Instituto de Física de Cantabria
(CSIC-Universidad de Cantabria)
y
Departamento de Física Moderna
(Universidad de Cantabria)

Medida de la sección eficaz de producción del quark top y búsqueda del bosón de Higgs del SM en estados finales dileptónicos con el detector CMS del LHC a $\sqrt{s} = 7$ TeV

Memoria de tesis presentada por
Patricia Lobelle Pardo
para optar al grado de Doctor

Dirigida por
Dr. Javier Cuevas Maestro
y siendo tutor
Dr. Alberto Ruiz Jimeno

Santander, Mayo de 2011

Instituto de Física de Cantabria
(CSIC-Universidad de Cantabria)
y
Departamento de Física Moderna
(Universidad de Cantabria)

Measurement of the top quark production cross section and search for the SM Higgs boson in dilepton final states with the CMS detector at the LHC pp collider at $\sqrt{s} = 7$ TeV

Memoria de tesis presentada por
Patricia Lobelle Pardo
para optar al grado de Doctor

Dirigida por
Dr. Javier Cuevas Maestro
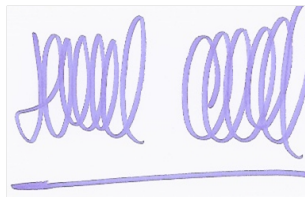y siendo tutor
Dr. Alberto Ruiz Jimeno

Santander, Mayo de 2011

Dr. Javier Cuevas Maestro, profesor titular de Física Atómica del departamento de Física de la Universidad de Oviedo , y **Dr. Alberto Ruiz Jimeno**, Catedrático de Universidad del área de Física Atómica, Molecular y Nuclear de la Facultad de Ciencias de la Universidad de Cantabria.

Certifican:

Que la presente memoria: **Medida de la sección eficaz de producción del quark top y búsqueda del bosón de Higgs del SM en estados finales dileptónicos con el detector CMS del LHC a $\sqrt{s} = 7$ TeV**, ha sido realizada bajo la dirección del Dr. Javier Cuevas, siendo tutor de la misma, el Dr. Alberto Ruiz, en el Departamento de Física Moderna de la Facultad de Ciencias de la Universidad de Cantabria por **Patricia Lobelle Pardo**, para optar al grado de Doctor en Ciencias Físicas.

Y para que así conste, en cumplimiento de la legislación vigente, firmo el presente certificado:

Santander, Mayo de 2011

Contents

1

Introduction

The Standard Model (SM) of Particle Physics describes the behaviour of all known subatomic particles and their interactions. It has successfully explained the experimental data so far, although there are still some remaining questions, as the nature of the mechanism responsible for the electroweak symmetry breaking and the origin of mass.

The Higgs mechanism is a candidate model that provides one of the simplest solutions, it requires the existence of a new scalar particle, the Higgs boson, that has yet to be detected. All the properties of the Higgs boson can be predicted in the SM except its mass, which is a free parameter of the theory.

The search for the SM Higgs boson has been carried out by other experiments and up to now limits has been put on its mass. The data analyzed by the four experiments at the Large Electron Positron collider (LEP) allowed to set a lower limit at 114.4 GeV/c^2 at 95% C.L. [1]. Currently, the Higgs search is also being performed by the CDF and D0 experiments at the Tevatron accelerator. They have reached enough sensitivity, making possible the exclusion of a SM Higgs boson with a mass between approximately 158 and 175 GeV/c^2 [2]. Also precision electroweak measurements constrain the mass to be less than 185 GeV/c^2 at 95% C.L. [3]. A light Higgs boson is preferred from all available experimental and theoretical bounds, and these constraints help to determine the range in which the search at the Large Hadron Collider (LHC) will be mainly focused.

The LHC has been designed to collide protons at a centre of mass energy of $\sqrt{s} = 14$ TeV, an energy seven times higher than the corresponding to the Tevatron. One of the main goals is to reveal the physical mechanism responsible for the electroweak symmetry breaking. If the Higgs mechanism is valid, the LHC experiments have the potential to find the Higgs boson, closing a search that started years ago and in which several experiments at different colliders have participated. The large centre of mass energies produced in proton-proton collisions at the LHC together with the high luminosities already achieved allow to develop a physics program that will give the final answer to this search. The LHC will cover a wide range of Higgs searches, from the minimal SM Higgs, to beyond the SM through the Minimal Supersymmetric SM(MSSM) extensions and other Super Symmetry (SUSY) models.

The top quark, the heaviest of all known elementary particles is considered to be a highly sensitive window for new physics searches. Besides, a better knowledge of the top quark properties such as its mass and the production cross-section will result in an ultimate confinement test of the SM.

The LHC started its operation in March 2010 at a centre of mass energy of $\sqrt{s} = 7$ TeV. SM measurements and in particular the observation of the $t\bar{t}$ production constituted an important part

1. INTRODUCTION

of the physics program as it is also a test of the good understanding of the CMS detector itself and requires the good performance of all its parts, which is needed before entering in the discovery mode.

This thesis focuses on the $t\bar{t}$ cross section measurement and the search for the SM Higgs boson in the WW channel, both in the $2l2\nu$ final state.

The $t\bar{t}$ process constitutes one of the major backgrounds for the searches and its control remains important, besides, the precise knowledge of the top quark mass together with the W mass also contribute to indirect constraints on the Higgs mass.

In proton-proton collisions at the LHC, top pairs are mostly produced in gluon-gluon fusion whereas at the Tevatron proton-antiproton collider it mainly occurs through $q\bar{q}$ annihilation, therefore its study at the LHC contributes to the understanding of the $t\bar{t}$ production mechanism.

In this work, the $t\bar{t}$ cross section measured at a centre of mass energy of 7 TeV in proton-proton collisions is presented. This was already possible with the first 3 pb^{-1} of data collected during the first months of the LHC running, after a long period of preparatory work with MC simulation. It was afterwards extended using the full luminosity collected in 2010, that corresponds to 36 pb^{-1}, allowing to reduce not only the statistical uncertainty, which is the dominant one but also the systematics. The measurement was performed in the dilepton channel, which is a clean final state, thus allowing to observe the signal already at an early stage. It presents two high energetic leptons, large amount of $\not{E}_T$ from the undetected neutrinos and two high energetic jets coming from the hadronization of the two b-quarks, this feature allows to select a more pure sample and it is studied in this thesis. The main backgrounds are those coming from $Z/\gamma^* \rightarrow l^+l^-$ (l=μ, e) production and the processes with fake, non-prompt leptons that are passing the selection, they are difficult to model properly in simulation and are estimated directly from data. Other background sources arise from diboson production, single top and $Z/\gamma^* \rightarrow \tau^+\tau^-$, which are estimated from simulation but their contribution is small.

The search of the Higgs boson with the full data sample collected in 2010 is also performed in this thesis, using the $gg \rightarrow H \rightarrow WW^* \rightarrow 2l2\nu$ final state. The gluon fusion process is the predominant Higgs production mode at the LHC and the WW channel is considered as one of the most sensitive to the Higgs boson search in the intermediate mass range between 120 and 200 GeV/c^2.

This channel has a clear signature characterized by the presence of two high p_T leptons and a significant amount of $\not{E}_T$ due to the undetected neutrinos, as well as little hard jet activity. The principal sources of background come from several SM processes that may have similar final states to those of the signal, mainly $W + jets$, $Z + jets$, $t\bar{t}$ and the almost irreducible WW. Other sources of background are WZ, ZZ and tW processes. Given the presence of neutrinos in the final state, that prevents the reconstruction of a clear Higgs mass peak, and the fact that the production cross-section for the Higgs signal is various orders of magnitude smaller than that of the backgrounds, the search relies on an accurate control of the dominant sources of background.

An efficient reconstruction, identification and isolation criteria is important to select the leptons from W bosons and is the key to reject $W + jets$ events when the W decays into a lepton. The contribution is estimated from data in a similar way as for the $t\bar{t}$ cross section measurement. The hadronic activity present in the $t\bar{t}$ events has to be controlled, and events with hard jets coming from b decays have to be vetoed, b-tagging tools are also used in this analysis to further reduce this background. The proper estimation of the missing E_T in events with no genuine missing E_T is crucial to reduce

the $Z + jets$ background to an acceptable level. The main variable to discriminate the signal from the WW background is the angle between the leptons in the transverse plane to the beam, $\Delta\phi_{ll}$. The leptons emitted by the W bosons coming from the decay of a Higgs scalar are close in the transverse plane due to spin correlations, while the angular separation of the leptons coming from WW decays tends to be larger. This is the most characteristic feature of the signal, and plays a major role in the discovery potential of the Higgs using $H \rightarrow WW^*$ process, especially in the mass range $\sim$160 GeV/c^2.

A multivariate analysis based on boosted decision trees has been developed in this thesis to achieve a better background rejection and to improve the sensitivity of the search.

The work presented here relies on the excellent performance of the CMS detector during 2010. No long after the startup different known SM processes started to become accessible and more detailed studies were done as the recorded luminosity was increasing, an illustration is presented in Figure 1.1, which shows the known resonances observed in the e^+e^- and $\mu^+\mu^-$ invariant mass distributions.
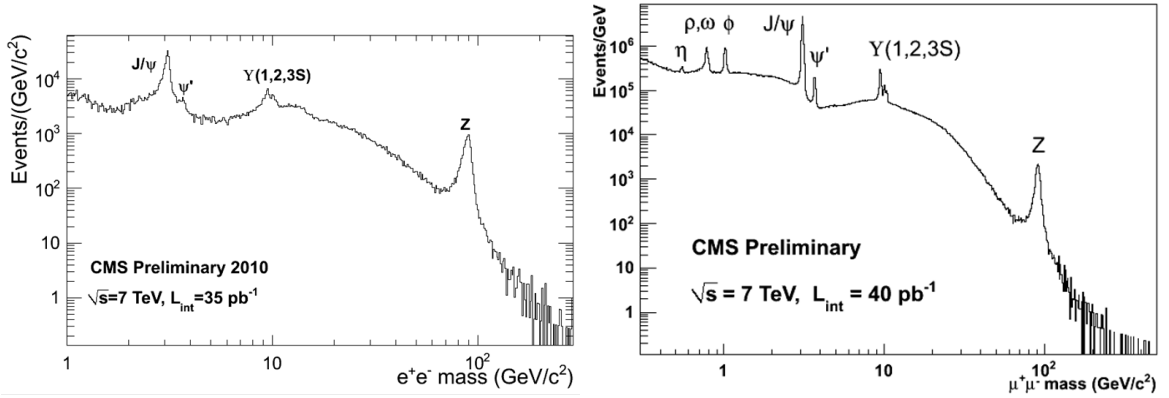
Figure 1.1: - Invariant mass distributions of e^+e^- (left) and $\mu^+\mu^-$ (right) where the known resonances are visible

The work presented in this thesis is organized as follows: In Chapter 2 a short introduction to the SM and the motivation for the Top and Higgs physics studies is given, together with the status of the Higgs searches performed up to now by previous experiments. Then, in Chapter 3 a brief description of LHC and the CMS experiment is presented. In Chapter 4 the main physics observables used in the two analysis and the way they are reconstructed in the CMS detector is described. Chapter 5 presents the $t\bar{t}$ cross section measurement, an introductory summary with the first measurement performed with 3.1 pb^{-1} is discussed and then the detailed study done with the 36 pb^{-1} collected during the 2010 run of the LHC. The description of the event selection, comparison between data and expectations from simulation, lepton efficiency determination, data-driven background estimation techniques and the study of the relevant sources of systematic error is developed. In Chapter 6 the search for the Higgs boson in the WW channel with 36 pb^{-1} is presented. First the specific selection is given, together with the description of the most important variables that help to discriminate the signal against the main backgrounds, then the study performed using multivariate techniques and finally results in terms of exclusion limits are presented as well. Finally, Chapter 7 contains a summary of the main conclusions extracted from the work described in this thesis.

1. INTRODUCTION

2

The Standard Model

2.1 The Standard Model

The Standard Model (SM) is a field theory that combines special relativity and quantum mechanics, compressing the present understanding of the fundamental constituents of the matter and their interactions. This description is the result of decades of worldwide fundamental scientific research based on a fruitful interplay between improved experimental data-taking and theoretical insights. After about 30 years of extensive testing, the SM is one of the best theories of modern physics. In fact, most of its building blocks have been tested up to a very high precision over a large range of energies and an agreement at the per mill level with the theoretical predictions has been found.

In this section only a short introduction and the most fundamental aspects of the SM will be presented, more detailed descriptions can be found for example in [4].

2.1.1 Elementary particles and interactions

The Standard Model (SM)[5; 6; 7] shows a bewitching simplicity when characterizing forces and matter at the most elementary level, relying on Elementary Particles.

Elementary Particles are point-like particles, not known to have sub-structure (up to the present limits of $10^{-18} - 10^{-19}$ m). The SM stands that there are two kinds of elementary particles: the ones that compose matter and the ones that mediate the interactions. Then, matter is composed by a limited set of primary blocks and interactions are described as the exchange of force-carrier particles.

There are two types of elementary particles that conforms matter: leptons and quarks; and each type have three generations. The well-known electron, e^-, as well as its heavier copies, the muon, μ^-, and the tau, τ^-, which are massive and electrically charged, accompanied by their almost massless, electrically neutral neutrinos, ν_e, ν_μ, ν_τ; plus their anti-particles, conform the group of leptons. Quarks (up u, down d, strange s, charm c, bottom b and top t; plus their anti-particles) have mass and fractional electric charge. Both leptons and quarks are fermions, as they have spin $s = 1/2$. An sketched overview of the Fermions can be seen in Table 2.1.

Quarks present an additional quantum number, the color, which can be of three types, generically denoted as q_i, $i = 1, 2, 3$. Color is not seen in Nature, and, because of that, it is not possible to individually observe quarks. However, their color-less bound states, called hadrons, can be observed. Hadrons are classified into baryons and *mesons*. Mesons are bosons made of one quark and one anti-

2. THE STANDARD MODEL

	Charge	First generation		Second generation		Third generation	
Leptons	-1	Electron	e^-	Muon	μ^-	Tau	τ^-
	0	Electron Neutrino	ν_e	Muon Neutrino	ν_μ	Tau Neutrino	ν_τ
Quarks	+2/3	Up	u	Charm	c	Top	t
	-1/3	Down	d	Strange	s	Bottom	b

Table 2.1: Overview of Fermions in the SM.

quark, $q\bar{q'}$, like for instance the pions, π^+, $(u\bar{d})$, and π^-, $(d\bar{u})$. The baryons are fermions composed by three quarks, being the proton, p, (uud), and the neutron, n, (udd), the most familiar examples.

The lightest and most stable particles make up the first generation, while the heavier and less stable particles belong to the second and third generations. All stable matter in the Universe is made from particles that belong to the first generation; as the atomic nucleus is formed by electrons, e^-, neutrons, n (udd), and protons, p (uud); any heavier particles quickly decay to the next most stable level.

The interactions of these 24 types of fermions (12 particles + 12 anti-particles) via three forces are enough to describe the elementary processes in particle physics. These forces are: The Electromagnetic Interaction (Quantum Electrodynamics, QED), corresponding to the classical Electrodynamics described by Maxwell; the Strong interaction, responsible for the atomic nuclei; and the Weak interaction, responsible for the nucleon decays. Other interactions, with special stress in the Gravitation, are also interesting but out of the scope of this thesis.

Interactions are mediated by the exchange of particles. Therefore, the second kind of elementary particles are the ones that intermediate these forces: bosons with spin s = 1.

The photon, γ, mass and charge-less, is exchanged in Electromagnetic interactions; eight gluons, g_α ($\alpha = 1,...8$), also mass and charge-less, but colored, mediate the Strong interactions amongst quarks; and the three weak bosons, the neutral Z^0 and the electrically charged $W^\pm$ mediate the Weak interactions. A small summary of the Bosons is presented in Table 2.2.

Interaction	Particle		Charge	Self-interacting
Electromagnetic	Photon	γ	0	No
Strong	Gluon	g_α $(\alpha = 1,...8)$	0	Yes
Weak		Z^0	0	
		W^-	-1	Yes
		W^+	+1	

Table 2.2: Overview of force-carrier Bosons in the SM.

Theoretically, the SM is a quantum field theory based on the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. This gauge group includes the symmetry group of strong interaction $SU(3)_C$, and the symmetry group of Electroweak interactions $SU(2)_L \times U(1)_Y$, where $U(1)_{em}$ appears as a sub-group; and is in this sense that the weak and electromagnetic interactions are said to be unified.

Hence, the gauge sector of the SM is composed by twelve gauge bosons, which are the force carriers

previously mentioned: the eight gluons g_α ($\alpha = 1, ...8$) which are the gauge bosons of $SU(3)_C$; and the photon, γ, plus the three weak bosons, Z^0, W^+ and W^-, which are the correspondent four gauge bosons of the $SU(2)_L \times U(1)_Y$ group.

Experimentally, it is found that the particles can have mass. The current knowledge of the masses of the fermions is presented in Table 2.3.

First generation		Second generation		Third generation	
Particle	Mass $[GeV/c^2]$	Particle	Mass $[GeV/c^2]$	Particle	Mass $[GeV/c^2]$
e^-	$0.5 \cdot 10^{-3}$	μ^-	0.105	τ^-	1.777
ν_e	$< 10^{-6}$	ν_μ	$< 10^{-6}$	ν_τ	$< 10^{-6}$
u	1.5 to $3.3 \cdot 10^{-3}$	c	1.27	t	171.2
d	3.5 to $6.0 \cdot 10^{-3}$	s	0.104	b	4.20

Table 2.3: Masses of the fermions.

The range of the interactions depends on the inverse of the masses of the gauge bosons that mediate them. For that, the range of the Electromagnetic interaction, mediated by photons ($m_\gamma = 0$), is infinite. A summary on the masses of the boson and the range of the interactions can be seen in 2.4.

Interaction	Particle	Mass $[GeV/c^2]$	Range of the Interaction $[m]$
Electromagnetic	γ	0	Infinite
Strong	g	< 0.20	10^{-15} ($\sim$ nucleus diameter)
Weak	Z^0	91.19	10^{-18}
	W	80.40	

Table 2.4: Masses of the bosons.

Masses and other properties of the different particles can be found in [8].

2.2   Gauge Field Theories

As said above, the SM is described within the context of modern Relativistic Quantum Field Theories[9].

The quantum field theories of the electromagnetic, weak, and strong interactions of the fundamental particles, all belong to the class of local gauge theories.

In short, gauge theories are dynamical theories based on local invariance. The Lagrangian is invariant under a certain continuous group of local transformations called gauge transformations, which conform the symmetry group of the theory. For each symmetry group, there is an associated algebra of generators; and for each group generator there is a corresponding vector field called gauge field. Gauge fields are included in the Lagrangian to ensure the invariance. In the case of quantum field theories, the quanta of the gauge fields are called gauge bosons.

2.2.1 Abelian Gauge Invariance (QED)

Quantum Electrodynamics (QED)[10], the original and most successful gauge field theory, is the perfect example of the application of the gauge principle and its physical implications.

As a representative example of matter we can consider a fermion field ψ with $s = \frac{1}{2}$, mass m, and electric charge Q_e. The corresponding Lagrangian is well known,

$$\mathcal{L}_0 = \bar{\psi}(x)(i\ \slashed{\partial} - m)\psi(x) \ ; \ \ \slashed{\partial} \equiv \partial_\mu \gamma^\mu \tag{2.1}$$

and the Dirac equation gives the corresponding equation of motion:

$$(i\ \slashed{\partial} - m)\psi(x) = 0 \tag{2.2}$$

It is immediate to show the invariance of the Lagrangian to the global $U(1)$ transformations which act on the fields and their derivatives like this:

$$\begin{aligned} \psi &\to e^{-iQ\omega}\psi \\ \partial_\mu \psi &\to e^{-iQ\omega}\partial_\mu \psi \end{aligned} \tag{2.3}$$

$Q\omega$ is the global phase, where ω is the transformation continuous parameter, an arbitrary constant number, independent of x.

By Noether's Theorem [11], this global invariance implies the conservation of the electromagnetic current J_μ; and therefore, the conservation of the electromagnetic charge

$$J_\mu = \bar{\psi}\gamma_\mu eQ\psi \ ; \ \partial_\mu J^\mu = 0 \ ; \ eQ = \int d^3x J_0(x) \tag{2.4}$$

If the transformation is promoted to local, i.e., if the parameter θ is allowed to depend on the space-time point x, the transformation of the fields and their derivatives are like this:

$$\begin{aligned} \psi &\to e^{-iQ\omega(x)}\psi \\ \partial_\mu \psi &\to e^{-iQ\omega(x)}[\partial_\mu \psi(x) - iQ(\partial_\mu \omega(x))\psi] \end{aligned} \tag{2.5}$$

the Lagrangian acquires an extra term that spoils the invariance.

To solve this, a gauge vector boson field $A_\mu(x)$ is introduced; it interacts with the field ψ and transforms properly canceling the extra term

$$A_\mu \to A_\mu - \frac{1}{e}\partial_\mu \omega(x) \tag{2.6}$$

making the total Lagrangian gauge invariant.

The Lagrangian is easily built replacing the derivative ∂_μ by the covariant derivative D_μ; where

$$D_\mu \psi \equiv (\partial_\mu + ieQA_\mu)\psi \tag{2.7}$$

which transforms covariantly, i.e., as the field itself

$$D_\mu \psi \to e^{-iQ\omega(x)}D_\mu \psi \tag{2.8}$$

To include the propagation of the photon field, the kinetic term, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, written in terms of the field strength tensor and also gauge invariant, has to be added; giving as a result a final $U(1)$ gauge-invariant Lagrangian which looks like

$$\mathcal{L}_{QED} = \bar{\psi}(x)(i\,\slashed{D} - m)\psi(x) - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \tag{2.9}$$

2.9 is the QED Lagrangian, and it contains the interactions in the term $\bar{\psi}i\,\slashed{D}\psi$

$$\bar{\psi}ieQA_\mu\gamma^\mu\psi \tag{2.10}$$

The gauge group for electromagnetism is the $U(1)_{em}$, with one generator, Q and one parameter θ. It has one gauge field, the Electromagnetic field, with the photon, γ, being the gauge boson. As the local gauge symmetry forbids mass terms for the gauge-field, the photons have to be mass-less to preserve the symmetry.

2.2.2   Gauge Theory of Strong Interactions (QCD)

Quantum Chromodynamics (QCD) is the gauge theory for strong interactions, and is based on the local color transformations which leave its Lagrangian invariant. The gauge symmetry group that is generated by these color transformation is the non-abelian Lie group $SU(3)_C$. Where C refers to colors and 3 refers to the three possible color states of the quarks.

The building of the QCD Lagrangian is done by following the same steps as in the QED case. The gauge principle is applied taking into account the particularities of the non-abelian group $SU(3)$. The global symmetry is promoted to local by replacing the derivative of the quark by its covariant derivative, which, in the QCD case is like this:

$$D_\mu q = (\partial_\mu - ig_s\frac{\lambda_\alpha}{2}A_\mu^\alpha)q \tag{2.11}$$

Where $q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}$, q_i are the quark fields, g_S is the strong coupling constant, $\frac{\lambda_\alpha}{2}$ are the $SU(3)$ generators and A_μ^α are the gluon fields.

The QCD Lagrangian is then written

$$\mathcal{L}_{QCD} = \sum_q \bar{q}(x)(i\,\slashed{D} - m_q)q(x)\frac{1}{4}F_{\mu\nu}^\alpha(x)F_\alpha^{\mu\nu}(x) \tag{2.12}$$

The gluon field strength is then

$$F_{\mu\nu}^\alpha(x) = \partial_\mu A_\nu^\alpha(x) - \partial\nu A_\mu^\alpha(x) + g_S f^{\alpha\beta\gamma}A_{\mu\beta}A_{\nu\gamma} \tag{2.13}$$

it contains a bilinear term as it correspond to a non-abelian gauge theory with structure constants $f^{\alpha\beta\gamma}$ with $\alpha, \beta, \gamma = 1, ..., 8$.

Similarly to QED, the gauge interactions among the quarks and gluons are contained in the $\bar{q}i\,\slashed{D}q$ term. There is, however, an important difference with the QED case. The gluon kinetic term $F_{\mu\nu}^\alpha F_\alpha^{\mu\nu}$ contains a three gluons term and a four gluons term. These are precisely, the self interaction gluon vertices, genuine of a non-abelian theory.

2.2.3   The Gauge Theory of Electroweak Interactions

The electroweak theory joins electromagnetism with the weak force in a single relativistic quantum field theory. As it was stated at the beginning of this chapter, Electromagnetism is a force of infinite range, while the influence of the charged-current weak interaction only spans distances shorter than about 10^{-15} cm. Two interactions so different in range and apparent strength are described to a common gauge symmetry, namely, the symmetry $SU(2)_L \times U(1)_Y$ which is required to be a local symmetry of the electroweak Lagrangian.

$SU(2)_L$ is the weak isospin group, which acts on left-handed fermions, and $U(1)_Y$ is the weak hypercharge group. The $SU(2)_L \times U(1)_Y$ group has four generators, three of which are the $SU(2)_L$ generators, $T_i = \frac{\sigma_i}{2}$, with $i = 1, 2, 3$, and the fourth one is the $U(1)_Y$ generator, $\frac{Y}{2}$. The conmutation relations for the total group are:

$$[T_i, T_j] = i\epsilon_{ijk}T_k;\ [T_i, Y] = 0;\ i, j, k = 1, 2, 3 \tag{2.14}$$

The left-handed fermions transform as doublets under $SU(2)_L$

$$f_L \rightarrow e^{i\overrightarrow{T}\overrightarrow{\theta}} f_L;\ f_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \begin{pmatrix} u_L \\ d_L \end{pmatrix}, ... \tag{2.15}$$

While the right-handed fermions transforms as singlets,

$$f_L \rightarrow f_R;\ f_R = e_R, u_R, d_R, ... \tag{2.16}$$

The relation between the quantum numbers of the fermions, $Q = T_3 + \frac{Y}{2}$, is also incorporated to the SM. The number of associated gauge bosons, being equal to the number of generators, is four; three weak bosons of $SU(2)_L$, W_μ^i, with $i = 1, 2, 3$, and the hypercharge boson of $U(1)_Y$, B_μ.

The building of the SM Lagrangian is done by following the same steps as in any gauge theory. In particular, the $SU(2)_L \times SU(1)_Y$ symmetry is promoted from global to local by replacing the derivatives of the fields by the corresponding covariant derivatives. For a generic fermion field f this covariant derivative is:

$$D_\mu f = (\partial_\mu - ig\overrightarrow{T}\overrightarrow{W_\mu} - ig'\frac{Y}{2}B_\mu)f \tag{2.17}$$

where, g is the coupling constant corresponding to $SU(2)_L$ and g' is the one corresponding to $U(1)_Y$. For example, the covariant derivative for a left-handed and right-handed electron are respectively:

$$D_\mu e_L = (\partial_\mu - ig\frac{\overrightarrow{\sigma}}{2}\overrightarrow{W_\mu} - ig'\frac{1}{2}B_\mu)e_L;\ D_\mu e_R = (\partial_\mu + ig'B_\mu)e_R \tag{2.18}$$

As in the previous cases of QED and QCD, the gauge invariant electroweak interactions are generated from the $\bar{f}i\ \slashed{D}f$ term.

2.2.4   The Standard Model Lagrangian

To get the Lagrangian of the Electroweak Theory it is necessary to add the previous fermion terms, the gauge boson kinetic terms and the gauge boson self-interacting terms.

$$\mathcal{L}_{SM} = \mathcal{L}_f + \mathcal{L}_G + \mathcal{L}_{SBS} + \mathcal{L}_{YW} \tag{2.19}$$

where the fermion Lagrangian is

$$\mathcal{L}_f = \sum_{l=l,q} \bar{f}i\ \slashed{D}f \tag{2.20}$$

and the Lagrangian for the gauge field is

$$\mathcal{L}_G = -\frac{1}{4}W^i_{\mu\nu}W^{\mu\nu}_i - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \mathcal{L}_{GF} + \mathcal{L}_{FP} \tag{2.21}$$

$\mathcal{L}_{GF}$ and $\mathcal{L}_{FP}$ are the gauge fixing and Faddeev Popov Lagrangians, needed in any gauge theory, previously omitted for QED and QCD; $\mathcal{L}_{SBS}$ and $\mathcal{L}_{YW}$ are the Symmetry Breaking Sector Lagrangian and the Yukawa Lagrangian, needed to provide masses to the weak bosons, m_W and m_Z and the fermions m_f.

The resulting Lagrangian contains the wanted self-interaction terms among the three electroweak bosons, and it can be shown that this is invariant under the $SU(2)_L \times U(1)_Y$ gauge transformations

The gauge bosons, $W^\pm_\mu$, Z_μ and A_μ, are obtained from the electroweak interaction eigenstates

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu) \\ Z_\mu &= c_\omega W^3_\mu - s_\omega B_\mu \\ A_\mu &= s_\omega W^3_\mu + c_\omega B_\mu \end{aligned} \tag{2.22}$$

Note that the mass terms $m^2_W W_\mu W^\mu$, $\frac{1}{2}m^2_Z Z_\mu Z^\mu$ and $m_f \bar{f}f$ are forbidden by $SU(2)_L \times U(1)_Y$ gauge invariance. The needed gauge boson masses have to be generated in a gauge invariant way. The spontaneous breaking of the symmetry via the Higgs Mechanism can provide the mass generation.

2.3 The Higgs Mechanism

In the previous section, it has been shown that, to preserve the invariance of the Lagrangian, the mass terms of the gauge bosons are explicitly forbidden; but this is contradicted by the experimental evidence, as the masses of the weak bosons, $W^\pm$ and Z^0 has been experimentally measured, and it is well-known that they are very heavy.

Hence, the Symmetry has to be broken. The Spontaneous Symmetry Breaking (SSB) is one of the key ingredients of the electroweak interactions. In Quantum Field Theory, a system is said to posses a symmetry that is spontaneously broken when the Lagrangian is invariant under the symmetry transformations, while the vacuum of the theory is not. The vacuum is the state in which the Hamiltonian expectation value is minimum.

In the context of the SM, when the SSB refers to a gauge symmetry, the Higgs Mechanism[12; 13; 14; 15] has to operate. The symmetry-breaking is realized linearly by a scalar field which acquires a non-zero vacuum expectation value. As a result, not only the masses of the vector-bosons and the fermions arise, but also de Higgs particle, a neutral scalar field which was not experimentally observed so far.

Considering the Higgs Mechanism as a responsible for the symmetry breaking in the SM, a consistent formulation exist, and any observable can be calculated pertubatively in the Higgs self-coupling constant. However, the concept of symmetry breaking is more general, and even when the Higgs mechanism is the simplest one, there are countless alternatives to replace the standard Higgs sector, following just three basic requirements:

2. THE STANDARD MODEL

1. Electromagnetism has to remain unbroken

2. The full symmetry has to contain the electroweak gauge symmetry

3. The symmetry breaking has to occur at about the energy scale $v = (\sqrt{2}G_F)^{-\frac{1}{2}} = 256$ GeV, being G_F the Fermi coupling constant.

Given that, the simplest choice for the SSB of the Electroweak theory in the SM is the Higgs Mechanism, in which an additional scalar field Φ is introduced:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^- \end{pmatrix} \tag{2.23}$$

Then, the Lagrangian of the SSB of the Electroweak theory will be:

$$\mathcal{L}_S BS = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \tag{2.24}$$

the covariant derivative is then:

$$D_\mu = (\partial_\mu - \frac{1}{2} ig \vec{\sigma} \overrightarrow{W_\mu} - \frac{1}{2} ig' B_\mu) \tag{2.25}$$

Here, Φ is a fundamental complex $SU(2)$ doublet with hypercharge $Y(\Phi) = 1$ and $V(\Phi)$ is the simplest renormalizable potential:

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \; ; \; \lambda > 0 \tag{2.26}$$

Depending on the mass parameter, $-\mu^2$, there are two possibilities that minimize the potential $V(\Phi)$, Figure 2.1:
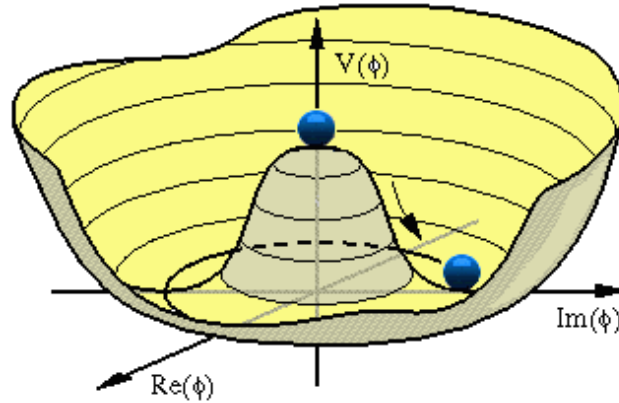
Figure 2.1: - Representation of the Higgs potential.

- If $(-\mu^2) > 0$, the minimum is at $< \Phi >= 0$. The vacuum is $SU(2)_L \times U(1)_Y$ symmetric, and therefore, no symmetry breaking occurs.

- If $(-\mu^2) < 0$, the minimum is at $< \Phi >= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$, for any arbitrary argument of Φ. Therefore, there are infinite degenerate vacuum choices, either of them $SU(2)_L \times U(1)_Y$ non symmetric and $U(1)_{em}$ symmetric. The simplest one, where $arg \; \Phi \equiv 0$, is taken.

$$< \Phi >= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \; ; \; v = \sqrt{\frac{\mu^2}{\lambda}} \tag{2.27}$$

2.3.1 The SM Lagrangian with Higgs

To get the particles spectra and the particle masses; the $SU(2)_L \times U(1)_Y$ gauge invariant Lagrangian should be written:

$$\mathcal{L}_{SM} = \mathcal{L}_f + \mathcal{L}_G + \mathcal{L}_{SSB} + \mathcal{L}_{YW} \tag{2.28}$$

$\mathcal{L}_f$ and $\mathcal{L}_G$ have been given previously, and $\mathcal{L}_{SSB}$ and $\mathcal{L}_{YW}$ are the SSB and the Yukawa Lagrangians respectively,

$$\mathcal{L}_{SSB} = (D_\mu \Phi)^\dagger (D_\mu \Phi) + \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 \tag{2.29}$$

$$\mathcal{L}_{YW} = \lambda_e \bar{l}_L \Phi_{eR} + \lambda_u \bar{q}_L \widetilde{\Phi} u_R + \lambda_d \bar{q}_L \Phi d_R + h.c. + 2^{nd} \text{ and } 3^{rd} \text{ families} \tag{2.30}$$

Where:

$$l_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} ; q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix} ; \Phi = \begin{pmatrix} \Phi^+ \\ \Phi^- \end{pmatrix} ; \widetilde{\Phi} = i\sigma_2 \Phi^* = \begin{pmatrix} \Phi_0^* \\ -\Phi^- \end{pmatrix}$$

$\mathcal{L}_{SSB}$ provides m_W and m_Z; $\mathcal{L}_f$ is needed to provide m_f, the masses of the fermions.

Then, in order to get the spectrum for $\mathcal{L}_{SM}$, a non-symmetric vacuum is fixed, $\langle \Phi \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}$. The physical spectrum is built performing small oscillations around the vacuum, parametrized by

$$\Phi(x) = e^{i\frac{\overrightarrow{\xi}(x)\overrightarrow{\sigma}}{v}} \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix} \tag{2.31}$$

noting that $\overrightarrow{\xi}(x)$ and $H(x)$ are small fields.

Some gauge transformations are made to eliminate the un-physical fields $\overrightarrow{\xi}(x)$. Then, the weak eigenstates are rotated to the mass eigenstates which define the gauge boson fields 2.22. Once that is done, is possible to read the masses from the terms of the SM Lagrangian, and get the three level predictions for the masses of the gauge bosons:

$$m_W = \frac{gv}{2} ; m_Z = \frac{\sqrt{g^2 + g'^2}v}{2} ; m_H = \sqrt{2}\mu \tag{2.32}$$

And for the fermions:

$$m_e = \lambda_e \frac{v}{\sqrt{2}} ; m_u = \lambda_u \frac{v}{\sqrt{2}} ; m_d = \lambda_d \frac{v}{\sqrt{2}} ;... \tag{2.33}$$

Finally, it is possible to re-write $\mathcal{L}_{SSB}$ and $\mathcal{L}_{YW}$, after the application of the Higgs Mechanism, in terms of the physical scalar fields, and get, not only the masses, but also the kinetic and interaction terms for the Higgs sector.

With some calculations all masses can be given in terms of a unique mass parameter v and the coupling constants: g, g', λ, λ_e, and so on. The interactions of H with fermions and gauge bosons are proportional to the gauge couplings and to the masses of the particles. The parameter v can be measured from μ-decay, identifying the predictions of the partial width $\Gamma(\mu \to \nu_\mu \bar{\nu}_e e)$ to low energies ($q^2 << m_W^2$) getting to:

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2} = \frac{1}{2v^2} ; v = (\sqrt{2}G_F)^{-\frac{1}{2}} = 246 GeV \tag{2.34}$$

The values of the masses of the $W^{\pm}$ and Z^0 were successfully anticipated before being measured experimentally, being the input parameters: the fine structure constant α, G_F, and θ_w[1]; the best measured electroweak parameters before LEP.

It is interesting to point out that the Higgs boson mass m_H and the Higgs self-coupling constant λ are undetermined by the SM, being related at tree level by $\lambda = \frac{m_H^2}{2v^2}$. The hierarchy in the fermion masses is also undetermined by the SM.

2.4   The search for the SM Higgs Boson

The introduction of the Higgs Mechanism in the SM gives a very elegant answer to the practical needs of the theory, and goes in an excellent agreement with the experimental evidence. Bosons and fermions acquire mass that can be calculated, and can also be experimentally measured. With every piece of Nature's subatomic puzzle getting in its place, something is still currently missing: the Higgs Boson.

The existence of the Higgs Boson has to be proved or excluded. The search strategy has to be precisely designed considering the Higgs boson production mode and its decay products, and that depends strictly on its mass.

Unfortunately the Higgs mass, m_H, is not predicted in the SM. The Higgs mechanism provides m_H as a function of the Higgs self-coupling λ and $v = 246\ GeV$ in this way:

$$m_H = \sqrt{2}\mu = \sqrt{2v^2\lambda} \tag{2.35}$$

As λ is also unknown m_H can take any value in the SM. The unique restrictions come from the consistence of the theory and the experimental observation.

The understanding of the mechanism responsible for the electroweak symmetry breaking and the search for the Higgs boson has been one of the major interests in particle physics over the past years and is the goal and one of the main tasks of the LHC program and CMS detector.

Previously to the LHC running, searches were carried out by other experiments, at LEP and at the Tevatron. The LEP program ended in 2001 without conclusive evidence for the Higgs boson but set a lower bound on its mass at 114.4 GeV. Since then, the Tevatron experiments are continuing the search but more data is still needed.

Also theoretical constraints can be derived from consistency conditions on the SM and indirect limits can be inferred from high precision electroweak data [3], constraining the mass of the Higgs boson via their sensitivity to loop corrections, which is expected to be lower than $185\ GeV/c^2$ if LEP results are also considered in the fit. All the available measurements favour a light Higgs boson and the LHC, being the highest energy accelerator ever built, have the potential to discover or definitively exclude its existence.

In this section, a brief description of the current knowledge about the Higgs boson from other experiments is given first, followed by the description of the Higgs phenomenology and search channels specific to the LHC.

---

[1]θ_w denotes the Weinberg mixing angle, $\tan\theta_w = \frac{g'}{g}$; which relates the masses of the weak bosons by $m_W = \cos\theta_w m_Z$.

2.4.1 Theoretical constraints on m_H

Although the Higgs mass cannot be predicted in the SM, there are several constraints deduced from the theory, imposing the known as unitarity, triviality and vacuum stability conditions. An upper limit on the Higgs boson mass can be given from the triviality and the unitarity bound, while the vacuum stability sets a lower limit.

Upper bounds can be derived from the requirement that the SM can be extended up to a scale Λ, before perturbation theory breaks down and new non-perturbative phenomena dominate the predictions of the theory. Unitarity provides an upper bound on m_H above which the standard Model is known to become non-perturbative. The Unitarity bound is related to the fact that the scattering probability can not be higher than one. It is possible to find the unitarity conditions for particular scattering cases, like for example, $W_L^+ W_L^-$ SM scattering, giving as a result different upper bounds on the Higgs mass, that combined, give a rough.

$$m_H < 700 - 800 GeV/c^2 \tag{2.36}$$

These bounds based on perturbative unitarity do not necessarily mean that the Higgs boson can't be heavier than these values.

In the case of large m_H, the perturbative approach is not valid anymore and non-perturbative techniques are required. For this heavy Higgs models, the Higgs self-interactions governed by the coupling λ become strong and new physics phenomena may appear at energies around 1 TeV. Hence, the existence of a non-perturbative regime for the scalar sector of the SM is still a possibility and this interesting phenomena can also be studied at the LHC.

Then, the Higgs self-coupling has to be finite, and another upper bound can derived considering Triviality. That means all the couplings being small enough for energies bellow some threshold. For example, if the energy scale is fixed to 10^{16} GeV, considering the *top* quark mass to be $m_t = 170$ GeV/c^2, $m_H < 170 GeV/c^2$ is derived following:

$$m_H^2 < \frac{4\pi^2 \nu^2}{3 \ln(\frac{\Lambda}{\nu})} \tag{2.37}$$

being Λ the cutoff scale.

A lower bound on the Higgs mass can be obtained from Vacuum Stability. The asymmetric vacuum of the $SU(2)_L \times U(1)_Y$ has to be stable under quantum corrections. The limit is resumed in this equation:

$$m_H^2 > \frac{4m_t^4}{\pi^2 \nu^2} \ln(\frac{\Lambda}{\nu}) \tag{2.38}$$

Recent studies requiring vacuum stability at two-loop level up to energies of the order of 10^{16} GeV, considering $m_t = 170$ GeV/c^2 and $\alpha_S = 0.117$ give a exclusion limit of $m_H > 132$ GeV/c^2.

To summarize, requiring the Higgs coupling to remain finite and positive up to an energy scale Λ, constraints can be derived on the Higgs mass m_H, such analysis exist at the two-loop level for both lower and upper Higgs mass bounds, and can be plotted as a function of the cutoff energy Λ. Taking the top quark mass to be 175 GeV/c^2 and a QCD coupling $\alpha_s(m_Z) = 0.118$ the result is shown in 2.2.
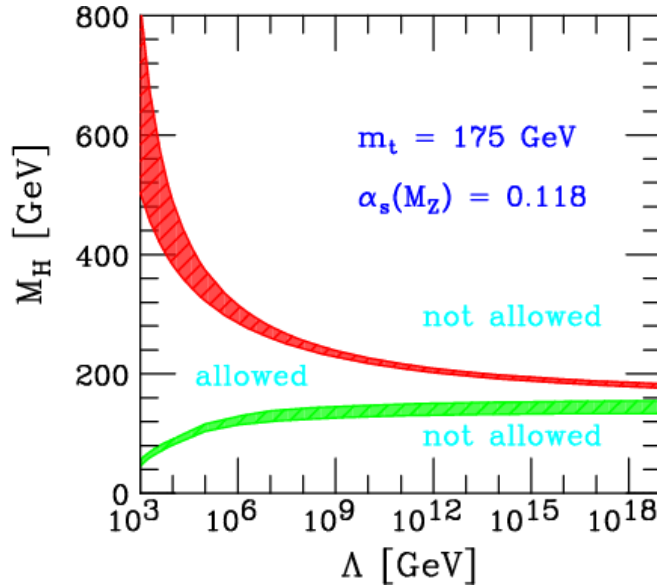
Figure 2.2: - Lower and upper limits on the Higgs boson mass as a function of the cut-off energy scale Λ, relying on the vacuum stability and triviality arguments.

2.4.2 Higgs searches at LEP

The Large Electron Positron Collider (LEP) at CERN was running from 1989 to 2000. Searches on the Higgs boson of the SM and beyond constituted an important part of the physics program, and was carried out by its four detectors ALEPH, DELPHI, L3 and OPAL [1]. There were two running periods, the first called LEP-I (1989-1995) and a second phase with increased centre of mass energy (1996-2000), called LEP-II.

At LEP-I the accelerator operated at energies close to the Z boson resonance. Because of the large production cross section for a low-mass Higgs boson in Z decays, LEP provided a very good environment to exclude small values of m_H. The dominant production mechanism was the Bjorken process, $e^+e^- \rightarrow HZ^* \rightarrow Hf\bar{f}$; with minor contributions from other processes, like the Wilczek process $e^+e^- \rightarrow H\gamma$ which suffered from larger background contamination.

Concerning the accessible decays, in the non-perturbative QCD domain ($m_H < 2\ GeV/c^2$), for very small Higgs boson masses of $m_H < 2m_e$, the Higgs boson can only decay to a pair of photons via a loop of W bosons. For masses below $2m_\mu$ it mostly decays to a pair of electrons and between $2m_\mu$ and $2m_\pi$, to a pair of muons. Above that threshold, the situation becomes slightly more difficult. For masses below 2 3 GeV/c^2, the Higgs boson decays to a pair of hadrons via its interaction with two gluons through a top-quark loop or its interaction with quarks. The hadronization of these gluons gets more complex with the mass of the Higgs boson.

The transition to the perturbative QCD domain comes above $\backsim 2\ GeV/c^2$. In this region, other decay modes start to gain importance. Bellow the $b\bar{b}$ threshold, the decays into the heaviest available fermion pair ($c\bar{c}$ or $\tau^+\tau^-$) prevail; above the $b\bar{b}$ threshold and for Higgs boson masses reachable at LEP (below $\sim 115\ GeV/c^2$), the dominant decay is $H \rightarrow b\bar{b}$ decay.

At LEP-I it was possible to achieve more than a 95% C.L. exclusion for a SM Higgs boson with a mass bellow 20 GeV/c^2 using three topologies of the dominant $e^+e^- \rightarrow Z^* \rightarrow ZH$ production mode,

one of them being background free, and the other, slightly affected by the $e^+e^- \to \gamma^* Z$ background. For higher masses, as the $b\bar{b}$ was the dominant Higgs decay, the huge amount of background coming from the hadronic decays of the Z and the small production cross-section for Higgs, make the study of Z^* decays involving hadrons or $\tau^{\pm}$ not suitable, hence the only studied channels were the decays to a pair of neutrinos or charged leptons (electrons or mons). A very small number of signal events was expected, and that, combined with the large Z decays collected, made necessary the use of more sophisticated techniques. The result of these studies led to a lower limit on the Higgs boson mass of 65.6 GeV/c^2.

The second period of running from 1996 is known as LEP II, increasing the centre of mass energy up to 206.6 GeV.The dominant production process has the direct coupling of the Higgs to W or Z vector-bosons. The kinematic threshold for Higgs boson production quickly decreased around $\sqrt{s} - m_Z \sim 10\ GeV$.Higgs physics at LEP were strongly dependent on the luminosity and cross-section.

Combining the final results from the four LEP experiments, ALEPH, DELPHI, L3 and OPAL, a lower bound of 114.4 GeV/c^2 is set on the mass of the SM Higgs boson at the 95 % confidence level (see Figure 2.3).

Towards the end of the LEP running in the year 2000, hints for the direct observation of Higgs signal, corresponding to a Higgs-boson mass around 116 GeV/c^2, were detected; a mass value well compatible with the constraints derived from the precision electroweak measurements. At the LHC, this mass will be investigated again, as well as a very wide mass range, covering both, the excluded and not yet excluded masses, that will lead to the definitive discover of exclusion of the SM Higgs boson.
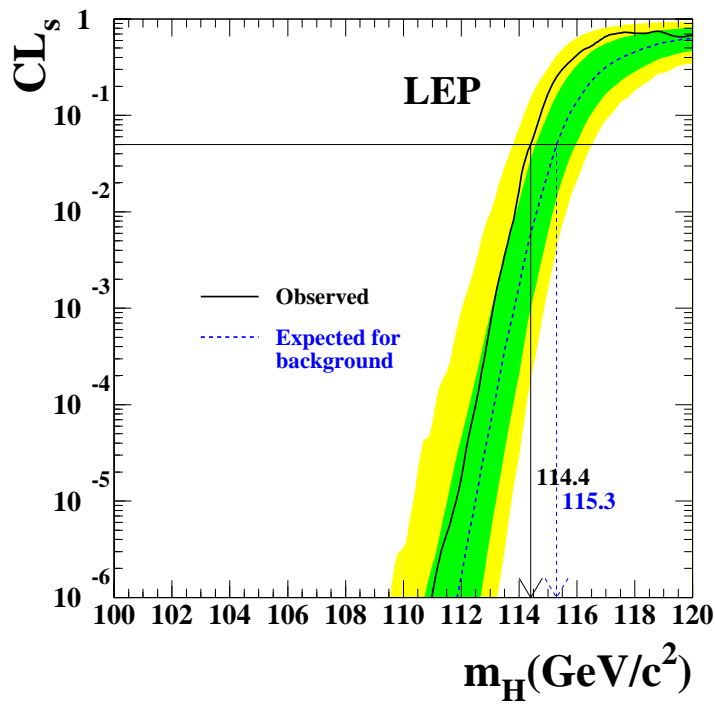
Figure 2.3: - The ratio $CL_s = CL_{s+b}/CL_b$ for the signal plus background hypothesis. Solid line: observation; dashed line: median background expectation. The dark and light shaded bands around the median expected line correspond to the 68% and 95% probability bands. The intersection of the horizontal line for $CL_s = 0.05$ with the observed curve is used to define the 95% confidence level lower bound on the mass of the SM Higgs boson. [1]

2.4.3 Higgs searches at the Tevatron

The Tevatron $p\bar{p}$ collider at the Fermi National Accelerator Laboratory has been running since 1992. In the first period, Run I, (1992-1996) operated at a centre of mass energy of 1.8 TeV and in the current run, called Run II, the energy has been raised to 1.96 TeV.

The Tevatron experiments CDF and D0 are also performing SM Higgs physics. While at LEP, the discover of the Higgs boson could have been achieved for certain masses, the Tevatron can only reach exclusion for some particular mass values. It was not designed specifically for SM Higgs physics, but for top quark physics. During all these years of successful running not only the top quark was discovered [16; 17] but the amount of recorded collision data has been enough to exclude a Higgs boson mass range around 160 GeV/c^2 combining different channels and taking advantage of the on-going development of multidimensional analysis.

The main production modes for a SM Higgs boson at the Tevatron are through the gluon-gluon fusion process and the HV (V = W or Z) associated production mechanism, and the main decay modes are the leptonic decays of the vector-bosons with H going to $b\bar{b}$, although many other signatures are being studied.

For a low mass Higgs boson ($m_H \leq 130\ GeV/c^2$) HV associated production with the Higgs decaying to $b\bar{b}$ is the optimal search channel, while for higher masses gg fusion with the Higgs decaying to WW^* (with subsequent decay to $2l2\nu$) is used. For intermediate masses HV production with $H \rightarrow WW^* \rightarrow 2l2\nu$ is also sensitive.

The last available combined CDF and D0 upper limits on SM Higgs boson production is shown in 2.4, reaching a 95% C.L. exclusion for SM Higgs in the region $158 < m_H < 173\ GeV/c^2$ [2]. The analysis combined here seek signals of Higgs bosons produced in association with vector bosons ($q\bar{q} \rightarrow$ W/ZH), through gluon-gluon fusion (gg $\rightarrow$ H), and through vector boson fusion (VBF) ($q\bar{q} \rightarrow q\bar{q}$H) corresponding to integrated luminosities up to 7.1 fb^{-1} at CDF and up to 8.2 fb^{-1} at D0.

2.4.4 Constraints from EWK precision data fits

A global fit of the data from the Z-pole experiments (ALEPH, DELPHI, L3, OPAL, SLD), from $nu - N$ deep inelastic scattering experiments (NuTeV, CHARM, CCFR, CDHS), from low energy $\nu - e$ scattering experiments (CHARM II) from atomic parity violation experiments (on cesium and thallium), and from top quark production (CDF, D0) provides very strong constraints on the range of m_H under the assumption that there is no physics other than that of the SM near to the weak scale. Assuming no physics beyond the SM, a value of $m_H = (98^{+57}_{-38})\ GeV/c^2$ is obtained.

The comparison between the indirect constraints and the direct measurements of m_t and m_W in the (m_t,m_W) plane is shown in Figure 2.5. The observed agreement is a crucial test of the SM. Since the SM is so successful in predicting the values of m_W and m_t, this type of analysis is now extended to predict the mass of the Higgs boson. As seen in the figure, both contours prefer low values for the mass of the Higgs boson.

Figure 2.5 shows the $\Delta\chi^2$ curve derived from high-Q^2 precision electroweak measurements, performed at LEP and by SLD, CDF, and D0, as a function of the Higgs-boson mass, assuming the SM to be the correct theory of nature. The preferred value for its mass, corresponding to the minimum of the curve, is at 89 GeV/c^2, with an experimental uncertainty of +35 and −26 GeV/c^2, at 68% confidence level derived from $\Delta\chi^2 = 1$ for the black line, thus not taking the theoretical uncertainty shown as the
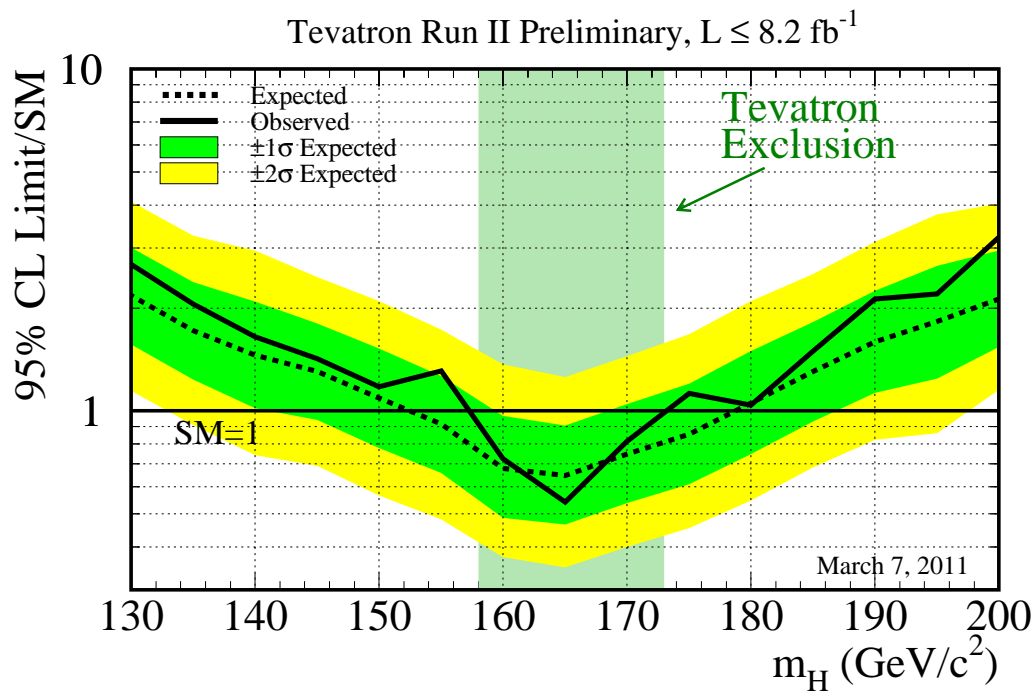
Figure 2.4: - Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on the ratios to the SM cross section, as functions of the Higgs boson mass for the combined CDF and D0 analysis. The limits are expressed as a multiple of the SM prediction for test masses (every 5 GeV/c^2 for which both experiments have performed dedicated searches in different channels. The points are joined by straight lines for better readability. 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 [2].

blue band into account. Precision electroweak measurements tell that the mass of the Standard-Model Higgs boson is lower than about 158 GeV/c^2 (one-sided 95% confidence level upper limit derived from $\Delta\chi^2 = 2.7$ for the blue band, thus including both the experimental and the theoretical uncertainty). The limit increases to 185 GeV/c^2 when including the LEP-2 direct search limit of 114 GeV/c^2 shown in yellow. These combinations are performed by the LEP Electroweak Working Group and updated results are collected in LEPEWWG.
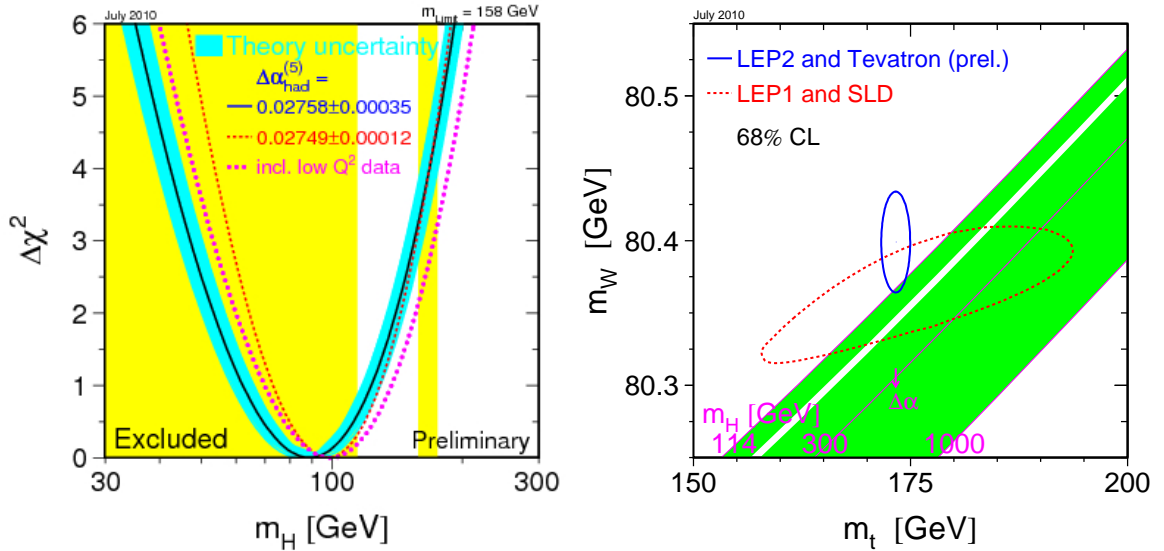
Figure 2.5: - Left: $\Delta\chi^2$ curve derived from high-Q^2 precision electroweak measurements, performed at LEP and by SLD, CDF, and D0, as a function of the Higgs-boson mass, assuming the SM to be the correct theory of nature. Right: Comparison of the indirect constraints on m_W and m_t based on LEP-I/SLD data (dashed contour) and the direct measurements from the LEP-II/Tevatron experiments (solid contour). In both cases the 68% CL contours are plotted. Also shown is the SM relationship for the masses as a function of the Higgs mass in the region favoured by theory ($< 1000\ GeV^2$) and not excluded by direct searches (114 GeV/c^2 to 158 GeV^2 and $> 175\ GeV^2$). The arrow labelled $\Delta\alpha$ shows the variation of this relation if $\alpha(m_Z^2)$ is varied between -1 and +1 standard deviation. This variation gives an additional uncertainty to the SM band shown in the figure.

