Università degli Studi di Trieste

Dipartimento di Fisica

Corso di Laurea Magistrale in Fisica Nucleare e Subnucleare

Measurement of the Z boson transverse momentum and pseudorapidity differential cross sections

Laureando:
Jacopo Magro

Relatore:
Prof. Giuseppe Della Ricca

Correlatore:
Dott. Emanuele Di Marco

Academic Year 2016-2017

Università degli Studi di Trieste

Dipartimento di Fisica

Corso di Laurea Magistrale in Fisica Nucleare e Subnucleare

Misura della sezione d'urto differenziale in momento trasverso e pseudorapidità del bosone Z

Laureando:
Jacopo Magro

Relatore:
Prof. Giuseppe Della Ricca

Correlatore:
Dott. Emanuele Di Marco

Anno Accademico 2016-2017

Seiryoku-zen'yō

Jita-kyō'ei

Jigoro Kano

Abstract

Il Modello Standard della fisica delle particelle è una teoria che descrive le particelle elementari e le interazioni a cui sono sottoposte, ed è una delle teorie più importanti della fisica moderna. Elaborato negli anni '60, nelle ultime cinque decadi è stato oggetto di numerosi test aventi lo scopo di verificarne l'accuratezza e la veridicità. I maggiori risultati ottenuti all'interno di questo modello furono la scoperta dei bosoni vettori W e Z nel 1983, da parte degli esperimenti UA1 e UA2, e del bosone di Higgs nel 2012 grazie agli esperimenti ATLAS e CMS. Sebbene descriva con successo molti processi di fisica delle particelle, rimangono escluse alcune evidenze sperimentali. Per cercare una risposta ad alcuni di questi quesiti, LHC, l'acceleratore di particelle più grande del mondo, continua il suo lavoro, migliorando la conoscenza che abbiamo del Modello Standard e cercando di superarlo. La misura della sezione d'urto dei processi del Modello Standard va proprio in tale direzione: fornisce importanti test sulla cromodinamica quantistica (QCD) e delimita alcuni parametri della teoria. In questa tesi si è studiata la sezione d'urto differenziale rispetto ad alcune osservabili cinematiche in eventi prodotti dal processo di Drell-Yan: un quark e un antiquark si annichiliscono creando un bosone Z o un fotone virtuale, che decade in una coppa di leptoni; tale studio ha permesso il confronto tra diverse funzioni di distribuzione partoniche (PDF). Queste funzioni esprimono la probabilità di trovare un partone (quark e gluoni) all'interno di un protone con una certa frazione di momento, ad una data scala di energia. Purtroppo non è possibile calcolare le PDF per via teorica, ma esse possono essere ricavate usando dati da differenti esperimenti. La conoscenza di tali funzioni è essenziale per poter determinare in modo accurato le condizioni iniziali delle particelle incidenti negli esperimenti di collisione. Le sezioni d'urto studiate in questa tesi sono espresse in funzione del momento trasverso, della pseudorapidità e della rapidità del bosone Z intermedio, e sono comparate con le previsioni teoriche fornite da tre diversi set di PDF: CTEQ6L1, CT10 e NNPDF21_100. Le coppie elettroniche usate nell'analisi sono state prodotte nelle collisioni protone-protone ad una energia nel centro di massa di 8 TeV, e sono state raccolte dall'esperimento CMS di LHC nel 2012 (corrispondente ad una luminosità integrata di 19.78 fb^{-1}). La tesi è divisa in cinque capitoli. Il primo capitolo presenta una breve introduzione al Modello Standard, descrivendo le interazioni elettrodeboli e forti, e concentrandosi in particolare sul bosone Z ed il processo di Drell-Yan, chiudendo con un sommario sui problemi irrisolti dal modello. Il secondo capitolo descrive il Large Hadron Collider e l'esperimento CMS, includendo l'esposizione dei vari sotto-rivelatori che compongono CMS: il tracciatore, i calorimetri, i rivelatori per i muoni e il magnete solenoidale. Nel terzo capitolo è presentato il set di dati utilizzato, con particolare attenzione per il sistema di ricostruzione delle particelle e la sua performance nella ricostruzione degli elettroni. Sono descritti inoltre i generatori della simulazione Monte Carlo, `MadGraph 5` e `PYTHIA 6`, e del rivelatore, GEANT4,

nonché i differenti set di PDF utilizzati. Il quarto capitolo è dedicato alla procedura di *unfolding*, ovvero la correzione degli effetti introdotti dal rivelatore nei dati, e la sua implementazione attraverso il pacchetto TUnfold del software ROOT, il quale è stato utilizzato per tutta l'analisi. Il quinto ed ultimo capitolo presenta l'analisi dei dati. All'inizio sono descritti i criteri di selezione e la stima del fondo, valutata dalla simulazione Monte Carlo: i principali contributi si devono alla produzione di bosoni vettori e di quark top. Sono quindi mostrati i grafici delle distribuzioni degli elettroni/positroni selezionati, in particolare quelli relativi alle quantità utilizzate nei tagli. Le distribuzioni dei bosoni Z ricostruiti sono riportate descrivendo anche le incertezze sistematiche associate. La procedura di unfolding è quindi applicata a queste distribuzioni dopo la sottrazione del fondo. Infine sono mostrate le sezioni d'urto differenziali in funzione del momento trasverso, della pseudorapidità e della rapidità del bosone Z, comparate con le previsioni teoriche. Il set CTEQ6L1 è risultato essere quello che descrive meglio i dati analizzati.

Contents

Introduction

The Standard Model of particle physics is a very important and fundamental theory of the modern physics, describing the elementary particles and their interactions. In the last 50 years it has been subject of several tests in order to validate it and put constraints on its parameters. The milestones of these validation tests are the discovery of the W and Z bosons in 1983 and of the Higgs boson in 2012. However, many aspects of the world around us are unclear and they are not explained in the Standard Model. In order to improve our understanding of some of these phenomena, the Large Hadron Collider carries on its work after the Higgs boson discovery, searching for new physics beyond the Standard Model and perfecting the model itself. The measurement of the cross section of Standard Model processes goes in this direction, providing sensitive tests of Quantum Chromodynamics and precise determination of Standard Model parameters. In this thesis the cross section for Drell-Yan events as a function of some kinematic variables is studied, allowing a comparison between different Parton Distribution Functions. The knowledge of these functions is useful not only in the Drell-Yan process, but in general in Standard Model processes and in searches for the physics beyond it. Three different differential cross sections are studied: the one as a function of the transverse momentum of the intermediate Z boson, the one as a function of the pseudorapidity of the intermediate Z boson and the one as a function of the rapidity of the intermediate Z boson. They are compared with three different sets of Parton Distribution Functions. Electron pairs provided from proton-proton collisions at a center of mass energy of $\sqrt{s} = 8$ TeV, delivered by LHC and recorded with the CMS experiment in 2012 (corresponding to an integrated luminosity of about 19.78 fb^{-1}) are used to perform the measurement of the differential cross sections. This thesis is divided in five chapters. The first chapter presents a brief theoretical introduction to the Standard Model, describing the electroweak interaction and the strong interactions sector, focusing on the Z boson and the Drell-Yan process. The chapter ends with a schematic summary of the unsolved problems of the model. The second chapter introduces the Large Hadron Collider and the CMS experiment, illustrated with its different subdetectors: the tracker, the electromagnetic and hadronic calorimeters, the muon system and the solenoidal magnet. The third chapter describes the data set employed in the analysis, the algorithm for the particle reconstruction and its performance. Moreover, the Monte Carlo generators and the different sets of PDFs employed are illustrated. The fourth chapter is dedicated to the unfolding procedure and its implementation using the TUnfold package. The fifth and last chapter reports the data analysis. First of all the cuts for the selection of the electron pairs from the data are exposed with the background estimation. The dielectron distributions are then shown, in particular the ones related to the selection cuts. The distributions of the reconstructed Z boson are displayed with a discussion on the different systematic uncertainties presented in the analysis. The unfolding procedure is then applied to the data and the results are shown. Finally the differential cross section as a function of the transverse momentum, pseudorapidity and rapidity of the Z boson is reported with a comparison between the theoretical predictions provided by the different PDFs sets.

Chapter 1

Standard Model

The Standard Model of particle physics (SM) is a gauge quantum theory that describes the elementary particles and interactions in terms of the quantum field dynamics. The SM was elaborated between 1960 and 1967 by S. Glashow [1], S. Weinberg [2] and A. Salam [3], and it can explain all the known particles, combining electroweak and strong interactions. Moreover, in the following years, the Higgs mechanism was included in the theory. Several important experimental confirmations to the SM were found starting from 80's, establishing this theory as the reference theory for particle physics. The most important, but not the only, confirmations were the discovery of the W and Z bosons at CERN by Rubbia and Van Der Meer in 1983 [4] and the discovery of the Higgs boson at CERN by ATLAS and CMS collaborations in 2012 [5]-[6]. In this chapter the Standard Model will be briefly introduced, discussing in particular the aspects that are more relevant for this thesis.

1.1 Particles and interactions

The fundamental constituents of matter interact through four types of forces: the electromagnetic, the weak, the strong and the gravitational one. The SM describes all these interactions except for the gravitational force. However, this force is completely negligible in particle physics experiments. The elementary particles described by the SM are collected in Figure 1.1. There are two types of particles: fermions, with semi-integer spin, and bosons, with integer spin. Fermions can be divided into quarks and leptons according to their interactions. Quarks can interact through the electromagnetic, weak and strong force, being the latter the dominant one; there are six different flavours of quarks: up, down, strange, charm, bottom and top. Leptons are grouped into three families formed by a charged lepton and the corresponding neutrino: electron and ν_e, muon (μ) and ν_μ, tau (τ) and ν_τ. Charged leptons can interact through weak and electromagnetic force, while neutrinos can interact only through weak force. The three interactions considered in the SM are mediated by the force-carrier gauge bosons: the electromagnetic force by the photon (γ), the weak force by $W_\pm$ and Z_0, and the strong force by gluons. Gluons and γ are massless bosons, while W and Z are massive: $M_W = 80.385 \pm 0.015$ GeV and $M_Z = 91.1876 \pm 0.0021$ GeV [7].

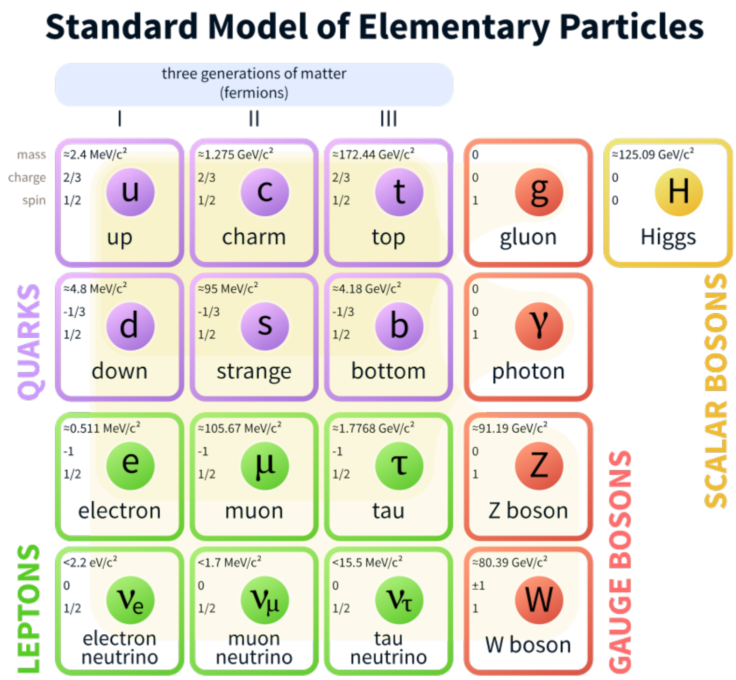

Figure 1.1: Scheme of the elementary particles in the SM [8].

1.2 Electroweak interaction

Electromagnetic and weak interactions are unified in the electroweak model describing the two different manifestations of the same fundamental interaction: $SU(2)_L \otimes U(1)_Y$, where $SU(2)_L$ is the weak isospin group which is non abelian and has three generators $T_{1,2,3} = \sigma_{1,2,3}$, that are the Pauli matrices, and $U(1)_Y$ is the weak hypercharge group which is abelian and has one generator: $Y/2$. The electromagnetic interaction is a subgroup of the electroweak group and its generator is a linear combination of the third component of the weak isospin with the weak hypercharge:

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

The weak interactions between leptons and their phenomenological aspects are described by the Fermi theory [9], in which one can consider that the Hamiltonian density of the interaction is the following:

$$\mathcal{H}_I(x) = \frac{G_F}{\sqrt{2}} J_\mu^\dagger J^\mu \tag{1.2}$$

i.e. the Fermi's coupling constant ($G_F = 1.16638 \pm 0.00001 \times 10^{-5}$ GeV^{-2} [7]) times the product between two charged currents. These currents can be expressed in terms of the Dirac spinors taking into account the parity violation:

$$J_\mu \equiv J_\mu^+ = \bar{u}_\nu \gamma_\nu \frac{1}{2}(1-\gamma^5) u_e = \nu_L \bar{\gamma}_\mu e_L \tag{1.3}$$

$$J_\mu^\dagger \equiv J_\mu^- = \bar{u}_e \gamma_\nu \frac{1}{2}(1-\gamma^5) u_\nu = e_L \bar{\gamma}_\mu \nu_L. \tag{1.4}$$

Using the spinors with an established helicity (a left-handed doublet L and a right-handed singlet R), $\psi_L = \begin{pmatrix} l \\ \nu_l \end{pmatrix}_L$ and $\psi_R = \begin{pmatrix} l \end{pmatrix}_R$, with $l = e, \nu, \tau$, the charged currents can be rewritten as:

$$J^+_\mu = \bar{\psi}_L \gamma_\nu \tau_+ \psi_L \tag{1.5}$$

$$J^-_\mu = \bar{\psi}_L \gamma_\nu \tau_- \psi_L, \tag{1.6}$$

where $\tau_\pm = \frac{1}{2}(\tau_1 \pm \tau_2)$. A neutral current J^3_μ and a weak hypercharged current J^Y_μ can be introduced in order to express the electromagnetic and the weak neutral currents as a linear combination of J^3_μ and J^Y_μ.

The weak isospin current

$$J^i_\mu = \bar{\Psi}_L \gamma_\mu \frac{1}{2} \tau_i \Psi_L \tag{1.7}$$

can be introduced in order to unify the weak and the electromagnetic interactions in a unique electroweak term:

$$-ig(J^i)^\mu W^i_\mu - i\frac{g'}{2}(J^Y)^\mu B_\mu, \tag{1.8}$$

where W^i_μ is an isotriplet of vector fields coupled with strength g to J^i_μ and B_μ is a single vector field coupled to the J^Y_μ current with strength $g'/2$. The gauge fields describing the massive charged bosons $W_\pm$ are therefore $W_\pm = \frac{1}{\sqrt{2}}(W^1_\pm \mp W^2_\pm)$.

The Lagrangian density expressing the unified electroweak theory is made up by an interaction term and a kinematic term:

$$\mathcal{L}_{EW} = \sum_f \bar{\Psi}_i \gamma^\mu D_\mu \Psi - \frac{1}{4} W_{\mu\nu} W^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{1.9}$$

where D_μ is the covariant derivative introduced in order to guarantee the invariance under local transformations.

In the following, the quark families that complete the electroweak theory are reported:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \; u_R, \; d_R; \; \begin{pmatrix} c \\ s \end{pmatrix}_L, \; c_R, \; s_R; \; \begin{pmatrix} t \\ b \end{pmatrix}_L, \; t_R, \; b_R.$$

As in the lepton case, the left-handed components of the fields are transformed as $SU(2)_L$ doublets and the right-handed fields are singlets of the group. The mass eigenstates for the left-handed doublets do not coincide with the eigenstates of the weak interaction of quarks; in fact, the quark fields enter into the charged current of the Standard Model in the form of mixed combinations. The transformation of the mass eigenstates (d, s, and b) into weak eigenstates (d', s' and b') is regulated by the CKM (Cabibbo-Kobayashi-Maskawa) matrix [10]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

a 3×3 unitary matrix, that can be parameterised by three mixing angles and a CP-violating phase. The transition probability between different quark flavours is described by the matrix elements, and it is proportional to $|V_{qq'}|^2$. Currently, the best determination of the magnitudes of the CKM matrix elements is [7]:

$$V_{CKM} \simeq \begin{pmatrix} 0.97434 & 0.22506 & 0.00357 \\ 0.22492 & 0.97351 & 0.0411 \\ 0.00875 & 0.0403 & 0.99915 \end{pmatrix}$$

so, as one can see, the off-diagonal elements are suppressed.

The electroweak model predicts massless particles: the introduction of explicit mass terms for fermions and gauge bosons inside the electroweak Lagrangian would violate the gauge symmetry. This prediction is in contradiction with observations, so it is necessary to consider a mechanism of spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y$ in order to explain particle masses and to preserve the gauge invariance. This mechanism was theorised by Higgs introducing a complex scalar field in the electroweak Lagrangian [11]. In this framework the $\mathrm{W}_{\pm}$ and the Z boson masses are originated, while the photon remains massless. Moreover, the masses of the fermions can be explained introducing also a Yukawa coupling term with the Higgs field. The Higgs mechanism led to the prediction of a new scalar particle, the Higgs boson, that was searched for over 50 years. Finally, as mentioned above, in 2012 the ATLAS and CMS collaborations discovered the Higgs boson at a mass close to 125 GeV [5]-[6].

1.3 Strong interactions

In addition to the electroweak theory, the Standard Model includes also the description of the strong interactions, through the Quantum Chromodynamics (QCD), a gauge theory based on the unbroken colour group $SU(3)$ [12]-[13].

The quarks are the fundamental constituents of the hadrons and they carry a fractional electric charge. The discovery of particles composed by three identical quarks, brought to the introduction of a new quantum number (colour) in order to avoid the violation of the Pauli exclusion principle: the quarks carry an additional degree of freedom which can take three different values (r, g, b). The colour symmetry $SU(3)_c$ implies new massless gauge bosons called gluons (the eight generators of $SU(3)$), which mediate the strong interaction; they couple only to colour charges.

The deep inelastic scattering experiments involving leptons and nuclei gave the first confirmation to the existence of a sub-structure in nucleons. In particular, some aspects were underlined: the quarks are fermions with 1/2 spin and carry colour charge and fractional electric charge, gluons exist and do not interact with electromagnetic or weak force.

1.3.1 Quantum chromodynamics

As stated before, the QCD is a gauge theory based on the group $SU(3)_c$. Using the quark Dirac spinors fields q_i^α (where α is the flavour index and i is the colour index), the Lagrangian of QCD can be written as:

$$\mathcal{L}_{QCD}(x) = \sum_{\alpha} \bar{q}_i^\alpha (i\gamma_\mu D_\mu - m_\alpha)_{ij} q_j^\alpha(x) - \frac{1}{4} F_{\mu\nu}^\alpha(x) F^{\alpha\mu\nu}(x) \tag{1.10}$$

where D_μ is the covariant derivative and $F_{\mu\nu}^\alpha$ is the field strength tensor.

The size of the strong coupling constant depends on the energy scale of the interaction: it is small for large momentum transfers (asymptotic freedom [14]), and large for small momentum transfers (colour confinement). At high energy quarks can be described as almost free particles, while the colour confinement at small energy binds quarks together, not allowing the observation of free quarks and gluons.

1.3.2 Parton distribution function

Considering two colliding protons, their structures can be described as a set of some point-like components (quarks and gluons), each carrying a fraction of the total proton momentum. The production cross section can be seen as the product of two terms: one at large momentum transfers (short distance) and one at small momentum transfers (long distance). The former can be analysed in the perturbative QCD frame, while for the latter one has to introduce a parametrisation through the parton distribution functions (PDFs).

For a proton A with a momentum p_A, the PDF is a function $f_{a/A}(x, Q^2)$ of the relative momentum ($x = \frac{p_a}{p_A}$) of the parton in the direction of the proton momentum, depending also on the energy scale of the process (Q^2). The proton-proton collision is then:

$$\sigma_{p_A p_B \to n} = \sum_q \int \mathrm{d}x_a \mathrm{d}x_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \sigma_{ab \to n}, \tag{1.11}$$

where A and B denote the two protons. A possible process that can happen in proton-proton collisions is the Drell-Yan process, that will be described in the next section.

It is possible to write the proton-proton cross section in terms of the strong coupling:

$$\sigma_{p_A p_B \to n} = \sum_q \int \mathrm{d}x_a \mathrm{d}x_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2)[\sigma_0 + \alpha_s(Q^2)\sigma_1 + \cdots], \tag{1.12}$$

where σ_0 is the tree-level parton-parton cross section, σ_1 is the first order QCD correction to the parton-parton cross section, and so on. In order to have a realistic model, higher orders of QCD corrections have to be employed: the virtual loop correction and the emission of quarks and/or gluons. The virtual loop correction does not affect the transverse momentum spectrum of the final state, while the emission implies an additional parton in the final state that enters in the balance of the transverse momentum.

As stated above, the partons can not be detected because of the colour confinement. Therefore, every parton generated in the interaction undergoes a hadronisation process that produces a jet of baryons and mesons that can be detected when interacts with matter. The direction of the hadrons produced in this process is the same of that of the initial parton and it is concentrated in a restricted region in η and ϕ. Moreover, the sum of the energy of these particles is equal to the energy of the starting parton.

The PDF is the probability density for finding a parton with a momentum fraction x_a of the proton momentum at a given factorisation scale μ_F. Their determination is important in order to better understand the initial state of the colliding particles. The PDFs can not be directly calculated, but can be determined using data from different experiments: deep inelastic scattering, Drell-Yan events and jet production. Three major groups, NNPDF, CTEQ and MRST, provide updates to

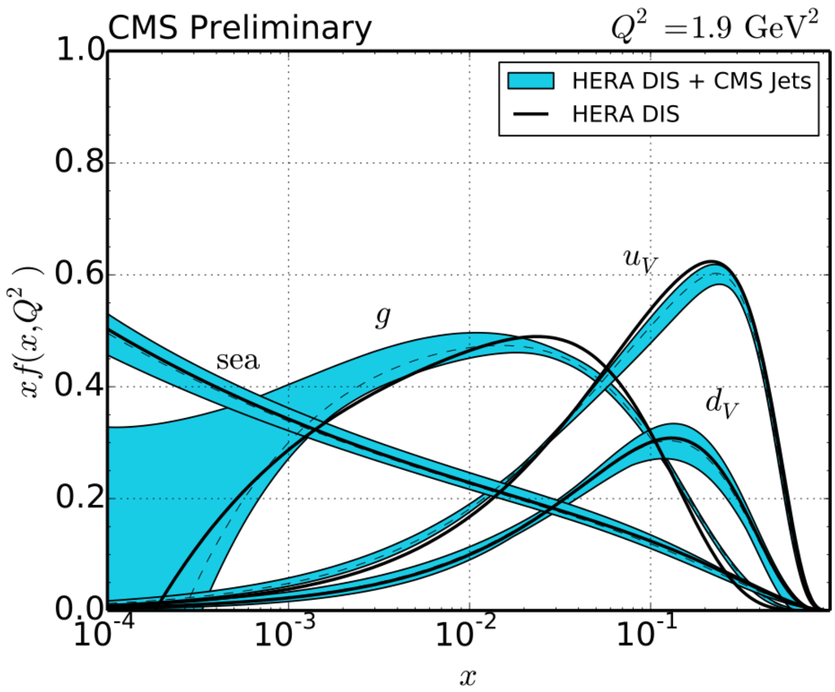

Figure 1.2: Overview of the gluon, sea, u valence, and d valence PDFs before (full line) and after (dashed line) the inclusion of the CMS inclusive jet data to DIS data from HERA into the global fit. The PDFs are shown at the starting scale $Q^2 = 1.9$ GeV^2. The total uncertainty, including the CMS inclusive jet data, is shown as a band around the central fit [17].

the PDFs when new data or theoretical developments become available. In addition, there are also PFDs available from the two HERA experiments, where the PDFs are studied using the deep inelastic scattering [15]. At LHC the study is done using Drell-Yan events and W/Z decays. The fits to various sets of experimental data are performed within the evolution scheme,

$$\frac{\partial f(x, Q^2)}{\partial log(Q^2)} = \frac{\alpha_s}{4\pi} \int_x^1 \frac{\mathrm{d}z}{z} [P_{aa'} f(\frac{x}{z}, Q^2)], \tag{1.13}$$

where $P_{aa'}$ is the Altarelli-Parisi splitting function [16] and gives the probability to have transformation of parton with momentum x into another with momentum z, as a consequence of the emission of one or more quarks or gluons. So, the PDFs measured at one scale can be used to predict the results of experiments at other scales using the evolution equation 1.13.

1.4 Z boson

The Z and W bosons are the mediators of the weak interaction. Discovered in 1983 [18], the Z boson mediates the neutral current interaction, in which the emission or absorption of a Z boson can only change the spin, momentum, and energy of the particle, but not the charges. It is its own antiparticle, so its quantum numbers and charges are zero. Having a mass of 91.1876 ± 0.0021 GeV [7], this puts a limit to the range of the weak force.

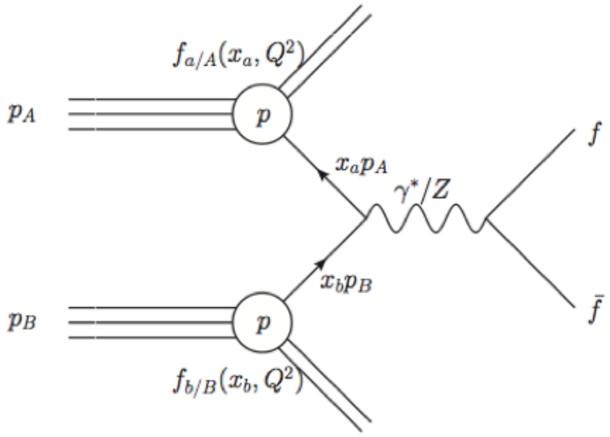

Figure 1.3: Schematic representation of the Drell-Yan process [19].

The Z boson decays into a fermion and its antiparticle. The experimental measured branching ratios are the following [7]:

- hadrons (69.91±0.06)%;
- neutrinos (20.00±0.06)%;
- e^+e^- (3.363±0.004)%;
- $\mu^+\mu^-$ (3.366±0.007)%;
- $\tau^+\tau^-$ (3.370±0.008)%.

At LHC, Z bosons are produced mostly in Drell-Yan process: a quark and an antiquark, belonging to the incoming protons, annihilate, creating a virtual photon or Z boson, that decays in a pair of opposite charged leptons $q\bar{q} \to Z/\gamma^* \to l^+l^-$ (see Figure 1.3).

The Z boson, since its discovery, has been object of several and detailed measurements, both at hadron and at e^+e^- colliders. In particular, many properties were investigated by the ALEPH, DELPHI, L3 and OPAL experiments at the large electron-positron collider (LEP) at CERN [20].

