretical basis and the fundamental steps of a generation algorithm are here presented. There is a comparison between various types of event generators, with special attention for the main features of the Monte Carlo generators which are relevant for this thesis.

2.1 Theoretical basis

In this Section there is an introduction to the theoretical concepts used in the following Chapters, with particular attention to the phenomenology of the hadronic interactions and to the concept of fixed order calculation.

2.1.1 Hadronic collisions

Considering, as an example, the production of two generic particles due to the hadronic interaction $h + h \to a + b$, the cross section can be expressed as:

$$\sigma = \sum_n \alpha_s^n(\mu_r^2) \sum_{i,j} \int \mathrm{d}x_1 \mathrm{d}x_2 f_{i/h_1}(x_1, \mu_F^2) f_{j/h_2}(x_2, \mu_F^2) \sigma^n_{i,j\to a+b}(x_1 x_2 s, \mu_r^2, \mu_F^2) + O\left(\frac{\Lambda^2}{Q^2}\right), \tag{2.1}$$

where s is the center-of-mass energy. In the equation 2.1 a convolution between the hard scattering cross section $\sigma^n_{i,j\to a+b}$ and the Parton Distribution Functions $f_{i/h}$ (PDFs) is performed. This is possible thanks to the Factorization Theorems [33] which state that the total cross section can be splitted in two parts: a

short-distance perturbative parton cross-section and a long-distance interaction encoded into the PDFs. These functions describe in a non-perturbative way the internal structure of hadrons and they allow to treat the interaction even when the coupling constant α_s becomes large. The parton distribution function $f_{i/h}(x, \mu_F^2)$ represents the probability density to find a parton i carrying a momentum fraction x at a specific energy scale, inside the hadron h. In high-energy interactions, such as those that occur in LHC, an accurate knowledge of the PDFs is necessary. However, the non-perturbative nature of the PDFs makes necessary their extrapolation from data and most of the time they are not measured at the energy scale of interest. Their evolution at a specific energy scale can be performed using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation [24, 25, 26]:

$$\mu_F^2 \frac{\partial f_i(x, \mu_F^2)}{\partial \mu_F^2} = \sum_j \frac{\alpha_s(\mu_F^2)}{2\pi} \int_x^1 \frac{\mathrm{d}z}{z} P_{i \leftarrow j}(z) f_j(\frac{x}{z}, \mu_F^2), \tag{2.2}$$

where $P_{i \to j}$ is *the splitting kernel* which describes the probability for a given parton to split into others.

In Figure 2.1, the parton distribution functions measured by the HERA Collaboration [27] are presented. Other sets of PDFs are available as the ones provided by the CTEQ [28] and the NNPDF [29] collaborations.

2.1.2 Collinear factorization

One of the most important consequences of the Factorization Theorems is that a matrix elements with multiple "legs", i.e. many external lines, can be computed as the product of a lower order matrix element and an additional branching term. Let's consider a process involving a number n of emitted partons. When the angle between two of these particles is small ($\theta \to 0$) the matrix element can be approximated as follow:

$$\lim_{\theta_{ij} \to 0} \mathrm{d}\phi_n |M_n^2(p_1, ..., p_n)| = \mathrm{d}\phi_{n-1} |M_{n-1}^2(p_1, ..., p_{n-1})| \frac{\alpha_s C_i}{\pi} \frac{d\theta_{ij}^2}{\theta_{ij}^2}, \tag{2.3}$$

where p_i are the partons momenta and $\mathrm{d}\phi_n$ corresponds to the phase space where the integration is performed. This is called the collinear limit of the parton production. In this approximation the formula has non-integrable divergences due to the angular term. For this reason the event may be dominated by many low-angle emissions. This factorization can be generalized using a generic variable t as the branching "hardness" instead of the emission angle θ^2. For example the transverse momentum p_T^2 is often used as the evolution scale t. By ordering the emissions using the t variable, it is possible to further reduce the number of legs of the matrix element, adding other branching terms.

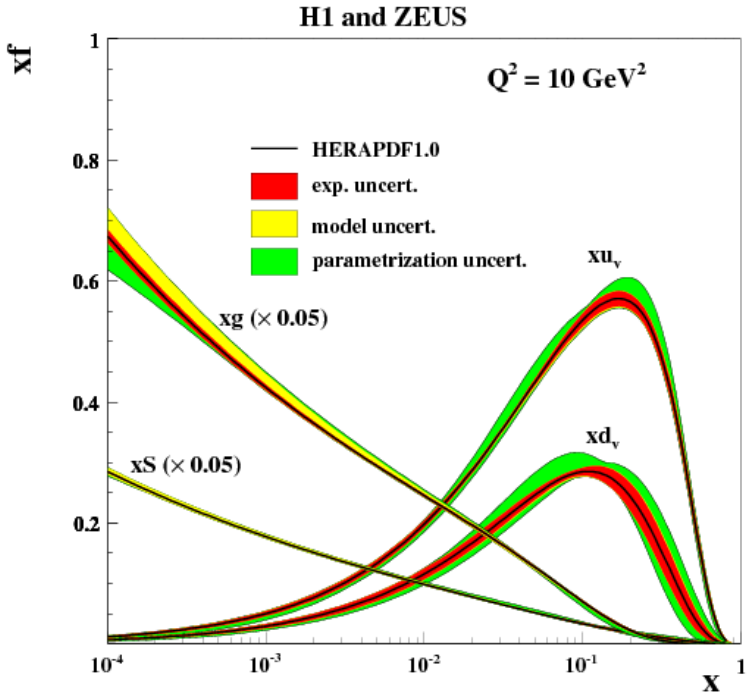

Figure 2.1: The parton distribution functions from HERAPDF1.0 at $Q^2 \simeq 10$ GeV, displayed as a function of the momentum fraction x [27].

An important quantity to define is the so-called Sudakov Form Factor (SFF):

$$\Delta(Q^2, t) = \exp\left[-\int_t^{Q^2} \mathrm{d}P_{a\to bc}(t')\right], \tag{2.4}$$

where $\mathrm{d}P_{a\to bc}(t)$ gives the probability for a parton "a" to branch into "bc" between the t and $t+\mathrm{d}t$ evolution scale values. The SFF represents the probability that a parton doesn't branch between the hard scale Q^2 and the evolution scale t.

2.1.3 Fixed Order QCD calculation

Calculating an all-order cross section, i.e. taking into account all the virtual and real emissions of a specific diagram, is beyond our computation capabilities. For this reason, in general, a limited number of diagrams is included in the computation and a fixed-order pQCD cross section is provided as the final result. Consider the production of a n-particle final-state ($2 \to n$) and the respective cross section, weighted for an observable $O(p_1, ..., p_n)$ which depends on the particle momenta. At the Leading-Order (LO) the cross section is defined as:

$$\sigma_{O,LO} = \int \mathrm{d}\phi_n \alpha_s^{n-2}(\mu_r^2)|M_n^2(p_1, ...p_n)|O(p_1, ..., p_n). \tag{2.5}$$

In this case the matrix element is calculated using the lowest order of perturbation theory and the corresponding Feynman diagram is called tree-level (or Born-level) diagram. The LO result can be corrected adding the Next-to-Leading-Order contributions. In this case the $2 \to n+1$ process is considered together with the $2 \to n$ virtual Feynman diagrams. In the NLO calculation the interference between the M_{LO} matrix element of the Born level diagram and the $M_{Virtual}$ matrix element of the virtual diagram must be taken into account:

$$\sigma_{NLO} \propto |M_{NLO}|^2 = |M_{LO}|^2 + 2Re(M_{LO}M_{Virtual}) + |M_{n+1}|^2. \tag{2.6}$$

The NLO approximation increases the complexity of the calculation. This is mainly due to the fact that loop amplitudes introduce divergent terms in the computation, which cancel out with the tree level divergences (Eq. 2.3) if the observable is infrared and collinear (IC) safe. An observable is IC safe if, after the addition of a collinear splitting or a soft emission, it stays unchanged:

$$\lim_{p_i \to 0} O_{n+1}(p_1, ...p_i, ..., p_n) = O_n(p_1, ..., p_n), \tag{2.7}$$

$$\lim_{p_i // p_j} O_{n+1}(p_1, ...p_i, p_j, ..., p_n) = O_n(p_1, ..., p_i, ..., p_n). \tag{2.8}$$

Examples of safe quantities are, for instance, inclusive cross-sections, while particle multiplicity and first jet transverse momentum are, instead, unsafe quantities.

By adding further orders of α_s, it is possible to compute quantities with an higher level of precision (NNLO etc.). In Figure 2.2, a $2 \to 2$ process and its virtual and real NLO corrections are represented.

A further clarification is appropriate. If some constraints are applied to the phase space, the cancellation of the divergences is less complete and logarithmic terms, which can spoil the convergence, remain in the perturbative series. In this cases, it can be shown that the series can be rearranged in terms of power of $\alpha_s L$ [32], where $L \sim \ln \frac{Q^2}{Q_0^2}$:

$$\sigma = \sigma_{tot} \sum_{n}^{\infty} \sum_{k}^{2n} c_{nk}\, \alpha_s^n L^k. \tag{2.9}$$

The Leading Logarithmic (LL) approximation correspond to the accounting of all the $k = 2n$ terms in equation 2.9, the NLL approximation adds the $k = 2n-1$ terms, etc.

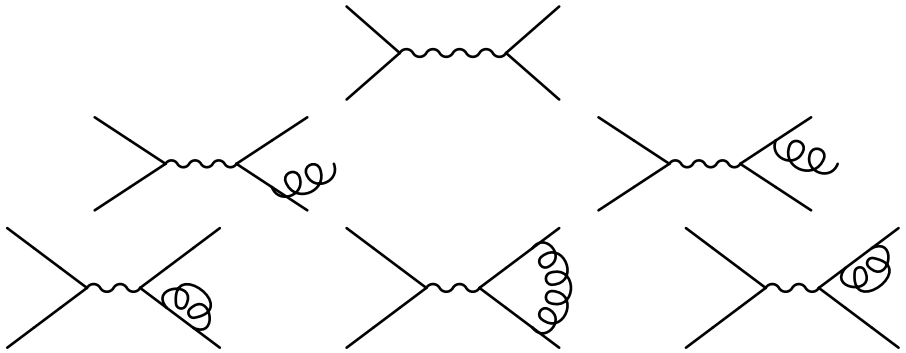

Figure 2.2: Born (top), LO (center) and NLO (bottom) diagrams for a $2 \to 2$ interaction. The initial state NLO corrections are neglected.

2.2 The event generation

The event generator process can be summarized in 5 steps:

- Hard process
- Parton Shower
- Hadronization
- Underlying event
- Unstable particle decay

The most general proton-proton (pp) collision usually concerns only the production of soft particles. The creation of high transverse-momentum particles, which are the most interesting ones, is very rare. For this reason the event generation needs to focus on the desired hard process, without simulating all possible pp collisions. On this basis, the simulation starts with the calculation of the hard scattering process: using the parton distribution functions of the incoming protons, the pQCD gives the distribution of the outgoing particles. The subsequent part of the simulation describes the emission of partons produced by the QCD interactions (the "parton shower") and the rearrangement of partons into hadrons (the "hadronizaton"). It is important to point out the fact that there could be other interactions, in parallel to the hard scattering. This is the "underlying event", which consists of contaminating secondary interactions produced by the proton remnants. The final step of the simulation involves the decay of the unstable particles. The simulation steps presented above are discussed in more detail in the following pages.

The outlined scheme is typical of the so-called General-Purpose Monte Carlo (GPMC) generators. These generators aim to completely simulate events, from the hard process to the particle decays. They are complex programs that need to take into account both for pQCD and for non-perturbative effects. More often Fixed-order Monte Carlos (FOMC) are used together with General-Purpose ones. In this case the hard scattering is simulated by a Fixed-order MC which gives an exact result to a given order in pQCD. The final-state partons are then passed to a General-Purpose MC to complete the event simulation.

When comparing the experimental data with theoretical predictions it is necessary also to add the simulation of the interactions with the detector, so as to completely reproduce the experimental environment. With this step a dependency on the detector is added to the simulated data, thus the events will be only comparable with that specific detector experimental results. For this reason it is often preferred to deconvolve the experimental data by removing detector resolution and acceptance effects. This is known as the "unfolding" procedure [36, 37]. In this way the theoretical predictions remain disjointed from detector effects and they can be compared with data at particle level.

2.2.1 The hard scattering

The first step of the event generation is the simulation of the hard scattering. As already mentioned, the process can be described by the perturbation theory (see Section 2.1.1). At this point all the secondary interactions are neglected. The perturbation order used in the calculation of the matrix element defines the order of the entire event generation. In general, more than one Feynman diagram will contribute to the specific process and they will have to be summed coherently.

2.2.2 The parton shower

The aim of the parton shower is to simulate the cascade of partons that are emitted by the accelerated colored particles involved in the hard scattering. The fundamental problem in this stage is to properly handle the soft and collinear divergences. To resolve the collinear limit, a factorization energy scale μ_F is defined. When the transverse momentum of one parton with respect to another is $k_T > \mu_F$ the parton are said to be "resolvable". If $k_T < \mu_F$ the partons are "unresolvable"; this means that it is impossible to distinguish them from a single particle with the same total momentum and, for the purpose of the simulation, no parton is actually emitted. The non-emission probability is evaluated using the Sudakov Form Factor defined in Section 2.1.2. The SFF, together with the DGLAP splitting functions, allow to construct the probability distribution of any final-state. This solves the collinear limit problem while the soft emission

divergences can be handled by simply ordering the parton emissions in the opening angle.

Care must be taken to the fact that the distribution of partons for both the final-state particles and the initial state ones must be generated. The radiation emitted in the first case is called Finale State Radiation (FSR), while the collection of particles emitted by the incoming partons is the Initial State Radiation (ISR). The ISR computation differs from the FSR one because there are some kinematic constraints, like the x-Bjorken and Q^2 of the incoming partons, that are absent in the FSR case. For this reason the ISR algorithm performs a backward evolution, considering only the parton distributions that lead to the specific values of Q^2 and x. In the end the two algorithms are very similar, but the Sudakov Form Factor of the FSR case must be replaced with $\frac{\Delta_{SFF}(Q^2,t)}{f_i(x,Q^2)}$, in the ISR calculation.

In the case where heavy quarks are involved in the parton shower, the angular distribution of the radiated particles is quite different. In Figure 2.3 the number of emitted gluons by light and heavy quarks is represented as a function of the opening angle.

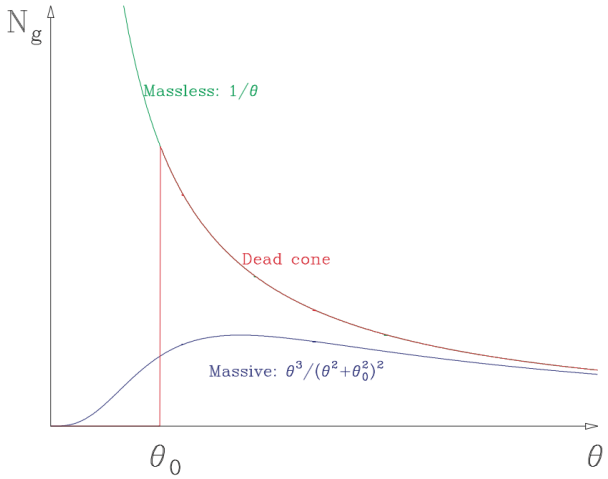

Figure 2.3: Number of emitted gluons as a function of the opening angle. The green line is the massless quark approximation, the blue one is the massive quark model. The red line is the "dead cone" approximation used by some algorithms to simplify the heavy quark case [96].

When quarks are much lighter than the confinement energy scale, radiation emitted by a parton diverges as the angle goes to 0. The number of radiated gluons

by a heavy quark, instead, rises until an angle $\theta_0 = \frac{m_q}{E_q}$ and then it is suppressed. Due to this behavior no heavy quark can radiate in the forward direction.

The parton shower model described above is the most common implementation used by the Monte Carlo generators. However there are also other approaches, such as the Color Dipole Model. In this shower model, $q\overline{q}$ color-anticolor pair are said to form a dipole. This color dipole radiates a gluon which divides the system into two dipoles, qg and $g\overline{q}$, that emit radiation in turn. In this way the cascade consists in $1 \to 2$ dipole splitting, instead that in $1 \to 2$ parton splitting.

2.2.3 Hadronization

Quarks and gluons are colored particles. Due to the already stated color confinement, they cannot exist individually. The formation of hadrons out of partons is called the Hadronization. After the final-state partons have originated by the primary scattering process, they necessarily have to rearrange into color-neutral hadrons. The process occurs at a low energetic scale ($\simeq$ 1 GeV); this corresponds to a large strong coupling which makes hadronization a non perturbative process. For this reason, the parton-to-hadron transition is not reproduced using pQCD calculations, but through the use of empirical models. In particular two hadronization models are vastly used in Monte Carlo generators: the cluster model and the string model [34, 35]. The first model splits all the gluons into quark-antiquark pairs. Color singlets are then chosen to form clusters, which decay into hadrons with the appropriate quantum numbers. If the clusters are light they directly decay into hadrons. In the case of heavy clusters, instead, the decay can be mediated by other lighter clusters before they finally convert into hadrons.

The string model, instead, represents the bond between a quark and an antiquark with a string. The string is defined by means of the strong force potential:

$$V(r) = \frac{4\pi}{3r} + kr, \tag{2.10}$$

where r is the radial distance between partons and k is a constant related to the amount of energy per unit of length which is necessary to keep the system bounded. As the quarks move away it becomes possible for the string to break into hadrons. Gluons create "kinks" on the string, which influence the angular and momentum distribution of the generated hadrons. In Figure 2.4 the different models are represented. The cone of particles created after the hadronization of a parton is a "jet". Jets are mainly composed by light mesons, leptons, neutrinos and photons while barions are a small fraction of the total. The techniques used to identify and cluster the jet particles are described in Section 5.4.3.

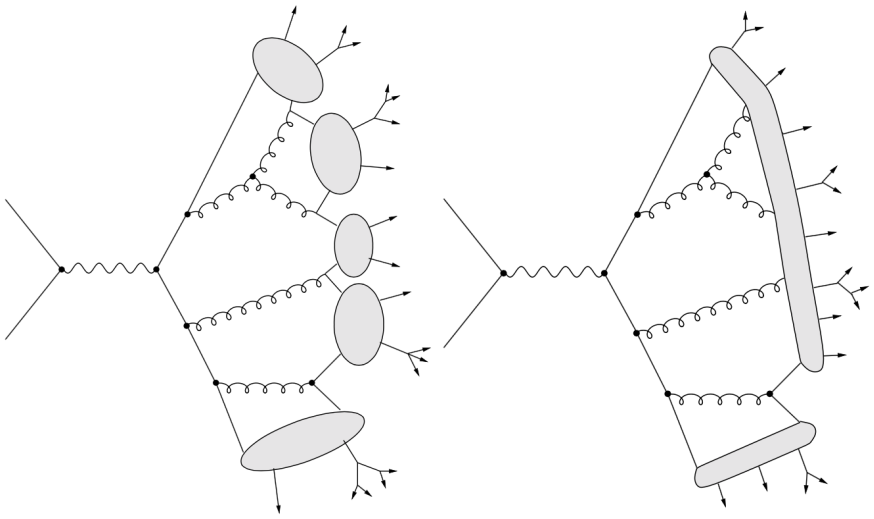

Figure 2.4: Cluster (left) and string (right) hadronization models [97].

2.2.4 The underlying event

After that the hard process initial-state partons are extracted from the protons, the other constituents (proton remnants) can interact with each other. The set of secondary processes arising from the interactions of partons that are not involved in the hard scattering is called Underlying Event. Without going further in the study of this process, it is important to emphasize the fact that the simulation should also take into account hadronization and parton emission subsequent to these interactions.

2.2.5 Particle decays

Hadrons that are created at the end of the hadronization process are often unstable particles. It is therefore necessary to include the decay tables into the simulation. It should also be emphasized that the concept of stable particle is relative. Since we often need to pass the particle collection to the detector simulation, it was decided to process at this level the decay of all the particles with a mean decay length $\gamma\beta c\tau < 10$ cm. More stable particles can reach the detector and interact with it.

2.3 Event generators

As already said, there are two broad categories of generators: General Purpose Monte Carlo (GPMC) generators and Fixed Order Monte Carlo (FOMC) gen-

erators. In the following a brief description of the features of the most used generators is proposed.

2.3.1 General Purpose Monte Carlo generators

The GPMC generators allow to simulate the entire event, from the hard scattering to the hadron decays. They start with the computation of a low-order matrix element, $2 \to 2$ in most cases, followed by the parton shower that approximates the higher order processes, i.e. $2 \to n$ with $n > 2$. The generators mainly differs in which matrix elements are available, in the evolution scale parameter t, in the choice of the hadronization model and the selection of decay tables. One of the most used GPMC is Pythia. Two different releases are currently living together: Pythia6 [38], a Fortran based version, and Pythia8 [39], a new C++ developed version. They both use the Lund string model for the hadronization and they have very similar Underlying Event models. However there are some differences: Pythia8 uses the dipole shower to deal with the parton emissions, while Pythia6 uses the standard parton shower. Furthermore Pythia8 implements up-to-date decay tables and PDFs. Also new Beyond the Standard Model processes are encoded in the new version. Herwig [40] and Sherpa [41] are other examples of GPMC generators.

2.3.2 Fixed Order Monte Carlo generators

In an attempt to create an all-inclusive event generator, the hard scattering process in the GPMCs has been thought in such a way as to involve a small number of particles in the final state. They depend on the parton shower which is designed to simulate soft and collinear emissions, therefore the GPMCs are not able to accurately simulate events that involve large angle emitted particles. In order to improve the accuracy of this kind of simulations new techniques have been developed. The matrix element for a basic $2 \to n$ process is computed by a Fixed Order event generator and then it is matched with a GPMC parton shower to deal with the soft emissions. Depending on the generator, the matrix element can be at Leading Order or NLO and can involve an additional number of real emitted partons. In the ME+PS approach the process starts by generating the Born level matrix elements for the basic scattering plus a certain number of partons. In this way the distribution of the first n jets is computed with tree level accuracy and the remaining ones are simulated by the parton shower. Examples of ME+PS generators are MadGraph [42] and ALPGEN [43]. In the NLO+PS method, instead, the basic process is generated at NLO, including also the virtual corrections. The hardest jet is therefore generated at the tree level, but the others are simulated by the GPMC parton shower. POWHEG [44] is an example of a NLO+PS implementation. Inclusive quantities are better simulated by NLO+PS

generators, while ME+PS give a more accurate description of many large angle emissions. The new frontier is to combine the two methods to be able to generate the first n jets with NLO accuracy. MadGraph5_aMC@NLO [45] is one of the first MC generators with this capability.

In the following some Fixed Order generators are presented.

- **MadGraph5.** MadGraph5 is a Python computer program which can generate LO matrix elements for generic $2 \to n$ processes. It is compatible with *FeynRules* [46] and it has several output formats, among which there are C++ libraries to guarantee the compatibility with Pythia8. The hard scattering can be simulated with tree level precision and must be completed using a subsequent parton shower simulation;

- **MadGraph5_aMC@NLO.** The MadGraph5_aMC@NLO package merges together the MadGraph5 and MC@NLO frameworks. It allows to compute diagram amplitudes with NLO accuracy and it can be matched with a General Purpose MC. In particular, it uses the FxFx merging technique (see Section 2.4.2) to deal with the matrix element overcounting. A specific feature of the MadGraph5_aMC@NLO generator is the possibility to perform an event reweighting. This peculiarity permits to apply corrections or to study systematic uncertainties without having to perform the entire event generation. During the simulation, the generator creates an array of weights for each event: the weights correspond to different settings, for example to different PDF sets or strong coupling values. Therefore selecting a specific index in the array, the corresponding configuration is applied;

- **POWHEG.** The POWHEG generator is capable of computing matrix elements with NLO accuracy. POWHEG stands for "POsitive Weight Hardest Emission Generator". The name underlies its ability to perform the event reweighting, similarly to the Mad Graph5_aMC@NLO approach. The latest version of the POWHEG BOX also includes the MiNLO (Multi-scale Improved NLO) [47] prescription which allows to include, in a high-multiplicity process evaluated at the NLO, the lower-multiplicity calculations. For example a NLO Z+jet process, simulated using the MiNLO approach, has NLO accuracy also for the boson inclusive production observables.

2.4 Matching and merging the matrix element

When the matrix element is generated by a Fixed Order generator and the parton shower is simulated by a GPMC, an overlap of the phase space between the two methods may occur. As an example take the case of a ME+PS generator where different LO exclusive processes are merged together. The additional real

emissions described by the matrix elements can be created also by the parton shower simulation. In Figure 2.5 a schematic diagram of the double counting between the two methods is represented.

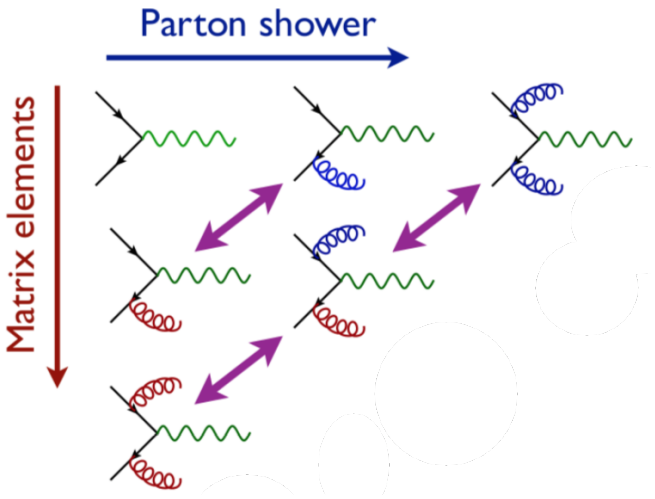

Figure 2.5: A representation of the diagrams corresponding to real emission processes which can be described both by the matrix element and by the parton shower. On the diagonal there are the over-counted diagrams. The matrix element diagram at the n+1 order is equal to the diagram at the n order where an additional leg is added by the parton shower [98].

Various merging techniques have been developed to solve this kind of problems. MLM [48] and FxFx [50] are two examples of algorithms that deal with LO and NLO generators respectively.

2.4.1 The MLM method

The MLM scheme is used in LO generators. The first step of the MLM approach is to determine the maximum multiplicity of the matrix element final state. Then the merging scale is defined: this parameter defines which part of the phase space is used by the matrix elements and which part is filled by the parton shower. For example, if the merging scale is an energy Q, the emission of partons with an energy greater than Q is computed by the matrix element, while the parton shower deals with lower energy partons. This value must be chosen with great care as too small values may spoil the perturbation approximation used for the calculation of the matrix element. Values that are too large, instead, imply that the parton shower works in areas of the phase space where the collinear approximation is no longer valid. At this point the events are showered using the merging scale as the upper cut-off of the shower evolution. The partons generated

by the showering are clustered in jets using a kT-algorithm (see Section 5.1.2) and then the jets are required to match with a matrix element parton.

2.4.2 The FxFx method

The FxFx procedure is the new standard in the Monte Carlo generators with NLO precision, as MadGraph5_aMC@NLO. When dealing with NLO calculations the problem is more complicated than the simple overcounting mentioned above. In the LO case, in fact, the only overcounting is due to the matrix element and parton shower overlap. In this case, instead, a further overlap arises due to the fact that different final state multiplicities are considered at the same time at the NLO. Consider for example the tree-level matrix element associated to a process which involves n partons in the final state. This element gives the Born contribution to the n parton production, but it will also give the first order real emission contribution to a process which involves only n-1 partons in the final state. The FxFx method uses an improved version of the MLM scheme: a merging scale is used to define the phase space areas filled by the Matrix element and the parton shower, and a kT-algorithm creates the parton clusters. Moreover some counter-terms are defined to deal with the additional overlap.

A detailed discussion of the MLM and the FxFx schemes goes beyond this thesis work. Further information can be found in the references [48, 50].

Chapter 3

Physics of the $Z+b$ process

The analysis of the production of a vector boson in association with heavy-flavored (HF) jets, originating from the hadronization of b quarks, is paramount for the Standard Model validation and also for the study of BSM processes. This kind of analysis allows to test the pQCD theoretical calculations and to improve the parametrization of the parton distribution functions. Moreover, at LHC, a sensitive production channel of the Higgs boson is in association with a Z boson, where the Higgs decays into a $b\bar{b}$ quark pair. The $Z+b$-jets associated production is therefore an irreducible background to this process. In addition to this, the $Z + b$ production is involved in many interesting Beyond the Standard Model processes; a better knowledge of their background could be fundamental to detect the first BSM signatures, especially in view of the 13 TeV second Run. The high energy and luminosity of the Large Hadron Collider provide in fact a perfect environment to make these precision measurements.

This Chapter starts with an introduction to the heavy flavor production within the pQCD framework; two methods, used to describe the $Z + b$ production, are here described: the 4 flavor and the 5 flavor schemes. Then there is an overview of the most important $Z + b$ signatures, with a focus on the measurement of the polarization asymmetry of the the $Z + b$ system. The forward-backward asymmetry is, in fact, the only Standard Model parameter which still present a consistent difference between the experimental value and the model prediction.

3.1 The QCD description of the $pp \to Z+b$ process

As mentioned, processes involving the b-quark production can be described using two different pQCD schemes: the 4-flavor scheme (4FS) and the 5 flavor scheme (5FS).

3.1.1 The 4 flavor scheme

In the four (or fixed) flavor scheme [51], the matrix element calculation is performed considering only the light quarks in the initial state. The b-quark does not contribute to the parton distribution functions nor in the calculation of the strong coupling constant. The basic tree-level diagrams for the associated production of $Z + b$ corresponds to the following processes:

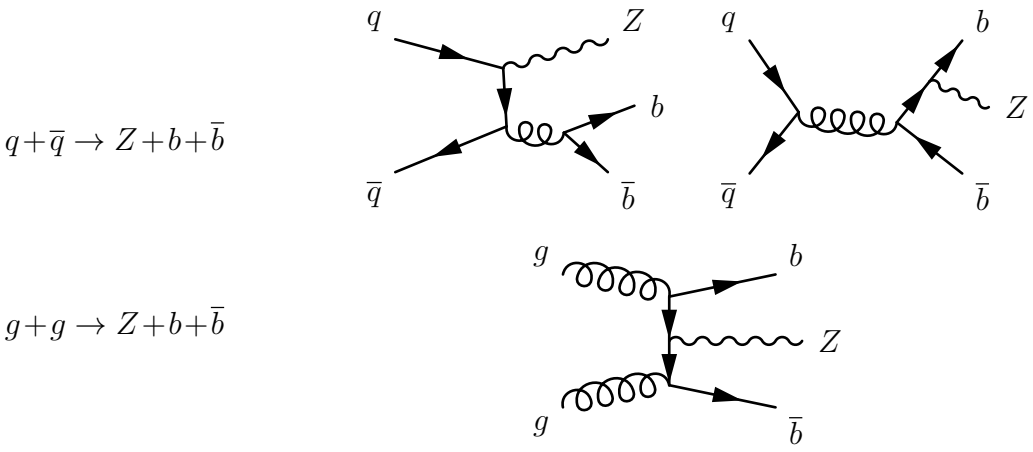

Figure 3.1: Examples of 4FS Z+b production at tree level.

In this approach all events have a $b\overline{b}$ pair in the final state. The scheme calculation is reliable under the assumption that the energy scale of the process is about the mass of the heavy quark ($m_b = 4.62$ GeV).

