UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA
Facoltà di Scienze MM. FF. NN.

Scuola di Dottorato di Scienze
Corso di Dottorato di Ricerca in Fisica ed Astronomia

CMS ECAL intercalibration with electrons from W and Z decays

and

Search for the Standard Model Higgs boson in the $H \to WW \to l\nu_l\, q\bar{q}$ channel with the CMS detector

Coordinatore della Scuola di Dottorato:	prof.	**Giuseppe Chirico**
Relatore:	prof.	**Marco Paganoni**
Correlatore:	dott.	**Pietro Govoni**

Tesi di dottorato di
Leonardo Di Matteo
Matricola: 068587

Ciclo XXV
Anno Accademico 2012-2013

Preface

Particle physics is now living an exceptional moment. After two decades of stall and headaches, the hunt for the Higgs boson has very recently come to a crucial point, culminated on July the 4^{th} 2012, when the European Organization for Nuclear Research (CERN) released an update of the results obtained by two of the experiments situated on the Large Hadron Collider (LHC), which independently observed the presence of a new particle state compatible with the Standard Model prediction.
This manuscript follows nevertheless an historical point of view, to better contextualize my research project.

Chapter 1 presents an extensive summary of the theoretical background of particle physics and follows the findings of particle colliders in the hunt for the Higgs boson. The state of art prior to the LHC era is presented and discussed, due to its importance in understanding why the LHC was built in the first place.

The LHC machine and the Compact Muon Solenoid (CMS), which were designed specifically to be sensitive to the the Higgs boson existence in the allowed mass range at the time of LHC construction, are described in Chapter 2.

Chapter 3 addresses the role of the CMS Electromagnetic Calorimeter (ECAL) in the search for the Higgs boson, in particular through the $H \to \gamma\gamma$ search. The first part of my research program, which aimed at the ECAL intercalibration with electrons from W and Z bosons decays, is described and results with 2011 data are reported.

Chapter 4, contains the main branch of my doctorate project, i.e. the search for the Standard Model Higgs boson with the $H \to WW \to l\nu_l q\bar{q}$ decay channel, which is one of the most sensitive channels in the high mass regime. Results obtained with the full LHC dataset up to fall 2012 are presented.

Finally, Chapter 5 presents the conclusions of this thesis.

Contents

Chapter 1

Introduction

In this chapter the current theory of particles and their interactions, i.e. the so called Standard Model (SM), is presented.

Section 1.1 presents a theoretical and experimental review of the SM.

The constraints on the Higgs boson mass, coming from theoretical motivations and experimental searches prior to the Large Hadron Collider (LHC), are presented in section 1.2.

Section 1.3 contains an overview of the Higgs production and decay mechanisms at the LHC as well as the main search strategies.

Finally, Section 1.4 presents the status of the theory in the light of the most recent findings.

1.1 The Standard Model

The Standard Model of particle physics is the most predictive theory of the electromagnetic, weak, and strong nuclear interactions, which mediate the dynamics of the known subatomic particles [1][2][3]. Mathematically, the Standard Model consists of a relativistic quantum field theory based on the group of symmetries $SU(3)_C \times SU(2)_L \times U(1)_Y$, where each group is associated to one of the the fundamental interactions. The theory is perturbative at sufficiently high energies and renormalizable [4].

1.1.1 How free particles evolve

The Standard Model fundamental matter constituents are spin-$\frac{1}{2}$ particles called *fermions*, which have a mathematical translation in a quantum field operator composed by four-components Dirac spinors, denoted as ψ. Introducing the Weyl representation of the γ matrices

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad \text{and} \quad \gamma^5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \tag{1.1}$$

where the σ^i are the 2×2 Pauli matrices, the adjoint spinor can be built as $\bar{\psi} = \psi^\dagger \gamma^0$, where the dagger operator "†" corresponds to the conjugate transpose.

For a free fermion of mass m, the Lagrangian and the associated equation of motion (Dirac

equation) have these expressions

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi \qquad \Rightarrow \qquad \left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi = 0 \quad . \tag{1.2}$$

For convenience, the Dirac spinor ψ is usually separated into the left-handed and the right-handed spinors $\psi = \psi_L + \psi_R$, obtained by applying the projection operators P_L and P_R

$$P_L = \tfrac{1}{2}\left(1 - \gamma^5\right) \qquad P_R = \tfrac{1}{2}\left(1 + \gamma^5\right) \quad , \tag{1.3}$$

where the left and the right handedness of the spinors is called *chirality*.

Other elements necessary to build up a complete theory are force carriers of mass m and spin zero, i.e. *scalar bosons*. Each of these is represented in QFT by a complex scalar field ϕ, whose dynamics is described by the Klein-Gordon Lagrangian, from which the Eulero-Lagrange law of motion can be easily obtained

$$\mathcal{L}_{\text{KG}} = \left(\partial_{\mu}\phi\right)^{\dagger}\left(\partial^{\mu}\phi\right) - m^2\phi^{\dagger}\phi \qquad \Rightarrow \qquad \left(\Box + m^2\right)\phi = 0 \quad . \tag{1.4}$$

Finally, vector (i.e. spin-1) bosons with mass m are associated to vector field operators A_{μ}, in the four dimensional Minkowski time-space, whose dynamics are described by

$$\mathcal{L}_{\text{Proca}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2A_{\mu}A^{\mu} \qquad \Rightarrow \qquad \left(\Box + m^2\right)A_{\mu} = 0 \quad , \tag{1.5}$$

where the kinetic term has been introduced through the antisymmetric tensor
$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$.

Up to now, only the free non-interacting theory has been introduced. The SM approach to account for interactions between particles is the requirement of local gauge invariance of the Lagrangian, as it is described in the following section.

1.1.2 How the Gauge Principle explains particle interactions

The requirement of a symmetry in the Lagrangian translates into a conservation of charges, via Noether's theorem [5], and allows the introduction of interactions and new fields, corresponding to the interaction (or force) carriers. A particular case of symmetry is given by gauge transformations, namely alterations of the phase of the fields, as a function of space and time, which are not changing the value of any measurable physical quantity.

In QFT, the simplest case of gauge transformation is given by an internal phase transformation of the form $\psi \to e^{iq\alpha}\psi$, where q is a given constant. If the (unmeasurable) phase α is constant in time and space, the transformation is called a *global gauge*, whereas if the phase is allowed to change differently from point to point, $\alpha = \alpha(x^{\mu})$, the gauge is *local*.
The Dirac Lagrangian (Equation 1.2), that describes a free fermion field, is evidently invariant under global phase transformations because $\partial_{\mu}\psi' = e^{iq\alpha}\partial_{\mu}\psi$ and the $e^{iq\alpha}$ terms times the $e^{-iq\alpha}$, coming from the adjoint spinor derivative, are giving the identity. Physically speaking, this implies that in the system there is a conserved current which, arranging the constants in the appropriate way, can be interpreted as the electromagnetic current $J^{\mu}_{q} = -q\bar{\psi}\gamma^{\mu}\psi$ and a conserved electric charge, obtained by integrating the fourth component of the current on the three-dimensional space.

Since the physical observables are not altered also with a local gauge, the Lagrangian can be studied again in the case where the phase α depends on the coordinates x^μ. In this case the transformation determines a change to the Dirac Lagrangian, which picks up an extra term and so looses its invariance under this transformation. Since the system dynamics, which is governed by the Lagrangian, must be independent from these transformation, the Dirac Lagrangian invariance must be restored. The minimal way to restore the Lagrangian invariance is to substitute each derivative with a covariant derivative, with the form

$$\mathcal{D}_\mu = \partial_\mu - iqA_\mu \quad , \tag{1.6}$$

where the new field A_μ changes under the transformation as $A_\mu \to A_\mu + \partial_\mu \alpha(x)$.

In the end, in order to preserve the invariance of the Lagrangian, a new vector field A_μ, called *gauge field*, must be introduced, together with its dynamics terms. The "local gauge invariant" Dirac Lagrangian then takes the form

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - J_q^\mu A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu \quad . \tag{1.7}$$

In order not to spoil the local gauge invariance, the new field A_μ is required to be massless (that is, the last term in Equation 1.7 must be dropped). What has been presented here is exactly the quantum field description of the electromagnetic interaction (QED), with the boson A_μ identified as the photon, arising from the invariance under the symmetry group $U(1)$.

More generally, all transformations that preserve the field unitarity must be considered, i.e. the transformations $\psi \to \psi' = U\psi$, where U is a $N \times N$ unitary matrix ($U^\dagger U = 1$). Since any unitary matrix can be written in the form $U = e^{iH}$, where H is Hermitian ($H^\dagger = H$), and since the most general Hermitian $N \times N$ matrix can be decomposed in the form $H = \theta\mathbb{I} + \sum_k \alpha_k \cdot t_k$, where $\mathbb{I}$ is the unit matrix and t_k are $N^2 - 1$ traceless hermitian matrices, it follows that that $U(N) = U(1) \times SU(N)$. Having already discussed the $U(1)$ group, the focus will be moved on the global and local $SU(N)$ transformations.

Every element of this group can be written as

$$S = S(\vec{\alpha}) = e^{ig\alpha_k t_k} \qquad k = 1, \ldots, N^2 - 1 \quad , \tag{1.8}$$

where the t_k matrices are identified with the generators of the group and g is a constant that will determine the interaction strength of the field. It is straightforward to prove the invariance under a global $SU(N)$ transformation of any Lagrangian which depends from some field ψ and from the derivatives $\partial_\mu\psi$. However, requiring the invariance to hold also locally turns out to be a little bit more subtle compared to the $U(1)$ case because, in general, the symmetry group is non-Abelian. This implies that the generators follow the non-trivial commutation relations $[t_i, t_j] = if_{ijk}t_k$, where the f_{ijk} are called "structure constants" of the group. Adapting the same idea seen before, the Lagrangian invariance can be restored by introducing a set of new vector fields A_μ^k, and by replacing the usual derivative with the covariant one

$$D_\mu = \partial_\mu - ig\mathcal{A}_\mu \qquad \text{with} \qquad \mathcal{A}_\mu = \sum_{k=1}^{N^2-1} t^k A_\mu^k \quad . \tag{1.9}$$

Substituting the covariant derivative into Equation 1.2, the Lagrangian expression becomes

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^{\mu}D_{\mu}\psi - m\bar{\psi}\psi \tag{1.10}$$

$$= i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi + g\bar{\psi}\gamma^{\mu}t^{k}A^{k}_{\mu}\psi \quad , \tag{1.11}$$

where the last term expresses the coupling between the fermion field and the new vector fields. To the N^2-1 gauge fields A^k_μ must be assigned a transformation rule such that $D_\mu\psi \to S(D_\mu\psi)$. This translates in the requirement that

$$\mathcal{A}_{\mu} \to \mathcal{A}'_{\mu} = S\mathcal{A}_{\mu}S^{-1} - \frac{i}{g}(\partial_{\mu}S)S^{-1} \quad , \tag{1.12}$$

that is

$$A^{k}_{\mu} \to A'^{k}_{\mu} = A^{k}_{\mu} + \partial_{\mu}\alpha^{k} - 2gf^{ijk}\alpha^{i}A^{j}_{\mu} \quad . \tag{1.13}$$

Finally, in order to give these gauge fields a free term, a tensor $F_{\mu\nu}$, antisymmetric in its spatial indexes, has to be introduced. The definition $-igF^k_{\mu\nu}t^k = [D_\mu, D_\nu]$ or, more explicitly

$$F^{i}_{\mu\nu} = \partial_{\mu}A^{i}_{\nu} - \partial_{\nu}A^{i}_{\mu} + gf^{ijk}A^{j}_{\mu}A^{k}_{\nu}$$

preserves the local gauge invariance. Again, the mass of the bosons must be set to zero to avoid the gauge symmetry breaking.

Experiments show that four fundamental forces are present in nature, but all attempts to include all of them into a single renormalizable quantum theory have failed. Since gravity is negligible at a sufficiently small scale, all efforts have been made to include the the electromagnetic, weak and strong interactions in a single theory, the Standard Model. This can be achieved by requiring the invariance under two non-Abelian symmetries, together with an Abelian one.

The first local gauge invariance, under $U(1)$, is associated to the field B_μ.
A second invariance, under a set of non-Abelian transformations that form a $SU(2)$ group, leads to the introduction of three W^i_μ fields ($i = 1, 2, 3$).
The third invariance, also non-Abelian, under a set of transformations that form an $SU(3)$ group, requires the introduction of eight G^a_μ fields ($a = 1, \ldots, 8$).
The general gauge transformation is then given by

$$U = \exp\left[i\left(g'\beta(x)\frac{Y}{2} + g\alpha^{i}(x)\frac{\tau^{i}}{2} + g_s\gamma^{a}(x)\frac{\lambda^{a}}{2}\right)\right] \quad , \tag{1.14}$$

and the covariant derivative, which ensures the invariance of the theory under all the three transformations, takes the form

$$D_{\mu} = \partial_{\mu} - ig'\frac{Y}{2}B_{\mu} - ig\frac{\tau^{i}}{2}W^{i}_{\mu} - ig_s\frac{\lambda^{a}}{2}G^{a}_{\mu} \quad , \tag{1.15}$$

where the scalar Y and the τ^i and λ^i matrices[1] are the generators for the $U(1)$ hypercharge, $SU(2)$ weak isospin and $SU(3)$ colour charge groups, respectively. The way fermions behave under gauge transformations depends on the charge they carry with respect to each interaction:

[1] The matrices τ^i and λ^a are the so called Pauli and Gell-Mann matrices respectively.

- $SU(3)_C$: fermions which have colour charge, are called *quarks* and they appear as colour triplets under $SU(3)$ transformations. The remaining fermions are called *leptons* and transform as colour singlets;
- $SU(2)_L$: recalling the chiral decomposition of ψ into the ψ_L and ψ_R spinors (Equation 1.3), the weak-isospin charge is experimentally found to be different for left and right-handed particles. Left-handed fermions transform as isospin doublets, while right-handed ones are singlets of zero weak isospin, and therefore do not interact with the gauge bosons of this symmetry group. Left-handed fermions are further subdivided according to the third component of the isospin T_3; quarks with $T_3 = +1/2$ are referred as *up-like quarks*, while quarks with $T_3 = -1/2$ are called *down-like quarks*. In the very same fashion, leptons appear as $T_3 = 1/2$ particles, i.e. *neutrinos*, and $T_3 = -1/2$ particles, i.e. *charged leptons*. The chiral nature of the weak isospin transformations has an immediate consequence. Fermion mass terms in the Lagrangian are written as

  $$-m\bar{\psi}\psi = -m\bar{\psi}\left(P_L + P_R\right)\psi = -m\left(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L\right) \quad , \tag{1.16}$$

  which manifestly violate gauge invariance, since ψ_L is a member of an isospin doublet and ψ_R is a singlet. It is then clear that that fermion mass terms cannot be included into the theory in this naive way;
- $U(1)_Y$: the $U(1)$ hypercharge induces transformations as singlets and is non-zero for all fermions except for right-handed neutrinos. As a convention, the corresponding quantum number for left-handed leptons is chosen to be $Y_L = -1$.

Since right-handed neutrinos do not couple to any of the previously introduced interactions, they can be regarded as "sterile" and are not included into the theory.

A part from the new fermion-bosons interactions terms, the Lagrangian must include terms for the free gauge fields

$$-\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^i_{\mu\nu}W^{\mu\nu}_i - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \quad , \tag{1.17}$$

where the second and the last term have important implications since give rises to terms proportional to $W^aW^bW^c$, $W^aW^bW^cW^d$ or $G^aG^bG^c$, $G^aG^bG^cG^d$, which show that gauge bosons associated to a non-Abelian symmetry are self interacting.

Unlike strong interactions, identified with the $SU(3)_C$ symmetry group, the $U(1)_Y$ and $SU(2)_L$ gauge interactions do not directly correspond, respectively, to the electromagnetic and weak forces. Instead, as it will be shown in the next section, the observed interactions are a manifestation of the two groups combination. Such combination is the effect of the spontaneous electroweak symmetry breaking, which causes the W^i and B bosons to coalesce together into four different bosons – the $W^{\pm}_\mu$, the Z^0_μ, and the photon A_μ - as follows

$$\begin{aligned} W^{\pm}_\mu &= \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right) \\ \begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} &= \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix} \quad , \end{aligned} \tag{1.18}$$

where θ_W is the weak mixing angle (Weinberg angle), having a measured value[2] of $\sin^2\theta_W = 0.23116 \pm 0.00012$ [6]. The $SU(2)_L$ and $U(1)_Y$ groups cannot therefore be considered separately, since the two components of the doublets have different electric charge. This fact leads to a relation between electric charge, hypercharge and weak isospin that is mathematically expressed by the Gell-Mann-Nishijima formula

$$Q = T_3 + \frac{Y}{2} \quad . \tag{1.19}$$

Up to this point, not only the fermions are forced to be massless, but also gauge bosons mass terms are not allowed, if the local gauge symmetry has to be preserved. A possible solution for the conflict between massless particles, as required by the theory, and massive fermions and vector bosons, as observed experimentally, can be provided by the spontaneous breaking of the symmetry.

1.1.3 How the Higgs Mechanism explains particle mass

If a theory is described by a Lagrangian which possesses a given symmetry, but its physical vacuum state does not, the symmetry is said to be spontaneously broken. In QFT, in case of broken continuous symmetries, the Nambu-Goldstone theorem [7] ensures that there is an appearance of N massless and spinless particles, where N is equal to the number of broken symmetries. These particles, called Goldstone bosons, can be "eaten" by the gauge bosons, so that the latter become massive and their new, longitudinal polarization is provided by the Goldstone boson.

To understand how the spontaneous symmetry breaking mechanism (SSB) could lead to a massive gauge boson, the case of the QED Lagrangian can be briefly discussed. The $U(1)$ gauge invariant kinetic term of the photon is given by

$$\mathcal{L}_{kin} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad . \tag{1.20}$$

If now the model is extended by introducing a complex scalar field with charge q that couples both to itself and to the photon, the Lagrangian becomes

$$\mathcal{L}_{kin} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \mathcal{L}_H = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \quad , \tag{1.21}$$

where $D_\mu = \partial_\mu - iqA_\mu$ and $V(\phi) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2$. It is easily discerned that this Lagrangian is invariant under the local $U(1)$ gauge transformations

$$A_\mu \rightarrow A_\mu + \partial_\mu \alpha(x) \tag{1.22}$$

$$\phi \rightarrow e^{iq\alpha(x)} \phi \quad . \tag{1.23}$$

If $\mu^2 < 0$, the state of minimum energy corresponds to $\phi = 0$ and the potential preserves the symmetries of the Lagrangian, as it is represented in Figure 1.1. In this case, the theory is simply QED with a massless photon and a charged scalar field with mass μ. However, if $\mu^2 > 0$, the field ϕ will acquire a vacuum expectation value (VEV), with

[2] This is the mixing angle value at a momentum transfer scale $Q = 91.2$ GeV/c, corresponding to the mass of the Z^0 boson.

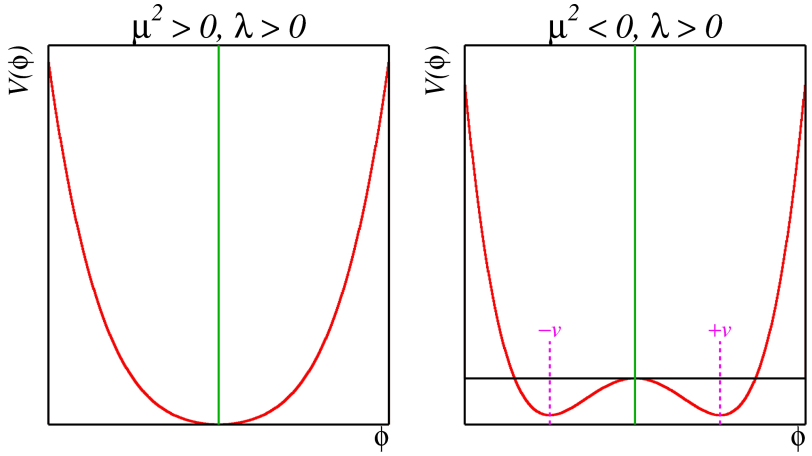


Figure 1.1: Form of the potential $V(\phi)$ depending on the sign of μ^2, positive on the left and negative on the right.

$$\langle 0|\phi|0\rangle = \phi_0 = -\frac{\mu^2}{2\lambda} = \frac{1}{2}v^2 \qquad \text{where} \qquad v \equiv \sqrt{-\frac{\mu^2}{\lambda}} \quad , \tag{1.24}$$

and the $U(1)$ gauge symmetry is spontaneously broken. It is convenient to parameterize ϕ as

$$\phi = \frac{v+h}{\sqrt{2}} e^{i\frac{\chi}{v}} \quad , \tag{1.25}$$

where h and χ, which are referred to as the Higgs boson and the Goldstone boson, respectively, are real fields which have no VEVs. By choosing a particular gauge, called unitary gauge, defined by

$$A_\mu \to A_\mu - \frac{1}{ev}\partial_\mu \chi \quad , \tag{1.26}$$

the Lagrangian becomes

$$\mathcal{L}_{kin} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{qv}{2}A_\mu A^\mu + \frac{1}{2}\partial_\mu h \partial^\mu h - \mu^2 h^2 + \text{higher order terms} \quad . \tag{1.27}$$

The Goldstone boson χ is completely disappeared from the theory and gives to the photon a mass. This Lagrangian now describes a theory with a photon of mass $m_A = qv$ and a scalar Higgs boson h with $m_h = \sqrt{2}\mu = \sqrt{2\lambda}v$. It is relevant to count the degrees of freedom (dof) before and after the spontaneous symmetry breaking has occurred. The original Lagrangian had one massless photon (2 dof) and a complex scalar field (2 dof) for a total number of 4 dof. After the SSB the Lagrangian contains one massive photon (3 dof) and a real scalar field h (1 dof), again for a total of 4 dof.

In the case of the Standard Model, in order to generate masses for the three gauge bosons $W^\pm$ and Z^0, without generating a photon mass, at least three extra degrees of freedom are needed. The simplest realization is to add a complex $SU(2)$ doublet of scalar fields of hypercharge

$Y = 1$, defined by

$$\mathbf{\Phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad . \tag{1.28}$$

This doublet has no colour charge and will therefore not affect the $SU(3)_C$ sector. In analogy to what has been presented for the case of $U(1)$, the Lagrangian for the Higgs field is given in Equation 1.21, with the simple substitution $\phi \to \mathbf{\Phi}$, which again retains its invariance under the gauge group, namely $SU(2)_L \times U(1)_Y$. Again, the spontaneous breaking of the gauge symmetry consists in choosing a particular ground state, around which the Higgs field $\mathbf{\Phi}(x)$ is expanded. In this case, the particular vacuum chosen is

$$\mathbf{\Phi}_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad . \tag{1.29}$$

It is clear that neither T_i nor Y cancels $\mathbf{\Phi}_0$, in particular

$$T_3 \mathbf{\Phi}_0 = -\frac{1}{2} \mathbf{\Phi}_0 \qquad \text{and} \qquad Y \mathbf{\Phi}_0 = \mathbf{\Phi}_0 \quad .$$

On the contrary, since $\mathbf{\Phi}_0$ is neutral, the $U(1)_Q$ symmetry remains unbroken, that is

$$Q\mathbf{\Phi}_0 = \left(T_3 + \frac{Y}{2} \right) \mathbf{\Phi}_0 = 0 \qquad \Rightarrow \qquad \mathbf{\Phi}_0 \to \mathbf{\Phi}_0' = e^{i\delta(x)\frac{Q}{2}} \mathbf{\Phi}_0 = \mathbf{\Phi}_0 \quad . \tag{1.30}$$

Thus, $SU(2)_L$ and $U(1)_Y$ are completely broken separately, but the product group $SU(2)_L \times U(1)_Y$ is not: after symmetry breaking, there remains a residual symmetry generated by Q. This pattern of symmetry breakdown is then described in formula by $SU(2)_L \times U(1)_Y \to U(1)_Q$.

By denoting the fluctuation of the ϕ_1, ϕ_2, ϕ_3 and ϕ_4 real scalar fields around the minimum as θ_2, θ_1, H and $-\theta_3$ then, in the small fluctuation approximation, the following formula holds

$$\begin{aligned} \mathbf{\Phi}(x) &= \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_2(x) + i\theta_1(x) \\ v + H(x) - i\theta_3(x) \end{pmatrix} \\ &\simeq \frac{1}{\sqrt{2}} \begin{pmatrix} 1 + i\theta_3/v & i(\theta_1 - i\theta_2)/v \\ i(\theta_1 + i\theta_2)/v & 1 - i\theta_3/v \end{pmatrix} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \\ &\simeq \frac{1}{\sqrt{2}} e^{i\frac{2\theta^i(x)}{v}\frac{\tau^i}{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad . \end{aligned} \tag{1.31}$$

As before, the massless Goldstone bosons $\theta_i(x)$ can be gauged away, with the transformation $U = \exp\left(-i\frac{2\theta^i(x)}{v}\frac{\tau^i}{2}\right)$, and they do not appear in the final Lagrangian. By expanding the scalar Higgs field in the Higgs Lagrangian around $\mathbf{\Phi_0}$ using

$$\mathbf{\Phi}(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.32}$$

and by reinstating the boson kinematic terms one finds

$$\begin{aligned}\mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{Higgs}} = &\frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2}2v^2\lambda H^2 - \frac{1}{3!}6v\lambda H^3 - \frac{1}{4!}6\lambda H^4 \\ &+ \frac{1}{2}\frac{v^2 g^2}{4} W_\mu^{-\dagger} W^{-\mu} + \frac{1}{2}\frac{v^2 g^2}{4} W_\mu^{+\dagger} W^{+\mu} \\ &+ \frac{1}{2}\frac{v^2(g^2+g'^2)}{4}\left(\frac{gW_\mu^3 - g'B^\mu}{\sqrt{g^2+g'^2}}\right)^2 + 0\cdot\left(\frac{g'W_\mu^3 + gB^\mu}{\sqrt{g^2+g'^2}}\right)^2 \\ &+ \frac{1}{4}\left(2vH + H^2\right)\left[g^2 W_\mu^- W^{+\mu} + \frac{1}{2}(g^2+g'^2)\left(\frac{W_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}}\right)^2\right] \quad .\end{aligned} \tag{1.33}$$

The first line, originated from the expansion of the potential $V(\mathbf{\Phi})$, contains the kinetic term for the Higgs boson, its mass term and the Higgs boson self-interaction terms. Even if v can be put in relation to the Fermi constant G_F and therefore estimated from precise muon lifetime measurements, i.e. $v = \left(\sqrt{2}G_F\right)^{-1/2} \simeq 247$ GeV, λ is a free parameter of the model, hence the Higgs mass is unknown.

In the second line, coming from the kinetic term $\left(\mathcal{D}_\mu \mathbf{\Phi}\right)^\dagger \left(\mathcal{D}^\mu \mathbf{\Phi}\right)$, the $\mathrm{W}^\pm$ vector bosons can be identified in the linear combination of the gauge bosons $W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2)$. The process of spontaneous symmetry breaking allows them to acquire mass.

The third line provides the mass terms for the observed Z^0 and γ vector bosons[3]. The first linear combination of the gauge fields W_μ^3 and B_μ comes with an appropriate mass term and it is therefore interpreted as the massive Z^0 boson. The second combination of fields is orthogonal to the first one and it is added "by hand" with an associated null mass.
Then, the results can be interpreted as

$$\begin{aligned} m_{\mathrm{W}} &= \frac{1}{2}vg && \text{with} && W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \\ m_{\mathrm{Z}} &= \frac{1}{2}v\sqrt{g^2+g'^2} && \text{with} && Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}} \\ m_\gamma &= 0 && \text{with} && A_\mu = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2+g'^2}} \quad . \end{aligned} \tag{1.34}$$

The gauge bosons have "eaten" the three massless Goldstone bosons, and have acquired mass. Again, the degrees of freedom of the Goldstone bosons are needed by the three gauge bosons, since once they become massive, an additional degree of freedom is required in order to allow them to have a longitudinal polarization. A last note is that the remaining unbroken $U(1)_Q$ symmetry causes the photon to remain massless.
The mixing of W_μ^3 and B_μ, yielding the physical force carriers, can be interpreted as a rotation

[3]The numerical factor $\sqrt{g^2+g'^2}$ has been introduced in order to normalize the combinations of the gauge fields $gW_\mu^3 - g'B_\mu$ and $g'W_\mu^3 + gB_\mu$.

of angle θ_W, where

$$\frac{g}{\sqrt{g^2+g'^2}} = \cos\theta_W \qquad \text{and} \qquad \frac{g'}{\sqrt{g^2+g'^2}} = \sin\theta_W \quad . \tag{1.35}$$

Therefore, the following relation between the weak bosons masses can be inferred

$$m_Z = \frac{m_W}{\cos\theta_W} \quad . \tag{1.36}$$

Finally, in the last line of the Lagrangian 1.33, the cubic and quartic couplings of the Higgs boson to the weak gauge bosons can be deduced. In particular, the coupling of one single Higgs boson to a pair of W or Z bosons is proportional to the square of m_W and m_Z respectively, and reads as

$$g_{HWW} = \frac{m_W^2}{v} \cdot 2 \tag{1.37}$$

$$g_{HZZ} = \frac{m_Z^2}{v} \quad . \tag{1.38}$$

From this, the following relation can be derived for the branching ratio $\mathcal{B}$ of the Higgs boson into a pair of vector bosons (valid at tree level for Higgs boson masses well above the kinematic threshold for the production of a diboson pair)

$$\frac{\mathcal{B}(H\to WW)}{\mathcal{B}(H\to ZZ)} = \left(\frac{g_{HWW}}{g_{HZZ}}\right)^2 = 4\frac{m_W^4}{m_Z^4} \simeq 2.4 \quad . \tag{1.39}$$

The full Standard Model Lagrangian (neglecting the colour part) can then be written as

$$\mathcal{L}_{SM} = \mathcal{L}_{GWS} + \mathcal{L}_{Higgs} \quad , \tag{1.40}$$

where the electroweak part of it (representing the Glashow-Weinberg-Salam model of electroweak unification), noting leptons as e and neutrinos as ν, is given by

$$\begin{aligned}\mathcal{L}_{GWS} = &-\frac{1}{4}W^i_{\mu\nu}W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ &+ i\bar{\nu}_L\gamma^\mu\partial_\mu\nu_L + i\bar{e}_L\gamma^\mu\partial_\mu e_L + i\bar{e}_R\gamma^\mu\partial_\mu e_R + \\ &- i\bar{f}_L\gamma^\mu\left(-ig\frac{\tau^i}{2}W^i - ig'\frac{Y}{2}B_\mu\right)f_L - i\bar{e}_R\gamma^\mu\left(-g'\frac{Y}{2}B_\mu\right)e_R \quad .\end{aligned} \tag{1.41}$$

Re-expressing the interaction part of the above Lagrangian in terms of the physics fields and writing explicitly the covariant derivative, one obtains

$$\begin{aligned}\mathcal{L}^{int}_{GWS} &= \mathcal{L}^{int}_{CC} + \mathcal{L}^{int}_{NC} \\ &= \left\{eJ^{em}_\mu A^\mu + \frac{g}{\cos\theta_W}J^Z_\mu Z^\mu\right\} + \left\{\frac{g}{\sqrt{2}}\left(J^+_\mu W^{+\mu} + J^-_\mu W^{-\mu}\right)\right\} \quad ,\end{aligned} \tag{1.42}$$

for the neutral and charged part respectively. The electromagnetic coupling constant e has been

introduced, identifying $e = g \sin\theta_W$. The following currents have also been defined

$$\begin{aligned} J_\mu^{em} &= Q\bar{f}\gamma_\mu f\,, \\ J_\mu^Z &= \frac{1}{2}\bar{f}\gamma_\mu\left(c_V^f - c_A^f\gamma_5\right) f \qquad \text{with} \qquad c_V^f = T_3 - 2Q\sin^2\theta_W,\ c_A^f = T_3 \\ J_\mu^+ &= \frac{1}{2}\bar{\nu}\gamma_\mu\left(1-\gamma_5\right) e \quad . \end{aligned} \tag{1.43}$$

1.1.3.1 Fermion masses

One of the most striking features of the Standard Model is that the same Higgs doublet which generates the W and Z masses is also sufficient to give mass to leptons and quarks. For the lepton sector, for instance, the following Lagrangian can be added (for each lepton generation ℓ)

$$\mathcal{L}_{\text{Yukawa}}^{\ell} = -G_\ell\left[(\bar{\boldsymbol{\ell}}_L\boldsymbol{\Phi})\ell_R + \bar{\ell}_R(\boldsymbol{\Phi}^\dagger\boldsymbol{\ell}_L)\right] \quad , \tag{1.44}$$

where the Higgs doublet has exactly the required $SU(2)_L \times U(1)_Y$ quantum numbers to couple to $\bar{\boldsymbol{\ell}}_L\ell_R$. After the breakdown of the symmetry, by inserting Equation 1.32 into 1.44, one obtains

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}}^{\ell} &= -\frac{G_\ell}{\sqrt{2}}\left\{(\bar{\nu}_\ell,\bar{\ell})_L\begin{pmatrix}0\\ v+H\end{pmatrix}\ell_R + \bar{\ell}_R\,(0, v+H)\begin{pmatrix}\nu_\ell\\ \ell\end{pmatrix}_L\right\} \\ &= -\frac{G_\ell}{\sqrt{2}}\left\{v(\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L) + (\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L)H\right\} \\ &= -\frac{G_\ell}{\sqrt{2}}\left\{v\bar{\ell}\ell + \bar{\ell}\ell H\right\} \quad . \end{aligned} \tag{1.45}$$

In Equation 1.45, G_ℓ (called Yukawa coupling) is chosen so that the required lepton mass term appears explicitly, that is

$$m_\ell = \frac{G_\ell v}{\sqrt{2}} \tag{1.46}$$

and the Yukawa Lagrangian for the lepton sector can be rewritten as

$$\mathcal{L}_{\text{Yukawa}}^{\ell} = -m_\ell\left\{\bar{\ell}\ell + \frac{1}{v}\bar{\ell}\ell H\right\} \quad . \tag{1.47}$$

This technique allows to generate a mass term for leptons and down-like quarks. For up-like quarks (and neutrinos), a different Higgs doublet has to be introduced, defined as $\boldsymbol{\Phi}_c = -i\tau_2\boldsymbol{\Phi}^*$.

Not only have fermions acquired mass, thanks to the spontaneous symmetry breaking, but a new coupling between the Higgs boson and the fermions has become manifest. As an important consequence, in case the Higgs boson is observed in an experiment, the amplitude of a Higgs decay process will be proportional to the second power of the mass of the fermion the Higgs decays into (see also Equations 1.38,1.47).

1.1.4 The Periodic Table of particles

Twelve fundamental fermions have been experimentally observed up to today. These particles come in three families or generations, that behave almost identically under interactions. The three known lepton families are the electron (e), the muon (μ) and the tau (τ). Each of them comes with its associated neutrino (ν_e, ν_μ and ν_τ). The six quark flavours are subdivided in up-like quarks, labeled as up (u), charm (c) and top (t) and down-like quarks named down (d), strange (s) and bottom (b). Each of these fermions is in addition accompanied by an anti-particle with opposite quantum numbers and exactly the same couplings of its counterpart.
Up to now, quarks have only been observed confined into bound states of $q\bar{q}$ pairs, called mesons, or qqq/ $\overline{qqq}$ aggregates, called baryons.
All fermion species are summarized, together with their main properties, in Table 1.1.

Similarly, four categories of gauge bosons have been observed: the massless photons (γ) and gluons (g) carry the electromagnetic and strong interactions respectively; the weak force carriers are massive and divided into charged bosons, $W^\pm$, and neutral bosons, Z. Their properties are summarized in Table 1.2.

	1st gen.		2nd gen.		3rd gen.		Q	Color
leptons	ν_e	~ 0	ν_μ	~ 0	ν_τ	~ 0	0	no
	e	$511\,\mathrm{keV}/c^2$	μ	$105.7\,\mathrm{MeV}/c^2$	τ	$1.777\,\mathrm{GeV}/c^2$	-1	no
quarks	u	$1.7-3.1\,\mathrm{MeV}/c^2$	c	$1.29^{+0.05}_{-0.11}\,\mathrm{GeV}/c^2$	t	$172.9^{+1.1}_{-1.1}\,\mathrm{GeV}/c^2$	2/3	yes
	d	$4.1-5.7\,\mathrm{MeV}/c^2$	s	$100^{+30}_{-20}\,\mathrm{MeV}/c^2$	b	$4.19^{+0.18}_{-0.06}\,\mathrm{GeV}/c^2$	-1/3	yes

Table 1.1: Spin-$\frac{1}{2}$ matter constituents. The second, third and fourth columns contain the fermion symbol and measured mass.

	mass	Q	Color
γ	0	0	no
g	0	0	yes
$W^\pm$	$80.385 \pm 0.015\,\mathrm{GeV}/c^2$	± 1	no
Z	$91.188 \pm 0.002\,\mathrm{GeV}/c^2$	0	no

Table 1.2: Spin-1 gauge bosons.

1.2 The search for the Higgs boson

As was presented in Section 1.1, the SM cannot predict the Higgs mass m_{H}, but nevertheless constraints can be set on m_{H} based upon theoretical considerations and experimental searches.

1.2.1 Theoretical constraints on the Higgs mass

A first upper constraint on m_{H} is found considering the weak boson scattering process $\mathrm{W}_L\mathrm{W}_L \rightarrow \mathrm{W}_L\mathrm{W}_L$ [8]. In the assumption that no Higgs boson is present in nature, the amplitude for this process would be proportional to the center-of-mass energy, and thus violate unitarity at high energy, which happens at $\sqrt{s} \sim 1$ TeV. Diagrams involving the exchange of an Higgs among the W_L bosons allow for cancellations so that the scattering amplitude is regularized and finite at all energies, provided that $m_{\mathrm{H}} \lesssim 700$ GeV/c^2.

Further constraints are coming from the fact that the Higgs sector runs its course between two problematic situations if it is to survive up to the reduced Planck scale $M_P \sim 2 \cdot 10^{18}$ GeV, where new physics associated with quantum gravity must surely appear [9]. If m_{H} is large enough, the renormalization-group equations (RGEs) of the SM drive the Higgs self-coupling into the non-perturbative regime at some scale $\Lambda < M_P$, entailing either new non-perturbative physics at a scale $\sim \Lambda$, or new physics at some scale $< \Lambda$, that prevents the Higgs self-coupling from blowing up (perturbativity or triviality bound).
On the other hand, if m_{H} is small enough, the RGEs drive the Higgs self-coupling to a negative value at some scale $\Lambda < M_P$, in which case the electroweak vacuum is only a local minimum and there is a new, deep and potentially dangerous minimum at scales $> \Lambda$ (stability bound). The electroweak vacuum can potentially become unstable and collapse (either because of zero-temperature (quantum) or thermal tunneling during the evolution of the Universe) into that deeper new vacuum with Higgs vacuum expectation value $> \Lambda$, unless there is new physics at some scale $< \Lambda$, that prevents the appearance of that vacuum. Below this stability bound, there is a region denoted as 'metastability' region, where the electroweak vacuum has a lifetime, for decay via either zero-temperature quantum fluctuations or thermal fluctuations, longer than the age of the Universe .

The triviality and stability bounds, together with the SM survival region, are represented in Figure 1.2 as a function of the Λ scale. As it can be seen, this plot suggests that if the Standard Model validity extends up to the scale of Grand Unification Theories ($\Lambda_{\mathrm{GUT}} \simeq 10^{16}$ GeV), the Higgs boson mass has to lie roughly in the $130 - 180$ GeV/c^2 range. Conversely, for an Higgs particle lighter than 130 or heavier than 180 GeV/c^2, new Physics is expected to exist at an energy scale inferior to Λ_{GUT}.

1.2.2 Indirect experimental constraints on the Higgs mass

Since the Higgs boson enters into one-loop radiative corrections in the Standard Model, precise electroweak measurements can bound the Higgs mass. Due to the logarithmic dependence on m_{H} in all one-loop electroweak parameters, the limits derived from this method are relatively weak. In contrast, the top quark contributes quadratically to many observables. From precision measurement at the Large Electron-Positron Collider (LEP) and the Stanford Linear Collider (SLC) of many electroweak observables, and direct pre-LHC measurements of m_{W} and m_{t} at LEP-II and Tevatron, the plots of Figure 1.3 can be obtained, where contour curves of 68%

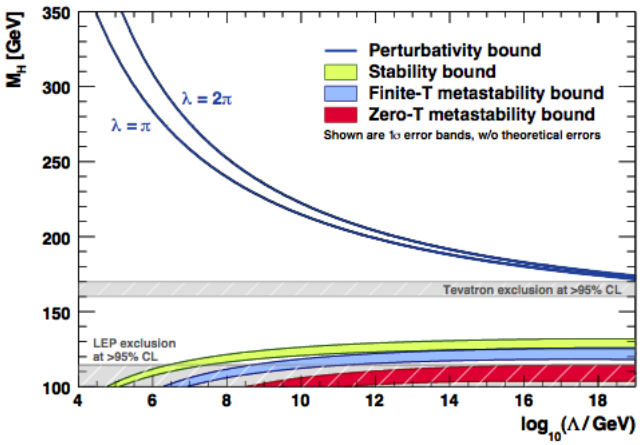
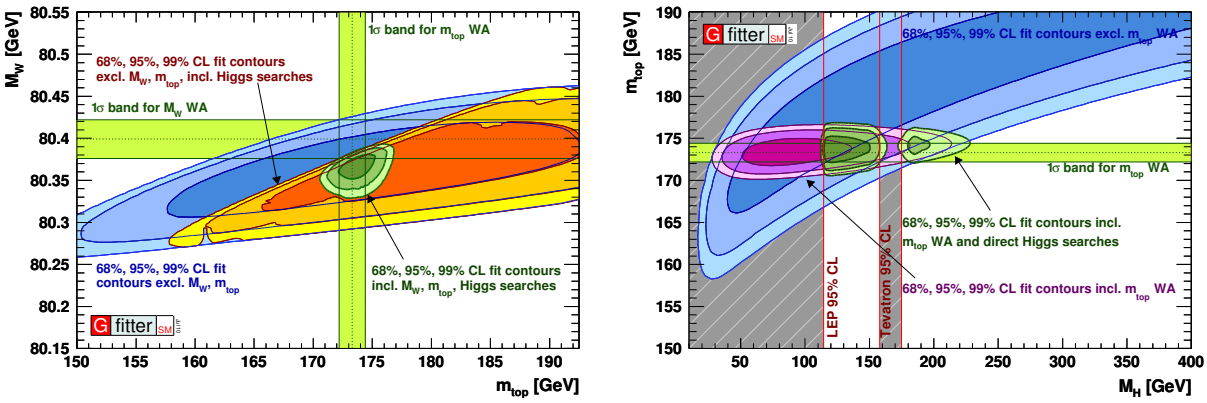


Figure 1.2: The scale Λ at which the two-loop RGEs drive the quartic SM Higgs coupling non-perturbative, and the scale Λ at which the RGEs create an instability in the electroweak vacuum ($\lambda < 0$). The width of the bands indicates the errors induced by the uncertainties in the top mass pole m_{top} and the strong coupling constant α_S (added quadratically). The perturbativity upper bound (referred to as 'triviality' bound) is given for $\lambda = \pi$ (lower blue bold line) and $\lambda = 2\pi$ (upper blue bold line). Their difference indicates the size of the theoretical uncertainty in this bound. The absolute vacuum stability bound is displayed by the light shaded green band, while the less restrictive finite-temperature and zero-temperature metastability bounds are medium blue and dark shaded red, respectively. The theoretical uncertainties for these bounds have been ignored in the plot.

probability are shown in $(m_{\mathrm{W}}, m_{\mathrm{t}})$ and $(m_{\mathrm{t}}, m_{\mathrm{H}})$ planes.



(a) (b)

Figure 1.3: Contour curves of 68% and 95% probability in (a) the $(m_{\mathrm{W}}, m_{\mathrm{t}})$ plane and (b) the $(m_{\mathrm{t}}, m_{\mathrm{H}})$ plane, based on various LEP, LEP-II, SLC, and Tevatron measurements. Contours are shown for different sets of input, i.e. without direct Higgs searches or $m_{\mathrm{W}}/m_{\mathrm{t}}$ measurements. The direct measurements of vector bosons masses are shown as green bands of width ± 1 standard deviation. The vertical grey band in (b) shows the 95% confidence level exclusion limit on m_{H}, derived from direct searches at LEP-II [10] and Tevatron [11].

The whole ensemble of electroweak measurements are used as input to perform a global fit. The fit consists in a χ^2 minimization, where the χ^2 is calculated comparing the measured values of 20 different variables and their uncertainties with their predictions calculated within the framework of the Standard Model. Out of the ensemble of the model parameters, the following are free to vary in the fit: m_Z, m_{H}, m_c, m_b, m_{t}, $\Delta\alpha_{had}(m_Z^2)$, $\alpha_S(m_Z^2)$ and four additional parame-

ters, accounting for theoretical uncertainties.
The standard fit converges at the global minimum value $\chi^2_{min} = 16.37$ for 13 degrees of freedom, giving an excellent quantitative measure of how well the Standard Model is able to describe the complete set of phenomena used as input parameters. Moreover, the fit yields a best value expected for the only unknown parameter m_{H}. The result of the fit is reported in Figure 1.4, where the $\Delta\chi^2(m_{\mathrm{H}}) = \chi^2{}_{\mathrm{min}}(m_{\mathrm{H}}) - \chi^2{}_{\mathrm{min}}$ curve is shown.

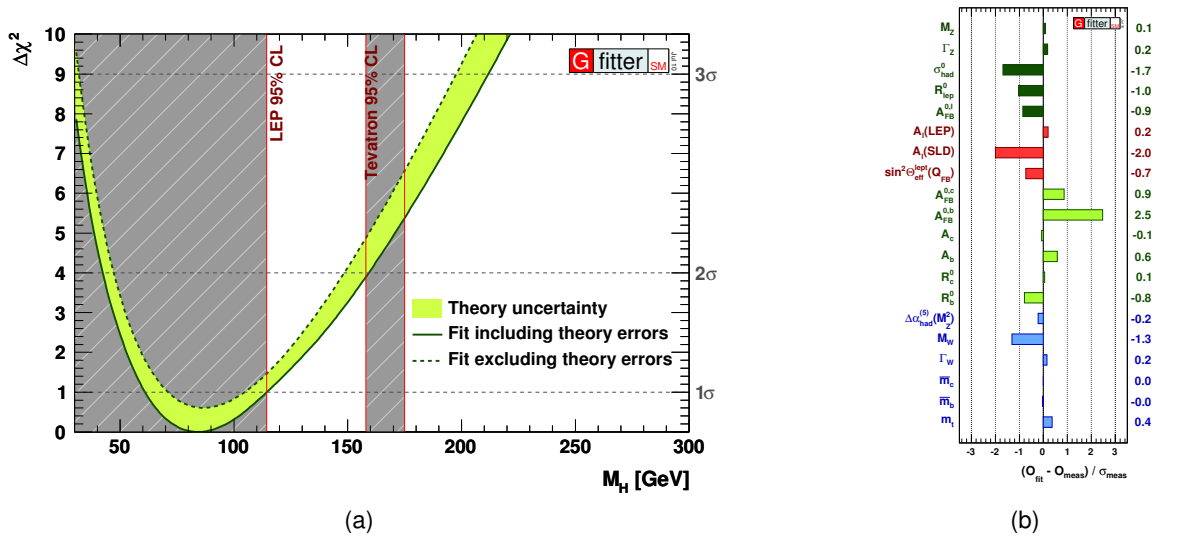


Figure 1.4: Global fit of all precision electroweak measurements. (a) The black continuous line shows $\Delta\chi^2(m_{\mathrm{H}}) = \chi^2{}_{\mathrm{min}}(m_{\mathrm{H}}) - \chi^2{}_{\mathrm{min}}$ as a function of m_{H}. The green area represents the associated theoretical uncertainty due to missing higher-order corrections. The grey shaded area corresponds to the 95% C.L. exclusion limit $m_{\mathrm{H}} > 114.4\,\mathrm{GeV}/c^2$ (LEP-II, 2003 [10]) and $m_{\mathrm{H}} \notin [158, 175]\,\mathrm{GeV}/c^2$ (Tevatron, 2010 [11]). (b) Comparison of all the 20 measurements with their SM expectations calculated using as model parameters the values in the minimum of the global χ^2 of the fit.

From the result of the fit it is clear that electroweak measurements seem to favour a light Higgs. It is also true that, since these results arises from loop corrections, they may be circumvented by some still unknown new Physics entering the same loops.

1.2.3 Direct constraints from pre-LHC experiments

Experimental bounds on m_{H} coming from pre-LHC era were provided by measurements at two different colliders.

The first direct search was carried out at the LEP accelerator located at the European Organization for Nuclear Research (CERN) [10], where data from e^+e^- collisions at a center-of-mass energy up to 209 GeV were recorded and analyzed by the four experiments installed on the accelerator. The main Higgs production mechanism at an e^+e^- collider is the so-called Higgs-strahlung, where an Higgs boson is radiated by a virtual Z boson, and the most probable final state consists in a couple of b-jets from the Higgs decay and another couple of jets from the Z boson decay (although also the leptonic decay modes of the Z were considered). The experiments combined the results obtained with the following channels (sorted for relative importance in

terms of discovery sensitivity):

$$\begin{aligned} &e^+e^- \to Z^*/\gamma^* \to \quad Z(\to q\bar{q}) \quad H(\to b\bar{b}) \\ &e^+e^- \to Z^*/\gamma^* \to \quad Z(\to \nu\bar{\nu}) \quad H(\to b\bar{b}) \\ &e^+e^- \to Z^*/\gamma^* \to \quad Z(\to l^+l^-) \quad H(\to b\bar{b}) \qquad \text{with} \qquad l = e,\mu,\tau \quad . \end{aligned} \tag{1.48}$$

No significant excess of events with an Higgs-compatible signal was found, thus allowing to set a lower bound on the Higgs boson mass at 95% confidence level (as shown in Figure 1.5a):

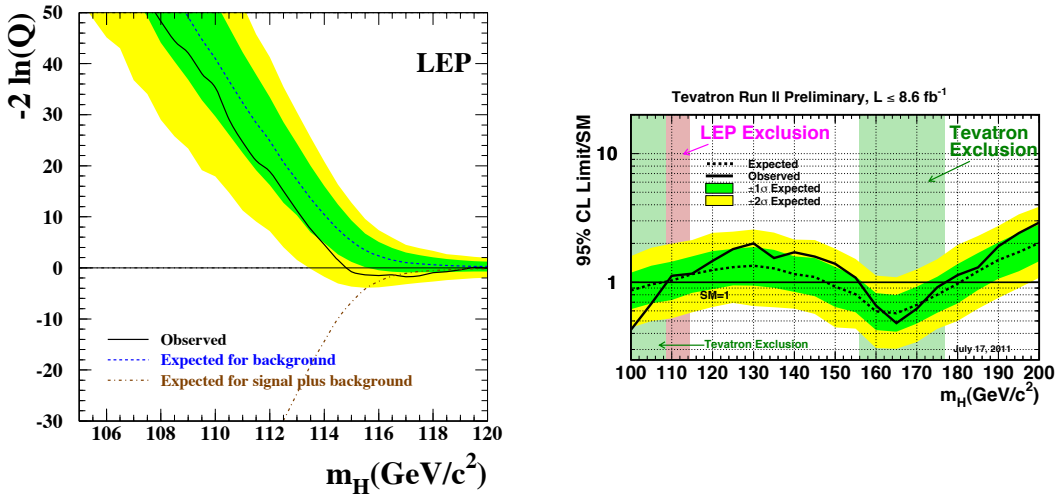
$$m_H > 114.4\,\text{GeV}/c^2 \quad \text{(LEP-II – 2003)} \quad . \tag{1.49}$$

The second direct search for the Standard Model Higgs particle was carried out at Tevatron, a $p\bar{p}$ collider with a center-of-mass energy of $\sqrt{s} = 1.96$ TeV. Similarly to LEP, at the Tevatron the main production mechanism consists in Higgs boson production in association with a vector boson (W or Z), whose decay products are used to tag the event. A large variety of Higgs decay channels was considered by the CDF and DØ experiments, though the driving ones were the following five:

$$\begin{aligned} &H \to b\bar{b} \\ &H \to \gamma\gamma \\ &H \to WW \to q\bar{q}\,q\bar{q} \\ &H \to WW \to l\nu_l\,q\bar{q} \\ &H \to WW \to l\nu_l\,l\nu_l \qquad \text{with} \qquad l = e,\mu,\tau \quad . \end{aligned} \tag{1.50}$$

As reported in Figure 1.5b, the combination of all analyses [12] leads to an exclusion of the SM Higgs boson particle in a mass region around 160 GeV/c^2 (mostly driven by $H \to WW$ final states) and at very low mass, in a region already covered by LEP results, namely

$$\begin{aligned} &m_H \notin [147, 179]\,\text{GeV}/c^2 \\ &m_H \notin [100, 106]\,\text{GeV}/c^2 \end{aligned} \quad \text{(Tevatron – 2011)} \quad . \tag{1.51}$$



(a) (b)

Figure 1.5: Pre-LHC era direct Higgs searches results. (a) Observed and expected behaviour of the test statistics $-2\ln Q$ as a function of the Higgs mass. The test statistics Q is defined as the ratio between the signal plus background likelihood and the background only likelihood. The result is the combination of the data collected by the four LEP experiments. Green and yellow shaded bands represent the 68% and 95% probability C.L. The observed trend in data (solid black line) is compatible with a background-only hypothesis (dashed blue line) up to 114.4 GeV/c^2. (b) Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on the ratios to the SM cross-section, as a function of the Higgs boson mass for the combined CDF and DØ analyses. The limits are expressed as a multiple of the SM prediction for test masses for which both experiments have performed dedicated searches in different channels. The bands indicate the 68% and 95% probability regions where the limits can fluctuate, in the absence of signal. The limits displayed in this figure are obtained with the Bayesian calculation.

1.3 Higgs searches at the Large Hadron Collider

The LHC is a $\sqrt{s} = 8$ TeV [4] proton-proton collider situated at CERN. In this section the main Higgs production modes at the LHC will be presented as well as an overview of the most interesting final states for the detection of the Higgs boson at the Compact Muon Solenoid (CMS). A more detailed description of LHC and CMS is reported in Chapter 2.

