Università degli Studi di Trieste

DIPARTIMENTO DI FISICA
Dottorato di Ricerca in Fisica
Coordinatore: Prof. Paolo Camerini

Study of the associated production of a Z boson and jets in pp collisions at $\sqrt{s} = 7$ TeV at CMS

Candidato
Damiana Montanino

Supervisori
Dott. Fabio Cossutti
Dott. Giuseppe Della Ricca

Revisori
Prof. Attilio Andreazza
Dott. Maarten Boonekamp

XXV ciclo di Dottorato

"A Kansas City Shuffle is when everybody looks right, you go left."
...
"Sorry about that, son.
But sometimes there's more to life than just livin'.
Besides, you can't have a Kansas City Shuffle without a body.
Mr. Goodkat

"Kore wa anata no Maneki Neko dewa arimasen.
Kore wa watashi no Kuro Neko desu.
from "Deliri in caffetteria"

Contents

Introduction

Between 1960 and 1970, a new theory, the so-called Standard Model (SM), was born in order to explain the interactions between the elementary particles. It has become well supported by many experimental observations like the discovery of the weak neutral currents in 1974 at CERN with the Gargamelle bubble chamber, the discovery of the W and Z bosons in 1983 at CERN with the S$p\bar{p}$S collider, the discovery of the top quark in 1995 at the Tevatron ($p\bar{p}$) collider, and since 2012 the observation of a new particle that could correspond to the Higgs boson, the particle that according to the SM theory is responsible of the mass of all other SM particles. The discovery of the Higgs boson was one of the main goals pursued at the LHC (Large Hadron Collider) and now the focus has been moved to the study of the properties of this new particle.

The fact that the SM theory is only a limit of a more general still unknown theory, valid at the energy accessible until now, is a common belief. The SM theory, in fact, is not able to provide an answer to relevant results from astrophysics and cosmological observations, as the lack of the presence of a particle candidate able to explain the dark matter. For these reasons several theories alternative to the SM were proposed which are subject to many studies at the LHC.

After two decades of development and construction, the LHC started its operation at the end of 2009. It is the most powerful proton-proton collider ever built and is designed to reach a luminosity of 10^{34} cm^{-2}s^{-1}. During the 2011 data taking the LHC reached $5 \cdot 10^{33}$ cm^{-2}s^{-1} with an energy of 7 TeV in the center of mass. The four main experiments associated to this collider are ALICE, ATLAS, LHCb, and CMS. This work exploits the data collected by the CMS experiment, for which the 2011 integrated luminosity was extremely high, being able to collect about 5 fb^{-1} of data.

In this thesis differential cross sections measurements of Z boson and jets associated production (Z + jets) will be presented; they are obtained by analysing a large part of the 2011 dataset that corresponds to a total integrated luminosity of 4.89 fb^{-1}.

This type of processes are central in the physics program of LHC. The study of jets produced with Z provides a stringent and important test of perturbative QCD. In addition, a precise measurement of the Z + jets cross sections is crucial since these processes are a significant background in searches for new physics.

This work is organized as follows. First an introduction to the SM theory, to the Z + jets physics and to the Monte Carlo generators is given. In the second chapter a summary of the main features of the CMS detector is presented. The description of how the reconstruction and the identification of the particles is done in CMS, with a particular attention to the electron and jets, is reported in the third chapter. In the fourth one, the performance of the electron reconstruction algorithms used in CMS are discussed. In the last chapter the measurement of the differential cross sections for the Z + jets associated production and the obtained results are described.

Chapter 1

Associated production of a Z boson and jets in the Standard Model

The understanding of the particles of all the matter surrounding us and their behaviour is based on two fundamental concepts, elementary components and symmetry conservation.

The former is introduced in order to simplify the description of the matter. With the combination of few elementary constituents following precise properties, it is possible to explain a large variety of composite objects and to describe their behaviour. The first attempt was done studying the multitude of particles discovered between the 40's and 50's and deriving hadrons from 3 basic quarks, up, down and strange.

The laws that establish the interactions between the constituents are derived from symmetry considerations coming from the latter concept. The particles are described by parameters that represent the sensitivity to the different interactions. They must be constant in order to identify univocally the particles. Thanks to the Noether's theorem, this indicates the association to a symmetry.

Considering these two elements together, it is possible to obtain a theory that provides a successful description of the experimental observation. A gauge theory is a physics theory based on the idea that the Lagrangian has to be globally and locally invariant if we apply a given transformation (symmetry). It means that the system has to be invariant for a given transformation even if the transformation is applied to a specific region of the spacetime. The Standard Model (SM) [1, 2, 3, 4] is a quantum field theory describing the elementary particles and is based on a $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ local gauge symmetry that introduces the fundamental interactions, where $SU(n)$ is the group of the unitary matrix n×n, with determinant=1, and $U(1)$ is a transformation of a complex phase. However, it does not provide a complete prediction of all the observable forces, in fact the gravity explanation is absent, and some contradictions have been observed.

In the following a brief overview of the SM is presented. The main properties

of the electroweak and strong interaction are described, then the main topic of this work, the associated production of Z boson and jets, is introduced. The last section describes the principal features of a Monte Carlo generator.

1.1 Elementary particles

The ordinary matter is composed by fermions classified in three generations, ordered according to their increasing masses, and divided in two categories, six spin-$\frac{1}{2}$ particles called leptons and six spin-$\frac{1}{2}$ particles called quarks, with their respective anti-particles. The anti-particles have the same mass than the particles but an opposite electric charge with respect to the corresponding particles. They are the constituents of the anti-matter. According to the current knowledge of the Universe composition, its visible part is built essentially of ordinary matter, and up to now the reason of the asymmetry between matter and anti-matter is still unclear.

Leptons are observed in free states while quarks exist only in bound states called hadrons, divided in two categories: mesons, composed of a quark and an anti-quark, and baryons composed of three quarks.

The SM of the electro-weak interactions unifies in a single model both electromagnetic phenomena, described by Quantum Electro Dynamics (QED) and the weak interactions, going beyond the Fermi theory. Quantum Chromo Dynamics (QCD) is the standard accepted theory to explain strong interactions phenomenology. The combination of these two is commonly referred to as the SM of elementary interactions. All the known elementary particles interact with each other through these three interactions, mediated by spin-1 particles called bosons. The three forces of the SM and the respective force carriers are summarised:

- strong interaction: only between quarks and gluons, mediated by eight massless gluons g.
- weak interaction: between leptons and/or $W^{\pm}$, Z^0 bosons as well as quarks, mediated by W^+, W^- and Z^0 massive bosons.
- electromagnetic interaction: between all electrically charged particles, mediated by a massless photon γ.

The gravity force is the only interaction not described by the SM; it is negligible for typical distances and masses involved in the ordinary space-time particle physics. In Fig. 1.1 a summary of the fundamental particles, with their properties and classification, is reported.

The first three left columns of the table in Fig. 1.1 contain the fermions composing the matter; in the last column the bosons, mediators of the interactions, are indicated. All the considered elementary particles have been observed and there is no experimental evidence for the existence of a fourth generation of particles [5, 6].

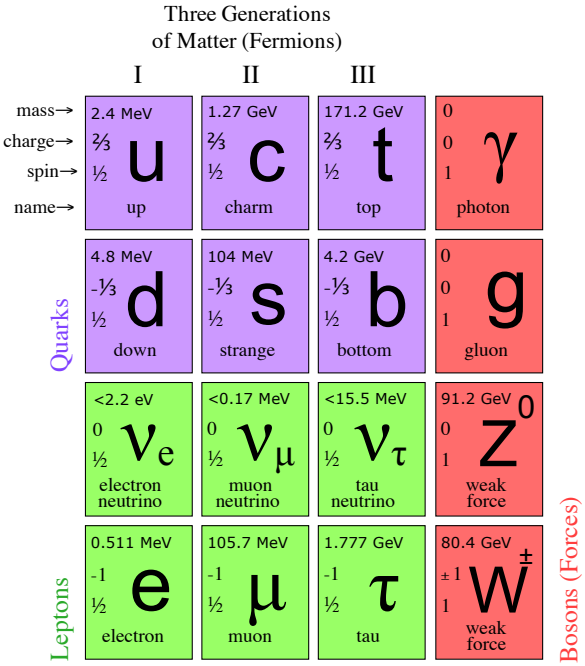


Figure 1.1: Schematic view of the Standard Model elementary particles.

1.2 The electroweak interaction

Quantum Electrodynamics (QED)

The electromagnetic force [7] is associated to the conservation of the electric charge, hence to the $U(1)$ symmetry and the propagation of the photon. In order to leave the Lagrangian invariant under the action of the transformation

$$\psi(x) \rightarrow e^{i\Lambda(x)}\psi(x), \tag{1.1}$$

a field $A_\alpha(x)$ is introduced which is modified by the gauge transformation $A_\alpha(x) \rightarrow A_\alpha(x) + \frac{1}{e}\partial_\alpha\Lambda$. The derivative must change to the covariant derivative $\partial_\alpha \rightarrow \mathcal{D}_\alpha = \partial_\alpha - ieA_\alpha$, in order to let the term $\bar{\psi}(i\gamma^\alpha\mathcal{D}_\alpha - m)\psi$ be invariant under the transformation 1.1, because the derivatives do not transform linearly under a transformation depending on a specific point of the spacetime x^μ.

The term corresponding to the propagation of the gauge field is also introduced,

$$-\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta}, \qquad \text{with} \tag{1.2}$$

$$F^{\alpha\beta} = \frac{i}{e}[\mathcal{D}_\alpha, \mathcal{D}_\beta] = \partial_\alpha A_\beta - \partial_\beta A_\alpha \tag{1.3}$$

where $F^{\alpha\beta}$ is the gauge field tensor.

The resulting Lagrangian is invariant under local gauge transformation:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\alpha\mathcal{D}_\alpha - m)\psi - \frac{1}{4}F_{\alpha\beta}F^{\alpha\beta}, \tag{1.4}$$

where the A_α represents the gauge boson, the photon, and the kinematic term its propagation.

The weak interaction

The first indications of the existence of the weak interaction were represented by the β decays, where the electrons are coupled with neutrinos through charged gauge bosons. Since this force was observed to act only on the left helicity particles, the helicity was chosen as the associated conserved parameter.

The simplest symmetry able to represent charged gauge boson is the $SU(2)$, a unitary non-abelian group composed by off-diagonal generators, τ_a, known as Pauli matrices:

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.5}$$

where τ_1 and τ_2 correspond to the contributions of the charged bosons $W^\pm$; τ_3 indicates the presence of an additional neutral boson.

The fermion field can be decomposed in its left- and righ-handed components

$$\psi = \psi_L + \psi_R, \tag{1.6}$$

where right-handed components are represented as singlets, while left-handed as doublets:

$$L_L = \frac{1}{2}(1-\gamma^5)\begin{pmatrix} \nu_l \\ l \end{pmatrix} = \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L, \quad L_R = \frac{1}{2}(1+\gamma^5)(l). \tag{1.7}$$

Equation 1.7 shows the general representation of leptons, where $L_{L,R}$ are the left/right handed component of the lepton spinor L and $\frac{1}{2}(1\pm\gamma^5)$ are the projectors on the chirality states, with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and γ^μ with μ=0,1,2,3 are the four Dirac matrices. Due to the parity transformation properties of the weak interaction [8, 9], only states with definite chirality can be involved in the interaction.

It can be demonstrated that for massless particles the definitions of chirality and helicity coincide, this is why only left-handed neutrinos can intervene [10]. In fact, even if the neutrino mass is not exactly null it is usually much smaller than its momentum, so the equivalence between helicity and chirality can be considered as valid with excellent approximation.

The quarks can be described with an equivalent representation, as shown in:

$$\begin{aligned} Q_L &= \frac{1}{2}(1-\gamma^5)\begin{pmatrix} u \\ d \end{pmatrix} = \begin{pmatrix} u \\ d \end{pmatrix}_L \\ u_R = \frac{1}{2}(1+\gamma^5)(u), &\qquad d_R = \frac{1}{2}(1+\gamma^5)(d) \end{aligned} \tag{1.8}$$

From now on, we will refer to leptons, but similar considerations can be done for the quark sector.

This description of the weak interaction was not satisfactory. While the contributions of the charged bosons $W^{\pm}$ can be related to the charged currents $J^{\pm}_{\mu} = \bar{L}_L \gamma_\mu \tau_{\pm} L_L$ with $\tau_{\pm} = \frac{1}{2}(\tau_1 \pm \tau_2)$, the third generator is not associated to any physical boson. It gives origin to the current $J^3_{\mu} = \bar{L}_L \gamma_\mu \tau_3 L_L$, which can not be identified as the weak neutral current, because it should involve both left-handed and right-handed components. The electromagnetic current satisfies this request, but does not couple with the chargeless neutrino.

In order to resolve the problem, a new symmetry was introduced, $U(1)_Y$, known as the hypercharged symmetry and corresponding to the current $J^Y_{\mu} = \bar{\psi}\gamma_\mu Y \psi$, where the hypercharge is $Y = 2(Q - \tau_3)$ and Q is the electric charge.

In this context the electromagnetic current can be expressed as a combination, correlating the $SU(2)_L$ and $U(1)_Y$ operators, $J^{em}_{\mu} = J^3_{\mu} + \frac{1}{2}J^Y_{\mu}$, resulting from the symmetry $SU(2)_L \otimes U(1)_Y$ [1, 2, 3] and representing the unification of the weak and the electromagnetic interactions.

The electroweak unification

The two interactions are expressed by the following Lagrangian, globally invariant under $SU(2)_L \otimes U(1)_Y$ gauge symmetry:

$$\mathcal{L}^{\text{lepton}}_{\text{EW}} = i \sum_{k=1}^{3} \left(\bar{L}^k_L \gamma^\mu \partial_\mu L^k_L + \bar{L}^k_R \gamma^\mu \partial_\mu L^k_R \right). \tag{1.9}$$

As for the electromagnetic case, two gauge fields are introduced. The B_μ field derives from the $U(1)_Y$ symmetry and behaves as shown before. The $SU(2)_L$ symmetry acts as $L_L \rightarrow e^{i\frac{\vec{\tau}}{2}\vec{\Lambda}} L_L$ and corresponds to $\overrightarrow{A}_\mu$ a vector field that transforms as $\overrightarrow{A}_\mu \rightarrow \overrightarrow{A}_\mu + \frac{1}{g}\partial_\mu \overrightarrow{\Lambda} + \overrightarrow{A}_\mu \times \overrightarrow{\Lambda}$.

As in the previous case, in order to switch from global to local gauge invariance we need to move to the covariant derivatives, $\mathcal{D}_\mu$, instead of the ordinary derivatives used in Eq. 1.9. Covariant derivatives are built to transform linearly under $SU(2)_L \otimes U(1)_Y$ symmetry and have the following form:

$$\partial_\mu \rightarrow \mathcal{D}_\mu = \partial_\mu - ig\frac{\overrightarrow{\tau}}{2} \cdot \overrightarrow{A}_\mu - \frac{1}{2} i g' Y B_\mu \tag{1.10}$$

where g and $g^{'}$ are the coupling constants for the gauge fields; Y is the weak hypercharge and $\overrightarrow{\tau}$ are the Pauli matrices.

The field tensors associated to the two $U(1)_Y$ and $SU(2)_L$ symmetry are

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.11}$$

$$\overrightarrow{E}_{\mu\nu} = \partial_\mu \overrightarrow{A}_\nu - \partial_\nu \overrightarrow{A}_\mu + g(\overrightarrow{A}_\mu \times \overrightarrow{A}_\nu), \tag{1.12}$$

and are used to define the kinetic term of the Lagrangian, having the following expression:

$$\mathcal{L}_{\mathrm{EW}}^{\mathrm{lepton}} = \bar{L} i\gamma^{\mu}\mathcal{D}_{\mu}L - \frac{1}{4}\overrightarrow{E}_{\mu\nu}\cdot\overrightarrow{E}^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.13}$$

When these variable substitutions are applied

$$\begin{aligned} W_{\mu}^{\pm} &= \frac{1}{\sqrt{2}}(A_{\mu 1} \mp iA_{\mu 2}), \\ Z_{\mu} &= \cos\theta_W A_{\mu 3} - \sin\theta_W B_{\mu}, \\ A_{\mu} &= \sin\theta_W A_{\mu 3} + \cos\theta_W B_{\mu}, \end{aligned} \tag{1.14}$$

where $\tan\theta_W = \frac{g'}{g}$ and θ_W is the Weinberg angle, we obtain the two interaction terms

$$\mathcal{L}_{\mathrm{EW}}^{cc} = \frac{g}{\sqrt{2}}\left(\bar{L}_L\gamma^{\mu}\tau^{+}L_L W_{\mu}^{+} + \bar{L}_L\gamma^{\mu}\tau^{-}L_L W_{\mu}^{-}\right), \tag{1.15}$$

$$\begin{aligned} \mathcal{L}_{\mathrm{EW}}^{nc} = &- i\left(g\sin\theta_W J_{\mu}^{3} + g'\cos\theta_W\frac{J_{\mu}^{Y}}{2}\right)A^{\mu} \\ &- i\left(g\cos\theta_W J_{\mu}^{3} - g'\sin\theta_W\frac{J_{\mu}^{Y}}{2}\right)Z^{\mu}. \end{aligned} \tag{1.16}$$

The Lagrangian in Eq. 1.15 describes the coupling between the charged leptons and the charged bosons and is called weak charged current interaction, while the one in Eq. 1.16, that involves vertices with zero net charge, is the weak neutral current interaction. The first term represents the contribution due to the electromagnetic force, the second the sole weak neutral part.

Using this description, the gauge fields of the $SU(2)_L \otimes U(1)_Y$ symmetry represent the three gauge weak bosons W^+, W^- and Z^0 and the γ for the electromagnetic interaction. The obtained Lagrangian describes two forces, the weak and electromagnetic forces, which are actually two manifestations of the same fundamental interaction, the electroweak interaction. However, this Lagrangian does not contain mass terms neither for the fermions nor for the massive gauge bosons, in obvious contrast with the experimental observation. Hence artificially mass terms should be included in the Lagrangian, which in principle could break the $SU(2)_L \otimes U(1)_Y$ gauge symmetry. This problem can be solved adding one more field, the Higgs field [11].

1.3 Higgs field and electroweak interaction mediators

The Higgs field is a complex scalar doublet and the Lagrangian, in its global gauge invariant form, is written as:

$$\mathcal{L}_{\mathrm{Higgs}} = (\partial_{\mu}\phi^{\dagger})(\partial^{\mu}\phi) - \mu^2\phi^{\dagger}\phi - \lambda(\phi^{\dagger}\phi)^2, \tag{1.17}$$

where ϕ is a spin$-\frac{1}{2}$ spinor, μ is a complex parameter and λ is a real positive parameter. As for the electromagnetic and weak fields, the Lagrangian in this form is not invariant under $SU(2)_L \otimes U(1)_Y$ local transformation, only the scalar potential has this property. The normal derivatives must be replaced with the covariant ones introduced before, obtaining:

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 - \frac{1}{4} \overrightarrow{E}_{\mu\nu} \cdot \overrightarrow{E}_{\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{1.18}$$

Minimizing the potential, the fundamental state is simply $\phi = 0$ when $\mu^2 > 0$, while if $\mu^2 < 0$ the minimal condition can be written as $\mu^2 + 2\lambda\phi^\dagger\phi = 0$, corresponding to infinite different solutions. Choosing a particular ground state, the symmetry is broken:

$$\phi_0 = \begin{pmatrix} 0 \\ \eta \end{pmatrix}, \quad \text{with} \quad \eta = \sqrt{-\frac{\mu^2}{2\lambda}}. \tag{1.19}$$

Expanding the Lagrangian around the ground state (with $\phi(x) = \eta + h(x)/\sqrt{2}$), the Higgs Lagrangian becomes:

$$\begin{aligned} \mathcal{L}_{\text{Higgs}} = & \frac{1}{2} \partial_\mu h \partial^\mu h + \mu^2 h^2 \\ & - \frac{1}{4} A_{\mu\nu} A^{\mu\nu} \\ & - \frac{1}{4} (W^{+\dagger}_{\mu\nu} W^{\mu\nu+} + W^{-\dagger}_{\mu\nu} W^{\mu\nu-}) + \frac{g^2\eta^2}{4} (W^{+\dagger}_{\mu} W^{\mu+} + W^{-\dagger}_{\mu} W^{\mu-}) \\ & - \frac{1}{4} Z_{\mu\nu} Z^{\mu\nu} + \frac{g^2\eta^2}{4\cos^2\theta_W} Z_\mu Z^\mu \\ & + \text{interaction} - \text{terms} \end{aligned} \tag{1.20}$$

where $A_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$. The first line of Eq. 1.20 represents the Higgs boson scalar field, with mass $\sqrt{-2\mu^2}$; the second line represents a massless field identified with the electromagnetic field; the third line represents $W^\pm$ with a mass of $\frac{g\eta}{\sqrt{2}}$; the fourth line represents the Z field with a mass of $\frac{M_W}{\cos\theta_W}$ with $\eta = \frac{1}{\sqrt{2\sqrt{2}G_F}} = 174$ GeV.

The Higgs field solves also the problem of the fermion mass term, by the introduction of Yukawa couplings, invariant for the $SU(2)_L \otimes U(1)_Y$ symmetry:

$$\mathcal{L}_Y = -\lambda_l (\bar{L}_R \phi^\dagger L_L + \bar{L}_L \phi L_R), \tag{1.21}$$

where λ_l is the lepton coupling constant. Exploiting the Eq. 1.19 and substituting the chosen ground state value for the Higgs field, the Yukawa coupling provides two terms:

$$\mathcal{L}_Y = -\lambda_l \eta (\bar{l}_L l_R + \bar{l}_R l_L) - \frac{\lambda_l}{\sqrt{2}} (\bar{l}_L l_R + \bar{l}_R l_L) h, \tag{1.22}$$

where the l_L and l_R are the left- and right-handed component of the charged leptons and $m_l = -\lambda_l \eta$.

The quark sector is slightly more complicated: since both the generation constituents have the right-handed components, u_R and d_R, a conjugate to the Higgs field is necessary, where $\phi_c = i\tau_2 \phi^*$. The resulting Yukawa couplings are:

$$-\lambda_u(\bar{u}_R \phi_c^\dagger Q + \bar{Q}\phi_c u_R) - \lambda_d(\bar{d}_R \phi^\dagger Q + \bar{Q}\phi d_R) \tag{1.23}$$

Each right-handed singlet can be characterized by its mass eigenstate. Since a state can not be defined as "up" for a component and "down" for the other, a new object must be introduced to identify the mass eigenstate for the left-handed doublets: the quark flavour mixing matrix, or CKM matrix (Cabibbo-Kobayashi-Maskawa). If the "up" component is defined by the following expression

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad , \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad , \quad \begin{pmatrix} t \\ b \end{pmatrix}_L \quad , \tag{1.24}$$

the "down" mass eigenstates are defined as d'_L, s'_L, b'_L where the CKM matrix V is obtained as:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_L = V \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \tag{1.25}$$

The matrix is measured experimentally and it is non diagonal, confirming the difference between the weak and the mass eigenstates.

1.4 The strong interaction

The quark model proposed by Gell-Mann [12] in the 1964 was not accepted peacefully, but was hampered because no experimental evidence of the existence of single quarks was available and some predicted particles, as the Δ^{++}, would been formed by three u quarks apparently identical to each other, violating the Pauli's exclusion principle.

The idea of the "colour" quantum number was introduced by Han and Nambu [4] in the 1965. Quarks can exist in three different color states, conventionally called green, red, and blue, and only colourless quark bound states can exist. In this way it is possible to explain the non-observation of free quarks and the existence of the Δ^{++}, that does not violate the Pauli's exclusion principle since the three quarks are not identical but differ from each other by their color charge.

The description of the strong interaction in terms of a gauge theory, the Quantum Chromo Dynamics (QCD), was obtained only in the 1973 by Fritzsch [13], Gross and Wilczec [14] and Weinberg [15]. It can be included in the Lagrangian, extending the local gauge symmetry $SU(2)_L \otimes U(1)_Y$ with the orthogonal $SU(3)_C$ one, obtaining the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry.

The strong interaction Lagrangian is written as:

$$\mathcal{L}^{\mathrm{QCD}} = \mathcal{L}_{\mathrm{inv}} + \mathcal{L}_{\text{gauge fix}} + \mathcal{L}_{\mathrm{ghost}} \tag{1.26}$$

where the terms $\mathcal{L}_{\text{gauge fix}}$ and $\mathcal{L}_{\mathrm{ghost}}$ are needed for technical reasons due to the used strategy to normalize the QCD Lagrangian [16].

The $\mathcal{L}_{\mathrm{inv}}$ term, invariant under local $SU(3)_c$ transformations, is expressed as:

$$\mathcal{L}_{\mathrm{inv}} = \sum_f \bar{\psi}_f \left(i\gamma_\mu \mathcal{D}^\mu - m_f\right) \psi_f - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.27}$$

where f runs over the six quark fields (flavours), $\mathcal{D}^\mu$ is the covariant derivative

$$\mathcal{D}_\mu = \partial_\mu + i g_s A_{\mu a} T_a \tag{1.28}$$

and $F_{\mu\nu}$ is the field tensor

$$F_{\mu\nu} = \partial_\mu A_{\nu a} - \partial_\nu A_{\mu a} - g_s C_{abc} A_{\mu b} A_{\nu c} \tag{1.29}$$

where $A_{\mu a}$ are the fields of the eight coloured gluons, T_a are the eight generators of $SU(3)$, C_{abc} are the structure constants that define the commutation rules of the $SU(3)$ generators, and $\alpha_s = g_s/4\pi$ is the strong coupling constant.

Describing a collision between hadrons, their composite nature needs to be taken into account. Computing calculations, it is important to consider that the objects that undergo the strong interaction are not directly the hadrons themselves but their constituents: valence quarks, Dirac sea quarks and gluons. In a $2 \to N$ process only two of them interact and in order to describe the event, the process needs to be factorized in two components. The hard interaction between the single partons can be calculated exploiting the perturbative theory, while the partons dynamics inside the hadron does not depend on the process but on the sole momentum fraction carried by the partons (the Bjorken factor x [17]).

Taking into account a proton-proton collision, the cross section can be expressed as:

$$d\sigma_{pp\to N}(Q^2, \mu_F^2) = \int_0^1 dx_1 \int_0^1 dx_2 f_1(x_1, \mu_F^2) f_2(x_2, \mu_F^2) d\hat{\sigma}_{12\to N}(Q^2, \mu_F^2), \tag{1.30}$$

where Q^2 is the scale of the interaction; μ_F^2 indicates the scale at which the separation between hard and soft processes takes place; x_1 and x_2 are the momentum fractions of the proton momentum carried by the colliding partons; f_1 and f_2 are the parton density functions (PDFs) representing the probability that the parton carries a momentum fraction x_1 and x_2; the $\hat{\sigma}$ is the parton level cross section.

An important aspect of the theory is that, measuring the PDFs $f(x, \mu)$ at a fix scale value $Q = \mu$, it is possible to calculate their value $f(x, \mu')$ at any other scale

μ', (they must be large enough such that both $\alpha_s(\mu)$ and $\alpha_s(\mu')$ are small). The evolution of the parton density function is described by the DGLAP equation [18]:

$$\mu^2 \frac{d}{d\mu^2} f_i(x, \mu, \mu_F^2) = \int_x^1 \frac{d\xi}{\xi} P_{ij}\left(\frac{x}{\xi}, \alpha_s(\mu_F^2)\right) f_j(\xi, \mu, \mu_F^2) \tag{1.31}$$

where P_{ij} is a matrix with $i, j = q, g$, its elements are known as splitting functions and represent the probability that the parton i with energy scale μ evolves to the energy scale μ', with a fraction ξ of its momentum due to the emission of a parton j with a fraction $1-\xi$ of the initial parton i momentum. The expression in Eq. 1.31 is an integral-differential equation requiring boundary conditions to be solved. These can be obtained measuring the $f_i(x, Q)$ at a fix scale $Q = \mu$.

1.5 Radiative corrections

A fundamental characteristic of a theory is the capability to make predictions in order to check them with the experimental results. To fulfil this requirement and to obtain cross sections, the calculation problem deriving from ultraviolet divergent loop integrals, present both for the electroweak and strong interaction, needs to be considered. The problematic integrals are expressed as

$$\Gamma(p) = \int \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2 - M^2(p))^2} \tag{1.32}$$

where $M(p)$ depends on the incoming momentum p. The integral tends to the term $\int d^4k/k^4$, producing the logarithmic divergence to cure, and, in order to reduce it to a finite one, a parameter μ is introduced. The Eq. 1.32 becomes

$$\Gamma_{ren}(p, \mu) = \frac{i}{(4\pi)^2} \ln\left(\frac{M^2(p)}{\mu^2}\right) \tag{1.33}$$

To obtain reliable results, it is necessary to verify the independence of the prediction from the cut-off μ. The variable that we want to predict $O(p)$, where p is a set of momenta, is defined as a series with coefficients $c_k(p, \mu)$ and a coupling $g(\mu)$. Measuring this observable at a given value of momentum, the $g(\mu)$ can be extracted and $O(p)$ can be predicted at every other values of p.

Considering the QCD theory and applying this procedure to a cross section, that depends on the transferred momentum and the strong coupling α_s, it is possible to show [19] that the observable can be expressed as a function of $\alpha_s(Q^2)$, hence depends only on the running value of the coupling.

One of the most crucial differences of the QCD with respect to the QED is the type of correlation between the coupling and the transverse momentum. It is well known that the strength of their interaction quickly decreases, if we increase the distance between two electric charges. It is in fact inversely proportional to the square of the distance between the two electric charges. Taking into account the more general case of the electroweak interaction, its coupling constant tends to

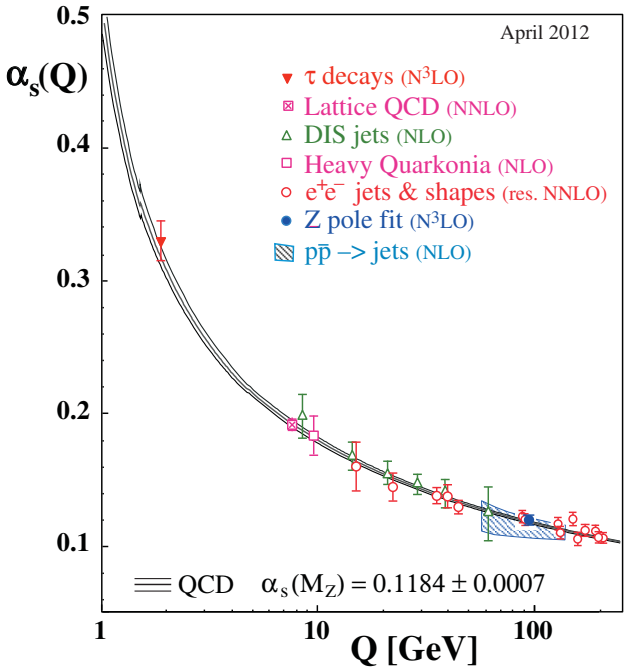


Figure 1.2: Illustration of the running of α_s in a theoretical calculation (shaded band) and in physical processes at different characteristic scales, from [20].

increase with greater energy, remaining sufficiently small at usual energies to apply the perturbative theory in order to compute predictive calculations.

On the contrary, the strong interaction exponentially increases with the distance between the color charges, therefore it decreases with higher scales. This effect is called asymptotic freedom. At the same time its value grows for low energies, explaining the experimental observation of sole bound quark states, the so-called confinement. The running of α_s is shown in Fig. 1.2. From an energetic point of view the parton will arrive at a distance where it will be more convenient to create new pairs of quark and anti-quark than to allow the original quarks to continue to move away. This process is known as hadronization and it is still one of the less understood processes of particle physics that is not analytically described.

For this reason in QCD only hard processes can be described by the perturbative method and their cross sections can be expressed as $\sigma = c_1\alpha_s + c_2\alpha_s^2 + \mathcal{O}(\alpha_s^3)$. As explained above, an observable defined as a series does not depend on the cut-off parameter μ, but this is not true in case of a truncated series. Indeed, QCD cross sections are calculated at different orders of the series, as leading order (LO), next-to-leading order (NLO) or next-to-next-to-leading order (NNLO). The latter two cases are the most useful in the context of precision measurements, having the lowest dependence.

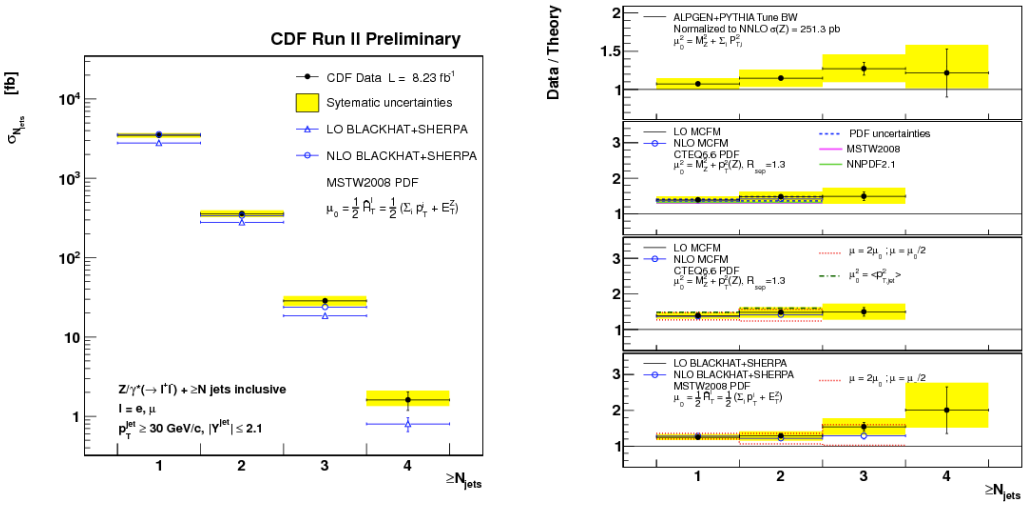


Figure 1.3: Measurement of the inclusive production cross sections of $Z + (n)jets$ for $n = 1 - 4$. Comparison is made to LO/NLO pQCD predictions from BLACKHAT+SHERPA and also presented as a ratio of data/theory in comparison with ALPGEN+PYTHIA predictions and LO/NLO pQCD predictions from MCFM for a variety of choices of scale and parton density functions.

1.6 Z + jets associated production

The analysis carried out in this thesis is focused on the characterization of the associated production of a Z boson and jets (Z+jets) in proton-proton collisions at LHC with the CMS detector. This study provides an important Standard Model (SM) test, a detector commissioning ground through physics, and last but not least a powerful probe for new phenomena.

In the context of the SM, the study of jets produced in association with Z allows for tests of perturbative QCD (pQCD) calculations. The leading order (LO) and next-to-leading order (NLO) predictions are in good agreement with data, but the latter, available for n up to 4 in the Z $+n$ jets final state, are only known with a precision varying from 10% up to 30% [21, 22], due to uncertainties on the parton distribution functions and on the perturbative nature of the calculations.

Z+jets studies were already presented at 1.96 TeV at Tevatron [23, 24, 25] and at 7 TeV with low statistics by ATLAS [26] and CMS [27]. As an example some results of the CDF collaboration are shown in Fig. 1.3.

The Z+jets production is an important background in searches on supersymmetry, in Higgs and Dark Matter signatures, and for studies of the top quark. Many extensions of the SM predict new particles with electroweak couplings that decay into SM gauge bosons accompanied by jets. In the case of Higgs physics, two processes can be taken as example. The first one involves the Higgs decay in two b quarks, that is an interesting channel due to its high branching ratio. In order to reduce significantly the QCD background, the production in association with a vector boson is also required, increasing the signal-to-noise ratio by a factor 10^3. The

production of the Z boson with light jets or heavy flavour ones represents one of the main backgrounds of this process, which is subtracted exploiting scale factors estimated by means of data-driven techniques, and distribution shapes taken directly from the Monte Carlo predictions.

Another interesting process is the Higgs decay in two Z bosons, considering their semileptonic decay that mimics the Z with 2 jets final state. With respect to the fully hadronic or leptonic decay, this represents the process with the highest cross section but also with the lowest signal-to-noise ratio. In order to improve this factor the requirement of b flavour jets can be also introduced, but this additional selection decreases the signal yield.

Considering supersymmetric phenomena, the gluino cascade can be mentioned. It consists in the production by means of the strong interaction of the supersymmetric partner of the gluon, the so-called gluino, that decays producing other new particles (even potential Dark Matter candidates). The detailed phenomenology of the process depends on the properties of the supersymmetric particles involved, but the final state is characterised by the presence of the Z boson decay products, jets and missing energy.

In all the examples described, the estimation of the background is an important and delicate operation due to the low expected signal yield. Hence the precise knowledge of the Z + jets physics represents a fundamental achievement in order to reduce the uncertainties on the processes mentioned above.

The main quantity exploited to characterize a process is the cross section, which, in the case of the associated production of a Z boson and jets, can be calculated in two steps: the Z boson creation and annihilation in a pair of lepton-antilepton described by the Drell-Yan process and the associated production of jets explained through the strong interaction.

1.6.1 Drell-Yan process

Considering a proton-proton collision, a pair of lepton-antilepton (l^+l^-) with large invariant mass, $M^2 = (p_{l^+} + p_{l^-})^2 \gg 1\ \mathrm{GeV}^2$, can be created through the annihilation of a quark-antiquark pair, as described by the Drell-Yan process (see Fig. 1.4):

$$pp \to q\bar{q} \to l^+l^- + X \tag{1.34}$$

where X represents a generic hadronic final state consistent with energy and momentum conservation.

In order to obtain the inclusive cross section $\sigma_{pp\to l^+l^-+X}$, all the possible subprocess cross sections $\hat{\sigma}_{q\bar{q}\to l^+l^-+X}$ have to be considered. Exploiting the general expression shown in Eq. 1.30 and summing over all the quark-antiquark combinations in the protons, it is possible to obtain the following σ_{DY}:

$$\sigma_{DY} = \sum_q \int dx_1 dx_2 f_q(x_1) f_{\bar{q}}(x_2) \hat{\sigma}_{q\bar{q}\to l^+l^-} \tag{1.35}$$

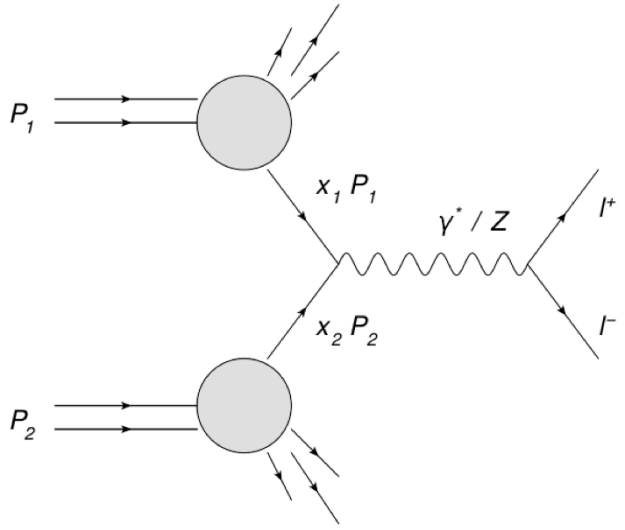


Figure 1.4: Leading order diagram of the Drell-Yan process $pp \to q\bar{q} \to l^+l^- + X$.

where the $f_q(x_1)$ and $f_{\bar{q}}(x_2)$ are the parton distribution functions, extracted from deep inelastic electron scattering experiments. The equation is formally valid in the limit where quarks are asymptotically free at zeroth order in α_s. The lowest-order total cross section for quark-antiquark annihilation into a lepton pair via a off mass-shell photon γ^* is given by:

$$\hat{\sigma}_{q(p_1)\bar{q}(p_2)\to l^+l^-} = \frac{4\pi\alpha^2}{2\hat{s}}\frac{1}{N_C}Q_q^2 \tag{1.36}$$

where $\hat{s} = (p_1 + p_2)^2 = x_1x_2s$, p_1 and p_2 are the parton four-momentum, N_C is the number of colours, Q_q^2 is the quark fractional charge and α is the electromagnetic coupling constant.

Since the incoming quark and antiquark can have a spectrum of collision energies $\sqrt{\hat{s}}$ rather than a fix value, it is more appropriate to consider the differential lepton pair mass distribution. The differential cross section for producing a lepton pair of invariant mass M is written as:

$$\frac{d\hat{\sigma}}{dM^2} = \frac{4\pi\alpha^2}{3M^2N_C}Q_q^2\delta(\hat{s} - M^2) \tag{1.37}$$

Substituting the Eq. 1.37 into the Eq. 1.35, it is possible to obtain the parton model differential cross section for the Drell-Yan process at the leading order:

$$\begin{aligned}\frac{d\sigma^{DY}}{dM^2} &= \int_0^1 dx_1dx_2 \sum_q f_q(x_1)f_{\bar{q}}(x_2) + (q \leftrightarrow \bar{q}) \times \frac{d\hat{\sigma}}{dM^2}(q\bar{q} \to l^+l^-) \\ &= \frac{4\pi\alpha^2}{3M^2N_C}\int_0^1 dx_1dx_2\delta(x_1x_2s - M^2) \times \left[\sum_q Q_q^2 f_q(x_1)f_{\bar{q}}(x_2) + (q \leftrightarrow \bar{q})\right]\end{aligned} \tag{1.38}$$

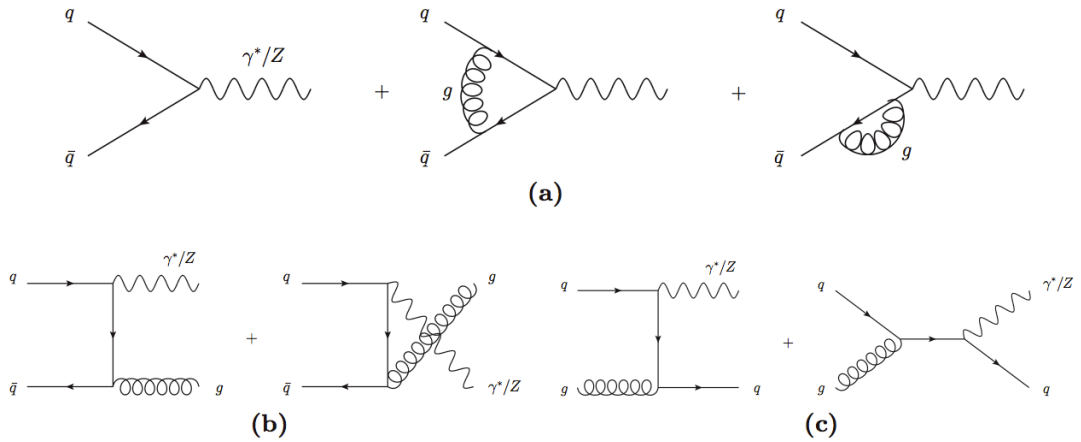


Figure 1.5: The leading- and next-to-leading-order diagrams for the Drell-Yan process.

In order to improve the precision of this LO expression, perturbative QCD corrections have to be considered. The diagrams which contribute at $\mathcal{O}(\alpha_s)$ are shown in Fig. 1.5 and three different kind of contributions are identifiable: a) virtual gluon corrections to the LO contribution; b) real gluon corrections; c) quark-gluon scattering process together with the corresponding $\bar{q}g$ contribution. It is worth to note that virtual corrections do not modify the final state, while real corrections appear in the form of additional jets in the final state.

As a consequence of these corrections, the PDFs acquire a logaritmic mass dependence, modifying the Eq. 1.35 as in the following:

$$\sigma_{DY} = K \sum_q \int dx_1 dx_2 f_q(x_1, M^2) f_{\bar{q}}(x_2, M^2) \hat{\sigma}_{q\bar{q}\to l^+l^-} \tag{1.39}$$

The mass-dependent PDFs are obtained from studies of inclusive inelastic electron scattering and K represents a constant factor. Currently, calculations at the next-to-next-to-leading order (NNLO) are available [28].

All the calculations described up to now took into account an intermediate state of γ^*; to adapt them for the case of the Z boson production and its leptonic decay, it is sufficient to substitute the $\hat{\sigma}_{q\bar{q}\to\gamma^*\to l^+l^-}$ with the cross section for the process $\hat{\sigma}_{q\bar{q}\to Z\to l^+l^-}$:

$$\hat{\sigma}_{q\bar{q}\to Z\to l^+l^-} = \hat{\sigma}_{q\bar{q}\to Z} \cdot BR(Z \to l^+l^-) \tag{1.40}$$

where $\hat{\sigma}_{q\bar{q}\to Z}$ is the production cross section of the Z boson and $BR(Z \to l^+l^-)$ is the branching ratio of its leptonic decay. Since the decay width of the Z boson is small ($\Gamma_Z = 2.5$ GeV) compared to its mass ($M_Z \simeq 91$ GeV), the production of effectively stable particles can be considered[19]. The production cross section can be approximated as if the Z boson was on shell:

$$\hat{\sigma}_{q\bar{q}\to Z} = \frac{\pi}{3}\sqrt{2} G_F M_Z^2 (V_q^2 + A_q^2) \delta(\hat{s} - M_Z^2) \tag{1.41}$$

where V_q^2 and A_q^2 are associated with the vector and the axial coupling constants of the neutral current interaction.

1.6.2 Multijet production

The total Z production cross section can be written as the sum of multijet cross sections with increasing order in α_s:

$$\sigma_Z = \sigma_{Z+0jets} + \sigma_{Z+1jet} + \sigma_{Z+2jets} + ... \tag{1.42}$$

where each cross section can be expressed as:

$$\begin{aligned} \sigma_{Z+0jets} &= a_0 + \alpha_s a_1 + \alpha_s^2 a_2 + ... \\ \sigma_{Z+1jet} &= \alpha_s b_1 + \alpha_s^2 b_2 + ... \\ \sigma_{Z+2jets} &= \alpha_s^2 c_2 + ... \\ &... \end{aligned} \tag{1.43}$$

The coefficients a_i, b_j, c_k, ... in these expansions are in general functions of the jet-definition parameters, for example the cone size used to cluster the partons into jets, the transverse momentum, rapidity and separation cuts imposed on the jets or the clusters. It is worth to note that the sum of the parameters at each order in perturbative theory

$$\begin{aligned} a_0 &= \delta_0 \\ a_1 + b_1 &= \delta_1 \\ a_2 + b_2 + c_2 &= \delta_2 \end{aligned} \tag{1.44}$$

is independent from the jet parameters and represents the perturbative expansion in power of α_s of the total cross section.