At Tevatron the single boson production was explored in hadron collisions with $\sqrt{s} = 1.96$ TeV by CDF [21] and D0 [22] experiments. Moreover, the production of a Z boson in association with jets and with heavy flavour quarks was investigated. Nowadays, the Large Hadron Collider gives the possibility to extend these studies to higher energy, allowing new tests of perturbative QCD and parton distribution functions.

Measurements of the cross section as a function of the transverse momentum, rapidity and pseudorapidity of the intermediate boson in Drell-Yan events provide a very sensitive test of QCD. Precise measurements of the differential cross section allow comparisons between calculations employing different sets of PDFs and underlying theoretical models [23]. This is a very crucial and interesting point because no theoretical prediction at hadron colliders is possible without PDFs; moreover, PDF uncertainties are essential for precision determination of Standard Model parameters. Furthermore, the understanding of Drell-Yan lepton pair production is important in the study of several Standard Model processes, as well as in searches for the physics beyond it.

The production of a Z boson in Drell-Yan events has a large rate and a very clear signature. Since it can also be calculated to high accuracy within the Standard Model, this process can thus be used to determine quantities of fundamental importance, such as PDFs. For example, the distribution of the transverse momentum of the Z boson is sensitive to the gluon and the light-quark PDFs in the not-so-well constrained intermediate Bjorken-x region. Also the Higgs production cross section at LHC is sensitive to the same PDFs in the same region: therefore the study derived from the Z boson reflects on the studies in the Higgs sector. The potential for intermediate bosons transverse momentum measurements to provide valuable constraints on PDF determinations has been already considered in Ref. [24].

1.5 Beyond Standard Model

The Standard Model, although describes successfully a lot of particle physics processes, does not explain some experimental observations:

- the instability of the value of the Higgs boson mass at energies far above the electroweak scale. This is called *"hierarchy problem"*;
- the SM does not propose any proper candidate for the presence of the Dark Matter in the universe. Cosmological observations reveal us that only 4% of the energy in the universe is bright matter, that can be described within the SM. About 27% should be Dark Matter, a type of matter that interacts only gravitationally but not electromagnetically. Possible candidates for the Dark Matter can be primordial black holes, supersymmetric particles and massive neutrinos (weakly interacting massive particles - WIMPs);
- neutrinos are treated as massless particles and their flavour mixing is not considered;
- the prevalence of matter over antimatter. In fact the SM predicts an equal amount for matter and antimatter, but there is obviousness of the opposite;
- the gravitational force is not included in the SM, although an explanation could be the graviton, an hypothetical particle, not discovered yet, that mediates the gravitational force.

There is so the need to new theories to go beyond the Standard Model frame. Some attempts are done in the Supersymmetry (SUSY) and in the Grand Unified Theory (GUT) frames, but there is still no experimental evidence sustaining these theories.

Chapter 2

LHC and the CMS experiment

The search for experimental confirmations to the Standard Model brought to the development of new technologies and experimental setups. As the precision tests were performed, providing more stringent constraints on the Standard Model parameters, new energy regimes required to be explored. The Large Hadron Collider (LHC) was built in order to ask to this need. LHC is a proton-proton collider started up in 2008 and it is the latest addition to CERN's accelerator complex. The accelerator characteristics will be summarised in the first part of this chapter, while the experiment Compact Muon Solenoid (CMS), that collected the data for this thesis, will be described in the second part.

2.1 Large Hadron Collider

The Large Hadron Collider [25] is the world's largest and most powerful particle accelerator. It was built between 1998 and 2008 near Geneva, Switzerland. It has been installed in the underground tunnel which housed the Large Electron-Positron Collider (LEP), a 27 km ring at an average depth of 100 m. Being a hadronic collider, LHC can reach a higher energy in the center of mass frame with respect to LEP, because less energy is lost due to the synchrotron radiation. It can accelerate two proton beams, circulating in opposite directions, at an energy of 14 TeV in the center of mass; it can also accelerate heavy ions (lead) beams at a center of mass energy of $\sqrt{s} = 5.5$ TeV. The beams circulate in two different pipes in a regime of ultra high vacuum (10^{-7} Pa), and they are bent by superconducting magnets cooled with liquid helium to a temperature of -271.3 °C.

There are four points of collision, where the following experiments are situated:

- ATLAS (A Toroidal LHC ApparatuS);
- CMS (Compact Muon Solenoid);
- ALICE (A Large Ion Collider Experiment);
- LHCb (Large Hadron Collider beauty).

There are also three further experiments: LHCf (Large Hadron Collider forward), TOTEM (TOTal Elastic and diffractive cross section Measurement) and, from 2010, MOEDAL (MOnopole and Exotical Detector At the LHC). The principal aim of these experiments is the exploration of the high energy physics:

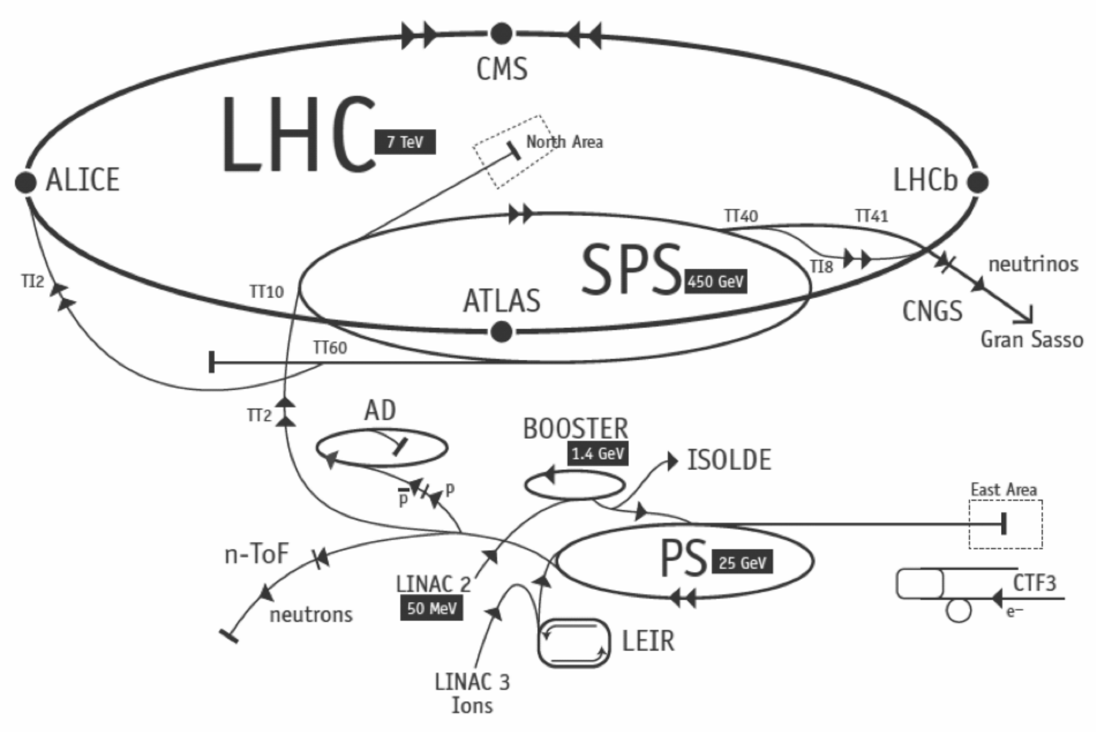

Figure 2.1: The complex of CERN's accelerators with the four main LHC experiments. Some lines are used to extract particles for non-LHC experiments [26].

- ATLAS and CMS have been designed to study as many as possible events related to the proton-proton collisions. They are both multifunction detectors that investigate mainly the Standard Model and the Higgs boson physics, and search for possible supersymmetry (SUSY) physics;
- ALICE has the goal of to study the strong interactions between heavy ions in high energy density regime (when matter is in a quark-gluon plasma state);
- LHCb has been built to examine CP violation in hadrons with bottom quarks;
- TOTEM is placed near CMS and its aim is to increase our knowledge on protons, measuring those that emerge from the collision region with a small angle and that the other experiments can not reveal;
- LHCf is situated near ATLAS and it has the purpose of a better understanding on cosmic rays. It uses the particles produced in collisions to simulate cosmic rays in laboratory conditions;
- MOEDAL is the last experiment built (near LHCb) and it searches for the magnetic monopole.

2.1.1 Accelerator structure

As one can see in Figure 2.1, the proton beam acceleration is performed in separated stages. The protons are produced by removing electrons from a source of hydrogen gas and then injected in the LINAC2, a 36 m long linear accelerator, that, using

Figure 2.2: Panoramic view of the LHC complex [26].

radiofrequency quadrupoles and focusing quadrupole magnets, brings protons to the energy of 50 MeV. The beam is then conducted to the Proton Synchrotron Booster (PSB), a 157 m long circular accelerator, in which the beam energy is increased up to 1.4 GeV. Then protons reach the Proton Synchrotron (PS), a ring of circumference equal to 628 m, that raises the protons energy up to 25 GeV. In these stages, the protons are divided in bunches interspersed by 25 ns (i.e. at the frequency of 40 MHz), in which there are about $1.1 \cdot 10^{11}$ particles per bunch. Before entering in the LHC ring, the beam is further accelerated by the Super Proton Synchrotron (SPS, a 7 km circumference synchrotron) to the energy of 450 GeV.

The particles in LHC are bent into a circular orbit by a 8.36 T magnetic field, transverse to beam direction, generated by 1232 magnetic dipoles. As stated before, the magnets are cooled to 1.9 K with a cryogenic system that uses superfluid ^{3}He, and their field is very complicated, since the beams are accelerated together in two different but nearby pipes (see Figure 2.3). The beams are also focused and collimated before the collision using quadripolor, sextupolar and octupolar magnets.

The energy of the bunches is increased in the accelerator by eight radiofrequency cavities, that admit a slightest spacing of 25 ns. The high density of particles in bunches produces a high average number of proton-proton interactions for each collision (bunch-crossing): this phenomenon, called *pile-up*, is very important in data analysis. In a single bunch-crossing there are about 20 interactions (8 TeV - 2012), so in total the frequency of collisions is about 10^9 Hz.

The collision rate is quantified in terms of the instantaneous luminosity (number of collisions per unit time and transverse section of the beams), which depends only on the beam parameters:

$$L = \frac{N_b^2 n_b f_r \gamma}{4\pi\epsilon_n\beta^*}F, \tag{2.1}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_r the revolution frequency, γ the relativistic gamma factor, ϵ_n the normalised transverse beam emittance, that describes the dispersion of the protons in the phase space, β^* the beta function at the collision point, that describes the oscillation of the bouches around the ideal orbit, and F the geometric luminosity reduction factor due to the crossing angle at the interaction point.

The total number of interactions in a data acquisition is given by the integrate luminosity, which depends on the instantaneous luminosity and on the acquisition time. In 2012 the integrated luminosity was 19.8 fb^{-1}, with an uncertainty of 2.6%.

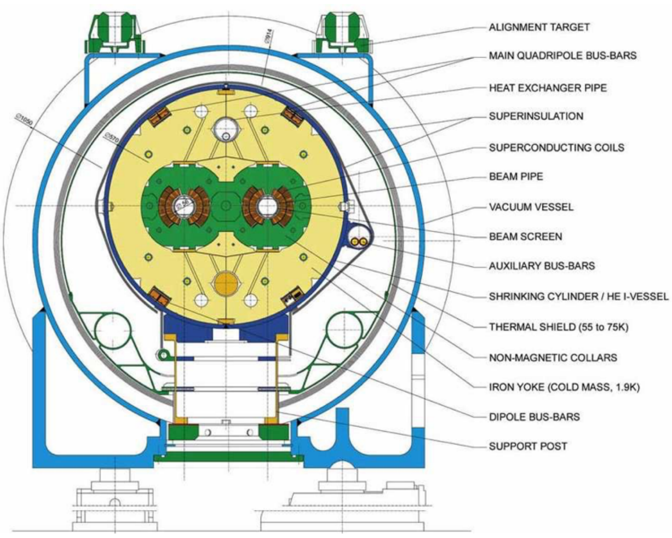

Figure 2.3: Section of one of the magnetic dipoles used to bend and stabilise the beam orbit [26].

2.1.2 Coordinate system

The coordinate system adopted by LHC is defined in the following way. In a collision point the x axis is identified by the intersection between the plane where is placed LHC and the plane perpendicular to the beam in the interaction point: the positive direction points to the ring center. The y axis is directed vertically upward and the z axis points along the beam direction (toward the Jura mountains from CMS).

The cylindrical symmetry of the LHC experiments (in particular of CMS) suggests to use a pseudo angular reference frame, given by the triplet (r, ϕ, η):

- $r = \sqrt{x^2 + y^2}$ is the distance from the z axis;
- $\phi = \cot \frac{x}{y}$ is the azimuthal angle;
- $\theta = \arctan \frac{y}{z}$ is the polar angle measured from the z axis;
- $\eta = -\ln\left(\tan \frac{\theta}{2}\right)$ is the pseudorapidity, a quantity that is invariant under Lorentzian boosts along the axis direction.

The pseudorapidity, for high energies, is a good approximation of the rapidity (y) of a particle, defined as:

$$y = \frac{1}{2} \ln\left(\frac{E + p_L}{E - p_L}\right), \tag{2.2}$$

where E is the particle's energy and p_L is the component of its momentum projected along the beam axis. The pseudorapidity is zero for $\theta = \frac{\pi}{2}$, grows up approaching the beam pipe, and tends to infinity for $\theta = 0$ as shown in Figure 2.4.

In hadron collisions the momenta of the partons inside the protons are unknown. It is assumed that the transverse component is negligible compared to the longitudinal component. In the transverse plane it is possible to apply the four momentum conservation and one can define the following transverse kinematic variables:

- $p_T = p \sin\theta$ is the transverse momentum of a particle;
- $E_T = E \sin\theta$ is the transverse energy of a particle.

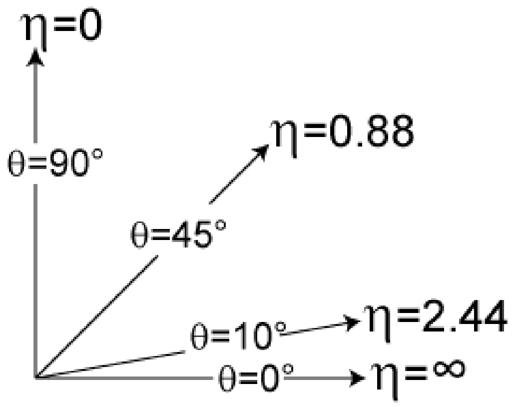

Figure 2.4: Variation of the pseudorapidity as a function of θ [27].

2.2 Compact Muon Solenoid

The CMS experiment [28] is a multi-purpose detector built in the underground cave of LHC, near Cessy, France. As stated before its aims are the exploration of the electroweak symmetry breaking in Standard Model, the measurement of QCD events and the search for new physics beyond the Standard Model (SUSY, extra dimensions, etc.). In order to study these phenomena CMS has to satisfy the following requirements:

- good muon identification and momentum resolution over a wide range of momenta and angles, good dimuons mass resolution (about 1% at 100 GeV), and the ability to determine unambiguously the muon charge with $p < 1$ TeV;

- good charged particle momentum resolution and reconstruction efficiency in the inner tracker. Efficient triggering and offline tagging of τ's and b-jets, requiring pixel detectors close to the interaction region;

- good electromagnetic energy resolution, good diphotons and dielectrons mass resolution (about 1% at 100 GeV), wide geometric coverage ($|\eta| < 2.5$), measurement of the direction of photons and correct localisation of the primary interaction vertex, π^0 rejection, and efficient photon and lepton isolation at high luminosities;

- good missing transverse energy and dijets mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage ($|\eta| < 5$) and with fine lateral segmentation.

In addition, CMS has to have an on-line event selection system to reduce the number of acquired collisions from 10^9 Hz to 100 Hz, a high granularity and good time resolution to reduce the pile-up effect and it must be radiation-hard to stand the high radiation level produced in collisions. CMS has a cylindrical structure, called barrel, closed by two endcaps: in this way the detector surrounds entirely the interaction point. The overall dimensions of the CMS detector are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12500 t.

As one can see in Figure 2.5 the detector presents a layer structure:

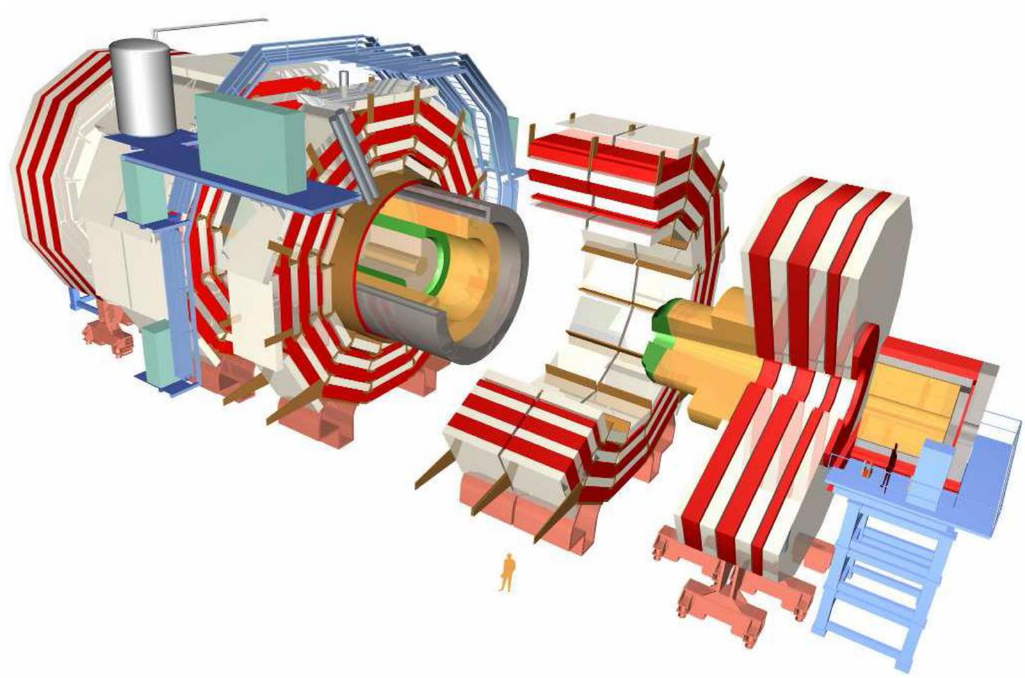

Figure 2.5: Overview of the CMS structure. In the inner part, the silicon tracker (in grey). Externally there are the electromagnetic calorimeter (in green) and the hadron calorimeter (in orange). Outside there are the superconducting solenoidal magnet (in dark grey) and the muon chambers (in red) housed inside the iron structure of the return yoke that encloses the magnetic field [29].

- Silicon Tracker. Placed in the region $r < 1.2$ m with $|\eta| < 2.5$, it consists of a silicon pixel vertex detector and a surrounding silicon microstrip detector. It is used to reconstruct charged particle tracks and vertices;

- Electromagnetic Calorimeter (ECAL). Placed in the region between 1.2 m and 1.8 m with $|\eta| < 3$, it consists of scintillating crystals of lead tungstate ($PbWO_4$) and it is used to measure the trajectory and the energy released by photons and electrons. The ECAL thickness, in radiation lengths, is larger than 25 X_0;

- Hadron Calorimeter (HCAL). Placed in the region between 1.8 m and 2.9 m with $|\eta| < 5$, it consists of brass layers alternated with plastic scintillators and it is used to measure the direction and the energy released by the hadrons. The HCAL thickness, in interaction lengths, varies in the range 7-11 λ_I;

- Superconducting Solenoidal Magnet. Placed in the region between 2.9 m and 3.8 m with $|\eta| < 1.5$, it generates a 4 T magnetic field in the inner part along the direction of the beams. The magnetic field is closed with an iron yoke, 21.6 m long with a diameter of 14 m, involved in a residual magnetic field of 1.8 T in the opposite direction with respect to the central field;

- Muon System. Placed in the region between 4 m and 7.4 m with $|\eta| < 2.4$, it consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC) is used both in the barrel and in the endcaps. The muon chambers are

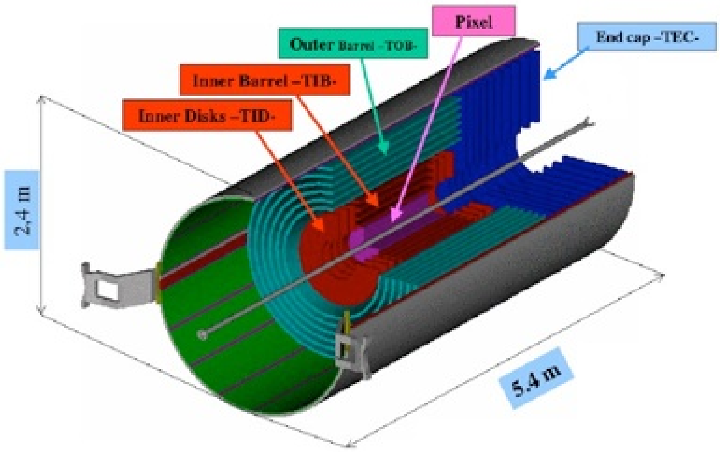

Figure 2.6: A representation of the inner tracking system of CMS [30].

housed inside the iron structure of the return yoke that encloses the magnetic field.

2.2.1 Tracker

The inner tracking system of CMS (see Figure 2.6) is designed to provide a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collisions, as well as a precise reconstruction of secondary vertices. It surrounds the interaction point and it is immersed entirely in the 4 T magnetic field of the solenoid. The requirements on granularity, speed and radiation hardness lead to a tracker design entirely based on silicon detector technology: three cylindrical layers of 100×150 μm^2 silicon pixels at a distance of 4.4, 7.3 and 10.2 cm from the beam, and a layer of silicon strips, for a total of 66 millions pixels and 15148 strip detector modules. The system is cooled to the constant temperature of -20 °C to reduce the damage from radiation.

The silicon strip detector is divided in four different sections. The Tracker Inner Barrel (TIB) is in the inner part and it is composed by four layers of strips with width variable between 80 μm and 120 μm; the TIB covers up to $|z| < 65$ cm with a resolution between 23 μm and 24 μm in r and ϕ. The Tracker Outer Barrel (TOB) is composed by six layers with larger dimensions and a resolution between 35 μm and 52 μm in r and ϕ; it covers up to $|z| < 65$ cm as the TIB. The Tracker EndCaps (TEC) and Tracker Inner Disk (TID) close the barrel, and they are composed by layers of different thickness.

2.2.2 Electromagnetic calorimeter

The electromagnetic calorimeter of CMS (ECAL) is a hermetic homogeneous calorimeter made of 61200 lead tungstate ($PbWO_4$) crystals mounted in the central barrel part, closed by 7324 crystals in each of the two endcaps, for a total of 75848 crystals. A preshower detector is placed in front of the endcap crystals. Avalanche photodiodes (APDs) are used as photodetectors in the barrel, while vacuum phototriodes (VPTs) are used in the endcaps, due to the different directions of the magnetic field.

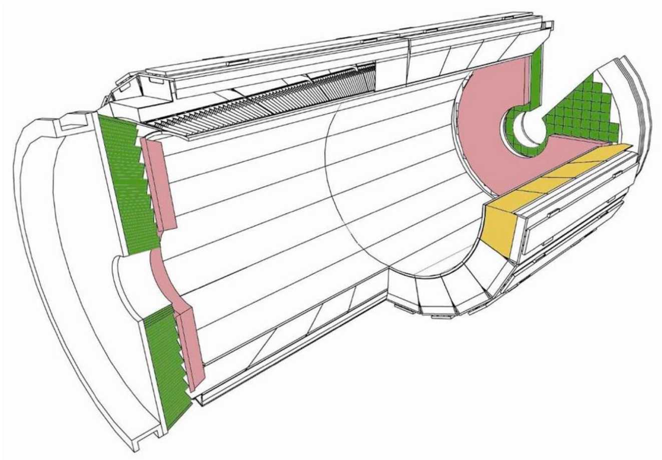

Figure 2.7: A representation of the electromagnetic calorimeter of CMS. The endcap crystals are represented in green, the barrel crystals in orange, and, in pink, the preshowers [28].

The choice of $PbWO_4$ is due to the high density (8.28 g/cm^3), short radiation length (0.89 cm) and small Molière radius (2.2 cm), in order to have a fine granularity and a compact calorimeter. The scintillation decay time is of the same order of magnitude as the LHC bunch-crossing time: about 80% of the light is emitted in 25 ns. The crystals emit a scintillation light with a peak at 425 nm.

ECAL is divided in barrel (EB) and endcaps (EE); the former covers the range $-1.479 < \eta < 1.479$, while the latter cover the range $1.479 < |\eta| < 3.0$. The crystals in barrel are grouped in submodules (5×2 =10 crystals), modules (400 or 500 crystals) and supermodules (20×85 = 1700 crystals). An aluminium lattice supports the structure's weight. Every crystal has the shape of a truncated pyramid with a front face cross section of 22×22 mm^2 and a rear face cross section of 26×26 mm^2, in order to subtend the same solid angle. The length is 230 mm, corresponding to 25.6 X_0. The crystals in endcaps are bigger (faces are 28.62×28.62 mm^2 and 30×30 mm^2, length is 220 mm) and are grouped in 5×5 arrays, supported by a carbon-fibre structure. Each endcap is divided into 2 halves. The preshower, which covers the range $1.653 < |\eta| < 2.6$, is a sampling calorimeter consisting of two lead converters followed by silicon strips with a pitch of less than 2 mm. The choice regarding the photodetector is due to the request of a fast response, radiation hardness and insensitivity to the magnetic field. The nominal operating temperature of the CMS ECAL is 18 °C maintained by a water cooling system.

The relative energy resolution can be estimated as [31]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E(GeV)}} \oplus \frac{12\%}{E(GeV)} \oplus 0.3\%, \tag{2.3}$$

where the three contributions correspond to the stochastic, noise, and constant terms. This result was confirmed by beam tests with 120 GeV electrons. The constant term, in particular, is the dominant contribution to the energy resolution for high energy electron and photon showers and it depends on non uniformity of the longitudinal light collection, energy leakage from the back of the calorimeter, single channel response uniformity and stability.

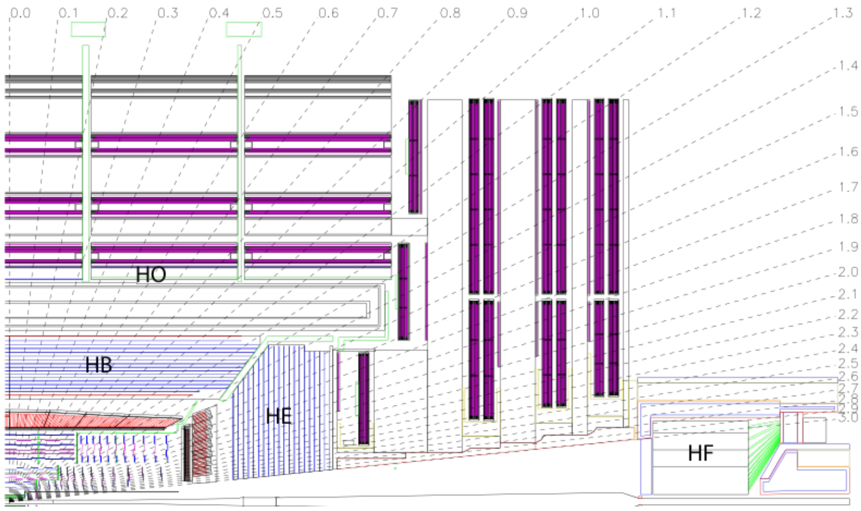

Figure 2.8: Longitudinal view of the CMS detector where the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters are highlighted [28].