1.3.1 Higgs boson production mechanisms

The SM Higgs boson main production mechanisms at the LHC are reported in Figure 1.6 at tree level. These diagrams correspond, sorted in order of relative cross-section magnitude, to:

- **gluon fusion** ($gg \to H$), where the Higgs production from two incoming gluons is mediated by an heavy quark triangle loop. Because of the Higgs couplings to fermions, the t-quark loop is the most important, with the secondary contributions being at least a factor $\mathcal{O}(m_t^2/m_b^2)$ smaller. Next to Leading Order (NLO) corrections have been shown to increase the LO cross-section of about $80-100\%$ at the LHC, both in the large-m_t limit ([13, 14]) and maintaining the full m_t dependence [15]. Also NNLO computations have become available recently in the large m_t limit, and they provide a further $\sim 25\%$ enhancement, depending on the Higgs mass [16];
- **vector boson fusion** (VBF, $qq \to qqH$), where the Higgs boson is created through the fusion of two weak bosons radiated off the incoming quarks. Although this process is about one order of magnitude below the gluon fusion, it can be easily tagged and discriminated from the backgrounds, since the quark pair hadronizes and produces two forward jets of high invariant mass. Moreover, another interesting property is the reduced hadronic activity in the region in between the tag jets, since they are colour disconnected. NLO Quantum ChromoDynamics (QCD) and ElectroWeak (EWK) corrections to this production process are found to be modest (order of 5-10%) and roughly of the same order [17]. Approximate NNLO QCD corrections to the total inclusive cross section for VBF have been presented [18] and shown to reduce the scale dependence of the result to only 1–2%;
- **Higgs-strahlung** ($q\bar{q}' \to WH$, $q\bar{q} \to ZH$) and $t\bar{t}$ **associated production** ($gg, q\bar{q} \to t\bar{t}H$) are processes where the Higgs is produced in association with a W/Z boson or a pair of top quarks. In both cases, their decay products can be used to tag the event, making these production modes experimentally significant, even if the cross-section is orders of magnitude smaller with respect to the gluon fusion one [19, 20]. The QCD corrections are large and the k-factors range from about 1.2 to 1.4, depending on the Higgs mass and on the parton distribution functions chosen for the calculations.

The total SM Higgs boson production cross-section at a pp hadron collider is reported in figure in Figure 1.7 [21] for a center-of-mass energy of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV.

[4] The machine operated at $\sqrt{s} = 7$ TeV during the 2011 run.

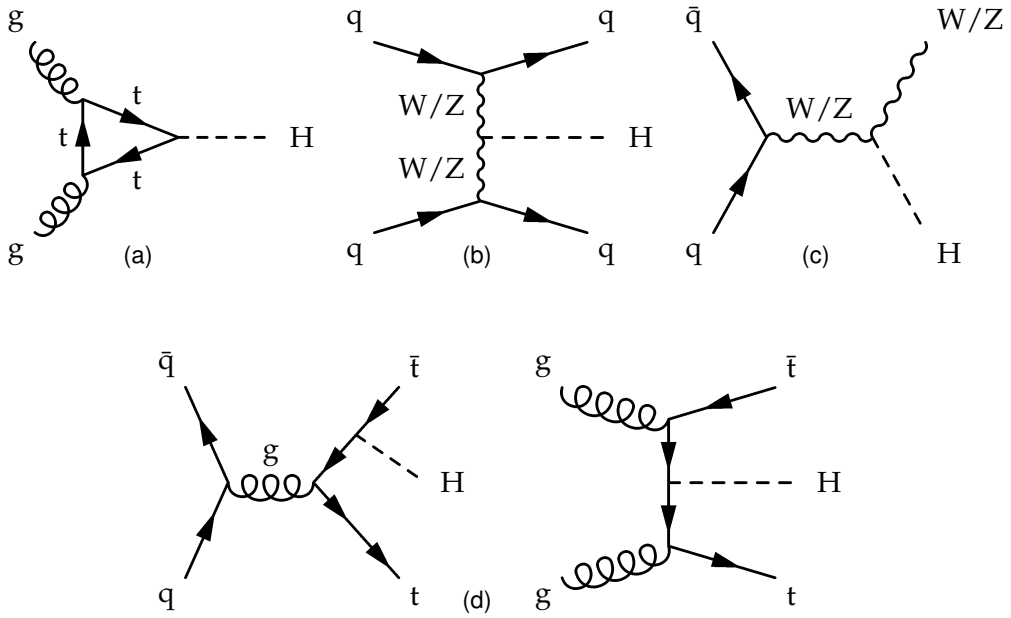


Figure 1.6: LO Feynman diagrams for the most important production processes of the SM Higgs boson at the LHC: (a) gluon fusion, (b) vector boson fusion, (c) Higgs-strahlung, (d) $t\bar{t}$ associated production.

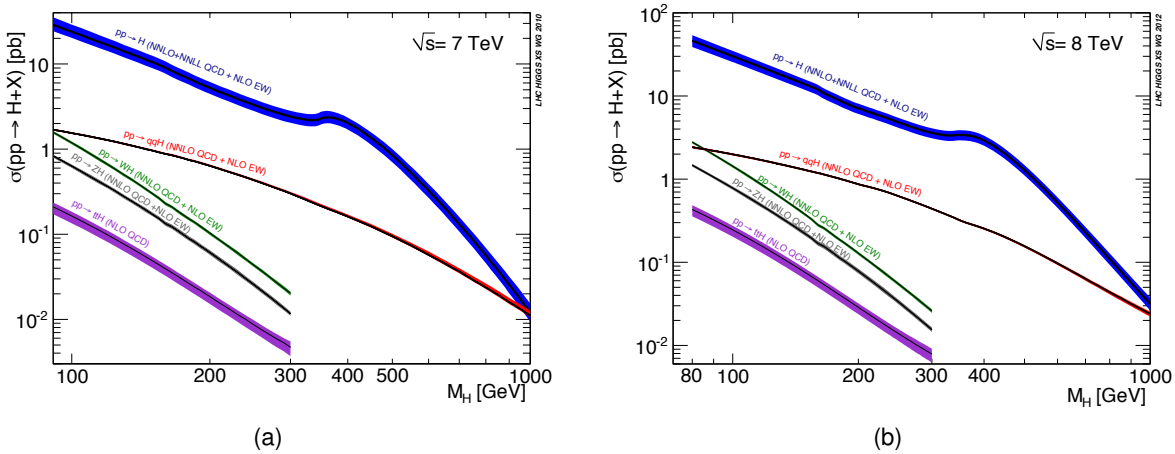


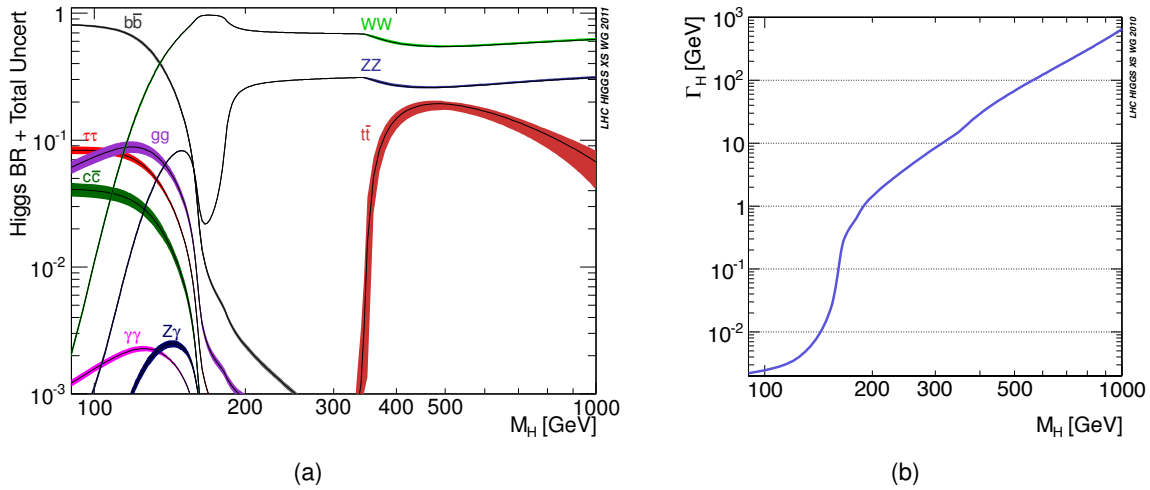
Figure 1.7: Higgs production cross-sections at $\sqrt{s} = 7$ TeV (a) and $\sqrt{s} = 8$ TeV (b) as a function of the Higgs mass for different production mechanisms. From top to bottom, sorted for their relevance: gluon fusion (blue), VBF (red), associate production with a W/Z boson (green/grey), $t\bar{t}$ associated production (violet). NNLO QCD corrections as well as NLO EKW corrections are taken into account.

1.3.2 Higgs boson decay modes

The Higgs total decay width and its different decay branching ratios depend on the Higgs couplings to the vector bosons and to the fermions in the Standard Model Lagrangian, given in the previous section. As a thumb rule, due to the dependence of the Higgs couplings on particle masses, the Higgs tends to decay into the heaviest particles which are kinematically allowed. In addition, whenever a decay into vector bosons becomes kinematically accessible,

it is favoured since the W/Z-Higgs coupling has a quadratic mass dependency to the vector bosons masses. This general overview is summarized in Figure 1.8a, which shows the Higgs decay branching ratios including also NLO QCD and EWK corrections [22]. Light-fermion decay modes contribute only in the low mass region (up to $\sim 150\,\mathrm{GeV}/c^2$), where the branching ratio is dominated by the channel $H \to b\bar{b}$. Once the decay into a pair of weak bosons is possible, it quickly dominates. A peak in the WW decay mode is visible, when the production of two on-shell W bosons becomes possible, while a Z pair is still not accessible. At high masses ($\gtrsim 350\,\mathrm{GeV}/c^2$) also $t\bar{t}$ pairs can be produced. Finally, even if the Higgs boson does not couple to photons and gluons at tree level, such couplings can arise via fermion loops and they give a tiny contribution in the low mass region.

The total width, given by the sum over all the possible decay channels, is shown in Figure 1.8b. It quickly increases with the Higgs mass due to the opening of new channels and it becomes almost as large as the Higgs mass itself around $1\,\mathrm{TeV}/c^2$, where the interpretation of the Higgs resonance as a particle loses meaning: such an heavy Higgs would be more suitably accommodated in theories extending the Standard Model.



(a) (b)

Figure 1.8: (a) Decay branching ratios of the SM Higgs boson in the different channels versus its mass. (b) Total decay width (in GeV) of the SM Higgs boson with respect to its mass.

1.3.3 Experimental features of the Higgs decay modes

The low mass Higgs driving discovery channel is the $H \to \gamma\gamma$ decay mode, which has a very clear experimental signature. The main cons are the small signal rate compared to the backgrounds coming both from two prompt photons (irreducible), and from those in which one or more of the photons are due to decay products or mis-identified particles in jets (reducible). The main feature is that the signal would appear as a very narrow mass peak above a large background, which can be determined from a sideband region in the $m_{\gamma\gamma}$ spectra.

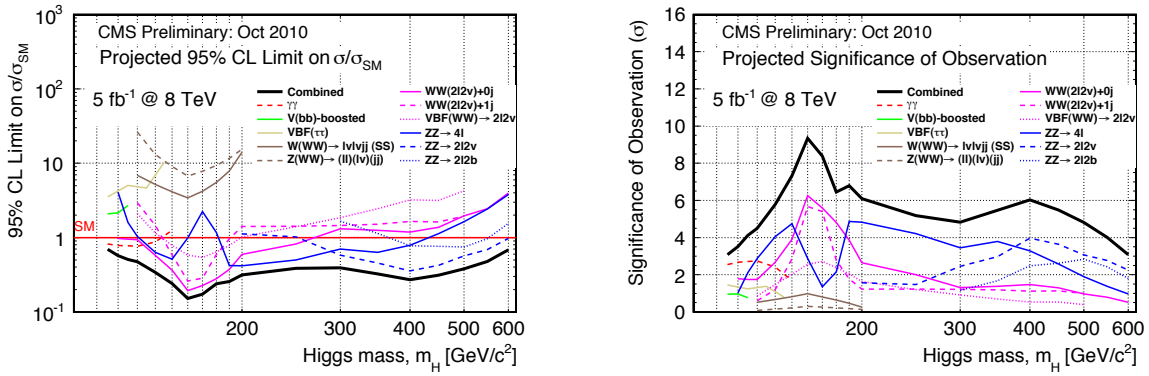
In the low-medium mass range, $H \to ZZ^{(*)} \to l^+l^-\,l^+l^-$ is considered the golden-channel for the Higgs search, due to its very clean signature with four isolated[5] leptons in the final state.

[5]The concept of lepton isolation, which is crucial for experimental searches, is extensively reviewed in Section 2.3.4.

The backgrounds to this channel are $ZZ^{(*)}$, $t\bar{t}$ and $Zb\bar{b}$ productions, and they can be suppressed in an efficient way by some requirements on the lepton isolation, transverse momentum and invariant mass, plus requirements on the event vertex. The main feature of this channel is that the Higgs mass peak can be fully reconstructed, hence giving access to the m_H parameter.

In the medium-high mass range, i.e. for $m_H \geq m_{WW}$, the $H \to WW^{(*)} \to l\nu_l l\nu_l$ is the other relevant channel. Again, the signature is very clean since two charged leptons and missing transverse energy leave as backgrounds only WW, $t\bar{t}$ and W+jets productions. These can be reduced by requirements on the lepton momentum and isolation, jet veto and by exploiting the small opening angle between the two leptons which is due to spin correlations. The channel is characterized by an extremely high signal event yield, which makes it the main tool for the Higgs production cross-section measurement.

Figure 1.9 summarizes quantitatively the considerations on the various channel discovery sensitivity, using a projection based on CMS Monte Carlo (MC) studies conducted in 2010. The general indication is that, in case a SM Higgs boson exists, an integrated luminosity[6] of $5\,\text{fb}^{-1}$, collected at $\sqrt{s} = 8$ TeV, would grant a discovery significance of at least 3σ (5σ in a wide part of the mass range) for all Higgs mass hypotheses between 114 and 600 GeV/c^2. Otherwise, the same amount of data would be enough to exclude the SM Higgs existence with a confidence level of at least 95% everywhere in the above mass range.



(a) (b)

Figure 1.9: The plots show (a) the projected exclusion limit and (b) the expected observation significance for a SM Higgs search at $\sqrt{s} = 8$ TeV and $\mathcal{L} = 5$ fb^{-1}. Contributions of individual channels used in the overall combination are also shown.

This picture is fundamental also to understand how this thesis project was designed. The available information in 2010, when I started to work at the thesis, enclosed two main considerations:

1. The low mass Higgs discovery potential of the LHC experiments is driven by a channel where the detector performance, in particular the photon energy resolution, is dramati-

[6] In accelerator physics, the luminosity $\mathcal{L}$ is an observable that represents the number of particles per unit area per unit time times the opacity of the target. It is a measurement of how well a collider performs, in terms of provided collisions per unit time. Using this variable, it is common to express the number of collisions delivered by some collider over a given amount of time using the relation $L = \int \mathcal{L}\, dt$, where L is the integrated luminosity.

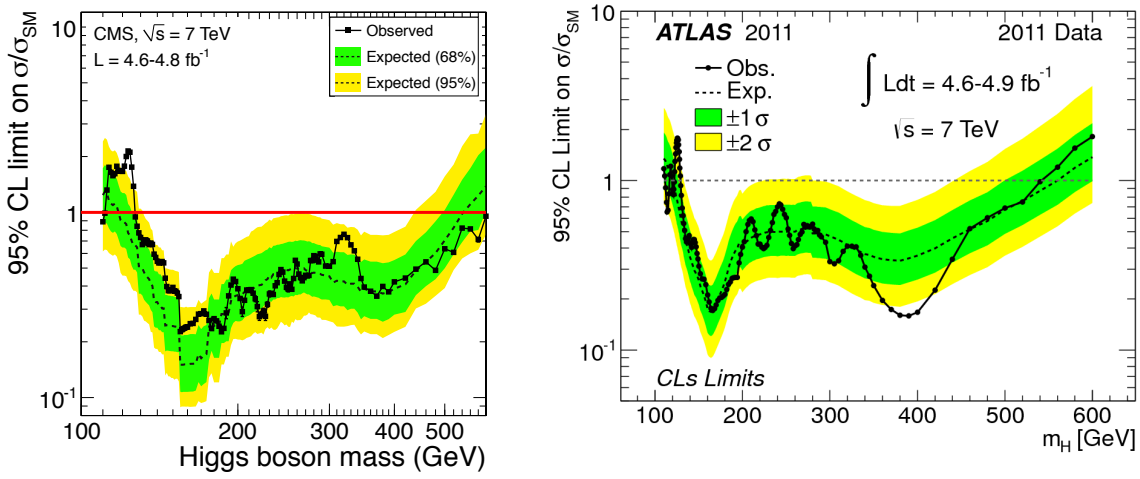
cally affecting the signal over background ratio. It is then imperative to guarantee ideal calibration conditions for the electromagnetic calorimeters of the experiments. This constitutes the first part of my doctorate thesis work and it is presented in Chapter 3.

2. The medium-high mass range sensitivity is driven by a channel, $H \to WW^{(*)} \to l\nu_l l\nu_l$, which lacks of the possibility to reconstruct the Higgs invariant mass. By letting one W decay hadronically, the signal yield can even be improved and, with a kinematic constraint, the reconstructed Higgs invariant mass can be accessed. The main challenge of the $H \to WW \to l\nu_l q\bar{q}$ search is that the background becomes an order of magnitude greater with respect to $H \to WW^{(*)} \to l\nu_l l\nu_l$, so that its mastering becomes critical to retain a good signal over background ratio and to keep the systematic uncertainties low. All these points are addressed in Chapter 4, which contains the second part of the doctorate studies.

1.4 The Standard Model today

The LHC experiments have been now recording and analyzing collisions since 2010. The first wave of striking results were published during early 2012, when both the ATLAS and CMS collaborations released the Higgs exclusion limits obtained with the dataset corresponding to LHC 2010-2011 proton-proton collisions at 7 TeV center of mass energy [23, 24].

With approximately $5\,fb^{-1}$ of integrated luminosity for each experiments, the results showed that the Standard Model Higgs boson is excluded, at 95% confidence level, in the [127,600] GeV/c^2 and [111.4,116.6]+[119.4,122.1]+[129.2,541] GeV/c^2 mass ranges for CMS and ATLAS respectively. Figure 1.10 graphically summarizes the limits for both the collaborations.



(a) (b)

Figure 1.10: The observed and expected 95% C.L. upper limits on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for the SM Higgs boson hypothesis as a function of the Higgs boson mass hypothesis for (a) the CMS experiment [24] and (b) the ATLAS experiment [23] with $\mathcal{L} \simeq 5\,fb^{-1}$.

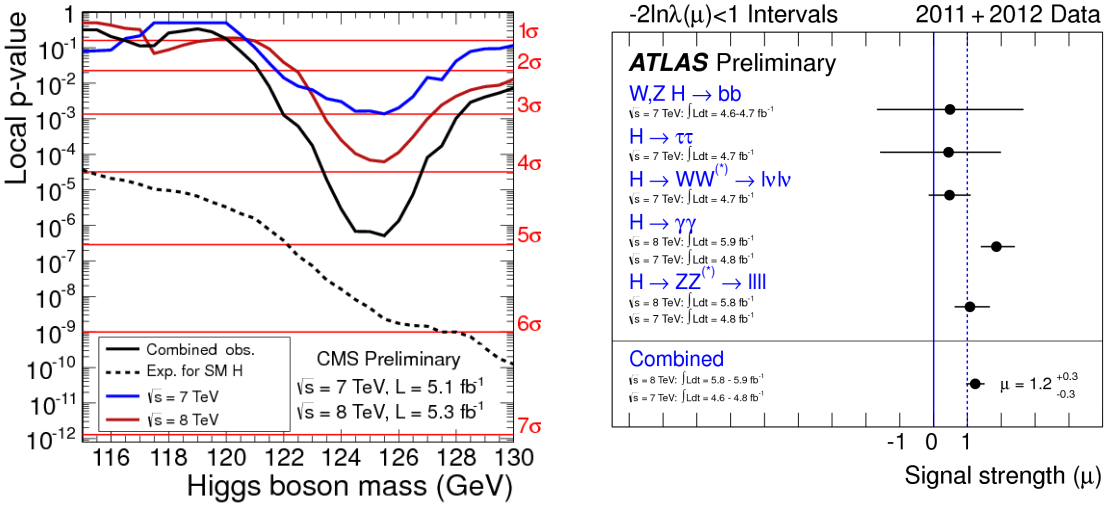
The results showed that the Higgs boson is not likely living in the medium-high mass regime. The search then was carried on with new data with particular care for the yet not excluded area,

around 125 GeV. The extreme high mass regime was also still opened for the boson existence, but here the electroweak fit is playing an important role since the Standard Model likelihood is compromised with an Higgs at very high mass.

A second, and more important, series of results were released during last summer. During the time-gap, the available statistics of collision data doubled, and, moreover, the effective statistical power of the dataset was more than doubled, since the switch of the LHC beam energy, from 3.5 to 4 TeV, increased the cross-section of the heavy products, such as the Higgs boson. Two important papers, by the CMS [25] and ATLAS [26] collaborations, enclosed the first statistically significant excess compatible with a SM Higgs Boson like particle.

The results can be summarized by two main facts:

- The two experiments detected a 5-sigma deviation from the background only hypothesis for a mass around 125 GeV (Figure 1.11a);
- The excess is compatible with a SM Higgs boson (Figure 1.11b).



(a) (b)

Figure 1.11: The observed local p-value p_0 for 7 TeV and 8 TeV CMS data [25], and their combination as a function of the SM Higgs boson mass (a). The dashed line shows the expected local p-value $p_0(m_H)$, should a Higgs boson with a mass m_H exist. In (b), values of $\hat{\mu} = \sigma/\sigma_{SM}$ for the combination (dashed vertical line) and for contributing channels (points), obtained with ATLAS 7 TeV plus 8 TeV data [26]. The horizontal line shows the overall $\hat{\mu}$ value of 1.2 ± 0.3. The horizontal bars indicate the $\pm 1\sigma$ uncertainties on the $\hat{\mu}$ values for individual channels; they include both statistical and systematic uncertainties.

The experiments are now refining the previous analyses and exploiting new data to measure various proprieties of the excess, a side from the cross-section, such as its spin, mass and branching ratios. The main goal is now to characterize the excess and see if it is fully compatible with the SM prediction. In any case, even if the recently found boson is behaving as expected from SM expectations, the search for new particle states is still mandatory to probe the completeness of the theory, since some of the most studied extensions of the Standard Model predict the appearance of several Higgs bosons [27, 28], one of which behaves exactly as predicted by the sole SM.

Finally, the CMS H $\rightarrow \gamma\gamma$ peak, in Figure 1.12, shows one of the most direct manifestations on data of the Higgs boson. It is striking to note that so many small efforts in different areas, from the detector design and construction, to the data acquisition and trigger, the detector alignment and calibration, the analysis design and validation, the Monte Carlo studies, are all converging to the very same objective, so simply and beautifully represented by the di-photon mass peak. This thesis essentially presents one of this little, but important, contributions in the quest for the Higgs boson.

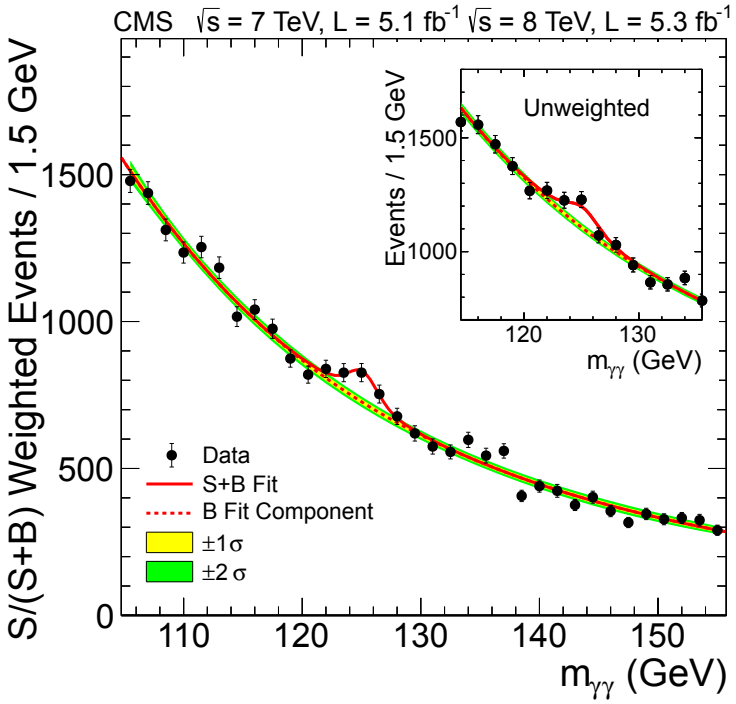


Figure 1.12: The diphoton invariant mass distribution with each event weighted by the S/(S+B) value of its category. The lines represent the fitted background and signal, and the coloured bands represent the ± 1 and ± 2 standard deviation uncertainties in the background estimate. The inset shows the central part of the unweighted invariant mass distribution.

Chapter 2

The LHC and the CMS experiment

In Chapter 1 the Standard Model was presented as the most accurate and predictive theory of elementary particles and their interactions. The theory evolved constantly as experiments provided new inputs, provided by complex machines that recorded particles collisions. The Large Hadron Collider represents the state of art of the collider technology and it is briefly reviewed in Section 2.1.

Section 2.2 contains a description of the Compact Muon Solenoid, which is one of the complex detectors sitting on the LHC ring and constitutes the experimental tool used in this thesis work.

Finally, Section 2.3 summarizes the CMS object reconstruction.

2.1 The Large Hadron Collider

The LHC [29] is a proton-proton circular collider with a length of 26.7 km located at CERN, near Geneva, in the already existing LEP tunnel. Four detectors are installed in the experimental caverns around the collision points: two general purpose experiments, ATLAS [30] and CMS [31, 32], the LHCb [33] experiment dedicated to B Physics and the ALICE [34] experiment where the physics of heavy ion collisions is investigated.

To supply the LHC with pre-accelerated protons, the pre-existing CERN facilities are employed, as it is shown in Figure 2.1, which contains a schematic view of the LHC accelerator and its injection chain. The protons are condensed into bunches, with $\approx 1.5 \cdot 10^{11}$ particles per bunch, in the Proton Synchrotron (PS), then accelerated by the Super Proton Synchrotron (SPS) up to an energy of 450 GeV, and finally injected into the LHC. Here, they are accelerated to an energy of 4 TeV in ≈ 1400 bunches[1]. Superconducting dipole magnets provide a magnetic field of 8.3 T to keep the protons on the orbit during the acceleration. The superconducting magnets are cooled using liquid helium at a temperature of 1.9 K.

At the interaction point, where the CMS experiment is located, collisions happen every 50 ns, corresponding to a bunch crossing frequency of 20 MHz. The total proton-proton cross-section at $\sqrt{s} = 8\,\text{TeV}$ is around 70 mb. Therefore, approximately 30 inelastic events (pile-up) per

[1] These are the actual values for the 2012 run. During 2011 the beam energy was of 3.5 TeV, while the number of bunches and the number of particles per bunch was approximately equal to the 2012 values.

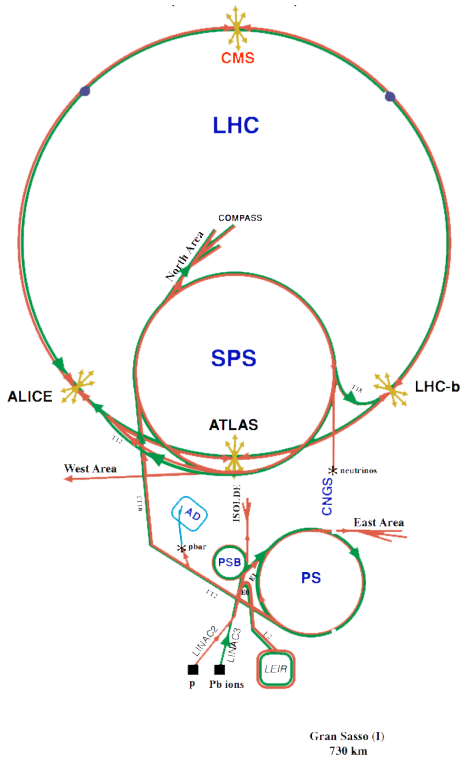


Figure 2.1: The LHC injection scheme. The proton bunches are injected into the Proton Synchrotron (PS), then accelerated by the Super Proton Synchrotron (SPS) up to an energy of 450 GeV, and finally redirected into the LHC, where they reach an energy of 4 TeV.

bunch-crossing are observed in the multi-purpose experiments at the current luminosity.

Historically, the first two years of LHC proton-proton collisions happened at a lower center-of-mass energy of $\sqrt{s} = 7$ TeV and at a significantly lower luminosity than what is foreseen in the original design. During the last year, the LHC team managed to bring up the energy to $\sqrt{s} = 8$ TeV and the instantaneous luminosity up to $\mathcal{L} = 7.54 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, close the the design value of $\mathcal{L} = 1 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The machine history is graphically summarized in Figure 2.2.

2.2 The CMS detector

The Compact Muon Solenoid, illustrated in Figure 2.3, is one of the two general-purpose detectors located at the LHC. The main features of the CMS detector are summarized by its name: it is very compact, so that the tracking and the calorimetry system are both inside the magnetic solenoid, it has an superb muon system assisted by an excellent central tracking detector, and the solenoidal magnet produces a magnetic field of 3.8 T, which allows an outstanding particle momentum measurement. Furthermore, the design of the CMS detector has focused on the hermeticity, crucial to be sensitive also to non-interacting high energetic neutrinos.

As most of the high energy detectors, it has a multi-layered cylindrical barrel structure enclosing the beam pipe, as it is shown in the transverse section schematic (Figure 2.4), and end-caps to complete the hermetic coverage. Starting at the beam interaction point, the constituent

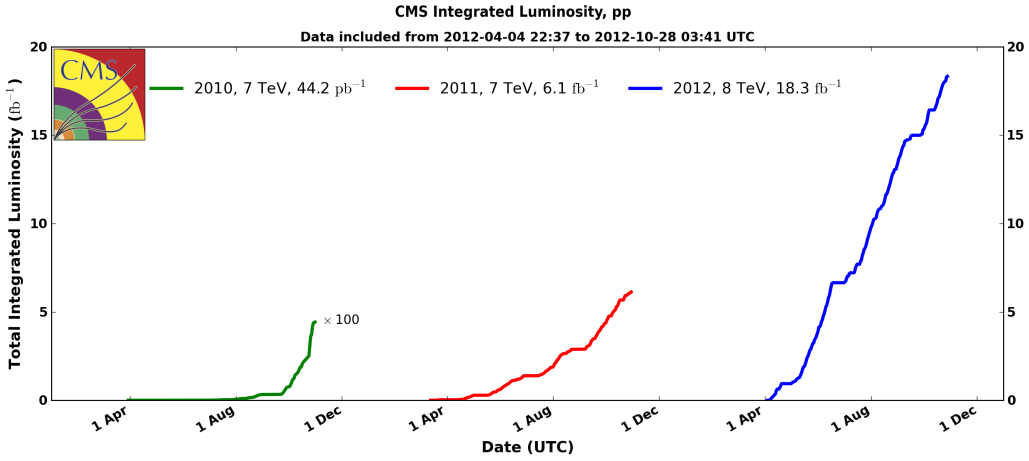


Figure 2.2: Total integrated luminosity vs. time since the startup of the LHC. Each color refers to a different year of the data taking. The slope of the curves gives the corresponding instantaneous luminosity $\mathcal{L}$, which dramatically increased up the current value of $7.54 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

compartments are the silicon tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the superconducting solenoidal magnet, followed by layers of muon chambers supported by the iron return yoke of the magnetic field. In the following paragraphs each of these sub-systems is introduced.

2.2.1 Coordinate conventions

In a proton-proton collision, the fractions x_a and x_b of the parent proton momentum carried by the interacting partons are in general different, and the rest frame of the hard collision is boosted along the beam line with respect to the laboratory frame. The reconstruction of the boost of the system requires the full reconstruction of the remnants of the colliding protons, which is in practice not feasible because of the presence of the beam-pipe and other passive instrumentation at small angles with respect to the beam-axis. The consequence is that proton collisions are usually studied in convenient coordinates, which retain some invariance under Lorentz boosts along the beam-line.

More specifically, the coordinate convention is such that the z-direction is parallel to the beam line, the y-direction is vertical, and the x-direction is horizontal and points to the center of the LHC ring. The center of the detector is taken as the origin of the coordinate system. Additionally, the azimuthal angle, ϕ, is measured around the beam axis with respect to the x-axis, and the polar angle, θ, is the angle of inclination from the beam axis. The polar angle is in fact expressed in terms of the pseudorapidity[2], η, defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \quad . \tag{2.1}$$

The advantage of using this coordinate comes from the Lorentz invariance of differences in η under Lorentz boosts along the beam-line. As a consequence, a solid angle in (η, ϕ) is also invariant under longitudinal boosts. For the same conveniency reasons, the momentum is

[2]The pseudorapidity η is in fact the high-energy limit of the rapidity y, defined as $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$.

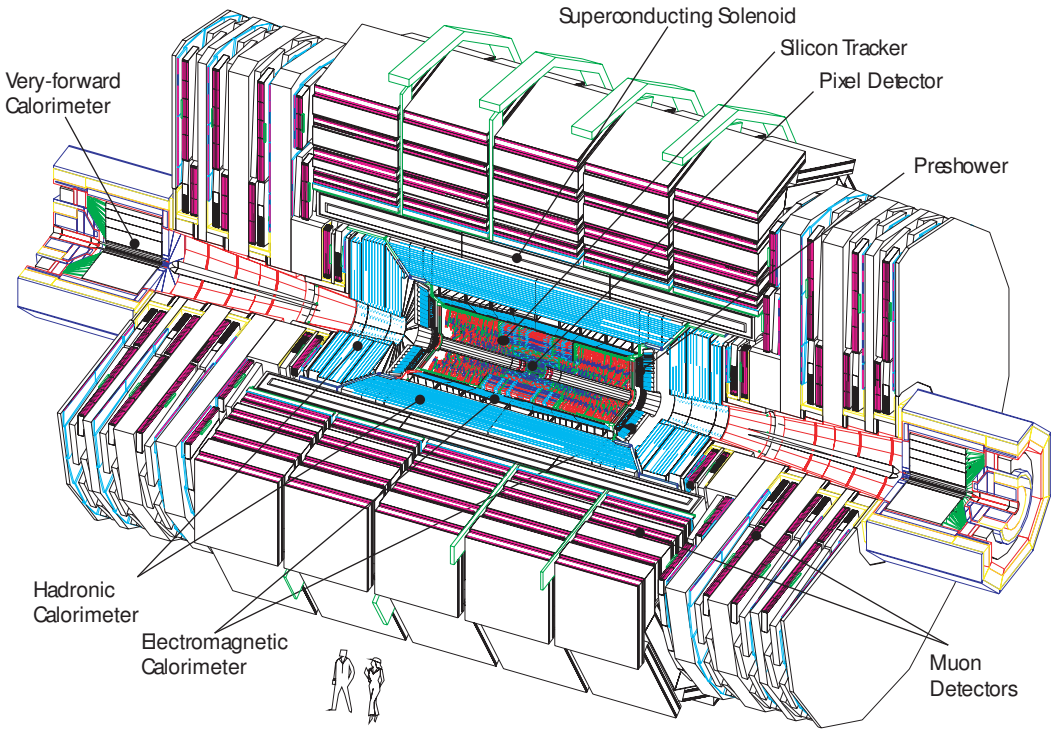


Figure 2.3: Pictorial view of the CMS detector.

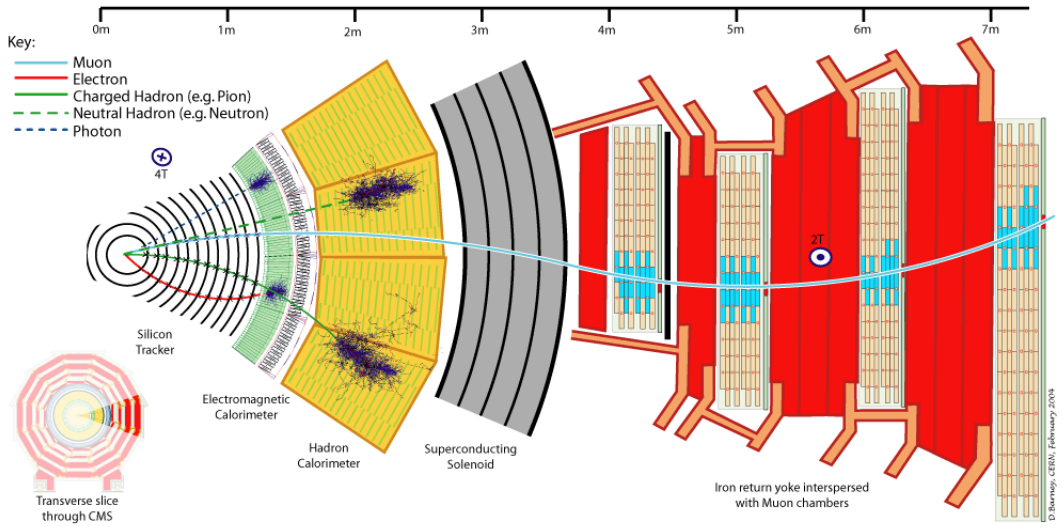


Figure 2.4: A slice of CMS: the picture shows the sub-detector sequence. Paths of different particles are superimposed.

particularly studied in the transverse plane.

2.2.2 The silicon tracker

The CMS silicon tracker [35] has been designed to be able to cope with the high particle flux present in the LHC proton-proton collisions. It is the largest silicon tracker ever built, having a diameter of 2.4 m and a length of 5.4 m. In order to ensure an effective pattern recognition, it has been designed such that the occupancy per event of the silicon sensor channels is small, ranging from 10^{-4} for pixel sensors, to the percent level for strip sensors. In order to achieve such a small hit occupancy, the entire system consists of a 66 M channel pixel detector and of 20000 silicon sensors which have altogether a surface of 210 m^2. The increasing size of the sensors is due to the energy flux, which decreases with the distance to the interaction point. The tracking system is subdivided into five main parts: the Tracker Outer Barrel (TOB), the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker End Caps (TEC) and the pixel detector, all illustrated in Figure 2.5. The barrel-endcap transition occurs at $0.9 < |\eta| < 1.4$, and the entire tracker provides rapidity coverage up to $|\eta| \simeq 2.5$.

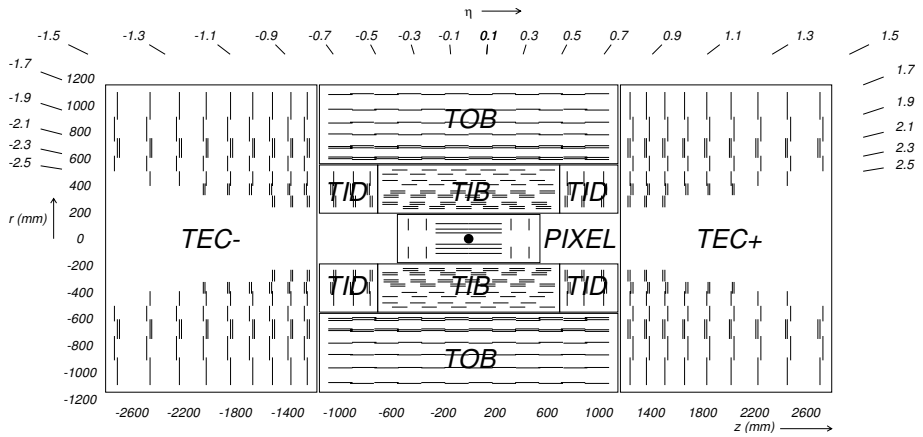


Figure 2.5: Schematic cross-section through the CMS tracker. Each line represents a detector module.

The pixel detector must be situated as close to the interaction region as possible in order to identify relatively long-lived particles such as bottom and charm hadrons and τs: the three pixel layers are positioned at a radial distance of 4.4, 7.3 and 10.2 cm from the beam axis. The layers are composed of ladders, which include 8 pixel sensors each. The pixel endcaps consist of two end disks, which extend from 6 to 15 cm in radius and are placed on each side at $|z| = 34.5$ cm and 46.5 cm. Each disc blade holds 24 pixel modules. In order to optimize the resolution of the pixel detector in both the $r\phi$ and z-directions, each module in the pixel end disks is rotated to take advantage of the electron Lorentz drift angle in the sensor layer due to the 3.8 T magnetic field. The hit resolution of the pixel detector is approximately 10 μm in $r\phi$ and 15-20 μm in z.

The barrel part of the strip detector consists of four layers of TIB modules and six layers of TOB modules. The modules are equipped with chips to read out energy deposition from the strips. Depending on the radii, the rectangular barrel modules have different sizes, and contain different numbers of strips, in order to keep the occupancy low, while sustaining the separation of strips to a $100 - 200$ μm level to ensure a good hit position resolution. The inner two layers of

TIB and TOB modules are made doublesided with two back-to-back sensors at a relative angle of 100 mrad, providing high precision two-dimensional hit detection.

Due to its multi-layered complex structure, the amount of material a particle traverses and the fraction of energy it may lose until it exits the tracker varies with position. One of the aims in the tracker design is to keep this so-called *material budget* minimal. In the case of the CMS silicon tracker, the radiation length of the material (X_0) crossed by a particle is strongly dependent of the particle direction. For the central region of the detector, the radiation length is about 0.4 X_0, but this number increases rapidly when moving to forward regions, as it can be seen in Figure 2.6. A maximum of 2 X_0 is found for the barrel-endcap transition region. This rather large amount of material in the tracker leads to significant energy loss, multiple scattering for electrons, and to photon conversions.

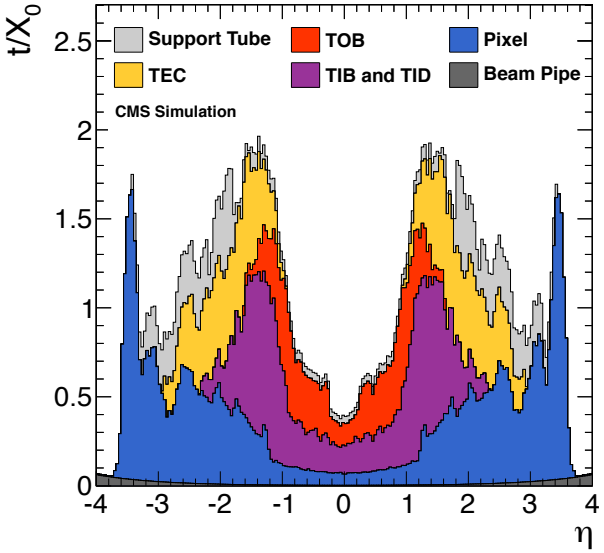


Figure 2.6: Material budget in units of radiation length as a function of pseudorapidity η for the different tracking sub-detectors.

2.2.3 The electromagnetic calorimeter

Since the electromagnetic calorimeter [36] plays an important role in reconstructing the di-photon decay of the Standard Model Higgs boson as well as in identifying electrons from vector boson and τ-lepton related channels, CMS has chosen for a very compact homogeneous scintillating crystal calorimeter designed for precision measurements of electron/photon energies and directions.

The ECAL barrel is positioned just outside the tracking system at a radius of only 120 cm from the interaction point. Consequently, to distinguish energy depositions from different sources and to achieve high spatial resolution, a material that can provide small lateral and longitudinal spread of the shower is needed. Due to its high density, small Molière radius, fast response[3] and radiation hardness, lead tungstate ($PbWO_4$) crystals were selected as the scintillation material to realize a very compact, high performance ECAL (energy resolution of 0.5% for a 50 GeV

[3] Around 80% of the light-yield is emitted within 25 ns from the particle-crystal interaction.

particle). The light produced in the crystals is gathered with silicon Avalanche Photo-Diodes (APDs) in the barrel and Vacuum PhotoTriodes (VPTs) in the endcaps.

Similar to the other sub-detectors, also the ECAL consists of a barrel and two endcap substructures (see figure 2.7). The ECAL barrel, which is symmetric around $\eta = 0$, contains a total of 61200 $PbWO_4$ crystals, each with a surface of 22 mm$\times$22 mm, where the edge length is similar to the Molière radius of the scintillation material. This surface translates in a 1 deg or 0.0174 rad coverage in η and ϕ, while its depth of 23 cm corresponds to a radiation length of 25.8 X_0. All the barrel crystals approximately point toward the interaction point, with an offset of 3 deg in order to minimize the detected energy loss if a particle transverses exactly between two crystals.

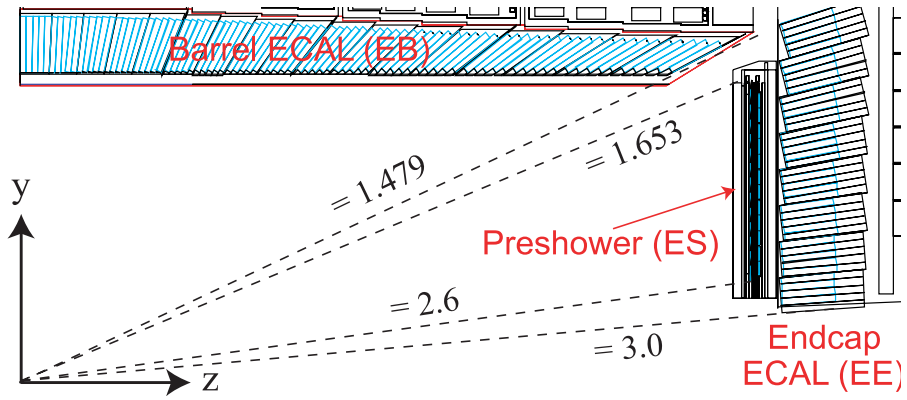


Figure 2.7: View in (r, z) of a quarter of the CMS electromagnetic calorimeter.

The 7324 $\times$ 2 endcap crystals have different dimensions from those used in the barrel, having a larger front face of size 29 mm$\times$29 mm and a shorter length of 22 cm, corresponding to $\sim$ 25 X_0. Additionally, pre-shower detectors are inserted between the tracker and the ECAL endcaps. The role of the pre-shower is to provide position measurement of the electromagnetic shower to high accuracy and, for instance, it is used to discriminate photons produced in Higgs boson decays from photons produced by $\pi^0 \to \gamma\gamma$. The pre-shower is located just in front of the ECAL endcap, where thin lead radiators are used to initiate the shower (providing 3 X_0 in total). After each radiator layer, silicon strip sensors are placed to measure the hit position of the shower.

The main drawback of $PbWO_4$ is due to irradiation: while the scintillation mechanism stays unaffected, crystal transparency is modified via creation of colour centers that absorb and scatter the light [36, 37]. To account for these changes, crystal transparency is measured using laser pulses injected into the crystals via optical fibers during LHC beam gaps. Two laser wavelengths are used to perform the monitoring. One, blue, at $\lambda = 440$ nm, is very close to the scintillation emission peak and it is used to follow the changes in transparency due to radiation; the other, near infra-red, at $\lambda = 796$ nm, is far from the emission peak, and very little affected by changes in transparency, so it can be used to verify the stability of other elements in the system.

2.2.4 The hadronic calorimeter

In the surroundings of the ECAL there is the Hadronic CALorimeter (HCAL) [38], which is responsible for energy measurements of hadrons and their products.

Hadronic calorimeters rely on nuclear interactions which result in both hadronic and electromagnetic showers. As the probability of a nuclear interaction is small but the energy deposited in the calorimeter is large, there are significant fluctuations in the measured energy in hadronic calorimeter showers, decreasing the overall precision of the detector. The CMS hadronic calorimeter design was strongly influenced by the decision to place the calorimeter inside the solenoid, leading to little space for the detector. The HCAL consists of three main parts: the Hadron Barrel (HB) and the Hadron Endcap (HE), which extend up to $|\eta| = 3$, and the Hadron Very Forward (HVF), located around the beam pipe outside the muon system at $|z| = 10.9$ m, that completes the coverage up to $|\eta| = 5.3$. The large pseudorapidity coverage ensures a very high hermeticity leading to a relative high precision to the reconstructed missing transverse energy. A schematic view of the calorimeter location in CMS is given in Figure 2.8.

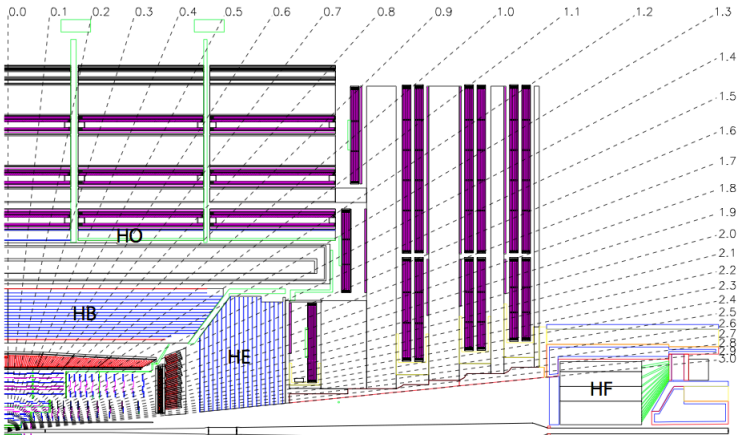


Figure 2.8: Longitudinal view in (r, z) of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HVF) hadronic calorimeters.

Both the HB and HE are built in a classical sandwich-like sampling calorimeter structure, with a brass absorber coupled to a plastic scintillator with wavelength-shifting fibers. Light collection is achieved through hybrid photodiodes. Scintillators behind the coil of the magnet constitutes the Hadron Outer (HO) detectors and effectively increase the calorimeter thickness to 10 interaction lengths. The scintillators are sub-divided into tiles of size $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 2$ and larger divisions at high pseudorapidity. These are to be used as the basic building blocks in the reconstruction of the hadronic showers.

For the construction of the HVF, steel was selected as absorber. The collection of the energy is accomplished using quartz fiber emitting Cerenkov light, a technology which is preferred to plastic because of its greater radiation tolerance. Additionally, relatively faster phototubes are selected in the HVF compartment instead of hybrid photodiodes.

According to the test-beam results, the expected energy resolution for single pions interacting in the central part of the calorimeter is

$$\frac{\sigma_E}{E} = \frac{94\,\%}{\sqrt{E}} \oplus 4.5\,\% \quad ,$$

where the energy is measured in GeV. An important degradation of the resolution is expected at $|\eta| = 1.4$, due to the presence of services and cables. The performance of the very forward calorimeter is expected to be

$$\frac{\sigma_E}{E_{\text{had}}} = \frac{172\,\%}{\sqrt{E_{\text{had}}}} \oplus 9\,\% \qquad \frac{\sigma_E}{E_{\text{em}}} = \frac{100\,\%}{\sqrt{E_{\text{em}}}} \oplus 5\,\% \quad .$$

2.2.5 The muon system

The muon system [39] is a muon-tracking device in the outermost region of the CMS detector; at this layer, only these kind of leptons, plus other non-interacting particles, manage to pass through the calorimeters without depositing a large fraction of their energy.

There are four layers of muon stations in the barrel and the endcaps, interleaved with the iron return yokes, as it is illustrated in Figure 2.9. Each of the stations is meant to provide track segments reconstructed from a few distributed hits. These will later be combined with information from the inner silicon tracking system to form the complete muon tracks.

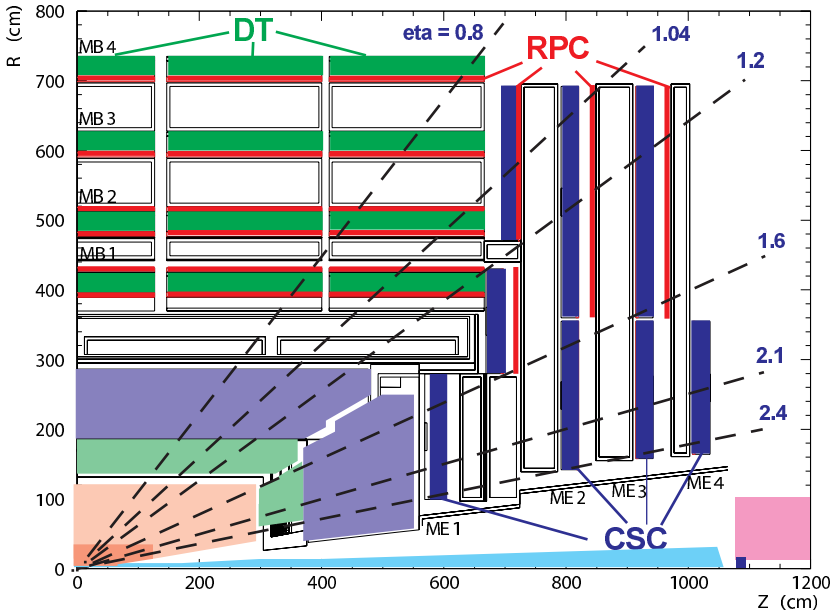


Figure 2.9: View in (r, z) of a quarter of the CMS muon system layout.

An important challenge of the muon system is to provide fast recognition and efficient reconstruction, which are required for triggering purposes. These goals have to be achieved with varying muon and neutron radiation flux, and an inhomogeneous magnetic field in forward detector regions, so the decision was to design a system exploiting different technologies for the barrel and endcap regions. Drift Tubes (DT) are employed in the barrel stations and Cathode Strip Chambers (CSC) in the endcap stations. Both types of detectors are assisted by Resistive Plate Chambers (RPC). Each of the stations contains respectively 12 DT layers for the barrel and 6 CSC layers in the endcap. Within a single station, the DT layers provide a final spacial resolution of 100 μm and the CSCs provide $80 - 450$ μm, depending on the location of the station. The

DT and CSC detectors are used to obtain a precise position measurement, while the RPCs are, due to their very fast response and time resolution of the order of 1 ns, dedicated to triggering purposes. The RPC plates in the endcaps extend up to $|\eta| < 1.6$ and the reconstruction of muon tracks in the region outside is carried out using the CSCs alone.