2.5 Higgs searches at the LHC

The combination of experimental and theoretical bounds on m_H just discussed, set the pace for the upcoming Higgs physics studies, providing the mass range in which to search for the Higgs boson at the LHC.

Higgs physics at the LHC are different than in previous colliders due mainly to the larger centre of mass energy considered. The design collision energy for the LHC is 14 TeV, while the second most energetic accelerator built, the Tevatron, has an energy of about 2 TeV. Particle production in the LHC is in a mass range from $10^2\ GeV/c^2$ and above and in a Feynman x region where the gluon density is much larger than the quark density. Therefore for Higgs and top physics it is useful to think of the LHC as a gluon collider. Also the high multijet environment of the LHC constitutes a challenge and will determine the channels feasible to carry out the analysis.

Figure 2.6 shows the production cross-sections for different processes, Higgs included, against the centre of mass energy.
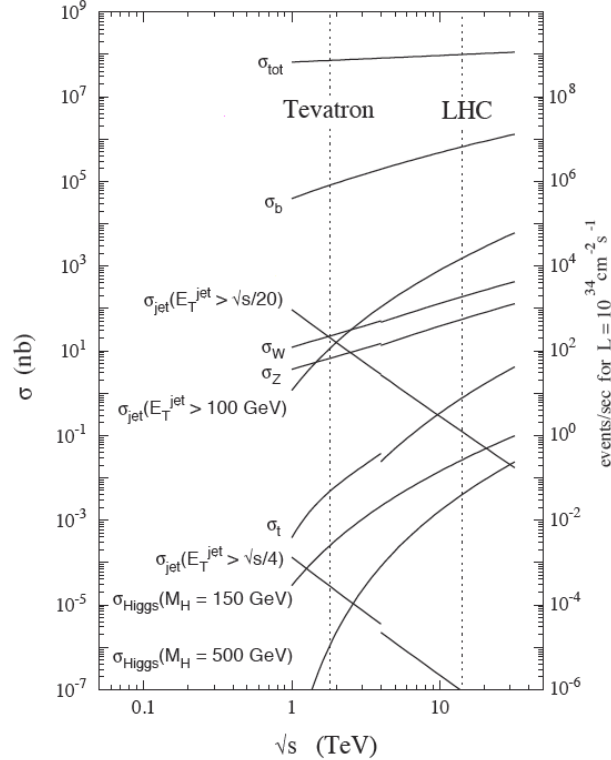
Figure 2.6: - Production cross-sections for different events at the Tevatron and at the LHC as a function of centre of mass energy.

A description of the main production modes and search channels will be given in the following.

2.5.1 Higgs boson production

The Higgs boson can be produced at the LHC through various processes depending on its mass, an illustration of the feynman diagrams can be seen in Figure 2.7. The Higgs production cross-section at $\sqrt{s} = 7$ TeV as a function of the Higgs boson mass is shown in Figure 2.9, where one can see as different production modes are opening up as the mass increases.

The dominant Higgs production mechanism at the LHC is the gluon-gluon fusion process [18], which has the largest production cross-section for the whole mass range. The process $gg \rightarrow h + X$ is an essentially strong-interaction process, that has attracted a large amount of theoretical work in recent years. Limits on the Higgs mass will rely upon QCD calculations of the cross-sections, and conversely, if a Higgs boson is discovered, discrepancies of its measured cross-section from QCD calculations may signal deviations of the Yukawa couplings from the SM predictions [19].

The second most important production mode is the vector boson fusion (VBF) $qq \rightarrow qqh + X$, the inelastic scattering of two quarks (antiquarks), mediated by t-channel W or Z exchange, with the Higgs boson radiated off the weak bosons. In this case, the Higgs boson is produced via the fusion of the weak bosons W and Z, and is accompanied by two forward jets that carry large transverse
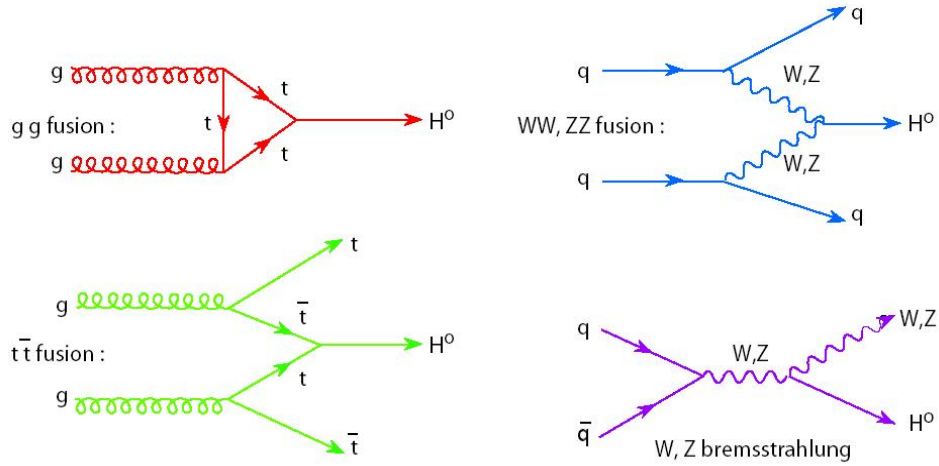
Figure 2.7: - Higgs production modes at hadron colliders: Gluon fusion (a), Vector Boson fusion (b), Associate Production with a W boson (c) and Associate production with a top pair (d).

momentum in the final state. This mode becomes competitive for larger Higgs boson masses and being also relevant in the intermediate mass range since the characteristic forward jets provide an additional signature to tag the event and hence reduce the background significantly.

For lower masses $m_H < 2m_Z$, the W,Z associated production $qq \rightarrow Wh + X$, $qq \rightarrow Zh + X$, also called Higgsstrahlung, provides alternative signatures for the Higgs boson search. Also the bottom fusion $bb \rightarrow h + X$ could be studied in this mass range.

The $t\bar{t}$ associated production $qq, gg \rightarrow tth + X$ at the LHC is an important search channel for Higgs masses below $\approx 125 \ GeV/c^2$. The single-top associated production $qb \rightarrow bth$ will also be accessible and studied in the LHC.
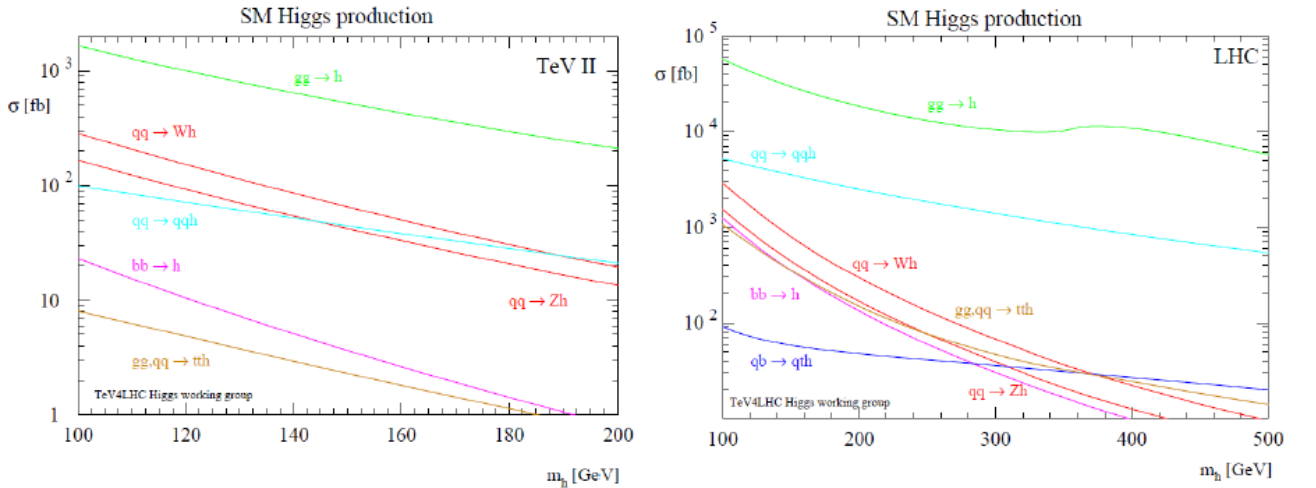
Figure 2.8: - SM Higgs production cross-sections at the Tevatron (left) and at the LHC (right).

Figure 2.8 displays the various SM Higgs cross-sections for the Tevatron and the LHC at design energy, with a much larger range of m_H for the second. The ratio between gluon fusion Higgs production and Higgsstrahlung at the LHC is then larger than at the Tevatron. In the Tevatron the large
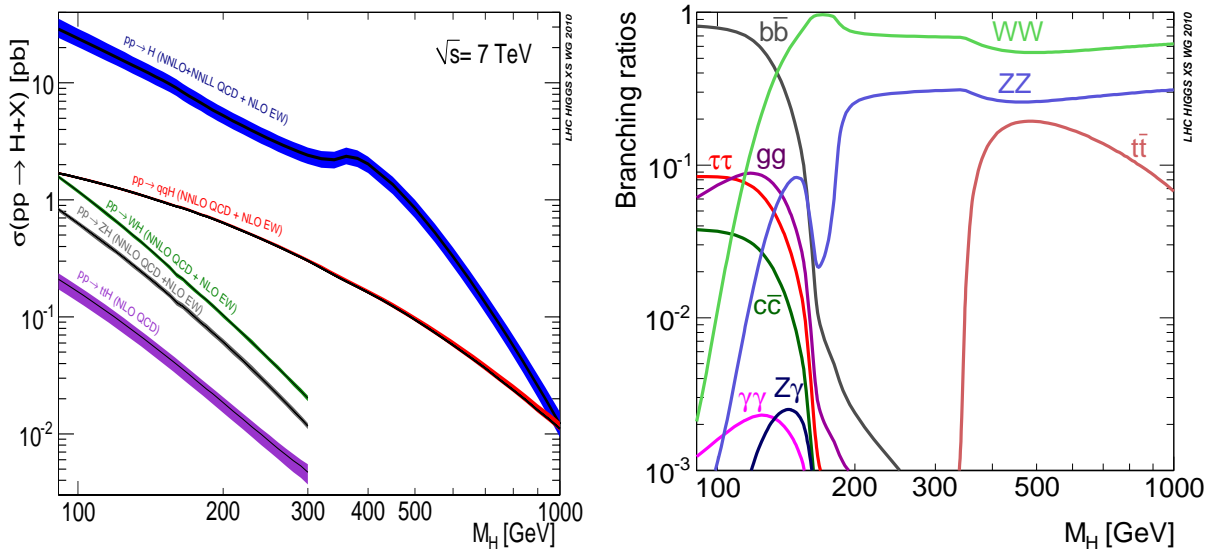
Figure 2.9: SM Higgs boson production cross section and branching ratios as a function of the Higgs boson mass at $\sqrt{s} = 7$ TeV

m_H cross-sections are small and that does not happen in the LHC.

2.5.2 Higgs boson decay modes

The Higgs boson decay modes depend on the Higgs mass, as it decays to the heaviest particles kinematically accessible. Therefore different strategies can be defined at several masses, taking into account that the search feasibility will be driven not only by the cross section but on the possibility of experimentally isolate the signal over the background. Figure 2.9 shows the branching ratios of the different Higgs decay channels depending on its mass.

For lower masses, below 135 GeV/c^2, the decay to $\bar{b}b$ dominates the branching ratio with 85% , although the huge QCD production at the LHC will make difficult to search in this channel, being the best possibilities to look for associated production with $t\bar{t}$, W and Z, that can provide leptons in the final state. In this range the second contribution comes from the τ pair decays, which is mainly considered through vector boson fusion because of its higher signal over background ratio. Despite the small branching ratio (0.2 %), the Higgs decaying to two photons is the most relevant channel at low masses, due to its clear signature of two isolated photons and the excellent resolution of the calorimeters.

For intermediate and higher masses, above 135 GeV/c^2, the main decay modes are those into WW and ZZ pairs, where one of the vector bosons is off-shell below the corresponding kinematical threshold. The WW is specially important around 160 GeV/c^2, with a branching ratio close to 100% . Both channels are considered the golden modes for the Higgs discovery at the LHC.

2.5.3 Higgs searches in CMS

Given the Higgs production and decay modes accessible at the LHC, many different analysis can be foreseen. The search strategy for several channels has been designed in CMS [20], see for example

[21], oriented to the discovery or exclusion for a SM Higgs boson in a mass range that starts where LEP stopped, around 114 GeV/c^2, and with special stress in masses that are not accessible by the Tevatron. The main SM channels to be studied, amongst others, are:

- $H \rightarrow \gamma\gamma$

- $H \rightarrow WW^* \rightarrow 2l2\nu$

- $H \rightarrow ZZ^* \rightarrow 4l/2l2\nu/2l2b$

- $H2\tau$ ($H \rightarrow \tau\tau$; $H \rightarrow ZZ^* \rightarrow 2l2\tau/4\tau...$)

- VBF (VBF $H \rightarrow WW^* \rightarrow l\nu jj$, VBF $H \rightarrow ZZ^* \rightarrow 2ljj$, VBF $H \rightarrow invisible$)

- ttH, $H \rightarrow b\bar{b}$

$H \rightarrow \gamma\gamma$ is one of the most promising channels for discovery of a Higgs boson at low masses between $120 < m_H < 140$ GeV/c^2 despite its small branching ratio (0.2% at $m_H = 120$ GeV/c^2), since it presents a very clean signal of two high E_T isolated photons. The main sources of background arise from events with two prompt isolated photons and from photon+jets or multijets events in which at least one jet is misidentified as photon, therefore it will be essential a good photon identification, and isolation criteria should be applied to reject this fake photons due to jets.The energy resolution of the calorimeter and the primary vertex reconstruction are essential in this channel.

It can be produced in an inclusive way or in association with hadronic high p_T jets. Once the signal is identified the Higgs mass peak can be reconstructed. With a good control of the backgrounds, Higgs to $\gamma\gamma$ analysis can lead to a low mass Higgs discovery with significant luminosity. In [21] the estimated luminosity needed for discovery using this channel was ~ 30 fb^{-1}.

For Higgs masses between 140 and 180 GeV/c^2 the channel that dominates is the $H \rightarrow WW^*$ that covers the most interesting SM Higgs mass range for a possible discovery [22]. The search strategy for this channel will be discussed in this thesis and more details will be given in Chapter 6. It will be mainly produced via gluon-fusion, but also the VBF Higgs to WW^*, that adds two forward jets to the final state, will be studied. In this case the cross-section is smaller than the gg-induced process but the signature is cleaner. In the intermediate mass range, between 130 GeV/c^2 and 200 GeV/c^2 and specially close to 160 GeV/c^2, the Higgs decaying to two W's is the most important channel. [23].

The $H \rightarrow ZZ^* \rightarrow 4l$ decay channel [24] was known as the Golden Higgs decay, as it brings the experimentally cleanest signature for discovery, with a narrow 4-lepton invariant mass peak on top of a smooth background. It is a powerful analysis in a wide mass range with a very high Branching ratio except for Higgs masses around two times the mass of the W, when the $H \rightarrow WW^*$ dominates. However, by indirect estimation of m_H based on EWK precision fits, see **??**, the preferred mass for a SM Higgs boson is around 150 GeV/c^2, and in that region, not only the WW Branching Ratio is higher, but also one of the Z bosons of the $H \rightarrow ZZ^*$ decay is off-shell. The analysis of this channel relies on the lepton identification and background control, and is sensitive up very high masses ~ 500 GeV/c^2 with a discovery potential for high energies of 5 fb^{-1} of integrated luminosity [24].

As mentioned in the previous section, the $H \rightarrow \tau\tau$[25; 26] is a significant channel for low masses despite its small cross-section, as it can be used for studying the properties of a possible Higgs boson, like the couplings to leptons. Different final states with leptons and jets in the final state (Vector-Boson fusion induced) are studied. The Higgs mass peak is reconstructed but it is not as straightforward as

in the case of Higgs to ZZ.

Higgs to $b\bar{b}$ is the only Higgs decay that can be studied at the LHCb experiment[27], and it is also considered in the Higgs program of CMS. It presents a large combinatorial background with more than 5 jets, and for that requires an excellent b-tagging performance.

Concerning the MSSM, the most significant channels at the LHC centre of mass energies are $\tau\tau$ and $\mu\mu$. Charged Higgs is interesting at low ($m_{H^\pm} < m_{top}$, $H^\pm \to \tau\nu$) and high ($m_{H^\pm} > m_{top}$, $H^\pm \to tb$) masses.

In the case of bbH/A, the reach can be extended down to $m_A \sim 150\ GeV/c^2$, $tan\beta \sim 10$. SM Higgs searches can be reinterpreted in MSSM ($h \to \gamma\gamma$, $qqh/H \to qq\tau\tau$) covering the low and intermediate $tan\beta$ region. These searches suffer from systematics, and many parameters and scenarios have to be studied.

2.6   Top quark physics

Since the discovery of the top quark [28], [29] in 1995, several of its properties have been extensively studied. This was only possible at the Tevatron proton-antiproton collider, but since the LHC startup it is possible to observe the top quark production for the first time outside the Tevatron and in proton-proton collisions at a center of mass energy of $\sqrt{s} = 7$ TeV. Additionally, the large amount of top pairs that can be produced at the LHC will give the opportunity to measure many of its properties with high precision

The top quark is also a very interesting particle. It is the heaviest known fundamental particle ($m_t = 173.3 \pm 1.1\ GeV/c^2$ from latest CDF and D0 available combined results [30]) and it has the strongest coupling to the SM Higgs boson (as it couples to fermions proportionally to its mass). Because of this, a detailed study of the top quark properties can provide constrains on the Higgs boson mass. Besides, due to its large mass, the top quark could play a role in the electroweak symmetry breaking and the generation of particle masses in alternative models to the Higgs mechanism.

It has a short lifetime, being the only quark that does not hadronize before decaying, which allows the study of the properties of a bare quark, its decay particles give the opportunity to directly measure its properties, as its mass, spin or charge.

Finally, in scenarios beyond the SM, new particles can decay into top-quarks, or top quark production can be the background for many new physics searches as well, so a detailed understanding of the top quark properties may provide a hint of new physics and to achieve as precise knowledge as possible is essential.

2.6.1   Top quark production

Top quarks are mostly produced in strong interactions as top-antitop pairs via gluon-gluon fusion and quark-antiquark annihilation (Figure 2.10). The importance of each of these processes depends on the center of mass energy and the type of particles in the beams.

At the LHC pp collider, the gluon-gluon scattering is the dominant production mechanism, which represents about a 90% of the produced top quarks, while at the Tevatron proton-antiproton collider, the $q\bar{q}$ scattering is the dominant one.
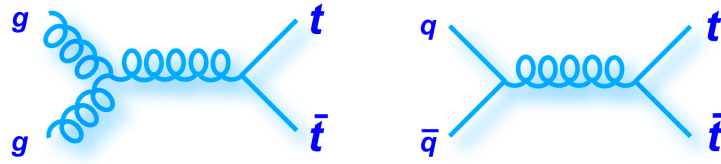
Figure 2.10: - Top-antitop production processes: (a) gluon-gluon fusion and (c) quark-anti quark annihilation.

At the LHC, the top quark is expected to be produced mainly via the strong interaction in $t\bar{t}$ pairs, but single top quarks can also be produced through electroweak interactions, these processes can be classified in t-channel (the top quark is produced through the exchange of a virtual W), s-channel (decay of a virtual W) (Figure 2.11) and tW channel (associated production with a W boson) (Figure 5.3).
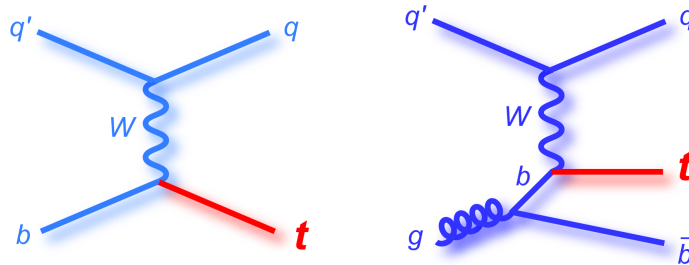
Figure 2.11: - Leading order feynman diagrams for the single top production in s-channel (left) and t-channel (right)
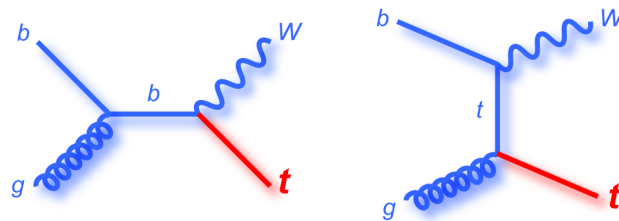
Figure 2.12: - Leading order feynman diagrams for the single top production in association with a W boson

2.6.2   Top quark decay

In the SM, the decay of top quarks takes place almost exclusively into a W boson and a b-quark. According to the W decay, several signatures can be identified experimentally (Figure 2.13):

- **Dileptonic channel**, which represents about 1/9 of the decays. Both W-bosons decay into a lepton and neutrino, resulting in an event with two leptons, $\not{E}_T$ and two b-jets, allowing to obtain a clean sample of $t\bar{t}$ events.

- **Semi-leptonic channel**: when only one of the W-bosons decays leptonically, and it represents

4/9 of the decays. The event presents 1 high p_{T} lepton, $\not{E}_T$ and at least 4 jets. This is an important channel for the top mass measurement.

- **Fully hadronic channel**: represents about 4/9 of the decays. Both W-bosons decay hadronically, which gives at least six jets in the event: 2 b-jets from the top quark decay and 4 light jets from the W decay. This is a challenging channel, as there is no high p_{T} lepton it's not easy to distinguish the signal from the huge QCD background.
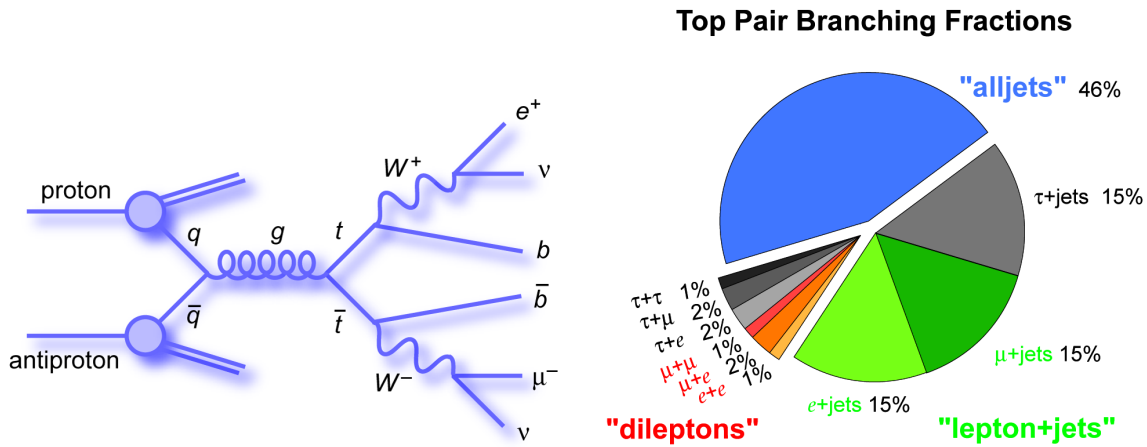
Figure 2.13: - $t\bar{t}$ branching ratios

3

The CMS experiment at LHC

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [31] is a circular particle accelerator of 26.7 km of circumference sited at CERN, near Geneva. It is installed in the existing LEP tunnel at a depth between 50 and 175 m underground, under French and Swiss territory. It was designed to provide proton-proton collisions at a center-of-mass energy of 14 TeV and a luminosity of $10^{34}\ cm^{-2}s^{-1}$, being the largest and highest energy accelerator ever built. It will also operate with heavy ion collisions (Pb-Pb) at a center-of-mass energy of more than 1000 TeV and a luminosity exceeding $10^{27}\ cm^{-2}s^{-1}$. Some design parameters of the LHC are shown in Table 3.1.

The design of the LHC has been driven by the goal to explore physics at the TeV scale. The increase in centre of mass energy and luminosity with respect to previous accelerators opens the possibility to discover new processes and objects of large mass and energy.

Large Hadron Collider	
circumference	26659 m
depth	50 - 175 m
total number of magnets	9600
number of main dipoles	1232
number of quadrupoles	392
temperature	1.9 K (-271.3C)
beam vacuum pressure	10^{-13} atm
nominal p energy	7 TeV
center-of-mass energy	14 TeV
design luminosity	$10^{34}\ cm^{-2}s^{-1}$
bunches per proton beam	2808
protons per bunch	1.1×10^{11}
turns per second	11245
collisions per second	600 million
length of each dipole	15 m
weight of each dipole	$\sim$ 35 t
dipole field	8.33 T

Table 3.1: Design parameters of the Large Hadron Collider.
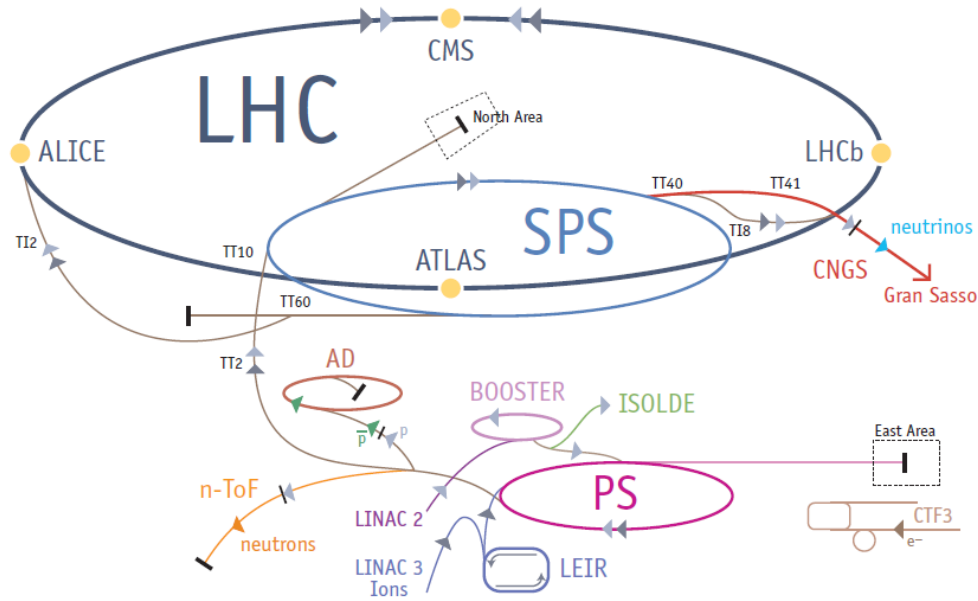
Figure 3.1: - Scheme of the CERN accelerator complex

The LHC is supplied with protons and *Pb* ions from the existing injector chain comprising Linear Accelerators, Proton Synchrotron Booster (PSB), Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). These accelerators were upgraded to meet the very stringent needs of the LHC. The Figure 3.1 shows a schematic view of the accelerator complex, in which the LHC is the last element.

One of the two linear accelerators, LINAC2, is used for the initial acceleration of protons, which are injected into the PSB, then the PS is used to provide a low energy beam (25 GeV) with the final bunch structure. Next, the beams are pre-accelerated using the SPS, and finally are injected into the LHC, where particles will be accelerated from 450 GeV to the nominal energy of 7TeV.

Inside the LHC, the beam of particles circulate in an ultrahigh vacuum tube, accelerated and focused by magnets. There are thousands of magnets of different varieties and sizes used to direct the beams around the ring. These include 1232 dipole magnets of 15 m length which are used to bend the beams, and 392 quadrupole magnets, each 57 m long, to focus the beams, as well as sextupoles, decapoles etc that are also used to correct the beam trajectories.

The acceleration of the beams from their injection energy of 450 GeV up to the nominal energy of 7 TeV will be performed by 8 so-called radio-frequency (RF) cavities, boosting the beams in total by 16 MeV per turn in an electric field of 5.5MV/m, oscillating at 400MHz. A serious difficulty comes from the high rigidity of a 7TeV proton beam, which requires a very strong magnetic field for the bending. The LHC exploits superconducting technology, with Nb-Ti conducting dipole magnets which are designed to produce a magnetic field of 8.33T to keep the beams on a circular trajectory at nominal energy and for that, they have to be maintained at 1.9 K,therefore, the LHC has the largest cryogenic system in the world, with superfluid helium pumped into the magnet systems. The refrigeration process happens in a sequential way, first helium is cooled down to 4.5 K, then it is injected into the magnet cold masses, and, once the magnets are filled, the refrigeration units bring the temperature to 1.9 K. The dipoles were the biggest technological challenge for the design of the

LHC.

The two proton beams circulate in two different vacuum chambers in opposite direction and are guided by 1232 of these superconducting dipoles. More than 2500 of other magnets are used to clean and focus the beams, correct their trajectories and finally collimate the beams to a transverse size of about 16 μm at the interaction points of the main experiments.

The protons circulate around the LHC ring in well-defined bunches. Under nominal operating conditions, LHC will have 2808 bunches in each beam, with around 10^{11} protons in each bunch. The size of the bunch is not constant around the ring, it gets squeezed and expanded around the LHC and, as they approach the collision points, the bunches get squeezed to about 16 μm, to allow the proton-proton collision to happen with higher probability. The luminosity of the machine increases with the number of bunches. At full luminosity, LHC will have a bunch spacing of 25 ns, that is, around 7 m, corresponding to a frequency of 40 MHz. The average crossing rate is equal to the total number of bunches multiplied by the number of turns round the LHC per second: $2808 \times 11245 = 31.6$ MHz. The particles are so tiny that the chance of any two colliding is very small. When the bunch cross, there will be a maximum of about 20 collisions between 200 billion particles. Bunches will cross about 30 million times per second, hence, at nominal conditions, LHC will generate up to 600 million particle collisions per second.

The exploration of the rich LHC physics program requires large particle detectors of great complexity, in order to be able to reconstruct the proton collisions in as much detail as possible.

The LHC physics program will be studied by means of the detectors installed at different points around the LHC ring. There are two general purpose detectors: ATLAS (A Toroidal LHC ApparatuS) [32] and CMS (Compact Muon Solenoid) [33], and other detectors specialized in different topics: ALICE (A Large Ion Collider Experiment)[34], dedicated to the study of Heavy Ions, LHCb (Large Hadron Collider beauty) [27], which focuses on the physics of b-quarks, LHCf(Large Hadron Collider forward) and TOTEM (TOTal Elastic and diffractive cross-section Measurement). ALICE, ATLAS, CMS and LHCb are installed in four huge underground caverns built around the four collision points of the LHC, while LHCf and TOTEM are close to the main detectors, ATLAS and CMS respectively.

On 10th September 2008 proton beams circulated for the first time in the LHC, but an incident occurred 9 days later during the commissioning of one of the main dipole circuits resulted in a number of magnets damaged, and almost a year had to be spent in repairing and consolidating the machine.

After the fully recommissioning of the machine, protons circulated again in the LHC on 20th November 2009 and the first pp collisions at injection energy ($\sqrt{s} = 0.90$ TeV) took place on 23rd November 2009, opening a new era in high energy physics. Not much later, on the 30th November 2009, the LHC became the world's highest energy accelerator, having accelerated the proton beams to an energy of 1.18 TeV, exceeding the previous world record of 0.98 TeV, which had been held by the Tevatron collider since 2001.

2010 was the first full year of operation of the LHC, which started with the first collisions at $\sqrt{s} = 7$ TeV that occurred on 30th March. Since then, the LHC was continuously running with some short technical stops until the 4th November, where it switched to heavy ion operations.

The running conditions were changing and evolving during this period, it was possible to in-

crease the number of protons per bunch and the luminosity, starting from about $8x10^{26}cm^{-2}s^{-1}$ in March to $2x10^{32}cm^{-2}s^{-1}$ at the end of the proton run in November, exceeding the planned goals. The experiments successfully recorded data with efficiencies higher than 90 % . The Compact Muon Solenoid (CMS) experiment has recorded integrated luminosity of over 43 $pb^{-1}$of proton-proton data at $\sqrt{s} = 7$ TeV. Figure 3.2 shows the integrated luminosity versus time delivered and recorded by CMS

The first Pb-Pb collisions in the LHC at a center of mass energy of 2.7 TeV were observed on 8th November 2010. In the following heavy-ion run until 6th December, the experiments recorded data corresponding to an integrated luminosity of about 10 μb.
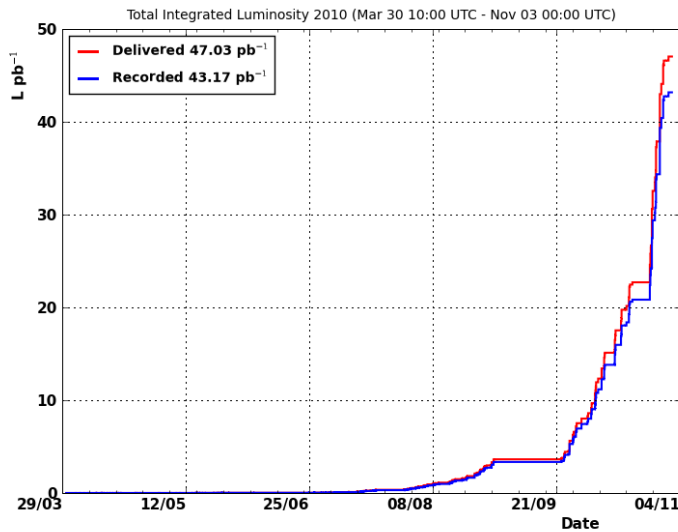
Figure 3.2: - Integrated luminosity versus time delivered (red) and recorded by CMS (blue) during stable beams at 7 TeV centre of mass energy during the 2010 LHC run

The 2010 run period was successful and it was possible to achieve many goals. After the early data essential for the calibration of the detectors, the increasing luminosity gave the experiments the chance to re-discover the SM physics: from quarkonia to electroweak and top physics, which improved the level of understanding of the detectors, getting ready for the new physics searches that will start to be possible with the accumulated luminosity of 1 fb^{-1}that it is expected to be collected before Summer 2011.

3.2 The CMS detector

The Compact Muon Solenoid, CMS [33], is a general purpose detector designed to cover the widest possible range of physics, and specially to perform new physics searches at the highest energies of the LHC. The overall dimensions of the detector are: 21.6 m height, a diameter of 14.6 m and a weight of 12500 tonnes.

The detector has been designed to achieve the main goals of the LHC physics program, which requires muon system with high performance to get a good muon identification and momentum resolution over a wide range of momenta, the best possible electromagnetic calorimeter, a high quality central tracking and a hadron calorimetry with good energy resolution and hermeticity with a large

geometric coverage.

In the CMS design, priority was given to the muon detection system. That, together with the desire to make it compact, led to the choice of a strong magnetic field. The only practical magnet able to generate a strong magnetic field in a compact scenario is a solenoid. It was then decided for CMS to have a large superconducting solenoid, of about 14 meters long with a radius of 3 meters, providing a 3.8 T magnetic field. That guarantees a good momentum resolution for high momentum muons (> 1 TeV) up to rapidities of 2.5. The coil does not affect the performance of the calorimeter, as the big radius of the magnet allows the tracking and the calorimeters to be nicely placed inside the solenoid.

The main distinguishing features of CMS are a high-field solenoid, a full-silicon-based inner tracking system, and a homogeneous scintillating-crystals-based electromagnetic calorimeter. The overall layout of CMS is shown in Figure 3.3. It is a cylindrical detector 13 meters long and 5.9 meters of inner diameter 3.8 T superconducting solenoid. The return field is large enough to saturate 1.5 m of iron, allowing 4 muon stations to be integrated, ensuring robustness and full geometric coverage. The magnet is large enough to accommodate the inner tracker and the calorimetry inside.
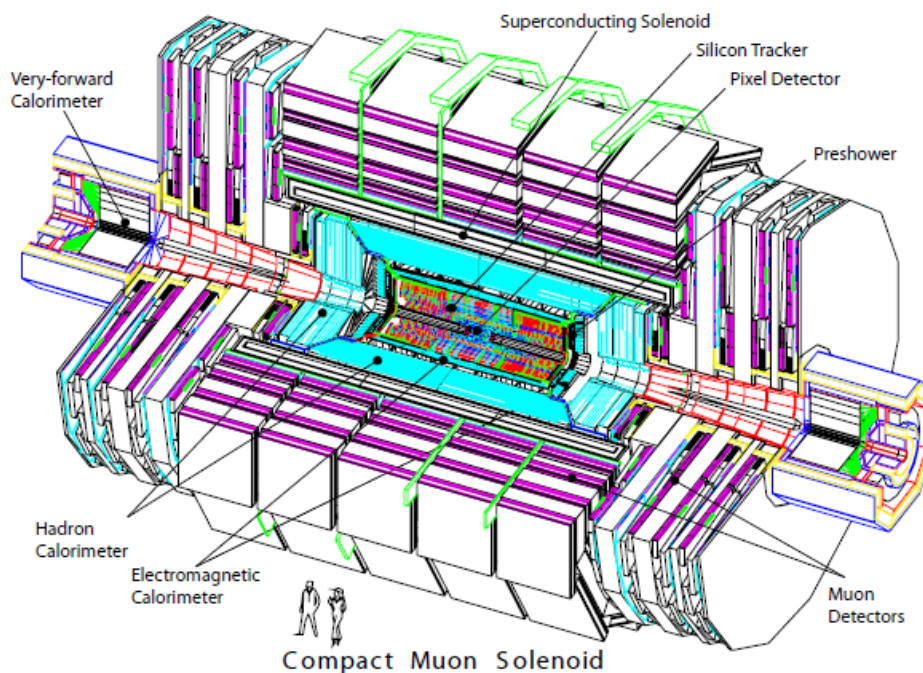
Figure 3.3: - A general view of the Compact Muon Solenoid

The coordinate system adopted by CMS, that will be used in the following sections, has the origin centered at the nominal collision point inside the experiment, the y-axis pointing vertically upward, the x-axis pointing radially inward toward the center of the LHC, and the z-axis along the beam direction towards the Jura mountains from the LHC Point 5. The azimuthal angle ϕ is measured from the x-axis in the $x - y$ plane, the polar angle θ is measured from the z-axis. Pseudorapidity η, is defined as $\eta = -\ln \tan(\frac{\theta}{2})$. The momentum and energy measured transverse to the beam direction, p_T and E_T, are computed from the x and y components.

A more detailed description of the different components of the CMS detector is given in the following subsections.

3.2.1 Magnet

The CMS magnet [35] is the central device around which the experiment is built, the magnet is used to bend the paths of particles emerging from high-energy collisions in the LHC. To ensure the appropriate performance of the muon system, a momentum resolution of $\Delta p/p \approx 10\%$ at $p = 1 \, TeV/c$ is achieved by CMS with a large Superconducting Solenoid which parameters are given in Table 3.2.

Superconducting Solenoid	
Field	4 T
Inner Bore	5.9 m
Length	12.9 m
Number of Turns	2168
Current	19.14 kA
Stored Energy	2.6 GJ
Hoop stress	64 atm

Table 3.2: Parameters of the CMS superconducting solenoid.

The CMS solenoid was built using the same technique previously used in the construction of the large solenoids, ALEPH and DELPHI at LEP and H1 at HERA, using a high-purity aluminium-stabilized conductor and indirect cooling by thermosyphon, together with a full epoxy impregnation. However, in this case, the large increase in parameters such as magnetic field, Ampere-turns, forces and stored energy, led to major changes. A four-layer winding has been adopted using a novel conductor with a larger cross-section that can withstand a hoop stress of 64 atmospheres. The conductor carries a current of 20 kA, and has a compound structure. The Rutherford-type cable is co-extruded with pure aluminium, that acts as thermal stabilizer, this is then electron-beam-welded to 2 plates, made of high-strength aluminium alloy, for the mechanical reinforcement. The overall conductor cross-section is $64 \times 22mm^2$, and it was manufactured in 20 continuous lengths, of 2.65 Km each, 4 lengths were wound to make each of the 5 coil modules. These modules were assembled and connected together, as well as the whole magnet, in the surface experiment hall, SX5 at Point 5.

3.2.2 Inner Tracking System

The CMS tracker [36] provides a precise and efficient reconstruction of the charged particles trajectories emerging from the LHC collisions, as well as the primary and secondary vertices reconstruction. The nominal momentum resolution is typically 0.7 (5.0)% at 1 (1000) GeV in the central region and the impact parameter resolution for high-momentum tracks is 10 μm .

The large amount of particles traversing the tracker and the fact that it has to lead with the presence of several overlapping pp collisions require from the tracker a high granularity and a fast response, in addition, as it is the inner most layer of the detector, it has to resist the high radiation from the intense particle flux. These requirements on granularity, speed and radiation hardness lead to a tracker design entirely based on silicon detector technology.

The tracker, shown in Figure 3.4, consists of two main detectors: a silicon pixel detector, with three barrel layers, covering the region from 4 to 15 cm in radius, and 49 cm on either side of the

collision point along the LHC beam axis, and a silicon strip detector with 10 barrel detection layers, covering the region from 25 to 110 cm in radius, and within 280 cm on either side of the collision point along the LHC beam axis. Each system is completed by endcaps with consist of 2 disk 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 $|\eta| < 2.5$. As particles travel through the tracker, the pixels and microstrips produce electric signals that are amplified and detected. With about 200 m^2 of active silicon area the CMS tracker is the largest silicon tracker ever built.

The silicon pixel detector is placed in the closest region to the interaction point, thus is essential for the reconstruction of secondary vertex from relatively long-lived particles such as bottom and charm hadrons and tau decays. It has 66 million active elements with a cell size of $100\mu m \times 150\mu m$, which are oriented with the smaller pitch in the azimuthal direction in the barrel and the radial direction in the disks , and cover a surface area of about 1 m^2. It is designed to provide the determination of three high precision three-dimensional points on track trajectories. The detector consists of three concentric cylindrical barrel layers and four fan-blade disks which close the barrel ends. The barrel layers have an active length of 53 cm and are located at average radii of 4.3, 7.3, and 10.2 cm. The endcap disks instrument the regions between radii 4.8 and 14.4 cm at mean longitudinal distances of 35.5 and 48.5 cm from the interaction point. The system provides efficient three-hit coverage in the region of pseudorapidity $|\eta| < 2.2$ and efficient two hit coverage in the region $|\eta| < 2.5$.

The silicon strip detector has 9.3 million active elements covering a surface area of 198 m^2. The detector consists of three large subsystems: The Tracker Inner Barrel and Disks (TIB/TID), Tracker Outer Barrel (TOB) and the Tracker Endcaps (TEC).

The TIB/TID extend in radius to 55 cm and are composed of four barrel layers, supplemented by three disks at each end. It delivers up to four $r - \phi$ measurements on a trajectory using 320 μm thick silicon microstrip sensors, which have their strips oriented parallel to the beam axis in the barrel and oriented radially in the disks. The strip pitch is 80 μm in the inner pair of TIB layers and 120 μm in the outer pair of TIB layers. In the TID, the mean pitch varies between 100 μm and 141 μm .

The TIB/TID is enclosed within the TOB, which has an outer radius of 116 cm. It consists of six barrel layers of 500 μm thick microstrip sensors with strip pitches of 183 μm in the first four layers and 122 μm in the last pair of layers.

The TOB extends to $\pm$ 118 cm in z. Beyond this z range, the Tracker EndCaps (TEC) instrument the region $124 < |z| < 280$ cm and $22.0 < r < 113.5$ cm. Each TEC is composed of nine disks that are instrumented with up to seven rings of radial strip silicon detectors. The sensor thicknesses are thin (320 μm) in the inner four rings and thick (500 μm) in the outer three rings; the average radial strip pitch varies from 97 μm to 184 μm . The inner two layers of the TIB and TOB, the inner two rings of the TID and TEC, and the fifth ring of the TEC include a second microstrip detector module that is mounted back-to-back at a stereo angle of 100 mrad and that enables a measurement of the orthogonal coordinate. Assuming fully efficient planes and not counting hits in stereo modules, there are from 8 to 14 high precision measurements of track impact points for $|\eta| < 2.4$.

The extended commissioning with cosmic rays in 2008 and 2009 allowed most calibrations to be completed and provided a good initial alignment for most of the detector. This allowed fast and reliable operation of the tracker in the first LHC collisions in December 2009. [37]. Also the alignment parameters determined in the cosmic ray studies were already adequate for the precise determination

of impact parameters and the reconstruction of secondary vertices, which can be exploited for b-hadron physics and top-quark studies.
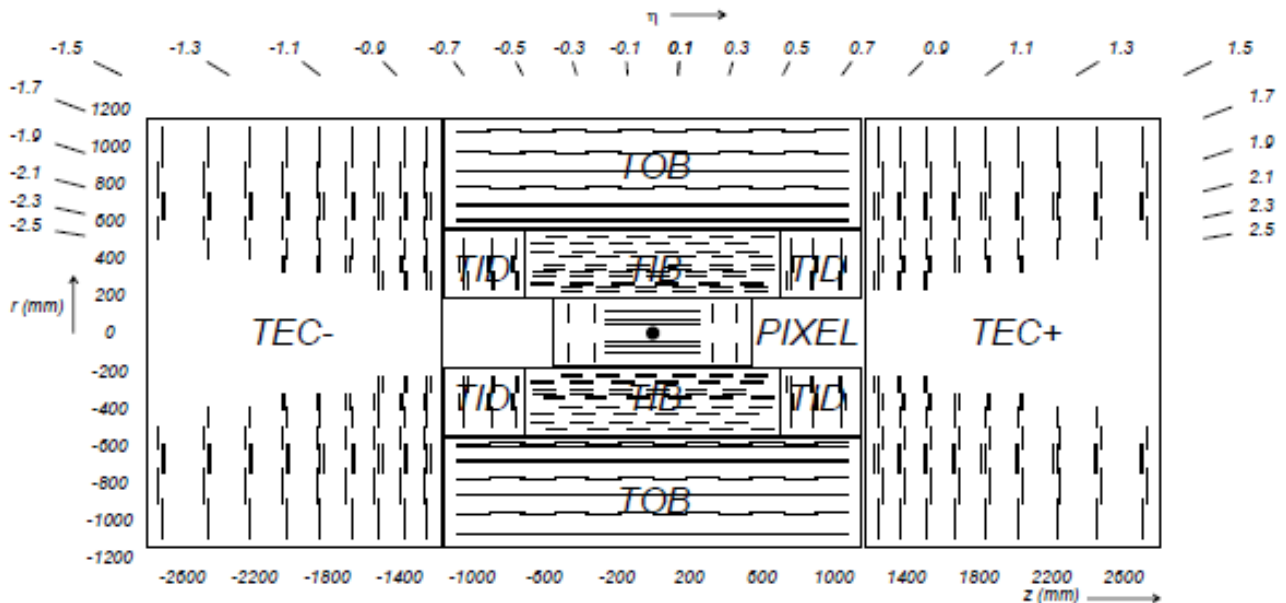
Figure 3.4: - Schematic cross-section through the CMS tracker, each line represents a detector module and double lines indicate back-to-back modules which deliver stereo hits.

3.2.3 Electromagnetic Calorimeter

To measure the energies of emerging particles and in particular to identify and measure electrons and photons, the electromagnetic calorimeter is the main component. It is designed with stringent requirements on energy resolution in order to be sensitive to the decay of a Higgs boson into two photons and other new physics signatures containing photons in the final states. Crystal calorimeters have the potential to provide fast response, radiation tolerance and excellent energy resolution.

The ECAL [38][39] is composed by lead tungstate crystal primarily made of metal. Heavier than stainless steel, it is highly transparent and scintillates when electrons and photons pass through it producing light in fast, short, well-defined photon bursts that allow for a precise, fast and fairly compact detector. Photodetectors specially designed to work within the high magnetic field, are glued onto the back of each crystal to detect the scintillation light and convert it to an electrical signal that is amplified and sent for analysis.

The light produced in the crystals is gathered with silicon Avalanche Photodiodes (APDs) in the barrel and Vacuum Phototriodes (VPTs) in the endcaps. The ECAL is hermetic and homogeneous, with a fine granularity, composed by 61200 lead tungstate ($PbWO_4$) crystals in the barrel part, and 7324 crystals in each of the two endcaps.

The characteristics of PbWO4 crystals make them an appropriate choice for operation at LHC. The high density (8.28 g/cm^3), short radiation length (0.89 cm) and small Moliere radius (2.2 cm) result in a fine granularity and a compact calorimeter, besides, they are fast (as around 80% light is emitted within 25 ns) and resistent to radiation, all important characteristics in the LHC environment.The layout of the ECAL is shown in Figure 3.5.
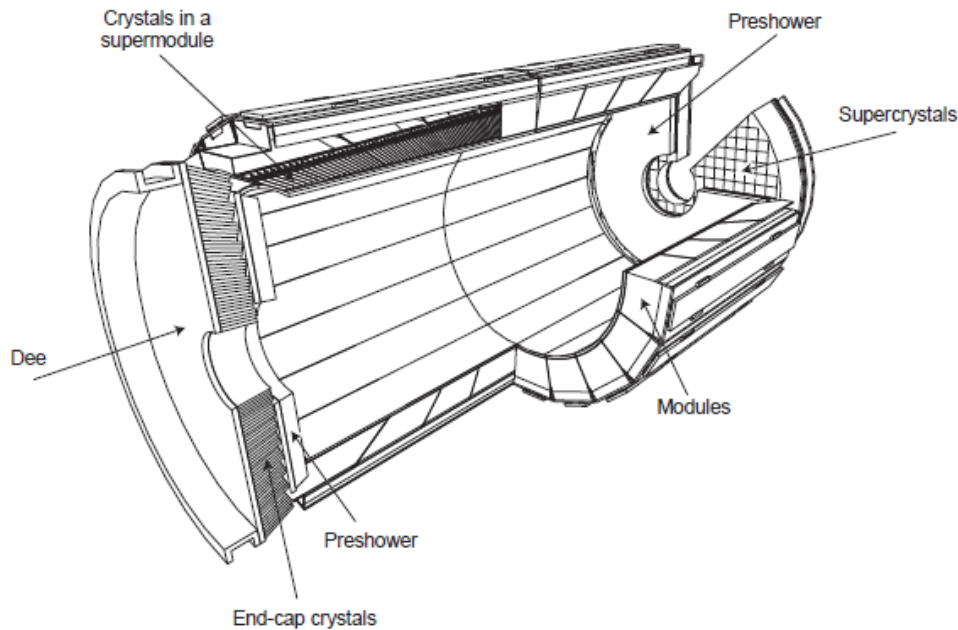
Figure 3.5: - Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front.

The barrel section (EB) has an inner radius of 129 cm. It is subdivided into 36 identical supermodules, each containing 1700 crystals and provides coverage in pseudorapidity up to $|\eta| < 1.479$. The crystals are quasi-projective and cover 0.0174 (that is 1°) in $\Delta\phi$ and $\Delta\eta$. The crystals have a front face cross-section of $\approx 22 \times 22$ mm^2 and a length of 230 mm, corresponding to 25.8 χ_0.

The endcaps section (EE) is located a distance of 314 cm from the vertex, covering a pseudorapidity range between $1.479 < |\eta| < 3.0$. They are each structured as two "Dees", consisting of semi-circular aluminium plates from cantilevered structural units of 5×5 crystals, known as supercrystals.
A preshower device is placed in front of the crystal calorimeter over much of the endcap pseudorapidity range, composed by 2 planes of silicon strip detectors, with a pitch of 1.9 mm, that lie behind disks of lead absorber at depths of $2\chi_0$ and $3\chi_0$.

3.2.4   Hadron Calorimeter

The Hadron Calorimeter (HCAL) [40] is important for the measurement of the energy of hadron jets and provides indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos, which escape detection and do not leave record of their presence in any part of the CMS detector. The determination of this missing energy will also form a crucial signature for new particles and phenomena, such as will be encountered in the searches for the supersymmetric partners of quarks and gluons. The hadron calorimeter will also help in the identification of electrons, photons and muons together with the electromagnetic calorimeter and the muon system.

The HCAL is a sampling calorimeter, it consists of active material inserted between copper absorber plates; it finds a particles position, energy and arrival time using alternating layers of absorber and fluorescent scintillator materials that produce a rapid light pulse when the particle passes through.

Optic fibres collect up this light and feed it into readout boxes where photodetectors amplify the signal. When the amount of light in a given region is summed up over many layers of tiles in depth, called a tower, this total amount of light is a measure of a particle's energy.

The HCAL is massive and thick, made of staggered layers, with no gaps in direct lines. Fitting the HCAL into CMS was a challenge, as the showers, cascades of particles produced when a hadron hits the dense absorber material, are large. The fact that most of the CMS calorimetry is located inside the magnet coil, strongly influenced the design, very dependant of the choice of magnet parameters. An important requirement of HCAL is to minimize the non-Gaussian tails in the energy resolution and to provide good containment and hermeticity for the missing E_T measurement.

The HCAL consists of three main parts: the Hadron Barrel (HB) and the Hadron Endcap (HE), which extend to $|\eta| = 3$, and the Hadron Forward (HF), which completes the coverage up to $|\eta| = 5.0$

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 (HO) or tail catcher is placed outside the solenoid complementing the barrel calorimeter. Beyond $|\eta| = 3$, the forward hadron calorimeters placed at 11.2 m from the interaction point extend the pseudorapidity coverage down to $|\eta| = 5$ using a Cherenkov-based, radiation-hard technology.

Both the barrel and endcap calorimeters experience the field of the CMS solenoid and hence are necessarily fashioned out of non-magnetic material (copper alloy and stainless steel).
Brass has been chosen as absorber material for its short interaction length, and for being easy to machine and non-magnetic. Maximizing the amount of absorber before the magnet requires keeping to a minimum the amount of space devoted to the active medium. The tile/fibre technology makes for an ideal choice. It consists of plastic scintillator tiles read out with embedded wavelength-shifting (WLS) fibres. The WLS fibres are spliced to high attenuation-length clear fibres outside the scintillator that carry the light to the readout system. The photodetection readout is based on multi-channel hybrid photodiodes (HPDs). The absorber structure is assembled by bolting together precisely machined and overlapping brass plates so as to leave space to insert the scintillator plates, which have a thickness of 3.7 mm. The overall assembly enables the HCAL to be built with essentially no uninstrumented cracks or dead areas in ϕ. The gap between the barrel and the endcap HCAL, through which the services of the ECAL and the inner tracker pass, is inclined at 53° and points away from the center of the detector.

The HB consists of 32 towers covering $-1.4 < \eta < 1.4$, resulting in 2304 towers with a segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The HB is constructed in 2 half barrels, and it is read out as a single longitudinal sampling. There are 15 brass plates, each with a thickness of about 5 cm, plus 2 external stainless steel plates for mechanical strength. Still in the barrel, the HO detector contains scintillators with a thickness of 10 mm, which line the outside of the outer vacuum tank of the coil and cover the region $-1.26 < |\eta| < 1.26$. The tiles are grouped in 30°-sectors, matching the ϕ segmentation of the DT chambers. They sample the energy from penetrating hadron showers leaking through the rear of the calorimeters and so serve as a tail-catcher after the magnet coil. They increase the effective thickness of the HCAL to over 10 interaction lengths, reducing the tails in the energy resolution function.

The HO also improves the missing E_T resolution. It is located inside the barrel muon system and is hence constrained by the geometry and construction of that system. It is divided into 5 rings, in the ring 0 it has 2 scintillator layers on either side of an iron absorber with a thickness of about 18
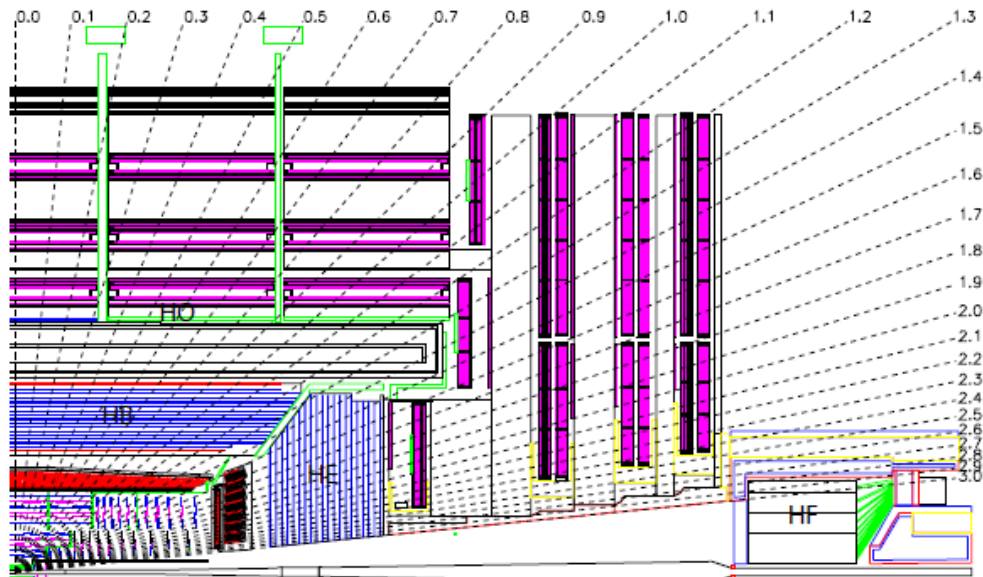
Figure 3.6: - Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters.

cm, at radial distances of 3.850 m and 4.097 m, respectively. The other mobile rings have single layers at a radial distance of 4.097 m. Each ring covers 2.5 m in z. HO scintillators follow the HCAL barrel tower geometry in η and ϕ.

