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

Higgs boson search in the $H \to WW \to l\nu_l q\bar{q}$ final state with the CMS detector at the LHC

Coordinatore della Scuola di Dottorato: prof. **Giuseppe Chirico**
Relatore italiano: prof. **Stefano Ragazzi**
Relatore francese: prof. **Yves Sirois**

Tesi di dottorato di
Andrea Davide Benaglia
Matricola: 055881

Ciclo XXIV
Anno Accademico 2011-2012

ÉCOLE POLYTECHNIQUE

Thèse présentée pour obtenir le grade de

Docteur en Sciences

par

Andrea Davide Benaglia

Higgs boson search in the $H \to WW \to l\nu_l\, q\overline{q}$ final state with the CMS detector at the LHC

Soutenue le 11/07/2012 devant le jury composé de :

M. Besançon	Président du Jury (CEA, Saclay)
G. Dissertori	Examinateur (ETH, Zürich)
C. Dionisi	Rapporteur (Università La Sapienza, Roma)
L. Moroni	Rapporteur (INFN, Milano-Bicocca)
S. Ragazzi	Directeur de thèse (Università di Milano-Bicocca)
Y. Sirois	Directeur de thèse (École Polytechnique, Palaiseau)

Summary

This thesis presents a three-year work in the CMS experiment, in the context of the first LHC proton-proton collisions. In particular, my studies concentrated on Higgs boson searches. This particle, whose existence is predicted by the Standard Model, has not been observed yet, and constitutes the main Physics target of many high-energy colliders. In the recent past, the LEP and Tevatron experiments allowed to exclude, at 95% confidence level, the existence of the Higgs boson for masses below 114 GeV/c^2 and in the $[147, 179]$ GeV/c^2 mass range. Therefore, at the moment the subject of this thesis was being decided, the very low and very high mass ranges emerged as the most interesting and promising regions.

With collision data collected in 2011 I have studied the response stability and uniformity of the electromagnetic calorimeter (ECAL). Isolated electrons from W boson decays into electron-neutrino pairs have been used to characterize the ECAL response (local uniformity corrections, crystal transparency corrections, ageing of the readout channels). These studies are of great relevance for Higgs boson searches in the di-photon decay channel, which is the reference analysis for low-mass Higgs hypotheses and demands a supreme energy resolution to profit from the very narrow Higgs resonance.

The other experimentally-allowed Higgs mass range has been tackled with the Physics analysis channel described in this thesis. In particular, I have studied the decay channel $H \to WW \to l\nu_l q\bar{q}$, for Higgs mass hypotheses well above the $2m_W$ threshold. This channel has the largest effective cross section ($\sigma \times \mathcal{B}$) for the Higgs boson, although the final state is contaminated by large background events from Standard Model sources. A complete analysis strategy has been defined, optimizing the selections that allow to enhance the signal presence in the final set of events. I have studied the performances of lepton reconstruction and identification and contributed to the development and characterization of a trigger, specifically designed for this channel. Finally, all sources of systematic uncertainties that affect the statistical interpretation of the result have been addressed. With no significant excess observed on the expected background yield, an upper limit on the Higgs boson production cross-section has been settled, for Higgs masses between about 320 and 400 GeV/c^2.

Résumé

Cette thèse présente le travail de trois ans effectué dans l'expérience CMS, dans le contexte des premières collisions proton-proton du LHC. L'étude porte en particulier sur la recherche du boson de Higgs. Cette particule, dont l'existence est prévue par le Modèle Standard, n'a encore jamais été observée, mais elle est toujours activement l'objet de recherches auprès de collisionneurs à très hautes énergies. Dans le passé récent, les expériences au LEP et Tevatron ont permis d'exclure l'existence du Higgs dans les domaines de masse $< 114\,\mathrm{GeV}/c^2$ et $[147, 179]\,\mathrm{GeV}/c^2$, avec un intervalle de confiance de 95%. Au moment où le sujet de cette thèse a été décidé, les domaines de masse très faible et très élevé étaient donc les régions les plus intéressantes.

Avec les données de collisions de 2011, je me suis occupé de la question de la stabilité et uniformité de la réponse du calorimètre électromagnétique (ECAL). Des électrons isolés, provenant de la désintégration de boson W en en voie électron-neutrino ont été utilisé pour caractériser la réponse du ECAL (corrections locales d'uniformité, corrections de transparence des cristaux, vieillissement des cellules de lecture). Ceci sert de référence pour l'analyse du Boson du Higgs se désintégrant en deux photons, qui est le canal favori pour des hypothèses de masse faible et qui exige une une résolution en énergie optimale pour mieux profiter de la résonance trés étroite du boson du Higgs pour ces masses.

L'analyse de physique effectuée dans cette thèse porte sur la recherche de bosons de Higgs dans l'autre région de masse encore permise par les contraintes expérimentales. En particulier, j'ai étudié le canal de désintégration $\mathrm{H} \to \mathrm{WW} \to \mathrm{l}\nu_\mathrm{l}\,\mathrm{q}\bar{\mathrm{q}}$ pour des masses du Higgs $m_\mathrm{H} > 2m_\mathrm{W}$. Ce canal est celui qui offre la plus grande section efficace de production ($\sigma \times \mathcal{B}$) pour le boson de Higgs, même s'il est pénalisé par des bruits de fonds de processus standards importants. Une stratégie d'analyse complète a été définie: j'ai étudié les performances pour la physique de la reconstruction et de l'identification des leptons isolés, fortement contribué à la définition et à la caractérisation du déclenchement (trigger) pour ce canal, et procédé à une analyse détaillée des sources d'incertitudes systématiques affectants l'interprétation statistique des résultats. En absence d'une déviation par rapport à l'attendue des bruits de fonds du modèle standard, des limites supérieures sur la masse du boson de Higgs ont été obtenu, entre 320 et $400\,\mathrm{GeV}/c^2$ environ.

Contents

Chapter 1

Introduction

In this chapter the main theoretical framework used for the studies in this thesis is presented. At first, the Standard Model of elementary particles is reviewed in Section 1.1. The electroweak symmetry breaking as originating spontaneously via the Brout-Englert-Higgs mechanism is briefly discussed. The existence of a scalar field with non-zero value of the ground state in the vacuum is considered responsible for the weak boson masses and at the origin of the elementary fermion masses.
The mass of the Higgs boson, the quantum associated to the Higgs field, is not predicted by the theory. The current knowledge on the Higgs boson mass, from both theoretical considerations and from direct or indirect experimental constraints is presented in Section 1.2.
An overview of the Higgs production and decay mechanisms at the LHC as well as the main search strategies is presented in Section 1.3.
Finally, the Section 1.4 presents an overview of the following chapters and the motivations for the studies conducted in this work.

1.1 The Standard Model

1.1.1 The non-interacting theory

The theory that summarizes the current experimental knowledge of elementary particles and their interactions is called Standard Model (SM) [1, 2, 3]. It is a relativistic quantum field theory based on the group of symmetries $SU(3)_C \times SU(2)_L \times U(1)_Y$. These three local symmetry groups dictate the three different type of interactions between the particles in the Standard Model. The theory is perturbative at sufficiently high energies and renormalizable.

According to the Standard Model, the matter constituents are spin-$\frac{1}{2}$ particles called fermions. The fundamental fermions observed up to today in experiments are subdivided in leptons and quarks. These two groups of 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 labelled as up (u), down (d), charm (c), strange (s), top (t) 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.

	1^{st} gen.		2^{nd} gen.		3^{rd} gen.		Q
leptons	ν_e	~ 0	ν_μ	~ 0	ν_τ	~ 0	0
	e	$511\,\mathrm{keV}/c^2$	μ	$105.7\,\mathrm{MeV}/c^2$	τ	$1.777\,\mathrm{GeV}/c^2$	-1
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
	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

Table 1.1: Spin-$\frac{1}{2}$ matter constituents.

The quantum field operators associated to fermions are 4-components Dirac spinors ψ. For a free fermion of mass m the Lagrangian and the associated equation of motion (Dirac equation) are

$$\mathcal{L}_{\mathrm{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \qquad \Rightarrow \qquad \left(i\gamma^\mu\partial_\mu - m\right)\psi = 0 \quad . \tag{1.1}$$

The adjoint spinor $\bar{\psi} = \psi^\dagger\gamma^0$ has been introduced. For further convenience, we also introduce the Weyl spinor representation[1], which allows to write

$$\psi = \psi_L + \psi_R = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \chi_R \end{pmatrix} \quad . \tag{1.2}$$

The two-components objects ψ_L and ψ_R are referred to as left-handed and right-handed Weyl spinors respectively, obtained from the spinor ψ via the projection operators $P_L = \frac{1}{2}\left(1-\gamma^5\right)$ and $P_R = \frac{1}{2}\left(1+\gamma^5\right)$. The left or right-handedness of spinors is called chirality.

Interactions between fermions are mediated through the exchange of spin-1 particles, called bosons, which arise from invariances of the theory under so-called gauge symmetries (see. Section 1.1.2). Three types of fundamental interactions between fermions have been observed[2].

The $SU(2)_L \times U(1)_Y$ group is associated to electroweak interactions, which are the unified description of electromagnetism and weak interactions. The long-range electromagnetic interaction is mediated by the massless photon, while the short-range weak force carriers are the massive W^+, W^- and Z^0 bosons. The $SU(2)_L$ gauge bosons (i.e. $W^\pm$) couple only to the left-handed components ψ_L of the fermion fields, while the $U(1)_Y$ gauge bosons (i.e. Z^0 and γ) couple to both ψ_L and ψ_R, thus leading to the observed parity-violation phenomena of weak interactions.

[1] Here the adopted Weyl representation of the γ matrices corresponds to $\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}$ and $\gamma^5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$.

[2] The fourth known fundamental interaction, gravity, is much weaker than the other three and cannot be accommodated yet into an unified theory together with electromagnetic, weak and strong forces.

The left-handed projections of the fermion fields form $SU(2)_L$ doublets

$$f_L = \left\{ \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L \right\} \quad , \tag{1.3}$$

while the right-handed components are $SU(2)_L$ singlets:

$$f_R = \{e_R,\ \mu_R,\ \tau_R,\ u_R,\ d_R,\ c_R,\ s_R,\ t_R,\ b_R\} \quad . \tag{1.4}$$

To each doublet it is associated a so-called weak isospin charge $T = \frac{1}{2}$: neutrinos and up-like quarks possess a third isospin component $T_3 = \frac{1}{2}$, whereas leptons and down-like quarks exhibit a $T_3 = -\frac{1}{2}$. The absence of right-handed neutrinos will be discussed afterwards.

The $SU(3)_C$ symmetry group is related to the strong interaction between quarks, which is governed by QCD. Each quark appears in three different colour states, thus belonging to a $SU(3)_C$ triplet, while leptons are colourless singlets. The gluons, which are the quanta of the strong interaction field, have zero mass and carry colour charge. They can interact among each other, which leads to the already mentioned phenomenon of quark confinement, restricting the strong force to nuclear distances.

A free bosonic force carrier of mass m and spin 0 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:

$$\mathcal{L}_{\text{KG}} = \left(\partial_\mu \phi\right)^\dagger \left(\partial^\mu \phi\right) - m^2 \phi^\dagger \phi \qquad \Rightarrow \qquad \left(\Box + m^2\right) \phi = 0 \quad . \tag{1.5}$$

For vector (i.e. spin 1) bosons, the associated operator is a vector field A_μ in the four dimensional Minkowski time-space, whose dynamics is described by

$$\mathcal{L}_{\text{Proca}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} m^2 A_\mu A^\mu \qquad \Rightarrow \qquad \left(\Box + m^2\right) A_\mu = 0 \quad . \tag{1.6}$$

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 explained in the following section.

1.1.2 Particle interactions: the gauge invariance principle

Since Maxwell's unification of electric and magnetic interaction, the concept of gauge invariance has played a strategic role in the understanding and the description of the fundamental forces of Nature. The requirement of a symmetry in the Lagrangian of a theory accounts for conservation of charges, via the Noether's theorem, and allows for the introduction of new fields and interactions in the theory. From another point of view, the existence of a symmetry implies that some quantity is unmeasurable. For example, translation invariance means that we cannot determine an absolute position in space.

In QFT it proves very convenient to require the Lagrangian invariance under gauge transforma-

tions, i.e. under an internal phase transformation of the form $\psi \to e^{i\alpha}\psi$. If the (unmeasurable) phase α is constant in time and space, we speak of "global gauge", whereas if we allow the phase to differ from point to point, $\alpha = \alpha(x)$, we deal with "local gauge" transformations.

Let us consider for instance a local gauge transformation of the form

$$\psi \to \psi' = U\psi = e^{i\alpha(x)}\psi \tag{1.7}$$

into some internal group G of generators T^a, with $a = 1, \ldots, n$. The group algebra is defined by the structure constants $f^{abc} = [T^a, T^b]$. The phase can be expressed as $\alpha(x) = \varepsilon^a(x)T^a$, with $\varepsilon^a(x)$ being the rotation parameter. The quantum-mechanical observables, which depend only on $|\psi|^2$ are invariant under Equation 1.7, whereas a Lagrangian such as the one in Equation 1.1 in general is not. This is due to the extra term $\partial_\mu\alpha(x)$ in the derivative transformation:

$$\partial_\mu\psi \to \partial_\mu\psi' = e^{i\alpha(x)}\partial_\mu\psi + i\psi\partial_\mu\alpha(x) \quad . \tag{1.8}$$

A possible way to make the theory manifestly invariant under the gauged symmetry is to introduce a set of new vector fields A^a_μ and replace the usual time-space derivative ∂_μ with the so-called covariant derivative, defined as

$$\mathcal{D}_\mu \doteq \partial_\mu - igT^a A^a_\mu \quad , \tag{1.9}$$

with an arbitrary parameter g that will determine the interaction strength of the field. Substituting the covariant derivative into Equation 1.1, the Lagrangian reads

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi - m\bar{\psi}\psi = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - ig\bar{\psi}\gamma^\mu T^a A^a_\mu\psi \quad , \tag{1.10}$$

where the last term expresses the coupling between the fermion field and the new vector fields. The substitution of the standard with the covariant derivative allows the Lagrangian to be invariant under the U transformation, provided that the new A^a_μ vector fields transform under U so as to exactly compensate the extra term in Equation 1.10. By demanding that $\mathcal{D}'_\mu\psi' = U(\mathcal{D}_\mu\psi)$, the following transformation laws for the vector fields A^a_μ can be derived:

$$A_\mu \to A'_\mu = UA_\mu U^\dagger - \frac{i}{g}(\partial_\mu U)U^\dagger \quad . \tag{1.11}$$

Finally, in order to give these gauge fields a kinetic term, a tensor $F_{\mu\nu}$ has to be introduced, which must be antisymmetric in its two spatial indexes. It is natural to define $-igF^a_{\mu\nu}T^a = [\mathcal{D}_\mu, \mathcal{D}_\nu]$, or more explicitly

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu \quad , \tag{1.12}$$

which preserves the local gauge invariance.

As an example, we can consider the case where $U = e^{i\alpha(x)}$ represents a $U(1)$ phase abelian transformation: the covariant derivative is given by $\mathcal{D}_\mu = \partial_\mu - igA_\mu$ and the gauge field transformation law is simply given by $A'_\mu = A_\mu + \frac{1}{g}\partial_\mu\alpha(x)$. The kinetic tensor has the form $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. We notice that no mass term of the form $\frac{1}{2}m_A^2 A^\mu A_\mu$ can be added without spoiling the symmetry, because of the A_μ transformation law: the gauge boson is intrinsically

massless. This is nothing but the quantum field description of the electromagnetic interaction (QED), with the boson A_μ identified as the photon.

Theories with a local non-abelian phase invariance are also possible and go under the name of Yang-Mills theories. To describe the experimental knowledge of the particles and their interactions at the quantum level, two such symmetries, together with an abelian symmetry, are necessary and sufficient. First of all, the Lagrangian exhibits a local $U(1)$ phase invariance. The gauge field associated to it is called 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 ($i = 1, 2, 3$), one for each of the generators $\tau^i/2$. 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 transformation is then given by

$$U = e^{i\left(\beta(x)\frac{Y}{2}+\alpha^i(x)\frac{\tau^i}{2}+\gamma^a(x)\frac{\lambda^a}{2}\right)} \quad , \tag{1.13}$$

and the covariant derivative which ensures all the three invariances of the theory takes the form

$$\mathcal{D}_\mu = \partial_\mu - ig'\frac{Y}{2}B_\mu - ig\frac{\tau^i}{2}W^i_\mu - ig_s\frac{\lambda^a}{2}G^a_\mu \quad , \tag{1.14}$$

where the scalar Y and the matrices[3] τ^i and λ^i 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:

- $SU(3)_C$: only quarks have colour charge, and appear as colour triplets under $SU(3)$ transformations. Other leptons transform as colour singlets;
- $SU(2)_L$: recalling the chiral decomposition into Weyl spinors (1.2), 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 0 weak-isospin, and therefore do not interact with gauge bosons. This chiral nature of the weak isospin transformations has an immediate consequence. Fermion mass terms in the Lagrangian are written as

  $$-m\bar{\psi}\psi = -m\bar{\psi}\left(P_L + P_R\right)\psi = -m\left(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L\right) \quad , \tag{1.15}$$

  which manifestly violates gauge invariance, since ψ_L is a member of an isospin doublet and ψ_R is a singlet. At this point, fermion mass terms must be excluded from the theory;
- $U(1)_Y$: the $U(1)$ hypercharge induce transformations as singlets and is non-zero for all fermions except for the right-handed neutrinos. As a convention the corresponding quantum number is chosen $Y = -1$ for left-handed leptons. Since right-handed neutrinos do not couple to any of the introduced interactions, they are sterile and do not form a part of the theory.

[3] By τ^i we indicate the set of 2×2 complex Hermitian and unitary matrices called Pauli matrices. The λ^a are the Gell-Mann traceless and Hermitian matrices.

Restricting to the sole electroweak sector, the Lagrangian must include kinetic terms for the gauge fields, which look like

$$-\frac{1}{4}W^i_{\mu\nu}W^{\mu\nu}_i - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{1.16}$$

and given the $SU(2)$ algebra, we can write

$$\begin{aligned} W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - ig\epsilon_{ijk}W^j_\mu W^k_\nu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \quad , \end{aligned} \tag{1.17}$$

From the above equation for $W_{\mu\nu}$, self-interaction terms among the gauge bosons are visible, due to the non-Abelian character of $SU(2)$ gauge symmetry.

Recalling Equation 1.11, the following relations can be derived, expressing the transformation law for the vector gauge fields:

$$\begin{aligned} B_\mu \to B'_\mu &= B_\mu + \frac{1}{g'}\partial_\mu\beta(x) \\ \vec{W}_\mu \to \vec{W}'_\mu &= \vec{W}_\mu + \frac{1}{g}\partial_\mu\vec{\alpha}(x) - \vec{\alpha}(x)\times\vec{W}_\mu \quad . \end{aligned} \tag{1.18}$$

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 to the electromagnetic and weak forces respectively. Instead, the observed interactions are a manifestation of the combined $SU(2)_L \times U(1)_Y$ gauge group, where the physical fields A_μ, Z_μ and $W^\pm_\mu$, for respectively the photon, the Z boson and the $W^\pm$ bosons, arise as combinations of the gauge fields according to

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right) \\ \begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} &= \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix} \quad , \end{aligned} \tag{1.19}$$

where θ_W is the weak mixing angle (Weinberg angle), measured[4] $\sin^2\theta_W = 0.23153 \pm 0.00016$. The $SU(2)_L$ and $U(1)_Y$ groups cannot therefore be considered separately, since the two components of doublets have different electric charge. The relation between electric charge, hypercharge and weak isospin is given by the Gell-Mann-Nishijima formula: $Q = T_3 + \frac{Y}{2}$.

Up to this point, not only are the fermions forced to be massless, but also gauge bosons mass terms are not allowed if the local gauge symmetry has to be preserved. The transformation of the gauge fields (see Equation 1.18) does not allow for an explicit term $\propto \frac{1}{2}W^i_\mu W^\mu_i$ or $\frac{1}{2}B_\mu B^\mu$. A possible solution of 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 breakdown of the symmetry.

[4]This is actually the value of $\sin^2\theta^{\text{lept}}_{\text{eff}}$, the effective Weinberg angle from Z-pole measurements in leptonic final states [4].

1.1.3 The origin of mass: the Higgs mechanism

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. A canonical example of a spontaneously broken symmetry is that of a ferromagnetic system. Above the Curie temperature T_C, the system shows a $SO(3)$ rotational symmetry, with all the dipoles randomly oriented in the three-dimensional space, yielding a null overall magnetization. For $T < T_C$ the configuration of minimum energy is reached when all the dipoles are aligned in some arbitrary direction (spontaneous magnetization) and the rotational symmetry is hidden. The system ground state then chooses a particular configuration among the infinite possible, but once a ground state is assumed, it cannot be changed, unless an infinite amount of energy is introduced into the system to re-orient the dipoles in a different direction.

An important consequence of the spontaneous symmetry breaking in QFT is the appearance of massless and spinless particles, when the original symmetry is continuous. This affirmation is in fact the result of a theorem, which goes under the names of Nambu and Goldston [5], and the newly appeared scalar fields are referred to as Goldstone bosons. The number of Goldstone bosons of the broken theory coincides to the number of continuous symmetries which are broken by the choice of a specific ground state.

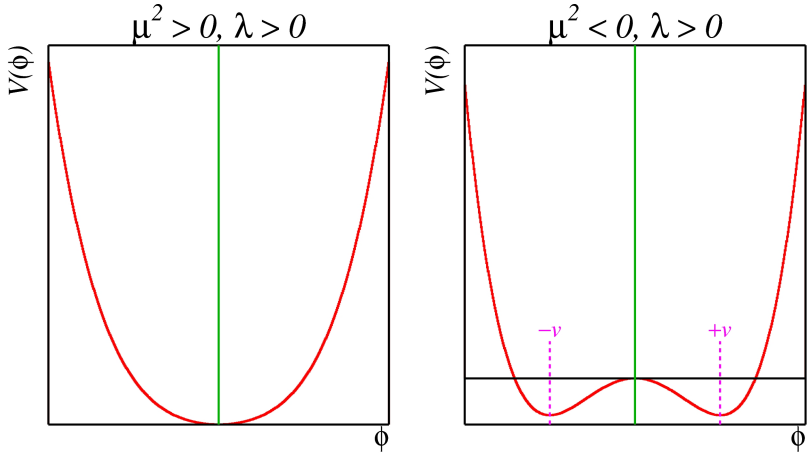


Figure 1.1: Form of the potential $V(\phi)$ depending on the sign of μ^2, positive on the left and negative on the right.

In the Standard Model, one needs an external field to break the electroweak gauge symmetry: it is called the Higgs field [6, 7, 8]. In order to generate masses for the three gauge bosons $W^{\pm}$ and Z^0, without generating a photon mass, at least three degrees of freedom are needed. The simplest realization is to add a complex $SU(2)$ doublet of scalar fields of hypercharge $Y = 1$:

$$\mathbf{\Phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad . \tag{1.20}$$

It has no colour charge and will therefore not affect the $SU(3)_C$ sector. The Lagrangian for the

Higgs field is given by

$$\mathcal{L}_{\text{Higgs}} = \left(\mathcal{D}^{\mu}\mathbf{\Phi}\right)^{\dagger}\left(\mathcal{D}_{\mu}\mathbf{\Phi}\right) - V(\mathbf{\Phi}) \qquad \text{with} \qquad V(\mathbf{\Phi}) = \mu^2\mathbf{\Phi}^{\dagger}\mathbf{\Phi} + \lambda\left(\mathbf{\Phi}^{\dagger}\mathbf{\Phi}\right)^2 \quad , \tag{1.21}$$

which is manifestly invariant under $SU(2)_L \times U(1)_Y$. The expected form of the potential is sketched in Figure 1.1: for $\mu^2 > 0$ the scalar potential has a global minimum at $\langle 0|\mathbf{\Phi}|0\rangle = \Phi_0 = 0$, which would not break the electroweak gauge symmetry. For $\mu^2 < 0$ the potential has a circle of degenerate minima at

$$\langle 0|\mathbf{\Phi}|0\rangle = \Phi_0 = -\frac{\mu^2}{2\lambda} = \frac{1}{2}v^2 \qquad \text{where} \qquad v \doteq \sqrt{-\frac{\mu^2}{\lambda}} \quad , \tag{1.22}$$

with v the vacuum expectation value of the field $\mathbf{\Phi}$ (VEV). The spontaneous breaking of the $SU(2)$ symmetry consists in choosing a particular ground state, around which the Higgs field $\mathbf{\Phi}(x)$ is expanded. The particular vacuum chosen is

$$\mathbf{\Phi}_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \quad . \tag{1.23}$$

It is clear that neither T_i nor Y cancels $\boldsymbol{\phi}_0$, 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

$$Q\mathbf{\Phi}_0 = \left(T_3 + \frac{Y}{2}\right)\mathbf{\Phi}_0 = 0 \qquad \Rightarrow \qquad \mathbf{\Phi}_0 \to \mathbf{\Phi}_0' = e^{i\delta(x)\frac{Q}{2}}\mathbf{\Phi}_0 = \mathbf{\Phi}_0 \quad . \tag{1.24}$$

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 $SU(2)_L \times U(1)_Y \to U(1)_Q$.

If we label 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 $-\theta_3$ we can write [5]

$$\begin{aligned}
\mathbf{\Phi}(x) &= \frac{1}{\sqrt{2}}\begin{pmatrix} \theta_2(x) + i\theta_1(x) \\ v + H(x) - i\theta_3(x) \end{pmatrix} \\
&\simeq \frac{1}{\sqrt{2}}\begin{pmatrix} 1 + i\theta_3/v & i(\theta_1 - i\theta_2)/v \\ i(\theta_1 + i\theta_2)/v & 1 - i\theta_3/v \end{pmatrix}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \\
&\simeq \frac{1}{\sqrt{2}}e^{i\frac{2\theta^i(x)}{v}\frac{\tau^i}{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad .
\end{aligned} \tag{1.25}$$

Thanks to the $SU(2)$ invariance of the Lagrangian, the three fields $\theta_i(x)$ in Equation 1.25 can be gauged away with a transformation $U = \exp\left(-i\frac{2\theta^i(x)}{v}\frac{\tau^i}{2}\right)$: these are the massless Goldstone bosons, which do not explicitly appear in the final Lagrangian. By expanding the scalar Higgs

[5] In the following expression we only consider small fluctuations around the minimum.

field Lagrangian 1.21 around $\mathbf{\Phi_0}$ using

$$\mathbf{\Phi}(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.26}$$

one finds [9]:

$$\begin{aligned} \mathcal{L}_{\text{Higgs}} = & \frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2} 2v^2 \lambda H^2 - \frac{1}{3!} 6v\lambda H^3 - \frac{1}{4!} 6\lambda H^4 \\ & + \frac{1}{2}\frac{v^2 g^2}{4} W_\mu^{-\dagger} W^{-\mu} + \frac{1}{2}\frac{v^2 g^2}{4} W_\mu^{+\dagger} W^{+\mu} \\ & + \frac{1}{2}\frac{v^2(g^2+g'^2)}{4} \left(\frac{g W_\mu^3 - g' B^\mu}{\sqrt{g^2+g'^2}} \right)^2 + 0 \cdot \left(\frac{g' W_\mu^3 + g B^\mu}{\sqrt{g^2+g'^2}} \right)^2 \\ & + \frac{1}{4}\left(2vH + H^2\right) \left[g^2 W_\mu^- W^{+\mu} + \frac{1}{2}(g^2+g'^2) \left(\frac{W_\mu^3 - g' B_\mu}{\sqrt{g^2+g'^2}} \right)^2 \right] \end{aligned} \tag{1.27}$$

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

In the second line, coming from the kinetic term $\left(\mathcal{D}_\mu \mathbf{\Phi}\right)^\dagger \left(\mathcal{D}^\mu \boldsymbol{\phi}\right)$, the $\text{W}^\pm$ 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 right mass terms for the observed Z^0 and γ vector bosons [6]. The first linear combination of the gauge fields W_μ^3 and B_μ comes with an appropriate mass term and 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.
We can therefore interpret the results as

$$\begin{aligned} m_{\text{W}} &= \frac{1}{2}vg && \text{with} && W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \\ m_{\text{Z}} &= \frac{1}{2}v\sqrt{g^2+g'^2} && \text{with} && Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}} \\ m_\gamma &= 0 && \text{with} && A_\mu = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2+g'^2}} \end{aligned} \tag{1.28}$$

The gauge bosons have "eaten" the three massless Goldstone bosons, and have acquired mass. The degrees of freedom of the Goldstone bosons are in fact needed by the three gauge bosons:

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

once they become massive, an additional degree of freedom is required in order to allow them to have a longitudinal polarization. We also notice that the unbroken $U(1)_Q$ symmetry causes the photon to remain massless.
The mixing of W^3_μ and B_μ yielding the physical force carriers can be interpreted as a rotation of parameter θ_W, where we have identified

$$\frac{g}{\sqrt{g^2+g'^2}} = \cos\theta_W \qquad \text{and} \qquad \frac{g'}{\sqrt{g^2+g'^2}} = \sin\theta_W \tag{1.29}$$

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

$$m_Z = \frac{m_W}{\cos\theta_W} \tag{1.30}$$

Finally, in the last line of the Lagrangian 1.27, the cubic and quartic couplings of the Higgs boson to the weak gauge bosons can be deduced. We notice in particular the coupling of one single Higgs boson to a pair of W or Z bosons to be proportional to m_W and m_Z respectively:

$$\begin{aligned} g_{HWW} &= g m_W \\ g_{HZZ} &= \frac{g}{2\cos\theta_W} m_Z \quad . \end{aligned} \tag{1.31}$$

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):

$$\frac{\mathcal{B}(H\to WW)}{\mathcal{B}(H\to ZZ)} = \left(\frac{g_{HWW}}{g_{HZZ}}\right)^2 = 4\cos^2\theta_W \frac{m_W^2}{m_Z^2} \simeq 2.7 \quad . \tag{1.32}$$

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

$$\mathcal{L}_{SM} = \mathcal{L}_{GWS} + \mathcal{L}_{Higgs} \quad , \tag{1.33}$$

where the electroweak part of it (representing the Glashow-Weinberg-Salam model of electroweak unification) is given by

$$\begin{aligned} \mathcal{L}_{GWS} = &-\frac{1}{4} W^i_{\mu\nu} W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \\ &+ i\bar{\nu}_L \gamma^\mu \partial_\mu \nu_L + i\bar{e}_L \gamma^\mu \partial_\mu e_L + i\bar{e}_R \gamma^\mu \partial_\mu e_R + \\ &- i\bar{f}_L \gamma^\mu \left(-ig\frac{\tau^i}{2} W^i - ig'\frac{Y}{2} B_\mu\right) f_L - i\bar{e}_R \gamma^\mu \left(-g'\frac{Y}{2} B_\mu\right) e_R \quad . \end{aligned} \tag{1.34}$$

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

$$\begin{aligned} \mathcal{L}^{int}_{GWS} &= \mathcal{L}^{int}_{CC} + \mathcal{L}^{int}_{NC} \\ &= \left\{ e J^{em}_\mu A^\mu + \frac{g}{\cos\theta_W} J^Z_\mu Z^\mu \right\} + \left\{ \frac{g}{\sqrt{2}} \left(J^+_\mu W^{+\mu} + J^-_\mu W^{-\mu} \right) \right\} \end{aligned} \tag{1.35}$$

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:

$$\begin{aligned} J_\mu^{em} &= Q\bar{f}\gamma_\mu f\,, \\ J_\mu^Z &= \frac{1}{2}\bar{f}\gamma_\mu\left(c_V^f - c_A^f\gamma_5\right) f \qquad \text{with} \qquad c_V^f = T_3 - 2Q\sin^2\theta_W,\ c_A^f = T_3\,, \\ J_\mu^+ &= \frac{1}{2}\bar{\nu}\gamma_\mu\left(1-\gamma_5\right) e \quad . \end{aligned} \tag{1.36}$$

1.1.3.1 Fermion masses

An attractive feature of the Standard Model is that the same Higgs doublet which generates 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):

$$\mathcal{L}_{\text{Yukawa}}^{\ell} = -G_\ell\left[(\bar{\boldsymbol{\ell}}_L\boldsymbol{\Phi})\ell_R + \bar{\ell}_R(\boldsymbol{\Phi}^\dagger\boldsymbol{\ell}_L)\right] \quad , \tag{1.37}$$

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, inserting Equation 1.26 into 1.37, one obtains

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}}^{\ell} &= -\frac{G_\ell}{\sqrt{2}}\left\{(\bar{\nu}_\ell,\bar{\ell})_L\begin{pmatrix}0\\ v+H\end{pmatrix}\ell_R + \bar{\ell}_R\,(0, v+H)\begin{pmatrix}\nu_\ell\\ \ell\end{pmatrix}_L\right\} \\ &= -\frac{G_\ell}{\sqrt{2}}\left\{v(\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L) + (\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L)H\right\} \\ &= -\frac{G_\ell}{\sqrt{2}}\left\{v\bar{\ell}\ell + \bar{\ell}\ell H\right\} \quad . \end{aligned} \tag{1.38}$$

In Equation 1.38, we choose G_ℓ (called Yukawa coupling) so as to generate the required lepton mass term, that is

$$m_\ell = \frac{G_\ell v}{\sqrt{2}} \tag{1.39}$$

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

$$\mathcal{L}_{\text{Yukawa}}^{\ell} = -m_\ell\left\{\bar{\ell}\ell + \frac{1}{v}\bar{\ell}\ell H\right\} \quad . \tag{1.40}$$

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 a future experiment, the amplitude of a Higgs decay process will be proportional to the second power of the mass of the particle the Higgs decays into (see also Equation 1.31).

1.2 The search for the Higgs boson

Some details about the current constraints on the Higgs boson mass are given here, driven both by theoretical assumptions and direct experimental searches.

1.2.1 Theoretical constraints on the Higgs mass

Although the Higgs boson mass is not predicted by the theory, both lower and upper limits can be set, based on theoretical backgrounds [10]. A first upper constraint is found considering the weak boson scattering process $W_L W_L \to W_L W_L$. In a scenario where no Higgs boson actually exists, the amplitude for such a process would be proportional to the center-of-mass energy, and thus violate unitarity at high energy, which happens at $\sqrt{s} \simeq 1.2\,\text{TeV}$. Diagrams involving the exchange of a Higgs boson among the W_L allow for cancellations so that the scattering amplitude is regularized and finite at all energies, provided that

$$m_H \lesssim 700\,\text{GeV}/c^2 \quad \text{(vector boson scattering)} \quad . \tag{1.41}$$

More restricting bounds on the Higgs mass can be set as a function of a cut-off energy scale Λ, up to which we assume the validity of the Standard Model (i.e. the scale up to which no new interactions and particles are expected). These bounds are derived from the renormalization group equation for the Higgs quartic coupling λ, which describes the evolution of the parameter with energy.

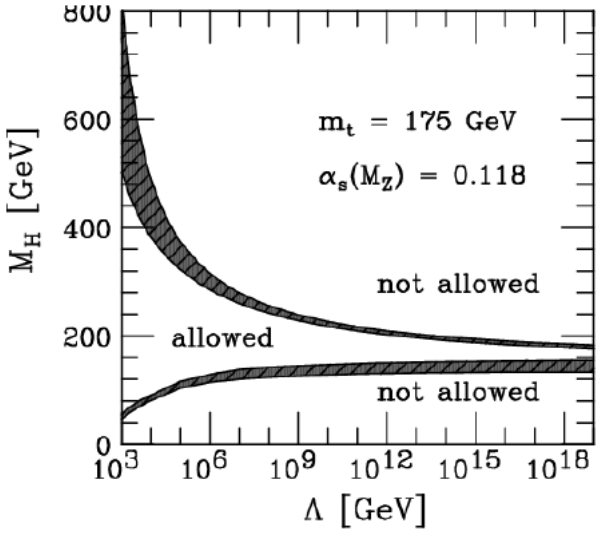


Figure 1.2: Upper and lower theoretical limits on the Higgs mass as a function of the energy scale Λ up to which the Standard Model is assumed to hold. The shaded area indicates the theoretical uncertainties in the calculation of the bounds. Here the values $m_t = 175\,\text{GeV}/c^2$ and $\alpha_S = 0.118$ have been used.

First of all, the Higgs potential described in Equation 1.21 is affected by radiative corrections from Higgs couplings to vector bosons and fermions (mostly top loops). These corrections may modify the shape of the potential in a way such that an absolute minimum no longer exists and no stable spontaneous symmetry breaking can occur. For $\lambda \ll 1$, and therefore for light Higgs particles, this requirement of vacuum stability, i.e. λ positive and large enough to avoid instabilities, implies a lower bound on m_H (stability bound).

Another limit can be imposed after observing that the coupling constant evolution with energy presents a singularity (Landau pole) for some energy value. In a simplified theory with only scalars, the evolution of λ is given by

$$\lambda(Q) = \frac{\lambda(Q_0)}{1 - \frac{3\lambda(Q_0)}{4\pi^2}\log\left(Q^2/{Q_0}^2\right)} \quad , \tag{1.42}$$

with Q_0 some reference scale. Since λ explodes in proximity of the pole, the theory is no more consistent with a perturbative approach. A possible way to make it meaningful for all energy scales is to set the value of λ to 0, which would result in a trivial, i.e. non interacting theory. Another possible approach is to consider the Standard Model as an effective theory up to an energy scale Λ, after which it leaves the perturbative domain. This requirement imposes an upper limit to the Higgs mass given by

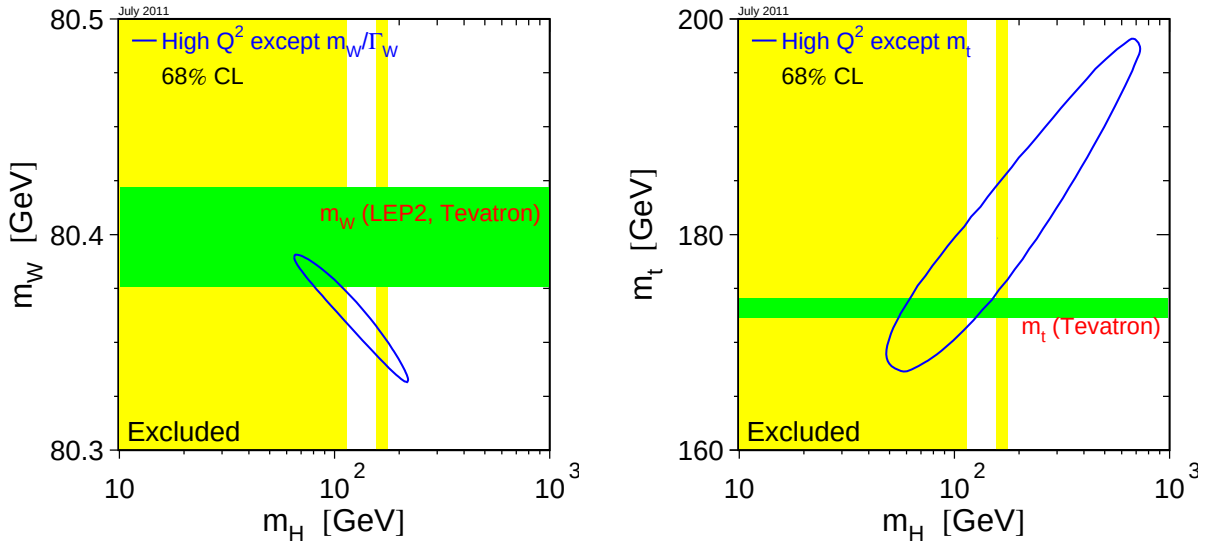
$$m_{\mathrm{H}}^2 < \frac{8\pi^2 v^2}{3\log\left(\Lambda^2/v^2\right)} \quad \text{(triviality bound)} \quad , \tag{1.43}$$

where we have assumed $Q_0 \equiv v$ and $\lambda(v) \equiv {m_{\mathrm{H}}}^2/\left(2v^2\right)$.
In a more correct approach, where also fermions are included in the theory, top-loop contributions cannot be neglected, and the result depends on m_{t} as well.
The theoretical constraints on m_{H} as a function of the cut-off scale Λ are shown in Figure 1.2, for a given choice of m_{t}. As it can be seen, this plot suggests that if the Standard Model validity extends up to the scale of Grand Unification Theories ($\Lambda_{\mathrm{GUT}} \simeq 10^{16}\,\mathrm{GeV}$), the Higgs boson mass has to lie roughly in the $150 - 180\,\mathrm{GeV}/c^2$ range. Conversely, for a Higgs particle lighter than 150 or heavier than $180\,\mathrm{GeV}/c^2$, new Physics is expected to exist at an energy scale inferior to Λ_{GUT}.

Finally, the Higgs boson enters into one-loop radiative corrections in the Standard Model, and so precise electroweak measurements can bound the Higgs mass. Since the dependance on m_{H} in all one-loop electroweak parameters is only logarithmic, the limits derived from this method are relatively weak. In contrast, the top quark contributes quadratically to many observables. From precision measurement at LEP and SLC of many electroweak observables, and direct 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% probability are shown in $(m_{\mathrm{W}}, m_{\mathrm{H}})$ and $(m_{\mathrm{t}}, m_{\mathrm{H}})$ planes.

In order to obtain the most stringent constraint on the mass of the SM Higgs boson, 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 18 different variables and their errors with their predictions calculated within the framework of the Standard Model as a function of five input parameters ($\Delta\alpha^{(5)}_{\mathrm{had}}({m_{\mathrm{Z}}}^2)$, $\alpha_S({m_{\mathrm{Z}}}^2)$, m_{Z}, m_{t}, m_{H}). This analysis procedure tests quantitatively how well the Standard Model is able to describe the complete set of measurements with just one value for each of the five 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.

From the result of the fit it is clear that electroweak measurements seem to favour a light Higgs,



(a) (b)

Figure 1.3: Contour curves of 68% probability in (a) the $(m_{\mathrm{W}}, m_{\mathrm{H}})$ plane and (b) the $(m_{\mathrm{t}}, m_{\mathrm{H}})$ plane, based on various LEP, LEP-II, SLC, and Tevatron measurements except the direct measurements of $m_{\mathrm{W}}, \Gamma_{\mathrm{W}}$ and the direct measurement of m_{t} respectively. The direct measurements of these excluded variables are shown as the horizontal bands of width ± 1 standard deviation. The vertical band shows the 95% confidence level exclusion limit on m_{H} of 114.4 GeV/c^2, derived from direct searches at LEP-II [11] and Tevatron [12].

with the one-sided 95% C.L. upper limit on m_{H} being

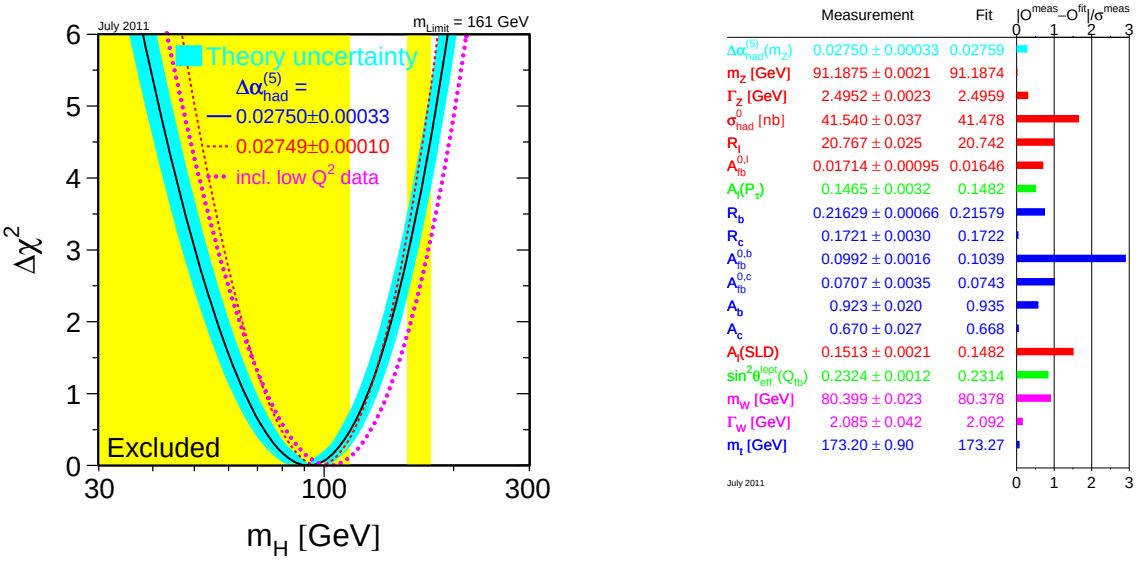
$$m_{\mathrm{H}} < 161\,\mathrm{GeV}/c^2 \quad \text{(electroweak fit)} \quad . \tag{1.44}$$

Of course, since these bound arises from loop corrections, it may be circumvented by some still unknown new Physics entering the same loops.

1.2.2 Direct constraints from previous experiments

Experimental bounds on the Higgs boson mass are provided by measurements at different experiments. The first direct search was carried out at the LEP accelerator at CERN [11]. Data from e^+e^- collisions at a center-of-mass energy up to 209 GeV were used to look for hints of the Higgs boson particle by the four experiments operative installed on the LEP accelerator. The main production mechanism at a e^+e^- collider was the so-called Higgs-strahlung, where a Higgs boson is radiated by a virtual Z boson, and the most probable final state consisted in a couple of b-jets from Higgs decay and another couple of jets from Z boson decay (although also the leptonic decay modes of the Z were considered). To summarize, the following channels were analyzed (sorted for relative importance):

$$\begin{aligned}
&\mathrm{e^+e^-} \to \mathrm{Z^*/\gamma^*} \to \quad \mathrm{Z}(\to \mathrm{q\bar{q}}) \quad \mathrm{H}(\to \mathrm{b\bar{b}}) \\
&\mathrm{e^+e^-} \to \mathrm{Z^*/\gamma^*} \to \quad \mathrm{Z}(\to \nu\bar{\nu}) \quad \mathrm{H}(\to \mathrm{b\bar{b}}) \\
&\mathrm{e^+e^-} \to \mathrm{Z^*/\gamma^*} \to \quad \mathrm{Z}(\to \mathrm{l^+l^-}) \quad \mathrm{H}(\to \mathrm{b\bar{b}}) \qquad \text{with} \qquad \mathrm{l} = \mathrm{e},\mu,\tau \quad .
\end{aligned} \tag{1.45}$$



(a) (b)

Figure 1.4: Global fit of all precision electroweak measurements. (a) The black continuous line shows $\Delta\chi^2(m_H) = \chi^2{}_{min}(m_H) - \chi^2{}_{min}$ as a function of m_H. The shaded cyan area represents the associated theoretical error due to missing higher-order corrections. The yellow shaded area corresponds to the 95% C.L. exclusion limit $m_H > 114.4\,\text{GeV}/c^2$ (LEP-II, 2003 [11]) and $m_H \notin [158, 175]\,\text{GeV}/c^2$ (Tevatron, 2010 [12]). (b) Comparison of all the 18 measurements with their SM expectations calculated for the five input parameters values in the minimum of the global χ^2 of the fit.

No significant excess[7] of events with a Higgs-compatible topology were found, thus allowing to set a lower bound on the Higgs boson mass at 95% confidence level (as shown in Figure 1.5):

$$m_H > 114.4\,\text{GeV}/c^2 \quad \text{(LEP-II – 2003)} \quad . \tag{1.46}$$

The search for the Standard Model Higgs particle continued at Tevatron, a $p\bar{p}$ collider with a center-of-mass energy of $\sqrt{s} = 1.96\,\text{TeV}$ which ended its operation in 2011. Also in this case, 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 final states was considered by the CDF and DØ experiments. As it can be seen in Figure 1.6, the combination of all analyses [13] leads to an exclusion of the SM Higgs boson particle in a mass region around $160\,\text{GeV}/c^2$ (mostly driven by $H \to WW$ final states) and at very low mass, in a region already covered by LEP results, namely

$$\begin{aligned} &m_H \notin [147, 179]\,\text{GeV}/c^2 \\ &m_H \notin [100, 106]\,\text{GeV}/c^2 \end{aligned} \quad \text{(Tevatron – 2011)} \quad . \tag{1.47}$$

[7] At a mass of $115\,\text{GeV}/c^2$, the ALEPH experiments reported an excess of four events in the four-jet final state, compatible with the production of a Standard Model Higgs boson. This excess was nevertheless not statistically significant, especially after combining it with the results of DELHPI, L3 and OPAL.

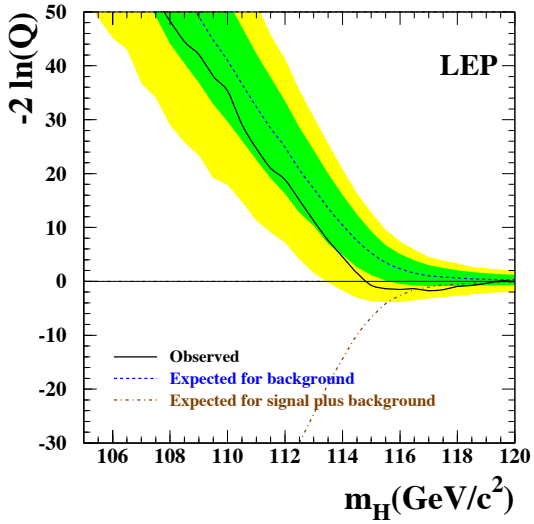


Figure 1.5: Observed and expected behaviour of the test statistics $-2\ln Q$ as a function of the Higgs mass. The test statistics Q is defined as the ratio between the signal plus background likelihood and the background only likelihood. The result is the combination of the data collected by the four 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. Picture from [11].

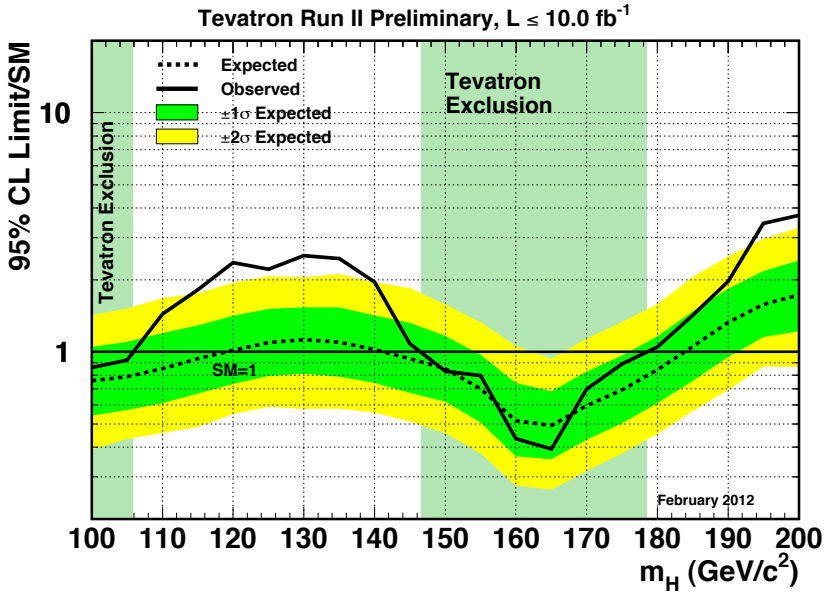


Figure 1.6: Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on the ratios to the SM cross section, as a function of the Higgs boson mass for the combined CDF and DØ analyses. The limits are expressed as a multiple of the SM prediction for test masses for which 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. Picture from [13].

1.3 Higgs boson searches at the LHC

The LHC is a machine situated at CERN that started colliding protons at center-of-mass energy of 7 TeV in 2010. Among its main Physics objectives is for sure the continuation – and hopefully the conclusion – of the Higgs boson hunt started by its predecessors. While a more detailed description of the accelerator and the CMS experiment is given in Chapter 2, a brief overview of the expected scenario of the production and the detection of a SM Higgs boson at such energies is given here below.

1.3.1 Higgs boson production mechanisms

The SM Higgs boson production cross-section at a pp hadron collider of center-of-mass energy $\sqrt{s} = 7$ TeV is shown in Figure 1.7 [14]. In Figure 1.8 the corresponding leading-order Feynman diagrams are drawn.

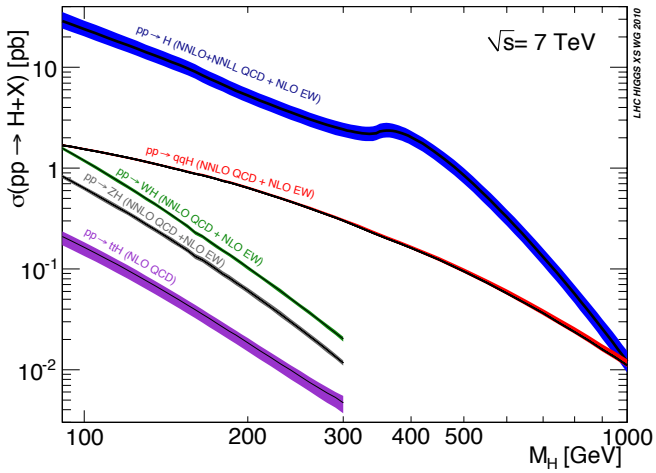


Figure 1.7: Higgs production cross-sections at $\sqrt{s} = 7$ TeV as a function of the Higgs mass for the different production mechanisms. From top to bottom, sorted for their relevance: gluon fusion (blue), VBF (red), associate production with a W/Z boson (green/grey), $t\bar{t}$ associated production (violet). NNLO QCD corrections as well as NLO EKW corrections are taken into account. Picture from [14].

The **gluon fusion** ($gg \to H$) is the dominating Higgs production process over the entire mass range accessible at the LHC. It proceeds with a heavy quark triangle loop, as shown in Figure 1.8. Because of the Higgs couplings to fermions, the t-quark loop is the most important, with the next-to-leading contribution being at least a factor $\mathcal{O}(m_t^2/m_b^2)$ smaller. 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 ([15, 16]) and maintaining the full m_t dependance [17]. Also NNLO have become available recently in the large m_t limit, and they provide a further $\sim 25\%$ enhancement, depending on the Higgs mass [18].

The second largest production mechanism of the Higgs boson is via **vector boson fusion** (VBF, $qq \to qqH$). In this process, which is about one order of magnitude below the gluon fusion, the Higgs boson is originated from the fusion of two weak bosons radiated off the incoming quarks. Their hadronization produces two forward jets of high invariant mass which can be used to tag the event and differentiate it from backgrounds. Another interesting property is the reduced hadronic activity in the region in between the tag jets, since they are colour disconnected. NLO

QCD and EWK corrections to this production process are found to be modest (order of 5–10%) and roughly of the same order [19]. Approximate NNLO QCD corrections to the total inclusive cross section for VBF have been presented [20] and shown to reduce the scale dependance of the result to only 1–2%.

In the **Higgs-strahlung** ($q\bar{q}' \to WH$, $q\bar{q} \to ZH$) and $t\bar{t}$ **associated production** ($gg, q\bar{q} \to t\bar{t}H$) processes the Higgs is produced in association with a W/Z boson or a pair of t quarks. In both cases, their decay products can be used to tag the event. Nevertheless the cross section is orders of magnitude smaller with respect to the gluon fusion one [21, 22]. 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.

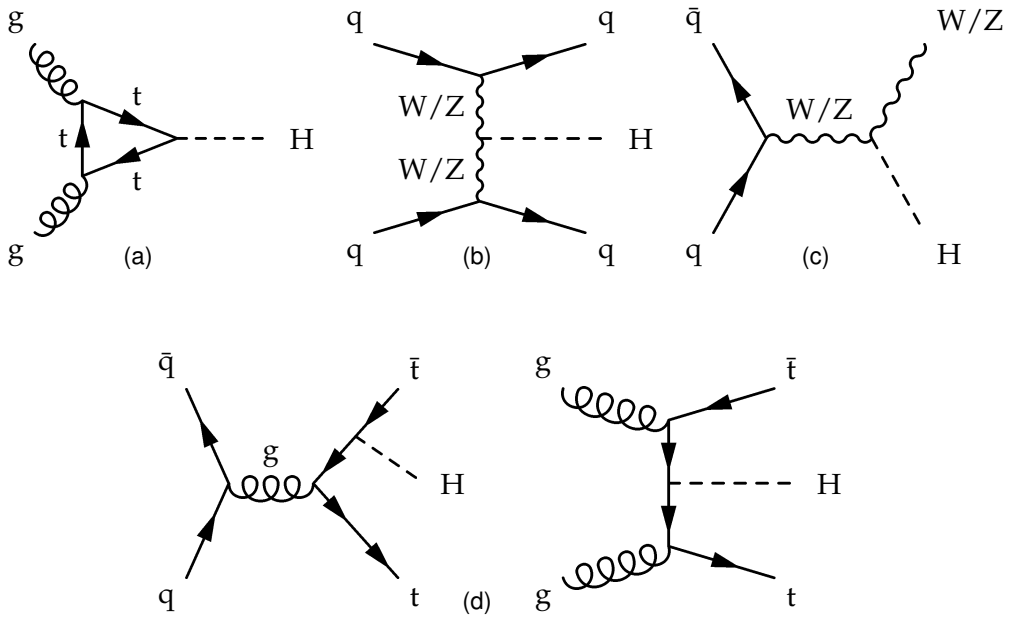


Figure 1.8: Feynman diagrams for the most important LO production processes of the SM Higgs boson: (a) gluon fusion, (b) vector boson fusion, (c) Higgs-strahlung, (d) $t\bar{t}$ associated production.

1.3.2 Higgs boson decay modes

Depending on the Higgs mass, different decay channels can be exploited to detect the particle. 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 paragraph. Due to the dependance of Higgs couplings on the particle masses, the Higgs tends to decay into the heaviest particles which are kinematically allowed. This behaviour is visible in Figure 1.9a, which shows the Higgs decay branching ratios including also NLO QCD and EWK corrections [23]. It can be seen how new channels open up when the Higgs mass is above a di-lepton/quark or vector boson threshold. 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 boson 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. The Higgs boson does not couple to photons and gluons at tree level, but such couplings can arise via fermion loops and they give a contribution in the low mass region.

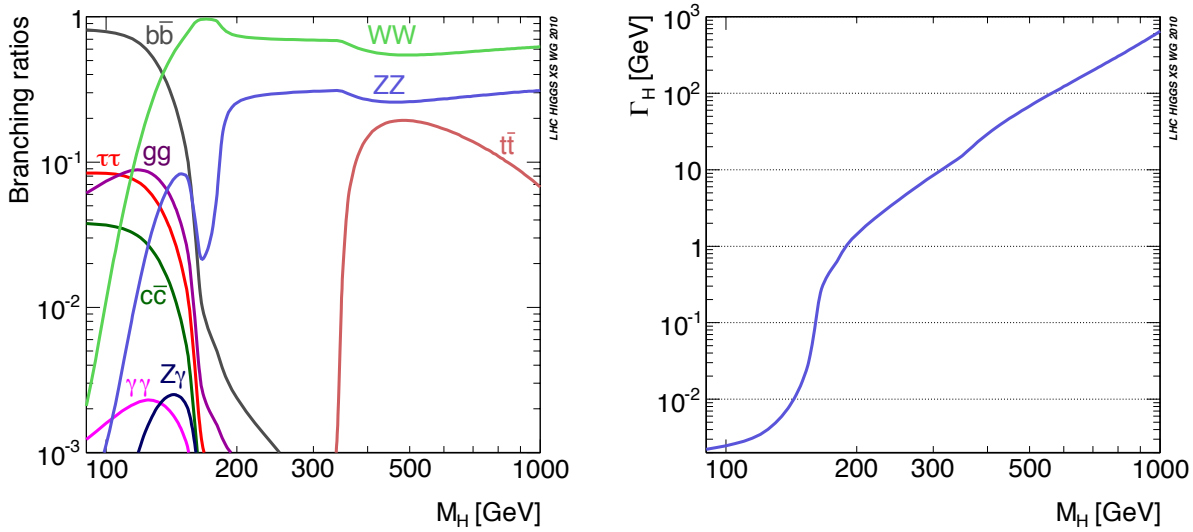


Figure 1.9: (a) Decay branching ratios of the SM Higgs boson in the different channels versus its mass. (b) Total decay width (in GeV) of the SM Higgs boson with respect to its mass. Pictures from [14].