2.2.6 The trigger system

The LHC design bunch crossing frequency for proton-proton interactions is 40 MHz, which results in a total event rate of the order of 10^9 Hz. Since the raw event size is of the order of 1 MB, and storing and processing the resulting amount of data exceeds the current data-storing and process-time limits, a large fraction of the total event rate needs to be discarded by means of an online event selection system. The goal of the CMS trigger system is to reduce the event rate to the order of 100 Hz, the maximum rate that can be archived by the online computer farm. Hence, typically only one out of 10^7 events can be stored, which fortunately leaves enough freedom to accommodate all signal channels of interest at the LHC.

Another aspect of the online selection is its speed to process the events. Since the input rate is 40 MHz, a decision must be taken every 25 ns. However, this time is too small even to read out all raw data from the detector. The accept/reject decision is therefore taken in several steps (levels) of increasing refinement, where each level takes a decision using only part of the available data. In the case of CMS, the trigger consists of two physical levels [40].

The Level-1 (L1) trigger is implemented on dedicated hardware, and has only access to data from the calorimeters and the muon detectors with coarse granularity. Hence, information of the central tracking system is not yet used in the L1 trigger decision. L1, on the basis of this limited information, has to reduce the input rate up to a level acceptable for the Data Acquisition system (DAQ): at LHC, the DAQ system is able to handle an event rate of up to 100 kHz. Once accepted, the L1 triggered objects are passed to the High Level Trigger (HLT), which is implemented in software running on a single farm of commercial processors.

The HLT event selection at the LHC is generally based on the fact that most proton-proton collisions produce soft hadrons with transverse momenta (p_{T}) of few GeV/c, or so-called "minimum bias" events, while signal events (e.g. decays of heavy objects like the Higgs boson and the top quark) typically contain high-p_{T} leptons, jets or missing transverse energy. Hence, the trigger decisions are made based upon the identification of the following physics objects:

- electrons/photons
- jets
- muons
- missing transverse energy

where each object is reconstructed in a specific detector sub-system or with combined information, as it is detailed in Section 2.3.

2.3 Object reconstruction at CMS

As was presented in Section 2.2, the CMS comprehends many subdetectors, each dedicated to a particular function. These subdetectors constitute the building blocks of the CMS high-level Physics objects reconstruction. This chapter contains an overview of the CMS reconstruction chain, with particular emphasis on the objects used in the ECAL calibration study (Chapter 3) and in the $H \to WW \to l\nu_l q\bar{q}$ search (Chapter 4).

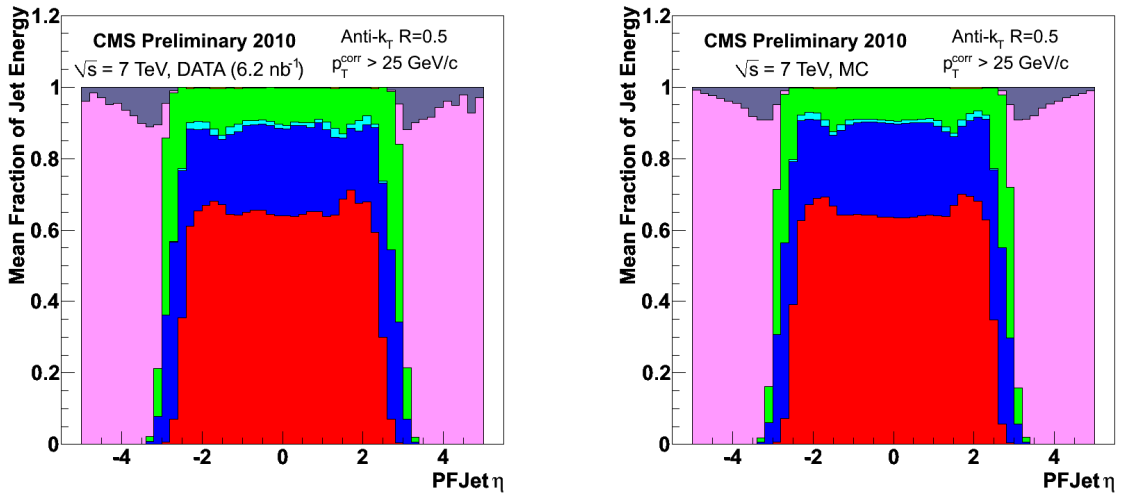
2.3.1 Particle-flow reconstruction

With its large silicon tracker immersed in a uniform axial magnetic field of 3.8 T provided by a superconducting solenoidal coil, CMS is the ideal detector to reconstruct charged-particle tracks with large efficiency and adequately small fake rate down to a transverse momentum (p_T) of 150 MeV/c, for pseudorapidities as large as ± 2.6. Thus, each charged particle history can be reconstructed starting from the very interaction point, and each of them can be identified through the combination of all sub-detectors inputs, so that an optimal determination of their direction, energy and type can be performed. In a similar way, exploiting calorimeter deposits, also neutral particles can be studied and classified, hence a full event reconstruction aimed at characterizing and identifying all stable particles in the event, can be implemented. This procedure is denoted as particle-flow (PF) event reconstruction and extensively described in here [41, 42]. As an additional feature, the particle-flow reconstruction builds up physics objects by picking, exclusively, from the whole poll of particle candidates, so that none single particle can live inside two distinct objects. The important consequence is that particle double counting and object overlapping are completely avoided.

The first physics objects reconstructed by the particle-flow sequence are muons, from a combination of the tracker and muon chamber information, as it is more detailed in Section 2.3.2.
Electrons are reconstructed by a combination of a track and of several energy deposits in the ECAL, from the electron itself and from possible Bremsstrahlung photons radiated by the electron in the tracker material on its way to the ECAL. More details are reported in Section 2.3.3.
Photons (e.g. coming from π^0 decays or from electron bremsstrahlung) are identified as ECAL energy clusters not linked to the extrapolation of any charged particle trajectory to the ECAL.
Charged hadrons are identified as charged particle tracks neither identified as electrons, nor as muons.
Neutral hadrons are identified as HCAL energy clusters not linked to any charged hadron trajectory, or as ECAL and HCAL energy excesses with respect to the expected charged hadron energy deposit.
Finally, the PF approach to the event reconstruction also allows for a natural definition of jet objects. Once final state, well isolated leptons and photons are excluded from the particle list, all that remains can be clustered into jets, as further described in Section 2.3.5.

Figure 2.10 shows the composition of a typical minimum-bias event in terms of different particle types. In the central part of the detector, where the tracker allows for charge measurements, the largest fraction of an event energy is carried by charged hadrons ($\sim 65\%$). Only about 2% is carried by electrons, with neutral hadrons and photons almost equally sharing the remaining part. Outside the tracker acceptance, instead, no distinction can be made between charged and neutral particles. Here, the vast majority of the event energy is carried by hadronic candidates,

with purely electromagnetic objects contributing at 10% level or less.



(a) (b)

Figure 2.10: Reconstructed jet energy fractions as a function of pseudorapidity in data (a) and in Monte Carlo (b). From bottom to top in the central region: charged hadrons, photons, electrons, and neutral hadrons. In the forward regions: hadronic deposits and electromagnetic deposits.

In terms of physics performance, while no substantial changes are expected for the reconstruction of high-energy electrons and muons (e.g. from W boson fragmentation), the particle-flow allows to significantly improve the resolution of jets and missing transverse energy, with respect to a standard, pure calorimetric jet reconstruction. Since only about the 15% of a jet energy is carried by neutral, long-lived hadrons (neutrons, Λ baryons, etc.), for the remaining 85% carried by charged particles, the coarse HCAL information can be combined with the more precise tracker momentum measurements, thus allowing for a largely better jet reconstruction.

2.3.2 Muon reconstruction

The standard muon reconstruction sequence is performed in three stages: local reconstruction, standalone reconstruction and global reconstruction [43].

The local/Level-1 muon reconstruction is performed in the entire muon system and uses patterns of segments reconstructed in the CSC and/or DT chambers. Starting from initial seeds, the compatible chambers are identified, and those aligned are used as building blocks for segment candidates.

The standalone/Level-2 muon reconstruction uses data from DTs, CSCs and RPCs. The latters, despite having a coarser spatial resolution, are used since they complement the tracking chambers, especially where the geometrical coverage is problematic (mostly in the barrel-endcap overlap region). The reconstruction starts with the track segments from the muon chambers obtained by the local reconstruction. The state vectors (i.e. track position, momentum, and direction) associated with the segments found in the innermost chambers are used to seed the muon trajectories, working from inside out, using the Kalman-filter technique. A suitable cut is applied in order to reject bad hits, mostly due to showering, delta rays and pair production. In

case no matching hits (or segments) are found, e.g. due to detector inefficiencies, geometrical cracks, or hard showering, the search is continued in the next station. The state is propagated from one station to the next by taking into account the muon energy loss in the material, the effect of multiple scattering and the varying magnetic field in the muon system. The track parameters and the corresponding uncertainties are updated at each step. The procedure is iterated until the outermost measurement surface of the muon system is reached. A backward Kalman-filter is then applied, working from outside in, and the track parameters are defined at the innermost muon station. Finally, the track is extrapolated to the nominal interaction point (defined by the beam-spot size: $\sigma_{xy} = 150$ μm and $\sigma_z = 5.3$ cm) and a vertex-constrained fit to the track parameters is performed.

The global/Level-3 muon reconstruction consists in extending the muon trajectories to include hits in the silicon tracker. A set of silicon layers compatible with the muon trajectory is exploited to perform regional track reconstruction. Inside the region of interest, initial candidates for the muon trajectory (regional seeds) are built from pairs of reconstructed hits. In addition, a relaxed beam-spot constraint is applied to track candidates, which must fulfill a minimum transverse momentum requirement, to obtain initial trajectory parameters. Starting from the regional seeds, a track reconstruction algorithm, based on the Kalman-filter technique, is used to reconstruct tracks inside the selected region of interest. The track-reconstruction algorithm consists of the following steps: trajectory building (seeded pattern recognition), trajectory cleaning (resolution of ambiguities) and trajectory smoothing (final fit). In addition, the trajectories are refitted using, at the same time, silicon tracker hits and hits in the innermost muon station (excluding RPC hits). The χ^2 probability of such fit is compared with the corresponding of the tracker-only trajectory, in order to detect muon bremsstrahlung or any kind of significant energy loss of the muon before the first muon station. This procedure gains more importance for the accurate momentum reconstruction of very high-p_{T} ($\mathcal{O}(1\,\mathrm{TeV})$) muons.

Figure 2.11 shows the final resolution achieved by the CMS muon reconstruction system, as well as the contributions of the single sub-detectors, silicon tracker and muon system.

2.3.3 Electron reconstruction

Electron and positron[4] reconstruction in CMS uses two complementary algorithms, different at the track seeding stage: *tracker driven* seeding, more suitable for low p_{T} electrons as well as performing better for electrons inside jets, and *ECAL driven* seeding. The ECAL driven algorithm was created and optimized for isolated electrons in the p_{T} range relevant for Z or W decays, which are essential for this thesis work, hence will be briefly detailed in this section.

The clustering of the energy deposits in the electromagnetic calorimeter is the first step of the electron reconstruction procedure. A single electron hitting ECAL generates a shower which develops in more than one crystal, and the bremsstrahlung emission is responsible for a further spread of the energy in the azimuthal direction. Therefore, to obtain an accurate measurement of the electron energy in correspondence of the primary vertex and minimize the cluster containment variations, it is essential to collect bremsstrahlung photons.

As a starting point, crystals with energy above a certain threshold, called *seeds*, are searched

[4] In the thesis, the word "electron" will be conventionally used to refer both the the e^- and e^+ particles.

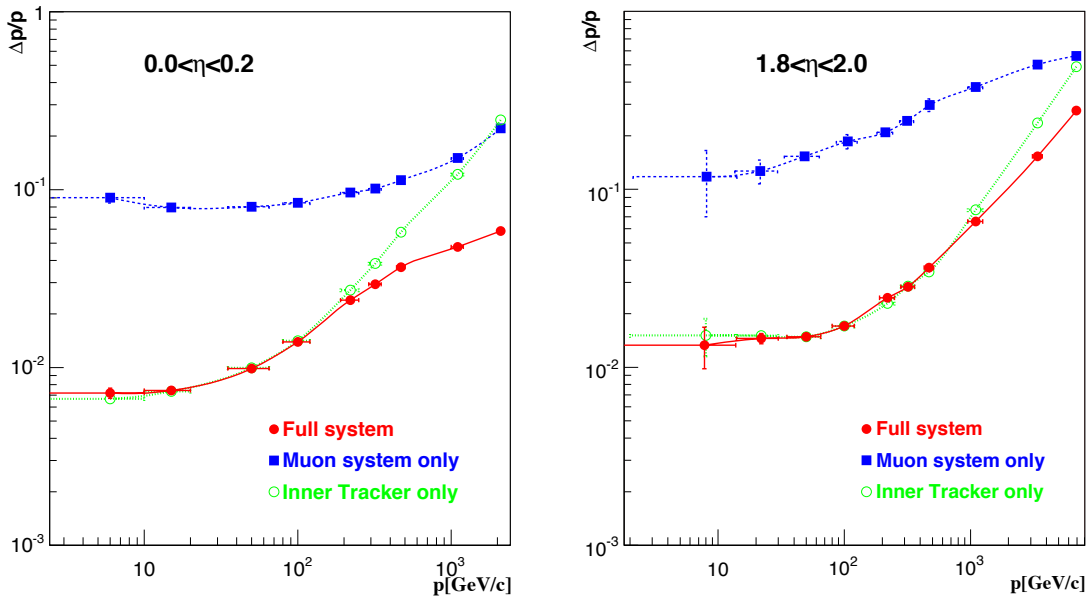


Figure 2.11: The momentum resolution for central (a) and forward (b) reconstructed muons. The final resolution is overlapped to the single system contributions, i.e. the silicon tracker and the muon system.

for. Starting from each seed, the energy deposits are then collected using two different reconstruction algorithms: the Hybrid and the Island algorithms [44], used respectively in the barrel and in the endcaps.
The Hybrid algorithm looks for 1×3 or 1×5 dominoes of crystals in the $\eta - \phi$ plane, each with a total energy of at least 100 MeV. Different dominoes, aligned along η, are clustered together in ϕ, if a valley with less than 100 MeV of deposited energy separates different clusters, up to an extension of ± 17 crystals away from the seed one. The so-obtained cluster of clusters goes under the name of super-cluster.
The Island algorithm moves in both directions in ϕ, starting from the seed position, and collects all the crystals until it finds an energy rise or a hole. Then, it moves one step in the η direction and makes another azimuthal scan. The stop condition for the η search coincides with the azimuthal one. Finally, the algorithm comes back to the seed and moves in the opposite η direction. When the procedure is done, a new crystal cluster is created. To recover bremsstrahlung photons, nearby clusters are merged into super-clusters as it is done for the previous algorithm.

After the clustering, the electron energy is computed as the sum of the deposits in the crystals belonging to a given super-cluster. A correction factor, depending on the number of crystals in the super-cluster, is applied to minimize the residual dependence of the energy scale both on the energy and on the pseudorapidity.

Following the clustering, the reconstruction proceeds with the track-building stage. Under both $+1$ and -1 charge hypotheses, the super-cluster energy and position are back-propagated in the tracker to the nominal vertex, to look for compatible hits in the pixel detector. Once a pair of compatible hits is found, an electron pre-track seed is built. Starting from seeds, compatible hits

are searched for on the next available silicon layers. In this pattern-recognition problem, the probability of major energy losses due to bremsstrahlung emission has to be taken into account. Therefore, a dedicated algorithm has been developed, where the electron energy loss pdf, well described by the Bethe-Heitler model [45], is approximated with a sum of Gaussian functions, in which different components model different degrees of hardness of the bremsstrahlung in the layer under consideration. This procedure, known as Gaussian Sum Filter (GSF) [46], is iterated until the last tracker layer, unless no hit is found in two subsequent layers. A minimum of five hits is required to create a track.

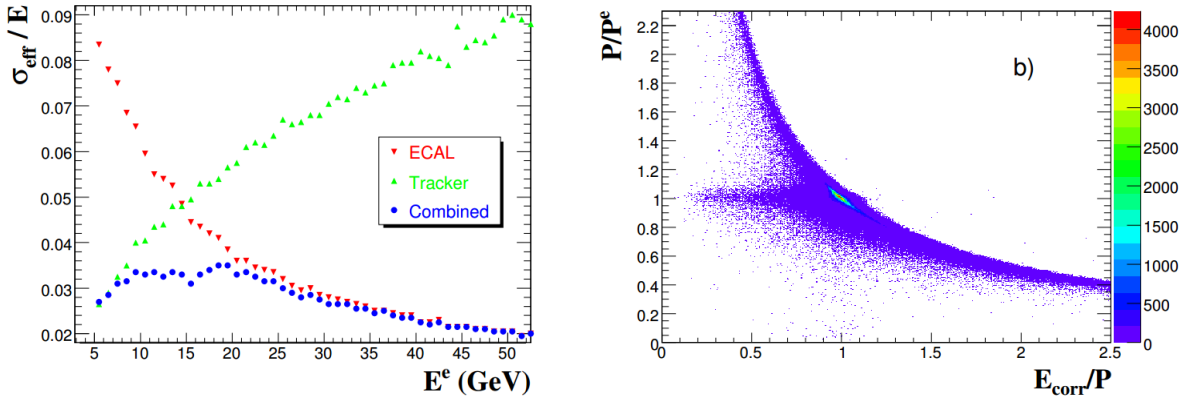


Figure 2.12: In the left plot the fractional resolution (effective RMS) is plotted as a function of generated energy E^e as measured with the ECAL supercluster (downward arrows), the electron track (upward arrows) and the combined track-supercluster (circles). In the right plot correlations between ECAL energy and tracker momentum measurements in the η range of the barrel are shown [43].

In the final stage, the super-cluster and track information are merged. The energy measurement E_{sc} provided by the electromagnetic calorimeter can be combined with the tracker momentum measurement p_{tk} to improve the estimate of the electron momentum at the interaction vertex. The improvement is expected to come both from the opposite behaviour of E and p with respect to the respective calorimetric and tracking resolutions, and from the fact that p_{tk} and E_{sc} are differently affected by the bremsstrahlung radiation (see Figure 2.12).

2.3.4 Lepton isolation

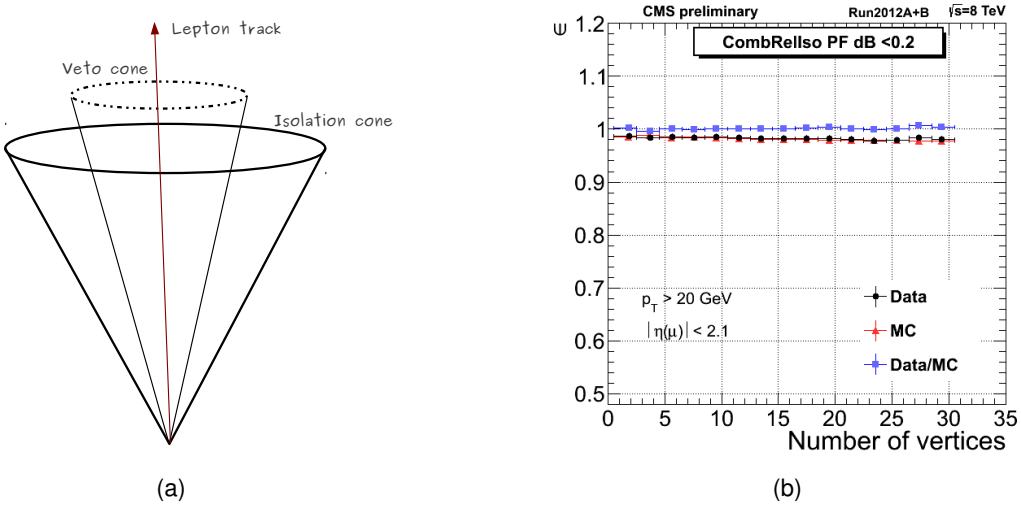
Although it lives at a more sophisticated level with respect to the bare object reconstruction, the *isolation* variable can be described generally for a typical CMS lepton as the measure of hadronic activity around a given lepton. This measurement proves as a powerful tool for physics searches, since leptons characterized by a neighboring high hadronic activity are likely to come from heavy hadrons decays or to be fake-leptons living inside hadronic jets, and not products of W or Z bosons decays. As a simple example, one can consider a lepton coming from a B-meson semileptonic decay, which will likely have spatially neighboring particles from the B-meson remnants. Those additional particles can be reconstructed, their transverse momentum can be summed up, and, if the sum is greater than a certain threshold, the lepton would then be considered not isolated.

Particle-flow based isolation

Particle-flow provides as an excellent tool for a natural definition of the isolation variable, which is defined as

$$I_{\text{rel}}^{\text{PF}} = \frac{I_{\text{CH}} + I_{\text{NH}} + I_{\text{PHOTON}} - f(\text{PU})}{p_{\text{T}}} \quad , \tag{2.2}$$

where I_{CH}, I_{NH} and I_{PHOTON} are the charged hadron, neutral hadron and photon isolation variables, defined as the sum of the transverse momenta of all particles falling inside a cone centered around the lepton track, and p_{T} is transverse momentum of the lepton itself. The $f(\text{PU})$ function is an additional term, depending on the amount of minimum-bias interactions, intended to correct the contribution to the isolation variable due to particles coming from pile-up vertexes, which will be discussed shortly. To exclude the lepton from its isolation variable computation, a veto cone, with radius smaller than the isolation cone, is established around the lepton, and those particles falling into it are not considered in the calculation[5]. A pictorial view of this cone scheme is presented in Figure 2.13a.



(a) (b)

Figure 2.13: (a) Definition of the isolation cones used in the lepton isolation calculation. Only for the particle candidates belonging to the isolation cone and not to the veto cone, the transverse momentum is contributing to the isolation variable. The value of the isolation variable is therefore depending on the activity around the lepton, which is typically low for leptons coming from W and Z bosons decays. (b) Probability that a muon from a Z decay is considered as "isolated" as a function of the number of spurious interactions. The plot demonstrates that the isolation variable defined in Equation 2.2 is invariant under different pile-up conditions. In addition, the plot shows that the isolation variable is well modeled in the Monte Carlo simulation.

The pile-up correction proceeds in two steps. Firstly, the charged hadrons, which are associated to some track, are taken into account only if coming from the primary interaction vertex, thus, I_{CH} is naturally pile-up invariant. Secondly, in order to correct the neutral part of the isolation

[5]The role of the veto cone is in fact more important for electrons. Electrons can emit low energy photons, close in space to the electron direction, and moreover part of their calorimetric deposit might not be clustered correctly and associated to the respective super-cluster. Hence, part of the energy is not linked to the electron object, but associated to low energetic objects very close in space to the electron position. In the case of electrons, the veto cone has been tuned to specifically address this issue.

variable, two methods are devised:

1. all particle-flow candidates are re-clustered into jets after adding a large number of very soft "ghost" particles uniformly to the event. Thus, a median energy density $\rho(\eta)$, expressed in unity of area, is obtained. Then, since this energy density contains also the charged contribution, which changes with the pseudorapidity, the Monte Carlo simulation is used to extract an effective area, $A_E(\eta)$, to be used as a multiplicative factor for $\rho(\eta)$ in data, and which translates the total energy density into a neutral energy density. The isolation variable then becomes

$$I_{\text{rel}}^{\text{PF}} = \frac{I_{\text{CH}} + \max(0, I_{\text{NH}} + I_{\text{PHOTON}} - (\text{A}_\text{E} \cdot \rho))}{E_{\text{T}}} \quad .$$

   This method is conventionally referred to as *effective area correction*;

2. one can estimate the energy coming from pile-up vertexes by looking at the charged hadron total energy from spurious-interactions. From Monte Carlo studies, it is found that the charged hadrons transport 2/3 of the total pile-up energy, in other words

$$I_{\text{NH}}^{\text{PU}} + I_{\text{PHOTON}}^{\text{PU}} = 0.5 \cdot I_{\text{CH}}^{\text{PU}} \quad .$$

   Therefore, using the last equation, a pile-up invariant isolation variable (Figure 2.13b) can be defined as

$$I_{\text{rel}}^{\text{PF}} = \frac{I_{\text{CH}} + \max(0, I_{\text{NH}} + I_{\text{PHOTON}} - 0.5 \cdot I_{\text{CH}}^{\text{PU}})}{p_{\text{T}}} \quad ,$$

   where this correction is usually referred to as *DeltaBeta correction*.

Sub-detector based isolation

The isolation variable definition based on particle-flow objects has been adopted in CMS only recently, namely during 2012 analyses. A previous definition, which still is commonly used to analyze 2011 $\sqrt{s} = 7$ TeV data and in general for triggering purposes, relies on sub-detector quantities, and it is interesting for the purpose of the electron studies contained in this thesis.

The sub-detector based isolation, for electrons, is based on three quantities:

- I_{TK}, i.e. tracker isolation. This is given by the p_T sum of all the tracks in a (η, ϕ) region defined by $\Delta R \in (0.04\,,\, 0.3)$ around the electron GSF track, with $p_T > 0.7$ GeV and $dz_{track} < dz_{ele} \pm 0.2$ [6];
- I_{EM}, i.e. electromagnetic isolation. This is the transverse energy of all ECAL deposits, with $|E| > 0.08$ GeV in EB and $|E| > 0.1$ GeV in EE, within a cone of radius $\Delta R = 0.3$ centered on the electron calorimetric position, excluding those in a inner 3-crystal radius cone [43];
- I_{H}, i.e. hadronic isolation, defined as the E_T sum of all HCAL calo-towers within a $\Delta R = 0.3$ cone centered on the electron position, excluding all the energy inside an inner cone of amplitude $\Delta R = 0.15$ [43].

[6] dz_{ele} is the minimum distance between GSF track and the nominal interaction point.

Additionally, a combined isolation variable, is defined as

$$I_{\text{rel}}^{\text{SD}} = \frac{I_{\text{TK}} + \max(I_{\text{EM}} + I_{\text{H}} - (\text{A}_{\text{E}} \cdot \rho))}{E_{\text{T}}} \quad ,$$

where the pile-up is taken into account using the effective area correction.

2.3.5 Jets and missing transverse energy

Jets are detectable in modern experiments as clusters of tracks and energy deposits in a defined region of the detector. Due to the intrinsic compositeness of such objects, a jet cannot be defined until an algorithmic procedure to recombine different daughter particles into a single mother jet is defined. In CMS, the adopted cluster algorithm is the so-called anti-k_{T} [47]; the algorithm has been chosen due to its computational speed and its collinear and infrared safeness, which ensures that theoretical predictions, made with this type of algorithm, are always finite.

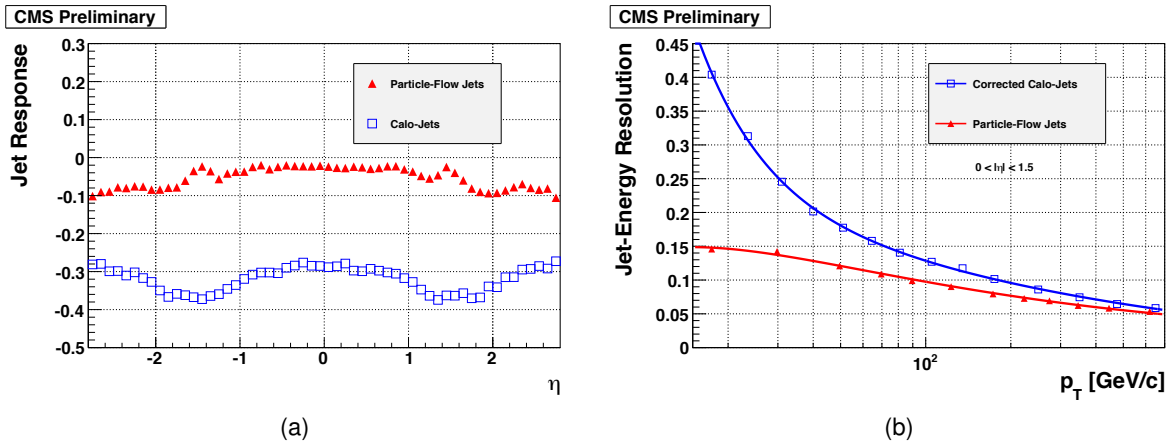
The jet momentum is determined as the vectorial sum of all particle momenta in it. Although important corrections are already applied at particle level during PF reconstruction, a set of further corrections have to be applied on reconstructed jets so that they can be used as high-level Physics objects. The jet correction scheme adopted in CMS is factorized into subsequent steps, each of them addressing a different physics aspect:

- Level 1 (offset) corrections: the purpose of this first step is to remove from the jet the additional energy coming from secondary pp interactions and the underlying event. This correction is determined both in data and in Monte Carlo on a event-by-event basis. Firstly, in the tracker acceptance region, only the charged particles coming from the primary interaction vertex are clustered in the jet. Secondly, to remove the contribution of neutral particles, or to take into account particles with $|\eta| > 2.4$, a different technique is used. All particle-flow candidates are re-clustered implementing a different algorithm (k_{T} instead of anti-k_{T}) and after adding a large number of very soft "ghost" particles uniformly to the event. The median energy density[7] ($\rho_{\text{PU}} = E/\Delta\eta/\Delta\phi$) of the many pseudo-jets so produced is taken as the estimate of the pile-up plus underlying event energy density for that event, and is subtracted from real jets, after being multiplied for the jet area (roughly πR^2) [48];
- Level 2 (relative) corrections: these corrections are meant to correct for non-uniformities in the different CMS sub-detectors by equalizing the jet response along η with respect to the center of the barrel;
- Level 3 (absolute) corrections: this last correction factor accounts for the non-linear response of the calorimeter to hadrons, the influence of the magnetic field, and undetected energy in the calorimeters. This absolute energy scale correction is derived from γ+jets events, where the event energy balance allows to compare the jet energy to the photon, precisely measured in ECAL.

Level 2 and 3 corrections are derived in simulated events, and further checked on real data via a closure test. Potential differences between data and Monte Carlo are accounted for with residual correction factors for jets in real data.

[7] The median is used since only little affected by the presence of few real, high-energy jets in the event.

Further details about the performances of PF jet reconstruction at CMS and their comparison with the standard calorimetric resolution can be found at [49]. As an example, Figure 2.14 reports the jet energy response and resolution expected from the simulation for PF and Calorimetric jets, reconstructed with an anti-k_T algorithm of R parameter 0.5, within the tracker acceptance.



(a) (b)

Figure 2.14: PF and Calo Jet response (a) and resolution (b) in the simulation. The jet response is defined by the ratio $(p_T^{rec} - p_T^{gen})/p_T^{gen}$, where p_T^{rec} and p_T^{gen} correspond to the reconstructed-level and generated-level jet transverse momenta. (b) Is obtained by requiring that the jet pseudorapidity η lies in the central part of the detector.

Finally, at the end of the particle-flow event reconstruction chain, it is in principle simple to determine $\vec{E}_T^{miss}$: it merely consists in forming the transverse momentum vector sum over all reconstructed particles in the event and then taking the opposite of this azimuthal two-vector momentum. Jet factorized energy corrections are propagated to the missing transverse energy computation to improve its resolution. A comparison of the particle-flow vs. the calorimetric missing transverse energy resolution is shown in figure 2.15.

$\vec{E}_T^{miss}$, in the hypothesis that all detectable particles are properly reconstructed, coincides with the vectorial sum of the momenta of all the undetectable particles (i.e. neutrinos, or BSM particles such as neutralinos in more exotic scenarios), since the incoming pp pair system has negligible transverse momenta ($\lesssim 1\,\text{GeV}$). In practice, since any experiment is not fully hermetical, a fraction of the final state energy is lost in the reconstruction process, so the $\vec{E}_T^{miss}$ is an approximation of the undetectable transverse energy.

In the experimental language, it is common to refer to the missing transverse energy E_T^{miss} (or MET) as the modulus of $\vec{E}_T^{miss}$. This convention will be used in the following part of the thesis.

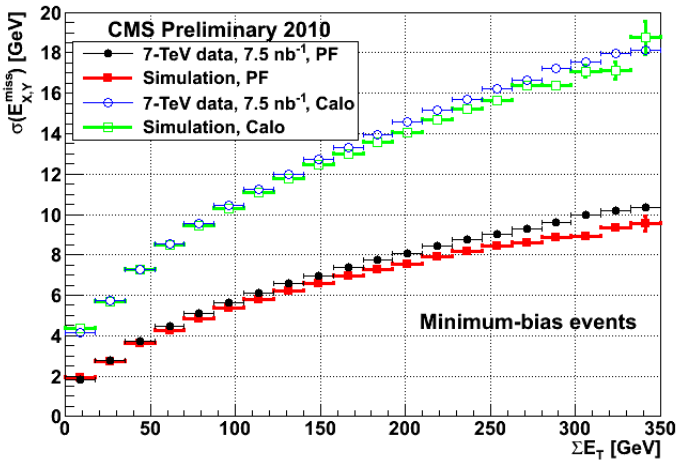


Figure 2.15: PF and Calo $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of the scalar sum of the event transverse energy for minimum bias events.

Chapter 3

CMS ECAL intercalibration with electrons from W and Z decays

In the $[110 \rightarrow 150]$ GeV/c^2 mass range, where the Higgs decay width is of the order of 10 MeV, the $H \rightarrow \gamma\gamma$ channel plays a prominent role in the search for the SM Higgs boson (see Chapter 1). The electromagnetic calorimeter of the CMS detector, ECAL, has been specifically designed to profit from such a small width, by providing an excellent energy resolution σ_E / E [36]. Recalling that

$$\frac{\sigma_E}{E} = \frac{A}{\sqrt{E(\text{GeV})}} \oplus \frac{B}{E(\text{GeV})} \oplus C,$$

where $A \sim 2\%$ and $B \sim 10\%$, it follows that the energy resolution for electrons and photons with $E_{e/\gamma} \geq 50$ GeV is dominated by the constant term C, as it is represented in figure 3.1. This constant term depends on the non-uniformity of the longitudinal light collection, the energy leakage from the rear face of the crystals, the instabilities in the operation of ECAL and the relative calibration constants accuracy: the targeted goal is to control all these aspects to set the C constant at 0.5% level.

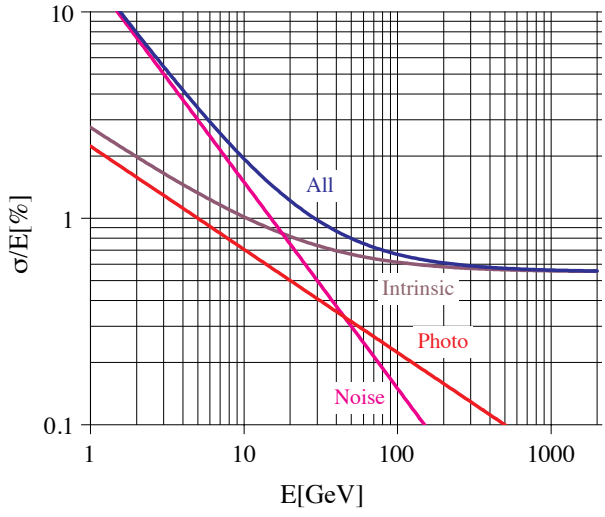
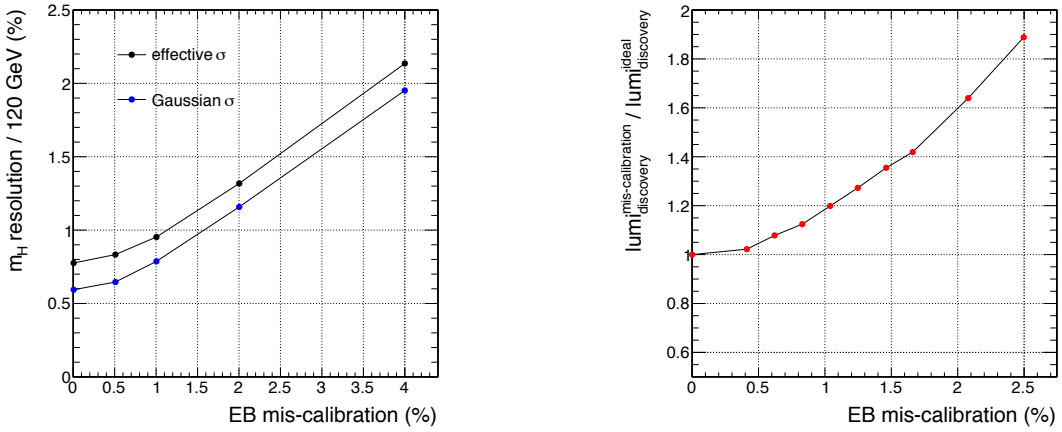


Figure 3.1: CMS ECAL Energy resolution and its breakdown in the different contributing sources.

In particular, the impact of the miscalibration on the physics analyses can be computed for the special case of the $H \to \gamma\gamma$ search, where the link between the ECAL energy resolution and the CMS physics reach is direct, as it is shown in figure 3.2.



(a) (b)

Figure 3.2: Effect of a given mis-calibration of ECAL barrel crystals on (a) the mass resolution of a 120 GeV/c^2 SM Higgs boson decaying into two photons and (b) on its discovery potential with respect to the target level calibrated detector. The gaussian resolution term, which depends on experimental conditions only, is added to the Higgs intrinsic width to obtain the effective σ of the expected signal mass spectrum.

Different techniques are used to fix the relative calibration of the ECAL crystals [50], and, in particular, this chapter presents the calibration method based on isolated electrons from W/Z decays. For convention, in this chapter, the relative calibration will be simply referred to as *intercalibration*.

A summary of the used datasets and Monte Carlo samples, as well as a short description of the trigger strategy used to select $W \to e\nu_e$ and $Z \to e^+e^-$ events is reported in Section 3.1. Then, a suitable set of selections necessary to isolate an almost pure W/Z sample is described in Section 3.2. The core of the calibration procedure, the L3 algorithm, is presented is Section 3.3, and Section 3.4 presents the closure test performed on MC samples, which also addresses the systematic limitation of the calibration procedure. Additional studies, which provide second order corrections, namely the momentum calibration (Section 3.5), are presented in Section 3.5. Then, Section 3.6 shows the results obtained with this method and the impact of the intercalibration on the $Z \to e^+e^-$ mass spectrum, and finally presents a digression on the role of these results for the $H \to \gamma\gamma$ search.

3.1 Datasets and Monte Carlo samples

3.1.1 Datasets used for the analysis

The data samples used in this analysis were recorded by the CMS experiment in 2011, at $\sqrt{s} = 7$ TeV. Only certified data are considered, which means that a good functioning of all CMS sub-detectors is required, so that the total analyzed statistics corresponds to an integrated luminosity of about 5 fb^{-1}.

A subset of the 2011 *Single Electron* dataset has been used for the analysis. In terms of content, all the events containing at least one electron are part of the used datasets, with some additional requirements, provided centrally, based on the same selections which will be listed in Section 3.2. Data have been processed with pre-existing calibration and alignment constants, based on early 7 TeV collision data. The specifics calibration sets used for the reconstruction are important for the calibration study under discussion, as will be highlighted in the following sections.

3.1.2 Monte Carlo samples

Although it is fairly easy to isolate in data an almost pure sample of W/Z events, it is nevertheless necessary to dispose of simulated events in order to validate the calibration procedure. Essentially, apart from W/Z signal process, it is interesting to simulate all those events which produces final state objects which are mis-identified as electrons (*fake electrons*), i.e. QCD multijet events. In general, the Monte Carlo event simulation provides the ideal tool for event modeling in the CMS environment (for more details see Section 4.1). In particular, for this study, MC events have been reconstructed with the very same calibration conditions of data, and data vs. MC differences in the pile-up modeling are properly taken into account with a dedicated procedure (see Section 4.2).

Table 3.1 summarizes the list of signal and background Monte Carlo samples employed for this study, together with their cross-section and the equivalent integrated luminosity L_{eq}.

MC Sample	cross-section (pb)	L_{eq} (fb^{-1})
/WJetsToLNu TuneZ2 7TeV-madgraph-tauola/	31314 ± 1600	2.55
/DYJetsToLL TuneZ2 M-50 7TeV-madgraph-tauola/	3048 ± 130	11.87
/TTJets TuneZ2 7TeV-madgraph-tauola/	163 ± 14	275.07
/QCD Pt-30to40 doubleEMEnriched TuneZ2 7TeV-pythia6/	~ 10868	~ 0.49
/QCD Pt-40 doubleEMEnriched TuneZ2 7TeV-pythia6/	~ 43571	~ 0.79
/QCD Pt-30to80 EMEnriched TuneZ2 7TeV-pythia/	~ 3625840	~ 0.018
/QCD Pt-80to170 EMEnriched TuneZ2 7TeV-pythia6/	~ 142814	~ 0.05
/QCD Pt-170to250 EMEnriched TuneZ2 7TeV-pythia6/	~ 3263	~ 0.92
/QCD Pt-250to350 EMEnriched TuneZ2 7TeV-pythia6/	~ 368	~ 0.72
/QCD Pt-350 EMEnriched TuneZ2 7TeV-pythia6/	~ 55	~ 43.82

Table 3.1: Summary of Monte Carlo samples used in the intercalibration procedure validation. Both the W and Z topologies are covered by the set of simulated samples.

3.1.3 Trigger strategy

W events are characterized by the presence of a real electron and a neutrino, which manifests itself experimentally through an unbalance in the missing transverse energy $E_{\text{T}}^{\text{miss}}$. The triggering matches this topology: events with at least one identified and isolated electron, with a given E_{T} threshold, are passing a first online filter; the online selection then probes the lepton-neutrino system by computing its transverse mass, which is defined as

$$m_{\text{T}} = \sqrt{p_{\text{T}}^{\text{lepton}} \cdot E_{\text{T}}^{\text{miss}} \cdot (1 - \cos\Delta\phi)} \quad , \qquad \text{with} \qquad \Delta\phi = \phi^{lepton} - \phi^{E_{\text{T}}^{\text{miss}}} \quad ,$$

and which is peaking around m_W for signal events and around zero for backgrounds. If the event m_T is greater than a fixed threshold, then it is selected and reconstructed offline.

Z events are triggered by the presence of at least one pair of identified and isolated electrons, with given asymmetric transverse momentum cuts.

In both cases, the selection criteria have been varying according to the LHC luminosity, which changed by an order of magnitude during 2011, up to $\mathcal{L} = 4 \cdot 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Since the online computing power and the DAQ bandwidth has been almost fixed, the triggers were re-tuned so that the global HLT rate remained constant, whilst keeping the signal selection efficiency as high as possible.

Both W and Z triggers are requiring quality cuts on electron candidates, based on the electromagnetic shower shape, the calorimetric energy deposition and isolation information from each sub-detector (tracker, ECAL and HCAL), adopting different working points. For most of the data taking, $W \rightarrow e\nu_e$ triggers employed a `WP80` working point (80% efficient on signal events), while $Z \rightarrow e^+e^-$ a looser `CaloTrk`[1] selection (90% efficient on real electrons). Alike the electron identification working points, thresholds on electron transverse momenta and m_T are changing, as it is shown in Figure 3.3. All these details are summarized in Table 3.2, where each object threshold is expressed in GeV.

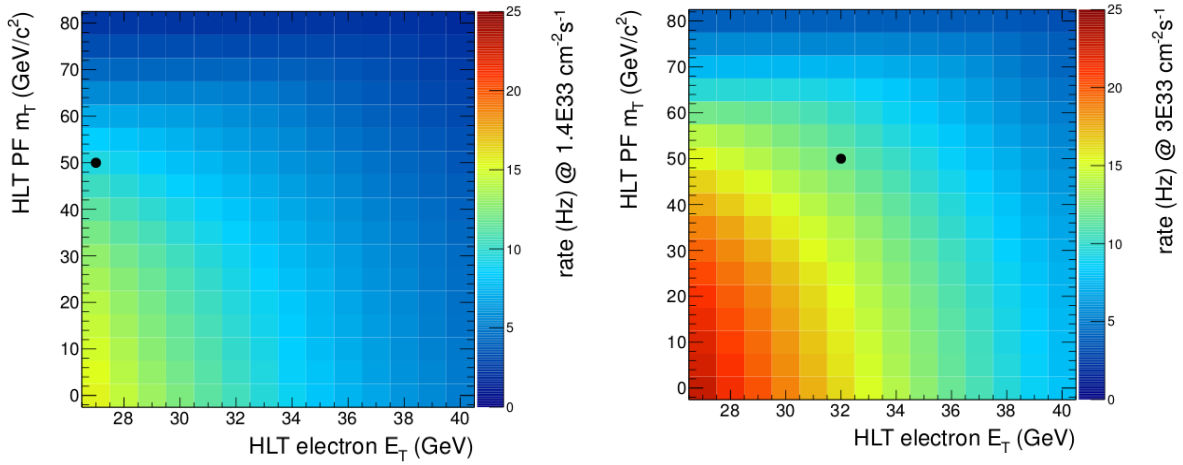


Figure 3.3: HLT Ele+m_T trigger expected rate as function of electron E_T and m_T for two different instantaneous luminosity: (Left) $1.4 \cdot 10^{33}\,cm^{-2}s^{-1}$, (Right) $3 \cdot 10^{33}\,cm^{-2}s^{-1}$. The black points represent the two working points, chosen by CMS, to cope with the two different luminosity regimes.

The trigger efficiency is taken into account in the Monte Carlo event generator by requiring the HLT emulation, which applies the HLT reconstruction on simulated events and matches the online requirements with the ones made on data. Additionally, to avoid problems related to the MC mismodeling on quantities that depend on soft-particles (i.e. isolation), the offline working points are chosen to be tighter than the online ones, in order to perform the analysis in a region where the HLT efficiency is uniform and approximately 100% (more details on the trigger efficiency are provided in Section 4.4).

[1]The full name of the working point is `CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL`, where each substring refers to a quality criterion imposed on the EM shower (Calo) and isolation (Iso) variables. T and VT stand for the "tight" and "very loose" working points respectively.

$W \to e\nu_e$ HLT	Run Range
Ele27_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT	160404-163869
Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT	165088-165633
Ele25_WP80_PFMT40	165970-166967
Ele27_WP80_PFMT50	167039-167913
Ele32_WP80_PFMT50	170249-180252
$Z \to e^+e^-$ HLT	Run Range
Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL	160404-167913
Ele17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL	170249-180252

Table 3.2: HLT paths used and relative run ranges for $W \to e\nu_e$ (Top) and $Z \to e^+e^-$ (Bottom) event selection.

3.2 The event selection

The event selection is aimed towards the identification of final states containing either an electron and a neutrino or two electrons, corresponding to W and Z decays respectively. In this section, the criteria applied on single objects are described and then requirements made at event-level are detailed.

3.2.1 Object reconstruction

The analysis relies on the standard reconstruction algorithms produced by the CMS community. The particle-flow reconstruction procedure (Section 2.3.1) was used to coherently define the collection of jets, leptons and E_T^{miss} considered in the event selection.

Electrons

A set of offline selections are applied to provide a sample of high purity electrons. Different sets of cuts are tuned on MC samples by optimizing the rejection of fake electrons at a given level of efficiency ϵ on real electrons, i.e. electrons coming from W/Z decays, and the naming convention refers to such efficiency (eg. `WP80` corresponds to ϵ = 80%).

A first list of cuts acts on the reconstructed electron proprieties, namely:

- $\Delta\phi_{in}$ and $\Delta\eta_{in}$, i.e. the difference, respectively in ϕ and in η, between the track position extrapolated to the ECAL surface, and the super-cluster (η, ϕ) energetic barycenter;
- $\sigma_{i\eta i\eta}$, i.e. the weighted energy cluster covariance, made along η inside a 5×5 matrix centered on the most energetic deposit position. this quantity value is typically smaller for real electrons than for QCD fakes, such as π^0;
- H/E, namely the ratio between HCAL deposits, within a ΔR=0.1 cone centered in (η_{SC}, ϕ_{SC}), and the super-cluster electron energy;
- missing/ambiguous track hits and displaced e^+e^- vertexes, i.e. conversion veto. The former are relevant since electrons coming from photon conversions might be produced outside internal tracker layers, hence have a track which lack the most internal hits. The latter is performed by looking at close GSF tracks corresponding to opposite charge leptons, where the closeness is measured through $|\Delta \cot\theta|$

and |dist|, which define the distance between the two tracks in the longitudinal and transverse planes respectively (see Figure 3.4).

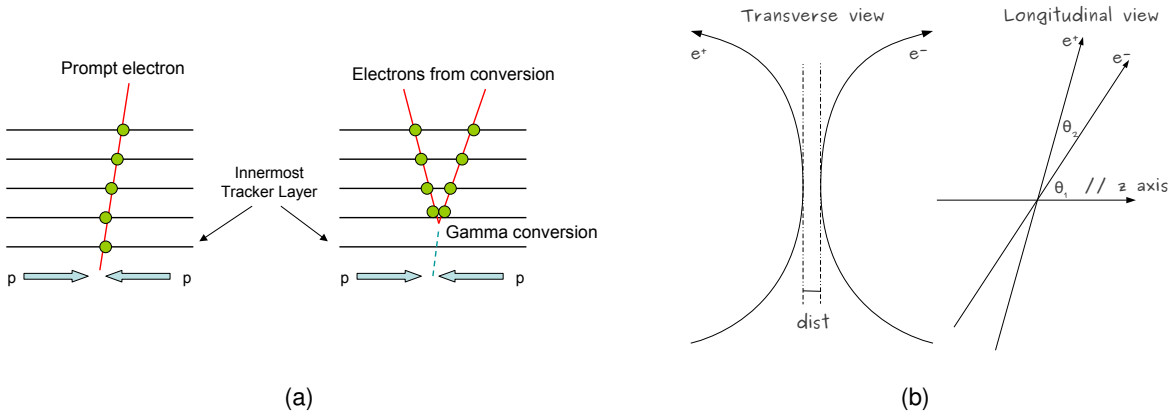


Figure 3.4: (a) Tracks associated to prompt electrons originating from the primary proton-proton collision vertex have detected hits on the innermost tracker layer. Conversely, tracks associated to electrons from converted photons often do not have hits on the first layer due to photons converting at larger radii. (b) Pictorial representation of the variables used for the tracks closeness measurement, used in the photon conversion veto. In the transverse plane, the minimum distance of the track defines the |dist| variable. In the longitudinal plane, the electron tracks have an associated angle θ with respect to the z axis, defined by the proton beam. By taking the absolute difference of the angles cotangents $|\Delta \cot\theta| = |\cot\theta_2 - \cot\theta_2|$, one can define a track distance in the longitudinal plane.

In addition, electrons are required to be isolated, i.e. not surrounded by high-energy activity, which is the typical situation for a reconstructed electron inside a jet. The isolation used for the calibration study is sub-detector based and it has already been defined in Section 2.3.4. As a short reminder, the formula for the combined isolation I^{SD}_{rel} reads as

$$I^{SD}_{rel} = \frac{I_{TK} + \max(I_{EM} + I_H - (A_E \cdot \rho))}{E_T} \quad ,$$

where I_{TK}, I_{EM} and I_H are, respectively, the tracker, electromagnetic and hadronic isolation variables and $A_E \cdot \rho$ represents the terms which accounts for isolation contributions coming from pile-up interactions.

Different working points, used for the different topology searches ($W \to e\nu_e$ vs $Z \to e^+e^-$), are chosen, as it is described in the following paragraphs and summarized in Table 3.3.

3.2.2 Event-level criteria

Events are selected with two main workflows, according to the target topology.

W selection

$W \to e\nu_e$ events are selected requiring:

- exactly one electron with $E_T > 30$ GeV and $|\eta| < 2.5$, passing the `WP70` electron ID;
- a veto on looser identified electrons, i.e. passing the `WP80` electron quality cuts;
- an azimuthal angle separation between the electron and E^{miss}_T greater than $\pi/2$;

	WP80		WP70	
Variable	EB	EE	EB	EE
$\lvert\Delta\eta_{in}\rvert$	0.004	0.007	0.004	0.005
$\lvert\Delta\phi_{in}\rvert$	0.060	0.030	0.030	0.020
$\sigma_{i\eta i\eta}$	0.010	0.030	0.010	0.030
H/E	0.040	0.025	0.025	0.025
I_{TK}	0.09	0.04	0.05	0.025
I_{EM}	0.07	0.05	0.06	0.025
I_{H}	0.10	0.025	0.03	0.02
$I^{\mathrm{SD}}_{\mathrm{rel}}$	0.070	0.060	0.04	0.03
$miss^{hits}_{tk}$	0	0	0	0
$\lvert\Delta\cot\theta\rvert$	0.02	0.02	0.02	0.02
$\lvert$dist$\rvert$	0.02	0.02	0.02	0.02

Table 3.3: The quality cuts applied on electrons for the WP80 and WP70 working points. Top, central and bottom group of cuts correspond to identification, isolation and conversion veto requirements.

- missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV and transverse mass $m_T = \sqrt{2E_{\mathrm{T}}^{\mathrm{miss}}E_T^e(1-\cos\Delta\phi)}$ greater than 50 GeV.

The latter cuts are specifically designed and optimized to reduce fake electrons contamination, which is due to QCD multijet events.

Z selection

$Z \to e^+e^-$ events are selected requiring:

- exactly two electrons with opposite charge passing WP80 electron ID;
- $E_{\mathrm{T}}^{\mathrm{miss}} < 40$ GeV and $60\,\mathrm{GeV} < m_{\mathrm{ee}} < 120$ GeV.

3.2.3 Data - Monte Carlo comparisons

To assess the quality of the modeling provided by the MC simulation, comparisons between the MC shapes, normalized to the collected integrated luminosity, and the data, after applying the event selection criteria listed above, are performed.

Contrarily to other studies, for calibration purposes the simulation is used only in specific tasks, since, in the end, the procedure is relying on data only. Therefore, the generic observable has not to be well modeled, since the precise selection efficiency measurement is not relevant, but some distributions, such as the energy over momentum ratio and one of the shower shape variables (R9, described later), are fed to the calibration algorithm and hence must be well described by the simulation if the studies on the calibration procedure presented in the following are to be meaningful. The data vs. MC agreement for these quantities is at 10% level, as it is shown in Figure 3.5. Additionally, Figure 3.6, shows a good agreement for the number of reconstructed vertexes, which is one of the variable most sensitive to PU interactions. This validate the use of the PU reweighting scheme in the Monte Carlo simulation.

On top of the intercalibration validation purposes, the other relevant use of the MC is the assessment of the background contamination in the signal region, which, as it can be seen in

the listed figures, is at 1% level for both the $W \to e\nu_e$ and $Z \to e^+e^-$ selections.