2.2.3 Hadron calorimeter

The hadron calorimeter (HCAL) aims the measurement of hadron jets and particles resulting in apparent missing transverse energy. In order to meet these requirements, a high hermeticity, a good transverse granularity and a good energy resolution are needed. Figure 2.8 shows the longitudinal view of HCAL: it is placed between the ECAL (r = 1.77 m) and the inner face of the solenoid (r = 2.95 m), with an outer hadron calorimeter placed outside the solenoid, complementing the barrel calorimeter.

As the electromagnetic one, the hadron calorimeter is divided in barrel (HB) and endcaps (HE). The HB is a sampling calorimeter covering the range $|\eta| < 1.3$; the absorber is a 40 mm thick front steel plate, followed by eight 50.5 mm thick brass plates, six 56.5 mm thick brass plates, and a 75 mm thick steel back plate, while the active medium is a plastic scintillator divided in about 70000, 3.7 mm thick, tiles. It has a total thickness of 7-10 interaction lengths λ_I. The HEs cover the range $1.3 < |\eta| < 3.0$, and they are composed by absorber plates alternated with scintillator disks. The scintillation light is collected by wavelength shifting (WLS) fibres and multipixel hybrid photodiodes (HPD) are used as photodetectors due to their low sensitivity to magnetic fields and their large dynamical range.

In the central pseudorapidity region, the combined stopping power of EB and HB does not provide sufficient containment for hadron showers: for this reason the hadron calorimeter is extended outside the solenoid with a *tail-catcher* called the HO (hadron outer calorimeter); it was built with a geometry and a composition similar to the barrel ones.

Two additional hadron calorimeters, the forward calorimeters (HF), are placed outside the magnet yoke to cover the region $3.0 < |\eta| < 5.0$. In order to sustain the very high radiation dose of these regions, HF are composed by 5 mm thick steel absorber alternated with quartz fibres as the active medium. The signal is generated when charged shower particles produce Cherenkov light.

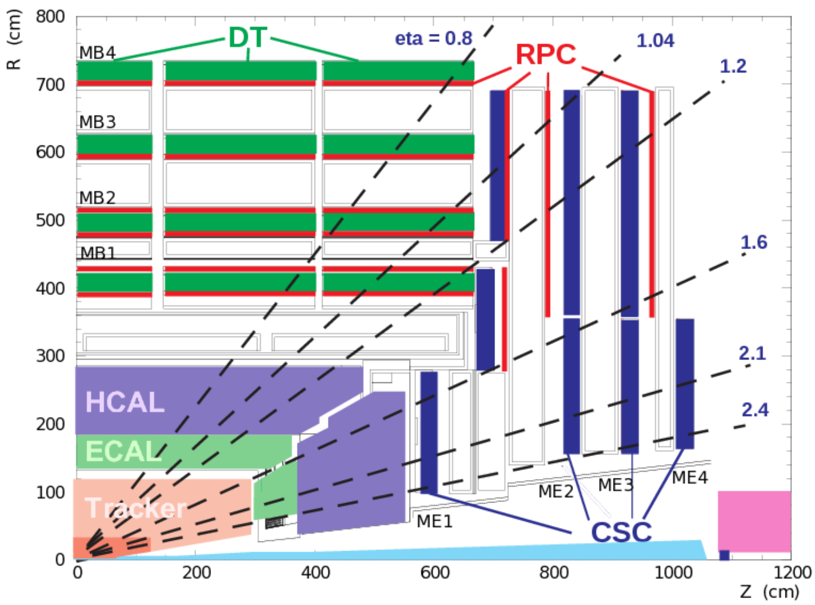

Figure 2.9: Overview of the CMS detectors where the muon system is highlighted [32].

2.2.4 Muon detectors

The muon system has 3 functions: muon identification, momentum measurement, and triggering. Good muon momentum resolution and triggering are enabled by the solenoid and its flux-return yoke. It is designed to reconstruct the muon momentum and charge over the entire kinematic range of the LHC. The tracker system does a first measurement, but for high momenta it can not be enough precise, so three type of detectors are placed in the region outside the solenoid (see Figure 2.9):

- Drift Tubes (DT) are divided in 4 stations and cover the barrel region $|\eta| < 1.2$. They are composed by a gas mixture of Ar (85%) and CO_2 (15%);
- Cathode Strip Chambers (CSC) are placed in the endcaps region and cover the region $0.9 < |\eta| < 2.4$. There are 4 stations of CSC for each endcap. Every CSC is a multiwire proportional chamber filled with a gas mixture of Ar (40%), CO_2 (50%) and CF_4 (10%);
- Resistive Plate Chambers (RPC) are a complementary subsystem both in barrel and in endcap regions. They are dedicated to the triggering operations. Every RPC is composed by a gas mix of $C_2H_2F_4$ (96.2%), iC_4H_{10} (3.5%) and SF_6 (0.3%), and, moreover, steam is added to the gas mixture to maintain a relative humidity of about 45%.

2.2.5 Trigger system

As mentioned above, the bunch-crossing frequency is 40 MHz (8 TeV - 2012). Obviously, such an amount of data produced from the collisions is impossible to be recorded, so it is necessary to perform a drastic cut on data before the storage. This is the task of the trigger system, that is structured in two steps: Level-1 Trigger (L1) and High Level Trigger (HLT). L1 is a hardware trigger, while HLT is a software trigger. The rate reduction is a factor 10^3 for both the levels, so the final frequency is about 100 Hz.

The L1 trigger uses the raw data from the calorimeters and muon chambers in order to cut the data flow from 40 MHz to 100 kHz, while holding the high-resolution data in pipelined memories in the front-end electronics. The L1 trigger decides in 3.2 μs if holding or discarding the data. The held ones are then processed by the HLT. The informations provided by the Calorimeter Trigger and the Muon Trigger are transferred to the Global Trigger, which, performing a first elementary muon and electron identification, takes the final accept-reject decision.

The HLT has access to the complete data and can therefore perform complex calculations similar to those made in the off-line analysis in order to select the interesting events and reduce the data flow to 100 Hz. Starting from the L1 accepted events, exploiting the high resolution data from all subdetectors the HLT performs a partial reconstruction of the event. Interaction vertices and high level objects, such as τ, are identified using simplified and faster versions of the algorithms used in the off-line data processing.

Chapter 3

Data and simulation settings

As already discussed in the first chapter, the study of the differential cross section of the intermediate boson in Drell-Yan events provides very sensitive tests of QCD and allows comparisons between different sets of PDFs.

In this chapter I will present the data set employed in this analysis in order to study the differential cross section of the Drell-Yan process, and I will describe the electron reconstruction in the CMS experiment. Moreover, the high level trigger (HLT) selection will be also presented.

In the last part of this section I will explain the generators and parameter settings for the Monte Carlo simulation, including the differences between the various sets of PDFs implicated in this analysis.

3.1 Data set

In this thesis I will present the measurement of the transverse momentum and pseudorapidity distributions of the e^+e^- final state in Drell-Yan events at the LHC center of mass energy of $\sqrt{s} = 8$ TeV, selected in the mass range of 71-111 GeV/c^2 using proton-proton collision data collected with the CMS experiment in 2012, corresponding to an integrated luminosity of 19.8 fb^{-1}. Analyses employing muon pairs have already been performed at $\sqrt{s} = 7$ TeV [33] and 8 TeV [34] by CMS, while ATLAS has analysed the transverse momentum distribution employing lepton pairs at the same energies [35]-[36]. The data set was preselected with a electron pair trigger selection, called `Double Electron`. The different sub-periods that generate the full sample, according to the different timescale and technical conditions of the machine, are described in Table 3.1.

The candidate electrons from the `Double Electron` data set are requested to satisfy one of the two specific high level triggers, selected with asymmetric transverse momentum cuts:

- `HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloId_CaloIsoVL`;
- `HLT_Ele8_CaloIdT_CaloIsoVL_Ele17CaloId_CaloIsoVL`.

These triggers require at least one electron to have $p_T \geq 17$ GeV and at least another electron with $p_T \geq 8$ GeV; both electrons have to pass some minimal calorimetric identification cuts to reduce the huge background of QCD-induced electrons.

In the following, the algorithm and the performance of electron reconstruction of CMS will be described.

Run	Number of events	Integrated luminosity [fb^{-1}]
2012 A	12964286	0.889
2012 B	23571931	4.429
2012 C	33843769	7.152
2012 D	34526899	7.318
2012 Total	104906885	19.78

Table 3.1: Number of events and relative integrated luminosity for the partial runs of the total 2012 CMS statistics employed for the `Double Electron` primary data set.

3.1.1 Particle reconstruction in CMS

The Particle Flow (PF) algorithm [37]-[38] is the most used and performing technique for object reconstruction in CMS, so the electrons involved in this thesis are reconstructed with this approach. It reconstructs and identifies all the stable particles in the event (electrons, photons, muons, charged and neutral hadrons) with an accurate combination of all CMS subdetectors towards an optimal determination of their direction, energy and type. For each event it is possible to determine leptons, jets, missing energy (i.e. neutrinos) and to define tracks, isolation variables and quark flavours discriminators. The initial point is to take the calorimeter clusters or the track measurements and then combining the other subdetector informations.

A single track is defined as the sum of all the hits inside the tracker. The reconstruction starts with this track: a "seed" is generated if it satisfies some quality requirements on the vertex, pixel and silicon detectors measurements. The seed is treated with a Kalman Filter, an iterative fit technique that estimates the statistical compatibility of the hit with the track and finally extracts the first estimation of its momentum (p_T) and pseudorapidity (η). A series of quality cuts are then applied in order to optimise the track identification efficiency. Finally the track is given to the PF algorithm as input.

The calorimetric information is essential to measure and identify electromagnetic particles and hadrons. The energy deposits released by the different particles are clustered together if they have a minimum energy threshold forming then a "cluster seed". Two cluster seeds are merged when they have at least one calorimetric cell in common and if they have an energy 2 σ away from the measured calorimeter noise energy.

The track seeds and the calorimetric clusters are connected in the Particle Flow algorithm providing new reconstructed objects called "blocks". Bremsstrahlung photons emitted by electrons are also taken into account by the PF algorithm. The different particles have peculiar characteristics that can be exploited in the discrimination:

- the main feature of an electron candidate is a short track associated with significant bremsstrahlung photon emission inside the tracker. Since electrons may have a great bending in the 4 T axial field generated by the solenoid, the emission is widely spread out in the ϕ direction. However this is dependent on the electron p_T, directly proportional to the magnetic field bending power;

- CMS uses three different strategies to build up a reconstructed muon, the association between track hits measured in the muon spectrometer chambers and in the tracker, the tracker-only measurement and the spectrometer-only measurement. A global fit to the information coming from all the three methods combined statistically gives rise to a "global muon" that is taken by the algorithm as input and, if the combined momentum is within 3 σ of the tracker-only reconstruction, the corresponding track is removed from the block;

- charged hadron candidates are reconstructed by charged tracks linked to any number of calorimeter (ECAL or HCAL) clusters, and which are not identified as electrons;
- if the energy of the closest ECAL and HCAL clusters linked to a track is significantly larger than the total associated charged particle momentum, neutral particles might be present. If this energy is found to be larger than the total ECAL energy, a photon is created with this ECAL energy and a neutral hadron is created with the remaining part of the energy. The remaining ECAL and HCAL clusters, either originally not linked to any track or for which the link was disabled, are identified as photons and neutral hadrons, respectively;
- jets are simply constructed by applying the chosen algorithm taking as input the PF candidates list of the event within the cone of clusterisation;
- the missing transverse energy is defined as the imbalance in the plane transverse to the colliding proton beams. The imbalance can be due to particles that escape from the detector, like neutrinos, detector effects, or physics not from the collision, like cosmic rays.

3.1.2 Performance of electron reconstruction

The performance of CMS in the reconstruction and selection of electrons at 8 TeV has been studied in detail in Ref. [39]. The offline electron reconstruction is based on a combination of the energy measured in the ECAL and the momentum measured in the tracker, in order to optimise the performance over a wide range of transverse momenta. Electrons are reconstructed by associating a track reconstructed in the silicon detector with a cluster of energy in the ECAL. The electron energy deposits usually spread over several crystals of the ECAL. This spread depends also on the energy lost via bremsstrahlung before reaching ECAL. Two clustering algorithms, the "hybrid" algorithm in the barrel and the "multi-5×5" algorithm in the endcaps, are used for this aim.

The hybrid algorithm exploits the geometry of the ECAL barrel and the properties of the shower shape: starting from a seed crystal, defined as the one containing most of the energy deposited in any considered region, arrays of 5×1 crystals are added around the seed crystal if their energies exceed a minimum threshold of $E^{min}_{array} = 0.1$ GeV. The contiguous arrays are grouped into clusters, with each distinct cluster required to have a seed array with energy greater than a threshold of $E^{min}_{seed-array} = 0.35$ GeV in order to be collected in the final global cluster, called the supercluster (SC).

The multi-5×5 algorithm is used instead in the ECAL endcaps. It starts with the seed crystal, the one with local maximal energy relative to its four direct neighbours, and the energy is collected in clusters of 5×5 crystals. The clusters are then grouped in a supercluster if the total transverse energy is greater than 1 GeV.

The supercluster position is calculated as the energy-weighted mean of the cluster positions.

The electron tracks can be reconstructed using the standard Kalman filter track reconstruction procedure used for all charged particles [40]. However, the large radiative losses for electrons in the tracker material compromise the operation and so a dedicated tracking procedure is used. In order to save time, the procedure is initialised from seeds that are likely to correspond to the initial electron trajectories. The first step is the selection of two or three hits in the inner layers from which the track can be initialised. For this procedure two different algorithms are combined: the ECAL-based seeding and the tracker-based seeding. The overall efficiency of the seeding is greater than 95% for simulated electrons from Z boson decay. The selected electron seeds are used to start electron-track building, followed by track fitting: for each electron seed it proceeds iteratively from the track parameters provided in each layer, including one-by-one the information from each successive layer [40]. The electron energy loss is modelled through a Bethe-Heitler function. Only one missing hit is accepted for a trajectory candidate. Once the hits are collected, a Gaussian sum filter (GSF) fit is applied to estimate the track parameters. The energy loss in each layer is approximated by a mixture of Gaussian distributions.

This procedure provides electron tracks that can be followed up to the ECAL, finding the track parameters at the ECAL surface. The fraction of energy lost through bremsstrahlung is estimated as $f_{br} = (p_{in} - p_{out})/p_{in}$, where p_{in} is the momentum at the point nearest to the beam spot, and p_{out} is the momentum at the ECAL surface. The electron candidates are constructed from the association of a GSF track and a cluster in the ECAL. The overall efficiency is about 93% for electrons from Z decay.

Bremsstrahlung photons can convert into e^+e^- pairs within the tracker and be reconstructed as electron candidates. Since the bremsstrahlung photons have a small probability to convert in a pair just after their point of emission they can be mostly discriminated.

A natural choice for the charge estimate is the sign of the GSF track curvature, but this estimate can be altered by the misidentification probability in presence of conversions. In order to avoid this problem, two other estimates are used: one is based on the Kalman filter track reconstruction, while the other exploits the supercluster position. The electron charge is then defined by the sign shared by at least two of the three estimates, and is referred to as the *"majority method"*. The misidentification probability of this algorithm is predicted by simulation to be 1.5% for reconstructed electrons from Z boson decays. The charge misidentification probability decreases strongly with more restrictive identification selection, becoming almost negligible with the HLT criteria.

The electron momentum is estimated using a combination of the tracker and ECAL measurements and the electrons are classified according to their bremsstrahlung pattern: for most of the electrons, the bremsstrahlung fraction in the tracker, f_{br}, is complemented by the bremsstrahlung fraction in the ECAL. Electrons are classified in the following categories (see Figure 3.1):

- *Golden* electrons are those with little bremsstrahlung and consequently provide the most accurate estimation of momentum;
- *Big-brem* electrons have a large amount of bremsstrahlung radiated in a single

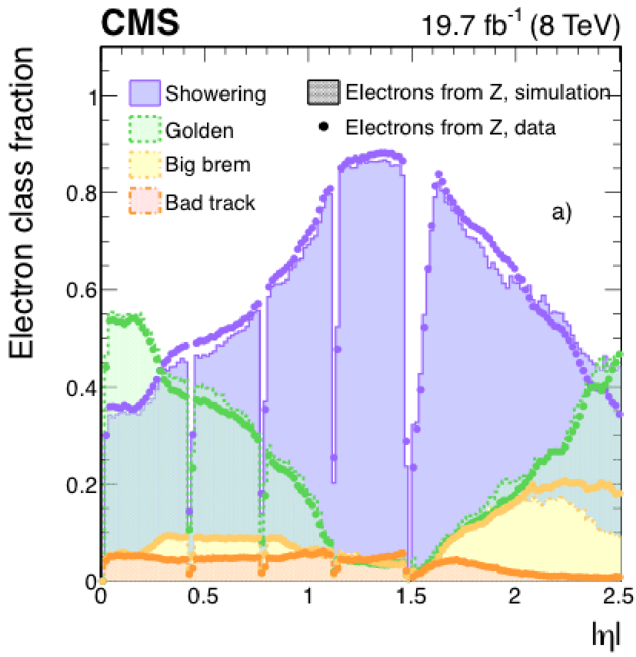

Figure 3.1: Fraction of population in different classes of electrons from Z boson decays as a function of $|\eta|$, for data (dots) and simulated (histograms) events. Crack electrons are not shown, but complement the proportion to unity [39].

step, either very early or very late along the electron trajectory;

- *Showering* electrons have a large amount of bremsstrahlung radiated all along the electron trajectory;
- *Crack* electrons are defined as electrons with the SC seed crystal adjacent to an η boundary;
- *Bad track* electrons have a poorly fitted track in the innermost part of the trajectory.

As analysed in Ref. [39], the overall momentum scale is calibrated with an uncertainty smaller than 0.3% in the p_T range from 7 to 70 GeV. For electrons from Z boson decays, the effective momentum resolution varies from 1.7%, for the barrel, to 4.5% for the endcaps. The reconstruction efficiency ranges from 88% to 98% in the barrel and from 90% to 96% in the endcaps in the p_T range from 10 to 100 GeV.

Several strategies are used in order to identify prompt isolated electrons (signal), and to separate them from background sources, mainly originating from photon conversions, jets misidentified as electrons, or electrons from semileptonic decays of b and c quarks. Simple and robust algorithms have been developed applying sequential selections.

A consistent fraction of background is due to misidentified jets or to electrons within a jet resulting from semileptonic decays of b or c quarks. In both cases, the electron candidates have significant energy flow near their trajectories, so, requiring electrons to be isolated from such nearby activity, greatly reduces these backgrounds. The isolation requirements are separated from electron identification: in this way they can be used as control of different sources of such backgrounds in data.

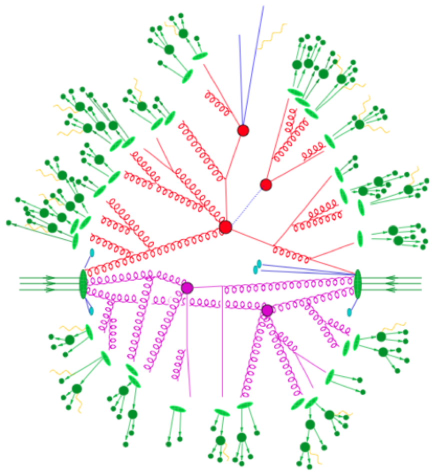

Figure 3.2: Schematisation of a proton-proton collision. The event represented contains the hard event followed by hard decays of two heavy unstable particles (red), and two more hard parton interactions (purple). The partons are dressed with secondary radiation as well, before the parton ensemble is transformed into primary hadrons which then decay further (green) [42].

Another important source of background to electrons arises from secondary electrons produced in conversions of photons in the tracker material. As stated before, in this case photons do not convert in the innermost layers of the tracker and the missing hits can be used to reject electron candidates from this background.

3.2 Monte Carlo generators

In order to validate the data results and to make a comparison between different sets of PDFs, a simulation of the signal events and of the various background sources is needed. The theoretical calculations described in the first chapter are implemented in Monte Carlo (MC) generators that provide an accurate description of the final state of the interactions considered. Different algorithms and generators are available, and for this thesis two of them are exploited: simulated events are produced by the matrix element calculation of `MadGraph 5`, while the hadronisation is simulated by `PYTHIA 6`. The detailed CMS detector simulation is based on the GEANT4 [41] program, that describes electromagnetic and hadronic interactions of particles with the CMS detector material. It is also the main tool for modelling the full CMS detector and geometry simulation. In simulation the proton-proton collision at the LHC is schematised in the following steps (see Figure 3.2):

- hard process;
- parton shower;
- hadronisation;
- underlying events;
- unstable particle decays.

Process	Cross section [pb]	Number of events
Z/γ^*	3531.8	$3.022836\cdot10^7$
$t\bar{t}$ + jets	225.197	6923750
$\bar{t}$ s-channel	1.76	139974
$\bar{t}$ t-channel	30.7	1935072
$\bar{t}$ tW-channel	11.73	493460
t s-channel	3.79	259961
t t-channel	56.4	3753227
t tW-channel	11.73	497658
W + jets	37509.0	$5.720585\cdot10^7$
WW + jets	5.995	1898738
WZ + jets	1.057·1.10	2007120
ZZ	8.059	9698911

Table 3.2: Theoretical predictions for the Standard Model cross sections [43]-[44] used in this analysis with the total number of generated events.

In Table 3.2 the complete list of simulated processes for signal and backgrounds employed in this analysis is presented, with the respective theoretical cross section and the number of events. The Drell-Yan process includes also the electrons coming from $Z \to \tau^+\tau^- \to e^+e^- + 4\nu$. This process is a background that will be evaluated and discriminated in the analysis step.

3.2.1 MadGraph

`MadGraph` [45] is a Monte Carlo event generator that simulates the hard scattering process instead of the whole event. So it needs to be completed by other generators (such as `PYTHIA`) for parton shower and hadronisation. The main advantage of using `MadGraph` is that it can simulate $2 \to n$ hard scattering processes without any approximation, unlike `PYTHIA`. The generator simulates matrix elements for decays of heavy particles, like Z boson, and $2 \to n$ hard scattering processes simultaneously creating the Feynman diagrams that correspond to these processes. The full reconstructed `MadGraph` simulation is used as default for the data to Monte Carlo agreement study, as well as for the background processes. The main input parameter settings for performing the calculations are as follows [46]:

- CTEQ6L1 PDF set;
- $\alpha_{QCD} = 0.007297352$;
- $m_Z = 91.1876$ GeV;
- $m_W = 80.398$ GeV;
- $m_t = 172.5$ GeV;
- $m_b = 4.8$ GeV;
- $m_c = 1.27$ GeV;
- $\Gamma_Z = 2.4952$ GeV;
- $\Gamma_W = 2.141$ GeV.

3.2.2 PYTHIA

`PYTHIA` [47] is a Monte Carlo event generator capable of modelling a large number of hard scattering processes. It is optimised for $2 \to 1$ and $2 \to 2$ hard scattering processes and it can also well approximate $2 \to n$ (>2) hard scattering processes. In each event simulation PYTHIA provides the incident particles, the initial state radiation, the hard scattering, the final state radiation, the hadronisation of coloured particles and the decay of the unstable particles, giving back a list of stable particles. Different parameters have to be adjusted to better fit some aspects of the data: a specific set of these parameters is referred to as "tune". For this work the Z2* tune has been used, exploiting the CTEQ6L1 parton distribution set.

3.2.3 GEANT

GEANT, acronym of GEometry ANd Tracking, is a toolkit for simulating the passage of particles through matter [41]. It covers the majority of the electromagnetic, hadronic and optical processes, and a wide range of materials. The energy scale starts from 250 eV to a few TeV. All aspects of the simulated process are included: the geometry of the system, the materials involved, the generation of primary particles of events, the tracking of particles through materials and external electromagnetic fields, the interactions between particles, the response of sensitive detector components, the generation of event data, the storage of events, the visualisation of the detector and particle trajectories, and the capture for subsequent analysis of simulation data. The version used in this analysis is GEANT4.

3.2.4 PDF sets

In this thesis the main set of parton distribution functions used is the CTEQ6L1 as suggested in Ref. [46]. In order to make a comparison, two other sets are considered: CT10 and NNPDF21_100 [48].

CTEQ6 is a LO set of PDFs (including CTEQ6L1) based on the deep inelastic scattering structure function measurements of H1 and ZEUS and on the inclusive jet cross section measurement of D0. These informations integrate the fixed-target DIS experiments, the CDF measurement of W boson asymmetry and the CDF measurement of inclusive jets, previously used in CTEQ5.

CT10 is an evolution of CTEQ, having 26 parameters, at the NLO, instead of the 22 of CTEQ. The CT10 analysis includes the precise experimental data in every major category of scattering processes: deep-inelastic scattering, vector boson production and single-inclusive jet production. It uses Run-I and Run-II (Tevatron) jet data sets, including other recent data from HERA and Tevatron experiments. The high luminosity Run-II boson asymmetry data by the D0 collaboration play a special role in this study. While being precise, they run into disagreement with some previous data sets and, in addition, they exhibit some tension among themselves. Because of these disagreements, two different PDF fits can be done: CT10, in which the D0 data on high luminosity are ignored, and CT10W, in which these data are exploited.

NNPDF is a NLO PDF set which employs, in addition, the LHC data on vector boson production and jets production. The approach adopted for the determination of parton distributions is based on a combination of a Monte Carlo method with the use of neural networks as basic interpolating functions.

Chapter 4

Unfolding

Under ideal conditions, with a perfect detector one could have a true measurement of physics quantities. However, with real detectors, distortions to distributions of the observables occur being subject to additional random fluctuations due to the limited resolution of the measuring device. The procedure of correcting for these distortions is known as unfolding [49]. By unfolding the distributions one provides a result which can directly be compared with those of other experiments as well as with theoretical predictions.

The principal distortions introduced by the detectors are:

- smearing due to the finite resolution and accuracy of the variable;
- non linear response of detector components;
- acceptance, limited and depending on the different variables to be measured;
- radiative effects at parton level (initial and final state radiation).

4.1 Unfolding procedure

Experiments are usually performed as counting experiments, where events are grouped into certain regions of phase-space, also called bins. However, as stated in the introduction of this chapter, the kinematic properties of each event are measured only at finite precision. As a consequence, events may be found in the wrong bin. Furthermore there is the presence of background, such that only a fraction of the events observed in a given bin originates from the process one is interested in.

In most of the cases, the use of algorithms and simulation generators can correct data for the detector effects making possible to confront the physics process modelled by the event generator with the background-subtracted data. However, often one is interested to report results independent of the detector simulation. In this case, the observed event counts have to be corrected for detector effects.