The greatest contribution to the $Z + jets$ cross section is given by the first coefficients of the exclusive multijet cross sections, a_i, b_j, c_k, ... They can be obtained from the Feynman diagrams of the partonic processes $xy \to Z + j_1...j_n$, where x, y, j_i are quarks and gluons. The explicit calculations of these coefficients were carried out by Berends and Giele up to a multiplicity of 4 jets [29]. They also investigated the ratio between the cross section with n jets and the $n-1$ jets one:

$$f_n(Z) = \frac{\sigma_{Z+n_{jets}}}{\sigma_{Z+(n-1)_{jets}}} \tag{1.45}$$

founding that this expression should be constant. It is possible to parametrized $f_n(Z)$ as

$$f_n(Z) = \alpha + \beta n_{jets}. \tag{1.46}$$

This function has been tested in different experiments, as an example the first results of the CMS Collaboration [27] are shown in Fig. 1.6.

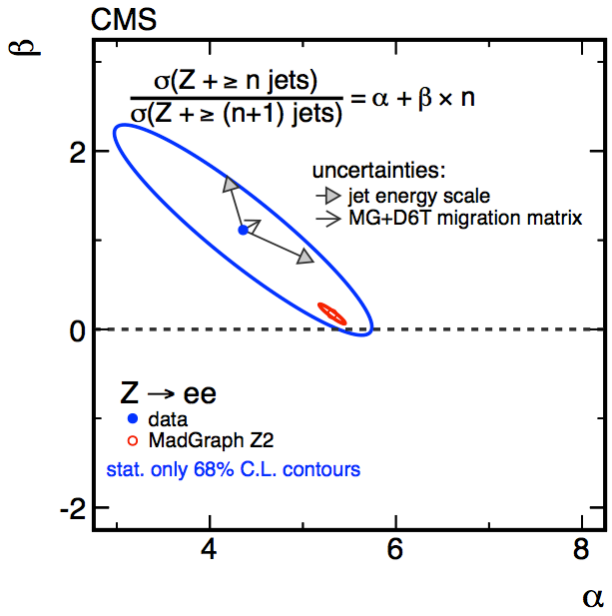


Figure 1.6: Fit results for the Berends-Giele scaling parameters α and β after pileup subtraction, efficiency corrections, and unfolding of detector resolution effects. The data are compared with the expectations from the `MadGraph` simulation with the Z2 tune. The ellipses correspond to 68% confidence level contours considering the statistical uncertainty only, for both data and simulation. The arrows show the displacement of the central value when varying each indicated parameter by its estimated uncertainty. The arrows labelled "MG+D6T migration matrix" correspond to the displacement when `MadGraph` simulation with the D6T tune is used for the unfolding.

1.7 Monte Carlo generators

In order to obtain a detailed description of the final state and due to the impossibility to perform the calculations analytically because of the hadronization, Monte Carlo event generators are exploited to simulate the processes. They are computer programs that subdivide the generation of the whole event in smaller and simpler steps that can be computed separately.

Several types of event generators are implemented and they distinguish from each other, for instance, because of the number of elements in the final state, the provided precision or the considered models. Usually, high precision calculations considering several orders in perturbation theory are available for a limited number of processes and in this way general predictions are hard to derive. However lower order calculations, usable on a high number of different processes, manage to describe with reasonable precision inclusive quantities.

The generator must be chosen carefully depending on the type of study carried out. `MadGraph` [30] and `PYTHIA` [31] are general purpose tree-level generators (LO in perturbative theory). The former is able to make precise calculations of the Matrix Elements (ME) for a number of processes, considering $2 \to N$ interactions. The latter completes the description of the event matching the ME outcome with approximated predictions of the radiation development exploiting the parton shower (PS) method and modelling the phenomenon of the hadronization. The principal steps of the generation procedure will be presented.

Generation phase

The complex process leading to the simulation of a hadron-hadron collision can be schematized as a sequence of simpler calculation steps as represented in Fig. 1.7:

- The cross sections of the considered hard process is obtained using the matrix element (it can be calculated at different perturbative orders) for a pair of incoming partons (quarks and gluons) inside the two colliding hadrons with given momenta.

- One parton out of each hadron is selected to enter the scattering process, according to the parton density functions. Final state partons and leptons are produced according to the calculated differential cross sections. Resonances produced in the hard event decay.

- The partons, that produce the hard event, can emit bremsstrahlung, the so-called Initial State Radiation (ISR), simulated with the Initial State Parton Showers. Also the final state partons can produce further radiation, known as Final State Radiation (FSR), simulated by the Final State Parton Showers.

- Partons which do not participate in the hard interaction can give rise to interactions with smaller transferred momentum, the Multiple Parton Interactions (MPI). They are a constituent of the underlying component of the event

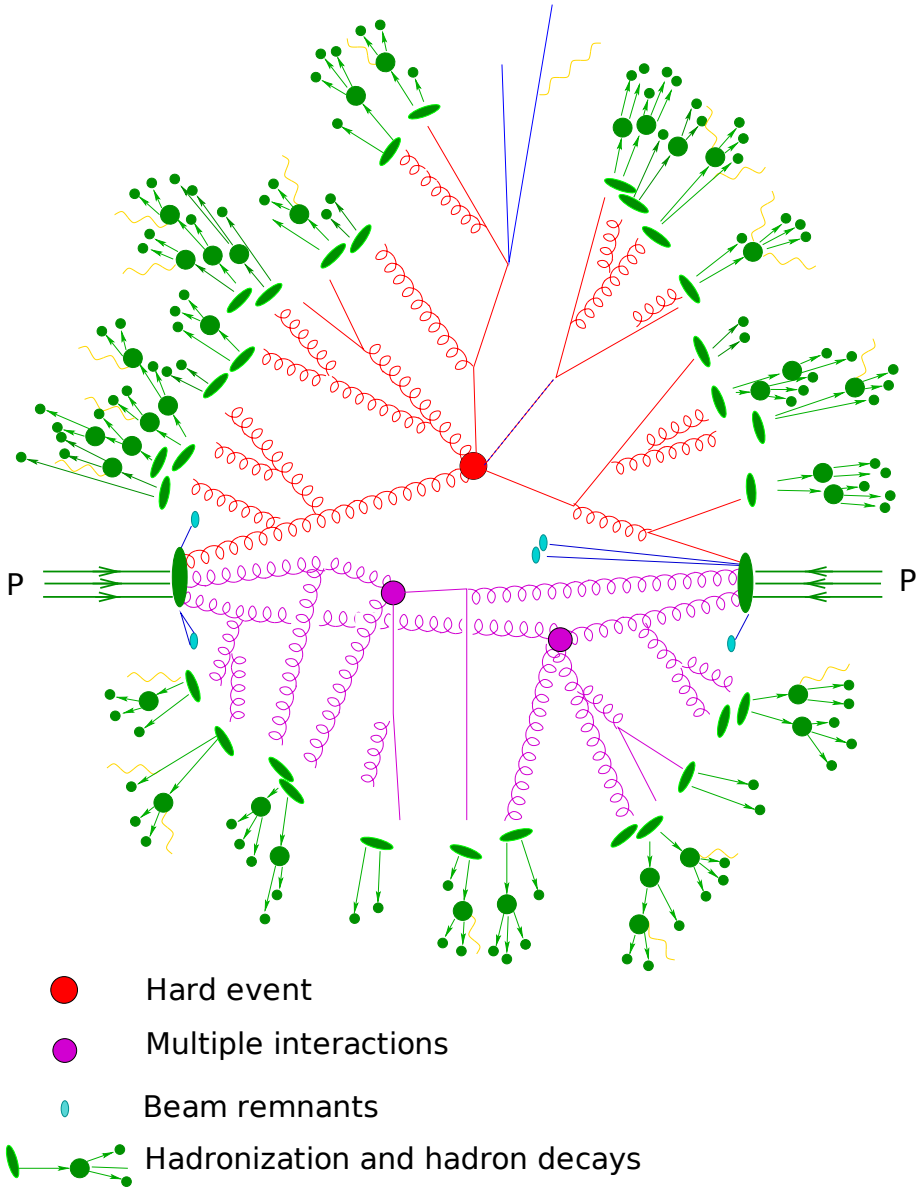


Figure 1.7: A schematic representation of the working chain of an event generator for a hadron collision. Partons from the two incoming hadrons participate in the hard scattering and in softer multiple interactions. Quarks and gluons hadronize and then hadrons decay.

and they also need to be corrected for ISR and FSR. Another constituent is composed by the remaining partons of the interacting hadrons, called beam remnants, that need to be simulated to balance the colour charge and for momentum conservation.

- The calculations described above are performed in the perturbative regime, but as the produced partons move apart from each other, the coupling constant increases giving rise to the confinement effect. When the coupling constant is strong enough, quark-antiquark pairs are produced from the vacuum and the partons hadronize. One of the most used model is the Lund string model [32] implemented in `PYTHIA`.

- At the end, the generator deals with the decay of τ leptons and all the other unstable particles.

Now the matrix element and the parton shower calculations followed by their matching procedure will be described in more detail.

Matrix Element calculations

As described before, one of the main ingredients for the description of an hadron collision is the calculation of the parton cross section of the hard interaction. Several different kinds of processes can be described by the existing general purpose event generators, even if it is often useful to interface such generators with dedicated hard process libraries when dealing with particular events.

Currently, the calculations produced exploiting the matrix element (ME) can be done at different orders both at LO and NLO, varying the number of element in the final state ($2 \to 2$ processes as in `PYTHIA` or $2 \to N$ processes as in `MadGraph`) exploiting the Feynman rules. Due to the complexity of these calculations, only a limited number of processes are described. Most of them are still computed only with the tree-level approach, which still plays an important role in the simulation of events produced at hadron colliders and can be performed up to several partons in the final state.

When the calculations are limited to only the tree-level diagrams, the loop corrections that would cancel the divergences due to soft and collinear emissions are not taken into account. In order to avoid them, only a limited phase space is considered, and the problematic regions are omitted. Since the resulting cross sections do not contain contributions due to this kind of processes, they need to be recovered using other techniques to be able to describe realistic events. This task is effectively taken care by parton showers calculations.

Parton Showers

As said before, in order to deal with complex events, as $2 \to N$ processes, the sole tree-level matrix element is not enough to cover the whole phase space. Since the ME

representing the parton evolution, up to the energy scale where the hadronization starts, is not known, the missing radiation description must be added.

In order to describe completely the interaction, parton shower algorithms (PS) [33, 34] represent an optimal alternative to the complex calculations needed to include the loop corrections corresponding to these divergences and to manage several successive branchings.

In order to describe the radiation produced by the accelerated coloured partons, their virtuality must be considered. Partons very far in time from the hard scattering are obviously on the mass shell, but this condition changes, due to the uncertainty principle when they become closer to the interaction. Approaching the hard scattering they start to go off-shell and to be able to emit harder gluons. This is the so-called Initial State Radiation. The virtuality of the emitting partons in such conditions is space-like.

After the hard interaction, the produced partons scatter far from each other allowing a lower virtuality therefore the emission of softer gluons gives rise to the process known as Final State Radiation. The emitting parton virtuality in these circumstances is time-like.

An important aspect of this approach is the resummation procedure. In perturbative QCD calculations, terms with the form $\alpha_s^n L^k$, with $k \leq 2n$ and $L = \ln q_{cut}/s$ where q_{cut} is the cutoff for resolved emission, need to be treated. Even if α_s is small enough to use the perturbative theory, the value of q_{cut} must be considered. It can be such small that the logarithm terms can spoil the convergence of the series. The resummation allows to handle these large logarithms exploiting an effective non analytical method, using the Sudakov form factors [35], describing the radiation exactly in the soft and collinear limit.

Matrix Element and Parton Showers matching

Analysing the different aspects of the two procedures described above, the optimal solution can be represented by the combined use of ME and PS to exploit the advantages of the two approaches in the phase space regions where each performs better.

The ME calculations are exact to a given order in perturbation theory but they need to be used only for well separated parton configurations, hence are not able to reproduce the internal structure of a jet. On the other hand, the PS is an approximation of the ME approach. It can handle soft and collinear divergences and can be used in the phase space not suitable for the ME calculations.

A solution is represented by the description of final states with n well separated partons with the corresponding n-partons tree-level matrix elements, adding the large logarithms resummation that characterizes the parton shower. The most difficult aspect of the procedure is to avoid holes in the phase space and possible double counting that can occur when a configuration with n partons emerging from the ME is produced also by an $(n-1)$-partons ME plus an additional hard emission coming from the PS.

The most common matching approaches, in case of matrix elements at LO order, are the so called MLM prescription [36] and the CKKW (Catani-Krauss-Kuhn-Webber) prescription [37]. The first one is implemented in the `MadGraph` generator (interfaced with the `PYTHIA` parton showers), while the second one is used in `Sherpa` [38].

Both matching prescriptions are based on the separation of the phase space in two regions, through a k_T measure cut-off. The first region is a "jet production" region, which is mainly described by the matrix element cross sections $\sigma_{n,i}$, calculated for each parton multiplicity n and for each different combination i of partons that contributes to multiplicity n. The cutoff on the final state partons is applied to avoid divergences. The second one is a "jet evolution" region, which is described by the parton shower calculations.

In the CKKW prescription, the shower corrections are set to zero above the scale, by vetoing any emissions above the cutoff, causing the matched result to be identical to the matrix element in that region, apart for higher-order corrections, needed to obtain a smooth transition to the phase space described by the PS. Below the matching scale, the pure shower is used to describe the evolution of the event.

The MLM prescription (the precise algorithm used in `MadGraph` and described in the following is called k_T-MLM), after the generation of the parton according to the ME, showers them without any constraint using PS algorithm. The parton collection that results from this step is clustered using a jet clustering, in particular a k_T algorithm; the resulting jets are matched to the ME partons if the jet measure k_T, $d_{parton,jet}$ written in Eq. 1.47, between the two objects is smaller than the cut-off.

$$\begin{aligned} d_{parton,jet} &= \min(k^2_{T,parton}, k^2_{T,jets}) \frac{\Delta^2_{parton,jet}}{R^2} \\ \Delta^2_{parton,jet} &= (y_{parton} - y_{jet})^2 + (\phi_{parton} - \phi_{jet})^2, \end{aligned} \tag{1.47}$$

Only those events in which all the jets match to the ME partons without any extra unmatched jets are retained (for the maximum ME parton multiplicity additional jets, softer than the matched ones, are allowed).

The MLM prescription is really convenient because it just requires a veto routine to kill events not fulfilling the matching criteria. On the other hand the user has to set up different parameters. The algorithm used for the matching has three parameters, namely the minimum jet p_T, the radius R, and the jet maximum pseudorapidity η.

Chapter 2

The Compact Muon Solenoid experiment at LHC

The Large Hadron Collider (LHC) [39] is the most powerful proton-proton collider ever built. Its installation was completed in summer 2008 and the operation started in September 2009. Since the end of March 2010 stable beams of protons collide at 7 TeV in the center of mass. The energy was increased to 8 TeV in 2012. Due to the high energy and luminosity it can provide, LHC is able to investigate processes produced with very small cross sections, down to the femtobarn.

It is possible to identify two main reasons that led to the choice of building an hadron collider instead of an electron collider: first of all, the energy in the center of mass achievable with a proton collider housed in the LEP tunnel is higher. The energy of electron collisions is limited by the synchrotron energy emission, a factor $O(10^{13})$ greater than the proton synchrotron emission. Moreover, due to the composite nature of protons, many low energy particles are generally produced, but occasionally a hard parton-parton scattering occurs allowing to explore parton collisions at a fraction of the proton-proton center of mass energy which can reach several tens of per cent. This is an important feature for an experiment involved in the discovery of unknown new physics.

In this chapter the CMS experiment is briefly described. Initially a introduction on the LHC accelerator is given, followed by the presentation of the different detectors composing CMS. The main features of the trigger are also listed.

2.1 The Large Hadron Collider

The LHC accelerator was built in the same tunnel that hosted the previous electron-positron collider, LEP, and two proton beams are accelerated in a 27 km circumference ring located on average 80 m deep, in part in Switzerland, in part in France.

In LHC the final acceleration is provided in several steps. The proton injection starts at the duo-plasmatron, which is the proton source. A linear accelerator (LINAC) boosts the protons to energy of 750 KeV using Radio Frequency Quadruples. Then in the Proton Synchrotron Booster (PSB) the energy of protons

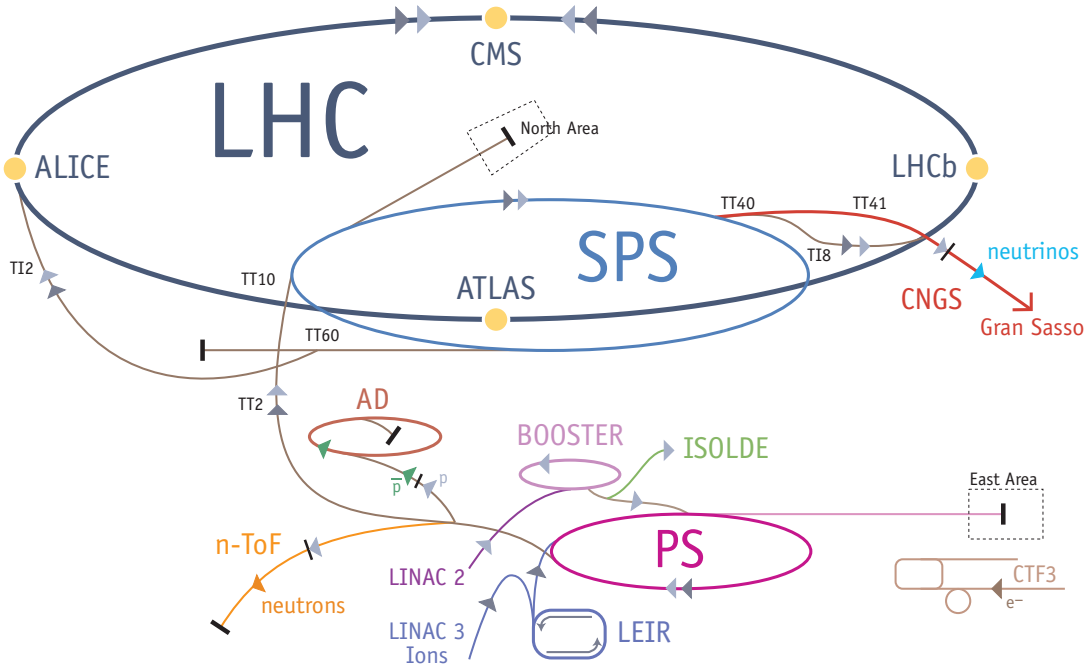


Figure 2.1: The LHC accelerator complex.

is increased up to 1.4 GeV. After that the LHC bunch train starts in the Proton Synchrotron (PS), in which the energy increases again up to 25 GeV. Then the protons are accelerated up to 450 GeV by the Super Proton Synchrotron (SPS), and finally they are injected into the LHC ring where the acceleration reaches the nominal proton energy of 3.5 TeV (the nominal proton energy during 2011, 4 TeV during 2012) through 8 Radio-Frequency (RF) cavities, using oscillating electric and magnetic fields. In order to keep the beam collimated, 7000 quadrupoles are used. In the Fig. 2.1 a schematic view of the different constituents of LHC is shown.

The two beams collide in four interaction points where their transverse dimension is reduced to 16 μm to maximize the luminosity and at which four experiments are located. Two general purpose experiments, ATLAS [40] and CMS [41], are dedicated to general Standard Model measurements and to the search for new physics; one experiment, LHCb [42], is dedicated to the b quark physics and to measurements of CP violation; one experiment, ALICE [43], is built to investigate heavy ion physics. LHC is able to investigate mass scales from the order of a few GeV, as in the case of B meson physics, up to a few TeV, for the discovery of new vector bosons or quark compositeness. In order to extend the LHC capability to explore rare processes an enormous effort has been made to increase the proton momentum as much as possible. In particular, a very sophisticated magnet system was designed in order to

keep such high momentum protons in the machine orbit. The formula that connects the bending radius with the charged particles momentum and the magnetic field is:

$$B[\mathrm{T}] = \frac{p[\mathrm{GeV}]}{0.3\rho[\mathrm{m}]} \tag{2.1}$$

where B is the magnetic field in Tesla, p the momentum in GeV and ρ the orbit radius in meters. The magnetic field needed to keep in orbit in LHC protons with an energy of 3.5 TeV is about 2.7 T, a value close to the current technological limits for superconducting magnets. The beams are bent and kept along the trajectory by 1232 superconductive dipoles and a sophisticated superfluid helium cooling system is used in order to maintain their temperature below 1.9 K. As we can see from Eq. 2.1 the beam energy is limited by the power of the magnetic field and by the circumference of the LHC. So the only way to increase the rate of rare and interesting events is to raise the luminosity $\mathcal{L}$. This variable is defined as:

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{2.2}$$

where n_1 and n_2 are the number of particles in beam 1 and 2 respectively, f is the collision frequency, and $\sigma_{x,y}$ are the transverse dimensions of the beam. The luminosity is related to the rate of events, n, for a given process with cross section σ by the following expression:

$$n = \mathcal{L}\sigma_{process}. \tag{2.3}$$

In Fig. 2.2 the production cross sections of many Standard Model processes are shown as a function of the center-of-mass energy. In the hard proton proton collision, the energy in the center of mass $\sqrt{\hat{s}}$ of the partonic level interaction is related to the total centre of mass energy, $\sqrt{s}$, by the following expression:

$$\hat{s} = \sqrt{x_1 x_2 s} \tag{2.4}$$

where x_1 and x_2 are the energy fractions of the two partons participating in the hard scattering (Bjorken factors [17]) distributed as the parton distribution functions, hence the center of mass of the two hardly interacting partons is not known a priori. As a consequence, the Lorentz boost invariant observables are very important to characterize the process. One of the most used quantities is the transverse momentum p_T, defined as the projection of the momentum vector on a plane perpendicular to the beam axis (Lorentz boost does not transform transverse coordinates of a four-vector when the boost is along the beam direction z). Another useful observable is the rapidity y defined as:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) = \tanh^{-1}\left(p_z/E\right) \tag{2.5}$$

where E is the energy of the particle, and p_z the projection of particle momentum along the beam direction. The differences of this variable are invariant for longitudinal boosts, since applying a Lorentz boost to y along z with speed β, the rapidity

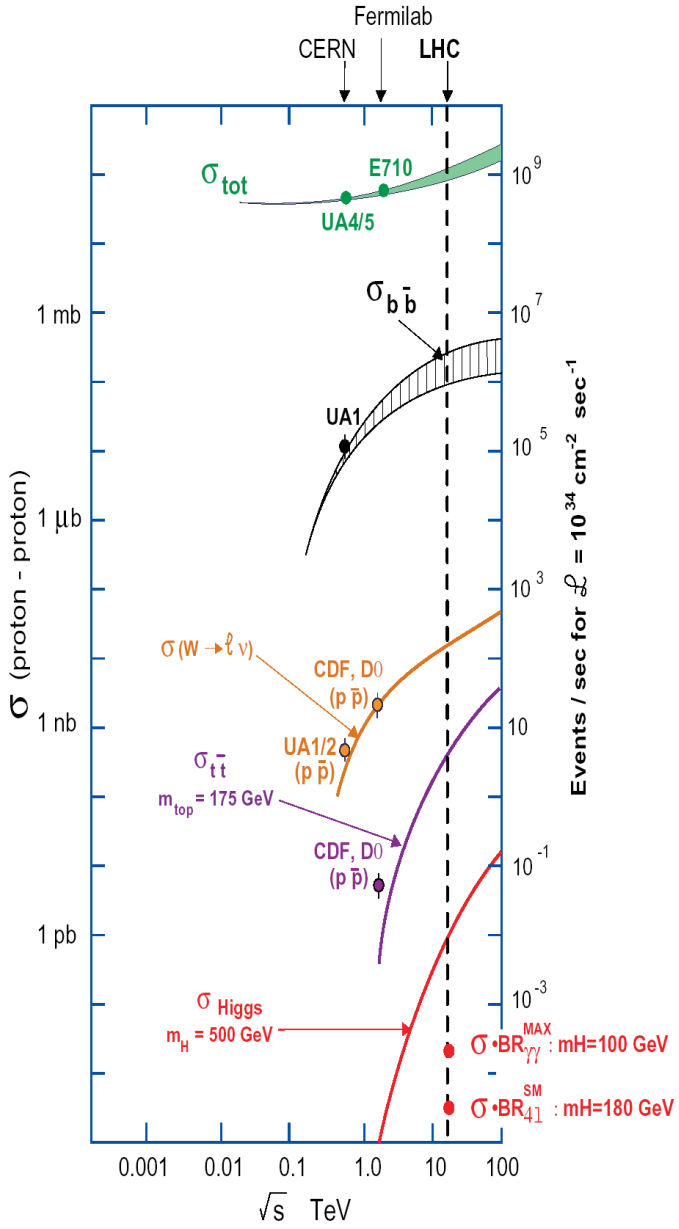


Figure 2.2: Expected cross section for several processes as a function of the center-of-mass energy of the pp collision.

y becomes $y = \tanh^{-1}(\beta)$.
In the ultra-relativistic approximation ($m \sim 0$ so $p_z \sim E$) the rapidity y is equal to the pseudorapidity η defined as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{2.6}$$

where θ is the polar angle between the direction of the momentum of the particle and the beam direction. When using the ultra-relativistic approximation, the pseudorapidity is often used to replace the rapidity because it depends only on the direction of the momentum of the particle.

The LHC schedule

The four LHC experiments registered the first pp collisions at 7 TeV in the center of mass at the end of 2009. In September 2010 the number of bunches per beam reached 368, corresponding to a luminosity of 2.1×10^{32} cm^{-2}s^{-1}. The final integrated luminosity delivered by the LHC in 2010 run was about 36 pb^{-1}.

The 2011 run started on March with three bunches per beam. The number of bunches increased up to 1380 and the instantaneous luminosity was of the order of 3×10^{33} cm^{-2}s^{-1} already in September. Thanks to the gradual increase of this property (up to 5×10^{33} in December 2011), CMS reached about 5 fb^{-1} at the end of 2011 (Fig. 2.3).

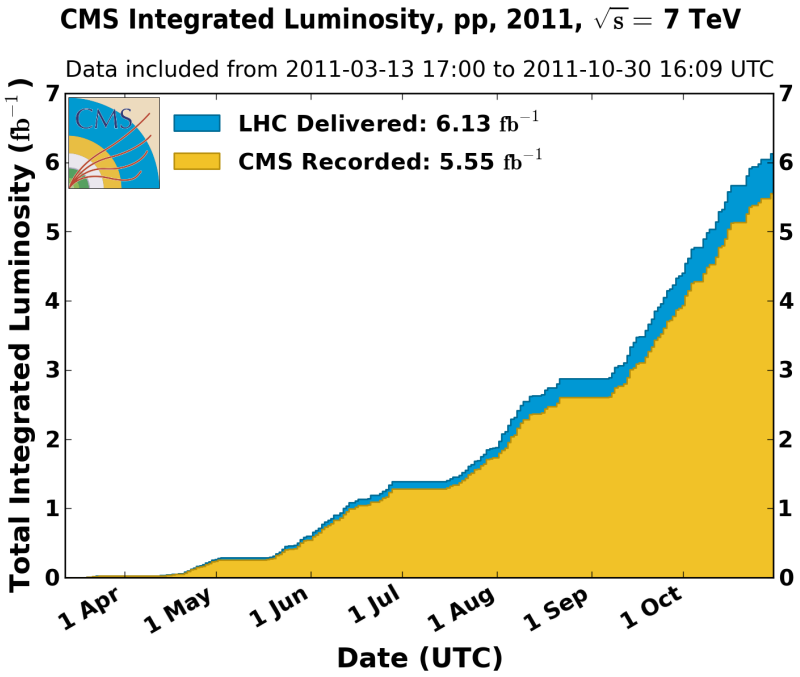


Figure 2.3: Total integrated luminosity recorded by the CMS experiment during the 2011 data taking.

.

The 2012 data taking started in March, registering the first collisions at 8 TeV in the center of mass. The number of bunches per beam was still the same as the previous year, allowing to exceed an instantaneous luminosity of about 7×10^{33} cm^{-2}s^{-1}. Due to the optimal performance of the system, at the end of 2012 the total integrated luminosity recorded by CMS was more than 20 fb^{-1}.

2.2 The CMS detector

The Compact Muon Solenoid (CMS) [41] is a general purpose experiment. It is characterized by the presence of a superconducting solenoidal magnet able to provide a magnetic field of 3.8 T, allowing a compact design of the detector. The Silicon Tracker, the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL) are in fact located inside its magnetic field. The Muon Chambers are located outside the solenoid, alternated by 1.5 m of iron saturated by the magnet return yoke. Its total dimensions are: 21.6 m of length, diameter of 15 m, and a total weight of 12500 tons.

The structure of CMS consists of several cylindrical detecting layers, coaxial with the beam direction called *barrel* region. The barrel layers are closed at both ends with disks, referred to as *endcap* regions. In Fig. 2.4 a schematic view of the CMS detector is presented. These characteristics determine the main features of the CMS experiment:

- good momentum resolution and reconstruction efficiency of charged particles;
- good muon identification, charged reconstruction and momentum resolution for $|\eta| < 2.5$;
- good electromagnetic energy resolution and measurement of electromagnetic shower properties for $|\eta| < 3$;
- good jet and missing transverse energy resolutions.

The CMS experiment uses a right-handed cartesian coordinate system. The x axis points towards the LHC centre, the y axis is directed vertically toward the surface and the z axis points along the beamline. Usually the cartesian coordinates are replaced by a pseudo-angular reference frame suggested by the cylindrical symmetry of CMS and by the invariant description of proton-proton collisions. The new set of reference variables is (r, ϕ, η), where r is the distance in the transverse plane, ϕ is the azimuthal angle, measured starting from the $+x$ axis, and η is the pseudorapidity.

2.2.1 The solenoid

The CMS magnet provides a magnetic field of 3.8 T and is the biggest superconducting solenoid in the world. It is composed by a superconductor cable surrounded by

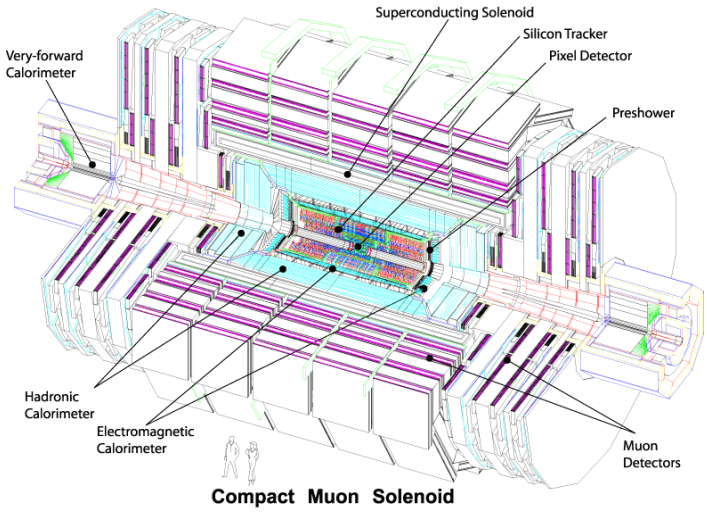


Figure 2.4: A view of the CMS detector. Each subdetector is labeled.

a layer of high purity aluminium that is used as a thermic stabilizer. All is welded by two layers of high-strength aluminium alloy for mechanical reinforcement. The total conductor cross section is 64×22 mm^2. It is kept at 4 K during operation by a liquid helium cooling system. The iron layers interspersed with the muon detectors are used to saturate the magnet return yoke. The residual magnetic field present in the yoke (around half of the field in the central region of the detector) is used to curve the muons in the muon detectors in order to facilitate the muon momentum measurement. The solenoid is the bearing structure of the whole apparatus and it is the main element in term of weight and dimensions, 5.9 m of diameter and 12.9 m of length. The energy stored in the magnet is about 2.7 GJ at full current. The energy can be dumped to resistors in only 200 ms in case of a quench, when a magnet loses its superconducting property. More information about the features of the CMS solenoid can be found in [44].

2.2.2 The tracker

The silicon tracker is placed close to the beam pipe and it covers the region $|\eta| < 2.4$, $r < 120$ cm. Its main goal is to provide a precise measurement of the momentum of the charged particles to allow a precise determination of the position of primary vertices, to distinguish superimposed events and secondary vertices. Due to the great multiplicity of products in the final state, the events are characterized by high complexity, hence the track reconstruction represents a complicated pattern recognition problem, solvable with two requirements. With a highly granular detector it is possible to keep low the detector occupancy; with a large number of detecting layers a large hit redundancy is provided. In order to do that the tracker consists of two main parts: a Silicon Pixel detector, and a Silicon Microstrip detector.

Due to the proximity to the beam pipe, pixel, micro-strips and readout electronic are subjected to a huge flux of radiation that can cause important damages. The pixel detector, which is exposed to the highest flux per unit area, will be replaced at least once during the LHC lifetime. In order to limit the effect of radiation damage on the sensor performances the tracker operates at low temperature (7°C).

The material budget[1] has to be limited since the electron energy loss due to bremsstrahlung and the presence of hadron nuclear interactions need to be kept as low as possible not to spoil the tracking performances and the calorimeter measurements. The tracker depth in terms of radiation length X/X_0 (X_0 is the distance over which a high energy electron reduces its energy to a fraction $1/e$ of the initial energy by bremsstrahlung emission) and in terms of interaction length λ/λ_0 (λ_0 is the mean free path of a hadron before having an interaction when traversing a material) as obtained from the full simulation of the tracker is shown in Fig. 2.5 as a function of η. The material budget is higher in the region $1< |\eta| <2$, the transition area between barrel and endcap.

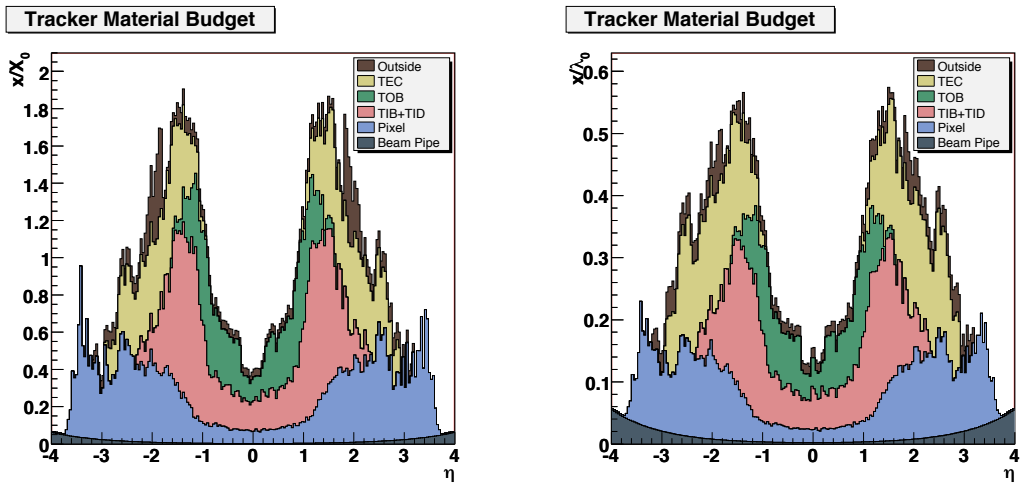


Figure 2.5: Radiation length (on the left) and interaction length (on the right) of the tracker as a function of η. Contributions from different components are put into evidence.

Another crucial element to obtain high precision spatial resolution is the alignment of the tracker modules. Deviations are caused by assembly inaccuracies, deformations due to cooling and stress from the magnetic field. The geometry was determined during assembly to an accuracy of 80 to 150 μm. An infrared laser system is used for continuous monitoring of the position of selected tracker modules. The final alignment is done with tracks from well known physics processes, e.g.

[1]The material budget expression indicates the amount of material needed to built the detector.

cosmic muons, or di-muons from Z decays.

The pixel vertex detector

The pixel detector [45] is the innermost part of the tracker. In Fig. 2.6 a schematic view of the CMS pixel detector is shown. It is made of three barrel layers (each one 53 cm long) positioned at $r = 4.4$, 7.3 and 10.2 cm, and two endcap wheels per each side, located at $|z| = 34.5$ cm and 46.5 cm to guarantee at least two crossed layers per track coming from the center of the detector within the fiducial acceptance $|\eta| < 2.5$.

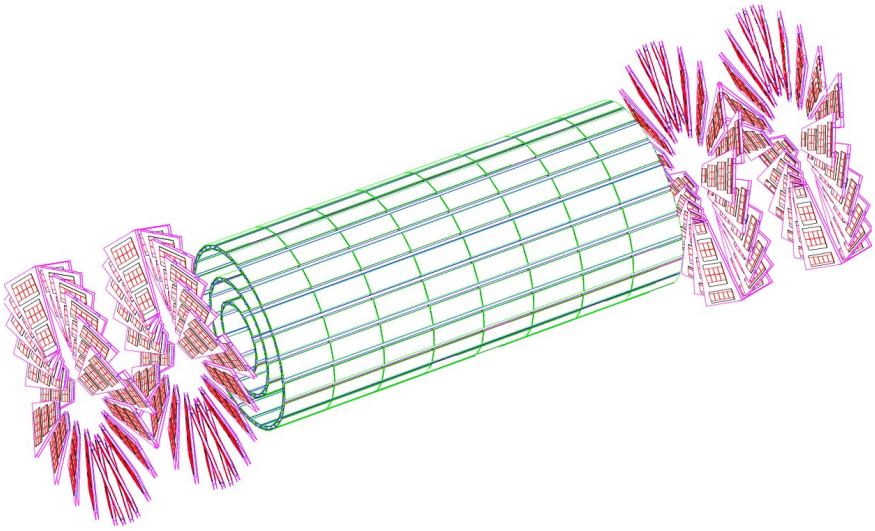


Figure 2.6: A schematic view of the CMS pixel detector.

The size of each pixel is 100×150 μm^2, with a total of about 66 million channels. The rectangular shape has been chosen in order to optimize the resolution of both coordinates, $r\phi$ and z. In particular the resolution is 10 μm for the $r\phi$ coordinate, while along the beam axis (z) is of 20 μm. These features are really important to provide a good measurement of the vertices.

The signals coming from each pixel are combined with analogical signal readout to profit of the charge sharing effect among pixels and improve the position resolution by interpolation. The charge sharing between pixels is enhanced by the Lorentz angle of the charge carriers, which is about 25° in the barrel and 4° in the endcaps for electrons in a 4 T magnetic field at 100 V of bias voltage, three times wider than for the holes. Therefore initial n-type substrate sensors were chosen to collect electron

signals on n^+ implants, which in turn are more radiation hard. In the barrel the pixels are tilted to induce significant charge sharing between neighbouring implants in the rϕ plane improving the intrinsic hit[2] resolution down to 10-15 μm, much lower than 150 μm, the width of each n^+ implant. In Fig. 2.7 the mechanism of the charge sharing induced by Lorentz drift is shown.

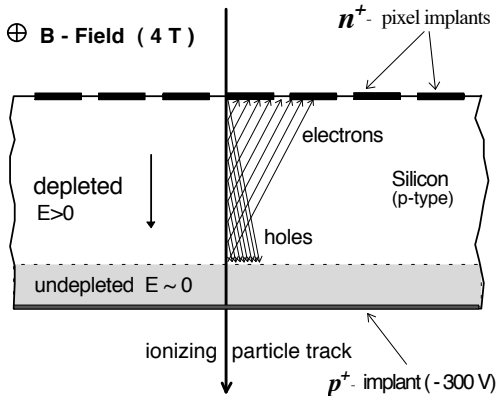


Figure 2.7: Charge sharing induced by Lorentz drift.

Charge sharing is present also along z direction for inclined tracks leading to a similar resolution. The detectors placed on the disks are rotated with an angle of 20° around the central radial axis to benefit of charge sharing improved both in r and $r\phi$ directions by induced Lorentz effects.

The silicon micro-strips detector

The outer region of the tracker, where the flux of particles has decreased enough to allow the use of this type of sensors, consists of the silicon double-layer micro-strip detector [46].

The detector is divided in four parts, as shown in Fig. 2.8, TIB (Tracker Inner Barrel), TID (Tracker Inner Disks), TOB (Tracker Outer Barrel) and TEC (Tracker EndCap) covering a tracking volume up to $r = 1.1$ m with a length of 5.4 m.

The detector unit is made of one or two sensors glued on a carbon fibre mechanical support together with the read out electronics, in the inner part the minimum size is 10 cm×80 μm, while in the outer one is 25 cm×180 μm. The sensor is a n-type phosphorus doped substrate with p^+ implant strips. Since the mean energy required to create an electron-hole pair in silicon is 3.6 eV, a minimum ionizing particle (*mip*) passing through a 300 μm thick sensor with an average energy loss per path length of 390 eV/μm should create about 32500 electron-hole pairs. The analogical signal coming from each strip is transmitted to ADCs located in the counting room via optical links. In order to decouple the readout electronics from the detector leakage

[2]The *hit* definition is not trivial, but at this level it can be defined as the trajectory point intersecting the detector layer plane.

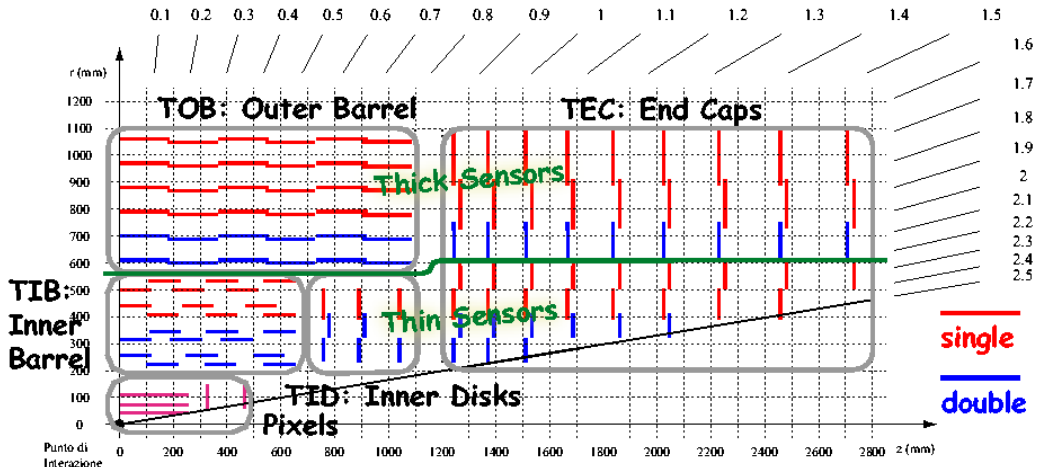


Figure 2.8: A schematic view $r - z$ of the CMS tracker subdetector.

current, insulating capacitor layers of dielectrics (SiO_2, Si_3N_4) are placed between the p^+ and the aluminium strips electrodes.

The expected resolution for the CMS tracker, for three different parameter of the tracks as function of pseudorapidity for muons with transverse momentum of 1, 10, 100 GeV respectively, is presented in Fig. 2.9.

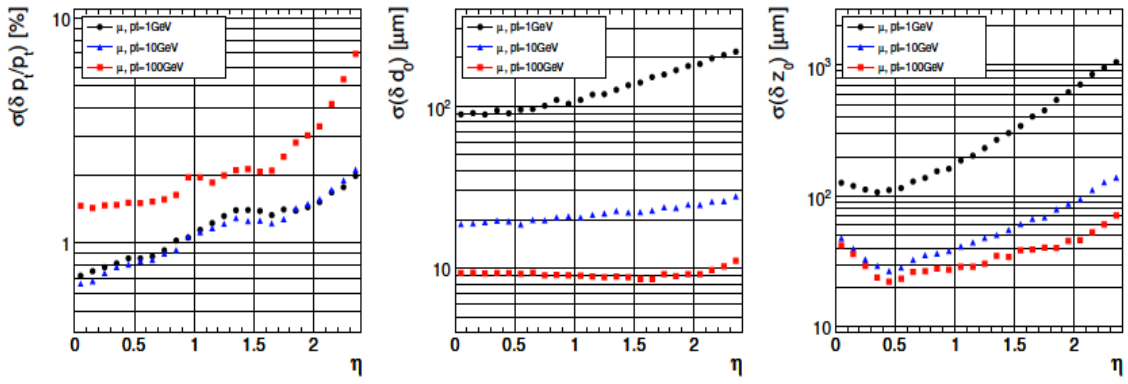


Figure 2.9: Resolution for three different track parameters for muons with p_T =1/10/100 GeV. Muons are reconstructed using only the tracker information: transverse momentum (left), transverse impact parameter (center), longitudinal impact parameter (right).

2.2.3 The electromagnetic calorimeter

The main goal of the Electromagnetic Calorimeter (ECAL) [47] is the precise measurement of energy of electrons and photons. The architecture of ECAL was chosen

taking into account the requests imposed by the H$\rightarrow \gamma\gamma$ analysis, which is performed looking for an excess in the di-photon invariant mass distribution. Since the experimental width of this decay is expected to be dominated by the resolution of the calorimeter, and to be of the order of 1%, high granularity is needed to improve the measurement of the angle between the two photons and to obtain a good $\pi^0 \rightarrow \gamma\gamma$ separation.

ECAL is an homogeneous, high granularity calorimeter divided into a barrel and two endcaps containing almost 76000 Lead Tungstate ($PbWO_4$) scintillating crystals. The $PbWO_4$ properties have led to its choice. The $PbWO_4$ crystals have a low light yield ($\sim$10 photo-electrons/MeV), with requires the use of high intrinsic gain detectors able to operate inside a magnetic field. In addition, its short radiation length ($X_0 = 0.89$ cm), due to its high density (8.28 g/cm^3), and low Moliere Radius ($R_M = 2.19$ cm) allow to contain the electromagnetic shower in a limited region, in order to build a compact calorimeter. Another important aspect is its fast response ($\tau = 10$ ns with about 80% of the light collected within 25 ns), that is a crucial issue in the high LHC rate. Since the $PbWO_4$ has a good intrinsic radiation hardness, it can work in the LHC environment. In Fig. 2.10 a schematic view of the ECAL detector is shown.