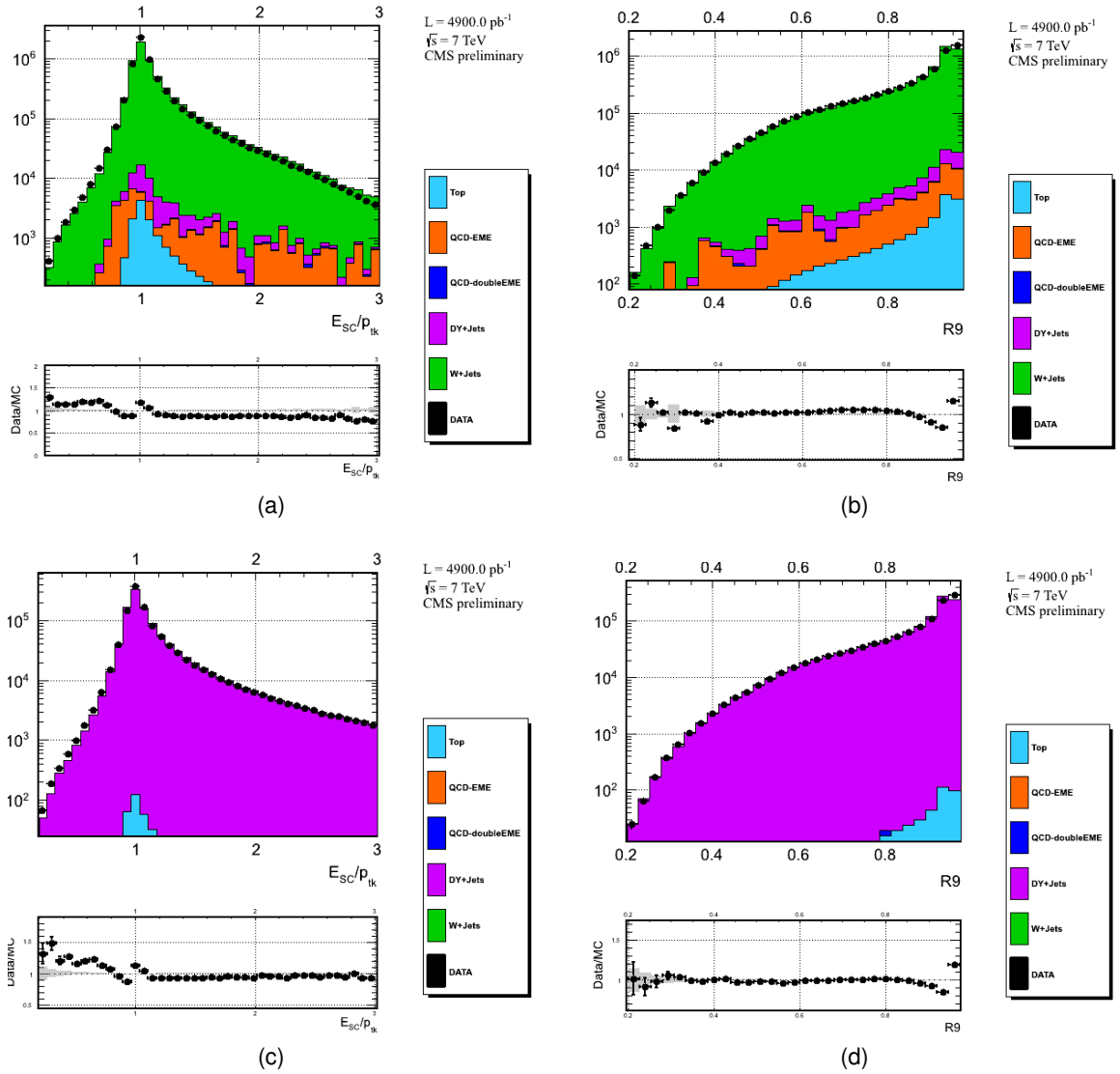


Figure 3.5: Data vs. MC comparison for the main variables used for the ECAL intercalibration. On the left, the energy over momentum ratio, on the right, the R9 variable, used to estimate the electron energy fraction emitted through bremsstrahlung. Top and bottom rows refer to $W \to e\nu_e$ and $Z \to e^+e^-$ selections respectively.

3.2.4 Special selections for the calibration procedure

In addition to the listed selections, additional cuts were designed specifically to improve the calibration procedure, namely the statistical precision and the convergence speed (Section 3.3). In particular, a selection based on the R9 variable has proven to be useful.

R9 is defined as E_{3x3}/E_{sc} and its value is typically high for non-radiating electrons, since most of the super-cluster energy is deposited in a 3x3 matrix around the highest energetic deposit (seed crystal) and not spread along ϕ. Thus, one can use this class of electrons and study the effects on the calibration: in the barrel, selecting high R9 electrons leads to worst performances, while in EE, tuning an R9 cut as function of η_{sc}, allows a clear improvement in the achieved

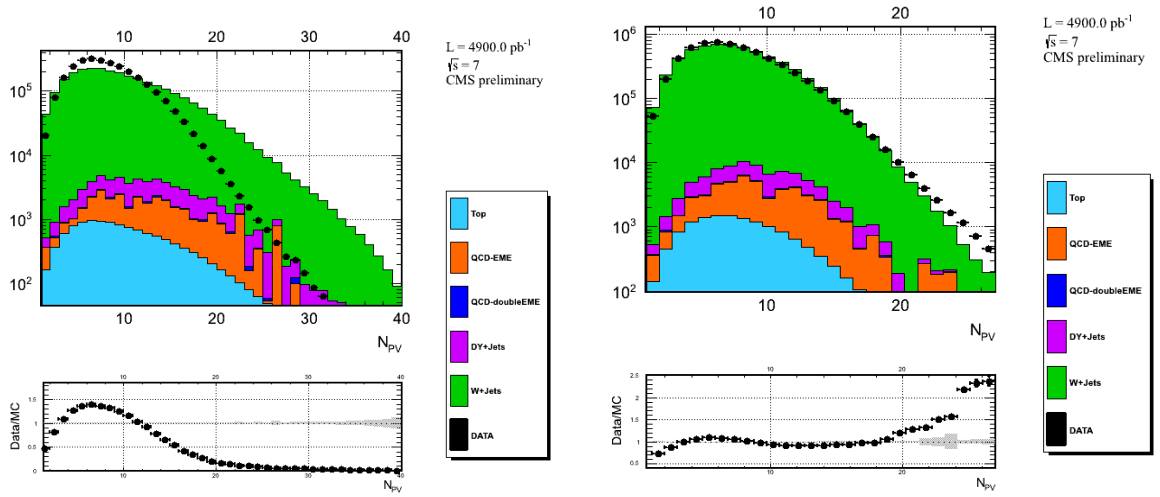


Figure 3.6: Data vs. MC comparison of the number of reconstructed primary vertices (a) before and (b) after the re-weighting for the pile-up is applied. The plots are obtained by looking at $W \to e\nu_e$ events..

intercalibration precision (defined in Section 3.4). This opposite effect is due to an interplay between the lowering in statistics, due to the R9 cut, and the improvement of the sample quality, since the selected electrons have a lower probability of having part of their energy which is not reconstructed[2], and hence of introducing a bias. Since the material budget increases with the electron pseudorapidity, the radiating probability is changing along η, and this determines the different trade-off in the barrel and in the endcaps.

The optimal cut values are found to be:

$$\begin{aligned} 0 < |\eta_{sc}| < 1.50 &\to \text{nocut} \\ 1.5 < |\eta_{sc}| < 1.75 \to R9 > 0.80 \quad &; \quad 1.75 < |\eta_{sc}| < 2.0 \to R9 > 0.88 \\ 2.0 < |\eta_{sc}| < 2.15 \to R9 > 0.92 \quad &; \quad 2.15 < |\eta_{sc}| < 2.5 \to R9 > 0.94 \quad . \end{aligned} \tag{3.1}$$

Figure 3.7 shows the seed crystal occupancy as a function of η-ring, normalized to 1 fb^{-1} of data. The plot reports the occupancy before and after the additional R9 cut. With the full selections, each crystal in ECAL barrel is hit, on average, by 20-25 good electrons for each fb^{-1} of collected luminosity. The number mildly increase at higher eta, due to the different solid angle coverage of each crystal, which is increasing with the pseudorapidity[3]. The visible dip at $\eta \approx 1.45$ corresponds to the barrel-endcap transition region, where the tracker material, hence the bremsstrahlung probability, is maximal.

3.3 The calibration algorithm

As was shown in Section 3.2, by applying some selections, an high purity sample of real electrons from W and Z decays can be obtained. W/Z electrons are particularly interesting since

[2]This could happen if the cluster corresponding to the radiated photon is not associated to the electron track.

[3]The crystal solid angle coverage at $\eta = 0$ corresponds to $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$, and increases progressively in the forward region up to $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$.

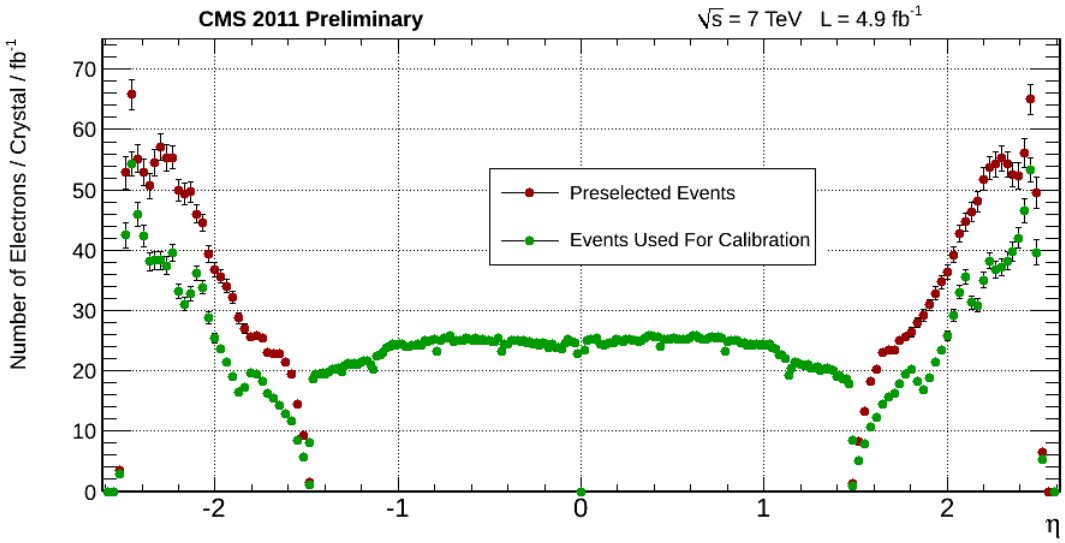


Figure 3.7: Average number of electrons for each ECAL crystal per fb^{-1} as a function of η_{sc}. Standard selections (red) are applied in the barrel, while in the endcap an additional requirement on electron R9 is performed (green).

their energy is such that they can be considered massless, so that the relation $E_\mu = p_\mu$ becomes a good approximation; thus, by measuring the momentum with the CMS tracker, the ECAL can be calibrated by constraining the measured super-cluster energy E_{sc} to be equal to the measured p_{tk}, in other words by imposing that $E_{sc}/p_{tk} = 1$.

First of all, one notices that, due to the different bremsstrahlung probability, and hence different reconstruction conditions (clustering, bremsstrahlung photon recovery, ecc...), the reconstructed energy is affected differently along η, therefore a relative pseudorapidity scale must be obtained. This can be done with $Z \to e^+e^-$ events, by considering di-electron pairs falling at the same pseudorapidity. Thus, for a given pseudorapidity η, a given mass spectrum $m_{ee}(\eta)$ is obtained. From this spectra, the invariant mass of the Z boson can be measured, and the electron energies can be corrected, with a function of the electron pseudorapidity, so that the measured mass matches the predicted one[4]. The correction gives the ECAL pseudorapidity scale. The above method will not be specifically addressed in this manuscript.

Secondly, by comparing the $E_{\mathrm{sc}}/p_{\mathrm{tk}}$ distributions for each crystal at a given pseudorapidity, one can fix the relative scale in the azimuthal direction. In practice, a relative scale, such that the different distributions overlap each other, is computed with a specific algorithm. Since the scale is relative, one can fix an arbitrary average value for $E_{\mathrm{sc}}/p_{\mathrm{tk}}$ for a given η, which is by convention chosen to be one. This particular procedure constitutes the object of this chapter.

This section firstly summarizes how the electron energy is built from single crystal deposits and then presents the core of the azimuthal intercalibration procedure, i.e. the L3 algorithm.

[4]The relation between the Z mass spectrum and the electron-positron energies is throughly discussed in Section 4.4.

3.3.1 ECAL energy reconstruction

The energy $E_{e,\gamma}$ is spread over several ECAL crystal, and has to be firstly re-clustered through dedicated algorithms (Section 2.3.3), then, in formulae, it is given by

$$E_{e,\gamma}^{\mathrm{ECAL}}\,(\,\mathrm{GeV}) = F_{e,\gamma}\,(E_{\mathrm{T}},\eta)\cdot G\cdot \sum_{i\,\in\,\mathrm{cluster}} S_i(t)\cdot c_i\cdot A_i \quad , \tag{3.2}$$

where the sum is over the crystals in a super-cluster. A_i are the reconstructed amplitudes in ADC counts, $S_i(t)$ the transparency correction factor [51] and G is the ECAL global energy scale. The factor $F_{e,\gamma}$ is defined as an additional energy correction which depends on the type of the particle, its energy and pseudorapidity, and in particular takes into account shower leakage and bremsstrahlung losses.

The interesting terms, for the thesis purpose, are the c_i, i.e. the intercalibration coefficients. These are meant to account for crystal-to-crystal variations by equalizing the different crystal responses.

3.3.2 The L3 algorithm

The intercalibration coefficients (c_i) are calculated using an iterative algorithm, which employs a principle introduced in the electromagnetic calorimeter calibration of the LEP L3 experiment[5]. At each iterative step, intercalibration coefficients are given by:

$$c_i^N = c_i^{N-1}\times \frac{\sum_{j=1}^{N_e}\omega_{ij}\cdot f\left(\frac{E_{sc}}{p_{tk}}\right)\Big|_j\cdot\left(\frac{p_{tk}}{E_{sc}}\right)_j}{\sum_{j=1}^{N_e}\omega_{ij}\cdot f\left(\frac{E_{sc}}{p_{tk}}\right)\Big|_j} \quad , \tag{3.3}$$

where :

- N is the iteration index;
- c_i is the single crystal intercalibration coefficient, identified by the crystal index i, or by the integer coordinates $(i\eta, i\phi)$, in EB, and (ix, iy), in EE;
- N_e is the number of selected electron candidates. This number is equal to $N_W + 2N_Z$, where N_W and N_Z correspond to the selected W and Z events respectively;
- ω_{ij} is a weight which accounts for the fact that the energy of the j-th electron is falling in different crystals with different fractions, defined by the ratios between the i-th crystal energy and the super-cluster one: $\omega_{ij} = \frac{E_{ij}^{recHit}}{E_j^{sc}}$;
- $\left(\frac{p_{tk}}{E_{sc}}\right)_j$ is the ratio between the tracker momentum and the super-cluster energy for the electron j;
- $f\left(\frac{E_{sc}}{P_{tk}}\right)$ is a weight on the electron j. It corresponds to the probability of finding an electron with energy E_{sc} and momentum p_{tk} in a given barrel or endcap ring, identified by η_{sc}.

The algorithm is run independently for the barrel and each endcap (EE+ and EE-), following

[5] For this reason this calibration procedure is called L3 method.

the same procedure for each iteration cycle.

First of all, the E_{sc}/p_{tk} distributions are built for each η-ring ($85 \times 2 + 1$ in EB and 40 for each endcap), and each electron is assigned to a specific ring by considering the coordinate of the most energetic crystal inside the electron super-cluster, $i\eta_{seed}$. At each step, the super-cluster energy is re-computed using the latest intercalibration coefficient values, through the formula:

$$E_j^{sc} = \sum_{k=0}^{N_{Hit}} E_{jk} \cdot F_e(\eta, p_T) \cdot c_k(i\eta, i\phi) \quad , \tag{3.4}$$

where k is the crystal index, E_{jk} is the single crystal energy corrected for transparency loss and $c_k(i\eta, i\phi)$ is the corresponding intercalibration value. Only energy deposits corresponding to healthy crystals are considered to build up E_{sc}, while noisy or dead channels are discarded. Also $F_e(\eta, p_T)$ is re-computed, as:

$$F_e(\eta, p_T) \approx \frac{E_{sc}}{E_{sc}^{Raw}} \text{ in EB} \quad ; \quad F_e(\eta, p_T) \approx \frac{E_{sc}}{E_{sc}^{Raw} + E_{ES}} \text{ in EE} \quad . \tag{3.5}$$

Once the E_{sc}/p_{tk} distributions are obtained, the numerator and the denominator of Equation 3.3 are calculated for each event, and the new intercalibration constants are computed. The iterations are repeated until the calibration constants converge.

Once the convergence has been reached, the c_i are re-normalized for each η-ring, such that the mean value of the c_i at a given $i\eta$ is equal to one.

3.4 L3 calibration performance studies on simulated events

The performance of the L3 algorithm is studied on the Monte Carlo samples, where additional mis-calibration effects, for eg. due to the laser corrections, are not present and where the value of the true calibration constants is known.

In practice, the calibration procedure is run on a single Monte Carlo sample, i.e. W+jets (see Section 3.1). The sample is sufficient to simulate the data after selections, since, as was explained in Section 3.2, the analysis cuts are designed to produce an almost pure W sample. The study on Z simulation is redundant, for calibration purposes, due to the fact that the Z events are equivalent to the W ones, except the number of clean electrons per event (2 against 1).

3.4.1 Convergence and statistical precision

To reduce computing-time consumption, the number of iterations is optimized by looking at the level of convergence, or, equivalently, at the intercalibration constants precision. The equivalence is in fact a first sanity check of the algorithm, since negligible corrections to the intercalibration constants (c_i convergence) must imply a saturation in the intercalibration precision.

To perform this test, the intercalibration precision has to be computed. To do so, for each η-ring[6], the precision is estimated through the Gaussian fit of the intercalibration constants

[6]The intercalibration constant precision is usually investigated as function of η and not of ϕ, since the ECAL response is expected to be uniform in ϕ but not in η. This can be related to different tracker material budget vs η, ECAL geometry and different radiation flux.

distributions. The intercalibration dispersion depends on the number of collected electrons per crystal, according to

$$\frac{\sigma_c}{c}(\eta) = \frac{a(\eta)}{\sqrt{n_e(\eta)}} \oplus b(\eta) \quad , \tag{3.6}$$

where

- σ_c and c are, respectively, the width and the mean of the gaussian distribution obtained by considering all the c_i for a given η-ring;
- $a(\eta)$ is the stochastic term;
- $n_e(\eta)$ is the average electron number per crystal, which is uniform within the same η-ring, since the azimuthal electron distribution is flat;
- $b(\eta)$ is a residual term, which depends on the systematic bias related to the algorithm itself, the momentum measurement inaccuracies and the accuracy of the initial calibration coefficients. The latter point is particularly important on data, where it can happen that the spread of the intercalibration after the L3 procedure is dominated by initial miscalibrations.

The latter point also shows that, in order to be decoupled from initial conditions, it is necessary to find a method to compute the statistical precision of the intercalibration constants.

The statistical precision is the first term of the Equation 3.6

$$\frac{\sigma_c}{c}(\eta) = \frac{a(\eta)}{\sqrt{n_e(\eta)}} \oplus b(\eta) = \frac{\sigma_c}{c}(\eta)\Big|_{stat} \oplus \frac{\sigma_c}{c}(\eta)\Big|_{res} \quad \rightarrow \quad \frac{\sigma_c}{c}(\eta)\Big|_{stat} = \frac{a(\eta)}{\sqrt{n_e(\eta)}} \quad . \tag{3.7}$$

Considering the random variable z, defined as

$$z(i\eta, i\phi) = \frac{c_{ev}(i\eta, i\phi) - c_{odd}(i\eta, i\phi)}{c_{ev}(i\eta, i\phi) + c_{odd}(i\eta, i\phi)} \quad , \tag{3.8}$$

where $c_{ev}(i\eta, i\phi)$ and $c_{odd}(i\eta, i\phi)$ are the coefficient maps obtained by running the calibration algorithm, independently, on even and odd sub-samples of W/Z events, one finds, for each ring, that

$$\sigma_z = \frac{2}{(c_{ev} + c_{odd})^2} \cdot \sqrt{c_{ev}^2\sigma_{odd}^2 + c_{odd}^2\sigma_{ev}^2 - 2(c_{ev}c_{odd})\sigma_{cov}} \tag{3.9}$$

$$\sigma_{cov} = \rho_{ev,odd} \cdot \sigma_{ev} \cdot \sigma_{odd} \quad ; \quad \sigma_{ev} \approx \sigma_{odd} = \sqrt{\sigma_{stat}^2 + \sigma_{sys}^2} \quad . \tag{3.10}$$

Since odd and even events are statistically independent, only the systematic term contributes to the covariance σ_{cov}, by means of a correlation factor equal to one. In addition, since the c_{ev} and c_{odd} are the means of the c_i distributions corresponding to a given ring, they can be considered equal to one, due to the freedom in the normalization choice. Therefore, one finds that

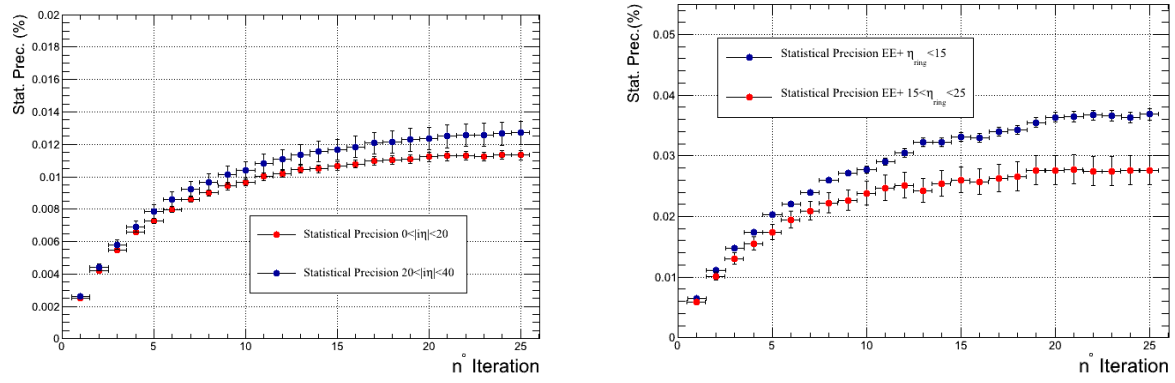
$$\sigma_z \approx \frac{2}{(c_{ev} + c_{odd})^2} \cdot \sqrt{c_{ev}^2 + c_{odd}^2} \cdot \sigma_{stat} = \frac{\sigma_{stat}}{\sqrt{2}} = \sigma_{stat}^{FULL} \quad , \tag{3.11}$$

where the last equality is obtained by reminding that σ_{stat} is the statistical precision of either

the even or odd subsamples, which is $\sqrt{2}$ greater than the un-merged dataset one σ_{stat}^{FULL}. This last equation states that the width of the random variable z allows a quantitative estimation of the statistical precision.

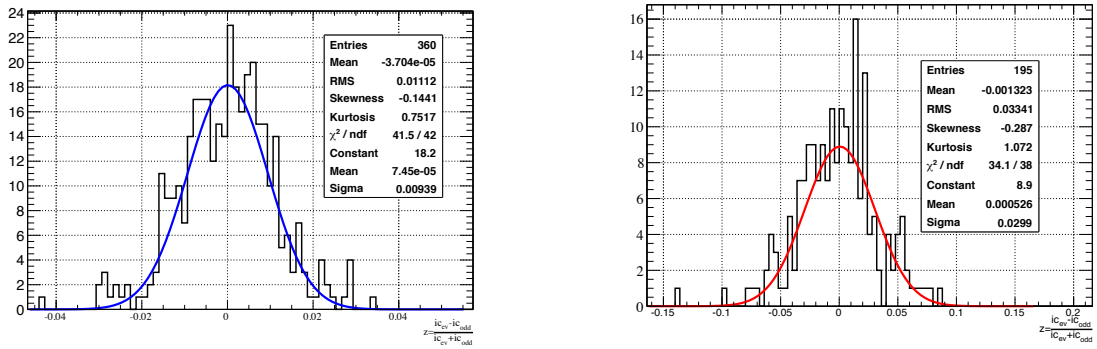
3.4.1.1 Results

Figure 3.8 shows the statistical precision as a function of the iteration cycle, for different η regions. Each point is obtained by fitting the $z = \frac{c_{ev}-c_{odd}}{c_{ev}+c_{odd}}$ distributions obtained using all the c_i constants, at each iteration, belonging to the different regions. Two examples of fits, for single η-rings, are reported in the plots contained in Figure 3.9. The fact that the initial intercalibration precision is better for a lower number of iterations is an artifact: initially all the c_i have the default value equal to the unity, and since the L3 algorithm takes some iterations to change significantly the value of the constants, the initial z variable shape is a delta function centered on zero, being $c_{ev} - c_{odd} \approx 0$.



(a) (b)

Figure 3.8: Calibration precision as a function of the iteration step in EB (a) and EE (b).



(a) (b)

Figure 3.9: Gaussian fit of $z = \frac{c_{ev}-c_{odd}}{c_{ev}+c_{odd}}$ for a central η-ring in EB (a) and a forward one in EE (b). The intercalibration constants used to obtain the z distributions are computed with 25 algorithm cycles.

The precision vs. $i\eta$ shows two proprieties:

- running the calibration algorithm 25 times saturates the statistical dispersion of the c_i, at any given pseudorapidity. This means that the convergence is ensured with a limited number of iterations;

- the statistical precision is increasing with $i\eta$. The explanation is twofold. Firstly, the number of reconstructed electrons per crystal decreases with increasing pseudorapidity, since the bremsstrahlung probability rises. Secondly, although at very high η the electron statistical population rises again due to the augmented crystal angular coverage, the dominating effect is the degradation in the tracker resolution [35].

The first propriety can be also cross-checked by looking at the E_{sc}/p_{tk} distributions for each η-ring as a function of the iteration. These can be thought as the convolution of each crystal E_{sc}/p_{tk}, which, if the calibration works as designed, will tend to overlap on each other hence reducing the width of the convoluted sum. In addition, due to the particular choice of the reference value of $E_{\mathrm{sc}}/p_{\mathrm{tk}}$, the convoluted sum will peak at one. These effects are verified to be true for the L3 algorithm, as it is shown in Figure 3.10.

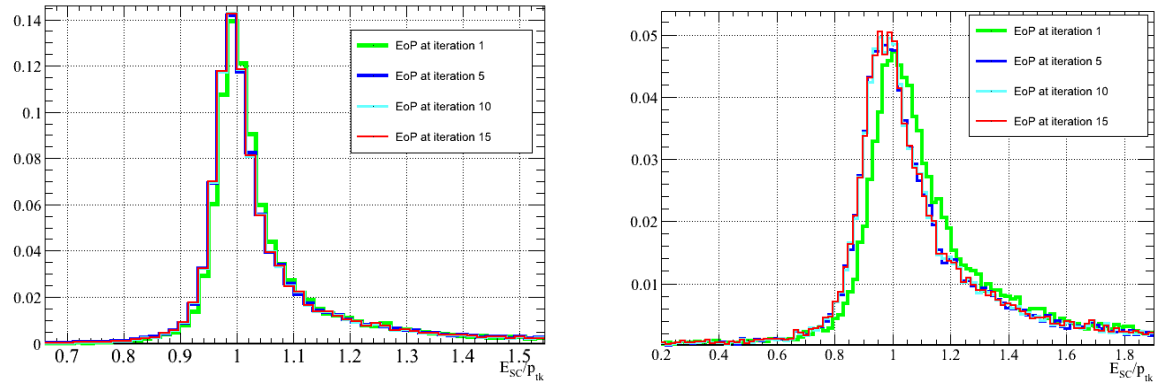


Figure 3.10: Evolution of E_{sc}/p_{tk} distribution, for a given η-ring, vs. L3 iteration in the central part of EB (a) and in the forward EE (b).

3.4.2 Systematic limits of the L3 algorithm

The Monte Carlo study has shown that statistical convergence holds for the L3 method. Then, on data, the method can be applied and the statistical precision can be extracted with the usual z variable. The other interesting parameter, to get the total intercalibration precision, is the systematic term in Equation 3.6. Unfortunately, on data, the residual dispersion of the intercalibration dispersion is not entirely associated with the L3 method itself, as other factors, like the initial miscalibration conditions, are contributing.

Monte Carlo simulation can again be exploited to reduce many of the sources of uncertainties present in real data, and hence studied to extract the algorithm systematic limitations. In principle, the correct way to estimate the statistical limitation of the procedure would be to dispose of hundreds of fb^{-1} of integrated luminosity L, so that, for different L values, the intercalibration dispersion could be evaluated. Then the obtained values could be fitted with a $a/\sqrt{(L)}+b$ curve, and the value b would provide the interesting term. In practice this is not possible, since the simulation has a little statistics, equivalent to 2.5 fb^{-1} of data.

On the other hand, the Monte Carlo can be used to perform another interesting and mandatory study, which is to check the convergence of the calibration procedure starting from a dataset which is mis-calibrated at a given level. To do so, the statistical precision σ of the method is

firstly computed on the Monte Carlo sample without any correction.
Secondly, the simulated electron energy is smeared with a gaussian miscalibration map $c_{\text{misc}}(i\eta, i\phi)$, with width equal to 5%, and the statistical precision σ_{misc} is recomputed. The convergence of the algorithm is verified to hold, with 25 iterations, also in this case.

If a difference between σ^{stat} and $\sigma^{stat}_{\text{misc}}$ is found, then it must be attributed to a bias of the L3 method itself, which is not capable of fully recovering a given miscalibration condition. This difference σ_{syst} is computed as a function of the crystal pseudorapidity (Table 3.4), and it is setting an upper limit on the final intercalibration precision[7].

η range	σ^{stat}	$\sigma^{\text{stat}}_{\text{misc}}$	$\sigma_{\text{syst}} = \sqrt{(\sigma^{\text{stat}}_{\text{misc}})^2 - (\sigma^{\text{stat}})^2}$
$0.00 < \lvert\eta\rvert < 0.35$	1.15%	1.20%	0.32%
$0.35 < \lvert\eta\rvert < 0.70$	1.27%	1.30%	0.27%
$0.70 < \lvert\eta\rvert < 1.05$	1.55%	1.58%	0.30%
$1.05 < \lvert\eta\rvert < 1.48$	3.32%	3.43%	0.85%
$1.53 < \lvert\eta\rvert < 1.80$	3.72%	3.83%	0.91%
$1.80 < \lvert\eta\rvert < 2.00$	2.78%	2.98%	1.07%
$2.00 < \lvert\eta\rvert < 2.50$	3.40%	3.77%	1.63%

Table 3.4: Systematic limitation on the intercalibration precision due to initial miscalibrations, computed as a function of the crystal pseudorapidity. Top group refers to the barrel, bottom to the endcaps. The values are obtained by considering an initial miscalibration level of 5%.

3.5 Tracker momentum bias and correction

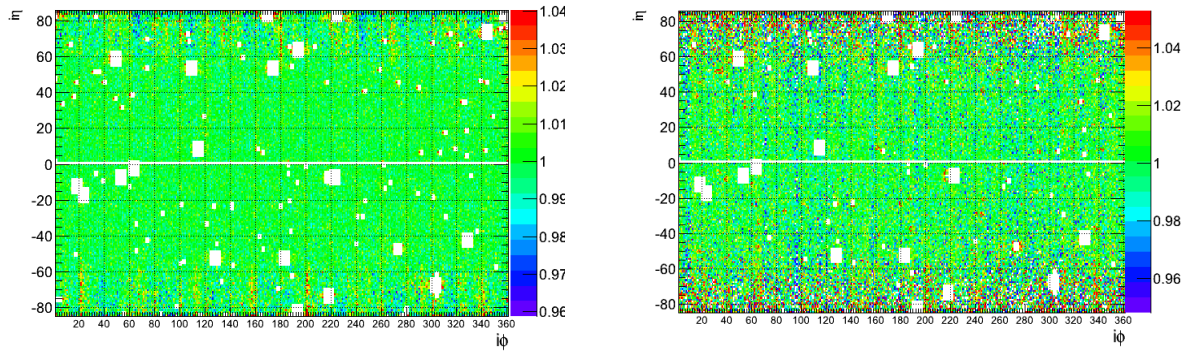
As was already noted, the intercalibration method presented here is based upon the measured E_{sc}/p_{tk} ratio, and it is therefore sensitive to energy and momentum measurement biases along the azimuthal direction.

Firstly, Monte Carlo has been used to check which of the measurement mostly affect the calibration procedure. The intercalibration procedure has been tested by exploiting the E_{sc}/E_{true} ratio, where E_{true} is the generator level quantity associated to the electron 4-vector. The resulting map is shown in Figure 3.11a and it can be compared with the standard map, obtained with E_{sc}/p_{tk} (Figure 3.11b). Since some structure are recurring in the azimuthal ($i\phi$) direction for the standard method based on E_{sc}/p_{tk}, and since no similar effect is seen in the E_{sc}/E_{true} map, this has to be imputable to some local scale in the momentum reconstruction. A cross-check is obtained by computing the p_{tk}/p_{true} map, which is directly correlated with the intercalibration map obtained with the $E_{sc}/p_{tk} = 1$ target: whenever p_{tk}/p_{true} is smaller than one, the E_{sc}/p_{tk} becomes greater than one, hence the L3 algorithm produce a coefficient which is smaller than the unity; vice versa for p_{tk}/p_{true} greater than one. The p_{tk}/p_{true} for the ECAL barrel is shown is Figure 3.12.

The same exercise is repeated for the endcaps, and a similar conclusion, i.e. that the tracker momentum effectively creates a pattern in the intercalibration map, can be derived, as it is

[7] This limit is based on a conservative estimation. In fact, due to the limited statistics of the simulated sample, the limit precision of the method is not yet achieved, hence, the difference between σ^{stat} and $\sigma^{stat}_{\text{misc}}$ might be overestimated.

shown in Figure 3.13: once again, a radial pattern, corresponding to a tracker material budget structure along the ϕ angle direction, emerges clearly once the L3 algorithm is fed with the E_{sc}/p_{tk} ratio.



(a) (b)

Figure 3.11: Intercalibration barrel map obtained using W+jets simulated events, using as target $E_{sc}/E_{true} = 1$ (a), or $E_{sc}/p_{tk} = 1$ (b).

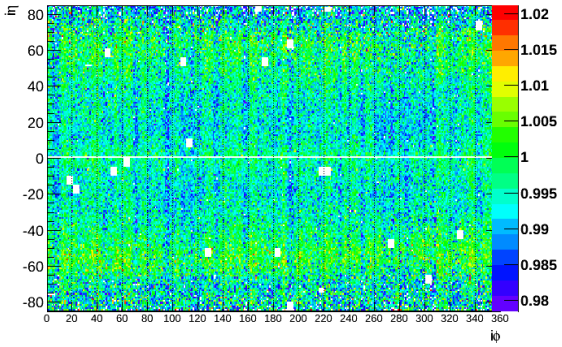


Figure 3.12: The p_{tk}/p_{true} map in the ECAL barrel obtained using W+jets simulated events.

The goal is then to find a method to measure, in real data, the momentum bias, or scale, as a function of the azimuthal angle, and then correct the intercalibration procedure.

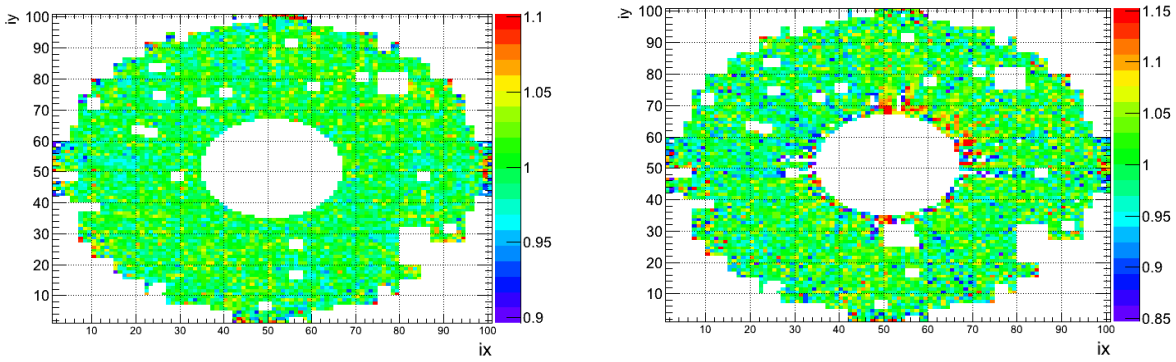
3.5.1 Tracker momentum bias measurement

All the $Z \rightarrow e^+e^-$ candidates are considered, and each of these events is used to determine the momentum scale in two different ϕ_1 and ϕ_2 coordinates, corresponding to the super-cluster seed of each Z electron. To get the correction, the Z invariant mass, as a function of ϕ_1 and ϕ_2, can be computed with the formula

$$m^2_{ee}(\phi_1, \phi_2) = 4 \cdot p^{tk}_1(\phi_1) \cdot E^{sc}_2(\phi_2) \cdot [1 - \cos(\theta)]^2 \quad , \tag{3.12}$$

where $E = p$ due to the relativistic regime where the electrons belong, $E^{sc}_i(\phi_2)$, $p^{tk}_i(\phi_1)$ refer to the electrons energies and momenta, and θ is the angle between the two electrons. One can then compute the distribution of m^2_{ee} as a function of one of the ϕ angles, for eg. ϕ_1, through a simple projection.

The parameters of this distribution are interesting, in particular the mean, which can be com-



(a) (b)

Figure 3.13: Intercalibration endcap minus map obtained using W+jets simulated events, using as input E_{sc}/E_{true} (a), or E_{sc}/p_{tk} (b).

puted as

$$\left\langle m_{ee}^2 \right\rangle (\phi_1) = \alpha_p(\phi_1) \left\langle 4 \cdot \hat{p}_1^{tk} \cdot \alpha_E(\phi_2) \hat{E}_2^{sc} \cdot [1 - \cos(\theta)]^2 \right\rangle \quad , \tag{3.13}$$

where the "hatted" quantities refer to values measured in ideal conditions, and the α_i correspond to the local scales of energy and momentum, which are only ϕ dependent since, as remarked before, all the intercalibration procedure studied here is a ring-by-ring intercalibration. The $\alpha_p(\phi_1)$ correction has been brought out of the $<>$ operator since the direction of ϕ_1 is fixed. The second member of Equation 3.13 can be further simplified, since $\left\langle 4 \cdot \hat{p}_1^{tk} \cdot \alpha_E(\phi_2) \hat{E}_2^{sc} \cdot [1 - \cos(\theta)]^2 \right\rangle$ is an expected value computed on a sample of electrons which are distributed along the ϕ direction with a wide probability distribution centered around $-\phi_1$. This is due to the mild anticorrelation of the electron directions, in the transverse plane, for $Z \to e^+e^-$ events. In addition, the electron super-cluster is spread on the ϕ direction for about 20 degrees, hence widening even more the angular region where the average is performed. The result is that the $\alpha_E(\phi_2)$ local scale is mediated in a wide angular region, therefore, being the ECAL globally uniform in the azimuthal direction, its effect, on the expected values, is negligible.

Due to these considerations, Equation 3.13 becomes

$$\left\langle m_{ee}^2 \right\rangle (\phi_1) = \alpha_p(\phi_1) \left\langle 4 \cdot \hat{p}_1^{tk} \cdot \hat{E}_2^{sc} \cdot [1 - \cos(\theta)]^2 \right\rangle \quad , \tag{3.14}$$

where the expected value in the second member now corresponds to a mass measurement with an ideally calibrated detector, therefore the equation reads as

$$\left\langle m_{ee}^2 \right\rangle (\phi) = \alpha_p(\phi) m_Z^2 \quad . \tag{3.15}$$

This simple example can be used to devise a method, to be used on data and Monte Carlo, to obtain the momentum scale.

First of all, a set of ϕ-bins is chosen, according to the available statistics of data and MC samples. In the barrel, the single crystal granularity is kept, while in the endcaps 120 ϕ-bins are used both in EE+ and EE-.

The m_{ee}^2/m_Z^2 distributions, obtained in each ϕ-bin, are fitted with a reference one, derived from Monte Carlo. Only one template reference is built in EB, integrating on all the electrons belonging to the barrel. On the contrary, two independent references are used for EE+ and EE-, to account for possible asymmetries.
The fit consists of a likelihood maximization, where the scaling parameter $k(\phi)$ acting on m_{ee}^2 of the template distributions is free to float. The $k(\phi)$ best-fit value is used as momentum scale for the corresponding ϕ bin.

Using the procedure previously described, results on both 2011 data and Monte Carlo are obtained for the barrel (Figure 3.14) and the endcaps (Figure 3.15). The figures show that the momentum correction is at the same level of the intercalibration constants magnitude, being at 1.5% level in EB and up to 4.5% in EE, therefore the bias introduced is not negligible. In addition, the same figures show that although a good data vs. MC agreement is observed, thus validating the use of the Monte Carlo for intercalibration studies, some local differences are visible. This is due to the material budget simulation, which is hard to synchronize with the real experimental conditions since the measurement of the material budget is hard to perform and the structural conditions of the tracking system are varying (due for eg. to the presence of a magnetic field, which is turned on and off from time to time, hence causing material stretches and dilations.).

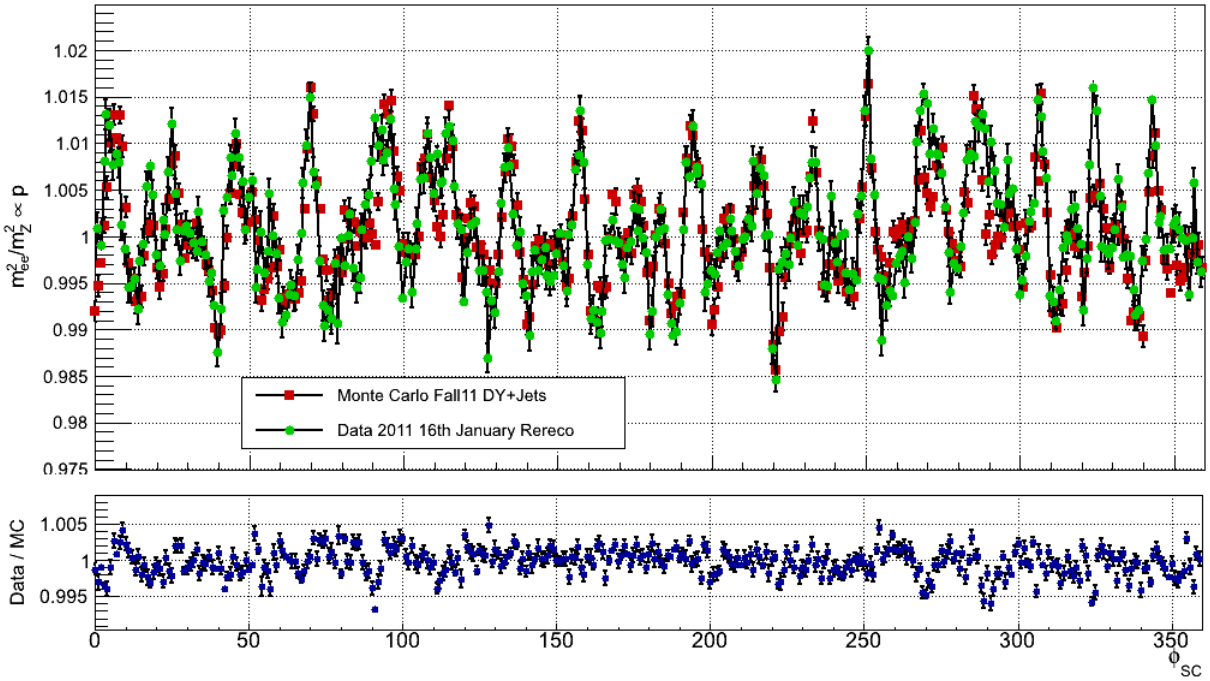
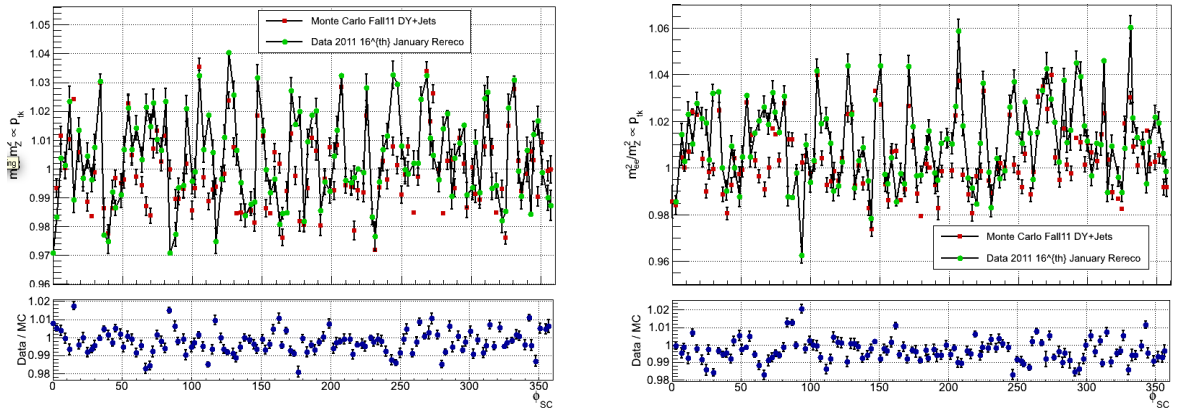


Figure 3.14: Momentum scale as a function of the super-cluster position along ϕ in EB, for 2011 data (green) and MC (red). The bottom frame shows the data/MC ratio.

3.5.2 Intercalibration results with momentum bias correction

Using the azimuthal momentum scale map $\alpha_p(i\phi)$, the intercalibration constants can be corrected according to

$$c_i^{corrected}(i\phi, i\eta) = c_i^{raw}(i\phi, i\eta) \cdot \left[\frac{\alpha_p(i\phi)}{< c_i^{raw} > (i\phi)}\right] \quad , \tag{3.16}$$

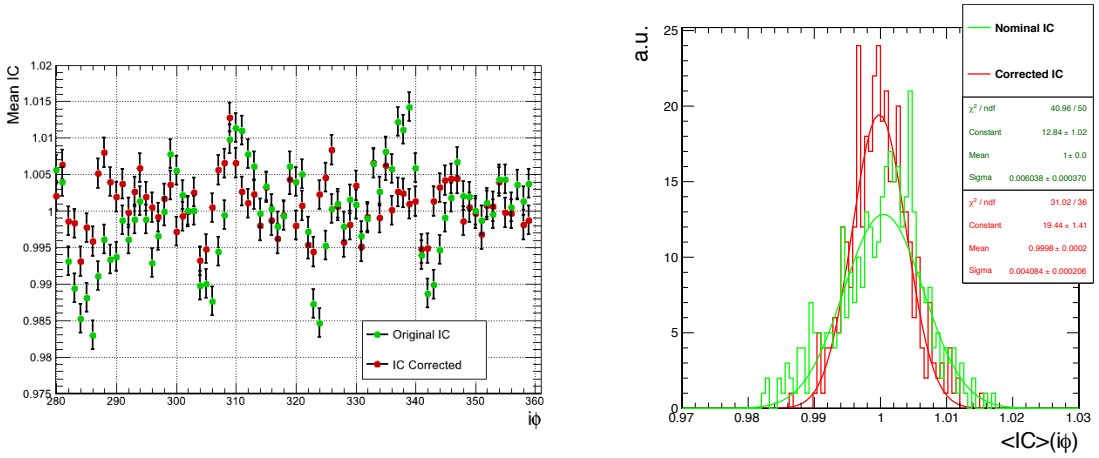


(a) (b)

Figure 3.15: Momentum scale as a function of the super-cluster position along ϕ in EE+ (a) and EE- (b), for 2011 data (green) and MC (red). The bottom frames show the data/MC ratio.

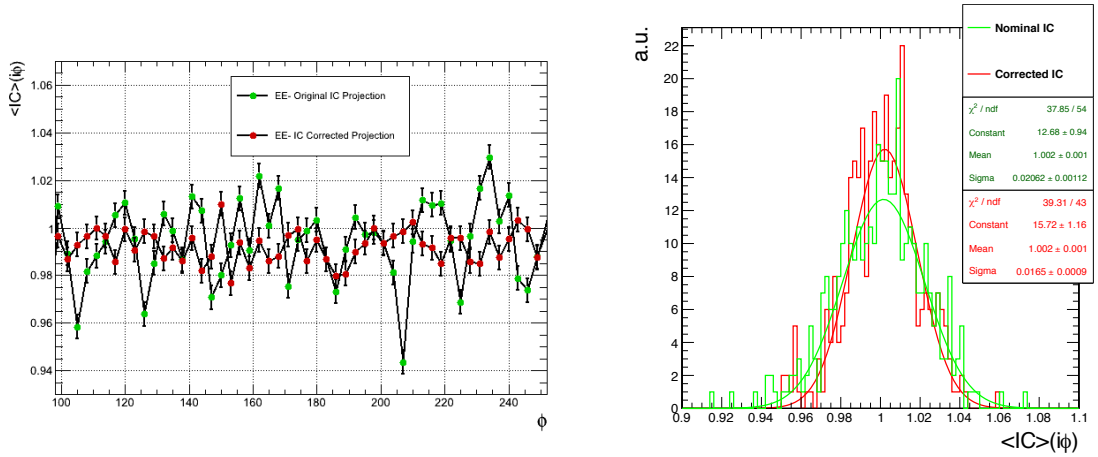
where $< c_i^{raw} > (i\phi)$ is the mean along η of the intercalibration coefficients, corresponding to a given $i\phi$ bin.

To cross check that this procedure is effectively smoothing the modular structure of the intercalibration map due to local momentum scale, the $< c_i > (i\phi)$ distribution before and after the correction, can be studied. In particular, the variance of such distributions, evaluated through a gaussian fit, is taken as a figure of merit to estimate the goodness of the correction. Figure 3.16 shows the effect of the momentum correction in EB, taking as inputs 2011 data. The fit demonstrates that the correction improves the ϕ spread of the intercalibration constants, since the width of the $< c_i > (i\phi)$ goes from 0.6% to 0.4%. The same test is performed in the endcaps, as it is shown in Figures 3.17 and 3.18. Once again the conclusion is that the correction is working as designed, bringing the spread of the $< c_i > (i\phi)$ from 2.1% and 2.4% down to 1.7% and 1.9%, for the EE- and EE+ respectively.



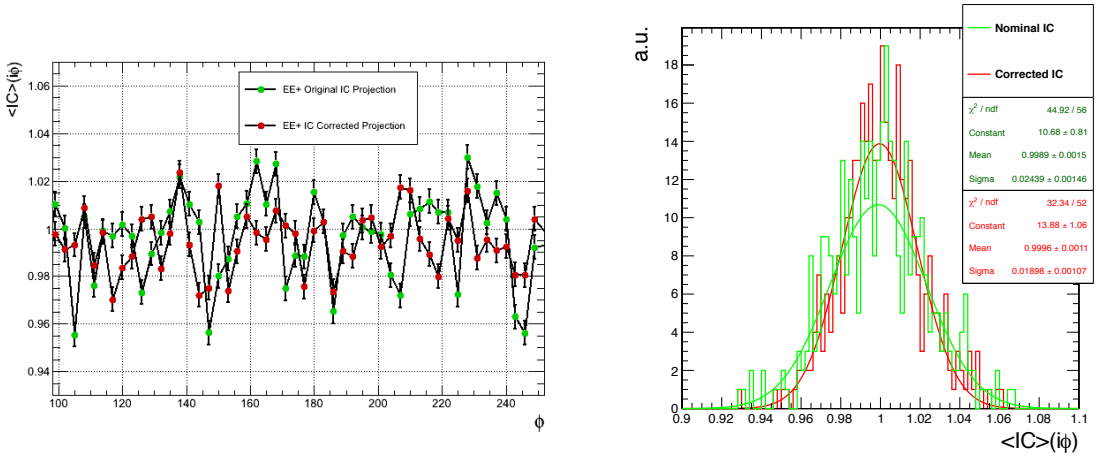
(a) (b)

Figure 3.16: The $< c_i > (i\phi)$ comparison before and after momentum scale correction in EB, for 2011 data. (a) example of comparison plot in the $i\phi$ range $\in$[280 , 360]. (b) gaussian fit of $< c_i > (i\phi)$ distributions, where the spread goes from 0.6% to 0.4%.



(a) (b)

Figure 3.17: The $< c_i > (i\phi)$ comparison before and after momentum scale correction in EE-, for 2011 data. (a) example of comparison plot in the $i\phi$ range $\in[100\,,250]$. (b) gaussian fit of $< c_i > (i\phi)$ distributions, where the spread goes from 2.1% to 1.7%.



(a) (b)

Figure 3.18: The $< c_i > (i\phi)$ comparison before and after momentum scale correction in EE+, for 2011 data. (a) example of comparison plot in the $i\phi$ range $\in[100\,,250]$. (b) gaussian fit of $< c_i > (i\phi)$ distributions, where the spread goes from 2.4% to 1.9%.

3.6 Intercalibration results and consequences

In this section, final intercalibration results, obtained using 2011 pp collision data, are presented.

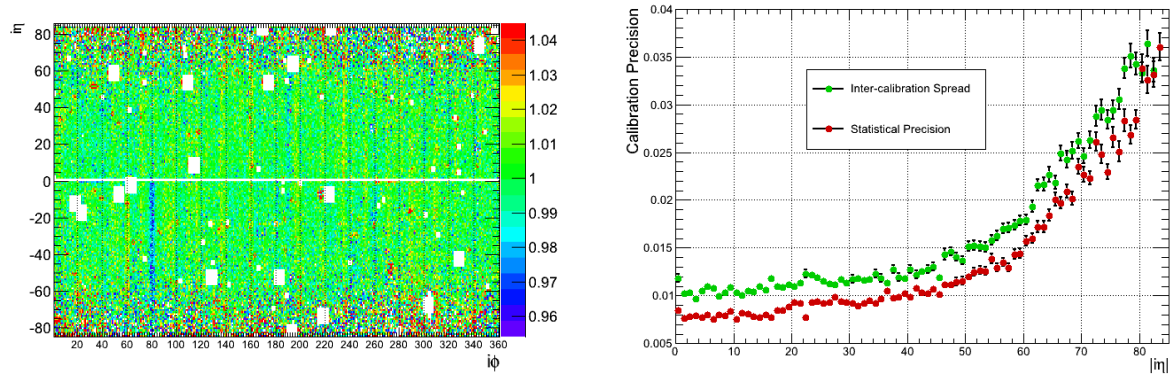
Firstly, the intercalibration maps, for barrel and endcaps, are shown. The c_i spread and statistical precision is analyzed as function of the ring pseudorapidity. The statistical precision is also studied as a function on the integrated luminosity, thus showing that the method, with the 2011 dataset, is still dominated by the limited amount of statistics.