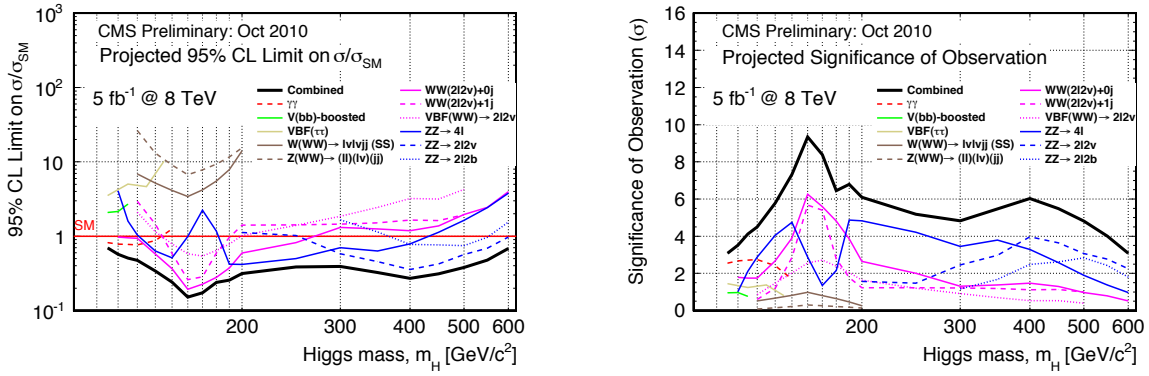
The total width, given by the sum over all the possible decay channels, is shown in Figure 1.9b. It quickly increases with the Higgs mass due to the opening of new channels and it becomes almost as large as the Higgs mass itself around $1\,\mathrm{TeV}/c^2$, thus making it problematic to identify the Higgs resonance as a particle: such a heavy Higgs would be more suitably accommodated in theories extending the Standard Model.

The decay $H \to \gamma\gamma$ is the main channel for the discovery of the Higgs boson at masses of about $140\,\mathrm{GeV}/c^2$ or below. The challenge here is the low branching ratio and therefore the small signal rate. Large backgrounds come from prompt photon pairs produced by quarks and gluons in the initial state, from one or two jets which fake the photon signature and from Drell-Yan production of electron pairs. The signal signature is two high-E_T isolated photons which can be well identified experimentally, and the Higgs boson can be detected as a narrow peak above a large background. With a cut-based analysis the Higgs boson can be discovered in this channel with a 5σ significance from the LEP limit up to $m_H = 140\,\mathrm{GeV}/c^2$ with an integrated luminosity lower than $30\,\mathrm{fb}^{-1}$.

Due to its very clean signature with 4 isolated leptons in the final state, $H \to ZZ^{(*)} \to l^+l^-\,l^+l^-$ is considered the golden-plated mode for the discovery of the Higgs boson. The backgrounds to this channel are $ZZ^{(*)}$, $t\bar{t}$ and $Zb\bar{b}$ productions, which can be suppressed in an efficient way by some requirements on the lepton isolation, transverse momentum and invariant mass and by requirements on the event vertex. The Higgs discovery in the $ZZ^{(*)}$ channel is possible with an integrated luminosity of $10\,\mathrm{fb}^{-1}$ in the whole range of masses between 130 and $500\,\mathrm{GeV}/c^2$, apart for the small region around $m_H \simeq 160\,\mathrm{GeV}/c^2$.

In this mass region the Higgs branching ratio into WW is close to one. This makes $H \to WW^{(*)} \to l\nu_l l\nu_l$ the discovery channel in this mass range. The signature is two charged leptons and missing energy. Since the mass peak can not be reconstructed due to the neutrinos in the final state, the search strategy is based on event counting, for which an accurate knowledge of all the possible backgrounds is needed. The main backgrounds are electroweak WW, $t\bar{t}$ and W+jets productions. They can be reduced by requirements on the lepton momentum and isolation, by a jet veto and exploiting the small opening angle between the two leptons which is due to spin correlations. The Higgs boson can be discovered in this channel with a 5σ significance with less than $10\,fb^{-1}$ in the $150 < m_H < 180\,GeV/c^2$ region [24].

As shown in Figure 1.10, CMS Monte Carlo studies conducted in 2008 indicate that, in case a SM Higgs boson exists, an integrated luminosity of $5\,fb^{-1}$, 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) Projected exclusion limit

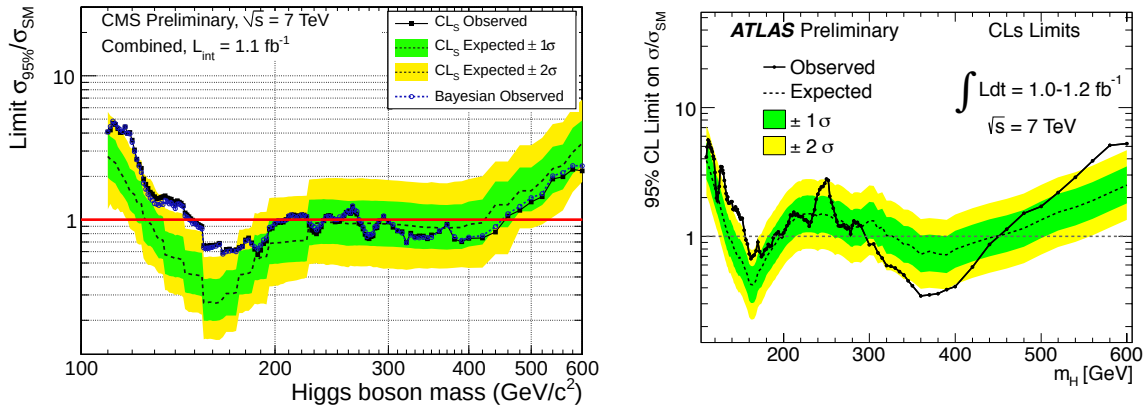
(b) Expected discovery significance

Figure 1.10: The plots show (a) the projected exclusion limit and (b) the expected observation significance for a SM Higgs search a $\sqrt{s} = 8\,TeV$ and $\mathcal{L} = 5\,fb^{-1}$. Contributions of individual channels used in the overall combination are also shown.

1.4 The state of the art

The quest for the Higgs boson started at the LHC already in 2010, profiting from the first $\sim 40\,pb^{-1}$ of integrated luminosity collected by the CMS and ATLAS experiments. Some meaningful constraints on Beyond-Standard-Model Higgs particles were provided by both experiments. In particular, from the absence of statistically-significant excesses in the $H \to W^+W^-$ and $H \to \tau^+\tau^-$ final states, it was possible to expand the constraints on the Higgs mass in a scenario where a fourth generation of highly massive fermions extends the SM framework or in the minimal supersymmetric extension to the Standard Model [25, 26, 27].
By Summer 2011, the LHC had already provided the CMS and ATLAS experiments with a large amount of collision data, about $1\,fb^{-1}$. At the EPS conference (Grenoble, July 21–27 2011) both experiments showed updated results on the SM Higgs boson exclusion limit [28, 29], largely

extending the pre-existing constraints on m_H. As can be seen in Figure 1.11, Tevatron results about the non-existence of a SM Higgs particle in a region around $m_H = 160\,\text{GeV}/c^2$ were confirmed and extended. Moreover, the Higgs boson was excluded, at 95% confidence level, in an intermediate-to-high mass range by both experiments ($m_H \notin [300, 440]\,\text{GeV}/c^2$ for CMS and $m_H \notin [295, 450]\,\text{GeV}/c^2$ for ATLAS).
Therefore, the very low mass region ($m_H \lesssim 140\,\text{GeV}/c^2$) and the very high mass region ($m_H \gtrsim 400\,\text{GeV}/c^2$) emerged as the most interesting mass ranges where to concentrate the studies. Moreover, the range between 216 and $400\,\text{GeV}/c^2$ remained only marginally excluded by the CMS experiment, with a $\pm 1\sigma$ band around $\sigma_{95\%}/\sigma_{SM} \sim 1$. These considerations drove the studies described in this work.



(a) (b)

Figure 1.11: 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 [28] and (b) the ATLAS experiment [29] with $\mathcal{L} = 1.0 - 1.2\,\text{fb}^{-1}$. These results were shown at the EPS conference in July 2011.

A light Higgs discovery is strongly favoured by all electroweak indirect constraints, yet this mass region is the most difficult to access for many different reasons. The only final states of practical usage in this mass range have a tiny branching ratio, so that the expected event yield is small despite the large production cross-section. The major player in this context is the $H \to \gamma\gamma$ channel. The most critical aspect to best profit from the very narrow Higgs resonance ($\Gamma_H < 1\,\text{GeV}$) is the energy resolution of the electromagnetic calorimeter, whose constant term has to be kept lower than 0.5%. An imperfect crystal-to-crystal intercalibration as well as instabilities in the ECAL response during its operation are factors that directly degrade the calorimeter resolution.
Chapter 3 describes the usage of isolated electrons from $W \to e\nu_e$ decays to develop a prompt monitoring tool of the response of the calorimeter and to provide an intercalibration technique complementary to the usage of $\pi^0 \to \gamma\gamma$ events.

Although disfavoured by the electroweak fit, searches for a heavy Higgs particle need to be directly performed in order to definitively rule out its presence if no excess is found. Moreover, the high mass range is worth studying since the unprecedented center-of-mass energy provided by the LHC has made this kinematic region accessible for the first time. Deviations from the SM expectations or even unexpected resonances may then show up as signals of new

physics.
Chapter 4 describe the study of a new Higgs decay channel, i.e. $H \to WW \to l\nu_l q\bar{q}$. This analysis was started in Spring 2011 as a natural extension of CMS most relevant publications in the $H \to W^+W^-$ channel, with the hope to bring in additional sensitivity around $m_H \sim 300\,\text{GeV}/c^2$, where CMS observed limit is weaker, and at high masses well above $400\,\text{GeV}/c^2$. The $l\nu_l q\bar{q}$ final state has the largest branching fraction among all Higgs decays (about six times larger that the golden-plated $H \to WW^{(*)} \to l\nu_l l\nu_l$). Another advantage with respect to the fully leptonic case is the possibility of fully reconstructing a mass peak, thanks to the presence of only one neutrino. The direct search of a peak appearance over the continuum background shape can then be performed. There is of course a price to pay: since final states with jets are much more likely to occur in a hadron collider than final states with leptons, the background contamination for such signals is much higher. The following chapters show the techniques that were put in place to control these backgrounds sufficiently well that the advantages of this final states are realized, and the channel can be competitive with other final states in the search for the Higgs boson at high mass.

Chapter 2

The CMS detector at the LHC

Despite the remarkable accuracy in the description of elementary particle interactions, the Standard Model does not provide an answer to some fundamental questions, which remain unsolved. As a consequence, the Physics community has focused its attention to high-energy hadron colliders as the more suitable facility to search for, and potentially discover, new particles and the associated Physics [30].

A brief overview on the LHC machine and its experiments is presented in the first section of this chapter, followed by a more detailed description of the CMS detector and its sub-parts.

2.1 The Large Hadron Collider

Particle Physics has entered a new era with the Large Hadron Collider starting its operation at CERN in 2010. Physicists are given the opportunity to explore the Physics beyond the Standard Model thanks to a center of mass energy of $\sqrt{s} = 7\,\text{TeV}$ (8 TeV since March 2012) and an instantaneous luminosity L which is still constantly increasing towards the design value of $1 \cdot 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$. Not only the energy and the luminosity of the colliding beams make the LHC an unprecedented project, but also in terms of cost, complexity of the detectors and human effort it exceeds any other existing experiment.

The physics motivations for such a challenging effort are manifold [31]. Besides the quest for the Higgs boson, another main field of activity is the search for new types of symmetries beyond the well-established $SU(2)_L \times U(1)_Y$ gauge symmetry, like supersymmetry (SUSY), which could extend the present theoretical framework of particle Physics. There are several reasons to believe that the Standard Model is not the ultimate theory, one for all the fact that it fails to provide a unique coupling constant for all the fundamental interactions at very high energy. On the other hand, unification succeeds in more general theories, such as SUSY (see Figure 2.1).
There are also a number of open questions that LHC could help clarify. For example, the possible existence of additional generations of quarks and leptons or new gauge bosons; the origin of the asymmetry between matter and antimatter in the universe or the origin of quark confinement; the "elementarity" of quarks and leptons, etc..
Thanks to the large production rate of W and Z bosons and t and b quarks at the LHC, the

considerable amount of statistics available also allows many precise measurements within the electroweak sector, despite the harsh environment.
Finally, LHC is also capable of accelerating lead ions and producing PbPb collisions at a center-of-mass energy of 2.76 TeV per nucleon pair. The creation of the so-called quark-gluon plasma in the interaction point recreates the temperature and density conditions that existed in the first microseconds after the Big Bang and promises to shed light on basic aspects of the strong interaction such as the nature of confinement, the mechanism of mass generation (chiral symmetry breaking, structure of the QCD vacuum) and hadronization, which still evade a thorough theoretical description due to their highly non-perturbative nature.

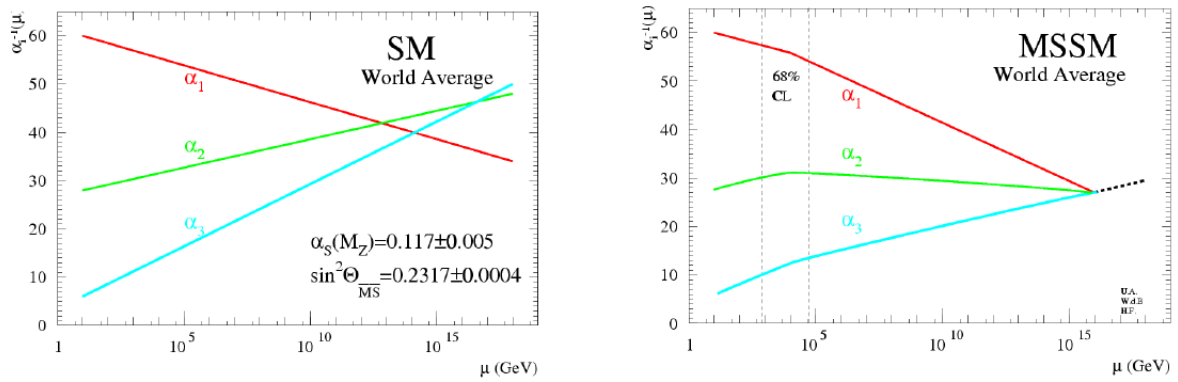


Figure 2.1: Evolution of the electroweak, strong and gravitational coupling constants with the energy scale of interactions μ for (left) the Standard Model and (right) the Minimal SuperSymmetric Model [32].

2.1.1 The machine

The Large Hadron Collider (LHC) [33] is a 27 km long proton collider installed 100 m underneath the ground level in the tunnel previously built to host the LEP e^+e^- accelerator. The choice to build a (hopefully!) 14 TeV center-of-mass energy collider and the radius of the collider circumference impose a magnetic field of $\sim$ 8.3 T to the 1232 super-conducting dipole magnets, in order to keep the beams on a circular trajectory. The Nb-Ti magnets are kept to a temperature of 1.9 K by means of super-fluid Helium. Since collisions occur between particles of the same charge, two separate beam pipes are required, with two opposite magnetic field configurations.

Collisions at a center-of-mass maximum energy of 14 TeV are the result of subsequent acceleration steps. First of all, each colliding bunch is formed in the Proton Synchrotron (PS) at an energy of 26 GeV, accelerated to 450 GeV by the Super Proton Synchrotron (SPS) and eventually injected into the LHC. Here, 8 radio-frequency resonant cavities oscillating at 400 MHz accelerate the bunches to their final energy with "kicks" of 0.5 MeV per turn. A schematic view of the acceleration complex can be seen in Figure 2.2.

The LHC is designed to ultimately reach an instantaneous luminosity of $L = 1 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, referred to as high luminosity, after a first few years of operation at lower luminosity. This corresponds to an accumulated statistics per year of $\mathcal{O}(100)\,\mathrm{fb}^{-1}$. To reach such high luminosities, up to 2808 bunches per beam, with about 1.1×10^{10} protons each and a small transverse size of

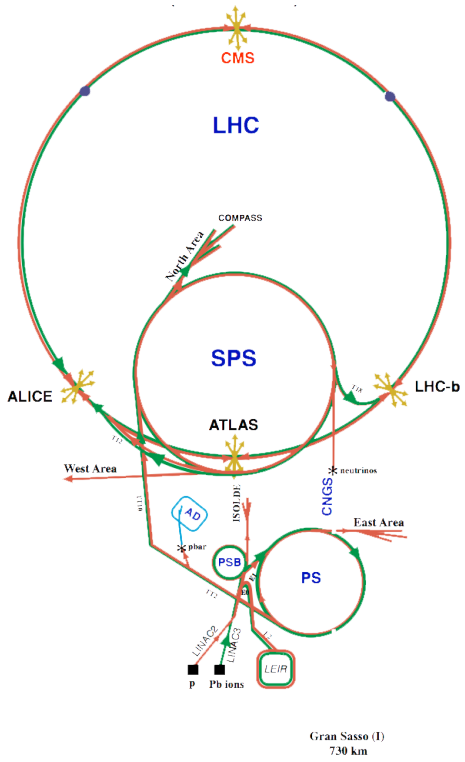


Figure 2.2: LHC injection scheme.

15 µm, will be collided every 25 ns. Up to 2012, though, pp collisions have been delivered with a minimal bunch time separation of 50 ns.

Four detectors are installed along the LHC tunnel, as shown in Figure 2.2, corresponding to the four bunch crossing points: CMS [34] and ATLAS [35], both multi-purpose detectors, LHCb [36], whose main field of study is b-physics and CP violation, and ALICE [37], devoted to heavy ions collision.

2.2 The CMS detector

The multi-purpose CMS detector is installed at interaction point n° 5 in the LHC tunnel. The heart of the detector, as can be seen in Figure 2.3, is a large superconducting solenoid, with a length of around 13 m and an inner diameter of about 6 m. The resulting axial magnetic field has a strength of 3.8 T: such an intense field is required in order to achieve a good momentum resolution, with the goal to retain a 10% resolution for muon transverse momenta of 1 TeV. Moreover the large bending power of charged particles results in small tracking devices, which are accommodated inside the solenoid together with all the calorimetry (ECAL + HCAL), hence the name "compact". The solenoid requires a return yoke, which is made of iron and can host four concentric muon detectors (muon stations).

In a typical 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,

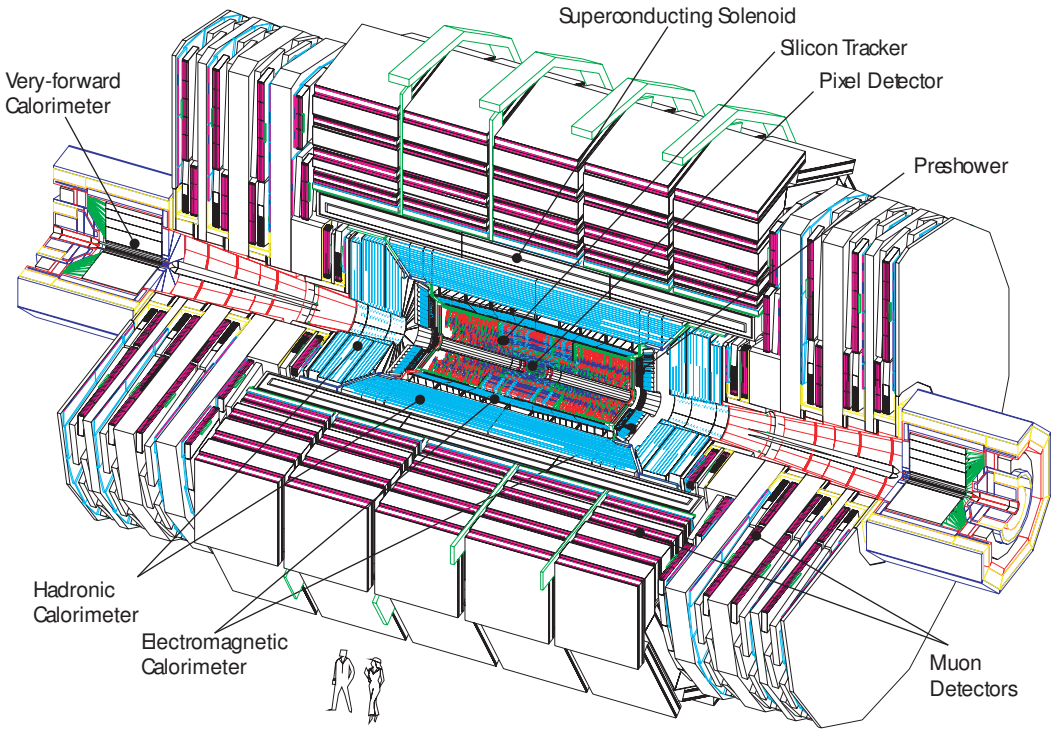


Figure 2.3: Pictorial view of the CMS detector.

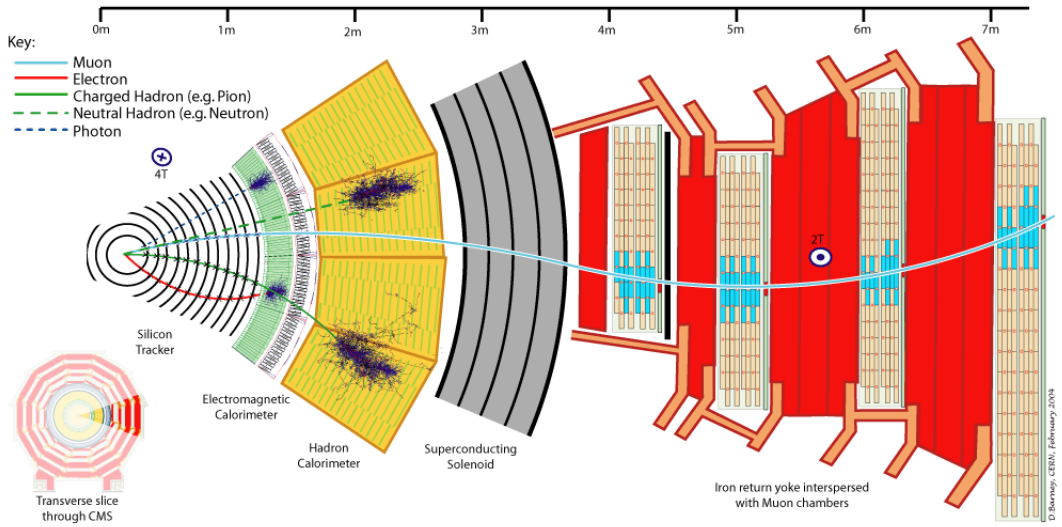


Figure 2.4: A slice of CMS: the picture shows the sub-detector sequence. Paths of different particles are also drawn.

which is in practice not possible because of the presence of the beam-pipe and instrumentation at small angles with respect to the beam-line.
Because of the unknown energy balance along the beam-line, proton collisions are usually studied in a convenient coordinate frame that has the z axis placed along the beam-line, and the x and y axes hence defining the transverse plane, with x pointing radially towards the center of the accelerator ring and the y axis upwards. The spherical coordinates (r, θ, ϕ) are replaced with (r, η, ϕ), where r is the radial distance from the z axis, ϕ is the azimuthal angle in the transverse plane and the pseudorapidity[1] η is defined as $\eta = -\log\left(\tan\frac{\theta}{2}\right)$. The advantage of this coordinate frame is the Lorentz invariance of transverse quantities and 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.

The structure of sub-detectors, as moving from the interaction point towards the outer part of CMS, is represented in Figure 2.4 and shortly described hereafter. A more complete description of the full CMS detector is available at [38].

2.2.1 The tracker system

The inner CMS tracking system [39] is designed to provide a precise and efficient measurement of the trajectories and the momentum of charged particles emerging from the LHC collisions, as well as a precise reconstruction of secondary vertices. It surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m.

At the LHC design luminosity an average of about 1000 particles come from more than 20 overlapping proton-proton interactions for each bunch crossing. Therefore a detector technology featuring high granularity and fast response is required so that trajectories can be identified reliably and attributed to the correct bunch crossing. The intense particle flux will also cause severe radiation damage to the tracking system. The main challenge in the design of the tracking system was then to develop detector components able to operate in this harsh environment for an expected lifetime of ten years. These requirements on granularity, speed and radiation hardness lead to a tracker design entirely based on silicon detector technology.

The CMS tracker is composed of a pixel detector with three barrel layers at radii between 4.4 cm and 10.2 cm and a silicon strip tracker with 10 barrel detection layers extending outwards to a radius of 1.1 m. Each system is completed by endcaps which consist of 2 disks in the pixel detector and 3 plus 9 disks in the strip tracker on each side of the barrel, extending the acceptance of the tracker up to a pseudorapidity of 2.5. A schematic view of the tracker is in Figure 2.5.

In the pixel detector, the estimated resolution on the single hit is of 10 μm for the (r, ϕ) coordinate and 15 μm for z in the barrel, while it is of 15 μm and 20 μm respectively in the endcaps. In total the pixel detector covers an area of about 1 m^2 and has 66 million pixels. For the silicon strip detector, the expected resolutions grow to $\sim$ 50 μm in (r, ϕ) and $\sim$ 500 μm along the z coordinate. The CMS silicon strip tracker has a total of 9.3 million strips and 198 m^2 of active silicon area, making it the largest entirely silicon-based tracker detector ever built.

[1] The pseudorapidity η is in fact the high-energy limit of the rapidity y, defined as $y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right)$. Differences of rapidities are Lorentz-invariant under boosts along the z axis.

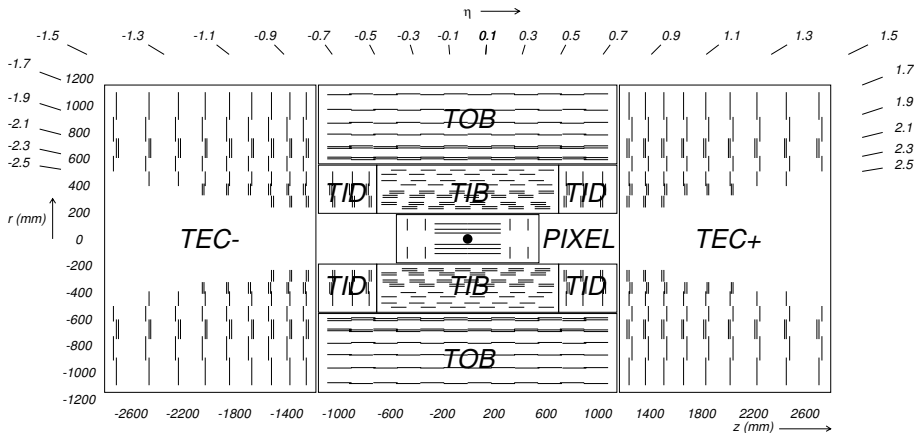


Figure 2.5: Schematic cross section through the CMS tracker. Each line represents a detector module.

One of the major constraints in the design of a tracking system is to reduce as much as possible the amount of material distribution in front of the subsequent calorimeters. For the CMS tracker, the material budget, shown in Figure 2.6, constitutes the main source of error in accurate calorimetric measurements of electrons (which emit bremsstrahlung radiation) and photons (which convert into e^+e^- pairs).

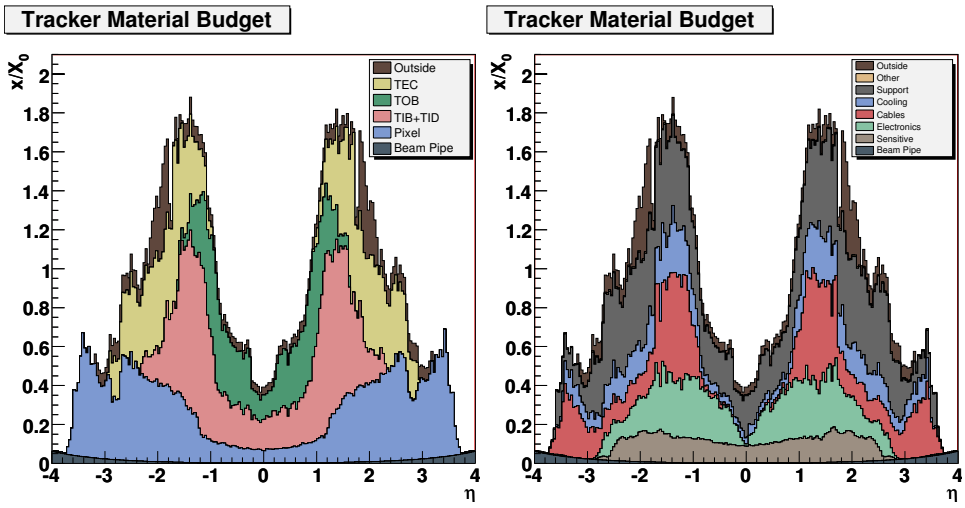


Figure 2.6: Material budget in units of radiation length as a function of pseudorapidity η for the different sub-detectors in left panel and broken down into the functional contributions in the right panel.

The supreme performance of the CMS tracker can be evicted at a glance from Figure 2.7, where the whole invariant mass spectrum of di-muon resonances with data collected in 2010 is shown. Above a large background continuum, many narrow resonances are reconstructed with good resolution. As an example, the Y resonances are shown in the right panel: three distinct peaks, originating from $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ decays into $\mu^+\mu^-$ pairs are clearly resolved.

2.2.2 The electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) [40] will play an essential role in the search for the Higgs boson by measuring, with high precision, the energy of photons arising in the $H \to \gamma\gamma$ channel and of electrons from the $H \to ZZ$ and $H \to W^+W^-$ decay chains. In order to

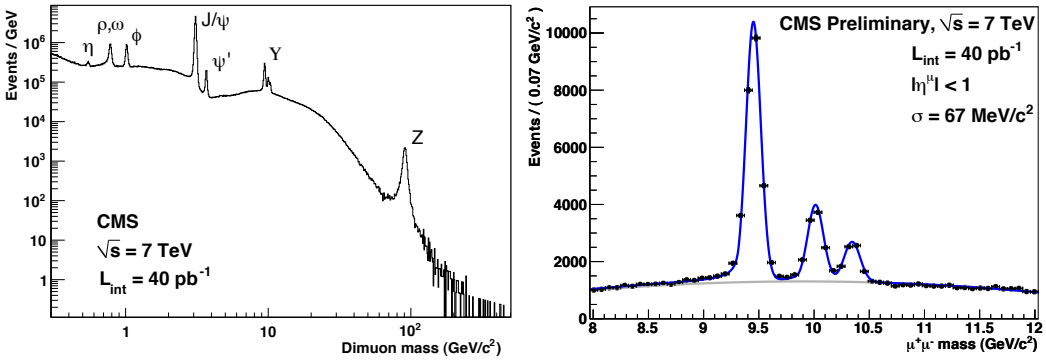


Figure 2.7: Invariant mass spectrum of di-muon resonances, using all pp collision data collected in 2010 ($40\,\mathrm{pb}^{-1}$). Data are triggered by di-muon HLT paths with no p_{T} threshold.

reach the high energy resolution these physics channels require, a homogeneous, scintillating crystal calorimeter has been chosen. The nearly 83000 lead tungstate ($PbWO_4$) crystal shape is that of a truncated pyramid, with a front face cross section of about $22 \times 22\,\mathrm{mm}^2$ ($\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174°$) and a length of 230 mm (which corresponds to 25.8 X_0). Thanks to the $PbWO_4$ short radiation length ($X_0 = 0.89\,\mathrm{cm}$) and the small Molière radius ($r_M = 21.9\,\mathrm{mm}$), most of an electron or photon energy can be collected within a small matrix of crystals.

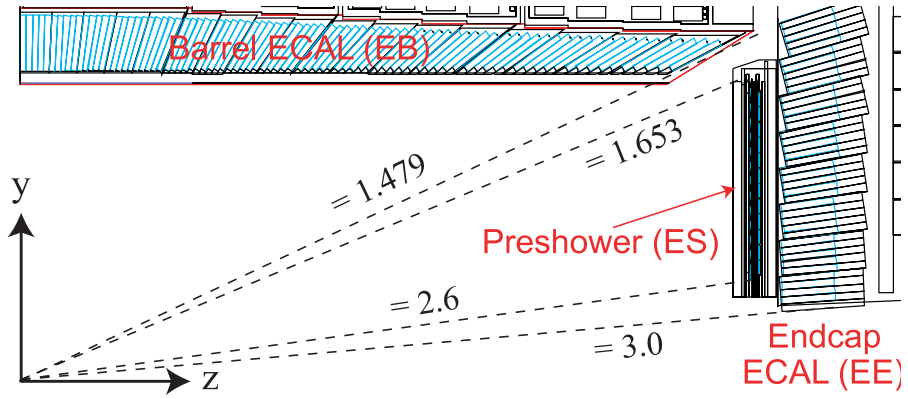


Figure 2.8: View in (r, z) of a quarter of the CMS electromagnetic calorimeter.

The ECAL crystals are divided into the two main parts of the calorimeter:

- the barrel (EB) is made of 61200 crystals and comes in the shape of a cylinder, with an inner radius of 1.290 m. It covers a pseudorapidity range of $0 < |\eta| < 1.479$ and its granularity is 360-fold in ϕ and 2×85-fold in η. EB crystals are grouped in arrays of 5×2 elements, called sub-modules: they are assembled in a thin glass-fibre alveolar structure, and all five alveoli in η contain crystals of the same size, thus reducing the number of crystal types to 17. Sub-modules are then arranged in a module, in the number of 40/50 and, eventually, 4 modules are assembled with metallic cross plates in between to form the biggest unity of the barrel part, the so-called supermodule. On the whole there are therefore 36 supermodules, 18 for each side of the interaction point. Supermodules in EB are mounted in a quasi-projective geometry to avoid cracks aligned with particle trajectories, so that the crystal axes make a small angle

(3°) with respect to the vector from the nominal interaction vertex, in both the ϕ and η projections;

- the two endcaps (EE) cover the pseudorapidity range $1.479 < |\eta| < 3.0$ and consist of identically shaped crystals, grouped into carbon-fiber structures of 5×5 elements, called supercrystals. Each endcap is divided into 2 halves, or Dees, holding 3662 crystals each;

In addition, a preshower (ES) is placed in front of EE crystals with the aim of identifying neutral pions in the endcaps and improving the position determination of electrons and photons. The preshower is a sampling calorimeter with two layers: passive lead radiators, that initiate electromagnetic showers from incoming photons and electrons, and active silicon strip sensors placed after each radiator, that measure the deposited energy and the transverse shower profiles. The total material depth corresponds to about $3X_0$ on a fiducial region of $1.653 < |\eta| < 2.6$.

Lead Tungstate has been chosen for the crystals as the scintillating material as its characteristics make it suitable for operation at LHC, in particular its rather weak though very fast emission and its resistance to irradiation. The scintillation decay time is comparable to the LHC bunch crossing time: at a rate of 40 MHz there will be an interaction every 25 ns, which is the time required for a crystal to emit about 80% of its scintillation light. The relatively low light yield of $\sim 30\,\gamma$/ MeV makes it necessary to use intrinsic high-gain photodetectors, capable of operating in high magnetic fields. Avalanche PhotoDiodes (APDs) and Vacuum PhotoTriodes (VPTs) are used, the former glued to the back side of each barrel crystal, the latter to endcap crystals.

According to the radiation damage models supported by direct measurements [40, 38], while the scintillation mechanism of $PbWO_4$ stays unaffected, irradiation modifies the $PbWO_4$ crystal transparency via creation of colour centers that absorb and scatter the light. At the ECAL working temperature (18° C) the damage anneals and the balance between damage and annealing results in a dose-rate dependent equilibrium of the optical transmission, if the dose rate is constant. In the varying conditions of LHC running the result is a cyclic transparency behaviour between LHC collision runs and machine refills. The performance of the calorimeter would be unacceptably degraded by these radiation induced transparency changes were they not measured and corrected for.
The evolution of the crystal transparency is measured using laser pulses injected into the crystals via optical fibers. Two laser wavelengths are used for the basic source. One, blue, at $\lambda = 440$ nm, is very close to the scintillation emission peak and 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, can be used to verify the stability of other elements in the system. The crystal response to laser light is normalized by the laser pulse magnitude measured using silicon PN photodiodes. Thus $R(t) = \mathrm{APD}(t)/\mathrm{PN}(t)$ is used as the measure of the crystal transparency.
To provide continuous monitoring during LHC runs, laser pulses are sent to ECAL detector elements in the LHC beam gaps of 3.17 μs, which are designed to rest the LHC kicker magnet in every 89.924 μs. Only about 1% of the available beam gaps are used for the ECAL monitoring data taking, with a corresponding data rate of about 100 Hz. The time needed to scan the entire ECAL is expected to be about 30 minutes. A schematic view of the ECAL laser monitoring

system is drawn in Figure 2.9.

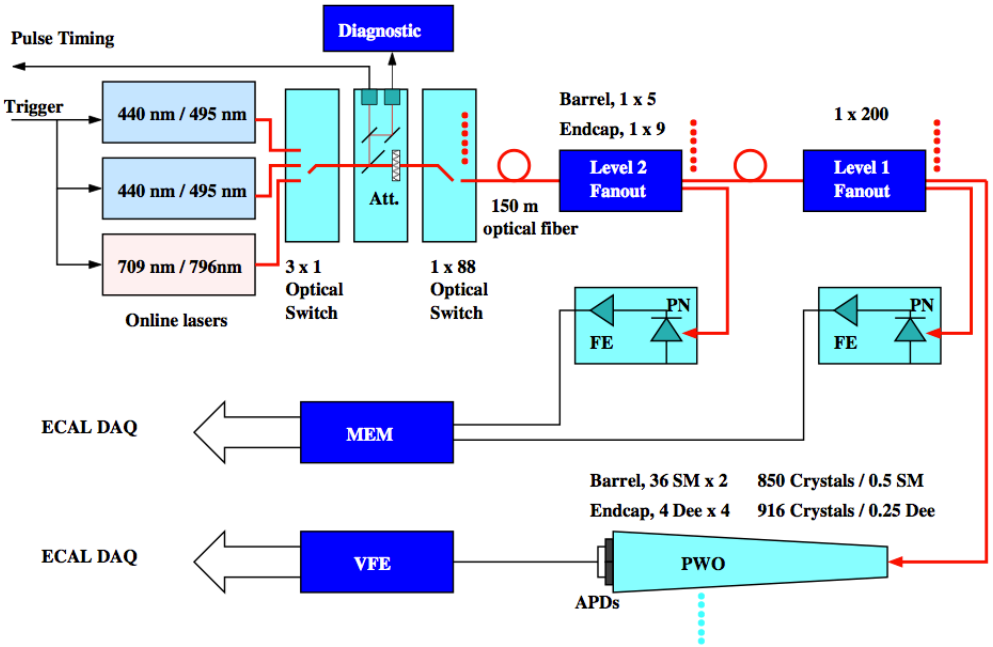


Figure 2.9: A schematic representation of the CMS ECAL laser monitoring system.

2.2.3 The hadron calorimeter

The goal of the hadronic calorimeter (HCAL) [41] is to measure the energy and the direction of hadronic jets as well as the missing transverse energy. HCAL is a sampling calorimeter whose barrel and endcaps sit behind the tracker and the electromagnetic calorimeter as seen from the interaction point. The hadron calorimeter barrel is radially restricted between the outer extent of the electromagnetic calorimeter (r = 1.77 m) and the inner extent of the magnet coil (r = 2.95 m). This constrains the total amount of material which can be put in to absorb the hadronic shower. Therefore, an outer hadron calorimeter is placed outside the solenoid complementing the barrel calorimetry. Beyond $|\eta| = 3$, the forward hadron calorimeters are placed at 11.2 m from the interaction point and extend the pseudorapidity coverage down to $|\eta| = 5.2$ using a radiation-hard technology.
A schematic view of the calorimeter location in CMS is given in Figure 2.10.

The central calorimeter is divided into a barrel part (HB, $0 < |\eta| < 1.3$) and two endcaps (HE, $1.3 < |\eta| < 3$), with a transverse granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087^\circ$ and $\Delta\eta \times \Delta\phi \simeq 0.17 \times 0.17^\circ$ respectively. The 3.7 mm thick active layers of plastic scintillators are interleaved with 5 to 8 cm thick brass absorbers. The scintillators have grooves in which there are wavelenght-shifter fibers inserted, used to bring out the scintillation light and read the signal. The total absorber thickness at 90° is 5.82 interaction lengths (λ_I). The HB effective thickness increases with polar angle θ as $1/\sin\theta$, resulting in 10.6 λ_I at the end of the barrel. In the endcaps, the total length of the calorimeter is about 10 interaction lenghts.
Since in the central pseudorapidity region the combined stopping power of EB plus HB does not provide sufficient containment for hadron showers, to ensure adequate sampling depth for the hadron calorimeter is extended outside the solenoid with a tail catcher (HO), using the

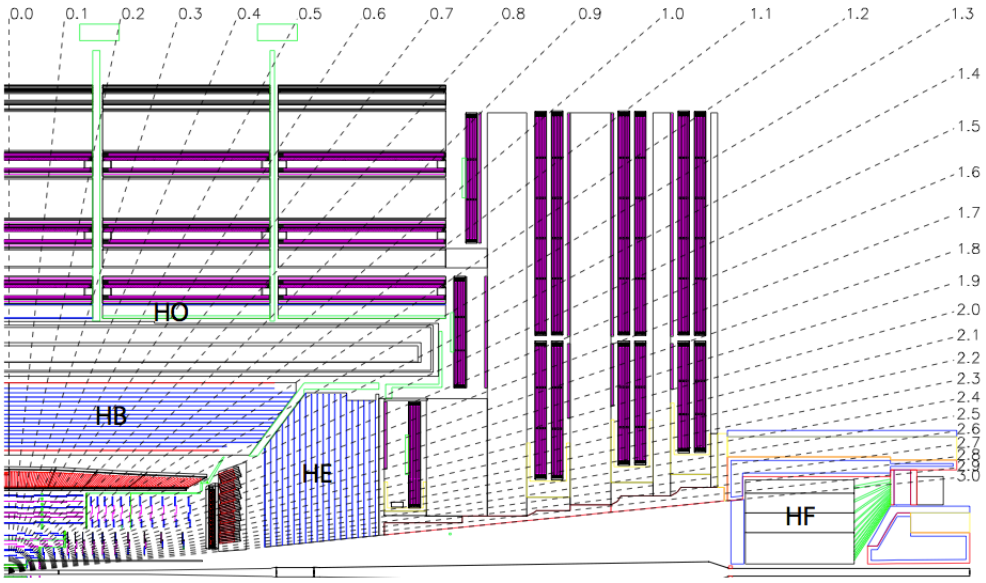


Figure 2.10: Longitudinal view in (r, z) of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters.

solenoid coil as an additional absorber equal to $1.4/\sin\theta$ interaction lengths.
The forward calorimeter (HF) experiences unprecedented particle fluxes (on average, 760 GeV per pp interaction is deposited into the two forward calorimeters, compared to only 100 GeV for the rest of the detector). The design of the HF calorimeter was guided by the necessity to survive in these hostile environment preferably for at least a decade (10 MGy of absorbed dose are expected at $|\eta| = 5$ after ten years of LHC operation). The calorimeter consists of a steel absorber structure composed of 5 mm thick grooved plates. Quartz fibers are inserted in these grooves and constitute the calorimeter active medium, detecting energy emitted by particles via Cherenkov radiation. For this reason, the calorimeter is mostly sensitive to the electromagnetic component of showers and practically insensitive to neutral hadrons. Its granularity is $\Delta\eta \times \Delta\phi = 0.175 \times 0.175°$.

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

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

2.2.4 The muon system

The CMS muon system [42] provides full geometric coverage for muon measurement up to $|\eta| = 2.4$. The detectors are embedded in the iron construction of the magnet return yoke, such that muon momentum and charge measurements can also exploit the strong magnetic return

field. This is particularly important for muons with transverse momentum in the TeV range, for which the complementary tracker measurements degrade.

In analogy to the previously described sub-detectors, the CMS muon spectrometer is composed of a barrel part ($|\eta| < 1.2$) and a forward region ($0.9 < |\eta| < 2.4$). The barrel consists of five wheels, in which drift-tube (DT) detectors and resistive-plate chambers (RPC) are placed in concentric muon stations around the beam line, with radius between about 4 and 7 m. The detector modules are arranged with respect to each other, such that a muon traverses at least three of the four detection layers.

In the forward region cathode-strip chambers (CSC) and RPC are mounted perpendicular to the beam line in overlapping rings on the endcaps. In Figure 2.11 an (r, z) view is given of the different parts of a quarter of the CMS muon system.

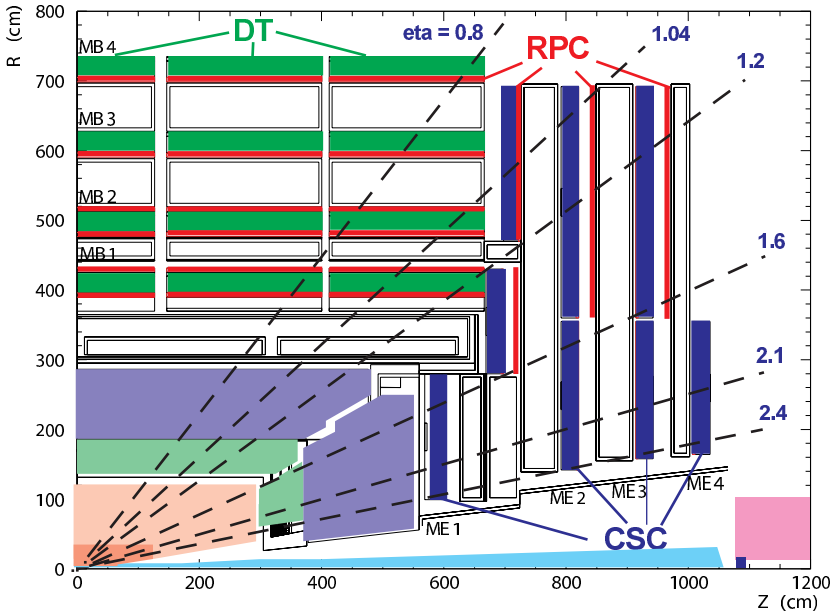


Figure 2.11: View in (r, z) of a quarter of the CMS muon system layout.

The CMS muon system uses three types of gaseous detectors. The choice of detector technologies has been driven by the need for fast triggers, excellent resolution, coverage of a very large total surface and operation in dense radiation environments.

- the drift tubes (DT) are located in the barrel part, where the muon rate is expected to be low. The CMS DTs are long aluminium cells filled with gas, and with an anode wire in the centre that collects ionization charges. In a DT chamber many of these drift tube cells are arranged in three (two in the outermost muon station) super-layers, consisting each of four layers of staggered drift tubes. Two of these super-layers have anode wires parallel to the beam line, providing a measurement of the r and ϕ coordinates; the third super-layer is placed perpendicular between the others, and provides the z measurement. Each station is designed to measure muon positions with about 1 mrad resolution in ϕ;
- the cathode strip chambers (CSC) are used in the endcap regions, where the magnetic field is very intense (up to several Tesla) and very inhomogeneous. CSCs are

multi-wire proportional chambers in which one cathode plane is segmented intro strips running across wires. An avalanche developed on a wire induces a charge on several strips of the cathode plane. In a CSC plane two coordinates per plane are made available by the simultaneous and independent detection of the signal induced by the same track on the wires and on the strips;

- the resistive plate chambers (RPC) consist of double-gap Bakelite chambers, with the 2 mm spacings filled with gas. The position resolution from the RPCs is poorer than for the DTs and CSCs, but the collection of charges on the strips is very fast, with a 1 ns time resolution. Therefore these chambers serve mainly for trigger purposes and for an unambiguous identification of the bunch crossing.

2.2.5 The trigger system

At the nominal LHC luminosity, the expected event rate is about 10^9 Hz. Given a typical raw event size of $\mathcal{O}(1)$ MB, it is not possible to store all collision events. In fact, the event rate is largely dominated by soft pp interactions, which are not interesting for the CMS Physics program. Therefore, a trigger system [43] has been devised with the purpose of providing a large rate reduction factor, whilst maintaining a high efficiency on potentially useful events.
The total output rate is reduced by about seven order of magnitudes to $\mathcal{O}(100)$ Hz thanks to a two-level system: a Level-1 (L1) Trigger, which consists of custom-designed, largely programmable electronics, and a High-Level Trigger (HLT), which is a software system implemented in a farm of about one thousand commercial processors. These two levels are briefly described here below.

The Level-1 trigger

The L1 trigger reduces the rate of selected events down to $\mathcal{O}(100)$ kHz. The full data are stored in pipelines of processing elements, while waiting for the trigger decision. The L1 decision about taking or discarding data from a particular bunch crossing has to be taken in 3.2 μs. If the L1 accepts the event, the data are moved to be processed by the High Level Trigger.
To deal with the 25 ns bunch crossing rate, the L1 trigger has to take a decision in a time too short to read data from the whole detector, therefore it employs the calorimetric and muons informations only, since the tracker algorithms are too slow for this purpose. The L1 trigger is organized into a calorimeter and a muon trigger, whose information is transferred to the global trigger which takes the accept-reject decision.

The calorimeter trigger is based on trigger towers, arrays of 5 crystals in ECAL which match the granularity of the HCAL towers. The trigger towers are grouped in calorimetric region of 4×4 trigger towers. The calorimeter trigger identifies the best four candidates of each of the following classes: electrons and photons, central jets, forward jets and τ-jets, identified from the shape of the deposited energy. The information of these objects is passed to the global trigger, together with the measured $E_{\mathrm{T}}^{\mathrm{miss}}$.
The muon trigger is performed separately for each muon detector. The information is then merged and the best four muon candidates are transferred to the global trigger, which takes the accept-reject decision exploiting both the characteristic of the single objects and their combination.

The High-Level Trigger

The HLT reduces the output rate down to $\mathcal{O}(100)$ Hz. The idea of the HLT software is the regional reconstruction on demand, that is only those objects in the useful regions are reconstructed and the uninteresting events are rejected as soon as possible. This leads to the development of three virtual trigger levels: at the first level only the full information of the muon system and of the calorimeters is used, in the second level the information of the tracker pixels is added and in the third and final level the full event information is available.
The use of a processor farm for all selections beyond Level-1 allows maximal benefit to be taken from the evolution of computing technology. Flexibility is maximized since there is complete freedom in the selection of the data to access, as well as in the sophistication of the algorithms, usually referred to as HLT paths. More details about the electron and muon trigger paths are given in Chapters 3 and 4.

2.3 Object reconstruction at CMS

A description of the the most important high-level Physics objects used in this work and the respective reconstruction efficiencies is presented here below.

2.3.1 Electron reconstruction

For a single electron reaching the ECAL surface, an electromagnetic shower starts within the first centimeters of the ECAL crystals and most of the electron energy is collected within a small matrix of crystals around the hit one. In general, the situation is found to be much more complicated. The tracker material budget causes electrons to loose part of their energy radiating photons by bremsstrahlung effect. As the electrons are degraded in energy, the effect of the magnetic field is to enhance the bending of their trajectories, which ultimately results in a spread of irradiated photons along the ϕ coordinate. 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.

This is the purpose of the first stage in the electron reconstruction sequence, which goes under the name of superclustering. Two algorithms, known as Hybrid and Island algorithms, are used to group electron crystals and recollect bremsstrahlung photons [44]. Both algorithms are seeded by single crystals with at least 1 GeV of deposited transverse energy.
The first one, used in the barrel, then 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. The dominoes are aligned with the seed crystal along η and extend up to ± 17 crystals away from the seed one. Different dominoes are then clustered together along ϕ, where a valley with less than 100 MeV of deposited energy separates different clusters. The so-obtained cluster of clusters goes under the name of supercluster.
The Island algorithm in the endcap, instead, builds clusters by connecting rows of crystals along ϕ containing energies decreasing monotonically when moving away from the seed crystal. Superclusters are built by collecting other Island clusters along a ϕ road in both directions around each Island clusters.

Once a supercluster is found, the reconstruction proceeds with the track-building stage. Under

both $+1$ and -1 charge hypotheses, the supercluster energy and position are back-propagated in the magnetic field 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 finally required to create a track.

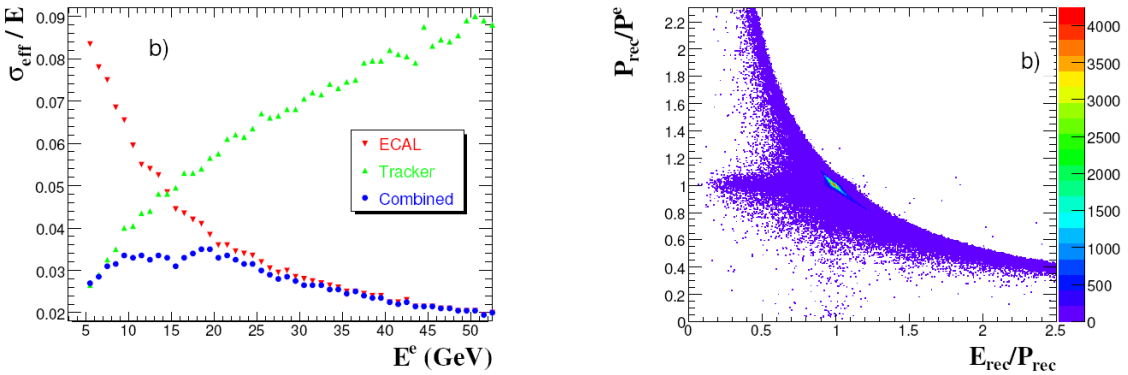


Figure 2.12: In the left plot the fractional resolution (effective RMS) is plotted as a function of generated energy E^e as measured with the ECAL supercluster (downward arrows), the electron track (upward arrows) and the combined track-supercluster (circles). In the right plot correlations between ECAL energy and tracker momentum measurements in the η range of the barrel are shown [47].

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

2.3.2 Muon reconstruction

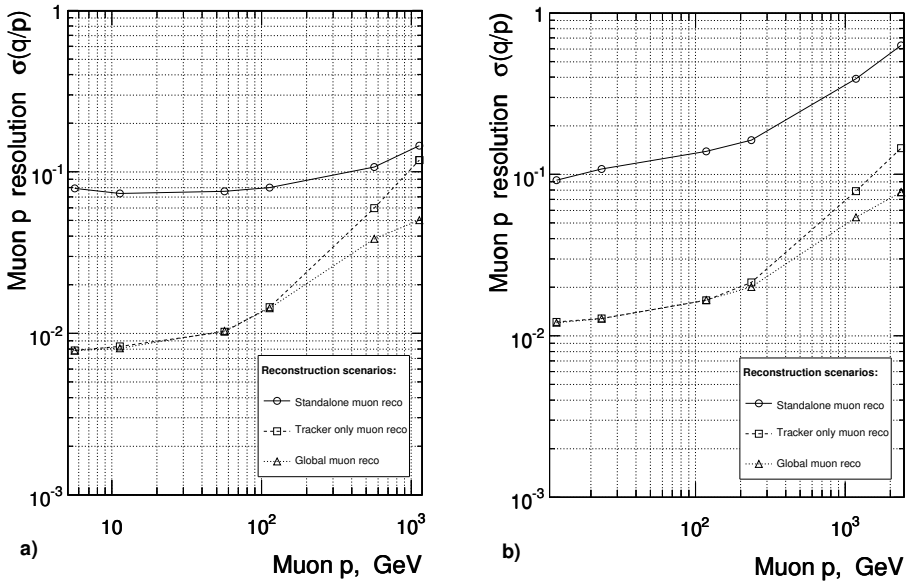
The standard muon reconstruction sequence is performed in two stages: a standalone reconstruction in the muon chambers and a global reconstruction in the whole detector [47].

The first stage starts from pre-built track segments (i.e. a set of aligned hits) in the muon chambers. The state vector associated to the segments found in the innermost chambers are used to seed the muon trajectory, working from inside out, using the Kalman-filter technique: the predicted state vector at the next measurement surface is compared with existing hits and updated accordingly. A suitable χ^2 cut is applied to reject bad hits and the procedure is iterated until the outermost surface of the muon system is reached. Finally the track is extrapolated to the nominal interaction point and a vertex-constrained fit is performed. Due to the large amount of material traversed to reach the muon spectrometers, the momentum resolution as

measured in the muon chambers is degraded by multiple scattering.

In the global muon reconstruction, muon trajectories are extended in order to include hits from the silicon tracker, starting from a standalone muon, allowing to significantly improve the momentum resolution of low- and medium-p_{T} muons.

As Figure 2.13 tells, the additional information provided by the muon tracking system is precious for the momentum reconstruction of high-energy muons ($p \gtrsim 100\,\mathrm{GeV}/c$), for which the tracker-only momentum measurement degrades. For lower momenta, instead, the resolution of the tracking system is dominating.



(a) (b)

Figure 2.13: Resolution of $1/p$ versus p for standalone, global and tracker-only reconstruction. (a) Barrel, $|\eta| = 0.5$; (b) Endcap, $|\eta| = 1.5$.

2.3.3 Particle Flow reconstruction for jets and missing transverse energy

The Particle Flow (PF) is a whole-event reconstruction technique whose purpose is the reconstruction and identification of each single particle produced in each pp collision with an optimized combination of all sub-detector information [48, 49]. In this process, the identification of the particle type (photon, electron, muon, charged hadron, neutral hadron) plays an important role in the determination of the particle direction and energy.
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 $E_{\mathrm{T}}^{\mathrm{miss}}$ 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.

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.

Electrons (e.g. coming from photon conversions in the tracker material or from hadron semi-leptonic decays) are identified as a primary charged particle track and potentially many ECAL energy clusters corresponding to this track extrapolation to the ECAL and to possible bremsstrahlung photons emitted along the way through the tracker material.
Muons (e.g. from hadron semi-leptonic decays) are identified as a track in the central tracker consistent with either a track or several hits in the muon system, associated with an energy deficit in the calorimeters.
Charged hadrons are identified as charged particle tracks neither identified as electrons, nor as muons.
Finally, 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.

The energy of photons is directly obtained from the ECAL measurement, corrected for zero-suppression effects. The energy of electrons is determined from a combination of the track momentum at the main interaction vertex, the corresponding ECAL cluster energy, and the energy sum of all bremsstrahlung photons attached to the track. The energy of muons is obtained from the corresponding track momentum. The energy of charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energy, corrected for zero-suppression effects, and calibrated for the nonlinear response of the calorimeters. Finally the energy of neutral hadrons is obtained from the corresponding calibrated ECAL and HCAL energy.