In the endcaps, each HE, consists of 14 η towers covering the region $1.3 < |\eta| < 3.0$. The segmentation varies in the outer and innermost parts, being he total number of HE towers is 2304. In the forward region, for pseudorapidities between 3.0 and 5.0, the steel/quartz fibre HF calorimeter operates.

To extend the hermeticity of the central hadron calorimeter system to pseudorapidity of five (as required for a good missing transverse energy measurement), CMS employs a separate forward calorimeter (HF) located 6m downstream of the HE endcaps. The HF calorimeter covers the region $3.0 < |\eta| < 5.0$. It uses quartz fibers as the active medium, embedded in a copper absorber matrix. The HF will be located in a very high radiation and a very high rate environment. Because of the quartz fiber active element, it is predominantly sensitive to Cerenkov light from neutral pions. This leads to its having the unique and desirable feature of a very localized response to hadronic showers.

3.2.5   Muon System

The CMS muon system is designed to have the capability of reconstructing the momentum and charge of muons over the entire kinematic range of the LHC.

Muon detection is a main task for CMS, as they appear in many interesting decays of new physics or in Higgs boson decay signatures that can be produced by the LHC. As muons can penetrate several meters of material without interacting, they are not stopped by any of the calorimeters. Therefore, chambers to detect muons are placed at the very edge of the experiment where they are the only particles likely to register a signal conforming the Muon detection System [41].

3. THE CMS EXPERIMENT AT LHC

Centrally produced muons are measured three times, in the inner tracker, after the coil, and in the return flux. In the muon system, a particle is measured by fitting a curve to hits among the four muon stations, which sit outside the magnet coil and are interleaved with iron return yoke plates. By tracking its position through the multiple layers of each station, combined with tracker measurements, the particle's path can be precisely traced.

The momentum is essentially determined by the muon bending angle at the exit of the 4T coil, taking the interaction point as the origin of the muon. The resolution of this measurement is dominated by multiple scattering in the material before the first muon station up to p_T values of 200 GeV/c, when the chamber spatial resolution starts to dominate. For low-momentum muons, the best momentum resolution is given by the resolution obtained in the silicon tracker. However, the muon momentum resolution for hight p_T muons can be improved, extrapolating the muon trajectory beyond the return yoke back to the beam-line, combining the inner tracker and muon detector measurements.

There are three types of gaseous detectors used to identify and measure muons in CMS. The choice of the detector technologies has been driven by the very large surface to be covered and by the different radiation environments. In the barrel ($|\eta| < 1.2$), where the neutron induced background is small and the muon rate and the residual magnetic field in the chambers are low, drift tube chambers (DT) are used. In the endcaps, where both the muon rate and the neutron induced background rate are high, and the magnetic field is also high, cathode strip chambers (CSC) are located, covering the region up to $|\eta| < 2.4$. In addition to this, resistive plate chambers (RPC) are used in both the barrel and the endcap regions. The RPCs are operated in avalanche mode to ensure good operation at high rates (up to 10 kHz/cm^2) and have double gaps with a gas gap of 2 mm. RPCs provide a fast response with good time resolution but with a coarser position resolution than the DTs or CSCs. RPCs can unambiguously identify the correct bunch crossing. The DTs or CSCs and the RPCs operate within the first level trigger system, providing 2 independent and complementary sources of information. The complete system results in a robust, precise and flexible trigger device. The Layout of the muon system is presented in Figure 3.7.

In the Muon Barrel (MB) region, 4 stations of detectors are arranged in cylinders interleaved with the iron yoke. The segmentation along the beam direction follows the 5 wheels of the yoke (labeled YB-2 for the farthest wheel in $-z$, and YB+2 for the farthest is $+z$). In each of the endcaps, the CSCs and RPCs are arranged in 4 disks perpendicular to the beam, and in concentric rings, 3 rings in the innermost station, and 2 in the others. In total, the muon system contains of the order of 25000 m^2 of active detection planes, and nearly 1 million electronic channels.

The Barrel Detector, consists of 250 chambers organized in 4 layers, called stations, labeled MB1, MB2, MB3 and MB4, being the MB1 the closest to the beam axis, and the MB4 the outermost; within 5 wheels, divided in 12 sectors of 30° azimuthal angle each. Chambers in different stations are staggered so that a high-p_T muon produced near a sector boundary crosses at least 3 out of the 4 stations. There are 12 chambers in each of the 3 inner layers. In the 4th layer, the top and bottom sectors host 2 chambers each, thus leading to a total of 14 chambers per wheel in this outermost layer. The MB1, 2 and 3 chambers consist of 12 planes of aluminium drift tubes; 4 $r - \phi$ measuring planes in each of the 2 outermost superlayers, separated by about 20 cm and sandwiching a z-superlayer comprising 4 z-measuring planes. The MB4 station does not contain the z-measuring planes. The maximum drift length is 2.0 cm and the single point resolution is ≈ 200 μm. Each station is designed to give a muon vector in space, with a ϕ precision better than 100 μm in position and approximately 1 $mrad$ in direction. Each DT chamber has 1 or 2 RPCs coupled to it before installation, depending on
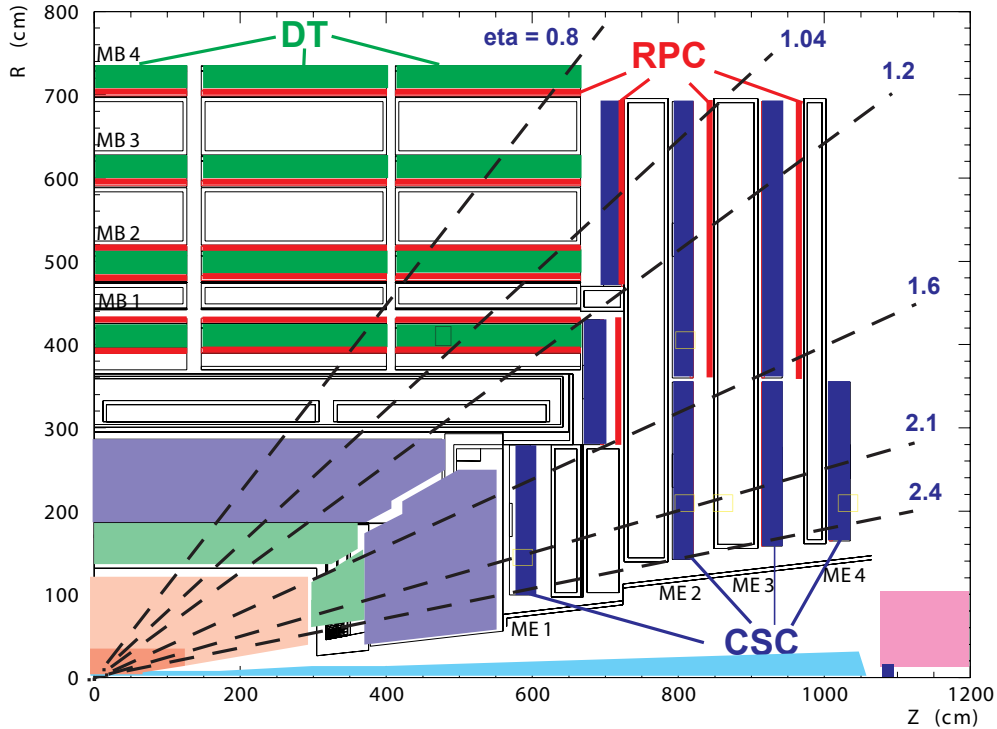
Figure 3.7: - Layout of the CMS muon system

the station. In stations MB1 and MB2, each package consists of 1 DT chamber sandwiched between 2 RPCs. In stations MB3 and MB4, each package comprises 1 DT chamber and 1 RPC, which is placed on the innermost side of the station. A high-p_T muon crosses up to 6 RPCs and 4 DT chambers, producing up to 44 measured points in the DT system from which a muon-track candidate can be built.

The Muon Endcap (ME) system comprises 468 CSCs in the 2 endcaps. Each CSC is trapezoidal in shape and consists of 6 gas gaps, each gap having a plane of radial cathode strips and a plane of anode wires running almost perpendicularly to the strips. All CSCs except those in the third ring of the first endcap disk (ME1/3) are overlapped in ϕ to avoid gaps in the muon acceptance. There are 36 chambers in each ring of a muon station, except for the innermost ring of the second through fourth disks (ME2/1, ME3/1, and ME4/1) where there are 18 chambers. The gas ionization and subsequent electron avalanche caused by a charged particle traversing each plane of a chamber produces a charge on the anode wire and an image charge on a group of cathode strips. The signal on the wires is fast and is used in the Level-1 Trigger. However, it leads to a coarser position resolution. A precise position measurement is made by determining the center-of-gravity of the charge distribution induced on the cathode strips. Each CSC measures up to 6 space coordinates (r, ϕ, z). The spatial resolution provided by each chamber from the strips is typically about 200 μm (100 μm for ME1/1). The angular resolution in ϕ is of order 10 $mrad$.

In the forward region, the system comprises 4 stations covering the pseudorapidity region up to $|\eta| < 2.1$. However, a shortfall of funds has led to the staging of the chambers sitting beyond $|\eta| > 1.6$. RPCs in the first endcap station are also used to help resolve ambiguities in the CSCs. There are 36 chambers mounted in each of 2 rings in each of the endcap stations.

3.3   Trigger

The Trigger is the start of the physics event selection process and it is needed to reduce the huge rate of collision events produced to the level at which one can write data to mass storage. This is a challenging task at LHC, where a few interesting events have to be efficiently selected among billions of backgrounds ones. In addition, this selection has to be kept as inclusive as possible in order to cover all the broad physics program of CMS.

The trigger system must reduce the rate of approximately 40 million collisions occurring every second during nominal running to no more than about 100 events per second for offline analysis, consistent with the disk storage capability of the order of O(100) MB/s.

In addition to that, the short time between bunch crossings, 25 ns, require from the trigger system a high event process speed. This time is even smaller than the needed to read out all the raw data from the detector. With this time constraint it is not possible to achieve the large rejection needed while keeping a high efficiency for the physics phenomena that CMS plans to study in a single processing step, so this reduction is carried out in two steps, namely the Level-1 Trigger (L1), which provides a fast decision based on hardware information, and the High-Level Trigger (HLT), which analyzes the full event to decide if it is stored or not.

The Level-1 (L1) trigger [42] analyzes each 25 ns crossing with a latency of 3 μs, it is build from custom electronics and only uses coarsely segmented data from calorimeter and muon detectors and the rate of events accepted on the basis of this limited information for further processing has to be less than 100 kHz. The Level-1 trigger system is designed to reduce the event rate down to at most 100 kHz.

During the latency of the Level-1 trigger the event data is stored in front-end pipelines. With such a high collision rate, new particles are being generated before the ones from the last event have even left the detector so the data is stored in pipelines that can retain and process information from many interactions at the same time. To not confuse particles from two different events, the detectors must have very good time resolution and the signals from the millions of electronic channels must be synchronized so that they can all be identified as being from the same event.

The minimum transit time for signals from the front-end electronics get to the services cavern housing the Level-1 trigger logic, reach a decision to keep or discard the event, and return back is around 3.2 μs. During the latency of the Level-1 trigger the event data is stored in front-end pipelines. With such a high collision rate, new particles are being generated before the ones from the last event have even left the detector so the data is stored in pipelines that can retain and process information from many interactions at the same time. To not confuse particles from two different events, the detectors must have very good time resolution and the signals from the millions of electronic channels must be synchronized so that they can all be identified as being from the same event. Of the total latency, the time allocated to Level-1 trigger calculations is less than 1 μs. The Level-1 triggers involve the calorimetry and muon systems, and its decision is based on the presence of objects such as photons, electrons, muons, and jets with a energy above some thresholds. It also employs global sums of E_T and missing E_T. Reduced-granularity and reduced-resolution data are used to form trigger objects. The overall scheme of the Level-1 trigger can be seen in Figure 3.8.

Upon receipt of a Level-1 trigger, after the fixed time interval of 3.2 μs, the data from the pipelines is transferred to front-end readout buffers. After further signal processing, zero suppression and/or
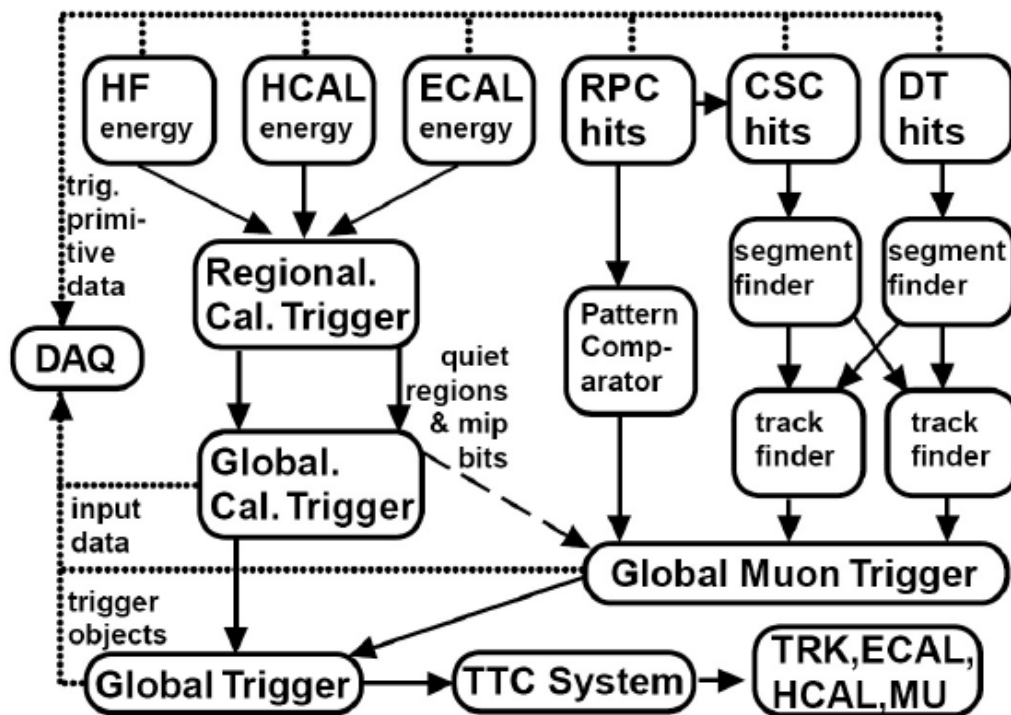
Figure 3.8: - Scheme of the Level-1 Trigger.

data-compression, the data are placed in dual-port memories for access by the DAQ system. Each event, with a size of about 1.5 MB in the case of pp interactions, is contained in several hundred front-end readout buffers. Through the event building switch, data from a given event is transferred to a processor. Each processor runs the same high-level trigger (HLT) software code to reduce the Level-1 output rate in a factor of 10^3. Various strategies guide the development of the HLT code. Rather than reconstruct all possible objects in an event, whenever possible only those objects and regions of the detector that are actually needed are reconstructed. Events are to be discarded as soon as possible. This leads to the idea of partial reconstruction and to the notion of many virtual trigger levels, e.g., calorimeter and muon information are used, followed by use of the tracker pixel data and finally the use of the full event information.

The High Level Trigger (HLT) [43] can make use in addition of the information from the tracking detectors. It is implemented in software running on a single processor farm, which consist of O(1000) processors. It is designed to reduce the maximum Level-1 accept rate to a final output rate of 100 Hz. Special care has been taken to have the trigger select not only standard model processes, but also the largest possible variety of exotic processes.

In the HLT a sequence of reconstruction and selection steps of increasing complexity is performed, setting up what is called a trigger path, which is designed to select a desired collection of events. The reconstruction modules and selection filters of the HLT use the software framework that is also be used for offline reconstruction and analysis.

Many such paths are executed for each event in parallel, forming the so called trigger menu, that contains several paths designed to select known standard model processes and also physics beyond de standard model.

Events accepted by the HLT are handled by a Storage Manager application, which streams event data on disk and eventually transfer the data files to the CMS Tier-0 computing center at CERN for permanent storage and offline processing. Based on the HLT trigger bits, events are grouped into a set of streams with distinct sets of event data files being written for each stream. The grouping of events is done according to their offline usage and is not exclusive, i.e. an event can be written to several streams.

The primary stream is the physics stream, which consists of events which satisfy the needs of most offline physics analysis. Other streams serve special purposes such as monitoring the performance of the HLT algorithms. The event content used for the physics stream comprises the full detector and trigger raw data, and the L1 and HLT trigger results.

Upon transfer to the CERN Tier-0 computing site, the data corresponding to the various streams is reformatted and split into various Primary Datasets (PDs). These datasets are defined, within a stream, by grouping a set of triggers that perform similar selections. The primary datasets are intended to provide small-sized samples suitable for specific physics analysis. A small fraction (10%) of the events selected by the HLT are sent to the so-called Express Stream (ES). They are processed through a quasi-online reconstruction and can be used for express analysis of the data.

3.3.1 Validation of the trigger software performance

It is important to ensure an efficient performance of the triggers that will be used to collect the signal events. Before the data taking, trigger rates and efficiency studies were performed to define the trigger tables to be run at different luminosities.

Each time a new trigger table is developed, the different triggers have to be validated in order to fix any possible issue. In simulation this is performed using a small set of dedicated samples, so-called RelVal, that are produced for every new software release before the final production is made, for the purpose of testing its performance. A set of quantities and distributions are regularly checked, such as trigger efficiency for the lepton trigger paths in final states with leptons and the differential distributions in η and p_{T} of the reconstructed leptons.

In order to optimize the validation procedure, the production of the validation results was included simultaneously with the production of the samples for an automatic detection and report of the performance discrepancies. This way, the plots can then be browsed through the Offline Data Quality Monitoring page of CMS, as illustrated in Figures 3.9 and 3.10.

3.4 CMS Software and Computing

The CMS software and computing systems cover a broad range of activities including design, evaluation, construction, and calibration of the detector; the storage, access, reconstruction and analysis of data; and the support of a distributed computing infrastructure for physicists engaged in these tasks. The storage, networking and processing power needed to analyze these data is well in excess of todays facilities and exceed any reasonably projected capabilities of CERNs central computing systems.

After the trigger system, CMS still produces a huge amount of data that must be analyzed, more than five petabytes per year when running at peak performance. To meet this challenge, the LHC
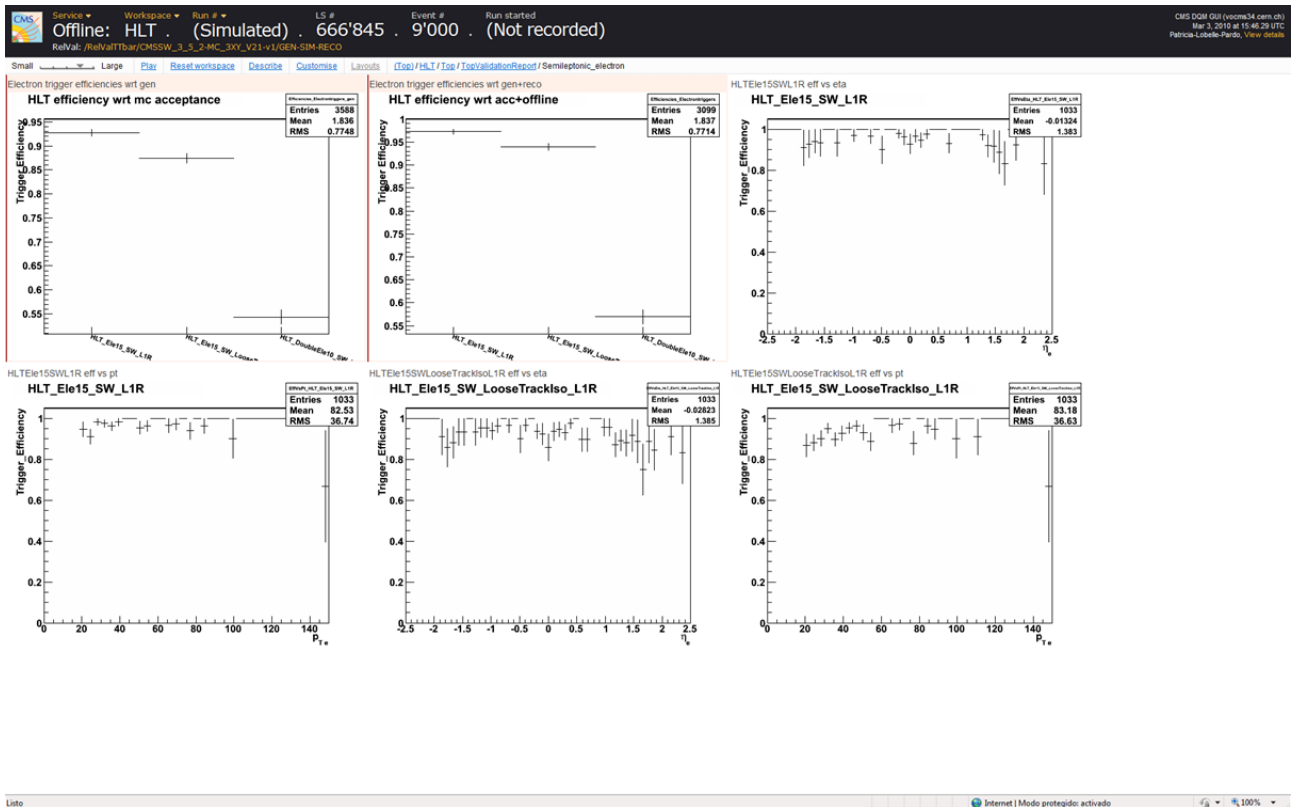
Figure 3.9: - Example of offline DQM system, in this case electron trigger efficiencies for top events
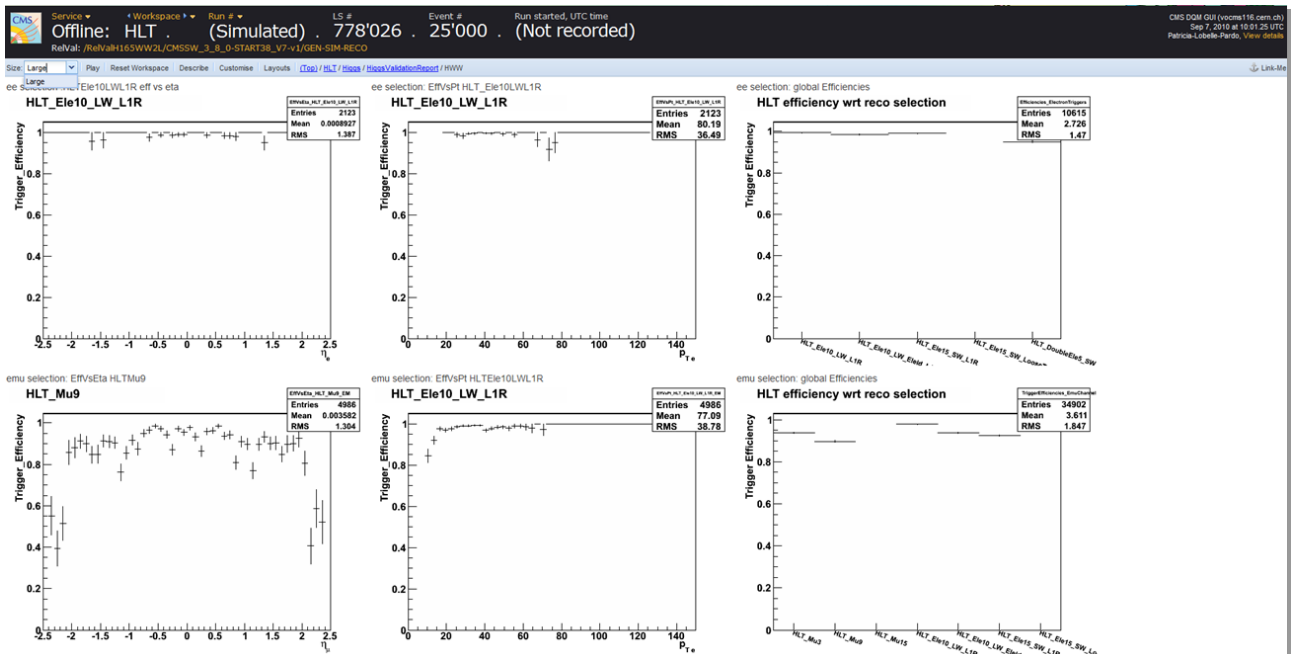
Figure 3.10: - Example of offline DQM system, in this case electron and muon trigger efficiencies for $H \rightarrow WW^* \rightarrow 2l2\nu$ events

employs the Worldwide LHC Computing Grid (WLCG), a distributed computing and data storage infrastructure. Giving access to data to thousands of scientists all over the world in a highly distributed way. It starts in the Tier 0 center at CERN reconstructs first the full collision events and data is started to be analyzed; but there is still a long way to go. Once CERN has made a primary backup of the data it is then sent to large Tier 1 computer centers, located in seven countries around the world: France, Germany, Italy, Spain, Taiwan, the UK and the US. There events are reconstructed again, using information from the experiment to improve calculations using refined calibration constants. The most complex events are then sent to a number of Tier 2 facilities, around 40, for further specific analysis tasks. In this way information branches out from each tier, and the analysis of the final data can be performed locally all around the world.

The CMS computing model [44] is therefore, highly distributed, with a primary Tier-0 center at CERN being supplemented by Tier 1 and Tier 2 computing centers at national laboratories and universities worldwide. Much software was and is still being developed and verified for simulation and physics analysis, as well as common libraries, tools and frameworks.
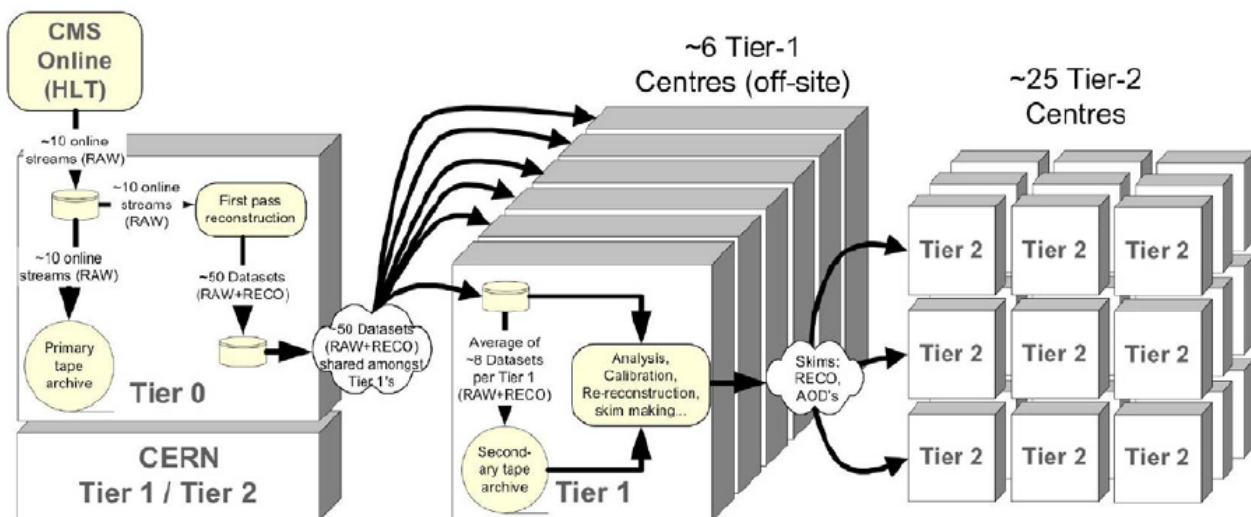
Figure 3.11: - Event data flow in the CMS Computing System.

In parallel to that, and before the data enters in the Tier system, calibration and express stream analysis are performed in the CERN Analysis Facility, CAF. The CAF is dedicated to latency critical activities like Calibration and Alignment, Detector/Trigger Commissioning or High Priority Physics Analysis. CAF access is restricted to users dedicated to these latency critical and high priority workflows.

With respect to the software, the CMS Software, CMSSW [45], consists of more than 2 Million lines of code not including external third party packages and has a very active developer community composed by a growing group of geographically dispersed collaborators. It performs a variety of event processing, selection and analysis tasks, and is used in both offline and online contexts. It was designed to be sufficiently modular to facilitate the development and maintenance. The chosen architecture consists of a common framework which is adaptable for each type of computing environment, physics modules which plug into the framework via a well-defined interface, and a service and utility toolkit which decouples the physics modules from details of event I/O, user interface, and other environmental

constraints.

The central concept of the CMS data model is the Event. The Event provides access to the recorded data from a single triggered bunch crossing, and to new data derived from it. The Event also contains information describing the origin of the raw data, and the provenance of all derived data products. Events are physically stored as persistent ROOT files. The Event is used by a variety of physics modules, which may read data from it, or add new data, with provenance information automatically included. Each module performs a well-defined function relating to the selection, reconstruction or analysis of the Event. Several module types exist, each with a specialized interface. These include: event data producers, which add new data products into the event; filters used in online triggering and selection; analyzers, producing summary information from an event collection; and input and output modules for both disk storage and DAQ. As modules execute independently from one another, and communicate only though the Event; they are developed and verified independently. A complete CMS application is constructed by specifying to the Framework one or more ordered sequences of modules through which each Event must flow, along with the configuration for each. The Framework configures the modules, schedules their execution, and provides access to global services and utilities.

To maintain the required level of data reduction, keeping the flexibility, CMS uses several event formats with differing levels of detail and precision. Other specialized event formats are used for heavy-ion data. The process of data reduction and analysis takes place in steps, typically carried out at different computer centers. The main types are, RAW, RECO, and AOD. RAW events contain the full recorded information from the detector, plus a record of the trigger decision and other meta-data. The RAW data is permanently archived in safe storage, and is designed to occupy around 1.5 $MB/event$ (2 $MB/event$ for simulated data, due to additional Monte Carlo truth information). The Reconstructed format, RECO data is produced by applying several levels of pattern recognition and compression algorithms to the RAW data. Reconstruction is the most CPU-intensive activity in the CMS data processing chain. The resulting RECO events contain high-level physics objects, plus a full record of the reconstructed hits and clusters used to produce them. Sufficient information is retained to allow subsequent application of new calibrations or algorithms without recourse to RAW data. RECO events are foreseen to occupy around 0.5 $MB/event$. Finally, AOD, Analysis Object Data is the compact analysis format, designed to allow a wide range of physics analysis whilst occupying sufficiently small storage so that very large event samples may be held at many centers. AOD data is produced by filtering of RECO data, and requires around 100 $kB/event$, small enough to allow a complete copy of the experimental data in AOD format to be held at computing centers outside CERN.

To facilitate development, CMS consolidates its code base regularly into releases. Different releases are grouped into release cycles dedicated to a specific purpose (that is, data taking and reconstruction in LHC collision period, integration of a new ROOT version, etc.) to aggregate specific feature sets of the software stack. The rapid development cycle of CMS and the resulting high number of releases requires a thorough quality assurance process. To guarantee stable and performant releases while supporting a high turn around and diverse development, CMS implemented an advanced central release validation process.

3.4.1 CMS Remote Analysis Builder

CMS Remote Analysis Builder, known as CRAB, is a Python program intended to simplify the process of creation and submission of CMS analysis jobs into a grid environment. It is then used to run jobs on the grid [46].

In order to store and manage a huge quantity of data, while assuring data access to physicists of CMS collaboration and computing power for analysis and simulation, CMS uses a distributed architecture based on Grid infrastructure that guarantees resources and data availability to allowed users.

The users' point of view involves analyze the distributed data as they were in the local farm but the analysis in a distributed environment is a complex computing task, hence, the CMS collaboration had developed CRAB to simplify the work of users to create, submit and manage their analysis job in Grid environments, handling handles data discovery, resources availability, job creation and submission, status monitoring and output retrieval. CRAB is an user-friendly tool that helps users to run their analysis code on data available at remote sites, hiding Grid and CMS infrastructure details.

3.4.2 ROOT

ROOT [47] is an analysis package written in an object-oriented structure in C++. It uses built-in functions and user-compiled code to produce graphics and histograms, as well as trees with data objects. ROOT files can be easily navigated using a GUI Object Browser.

The CMS event data model (EDM) is based around the idea that people will use ROOT both in a "framework application", and by interactive manipulation of tree objects and histograms. Interactive ROOT sessions are conducted using an extension to C++ called CINT.

The backbone of the ROOT architecture is a layered class hierarchy with, currently, around 1200 classes grouped in about 60 frameworks (libraries) divided in 19 main categories (modules). This hierarchy is organized in a mostly single-rooted class library, that is, most of the classes inherit from a common base class TObject. It enables the implementation of some essential infrastructure inherited by all descendants of TObject. However, there can also be classes not inheriting from TObject when appropriate.

3.5 Workflow of the Analysis

To perform the analysis described further in this thesis, the whole CMS computing system was used. Monte Carlo simulated samples and real data were equally treated. While the first is generated in the Tier 0, real data was recorded directly in the experiment and then transferred to the Tier 0. The format of the original samples (Monte Carlo and data) is RAW, so these samples are large and contain all the information possible. Afterwards the RAW data is transferred to the Tier-1 sites, in where the re-reconstruction, skimming and selection is performed. The resulting data, in RECO, is then transferred to the Tier-2 in AOD format.

The final physics analysis can be performed in many different ways, in the case of this thesis, it was done mainly in the so-called Analyzer mode. CMSSW provides different modules to perform, for example, the pre-selection of interesting events, sophisticated muon identification and isolation, Jet energy and missing E_T corrections, and so on. The Analyzer mode allows these modules to be added to the user code, in which a concrete kinematical selection oriented to, in this particular case, extract signal events from background in the most efficient way.

Once the Analyzer is tested, jobs are created and submitted using CRAB, accessing then the data stored in the Tier-2 sites worldwide. In this particular case, the Tier-2 site most accessed is the one built and maintained by IFCA (Instituto de Física de Cantabria), in which top and Higgs-related

datasets are stored.

The analysis can be done locally in the Tier-3 machines up to a certain level. It was therefore, partially performed in the Tier-3 in Oviedo, where skimmings of the interesting data streams are stored.

The output of the analyzer is usually a rootfile with histograms, that is opened for further analysis with ROOT. Also some other studies outside the regular analysis chain, are done using ROOT locally.

4

Physics observables

The capabilities of the CMS detector for an efficient and precise reconstruction of the different observables is a key aspect for performing all the planned measurements and even more in the new physics searches.

As the signal signatures studied in this thesis: $t\bar{t} \rightarrow l^+\nu_l l^-\bar{\nu}_l$ and $gg \rightarrow H \rightarrow WW^* \rightarrow l^+\nu_l l^-\bar{\nu}_l$ present 2 high p_T leptons, a significant amount of $\not{E}_T$ and either 2 jets in the first case or small jet activity in the second one, the high level objects used in the analysis are Muons, Electrons, Jets and $\not{E}_T$, the way they are reconstructed and identified by the CMS detector is described in this chapter.

4.1 Muons

4.1.1 Muon reconstruction

The ability to efficiently reconstruct muons over a wide range of energies and in the whole geometric acceptance of the detector is one of the main goals of CMS design.

Muon reconstruction [48] is performed essentially with the silicon tracker and with the specific muon detectors already described in Chapter 3: DTs in the barrel ($|\eta| < 1.2$) and CSCs in the endcaps ($0.9 < |\eta| < 2.4$), complemented by RPCs, which cover both regions up to $|\eta| < 1.6$.

The muon reconstruction starts with track segments from the muon chambers obtained by the local reconstruction, which are formed by matching hits within each chamber, that are reconstructed from digitized electronic signals. Track segments held information about position and direction of muon crossing one chamber.

In the offline reconstruction, these segments reconstructed in the muon chambers are used to generate seeds consisting of position and direction vectors and an estimate of the muon transverse momentum. These initial estimates are used as seeds for the track fits in the muon system, which are performed using segments and hits from DTs, CSCs and RPCs and are based on the Kalman filter technique. The result is a collection of track objects reconstructed in the muon spectrometer, which are referred to as Standalone Muons.

Based on the independently reconstructed tracks in the silicon tracker (tracker tracks) and in the muon spectrometer (standalone-muon tracks), two main reconstruction approaches are used [49]:
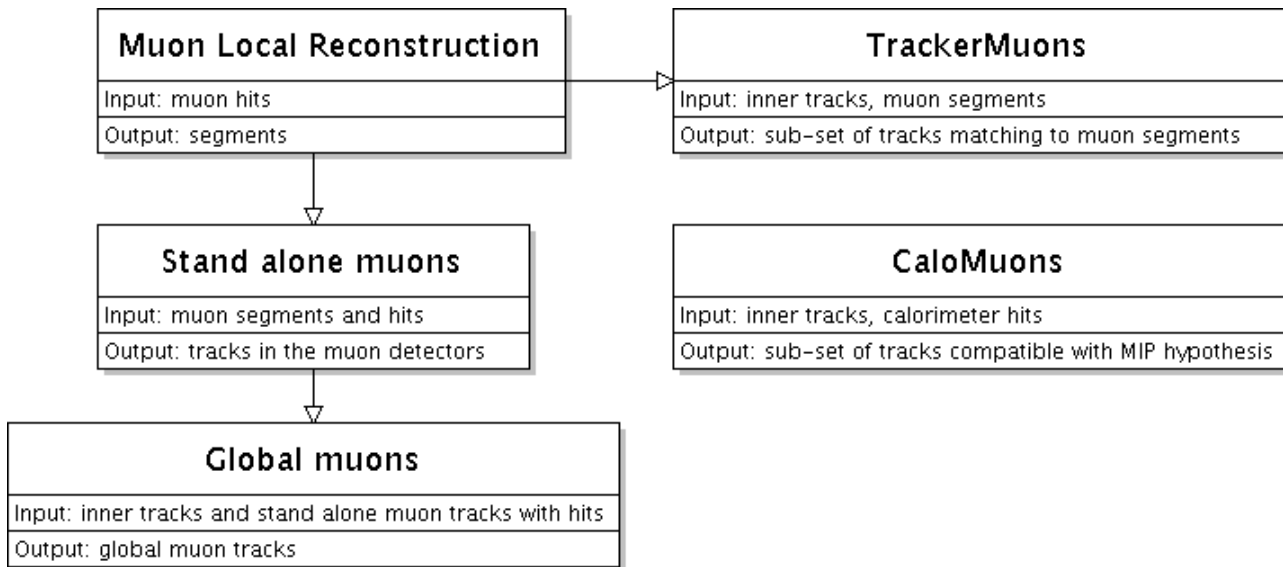
Figure 4.1: - Muon reconstruction overview.

- **Global Muon reconstruction**

  As each sub-detector is able to measure a part of the properties of the muon, the Global Muon reconstruction combines information from multiple sub-detectors in order to obtain the most accurate possible description of the muon, where the track parameters are measured in two sub-detectors: the inner tracker, and the muon system.

  Starting from a standalone reconstructed muon in the muon system, a matching tracker track is found, extrapolating the muon trajectory from the innermost muon station to the outer tracker surface, taking into account the muon energy loss in the material and effect of multiple scattering. Then a global muon track is fitted combining hits from the tracker track and standalone muon track.

  The momentum resolution of muon tracks up to $p_\mathrm{T} = 200$ GeV/c reconstructed in the muon system alone is dominated by multiple scattering, while at low momentum the best resolution is obtained from the silicon tracker. At higher momentum the muon momentum resolution can be improved by combining the muon track from the silicon detector, tracker track, with the muon track from the muon system, standalone muon, into a global muon track.

- **Tracker Muon reconstruction**

  The general idea of Tracker Muons is to reconstruct and identify muons in CMS starting from a silicon tracker track and then searching for compatible segments in the muon detectors. The energy deposition in the calorimeter can also be used for muon identification.

  All tracker tracks with $p_\mathrm{T} > 0.5$ GeV/c and p > 2.5 GeV/c are considered as possible muon candidates and are extrapolated to the muon system, taking into account the expected energy loss and the uncertainty due to multiple scattering. If at least one muon segment matches the extrapolated track in position, the corresponding tracker track qualifies as a tracker-muon track.

  The tracker-muon algorithm is particularly useful for the identification of low p_T muons (with p_T of the order of several GeV/c), which may not leave enough hits in the muon stations for a standalone muon to be reconstructed. At low momentum (roughly p < 5 GeV/c) this approach is

more efficient than the global muon reconstruction, since it requires only a single muon segment in the muon system, while global muon reconstruction typically becomes efficient with two or more segments.

The combination of the different algorithms provides a robust and efficient muon reconstruction.

4.1.2 Muon identification

The standard CMS reconstruction provides additional information for each muon, useful for muon quality selection and identification in physics analysis [50].

To suppress non-prompt muons there is additional track-quality selections cuts on different variables that can be used, as the number of hits associated to the tracker track, standalone muon track and global-muon track, the χ^2 of the various fits, the number of good hits (those having small residual with respect to the track) and the number of layers where a hit would be expected but no hit was found.

The impact parameter of the silicon fit, defined as the distance of closest approach of a track to the primary vertex, can be also used to select prompt-muons. Muons coming from W decays have small values, while it is larger for those coming from secondary vertices, in heavy quark decays.

In the software of CMS there are several muon categories that correspond to different quality requirements. These selectors are based either on parameters used in the muon identification, like the impact parameter, or in the compatibility with different properties in various subdetectors. The following selection types are available:

- *AllGlobalMuons*: checks if the muon is a Global muon

- *AllStandAloneMuons*: checks if the muon is a Standalone muon

- *AllTrackerMuons*: checks if the muon is a Tracker muon

- *TrackerMuonArbitrated*: resolves ambiguity of sharing segments

- *AllArbitrated*: all muons with the Tracker muon arbitrated

- *GlobalMuonPromptTight*: Global muons with tighter fit requirements

- *TMLastStationLoose*: penetration depth loose selector

- *TMLastStationTight*: penetration depth tight selector

- *TM2DCompatibilityLoose*: likelihood based loose selector

- *TM2DCompatibilityTight*: likelihood based tight selector

- *TMOneStationLoose*: requires one well matched segment

- *TMOneStationTight*: require one well matched segment with a tighter criteria

- *TMLastStationOptimizedLowPtLoose*: combination of *TMLastStation* and *TMOneStation*

- *TMLastStationOptimizedLowPtTight*: combination of *TMLastStation* and *TMOneStation*

- *GMTkChiCompatibility*: require tracker stub to have good χ^2 relative to global track

- *GMStaChiCompatibility*: require standalone stub to have good χ^2 compatibility relative to global track

- *GMTkKinkTight*: require a small kink value in the tracker stub

- *TMLastStationAngLoose*: TMLastStationLoose with additional angular cuts

- *TMLastStationAngTight*: TMLastStationTight with additional angular cuts

- *TMOneStationAngLoose*: TMOneStationLoose with additional angular cuts

- *TMOneStationAngTight*: TMOneStationTight with additional angular cuts

- *TMLastStationOptimizedBarrelLowPtLoose*: combination of TMLastStation and TMOneStation with low p_T optimization in barrel only

- *TMLastStationOptimizedBarrelLowPtTight*: combination of TMLastStation and TMOneStation with low p_T optimization in barrel only

The muon identification algorithms recommended by the muon POG are TMLastStationAngTight for muons with p_T below a few GeV and GlobalMuonPromptTight for higher p_T muons.

GlobalMuonPromptTight is designed to suppress hadronic punch-throughs and muons from decays in flight and consists of the following requirements:

- The track is identified as a global muon

- $\chi^2/ndof$ of the global muon fit is smaller than 10

- Number of valid muon-detector hits used in the global fit is greater than 0

4.1.3   Muon isolation

Muons coming from the signal are also expected to be isolated and therefore with no substantial deposits in the tracker and in the calorimeters. The requirement on lepton isolation helps to significantly reduce the background contamination, specially from QCD, $W + jets$ and non-dileptonic $t\bar{t}$.

Tracker and calorimeter depositions within a cone defined around the muon, as shown in Figure 4.2,are taken. The muon is considered isolated if these depositions do not exceed some thresholds.

Different isolation criteria have been studied in CMS. The choice made for the analysis described in this thesis is a relative combined isolation criteria, based on information in the tracker, HCAL and ECAL.

The tracker-based isolation variable is defined as the scalar sum of the p_T of the tracks in the $\eta \times \phi$ plane inside a cone of ΔR 0.3 around the muon (concretely $0.01 < \Delta R < 0.3$ to veto the muon track).

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{4.1}$$

$$SumP_T = \sum_{0.01 < \Delta R < 0.3}^{tracker} P_T^{tracks} \tag{4.2}$$
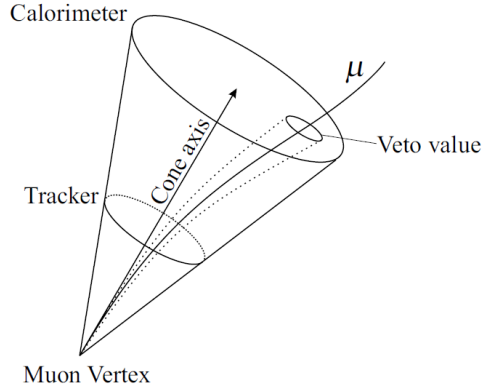
Figure 4.2: - Schematic illustration of the isolation cone. The muon direction at the vertex defines the cone axis.

In the calorimeters, the $hadE_T$ and emE_T variables, corresponding to the HCAL and ECAL respectively, are calculated as the scalar sum of the transverse energy of the calorimeter towers in a cone of ΔR 0.3 around the muon ($0.01 < \Delta R < 0.3$).

$$
\begin{aligned}
hadE_T &= \sum_{0.01<\Delta R<0.3}^{HCAL} E_T^{towers} \\
emE_T &= \sum_{0.01<\Delta R<0.3}^{ECAL} E_T^{towers}
\end{aligned}
\tag{4.3}
$$

Both tracker and calorimeter based information is used to define a relative muon isolation variable as:

$$
Iso_\mu = \frac{SumP_T + emE_T + hadE_T}{p_{T\mu}}
\tag{4.4}
$$

Details on the specific muon identification and isolation criteria chosen for the needs of $t\bar{t}$ and Higgs analysis will be described when discussing the analysis strategies for each of them in Chapters 5 and 6 respectively.

4.2 Electrons

4.2.1 Electron reconstruction

The electron reconstruction [51] is performed using the information from the pixel detector, the silicon strip tracker and the electromagnetic calorimeter, relying on the association of ECAL superclusters with a reconstructed track.

The electromagnetic showers initiated by electrons or photons deposit their energy in several crystals of the ECAL. For a single electron or photon reaching the ECAL, most of the energy is collected

in a small number of crystals.

The presence of the tracker material in front of the calorimeter difficult the energy measurement as it causes electron bremsstrahlung and photon conversions(losing about 70% of their energy). Also because of the strong magnetic field aligned with the beam axis, the energy of the electrons reaching the calorimeter is spread in ϕ, so this is collected using the super-cluster algorithms, this means building a cluster of clusters (supercluster) which is extended in ϕ to minimize the cluster containment variations.

Electrons are built starting from clusters in ECAL. Electron track seeds are built matching superclusters with hits from the standard trajectory seeds. The electron reconstruction relies on a dedicated Gaussian Sum Filter (GSF) method [52; 53] using track and clusters in the ECAL.

There are two complementary algorithms used for the track seeding: Tracker driven, more suitable for low p_{T} electrons and electrons inside jets and ECAL driven. The ECAL driven algorithm starts by the reconstruction of ECAL superclusters of transverse energy $E_T > 4$ GeV and is optimized for isolated electrons in the p_{T} range relevant for Z o W decays and down to p_{T} of 5 GeV/c.

4.2.2 Electron identification

A good electron identification is needed to select electrons from W decays and reject fakes from jets or conversions. This can be achieved by means of variables based on the shower shape and track-cluster matching, as well as the amount of relative hadronic activity. A set of these variables is combined to perform the electron identification, using cuts or multivariate techniques, such as a likelihood or neural network, which is defined by the Electron POG.

For early analysis, the selection has to be as simple, efficient and robust as possible, and the one based on simple cuts is the used in this thesis, which relies on cuts applied in the following four variables:

- H/E: The ratio of the energy deposited in HCAL behind superCluster to superCluster energy. Electrons tend to fully deposit their energy in the electromagnetic calorimeter, while hadrons leave a large fraction of their energy in the HCAL. The ratio is therefore close to zero for electrons and it is higher for hadrons.

- $\Delta\eta$ between superCluster position and track direction at vertex extrapolated to ECAL assuming no radiation

- $\Delta\phi$ between superCluster position and track direction at vertex extrapolated to ECAL assuming no radiation

  These two variables account for the geometrical matching between the track parameters at the interaction vertex extrapolated to the supercluster and the measured supercluster position.

- $\sigma_{i\eta i\eta}$: Cluster shape covariance. It exploits the fact that electromagnetic showers are narrower than hadronic ones. It is a measure of the shower spread in η and is defined as

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i (\eta_i \times 0.0175 + \eta^{seed} - \bar{\eta}_{5\times5})^2 w_i}{\sum_i w_i}}, \qquad (4.5)$$

  where the index i runs over all the crystals on the 5×5 matrix around the seed, 0.0175 is the average crystal η, $\eta_{5\times5}$ is the energy weighted mean η of the 5×5 block of crystals and w_i is the

weight of the ith crystal and is defined as

$$w_i = 4.2 + \ln E_i/E_{5 \times 5} \tag{4.6}$$

where E_i and $E_{5 \times 5}$ are the energy of the ith crystal and the 5×5 block of crystals respectively.

Different cuts are used for electron candidates found in the ECAL barrel and ECAL endcaps.

The Electron POG produced a series of reference selections using Monte Carlo samples of graded severity with efficiency for prompt electrons of $E_T >$20 GeV having nominal values of 95% , 90% , 85% , 80% and 70% . The criteria are optimized in simulation for inclusive $W \rightarrow e\nu$ events and are designed to maximize the rejection of electron candidates from QCD multijet production while maintaining the desired efficiency for electrons from the decay of W/Z leptons.

The analysis described in this thesis use the 90% and 80% working points (they will be refereed as WP90 and WP80) [54] and the cut values are summarized in Table 4.1.

Efficiencies	90%	80%
BARREL		
$\sigma_{i\eta i\eta}$	0.01	0.01
$\Delta\phi$	0.8	0.06
$\Delta\eta$	0.007	0.004
HtoE	0.12	0.04
ENDCAPS		
$\sigma_{i\eta i\eta}$	0.03	0.03
$\Delta\phi$	0.7	0.03
$\Delta\eta$	–	–
HtoE	0.05	0.025

Table 4.1: Cut values applied in the electron ID variables for the two working points used in the analysis presented in this thesis

Other variables are also needed to reject electrons from conversions, as due to the large tracker volume behind the ECAL, this source is not negligible:

- The transverse impact parameter is used to discriminate electrons from conversions as they will have, on average, a greater distance to the beam position.

- An electron from a conversion is also expected to have missing hits in the innermost tracker layers.

- Electrons from conversions are identified by finding the partner conversion track. This is done by looking for an oppositely charged track whose polar angle ($\Delta \cos\theta$) is close to the one of the electron track and whose distance ($Dist$) at the point where the two tracks are parallel is small.

4.2.3   Electron isolation

As for muons, electron isolation is needed to select prompt electrons from W decays and reduce the fake lepton contamination.

The tracker, ECAL and HCAL are also used to compute the isolation variable and the sum of the E_T and p_T components is performed in a cone of 0.3 around the electron direction.

ECAL isolation is calculated using ECAL crystals and additional vetoes are applied to remove tracks and ECAL crystals from the isolation sums to better account for the electron footprint. In the barrel region ($|\eta| < 1.479$) 1 GeV of ECAL energy deposition is subtracted if it is more than 1 GeV to account for the noise pedestal.

The relative combined isolation variable is then defined as:

$$Iso_e = \begin{cases} \frac{\sum_{tracks} P_T + max(0, \sum_{EB} E_T - 1) + \sum_{HCAL} E_T}{max(20., p_{Te})} & \text{for the barrel} \\\\ \frac{\sum_{tracks} P_T + \sum_{EE} E_T + \sum_{HCAL} E_T}{max(20., p_{Te})} & \text{for the endcaps} \end{cases}$$

where EB and EE denotes ECAL barrel and ECAL endcaps respectively.

4.3 Jet reconstruction

Jets are experimental signatures of quarks and gluons. They cannot be observed directly although they can be detected in the tracking and calorimeter systems as they fragment into stable hadrons, appearing in the detector as jets of particles [55] .
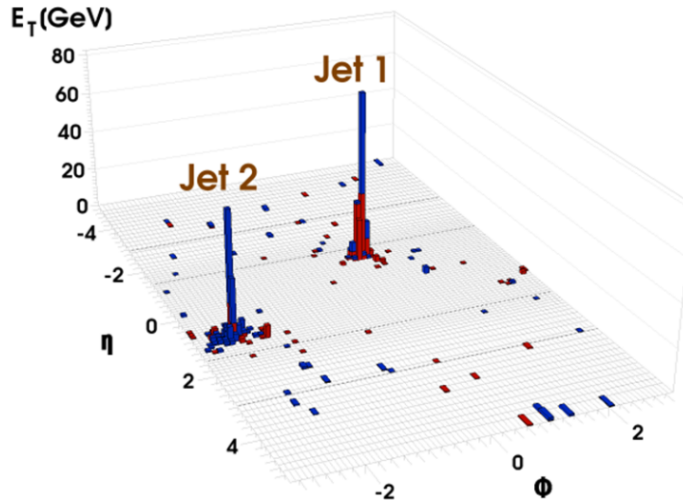
Figure 4.3: - View of the first di-jet event recorded in CMS with $\sqrt{s} = 7$ TeV collision data.

The jet identification and reconstruction is a complicated and sensitive task that involves the calorimeters and the tracking systems. Three main types of jets are reconstructed in CMS, which differently combine individual contributions from subdetectors to form the inputs to the jet clustering algorithm: Calorimeter jets, Jet Plus Track (JPT) jets and Particle Flow (PF) jets:

- **Calorimeter Jets** (CaloJets) are reconstructed using energy deposits in the electromagnetic (ECAL) and hadronic (HCAL) calorimeter cells, combined into calorimeter towers. [56]. A calorimeter tower consists of one or more HCAL cells and the geometrically corresponding ECAL

crystals. The unweighted sum of energy deposits of one single HCAL cell and 5×5 ECAL crystals form a projective tower in the barrel region, $|\eta| < 1.4$. A more complex association between HCAL cells and ECAL crystals is required in the forward region.

Calorimeter jets bring a good description of both the parton-level and the hadron showers emerging from the hard interaction.

- The **Jet Plus Tracks**(JPT) [57] algorithm exploits the excellent performance of the CMS tracking detectors to improve the p_{T} response and resolution of calorimeter jets. Calorimeter jets are reconstructed first as described above, then charged particle tracks are associated with each jet based on spatial separation in $\eta - \phi$ between the jet axis and the track momentum measured at the interaction vertex. The momenta of charged tracks are then used to improve the determination of the energy and direction of the associated calorimeter jet.

- The **Particle Flow**(PFJets) algorithm combines the information from all CMS sub-detectors to identify and reconstruct all particles in the event, namely muons, electrons, photons, charged hadrons and neutral hadrons. The detailed description of the algorithm and its performance can be found in [58]. Charged hadrons, in particular, are reconstructed from tracks in the central tracker. Photons and neutral hadrons are reconstructed from energy clusters in the electromagnetic and hadron calorimeters. Clusters separated from the extrapolated position of tracks in the calorimeters constitute a clear signature of these neutral particles. A neutral particle overlapping with charged particles in the calorimeters can be detected as a calorimeter energy excess with respect to the sum of the associated track momenta. PFJets are then reconstructed from the resulting list of particles. The jet momentum and spatial resolutions are improved with respect to calorimeter jets as the use of the tracking detectors and of the excellent granularity of the ECAL allows to resolve and precisely measure charged hadrons and photons inside jets, which constitute $\sim 90\%$ of the jet energy.

- **Track Jets** are reconstructed from tracks of charged particles measured in the central tracker. Only well measured tracks, based on their association with the primary vertex and their quality, are used by the algorithm. Detailed description of the track jet performance is given in [59]. The calorimeters, in which energy is deposited, collect both the charged and the neutral component of the parton hadronization, however, the strong inner CMS magnetic field, causes the Calo-Towers to fail in collecting the energy of low momentum charged particles which do not reach the calorimeters.

The tracking momentum measurements are more accurate than the calorimeter measurements for charged particles with energies up to several hundreds of GeV, and the direction of charged particles at the interaction point is extremely well determined by the track reconstruction. For these reasons a multi-jet event in CMS is expected to be cleaner when looking at the tracks instead of the CaloTowers, with less overlap and interference, and less background (e.g. pile-up events). On the other hand, as just $\frac{2}{3}$ of the energy of the jet are carried by charged particles, TrackJets are not optimal to measure jet energies, leading to a worse resolution than PF or Calo Jets.