Secondly, the residual term of the intercalibration precision, obtained by taking out the statistical contribution from the c_i spread, is computed for each η-ring and compared with the intercalibration precision obtained with alternative calibration methods.

Finally, the effect of the combined intercalibration is shown on one of the most important Higgs searches, $H \to \gamma\gamma$, through the analysis of the $Z \to e^+e^-$ invariant mass reconstruction.

3.6.1 Final intercalibration results on 2011 data

Figure 3.19 shows the final barrel intercalibration map, together with the c_i spread and statistical precision. Analog results, for the endcaps, are shown in Figure 3.20.



(a) (b)

Figure 3.19: (a) Final set of intercalibration constants, obtained on the 2011 pp collision dataset. The momentum scale corrections are applied. (b) The c_i spread (green) and the statistical precision (red) as a function of the ring pseudorapidity, obtained after folding the two halves of the barrel on the same plane.

In the barrel, the statistical precision varies from about 0.8%, in the central rings, up to order 3% in the forward region. In the endcaps, the statistical uncertainty is higher (about 6%) for the rings nearest to the EB-EE transition region, due to the higher material budget, and then decreases down to about three 3%. Then, in the very forward rings, the statistical precision worsens again, due to another rise of the material budget. The figures, for the barrel and the endcap, are obtained after folding two pseudorapidity regions in one, i.e. with a map that merges the content of the two rings corresponding to the same absolute eta index value.

The statistical precision is studied also as a function of the collected integrated luminosity, by considering an increasing amount of data, i.e. five points corresponding to 1,2,3,4 and 5 fb^{-1}. Due to the limited statistics, the study is meaningful only in the barrel. Figure 3.21

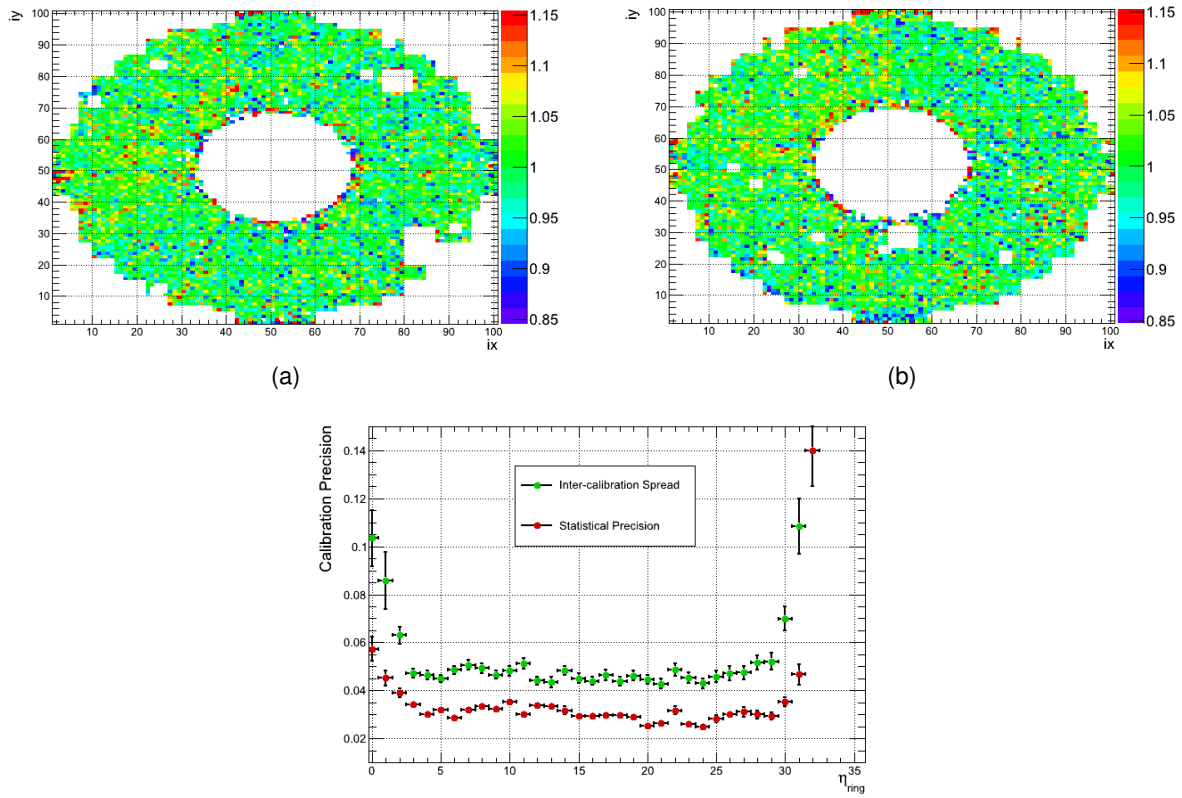


Figure 3.20: Final set of intercalibration constants for the EE+ (a) and EE- (b), obtained on the 2011 pp collision dataset. The momentum scale corrections are applied. (c) The c_i spread (green) and the statistical precision (red) as a function of the ring pseudorapidity, obtained after folding the two endcaps on the same plane.

shows the intercalibration statistical precision for the chosen data points, with the fitted function $a/\sqrt{L} + b$ overlapped. The fit demonstrates that the intercalibration precision of the L3 method is still dominated by the limited availability of electrons. Additionally, the fit shows that, in general, the limit statistical precision, represented by the b parameter, is compatible with zero, as expected. For the most central point, corresponding to the pseudorapidity coverage up to $\eta = 0.35$, the compatibility with zero appears not likely (at the level of three standard deviations). This effect is caused by the correlation of the different points, which are defined as subsets of the same dataset. As a consequence, the points are not fully independent, and the errors on the fit parameter are underestimated.

The study also demonstrates that, since the LHC Run I, corresponding to the entire 2011-2012 datasets, is bringing about 25 fb^{-1} of data, the projected statistical precision of the intercalibration coefficients is around 0.6% in the central barrel, a value close to the ECAL design intercalibration goal of 0.5%.

3.6.2 Intercalibration methods comparisons

The results of the E_{sc}/p_{tk} W/Z electron method can be compared with other ones, presented here [50], based on the decays of π^0/η into two photons and on the azimuthal symmetry of the average energy deposition. Figure 3.22 shows the precision of each method as a function

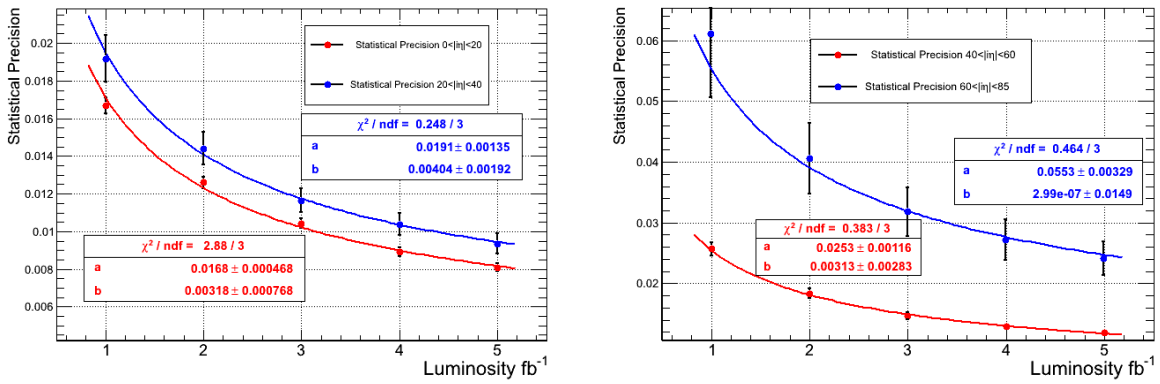


Figure 3.21: Intercalibration statistical precision as a function of the integrated luminosity, for four different regions in the ECAL barrel. The chosen fit function is of the form $a/\sqrt{L} + b$.

of the ECAL ring pseudorapidity, as well as the combined precision. Each point corresponds to the sum of the statistical and systematic uncertainty of each method. The plots provide also the final intercalibration precision that can be achieved by combining the different methods. Although only the 2011 combination includes the W/Z electrons method, the overall precision decreases, since the experimental conditions are in general worse in 2011 with respect to 2010, namely because of the higher radiation levels and number of minimum bias interactions.

Finally, it is interesting to analyze the relation between the W/Z electron residual spread and the 2011 dataset calibration initial conditions. These conditions are essentially given by the 2011 π^0/η method precision, since the ϕ symmetry is significantly less performing but in the forward endcaps region, and since a limited amount of statistics is needed to reach the limit of the π^0/η intercalibration. The comparison shows effectively that the residual term of the W/Z electron method is comparable with the initial 2011 dataset miscalibration conditions. The additional term is due to the W/Z electron systematic uncertainty and to the miscalibration contributions which are not corrected by any of these procedures, for instance due to time-dependent effects like the crystal transparency loss.

3.6.3 Impact of intercalibration in the Higgs search

The search for a light Higgs boson decaying into two photons is not specifically studied in this manuscript, but as was already introduced before, the $H \to \gamma\gamma$ analysis sensitivity is strictly dependent on the mass resolution of a narrow resonance decaying into two photons.

Although the $H \to \gamma\gamma$ search was designed for the 110-140 GeV/c^2 mass range, and a new boson has been seen at 125 GeV/c^2, the Z peak proves extremely useful to understand the performance of the Higgs to di-photon search. The $Z \to e^+e^-$ events can be reconstructed with the ECAL only, i.e. no tracking inputs in the reconstruction, hence, can be used as an approximation of a pure sample of di-photons coming from a narrow width at around 90 GeV/c^2 of mass. Thus, $Z \to e^+e^-$ events represent the best physics benchmark for testing the achieved ECAL resolution in the view of the $H \to \gamma\gamma$ search.

Firstly, the improvement in resolution in the di-photon due to the introduction of intercalibrations and transparency corrections is clearly shown in Figure 3.24, where the $Z \to e^+e^-$ mass peak is reconstructed using only ECAL inputs, and where the lineshape is plotted for an

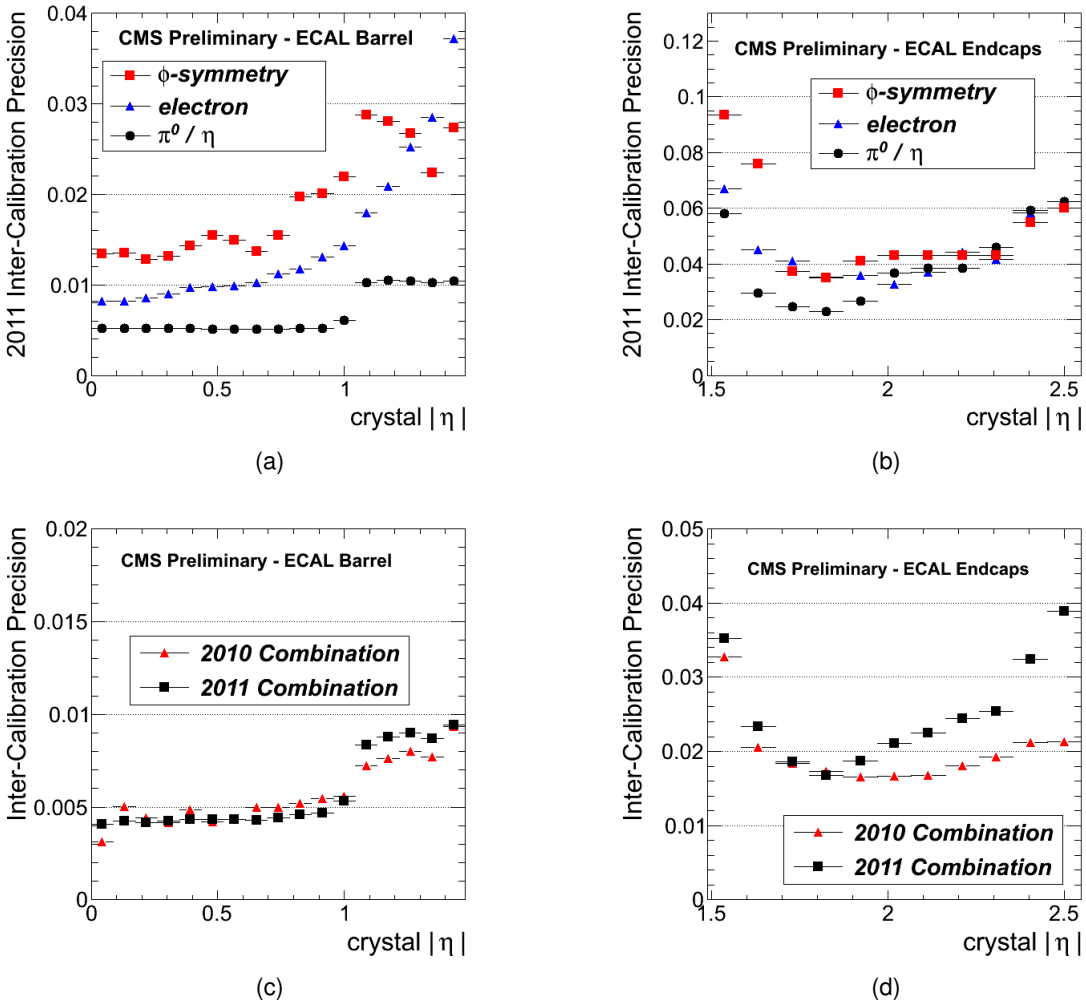
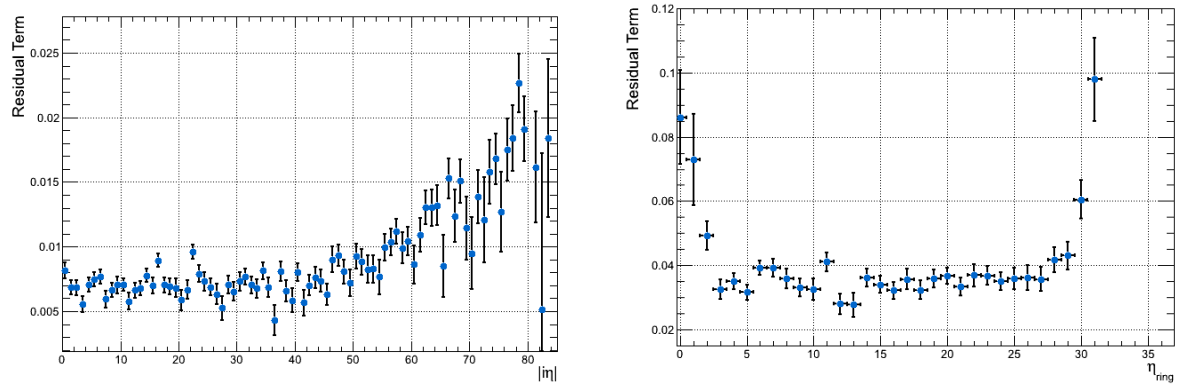


Figure 3.22: Comparison of intercalibration precision, as a function of the ECAL ring pseudorapidity, achieved with three different methods: ϕ symmetry (red), W/Z electrons (blue) and π^0/η (black). The results are obtained using the 2011 datasets and shown separately for the barrel (a) and endcap (b). Each method precision is combined, as it is shown in figures (c) and (d), for barrel and endcap respectively. For more details on the ϕ symmetry and π^0/η intercalibration see [50].

increasing level of corrections. These plots show in particular that, by applying the sole intercalibrations, the di-photon mass resolution improves from about 10% to 5% in the barrel and from about 20% to 10% in the endcaps, practically halving the raw peak width values in the whole ECAL acceptance.

Finally, a similar test is done before and after the improvement in the ECAL reconstruction conditions, namely improved single channel intercalibration, due to the inclusion of the W/Z electron derived constants discussed in this thesis, and new transparency corrections, together with a new, BDT-based approach to derive supercluster $F_{e,\gamma}(E_T, \eta)$ corrections, not specifically addressed in this work. Figure 3.25 show the expected improvement in resolution, for a simulated Higgs signal with mass equal to 120 GeV/c^2, obtained with these new conditions.

This last test is particularly interesting for the purpose of the discussion since it permits to link directly the improvement in the ECAL intercalibration with the $H \to \gamma\gamma$ signal modeling,



(a) (b)

Figure 3.23: W/Z electron intercalibration residual precision, obtained from the spread after subtracting the statistical term. On the left, barrel results, on the right, endcap ones.

which enters directly in the analysis that produced the fundamental physics results already discussed in Section 1.4. Although a precise quantification of the effect of the new intercalibration constant on the final results of the $H \to \gamma\gamma$ analysis is not trivial, some relevant conclusions can be derived by making reasonable assumptions. In the most simple model, the signal significance of a di-photon peak over a given background is given by

$$\mathcal{S}(\Delta) = \frac{\int_{m_H-\Delta}^{m_H+\Delta} f_S(m)\,dm}{\sqrt{\int_{m_H-\Delta}^{m_H+\Delta} f_B(m)\,dm}} = \frac{n_S(\Delta)}{\sqrt{n_B(\Delta)}} \quad , \tag{3.17}$$

where $\mathcal{S}$ is the signal statistical significance, Δ is the optimal counting window centered around the Higgs mass value m_H, and f_S, f_B are the signal and background expected counts per di-photon mass value m. If systematic uncertainties on the background determination are significant, then the previous formula reads as

$$\mathcal{S}(\Delta) = \frac{n_S(\Delta)}{\sqrt{n_B(\Delta) + \beta n_B^2(\Delta)}} \quad , \tag{3.18}$$

where β is the magnitude of the systematic uncertainty measured in unity of the statistical one (a systematic uncertainty equal to the statistical uncertainty means $\beta = 1$). The case of the $H \to \gamma\gamma$ search presents a further simplification, since the signal is narrow with respect to the background falling parameter, hence the background can be considered, under the peak, well described by a linear model. Under these assumptions

$$n_B(\Delta) = \int_{m_H-\Delta}^{m_H+\Delta} f_B(m)\,dm = f_B(m_H) \cdot 2\Delta \quad , \tag{3.19}$$

so that Equation 3.18, reads as

$$\mathcal{S}(\Delta) = \frac{n_S(\Delta)}{\sqrt{2\Delta f_B(m_H) \cdot (1 + \beta \cdot 2\Delta \cdot f_B(m_H))}} \quad . \tag{3.20}$$

Improving the signal resolution means that the same level of signal efficiency can be reached with a smaller mass window. In the hypothesis of a counting window equal to the resolution

($\Delta = \sigma$), $n_S(\sigma_{old}) = n'_S(\sigma_{new})$, where n_S and n'_S represent, respectively, the integral of the old and new di-photon signal peaks in the respective windows σ_{old} and σ_{new} ($\sigma_{old} > \sigma_{new}$). At the same time, in the conservative assumption that the shape of the background is unchanged by the resolution improvement[8], the simple narrowing of the counting window reduces the background yield by a factor $\sigma_{old}/\sigma_{new}$. Then, Equation 3.18 can be used to conclude that the significance improves by a factor

$$\frac{\mathcal{S}'}{\mathcal{S}} = \sqrt{\frac{\sigma_{old}}{\sigma_{new}}} \cdot \sqrt{\frac{1+\beta\cdot 2\sigma_{old}}{1+\beta\cdot 2\sigma_{new}}} \quad , \tag{3.21}$$

where $\mathcal{S}'$ and $\mathcal{S}$ represent the signal significance obtained with the newer and older reconstruction conditions respectively. Since $H \to \gamma\gamma$ is currently limited by statistical effects, β in Equation 3.21 can be considered null, then the signal sensitivity is improving with the square root of the di-photon mass resolution improvement. As it is shown in Figure 3.24, the W/Z intercalibration helps to bring down the di-photon mass resolution from 2.4 GeV/c^2 to 1.82 GeV/c^2, thus leading an improvement in the $H \to \gamma\gamma$ signal sensitivity of 15%.

With the increasing size of the proton-proton collisions dataset, systematics will play a prominent role in the assessment of the background yields, and once again, by letting $\beta \to +\infty$, Equation 3.21 can be used to compute the significance improvement, which in this condition grows linearly with the resolution improvement factor. Hence, as the analysis will approach its intrinsic systematic limitations, the di-photon resolution improvements, and therefore the intercalibration efforts, will play an increasingly important role in the $H \to \gamma\gamma$ performances.

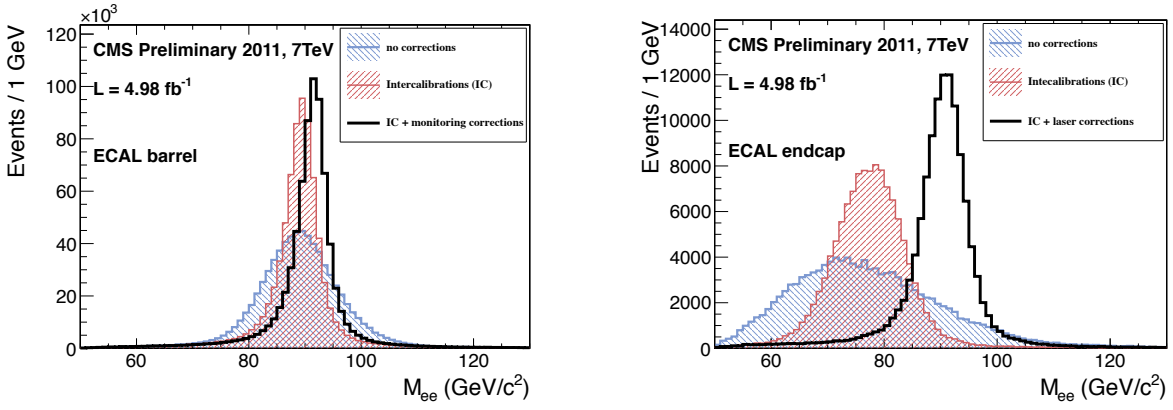


Figure 3.24: The $Z \to e^+e^-$ mass peak reconstructed using only ECAL inputs, which gives a measure on data of the di-photon mass reconstruction performance in a phase space similar to the one typically scanned in the $H \to \gamma\gamma$ search. The lineshapes are produced for different reconstruction conditions, including channel intercalibrations and transparency loss corrections, for barrel (a) and endcaps (b) electrons.

[8]The ECAL resolution improvement moves the background events towards the low energy pole, thus reducing the yield in the high di-photon mass region where the signal has been observed.

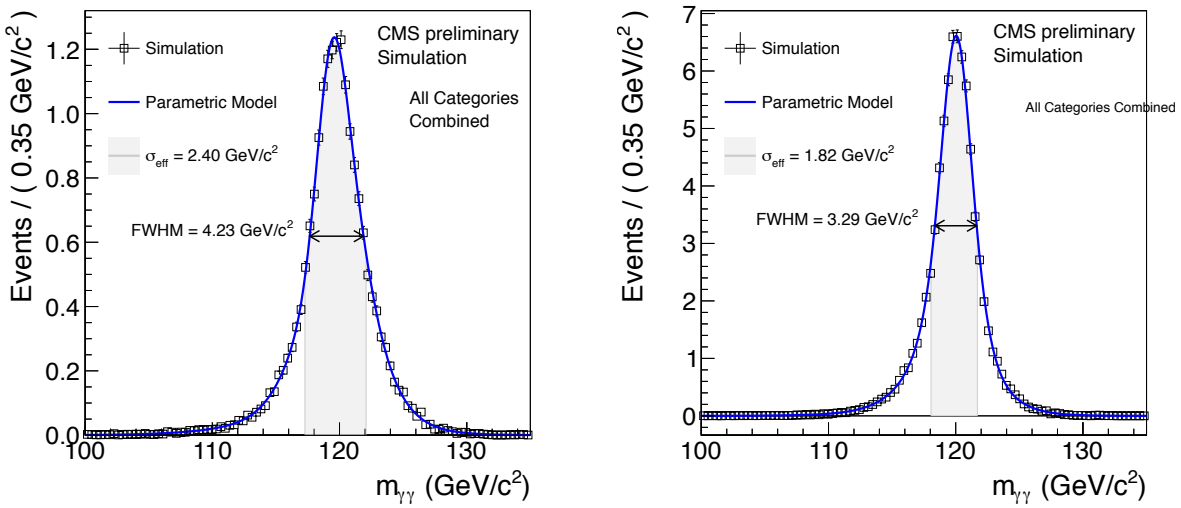


Figure 3.25: Expected Higgs boson mass ($m_H = 120$ GeV/c^2) from the Monte Carlo simulation, before (a) and after (b) the improvement in the ECAL reconstruction conditions, due, among other factors, to the inclusion of the W/Z electron derived constants in the intercalibration combination.

Chapter 4

Search for the the Standard Model Higgs Boson in the $H \to WW \to l\nu_l q\bar{q}$ decay channel

As was described in Chapter 1, the decay of the Higgs boson to two gauge bosons provides an important discovery signature at the LHC. The WW channel, where one W decays leptonically and allows for triggering the event, while the other decays hadronically, has a larger branching fraction than the two-lepton final state and has a reconstructible Higgs mass peak. The main experimental challenge is to control the large W+jets background.

This chapter contains the analysis that sets a limit on the Higgs boson cross-section based on this decay mode, performed on data acquired by CMS at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV during the years 2011-2012. The main part of the measurements and studies presented in this manuscript are referred to the $\sqrt{s} = 8$ TeV dataset and simulation. Whenever a significant difference between the 7 TeV and 8 TeV analysis is present, it will be explicitly detailed. In this chapter, the final state will conventionally be addressed as $l\nu_l q\bar{q}$ or $l\nu_l jj$. For both the naming, some comments are necessary. The first, $l\nu_l q\bar{q}$, should in fact be reported as $l\nu_l q_a\overline{q_b}$, since the quarks, coming from a W decay, cannot have the same flavour. The second, $l\nu_l jj$, implicitly recalls that the final state quarks are not detectable, because of the QCD confinement, and in fact manifest themselves as hadronic jets, denoted with the symbol "j".

Signal and background expectations are discussed in Section 4.1, followed by a summary of the used datasets and Monte Carlo samples (Section 4.2). Section 4.3 contains a description of the physics objects used for this search and details the online and offline selections requirements. Section 4.4 focuses on the leptons and presents the relevant measurements used to cross-check the validity of the lepton Monte Carlo simulation. The data-driven determination of QCD events yield is presented in Section 4.5 and the procedure followed to reconstruct the Higgs invariant mass is shown in Section 4.6. Detailed comparisons of data and MC predictions are then shown in Section 4.7, for the set of preselected events.

After the preselections, the signal-over-background ratio is enhanced by means of a selection on a MVA discriminant, designed to control the background while preserving as much as possible the difference in shape with respect to the signal, as it is detailed in Section 4.8. In the following,

the normalization of the backgrounds in the signal region is fixed through a sideband fit, as it is shown in Section 4.9. The signal extraction mechanism, together with the main background data-driven study, to obtain the W+jets $l\nu_l jj$ mass shapes, is described in Section 4.10. Finally, systematic uncertainties are described in Section 4.11 and Section 4.12 contains the obtained limits on the Standard Model Higgs cross-section.

4.1 Signal and background expectations

Scattering processes at high energy hadron colliders can be classified as either hard or soft ones. Quantum Chromodynamics is the underlying theory for all such processes, but the approach and level of understanding is very different for the two cases. For soft processes, which constitute by far the majority of collision events at the LHC, the rates and properties are dominated by non-perturbative QCD effects, which are not well modeled. Luckily, since such events are characterized by small momentum transfer between the interacting partons ($\hat{p}_T \simeq 500\,\mathrm{MeV}/c$), they do not yield interesting physics for most of the LHC searches, as is the case of this thesis, and are simply discarded. For hard processes, e.g. Higgs boson or high p_T jet production, the rates and event properties can be predicted with good precision using perturbation theory. Nevertheless, for most of the hard processes, soft (or minimum-bias) interactions are occurring along with the hard interactions and their effects must be understood for comparisons to be made to perturbative predictions.

The most predictive scheme is given by the QCD factorization theorem [52]. The core concept of factorization is that the cross-section can be computed through a product of probability functions, namely parton distribution functions (PDFs), describing the probability to extract a quark or gluon from the protons in the initial state, a perturbative cross-section for the hard scattering, and a probabilistic description of the final state by a parton shower Monte Carlo. Figure 4.1 is a pictorial representation of the same concept.

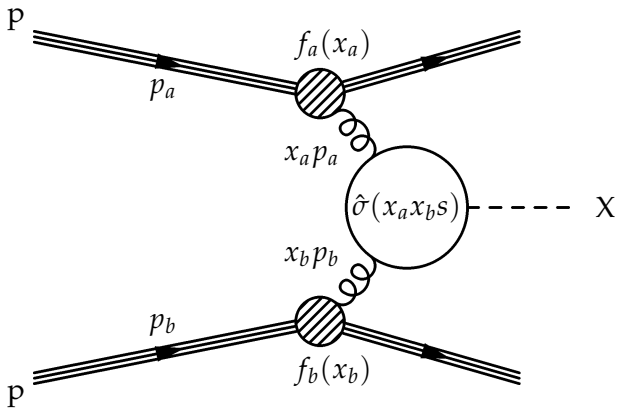


Figure 4.1: Diagrammatic structure of a generic hard scattering process. The partons, extracted from the colliding p^+p^- pair, carry a momentum fraction with respect to the proton energy described by a PDF. The scattering of the partons is then analyzed with a perturbative computation, and hence the kinematic proprieties of the final state object X are predicted.

Mathematically, the generic $\sigma_{pp\to X}$ LHC cross-section can be written as

$$\begin{aligned}\sigma_{pp\to X} &= \sum_{a,b \in \text{ partons}} \int_0^1 dx_a dx_b \; f_a(x_a) f_b(x_b) \; \hat{\sigma}_{ab\to X}(\hat{s}) \\ &= \sum_{a,b \in \text{ partons}} \int_0^1 dx_a dx_b \; f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \times \left[\hat{\sigma}_0(\hat{s}) + \alpha_S(\mu_R^2)\hat{\sigma}_1(\hat{s}, \mu_F^2) + \dots\right]_{ab\to X} \quad ,\end{aligned} \tag{4.1}$$

where f_i $(i = a, b)$ are the standard PDFs for partons $a, b = \{g, u, \bar{u}, d, ...\}$ carrying fractions x_a, x_b of the proton longitudinal momentum, and $\sigma_{pp\to X}$ is the partonic cross-section to scatter i and j calculated in fixed-order perturbation theory. The μ_F coefficient is the factorization scale, which can be thought of as the scale that separates the long and short-distance physics, and μ_R is the renormalization scale for the QCD running coupling. Formally, the cross-section calculated to all orders in perturbation theory is invariant under changes in these parameters, but in practice it is not possible to dispose of a complete set of higher order corrections. It is then necessary to make a specific choice for the two scales in order to make cross-section predictions. The usual prescription consists in choosing a central value μ_0 for both scales equal to some sensible energy scale in the process (e.g. m_H for Higgs boson production, m_Z for Drell-Yan events, etc.). A range of variation of the renormalization and factorization scales of $\mu_0/2 \leq \mu_R, \mu_F \leq 2 \cdot \mu_0$ is used to determine the uncertainty in the cross-section calculation due to missing higher-order QCD radiative corrections.

Another important source of uncertainty in the cross-section calculation is due to the limited knowledge of the PDFs. The recommendations of the PDF4LHC working group for the use of PDF and of PDF uncertainties at the LHC are followed [53]. In summary, for uncertainties of next-to-leading order cross-sections, a 68% C.L. envelope provided by the central values and PDF$\pm\alpha_S$ uncertainties from the MSTW08 [54], CTEQ6.6 [55] and NNPDF2.0 [56] sets is built. As a central value, the midpoint of this envelope is used. At next-to-next-to leading order, MSTW08 prediction is used as central value. As an uncertainty, the same relative uncertainty on this NNLO prediction is taken as found using the NLO uncertainty prescription given above. In both cases, the uncertainty envelope is obtained by summing in quadrature the errors from the α_S variations and the ones related to the PDF fit. The latters are computed by considering the effect on the PDFs due to the limited knowledge of the PDFs fit minima, through the variation of the PDFs eigenvectors around the best fit values in a 68% CL interval.

4.1.1 The SM Higgs signal

As already presented in Section 1.3, the primary production mechanism of the Higgs boson at the LHC is gluon-gluon fusion, with a small but measurable contribution from vector boson fusion and more rare contributions from other processes. The $H \to W^+W^-$ branching ratio has to be taken into account when computing the effective cross-section for the channel under study, and the list of obtained values for different Higgs mass hypotheses is reported in Table 4.1. For an heavy Higgs, i.e. for masses above 500 GeV/c^2, the effect of interference between the gluon fusion Higgs production and the non-resonant $gg \to WW$ production is taken into account using the method proposed here [57], which is briefly summarized in Section 4.2.3.

The $l\nu_l q\bar{q}$ final state is searched for, experimentally, as a minimal signature given by an isolated lepton, missing transverse energy and two jets. Events with an extra jet, coming both from a gluon-gluon fusion process with an hard QCD emission or from a VBF produced event, are kept. The $l\nu_l q\bar{q}$ experimental signature introduces a bias in the search:

1. the $W \to l\nu_l$ decay provides the lepton used to trigger the events. CMS single lepton triggers rely on threshold above 24 and 27 GeV for muons and electrons respectively. This implies that events with a soft virtual leptonic-decaying W boson are not efficiently triggered;

2. the $W \to q\bar{q}$ decay is experimentally reconstructed as a pair of jets. As presented in Section 2.3.5, CMS jet resolution is drastically decreasing for a reconstructed p_T below 30 GeV. Moreover, at low p_T, the contamination from jets coming from minimum-bias interactions is significant. Practically, soft jets are not experimentally reliable objects. Thus, only real hadronic-decaying Ws are reconstructible;

3. since an heavy Higgs boson creates heavily boosted W pairs, the boosted $W \to q\bar{q}$ might not give rise to two distinct reconstructed jets if the two are geometrically separated by less than the jet radius. Quantitatively, assuming massless jets, one finds that ΔR_{jj} is about $2m_W/p_W^T$ and, reminding the thumb rule $p_W^T \lessapprox m_H/2$ and that the standard CMS jet radius is 0.5, the search must be limited for an Higgs mass hypothesis below 600 GeV/c^2.

Points 1 and 2 imply that both the bosons coming from the Higgs decay must be real to efficiently reconstruct the signal events, while the last point limits the search below 600 GeV/c^2. As a result, this analysis is limited in the Higgs mass range $170 \to 600$ GeV/c^2. In addition, due to the unique nature of the τ leptons[1], the search is limited to the case where the leptonic W decays into an electron or a muon.

4.1.2 The background processes

Several SM processes have a final state with one lepton, two jets and missing transverse energy, therefore produce events which can be reconstructed and interpreted as signal events. Additionally, also other processes which do not produce all the objects characterizing the signal final state can, through experimental effects, give rise to events which are reconstructed as signal candidates. All these processes are then to be considered as background for this search, and the most important ones are listed in order of relevance:

- **$W(\to l\nu)$+jets**, i.e. W bosons produced in association with quark or gluon radiation (Figure 4.2). The leptonic decay of the W and the presence of jets due to the showering of the final state colored particles give rise to a signature matching the signal one. Because of its large cross-section, this is by far the most important background to the analysis.
- **Drell-Yan $Z/\gamma^*(\to l^+l^-)$+jets**. Z/γ^* boson production in association with quarks (Figure 4.2) or gluons may mimic the $l\nu_l q\bar{q}$ signature when one lepton, coming from

[1] The lifetime of τ leptons is short enough that they decay before reaching the detector elements. Thus, this lepton species is reconstructed through the identification of its decay products. This property makes the $\tau\nu_\tau q\bar{q}$ events much different, experimentally, from the $e\nu_e q\bar{q}$ or $\mu\nu_\mu q\bar{q}$ ones.

m_{H} (GeV/c^2)	$\sigma_H^{7\,\mathrm{TeV}} \times \mathcal{B}(\mathrm{H} \to \mathrm{l}\nu_{\mathrm{l}}\,\mathrm{q}\bar{\mathrm{q}})$ (pb)	$\sigma_H^{8\,\mathrm{TeV}} \times \mathcal{B}(\mathrm{H} \to \mathrm{l}\nu_{\mathrm{l}}\,\mathrm{q}\bar{\mathrm{q}})$ (pb)
170	$2.52^{+0.35}_{-0.37}$	$3.28^{+0.45}_{-0.47}$
180	$2.17^{+0.30}_{-0.32}$	$2.80^{+0.38}_{-0.40}$
190	$1.60^{+0.22}_{-0.23}$	$2.09^{+0.28}_{-0.29}$
200	$1.34^{+0.18}_{-0.20}$	$1.78^{+0.24}_{-0.26}$
250	$0.84^{+0.11}_{-0.12}$	$1.14^{+0.15}_{-0.16}$
300	$0.60^{+0.08}_{-0.09}$	$0.83^{+0.11}_{-0.12}$
350	$0.53^{+0.07}_{-0.08}$	$0.74^{+0.10}_{-0.10}$
400	$0.38^{+0.06}_{-0.05}$	$0.54^{+0.08}_{-0.07}$
450	$0.25^{+0.04}_{-0.04}$	$0.36^{+0.05}_{-0.05}$
500	$0.16^{+0.02}_{-0.02}$	$0.23^{+0.03}_{-0.03}$
550	$0.10^{+0.02}_{-0.01}$	$0.15^{+0.02}_{-0.02}$
600	$0.07^{+0.01}_{-0.01}$	$0.10^{+0.02}_{-0.01}$

Table 4.1: Cross-section values and their uncertainties for Higgs boson production at $\sqrt{s} = 7$ TeV (left) and $\sqrt{s} = 8$ TeV (right), multiplied by the branching ratio of the $\mathrm{l}\nu_{\mathrm{l}}\,\mathrm{q}\bar{\mathrm{q}}$ final state (with $\mathrm{l} = \mathrm{e},\mu$), as a function of the Higgs mass. Values taken from [21].

a Z decay, is undetected because of acceptance or inefficiency effects, hence giving rise to missing transverse energy, and the hadronic activity produces high p_{T} jets.

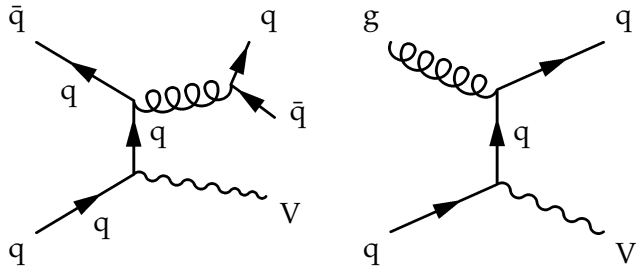


Figure 4.2: Leading order Feynman diagrams for the Vector boson+jets production in pp collisions.

- $\mathrm{t\bar{t}}$ **+jets**. Top quark pairs are produced at LHC via the gluon fusion process $gg \to \mathrm{t\bar{t}}$ or via QCD quark annihilation $qq \to \mathrm{t\bar{t}}$ (Figure 4.3). The semi-leptonic sample, where one W decays hadronically and the other leptonically, gives rise to a major background. It can be handled by requests on the number of jets and by means of jet tagging techniques. The less relevant fully leptonic decay can be suppressed by imposing a second lepton veto on the final state.
- **VV**. Diboson pair production (Figure 4.4), albeit being characterized by a lower cross-section with respect to other backgrounds, exhibits the same signature of signal and therefore constitutes an irreducible background, with the only distinctive feature given by the non-resonating m_{WW} shape. Three different production chan-

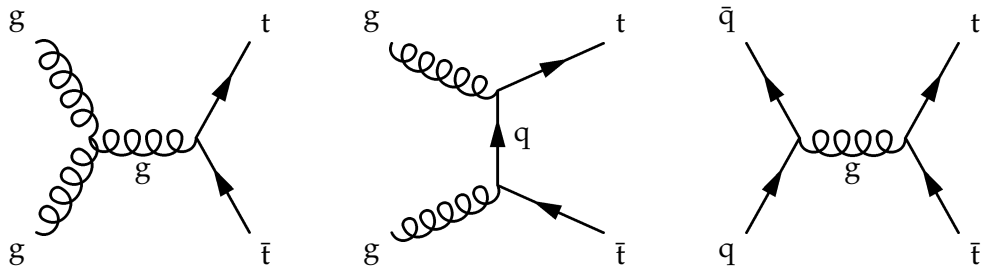


Figure 4.3: Feynman diagrams for the $t\bar{t}$ pair production at the LHC.

nels are considered:

1. WW, for the semileptonic decay mode;
2. WZ, where the Z decays hadronically and the W leptonically, or where the W decays hadronically and one of the two Z leptons is not reconstructed;
3. ZZ, where one Z decays hadronically and the other leptonically, with one lepton failing to be reconstructed.

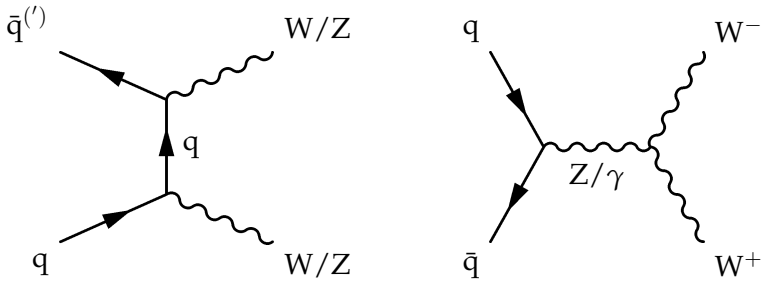


Figure 4.4: Diboson pair production diagrams in pp collisions.

- **Single top**. Single top production (Figure 4.5), although having an order of magnitude smaller cross-section with respect to $t\bar{t}$ pairs, may give rise to $W \to l\nu_l$ decays associated with hadronic activity, which can resemble $W \to q\bar{q}$ decays. This process is created in three different ways:

  1. *t*-channel, where the top is produced after a quark-quark interaction with the exchange of a virtual W;
  2. *s*-channel, characterized by an annihilation of a pair of quarks, through a weak vertex and a creation of a pair of top and bottom quarks;
  3. *t*W-channel, where a gluon-bottom pair in the initial state exchanges a bottom quark and produces a top quark associated with a W boson.

- **QCD**, i.e. multi-jet events. These events, in principle topologically different from Higgs decays, are produced with an extremely high cross-section at hadron colliders. Some of the jets in the final state could fail the detector acceptance region, giving rise to an unbalance in the transverse energy. In addition, lepton identification algorithms have a non-zero probability to classify a jet as a lepton, especially for electrons. Hence, QCD events constitute a detectable source of background, in particular for the $e\nu_e q\bar{q}$ final state.

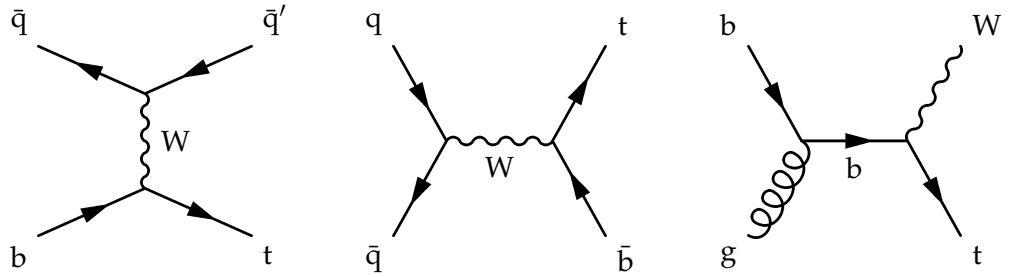


Figure 4.5: Single top production in pp collisions. From left to right: *t*-channel, *s*-channel and *t*W-channel production modes.

The cross-section for the backgrounds, multiplied by the branching ratio when meaningful, are reported in Table 4.2.

process	$\sigma^{7\,\text{TeV}}$ (pb)	$\sigma^{8\,\text{TeV}}$ (pb)
W($\to$ lν)+jets	31300 ± 1600	36300 ± 1800
Z($\to l^+l^-$)+jets	3050 ± 130	3350 ± 150
WW	47.0 ± 1.5	57.1 ± 1.9
WZ	18.2 ± 0.7	32.2 ± 1.1
ZZ	7.10 ± 0.15	8.3 ± 0.17
$t\bar{t}$ +jets	163 ± 14	225.2 ± 16
t (*t*-channel)	41.9 ± 1.8	56.4 ± 2.1
$\bar{t}$ (*t*-channel)	22.6 ± 1.0	30.7 ± 1.1
t (*t*W-channel)	7.9 ± 0.6	11.1 ± 0.7
$\bar{t}$ (*t*W-channel)	7.9 ± 0.6	11.1 ± 0.7
t (*s*-channel)	3.19 ± 0.14	3.79 ± 0.15
$\bar{t}$ (*s*-channel)	1.44 ± 0.07	1.76 ± 0.08
QCD (e-enriched)	~ 6740000	~ 7530753
QCD (μ-enriched)	~ 84700	~ 134680

Table 4.2: Cross-section values for the backgrounds, multiplied by the branching ratio when meaningful (with l = e,μ,τ in this table), at $\sqrt{s} = 7$ TeV (left) and $\sqrt{s} = 8$ TeV (right).

4.2 Datasets and Monte Carlo samples

4.2.1 Datasets used for the analysis

The data samples used in this analysis were recorded by the CMS experiment in 2011, at $\sqrt{s} = 7$ TeV, and in 2012, at $\sqrt{s} = 8$ TeV. Only certified data are considered, which means that a good functioning of all CMS sub-detectors is required. The total analyzed statistics corresponds to an

integrated luminosity of about 5 and 12 fb^{-1} for 2011 and 2012 respectively. LHC conditions, which varied greatly during 2011 run, when the instantaneous luminosity $\mathcal{L}$ changed by an order of magnitude, were almost stable in 2012, with a value of $\mathcal{L}$ around $7 \cdot 10^{33}\,cm^{-2}\,s^{-1}$, as it is shown in Figure 4.6.

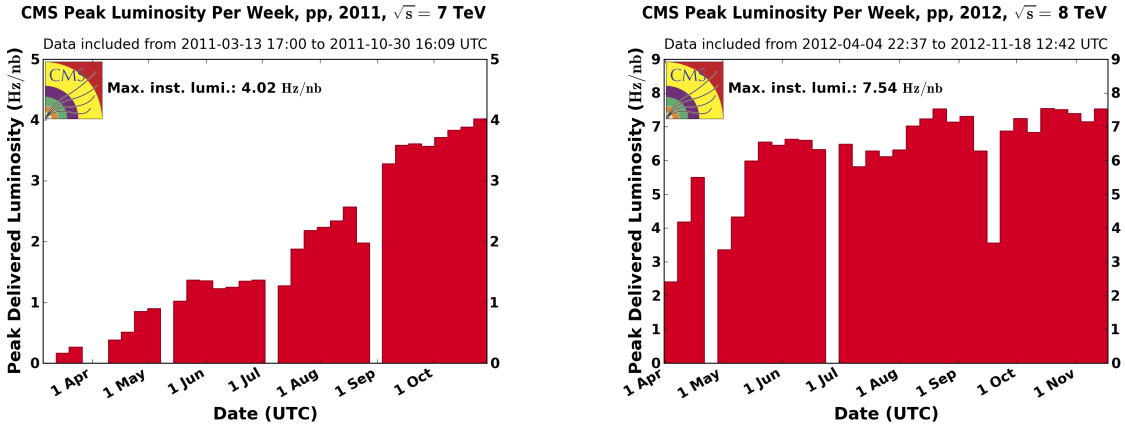


Figure 4.6: Instantaneous luminosity versus time for 2011 (left) and 2012 (right) pp collision runs. The plotted values correspond to weekly maxima.

Single Electron and *Single Muon* datasets have been used for the analysis, for both the 2011 and 2012. In terms of content, all the events containing at least one lepton are part of the used datasets, with some minor differences between 2011 and 2012, related to the trigger and reconstruction conditions, which will be highlighted in the following.

4.2.2 Monte Carlo samples

As introduced in Section 4.1, factorization theorem and MC techniques provide a viable way to simulate collisions in the complex environment of an hadron machine like the LHC. Different MC event generators can be used to get the final configuration of generated particles and, similarly, to get jets in case of final state quarks and gluons. On top of that, an additional layer is introduced, to simulate the interaction between particles and detector elements, through a dedicated CMS Software [43]. As a result, the simulated events are translated into elements of detector response and can be fed to the very same reconstruction algorithms used for data.

More in detail, the POWHEG-BOX NLO generator [58, 59, 60, 61] has been used to produce signal events, and the showering has been performed with PYTHIA [62]. For this analysis, samples with Higgs mass hypotheses ranging from 170 to 600 GeV/c^2 have been used (Tables 4.3 and 4.4). Only gluon fusion production mode samples have been considered, due to computing time limitations. To consider the contribution of the VBF production mode, the samples are associated to a cross-section which is the sum of the gluon fusion and VBF channels[2].

For the main backgrounds, i.e. W+jets, Z+jets, and $t\bar{t}$+jets processes, the leading-order MADGRAPH generator [64] is used, which is capable to generate vector bosons/$t\bar{t}$ pairs with up to four partons at the matrix element level; MADGRAPH is also designed to match the final

[2]Past studies [63] proved that the selection efficiency on VBF events is higher than gluon fusion ones, hence the total signal yield is in fact underestimated. This assumption, which is simply conservative for the limit setting, should be revised in case an excess is found and a characterization of its cross-section should be performed.

state partons with the parton showers. For single-top events POWHEG-BOX is chosen, while for diboson events the LO PYTHIA generator is employed.

In general, for all samples, the showering and fragmentation of the hard events is performed with version 6.442 of PYTHIA, using the Z2* tune and the CTEQ6L PDF set [65]. The response to final-state particles of each sub-detector part is then simulated by means of a detailed description of the whole CMS detector within the GEANT4 toolkit [66].

The complete list of used samples is reported in Tables 4.3 and 4.4.

4.2.3 Signal sample weighting

Signal samples are reweighted in order to match the correct Higgs line shape, according to the complex pole reweighting scheme [67, 68]. The weights are cross-section neutral and computed on an event by event basis for the signal MC samples only.

Interference between the signal and the SM $gg \to$ WW production process is an increasingly important effect: the first order effect is a modification of the cross-section, up to $O(10\%)$ for an Higgs boson with mass equal to 600 GeV/c^2; more importantly, the m_{WW} spectrum, which constitutes the most important observable in the view of this search, is varied, with an enhancement for $m_{\mathrm{WW}} \lesssim m_{\mathrm{H}}$ and a reduction for $m_{\mathrm{WW}} \gtrsim m_{\mathrm{H}}$. A practical solution to include interference effects in the signal modeling is to reweight the signal samples, by means of a scheme proposed in [57], which addresses this issue in the case of a gluon fusion produced Higgs boson.

In summary, one starts by noting that the total WW cross-section can be written as

$$\sigma_{gg\to \mathrm{WW}} = \sigma_{gg\to \mathrm{WW}}(S) + \sigma_{gg\to \mathrm{WW}}(I) + \sigma_{gg\to \mathrm{WW}}(B) \quad , \tag{4.2}$$

where S, I and B stand for signal ($gg \to \mathrm{H} \to$ WW), interference and background ($gg \to$ WW). All the terms in last equation are known at LO, but only the signal has been computed with higher order corrections (NNLO). Considering the observable $D = \mathrm{d}\sigma/\mathrm{d}m_{\mathrm{WW}}$, one can consider two different options:

- **additive**, where one computes

$$D^{\mathrm{NNLO}}_{\mathrm{eff}} = D^{\mathrm{NNLO}}(S) + D^{\mathrm{LO}}(I) + D^{\mathrm{LO}}(B) \quad ; \tag{4.3}$$

- **multiplicative**, where one computes

$$D^{\mathrm{NNLO}}_{\mathrm{eff}} = K_D[D^{\mathrm{LO}}(S) + D^{\mathrm{LO}}(I)] + D^{\mathrm{LO}}(B) \quad , \quad K_D = \frac{D^{\mathrm{NNLO}}(S)}{D^{\mathrm{LO}}(S)} \quad , \tag{4.4}$$

  where K_D is the differential K-factor for the m_{WW} distribution.

The differential K-factor for the WW-invariant mass distribution is always greater than one in the region of interest for this search. In fact, it is a slowly increasing function of m_{WW}, going from about 2 at $m_{\mathrm{WW}} = 210$ GeV/c^2 to about 2.5 at $m_{\mathrm{WW}} = 1$ TeV/c^2. Unfortunately, these two options suffer from an obvious problem: they are spoiling the unitarity cancellation between signal and background for $m_{\mathrm{WW}} \to \infty$. Therefore, neither of the two options alone can be used for too high values of the WW-invariant mass, which is the case of interest for this work.

A viable alternative is offered by considering a third option:

- **intermediate**, where one computes

$$D_{\text{eff}}^{\text{NNLO}} = K_D D^{\text{LO}}(S) + \sqrt{K_D^{\text{real}}} D^{\text{LO}}(I) + D^{\text{LO}}(B) \quad , \quad K_D = K_D^{\text{virt}} + K_D^{\text{real}} \quad , \tag{4.5}$$

  where K_D^{virt} and K_D^{real} are the virtual and real contributions to the K-factor K_D.

This third option simulates the inclusion of K-factors at the amplitudes level. The other two options represent the extreme cases where only the signal (4.3) or both signal and interference (4.4) are reweighted, and are used to determine the theoretical uncertainty band of the interference prediction. In fact, the difference between the intermediate option and the median of the band is found to be always small but far away from the Higgs mass value, where, in any case, the signal yield is negligible. To summarize, the intermediate option gives the central value, while the band between the multiplicative and the additive options gives the uncertainty.