The barrel (EB) covers the $|\eta| < 1.479$ region and has an inner radius of 129 cm. It consists of 36 identical supermodules built as a matrix of 20 crystals in ϕ and 85 crystals in η, covering an azimuthal angle of 20°. The supermodules are divided along η in 4 modules, with submodules consisting of 5×2 crystals as the basic unit of ECAL. Each crystal has a tapered shape, with 2.2 cm × 2.2 cm front face and 23 cm length corresponding to 25.8 X_0. The barrel granularity is $\Delta\phi \times \Delta\eta = 0.0175 \times 0.0175$, the crystals are grouped into 5×5 arrays called trigger towers, providing information to the trigger system. To avoid that cracks might align with the particles trajectories, the crystal axes are tilted with respect to the direction from the interaction point, both in ϕ and in η. Avalanche photodiodes (APDs) are used to collect the produced light. Each APD has an active area of 5×5 mm^2; a pair is mounted on each crystal. They are operated at gain 50 and read out in parallel.

Each endcap (EE) consists of two semicircular plates called Dees and covers the $|\eta|$ region between 1.48 and 3. The crystals have a 2.86×2.86 cm^2 front face, are 22 cm long, corresponding to 24.7 X_0, and they are grouped in structures of 5×5 crystals called super-crystals. The granularity varies from $\Delta\phi \times \Delta\eta = 0.0175 \times 0.0175$ to 0.05×0.05. In the endcaps the photodetectors used are vacuum phototriodes (VPTs). Each one is 25 mm in diameter, with an active area of about 280 mm^2; one VPT is glued to the back of each crystal. They have a single gain stage, with a value of 10.2 without magnetic field, a 4 T magnetic field lowers this value by less than 10%.

On the front of each endcap, a preshower (ES) is installed between $|\eta|$=1.6 and $|\eta|$=2.6 in order to improve the π_0/γ separation and the vertex identification. It works at a temperature of 5°C. It consist of two lead radiators (the first 2 X_0 and the second 1 X_0 thick) followed by silicon strip sensors positioned with orthogonal orientation with a pitch of less than 2 mm. Each sensor measures 63 × 63 mm^2,

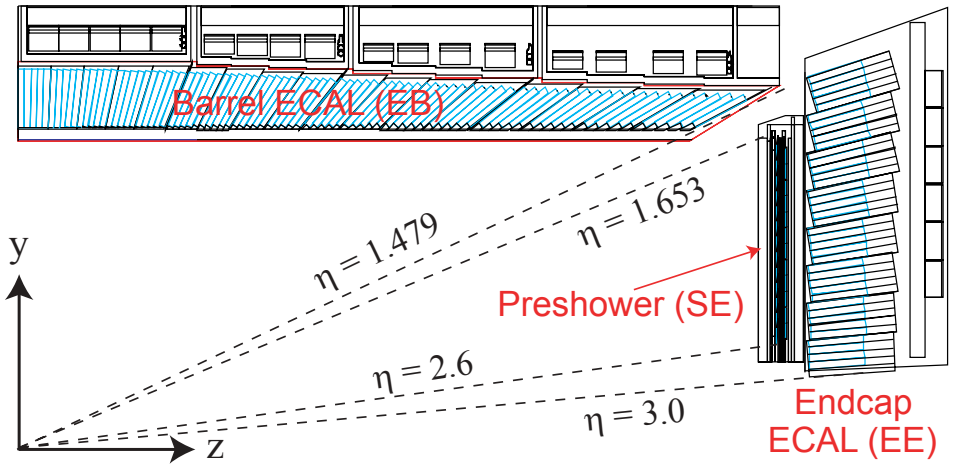


Figure 2.10: A schematic view of the CMS electromagnetic calorimeter (ECAL).

with an active area of 61 × 61 mm^2, divided into 32 strips. The nominal thickness of the silicon is 320 μm. Important requirements are imposed on the front-end (FE) electronics since it has to be fast enough to sustain the 25 ns LHC crossing rate and it has to be radiation hard.

2.2.4 The hadronic calorimeter

The CMS hadron calorimeter (HCAL) [48] works together with ECAL in order to measure the energy and direction of charged and neutral hadrons, and the energy imbalance in the transverse plane E_T^{miss}. It has been structured in order to contain hadronic showers, much deeper and wider that the electromagnetic ones, i.e. it must be as hermetic as possible. The central barrel and endcap HCAL subdetectors are placed in the high magnetic field of the solenoid and this is another element that influenced its design. It provides a good segmentation, a quite fine energy resolution and it covers the $|\eta| < 5$ region.

The HCAL is made of four subdetectors (see fig. 2.11):

- Barrel Hadronic calorimeter (HB): it is placed inside the magnetic coil and it covers the central pseudorapidity region, up to $|\eta| = 1.3$. The barrel hadron calorimeter consists of two half barrels, each composed of 18 identical 20° wedges in ϕ. Each wedge is composed of 5 cm thick brass stainless steel absorber plates parallel to the beam axis and interleaved with 3.7 mm thick plastic scintillators. The signal is readout through wavelength-shift fibres and hybrid photodiodes (HPD). The granularity is of the order of $\Delta\phi \times \Delta\eta$=0.087×0.087. The HB has an energy resolution for single pions of approximately $120\%/\sqrt{E}$. The minimum depth is about 5.8 λ.

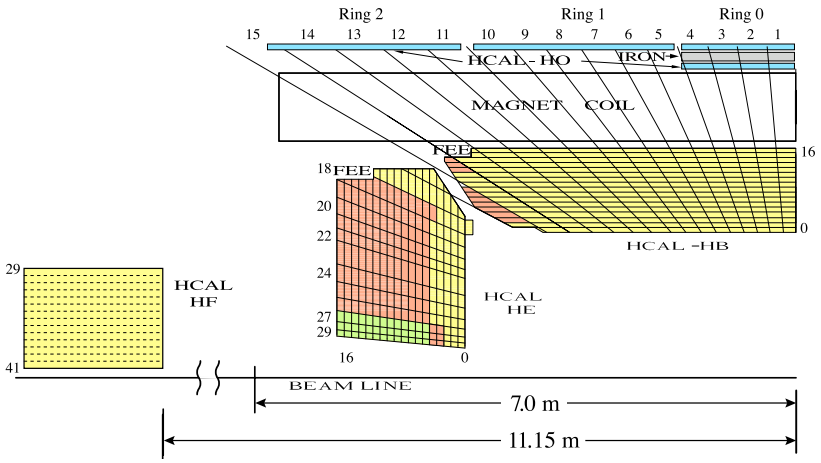


Figure 2.11: A schematic view of the CMS hadron calorimeter (HCAL).

- The Endcap Hadronic calorimeter (HE) is located as well inside the magnetic coil and it is made of two endcaps extending the angular coverage up to $|\eta|$ = 3. The HE has the same architecture of HB for the same granularity. The signal is read through wavelength-shift fibres and hybrid photodiodes. The minimum depth is of 10 λ.

- The Outer Hadronic calorimeter (HO) is placed in the barrel region, added outside the magnetic coil in order to extend the depth of the calorimeter in terms of nuclear interaction length λ. The total depth for $|\eta| < 1.26$ is thus extended to about 11.8 λ. It consist of two scintillator layers with the same granularity than the ones of HB.

- The Forward Hadronic calorimeter (HF) consists of two units placed outside the magnetic coil, at $\pm$11.2 m from the interaction point along the beam direction that receive the most part of the energy produced in the interactions, represented by the particles at small angle. It extends the pseudorapidity coverage up to $|\eta| = 5$. The material of the two units, steel absorbers and embedded radiation hard quartz fibers, provide a fast collection of Cherenkov light. The granularity of HF is $\Delta\phi \times \Delta\eta = 0.17 \times 0.1745$.

In order to obtain a reference calibration and to measure the characteristics of the different parts of HCAL, these were exposed to beams of electrons, pions, protons and muons. An ECAL module was also included in the test beam setup [49, 50, 51, 52].

2.2.5 The muon system

The CMS muon system [53] is dedicated to the identification and the measurement of high p_T muons, in combination with the tracker. The system is placed outside the magnetic coil, embedded in the return yoke, in order to exploit the returning flux of about 1.8 T. Given the shape of the solenoid, the muon system is divided into a cylindrical barrel section and two planar endcap regions. The first one is easier to operate, having less background and low muon rate, and indeed the choice of detectors was influenced by these features.

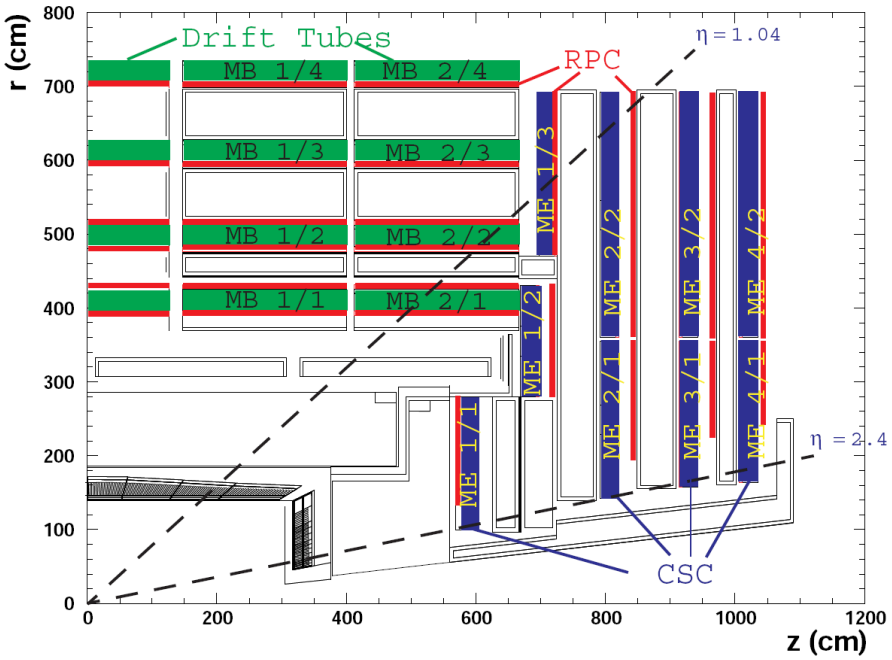


Figure 2.12: A quadrant of the CMS muon system.

The system consists of three independent subsystems (Fig. 2.12):

- Drift Tubes (DT) are placed in the barrel region, covering $|\eta| < 1.2$, where the hits occupancy is relatively low (< 10 Hz/cm^2).
- Cathode Strip Chambers (CSC) are located in the endcap regions, $0.9 < |\eta| < 2.4$, with higher occupancy (> 100 Hz/cm^2) and large and non-uniform magnetic field.
- Resistive Plate Chambers (RPC) are placed both in the barrel and in the endcaps, covering the region $|\eta| < 1.6$.

The Drift Tube system is made of 250 chambers consisting of twelve layers of drift tubes. Each group of layers is packed in three independent super-layers, for a

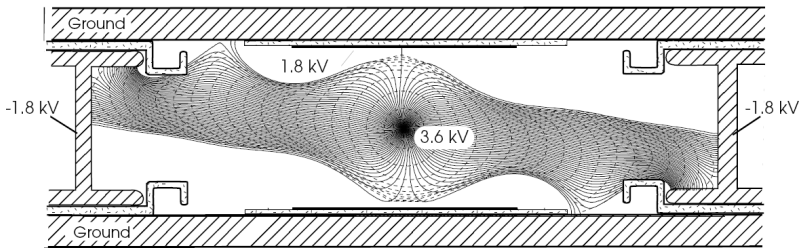


Figure 2.13: A schematic representation of a drift tube chamber. The drift lines in presence of the magnetic field are shown.

total of four chambers with three super-layers per chamber. The structure is divided along the beam line in 5 rings, each of them composed by 12 sectors covering an azimuthal angle of 30°. In each chamber two super-layers have anode wires parallel to the beam axis, and one has perpendicular wires. Thus, each chamber can provide two measurements of the $r-\phi$ coordinate and one measurement of the z coordinate of the track hit positions. The maximum drift length is 2 cm and the position resolution is about 100 μm in both $r-\phi$ and $r-z$. In Fig. 2.13 a representation of a drift tube chamber is shown. Each chamber is filled with a gas mixture of Ar(85%) and CO_2(15%).

The Fig. 2.14 shows the structure of the Cathode Strip Chambers, multi-wire proportional chambers with segmented cathodes. There are 468 trapezoidal chambers filled with a gas mixture of Ar(40%), CO_2(50%), CF_4(10%), each of them providing both hit position coordinates. The chambers are placed perpendicular to the beam line and provide a precision measurement in the $r-\phi$ bending plane, the spatial resolution is about 80-85 μm.

In order to improve the efficiency, a complementary system was added in both barrel and endcap regions. A schematic view of the Resistive Plate Chambers is shown in Fig. 2.15. The chambers are made of parallel bakelite planes, with a bulk resistivity of $10^{10} \div 10^{11}$ Ωcm and filled with a mixture of $C_2H_2F_4$ (96.2%), iso-C_4H_{10} (3.5%) and SF_6 plus water vapour (0.3%). They operate in avalanche mode to ensure good operation at high rate. The chambers have limited spatial resolution, but since they have excellent timing performances they can provide an independent trigger system and help to reduce ambiguities in the track reconstruction.

2.3 The trigger system

The proton bunches at LHC cross at a rate of about 40 MHz, but it is impossible to register all of them due to the limited disk space and the limited speed of the readout electronics. Beside that, only a small fraction of these collisions represents

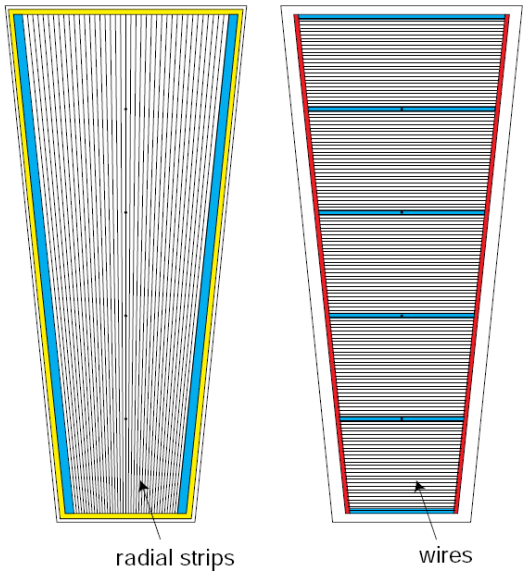


Figure 2.14: A schematic representation of the CSC cathode panel (on the left) and anode panel (on the right).

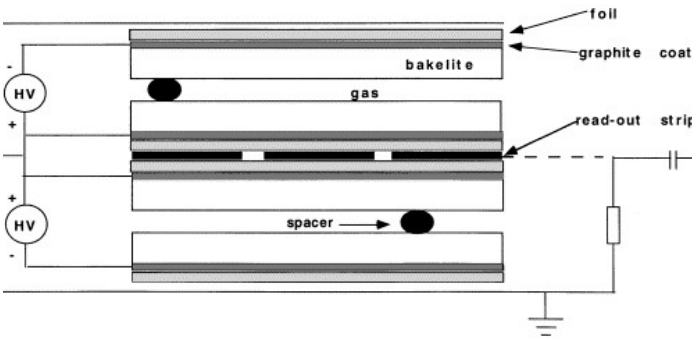


Figure 2.15: Schematic view of the RPC double-gap structure. The read-out strips in the Barrel chambers run along the beam direction.

interesting events worth to be written. Hence a trigger system is needed in order to reject by a factor 10^7 the collisions and to select in a short time the interesting physics events with high efficiency.

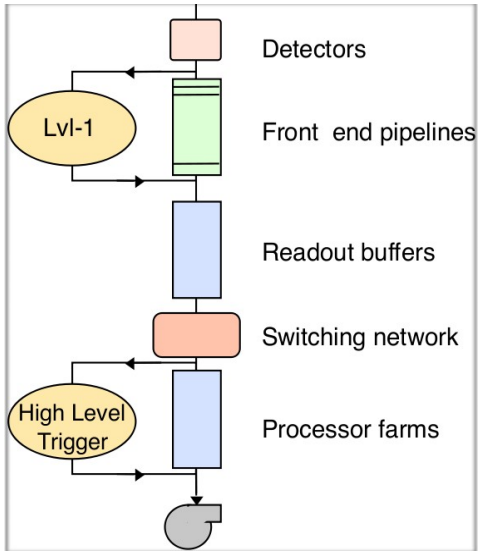


Figure 2.16: A schematic representation of the CMS trigger system.

CMS has chosen a two-level trigger system, composed by a Level-1 Trigger (L1) [54] and a High Level Trigger (HLT) [55]. In Fig. 2.16 a schematic view of the CMS trigger system is presented. The L1 trigger runs on dedicated processors and accesses coarse level granularity information from the calorimeter and muon system. Its task is to reduce the flux of data from 40 MHz to 100 kHz, taking decision for each bunch crossing within 3.2 μs.

The HLT must reduce the L1 output rate down to the nominal rate of 100 Hz. During the 2011 data taking the output rate of the HLT was around 300 Hz on average. It runs on a farm of commercial processors and can access the full granularity information of all the subdetectors.

2.3.1 Level-1 Trigger

The L1 trigger is responsible for the identification of electrons, muons, photons, jets and missing transverse energy. It has to have a high and carefully understood efficiency. Its output rate and speed are limited by the readout electronics and by the performances of the Data Acquisition (DAQ) system. The organization of the CMS L1 Trigger is summarized in Fig. 2.17.

It consists of three main subsystems:

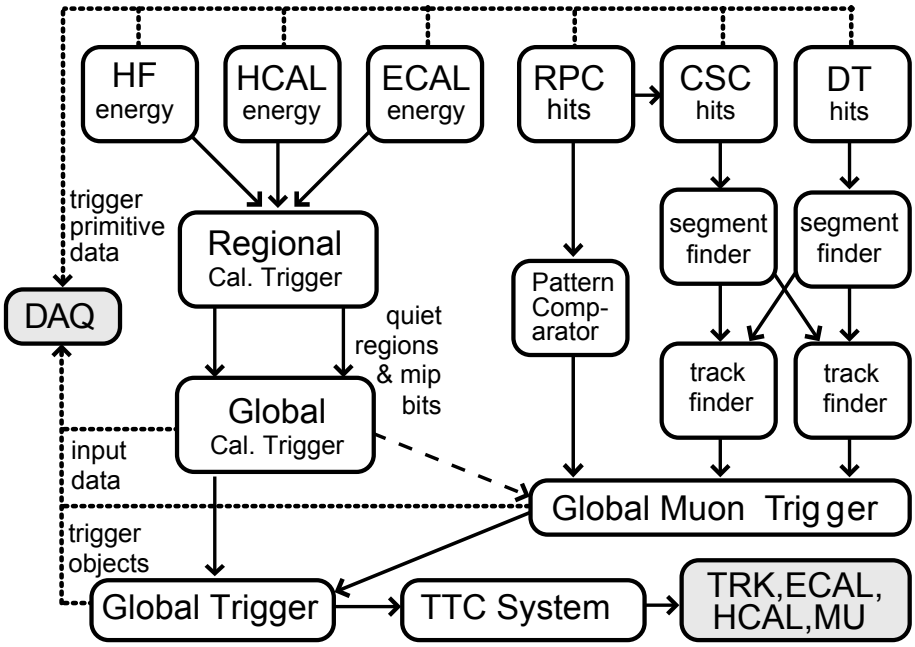


Figure 2.17: Overview of the L1 trigger system.

- **L1 Calorimeter Trigger**. It receives information from both the calorimeters, in the form of "towers", 5×5 ECAL crystals and a single HCAL segment. The Regional Calorimeter Trigger (RCT) identifies electron, photon, τ and jet candidates along with their transverse energy. The Global Calorimeter Trigger (GCT) sort the candidates according to their transverse energy and sends the first four to the L1 Global Trigger.

- **L1 Muon Trigger**. It is composed of three subsystems: the RPC, CSC and DT triggers. Each of them builds track segments, assigning them a p_T estimation and a quality flag. All the information is sent to the Global Muon Trigger (GMT), that sorts and tries to combine the muon tracks. The final set of muons is sorted according to the quality and the best four tracks are passed to the L1 Global Trigger.

- **L1 Global Trigger**. It is responsible for collecting the information from the calorimeter and the muons systems and for making a decision whether to save the event or not. All the ranked trigger objects produced by the calorimetry and muon system are sorted and the Global Trigger, on the base of their properties, determines to either accept or reject the event. The decision is transmitted to the Timing Trigger and Control system (TTC), which commands the readout of the remaining subsystems in case of positive response.

2.3.2 High Level Trigger

The main goal of the HLT is to reduce the L1 output rate up to $\mathcal{O}(100 - 1000)$ Hz. Only the HLT selected events will be definitively written to mass storage. The information from all the subdetectors is assembled by a builder unit and then assigned to a switching network that dispatches events to a commercial processor farm, where the HLT performs the reconstruction.

This simple design ensures maximum flexibility to the system with the limitation of the total bandwidth and the number of processors, that can be easily upgraded adding new machines or replacing the existing ones with newer and faster ones. Since the HLT algorithm implementation is fully software, its improvements do not require any hardware substitution. Each event is processed on a single machine and the average HLT decision must be made in 100 ms to avoid event dropping. Since the online nature of this selection, the algorithm can use a limited amount of resources, but its reliability is of extreme importance: the events not saved are lost.

In order to process events efficiently, the HLT code has to reject not interesting collisions immediately; computationally expensive algorithms must be ran only on good candidate events. For this reason the HLT is organized in three virtual levels, making use of the partial reconstruction idea. In a first phase, only the objects positioned in specific regions of the detector, very sensitive to the data selection, are reconstructed. In particular, in the Level 2 (L2) only the muons and the calorimetry information is used, in the Level 2.5 (L2.5) the pixel data are added and in the Level 3 (L3), considering the information from all the tracking detector, the event is completely reconstructed.

Each step reduces the number of events to be processed in the following one. The most computationally expensive tasks are executed at L3; time consuming algorithm such as track reconstruction are only executed in the regions of the detector where good candidates from L1, L2 and L2.5 are found. Besides that, since a high precision is not required at HLT, the track reconstruction is performed on a limited set of hits and is stopped when the desired resolution is achieved.

Chapter 3

Particle reconstruction in CMS

In this chapter the particle reconstruction used in CMS is described. In the first sections the algorithms used for the event reconstruction are presented and the "Particle Flow" reconstruction technique is introduced.

The focus is then put on the main ingredients for the reconstruction of the associated production of jets and the Z boson decaying in the electron channel, i.e. the jets reconstruction together with the jet energy corrections and the electron selection variables used to identify leptons originated by the Z boson decay.

3.1 The Particle Flow based reconstruction

The "Particle Flow" (PF) [56, 57] is an algorithm for a global event description using the combined information provided by all CMS sub-detectors, like calorimeter clusters and track segments, for an optimal determination of the direction, energy and type of all the stable particles in the event, like electrons, muons, photons, charged and neutral hadrons (see Fig. 3.1). The single subdetector giving one of the most important contributions to the PF algorithm is the tracker, since it allows a good reconstruction of charged-particle tracks, with large efficiency and small fake rate down to a transverse momentum p_T of 150 MeV/c, for a pseudo-rapidity range of $|\eta| < 2.4$.

The set of all the identified particles is then used for many different purposes, i.e. to build the jets (from which the quark and gluon energies and directions are determined), to quantify the charged lepton isolation with respect to other particles, to determine the missing transverse energy E_T^{miss} estimating the direction and energy of the neutrinos and other non-interacting particles, in order to obtain a consistent reconstruction and description of the whole event.

3.1.1 Iterative tracking and calorimeter clustering

It is particularly important to be able to reconstruct all the particles present in the final state, even those with small p_T, because the energy of the non-identified objects could be counted as missing energy. Indeed, even in proton-proton collisions at LHC

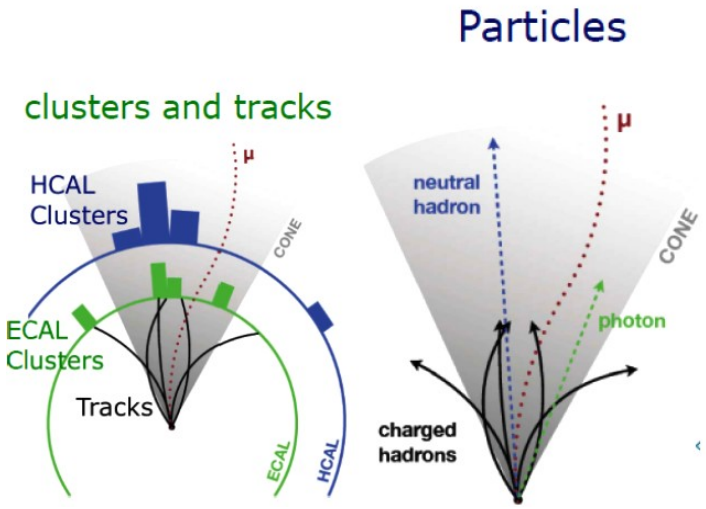


Figure 3.1: The PF algorithm starts from basic objects as tracks and calorimeter clusters (left), then all the information from the different subdetectors are processed in order to reconstruct the particles produced in the event (right).

with large energy in the center-of-mass, a sizeable fraction of the particles has low transverse momentum. For this reason, each element used by the PF algorithm, like calorimeter clusters or tracks, has to be reconstructed with high efficiency and low fake rate in the high-density events obtained in LHC. Advanced tracking and clustering algorithms were developed as consequences of these requests and they are described in the following.

The iterative tracking

The main purpose of the tracking algorithm is to provide a precise measurement of the direction of the charged particles making use of the information obtained by the tracker detector.

The standard track reconstruction is organized in five steps:

local reconstruction: the signals from the tracker are clustered into hits;

seed generation: initial track candidates (seeds) are created using three constraints which can be obtained from pixel, strip, vertex or beam spot measurements; they define the initial parameters for the next step;

pattern recognition: starting from the seeds, a global Kalman filter is used to find the track candidates corresponding to charged particles of interest;

final track fit: the final precise parameters of the tracks are estimated;

track selection: only the tracks passing a minimum set of quality cuts are considered.

Indeed, this procedure is essential for the PF event reconstruction, considering that about two thirds of the energy of a jet is carried by charged particles. Moreover, the momentum of charged hadrons is measured in the tracker with a higher resolution than that of the calorimeters for p_T up to several hundreds of GeV/c and each charged hadron missed by the tracking algorithm would be only detected by the calorimeters, with reduced efficiency, degraded energy resolution and biased direction.

The tracking efficiency must be as close to 100% as possible and the tracking fake rate must be kept small. For this purpose an iterative-tracking algorithm [58] has been developed.

- The first step consists in the seeding and reconstruction of the tracks applying very tight identification criteria (the p_T, the number of hits and the χ^2 of the tracks are taken into account). In this way, even if the obtained tracking efficiency is moderate, a low fake rate is kept (under 1%).

- In the second step the hits unambiguously assigned to the tracks found in the previous iteration are removed from the list of hits.

- Then the track seeding criteria are loosen in order to increase the tracking efficiency. The algorithm is run again but in this step, due to the hit removal, the combinatorial results are reduced and the fake rate is kept low.

- In order to reconstruct secondary charged particles, as particles originating from photon conversion, in the fourth and fifth iterations the constraints on the origin vertex are relaxed.

With this iterative technique, tracks of charged hadrons in jets originating from within a thin cylinder around the beam axis (that includes the beam pipe) are found with an efficiency larger than 90%; in the end the fake rate is at the per-cent level even for charged particles with only three hits and a p_T of about 150 MeV/c.

The calorimeter clustering

The goal of the calorimeter clustering algorithm is the detection and measurement of energy and direction of stable neutral particles, the separation of neutral particles from charged hadrons, the reconstruction and identification of electrons and all accompanying bremsstrahlung photons and the improvement of the energy resolution of high p_T charged hadrons, for which the track parameters are less precisely measured and for which the calorimeter energy resolution is more accurate.

In order to satisfy these characteristics, a dedicated clustering algorithm running for each sub-detector (ECAL barrel, ECAL endcap, HCAL barrel, HCAL endcap, ES first layer and ES second layer) has been developed for the PF reconstruction.

- The calorimeter cells are identified as “cluster seeds” if the energy is above a given threshold.

- The "topological clusters" are built starting from the seeds by aggregating cells with at least one side in common with a cell in the cluster, and with an energy above a given threshold. These thresholds represent two standard deviations of the electronics noise in the ECAL (80 MeV in the barrel and up to 300 MeV in the endcaps) and are about 800 MeV in the HCAL, producing the so-called PF seeds.

- The final energy of the topological cluster is calculated giving to each cell a weight according to its distance from the seed. In this way, if the cluster contains more than one seed because of the passage of two close particles, the possible overlap between them can be taken into account.

3.1.2 The link algorithm

Any PF object, corresponding to a given particle, is composed by several basic elements representing the responses of the different subdetectors, like a track, calorimeter clusters and muons tracks, as shown in Fig. 3.2. These objects have to be connected to each other by a link algorithm in order to reconstruct each single particle, avoiding any possible double counting between the different subdetectors.

The final result of the PF algorithm can be visualized as a chain, where the unitary component is given by a pair of elements linked to each other. There are different kind of components, depending on the type of objects taken into account (as track-ECAL cluster, ECAL-HCAL clusters). A single element can be connected to more than one, creating a so-called block, and in this case the quality of each pair is represented by the relative distance between the two linked elements. Thanks to the granularity of the CMS detector, these structures contain only few elements, allowing to keep procedure independent from the number of particles in the event. These blocks are the inputs for the particle reconstruction and identification algorithm.

In detail, a link between a charged particle track and a calorimeter cluster is performed with the following procedure:

- The track is extrapolated from the last measured hit in the tracker to ECAL and, if in the endcaps, to the two layers of ES, at a depth corresponding to the expected maximum of a typical longitudinal electron shower profile;

- The track is extrapolated to HCAL, at a depth corresponding to one interaction length, typical of a hadron shower;

- The track is linked to any cluster if the extrapolated position in the corresponding calorimeter is within the cluster boundaries. In order to consider effects of the presence of gaps between calorimeter cells, or calorimeter modules, uncertainties of the position of the shower maximum and multiple scattering in case of low momentum charged particles, the cluster can be extended up to the size of a cell in all directions. The distance in the (η, ϕ) plane [1] between

[1] The distance in the (η, ϕ) plane is defined as $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$

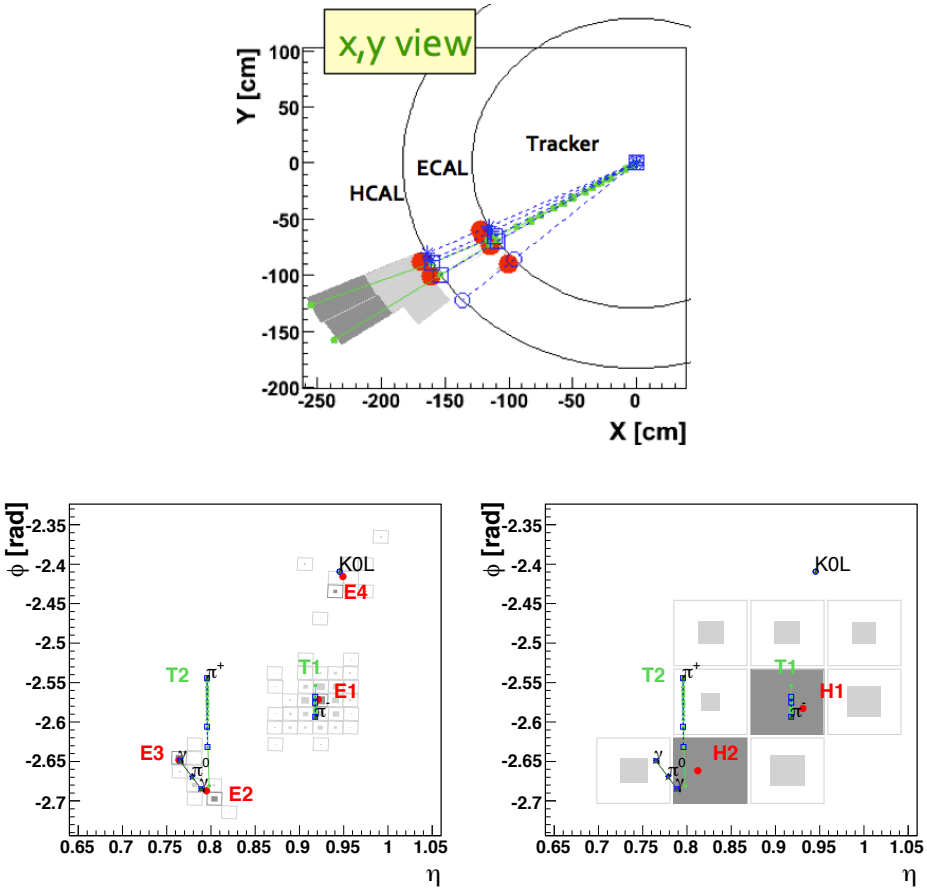


Figure 3.2: An event display of a simple hadronic jet in the transverse plane (top) and in the (η, ϕ) view, both on the ECAL surface (bottom left) and the HCAL surface (bottom right). The K_L^0, the π^- and the two photons from the π_0 decay are detected as four well separated ECAL clusters while the π^+ leaves no energy in this detector. The two charged pions are reconstructed as charged-particle tracks, appearing as vertical solid lines in the (η, ϕ) views. These tracks point towards two HCAL clusters. The cluster positions are represented by dots, the simulated particles by dashed lines, and the position of their impact on the calorimeter surfaces by various open markers.

the extrapolated track position and the cluster position defines the quality of the link.

An important feature of the PF algorithm is the recovery of bremsstrahlung photons emitted by electrons. For each track, starting from the intersection points with the tracker layers, tangents are extrapolated to ECAL. If clusters are found by the intersection of a track tangent with the calorimeter, they are linked to the starting track as originated by bremsstrahlung photons. The link quality corresponds to the distance in (η, ϕ) between the cluster position and the extrapolated point of the tangent in ECAL. All the clusters already directly associated to charged particle tracks are vetoed for this procedure, in order to avoid possible double counting.

Two calorimeter clusters, either an HCAL and an ECAL cluster or an ECAL and a ES cluster, are linked when the cluster position in the more granular calorimeter (ES or ECAL) is within the cluster boundaries in the less granular calorimeter (ECAL or HCAL). As for the track-cluster link, the cluster dimension can be slightly enlarged.

A link between a charged-particle track in the tracker and a muon track in the muon system is defined when a global fit between the two tracks returns a χ^2 (the link quality) less than a given threshold. This link gives rise to a global muon and if there are more than a charged-particle track associated to a muon one, only the global muon with the smallest χ^2 is retained.

3.1.3 Particle reconstruction and identification

Starting from the set of blocks derived from the linking procedure, the reconstruction algorithms extract a set of particles that provide a global description of the events to be used in the analysis. The different objects are identified in a specific order as described in the following:

- **muons:** Starting from a global muon[2], a "particle-flow muon" is reconstructed if the combined momentum is within three standard deviations of the one determined only from tracker. The corresponding track is removed from the block. The contribution of cluster energy information is negligible.

- **electrons:** A pre-identification procedure is first applied to each track exploiting the tracker as a preshower. The electron tracks are usually short and suffer of energy losses due to the bremsstrahlung photon emitted in the tracker material. The tracks having these characteristics undergo another fit procedure with a Gaussian-Sum Filter (GSF) [59]. This is a suitable filter to reconstruct tracks with kinks, produced by bremsstrahlung photons, in order to extrapolate with higher precision the trajectory in ECAL, obtaining the

[2]In CMS the track associated to a muon can be obtained in three different ways: the tracker fit uses only the information from the pixel and the strip detectors, the standalone fit is provided by the sole muon system, and the global fit is obtained by combining all the information. A global muon derives from a global fit track.

so-called main or GSF track. In order to take into account the energy deposits produced by these photons, clusters associated with the tangents to the main track are considered. In addition, the possibility of their conversion in the material is studied. If both the track and the selected clusters are isolated[3], energy deposits obtained with the standard (STD) reconstruction are included. In this case the reconstruction procedure is based on the idea of searching for clusters in a window in a given (η, ϕ) region around the cluster seed. In case of non-isolated electrons, this method can be strongly influenced by nearby particles. The final identification is performed by combining the tracker and calorimeter information with a proper weight in order to match the expected energy resolutions. The track and ECAL clusters (also those representing bremsstrahlung photons) corresponding to the identified one (PF electron) are removed from the blocks.

- **charged hadrons:** In order to improve the rejection of fake charged hadrons, the tracks are kept only if the relative uncertainty on the measured p_T is smaller than the relative calorimetric energy resolution expected for charged hadrons. If the total calibrated calorimetric energy is smaller than the total track momentum by more than three standard deviations, muons and fake tracks are searched for. All global muons that failed the previous selection but have a momentum precision greater than 25% are now identified. Moreover tracks are rejected one by one according to their measured transverse momentum uncertainty until the uncertainty of the remaining tracks is smaller than 1 GeV/c, or when the removal of a track makes the total track momentum smaller than the calibrated calorimetric energy. Charged hadrons are reconstructed with the remaining tracks.

- **photon and neutral hadrons:** If the calibrated energy of the closest ECAL and HCAL clusters linked to a track is significantly larger than the total associated charged-particle momentum, the search for neutral particles is performed. In particular if the energy excess 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 excess. 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.

3.2 Reconstruction of the Z → ee + jets event

The Z boson can decay both in the hadronic and in the leptonic channel. In this study only the electron decay, $Z \to ee$, is taken into account. The final state is characterized by the presence of two isolated electrons and, in the particular case in which we are interested, some energetic jets.

[3]If in a cone of radius $R = 0.3$ (for the track also $|\Delta\eta| > 0.05$) the total energy of other tracks (clusters) normalized by the analysed track (clusters) energy is below a given threshold, the object is isolated.

3.2.1 Electron identification

There are two possible types of reconstructed electrons: the one made by the Particle Flow algorithm (PF), using all the information of the detector as described before, and the standard algorithm (STD) [59], using a gaussian sum procedure over the reconstructed hits. A detailed study on the performance of the two algorithms (see chapter 4 for more details) was carried on, in which a good agreement between the two approaches was found.

In order to identify electrons coming from the Z boson decay, a list of selections is applied:

identification variables

- shape of the energy deposit, requiring the energy dispersion along η to be limited;
- alignment between energy deposit and track, requiring the barycentre of the cluster and the extrapolation of the track position at the calorimeter surface to match within the cuts in η and ϕ;
- hadronic to electromagnetic energy ratio, required to be lower than the threshold to discriminate hadrons;

isolation variables

- particle-based isolation, used to exclude electrons produced by hadronization, the sum of p_T of charged hadron (neutral hadrons or photons) in a cone of radius $R = 0.3$ around the electron normalized by the electron p_T must be lower than the threshold;

rejection conversion variables

- missing hits, the number of crossed tracker layers without compatible hits with the track is at most one;
- partner conversion track, looking for an oppositely charged track whose polar angle is close to the one of the electron track, the distance at the point where the two tracks are parallel must be within the cuts.

3.2.2 Jet reconstruction

From an operative point of view, the jet reconstruction starts from a list of pseudo particles, like tracks or calorimeter clusters, or particles at generator level when working on MC studies, which are organized in more complex objects called jets [60], following precise and well defined rules. Different kind of inputs to the jet clustering algorithm are used at CMS [61], resulting in four type of jets: calorimeter jets, Jet-Plus-Track (JPT) jets, Particle-Flow (PF) jets, and track jets.

The calorimeter jets are reconstructed combining the information obtained with ECAL and HCAL only, properly corrected in order to reduce the electronic noise and

the pile-up effects. They are mostly used at trigger level, because they represent the best compromise between resolution and computational time. The Jet-Plus-Track jets are essentially resulting from the same procedure as the calorimeter jets, but the energy of the charged particles inside the jet is corrected by taking into account the momentum measurement provided by the tracking system. The starting collection for the track jets is composed by all the tracks, but in this way only charged particles are considered and the neutral contribution is lost. The PF jets are obtained starting from the list of particles identified by the PF algorithm hence they are the result of the combination of information coming from the whole detector. For this reason they have the best resolution and are the most used at analysis level.

There are different jet reconstruction procedures developed and used in the last years. In the analysis considered in this work, jets are reconstructed using the anti-k_T algorithm [62], a sequential recombination procedure. It is based on two distance definitions: d_{ij} between elements i and j and d_{iB} between element i and the beam (B). They are defined by the following expressions:

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \qquad with \qquad \Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2, \tag{3.1}$$

$$d_{iB} = k_{Ti}^{2p}, \tag{3.2}$$

where k_{Ti}, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i. The radius parameter R corresponds to the maximum distance between two elements inside a jet and is set equal to 0.5. The parameter p is needed to generalise the distance definition, with $p = -1$ corresponding to the anti-k_T algorithm. For $p = 1$ the inclusive k_T algorithm is obtained, while $p = 0$ corresponds to the inclusive Cambridge/Aachen algorithm. Thanks to the definition of the distance with $p = -1$, the low k_T particles do not cluster among themselves, but are recombined with the high k_T ones, not distorting the general jet shape. In this way the resulting jets have a more regular configuration with respect to the other procedures, making the algorithm more robust for non-perturbative effects, like hadronization and underlying event characteristics, improving also the momentum resolution. In Fig. 3.3 an example of results obtained with the different algorithms is shown.

Starting from the set of objects chosen for the jet reconstruction, the distances between pseudo-particle are calculated and the algorithm identifies the smallest one. In the case of d_{ij}, elements i and j are recombined, summing their quadrimomentum: the previous elements are removed and a new one is created. In the case of d_{iB}, the element i is identified as a jet and is removed from the list of pseudo-particles. The distances are recalculated and the procedure is repeated until no element is left.

Reclustering

Due to their large energy, electrons from the Z decay can be reconstructed as jets on their own: it is necessary to identify and to reject them before the clusterization.

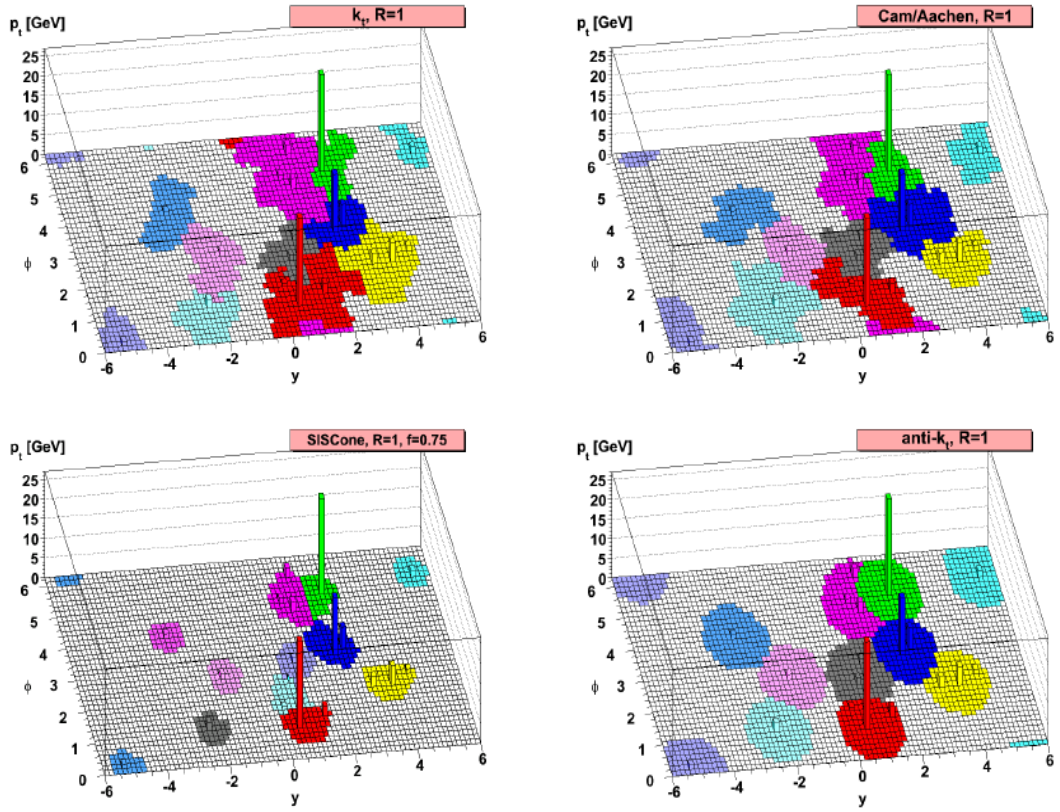


Figure 3.3: Example of jets obtained with four different algorithms in a simulated event with a few high energetic jets and many low k_T particles. Jets clustered with the anti-k_T algorithm (bottom right) have a more regular shape with respect to the others: k_T with $R = 1$ and $p = 1$ (upper left), Cambridge/Aachen with $R = 1$ and $p = 0$ (upper right), SisCone with $R = 1$ and $p = 0.75$ (bottom left).