The problem can be written as:

$$\tilde{y}_i = \sum_{j=1}^{m} A_{ij}\tilde{x}_j, 1 \leq i \leq n, \tag{4.1}$$

where the m bins $\tilde{x}_j$ represent the true distribution, A_{ij} is a matrix of probabilities describing the migrations from bin j to any of the n bins on detector level (response

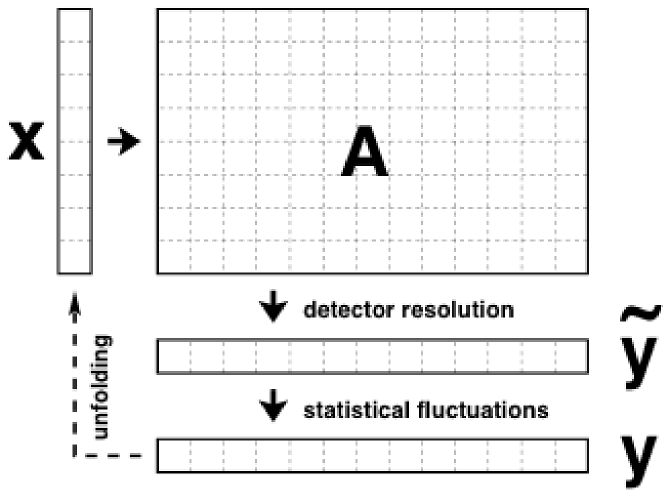

Figure 4.1: Schematic view of the unfolding procedure [50].

matrix) and $\tilde{y}_i$ is the average expected event count at detector level (see Figure 4.1). The observed event counts y_i may be different from the average $\tilde{y}_i$ due to statistical fluctuations.

In the case of background $\tilde{y}_i$ has an additional term:

$$\tilde{y}_i = \sum_{j=1}^{m} A_{ij}\tilde{x}_j + b_i, 1 \leq i \leq n \tag{4.2}$$

where b_i is the background contribution in the i-th bin. Both b_i and A_{ij} are affected by systematic uncertainties which have to be considered in addition to the statistical uncertainties.

Replacing $\tilde{y}_i \to y_i$ and $\tilde{x}_i \to x_i$ allows to solve simply by inverting the matrix of probabilities. Unfortunately, the statistical fluctuations of the y_i are amplified when calculating the x_j. Such fluctuations can be damped by imposing certain smoothness conditions on x_j, using the so called "regularisation" procedure.

The estimate for x_j obtained by inverting the response matrix is characterised by the acceptability of the solution: a region of the x_j-space around the maximum likelihood solution. In addition the measure of the smoothness of the solution is defined introducing a regularisation function $S(x_j)$. The general strategy is to choose the solution with the highest degree of smoothness out of the acceptable solutions. This strategy can be written as the maximisation of the following quantity with respect to α and x_j:

$$\alpha[\log L(x_j) - (\log L_{max} - \Delta \log L)] + S(x_j) \tag{4.3}$$

in which α is a Lagrange multiplier and it is the regularisation parameter. A commonly used measure of smoothness is the mean value of the square of some derivative of the true distribution. This technique is called Tikhonov regularisation [51]. Other common choices are regularisations based on the entropy, in particular on the Shannon-Jaynes entropy [49]. The method to determinate α depends on the numerical implementation.

4.2 TUnfold

The TUnfold algorithm [51] is a tool for correcting migration and background effects in high energy physics for multi-dimensional distributions [50]. It is based on a least square fit with Tikhonov regularisation; the L-curve method and scans of global correlation coefficients are implemented in order to determine the strength of the regularisation parameter. The program is interfaced to the ROOT analysis framework [52] and it is recommended by the CMS Statistics Committee [53]. In order to obtain best results, the number of degrees of freedom, $n - m$, has to be larger than zero: it means that the data distribution y_i has to be measured in finer bins than the simulated one, in contrast to some other commonly used unfolding methods.

4.2.1 Algorithm

The TUnfold algorithm gives an estimator of a set of truth parameters, using a single measurement of a set of observables, described by a vector of random variables, $\mathbf{y}$. The random variables $\mathbf{y}$ are taken to have a multivariate Gaussian distribution with mean $\tilde{\boldsymbol{y}} = \mathbf{A}\tilde{\boldsymbol{x}}$, where $\tilde{\boldsymbol{x}}$ is a vector corresponding to the set of the truth parameters and $\mathbf{A}$ is a matrix. The covariance matrix of $\mathbf{y}$ is $\boldsymbol{V_{yy}}$. The algorithm only works if the dimension of $\tilde{\boldsymbol{x}}$ is less or equal to the dimension of $\tilde{\boldsymbol{y}}$.

TUnfold determines the stationary point of the following Lagrangian:

$$L(x, \lambda) = L_1 + L_2 + L_3 \tag{4.4}$$

where

$$L_1 = (\mathbf{y} - \mathbf{A}\mathbf{x})^T \mathbf{V}_{yy}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x}) \tag{4.5}$$

$$L_2 = \tau^2 (\mathbf{x} - f_b x_0)^T (\mathbf{L}^T \mathbf{L})(\mathbf{x} - f_b x_0) \tag{4.6}$$

$$L_3 = \lambda(Y - e^T \mathbf{x}) \tag{4.7}$$

$$Y = \sum_i y_i \tag{4.8}$$

$$e_j = \sum_i A_{ij}. \tag{4.9}$$

The term L_1 comes from the least square minimisation. $\mathbf{y}$ has n rows, while $\mathbf{x}$ has m rows and correspond to the unfolding result. The covariance matrix $\boldsymbol{V_{yy}}$ is in general diagonal, but also non-diagonal matrices are supported. The elements A_{ij} of the response matrix describe for each row j of $\mathbf{x}$ the probabilities to migrate to bin i of $\mathbf{y}$.

The term L_2 describes the regularisation. The parameter τ^2 is the strength of the regularisation and it is considered as a constant. The matrix $\mathbf{L}$ has n columns and n_R rows, corresponding to the number of the regularisation conditions. $f_b x_0$ is a bias term with a normalisation factor (f_b).

The term L_3 is an optional area constraint, with a Lagrangian parameter λ and an efficiency vector $\mathbf{e}$. It is useful to limit biases on the normalisation.

The minimum of the stationary points of $L(x, \lambda)$ is determined by setting the first derivatives to zero, as explained in Ref. [50].

Generally, the response matrix $\mathbf{A}$ is determined from Monte Carlo simulations; an extra row is implemented in order to take into account generated events that are not reconstructed.

4.2.2 Choice of the regularisation strength

In the unfolding procedure the strength of the regularisation parameter, τ^2, is unknown. If it is too small, the unfolding result has large fluctuations, while, if it is too large, the result is biased by the bias term $f_b x_0$. In TUnfold the L-curve method is implemented in order to determine the best choice of τ^2.

The L-curve method searches in the graph of two variables (L_x^{curve} and L_y^{curve}) the point were the curvature is maximal. The definitions of these variables are the following:

$$L_x^{curve} = \log(L_1) \tag{4.10}$$

$$L_y^{curve} = \log(\frac{L_2}{\tau^2}). \tag{4.11}$$

In TUnfold this algorithm is implemented as the repetition of the unfolding for a number of points in $t = log\tau$, thus scanning the L-curve. The curvature of the L-curve is defined as

$$\mathcal{C} = \frac{d^2 L_y^{curve} dL_x^{curve} - d^2 L_x^{curve} dL_y^{curve}}{((dL_x^{curve})^2 + (dL_y^{curve})^2)^{\frac{3}{2}}}. \tag{4.12}$$

The first and second derivatives are approximated using cubic spline parametrisations of the scan results. The maximum of $\mathcal{C}$ is determined with the help of a cubic spline parametrisation.

A method of minimising global correlation coefficients is also implemented. For each point of the scan the correlation is calculated and the minimum is determined using again a cubic spline interpolation.

4.2.3 Choice of regularisation conditions

In TUnfold, three different types of regularisation are available:

- rows of $\mathbf{L}$ where only one element is non-zero, corresponding to a regularisation of the amplitude or size of $\mathbf{x}$;
- rows of $\mathbf{L}$ where two elements are non-zero, corresponding to a regularisation of the first derivative of $\mathbf{x}$;
- rows of $\mathbf{L}$ where three elements are non-zero, corresponding to a regularisation of the second derivative (curvature) of $\mathbf{x}$.

The first derivatives are approximated by differences of event counts in adjacent bins, $x_{i+1} - x_i$. Similarly, the second derivatives are approximated by $(x_{i+1} - x_i) - (x_i - x_{i-1})$. Every type can be chosen for all bins or it is possible to chose a different type bin by bin.

An example of the unfolding with the TUnfold algorithm is shown in the following Figure 4.2.

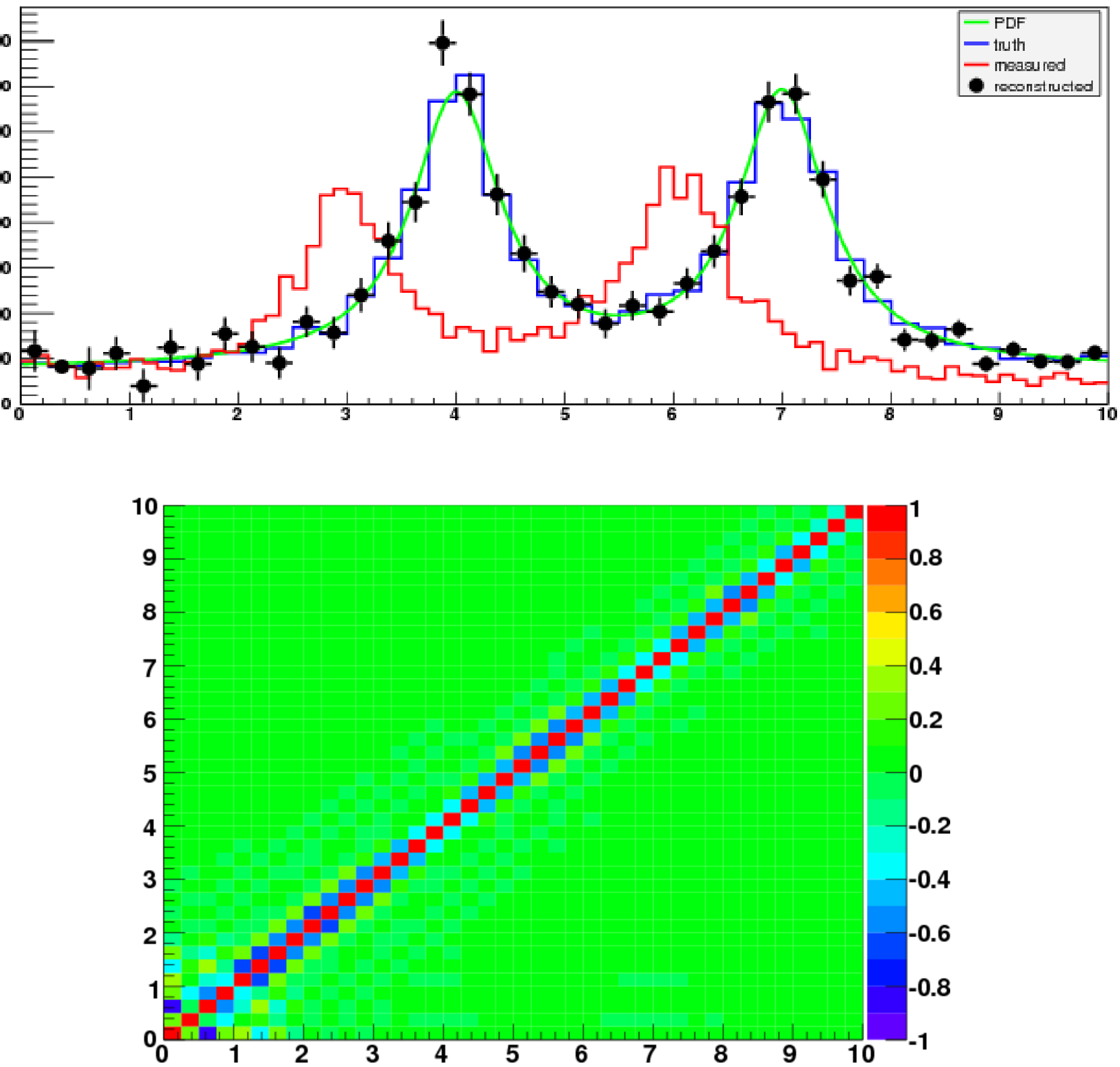

Figure 4.2: Unfolding with the TUnfold algorithm. On the top, a double Breit-Wigner PDF on a flat background (green curve) is used to generate a test "truth" sample (upper histogram in blue). This is then smeared, shifted, and a variable inefficiency applied to produce the "measured" distribution (lower histogram in red). Applying the TUnfold algorithm on this latter gave the unfolded result (black points), shown with errors from the diagonal elements of the error matrix. The bin by bin correlations from the error matrix are shown on the bottom. Here two measurement bins for each truth bin are used [54].

Chapter 5

Data analysis

The aim of this thesis is the study of the Z boson differential cross section in transverse momentum, pseudorapidity and rapidity using the data set and the Monte Carlo simulation described in Chapter 3. In this chapter I will describe the procedure of data selection, reporting the various cuts imposed to the data and the simulation. In the second part the background sources and their estimate will be described. Moreover the distributions of the kinematic observables used in this analysis will be presented with the description of the systematic uncertainties. Later, the procedure of unfolding as described in Chapter 4 will be applied to the distributions in order to take into account the distortions introduced by the detector. Finally I will present the Z boson differential cross sections.

All the analysis has been done using the software ROOT, version 6.08/02 [52].

5.1 Data selection

The data selected with the HLT from the `Double Electron` data set, as indicated previously in Section 3.1, need to be further discriminated in order to reduce the background contribution and improve the selection. Events are required to have a Z boson reconstructed through its decay products, in this case an electron pair. Physics objects are reconstructed as described in Chapter 3.

Dielectrons coming from a Z boson are selected if they are correctly identified as a electron-positron pair and have opposite charges. Asymmetric minimum thresholds on their transverse momenta are imposed: the most energetic lepton must have $p_T^\ell > 35$ GeV/c, while the second one is accepted if $p_T^\ell > 25$ GeV/c. The values of these cuts are well above the HLT ones (see Chapter 3.1) in order to guarantee 100% of efficiency of the trigger selection, with respect to the offline selection. Both the leptons have to be produced inside the geometrical region of $|\eta_\ell| < 2.5$.

Dielectrons have also to satisfy electron isolation criteria which use particles within a cone around the lepton direction with radius

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.3 \tag{5.1}$$

where ϕ and η are defined as in Section 2.1.2, in order to reduce the possible contamination from hadron jets in the energy reconstruction. The relative isolation is defined by

$$I_{rel} = (I_{charge} + I_{photon} + I_{neutral})/p_T^\ell \tag{5.2}$$

Variable	First lepton cut	Second lepton cut
N. of leptons	≥ 2	
pdgId	± 11	∓ 11
MVAPreselId	1	1
MVAId	≥ 2	≥ 2
Missing hits	1	1
I_{rel}	< 0.15	< 0.15
$\lvert\eta_l\rvert$	< 2.5	< 2.5
p_T^l	> 35	> 25

Table 5.1: Summary of the cuts applied to the electron pairs. Refer to the text for further details.

where I_{charge} is the scalar transverse momentum sum of all the charged hadrons, I_{photon} is the scalar transverse momentum sum of all the photons, $I_{neutral}$ is the scalar transverse momentum sum of all the neutral hadrons in the cone of interest and p_T^ℓ refers to the transverse momentum of the selected electron/positron. For both the leptons a relative isolation lower than 0.15 is asked. Other cuts are imposed to the multivariate analysis selection (MVA), which combines observables sensitive to the bremsstrahlung with the geometrical and momentum-energy matching between the electron trajectory and the associated supercluster, as well as ECAL shower-shape observables, to achieve a better discrimination. Finally is requested that no more than one missing hit is present in the track of the electron candidate in order to discriminate the electron pairs coming from the photon conversion. In addition, the preliminary cut on the number of leptons (≥ 2) is implemented in order to select lepton pairs from the `Double Electron` data set. The summary of all the cuts applied is presented in Table 5.1. The distributions of the main observables used in these selection cuts will be shown in Section 5.3.

In the following analysis the electron pairs are selected only if the dielectrons invariant mass is reconstructed in the range $[71; 111]$ GeV/c^2: a neighbourhood of the Z boson mass.

5.2 Background estimation

Several processes inside the Standard Model framework can emulate the Drell-Yan process, the subject of this thesis, and so these processes have to be considered as background sources. The production of vector bosons associated with jets and the top production can mimic the Drell-Yan final state if some particles are misidentified or non-reconstructed. Moreover, as already mentioned, the Drell-Yan process includes the background electrons coming from $Z \to \tau^+\tau^- \to e^+e^- + 4\nu$.

The different background processes have been already listed in Table 3.2 and they are reported also in Table 5.2, where the uncertainties associated to the cross sections for the simulated samples used are also presented. All these errors come from theoretical uncertainties and correspond to the 15% of the cross section, except for the $t\bar{t}$ process, for which the experimental measurement was available and it was associated with an experimental relative error of 7%, more precise than the corresponding theoretical uncertainty.

Process	Cross section [pb]	Relative uncertainty
$t\bar{t}$ + jets	225.197	7%
$\bar{t}$ s-channel	1.76	15%
$\bar{t}$ t-channel	30.7	15%
$\bar{t}$ tW-channel	11.73	15%
t s-channel	3.79	15%
t t-channel	56.4	15%
t tW-channel	11.73	15%
W + jets	37509.0	15%
WW + jets	5.995	15%
WZ + jets	1.057·1.10	15%
ZZ	8.059	15%
$Z \to \tau^+\tau^-$	119.02	15%

Table 5.2: Theoretical predictions for the Standard Model cross sections of the background sources used in this analysis with the relative errors [43]-[44].

The backgrounds can be divided in "resonant" and "non-resonanat". The former ones produce a peak in correspondence of the Z boson mass in the reconstructed invariant mass plot, while the latter ones are distributed more uniformly along the plot range. All these background processes have been evaluated from Monte Carlo simulations using `MadGraph 5` and `PYTHIA 6` generators, and their contributions have been subtracted from the data. This procedure has been validated for the non-resonant backgrounds using a data-driven method employing an electron-muon data control sample.

The following processes can contribute as background sources:

- W^+W^-+ jets $\to 2\ell^\pm + 2\nu$ + jets, in which the neutrinos are not detected;
- $W^\pm$+ jets, where a secondary lepton is wrongly identified;
- $W^\pm Z$ + jets $\to 3\ell^\pm + \nu$ + jets;
- $Z/\gamma^* \to \tau^+\tau^-$+ jets $\to 2\ell^\pm + 4\nu$ + jets has been identified through its $\tau^+\tau^-$ intermediate step, isolated and subtracted from the DY sample;
- the production of $ZZ \to 4\ell$;
- the general QCD production has been shown to be negligible [55];
- the single top quark production is a small background for the Drell-Yan events. A top or an anti-top can be produced via the s-channel, the t-channel or the tW-channel [56] through the weak interactions during the hard scattering. In the following, all these single top (anti-top) production channels are merged as an unique "single top" background;
- the top-anti-top production ($t\bar{t} \to W^+W^-b\bar{b} \to \ell^+\ell^-\nu\bar{\nu}\ b\bar{b}$) is the dominant background in the analysis. Fortunately, as mentioned before, the relative error associated to this background is bounded from experimental measurements to 7% instead of the theoretical 15% of the other backgrounds.

5.3 Dielectrons distributions

In this section the distributions of all the kinematic variables that characterise the dielectrons of the Drell-Yan events are reported. The characterising distributions are the transverse momenta (p_T) of the most energetic (leading) lepton and of the second (subleading) one, their distribution in pseudorapidity (η) and their relative isolation: the event selection is in fact based on these variables. All these distributions are presented together with the Monte Carlo simulation for the signal and for all the background contributions. The unfolding procedure is not yet implemented at this stage, and it will be examined in Section 5.6. The MC distributions are corrected for differences in electron reconstruction efficiency between data and simulation using scale factors obtained with the Tag and Probe method as a function of the electron transverse momentum and pseudorapidity. The simulated events are re-weighted in order to take into account the differences in the pile-up distributions between data and simulation, and they are also weighted in order to take into account the conditions of luminosity of the experiment: $weight = \sigma \cdot L/N_{ev}$, where σ and N_{ev} are reported in Table 3.2 and 5.2, and $L = 19.78$ fb^{-1}. Another important distribution is the one of the R_9 variable. This is defined as the ratio between the energy found in a 3×3 crystal array and the energy found in a ECAL supercluster [57], and it provides measures of the lateral shower shape. Using this variable one can group the electrons and positrons in two categories: the "golden" electrons with $\mathrm{R}_9 > 0.94$ and the "non-golden" electrons with $\mathrm{R}_9 < 0.94$. As mentioned in Section 3.1.2 golden electrons are those with little bremsstrahlung and consequently provide the most accurate estimation of momentum. Also the jet multiplicity distribution is presented in this section.

All the selection cuts are implemented in the distributions shown in this chapter, except for those of the invariant mass, in which the cut $\mathrm{M}_{\ell\ell} \in [71; 111]$ GeV/c^2 is not present in order to show the whole spectrum.

In Figure 5.1 and 5.2 the distributions of the transverse momentum of the two leptons are displayed. The asymmetric cuts are well visible; the shapes present a peak at almost 50 GeV/c, and then a long slope extending to high momenta, in particular for the leading electron. This slope contains almost exclusively events from the Drell-Yan process, while the backgrounds contribute mainly in the peak region, although, as in the other distributions, the backgrounds are almost 3 orders of magnitude smaller than the DY signal. As one can see from the ratio plots, the data are well described by the simulation, in particular for the leading electron, while for the subleading one the ratio is anyway compatible with unity within the errors.

In Figure 5.3 and 5.4 the distributions of the pseudorapidity of the two electrons are displayed. The cuts at $|\eta| < 2.5$ are well visible, as the transition from the barrel to the endcaps (the incision at $|\eta| = 1.479$). The majority of the leptons are emitted in the barrel region, in particular for the leading particle, where the difference between the two detector regions is particularly clear. On the opposite, for the subleading particle the transition is nearly smooth. For both the distributions the ratio plot is flat to the unity in the barrel region, while it decreases in the endcap regions, especially in the first case. Although, at first sight, the distributions seem symmetric respect to $\eta = 0$, the ratio plots show that there is an asymmetry between the two endcap regions. This observation points to a possible need of a more refined

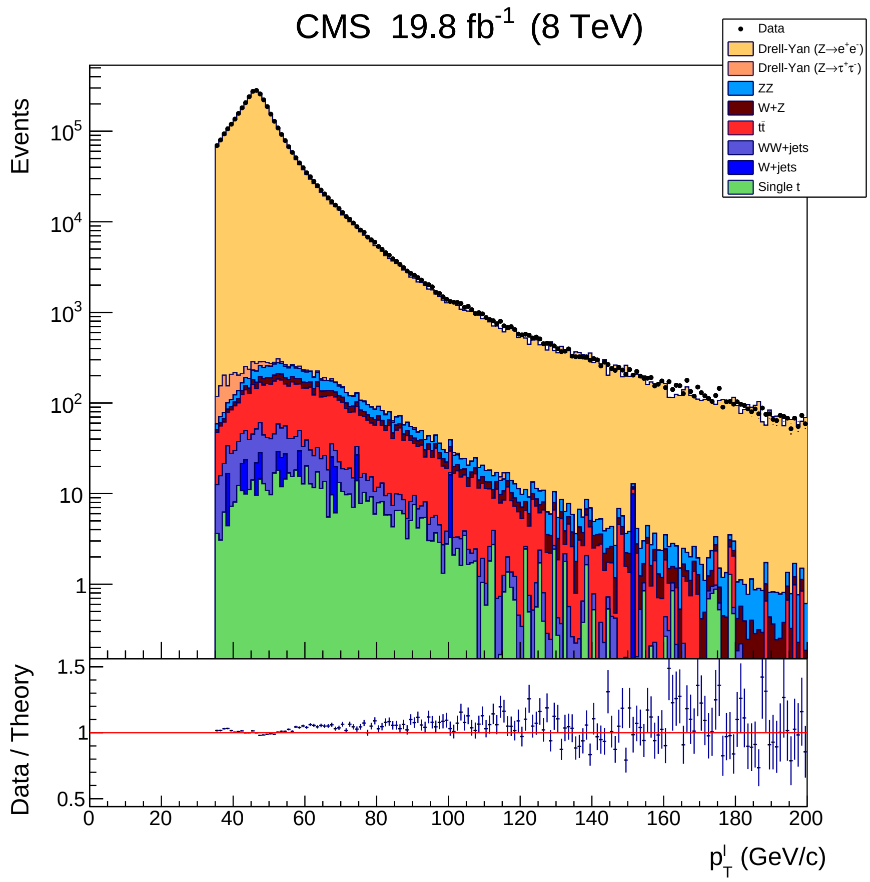

Figure 5.1: Distribution of the transverse momentum (p_T) of the leading lepton.

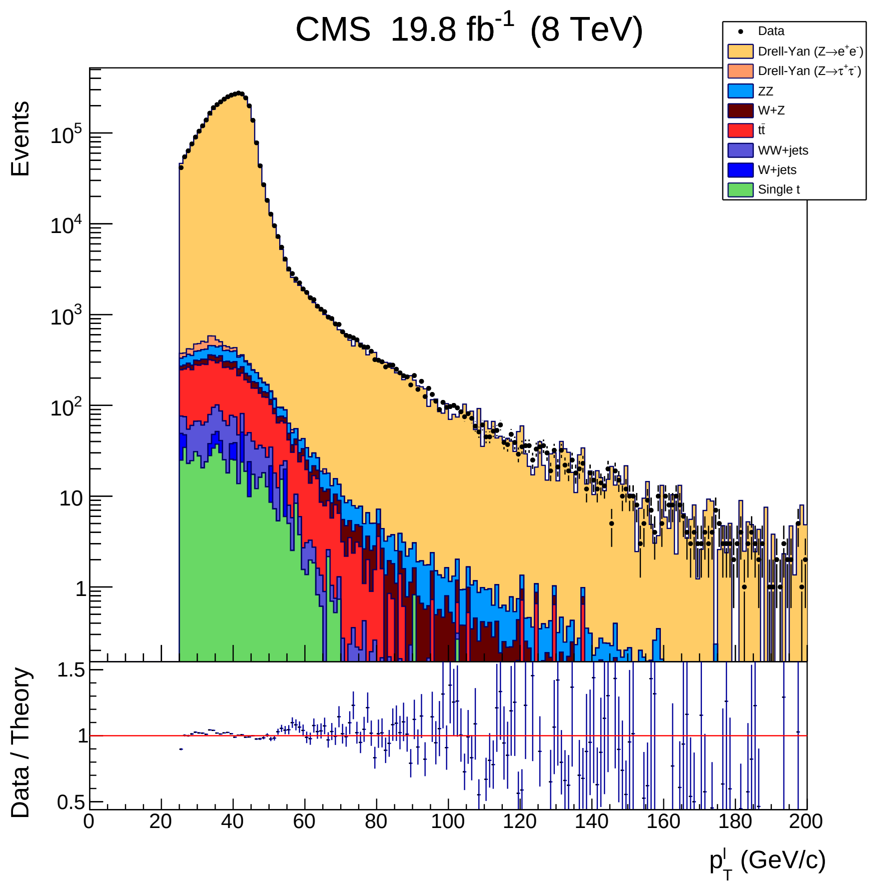

Figure 5.2: Distribution of the transverse momentum (p_T) of the subleading lepton.