According to this correction, the signal, which already contains NNLO corrections, must be reweighted by the the factor

$$\frac{\sqrt{K_D^{\text{real}}}}{K_D} \cdot D^{\text{LO}}(I) \quad . \tag{4.6}$$

Since this weight is m_{WW} dependent, one must also verify that other relevant observables are correctly shaped by the reweighting procedure. To perform this check a conservative test is devised. A very high mass signal sample, with $m_{\text{H}} = 700$ GeV/c^2, is chosen, since for this mass the interference effects are maximal ($O(20\%)$ on the cross-section), and then it is reweighted with the procedure explained above. In parallel, the NLO MCFM Monte Carlo generator [69] is used to generate the same Higgs mass point with interference effects included at LO. The relevant distributions obtained with the MCFM sample are compared with the ones obtained from the standard sample, reweighted to account for the interference, and a good compatibility is found, as it shown for example in Figure 4.7. The reweighting procedure is thus validated and implemented for all the Higgs signal samples.

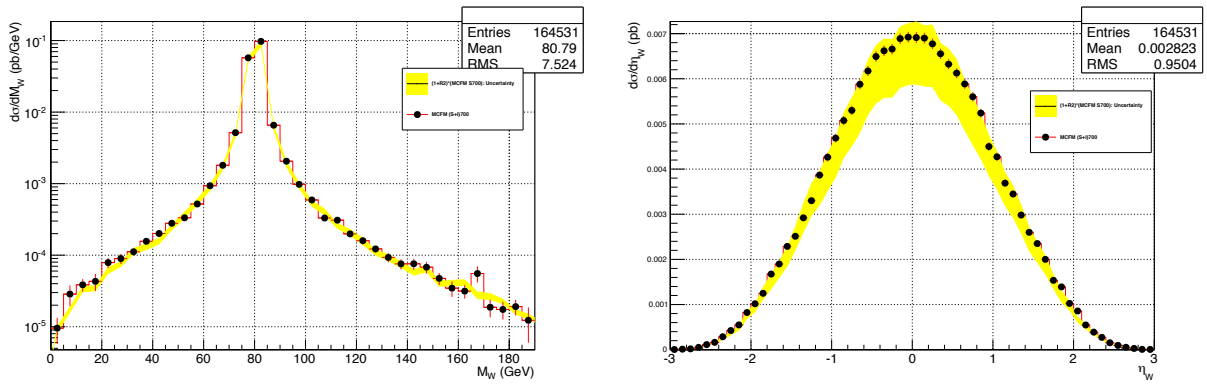
4.2.4 Pile-up sample weighting

In actual-data taking conditions, there is a significant probability that more than two protons interacts, and additionally out-of-time pile-up, i.e. spurious interactions from a contiguous bunch crossing, anticipated or delayed of 50 ns, may happen. Therefore, in the MC samples, a generation procedure for pile-up (PU) events is added on top of the hard scattering.

In data, the average expected number of such PU interactions, in a given bunch crossing i, is expressed by the following formula:

$$(N_{\text{PU}})_i = \frac{\mathcal{L}_i \cdot \sigma_{\text{min. bias}}}{\nu_{\text{orbit}}} \quad , \tag{4.7}$$

where $\mathcal{L}_i$ is the instantaneous luminosity of that bunch crossing, $\sigma_{\text{min. bias}}$ is the cross-section of minimum-bias interactions and ν_{orbit} is the LHC orbit frequency (11246 Hz). An example of the measured distribution of $N_{\text{PU}}^{\text{true}}$ for the whole 2012 dataset is shown in Figure 4.8. The minimum-bias cross-section $\sigma_{\text{min. bias}}$, used above, is measured using 7 TeV $Z \to \mu^+\mu^-$ events,



(a) (b)

Figure 4.7: Interference reweighting closure test. The plots contain the distributions predicted with the interference included at generator level (black points) and the ones obtained starting from the standard sample reweighted with the scheme summarized in Section 4.2.3 (yellow band). In particular, for the latters, the 68% CL band is shown, where the corresponding nominal value lies approximately in the band mid-point. The comparison is shown for the mass (a) and the pseudorapidity (b) of the final state W bosons (two entries per event).

as a result of a best-fit analysis comparing the distribution of the number of reconstructed vertices in data and Monte Carlo, and 8 TeV minimum bias events, through the pixel cluster counting method [70]. The two methods lead to the two measurements $\sigma^{7\,\mathrm{TeV}}_{\text{min. bias}} = (68.0 \pm 5.0\%)$ mb and $\sigma^{8\,\mathrm{TeV}}_{\text{min. bias}} = (69.4 \pm 4.4\%)$ mb, where in both cases the uncertainty is dominated by systematic effects.

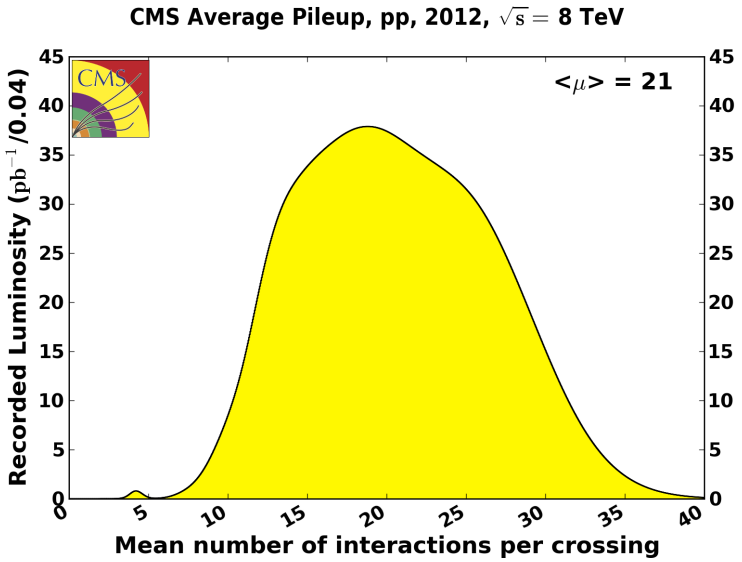


Figure 4.8: Mean number of interactions per bunch crossing for the 8 TeV LHC run.

The simulated samples are generated with a number of spurious-interactions which follows a poissonian distribution centered around N_{PU}, and in order to represent the distribution of number of pp interactions per bunch crossing (pile-up) as measured in the data, the simulated samples are reweighted. To cross-check the effectiveness of the reweighting procedure, the distribution of reconstructed vertexes can be looked at, since it is a variable which strongly

depends on the number of pile-up interactions. Figure 4.9 shows the data vs MC comparison before and after the reweighting procedure.

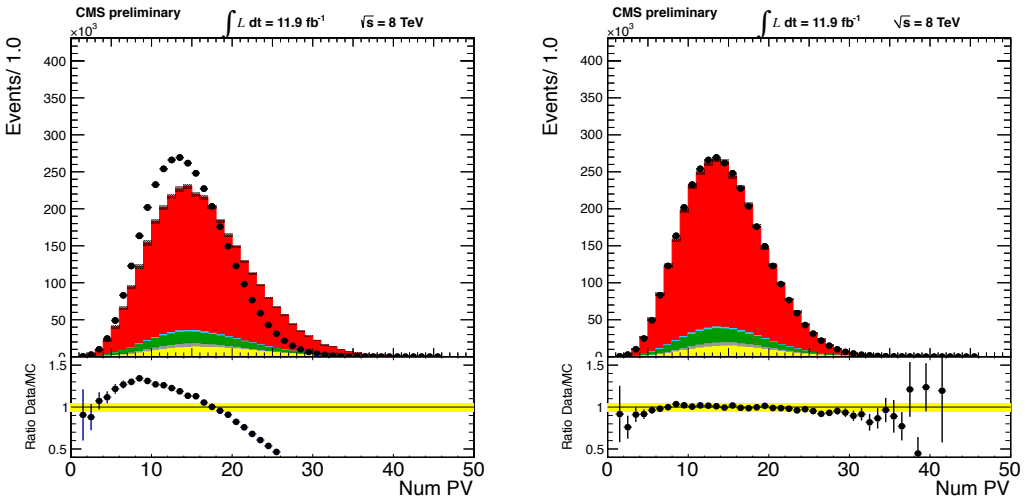


Figure 4.9: Data vs. MC comparison of the number of reconstructed primary vertices (a) before and (b) after the re-weighting for the pile-up is applied. The plots are obtained by looking at 2012 data, for the $e\nu_e\, q\bar{q}$ final state.

4.2.5 Trigger strategy

All the signal events, both in the $e\nu_e\, q\bar{q}$ and $\mu\nu_\mu\, q\bar{q}$ channels, contain a clean lepton coming from a W decay. The trigger for the analysis is designed to store events which contain at least one of such leptons. To reduce the trigger rate whilst keeping a good efficiency on signal, some selections are required on the online reconstructed leptons, based on the $W \to l\nu_l$ topology. The leptons are required to pass a set of quality cuts, to be isolated[3] and to have an high transverse momentum. For most of the data taking, both in 2011 and 2012, muons/electrons p_T cut is set at 24/27 GeV/c.

No trigger emulation request is applied on MC samples, but trigger lepton efficiencies, measured on data by means of a tag-and-probe technique (Section 4.4), are applied as weights on simulated events.

The trigger strategy was not as simple during 2011, for the $e\nu_e\, q\bar{q}$ final state, since the CMS triggering of $W \to e\nu_e$ was previously based on a combination of requirements on the electron and on the transverse mass of the online reconstructed electron-E_T^{miss} system. In this case, the corresponding trigger efficiency, to be applied on MC samples, is given by the product of the lepton and E_T^{miss} trigger efficiencies.

[3] For more details on the online isolation variable see Section 2.3.4

/GluGluToHToWWToLNuQQ_M-*_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B-v1/AODSIM
/VBF_HToWWToLNuQQ_M-*_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B-v1/AODSIM
/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/WW_TuneZ2_7TeV_pythia6_tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/WZ_TuneZ2_7TeV_pythia6_tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/TTJets_TuneZ2_7TeV-madgraph-tauola/Fall11-PU_S6_START42_V14B-v2/AODSIM
/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/T_TuneZ2_s-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/T_TuneZ2_t-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/T_TuneZ2_tW-channel-DS_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/Tbar_TuneZ2_s-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/Tbar_TuneZ2_t-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM
/Tbar_TuneZ2_tW-channel-DS_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM

Table 4.3: Summary of Monte Carlo samples used in the 7 TeV analysis. Signal samples are reported in the top box.

/LQ-ggh*_SIM/*-SQWaT_PAT_53X_ggH*_v1-*/USER
/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/WW_TuneZ2star_8TeV_pythia6_tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/WZ_TuneZ2star_8TeV_pythia6_tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/T_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/T_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM

Table 4.4: Summary of Monte Carlo samples used in the 8 TeV analysis. Signal samples are reported in the top box.

4.3 The event selection

The final state of the Higgs decay is characterized by a charged lepton, large missing transverse energy and two hadronic jets. In this section, the criteria applied on single objects are described and then requirements at event-level are detailed.

4.3.1 Object reconstruction

The analysis relies on the standard reconstruction algorithms produced by the CMS community. The particle-flow reconstruction procedure (Section 2.3.1) is used to coherently define the collection of jets, leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ considered in the event selection.

Electrons

Electrons are required to pass electron identification (ID) cuts according to a multi-variate selection, designed to fully exploit the variables, and their correlations, used for the standard cut based identification in the real-fake electron discrimination (see Chapter 3). They are also required to be isolated. The chosen isolation variable is particle-flow based, with an isolation cone of 0.3. This quantity has already been described in Section 2.3.4, and shorty, it is given mathematically by the relation

$$I_{\mathrm{rel}}^{\mathrm{PF}} = \frac{I_{\mathrm{CH}} + \max(0, I_{\mathrm{NH}} + I_{\mathrm{PHOTON}} - (\mathrm{A_E} \cdot \rho))}{E_{\mathrm{T}}} \quad ,$$

where I_{CH}, I_{NH} and I_{PHOTON} are the charged hadron, neutral hadron and photon isolation variables and $\mathrm{A_E} \cdot \rho$ represents the term which accounts for isolation contributions coming from pile-up interactions.

The ID and isolation cuts used are shown in Table 4.5 and are tuned to give the same bin-by-bin signal efficiency with respect to the working point (WP) used for the 2011 analysis.

Additionally, some requirements are added:

- electron $E_{\mathrm{T}} > 30$ GeV, to be in a phase space where the single electron trigger is fully efficient;
- pseudorapidity restricted to the ECAL coverage region $|\eta| < 2.5$, with an exclusion range due to the ECAL barrel-endcap transition region, defined by $1.4442 < |\eta_{\mathrm{sc}}| < 1.566$, where η_{sc} is the pseudorapidity of the ECAL supercluster;
- in order to make sure that the selected electron does not come from a minimum bias interaction, the absolute value of the impact parameter calculated with respect to the primary vertex (PV) is required to be $d_0(\mathrm{PV}) < 0.02$ cm. In addition, the z coordinate of the PV has to lie within a distance of less than 0.1 cm from the electron production vertex, identified with the point of minimum distance between the electron track and the beam axis;
- in order to reject events in which the electron candidate actually originates from a conversion of a photon into an e^+e^- pair, the vertex fit probability of fully reconstructed PV-displaced conversions must be compatible with zero. Additionally, the number of missed inner tracker layers of the electron track must be exactly zero (i.e. there are no missed layers before the first hit of the electron track from the beam

line).

Lepton η	$\lvert\eta\rvert < 0.8$	$0.8 < \lvert\eta\rvert < 1.479$	$1.479 < \lvert\eta\rvert < 2.5$
ID MVA cut value (tight lepton)	0.913	0.964	0.899
Isolation cut value (tight lepton)	0.105	0.178	0.150
ID MVA cut value (loose lepton)	0.877	0.811	0.707
Isolation cut value (loose lepton)	0.426	0.481	0.390

Table 4.5: Cut values for the electron identification MVA output and for the isolation which are tuned to give the same efficiency as the 2011 cut-based working point WP80, used for the tight electron selection, and WP90, used in the loose electron selection.

The same requirements are applied on the 2011 dataset, with some minor changes in the definitions. The identification is cut based (see Section 3.2.1) and the isolation is sub-detector based (see Section 2.3.4). Additionally, the cut on the electron E_{T} is set to 35 GeV to cope with more stringent trigger requirements.

Muons

Alike electrons, reconstructed muons have to fulfill quality and isolation requirements. The particle-flow based relative isolation for muons slightly differs from the electron definitions, since it makes use of the DeltaBeta correction (see Section 2.3.4) and exploits an isolation cone with amplitude 0.4, and it reads as

$$I_{\mathrm{rel}}^{\mathrm{PF}} = \frac{I_{\mathrm{CH}} + \max(0, I_{\mathrm{NH}} + I_{\mathrm{PHOTON}} - 0.5 \cdot I_{\mathrm{CH}}^{\mathrm{PU}})}{p_{\mathrm{T}}} \quad .$$

The muon is required to have $I_{\mathrm{rel}}^{\mathrm{PF}} < 0.12$ in order to be considered isolated.

The quality cuts criteria for muons read as:

- muon reconstructed both in the tracker and in the muon chambers;
- number of pixel hits of the tracker track ≥ 1;
- number of muon system hits of the global track ≥ 1;
- normalized χ^2 of the global track < 10.0;
- muon $p_{\mathrm{T}} > 30$ GeV, to be in the region where the HLT is fully efficient;
- pseudorapidity restricted to the muon system coverage region, i.e. $|\eta| < 2.1$;
- number of tracker layers with hits from the muon track has to be greater than five;
- in order to make sure that the selected muon does not come from a minimum bias interaction, the absolute value of the impact parameter calculated with respect to the primary vertex is required to be $d_0(\mathrm{PV}) < 0.02\,\mathrm{cm}$. In addition, the z coordinate of the PV has to lie within a distance of less than 0.5 cm from the muon production vertex, identified with the point of minimum distance between the muon track and the beam axis.

The same requirements are applied on the 2011 dataset, with the only change in the definition of isolation variables, which is sub-detector based.

Loose Electron and Muon

For the purposes of rejecting events with more than one lepton, a loose electron and muon are defined.

Those electrons which have $p_T > 20$ GeV/c, $|\eta| < 2.5$, and which satisfy electron I_{rel}^{PF} and MVA ID cuts, are considered as *loose electrons*. The cut values for the electron ID and isolation used in the analysis can be found in Table 4.5. The same working points are used for the 7 TeV analysis, but with cut based identification and sub-detector based isolation.

Global muons which have $p_T > 10$ GeV/c, $|\eta| < 2.5$, and $I_{rel}^{PF} < 0.2$, are defined as *loose muons*. Loose muons for the 2011 selection retain the same definition with a sub-detector based isolation.

Jets

The goal is to identify jets coming from a W decay or an hard QCD emission. First of all, some loose quality requirements (98%–99% efficiency on real jets) are set on jets to discard badly reconstructed and calorimeter noise jets:

- jets must be made of more than one single constituent, and at least one of them must be a charged particle;
- the jet energy fraction carried by charged hadrons must be greater than zero, while neutral hadrons must carry at most 99% of the total jet energy;
- the jet energy fraction carried by charged and neutral electromagnetic objects is required to be smaller than 0.99.

Additionally, pile-up jets are identified and discarded by means of a dedicated multivariate analysis. The pile-up identification is founded on three types of proprieties of the jets:

- inside the tracker acceptance, the trajectories of tracks associated to the jets can be used to establish the compatibility of the jet with the primary interaction vertex;
- the topology of the jet shape can be used in order to disentangle jets arising from the overlap of multiple interactions from true hard jets;
- the object multiplicity can be used as an additional handle.

The strategy is to combine the above information to retain more than 98% of jets from PV, while cutting away almost 80% of pile-up jets.

Finally, only jets with measured (corrected) p_T greater than 30 GeV/c and within the tracker acceptance, $|\eta| < 2.4$, are retained.

4.3.2 Event-Level criteria

Events are selected in two main steps. First of all, an event pre-selection is applied to go in a region where perturbative calculations hold, i.e. where simulation is reliable, the trigger is fully efficient and the dominant topology is ($W \rightarrow l\nu_l$)+jets events. After that, an MVA approach, different for each Higgs mass hypothesis, is followed, in order to drastically reduce the background. This step will be described in detail in Section 4.8.

Pre-selection

The pre-selection aims at requiring the final-state objects typical of the searched signature and applying minimum criteria to sit as much as possible on trigger efficiency plateaus. At the same time, the cuts are designed to enhance the signal over the large background rate. The list of requirements is detailed below:

- the event should have a good primary vertex. This means selecting the primary vertex with the highest sum of p_T^2 of the tracks associated with it and requiring it to have a number of degrees of freedom (NDF) ≥ 4, where NDF is proportional to the number of tracks used for the construction of the PV. In addition, the PV must lie in the central detector region of $|z| \leq 24$ cm and $\rho \leq 2$ cm around the nominal interaction point;
- the event is required to have exactly one tight lepton candidate (electron or muon). A veto is applied on the events containing one or more additional loose leptons. This requirements reduces the Drell-Yan contamination and the overlap with other searches based on the $H \to W^+W^-$ fully leptonic decays;
- to reduce multi-jet and Drell-Yan backgrounds, the events are required to have large missing transverse energy from the undetected neutrino, that is $E_T^{miss} > 25(30)$ GeV for electrons(muons). Additionally, the leptonic W transverse mass has to be $m_T > 30$ GeV/c^2, where the transverse mass is defined as

  $$m_T = \sqrt{p_T^{lepton} \cdot E_T^{miss} \cdot (1 - \cos\Delta\phi)} \quad ,$$

  where $\Delta\phi$ is the angle between the lepton momentum $\vec{p}_T^{lepton}$ and the $\vec{E}_T^{miss}$ in the transverse plane. In the 2011 selection, for $e\nu_e q\bar{q}$ events, a higher threshold at 50 GeV/c^2 is applied to cope with the more stringent HLT requirements;
- events are required to have exactly two or three jets. The two jets with highest p_T are considered as W decay product candidates. This criterion, instead of choosing the jet pair whose invariant mass is closest to m_W, is used in order to minimize the bias introduced for background events due to jet combinatorics. According to simulation, in the case of 2-jet events, the correct jet combination rate, chosen with this criterion, varies from 68% for $m_H = 200$ GeV/c^2 to 88% for $m_H = 600$ GeV/c^2. For 3-jet events, the corresponding rates are 26% and 84%, respectively.

4.4 Lepton reconstruction, selection and trigger efficiencies

Since the lepton reconstruction, selection and trigger efficiencies can be slightly different between data and simulation, correction factors have to be applied to the MC to account for these differences. The efficiencies are calculated using a tag-and-probe technique exploiting Z bosons decays to a pair of electrons or muons. One of the leptons is used as tag and has to pass a tight selection, while the second one is used as probe if the tag-probe pair is compatible to the Z boson mass. A detailed review of the tag-and-probe technique is presented here [71].

The total lepton efficiency can be factorized into three components:

$$\epsilon_{\text{total}} = \epsilon_{\text{Reco}} \cdot \epsilon_{\text{Id}} \cdot \epsilon_{\text{HLT}} \quad , \tag{4.8}$$

where ϵ_{Reco}, ϵ_{Id}, ϵ_{HLT}, refer, respectively, to the lepton reconstruction, selection and trigger efficiencies. Since no HLT emulation is applied on Monte Carlo simulation, the HLT efficiency measured on data is applied directly in the analysis of MC samples, while the other two efficiency components are calculated both for data and MC, so that a data/MC scale factor has to be applied. In the following, the value and the meaning of each component will be discussed, with a focus on results from 2012 data.

4.4.1 Electron efficiencies

In the electron case, the reconstruction efficiency ϵ_{Reco} characterizes the transition from a super-cluster in the electromagnetic calorimeter to a reconstructed particle-flow electron. The probability of a reconstructed electron to pass the offline selection, consisting of the isolation and identification criteria listed in Section 4.3, is given by the identification efficiency ϵ_{Id}. Finally, the selected electron has a certain probability to fire the high level trigger and the efficiency to fulfill the HLT requirements is parametrized as ϵ_{HLT}.

In general, since the efficiency depends both on p_{T} and η of the electron, the measurement is binned in p_{T} as (30, 35, 40, 45, 50, 200) GeV/c and in η as (-2.5, -1.5, 0.0, 1.5, 2.5) of the probe electron. The resulting efficiencies and scale factors are summarized in Table 4.6 and shown in Figure 4.10.

$p_{\text{T,min}}$ [GeV/c]	$p_{\text{T,max}}$ [GeV/c]	η_{min}	η_{max}	$\epsilon_{\text{Reco,data}}/\epsilon_{\text{Reco,mc}}$	$\epsilon_{\text{ID,data}}/\epsilon_{\text{ID,mc}}$	$\epsilon_{\text{HLT,data}}$
30	35	-2.5	-1.5	1.000 ± 0.002	0.986 ± 0.006	0.641 ± 0.003
30	35	-1.5	0	1.005 ± 0.003	0.986 ± 0.003	0.875 ± 0.002
30	35	0	1.5	1.001 ± 0.001	0.981 ± 0.003	0.874 ± 0.001
30	35	1.5	2.5	1.001 ± 0.002	0.995 ± 0.006	0.652 ± 0.003
35	40	-2.5	-1.5	0.998 ± 0.002	0.983 ± 0.004	0.687 ± 0.002
35	40	-1.5	0	1.000 ± 0.001	0.985 ± 0.002	0.896 ± 0.001
35	40	0	1.5	1.001 ± 0.001	0.987 ± 0.002	0.892 ± 0.001
35	40	1.5	2.5	1.002 ± 0.001	0.994 ± 0.004	0.693 ± 0.002
40	45	-2.5	-1.5	1.003 ± 0.001	0.978 ± 0.002	0.710 ± 0.002
40	45	-1.5	0	1.000 ± 0.001	0.989 ± 0.002	0.909 ± 0.001
40	45	0	1.5	1.001 ± 0.002	0.988 ± 0.030	0.906 ± 0.001
40	45	1.5	2.5	1.001 ± 0.001	0.984 ± 0.002	0.721 ± 0.002
45	50	-2.5	-1.5	0.999 ± 0.002	0.996 ± 0.004	0.725 ± 0.002
45	50	-1.5	0	1.000 ± 0.001	0.996 ± 0.002	0.917 ± 0.001
45	50	0	1.5	1.000 ± 0.001	0.994 ± 0.002	0.910 ± 0.001
45	50	1.5	2.5	1.000 ± 0.001	0.989 ± 0.004	0.734 ± 0.002
50	200	-2.5	-1.5	1.003 ± 0.002	0.984 ± 0.004	0.735 ± 0.003
50	200	-2.5	-1.5	1.003 ± 0.002	0.984 ± 0.004	0.735 ± 0.003
50	200	-1.5	0	1.000 ± 0.001	1.001 ± 0.002	0.924 ± 0.001
50	200	0	1.5	1.001 ± 0.001	0.993 ± 0.002	0.920 ± 0.001
50	200	1.5	2.5	1.000 ± 0.002	0.995 ± 0.004	0.746 ± 0.003

Table 4.6: Electron efficiency and data/MC scale factors for super-cluster to reconstructed electrons (ϵ_{Reco}), reconstructed to selected electrons (ϵ_{ID}) and selected to HLT electrons (ϵ_{HLT}). The uncertainties are statistical only.

4.4.2 Muon efficiencies

In the muon case, the reconstruction efficiency ϵ_{Reco} describes the ability to reconstruct a particle-flow muon starting from a muon track, and it can be assumed to be one [72]. The identification

efficiency ϵ_{Id} gives an estimate for a reconstructed muon to pass the offline selection criteria, listed in Section 4.3. The trigger efficiency ϵ_{HLT} is the fraction of selected muons fulfilling the HLT requirements.

The efficiency measurement is binned both in p_{T} and η of the probe muon covering the relevant intervals (25, 30, 35, 40, 45, 50, 200) GeV/c in p_{T} and (-2.1, -1.6, -1.1, -0.6, 0.0, 0.6, 1.1, 1.6, 2.1) in η. The resulting selection and trigger efficiencies and scale factors are summarized in Table 4.7 and Figure 4.10.

$p_{\mathrm{T,min}}$ [GeV/c]	$p_{\mathrm{T,max}}$ [GeV/c]	η_{min}	η_{max}	$\epsilon_{\mathrm{ID,data}}/\epsilon_{\mathrm{ID,mc}}$	$\epsilon_{\mathrm{HLT,data}}$
25	30	-2.1	-1.6	0.991 ± 0.003	0.746 ± 0.003
25	30	-1.6	-1.1	0.991 ± 0.003	0.825 ± 0.003
25	30	-1.1	-0.6	0.988 ± 0.003	0.894 ± 0.002
25	30	-0.6	0	0.986 ± 0.002	0.923 ± 0.002
25	30	0	0.6	0.986 ± 0.002	0.926 ± 0.002
25	30	0.6	1.1	0.992 ± 0.003	0.895 ± 0.002
25	30	1.1	1.6	0.993 ± 0.003	0.797 ± 0.003
25	30	1.6	2.1	0.995 ± 0.003	0.811 ± 0.003
30	35	-2.1	-1.6	0.989 ± 0.002	0.765 ± 0.002
30	35	-1.6	-1.1	0.991 ± 0.002	0.830 ± 0.002
30	35	-1.1	-0.6	0.986 ± 0.002	0.901 ± 0.002
30	35	-0.6	0	0.984 ± 0.002	0.932 ± 0.001
30	35	0	0.6	0.985 ± 0.002	0.936 ± 0.001
30	35	0.6	1.1	0.989 ± 0.002	0.902 ± 0.002
30	35	1.1	1.6	0.989 ± 0.002	0.804 ± 0.002
30	35	1.6	2.1	0.995 ± 0.002	0.830 ± 0.002
35	40	-2.1	-1.6	0.989 ± 0.002	0.776 ± 0.002
35	40	-1.6	-1.1	0.992 ± 0.002	0.832 ± 0.002
35	40	-1.1	-0.6	0.988 ± 0.002	0.906 ± 0.001
35	40	-0.6	0	0.986 ± 0.001	0.937 ± 0.001
35	40	0	0.6	0.987 ± 0.001	0.941 ± 0.001
35	40	0.6	1.1	0.989 ± 0.002	0.906 ± 0.001
35	40	1.1	1.6	0.989 ± 0.002	0.806 ± 0.002
35	40	1.6	2.1	0.996 ± 0.002	0.835 ± 0.002
40	45	-2.1	-1.6	0.993 ± 0.002	0.783 ± 0.002
40	45	-1.6	-1.1	0.989 ± 0.001	0.836 ± 0.001
40	45	-1.1	-0.6	0.990 ± 0.001	0.908 ± 0.001
40	45	-0.6	0	0.987 ± 0.001	0.942 ± 0.001
40	45	0	0.6	0.988 ± 0.001	0.946 ± 0.004
40	45	0.6	1.1	0.990 ± 0.001	0.908 ± 0.001
40	45	1.1	1.6	0.992 ± 0.001	0.810 ± 0.002
40	45	1.6	2.1	0.996 ± 0.001	0.844 ± 0.002
45	50	-2.1	-1.6	0.993 ± 0.002	0.791 ± 0.002
45	50	-1.6	-1.1	0.990 ± 0.002	0.840 ± 0.039
45	50	-1.1	-0.6	0.988 ± 0.001	0.912 ± 0.001
45	50	-0.6	0	0.988 ± 0.001	0.943 ± 0.001
45	50	0	0.6	0.989 ± 0.015	0.949 ± 0.001
45	50	0.6	1.1	0.991 ± 0.012	0.911 ± 0.001
45	50	1.1	1.6	0.992 ± 0.001	0.814 ± 0.002
45	50	1.6	2.1	0.995 ± 0.002	0.853 ± 0.034
50	200	-2.1	-1.6	0.991 ± 0.002	0.795 ± 0.003
50	200	-1.6	-1.1	0.989 ± 0.002	0.841 ± 0.002
50	200	-1.1	-0.6	0.988 ± 0.002	0.913 ± 0.002
50	200	-0.6	0	0.987 ± 0.002	0.946 ± 0.001
50	200	0	0.6	0.989 ± 0.002	0.948 ± 0.001
50	200	0.6	1.1	0.992 ± 0.002	0.913 ± 0.002
50	200	1.1	1.6	0.993 ± 0.002	0.816 ± 0.002
50	200	1.6	2.1	0.995 ± 0.002	0.856 ± 0.002

Table 4.7: Muon efficiency and data/MC scale factors for reconstructed to selected muons (ϵ_{ID}) and selected to HLT muons (ϵ_{HLT}). The uncertainties are statistical only.

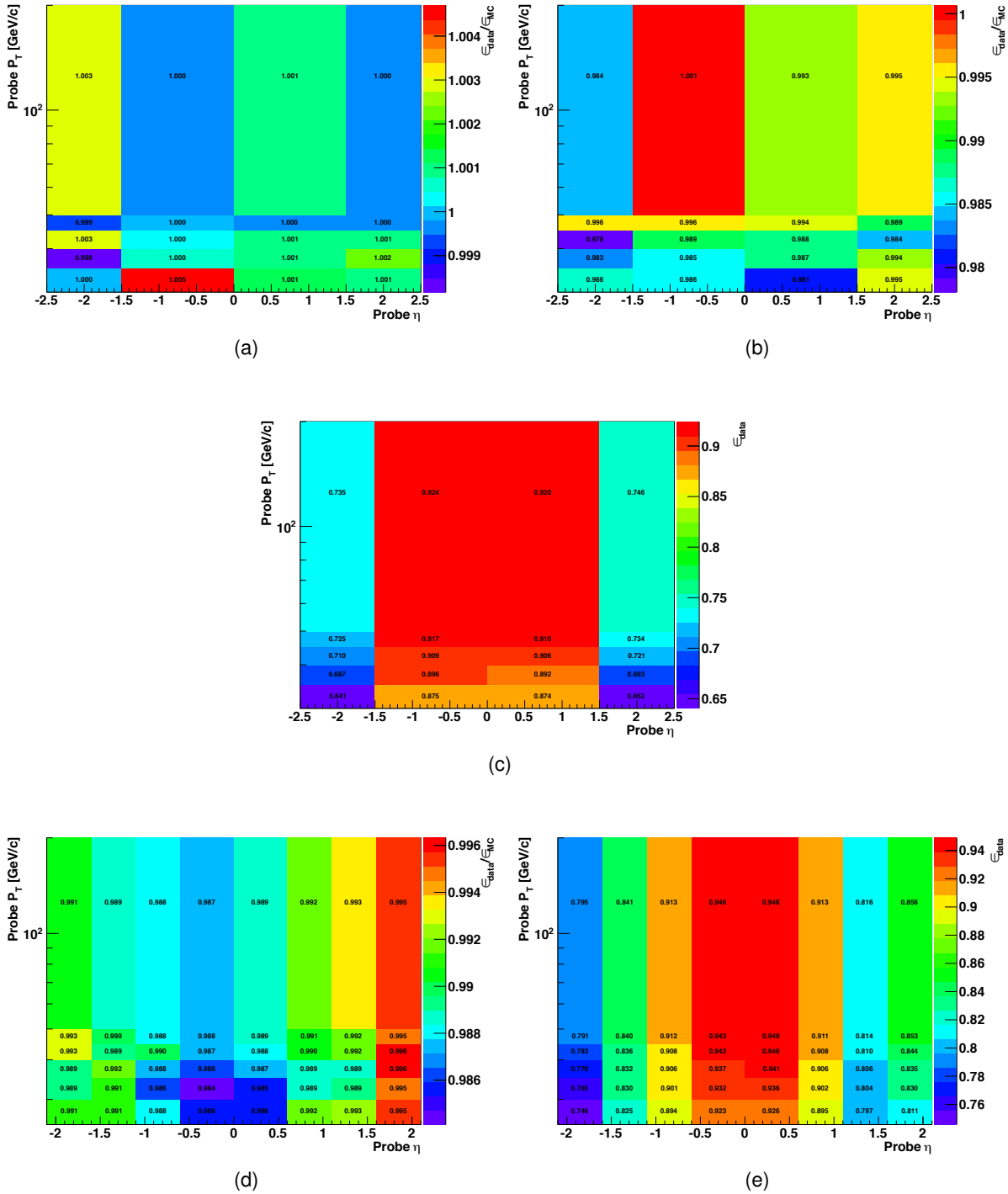


Figure 4.10: Electron efficiency and data/MC scale factors for super-cluster to reconstructed electrons ϵ_{Reco} (a), reconstructed to selected electrons ϵ_{Id} (b) and selected to HLT electrons ϵ_{HLT} (c).
Muon scale factors for reconstructed to selected muons $\epsilon_{\text{ID,data}}/\epsilon_{\text{ID,mc}}$ (a) and efficiency for selected to HLT muons $\epsilon_{\text{HLT,data}}$ (b).

4.4.3 Electron scale and resolution

As it is detailed in Chapter 3, the ECAL crystals transparency loss makes the reconstruction conditions in data not stable against time, hence, some differences might be introduced between data and MC simulation for the electron objects.

The same class of $Z \to e^+e^-$ events used for the efficiencies studies can be used to measure other electron proprieties. To make a quantitative estimation of such differences, $Z \to e^+e^-$ are fitted in data and MC (Figure 4.11), with a Crystal-Ball (CB) resolution function convoluted with the Z boson Breit-Wigner (BW), defined by theoretical predictions. This fit has direct connections to the electron reconstruction.

By noting that the mass of the di-electron can be written as

$$m_{ee} = \sqrt{2E_1E_2[1-\cos(\theta)]} \quad , \tag{4.9}$$

where E_1, E_2 refer to the electrons energies and θ is the angle between the two electrons, it can be derived, assuming the θ angle is measured with a negligible uncertainty, that

$$\frac{\sigma(m_{ee})}{m_{ee}} = \frac{1}{\sqrt{2}}\frac{\sigma(E)}{E} \quad , \tag{4.10}$$

where $\sigma(m_{ee})$ and $\sigma(E)$ are the uncertainties on the di-electron mass and electron energy respectively. Additionally, the scale on the reconstructed electron energy enters linearly in the invariant mass. In mathematical terms, if $E_{ele} \to \alpha E_{ele}$, then $m_{ee} \to \alpha m_{ee}$.

Since the gaussian core of the Crystal-Ball function, used for the Z peak fit, can be written as

$$\text{Gaus}(m_{ee}) \propto \exp\left[\frac{(m_{ee}-\Delta m)^2}{2(\sigma_{reco} \bigoplus \sigma_{smear})}\right] = \exp\left[\frac{(m_{ee}-\Delta m)^2}{2\sigma_{CB}}\right], \tag{4.11}$$

where Δm represents the reconstructed invariant mass bias, σ_{reco} the electron resolution in the simulation and σ_{smear} an additional smearing present in data, one can find the scale and resolution smearing, to be applied on top-of the simulation, in order to match the electron modeling in the MC with the data. Once the fits on data and MC are done, the relevant quantities are obtained with the following:

$$\alpha = \frac{\Delta m^{DATA}}{\Delta m^{MC}} \tag{4.12}$$

$$\sigma_{smear} = \sqrt{2\left(\frac{\sigma_{CB}^{DATA}}{\text{peak}_{CB}^{DATA}}\right)^2 - \left(\frac{\sigma_{CB}^{MC}}{\text{peak}_{CB}^{MC}}\right)^2} \quad , \tag{4.13}$$

then, the corrected electron energy in the MC simulation is obtained in this simple fashion:

$$E_{ele} \to \text{Gaus}(\alpha E_{ele}, \sigma_{smear}) \quad . \tag{4.14}$$

These corrections, summarized in Table 4.8, have been applied to the analysis presented in the following.

	α	σ_{smear}
EB	1.003 ± 0.002	$(1.38 \pm 0.01)\%$
EE	0.995 ± 0.004	$(2.78 \pm 0.02)\%$

Table 4.8: Electron scale and resolution corrections. These values, reported separately for barrel and endcaps electrons, are applied on the electron energy in the MC simulation to better model the data, using the simple formula $E_{\text{ele}} \to \text{Gaus}(\alpha E_{\text{ele}}, \sigma_{\text{smear}})$.

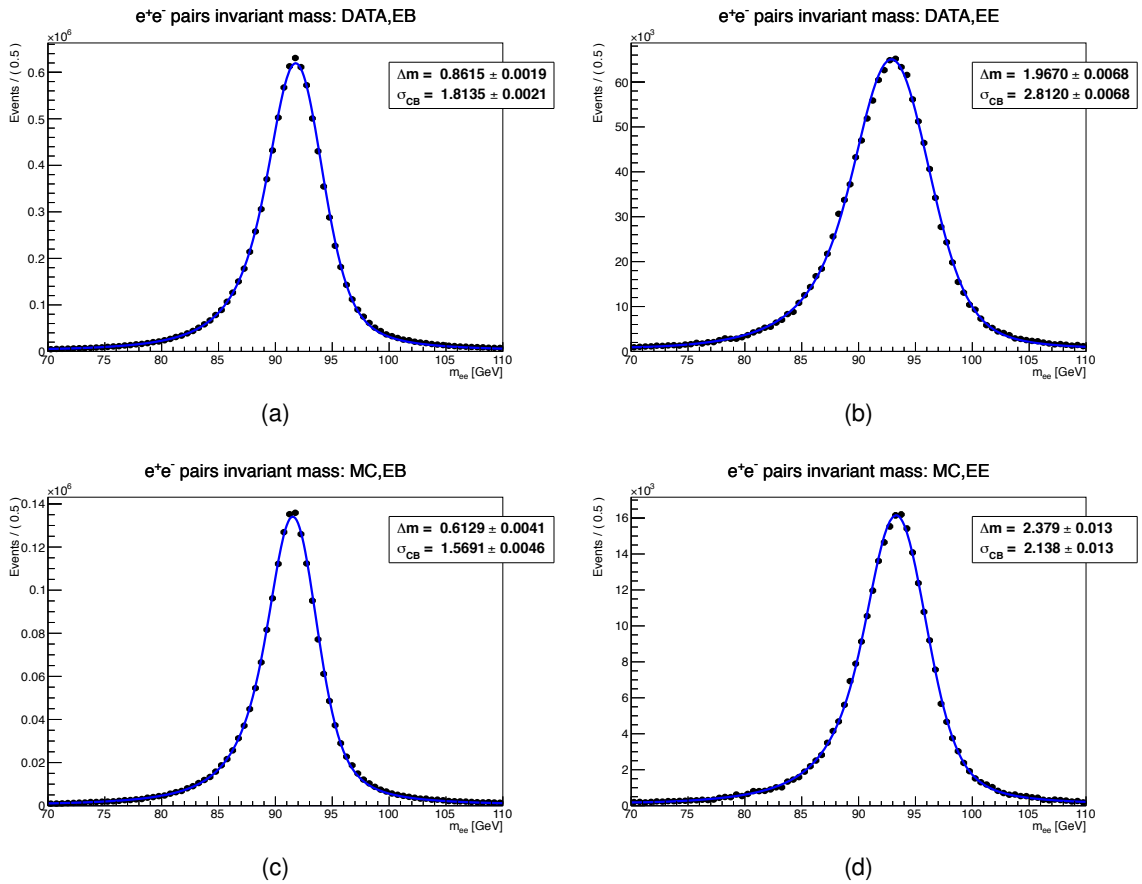


Figure 4.11: Invariant mass of the di-electron for $Z \to e^+e^-$ events in data (top) and MC (bottom), in the barrel (left) and endcaps (right). The spectra are fitted with a Crystal-Ball resolution function convoluted with the Z boson Breit-Wigner. The legend box shows the values of the center and width of the Crystal-Ball gaussian core.

4.5 Data driven QCD determination

A background from QCD multijet events comes from 3 or 4-jet events with one jet passing the lepton criteria as false positive, i.e. a *fake*. Since it is not practical to generate sufficient MC events to create a statistically significant sample that passes the selection criteria, a data-driven approach is exploited, in which the isolation-inverted samples from data, which mirror the QCD background, are used instead. Specifically, a two-component simultaneous fit to data of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is performed, in order to obtain the fraction of QCD events in the data; the two components are a data-based QCD sample and a MC-based W+jets sample. In practice, the two different $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions are associated to two normalizations which are free to float, namely N_{QCD} and $N_{\mathrm{W+jets}}$. The sum of the two normalized distributions is used as input for the maximum likelihood fit to the data, which is performed with the additional constraint $N_{\mathrm{QCD}} + N_{\mathrm{W+jets}} = N_{\mathrm{DATA}}$.

The electron QCD sample is obtained by selecting events in the data with the isolation larger than 0.3 (default selection for loose electrons uses $I_{\mathrm{rel}}^{\mathrm{PF}} \lesssim 0.2$, for muons $I_{\mathrm{rel}}^{\mathrm{PF}} < 0.12$). In order to increase the statistics of the QCD sample, the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut is relaxed from 30 GeV to 20 GeV, and the requirements on Electron MVA/Tight Muons are removed. Figure 4.12a demonstrates that taking $I_{\mathrm{rel}}^{\mathrm{PF}} > 0.3$ (rather than simply inverting the isolation cut), relaxing the $E_{\mathrm{T}}^{\mathrm{miss}}$ as well as the Electron MVA, provides a falling m_{T} spectrum (as opposed to a W-like one, which contains a peak near $m_{\mathrm{T}} = 80$ GeV). The MC W+jets and target data samples are obtained by applying the default cuts (Section 4.3).

An example of such a fit is shown for the electrons+2-jets channel in Figure 4.13.

Due to these assumptions, to account for discrepancies in template modeling (e.g. using W+jets MC as a proxy for all non-QCD processes) as well as the fact that the QCD fraction is estimated prior to the MVA cut, a very large uncertainty is conservatively assumed. The final fraction of QCD events in data, after accounting for the change in acceptance due to the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, is given in Table 4.9. The QCD selection is used also to produce the shapes used for the m_{jj} fit (Section 4.9) and the the final four-body total background determination in preparation for the limit setting procedure.

	2 jets	3 jets
electron	$10.2 \pm 0.2\%$	$5.1 \pm 0.3\%$
muon	$0.2 \pm 0.4\%$	$0.0 \pm 0.4\%$

Table 4.9: Estimates of the percentage of QCD in data (and the fit uncertainty) for the muon and electron datasets after the pre-selection.

4.5.1 QCD uncertainties

When performing the m_{jj} fit, the QCD yield is Gaussian-constrained with a mean given by the values shown in Table 4.9. The fit uncertainty on the QCD fraction is (conservatively) estimated to be one half of the expected value, i.e. a relative error equal to 50%.

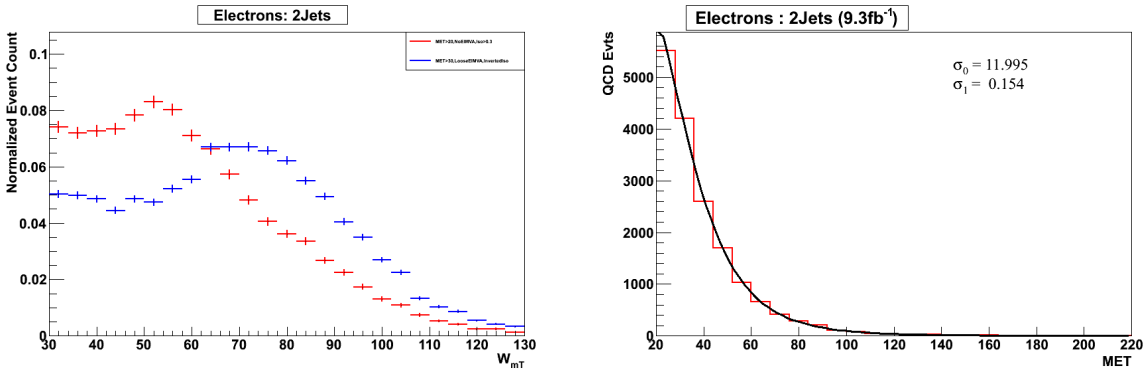
4.5.2 Cross-checks

In order to ensure that the inverted selection provides a consistent representation of QCD events, the QCD is fitted with a Raileigh Function: $xe^{-x^2/2(\sigma_0+\sigma_1 x)^2}$, used during the CMS W/Z inclusive cross-section measurements [73]. As it can be seen from Figure 4.12b, the function accurately fits the overall shape, and the best-fit value for σ_0 corresponds to the intrinsic E_T^{miss} resolution ($\sigma_0 \simeq 12$ GeV), i.e. no real source of E_T^{miss} is present in the sample.

In addition, the above procedure is compared to the fit with the remaining backgrounds included by:

- fixing the relative ratios based on the expected cross-sections and fitting the E_T^{miss} with the combination, instead of W+jets;
- fixing the additional backgrounds to their expected values and only allowing W+jets (and QCD) to float during the fit.

The results (Figure 4.14) are consistent, within 20% (below the systematic uncertainty for the m_{jj} fit), with the default approach.



(a) (b)

Figure 4.12: On the left, the W transverse mass shapes with $I_{rel}^{PF} > 0.3$, $E_T^{miss} > 20$ and no Electron MVA cut (red) vs inverted loose isolation, $E_T^{miss} > 30$ and loose Electron MVA cuts (blue), for data selected in the electron+2jets category. On the right, the E_T^{miss} distribution fitted with a Raileigh function for same data sample, selected with the looser selection.

4.6 Higgs mass reconstruction

The Higgs decay into the $l\nu_l q\bar{q}$ final state presents an issue whenever trying to reconstruct the Higgs mass, which is represented by the presence of the neutrino. Although the good hermeticity of the CMS experiments grants the ability to reconstruct the transverse components of the neutrino ν, the proton PDFs and the non-complete experimental coverage in the forward region practically inhibit the access to the z-axis component of the ν momentum.

A workaround is possible for the Higgs mass range considered by this analysis, since both the Ws produced by the Higgs decay are real. The consequence is that the invariant mass of the lepton-neutrino system must be equal to m_W. Solving the resulting second-order equation one

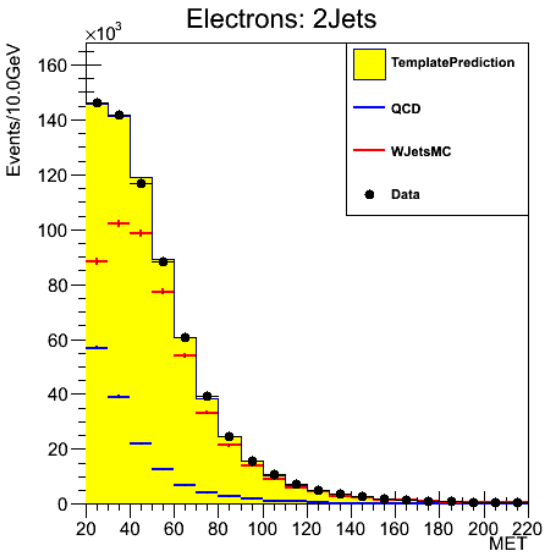


Figure 4.13: The $E_{\mathrm{T}}^{\mathrm{miss}}$ fit to the data used to extract the signal contamination coming from multijet events, for the electron+2jets category. The normalizations of QCD and W+jets templates, corresponding to the blue and red points, are varied to fit the data distribution.

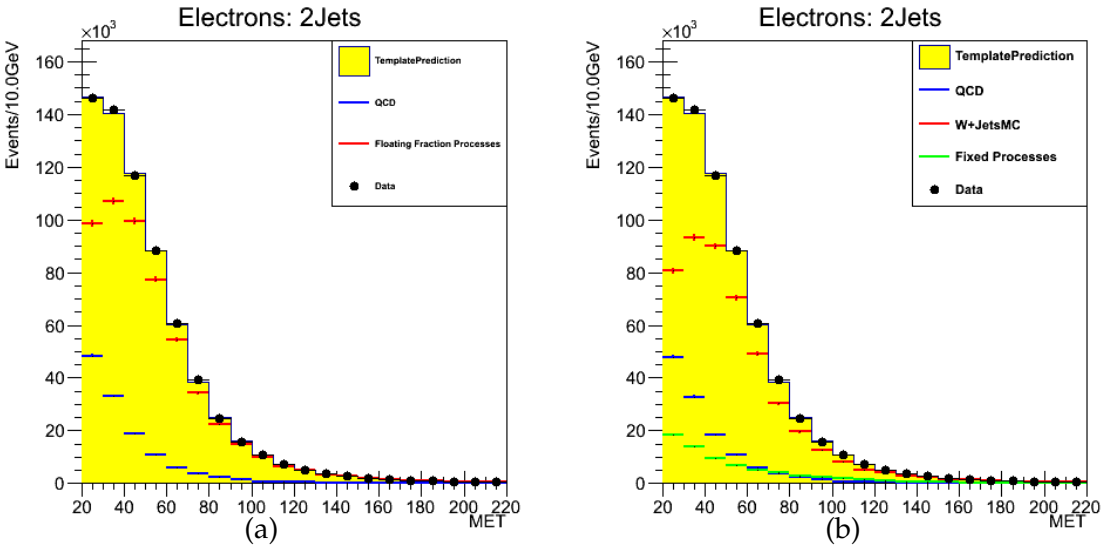


Figure 4.14: Electron $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution fit with: (a) fractions of the non QCD processes fixed relative to each other and the overall coefficient allowed to float, (b) the additional backgrounds (i.e. processes which are not W+jets or QCD) fixed to their expected yields and W+jets (as well as QCD) fraction allowed to float. The resultant fraction of QCD events is consistent with the default approach (Figure 4.13).

finds:

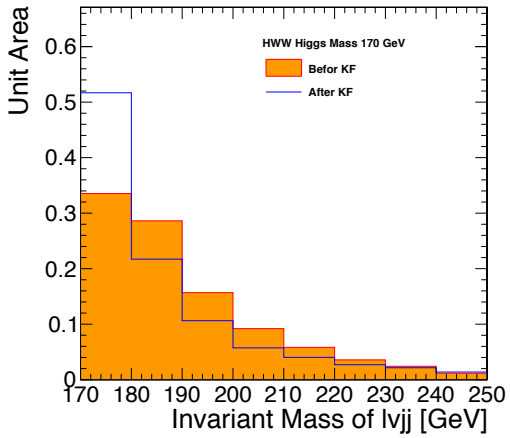
$$
\begin{aligned}
(p_z^{\nu})_{1,2} &= \frac{p_z^{\mathrm{lepton}} \pm \sqrt{\Delta}}{\left(p_{\mathrm{T}}^{\mathrm{lepton}}\right)^2} \quad , \quad \text{where} \\
p_z^{\mathrm{lepton}} &= \sqrt{\vec{p}_{\mathrm{T}}^{\,\mathrm{lepton}} \cdot \vec{E}_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{W}}^2/2} \quad \text{and} \\
\Delta &= \left(\vec{p}_{\mathrm{T}}^{\,\mathrm{lepton}} \cdot \vec{E}_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{W}}^2/2\right)^2 - \left(p_{\mathrm{T}}^{\mathrm{lepton}}\right)^2 \left(E_{\mathrm{T}}^{\mathrm{miss}}\right)^2 .
\end{aligned}
\tag{4.15}
$$

Due to the finite detector resolution and uncertainties in the reconstruction of E_T^{miss}, it can happen that the discriminant Δ is smaller than zero, in which cases it is set to zero and a unique solution is obtained for p_z^ν. For the remaining events, the discriminant Δ is larger than zero, and two distinct solutions exist. In this case, the one yielding the smaller value of $|p_z^\nu|$ for each event is chosen, after verifying via matching with the Monte Carlo generated particle that this choice corresponds to the correct solution in the majority of the cases (61-65%, depending on the Higgs mass).