3.1.2 The 5 flavor scheme

Unlike the previous scheme, the five (or variable) flavor scheme [52] takes into account the b-quark contribution to the parton distribution functions of the proton, therefore $q + g \to Zq$ processes can be considered at the Leading Order. In this scheme the energy scale of the process is assumed to be higher than the b-quark mass which is neglected (sometimes m_b could take a small value different from zero if it is necessary to make the calculation finite). In Figure 3.2 the Leading Order diagrams for the $Z + b$ production, within the 5 flavor scheme, are represented.

In the 5FS, the matrix element calculation is easier than in the 4FS because a $2 \to 2$ process is used in place of a $2 \to 3$ process and the b mass is set to zero. Moreover, given the fact that the b-quark contributes to the parton distribution function of the initial state proton, it is possible to include logarithmic terms

Figure 3.2: Examples of 5FS Z+b production at tree level.

($\alpha_s \log \frac{Q^2}{m_b^2}$), arising from the gluon splitting, into the bottom PDF. For this reason the perturbative series converges more easily in the 5 flavor scheme.

In the end the two approaches lead to the same result when accounting for all the orders in pQCD. Nevertheless, when dealing with fixed order computations, some differences appear. At the LHC energies this discrepancy is small: the 4FS behaves better when dealing with exclusive quantities, while the 5FS is more appropriate to describe inclusive quantities. However it should be kept in mind that the 5FS is an approximation of the full kinematic process and only the 4 flavor scheme takes into account all the quark mass effects in the final state. The standard approach used by Monte Carlo generators is the 5 flavor scheme.

3.2 The state of art

The associated production of a Z boson and b-quarks has already been studied by different experiments. The CDF [53] and the D0 [54] collaborations have published the results obtained at FNAL, using the Tevatron collider, with a center of mass energy $\sqrt{s} = 1.96$ TeV. Their results showed a general agreement with the theoretical predictions but at the time the full NLO calculation was not disposable therefore a more detailed study was needed. Both the CMS [55] and ATLAS [56] experiment have measured the cross section of the $Z + b$ process, analyzing the inclusive process together with the production of a single b-jet and at least 2 b-jets. Both the final states are interesting to test the QCD predictions and they should preferably studied together because one sample could contaminate the other if the jet b-tagging is inefficient. Both the 4FS and the 5FS approaches have been used: as expected the 4FS provides a better description of the $Z + 2b$ exclusive process, while the inclusive cross section is well modeled by the 5 flavor scheme. At the moment the study has been performed using the data collected at $\sqrt{s} = 7$ TeV therefore it is interesting to compare the theoretical predictions at higher energies and luminosity.

3.3 The $Z+b$ final state

Below there is an introduction to some of the processes involving the production of a Z boson in association with b-quarks in the Standard Model and in some BSM models.

3.3.1 The Higgs decay into $b\bar{b}$

In the Standard Model the coupling between the Higgs boson and the other particles is proportional to their masses. Consequently, the Higgs will preferentially decay into heavy particles. The top quark is the one with the greatest mass but, since the Higgs boson has a mass of 125 GeV, the decay into a $t\bar{t}$ pair is kinematically forbidden. The main decay channel is therefore $H \to b\bar{b}$ which has a branching ratio of $\simeq 56\%$ [57]. Unfortunately the detection of this process is very difficult because there is a huge QCD background, which covers the Higgs decay signature. To avoid this problem the Higgs boson production is often studied in association with a vector boson ($pp \to Z^* \to ZH \to b\bar{b}l\bar{l}$). The corresponding Feynman diagram is represented in Figure 3.3.

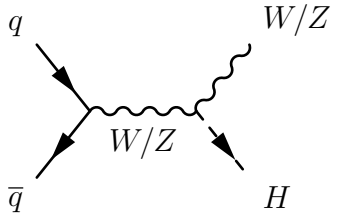

Figure 3.3: The associated production of a Higgs and a vector boson.

In this case the $Z + b$ production is a main background to the process. Its study is therefore fundamental to gain a clean signature of the Higgs decay into heavy quarks. Moreover, when dealing with supersymmetric extensions of the SM, the theory provides for the existence of many Higgses whose coupling with the b-quark could be even stronger. Again the $Z + b$ process would be an irreducible background to the decay of HZ states.

3.3.2 The fourth generation of heavy quarks

An open question of the Standard Model is why there are only three fermion families. In fact the only constraint, produced by the QCD, is that the number of families must be lower than 9, therefore the existence of a 4^{th} generation is theoretically possible. The 4^{th} generation quarks are generally indicated with t' and b' and they necessarily must be heavier than the Standard Model top and bottom quarks. Many models of this kind have already been excluded, but there

are still some that are under experimental examination. An interesting decay channel of the 4$^{\text{th}}$ generation b'-quark is described by the following diagram:

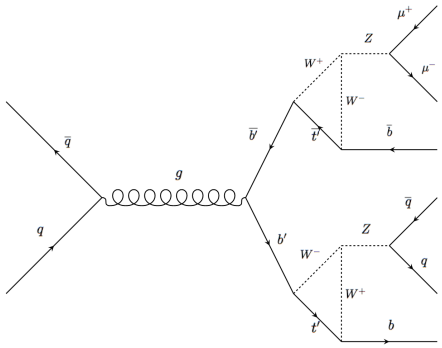

Figure 3.4: The decay of the 4^{th} generation b' quark into a SM bottom quark and a Z boson.

The process has the same final-state of the Standard Model $Z + b$ production which is again a background to the decay. At the moment, both the ATLAS and CMS experiments have not found any evidence of a new family of heavy quarks [58, 59].

3.3.3 The polarization asymmetry

Over the past decades, the Standard Model predictions have always been in good agreement with the experimental data. As proof of the effectiveness of the model, recently the ATLAS and CMS experiments have demonstrated the existence of the Higgs boson [9, 10]. However there are some observables whose measurements are still in tension with the theoretical predictions. In particular the forward-backward asymmetry A^b_{FB} of the bottom quark evaluated at the Z boson resonance peak differs by over 2 standard deviations from the value predicted by the SM [60]. In Figure 3.5 the Electroweak Global Fit for the main SM parameters is represented. The A^b_{FB} presents the biggest discrepancy with respect to the SM prediction:

$$A^b_{FB,exp} = 0.0992 \pm 0.0016, \tag{3.1}$$

$$A^b_{FB,th} = 0.1037 \pm 0.008. \tag{3.2}$$

This problem was already noticed at LEP and at SLC, studying the $e^+e^- \to b\bar{b}$ reaction. The polarized forward-backward asymmetry for this process is defined as:

$$A_{FB}^{b,pol} = \frac{(\sigma_{e_L\, b_F} - \sigma_{e_R\, b_F}) - (\sigma_{e_L\, b_B} - \sigma_{e_R\, b_B})}{\sigma_{e_L\, b_F} + \sigma_{e_R\, b_F} + \sigma_{e_L\, b_B} + \sigma_{e_R\, b_B}}, \tag{3.3}$$

where $b_{F,B}$ stands for the outgoing forward (backward) b-quark:

$$\sigma_F = \int_0^1 \frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta}\mathrm{d}\cos\theta \quad \sigma_B = \int_{-1}^0 \frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta}\mathrm{d}\cos\theta. \tag{3.4}$$

At the Z-mass resonance peak, this quantity is strictly connected to the couplings $g_{L,R}$ between the Z boson and the left (right) bottom quarks:

$$A_{FB}^{b,pol} = \frac{3}{4}\frac{g_L^2 - g_R^2}{g_L^2 + g_R^2} = \frac{3}{4}A_b. \tag{3.5}$$

The unpolarized asymmetry, which is the SM parameter generally used as a reference, can be easily defined as:

$$A_{FB}^{b} = \frac{3}{4}A_e A_b, \tag{3.6}$$

where A_e is the longitudinal electron asymmetry [61]. As outlined above, the LEP and SLC measurements presented a consistent discrepancy with the A_{FB}^b prediction of the Standard Model. At present time the problem persists and it is in the best interests of the collaborations working at the LHC to perform precision measurements to improve the experimental value of the A_{FB}^b quantity. A direct measurement is no longer possible since the initial state is constituted by quarks and gluons. It is however possible an indirect evaluation, for example, through the measurement of the Z boson polarization asymmetry A_Z, when the vector boson is produced in association with a b-quark. This quantity is defined as:

$$A_Z = \frac{\sigma_{Z_R} - \sigma_{Z_L}}{\sigma_{Z_R} + \sigma_{Z_L}} = C\frac{g_L^2 - g_R^2}{g_L^2 + g_R^2}, \tag{3.7}$$

where C is a convolution factor which takes into account the PDF of the b-quark and of the gluons. In this case, the Forward and Backward directions are defined in terms of the b-quark directions with respect to the Z boson reference frame. In the end, the A_Z quantity is simply proportional to A_b, therefore A_{FB}^b could be indirectly determined starting from A_Z [62].

Many BSM models have been proposed to improve the compatibility between the measurement and the theoretical prediction. Some of these require the existence of a heavy quark b' which couples with the Z vector boson. The existence of this partner of the bottom quark generates important correction to the $Z + b$ cross section and to the Z polarization asymmetry, reducing the discrepancy with the experimental data [63].

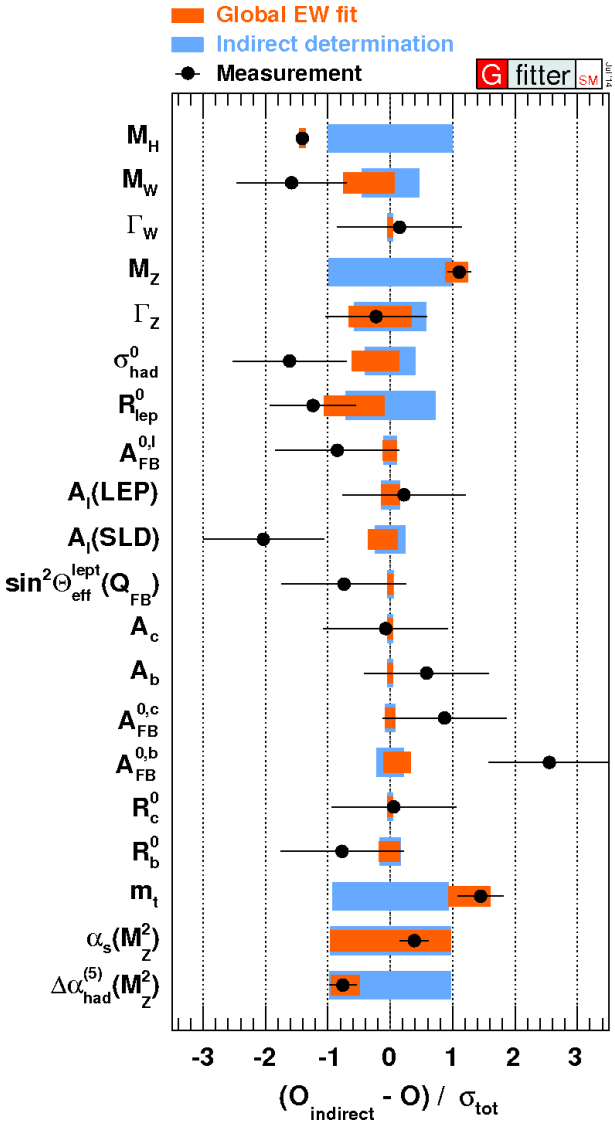

Figure 3.5: Comparison of the fit results with direct measurement and indirect determination of the SM parameters [99].

Chapter 4

The LHC and the CMS detector

In this Chapter there is a brief description of the experimental tools employed to collect the data used in the thesis. At first there is a short introduction to the CERN and the Large Hadron Collider, while the Compact Muon Solenoid experiment is described with particular attention in the second part.

4.1 The European Organization for Nuclear Research

The Conseil Européen pour la Recherche Nucléaire (CERN), in English the European Organization for Nuclear Research, is the world biggest laboratory in the field of particle physics. It was founded in 1954 by 12 member states but today there are more than 20 countries and 10000 people which are cooperating in the research. The laboratories are located on the border between France and the Geneve area, in Switzerland. The aim of the CERN foundation is to provide the environment and the fundamental tools which are necessary for particle physics research. In particular, 5 accelerators have been built: each one, after having accelerated the particles up to the requested energy, injects the particle beam into the next accelerator. This "chain" is organized in the following way: a linear accelerator (LINAC) injects 50 MeV protons into the PS Booster which in turn increases the particle energy up to 1.5 GeV. The subsequent Proton Synchrotron (PS) and Super Proton Synchrotron (SPS) accelerate the particles until a final energy of 450 GeV. At this point the beam is delivered to the last accelerator, the Large Hadron Collider. A scheme of the acceleration complex is presented in Figure 4.1

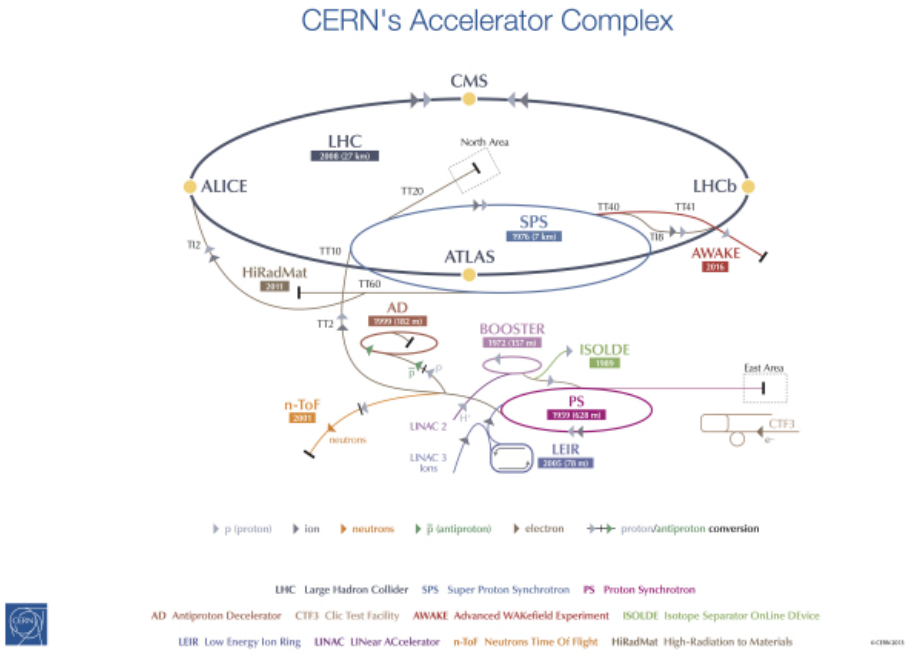

Figure 4.1: The complex of CERN accelerators that inject protons into the LHC [100].

4.2 The LHC

The LHC (Large Hadron Collider) [64] is the biggest and most powerful particle accelerator in the world. It can accelerate protons or heavy ions up to a center-of-mass energy of $\sqrt{s} = 14$ TeV with a design instantaneous luminosity (see Section 1.5) of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. It is a synchrotron, namely a circular accelerator in which the cyclic magnetic and electric fields, respectively responsible to maintain the particles along the path and to accelerate them, are time dependent, i.e they are synchronized with the increasing energy of the beam. LHC is located in a circular underground tunnel of 27 km which depth goes from 50 to 175 meters. It has been built between 1998 and 2008, substituting the previous accelerator which was housed in the same tunnel, LEP [65, 66] .The LHC collider consists of 2 parallel beamlines where the particles travel inside the vacuum. The particle beams move in opposite directions so as to collide in the four interaction points where the beamlines intersect.

LHC is a semi-circular accelerator made by 8 arcs, containing the magnets which drive the particle in its trajectory, and 8 straight areas. These zones are used to inject the beam from the pre-accelerator and to host the radiofrequency

cavities and the experiments. In the following a brief description of the main systems of the accelerator is provided:

- **Magnets.** More than 9600 different magnets are used to maintain the particles in their trajectory within the collider. Among these there are dipoles, quadrupoles, sextupoles etc. The dipoles curve the trajectory of the particles, while the other types are responsible for focusing the beam and providing minor corrections. The need to have dipolar magnets derives from the fact that, in the LHC, two beams of protons move in opposite directions. Therefore protons are kept in two separate vacuum tubes, one for the particles that circulate in one direction, in which the magnetic field has positive polarity, and the other for the particles that circulate in the opposite direction, in which the magnetic field has negative polarity. The dipoles are superconductive magnets whose magnetic field has an intensity of 8.3 Tesla. Due to the superconductive nature of these magnets, they must be continuously cooled by a liquid helium cooling system which keep them at 1.9 Kelvin.

- **Vacuum system.** 3 different vacuum systems are used at the LHC. Two of them are used to insulate the superconductive magnets and the liquid helium supply system. The third, instead, creates the vacuum inside the beamlines, keeping the pressure under 10^{-13} atm.

- **Radiofrequency cavities.** The radiofrequency (RF) cavities of LHC are tools that accelerate the protons and assort them in compact bunches of 10^{11} particles so as to maximize the number of interactions. When fully operational, these cavities provides bunches temporally separated by 25 ns for a total of more than 2800 bunches circulating inside the accelerator.

Along the LHC there are several detectors among which the main experiments are ALICE [67], LHCb [68], ATLAS [69] and CMS [70]. ALICE and LHCb are designed for heavy ions physics and flavor physics respectively, while ATLAS and CMS are multipurpose detectors.

4.3 The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) is one of the two multi-purpose experiments at the LHC, together with ATLAS. It is located at Cessy, in France, and it consists of a 21 m long detector, with 15 m of diameter and a weight of 12500 t. Being a general-purpose detector, it is able to study many aspects of the high energy interactions between protons, from the Standard model parameters to the New Physics theories and their associated particles. To deal with such a wide range of physics phenomena the detector has been built requiring its ability to reveal

many kinds of neutral and charged particles. With this aim, it has been structured with a series of cylindrical sub-detectors able to detect different interactions. The internal structure of the detector is depicted in Figure 4.2.

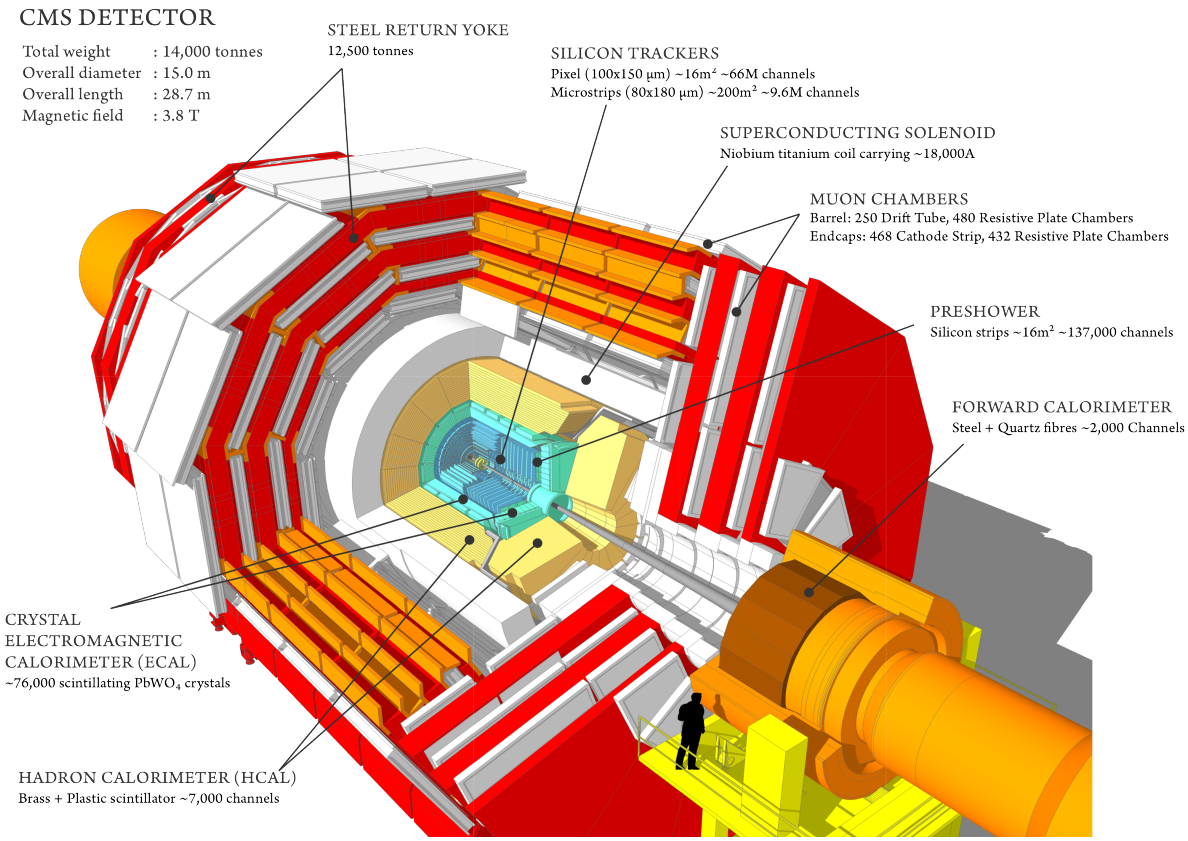

Figure 4.2: The internal structure of the CMS detector [101].

The inner detector, closest to the beam pipe, is the silicon tracker. Around the tracker, there is the electromagnetic calorimeter (ECAL), followed by the hadron calorimeter (HCAL). The tracker and the calorimeters are placed inside the solenoidal superconducting magnet that generates a magnetic field up to 4 Tesla. The last sub-detector is the muon detection system, placed in the iron yoke which closes the lines of force of the magnetic field and supports the entire detector. The central cylindrical part of CMS is usually called "barrel" while the two caps that close the sides of the structure are the "endcaps".

The CMS detector has been designed observing severe requirements:

- An intense magnetic field to measure the momenta of charged particles as electrons and muons.

- Good tracker resolution in order to measure the charged particle momenta and to reconstruct their tracks. Efficient tagging of heavy flavour jets, i.e. high density of the tracker inner sensors.
- Excellent resolution in the measurement of electromagnetic showers and large coverage of the electromagnetic calorimeter.
- Hermetic hadronic calorimeter to properly estimate the missing energy.
- High efficiency of the muon detection system.

In the following a brief overview of the main features of the CMS detector is presented.

4.3.1 The CMS coordinate system

The cartesian coordinate system (x, y, z) in CMS has the origin fixed at the interaction point, namely in the center of symmetry of the detector. The z-axis corresponds to the beamline direction, the x-axis points to the center of the LHC ring and the y-axis points upward to the surface. However, given the cylindrical symmetry of the machine, the Cartesian coordinate system is generally replaced by the three variables (r, η, ϕ). In this case the r variable is the radial distance from the origin, ϕ is the azimuthal angle in the xy plane and η, the pseudorapidity, is defined as

$$\eta = -\ln\left[\tan\frac{\theta}{2}\right] \tag{4.1}$$

where θ is the polar angle with respect to the z-axis. The η dependence over the polar angle is depicted in Figure 4.3. The pseudorapidity is highly preferred over the θ polar angle because differences in $\Delta\eta$ are Lorentz invariant. $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is also a Lorentz invariant measure often used to evaluate the angular separation between two particles.

4.3.2 The interaction point

The interaction point is the place where collisions take place between particles. When the two beams collide, they are slightly inclined (280 μrad) and they have a transverse extension of 17 μm. This thesis is based on the data collected in 2012, when the maximum energy achieved in the center of mass was 8 TeV and the maximum luminosity was $7 \times 10^{33}\,\mathrm{cm^{-2}s^{-1}}$. As already mentioned, in 2015, after a two years long technical stop, the energy of 13 TeV has been reached. Later this year it is also expected that the luminosity will rise up to the design value of $10^{34}\,\mathrm{cm^{-2}s^{-1}}$. The consequence of a so high collision rate is a huge pile-up, i.e. a large number of superimposed secondary interactions. This effect makes the

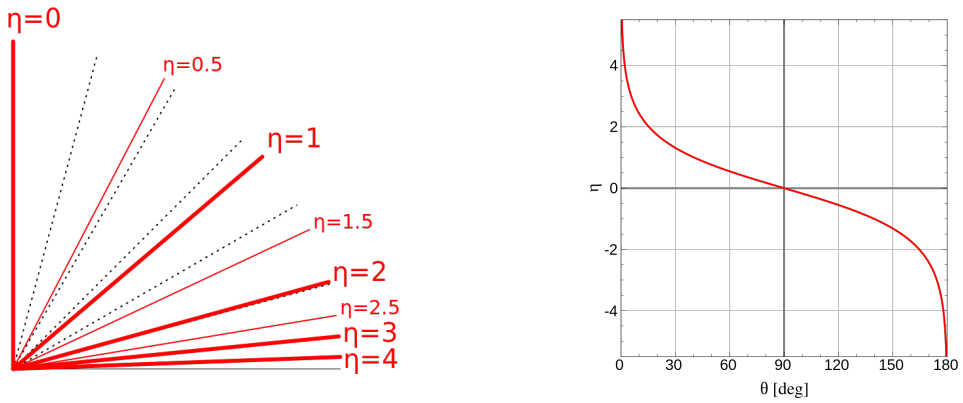

Figure 4.3: Pseudorapidity as a function of the θ polar angle [102].

event reconstruction even more difficult. To deal with this problem the detector must have a good time resolution and a high granularity to reduce the particle occupancy. With this in mind, the inner part of CMS, the tracker, was built using silicon sensors which are particularly fast (see the next Section). An example of a proton-proton collision with 78 interaction vertexes is depicted in Figure 4.4.

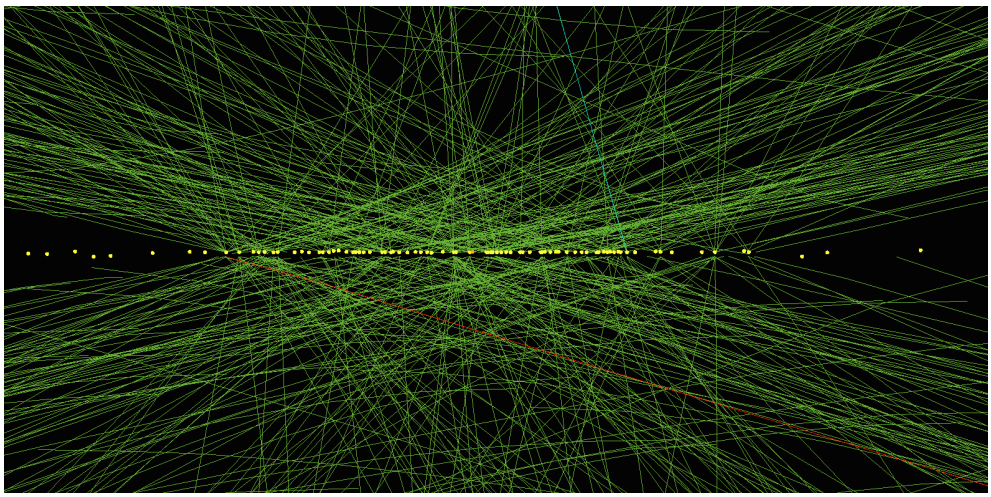

Figure 4.4: Example of a pp collision with 78 interaction vertexes [103].

4.3.3 The tracker

The inner sub-detector of CMS is the tracker [71]. It is the fundamental tool to track the particles trajectory and to reconstruct their momentum. The tracker is capable of reconstructing the trajectory of muons, electrons, hadrons and it can

also distinguish tracks created by particles with a very low average life such as the b-quark. It is divided in three zones with different technical specifications. The closest part ($r < 11$ cm) to the interaction point consists of three layers of silicon pixel-detectors. There is a total of 65 millions pixels, each one with a size of $100 \times 150\,\mu\text{m}^2$. This high granularity permits a clean reconstruction of the tracks even in case of an high level of pile-up. Such a large number of sensors also ensures a proper identification of the track primary vertexes. The middle ($r < 55$ cm) and the external ($r < 1.3$ m) regions of the tracker are composed by silicon microstrips. These sensors are larger than the pixels and provide a lower resolution. There are more than 24000 silicon strip sensors; they have a size of $10\,cm \times 80\,\mu$m in the intermediate region and $25\,cm \times 180\,\mu$m in the external one. In Figure 4.5 there is a schematic representation of the tracker structure. The tracker covers a limited pseudorapidity range $|\eta| < 2.4$.

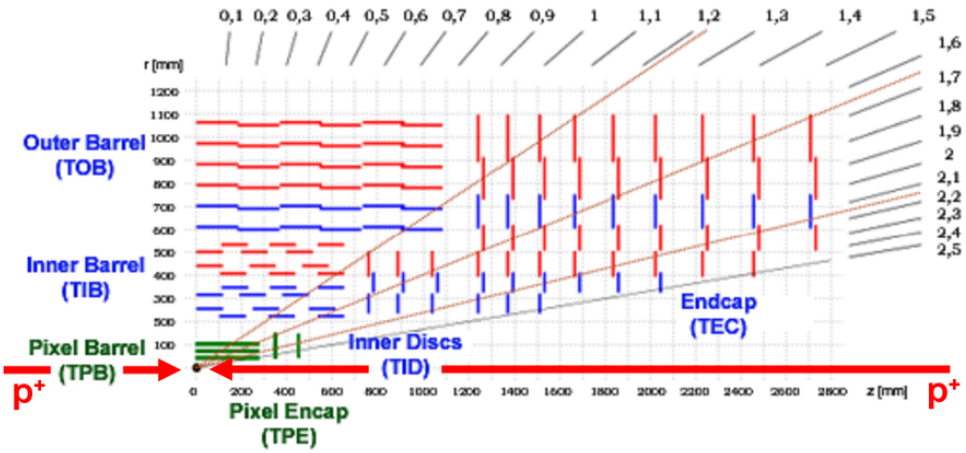

Figure 4.5: The internal structure of the CMS tracker subdetector [104].

4.3.4 The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [72] is the CMS sub-detector that measures the energy of electrons and photons. The calorimeter is made of 76000 lead tungstate ($PbWO_4$) crystals, distributed between the barrel (61000) and the endcaps (15000). When a particle pass through the crystals, scintillation light is produced, proportionally to its energy. On the back of each crystal there is a photodetector which converts the light radiation into an electric signal. In particular avalanche photodiodes (APD) and vacuum phototriodes (VPT) are used in the barrel and in the endcaps respectively. The presence of amplified photomultiplier tubes is necessary to compensate the low light yeld of the lead-tungstate crystals (30 p.e./MeV). The choice of using $PbWO_4$ for the crystals is due to the

fact that it has a small radiation length $X_0 \simeq 0.89$ cm. This feature has allowed to reduce the size of ECAL and place it inside the magnet while maintaining a good calorimeter hermeticity. The crystals have a length that corresponds to more than $24X_0$. The $PbWO_4$ has a good radiation hardness therefore the crystals are supposed to work for years without high deterioration. The presence of the "preshower detector" must also be mentioned. This is a sub-detector placed between the tracker and ECAL, in the endcap region. It consists of two layers of lead and silicon strips that allow to distinguish between single and multiple photons entering the calorimeter. The energy resolution of the CMS calorimeter (barrel) is:

$$\frac{\sigma_e}{E} = \frac{12\%}{\sqrt{E}} + \frac{2.8\%}{E} + 0.3\%, \tag{4.2}$$

where the contributions are the noise, the stochastic and the constant term respectively. The result was obtained using an electron beam test [73] .

Figure 4.6: A $PbWO_4$ ECAL crystal with a vacuum phototriodes at the end [105].

4.3.5 The hadronic calorimeter

The hadronic calorimeter (HCAL) [74] measures the energy and the momentum of hadrons. It is also a key element in the reconstruction of the event missing energy, i.e. the energy that is lost because of the neutrino component of jets. HCAL is a sampling calorimeter; it consists of alternating layers of a dense absorbing material, brass, and of a plastic scintillator. To ensure the hermeticity, the calorimeter covers a very wide pseudorapidity range $|\eta| < 5$. This is possible thanks to the forward (HF) sections of the calorimeter that are positioned at either ends of the detector ($3 < |\eta| < 5$). Moreover, in addition to the barrel (HB) and endcap (HE) sections, there is an another section (HO) to the outside of the magnet to detect an eventual energy leakage. The total thickness of HCAL is of 10 interaction lengths λ_i ($3\lambda_i$ are added thanks to the HO section). The energy resolution is

$$\frac{\sigma_e}{E} = \frac{100\%}{E} + 8\%. \tag{4.3}$$

where the values have been measured using a π^0 test beam [75].