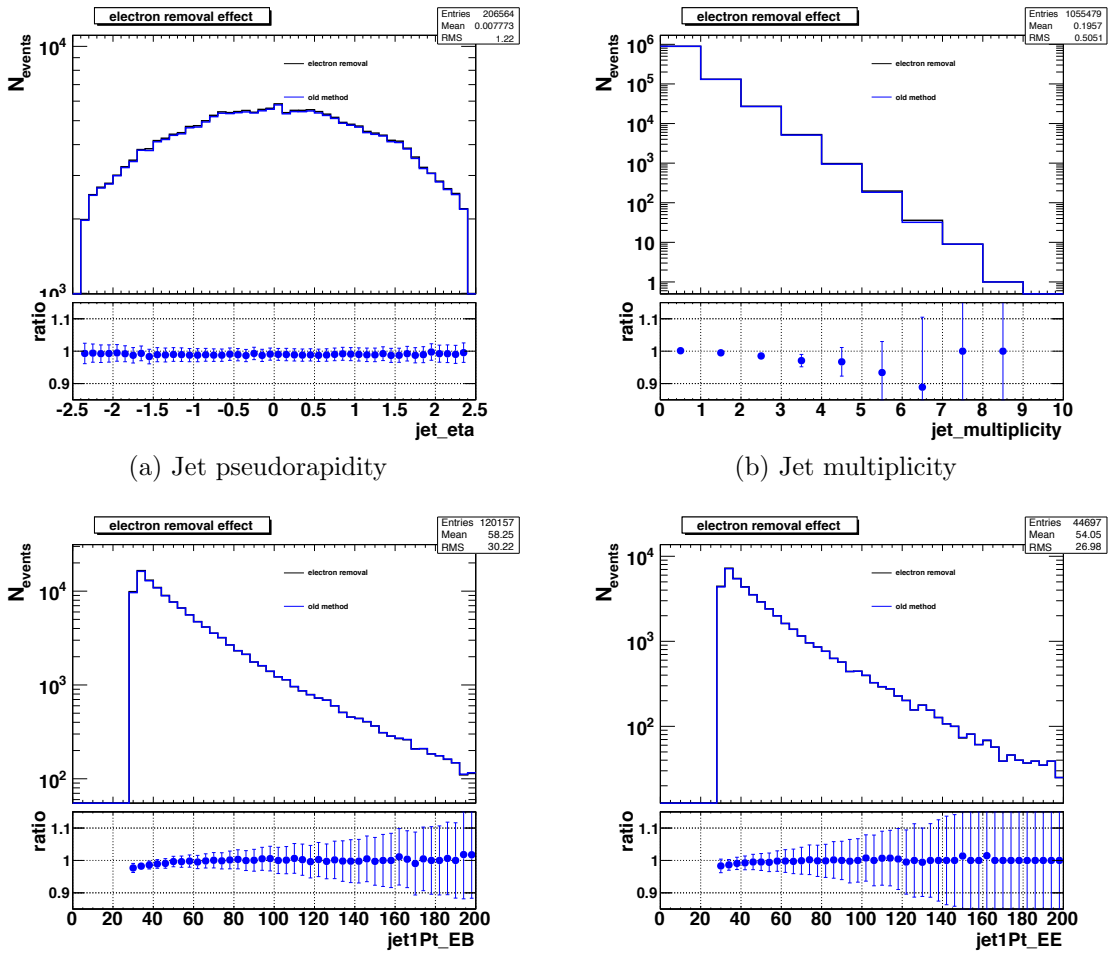
In other analyses [27] this procedure was realized by the removal of all the jets in a cone of radius $R = 0.3$ around the lepton. These jets are constituted by the lepton, that represents the main energy contribution, and by other particles, as photons and hadrons. In addition to the electron, also these other particles are subtracted from the total jet energy and their contribution in the whole event is lost, interfering with the counting of the number of jets.

In this work a different approach was used to remove these "fake" jets, discarding the electron from the particle collection before the jet reconstruction, avoiding to remove additional particles that must be included in jets. All the leptons that pass the selections (see Sec. 5.2.2) but the transverse momentum and invariant mass cuts are identified and discarded from the particle list.

In Fig. 3.4 jet distributions obtained with the standard method and with the removal of the electron before the clusterization are shown. The results are compatible; using the second method, an increase of the number of total jets is observed, as expected. The additional objects are equally distributed in pseudorapidity (Fig. 3.4a), but they are present especially in events with high multiplicity (Fig. 3.4b) and are characterized by low momentum (Fig. 3.4c and Fig. 3.4d). In events with high number of jets, if not subtracted before, there is a greater probability to cluster the lepton with particles belonging to other jets, decreasing the energy of the real jets. If the jets are close to the transverse momentum threshold, they could fall outside the allowed p_T range, decreasing the number of available objects in the analysis.

Jets at generator level

A jet collection at generator level (genJets) is needed in order to compare in the unfolding procedure the Monte Carlo truth with the reconstruction events after the simulation of the interaction with the detector (these jets will be called recoJets). In order to create objects directly comparable, similar clustering algorithms must be applied to both jet types and pseudo-particle collections with equivalent characteristics have to be considered. The standard set of particles used is composed by all stable ($c \cdot \tau > 1$ cm) particles generated in the simulation, excluding resonances and neutrinos. The selection cuts applied to genJets are the same as for the recoJets. In order to proceed in parallel with the selection in the analysis, also leptons which could derive from a Z decay and the corresponding photons from final state radiation (FSR) that are included in the lepton through the reconstruction algorithm are removed. The electron energy cluster extension in azimuthal angle reflects the bremsstrahlung emission along the lepton trajectory deviated by the magnetic field. The pseudorapidity dimension, instead, could give an indication of the region in which the FSR is included in the final object called lepton [63]. At generator level the object considered in the study is composed by the electron and all the photons in a cone around it. The value used for the radius is 0.1. The new object is obtained as sum of the quadrimomenta of its constituents. The new e^+ and e^- with the greatest p_T are selected and their constituents (electrons and photons) are removed from the particle collection used for the genJets reconstruction. The pseudorapidity, transverse momentum and invariant mass (only for the $e + \gamma s$ objects) of the removed



(a) Jet pseudorapidity

(b) Jet multiplicity

(c) Jet transverse momentum in EB

(d) Jet transverse momentum in EE

Figure 3.4: Comparison between jet distributions obtained with the removal of jets in the cone around the electron (blue) and the electron removal before the clusterization (black).

particles are shown in Fig. 3.5.

The remaining particles, after the exclusion of the selected electrons and photons, are clustered into jets using the same algorithm and jet selection cuts as applied for real data reconstruction.

3.2.3 Jet energy corrections

The energy of the reconstructed jets is usually different from the true energy of the clustered particles. The main sources of this effect are the non-uniform and non-linear response of the calorimeters, electronics noise and presence of pile-up events. The energy of each PF particle is calibrated depending on its type, hence, exploiting this PF algorithm feature, the final jets need only small energy corrections in order to match the real energy of their constituents with respect to the other types of jets.

The application of the jet energy corrections (JEC) corresponds to a factorized multi-step procedure [64], structured in subsequent steps of correction:

- the **offset** correction (L1) corrects the jet energy spoiled by effects of electronic noise and pile-up;
- the **relative** correction (L2) uniforms the jet response as a function of the jet pseudorapidity, choosing a central control region as reference;
- the **absolute** correction (L3) uniforms the jet response as a function of the jet transverse momentum
- the **residual** correction (RES) takes care of remaining differences between data and Monte Carlo coming from an imperfect simulation.

In CMS the JEC are provided with two complementary approaches. The first one relies on the MC truth information (MC truth JEC) and is used to provide a first evaluation of L2 and L3 corrections; the second one uses the experimental data, exploiting physics processes from hadron collisions, in order to derive the L1 and the residual L2 and L3 miscalibration.

The application of the jet energy corrections can be expressed as:

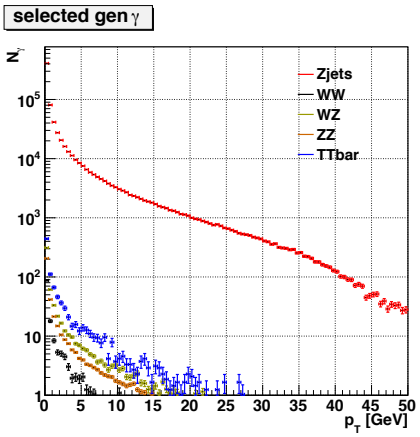
$$p_{\mu}^{cor} = C \cdot p_{\mu}^{raw} \tag{3.3}$$

$$C = C_{L1}(p_T^{raw}) \cdot C_{MC^{L2L3}}(\eta, p_T^{'}) \cdot C_{RES^{L2}}(\eta) \cdot C_{RES^{L3}}(p_T^{''}) \tag{3.4}$$

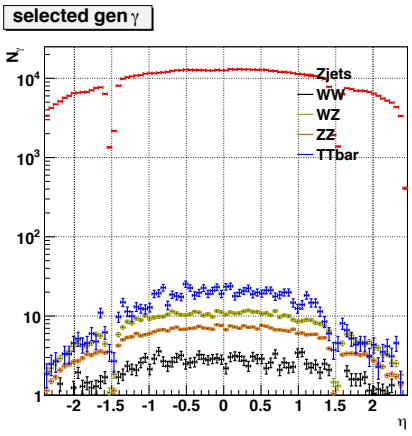
where $p_T^{'}$ is the transverse momentum of the jet corrected for L1, $p_T^{''}$ is the transverse momentum of the jet after all previous corrections and the C factors represent the different corrections.

Offset correction

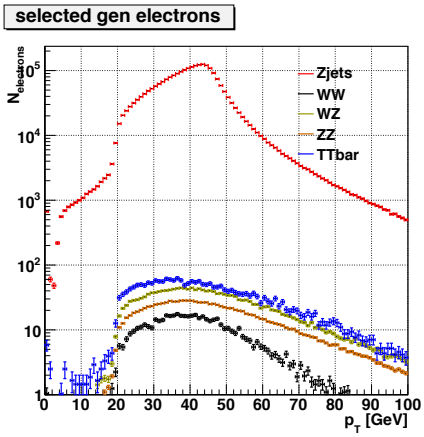
It is the first factor of the correction chain, with the goal of subtracting the energy not related with the main interaction, using the jet area strategy [65], based on the evaluation of the jet area A_j.



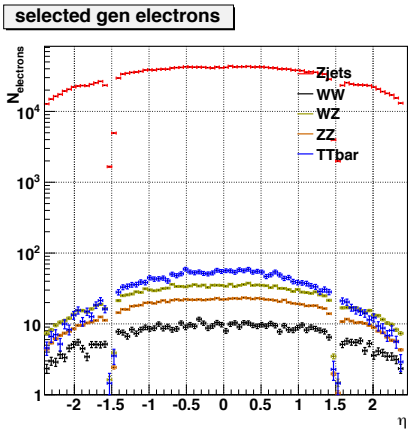
(a) Photons transverse momentum



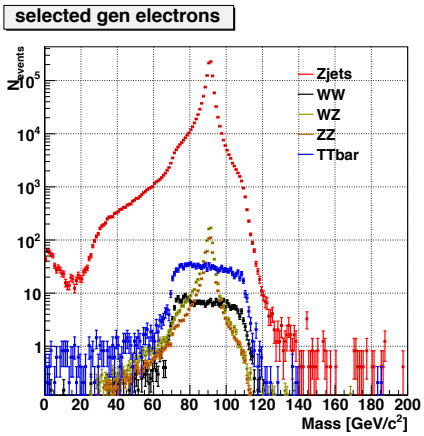
(b) Photons pseudorapidity



(c) Electron transverse momentum



(d) Electron pseudorapidity



(e) Electron invariant mass

Figure 3.5: Transverse momentum, pseudorapidity and invariant mass distributions for electrons and photons removed from the list of particles used to create jets at generator level.

A large quantity of soft artificial particles, not able to modify the properties of the jet, are clustered together with the true jet constituents. The region in the (η, ϕ) space covered by the soft component in each jet provides the value of the active jet area. For each event, an average transverse momentum density per unit area ρ is calculated with the k_T jet clustering algorithm with a distance parameter $R = 0.6$. It is defined as the median of the distribution of the variable p_{T_j}/A_j, where j runs over all the jets in the event.

Fig. 3.6 (left) shows the transverse momentum density ρ as a function of the leading jet p_T in QCD events for various pile-up conditions. The ρ variable scales linearly with the number of vertices, and since it does not depend on the hard scale of the event, it is a good tool to measure the soft jet activity.

In order to take care also for dependence on the pseudorapidity, a sample of zero bias events is considered. They are collected using random triggers in presence of beam crossing. The energy inside a cone of radius $R = 0.5$ in the (η, ϕ) space is summed. The total average offset is then classified according to the number of vertices.

The correction is a factor applied jet-by-jet that can be defined as:

$$C_{offset}(p_T^{raw}, A_j, \rho) = 1 - \frac{(\rho - \langle \rho_{PV1} \rangle) \cdot \beta(\eta) \cdot A_j}{p_T^{raw}}, \tag{3.5}$$

where p_T^{raw} is the uncorrected jet transverse momentum, $\langle \rho_{PV1} \rangle$ is the p_T-density measured in events with only one reconstructed primary vertex (without pile-up), and $\beta(\eta)$ is a multiplicative factor describing the modulation of the average offset in p_T as a function of η, as in Fig 3.6 (right).

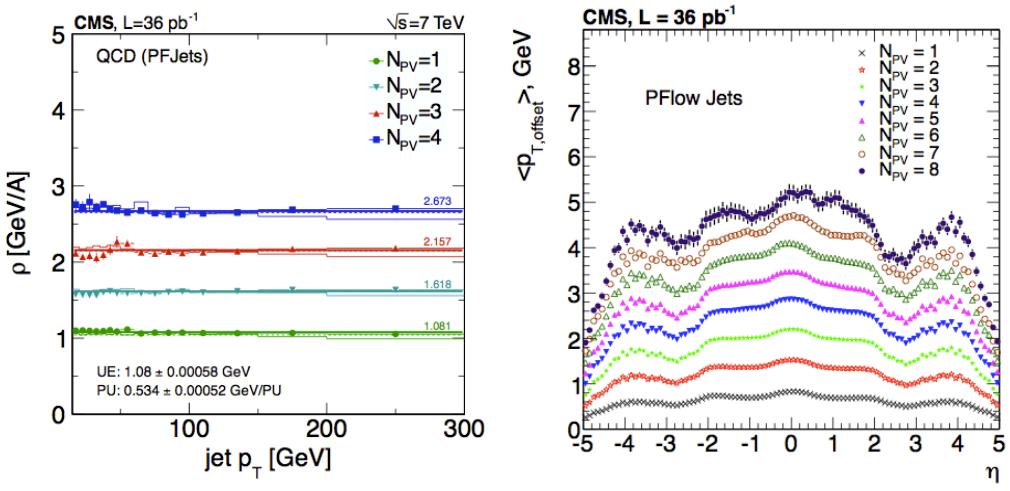


Figure 3.6: Left: transverse momentum density ρ as a function of the leading jet p_T in QCD events for different pile-up conditions (N_{PV} is the number do reconstructed vertices). Right: average offset in p_T as a function of η for different pile-up conditions.

Monte Carlo truth jet energy corrections

The Monte Carlo truth jet energy corrections are obtained from a sample of simulated QCD events at $\sqrt{s}$= 7 TeV in order to remove the bulk of the non-uniformity in η and the non-linearity in p_T. The procedure works by steps:

- the different type of jets are reconstructed, calorimeter, JPT and PF jets (recoJets) as well as genJets[4];
- the genJets are matched with the recoJets in the (η, ϕ) space by requiring their distance to be $R < 0.25$ and the quantity $p_T^{recoJet}/p_T^{genJet}$ is studied to extract the correction factors as a function of p_T and η;
- the L2^{MC} is derived comparing the response at a given pseudorapidity with the one of central jets $|\eta| < 1.3$;
- the L3^{MC} is computed, bringing the p_T dependence of the jet response to unity.

The combined correction factor $C_{MC}(p_T, \eta)$ is given by the following expression:

$$C_{MC}(p_T, \eta) = C_{MC^{L2}}(\eta, p_T) \cdot C_{MC^{L3}}(p_T \cdot C_{MC^{L2}}(\eta, p_T)). \tag{3.6}$$

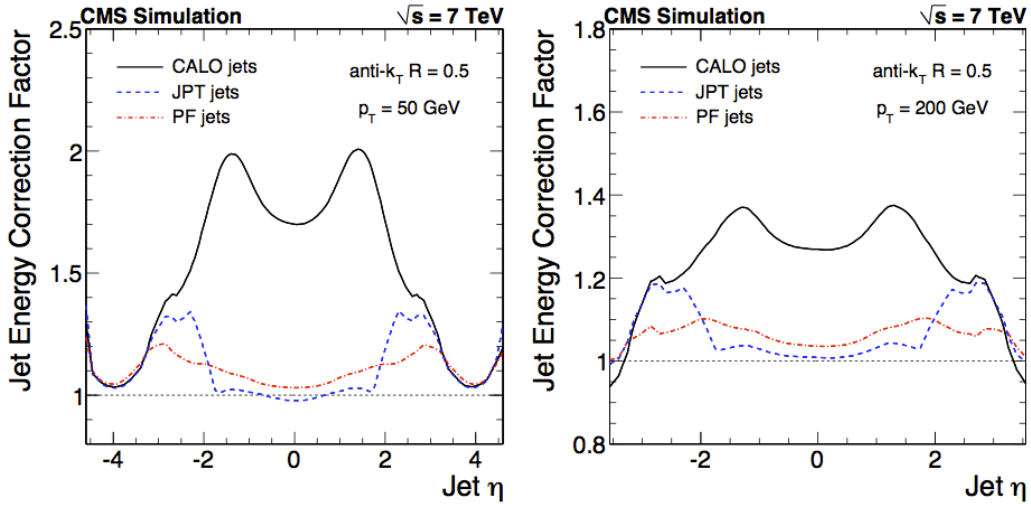


Figure 3.7: Total Monte Carlo jet energy correction factors $C_{MC}(p_T, \eta)$ for different jet types, as a function of jet pseudorapidity, for jets with $p_T = 50$ GeV (left) and $p_T = 200$ GeV (right).

Fig. 3.7 shows the MC-truth jet energy correction factors for calorimeter, JPT and PF jets, as a function of η at two different corrected jet p_T values, while Fig. 3.8 shows the average correction in $|\eta| < 1.3$ as a function of the corrected jet p_T.

A structure in the pseudorapidity dependence of $C_{MC}(p_T, \eta)$ is observed and is most evident for the calorimeter jets, due to the non-linear response of ECAL

[4]As previously described, they are obtained with MC stable particles at generator level.

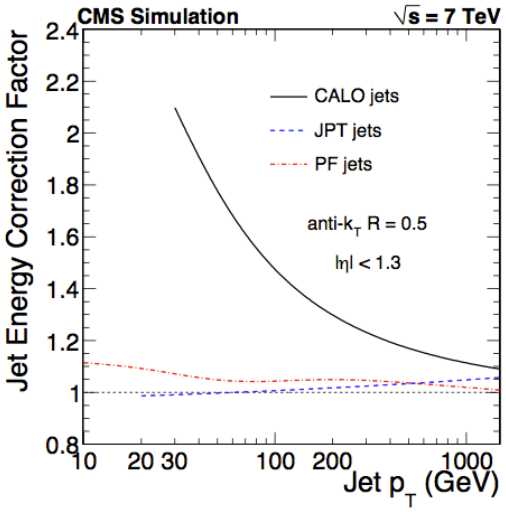


Figure 3.8: Total Monte Carlo jet energy correction factors $C_{MC}(p_T, \eta)$ for different jet types, as a function of jet transverse momentum.

to hadronic jets. It is possible to divide the detector in three calorimeter regions with different behaviours: barrel ($|\eta|$ <1.3), endcaps ($1.3 < |\eta| < 3$) an forward ($3 < |\eta| < 5$). Overall correction factors are smaller for JPT and PF jets, because the charged component of the jet profits from the tracking information; this is shown particularly in the pseudorapidity region $|\eta| < 2.4$. In the forward region, since the three jet types converge to simple calorimeter objects, the corrections are similar.

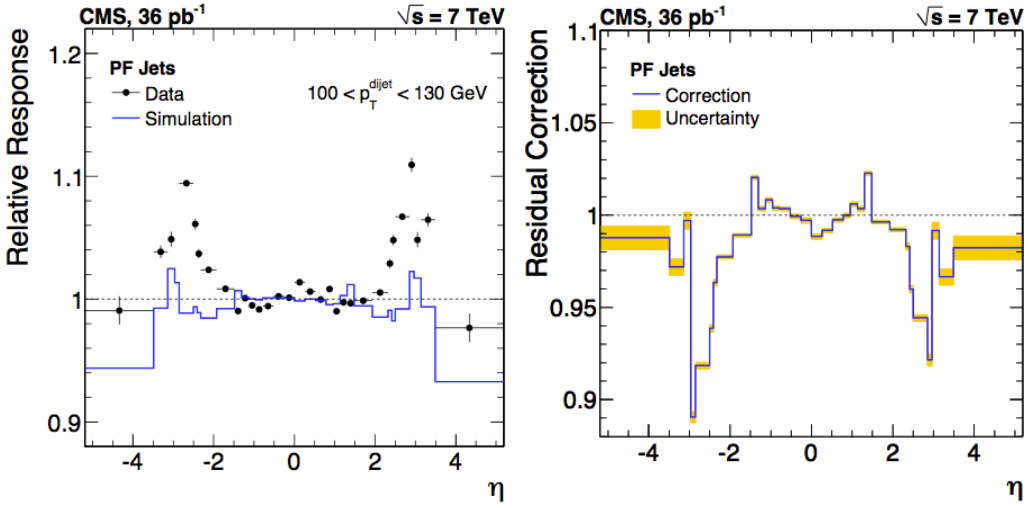


Figure 3.9: Left: Relative jet energy response as a function of η, measured with the dijet balance method. Right: Relative jet energy residual correction as a function of η with its uncertainties (band).

Residual correction

The relative residual correction is equal to the difference between the relative response in data with respect to the MC, expressed in terms of the jet $|\eta|$. In data the dijet p_T balance technique is used in back-to-back dijet events (with at least two jets) to measure the response of a jet at any η (probe jet) relative to the jet energy response in $|\eta| < 1.3$ (barrel jet). The central region is chosen as reference because of the uniformity of the detector and because it has highest jet transverse momentum reach. The two leading jets must be azimuthally separated by $\Delta\phi > 2.7$ and no additional jets with $p_T^{3rdJet}/p_T^{ave} > 0.2$ are allowed, where $p_T^{ave} = (p_T^{probe} + p_T^{barrel})/2$ is the average uncorrected p_T of the two leading jets.

The balance quantity B is defined as:

$$B = \frac{p_T^{probe} - p_T^{barrel}}{p_T^{ave}}. \tag{3.7}$$

The average value of the B distribution, $< B >$, in a given η^{probe} and p_T^{ave} bin is used to determine the relative response, see Fig. 3.9 (left):

$$R(\eta^{probe}, p_T^{dijet}) = \frac{2+ < B >}{2- < B >}. \tag{3.8}$$

The correction is tipically of the order of few percent with the exception of the region $2.5 < |\eta| < 3.0$ where higher values are reached as shown in Fig. 3.9 (right).

Fig. 3.10 shows the data/MC ratio for the relative response before and after the residual correction application, demonstrating they establish a good agreement between data and simulation.

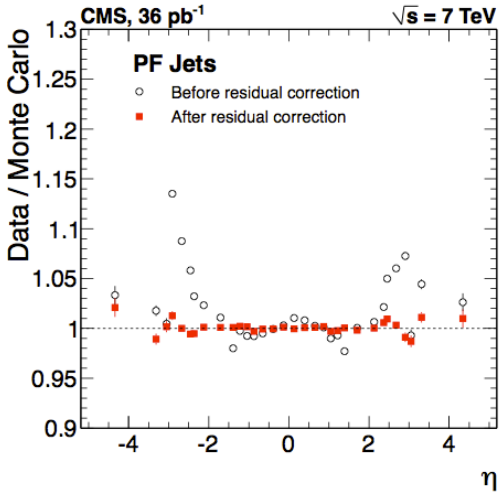


Figure 3.10: Relative response ratio between data and Monte Carlo simulation before and after the residual correction.

The last step is the calculation of the residual absolute correction factor. It is performed with a MPF (Missing E_T Projection Fraction) method applied to a sample of γ + jets events. The MPF method is based on two assumptions:

- the γ + jets events have no intrinsic missing transverse energy E_T^{miss};
- the photon is perfectly balanced by the hadronic recoil in the transverse plane, $\overrightarrow{p}_T^{\gamma} + \overrightarrow{p}_T^{recoil} = 0$

For reconstructed objects, the following equation can be written:

$$R_\gamma \overrightarrow{p}_T^{\gamma} + R_{recoil} \overrightarrow{p}_T^{recoil} = -\overrightarrow{E}_T^{miss}, \tag{3.9}$$

where the R_γ and R_{recoil} are the detector responses to the photon and the hadronic recoil. Solving the equation above, R_{recoil} is obtained:

$$R_{recoil} = R_\gamma + \frac{\overrightarrow{E}_T^{miss} \cdot \overrightarrow{p}_T^{\gamma}}{(\overrightarrow{p}_T^{\gamma})^2} \equiv R_{MPF} \tag{3.10}$$

The jet energy response has to be extracted from the measured R_{MPF}, considering that $R_{recoil} = R_{leading}$ is a good approximation if particles not clustered in the leading jet have a response similar to the ones inside the jet, or if these particles are in a direction perpendicular to the photon axis.

The total jet energy correction factor and its uncertainty are shown in Fig. 3.11 as a function of η, for two different jet p_T values.

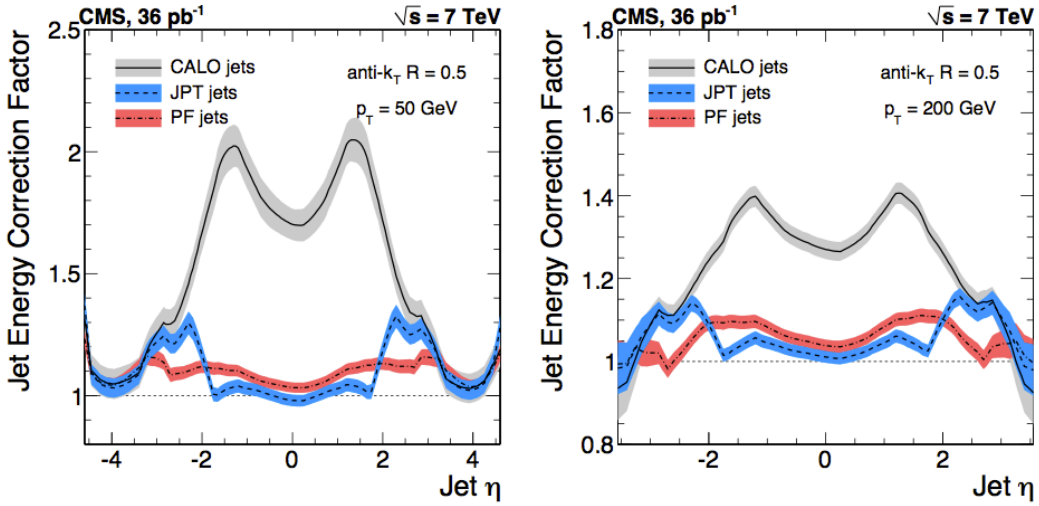


Figure 3.11: Total jet energy correction factor and its uncertainty (band) as a function of jet η for two jet p_T values.

Chapter 4

Electron reconstruction with a Particle Flow based approach

The CMS framework includes two algorithms for the electron reconstruction which produce different objects: the standard electrons (GSF) [66], [67] and the Particle Flow ones [68] (already introduced in the previous chapter).

The previous analysis of the associated production of vector bosons and jets [27] has used electrons from the standard reconstruction, because the GSF algorithm is the first one developed for electrons and this type of electrons and super clusters are the objects used by the trigger. On the other hand, the jets used in the analysis are composed by particles from the PF reconstruction. Since the GSF and PF algorithms exploit raw detector information using different strategies, they are not fully integrated. For this kind of analysis, it is essential to have a consistent reconstruction of the whole event in order to avoid a potential double counting of ECAL energy deposits or generation of artificial missing energy. A possible solution is to move to a reconstruction fully based on the Particle Flow algorithm. The goal of this study is to test the reconstruction performance of the PF object, to check if using the PF electrons the results are consistent with the ones obtained with the GSF electrons and to choose the best solution for the Z boson reconstruction.

The basic steps of the algorithms and their main differences are summarized in the following list (1 is referring to the GSF electrons, 2 to the PF ones):

- the selection of the seed from which the procedure starts:
    1. ECAL energy cluster,
    2. simple tracks;
- the super cluster[1] (SC) reconstruction:
    1. searches other ECAL energy clusters in a window around the seed,

[1]The SC is a composite energy deposit, made by several energy clusters, with the goal of reconstructing the total energy of the original electron by means of the energy of all its products.

2. searches clusters that match trajectories tangent to the main track (clusters associated to other tracks are vetoed), recovers from converted bremsstrahlung photons and use ECAL SC information to recover additional clusters under specific isolation conditions;

- the track reconstruction:

1. the hits in the tracker are combined with a filter suitable for reconstruction of tracks with kinks (Gaussian-Sum-Filter),
2. a simplified version of the full GSF filter;

- the link of the two objects:

1. from ECAL SC to tracks,
2. from tracks to ECAL deposits;

- the determination of the electron momentum, combining track and energy information or using only one of them:

1. depends especially on the E/p ratio;
2. depends on the number of track hits, and thus on the GSF track quality, then checks the sensitivity to bremsstrahlung through the difference between track and calorimeter energy.

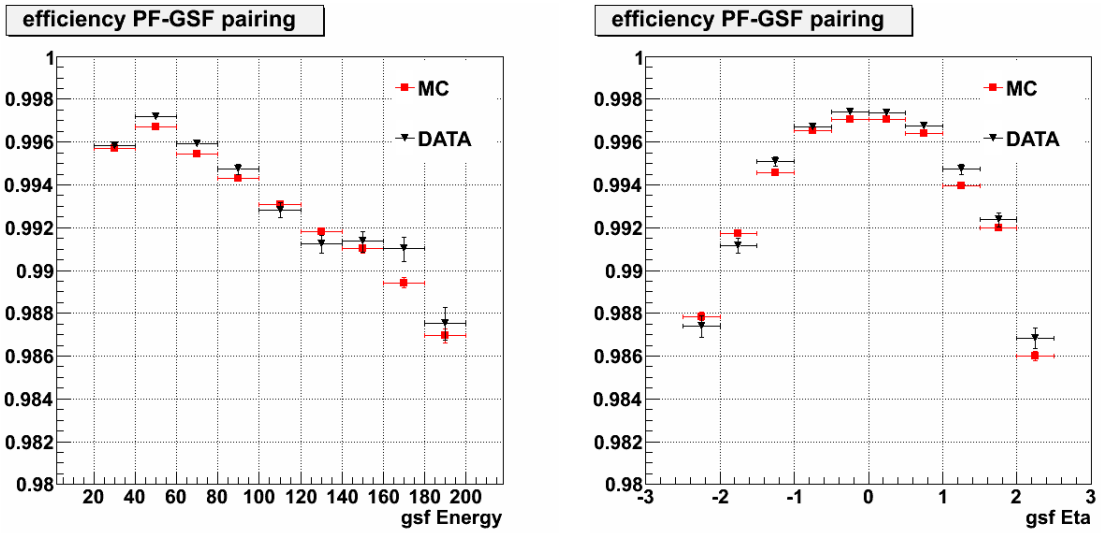


Figure 4.1: The association efficiency of PF to GSF electron as a function of the GSF electron energy (left) and pseudorapidity (right). Black is used for data, red represents Monte Carlo.

Only the events passing the same triggers and requirements presented in Sec. 5.2, used in the Z + jets measurement are studied. Similarly, the selected electrons are

the leptons which can be used for Z boson candidates reconstruction in the same analysis.

In the following, tracks are used in order to associate a GSF and a PF electron: if a GSF and a PF electron have the same main track, they are considered as describing the same object. The pairing efficiency of PF to GSF electron is shown in Fig. 4.1. The fraction of GSF electron associated is very close to one, only for less than 1% of the set it is not possible to find the corresponding reconstructed PF electron. The agreement data-MC is good.

The study is performed in steps. Initially the basic objects which take part in the electron candidate construction are considered: tracks, to associate the two different reconstructed electrons, and super clusters to compare their properties. Then the final momentum determined by the two algorithms is studied. Finally the Z boson is reconstructed using electron candidate pairs, and the boson masses obtained with the two methods are compared.

4.1 Super cluster properties comparison

Previously it was said that a GSF and a PF electron are describing the same physical object if they have the same main track. Hence the other basic component that can be studied is the ECAL SCs produced by the two algorithm.

The energies of the leptons are measured as a function of the pseudo rapidity range. Fig. 4.2 shows the distributions of the ratios of the GSF and PF SC energy for deposits in the ECAL Barrel (EB, on the left) and in the ECAL Endcap (EE, on the right) for data (in black) and Monte Carlo (MC, in red). The results for data and MC are compatible, only differences of few per mil are observed. Fig. 4.3 shows the distributions of the energy ratios for SC composed by only one basic cluster, in order to study the cluster energy itself regardless of the details of the SC reconstruction; a shift of the order of a per mil is still observed, due to the calibration and correction procedure. In all cases GSF SCs have energies greater than PF ones, especially in the endcaps, where the PF SC algorithm does not manage to recover all the electron energy.

The data energy ratios, as a function of the GSF SC energies, are shown in Fig. 4.4. In EB, the differences between the GSF and the PF SCs increase with energy up to 100 GeV becoming stationary at higher energies. For low energy the spread in η of the single clusters of a SC is greater and the windows size in the GSF algorithm could be not enough to collect all the contributions; the PF method, looking for deposits corresponding to tangent trajectory directions, is more suitable to gather also the farthest clusters. In EE the differences are greater as seen before. In the MC, instead, the performances in EB and EE for high energy are similar. In Fig. 4.5 the average energy ratios as a function of η for data (in black) and MC (in red) are presented. In EE there are higher discrepancies between GSF and PF energy, as expected from Fig. 4.2 histograms. In the EB, in both the cases, the average values are close to one.

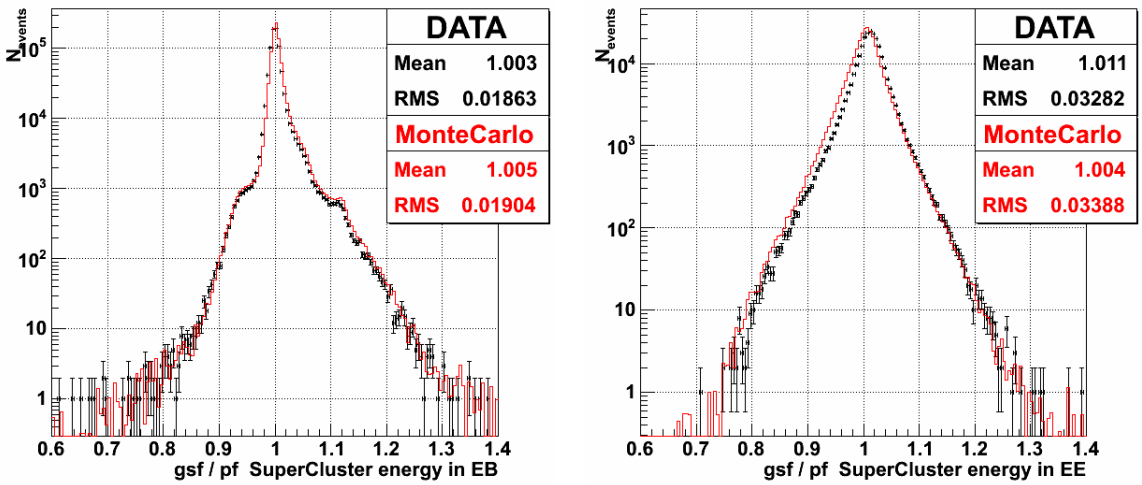


Figure 4.2: The ratios of the GSF and the PF SCs energies for deposits in the EB (left) and in the EE (right) for data (in black) and MC (in red) are shown.

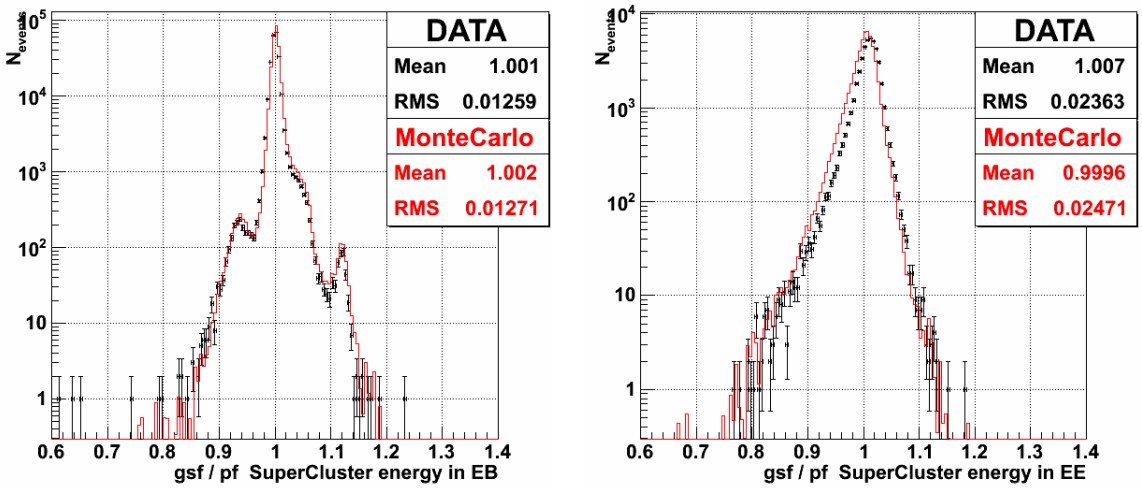


Figure 4.3: The ratios of the GSF and the PF SCs energies for deposits composed by only one basic cluster in the EB (left) and in the EE (right) for data (in black) and MC (in red) are shown.

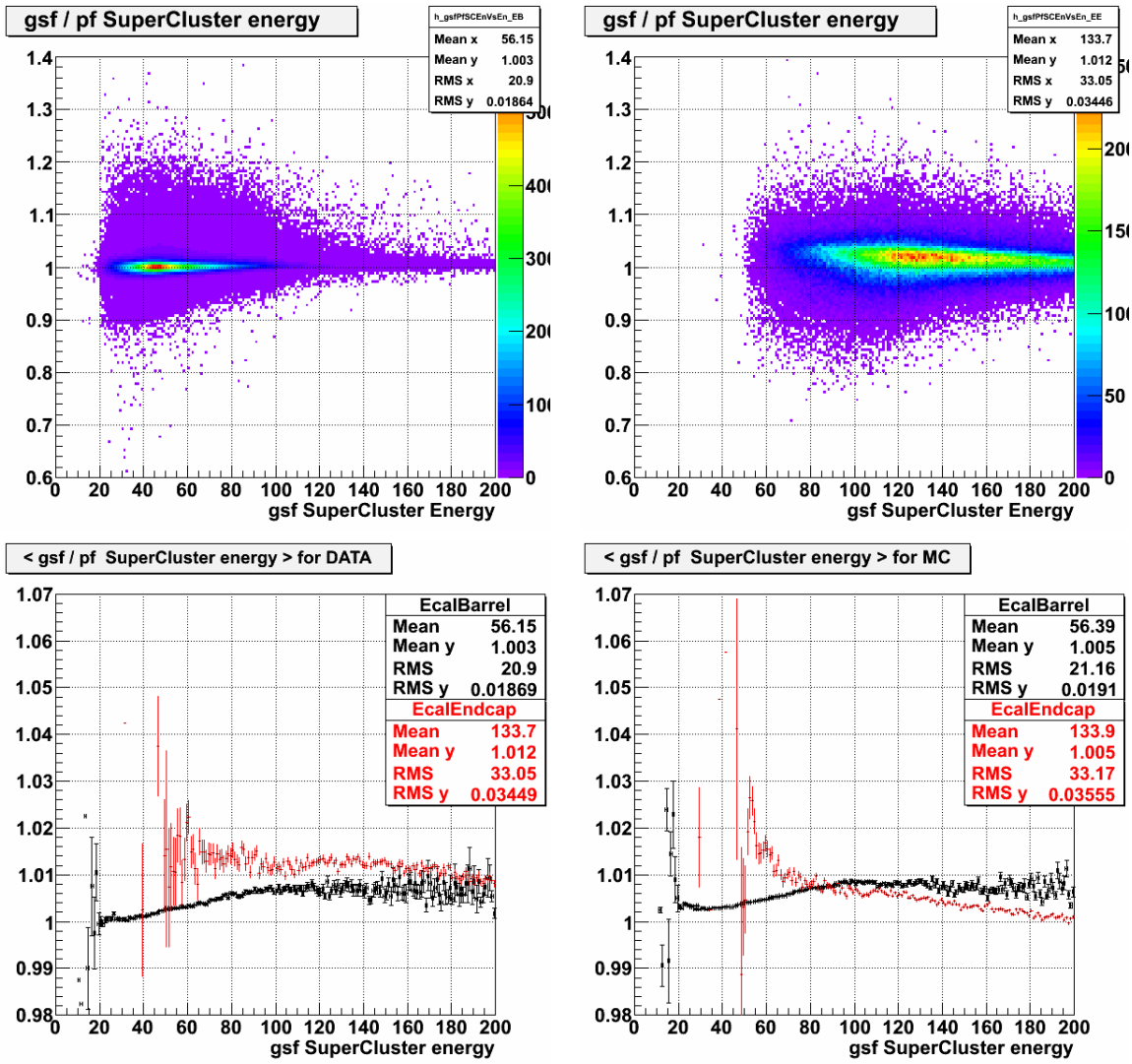


Figure 4.4: On top the ratios of the GSF and the PF SCs energies as a function of the GSF SC energy on data for deposits in the EB (left) and in the EE (right). On the bottom the average ratios of both EB (in black) and EE (in red) for data on the left and for MC on the right are presented.

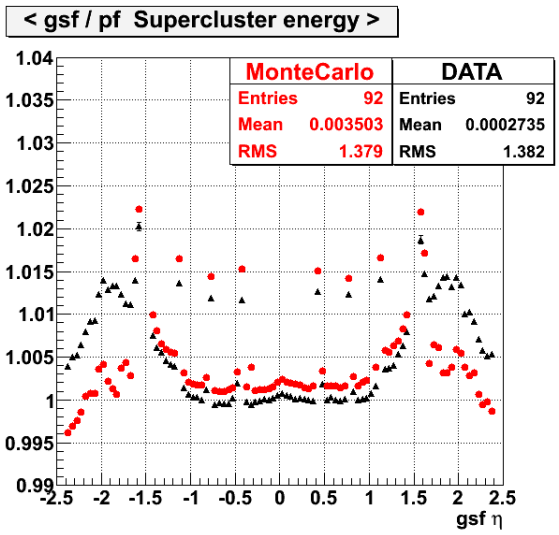


Figure 4.5: Average ratios of the GSF and the PF SCs energies of data (in black) and MC (in red) as a function of η.

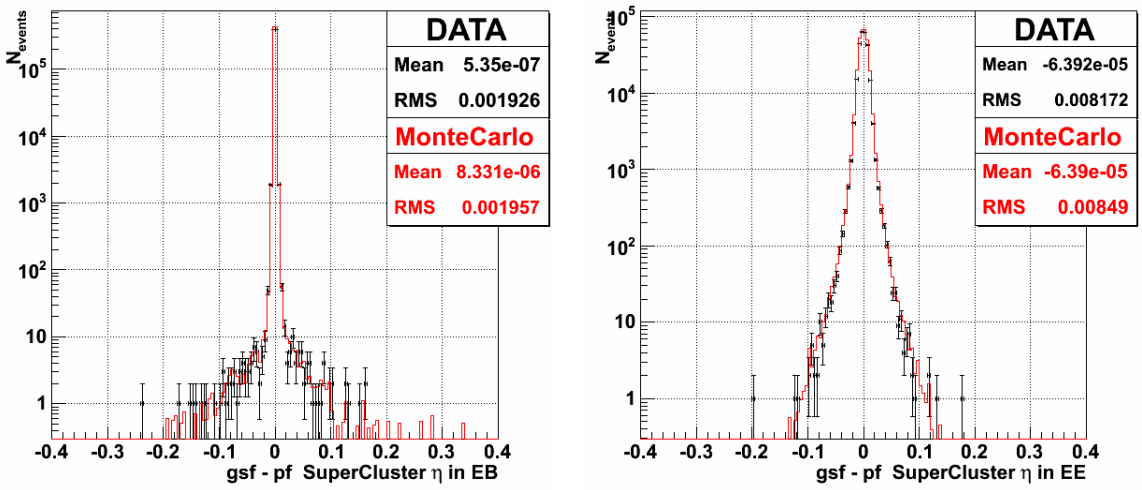


Figure 4.6: Differences of the GSF and the PF SCs pseudorapidity for deposits in the EB (left) and in the EE (right) for data (in black) and MC (in red).

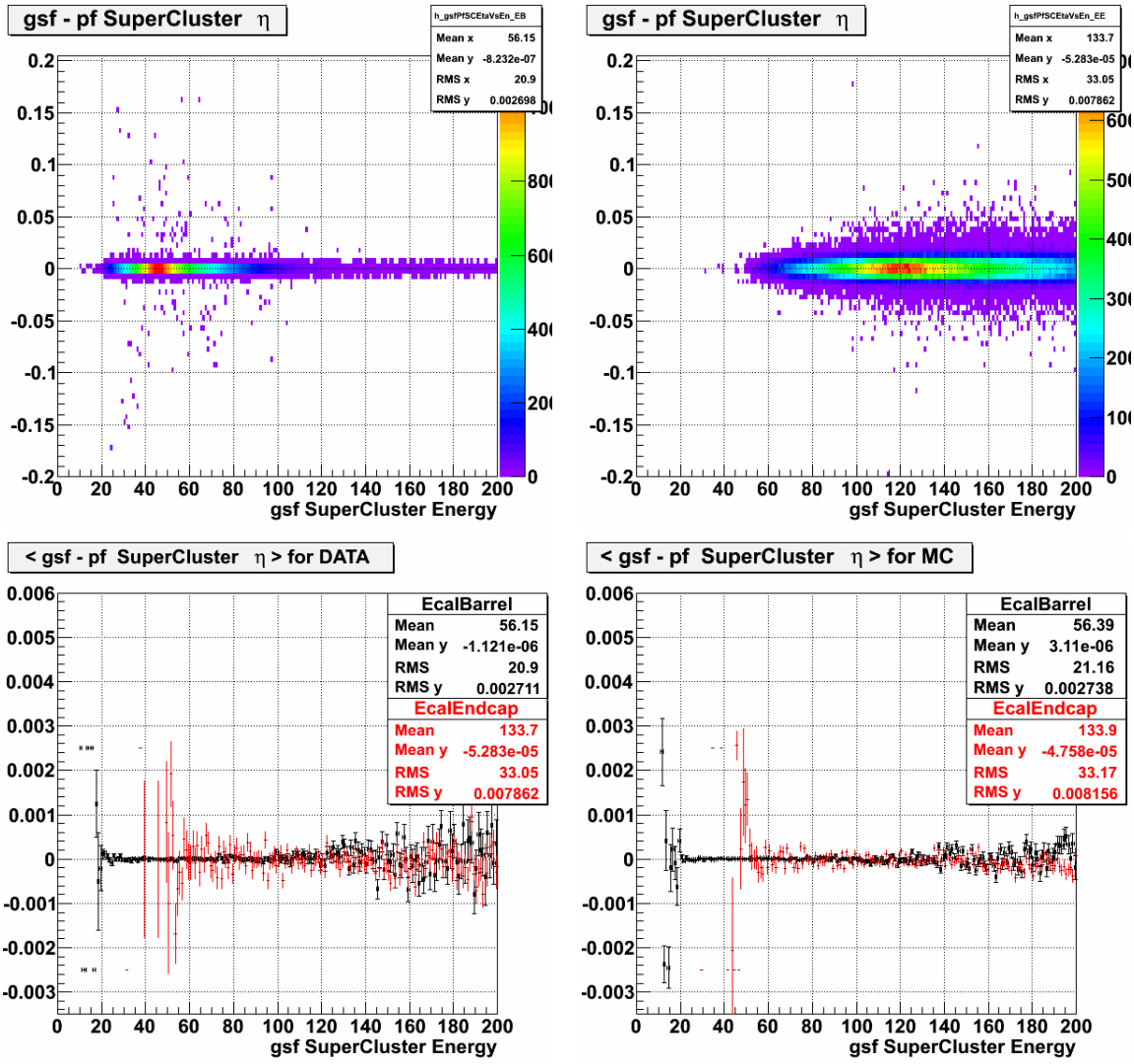


Figure 4.7: On top: differences of the GSF and the PF SCs pseudorapidity as a function of the GSF SC energy only for data for deposits in the EB (left) and in the EE (right). On bottom: average differences for deposits in the EB (in black) and in the EE (in red) as a function of the GSF SC energy, for data on the left and for MC on the right.