The list of particles resulting from the operation of the PF algorithm on a whole event represents the best description of the event at the particle level, according to the information provided by the CMS detector and the intrinsic energy and position resolutions of the different sub-detectors. Figure 2.14 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 a 10% or less.

The PF approach to the event reconstruction also allows for a natural definition of jet objects. Once final state, well isolated leptons are excluded from the particle list, all that remains can be clustered into jets, as further explained in the following paragraph. In this approach, jets and leptons are naturally disentangled, since the same energy deposits or tracker hits cannot have contributed to the reconstruction of distinct objects.

Jets

A high-energy, coloured quark or gluon emitted in a hard proton-proton collision does not in the end appear in the detector. As it reaches large distances from the rest of the proton, the strong force potential favours the radiation of softer, often collinear gluons and quarks, until a point where a non-perturbative transition causes the partons combine into colourless hadrons. The result is a spray of more-or-less collimated particles, referred to as jet, which, due to energy

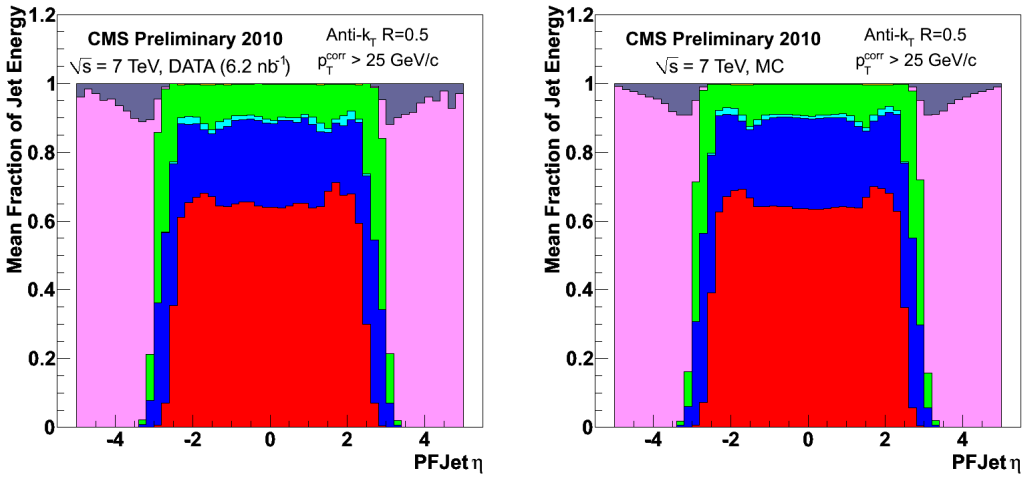


Figure 2.14: Reconstructed jet energy fractions as a function of pseudorapidity (a) in data (b) and in Monte Carlo. From bottom to top in the central region: charged hadrons, photons, electrons, and neutral hadrons. In the forward regions: hadronic deposits, electromagnetic deposits.

conservation, reflects at some level the energy and the flight direction of the initial parton.

Jets are detectable in modern experiments as a cluster 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.
Given the infinite probability that a collinear or soft gluon be emitted by a parton, jet algorithms must satisfy a few requirements, so that they can be used to provide finite theoretical predictions. The two conditions to be respected are the following:

- collinear safety: the outcome of the jet algorithm must not change if a particle of momentum p is substituted by two collinear particles of momentum $p/2$;
- infrared safety: the outcome of the jet algorithm must not change if an infinitely soft particle is added (or subtracted) to the list of particles to be clustered.

In CMS, the adopted clustering algorithm, which respects the two criteria above described, is the so-called anti-k_{T} [50]. This algorithm proceeds via the definition of two distances for each particle i in the list of particles, namely

$$\begin{aligned} d_{ij} &= \min\left(\frac{1}{p^2_{\mathrm{T}i}}, \frac{1}{p^2_{\mathrm{T}j}}\right)\frac{\Delta R^2_{ij}}{R^2} \\ d_{iB} &= \frac{1}{p^2_{\mathrm{T}i}} \end{aligned} \tag{2.1}$$

In the above equation, d_{ij} can be interpreted as the "distance" between the particle i and a generic other particle j among those still to be clustered, while d_{iB} represents the "distance" between the particle i and the beam line. ΔR_{ij} is the distance between the two particles in the $\eta - \phi$ plane, while R is the algorithm radius parameter.
The algorithm looks, for each particle i, if there is another particle j such that d_{ij} is smaller

than d_{iB}. If this happens, then particles i and j are recombined by adding together their four-momenta, otherwise the i particle is promoted to jet. The whole procedure is iterated and the algorithm stops when only jets are left.
It can be easily seen that particles at a distance greater than R from the jet axis are not clustered together with the jet itself, thus leading to the construction of cone-shaped jets. The standard radius parameter adopted in CMS, and then the approximate jet size in the $\eta - \phi$ plane, is 0.5.

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 physic aspect.

- Level 1 (offset) corrections: the purpose of this first step is to remove from the jet the additional energy coming from spurious particles produced in secondary pp interactions within the same bunch crossing or from the underlying event that randomly overlaps with the jet area. This correction is determined both in data and in Monte Carlo on a event-by-event basis. First of all, the charged component of a jet within the tracker acceptance can be removed from the jet calculating the impact parameter of all jet particles: those which are not compatible with the event primary vertex are not considered in the jet clustering algorithm. To further remove the contribution of neutral particles, or to correct jets 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[2] ($\rho_{\mathrm{PU}} = E/\Delta\eta/\Delta\phi$) 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) [51];
- 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 to the center of the barrel;
- Level 3 (absolute) corrections: this last correction factor correctly sets the jet absolute energy scale, and is derived from γ+jet 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.

Further details about the performances of PF jet reconstruction at CMS with the first 36 pb^{-1} of data collected in 2010 can be found at [52]. As an example, Figure 2.15 reports the jet energy resolution expected from the simulation and measured in data for PF jets reconstructed with an anti-k_{T} algorithm of R parameter 0.5 within the tracker acceptance.

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

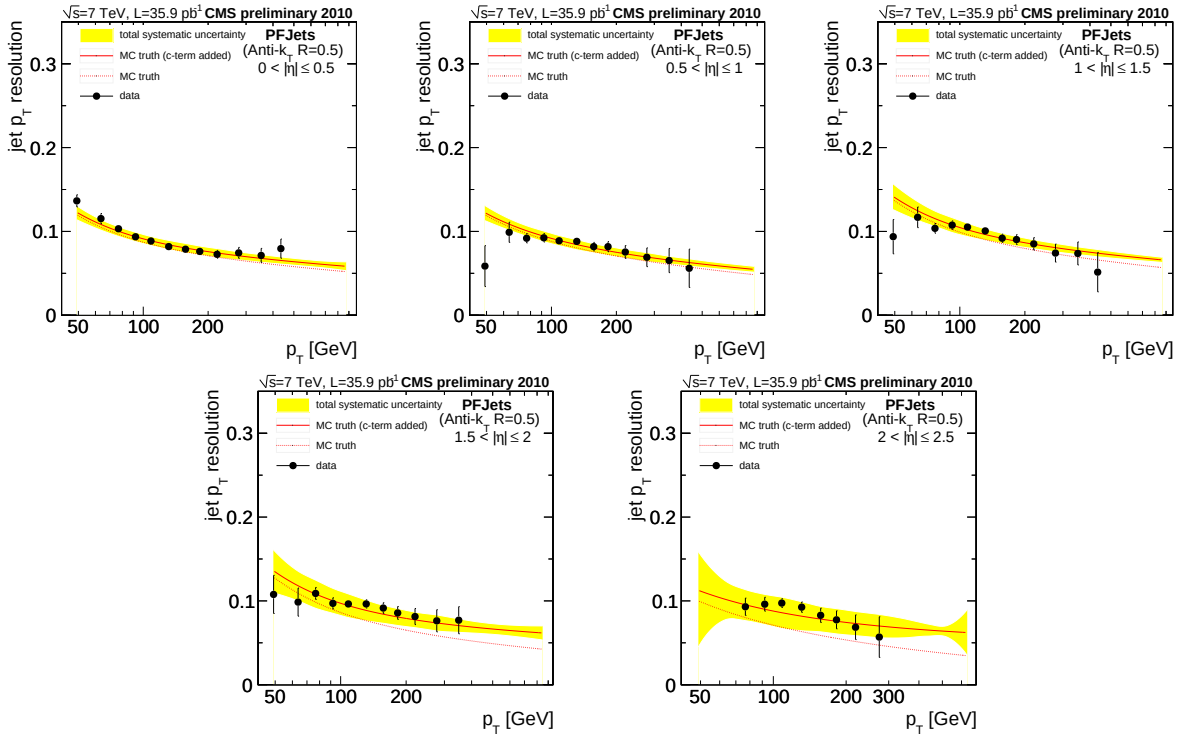


Figure 2.15: Data resolution measurements compared to the MC truth resolution before (red dashed line) and after correction for the residual discrepancy between data and simulation (red solid line) for PF jets in different η ranges.

Missing transverse energy

In general, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative of the vector sum of the transverse momenta of all final-state particles in the event. In the hypothesis that all detectable particles are properly reconstructed, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ coincides with the sum of the four-momenta of all undetectable particles (i.e. neutrinos, or BSM particles such as neutralinos in more exotic scenarios), since the initial pp collision occurs between two particles of negligible transverse momentum ($\lesssim 1\,\mathrm{GeV}$). In practice, this is not possible, since a fraction of the total event energy is unavoidably lost in the beam pipe or only coarsely reconstructed in the forward calorimeters. As a result, the measured $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is only an approximation of the neutrino transverse momenta.

Since all detector information is included in the PF-based reconstruction, it is simple to define PF $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ as the negative of the vector sum, over all PF candidates particles, of their transverse momenta. Jet factorized energy corrections are propagated to the missing transverse energy computation to improve its resolution.

Chapter 3

ECAL stability and calibration studies with isolated electrons

As anticipated in Section 1.4, the $H \to \gamma\gamma$ decay channel provides a clean final-state topology with a mass peak that can be reconstructed with high precision. In the mass range $110 < m_H < 150\,\text{GeV}$, the di-photon final state is one of the most promising channels for Higgs searches at the LHC. Since in this mass range the SM Higgs intrinsic width varies between 2.82 and 17.3 MeV, the electromagnetic calorimeter of CMS has been specifically designed to provide optimal energy resolution, so that the advantages of such a narrow resonance are maintained. Recalling the parameterization of the ECAL energy resolution

$$\frac{\sigma_E}{E} = \frac{A}{\sqrt{E(\text{GeV})}} \oplus \frac{B}{E(\text{GeV})} \oplus C \tag{3.1}$$

and the values of the stochastic term $A \simeq 2.8\%$ and the noise term $B \simeq 12\%$, it follows that the energy resolution for electrons or photons with $E_{e,\gamma} \gtrsim 50\,\text{GeV}$ is dominated by the constant term C (see Figure 3.1). This last term, whose target value in CMS is 0.55%, depends on the non-uniformity of the longitudinal light collection, the energy leakage from the rear face of the crystals, instabilities in the operation of ECAL and the intercalibration constants accuracy.

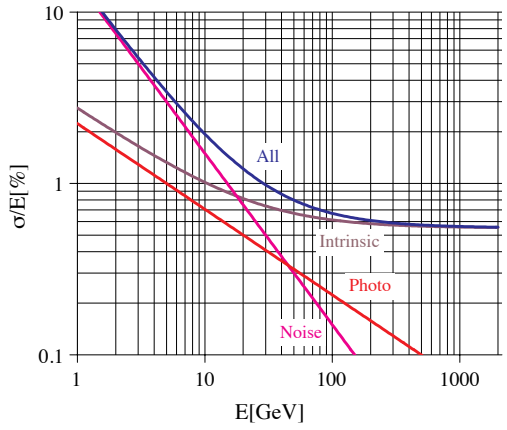
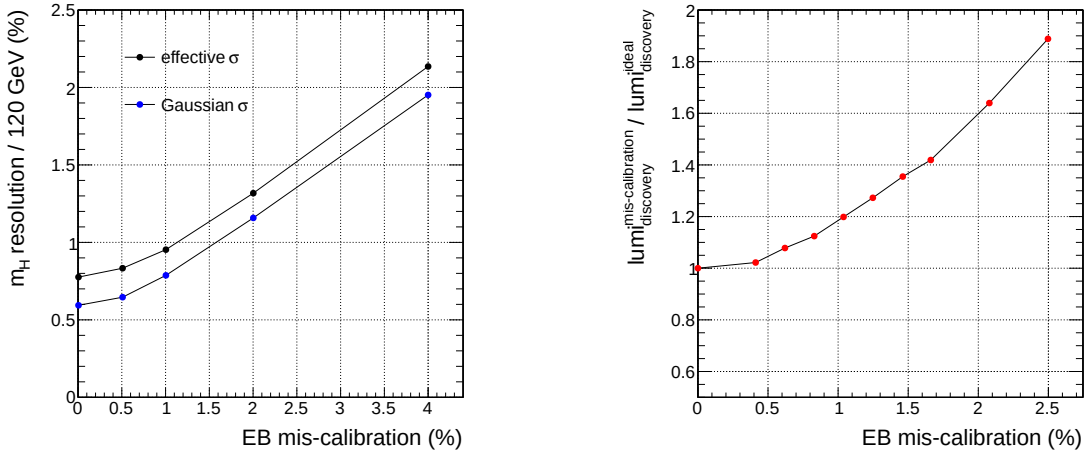


Figure 3.1: Different contributions to the energy resolution of the CMS $PbWO_4$ electromagnetic calorimeter.

As an example of the impact of the constant term of the energy resolution on a physics case, Figure 3.2 shows how an imperfect intercalibration of ECAL crystals affects Higgs searches in the $\gamma\gamma$ channel by worsening the resolution of the mass peak and, as a consequence, making it necessary to integrate a higher luminosity for a 5σ-significance discovery.



(a) (b)

Figure 3.2: Effect of a given mis-calibration level 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 a perfectly calibrated detector.

In this context, isolated electrons from $W \to e\nu_e$ decays have proven to be a valuable ensemble of events to study the ECAL response stability and uniformity and to provide a set of intercalibration coefficients complementary to the usage of other classes of events.
After a brief review of the ECAL energy calibration scheme, given in Section 3.1, the usage of such events is presented. Firstly, Section 3.2 shows the study and development of a dedicated trigger path, needed to allow the storage of $W \to e\nu_e$ events throughout the whole 2011 data-taking period. Section 3.3 is devoted to a general description of the analysis method, while in Sections 3.4 and 3.5 the results on the ECAL response stability and uniformity are presented. The impact of the relative scale corrections obtained with $W \to e\nu_e$ events on the ECAL energy resolution is assessed reconstructing the Z peak in $Z \to e^+e^-$ events, as explained in Section 3.6. Section 3.7 is devoted to the description of the intercalibration technique with isolated electrons. Finally, the $H \to \gamma\gamma$ discovery channel is briefly presented in Section 3.8 as the analysis where these efforts have channelled into.

3.1 ECAL energy reconstruction and calibration

We want here to briefly review how the energy $E_{e,\gamma}$ of electrons and photons is reconstructed in the electromagnetic calorimeter (an extensive description of this sub-detector and the rudiments of the clustering algorithms were given in Section 2.2.2). In formulae, the ECAL energy is given by

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

Here follows a brief description of each term in Equation 3.2.

- A_i: the energy deposited in each ECAL crystal is digitized by a sampling ADC. The result in each channel is a series of ten samples separated by 25 ns, for every accepted trigger. The delay of the read-out pipeline is such that the signal pulse is expected to start from the fourth digitized sample and the baseline pedestal value can be estimated from the first three ones. The amplitude A_i of the signal in channel i is then reconstructed using a linear combination of these samples, i.e. $A_i = \sum_{j=1}^{10} w_j \cdot s_{ij}$, where s_{ij} is the sample value in ADC counts and w_j is a weight. The weights are optimized using a reference pulse shape and they implicitly subtract the pedestal from the first three pre-samples [53];
- c_i: the intercalibration coefficient is meant to account for crystal-to-crystal variations in the response by equalizing the different crystal responses and normalizing them to an arbitrary reference value;
- $S_i(t)$: the monitoring correction factor to each crystal signal amplitude is needed because of the radiation-induced transparency change with time t;
- $\sum_i$: the ECAL crystals are approximately one Molière radius in transverse size, thus high energy electromagnetic showers spread over a few crystals. Clustering algorithms [47, 40] are used to add together energy deposits in adjacent crystals in order to reconstruct the original energy of electron or photon showers in the calorimeter. Furthermore, the presence of material in front of the electromagnetic calorimeter (corresponding to $1 - 2X_0$) causes conversion of photons and bremsstrahlung from electrons and positrons. The strong magnetic field of the experiment tends to spread this radiated energy in ϕ. The ECAL clustering algorithms are also designed to recover this energy. To summarize, the sum in Equation 3.2 is intended over ECAL crystals that are associated by the clustering algorithm;
- G: the global ECAL energy scale converts ADC counts into GeV. It is determined separately for the ECAL barrel and the endcaps from di-electron resonances such as $J/\psi \to e^+e^-$ and $Z \to e^+e^-$;
- $F_{e,\gamma}(E_T, \eta)$: small algorithmic energy corrections are required to correct for residual cluster non-containment effects, particularly shower leakage and bremsstrahlung losses which are dependent on the type of the particle, its momentum, direction and impact point position.

The studies presented in this chapter make use of isolated electrons to control the transparency corrections, needed to follow the radiation damage and recovery cycles during LHC operations (S_i term), provide valuable information on the response uniformity across the calorimeter ($F_{e,\gamma}$ term) and help improve the accuracy of the intercalibration coefficients (c_i terms).

3.2 Triggering $W \to e\nu_e$ events

As explained in Section 2.2.5, CMS relies on a two-level Trigger System to select the interesting events to be stored. The first step is the Level-1 (L1) trigger, whose purpose is to reduce the rate of selected events down to about 100 kHz.

In 2011, single electromagnetic objects of increasing energy have been used as L1 seeds for electrons from W decay. At this stage, the trigger choice is based only on the transverse energy of the calorimetric deposit, coarse shower shape requirements and the fraction of the total L1 seed energy belonging to the hadronic calorimeter, required to be smaller than 5%. Table 3.1 summarizes the evolution of L1 triggers used for the different instantaneous luminosities conditions, and the observed rates.

L1 seed	L ($\mathrm{cm^{-2}\,s^{-1}}$)	rate (kHz)
`L1_SingleEG15`	$5.0 \cdot 10^{32}$	2.9
`L1_SingleEG15`	$1.0 \cdot 10^{33}$	5.5
`L1_SingleEG15`	$1.4 \cdot 10^{33}$	7.1
`L1_SingleEG20`	$2.0 \cdot 10^{33}$	3.8
`L1_SingleEG20`	$3.0 \cdot 10^{33}$	5.5
`L1_SingleEG20`	$5.0 \cdot 10^{33}$	8.9

Table 3.1: Level-1 single electromagnetic objects used and their observed rates for the reference instantaneous luminosities of the 2011 trigger menu. The wording `L1_SingleEGXY` indicates a L1 seed of transverse energy threshold $E_{\mathrm{T}} > XY$ GeV.

In the second step of the CMS Trigger System, the High-Level Trigger (HLT), a fast version of the offline reconstruction sequence allows to implement precise selection algorithms based on finer granularity and higher resolution information provided by all CMS sub-detectors, with the intent of reducing the total output rate to about 300 Hz.
To trigger $\mathrm{W} \to \mathrm{e}\nu_{\mathrm{e}}$ events, HLT electron objects are built. At first, ECAL clusters are reconstructed (in regions around the Level-1 EG seeds previously identified) and merged together along the ϕ direction in superclusters to collect the energy radiated via bremsstrahlung photons. The supercluster transverse energy is required to exceed a threshold, with a value dependent on the HLT path. To reduce the contamination from QCD jets faking an electron object, cluster shape requirements are applied and the electron candidate is required to be isolated, i.e. the sum of energy deposits in ECAL and HCAL in a cone around the supercluster is required to be smaller than a given threshold.
Secondly, the energy and the position of the supercluster are back-propagated through the magnetic field all the way to the interaction point to search for a couple of compatible hits in the first layers of the pixel detector. The pixel hits are used as seeds for the whole track reconstruction, which makes use of a Kalman Filter technique[1]. Finally, once the electron track is calculated, additional cuts are applied on the electron candidate, both requiring the match between the track and the supercluster relative positions, and a further tracker-based isolation.

The sole requirement of an HLT electron object with $E_{\mathrm{T}} > 27(32)$ GeV and with the cut values defined in the first column of Table 3.2, labelled `VT-T`, allowed to efficiently select and save isolated electron events only in the very first part of the 2011 data taking, when the instantaneous

[1]Due to timing issues, the Kalman Filter technique is used at HLT reconstruction level, at variance with the full offline event reconstruction, when a GSF technique is used, as described in Seciton 2.3.1. This can originate differences between the online and the offline track parameters.

luminosity provided by LHC was $\sim 5 \cdot 10^{32}\,cm^{-2}\,s^{-1}$ or lower.
At higher instantaneous luminosities, triggering electrons from W decays becomes arduous. Recalling the inclusive production cross-section of W bosons at $\sqrt{s} = 7\,TeV$, which is (31300 ± 1600) pb, a bare rate of (31 ± 2) Hz is expected for $W \to e\nu_e$ events at an instantaneous luminosity of $3 \cdot 10^{33}\,cm^{-2}\,s^{-1}$. This estimate, which does not include the additional spurious rate due to fake electrons from multi-jet production, is already about half of the total CMS bandwidth. The implementation of different, lower rate HLT paths emerged as a priority to save as many W events as possible and avoid turning to prescaled[2] triggers.

3.2.1 The electron+PFm_T trigger

The strategy adopted to contain the rate was twofold. First of all, tighter electron identification and isolation requirements[3] were imposed on the electron leg of the trigger, mostly to reduce the contamination from fake QCD events and improve the purity of the HLT path. Two different working points were inspected. The looser one, labelled `WP80`, was used up to $L = 1.4 \cdot 10^{33}\,cm^{-2}\,s^{-1}$, while the tighter one, called `WP70` was required when the luminosity was raised up to $5 \cdot 10^{33}\,cm^{-2}\,s^{-1}$ at the end of the 2011 run. These working points, defined in such a way that their efficiency on real electrons is respectively 80 and 70%, are always looser or at least as tight as the cuts applied in the offline analysis, as will be outlined in Section 3.2. The exact cut values on the discriminating variables are reported here below (Table 3.2). Table 3.3 instead shows the improved purity of the trigger when using the tighter working points with respect to the paths implemented at the beginning of the year.

variable	`VT-T` cuts		`WP80` cuts		`WP70` cuts	
	EB	EE	EB	EE	EB	EE
$\sigma_{i\eta i\eta}$	0.011	0.031	0.010	0.030	0.010	0.030
$\lvert\Delta\phi_{tk-sc}\rvert$	0.070	0.050	0.060	0.030	0.030	0.020
$\lvert\Delta\eta_{tk-sc}\rvert$	0.008	0.008	0.004	0.007	0.004	0.005
H/E	0.050	0.050	0.040	0.025	0.025	0.025
I_{tk}/E_T	0.125	0.075	0.090	0.040	0.050	0.025
I_{em}/E_T	0.125	0.075	0.070	0.050	0.060	0.025
I_{had}/E_T	0.125	0.075	0.100	0.025	0.030	0.020

Table 3.2: Cut values for electron identification and isolation variables applied at trigger level on electron candidates for ECAL barrel (EB) and ECAL endcaps (EE). Three different working points are shown: `VT-T` stands for `CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT`, while `WP80(70)` are defined so to have an efficiency on real electrons of about 80(70)%.

The second expedient put in place to contain the rate to within ~ 10 Hz is the requirement of a reconstructed W transverse mass above a certain threshold. The event missing transverse

[2] By prescale, a random reduction factor of the trigger rate is meant, with the purpose of containing the trigger rate within affordable values.

[3] A description of the variables used to identify real electrons is postponed to the $H \to WW \to l\nu_l q\bar{q}$ analysis Chapter (Section 4.2).

working point	HLT rate (Hz)	$W \to e\nu_e$ rate (Hz)	purity
VT-T	80.4 ± 0.2	9.3	12%
WP80	35.5 ± 0.1	7.9	22%
WP70	22.8 ± 0.1	6.5	29%

Table 3.3: Expected total trigger rate and bare $W \to e\nu_e$ rate for an instantaneous luminosity of $3 \cdot 10^{33}\,cm^{-2}\,s^{-1}$, shown for three different identification and isolation requirements. On top of these selections, a $E_T^{ele} > 27\,GeV$ cut is assumed. From the ratio of the two rate estimates, the trigger purity is computed as $(W \to e\nu_e \text{ rate})/(\text{HLT rate})$.

energy is needed for the m_T calculation, and it is crucial to have it reconstructed with the best energy resolution attainable at trigger level, in order for the trigger to have a high efficiency on real $W \to e\nu_e$ events.
Since the complexity of the full Particle Flow (PF) algorithm prevents it from running at HLT level, a lighter version is run to reconstruct jets. At first, a fast calorimetric jet reconstruction is performed using only the energy deposits in ECAL and HCAL and neglecting the tracker information. Then, to save CPU time, the PF reconstruction algorithm is operated only in those regions where a calorimetric jet above a certain energy threshold is found: information from the calorimeters as well as the tracker and the muon systems are combined at this stage to build PF candidates, which are successively clustered using an anti-k_T algorithm to form jets. The PF-based missing transverse energy[4] H_T^{miss} is then obtained as the negative sum of all reconstructed PF jets. The HLT electron object and PF H_T^{miss} are combined to get the W transverse mass according to the following formula:

$$m_T = \sqrt{2E_T^{ele} H_T^{miss} \cdot (1 - \cos\Delta\phi)} \quad , \tag{3.3}$$

where $\Delta\phi$ is the angle between the electron momentum $\vec{p}_T^{ele}$ and the $\vec{H}_T^{miss}$ in the transverse plane.

In order to determine the optimal trigger working point and define the operating electron E_T and PF m_T thresholds for the forthcoming LHC runs at higher luminosity, LHC data taken at the beginning of 2011 are used, in particular, data from CMS run 163374, collected at an instantaneous luminosity of $3.44 \cdot 10^{32}\,cm^{-2}\,s^{-1}$. While a trigger path with an electron E_T threshold of 27 GeV/c and a VT-T working point is actually implemented in that run, an offline emulation of the High-Level Trigger is used to study the performance and compute the new expected rates after applying different E_T cuts, alternative working points or an additional cut on m_T. The rates of the new tested HLT paths are linearly projected to higher instantaneous luminosities, thus neglecting non-linear effects (e.g. the contribution of pile-up, which constantly increased throughout the 2011 data taking). Figure 3.3 shows the rate dependance as a function of the HLT electron E_T or W transverse mass m_T, for the three different working points under study. The rate reduction granted by the tighter identification and isolation requirements is sizable, although not enough to contain the rate within 10 Hz in a high-luminosity scenario. A combined requirement on E_T and m_T is necessary, and the optimum working point is defined from

[4] H_T^{miss} differs from the full event missing transverse energy E_T^{miss} since it is computed as the negative sum only of PF jets above a certain energy threshold.

the plots in Figure 3.4 and visualized by a black marker in the figures.

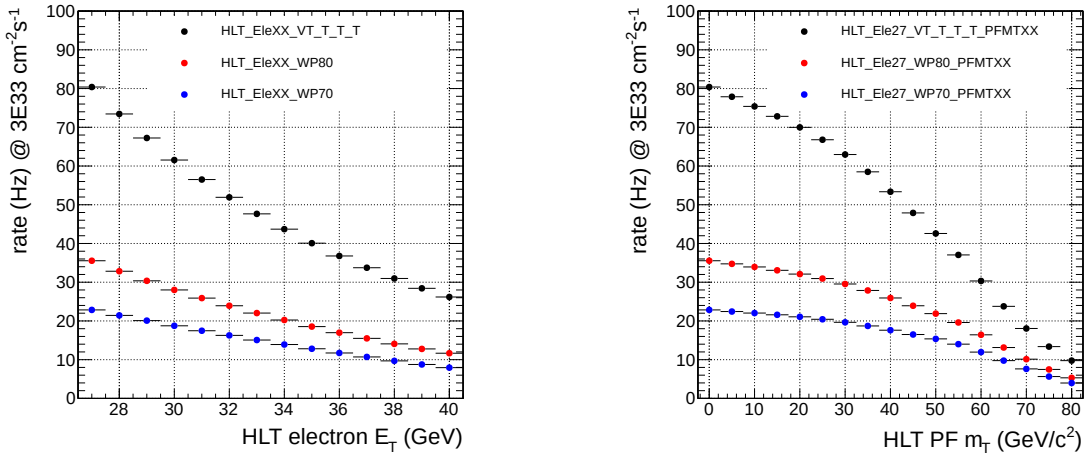


Figure 3.3: Rate evolution in dependence of (a) HLT electron E_T and (b) HLT PF-based m_T for three different identification and isolation working points (`VT-T` in black, `WP80` in red and `WP70` in blue). Rate estimates correspond to a projected instantaneous luminosity of $3.0 \cdot 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

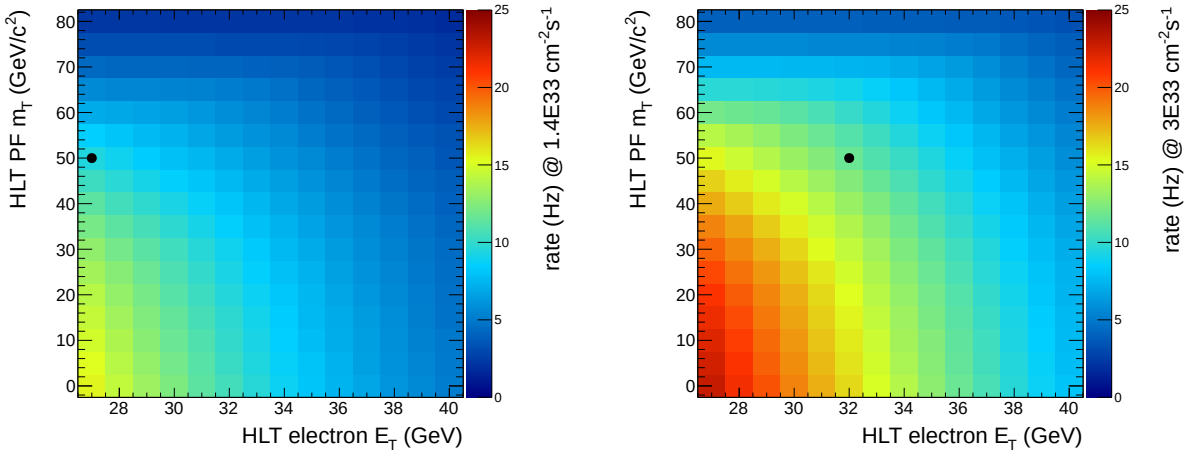


Figure 3.4: Rate evolution as a combined function of HLT electron E_T and m_T for a projected instantaneous luminosity of (a) $1.4 \cdot 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and (b) $3.0 \cdot 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In both cases, a `WP70` working point is assumed. Black dots indicate the final chosen configuration.

3.2.2 Summary of the trigger strategy

Thanks to the studies here reported, an optimal configuration was defined for each evolution of the CMS trigger menu, following the increases in instantaneous luminosity provided by the LHC mahine. In Table 3.4, a summary of the final chosen working points is reported. Once deployed online, it was possible to measure the effective rate of the new defined HLT paths on data: as shown in Figure 3.5, the measurements turned out to be consistent with the projections, within the intrinsic uncertainty of the method. In particular, the disagreement

observed towards the end of the data taking can be ascribed to the high pile-up conditions: rate estimates are in fact based on a low pile-up run, and an elevated number of secondary interactions can reduce the efficiency of the H/E and the isolation cuts on the HLT electron object, thus resulting in a lower than expected rate.

HLT path	L ($\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$)
`HLT_Ele27_VT-T_v1(v2,v3)`	$5.0 \cdot 10^{32}$
`HLT_Ele32_VT-T_v3`	$1.0 \cdot 10^{33}$
`HLT_Ele25_WP80_PFMT40_v1`	$1.0 \cdot 10^{33}$
`HLT_Ele27_WP80_PFMT50_v1`	$1.4 \cdot 10^{33}$
`HLT_Ele32_WP70_PFMT50_v3`	$2.0 \cdot 10^{33}$
`HLT_Ele32_WP70_PFMT50_v3(v4)`	$3.0 \cdot 10^{33}$
`HLT_Ele32_WP70_PFMT50_v4(v8,v9)`	$5.0 \cdot 10^{33}$

Table 3.4: HLT single electron paths used for the reference instantaneous luminosities of the 2011 trigger menu. The wording `HLT_EleXY` indicates that a transverse energy threshold $E_{\mathrm{T}} > XY$ GeV is applied on the electron object. The cut on the transverse mass, when present, is indicated as `PFMTXY`. The electron identification and isolation working point `VT-T` is detailed in Table 3.2, as well as the `WP80(70)` working points.

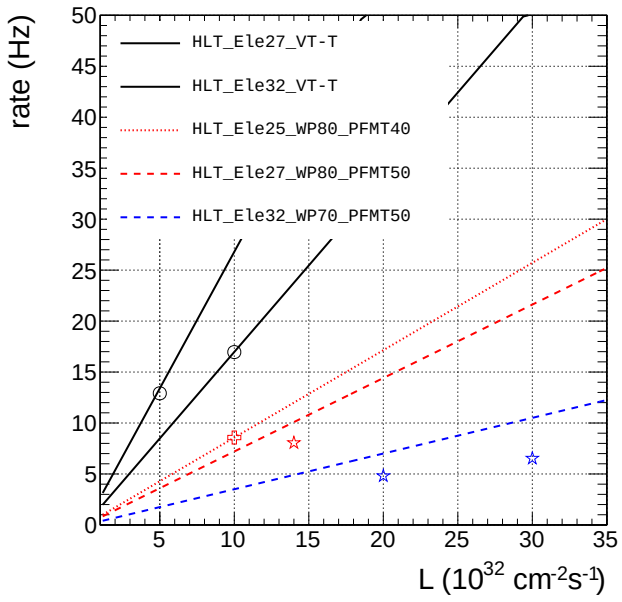


Figure 3.5: Rate projections of different HLT paths. Solid lines do not include a cut on m_{T}, while dotted (dashed) lines correspond to a $m_{\mathrm{T}} > 40(50)$ GeV/c^2 cut. Colors identify different identification/isolation working points: `VT-T` is in black, `WP80` in red and `WP70` in blue. Markers correspond to the measured rates on real data for some reference instantaneous luminosities.

3.3 The analysis method

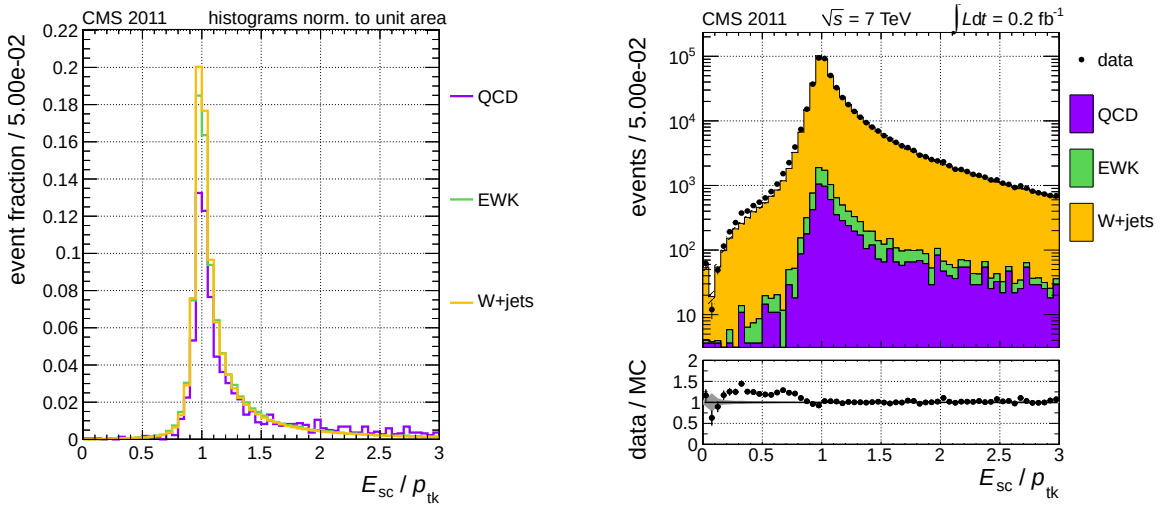
3.3.1 Event selection

With the trigger menu above described, it was possible to collect isolated electrons from W boson decays in a non prescaled fashion throughout the whole 2011 LHC run. The data collected correspond to an integrated luminosity of $5.0\,\text{fb}^{-1}$. As will be shortly described in detail, the analysis technique relies on the comparison of the electron energy as measured by the ECAL supercluster E_{sc} and by the tracker momentum p_{tk}. The event selection is then driven by the necessity of suppressing all background sources for which the $E_{\text{sc}}/p_{\text{tk}}$ distribution is broader than for real electrons. In particular, the following requirements are applied:

- exactly one electron with $p_{\text{T}} > 30\,\text{GeV}/c$ and satisfying the offline `WP70` electron identification criteria (see Table 3.2);
- electron isolation performed at HLT is strengthened by requiring a relative combined isolation (i.e. the sum of the tracker, ECAL, and HCAL deposits in a cone of size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.03$ around the electron track divided by the electron p_{T}) is required to be smaller than 0.04 (0.03) for electrons in the ECAL barrel (endcaps); the isolation is corrected on an event-by-event basis subtracting the spurious contribution from pile-up events, computed as $\rho_{\text{PU}}\pi(\Delta R)^2$ (for the definition of ρ_{PU}, see Section 2.3.3);
- in order to kill most of the contamination of fake electrons from QCD events, a cut on the event missing energy $E_{\text{T}}^{\text{miss}}$ and the W boson transverse mass m_{T} is applied, namely $E_{\text{T}}^{\text{miss}} > 25\,\text{GeV}$ and $m_{\text{T}} > 50\,\text{GeV}/c^2$. The full Particle Flow algorithm is used to construct the offline missing transverse energy object;
- a cut on the opening angle $\Delta\phi$ between the electron and the $E_{\text{T}}^{\text{miss}}$ direction in the transverse plane, required to be greater than $\pi/2$, further reduce the QCD contamination by a factor of about two, being its distribution nearly flat in $\Delta\phi$. $\text{W} \to \text{e}\nu_{\text{e}}$ events, instead, are mostly produced with no other associated hard jets (the reduction factor in the W cross-section for each extra jet is naively $\alpha_S \simeq 0.12$), and therefore the electron and the $E_{\text{T}}^{\text{miss}}$ tend to be emitted back-to-back in the transverse plane.

The first $211\,\text{pb}^{-1}$ collected at the beginning of 2011 were triggered by the `HLT_Ele27_VT-T` path, with no requirements on m_{T} and with loose requirements on electron identification and isolation. Therefore they constitute a useful sample to check the efficiency of the described selections on real electrons, and to assess the level of fakes in the final analysis sample. In Figure 3.6a, the $E_{\text{sc}}/p_{\text{tk}}$ distribution is compared for $\text{W} \to \text{e}\nu_{\text{e}}$ events and other background sources. Events with a jet from QCD production being reconstructed as a fake electron constitute a potentially dangerous background, being the $E_{\text{sc}}/p_{\text{tk}}$ shape substantially broader than for real electrons. Anyway, as Figure 3.6b shows, the QCD contamination in the final sample amounts to a negligible fraction of the total selected events ($\sim 1\%$). To get the proper shape and normalization of QCD events, a data-driven approach has been followed. It consists in the inversion of the isolation requirements to define a sample enriched in multi-jet events, from which the QCD template shape is obtained, and in the fit of the $E_{\text{T}}^{\text{miss}}$ distribution to obtain the event yield. A similar technique is used and further described later in Section 4.7.
As estimated by the Monte Carlo simulation, the event selection described allows to retain

about 22% of all $W \to e\nu_e$ events.



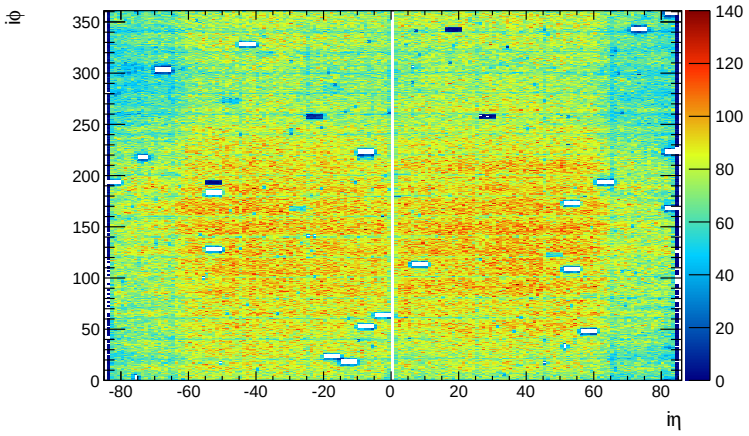
(a) MC (b) data/MC comparison

Figure 3.6: (a) E_{sc}/p_{tk} distribution for $W \to e\nu_e$ events, QCD fake electrons and all other electroweak background sources summed together (mainly Z+jets and $t\bar{t}$) as expected from the Monte Carlo simulation (the QCD shape is data driven). Each histogram is normalized to unity area so to allow a shape comparison of the different event topologies. (b) the data vs. Monte Carlo comparison for the E_{sc}/p_{tk} distribution is shown for the first 211 pb^{-1} of data collected in 2011. As can be noticed, the contamination from non $W \to e\nu_e$ events is kept at a negligible level.

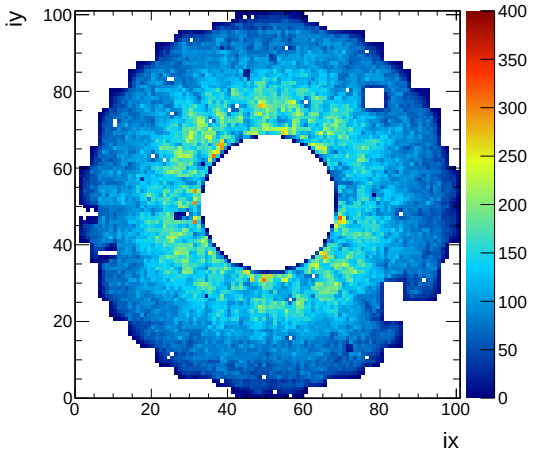
In Figure 3.7, the occupancy of electron seeds is shown for the full 2011 dataset after applying the offline $W \to e\nu_e$ selections described above. A total of 4.6 million events in the barrel and 1.3 million events in the two endcaps are at disposal for the analysis. The observed modulation of the occupancy in the barrel along the ϕ direction can traced back to noise asymmetries in the ECAL and HCAL calorimeters. The missing transverse energy in $W \to e\nu_e$ events gets contribution from the electron transverse energy, hadronic activity in the event and calorimeter noise. The ϕ distribution of the latter term being asymmetric, the final effect is an asymmetry in our selection efficiency. This inefficiency, anyway, does not induce any bias in the analysis of the ECAL stability and uniformity.

Another source of real electrons, in addition to $W \to e\nu_e$ events, are Z boson decays. Since the statistics added by $Z \to e^+e^-$ to the large W electron sample is modest, these electrons have rather been used in the following studies as a complementary control sample to test the effectiveness of the corrections computed. Given the practically negligible level of QCD contamination in di-electron final states, the selections applied to save $Z \to e^+e^-$ events are looser. In particular, the following requirements are applied:

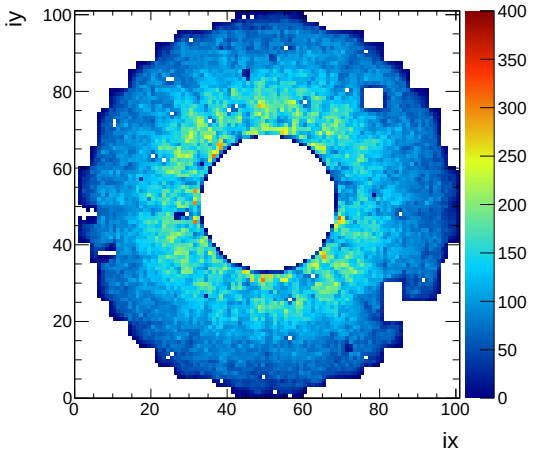
- exactly two electrons with $p_T^{max} > 20\,GeV/c$, $p_T^{min} > 12\,GeV/c$ and satisfying offline `WP80` electron identification criteria (see Table 3.2);
- combined relative isolation (after pile-up correction) smaller than 0.07 (0.06) for electrons in the barrel (endcaps);
- $E_T^{miss} < 40\,GeV$;
- the two electrons must sit on the Z resonance, i.e. $60 < m_{ee} < 120\,GeV/c^2$.



(a) EB



(b) EE+



(c) EE-

Figure 3.7: Electron seed occupancy in EB, EE+ and EE- corresponding to the full 2011 dataset after applying $W \to e\nu_e$ offline selections (see text for details).

3.3.2 The template fit

The ratio of the supercluster (sc) energy of an electron measured by ECAL and the momentum measured in the tracker (tk) has been exploited to monitor the stability and uniformity of the ECAL response. The analysis gives an unbiased answer on the electron energy measurement, assuming that the measurement of the electron momentum is stable in time and uniform. This assumption – definitely not straightforward in the case of uniformity studies versus η, given the structure of the tracker – will be challenged in the discussion of the analysis results.

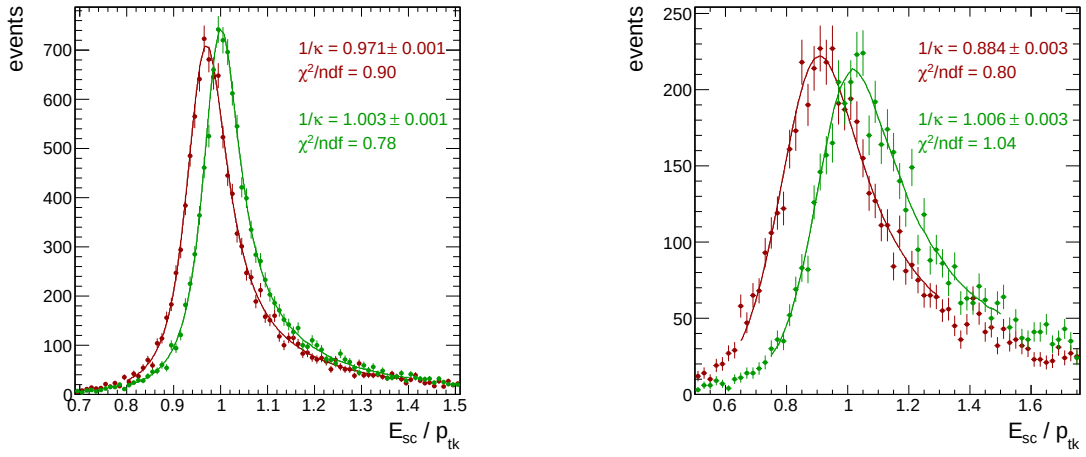
We are interested in relative variations of the response rather than in the absolute response, the latter being fixed in ECAL via the invariant mass of di-electron resonances. The analysis strategy therefore relies on the construction of a reference distribution $\mathcal{T}$, often referred to as "the template", describing the E_{sc}/p_{tk} observable at a given time or position. This distribution is then scaled to fit subsets of data, properly partitioned in time or position, in order to measure the response in each subset relative to the reference. To avoid biases related to the imperfect description of the data by the simulation, the reference distribution has been in general derived

from data themselves. Monte Carlo samples have been sometimes used for cross checks.
In formulae, we can describe the generic function f adopted in the analysis to fit the data in a subset with N events as

$$f(x;\kappa) = N \cdot \kappa \ (\mathcal{T} * \mathcal{R})(\kappa x) \tag{3.4}$$

where $x = E_{sc}/p_{tk}$, $\mathcal{T}$ is the reference distribution and κ is a dilation factor, whose reciprocal $1/\kappa$ can be interpreted as the ECAL energy scale of the subset relative to the template, assuming the invariance of p_{tk}. The reference distribution is assumed to have unit area, i.e. $\int_0^\infty \mathcal{T}(x)\mathrm{d}x = 1$, so that f is naturally normalized to the population of the subset and κ is the only parameter floated in the fit. The template function is convoluted with a proper normalized resolution function $\mathcal{R}$ that allows the fit function to describe subset of data with different resolutions than the reference distribution (provided the latter is properly sampled from a dataset with better resolution). In general, the $\mathcal{R}$ term in Equation 3.4 has been collapsed to a Dirac's delta. Nevertheless, it is useful when data are analyzed in regions where electrons are poorly reconstructed (e.g. intermodule cracks) or when fitting events with large energy scale variations within the same subset. This cases will be specifically described.

The shape of the E_{sc}/p_{tk} distribution can vary with varying datasets and selections. It is thus important that the template is built using a consistent set of selection, and that its shape is representative of the whole set of sub-samples being analyzed. This has been measured by monitoring the χ^2 of the best-fits. The analysis has been typically split in ECAL barrel (EB) and ECAL endcaps (EE), as the energy and the momentum resolution are noticeably different in these two regions, and strongly η dependent. In some cases, further splittings in smaller η regions have been necessary. Examples of E_{sc}/p_{tk} template fits in two representative subsets from EB and EE are shown in Figure 3.8.



(a) EB (b) EE

Figure 3.8: Examples of E_{sc}/p_{tk} distributions in (a) ECAL barrel and (b) ECAL endcaps with the best-fit reference distribution superimposed (line). In this examples, the reference distribution is sampled from a calibrated set of data and fitted to a subset of the same data before calibration (red) and after calibration (green).

The method outlined above in this section can easily be extended to any observable sensitive

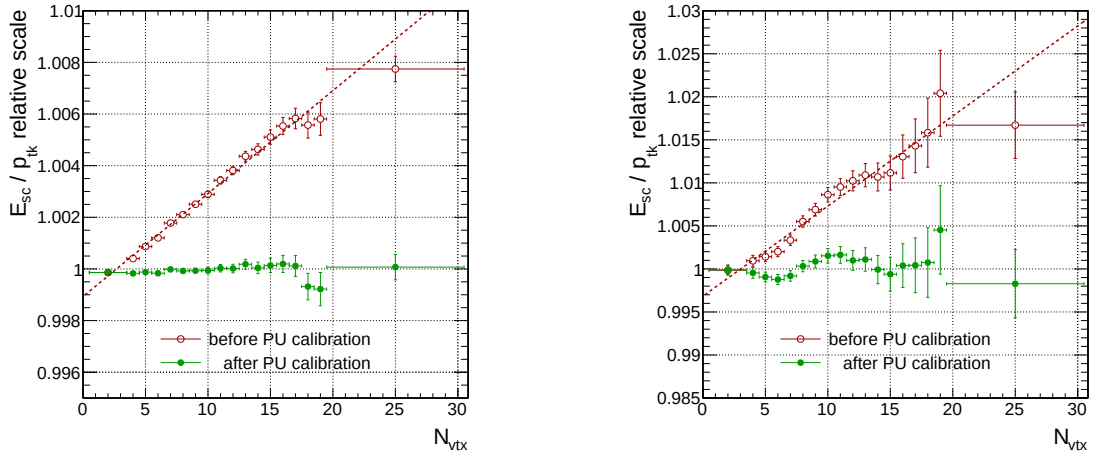
to the energy response of ECAL. Some of them have been used for cross checks, as mentioned where appropriate.

3.3.3 Calibration of the fitting procedure

Pileup correction

The beam conditions have varied throughout the 2011 run. The number of pile-up (PU) interactions per beam crossing has increased from a few units at an instantaneous luminosity $L \lesssim 5.0 \cdot 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at the beginning of 2011, to about 15 at $L \simeq 5.0 \cdot 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at the end of the LHC run. The probability that energy deposits in ECAL generated by particles from PU vertexes randomly overlap with the electron shower increases with the number of PU interactions. This affects the mean value of the E_{sc}/p_{tk} ratio and results in a biased determination of the ECAL response, if not properly corrected for.

The effect is visible in Figure 3.9, where the evolution of the relative E_{sc}/p_{tk} scale as a function of the number of reconstructed vertices in each event (N_{vtx}) is shown, N_{vtx} being directly proportional to the number of PU interactions per beam crossing. The relative scale variation has been measured with the analysis procedure described above: first of all, a reference E_{sc}/p_{tk} distribution is built using all the events collected in the year in which $N_{vtx} \leq 3$; then, data are split in subsets of events with a fixed number of reconstructed vertices, and the reference distribution is fitted to the E_{sc}/p_{tk} distribution in each subset. For these fits, a Gaussian function centered in zero and with varying σ has been used as resolution function, i.e. $\mathcal{R} \sim \mathcal{G}(0,\sigma)$, to account for energy resolution degradation with an increasing number of spurious interactions. The σ parameter is fitted simultaneously with κ in each N_{vtx} bin. The energy scale has been found to vary by about 0.04% and 0.1% per reconstructed vertex in the ECAL barrel and endcap respectively.



(a) EB (b) EE

Figure 3.9: E_{sc}/p_{tk} relative scale in (a) ECAL barrel and (b) ECAL endcap as a function of the number of reconstructed vertices. Points in red show the trend of the ECAL relative scale vs N_{vtx} if no correction for the pile-up is applied, while green points show the effectiveness of the correction.

The impact of pile-up depends on the cluster size and, in turn, on the electron quality. The clustering in ECAL [47, 40] is not PU-resilient: the dynamic addition of secondary clusters

along the ϕ direction to the electron supercluster, in order to recover the energy emitted by the electron via bremsstrahlung in the tracker material, is prone to the collection of PU clusters. PU-resilient clustering algorithms, intended to mitigate the effect of the pile-up on the photon resolution, are being developed, although they still need a further tuning.
For the purpose of the following analyses, however, an effective correction derived from an empirical fit of the data points in Figure 3.9 is sufficient.

Calibration of the momentum scale

At variance with the stability studies, the study of the response uniformity across ECAL with the E_{sc}/p_{tk} method requires a precise calibration of the relative response of the tracker. The tracker material varies sizably as a function of the pseudorapidity η, and the tracker structure has also a modularity in the azimuth direction ϕ. Both effects can have an impact on the momentum reconstruction. The response to muons and heavier particles cannot be directly applied to electrons, as a different tracking algorithm is used for the latter to account for the possible bremsstrahlung emission in the tracker material.

In order to derive a position-dependent calibration of the tracker response to electrons, di-electron pairs from $Z \rightarrow e^+e^-$ decays have been assigned to a given η (or ϕ) position if at least one electron was within a given acceptance around that position. The invariant mass of the pairs have been reconstructed by using the tracker momentum for that electron and the supercluster energy for the other electron, on which no acceptance restrictions were placed. The momentum scale at the given position is thus proportional to the square of the invariant mass according to:

$$m^2_{e^+e^-} = 4 \cdot E_{sc} \cdot p_{tk}(i\eta) \cdot \sin^2(\theta/2) \quad , \tag{3.5}$$

where we have specialized to the case of the η coordinate. In Equation 3.5, θ is the opening angle between the electron and the positron tracks; $i\eta$ is the pseudorapidity index, identifying one ring of ECAL crystals at the same η, to which the supercluster seed belongs. A momentum calibration factor for the E_{sc}/p_{tk} method can then be obtained from the relative $m^2_Z/m^2_{e^+e^-}$ scale which, in turn, is proportional to $1/p_{tk}$.
As the momentum resolution varies sizably along η, seven reference distributions have been derived by assigning e^+e^- pairs to seven pseudorapidity regions. These regions, symmetric around $\eta = 0$ correspond to the four modules of the ECAL barrel (up to $|i\eta| = 85$), and to three additional regions in the endcaps made by ten-crystal wide rings on both sides of ECAL (starting from $|i\eta| = 88$). The momentum scale at each $i\eta$ relative to the reference has been derived by fitting the proper reference distribution (the one of the closest $|\eta|$ region) to the invariant mass distributions reconstructed in each $i\eta$ ring, both in data and Monte Carlo.

The result of this study is shown in Figure 3.10, where a fair agreement between data and Monte Carlo is visible, with the exception of the positive side of the endcap ($i\eta > 85$). Peculiar structures are observed in both the endcaps and in the fourth modules of the barrel, where the material budget is large. In these regions, the electron momentum mis-calibration is at the percent level. Artificial discontinuities in the analysis are also induced by the usage of a discrete set of reference distributions. As the same modularity is adopted in the E_{sc}/p_{tk} analysis (see Section 3.5), these discontinuities cancel out in the derivation of the ECAL regional scale

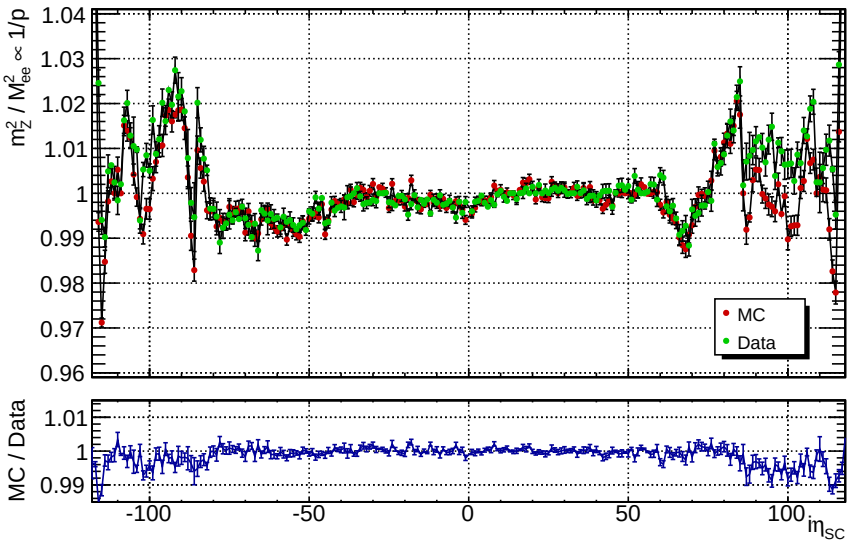


Figure 3.10: Momentum scale as a function of the supercluster position along η in data and Monte Carlo and data/MC ratio. The vertical lines mark the intervals inside which the same reference distribution was used.

relative to the Monte Carlo.

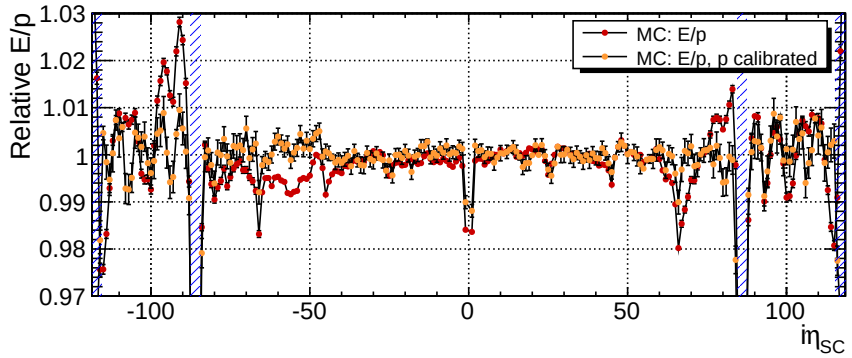


Figure 3.11: E_{sc}/p_{tk} relative scale versus η in Monte Carlo before and after the calibration of the tracker momentum. The events in the shaded regions at the limit of the ECAL or tracker acceptance are excluded from further analysis.