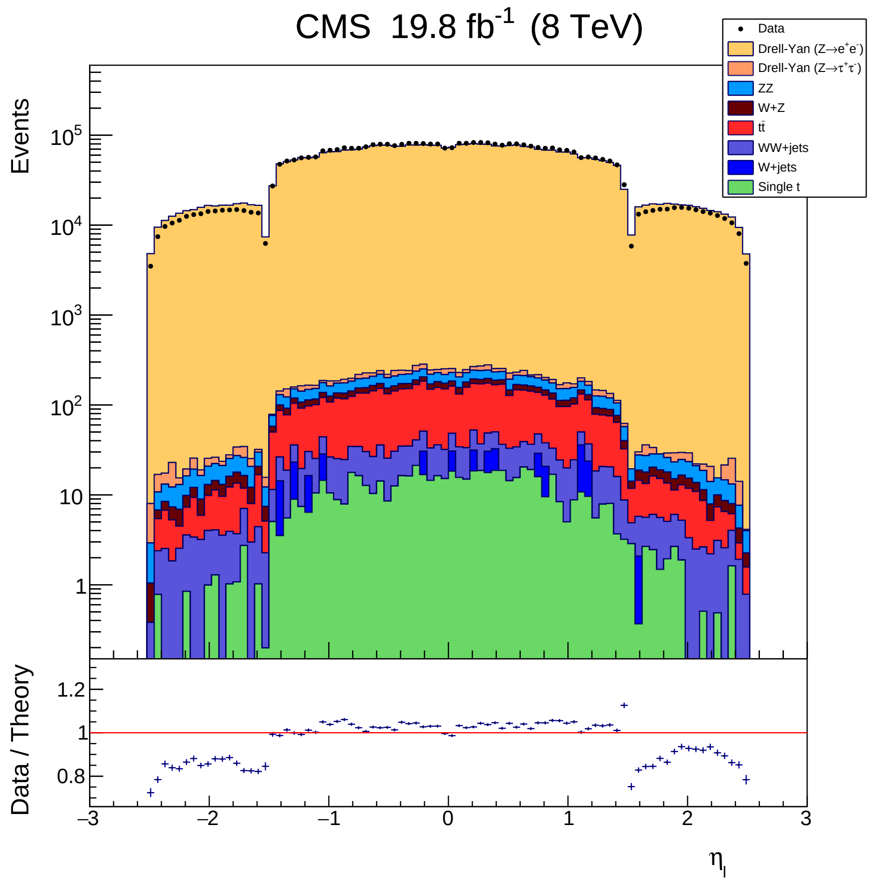

Figure 5.3: Distribution of the pseudorapidity (η) of the leading lepton.

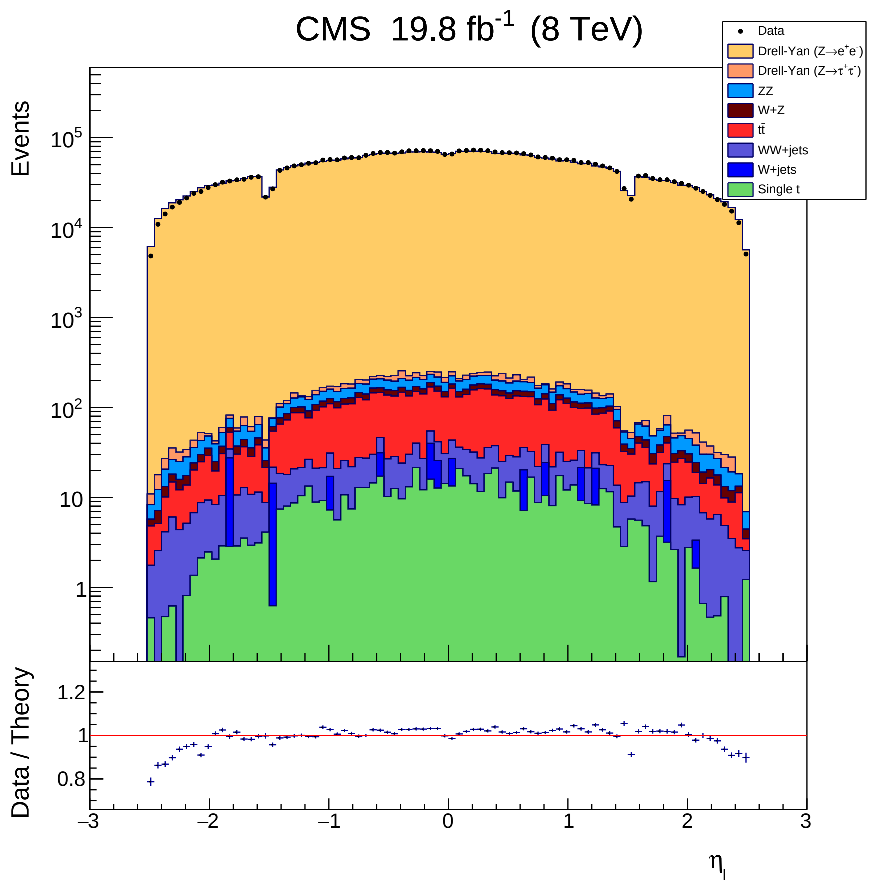

Figure 5.4: Distribution of the pseudorapidity (η) of the subleading lepton.

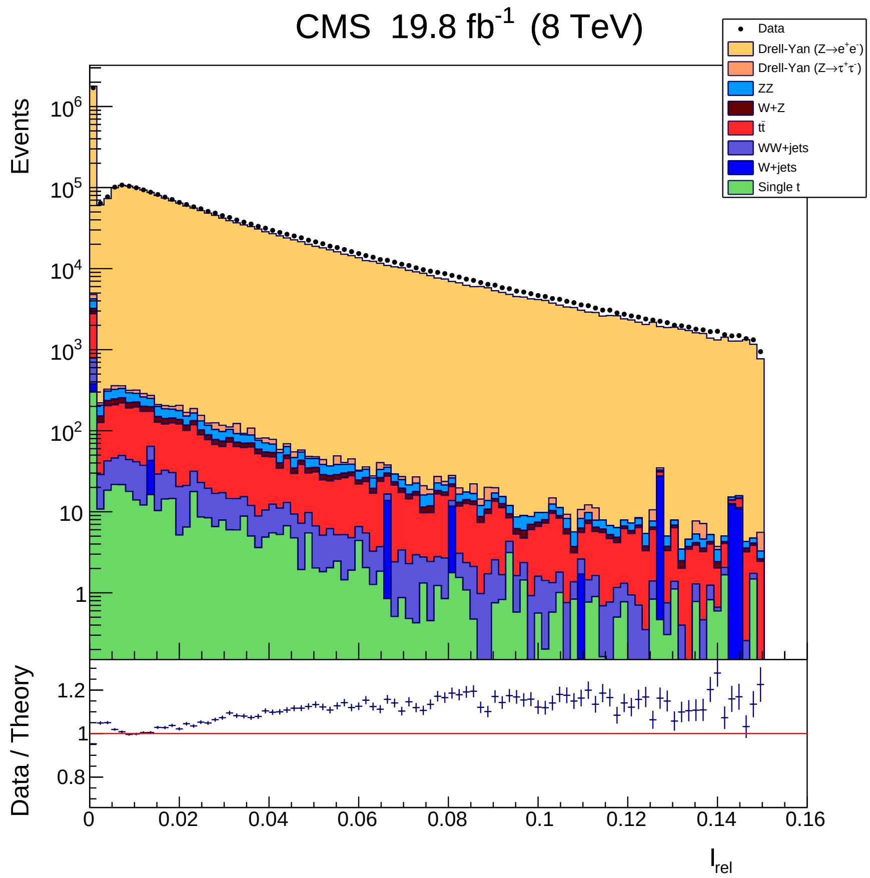

Figure 5.5: Distribution of the relative isolation (I_{rel}) of the leading lepton.

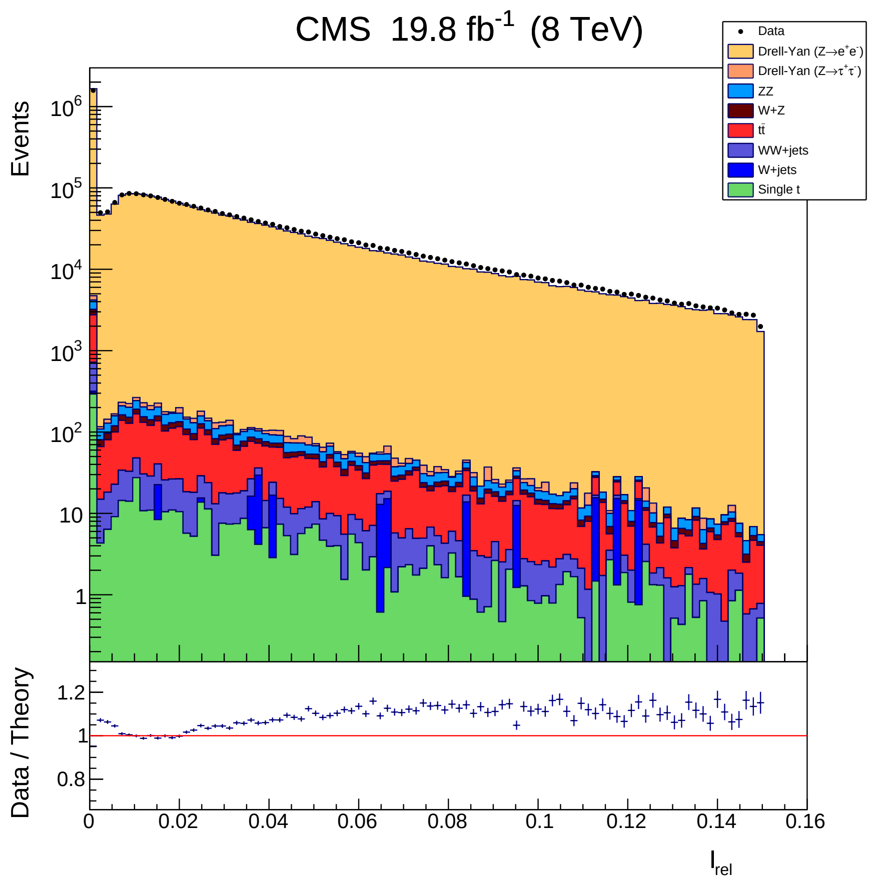

Figure 5.6: Distribution of the relative isolation (I_{rel}) of the subleading lepton.

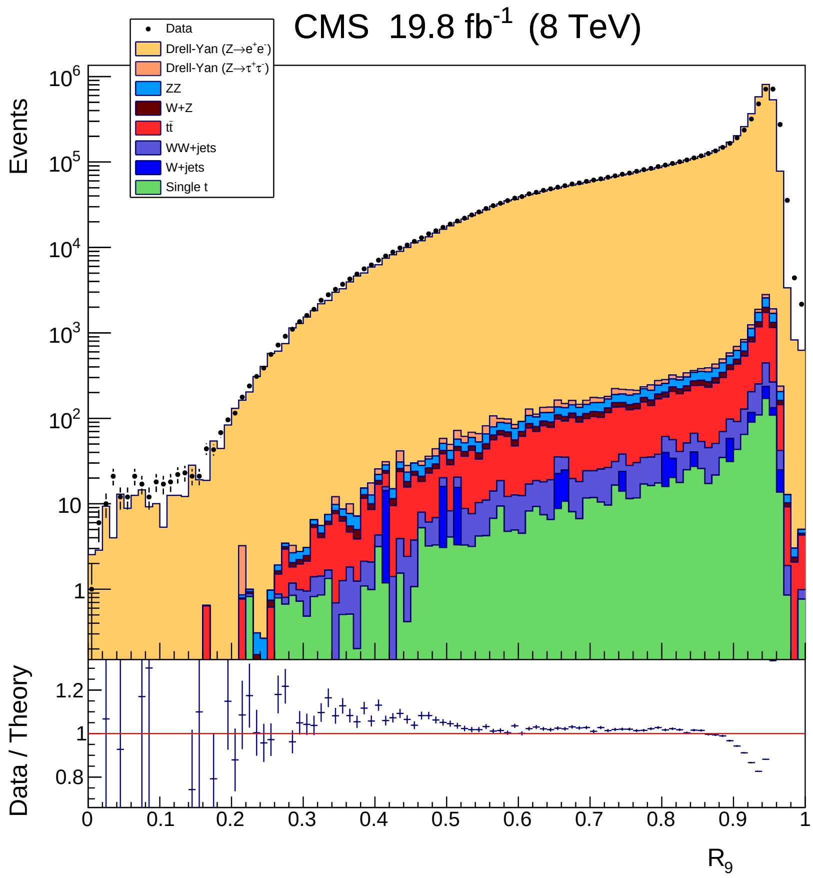

Figure 5.7: R_9 variable distribution for electrons/positrons in the barrel.

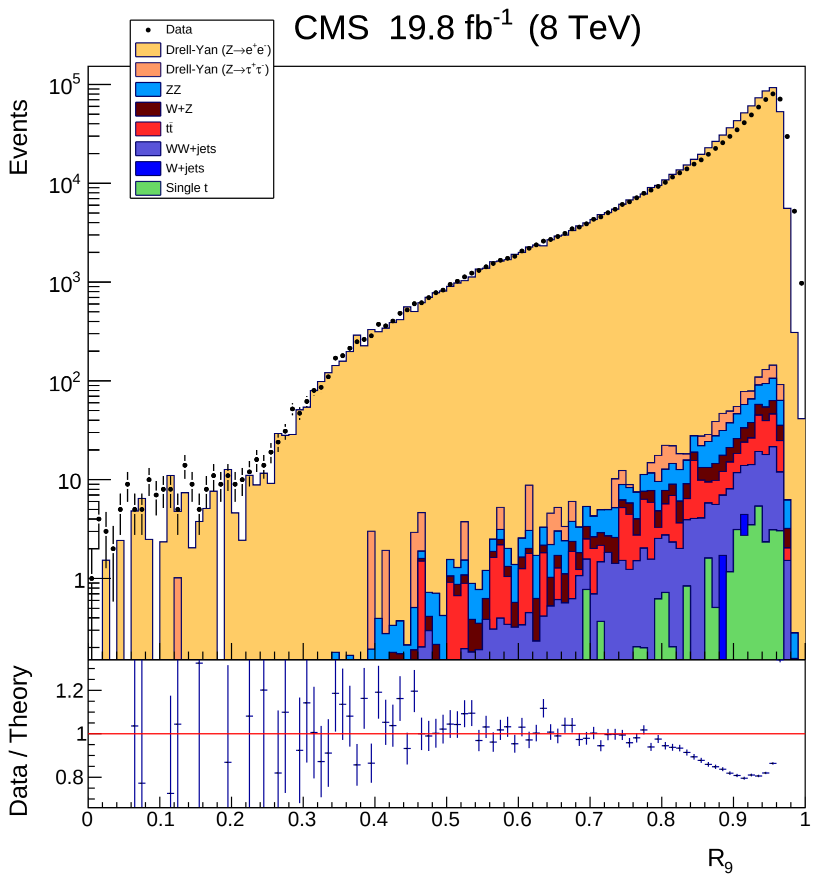

Figure 5.8: R_9 variable distribution for electrons/positrons in the endcaps.

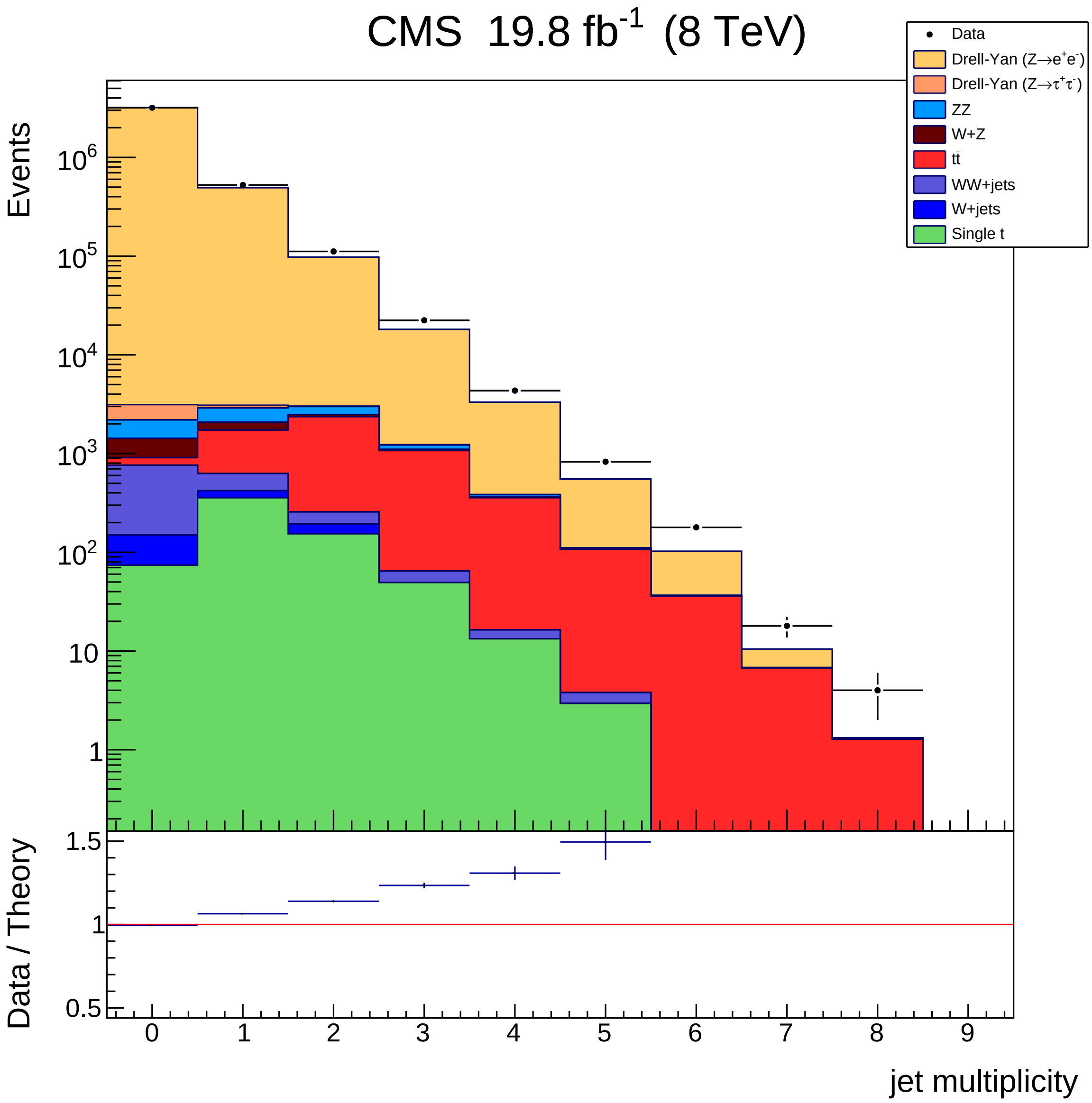

Figure 5.9: Jet multiplicity distribution in Drell-Yan events.

study of the differences between data and MC in these regions.

In Figure 5.5 and 5.6 the distributions of the relative isolation of the two electrons are displayed. Both the plots show an agreement within 20%, with a disagreement at high isolation likely coming from the neutral component in the pile-up events.

In Figure 5.7 and 5.8 the distributions of the R_9 variable are displayed, distinguishing the case in which both the particles were identified in the barrel from that where the two leptons were identified in the endcaps. As one can see there is a peak at $R_9 = 0.94$. The ratio plots show as the simulation overestimates the data before such peak and underestimates them beyond it, possibly due to a not perfect simulation of the detector material. This discrepancy may induce a systematic effect on the energy scale of the transverse momentum, and it has to be taken into account.

In Figure 5.9 the distribution of the jet multiplicity of the events is displayed. For this plot two additional cuts have been applied: $p_T^{jet} > 30$ GeV/c and $|\eta_{jet}| < 2.4$. In most cases the event is associated with no jets, both in the data and in the simulation. However it is clear that the prediction underestimates the jet multiplicity when two or more jets are involved. At high multiplicity the main process that contributes to the jet production is the $t\bar{t}$ process, while the other backgrounds and the Drell-Yan simulation gradually vanish. The discrepancy can be attributed to a not well simulated jet production in the DY Monte Carlo. However this analysis is inclusive in the jet multiplicity and so this discrepancy can be not considered.

5.4 Z boson distributions

In Figure 5.10 the whole spectrum of the distribution of the reconstructed invariant mass of the dielectrons is presented. There is a big discrepancy between data and simulation in the region below 50 GeV/c^2, but this has a simple explanation: while the data are selected by the HLT, the Drell-Yan events simulation has been generated with an additional cut on the invariant mass that drops rapidly the contribution of the distribution below 50 GeV/c^2 to zero. The shape shows a high peak at the Z boson mass confirming that the electron pairs are produced truly in the Z boson decay, except for the ones coming from the $Z \to \tau^+\tau^-$ process, where the electrons come from the leptonic decays of τ, so there is a part of the energy not reconstructed due to the neutrinos, or from the hadronic decays of τ, where a portion of the jet is misidentified as an electron. The background contributions are very small and their principal components are the $t\bar{t}$ process all over the spectrum, the WZ and ZZ productions which instead have a peak corresponding to the Z boson. The ratio plot shows that there is a quite good agreement between the data and the Monte Carlo simulation in region above 50 GeV/c^2.

In Figure 5.11 the distribution of the reconstructed Z boson transverse momentum is displayed. A variable binning has been employed: the bin width enlarges at increasing transverse momenta. This choice, useful to show the distribution, creates the fake peaks visible along the spectrum. The ratio plot shows a general reasonable agreement: the difference is not above the 10%.

In Figure 5.12 the distribution of the reconstructed Z boson pseudorapidy is displayed. The background contribution is concentrated in the $-6 < \eta < 6$ region and it does not show the same shape of the DY process. The ratio plot shows that the agreement is good in this region, while it is less good, although compatible, in the external regions, possibly due to a not perfect simulation of the detector.

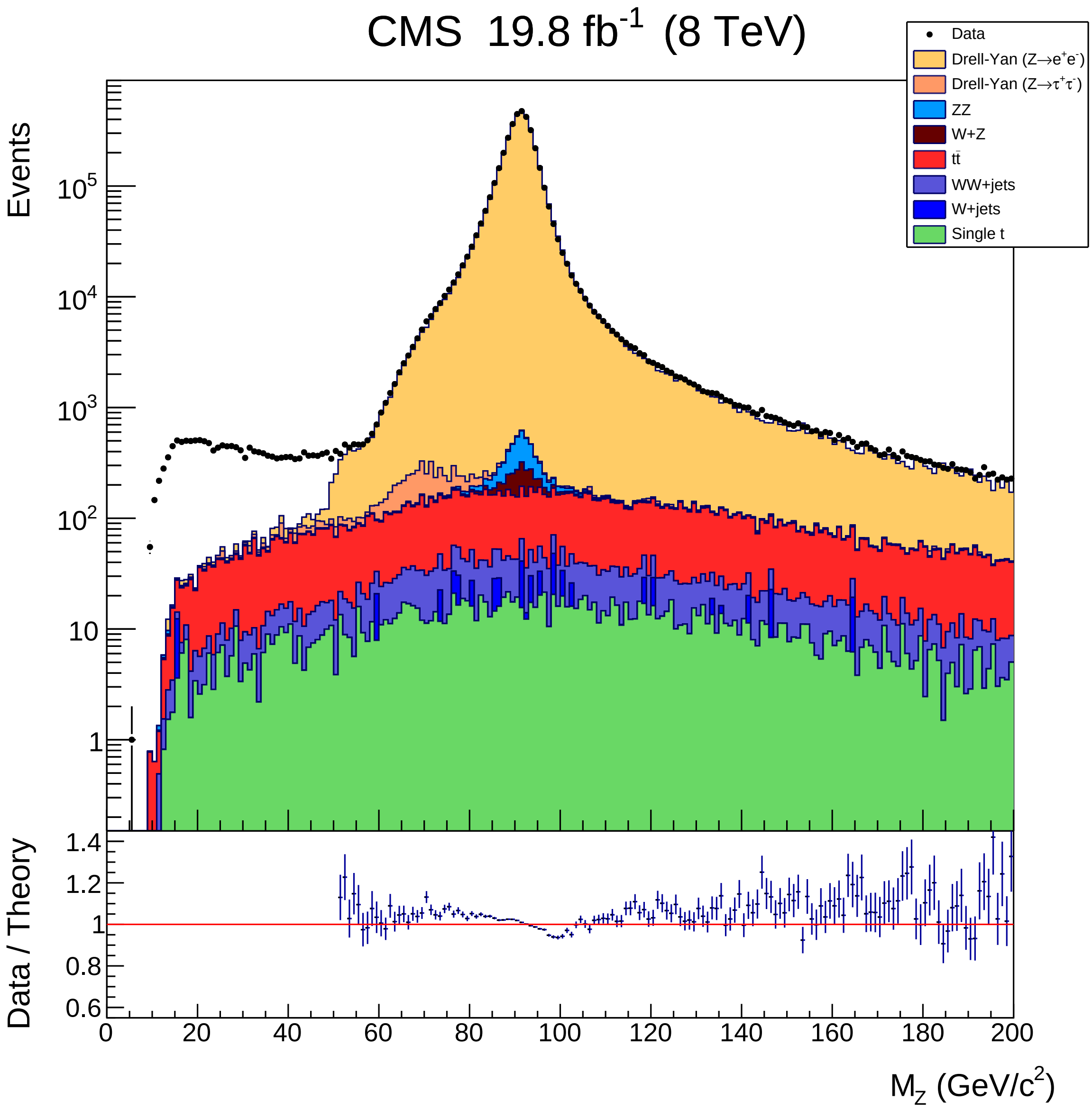

Figure 5.10: Invariant mass distribution of the reconstructed Z boson.