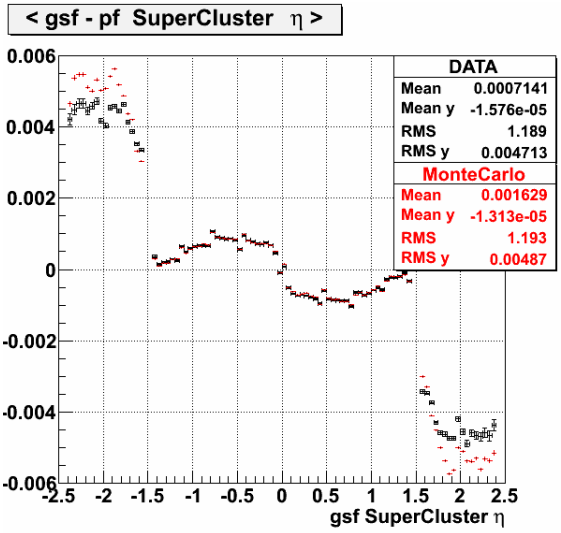


Figure 4.8: Average differences of the GSF and the PF SCs pseudorapidity as a function of the GSF SC η for data (in black) and for MC (in red).

Fig. 4.6 shows the differences between GSF and PF SCs pseudorapidity for deposits in EB (left) and in EE (right) for data (in black) and MC (in red). The agreement between data and MC is better than for the energy. Also the dispersion is better, in particular in EB. In the bottom part of Fig. 4.7 the average difference of pseudorapidity as a function of energy for deposits in EB and EE is compared. In both cases no particular trend is observed, both for data and MC. The same distribution as a function of the pseudorapidity is shown in Fig. 4.8. The disagreement between GSF and PF electrons is greater in the endcaps, where also the energy presents the bigger discrepancies. It is worth to note that it is only a few per mil effect.

In Fig. 4.9 the scale of the SC energy, pseudorapidity and azimuthal angle for GSF and PF electrons as a function of the number of vertices are shown. For what concerns η and ϕ the performances are almost the same. Differences of the order of per cent are seen for the energy scale. In the barrel region both the reconstructed electrons have a worse performance with higher number of vertices (less than per cent). In the endcap regions, in the case of GSF electron there is a slight deterioration of the scale with increasing pile up; in the case of PF electrons the performances are stable on the range taken into account. From these distributions the PF super cluster performances are better than the GSF ones, especially for increasing luminosity.

4.2 Reconstructed electron momentum comparison

In this section the comparison of the basic properties of the final objects obtained by the two reconstruction algorithms is presented.

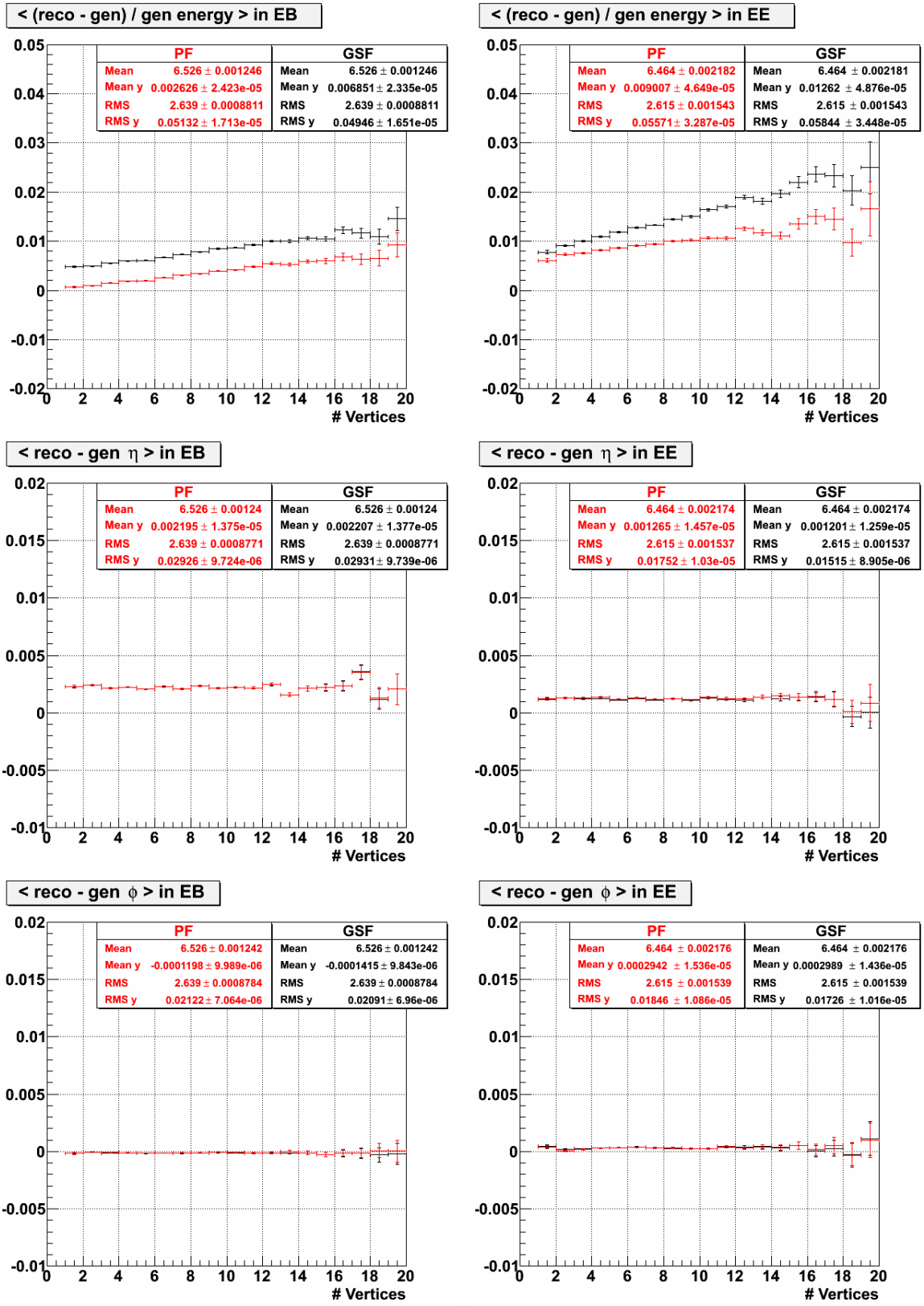


Figure 4.9: Scale of the super cluster energy (first row), η (second row) and ϕ (third row) as a function of the number of vertices, in black for GSF and in red PF electrons, for deposits in EB (left) and in EE (right).

4.2.1 GSF-PF comparison

A simple ratio between the GSF and PF electrons, considering energy, pseudorapidity and transverse momentum, is analysed. On top of Fig. 4.10 one can see the energy ratios between the two algorithms as a function of energy for data (on the left) and MC (on the right). The bottom of the same Figure shows the comparison of data and MC average values. As seen in the previous section, the GSF electron have on average greater energy, in particular for data. This is more evident for high values of energy, where the biggest discrepancies with the MC results is observed. If for the SC the differences GSF-PF increase up to 100 GeV and then become stationary, after the reconstruction of the final electron three regions are visible. For very low energy the GSF energy loss due to the pseudorapidity spread is clearer than in the SC case; a stable trend very close to one is present up to about 100 GeV, where there is a step in the ratio reporting higher values.

This variable as a function of the electron pseudorapidity is also studied and the results are shown in Fig. 4.11. A specific trend describes the subdivision of ECAL in modules and the cracks present in the detector[2]; it is reproduced by the MC. The biggest differences between the two algorithms as well as the strongest discrepancies with the MC lie in the endcaps. These differences may correspond to the high energy electrons that show the largest disagreement in data-MC comparison in the previous Fig. 4.10.

The last studied variable is the transverse momentum. (GSF-PF) / GSF p_T as a function of the p_T^{GSF} is shown in Fig. 4.12. Also in this case GSF values are in general higher than PF ones, except at very low p_T. The agreement between data and MC is better for the p_T, than in the energy case, where also the longitudinal component and the different η spread of the two algorithms are considered.

4.2.2 GEN-RECO comparison

In this subsection the reconstructed (*reco*) electrons are compared with the MC truth. Given the *reco* objects, we search for the nearest electron at generator level (*gen*) in a cone of $R < 0.05$. The association efficiency of *gen* and *reco* level objects is shown in Fig. 4.13: it agrees with one at the per-mil level everywhere, and is consistent between GSF and PF electrons.

In Fig. 4.14 the ratios (*reco* - *gen*) / *gen* for energy and transverse momentum, and differences (*reco* - *gen*) for pseudorapidity and azimuthal angle with GSF (in black) and PF (in red) results are shown. The mean discrepancy between the MC truth and the reconstructed objects for the energetic variables is of the order of per mil with a variance of a few per cent. The reconstructed level objects have greater values than the generator level ones. As far as the directional variables are concerned, the differences are smaller: of about one order of magnitude for η and two for ϕ.

[2]ECAL crack corrections are included for the GSF SCs; a similar correction is not currently available for the PF SCs.

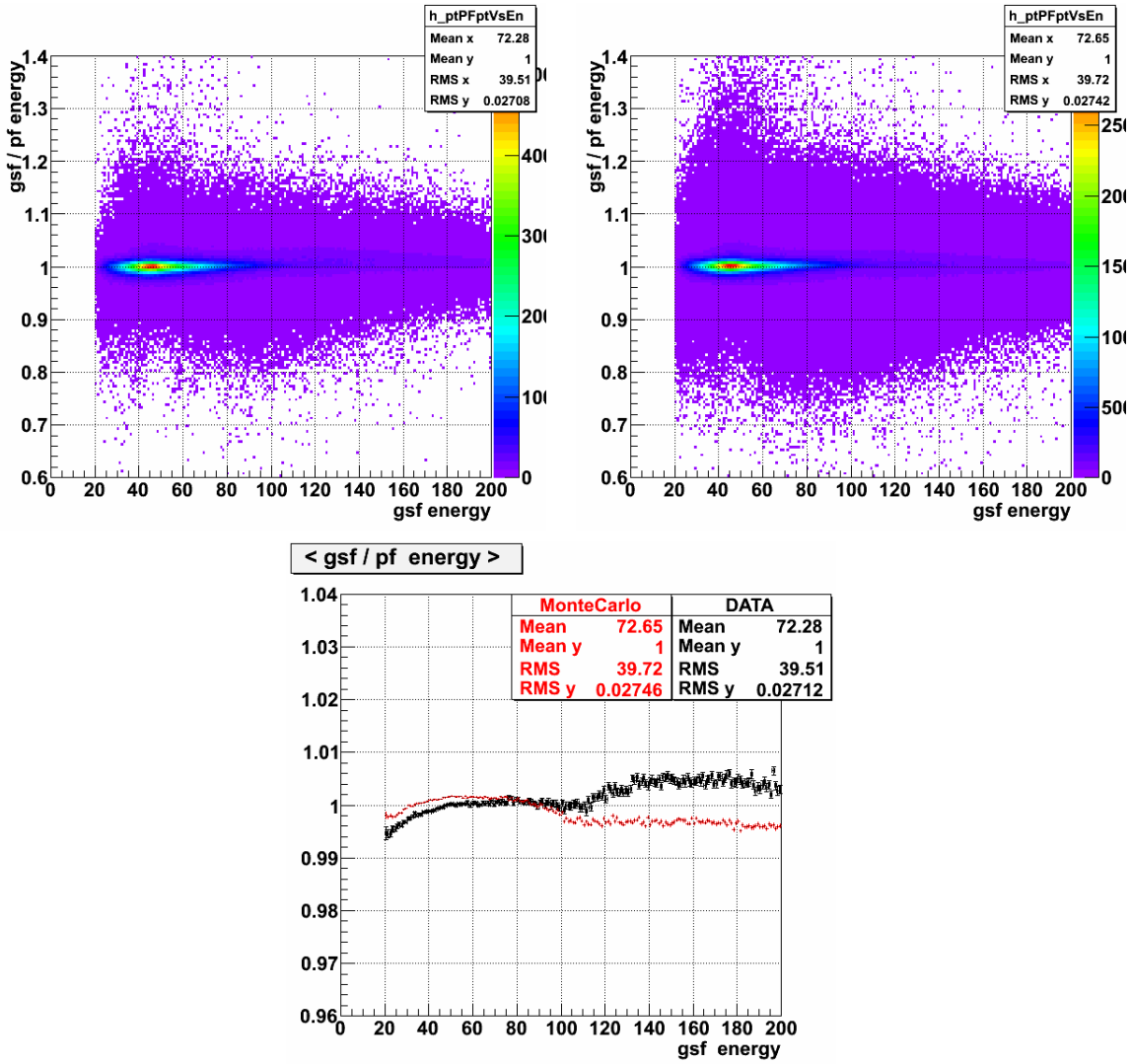


Figure 4.10: On top: ratio of GSF and PF electron energies as a function of GSF energy for data (left) and MC (right). On bottom: comparison of mean values for data and MC.

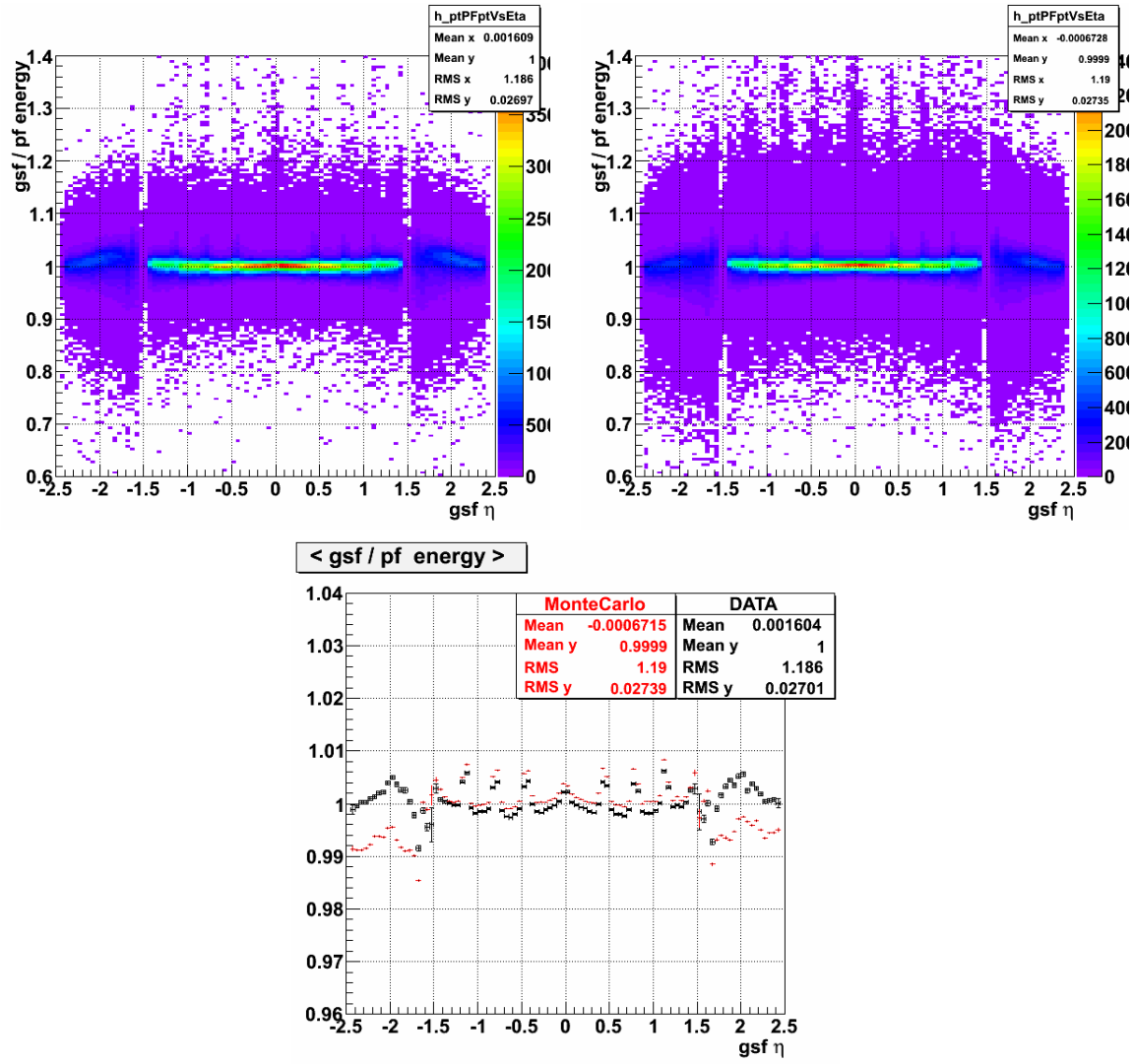


Figure 4.11: On top: the ratios of the GSF and PF electron energy as a function of the GSF pseudorapidity for data (left) and MC (right). On bottom: comparison of the mean values for data and MC.

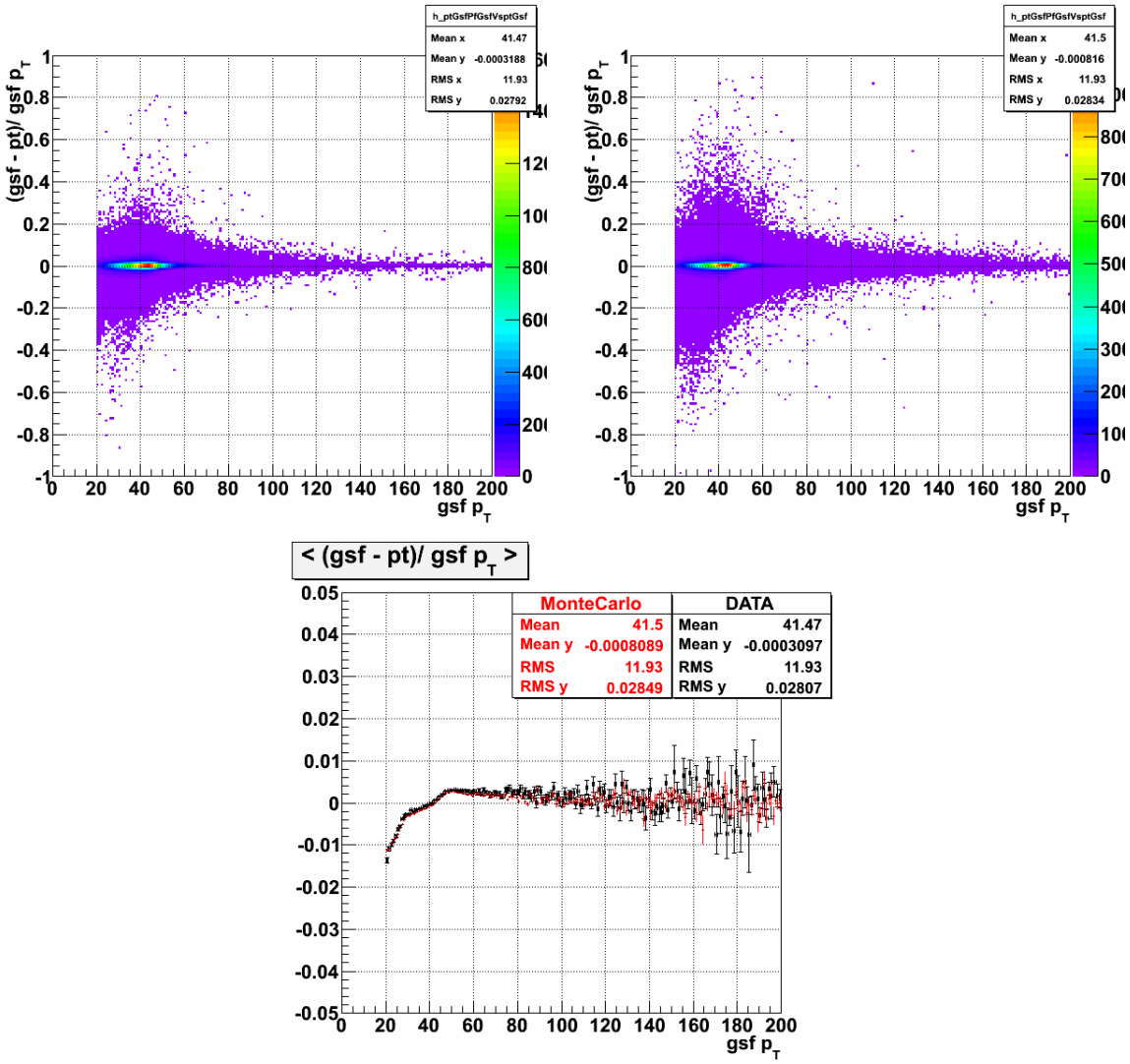


Figure 4.12: On top: the distribution of (GSF - PF)/GSF electron transverse momentum as a function of the GSF p_T for data (left) and MC (right). On bottom: comparison of the mean values for data and MC.

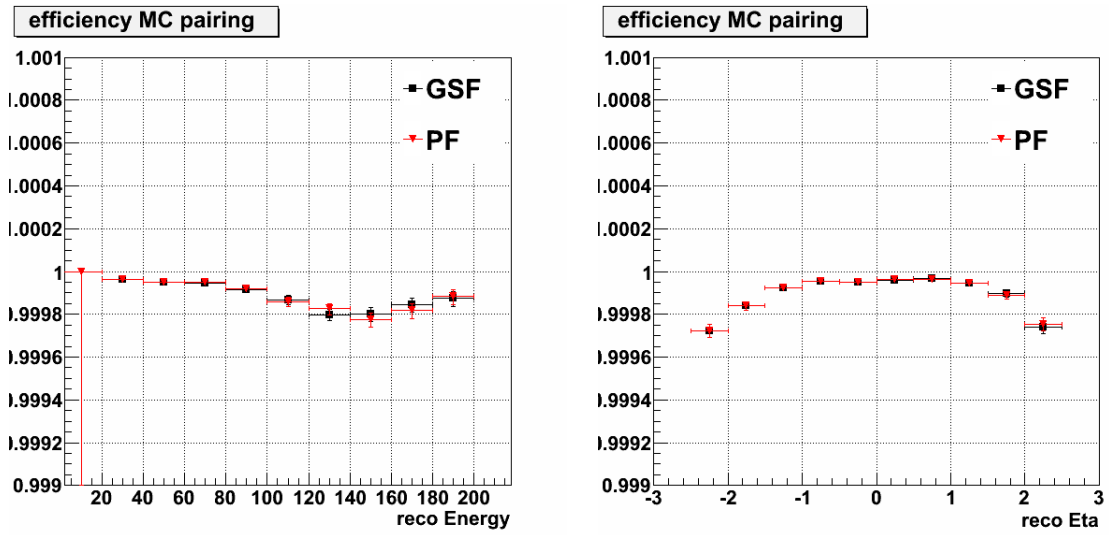


Figure 4.13: The association efficiency of *gen* and *reco* electrons as a function on the reco electron energy (left) and pseudorapidity (right). The GSF results are in black, PF results are in red.

The effects observed above can be studied in more detail by taking into account some dependencies. In Fig. 4.15 the energy ratios are shown as a function of energy and pseudorapidity. In the first case the dominant differences are at high energy, where both the PF and the GSF reconstruction overestimate the particle energy. Looking at the second type of plots, it is possible to see that the performances of the algorithms differ especially in the barrel region, where the PF reconstruction shows the influence of the ECAL cracks while the GSF one are corrected for this effect. This was seen also in the energy ratio GSF/PF in the previous subsection.

Since in the GSF/PF comparison the MC-data agreement is better in the case of the transverse momentum than for the energy, the p_T variable is also studied. The transverse momentum ratios as a function of p_T, η and ϕ are shown in Fig. 4.16. There is no evidence of discrepancies on p_T and ϕ for that variable dependency. Both the reconstruction algorithms tend to overestimate the value of the p_T for electrons with transverse momentum lower than 40 GeV. As a function of η, the PF algorithm looks more sensitive to variation in the material budget and to the presence of the calorimeter cracks, especially in the barrel region. The GSF reconstruction takes into account this effect and corrects the results, as for the energy distributions. Both the algorithms underestimate the transverse momentum of the electrons near the border region between EB and EE and overestimate it elsewhere.

Since the *gen* - *reco* association is based on the distance in the space (η, ϕ), comparisons concerning η or ϕ variables are slightly biased. In Fig. 4.17 the differences of the pseudorapidity as a function of p_T, η and ϕ are shown. There is no transverse momentum dependency on the pseudorapidity reconstruction. On the opposite, the performance depends on η itself: in the endcap region η at reco level overestimates the MC truth. There is also a slight dependence on ϕ, only of the order of 0.5 per

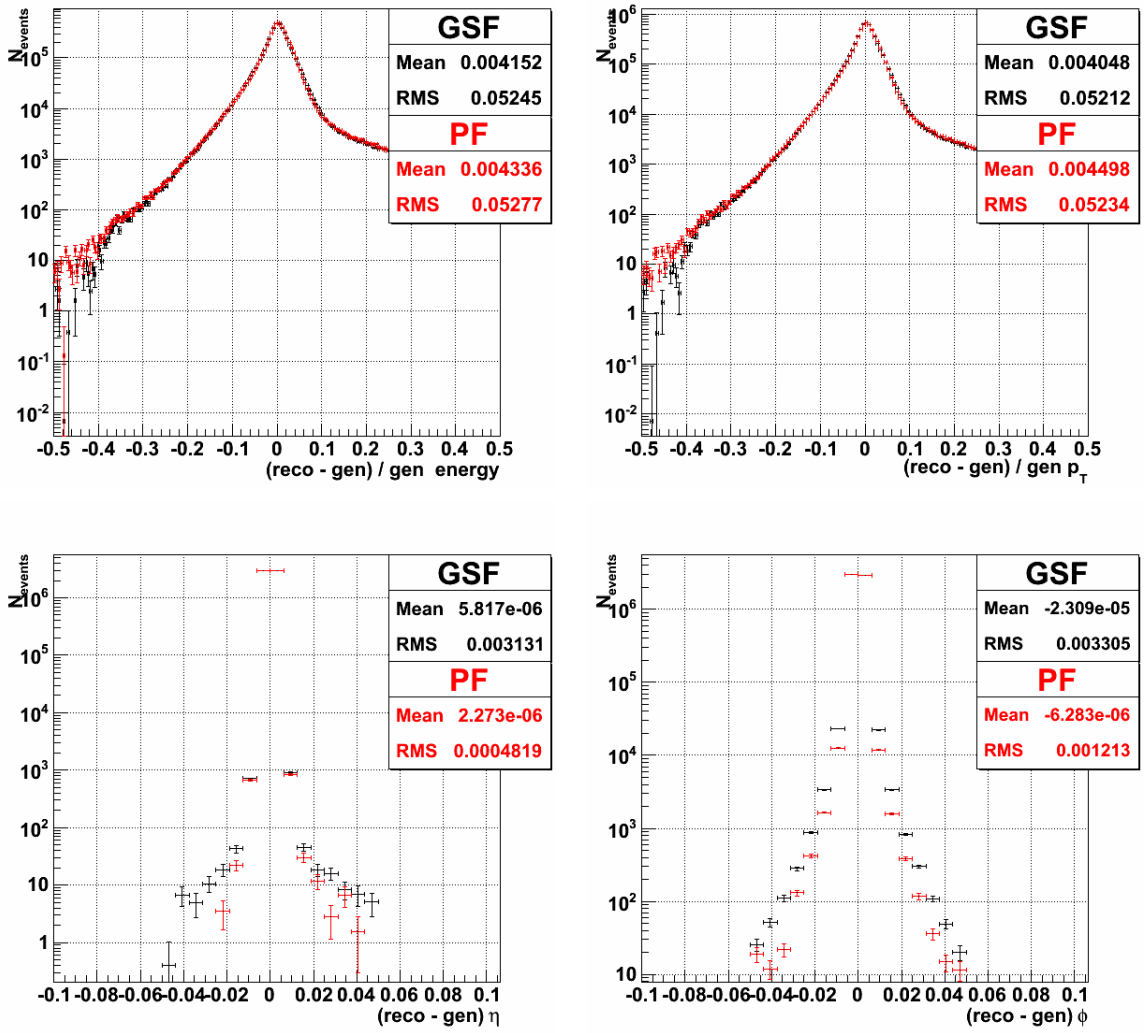


Figure 4.14: Comparison between *gen* and *reco* objects for GSF (black) and PF electrons (red). On top: ratios of energies and transverse momenta; on bottom: differences of pseudorapidities and azimuthal angles.

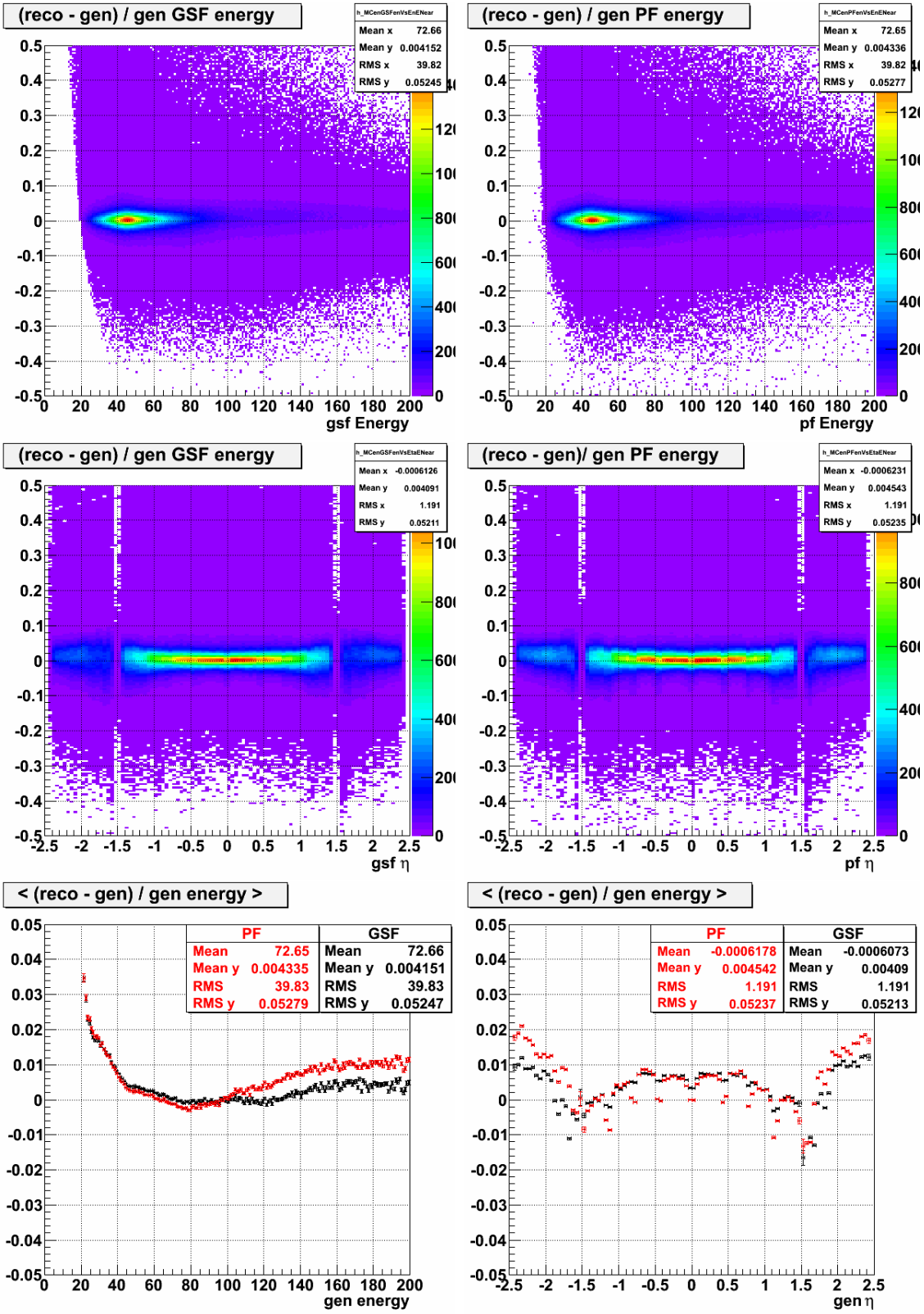


Figure 4.15: Energy comparison between *gen* and *reco* objects, for GSF (left) and PF (right) as a function of *gen* energy (first row) and η (second row). Last row: average of distributions as a function of *gen* energy and η.

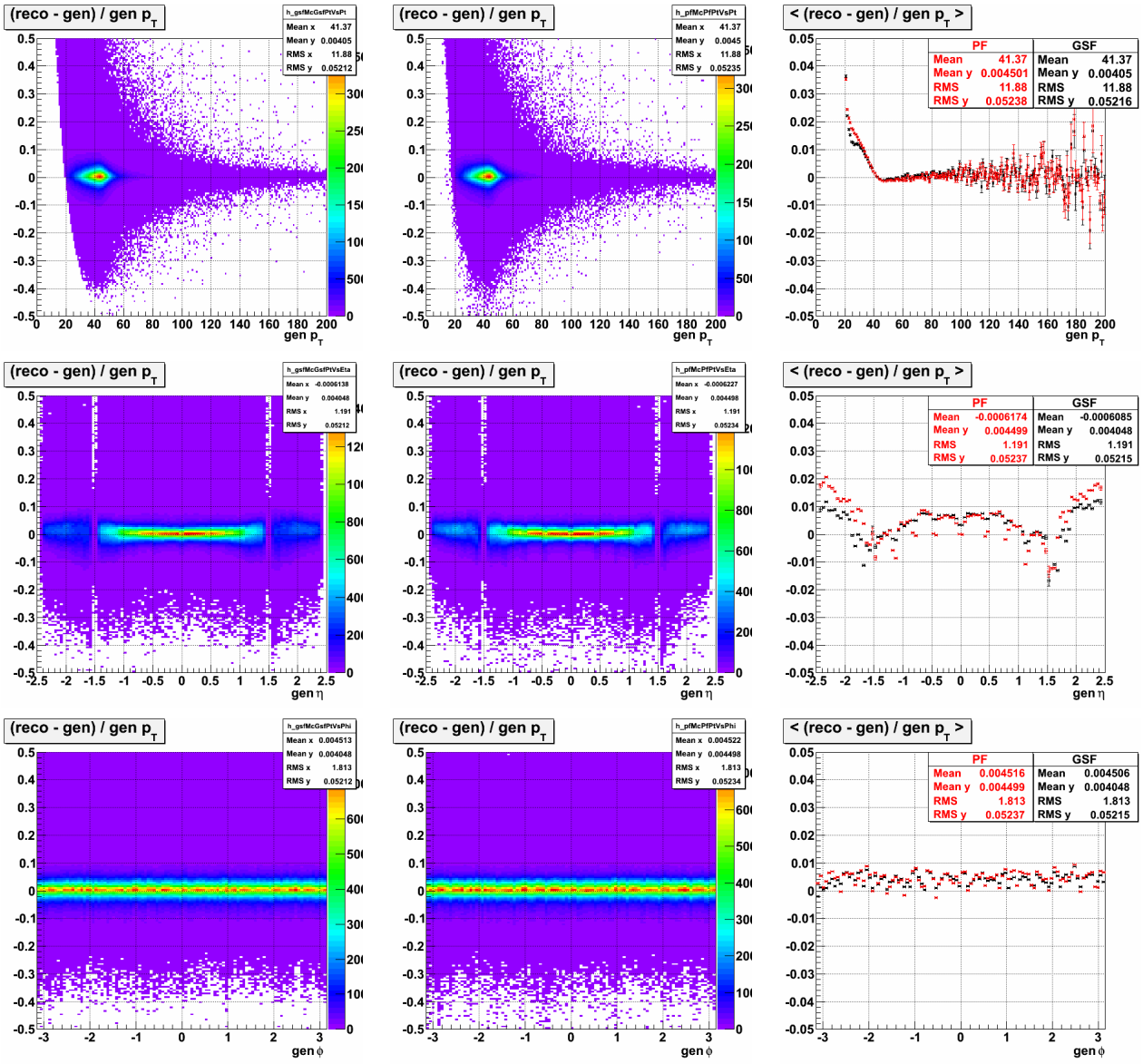


Figure 4.16: Distribution of $(p_T^{reco}$ - $p_T^{gen})/p_T^{gen}$ for GSF (1st column) and PF (2nd column) electrons, and average distributions (3rd column), as a function of gen p_T (1st row), η (2nd row) and ϕ (3rd row).

mill. The results are the same for both GSF and PF algorithms.

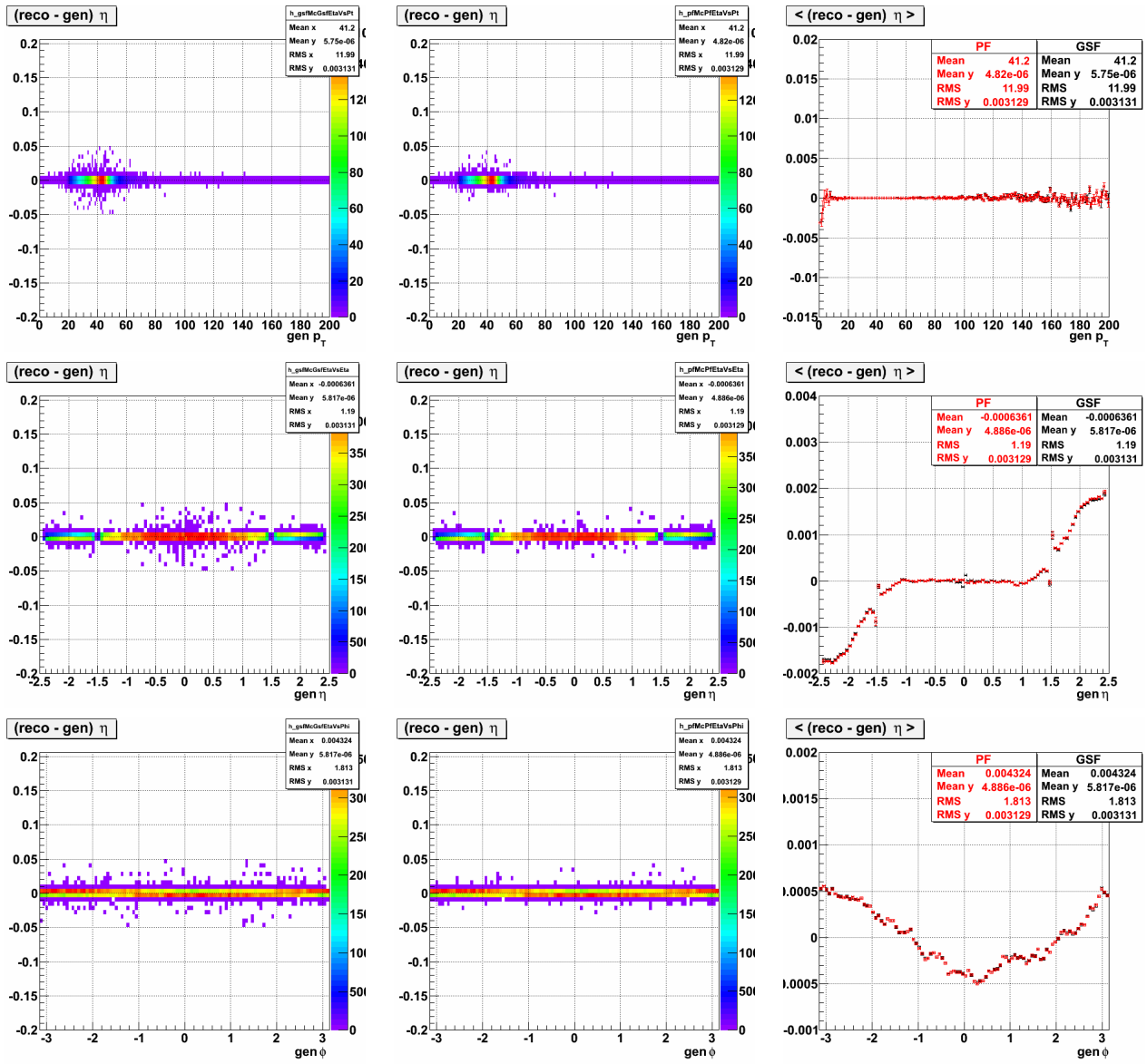


Figure 4.17: Results concerning the pseudorapidity comparison between *gen* and *reco* objects, for GSF (1st column) and PF (2nd column) electrons as function of p_T (1st row), η (2nd row) and ϕ (3rd row). On third column the average results for both algorithms are superimposed (GSF in black and PF in red).

Finally reconstruction performances on azimuthal angle are shown in Fig. 4.18. This variable is quite well reproduced. Both algorithms have similar results, they describe rather well the MC truth, without any particular dependency on the quantities taken into account: p_T, η and ϕ.

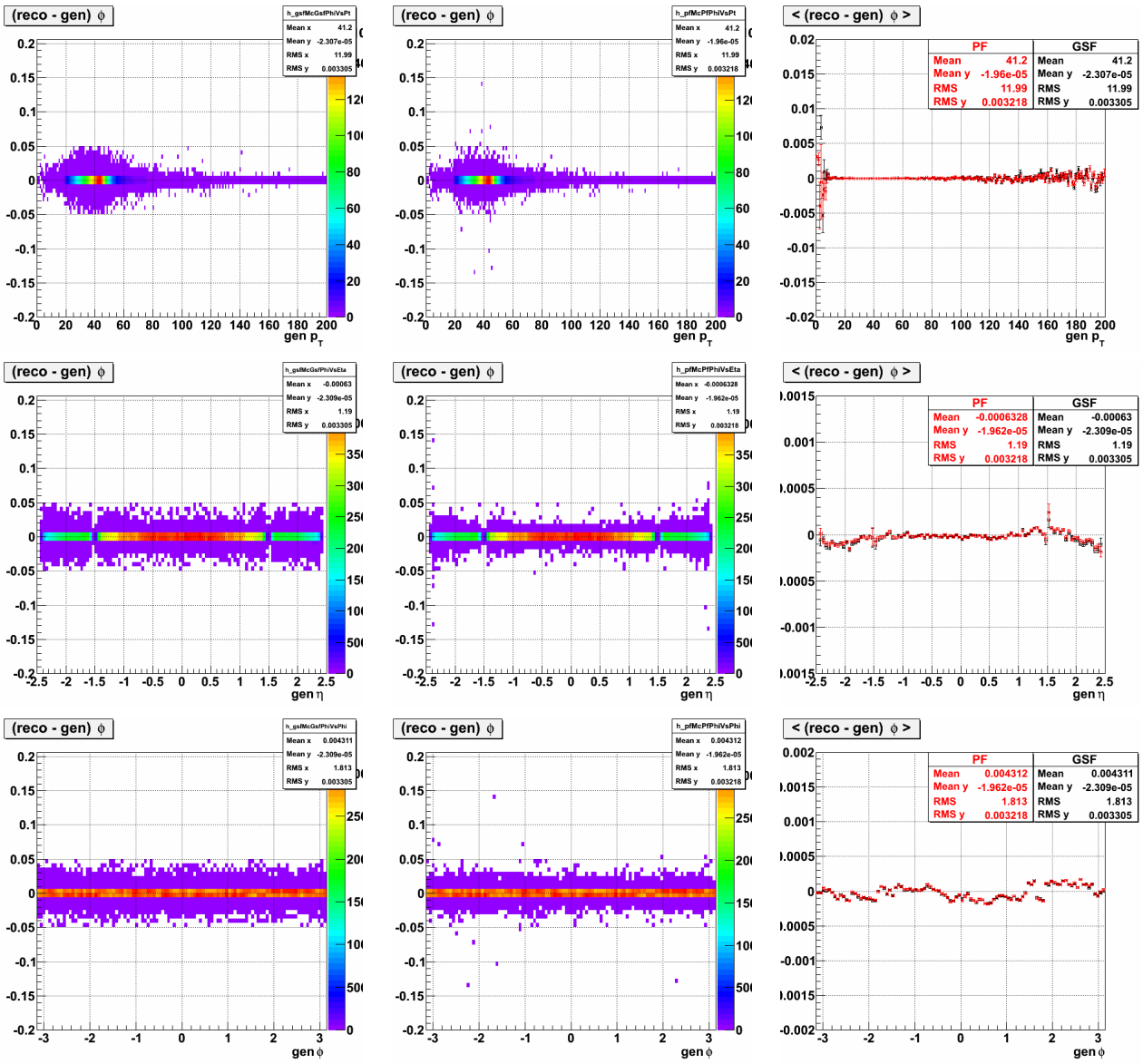


Figure 4.18: Azimuthal angle comparison between *gen* and *reco* objects, for GSF (1st column) and PF (2nd column) electrons as function of p_T (1st row), η (2nd row) and ϕ (3rd row). On third column, the average results for both algorithms are superimposed (GSF in black, PF in red).