As a closure test, the E_{sc}/p_{tk} analysis as a function of η has been performed on a simulated sample with and without applying the momentum calibration procedure (see Figure 3.11). Noteworthy, the relative scale as a function of η is flat in Monte Carlo after p_{tk} calibration [5]. This proves that the structures observed in the E_{sc}/p_{tk} distribution are related to p_{tk}, and that the energy reconstruction in ECAL is consistently tuned in the Monte Carlo. This result indicates that the momentum calibration technique is applicable to data, and that a rescaling (where needed) of regional ECAL energy scale in data to match the regional scale in Monte Carlo is equivalent to a regional calibration of the ECAL detector response.

[5] Up to a possible small residual modulation in the endcaps with a period of about 7 crystals.

3.4 ECAL response stability

3.4.1 Correlation between transparency loss and signal loss

The most critical aspect in terms of the ECAL stability is represented by the monitoring of the crystal transparency variation under irradiation, as already pointed out in Section 2.2.2. While The $PbWO_4$ scintillation mechanism is not affected by the irradiation, the crystal transparency deteriorates, depending on the dose rate. The ECAL laser monitoring (LM) system [40, 54, 38] is specifically designed to accurately monitor and correct for changes in the response due to the radiation-induced crystal transparency variation. The crystal transparency is measured injecting laser light through optical fibers in each ECAL barrel and endcap crystal (respectively from the crystal front and rear face) and normalizing each crystal response to the laser pulse magnitude, measured using silicon PN photodiodes.

The number of reflections inside the crystal before reaching the photo-detector is different in the case of the laser monitoring light and of the scintillation light. In this condition, electromagnetic shower signals S and the crystal transparency R at a given time t (i.e. after a given irradiation) are not simply proportional, their relation being better described by a power law:

$$\frac{S(t)}{S(t_0)} = \left[\frac{R(t)}{R(t_0)}\right]^{\alpha} \quad , \tag{3.6}$$

holding while the loss $\Delta S = S(t_0) - S(t)$ is small (a few percent) [38]. Here, t_0 represent an initial time before irradiation. The α index has to be determined empirically and depends on the different optical paths, absorption cross sections, and light collection efficiencies of the monitoring and scintillation lights. Samples of a few tens of crystals from different production batches and in their final assembly have been irradiated at test beams, and the average spectral index of Equation 3.6 has been measured to be $\alpha \simeq 1.52$ for crystals produced at BTCP[6] and $\alpha \simeq 1.00$ for crystals produced at SIC[7] [38]. Consistent values have been obtained for the crystal in the ECAL barrel and endcaps, notwithstanding the differences in geometry and photon detectors. The α index has been found to be fairly reproducible in different crystals: the RMS spread of the measurements is less than 10% for the crystals in the ECAL barrel, and somewhat larger (about 20%) in ECAL endcap crystals.

Assuming the mean value and RMS spread of the available measurements of α as representative of the entire crystal population, and given the observed transparency losses in 2011, we estimate that the correction precision with one fixed value of α for all the BTCP crystals (the vast majority) and one for the SIC crystals should be better than about 0.2% in ECAL barrel, but not better than about 2% in the endcaps. Moreover, Equation 3.6 is strictly valid for small losses. An *in situ* measurement of α at the single crystal level, and as a function of the observed loss, is thus of utmost importance.
As an additional complication, the monitoring response is actually sensitive both to transparency variations and to changes in the response of the photo-detectors. The latter can still be parameterized with Equation 3.6, but with $\alpha = 1$. This would result in an effective α lower than 1.52 for BTCP crystals. Transparency losses and photodetector response variations are not

[6] Bogoroditsk Plant of Technochemical Products
[7] Shanghai Institute of Ceramics

easily disentangled by the current monitoring system.

3.4.2 Time stability results

The correctness of the correction procedure has been verified studying the stability of the ECAL response as a function of time. Since in this analysis we are interested only in relative variations of the ECAL response, the reference distributions (one in each analysis region) have been derived from data themselves exploiting the entire 2011 data sample. The same data have been partitioned in time-ordered subsets containing approximately the same number of events each[8] and the relative ECAL scale in each bin has been derived by fitting the appropriate reference distribution to each subset.
Results for both the ECAL barrel and endcaps obtained with the latest calibration conditions[9] over the entire 2011 data sample are shown in Figure 3.12. In each time bin, both transparency-corrected data (solid green dots) and data without LM corrections applied (empty red dots) are fitted to get the relative energy scale. Uncorrected data help visualize the ECAL crystal transparency loss trend with time, while from the spread of corrected data the LM contribution to the ECAL constant resolution term can be derived.
Especially in the endcaps, where the average transparency losses are large, events populating the same time bin can have undergone different transparency losses. The proper reference distribution for uncorrected data is therefore obtained upon using the distribution of LM corrections in that bin as resolution distribution $\mathcal{R}$ in Equation 3.4. Moreover, in order to mitigate the potential bias of pile-up, whose conditions changed heavily over time, the effective correction described in the previous section has been applied.

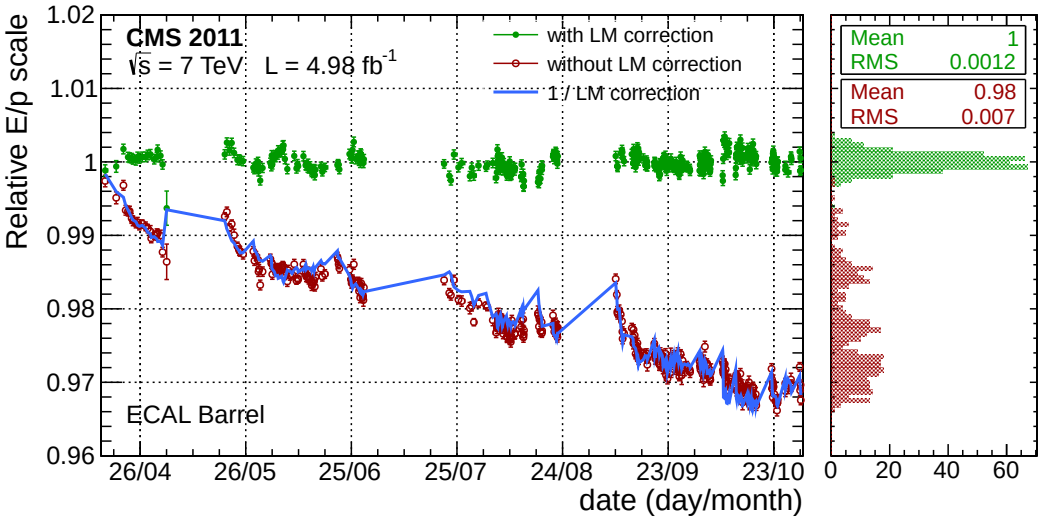
The analysis has been repeated and refined several times, and proved to be one of the pillars of the commissioning phases of the transparency corrections during 2010 and 2011. Upon application of the latest transparency corrections, the RMS stability of the ECAL barrel response, averaged over all the channels, is measured 0.12% (0.45%) over the $5.0\,\text{fb}^{-1}$ of integrated luminosity of the 2011 run in the ECAL barrel (endcaps). In the endcaps, the somewhat larger spread is the result of a systematic trend giving an overcorrection of about 1.5% at the end of the year as compared to the start of the data taking, indicating the need of further refinements of the calibration procedure.

3.4.3 *In situ* α measurement

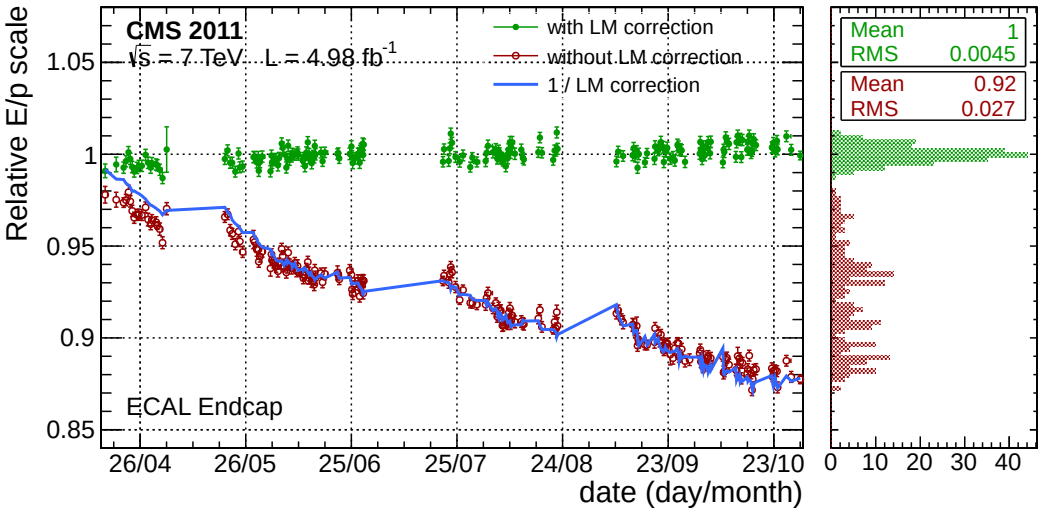
An *in situ* measurement of α is possible through the study of the correlation between the transparency loss seen by the monitoring system and the signal loss observed on a reference signal. To this purpose, the single electron data sample has been exploited as follows:

- the signal attenuation $S(t)/S(t_0)$ due to crystal transparency loss is measured by reconstructing electrons from $W \to e\nu_e$ decays and using $E_{sc}^{unc.}/p_{tk}$ as test variable, where $E_{sc}^{unc.}$ is the energy of the supercluster reconstructed without applying LM corrections;
- for each electron candidate, an effective transparency loss $R(t)/R(t_0)$ is obtained from the weighted average of the response losses measured by the laser monitoring

[8] The approximation comes form the constraint that none of the subset should span over a technical stop.
[9] GlobalTag `FT_R_42_V24` used for $H \to \gamma\gamma$ data re-processing of January16th 2012



(a) EB



(b) EE

Figure 3.12: ECAL relative energy scale evolution obtained from electron E_{sc}/p_{tk} fits. The plot shows the results for (a) ECAL barrel and (b) ECAL endcaps during the whole 2011 LHC run. Uncorrected data (red empty dots), data corrected for the observed transparency loss (green solid dots) and the inverse of the correction derived from monitoring data (blue line) are displayed.

system in all the crystals of the supercluster, where the fraction of the supercluster energy seen by the crystal is used as weight for that crystal;

- the dataset is divided in subsets ordered as a function of the effective transparency loss and a reference distribution, derived from the same data upon inclusion of LM corrections, is fitted to each subset to extract the relative scale of uncorrected data $1/\kappa^{\text{unc.}}$. Different reference distributions have been used in different η regions;
- the parameter α is the extracted from these data after linearization of the power law

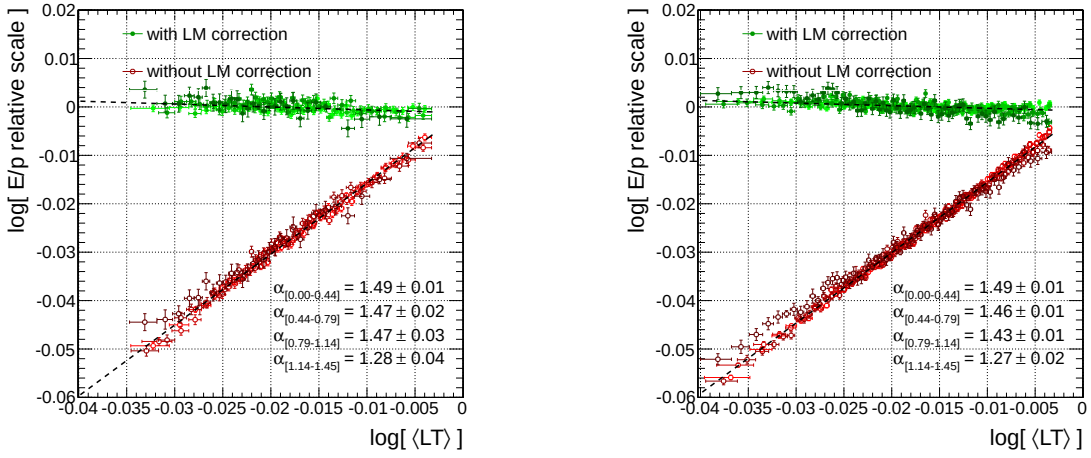
of Equation 3.6:

$$\log\left[\frac{S(t)}{S(t_0)}\right] = \alpha \cdot \log\left[\frac{R(t)}{R(t_0)}\right] \quad \Rightarrow \quad \log\left[\frac{1}{\kappa^{\text{unc.}}}\right]_i = \alpha \cdot \log\left[\langle LT\rangle\right]_i \tag{3.7}$$

A stable and precise measurement of α at the single-crystal level has not been achieved with the available statistics. Thus we limit the discussion to sanity checks of the calibration and correction procedure, through the verification of the average value of α over many channels, grouped by crystal producer and region. As different crystals of different radiation hardness populate the curve described by Equation 3.7 at different values of $\langle LT\rangle$, the value of α extracted by this method is not independent of the intercalibration and regional scale precision. In particular, if intercalibrations are derived from data collected over the same time range (or transparency loss range) by using transparency corrected data with a wrong value of α, results tend to be biased towards that value. Thus, the value of α here extracted has to be interpreted as an effective value, and relative comparisons are more robust than absolute results.

ECAL barrel

Summary results for the ECAL barrel are shown in Figure 3.13. Four different η regions, corresponding to the four ECAL modules, have been analyzed separately, and they are displayed in the same plot with different colour shades.



(a) September 2011 data

(b) Whole 2011 dataset

Figure 3.13: Uncorrected (empty red dots) and corrected (solid green dots) $E_{\text{sc}}/p_{\text{tk}}$ relative scale versus the transparency loss in the four modules of the ECAL barrel (darker colour shades indicate higher η regions.

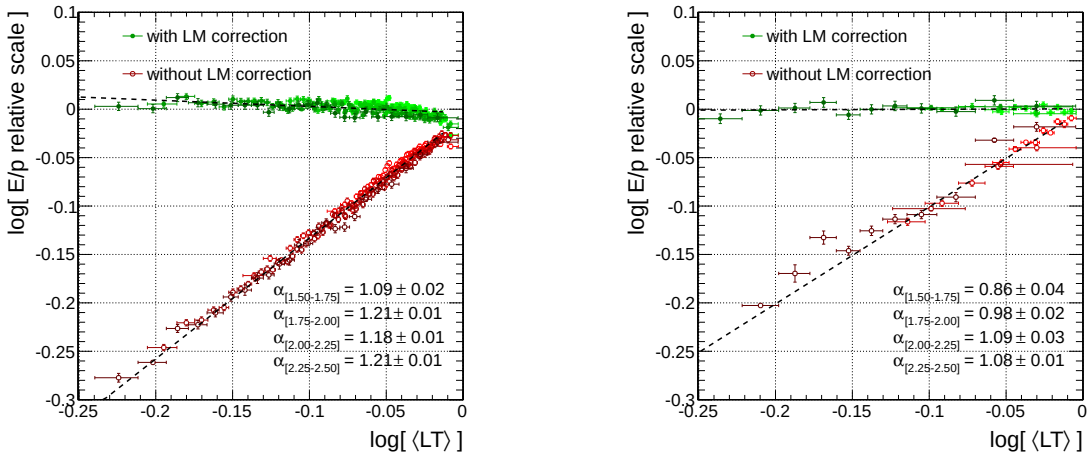
As the number of SIC crystals in the ECAL barrel is small, the analysis has been limited to BTCP crystals for which a mean value of $\alpha = 1.47 \pm 0.01(\text{stat.})$ is found. Only data collected in September 2011 have been used to derive this value, since in this the same period intercalibration constants from π^0 and $\eta \to \gamma\gamma$ samples were consistently derived. This value varies within its error if the analysis is expanded to the whole 2011 dataset. While α is systematically lower in the fourth module, consistent vales of α are measured elsewhere in the other three modules:

$$|\eta| \in [0.00, 0.44]: \quad \alpha_{\mathrm{BTCP}} = 1.49 \pm 0.01(\mathrm{stat.})$$
$$|\eta| \in [0.44, 0.79]: \quad \alpha_{\mathrm{BTCP}} = 1.47 \pm 0.02(\mathrm{stat.})$$
$$|\eta| \in [0.79, 1.14]: \quad \alpha_{\mathrm{BTCP}} = 1.47 \pm 0.03(\mathrm{stat.})$$
$$|\eta| \in [1.14, 1.45]: \quad \alpha_{\mathrm{BTCP}} = 1.28 \pm 0.04(\mathrm{stat.})$$

Though the exact interpretation of the results is outstanding, we cannot exclude a bias related to the calibration precision in the fourth ECAL module. The observed value of α is lower than the default used in the event reconstruction, thus resulting in a small overcorrection.

ECAL endcaps

The analysis in the ECAL endcap has been similarly performed in four different pseudorapidity regions of width $\Delta\eta = 0.25$ from $|\eta| = 1.5$ to $|\eta| = 2.5$. Moreover, BTCP and SIC crystals have been analyzed independently. Results span the whole 2011 dataset and are shown in Figure 3.14.



(a) BTCP crystals (b) SIC crystals

Figure 3.14: Uncorrected (empty red dots) and corrected (solid green dots) $E_{\mathrm{sc}}/p_{\mathrm{tk}}$ relative scale versus the transparency loss in four pseudorapidity regions of ECAL endcaps (darker colour shades indicate higher η regions.

While for BTCP crystals the value of the effective α is significantly lower than the pre-LHC measured value of 1.52, this behaviour is less evident in case of SIC crystals, which are compatible with an average value of $\alpha = 1.00 \pm 0.01(\mathrm{stat.})$.

$$|\eta| \in [1.50, 1.75]: \quad \alpha_{\mathrm{BTCP}} = 1.09 \pm 0.02(\mathrm{stat.}) \quad \alpha_{\mathrm{SIC}} = 0.86 \pm 0.04(\mathrm{stat.})$$
$$|\eta| \in [1.75, 2.00]: \quad \alpha_{\mathrm{BTCP}} = 1.21 \pm 0.01(\mathrm{stat.}) \quad \alpha_{\mathrm{SIC}} = 0.98 \pm 0.02(\mathrm{stat.})$$
$$|\eta| \in [2.00, 2.25]: \quad \alpha_{\mathrm{BTCP}} = 1.18 \pm 0.01(\mathrm{stat.}) \quad \alpha_{\mathrm{SIC}} = 1.09 \pm 0.03(\mathrm{stat.})$$
$$|\eta| \in [2.25, 2.50]: \quad \alpha_{\mathrm{BTCP}} = 1.21 \pm 0.01(\mathrm{stat.}) \quad \alpha_{\mathrm{SIC}} = 1.08 \pm 0.01(\mathrm{stat.})$$

A possible interpretation would require assuming that only part of the loss measured by the monitoring system is due to crystal transparency variations, and the remaining fraction of it is related to the ageing of the readout photodetectors (VPTs) [55]. Photon detector ageing would

impact linearly on the signal loss, or in other words with $\alpha = 1$ in Equation 3.6. Thus, for SIC crystals both the transparency and the photon-detector loss would affect the signal loss linearly with the same power law, yielding an effective $\alpha = 1$. At variance with this, the effective α for BTCP crystals would be lower than the expected value for transparency losses, and increasingly lower at larger η assuming that the VPT ageing is dose related. A combination of the two effects may also be present.

The correction model adopted for the endcaps in the last set of reconstruction conditions of 2011 includes a non-linear term in the form of $\alpha = \alpha[R(t)/R(t_0)]$, which is not inconsistent with the current analysis. The usage of an effective α value of roughly 1.17, suggested by the E_{sc}/p_{tk} analysis by simply taking the average of the fitted α values in the four endcap sub-regions, results in an improved stability of the EE relative response in 2011: the linearly increasing trend observed with the default reconstruction is practically removed, with a consequent improvement of the stability RMS down to 0.35% (see Figure 3.15).
It is difficult, however, to draw definitive conclusions from the E_{sc}/p_{tk} analysis alone, and further information must be gathered elsewhere. Viable options are the usage of $Z \to e^+e^-$ events, which allow to extract the best-α value (possibly in η rings) from a fit to the expected Z line shape, or the study of single crystal energy responses with energy flow measurements from minimum-bias events, allowing to retain a large statistics for each single ECAL channel.

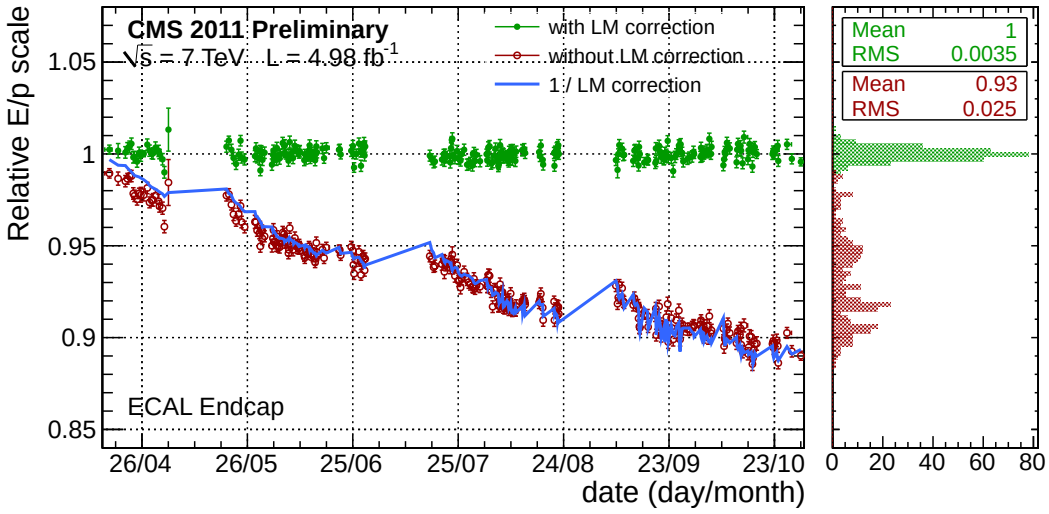


Figure 3.15: ECAL relative energy scale evolution obtained from electron E_{sc}/p_{tk} fits. The plot shows the results for ECAL endcaps during the whole 2011 LHC run. Uncorrected data (red empty dots), data corrected for the observed transparency loss (green solid dots) and the inverse of the correction derived from monitoring data (blue line) are displayed. Energy corrections for crystal transparency losses are obtained using an effective value of the crystal spectral index $\alpha = 1.18$, instead of the radiation-dependent value implemented in the default reconstruction shown in Figure 3.12.

3.5 ECAL response uniformity

The energy reconstructed in a supercluster is degraded by the presence of inter-modules or sub-modules cracks and gaps between crystals. E_{sc}/p_{tk} data are sensitive to these effects and can be used to derive correction functions or to validate existing sets of corrections entering the

$F_{e,\gamma}$ term of the ECAL energy reconstruction scheme (see Equation 3.2).

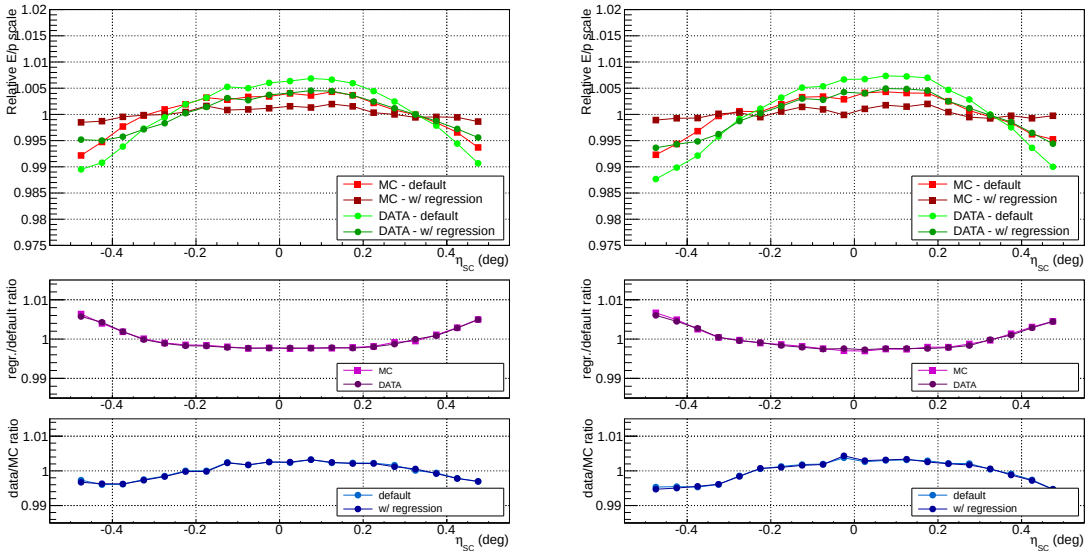
3.5.1 Local corrections

The local shower containment varies depending on the position of the impact point of the particle on the crystal. The energy scale dependence on local supercluster coordinates (i.e. position of the supercluster with respect to the center of the crystal) is studied considering E_{sc}/p_{tk} distributions for different impact positions. A different template is built for each one of the four ECAL modules separately in order to account for E_{sc}/p_{tk} shape variation along η. The focus of this section is the dependence of the reconstructed energy on the local coordinates, thus the procedure (including the building of the reference distributions) is performed independently on data and Monte Carlo. This implies that the results are by construction not sensitive to an overall scale factor between data and Monte Carlo, which is instead addressed in the following section. Only the crystals far from the ECAL supermodule and module boundaries are considered, and crystals belonging to the same ECAL module are folded in ϕ and η.

Figure 3.16 shows the relative E_{sc}/p_{tk} scale as a function of the local coordinate along η. Variations in the relative scale are at the level of 1-3% at the crystal edges with respect to the response at the center of the crystal, with a larger dependence on the local coordinate in the fourth ECAL module. This effect is more pronounced in data than in Monte Carlo, thus resulting in an additional spread of data with respect to Monte Carlo of a few per-mill (from 0.24% in the first module to 0.48% in the fourth module).

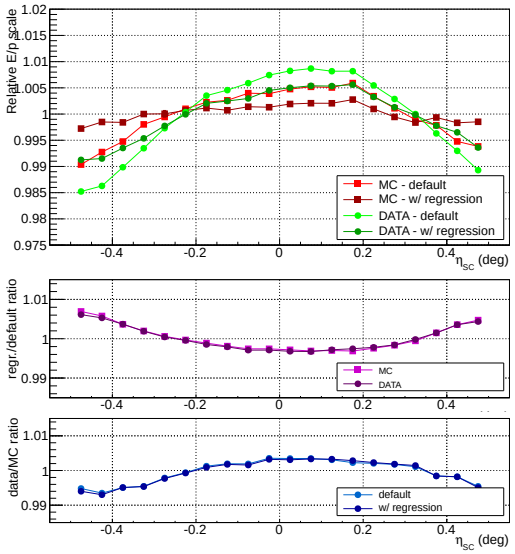
Similarly, the relative E_{sc}/p_{tk} scale as a function of the local coordinate along ϕ is shown in Figure 3.17. In this case, variations are below 1% with a very similar behaviour in data and Monte Carlo and no strong dependence on the η region of the calorimeter.

The local containment has been studied also for superclusters made of a different number of basic clusters n_{BC}. The effect along the local η coordinate is larger for superclusters with $n_{BC} > 1$ than for $n_{BC} = 1$ when default cluster corrections are applied (Figure 3.18). Along the ϕ direction, no particular trend is observable as a function of the basic clusters number.
Ideally each basic cluster can be considered as an independent shower and it is reasonable to apply local containment corrections to each basic cluster. We parametrized the local containment correction along η derived using the E_{sc}/p_{tk} relative scale of electrons with supercluster made by a single basic cluster and we verified, on an independent sample, that the correction is effective both in the $n_{BC} = 1$ and $n_{BC} > 1$ case, as visible in Figure 3.19). The effect of this basic cluster local containment corrections on resolution will be discussed in Section 3.6.

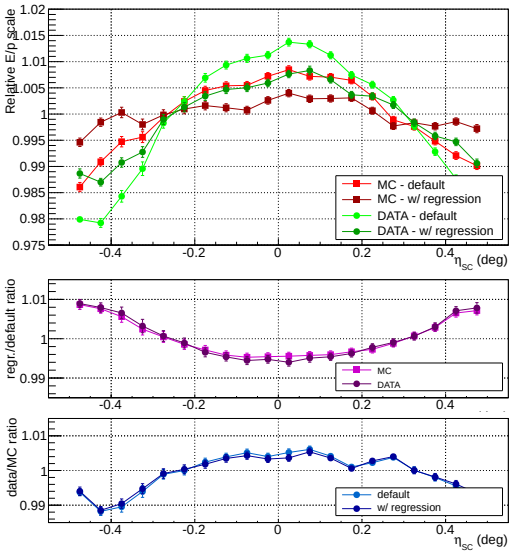


(a) Module 1

(b) Module 2

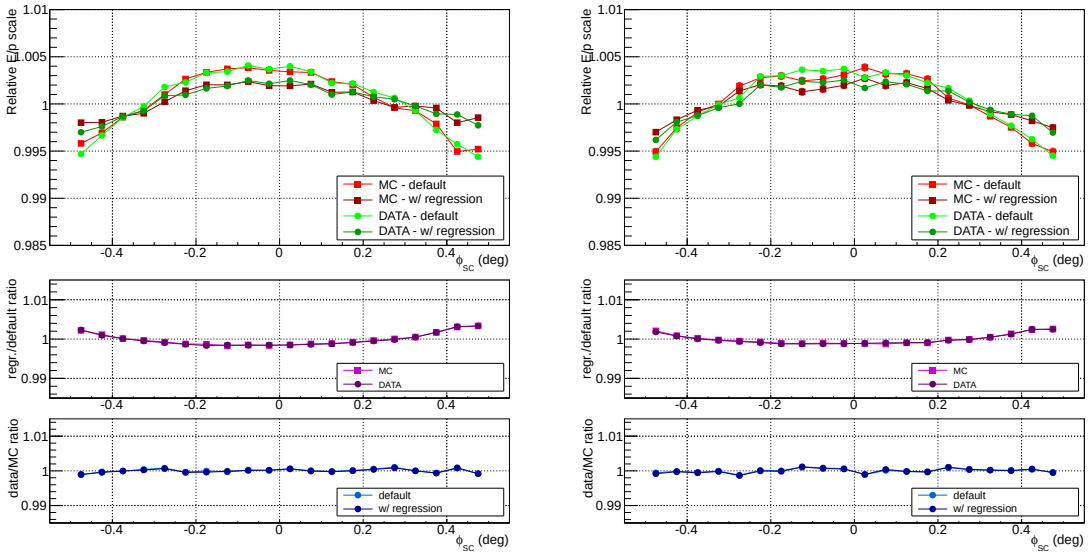


(c) Module 3



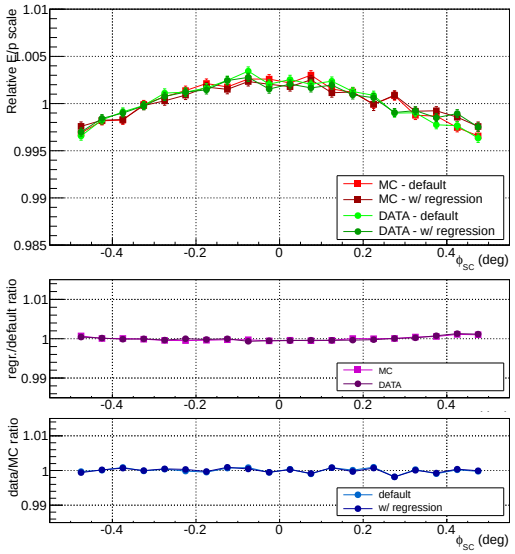
(d) Module 4

Figure 3.16: E_{sc}/p_{tk} relative scale as a function of the local η coordinate in the four ECAL modules for data (green dots) and Monte Carlo (red dots).

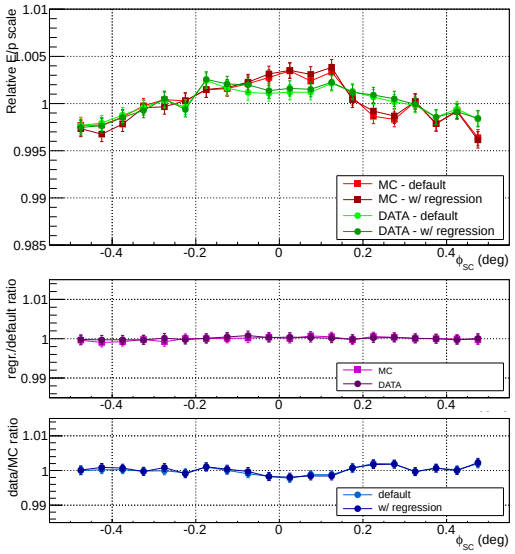


(a) Module 1

(b) Module 2

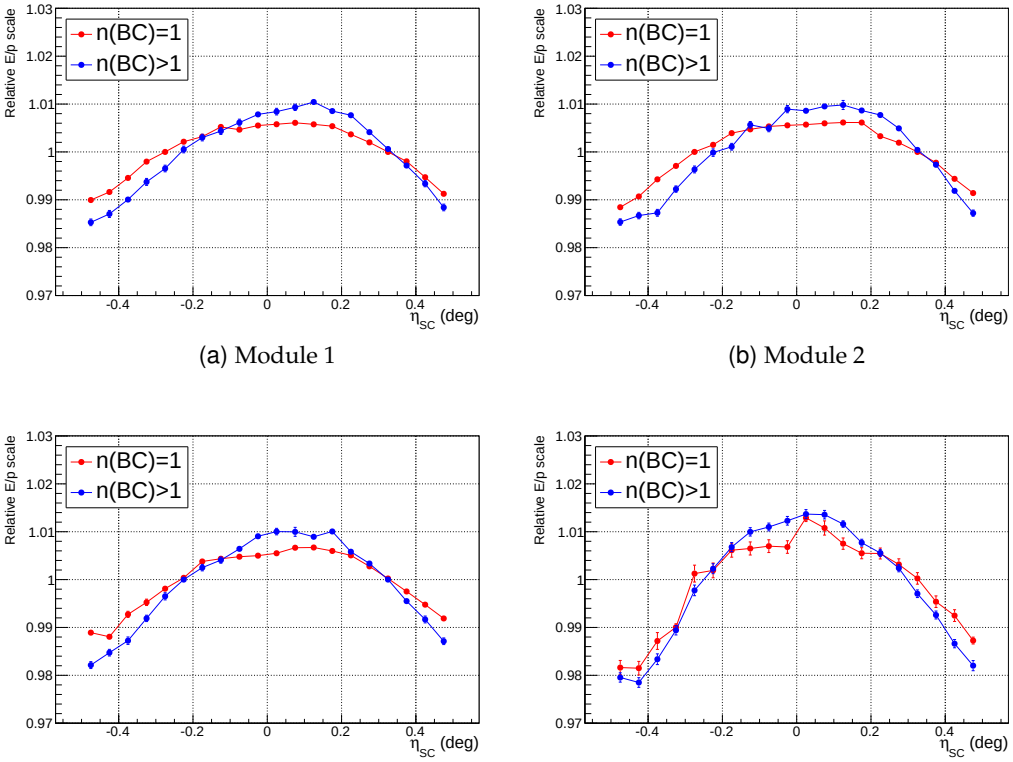


(c) Module 3



(d) Module 4

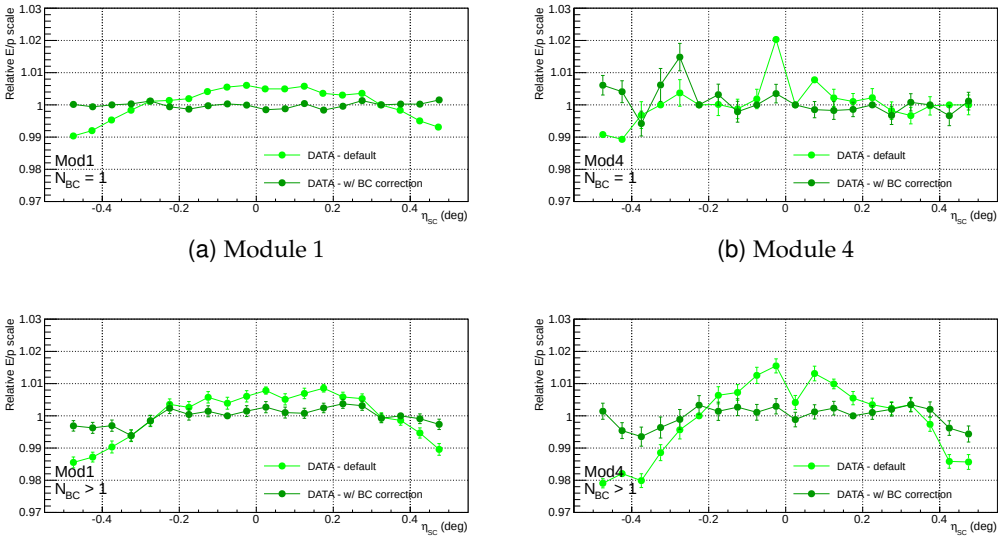
Figure 3.17: E_{sc}/p_{tk} relative scale as a function of the local ϕ coordinate in the four ECAL modules for data (green dots) and Monte Carlo (red dots).



(a) Module 1 (b) Module 2

(c) Module 3 (d) Module 4

Figure 3.18: E_{sc}/p_{tk} relative scale in data as a function of the local η coordinate for superclusters made of one basic cluster (red) or more than one basic cluster (blue).



(a) Module 1 (b) Module 4

(c) Module 1 (d) Module 4

Figure 3.19: E_{sc}/p_{tk} relative scale in $Z \to e^+e^-$ events as a function of the local η coordinate for default cluster corrections and after applying the local containment correction derived from electrons from $W \to e\nu_e$ events with supercluster made by a single basic cluster.

In Figure 3.20 the E_{sc}/p_{tk} scale is shown for events folded in 4 crystals as a function of the local ϕ. The modulation every second crystal in Monte Carlo is due to the ECAL internal module structure. In data this effect is not visible as it is compensated by intercalibration constants. This implies that MC-tuned corrections are over-correcting for this effect when applied to data. The corresponding additional spread in data with respect to Monte Carlo due to this effect is about 0.16%.

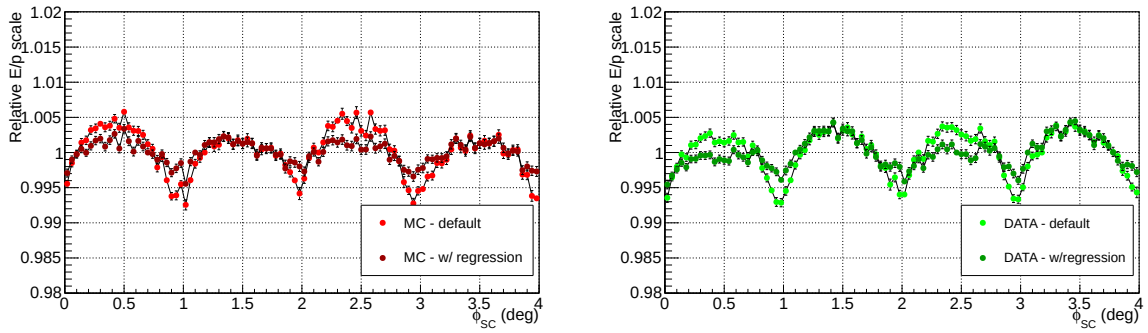


Figure 3.20: E_{sc}/p_{tk} relative scale as a function of the local ϕ coordinate in data (green) and Monte Carlo (red) for electrons and positrons with $|\eta| < 0.8$ and $\eta \times \text{charge} > 0$. All events are folded in 4 crystals.

3.5.2 Crack corrections along ϕ

In Figure 3.21, the effect of inter-module ϕ cracks is shown. Electrons and positrons with $|\eta| < 0.8$ and with $\eta \times \text{charge} > 0$ are used. Crack corrections, which are derived from Monte Carlo, are effective also in correcting data, with a small residual effect at the per mill level.

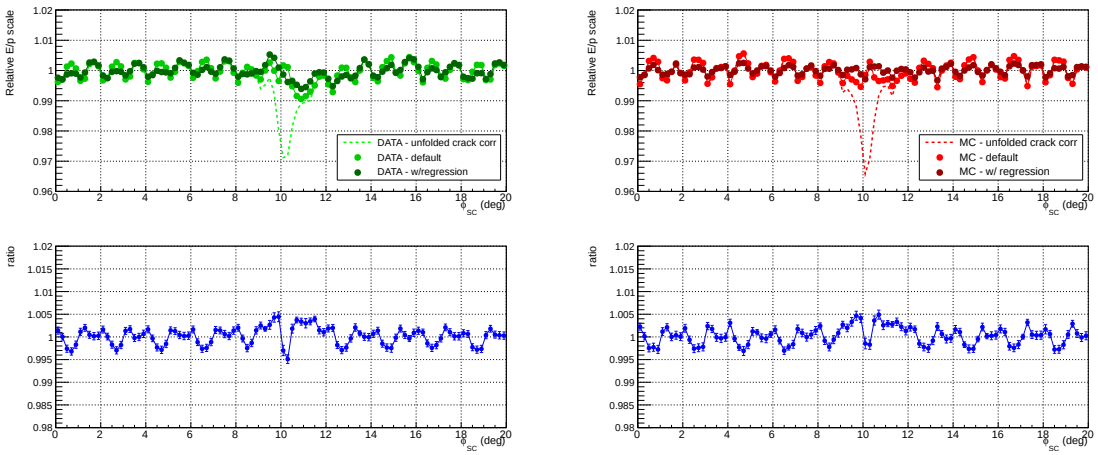


Figure 3.21: E_{sc}/p_{tk} relative scale as a function of the ϕ coordinate in data and Monte Carlo. All the events are folded in one supermodule to show the relative scale behaviour around the cracks between two adjacent supermodules.

3.5.3 Regional scale

The relative E_{sc}/p_{tk} scale can be studied as a function of the supercluster position along η, though some caveats must be taken into account.
Firstly, the shape of the E_{sc}/p_{tk} distribution is strongly dependent on η (e.g. see Figure 3.8)

and variations in the shape can induce biases in the fitted scale. To limit this effect, we adopted different templates in different η regions: four templates in the barrel, each corresponding to one ECAL module, and three templates in the endcaps, as defined in Section 3.3.3.
Secondly, the E_{sc}/p_{tk} scale is sensitive to the momentum scale. As we discussed in the same Section, a calibration of the momentum scale along η has been derived from data and is applied in the following analysis. Figure 3.22 shows the relative scale as a function of η in data and Monte Carlo and their ratio.

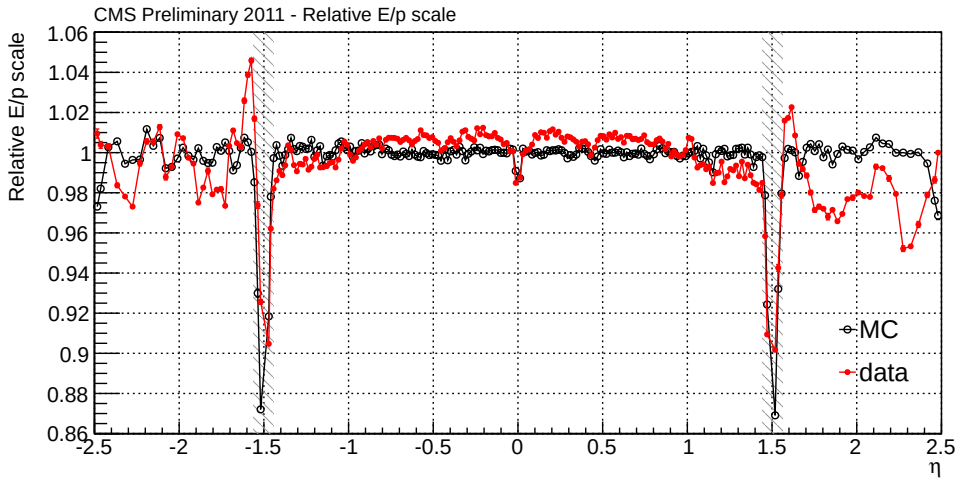


Figure 3.22: E_{sc}/p_{tk} relative scale as a function of the supercluster position along η in data and Monte Carlo and their ratio.

A noticeable feature, which is present in the data and not in the Monte Carlo, is the dependence of the relative energy scale on η in the first module around $\eta \sim 0$. It is observed both on low R_9 and high R_9 electrons, suggesting that it is a genuine scale calibration effect rather than a material effect, and deserves further investigation.

3.6 ECAL resolution studies

3.6.1 Resolution on Z events after regional local containment corrections

We applied the local containment corrections on single basic clusters derived from E_{sc}/p_{tk} data to all basic cluster of electrons in $Z \to e^+e^-$ events. An improvement in mass resolution of 0.6% is observed, as can be observed in Figure 3.23.

3.6.2 Resolution on Z events after regional scale calibration

We observe an improvement in the resolution on the $Z \to e^+e^-$ invariant mass where both electrons are in the endcap, after applying regional scale corrections based on the E_{sc}/p_{tk} method (see Section 3.5). The effect is visible in Figure 3.24, and corresponds to an improvement in the mass resolution of about 1.2% in quadrature, as estimated by the σ parameter of a Crystal-Ball resolution function convoluted with the Z boson Breit-Wigner distribution fitted to the data.

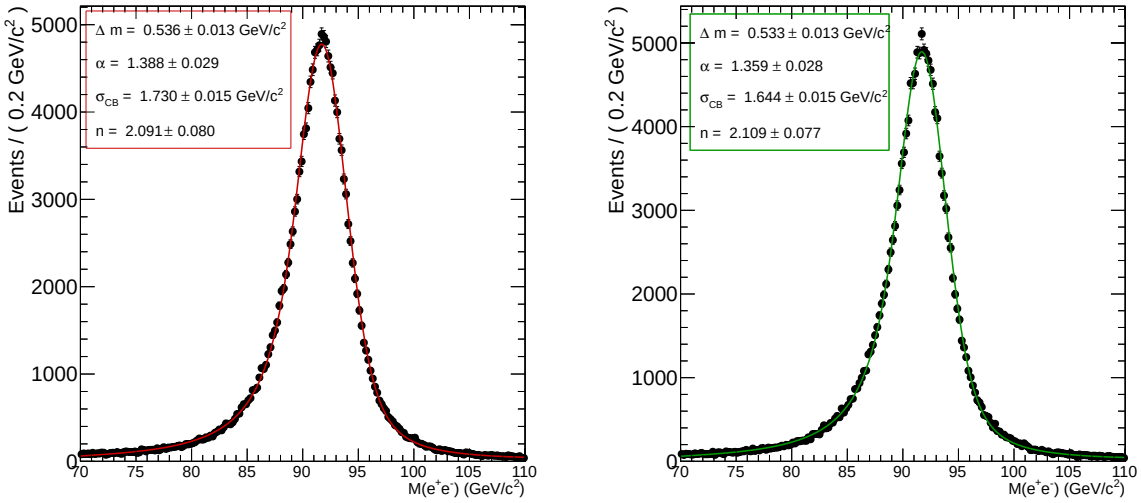


Figure 3.23: Invariant mass of the electron pairs in $Z \to e^+e^-$ decays, with both electrons in the barrel: without local containment corrections along η (left) and after applying local containment corrections along η to basic clusters (right).

3.6.3 Regional relative resolution

Exploiting the resolution term in Equation 3.4, the relative resolution of different ECAL regions can be studied. In particular, the impact of cracks between supermodules every 20 degrees in ϕ has been studied as follows: i) an E_{sc}/p_{tk} reference distribution is formed using data selected with the request that the seed was at least 3° away from a ϕ crack (i.e. three crystals); ii) the entire data sample is divided in subsamples identified by the local ϕ of the seed superclusters (ranging from 0° to 20° inside a supermodule), with a bin size of $\Delta\phi = 0.2°$; iii) the reference distribution of ii) was then fitted to each of the latter distributions, after convolution with a Gauss resolution curve as described by Equation 3.4. Results, averaged over all the supermodules, are shown in Figure 3.25, and indicate that the resolution in ECAL is rather uniform inside the supermodule, but degrades considerably in the crack region, where an additional constant term of up to 2–3% is needed to describe data[10]. The impact of the crack, however, is limited to a narrow ($\sim 1°$) region around it.

The method has been applied to data with the default reconstruction, including or excluding the default crack corrections, and with the regression correction of cracks, local containment and tracker material effects. In all the cases, the reference distribution was built by consistently using the reconstruction method adopted for the subsets. Thus, by construction, no need of extra smearing in the centre of the supermodule is needed in any of the cases. Differences are however visible in the crack region, where the regression reconstruction shows a slightly improved resolution relative to the center of the supermodule.

[10]Outside the crack region the additional smearing is not exactly zero, as zero is not allowed by the fit for reasons of numerical stability.

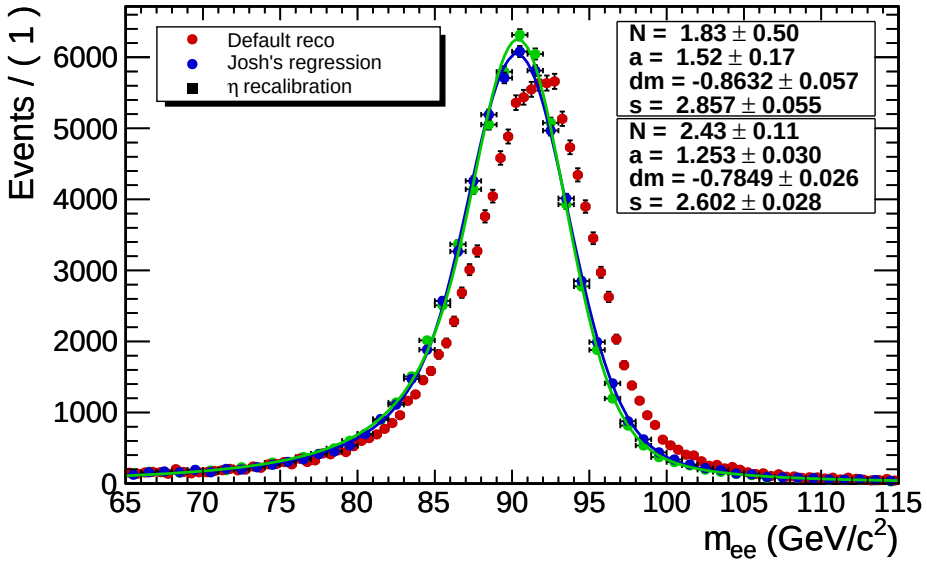


Figure 3.24: Invariant mass of the electron pairs in $Z \to e^+e^-$ decays, with both electrons in the endcap; the energy from the default reconstruction (red), with the regression correction (blue) and after recalibration of the regional energy scale (green) are compared.

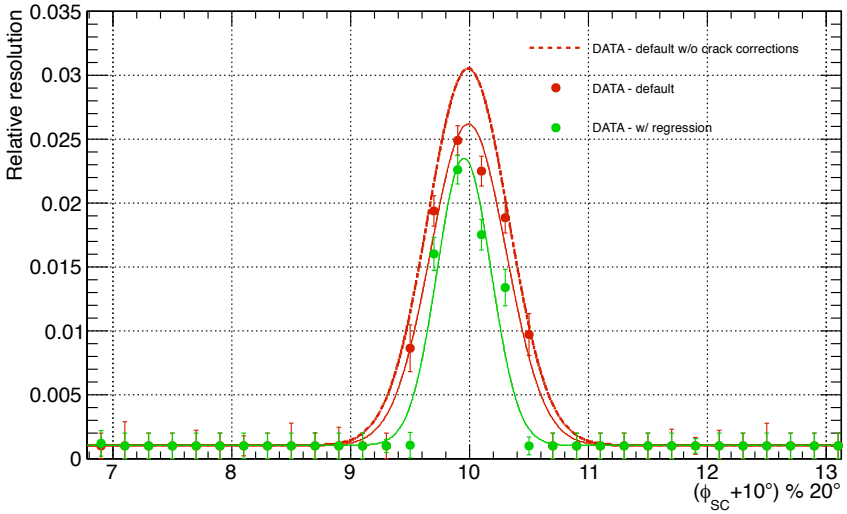


Figure 3.25: Relative electron resolution as a function of the local ϕ inside a supermodule (shifted by 10° to center the crack in the middle of the x-axis range): the resolution of the regression (green), the default (red points and continuous line) and of the default reconstruction without crack correction (red dashed line) are displayed.

3.7 ECAL single channel intercalibration

The CMS ECAL channels have been pre-calibrated prior to installation with laboratory measurements of crystal light yield and photo-detector gain during the construction phase (all EB and EE channels), with test-beam electrons (nine out of 36 EB supermodules and about 500 EE crystals) and with cosmic ray muons for all EB channels [56]. After installation in the LHC, the

beam dump events have been used to improve the pre-calibration precision in the EB and EE [57].
Inter-calibration constants from different sources have been cross-checked and then combined. The channel-by-channel calibration precision at the start-up at 7 TeV data taking was estimated to be:

- EB: about 0.5% for the 9 supermodules calibrated at test beams and 1.4% to 1.8%, depending on pseudorapidity, for the other 27 supermodules;
- EE: below 1% in the ~ 500 crystals calibrated with beam and about 5% elsewhere;
- ES: about 2.5% in all silicon modules from calibration with cosmic rays prior to installation.

The strategy to inter-calibrate with collision data involves the use of different methods as described in [58]:

- the ϕ-symmetry inter-calibration is intended to provide a fast inter-calibration exploiting the azimuthal invariance around the beam axis of the energy flow in minimum bias events;
- the π^0 and η calibration uses the invariant mass of photon pairs from low mass resonances to calibrate the crystal response;
- isolated electrons from $W \to e\nu_e$ and $Z \to e^+e^-$ decays can be used to compare the energy measured in ECAL to the track momentum measured in the silicon tracker.

The latter method profits from the analysis techniques described in this Chapter. Following the criteria of Section 3.3, the sample of isolated electrons used to derive a set of intercalibration constants for the ECAL barrel and endcaps has been defined. In the endcaps, a further requirement of $R_9 > 0.9$[11] is applied in order to restrict the data sample to low-bremsstrahlung electrons and improve the calibration precision. With these criteria, each crystal in ECAL barrel has on average 20–25 good electron seeds for each fb^{-1} of collected luminosity. This number reduces sizably in the barrel–endcap transition region due to the large amount of material upstream the calorimeter and the consequently large probability of electrons to emit bremsstrahlung photons (see Figure 3.26).

Once electron candidates are selected, the L3 recursive algorithm is used to calculate the intercalibration coefficients (the name of this technique comes from the L3 LEP experiment, where it was used for the first time to intercalibrate the BGO calorimeter [59]). For each l^{th} crystal being the seed crystal of N good electrons, the intercalibration coefficient at the n^{th} iteration of the procedure is linked to the coefficient at the $(n-1)^{\text{th}}$ step by the following formula:

$$c_l^n = c_l^{n-1} \times \frac{\sum_{m=1}^{N} w_{lm,n-1} \cdot \mathcal{T}_\eta \left[\left(E_{\text{sc}} / p_{\text{tk}} \right)_{m,n-1} \right] \cdot \left(p_{\text{tk}} / E_{\text{sc}} \right)_{m,n-1}}{\sum_{m=1}^{N} w_{lm,n-1} \cdot \mathcal{T}_\eta \left[\left(E_{\text{sc}} / p_{\text{tk}} \right)_{m,n-1} \right]} \quad \text{with} \quad w_{lm,n-1} = \frac{E_{lm,n-1}}{E_{\text{sc}}^{m,n-1}} \quad . \tag{3.8}$$

[11] The R_9 variable is defined as the ratio between the energy in a 3×3 matrix of crystals around the most energetic one and the supercluster energy.

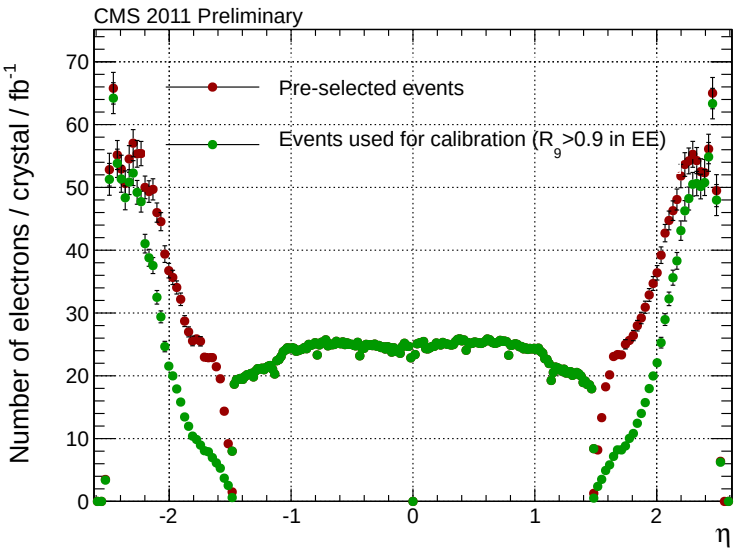


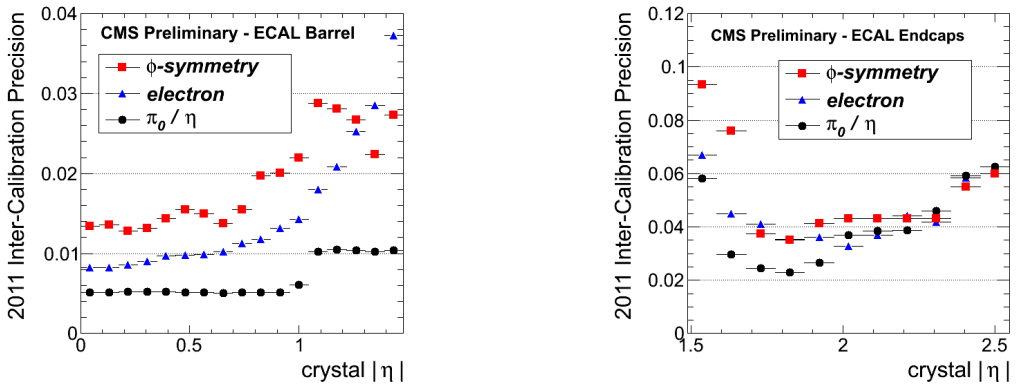
Figure 3.26: Distribution of the average number of electron seeds available for each ECAL crystal in 1 fb^{-1} of collected data as a function of η. While in the barrel no further selections are applied on top of those described in Section 3.3, a cut on R_9 is applied in the endcaps.

Each single crystal is assigned a weight w corresponding to the fraction of the total supercluster energy deposited in that particular channel. At each iteration, and for each electron, the supercluster energy is re-computed following Equation 3.2 and using the calibration coefficients of the previous iteration step for all the crystals concerned. The algorithm allows to recursively find the best values of c_i that yield a narrow E_{sc}/p_{tk} distribution, according to $\mathcal{T}_\eta$. These PDFs are different for groups of η rings, given the strong dependence of E_{sc}/p_{tk} on the η coordinate. The intercalibration coefficients are evaluated for each single crystal and normalized to the η-ring average intercalibration. The η ring scale is then fixed through E_{sc}/p_{tk} studies along the η coordinate with $Z \to e^+e^-$ events, as explained in Section 3.3.3.
A similar approach is followed to derive a ϕ-dependent correction from E_{sc}/p_{tk} studies to account for modulations in ϕ of the tracker-reconstructed electron momentum.