4.3.1 Jet clustering algorithms

Two broad classes of jet algorithms exist, cone and sequential recombination algorithms. In the Cone algorithms, jets are defined as dominant directions of energy flow. A stable cone is defined as a circle of fixed radius R in the $\eta - \phi$ plane such that the sum of all the momenta of the particles

within the cone points in the same direction as the center of the circle. Cone algorithms attempt to identify all the stable cones in the event, and most implementations use a seeded approach to do so, in which iterations are done starting from one seed for the center of the cone until the cone is stable. The set of seeds can be taken as the set of initial particles, like over a P_T threshold or as the midpoints between previously-found stable cones. This iterative method fails to identify all the stable cones, leading to infrared or collinear unsafety in the perturbative computations. Another class of jet algorithms, sequential recombination jet algorithms, define the distance between pairs of particles, performing successive recombinations of the pair of closest particles and stops when all resulting objects are too far apart. Algorithms within that class differ by the definition of the distance. By design, these clustering algorithms are infrared and collinear safe to all orders of calculation.

Several basic algorithms are currently implemented and used in CMS [60]:

- Midpoint Cone algorithm

- Iterative Cone algorithm

- Seedless Infrared Safe Cone algorithm

- k_T algorithm

- Anti-k_T algorithm

4.3.1.1 Iterative Cone

The Iterative Cone algorithm is a simple cone-based algorithm. It has a short and predictable execution time in CMS. Calorimeter towers and particles with $E_T > 1$ GeV are considered in descending order as starting points, seeds, for an iterative search for stable cones such that all inputs with $\Delta R \leq$ that certain value are considered. The cone size parameter, ΔR, is defined as follows:

$$\Delta R = \sqrt{\Delta \Phi^2 + \Delta \eta^2} \tag{4.7}$$

A cone is considered stable if its geometric center agrees with the (Φ, η) location of the sum of the constituent four vectors within a certain tolerance. Once a stable cone is found, it is declared a jet and its constituents are removed from the remaining inputs. The algorithm is neither collinear- nor infrared-safe.

4.3.1.2 Midpoint Cone

The Midpoint Cone algorithm is based as well on an iterative procedure to find stable cones. Unlike in the Iterative Cone algorithm, the infrared safety is addressed by considering the midpoints between each pair of proto-jets which are closer than twice the cone radius R as additional seeds. Each input object, like a tower, particle, etc. can initially be associated with several proto-jets, and a splitting and merging algorithm is applied afterwards to ensure each input object appears in one jet only. Despite these improvements to the cone-based clustering procedure, the algorithm has been shown not to be infrared-safe for QCD orders beyond NLO.

4.3.1.3 Seedless Infrared-Safe Cone

The Seedless Infrared-Safe Cone (SISCone) [61], jet algorithm is the first cone algorithm that is at the same time infrared and collinear safe to all orders of QCD, while fast enough to be used in experimental analysis.

4.3.1.4 k_t

The k_t algorithm [62; 63; 64] is a recombination algorithm that successively merges pairs of nearby objects, namely partons, particles, or calorimeter towers, in order of increasing relative transverse momentum. A single parameter, D, which approximately characterizes the size of the resulting jets, determines when this merging stops. No splitting or merging is involved because each object is uniquely assigned to a jet.

In the case of the k_t, the definition of the distance between the objects that determines the algorithm is given by:

$$d_{ij}^2 = min(k_{t,i}^{2p}, k_{t,j}^{2p})(\Delta y_{ij}^2 + \Delta \Phi_{ij}^2) \tag{4.8}$$

where $p = 1$. If $p = 0$, it results in the Cambridge/Aachen [65; 66]algorithm.

4.3.1.5 **Anti-**k_t

The Anti-k_t algorithm [67] is a third variation of the k_t. Considering again the distance:

$$d_{ij}^2 = min(k_{t,i}^{2p}, k_{t,j}^{2p})(\Delta y_{ij}^2 + \Delta \Phi_{ij}^2) \tag{4.9}$$

in the Anti-k_t, $p = -1$.

Here, soft particles are recombined with hard ones before recombining among themselves, resulting in regular, soft-resilient, hard jets. This the hallmark of the iterative cone and, in that respect, the Anti-k_T can be seen as an infrared and collinear safe replacement.

The Anti-k_t is the current CMS recommendation for physics analysis with jets, and therefore will be used by default in this study.

4.3.2 Jet energy scale corrections

As the calorimeter response to particles is not linear, the energy of a jet that is reconstructed and measured from the detector is different than that of the corresponding particle jet, which is obtained in the simulation by clustering, with the same algorithm, the stable particles produced during the hadronization process that follows the hard interaction. Besides, electronic noise and pile-up events can lead also to extra unwanted energy.

The purpose of the jet energy calibration is to relate, on average, the energy measured in the detector jet to the energy of the corresponding particle jet [56]. The correction is applied as a multiplicative factor : $P_T^{corrected} = C(P_T^{RAW}, \eta) \cdot P_T^{RAW}$.

CMS has adopted a factorized solution to the problem of jet energy corrections, where each level of correction takes care of a different effect, shown in Figure 4.4. Each level of correction is essentially a scaling of the jet four momentum with a factor depending on various jet-related quantities, such as P_T, η, flavor, etc., and they are applied sequentially with a fixed order.

The energy of jets considered for the analysis developed in this thesis, as well as for any physics analysis to be performed in CMS, has to be corrected. The combination of L2(Relative) + L3(Absolute) jet corrections is currently the default correction in CMS, and is therefore required here.
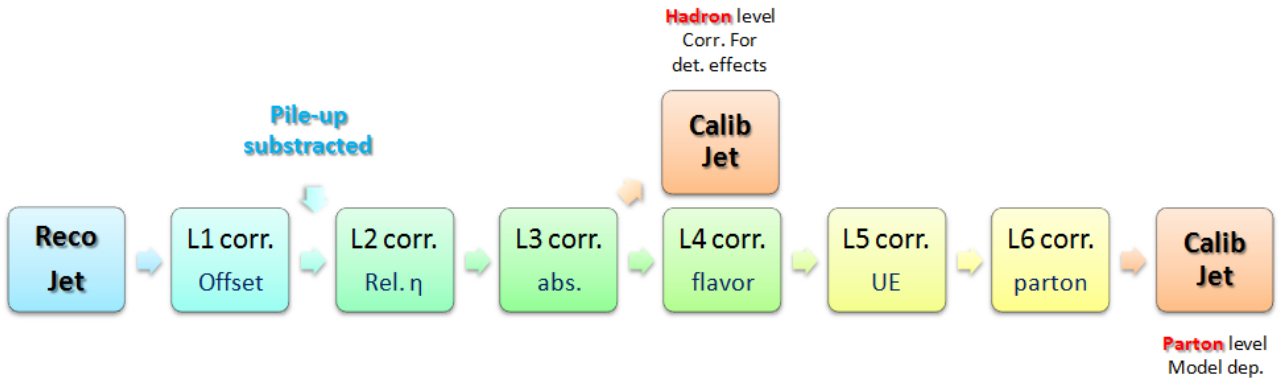
Figure 4.4: - Sequential application of the different levels of correction from Reconstructed to Calibrated Jet in CMS.

The L2 Relative correction makes the jet response flat vs η. The uniformity in pseudorapidity is achieved by correcting a jet in arbitrary η relative to a jet in the central region ($|\eta| < 1.3$). The derivation of the Relative correction is done either by using MC truth or by employing a data driven method (di-jet balance).

The L3 Absolute correction makes the jet response flat vs p_T. Once a jet has been corrected for η dependence (L2 relative correction), it is corrected back to particle level (this means that the corrected CaloJet p_T is equal on average to the Generated Jet p_T). The derivation of the Absolute correction is also done either by using MC truth information or by employing data driven techniques ($Z/\gamma + jet$ balance).

The combination of L2(Relative)+L3(Absolute) MC derived jet corrections is required for MC, while for data an additional L2 L3 Residual correction derived from data is applied.

The detailed understanding of the jet energy calibration is of crucial importance and a leading source of systematic uncertainty for many physics analysis at LHC.

The determination of the jet energy calibration in CMS has been studied by the JetMET POG, results can be found in [56], where detailed studies of the jet calibration precision performed with 3 pb^{-1} of data yield a 3 - 6 % uncertainty of the overall jet energy scale in a wide region of jet p_T from 30 (20, 15) GeV up to 2 TeV for Calo (JPT, PF) jets. This can be observed in Figures 4.5 and 4.6.

4.4 Missing E_T

Neutral weakly interacting particles such as neutrinos normally escape from a detector without producing any direct response in the detector elements. The presence of such particles in collider experiments must be inferred from the imbalance of total momentum [68].

The missing E_T ($\not{E}_T$) is calculated as the negative of the vector sum of the components of momentum transverse to the beam axis of all final-state particles reconstructed in the detector. There are several methods developed in CMS to reconstruct the $\not{E}_T$: the standard calorimeter $\not{E}_T$ (Calo$\not{E}_T$), the Track-corrected $\not{E}_T$ (tc$\not{E}_T$), and the Particle Flow $\not{E}_T$ (PF$\not{E}_T$). In the following the characteristics and main differences of each one will be presented.

Figure 4.5: - Overall uncertainty on the absolute jet energy scale for CALO jets (left) and JPT jets (right). The uncertainty for each algorithm is shown down to the lowest recommended p_T value.

Figure 4.6: - Overall uncertainty on the absolute jet energy scale for PF jets. The uncertainty for each algorithm is shown down to the lowest recommended p_T value.

4. PHYSICS OBSERVABLES

4.4.1 Calorimeter Missing E_T
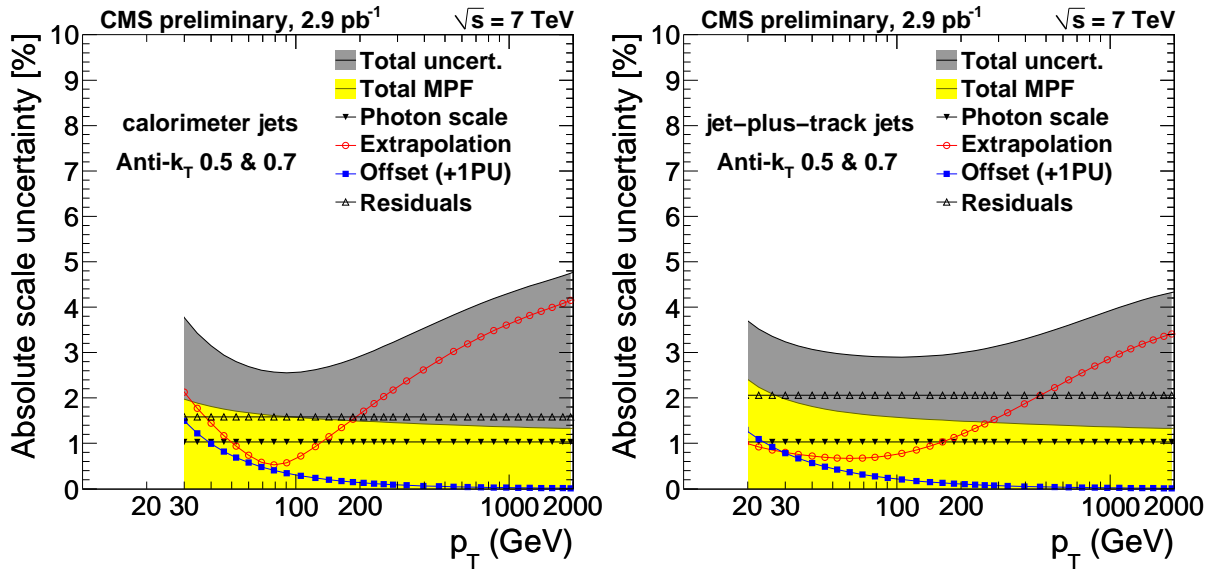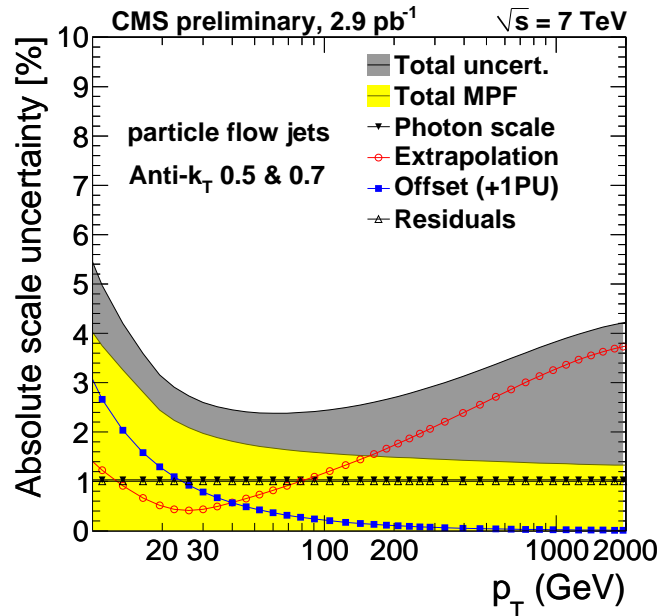
The calorimeter $\not{E}_T$ (Calo$\not{E}_T$) is determined from the transverse vector sum over energy deposits in uncorrected projective Calorimeter Towers, in both the ECAL and the HCAL [69], considering only energy deposits above noise threshold.

This calculation can be improved by correcting it for several effects. Since a muon deposits only about 2 GeV in the calorimeter, independent of its momentum, the muon p_T measured by the inner and outer spectrometer must be added into the Calo$\not{E}_T$ while simultaneously removing the small calorimetric energy deposit associated to the muon track. No correction is needed for the electron, whose energy is reconstructed without bias by the electromagnetic calorimeter.

There are many other corrections to the calorimeter $\not{E}_T$. The jet energy scale corrections (called Type I corrections), can be applied to account for the nonlinearity response of the calorimeter to hadronic particles. In addition, towers not associated to any jet can be corrected for calorimeter nonlinearity (called Type-II corrections).

4.4.2 Track Corrected Missing E_T

Using not just the calorimeters, but other CMS subdetectors, fundamental improvements to calorimeter-based $\not{E}_T$ can be made. That is the case of the Track corrected $\not{E}_T$ (tcMET) [70], in which tracker information is used.

The Track Corrected $\not{E}_T$ improves the resolution and tails replacing the expected energy deposition in the calorimeter of all well reconstructed tracks by their measured momentum in the tracker. Since at the momentum scale of the order of 1 GeV the tracker has a superior resolution compared to the calorimeter, this leads to significant improvements.

It starts from the Calo$\not{E}_T$ algorithm described above, and adds in the p_T of the remaining tracks which have been reconstructed in the inner tracker, while subtracting the expected calorimetric energy deposit of each track. All tracks are treated as pions in this process. The calorimetric energy deposit has been estimated from a single-pion Monte Carlo sample, in bins of p_T and η and taking into account the track extrapolation in the CMS magnetic field to determine the expected position of the calorimetric energy deposit. No correction is applied for very high-p_T tracks (pt $> 100 GeV/c$), whose energy is already well measured by the calorimeters. Low-p_T tracks ($p_T < 2 GeV/c$), are fully compensated for, assuming no response from the calorimeter.

4.4.3 Particle Flow Missing E_T

A more refined approach consists of using the individual particles that have been reconstructed using the Particle Flow technique [71].

As in the jet reconstruction, the particle-flow $\not{E}_T$ reconstruction algorithm is intended to provide a complete and unique event description at the level of individually reconstructed particles, with an optimal combination of information across all CMS sub-detectors. The reconstructed and identified individual particle list includes muons, electrons (with individual Bremsstrahlung photons), photons (either unconverted or converted), charged hadrons (without or with a nuclear interaction in the tracker material), as well as stable and unstable neutral hadrons. The PF $\not{E}_T$ is then simply the negative vector sum of all such reconstructed particles in the event.

4.5   b-tagging

The identification of jets coming from the hadronization of a b-quark is an important tool for a wide range of physics analysis in the SM and beyond, among them the $t\bar{t}$ which contains real b-quarks and also the Higgs analysis in which the $t\bar{t}$ constitutes an important background and is essential to reject it.

The CMS detector, with its precise charged particle tracking system, robust lepton identification and finely segmented calorimetry, is well matched to the task of b-jet identification (b-tagging).

There are several algorithms (taggers) of b-jet identification in CMS, which exploit the long lifetime and the presence of decay leptons to discriminate these jets from light-jets (those originating from u,d,s quarks or gluons).

These taggers are based on several observables that can be used to distinguish b-jets from light-jets.

The impact parameter (IP) is the most powerful single-track observable, which is the distance between the track and the vertex at the point of closest approach. (For B hadrons with finite lifetime, the IP is Lorentz invariant and the typical scale is set by $c\tau$ approx 480 μm). The IP can be calculated either in the transverse plane or in 3D, which can be used thanks to the good z resolution provided by the CMS pixel system. Given that the uncertainty can be of the same order of magnitude as the IP, a better observable for b-tagging is the impact parameter significance defined as IP/σ_{IP} . The IP is "life time signed". The IP sign is obtained from the sign of the scalar product of the IP segment with the jet direction. A sign flip can happen due to differences between the reconstructed jet axis and the true B hadron flight direction. For decays without a sizeable lifetime, the IP is expected to be symmetric with respect to zero, while for B hadrons decaying weakly, it is mostly positive.

This can be also used to define a track by track probability (P_{tr}), by extracting the probability density function for tracks not coming from b-jets.

For a set of tracks, one can directly search for a secondary vertex from the B hadron decay. The same tool used for the primary vertex finding is used on all the tracks associated to the jet. Vertices with at least 65% of tracks shared with the primary vertex are removed from the list

The several b-tagging algorithms developed by CMS are described in [72]. Each produces as output a numerical discriminator, which can be used to select jets according to the priority for efficiency versus purity of the resulting sample. This discriminator can be a physical quantity like the IP significant for some taggers or a complex variable like the output of a likelihood ratio or neural network.

- **The Track Counting** approach identifies a jet as a b-jet if there are at least N tracks each with a significance of the impact parameter exceeding S. The way of producing a continuous discriminator for this algorithm is to fix the value of N, and consider as discriminating variable the impact parameter significance of the Nth track (ordered in decreasing significance). If one is interested in a high efficiency for b-jets, the second track can be used; for higher purity selections the third track is a better choice. The discriminators associated with N = 2 and N = 3 are called Track Counting High Efficiency (TCHE) and Track Counting High Purity (TCHP), respectively.

- **The jet probability** algorithms combine information from all selected tracks in the jet. There are two different discriminators available called Jet probability and Jet B probability. The first one is related to the combined probability that all the tracks in the jet come from the primary vertex, defined as $P_{jet} = \prod \cdot \sum_{j=0}^{N-1} \frac{(-ln\pi)^j}{j!}$ where $\pi = \prod_{i=1}^{N} P_{tr}(i)$. The second one estimates
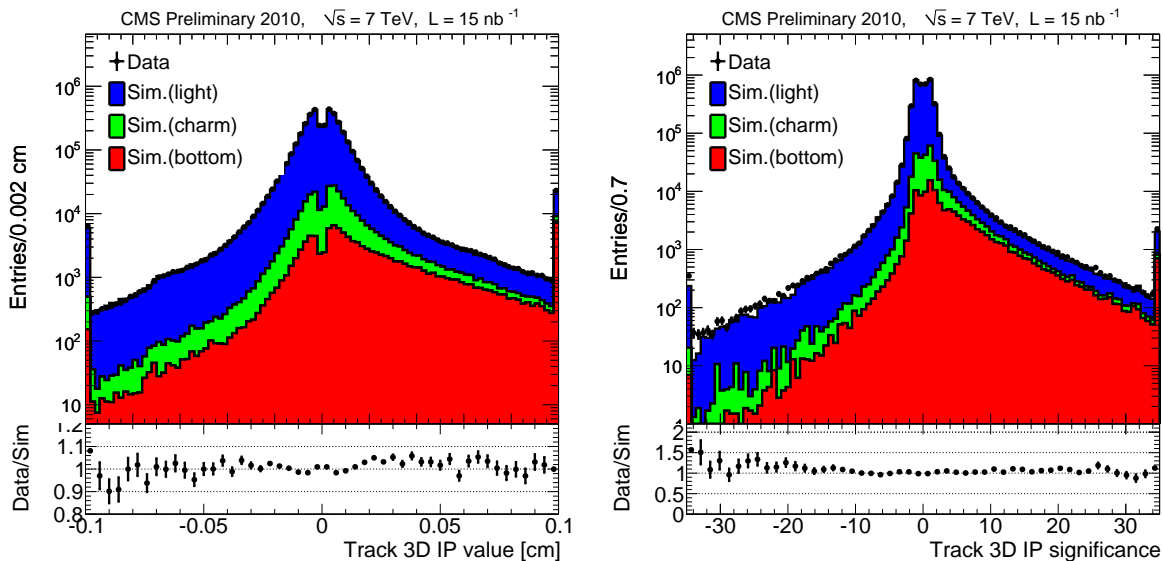
Figure 4.7: - IP value and IP significance for all selected tracks with 15 nb^{-1} of data

how likely is that the four most displaced tracks are compatible with the primary vertex, four being the average number of reconstructed tracks from a b-hadron decay.

- **Lepton-based tagging** algorithms identify b hadrons via their semileptonic decays, as the presence of a lepton close to the jet is already a hint of a weak decay of a B hadron. Discrimination between b- and light-flavor jets can be achieved based on the p_T of the lepton with respect to the jet direction, the impact parameter of the lepton, or both.

- **The Simple Secondary Vertex (SSV)** algorithm is based on the reconstruction of at least one secondary vertex. The significance of the 3D flight distance is used as a discriminating variable for this tagger. There two variants based on the minimum number of tracks attached to the vertex: High Efficiency (SSVHE) if there are at least 2 tracks and High Purity (SSVHP) if there are at least 3.

- A more complex approach involves the use of secondary vertices, together with other lifetime information, like the IP significance or decay lengths. By using these additional variables, the Combined secondary vertex algorithm provides discrimination even when no secondary vertices are found, so the maximum possible b-tagging efficiency is not limited by the secondary vertex reconstruction efficiency.

For each of these taggers, several operating points are defined by the b-tag POG by choosing three values for each discriminant, Loose (L), Medium (M), and Tight (T), being the value at which the acceptance of light partons is estimated from MC to be 10% , 1% , or 0.1% , respectively. B-tagging and mistag efficiencies for several of the above mentioned taggers is shown in Figures 4.9.

The b-tag POG takes care of measuring the efficiencies and mistag rates from data and then provide data to MC scale factors, together with their uncertainties, for several of the defined working points, to be later used in the analysis [73].
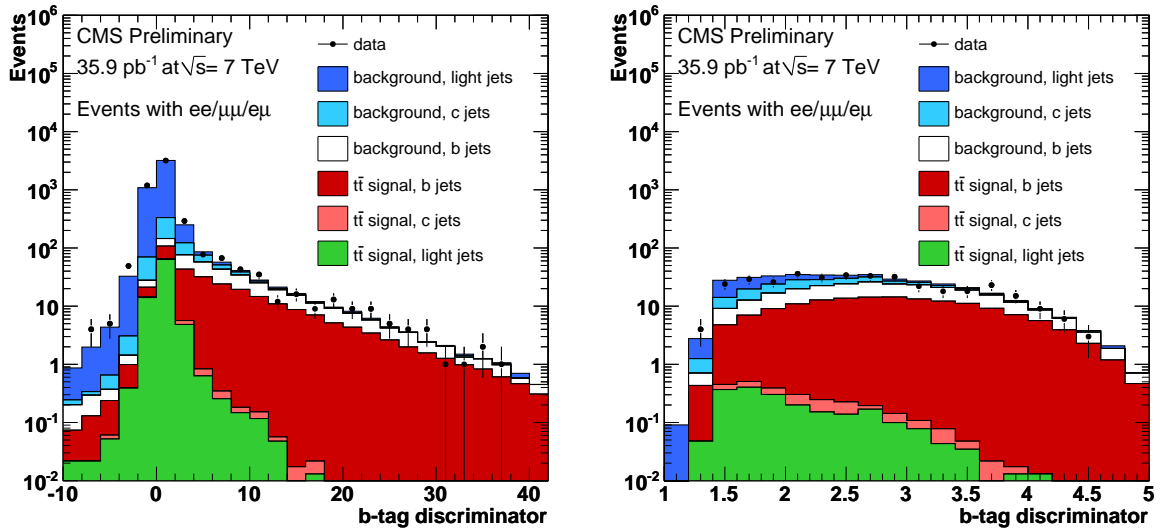
Figure 4.8: - Distribution of the TCHE(left) and SSVHE(right) discriminants, comparing data and simulation with 35.9 pb^{-1}
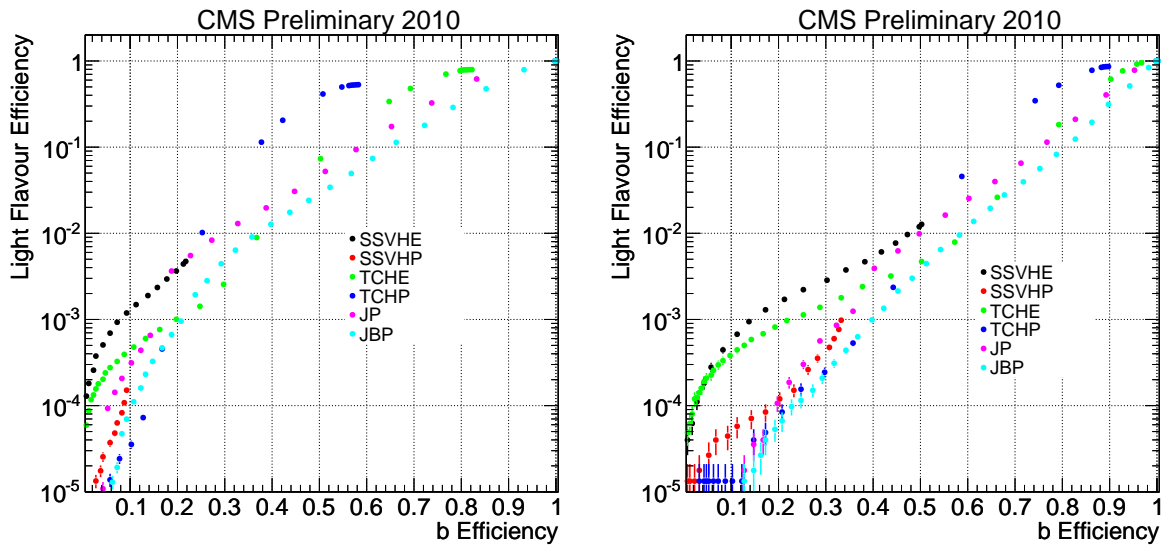
Figure 4.9: - Efficiency for obtaining a b tag of a non-b vs true b jet for each of the tagging algorithms. Left: For jets with $10 < p_T < 30$ GeV/c, Right: For jets with $30 < p_T < 50$ GeV/c

5

Measurement of the $t\bar{t}$ cross section with b-tagging

5.1 Introduction

In this thesis the measurement of the $t\bar{t}$ cross-section in dilepton final states: $\mu^+\mu^-$, e^+e^- and $e^\pm\mu^\mp$ (including leptons from τ decays) is described, using the full dataset collected by the CMS detector during the 2010 run of the LHC, which corresponds to 35.9 ± 1.4 pb^{-1} of integrated luminosity.

Next to leading order (NLO) calculations [74], [75] of the $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using MCFM [76], [77] predict a cross section value of $\sigma_{t\bar{t}} = 157.5^{+23.2}_{-24.4}$ pb for a top quark mass of $m_t = 172.5$ GeV/c^2 , which is dominated by the gluon gluon fusion mechanism, as described in section 2.6.1.

The dilepton signature is quite clean, thus the event selection used to extract the signal is based on simple sequential cuts and the cross section is computed using a counting method, the events observed in data are compared to the SM backgrounds and the excess is assumed to come from the $t\bar{t}$ signal.

A summary of the data samples used, the trigger and object selection, the observed yields and expectation from Monte Carlo simulation is shown first, followed by the description of methods used to estimate the background from data ($Z/\gamma^* \to l^+l^-$ and fake leptons). Then all the sources of systematic uncertainties considered are discussed and finally the measured cross section value is presented.

Top quark events play an important role in the commissioning of the detector, as they contain most of the relevant objects that need to be understood since the beginning of the data taking: its decay signature involves leptons, jets and $\not{E}_T$, requiring the precise understanding and operation of all parts of the CMS detector.

Rediscovery of the SM is a major goal in the early phase of the LHC, understanding of the top quark properties is a test of the performance of the detector and a proof of the readiness for starting new physics searches beyond the SM, so top quark analysis have been defined as one of the benchmark analysis and are a crucial component of the CMS physics program, a primary goal is the first observation of top-quark pair production and the measurement of its production cross section, which will be followed by precision measurements of its properties.

The path towards the first $t\bar{t}$ cross section measurements is described in this chapter, from a brief

introduction describing the preparations based on simulation to the analysis of the first data and finally the measurement using the full dataset collected during 2010.

5.2   Signal and background processes

The $t\bar{t}$ dilepton is a quite clean final state, as described in Section 2.6 it is characterized by the presence of two high p_{T} and isolated leptons coming from the W boson decay, a large amount of real $\not{E}_T$ due to the 2 undetected neutrinos, and also 2 b-jets, whose presence can be exploited to reduce the amount of background to a small level and thus obtain a pure sample of $t\bar{t}$ events.

The main sources of background arise from events with mis-identified leptons, fake leptons passing the selections, (as W+jets or QCD. $t\bar{t}$ semileptonic events also contribute here) and events with mis-measured MET (as Z+jets). Contributions from backgrounds with real leptons and $\not{E}_T$ are irreducible, and are mainly dibosons or single top tW production. The backgrounds considered in the analysis are the following:

- **Drell Yan** $Z/\gamma^* \to l^+l^-$
  It presents 2 high p_{T} , opposite charged and isolated leptons, but it doesn't have real $\not{E}_T$, as there are no neutrinos in the final state. Therefore, the main variable to reject it is the $\not{E}_T$, so a good control and understanding of its performance is essential as this background has one of the largest cross sections: 18417 pb. The presence of high p_{T} jets in the event also helps to reduce it, as it has less hadronic activity, this feature also helps to reduce the $Z/\gamma^* \to \tau^+\tau^-$, which has real $\not{E}_T$ coming from the neutrinos produced in the tau decays.
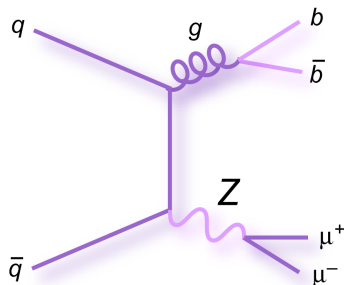
Figure 5.1: - Example of a feynman diagram for the Z+jets process

- **Dibosons: WW, WZ, ZZ**
  WW production, when both W bosons decay leptonically also presents 2 isolated leptons and real $\not{E}_T$, the only difference with respect to the $t\bar{t}$ is the presence of high p_{T} jets coming from the hadronization of the 2 b-quarks, so this can be used to reduce this background, together with the b-tagging requirement.

  WZ and ZZ can present slightly different signatures depending on how both bosons decay, the $W \to l\nu$ branching ratio is about 10% per lepton flavour and 3% in the case of $Z \to l^+l^-$, so the main contributions will come from:

  $W^{\pm}Z \to q\bar{q}l^+l^-$, $W^{\pm}Z \to l^{\pm}\nu l^+l^-$, $ZZ \to l^+l^- q\bar{q}$, $ZZ \to l^+l^- \nu\bar{\nu}$, $ZZ \to l^+l^-l^+l^-$.

  In both cases the $\not{E}_T$ requirement will help to reduce these backgrounds, as well as the b-tagging. Anyway, the cross section is small compared to the signal: 4.51 pb the WW decaying to 2 leptons, 0.61 pb the WZ decaying to 3 leptons and 7.4 pb the inclusive ZZ production.
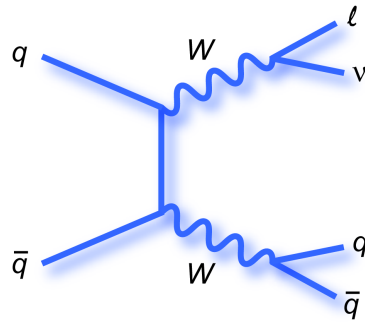
Figure 5.2: - Example of a feynman diagram for the WW process

- **Single Top**
  It is one of the main remaining backgrounds after the full selection is applied. The tW channel is the one relevant and it has almost the same signature as the $t\bar{t}$ process, it presents 2 isolated leptons, $\not{E}_T$ from the neutrinos and jets, although it has only one b-jet, so the b-tagging helps to reduce this background, specially if the presence of at least 2 b-tagged jets is required in the selection. The cross section is about one order of magnitude smaller than the signal: 10.6 pb.
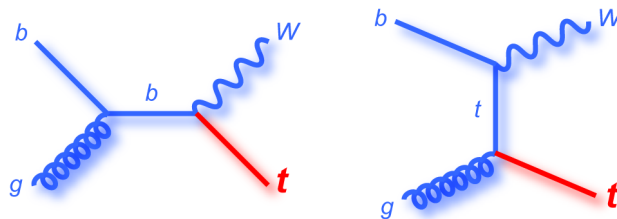
Figure 5.3: - Example of a feynman diagram for the single top- tW channel production

- **W+jets**
  This background has only 1 isolated lepton, but due to its huge cross section (31314 pb), it can represent a dangerous background when one non-prompt lepton is passing the selection, so a good control on the identification and isolation must be performed to reduce the fake lepton contamination as much as possible.
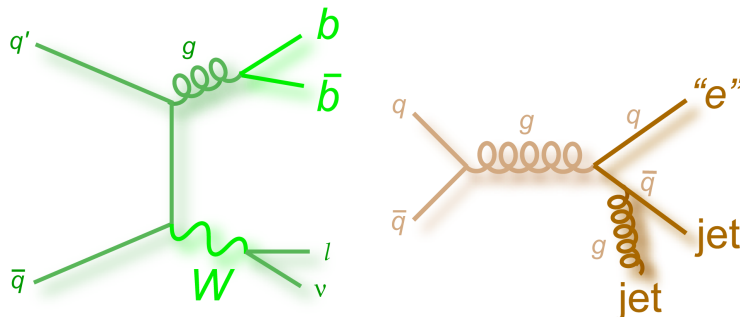
Figure 5.4: - Example of a feynman diagram for the W+jets and multijets production

In order to select electrons and muons from W bosons from $t\bar{t}$ decays, events are required to have two opposite charged, energetic, well identified and isolated leptons. To suppress the Z+jets

contribution the e^+e^- and $\mu^+\mu^-$ events are required to have a reconstructed invariant mass outside the Z mass window, and also to have a significant $\not{E}_T$ typical of signal due to presence of neutrinos in the final state. Further suppression of all backgrounds is achieved by requiring presence of energetic jets in the event, even more if they are requited to be b-tagged.

5.3   Studies based on MC before the start of the data taking

The potential of the CMS detector to measure the $t\bar{t}$ production cross section has been studied in detail based on simulation over the past years. The analysis strategy has evolved and has been tuned according to the different proposed scenarios and changes in the planned LHC running conditions (from the design energy of 14 TeV, to 10 TeV and finally 7 TeV), as well as to the better knowledge of the detector.

The initial centre of mass energy for the LHC was planned to be 10 TeV in 2009, therefore all the analysis were optimized accordingly and assuming of the order of 100-200 pb^{-1} of integrated luminosity. Finally, after the announcement of the new running conditions for 2010, the analysis was adapted to 7 TeV and optimized for the first 10 pb^{-1} of data.

The main impact of changing the energy is the reduction of the cross section, so more luminosity is needed to collect the same number of $t\bar{t}$ events. A comparison of the expected values at $\sqrt{s} =$ 14, 10 and 7 TeV is shown in Table 5.1.

Figure 5.5 shows the b-tag multiplicity in events passing full dilepton selection at $\sqrt{s} = 10$ TeV and 7 TeV.

This section summarizes some of the MC-based studies done in preparation for the data taking. Tables 5.2, 5.3 and 5.4 show the number of expected events in $\mu^+\mu^-$, e^+e^- and $e^\pm\mu^\mp$ channels respectively for 100 pb^{-1} of integrated luminosity at $\sqrt{s} = 10$ TeV.

$\sqrt{s}$	NLO Cross section (pb)
14	908
10	414
7	157.5

Table 5.1: $t\bar{t}$ production cross section at different centre of mass energies

Analysis selection, trigger efficiencies, data-driven techniques for background estimation and methods for systematic uncertainty estimates were studied and defined in simulation [78]. Since then, the lepton identification criteria has improved, also the isolation cuts have been tuned and more jet and $\not{E}_T$ reconstruction algorithms have been developed in CMS, this improvements have been systematically incorporated to the analysis.

Different studies were done based on the official CMS MC production at $\sqrt{s} = 7$ TeV in the Summer of year 2009, in order to optimize the analysis previously to the data taking.

Several jet and $\not{E}_T$ algorithms and different approaches were explored: Calorimeter based analysis (using caloJets and calo$\not{E}_T$), particle flow based (using PFJets and PF$\not{E}_T$) and track-based (using
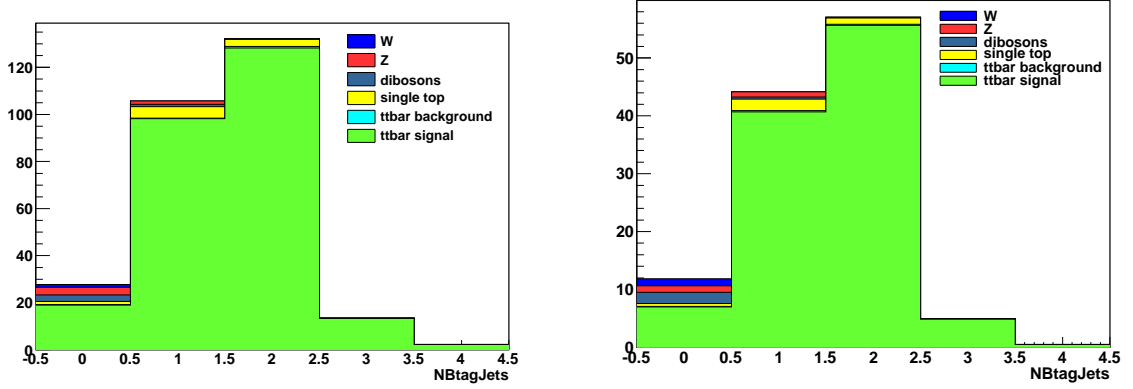
Figure 5.5: - Number of b-tagged jets in $e^{\pm}\mu^{\mp}$events passing full selection criteria before b-tagging cut expected from simulation in 100 $pb^{-1}$of integrated luminosity at $\sqrt{s} = 10$ TeV (left) and $\sqrt{s} = 7$ TeV (right)

	diLepton	Z Veto	≥ 2 jets	MET	≥ 1 btag	≥ 2 btag
$t\bar{t}$ signal	261± 4	201± 4	148± 3	97± 3	90± 3	54± 2
$t\bar{t}$ bkg	0.4± 0.2	0.3± 0.2	0.3± 0.2	0.3± 0.2	0.2± 0.1	0.08± 0.08
WW	37.4± 0.2	28.9± 0.2	1.93± 0.04	1.14± 0.03	0.39± 0.02	0.074± 0.008
WZ	46.0± 0.2	4.85± 0.06	1.16± 0.03	0.29± 0.01	0.097± 0.008	0.016± 0.003
ZZ	30.2± 0.1	3.28± 0.03	1.22± 0.02	0.170± 0.008	0.074± 0.005	0.021± 0.003
$Z\mu\mu$	50343± 78	4530± 23	174± 5	9± 1	3.5± 0.6	0.5± 0.2
$Z\tau\tau$	118± 4	112± 4	5.1± 0.8	1.8± 0.5	0.5± 0.2	0.1± 0.1
WJets	1.0± 0.7	0.5± 0.5	0	0	0	0
Single top	20.0± 0.4	15.5± 0.3	5.2± 0.2	3.4± 0.2	3.0± 0.1	1.25± 0.09

Table 5.2: Number of expected events for 100 pb^{-1}at 10 TeV in $\mu^{+}\mu^{-}$channel.

	diLepton	Z Veto	≥ 2 jets	MET	≥ 1 btag	≥ 2 btag
$t\bar{t}$ signal	164± 4	127± 3	90± 3	57± 2	54± 2	32± 2
$t\bar{t}$ bkg	1.2± 0.3	1.0± 0.3	0.9± 0.3	0.7± 0.2	0.7± 0.2	0.5± 0.2
WW	23.6± 0.1	18.3± 0.1	1.18± 0.03	0.67± 0.02	0.24± 0.01	0.036± 0.006
WZ	28.7.0± 0.1	3.06± 0.04	0.90± 0.02	0.24± 0.01	0.078± 0.007	0.008± 0.002
ZZ	18.29± 0.08	1.95± 0.03	0.89± 0.02	0.091± 0.006	0.032± 0.003	0.008± 0.002
Zee	30555± 52	2427± 15	93± 3	4.1± 0.6	0.9± 0.3	0.09± 0.09
$Z\tau\tau$	63± 3	60± 3	3.1± 0.6	1.5± 0.4	0.7± 0.3	0.1± 0.1
WJets	17± 3	14± 3	1.6± 0.9	0.5± 0.5	0	0
Single top	13.7± 0.3	10.8± 0.3	3.6± 0.2	2.2± 0.1	1.9± 0.1	0.68± 0.07

Table 5.3: Number of expected events for 100 pb^{-1}at 10 TeV in $e^{+}e^{-}$channel.

	diLepton	≥ 2 jets	MET	≥ 1 btag	≥ 2 btag
$t\bar{t}$ signal	413$\pm$ 6	300$\pm$ 5	261$\pm$ 4	242$\pm$ 4	144$\pm$ 3
$t\bar{t}$ bkg	1.5$\pm$ 0.3	1.2$\pm$ 0.3	1.1$\pm$ 0.3	0.9$\pm$ 0.3	0.7$\pm$ 0.2
WW	60.8$\pm$ 0.2	3.94$\pm$ 0.06	3.15$\pm$ 0.05	1.01$\pm$ 0.03	0.17$\pm$ 0.01
WZ	5.55$\pm$ 0.06	0.96$\pm$ 0.02	0.70$\pm$ 0.02	0.19$\pm$ 0.01	0.029$\pm$ 0.004
ZZ	1.22$\pm$ 0.02	0.33$\pm$ 0.01	0.133$\pm$ 0.007	0.040$\pm$ 0.004	0.008$\pm$ 0.002
$Z\mu\mu ee$	10$\pm$ 1	0.4$\pm$ 0.2	0	0	0
$Z\tau\tau$	184$\pm$ 5	7.3$\pm$ 0.9	4.7$\pm$ 0.7	1.5$\pm$ 0.4	0.1$\pm$ 0.1
WJets	19$\pm$ 4	1.2$\pm$ 0.9	1.2$\pm$ 0.9	0	0
Single top	33.0$\pm$ 0.5	11.1$\pm$ 0.3	9.4$\pm$ 0.2	8.0$\pm$ 0.2	3.1$\pm$ 0.1

Table 5.4: Number of expected events for 100 pb^{-1}at 10 TeV in $e^{\pm}\mu^{\mp}$channel

trackJets and track-corrected $\not{E}_T$). This is reported in Tables 5.5,5.6 and 5.7.

Figure 5.6 shows the performance of the different $\not{E}_T$ reconstruction algorithms for the $t\bar{t}$ signal and $Z/\gamma^* \to l^+l^-$ events. In the first case, as it presents real $\not{E}_T$ the behaviour of the 3 algorithms is similar, while in the second case with no genuine $\not{E}_T$, track$\not{E}_T$ and PF$\not{E}_T$ are performing better.

	diLepton	Z Veto	≥ 2 CaloJets	≥ 2 PFJets	≥ 2 TrackJets
$t\bar{t}$ signal	11.8$\pm$ 0.2	8.9$\pm$ 0.2	6.5$\pm$ 0.1	6.4$\pm$ 0.1	6.0$\pm$ 0.1
$t\bar{t}$ bkg	0.04$\pm$ 0.01	0.022$\pm$ 0.008	0.020$\pm$ 0.007	0.020$\pm$ 0.007	0.014$\pm$ 0.006
WW	2.5$\pm$ 0.1	1.88$\pm$ 0.09	0.09$\pm$ 0.02	0.09$\pm$ 0.02	0.06$\pm$ 0.02
WZ	2.89$\pm$ 0.07	0.30$\pm$ 0.02	0.06$\pm$ 0.01	0.06$\pm$ 0.01	0.06$\pm$ 0.01
ZZ	1.91$\pm$ 0.03	0.213$\pm$ 0.009	0.073$\pm$ 0.005	0.078$\pm$ 0.006	0.075$\pm$ 0.006
$Z\mu\mu$	3732$\pm$ 5	343$\pm$ 2	11.3$\pm$ 0.3	7.7$\pm$ 0.2	6.3$\pm$ 0.2
$Z\tau\tau$	9.0$\pm$ 0.3	8.6$\pm$ 0.3	0.27$\pm$ 0.04	0.22$\pm$ 0.04	0.21$\pm$ 0.04
WJets	0.16$\pm$ 0.08	0.11$\pm$ 0.06	0	0	0
Single top	0.82$\pm$ 0.01	0.64$\pm$ 0.01	0.223$\pm$ 0.007	0.21$\pm$ 0.007	0.196$\pm$ 0.007

Table 5.5: Expected number of events in $\mu^+\mu^-$channel for 10 pb^{-1}at 7 TeV. Different jet reconstruction algorithms are considered: caloJets, PFJets and TrackJets

Besides, different scenarios have been tested: the assumption under which the $\not{E}_T$ is not properly modeled at the beginning of the data taking and the analysis should not rely on it but on b-tagging, or the case in which both observables can be used in the analysis. An example is shown in Tables 5.11 5.12, 5.13 for an analysis performed using PFJets and PFMet.

5.4 Results with the data taken in the first run period of the LHC in 2010

Since the beginning of the data taking in March 2010 there have been continuums efforts to understand the results and check that what was observed in data was well represented by the MC predictions, as well as to proof the analysis selections and data-driven techniques developed in simulation was performing as expected. A $t\bar{t}$ candidate is shown in Figure 5.7.
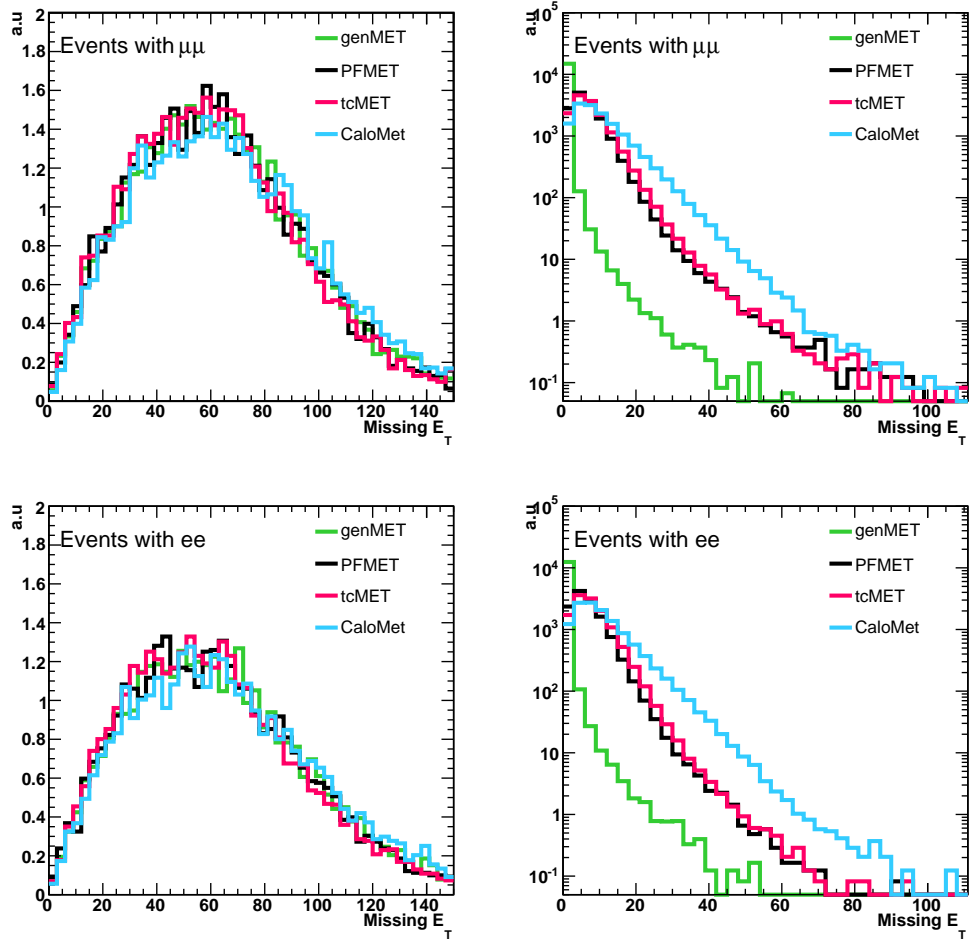
Figure 5.6: - Distributions of the different $\not{E}_T$ algorithms for $t\bar{t}$ events (left) and Drell-Yan events (right)

	diLepton	Z Veto	$\geq$ 2 CaloJets	$\geq$ 2 PFJets	$\geq$ 2 TrackJets
$t\bar{t}$ signal	8.7$\pm$ 0.2	6.7$\pm$ 0.1	4.8$\pm$ 0.1	4.9$\pm$ 0.1	4.5$\pm$ 0.1
$t\bar{t}$ bkg	0.13$\pm$ 0.02	0.10$\pm$ 0.02	0.08$\pm$ 0.02	0.08$\pm$ 0.02	0.08$\pm$ 0.02
WW	2.01$\pm$ 0.09	1.61$\pm$ 0.08	0.09$\pm$ 0.02	0.09$\pm$ 0.02	0.09$\pm$ 0.02
WZ	2.23$\pm$ 0.06	0.22$\pm$ 0.02	0.06$\pm$ 0.01	0.06$\pm$ 0.01	0.046$\pm$ 0.009
ZZ	1.39$\pm$ 0.02	0.147$\pm$ 0.008	0.062$\pm$ 0.005	0.050$\pm$ 0.004	0.042$\pm$ 0.004
Zee	2682$\pm$ 4	225$\pm$ 1	6.9$\pm$ 0.2	5.4$\pm$ 0.2	4.8$\pm$ 0.2
Z$\tau\tau$	6.7$\pm$ 0.2	6.3$\pm$ 0.2	0.24$\pm$ 0.04	0.22$\pm$ 0.04	0.15$\pm$ 0.03
WJets	2.0$\pm$ 0.3	1.8$\pm$ 0.2	0.03$\pm$ 0.03	0.03$\pm$ 0.03	0.03$\pm$ 0.03
Single top	0.64$\pm$ 0.01	0.49$\pm$ 0.01	0.171$\pm$ 0.006	0.169$\pm$ 0.006	0.152$\pm$ 0.006

Table 5.6: Expected number of events in e^+e^- channel for 10 pb^{-1}at 7 TeV. Different jet reconstruction algorithms are considered: caloJets, PFJets and TrackJets

75

	diLepton	$\geq$ 2 Calojets	$\geq$ 2 PFJets	$\geq$ 2 TrackJets
$t\bar{t}$ signal	20.0$\pm$ 0.2	14.6$\pm$ 0.2	14.7$\pm$ 0.2	13.5$\pm$ 0.2
$t\bar{t}$ bkg	0.20$\pm$ 0.02	0.16$\pm$ 0.02	0.16$\pm$ 0.02	0.14$\pm$ 0.02
WW	4.2$\pm$ 0.1	0.25$\pm$ 0.03	0.22$\pm$ 0.03	0.20$\pm$ 0.03
WZ	0.37$\pm$ 0.02	0.06$\pm$ 0.01	0.052$\pm$ 0.009	0.054$\pm$ 0.009
ZZ	0.097$\pm$ 0.006	0.030$\pm$ 0.003	0.043$\pm$ 0.04	0.042$\pm$ 0.004
$Z\mu\mu ee$	1.17$\pm$ 0.09	0.026$\pm$ 0.01	0.046$\pm$ 0.02	0.056$\pm$ 0.02
$Z\tau\tau$	15.6$\pm$ 0.3	0.47$\pm$ 0.06	0.40$\pm$ 0.06	0.33$\pm$ 0.05
WJets	2.45$\pm$ 0.3	0.16$\pm$ 0.08	0.14$\pm$ 0.07	0.11$\pm$ 0.06
Single top	1.44$\pm$ 0.02	0.48$\pm$ 0.01	0.47$\pm$ 0.01	0.435$\pm$ 0.01

Table 5.7: Expected number of events in $e^{\pm}\mu^{\mp}$ channel for 10 pb^{-1} at 7 TeV. Different jet reconstruction algorithms are considered: caloJets, PFJets and TrackJets

	Calo	PF	Track
$t\bar{t}$ signal	4.4$\pm$ 0.1	5.4$\pm$ 0.1	5.0$\pm$ 0.1
$t\bar{t}$ bkg	0.014$\pm$ 0.006	0.017$\pm$ 0.007	0.014$\pm$ 0.006
WW	0.06$\pm$ 0.02	0.07$\pm$ 0.02	0.06$\pm$ 0.02
WZ	0.006$\pm$ 0.003	0.019$\pm$ 0.006	0.021$\pm$ 0.006
ZZ	0.008$\pm$ 0.002	0.008$\pm$ 0.002	0.013$\pm$ 0.002
$Z\mu\mu ee$	0.52$\pm$ 0.06	0.30$\pm$ 0.05	0.76$\pm$ 0.08
$Z\tau\tau$	0.08$\pm$ 0.02	0.11$\pm$ 0.03	0.09$\pm$ 0.03
WJets	0	0	0
Single top	0.142$\pm$ 0.006	0.174$\pm$ 0.006	0.159$\pm$ 0.006

Table 5.8: Expected number of events in $\mu^{+}\mu^{-}$ channel for 10 pb^{-1} at 7 TeV, comparing the three possible approaches: At least 2 Calojets and calo$\not{E}_T > 50$ GeV, at least 2 PFJets and PF$\not{E}_T > 30$ GeV, at least 2 trackJets and tc$\not{E}_T > 30$ GeV

	Calo	PF	Track
$t\bar{t}$ signal	3.2$\pm$ 0.1	3.1$\pm$ 0.1	3.8$\pm$ 0.1
$t\bar{t}$ bkg	0.04$\pm$ 0.01	0.04$\pm$ 0.01	0.06$\pm$ 0.01
WW	0.06$\pm$ 0.01	0.04$\pm$ 0.01	0.08$\pm$ 0.02
WZ	0.018$\pm$ 0.005	0.011$\pm$ 0.004	0.025$\pm$ 0.006
ZZ	0.004$\pm$ 0.001	0.002$\pm$ 0.001	0.011$\pm$ 0.002
$Z\mu\mu ee$	0.26$\pm$ 0.04	0.13$\pm$ 0.009	1.0$\pm$ 0.08
$Z\tau\tau$	0.05$\pm$ 0.02	0.05$\pm$ 0.02	0.06$\pm$ 0.02
WJets	0	0	0
Single top	0.113$\pm$ 0.005	0.105$\pm$ 0.005	0.130$\pm$ 0.005

Table 5.9: Expected number of events in $e^{+}e^{-}$ channel for 10 pb^{-1} at 7 TeV, comparing the three posible approaches: At least 2 Calojets and calo$\not{E}_T > 50$ GeV, at least 2 PFJets and PF$\not{E}_T > 30$ GeV, at least 2 trackJets and tc$\not{E}_T > 30$ GeV

	Calo	PF	Track
$t\bar{t}$ signal	12.5± 0.2	13.6± 0.2	12.4± 0.2
$t\bar{t}$ bkg	0.13± 0.02	0.15± 0.02	0.13± 0.02
WW	0.20± 0.03	0.20± 0.03	0.17± 0.03
WZ	0.043± 0.008	0.041± 0.008	0.049± 0.009
ZZ	0.013± 0.002	0.011± 0.002	0.017± 0.003
$Z\mu\mu ee$	0.006± 0.006	0.008± 0.008	0.006± 0.006
$Z\tau\tau$	0.26± 0.04	0.29± 0.05	0.19± 0.04
WJets	0.16± 0.08	0.11± 0.06	0.11± 0.06
Single top	0.41± 0.01	0.43± 0.01	0.401± 0.009

Table 5.10: Expected number of events in $e^{\pm}\mu^{\mp}$ channel for 10 pb^{-1} at 7 TeV, comparing the three posible approaches: At least 2 Calojets and calo$\not{E}_T > 30$ GeV, at least 2 PFJets and PF$\not{E}_T > 20$ GeV, at least 2 trackJets and tc$\not{E}_T > 20$ GeV

	MET+1btag	NoMET+1btag
$t\bar{t}$ signal	5.1± 0.1	6.0± 0.1
$t\bar{t}$ bkg	0.014± 0.006	0.017± 0.007
WW	0.04± 0.01	0.06± 0.02
WZ	0.006± 0.003	0.014± 0.005
ZZ	0.003± 0.001	0.030± 0.003
$Z\mu\mu ee$	0.12± 0.03	2.2± 0.1
$Z\tau\tau$	0.04± 0.02	0.07± 0.02
WJets	0	0
Single top	0.154± 0.006	0.185± 0.006

Table 5.11: $\mu^{+}\mu^{-}$ channel 10 pb^{-1} at 7 TeV. Calojets + calo$\not{E}_T > 30$ GeV, PFJets+PF$\not{E}_T > 20$ GeV, trackJets+ tc$\not{E}_T > 30$ GeV

	MET+1btag	NoMET+1btag
$t\bar{t}$ signal	3.9± 0.1	4.6± 0.1
$t\bar{t}$ bkg	0.06± 0.01	0.08± 0.02
WW	0.02± 0.01	0.03± 0.01
WZ	0.008± 0.004	0.021± 0.006
ZZ	0.0020± 0.0009	0.016± 0.002
$Z\mu\mu ee$	0.10± 0.02	1.5± 0.1
$Z\tau\tau$	0.03± 0.02	0.05± 0.02
WJets	0	0
Single top	0.126± 0.005	0.149± 0.006

Table 5.12: $e^{+}e^{-}$ channel 10 pb^{-1} at 7 TeV. Calojets + calo$\not{E}_T > 30$ GeV, PFJets+PF$\not{E}_T > 20$ GeV, trackJets+ tc$\not{E}_T > 30$ GeV

	MET+1btag	NoMET+1btag
$t\bar{t}$ signal	12.8± 0.2	13.9± 0.2
$t\bar{t}$ bkg	0.12± 0.02	0.13± 0.02
WW	0.08± 0.02	0.09± 0.02
WZ	0.003± 0.002	0.005± 0.003
ZZ	0.004± 0.001	0.006± 0.002
$Z\mu\mu ee$	0	0.02± 0.01
$Z\tau\tau$	0.13± 0.03	0.15± 0.03
WJets	0	0
Single top	0.377± 0.009	0.41± 0.01

Table 5.13: $e^{\pm}\mu^{\mp}$ channel 10 pb^{-1} at 7 TeV. Calojets + calo $\not{E}_T > 30$ GeV, PFJets+PF$\not{E}_T > 20$ GeV, trackJets+ tc$\not{E}_T > 30$
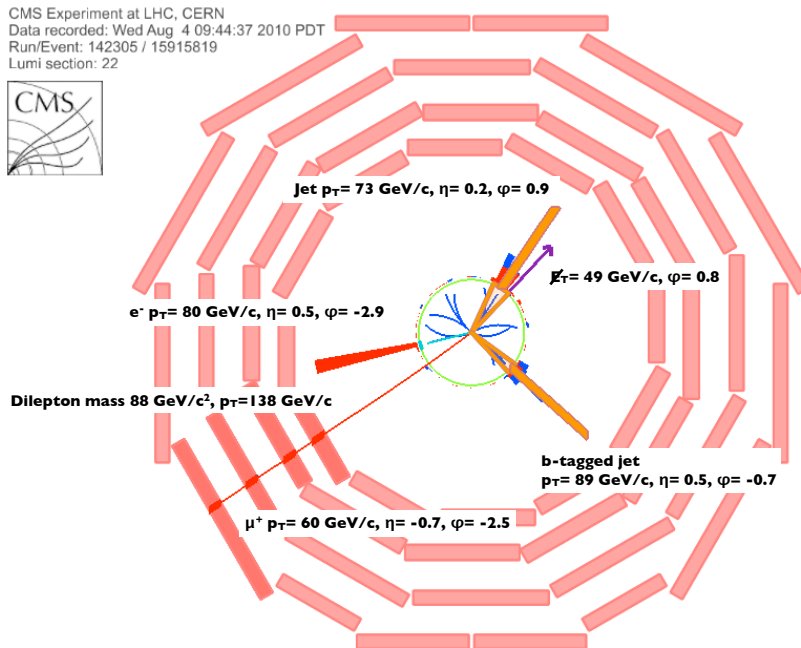
Figure 5.7: - Display of a $t\bar{t}$ $e^{\pm}\mu^{\mp}$ candidate

The first observation of the top quark production constituted a major goal for the experiment, as it demonstrated the excellent performance of the detector. Already with the first 3.1 ± 0.3 pb^{-1} collected between March and August 2010 it was possible to establish the $t\bar{t}$ signal in the dilepton channel.