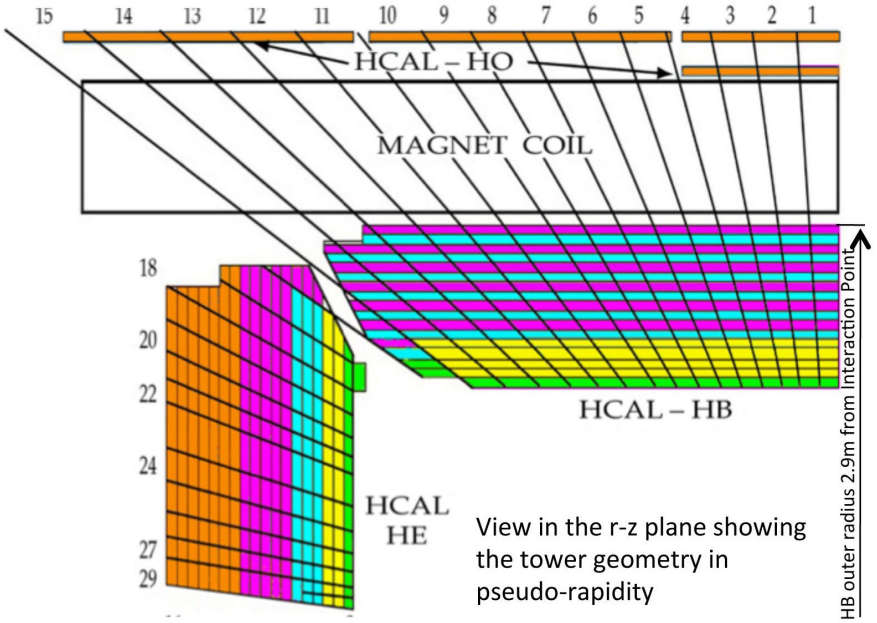

Figure 4.7: The side view of HCAL [106].

4.3.6 The Magnet

The magnet [76] consists in a superconductive cylindrical solenoid cooled to $-268.5°$, and it contains all the instrumentation described so far (apart from the HO region of the hadronic calorimeter). It is 13 m long and it has a diameter of 6 m. The produced magnetic field is about 3.8 T, i.e. almost 100,000 times more powerful than the Earth's field. The magnet bends the trajectory of the charged particles: tracing their trajectory, and determining the corresponding radius, it is then possible to measure the particle momenta.

4.3.7 The muon detection system

Outside the solenoid there is the last part of the CMS detector, which is the muon detecting system [77]. This sub-detector is dedicated to the muon identification and to the measurement of their transverse momentum. This is the outer part of CMS because the muon is the only particle that interacts so little to be able to overcome all the other layers. The system is embedded in the iron return yoke

of the magnet inside a 1.8 T magnetic field. The detection system consists in four layers of muon stations in the barrel and six layers in the endcaps. Each station is composed by various detection modules: in particular drift tubes are used in the barrel while cathode strip chambers are used in the endcaps. There are also 612 resistive plate chambers distributed among both the barrel and the endcap regions. A transverse section of the barrel detection system is represented in Figure 4.8.

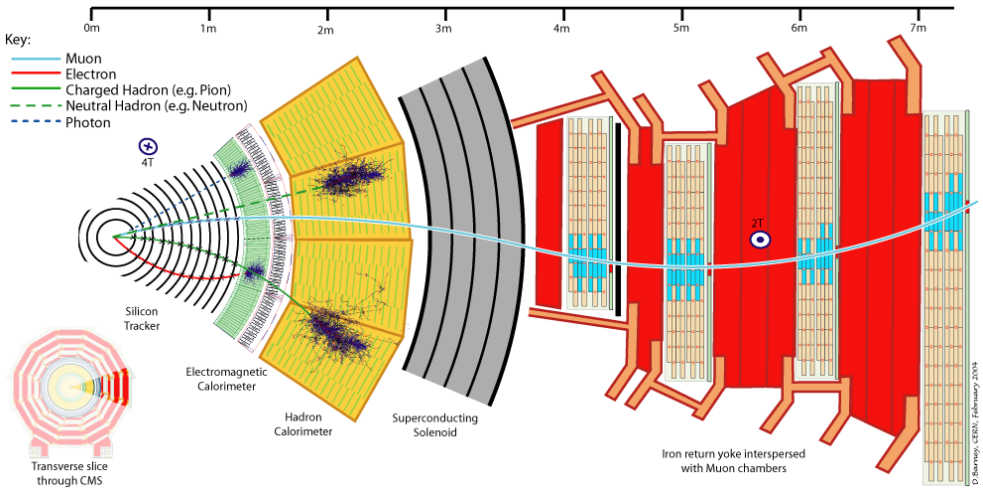

Figure 4.8: A transverse section of the CMS detector [107].

4.3.8 The trigger system

During the data taking, there are about 10^9 interactions per second inside the detector. It is impossible to collect, store and analyze such an amount of data. However most events are "minimum bias", i.e. they consists of soft interactions that are irrelevant to the research purpose of CMS. It is therefore more beneficial to filter these events and collect only the relevant data. To accomplish this skimming procedure a trigger system has been created. There are two distinct levels of trigger. The Level-1 trigger is a very fast process which automatically identify possible interesting events, such as high energy peaks, and select about 100,000 events per second from the initial billion. It is hardware-based and, to quickly process the data, only the information from the electromagnetic calorimeter and the muon system is used. In particular the Level-1 trigger is divided in two steps: at first the information coming from the calorimeter and the muon sub-detector is processed separately by the Calorimeter Trigger and the Muon Trigger and then a Global Trigger analyzes the merged information before passing it to the next trigger level. During this phase, the event rate goes from 40 MHz to 100 kHz. A

second decrease in the number of events is managed by the High-Level Trigger (HLT) which reduces the rate down to 100 Hz. At this stage the entire information, coming from the detector, is used to take the final decision about keeping or discarding the specific event. The last step of the triggering process is to perform a first event reconstruction: various datasets, corresponding to different physical processes, are created to simplify the event identification.

Chapter 5

Analysis of the $Z+b$ events

As already mentioned in Chapter 3, the study of the associated production of b quarks and a Z boson is important both in terms of the Standard Model validation and of the search for new physics. The main objective of this thesis is to test the NLO effects of the $Z + b$ production, comparing the experimental data with the theoretical predictions provided by some Monte Carlo Generators, with particular attention to the related theoretical uncertainties. In the first part of the Chapter, the technique used for the reconstruction of physical objects in CMS is presented, before focusing on the strategy adopted to identify events at particle level. The second part of the Chapter is instead dedicated to a series of comparative studies performed to evaluate the stability and consistence of the choices made in the selection of the generated events.

5.1 Physics object reconstruction

The key element used in the reconstruction of physical objects in CMS is the Particle Flow (PF) algorithm [79]. It combines the measurements from all the sub-detectors of CMS (tracker, calorimeters, muon chambers) to characterize the final particle momentum and its direction. It is able to distinguish between leptons and jets, to assess the missing energy fraction, and to determine the jet flavor. The general procedure followed by the PF algorithm can be schematized as follows: at first the muon identification is performed and the corresponding tracks are removed from the collection. Then the energy deposits in the hadronic calorimeter are matched with the remaining tracks to build the hadrons. Subsequently, electrons are reconstructed using the information of the tracker and ECAL. At the end the remaining ECAL deposits are associated with photons.

On the next pages there is a focus on the reconstruction of the objects which are most relevant to this thesis, namely the Z boson reconstruction and the identification of the b-jets, i.e. jets coming from the hadronization of b-quarks.

5.1.1 Reconstruction of the Z boson

The Z boson reconstruction starts selecting an electron or muon pair. The presence of two opposite charge leptons with a transverse momentum $p_t > 20$ GeV, and in the geometrical range of $|\eta| < 2.4$, is required. The four-momenta of the two particles are summed up to form the Z boson and the event is accepted only if the Z invariant mass is in the range $[71, 111]$ GeV. Additional criteria are also applied to the reconstructed leptons: a minimum number of hits in the tracker (and also in the muon chambers if the lepton is a muon), a specific ratio between the energy deposited in the hadronic and the electromagnetic calorimeters (H/E), a good χ^2 of the track global fit and a maximum value for the impact parameters of the lepton, i.e. the transverse (d_{xy}) and longitudinal (d_z) distances to the primary vertex. Moreover it is required that the leptons are isolated to reject the electrons or muons coming from the semileptonic decay of heavy quarks inside a jet (see Figure 5.1).

A lepton candidate is said to be "isolated" if the isolation variable $I_{e,\mu}$ has a value lower than 0.15 and 0.2 for electrons and muons respectively. $I_{e,\mu}$ is defined as the sum of the Particle Flow reconstructed energies of all the CMS subdetectors inside a cone $\Delta R < 0.3$ around the lepton:

$$I_{e,\mu} = \frac{1}{p_T^{e,\mu}} \left[\sum_{tracks} E_T + \sum_{ECAL} E_T + \sum_{HCAL} E_T \right]. \tag{5.1}$$

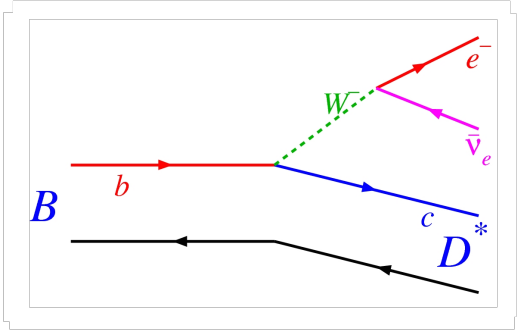

Figure 5.1: The semileptonic decay of a B meson [108].

5.1.2 Jet reconstruction

The jet reconstruction consists in clustering together the particles which come from the Particle Flow candidate list. In order to properly reconstruct the jet set, it is necessary to decide an appropriate jet definition, which correctly maps

the hadrons, avoiding overcounting and which creates jets that are infrared and collinear safe. The algorithms which perform the jet reconstruction can be divided in two categories: cone algorithms and sequential algorithms. The algorithm used in this thesis belongs to the second class; these algorithms are based on the definition of two effective distances:

$$d_{i,j} = min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta_{i,j}^2}{R^2}, \tag{5.2}$$

$$d_{i,B} = k_{T,i}^{2p}, \tag{5.3}$$

where $\Delta_{i,j} = \sqrt{|\eta_i - \eta_j|^2 + |\phi_i - \phi_j|^2}$ is the the distance between the particles i,j in the rapidity-azimuthal plane, $k_{T,i}$ is the transverse momentum of the specific particle, and R is an input parameter that characterize the maximum angular distance between two particles.

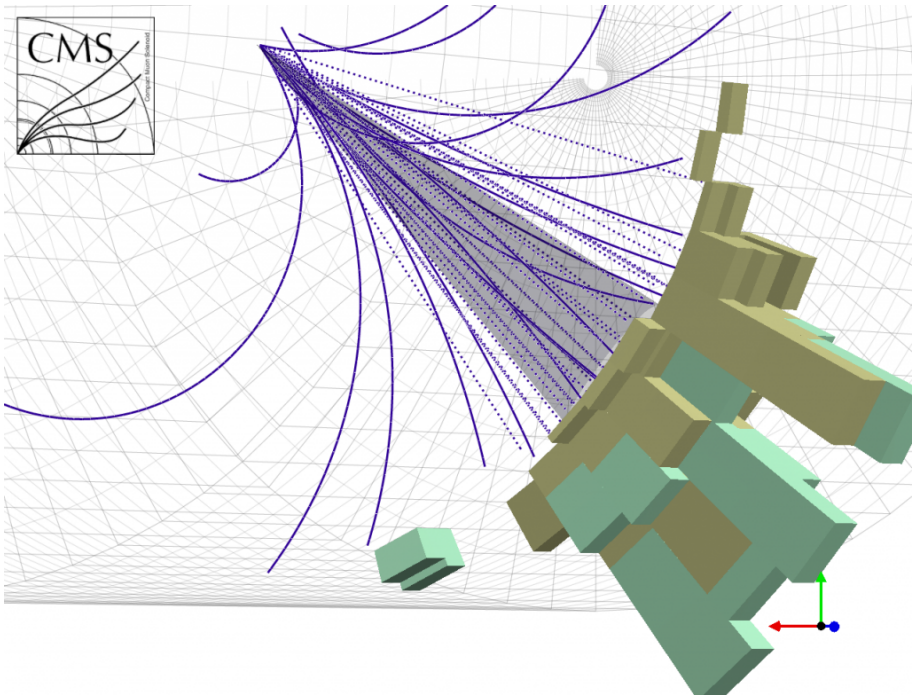

Figure 5.2: A graphical representation of the anti-kT algorithm jet clustering in the CMS experiment [109].

The algorithm identifies the smallest effective distance between the $d_{i,j}$ and $d_{i,B}$; if a $d_{i,j}$ is the smallest, the i,j particles are merged together, otherwise if it is the $d_{i,B}$, the i particle is removed from the collection and it forms a new jet. The p-parameter identifies the kind of algorithm: $p = 0$ is the Cambridge-Aachen algorithm, $p = 1$ is the kT-algorithm, while $p = -1$ is the anti-kT algorithm.

In the Cambridge-Aachen algorithm the constituents are clustered taking into account only their spatial separation, in the kT algorithm the soft objects are merged together at first, while, in the anti-kT algorithm, the hard particles are the first to be clustered. The anti-kT is the standard algorithm at the LHC and it is the one used in this work (the radius parameter R has been set to $R = 0.5$).

After the particle clustering, the jets have to fulfill the following requirements to pass the selection:

- they must be formed by at least two constituents among which one charged constituents;
- the charged hadronic fraction must be grater than 0%;
- the neutral hadronic fraction, the neutral and charged electromagnetic fractions must be lower than 99%;
- the jet transverse momentum must be grater than 30 GeV;
- the jet must be placed in the η range $[-2.4; 2.4]$.

5.1.3 The b-tagging

There are many algorithms which performs the jet b-tagging. They all exploit the unique characteristics of the B hadrons, in particular their "long" lifetime. The B lifetime is in fact around 1.5 ps, which corresponds to a decay length of $\sim 450\,\mu$m. Thanks to the high resolution of the tracker, the CMS experiment is capable of observing such small distances, therefore it can distinguish between the primary vertex of the hard collision and the secondary vertex created by the B decay (see Figure 5.3). In addition to this, the B hadrons have a large mass (~ 5 GeV) and this results in a high multiplicity of the charged decay products (5 tracks in average). Another important feature of the B-hadrons, which is exploitable to increase the flavor tagging accuracy, is the fact that they present a semileptonic decay about the 10.7% of the times. For a detailed description of the available b-tagging algorithms see Ref. [80]. The algorithm used to b-tag the reconstructed jets shown in this analysis is the Combined Secondary Vertex (CSV) [81], which uses the flight distance of the b-quark and the invariant mass of the particle associated to the secondary vertex as discriminators.

5.2 The event selection at particle level

The ultimate test for any Monte Carlo event generator is whether it can reproduce the experimental data with a good approximation. In addition to a correct

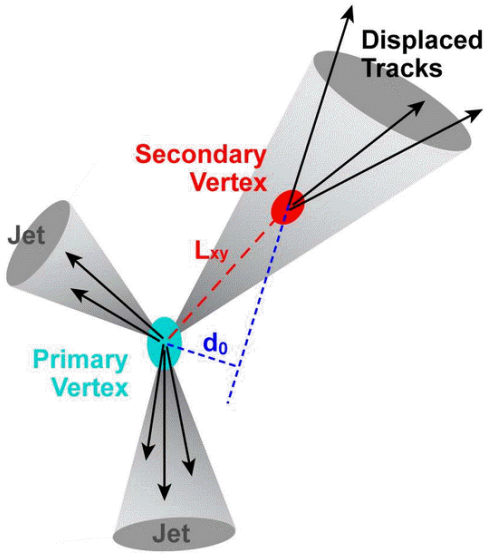

Figure 5.3: A schematic representation of the b-jet secondary vertex production [110].

modeling of the hard scattering, the parton shower and the hadronization, also the selection applied to accept or reject the generated events is a fundamental piece of the simulation process. In general the selection applied to the experimental data and the one used for the generated events almost match. In this way it is not necessary to extrapolate the data distribution in areas of the phase space where the detector is not capable of revealing particles, for resolution or acceptance reasons.

To perform the event selection at particle level, a Rivet [82] analysis has been built (see Appendix A). Rivet (Robust Indipendent Validation of Experiment and Theory) is a C++ class library which contains the necessary tools for the analysis of data in high-energy interaction experiments. Rivet operates using the "HepMC" standard for the event record, while the data saving is performed using the "Yoda" proprietary format, whose conversion in Root is guaranteed by the package *yoda2root*. A primary feature of Rivet is the presence of the projections. These are functions that allow to "project out" the particles of interest, or more generally the observables, from the event collection. For example, the projection "final state" returns only the particles of the final state, the "chargedfs" only the charged particle subgroup etc.

In this analysis the goal is to select $Z+b$ events, where the Z boson decays into electrons or muons. The cuts used at particle level are the same ones that were

applied during the reconstruction of the physical objects. It is therefore requested that the leptons and the jets have a minimum p_T of 20 and 30 GeV respectively and that their direction is inside the $|\eta| < 2.4$ range. The jet collection is created using the anti-kT algorithm mentioned in Section 5.1.2, with $R = 0.5$.

The particle identification is performed using its PID. In the event simulation, the Monte Carlo generator associates to every particle a number which identifies the particle type; this number is called Particle Identification (PID). A complete table of the PIDs can be found at [32], while the most relevant PIDs used in this work are listed in Table 5.1.

Particle (anti-particle)	e^- (e^+)	μ^- (μ^+)	γ	b ($\bar{b}$)
PID	11 (-11)	13 (-13)	22	5 (-5)

Table 5.1: Particle PIDs.

The analysis of the events starts selecting leptons of opposite charge. The momentum requirement ($p_T > 20$ GeV) is not applied onto this "bare" particles, but onto the "dressed" leptons. Electrons and muons, in fact, may emit neutral radiation (photons), due to final state radiation effects, and they must be accounted when reconstructing the total momentum of the particle. Therefore a "dressing" procedure is performed: all the photons within a $\Delta R < 0.1$ cone around a lepton are merged together with the lepton itself. The two dressed leptons with the highest p_T are used to reconstruct the Z boson. If the leptons do not satisfy the momentum or the range requirements and if the mass of the Z boson does not fall in the mass window $[71; 111]$ GeV the event is vetoed.

At this point, photons and leptons used to identify the Z are removed from the particle collection. In this way there is no overcounting between the Z associated particles and the jets. Also the neutrinos are removed from the particle collection to reproduce the fact that they are not actually detected and they are not accounted in the experimental data distributions. Then the anti-$kT(0.5)$ algorithm is applied on the remaining particle. The jets must have a transverse momentum grater than 30 GeV and they must be well separated from the Z-leptons. This condition is implemented requiring an angular distance between jets and leptons greater than $\Delta R = 0.5$. In this way the Final State Radiation (FSR) remnants of the lepton outside the 0.1 cone do not alter the jet structure. Moreover this selection cut helps to exclude events in which the electron or the muon come from the semileptonic decay of a heavy hadron. The final stage of the event selection is the identification of the b-jets. This procedure is performed looking for a B-hadron between the ancestors of the jet particle collection. For every particle, the last B-hadron is selected, i.e. the hadron which does not have another b-particle as a daughter. If the B-hadron direction is within a $\Delta R < 0.5$ cone around the jet direction, the jet is b-tagged. The B-hadron is identified looking for a particle whose PID contains a 5, i.e. a particle which contains a b-quark.

The final result, after the event selection, consists in many differential count distribution. To obtain the differential cross sections, all the histograms are multiplied by the factor $f_{scale} = \frac{\sigma_{th}}{N_{gen}}$, which takes into account the total number of generated events and the theoretical cross section. The production of $Z + b$ is studied in the inclusive case, when the number n_b of b-jets is greater than 0, the single b case, $n_b = 1$, and the case of at least two b-jets, $n_b \geq 2$. The results are presented combining the $Z \to e^+e^-$ and the $Z \to \mu^+\mu^-$ processes with the weighted cross section of the single channel.

The distributions used to characterize the $Z+b$-jet final state are the following:

- The transverse momentum of the Z boson;
- The p_t and η of the leading b-tagged jet. "Leading" means the jet with the highest transverse momentum;
- The b-jets multiplicity;
- The scalar sum of all the b-jet transverse momenta, $H_t = \sum_i p_{T,i}$;
- The azimuthal angle $\Delta\phi(Zb)$ between the Z boson and the leading b-jet.

In the case of at least two b-jets also other observables are measured. Many angular correlation variables are investigated: they allow to test the theoretical prediction related to small angle emissions, which are still subject to significant uncertainties.

- The p_t and η of the subleading b-tagged jet;
- The angular separation between the leading and the subleading b-jets in the (η, ϕ) plane, ΔR_{bb}, and in the transverse plane, $\Delta\phi_{bb}$. The ΔR_{bb} is a direct probe of gluon splitting sub-processes, while $\Delta\phi_{bb}$ is used to test the b-quark back-to-back configuration;
- The angular separation between the Z boson and the closest b-jet, ΔR_{Zb}^{min};
- The angular separation between the Z boson and the farthest b-jet, ΔR_{Zb}^{max};
- The asymmetry between the direction of the Z boson and the b-jets,

$$A_{Zbb} = \frac{\Delta R_{Zb}^{max} - \Delta R_{Zb}^{min}}{\Delta R_{Zb}^{max} + \Delta R_{Zb}^{min}}.$$

  It is an indirect way to evaluate the modeling of higher order QCD processes. Its value is 0 when the b-hadrons are emitted symmetrically; a deviation it expected when additional gluons are radiated by the hadrons.

Also a couple of distributions related to new physics searches are investigated:

- The invariant mass distribution of the two leading b-jets, M_{bb};
- The invariant mass distribution of the Z boson and the two leading b-jets, M_{Zbb}.

5.3 Data and Monte Carlo samples

In the next Chapter, the results obtained by Monte Carlo simulations will be compared with experimental data. The data were collected with the CMS experiments during the 2012, when the center of mass energy of the LHC collider was $\sqrt{s} = 8\,\text{TeV}$. The datasets used in this work are enlisted in Table 5.2: they are divided in run periods and in electron and muon samples. The total integrated luminosity of the datasets is $19.78\,\text{fb}^{-1}$ for the $Z \to e^+e^-$ set and $19.75\,\text{fb}^{-1}$ for the $Z \to \mu^+\mu^-$.

Run Range	Primary Datasets
Run2012A	/DoubleElectron/Run2012A-22Jan2013-v1
	/DoubleMu/Run2012A-22Jan2013-v1
Run2012B	/DoubleElectron/Run2012B-22Jan2013-v1
	/DoubleMuParked/Run2012B-22Jan2013-v1
Run2012C	/DoubleElectron/Run2012C-22Jan2013-v1
	/DoubleMuParked/Run2012C-22Jan2013-v1
Run2012D	/DoubleElectron/Run2012D-22Jan2013-v1
	/DoubleMuParked/Run2012D-22Jan2013-v1

Table 5.2: List of Primary Datasets, divided in run periods, for electrons and muons samples.

Regarding the Monte Carlo simulation, different generated samples have been used to simulate the signal and the background. The latter have been used to extract the signal from the measured data before applying the unfolding procedure. All the simulated data samples have been generated using the same energy of the measured data (8 TeV) and they have been rescaled to the higher order cross section available (NNLO for Drell-Yan Z+jets). The cross sections used for the normalization of signal and background are listed in Appendix B. Three Drell-Yan Z+jets samples have been simulated to reproduce the signal, using MadGraph5+Pythia6, POWHEG(MiNLO)+Pythia8 and MadGraph5_aMC@NLO+ Pythia8. The parton distribution functions used are the CTEQ6L1 [28] for the MadGraph5 sample, NNPDF3.0 [29] for MadGraph5_aMC@NLO and for POWHEG.

The POWHEG(MiNLO) Z+2jets sample has been produced by our group and it is used for the first time in this analysis. The total cross section of the process is $2312.82\,\mathrm{pb}^{-1}$. Only 2 millions of dielectron and dimuon events have been produced at the moment. The production of a larger sample using the Worldwide LHC Computing GRID is programmed in the short term. The MadGraph5 and the MadGraph5_aMC@NLO samples are taken from the official CMS repository and they contain about 30 millions Drell-Yan Z+jets events with a total cross section of $3531.8\,\mathrm{pb}^{-1}$.

The inclusive jet multiplicity and the b-jet multiplicity of the Monte Carlo samples described above are represented in Figure 5.4, together with the statistical error. Also the ratio plot is reported, using the MadGraph5 sample for reference. The POWHEG and MadGraph5_aMC@NLO samples seem to be in good agreement; in particular they almost coincide for a b-jet multiplicity lower than 3.

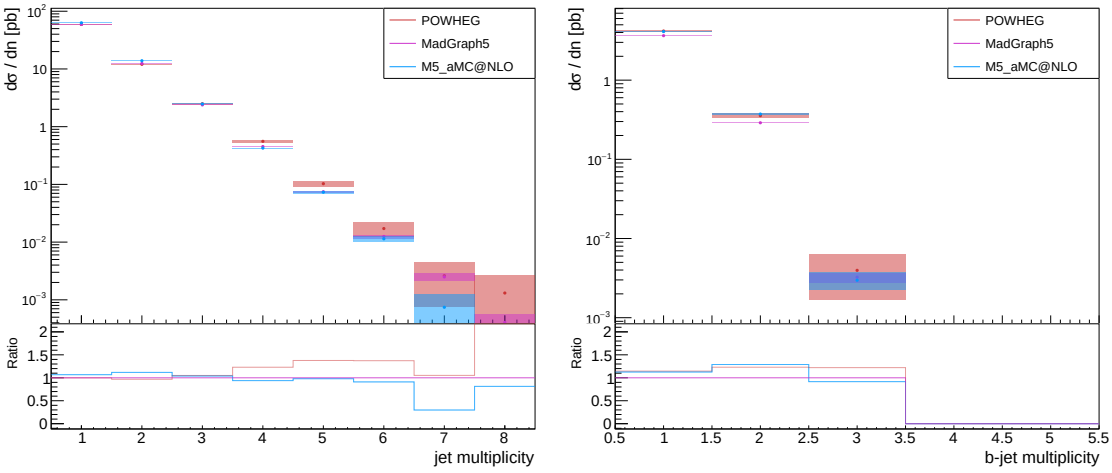

Figure 5.4: The inclusive jet multiplicity (left) and the b-jet multiplicity (right) of the MadGraph5, POWHEG and MadGraph5_aMC@NLO samples.

As already mentioned (see Section 2.2), there are two ways to compare experimental data and theoretical prediction. The first is to pass the generated events through the detector simulation. At the end of this procedure the different background contributions and the signal can be summed together and a direct comparison with the experimental data is possible. Also pile-up events must be superimposed to the generated signal. In Figure 5.5 the leading b-jet momentum distribution is represented together with the simulated data at detector level, using the GEANT4 [83] package. The different colors represent the various background contributions.

However, it is often preferred to subtract the background contribution and to extrapolate the "true" distribution from the observed one, applying the unfolding procedure. This method allows to remove the resolution and acceptance effects of the detector from the measured distribution to get an approximation of what an ideal detector would have revealed. Different approaches are available: the data shown in this thesis have been unfolded applying the Singular Value Decomposition (SVD) [84], using the RooUnfold [85] package. In the SVD approach a response matrix is built for every observable. This matrix encodes the detector effects corresponding to the specific distribution and describes how an histogram changes moving from the detector level distribution to the particle level one. From now on, all the comparisons will be done between simulated data and unfolded measured data. As an example, in Figure 5.6, the leading b-jet momentum distribution is represented, for the electron and muon sample, before and after the unfolding.

The unfolded distributions of the experimental data will always be presented with the total (systematic and statistical) error. The sources of statistical uncertainties are the $t\bar{t}$ background and the finiteness of the data and Monte Carlo samples. Regarding the systematical uncertainties, instead, the main contributions come from the b-tagging efficiency, the LHC luminosity, the jet reconstruction (energy resolution and energy correction effects) and the uncertainties related to the number of interactions and to the pile-up models. In addition, the unfolding method and the b-purity extraction give both statistical and systematic contributions (the b-purity is a measurement of how many jets are misidentified as b-jets). A detailed description of the error sources and the data error correlation can be found in Ref. [88].

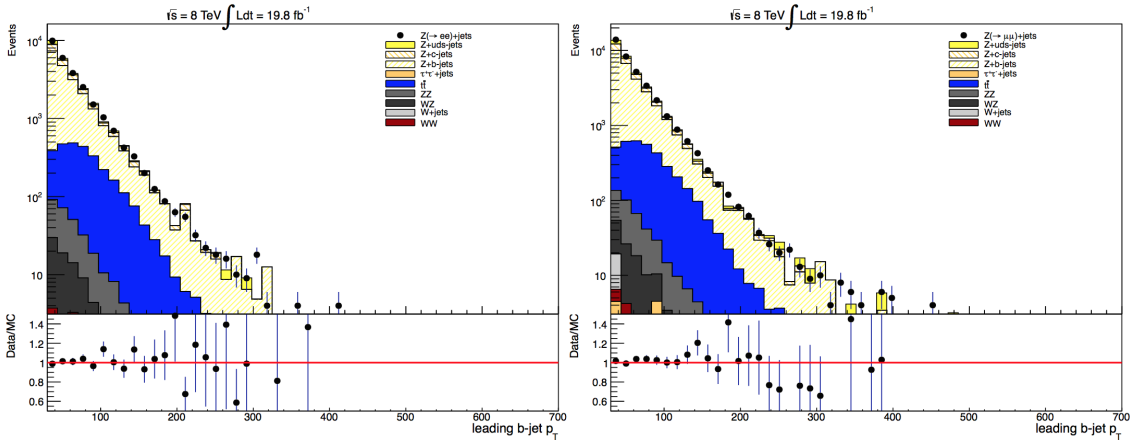

Figure 5.5: The leading b-jet momentum distribution of the measured data with the superimposition of the simulated background and signal. $Z \to e^+e^-$ on the left and $Z \to \mu^+\mu^-$ on the right.

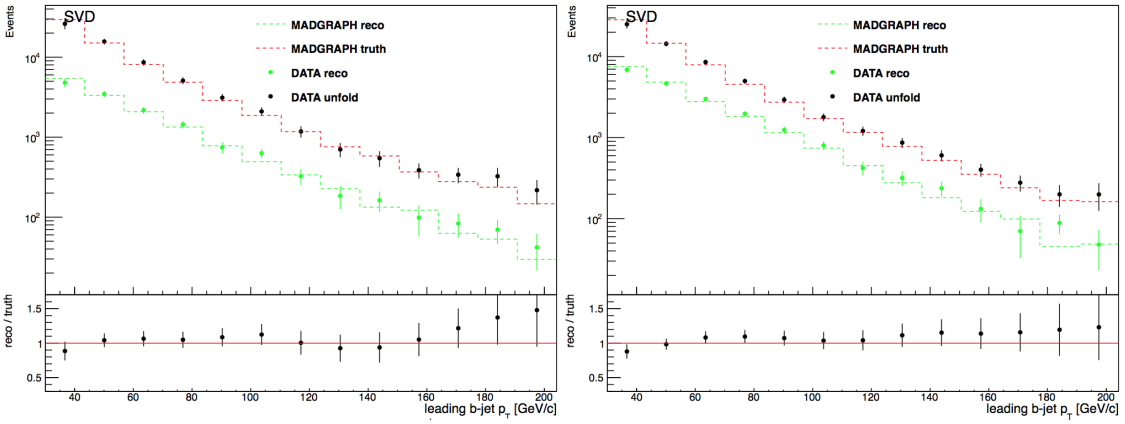

Figure 5.6: The leading b-jet momentum distribution before and after the unfolding procedure. All the background contributions have been subtracted from the signal. Green line and dots are the simulated and measured data at detector level, i.e. before the unfolding. The red line and the black dots, instead, are the simulated and measured data after the unfolding.