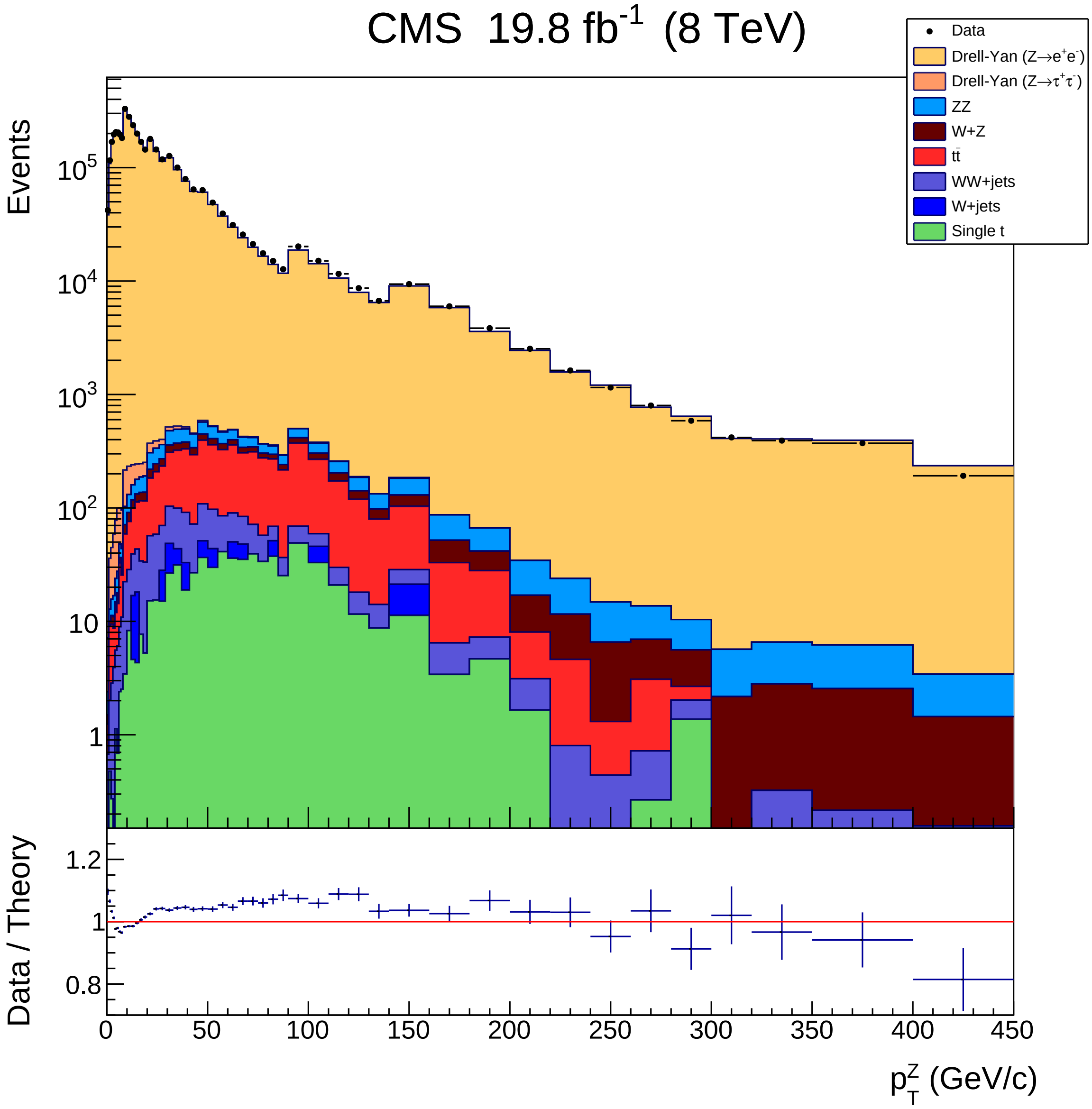

Figure 5.11: Distribution of the transverse momentum (p_T) of the reconstructed Z boson.

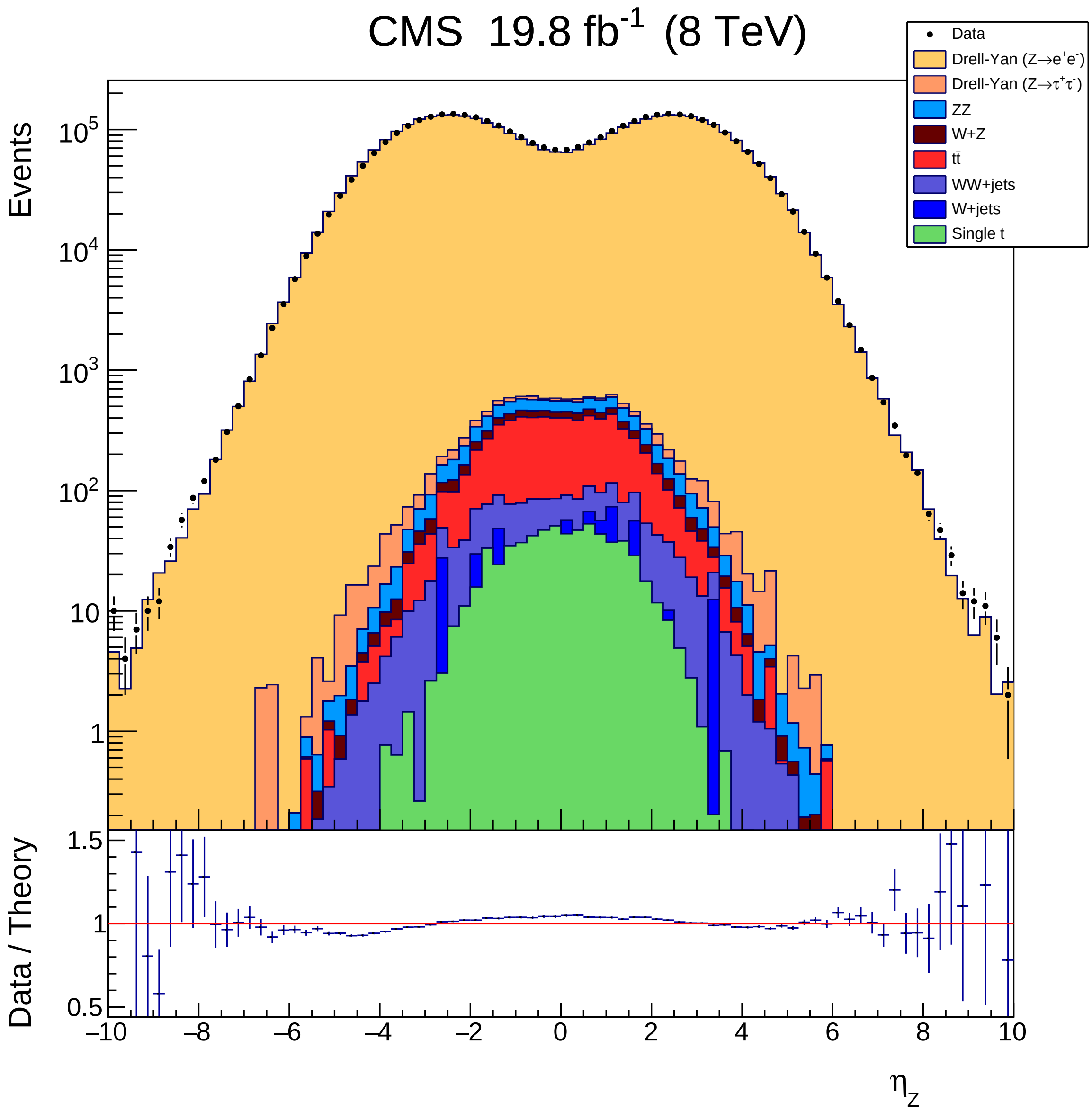

Figure 5.12: Distribution of the pseudorapidity (η) of the reconstructed Z boson.

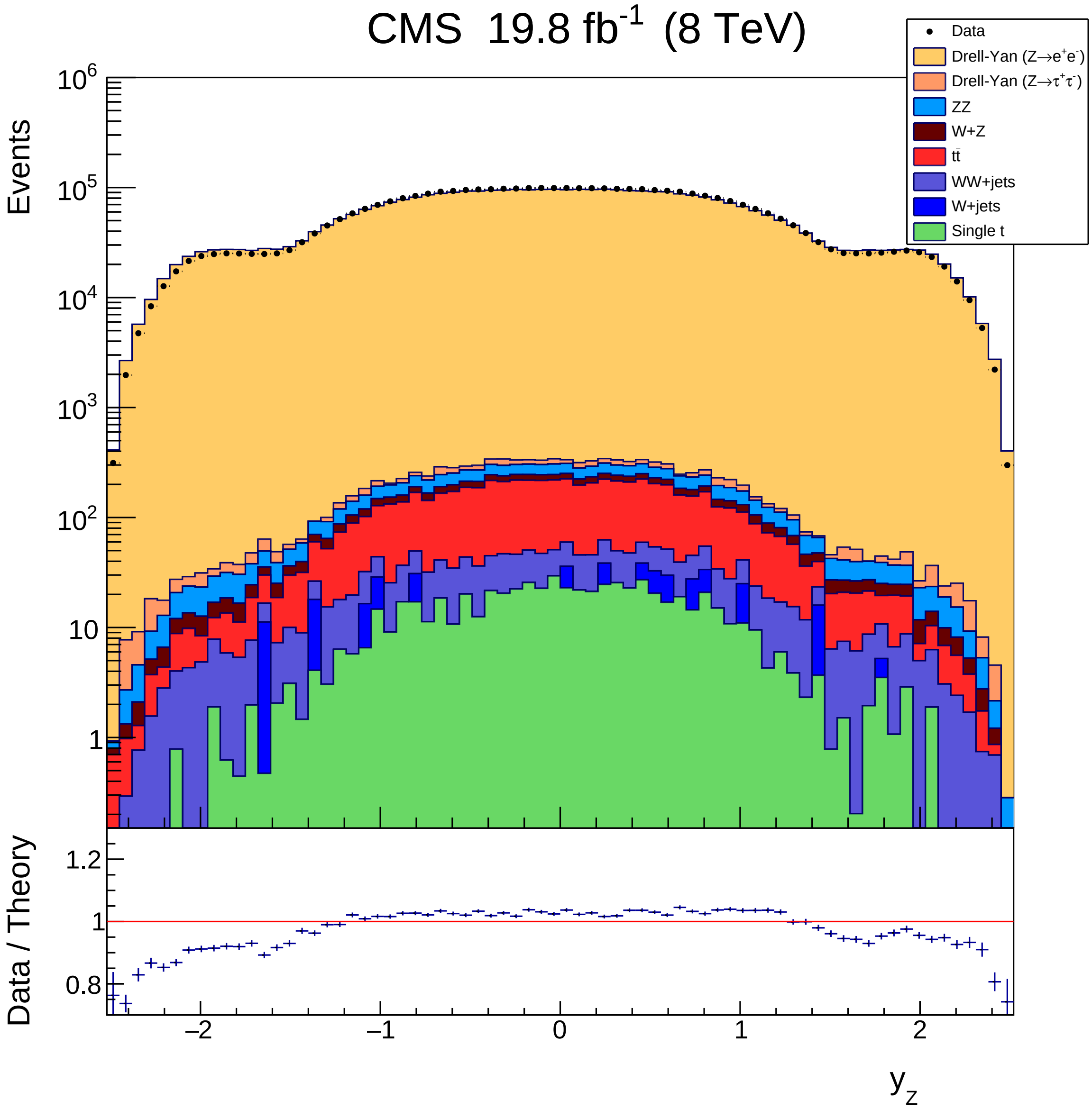

Figure 5.13: Distribution of the rapidity (y) of the reconstructed Z boson.

The large range was needed for the unfolding procedure and it was reduced to a smaller range for the final distribution of the differential cross section respect to η. The ratio plot displays also a faint asymmetry of unknown origin between $-6 < \eta < -2$ and $2 < \eta < 6$.

In Figure 5.13 the distribution of the reconstructed Z boson rapidity is displayed. The background contribution is spread along all the range. The ratio plot shows that the agreement is good in the central region (below 5% of discrepancy), while it gets worse in the regions at $|y| > 1.4$, in which there is also an asymmetry at around $|y| = 2$.

5.5 Systematic uncertainties

Different sources of systematic uncertainties have been considered [33]:

- the uncertainty in the measurement of the integrated luminosity is 2.6% [58];
- the reconstruction inefficiency uncertainty for electrons, as mentioned previously, is in general about 1.5%;
- the uncertainty on the momentum resolution due to the R_9 variable distribution will be analysed in the following, and it is about 0.5%;
- the uncertainties on the background cross sections are evaluated as 7% for the $t\bar{t}$ and 15% for the other sources, as mentioned before;
- the pile-up uncertainty is smaller than 0.5% and can be considered as negligible with respect to the other uncertainties.

As seen in Figure 5.7 and 5.8 the distributions of the R_9 variable present a discrepancy between data and simulation. In order to investigate if that situation produces a systematic uncertainty on the transverse momentum, the distribution of the reconstructed Z boson transverse momentum has been divided in four different cases, visible in Figure 5.14: electrons that were both detected in the barrel and were both golden electrons, electrons that were both detected in the barrel but, both, non-golden, and the same for electrons that were both detected in the endcaps. As one can see, the difference between data and simulation is due to a scale factor, in particular for the first histogram (golden electrons in the barrel). This scale factor is introduced exactly by the R_9 discrepancy. The plots show that the ratio is quite flat, apart from the scale factor

In order to estimate the systematic error in the energy scale the same sub-division has been applied to the invariant mass (see Figure 5.15 and 5.16) and the position of the peak has been fitted. The fit was done with a Cruijff function, i.e. an asymmetric Gaussian function, summed with a polynomial for the background:

$$f(x, \alpha, \mu, \sigma_l, \sigma_r, \alpha_l, \alpha_r, a, b, c) = \alpha \exp\left(-\frac{(x-\mu)^2}{2\sigma_i^2 + \alpha_i (x-\mu)^2}\right) + a + bx + cx^2 \quad (5.3)$$

$$\begin{cases} i = l \ (x-\mu) < 0 \\ i = r \ (x-\mu) > 0 \end{cases} \quad (5.4)$$

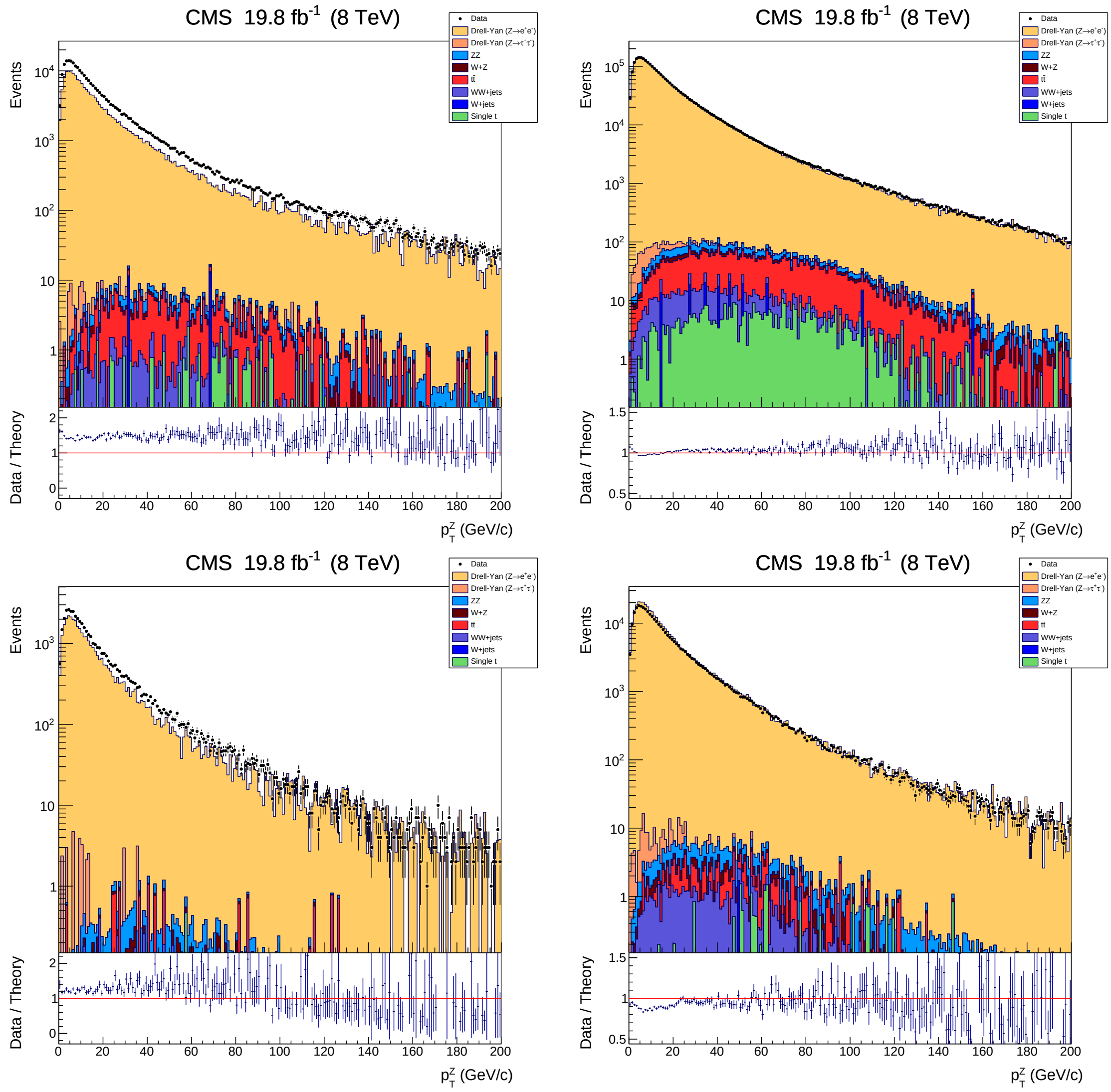

Figure 5.14: Distribution of the Z boson transverse momentum in four different cases: at the top left when both the leptons were golden electrons and they were reconstructed both in the barrel, at the top right when both the leptons were non-golden electrons and they were reconstructed both in the barrel, at the bottom left when both the leptons were golden electrons and they were reconstructed both in the endcaps, at the bottom right when both the leptons were non-golden electrons and they were reconstructed both in the endcaps. The normalisation discrepancy is introduced by R_9.

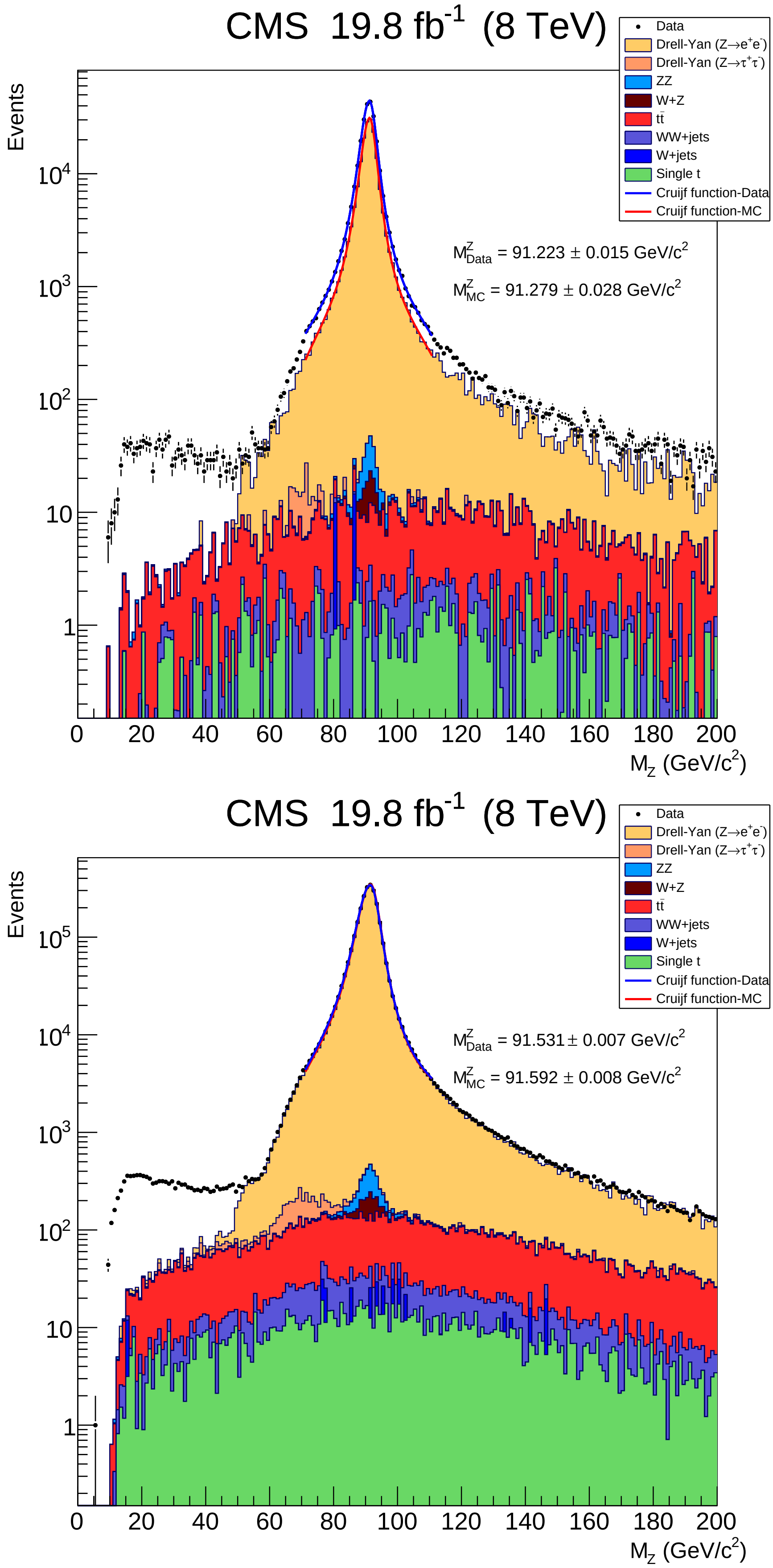

Figure 5.15: Distribution of the invariant mass for golden (top) and non golden (bottom) dileptons detected in the barrel. The fitted position of the peak is indicated. The normalisation discrepancy is introduced by R_9.

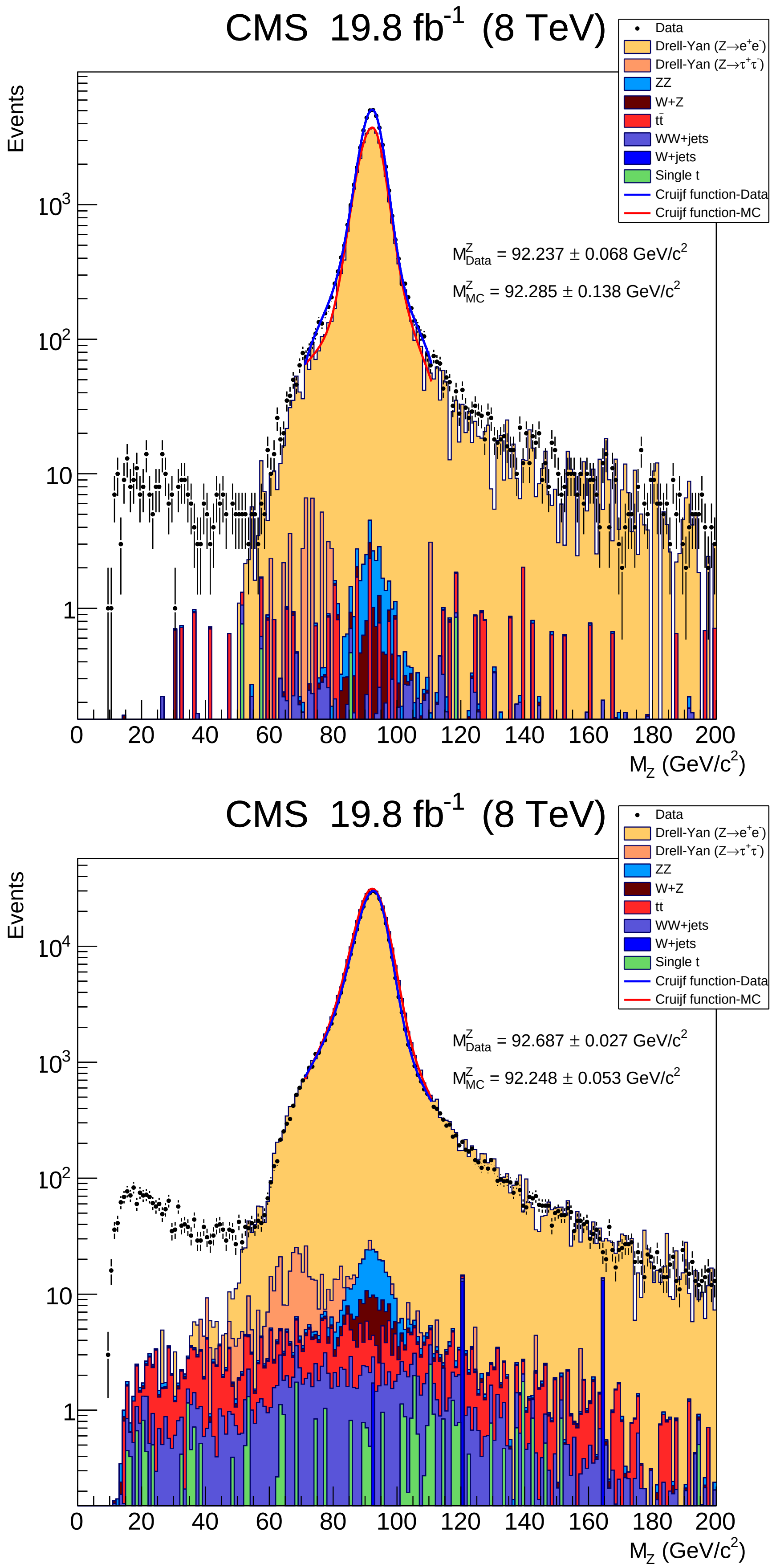

Figure 5.16: Distribution of the invariant mass for golden (top) and non golden (bottom) dileptons detected in the endcaps. The fitted position of the peak is indicated. The normalisation discrepancy is introduced by R_9.

	Data		Monte Carlo	
Parameter	*Value*	*Error*	*Value*	*Error*
α	$4.566{\cdot}10^4$	$0.029{\cdot}10^4$	$3.230{\cdot}10^4$	$0.038{\cdot}10^4$
α_l	0.2905	0.0078	0.296	0.015
$2\sigma_l^2$	7.23	0.11	6.82	0.20
α_r	0.302	0.008	0.304	0.016
$2\sigma_r^2$	4.87	0.08	4.62	0.15
μ	91.223	0.015	91.279	0.028
a	$-3.36{\cdot}10^3$	$0.76{\cdot}10^3$	$-3.43{\cdot}10^3$	$0.98{\cdot}10^3$
b	47	19	54	25
c	-0.28	0.11	-0.30	0.14

Table 5.3: Fit parameters for the Figure 5.15-top: two golden leptons in the barrel.