4.3 Z mass reconstruction

After having probed the performances of the electron reconstruction the level of basic physical quantities, a test on the energy scale and resolution is done by reconstructing the Z boson invariant mass using dielectron events. Considering the Z mass reconstructed with GSF or PF electrons, the results are compared in order to understand if there are some significant differences.

In Fig. 4.19 the Z mass obtained using GSF electrons (left) and PF ones (right) is shown. The data are superimposed to the MC, considering also the most relevant sources of background for this process: QCD, $t\bar{t}$, W + jets and di-bosons (WW, WZ, ZZ) decaying into leptons (see [27]). The distribution of the background is rather flat except in correspondence to the Z boson peak, due to the ZZ and WZ samples. However the ratio S/B varies from about 10^2 to 10^3 under the peak. For this reason it is not taken into account in what follows. There are some differences between the mean values and the sigma of the distributions, but in order to obtain quantitative results a fit of the reconstructed Z line shape is needed.

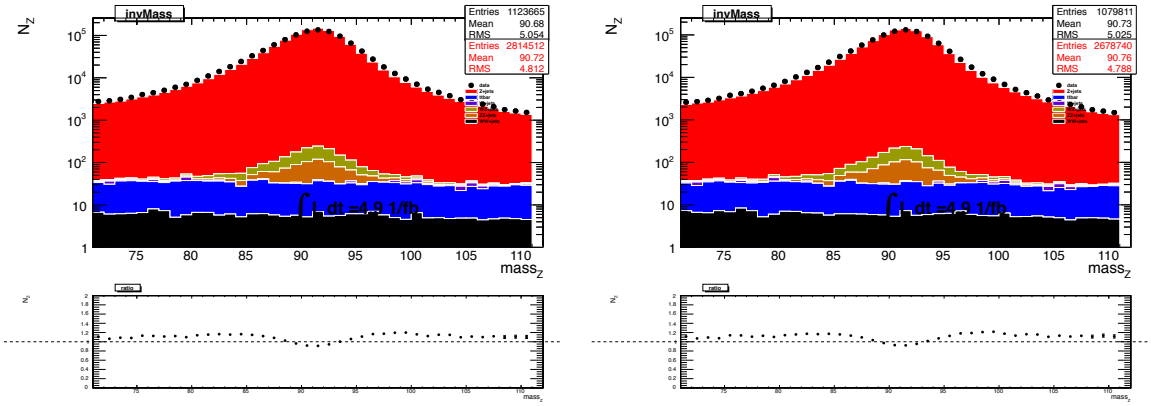


Figure 4.19: Distributions of the Z boson mass reconstructed with GSF electrons (left) and PF ones (right). The black dots represent data, the red is MC simulation of the signal. Other colors are from relevant MC simulated background.

4.3.1 Fit functions

To extract the signal from the background, two different fit functions are tested. The first is the Cruijff function (4.1), characterized by two different tails, described by σ_L and σ_R, and a mean value μ. It is associated with a exponential background (4.2). This is the function used in [27] for old "*Tag&Probe*" studies. It is an effective fast approach with some limitations that we will analyse later.

$$signal(x) = \begin{cases} \exp\left[-\dfrac{(x-\mu)^2}{\sigma_L^2 + \alpha_L (x-\mu)^2}\right], & x < \mu \\ \exp\left[-\dfrac{(x-\mu)^2}{\sigma_R^2 + \alpha_R (x-\mu)^2}\right], & x > \mu \end{cases} \tag{4.1}$$

$$bkg(x) = \exp(cx) \tag{4.2}$$

The second attempt is represented by the convolution of the Breit Wigner (BW) and Crystal Ball (CB) functions shown respectively in Eq. 4.3 and Eq. 4.4.

$$BW(x) = \frac{\Gamma_Z}{2\pi\left[(x-m_Z)^2 + (\Gamma_Z/2)^2\right]} \tag{4.3}$$

$$CB(x-\Delta m) = \begin{cases} \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\Delta m}{\sigma_{CB}}\right)^2\right], & \dfrac{x-\Delta m}{\sigma_{CB}} > -\alpha \\ \left(\dfrac{n}{\alpha}\right)^n \exp\left[-\dfrac{n^2}{2}\right]\left(\dfrac{n}{\alpha} - \alpha - \dfrac{x-\Delta m}{\sigma_{CB}}\right)^{-n}, & \dfrac{x-\Delta m}{\sigma_{CB}} \leq -\alpha \end{cases} \tag{4.4}$$

The BW function describes the physics of the process, the electronic decay of the Z boson, and has all the parameters fixed to the PDG [69] values.

$$m_Z^{PDG} = 91.1876 \text{ GeV} \qquad \Gamma_{\mathrm{Z}}^{\mathrm{PDG}} = 2.4952 \text{ GeV}$$

The CB function describes the detector effects, trying to describe the detector resolution effect. From this point of view it is a perfect probe to study the differences between the two electron reconstruction algorithms with all parameters left free. Their meaning can be summarised as follows:

- α and n define when the curve stops to behave as a Gaussian;
- Δm and σ_{CB} are the mean and the variance of the distribution in the Gaussian case.

4.3.2 Inclusive results

In order to get clear results on the deviation from the reference value of the Z boson and the differences between the performance of the two reconstruction algorithms, the distribution studied in the following is the Z mass minus the reference value m_Z^{PDG}. In this way a distribution peaked at zero should be ideally obtained in case of no bias in the Z boson line shape reconstruction. In Fig. 4.20 (Fig. 4.21) the results obtained with the Cruijff (BW⊗CB) fit are presented. In Tab. 4.1 the most interesting parameters resulting from the fit procedure are summarised.

As said previously, the Cruijff fit is a first, fast and effective approach, and indeed the resulting fits are not so good and the peak is not properly described,

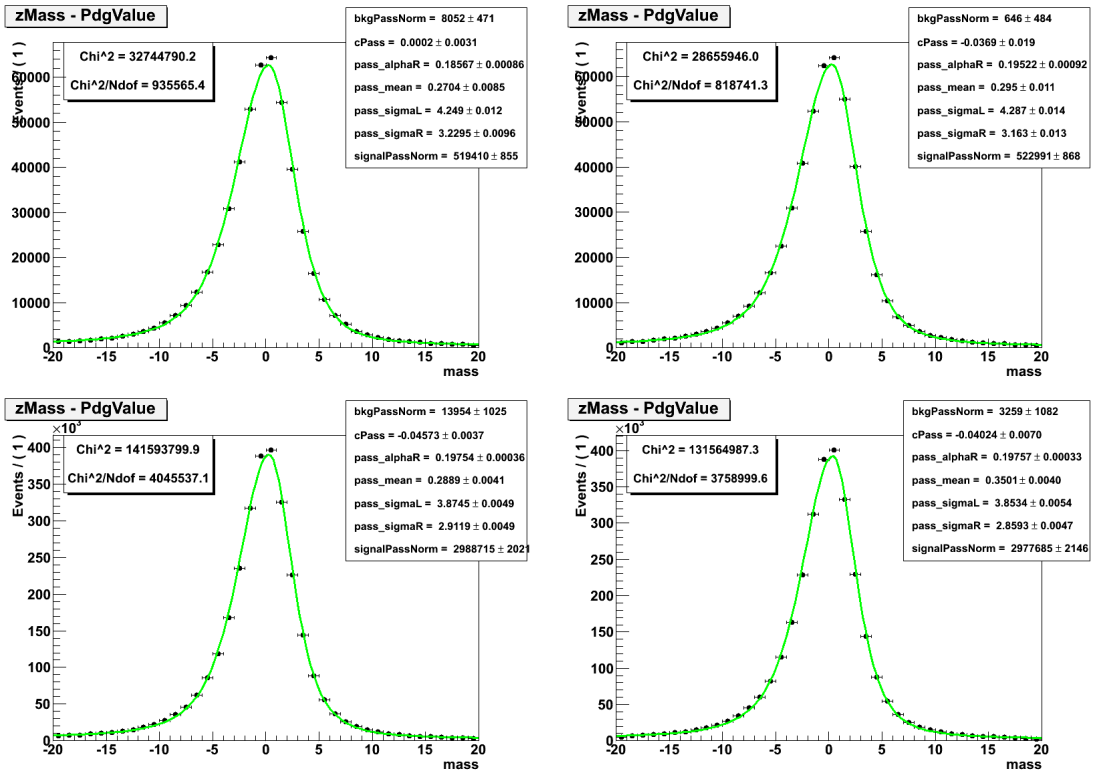


Figure 4.20: $m_Z - m_Z^{PDG}$ distribution fitted with the Cruijff function for GSF (left) and PF electrons (right). On the top there are the results obtained with the data, on the bottom with the MC.

		DATA		MonteCarlo	
		GSF	PF	GSF	PF
Cruijff	μ	0.270 ± 0.009	0.295 ± 0.011	0.289 ± 0.004	0.350 ± 0.004
	σ_L	4.249 ± 0.012	4.287 ± 0.014	3.875 ± 0.005	3.853 ± 0.005
	σ_R	3.230 ± 0.010	3.163 ± 0.013	2.912 ± 0.005	2.859 ± 0.005
BW⊗CB	Δm	0.027 ± 0.005	0.086 ± 0.007	0.047 ± 0.003	0.131 ± 0.003
	σ_{CB}	2.009 ± 0.007	1.933 ± 0.008	1.768 ± 0.003	1.686 ± 0.003

Table 4.1: The parameters obtained from the fit using the Cruijff function (on the top) and the BW convoluted with the CB (on the bottom) are summarised.

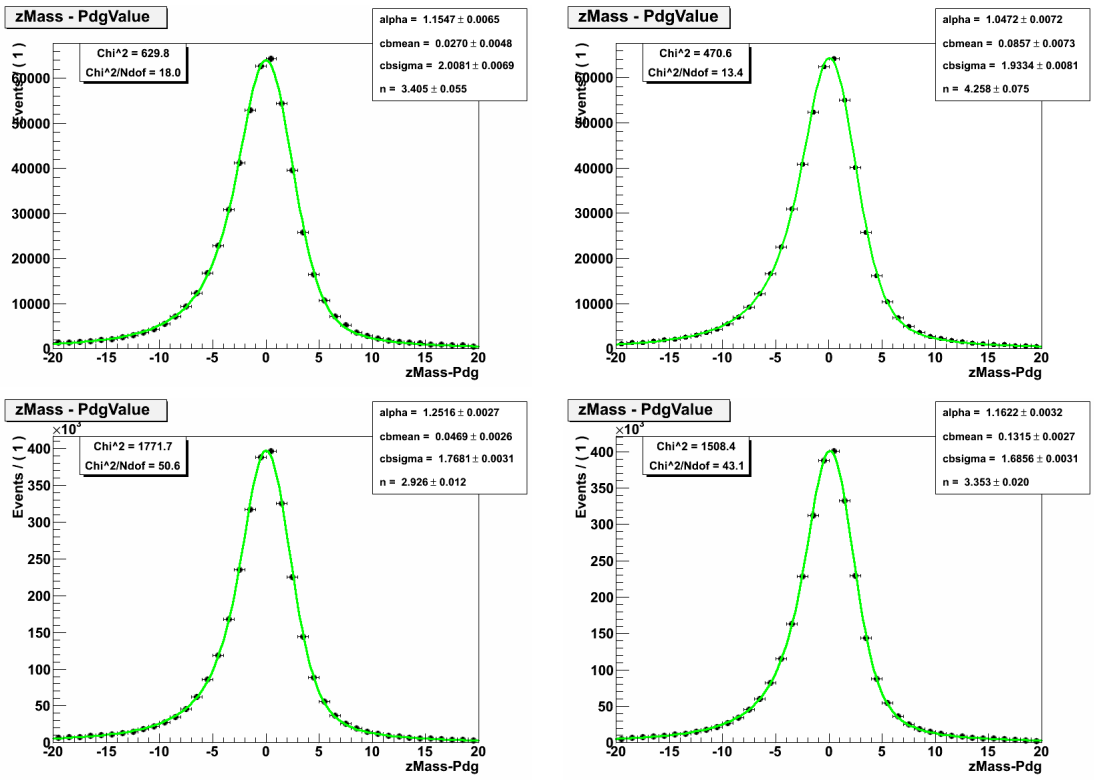


Figure 4.21: $m_Z - m_Z^{PDG}$ distribution fitted with the BW⊗CB function for GSF (left) and PF electrons (right). On the top there are the results obtained with the data, on the bottom with the MC.

both for data and MC. Looking at the mean values, both in data and in MC the GSF electrons reconstruct a Z mass closer to the reference value.

The second attempt with a more complex function works better, the BW⊗CB fit reproduces quite well the Z peak. The results are similar to the previous ones: the results closer to the reference value are the GSF ones, both in data and MC. However it is worth to note that the differences are all of the order of per mil.

In order to understand better the results obtained from the two reconstruction algorithms, a more detailed study, using the BW convoluted with the CB function, is carried out, analysing the variation of the Z mass fits as a function of the calorimeter region in which the electrons are detected and the number of jets reconstructed in association with the boson.

4.3.3 Pseudorapidity dependence

In this paragraph the reconstruction performance of the two algorithms is studied as a function of the electromagnetic calorimeter region (EB or EE) in which the electrons used to reconstruct the Z boson are detected. In Fig. 4.22 the results obtained for data and for MC using electron reconstructed in the EB are shown; the most important parameters are summarised in Tab. 4.2.

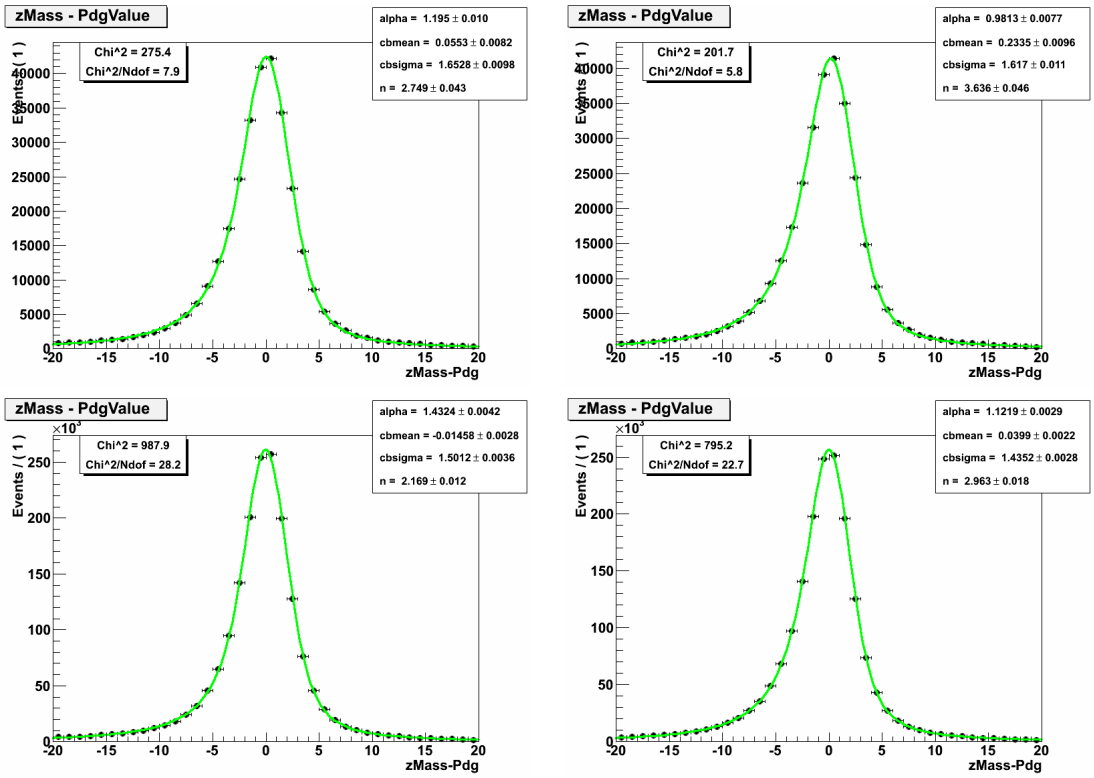


Figure 4.22: $m_Z - m_Z^{PDG}$ distribution fitted with the BW⊗CB function for GSF (left) and PF electrons (right) obtained with data (first raw) and MC (second raw) relative to reconstructed Z with both electrons in EB.

			EB-EB	EE-EE	EB - EE
DATA	GSF	Δm	0.055 ± 0.008	-0.108 ± 0.022	-0.068 ± 0.001
		σ_{CB}	1.653 ± 0.010	2.892 ± 0.027	2.544 ± 0.001
	PF	Δm	0.234 ± 0.001	-0.290 ± 0.024	-0.158 ± 0.003
		σ_{CB}	1.617 ± 0.011	2.643 ± 0.026	2.400 ± 0.005
MC	GSF	Δm	-0.015 ± 0.003	0.119 ± 0.013	0.160 ± 0.004
		σ_{CB}	1.501 ± 0.004	2.275 ± 0.013	2.138 ± 0.005
	PF	Δm	0.040 ± 0.002	0.583 ± 0.015	0.311 ± 0.005
		σ_{CB}	1.435 ± 0.003	1.968 ± 0.014	1.946 ± 0.006

Table 4.2: The parameters obtained from the fit using the BW convoluted with the CB function for different ranges of pseudorapidity are summarised.

The parameters obtained for GSF and PF electrons are slightly different both in data and MC, with discrepancies of the order of per mil that depend on the pseudorapidity: increasing from EB to EE in MC and decreasing in data. The same behaviour is observed also for differences between data and MC especially for the PF electrons. Another value influenced by the pseudo-rapidity is the resolution of the distribution. As expected, it increases with the pseudorapidity, it is best for EB-EB electrons and has its highest value for EE-EE electrons; this is observed both for data and MC. The value obtained with MC are smaller than in data and the PF resolution is better both in MC and data. In general a shift of about several hundreds of MeV is observed between data and MC, and the GSF electrons are always the nearer to the reference value.

4.3.4 Jet multiplicity dependence

In this paragraph the change of the performances of the two algorithms are studied as a function of the multiplicity of the jets reconstructed in association with the Z boson. The results concerning events reconstructed with Z boson associated to one jet (for data and MC) are shown in Fig. 4.23; the most important parameters obtained from all the considered multiplicities are summarised in Tab. 4.3.

Altogether the parameters obtained in data for GSF and PF electrons are quite in agreement, with differences of the order of per mil rather independent from the jet multiplicity. The absolute shift from the PDG value, instead, seems to increase with jets multiplicity, it starts from some tens of MeV in the zero jets case and increases to about a hundred of MeV in events with 3 jets. It is higher for PF results. The resolution is rather flat as a function of the number of jets for data; in MC it seems to decrease (except for the 4-jets case). In both data and MC results the resolution obtained with PF electrons is better with respect to the standard electron distributions.

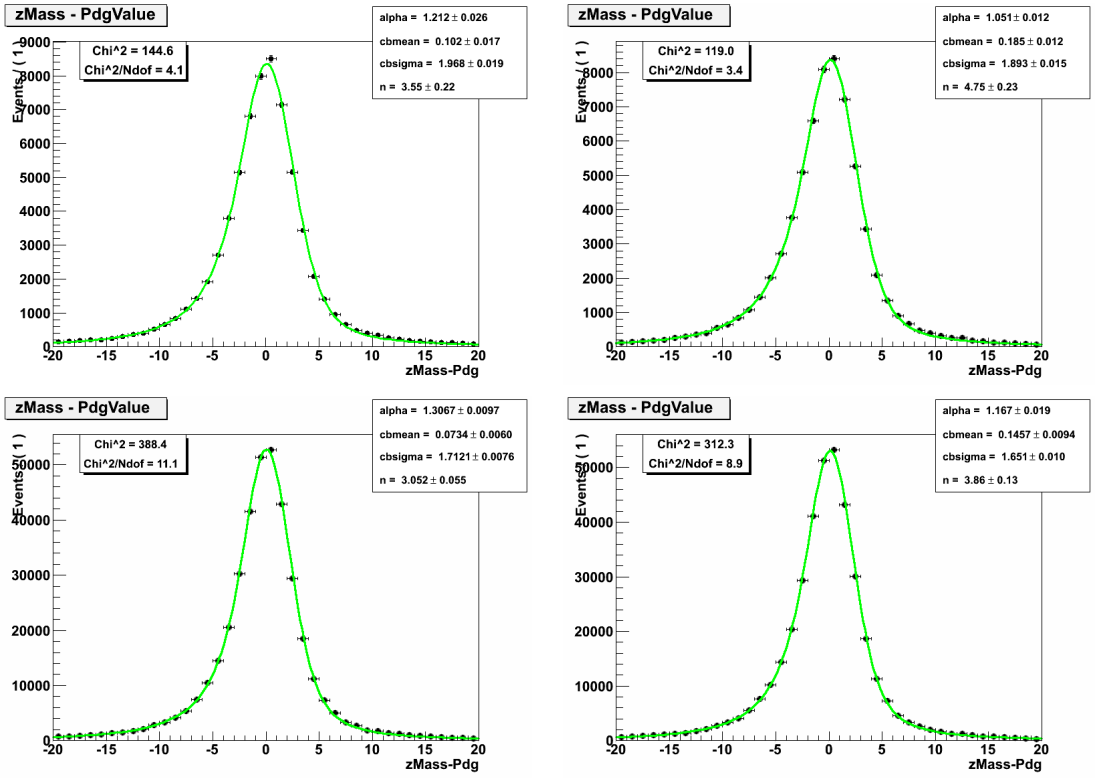


Figure 4.23: $m_Z - m_Z^{PDG}$ distribution fitted with the BW⊗CB function for GSF (left) and PF electrons (right) obtained with data (top) and MC (bottom), relative to events reconstructed with a Z boson and one jet.

			0 jets	1 jet	2 jets	3 jets	4 jets
DATA	GSF	Δm	0.012 ± 0.005	0.102 ± 0.017	0.099 ± 0.038	0.029 ± 0.075	0.25 ± 0.19
		σ_{CB}	2.014 ± 0.008	1.968 ± 0.019	2.041 ± 0.046	2.130 ± 0.090	1.72 ± 0.24
	PF	Δm	0.067 ± 0.008	0.185 ± 0.012	0.145 ± 0.040	0.075 ± 0.063	0.554 ± 0.037
		σ_{CB}	1.939 ± 0.009	1.893 ± 0.015	1.949 ± 0.048	2.127 ± 0.089	1.38 ± 0.79
MC	GSF	Δm	0.040 ± 0.003	0.073 ± 0.006	0.079 ± 0.014	0.193 ± 0.037	-0.030 ± 0.079
		σ_{CB}	1.781 ± 0.003	1.712 ± 0.008	1.682 ± 0.018	1.543 ± 0.048	1.857 ± 0.096
	PF	Δm	0.127 ± 0.003	0.146 ± 0.009	0.135 ± 0.013	0.227 ± 0.031	0.055 ± 0.068
		σ_{CB}	1.694 ± 0.004	1.651 ± 0.010	1.653 ± 0.016	1.498 ± 0.039	1.787 ± 0.083

Table 4.3: The parameters obtained from the fit using the BW convoluted with the CB function for different jet multiplicities are summarised.

4.4 Conclusions on electron reconstruction performance

In this analysis the comparison of GSF and PF electron reconstruction performance are presented, in order to understand which is the most suitable one in the context of the precision measurement described in Chapter 5. Different observables associated to the electrons are taken into account, handling the problem in steps: first the components used by the reconstruction, then the final object obtained from the algorithms and in the end the Z boson reconstructed by the two types of electrons.

Summing up the results of the first two checks, it is possible to note that the distributions of the kinematic observables of GSF and PF electrons are quite similar, but on average the GSF ones tend to be more energetic than the PF ones. The differences between the two algorithms are almost well described by the MC and the greater discrepancy is seen in the energy distributions. Taking into account the Z mass distributions produced by the two kinds of electrons, the results obtained with data and MC are rather compatible, showing differences of the order of per mil.

In conclusion the two reconstruction algorithms show slight differences in their performances. It is worthwhile to note the fact that up to now the Particle Flow reconstruction does not apply any kind of crack corrections, as GSF does. This produces absolute point-like differences of about one per cent. Even considering this effect, in the end it is confirmed that the two types of electrons have comparable performance, in fact the average discrepancies between their properties are of the order of per mil.

Chapter 5

Measurement of the jet production in association to a Z boson

In this chapter the measurement of differential cross sections of the associated production of a Z boson and jets is presented. The differential jet rate cross section is measured, as well as the transverse momentum p_T and pseudorapidity η distributions for the four highest transverse momentum jets. The distribution of the scalar sum of jet transverse momenta H_T is also measured as a function of the jet multiplicity. This study has used the 7 TeV data collected by the CMS experiment in the 2011 using jets and Z bosons detected through their decays into electron-positron pairs. The corresponding total integrated luminosity is measured to be 4.89 fb^{-1}.

Events are selected requiring the presence of a well-identified and isolated high-energy electron-positron pair[1] as signature of a Z boson decay. The jets, considered in the study, are reconstructed using a sequential clustering algorithm. All the distributions are corrected for the efficiency of the trigger, selection, reconstruction and isolation criteria calculated with the *Tag&Probe* method. The final results, corrected applying an unfolding procedure to deconvolve physics from detector effects, are presented with a direct comparison with pQCD theoretical predictions.

This chapter is structured as follows: in Sec. 5.1 details about datasets and MC sample used in this analysis are provided. In Sec. 5.2 the event selection and the electrons and jets requirements are discussed. The uncorrected distributions are presented in Sec. 5.3, while in Sec. 5.4 the efficiency measurement procedure exploiting the *Tag&Probe* method is described. The unfolding correction is presented in Sec. 5.5 and the systematic uncertainties estimation is discussed in Sec. 5.6. Finally, the corrected distributions are presented in Sec. 5.7.

[1] In the following, electron means either an electron or a positron, unless the contrary is explicitly stated.

5.1 Data and Monte Carlo samples

The analysis is performed using proton-proton collisions data collected between March and October 2011 by the CMS experiment at a center-of-mass energy of 7 TeV. The first period of data taking is called "RunA", and took place before the August technical stop. The following data taking run, "RunB" (September-October), was characterized by higher instantaneous luminosity and pile-up. The data sets and the relative run ranges are listed in Table 5.1 .

data sets	run ranges
/DoubleElectron/Run2011A-ZElectron-08Nov2011-v1/	160329-175770
/DoubleElectron/Run2011B-ZElectron-19Nov2011-v1/	175832-180252

Table 5.1: The data sets used in the analysis with the relative run ranges.

The data samples processed amount to an integrated luminosity of 4.89 fb^{-1}, estimated as described in [70].

A complete list of Monte Carlo samples used in this analysis can be found in Table 5.2. The signal is modelled with a Drell-Yan MC sample including up to four generated jets events (Z+jets), restricted to Z leptonic decays, generated with the matrix element generator `MadGraph` [30] (CTEQ6l1 parton density function [71]). The residual QCD radiation, described through a parton shower algorithm, and the hadronization which turns partons into physical particles are described by `Pythia6` [31], using the Z2 tune [72].

Electroweak background processes, including W+jets and di-bosons (WW,WZ,ZZ) decaying into leptons, as well as top-pair production, are also modelled by `MadGraph` plus `Pythia6`.

The QCD background is described by two different samples, both generated using `Pythia6`:

- the `QCD_EMenr_XtoY` sample is made of QCD events where at least one real isolated lepton or photon has been produced as part of a QCD diagram;

- the `QCD_BCtoE_XtoY` is similar to QCD_EM, with the requirement that the parent of the lepton or photon is a bottom or charm quark.

The QCD contribution has been found to be negligible, and therefore no QCD background is further considered in this analysis.

All samples have been rescaled to the NLO predictions (or NNLO-NNLL where available, see Table 5.2.),

The samples used in this analysis were processed through the CMS detector simulation obtained with `Geant4` and the trigger emulation, and reconstructed with the same sequence of algorithms used for real data.

MC sample	Cross Section (pb)	Integrated Lumi (fb^{-1})	Scaling order
DYToLL	3048	11.902	NNLO
ttbar	165	22.436	NNLL
W+Jets	31314	2.597	NNLO
ZZ	5.9	710.346	NLO
ZW	18.2	234.354	NLO
WW	43.0	98.277	NLO

Table 5.2: Monte Carlo data sets used in this analysis.

5.1.1 Pile-up simulation

The instantaneous luminosity grew systematically throughout the 2011 data taking. Under this condition, an a-priori knowledge of the pile-up (PU) final distribution was not accessible, thus MC samples could not be generated with the right PU conditions. To overcome this problem, the various samples used in this analysis have been generated according to a more inclusive distribution, and then reweighted to match the real distribution of the number of vertices measured in data.

The Monte Carlo samples have been simulated, for the number of pileup interactions, with a distribution characterized by two regions: up to a given pileup multiplicity it is flat, then it is described by a Poisson tail ("flat+Poisson"). The real distribution is quite different from the generated one: the plateau of the "flat+Poisson" distribution has been chosen to maximize the statistics in the interval [0,10], where it was expected to have the largest fraction of events. Due to the out-of-time pileup, this is not the distribution of the number of observed interactions in MC, that is represented by the previous curve convoluted with a poisson distribution at every bin. In Fig. 5.1 the "ideal" distribution and the "observed" distribution of the number of pileup interactions in MC are shown in blue and red respectively.

The same distinction occurs in data. The real number of interactions is obtained by means of the instantaneous luminosity measurement and the "observed" pileup distribution is generated filling a histogram with a poisson distribution whose mean corresponds to the expected number of interactions and whose magnitude is given by the integrated luminosity normalized by the lifetime for each lumisection. In Fig. 5.2 the "true" number of interactions (blue) is compared with the expected observed number (red) for one of the early 2011 data-taking periods.

The reweighting procedure, applied to adapt the simulated number of vertices in MC to the data, is based on the assignment to each event of a weight w_i, either greater or lower than 1. The set of weights is derived by normalizing the "observed" pileup distribution in data and MC and dividing the two distributions bin by bin:

$$w_i = \frac{N_i^{DATA}}{N_i^{MC}} \tag{5.1}$$

where N_i^{DATA} refers to the number of pileup events in the bin i in the data

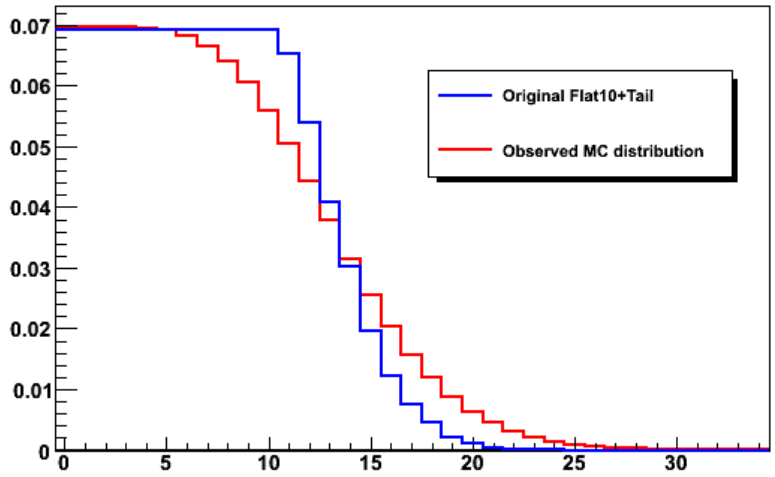


Figure 5.1: Distribution of the number of interactions in MC, the "true" one in blue and the "observed" one in red.

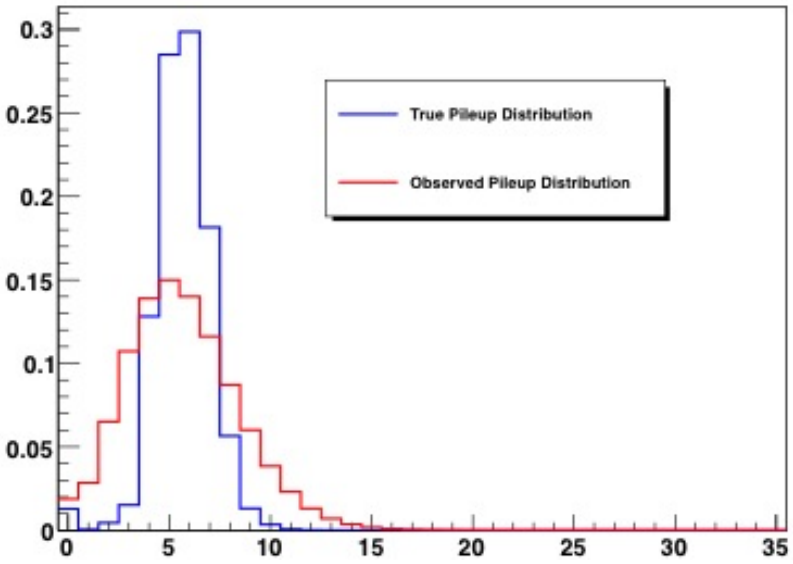


Figure 5.2: Distribution of the number of interactions in data, the "true" one in blue and the "observed" one in red.

distribution, and N_i^{MC} to the number of pileup events in MC. These weights are then applied to the simulated events by weighting each event by the weight corresponding to the number of simulated interactions.

5.2 Event selection

In this section the event selection used in the analysis is described and is developed in three steps: HLT, electron and jet selection. The HLT selection methods and paths needed to isolate event topologies of interest for the final state under study are presented. The final event selection at offline reconstruction level is fully based on the global event description provided by the particle flow algorithm. The pile-up subtraction, needed to correct the selection variables, is also discussed.

5.2.1 HLT

Triggering data at LHC is one of the most challenging experimental aspects: during the whole 2011, trigger menus were changed progressively adapting them to the higher luminosities delivered by the LHC. To extract the interesting events, a set of triggers corresponding to the lowest-threshold unprescaled double-electron trigger available are used. The HLT trigger paths used for the present analysis are:

- HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_*

- HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_
  Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*

At the HLT stage the electron candidates are built up requiring a transverse energy E_T above a certain threshold: the leading one must have E_T greater than 17 GeV, while the second one has a threshold equal to 8 GeV.

The list of cuts indicated in the trigger names after the energy of the electrons are additional constraints applied to the leptons in order to keep the rate of registered events acceptable, due to the heavy contamination from fake candidates. They represent calorimetry and tracker-based electron identification and isolation criteria. The used cuts (CaloId, CaloIso, TkId and TkIso) are summarized in Table 5.3.

5.2.2 Electrons

The selection of the analysed events is based on the search of a Z boson, reconstructed searching for a pair of well-identified and isolated high energy electrons as signature of its electronic decay, which is easily identifiable.

Electron candidates are built by combining the tracks reconstructed in the tracker and the energy deposits in ECAL, suitably clustered. As discussed in the previous chapter, the HLT trigger selection is based on the GSF electrons, but for the final

Name	Selections
CalIdL	H/E < 0.15 (0.10)
CalIdL	$\sigma_{i\eta i\eta} < 0.014$ (0.035)
CalIdT	H/E < 0.10 (0.075)
CalIdT	$\sigma_{i\eta i\eta} < 0.011$ (0.031)
CalIsoVL	ECalIso/$E_T < 0.2$ (0.2)
CalIsoVL	HCalIso/$E_T < 0.2$ (0.2)
TrkIdVL	$d\eta < 0.01$ (0.01)
TrkIdVL	$d\phi < 0.15$ (0.10)
TrkIsoVL	TrkIso/$E_T < 0.2$ (0.2)

Table 5.3: Calorimetry and tracker-based electron identification and isolation criteria applied at the HLT level.

electron reconstruction the PF algorithm is used, as it naturally combines with the Particle Flow based jet reconstruction to provide mutually independent objects in a global event description.

Reconstructed electrons are required to have a minimum transverse energy of 20 GeV: this threshold is chosen in order to be adequately far from the HLT energy thresholds (17 and 8 GeV) avoiding a possible bias. A set of cuts are used to reduce the misidentification rate for the candidate, and their values are optimized for different pseudorapidity ranges:

- ($|\eta| < 1.4442$): barrel
- ($1.566 < |\eta| < 2.4$): endcap
- $1.4442 < |\eta| < 1.566$: corresponding to the EB-EE transition region, is excluded due to decreased reconstruction efficiency.

The electron selection is based on the requirement of a 90% efficiency working point (WP90 cuts - WP loose) [73] made basically by three different selection blocks: identification, conversion rejection and isolation observables. The first block is used to reject particles misidentified as electrons. The alignment between the supercluster and the associated reconstructed track is required, checking the $\Delta\eta$ and $\Delta\phi$ variables, namely the difference between the angular coordinates from these two objects. A cut on the spread of the shower shape of the electromagnetic particles distributions along the η direction is also performed. The last observables considered in order to discriminate electron from hadrons is the energy ratio between hadronic and calorimeter deposits.

The second selection block rejects electrons from photon conversion applying a cut on the inner detector tracks consistent with a photon conversion partner near the electron, and requiring that each electron track has one hit in the innermost pixel layer.

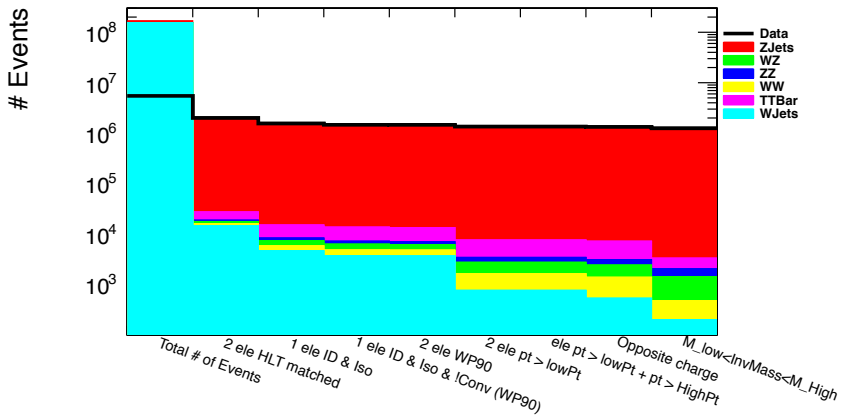


Figure 5.3: Effects of the steps of the selection over the yield of events.

Considering the last block, electron isolation criteria exploit the full Particle Flow based event reconstruction, using particles within a cone around the electron direction of $R = 0.3$. An isolation cut is defined as

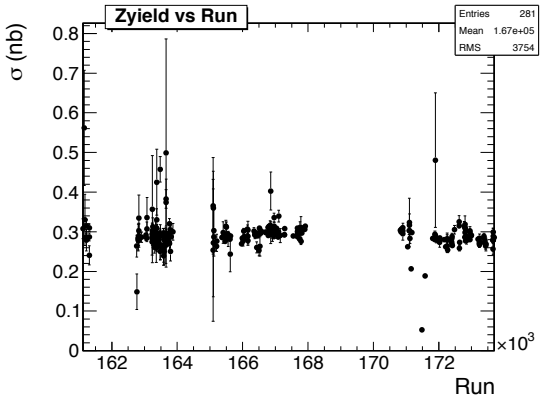
$$I_{rel} = (I_{charged} + I_{photon} + I_{neutral})/p_T^e, \tag{5.2}$$

where $I_{charged}$ is the p_T sum of all the charged hadrons, I_{photon} is the p_T sum of all the photons and $I_{neutral}$ is the p_T sum of all the neutral hadrons in the cone of interest. The value of the isolation cut is corrected for pile-up contamination as described later (see par. 5.2.4). All the cuts applied are summarized in Table 5.4.

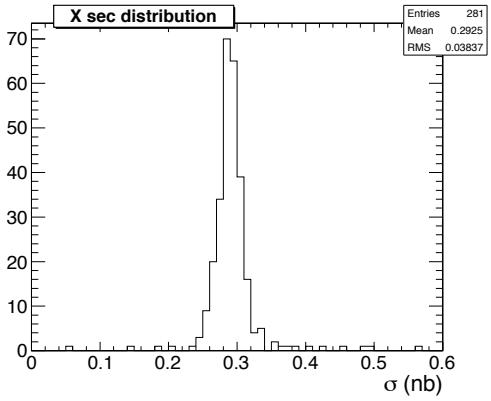
	BARREL	ENDCAP
MissingHits $\leq$	1	1
Dist $\geq$	0.02	0.02
Dcot $\geq$	0.02	0.02
$\Delta\eta <$	0.007	0.009
$\Delta\phi <$	0.8	0.7
$\sigma(i\eta i\eta) <$	0.01	0.03
H/E $<$	0.12	0.15
$I_{rel}^{PUR} <$	0.15	0.15

Table 5.4: Rejection conversions, identification and isolation cuts for the WP90 selection.

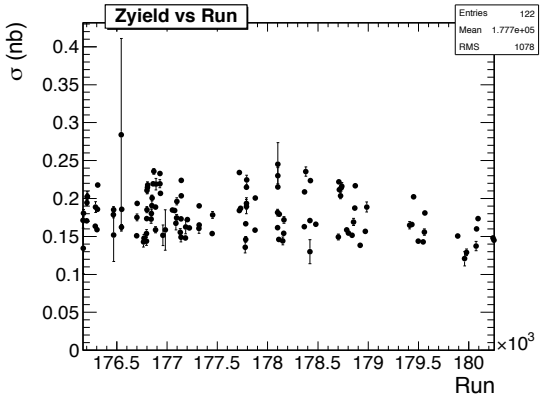
It is important to control the selections consistency during the data taking period, for example to verify that they are not influenced by the changes in the pile-up conditions. The stability of conditions in the selected data samples used has been checked by monitoring the measured Z yield normalized to the integrated luminosity as a function of the run number. Figure 5.4 shows a good stability within uncertainties over the whole time range.



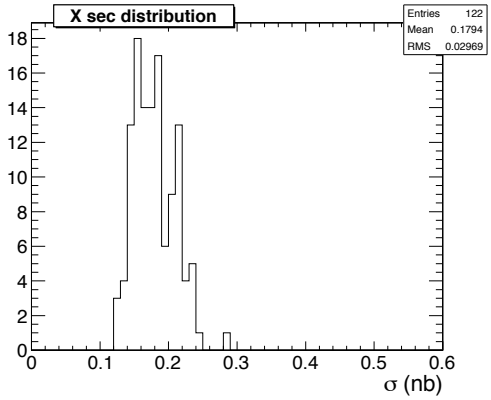
(a) Z yield normalized to the integrated luminosity per run as a function of the run number (Run A).



(b) Distribution of the uncorrected Z cross section per run (Run A).



(c) Z yield normalized to the integrated luminosity per run as a function of the run number (Run B).



(d) Distribution of the uncorrected Z cross section per run (Run B).

Figure 5.4: Z boson (e^+e^- decays) cross section per run as a function of the run number, without any efficiency correction, during Run A (a and b) and Run B (c and d). In Run B the average Z boson production cross section is smaller due to lower HLT efficiency.

In order to reconstruct the Z boson candidates, the two electrons passing the mentioned selections with the highest transverse momenta are considered. The formed pairs must be constituted by opposite charge electrons and their invariant mass M is required to lie between 71 and 111 GeV.

Table 5.5 presents a summary of the results of the selection discussed before. Fig. 5.3 summarizes the effects of the selection and represents the intermediate event yields after each step of the selection.

Cut	**Data**	**MC signal**	**ww**	**ttbar**	**wz**	**zz**	**WJets**
Total # of events	$5.4{\cdot}10^6$	$14.9{\cdot}10^6$	88998	28851	210270	767730	$153.1{\cdot}10^6$
2 e HLT matched	1997972	1898065	1670	1060	1511	8621	14512
e ID	1551086	1544135	1529	948	1179	6777	4652
e not converted	1463410	1465809	1454	902	1117	6490	3713
2 e	1461154	1464059	1360	862	1116	6370	3680
second e $p_T > 8$ GeV	1347192	1328538	1113	690	868	4159	764
first e $p_T > 17$ GeV	1347192	1328538	1113	690	868	4159	764
opposite charge	1328696	1312812	1032	670	854	4098	533
$\lvert M_{ee} - M_Z \rvert$ <20 GeV	1246130	1243567	949	614	277	1272	201
≥ 1 jet	196600	201084	68	1309	663	436	66
Total	**196900**	**201084**	**68**	**1309**	**663**	**436**	**66**

Table 5.5: Number of events selected after the list of cuts. Considering the MC samples, no scaling factor is applying to correct for data, only luminosity weighting exploiting the theoretical cross sections reported in Tab. 5.2.

5.2.3 Jets

In order to have a mutually exclusive reconstruction of electrons and jets, particle candidates reconstructed by the Particle Flow algorithm are clustered after removing from the particle collection all the electrons passing all the selections (see Sec. 5.2.2) but the transverse momentum and invariant mass cuts. In this way no fake jet corresponding to leptons is built, avoiding to erroneously attach to it soft particles from other jets and the underlying event (see Sec. 3.2.2).

The jets are clustered using the anti-k_T algorithm [62] with a resolution parameter $R = 0.5$. In order to ensure the best quality of the tracking information, only jets with a pseudorapidity $|\eta| < 2.4$, i.e. within the silicon tracker acceptance, are selected. A minimum threshold on the jet transverse momentum of $p_T > 30$ GeV is required, in order to reduce the contamination from the underlying event to the final state studied and to ensure a good reconstruction. In order to provide good jet quality selection and good noise jet rejection, several cuts on the jet composition are applied: on the number of jet constituents, on the fractions of different particle types (neutral and charged hadrons and electromagnetic particles) with respect to the total number of jet constituents and on the amount of charged particles in the jet (charged multiplicity). They are reported in Tab. 5.6.

Neutral Hadron Fraction	< 0.99
Neutral Electromagnetic Fraction	< 0.99
Number of constituents	> 1
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged Electromagnetic Fraction	< 0.99

Table 5.6: Jet identification criteria used in the analysis.

5.2.4 Pile-up removal

An important phenomenon to consider in the analysis, which can contaminate the obtained results, is the pile-up, that is generated by overlapping minimum-bias events coming from the same (in-time pile-up) and the previous (out-of-time pile-up) bunch crossing, and the underlying event. It is the most problematic effect that influences the selection efficiency especially in high luminosity conditions. In order to take into account the high pile-up scenario at LHC, with an average of 10 additionally vertices (2011 Run), the selection variables that are more sensitive to the pile-up effect require to be corrected. Considering the jets, this is performed by the offset corrections discussed in the Sec. 3.2.3. For the electrons, the identification and the conversion rejection variables have really little contamination, while the highest effect is present in the isolation variable (last line in Table 5.4) that is the result of the sum of all the contributions in a specific area, considering photons, neutral and charged hadrons.