5.4 Comparative studies on the event selection

In this Section there is a comparison between the different results obtained by varying the conditions on the event selection at particle level. This type of studies allow to evaluate the consistence of the applied cuts and to define the phase space used in the analysis.

5.4.1 The lepton dressing

As explained in Section 5.2, in the standard Rivet analysis the Z boson is reconstructed starting from the dressed leptons. In particular the two leptons with the higher transverse momentum are used as candidates for the Z decay. While dressing a lepton, all the photons within a $\Delta R < 0.1$ cone are merged with the electron or the muon and the four-momentum used to reconstruct the Z boson is the sum of all the momenta of the particles inside the cone. This is a simple procedure which allows a fast dressing of the leptons. However, this approach does not take into account the following situation. If two leptons are very close one to another, i.e. their angular separation is lower than 0.1, the photons are associated with both the leptons, disregarding which of the two is closer to the specific photon. This can cause an overcounting of the photons as they can be associated to two different particles. There is a small probability that this situation occur between the two most energetic leptons; however, it could happen that one of the leptons used to reconstruct the Z is close to another soft lepton. To understand what is the extent of the problem, the standard method has been faced with an alternative procedure. One possibility would be to bind the photons to the closer lepton. This would be limiting as it is much more likely that a photon is emitted by a energetic lepton rather than by a less energetic particle close to it. It was therefore decided to apply a sort of anti-kT algorithm; when a photon is found within an angular distance of 0.1 from a lepton, an effective distance between the photon and the lepton is defined:

$$d_{pl} = \frac{\Delta R_{pl}}{p_{t,l}}, \tag{5.4}$$

where ΔR_{pl} is the angular distance between the lepton and the photon and $p_{t,l}$ is the lepton transverse momentum. This distance is evaluated between the photon and all the leptons of the particle final-state and the photon is merged with the specific lepton only if their distance is the smallest. In this way the photons are merged with electrons or muons taking into account both the distance and energy information.

In the following pages, some of the most relevant distributions, for the case of at least 1 b-jet ($n_b \geq 1$) and at least 2 b-jets ($n_b \geq 2$), are presented. To-

gether with the distributions there is the ratio, bin to bin, between the cross sections obtained with the new approach and standard methodology. The presented results are obtained using the already cited Monte Carlo generators, namely POWHEG+Pythia8, MadGraph5+Pythia6 and MadGraph5_aMC@NLO+Pythia8. The statistical error is reported for all the plots.

It is possible to note that there is no substantial change in the number of events. All the ratio distributions present mean values centered on 1. The discrepancy is of the order of 0.01% for the inclusive case and about the 0.1% for the $n_b \geq 2$ case. Looking to the bin to bin ratio, it is noticeable that some bins present a decrease of the number of entries while others present an increase. The first effect is probably due to the reduction of the transverse momentum of some dressed leptons below the 20 GeV threshold, which causes the event to be vetoed. The second effect is plausibly linked to the fact that the procedure reduces the dressing cone opening angle and therefore some photons can be clustered to the jets, increasing their p_T. Also a change of direction of the leptons could cause the event to be accepted or rejected.

In conclusion it can be stated that the two methods are almost equivalent and that the mentioned effect is almost negligible.

MC generator	Z+b ratio (anti-kT/Standard)	Z+2b ratio (anti-kT/Standard)
POWHEG	1.0003 ± 0.0002	1.0000 ± 0.0001
MadGraph5	1.0003 ± 0.0002	1.0016 ± 0.0004
MadGraph5_aMC@NLO	1.0002 ± 0.0001	1.0013 ± 0.0004

Table 5.3: Fraction of events lost or gained by applying the anti-kT method for the lepton dressing.

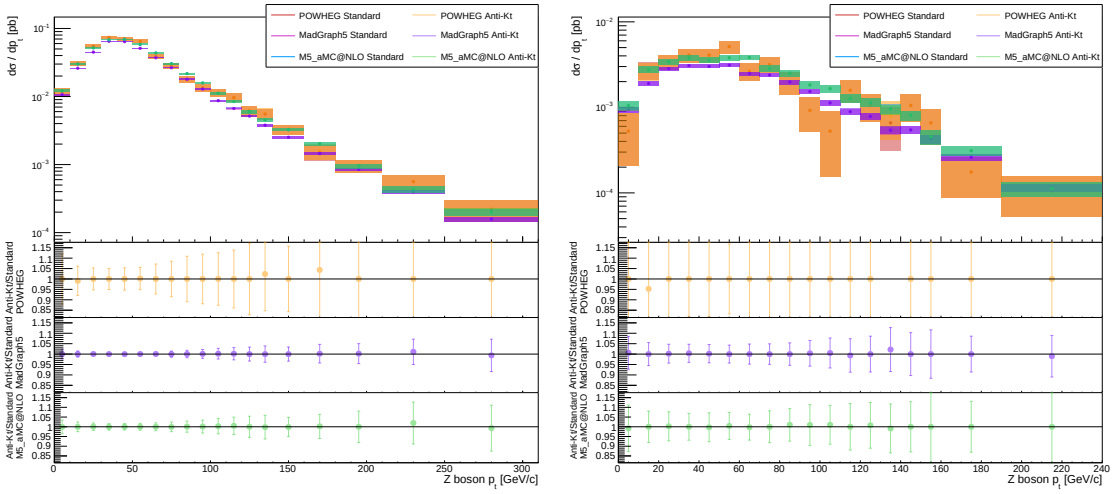

Figure 5.7: The differential $Z+b$ inclusive cross section (left) and the $Z+2b$ cross section (right) as a function of the Z boson transverse momentum. The predictions of different generators are compared when the lepton dressing is performed with the standard method or with an anti-kT technique.

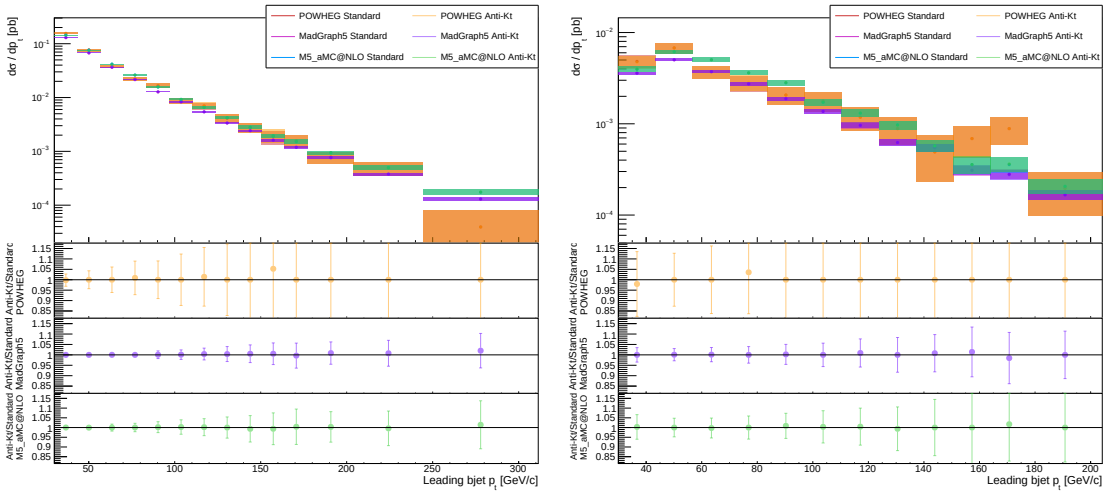

Figure 5.8: The differential $Z+b$ inclusive cross section (left) and the $Z+2b$ cross section (right) as a function of the leading b-jet transverse momentum. The predictions of different generators are compared when the lepton dressing is performed with the standard method or with an anti-kT technique.

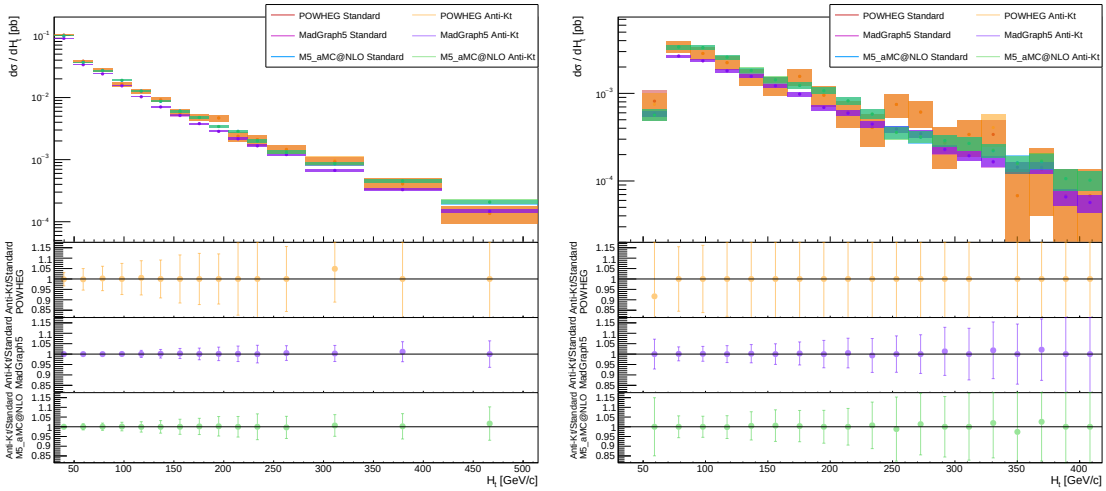

Figure 5.9: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the H_t variable. The predictions of different generators are compared when the lepton dressing is performed with the standard method or with an anti-kT technique.

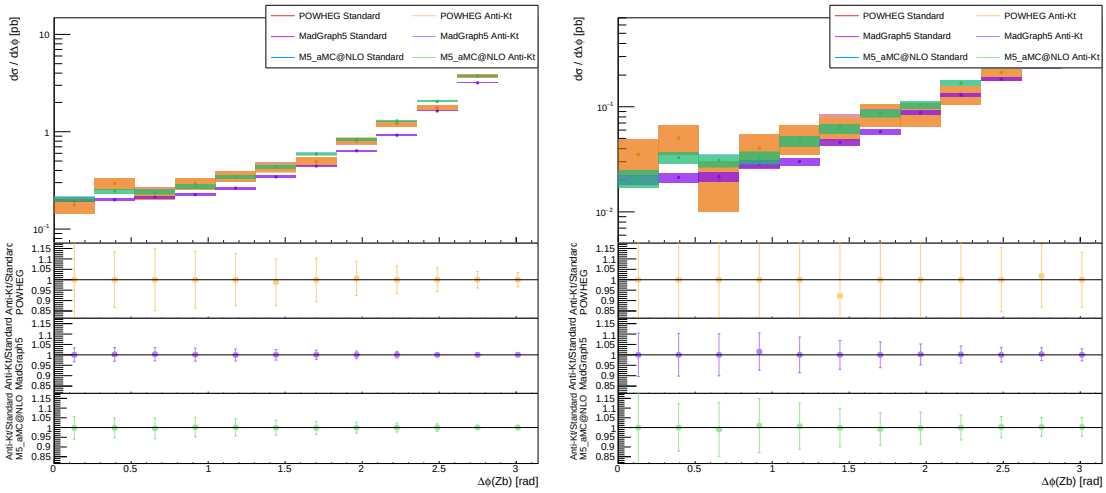

Figure 5.10: The differential $Z+b$ inclusive cross section (left) and the $Z+2b$ cross section (right) as a function of the $\Delta\phi$ between the Z boson and the leading b tagged jet. The predictions of different generators are compared when the lepton dressing is performed with the standard method or with an anti-kT technique.

5.4.2 Semileptonic decays

As mentioned in Section 5.1.1, a hadron may decay through the electroweak interaction, producing a lepton, the corresponding neutrino and additional hadrons. Examples of semileptonic decays are:

$$K^0 \to e^- + \overline{\nu}_e + \pi^+, \tag{5.5}$$

$$B \to e^- + \overline{\nu}_e + D^*. \tag{5.6}$$

Semileptonic decays involving B hadrons, as the last one listed above, are particularly interesting for the purpose of this analysis. The electroweak decays of a B hadron [86] are mediated by a W boson and they occur about the 10% of the times. The decay rate depends on the modulus of the CKM matrix element corresponding to the quark transition, $|V_{cb}|$ in the case of the aforementioned decay.

Regarding the event selection performed in this work, the problem is that a final state lepton coming from the semileptonic decay of an hadron could be selected as a candidate for the Z boson decay. In this case, in addition to reconstructing a fake Z boson, also the b-jet associated to the semileptonic decay will be distorted and eventually lost. As already stated, in fact, the particles associated to the dressed leptons are removed from the particle collection before that the jet algorithm is applied. Therefore a lepton coming from an hadron decay would be removed causing a mis-reconstruction of the jet associated to the same hadron.

To dampen this problem a safety cut is applied: all the jets whose angular separation from the leptons is lower than $\Delta R < 0.5$ are not considered. Thanks to this cut, the jet produced by the hadronization of the B hadron, which decay in leptons, should be removed from the final jet collection. Therefore the number of events in which the jets are deformed should be small and the selection should be protected against this kind of occurrences. The downside is that the removal of a single jet may lead to recognize events with two jets as single-jet events.

Since the selection on the generated event has as objective to reproduce the experimental observables, in the standard analysis there is no control carried out to avoid the occurrence of this events. Nevertheless it is interesting to understand what is the fraction of events in which the leptons with the highest transverse momentum come from a semileptonic decay. To determine the percentage of events affected by this problem, an alternative Rivet analysis has been written. In this case whenever one of the two leptons with the highest p_T comes from the semileptonic decay of a hadron, the event is vetoed. This is performed looking for hadronic PIDs in the history of the leptons. All the rest of the program remains unchanged.

Also in this case some relevant distribution in the case of $n_b \geq 1$ and $n_b \geq 2$ are reported in the following pages. The bin to bin ratio is again evaluated. The events have been simulated using POWHEG+Pythia8, MadGraph5+Pythia6 and MadGraph5_aMC@NLO+Pythia8. The results are presented in Table 5.4. The generators provide analogue results: both for the inclusive case and for the production of at least two jets, the fraction of lost events is lower than 0.5%. This is a confirmation of the fact that the $\Delta R = 0.5$ safety cut does an appropriate work and the b-jets overlapped to the leptons are effectively removed. The fraction of lost events does not present an evident dependence in terms of the reported observables. Someone could notice the fact that, even if we are vetoing events, some bins of the ratio plot of the Madgraph5_aMC@NLO Monte Carlo present an increase of the entries number. This is due to the reweighting technique used by the generator: if a negative weighted event is vetoed, the bin height increases.

MC generator	Z+b ratio (No decay/Standard)	Z+2b ratio (No decay/Standard)
POWHEG	0.997 ± 0.002	0.996 ± 0.002
MadGraph5	0.999 ± 0.001	0.998 ± 0.001
MadGraph5_aMC@NLO	0.998 ± 0.001	0.999 ± 0.001

Table 5.4: Fraction of events lost by vetoing events in which the Z boson leptons come from a semileptonic decay.

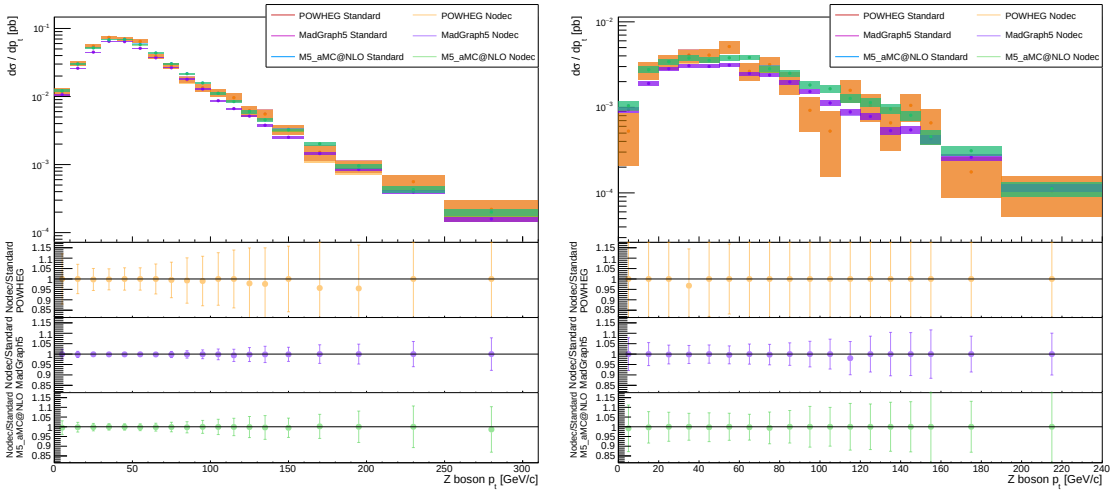

Figure 5.11: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the Z boson transverse momentum. The predictions of different generators are compared when the lepton coming from the semileptonic decay of hadrons are accepted or vetoed.

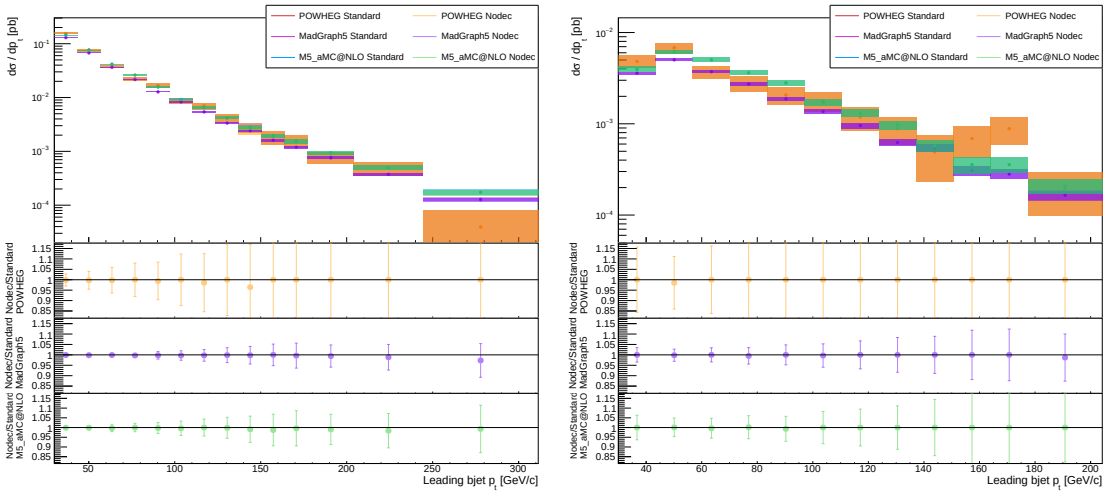

Figure 5.12: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the leading b-jet transverse momentum. The predictions of different generators are compared when the lepton coming from the semileptonic decay of hadrons are accepted or vetoed.

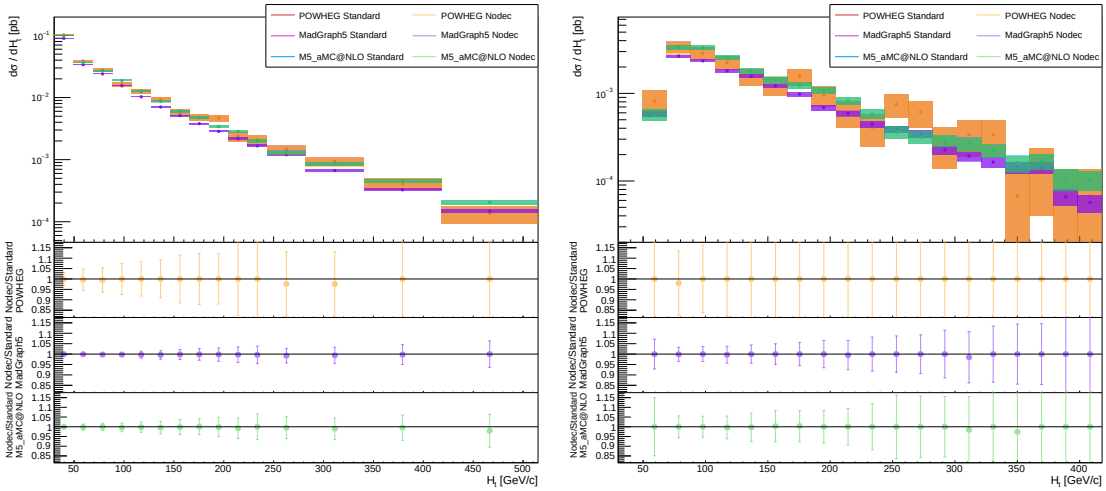

Figure 5.13: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the H_t variable. The predictions of different generators are compared when the lepton coming from the semileptonic decay of hadrons are accepted or vetoed.

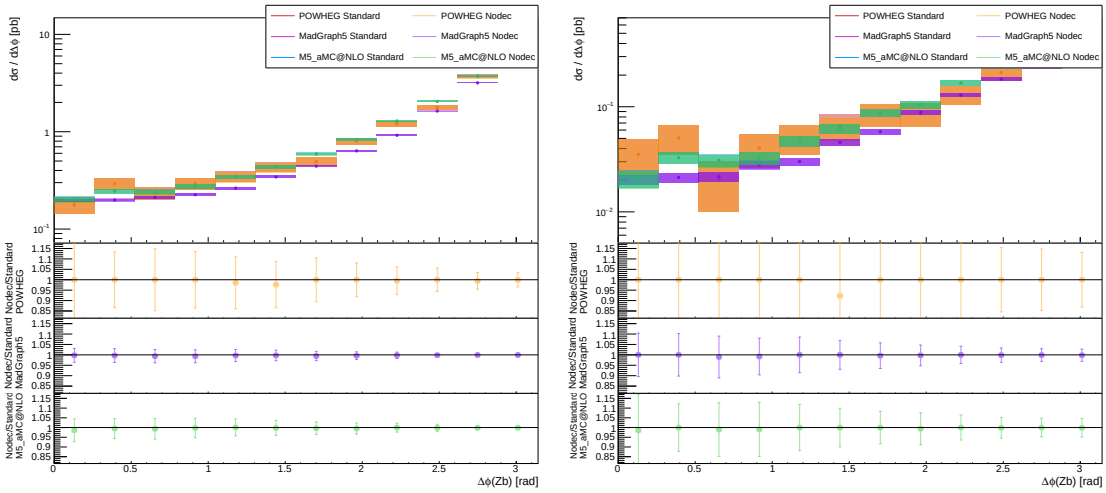

Figure 5.14: The differential $Z+b$ inclusive cross section (left) and the $Z+2b$ cross section (right) as a function of the $\Delta\phi$ between the Z boson and the leading b tagged jet. The predictions of different generators are compared when the lepton coming from the semileptonic decay of hadrons are accepted or vetoed.

5.4.3 Jet clustering

The jet clustering is not performed over the entire final state particle collection. The particles used to feed the anti-kT algorithm are all those which have not been used during the lepton dressing with the exception of the invisible final state, the neutrinos. The decision to exclude the neutrinos is due to the fact that they are not measured directly. Eventually the neutrinos could have been considered and clustered to the jets. Their contribution leads to a general raising of the energy, and therefore of the transverse momentum, of the jets. In particular there is a gain of about the 10% of events due to the fact that more jets pass the p_T threshold (see Figure 5.15). Nevertheless to compare the experimental data with the theoretical prediction a correction factor for the measured jet energy should have been considered and therefore it has been preferred not to modify the experimental data and to keep the procedure as simple as possible.

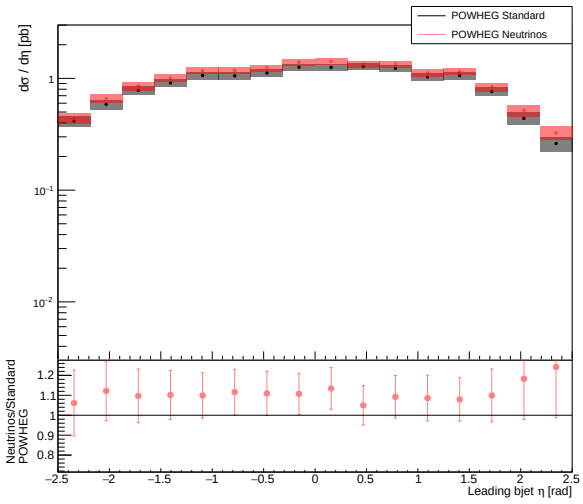

Figure 5.15: The differential $Z + b$ inclusive cross section as a function of the leading b-jet η. The prediction of the POWHEG generator is compared when the neutrinos are used to reconstruct the jets and when they are not.

Another issue related to the jet clustering is the fact that the removal of some particles from the final state collection may cause a deformation of the reconstructed jets. In fact, removing the dressing lepton associated particles, a sort of hole is formed in terms of the angular distribution of the particles and a jet, which should have included those particles, will be distorted. This effect should not be overly relevant because the safety cut $\Delta R = 0.5$ should remove the majority of these malformed jets as they are very close to the dressed leptons. It is possible to assess the importance of this effect by not removing the particles before the

clustering and by relying on the safety cut for their removal. The expected result is that the deformed jets should come closer to the leptons and therefore a small percentage of events could be removed. Also if one of the two leptons used for the Z reconstruction is a decay product of the semileptonic decay of an hadron, with the standard procedure, it is possible that the associated jet direction would be modified such as to keep the jet in the final collection. Not removing the particles, the lepton associated jet would be exactly reconstructed and hopefully removed.

In Table 5.5 the results obtained with this alternative procedure are presented. Again the approach has been tested using POWHEG+Pythia8, MadGraph5+Pythia6 and MadGraph5_aMC@NLO+Pythia8 and some distribution with the respective ratio plot and the statistical errors are disposable in the next pages. As expected there is a reduction of the event number: this is more evident in the case of $n_b \geq 2$ when there is an higher probability of having one jet deformed. Looking at the distributions it can be noticed that the greatest fraction of events lost involves jets close to the Z boson direction (small $\Delta\phi(Zb)$) and with a small p_T. This is due to the fact that a jet with a small momentum undergoes more easily to a change of direction.

MC generator	Z+b ratio (No removal/Standard)	Z+2b ratio (No removal/Standard)
POWHEG	0.992 ± 0.003	0.982 ± 0.004
MadGraph5	0.991 ± 0.003	0.977 ± 0.005
MadGraph5_aMC@NLO	0.993 ± 0.003	0.985 ± 0.004

Table 5.5: Fraction of events lost by not removing the dressed lepton associated particles before the jet clustering.

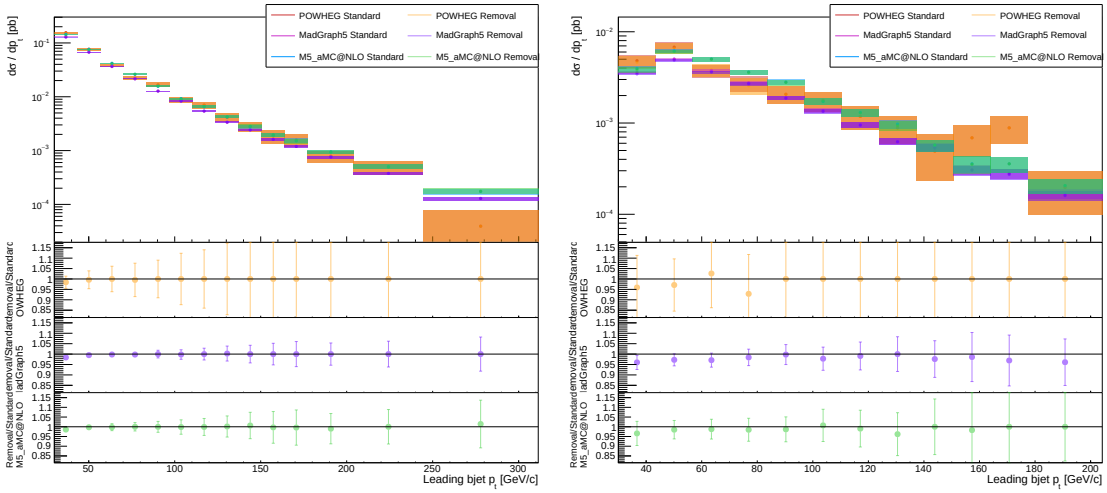

Figure 5.16: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the leading b-jet transverse momentum. The predictions of different generators are compared when the particles are not removed before the jet clustering or not.

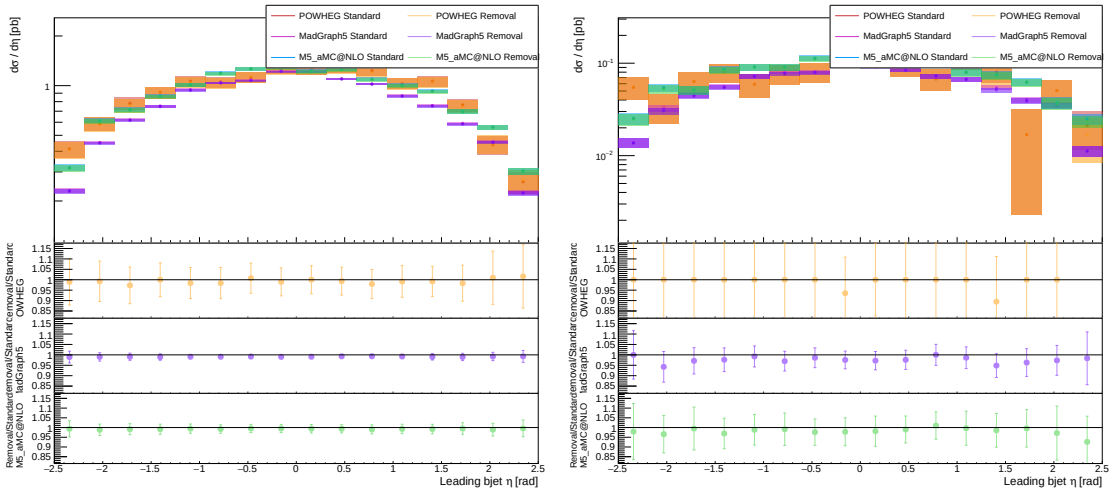

Figure 5.17: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the leading b-jet η. The predictions of different generators are compared when the particles are not removed before the jet clustering or not.