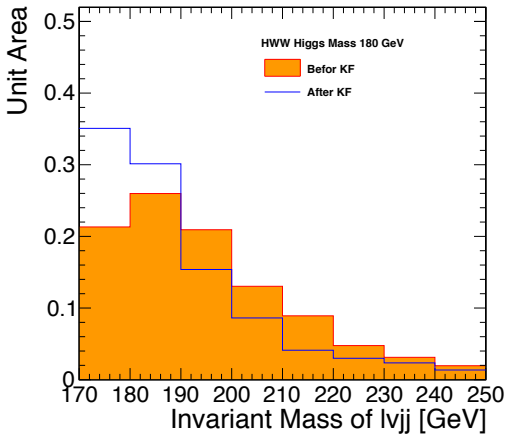
In addition, in order to improve the mass resolution in the four body mass measurement, a two constraints fit with one unknown parameter, the neutrino z momentum component, is made. The constraints are that the lepton-neutrino pair and the di-jet system both separately form an invariant mass m_W within the W known width Γ_W. The infrastructure of the fit uses standard CMS kinematic tools [74]. The four particles are part of the fit, with a covariance matrix supplied for the jets, E_T^{miss} and the lepton, according the their experimental resolutions. The starting value input to the fit for the x and y components of the neutrino is the measured missing transverse energy along the x and y axes, while the z component is found by solving the second-order equation as described above. Mathematically, the best value for the neutrino p_z is found through the minimization of a χ^2 variable of this form:

$$\chi^2 = \sum_{i=\text{lepton},\nu,\text{jet}_1,\text{jet}_2} \frac{\left(p_i^{\text{fit.}} - p_i^{\text{meas.}}\right)^2}{\sigma_i^2(p_T,\eta)} + \frac{\left(m_{l\nu} - m_W\right)^2}{\Gamma_W^2} + \frac{\left(m_{jj} - m_W\right)^2}{\Gamma_W^2} \quad , \tag{4.16}$$

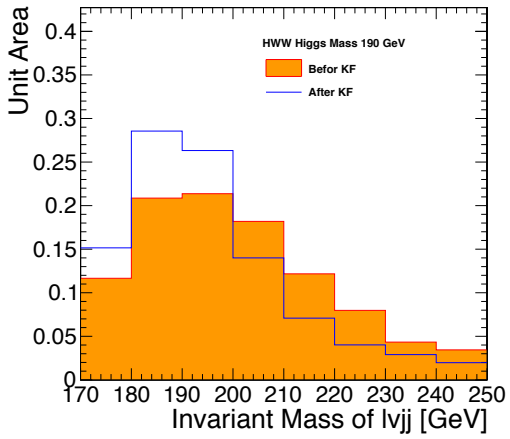
and events which fail the minimization convergence are discarded. Figure 4.15 shows the effect of the kinematic fit on the four-body mass distribution for the signal at the four lowest mass points, as well as for the non-resonant WW and W+Jets backgrounds. The distributions obtained before and after the kinematic fit are reported. The effect of the kinematic fit becomes less pronounced for large signal masses, where the width of the Higgs mass peak is dominated by the intrinsic decay width Γ_H and not by detector resolution effects.



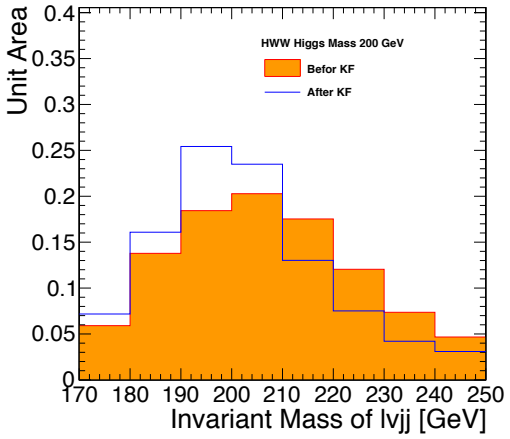
(a) Higgs M=170 GeV



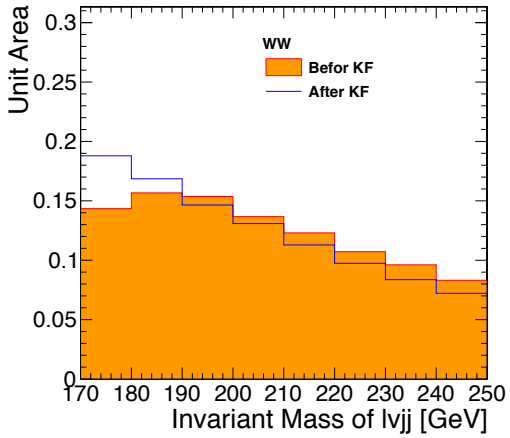
(b) Higgs M=180 GeV



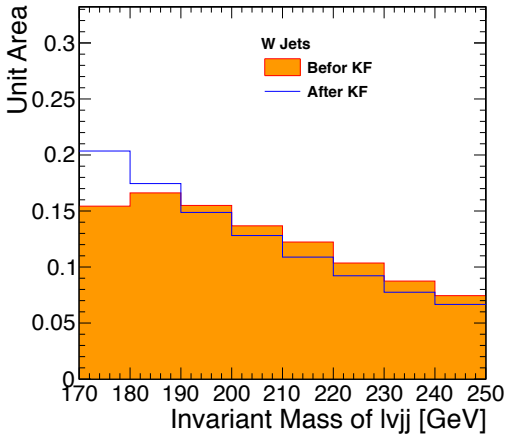
(c) Higgs M=190 GeV



(d) Higgs M=200 GeV



(e) WW



(f) W+Jets

Figure 4.15: The four-body mass distributions, obtained before and after the kinematic fit, are reported for the four lowest Higgs mass hypotheses as well as for WW and W+Jets backgrounds. The mass range is constrained to that used for subsequent template modeling and limit setting.

4.7 Data - Monte Carlo comparisons

To assess the quality of the modeling provided by the MC simulation, comparisons between the MC shapes, normalized to the collected integrated luminosity, and the data, after applying the event selection criteria listed in Section 4.3, are performed. The MC is corrected for lepton reconstruction and trigger efficiencies as well as for scale and resolution mismodeling effects, and the QCD contribution is taken from the data-driven estimation reported in Section 4.5. For all the distributions, the errors correspond to the combined statistical uncertainty of the data and MC samples, while, in the data/MC ratio plots, the yellow band is related to the uncertainty in the integrated luminosity measurement, which directly affects the total MC normalization.

First of all, the impact of the selections, on data and MC, is reported in Figure 4.16 for the inclusive final state, i.e. 2j plus 3j events. For the same class of events, distributions of various kinematic variables for the relevant final state objects are shown in Figures 4.17-4.19 for the muon+jets sample. Finally, Figures 4.20 and 4.21 show the two-body and the four-body invariant mass, for electrons (on the left) and muons (on the right).

In general, seeing reasonable agreement gives confidence in the qualitative aspects of the MC modeling. Anyway, from the non-perfect description of the Monte Carlo simulation of some key variables, such as the reconstructed four-body invariant mass, it is clear that a data-driven approach has to be followed for a correct background evaluation. Concerning the signal modeling, where no data-driven can be used, another control region is used to cross-check the goodness of the simulation.

The $t\bar{t}$ +jets events provide the ideal environment for this purpose, since a large statistics sample containing WW pairs can be identified. A special selection, where 2/3 non b-tagged jets and 2 b-tagged jets are required on top of the standard selections, can provide high $t\bar{t}$ purity. Then, the control plots can be produced on this final state without considering the 2 b-jets, in particular the four-body invariant mass, as it is shown in Figure 4.22. These last plots prove that, in a control region similar to the one where the signal lives, the MC is well modeling the data, hence validate the use of the simulation to compute signal expectations.

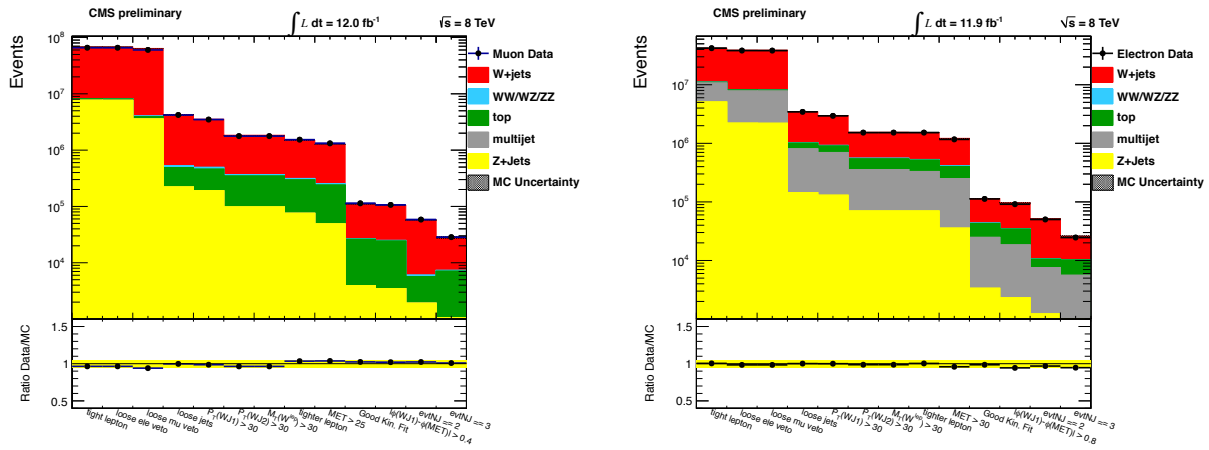


Figure 4.16: Cut-flow of 8 TeV data samples superimposed to the expected yields of events from Monte Carlo simulation in 12.0 fb^{-1}. Each plot shows the inclusive final state (2/3 jets), for $\mu\nu_\mu$ jj (left) and $e\nu_e$ jj (right).

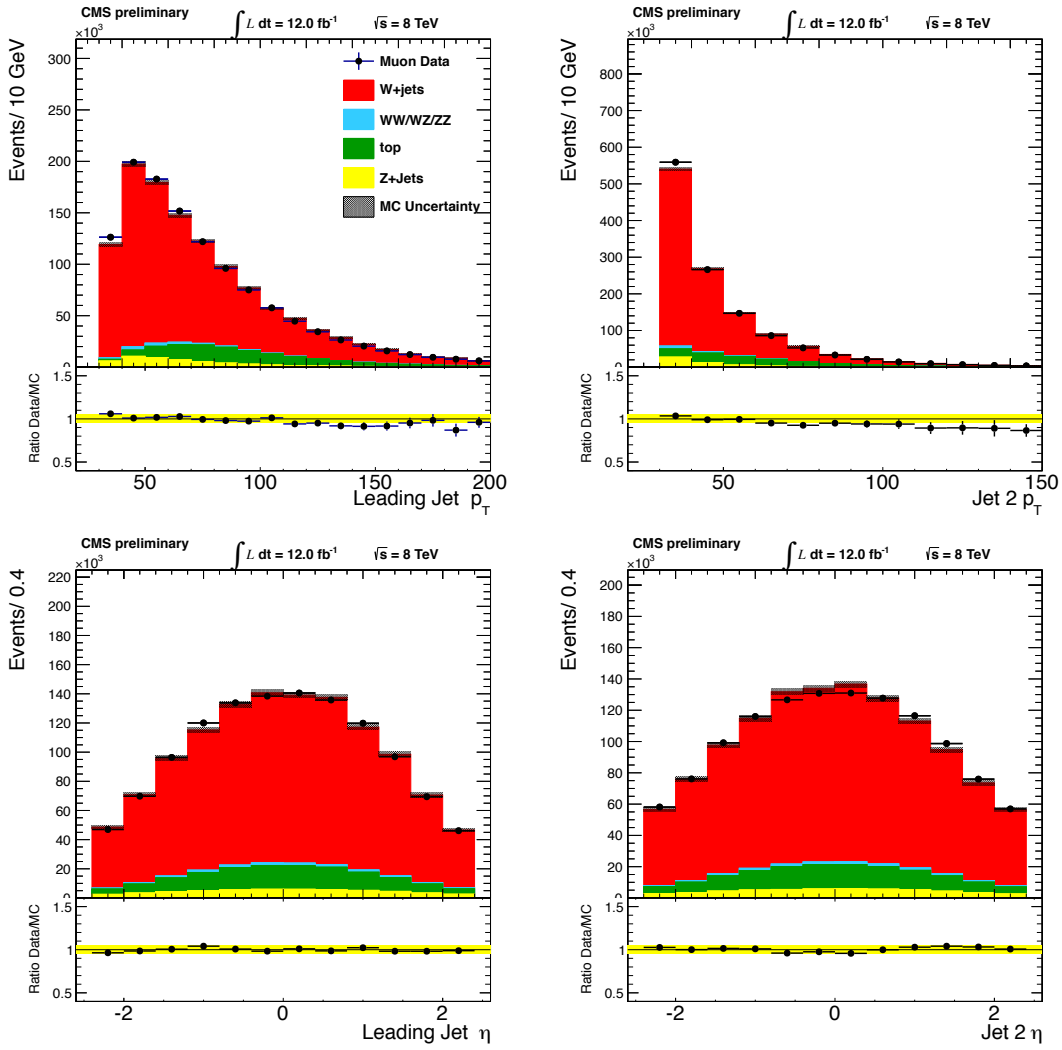


Figure 4.17: Data vs MC comparison of leading (left) and trailing jet (right) for the p_T (top) and η (bottom) distributions, for the muon+jets selection.

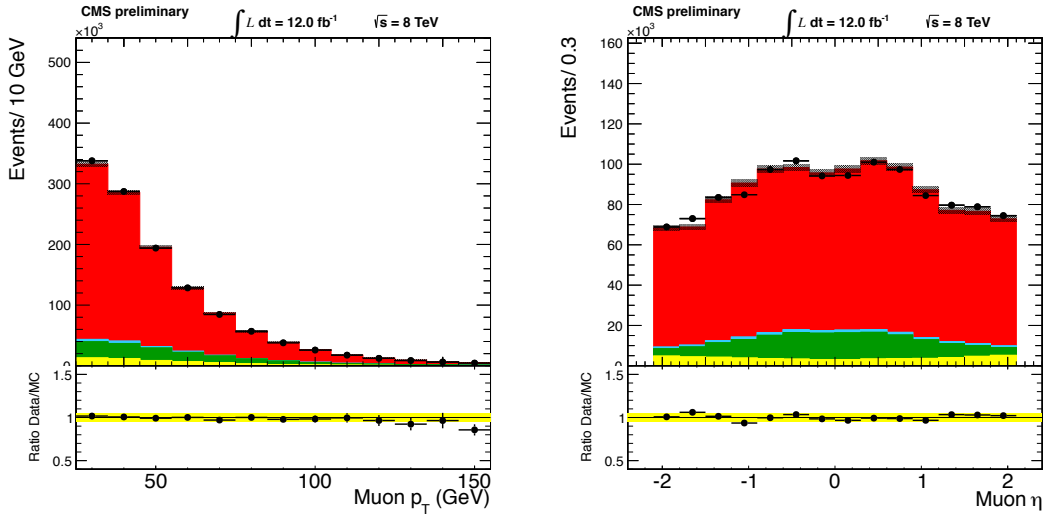


Figure 4.18: Comparison of the muon p_T (left) and the muon η (right) distributions from data and MC, for the muon+jets selection.

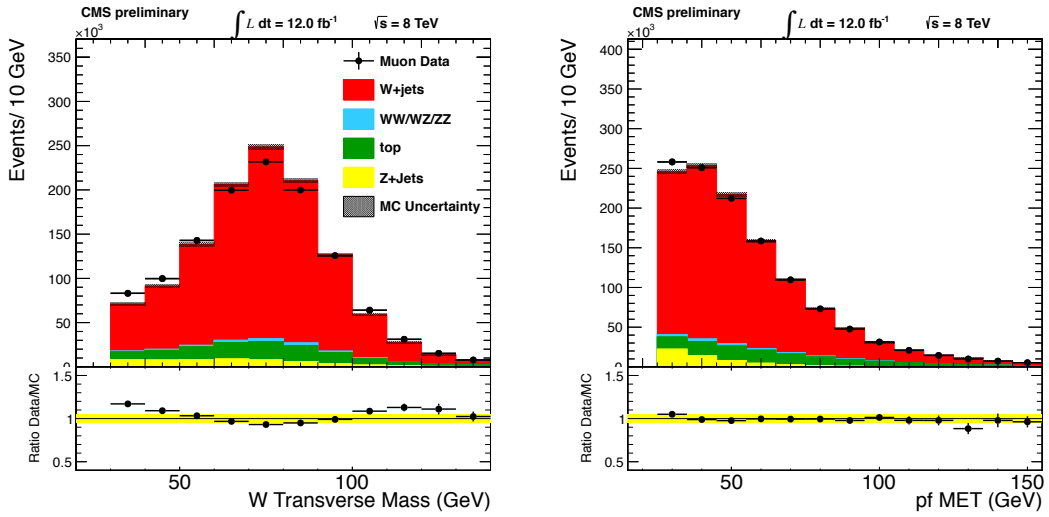


Figure 4.19: Comparison of the distributions from data and MC of the transverse mass of the muon-$E_{\mathrm{T}}^{\mathrm{miss}}$ system (left) and the $E_{\mathrm{T}}^{\mathrm{miss}}$ (right), for the muon+jets selection.

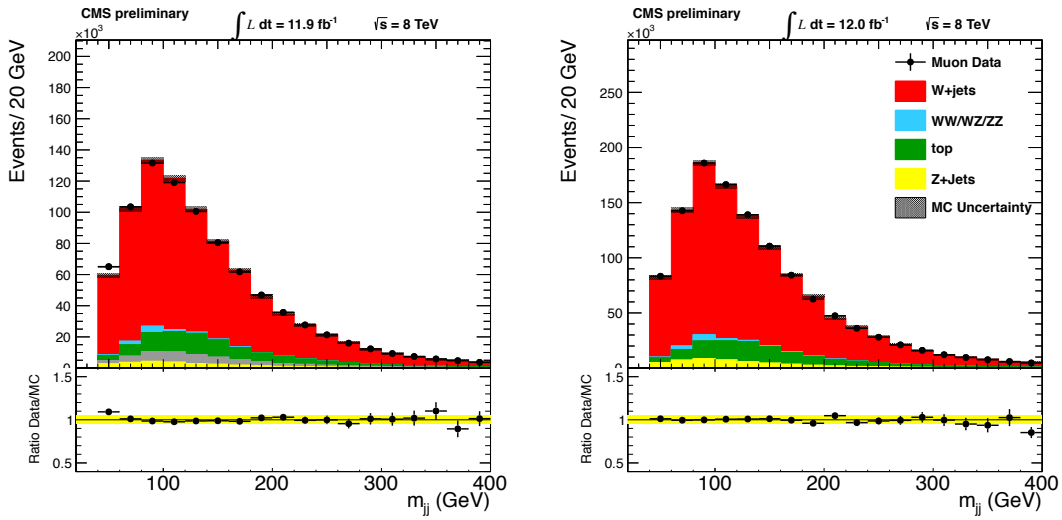


Figure 4.20: Comparison of the dijet mass (m_{jj}) distributions from data and MC, for the electron+jets (left) and muon+jets (right) selection.

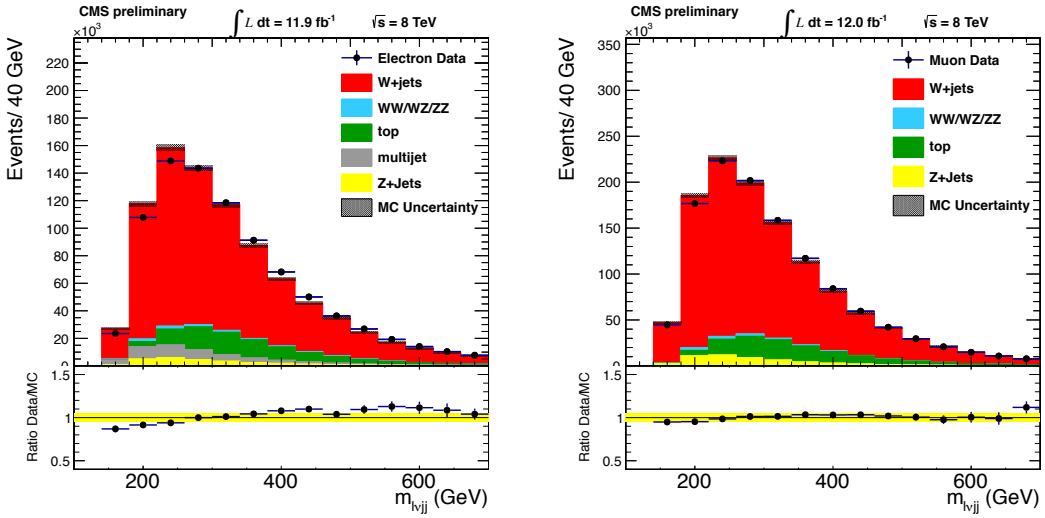


Figure 4.21: Comparison of the four-body invariant mass from data and MC for the electron+jets selection (left) and muon+jets selection (right). The disagreement seen here between data and MC for W+jets background is the motivation for using data-driven shape for W+jets as described in Section 4.9.

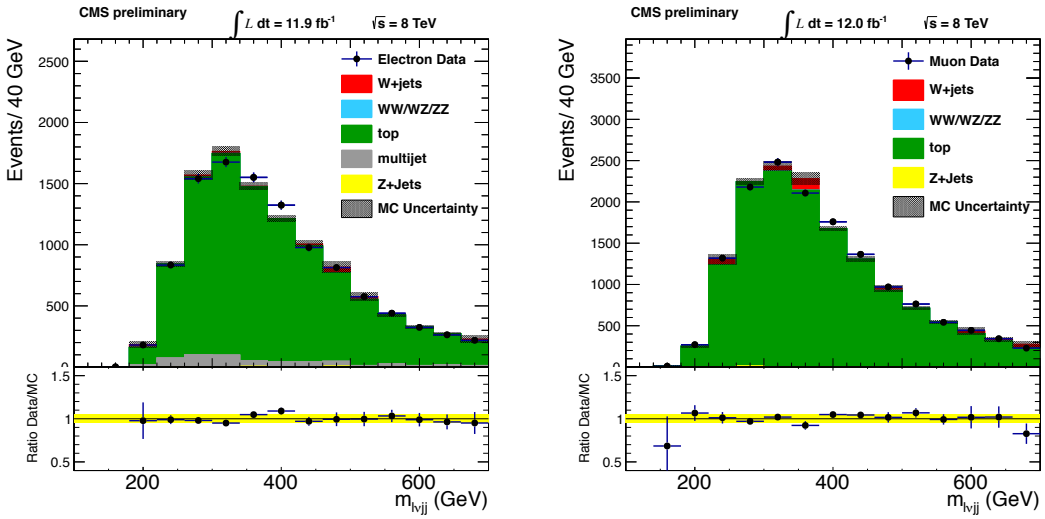


Figure 4.22: Comparison of the four-body invariant mass from data and MC, for the electron+jets selection (left) and muon+jets selection (right), in the top control region.

4.8 MVA optimization

Using a complete set of mostly uncorrelated variables, a dedicated optimization is performed for each Higgs mass hypothesis, to distinguish between the Higgs signal and the dominant W+jets background.

To achieve a better separation between signal and background, different MVAs are trained separately on events with two jets and three jets in the final state, because the background composition and kinematics are different for the two categories. In addition, the training is performed independently for the two lepton species, i.e., muons and electrons.

4.8.1 Training and validation method

The Toolkit for Multivariate Data Analysis with ROOT (TMVA) [75] is used to carry out a multivariate analysis based on a likelihood estimator.

The MVA is trained separately for the following 12 Higgs mass points: 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 550, and 600 GeV/c^2. Exactly 50% of the events in each category are used for training the classifier and the other 50% are used for testing and validation.

A minimum number of loosely correlated variables necessary to describe the event topology are used. The set is found by noting that, at leading order, the final state is fully characterized [76, 77] by

$$\{m_{l\nu jj}, m_{jj}, \cos\theta_1, \cos\theta_2, \Phi, \cos\theta^*, \Phi_1\} \quad , \tag{4.17}$$

where the angular variables are defined in Figure 4.23. The angle θ^* is the polar angle between

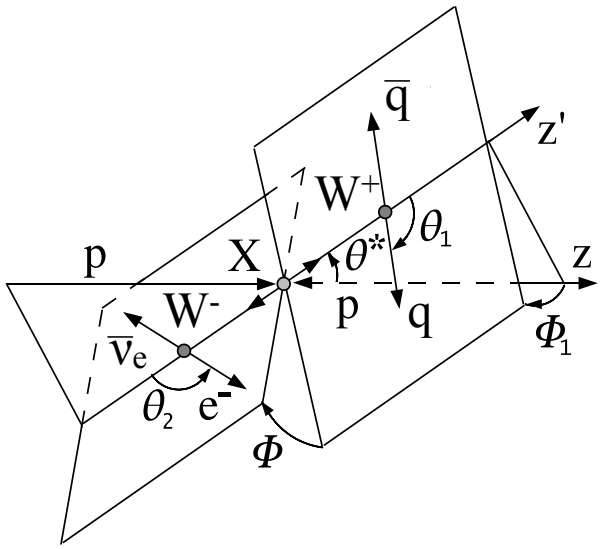


Figure 4.23: Angular variables definitions for the $H \to WW \to l\nu_l q\bar{q}$ process.

the parton collision axis z and the Higgs decay axis z', both defined in the Higgs rest frame. The angle Φ_1 is the azimuthal angle between the zz' plane and the decay plane of the hadronic W. The two are called production angles because they depend on the production mechanism, gg or $q\bar{q}$. For the SM Higgs, which is a spin-zero particle, the production angles are flat (before acceptance selections). The angle Φ is the angle between the decay planes of the two W systems

in the Higgs rest frame. The angle θ_i is the angle between the direction of one fermion f, from the $W \to f\bar{f}$ decay, and the direction opposite to the Higgs, in the W_i rest frame, where the index i refers to the W bosons. The angle θ_1, referred to the hadronic W, is ambiguously defined since, experimentally, the two quarks appears as identical hadronic jets. Then, by convention, the angle is defined from 0 to π for the leading p_T jet. Contrarily, for the angle θ_1, referred to the leptonic W, the only possible choice is to measure the angle between the lepton and the WW direction, since the neutrino 3-d direction is unknown. The angles Φ, θ_1 and θ_2 do not depend on the production mechanism and are denoted as the helicity angles.

The observables m_{lvjj} and m_{jj} are excluded from the MVA inputs:

- m_{jj} because it is used to find the background normalization;
- m_{lvjj} because it is used in the extraction of the cross-section upper limit.

On the other hand, even if the m_{jj} and m_{lvjj} are excluded from the MVAs, they do enter indirectly in the MVA definition, since angles and masses are mildly correlated and the angular variables are defined in a loose window around $m_{lvjj} = m_H$.

This leaves the five angular variables available as inputs to the multivariate discriminant. In addition to the five angular variables, the p_T of the WW system $p_{T,WW}$ and longitudinal boost (pseudorapidity) η_{WW} are also used. The lepton charge is also included to give some more discrimination power since the W+jets background is asymmetric with respect to charge while the Higgs signal is symmetric.

Combining all together, the final set of inputs to the multivariate discriminant is:

$$\{\cos\theta_1, \cos\theta_2, \Phi, \cos\theta^*, \Phi_1, p_{T,WW}, \eta_{WW}, \text{lepton charge}\} \quad . \tag{4.18}$$

Figure 4.24 shows the data vs. MC comparisons for the MVA input variables, after the preselection, for the muon plus jets final state for 8 TeV data. Similarly, Figure 4.25 shows the response of the data to the MVA, superimposed with background and signal expectations. Once again, the simulation does not well reproduce the data in the W+jets dominated region for the WW observables, hence the output of the discriminant is not perfectly modeled. As it will be shown later (Section 4.11), the simulation does well reproduce the data in a signal like region, obtained through the $t\bar{t}$ event selection, and the background mis-modeling has no impact in the limit setting procedure, since the background is fully re-derived after the MVA selection with data-driven approaches (Section 4.9 and 4.10).

4.8.2 Optimization method

The discriminant is exploited by means of a single cut on it, chosen for each of the working points separately, on the basis of a scan over a range of MVA cut values at discrete intervals. The figure of merit is chosen to be the expected median exclusion limit, obtained for each given cut value. The same limit setup used for the final results (Section 4.12) is used in the scan, and a reasonable assumption on the systematics (Section 4.11) is made, namely:

- full signal theoretical uncertainties;
- all lepton related uncertainties;

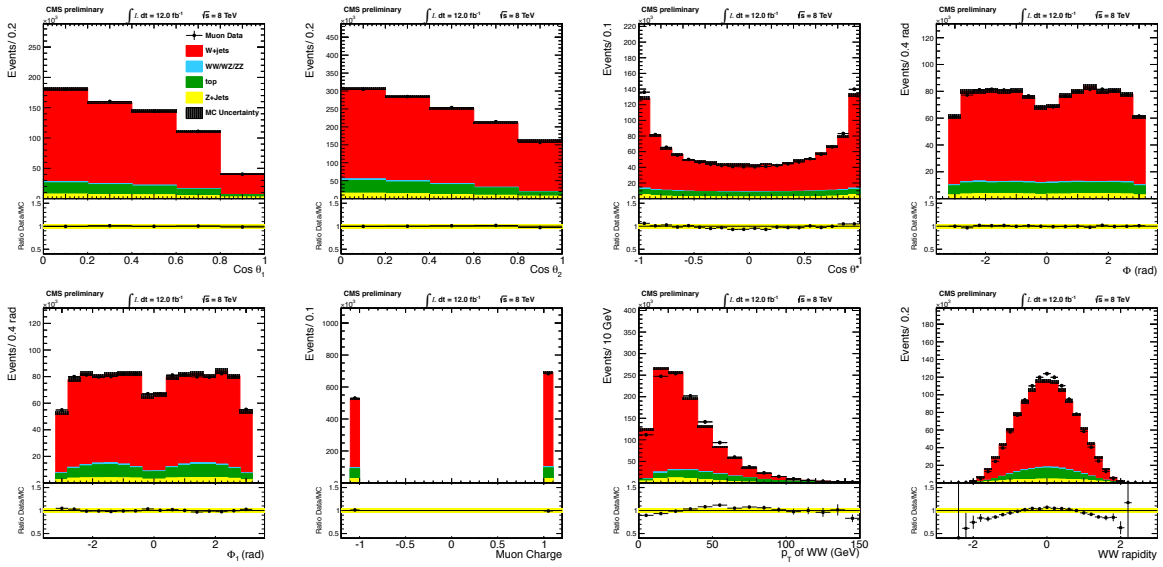


Figure 4.24: Data vs. MC Comparison of MVA input distributions for the muon-2jets selection, for the 8 TeV dataset. From left to right, top to bottom row, $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$, Φ, Φ_1, lepton charge, WW p_T and WW η.

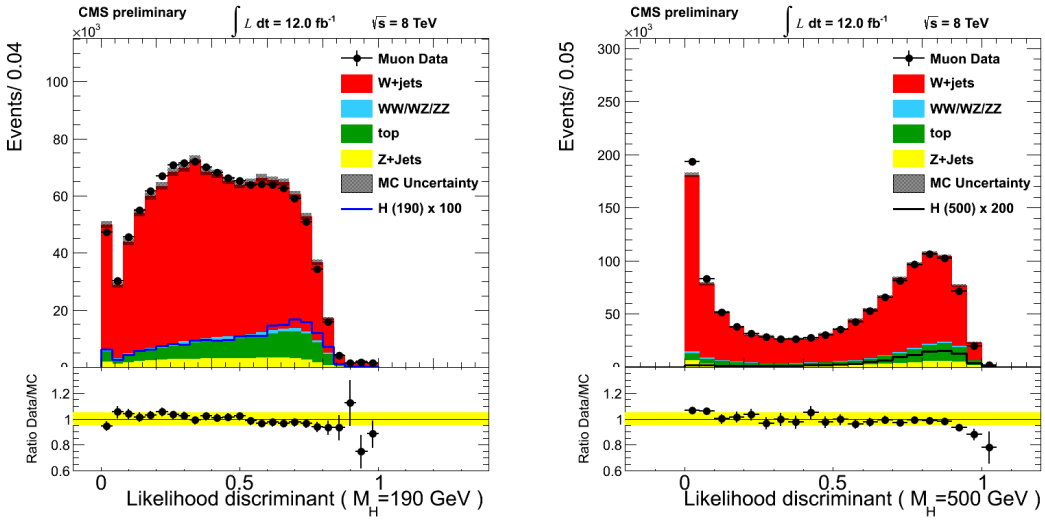


Figure 4.25: Data vs. MC Comparison of MVA output distributions for the muon-2jets selection, for the 8 TeV dataset. On the left (right) an example of a low (high) Higgs mass optimized likelihood discriminant. The signal regions are defined by the [0.6,1] (left) and [0.5,1] (right) ranges.

- jet and missing transverse energy normalization uncertainties;
- uncertainties on background normalization and shape.

4.9 Determination of background normalization from m_{jj}

For each mass hypothesis, the MVA cut is applied and a loose window in m_{lvjj}, centered around m_{H}, is selected. Then, background yields are extracted from an unbinned maximum likelihood fit to the dijet invariant mass distribution m_{jj}, excluding the signal region (65 GeV < m_{jj} < 95 GeV) to retain signal sensitivity. In this fit all the backgrounds are considered. The W+jets modeling is described in Section 4.9.1, the QCD shape is obtained using the dedicated selection described in Section 4.5, while the other background shapes come from the simulation. During the fit, the Jet Energy Scale (JES), which acts linearly on m_{jj}, is free to float in a 1-σ window around the best value found with the $\mathrm{t\bar{t}}$ control region fit presented is Section 4.11. The dijet mass spectrum fit fixes therefore the yields of the physics processes. Table 4.10 shows in a schematic view how the shape of each component is determined, and what constraints are applied on the normalization during the fit.

Process	Shape	Shape syst.	Normalization	Norm. syst.
W+jets	data	—	Unconstrained	Unconstrained
diboson	MC	JES	Constrain: NLO	Gauss $\sigma = 10\%$
$\mathrm{t\bar{t}}$	MC	JES	Constrain: NLO	Gauss $\sigma = 6.3\%$
single top	MC	JES	Constrain:NLO	Gauss $\sigma = 5\%$
Z+jets	MC	JES	Constrain: NLO	Gauss $\sigma = 4.3\%$
QCD	data	JES	Constrain: $E_{\mathrm{T}}^{\mathrm{miss}}$ fit in data	Section 4.5.1

Table 4.10: Determination of the m_{jj} shape and normalization.

4.9.1 Modeling of W+jets background shape in m_{jj}

To determine the expected number of W+jets events in the signal region, it is necessary to know its shape in the m_{jj} variable.

Because of inadequate statistics in the W+jets MC and the overall poor agreement between W+jets MC and data distributions, an empirical description of the W+jets shape is employed. This description is a kinematic turn on and a power law tail:

$$\mathcal{F}_{\mathrm{W+jets}} = \mathrm{erf}(m_{jj}; m_0, \sigma) \times \left[(m_{jj})^{-\alpha - \beta \ln(m_{jj}/\sqrt{s})} \right] , \tag{4.19}$$

where m_0 is the value of the turn on and σ is the width of this turn on. The parameters m_0, σ, α and β are determined in the fit to the data m_{jj} sideband. For mass points below 250 GeV, the parameters of the kinematic turn-on cannot be reliably determined because the events in the excluded signal region fall into this area. In this case, the W+jets MC is used to fix the m_0 parameter of the turn-on, while in the data fit the turn-on width is still allowed to float in a 1-σ window corresponding to the MC fit.

To test how a different choice of a functional form impacts on the W+jets estimation in the signal region, a different turn model (Fermi function) and a different falling shape (second order exponential and power law) are tested. The variation of the W+jets yield is found to be negligible with respect to the fit uncertainty and therefore neglected.

As a sanity check, the fit is re-done while releasing the constraint on the WW cross-section, and then comparing the WW yield found with the fit with the one from the MC prediction.

The values are found to be compatible within uncertainties. This fit, with the unconstrained diboson contribution, constitutes the core of the WW cross-section measurement in the semi-leptonic final state, published in 2012 by the CMS collaboration [78].

4.9.2 Fit examples

Figure 4.26 contains three examples of fits, for a low, medium and high Higgs mass hypothesis, for the muon+2jets category. Each plots contains the background estimation after the fit, as well as the signal contribution expected for 12 fb^{-1} of 8 TeV collision data magnified by a factor ten.

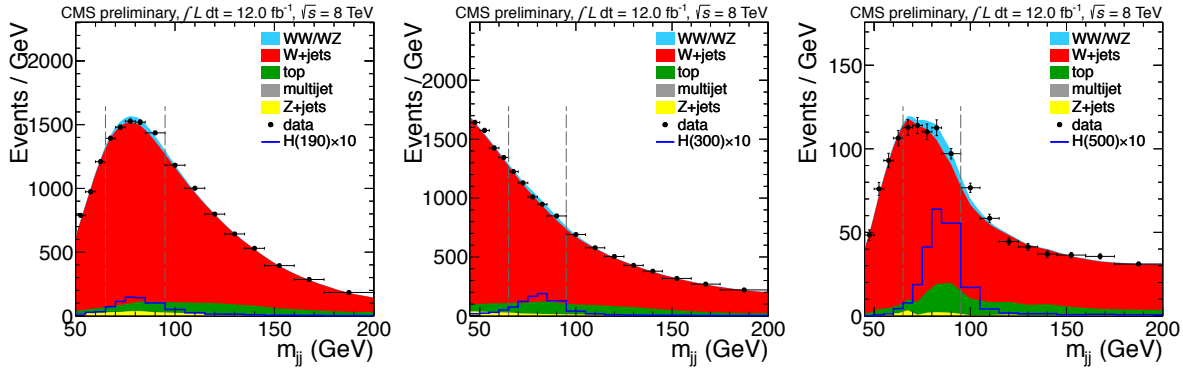


Figure 4.26: The di-jet invariant mass distribution with the fit projections, for the muon 2-jet category, after selections optimized for the Higgs mass hypotheses of 190 GeV/c^2 (left), 300 GeV/c^2 (center), and 500 GeV/c^2 (right). The fit does not include data points in the test region between the vertical lines.

4.10 Use of four-body mass to extract Higgs limits

Having determined the yield of the ensemble of physics processes, using the dijet sidebands, the four body mass spectrum, for events where the dijet mass lies in the W mass window, is explored and used for the limit setting. The most important feature of this variable is that the signal is peaking around m_{H}, while all the backgrounds are essentially characterized by a falling shape.

For the signal and the minor backgrounds, the shapes are obtained from the simulation. For the W+jets background, the WW four body mass is derived by data-driven means as described in Section 4.10.1.

Figures 4.29 and 4.30 show an example of the final four-body mass distributions for background and signal events. Data are superimposed.

4.10.1 Determination of the W+jets shape in $m_{\ell\nu jj}$

The four-body mass shape of the W+jets background in the signal region is estimated in a data driven way from sidebands in the m_{jj} region.

For each of the 48 working points of the MVA optimization, three regions in m_{jj} are looked at:

- lower sideband region (SBL): $m_{jj} \in [55, 65]$ GeV;
- signal region: $m_{jj} \in [65, 95]$ GeV;

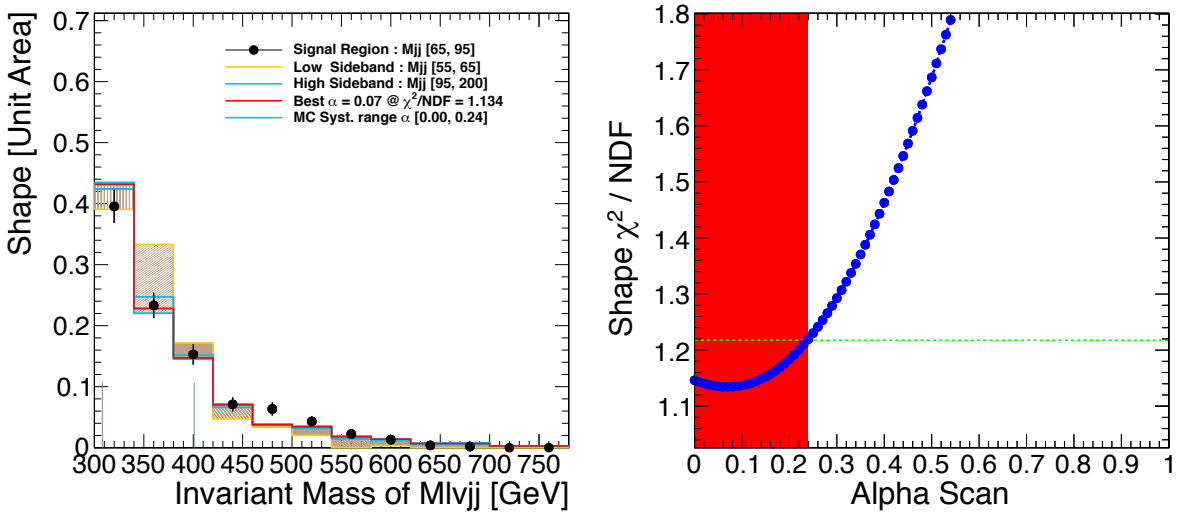


Figure 4.27: The optimal α value from W+jets MC for the SM Higgs mass of 350 GeV for the 2-jet $W \to \mu\nu$ category.

- upper sideband region (SBH): $m_{jj} \in [95, 115]$ for $m_{\mathrm{H}} < 250$ GeV, $[95, 200]$ GeV for $m_{\mathrm{H}} \geq 250$ GeV.

In the Monte Carlo, the $m_{\ell\nu jj}$ shapes in the upper and lower sidebands are compared to the one in the signal region, to find the best mixture of the first two that reproduce the latter. Therefore an α parameter is searched for, such that:

$$m^j_{\mathrm{WW}_i} = (1 - \alpha^j) \cdot m^j_{\mathrm{SBH}_i} + \alpha^j \cdot m^j_{\mathrm{SBL}_i} \,, \tag{4.20}$$

where the index j refers to each of the 48 mass spectra and i to the bins in the four-body mass. In this way, the technique is largely data driven but the precise extrapolation depends on the Monte Carlo model. The value for the best α in W+jets MC and the χ^2/NDF scan of the shapes are shown in Figs. 4.27 as an example for the SM Higgs mass of 350 GeV for the 2-jet $W \to \mu\nu$ category. The α 68% confidence interval is then used to derive a systematic uncertainty associated to the method as detailed in Section 4.11.

The W+jets shape derived from the sidebands is then smoothed with a parametric fit to an exponential shape, modulated with a turn-on when necessary, i.e. for $m_{\mathrm{H}} < 250$ GeV/c^2. The distribution of the extrapolated W+jets background in the signal region is reported for four working points in Figure 4.28. The black dots represent the extrapolated background, the solid line shows the fitting function and the dashed lines the uncertainty from the fit.

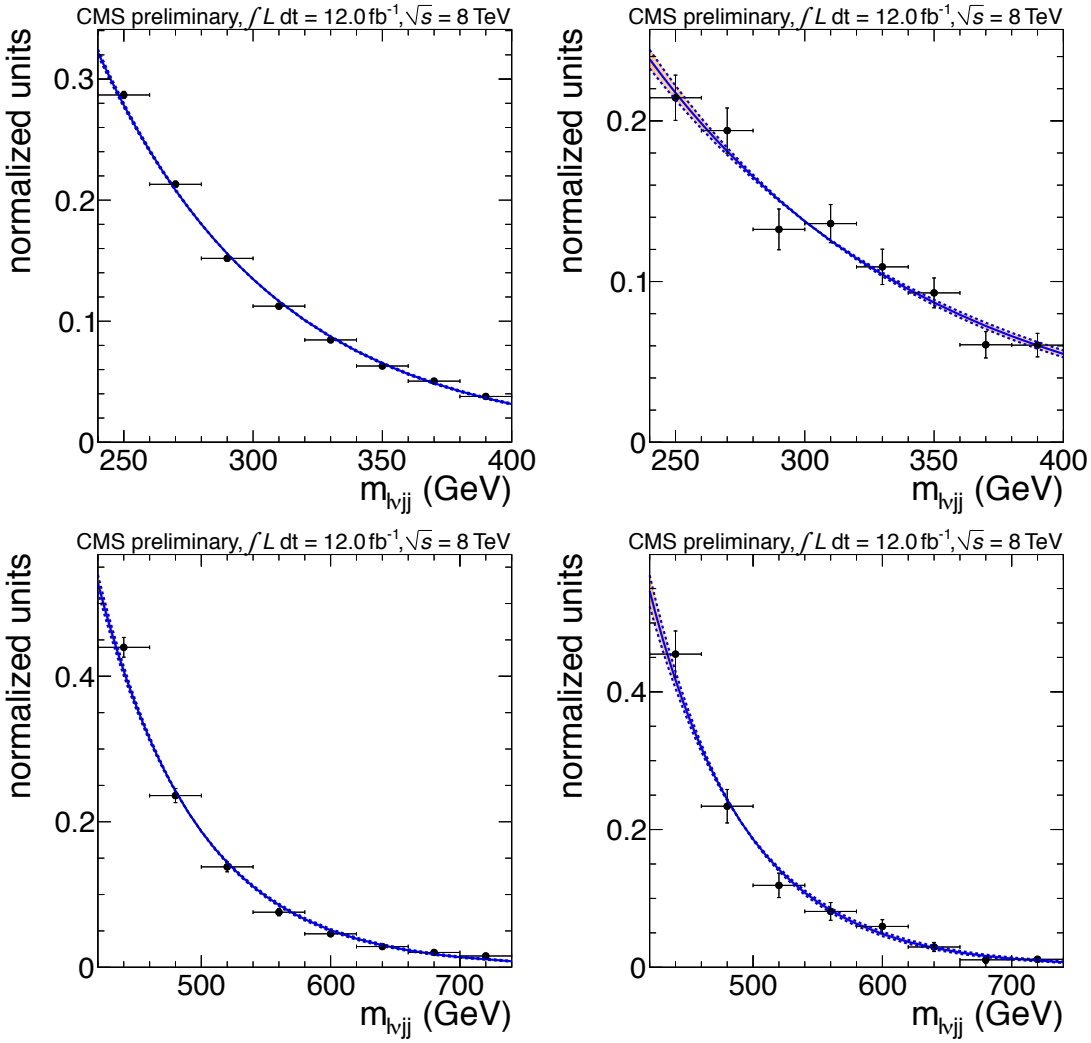
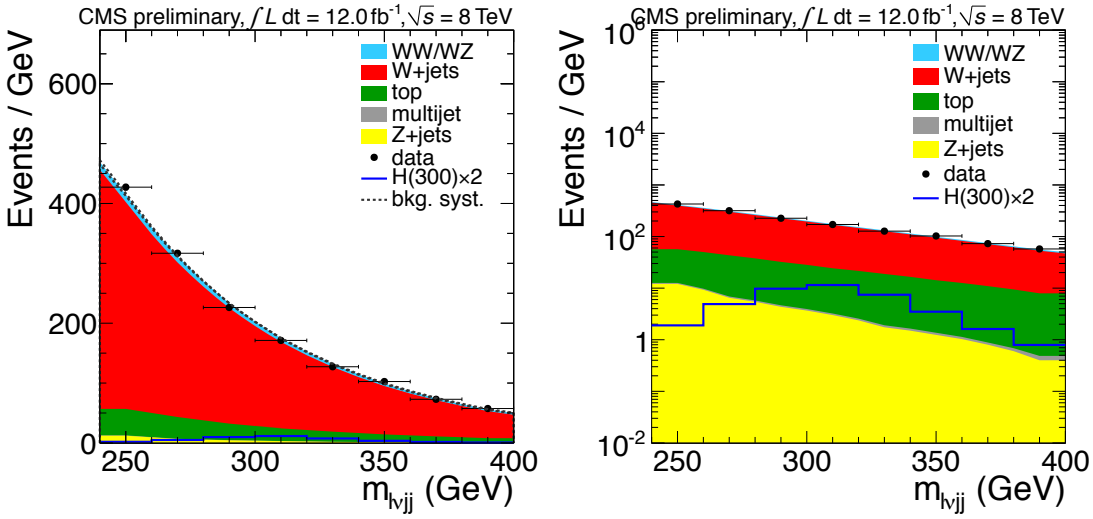
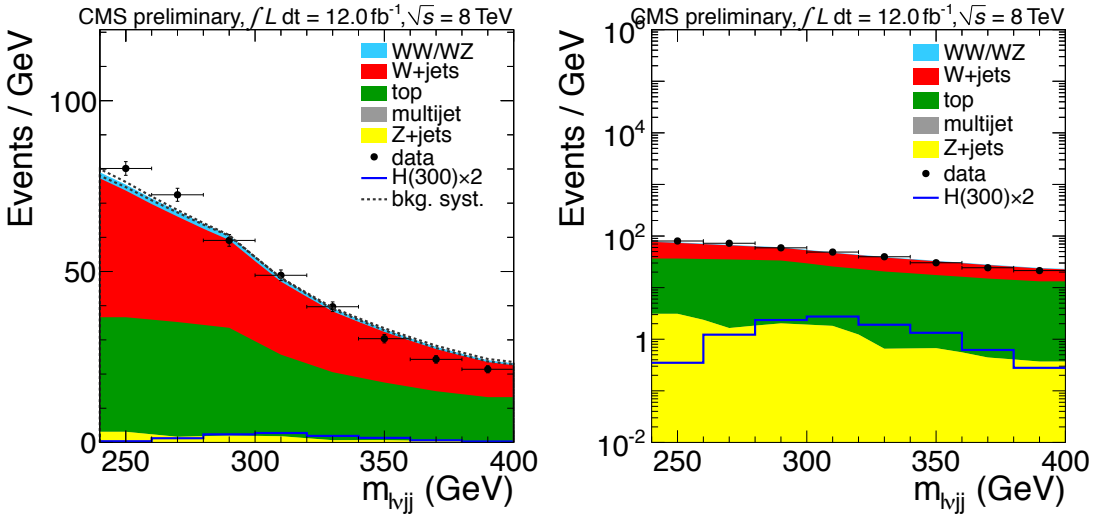


Figure 4.28: The distribution of the extrapolated background in the signal region is reported for the Higgs mass hypotheses of 300 GeV (top row) and 500 GeV (bottom row). The left and right columns display results for the muon 2jet and 3jet cases, respectively. The points represent the data-driven W+jets distribution, while the blue line shows the fitting function and the blue dashed lines correspond to the uncertainty on the shape.

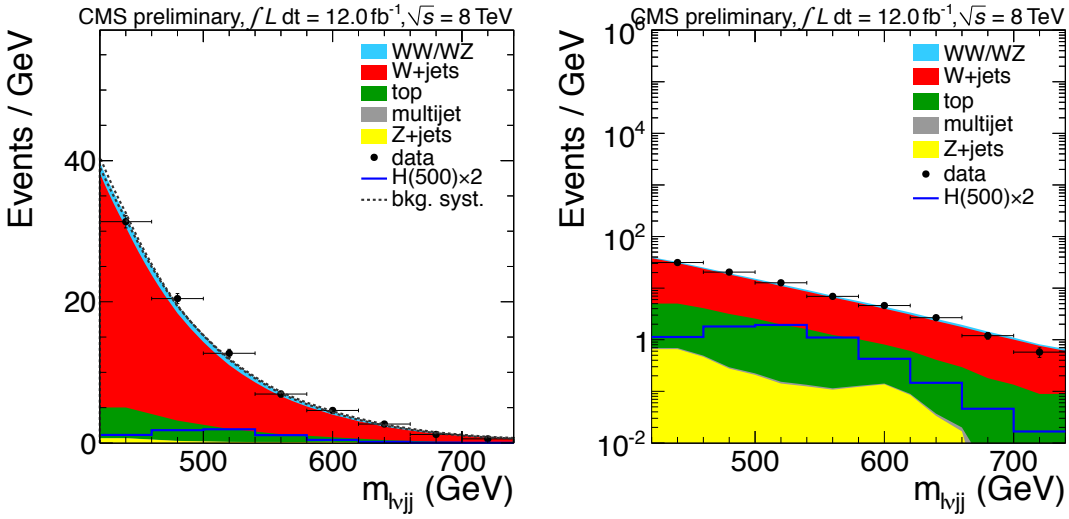


(a) muon 2 jets, 300 GeV Higgs

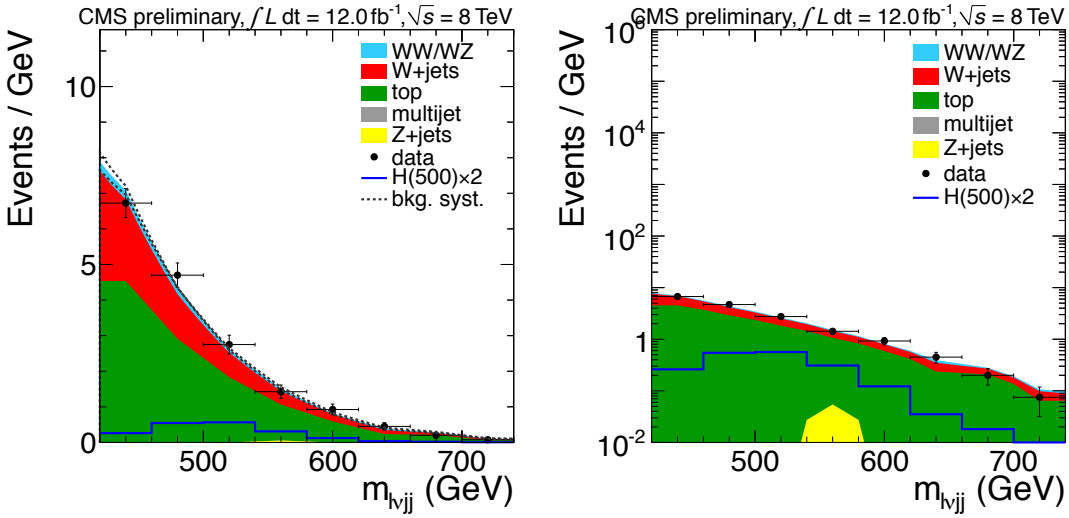


(b) muon 3 jets, 300 GeV Higgs

Figure 4.29: The four-body mass distribution of background events in the signal region for the mass point m_{H} =300 GeV/c^2. Data points are superimposed.



(a) muon 2 jets, 500 GeV Higgs



(b) muon 3 jets, 500 GeV Higgs

Figure 4.30: The four-body mass distribution of background events in the signal region for the mass point m_H =500 GeV/c^2. Data points are superimposed.

4.11 Systematics sources

Since the analysis is relying on data-driven approaches, systematics are affecting differently signal and background modeling.

Background normalization is obtained from a sideband data fit and subsequently $m_{\ell\nu jj}$ is used as final discriminating variable. Hence, only those systematics which have a significant impact on the m_{jj} and $m_{\ell\nu jj}$ distributions are effectively changing the relative background normalization and four-body mass modeling, and therefore the signal extraction. This happens, for example, with the jet energy scale, which is discussed in section 4.11.1.

The remaining effects on background are absorbed, in general, in the normalization uncertainty, as discussed in Section 4.11.2, and in the four-body mass shape uncertainty for the main background, i.e. W+jets (Section 4.11.3).

For the signal modeling, which is relying entirely on simulation, also other experimental effects and theoretical prediction uncertainties have to be taken into account, as it is described in Sections 4.11.5-4.11.10. On top of that, since the MVA selection efficiency ϵ_{MVA} is again taken purely from simulation, an additional uncertainty related to the limited knowledge of ϵ_{MVA} has to be added, as it is discussed in Section 4.11.4.