After applying the event selection criteria (presence of at least one pair of oppositely charged leptons, dilepton invariant mass inconsistent with the Z boson in the e^+e^- and $\mu^+\mu^-$ modes and $\not{E}_T > 30$ GeV (e^+e^-, $\mu^+\mu^-$) and 20 GeV ($e^{\pm}\mu^{\mp}$) as well as at least 2 jets, as will be described in the next section), there are 11 events observed in data (3 e^+e^-, 3 $\mu^+\mu^-$ and 5 $e^{\pm}\mu^{\mp}$).

Backgrounds from Drell-Yan and non-W/Z boson production are estimated from data, while the

remaining backgrounds are taken from simulation, this yields 2. ± 1.0 expected background events.

Figures 5.8 and 5.9 show the distribution of some of the main variables used in the analysis comparing data and simulation, which appear to be in good agreement.
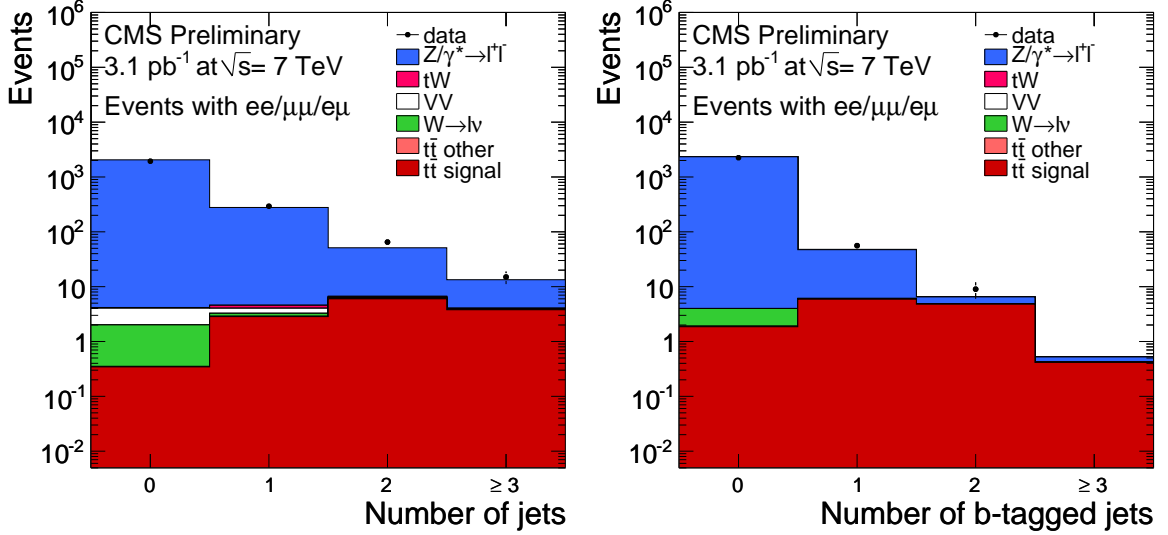
Figure 5.8: - Jet multiplicity (left) and number of b-tagged jets (right) in events passing all dilepton selection criteria for all three dilepton modes combined in 3.1 pb^{-1}, compared to signal and background predictions.

The top-quark pair production cross section is determined from the ratio of the number of observed events in the data after background subtraction with the product of the signal acceptance, selection efficiency, the branching fractions, and the integrated luminosity. From the simulated $t\bar{t}$ sample, the acceptance times efficiency is found to be $(23.0 \pm 1.4)\%$ for events contributing to the $e^+e^-, \mu^+\mu^-$ and $e^\pm\mu^\mp$ modes combined. The total branching fraction for tt to the three modes of the dilepton final state is $(6.45 \pm 0.11)\%$ [8].

Several sources of systematic uncertainty were studied, for the signal an overall uncertainty of 6.4% was obtained (4.4% from the selection efficiency, 3.7% due to jets and $\not{E}_T$ reconstruction and 2.8% from the uncertainties on simulation of the signal). Additionally, the contribution to the systematic uncertainty on the cross section from the uncertainties from the background estimates is 11% .

The measured cross section is 194 ± 72 (stat.) ± 24 (syst.) ± 21 (lumi.) pb [79], which is consistent with the next-to-leading order theoretical predictions, probing the readiness of the CMS detector.

With 3.1 pb^{-1} the uncertainty on the cross section is dominated by the statistical uncertainty, which is 37% , this is expected to be reduced when performing the analysis with the full 2010 dataset.

The property that the two jets expected in dilepton $t\bar{t}$ events both originate from b quarks is exploited to further confirm the top-quark signal. The multiplicity of jets satisfying these b-tagging criteria in events passing full dilepton event selection is shown in Figure 5.10, which provides additional confidence on the top-likeliness of the selected candidates.
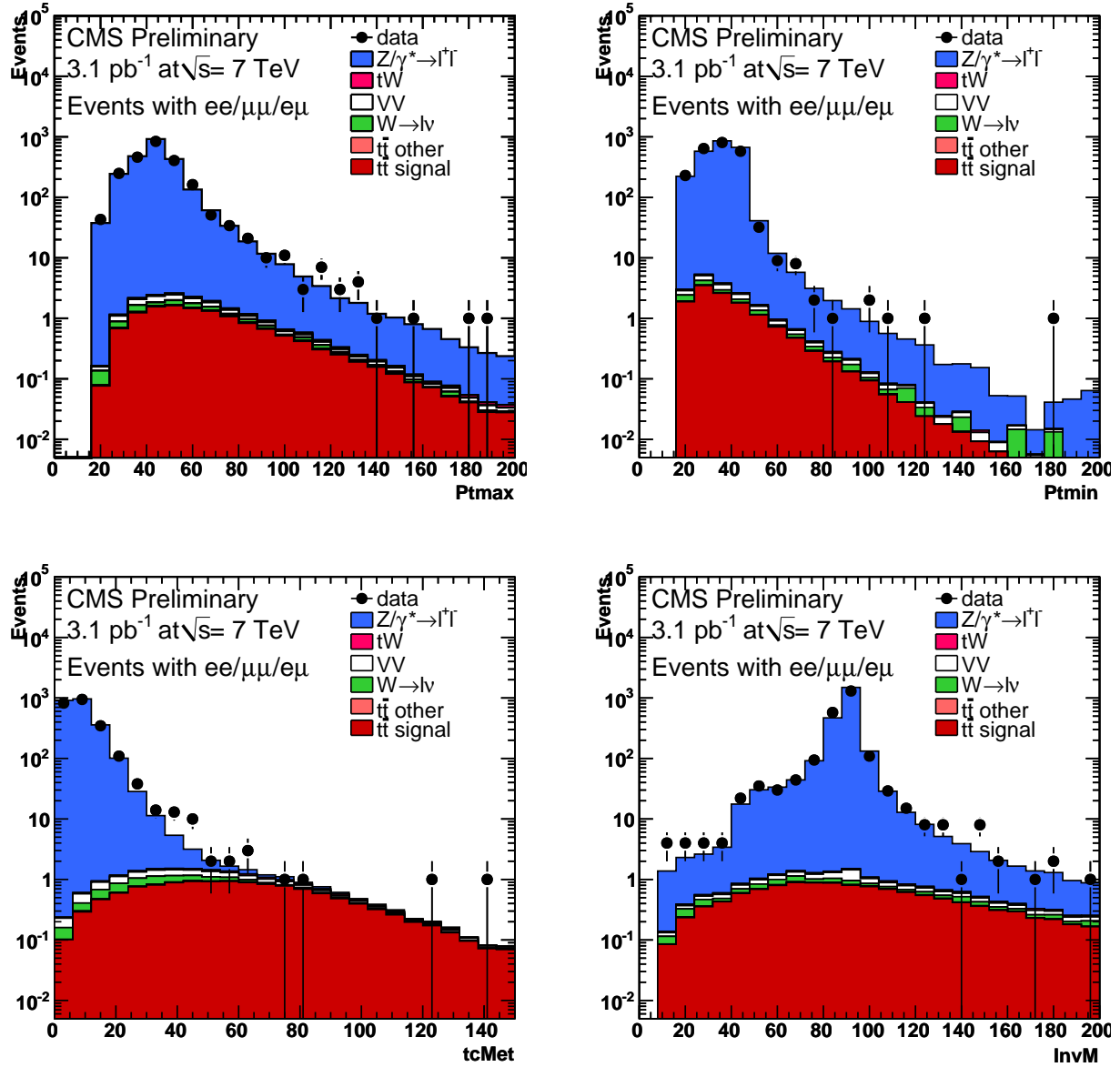
Figure 5.9: - Distribution of the main variables used in the analysis: p_T of the leptons, $\not{E}_T$ and dilepton invariant mass in events passing all dilepton selection criteria for all three dilepton modes combined in 3.1 pb^{-1}, compared to signal and background predictions
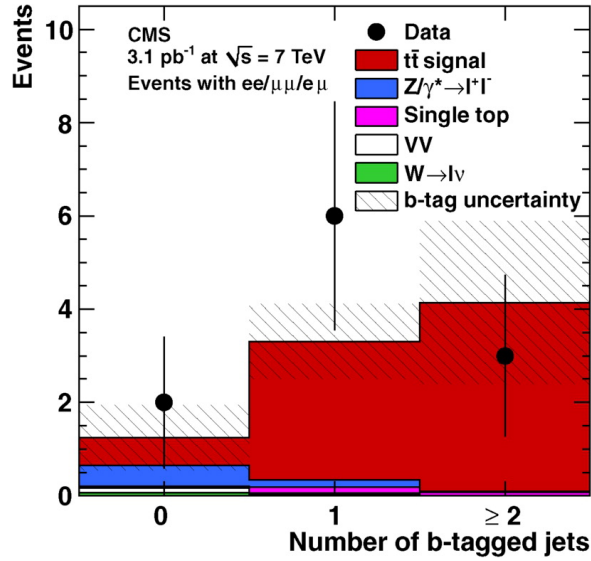
Figure 5.10: - Number of b-tagged jets in events passing all dilepton selection criteria for all three dilepton modes combined in 3.1 pb^{-1}, compared to signal and background predictions. The hatched bands reflect the expected uncertainties on the b-tag efficiency for signal events.

This study has been extended to the full 2010 luminosity to achieve a more precise measurement and b-tagging techniques has been included in the selection to further reduce the backgrounds and obtain a rather pure sample of $t\bar{t}$ events.

5.5 Results with full 2010 luminosity

5.6 Datasets and cross sections

The full list of data and simulation samples used in this analysis is summarized in Tables 5.14, 5.15

Dataset Description	Dataset Name
tt Selection Samples	
Run2010A Muon Nov4 ReReco	/Mu/Run2010A-Nov4ReReco_v1/AOD
Run2010A Electron Nov4 ReReco	/EG/Run2010A-Nov4ReReco_v1/AOD
Run2010B Muon Nov4 ReReco	/Mu/Run2010B-Nov4ReReco-v1/AOD
Run2010B Electron Nov4 ReReco	/Electron/Run2010B-Nov4ReReco-v1/AOD
Fake Rate Measurement Samples	
Run2010A JetMETTau	/JetMETTau/Run2010A-Nov4ReReco_v1/AOD
Run2010A JetMET	/JetMET/Run2010A-Nov4ReReco_v1/AOD
Run2010B Jet	/Jet/Run2010B-Nov4ReReco-v1/AOD
Run2010B Electron PromptReco	/Electron/Run2010B-PromptReco-v2
Run2010A EG Sep17 ReReco	/EG/Run2010A-Sep17ReReco_v2

Table 5.14: Summary of data datasets used

The data samples used in the analysis correspond to data collected by CMS at the center of mass

With Pileup: Processed dataset name is always		
/Fall10-E7TeV_ProbDist_2010Data_BX156_START38_V12-v*/AODSIM		
OR		
Without Pileup: Processed dataset name is always		
/Fall10-START38_V12-v*/AODSIM		
Dataset Description	Primary Dataset Name	cross-section (pb)
$t\bar{t}$	/TTJets_TuneZ2_7TeV-madgraph-tauola/	157.5
WW	/WWTo2L2Nu_TuneZ2_7TeV-pythia6/	4.51
WZ	/WZTo3L3Nu_TuneZ2_7TeV-pythia6-tauola/	0.61
ZZ	/ZZtoAnything_TuneZ2_7TeV-pythia6-tauola/	7.4
Single top- tW channel	/TToBLNu_TuneZ2_tW-channel_7TeV-madgraph/	10.6
W $\rightarrow$ eν	/WToENu_TuneZ2_7TeV-pythia6/	10438
W $\rightarrow$ $\mu\nu$	/WToMuNu_TuneZ2_7TeV-pythia6/	10438
W $\rightarrow$ $\tau\nu$	/WToTauNu_TuneZ2_7TeV-pythia6-tauola/	10438
$Z/\gamma^* \rightarrow e^+e^-$ m_{ll} ϵ [10-20]	/DYToEE_M-10To20_CT10_TuneZ2_7TeV-pythia6/	3457
$Z/\gamma^* \rightarrow \mu^+\mu^-$ m_{ll} ϵ [10-20]	/DYToMuMu_M-10To20_CT10_TuneZ2_7TeV-pythia6/	3457
$Z/\gamma^* \rightarrow \tau^+\tau^-$ m_{ll} ϵ [10-20]	/DYToTauTau_M-10To20_CT10_TuneZ2_7TeV-pythia6-tauola/	3457
$Z/\gamma^* \rightarrow e^+e^-$ m_{ll} ϵ [20-50]	/DYToEE_M-20_TuneZ2_7TeV-pythia6	1666
$Z/\gamma^* \rightarrow \mu^+\mu^-$ m_{ll} ϵ [20-50]	/DYToMuMu_M-20_TuneZ2_7TeV-pythia6/	1666
$Z/\gamma^* \rightarrow \tau^+\tau^-$ m_{ll} ϵ [20-50]	/DYToTauTau_M-20_TuneZ2_7TeV-pythia6-tauola/	1666
$Z/\gamma^* \rightarrow l^+l^-$ m_{ll} ϵ [50-inf]	/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/	3048
$Z/\gamma^* \rightarrow l^+l^-$ m_{ll} ϵ [50-inf]	/DYJetsToLL_TuneD6T_M-50_7TeV-madgraph-tauola/	3048

Table 5.15: Summary of Monte Carlo datasets used

energy of $7\,TeV$. They are provided by the Muon and Electron Primary Dataset based on HLT trigger conditions. These samples are cleaned by requiring that luminosity section were validated by the DQM and PVT groups according to the validation criteria applied for each sub-detector. The good luminosity sections are provided in JSON files applied at the beginning of the analysis chain. The corresponding integrated luminosity is given by the convolution of these JSON files with the analyzed samples (depending of reprocessing). Overlap between datasamples are excluded.

Beam scrapping events are vetoed by selecting a significant fraction of high purity tracks w.r.t. the total number of tracks ($> 25\%$) when the event has at least 10 tracks. Events with anomalous HCAL noise are also rejected. Finally, the presence of at least one well reconstructed primary vertex is required, with $|PV_z| < 24$ cm and with at least four tracks considered in the vertex fit. In addition, the radial position of the primary vertex has to lie wihtin the beam pipe $|PV_\rho| < 2$ cm.

5.7 Selection

This section describes the main physics observables and selections used in the analysis to extract the signal and reduce as much as possible the backgrounds. The first step in the selection chain is the trigger, then the leptons (muons and electrons), jets and $\not{E}_T$ objects have to be reconstructed, a more detailed description of the selection is presented in the following subsections.

5.7.1 Trigger

Different trigger menus run depending on the LHC instantaneous luminosity and are designed to accept most of the signal-like events, keeping the rates at acceptable levels (within $\sim$150 Hz).

As the signal presents 2 high p_{T} leptons, leptonic triggers are used to select the data samples, specifically, single lepton triggers help to efficiently collect dilepton events and are used whenever possible. At the beginning of the data taking, simple and relaxed triggers based on the requirement of the presence of leptons above certain p_{T} threshold were used, but as the luminosity was increasing, these triggers became prescaled, new menus had to be designed and tighter conditions were imposed in the trigger selection, leading mainly to tighter p_{T} thresholds, and in the case of electrons, also tighter identification or isolation criteria, as the electron rates are higher and need to be controlled. If one wants to keep the p_{T} threshold reasonably low with loose cuts, it's necessary to move to double lepton triggers, which combined with the single leptons provide close to 100% efficiencies.

Several High Level inclusive triggers have been considered according to the dilepton channel studied. Tables 5.16, 5.17 show a list of the lepton triggers used and the corresponding run range where the were present.

Muon Trigger path	Run range
HLT_Mu9	< 145000
HLT_Mu11	145000 - 147120
HLT_Mu15_v1	> 147120

Table 5.16: Summary of muon trigger paths used in the analysis

The e^+e^- selection is based on the electron triggers described in Table 5.17 and the $\mu^+\mu^-$ selection is based on the muon triggers described in Table 5.16, while for the $e^\pm\mu^\mp$ study the events are required to have fired either the electron or the muon triggers, this helps to select almost 100 % of the events.

As described before, the trigger menus were evolving as the luminosity conditions were changing so different triggers were used during different periods. In the case of the MC samples, not all the triggers were simulated and the ones present were those corresponding to the trigger menu used for an instantaneous luminosity of $\sim 10^{31}\ cm^{-2}s^{-1}$, since the MC production was launched at around August 2010 with the HLT menu developed at that moment. So in simulation the lepton triggers required are the HLT_Mu9 for muons and HLT_Ele10_SW_L1R with the matching HLT p_{T} object greater than 15 GeV/c for electrons.

5.7.2   Muon selection

Muon candidates are required to be successfully reconstructed in both tracker and muon systems being globalMuon and trackerMuon. Once the muons are reconstructed, the following identification criteria is applied in order to select prompt muons from W decays:

- Identification as GlobalMuonPromptTight

- $p_T > 20\ GeV/c$ and $|\eta| < 2.4$;

- The track associated with the muon candidate is required to have a minimum number of hits in the silicon tracker and to have a high-quality global fit including a minimum number of hits in the muon detector. Number of valid hits in the inner tracker must be greater than 10 and $\chi^2/ndof < 10$ for the global muon fit.

Electron Trigger path	Run range
HLT_Ele10_LW_L1R AND matching HLT object pt>15 GeV/c	< 138000
HLT_Ele15_LW_L1R	138000-141900
HLT_Ele15_SW_L1R	141900-144000
OR of HLT_Ele15_SW_CaloEleId_L1R HLT_Ele20_SW_L1R HLT_DoubleEle10_SW_L1R	144000-144114
OR of HLT_DoubleEle10_SW_L1R HLT_Ele17_SW_CaloEleId_L1R	146000-147120
OR of HLT_DoubleEle15_SW_L1R_v1 HLT_Ele17_SW_TightCaloEleId_SC8HE_L1R_v1 HLT_Ele17_SW_TightEleId_L1R	147120-148100
OR of HLT_DoubleEle17_SW_L1R_v1 HLT_Ele17_SW_TightCaloEleId_Ele8HE_L1R_v2 HLT_Ele17_SW_TightCaloEleId_Ele8HE_L1R_v1 HLT_Ele22_SW_TighterEleId_L1R_v3 HLT_Ele22_SW_TighterEleId_L1R_v2 HLT_Ele17_SW_TighterEleIdIsol_L1R_v3 HLT_Ele17_SW_TighterEleIdIsol_L1R_v2	> 148100

Table 5.17: Summary of electron trigger paths used in the analysis

- The distance of the closest approach to the beam line should be below 0.02 cm in direction transverse to the beam line in order to have a muon candidate track consistent with originating from the beam spot. In addition, the selected leptons have to be close to the good primary vertex: the difference between the muon track at its vertex and the PV along the Z position is less than 1 cm.

Finally, the muons have to be isolated in order to reduce the contamination from fake muons originating from jets using the variable presented in Section 4.1.3, which is required to be $Iso_\mu < 0.15$.

5.7.3 Electron selection

Electron candidates are reconstructed by combining electromagnetic calorimeter towers and track seeds reconstructed from pixel hits and are required to fulfill the following criteria:

- Electron candidates are required to have $p_T > 20\ GeV/c$ and $|\eta| < 2.5$;

- The electron candidate track is required to be consistent with originating from the beam spot: the distance of the closest approach to the beam line should be below 0.04 cm in direction transverse to the beam line; the distance between the point of the closest approach to the beam line and a primary vertex should be less than 1 cm along the z-coordinate direction, as also used for muons.

- The electron identification algorithm is based on cuts applied on discriminating variables between real and fake electrons, as described in Section 4.2.2 and WP90 is the working point chosen.

- Electron candidates consistent with photon conversions are rejected as described in Section 4.2.2, based on either reconstruction of a conversion partner in the silicon tracker, or based on consistent presence of hits in the pixel tracker along the electron candidate trajectory. The number of lost hits in the tracker should be 0, the minimal distance between the electron and its closest opposite sign track should be $|\Delta \cos \theta| > 0.02$ and $Dist > 0.02$ in the $r - \phi$ plane

- Electron candidates within $dR = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.1$ of a tracker-based or globally-fitted muon are rejected to remove fake electron candidates due to muon inner bremsstrahlung (prompt collinear final state radiation).

- Electron is required to be isolated by imposing a requirement on the combined relative isolation presented in Section 4.2.3. This variable is required to be $Iso_e < 0.15$.

5.7.4 Jet selection

The $anti - k_T$ algorithm [67] with R = 0.5 is used for jet clustering. Jets are reconstructed based on the calorimeter, tracker, and muon system information combined in a particle level view of the event using the particle flow reconstruction. Further corrections are applied to the raw jet momenta to establish a relative uniform response of the calorimeter in jet η and an absolute uniform response in jet p_{T} .

Reconstructed jet candidates have to fulfill the following requirements:

- $p_{\mathrm{T}} > 30 GeV/c, |\eta| < 2.5$

- Loose jet identification ("FIRSTDATA"). A jet quality criteria was developed in CMS to select most of the real jets while rejecting most of the fake jets arising from calorimeter or readout electronics noise.

- Jet lepton cleaning: exclusion of jets overlapping with the 2 identified and isolated lepton candidates (electron or muon) used in the analysis if $\Delta R(jet, lepton) < 0.4$.

- Use of TCHE as b-tagging algorithm.

5.7.5 Missing E_T selection

As described in section 4.4, there are several techniques for computing the $\not{E}_T$ in CMS, in this analysis the Particle Flow algorithm is the one chosen.

5.7.6 Event selection

Events having at least 2 selected leptons are classified in e^+e^-, $\mu^+\mu^-$or $e^\pm\mu^\mp$according to the lepton pair with opposite charge which maximizes the sum of the transverse momentum.

In addition, at this step the dilepton invariant mass is required to be above 12 GeV/c^2, this is applied to all modes, to reduce low mass resonances and also because the lower mass range is not well simulated. Also a trigger fiducial selection is applied to $\mu^+\mu^-$and e^+e^-channels, requiring at least one muon to be in $|\eta| < 2.1$ in $\mu^+\mu^-$final state and rejecting those events in which the electron

supercluster $E_T < 17 GeV$ and the muon $|\eta| > 2.1$ in $e^{\pm}\mu^{\mp}$ final state.

In summary, the following sequential cuts are applied to reject the backgrounds while keeping a good signal efficiency:

- Presence of a selected lepton pair;

- Dileptonic invariant mass $\notin [76; 106]\ GeV$ for the e^+e^- and $\mu^+\mu^-$ channels.

- Jet multiplicity ≥ 2;

- Transverse missing energy $\not{E}_T > 30\ GeV$ for e^+e^- and $\mu^+\mu^-$ channels. No cut is applied in the $e^{\pm}\mu^{\mp}$ channel.

- Number of loose b-tagged jets ≥ 1. The b-tagging algorithm considered is the TCHE.

5.8 Lepton efficiency determination

5.8.1 Introduction

Lepton efficiencies can be measured in data using the so called *tag and probe* method [80], where a well known resonance is used. As in this analysis we are interested in high p_T leptons, the Z mass resonance is the one chosen, allowing to provide a pure sample of leptons with which to measure the efficiency of a given selection cut.

One of the leptons, which will be called *tag* , is required to pass tight selection criteria in order to have as low background contamination as possible, allowing to select an enriched Z sample, while the other lepton, which will be called *probe* will be the one used to measure the efficiency, therefore it is only required to pass a certain set of cuts whose efficiency one wants to study. The reconstructed invariant mass between the *probe* and the *tag* candidates is required to be consistent with the Z resonance.

This method performs an unbinned maximum-likelihood fit of the *tag* and *probe* pairs invariant mass peak in the cases in which the *probe* passes the requirement and in which fails, so the efficiency parameter is extracted. An example of the mass distribution for this *tag –probe* pairs can be seen in the Figure 5.11. The function used to model the signal is a Voigtian and an exponential is used to fit the background.

This method is used to measure the lepton reconstruction, identification and isolation efficiencies, as well as the trigger efficiency.

The total lepton selection efficiency can be factorized as:

$$\varepsilon_{TOT} = \varepsilon_{RECO} \cdot \varepsilon_{ID} \cdot \varepsilon_{ISO} \cdot \varepsilon_{trig}$$

where ε_{RECO} is the efficiency for a lepton to be reconstructed, ε_{ID} the identification efficiency relative to an offline reconstructed lepton, ε_{ISO} the isolation efficiency relative to a reconstructed and well identified lepton and finally ε_{trig} the trigger efficiency relative to the offline identified and isolated lepton.

Efficiencies using this method are performed in both data and MC and data to MC scale factors are derived and later applied to correct the MC event yields when computing the cross section.
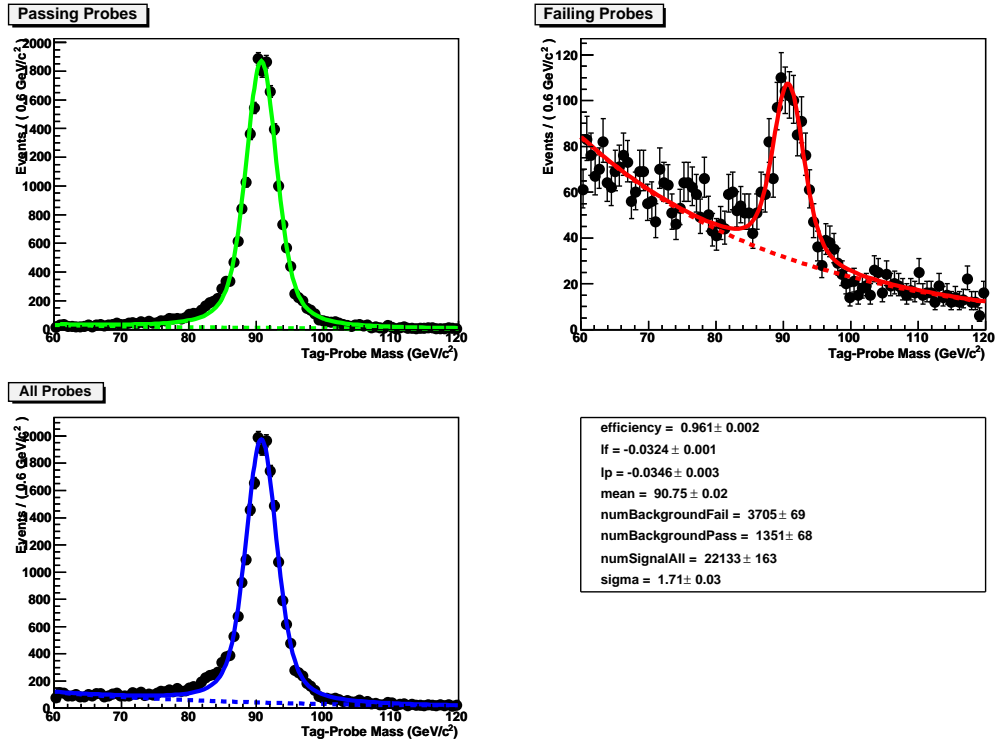
Figure 5.11: - Invariant mass distributions for tag-probe pairs: all pairs (blue), tag-passing probes (green) and tag-failling probes (red)

5.8.2 Muon efficiencies

To estimate the muon efficiencies, the *tag* muon is required to fulfill quality criteria close to the applied for the analysis:

- Muon identification: *isTrackerMuon* and *GlobalMuonPromptTight*

- Number of valid hits for the inner track > 10, and valid of valid hits in the Pixel tracker > 0

- Number of valid hits for the global track > 0

- Number of segments matched in the muon system > 1

- Tracker isolation < 3 GeV within a cone $\Delta R = 0.3$

- Kinematical cuts: $p_T > 20$ GeV/c and $|\eta| < 2.1$

- The tag must fire one of the following trigger paths: HLT_Mu9, HLT_Mu11 or HLT_Mu15_v1

The probes are different in each case, depending on the efficiency definition.

Finally, for the tag and probe pairs, the primary vertex used to estimate the distance cuts is the one that fulfill the following requirements:

- Is valid and is not fake

- Number of degrees of freedom >= 4, $|\rho| < 2$cm and $|z| < 4$ cm

5.8.2.1 Muon reconstruction efficiencies

The efficiencies for the muon reconstruction are calculated using tracker tracks with $p_T > 5\ GeV$ as probes. The passing probes are the ones that also fulfill the GlobalMuon ID and TrackerMuon ID requirements. This efficiency is estimated for different detector regions: $|\eta| < 1.5$ and most forward region $(1.5 < |\eta| < 2.4)$.

In Figure 5.12 the efficiency plots for these two regions are shown comparing Data and MC, and Tables 5.18 and 5.19 summarize these efficiencies.

Figure 5.12: - Muon reconstruction efficiencies for Data and MC, for $|\eta| < 1.5$ (left) and $1.5 < |\eta| < 2.4$ (right) regions

p_T^μ(GeV/c) $\backslash \eta^\mu$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.981\pm^{0.003}_{0.003}$	$0.990\pm^{0.005}_{0.005}$
(40.0, 150.0)	$0.985\pm^{0.002}_{0.002}$	$0.980\pm^{0.004}_{0.004}$

Table 5.18: Muon reconstruction efficiencies for Data Z resonance

p_T^μ(GeV/c) $\backslash \eta^\mu$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.985\pm^{0.001}_{0.001}$	$0.996\pm^{0.001}_{0.001}$
(40.0, 150.0)	$0.985\pm^{0.001}_{0.001}$	$0.994\pm^{0.001}_{0.001}$

Table 5.19: Muon reconstruction efficiencies for MC Z resonance

The average muon reconstruction efficiency in data is found to be 0.985 ± 0.002, while for MC is 0.988 ± 0.001, then the scale factor is 0.997 ± 0.002. The efficiency has some dependency on the kinematics and although the overall agreement data- simulation is found to be good, it shows 1% lower values in the endcap region. To account for this discrepancy due to kinematics, a 20% of the maximum difference in bins is added to the scale factor uncertainty, leading to a muon reconstruction scale factor of $SF_{RECO}^\mu = 0.997\pm 0.005$.

5.8.2.2 Muon identification efficiencies

The efficiencies for the muon identification are calculated using muons that pass the GlobalMuon and TrackerMuon ID criteria as probes. The passing probes are those that also fulfill all the following identification cuts:

- *isTrackerMuon*

- *isGlobalMuon*

- $d_{0BS} < 0.02$ cm

- Number of valid hits for the inner track > 10

- Normalized χ^2 for the global fit < 10

- Number of valid muons hits >0

This efficiency is estimated for different detector regions: $|\eta| < 1.5$ and most forward region $(1.5 < |\eta| < 2.4)$. In Figure 5.13 the efficiency plots for these two regions are shown comparing Data and MC, and Tables 5.20 and 5.21 summarize these efficiency plots.
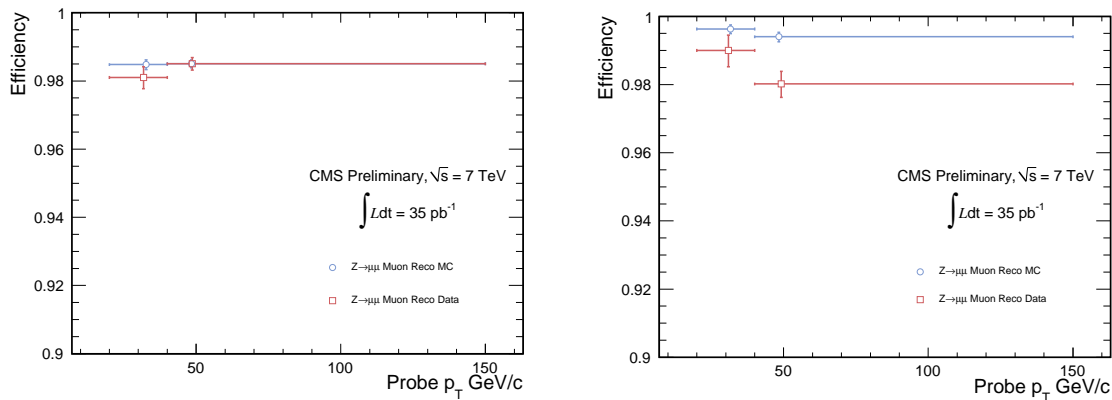

Figure 5.13: - Muon identification efficiencies for Data and MC, for $|\eta| < 1.5$ (left) and $1.5 < |\eta| < 2.4$ (right) regions

p_T^{μ}(GeV/c) $\setminus \eta^{\mu}$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.992 \pm^{0.001}_{0.001}$	$0.978 \pm^{0.003}_{0.003}$
(40.0, 150.0)	$0.989 \pm^{0.001}_{0.001}$	$0.978 \pm^{0.003}_{0.003}$

Table 5.20: Muon identification efficiencies for Data Z resonance

p_T^{μ}(GeV/c) $\setminus \eta^{\mu}$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.994 \pm^{0.001}_{0.001}$	$0.983 \pm^{0.002}_{0.003}$
(40.0, 150.0)	$0.992 \pm^{0.001}_{0.001}$	$0.981 \pm^{0.002}_{0.003}$

Table 5.21: Muon identification efficiencies for MC Z resonance

The average muon identification efficiency in data is found to be 0.987 ± 0.001, while for MC is 0.990 ± 0.001, so the scale factor is 0.997 ± 0.002. The agreement between data and simulation is found to be good and not too dependent on the kinematics, so no additional uncertainty is added.

5.8.2.3 Muon isolation efficiencies

The muon isolation efficiencies are calculated using muons fulfilling all the quality criteria as probes. The passing probes are those that are also isolated with relative isolation < 0.15:

This efficiency is estimated for different detector regions: $|\eta| < 1.5$ and most forward region $(1.5 < |\eta| < 2.4)$. In the Figure 5.14 the efficiency plots for these two regions are shown comparing Data and MC, and Tables 5.22 and 5.23 summarize these efficiency plots.
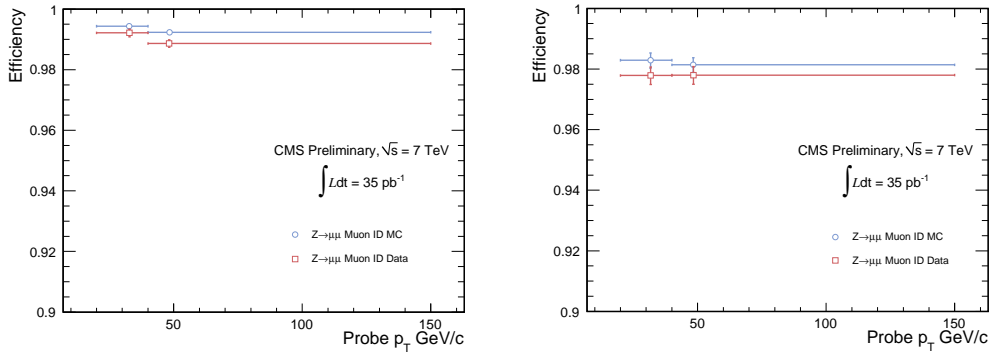

Figure 5.14: - Muon isolation efficiencies for Data and MC, for $|\eta| < 1.5$ (left) and $1.5 < |\eta| < 2.4$ (right) regions

p_T^μ(GeV/c) $\backslash \eta^\mu$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.977 \pm^{0.002}_{0.002}$	$0.981 \pm^{0.003}_{0.003}$
(40.0, 150.0)	$0.994 \pm^{0.001}_{0.001}$	$0.998 \pm^{0.001}_{0.001}$

Table 5.22: Muon isolation efficiencies for Data Z resonance

p_T^μ(GeV/c) $\backslash \eta^\mu$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.981 \pm^{0.002}_{0.002}$	$0.988 \pm^{0.002}_{0.002}$
(40.0, 150.0)	$0.996 \pm^{0.001}_{0.001}$	$0.997 \pm^{0.001}_{0.001}$

Table 5.23: Muon isolation efficiencies for MC Z resonance

The average muon isolation efficiency in data is found to be 0.9872 ± 0.001, while for MC is 0.990 ± 0.001, so the scale factor is 0.997 ± 0.002. The isolation efficiency in data is also found to be close to simulation and similar values are obtained in the different p_T, η bins.

The overall efficiency for the muon reconstruction, identification and isolation on data is 0.961 ± 0.002, while for MC is 0.969 ± 0.001, thus the scale factor (Data/MC) is 0.992 ± 0.003 . Including systematics $SF^\mu_{NoTrig} = 0.992 \pm 0.005$

5.8.2.4 Trigger efficiencies

As described in section 5.7, HLT_Mu9 was the trigger used at the beginning of the data taking, while as the luminosity increased it had to be prescaled and replaced by tighter triggers as the HLT_Mu11 first and the HLT_Mu15_v1 afterwards. The efficiencies of these triggers are estimated in

data using the tag and probe method described before and compared to the MC expectations, as the MC sample used only has the HLT_Mu9 simulated, the data-MC scale factor is computed with respect to the efficiency of this trigger.

For the estimation of the trigger efficiencies the muons that have passed the muon identification plus isolation requirements are taken as probes. They are then matched within a cone of $\Delta R = 0.1$ with the L3 trigger particles. The results are shown in Figure 5.15 and Tables 5.24, 5.25, 5.26 and 5.27 for the same 2 regions as for the muon identification.
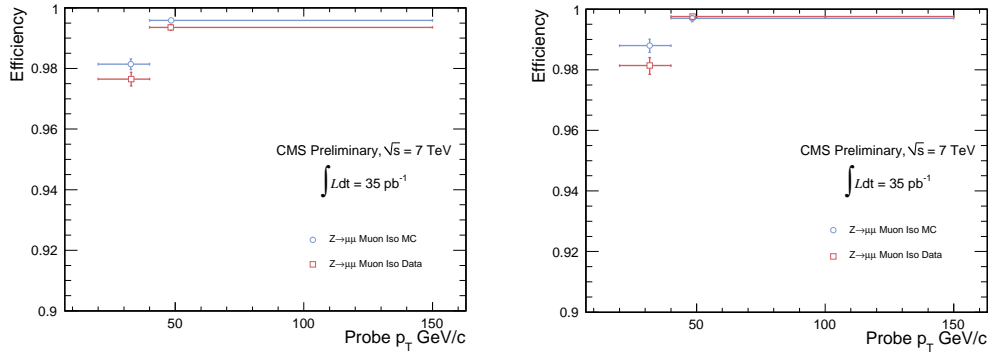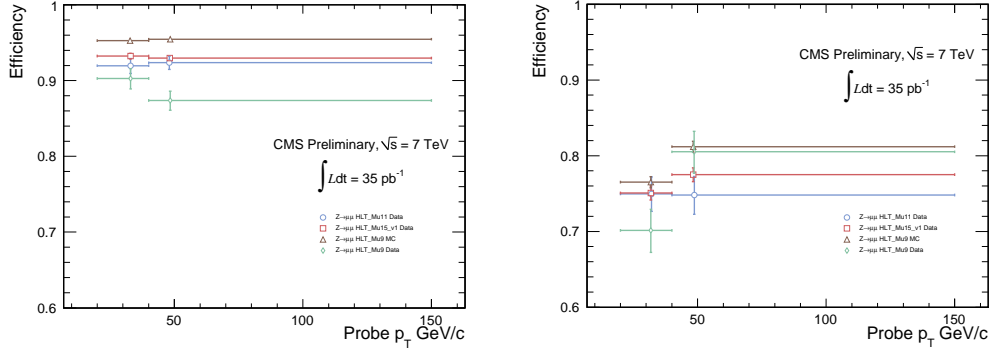

Figure 5.15: - Trigger efficiencies for Data and MC, for $|\eta| < 1.5$ (left) and $1.5 < |\eta| < 2.4$ (right) regions

p_T^μ(GeV/c) $\backslash \eta^\mu$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.903\pm^{0.013}_{0.014}$	$0.701\pm^{0.028}_{0.029}$
(40.0, 150.0)	$0.874\pm^{0.012}_{0.013}$	$0.805\pm^{0.027}_{0.028}$

Table 5.24: Trigger efficiencies for HLT_Mu9 for Data Z resonance

p_T^μ(GeV/c) $\backslash \eta^\mu$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.920\pm^{0.010}_{0.010}$	$0.749\pm^{0.022}_{0.023}$
(40.0, 150.0)	$0.924\pm^{0.008}_{0.009}$	$0.748\pm^{0.025}_{0.025}$

Table 5.25: Trigger efficiencies for HLT_Mu11 for Data Z resonance

p_T^μ(GeV/c) $\backslash \eta^\mu$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.932\pm^{0.004}_{0.004}$	$0.751\pm^{0.009}_{0.009}$
(40.0, 150.0)	$0.930\pm^{0.003}_{0.003}$	$0.775\pm^{0.009}_{0.009}$

Table 5.26: Trigger efficiencies for HLT_Mu15_v1 for Data Z resonance

The overall efficiency on data for HLT_Mu9 is 0.873 ± 0.007, for HLT_Mu11 is 0.848 ± 0.009 and for HLT_Mu15_v1 is 0.884 ± 0.003 while for MC HLT_Mu9 is 0.909 ± 0.002.

A summary of the muon trigger bits used during the 2010 data taking and the corresponding integrated luminosity accumulated in each period is summarized in Table 5.28. To estimate the

p_T^μ(GeV/c) $\backslash\eta^\mu$	(0.0, 1.5)	(1.5, 2.4)
(20.0, 40.0)	$0.953^{+0.002}_{-0.002}$	$0.765^{+0.007}_{-0.007}$
(40.0, 150.0)	$0.955^{+0.002}_{-0.002}$	$0.812^{+0.007}_{-0.007}$

Table 5.27: Trigger efficiencies for HLT_Mu9 for MC Z resonance

trigger efficiency in data, the efficiencies of each trigger path has been weighted by the luminosity, obtaining an average value of 0.878 ± 0.003 per muon.

Trigger path	Efficiency	Luminosity (pb^{-1})
HLT_Mu9	0.873 ± 0.007	3.1
HLT_Mu11	0.848 ± 0.009	5.0
HLT_Mu15	0.884 ± 0.003	28.0
Average	0.878 ± 0.003	36.1

Table 5.28: Trigger efficiencies measured in data for the different muon trigger bits used in the analysis according to the run range where they were unprescaled and the corresponding integrated luminosity.

The efficiency for a dilepton event to pass the trigger is given by:
$\varepsilon_{l_1 l_2} = 1 - (1 - \varepsilon_{l_1})(1 - \varepsilon_{l_2})$.

According to that, the luminosity weighted dimuon trigger efficiency is estimated to be $\varepsilon_{\mu\mu}^{DATA} = 0.985 \pm 0.0007$. The dimuon trigger efficiency in simulation is found to be $\varepsilon_{\mu\mu}^{MC} = 0.992 \pm 0.0004$. So the trigger scale factor per dimuon event is 0.993 ± 0.001.

According to [81], another contribution not covered by the tag and probe method has to be taken into account in the muon trigger efficiency. It refers to the case where the L1 muon trigger pre-fires, this means that the L1 generates the event readout in the bunch crossing earlier than where the actual collision occurred, so the tracker hits of the actual collision are not present in the previous and no muons can be reconstructed. In most of the data taking period the rate of pre-firing is 0.8% in DT and 0.1% in CSC, so the average inefficiency due to prefiring is estimated to be $0.5 \pm 0.3\%$, corresponding to a scale factor of 0.995 ± 0.003 per muon.

The trigger efficiency has a large dependency on the pseudorapidity, muons in the endcap region have lower efficiencies than in the barrel, so a systematic uncertainty is added to account for this effect. About 30% of the times, one of the muons is in $|\eta| > 2.1$, so this is added as a systematic uncertainty. Including also the pre-firing effect, the muon trigger scale factor per dimuon events is $SF_{Trig}^{\mu\mu} = 0.983 \pm 0.007$

5.8.3 Electron efficiencies

Electron reconstruction, identification and isolation selection efficiencies can be also measured using Z events in a similar way as described for muons. This has been studied in [82] [81] [83] and a summary of these results is quoted here for completeness.

The reconstruction efficiency is of the order of 99% [54], while the combination of all quality requirements has an efficiency in the range of 85% to 95% with a small dependence on momentum and

a more significant difference between the ECAL barrel and the ECAL endcaps. The average isolation selection efficiency in Z events in data is in good agreement with the value in simulation: it is approximately 99% for electrons. The electron charge mis-identification is approximately 0.8% , changing from close to 0.5% in the ECAL barrel and increasing to approximately 2% in the ECAL endcaps. Based on the combination of the measured electron efficiencies in data and simulation, the rate of events selected in simulation should be corrected by 0.961 ± 0.009 per electron, where the correction also accounts for differences between data and simulation in the electron isolation selection as well as effects due to charge mis-identification and energy scale.

The trigger efficiency on dielectron events is 99.91%. In data the efficiency in the last data taking period is $99.85 \pm 0.02\%$. The luminosity weighted efficiency from the single electron triggers is $99.88 \pm 0.01\%$. The trigger efficiency scale factor on dielectron events passing identification and isolation requirements is $SF^{ee}_{Trig} = 1.000 \pm 0.001$.

The trigger efficiency on electron-muon events is 99.7% . The luminosity averaged trigger efficiency in data is $99.6 \pm 0.1\%$. Combined with the pre-firing effect, the trigger efficiency scale factor on electron-muon events passing identification and isolation requirements is $SF^{e\mu}_{Trig} = 0.994 \pm 0.003$.

5.8.4 Summary of lepton efficiencies and scale factors

In this section the data to MC scale factors used to correct the differences in lepton efficiencies are summarized.

SF^{μ}_{RECO}	SF^{μ}_{ID}	SF^{μ}_{ISO}	SF^{μ}_{NoTrig}
0.997 ± 0.005	0.997 ± 0.002	0.997 ± 0.002	0.992 ± 0.0054

Table 5.29: Summary of the reconstruction, identification and isolation scale factors per muon

SF^{e}_{RECO}	SF^{e}_{IDISO}	SF^{e}_{qID}	SF^{e}_{eScale}	SF^{e}_{NoTrig}
0.994 ± 0.006	0.9724 ± 0.0061	0.999 ± 0.002	0.995 ± 0.002	0.9608 ± 0.0088

Table 5.30: Summary of the reconstruction, identification, isolation, charge misidentification and electron scale effects scale factors per electron. [82] [81] [83]

$SF^{\mu\mu}_{Trig}$	SF^{ee}_{Trig}	$SF^{e\mu}_{Trig}$
0.983 ± 0.007	1.000 ± 0.001	0.994 ± 0.003

Table 5.31: Summary of the trigger scale factors per dilepton channel

Combining the trigger and non-trigger related contributions, the following lepton efficiency scale factors are derived per dilepton event:

- $\mu^+\mu^-$ final state: $SF^{\mu\mu} = 0.9673 \pm 0.0125$

- e^+e^- final state: $SF^{ee} = 0.9231 \pm 0.0176$

- $e^\pm\mu^\mp$ final state: $SF^{e\mu} = 0.9474 \pm 0.0105$

5.8.5 Lepton efficiencies in Z and top

As described before, lepton reconstruction, identification, isolation and trigger efficiencies have been measured in data using Z events. The topology of this kind of events may be a bit different from $t\bar{t}$ events, which for example contain more jets. This can affect specially the isolation efficiency, so to check this a comparison between the lepton efficiencies in simulated Z and $t\bar{t}$ samples have been performed for both inclusive and 2 jets selections.

The results are summarized in Table 5.32 and the behaviour of the efficiencies as a function of p_{T} and η of the leptons is shown in Figures 5.16 and 5.17.
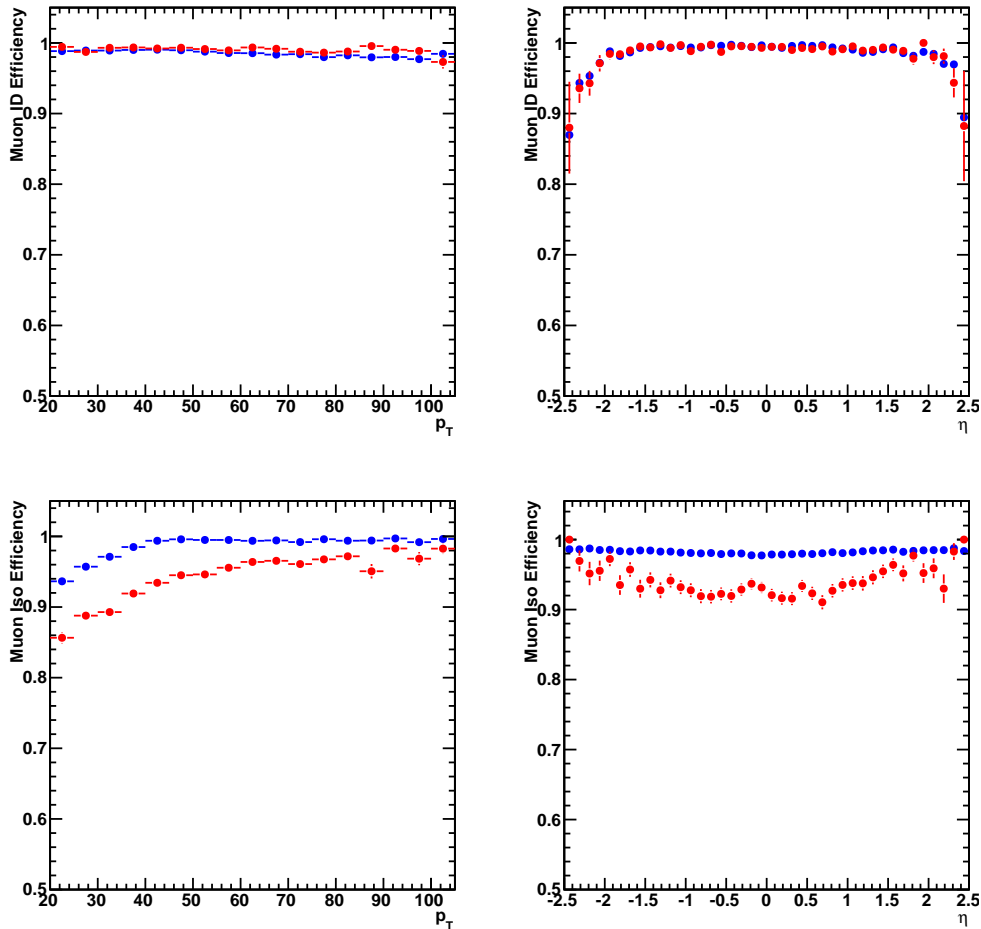
Figure 5.16: - Muon identification and isolation efficiencies estimated using $Z + jets$ and $t\bar{t}$ samples

As expected, no significant difference is observed in the identification efficiencies while from 4% to 5% difference in efficiency is found between the two samples, which is taken as a systematic uncertainty due to the lepton selection modeling.
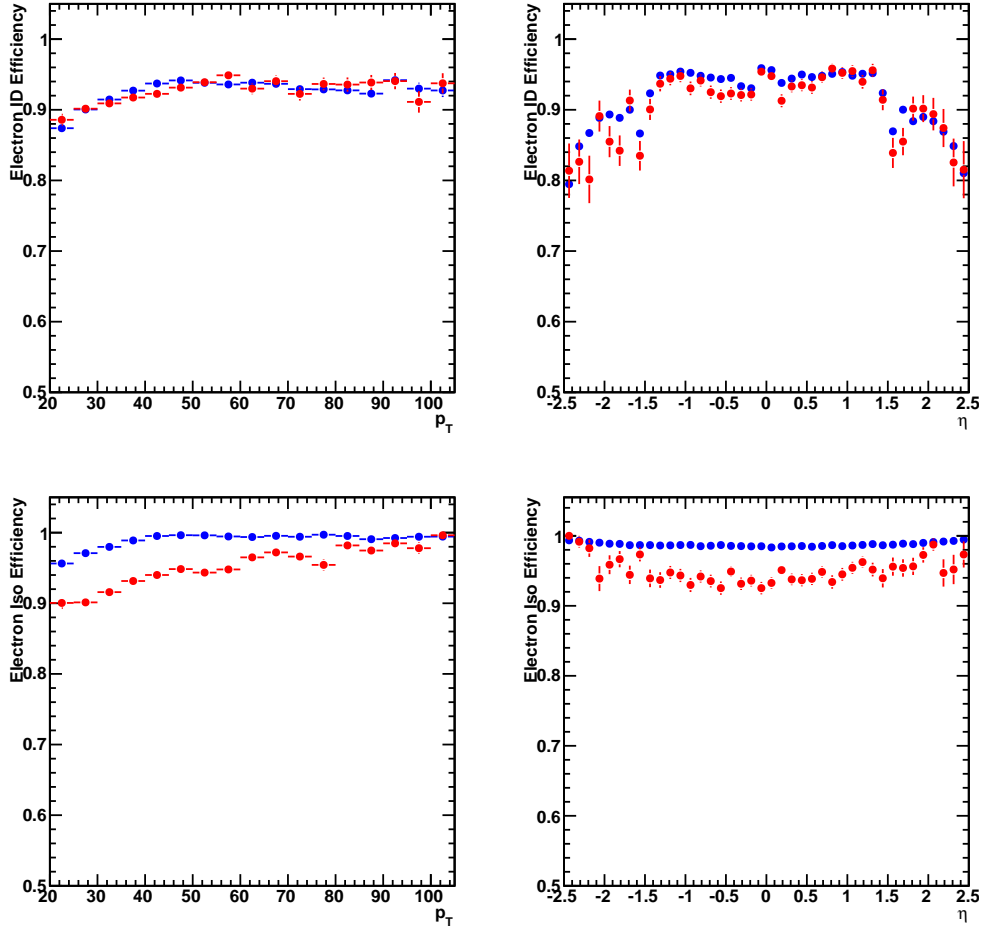
Figure 5.17: - Electron identification and isolation efficiencies estimated using $Z + jets$ and $t\bar{t}$ sample

	MuonID	MuonISO	ElectronID	ElectronISO
Inclusive				
Z+Jets	98.91 ± 0.01	98.19 ± 0.02	92.42 ± 0.03	98.70 ± 0.01
$t\bar{t}$	99.07 ± 0.07	93.26 ± 0.2	92.30 ± 0.2	94.34 ± 0.2
$\geq 2jets$				
Z+Jets	98.99 ± 0.02	97.35 ± 0.03	91.84 ± 0.05	98.03 ± 0.03
$t\bar{t}$	99.06 ± 0.07	93.05 ± 0.2	92.23 ± 0.2	94.21 ± 0.2

Table 5.32: Comparison of the muon and electron efficiencies in Z+jets and $t\bar{t}$ events

5.9   Observed data events and MC expectation with 35.9 pb^{-1}

In this section it is summarized the number of signal and background events expected from simulation, as well as the observed candidates in data that are passing each of the selection steps described in previous sections for $\mu^+\mu^-$ (Table 5.33), ee (Table 5.35) and $e^\pm\mu^\mp$ (Table 5.35) channels. No corrections are yet applied to the MC.

	diLepton	Z Veto	$\geq$ 2 jets	MET	$\geq$1 btag	$\geq$2 btag
$t\bar{t}$ sig	41.5 ±0.4	32.0 ±0.4	23.6 ±0.3	19.8 ±0.3	18.5 ±0.3	10.5 ±0.2
$t\bar{t}$ bkg	0.16 ±0.03	0.13 ±0.02	0.10 ±0.02	0.08 ±0.02	0.06 ±0.02	0.022 ±0.010
$DY\mu\mu,ee$	15078 ±25	1367 ±7	41 ±1	3.5 ±0.4	1.0 ±0.2	0.19 ±0.09
$DY\tau\tau$	41 ±1	38 ±1	1.3 ±0.2	0.5 ±0.2	0.2 ±0.1	0.04 ±0.04
WJets	0.7 ±0.2	0.6 ±0.2	0	0	0	0
VV	22.0 ±0.1	8.8 ±0.1	0.69 ±0.02	0.32 ±0.02	0.086 ±0.009	0.012 ±0.003
Single top	3.01 ±0.05	2.35 ±0.04	0.80 ±0.02	0.66 ±0.02	0.56 ±0.02	0.19 ±0.01
Total MC	15187 ±25	1449 ±7	68 ±1	24.9 ±0.5	20.5 ±0.4	11.0 ±0.2
Data	15056	1463	83	28	24	15

Table 5.33: Expected and observed number of events with 35.9 pb^{-1} in the $\mu\mu$ channel

	diLepton	Z Veto	$\geq$ 2 jets	MET	$\geq$1 btag	$\geq$2 btag
$t\bar{t}$ sig	35.1 ±0.4	27.0 ±0.3	19.8 ±0.3	16.5 ±0.3	15.1 ±0.3	8.6 ±0.2
$t\bar{t}$ bkg	0.88 ±0.06	0.63 ±0.05	0.48 ±0.05	0.38 ±0.04	0.35 ±0.04	0.18 ±0.03
$DY\mu\mu,ee$	12621 ±23	1106 ±7	34 ±1	1.5 ±0.3	0.4 ±0.1	0.16 ±0.08
$DY\tau\tau$	37 ±1	35 ±1	1.2 ±0.2	0.7 ±0.2	0.17 ±0.09	0.04 ±0.04
WJets	12.3 ±1.0	9.8 ±0.8	0.4 ±0.2	0.2 ±0.1	0	0
VV	18.4 ±0.1	7.36 ±0.09	0.61 ±0.02	0.29 ±0.02	0.085 ±0.010	0.011 ±0.003
Single top	2.64 ±0.05	2.06 ±0.04	0.68 ±0.02	0.56 ±0.02	0.48 ±0.02	0.16 ±0.01
Total MC	12727 ±23	1188 ±7	57 ±1	20.2 ±0.4	16.7 ±0.3	9.2 ±0.2
Data	11974	1272	65	23	15	7

Table 5.34: Expected and observed number of events with 35.9 pb^{-1} in the ee channel

	diLepton	$\geq$ 2 jets	$\geq$1 btag	$\geq$2 btag
$t\bar{t}$ sig	77.8 $\pm$0.6	57.9 $\pm$0.5	53.5 $\pm$0.5	30.9 $\pm$0.4
$t\bar{t}$ bkg	1.28 $\pm$0.07	1.00 $\pm$0.07	0.91 $\pm$0.06	0.42 $\pm$0.04
$DY\mu\mu, ee$	3.8 $\pm$0.4	0.4 $\pm$0.1	0.21 $\pm$0.09	0
$DY\tau\tau$	78 $\pm$2	2.5 $\pm$0.3	0.7 $\pm$0.2	0.09 $\pm$0.06
WJets	16 $\pm$1	0.5 $\pm$0.2	0.07 $\pm$0.07	0
VV	18.8 $\pm$0.2	0.92 $\pm$0.03	0.23 $\pm$0.02	0.024 $\pm$0.005
Single top	5.67 $\pm$0.07	1.86 $\pm$0.04	1.56 $\pm$0.03	0.53 $\pm$0.02
Total MC	202 $\pm$2	65.1 $\pm$0.7	57.2 $\pm$0.6	32.0 $\pm$0.4
Data	176	60	51	30

Table 5.35: Expected and observed number of events with 35.9 pb^{-1} in the $e\mu$ channel

The distribution of the main variables used in the analysis, comparing data and MC, is illustrated in Figures 5.18, 5.19, 5.20 and 5.21.