Figure 3.27 shows the statistical precision of the intercalibration for the ECAL barrel and endcaps, estimated dividing the data sample in two independent sub-samples and comparing the obtained intercalibration coefficients. With the 2011 statistics, a precision of about 1% in the central barrel and 2–4% at large η is achieved in the barrel, while endcap crystals can be intercalibrated with this method only the level of 4–5%.
This technique provides an important cross-check of the more precise π^0 calibration technique in the barrel. Moreover it allows to improve significantly the calibration precision of the ECAL endcap crystals when combined with the other techniques available.

3.8 Higgs searches in the $H \to \gamma\gamma$ channel at CMS

The search for a light Higgs boson decaying into two photons is not specifically studied in this work. The analysis of this decay channel is only shortly summarized here as the Physics motivation for the ECAL stability and resolution studies described in the previous Sections. A



(a) EB (b) EE

Figure 3.27: Precision of the various calibration sets used in 2011: ϕ symmetry, electron E_{sc}/p_{tk} and π^0/η for (a) the ECAL barrel and (b) the endcaps.

complete description of the analysis strategy can be found here [60, 61].

The analysis is based on the search for two well reconstructed and isolated photons within the ECAL fiducial region and makes wide usage of multi-variate techniques (boosted decision trees, or BDT) to enhance the small signal expected over a large background, mainly constituted by irreducible prompt di-photon production and QCD production of γ+jet or even di-jet events, where one or more object is reconstructed as a fake photon.

The mass resolution of a narrow resonance decaying into two photons is driven by two factors: the photon energy resolution and the resolution in measuring the opening angle between the two photons. Since photons, due to the absence of a reconstructed track, cannot be readily assigned to a given interaction vertex, the opening angle resolution strongly depends on the determination of the interaction vertex where the di-photon pair was produced. This becomes critical considering that, during the 2011 data taking, the number of interactions per bunch crossing (pile-up) constantly increased, yielding an year-average of about ten simultaneous pp interactions. These are spread over a luminous region of about 6 cm long: only if the interaction point is known to better than about 10 mm, then the resolution on the opening angle between the photons makes a negligible contribution to the mass resolution, as compared to the ECAL energy resolution.

A BDT is used to assign the photon pair to a specific reconstructed vertex, making use of three discriminating variables, constructed from the measured scalar or vector transverse momenta of the tracks associated with each vertex and the transverse momentum of the di-photon system. The overall vertex-finding efficiency (within 10 mm) for a Higgs boson of mass 120 GeV/c^2, integrated over its p_T spectrum, is found to be $83.0 \pm 0.2(\text{stat.}) \pm 0.4(\text{syst.})\%$. This number is measured in the simulation and further checked on real data with $Z \to \mu^+\mu^-$ events, where the algorithm is run after artificially removing the muon tracks.

The second place where a multi-variate approach is used is the identification of real photons. Shower topology variables, as well as isolation and pseudorapidity information are combined to discriminate real prompt photons from QCD fake objects. The number of reconstructed vertices is used as well to adjust the pile-up dependence in the isolation variables. Also in

this case, the efficiency of this selection is studied on real data making use of $Z \to e^+e^-$ and $Z \to \mu^+\mu^-\gamma$ events.

Finally, the output of the photon identification BDT, together with the di-photon kinematic variables and expected mass resolution, are combined together in a third, event-level BDT discriminator.
For the final result in this channel, it proves convenient to use the output of this di-photon BDT to define four event categories, which are handled as independent channels in the final extraction of a limit on the SM Higgs boson cross-section. This approach allows to best profit from the different Higgs mass resolution and S/B significances expected across ECAL. This can be appreciated in the BDT discriminator output of Figure 3.28.

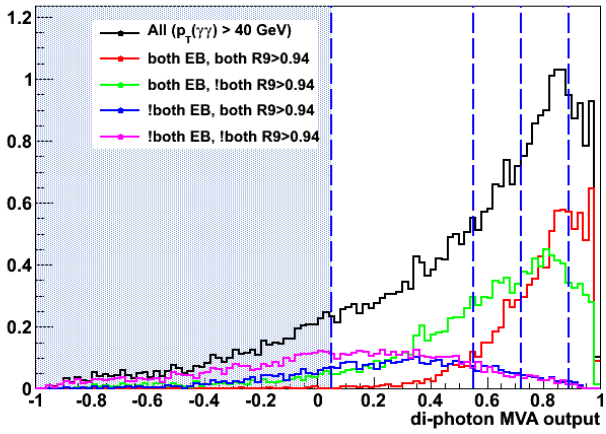
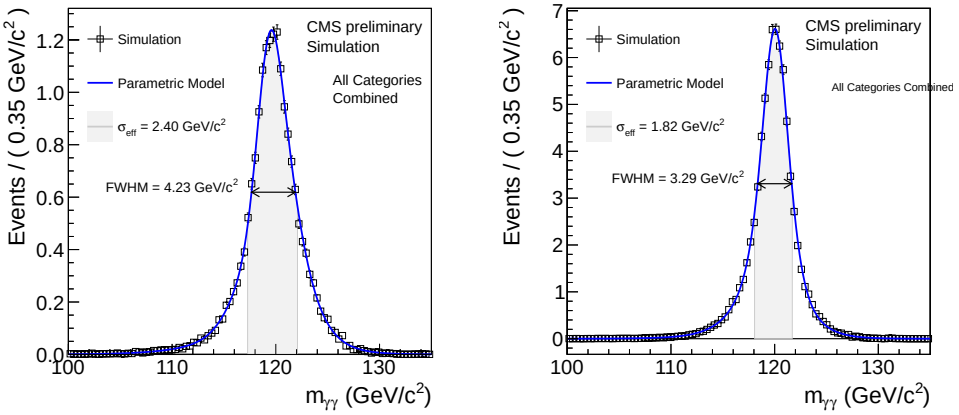


Figure 3.28: Di-photon BDT output distribution for events in a simulated Higgs boson signal (m_H = 120 GeV/c^2). The events are shown classified according to the four η/R_9 event classes. The event class boundaries based on the di-photon BDT output are shown as vertical blue lines, and events in the blue shaded region are not used in the statistical treatment of the analysis. It can be seen how events that events contributing the most, in the two highest BDT output categories, correspond almost entirely to EB photons with either high or low R_9 events.

A good understanding of the expected signal shape is a crucial input to the limit extraction procedure. The width of the reconstructed Higgs mass peak in the low mass range is driven by the detector resolution, whenever the correct vertex is taken. It is thus necessary to optimize the agreement between Monte Carlo and the observed data, so that the simulated events match the observed detector performance. As extensively discussed in the previous chapters, the main ingredients affecting the determination of the photon energy are the level of channel-to-channel intercalibration, the stability of the ECAL response thanks to the transparency loss corrections and the uniformity of the response provided by cluster energy corrections.
Even correcting the data for all the above effects to the best of our (present) knowledge, some differences with respect to simulation are still present. These discrepancies are dealt with by applying a run-dependent energy scale correction factor to simulated energies, obtained from the comparison of $Z \to e^+e^-$ events in data and Monte Carlo. An additional energy smearing factor is also derived so that the photon energy resolution in the simulation matches the one observed in data.
As an example to show the many progresses in the always-improving understanding of ECAL,

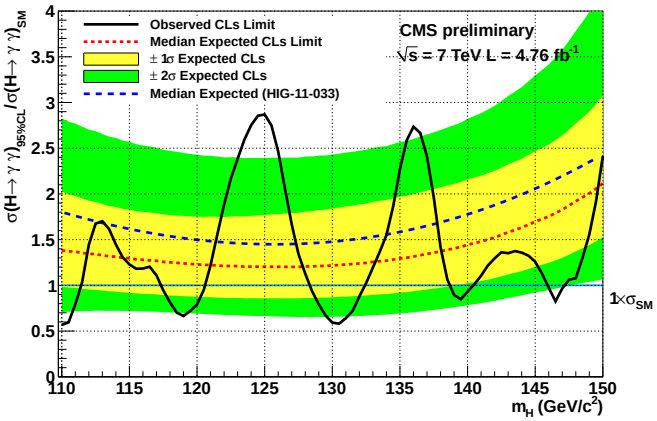
the Higgs boson mass as expected from the simulation, after all correction factors and extra-smearing terms are included, is shown in Figure 3.29 for two successive key moments in the Higgs boson search campaign: from July 2011 (EPS conference) to March 2012 (Moriond conference), the improvement is sizable. Improved single channel intercalibration and transparency corrections contribute to this result, together with the relative scale local corrections of Section 3.5 and a new, BDT-based approach to derive supercluster $F_{e,\gamma}(E_T, \eta)$ corrections (see Equation 3.2), not specifically addressed in this work.



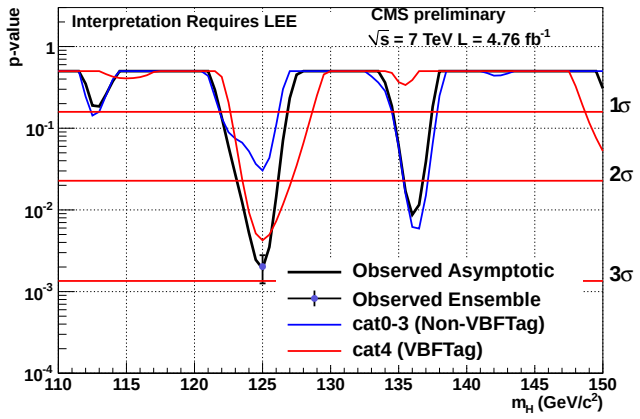
(a) July 2011 (b) March 2012

Figure 3.29: Expected Higgs boson mass from the Monte Carlo simulation, after scale correction and energy smearing factors, obtained from $Z \to e^+e^-$ events, are applied to match the observed performances in data, as shown at (a) the EPS conference (July 2011) and (b) the Moriond conference (March 2012). All event categories are included in these plots.

Although a complete discussion on the statistical methods to interpret the observed results is postponed to Section 4.10, the observed limit on the SM Higgs cross section in the $H \to \gamma\gamma$ channel is reported in Figure 3.30a. Here all four categories are combined, plus a fifth category where two forward jets are found in the event, meant to tag the VBF production channel. While the expected signal is modeled with the smeared Monte Carlo shapes, the background is estimated through a fit of the observed data. The $H \to \gamma\gamma$ analysis exclude, at 95% confidence level, the existence of a SM Higgs particle in a few narrow regions between 110 and 150 GeV/c^2, while an intriguing excess of events is observed around 125 GeV/c^2. At this mass, the local p-value, i.e. the probability of background to fluctuate upwards and originate a fluctuation as large as the one observed is computed $2 \cdot 10^{-3}$ (2.9σ), and most of it is due to the VBF category of events (see Figure 3.30b). Although this result encourages exciting hypotheses on the existence of a Higgs particle in that mass region, the observed excess is not statistically significant. After taking into consideration the so-called look-elsewhere effect, telling the probability that a similar or larger excess is found anywhere in the considered mass spectrum, the local p-value is reduced to 1.6σ. Only the analysis of 2012 data at $\sqrt{8}$ TeV will allow a correct interpretation of this result.



(a)



(b)

Figure 3.30: (a) Exclusion limit on the cross section of a SM Higgs boson decaying into two photons as a function of the boson mass relative to the SM cross-section, where the theoretical uncertainties on the cross section have been included in the limit setting. All four event classes and the VBF one are combined. (b) Observed local p-values, for the combined event classes, and also for the di-jet-tagged class and the combination of the other four classes.

Chapter 4

Higgs boson search in the $H \to WW \to l\nu_l q\bar{q}$ channel

The Standard Model (SM) of particle physics, briefly reviewed in Chapter 1, successfully describes the majority of high-energy experimental data [62]. One of the key remaining questions is the origin of the masses of fermions and weak vector bosons. In the simplest implementation of the Standard Model, these masses are attributed to the spontaneous breaking of electroweak symmetry caused by a new scalar field. The existence of the associated field quantum, the Higgs boson, has yet to be experimentally confirmed. Therefore, the search for the Higgs boson is arguably one of the most important studies being done at the LHC.

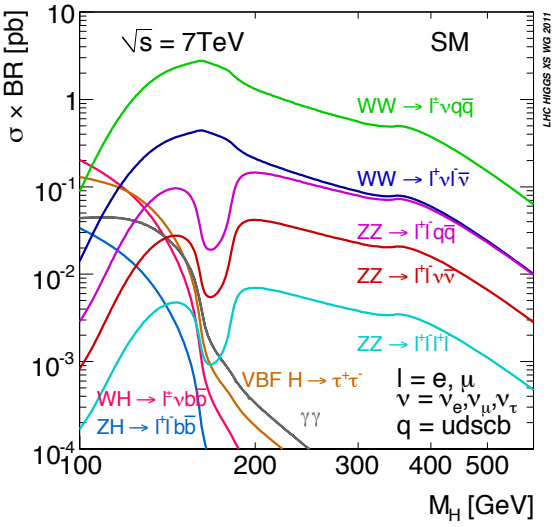


Figure 4.1: Accessible final states of Higgs boson decays and relative cross-section times branching ratio (with $l = e,\mu$). Picture from [14].

For Higgs masses that are sufficiently high to kinematically allow the Higgs decay into a pair of W or Z bosons, these decay modes are the dominant ones (see Figure 1.9a). The experimentally accessible final states are shown in Figure 4.1 and listed in Table 4.1, where the subsequent decays of the vector bosons into the different final state particles are taken into account. It is clear that the final state where the Higgs decays into a pair of W bosons, of which one decays

H decay mode	$\mathcal{B}$	$\mathcal{B}/\mathcal{B}(H \to l\nu_l q\bar{q})$
$WW \to l\nu_l q\bar{q}$	$1.7 \cdot 10^{-1}$	1
$WW \to l\nu_l l\nu_l$	$2.6 \cdot 10^{-2}$	1/6
$ZZ \to l^+l^- q\bar{q}$	$2.5 \cdot 10^{-2}$	1/7
$ZZ \to l^+l^- \nu\bar{\nu}$	$7.2 \cdot 10^{-3}$	1/24
$ZZ \to l^+l^- l^+l^-$	$1.2 \cdot 10^{-3}$	1/144

Table 4.1: Absolute and relative-to-$l\nu_l q\bar{q}$ branching ratio values for Higgs boson decays into a couple of vector bosons.

leptonically and the other decays hadronically, has the largest branching ratio $\mathcal{B}$ and exhibits a reconstructable Higgs mass peak [63].
In this work, the analysis that sets a limit on the Higgs boson cross-section based on this decay mode is described. The analysis selects events with one well identified and isolated lepton (electron or muon) and large missing transverse energy as the decay products of the leptonic W, and two high-p_T jets originating from the fragmentation of the $q\bar{q}$ pair from the hadronic W decay. To increase the signal efficiency, events with a third additional jet are retained too. The main experimental issue is to control the large W+jets background sufficiently well that the advantages of the large branching ratio of this final state are realized.

This Chapter is structured as follows. A discussion about the signal and background processes relevant for the analysis is presented in Section 4.1 while the datasets analyzed as well as the trigger strategy are described in Section 4.2. The physics object reconstruction and the lepton and jet pre-selections are described in detail in Section 4.3, the reconstruction of the Higgs mass is detailed in Section 4.4. The measurement of the reconstruction, identification and trigger efficiencies for the lepton leg are described in Section 4.5, while Section 4.6 describes the measurement of the trigger efficiency for the transverse mass cut of the electron trigger. With these ingredients, LHC data are compared to Monte Carlo expectations in Section 4.7 to assess the good understanding of the key variables for the analysis and validate the subsequent studies to extract the limit. The systematic uncertainties affecting the background and signal evaluation are described in Section 4.9. Finally, Section 4.10 describes the limits obtained on the Standard Model Higgs cross-section.

4.1 Signal and background expectations

The total pp cross-section at the LHC is about 100 mb, which implies an interaction rate of about 10^8 Hz when running at an instantaneous luminosity of $1 \cdot 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In most of cases, events are generated by long-distance collisions between the two incoming protons. The momentum transfer of the interaction is small and therefore particle scattering at large angle is suppressed. The particles produced in the final state of such interactions have large longitudinal momentum but small transverse momentum ($p_T \simeq 500\,\mathrm{MeV}/c$), and most of the collision energy escapes down the beam pipe. These events, called "minimum-bias" events, represent by far the majority of pp interactions, but are of no interest.
Occasionally, head-on collisions occur between two partons of the incoming protons. These are interactions at small distances, and therefore are characterized by large momentum transfers. In this case, particles in the final state can be produced at large angles with respect to the

beam line (high p_T) and massive particles can be created. These are the interesting, though rare physics events. A schematic view of such events is shown in Figure 4.2.

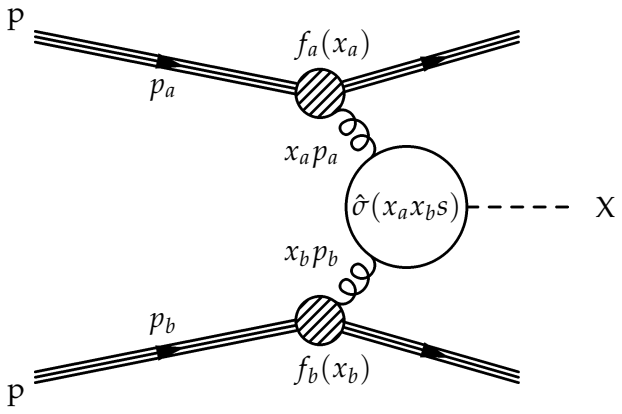


Figure 4.2: Diagrammatic structure of a generic hard scattering process.

Using the parton model formalism, the cross-section of a generic hard-scattering interaction at the LHC hadron collider can be written as [64]

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

In the above equation, the total QCD cross-section has been factorized into the convolution of two distinct terms:

- a hard, process-dependent partonic cross-section $\hat{\sigma}$ that depends on the effective center-of-mass energy $\hat{s} = x_a x_b s$. It can be calculated in the perturbative-QCD framework;
- two non-perturbative, process-independent parton distribution functions (PDFs) f_i ($i = a, b$) that represent the probability of finding a parton i inside the proton carrying a fraction x_i of the proton longitudinal momentum $\sqrt{s}/2$. The PDFs are extracted from data.

The factorization scale μ_F has been introduced in order to regularize the next-to-leading order correction terms from real and virtual gluon emission. Since collinear divergences for initial state partons are not cancelled by virtual diagrams, this cut-off scale is introduced and the divergent contribution is moved from the perturbative hard cross-section $\hat{\sigma}_1(\hat{s}, \mu_F^2)$ to the non-perturbative proton dynamics via the PDFs $f_i(x_i, \mu_F^2)$, which acquire a μ_F dependence. The fact that the PDF redefinition absorbing divergences is universal and not process dependent is a major feature of QCD.
Figure 4.3 shows the predictions for some important Standard Model cross sections at pp and $\mathrm{p\overline{p}}$ colliders, calculated using the above formalism (at next-to-leading order in perturbation

theory, i.e. including also the σ_1 term in Equation 4.1.)

Since in the absence of a complete set of higher order corrections, the cross-section dependence on μ_R and μ_F is explicit, it is necessary to make a specific choice for the two scales in order to make cross-section and uncertainties calculations. 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.

The second important source of uncertainty in the cross-section calculation is related 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 [65]. For uncertainties of next-to-leading order (NLO) cross-sections, a 68% C.L. envelope provided by the central values and PDF+α_S errors from the MSTW08, CTEQ6.6 and NNPDF2.0 sets is built. As a central value, the midpoint of this envelope is used. At next-to-next-to leading order (NNLO), 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.

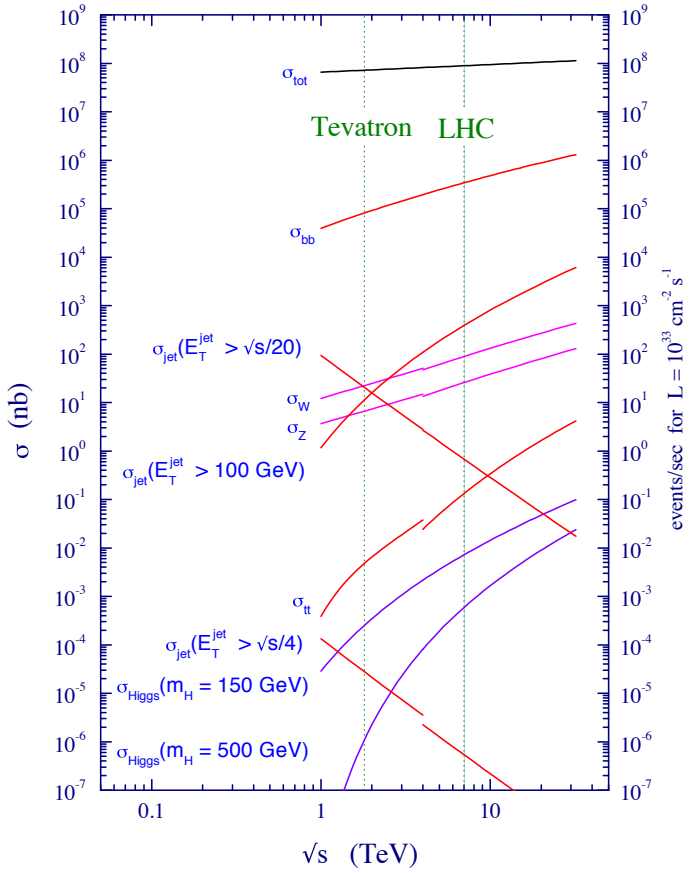


Figure 4.3: Standard Model cross sections at the Tevatron and LHC colliders for some reference processes.

4.1.1 The background processes

All the processes with one lepton, two jets and missing transverse energy in the detector have to be considered as possible sources of background. The most relevant are listed here below:

- **W($\rightarrow l\nu$)+jets**: this is the production of single W vector boson decaying leptonically, in association with the radiation of quarks or gluons that can mimic the final state signature when reconstructed as the hadronic W decay products. Because of a huge production cross-section, this background is by far the most dangerous for the analysis. The fully differential cross-section of W production through $\mathcal{O}(\alpha_S^2)$ is known via the FEWZ program [66], with NNLO QCD corrections being at the level of a few percent.
- **Drell-Yan Z/γ^*($\rightarrow l^+l^-$)+jets**: this is the production of single Z/γ^* bosons in association with quarks or gluons, where one lepton is undetected because of acceptance or inefficiency effects, and the hadronic activity mimics the final state signature and the hadronic W decay products. Also Z boson production is calculated with FEWZ at NNLO accuracy in QCD [66].
- **Diboson**: the production of vector boson pairs constitutes an irreducible background for the analysis. Three different diboson production channels are considered:

  1. WW: the final state originating from the non-resonant production of W boson pairs is topologically indistinguishable from signal events. Only kinematic selections can help in its reduction. The inclusive production cross-section is known at NLO in QCD, both considering the $q\bar{q} \rightarrow W^+W^-$ and the $gg \rightarrow W^+W^-$ production mechanisms;
  2. WZ: in case the Z decays hadronically, or the W decays hadronically and one lepton from Z decay is not identified by the detector, this process also contributes to the backgrounds. The production of such events proceeds via $q\bar{q}' \rightarrow W^{\pm}Z$, and its cross-section is known at NLO in QCD;
  3. ZZ: the production of Z pairs can originate the same final state of the analysis in case one Z decays hadronically, and one lepton from the other Z boson is not identified by the detector. Both the $q\bar{q} \rightarrow ZZ$ and the $gg \rightarrow ZZ$ production mechanisms are included in the cross-section calculation at NLO accuracy. The contamination level from such events is nevertheless irrelevant due to its tiny cross-section, when compared to the other leading backgrounds.

  All diboson production cross-section calculations are obtained with MCFM, further details can be found at [67].
- **$t\bar{t}$+jets**: top quarks pairs are produced at LHC via the gluon fusion process $gg \rightarrow t\bar{t}$ or via quark annihilation $qq \rightarrow t\bar{t}$. The semi-leptonic final states can be reduced by requests on the number of jets and vetoing the presence of b-like jets, while the fully leptonic ones are reduced by requiring the presence of only one good lepton in the event. In any case, because of acceptance and inefficiencies, this background still contaminates the signal phase space, and it is the second leading one after W+jets, especially for final states with an extra jet. Calculations for this process are available

at approximate NNLO order (NLO + next-to-next-to-leading logarithm resummations of soft-gluon corrections, NNLL) [68].

- **Single top**: it proceeds through three separate channels:
  1. *t*-channel: top is produced after a quark-gluon interaction with the exchange of a virtual W [69];
  2. *t*W-channel: top is produced in association with a charged vector boson in a weak process, from a gluon-bottom pair in the initial state [70];
  3. *s*-channel: top is produced in association with an anti-bottom, after the annihilation of a pair of quarks in a weak vertex [71].

  All single-top cross-sections calculations are available at NNLL accuracy order.
- **QCD**: multi-jet events can generate a background because of the combination of a huge production cross-section at a hadron collider and the the non-zero probability of a jet to be mis-reconstructed as a lepton. Since the fraction of fake muons reconstructed from jets is much lower than the probability of fake electrons, QCD contamination is important only in the $e\nu_e q\bar{q}$ final state.

The cross-section for the backgrounds, multiplied by the branching ratio when meaningful, are reported in Table 4.2. Each different process includes all the decays taken into account in the effective cross-section calculation.

process	σ (pb)
W($\to l\nu$)+jets	31300 ± 1600
Z($\to l^+l^-$)+jets	3050 ± 130
WW	47.0 ± 1.5
WZ	18.2 ± 0.7
ZZ	7.10 ± 0.15
$t\bar{t}$+jets	163 ± 14
t (*t*-channel)	41.9 ± 1.8
$\bar{t}$ (*t*-channel)	22.6 ± 1.0
t (*t*W-channel)	7.9 ± 0.6
$\bar{t}$ (*t*W-channel)	7.9 ± 0.6
t (*s*-channel)	3.19 ± 0.14
$\bar{t}$ (*s*-channel)	1.44 ± 0.07
QCD (e-enriched)	~ 6740000
QCD (μ-enriched)	~ 84700

Table 4.2: Cross-section values for the backgrounds, multiplied by the branching ratio when meaningful (with $l = e,\mu,\tau$ in this table).

4.1.2 The SM Higgs signal

The expected SM Higgs production cross-section is a factor of $10^4 - 10^5$ smaller than the electroweak production of W bosons (see Figure 4.3). Multiplying the cross-section by the branching ratio of the Higgs boson decay into a pair of W bosons and subsequently into a lepton-neutrino and a di-jet pair, it translates into the effective cross-section for the signal processes relevant for this analysis, as listed in Table 4.3 as a function of the Higgs mass. Cross-section

and branching ratio values are provided by the "LHC Higgs Cross Section Working Group" [14]. Further details on the SM Higgs production cross-section have already been addressed in Section 1.3.

Of the four distinct production mechanisms foreseen at the LHC, the Higgs-strahlung and the $t\bar{t}$ associated production have not been considered in the analysis here described, being their contribution not significant when compared to the total cross section. With at most one extra jet in addition to the two hadronic W decay products, the analysis is specifically designed for Higgs boson searches in the gluon fusion production channel, since for Higgs particles produced via vector boson fusion, an additional couple of forward high-p_T jets is expected. Nevertheless, the VBF production channel contributes a non negligible factor at the end of the event selection ($\sim$ 10%, see Table 4.11) and is therefore considered together with the gluon fusion channel.

In order to have both W bosons from Higgs decay produced on shell, only Higgs boson masses well above the kinematic threshold of $2 \times m_W$ are considered in the following studies ($m_H \geq 250\,\text{GeV}/c^2$).

m_H (GeV/c^2)	$\sigma_{gg} \times \mathcal{B}(H \to l\nu_l q\bar{q})$ (pb)	$\sigma_{VBF} \times \mathcal{B}(H \to l\nu_l q\bar{q})$ (pb)
250	$0.67^{+0.11}_{-0.10}$	$0.087^{+0.003}_{-0.002}$
300	$0.48^{+0.08}_{-0.07}$	$0.060^{+0.003}_{-0.002}$
350	$0.45^{+0.09}_{-0.07}$	$0.0416^{+0.002}_{-0.0012}$
400	$0.34^{+0.05}_{-0.06}$	$0.0272^{+0.0016}_{-0.0008}$
450	$0.22^{+0.03}_{-0.04}$	$0.0197^{+0.0013}_{-0.0006}$
500	$0.13^{+0.02}_{-0.02}$	$0.0150^{+0.0011}_{-0.0005}$
550	$0.084^{+0.016}_{-0.015}$	$0.0117^{+0.0009}_{-0.0004}$
600	$0.053^{+0.010}_{-0.010}$	$0.0093^{+0.0008}_{-0.0004}$

Table 4.3: Cross-section values and their uncertainties for Higgs boson production via the gluon fusion (left) or vector boson fusion (right) production mechanism, multiplied by the branching ratio of $l\nu_l q\bar{q}$ final state (with $l = e,\mu$), as a function of the Higgs mass. Values taken from [14].

4.2 Datasets and triggers

4.2.1 Data samples

The data sample used in this analysis was recorded by the CMS experiment in 2011, and can be divided in two main running periods, called `Run2011A` and `Run2011B`. The former extends up to the September technical stop, and can be further divided into four sub-periods, while the latter covers the data taking operations from September to the end of the proton run.
The conditions of the LHC machine changed significantly throughout the data-taking period. In particular, the instantaneous luminosity was less than $5 \cdot 10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$ at the beginning of 2011 and gained a factor of about ten at the end of the year. Also the average number of pile-up (PU) interactions per bunch crossing $\langle N_{PU} \rangle$ was not constant, with a sizable increase in

Run2011B with respect to Run2011A.
Only certified runs and luminosity sections are considered in this work, which means that a good functioning of all CMS sub-detectors is required. The total statistics analyzed correspond to an integrated luminosity of $(5.0 \pm 0.1)\,\text{fb}^{-1}$.

Events relevant for this analysis were stored in two Primary Datasets, called SingleElectron and SingleMu, depending on the flavour of the final-state lepton and processed using the official CMS reconstruction software[1] (see the following paragraphs for more details on the trigger selections).

The 2011 evolution of the integrated luminosity delivered by LHC and usable for Physics analysis is represented in Figure 4.4, together with $\langle N_{\text{PU}}^{\text{true}} \rangle$ versus time. In Table 4.4a, the different running periods and their corresponding run ranges and integrated luminosities are summarized.

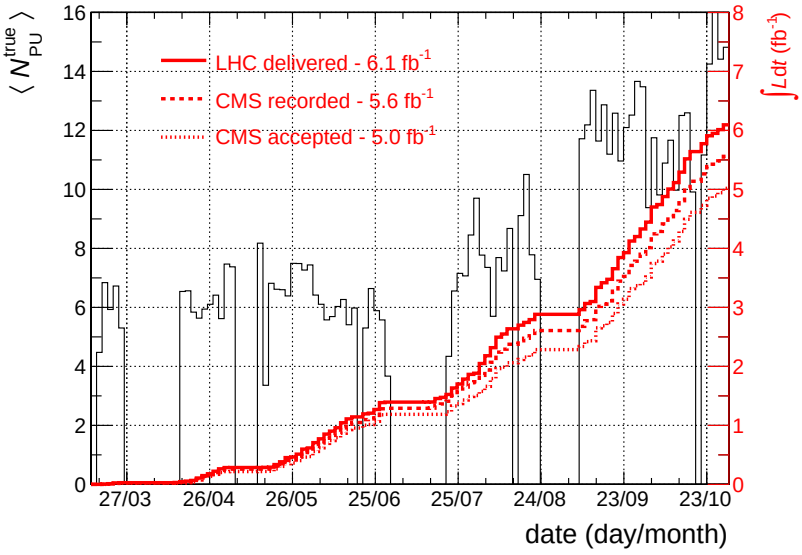


Figure 4.4: Integrated luminosity versus time: delivered to (solid red line), recorded by (dotted red line) and accepted by CMS (dashed red line) during stable beams for pp running at 7 TeV centre-of-mass energy. On the same plot, the expected number of PU interactions is shown versus time, averaging on a time lapse of two days (black line). $N_{\text{PU}}^{\text{true}}$ is calculated from the instantaneous luminosity of each luminosity section as in Equation 4.2.

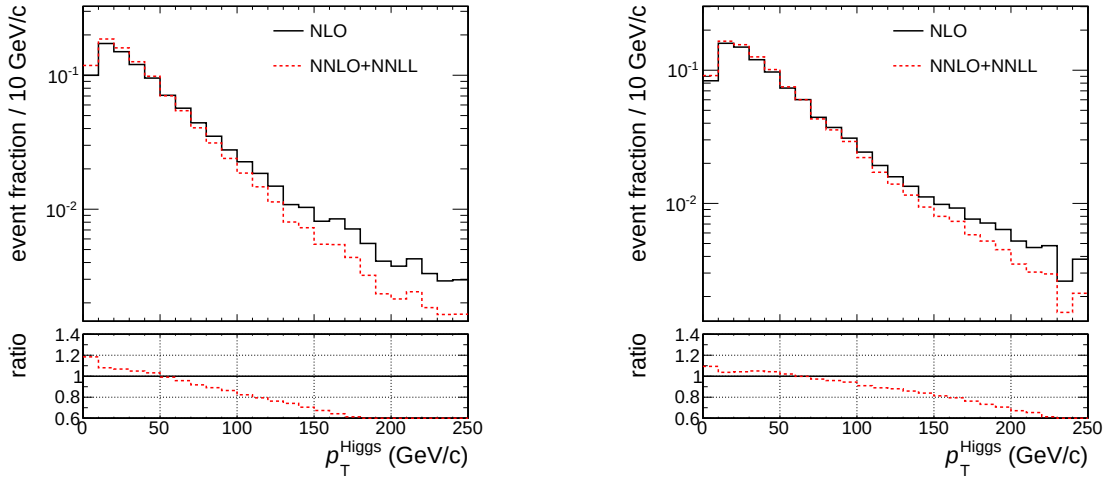
4.2.2 Monte Carlo samples

Standard Model Higgs boson samples, as well as samples for the electroweak and QCD-induced background sources of Table 4.2, have been generated and showered using different Monte Carlo generators[2]. The list of signal and background samples used in the analysis is in Table 4.4b, together with the equivalent luminosity available for the study.

The POWHEG-BOX generator [72, 73, 74] is used to produce signal Higgs events [75, 76] with NLO accuracy. For this analysis, samples with Higgs mass hypotheses ranging from 250 to 600 GeV/c^2 are considered, produced both via the gluon fusion and the vector boson fusion

[1] A CMSSW_4_2_X version of the CMS software was used.
[2] Monte Carlo samples from the so-called "Fall11" CMS production campaign have been actually used.

mechanism. Although the analysis requires the presence of an electron or a muon in the final state, the samples in which the leptonic W boson goes into a $\tau\nu_\tau$ pair are taken into account, too. These events can have the same final state searched for in this analysis when the τ decays leptonically. Anyway, given the small branching ratio of such decays, and the softer p_T spectrum of the resulting electron or muon, they contribute only a small fraction of the total signal yield at the end of the event selection (see 4.11 for a quantitative estimate).
Due to details in the implementation of the calculation, the resulting Higgs p_T spectrum for the gluon fusion production mechanism is harder compared with the most precise calculation available (NNLO with re-summation to NNLL order), obtained with MCFM. Therefore, an event-by-event re-weighting is applied to POWHEG-BOX signal events, based on the generated Higgs p_T. Figure 4.5 shows the effect of such a re-weighting on the p_T spectrum of the generated Higgs boson for a couple of representative mass points.



(a) $m_H = 350\,GeV/c^2$ (b) $m_H = 500\,GeV/c^2$

Figure 4.5: Gluon-fusion produced Higgs p_T spectrum as generated at NLO with POWHEG-BOX (black continuous line) and after a re-weighting to account for NNLO correction (red dashed line). A Higgs mass of 350 GeV/c^2 is assumed in the left plot, of 500 in the right one.

For W+jets, Z+jets, and $t\bar{t}$+jets processes the leading-order MADGRAPH generator is used, able to generate vector bosons/$t\bar{t}$ pairs together with up to four partons at the matrix element level, successively matched to the parton shower. Instead, POWHEG-BOX is chosen for single-top events, generated at next-to-leading order. Finally, diboson and QCD events are generated with PYTHIA [77].

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 [78]. 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 [79].

dataset name: /[SingleElectron][SingleMu]/XXX/AOD	run range	$\mathcal{L}$ (fb^{-1})
/Run2011A-May10ReReco-v1/	160431 – 163869	0.211 ± 0.005
/Run2011A-PromptReco-v4/	165088 – 167913	1.0 ± 0.2
/Run2011A-05Aug2011-v1/	170249 – 172619	0.390 ± 0.009
/Run2011A-PromptReco-v6/	172620 – 173692	0.71 ± 0.02
/Run2011B-PromptReco-v1/	175832 – 180252	2.72 ± 0.06
Total	160431 – 180252	5.0 ± 0.1

(a) Data samples

sample name: /XXX/Fall11-PU_S6_START42_V14B-v1/AODSIM	$\mathcal{L}^{eq.}$ (fb^{-1})
/GluGluToHToWWTo[LNu][TauNu]QQ_M-*_7TeV-powheg-pythia6/	~ 1000
/VBF_HToHToWWTo[LNu][TauNu]QQ_M-*_7TeV-powheg-pythia6/	~ 9000
/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola/	2.5
/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/	11.3
/WW_TuneZ2_7TeV_pythia6_tauola/	98.3
/WZ_TuneZ2_7TeV_pythia6_tauola/	234.0
/ZZ_TuneZ2_7TeV_pythia6_tauola/	710.0
/TTJets_TuneZ2_7TeV-madgraph-tauola/	23.5
/T_TuneZ2_t-channel_7TeV-powheg-tauola/	93.0
/Tbar_TuneZ2_t-channel_7TeV-powheg-tauola/	85.9
/T_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/	81.5
/Tbar_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/	41.1
/T_TuneZ2_s-channel_7TeV-powheg-tauola/	103.0
/Tbar_TuneZ2_s-channel_7TeV-powheg-tauola/	95.8
/QCD_Pt-*to*_EMEnriched_TuneZ2_7TeV-pythia6/	0.03
/QCD_Pt-*to*_BCtoE_TuneZ2_7TeV-pythia6/	0.02
/QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6/	0.28

(b) Monte Carlo samples

Table 4.4: (a) Summary of datasets used and run ranges of applicability, together with the corresponding integrated luminosity for each sub-period. (b) Summary of Monte Carlo samples used in the analysis and their equivalent integrated luminosity. Higgs masses range from 250 to 600 GeV/c^2. QCD samples are divided in 20-30, 30-80, and 80-170 GeV/c $\hat{p}_T$ bins.

Pile-up re-weighting

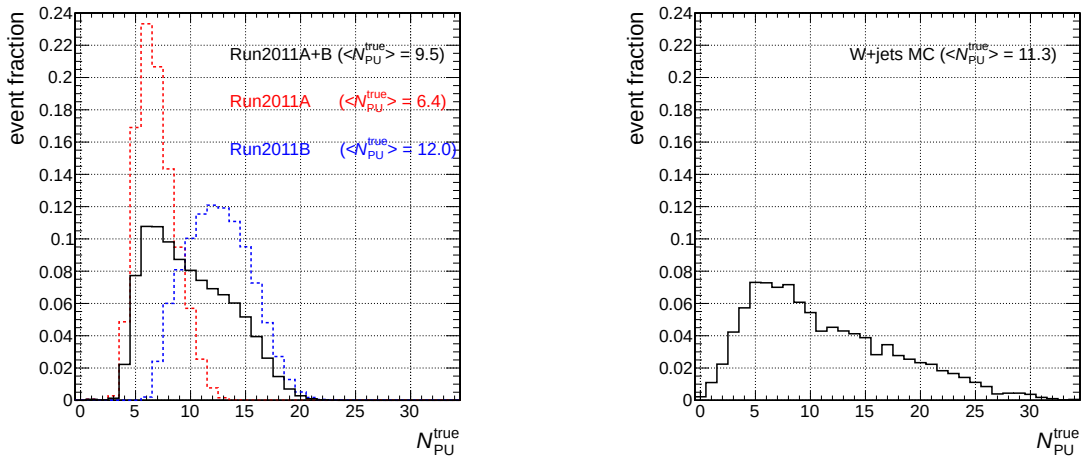
To better reproduce the actual data taking conditions, where the probability that multiple long-distance interactions occur in the same bunch crossing (in-time pile-up) of an interesting head-on collision is significant, minimum-bias simulated events are added on top of the hard scattering in Monte Carlo samples. Out-of-time pile-up, i.e. spurious interactions from a contiguous bunch crossing, and therefore anticipated or delayed of 50 ns, are simulated as well.

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

$$\left(N_{\mathrm{PU}}^{\mathrm{true}}\right)_i = \frac{L_i \cdot \sigma_{\text{min. bias}}}{\nu_{\mathrm{orbit}}} \quad , \tag{4.2}$$

where 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). Figure 4.6a shows the distribution of such $N_{\mathrm{PU}}^{\mathrm{true}}$ for the whole 2011 dataset and for the `Run2011A` and `Run2011B` sub-periods separately. A value of 68 mb is assumed for $\sigma_{\text{min. bias}}$, as a result of a best-fit analysis comparing the distribution of the number of reconstructed vertices in data and Monte Carlo for $Z \to \mu^+\mu^-$ events. This value is compatible with a direct measurement of the total inelastic cross-section performed by CMS in 2011, using the forward calorimetry [80].

In Monte Carlo, the actual number of PU interactions superimposed to the hard scattering in each event is randomly taken from a Poisson distribution, having $N_{\mathrm{PU}}^{\mathrm{true}}$ as mean parameter. The distribution of $N_{\mathrm{PU}}^{\mathrm{true}}$ in MC samples is therefore chosen to be similar to the one observed in data. In Figure 4.6b, such a distribution for a W+jets sample is shown.
In order to account for the residual differences, a weight is applied to simulated events so that the two distributions match exactly.

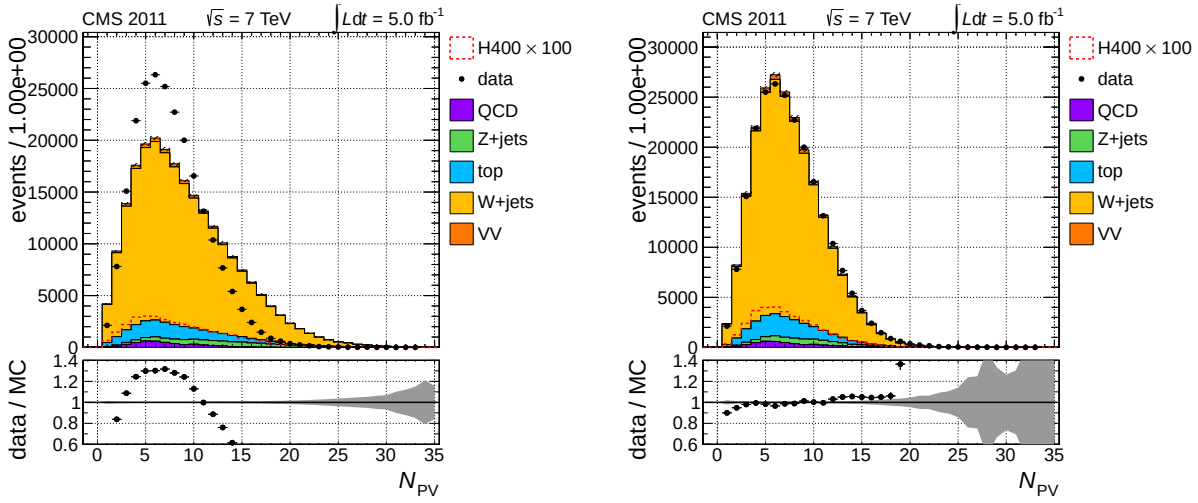


(a) Data (b) Monte Carlo

Figure 4.6: Distribution of $N_{\mathrm{PU}}^{\mathrm{true}}$ in (a) data and (b) Monte Carlo.

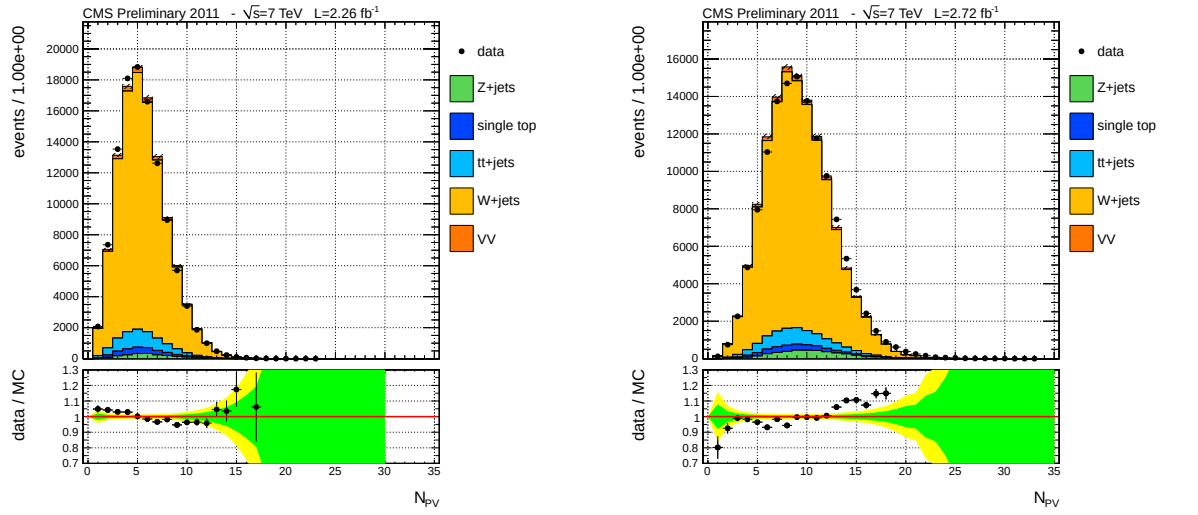
As a check of the effectiveness of this re-weighting procedure, in Figure 4.7 and 4.8 we compare, in data and Monte Carlo, the distribution of the number of vertices reconstructed from pixel tracks, being this quantity closely related to the number of pile-up interactions superimposed

to the main event.



(a) Without PU re-weighting (b) With PU re-weighting

Figure 4.7: Data/MC comparison of the number of reconstructed primary vertices (a) before and (b) after the re-weighting for the pile-up is applied.



(a) `Run2011A` (b) `Run2011B`

Figure 4.8: Data/MC comparison of the number of reconstructed primary vertices, divided in (a) `Run2011A` and (b) `Run2011B`, after applying the pile-up re-weighting factor. The stack of Monte Carlo samples is normalized to the number of events in data.

4.2.3 Trigger strategy on data

The eν$_e$ qq̄ final state is triggered by single electron triggers only for the first period of the data taking, while for higher instantaneous luminosities events are selected with the cross trigger based on one electron plus transverse mass m_T thoroughly described in Section 3.2, and only briefly recalled here. The transverse mass is calculated using the decay product of the leptonic

W at HLT level, i.e. the online electron, and the HLT Particle Flow missing transverse energy $H_{\mathrm{T}}^{\mathrm{miss}}$.

The electron E_{T} threshold at trigger level is 32 GeV for most of the data-taking period, except for the very beginning of `Run2011A`, when it was possible to trigger electrons down to 27 GeV, and for other sub-periods before the September technical stop.
Electron candidates at trigger level fulfill some identification and isolation criteria. The so-called `VT_T_T_T` working point is used for the first part of 2011, while two working points called `WP80(WP70)` (efficiency on real electrons $\sim 80(70)\%$) are used for the remaining periods. The three different sets of conditions have been detailed in Table 3.2.
Depending on the data taking period, the PF m_{T} was requested to be larger than 40(50) GeV/c^2. In Table 4.5a a summary of the electron trigger path names used and their range of validity is reported.

For the $\mu\nu_\mu\,q\bar{q}$ final state, single isolated muon triggers with a p_{T} threshold of 24 GeV/c can be used throughout the full period of data taking, as detailed in Table 4.5b. To cope with the increasing interaction rate at the end of 2011, it is sufficient to restrict the pseudorapidity range of muon candidates within $|\eta| < 2.1$. This is anyway harmless for the analysis, since final state muons are required to be central (see Section 4.3).

4.2.4 Trigger strategy on Monte Carlo

Although Monte Carlo events do not need to fire any trigger to be saved, an accurate reproduction of the online trigger requirements is needed in order for MC samples to faithfully describe the observed data, due to inefficiencies and/or inhomogeneities of the CMS trigger system response. Because of the complex HLT evolution during the 2011 data taking, it has not been possible to reproduce all the trigger paths that were run online in the MC simulations. To compensate for the missing path emulations, and to correct for possible differences between data and Monte Carlo, trigger efficiencies have been measured on data. The efficiency of the electron and muon trigger legs have been computed by means of a tag-and-probe technique, as described in Section 4.5. In the $e\nu_e\,q\bar{q}$ final state, the efficiency of the m_{T} requirement of the trigger is measured on data as well, as described in Section 4.6.

For the electron channel, the triggers paths available in Monte Carlo are `HLT_Ele32_VT_T_T_T`, `HLT_Ele27_WP80_PFMT50` and `HLT_Ele32_WP70_PFMT50`. They correspond to periods e-ii and e-iv to e-vi (cfr. Table 3.4 for the naming convention). In periods e-i and e-iii, the trigger in Monte Carlo is almost identical to the one required on data, apart from a different electron E_{T} threshold or a different m_{T} cut. Therefore, simulated events are required to have fired these HLT paths, and an efficiency scale factor is further applied to account for differences in the HLT implementation (E_{T} or m_{T} thresholds) or for more general mis-modeling of Monte Carlo events.

For the muon channel, the HLT paths `HLT_IsoMu24` and `HLT_IsoMu24_eta2p1` are emulated in Monte Carlo, covering the full data taking period. Therefore, also in this case simulated events are required to have fired these trigger paths, and efficiency correction factors are further applied to account for data/MC differences.

period	run range	dataset name	trigger path name
(e-i)	160404 - 163869	Run2011A-May10ReReco-v1	HLT_Ele27_VT_T_T_T_v1/2/3
(e-ii)	165088 - 165633	Run2011A-PromptReco-v4	HLT_Ele32_VT_T_T_T_v3
(e-iii)	165970 - 166967	Run2011A-PromptReco-v4	HLT_Ele25_WP80_PFMT40_v1
(e-iv)	167039 - 167913	Run2011A-PromptReco-v4	HLT_Ele27_WP80_PFMT50_v1
(e-v)	170249 - 172619	Run2011A-05Aug2011-v1	HLT_Ele32_WP70_PFMT50_v3
(e-vi)	172620 - 173692	Run2011A-PromptReco-v6	HLT_Ele32_WP70_PFMT50_v3/4
(e-vii)	175832 - 180252	Run2011B-PromptReco-v1	HLT_Ele32_WP70_PFMT50_v4/8/9

(a) Trigger paths for electrons

period	run range	dataset name	trigger path name
(μ-i)	160404 - 163869	Run2011A-May10ReReco-v1	HLT_IsoMu24_v1/2
(μ-ii)	165088 - 167913	Run2011A-PromptReco-v4	HLT_IsoMu24_v4/5/6/7
(μ-iii)	170249 - 172619	Run2011A-05Aug2011-v1	HLT_IsoMu24_v8
(μ-iv)	172620 - 173198	Run2011A-PromptReco-v6	HLT_IsoMu24_v8
(μ-v)	173236 - 173692	Run2011A-PromptReco-v6	HLT_IsoMu24_eta2p1_v3
(μ-vi)	175832 - 180252	Run2011B-PromptReco-v1	HLT_IsoMu24_eta2p1_v3/6/7

(b) Trigger paths for muons

Table 4.5: List of trigger paths used to select events for this analysis, in the final state with (a) electrons and (b) muons. VT_T_T_T stands for CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT and is made explicit in Table 3.2, as well as the WP80(70) working points.

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 from a W boson decay. In this section the criteria applied to select final-state objects in the event are described first, then the requirements made at the event level are described.

4.3.1 Object reconstruction

The analysis relies on the standard reconstruction algorithms produced by the CMS community. Particle Flow (PF) objects are used for lepton, jet, and missing transverse energy reconstruction (more details have been given in Section 2.3.3).

Electrons

For the $e\nu_e q\bar{q}$ final state, the electron candidates have to pass several requirements in order to be identified as real electrons:

- candidates are required to pass electron identification cuts according to a simple cut-based selection, which has roughly an efficiency of 70% on real electrons (WP70). This electron ID relies on four variables with different cut values for the barrel and the endcap regions. The four variables are:
  - $\sigma_{i\eta i\eta}$: electron supercluster (sc) η width;
  - $|\phi_{sc} - \phi_{tk}|$: difference between the ϕ of the supercluster back-propagated to the vertex and the ϕ of the track;
  - $|\eta_{sc} - \eta_{tk}|$: difference between the η of the supercluster back-propagated to the vertex and the η of the track;
  - H/E: fraction of deposited energy in the hadronic calorimeter and the electromagnetic calorimeter by the electron candidate.

  The cut values used in the analysis can be found in Table 4.6 and are chosen to be at least as tight as the cuts applied at trigger level (cfr. Table 3.2);
- the electron transverse momentum is always chosen to be at least 3 GeV/c larger than the HLT threshold to avoid turn-on effects because of the trigger not being fully efficient yet. The electron p_T cut is then 30–35 GeV/c, depending on the trigger period;
- the electron pseudorapidity spans over all the tracker acceptance region, i.e. $|\eta| < 2.5$. There is an exclusion range due to the ECAL barrel-endcap transition region, defined by $1.4442 < |\eta_{sc}| < 1.566$ where, η_{sc} is the pseudorapidity of the ECAL supercluster;
- in order to make sure that the selected electrons come from the primary hard interaction and not from a pile-up vertex, we require that the z coordinate of the primary vertex and the z coordinate of the electron inner track vertex lie within a distance of less than 0.2 cm. We also cut on the absolute value of the transverse impact parameter calculated with respect to the primary vertex, requiring it to be smaller than 0.02 cm;

- the selected electron candidates have to be isolated, i.e. low hadronic activity is expected around them since they come from the decay of a boson of the weak interaction. A generic isolation variable I can be defined adding together other particles' energy deposits in a cone around the electron track. The cone is defined in the $\eta - \phi$ plane by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \leq 0.3$.
  First of all, we make use of tracker isolation I_{tk}, which is the sum of p_{T} of the tracks inside the cone. The actual cut is applied on the relative isolation variable, defined as $I_{\text{tk}}^{\text{rel}} = I_{\text{tk}}/p_{\text{T}}^{\text{ele}}$, which is required to be smaller than 0.1. This cut is chosen since the isolation variable computed with tracker deposits is only mildly dependent on the pile-up.
  We further require the electron to be isolated simultaneously in the tracker and in the electromagnetic and hadronic calorimeters. The fraction of energy deposited in the calorimeters within the isolation cone by particles from pile-up interactions can become sizable in conditions of high pile-up, as it was in the second part of the data taking period. We therefore apply a correction subtracting the average pile-up energy from the isolation cone. The combined relative isolation is then defined as

$$I_{\text{comb}}^{\text{rel}} = \frac{I_{\text{trk}} + I_{\text{em}} + I_{\text{had}} - \rho_{\text{PU}} \cdot \pi\Delta R_{\text{eff}}^2}{p_{\text{T}}^{\text{ele}}} \quad . \tag{4.3}$$

  In Equation 4.3, ρ_{PU} is the average density of energy from pile-up particles, estimated on an event-by-event basis as the median energy density in the $\eta - \phi$ plane of all jets[3], while $\pi\Delta R_{\text{eff}}^2$ is the isolation cone effective area, to which the pile-up energy density is subtracted. To determine the cone effective area, the pile-up energy density as well as the combined isolation variable are studied as a function of the number of reconstructed vertices in the event (i.e., as a function of the number of PU interactions), as shown in Figure 4.9. Both quantities exhibit a linearly increasing trend with the number of vertices, although with different slopes. The ratio of the slopes of two linear fits to these quantities can be interpreted as $\pi\Delta R_{\text{eff}}^2$. A value of about 0.20 is found in data for the effective cone size, while in Monte Carlo $\Delta R_{\text{eff}} \simeq 0.18$.
  Finally, with this correction applied, the electron candidate is required to have $I_{\text{comb}}^{\text{rel}} < 0.05$ in order to be considered isolated;
- in order to reject events where the electron originates actually from a conversion of a photon into an e^+e^- pair in the tracker material, the number of missed inner tracker layers of the electron track is required to be exactly zero, i.e. there are no missed layers before the first hit of the electron track from the beam line. In addition any event in which the selected electron is close in space to a partner track compatible with a photon conversion is rejected ($|\Delta \cot\theta| < 0.02$ and $|\text{dist}| < 0.02$, being these quantities the distance of the two tracks in the longitudinal and transverse plane respectively).

[3] Particle Flow jets built with a k_{T} algorithm with R parameter of 0.6 are used for ρ_{PU} evaluation. A large number of very soft "ghost" particles is added uniformly to the event before jet clustering in order to allow a uniform tessellation of the $\eta - \phi$ plane.

variable	WP70 cuts		WP95 cuts	
	EB	EE	EB	EE
$\sigma_{i\eta i\eta}$	0.010	0.030	0.010	0.030
$\lvert\Delta\phi_{\text{tk-sc}}\rvert$	0.030	0.020	0.800	0.700
$\lvert\Delta\eta_{\text{tk-sc}}\rvert$	0.004	0.005	0.007	0.010
H/E	0.025	0.025	0.150	0.070

Table 4.6: Cut values of electron identification variables for WP70 (barrel and endcap), used for the tight electron selection, and WP95 (barrel and endcap), used in the loose electron selection.

Muons

In the $\mu\nu_\mu\,q\bar{q}$ final state, the candidates of the reconstructed muon collection have to pass several requirements in order to be identified as real muons:

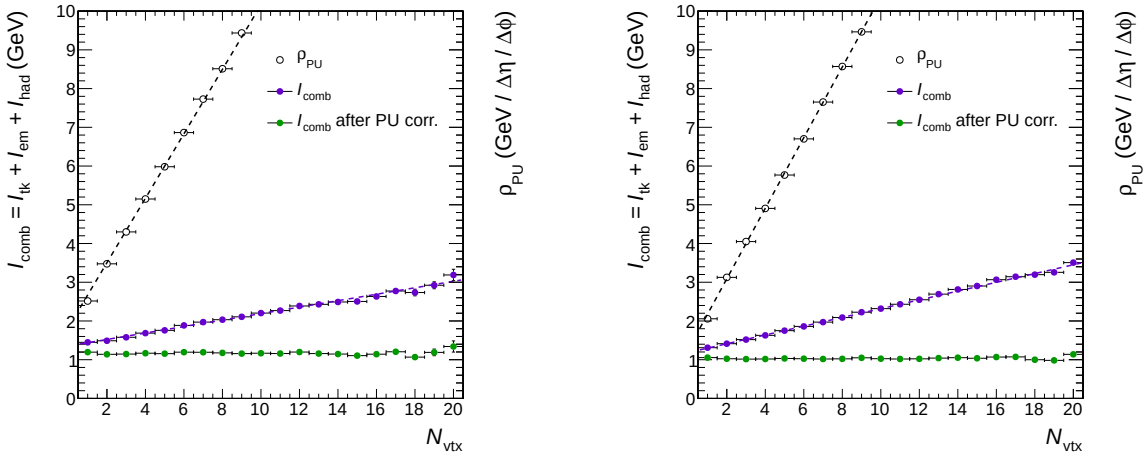
- the muon candidate has to be reconstructed both in the tracker and in the muon chambers;
- the HLT conditions allows to keep the muon transverse momentum requirement as low as 27 GeV/c for the whole running period;
- the muon pseudorapidity spans the muon trigger geometrical acceptance region, that is $|\eta| < 2.1$;
- as done for electrons, cuts on the muon track impact parameter along the z coordinate and in the transverse plane of respectively 0.2 cm and 0.02 cm ensure that the muon does not originate from a pile-up vertex;
- muon identification helps rejecting fake muons from QCD jets and consists in cutting on track quality parameters. In particular, we require at least one pixel hit found for the inner track of the muon and the total number of tracker hits larger than 10. Finally, the global track reduced chi-square must be smaller than 10.
- the selected muon candidates also have to be isolated. As for electron candidates, we require both a tracker relative isolation and a combined, PU-corrected relative isolation. The value of the effective cone size for muons is $\Delta R_{\text{eff}} \simeq 0.17$ in data and Monte Carlo. The cuts applied are then $I^{\text{rel}}_{\text{trk}} < 0.05$ and $I^{\text{rel}}_{\text{comb}} < 0.1$.