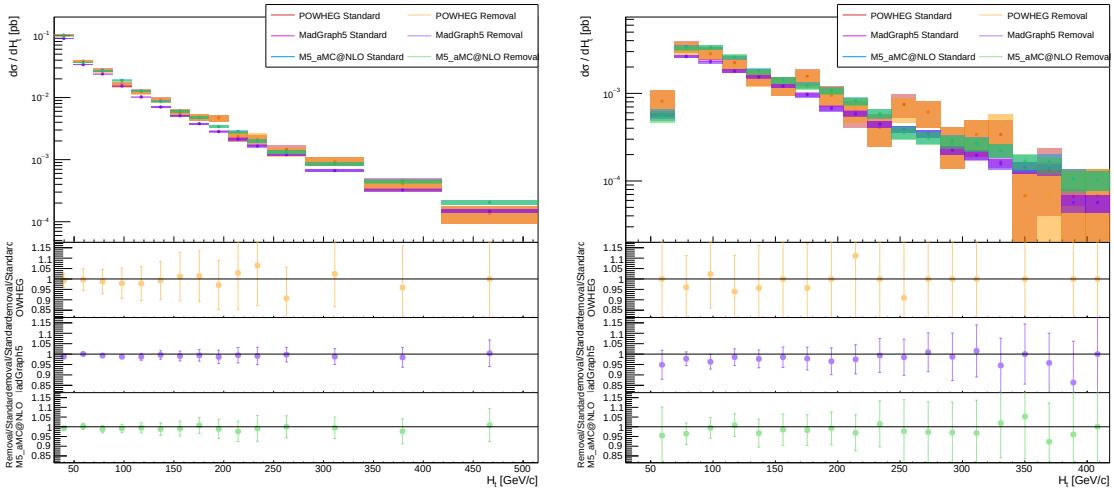

Figure 5.18: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the H_t variable. The predictions of different generators are compared when the particles are not removed before the jet clustering or not.

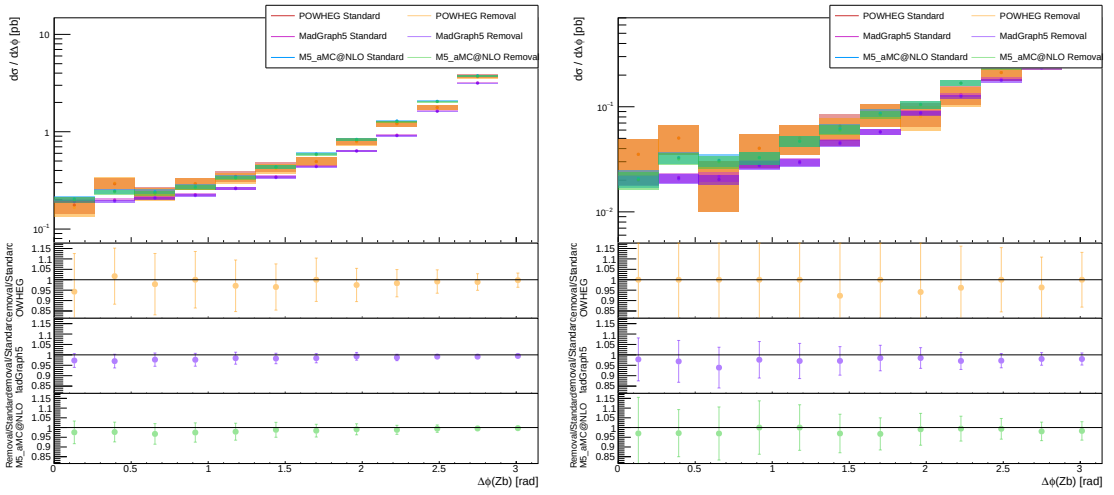

Figure 5.19: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the $\Delta\phi$ between the Z boson and the leading b tagged jet. The predictions of different generators are compared when the particles are removed before the jet clustering or not.

5.4.4 The b-tagging

The ability of b-tag heavy flavor jets produced by the hadronization of heavy particles is of primary importance to identify the signatures of new physics processes. There are several possible definitions of a flavored jet and different techniques to perform their identification. In this Section a comparison between two of the most common techniques is presented. In the standard analysis the procedure to b-tag a jet is the following. Firstly the last B-hadron in the history of every constituent of the jet is selected. The particle is a B-hadron if its PID contains a 5, i.e. a b-quark. Here "last B-hadron" means the youngest particle between the ancestors of the constituents. This is the last b-particle in chronological order, i.e. a hadron which does not have any other B-hadron as daughters. After all the B-hadrons have been identified in this way, the angular distance between their direction and the jet direction is evaluated. If this distance is smaller than 0.5, the jet is b-tagged, otherwise it is not.

Another approach is possible, known as the "ghost b-tagging". In this case the first step is to identify all the B-hadrons which are the mothers of the final-state particles. They have to respect all the requirements mentioned above, in particular only the particle without other B-hadron daughters are selected. At this point their momentum is rescaled for a small factor (10^{-18} is the default value) and they are added to the final-state particle collection. They are then clustered together with the other jet constituents by the anti-kT algorithm. Since they are very soft, they will not contribute in any way to the jet four-momentum and the resulting collection will be practically identical to the standard one. For this reason they are known as "ghosts". At this point the jet flavor is determined identifying the presence of a ghost between the jet constituents. In this way it is impossible that more than one jet is associated to the same ancestor.

The same procedure is generically applicable to assign the correct flavor to any jets, but in the case of this analysis only the B-hadrons have been used as ghosts. One important thing to point out is that, depending on the kind of algorithm used for the jet clustering, the presence of ghosts may spoil the infrared safety of the procedure. This is the case of the Cambridge/Aachen and kT clustering algorithms, where it is possible that ghosts are merged together in pseudo-jet before being paired with the actual jet constituents. The anti-kT algorithm stays infrared safe.

As usual in the following pages the results obtained using the two method are presented both for the inclusive case and for the case of at least two b-jets. The event generator used are POWHEG+Pythia8, MadGraph5+Pythia6 and MadGraph5_aMC@NLO+Pythia8. The fraction of events lost or gained using the ghost b-tagging is summarized in Table 5.6. The results provided by the three generators are very similar: for the $n_b \geq 1$ case the difference between the

two method is about 1%, while the variation for the case of at least two b-jet is smaller ($\leq 1\%$). There are not evident variations in the shape of the b-jet related observables both for the $n_b \geq 1$ and for the $n_b \geq 2$ case.

The slight gain in terms of number of events is explainable considering the fact that the anti-kT algorithm cluster the particles inside a cone with an opening angle equal to 0.5. The angle is evaluated with respect to the direction of the pseudojet which is being created, but at every step of the clustering process this direction changes, therefore the cone border moves together with the jet. Due to this behavior the jet might be tagged even if the distance between the b-hadron and the jet is slightly larger than $\Delta R = 0.5$.

MC generator	Z+b ratio (Ghost/Standard)	Z+2b ratio (Ghost/Standard)
POWHEG	1.013 ± 0.004	1.010 ± 0.003
MadGraph5	1.008 ± 0.003	1.005 ± 0.002
MadGraph5_aMC@NLO	1.009 ± 0.003	1.002 ± 0.001

Table 5.6: Fraction of events lost or gained by applying the ghost b-tagging procedure to identify the jet flavor.

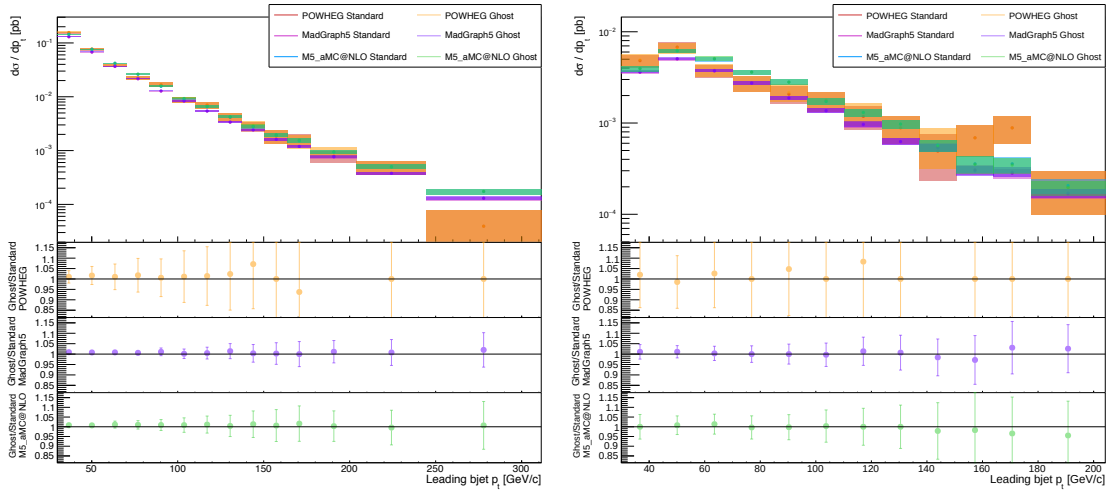

Figure 5.20: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the leading b-jet transverse momentum. The predictions of different generators are compared when the b-tagging is performed with the standard method or with the ghosting procedure.

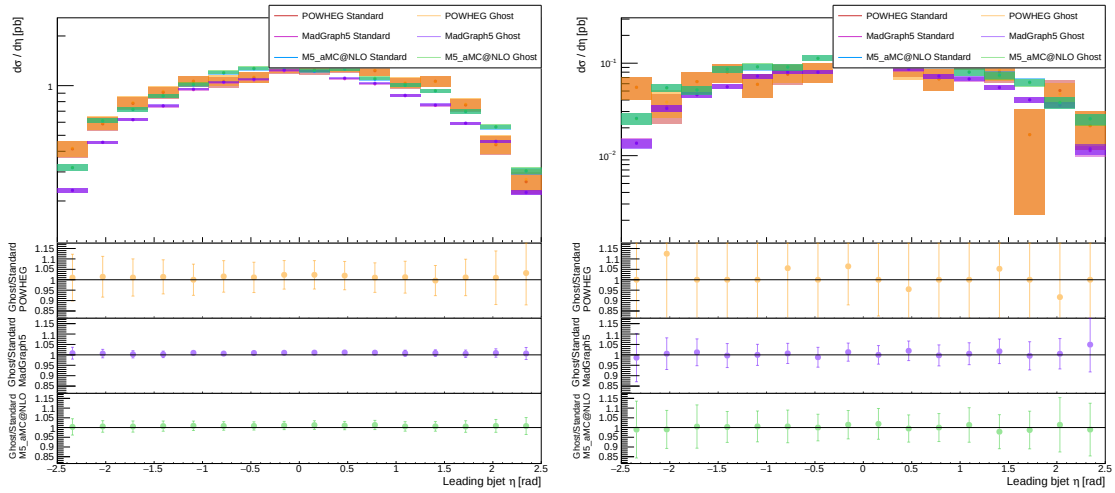

Figure 5.21: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the leading b-jet η. The predictions of different generators are compared when the b-tagging is performed with the standard method or with the ghosting procedure.

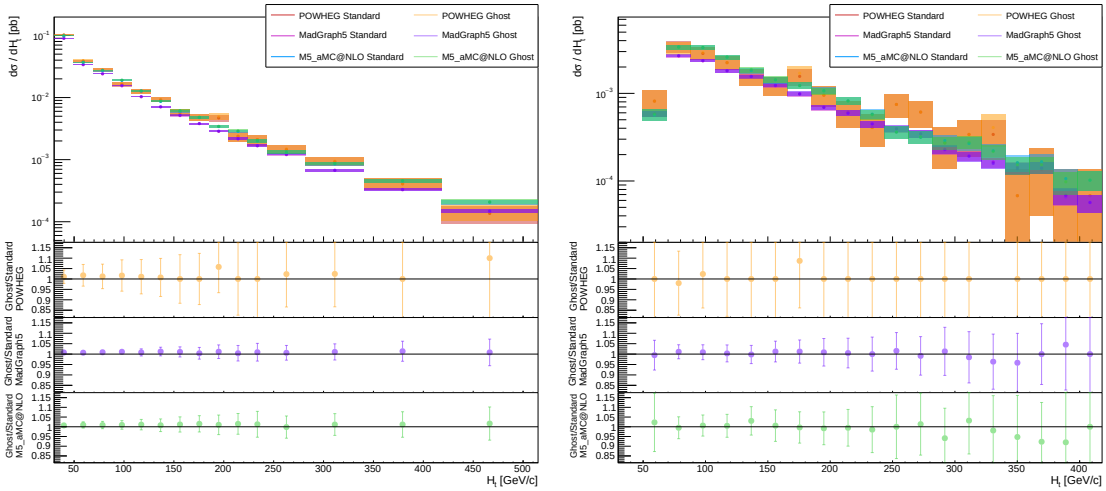

Figure 5.22: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the H_t variable. The predictions of different generators are compared when the b-tagging is performed with the standard method or with the ghosting procedure.

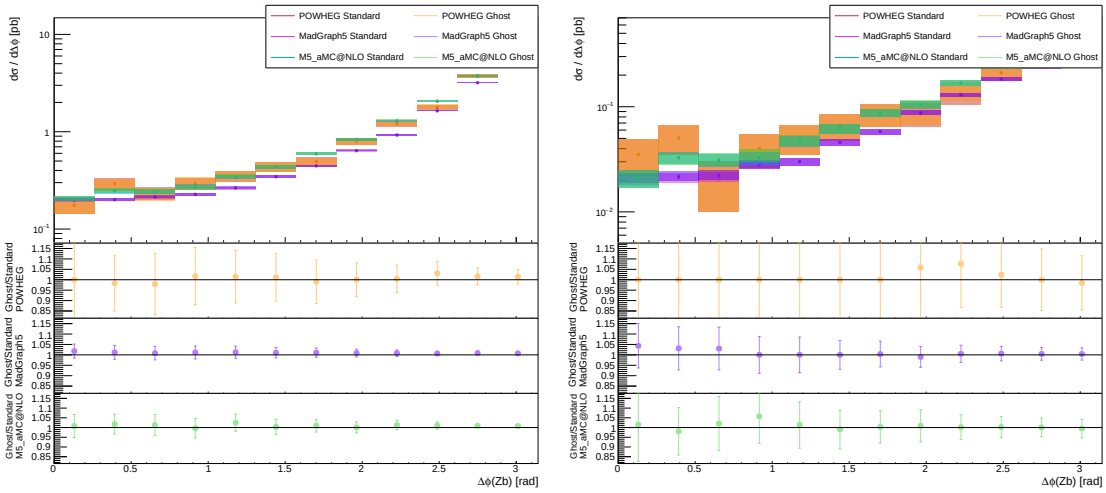

Figure 5.23: The differential $Z + b$ inclusive cross section (left) and the $Z + 2b$ cross section (right) as a function of the $\Delta\phi$ between the Z boson and the leading b tagged jet. The predictions of different generators are compared when the b-tagging is performed with the standard method or with the ghosting procedure.

5.4.5 Summary

Different approaches for the event selection have been discussed in this Chapter. The results clearly show that the standard method used for the selection is stable: the variations in terms of percentage of selected events are small in all the cases under analysis. Both for the inclusive case and for the case of at least two b-jets the difference is generally smaller than the 1% and in most of the cases the associated statistical error is much larger. Our group is actually considering the implementation of the discussed selection techniques in the official analysis used for the study of the $Z + b$ associated production. In particular the standard lepton dressing will be probably substituted to avoid the eventual overcounting of photons, even if the occurrence of this kind of events is very small. Also the use of the ghost b-tagging algorithm is under examination and it will be likely implemented soon.

Chapter 6

The $Z+b$ differential cross sections

Comparing the experimental and simulated data, even small deviations in the statistical analysis can be vital to identify signals of new physics or, conversely, to exclude some theoretical models. With the progression of the detector technology and the measurement techniques, the uncertainties on the measured data are becoming smaller, hence the uncertainties related to the simulated data and theoretical models are increasingly important. For this reason modern Monte Carlo generators include high-order corrections in perturbation theory, so as to have smaller uncertainties and a better description of the data.

In the previous pages, the Monte Carlo generators used to simulate the $Z + b$ final state have been presented and the event selection procedure at particle level has been described. A comparison between the differential cross sections of the $Z + b$ production with the theoretical predictions, as a function of the different observables, has already been presented in previous works [87, 88]. Nevertheless a complete analysis of the theoretical uncertainties related to the Monte Carlo generators was not included. The goal of this thesis is precisely to provide a comparison between the NLO predictions of the available Monte Carlo generators, estimating the systematic uncertainties related to the scale and PDF variations in the $Z+b$ final state. In this Chapter the role of the factorization and renormalization scales in MC generators is succinctly overviewed and the available PDF sets for the event generation are presented. The methodology used to evaluate the theoretical uncertainties is described, presenting, at the end, the comparison between data and theory and discussing possible improvements to the measure.

6.1 Monte Carlo parameters

6.1.1 The renormalization scale

In QCD and other quantum field theories, ultraviolet (UV) divergences can arise during the evaluation of the perturbation theory integrals. These divergences originate from Feynman diagrams made by one or more closed loops of virtual particles, since the momentum of these particles may tend to infinity. The ensamble of techniques used to solve this problem, in the calculation of physical quantities, is called "Renormalization". The first step to obtain a renormalized physical theory is to define an energy scale, the renormalization scale μ_R, which can be interpreted as an energy cut-off. Thanks to this cut-off the divergent integrals are splitted in two parts: a finite integral which contains all the radiative correction up to the renormalization energy scale, and an infinite part that the theory predicts to be canceled by some counter-terms that originate from the particle propagators. In most Monte Carlo generators the value of the renormalization scale is set by default to $\mu_R = \sum_i E_i/2$, where the E_i is the transverse energy of a final state particle i.

6.1.2 The factorization scale

As already mentioned (Section 2.2.2), the factorization scale is a collinear cut-off and it can be interpreted as the resolution used for the hadron identification. When evaluating the cross section of a process, both the PDFs $f_{i/h}(x, \mu_F^2)$ and the hard scattering cross section $\sigma^n_{i,j\to a+b}(x_1 x_2 s, \mu_r^2, \mu_F^2)$ depend on the factorization scale. This dependence is strong for a low-order approximation of the scattering process and it becomes weaker when dealing with higher orders of the perturbation theory. By default the factorization scale value is set equal to the renormalization one.

6.1.3 The PDFs set

The determination of the parton distribution functions of the proton is critical both to the precision phenomenology of the Standard Model and to new physics measurements. In recent years there have been significant developments in this regard, which led to a deeper knowledge of the PDFs and to quantify the related uncertainties. The experimental uncertainties related to the PDFs depend on the specific set which has been chosen. Various PDF sets are available at the moment: HERAPDF [27], CTEQ/CT (CTEQ6.6 [28], CT10 [30]), NNPDF (NNPDF3.0 [29]), MMHT [31] and many others. They mainly differ because of the included data set, the uncertainty estimator, and the parton parametrization, i.e. the number of free parameters.

In terms of data sets, the choice that might seem more intuitive is to use the largest data set possible. This choice allows to maximize the available information but at the same time it could lead to inconsistencies due to the fact that data sets coming from different experiments may be partly incompatible. For this reason different groups have made different choices: the HERAPDF includes only the HERA deep inelastic scattering data, the NNPDF and the CT10 uses the HERA data together with the LHC and Tevatron datasets respectively. MMHT uses the LHC, Tevatron and Hera data.

Regarding the statistical treatment, the PDF can be determined using the Hessian or the Monte Carlo method. The Hessian approach consists in the minimization of a χ^2 function. Given a specific data point, x_i, and the corresponding theoretical prediction, $\overline{x}_i$, the following χ^2 function is defined:

$$\chi^2 = \frac{1}{N}\sum_{i,j}(x_i - \overline{x}_i)V_{i,j}(x_j - \overline{x}_j), \tag{6.1}$$

where N is the total number of data points and $V_{i,j}$ is the covariance matrix. The PDF best fit corresponds to the χ^2 minimum value, while the related uncertainties are determined increasing the χ^2 of a specific value with respect to the minimum and finding the corresponding range for each eingenvector of the χ^2 Hessian matrix:

$$H_{i,j} = \frac{\partial^2\chi^2}{\partial x_i \partial x_j}. \tag{6.2}$$

A 1σ variation is usually performed. CTEQ and MMHT use this method. The NNPDF group, instead, uses the Monte Carlo method. In this case a large number N_s of data samples are produced using a Monte Carlo generator. The simulated data are built in such a way that the mean of a specific point $\overline{x}_i$ is equal to the central value of the data point x_i. Analogously to the Hessian method a χ^2 function is defined and, through its minimization, for each sample, a PDF_i is found. The final PDF is defined using the central values, the variances and the covariances of the PDF_i.

6.2 Uncertainties computation

6.2.1 Scale uncertainties

The renormalization and factorization scale uncertainties are usually determined taking into account a variation of a factor two, upwards and downwards, of the μ_r and μ_f default values. A total of nine pairs of scale values are evaluated. Given a specific distribution, the theoretical error associated to a point is then defined

as the maximum variation from the central value ($\mu_r = 1, \mu_f = 1$). In this work, the values corresponding to the two most extreme variations, ($\mu_r = 2, \mu_f = 2$) and ($\mu_r = 0.5, \mu_f = 0.5$), are not taken into account. In general the final error associated to each point will be asymmetric.

6.2.2 PDF uncertainties

The error computation for the PDF depends on the method used for the creation of the specific set.

In general, when dealing with an Hessian PDF, both the central set and the error sets are provided. Each error set corresponds to a variation of 1 sigma (or 90% C.L.) with respect to the central value of a specific eingenvector of the Hessian matrix. The variation is obviously performed both in the positive and in the negative direction, therefore the number of error sets is equal to double the eingenvector number. The deviation of a variable O from its central value O_0 can be determined using the following formulas:

$$\Delta O^+ = \sqrt{\sum_i^N \left[max\left(O_i^+ - O_0, O_i^- - O_0, 0\right)\right]^2}, \tag{6.3}$$

$$\Delta O^- = \sqrt{\sum_i^N \left[max\left(O_0 - O_i^+, O_0 - O_i^-, 0\right)\right]^2}, \tag{6.4}$$

where O_i^- (O_i^+) is the O value using the PDF corresponding to the positive (negative) increment of the i eingenvector. N is the total number of eingenvectors. The values can be added in quadrature because of the orthonormal nature of the eingenvectors. Also a symmetric error can be defined:

$$\Delta O = \frac{1}{2}\sqrt{\sum_i^N \left(O_i^+ - O_i^-\right)^2}. \tag{6.5}$$

When a Monte Carlo approach is used for the PDF computation a different method must be used. Consider an observable O whose expectation value can be evaluated as the average over the O_i values corresponding to the different sample's PDF$_i$:

$$\overline{O} = \frac{1}{N_s}\sum_i^{N_s} O_i. \tag{6.6}$$

In this case, the uncertainty associated to the observable can simply be calculated as its standard deviation:

$$\sigma_O = \left[\frac{1}{N_s - 1} \sum_{i}^{N_s} (O_i - \overline{O})^2 \right]^{\frac{1}{2}}. \tag{6.7}$$

6.2.3 Strong coupling uncertainties

Selecting an appropriate α_S value is of primary importance because it is highly correlated to the parton distribution functions. There are two common approaches to decide the standard value of the strong coupling. The first is to take α_S as a fit parameter and to determine its central value through a fitting procedure, as for the other PDF parameters. Another possibility is to treat α_S as an external parameter, similarly to the mass of the heavy quarks. In this case the user can decide the specific central value. In the first case the associated uncertainty is encoded in the error sets, while in the second case the error can be evaluated changing α_S in a specific range. It must be noticed that a variation of the α_S value causes a modification of the cross section of a process both because of the direct dependence of σ on the strong coupling value and because of the parton distribution function dependence on α_S. For this reason, when it is possible, the combined PDF+α_S error is provided as the result.

If the χ^2 dependence on the parameters can be approximated as quadratic (which is true in most of the cases), the total uncertainty related to the PDF+α_S variation can be calculated as:

$$\Delta O = \sqrt{(\Delta O_\alpha)^2 + (\Delta O_{PDF})^2}, \tag{6.8}$$

where ΔO_α is the O error related to the 1-σ variation of α_S when the PDF parameters are fixed at the central values and ΔO_{PDF} is the O error related to the 1-σ variation of the PDF when the α_S value is fixed. In the following all the PDF sets used have the strong coupling fixed as an external parameter at the standard value: $\alpha_s^0 = 0.118$.

When the coupling is an external parameter the uncertainties related to a specific observable O can be determined as:

$$\Delta_\alpha^+ = O(\alpha_S^0 + \delta_\alpha^{1\sigma}) - O(\alpha_S^0), \tag{6.9}$$

$$\Delta_\alpha^- = O(\alpha_S^0 - \delta_\alpha^{1\sigma}) - O(\alpha_S^0). \tag{6.10}$$

Therefore the PDF+α_S uncertainties will be:

$$\Delta_{PDF+\alpha}^+ = \sqrt{(\Delta_\alpha^+)^2 + (\Delta_{PDF}^{\alpha^0,+})^2}, \tag{6.11}$$

$$\Delta_{PDF+\alpha}^- = \sqrt{(\Delta_\alpha^-)^2 + (\Delta_{PDF}^{\alpha^0,-})^2}. \tag{6.12}$$

A more detailed discussion about the currently available PDF sets and the related uncertainties and correlations can be found at [90].

6.3 Results

Various Monte Carlo generators are available for the simulation of $Z+$jets events. In the following analysis two NLO Monte Carlo generators have been used: MadGraph5_aMC@NLO and POWHEG+MiNLO. The MC generators have been interfaced with the Pythia8 parton shower. Both MadGraph5_aMC@NLO and POWHEG+MiNLO have encoded an event reweighting technique which allows to change some MC parameters without the need of multiple simulation runs. During the event generation a weight-vector is associated to each event. This vector contains the central weight, which corresponds to the standard parameters, and some other weights which are associated to the various PDF sets or to the scale variations. In this way, when the event analysis is performed, the choice of a different weight corresponds to a different setting. It must be noticed that the final distributions are not normalized therefore they have to be rescaled dividing by the sum of weights. The parameter variations considered for each Monte Carlo are summarized in Table 6.1.

Parameters	MadGraph5_aMC@NLO	POWHEG+MiNLO
Scale (μ_R, μ_F)	✓	✓
PDF:NNPDF3.0	✓	✓$(+\alpha_S)$
PDF:CT10	-	✓$(+\alpha_S)$
PDF:MMHT	-	✓$(+\alpha_S)$

Table 6.1: Summary of the scale and PDF variations available for the MadGraph5_aMC@NLO and POWHEG+MiNLO Monte Carlo generators.

The standard PDF set used by MadGraph5_aMC@NLO and POWHEG+ MiNLO is the NNPDF3.0 set, while the central value for the strong coupling constant is chosen at $\alpha_s = 0.118$. Both the samples allow to perform the scale variations with the PDF fixed at the NNPDF central value. Conversely all the PDF variations are performed with the standard values of the Renormalization and Factorization scales. The only PDF set available for the MadGraph5_aMC@NLO sample is NNPDF3.0 while POWHEG contains also the CT10 and MMHT sets.

In the following pages all the distributions used to characterize the $Z + b$ final state are presented together with the theoretical uncertainties. Every plot is reported 4 times. The first two versions (top left and top right) correspond to

the scale and PDF variations respectively. The errors have been computed as described in Section 6.2. As already reported the scale uncertainties have been calculated without considering the two most extreme variations of the renormalization and factorization scales. Regarding the uncertainties related to the PDF variation, the reported error consider the PDF+α_S contribution when available. The pure PDF error contribution is evaluated with the symmetric formula (see equation 6.5). Also the distributions with the pure statistical error are reported (bottom left) and with the total (statistical+theoretical) error (bottom right). The unfolded differential cross section of experimental data is superimposed to each plot, together with the associated total error (the specific systematic and statistical contributions to the data uncertainties have been described in Section 5.3). The plots are reported at the end of this Chapter:

Process	Observable	Figures
$(Z + \text{jets})$	Leading jet transverse momentum	6.1, 6.2
$(Z + \text{jets})$	Leading jet η	6.3, 6.4
$(Z + b)$	Z boson transverse momentum	6.5, 6.6
$(Z + b)$	H_t	6.7, 6.8
$(Z + b)$	Leading b-jet transverse momentum	6.9, 6.10
$(Z + b)$	Leading b-jet η	6.11, 6.12
$(Z + b)$	$\Delta\phi(Zb)$	6.13, 6.14
$(Z + b)$	M_{Zb}	6.15, 6.16
$(Z + 2b)$	Z boson transverse momentum	6.17, 6.18
$(Z + 2b)$	H_t	6.19, 6.20
$(Z + 2b)$	Leading b-jet transverse momentum	6.21, 6.22
$(Z + 2b)$	Leading b-jet η	6.23, 6.24
$(Z + 2b)$	Sub-leading b-jet transverse momentum	6.25, 6.26
$(Z + 2b)$	Sub-leading b-jet η	6.27, 6.28
$(Z + 2b)$	$\Delta R(bb)$	6.29, 6.30
$(Z + 2b)$	$\Delta\phi(bb)$	6.31, 6.32
$(Z + 2b)$	A_Z^b	6.33, 6.34
$(Z + 2b)$	M_{bb}	6.35, 6.36
$(Z + 2b)$	M_{Zbb}	6.37, 6.38

6.4 Outcome interpretation

The measured cross section for the associated production of a Z boson and b-jets seems to be compatible with the NLO predictions provided by the Mad Graph5_aMC@NLO and the POWHEG generators. Both in the case of the production of at least one b-jet and in the case of at least two b-jets, the total

uncertainties associated to the Monte Carlo samples and to the data distribution make the predictions consistent with the experimental data within two standard deviations. Nevertheless a general overestimation of the cross sections estimated by both the Monte Carlo generators is evident. In the inclusive $Z+$jets case, this overestimation is about the 10%, while in the case of the production of a Z boson and at least 1 or 2 b-jets, the theoretical cross section and the one extracted by the unfolded data show discrepancies of about the 20%. It is possible to talk of a "general" overestimation because the shape of each differential cross section is well reproduced and the discrepancy seems related to the normalization of the cross section itself. This is true for most of the distributions, although, for certain differential cross sections, also the shape of the unfolded data is not completely reproduced. This is the case of the $Z + 2b$ differential cross section in function of the leading and the sub-leading b-jet transverse momentum, where higher values of the b-jet p_T show larger discrepancies, or the case of the ΔR angular separation between the two b-jets, where a clear disagreement is present for large opening angles. The normalization issue seems proper of the NLO calculation, in fact it was not evident in the comparison with LO calculations [88].

Regarding the single contributions to the total theoretical error, it is possible to notice how the scale uncertainties are absolutely dominant. In fact the relative error associated to the scale variations is generally about the $10-15\%$, while the PDF uncertainties are much smaller ($2-3\%$). When comparing the POWHEG and MadGraph5_aMC@NLO MC generators, it is also possible to point out that the PDF errors are of the same order, while the scale uncertainties are much larger in the POWHEG case. This difference could be attributable to the smaller number of events included in the POWHEG sample, therefore an analysis with a larger number of events is indeed auspicable.

It is also interesting to note how different PDF sets, in the POWHEG sample, produce slightly different results. In particular the CT10 set is the one that shows the best agreement with the experimental data, while the standard NNPDF set seems to produce the worst result.

Further improvements on the measurement are surely possible. First of all, a larger POWHEG sample has to be produced to increase the number of events and lower the associated statistical error. It would also be interesting to include other PDF sets in the MadGraph5_aMC@NLO sample to understand how the results are influenced by different choices. A further investigation is also needed for the phase space areas which show the largest discrepancies between data and predictions, as the high p_T part of the b-jet transverse momentum spectra.