Figure 5.22 shows the number of jets present in events with at least two leptons passing the identification and isolation criteria.

Figure 5.18: - Distributions of the p_{T} of the leading lepton for events passing the dilepton selection criteria, compared to expectations from simulation.
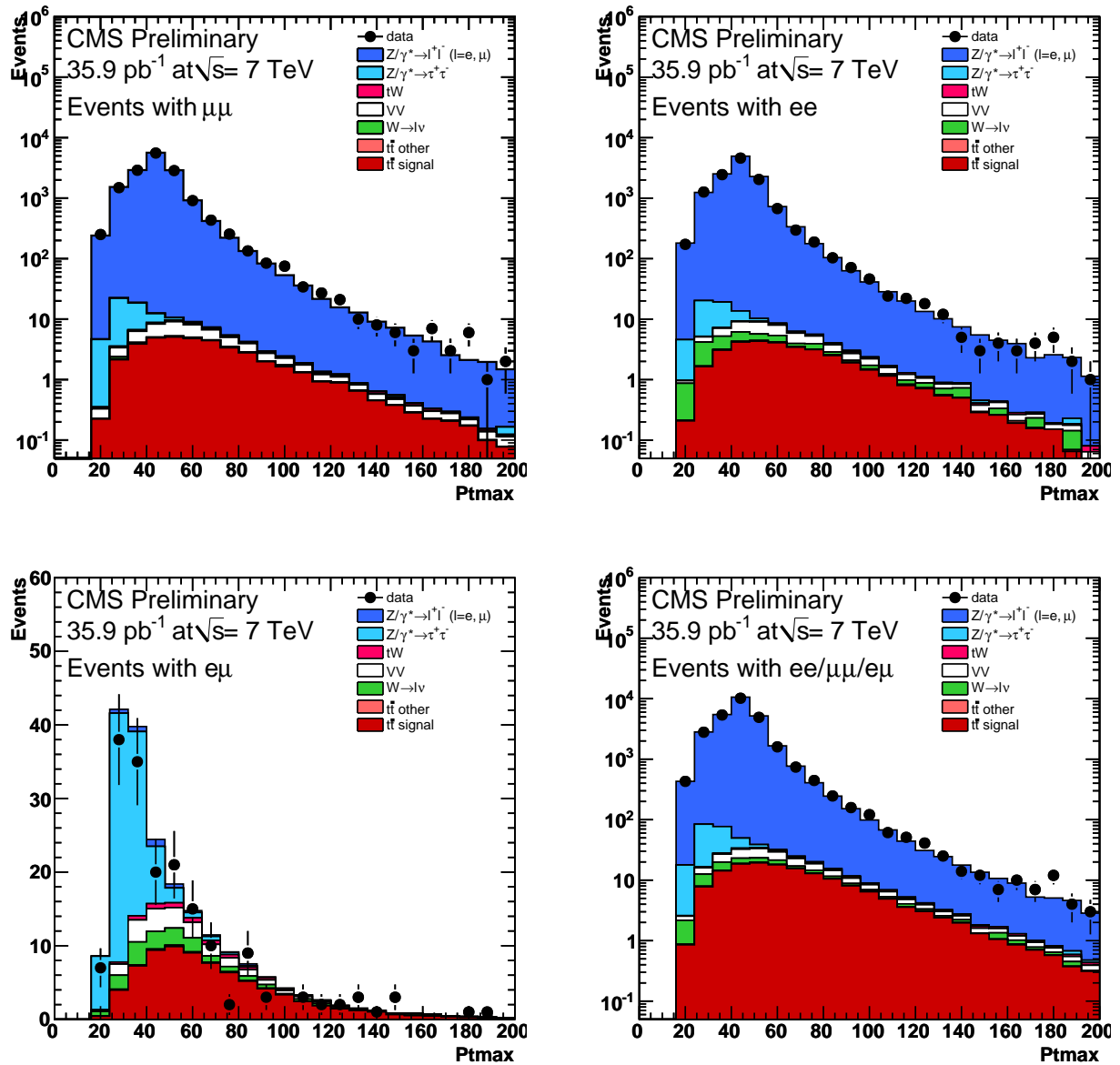
Figure 5.19: - Distributions of the p_T of the second lepton for events passing the dilepton selection criteria, compared to expectations from simulation.
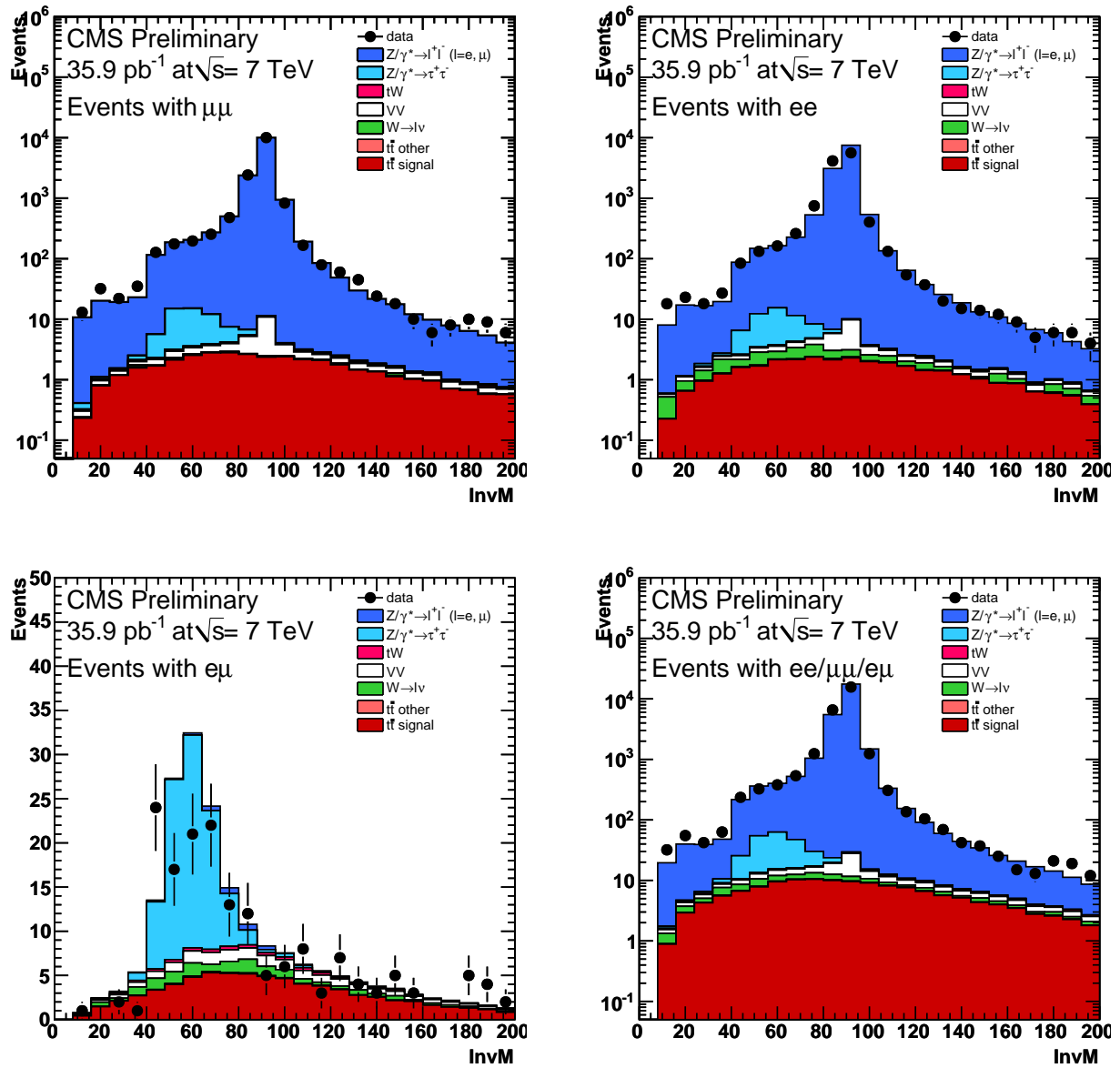
Figure 5.20: - Distributions of the dilepton invariant mass of the leading lepton for events passing the dilepton selection criteria, compared to expectations from simulation.
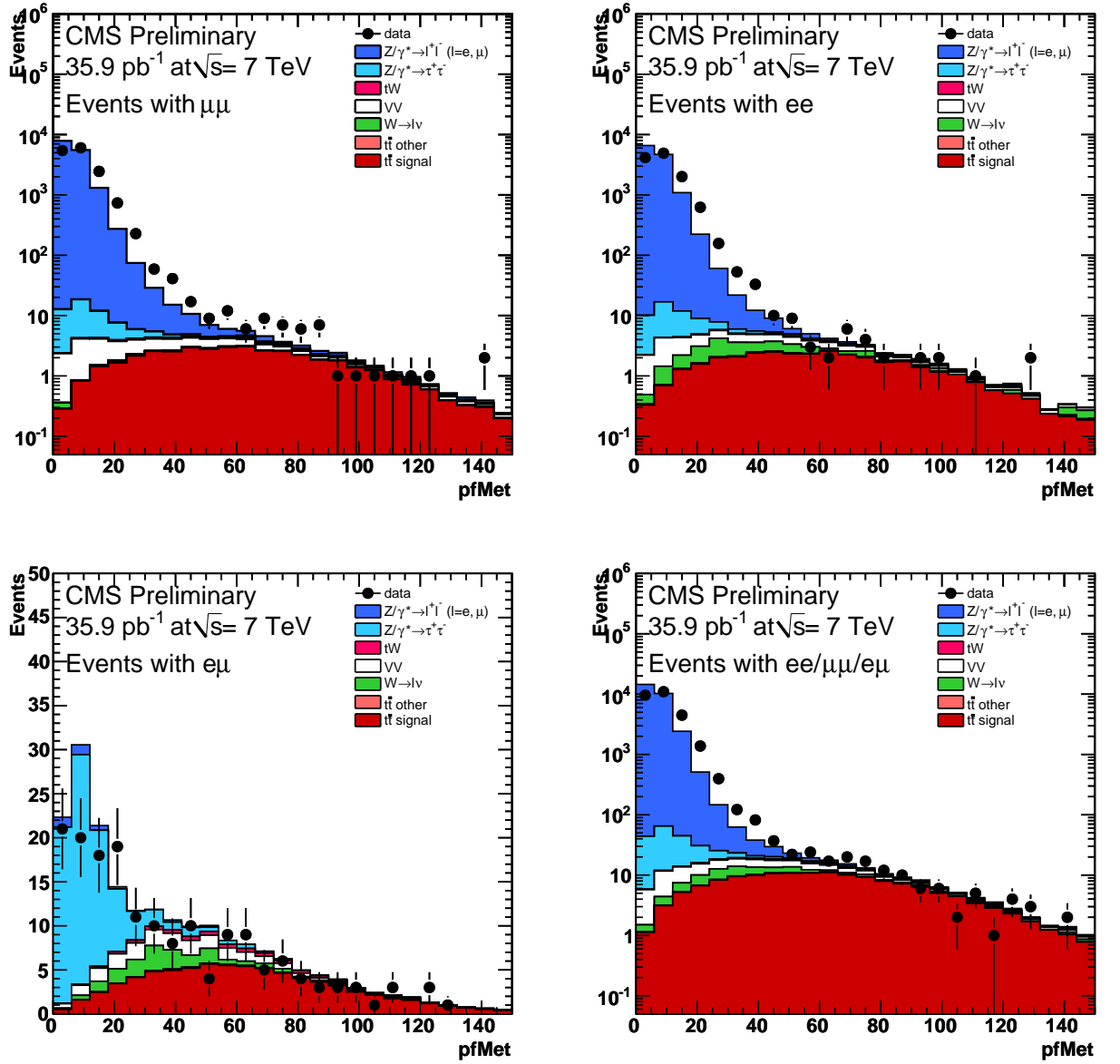
Figure 5.21: - Distributions of the particle flow E_T of the leading lepton for events passing the dilepton selection criteria, compared to expectations from simulation.

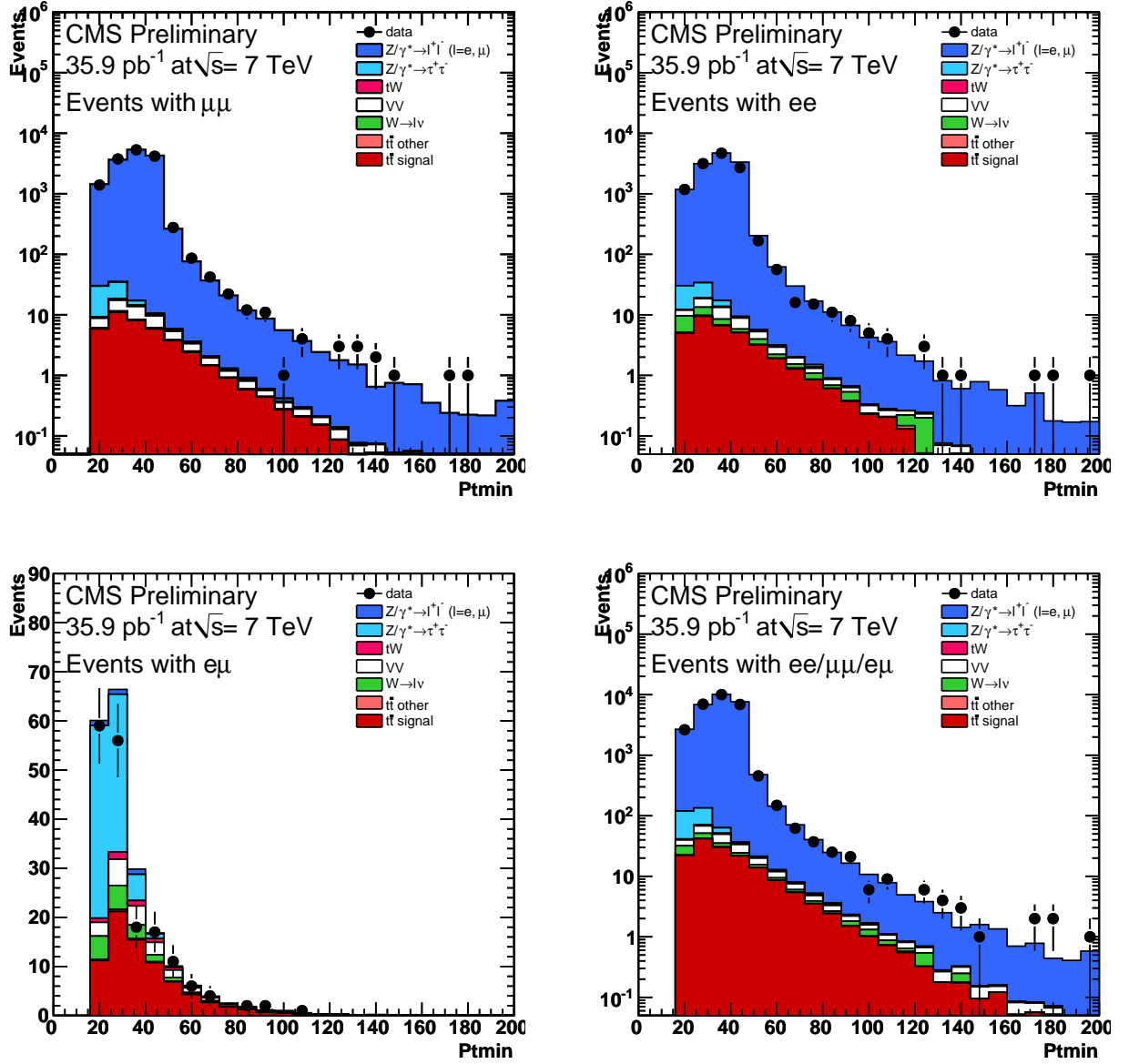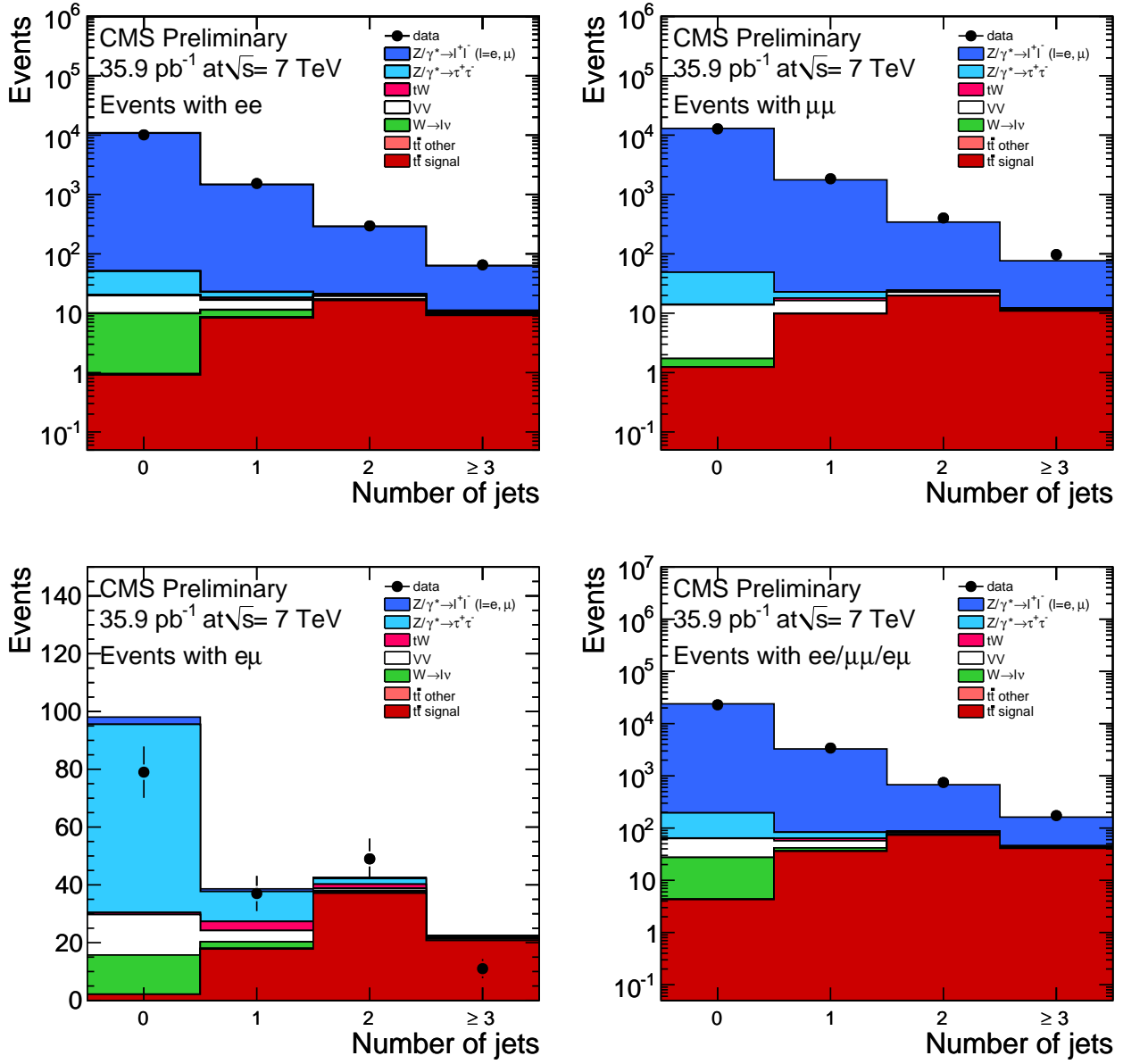**Figure 5.22:** - Distributions of the jet multiplicity for events passing the dilepton selection criteria, compared to expectations from simulation.
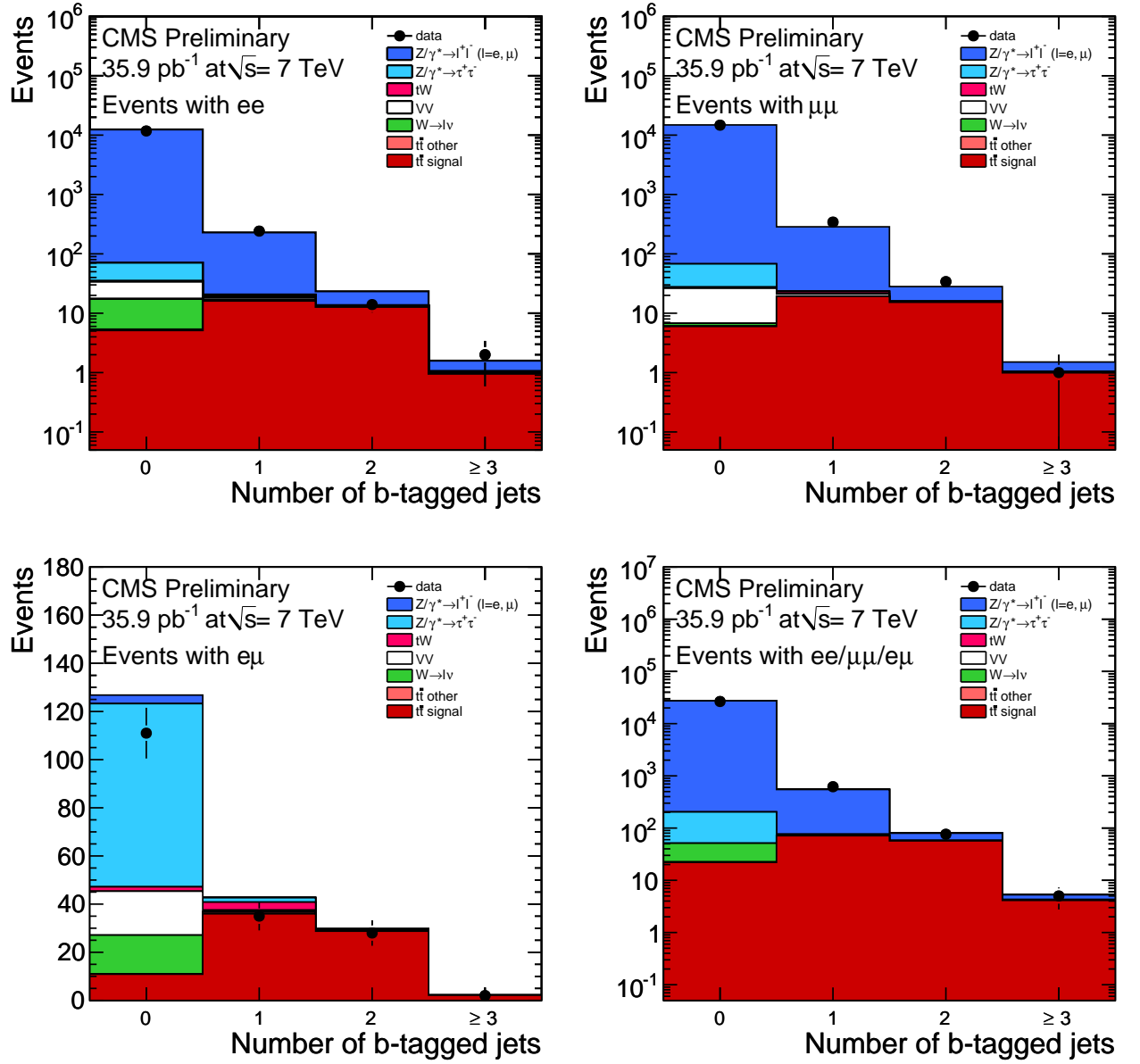
Figure 5.23: - Distributions of the b-tagged jet multiplicity for events passing the dilepton selection criteria, compared to expectations from simulation.

5.10 Background estimation from data

In this section it is described the methods followed in this analysis to estimate the background contribution from data, which will help to rely as less as possible on simulation, where they may not be fully well modeled, the processes considered are Drell Yan events decaying into 2 leptons (e^+e^- or $\mu^+\mu^-$) and the background from fake leptons (W+jets, QCD and non-dileptonic $t\bar{t}$). The other processes: dibosons, single top and $Z/\gamma^* \to \tau^+\tau^-$ are directly taken from simulation, their contribution is expected to be small and are expected to be reasonably well modeled in simulation.

5.10.1 Data driven Drell-Yan estimation

The $Z/\gamma^* \to l^+l^-$ (with l=μ ,e) process does not have genuine $\not{E}_T$ so a significant amount of $\not{E}_T$ is required in the analysis selection in order to reject this background, however there are still events surviving this cut, mainly coming from events with mismeasured leptons.

In addition to the $\not{E}_T$ selection, events near the Z mass peak are also rejected so they can be used used as a control region to estimate this background from data [84].

In the control region, a scale factor between the events in data and simulation can be computed, and the estimated number of events outside the control region can be expressed as:

$$N_{DY}^{out} = \frac{N_{DY\,data}^{in}}{N_{DY\,MC}^{in}} \cdot N_{DY\,MC}^{out} \tag{5.1}$$

The term coming from simulation can be represented as $R_{out/in} = \frac{N_{DY\,MC}^{out}}{N_{DY\,MC}^{in}}$ and:

$$N_{DY}^{out} = R_{out/in} \cdot N_{DY\,data}^{in} \tag{5.2}$$

The control region is dominated by Z/γ^* events but there is still some contamination from other backgrounds that has to be subtracted, specially in events with large $\not{E}_T$ and jets in which Z/γ^* may not be so dominant to neglect other sources of background, as can be observed for instance in Figures 5.24,5.25 and 5.26.

The background contribution in the Z control region can be classified as:

- **Peaking backgrounds**: WZ and ZZ give a peak in the reconstructed dilepton invariant mass if the 2 selected leptons are coming from Z in the case of WZ or the same Z in the case of ZZ. The dominant ZZ contribution is coming from $l^+l^-\nu_l\bar{\nu}_l$ and this contribution can be estimated together with the Z/γ^* .

- **Non peaking backgrounds**: WW, $t\bar{t}$, tW and W+jets which give a continuum distribution in the invariant mass. This contribution can be estimated from data by measuring the number of events in the control region in the $e^{\pm}\mu^{\mp}$ final state, scaled according to the event yields in the ee and $\mu^+\mu^-$ channels, taking into account the differences in reconstruction between muons and electrons.

As contamination from non-DY backgrounds can still be present in the control region, this contribution is subtracted from the $e^{\pm}\mu^{\mp}$ channel and then scaled according to the event yields in e^+e^- and $\mu^+\mu^-$ channels.

Then the number of events outside the Z-veto can be measured from data as:

$$N_{out}^{l^+l^-,obs} = R_{out/in}^{l^+l^-}(N_{in}^{l^+l^-} - 0.5N_{in}^{e\mu}k_{ll}) \tag{5.3}$$

where $ll = \mu\mu$ or ee.

k is a corrector factor that must be applied to take into account the differences between electron and muon reconstruction, it is calculated using the events in the Z peak region passing the standard dilepton and jet selection but without any $\not{E}_T$ requirement, and can be expressed as:

$$k_{ee} = \sqrt{\frac{N^{e^+e^-}_{in,loose}}{N^{\mu^+\mu^-}_{in,loose}}} \ k_{\mu\mu} = \sqrt{\frac{N^{\mu^+\mu^-}_{in,loose}}{N^{e^+e^-}_{in,loose}}}$$
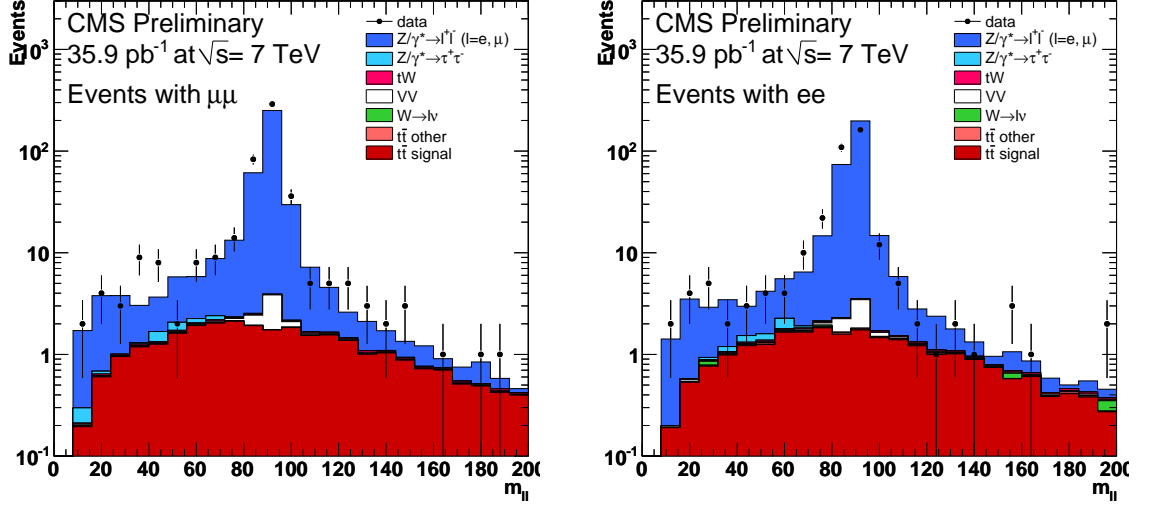
Figure 5.24: - Dilepton invariant mass distribution for events with two leptons and at least two jets
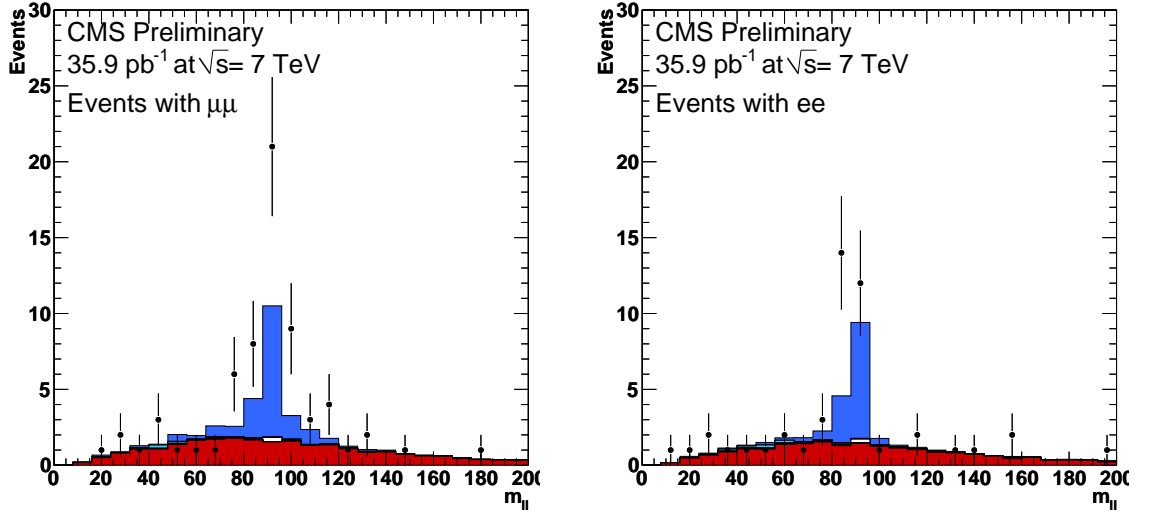
Figure 5.25: - Dilepton invariant mass for events with two leptons, at least two jets and $\not{E}_T > 30$ GeV

The presence of pile-up events can affect the selection, specially the jet and $\not{E}_T$ distributions. There are two sets of simulated samples available, one in which no pile-up events are present and another one which contains extra pp collisions, although they contain more than what it is actually in data. As the use of these samples can be somewhat pessimistic, it was decided to use both set of samples and take the final result as the average between them. Results are summarized in table 5.36.
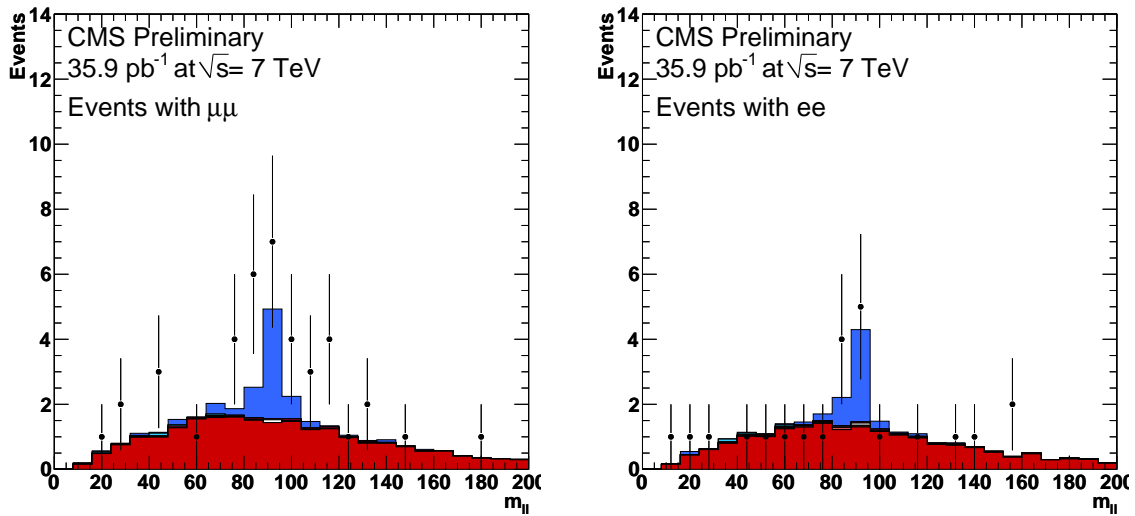
Figure 5.26: - Dilepton invariant mass for events with two leptons, at least two jets, $\not{E}_T > 30$ GeV and at least one b-tagged jet

Figure 5.27: - $R_{out/in}$ as a function of the cut applied on MET for events with two good leptons and at least two jets (left plot) and as a function of the number of jets for events with two good leptons (right plot) (No PU samples)

As can be observed in table 5.36 $R_{out/in}$ varies with the selection, specially after the $\not{E}_T$ requirement and is more noticeable in the $\mu^+\mu^-$ channel. The stability of $R_{out/in}$ with the number of jets and the $\not{E}_T$ is represented in Figures 5.27 and 5.28 for non pile-up and pile-up samples. The value is found to be stable with the jet multiplicity but fluctuates towards higher values in events with larger MET, specially for muons. This can be observed also in Figure 5.29. Events can migrate outside the Z mass window due to the mismeasurements of leptons from Z decays, which cause the invariant mass of the lepton pair to be outside the Z mass window, as well as high $\not{E}_T$. Muons are more affected by this, leading to a invariant mass distribution which is less sharp compared to electrons.

	$\geq$ **2 jets**	$\not{E}_T > 30$	$\geq$ **1 b-tag**
e^+e^- no pile up			
DY MC	33.5 ± 1.2	1.6 ± 0.2	0.45 ± 0.1
DY Data driven estimate	33.6 ± 2.4	3.0 ± 0.96	0.67 ± 0.6
$R_{out/in}$	0.116 ± 0.004	0.13 ± 0.02	0.11 ± 0.04
e^+e^- pile up			
DY MC	38.2 ± 1.2	1.8 ± 0.3	0.51 ± 0.1
DY Data driven estimate	34.7 ± 2.4	2.7 ± 0.8	0.64 ± 0.6
$R_{out/in}$	0.120 ± 0.004	0.12 ± 0.02	0.10 ± 0.03
e^+e^- combined			
DY MC	35.8 ± 1.2	1.7 ± 0.2	0.48 ± 0.1
DY Data driven estimate	34.2 ± 2.4	2.8 ± 0.9	0.66 ± 0.6
$R_{out/in}$	0.118 ± 0.004	0.12 ± 0.02	0.10 ± 0.04
$\mu^+\mu^-$ no pile up			
DY MC	41.2 ± 1.3	3.5 ± 0.4	1.05 ± 0.2
DY Data driven estimate	49.2 ± 3.0	9.2 ± 2.2	2.9 ± 1.4
$R_{out/in}$	0.121 ± 0.004	0.26 ± 0.03	0.20 ± 0.04
$\mu^+\mu^-$ pile up			
DY MC	46.0 ± 1.4	3.8 ± 0.4	1.1 ± 0.2
DY Data driven estimate	49.3 ± 2.9	7.1 ± 1.7	2.4 ± 1.1
$R_{out/in}$	0.121 ± 0.004	0.20 ± 0.02	0.16 ± 0.03
$\mu^+\mu^-$ combined			
DY MC	43.6 ± 1.4	3.6 ± 0.4	1.1 ± 0.2
DY Data driven estimate	49.2 ± 3	8.1 ± 2	2.6 ± 1.2
$R_{out/in}$	0.121 ± 0.004	0.23 ± 0.02	0.18 ± 0.04

Table 5.36: Drell Yan background data driven estimation in the $\mu^+\mu^-$ and e^+e^- channels compared to simulation for $35.9pb^{-1}$ at several steps of the analysis: after the 2 jet, MET or 1 btag requirements. Errors are statistical only
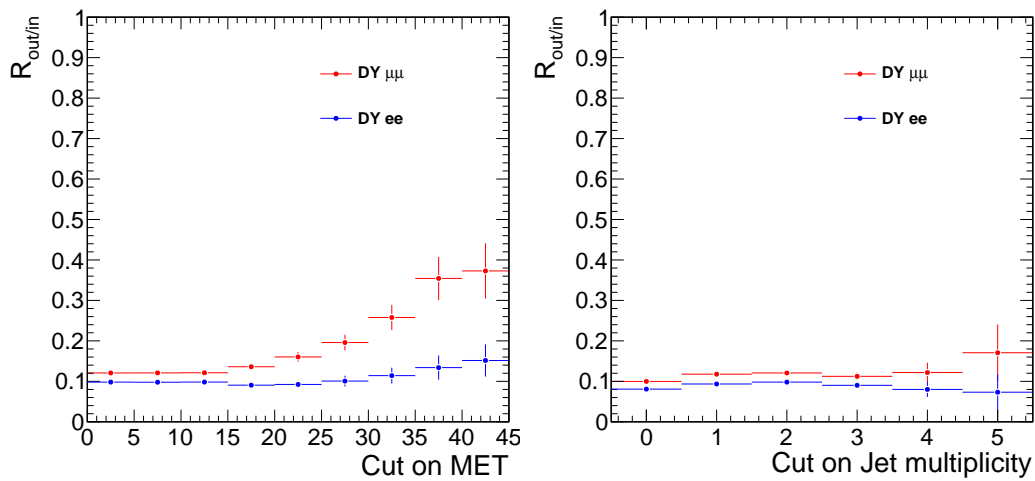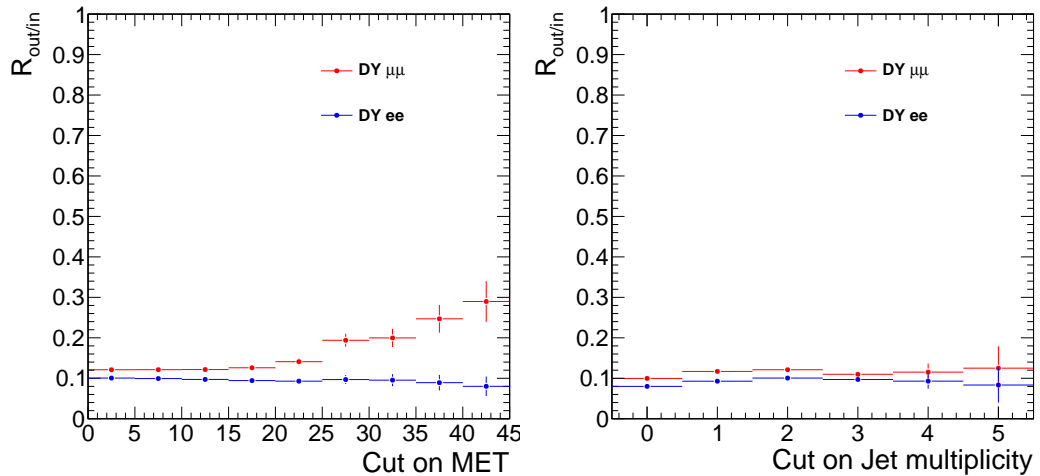

Figure 5.28: - $R_{out/in}$ as a function of the cut applied on MET for events with 2 good leptons and at least 2 jets (left plot) and as a function of the number of jets for events with 2 good leptons (right plot) (PU samples)
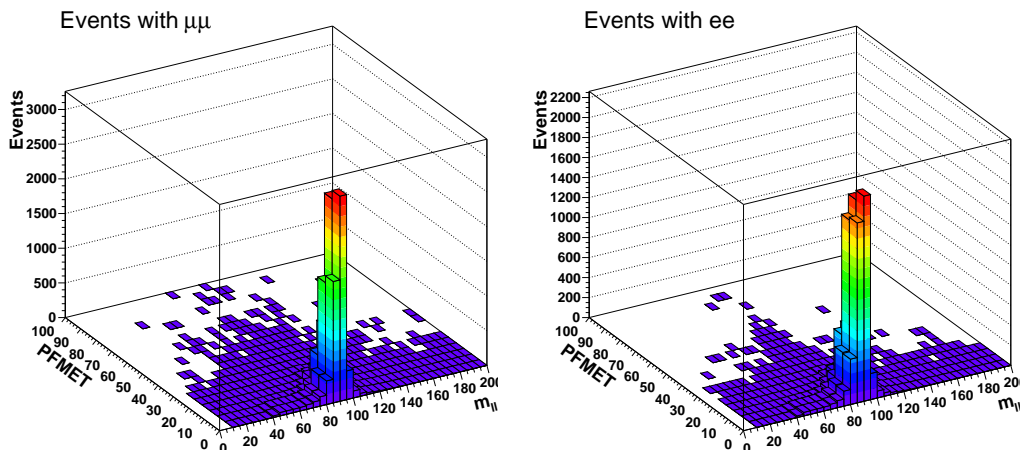
Figure 5.29: - MET vs InvMass for dimuon (left) and dielectron (right) events (NoPU samples)

$R_{out/in}$ is the only quantity taken from simulation and this variation with the selection can be taken as a systematic uncertainty of the method. A 50% uncertainty is assigned.

5.10.2 Background from fake leptons

The goal is to estimate the fake lepton contribution to a signal lepton selection. This estimate is derived from data only and thus does not rely on MC simulation. Any lepton that does not come from a prompt decay such as from W or Z is considered a fake lepton. These include true fakes, e.g. jets taken as electrons, but also real leptons from a heavy flavour decays or electrons from γ conversions.

The method used to study this background follows the formulae for the calculation of contamination due to fake leptons, based on the so-called fakeable method described in [85] and [86].

5.10.2.1 Muon fake rate

The method consists of a determination and an application part where both parts use statistically independent data samples. In the determination part, an inclusive jet trigger is used to select a fake dominated background sample. This sample is used to measure the probability of a loose defined lepton, called Fakeable Object (FO), to pass the signal selection.

The looser definition, relaxes the muon isolation, from relative isolation less than 0.l5 to less than 1, and identification, from global muons to tracker muons, cuts w.r.t to the signal cuts as described in section 5.7. This fake rate is expected to vary within the considered phase space, therefore it is extracted in bins of η and $p_{\rm T}$ of the fake leptons. This ratio is highly dependent on the definition of the denominator. By definition, the numerator, any lepton candidate satisfying the signal requirements, is a subset of the denominator.

For this study several HLT jet triggered data samples were used 5.14. With a 20 GeV/c $p_{\rm T}$ threshold, the data sample is dominated by QCD background and Drell-Yan events. Then, in order to select almost a clean QCD sample, it is required to pass different HLT jet triggers. The HLT triggers used for this analysis are HLT_Jet30U, HLT_Jet50U, HLT_Jet70U and HLT_100U.

Jet triggered samples may content real leptons from W or Z leptonic decays. Lepton from W's are removed by requiring the event to have less than 20 missing transverse energy and less than 20 transverse mass. Leptons from Z's are removed rejecting events with two opposite charge dileptons within the Z mass window [76,106]. An aditional cut mass > 12 GeV is required. In order to avoid any bias from the jet trigger requirement, muons matching a jet in a $\Delta R < 1$ are removed from the estimation of the FR.

The FR is measured as function of the p_T and η of the FO. Figure 5.30 shows a comparison of the FR estimation from the four different HLT jet triggers used. Figure 5.31 shows the comparison of the FR estimation from data and from a enriched muon Monte Carlo sample for a jet threshold of 30 GeV and 50 GeV.

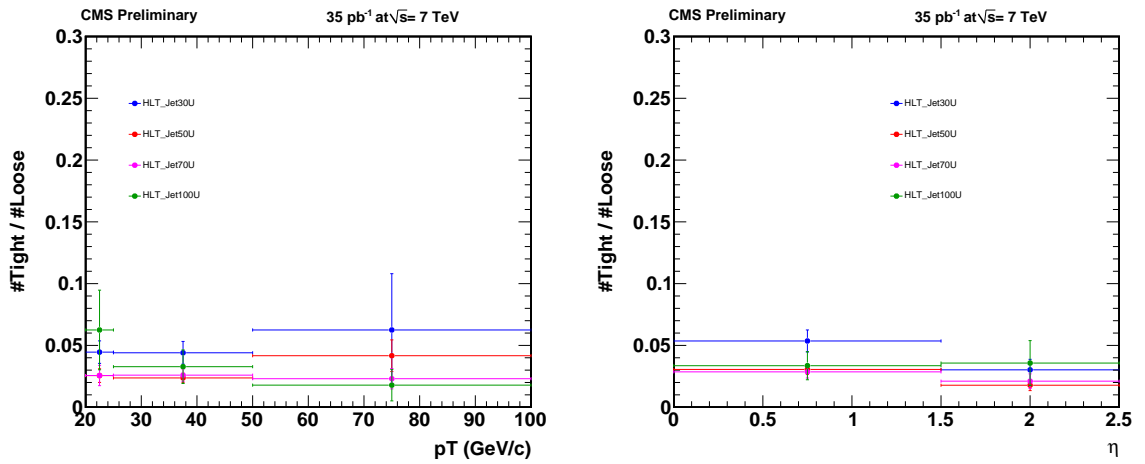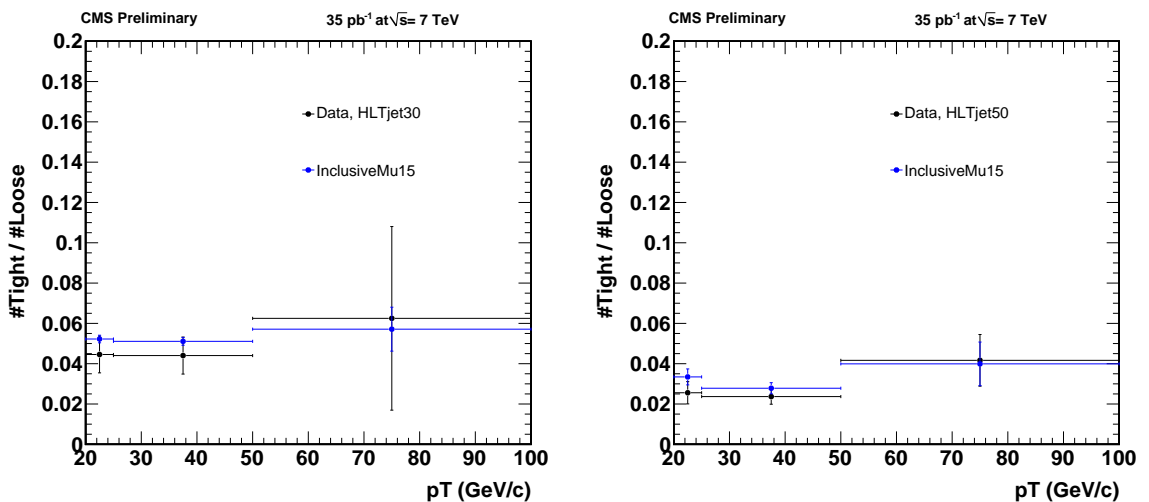**Figure 5.30:** - Tight to loose ratio for muons as a function of pT and η for different HLT jet triggers

Figure 5.31: - Tight to loose ratio for muons from data and Monte Carlo as a function of pT for a HLT_Jet 30 and HLT_Jet 50 triggers

The FR value is taken as the weighted mean of the FR from the different jet triggers and the error assigned is the maximum different between them. Because the fake rate is computed using different data sample than the one to which it will be applied, there will be some systematic error to that calculation. The systematic error is estimated as the maximum difference between observed and predicted values estimated from the different jet samples. Results are shown in table 5.37, a 30% final systematic is assigned to the fake rate estimation.

	Observed	Predicted	(Obs.-Pred.)/Obs.	Error
Jet 30	57.0±7.5	39.0±6.2	0.32	32%
Jet 50	88.0±9.4	107.7±10.4	0.22	22%
Jet 70	55.0±7.4	71.3±8.4	0.30	30%
Jet 100	15.0±3.9	15.4±3.9	0.03	3%

Table 5.37: Predicted and observed number of fakes in the different HLT triggered jet samples used, for the $t\bar{t}$ dilepton analysis. Errors are statistical only.

5.10.2.2   Electron fake rate

The fake electrons originating from jets have typically quite different profile of electromagnetic shower than prompt electrons, while the fake electrons coming from heavy flavor decays can be distinguished from real electrons by their isolation. For this reason, in order to obtain a reasonably good control on the rate of both of these fake electrons, a loose electron has been defined by relaxing both the isolation cut and electron identification cuts.

Loose electrons are defined as those which pass electron ID and conversion rejection cuts which correspond to the so called 90% efficiency working point (WP90). In addition, a relative isolation smaller than 1 is required.

This fake rate is expected to vary within the considered phase space, therefore it is extracted in bins of η and $p_{\rm T}$ of the fake leptons. This ratio is highly dependent on the definition of the denominator. By definition, the numerator, any lepton candidate satisfying the signal requirements, is a subset of the denominator.

The electron fake ratio is measured on the Electron/EG dataset requiring the same electron triggers as those used in the signal selection region. The main motivation for this approach comes form the fact that definitions of several used electron HLT triggers contain tight requirements on electron identification (or even isolation) which bias the effective electron fake rate in the signal selection region. Leptons from W's are removed by requiring the event to have less than 20 missing transverse energy and less than 20 transverse mass. Leptons from Z's are removed rejecting events with two opposite charge dileptons within the Z mass window [76,106].

The measured electron fake rate is shown in figure 5.32 in $p_{\rm T}$ bins. For comparison the electron fake rate from Monte Carlo is also shown.

5.10.2.3   Fake lepton background prediction

In the application part the dilepton sample used to predict the fake background will consist of true dilepton events, e.g., $t\bar{t}$, as well as events with fake leptons, e.g., QCD, W + jets or semileptonic $t\bar{t}$.
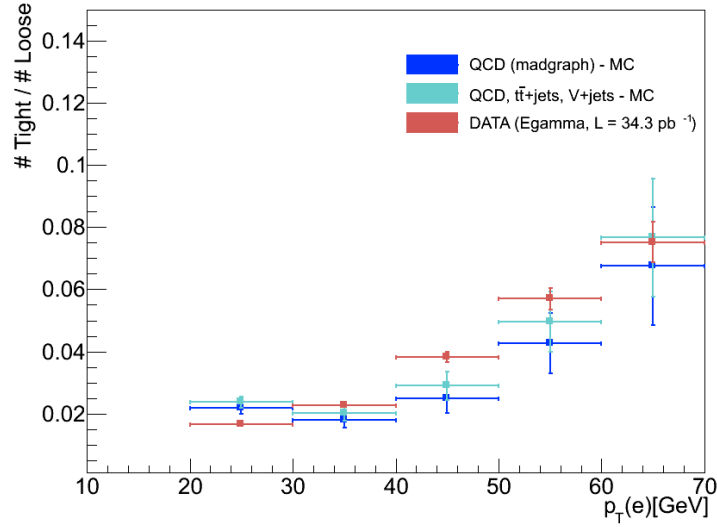
Figure 5.32: - Tight to loose ratio for electrons as a function of pT and η

The existence of true dileptons in the sample complicates the calculation somewhat, since these true dileptons are a major source of fakeable objects. Since true lepton tight-to-loose cut efficiency (p) is not 100% , they will contribute to the prompt+fakeable or fakeable+fakeable selection, and will lead to an overestimate of the fake contribution.

The well known approach is generalized by including the probability for 'prompt leptons' passing loose selection criteria to also pass tight ones. This efficiency is estimated directly from data selecting leptons within the Z mass window [76,106]. In case of muons the efficiency is $\sim$95% , except in the p_{T} range between 20-25 GeV where it drops to $\sim$90% . In case of electrons the average efficiency is $\sim$97% .