	Data		Monte Carlo	
Parameter	*Value*	*Error*	*Value*	*Error*
α	$3.499{\cdot}10^5$	$0.008{\cdot}10^5$	$3.545\ {\cdot}10^5$	$0.063{\cdot}10^5$
α_l	0.270	0.003	0.268	0.001
$2\sigma_l^2$	15.11	0.07	14.54	0.08
α_r	0.2616	0.0033	0.2584	0.0018
$2\sigma_r^2$	7.735	0.049	7.630	0.055
μ	91.531	0.007	91.592	0.008
a	$-2.46{\cdot}10^4$	$0.34{\cdot}10^3$	$-1.94{\cdot}10^4$	$0.30{\cdot}10^4$
b	$3.18\ {\cdot}10^2$	$0.85\ {\cdot}10^2$	$2.00{\cdot}10^2$	$0.03\ {\cdot}10^2$
c	-1.40	0.46	-0.75	0.03

Table 5.4: Fit parameters for the Figure 5.15-bottom: two non-golden leptons in the barrel.

	Data		Monte Carlo	
Parameter	*Value*	*Error*	*Value*	*Error*
α	$4.937{\cdot}10^3$	$0.055{\cdot}10^3$	$3.665{\cdot}10^3$	$0.052{\cdot}10^3$
α_l	0.135	0.016	0.081	0.019
$2\sigma_l^2$	19.14	0.85	27.0	2.1
α_r	0.113	0.017	0.143	0.021
$2\sigma_r^2$	15.43	0.74	17.3	1.5
μ	92.237	0.068	92.28	0.14
a	$-2.14{\cdot}10^3$	$0.50{\cdot}10^3$	$-6.97{\cdot}10^2$	$0.21{\cdot}10^2$
b	50	12	18.05	0.04
c	-0.275	0.064	-0.104	0.002

Table 5.5: Fit parameters for the Figure 5.16-top: two golden leptons in the endcaps.

	Data		Monte Carlo	
Parameter	*Value*	*Error*	*Value*	*Error*
α	$2.938{\cdot}10^4$	$0.009{\cdot}10^4$	$3.060{\cdot}10^4$	$0.033{\cdot}10^4$
α_l	0.1825	0.0038	0.171	0.016
$2\sigma_l^2$	29.66	0.41	30.10	0.84
α_r	0.1154	0.0045	0.088	0.014
$2\sigma_r^2$	20.33	0.34	28.56	0.82
μ	92.687	0.027	92.248	0.053
a	$-6.903{\cdot}10^3$	$0.069{\cdot}10^3$	$-8.8{\cdot}10^3$	$3.7{\cdot}10^3$
b	$1.626{\cdot}10^2$	$0.009{\cdot}10^2$	$2.05{\cdot}10^2$	$0.88{\cdot}10^2$
c	-0.877	0.007	-1.11	0.49

Table 5.6: Fit parameters for the Figure 5.16-bottom: two non-golden leptons in the endcaps.

where α is a normalisation parameter, μ is the position of the peak, σ_i are the asymmetric standard deviations, α_i are offset parameters, a, b and c are the parameters of the second degree polynomial. All these parameters have been left floating in the fitting procedure.

The results of the fitting procedure are reported divided in the four cases in Tables 5.3 $\div$ 5.6. As one can see from the fitting procedure, there is a small displacement between the peak fitted in the data distribution and the one in the Monte Carlo distribution. The maximum discrepancy can be located in the case of the non-golden electrons in the endcaps: 0.439 GeV/c^2. Dividing for the fitted invariant mass one obtains a systematic uncertainty on the energy scale of the transverse momentum of about 0.5%. The way to take into account this uncertainty involves the unfolding procedure, and so it will be explained in the next section.

5.6 Data unfolding

The unfolding procedure has been implemented using the TUnfold package inside the ROOT framework, as considered in Chapter 4. In order to create the matrix that describes the bins migration of the distribution of a chosen observable it is necessary to know the "truth" distribution. In fact `MadGraph` and `PYTHIA` generate events with some distributions of observable ("truth" or "generated" distributions) that are further simulated with GEANT4 to reproduce the processes that happen in the detector, its geometry and materials and to match with the detected data ("reconstructed" distributions, those shown in the previous sections). The response matrix links the two distributions, and is used with the unfolding procedure to go back to the "truth" data distributions.

The response matrices for the transverse momentum, pseudorapidity and rapidity are produced from the Drell-Yan simulation. The generated distributions are filled if the cuts reported in Table 5.7 are satisfied: they are equivalent to the cuts on the reconstructed level described in Section 5.1. The reconstructed distributions, instead, are filled if, in presence of a generated event, the reconstructed one satisfies also the cuts reported in Table 5.1. In this way the response matrices are created with the truth distribution along the x axis and the reconstructed distribution along

Variable	First lepton cut	Second lepton cut
N. of leptons	≥ 2	
$M_{\ell\ell}^{gen}$	$\in [71; 111]$	
pdgId	± 11	∓ 11
$\lvert\eta_l\rvert$	< 2.5	< 2.5
p_T^l	> 35	> 25

Table 5.7: Summary of the cuts applied to the generated electron pairs.

the y axis. If there is a generated (reconstructed) event, but the corresponding reconstructed (generated) event is not present, a misidentification occurs and these cases are registered in the underflow and overflow bins of the matrix (not visible). The response matrices for the transverse momentum, pseudorapidity and rapidity are displayed in Figures 5.17 $\div$ 5.19. In the case of p_T the matrix is also shown using a logarithmic scale both in the x axis and in the y axis: since a variable binning has been employed in this distribution, the first part, below 50 GeV/c, is not well clear; the logarithmic scales allow a better view in this region. The matrices result almost diagonal, in particular for the rapidity, but also for the transverse momentum. For the pseudorapidy, instead, the distribution is more spread with a double drop shape.

In order to use the TUnfold algorithm (see Chapter 4) the truth distributions are rebinned to answer the request that the reconstructed distributions have to be measured in finer bins. Also in the matrix this procedure is needed. A factor two has been chosen for this rebinning procedure. For the reconstructed distributions, the background components have been subtracted before the unfolding. For the regularisation condition, the one of the first derivative of $\mathbf{x}$ has been chosen (see Chapter 4.2.3). The TUnfold algorithm requires the generated distribution, the reconstructed distribution and the response matrix, built as described. It gives back the unfolded distribution with a statistical and a total (statistical+systematic) error. The first one has been calculated also using the response matrices produced in simulations with other PDF sets: they allow to estimate the effect of using a different simulation for the response matrix.

Figures 5.20 $\div$ 5.22 display the result of the unfolding procedure. In the case of the transverse momentum, a logarithmic scale is used in both the axes in order to better show the distribution. The truth and reconstructed Monte Carlo distributions are also reported with the measured background-subtracted data. The discrepancies between the unfolded data and the generated simulations affect the differential cross section distributions and they will be examined in the next section.

As stated in the previous section there is a systematic uncertainty on the energy scale of the transverse momentum. In order to take into account this systematic effect, the content of the bins of the reconstructed transverse momentum distribution has been multiplied by 1.0047 (1 plus the uncertainty on the energy scale, 0.0047), and the same in the response matrix (only the reconstructed distribution has been multiplied). These new distributions are then used to unfold the data, obtaining "modified" unfolded distributions. One can then take bin by bin the absolute value of the difference between the unity and the ratio (r_i) between the two different unfolded distributions (the normal and the modified ones), obtaining the systematic uncertainty on the energy scale of the transverse momentum, $\delta p_{T,i}^{syst} = |1 - r_i|$, that can be summed in quadrature with the other errors.

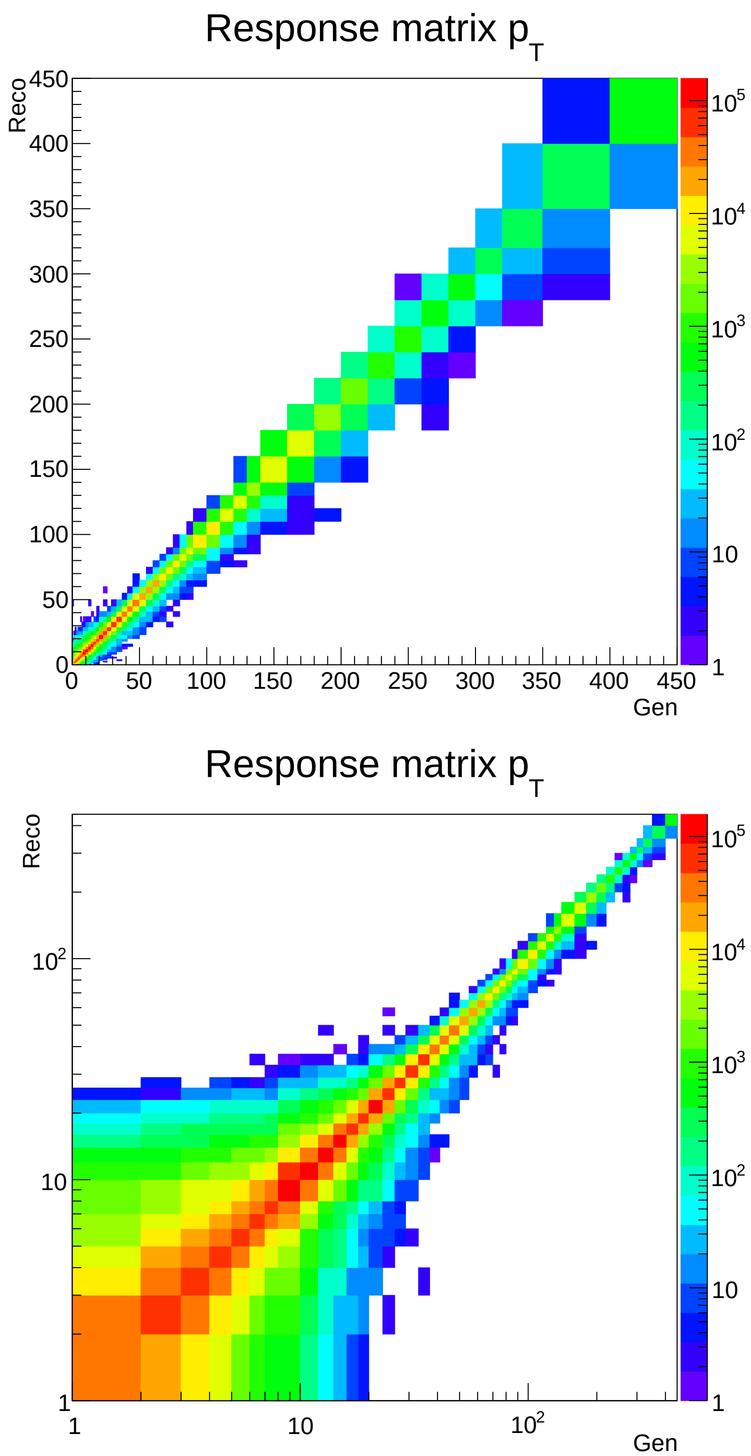

Figure 5.17: Response matrix for the transverse momentum (p_T) of the Z boson. Having used a variable binning, the matrix with logarithmic scales is reported to better show the region at low transverse momenta.

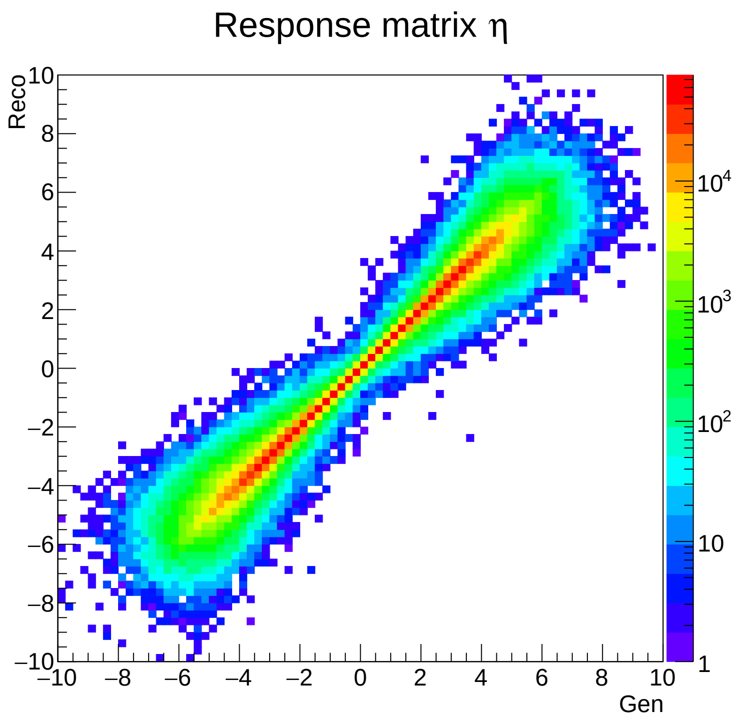

Figure 5.18: Response matrix for the pseudorapidity (η) of the Z boson.

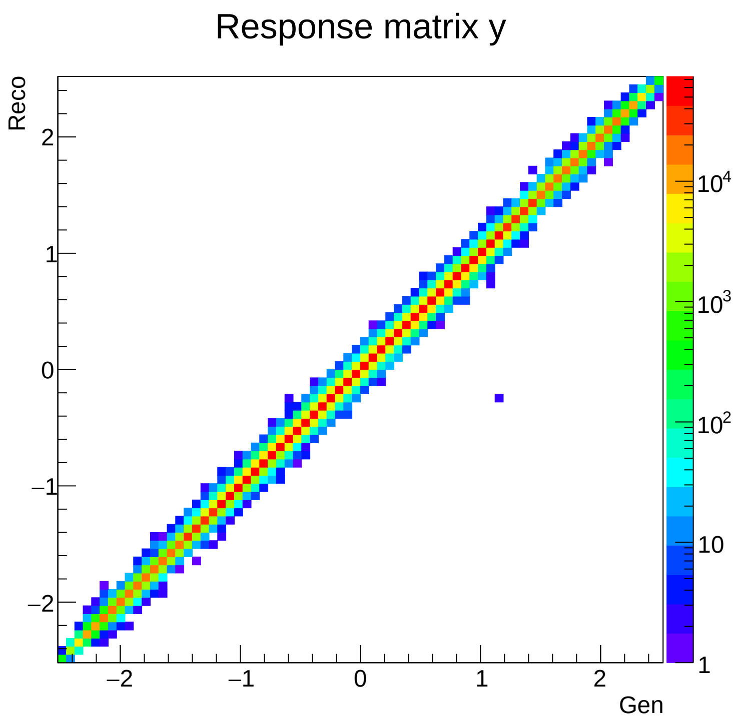

Figure 5.19: Response matrix for the rapidity (y) of the Z boson.

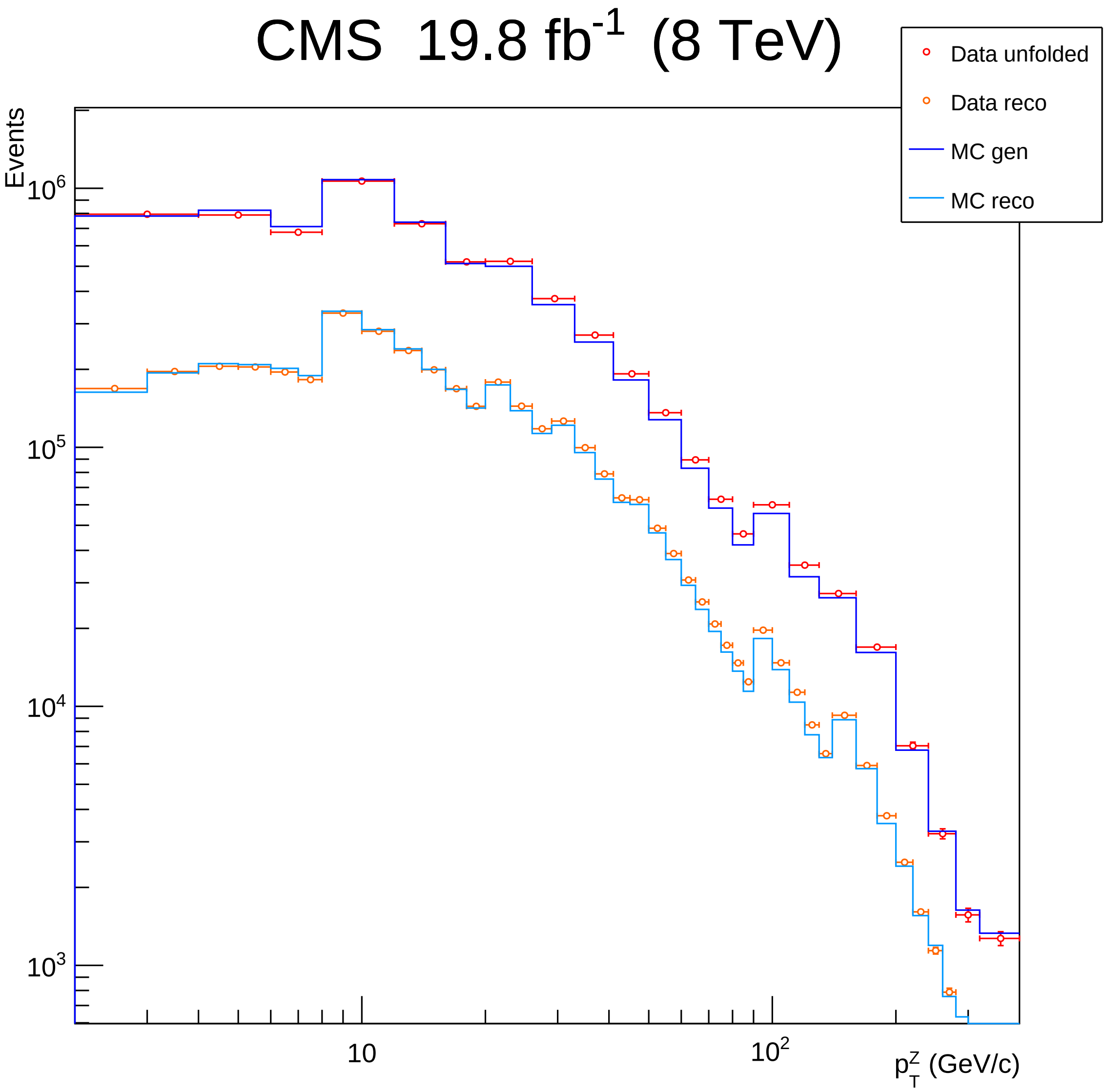

Figure 5.20: Results of the unfolding procedure for the transverse momentum (p_T) of the Z boson (red dots). In addition, the distribution of the measured data (the reconstructed distribution, in orange dots), and the Monte Carlo generated (blue histogram) and reconstructed distributions (azure histogram) are also reported.

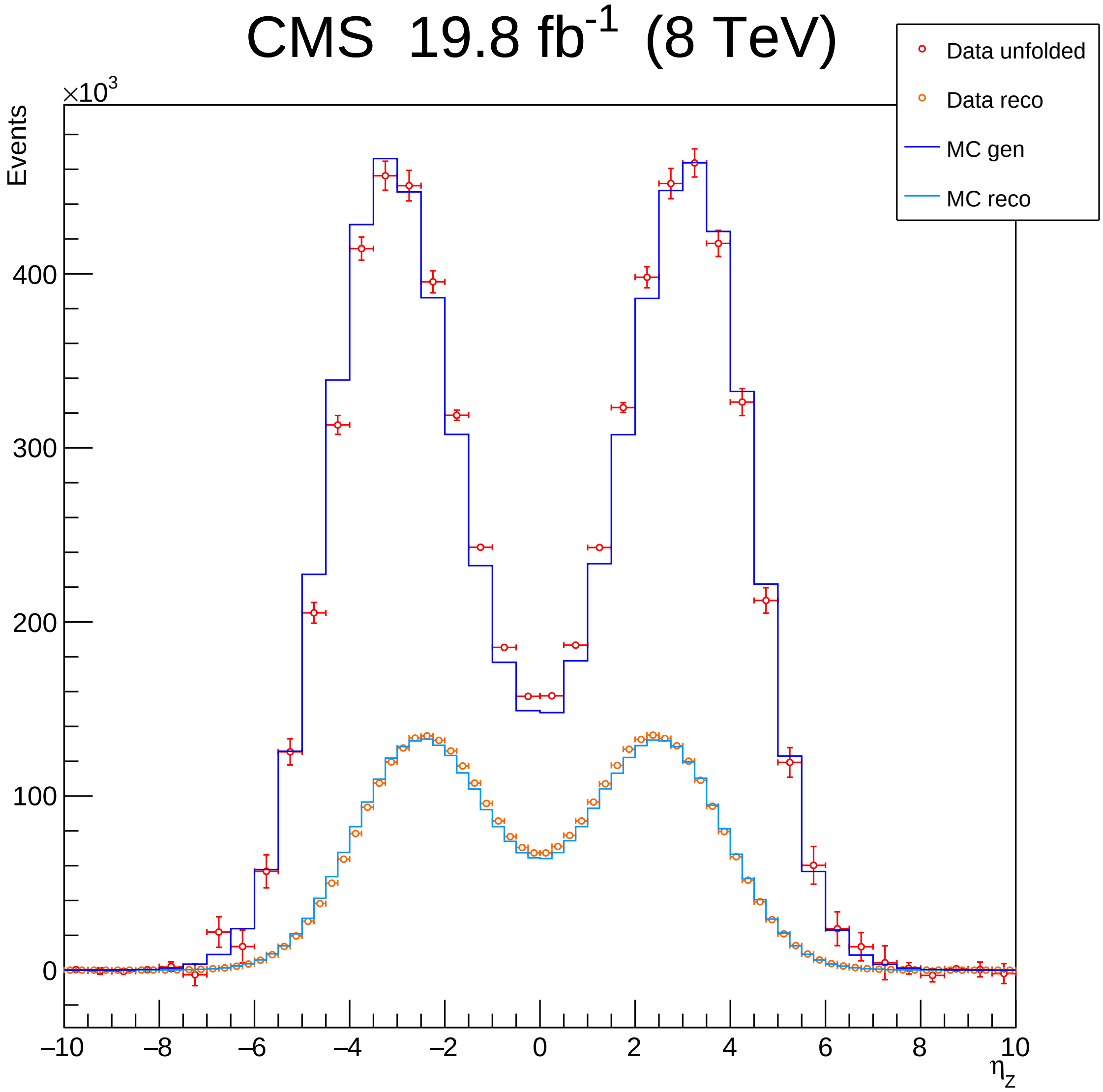

Figure 5.21: Results of the unfolding procedure for the pseudorapidity (η) of the Z boson (red dots). In addition, the distribution of the measured data (the reconstructed distribution, in orange dots), and the Monte Carlo generated (blue histogram) and reconstructed distributions (azure histogram) are also reported.

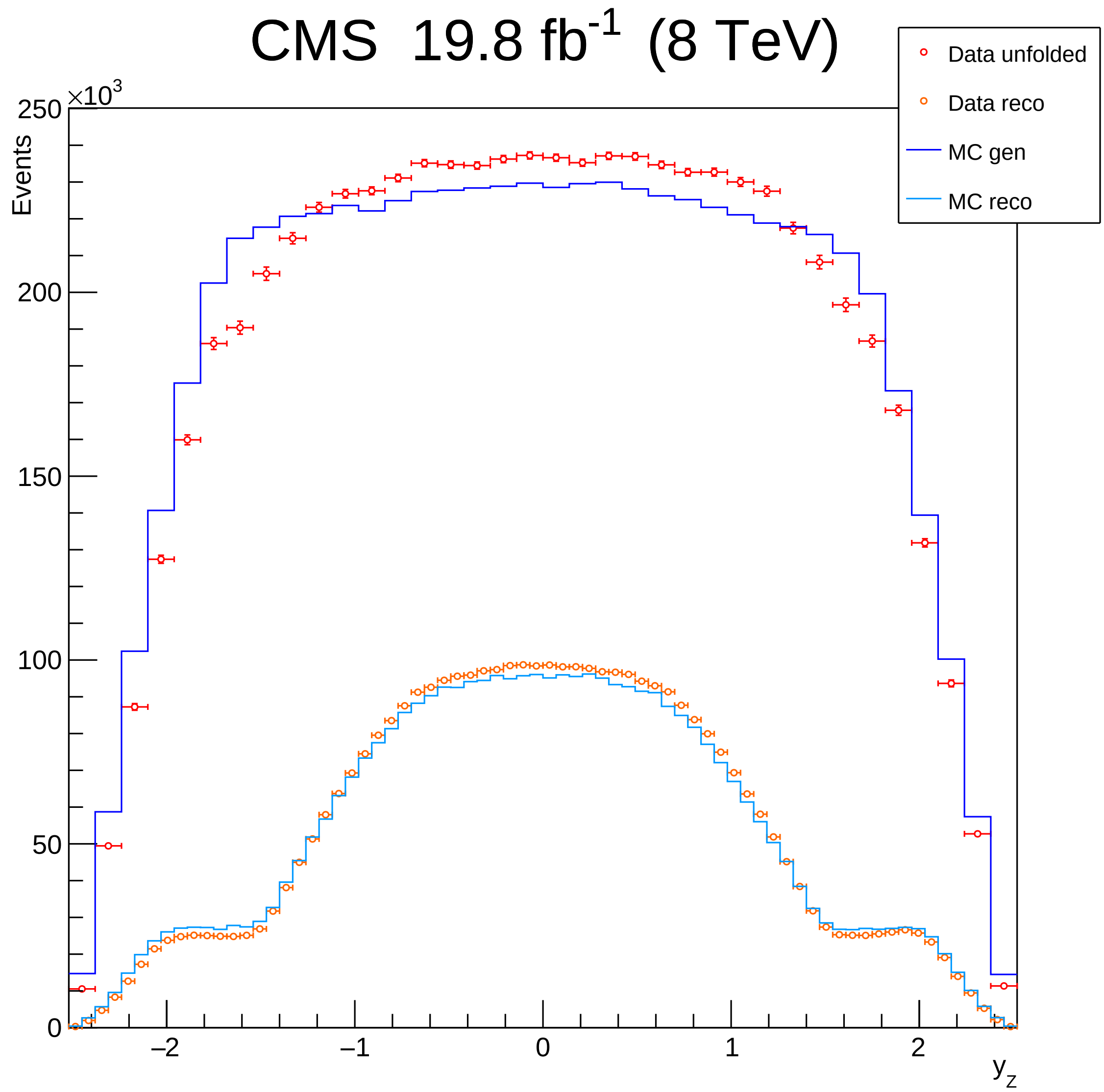

Figure 5.22: Results of the unfolding procedure for the rapidity (y) of the Z boson (red dots). In addition, the distribution of the measured data (the reconstructed distribution, in orange dots), and the Monte Carlo generated (blue histogram) and reconstructed distributions (azure histogram) are also reported.

5.7 Differential cross sections

In Figures 5.23 ÷ 5.25 the differential cross sections distributions are shown: the unfolded data distributions (red dots) and the Drell-Yan Monte Carlo generated simulations (violet colour) are divided bin by bin for the bin width and the luminosity (19.78 fb^{-1}) in order to obtain the differential cross sections measured in pb. The unfolded data are then compared with the other Monte Carlo generated distributions, likewise divided for the bin widths and the luminosity: the ones from the CT10 set (beige colour) and the ones from the NNPDF21_100 set (green colour). For the data, in particular for the ratio plots, both the statistical errors (black bars), calculated from the unfolding, and the total errors (red bars) are displayed. For the simulations, instead, only the total errors are shown.