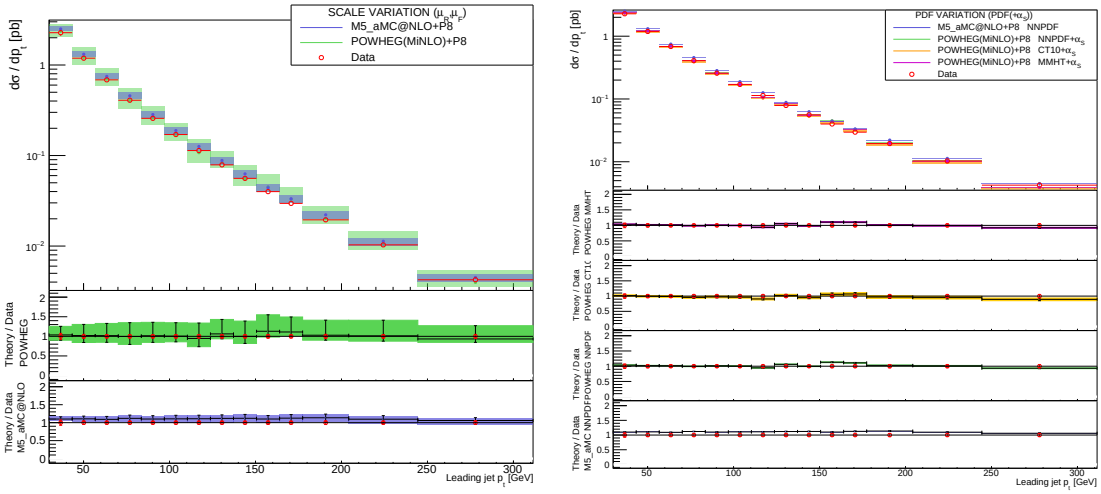

Figure 6.1: The differential $Z + jet$ inclusive cross section as a function of the leading jet transverse momentum. The theoretical errors associated to the scale variations (left) and PDF variation (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

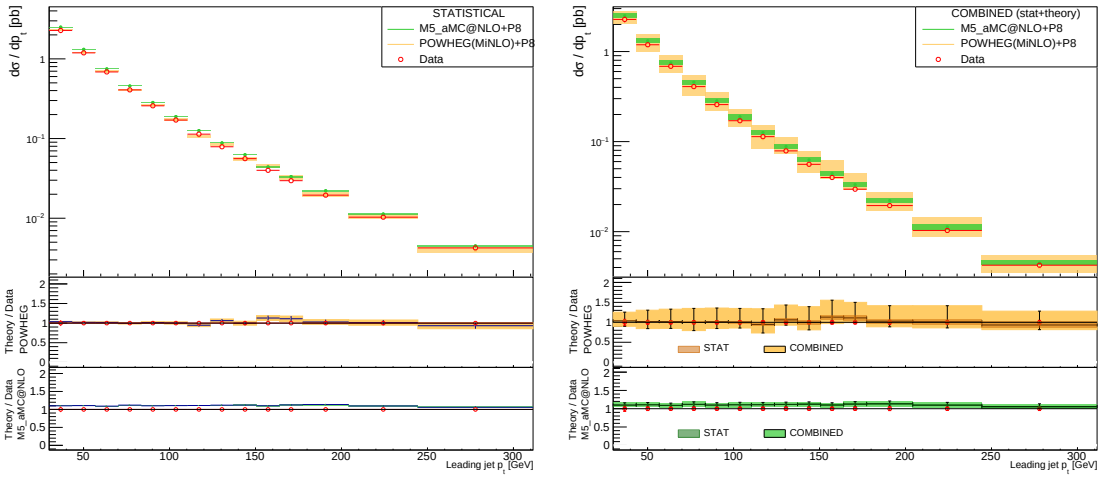

Figure 6.2: The differential $Z+jet$ inclusive cross section as a function of the leading jet transverse momentum.The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

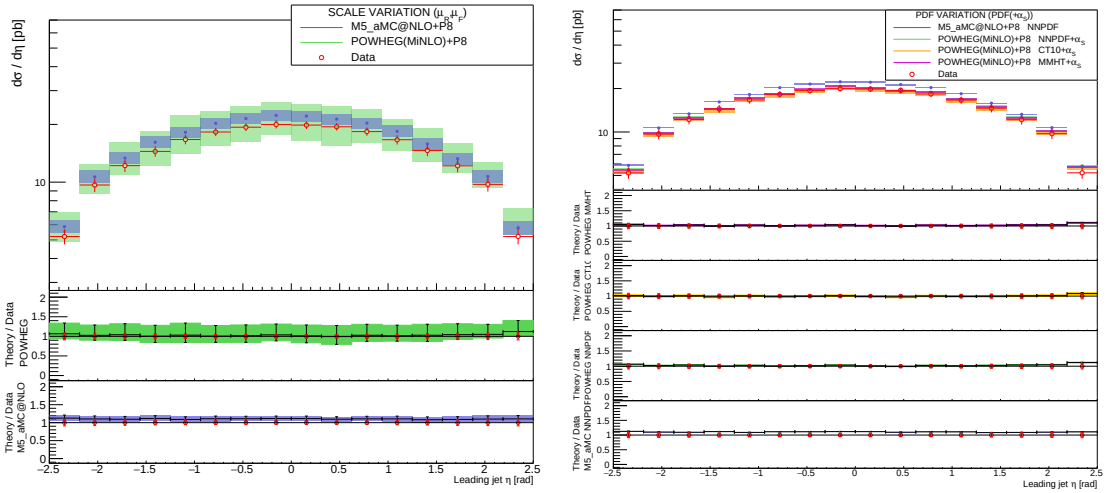

Figure 6.3: The differential $Z + jet$ inclusive cross section as a function of the leading jet η. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

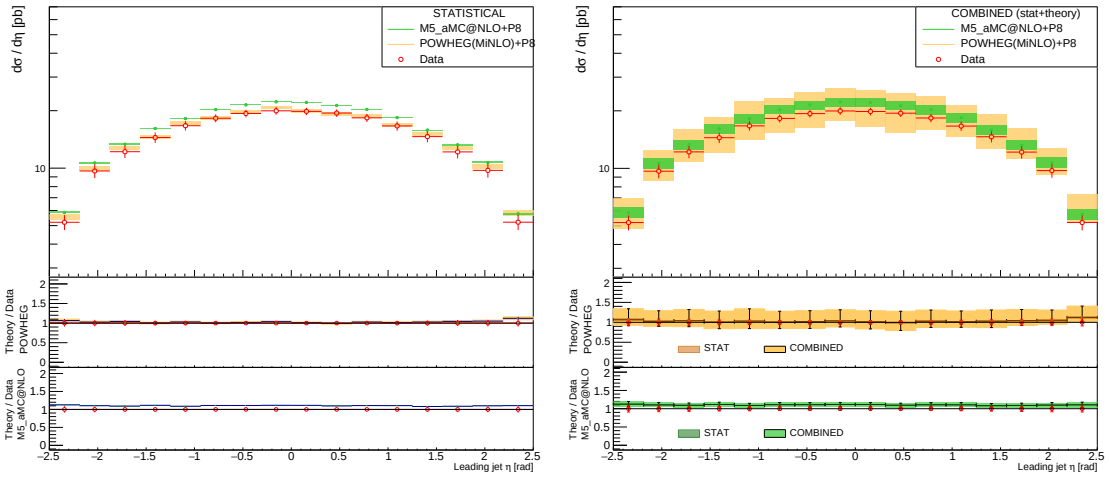

Figure 6.4: The differential $Z + jet$ inclusive cross section as a function of the leading jet η. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

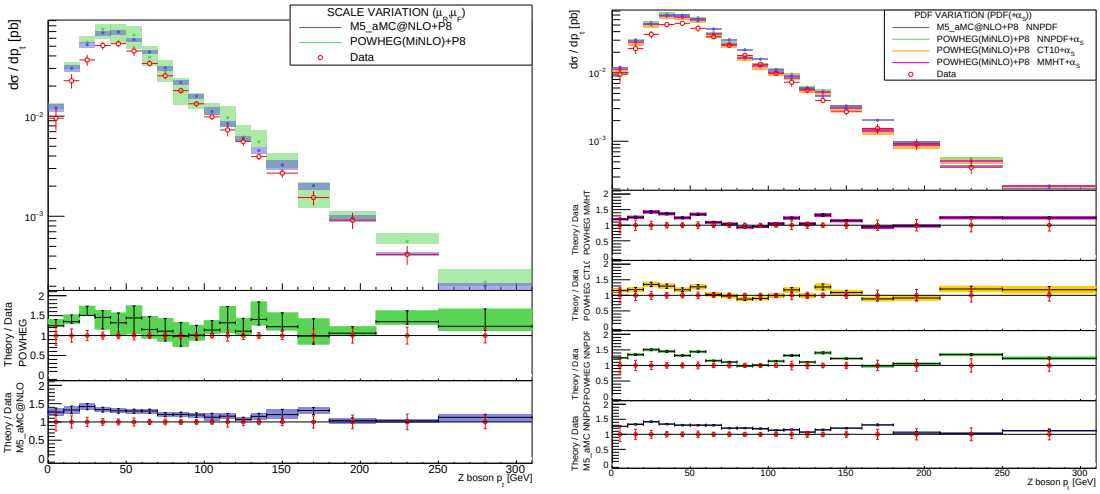

Figure 6.5: The differential $Z+b$ inclusive cross section as a function of the Z boson transverse momentum. The theoretical errors associated to the scale variation (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

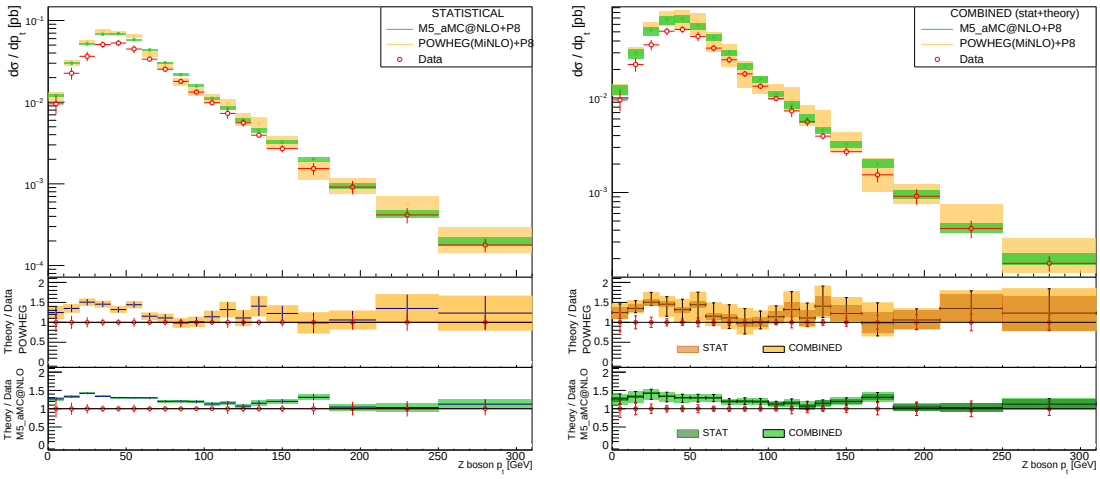

Figure 6.6: The differential $Z+b$ inclusive cross section as a function of the Z boson transverse momentum. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

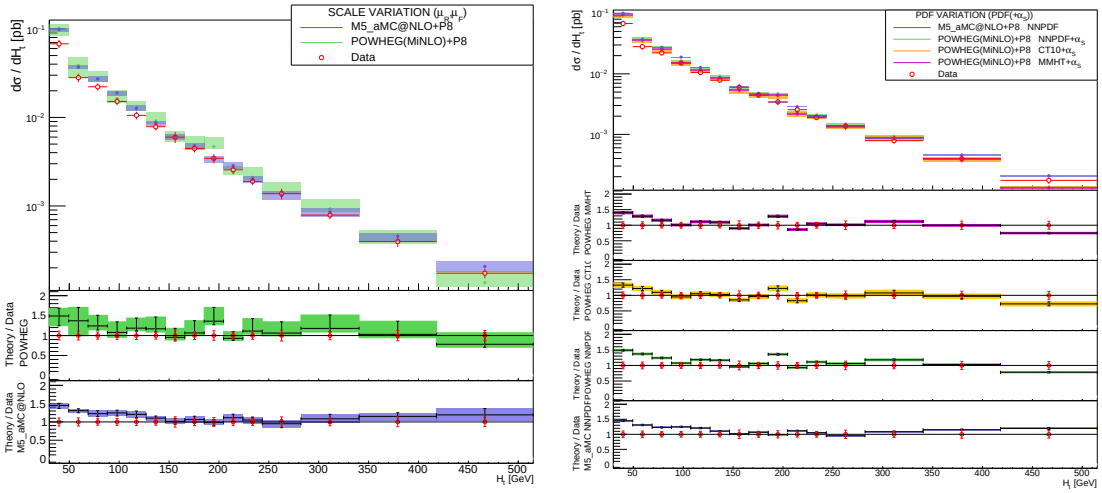

Figure 6.7: The differential $Z + b$ inclusive cross section as a function of the H_t variable. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

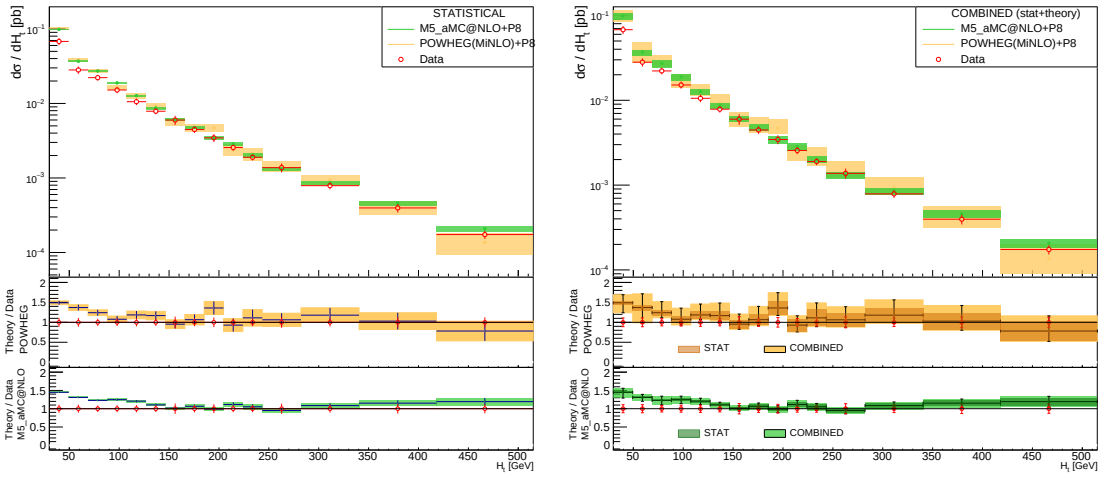

Figure 6.8: The differential $Z+b$ inclusive cross section as a function of the H_t variable. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

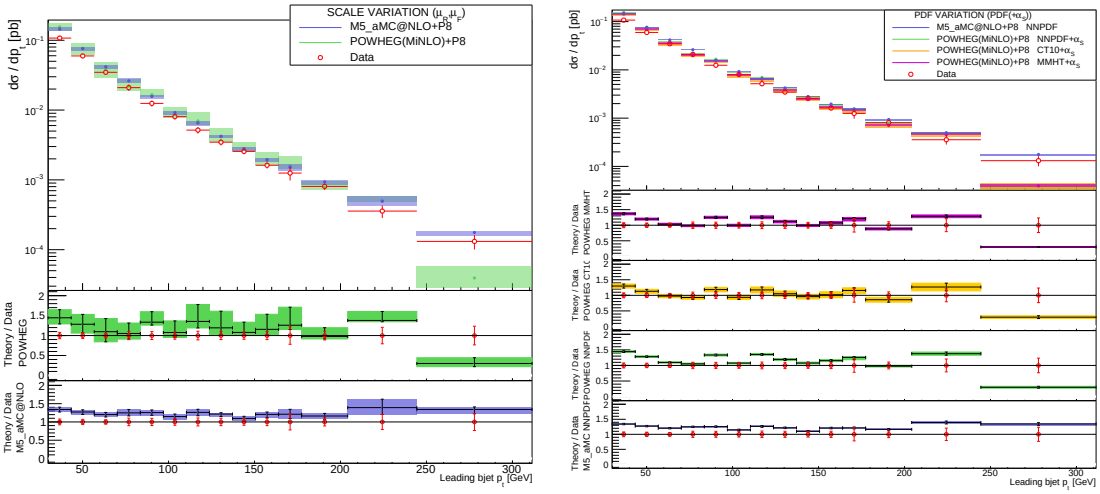

Figure 6.9: The differential $Z + b$ inclusive cross section as a function of the leading b-jet transverse momentum. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

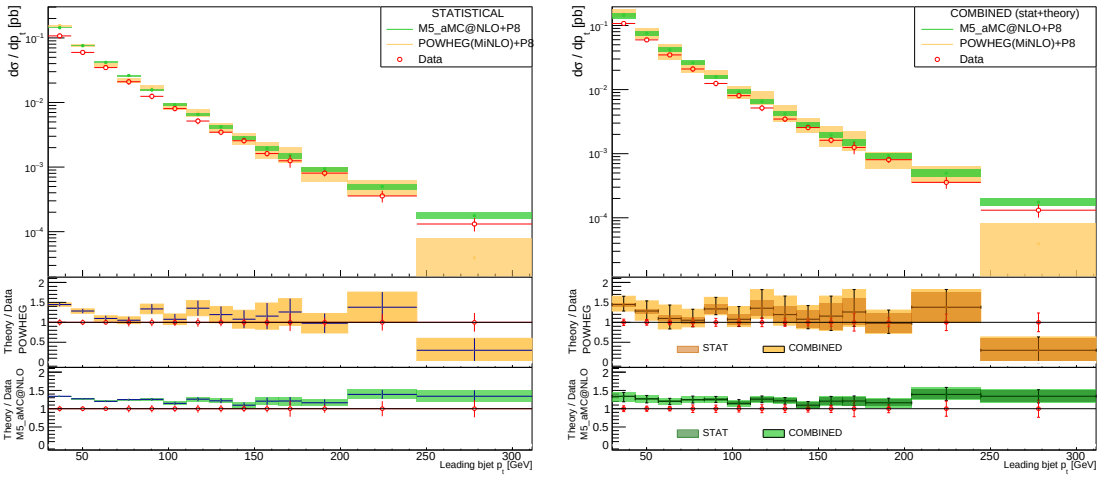

Figure 6.10: The differential $Z+b$ inclusive cross section as a function of the leading b-jet transverse momentum.The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

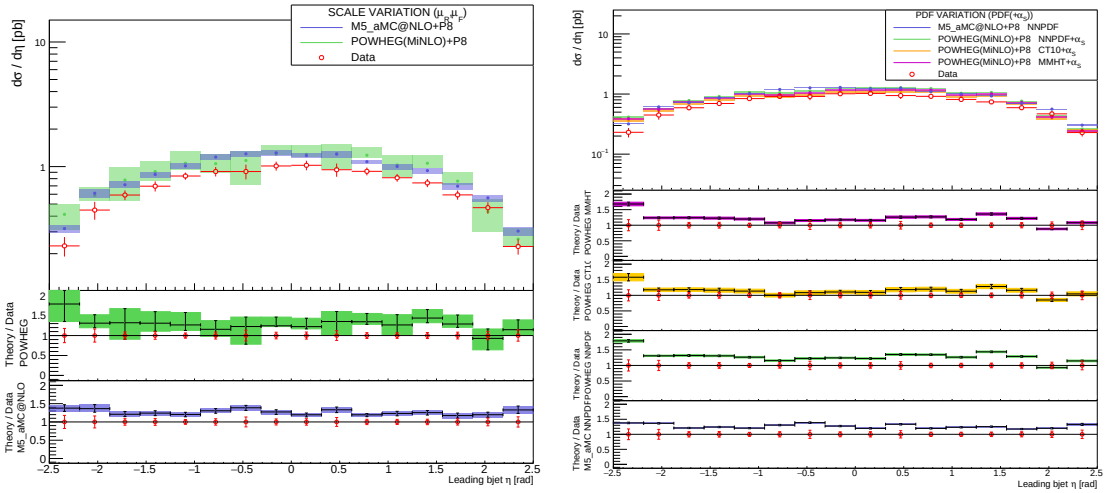

Figure 6.11: The differential $Z + b$ inclusive cross section as a function of the leading b-jet η. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

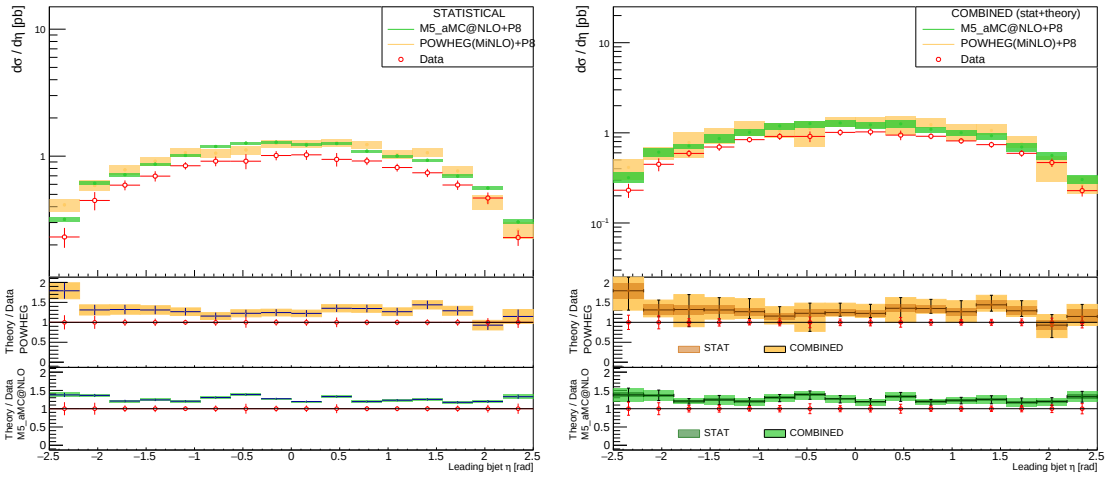

Figure 6.12: The differential $Z + b$ inclusive cross section as a function of the leading b-jet η. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

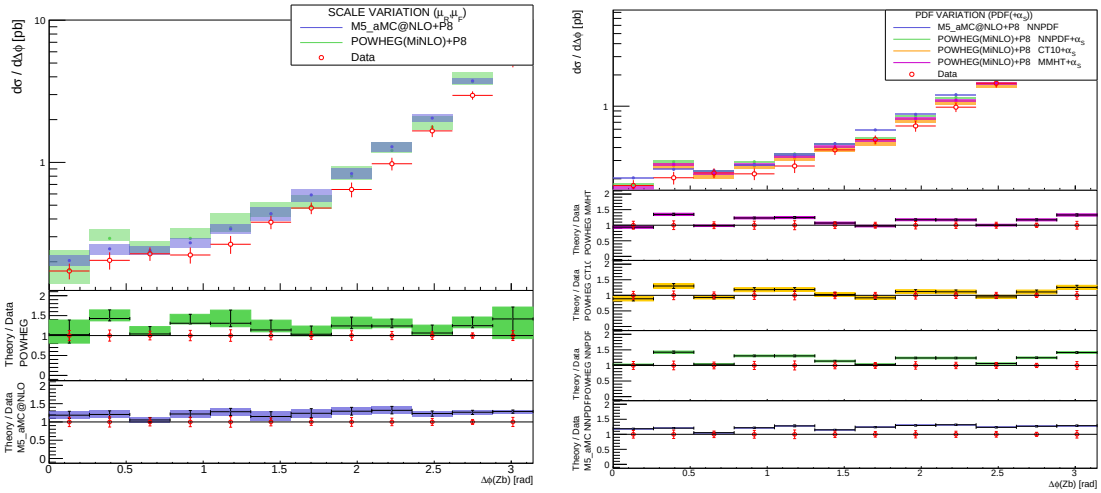

Figure 6.13: The differential $Z + b$ inclusive cross section as a function of the $\Delta\phi(Zb)$ angular separation between the leading b-jet and the Z boson. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

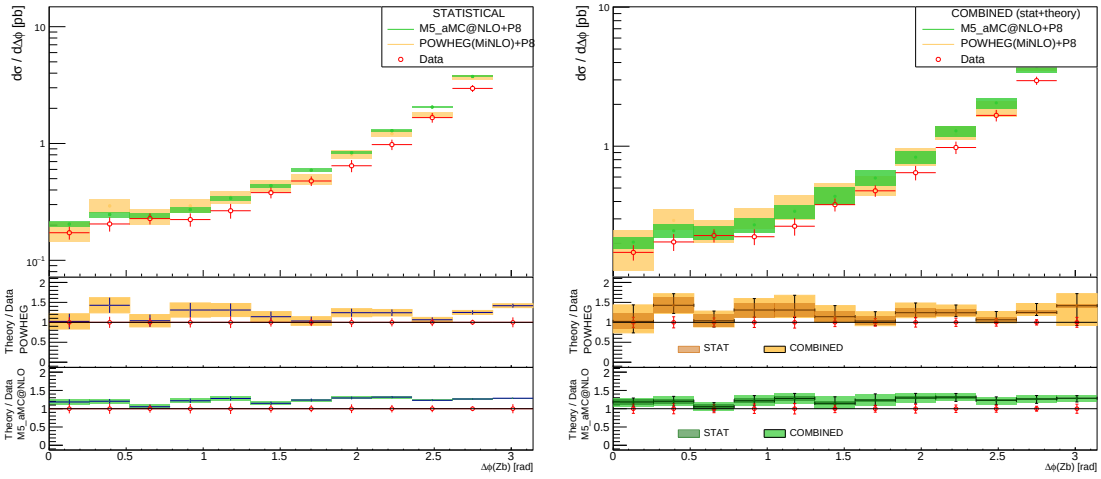

Figure 6.14: The differential $Z + b$ inclusive cross section as a function of the $\Delta\phi(Zb)$ angular separation between the leading b-jet and the Z boson. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

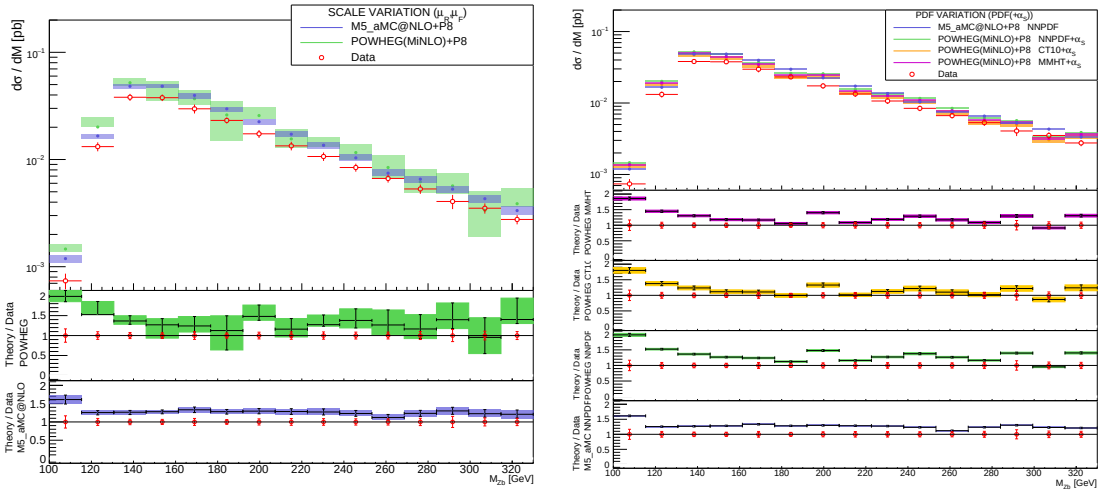

Figure 6.15: The differential $Z + b$ inclusive cross section as a function of the invariant mass M_{Zb}. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

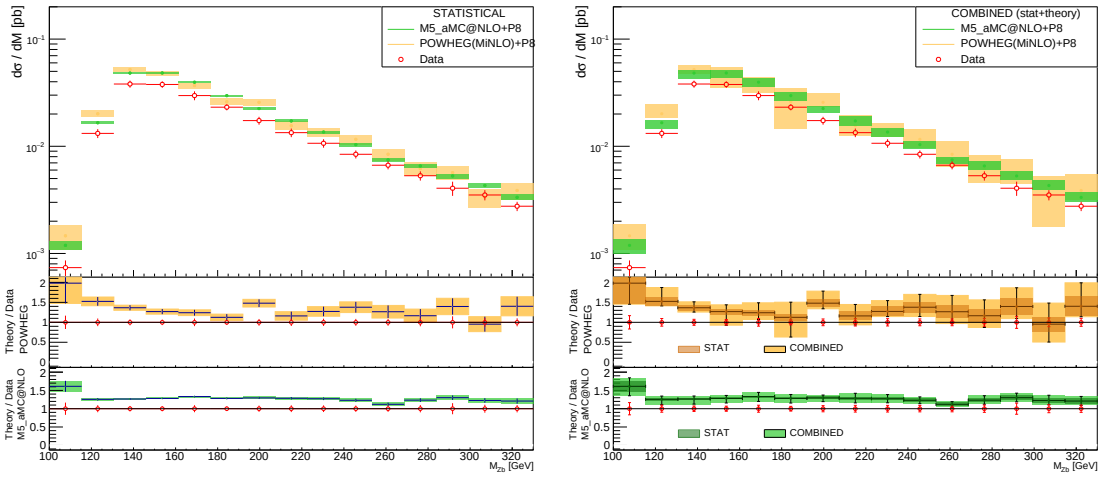

Figure 6.16: The differential $Z + b$ inclusive cross section as a function of the invariant mass M_{Zb}. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

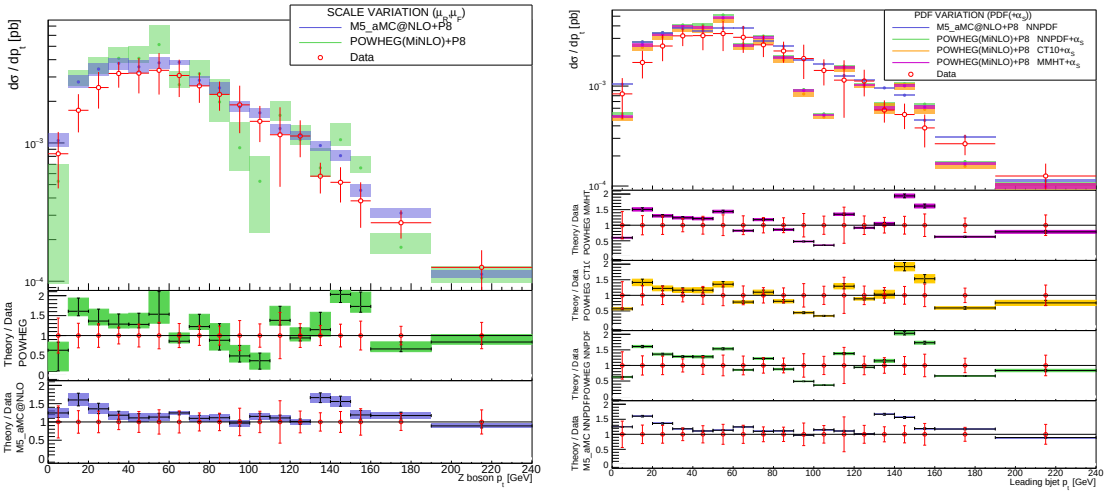

Figure 6.17: The differential $Z + 2b$ cross section as a function of the Z boson transverse momentum. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