Estimates of fake lepton contribution in events with two identified and isolated leptons of opposite charge are shown in tables 5.38, 5.39 and 5.40 at different selection level for a selection analysis. The simulation with fakes is a total of non-dileptonic $t\bar{t}$ and W+jets.

	W+jets	Semileptonic $t\bar{t}$	Data (p = 1)	Data (p from Z)
MET	0	0.08 ± 0.02	$0.27 \pm 0.08 \pm 0.08$	$0.17 \pm 0.09 \pm 0.05$
>= 1btag	0	0.07 ± 0.02	$0.16 \pm 0.06 \pm 0.05$	$0.07 \pm 0.07 \pm 0.02$
>= 2btag	0	0.022 ± 0.010	$0.06 \pm 0.04 \pm 0.02$	$0.01 \pm 0.04 \pm 0.003$

Table 5.38: Estimates of the fake muon fake lepton contribution to the $t\bar{t}$ muon-muon channel. Statistical and systematic uncertainties are quoted

	W+jets	Semileptonic $t\bar{t}$	Data (p = 1)	Data (p from Z)
MET	0.2 ± 0.1	0.38 ± 0.04	$0.25 \pm 0.11 \pm 0.13$	$0.19 \pm 0.12 \pm 0.10$
>= 1btag	0	0.35 ± 0.04	$0.16 \pm 0.11 \pm 0.08$	$0.13 \pm 0.10 \pm 0.07$
>= 2btag	0	0.18 ± 0.003	$0.08 \pm 0.08 \pm 0.04$	$0.08 \pm 0.08 \pm 0.04$

Table 5.39: Estimates of the fake muon fake lepton contribution to the $t\bar{t}$ electron-electron channel. Statistical and systematic uncertainties are quoted

	W+jets	Semileptonic $t\bar{t}$	Data (p = 1)	Data (p from Z)
Njets >= 2	0.5 ± 0.2	1.01 ± 0.07	$1.75 \pm 0.28 \pm 0.80$	$1.40 \pm 0.29 \pm 0.70$
>= 1btag	0.07 ± 0.07	0.92 ± 0.42	$0.88 \pm 0.21 \pm 0.44$	$0.68 \pm 0.21 \pm 0.34$
>= 2btag	0	0.42 ± 0.04	$0.15 \pm 0.06 \pm 0.07$	$0.050 \pm 0.06 \pm 0.03$

Table 5.40: Estimates of the fake muon fake lepton contribution to the $t\bar{t}$ electron-muon channel. Both statistical and systematic uncertainties are quoted

5.11 Systematic uncertainties

Several sources of systematic uncertainties are considered in the analysis: detector performance and experimental effects and biases which affect the selection efficiency of both signal and background events, the determination of the Jet Energy Scale (JES) and Jet Energy Resolution (JER), the imperfect understanding of b-tagging and mistag efficiencies, the modeling of the reconstruction and selection of events (mainly lepton identification and isolation), and generator parameters describing process kinematics and hadronization.

The uncertainties specific from the methods used to estimate the background contamination are also taken into account, as well as the uncertainty on the determination of the integrated luminosity, which is currently 4% and is considered separately.

Details on the determination of these systematic uncertainties and their effect on the cross section measurement are described in this section.

5.11.1 Lepton selection

Systematic uncertainties due to lepton and trigger efficiencies are taken from the studies described in Section 5.8 based on tag and probe methods, which rely on the differences observed between the efficiencies measured in data and simulation.

The scale factors obtained per dilepton channel are $SF^{\mu\mu} = 0.9673\pm 0.0125$, $SF^{ee} = 0.9231\pm 0.0176$ and $SF^{e\mu} = 0.9474\pm 0.0105$. Therefore, the systematic uncertainty assigned to the lepton selection is 1.3% , 1.9% and 1.1% for $\mu^+\mu^-$, e^+e^- and $e^\pm\mu^\mp$ events respectively.

5.11.1.1 Lepton selection model

In addition to the differences observed between data and simulation in Z events, and as described in Section 5.8, the discrepancies due to the different topologies of $t\bar{t}$ and Z+jets events have to be taken into account. Considering the average of the difference in isolation efficiencies found between these two samples, a 4% uncertainty is added to account for this effect.

5.11.2 Decay branching ratio

The branching fraction of $W \longrightarrow l\nu$ decays used in the MADGRAPH [87] sample is set to its LO value of 1/9. The current world average is 0.1080 ± 0.0009 [8]. To correct for this, a scale factor has to be applied, and $t\bar{t}$ dilepton events have to be weighted with $(0.108*9)^2 = 0.9448$. The contribution to the relative error on the cross section is 1.7% .

5.11.3 Jet Enery Scale

This is one of the main sources of systematic uncertainty on the $t\bar{t}$ cross section measurement. An uncertainty in the Jet Energy Scale (JES) affects the p_{T} distribution of the jets which can bias the selection, besides, this has also a potential impact on the $\not{E}_T$ selection.

The uncertainty of the overall jet energy scale in a region of p_{T} from 15 GeV up to 2 TeV was found to be from 3 to 6% according to the studies performed by the Jet Energy Correction group in CMS [56]. These uncertainties have been provided in a set of p_{T} -η bins and are accessible from a database. On top of that, additional uncertainties are recommended to be added to account for other

effects: 1.5% due to differences in software releases and calibration changes and between 2% and 3% (depending on the jet p_T) for the b-jet scale, which is applied to those jets identified as coming from a b-quark in simulation.

The p_T of each jet is scaled up and down by the total uncertainty δ and the difference between the nominal selection and the shifted one is then taken as the systematic uncertainty:
$\delta = \sqrt{unc^2 + cor^2}$, where unc = (p_T ,η) uncertainties and $cor^2 = c_{SW}^2 + c_{bJES}^2$

- $c_{SW} = 0.015$
- $c_{bJES} = 0.02$ for jets with ($50 < p_T < 200$ GeV and $|\eta| < 2.0$) and $c_{bJES} = 0.03$ otherwise

This uncertainty is also propagated to the $\not{E}_T$, through the vectorial component of the p_T shifted jets. The 2 jets that are matched to the selected leptons within $\Delta R < 0.4$ are removed from the $\not{E}_T$ computation.

Results are summarized in Table 5.41. As can be observed, the systematic uncertainties are about 3.6% in e^+e^-, $\mu^+\mu^-$channels and 2.8% in the $e^\pm\mu^\mp$channel. The uncertainty is smaller in the $e^\pm\mu^\mp$channel as there is no $\not{E}_T$ cut applied.

$t\bar{t}$ signal	MET	≥ 1 b-tag
$\mu\mu$	+3.21%	+3.14%
	−3.93%	−3.68%
ee	+3.57%	+3.54%
	−4.06%	−3.46%
$\mu\mu$ +**ee**	+3.37%	+3.32%
	−3.99%	−3.58%
$e\mu$	+2.68%	+2.67%
	−2.84%	−2.85%

Table 5.41: JES systematics estimated taking the p_T , η uncertainties from the database and adding in quadrature the contribution due to software and b-jet scale

Another contribution that has to be taken into account is the uncertainty due to the unclustered $\not{E}_T$, to estimate this effect the part of the $\not{E}_T$ not clustered in jets is scaled by up and down by 10% . Results are given in Table 5.42. This affects only the $e^+e^-$and $\mu^+\mu^-$channels, where a $\not{E}_T$ cut is applied, and it is treated as independent to the Jet Energy Scale. Anyway the effect is small, and it is estimated to be of the order of 1 % . Both uncertainties, the one coming from the Jet Energy Scale and the one coming from the unclustered energy, are added in quadrature and a final uncertainty of 3.8% for e^+e^-, $\mu^+\mu^-$and 2.8% for $e^\pm\mu^\mp$channel is derived.

The impact on the selection due to the uncertainty on the Jet Energy Scale has been also estimated for those backgrounds that are taken from simulation, results are presented per channel in Table for all channels combined in Table 5.43. Because of the small number of background events remaining after the full selection is applied, all channels together are considered in the estimation of the uncertainty. A 12% uncertainty is obtained for $Z/\gamma^* \to \tau^+\tau^-$ events, and about 7% for dibosons and single top.

$t\bar{t}$ signal	MET	≥ 1 b-tag
$\mu\mu$	-1.11%	-1.10%
	$+1.03\%$	$+1.10\%$
ee	-0.52%	-0.54%
	$+1.26\%$	$+1.34\%$
$\mu\mu$ +**ee**	-0.85%	-0.85%
	$+1.13\%$	$+1.21\%$

Table 5.42: Systematics due to unclustered MET, obtained considering a 10% uncertainty

	MET	≥ 1 b-tag
$Z/\gamma^* \to \tau^+\tau^-$	$+13.90\%$	$+11.83\%$
	-12.74%	-7.89%
VV	$+9.49\%$	$+6.68\%$
	-7.52%	-7.42%
Single top	$+6.55\%$	$+6.00\%$
	-5.97%	-5.79%

Table 5.43: JES systematics for the backgrounds from simulation, estimated taking the p_T, η uncertainties from the database and adding in quadrature the contribution due to software and b-jet scale. All channels combined are used: $e^+e^- + \mu^+\mu^- + e^\pm\mu^\mp$

5.11.4   Jet Energy Resolution

Another source of uncertainty affecting the jets arises from the jet energy resolution. According to the studies performed by the JetMET group [88] the resolution appears to be about 10% worse in data compared to simulation: $\sigma_{data}(jetp_T) \sim 1.1 \cdot \sigma_{MC}(jetp_T)$ and also it is expected that the uncertainty on the value of the jet energy resolution is of the same order : 10% of $\sigma_{data}(\text{jet } p_T)$.

To estimate the impact on the analysis, the reconstructed jet p_T relative to the generator-level jet is scaled up by 10% (this will be taken as the nominal value), then it's scaled up and down by 10% and the variation with respect to the nominal value is taken as an estimation of the systematic uncertainty. Results are summarized in Table 5.44 and, as can be observed, the effect is small, of the order of 1%

$t\bar{t}$ signal	Before b-tag	≥ 1 b-tag
$\mu\mu$	-0.77%	-0.83%
	$+1.15\%$	$+1.02\%$
ee	-0.90%	-0.92%
	$+0.98\%$	$+1.07\%$
$e\mu$	-0.81%	-0.91%
	$+0.77\%$	$+0.82\%$
$\mu\mu$ +**ee**	-0.83%	-0.87%
	$+1.07\%$	$+1.04\%$

Table 5.44: Systematic uncertainties due to the Jet Energy Resolution

5.11.5   Pile-up effect

The presence of pile-up events affects the lepton isolation and the jet and $\not{E}_T$ distributions. The effect on the lepton efficiencies is implicitly taken into account in the tag and probe method, while the effect on the jet and $\not{E}_T$ reconstruction can be estimated comparing simulated samples generated with and without pile-up and a scale factor can be derived to account for this effect.

Considering all the modes combined, a $1.57\pm 0.60\%$ increase in events passing full selection is found in the sample with pile-up [81]. The pile-up MC samples available were generated with approximately 25% extra pp collisions than those present actually in data so this value has to be corrected when computing the scale factor. After taking this into account the scale factor due to pileup effects is estimated to be $SF_{PU} = 1.0126 \pm 0.0048 \pm 0.0063 \approx 1.013 \pm 0.008$, where a half of the shift is assigned as a systematic uncertainty [81].

5.11.6   b-tag systematics

Another important source of systematic uncertainty is due to the limited knowledge of b-tagging and mistag rates. This analysis is using the most simple b-tagger developed by the CMS Collaboration and b-tagging remains important to discriminate the signal and reject backgrounds almost completely.

As described in Section 5.7, the b-tagging algorithm used in this analysis is the TrackCountingHighEfficiency, which uses the impact parameter significance of the second track in the jet as the discriminant variable to identify a jet as a b-jet. The shape of the discriminator is shown in Figure 5.33 separately for b-jets, c-jets and light-jets (u,d,s and gluons).
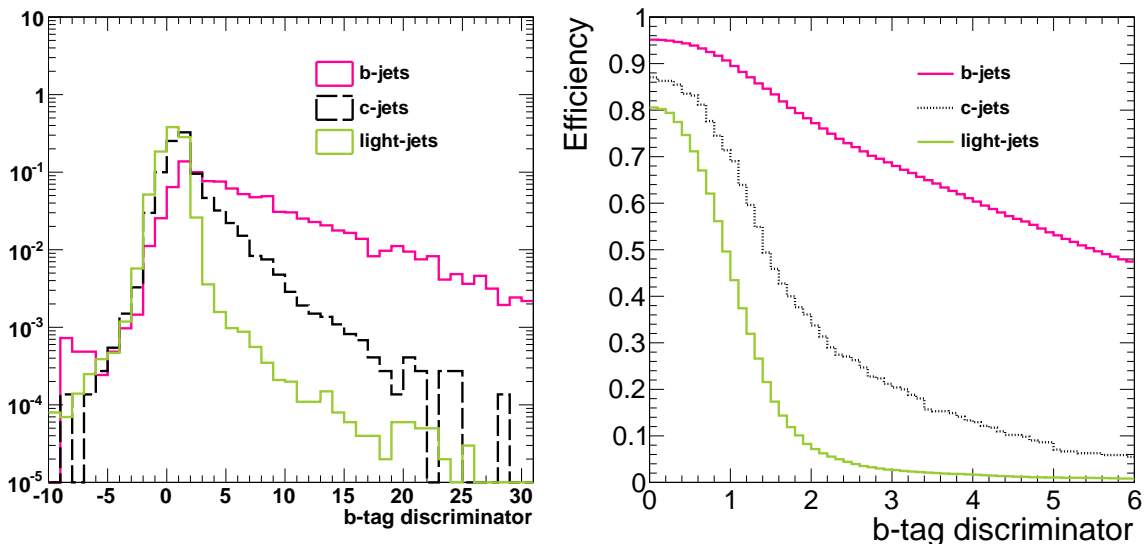
Figure 5.33:   - Distribution of the b-tagging discriminator for b,c and light jets (left) and b-tagging efficiency and mistag rates as a function of the b-tag discriminator cut (right) for a $t\bar{t}$ MC sample

The performance curves versus the b-tag discriminator cut are shown in Figure 5.33 for a simulated sample of $t\bar{t}$ dilepton events. As can be inferred from the plot, the efficiencies associated to the working point used in the analysis (1.7) are about 0.81 for b-jets, 0.38 for c-jets and 0.12 for light jets, which may be different from the efficiencies measured in data.

The data-MC agreement of the TCHE discriminator at different levels of the selection is shown in Figures 5.34, 5.35 and 5.36.

Tagging efficiencies and mistagging rates are measured in data by the b-tag POG, who provide data to MC scale factors for several specific working points in each of the b-tagging discriminants that exist in CMS. Relying on these studies, the following scale factors and uncertainties are considered in this analysis: $SF_b = 1.0 \pm 0.1$ for b-jets, $SF_c = 1.0 \pm 0.2$ for c-jets and $SF_l = 1.0 \pm 0.25$ for light-jets.

To determine the b-tag systematic uncertainty, the following approach has been used: First of all, the scale factors are applied to the above mentioned efficiencies associated to the working point 1.7, according to the jet flavour, leading to new efficiencies that, taking the performance curve, will correspond to new b-tag discriminant cuts. The analysis is performed applying these cuts for each jet flavour and the number of events obtained this way will become the nominal selection.

Then, these new efficiencies and mistag rates associated to these new cut values are shifted up and down by their corresponding uncertainties (10% for b-jets, 20% for c-jets and 25% for light-jets) to derive a new set of b-tag cuts. Finally, the analysis is repeated using this set of b-tag cuts and the difference between the number of events obtained after applying these shifts and the nominal number of events obtained after applying the scale factors is taken as the systematic uncertainty.

b,c efficiencies are varied simultaneously and independently to the mistag, which is finally added in quadrature and a total positive and negative systematic uncertainty is reported. Results are summarized in Table 5.45 for the $t\bar{t}$ signal and events with at least 1 b-tagged and 2 b-tagged jets. Uncertainties per b-tag bin have been also estimated and are presented in Table 5.46.

An illustration of the variation in the number of b-tagged jets after applying the scale factors and after applying the shifts is shown in Figure 5.37. As expected, in $t\bar{t}$ events the main contribution comes from the b-jet uncertainty and the effect due to mistags is very small.
For the signal, it is found an average value of the systematic relative error of 5% for events with at least 1 b-tagged jet.

Figure 5.38 shows the number of b-tagged jets after applying the full selection without b-tagging, the uncertainties on the b-tag efficiency for signal events previously described are reflected by the hatched bands.

The b-tagging systematic uncertainty has also been estimated for the backgrounds that are taken from simulation following the same procedure described before. The results are given in Table 5.47 for all the final states combined. A 21 % uncertainty is estimated for $Z/\gamma^* \rightarrow \tau^+\tau^-$ and dibosons, while a 8 % is found for single top events.
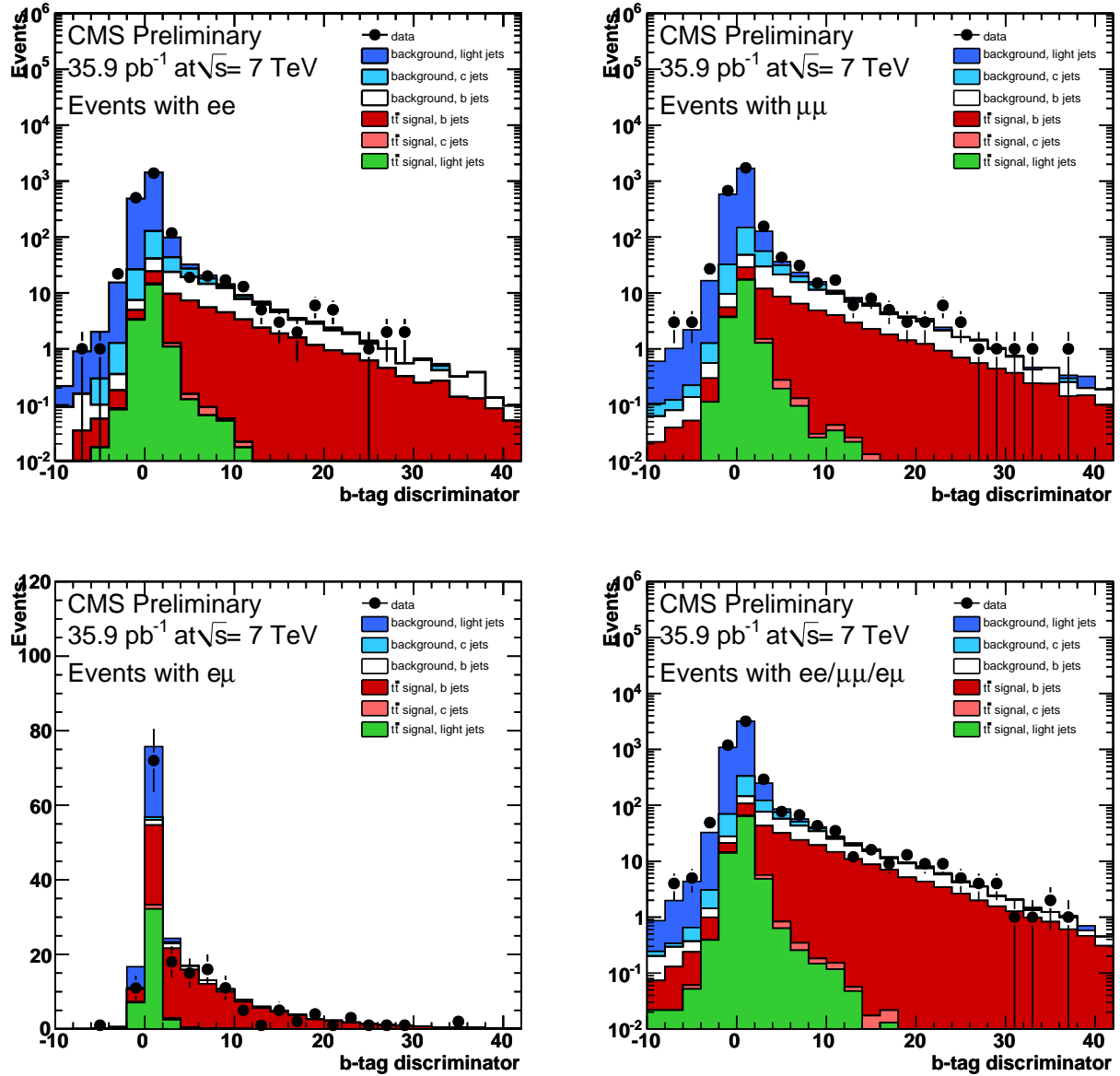
Figure 5.34: - Distribution of the b-tagging discriminant in events with at least one jet and two oppositely charged leptons compared to signal and background expectations from simulation. All background contributions are combined and displayed separately based on the flavor of the simulated jet. Essentially all the visible background is dominated by the $Z/\gamma^* \to l^+l^-$ contribution. The distributions are for e^+e^-, $\mu^+\mu^-$, $e^\pm\mu^\mp$ and all channels combined
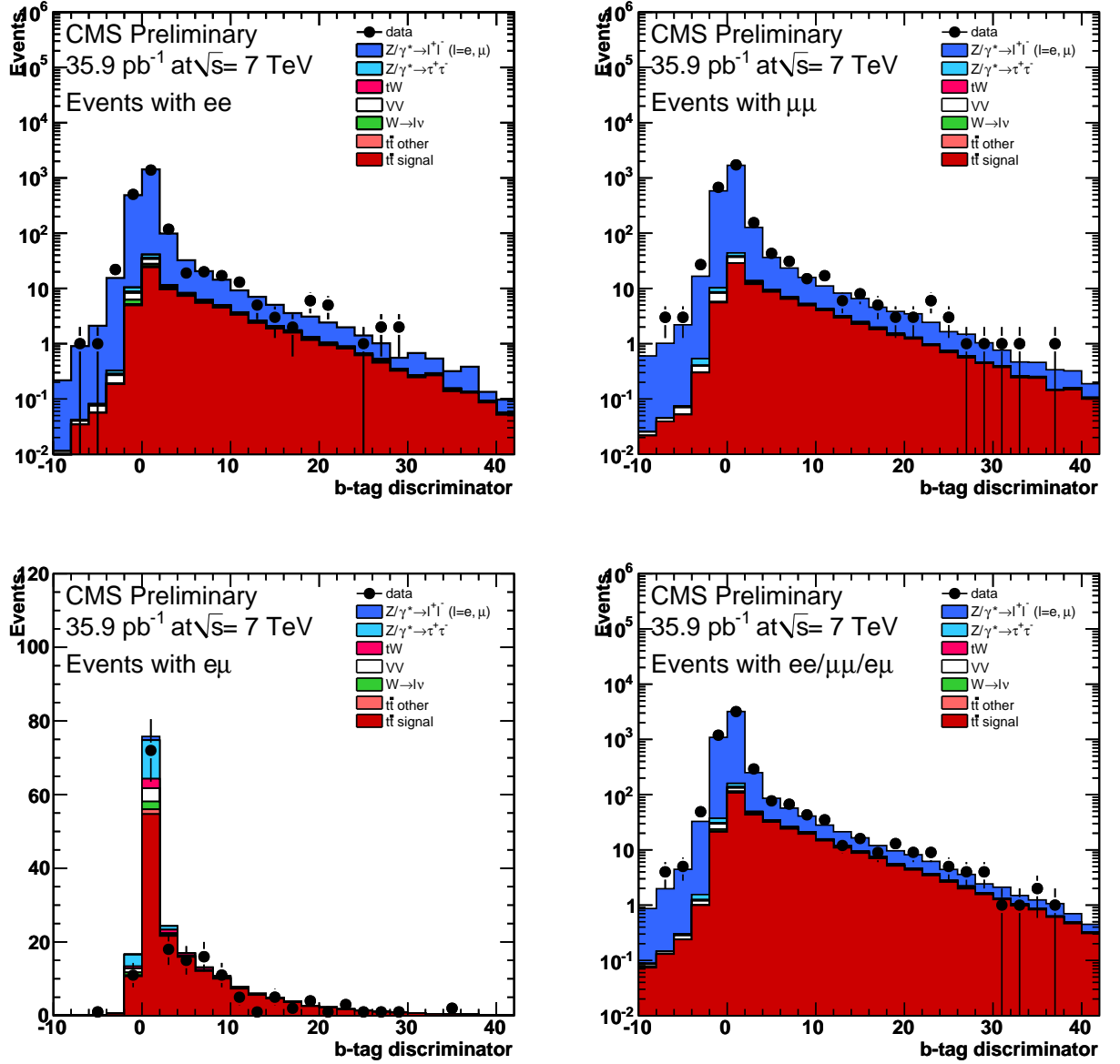
Figure 5.35: - Distribution of the b-tagging discriminant in events with at least one jet and two oppositely charged leptons compared to signal and background expectations from simulation. The distributions are for e^+e^-, $\mu^+\mu^-$, $e^\pm\mu^\mp$ and all channels combined
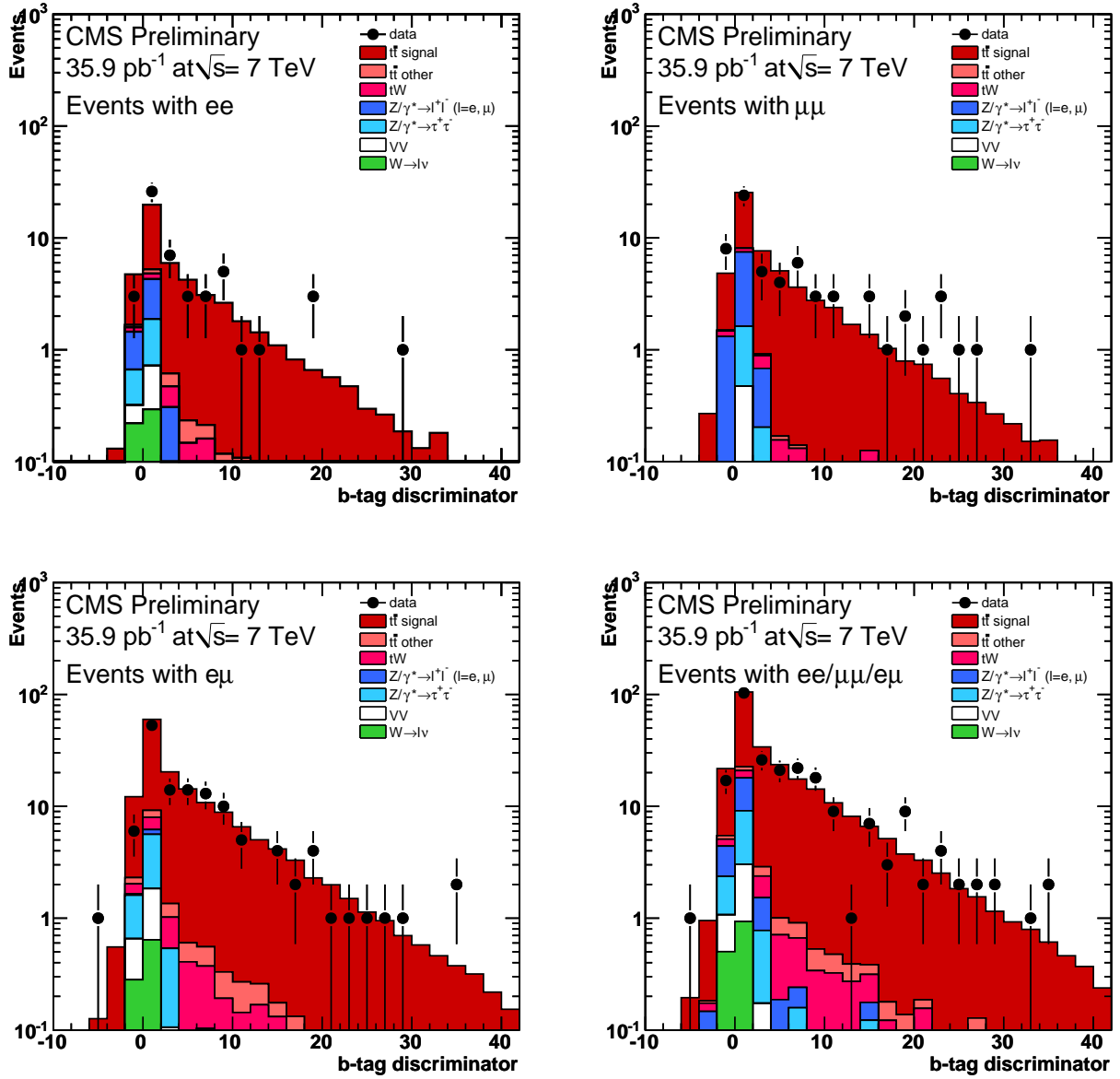
Figure 5.36: - Distribution of the b-tagging discriminant in events with at least two jets and passing full analysis selection criteria prior to applying a b-tagging requirement compared to signal and background expectations from simulation. The distributions are for e^+e^-, $\mu^+\mu^-$, $e^\pm\mu^\mp$ and all channels combined
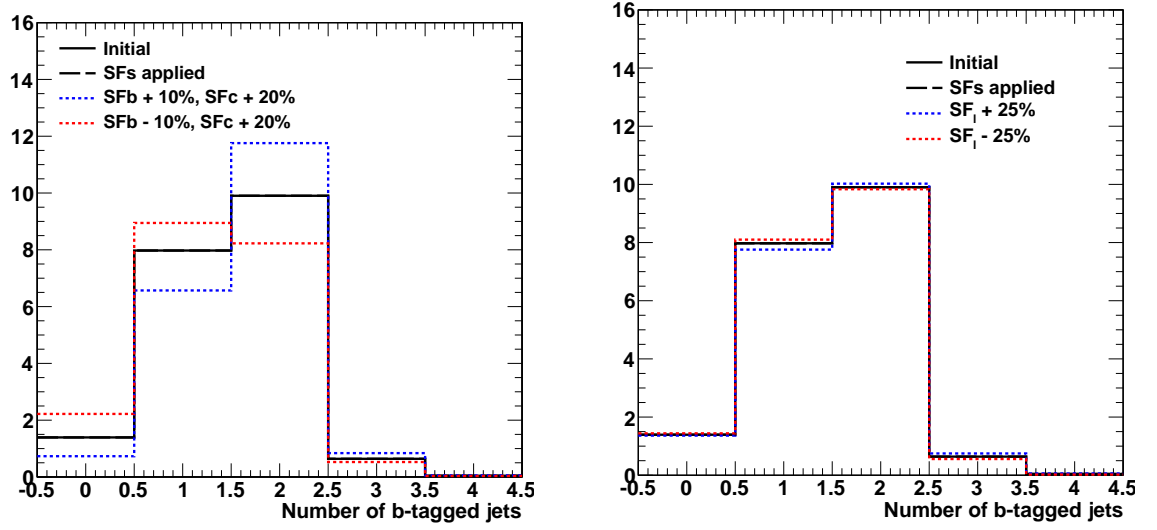
Figure 5.37: - Number of b-tagged jets before the b-tagging cut obtained after applying the scale factors on b, c and mistag efficiencies and after varying these efficiencies by their relative uncertainties.

$t\bar{t}$ signal	$\pm10\%$ b-eff $\pm20\%$ c-eff	$\pm25\%$ mistag	Total uncertainty
$\mu\mu$ $\geq$ **1 btag**	$+3.71\%$ -4.63%	$+0.35\%$ -0.29%	$+3.73\%$ -4.64%
$\geq$ **2 btag**	$+18.93\%$ -16.75%	$+2.14\%$ -2.09%	$+19.05\%$ -16.85%
ee $\geq$ **1 btag**	$+4.06\%$ -4.52%	$+0.32\%$ -0.35%	$+4.07\%$ -4.53%
$\geq$ **2 btag**	$+20.10\%$ -17.38%	$+2.39\%$ -2.03%	$+20.24\%$ -17.50%
$e\mu$ $\geq$ **1 btag**	$+3.75\%$ -4.74%	$+0.36\%$ -0.31%	$+3.77\%$ -4.75%
$\geq$ **2 btag**	$+19.53\%$ -17.31%	$+2.24\%$ -2.10%	$+19.66\%$ -17.44%
$ee + \mu\mu + e\mu$ $\geq$ **1 btag**	$+3.80\%$ -4.68%	$+0.35\%$ -0.31%	$+3.82\%$ -4.69%
$\geq$ **2 btag**	$+19.50\%$ -17.20%	$+2.25\%$ -2.08%	$+19.63\%$ -17.32%

Table 5.45: b-tagging systematic uncertainties for the signal, obtained applying $SF_b = 1.0$, $SF_c = 1.0$ and $SF_l = 1$ and considering a 10% uncertainty on b-efficiency, 20% uncertainty on c-efficiency and 25% uncertainty on mistag rate.

$t\bar{t}$ signal	$\pm10\%$ b-eff $\pm20\%$ c-eff	$\pm25\%$ mistag	Total uncertainty
$\mu\mu$			
0 btag	-45.97% $+57.27\%$	-4.34% $+3.58\%$	-46.17% $+57.38\%$
1 btag	-16.64% $+11.59\%$	-2.04% $+2.12\%$	-16.76% $+11.78\%$
2 btag	$+18.71\%$ -16.48%	$+0.47\%$ -0.71%	$+18.72\%$ -16.50%
$\geq$ 3 btag	$+21.21\%$ -19.58%	$+19.58\%$ -16.51%	$+28.87\%$ -25.61%
ee			
0 btag	-47.20% $+52.47\%$	-3.75% $+4.08\%$	-47.35% $+52.63\%$
1 btag	-17.11% $+12.46\%$	-2.40% $+1.87\%$	-17.28% $+12.60\%$
2 btag	$+19.83\%$ -17.11%	$+0.89\%$ -0.61%	$+19.85\%$ -17.12%
$\geq$ 3 btag	$+23.21\%$ -20.45%	$+18.81\%$ -17.69%	$+29.88\%$ -27.0%
$e\mu$			
0 btag	-45.62% $+57.65\%$	-4.34% $+3.77\%$	-45.83% $+57.78\%$
1 btag	-17.39% $+12.11\%$	-2.17% $+2.08\%$	-17.53% $+12.29\%$
2 btag	$+19.15\%$ -16.93%	$+0.55\%$ -0.53%	$+19.16\%$ -16.94%
$\geq$ 3 btag	$+23.65\%$ -21.49%	$+21.52\%$ -19.90%	$+31.98\%$ -29.29%
$ee + \mu\mu + e\mu$			
0 btag	-45.99% $+56.61\%$	-4.23% $+3.79\%$	-46.18% $+56.74\%$
1 btag	-17.18% $+12.06\%$	-2.19% $+2.05\%$	-17.34% $+12.23\%$
2 btag	$+19.18\%$ -16.86%	$+0.59\%$ -0.58%	$+19.19\%$ -16.87%
$\geq$ 3 btag	$+23.02\%$ -20.87%	$+20.59\%$ -18.73%	$+30.89\%$ -28.0%

Table 5.46: b-tagging systematic uncertainties for the signal, obtained applying $SF_b = 1.0$, $SF_c = 1.0$ and $SF_l = 1$ and considering a 10% uncertainty on b-efficiency,20% uncertainty on c-efficiency and 25% uncertainty on mistag rate.
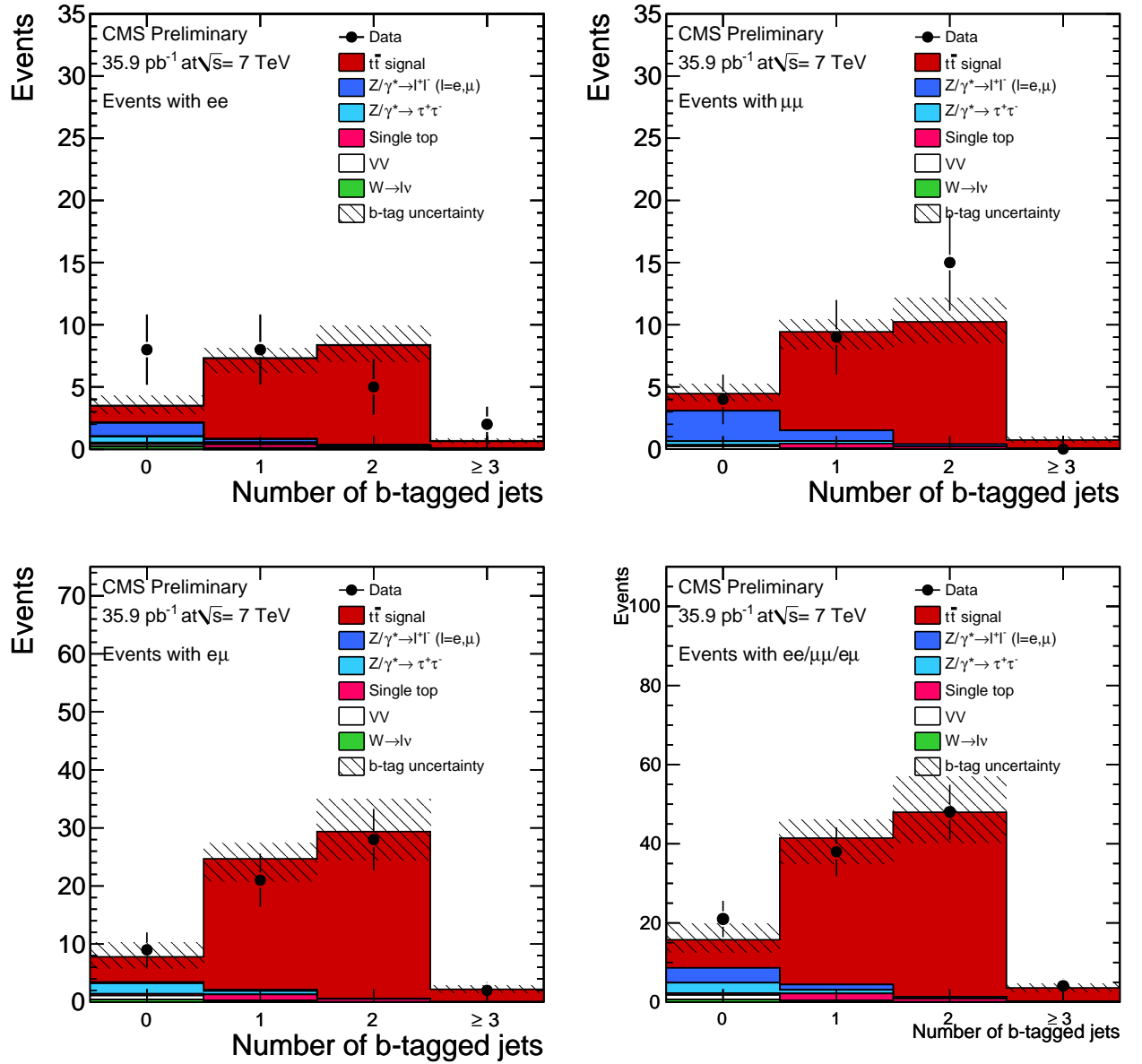
Figure 5.38: - Multiplicity of b-tagged jets in events passing full dilepton selection criteria with at least two jets compared to signal and background expectations from simulation. The uncertainty on the number of signal events corresponding to the uncertainty in the selection of b-tagged jets is displayed by the shaded area. The distributions are for e^+e^-, $\mu^+\mu^-$, $e^\pm\mu^\mp$ and all final states combined.

	±10% b-eff ±20% c-eff	±25% mistag	Total uncertainty
DY $\tau\tau$ $\geq$ **1 btag**	+3.94% −3.94%	+11.87% −19.72%	+12.51% −20.11%
$\geq$ **2 btag**	+25.0% −0.0%	+0.0% −0.0%	+25.0% −0.0%
VV $\geq$ **1 btag**	+2.54% −4.15%	+13.17% −20.22%	+13.41% −20.64%
$\geq$ **2 btag**	+7.42% −15.73%	+25.92% −30.19%	+26.96% −34.04%
tW $\geq$ **1 btag**	+7.06% −7.45%	+0.89% −0.30%	+7.12% −7.46%
$\geq$ **2 btag**	+22.04% −20.19%	+4.74% −4.65%	+22.54% −20.72%

Table 5.47: b-tagging systematic uncertainties for the backgrounds, obtained applying $SF_b = 1.0$, $SF_c = 1.0$ and $SF_l = 1.0$ and considering a 10% uncertainty on b-efficiency, 20% uncertainty on c-efficiency and 25% uncertainty on mistag rate. All channels combined are considered

5.11.7 Theoretical systematic uncertainties

Uncertainties due to the signal modeling are evaluated by comparing the nominal sample with additional ones that have been produced with varied parameters and configurations. The most important effects arise from the presence of additional hadronic jets produced by the QCD radiation in the initial and final states, the uncertainty on the Q^2 scale determining the value of the strong coupling constant, on the modeling of the decay particles, hadronization and showering model, the uncertainty on the top quark mass or in the parton distribution functions. A summary of this effects is given in this section.

The effect of the initial and final state radiation (ISR/FSR) on the signal efficiency was evaluated by using MC samples generated with more and less initial and final state radiation. The effect of the factorization scale is estimated using samples with the scale increased and decreased by a factor of 2. Dedicated samples are also used to estimate the uncertainty due to the parton-jet matching.

The impact of the top mass uncertainty on the signal selection efficiency is also estimated using 2 samples with a top mass of 166.5 GeV/c^2 and 178.5 GeV/c^2, which are compared to the default sample with a generated mass of 172.5 GeV/c^2. Therefore, per 6 GeV/c^2 a 5% uncertainty is obtained, although this is too pessimistic as the uncertainty on the top mass is less than 2 GeV/c^2. Scaling the result to 2 GeV/c^2, a 1.7% systematics is derived.

A summary of the results is given in Table 5.48 .

$t\bar{t}$ signal	$e\mu$	$\mu\mu$ +**ee**
ISR/FSR	1.33%	0.84%
$Q^2 scale$	1.78%	3.68%
Parton-jet matching	0.075%	1.94%
Top Mass (6 GeV)	4.76%	5.56%

Table 5.48: Theoretical systematics estimated after full selection with b-tag

Systematic effects due to jet and $\not{E}_T$ and shower modeling are estimated using samples produced with different generators: ALPGEN [89] and POWHEG [90] and compared to the default used in the analysis (MADGRAPH [87]).

Efficiencies of the jet and $\not{E}_T$ selection with respect to the selected dilepton events for different generators is shown in Table 5.49.

$t\bar{t}$ signal	$\mu\mu$	**ee**	$e\mu$	$\mu\mu$ +**ee**
MADGRAPH	47.77%	46.98%	74.40%	47.41%
ALPGEN	49.35%	48.72%	74.72%	49.06%
POWHEG PYTHIA	48.91%	48.89%	74.69%	48.90%
POHEG HERWIG	49.16%	49.20%	75.27%	49.18%

Table 5.49: Jet and $\not{E}_T$ selection efficiencies obtained using simulated samples produced with different generators

The relative variation between the different generators considered after jet and $\not{E}_T$ selections is

summarized in Table 5.50.

$t\bar{t}$ signal	$\mu\mu$	**ee**	$\mu\mu$ +**ee**	$e\mu$
ALPGEN wrt MADGRAPH	3.3%	3.7%	3.5%	0.43%
POWHEG-PYTHIA wrt MADGRAPH	2.4%	4.1%	3.1%	0.39%
POWHEG PYTHIA wrt HERWIG	-0.51%	-0.63%	-0.57%	-0.77%

Table 5.50: Jet and $\not{E}_T$ Systematic uncertainties derived by comparing simulated samples produced with different MC generators

Systematic uncertainty due to the jet and $\not{E}_T$ modeling is estimated comparing ALPGEN, POWHEG-PYTHIA and MADGRAPH samples, taken an average value of the difference observed between each sample and the MADGRAPH one used in the analysis. Systematic associated to the showering model is estimated comparing the two POWHEG samples available, one showered by PYTHIA [91] and the other one by HERWIG [92].

5.11.8 Background estimation

Uncertainties on the background estimation in data for $Z/\gamma^* \to l^+l^-$ and events with fake leptons have been already discussed in Sections 5.10 and 5.10.2, so in the following it will be described the uncertainties from the backgrounds estimated from simulation: single-top (tW), diboson production (WW, WZ, ZZ) and $Z/\gamma^* \to \tau^+\tau^-$ where both taus decay to electrons or muons.

The background events are normalized to the predicted NLO cross section value and a conservative 30% uncertainty is considered in this calculation. Other systematic effects are coming from the same sources as for the $t\bar{t}$ signal. The main effect arises from the jet energy scale, which was estimated in Section 5.11.3 (12% for $Z/\gamma^* \to \tau^+\tau^-$, 7% for dibosons and tW), and from the b-tagging, estimated in Section 5.11.6 (21% for $Z/\gamma^* \to \tau^+\tau^-$ and dibosons, 8% for tW). A 4% uncertainty has to be also added due to the luminosity.

Considering all these effects, a 38% uncertainty is assigned to $Z/\gamma^* \to \tau^+\tau^-$ and dibosons and 32% for tW.

5.11.9 Luminosity

The primary method to normalize the data sample at CMS is using a method of counting collision events using a minimal requirement for event selection by the data acquisition system. A presence of a proton-proton collision can be identified from a reconstructed collision vertex using charged particle tracks or from a significant measurement in the calorimeters. The measured rate of proton-proton collisions is normalized to the absolute instantaneous luminosity estimate using Van der Meer method in a special fill of the LHC. The uncertainty on the total integrated luminosity is 4% [93].

5.11.10 Summary of the systematic effects

A summary of all the systematic uncertainties that are considered in the analysis is given in Table 5.51 for the signal and 5.52 for the backgrounds.

Source	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$
Lepton selection	1.9	1.3	1.1
Energy scale	3.8	3.8	2.8
Pile-up	0.8	0.8	0.8
b-tagging ($\geq$ 1 b-tag)	5.0	5.0	5.0
Lepton selection model	4.0	4.0	4.0
Branching ratio	1.7	1.7	1.7
Decay model	2.0	2.0	2.0
Event Q^2 scale	2.3	2.3	1.7
Top quark mass	2.6	2.6	1.5
Jet and $\not{E}_T$ model	3.2	3.2	0.4
Shower model	0.7	0.7	0.7
Total without luminosity	9.5	9.4	8.0
Luminosity	4.0	4.0	4.0

Table 5.51: Summary of systematic uncertainties on the signal selection considered in the analysis

Source	$Z/\gamma^* \to \tau^+\tau^-$	dibosons	tW
Cross section	30	30	30
Energy Scale	12	7	6
b-tagging ($\geq$ 1 b-tag)	21	21	8
Total without luminosity	38	37	32
Luminosity	4.0	4.0	4.0

Table 5.52: Summary of systematic uncertainties on the backgrounds estimated from simulation

5.12 Cross section measurement

As described in Section 5.2, the $t\bar{t}$ dilepton presents a clean final state, therefore a simple counting experiment is used to compute the cross section. The events observed in data that pass the selections are compared to the expected background estimates and the excess is attributed to the $t\bar{t}$ dilepton production.

The $t\bar{t}$ production cross section can be measured using the expression:

$$\sigma_{t\bar{t}} = \frac{N - B}{A \cdot L} \tag{5.4}$$

where N is the number of observed events in data, B is the number of estimated background events, A is the total acceptance including the geometric acceptance and the event selection efficiency, already corrected for differences between data and simulation, and L is the integrated luminosity used to normalize the sample.

The $t\bar{t}$ signal yields reported in Section 5.9 are corrected with a scale factor, that is computed as the product of the individual scale factors that has been discussed previously and that account for the differences in lepton efficiencies (Section 5.8.4), branching ratio (Section 5.11.2) and pile-up effect (Section 5.11.5). The combination of all these contributions yields the following scale factors per channel:

- $\mu^+\mu^-$ final state $\rightarrow SF_{all}^{\mu\mu} = 0.926$

- e^+e^- final state $\rightarrow SF_{all}^{ee} = 0.883$

- $e^{\pm}\mu^{\mp}$ final state $\rightarrow SF_{all}^{e\mu} = 0.906$

A summary of the expected number of signal and background events is compared with the number of events observed in data in Table 5.53. As can be observed, a good agreement is found between the expectation and the number of events in data in all channels. The expected signal events are corrected as described above and two uncertainties are quoted, the first value is coming from the total systematics estimated in Table 5.51 and the second one is a combination of the 15% uncertainty on the NLO cross section and a 4% on the luminosity.

Contribution from Drell-Yan $ee, \mu\mu$ and events with non-W/Z leptons is estimated from data as described in sections 5.10 and 5.10.2 respectively, and statistical and systematic uncertainties are shown.
Systematics are also quoted in the backgrounds estimated from simulation (VV, Single top and Drell-Yan $\tau\tau$), which are assigned a 38% , 38% and 32% respectively, as presented in Table 5.52.

A summary of the $t\bar{t}$ signal production cross section measured in separate event selection channels is shown in Table 5.54. The number of events observed in data, the total background expectations, the total signal acceptance (includes geometric detector acceptance and event selection efficiency) with systematic uncertainties, and the cross section measurements are shown separately. The uncertainties on the cross section include statistical, systematic, and luminosity normalization uncertainties. Statistical and systematic uncertainties are computed according to the expressions 5.5 and 5.6 respectively.

$$\delta_{stats}(\sigma_{t\bar{t}}) = \sigma_{t\bar{t}} \cdot \frac{\sqrt{N}}{N - B} \tag{5.5}$$

Source	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$
Dilepton $t\bar{t}$	$13.3 \pm 1.2 \pm 2.1$	$17.2 \pm 1.5 \pm 2.7$	$48.5 \pm 4.0 \pm 7.5$
VV	0.09 ± 0.03	0.09 ± 0.03	0.2 ± 0.08
Single top - tW	0.5 ± 0.2	0.6 ± 0.2	1.6 ± 0.5
Drell-Yan $\tau\tau$	0.17 ± 0.06	0.2 ± 0.08	0.7 ± 0.3
Drell-Yan $ee, \mu\mu$	$0.66 \pm 0.6 \pm 0.3$	$2.6 \pm 1.2 \pm 1.3$	N/A
Non-W/Z leptons	$0.13 \pm 0.10 \pm 0.07$	$0.07 \pm 0.07 \pm 0.02$	$0.68 \pm 0.21 \pm 0.34$
Total backgrounds	1.55 ± 0.7	3.56 ± 1.8	3.2 ± 0.7
Data	15	24	51

Table 5.53: Expected signal and background contributions compared to the number of events observed in data passing full signal selection with at least one b-tagged jet required.

$$\delta_{sys}(\sigma_{t\bar{t}}) = \sigma_{t\bar{t}} \cdot \sqrt{(\frac{\delta B}{N - B})^2 + (\delta S)^2} \qquad (5.6)$$

Final state	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$
Events in data	15	24	51
All backgrounds	1.55 ± 0.7	3.56 ± 1.8	3.2 ± 0.7
Total acceptance, %	0.235 ± 0.021	0.304 ± 0.026	0.858 ± 0.071
Cross section, pb	$159.4 \pm 46 \pm 17 \pm 6$	$187.3 \pm 45 \pm 23 \pm 8$	$155.2 \pm 23 \pm 13 \pm 6$

Table 5.54: Summary of $t\bar{t}$ signal production cross section measured in separate event selection channels. The number of events observed in data, the total background expectations, the total signal acceptance (includes geometric detector acceptance and event selection efficiency) with systematic uncertainties, and the cross section measurements are shown separately. The uncertainties on the cross section include statistical, systematic, and luminosity normalization uncertainties.

Combining the 3 channels e^+e^-, $\mu^+\mu^-$ and $e^\pm\mu^\mp$ (see Table 5.55) the $t\bar{t}$ production cross section is found to be:

$$\sigma_{t\bar{t}} = 162.5 \pm 19(stats.) \pm 15(sys.) \pm 6(lumi.) \qquad (5.7)$$

which is in good agreement with the NLO predicted cross section. As expected, with 35.9 pb^{-1} this measurement is dominated by the statistical uncertainty.

In addition to the measurement described in this thesis, performed requiring at least one b-tagged jet in the analysis selection, the cross section has been also measured without any b-tagging requirement both in the at least 2 jets and in the 1 jet bins. Details on these analysis are documented in [81],[94].

All these measurements are combined assuming they correspond to the same physical quantity of the total $t\bar{t}$ production cross section. The combination is done using the Best Linear Unbiased Estimator (BLUE) technique [95], which accounts for correlations between different contributions to the measurements. This combination includes statistically correlated contributions from the events selected with at least two jets with and without a b-tagging requirement and the procedure followed

Final state	$e^+e^-+ \mu^+\mu^-+ e^{\pm}\mu^{\mp}$
Events in data	90
All backgrounds	8.31 ± 2.0
Total acceptance, %	1.40 ± 0.12
Cross section, pb	$162.5 \pm 19 \pm 15 \pm 6$

Table 5.55: Summary of ttbar signal production cross section measured in the 3 final states combined. The number of events observed in data, the total background expectations, the total signal acceptance (includes geometric detector acceptance and event selection efficiency) with systematic uncertainties, and the cross section measurements are shown separately. The uncertainties on the cross section include statistical, systematic, and luminosity normalization uncertainties.

is documented in [96].

The combined value of the cross section is found to be:

$$\sigma_{t\bar{t}} = 168 \pm 18(stats.) \pm 14(sys.) \pm 7(lumi) \tag{5.8}$$

This has been also combined with the measurement performed in the lepton+jets channel [97] and compared with theory [98]. Results are shown in Figures 5.39 and 5.40. Previous measurement done in dilepton channel with 3 pb^{-1}, as well as other measurements done in lepton+jets channels [99][97] are also displayed here for comparison, together with ATLAS combination [100].

Figure 5.39 shows a summary of the various existing $t\bar{t}$ production cross sections at 7 TeV LHC energy, and comparison with theory. The CMS combined cross section using 36 pb^{-1} is presented, as well as the various individual CMS measurements. For comparison, also the previous ATLAS and CMS results using 3 pb^{-1} are shown. All measurements are in good agreement with the theoretical prediction, within uncertainties. The experimental precision is already better than the theoretical uncertainty of the NLO calculation.

Figure 5.40 shows the $t\bar{t}$ cross section as a function of the center-of-mass energy $\sqrt{s}$ of the collision. It compares measurements made at the Tevatron $p\bar{p}$ collider with the LHC measurements. Again, good agreement between the CMS measurement and theory is observed.
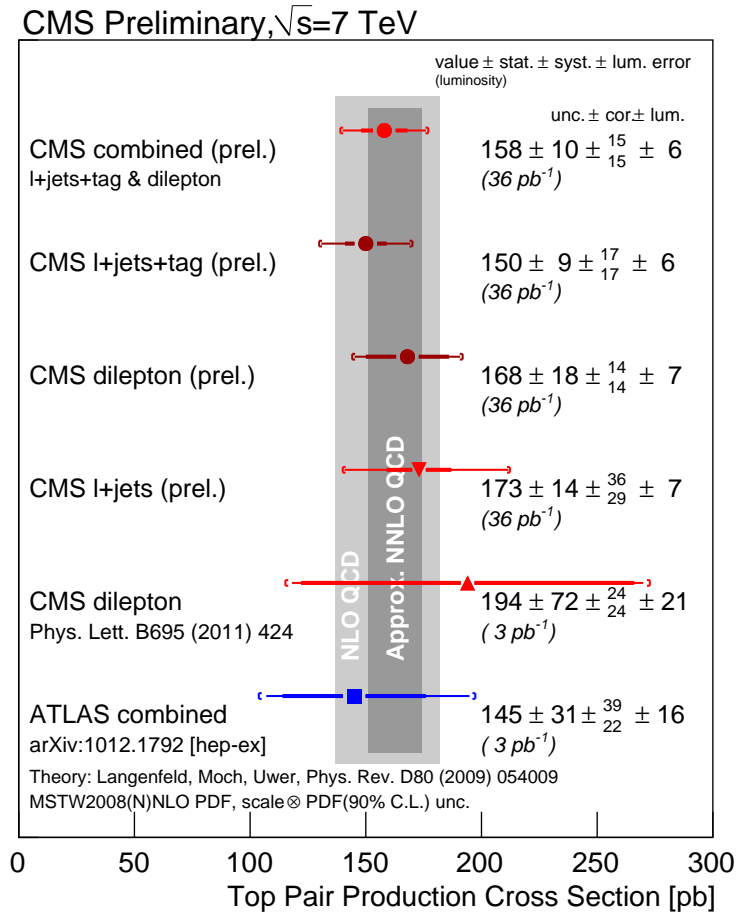
Figure 5.39: - Summary of various inclusive top pair production cross section measurements made in 7 TeV proton-proton collisions. The inner error bars of the data points correspond to the statistical uncertainty, while the outer (thinner) error bars correspond to the quadratic sum of statistical and systematic uncertainties. The outermost brackets correspond to the total error, including a luminosity uncertainty of 11% (4%) for the 3 (36) pb^{-1} results, respectively, which is also added in quadrature [98].
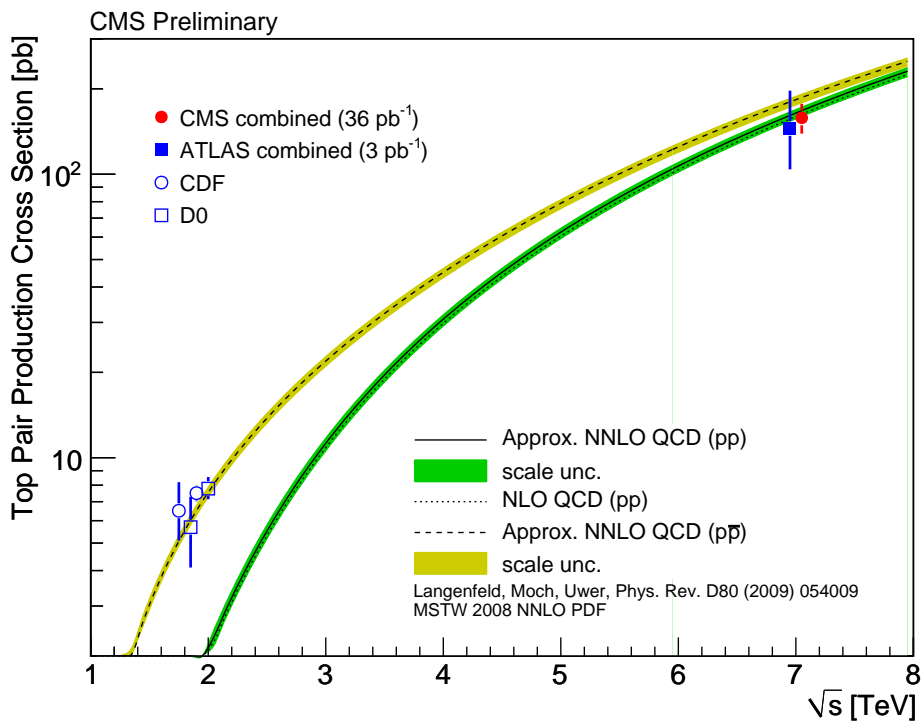
Figure 5.40: - Top pair production cross section as a function of $\sqrt{s}$, for both $p\bar{p}$ and pp collisions [98]. Tevatron measurements made at $\sqrt{s} = 1.8$ TeV are taken from [101; 102], while those made at $\sqrt{s} = 1.96$ TeV are taken from [103; 104]. The CMS combined measurement is shown, as well as the ATLAS measurement from [100]. Data points are slightly displaced horizontally for better visibility. Theory predictions at approximate NNLO are obtained using [105; 106]. The error band of the prediction corresponds to the scale uncertainty.