Loose leptons

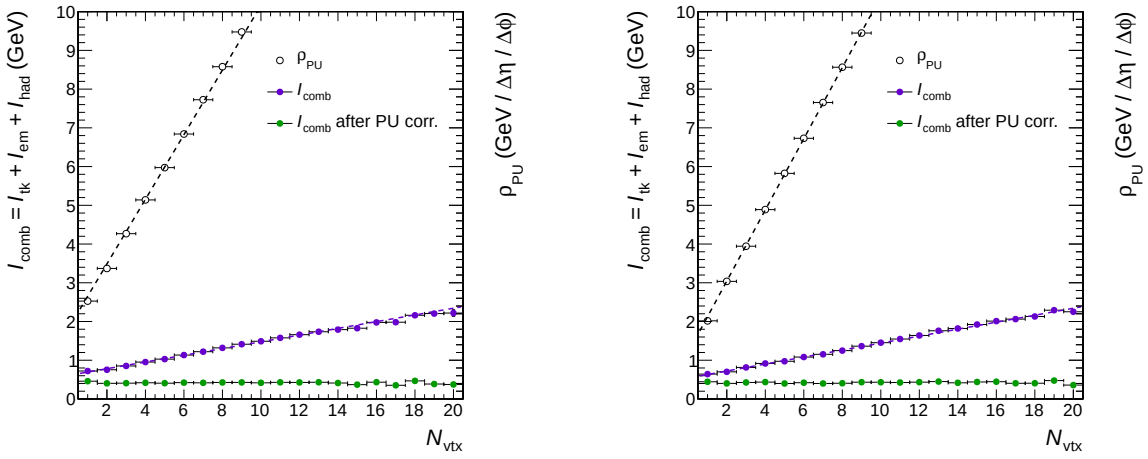
For the purpose of rejecting events with more than one lepton we define loosely-identified leptons by relaxing some of the requirements previously described. We consider electrons which have $p_{\text{T}} > 15\,\text{GeV}/c$, $|\eta| < 2.5$, and $I^{\text{rel}}_{\text{trk}} < 0.2$ and which satisfy electron identification cuts according to WP95 to be loose. The cut values for the identification variables used in the analysis can be found in Table 4.6.
Similarly, we consider all global muons which have $p_{\text{T}} > 10\,\text{GeV}/c$, $|\eta| < 2.5$, and $I^{\text{rel}}_{\text{trk}} < 0.2$ to be loose muons.

Jets

The whole event information is used to assign all energy deposits and tracker hits to individual particles, thanks to the Particle Flow algorithm. Once final-state isolated electrons or muons are



(a) Electrons



(b) Muons

Figure 4.9: Effective area calculation: the bare combined isolation variable I_{comb} (purple dots in the plot), as well as the energy density from PU particles ρ_{PU} (empty dots in the plot) are studied as a function of the number of vertices in the event N_{vtx} (data on the left, Monte Carlo on the right). The cone effective area is obtained as the ratio of the slopes of the two quantities. Pile-up corrected isolation is represented in the plot (green dots) and shows the effectiveness of the empiric correction.

excluded from the list, remaining particles are clustered together to form jets. The clustering algorithm used is the anti-k_{T}, with R parameter 0.5. Only jets with $p_{\text{T}} > 30\,\text{GeV}/c$ and $|\eta| < 2.4$ are considered in the analysis, that is only jets within the tracker acceptance region are counted and selected. The L1 (offset), L2 (relative) and L3 (absolute) corrections are applied both in data and in the simulated samples, while residual corrections applied on data account for additional data/MC differences. More details can be found in Section 2.3.3.

A few requirements are applied on reconstructed jets with the goal to reject fake, badly reconstructed and calorimeter noise jets, while retaining more than 98%–99% of real jets. The quality cuts are the following:

- 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 exactly 0, while neutral hadron 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.

In order to account for good leptons which are also reconstructed as jets, we finally require the jets to lie outside a cone of radius $\Delta R = 0.5$ around the selected lepton.

4.3.2 Event-level cuts

The event selection can be subdivided into two logical steps. A first one consists in a basic pre-selection with the aim of requiring the final-state objects and applying minimum criteria to sit as much as possible on trigger efficiency plateaus. Secondly, further cuts on the pre-selected events have the purpose of enhancing the signal over the large background rate. Both steps are detailed here below:

Pre-selection

- The events are required to have a good primary vertex (PV). The primary vertex is selected as the one with the highest sum of p_{T}^2 of the associated tracks, and it is required to have a number of degrees of freedom $N_{\mathrm{dof}} \geq 4$. In addition, the PV must lie in the central detector region of $|z| \leq 24\,\mathrm{cm}$ and $\rho \leq 2\,\mathrm{cm}$ around the nominal interaction point;
- the events are required to have exactly one tight lepton candidate (electron or muon) fulfilling the criteria described in Section 4.3.1. We reject events containing one or more additional loose leptons of the same flavour or any other lepton of the other flavours;
- in both the electron and the muon+jets final states events are required to have large missing transverse energy from the undetected neutrino, that is $E_{\mathrm{T}}^{\mathrm{miss}} > 30\,\mathrm{GeV}$ and to have a leptonic W transverse mass $m_{\mathrm{T}} > 30\,\mathrm{GeV}/c^2$, where the transverse mass is defined as

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

  where $\Delta\phi$ is the angle between the lepton momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{lepton}}$ and the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ in the transverse plane. These cuts are designed to reduce the background contamination from QCD multi-jet production and Drell-Yan. For $\mathrm{e}\nu_{\mathrm{e}}\,\mathrm{q}\bar{\mathrm{q}}$ final states, a higher threshold at 50 GeV/c^2 is applied since when the same selection appears online in the HLT requirement;
- events are required to have exactly two or three jets passing the cuts described in Section 4.3.1. 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.
  Since in final states with three jets the $\mathrm{t\bar{t}}$ contribution to the total background is significant, a b-jet veto is applied in those events to reduce its contamination. Events are discarded if one or more jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}/c$ and $|\eta| < 2.4$ is identified to

be generated from a b quark, using a b-tag discriminator based on the counting of tracks with large impact parameter within the jet[4];

Final selection

- The most effective selection in reducing the large W+jets contamination profits from the existence of the hadronic W boson resonance. In W+jets events the jet pair originates from initial and/or final state gluon or quark radiation. Due to the kinematic selections, the di-jet invariant mass has a broad peak around $80\,\mathrm{GeV}/c^2$ with large low- and high-energy tail from jet combinatorics. In signal events, on the contrary, the di-jet system invariant mass resonates around m_W, within jet energy resolution. For the final analysis, only a narrow region around m_W is retained as the signal phase space. Events which fail the di-jet mass requirement are nevertheless kept and categorized as sideband events. Their usage will be addressed in Section 4.8 To summarize, the signal and the sidebands phase spaces are defined as

$$\begin{aligned} \text{signal region:} \quad & 65 < m_\mathrm{jj} < 95\,\mathrm{GeV}/c^2 \\ \text{sideband region:} \quad & 55 < m_\mathrm{jj} < 65\,\mathrm{GeV}/c^2 \quad \vee \quad 95 < m_\mathrm{jj} < 115\,\mathrm{GeV}/c^2 \end{aligned}$$

- in signal events, the Higgs decay products tend to be emitted in the central part of the detector, at variance with the behaviour of W+jets event. Centrality cuts can therefore be applied on the η of the lepton and of the di-jet system. In particular, we require $|\eta_\mathrm{lepton}| < 1.5$ and $|\eta_\mathrm{jj}| < 3$;
- the W bosons from Higgs decay can have a large transverse momentum for medium-to-high Higgs masses. Since the W decay products are emitted at small opening angles for large boosts, we require $|\Delta\eta_\mathrm{jj}| < 1.5$. This cut reduces the contamination from W+jets background without cutting much in the signal phase space and preserving the $\mathrm{l}\nu_\mathrm{l}\,\mathrm{jj}$ invariant mass shape, at variance with the effect of a cut on $\Delta\phi_\mathrm{jj}$, for example. This is important to correctly model the background for the Higgs cross-section limit extraction, as explained in Section 4.10.

The aforementioned cuts have been tuned using $S/\sqrt{S+B}$ as figure of merit, where S and B are respectively the number of signal and background events expected from the Monte Carlo simulation in $5.0\,\mathrm{fb}^{-1}$. Since the analysis strategy consists in looking for a signal peak appearance over a large background continuum, only those events for which the reconstructed four-body mass $m_{\mathrm{l}\nu_\mathrm{l}\,\mathrm{jj}}$ lies in a region around the nominal Higgs mass have been considered in the significance computation (exact values of the mass windows are listed in Table 4.7). A detailed description of the techniques applied to reconstruct the Higgs invariant mass is postponed to next section.

The result of this cut optimization is presented in Figure 4.11, while Figure 4.10 shows the expected significance at pre-selection level and after the cuts here described are applied.

[4]The so-called TrackCountingHighEfficiency algorithm is used, and the cut value of 3.3 correspond to the medium operating point (TCHEM) defined by the b-tag working group [81]

m_H (GeV/c^2)	250	300	350	400	450	500	550	600
$m^{min}_{l\nu_l jj}$ (GeV/c^2)	225	270	310	350	390	430	470	515
$m^{max}_{l\nu_l jj}$ (GeV/c^2)	275	330	390	450	510	575	640	700

Table 4.7: Definition of reconstructed $l\nu_l jj$ mass windows for the different Higgs mass hypotheses. Values are extracted from the study of the $S/\sqrt{S+B}$ figure of merit as a function of the window size, shown in Figure 4.14.

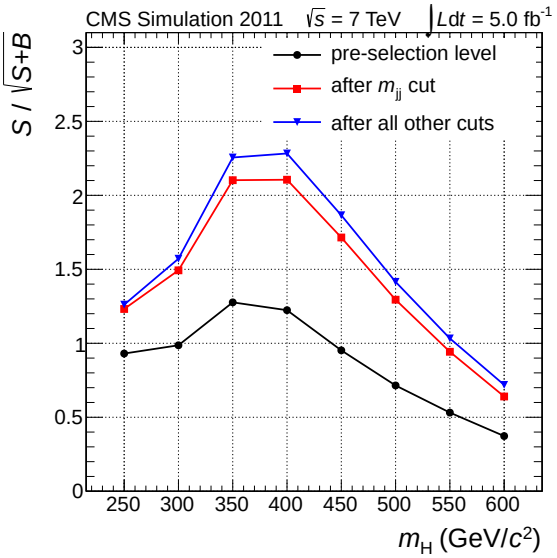


Figure 4.10: Higgs signal significance, expressed in terms of $S/\sqrt{S+B}$, for three different analysis steps: just after preselection (in black), after cutting on the di-jet invariant mass (in red), and after centrality and $\Delta\eta_{jj}$ cuts (in blue). S and B numbers of events are counted in mass windows of $m_{l\nu_l jj}$, as defined in Table 4.7.

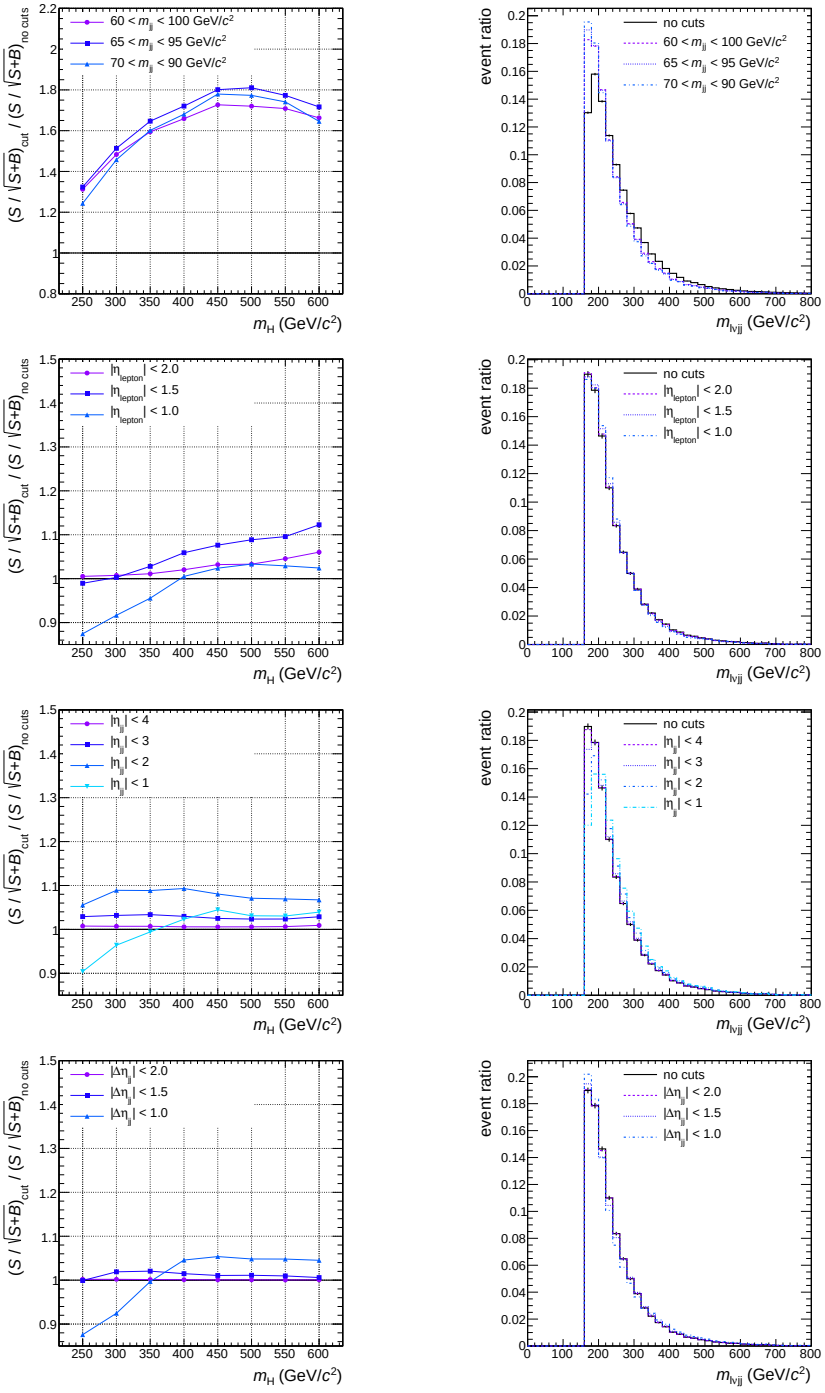


Figure 4.11: Left plots: cut optimization study, using $S/\sqrt{S+B}$ as a figure of merit (f.o.m.). The relative impact on the significance of different cuts is represented with respect to a reference significance. In the top plot the reference value of the f.o.m. is computed just after pre-selection, while in all other plots the cut on m_{jj} is assumed. S and B numbers of events are always counted in mass windows of $m_{\mathrm{l\nu_l\,jj}}$, as defined in Table 4.7. Right: sculpting of the reconstructed Higgs invariant mass $m_{\mathrm{l\nu_l\,jj}}$ after applying the different cuts. The four rows of plots, from top to bottom, refer to cuts on m_{jj}, η_{lepton}, η_{jj}, and $\Delta\eta_{\mathrm{jj}}$ respectively.

4.4 Higgs mass reconstruction

Despite the presence of missing transverse energy from the undetected neutrino, the full invariant mass of the Higgs candidates can be reconstructed in this channel. The two selected jets fully reconstruct the hadronic W boson, while the leptonic W boson is reconstructed adding together the selected lepton, the missing transverse energy, and imposing a further constraint to recover the undetected neutrino p_z, as described here below.

4.4.1 Neutrino p_z determination

The transverse components of the neutrino momentum correspond to the transverse missing energy (within $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction resolution), since no other undetectable particle is present in the final state. There is only one unknown parameter, i.e. the neutrino p_z, which can anyway be determined *a posteriori*, imposing that the the lepton and the neutrino momenta add together to form a W boson of mass $m_{\mathrm{W}} = 80.399\,\mathrm{GeV}/c^2$. Solving the resulting second-order equation one finds:

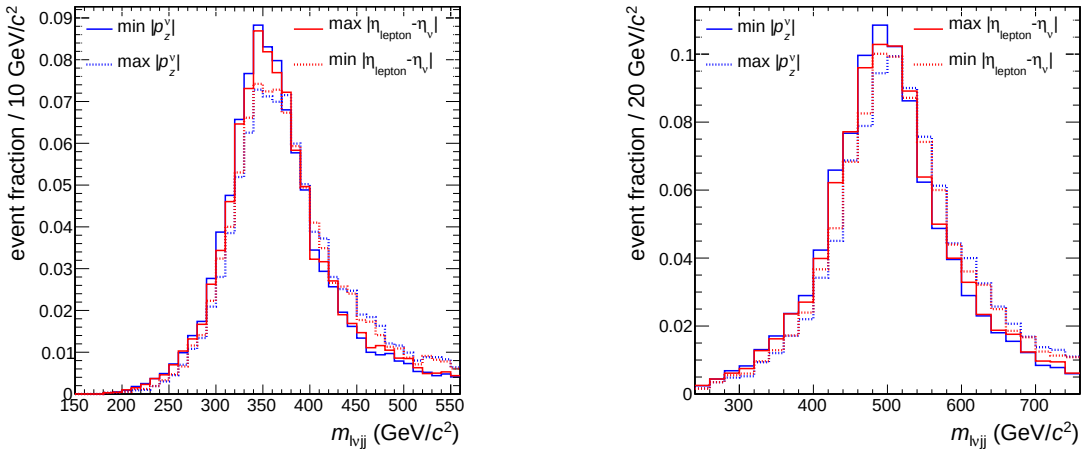
$$
\begin{aligned}
(p_z^{\nu})_{1,2} &= \frac{p_z^{\mathrm{lepton}} \left(\vec{p}_{\mathrm{T}}^{\,\mathrm{lepton}} \cdot \vec{E}_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{W}}^2/2 \right) \pm \sqrt{\Delta}}{\left(p_{\mathrm{T}}^{\mathrm{lepton}} \right)^2} \quad , \quad \text{where} \\
\Delta &= \left(\vec{p}_{\mathrm{T}}^{\,\mathrm{lepton}} \cdot \vec{E}_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{W}}^2/2 \right)^2 - \left(p_{\mathrm{T}}^{\mathrm{lepton}} \right)^2 \left(E_{\mathrm{T}}^{\mathrm{miss}} \right)^2 .
\end{aligned}
\tag{4.4}
$$

Due to the finite detector resolution and errors in the reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$, it can happen that the discriminant Δ is smaller than 0, in which cases we force it to 0 and a unique solution is obtained for p_z^{ν}. Depending on the Higgs mass, this happens in about 31–39% of cases. For the remaining events, the discriminant Δ is larger than 0, 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 truth that this choice corresponds to the correct solution in the majority of cases (61–65% depending on the Higgs mass). In Figure 4.12 the effect on the reconstructed $\mathrm{l}\nu_{\mathrm{l}}\mathrm{jj}$ mass of different criteria for the ambiguity resolution is reported, confirming the good performance of the chosen approach.

4.4.2 Kinematic fit

The possibility to fully reconstruct a mass peak is, together with its high branching ratio, one of the main peculiarities of this Higgs decay channel. The significance of an hypothetical signal peak depends directly on the width of the resonance. This, in turn, is the convolution of two terms: an irreducible component, originating from the Higgs intrinsic decay width Γ_{H}, and an additional smearing caused by detector resolution effects. While the Higgs intrinsic width can become very large at high mass and dominate the peak width, the impact of finite jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is sizable for a light Higgs particle.

A valuable piece of information in the Higgs decay chain is provided by the presence of two intermediate narrow resonances, that is the two W vector boson: the *a-priori* knowledge of their mass can be used to improve the mass resolution in the four-body mass reconstruction. The most efficient way to do so is to implement a two-constraint kinematic fit on both the leptonic and the hadronic side of the Higgs decay chain. On an event-by-event basis, the fit uses the



(a) $m_H = 350\,\mathrm{GeV}/c^2$ (b) $m_H = 500\,\mathrm{GeV}/c^2$

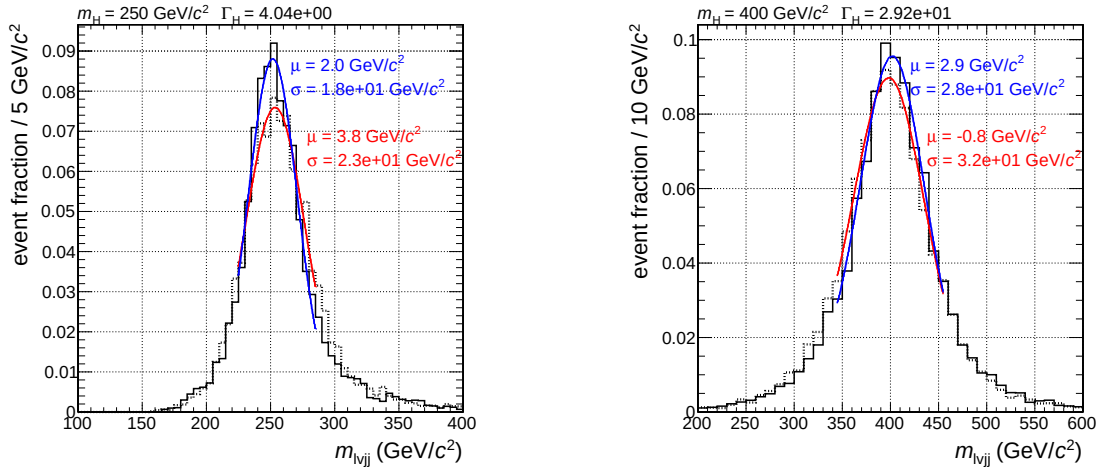
Figure 4.12: Alternative ways to choose among the two distinct solutions of the second-order equation for p_z^ν are compared in terms of resulting four-body mass resolution. Solid (dashed) red lines correspond to choosing the solution that yields the minimum (maximum) $|p_z^\nu|$. Solid (dashed) blue lines correspond to the maximum (minimum) $|\eta_{\mathrm{lepton}} - \eta_\nu|$ criterion.

quadrimomenta p_i of the lepton, neutrino and jets as inputs and modifies them so as to accommodate the external constraints. Each particle quadrimomentum is changed accordingly to the known resolution, parametrized as a function of its transverse momentum and pseudorapidity, and the best value is found through the minimization of a χ^2 variable of this form:

$$\chi^2 = \sum_{i=\mathrm{lepton},\nu,\mathrm{jet}_1,\mathrm{jet}_2} \frac{\left(p_i^{\mathrm{fit.}} - p_i^{\mathrm{meas.}}\right)^2}{\sigma_i^2(p_T,\eta)} + \frac{\left(m_{l\nu} - m_W\right)^2}{\Gamma_W^2} + \frac{\left(m_{jj} - m_W\right)^2}{\Gamma_W^2} \quad , \tag{4.5}$$

where the two constraints have been introduced as Gaussian constraints of width equals to the W boson decay width. The infrastructure of the fit uses the standard CMS kinematic fit tools [82].

The so-obtained four-body mass well reproduces the generator Higgs mass for all mass hypotheses, and the improved resolution obtained with the kinematic fit is one of the ingredients of the Higgs search. Figure 4.13 shows the effect of the kinematic fit on the four-body mass spectrum for signal Higgs mass hypothesis of 250 and 400 GeV/c^2. The distributions obtained before and after the kinematic fit are reported. Each of them is fitted with a Breit-Wigner function convoluted with a gaussian function. The parameters of the Breit-Wigner are fixed to the mass and width expected for that Higgs hypothesis, while the parameters of the Gaussian are left free in the fit and represent the mass shift and extra smearing due to jet energy resolution. In Figure 4.14 the usual figure of merit $S/\sqrt{B}$ is used to perform a signal significance scan as a function of the counting mass window. For low and intermediate Higgs mass points, the advantages of a kinematic-fit based reconstruction of the $l\nu_l jj$ mass are sizable. For Higgs masses $\geq 450\,\mathrm{GeV}/c^2$, little margin of improvement is expected, for the total peak width us dominated by the Higgs intrinsic decay width.



(a) $m_H = 250\,\text{GeV}/c^2$ (b) $m_H = 400\,\text{GeV}/c^2$

Figure 4.13: Four-body mass distributions obtained before (dashed black line) and after (solid black line) the kinematic fit, for Higgs mass hypotheses of 250 and 400 GeV/c^2. Reported on the figure is the result of a fit of the mass spectrum, using a Breit-Wigner function of fixed parameters m_H and Γ_H convoluted with a gaussian function of free parameters μ and σ. The resolution improvement, in terms of Gaussian σ, is sizable (about 15 GeV/c^2 subtracted in quadrature).

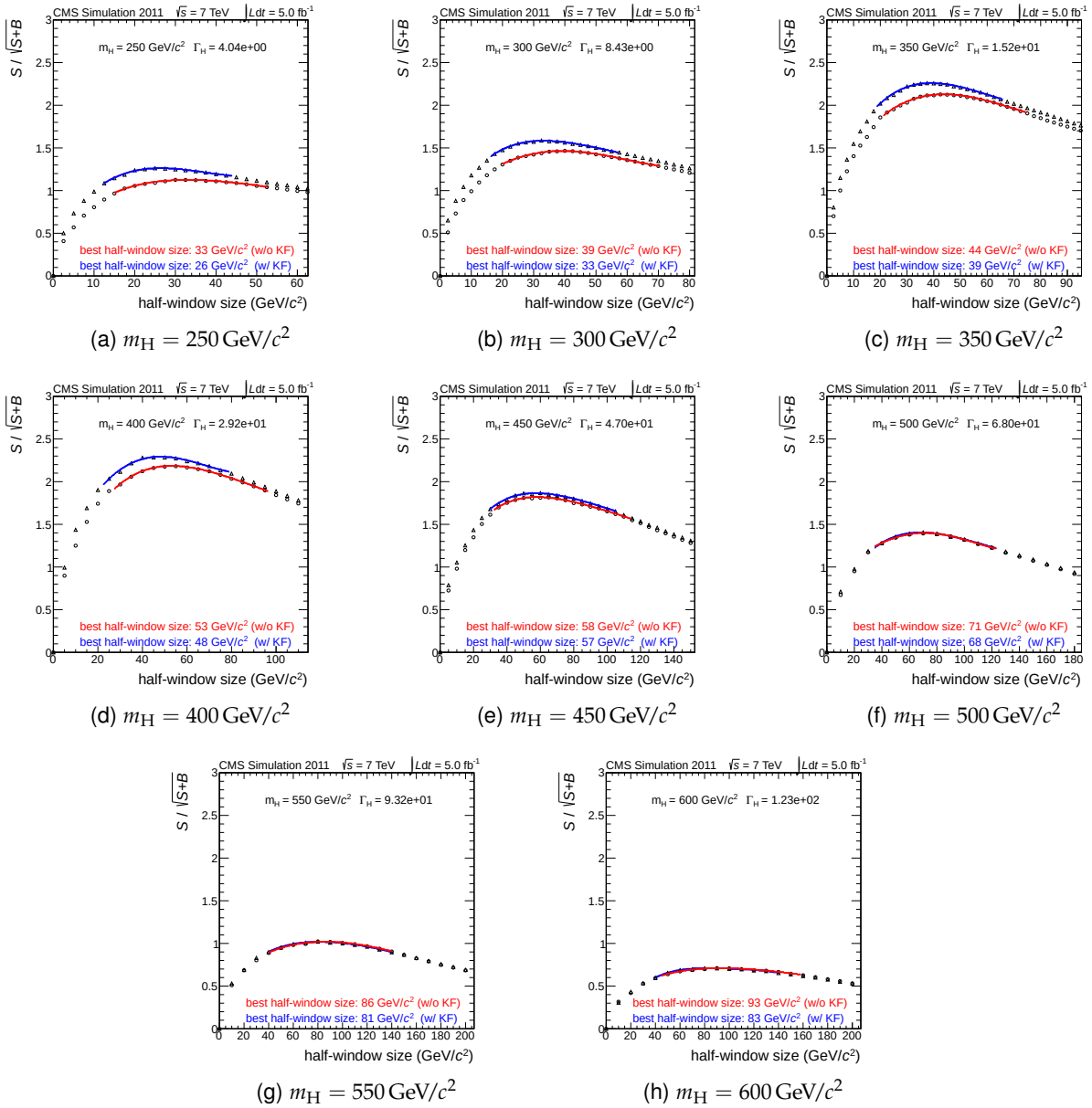
4.5 Lepton reconstruction, selection and trigger efficiencies

Lepton reconstruction, selection and trigger efficiencies are computed using a tag-and-probe technique on $Z \to l^+l^-$ events ($70 < m_{l^+l^-} < 110\,\text{GeV}/c^2$). While an exhaustive description of the method can be found at [83], we will only briefly remind here its most important ingredients.

The tag-and-probe method profits from the Z resonance to select particles of the desired type (electrons or muons), and probe the efficiency of a particular selection criterion on those particles.
In general, one of the two leptons from the Z decay is taken as the "tag" particle, and satisfies a set of very tight selection criteria designed to isolate the required particle type.
The "probe" is then selected with identification criteria that are kept looser than the one under study. To ensure that also the probe is a particle of the required type, and not a fake lepton for instance, a pairing with the tag is required, such that the invariant mass of the combination is consistent with the mass of the Z. The sample of tag-probe pairs under the Z peak is, for most of practical cases, almost background free.
The efficiency itself is then measured by counting the number of probes passing or failing the criteria under study with respect to the total number of probes, i.e.

$$\epsilon = \frac{N_{\text{passing probes}}}{N_{\text{passing probes}} + N_{\text{failing probes}}} \quad . \tag{4.6}$$

Although the requirement of a tag-probe pair close to the Z resonance allows to have a highly pure sample, a possible source of systematic errors in the computation of the efficiencies is the contamination of the probe ensemble by fake objects, which present a higher probability to fail the selection criteria under study than the real objects. Since the rate of fakes is expected



(a) $m_{\mathrm{H}} = 250\,\mathrm{GeV}/c^2$ (b) $m_{\mathrm{H}} = 300\,\mathrm{GeV}/c^2$ (c) $m_{\mathrm{H}} = 350\,\mathrm{GeV}/c^2$

(d) $m_{\mathrm{H}} = 400\,\mathrm{GeV}/c^2$ (e) $m_{\mathrm{H}} = 450\,\mathrm{GeV}/c^2$ (f) $m_{\mathrm{H}} = 500\,\mathrm{GeV}/c^2$

(g) $m_{\mathrm{H}} = 550\,\mathrm{GeV}/c^2$ (h) $m_{\mathrm{H}} = 600\,\mathrm{GeV}/c^2$

Figure 4.14: The figure of merit $S/\sqrt{S+B}$ is plotted as a function of the reconstructed mass window inside which the event counting is performed. Standard mass reconstruction (circles) and kinematic fit based one (triangles) are shown. The kinematic fit guarantees an improved resolution for masses as low as $\sim 450\,\mathrm{GeV}/c^2$: both higher significances and smaller effective mass windows are obtained. These f.o.m. calculations are performed after all cuts of Section 4.3.

to increment away from the Z peak, the efficiency calculation is repeated with two different requirements on the tag-probe invariant mass, one tighter ($80 < m_{l^+l^-} < 100\,\mathrm{GeV}/c^2$) and one looser ($60 < m_{l^+l^-} < 120\,\mathrm{GeV}/c^2$) with respect to the nominal one. The maximal excursion of the three different efficiency values is taken as the associated systematic uncertainty.

The tag-and-probe procedure can be exercised on both data, triggered by electron+supercluster and single-muon HLT paths, and on a Drell-Yan Monte Carlo sample. From the efficiency measurements in data and Monte Carlo, scale factors $\rho = \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{MC}}$ can be calculated and used

as weighting factors in simulated samples to remove any imperfections in the latter ones.

4.5.1 Electron efficiencies

The electron efficiency can be factorized in the product of three distinct components: the reconstruction efficiency ϵ_{RECO}, the identification and isolation efficiency ϵ_{ID} and the trigger efficiency ϵ_{HLT}. In formulae, we have

$$\epsilon_{\text{ele}} = \epsilon_{\text{RECO}} \times \epsilon_{\text{ID}} \times \epsilon_{\text{HLT}} \quad . \tag{4.7}$$

The reconstruction efficiency is the probability of finding a reconstructed GSF track when the electron supercluster is within the ECAL fiducial volume. The probe is selected as an ECAL supercluster of reconstructed transverse energy greater than 20 GeV. To reduce backgrounds, which are not insignificant at this step, we use a tight selection on the tag and require the probe to pass additional loose shower shape and isolation requirements; these are known from simulations to be uncorrelated with the supercluster reconstruction efficiency, which can be safely assumed to be equal to 100%. The measured reconstruction efficiency is then the fraction of probes reconstructed as electron tracks.
The isolation and identification efficiency is given by the fraction of probes surviving the so-called `WP70` working point and the isolation cuts defined in Section 4.3. For this step, electron tracks matched to a supercluster are considered as probes.
Finally, the HLT trigger efficiency is given by the probability of well identified electrons to fire the trigger paths defined in Table 4.5a.

The efficiency measurements are performed in exclusive bins of p_{T} and η of the probe electron, covering the intervals relevant for the analysis ($[30, 35, 40, 45, 50, 60, 75, 200]$ GeV/c in p_{T} and $[-1.5, 0, 1.5]$ in η). The whole 2011 data taking period is divided into five sub-periods for this measurement, according to changing HLT and pile-up conditions, and corresponding to periods e-i, e-ii, e-iii to e-iv, e-v to e-vi, and e-vii as listed in Table 4.5a. The results are then the luminosity-averaged efficiencies in the different sub-periods, as listed in Table 4.8.
In Figure 4.15, the efficiency of the isolation plus identification and HLT steps measured on data is shown as a function of the probe electron p_{T}, for probes with η between 0 and 1.5, averaging over the 2011 data taking. As can be noticed, the chosen working point (`WP70`) has an efficiency of about 70% on real electrons of low p_{T} ($\sim 30\,\text{GeV}/c$), and it increases up to about 85% thanks to the relative isolation cut being less and less tight at high p_{T}. The HLT efficiency plot shows a clear turn-on behaviour around the trigger p_{T} cut value (32 GeV/c for most of the data), which is the combination of L1 and HLT resolution effects, and reaches a very high plateau efficiency at high p_{T}.

4.5.2 Muon efficiencies

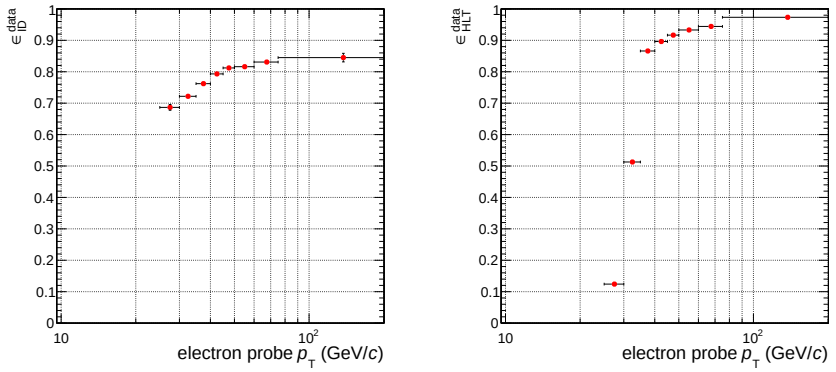
The muon efficiency can be factorized in the product of two distinct components: the identification and isolation efficiency ϵ_{ID} and the trigger efficiency ϵ_{HLT}. In formulae, we have

$$\epsilon_{\text{muon}} = \epsilon_{\text{ID}} \times \epsilon_{\text{HLT}} \quad . \tag{4.8}$$

In the above formula, the reconstruction efficiency has been neglected, since the probability for a muon across the CMS detector to originate from a reconstructed track is 100% (see e.g. [84]).

p_T(GeV/c)	η	ϵ^{data}_{RECO}	ρ_{RECO}	ϵ^{data}_{ID}	ρ_{ID}	ϵ^{data}_{HLT}	ρ_{HLT}
[30, 35]	[−1.5, 0]	0.968±0.006	0.994±0.010	0.731±0.007	0.968±0.022	0.511±0.005	0.742±0.014
	[0, 1.5]	0.967±0.006	0.995±0.011	0.722±0.008	0.952±0.022	0.513±0.004	0.753±0.013
[35, 40]	[−1.5, 0]	0.990±0.002	0.998±0.003	0.761±0.003	0.942±0.009	0.866±0.002	0.951±0.004
	[0, 1.5]	0.989±0.002	0.996±0.004	0.762±0.003	0.941±0.009	0.866±0.002	0.953±0.004
[40, 45]	[−1.5, 0]	0.996±0.001	0.999±0.002	0.797±0.003	0.943±0.006	0.895±0.001	0.953±0.003
	[0, 1.5]	0.996±0.001	0.998±0.001	0.793±0.002	0.936±0.006	0.896±0.001	0.957±0.003
[45, 50]	[−1.5, 0]	0.997±0.001	0.998±0.001	0.809±0.003	0.928±0.007	0.914±0.002	0.957±0.003
	[0, 1.5]	0.996±0.001	0.998±0.002	0.812±0.003	0.932±0.007	0.917±0.002	0.959±0.004
[50, 60]	[−1.5, 0]	0.997±0.001	0.998±0.002	0.819±0.003	0.926±0.008	0.930±0.003	0.958±0.005
	[0, 1.5]	0.996±0.001	0.998±0.002	0.816±0.003	0.921±0.008	0.933±0.002	0.962±0.004
[60, 75]	[−1.5, 0]	0.994±0.004	0.997±0.007	0.828±0.007	0.930±0.016	0.949±0.007	0.972±0.011
	[0, 1.5]	0.992±0.004	0.994±0.007	0.831±0.007	0.933±0.016	0.944±0.004	0.965±0.008
[75, 200]	[−1.5, 0]	0.992±0.004	0.996±0.008	0.851±0.009	0.958±0.020	0.968±0.005	0.984±0.010
	[0, 1.5]	0.995±0.004	1.000±0.008	0.845±0.014	0.954±0.030	0.973±0.006	0.989±0.012

Table 4.8: Luminosity-weighted reconstruction, identification and trigger efficiencies (ϵ) for electrons, and corresponding scale factors (ρ). The statistic and systematic uncertainties are combined.



(a) (b)

Figure 4.15: Efficiency measurements on electron data for the (a) identification and isolation and (b) the HLT step, using the tag-and-probe method. The result is shown as a function of the probe electron p_T, for probes within $0 < \eta < 1.5$.

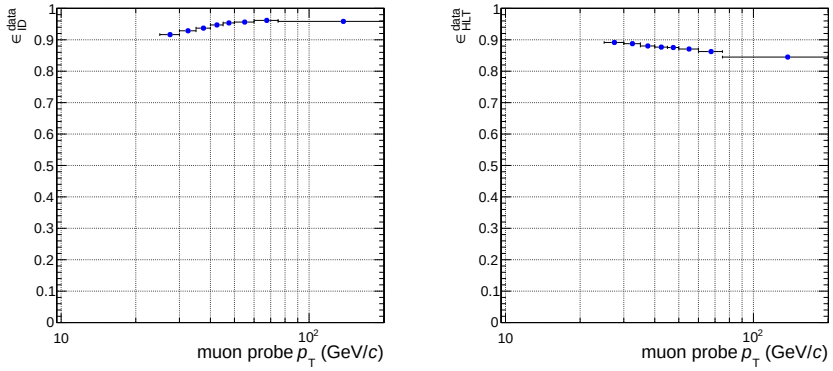
The isolation and identification efficiency is the probability of a reconstructed muon track to survive the good muon requirements listed in Section 4.3.
Finally, the HLT trigger efficiency is given by the probability of well identified muons to fire the trigger paths defined in Table 4.5b.

As done for electrons, also for muons the measurement is binned both in p_T and η of the probe muon, covering the intervals relevant for this analysis. The whole 2011 data taking period is divided into three sub-periods for this measurement, according to changing HLT and pile-up conditions, and corresponding to periods μ-i to μ-iv, μ-v, and μ-vi as listed in Table 4.5b. The results are then the luminosity-averaged efficiencies in the different sub-periods, and can be found in Table 4.9.

In Figure 4.16, the efficiency of the isolation plus identification and HLT steps measured on data is shown as a function of the probe muon p_{T}, for probes with η between 0 and 1.5. As can be noticed, the working point chosen for muons has an vey high efficiency on real muons, above 90% at all p_{T}. Instead, a negative trend with the probe p_{T} is exhibited by the HLT efficiency calculation. This effect originates in a different definition of the isolation cut for the HLT selection and the working point chosen in this analysis. While offline muons are required to survive a relative isolation cut – which is then less and less tight as the muon p_{T} increases – the requirement at HLT level consists of an absolute cut on the isolation variable, irrespective of the muon p_{T}.

p_{T}(GeV/c)	η	$\epsilon_{\mathrm{ID}}^{\mathrm{data}}$	ρ_{ID}	$\epsilon_{\mathrm{HLT}}^{\mathrm{data}}$	ρ_{HLT}
[30, 35]	$[-1.5, 0]$	0.928±0.001	0.991±0.001	0.890±0.001	1.000±0.001
	[0, 1.5]	0.929±0.001	0.995±0.001	0.888±0.001	0.999±0.001
[35, 40]	$[-1.5, 0]$	0.938±0.001	0.992±0.001	0.883±0.001	0.999±0.001
	[0, 1.5]	0.937±0.001	0.995±0.001	0.880±0.001	0.997±0.001
[40, 45]	$[-1.5, 0]$	0.948±0.001	0.992±0.001	0.881±0.001	1.001±0.001
	[0, 1.5]	0.947±0.001	0.994±0.001	0.877±0.001	0.998±0.001
[45, 50]	$[-1.5, 0]$	0.955±0.001	0.993±0.001	0.879±0.001	1.002±0.001
	[0, 1.5]	0.954±0.001	0.994±0.001	0.875±0.001	1.000±0.001
[50, 60]	$[-1.5, 0]$	0.959±0.001	0.993±0.001	0.873±0.001	1.000±0.002
	[0, 1.5]	0.956±0.001	0.994±0.001	0.871±0.001	1.000±0.002
[60, 75]	$[-1.5, 0]$	0.959±0.001	0.992±0.001	0.860±0.002	0.999±0.003
	[0, 1.5]	0.962±0.001	0.997±0.001	0.863±0.002	1.000±0.003
[75, 200]	$[-1.5, 0]$	0.964±0.002	0.993±0.002	0.844±0.003	0.994±0.004
	[0, 1.5]	0.959±0.002	0.990±0.002	0.845±0.003	0.997±0.004

Table 4.9: Luminosity-weighted identification and trigger efficiencies (ϵ) for muons, and corresponding scale factors (ρ). The statistic and systematic uncertainties are combined.



(a) (b)

Figure 4.16: Efficiency measurements on muon data for the (a) identification and isolation and (b) the HLT step, using the tag-and-probe method. The result is shown as a function of the probe muon p_{T}, for probes within $0 < \eta < 1.5$.

4.6 Electron + PFm_{T} trigger efficiency

In general, the trigger efficiency is modeled by the HLT emulation present in the simulated samples. In order to account for differences in the HLT emulation with respect to the HLT deployed during the data taking, a data-driven correction to the Monte Carlo trigger efficiency is applied.

For the electron+PFm_{T} trigger, the efficiency depends on the electron leg of the trigger and the transverse mass cut. The two can be factorized, yielding

$$\epsilon_{\mathrm{HLT}} = \epsilon^{\mathrm{ele}}_{\mathrm{HLT}} \times \epsilon^{PFm_{\mathrm{T}}}_{\mathrm{HLT}}, \tag{4.9}$$

where ϵ_{HLT} is the total HLT efficiency, $\epsilon^{\mathrm{ele}}_{\mathrm{HLT}}$ is the electron leg HLT efficiency, addressed in Section 4.5 and $\epsilon^{\mathrm{PF}m_{\mathrm{T}}}_{\mathrm{HLT}}$ is the trigger transverse mass cut efficiency, whose computation is described here below.

4.6.1 HLT PFm_{T} efficiency

The online PFm_{T} efficiency $\epsilon^{\mathrm{PF}m_{\mathrm{T}}}_{\mathrm{HLT}}$ can be modeled as a function of the offline electron+missing energy transverse mass m_{T}. To measure this turn-on function, events surviving the offline selection and firing the electron part of the cross trigger are used as reference events. Then the HLT PFm_{T} efficiency is computed as the ratio, binned in offline m_{T} bins, of the events that pass the PFm_{T} cut and the total number of reference events. In formulae, for each i-th m_{T} bin:

$$\epsilon^{\mathrm{PF}m_{\mathrm{T}}}_{\mathrm{HLT}}\Big|_i = \frac{\#(\text{offline selected} \wedge \text{ref. HLT} \wedge \mathrm{PF}m_{\mathrm{T}}\ \text{cut})_i}{\#(\text{offline selected} \wedge \text{ref. HLT})_i}$$

The resulting histogram can be fitted with an erf-like function of the offline transverse mass, defined as

$$\mathcal{T}^{\mathrm{PF}m_{\mathrm{T}}}_{\mathrm{HLT}}(m_{\mathrm{T}}) = p_{\mathrm{low}} + \frac{p_{\mathrm{high}} - p_{\mathrm{low}}}{2} \cdot \left[1 + \mathrm{erf}\left(\frac{m_{\mathrm{T}} - \mu}{\sigma\sqrt{2}}\right)\right] \quad .$$

In the trigger turn-on function $\mathcal{T}$, the parameters p_{high} and p_{low} represent the higher and the lower trigger plateau efficiencies respectively, while the μ and σ parameters tell about the differences in the online and offline reconstruction of the transverse mass (in an ideal world, where HLT variables have the same value and resolution of offline ones, μ would coincide with the trigger PFm_{T} cut value and σ would be nearly 0, the turn-on curve collapsing into a step function).

The $\mathcal{T}(m_{\mathrm{T}})$ curve has been measured in data and evaluated in Monte Carlo, separately for each sub-period of the data taking in which the running conditions were stable, and the ratio of the two curves has been used as a correction factor for the simulated samples. The different periods taken into account are listed in Table 4.10.
The systematic uncertainty related to the precision of this measure is accounted for by calculating the 68% confidence belts of the scaling factors from the fit parameter errors. Figure 4.17 shows the measured turn-on curves in data and Monte Carlo, while the data/MC scaling factors to be applied on simulated samples, and the relative uncertainties, are shown in Figure 4.18.

Run2011A - Periods e-iii to e-iv	
Reference HLT	HLT_Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v4/5
Test HLT 1)	HLT_Ele25_WP80_PFMT40_v1
Test HLT 2)	HLT_Ele27_WP80_PFMT50_v1

Run2011A - Periods e-v to e-vi	
Reference HLT	HLT_Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v6/7
Test HLT	HLT_Ele32_WP70_PFMT50_v3/4

Run2011B - Period e-vii	
Reference HLT 1)	HLT_Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v7
Reference HLT 2)	HLT_Ele32_WP70_v2/3
Test HLT	HLT_Ele32_WP70_PFMT50_v4/8/9

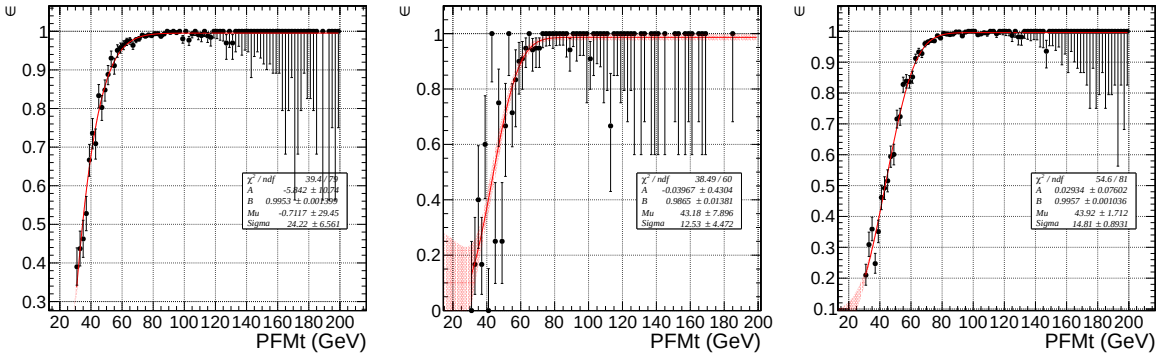
(a) Data

Monte Carlo 1)	
Reference HLT	HLT_Ele15_CaloIdT_CaloIsoVL_TrkIdT_TrkIsoVL_v2
Test HLT	HLT_Ele27_WP80_PFMT50_v4

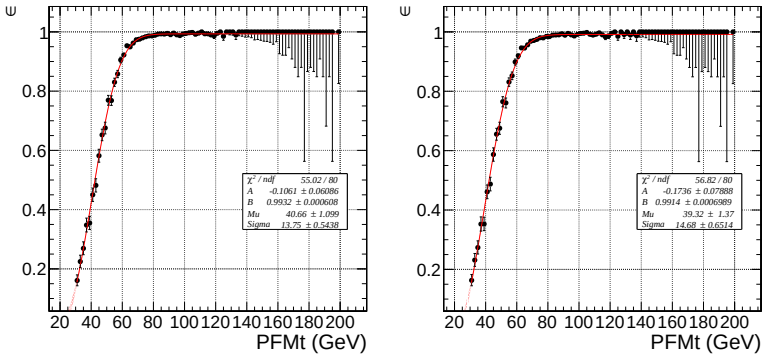
Monte Carlo 2)	
Reference HLT	HLT_Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v7
Test HLT	HLT_Ele32_WP70_PFMT50_v4

(b) Monte Carlo

Table 4.10: Reference and test trigger paths taken into account in the different periods for the PFm_{T} turn-on curve measurement in (a) data and (b) Monte Carlo.



(a) Data



(b) Monte Carlo

Figure 4.17: (a) Turn-on curves for data during `Run2011A` and `Run2011B` (shown from left to right in the same order of Table 4.10a). (b) Turn-on curves measured in Monte Carlo samples, corresponding to the two HLT paths of table 4.10b.

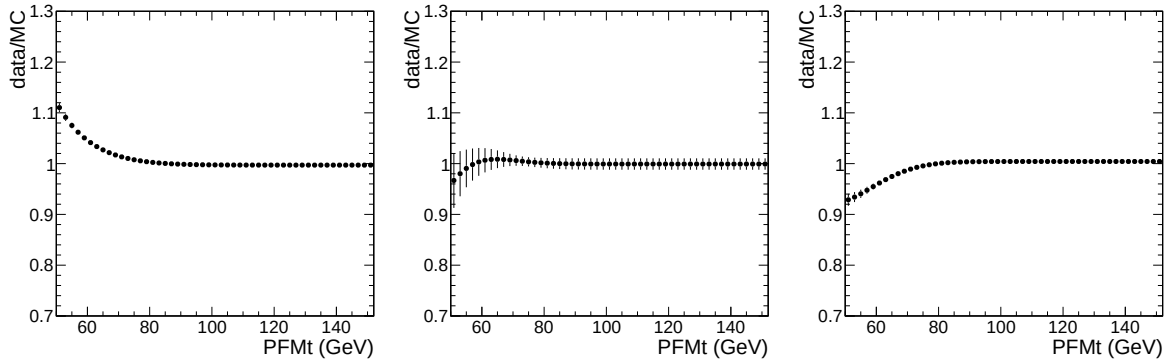


Figure 4.18: Data-driven scale factors to be applied on simulated samples. From left to right, the plots correspond to the periods reported in Table 4.10a.

4.7 Data - Monte Carlo comparisons

Following the event selection criteria described in Section 4.3 and using the Monte Carlo samples listed in Section 4.2, together with their NLO/NNLO cross-section, the most important kinematic variables in data and Monte Carlo are compared. All the various ingredients previously described enter this comparison. First of all, the event-by-event re-weighting to account for the different pile-up conditions in data and Monte Carlo, as explained in Section 4.2. In addition to that, simulated events are also assigned weighting factors that account for differences between data and Monte Carlo in the reconstruction, identification and trigger efficiencies, as addressed in Sections 4.5 and 4.6.
Before showing the table with event yields and the comparison plots, a data-driven approach is presented to estimate the contamination of QCD background in the analysis without having to rely on the – insufficient – Monte Carlo simulation.

4.7.1 The QCD estimation

Due to a huge production cross-section (and to finite computing resources), it is not possible to simulate a sufficient number of QCD multi-jet events surviving the selections of this analysis. To estimate its shape and contribution, a data-driven approach must be envisaged.

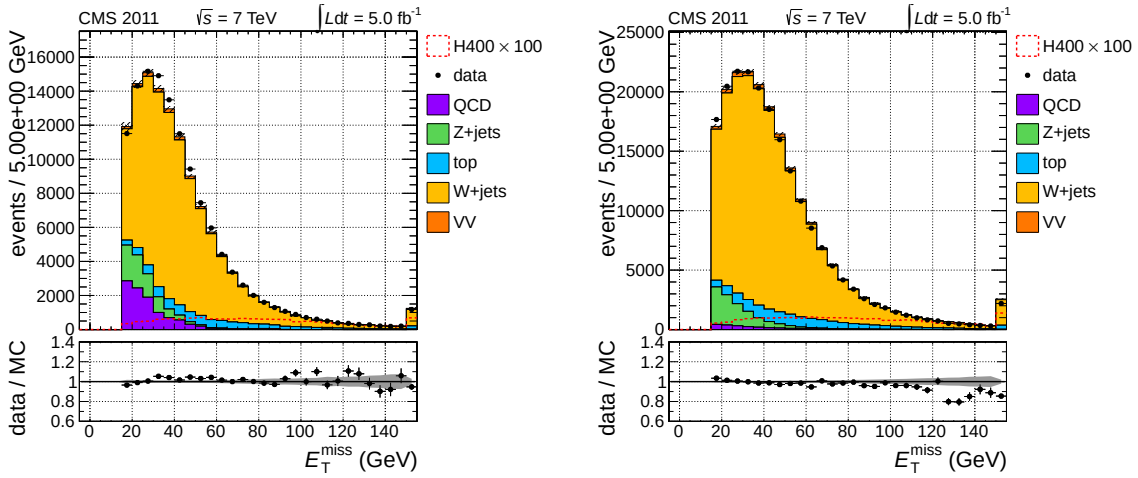
The lepton isolation variable provides a powerful tool to isolate in data a high-purity QCD sample: while in electroweak backgrounds leptons come from the decay of a weak bosons and are truly isolated, i.e. disconnected from hadronic activity, most leptons in multi-jet events are in fact jets being reconstructed as leptons in the detector, with plenty of hadronic activity in their surrounding[5].
Therefore, the same data entering the final signal phase space can be looked at, with the lepton cut on the tracker-based isolation variable reversed, namely $I_{\text{tk}} > 0.1(0.05)$ for electrons (muons), and the combined isolation selection released. To enrich the resulting sample in QCD events, the identification criteria for the $e\nu_e jj$ final state are relaxed to `WP95` too, and the $E_{\text{T}}^{\text{miss}}$ cut is lowered to 15 GeV in both channels. With these modifications, the data sample constitutes a good proxy for QCD events, upon verifying, by means of a Monte Carlo simulation, that the shape in the signal phase space of any interesting variable is preserved in the anti-selected sample.

Thanks to a good separation between QCD events, mostly at low $E_{\text{T}}^{\text{miss}}$, and W+jets events, exhibiting a Jacobian peak around 40 GeV, the missing transverse energy is a good candidate variable that can be used to fit the QCD shape template described above to the data, together with the Monte Carlo simulation of all other background shape and yields, in order to extract the QCD relative normalization. Results of such fits for the electron and muon channels are reported in Figure 4.19. On the full 2011 sample, the fraction of QCD events resulting from the fit amounts to $\sim 8\%$ for electron data and $\sim 1\%$ for muon data.

Although large uncertainties have to be assumed on these numbers, a thorough estimation of their systematic uncertainty has not been performed, since the QCD yield and shape are not actually used for any Physics result presented in this work (except for the data to Monte Carlo

[5] Real electrons or muons in jet events can be produced in semi-leptonic decays of b and c mesons contained in jets, but the same conclusions about their poor isolation hold.

comparison plots of this section). The purpose of this study is mainly to reassure the reader that no out-of-control contamination from multi-jet events is affecting the analysis.



(a) $e\nu_e$ jj final state (b) $\mu\nu_\mu$ jj final state

Figure 4.19: Result of the QCD template fit for (a) the electron and (b) the muon final state. The QCD shape is taken directly from data, as explained above, while its normalization is free in the fit. Other background have shape and normalization fixed to the Monte Carlo simulation. To ease the fit, the E_T^{miss} cut has been lowered to 15 GeV.

4.7.2 Yields

The impact of the selections described in the previous pages is reported in Table 4.11 for a couple of candidate signal samples. In this table, the event yields of the different production mechanisms (gluon fusion and weak vector boson fusion), as well as the different decay modes of the leptonic W boson (into electron or muon and tau) are detailed. Gluon fusion produced Higgs particles, decaying into electron or muon and a di-jet pair constitute the vast majority of the total signal, roughly the 90%. Although not optimized for a VBF-like signal, the analysis receives a not negligible contribution from the VBF production mechanism, its fraction at the end of the event selection being about the 5–10% of the total expectation. Since one good electron or muon is searched for, $\tau\nu_\tau$ jj final states are relevant only if the τ decays leptonically ($\mathcal{B}(\tau \to l\nu_l\nu_\tau) \sim 35\%$). In the end, these events constitute only a few percent of the total.