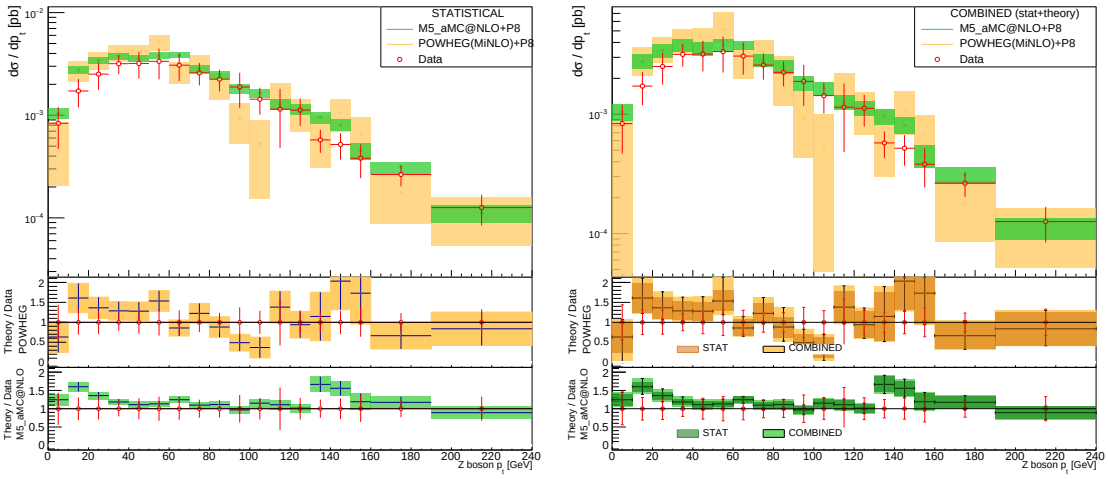

Figure 6.18: The differential $Z + 2b$ cross section as a function of the Z boson transverse momentum. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

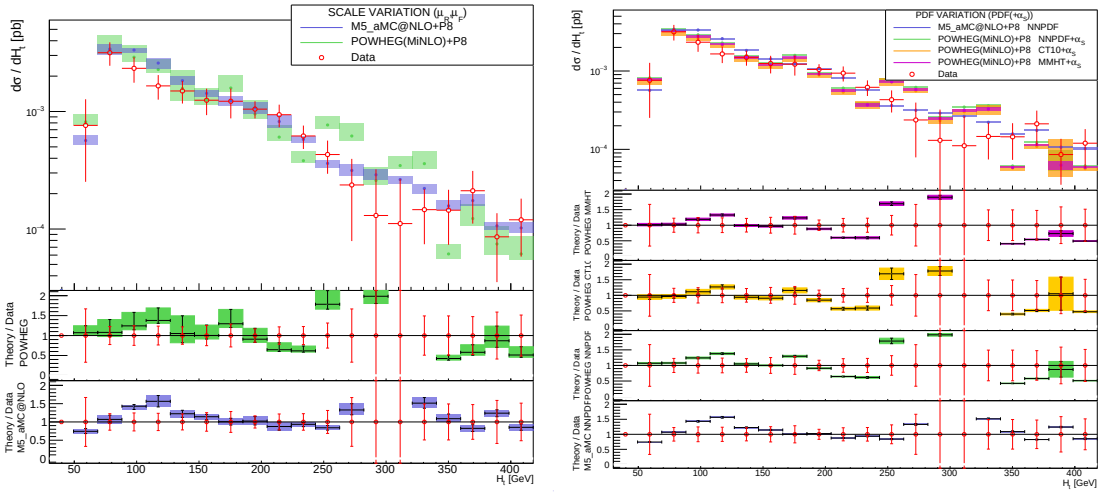

Figure 6.19: The differential $Z + 2b$ cross section as a function of the H_t variable. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

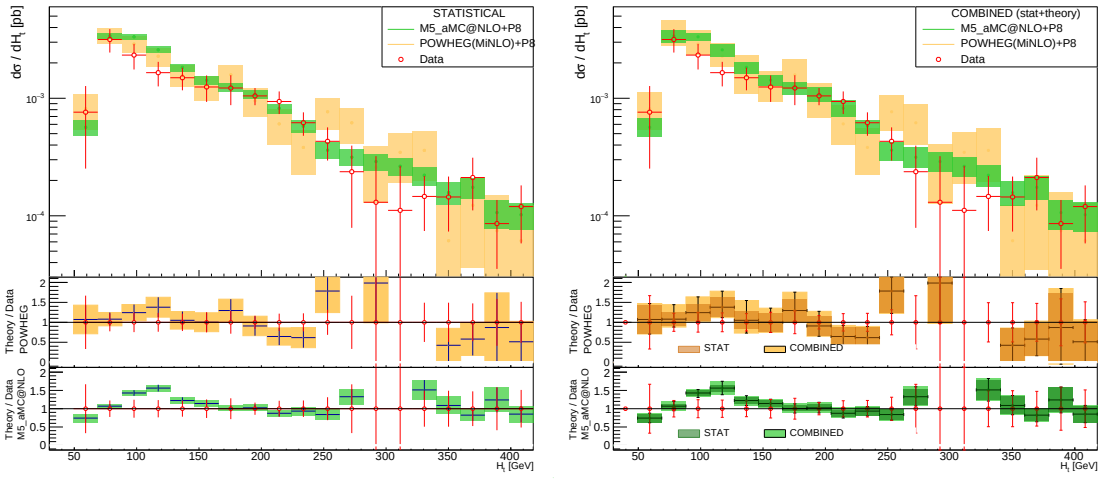

Figure 6.20: The differential $Z + 2b$ cross section as a function of the H_t variable. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

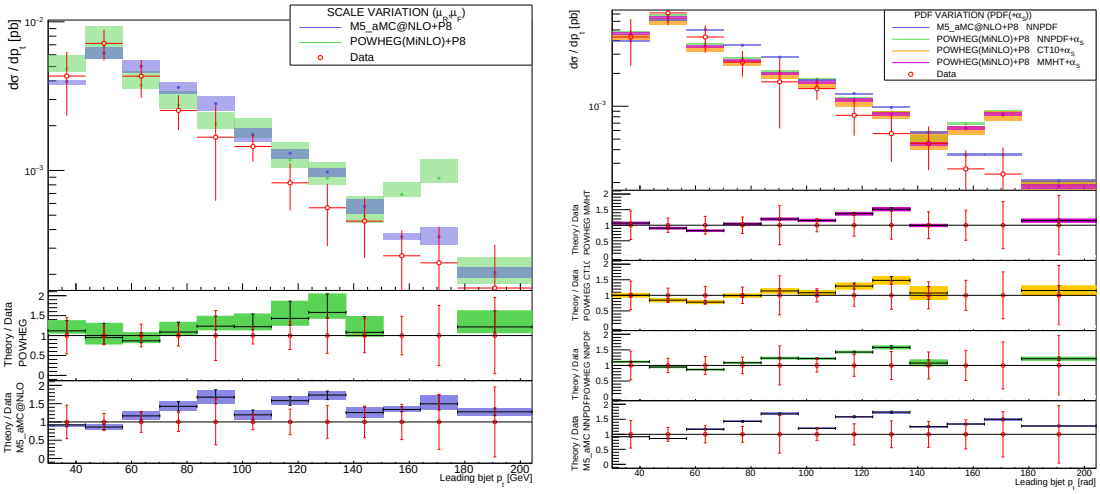

Figure 6.21: The differential $Z+2b$ cross section as a function of the leading b-jet transverse momentum. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

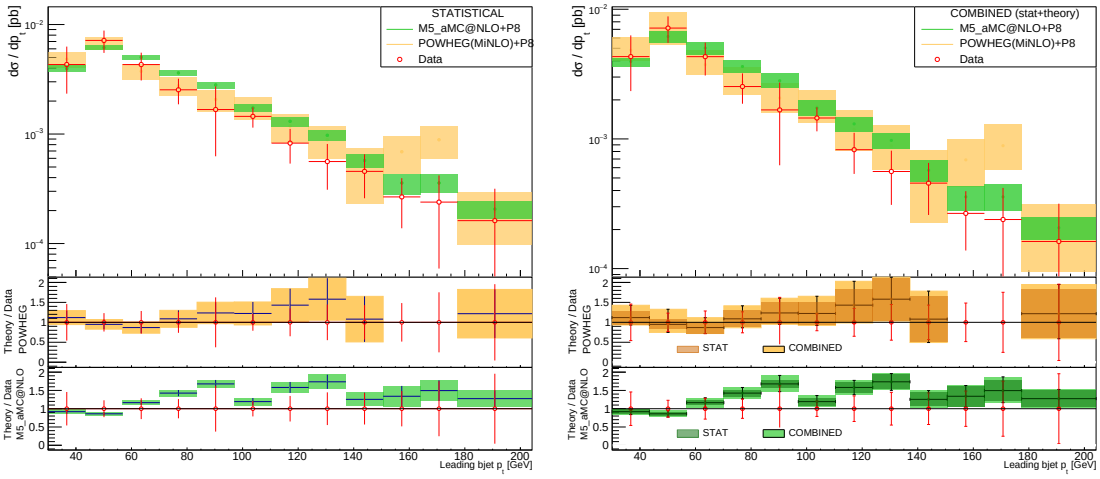

Figure 6.22: The differential $Z+2b$ cross section as a function of the leading b-jet transverse momentum. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

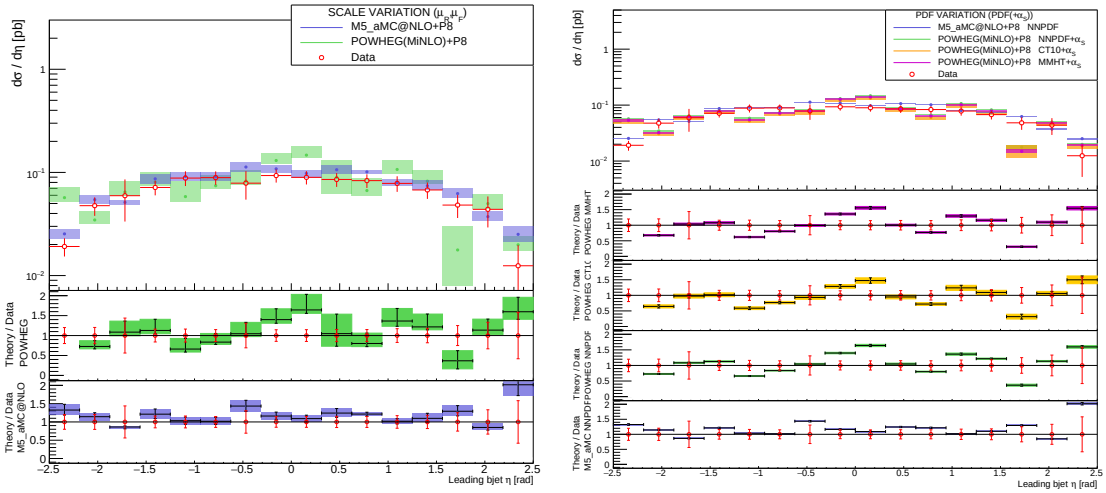

Figure 6.23: The differential $Z + 2b$ cross section as a function of the leading b-jet η. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

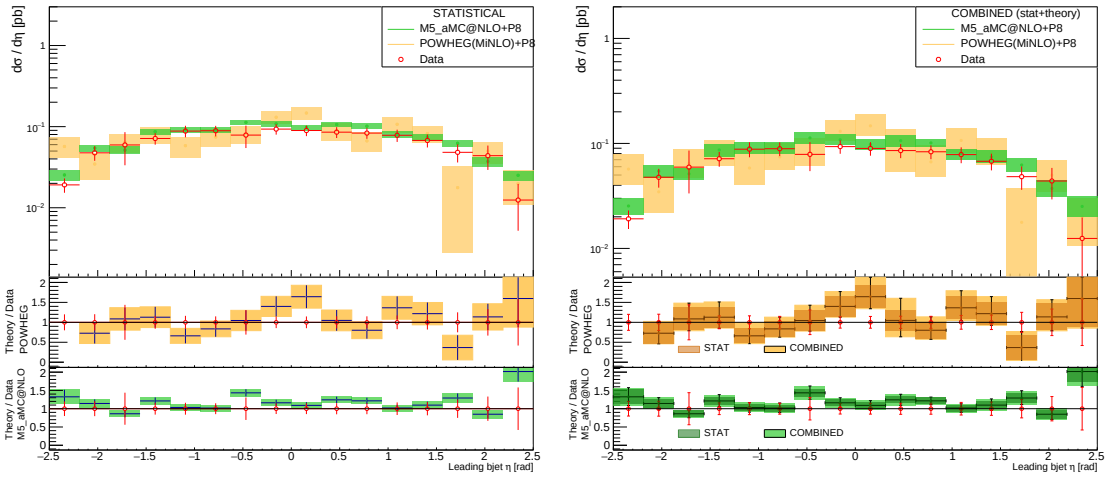

Figure 6.24: The differential $Z + 2b$ cross section as a function of the leading b-jet η. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

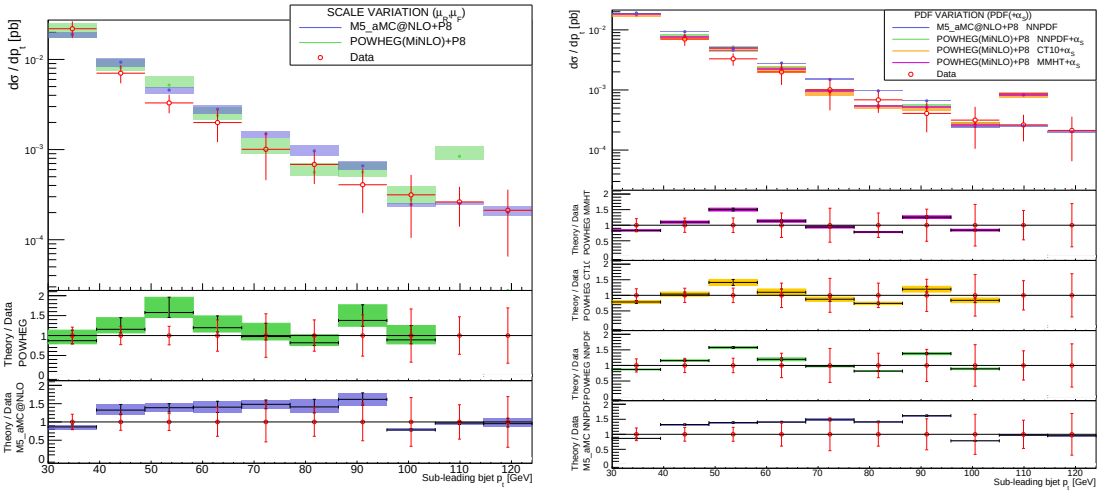

Figure 6.25: The differential $Z+2b$ cross section as a function of the sub-leading b-jet transverse momentum. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples

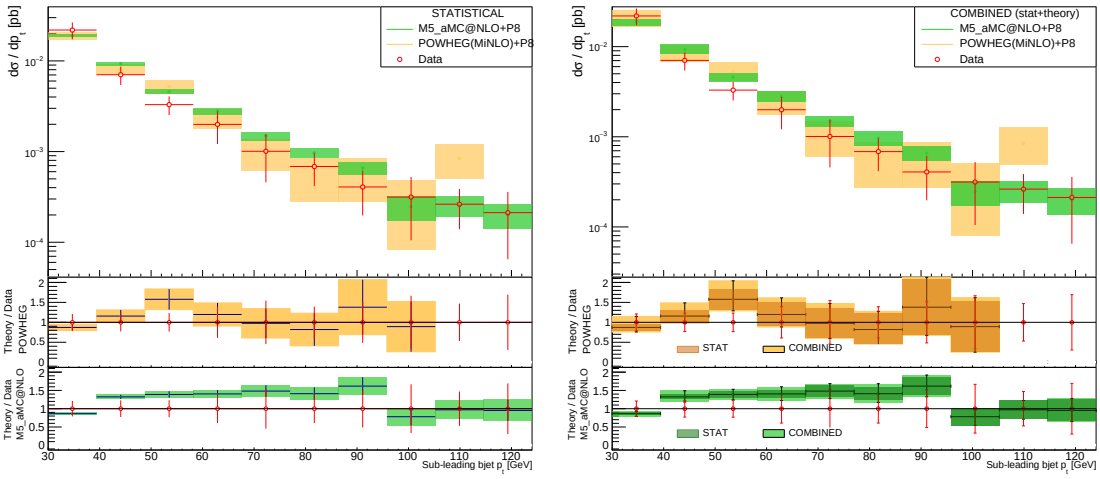

Figure 6.26: The differential $Z+2b$ cross section as a function of the sub-leading b-jet transverse momentum. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

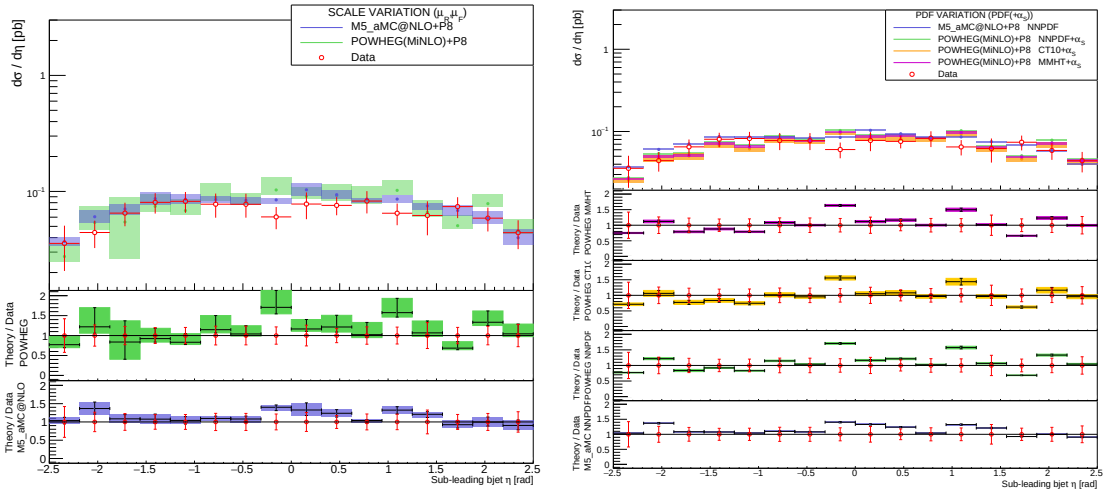

Figure 6.27: The differential $Z + 2b$ cross section as a function of the sub-leading b-jet η. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

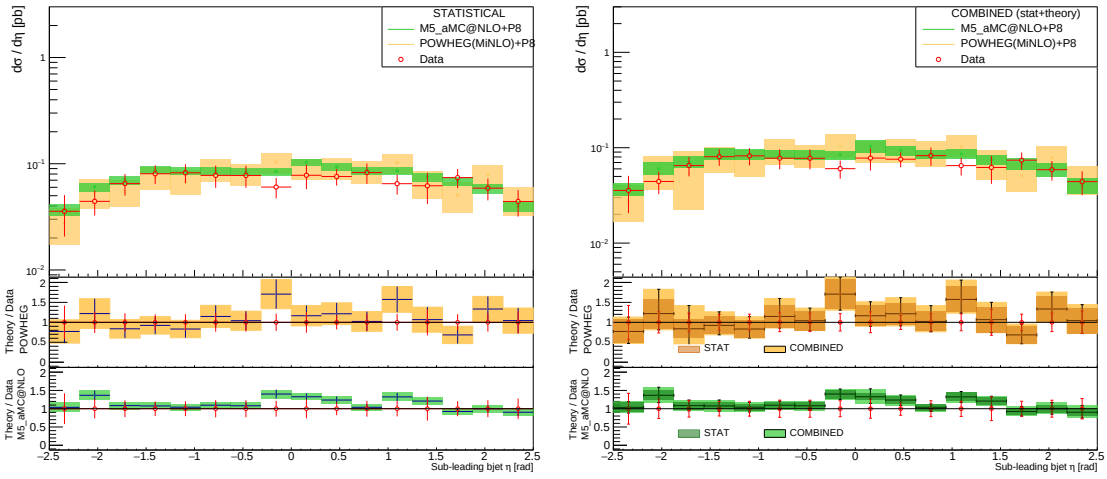

Figure 6.28: The differential $Z + 2b$ cross section as a function of the sub-leading b-jet η. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

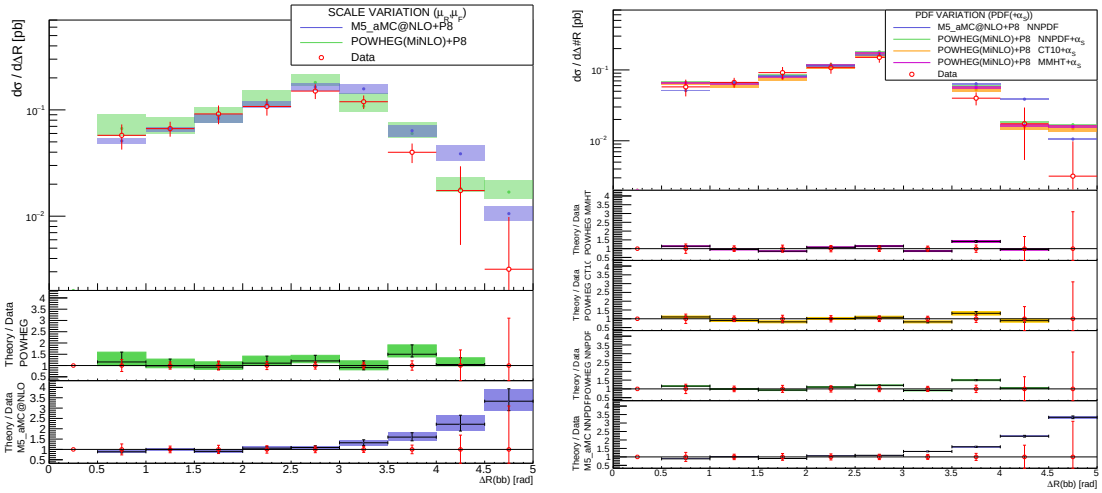

Figure 6.29: The differential $Z + 2b$ cross section as a function of the ΔR angular separation between the leading and sub-leading b-jets. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

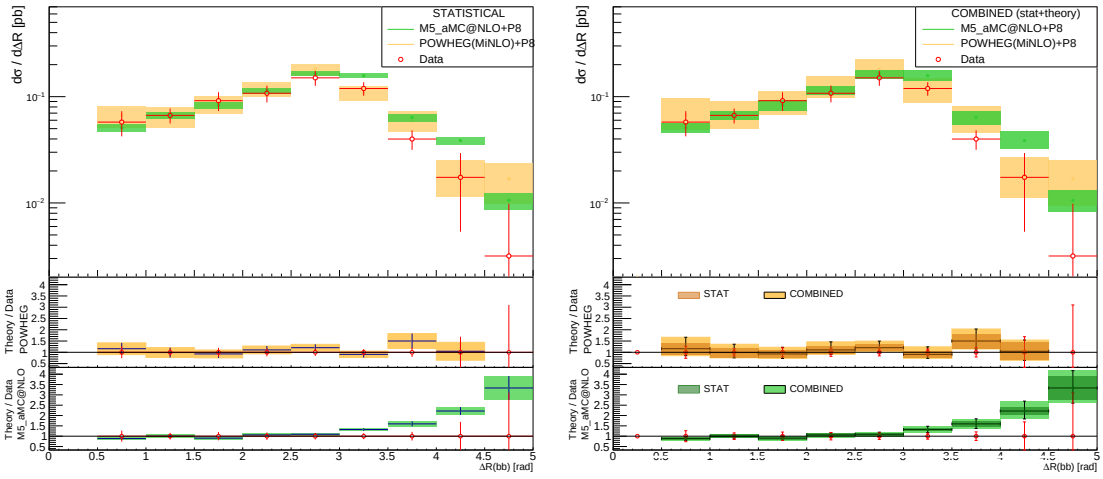

Figure 6.30: The differential $Z + 2b$ cross section as a function of the ΔR angular separation between the leading and sub-leading b-jets. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

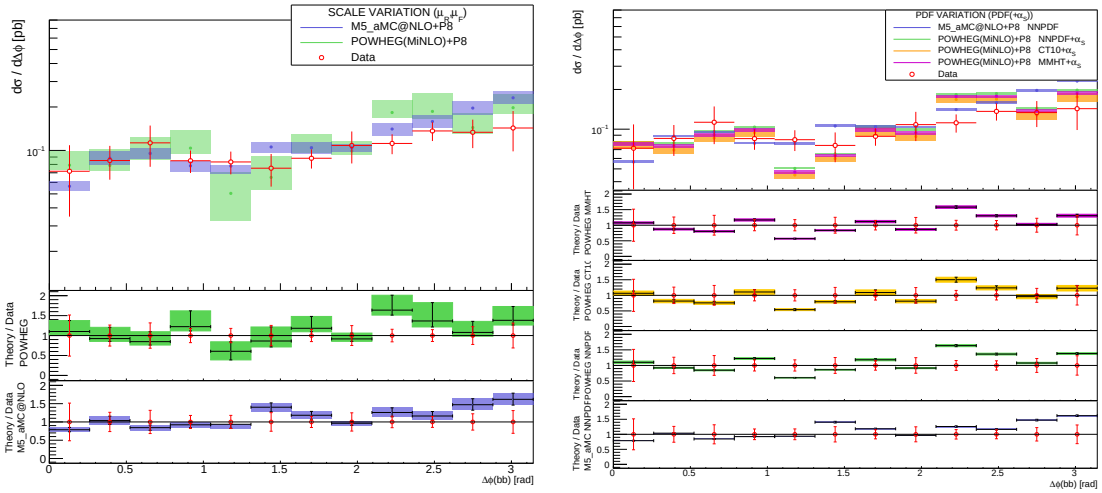

Figure 6.31: The differential $Z + 2b$ cross section as a function of the $\Delta\phi$ angular separation between the leading and sub-leading b-jets. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

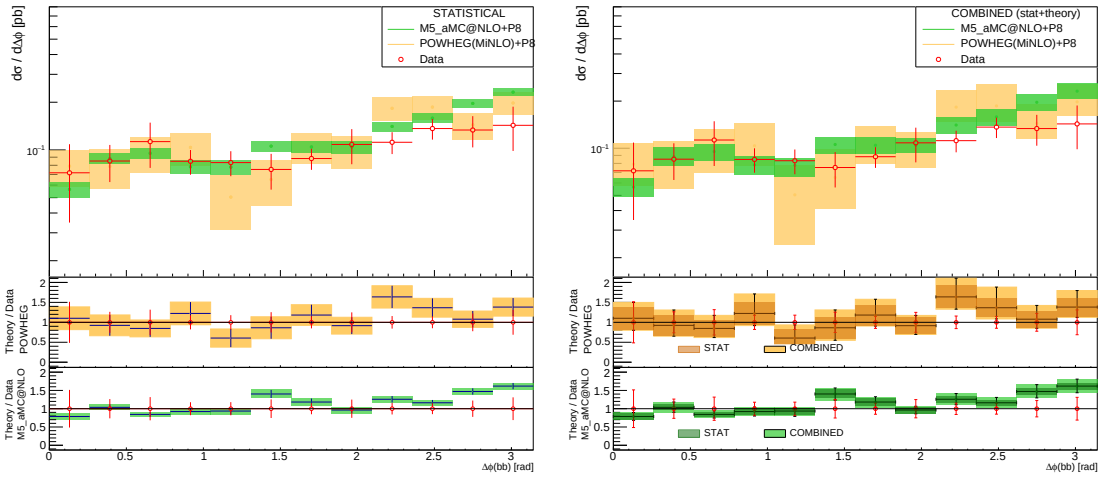

Figure 6.32: The differential $Z + 2b$ cross section as a function of the $\Delta\phi$ angular separation between the leading and sub-leading b-jets. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

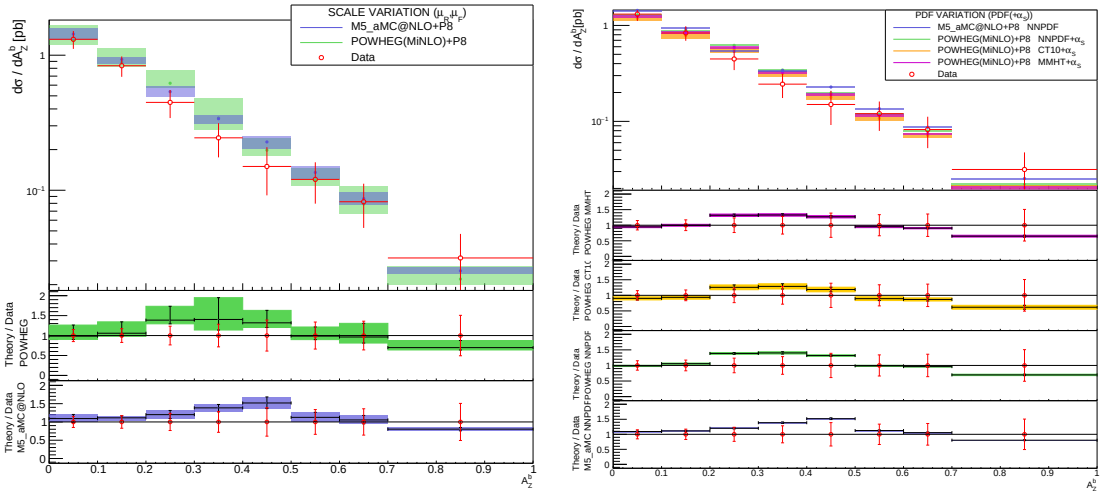

Figure 6.33: The differential $Z + 2b$ cross section as a function of the A_Z^b asymmetry. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

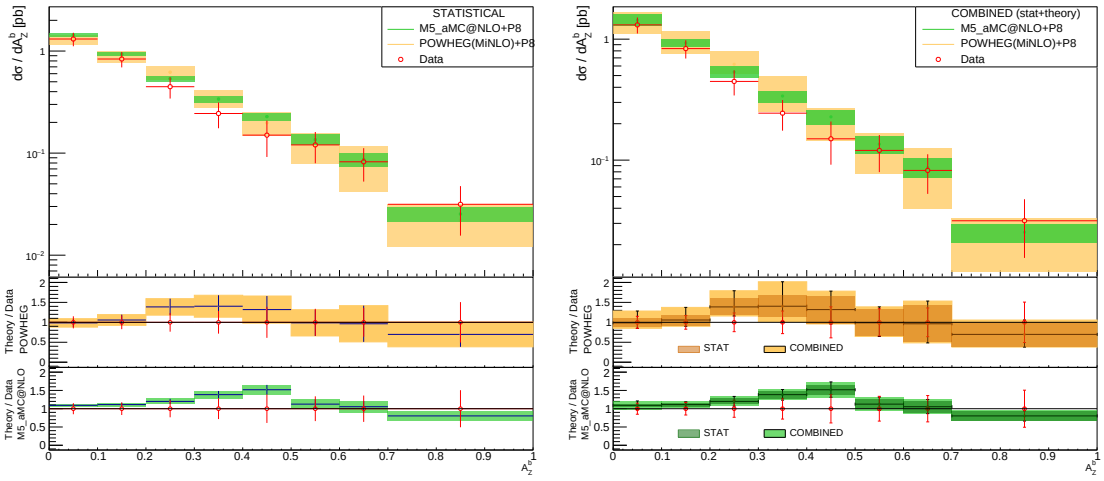

Figure 6.34: The differential $Z + 2b$ cross section as a function of the A_Z^b asymmetry. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