Taking into account charged particles reconstructed with the PF algorithm, it is possible to identify the vertex they are originated from and discard the particles coming from secondary vertices. It is worth to note that for photons and neutral hadrons isolation variable this procedure does not work. Hence the corresponding isolation variable needs a different pileup correction.

This effect is taken into account using the "RhoFastJet method" [65] by evaluating an event energy density to the hard interaction activity, as $\rho = median(p_{jet}/Area_{jet})$, assumed to be a constant in the event. The quantity $\rho \cdot A_{eff}$ is subtracted from the photon and the neutral isolation variables, in order to correct the energy deposits around the electron, as shown in eq. 5.3:

$$I_{corr} = (I_{charged} + \max(0, I_{neutral} - \rho \cdot A^{eff}_{neutral}) + \max(0, I_{photon} - \rho \cdot A^{eff}_{photon}))/p^e_T \quad (5.3)$$

where $I_{charged}$, $I_{neutral}$ and I_{photon} represent respectively the sum of the p_T of all the charged hadrons, neutral hadrons and photons in a cone of radius $R = 0.3$ around the selected electron, and p^e_T is the transverse momentum of the electron.

The values of A_{eff} used in the correction for different pseudorapidity regions are reported in Table 5.7.

η	≤ 1	≤ 1.479	≤ 2.0	≤ 2.2	≤ 2.3	≤ 2.4
photon	0.081	0.084	0.048	0.089	0.092	0.097
neutral hadrons	0.024	0.037	0.037	0.023	0.023	0.021

Table 5.7: Values of A_{eff} used for the pile-up effect correction.

5.3 Uncorrected distributions

The physical observables presented in this analysis are differential cross sections measured as a function of:

1. exclusive jet multiplicity;
2. transverse momentum and pseudorapidity of jets, up to the fourth highest p_T jet;
3. H_T, defined as the scalar sum of the jet transverse momenta, up to the fourth highest p_T jet, in events with at least N jets ($1 \leq N \leq 4$).

In this section the uncorrected distributions for those observables are presented, for events passing the selection criteria described in the previous section. The results are not deconvolved from the detector effects and no efficiency correction is applied. A comparison to the predictions from simulation, including both signal and background estimates (see Sec. 5.1), is provided.

The measurement of the inclusive jet multiplicity distribution is shown in Fig. 5.5. The agreement between data and Monte Carlo prediction is very good, especially up to 4 jets, where the theoretical calculation based on the Leading Order matrix element up to 4 partons ensures an accurate description of the p_T structure, and the collected data statistics is relatively large. In the studied data set up to nine jets can be observed in an event.

Differential cross sections as for the 4 highest p_T jets are shown, as a function of the p_T itself ($\frac{d\sigma}{dp_T^{jet}}$) in Fig. 5.6, as a function of η, ($\frac{d\sigma}{d\eta}$) in Fig. 5.7, and as a function of H_T, $\frac{d\sigma}{dH_T}$, in Fig. 5.8.

In order to reduce the sensitivity of these measurements to external inputs like the luminosity, the differential cross sections are normalized to the total number of Z boson events with at least N jets associated, where N represents the number of jets in the event. In this way the error related to the LHC luminosity measurement vanishes in the ratio.

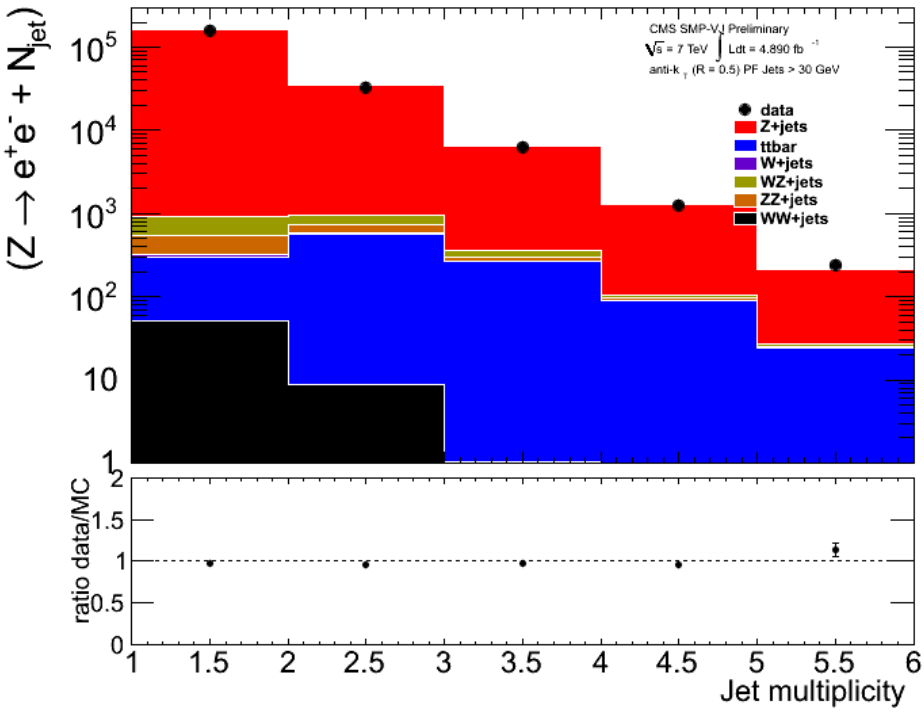
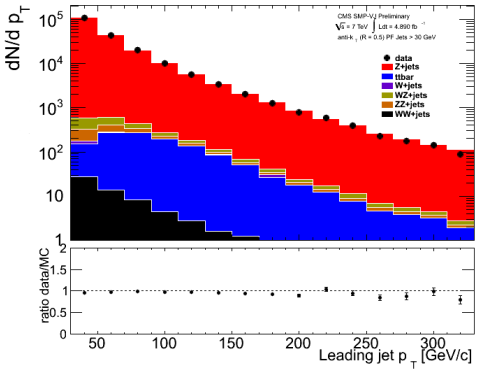
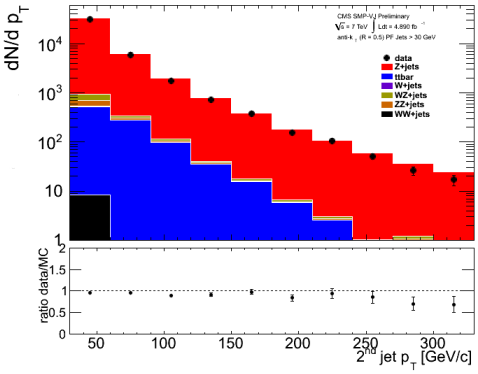


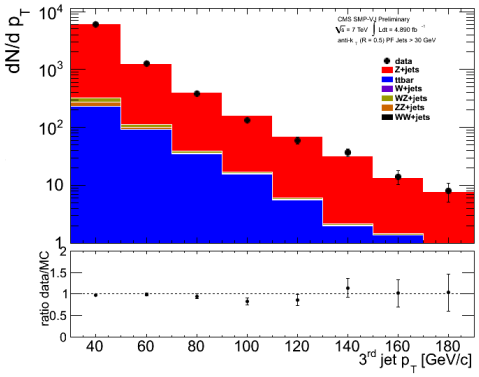
Figure 5.5: Inclusive jet multiplicity distribution (uncorrected distribution).



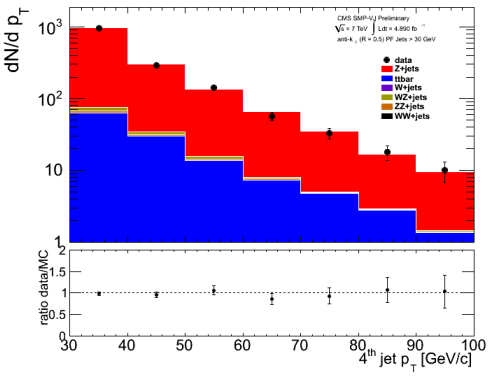
(a) leading jet p_T distribution.



(b) second leading jet p_T distribution.

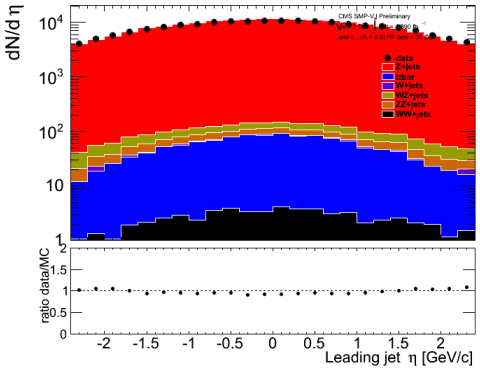


(c) third leading jet p_T distribution.

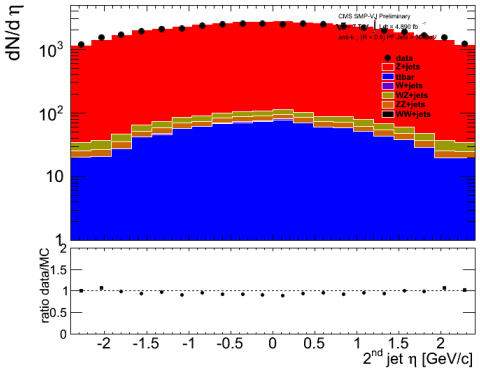


(d) fourth leading jet p_T distribution.

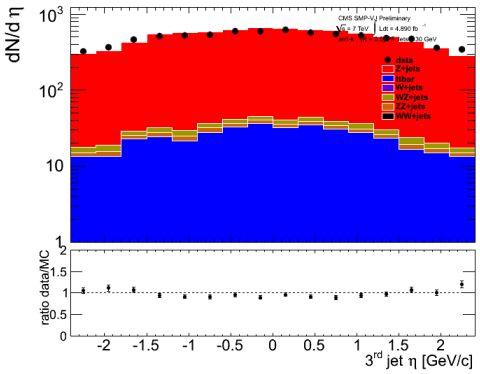
Figure 5.6: Jet p_T (for the 4 highest p_T jets) data distribution compared to the MC prediction (uncorrected distributions).



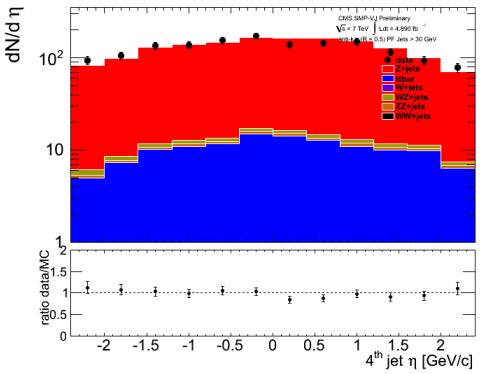
(a) leading jet eta distribution.



(b) second leading jet eta distribution.

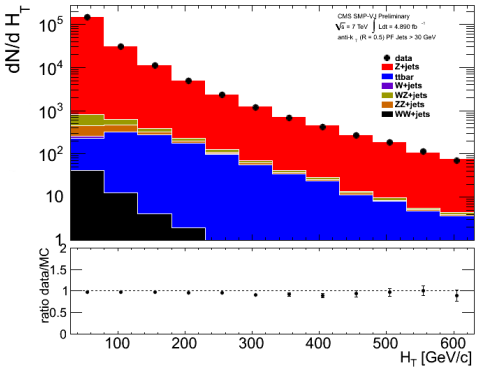


(c) third leading jet eta distribution.

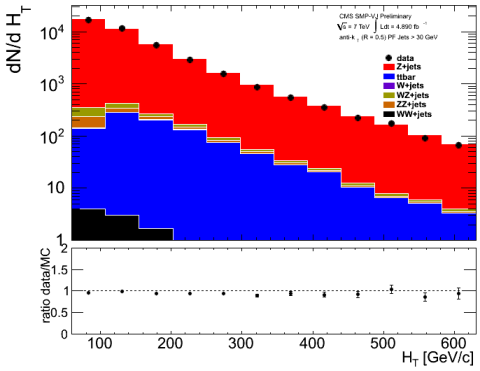


(d) fourth leading jet eta distribution.

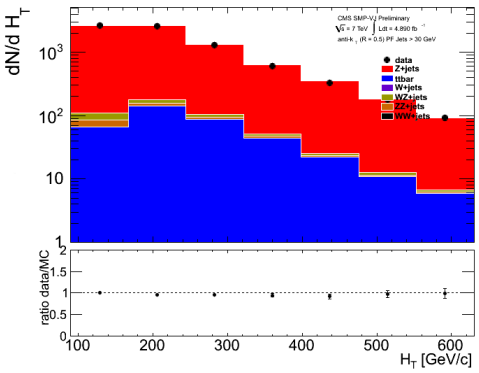
Figure 5.7: Jet η (for the 4 highest p_T jets) data distribution compared to the MC prediction (uncorrected distributions).



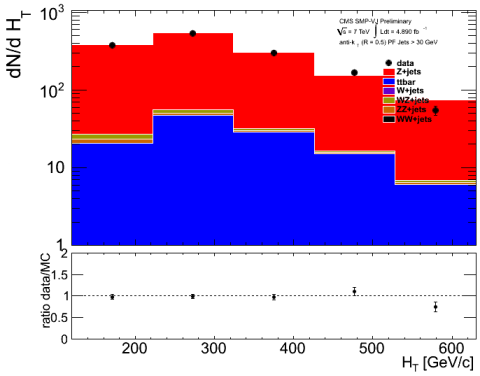
(a) H_T distribution with at least one jet



(b) H_T distribution with at least two jets



(c) H_T distribution with at least three jets



(d) H_T distribution with at least four jets

Figure 5.8: Jet H_T (for the 4 highest p_T jets) data distribution compared to the MC prediction (uncorrected distributions).

5.4 Efficiency evaluation

In order to get the correct measurement of the Z $\rightarrow ee$ + jets differential cross sections, it is necessary to accurately estimate the detection efficiencies of the decay products of the Z bosons used for the event selection. The *Tag&Probe* method [74] is a data driven procedure based on the selection of a well-known resonance, which exploits its decay in the pair of particles that are under study. It has been used to estimate efficiencies both on real data and simulated data in order to extract the data-to-simulation scale factors. They provide corrections for differences between data and Monte Carlo, in order to subtract the background considering the Monte Carlo predictions.

5.4.1 *Tag&Probe* method

In this particular analysis the invariant mass distribution of the two selected electrons around the Z mass peak is used to extract the signal and background components by means of a binned extended Maximum-Likelihood fit. Tight offline selection criteria are required on one of the two leptons, the tag, and loose cuts are applied on the other one, the probe.

The signal part of the distribution is modelled on top of the signal Monte Carlo sample, considering the simulated electrons which are direct decay products of a Z boson. Such a sample is obtained by matching electron candidates to the Monte Carlo truth information. The probability density function template taken from the Monte Carlo signal distribution is then convoluted with a Gaussian function: its purpose is to take into account statistical fluctuations in the Monte Carlo sample and possible small differences between the smearing effects of the detector and its simulation.

The background fraction of the global distribution is modelled by an exponential function multiplied by an error function describing the kinematic cut-off due to the binning on the probe lepton transverse momentum.

The two tag and probe electrons are chosen in a random and unbiased way among the selected Particle Flow candidates that lie in the $[60, 120]$ GeV invariant mass window. Moreover the tag lepton should be matched to a single electron trigger in order to minimize the fraction of background events, without introducing any bias. No unprescaled single-electron trigger paths with a useful p_T threshold are available for the entire 2011 dataset. Therefore cross triggers, characterized by cuts on more than one physics object with lower thresholds and smaller bias on the loose leg than double triggers, that require in particular an electron and a raw ECAL SC, are used:

- HLT_Ele32_CaloIdL_CaloIsoVL_SC17
- HLT_Ele32_CaloIdT_CaloIsoT_TrkIdT_TrkIsoT_SC17.

Two different distributions are filled when the probe electron passes or fails the efficiency test. Both the distributions are simultaneously fitted in the $[60, 120]$ GeV invariant mass window in order to compute the efficiency.

The overall electron efficiency can be divided into three main contributions associated to different processes: the trigger selection, the reconstruction and the off-line selection. The global efficiency corresponds to the product of the three contributions. For every step of the whole chain, the efficiency has been measured as a function of the electron transverse momentum and pseudorapidity, with the following binning scheme:

- p_T (GeV): $[20, 30]$, $[30, 40]$, $[40, 50]$, $[50, \text{inf}]$.
- $|\eta|$: $[0, 0.8]$, $[0.8, 1.4442]$, $[1.566, 2.0]$, $[2.0, 2.4]$

In the following sections, the measurement of each contribution will be described separately.

5.4.2 Fitting procedure validation

The fitting procedure has been validated by comparing the fitted efficiencies from the inclusive Monte Carlo dataset (non Drell-Yan backgrounds included) with the exact efficiencies obtained by counting the events with simulated electron from a Z boson decay in the Drell-Yan sample. The result, expressed in terms of ratios of these two efficiency evaluations and shown in Fig. 5.9, shows a good agreement between the values obtained through the two methods.

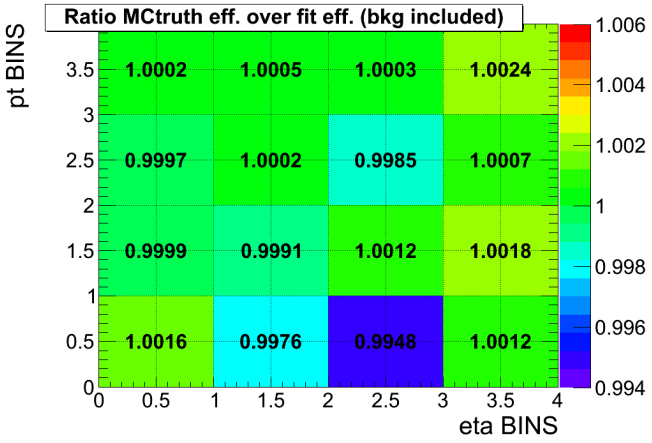


Figure 5.9: Ratio between fitted and exact Monte Carlo efficiencies.

5.4.3 HLT efficiency

The asymmetric cross trigger paths used to select the events have a very similar structure, thus their efficiency (ϵ_{HLT}) has been measured at the same time. The electron trigger paths are built in such a way for every event to contain at least one electron satisfying a set of tight isolation and identification requirements and at least one electron satisfying a set of loose requirements.

The efficiency of the tight leg of the electron trigger ϵ_{tight} is evaluated as a function of the electron η and p_T. The small fraction of electrons that are tagged as loose trigger legs but not as tight ones ($\epsilon_{loose!tight}$) is estimated. The overall trigger efficiency is computed as follows:

$$\begin{aligned}\epsilon_{HLT} = &\epsilon_{tight}(p_T(1),\eta(1)) * \epsilon_{tight}(p_T(2),\eta(2)) \\ &+ \epsilon_{tight}(p_T(1),\eta(1)) * \epsilon_{loose!tight}(p_T(2),\eta(2)) \\ &+ \epsilon_{loose!tight}(p_T(1),\eta(1)) * \epsilon_{tight}(p_T(2),\eta(2)),\end{aligned} \tag{5.4}$$

where $p_T(i)$ and $\eta(i)$ are respectively the transverse momentum and pseudorapidity of the electron i.

5.4.4 Reconstruction efficiency

The reconstruction efficiency (ϵ_{RECO}) has been measured over a set of probes chosen in such a way to introduce the least possible bias into the efficiency measurement itself. The probes are randomly selected among the collection of ECAL SCs in the acceptance with $E_T > 20$ GeV. Since no single-electron unprescaled trigger is available for the whole analysed data taking period, a cross trigger with the loosest possible criteria over the second leg, corresponding to an ECAL SC with $E_T > 17$ GeV, and with the tag electron matched to the tight leg of this trigger is used in order to avoid biases in the probe sample.

The reconstruction efficiency for every event is then computed as follows:

$$\epsilon_{RECO} = \epsilon_{RECO}(p_T(1),\eta(1)) * \epsilon_{RECO}(p_T(2),\eta(2)). \tag{5.5}$$

5.4.5 Selection efficiency

The offline selection efficiency (ϵ_{WP90}) is evaluated with a procedure similar to the previous ones. The probe electrons are chosen among the standard reconstructed electrons with $p_T > 20$ GeV in ECAL acceptance. The offline selection efficiency for every event is computed with the same formula as for the reconstruction efficiency:

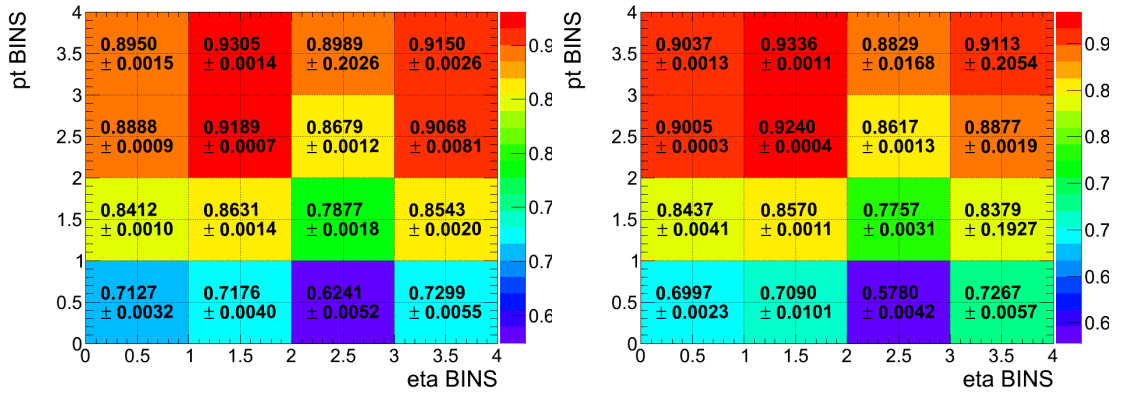
$$\epsilon_{WP90} = \epsilon_{WP90}(p_T(1),\eta(1)) * \epsilon_{WP90}(p_T(2),\eta(2)) \tag{5.6}$$

5.4.6 Final efficiencies

The product of each part (HLT, RECO, WP90) contributes to the global efficiency as a weight assigned to each event as a function of the selected electron pair η and p_T:

$$\epsilon_{global}(p_T(1),\eta(1),p_T(2),\eta(2)) = \epsilon_{HLT} * \epsilon_{RECO} * \epsilon_{WP90} \tag{5.7}$$

The global efficiencies for electrons have been measured both on the data and on the inclusive MC sample and are shown in Fig. 5.10.



(a) data (b) MC

Figure 5.10: The global efficiencies for electrons, calculated as a function of electron η and p_T, for MC and data.

5.4.7 Scaling factors and background removal

The comparison between data and simulation efficiencies allows to derive scaling factors needed to correct the simulated samples in order for them to be consistent with data. Thanks to these corrections it is possible to use the Monte Carlo predictions to subtract the background processes contribution to the event yield for the cross section measurement. These scaling factors are defined as follows:

$$w_{data/MC}(p_T(1),\eta(1),p_T(2),\eta(2)) = \frac{\epsilon^{DATA}_{global}(p_T(1),\eta(1),p_T(2),\eta(2))}{\epsilon^{MC}_{global}(p_T(1),\eta(1),p_T(2),\eta(2))} \tag{5.8}$$

The contamination from backgrounds in the Z+jets sample is small and has both a resonant (WZ, ZZ) and a non resonant ($t\bar{t}$, WW, non peaking Drell-Yan, QCD) component. Both the contributions are subtracted from the data yield exploiting the pure Monte Carlo prediction. The subtraction procedure is performed scaling the number of background events to the data integrated luminosity by means of the corresponding cross section for each background process. The cross section values of each considered process can be found in the Table 5.2. Finally, the corresponding number of background events from the rescaled Monte Carlo sample is subtracted from every bin of the studied observables.

5.5 Unfolding

In order to compare the results with any theoretical prediction for the measured observables, the distributions presented in Sec. 5.3 are corrected for the efficiencies discussed in Sec. 5.4 and for experimental resolution effects. This is done using well known unfolding techniques, allowing to obtain distributions at particle level that

reflect the event configuration before the interaction with the detector. The exploited methods are based on the two most robust mathematical algorithms available:

1. The first method is based on the singular value decomposition (SVD) of the response matrix [75] exploiting a linear regularization algorithm through a parameter k_{SVD}.
   The SVD of a real matrix is given by its factorization in the form $A = U \cdot S \cdot V^T$ where U and V are orthogonal matrices and S is a diagonal matrix with non-negative diagonal elements. If the vector x represents the true distribution values and $\hat{A}$ is the response matrix that describes the detector effects, the measured distribution values y can be expressed as $y = \hat{A} \cdot x$. In order to trace back to the particle level distribution, the inversion of the matrix is required. Using the SVD method this step is reduced to the solution of a system of linear equation with $\hat{A}^{-1} = V \cdot S^{-1} \cdot U^T$.
   The important problem concerning the stability of the solutions arises in real situations dealing with measurements and their uncertainties. The exact solutions usually lead to rapidly oscillating distributions and this oscillatory component should be suppressed exploiting the a-priori knowledge of the solutions: they should be smooth with small bin-to-bin variation. This regularization procedure is performed using a smooth cut-off on singular value contributions ($s_i^2 \to s_i^2/(s_i^2+s_k^2)$, where the k-th singular value, defining the cut-off, is chosen evaluating the oscillation of the solutions).

2. The Bayesian method is based on the Bayes theorem [76]. If the true generated variable values are called cause C_i ($i = 1, ..., n_C$) and the observed variable values effect E_j ($j = 1, ..., n_E$), thus the probability $P(E_j|C_i)$ that having a certain generated value C_i , the observed one will be E_j is known as the migration matrix. Exploiting the Bayes' theorem, it is possible to calculate the conditional probability $P(C_i|E_j)$, called smearing matrix, that an observed value E_j is obtained starting from a generated one C_i, using a certain hypothesis $P_0(C_i)$ for the true distribution. The smearing matrix is expressed as in the following:

$$P(C_i|E_j) = \frac{P(E_j|C_i)P_0(C_i)}{\sum_{k=1}^{n_C} P(E_j|C_k)P_0(C_k)}. \tag{5.9}$$

   If $n(E_j)$ events with effect E_j are observed, the expected number of events associated to each cause is $\hat{n}(C_i) = n(E_j)P(C_i|E_j)$. The estimated true distribution will be $\hat{P}(C_i) = \hat{n}(C_i)/\sum_{k=1}^{n_C} \hat{n}(C_k)$. Since this is an iterative method, the estimated value $\hat{P}(C_i)$ is used as new hypothesis for the true distribution, substituting $P_0(C_i)$.
   This method is also dependent on a regularization parameter that represents the number of the iterations, because truncating them, the effects of statistical fluctuations are removed.

Both methods are used to cross check the results and to estimate a systematic error of the procedure. They are both implemented by the RooUnfold package [77].

In both methods, a response matrix describing the changes in each bin of a distribution when moving from the detector level (corrected for the various efficiencies) to the original particle one is computed using the Monte Carlo simulation. Reconstructed level objects are associated to the corresponding particles at event generator level, accounting for any possible smearing induced by detector resolution effects. This matrix is applied to real data, corrected for the efficiencies, obtaining in this way the unfolded data distributions. The details of the validation procedure adopted and of the results of the closure tests performed are described in par. 5.5.2.

A crucial point of this procedure is the selection of the best value of the regularization parameter k, both for SVD and Bayesian algorithm. For this reason the dependency of the unfolded results on varying the parameter $k_{SVD/Bayes} \in [1,\ n_{bins}]$ is determined. The optimal value is selected by finding the largest value up to which the errors associated to the unfolding procedure remain small (compared to the statistical ones) and stable: small values would bias the unfolding result towards the MC truth input, while a too-large value would give a result that is dominated by un-physically enhanced statistical fluctuations [75].

In order to perform the unfolding procedure, a definition of electrons and jet objects at the generator level (genJets) is needed. In Sec. 3.2.2 the genJets strategy reconstruction is described: to account for the QED final state radiation (FSR) at generator level, the generated electron object is composed by the bare electron and all the photons in a cone around it and the selected objects are removed from the particle collection used for the genJets clusterization.

5.5.1 Response matrix calculation

Generated jets are considered only if they satisfy the following requirements:

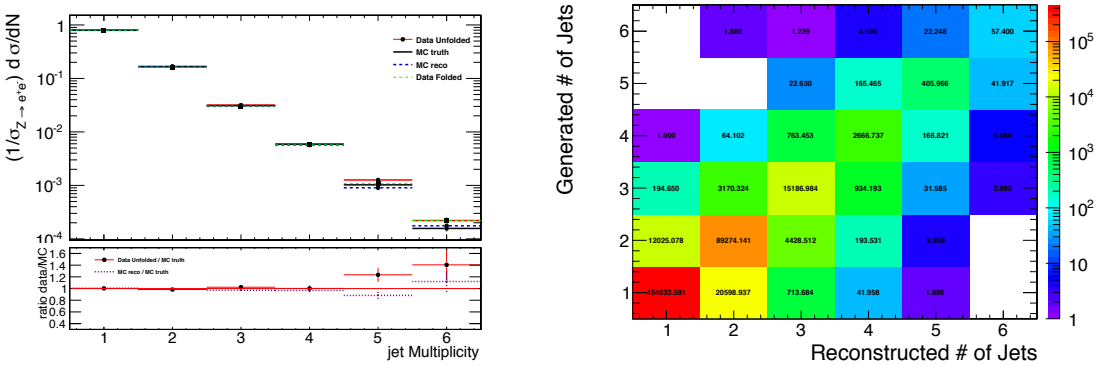
- only generated jets with a transverse momentum $p_T^{jet} > 30$ GeV/c are considered, within $|\eta| \leq 2.4$.
- Jet charge multiplicity ≥ 1, i.e. at least a charged track with $p_T \geq 250$ MeV.

To evaluate the migration effect (including over/underflows), the response matrix used in the unfolding has been filled with jets passing less tight selections ($p_T^{jet} > 15$ GeV/c and $\eta \leq 2.5$). Some examples of the results of the unfolding procedure for both the methods used are shown in Fig. 5.11 and Fig. 5.12.

No unfolding is applied to correct jet η distributions: the effect was almost negligible (thus the matrix response is diagonal), given the very small uncertainty in the η measurement.

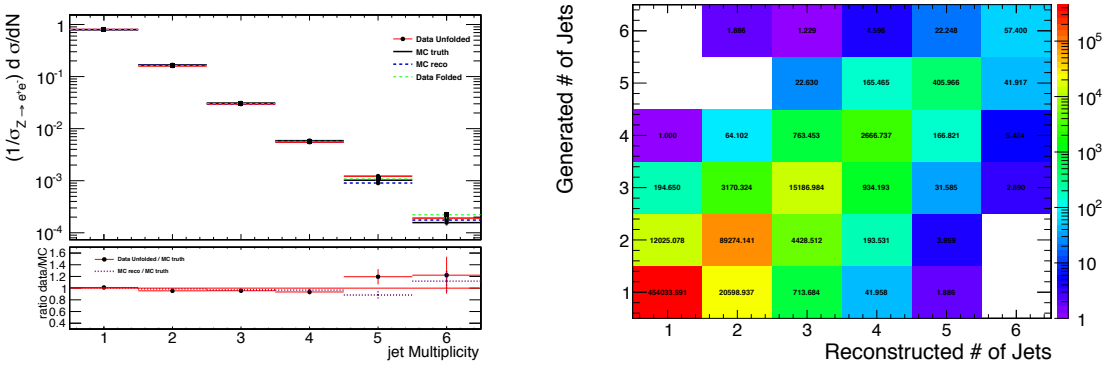
5.5.2 Validation of unfolding method

In order to demonstrate that the unfolding algorithm does not introduce any additional bias, three different tests have been carried out:



(a) Unfolding results (b) Unfolding response matrix

Figure 5.11: Unfolding of the jet multiplicity distribution with the SVD method.



(a) Unfolding results (b) Unfolding response matrix

Figure 5.12: Unfolding of the jet multiplicity distribution with the Bayesian method.

1. The response matrix is applied to the same MC (`Madgraph Z2`) used to built the matrix itself. While doing this, it is expected to obtain again the MC truth distribution and the ratio unfolded result over MC truth exactly equal to 1. This test can be seen in Fig. 5.13a and Fig. 5.13d done respectively with SVD and Bayesian methods for the differential cross sections as a function of the jet multiplicity.

2. MC sample used above is splitted into two subsamples of different size, one (subsample A) having an effective integrated luminosity approximately equal to the 2011 integrated luminosity 5 fb^{-1} and one (subsample B) obtained from the remaining part. Then, the matrix response calculated from subsample A is applied to unfold the subsample B. In Fig. 5.13b and Fig. 5.13e good agreement between the unfolding and the truth MC is shown.

3. In the last test, the unfolding procedure is applied to the distribution obtained from a signal Monte Carlo that is not identical to the one used to derive the response matrix (`Pythia 6`). This test produced a very good agreement between unfolded and truth distribution, showing no explicit dependence on the type of MC used for the unfolding (Fig. 5.13c and Fig. 5.13f).

Each test is made by varying the k regularization parameter from 1 to the maximum number of bins. All tests are computed for both Bayesian and SVD algorithm, in order to additionally test the consistency of the different approaches. All tests done show that the unfolding procedure is consistent with our expectations, and that the two different statistically approaches lead us to compatible results.

5.6 Systematic uncertainties

The systematic uncertainties considered in the evaluation of the studied observables are essentially caused by four different sources:

- jet energy corrections (JEC);

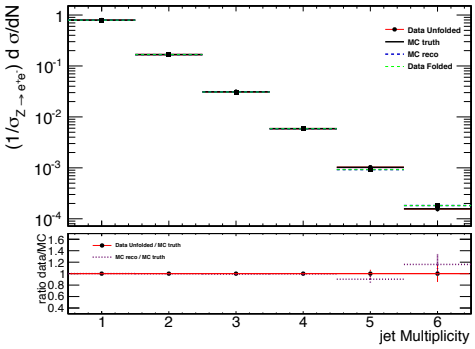
- unfolding procedure;

- pile-up subtraction;

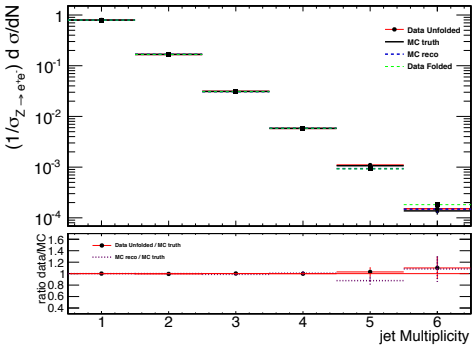
- efficiency correction and background subtraction.

The first kind of uncertainties are related to the jet energy reconstruction. As explained in Sec. 3.2.3, the energy of the reconstructed jets can differ from the true energy of its constituents due to several different factors, such as the non-uniformity and non-linearity of the response of the detector in pseudorapidity and momentum in the energy reconstruction or the pile up subtraction.

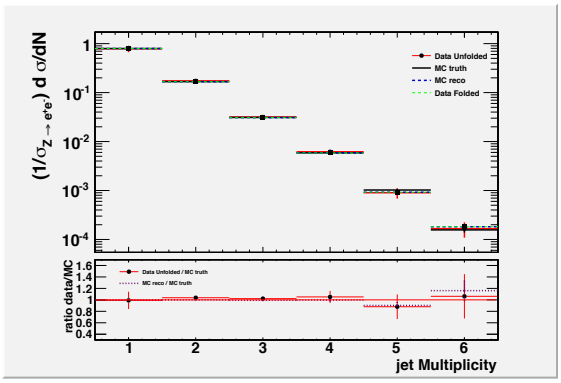
Each contribution is evaluated singularly as explained in [64] and the overall effect is calculated by adding in quadrature the individual contributions. The systematic effect due to the correction of the jet energy is estimated by comparing the



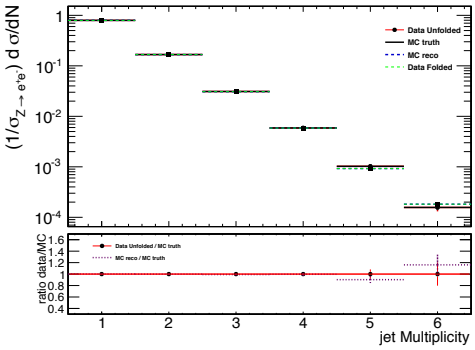
(a) result of check 1 for SVD method



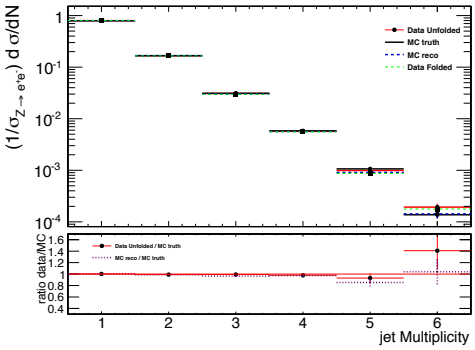
(b) result of check 2 for SVD method



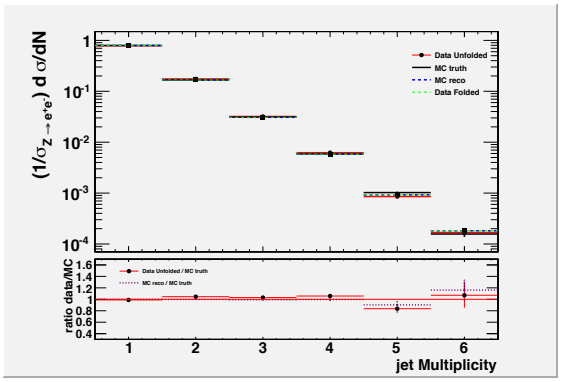
(c) result of check 3 for SVD method



(d) result of check 1 for Bayesian method



(e) result of check 2 for Bayesian method



(f) result of check 3 for Bayesian method

Figure 5.13: Closure test for the unfolding of the jet multiplicity distribution.

distributions obtained using the standard JEC and the ones resulting by the application of the JEC modified by their uncertainty. The size of the variation depends on the η and p_T of the jet itself as shown in Fig. 5.14. The JEC are the greatest source of systematics error in this analysis and in Fig. 5.15 the contribution of the JEC uncertainties on some of the studied observables is shown.

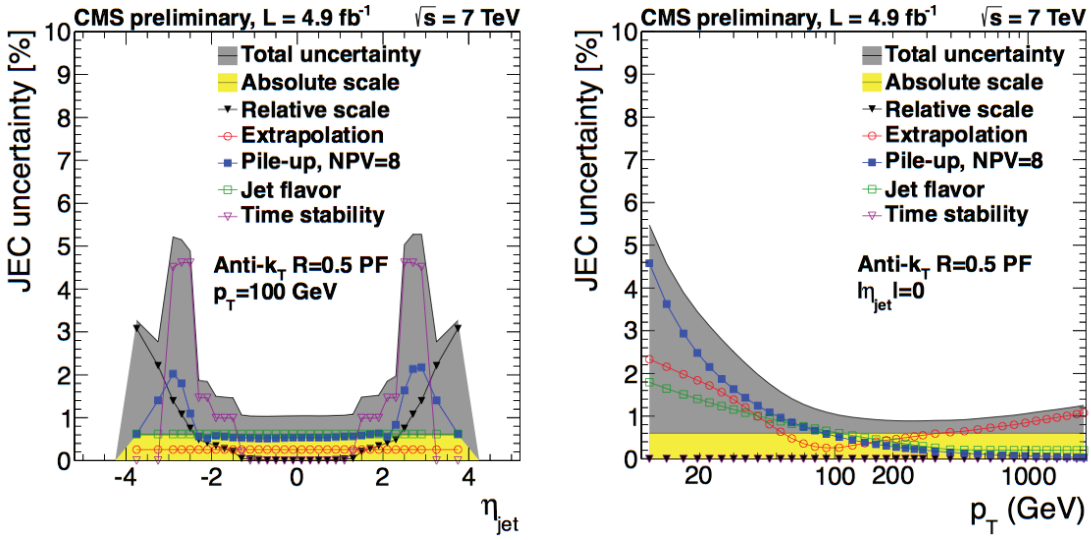


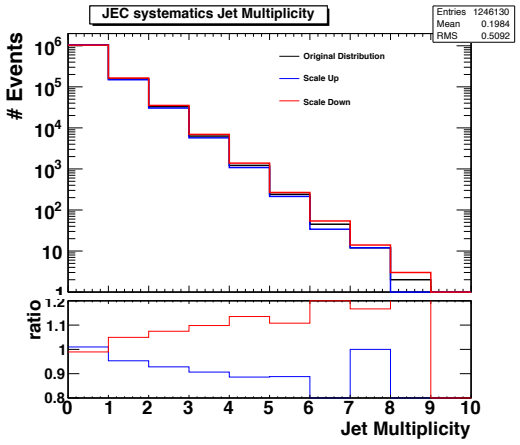
Figure 5.14: The JEC uncertainties as a function of η (left) and p_T (right) of the jet itself are reported, showing both the total effect and the single contributions.

The systematic uncertainty associated to the unfolding procedure is given by different contributions. The first one is related to the sensitivity of the response matrix to the choice of the model used to generated the Monte Carlo events. This dependence can be estimated comparing the resulting unfolded distribution got using the response matrix based on `MadGraph` and the one based on the pure `Pythia6`. The relevance of this contribution is very small as seen in the distributions shown in Figs. 5.13c and 5.13f.

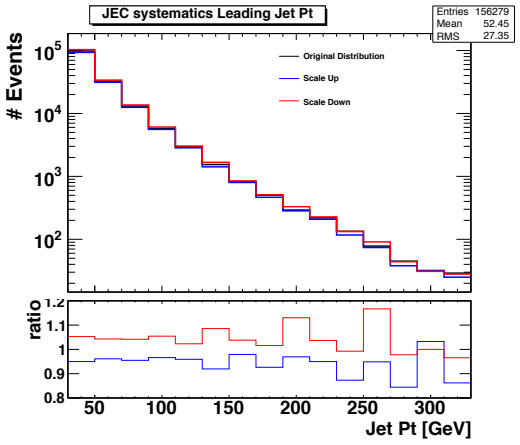
A contribution to the total unfolding uncertainty is also given by the choice of the normalization parameter used in the algorithm.

The most relevant contribution is related to the choice of the unfolding algorithm and it is evaluated comparing the results obtained with the SVD and the Bayesian methods. In Fig. 5.16 the systematic uncertainties caused by the unfolding procedure are shown for some of the analysed observables.

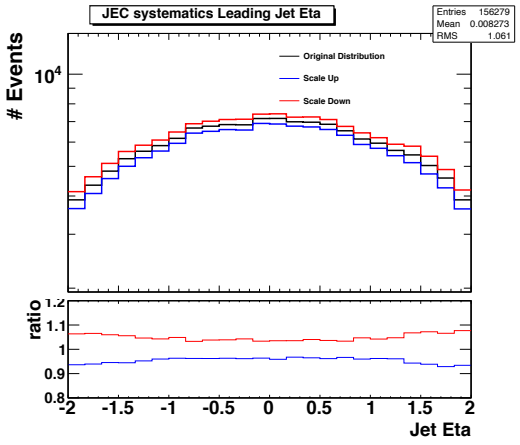
Another effect to take into account considering the systematic uncertainty is related to the pile-up subtraction. Its effect is evaluated calculating the differences between the distributions obtained modifying by $\pm 8\%$ the Minimum-Bias cross section used to reweight the samples for the pile-up. Some example of the uncertainty values obtained for some of the considered observables are reported in Fig. 5.17.



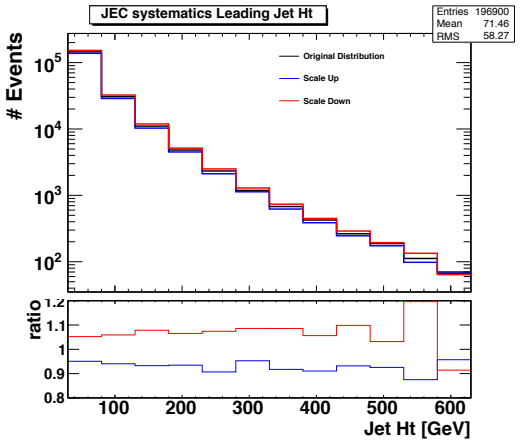
(a) Jet multiplicity distribution



(b) Leading jet p_T distribution

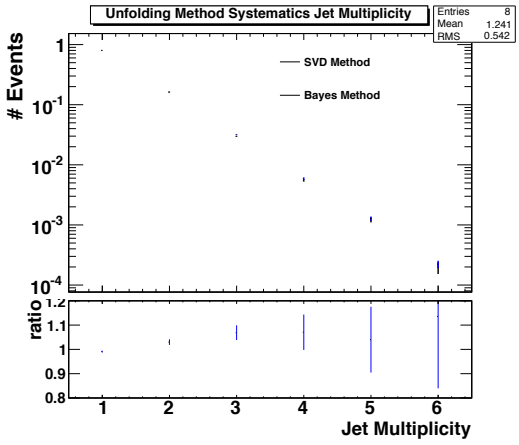


(c) Leading jet *eta* distribution

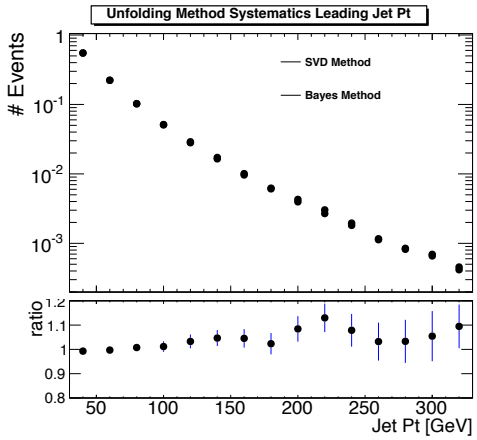


(d) Leading jet H_T distribution

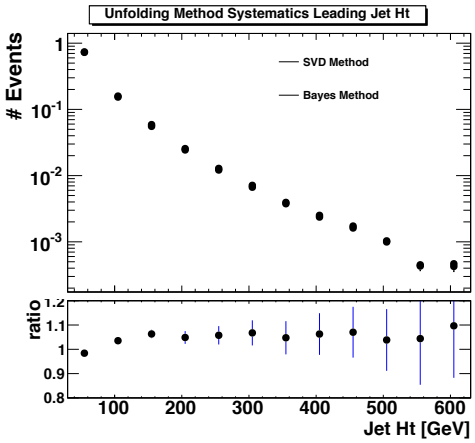
Figure 5.15: JEC uncertainties for some of the studied observables.