Yields of all background Monte Carlo samples and data are compared for the electron and the muon channel in Table 4.12 at the final-analysis level. On top of the selections described in Section 4.3, a further cut is applied on the reconstructed Higgs mass. Even though the full $m_{l\nu_l jj}$ spectrum is used to estimate the background and extract the limit on the signal cross-section, it makes more sense to show event yields in a region around the nominal Higgs mass, since the analysis result is mostly sensitive to the number of events lying in the surroundings of the resonance, with the sidebands of $m_{l\nu_l jj}$ actually being used to determine the global background normalization (more details are given in Section 4.10, where the signal extraction technique is thoroughly described).

cut	$m_H = 350\,GeV/c^2$				$m_H = 500\,GeV/c^2$				W+jets
	$l\nu_l\,q\bar{q}$		$\tau\nu_\tau\,q\bar{q}$		$l\nu_l\,q\bar{q}$		$\tau\nu_\tau\,q\bar{q}$		
	gg	VBF	gg	VBF	gg	VBF	gg	VBF	
pre-selection	4.66e+02	4.71e+01	2.25e+01	2.24e+00	1.56e+02	1.79e+01	8.24e+00	9.43e-01	4.39e+05
centrality cuts	4.05e+02	4.11e+01	1.96e+01	1.96e+00	1.38e+02	1.59e+01	7.25e+00	8.43e-01	3.06e+05
$\Delta\eta_{jj}$ cut	3.56e+02	2.75e+01	1.75e+01	1.33e+00	1.24e+02	1.17e+01	6.53e+00	6.39e-01	2.01e+05
m_{jj} cut	2.26e+02	1.59e+01	1.11e+01	7.55e-01	8.16e+01	7.75e+00	4.33e+00	4.34e-01	5.56e+04

Table 4.11: Expected event yields in $5.0\,fb^{-1}$ at different steps in the selection flow for a couple of signal points, divided in the different production and decay modes, and for the leading background W+jets. In the table, $l\nu_l\,q\bar{q}$ stands for Higgs samples generated with the leptonic W decaying into electron or muon, while $\tau\nu_\tau\,q\bar{q}$ indicates samples where the W decays into a $\tau\nu_\tau$ pair, and all decays of the τ lepton are considered. In the event counting, both the $e\nu_e\,jj$ and the $\mu\nu_\mu\,jj$ final states are added together.

$m_H\,(GeV/c^2)$	W+jets	Z+jets	top	diboson	QCD
250	4.09e+03/7.14e+03	1.44e+02/2.41e+02	4.33e+02/7.20e+02	1.68e+02/3.11e+02	1.11e+02/7.37e+01
300	2.64e+03/4.86e+03	6.87e+01/1.16e+02	3.48e+02/5.93e+02	1.24e+02/2.19e+02	9.75e+01/3.79e+01
350	1.78e+03/3.32e+03	3.84e+01/6.15e+01	2.86e+02/4.82e+02	1.01e+02/1.81e+02	8.36e+01/1.90e+01
400	1.22e+03/2.17e+03	2.13e+01/2.77e+01	2.07e+02/3.71e+02	7.93e+01/1.39e+02	9.75e+01/1.24e+01
450	9.01e+02/1.49e+03	1.06e+01/1.68e+01	1.50e+02/2.64e+02	6.27e+01/1.02e+02	9.75e+01/8.03e+00
500	6.27e+02/1.08e+03	7.99e+00/1.54e+01	1.09e+02/1.94e+02	4.69e+01/7.73e+01	8.36e+01/4.74e+00
550	4.64e+02/7.85e+02	4.89e+00/1.07e+01	7.14e+01/1.24e+02	3.78e+01/5.62e+01	5.57e+01/2.92e+00
600	3.21e+02/4.87e+02	2.02e+00/5.88e+00	4.48e+01/7.71e+01	2.69e+01/4.11e+01	1.39e+01/1.82e+00

$m_H\,(GeV/c^2)$	signal	$\sum$ bkg	data
250	5.29e+01/9.30e+01	4.95e+03/8.49e+03	5.11e+03/8.74e+03
300	5.68e+01/9.33e+01	3.28e+03/5.83e+03	3.46e+03/5.79e+03
350	6.80e+01/1.13e+02	2.29e+03/4.06e+03	2.59e+03/4.11e+03
400	5.89e+01/9.23e+01	1.63e+03/2.72e+03	1.88e+03/2.89e+03
450	3.97e+01/6.42e+01	1.22e+03/1.88e+03	1.29e+03/2.04e+03
500	2.50e+01/4.18e+01	8.74e+02/1.37e+03	9.22e+02/1.38e+03
550	1.57e+01/2.55e+01	6.34e+02/9.78e+02	6.29e+02/9.31e+02
600	8.76e+00/1.43e+01	4.09e+02/6.12e+02	3.94e+02/5.54e+02

Table 4.12: Expected yields of events from Monte Carlo simulation in $5.0\,fb^{-1}$ for the various background sources and for the signal. Each entry of the table shows the expected number of events for the $e\nu_e\,jj/\mu\nu_\mu\,jj$ final states. The yields are calculated at the end of the event selection (see Section 4.3), and further applying, for each considered Higgs mass point, a cut on the kinematic-fitted four-body invariant mass, using the mass windows defined in Table 4.7. In the last two columns, the sum of all background contributions as well as the number of events seen in data are reported.

4.7.3 Plots

To assess the quality of the Monte Carlo modelling of the data in this signature, comparisons in the full m_{jj} range between the simulated shape of different variables and the data are here presented. All other cuts of the selection flow are instead applied. We show lepton-related variables in Figure 4.20, E_T^{miss} related ones in Figure 4.21, and jet distributions in Figure 4.22.

The distributions of the di-jet system are shown in Figure 4.23, while the four-body invariant mass spectrum is presented in Figure 4.24.

For all plots, the following considerations hold:

- plots on the left are for the $e\nu_e jj$ final state, plots on the right for the $\mu\nu_\mu jj$ one;
- the stack of Monte Carlo samples is normalized to the integrated 2011 luminosity ($5.0\,fb^{-1}$).

The fair agreement observed in the bulk of all the distributions renders us confident that the signal hypothesis can be described through its simulation. 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.
A large discrepancy is observed in the distribution of ϕ variable of the the missing transverse energy. The origin of such effect is known, and resides in an imperfect modelization in the simulation of the HCAL detector noise, which is not symmetric along the ϕ coordinate. Nevertheless, it has been verified that no other variables are affected by this effect.

Lepton variables

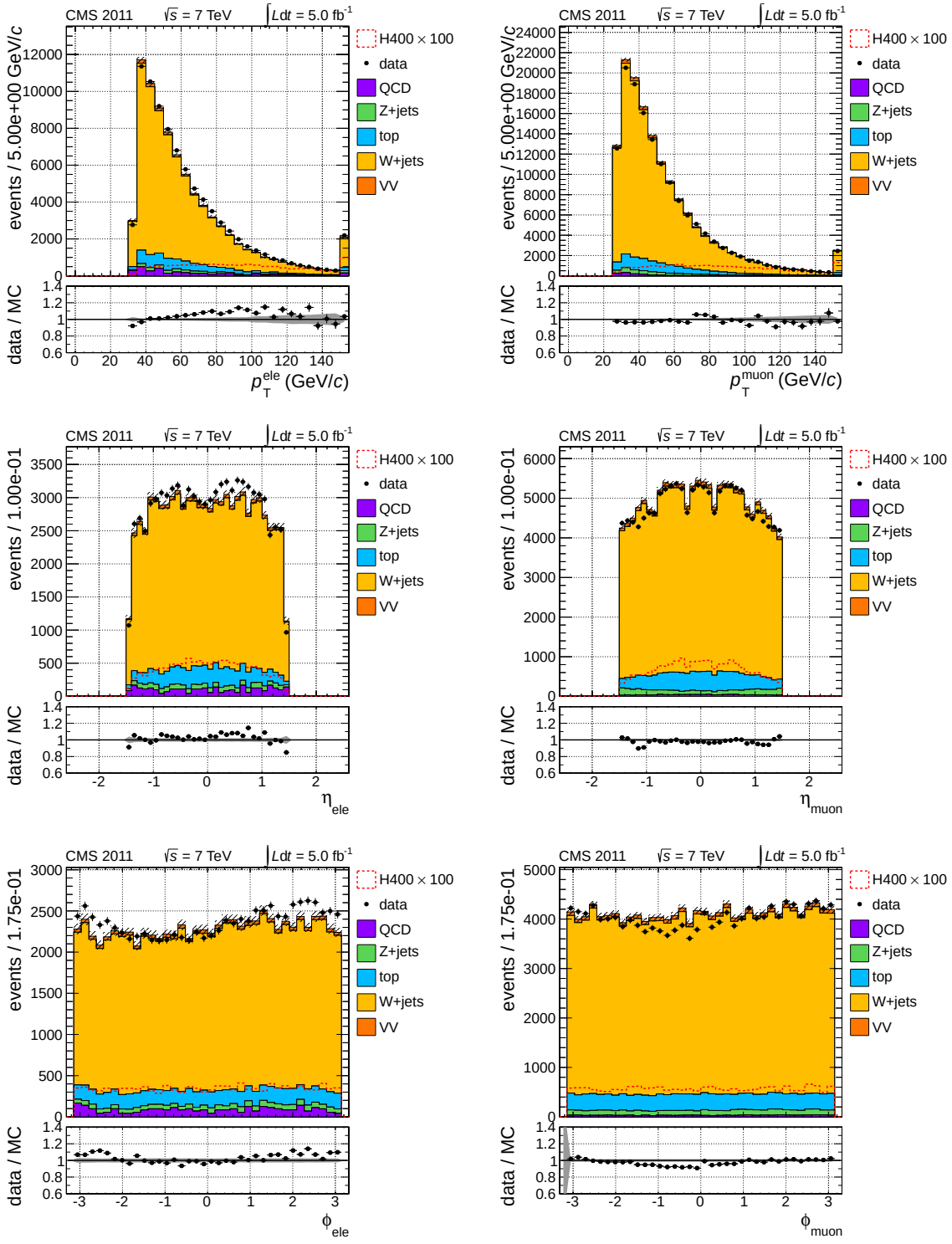


Figure 4.20: Lepton-related variables. From top to bottom: lepton p_{T}, η, ϕ.

E_T^{miss} variables

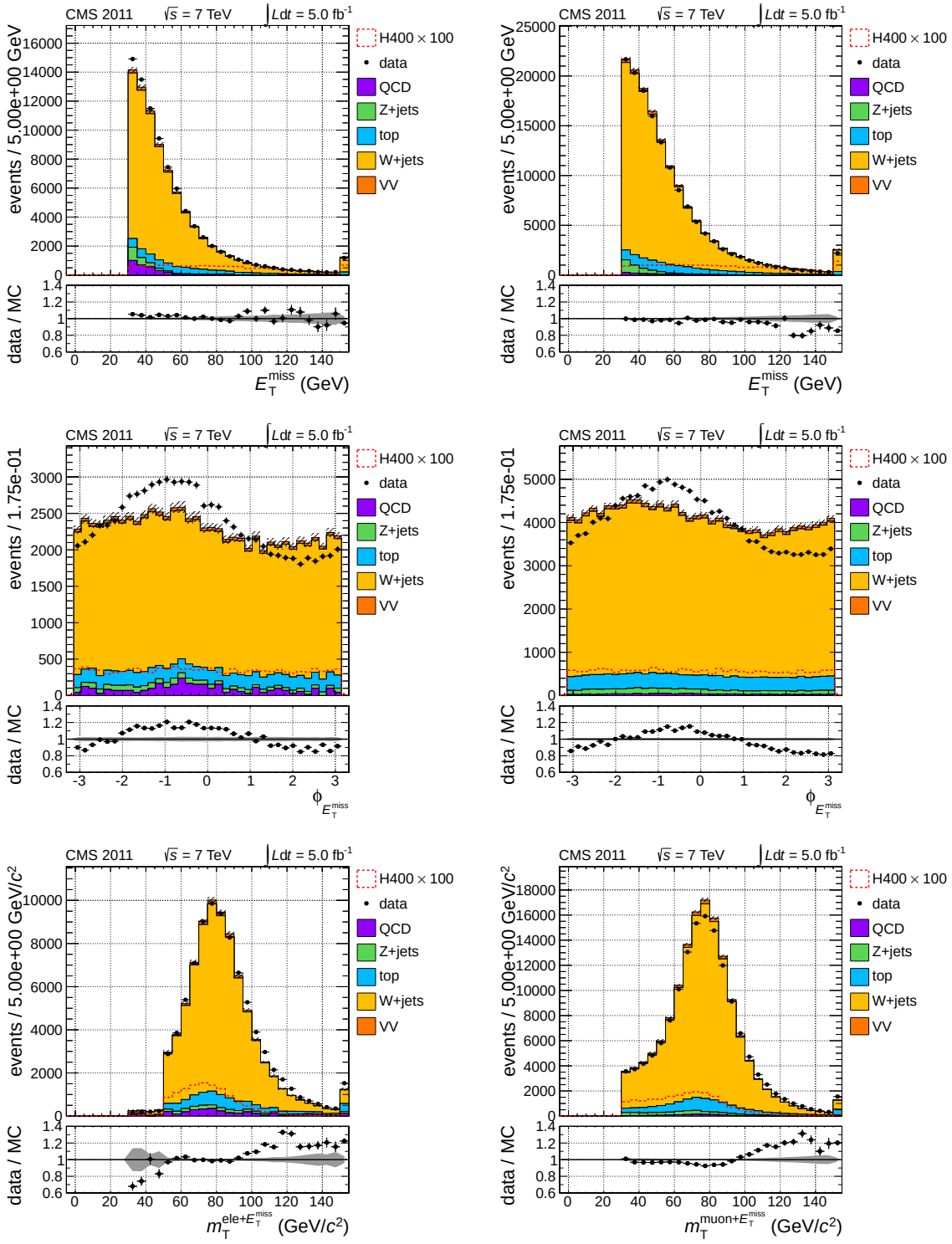


Figure 4.21: E_T^{miss}-related variable. From top to bottom: E_T^{miss}, E_T^{miss} ϕ, and lepton + E_T^{miss} m_T.

Jet variables

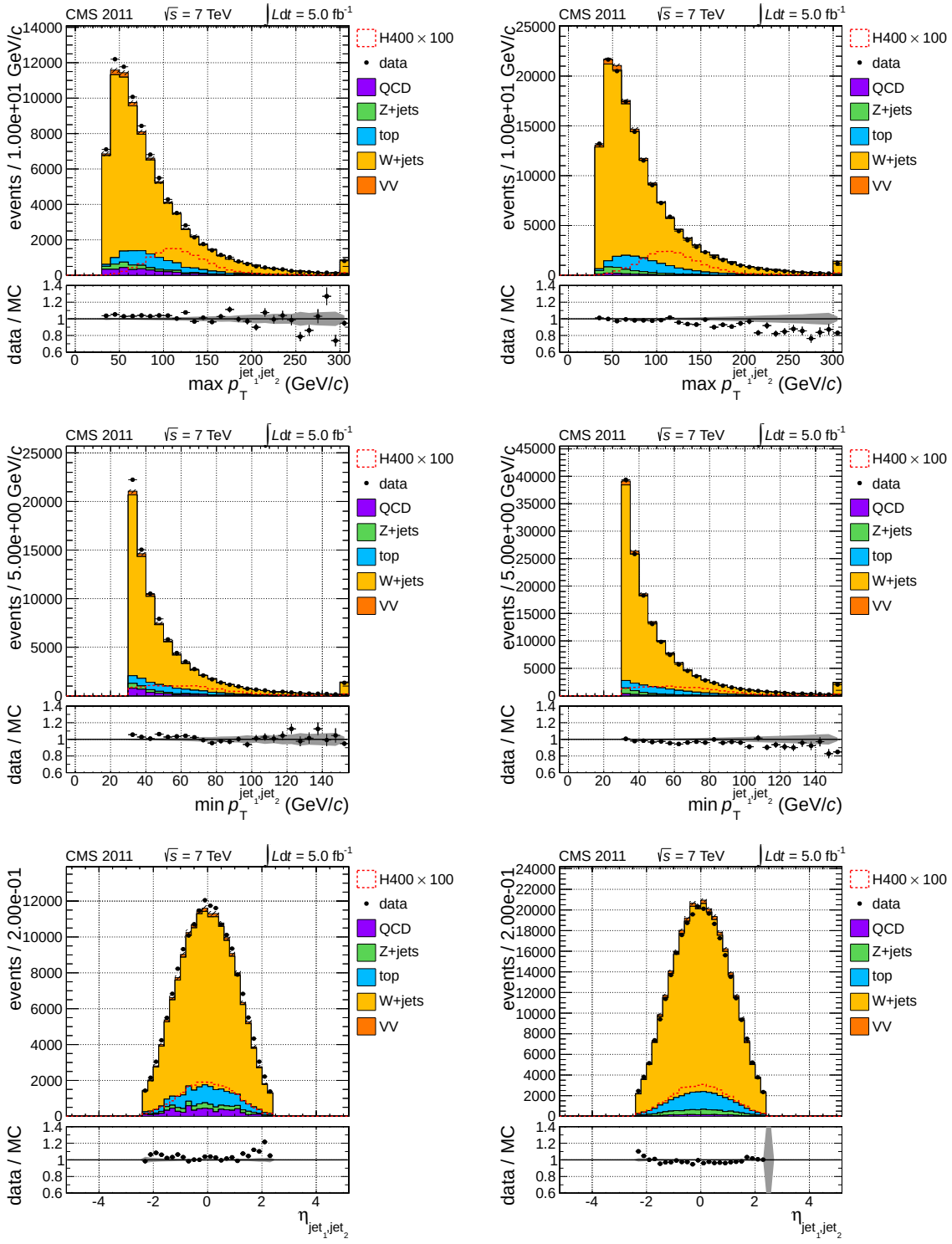


Figure 4.22: Jet-related variables. From top to bottom: leading jet p_T, trailing jet p_T, and jet η distribution.

Di-jet system variables

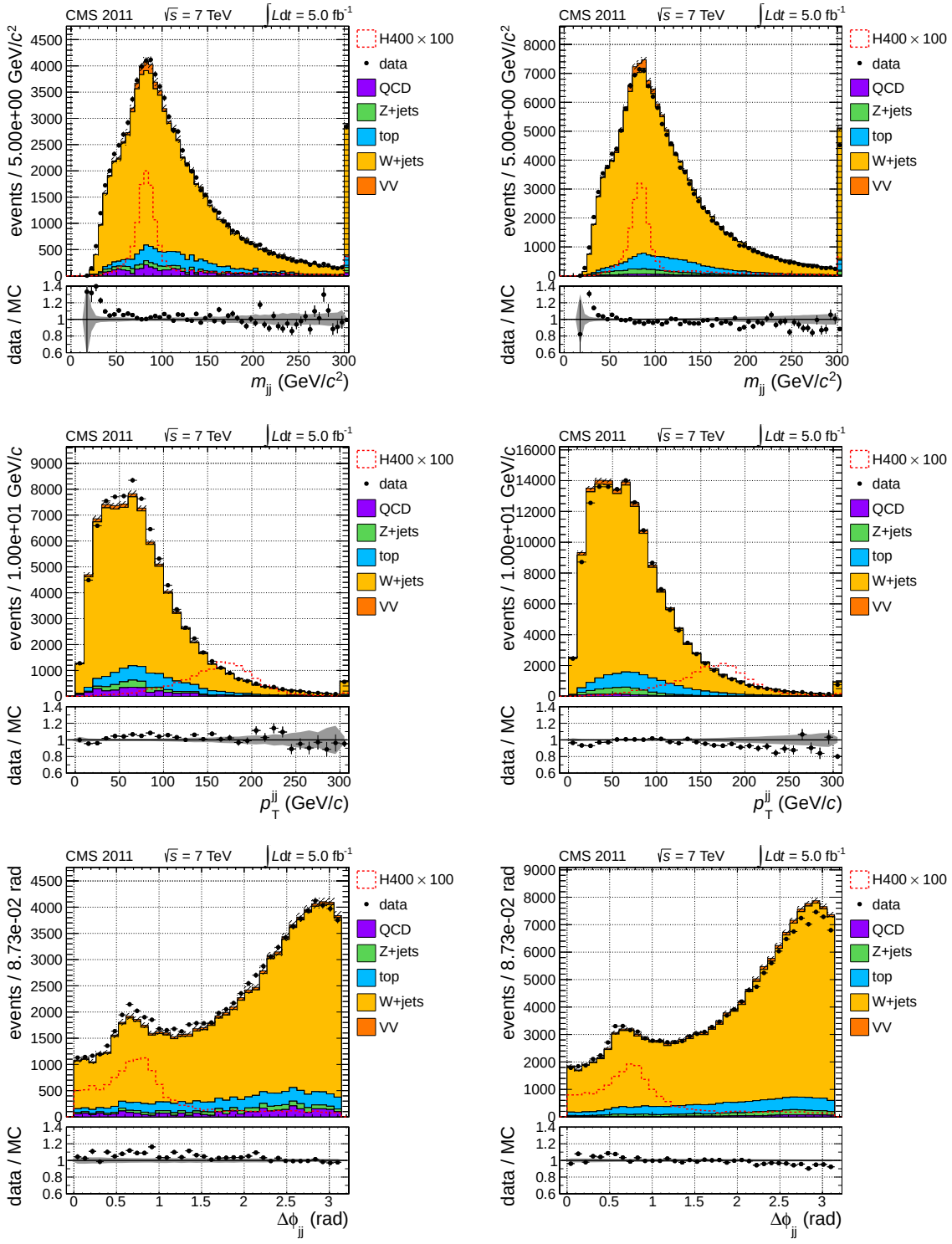


Figure 4.23: Di-jet system-related variables. From top to bottom: m_{jj}, di-jet p_T, and $\Delta\phi$ di-jet separation in the transverse plane.

Higgs system variables

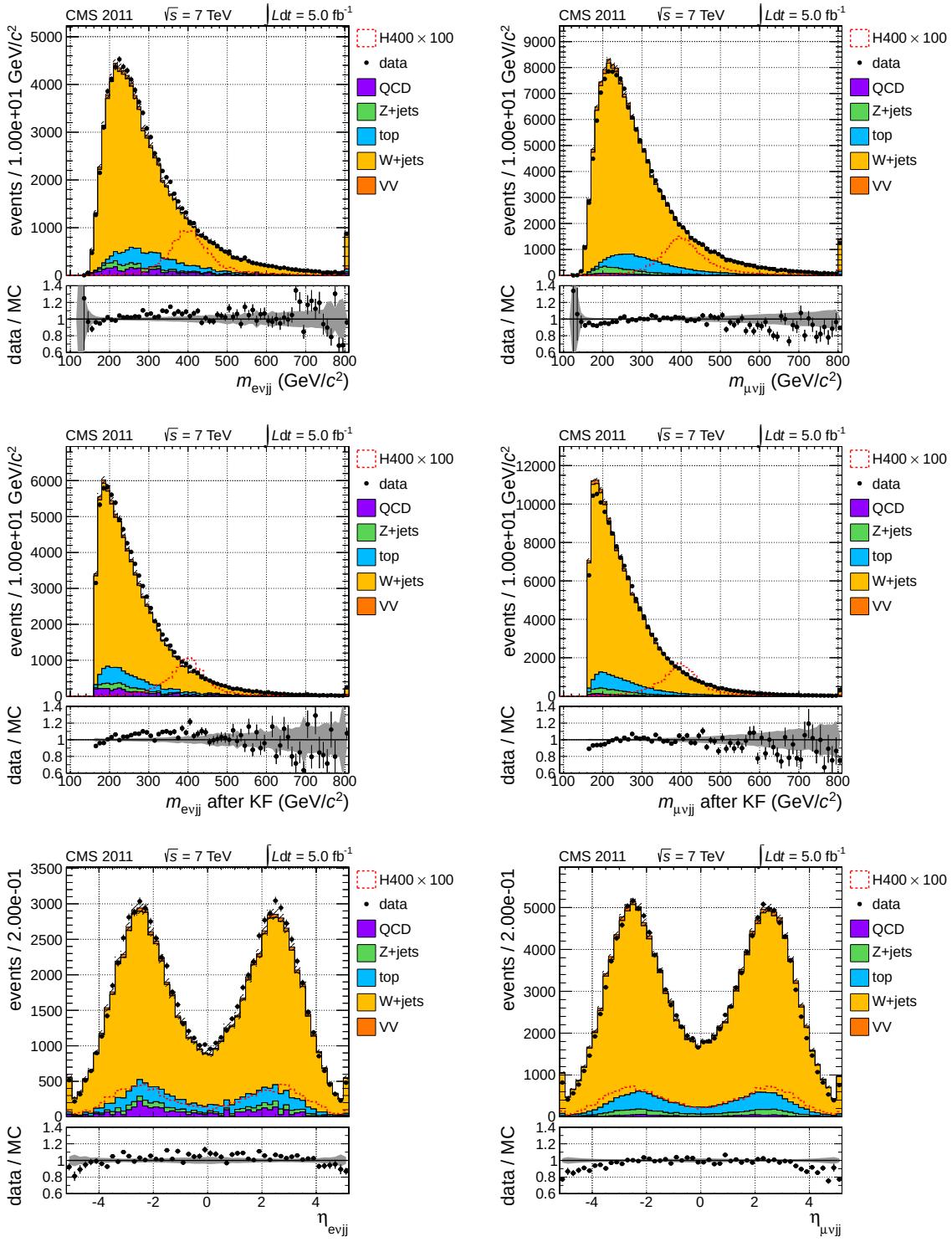


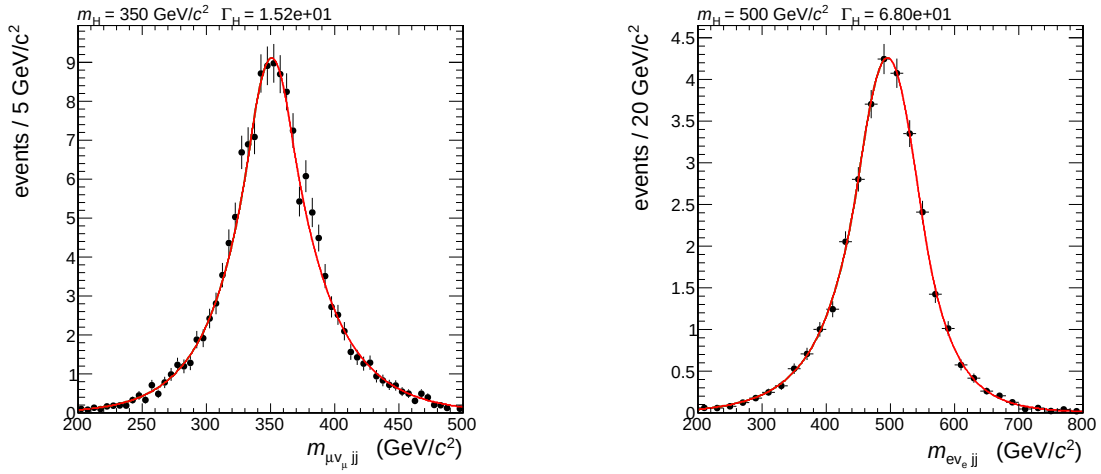
Figure 4.24: Higgs system-related variables. From top to bottom: plain $m_{l\nu_l jj}$, kinematic fitted $m_{l\nu_l jj}$, and Higgs reconstructed η.

4.8 Background and signal models

Previous sections have been devoted to the description of the selections that allow to retain events with a lepton, missing transverse energy and a couple of jets in the final state, compatible with Higgs boson production and decay into $l\nu_l q\bar{q}$ channel. A sizable number of events of an hypothetical SM Higgs signal (a few tens to a few hundreds, depending on the Higgs mass) would survive the cut flow, although accompanied by a large background, mostly dominated by W+jets production (see Table 4.11). The possibility to fully reconstruct the parent invariant mass from the four final state objects is a major feature of this channel, and it allows to configure the analysis as the search for a peak appearance above a steeply falling background continuum. In order to be able and draw any conclusion on the Higgs existence in this channel, two ingredients are then necessary: the signal expectations and a solid and precise estimation of the shape and the level of background events observed in data.

4.8.1 The signals

The expected number of signal events at the end of the analysis is estimated counting the number of simulated Higgs events that survive the event selection and properly multiplying this number by the predicted Higgs effective cross-section for this particular decay channel. Weighting factors to account for the different data to Monte Carlo pile-up condition, reconstruction, identification and trigger efficiencies and for the NNLO to NLO Higgs p_T spectrum are taken into account in the yield computation.



(a) $m_H = 350\,\text{GeV}/c^2 - \mu\nu_\mu$ jj final state (b) $m_H = 500\,\text{GeV}/c^2 - e\nu_e$ jj final state

Figure 4.25: Higgs boson invariant mass shape, as obtained from the simulation, with all the proper weighting factors taken into account. The number of events in the two histograms correspond to the expected signal yield in 5.0 fb^{-1} at the end of the selection chain. The spectrum is fitted with the function described in the paragraph above. This smooth function is used as signal input for the the limit extraction. A Higgs mass of 350 GeV/c^2 (500 GeV/c^2) is assumed in the left (right) plot.

The signal $m_{l\nu_l jj}$ shape is described with a fit on the Monte Carlo distributions. The fit is performed independently for each mass hypothesis and for each systematic effect that involves a shape change. Moreover, the gluon fusion and the vector boson fusion production mechanisms are considered as distinct signal channels, since they are affected by significantly different sys-

tematic uncertainties. As a consequence, each of them has its own independent mass fit.
Empirically, a double Crystal-Ball function (i.e. a Gaussian core with power law tails on both sides) has been found to adequately parametrize the signal shapes. Moreover, a Fermi turn-on is used to model the kinematic-fit threshold at 160 GeV/c^2. Examples of such fits can be found in Figure 4.25, for Higgs mass hypotheses of 350 and 500 GeV/c^2.

4.8.2 The background fit

A fully Monte Carlo-based background model presents several difficulties. First of all, the total generated events of the leading background source amount to an equivalent luminosity of about 2.5 fb^{-1} (see Table 4.4b), that is about one half of the collected data. The intrinsic statistical uncertainty of this method would then be larger than the fluctuations observed in data themselves. Moreover, it has been shown in Section 4.7 that the Monte Carlo simulation is capable of describing data collected in 2011 at a good, albeit not perfect, level. In particular, the shape of the four-body invariant mass presents, around the kinematical threshold of $m_{\mathrm{l}\nu_{\mathrm{l}}\,\mathrm{jj}} \sim 160\,\mathrm{GeV}/c^2$, a broader turn-on in data than in Monte Carlo, and also the falling part of the distribution is not well reproduced by the simulation. Given the large amount of background, even a few percent error in its estimate is enough to swamp a true signal or produce fake one. It is clear that such a technique is not viable.

A different, fully data-driven approach has been pursued, based on an analytic description of the $m_{\mathrm{l}\nu_{\mathrm{l}}\,\mathrm{jj}}$ spectrum. A suitable functional form has been identified, deriving the best-parameter values from a fit on data itself. In the following, studies performed to choose the optimal function and to evaluate the associated systematic uncertainty are reported.

After the event selection is applied to data, four distinct invariant mass spectra can be identified, one for each lepton flavour (electron or muon) times one for the two- or three-jet final states. The shapes of these four distinct cases as seen in the simulation are shown in Figure 4.26a. It can be noticed how the different kinematic selections in the electron and muon channels reflect into a different four-body shape, especially at low invariant mass values. To improve the fit stability, and since electrons and muons have in general different reconstruction systematics associated, it proves convenient to consider the two lepton flavours as distinct channels, with independent background models. The two- and three-jet bins, on the contrary, are not separated into distinct channels. Although their shape is considerably different, the three-jet case is about eight times less populated than the two-jet final state, and its background fit would be affected by a large statistical uncertainty if considered alone.
To summarize, since the event selection adopted does not depend on the Higgs mass hypothesis, one single background fit for each lepton flavour is sufficient for all Higgs searches from 250 to 600 GeV. All that changes is the signal expectation and shape, as addressed in the next paragraph. These two background shapes, as expected from simulation, are shown in Figure 4.26b.

Choice of the function

To avoid any source of bias from the usage of real data in the signal phase space, the choice of the appropriate analytic function to model the background shape is made on the basis of compatibility with the simulation.

First of all, a set of candidate functional families is defined. The functional forms are intended

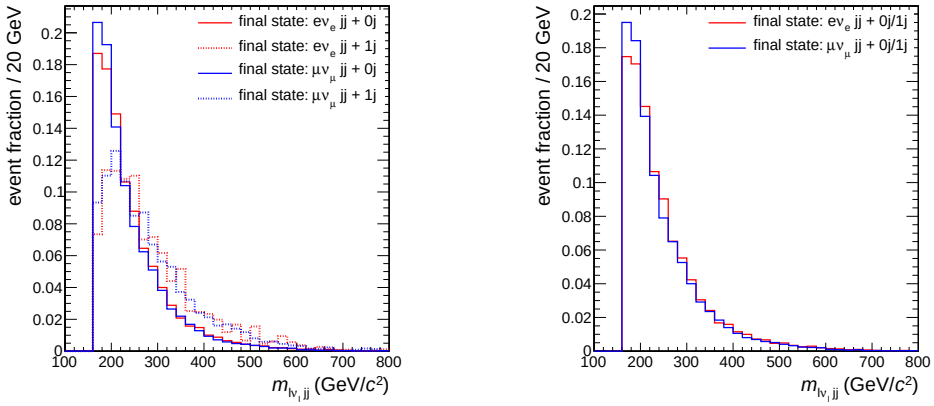


Figure 4.26: Background $m_{l\nu_l jj}$ shapes from Monte Carlo simulation are compared: (a) electron and muon final states, as well as the two- and three-jet cases considered independently; (b) two- and three-jet bins merged together. The tighter kinematic selections of $e\nu_e jj$ events reflect into a harder $m_{l\nu_l jj}$ spectrum.

to span all the possibilities of the true distribution, i.e. the distribution to which signal-free data would tend if if we disposed of a nearly infinite statistics. The set is defined as follows:

$$\begin{aligned}
&\bullet \text{ polynomials:} && \mathcal{T}(x;\mu,kT)\times N\cdot\left(1+\sum_{i=1}^{m}a_i x^i\right)\\
&\bullet \text{ exponentiated polynomials:} && \mathcal{T}(x;\mu,kT)\times N\cdot\exp\left(\sum_{i=1}^{m}a_i x^i\right)\\
&\bullet \text{ exponentials:} && \mathcal{T}(x;\mu,kT)\times\sum_{i=1}^{m}N_i\cdot e^{-\lambda_i x}\\
&\bullet \text{ power laws:} && \mathcal{T}(x;\mu,kT)\times\sum_{i=1}^{m}N_i\cdot\frac{1}{(x-a_i)^{n_i}}
\end{aligned} \quad , \tag{4.10}$$

where we have identified $x \equiv m_{l\nu_l jj}$ and all the parameter values are set through the fit. All functions are multiplied by a Fermi function intended to model the low invariant mass turn-on: $\mathcal{T}(x;\mu,kT) = 1/[e^{-(x-\mu)/kT}+1]$.
For each family, a single candidate is chosen by increasing the functional order until the extra degrees of freedom provided by the higher order do not bring in a significant improvement in the description of the Monte Carlo distribution, that is until the goodness of the fit expressed, in terms of reduced chi-square, does not significantly increase. In Figure 4.29 and 4.30, the result of such fits for the different function families in the electron and muon channels respectively are shown, while Table 4.13 summarizes the results.

The resulting set of candidates is composed by a five-order polynomial, a third-order exponentiated polynomial, a second-order exponential and a first-order power law[6]. From these studies, the latter emerges as the best candidate, given its ability to well reproduce the back-

[6]In a couple of cases, the highest order (N) function taken into account shows a not-yet saturated $\tilde{\chi}^2$. It has anyway been decided to stop the function scan at the N-1 level upon verifying that the $\tilde{\chi}^2$ improvements are dominated by high-mass tails fluctuations, where the MC statistics is largely insufficient.

family	**polynomials**		
order/d.o.f.	$4^{th}/7$	$5^{th}/8$	$6^{th}/9$
$e\nu_e jj$	$\tilde{\chi}^2 = 2.21$	$\tilde{\chi}^2 = 0.73$	$\tilde{\chi}^2 = 0.72$
$\mu\nu_\mu jj$	$\tilde{\chi}^2 = 6.16$	$\tilde{\chi}^2 = 1.37$	$\tilde{\chi}^2 = 1.06$

family	**exp. polynomials**		
order/d.o.f.	$2^{nd}/5$	$3^{rd}/6$	$4^{th}/7$
$e\nu_e jj$	$\tilde{\chi}^2 = 1.08$	$\tilde{\chi}^2 = 0.79$	$\tilde{\chi}^2 = 0.70$
$\mu\nu_\mu jj$	$\tilde{\chi}^2 = 1.34$	$\tilde{\chi}^2 = 1.11$	$\tilde{\chi}^2 = 1.19$

family	**exponentials**		
order/d.o.f.	$1^{st}/4$	$2^{nd}/6$	$3^{rd}/8$
$e\nu_e jj$	$\tilde{\chi}^2 = 1.06$	$\tilde{\chi}^2 = 0.68$	$\tilde{\chi}^2 = 0.70$
$\mu\nu_\mu jj$	$\tilde{\chi}^2 = 1.32$	$\tilde{\chi}^2 = 1.09$	$\tilde{\chi}^2 = 1.06$

family	**power laws**	
order/d.o.f.	$1^{st}/5$	$2^{nd}/8$
$e\nu_e jj$	$\tilde{\chi}^2 = 0.73$	$\tilde{\chi}^2 = 0.73$
$\mu\nu_\mu jj$	$\tilde{\chi}^2 = 1.03$	$\tilde{\chi}^2 = 1.07$

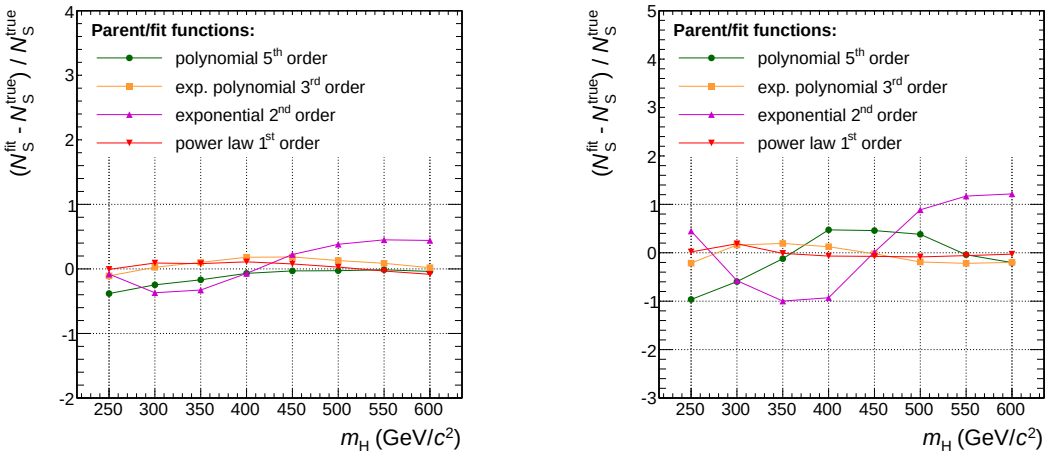
Table 4.13: Evolution of the reduced chi-square of the Monte Carlo background fits increasing the fit function degrees of freedom (labelled d.o.f. in the table) for the four different function species. A saturation of the $\tilde{\chi}^2$ indicates that no extra free parameters are needed to improve the background shape description.

ground shape with a limited number of free parameters (three, plus two for the turn-on). On the contrary, all other functions need at least one extra degree of freedom to achieve a similar performance.

The Physics reach of this analysis depends on the ability of the chosen background model to distinguish between a background-only or a signal-plus-background scenario. Given the little signal yield and the considerable width of the Higgs resonance, especially at high mass, some care must be put in the choice of the fit function to ensure that a potential resonance can not be accommodated in a background-only model with a different parameterization of the background shape. Therefore, a study is made to test the bias introduced by the aforementioned functions in the signal yield evaluation.
For each function, several toy datasets are generated using the function itself as the truth shape for the background and superimposing signal events for all Higgs masses. The dataset background population equals the number of events observed in data in the electron/muon channels, while the number of signal events is amplified by a variable factor depending on the Higgs mass[7]. A composite pdf, made by the sum of the function itself and the expected signal shape is then used to fit each toy dataset. All the function parameters, as well as the total number of background and signal events, are determined in the fit. A graphical representation of the result is shown in Figure 4.27: at each different mass hypothesis, the difference between the fitted signal averaged over many toys $\langle N_S^{\text{fit}} \rangle$ and the true number of signal events in each dataset N_S^{true} is shown for all the selected functions, relative to N_S^{true} itself. In general, when functions with a higher number of degrees of freedom are used as background model, there is a larger tendency to underestimate or overestimate the correct signal yield, as one could naively expect.
The constant small error associated with the power law confirms this as the chosen parametric shape for the background.

Up to this point, only simulated events have been used, and the function has not been tested on

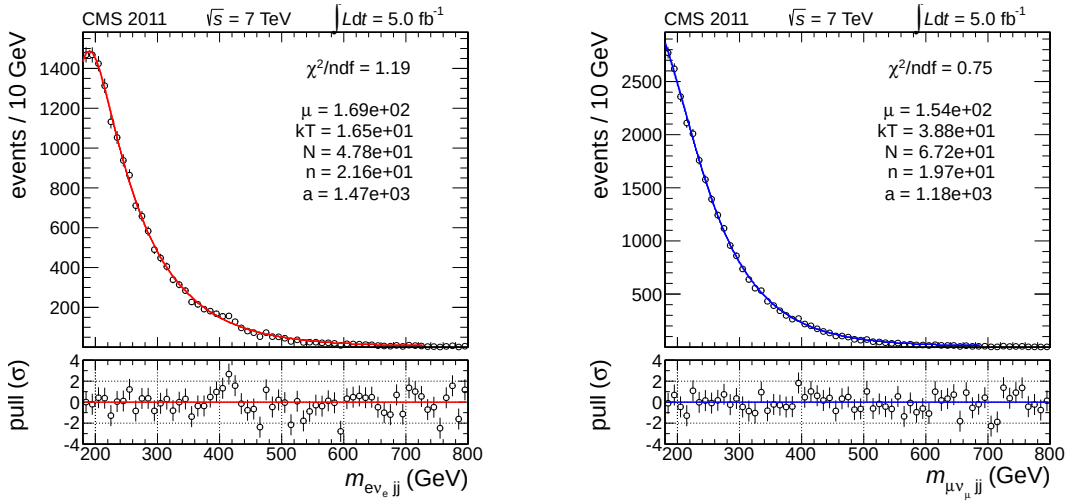
[7]This is done in order to have roughly the same sensitivity at all Higgs mass points. The amplification factor comes from the result of the analysis in this channel, and corresponds to the limit on the SM cross-section, discussed in Section 4.10.



(a) eν$_e$ jj final state (b) μν$_\mu$ jj final state

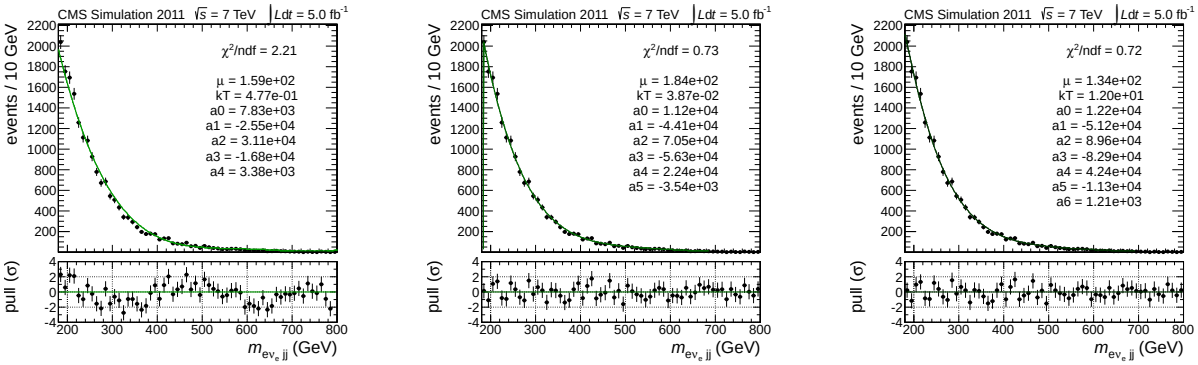
Figure 4.27: Toy datasets with signal events superimposed on background events have been generated and the four-body invariant mass distribution fitted with a signal-plus-background pdf. The plot shows, for different background shape hypotheses, the average difference between the number of signal events estimated by the fit and the true number of signal events present in the dataset, normalized to the latter.

real data. To gain confidence with the background fit, still without disclosing the phase space where the Higgs signal is expected, events falling in the m_{jj} sidebands, defined in Section 4.3 as $55 < m_{jj} < 65\,\text{GeV}/c^2$ and $95 < m_{jj} < 115\,\text{GeV}/c^2$, can be used to test the fit function on real data, even if within a different kinematic regime of the di-jet system and, as a consequence, with a different $m_{l\nu_l\,jj}$ shape than the final analysis selection. The result of the fit in the sidebands is shown in Figure 4.28. The very good description of observed data is expressed by the fit reduced chi-squares, close to unity in both the electron and muon channels.

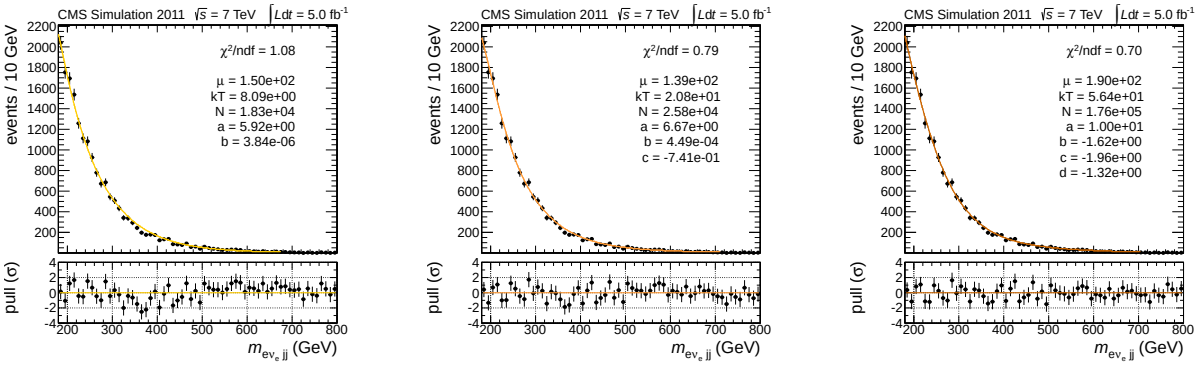


(a) eν$_e$ jj final state in the sideband region (b) μν$_\mu$ jj final state in the sideband region

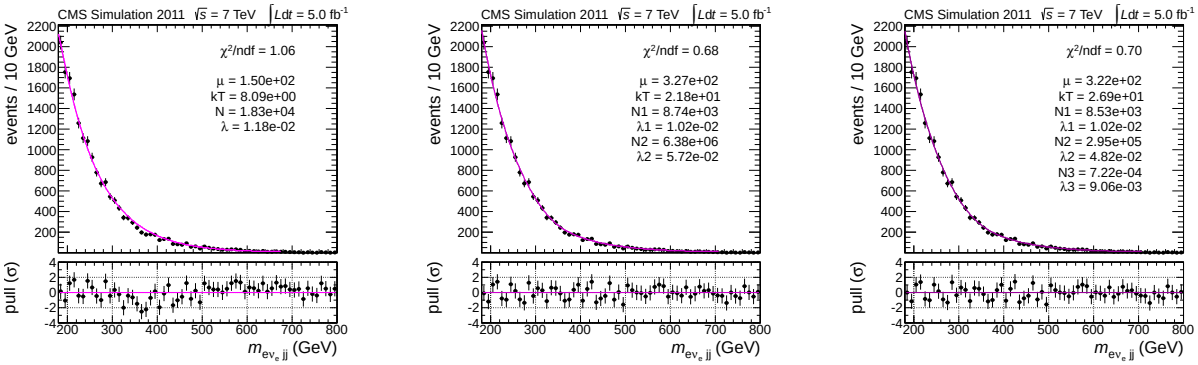
Figure 4.28: Background-only fits of the four-body invariant mass distribution in data for the m_{jj} sideband region. A first-order power law function multiplied by a Fermi turn-on is used as parametric shape.



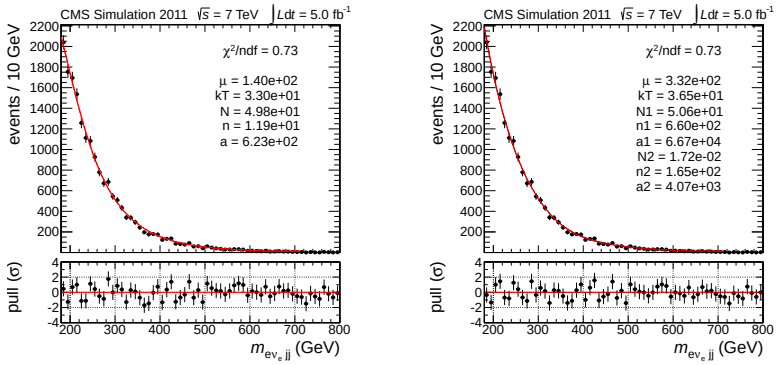
(a) Polynomials



(b) Exponentiated polynomials

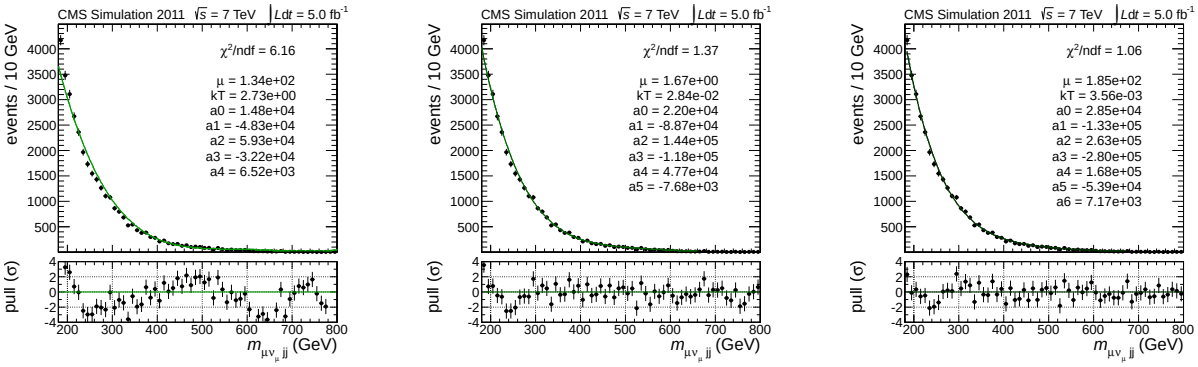


(c) Exponentials

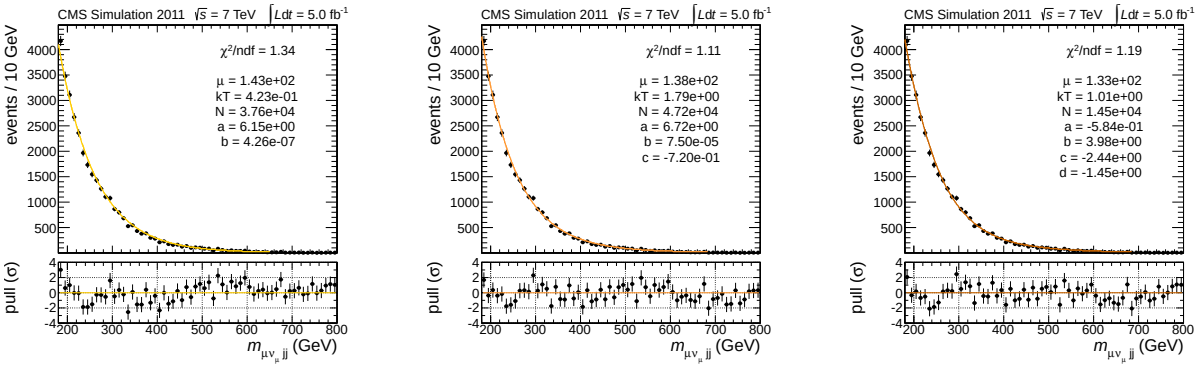


(d) Power laws

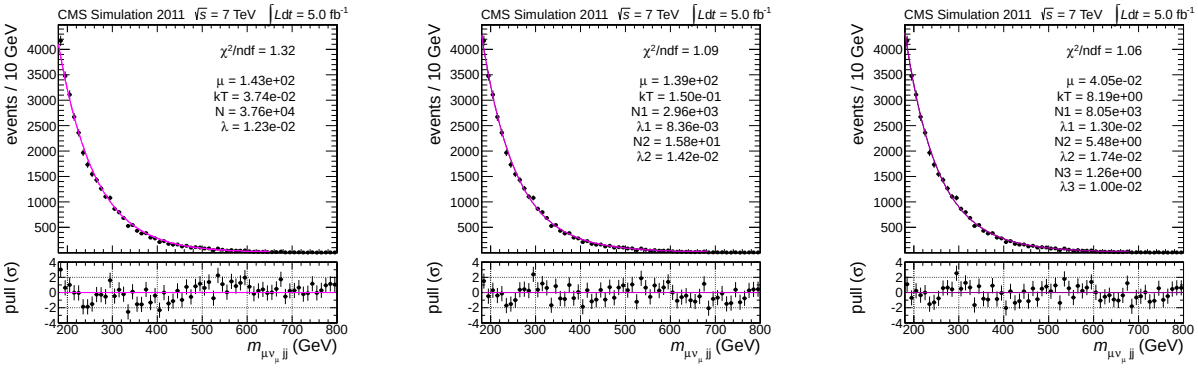
Figure 4.29: Fits of simulated $m_{l\nu_l jj}$ distribution with different functional forms of increasing order from left to right for the electron final state.



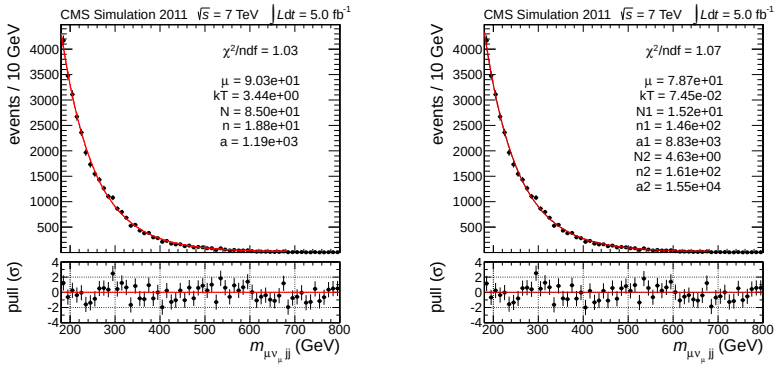
(a) Polynomials



(b) Exponentiated polynomials



(c) Exponentials



(d) Power laws

Figure 4.30: Fits of simulated $m_{l\nu_l jj}$ distribution with different functional forms of increasing order from left to right for the muon final state.

Bias study

It has been shown in the previous paragraph that a first-order power law as a fit function allows to make an unbiased estimate of the amount of signal events that appear on top of a background continuum with the same shape. Since the true shape of the $m_{\mathrm{l}\nu_{\mathrm{l}}\,\mathrm{jj}}$ spectrum in data is not known *a priori*, a further study is made to estimate the possible bias introduced by the choice of a specific function on the background evaluation. The procedure is as follows:

- one among the candidate functions above defined is chosen and taken as truth model $T(x)$;
- a toy dataset is generated using $T(x)$ as parent distribution and corresponding to a very large amount of accumulated statistics[8]. The parameters of $T(x)$ used to generate the pseudo-data are determined with a fit on real data[9];
- the toy dataset is fitted with all the candidate functions $F(x)$ from the list above. For each of them, and for each Higgs mass window $[x_{\mathrm{low}}, x_{\mathrm{high}}]$, we compute $\Delta_{FT} = N_B^{\mathrm{fit}} - N_B^{\mathrm{true}}$, where $N_B^{\mathrm{fit}} = \int_{x_{\mathrm{low}}}^{x_{\mathrm{high}}} F(x)\mathrm{d}x$ and $N_B^{\mathrm{true}} = \int_{x_{\mathrm{low}}}^{x_{\mathrm{high}}} T(x)\mathrm{d}x$. Higgs mass windows are defined in Table 4.7;
- the procedure is iterated for all $T(x)$.

In the end, for each combination of a test fit function $F_i(x)$ and a parent function $T_j(x)$, a value of Δ_{ij} is obtained for each one of the 8 different Higgs mass hypotheses. The distribution of Δ_{ij} obtained when varying the parent function $T(x)$ for a given choice of the fit function $F(x)$, can be taken as a a quantitative indication of the bias introduced on the background evaluation by choosing $F(x)$ as the background model. The bias Δ_{ij} can be translated into a systematic error on the signal yield estimation dividing it by the number of signal events expected, for each mass point, in the same mass window $[x_{\mathrm{low}}, x_{\mathrm{high}}]$, in formulae $\delta_{ij} = \Delta_{ij}/N_S^{\mathrm{true}}$. Plots in Figure 4.31 show the bias δ_{ij} relative to the signal yield when the first-order power law is taken as fit function.

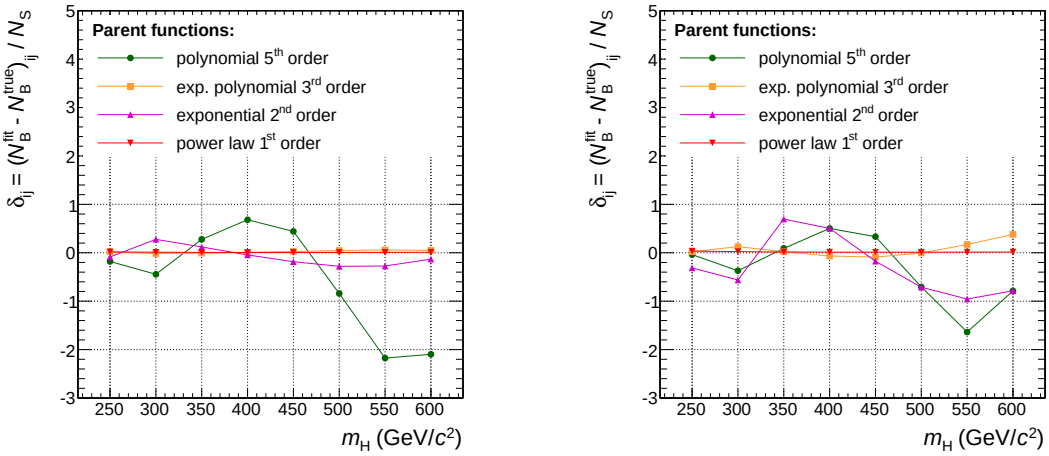
This study allows to define a systematic uncertainty on the expected number of signal events, as addressed in Section 4.9.2.

4.9 Evaluation of systematics

In this Section, the possible sources of systematic errors associated with this measurement are discussed. In particular, the signal yield and shape estimation can be affected by theoretical (production cross-section, choice of PDF and their acceptances, Higgs boson width modeling at High mass) as well as instrumental uncertainties (LHC luminosity, pile-up model in the Monte Carlo, lepton energy scale, jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ energy scale, lepton reconstruction and trigger efficiencies, b-tag).
Given that the background is extracted directly from data, all systematic effects arising from the imperfect knowledge of the experimental apparatus and LHC running conditions do not

[8] This is done to reduce the impact of statistical fluctuations in the determination of the bias and to avoid a double counting of the statistical uncertainty of the background model in the limit extraction

[9] Data is preferred to the Monte Carlo for the determination of the parameters of the parent functions since the Monte Carlo statistics is significantly lower than accumulated data, and this can lead to an artificially larger bias estimation.