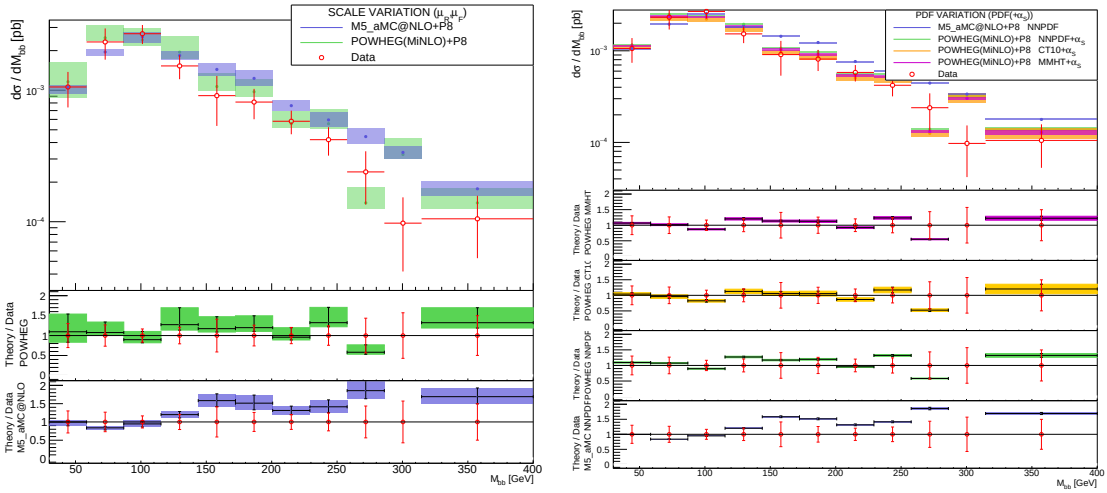

Figure 6.35: The differential $Z + 2b$ cross section as a function of the double b-jets M_{bb} invariant mass. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

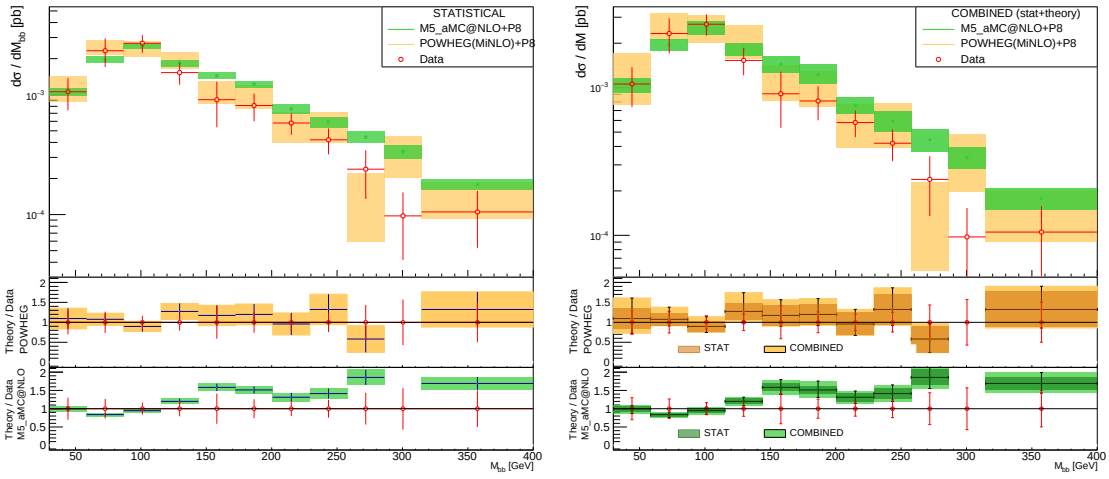

Figure 6.36: The differential $Z + 2b$ cross section as a function of the double b-jets M_{bb} invariant mass. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

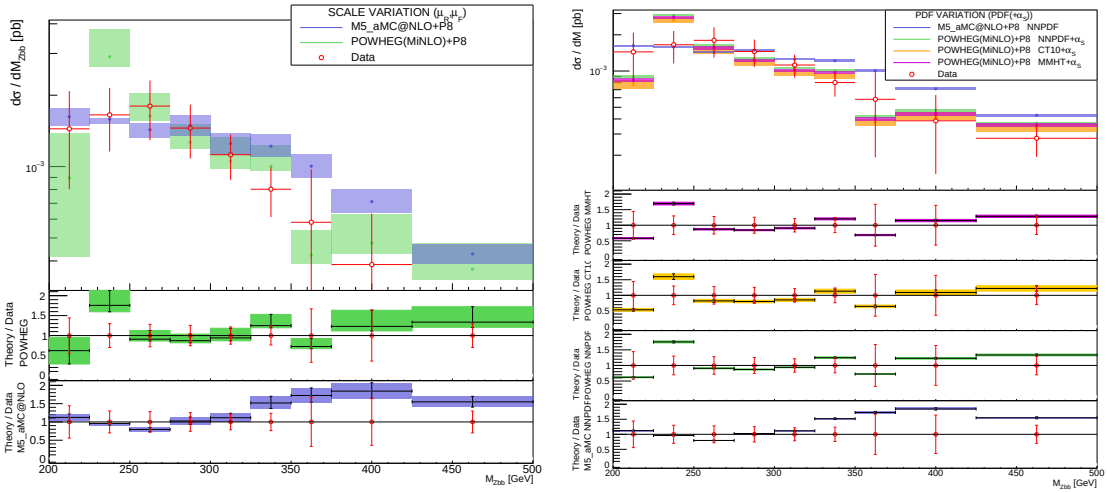

Figure 6.37: The differential $Z + 2b$ cross section as a function of the M_{Zbb} invariant mass. The theoretical errors associated to the scale variations (left) and PDF variations (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

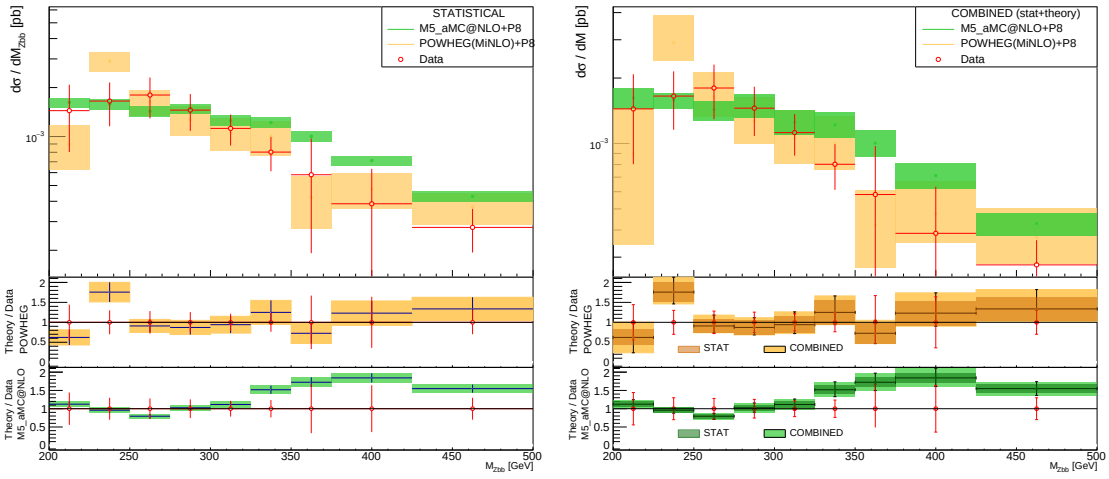

Figure 6.38: The differential $Z+2b$ cross section as a function of the M_{Zbb} invariant mass. The statistical errors (left) and the combined (statistical+systematic) errors (right) are reported for the MadGraph5_aMC@NLO and POWHEG+MiNLO samples.

Conclusions

The analysis of the associated production of a Z boson and b-jets has been presented, comparing the experimental data collected by the CMS experiment and the Next-To-Leading order predictions of the MadGraph5_aMC@NLO and the POWHEG generators. The data have been collected at the LHC during the 2012, at a center of mass energy of 8 TeV and with a total integrated luminosity of $19.8\, fb^{-1}$. Both the Monte Carlo (MC) generators used in the analysis have been interfaced with Pythia8 for the hadronization and they both use the 5 flavor scheme for the matrix element computation. Several kinematic observables have been investigated for the associated production of a Z boson and at least one or two b-jets, comparing the unfolded experimental cross sections with the theoretical ones. The results show a clear overestimation ($\sim 20\%$) of the normalization in both the MC generators. Nevertheless the experimental data and the theoretical predictions are compatible within two standard deviations and the shape of the distributions is well reproduced in most of the cases. The total (systematic and statistical) error is considered for the experimental data and the theoretical distributions. In particular the theoretical uncertainties related to the Parton Distribution Functions and scale variations have been estimated for the MC samples. Also different PDF sets have been investigated in the POWHEG case: the standard NNPDF set is the less compatible with the experimental results, while, using the CT10 set, the normalization issue seems mitigated.

The POWHEG sample was the only one to include multiple PDF sets and the strong coupling variations. As a future development to the analysis, it could be interesting to compare different PDFs also with the MadGraph5_aMC@NLO generator and to complete the associated error computation including the α_S contribution. The production of a larger POWHEG sample is also fundamental to reduce the associated statistical error and increase the accuracy of the prediction.

Appendix A

Rivet code

The following is the Rivet code used for the generated event selection.

```
// Book histograms and initialize projections before the run
void init() {

        //Finalstate declaration
        const FinalState fs;
        addProjection(fs, "FS");

        // Projection to find the electrons and muons
        IdentifiedFinalState part(fs);
        part.acceptIdPair(11);
        part.acceptIdPair(-11);
        part.acceptIdPair(13);
        part.acceptIdPair(-13);
        addProjection(part, "part");

        // Get photons to dress leptons
        IdentifiedFinalState photons(fs);
        photons.acceptIdPair(22);
        addProjection(photons, "photons");

        //Histograms booking
        //...

}


// Perform the per-event analysis
void analyze(const Event& event) {

        //Event weight
        const double MyWeight = event.weight();
        double Ht = 0;

        //Structure to contain original particle, dressed lepton and photons used for dressing
        struct pt_and_particles {
                FourMomentum p_part;
                vector < FourMomentum >lepton_photon;
        };

        struct pt_and_particles ele_dres;
        struct pt_and_particles pos_dres;
        struct pt_and_particles muon_dres;
        struct pt_and_particles antimuon_dres;

        //Apply lepton and photon projections
        vector<Particle> part;
        part = applyProjection<IdentifiedFinalState>(event, "part").particles();
        vector<Particle> photons;
        photons = applyProjection<IdentifiedFinalState>(event, "photons").particles();

        //Store lepton and photons fourmomentum
        FourMomentum partm;
        vector <FourMomentum> photonsm;

        //if there are no leptons veto
```

```
if (part.size()==0) vetoEvent;

//dressing of each lepton
foreach (const Particle& e, part){
        partm=e.momentum();
        photonsm.push_back(e.momentum());
        foreach (const Particle& p, photons){
                if  (deltaR(p.momentum(),e.momentum())<0.1) {
                        partm+=p.momentum();
                        photonsm.push_back(p.momentum());
                }
        }

        //saving of dressed lepton into the right ID structure
        if(e.pdgId()==11 && partm.pt()>20*GeV && partm.pt()>ele_dres.p_part.pt() &&
        partm.abseta()<2.4){
                ele_dres.p_part=partm;
                ele_dres.lepton_photon=photonsm;
        }
        if(e.pdgId()==-11 && partm.pt()>20*GeV && partm.pt()>pos_dres.p_part.pt() &&
        partm.abseta()<2.4){
                pos_dres.p_part=partm;
                pos_dres.lepton_photon=photonsm;
        }
        if(e.pdgId()==13 && partm.pt()>20*GeV && partm.pt()>muon_dres.p_part.pt() &&
        partm.abseta()<2.4){
                muon_dres.p_part=partm;
                muon_dres.lepton_photon=photonsm;
        }
        if(e.pdgId()==-13 && partm.pt()>20*GeV && partm.pt()>antimuon_dres.p_part.pt() &&
        partm.abseta()<2.4){
                antimuon_dres.p_part=partm;
                antimuon_dres.lepton_photon=photonsm;
        }

        photonsm.clear();
}

//if there are no dressed lepton veto
if ((ele_dres.lepton_photon.empty() || pos_dres.lepton_photon.empty()) &&
(muon_dres.lepton_photon.empty() || antimuon_dres.lepton_photon.empty())) vetoEvent;

//Store Z boson fourmomentum
FourMomentum ze(add(ele_dres.p_part, pos_dres.p_part));
FourMomentum zm(add(muon_dres.p_part, antimuon_dres.p_part));

if (ze.mass2()<0 || zm.mass2()<0) vetoEvent;

//Event identification depending on mass window
bool ee_event=false;
bool mm_event=false;

if (!ele_dres.lepton_photon.empty() && !pos_dres.lepton_photon.empty() && ze.mass()>71
&& ze.mass()<111) ee_event = true;
if (!muon_dres.lepton_photon.empty() && !antimuon_dres.lepton_photon.empty() && zm.mass()>71
&& zm.mass()<111) mm_event = true;

//If there is no Z boson veto
if (!ee_event && !mm_event) vetoEvent;

// Pseudojet building procedure
const FinalState& fs=applyProjection < FinalState > (event, "FS");
std::vector < fastjet::PseudoJet > vecs;
vector < Particle > part_jets;
int l=0;

//If the particle is not a neutrino and it isn't already used during
//the dressing a pseudojet is created
foreach (const Particle& p, fs.particles()) {
        if (fabs(p.pdgId())!=12 && fabs(p.pdgId())!=14 && fabs(p.pdgId())!=16  ) {
                bool overcount=false;
                if (mm_event){
                int psize1=muon_dres.lepton_photon.size();
                int psize2=antimuon_dres.lepton_photon.size();
                        for (int j=0;j<psize1;j++){
                                if (p.momentum()==muon_dres.lepton_photon[j]) overcount=true;
                        }
                        for (int j=0;j<psize2;j++){
                                if (p.momentum()==antimuon_dres.lepton_photon[j]) overcount=true;
                        }
                }
                if (ee_event) {
                        int psize3=ele_dres.lepton_photon.size();
                        int psize4=pos_dres.lepton_photon.size();
                        for (int j=0;j<psize3;j++){
                                if (p.momentum()==ele_dres.lepton_photon[j]) overcount=true;
                        }
```

```
                                for (int j=0;j<psize4;j++){
                                        if (p.momentum()==pos_dres.lepton_photon[j]) overcount=true;
                                }
                        }
                        if(overcount==false && (ee_event || mm_event)){
                                fastjet::PseudoJet pseudoJet(p.momentum().px(), p.momentum().py(),
                                p.momentum().pz(), p.momentum().E());
                                pseudoJet.set_user_index(l);
                                vecs.push_back(pseudoJet);
                                part_jets.push_back(p);
                                l++;
                        }
                }
        }

        //Jet building procedure
        fastjet::ClusterSequence cseq(vecs, fastjet::JetDefinition(fastjet::antikt_algorithm, 0.5));
        vector < fastjet::PseudoJet > jets = sorted_by_pt(cseq.inclusive_jets(30.0));
        vector < fastjet::PseudoJet > jet_final;
        vector < fastjet::PseudoJet > jb_final;

        //Identification of "good" jets and b-jets
        foreach (const Jet& j, jets) {
                bool bjet_found = false;
                bool bjet_found1 = false;
                if ((j.abseta() < 2.4 && j.perp() > 30) &&
                ((ee_event && deltaR(j.momentum(),ele_dres.p_part)>0.5 &&
                deltaR(j.momentum(),pos_dres.p_part)>0.5) ||
                (mm_event && deltaR(j.momentum(),muon_dres.p_part)>0.5 &&
                deltaR(j.momentum(),antimuon_dres.p_part)>0.5))) {
                        jet_final.push_back(j);
                        Ht = Ht + j.perp();
                        FourMomentum pim=j.momentum();
                        foreach (const fastjet::PseudoJet& c, cseq.constituents(j)) {
                                const Particle & part = part_jets.at(c.user_index());
                                const PdgId pid = part.pdgId();
                                if (((abs(pid)/100)%10 == 5 ||(abs(pid)/1000)%10 == 5) &&
                                 deltaR(j,part.momentum().eta(),part.momentum().phi())<0.5)
                                bjet_found = true;

                                HepMC::GenVertex * gv = part.genParticle()->production_vertex();
                                if (gv) {
                                        foreach (const GenParticle * pi, Rivet::particles(gv, HepMC::ancestors))
                                                const PdgId pid2 = pi->pdg_id();
                                                if (((abs(pid2)/100)%10 == 5 || (abs(pid2)/1000)%10 == 5) &&
                                                (pi->status()>0 && pi->status()<4)) {
                                                        bjet_found = true;
                                                        pim=pi->momentum();
                                                }
                                        }
                                }
                                if (bjet_found && deltaR(j,pim.eta(),pim.phi())<0.5) bjet_found1=true;
                        }
                        if (bjet_found1) {
                                jb_final.push_back(j);
                        }
                }
        }

        //Histogram filling
        //...
}

/// Normalize histograms etc., after the run
```

Appendix B

Cross section normalization

In the following Table there is a summary of the signal and background contributions, to the $Z + b$ production, which are relevant for this analysis. In the central column there is the Leading order cross section, while in the right column there is the normalized cross section used for the specific sample. The backgrounds have been used to extract the signal from the measured data before applying the unfolding procedure. The dominant background in the $Z + b$ analysis is the top-antitop $t\bar{t}$, in particular for the $Z + 2b$ final state. The top quark, in fact, has an almost 1 probability of decaying into a b-quark and a W vector boson because the V_{tb} CKM matrix element is $\simeq 1$. Subsequently the W bosons decay into leptons therefore, when a $t\bar{t}$ state is produced, the final particles will be 2 opposite charge leptons and 2 b-jets, which exactly mimic the $Z + 2b$ final state.

Signal	σ_{LO} (pb)	σ_{NORM} (pb)
Drell-Yan Z + jets	2950	3531.8 (NNLO)
Background	σ_{LO} (pb)	σ_{NORM} (pb)
$Z + b\bar{b}$	76.75	76.75 (LO)
W + jets	30400	36703.2 (NNLO)
$t\bar{t}$	136.3	225.197 (NLO)
ZZ	5.196	8.059 (NLO)
WZ	12.63	33.21 (NLO)
WW	34.01	54.838 (NLO)
single top t^+ (s-channel)	1.57	1.76 (NNLO)
single top t^+ (t-channel)	25	30.7 (NNLO)
single top tW	10.7	11.1 (NNLO)
single top t^- (s-channel)	2.82	3.79 (NNLO)
single top t^- (t-channel)	47	56.4 (NNLO)

Bibliography

[1] S. Weinberg, *A Model of Leptons*, Phys. Rev.Lett., 19 1264-1266, 1967.

[2] A. Salam, *Weak and Electromagnetic Interactions* Conf.Proc. C680519:367–577, 1968.

[3] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys., 22:579–588, 1961.

[4] DONUT Collaboration, *Observation of Tau Neutrino Interactions*, Phys. Lett. B 504 (2001), 2000.

[5] DØ Collaboration, *Observation of the Top Quark*, Phys. Rev. Lett. 74 (1995) 14, 1995.

[6] CDF Collaboration, *Observation of Top Quark Production in pp Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. 74 (1995) 14, 1995.

[7] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at* $\sqrt{s} = 540$ *GeV*, Phys.Lett. B122 (1983) 103-116, 1983.

[8] UA2 Collaboration *Evidence for* Z_0 *to* e^+e^- *at the CERN anti-pp Collider*, Phys.Lett., B129:130–140, 1983.

[9] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (2012) 1, 2012.

[10] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 (2012) 1, 2012.

[11] G. Dissertori, I. Knowles, M. Schmelling, *Quantum Chromodynamics, High Energy Experiments and Theory*, Oxford University Press, 2009.

[12] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog.Theor.Phys., 49:652–657, 1973.

[13] P. W. Higgs, *Broken symmetries, massless particles and gauge fields* Phys.Rev.Lett., 12(2):132–133, 1964.

[14] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons* Phys.Rev.Lett., 13:321–323, 1964.

[15] Abdelhak Djouadi, *The Anatomy of ElectroWeak Symmetry Breaking Tome I: The Higgs boson in the Standard Model*, arXiv:hep-ph/0503172, 2005.

[16] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. 51.1945, 1983.

[17] Nicola Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys.Rev.Lett., 10:531–533, 1963.

[18] C. Jarlskog, *Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation*, Phys. Rev. Lett. 55.1039, 1985.

[19] J. Beringer et al., Review of Particle Physics (RPP). Phys.Rev., D86:010001, 2012.

[20] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rep. 97, 31–145, 1983.

[21] R.K. Ellis, W.J. Stirling, and B.R. Webber, *QCD and collider physics*, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. 81,1996.

[22] Super-Kamiokande Collaboration, *Evidence for oscillation of atmospheric neutrinos*, Phys.Rev.Lett., 81:1562–1567, 1998.

[23] OPERA Collaboration, *Search for $\nu_\mu \to \nu_e$ oscillations with the OPERA experiment in the CNGS beam*, JHEP, 1307:004, 2013.

[24] G. Altarelli and G. Parisi, Nucl.Phys. B126:298,1977.

[25] Yu.L. Dokshitzer, Sov.Phys. JETP 46:641,1977.

[26] V.N. Gribov, L.N. Lipatov, Sov.J.Nucl.Phys. 15:438,1972.

[27] E. Lancu, *QCD in heavy ion collisions*, arXiv:1205.0579.

[28] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP, 07:012, 2002.

[29] The NNPDF Collaboration, *Parton distributions with QED corrections*, arXiv:hep-ph/1308.0598, 2013.

[30] Marco Guzzi et al., *CT10 parton distributions and other developments in the global QCD analysis*, arXiv:1101.0561, 2011.

[31] L. A. Harland-Lang et al., *Parton distributions in the LHC era: MMHT 2014 PDFs*, arXiv:1412.3989, 2014.

[32] K.A. Olive et al., *Review of Particle Physics (RPP)*, Phys.Rev., D86:010001,2014.

[33] R. K. Ellis, H. Georgi, M. Machacek, H. D. Politzer and G. G. Ross, *Factorization and the Parton Model in QCD*, Phys.Lett., B78:281, 1978.

[34] Jan-Christopher Winter et al., *A modified cluster-hadronization model*, http://arxiv.org/abs/hep-ph/0311085, 2003.

[35] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rep. 97, 31-145, 1983.

[36] Andreas Hocker and Vakhtang Kartvelishvili, *SVD approach to data unfolding*. Nucl. Instrum. Meth., A372:469–481, 1996.

[37] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem.* Nucl.Instrum.Meth., A362:487–498, 1995.

[38] T. Sjostrand et al., *PYTHIA 6.4 Physics and Manual*, JHEP 0605 (2006) 026, arXiv:hep-ph/0603175, 2006.

[39] T. Sjostrand et al., *An Introduction to PYTHIA 8.2*, arXiv:hep-ph/arXiv:1410.3012, 2014.

[40] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C58, arXiv:hep-ph/0803.0883, 2008.

[41] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP 0902 (2009) 007, arXiv:hep-ph/0811.4622, 2008.

[42] J. Alwall et a., *MadGraph 5 : Going Beyond*, arXiv:1106.0522, 2007.

[43] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP, vol. 07, hep-ph/0206293, 2003.

[44] S. Frixione et al., *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, arXiv:0709.2092, 2007.

[45] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP 07 (2014) 079, arXiv:hep-ph/1405.0301, 2014.

[46] N. D. Christensen and C. Duhr, *FeynRules - Feynman rules made easy*, arXiv:0806.4194, 2009.

[47] Keith Hamilton et al., *MINLO: Multi-scale improved NLO*, arXiv:1206.3572, 2012.

[48] M. Mangano et al., *Multijet matrix elements and shower evolution in hadronic collisions: W b $\bar{b}$ + n jets as a case study*, Nucl. Phys. B, arXiv:hep-ph/0108069, 2002.

[49] S. Catani et al., *QCD matrix elements + parton showers*, JHEP 0111 (2001) 063, 2011.

[50] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, arXiv:hep- ph/1209.6215,2012.

[51] John M. Campbell, R. Keith Ellis, F. Maltoni, and S. Willenbrock, *Associated production of a Z Boson and a single heavy quark jet* Phys.Rev., D69:074021, 2004.

[52] Fabio Maltoni, Giovanni Ridolfi, and Maria Ubiali, *b-initiated processes at the LHC: a reappraisal* JHEP, 1207:022, 2012.

[53] T. Aaltonen et al., *Measurement of cross sections for b jet production in events with a Z boson in $p\overline{p}$ collisions at $\sqrt{s} = 1.96\,TeV$.* Phys.Rev., D79:052008, 2009.

[54] D0 Collaboration, *Measurement of the ratio of differential cross sections $\sigma(p\overline{p} \to Z + bjets)/(p\overline{p} \to Z + jets)$ in $p\overline{p}$ collision at $\sqrt{s} = 1.96$ TeV*, Phys.Rev., D87(9):092010, 2013.

[55] CMS Collaboration, *Measurement of the production cross sections for a Z boson and one or more b jets in pp collisions at $\sqrt{s} = 7$ TeV*, arXiv hep-ex/, 1402.1521, 2014.

[56] 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*, Journal of High Energy Physics, 10(10):141–49, 2014.

[57] S. Heinemeyer et al., *Handbook of LHC Higgs cross sections: 3. Higgs properties* CERN Report CERN-2013-004, 2013.

[58] Serguei Chatrchyan et al., *Search for a Vector-like Quark with Charge 2/3 in $t + Z$ Events from pp Collisions at $\sqrt{s} = 7$ TeV*, Phys.Rev.Lett., 107:271802, 2011.

[59] Georges Aad et al., *Search for down-type fourth generation quarks with the ATLAS detector in events with one lepton and hadronically decaying W bosons*, Phys.Rev.Lett., 109:032001, 2012.

[60] M. Baak, R. Kogler, *The global electroweak Standard Model fit after the Higgs discovery*, arXiv:1306.0571,2013.

[61] B. W. Lynn and C. Verzegnassi, *Longitudinal e^- beam-polarization asymmetry in $e^+e^- \to$ hadrons* Phys. Rev. D 35 (1987) 3326, 1987.

[62] M. Beccaria, N. Orlando, G. Panizzo, F.M. Renard and C. Verzegnassi, *The relevance of polarized bZ production at LHC*, arXiv:1204.5315v2, 2012.

[63] Chong-Xing Yue, Qing-Guo Zeng, Qiu-Yang Shi, Meng-Ying Liao, *Bottom partner B' and Zb production at the LHC*, arXiv:1212.3080v2, 2012.

[64] L. Evans and P. Bryant, *LHC Machine*, JINST, 3:S08001, 2008.

[65] CERN, *LEP design report. 1. The LEP injector chain*, 1984.

[66] CERN, *LEP Design Report: Vol. 2. The LEP Main Ring*, 1984.

[67] K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST, 3: S08002, 2008.

[68] Jr. Alves, A. Augusto et al., *The LHCb Detector at the LHC*, JINST, 3: S08005, 2008.

[69] G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider* JINST, 3:S08003, 2008.

[70] S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, JINST, 3:S08004, 2008.

[71] CMS Collaboration, *CMS tracker design report: Technical Design Report*, Technical Design Report CMS, CERN, 1997.

[72] CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, 1997

[73] CMS Collaboration, *Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7TeV$*, arXiv:1306.2016, 2013.

[74] CMS Collaboration, *The CMS hadron calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, 1997.

[75] V.D.Elvira, *Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment* CMS-NOTE-2004-020, 2004.

[76] CMS Collaboration, *The CMS magnet project: Technical Design Report*, Technical Design Report CMS, CERN, 1997.

[77] CMS Collaboration, *The CMS muon project: Technical Design Report* Technical Design Report CMS, CERN, 1997.

[78] CMS Collaboration, *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*, Technical Design Report CMS, CERN, 2002.

[79] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, Number CMS-PAS-PFT-09-001. 2009.

[80] Seguei Chatrchyan et al., *Identification of b quark jets with the CMS experiment*, JINST, 8:P04013, 2013.

[81] Christian Weiser, *A combined secondary vertex based b-tagging algorithm in CMS*, Technical Report CMS-NOTE, 014, 2006.

[82] Andy Buckley et al., *Rivet User Manual*, arXiv:1003.0694, 2010.

[83] S. Agostinelli et al., *Geant 4 simulation toolkit*, Nucl.Inst.Meth.A, 506(3):250 – 303, 2003.

[84] Andreas Hocker and Vakhtang Kartvelishvili, *SVD approach to data unfolding*, Nucl. Instrum. Meth., A372:469–481, 1996.

[85] T. Adye, *Unfolding algorithms and tests using RooUnfold*, arXiv:1105.1160, 2011.

[86] Giulia Ricciardi, *Semileptonic B decays*, arXiv:1301.4389.

[87] *Measurement of the associated production of a Z boson and b quarks in proton-proton collisions at* $\sqrt{s} = 8\,TeV$ *with the CMS experiment at LHC*, 2013.

[88] Chiara La Licata, *Measurement of the production cross section of a Z boson in association with exactly one or at least two b-jets with the CMS experiment at LHC*, 2014.

[89] P. M. Stevenson, *Optimized perturbation theory*, Phys. Rev. D23 (1981) 2916, 1981.

[90] Sergey Alekhin et al., *The PDF4LHC Working Group Interim Report*, arXiv:1101.0536, 2011.

[91] MissMJ and Nasfarley88, *Standard Model of Elementary Particles*, in: Wikimedia Commons, URL: `http://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`.

[92] X. Rouby, *Introduction a la detection des particules*, URL: http://cp3.irmp.ucl.ac.be/~rouby/files/XavierRouby_Cours1.pdf.

[93] Siegfried Bethke, *Experimental Tests of Asymptotic Freedom*, arXiv:hep-ex/0606035.

[94] E2m, *Drell–Yan process*, in: Wikimedia Commons, URL: https://en.wikipedia.org/wiki/Drell\OT1\textendashYan_process#/media/File:Drell-Yan.svg.

[95] Elton J. G. Santos and E. Kaxiras, *Electric field dependence of the effective dielectric constant in graphene materials*, URL: http://scholar.harvard.edu/files/efthimios_kaxiras/files/santos-dielectric-talk_v032213mac.pdf.

[96] M. H. Seymour and M. Marx, *Monte Carlo Event Generators*, arXiv:1304.6677v1.

[97] B. Isildak, *Measurement of the differential djet production cross section in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, arXiv:1308.6064v1.

[98] F. Maltoni, *Predictive Monte Carlo tools for the LHC*, CERN Academic Training Lectures, 2012.

[99] G Fitter SM, *Results for the Global Electroweak Standard Model Fit*, URL: http://gfitter.desy.de/Figures/Standard_Model/2014_07_16_TablePlot_logo_large.gif.

[100] CERN, *The CERN accelerator complex*, URL: http://cds.cern.ch/record/1621894.

[101] CMS Collaboration, *CMS detector design*, URL: http://cms.web.cern.ch/news/cms-detector-design

[102] Pac72 and Mets501, *Pseudorapidity*, in: Wikimedia Commons, URL: https://en.wikipedia.org/wiki/Pseudorapidity.

[103] Kevin Stenson and Achintya Rao, *Reconstructing a multitude of particle tracks within CMS*, URL: http://cms.web.cern.ch/news/reconstructing-multitude-particle-tracks-within-cms.

[104] HEPHY,Institut fur Hochenergiephysik, *CMS inner tracker*, URL: http://www.hephy.at/en/research/projects/cms-inner-tracker/

[105] CMS Collaboration, *CMS ECAL Endcaps photographs*, URL: http://hepwww.rl.ac.uk/CMSvpt/bestphotos/UK/slides/UK$%$20RAL$%$20-$%$20VPT$%$20and$%$20Crystal.htm.

[106] CMS Collaboration, *CMS prepares for Pixel and HCAL upgrades*, URL: http://cms.web.cern.ch/news/cms-prepares-pixel-and-hcal-upgrades.

[107] CMS Italia, *Il sistema per muoni*, URL: https://web.infn.it/cms/images/stories/cms_slice.gif

[108] Ian C. Brock, Bonn Universitat, *B to D* l nu decay*, URL: http://www-zeus.physik.uni-bonn.de/~brock/feynman/vtp_ss06/chapter04/b_dstar_lnu.jpg

[109] CMS Collaboration, URL: http://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/.

[110] F.-P. Schilling, *Track Reconstruction and b-jet Identification*, URL:http://fpschill.web.cern.ch/fpschill/images/btag.png.