(a) Jet multiplicity distribution

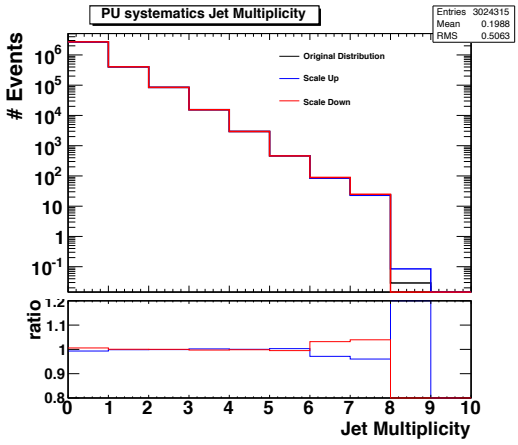


(b) Leading jet p_T distribution

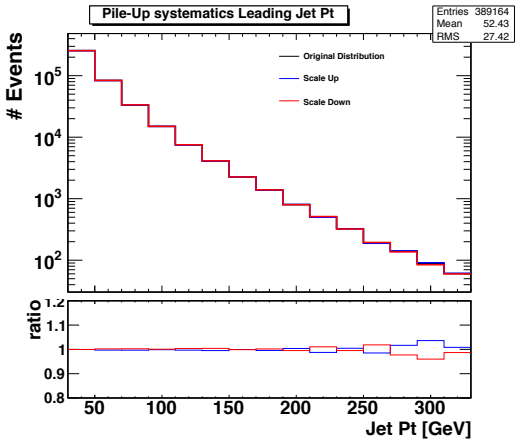


(c) Leading jet H_T distribution

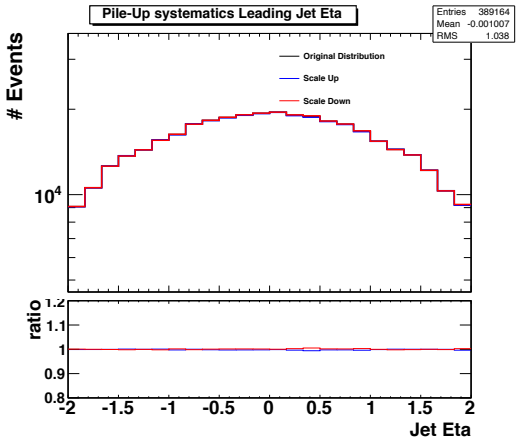
Figure 5.16: Unfolding uncertainties for some of the studied observables.



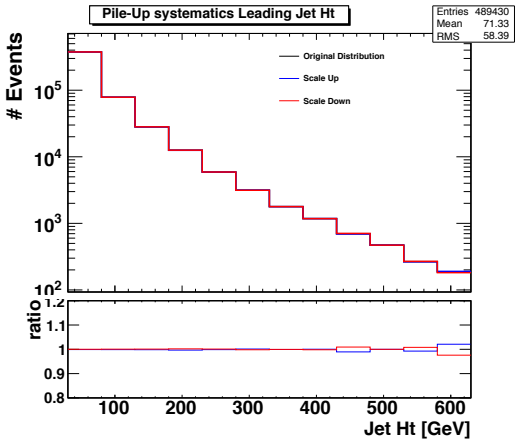
(a) Jet multiplicity distribution



(b) Leading jet p_T distribution



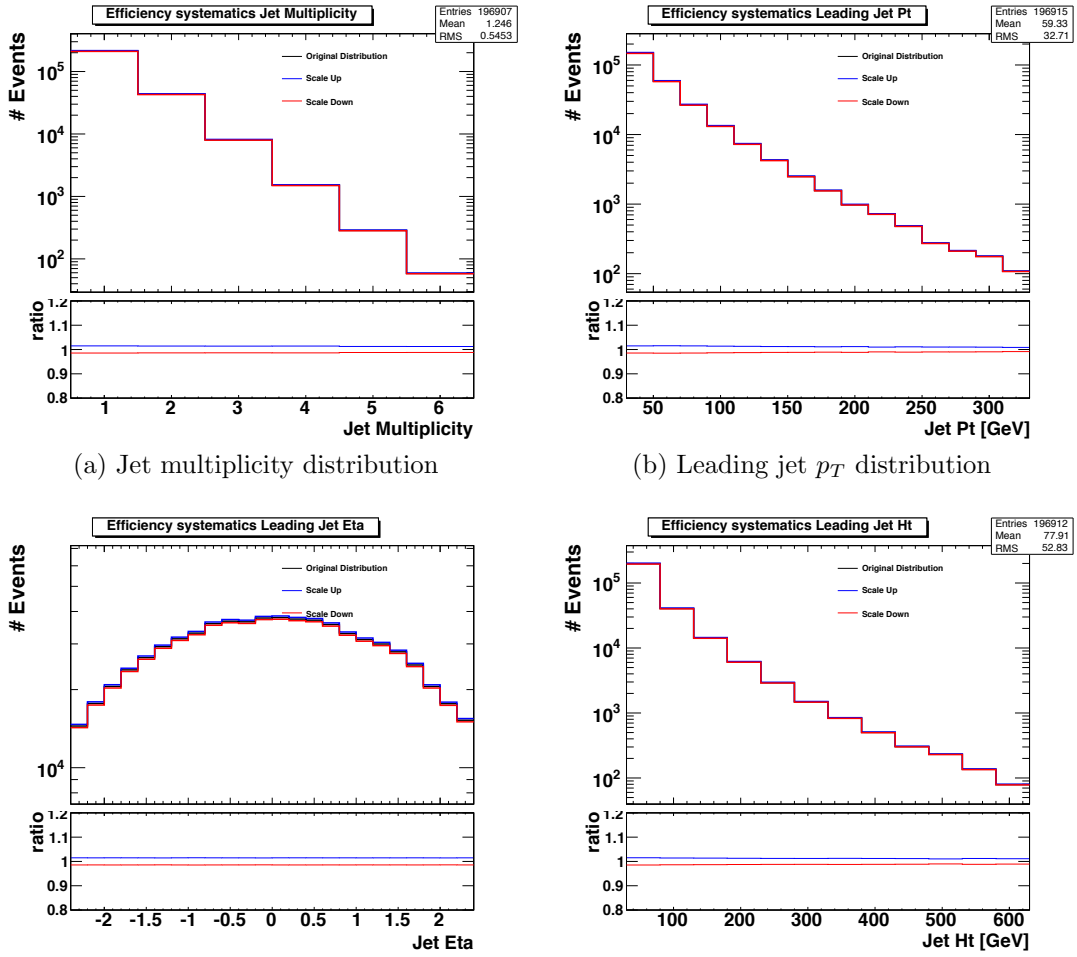
(c) Leading jet *eta* distribution



(d) Leading jet H_T distribution

Figure 5.17: Pile-up uncertainties for some of the studied observables.

Finally, the last contributions to the systematics uncertainties are given by the application of the efficiency corrections and the background subtraction process. The efficiency factors are calculated with an associated error and, as for the jet energy correction, this uncertainty must be propagated to the interesting observables. In order to estimate the systematics, the distribution are corrected with two different sets of efficiencies obtained by adding and subtracting the statistical error from the fits and the final results are compared. In Fig. 5.18 the uncertainties related to the efficiency corrections are shown for some distributions.



(a) Jet multiplicity distribution

(b) Leading jet p_T distribution

(c) Leading jet *eta* distribution

(d) Leading jet H_T distribution

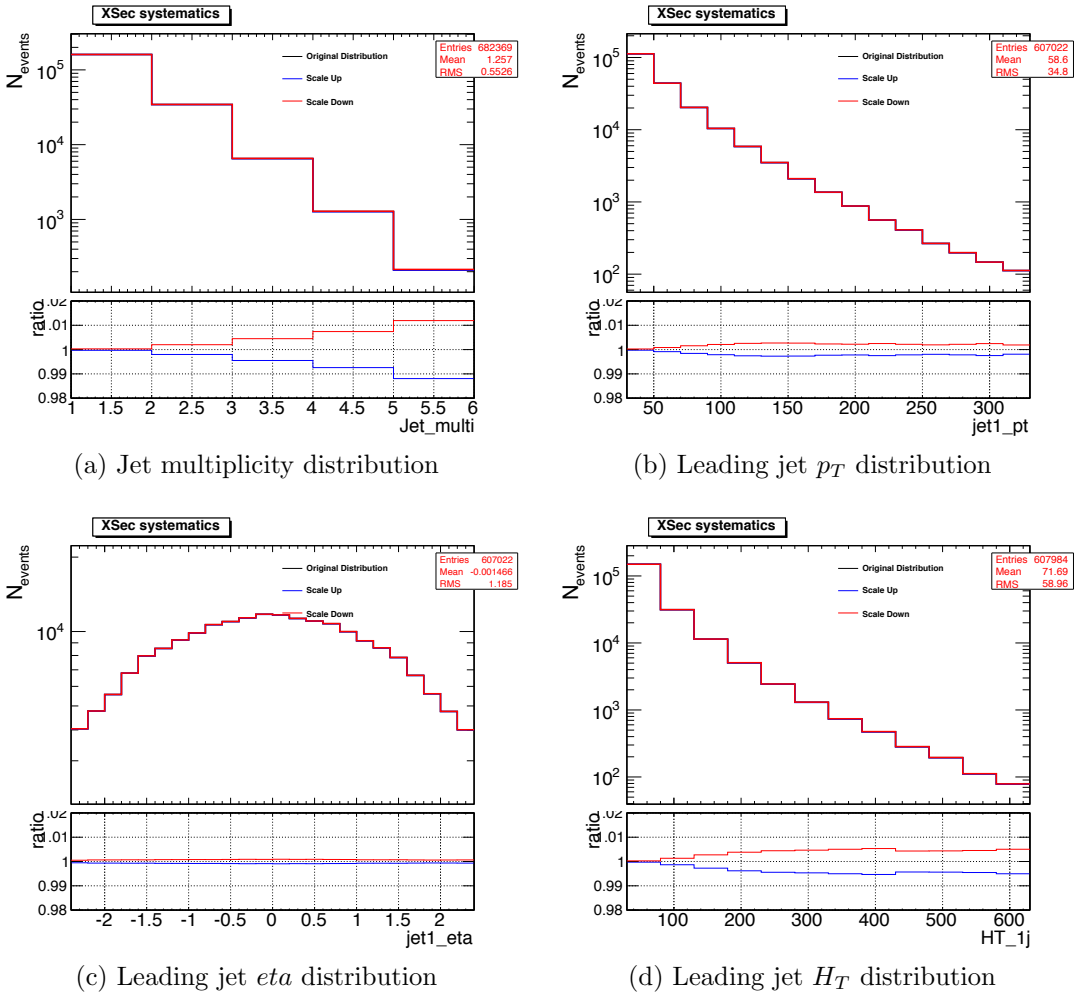
Figure 5.18: Efficiency uncertainties for some of the studied observables.

The background subtraction effect has the smallest impact on the overall systematics uncertainty and can be divided in two different contributions. The first one comes from the cross sections used to rescale the background samples. The impact of this source of error is straightforwardly obtained from the theoretical errors associated to the corresponding cross sections. Taking into account the worst possible scenario, all the cross sections are increased and decreased simultaneously, as reported in Tab. 5.8. The results are compared with Monte Carlo distributions

process	ttbar	W	WW	WZ	ZZ
σ	165	31314	43.0	18.2	5.90
$\sigma^{scaleUp}$	175	32872	44.5	18.9	6.05
$\sigma^{scaleDown}$	155	29756	41.5	17.5	5.75

Table 5.8: Cross sections used to scale the background. The standard values used in the analysis are listed in the first row; in the second and the third the σ used to calculate the systematics concerning the evaluation of the background are shown. All the values are expressed in pb.

obtained with the standard values in order to evaluate the related uncertainty. The impact of this systematics on some of the studied observables is reported in Fig. 5.19.



(a) Jet multiplicity distribution

(b) Leading jet p_T distribution

(c) Leading jet *eta* distribution

(d) Leading jet H_T distribution

Figure 5.19: Background uncertainties for some of the studied observables.

The second component is due to the fact that the efficiencies used to correct the event yields, both for data and MC, are subjected to a different background

subtraction technique. In the first case a fit procedure is applied, while in the second the information given by the Monte Carlo truth are exploited. The relevance of this source of systematics is estimated by comparing the efficiency obtained by applying the fit procedure on a Monte Carlo sample and the corresponding exact efficiency from the Monte Carlo truth information. This contribution is totally negligible as seen in Fig. 5.9.

The detailed breakdown of the uncertainties on the measured distributions is presented in Table 5.9, 5.10 and 5.11. The total error is calculated by adding in quadrature all the listed components.

Table 5.9: Source of systematics errors on the differential cross sections as a function of the jet multiplicity and p_T, up to the four most energetic jet (in percentage).

Systematic source [%]	$\frac{d\sigma}{dN}$	$\frac{d\sigma}{dp_T}$ 1st jet	$\frac{d\sigma}{dp_T}$ 2nd jet	$\frac{d\sigma}{dp_T}$ 3rd jet	$\frac{d\sigma}{dp_T}$ 4th jet
Unfolding	0.9 - 12	0.3 - 11.5	0.15 - 4.8	0.2 - 6.6	>0.1 - 5.6
Efficiency	1.2 - 1.5	0.9 - 1.6	1.2 - 1.6	1.0 - 1.5	1.1 - 1.5
JEC	1 - 11	3.2 - 17	3 - 17	5.8 - 13	9 - 15
PU	0.5 - 0.7	0.1 - 4	0.1 - 1	0.1 - 1	0.1 - 1
Bkg XSec	0.03 - 1.2	0.08 - 0.3	0.2 - 0.5	0.4 - 1.0	0.7 - 1.7
Total	2.1 - 16.4	4.6 - 17	6.2 - 17	6.7 - 15	10 - 16

Table 5.10: Source of systematics errors on the differential cross sections as a function of H_T, up to the four most energetic jet (in percentage).

Systematic source [%]	$\frac{d\sigma}{dH_T}$ 1st jet	$\frac{d\sigma}{dH_T}$ 2nd jet	$\frac{d\sigma}{dH_T}$ 3rd jet	$\frac{d\sigma}{dH_T}$ 4th jet
Unfolding	2 - 9	3 - 13	4 - 14	6 - 30
Efficiency	1.0 - 1.5	1.0 - 1.5	1.1 - 1.5	1.1 - 1.5
JEC	5 - 14.8	4 - 21	7 - 16	14 - 18
PU	0.1 - 2	0.1 - 2	0.2 - 1	0.1 - 0.7
Bkg XSec	0.1 - 0.5	0.1 - 0.6	0.3 - 0.7	0.6 - 1.0
Total	5.6 - 20	10 - 33	10 - 18	14 - 38

5.7 Results

In this section the distributions of the measured observables, corrected for efficiencies and detector effects, are presented. Jets and electrons are defined in the geometrical acceptance as described in the previous sections. All the results are compared to theoretical distributions obtained with the analysis independent `Rivet` [78] tool at

Table 5.11: Source of systematics errors on the differential cross sections as a function of η, up to the four most energetic jet (in percentage).

Systematic source [%]	$\frac{d\sigma}{d\eta}$ 1st jet	$\frac{d\sigma}{d\eta}$ 2nd jet	$\frac{d\sigma}{d\eta}$ 3rd jet	$\frac{d\sigma}{d\eta}$ 4th jet
Efficiency	1.4 - 1.5	1.3 - 1.5	1.3 - 1.4	1.1 - 1.5
JEC	3.4 - 7.7	6 - 9.5	7.7 - 12.5	11.7 - 17
PU	0.1 - 0.6	0.1 - 0.5	0.1 - 1	0.1 - 1.5
Bkg XSec	0.05 - 0.09	0.2 - 0.3	0.4 - 0.6	0.6 - 1.0
Total	3.7 - 8	6 - 10	8 - 13	12 - 17

particle level: theory results are computed with the `MadGraph` [30] plus `Pythia6` [31] and `Sherpa` [38] generator setups. The overall cross section has been rescaled to the NNLO prediction available [21]. Theoretical uncertainties on these predictions have been evaluated by varying consistently the factorization and normalization scale by a factor of 2 in both directions, and by testing the changes induced by alternative choices of PDF (for `Madgraph` the basic choice is CTEQ6L1, for `Sherpa` it is CT10 [79], while in both cases the considered alternatives are MSTW2008lo68cl [80] and NNPDF21 [81]).

5.7.1 Jet multiplicity

In Fig. 5.20 the measured cross section as a function of the exclusive jet multiplicity, for Z decaying into electrons, is shown. Up to 6 jets in the final state are considered. The data are normalized to the total $Z \to ee$ + jets cross section, and in the plots are reported also the statistic and the total (statistic + systematic) error, after the unfolding procedure described in the previous section. The trend of the jet multiplicity represents the expectation of the pQCD prediction for an exponential decay with the number of jets.

5.7.2 Differential cross sections

The jet differential cross section $\frac{d\sigma}{dp_T}$ for the 4 highest p_T jets in the event is presented in Fig. 5.21, in events with at least one jet in the final state. The cross sections are normalized to the cross section for production of Z decays with at least 1 jet, measured in the same kinematic region for electrons and consistent with the results in Sec. 5.2, with the aim of cancelling systematic uncertainties related to electron identification and luminosity. Total and statistic only uncertainties are included.

Jet differential cross sections as a function of the pseudorapidity of the jets, until the fourth one in p_T ranking, are presented in Fig. 5.22 for unfolded distributions in the acceptance region defined in Sec. 5.2. Jet differential cross sections as a function of H_T for different jet multiplicities are also presented in Fig. 5.23. Again both the total error and the statistic one are reported.

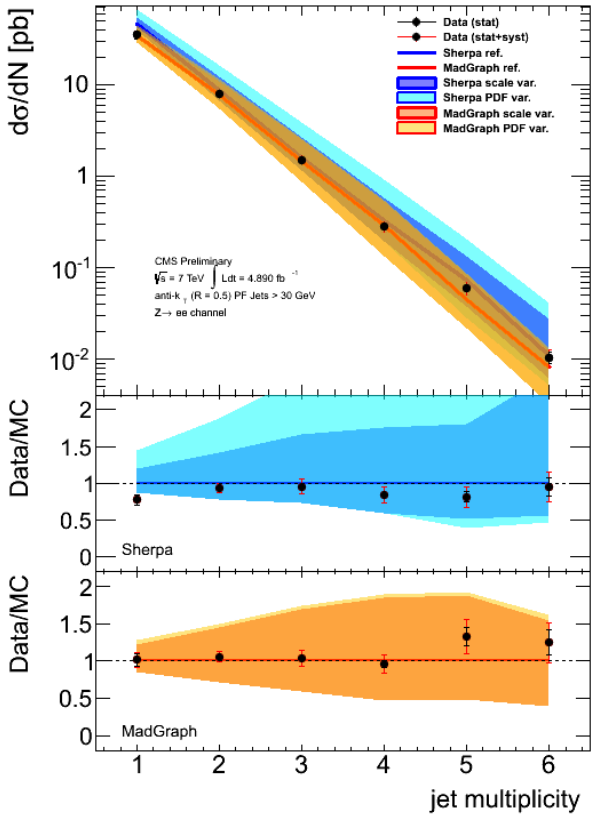
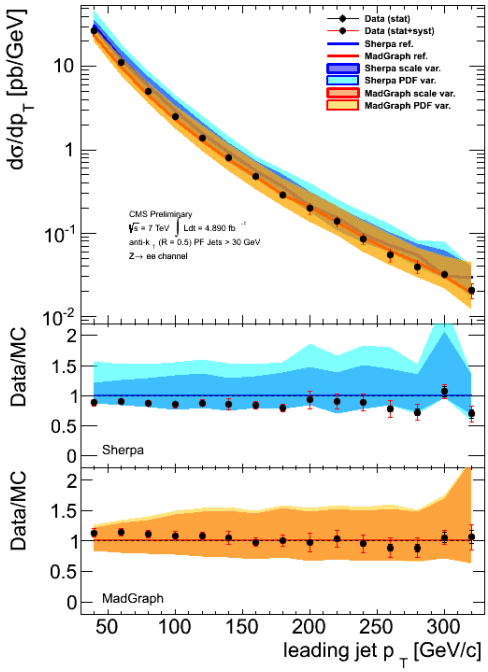
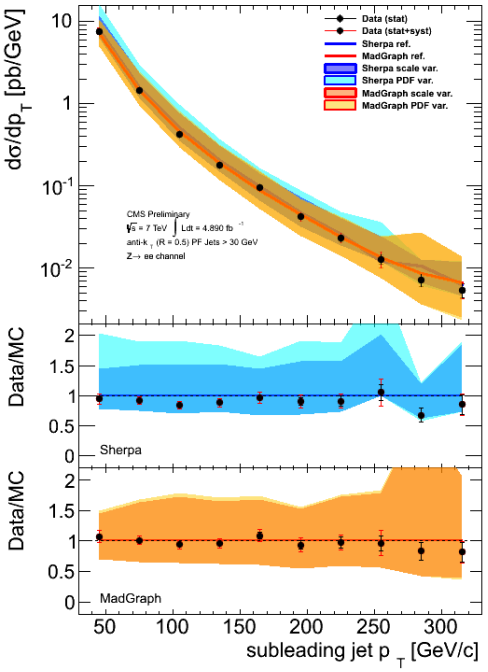


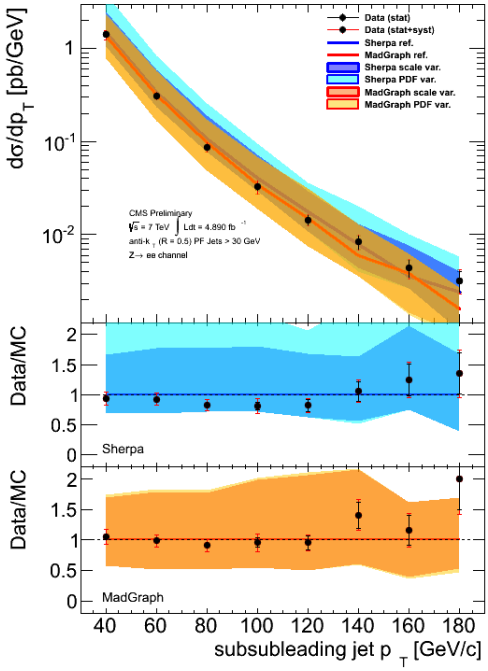
Figure 5.20: Inclusive jet multiplicity distribution.



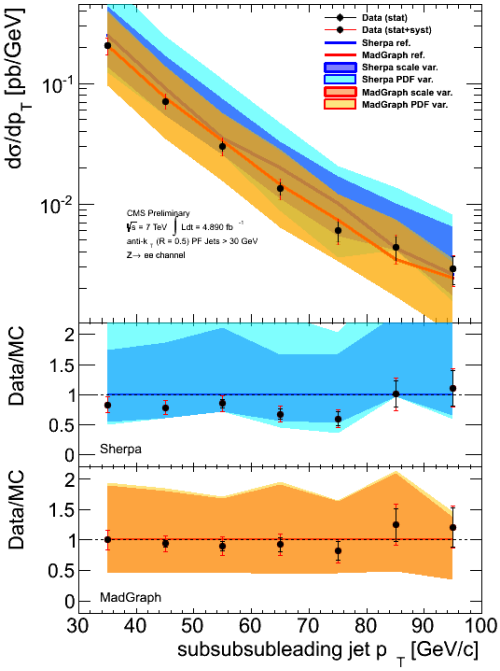
(a) leading jet p_T distribution



(b) second leading jet p_T distribution

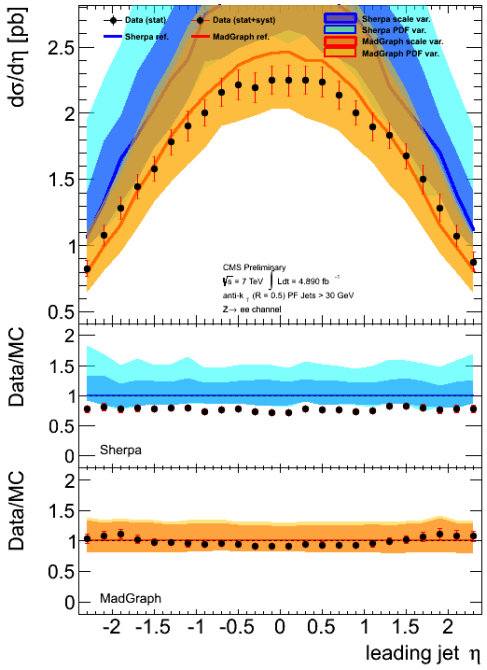


(c) third leading jet p_T distribution

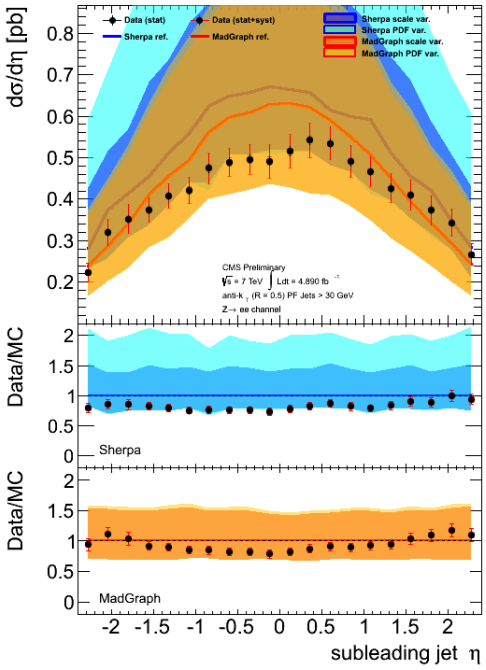


(d) fourth leading jet p_T distribution

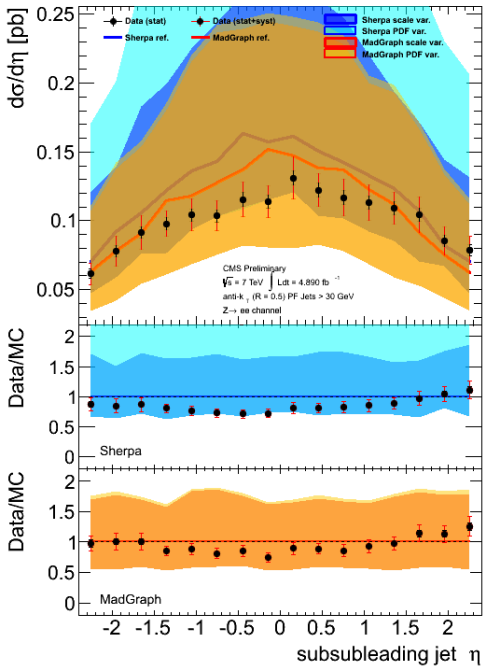
Figure 5.21: Unfolded differential cross sections $\frac{d\sigma}{dp_T}$ as a function of p_T, for the four highest p_T jets, including systematic and statistical uncertainties.



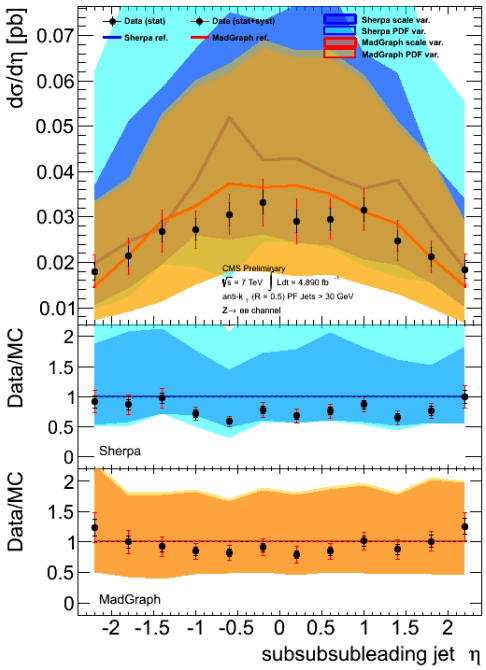
(a) leading jet eta distribution



(b) second leading jet eta distribution

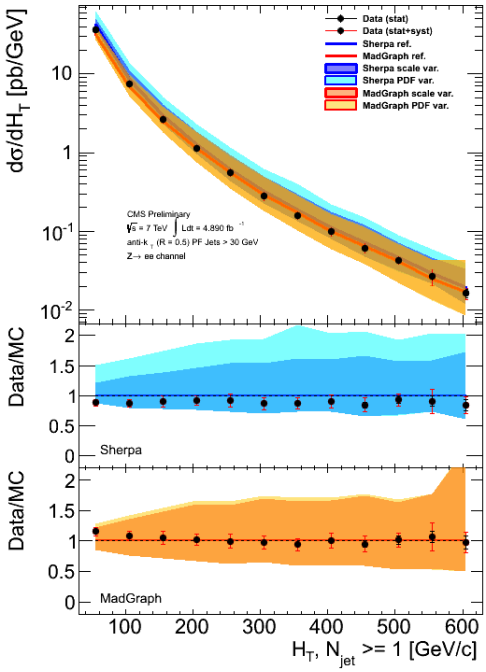


(c) third leading jet eta distribution

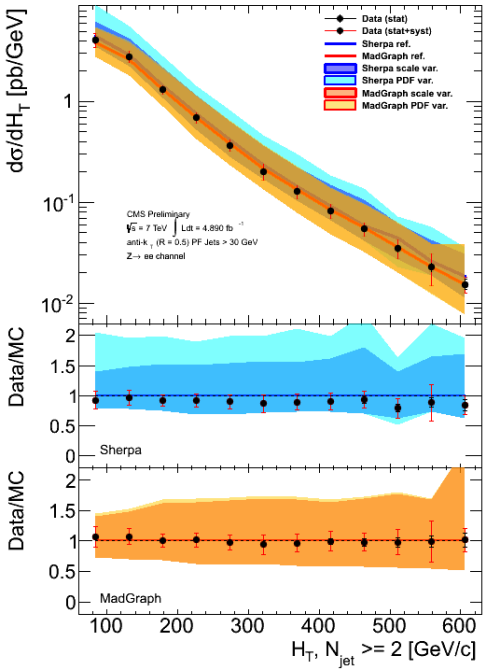


(d) fourth leading jet eta distribution

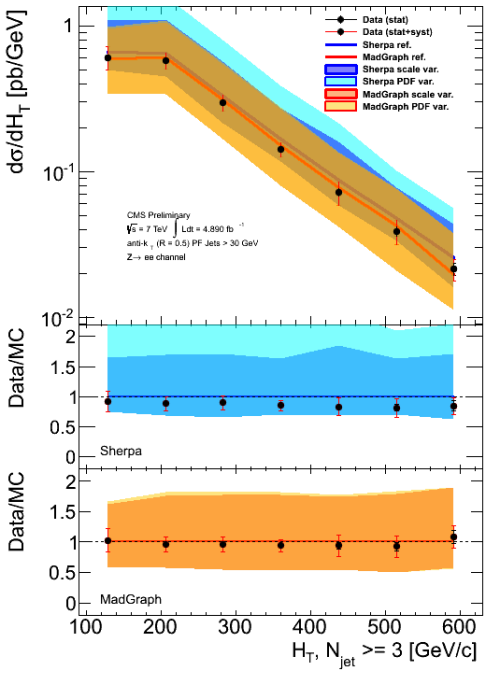
Figure 5.22: Unfolded differential cross sections $\frac{d\sigma}{d\eta}$ as a function of η, for the four highest p_T jets, including systematic and statistical uncertainties.



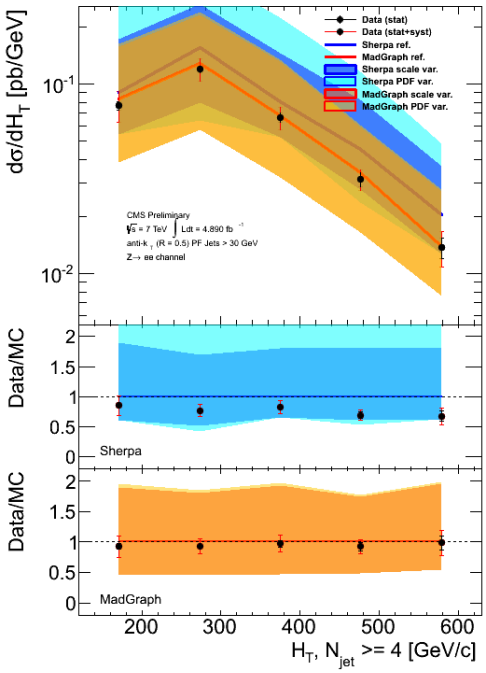
(a) H_T distribution with at least one jet



(b) H_T distribution with at least two jets



(c) H_T distribution with at least three jets



(d) H_T distribution with at least four jets

Figure 5.23: Unfolded differential cross sections $\frac{d\sigma}{dH_T}$ as a function of H_T, for the four highest p_T jets, for different jet multiplicities, including systematic and statistical uncertainties.

Conclusions

The 2011 year was characterized by interesting physics results for the LHC and for the CMS experiment, that managed to collect about 5 fb^{-1} of data, due to the good performance of the accelerator. The instantaneous luminosity provided by the LHC increased very fast, from 10^{32} to more than $5 \cdot 10^{33}$ $cm^{-2}s^{-1}$. At the same time, the pileup also increased from few events on average in early 2011 to about more than ten events at the end of the 2011 data taking, requiring a special effort in order to rapidly adapt both trigger and reconstruction to the changes of the collision conditions. CMS was able to carry out many of its physics goals, concerning Higgs and new physics searches as well as SM measurements.

In this thesis the measurement of the differential cross sections of Z boson and jets associated production, considering the electronic decay of the vector boson, is presented. They are studied as a function of different jet properties, as multiplicity, transverse momentum, pseudorapidity and sum of the transverse momentum of the jets in the event, called H_T. The jet multiplicity and transverse momentum can give directly information about the ability of the ME and PS in the description of the jet production and evolution. The pseudorapidity is the most sensitive variable to the influence of the choice of PDF. Finally, H_T is an interesting observable since it is exploited in the selection of multijet events in searches. Moreover this is the first time H_T is measured at LHC.

In order to perform the most precise and accurate measurement as possible, a careful study of the best choice of the event components reconstruction is performed. In the CMS framework two different algorithms are exploited for the electron reconstruction. A dedicated study on their performance is carried out, showing similar results for both methods. On the basis of this evaluation and in order to obtain a consistent description of the whole event, a full PF-based reconstruction is used in this thesis.

The jets represent the other main character of this work. Exploiting the PF-based reconstruction, it is possible to discard the contamination of electrons originated by the Z boson decay from the hadronic jets in a simple way, without the removal of additional particles as the one applied in previous similar analysis.

After the choice of the objects to analyse, the selection of the interesting events is carried out applying a list of simple cuts involving kinematic and quality properties: two energetic, isolated and well-identified electrons together with hadronic jets are required.

The results are corrected for experimental efficiencies, resolutions and estimated

backgrounds. A detailed evaluation of the systematic uncertainties is presented. The main contribution to these systematics is caused by the jet energy corrections, needed to take care of pile-up effects, non-uniformities and linearities in detector response and remaining differences in data and simulation.

The final distributions are compared to perturbative QCD predictions obtained with different generators at leading order: in the configuration used in the analysis `MadGraph+Pythia6` is associated to a leading order parton density function and `Sherpa` to a next-to-leading order PDF. To evaluate the theoretical uncertainties, two elements are considered: the normalization and factorization scale are modified by a factor 2 and in order to calculate the contribution of the PDF, the results obtained with two different distribution functions are taken into account.

In general, the leading-order calculations reproduce well the shape of the distributions, but they suffer of great scale uncertainty with respect to the errors associated to the measured results. `MadGraph+Pythia6` describes better the data than `Sherpa` due to the bigger shift in the normalization of the distributions. In the differential cross sections as a function of the pseudorapidity larger discrepancies are observed with respect to the other results. They are strongly influenced by the choice of the PDFs, that, as said before, are different in the standard configuration of the two generators. However, an overall good agreement is globally found on all the measured distributions, considering the large uncertainties on the predictions due to the choice of the parton density functions and the scale of the interaction.

For future developments a significant contribution will be given by the comparison of the results with theoretical predictions at next-to-leading order. It will represent a considerable improvement, in particular for the uncertainties associated to the interaction scale that will decrease considerably. Also the analysis update with the 8 TeV data, available with a greater statistics, will be an interesting development in order to reduce the associated errors and put more stringent constraints to the predictions.

References

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

[2] Abdus Salam and John Clive Ward. Electromagnetic and weak interactions. *Phys.Lett.*, 13:168–171, 1964.

[3] Steven Weinberg. A Model of Leptons. *Phys.Rev.Lett.*, 19:1264–1266, 1967.

[4] M.Y. Han and Yoichiro Nambu. Three Triplet Model with Double SU(3) Symmetry. *Phys.Rev.*, 139:B1006–B1010, 1965.

[5] The CMS Collaboration. Inclusive search for a fourth generation of quarks with the CMS experiment. 2011.

[6] S. Nektarijevic. Fourth Generation Searches at ATLAS. 2012.

[7] R. P. Feynman. *Quantum Electrodynamics*. Benjamin, New York, 1962.

[8] Richard L. Garwin, Leon M. Lederman, and Marcel Weinrich. Observations of the failure of conservation of parity and charge conjugation in meson decays: the magnetic moment of the free muon. *Phys. Rev.*, 105:1415–1417, Feb 1957.

[9] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity convservation in beta decay. *Phys. Rev.*, 105:1413–1414, 1957.

[10] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of neutrinos. *Phys. Rev.*, 109:1015–1017, Feb 1958.

[11] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys.Rev.Lett.*, 13:508–509, 1964.

[12] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys.Lett.*, 8:214–215, 1964.

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

[14] D.J. Gross and Frank Wilczek. Asymptotically Free Gauge Theories. 1. *Phys.Rev.*, D8:3633–3652, 1973.

[15] Steven Weinberg. Nonabelian Gauge Theories of the Strong Interactions. *Phys.Rev.Lett.*, 31:494–497, 1973.

[16] Raymond Brock et al. Handbook of perturbative QCD: Version 1.0. *Rev.Mod.Phys.*, 67:157–248, 1995.

[17] J. D. Bjorken. Asymptotic sum rules at infinite momentum. *Phys. Rev.*, 179:1547–1553, Mar 1969.

[18] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl.Phys.*, B126:298, 1977.

[19] R. Keith Ellis, W. James Stirling, and B.R. Webber. QCD and collider physics. *Camb.Monogr.Part.Phys.Nucl.Phys.Cosmol.*, 8:1–435, 1996.

[20] Siegfried Bethke. World Summary of α_s (2012). 2012.

[21] Kirill Melnikov and Frank Petriello. Electroweak gauge boson production at hadron colliders through O(alpha(s)**2). *Phys.Rev.*, D74:114017, 2006.

[22] H. Ita, Z. Bern, L.J. Dixon, Fernando Febres Cordero, D.A. Kosower, et al. Precise Predictions for Z + 4 Jets at Hadron Colliders. *Phys.Rev.*, D85:031501, 2012.

[23] T. Aaltonen et al. Measurement of inclusive jet cross-sections in Z/gamma*($\rightarrow e^+e^-$) + jets production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96-TeV. *Phys.Rev.Lett.*, 100:102001, 2008.

[24] V.M. Abazov et al. Measurements of differential cross sections of Z/gamma*+jets+X events in proton anti-proton collisions at s**(1/2) = 1.96-TeV. *Phys.Lett.*, B678:45–54, 2009.

[25] Darren D. Price. W/Z properties and V+jets at the Tevatron. *EPJ Web Conf.*, 28:06006, 2012.

[26] Georges Aad et al. Measurement of the production cross section for Z/gamma* in association with jets in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector. *Phys.Rev.*, D85:032009, 2012.

[27] The CMS Collaboration. Jet Production Rates in Association with W and Z Bosons in pp Collisions at sqrt(s) = 7 TeV. *JHEP*, 1201:010, 2012.

[28] Pavel Nadolsky, Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, et al. Progress in CTEQ-TEA PDF analysis. 2012.

[29] Frits A. Berends, H. Kuijf, B. Tausk, and W.T. Giele. On the production of a W and jets at hadron colliders. *Nucl.Phys.*, B357:32–64, 1991.

[30] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: Going Beyond. *JHEP*, 1106:128, 2011.

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

[32] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys.Rept.*, 97:31–145, 1983.

[33] Torbjorn Sjostrand. A Model for Initial State Parton Showers. *Phys.Lett.*, B157:321, 1985.

[34] Torbjorn Sjostrand. Monte Carlo Generators. pages 51–74, 2006.

[35] V.V. Sudakov. Vertex parts at very high-energies in quantum electrodynamics. *Sov.Phys.JETP*, 3:65–71, 1956.

[36] Stefan Hoeche, Frank Krauss, Nils Lavesson, Leif Lonnblad, Michelangelo Mangano, et al. Matching parton showers and matrix elements. 2006.

[37] S. Catani, F. Krauss, R. Kuhn, and B.R. Webber. QCD matrix elements + parton showers. *JHEP*, 0111:063, 2001.

[38] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al. Event generation with SHERPA 1.1. *JHEP*, 0902:007, 2009.

[39] Yves Baconnier, Giorgio Brianti, P Lebrun, A G Mathewson, R Perin, and Yves Baconnier. *LHC: the Large Hadron Collider accelerator project*. CERN, Geneva, 1993.

[40] The ATLAS collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *J. Instrum.*, 3:S08003. 437 p, 2008.

[41] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[42] The LHCb collaboration. *LHCb : Technical Proposal*. Tech. Proposal. CERN, Geneva, 1998.

[43] The ALICE collaboration. *A large ion collider experiment technical proposal design*. lhcc 95-71, lhcc/p3,1995. Cern, Geneva.

[44] The CMS collaboration. *The CMS magnet project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[45] The CMS collaboration. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[46] The CMS collaboration. *The CMS tracker: addendum to the Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2000.

[47] The CMS collaboration. *The Electromagnetic Calorimeter Project: Technical Design Report*. Technical Design Report CMS. CERN/LHCC 97-33, Geneva, 1997.

[48] The CMS collaboration. *The Hadron Calorimeter Project: Technical Design Report*. Technical Design Report CMS. CERN/LHCC 97-31, Geneva, 1997.

[49] S. Abdullin et al. Design, performance, and calibration of CMS hadron-barrel calorimeter wedges. *Eur.Phys.J.*, C55:159–171, 2008.

[50] G. Baiatian et al. Design, performance, and calibration of CMS hadron endcap calorimeters. 2008.

[51] S. Abdullin et al. Design, performance, and calibration of the CMS Hadron-outer calorimeter. *Eur.Phys.J.*, C57:653–663, 2008.

[52] G. Bayatian, A. Sirunian, I. Emelyanchik, V. Massolov, N. Shumeiko, et al. Design, performance and calibration of the CMS forward calorimeter wedges. *Eur.Phys.J.*, C53:139–166, 2008.

[53] The CMS collaboration. *The Muon Project: Technical Design Report*. Technical Design Report CMS. CERN/LHCC 97-32, Geneva, 1997.

[54] The CMS collaboration. *CMS TriDAS project: Technical Design Report; 1, the trigger systems*. Technical Design Report CMS.

[55] The CMS collaboration. *CMS trigger and data-acquisition project: Technical Design Report*. Technical Design Report CMS.

[56] The CMS collaboration. Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV. *CMS PAS*, PFT-10-002, 2010.

[57] The CMS collaboration. Commissioning of the particle-flow event reconstruction with leptons from J/Ψ and W decays at 7 TeV. *CMS PAS*, PFT-10-003, 2010.

[58] The CMS collaboration. Tracking and Primary Vertex Results in First 7 TeV Collisions. *CMS PAS*, TRK-10-005, 2010.

[59] Wolfgang Adam, R. Fruhwirth, Are Strandlie, and T. Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. 2005.

[60] Gavin P. Salam. Towards Jetography. *Eur.Phys.J.*, C67:637–686, 2010.

[61] The CMS collaboration. Jet performance in pp collisions at 7 tev. *CMS PAS*, JME-10-003, 2010.

[62] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[63] The CMS Collaboration. Characterization of the final state radiation in z boson decay to electrons. CMS Analysis Note CMS-AN-2012-204, 2012.

[64] The CMS collaboration. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *JINST 6 (2011) 11002*, Jul 2011.

[65] M. Cacciari and G. P. Salam. Pileup subtraction using jet areas. *Physics Letters B*, 659:119–126, 2008.

[66] The CMS Collaboration. Electron identification in cms. CMS Analysis Note CMS AN-09-178, 2009.

[67] The CMS Collaboration. Electron reconstruction in cms. CMS Analysis Note CMS AN-09-164, 2009.

[68] The CMS Collaboration. Electron reconstruction within the particle flow algorithm. CMS Analysis Note CMS AN-10-034, 2010.

[69] K. Nakamura et al. Particle data group. *J. Phys. G*, 37:075021, 2010.

[70] The CMS Collaboration. Absolute calibration of luminosity measurement at cms: Summer 2011 update. CMS Physics Analysis Summary CMS-PAS-EWK-11-001, 2011.

[71] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, Pavel M. Nadolsky, et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 0207:012, 2002.

[72] Vardan Khachatryan et al. Charged particle multiplicities in pp interactions at sqrt(s) = 0.9, 2.36, and 7 TeV. *JHEP*, 1101:079, 2011.

[73] The CMS Collaboration. Electron reconstruction and selection studies with first cms 7 tev data. CMS Physics Analysis Summary CMS-PAS-EGM-10-004, 2010.

[74] The CMS Collaboration. Tag and probe methodology for analyses using electrons and photons. CMS Analysis Note CMS-AN-2012-116, 2012.

[75] Andreas Höcker, Vakhtang Kartvelishvili. SVD approach to data unfolding. *Nuclear Instruments and Methods in Physics Research A*, 362:469–481, 1996.

[76] G. D'Agostini. A multidimensional unfolding method based on Bayes' theorem. *Nuclear Instruments and Methods in Physics Research A*, 362:487–498, 1995.

[77] Tim Adye. Unfolding algorithms and tests using RooUnfold. 2011.

[78] A. Buckley et al. Rivet user manual. 2011.

[79] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, et al. New parton distributions for collider physics. *Phys.Rev.*, D82:074024, 2010.

[80] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009.

[81] Francesco Cerutti. The NNPDF2.1 Parton Set. 2011.