(a) eν$_e$ jj final state (b) μν$_\mu$ jj final state

Figure 4.31: Toy datasets with signal events superimposed on background events have been generated and the four-body invariant mass distribution fitted with a signal-plus-background pdf. The plot shows, for different background shape hypotheses, the average difference between the number of signal events estimated by the fit and the true number of signal events present in the dataset, normalized to the latter.

play a role in its estimation. Nevertheless, we profit from the dedicated study of Section 4.8.2 to define a systematic uncertainty associated with the specific choice of the background model.

While the results are summarized in Table 4.14, all of these aspects are detailed singularly in the next paragraphs.

4.9.1 Signal systematics

All source of systematic errors affecting directly the signal yield or shape determination are detailed here below.

Cross-section prediction

The inclusive cross-sections, used for the Higgs signal yield determination and calculated by the "Higgs Cross Section Working Group" [14] come with an associated uncertainty that stems from the variation of the QCD normalization and factorization scales and the uncertainties of the PDF used for the calculation (see also Section 4.1). The overall effect is of the order of the order of 13-19%, depending on the Higgs boson mass. The total cross-section uncertainty is driven by the level of refinement in the calculation of Higgs production via the gluon fusion production mechanisms, since this category of signal events dominate the final state. Nevertheless, since the gg and VBF channels are treated independently in the limit setting procedure, each one is assigned its own cross-section systematics.

The uncertainty associated to the PDF above described only accounts for a global normalization of signal events. The effect of the choice of different PDF sets – and of their intrinsic uncertainties – on the event selection efficiency is analysis-dependent and has to be specifically studied. The PDF4LHC recommendations [65] have been followed and the systematic uncertainty has been estimated as the envelope of the signal acceptance differences calculated for three sets of PDFs, namely the CTEQ6.6, MSTW08 and NNPDF2.0. Table 4.15 shows the values obtained.

source of uncertainty	gg		VBF	
	$e\nu_e jj$	$\mu\nu_\mu jj$	$e\nu_e jj$	$\mu\nu_\mu jj$
cross-section (QCD scale)	$7.3 - 8.7\%$		$0.4 - 1.5\%$	
cross-section (pdf)	$8.1 \rightarrow 10\%$		$2.9 \rightarrow 4.6\%$	
pdf acceptances	$1.5 - 3.6\%$		$0.6 - 1.1\%$	
Higgs line-shape	$2.3 \rightarrow 32\%$		/	
LHC luminosity	$\longleftarrow$	2.2% $\longrightarrow$		
pile-up	$\longleftarrow$	$< 1\%$ $\longrightarrow$		
jet/E_T^{miss} energy scale	$1.0 - 3.8\%$		$0.2 - 3.4\%$	
lepton energy scale	$0.9 \leftarrow 1.7\%$	$< 1\%$	$0.9 \leftarrow 1.7\%$	$< 1\%$
lepton efficiencies	$2.8 - 4.7\%$	$< 1\%$	$2.8 - 4.7\%$	$< 1\%$
PFm_T trigger efficiency	$< 1\%$	/	$< 1\%$	/
b-tag veto	$\longleftarrow$	$< 1\%$ $\longrightarrow$		
background modeling	$7.6 - 63\%$	$10 - 69\%$	$7.6 - 63\%$	$10 - 69\%$

Table 4.14: Sources of systematics considered in the analysis, with the corresponding effect on the signal yields. The range of variation is intended to span the different Higgs mass points. An arrow between to values indicate that the systematic uncertainty has a trend with the Higgs mass. The uncertainties are reported separately for the gluon fusion and vector boson fusion production mechanisms and for the electron and muon final states.

m_H (GeV/c^2)	syst.	
	gg	VBF
250	1.5%	1.1%
300	2.0%	0.9%
350	2.2%	0.8%
400	2.4%	0.6%
450	2.7%	0.7%
500	2.9%	0.9%
550	3.2%	0.9%
600	3.6%	0.7%

Table 4.15: Acceptance uncertainty for the signal selection efficiency related to PDF variation, as a function of the mass hypothesis, for gluon fusion (left) and vector boson fusion (right) production mechanisms.

Finally, in the Monte Carlo generators used to produce signal events, the so-called zero-width approximation is used, which means that Higgs bosons are produced on-shell, and successively decayed via an ad-hoc Breit-Wigner distribution implemented in the MC simulation program. This approximation, valid at a good level for light Higgs particles, may be inadequate for heavy Higgs searches. Recent studies [85, 86] have become available, showing that effects due to off-shell Higgs boson production and decay, and to the interference with the SM backgrounds,

may become sizable for Higgs boson masses $m_H \gtrsim 300\,\text{GeV}/c^2$, affecting in a sizable way the line-shape of the resonance. A conservative systematic uncertainty has been proposed to cover the inappropriate Monte Carlo description of Higgs boson production and decay at high mass. It is expressed by $150\% \cdot m_H^3$, where the Higgs mass is expressed in TeV/c^2. As it can be easily seen, this effect amounts to about 4% for a $300\,\text{GeV}/c^2$-mass Higgs, and the systematic grows rapidly with mass, contributing a $\sim 30\%$ effect at $600\,\text{GeV}/c^2$.

LHC luminosity

The latest recommendation for the CMS luminosity uncertainty has been followed, and it amounts to 2.2% [87]. This value improves a prior one thanks to the recent employment of the pixel detector, in place of forward calorimeters, to measure the offline luminosity. This value is dominated by the afterglow correction uncertainty (by which also pixel sensors are affected, even if to a lesser extent than HF) and by the scan-to-scan variations in the Van der Meer scans, needed to define the absolute pixel cross-section.

Pile-up modelization

The presence of pile-up events from secondary proton-proton interactions overlaid to the primary hard interaction can affect the event reconstruction in different manners: spurious particles can be clustered in jets or fall within the lepton isolation cones, thus modifying respectively the jet measured energy or the isolation variable. Moreover, additional jets, entirely from pile-up particles can originate.
Different strategies have been adopted in the analysis to mitigate the disturbing effect of pile-up, namely L1-Offset correction for jets and effective-area energy subtraction from lepton isolations cone. In addition to that, Monte Carlo events are generated with pile-up events superimposed to the hard scattering, and an ad-hoc weight is applied to each simulated event in order to match the pile-up conditions observed in data.
The residual source of systematic errors has two main different origins, i.e. the uncertainty in the number of interactions observed in data and uncertainties in the Monte Carlo model used to describe the pile-up topology.

Recalling Equation 4.2, systematic errors in the estimation of the number of pile-up interactions in data may come from the uncertainty on the luminosity measurement, currently 2.2% as it was stated above, and on the minimum-bias cross-section. The minimum-bias cross-section is derived from a best-fit analysis, comparing the distribution of the number of reconstructed vertices in data and Monte Carlo for $Z \to \mu^+\mu^-$ events, for which a central value of 68 mb was obtained. The combination in quadrature of the luminosity and $\sigma_{\text{min. bias}}$ errors yields a total uncertainty of approximately 3.6% on the estimated number of interactions.

A precise estimation of the impact of the potential differences in the pile-up model implemented in PYTHIA and what is observed in data is currently not available. Anyway, a total systematic of 5% should be safe enough to cover all sources of uncertainties in the pile-up modeling. A variation of $\pm 5\%$ in the number of interactions is then propagated to the re-weighting procedure for signal samples, and the obtained distributions of the four-body invariant mass are inspected to look for potential sculpting effects on the shape of the distribution. Figure 4.32 shows a typical variation for a Higgs mass hypothesis of $400\,\text{GeV}/c^2$. The overall effect on the

signal yields are found to be very small at all Higgs mass hypotheses ($\lesssim 1\%$), and no significant modification of the reconstructed Higgs mass shape is observed. Shape effects are therefore neglected.

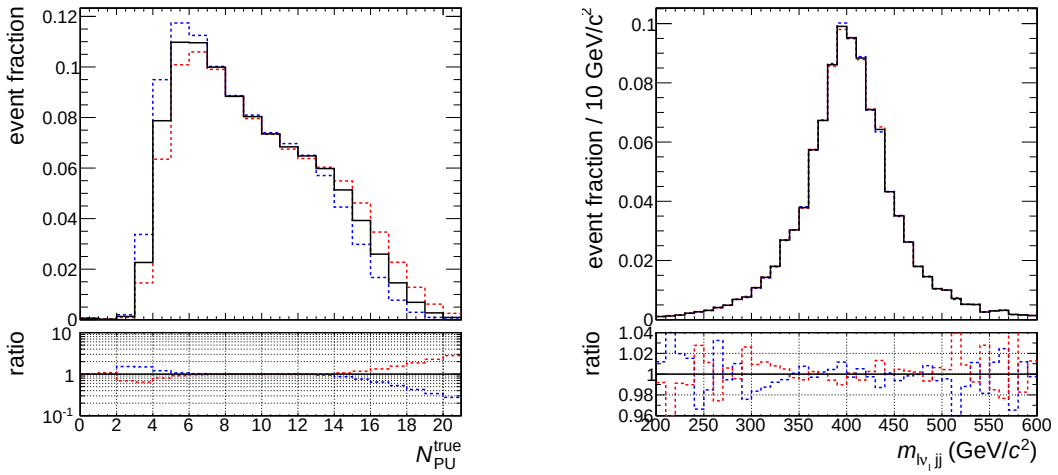


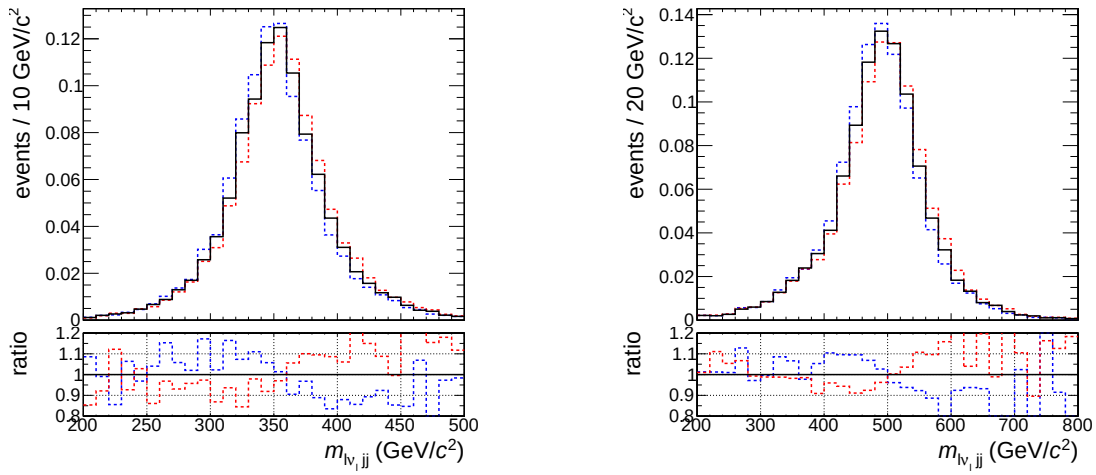
Figure 4.32: (a) Variation of $\pm 5\%$ of the number of interactions in data. (b) Effect of pile-up systematic on the Higgs mass peak after all analysis cuts (Higgs mass of 400 GeV/c^2).

Jet Energy Scale

A non-perfect knowledge of the energy scale of jets can affect both the signal efficiency, due to a different acceptance for jets with a given p_T threshold, and the shape of the jet-related quantities, namely the di-jet invariant mass m_{jj} and, most importantly, the four-body invariant mass, used in the final part of the analysis for the signal extraction.

The systematic effect due to the jet, and consequently E_T^{miss}, energy scale uncertainty (JES) has been taken into account by coherently varying the p_T of all reconstructed jets by $\pm 1\sigma$ of the uncertainty at which the energy scale is known by CMS measurements [88]. The JES uncertainty is a function of η and p_T and is known at 5% level or better in the region of interest for this analysis.
The missing transverse energy is varied coherently, in a fully correlated way. Recalling that $\vec{E}_T^{miss}$ is the negative vectorial sum of jet and other particle momenta in the event, all jets with $p_T > 15\,\text{GeV}/c$ are added to $\vec{E}_T^{miss}$, scaled according to the JES uncertainty, and successively summed again with an overall negative sign. The fraction of E_T^{miss} which is not clustered into jets with this p_T threshold is scaled by a conservative 5% factor. Figure 4.33 shows an example of the signal shape morphing when the jet energy scale is varied, while Table 4.16 summarizes the effect on signal yields of jet energy scale variation. Given the sizable modification in the four-body invariant mass shape, this source is handled as a yield and shape systematic in the limit extraction process.

As an additional source of systematics, the uncertainties in the knowledge of the jet energy resolution should be considered as well. For the p_T and η regions of interest for this analysis, the transverse energy resolution of PF anti-k_T jets amount to roughly 10%, with some dependence



(a) $m_H = 350\,\text{GeV}/c^2$ (b) $m_H = 500\,\text{GeV}/c^2$

Figure 4.33: Morphing of the signal four-body mass shape due to a $\pm 1\sigma$ jet energy scale variation for (a) a $350\,\text{GeV}/c^2$ and (b) a $500\,\text{GeV}/c^2$ Higgs.

m_H (GeV/c^2)	syst. gg	syst. VBF
250	2.2%	0.6%
300	1.0%	0.7%
350	1.2%	0.7%
400	2.0%	1.3%
450	2.3%	1.7%
500	2.9%	2.2%
550	3.6%	3.0%
600	3.6%	2.8%

Table 4.16: Systematic uncertainties related to the jet energy scale variations as a function of the Higgs mass and listed separately for gluon fusion and vector boson fusion production mechanisms.

on the jet transverse momentum and pseudorapidity. The resolution is known at a level which is everywhere smaller than 5%. We can therefore expect that the jet p_T variation needed to account for this systematic effect is at the 0.5% level or below. This is a tiny effect with respect to the above mentioned jet energy scale variation. Moreover, thanks to the di-jet mass kinematic fit, the final four-body invariant mass resolution is less sensitive to an imperfect description of the jet resolution in the Monte Carlo. For these reasons, this systematic effect has been neglected.

Lepton energy scale

As it is done for jets, the energy scales of electrons and muons are scaled too by their respective uncertainties, and the their impact on the signal estimations are calculated.

For $e\nu_e jj$ final states, a flat energy scale uncertainty of 2% has been assumed for electrons in the barrel ($|\eta| < 1.5$). It has been shown in Figure 3.12 that the ECAL energy scale is known to be stable to a much better level than 2% in whole 2011. Anyway, that plot is the result of

multiple iterations of the laser monitoring corrections and improvements of the ECAL channel intercalibration, successive to prompt data taking. For mere computing resources reasons, only datasets inherent to the physics analyses that demand an excellent energy resolution (e.g. the $H \to \gamma\gamma$ channel) have been re-processed with the latest improved ECAL conditions. In the `SingleElectron` dataset, used for this analysis, the data-taking laser and intercalibration models were used, for which a conservative 2% systematic uncertainty on the ECAL scale is appropriate. The impact of this sub-optimal electron reconstruction on the analysis, as shown in Table 4.17, is anyhow marginal.

For $\mu\nu_\mu jj$ final states, instead, the muon scale uncertainty has been assumed at the level of 1%. Table 4.17 reports the result for the muon channel as well.

m_H (GeV/c^2)	syst.	
	$e\nu_e jj$	$\mu\nu_\mu jj$
250	1.6%	0.6%
300	1.4%	0.5%
350	1.4%	0.4%
400	1.2%	0.4%
450	1.1%	0.3%
500	1.1%	0.3%
550	1.0%	0.3%
600	0.9%	0.3%

Table 4.17: Systematic uncertainty on signal yields due to lepton energy scale variations, shown for (left) electron and (right) muon final states.

Lepton offline and online efficiency

As reported in Sections 4.5 and 4.6, lepton reconstruction, selection and trigger efficiencies are computed using a Tag and Probe technique on $Z \to l^+l^-$ events in both data and Monte Carlo. From this measurement, data and MC efficiencies are derived and, from their ratio, scale factors ρ are obtained to re-weight the simulated samples. The scale factors come with an associated uncertainty σ_ρ, which is the result of the propagation of each single efficiency measured uncertainty. Such σ_ρ has been propagated through to the full analysis to check its impact on the signal shapes and yields.

Recalling that the total lepton efficiency is factorized in the product of reconstruction, identification/isolation and trigger efficiency, the global scale factor $\rho_{TOT} = \rho_{RECO} \times \rho_{ID} \times \rho_{HLT}$ has been conservatively scaled by $\pm 1\sigma_\rho$, which corresponds to consider the efficiencies of the three single steps fully correlated. The result of this systematic effect on the Higgs signal yields is reported in Table 4.18, upon verifying that the Higgs mass shapes are not affected by variations of lepton efficiencies.

In the very same fashion, the systematic uncertainty associated with the transverse mass cut part of the electron trigger has been estimated independently. The result on the signal yields is flat with the Higgs mass and everywhere below 1%.

m_H (GeV/c^2)	syst.	
	$e\nu_e jj$	$\mu\nu_\mu jj$
250	2.9%	0.6%
300	3.6%	0.6%
350	4.0%	0.5%
400	4.2%	0.5%
450	4.4%	0.5%
500	4.5%	0.5%
550	4.5%	0.5%
600	4.7%	0.5%

Table 4.18: Systematic uncertainty due to isolation, identification and trigger efficiency uncertainties for the signal rates, as a function of the Higgs mass hypothesis, for the (left) electron and (right) muon final states.

b-tagging

The tagging efficiencies for b and light flavour jets are measured in data and in Monte Carlo [81]. The ratios of data to Monte Carlo efficiencies (scale factors) are then calculated. The efficiencies and scale factors are provided in bins of jet p_T and η and can be used to correct for the mismatch between the efficiencies in data and Monte Carlo. These measurements have an associated uncertainty which we propagate to the analysis. For instance, the average b-tag efficiency scale factor for b jets is 0.94 ± 0.094 for the tagging algorithm and working points used in the analysis (TCHEM).

The efficiency corrections to simulated events are applied as follows. The flavor of each jet is identified using matching the jet to a b/c quark or to the hardest parton in a $\Delta R < 0.3$ cone. If no match is found, the jet is identified as a light-flavour jet. For each jet, a random number n is generated between 0 and 1, assuming a flat prior, and the jet is tagged according to the following criteria depending on the flavour:

- b and c jets are tagged if the discriminant value passes the cut defined by the TCHEM working point and n is lower than b-tag scale factor, so only a fraction ~ 0.94 of the heavy flavour jets that are claimed to be tagged in the simulation is kept;
- for light flavour and gluon-matched jets, the b-tagging discriminant output is not considered, but instead, to achieve the tagging efficiency measured in data, the jet is tagged if n is lower than the mis-tag efficiency measured in data.

For light flavour and gluon jets we use the measured efficiency rather than the scale factor since the scale factor for light flavour is larger than unity. This would introduce technical complications when using the above method. The number of b-tags is then used, as in the standard event selection, to veto top-like events. In particular, events with three central jets of $p_T > 30\,\text{GeV}/c$ in the final state are required to have no b-tags.

The uncertainty on this correction is propagated to the analysis by varying the scale factor or efficiency correction by $\pm 1\sigma$ in the above procedure. Since the uncertainty on the c-jet tagging efficiency is unknown, we conservatively apply twice the b-tag efficiency uncertainty to these jets, in accordance with the b-tag group recommendations. The variation in the signal event

yield is then taken as the normalization uncertainty, as shown in Tables 4.19. The effect of this uncertainty on the shape of $m_{l\nu_l jj}$ was found to be negligible and so is not considered in the following steps of the analysis

m_H (GeV/c^2)	syst.	
	$e\nu_e jj$	$\mu\nu_\mu jj$
250	0.3%	0.2%
300	0.4%	0.4%
350	0.5%	0.4%
400	0.6%	0.5%
450	0.6%	0.7%
500	0.7%	0.7%
550	0.8%	0.8%
600	0.8%	0.7%

Table 4.19: Systematic uncertainty due to uncertainties in the b-tag data/MC scale factors and efficiencies. The impact on the signal rates, as a function of the Higgs mass hypothesis, is shown for the (left) electron and (right) muon final states.

4.9.2 Background systematics

In Section 4.8.2, the fit procedure implemented to estimate the background level of contamination in the final data sample has been thoroughly described. An intrinsic source of uncertainty in this measurement depends on the fluctuations of the data sample, and is reflected in the best-fit parameter uncertainties. This statistical component of the total background uncertainty is naturally considered in the limit setting procedure described in Section 4.10: the fit function parameters are handled as nuisance parameters, on the same page of all other systematic uncertainties, and simultaneously fitted to the data.

An additional, systematic component of the total background uncertainty arises from our ignorance of the exact parameterization of the $m_{l\nu_l jj}$ shape in data. The dedicated study of Section 4.8.2 has been devoted to the assessment of the bias introduced in the estimation of the local number of background events (i.e. in a specific Higgs mass window) when a particular function is chosen as background model (a Fermi-attenuated power law, in this analysis).
In the Higgs cross-section limit extraction procedure, though, the total number of background events is fitted to the data, and therefore it is not meaningful to assign a systematic uncertainty to it. Since the bias due to potential shape differences between our chosen model and true data corresponds to generating fake signals or artificially depleted regions in the invariant mass spectrum, the bias can be taken into account as an additional systematic uncertainty affecting the signal. The bias, calculated in terms of a difference of background events Δ, is translated into a signal relative error δ by quoting the background event difference relative to the number of expected signal events in the same mass window.

The amount of such a systematic uncertainty is derived from the results shown in Figure 4.31. The four different background models used to generate the pseudo-dataset are all equivalent, i.e. they are all capable of well reproducing the four-body invariant mass shape. Since there are no valid motivations to prefer any of them (or any other function we may think of), singularly for each Higgs mass point the average of the absolute value of the bias obtained with the four

different truth models is taken as the additional systematic uncertainty on the signal for that mass, i.e. $\sum_{j=1}^{4} |\delta_{ij}|$, with i identifying the chosen fit model and j spanning the different truth models.

The result is graphically shown in Figure 4.34, and exact numbers are listed in Table 4.20. While a coherent picture emerges for electron and muon final states, it is clear that this method for the bias determination suffers from a certain degree of arbitrarity, for instance due to the limited set of functions used as truth model.
The rather large uncertainties, especially for heavy Higgs boson hypotheses, are expected to well cover all systematic effects related to the usage of a specific fit function, which not already included in the fit parameter error evaluation.

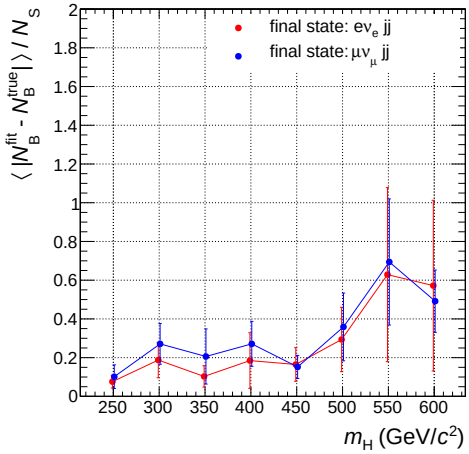


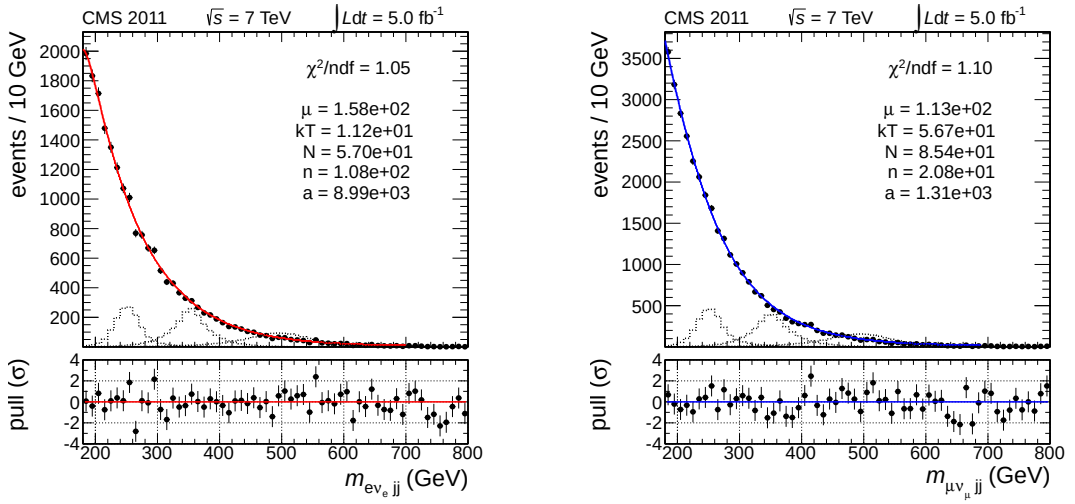
Figure 4.34: Signal systematic uncertainty due to the function choice bias, shown versus the Higgs mass hypothesis for the (red) electron and (blue) muon final states. The error bands correspond to the RMS of the absolute bias distribution, for each mass hypothesis.

m_H (GeV/c^2)	syst.	
	eν$_e$ jj	μν$_μ$ jj
250	7.6%	10%
300	19%	27%
350	10%	21%
400	19%	27%
450	17%	15%
500	29%	36%
550	63%	69%
600	57%	49%

Table 4.20: Additional signal systematic uncertainty due to the background shape bias for (left) electron and (right) muon final states.

4.10 Setting a limit on the SM Higgs boson cross section

The observed data in the final signal phase space are presented in Figure 4.35 for the electron and muon final states separately. The expected contribution of an hypothetical Higgs signal are superimposed as well, for three reference masses (signal yields have been enhanced by some multiplicative factor to ease the comparison with the background shape).
The background fits, performed with the chosen fit function, are overlaid: in both channels, the observed shape in data is well reproduced, and the fits converge with $\tilde{\chi}^2$ values close to one. These fit results are here reported only for illustrative purposes, and are not used anywhere in the analysis. As a matter of fact, the background fit parameters are determined during the limit extraction procedure itself, as we are to describe here below.



(a) $e\nu_e$ jj final state in the signal region (b) $\mu\nu_\mu$ jj final state in the signal region

Figure 4.35: Background-only fits of the four-body invariant mass distribution in data for the m_{jj} signal region. A first-order power law function multiplied by a Fermi turn-on is used as parametric shape. Higgs boson signals expected from the Monte Carlo simulation for three candidate mass hypotheses, 250,400 and 500 GeV/c^2 are superimposed as well, enhanced by a factor of 20 (40 for $m_H = 500$ GeV/c^2) to make them more clearly visible.

4.10.1 The statistical analysis

The purpose of this analysis is to infer a constraint on the existence of the Higgs boson for a set of different Higgs mass hypotheses. In particular, if no significant excess is observed in data, compatible with the expected Higgs resonance, an upper limit on the Higgs boson cross-section can be put, up to a certain degree of belief. For example, it can be said that there is a 95% C.L. that the Higgs boson cross-section is not larger than some value $\sigma_{95\%\,\text{C.L.}}$; otherwise said, if the cross-section was larger that that, a statistically-significant excess of data would have emerged over the background. If the upper limit $\sigma_{95\%\,\text{C.L.}}$ is as large as the production cross-section expected from the Standard Model σ_{SM} or smaller, it can be said that a SM Higgs boson of that mass is excluded at 95% C.L..

The CMS community has agreed upon the usage of the so-called modified frequentist (CL_s) method to calculate the cross-section upper limit on the observed data. While a thorough de-

scription of such method can be found e.g. at [89], the basic ingredients are summarized here below.

To construct the test statistics needed to define CL_s, a likelihood function is defined as follows:

$$\mathcal{L}(\text{data}|\mu,\theta) = \mathcal{P}(\text{data}|\mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta) \quad . \tag{4.11}$$

In the above equation, $s(\theta)$ and $b(\theta)$ are respectively the signal and background expectations, depending on some set of nuisance parameters θ of observed value $\tilde{\theta}$. The signal strength modifier μ has been introduced with the purpose of changing the cross-section of all signal processes described in s by the same scale, that is $\sigma = \mu \cdot \sigma_{SM}$. The likelihood then reads as the Poisson probability[10] of observing a certain amount of data when the expected yield is $\mu \cdot s(\theta) + b(\theta)$, times the probability of measuring a value $\tilde{\theta}$ for the nuisance parameter θ. Note that, in this likelihood definition, "data" stands for a generic dataset, either experimental or a pseudo-dataset generated randomly.
The appropriate test statistics to quantify an absence of signal is then defined as

$$q_\mu = -2 \log \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})} \quad , \qquad \text{with} \qquad 0 \leq \hat{\mu} \leq \mu \quad . \tag{4.12}$$

Both the numerator and the denominator in Equation 4.12 are maximized ($\hat{}$ superscript), and the subscript $_\mu$ indicates that the maximization of the likelihood in the numerator has been done with a fixed value of μ. The lower and upper constraints on the maximized $\hat{\mu}$ in the denominator force the limit to be one-sided and exclude the possibility of negative signal yields.
The value of the nuisance parameters that best describe experimental data are obtained through a maximum-likelihood fit in the background-only and in the signal-plus-background hypothesis (i.e. setting μ to 0 and to some level μ in $\mathcal{L}$), from which $\hat{\theta}_0^{obs}$ and $\hat{\theta}_\mu^{obs}$ 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 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 real data for a CMS Higgs combination with the first inverse femtobarn of data is shown in Figure 4.36.
Finally, the test statistics can be evaluated on real data, obtaining q_μ^{obs}. From the test statistics pdfs, two p-values associated with the observation can be calculated, one for the signal-plus-background (p_μ) and one for the background-only hypothesis (p_0), namely

$$\begin{aligned} p_\mu \equiv CL_{s+b} &= P(q_\mu \geq q_\mu^{obs} | \mu \cdot s(\hat{\theta}_\mu^{obs}) + b(\hat{\theta}_\mu^{obs})) \\ p_0 \equiv CL_b &= P(q_\mu \geq q_\mu^{obs} | b(\hat{\theta}_0^{obs})) \end{aligned} \quad . \tag{4.13}$$

From these p-values, CL_s (μ) can be calculated as

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad . \tag{4.14}$$

If $CL_s = \alpha$ for a signal-strength modifier value $\mu = 1$ ($\sigma \equiv \sigma_{SM}$), we say that a SM Higgs boson

[10] $\mathcal{P}$ 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).

is excluded at 95% C.L.. In order to quote, as conventionally done, 95% C.L. observed upper limits, the full procedure is iterated for different values of μ, until $CL_s = 0.05$ is found, in which case we say that a Higgs boson with a cross-section μ-times larger than the one predicted in the Standard Model is excluded.

In order to derive the median-expected upper limit, and the associated ± 1 and 2σ bands, a large set of background-only pseudo-data can be generated and, for each of them, CL_s and $\mu_{95\%\,\mathrm{C.L.}}$ are calculated. From the cumulative distribution of $\mu_{95\%\,\mathrm{C.L.}}$, the median value is taken as the expected limit in case no signal is present. $\pm 1(2)\sigma$ uncertainty bands on the expected limits are extracted from the values of the 16%(2.5%) and 84%(97.5%) quantiles.

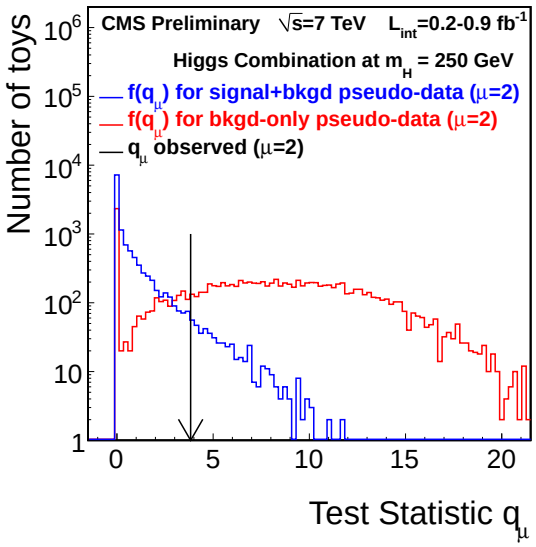


Figure 4.36: Frequentist tests statistic q_μ distributions for ensembles of pseudo-data generated for background-only and signal+background hypotheses. The signal strength assumed in this example is $\mu = 2$. The observed value of the test statistic is indicated by the arrow.

4.10.2 The result

The statistical framework above discussed has been applied to the analysis described in this work, using the all data collected in 2011 and the Monte Carlo signal expectations at the end of the analysis chain. The maximum likelihood fit has been done combining the event counts of the reconstructed Higgs invariant mass, binned in 5 GeV/c^2-wide bins, from 180 to 800 GeV/c^2. All systematic uncertainties described in Section 4.9 have been included as nuisance parameters θ. In particular, the jet energy scale effect has been properly accounted for not only for its effect on the overall signal normalization, but also for the shape modifications that it induces on the reconstructed Higgs mass by taking into account bin-to-bin correlations.
The parameters of the background fit model are handled as additional nuisance parameters, whose best value is determined, together with all other nuisances, when the numerator and the denominator of q_μ are maximized to calculate q_μ^{obs}. Also, they are independently determined when $\mathcal{L}$ is maximized in the $\mu \cdot s + b$ and b only hypotheses to construct the pseudo-dataset for the test statistics pdfs determination.

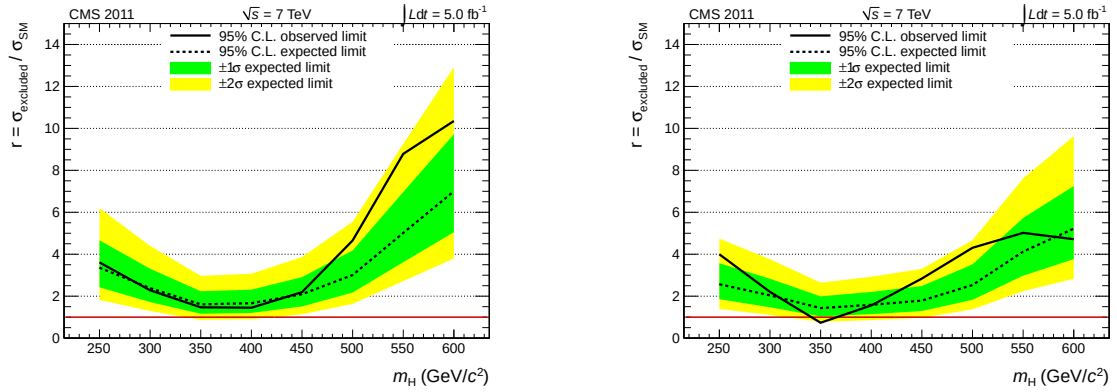
The obtained limit is reported in Figure 4.37 for the electron and the muon final states separately. In Figure 4.38, the statistical combination of the two final states is also shown[11].
The expected limit and the associated uncertainty bands from the generation of pseudo-experiments are represented with a dashed line and coloured areas respectively, while the observed limit is represented with a solid black line. It can be noticed that the expected limit on the SM Higgs cross-section approaches the sensitivity necessary to exclude its existence in a wide area between about 350 and 425 GeV/c^2, while the sensitivity degrades especially at high mass, due to the increasingly lower Higgs production cross-section, the sizable width of the resonance and the larger systematics. The observed limit agrees in general at a good level with expectations. A downward fluctuation is observed in the region of maximal sensitivity of the analysis, allowing for a SM Higgs exclusion in the mass region $[325 - 400]$ GeV/c^2 by this channel alone.

Figure 4.39 summarizes the CMS most up-to-date Physics reach of SM Higgs boson searches, obtained combining all experimentally-accessible production and decay modes of the Higgs boson [90]. In particular, five distinct decay modes enter the statistical combination, namely $H \to b\bar{b}$, $H \to \tau^+\tau^-$, $H \to \gamma\gamma$, $H \to ZZ$ and $H \to W^+W^-$. This result, presented at the "Rencontres de Moriond" conference (La Thuile, 3–10 March 2012), shows that CMS alone is able to exclude a SM Higgs boson with at least 95% C.L. all the way from 126 to 600 GeV/c^2. An interesting excess of events is observed at 124 GeV/c^2, where the sensitivity is mostly driven by the $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to l^+l^- l^+l^-$ channels (see below).
In Figure 4.40, the contribution of each single Higgs decay mode to the total exclusion limit can be deduced.

The result of this work has not been taken into account in the CMS combination yet. From Figure 4.40, anyway, it can be noticed that the $H \to WW \to l\nu_l q\bar{q}$ final state will bring in additional sensitivity in the medium-to-high mass range, where its contribution is at the same level of all other $H \to W^+W^-$ channels combined. Interestingly enough, the result of this analysis allows to exclude the SM Higgs boson in the region around 350 GeV/c^2, where all other channels combined only marginally observe an exclusion. The contribution of this channel in that mass region is therefore sizable.

The analysis described in this work has recently become a public CMS document [91]. Also the ATLAS experiment has published a result on this very same final state [92]. CMS Higgs searches in this channel have already started with recently collected data at $\sqrt{s} = 8$ TeV, and the $H \to WW \to l\nu_l q\bar{q}$ channel will be taken into account in future combinations of CMS Higgs results.

[11] In these plots, an asymptotic approximation of the q_μ test statistics is used, which allows to model the q_μ pdf with a χ^2 distribution for one degree of freedom, and save computational time, not having to generate pseudo-data events



(a) $e\nu_e$ jj final state (b) $\mu\nu_\mu$ jj final state

Figure 4.37: The Standard Model Higgs exclusion limit, obtained from a fit to the data after selections described in Section 4.3 for the (a) the electron and (b) the muon final states. The statistical result is obtained using the asymptotic approximation of the test statistics.

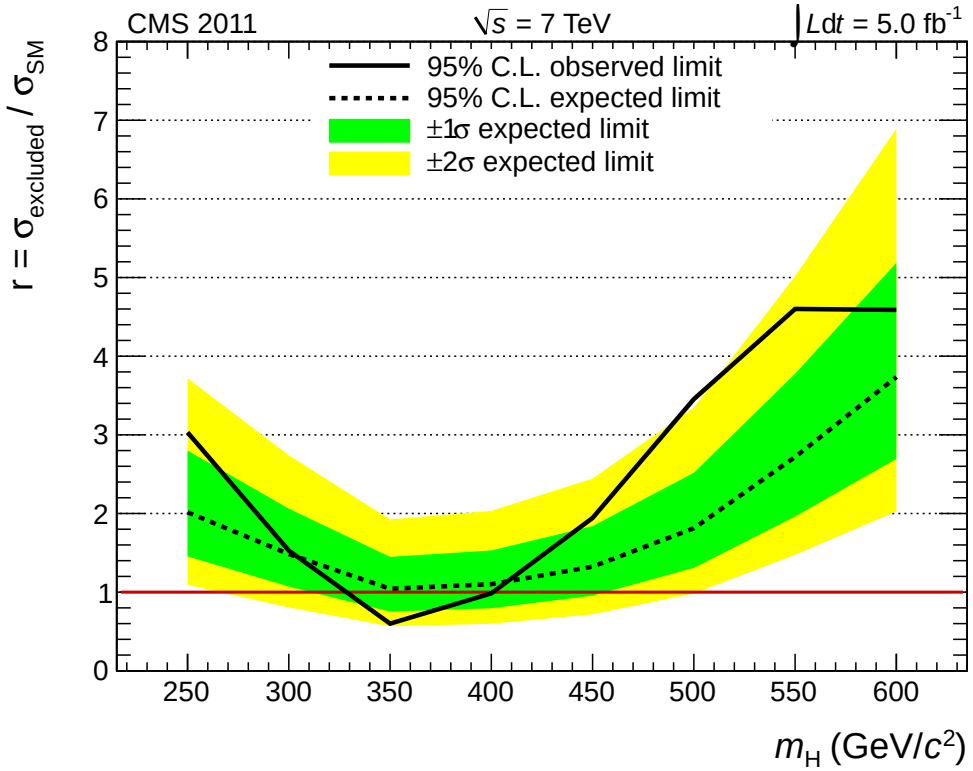
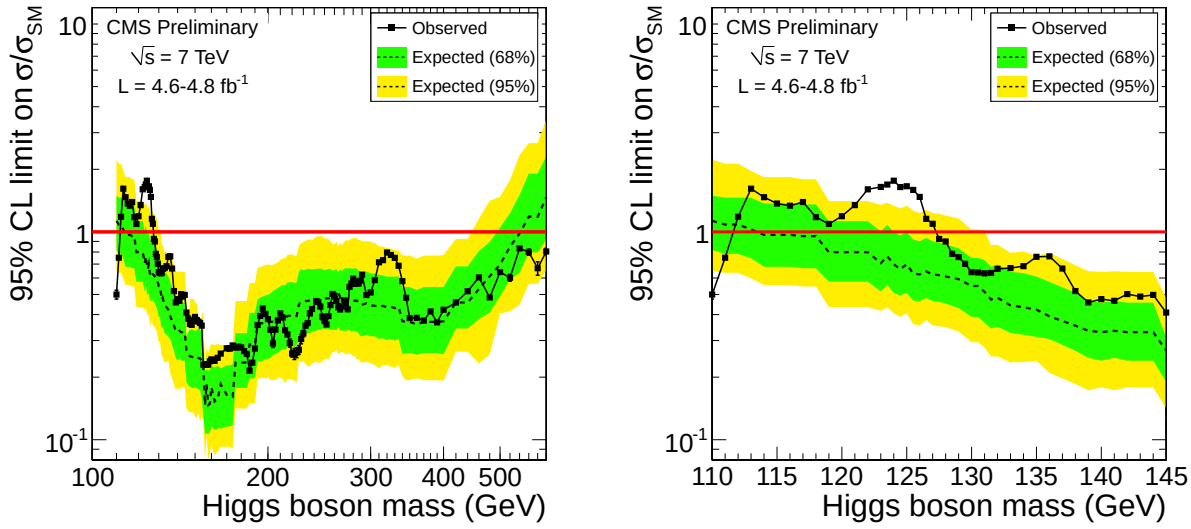


Figure 4.38: The Standard Model Higgs exclusion limit, obtained from a fit to the data after selections described in Section 4.3 and combining the electron and muon final states. The statistical result is obtained using the asymptotic approximation of the test statistics.



(a) (b) $\mu\nu_\mu$ jj final state

Figure 4.39: 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 in (left) the 110-600 GeV/c^2 and (right) the 110-145 GeV/c^2 range.

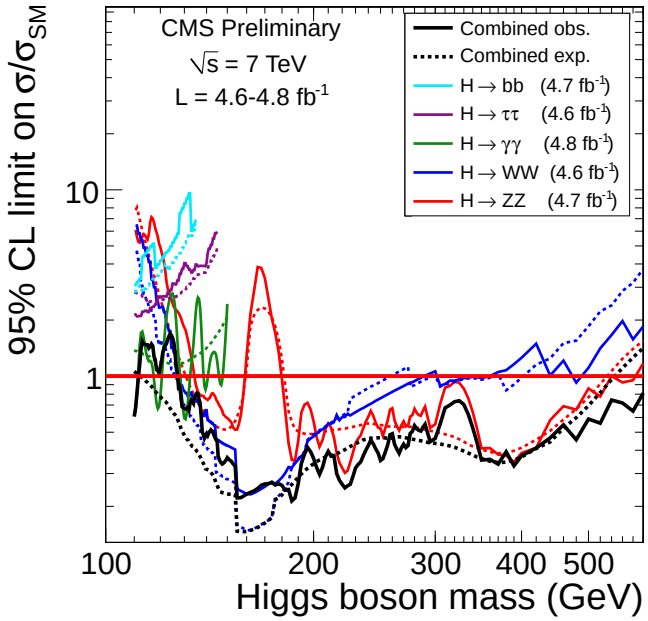


Figure 4.40: The observed and expected 95% C.L. upper limits on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for the SM Higgs boson hypothesis as a function of the Higgs boson mass in (left) the 110-600 GeV/c^2 and (right) the 110-145 GeV/c^2 range, for the five explored Higgs decay modes and their combination.

Chapter 5

Conclusions

The most relevant achievements of my three-year Ph.D. work in CMS have been presented. The field of interests of my research activity has been twofold. On the detector side, I have followed different aspects connected with the electromagnetic calorimeter and calibration and commissioning. On the Physics analysis side, I have significantly contributed to the development of the first CMS analysis that searches for a SM Higgs boson in the $H \to WW \to l\nu_l q\overline{q}$ decay channel.

Although not specifically addressed in this thesis, my work on ECAL began before the first proton-proton beams were delivered by the LHC. During Fall 2008, after reaching its final position 100 m underground in the cavern, the whole detector has gone through a month-long data-taking exercise exploiting cosmic-ray muons and all the sub-detectors have been thoroughly tested. Muon track reconstruction and momentum measurement in the inner silicon tracker and the ECAL measurement of the energy loss in the lead tungstate crystals allowed for a measurement of specific energy loss of muons in $PbWO_4$, namely dE/dx versus p_μ. The average measured stopping power in the ECAL barrel has been compared with the expected values for the muon stopping power in lead tungstate taken from literature. Other than allowing for a check of the global energy scale of the calorimeter (validated at 2% level at the $\mathcal{O}(100\,\mathrm{MeV})$ energy scale), the first measurement of the muon critical energy in $PbWO_4$ has been obtained as well. These results are described in one of the first CMS Physics papers [93].
With the first collisions data collected by CMS in 2010 and 2011, Physics events from $W \to e\nu_e$ decays have been used to implement a prompt and reliable monitoring tool to allow for quasi on-line checks of the ECAL stability. Given the harsh radiation environment, ECAL crystals are affected by a significant loss of transparency, that in 2011 amounts on average to about 2.5% in the barrel and 10% in the endcaps. A laser monitoring system has been installed to monitor and provide corrections for changes in the ECAL reponse due to the radiation-induced crystal transparency variation. An analysis was developed, based on the comparison of the electron ECAL-measured energy and their momentum as measured in the tracker system. The E_{sc}/p_{tk} technique has proven to be one of the pillars of the commissioning phases of the transparency corrections during 2010 and 2011, allowing to promptly spot problems and driving the necessary actions for their solution.
It was also possible to profit from the very same analysis technique for a large variety of different purposes. In particular, a first *in-situ* measurement of the crystal spectral index was

possible, indicating the need for a different (lower) value than the default one from test-beam measurements. Electrons from $W \to e\nu_e$ and $Z \to e^+e^-$ events have also allowed to check the electron local energy scale. In some cases, an *ad-hoc* correction has been derived, bringing to an improvement of the energy resolution.
All these studies act as a reference for the $H \to \gamma\gamma$ decay channel. This is the golden-plated mode for very light Higgs mass hypotheses, and it is among the most demanding in terms of energy resolution, to fully profit from the Higgs narrow width at those masses.

As regarding the Physics analysis channel studied in this thesis, the $H \to WW \to l\nu_l q\bar{q}$ decay mode of the SM Higgs was inspected. The motivations for this analysis reside in its large cross-section times branching ratio - in fact the largest among all experimentally-accessible final states - and in the possibility to fully reconstruct its invariant mass. Drawback to this is the large SM bacgkround contamination in the final state.
I have developed a full analysis chain. Firstly, some effort was put in the definition of a dedicated trigger path, that allowed to retain as much $e\nu_e jj$ events as possible, albeit the constantly-increasing instantaneous luminosity.
The signal and Monte Carlo expectations in this final state have been thoroughly studied, in order to define a selection strategy that maximizes the sensitivity to Higgs boson events. Simulation studies showed that this final state can actually be competitive with other CMS analyses. This analysis was configured as the search for a signal peak appearance over the large background continuum, also thanks to the mass resolution improvements allowed by a kinematic fit of the Higgs decay chain.
All possible source of systematic uncertainties affecting the signal have been deeply investigated, and their effect on the Higgs signal expectations have been evaluated.
All data collected in 2011 have been then analyzed in this final state. 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. In general, the analysis sensitivity is such to almost allow the exclusion of a SM Higgs boson at 95%C.L. in a wide mass region around $350\,\mathrm{GeV}/c^2$. The observed limits on the excluded cross-section, relative to the SM expected one, allow to exclude the existence of the Higgs boson between 325 and $400\,\mathrm{GeV}/c^2$ with this final state alone. Also outside this mass range, this analysis can anyhow add a sizable significance to the combination of all CMS analyses, being its sensitivity e.g. of the same order of other $H \to W^+W^-$ decay modes.

Bibliography

[1] A. Djouadi, "The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model", *Phys.Rept.* **457** (2008) 1–216, arXiv:hep-ph/0503172. doi:10.1016/j.physrep.2007.10.004. 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] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group Collaboration, "Precision electroweak measurements on the Z resonance", *Phys.Rept.* **427** (2006) 257–454, arXiv:hep-ex/0509008. doi:10.1016/j.physrep.2005.12.006. 6

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

[6] P. W. Anderson, "Plasmons, Gauge Invariance, and Mass", *Phys. Rev.* **130** (Apr, 1963) 439–442. doi:10.1103/PhysRev.130.439. 7

[7] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323. doi:10.1103/PhysRevLett.13.321. 7

[8] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509. doi:10.1103/PhysRevLett.13.508. 7

[9] H.-K. Quang and X.-Y. Pham, "Elementary Particles and their Interactions". Springer-Verlag Berlin, 1998. 9

[10] G. Altarelli and G. Isidori, "Lower limit on the Higgs mass in the standard model: An Update", *Phys.Lett.* **B337** (1994) 141–144. doi:10.1016/0370-2693(94)91458-3. 12

[11] 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, 16

[12] 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

[13] 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. 15, 16

[14] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", arXiv:1101.0593. 17, 19, 79, 85, 128

[15] S. Dawson, "Radiative corrections to Higgs boson production", *Nucl.Phys.* **B359** (1991) 283–300. doi:10.1016/0550-3213(91)90061-2. 17

[16] 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. 17

[17] 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. 17

[18] 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. 17

[19] 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

[20] M. Zaro, P. Bolzoni, F. Maltoni et al., "Higgs production via vector-boson fusion at NNLO in QCD", *PoS* **DIS2010** (2010) 211. 18

[21] 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

[22] 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

[23] 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. 18

[24] C. Rovelli, "Standard Model Higgs searches with the CMS detector", *CMS Conference reports* **CR-07-049** (2007). 20

[25] CMS Collaboration, "Measurement of W+W- Production and Search for the Higgs Boson in pp Collisions at sqrt(s) = 7 TeV", *Phys.Lett.* **B699** (2011) 25–47, arXiv:1102.5429. 20

[26] CMS Collaboration Collaboration, "Search for Neutral MSSM Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV", *Phys.Rev.Lett.* **106** (2011) 231801, arXiv:1104.1619. 20

[27] ATLAS Collaboration, "Limits on the production of the Standard Model Higgs Boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Eur.Phys.J.* **C71** (2011) 1728, arXiv:1106.2748. 20

[28] CMS Collaboration, "Search for standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summaries* **HIG-11-011** (2011). 20, 21

[29] ATLAS Collaboration, "Combination of the Searches for the Higgs Boson in $\sim 1\,\text{fb}^{-1}$ of Data Taken with the ATLAS Detector at 7 TeV Center-of-Mass Energy", *ATLAS Notes* **ATLAS-CONF-2011-112** (2011). 20, 21

[30] S. Alekhin, G. Altarelli, N. Amapane et al., "HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A", arXiv:hep-ph/0601012. 23

[31] ATLAS Collaboration, "Collider Physics: LHC", *ATLAS Conference Talks* **ATLAS-CONF-2000-001** (2000). 23

[32] U. Amaldi, W. de Boer, and H. Furstenau, "Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP", *Phys.Lett.* **B260** (1991) 447–455. doi:10.1016/0370-2693(91)91641-8. 24

[33] O. S. Brüning, P. Collier, P. Lebrun et al., "LHC Design Report". CERN, Geneva, 2004. 24

[34] CMS Collaboration, "CMS, the Compact Muon Solenoid: technical proposal", *LHC Tech. Proposal* **CERN-LHCC-94-38** (1994). 25

[35] 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

[36] LHCb Collaboration, "LHCb: Technical Proposal", *Tech. Proposal* **CERN-LHCC-98-004** (1998). 25

[37] ALICE Collaboration, "ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC", *LHC Tech. Proposal* **CERN-LHCC-95-71** (1995). 25

[38] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004. 27, 30, 58

[39] CMS Collaboration, "The CMS tracker system project: Technical Design Report", *Technical Design Report CMS* **CMS-TDR-005** (1997). 27

[40] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", *Technical Design Report CMS* **CMS-TDR-004** (1997). 28, 30, 45, 55, 58

[41] CMS Collaboration, "The CMS hadron calorimeter project : Technical Design Report", *Technical Design Report CMS* **CMS-TDR-002** (1997). 31

[42] CMS Collaboration, "The CMS muon project : Technical Design Report", *Technical Design Report CMS* **CMS-TDR-003** (1997). 32

[43] CMS Collaboration, "CMS TriDAS project : Technical Design Report; 1, the trigger systems", *Technical Design Report CMS* **CMS-TDR-006-** (2000). 34

[44] E. Meschi, T. Monteiro, C. Seez et al., "Electron Reconstruction in the CMS Electromagnetic Calorimeter", *CMS Notes* **CMS-NOTE-2001-034** (2001). 35

[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. 36

[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. 36

[47] CMS Collaboration, "CMS Physics Technical Design Report Volume I : Detector Performance and Software", *Technical Design Report CMS* **CMS-TDR-008-1** (2006). 36, 45, 55

[48] 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). 37

[49] 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). 37

[50] 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. 39

[51] 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. 40

[52] CMS Collaboration, "Jet Energy Resolution in CMS at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summaries* **JME-10-014** (2010). 40

[53] P. Adzic, R. Alemany-Fernandez, C. Almeida et al., "Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter", *Eur.Phys.J.* **C46S1** (2006) 23–35. doi:10.1140/epjcd/s2006-02-002-x. 45

[54] 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. 58

[55] D. Elizabeth, "The effect of pulse rate on VPT response and the use of an LED light to improve stability", *CMS Conference Reports* **CMS-CR-2009-284** (2009). 62

[56] R. Paramatti, "Calibration of CMS electromagnetic calorimeter at LHC startup", *J.Phys.Conf.Ser.* **293** (2011) 012045. 71

[57] CMS Collaboration, "Commissioning of the CMS Experiment and the Cosmic Run at Four Tesla", *JINST* **5** (2010) T03001, arXiv:0911.4845. doi:10.1088/1748-0221/5/03/T03001. 72

[58] CMS Collaboration, "Electromagnetic calorimeter calibration with 7 TeV data", *CMS Physics Analysis Summaries* **EGM-10-003** (2010). 72

[59] U. Chaturvedi, A. Favara, M. Gataullin et al., "Results of L3 BGO calorimeter calibration using an RFQ accelerator", *IEEE Trans.Nucl.Sci.* **47** (2000) 2101–2105. doi:10.1109/23.903855. 72

[60] CMS Collaboration, "A search using multivariate techniques for a standard model Higgs boson decaying into two photons", *CMS Physics Analysis Summaries* **HIG-12-001** (2012). 74

[61] CMS Collaboration Collaboration, "Search for the standard model Higgs boson decaying into two photons in pp collisions at $\sqrt{s} = 7$ TeV", *Phys.Lett.* **B710** (2012) 403–425, arXiv:1202.1487. 74

[62] Nakamura, K. et al. (Particle Data Group), "Review of particle physics", *J.Phys.G* **G37** (2010) 075021. doi:10.1088/0954-3899/37/7A/075021. 79

[63] 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. 80

[64] 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. 81

[65] M. Botje, J. Butterworth, A. Cooper-Sarkar et al., "The PDF4LHC Working Group Interim Recommendations", arXiv:1101.0538. 82, 128

[66] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $O(\alpha_S^2)$", *Phys.Rev.* **D74** (2006) 114017, arXiv:hep-ph/0609070. doi:10.1103/PhysRevD.74.114017. 83

[67] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **1107** (2011) 018, arXiv:f1105.0020. 83

[68] N. Kidonakis, "Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution", *Phys.Rev.* **D82** (2010) 114030, arXiv:1009.4935. doi:10.1103/PhysRevD.82.114030. 84

[69] N. Kidonakis, "Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production", *Phys.Rev.* **D83** (2011) 091503, arXiv:1103.2792. doi:10.1103/PhysRevD.83.091503. 84

[70] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-", *Phys.Rev.* **D82** (2010) 054018, arXiv:1005.4451. doi:10.1103/PhysRevD.82.054018. 84

[71] N. Kidonakis, "NNLL resummation for s-channel single top quark production", *Phys.Rev.* **D81** (2010) 054028, arXiv:1001.5034. doi:10.1103/PhysRevD.81.054028. 84

[72] 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. 86

[73] 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. 86

[74] 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. 86

[75] S. Alioli, P. Nason, C. Oleari et al., "NLO Higgs boson production via gluon fusion matched with shower in POWHEG", *JHEP* **0904** (2009) 002, arXiv:hep-ph/0812.0578. doi:10.1088/1126-6708/2009/04/002. 86

[76] 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. 86

[77] 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. 87

[78] CMS Collaboration, "Measurement of the underlying event in the Drell-Yan process in proton-proton collisions at $\sqrt{s} = 7$ TeV", arXiv:1204.1411. 87

[79] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl.Instrum.Meth.* **A506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8. 87

[80] CMS Collaboration, "Measurement of the inclusive production cross sections for forward jets and for dijet events with one forward and one central jet in pp collisions at $\sqrt{s} = 7$ TeV", arXiv:1202.0704. 89

[81] CMS Collaboration, "Status of b-tagging and vertexing tools for 2011 data analysis", *CMS Physics Analysis Summaries* **BTV-11-002** (2011). 98, 134

[82] J. D'Hondt, S. Lowette, O. L. Buchmüller et al., "Fitting of Event Topologies with External Kinematic Constraints in CMS", *CMS Notes* **CMS-NOTE-2006-023** (2006). 102

[83] 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. 103

[84] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7TeV$"", *CMS Physics Analysis Summaries* **MUO-10-002** (2010). 105

[85] G. Passarino, C. Sturm, and S. Uccirati, "Higgs Pseudo-Observables, Second Riemann Sheet and All That", *Nucl.Phys.* **B834** (2010) 77–115, arXiv:1001.3360. doi:10.1016/j.nuclphysb.2010.03.013. 129

[86] C. Anastasiou, S. Buehler, F. Herzog et al., "Total cross-section for Higgs boson hadroproduction with anomalous Standard Model interactions", *JHEP* **1112** (2011) 058, arXiv:1107.0683. doi:10.1007/JHEP12(2011)058. 129

[87] CMS Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update"", *CMS Physics Analysis Summaries* **SMP-12-008** (2012). 130

[88] CMS Collaboration Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, arXiv:1107.4277. 131

[89] CMS Collaboration Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at sqrt(s) = 7 TeV", *Phys.Lett.* **B710** (2012) 26–48, arXiv:1202.1488. 138

[90] CMS Collaboration, "Combination of SM, SM4, FP Higgs boson searches", *CMS Physics Analysis Summaries* **HIG-12-008** (2012). 140

[91] 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). 140

[92] ATLAS Collaboration, "Search for the Higgs boson in the $H \to WW \to l\nu_l jj$ decay channel using 4.7 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *ATLAS Conference Talks* **ATLAS-CONF-2012-018** (2000). 140

[93] CMS Collaboration, "Measurement of the Muon Stopping Power in Lead Tungstate", *JINST* **5** (2010) P03007, arXiv:0911.5397. 143