4.11.1 Jet energy scale

To evaluate the uncertainty due to jet energy scale in events with signal-like topology, the top data control sample is used. As already described, a semileptonic top sample is a good proxy for signal for the purposes of this study, since the top quark pairs are produced by gluon fusion and decay to two W bosons. The semileptonic top events are selected by requiring exactly four jets in the event, out of which two are b-tagged and the other two are anti-btagged. The hadronic W candidates are formed from the two anti-btagged jets. The invariant mass of the hadronic W candidates in the combined channels is shown in Figure 4.31, for data and Monte Carlo, together with a gaussian fit on the peak of each distribution. The relative difference between the gaussian means in data and Monte Carlo is used as jet energy scale uncertainty, and propagated through the template fits in the backgrounds, as well as to the signal shapes in the limit setting. These results for 2012 are consistent with those found for 2011, where the typical effect is of the order of less than a percent, as it is shown in Figure 4.32.

Additionally, the experimental effect on $E_{\mathrm{T}}^{\mathrm{miss}}$ related to the limited knowledge of the JES is taken into account, through the propagation of the JES uncertainties in the computation of the missing transverse energy, as it is detailed in Section 4.11.7.

4.11.2 Background normalization

The uncertainties on the total background normalization are derived from the unbinned maximum likelihood fit on the dijet invariant mass described in Section 4.9. The non-Poisson fractional errors for the 48 mass, lepton flavor, jet bin combinations are shown in Table 4.11. These are taken as a systematic uncertainty on the background normalization in the signal region.

The errors are computed using:

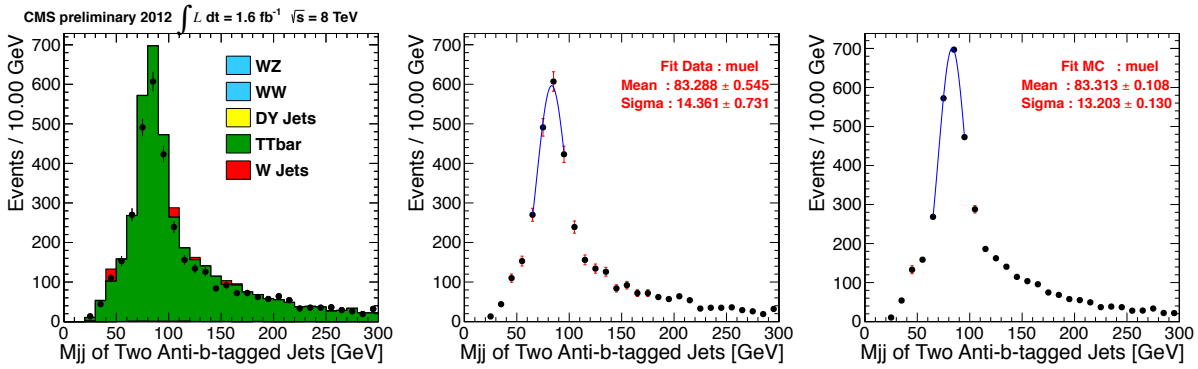


Figure 4.31: The invariant mass distribution of the hadronic W candidates in the semileptonic top sample (electron and muon combined). The left plot shows a good agreement between the data and MC. The distributions are fitted with a Gaussian to extract the peak location for the data (middle) and MC (right).

$$\text{non-Poisson fractional error} \equiv \frac{\sqrt{\sigma^2_{N_{\text{bkg}}} - N_{\text{bkg}}}}{N_{\text{bkg}}} \quad .$$

Poisson errors are included in the limit setting package and this additional systematic uncertainty contains the additional statistical errors derived in the dijet mass fit that are above and beyond the Poisson errors alone.

m_H (GeV)	electron 2-jet (%)	electron 3-jet (%)	muon 2-jet (%)	muon 3-jet (%)
170	0.2	0.3	0.2	0.2
180	0.5	0.3	0.5	0.2
190	0.3	0.3	0.3	0.2
200	0.6	0.4	0.4	0.3
250	0.3	0.4	0.3	0.3
300	0.3	0.5	0.3	0.7
350	0.9	0.8	2.5	1.3
400	0.4	0.8	0.5	0.8
450	0.6	2.9	0.7	1.8
500	0.9	1.9	1.0	1.7
550	0.9	4.1	1.2	1.9
600	1.6	0.9	1.3	1.6

Table 4.11: Systematic uncertainties on the total background normalization.

4.11.3 W+jets shape

The $m_{\ell\nu jj}$ shape for W+jets events is taken from the data sidebands. To get a smooth shape, this data-driven distributions are parametrized using an exponential function. The decay parameter of this parameterization has an associated uncertainty with it. The falling parameter is varied up and down to get shape variations on the W+jets four-body mass model. The shapes that are produced, corresponding to the different systematic variations on the parameters, are propagated to the limit setting as a systematic uncertainty.

The uncertainty on the α parameter, used to combine the two m_{jj} sidebands, is also contributing

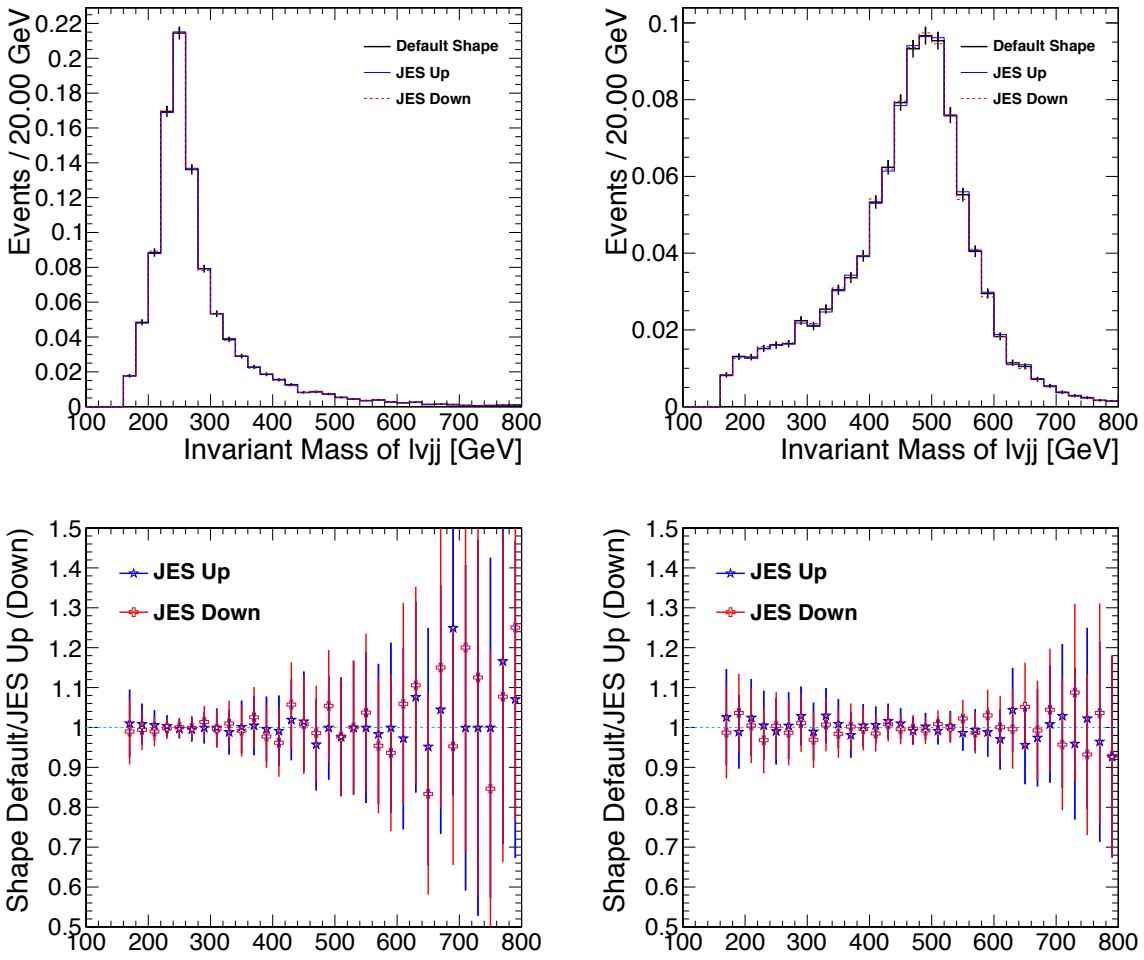


Figure 4.32: The Higgs signal shape comparison between the nominal shape and the shape by shifting JES up/down by its uncertainty. The left plots are for Higgs mass of 250 GeV and the right plots are for an Higgs mass of 500 GeV.

to a variation in shape. The errors on α are propagated to the W+jets shape and combined with the uncertainty on the shape that arises from other systematic uncertainties and from the limited statistical power of the sideband data samples.

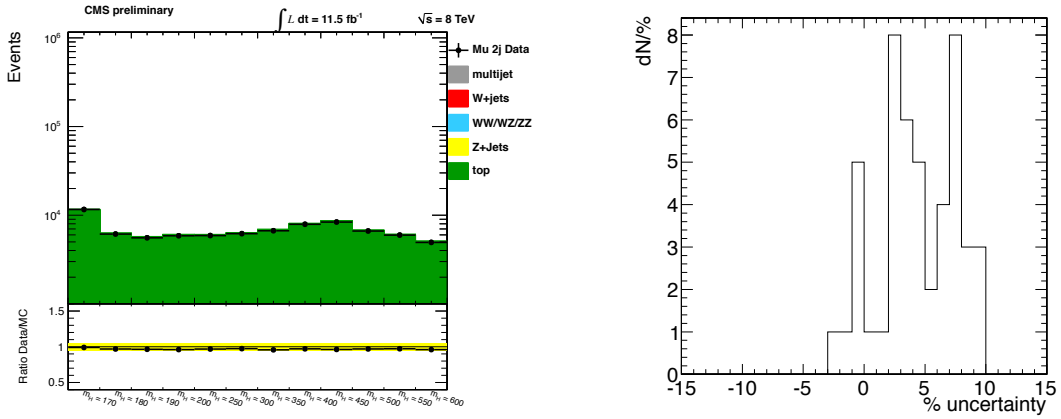
4.11.4 MVA selection efficiency

The systematic uncertainty associated with the efficiency on the final selection of the MVA output is studied by using the same top pair events as described above. There is reasonable agreement between the Monte Carlo and the data in the top control region, for the MVA output variable. The differences in selection efficiency are used to measure the potential uncertainty in the signal efficiency for each mass point and channel combination. The uncertainty is then taken as

$$100\% \times \left(1 - \frac{\epsilon_{data}}{\epsilon_{MC}}\right) \quad .$$

The distribution of measured uncertainties per mass point/channel combination is shown in

Figure 4.33. The measured efficiency uncertainties vary from less than 1% to 10%. Therefore, as a conservative choice, 10% is taken as the signal selection efficiency uncertainty for all channels and mass points.



(a) (b)

Figure 4.33: (a) The number of events passing the MVA cuts in the muon/2jets category, evaluated with the top control sample. (b) The distribution of measured uncertainties on signal selection efficiency, one entry per channel/mass point combination.

4.11.5 LHC luminosity

As other observables, luminosity measurement is affected by uncertainties, as is detailed here [70]. As a result, a luminosity uncertainty of 4.4% is applied to signal processes.

4.11.6 Lepton scale, selection and trigger efficiency

The lepton scale (and resolution) is varied according to the $Z \to e^+e^-$ fit uncertainty for electrons and to the nominal errors for the muons [72], and the MC samples are reweighted according to these variations. The reweighted samples two-body and four-body mass shapes are found to be compatible with the nominal ones, while the variation in the selection efficiency is found to be less than 1% for both the electron and muon 2jets/3jets categories.

In parallel, since all lepton efficiencies are measured with the tag-and-probe technique on $Z \to l^+l^-$ events, which is affected by uncertainties related to the limited statistics of the datasets, an additional systematic uncertainty has to be considered. The systematic uncertainties in the trigger efficiencies are of the order of 1%. The systematic uncertainties in the lepton reconstruction and identification efficiency scale factors are of the order of 2%. These uncertainties are summed in quadrature with the ones coming from the scale variation and accounted for in the final systematics on the signal model.

4.11.7 E_T^{miss} uncertainties

The E_T^{miss} reconstruction is mainly affected by the limited knowledge of the JES, as was already discussed in Section 4.11.1. In addition to JES effects, the E_T^{miss} distribution in the data is $\simeq$3% wider than the MC, and placing an hard cut on it creates an uncertainty on the signal yield. To estimate it, the E_T^{miss} is smeared for each event by a Gaussian with a $\sigma = 0.03 \times E_T^{miss}$ and the

number of signal events passing the cut are recomputed. As a result, the combined effect of the JES and smearing uncertainties on the $E_{\mathrm{T}}^{\mathrm{miss}}$, brings an additional 2% uncertainty on the signal yield.

Effects of the $E_{\mathrm{T}}^{\mathrm{miss}}$ smearing on the signal $\mathrm{l}\nu_{\mathrm{l}}\,\mathrm{jj}$ shape modeling are less important, since the four-body invariant mass distribution is obtained through a kinematic fit, which can recover most of the resolution degradation.

4.11.8 Pile-up model

The average number of pile-up interactions in a given bunch crossing is given by the following formula:

$$(N_{\mathrm{PU}})_i = \frac{\mathcal{L}_i \cdot \sigma_{\text{min. bias}}}{\nu_{\text{orbit}}} \quad , \tag{4.21}$$

where $\mathcal{L}$ is the instantaneous luminosity, $\sigma_{\text{min. bias}}$ is the cross-section of minimum bias interactions and ν_{orbit} is the LHC orbit frequency (11246 Hz). The uncertainty in the estimation of the number of pile-up interactions in data, comes from the uncertainty on the luminosity and the one on the minimum-bias cross-section. The combination in quadrature of these two, yields a total uncertainty of approximately 5% on the estimated number of interactions.

This 5% uncertainty in the number of interactions is propagated to the re-weighting procedure for signal samples, and the obtained variation in the signal yield is used as systematic error on the signal. The typical effect is less than a percent and therefore neglected.

4.11.9 Cross-section prediction

As of this writing, the inclusive cross-sections used for the Higgs signal at 7/8 TeV center-of-mass energy, which have been calculated by the Higgs Cross Section Working Group [21], come with an associated uncertainty caused by variations of the QCD normalization and factorization scales and the uncertainties of the PDF used for the calculation. The overall effect is of the order of 5-20%, depending on the production mode and the Higgs mass hypothesis.

In addition, the acceptance effect due to the PDF choice has been studied by following the PDF4LHC recipe [79], that considers as uncertainty the envelope of the errors calculated for three sets of PDFs, namely CT10, NNPDF and MSTW. Table 4.12 shows the values obtained. For the purposes of the limits calculation, these systematics are added in quadrature to the inclusive cross-section uncertainties.

Finally, there are uncertainties associated to the exclusive jet binning used in this analysis. A detailed description of the source of this uncertainty and how to calculate it is described in [80]. For this analysis, the numbers calculated by the $\mathrm{H} \to \mathrm{WW}^{(*)} \to \mathrm{l}\nu_{\mathrm{l}}\,\mathrm{l}\nu_{\mathrm{l}}$ group ([81], Section 8.1) are adopted.

4.11.10 Uncertainties related to interference

As was already summarized in Section 4.2.3, signal samples are corrected for the effect of interference between the signal and the SM $gg \to \mathrm{WW}$ production process. The procedure contains also a recipe for estimating the uncertainty band associated with the weight factors.

Table 4.12: Acceptance uncertainty related to the PDFs variation, for the signal rate, as a function of the mass hypothesis.

ggF		VBF	
m_H	unc.	m_H	unc.
170	2.0%	170	2.0%
180	2.0%	180	2.0%
190	2.0%	190	2.0%
200	2.0%	200	2.0%
250	1.5%	250	1.1%
300	2.0%	300	0.9%
350	2.2%	350	0.8%
400	2.4%	400	0.6%
450	2.7%	450	0.7%
500	2.9%	500	0.9%
550	3.2%	550	0.9%
600	3.6%	600	0.7%

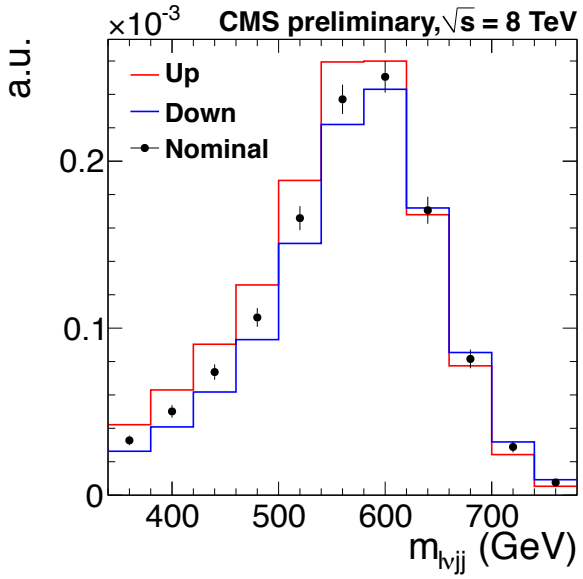


Figure 4.34: The shape variation for the signal $m_{\ell\nu jj}$ distribution, for a Higgs mass of 600 GeV/c^2, muon 2-jet category.

Since the effect of interference enhances the cross-section of the signal, this uncertainty band propagates through the analysis in two ways: it changes the yield of signal events passing the analysis selections, and it alters the shape of the signal invariant mass distribution that is used to set the upper limit. These effects are treated simultaneously by applying the up- and down-fluctuated weights to the signal distribution, in the same way that the nominal weights are applied. These alternative shapes are also propagated to the limit setter. Figure 4.34 shows an example of this variation for a Higgs mass of 600 GeV/c^2, for which, of all the mass hypotheses considered, the effect is largest.

4.11.11 Summary of systematic uncertainties

Table 4.13 summarizes the main systematic uncertainties, considered for the limit extraction, for the background and signal processes. Values are referring to the 8 TeV analysis.

Source of uncertainty	Magnitude
Background $m_{\ell\nu jj}$ shape	See Section 4.11.3
Background normalization	0.5-2%
Scale uncertainties from jet binning	4-28%
Higgs boson cross-section	13-15%
Signal shape (interference)	See Figure 4.34
Likelihood selection	10%
Luminosity	4.4%
Jet energy scale, resolution, and $E_{\mathrm{T}}^{\mathrm{miss}}$	2%
Lepton scale/selection/trigger	1-2%
Theory acceptances (PDF)	1-2%

Table 4.13: Sources of systematic uncertainties considered in the analysis, with the corresponding magnitude, for the limit extraction.

4.12 Extraction of the limit on the cross-section

No significant excesses are observed in data, hence an upper limit on the Higgs boson cross-section can be set, with a given confidence level.

Conventionally, in high energy physics, it is common to compute and show results corresponding to a confidence interval of 95%. In other words, data are used to exclude the presence of an Higgs boson with a cross-section greater than $\sigma_{95\%\,\mathrm{C.L.}}$ at 95% confidence level (yet leaving a 5% chance that data are compatible with a signal like Higgs particle with cross-section greater than $\sigma_{95\%\,\mathrm{C.L.}}$). $\sigma_{95\%\,\mathrm{C.L.}}$ is then compared with the Standard Model cross-section expectation σ_{SM}, and if the former is greater than the latter, then the Higgs boson is said to be excluded at 95% confidence level. This computation is then iterated for each Higgs mass hypothesis, and finally the Higgs boson can be quantitatively excluded in a given mass range.

The presented exclusion is valid on a SM-like Higgs object with width Γ_{SM} and decay rate $\mathcal{B}(\mathrm{H} \to \mathrm{W}^+\mathrm{W}^-)_{\mathrm{SM}}$. As a consequence, the results are valid even in a Beyond Standard Model context, where the new particle width is equal to Γ_{SM}, and its cross-section times branching ratio is equal to $\sigma_{\mathrm{SM}} \times \mathcal{B}(\mathrm{H} \to \mathrm{W}^+\mathrm{W}^-)_{\mathrm{SM}}$.

4.12.1 Statistical methodology for limit computation

For calculations of exclusion limits, the modified frequentist criterion CLs [82, 83] is adopted, which is briefly summarized here. First of all a likelihood function is defined as follows:

$$\mathcal{L}(\mathrm{data}|\mu,\theta) = \mathcal{P}(\mathrm{data}|\mu \cdot s(\theta) + b(\theta)) \cdot p(\theta|\tilde{\theta}) \quad . \tag{4.22}$$

The $s(\theta)$ stands for the signal expected under the SM Higgs hypothesis, $b(\theta)$ stands for backgrounds, μ is a signal strength modifier introduced to accommodate deviations from SM Higgs predictions, and θ are nuisance parameters describing systematic uncertainties of nominal

value $\tilde{\theta}$. The likelihood is in fact the Poisson probability[4] of observing a given dataset when the expected yield is $\mu \cdot s(\theta) + b(\theta)$, multiplied by the probability of measuring a value θ for the nuisance parameter with nominal value $\tilde{\theta}$.

The chosen test statistic q_μ, used to quantify the absence of signal, is based on the profile likelihood ratio, defined as:

$$q_\mu = -2\log\frac{\mathcal{L}(\text{data}|\mu,\hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\theta})} \quad , \qquad \text{with} \qquad 0 \leq \hat{\mu} \leq \mu \quad , \tag{4.23}$$

where the numerator and the denominator in Equation 4.23 are maximized, and the subscripted $\hat{\theta}_\mu$ indicates that the maximization of the likelihood in the numerator has been done with a fixed value of μ. In the maximization procedure, upper and lower constraints on the maximized $\hat{\mu}$ in the denominator force the limit to be one-sided and exclude the possibility of negative signal yields. Nuisance parameters are varied to best describe experimental data through a maximum-likelihood fit in the background-only and in the signal-plus-background hypothesis, from which $\hat{\theta}_0^{\text{obs}}$ and $\hat{\theta}_\mu^{\text{obs}}$ are obtained respectively. These values are used to generate pseudo-data in the b and $\mu \cdot s + b$ scenarios, where a signal strength μ is assumed for the latter. The pseudo-data of both hypotheses are then used to evaluate the test statistic distributions and obtain the pdf for q_μ in the case of no signal and a signal of strength μ. An example of these pdf obtained on CMS data is shown in Figure 4.35.

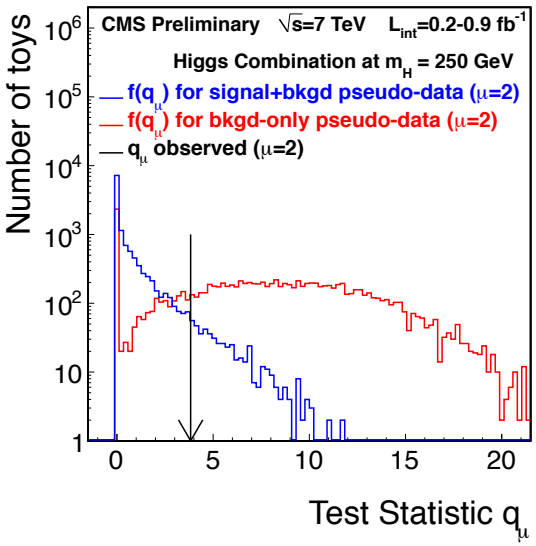


Figure 4.35: Frequentist test statistic q_μ distributions for ensembles of pseudo-data generated for background-only and signal+background hypotheses. The signal strength assumed in this example is $\mu = 2$. An example of the observed value of the test statistic is indicated by the arrow.

Once the q_μ pdfs are built, the test statistics is evaluated on data and two p-values, one for the

[4] $\mathcal{P}$ is in fact the product of the Poisson probabilities of all independent channels considered in the calculation (e.g. different final-state flavours, or the different bin content of a distribution for a shape-based limit extraction, as for this analysis).

signal-plus-background (p_μ) and one for the background-only hypothesis (p_0), are computed:

$$\begin{aligned} p_\mu \equiv \mathrm{CL_{s+b}} &= P(q_\mu \geq q_\mu^{\mathrm{obs}} | \mu \cdot s(\hat{\theta}_\mu^{\mathrm{obs}}) + b(\hat{\theta}_\mu^{\mathrm{obs}})) \\ p_0 \equiv \mathrm{CL_b} &= P(q_\mu \geq q_\mu^{\mathrm{obs}} | b(\hat{\theta}_0^{\mathrm{obs}})) \end{aligned} \quad . \tag{4.24}$$

From these p-values, $\mathrm{CL_s}$ (μ) can be calculated as

$$\mathrm{CL_s} = \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}} \quad . \tag{4.25}$$

If the $\mathrm{CL_s}(\mu) \leq \beta$, then a signal, with cross-section equal to μ times σ_{SM}, is excluded with a $1-\beta$ confidence level ($\beta = 0.05$ in this context). A particularly interesting example is when, for $\mu = 1$, $\mathrm{CL_s} \leq 0.05$, which is the mathematical translation of a SM Higgs boson exclusion at 95% C.L. .

Using the very same pdfs, the median-expected upper limit can be computed, alike the ± 1 and 2σ bands. Toy experiments are generated using the background model, and, for each of them, $\mathrm{CL_s}$ and $\mu_{95\%\,\mathrm{C.L.}}$ are calculated, so that a $\mu_{95\%\,\mathrm{C.L.}}$ probability distribution is computed. This distribution is then used to extract the median and the 16%(2.5%) and 84%(97.5%) quantiles.

4.12.2 Results

To set limits the full shape information of the $m_{\ell\nu jj}$ distribution is used. The four-body mass window is set roughly by the position and width of the signal distributions for the different Higgs mass points. All of the distributions are produced for each lepton flavor and jet bin, which represent independent channel inputs to the limit setter.

Recalling Equation 4.23, the numerator and denominator are built starting from these distributions. Each bin of the distributions is used to compute a likelihood term, and then, for each mass point, all these terms, $2 \times 2 \times \mathrm{N_{bins}}$ (number of lepton flavours times number of jet categories times number of $m_{\ell\nu jj}$ bins) are multiplied to obtain the global likelihood. The likelihoods is maximized in two different ways, i.e. with/without fixing the Higgs signal strength to a given value μ, so that the test statistics q_μ can be built. Finally, the p-values $\mathrm{CL_{s+b}}$ and $\mathrm{CL_b}$, are computed and the limit can be set.

The systematics described in Section 4.11 are treated as follows when being input to the limit setter:

- the main background systematics are the total background shape uncertainty and the total background normalization uncertainty. Both shape and normalization uncertainties are treated as uncorrelated across all channels, since they are derived from fits performed on independent samples;
- uncertainties on the signal deriving from lepton reconstruction and selection as well as trigger efficiency are treated as 100% correlated across the same-flavor lepton channels, but uncorrelated between electron and muon channels;
- uncertainties on the signal deriving from parton distribution functions (Table 4.12), luminosity, JES-$E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainty, pileup, interference, and theoretical cross-section uncertainty are treated as 100% correlated across all channels;

- uncertainties on the signal deriving from the final selection efficiency are treated as uncorrelated across different channels, but as correlated between quark-quark and glu-glu signal processes for the same channel.

The 95% exclusion limits are normalized to the SM expectation and plotted in Figure 4.36. The plots provide the exclusion range as the area where the given line (solid for the observation, dashed for the median expectation) lies below the value one.

From the analysis of 8 TeV data, the SM Higgs boson is excluded at 95% confidence level in the mass ranges 225–485 GeV/c^2 and 550–600 GeV/c^2, while the median expected exclusion range is 220–560 GeV/c^2. Combining with 7 TeV data, the SM Higgs boson is excluded in the mass ranges 215–490 GeV/c^2 and 525–600 GeV/c^2 at 95% confidence level, while the median expected one becomes 170–585 GeV/c^2.

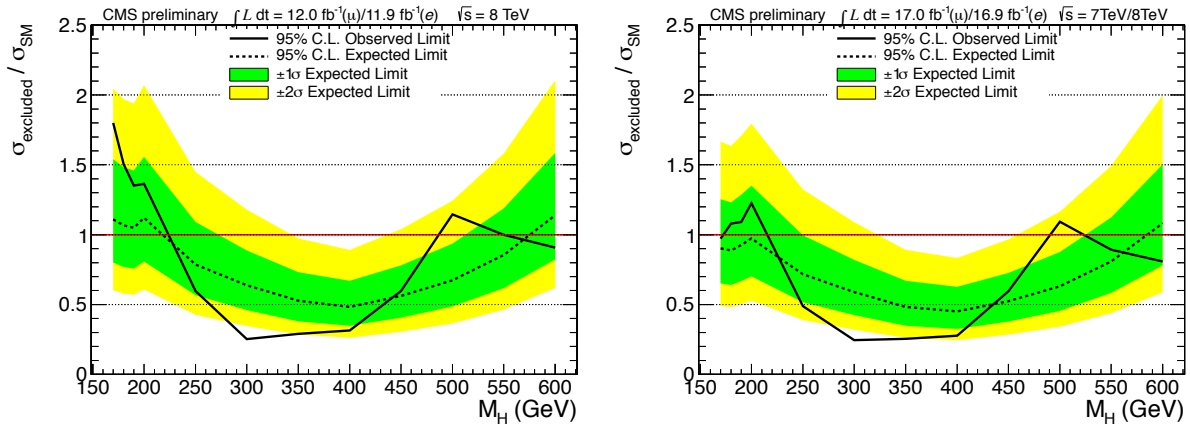


Figure 4.36: Observed (solid) and expected (dashed) 95% CL upper limit on the ratio of the production cross-section to the SM expectation for the Higgs boson obtained using the asymptotic CL_S technique. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid line at 1 indicates the SM expectation. The limit derived from 8 TeV data is shown on (a), while the combined limit using both 7 TeV and 8 TeV data is shown on (b).

The observed limit agrees, in general, with expectations in the null hypothesis, in other words confirming the lack of any statistical significant excesses. To better understand the compatibility, one can count the observed limit values in the different expected exclusion bands. In the combined limit plot (Figure 4.36b), there are 7 points falling in the 1-sigma band, 4 in the 2-sigma band, and 1 in the outer region. Assuming that the points are completely uncorrelated and that the dominant uncertainties are statistical, one would expect to observe 8 points in the 1-sigma band, 3 points in the 2-sigma band and 1 in the outer region. The fact that the downward fluctuation, observed around the Higgs mass hypothesis of 300 GeV/c^2, is wide, is a simple effect due to the correlation that is present among different points due to systematic uncertainties and, for neighboring mass hypothesis, additionally due to the selected events overlap, since the MVA selection is smoothly changing as m_H grows.

The results obtained with this channel are combined with other searches performed at CMS, namely the $H \to \gamma\gamma$, $H \to \tau^+\tau^-$, $H \to b\bar{b}$, $H \to ZZ$, as well as the $H \to W^+W^-$ fully leptonic, and the resulting combination is shown in Figure 4.37. The results show that the CMS

detector alone is able to exclude the Standard Model Higgs boson up to almost 700 GeV/c^2. The observed exclusion limit exhibits an excess in the low mass region, around 125 GeV/c^2, where a statistical significant signal has been found[5]. In the mid-high mass region, the observed curve lies around the expected limit calculated in the null hypothesis, thus excluding a SM-like boson particle existence, at 95% CL, up to 710 GeV/c^2. The contribution of the $H \to WW \to l\nu_l q\bar{q}$ search is particularly significant in this mass range, where, together with the $H \to WW^{(*)} \to l\nu_l l\nu_l$ and $H \to ZZ^{(*)} \to l^+l^- l^+l^-$ channels, it constitutes the driving contribution for the CMS Higgs boson search effort.

The analysis described in this chapter is also the subject of a public CMS document [84]. An additional result based on this very same final state has been published also by the ATLAS collaboration [85].

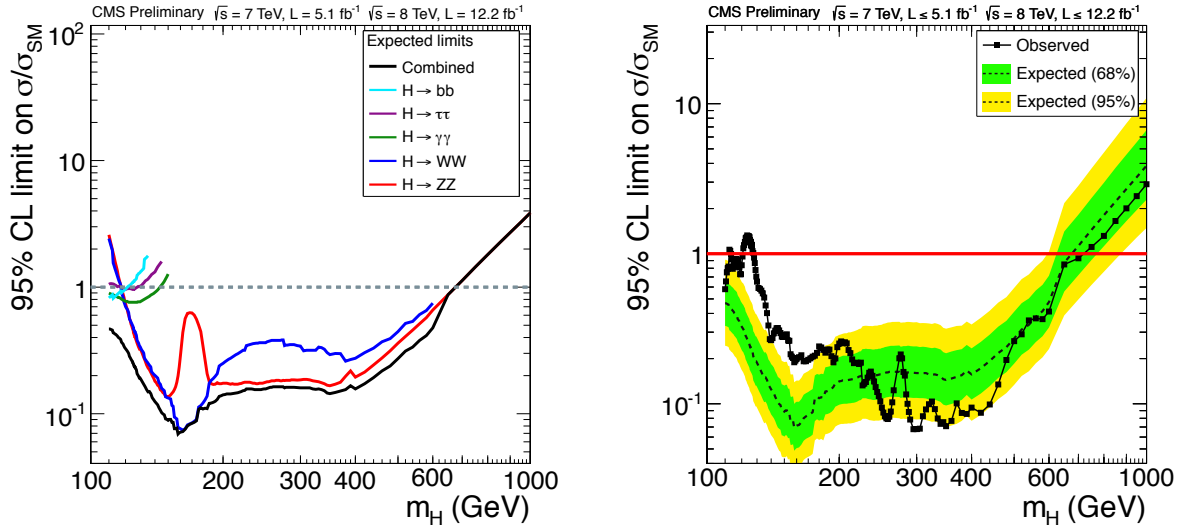


Figure 4.37: (a) The median expected 95% CL upper limits on the cross section ratio σ/σ_{SM} in the absence of a Higgs boson as a function of the SM Higgs boson mass in the range 110–1000 GeV/c^2, for the most relevant Higgs boson decay channels. (b) The corresponding combined limit, with the observed values superimposed (solid line). The dashed line indicates the expected median of results for the background only hypothesis, while the green (dark) and yellow (light) bands indicate, respectively, the ranges that are expected to contain 68% and 95% of all observed excursions from the median.

[5]The excess is in fact spreading above 125 GeV/c^2, in the low-mid mass region. This effect is driven by the $H \to WW^{(*)} \to l\nu_l l\nu_l$ analysis, which is characterized by a low mass resolution, hence sees the signal as a wide object covering the $[110, 160]$ GeV/c^2 range.

Chapter 5

Conclusions

The results of the studies that I have initiated and developed during my research doctorate have been presented in this manuscript. From a broad perspective, the project that I have worked on aimed at improving our knowledge of the electro-weak symmetry breaking mechanism, through the search of a scalar particle compatible with the one predicted by the Standard Model. Due to the peculiar phenomenology of the Higgs boson, such a search has to be performed in a wide mass range and in different final states. Therefore, the project I worked on consists of two main branches: a contribution to the CMS $H \to \gamma\gamma$ search, through the improvement of the experimental conditions relevant for the analysis, namely the calibration of the electromagnetic calorimeter, and the development of the CMS $H \to WW \to l\nu_l q\bar{q}$ analysis. These two searches are complementary, since the former covers the low mass range, and the latter the mid-high one, thus either of them is capable of the detecting the Higgs boson in case it exists. This capability has proven to be true on July the 4^{th} 2012, when the ATLAS and CMS experiments released an update of their results, driven by the $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to l^+l^-l^+l^-$ searches, which independently confirmed the presence of a new particle state, compatible with the Standard Model prediction.

As extensively explained in Chapter 3, the $H \to \gamma\gamma$ search is greatly affected by the photon energy resolution achieved by the CMS, which is ultimately depending on the electromagnetic calorimeter calibration conditions. As soon as the LHC started to provide proton-proton collision, the ECAL community developed several means to exploit data for calibration purposes. Among those collision events, a great deal of W/Z bosons can be found, and, with some simple requirements, they can be isolated from the remaining part of data. Firstly Monte Carlo studies, and subsequently the analysis of data, showed that with only 5 fb^{-1} of collected integrated luminosity at least one hundred good electron are hitting each of the 75,848 crystals that make up the calorimeter. In other words, the 2011 pp collision dataset alone provides an invaluable source of information to control the behavior of each of the ECAL crystals, since for the typical electron produced from a W/Z decay, relativity grants that its energy, measured with the ECAL, must be equal to its momentum, measured with the tracker. From this consideration, the E_{sc}/p_{tk} technique, used to equalize the crystal response, was developed and applied on the 2011 dataset. The introduction of this method, along with other improvements, substantially helped to improve the di-photon mass resolution: the effect on a 120 GeV/c^2 Higgs boson decaying into two photons is that the mass width goes from 2.4 GeV/c^2 to 1.82 GeV/c^2, thus

leading to an improvement in the $H \to \gamma\gamma$ signal sensitivity of 15%.

Subsequently, in Chapter 4, the motivations, the methods, and the results of the $H \to WW \to l\nu_l q\bar{q}$ search have been shown. The interest in this channel is driven by its features, namely its large cross-section times branching ratio - the largest among all experimentally accessible final states - and its final state closed kinematics, which grants the possibility to fully reconstruct the Higgs boson invariant mass. My research work on this topic started two years ago, when we first attacked the main experimental challenge of this analysis, which is to control the large W+jets background. Once the first Monte Carlo study was done, although not explicitly reported in this manuscript, I contributed to the deployment of the first CMS trigger dedicated to the $l\nu_l jj$ final state and to the development of a signal extraction method based on a peak hunt on a smoothly falling spectra, given by the $l\nu_l jj$ invariant mass [63]. Then, I started contributing to a multivariate analysis approach which significantly improved the signal over background ratio and which constitutes the core of this manuscript. From its beginning, this approach relied on data-driven means to keep track of the background processes contamination in the signal region. This has the important feature of avoiding relying on the simulation for the main background, W+jets, which is poorly modeled in the phase space where the signal over background ratio is good, and additionally reduces the effect of systematic uncertainties on the background estimation. Conversely, the Higgs events modeling is coming purely from the Monte Carlo simulation, and therefore all systematic uncertainties affecting the signal have been deeply investigated. All data collected in 2011 and 2012 have been then analyzed with this method. From their statistical analysis, and from the absence of a significant excess of events observed in data, an upper limit on the Higgs boson cross section was placed: the SM Higgs boson has been excluded in the mass ranges 215–490 GeV/c^2 and 525–600 GeV/c^2 at 95% confidence level. A part from this important result, the $H \to WW \to l\nu_l q\bar{q}$ search has been particularly significant in the CMS global channel combination, where, together with the $H \to WW^{(*)} \to l\nu_l l\nu_l$ and $H \to ZZ^{(*)} \to l^+l^- l^+l^-$ channels, constituted the driving contribution for the CMS Higgs boson exclusion in the medium-high mass range, which today covers the 128-710 GeV/c^2 interval.

List of acronyms

The following table contains some relevant acronyms, widely used in this manuscript. The acronyms are reported according to their alphabetical ordering, from top to bottom, left to right.

Acronym	Full phrase or word	Acronym	Full phrase or word
ATLAS	A Toroidal LHC Apparatus	LO	Leading Order
APD	Avalanche Photo-Diodes	MC	Monte Carlo
BDT	Boosted Decision Tree	MET	Missing Transverse Energy
CDF	Collider Detector at Fermilab	MVA	Multi-Variate Analysis
CERN	European Organization for Nuclear Research	NDF	Number of Degrees of Freedom
CMS	Compact Muon Solenoid	NLO	Next to Leading Order
CSC	Cathode Strip Chambers	NNLO	Next-to-Next to Leading Order
DAQ	Data Acquisition system	PDF	Parton Distribution Functions
DT	Drift Tubes	PF	Particle-flow event reconstruction
ECAL	Electromagnetic Calorimeter	PS	Proton Synchrotron
EB	ECAL Barrel	PU	Pile-Up
EE	ECAL Endcaps	PV	Primary Vertex
EWK	ElectroWeak	QCD	Quantum ChromoDynamics
GSF	Gaussian Sum Filter	QFT	Quantum Field Theory
HB	Hadron Barrel	RGE	Renormalisation Group Equation
HE	Hadron Endcaps	RPC	Resistive Plate Chambers
HCAL	Hadronic CALorimeter	SLC	Stanford Linear Collider
HLT	High Level Trigger	SM	Standard Model
HO	Hadron Outer	SPS	Super Proton Synchrotron
HVF	Hadron Very Forward	SSB	Spontaneous Symmetry Breaking mechanism
IC	Intercalibration Constants	TEC	Tracker End Caps
ID	Identification	TIB	Tracker Inner Barrel
JES	Jet Energy Scale	TID	Tracker Inner Disks
L1	Level-1 Trigger system	TOB	Tracker Outer Barrel
LEP	Large Electron-Positron collider	VEV	Vacuum Expectation Value
LHC	Large Hadron Collider	VPT	Vacuum PhotoTriodes

Bibliography

[1] D. Griffiths, "Introduction to Elementary Particles". John Wiley & Sons, New York, USA, 1987. 1

[2] M. E. Peskin and D. V. Schroeder, "An Introduction To Quantum Field Theory (Frontiers in Physics)". Westview Press, 1995. 1

[3] F. Halzen and A. D. Martin, "Quarks and Leptons". Wiley, 1984. 1

[4] J. D. Bjorken and S. D. Drell, "Relativistic Quantum Fields". McGraw-Hill, 1965. 1

[5] H. Goldstein, "Classical Mechanics". Addison-Wesley Publishing Company, Reading, MA, 2nd edition, 1980. 2

[6] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **86** (Jul, 2012) 010001. doi:10.1103/PhysRevD.86.010001. 6

[7] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries", *Phys.Rev.* **127** (1962) 965–970. doi:10.1103/PhysRev.127.965. 6

[8] B. W. Lee, C. Quigg, and H. B. Thacker, "Weak interactions at very high energies: The role of the Higgs-boson mass", *Phys. Rev. D* **16** (Sep, 1977) 1519–1531. doi:10.1103/PhysRevD.16.1519. 13

[9] J. Ellis, J. Espinosa, G. Giudice et al., "The probable fate of the Standard Model", *Physics Letters B* **679** (2009), no. 4, 369 – 375. doi:10.1016/j.physletb.2009.07.054. 13

[10] LEP Working Group for Higgs boson searches, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration Collaboration, "Search for the standard model Higgs boson at LEP", *Phys.Lett.* **B565** (2003) 61–75, arXiv:hep-ex/0306033. doi:10.1016/S0370-2693(03)00614-2. 14, 15

[11] CDF Collaboration, DØ Collaboration Collaboration, "Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 6.7 fb^{-1} of Data", arXiv:1007.4587. 14, 15

[12] CDF Collaboration, DØ Collaboration, "Combined CDF and DØ Search for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data", arXiv:1203.3774. 16

[13] S. Dawson, "Radiative corrections to Higgs boson production", *Nucl.Phys.* **B359** (1991) 283–300. doi:10.1016/0550-3213(91)90061-2. 18

[14] A. Djouadi, M. Spira, and P. Zerwas, "Production of Higgs bosons in proton colliders: QCD corrections", *Phys.Lett.* **B264** (1991) 440–446. doi:10.1016/0370-2693(91)90375-Z. 18

[15] M. Spira, A. Djouadi, D. Graudenz et al., "Higgs boson production at the LHC", *Nucl.Phys.* **B453** (1995) 17–82, arXiv:hep-ph/9504378. doi:10.1016/0550-3213(95)00379-7. 18

[16] S. Catani, D. de Florian, and M. Grazzini, "Higgs production in hadron collisions: Soft and virtual QCD corrections at NNLO", *JHEP* **0105** (2001) 025, arXiv:hep-ph/0102227. 18

[17] M. Ciccolini, A. Denner, and S. Dittmaier, "Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC", *Phys.Rev.* **D77** (2008) 013002, arXiv:0710.4749. doi:10.1103/PhysRevD.77.013002. 18

[18] M. Zaro, P. Bolzoni, F. Maltoni et al., "Higgs production via vector-boson fusion at NNLO in QCD", *PoS* **DIS2010** (2010) 211. 18

[19] T. Han and S. Willenbrock, "QCD correction to the pp $\to$ WH and ZH total cross-sections", *Phys.Lett.* **B273** (1991) 167–172. doi:10.1016/0370-2693(91)90572-8. 18

[20] W. Beenakker, S. Dittmaier, M. Kramer et al., "NLO QCD corrections to t anti-t H production in hadron collisions", *Nucl.Phys.* **B653** (2003) 151–203, arXiv:hep-ph/0211352. doi:10.1016/S0550-3213(03)00044-0. 18

[21] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", arXiv:1101.0593. 18, 77, 117

[22] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension", *Comput.Phys.Commun.* **108** (1998) 56–74, arXiv:hep-ph/9704448. doi:10.1016/S0010-4655(97)00123-9. 20

[23] ATLAS Collaboration Collaboration, "Combined search for the Standard Model Higgs boson in pp collisions at $\sqrt{s}$=7 TeV with the ATLAS detector", *Phys. Rev. D* **86** (Aug, 2012) 032003. doi:10.1103/PhysRevD.86.032003. 22

[24] CMS Collaboration Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at", *Physics Letters B* **710** (2012), no. 1, 26 – 48. doi:10.1016/j.physletb.2012.02.064. 22

[25] CMS Collaboration Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* **716** (2012), no. 1, 30 – 61. doi:10.1016/j.physletb.2012.08.021. 23

[26] ATLAS Collaboration Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B* **716** (2012), no. 1, 1 – 29. doi:10.1016/j.physletb.2012.08.020. 23

[27] H. Nilles, "SUPERSYMMETRY, SUPERGRAVITY AND PARTICLE PHYSICS", *PHYSICS REPORTS* **110** (1984) 1–162. 23

[28] G. L. K. H. E. Haber, "The search for supersymmetry: Probing physics beyond the standard model", *Physics Reports* **117** (1985) 75–263. 23

[29] O. S. Bruning, P. Collier, P. Lebrun et al., "LHC Design Report". CERN, Geneva, 2004. 25

[30] ATLAS Collaboration, "ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN", *LHC Tech. Proposal* **CERN-LHCC-94-43** (1994). 25

[31] CMS collaboration Collaboration, "CMS Physics Technical Design Report Volume I: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items. 25

[32] CMS collaboration Collaboration, "CMS physics Technical Design Report, Volume II: Physics Performance . oai:cds.cern.ch:942733", *J. Phys. G* **34** (2006), no. CERN-LHCC-2006-021. CMS-TDR-008-2, 995–1579. 669 p. revised version submitted on 2006-09-22 17:44:47. 25

[33] LHCb Collaboration, "LHCb: Technical Proposal", *Tech. Proposal* **CERN-LHCC-98-004** (1998). 25

[34] ALICE Collaboration, "ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC", *LHC Tech. Proposal* **CERN-LHCC-95-71** (1995). 25

[35] CMS Collaboration, "The CMS tracker system project: Technical Design Report", *Technical Design Report CMS* **CMS-TDR-005** (1997). 29, 59

[36] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", *Technical Design Report CMS* **CMS-TDR-004** (1997). 30, 31, 45

[37] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004. 31

[38] CMS Collaboration, "The CMS hadron calorimeter project : Technical Design Report", *Technical Design Report CMS* **CMS-TDR-002** (1997). 32

[39] CMS Collaboration, "The CMS muon project : Technical Design Report", *Technical Design Report CMS* **CMS-TDR-003** (1997). 33

[40] CMS Collaboration, "CMS TriDAS project : Technical Design Report; 1, the trigger systems", *Technical Design Report CMS* **CMS-TDR-006-** (2000). 34

[41] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", *CMS Physics Analysis Summaries* **PFT-10-001** (2010). 35

[42] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", *CMS Physics Analysis Summaries* **PFT-10-002** (2010). 35

[43] CMS Collaboration, "CMS Physics Technical Design Report Volume I : Detector Performance and Software", *Technical Design Report CMS* **CMS-TDR-008-1** (2006). 36, 39, 41, 80

[44] E. Meschi, T. Monteiro, C. Seez et al., "Electron Reconstruction in the CMS Electromagnetic Calorimeter", *CMS Notes* **CMS-NOTE-2001-034** (2001). 38

[45] H. Bethe and W. Heitler, "On the Stopping of fast particles and on the creation of positive electrons", *Proc.Roy.Soc.Lond.* **A146** (1934) 83–112. 39

[46] W. Adam, R. Fruhwirth, A. Strandlie et al., "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC", *eConf* **C0303241** (2003) TULT009, arXiv:physics/0306087. doi:10.1088/0954-3899/31/9/N01, 10.1088/0954-3899/31/9/N01. 39

[47] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k_T jet clustering algorithm", *JHEP* **0804** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063. 42

[48] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys.Lett.* **B659** (2008) 119–126, arXiv:0707.1378. doi:10.1016/j.physletb.2007.09.077. 42

[49] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET",. 43

[50] L. Di Matteo, "Calibration of the CMS Electromagnetic Calorimeter with $\sqrt{s} = 7$ TeV LHC collisions data", Technical Report CMS-CR-2011-290, CERN, Geneva, (Nov, 2011). 46, 67, 69

[51] M. Anfreville, D. Bailleux, J. Bard et al., "Laser monitoring system for the CMS lead tungstate crystal calorimeter", *Nucl.Instrum.Meth.* **A594** (2008) 292–320. doi:10.1016/j.nima.2008.01.104. 55

[52] J. M. Campbell, J. Huston, and W. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics", *Rept.Prog.Phys.* **70** (2007) 89, arXiv:hep-ph/0611148. doi:10.1088/0034-4885/70/1/R02. 74

[53] M. Botje, J. Butterworth, A. Cooper-Sarkar et al., "The PDF4LHC Working Group Interim Recommendations", arXiv:1101.0538. 75

[54] A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, "Parton distributions for the LHC", arXiv:0901.0002. 75

[55] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, P. Nadolsky, W.K. Tung, "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis", arXiv:hep-ph/0201195. 75

[56] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, Maria Ubiali, "A first unbiased global NLO determination of parton distributions and their uncertainties", arXiv:1002.4407. 75

[57] G. Passarino, "Higgs Interference Effects in $gg \to ZZ$ and their Uncertainty", arXiv:1206.3824. 75, 81

[58] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **0411** (2004) 040, arXiv:hep-ph/0409146. doi:10.1088/1126-6708/2004/11/040. 80

[59] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **0711** (2007) 070, arXiv:hep-ph/0709.2092. doi:10.1088/1126-6708/2007/11/070. 80

[60] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **1006** (2010) 043, arXiv:hep-ph/1002.2581. doi:10.1007/JHEP06(2010)043. 80

[61] P. Nason and C. Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG", *JHEP* **1002** (2010) 037, arXiv:hep-ph/0911.5299. doi:10.1007/JHEP02(2010)037. 80

[62] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **0605** (2006) 026, arXiv:hep-ph/0603175. doi:10.1088/1126-6708/2006/05/026. 80

[63] CMS Collaboration, "Search for the Standard Model Higgs boson in the $H \to WW \to l\nu_l jj$ decay channel", *CMS Physics Analysis Summaries* **HIG-12-003** (2011). 80, 126

[64] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, Tim Stelzer, "MadGraph 5 : Going Beyond", arXiv:1106.0522. 80

[65] CMS Collaboration, "Measurement of the underlying event in the Drell-Yan process in proton-proton collisions at $\sqrt{s} = 7$ TeV", arXiv:1204.1411. 81

[66] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl.Instrum.Meth.* **A506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8. 81

[67] N. Kauer and G. Passarino, "Inadequacy of zero-width approximation for a light Higgs boson signal", *JHEP* **1208** (2012) 116, arXiv:1206.4803. doi:10.1007/JHEP08(2012)116. 81

[68] S. Goria, G. Passarino, and D. Rosco, "The Higgs Boson Lineshape", *Nucl.Phys.* **B864** (2012) 530–579, arXiv:1112.5517. doi:10.1016/j.nuclphysb.2012.07.006. 81

[69] R. K. E. John M Campbell, "MCFM for the Tevatron and the LHC", *Nucl.Phys.Proc.* (2010) arXiv:1007.3492. doi:10.1016/j.nuclphysbps.2010.08.011. 82

[70] CMS Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update"", *CMS Physics Analysis Summaries* **SMP-12-008** (2012). 83, 116

[71] CMS Collaboration, "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1110** (2011) 132, arXiv:1107.4789. 89

[72] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7TeV$"", *CMS Physics Analysis Summaries* **MUO-10-002** (2010). 90, 116

[73] CMS Collaboration, "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **2** (2011) 2–40, arXiv:hep-ph/10122466. doi:10.1007/JHEP01(2011)080. 96

[74] J. D'Hondt, S. Lowette, O. L. Buchm?ller et al., "Fitting of Event Topologies with External Kinematic Constraints in CMS", *CMS Notes* **CMS-NOTE-2006-023** (2006). 98

[75] TMVA Collaboration, "TMVA Users Guide", *Sourceforge* **CERN-OPEN-2007/007** (2009). 104

[76] Y. Gao, A. V. Gritsan, Z. Guo et al., "Spin determination of single-produced resonances at hadron colliders", *Phys.Rev.* **D81** (2010) 075022, arXiv:1001.3396. doi:10.1103/PhysRevD.81.075022. 104

[77] B. A. Dobrescu and J. D. Lykken, "Semileptonic decays of the standard Higgs boson", *JHEP* **1004** (2010) 083, arXiv:0912.3543. doi:10.1007/JHEP04(2010)083. 104

[78] CMS Collaboration, "Measurement of the sum of WW and WZ production with W+dijet events in pp collisions at $\sqrt{s} = 7$ TeV", arXiv:1210.7544. 108

[79] M. R. Whalley, D. Bourilkov, and R. C. Group, "The Les Houches Accord PDFs (LHAPDF) and Lhaglue", arXiv:hep-ph/0508110. 117

[80] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in Summer 2011", *ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005* (2011). 117

[81] L. Bauerdick et al., "A Higgs Boson Search in the Fully Leptonic W^+W^- (update for ICHEP2012 conference)", *CMS AN* **AN-12-228** (2012). 117

[82] T. Junk, "Confidence level computation for combining searches with small statistics", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **434** (1999), no. 23, 435 – 443. doi:10.1016/S0168-9002(99)00498-2. 119

[83] A. L. Read, "Presentation of search results: the CL s technique", *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10, 2693. 119

[84] "Search for the Standard Model Higgs boson in the H to WW to lnujj decay channel in pp collisions at the LHC",. 123

[85] ATLAS Collaboration, "Search for the Higgs boson in the H->WW->lnujj decay channel at $\sqrt{s} = 7$ TeV with the ATLAS detector", arXiv:1206.6074. 123