The total error for the i-th bin of the data distributions is calculated in the following way:

$$\Delta O_i^{tot} = \Delta O_i^{unf} \oplus (2.6\% \cdot O_i) \oplus (1.5\% \cdot O_i) \oplus \epsilon(|1 - r_i| \cdot O_i). \tag{5.5}$$

The total error (systematic+statistical) calculated in the unfolding procedure, ΔO_i^{unf}, is summed in quadrature with the uncertainty on the luminosity (2.6% multiplied by the value of the bin O_i), with the inefficiency on the reconstruction for electrons (1.5% multiplied by O_i) and, if the distribution is the one of the transverse momentum, with the systematic uncertainty on the energy scale, calculated as mentioned above: ϵ has value 0 and takes value 1 only for p_T distribution, and r_i is the ratio between the two methods (the normal one and the modified one) of unfolding the distribution for the i-th bin.

The error shown for the CTEQ6L1 simulation is the statistical one only. Instead, the total errors for the other simulations are calculated in different ways. For the CT10 and the NNPDF21_100 PDFs, the events include both the statistical and the systematic error from the PDF set, and the computation depends on the method used for the creation of the set. For every set, different samples were available and they can be used to correctly determine the errors.

For the CT10 set, which has 53 different samples, for the i-th bin, the statistical error is summed in quadrature with:

$$\Delta O_i = \frac{1}{2}(O_i^{max} - O_i^{min}) \tag{5.6}$$

where O_i^{max} is the maximum value reached in the i-th bin among all the samples and O_i^{min} the minimum one.

For the NNPDF21_100 set, a Monte Carlo approach was employed, so 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. 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. In this case, the uncertainty associated to an observable can simply be calculated as its standard deviation, so, for the i-th bin, the statistical error is summed in quadrature with:

$$\Delta O_i = \sqrt{\frac{1}{N_s - 1}\sum_j^{N_s}(O_{i,j} - \overline{O}_i)^2} \tag{5.7}$$

where N_s is the number of the samples available for the PDF set, i.e. 101, $O_{i,j}$ is the value of the i-th bin for the j-th sample and $\overline{O}_i$ is the average value for the i-th bin.

In Figure 5.23 the distribution of the Z boson transverse momentum differential cross section is displayed. A logarithmic scale is chosen for the x axis in order to better show the whole spectrum, including the part at low p_T. The distribution reveals as the differential cross section is a decreasing function in this observable. The ratio plots show that there is a good agreement for the data with the CTEQ6L1 set, the one used as reference in the Monte Carlo simulations. The ratio is compatible with the unity within the error bars for almost every bin. For the other sets, instead, the agreement is not so good, although in the region $30 < p_T < 120$ GeV/c there is a better agreement to the respect of the CTEQ6L1 set. Apart for a different normalisation factor of about 15% for the CT10 and the NNPDF21_100 sets, due to the different data sets employed and to the different ways of computation, there is a spread of the ratio plots bigger than in the CTEQ6L1 case, and the agreement at low transverse momentum is very bad.

In Figure 5.24 the distribution of the Z boson pseudorapidity differential cross section is displayed. The distribution reveals an asymmetry of unknown origin between the regions $\eta < -4$ and $\eta > 4$. There are two maxima at about $|\eta| = 3$ and a relative minimum at $\eta = 0$. As for the previous distribution, the ratio plots show that the CTEQ6L1 set has the best agreement with the data, while the other sets have the normalisation factor and a bigger spread that make the ratio non compatible with the unity; only in the external regions the agreement occurs.

In Figure 5.25 the distribution of the Z boson rapidity differential cross section is displayed. The distribution shape has a central flat region with external decreasing regions. In this case a small asymmetry is present only in the tails of the distribution. The central region ($|y| < 1.5$) is well fitted by the CTEQ6L1 set, while the agreement gets worse moving out from this region. The CTEQ6L1 set slightly underestimates the central region (although compatible within the errors), and a wider flat central region is present producing the discrepancy with the data: the decreasing starts later causing the overestimates shown in the ratio plot. The CT10 and the NNPDF21_100 sets, instead, a part for the normalisation factor, seem to be less spread and to have a less wide central plateau.

Considering all the distributions here displayed, the CTEQ6L1 parton distribution functions seem to be the best set to fit the differential cross sections at 8 TeV. The CT10 and the NNPDF21_100 sets, although in some cases have a better agreement, in general are not so good, taking also into account that a normalisation factor has to be introduced.

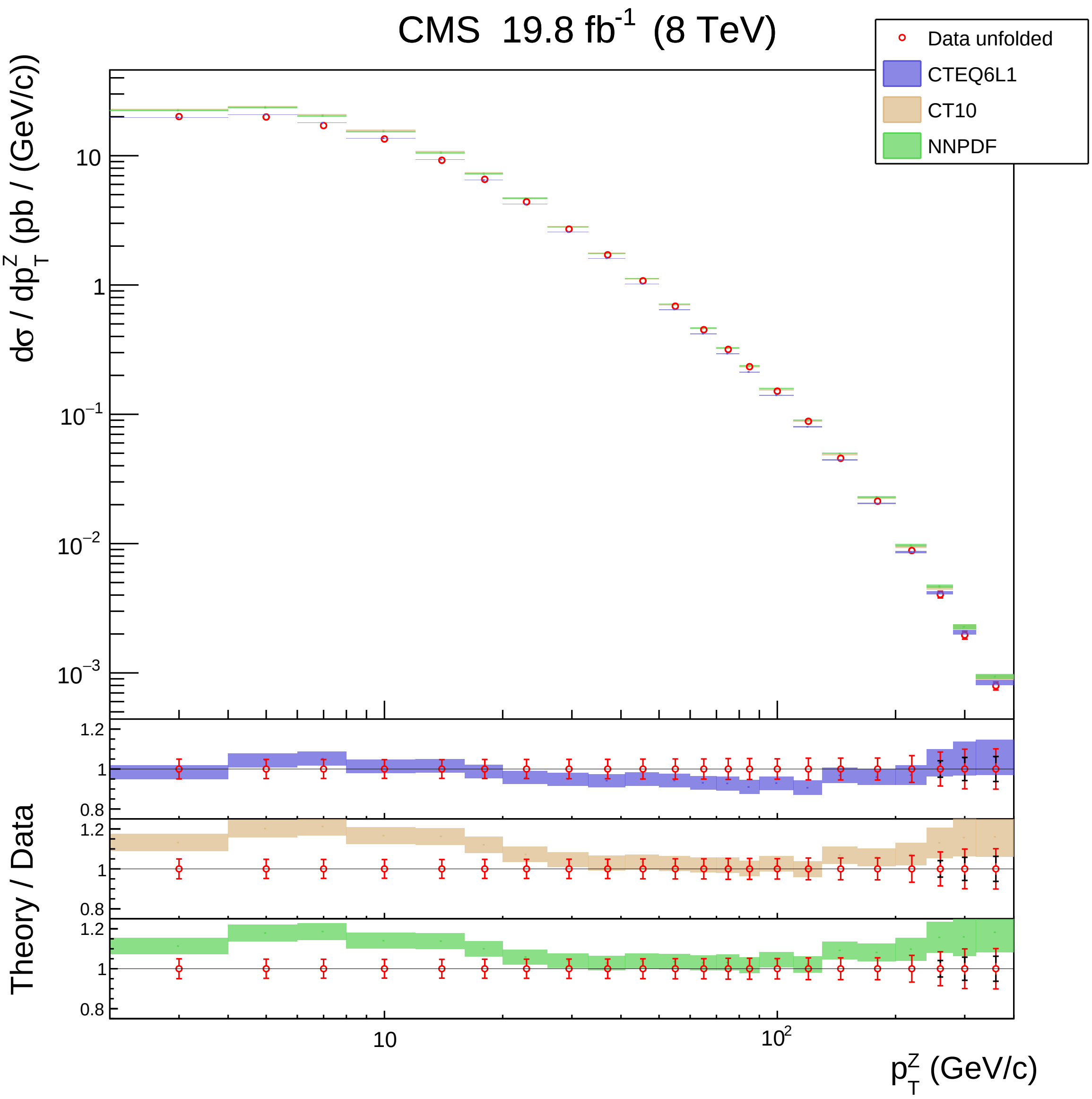

Figure 5.23: Unfolded differential cross section as a function of the transverse momentum (p_T) of the Z boson (red circles) compared with the theoretical prediction made with three different sets of PDFs. The ratio plots between the data and the theory are also displayed. The black error bars take into account only the statistical error, while the red error bars take into account also the systematic error.

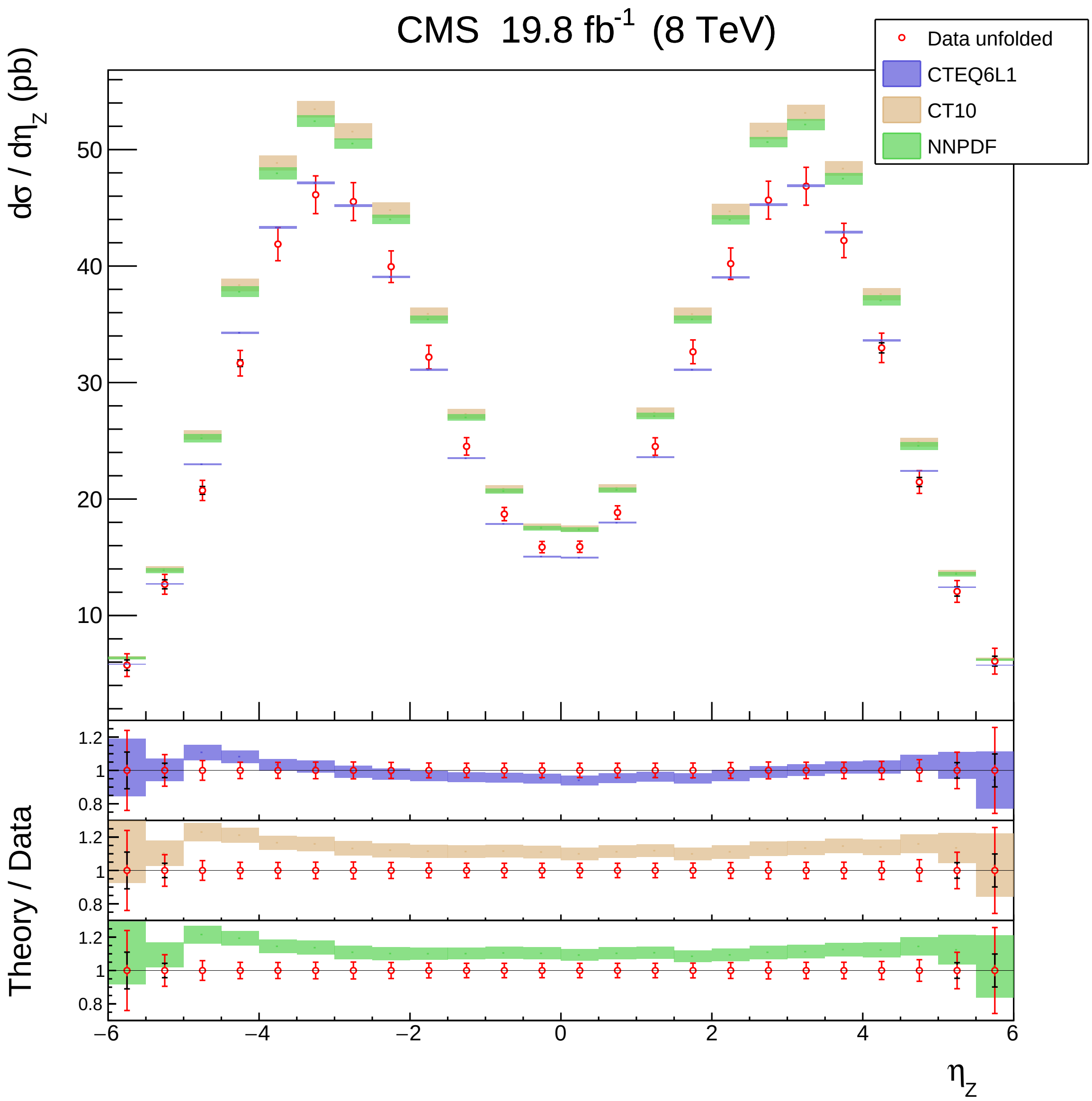

Figure 5.24: Unfolded differential cross section as a function of the pseudorapidy (η) of the Z boson (red circles) compared with the theoretical prediction made with three different sets of PDFs. The ratio plots between the data and the theory are also displayed. The black error bars take into account only the statistical error, while the red error bars take into account also the systematic error.

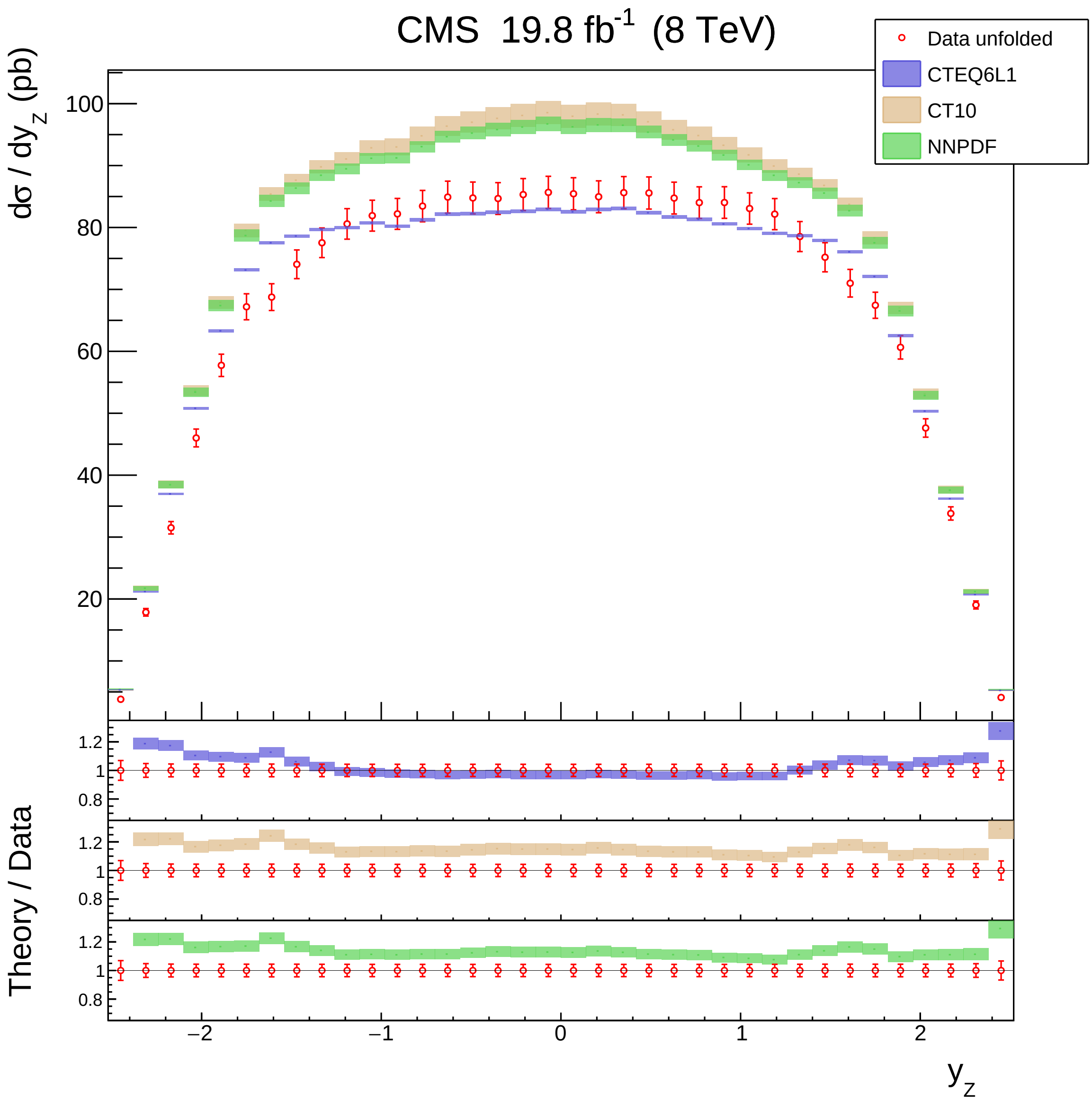

Figure 5.25: Unfolded differential cross section as a function of the rapidity (y) of the Z boson (red circles) compared with the theoretical prediction made with three different sets of PDFs. The ratio plots between the data and the theory are also displayed. The black error bars take into account only the statistical error, while the red error bars take into account also the systematic error.

Conclusions

In this thesis the differential cross section as a function of the transverse momentum, pseudorapidity and rapidity of the Z boson in Drell-Yan events has been studied at an energy of 8 TeV in the center of mass frame, using electron pairs collected by the CMS experiment. The results shown have been corrected for the detector effects with the unfolding procedure provided by the TUnfold package. Backgrounds have been studied from a Monte Carlo simulation provided by `MadGraph 5` and `PYTHIA 6` generators, employing the CTEQ6L1 PDF set. They are subtracted from the data distributions before the unfolding procedure. Two other theoretical predictions have also been used: they utilise the CT10 PDF set and the NNPDF21_100 PDF set. Except for the CTEQ6L1 simulation, in which the error is the statistical one only, in the other simulations and data distributions, the total error takes into account both the statistical and the systematic errors, calculated in the correct way for each distribution. The comparisons displayed in the last chapter show that for the observables examined in this work the data are well described by the CTEQ6L1 simulation, while for the other two theoretical predictions the agreement is not so good, in particular for the pseudorapidity and rapidity distributions. The normalisation factor correction may improve the agreement for the CT10 and NNPDF21_100 simulations, although they show also a bigger spread of the discrepancies in the ratio plots. In order to improve this measurement some further steps can be done. A comparison with other Monte Carlo simulations, produced both with the same and new PDF sets, can be useful, as well as an improved description of the detector. Moreover, the analysis may be repeated with the 13 TeV data, and the results compared.

Bibliography

[1] S. L. Glashow, *"Partial symmetries of weak interactions"*, Nuclear Physics 22: 579-588, 1961.

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

[3] A. Salam, *"Elementary Particle Physics: Relativistic Groups and Analyticity"*, Eighth Nobel Symposium: 367, 1968.

[4] G. Arnison et al., *"Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c^2 at the CERN SPS Collider"*, Phys. Lett., 126B: 398-410, 1983.

[5] 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(1): 1-29, 2012.

[6] CMS Collaboration, *"Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC"*, Phys. Lett. B, 716(1): 30-61, 2012.

[7] C.Patrignani et al. (Particle Data Group), *"Review of Particle Physics"*, Chin. Phys. C, 40, 100001 (2016).

[8] *https://it.wikipedia.org/wiki/Modello_standard.*

[9] E. Fermi, *"Trends to a Theory of beta Radiation"*, Nuovo Cim., 11: 1-19, 1934.

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

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

[12] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *"Advantages of the Color Octet Gluon Picture"*, Phys. Lett., B47: 365-368, 1973.

[13] S. Weinberg, *"Nonabelian Gauge Theories of the Strong Interactions"*, Phys. Rev. Lett., 31: 494-497, 1973.

[14] D. J. Gross and F. Wilczek, *"Asymptotically Free Gauge Theories"*, Phys. Rev., D8: 3633-3652, 1973.

[15] H1 and ZEUS Collaboration, *"Inclusive diffraction at HERA"*, Nucl. Phys. Proc. Suppl., 117: 403-407, 2003.

[16] G. Altarelli and G. Parisi, *"Asymptotic Freedom in Parton Language"*, Phys. Lett. B: 126:298, 1977.

[17] CMS Collaboration, *"PDF constraints and extraction of the strong coupling constant from the inclusive jet cross section at 7 TeV"*, Technical Report, arXiv:1410.6765.

[18] UA2 Collaboration, *"Evidence for Z_0 to e^+e^- at the CERN anti-p p Collider"*, Phys. Lett. B, 129: 130-140, 1983.

[19] $https://en.wikipedia.org/wiki/Drell-Yan\_process.$

[20] ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak, and Heavy Flavour Groups. *"Precision Electroweak Measurements on the Z Resonance"*, Phys. Rept., 427:257, 2006.

[21] CDF Collaboration, *"Measurements of inclusive W and Z cross sections in p anti-p collisions at s**(1/2) = 1.96-TeV"*, J. Phys., G34: 2457-2544, 2007.

[22] D0 Collaboration, *"Measurement of W and Z boson production cross sections"*, Phys. Rev., D60:052003, 1999.

[23] R. Boughezal, A. Guffanti, F. Petriello, M. Ubiali, *"The impact of the LHC Z-boson transverse momentum data on PDF determinations"*, Submitted to JHEP, arXiv:1705.00343v1.

[24] S. A. Malik and G. Watt, *"Ratios of W and Z Cross Sections at Large Boson p_T as a Constraint on PDFs and Background to New Physics"*, JHEP 02 (2014) 025, arXiv:1304.2424.

[25] O. Breuning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, P. Proudlock, *"LHC Design Report, Volume 1: the LHC main ring"*, CERN, Geneva - 2004.

[26] $https://en.wikipedia.org/wiki/Large\_Hadron\_Collider.$

[27] $https://it.wikipedia.org/wiki/Pseudorapidit\grave{a}.$

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

[29] $http://lhcitalia.infn.it/index.php/cms.$

[30] $http://www1b.physik.rwth-aachen.de/\ schael/CMS\_Tracker.html.$

[31] P. Adzic et al, *"Energy resolution of the barrel of the CMS Electromagnetic Calorimeter"*, JINST, 2:P04004, 2007.

[32] CMS Collaboration, *"CMS reconstruction improvement for the muon tracking by the RPC chambers"*, JINST 8 (2013) T03001, arXiv:1209.2646.

[33] CMS Collaboration, *"Measurement of the rapidity and transverse momentum distributions of Z bosons in pp collisions at $\sqrt{s}$ = 7 TeV"*, Phys. Rev. D 85 (2012) 032002, arXiv:1110.4973.

[34] CMS Collaboration, *"Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton-proton collisions at 8 TeV"*, Phys. Lett. B 749 (2015) 187, arXiv:1504.03511.

[35] ATLAS Collaboration, *"Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector"*, JHEP 09 (2014) 145, arXiv:1406.3660.

[36] ATLAS Collaboration, *"Measurement of the transverse momentum and ϕ^*_η distributions of Drell-Yan lepton pairs in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*, Eur. Phys. J. C 76 (2016), no. 5 291, arXiv:1512.02192.

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

[38] CMS Collaboration, *"Particle-Flow commissioning with muons and electrons from J/ψ and W events at 7 TeV"*, Number CMS-PAS-PFT-10-003. 2010.

[39] CMS collaboration, *"Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV"*, JINST 10 (2015) 06 P06005, arXiv:1502.02701.

[40] CMS Collaboration, *"Description and performance of track and primary vertex reconstruction with the CMS tracker"*, JINST 9 (2014) P10009, arXiv:1405.6569.

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

[42] *https://www.scientificcomputing.com/news/2014/06/high-energy-physics-predicting-emergence-jets.*

[43] *https://github.com/cms-ts/ZbAnalysis/blob/master/ZbSkim/macros/LumiInfo_v14.h.*

[44] *https://github.com/CERN-PH-CMG/cmg-cmssw/blob/CMG_PAT_V5_18_from-CMSSW_5_3_14/CMGTools/TTHAnalysis/python/plotter/mca.txt.*

[45] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *"Mad-Graph 5: Going Beyond"*, JHEP, 1106:128, 2011.

[46] *https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV.*

[47] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands, *"PYTHIA 6.4 Physics and Manual"*, JHEP 0605:026, 2006.

[48] S. Forte and G. Watt, *"Progress in the Determination of the Partonic Structure of the Proton"*, Annu. Rev. Nucl. Part. Sci. 63 (2013) 291-328, arXiv:1301.6754.

[49] G. Cowan, *"Statistical Data Analysis"*, Oxford University Press, 1998.

[50] S. Schmitt. *"TUnfold, an algorithm for correcting migration effects in high energy physics"*, JINST 7 (2012) T10003, arXiv:1205.6201v4.

[51] S. Schmitt, *TUnfold version 17.0, http://www.desy.de/ sschmitt/tunfold.html.*

[52] R. Brun and F. Rademakers, *"ROOT-An object oriented data analysis framework"*, Nucl. Instrum. Meth. A 389 (1997) 81-86.

[53] *https://twiki.cern.ch/twiki/bin/viewauth/CMS/ScrecUnfolding.*

[54] T. Adye *"Unfolding algorithms and tests using RooUnfold"*, PHYSTAT 2011 Workshop, arXiv:1105.1160.

[55] CMS Collaboration, *"Measurement of the transverse momentum spectra of weak vector bosons produced in proton-proton collisions at $\sqrt{s} = 8$ TeV"*, JHEP 1702 (2017) 096, arXiv:1606.05864.

[56] Hurng-Chun Lee, *"Single Top Quark Production at the LHC"*, GVO drukkers & vormgevers B.V.

[57] *https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideEgammaShowerShape.*

[58] CMS Collaboration, *"CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update"*, CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013.

Ringraziamenti

A conclusione di questa tesi ci tenevo a ringraziare alcune persone che mi hanno aiutato e supportato in questi anni universitari e non solo. Innanzi tutto i miei genitori, Carlo e Graziella, che hanno reso possibile questa scelta e che mi sono sempre stati accanto. In secondo luogo i Nonni per il supporto e l'affetto datomi fin da quando sono nato. Ringrazio zii, zie e cugini di ogni latitudine, per suggerimenti e consigli, in particolare Sebastiano e Giulia che mi hanno"aperto la strada" e in parte guidato negli anni appena trascorsi. Un ringraziamento di cuore va ad Ivana per l'aiuto, la tanta pazienza nei miei confronti e per tutte le cose belle che hai fatto per me. Ringrazio il mio relatore, Giuseppe Della Ricca, per il lavoro eccezionale che ha fatto e la grande disponibilità mostrata. E per gli innumerevoli pranzi leggendo "Il Piccolo", che hanno spezzato con allegria anche i momenti più stressanti. Non posso dimenticare Federico, per la compagnia in ufficio e i preziosi consigli negli ultimi mesi, e Alessandro, con cui ho diviso le fortune e le sfortune della magistrale e della tesi. Ringrazio Emanuele Di Marco, il mio correlatore, per il lavoro fatto e le dritte fornite, Chiara Rovelli e Fabio Cossutti per i preziosi suggerimenti. In questa lista non possono poi mancare gli amici che non hanno a che fare con il mondo della fisica, ma che sono stati comunque presenti in qualche modo: Enea, Enrico, Erik, Francesco, Jacopo, Matteo, Thomas, che hanno contribuito in maniera senz'altro originale a questi miei ultimi anni. Un grande ringraziamento va al mio maestro di judo, Gastone, che mi ha insegnato tanto da quando avevo 7 anni, non solo sportivamente. Grazie anche ai compagni di allenamento del corso amatori della SGT, in particolare al maestro Michele, ai miei "uke" Cristiana e Maurizio. Grazie di cuore a tutti voi.

In memoria di Twister, il quarto cugino.