6

SM Higgs boson search in the $H \to WW^* \to 2l2\nu$ channel

6.1 Introduction

As described previously (Section 2.5), $H \to WW^*$ is the most sensitive channel for Higgs discovery in the mass region around 160 GeV/c^2, where the Higgs boson decays almost exclusively to 2 W's and presents a high cross section. Despite of that, this value is so tiny compared to the background cross sections, that it is primordial an excellent performance and understanding of the whole CMS detector, as well as the development and detailed optimization of specific analysis techniques to achieve sensitivity, specially in the lower mass regions.

All parts of the analysis have to be carefully studied and defined. The design of the triggers is important to efficiently collect events without loosing potential signal candidates. An adequate lepton reconstruction, identification and isolation is also essential, specially for the low mass range, when lower p_T leptons have to be reconstructed and W+jets background contamination arises. A good jet identification is important to reject multi-jet events as $t\bar{t}$ and the reduction of Drell Yan events relies on a proper $\not{E}_T$ reconstruction. These tools are the key for the success of the analysis.

On the other hand, the characteristics of the event change depending on the Higgs mass, as when it is below the $2 \times m_W$ threshold, one of the W bosons is off-shell, which is reflected in the kinematics, so the analysis has to be optimized accordingly also considering different mass hypothesis.

The search strategy will be explained in this chapter, together with the first results obtained using the full dataset collected during the 2010 run of the LHC.

6.2 Signal and background processes

The $H \to WW^* \to 2l2\nu$ signal presents a quite clean final state with 2 high p_T isolated leptons, a significant amount of $\not{E}_T$ and little hadronic activity. Due to the presence of the undetected neutrinos a mass peak can not be reconstructed in this channel so it is very important to have a good control and understanding of the backgrounds. Figure 6.1 shows the feynman diagram for this process.

The backgrounds considered in the search are all the possible processes that can yield similar signatures: real or fake multi-lepton final states plus $\not{E}_T$ .The Feynman diagrams of some of the
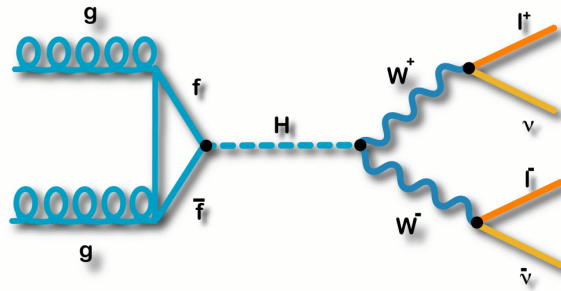
Figure 6.1: - Feynman diagram of the gluon-fusion induced Higgs to WW^* decaying to two leptons and two neutrinos.
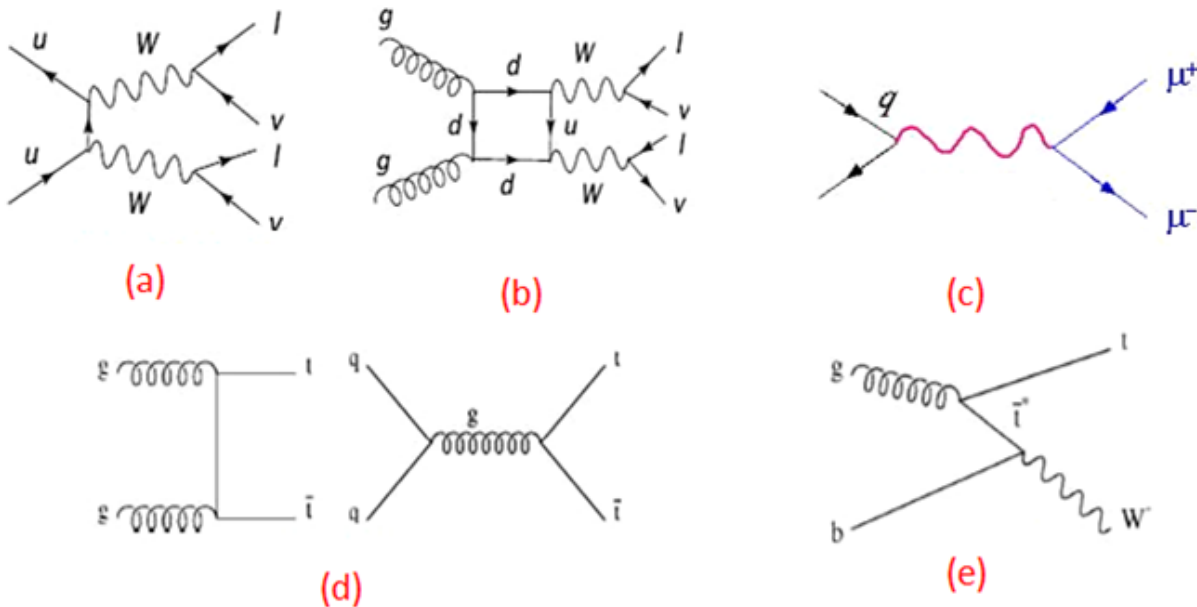
backgrounds are presented in Figure 6.2.

Figure 6.2: - Background processes:, qq (a) and gg (b) induced WW production, Drell-Yan (c), $t\bar{t}$ production (d) and single-top tW (e), others would be WZ, ZZ, $W + jets$ and QCD.

- The continuum WW production is almost identical to the signal and the only feature that can be used to distinguish them is the angle between the leptons in the transverse plane to the beam. For signal, as they come from the decay of a spin 0 particle, the angle between the leptons in the transverse plane takes small values, as it is presented in Figure 6.3. This does not happen for WW background, where the leptons tend to be emitted back-to-back. This property is the key of the Higgs to WW^* analysis.

- The large $t\bar{t}$ production at the LHC is another important background characterized by the hadronic activity. As energetic jets are present in the $t\bar{t}$ final state, this background can be reduced by rejecting events with jets, also the fact that they are coming from the hadronization
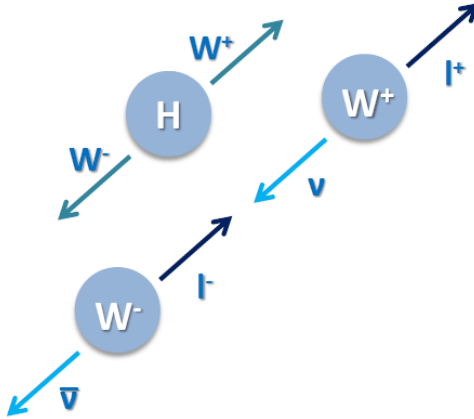
Figure 6.3: - Scheme of the leptonic decay of a Higgs to WW^* event.

of b-quarks can be used to further reduce it.

- $Z+jets$ process when the Z boson decays into two leptons,$Z/\gamma^* \to l^+l^-$, represents an important background as it has a huge cross section compared to the signal. The main distinctive property is that it has no genuine $\slashed{E}_T$ and this is the essential observable in which the analysis has to rely on to reduce this background, so it is a serious problem if this quantity is not properly estimated.

- $W + jets$ can be a source of background when a jet is misidentified as an isolated lepton and then two leptons are reconstructed in the final state. It has an even larger cross-section than $Z + jets$, and requires a detailed control of the fake leptons that can lead to di-leptonic final states. The way to reject this background is then, to unambiguously identify prompt leptons coming from W decays, for which an efficient identification and isolation is primordial.

- More backgrounds that are considered in the analysis are other diboson events as WZ and ZZ or single-top production (tW channel). These processes present lower cross-sections and different kinematics, like having more than two leptons in the final state. For example $WZ \to l\nu ll$ or $ZZ \to 4l$ decays can be reduced asking for only 2 good leptons in the event. The single top production is also controlled by the same tools used against $t\bar{t}$ process.

6.3 Searches performed with full 2010 luminosity

This section describes the first search on SM Higgs boson in the decay channel $WW^* \to 2l2\nu$ for Higgs boson masses between 120-600 GeV/c^2, performed with 36 $pb^{-1}$of data, the full luminosity collected during the 2010 LHC run.

The main analysis strategy is to select events with two opposite charged leptons, large missing energy and little jet activity. The two leptons are required to be isolated electrons or muons, of hight p_{T} and with a small opening angle in the transverse plane.

Part of the event selection and most of the data-driven techniques applied in the analysis are common to the ones used in the first WW cross section measurement [107], which was measured to

be $\sigma_{WW} = 41.1 \pm 15.3(stat.) \pm 5.8(syst.) \pm 4.5(lumi.)pb.$, consistent with the next to leading order predictions, which provides confidence on the robustness of the analysis tools.

6.4   Datasets and cross sections

The full list of data and simulated samples used in this analysis is summarized in Tables 6.1, 6.2 6.3 and 6.4.

With Pileup: Processed dataset name is always /Fall10-E7TeV_ProbDist_2010Data_BX156_START38_V12-v*/AODSIM OR Without Pileup: Processed dataset name is always /Fall10-START38_V12-v*/AODSIM	
Dataset Description	Dataset Name
Signal Selection Samples	
$gg \to H \to W^+W^- \to 2l2\nu$	/GluGluToHToWWTo2L2Nu_M-*_7TeV-powheg-pythia6/
$gg \to H \to W^+W^- \to l\nu\tau\nu$	/GluGluToHToWWTo2L2Nu_M-*_7TeV-powheg-pythia6/
$gg \to H \to W^+W^- \to 2\tau2\nu$	/GluGluToHToWWTo2L2Nu_M-*_7TeV-powheg-pythia6/

Table 6.1: Summary of signal datasets used

Mass (GeV/c^2)	$\sigma(gg \to H) \times BR(H \to WW^* \to 2l2\nu)(pb)$
120	0.25683
130	0.46511
140	0.66026
150	0.79267
160	0.89140
170	0.80552
180	0.67832
190	0.50050
200	0.42007

Table 6.2: Signal cross section for different masses [108]

Dataset Description	Dataset Name
Selection Samples	
Run2010A Muon Nov4 ReReco	/Mu/Run2010A-Nov4ReReco_v1/AOD
Run2010A Electron Nov4 ReReco	/EG/Run2010A-Nov4ReReco_v1/AOD
Run2010B Muon Nov4 ReReco	/Mu/Run2010B-Nov4ReReco-v1/AOD
Run2010B Electron Nov4 ReReco	/Electron/Run2010B-Nov4ReReco-v1/AOD
Fake Rate Measurement Samples	
Run2010A JetMETTau	/JetMETTau/Run2010A-Nov4ReReco_v1/AOD
Run2010A JetMET	/JetMET/Run2010A-Nov4ReReco_v1/AOD
Run2010B Jet	/Jet/Run2010B-Nov4ReReco-v1/AOD
Run2010B Electron PromptReco	/Electron/Run2010B-PromptReco-v2
Run2010A EG Sep17 ReReco	/EG/Run2010A-Sep17ReReco_v2

Table 6.3: Summary of data datasets used

With Pileup: Processed dataset name is always /Fall10-E7TeV_ProbDist_2010Data_BX156_START38_V12-v*/AODSIM OR Without Pileup: Processed dataset name is always /Fall10-START38_V12-v*/AODSIM		
Dataset Description	Primary Dataset Name	cross-section (pb)
$t\bar{t}$	/TTJets_TuneZ2_7TeV-madgraph-tauola/	157.5
$q\bar{q} \rightarrow$ WW	/VVJetsTo4L_TuneD6T_7TeV-madgraph-tauola/	43.0
$gg \rightarrow$ WW	/GluGluToWWTo4L_TuneZ2_7TeV-gg2ww-pythia6/	0.153
WZ	/WZToAnything_TuneZ2_7TeV-pythia6-tauola/	18.2
ZZ	/ZZtoAnything_TuneZ2_7TeV-pythia6-tauola/	5.9
Single top- tW channel	/TToBLNu_TuneZ2_tW-channel_7TeV-madgraph/	10.6
W $\rightarrow$ eν	/WToENu_TuneZ2_7TeV-pythia6/	10438
W $\rightarrow$ $\mu\nu$	/WToMuNu_TuneZ2_7TeV-pythia6/	10438
W $\rightarrow$ $\tau\nu$	/WToTauNu_TuneZ2_7TeV-pythia6-tauola/	10438
$Z/\gamma^* \rightarrow e^+e^-$ m_{ll} ϵ [10-20]	/DYToEE_M-10To20_CT10_TuneZ2_7TeV-powheg-pythia/	3892.9
$Z/\gamma^* \rightarrow \mu^+\mu^-$ m_{ll} ϵ [10-20]	/DYToMuMu_M-10To20_CT10_TuneZ2_7TeV-powheg-pythia/	3892.9
$Z/\gamma^* \rightarrow \tau^+\tau^-$ m_{ll} ϵ [10-20]	/DYToTauTau_M-10To20_CT10_TuneZ2_7TeV-powheg-pythia-tauola/	3892.9
$Z/\gamma^* \rightarrow e^+e^-$ m_{ll} ϵ [20-inf]	/DYToEE_M-20_TuneZ2_7TeV-powheg-pythia/	1666.0
$Z/\gamma^* \rightarrow \mu^+\mu^-$ m_{ll} ϵ [20-inf]	/DYToMuMu_M-20_TuneZ2_7TeV-powheg-pythia/	1666.0
$Z/\gamma^* \rightarrow \tau^+\tau^-$ m_{ll} ϵ [20-inf]	/DYToTauTau_M-20_TuneZ2_7TeV-powheg-pythia-tauola/	1666.0

Table 6.4: Summary of Monte Carlo datasets used

The same JSON files, overlapping events removal and event cleaning described in Section 5.6 are applied to these datasets.

6.5 Selection

This section describes the main physics observables and selections used in the analysis to extract the signal and reduce as much as possible the backgrounds. Although they are not too different to what has been described for the $t\bar{t}$ cross section measurement, the selection has to be optimized taken into account the characteristics of the Higgs signal. A detailed description of the lepton (muon and electron), jet and $\not{E}_T$ selection is presented in the following subsections.

6.5.1 Trigger

The trigger requirement is the same as for the $t\bar{t}$ analysis and the set of single muon and single electron paths already described in Section 5.7.1 is used here.

6.5.2 Muon selection

Muon candidates are required to be successfully reconstructed in both tracker and muon systems being globalMuon and trackerMuon. Once the muons are reconstructed, the following identification criteria is applied in order to select prompt muons from W decays:

- Identification as GlobalMuonPromptTight

- The tracker muon must have at least two matches to muon segments in different muon stations.

- $p_T > 20\ GeV/c$ and $|\eta| < 2.4$. The relative p_T resolution should be better than 10% .

- The track associated with the muon candidate is required to have a minimum number of hits in the silicon tracker and to have a high-quality global fit including a minimum number of hits in the muon detector: Number of valid hits in the inner tracker > 10, at least one pixel hit and $\chi^2/ndof < 10$ for the global muon fit.

- The distance of the closest approach to the beam line should be below 0.02 cm in direction transverse to the beam line in order to have a muon candidate track consistent with originating from the beam spot. In addition, the selected leptons have to be close to the good primary vertex: the difference between the muon track at its vertex and the PV along the Z position should be less than 1 cm.

Finally, the muons have to be isolated in order to reduce the contamination from fake muons originating from jets using the variable presented in Section 4.1.3, which is required to be $Iso_\mu < 0.15$.

6.5.3 Electron selection

Electron candidates are reconstructed by combining electromagnetic calorimeter towers and track seeds reconstructed from pixel hits and are required to fulfill the following criteria:

- Electron candidates are required to have $p_\mathrm{T} > 20 \; GeV/c$ and $|\eta| < 2.5$;

- The electron candidate track is required to be consistent with originating from the beam spot: the distance of the closest approach to the beam line should be below 0.04 cm in direction transverse to the beam line; the distance between the point of the closest approach to the beam line and a primary vertex should be less than 1 cm along the z-coordinate direction, as also used for muons.

- The electron identification algorithm is based on cuts applied on discriminating variables between real and fake electrons, as described in Section 4.2.2 and the WP80 is the working point chosen.

- Electron candidates consistent with photon conversions are rejected as described in Section 4.2.2, based on either reconstruction of a conversion partner in the silicon tracker, or based on consistent presence of hits in the pixel tracker along the electron candidate trajectory. The number of lost hits in the tracker should be 0, the minimal distance between the electron and its closest opposite sign track should be $|\Delta \cos\theta| > 0.02$ and $Dist > 0.02$ in the $r - \phi$ plane

- Electron candidates within $dR = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.1$ of a tracker-based or globally-fitted muon are rejected to remove fake electron candidates due to muon inner bremsstrahlung (prompt collinear final state radiation).

- Electron is required to be isolated by imposing a requirement on the combined relative isolation as presented in Section 4.2.3. This variable is required to be $Iso_e < 0.10$.

6.5.4 Jet selection and Jet veto

The anti-k_T algorithm [67] with R = 0.5 is used for jet clustering. Jets are reconstructed based on the calorimeter, tracker, and muon system information combined in a particle level view of the event using the particle flow reconstruction. Further corrections are applied to the raw jet momenta to establish a relative uniform response of the calorimeter in jet η and an absolute uniform response in jet p_T .

Reconstructed jet candidates have to fulfill the following requirements:

- $p_T > 25$ GeV/c, $|\eta| < 5.0$

- Jet lepton cleaning: exclusion of jets overlapping with the 2 identified and isolated lepton candidates (electron or muon) used in the analysis if $\Delta R(jet, lepton) < 0.3$.

Once jets are reconstructed, events having at least 1 jet passing the above selection criteria are vetoed, in order to suppress the $t\bar{t}$ background production, which is 2 orders of magnitude higher than the SM Higgs boson cross section. The distribution of the number of jets in the event is shown in Figure 6.4.

The ability of CMS to properly reconstruct and identify these jets will determine the amount of $t\bar{t}$ background reduction achieved. As the main difference with the signal is the presence of 2 high p_T jets coming from b-quarks, it becomes a dangerous background when these two jets are lost, either because their energy is behind the p_T threshold or just because they lay outside the acceptance.
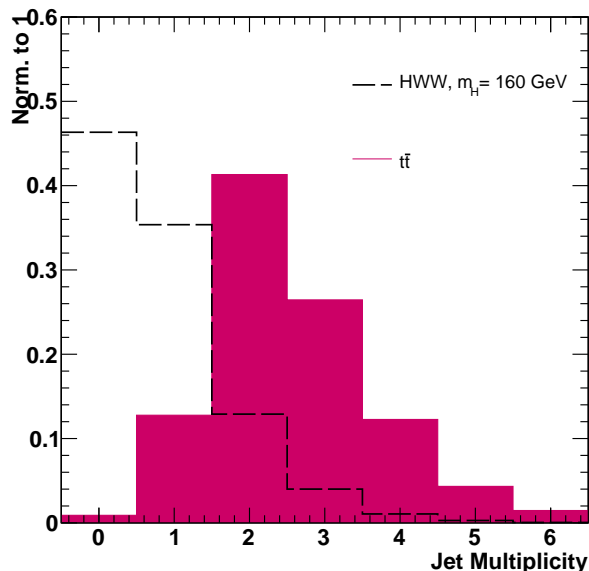

Figure 6.4: - Distribution of the number of jets for the $H \rightarrow WW^*$ signal (m_H=160 GeV/c^2) and $t\bar{t}$ background

6.5.5 Top tagging

The jet veto already rejects a large fraction of $t\bar{t}$ background, but the fact that jets in $t\bar{t}$ events are coming from the hadronization of b-quarks can be exploited to further reduce this background. This is done requiring that there are no b-tagged jets in those events remaining after the jet veto cut. A jet is considered b-tagged if it has a value of the TCHE tagger greater than 2.1, and this requirement is applied to any calojet.

In addition to the b-tagging part, also a soft muon tagging is applied to reduce $t\bar{t}$ events. This is useful to get rid of $t\bar{t}$ events in which one of the two b-quarks decays into a muon. In this case, as it is not isolated this muon will not be selected in the first step of the analysis, and the event can go further in the analysis chain. A soft-muon veto can be applied to suppress these events The soft muon selection requirements are:

- $p_{\mathrm{T}} > 3$ GeV/c

- is a TrackerMuon;

- passed TMLastStationAngTight muon id requirements;

- number of valid inner tracker hits is more than 10;

- impact parameter in the transverse plane $|d0| < 2mm$, calculated with respect to the primary vertex;

- muons with $p_{\mathrm{T}} > 20$ GeV/c have to be non-isolated with $Iso_\mu > 0.1$

6.5.6 $\not{E}_T$ selection

The missing transverse energy is used to reject background events where there is no natural source of missing energy like Drell-Yan and QCD. This is the main observable to reduce the Drell-Yan background and as it is one of the variables most affected by the performance of the detector, its control is very important for the analysis.

As discussed in Section 4.4 , there are several methods to compute the $\not{E}_T$ in CMS and the different estimators were studied during 2010 and previously based on simulation.

The Track Corrected $\not{E}_T$ is the one used in the analysis, where the precise measurement of charged particles in the tracker system is used to correct the calorimeter-based measurements.

To reject $Z/\gamma^* \to l^+l^-$ background events with the small opening angle between $\not{E}_T$ and one of the leptons, a variable called projected $\not{E}_T$ is defined as:

$$projected\ \not{E}_T = \begin{cases} \not{E}_T & \text{if } \Delta\phi_{min} \geq \frac{\pi}{2} \\[2mm] \not{E}_T \cdot \sin(\Delta\phi_{min}) & \text{if } \Delta\phi_{min} < \frac{\pi}{2} \end{cases}$$

where $\Delta\phi_{min} = \min(\Delta\phi(l_1, \not{E}_T), \Delta\phi(l_2, \not{E}_T))$.

Basically due large difference in tau and Z mass, taus are produced with large boost and their decay products, including neutrinos, are aligned with the leptons. In such case a transverse component of missing energy with respect to the leptons is a better measure of true missing energy in the event, not originating from tau decay.

Figures 6.5 show the $\not{E}_T$ and projected $\not{E}_T$ distributions for $H \to WW^*$ signal and $Z/\gamma^* \to l^+l^-$ events.

6.6 Event selection and analysis strategies

Events having 2 opposite sign isolated leptons passing the criteria described before are classified in e^+e^-, $\mu^+\mu^-$ or $e^\pm\mu^\mp$ according to the lepton pair with opposite charge which maximizes the sum of the transverse momentum. In addition, the dilepton invariant mass is required to be above 12 GeV/c, this is applied to all modes to reject backgrounds from low mass resonances.
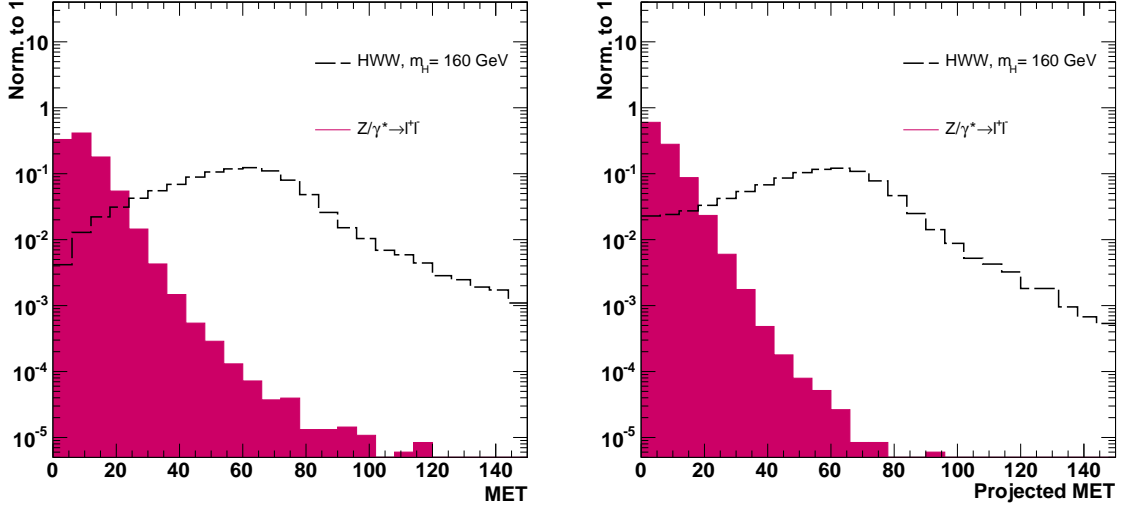
Figure 6.5: - Distribution of the $\not{E}_T$ (left) and Projected $\not{E}_T$ (right) for the $H \to WW^*$ signal and $Z/\gamma^* \to l^+l^-$ events.

Then further cuts are applied to reduce the dominant $Z/\gamma^* \to l^+l^-$ and $t\bar{t}$ backgrounds and to obtain a WW enriched sample: the two leptons are required not to be consistent with a Z decay, $t\bar{t}$ and single top productions are reduced by vetoing events which contain a high p_T jet and a significant amount of $\not{E}_T$ is required. In summary:

- A cut on $m_{ll} > 12$ is applied to reject the low mass resonances

- Dileptonic invariant mass $\notin [76; 106]$ GeV for the e^+e^- and $\mu^+\mu^-$ channels.

- Projected $\not{E}_T > 35$ GeV for e^+e^- and $\mu^+\mu^-$ channels and projected $\not{E}_T > 20$ GeV in the $e^\pm\mu^\mp$ channel.

- Jet Veto

- Top tagging

- Soft muon and extra lepton veto. Events are rejected if they have an additional fully identified lepton, which is useful to further reject the background from diboson production.

Then the selection is optimized for the Higgs search, following two different approaches: a standard sequential analysis, based on cuts on the main discriminant variables and a multivariate analysis using boosted decision trees, both optimized according to different Higgs boson masses hypothesis. The details of both analysis will be explained in the following sections.

141

6.7 Cut-based analysis

Once the major background contribution from $t\bar{t}$ and $Z/\gamma^* \to l^+l^-$ has been removed as explained before, the remaining background specially WW are reduced as much as possible by applying a sequential set of cuts on the kinematical variables that have been found to have a good discriminant power against the different sources of background.

- The angle $\Delta\phi_{ll}$ between the two selected leptons in the transverse plane. As explained before, this is the main discriminant variable against the irreducible WW background. Due to spin correlations, this angle tends to be small for the signal, while leptons from WW decays tend to be preferentially emitted back-to back and therefore with large $\Delta\phi_{ll}$ values. This is because the Higgs boson is a scalar particle, so the Ws from Higgs decays are produced with their polarizations opposite each other.

  The importance of this variable disappears for high masses, since the high boost of the boson makes the angular correlation to be lost. Due to the scalar nature of the higgs boson and of the v-a structure of the weak interaction, for higgs masses close to 2MW, the WW spin correlations play in favour of smaller opening angles between the 2 leptons.

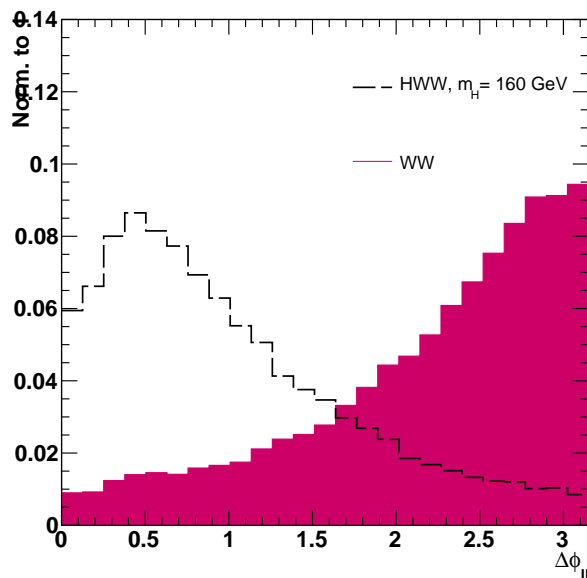 The distribution of $\Delta\phi_{ll}$ is represented in Figure 6.6 for the signal and WW background.

Figure 6.6: - $\Delta\phi_{ll}$ distribution for $H \to WW^*$ (m_H= 160 GeV/c^2) and WW

- The dilepton invariant mass. An upper cut on the dilepton invariant mass distribution is applied to reduce the remaining contamination of leptons coming from Z decays. Other backgrounds with high m_{ll} values are reduced as well by the application of this cut. The distribution of m_{ll} for Higgs signal and $Z/\gamma^* \to l^+l^-$ background is presented in Figure 6.7.

- The $p_{\rm T}$ of the leptons. A lower cut on the transverse momenta of the harder (p_{Tmax}^{lep}) and softer (p_{Tmin}^{lep}) lepton is applied as is useful to further reject other backgrounds. The $p_{\rm T}$ of leptons will be determined by the Higgs boson mass hypothesis considered, which can be used to further
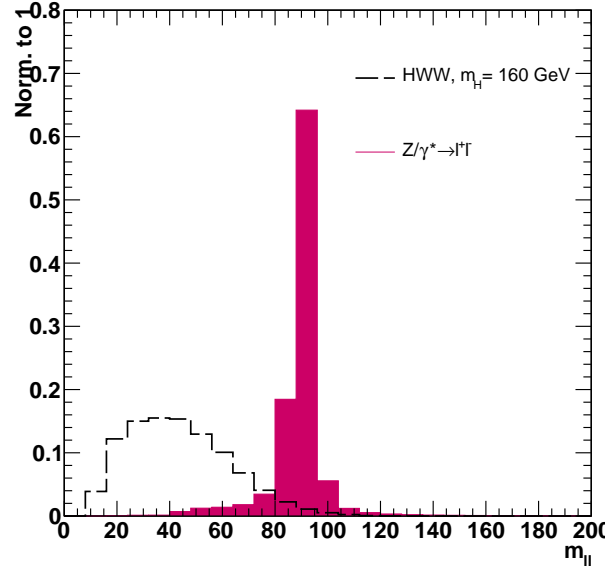
Figure 6.7: - Dilepton invariant mass distribution for $H \to WW^*$ ($m_H = 160~GeV/c^2$) and $Z/\gamma^* \to l^+l^-$

reduce the background. An example is illustrated in Figure 6.8 for the signal with 160 GeV/c^2 and diboson background.

The optimization of the analysis is performed independently for each Higgs boson mass hypothesis. The optimal set of cuts is determined by maximizing the statistical significance, defined as:

$$S = \frac{N_S}{\sqrt{N_B + (\Delta N_B)^2}} \tag{6.1}$$

where N_S and N_B represent the number of Higgs signal and total background events for $36pb^{-1}$ and assuming the SM Higgs boson cross section. The relative error on the background estimation is assumed to be 35% .

The full list of cuts as a function of Higgs mass is presented in Table 6.5.

Event yields and MC expectations are summarized in Tables 6.6 to 6.17, for $\mu^+\mu^-$,e^+e^- $e^\pm\mu^\mp$and all channels combined and different mass hypothesis: $m_H = 130$, 160 and 190 GeV/c^2.
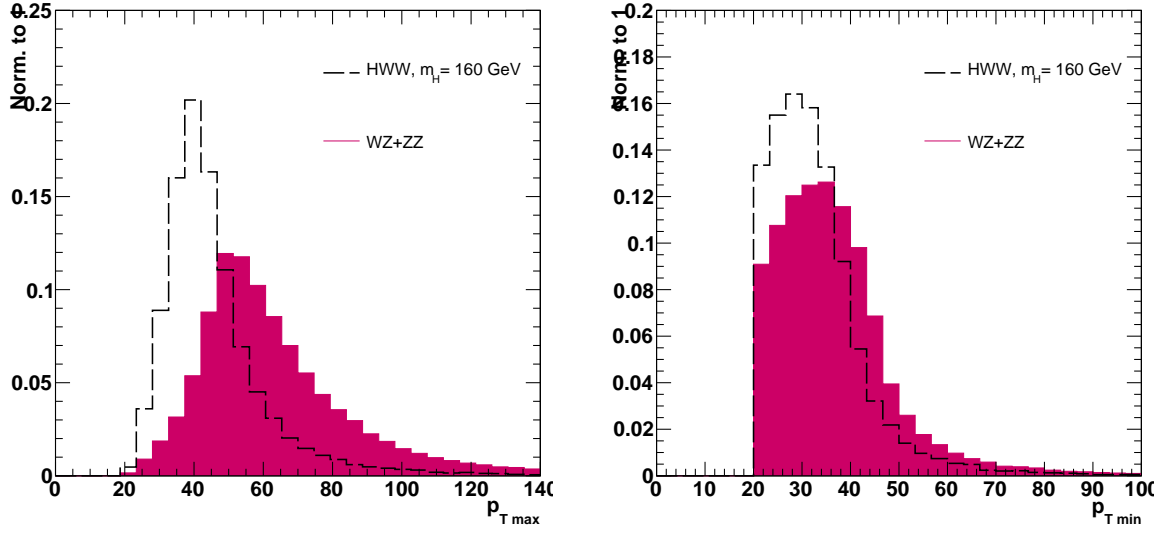
143

Figure 6.8: - Distribution of the harder (left) and softer(right) lepton for $H \rightarrow WW^*$ (m_H= 160 GeV/c^2) and diboson events (WZ+ZZ)

m_H[GeV]	p_{Tmax}^{lep} [GeV] $>$	p_{Tmin}^{lep} [GeV] $>$	m_{ll} [GeV] $<$	$\Delta\phi_{ll}$ [dg.] $<$
120	20	20	40	60
130	25	20	45	60
140	25	20	45	60
150	27	25	50	60
160	30	25	50	60
170	34	25	50	60
180	36	25	60	70
190	38	25	80	90
200	40	25	90	100
210	44	25	110	110
220	48	25	120	120
230	52	25	130	130
250	55	25	150	140
300	70	25	200	175
350	80	25	250	175
400	90	25	300	175
450	110	25	350	175
500	120	25	400	175
550	130	25	450	175
600	140	25	500	175

Table 6.5: Optimized selection cuts for different Higgs boson mass hypothesis

	HWW130	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	0.529 ±0.006	41.4 ±0.4	15152 ±21	105.9 ±0.6	11.07 ±0.05	10.58 ±0.06	6.90 ±0.03	2.95 ±0.05	15331 ±21	14834
Z veto	0.514 ±0.006	31.9 ±0.4	1448 ±7	62.6 ±0.5	7.63 ±0.04	0.89 ±0.02	0.691 ±0.008	2.31 ±0.04	1554 ±7	1437
projectedMET	0.322 ±0.005	19.2 ±0.3	2.2 ±0.3	0.2 ±0.1	4.08 ±0.03	0.221 ±0.008	0.110 ±0.003	1.41 ±0.03	27.4 ±0.4	32
CJV	0.165 ±0.003	0.18 ±0.03	0.12 ±0.06	0.011 ±0.005	2.50 ±0.02	0.101 ±0.006	0.059 ±0.002	0.084 ±0.008	3.06 ±0.07	1
Softmuon	0.164 ±0.003	0.17 ±0.03	0.12 ±0.06	0.011 ±0.005	2.49 ±0.02	0.074 ±0.005	0.058 ±0.002	0.071 ±0.007	3.00 ±0.07	1
Extralepton veto	0.164 ±0.003	0.17 ±0.03	0.12 ±0.06	0.011 ±0.005	2.49 ±0.02	0.070 ±0.005	0.057 ±0.002	0.071 ±0.007	2.99 ±0.07	1
Top veto	0.163 ±0.003	0.10 ±0.02	0.12 ±0.06	0.011 ±0.005	2.48 ±0.02	0.070 ±0.005	0.057 ±0.002	0.058 ±0.007	2.89 ±0.07	1
m_{ll}	0.115 ±0.003	0.005 ±0.005	0.06 ±0.04	0.003 ±0.003	0.60 ±0.01	0.012 ±0.002	0.014 ±0.001	0.008 ±0.002	0.70 ±0.04	0
p_{Tmax}	0.108 ±0.003	0.005 ±0.005	0.06 ±0.04	0.003 ±0.003	0.57 ±0.01	0.011 ±0.002	0.013 ±0.001	0.007 ±0.002	0.67 ±0.04	0
p_{Tmin}	0.108 ±0.003	0.005 ±0.005	0.06 ±0.04	0.003 ±0.003	0.57 ±0.01	0.011 ±0.002	0.013 ±0.001	0.007 ±0.002	0.67 ±0.04	0
$\Delta\phi$	0.077 ±0.002	0.005 ±0.005	0.06 ±0.04	0.003 ±0.003	0.406 ±0.009	0.009 ±0.002	0.011 ±0.001	0.006 ±0.002	0.50 ±0.04	0

Table 6.6: Expected number of events for 36.1 pb^{-1} in the $\mu\mu$ channel for the m_H= 130 GeV/c^2 selection

	HWW130	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	0.340 ±0.005	28.5 ±0.4	9952 ±17	66.9 ±0.6	7.25 ±0.04	7.24 ±0.05	4.65 ±0.02	2.10 ±0.04	10069 ±17	9227
Z veto	0.331 ±0.005	22.2 ±0.3	872 ±5	40.6 ±0.5	4.95 ±0.03	0.65 ±0.01	0.443 ±0.007	1.66 ±0.04	942 ±5	932
projectedMET	0.203 ±0.004	13.2 ±0.3	1.2 ±0.2	0.8 ±0.2	2.57 ±0.02	0.159 ±0.007	0.066 ±0.003	0.97 ±0.03	18.9 ±0.4	16
CJV	0.101 ±0.003	0.12 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.046 ±0.004	0.034 ±0.002	0.057 ±0.007	2.2 ±0.2	1
Softmuon	0.101 ±0.003	0.10 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.043 ±0.004	0.033 ±0.002	0.051 ±0.006	2.2 ±0.2	1
Extralepton veto	0.101 ±0.003	0.10 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.034 ±0.003	0.033 ±0.002	0.051 ±0.006	2.2 ±0.2	1
Top veto	0.100 ±0.003	0.06 ±0.02	0.06 ±0.04	0.3 ±0.1	1.54 ±0.02	0.034 ±0.003	0.033 ±0.002	0.039 ±0.005	2.1 ±0.2	1
m_{ll}	0.071 ±0.002	0.005 ±0.005	0 ±0.00000	0.2 ±0.1	0.354 ±0.009	0.006 ±0.001	0.0083 ±0.0009	0.007 ±0.002	0.6 ±0.1	0
p_{Tmax}	0.067 ±0.002	0.005 ±0.005	0 ±0.00000	0.2 ±0.1	0.340 ±0.009	0.006 ±0.001	0.0079 ±0.0009	0.007 ±0.002	0.6 ±0.1	0
p_{Tmin}	0.067 ±0.002	0.005 ±0.005	0 ±0.00000	0.2 ±0.1	0.340 ±0.009	0.006 ±0.001	0.0079 ±0.0009	0.007 ±0.002	0.6 ±0.1	0
$\Delta\phi$	0.048 ±0.002	0.005 ±0.005	0 ±0.00000	0.003 ±0.003	0.251 ±0.007	0.005 ±0.001	0.0068 ±0.0008	0.006 ±0.002	0.276 ±0.010	0

Table 6.7: Expected number of events for 36.1 pb^{-1} in the ee channel for the m_H= 130 GeV/c^2 selection

	HWW130	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	0.856 ±0.008	68.3 ±0.6	61 ±1	3.2 ±0.4	15.88 ±0.06	0.77 ±0.02	0.110 ±0.003	4.96 ±0.06	154 ±2	134
Z veto	0.856 ±0.008	68.3 ±0.6	61 ±1	3.2 ±0.4	15.88 ±0.06	0.77 ±0.02	0.110 ±0.003	4.96 ±0.06	154 ±2	134
projectedMET	0.708 ±0.007	52.0 ±0.5	1.4 ±0.2	1.9 ±0.3	11.43 ±0.05	0.52 ±0.01	0.033 ±0.002	3.85 ±0.05	71.1 ±0.6	64
CJV	0.338 ±0.005	0.50 ±0.05	0.12 ±0.06	1.0 ±0.2	7.10 ±0.04	0.183 ±0.008	0.0084 ±0.0009	0.25 ±0.01	9.1 ±0.3	9
Softmuon	0.337 ±0.005	0.44 ±0.05	0.12 ±0.06	1.0 ±0.2	7.07 ±0.04	0.158 ±0.007	0.0073 ±0.0009	0.22 ±0.01	9.0 ±0.3	9
Extralepton veto	0.337 ±0.005	0.44 ±0.05	0.12 ±0.06	1.0 ±0.2	7.07 ±0.04	0.135 ±0.007	0.0068 ±0.0008	0.22 ±0.01	9.0 ±0.3	9
Top veto	0.334 ±0.005	0.29 ±0.04	0.12 ±0.06	1.0 ±0.2	7.03 ±0.04	0.135 ±0.007	0.0067 ±0.0008	0.18 ±0.01	8.8 ±0.3	9
m_{ll}	0.202 ±0.004	0.04 ±0.01	0.03 ±0.03	0.2 ±0.1	0.97 ±0.01	0.020 ±0.003	0.0014 ±0.0004	0.028 ±0.005	1.3 ±0.1	0
p_{Tmax}	0.187 ±0.004	0.04 ±0.01	0.03 ±0.03	0.2 ±0.1	0.91 ±0.01	0.019 ±0.002	0.0012 ±0.0003	0.028 ±0.005	1.2 ±0.1	0
p_{Tmin}	0.187 ±0.004	0.04 ±0.01	0.03 ±0.03	0.2 ±0.1	0.91 ±0.01	0.019 ±0.002	0.0012 ±0.0003	0.028 ±0.005	1.2 ±0.1	0
$\Delta\phi$	0.129 ±0.003	0.03 ±0.01	0.03 ±0.03	0.2 ±0.1	0.61 ±0.01	0.013 ±0.002	0.0009 ±0.0003	0.026 ±0.004	0.9 ±0.1	0

Table 6.8: Expected number of events for 36.1 pb^{-1} in the $e\mu$ channel for the m_H= 130 GeV/c^2 selection

	HWW130	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	1.73 ±0.01	138.2 ±0.8	25165 ±28	176.1 ±0.9	34.20 ±0.09	18.59 ±0.08	11.66 ±0.03	10.00 ±0.09	25554 ±28	24195
Z veto	1.70 ±0.01	122.4 ±0.8	2381 ±9	106.4 ±0.8	28.46 ±0.08	2.31 ±0.03	1.24 ±0.01	8.92 ±0.08	2651 ±9	2503
projectedMET	1.232 ±0.009	84.5 ±0.6	4.8 ±0.4	2.9 ±0.4	18.07 ±0.06	0.90 ±0.02	0.208 ±0.005	6.22 ±0.07	117.5 ±0.9	112
CJV	0.604 ±0.006	0.80 ±0.06	0.30 ±0.10	1.3 ±0.3	11.15 ±0.05	0.33 ±0.01	0.102 ±0.003	0.39 ±0.02	14.4 ±0.3	11
Softmuon	0.602 ±0.006	0.70 ±0.06	0.30 ±0.10	1.3 ±0.3	11.11 ±0.05	0.275 ±0.009	0.098 ±0.003	0.34 ±0.02	14.2 ±0.3	11
Extralepton veto	0.602 ±0.006	0.70 ±0.06	0.30 ±0.10	1.3 ±0.3	11.11 ±0.05	0.240 ±0.009	0.097 ±0.003	0.34 ±0.02	14.1 ±0.3	11
Top veto	0.598 ±0.006	0.45 ±0.05	0.30 ±0.10	1.3 ±0.3	11.04 ±0.05	0.238 ±0.009	0.096 ±0.003	0.27 ±0.01	13.7 ±0.3	11
m_{ll}	0.388 ±0.005	0.05 ±0.02	0.09 ±0.05	0.4 ±0.2	1.92 ±0.02	0.038 ±0.003	0.024 ±0.002	0.043 ±0.006	2.6 ±0.2	0
p_{Tmax}	0.362 ±0.005	0.05 ±0.02	0.09 ±0.05	0.4 ±0.2	1.82 ±0.02	0.035 ±0.003	0.022 ±0.002	0.042 ±0.006	2.5 ±0.2	0
p_{Tmin}	0.362 ±0.005	0.05 ±0.02	0.09 ±0.05	0.4 ±0.2	1.82 ±0.02	0.035 ±0.003	0.022 ±0.002	0.042 ±0.006	2.5 ±0.2	0
$\Delta\phi$	0.254 ±0.004	0.04 ±0.01	0.09 ±0.05	0.2 ±0.1	1.27 ±0.02	0.027 ±0.003	0.018 ±0.001	0.038 ±0.005	1.7 ±0.1	0

Table 6.9: Expected number of events for 36.1 pb^{-1} in the $\mu\mu + ee + e\mu$ channel for the m_H= 130 GeV/c^2 selection

	HWW160	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	1.89 ±0.02	41.4 ±0.4	15152 ±21	105.9 ±0.6	11.07 ±0.05	10.58 ±0.06	6.90 ±0.03	2.95 ±0.05	15331 ±21	14834
Z veto	1.79 ±0.02	31.9 ±0.4	1448 ±7	62.6 ±0.5	7.63 ±0.04	0.89 ±0.02	0.691 ±0.008	2.31 ±0.04	1554 ±7	1437
projectedMET	1.46 ±0.01	19.2 ±0.3	2.2 ±0.3	0.2 ±0.1	4.08 ±0.03	0.221 ±0.008	0.110 ±0.003	1.41 ±0.03	27.4 ±0.4	32
CJV	0.722 ±0.010	0.18 ±0.03	0.12 ±0.06	0.011 ±0.005	2.50 ±0.02	0.101 ±0.006	0.059 ±0.002	0.084 ±0.008	3.06 ±0.07	1
Softmuon	0.719 ±0.010	0.17 ±0.03	0.12 ±0.06	0.011 ±0.005	2.49 ±0.02	0.074 ±0.005	0.058 ±0.002	0.071 ±0.007	3.00 ±0.07	1
Extralepton veto	0.719 ±0.010	0.17 ±0.03	0.12 ±0.06	0.011 ±0.005	2.49 ±0.02	0.070 ±0.005	0.057 ±0.002	0.071 ±0.007	2.99 ±0.07	1
Top veto	0.713 ±0.010	0.10 ±0.02	0.12 ±0.06	0.011 ±0.005	2.48 ±0.02	0.070 ±0.005	0.057 ±0.002	0.058 ±0.007	2.89 ±0.07	1
m_{ll}	0.537 ±0.008	0.015 ±0.008	0.06 ±0.04	0.003 ±0.003	0.74 ±0.01	0.017 ±0.002	0.017 ±0.001	0.012 ±0.003	0.86 ±0.05	0
p_{Tmax}	0.491 ±0.008	0.015 ±0.008	0.03 ±0.03	0	0.60 ±0.01	0.013 ±0.002	0.012 ±0.001	0.009 ±0.003	0.68 ±0.03	0
p_{Tmin}	0.395 ±0.007	0.015 ±0.008	0.03 ±0.03	0	0.349 ±0.009	0.008 ±0.002	0.0085 ±0.0009	0.006 ±0.002	0.42 ±0.03	0
$\Delta\phi$	0.315 ±0.006	0.015 ±0.008	0.03 ±0.03	0	0.229 ±0.007	0.007 ±0.001	0.0070 ±0.0008	0.005 ±0.002	0.29 ±0.03	0

Table 6.10: Expected number of events for 36.1 pb^{-1} in the $\mu\mu$ channel for the $m_H = 160 \ GeV/c^2$ selection

	HWW160	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	1.24 ±0.01	28.5 ±0.4	9952 ±17	66.9 ±0.6	7.25 ±0.04	7.24 ±0.05	4.65 ±0.02	2.10 ±0.04	10069 ±17	9227
Z veto	1.18 ±0.01	22.2 ±0.3	872 ±5	40.6 ±0.5	4.95 ±0.03	0.65 ±0.01	0.443 ±0.007	1.66 ±0.04	942 ±5	932
projectedMET	0.96 ±0.01	13.2 ±0.3	1.2 ±0.2	0.8 ±0.2	2.57 ±0.02	0.159 ±0.007	0.066 ±0.003	0.97 ±0.03	18.9 ±0.4	16
CJV	0.467 ±0.008	0.12 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.046 ±0.004	0.034 ±0.002	0.057 ±0.007	2.2 ±0.2	1
Softmuon	0.465 ±0.008	0.10 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.043 ±0.004	0.033 ±0.002	0.051 ±0.006	2.2 ±0.2	1
Extralepton veto	0.465 ±0.008	0.10 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.034 ±0.003	0.033 ±0.002	0.051 ±0.006	2.2 ±0.2	1
Top veto	0.461 ±0.008	0.06 ±0.02	0.06 ±0.04	0.3 ±0.1	1.54 ±0.02	0.034 ±0.003	0.033 ±0.002	0.039 ±0.005	2.1 ±0.2	1
m_{ll}	0.350 ±0.007	0.005 ±0.005	0	0.2 ±0.1	0.440 ±0.010	0.007 ±0.001	0.010 ±0.001	0.008 ±0.002	0.7 ±0.1	0
p_{Tmax}	0.318 ±0.006	0.005 ±0.005	0	0.2 ±0.1	0.362 ±0.009	0.005 ±0.001	0.0088 ±0.0009	0.008 ±0.002	0.6 ±0.1	0
p_{Tmin}	0.256 ±0.006	0.005 ±0.005	0	0	0.221 ±0.007	0.0022 ±0.0008	0.0060 ±0.0008	0.004 ±0.002	0.238 ±0.009	0
$\Delta\phi$	0.211 ±0.005	0.005 ±0.005	0	0	0.151 ±0.006	0.0019 ±0.0008	0.0049 ±0.0007	0.003 ±0.002	0.166 ±0.008	0

Table 6.11: Expected number of events for 36.1 pb^{-1} in the ee channel for the $m_H = 160 \ GeV/c^2$ selection

	HWW160	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	3.06 ±0.02	68.3 ±0.6	61 ±1	3.2 ±0.4	15.88 ±0.06	0.77 ±0.02	0.110 ±0.003	4.96 ±0.06	154 ±2	134
Z veto	3.06 ±0.02	68.3 ±0.6	61 ±1	3.2 ±0.4	15.88 ±0.06	0.77 ±0.02	0.110 ±0.003	4.96 ±0.06	154 ±2	134
projectedMET	2.78 ±0.02	52.0 ±0.5	1.4 ±0.2	1.9 ±0.3	11.43 ±0.05	0.52 ±0.01	0.033 ±0.002	3.85 ±0.05	71.1 ±0.6	64
CJV	1.33 ±0.01	0.50 ±0.05	0.12 ±0.06	1.0 ±0.2	7.10 ±0.04	0.183 ±0.008	0.0084 ±0.0009	0.25 ±0.01	9.1 ±0.3	9
Softmuon	1.33 ±0.01	0.44 ±0.05	0.12 ±0.06	1.0 ±0.2	7.07 ±0.04	0.158 ±0.007	0.0073 ±0.0009	0.22 ±0.01	9.0 ±0.3	9
Extralepton veto	1.33 ±0.01	0.44 ±0.05	0.12 ±0.06	1.0 ±0.2	7.07 ±0.04	0.135 ±0.007	0.0068 ±0.0008	0.22 ±0.01	9.0 ±0.3	9
Top veto	1.32 ±0.01	0.29 ±0.04	0.12 ±0.06	1.0 ±0.2	7.03 ±0.04	0.135 ±0.007	0.0067 ±0.0008	0.18 ±0.01	8.8 ±0.3	9
m_{ll}	0.93 ±0.01	0.04 ±0.01	0.03 ±0.03	0.2 ±0.1	1.25 ±0.02	0.027 ±0.003	0.0021 ±0.0005	0.029 ±0.005	1.6 ±0.1	0
p_{Tmax}	0.84 ±0.01	0.03 ±0.01	0	0.10 ±0.07	0.94 ±0.01	0.020 ±0.003	0.0016 ±0.0004	0.027 ±0.005	1.12 ±0.08	0
p_{Tmin}	0.664 ±0.009	0.015 ±0.008	0	0.008 ±0.005	0.54 ±0.01	0.012 ±0.002	0.0007 ±0.0003	0.017 ±0.004	0.60 ±0.02	0
$\Delta\phi$	0.534 ±0.008	0.010 ±0.007	0	0	0.352 ±0.009	0.008 ±0.002	0.0006 ±0.0002	0.016 ±0.004	0.39 ±0.01	0

Table 6.12: Expected number of events for 36.1 pb^{-1} in the $e\mu$ channel for the m_H= 160 GeV/c^2 selection

	HWW160	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	6.19 ±0.03	138.2 ±0.8	25165 ±28	176.1 ±0.9	34.20 ±0.09	18.59 ±0.08	11.66 ±0.03	10.00 ±0.09	25554 ±28	24195
Z veto	6.02 ±0.03	122.4 ±0.8	2381 ±9	106.4 ±0.8	28.46 ±0.08	2.31 ±0.03	1.24 ±0.01	8.92 ±0.08	2651 ±9	2503
projectedMET	5.19 ±0.03	84.5 ±0.6	4.8 ±0.4	2.9 ±0.4	18.07 ±0.06	0.90 ±0.02	0.208 ±0.005	6.22 ±0.07	117.5 ±0.9	112
CJV	2.52 ±0.02	0.80 ±0.06	0.30 ±0.10	1.3 ±0.3	11.15 ±0.05	0.33 ±0.01	0.102 ±0.003	0.39 ±0.02	14.4 ±0.3	11
Softmuon	2.51 ±0.02	0.70 ±0.06	0.30 ±0.10	1.3 ±0.3	11.11 ±0.05	0.275 ±0.009	0.098 ±0.003	0.34 ±0.02	14.2 ±0.3	11
Extralepton veto	2.51 ±0.02	0.70 ±0.06	0.30 ±0.10	1.3 ±0.3	11.11 ±0.05	0.240 ±0.009	0.097 ±0.003	0.34 ±0.02	14.1 ±0.3	11
Top veto	2.49 ±0.02	0.45 ±0.05	0.30 ±0.10	1.3 ±0.3	11.04 ±0.05	0.238 ±0.009	0.096 ±0.003	0.27 ±0.01	13.7 ±0.3	11
m_{ll}	1.82 ±0.02	0.06 ±0.02	0.09 ±0.05	0.4 ±0.2	2.42 ±0.02	0.050 ±0.004	0.029 ±0.002	0.049 ±0.006	3.1 ±0.2	0
p_{Tmax}	1.65 ±0.01	0.05 ±0.02	0.03 ±0.03	0.3 ±0.1	1.91 ±0.02	0.039 ±0.003	0.023 ±0.002	0.043 ±0.006	2.4 ±0.2	0
p_{Tmin}	1.32 ±0.01	0.03 ±0.01	0.03 ±0.03	0.008 ±0.005	1.11 ±0.02	0.023 ±0.003	0.015 ±0.001	0.027 ±0.005	1.25 ±0.04	0
$\Delta\phi$	1.06 ±0.01	0.03 ±0.01	0.03 ±0.03	0	0.73 ±0.01	0.016 ±0.002	0.012 ±0.001	0.025 ±0.004	0.85 ±0.04	0

Table 6.13: Expected number of events for 36.1 pb^{-1} in the $\mu\mu + ee + e\mu$ channel for the m_H= 160 GeV/c^2 selection

	HWW190	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	1.214 ±0.009	41.4 ±0.4	15152 ±21	105.9 ±0.6	11.07 ±0.05	10.58 ±0.06	6.90 ±0.03	2.95 ±0.05	15331 ±21	14834
Z veto	0.917 ±0.008	31.9 ±0.4	1448 ±7	62.6 ±0.5	7.63 ±0.04	0.89 ±0.02	0.691 ±0.008	2.31 ±0.04	1554 ±7	1437
projectedMET	0.734 ±0.007	19.2 ±0.3	2.2 ±0.3	0.2 ±0.1	4.08 ±0.03	0.221 ±0.008	0.110 ±0.003	1.41 ±0.03	27.4 ±0.4	32
CJV	0.349 ±0.005	0.18 ±0.03	0.12 ±0.06	0.011 ±0.005	2.50 ±0.02	0.101 ±0.006	0.059 ±0.002	0.084 ±0.008	3.06 ±0.07	1
Softmuon	0.348 ±0.005	0.17 ±0.03	0.12 ±0.06	0.011 ±0.005	2.49 ±0.02	0.074 ±0.005	0.058 ±0.002	0.071 ±0.007	3.00 ±0.07	1
Extralepton veto	0.348 ±0.005	0.17 ±0.03	0.12 ±0.06	0.011 ±0.005	2.49 ±0.02	0.070 ±0.005	0.057 ±0.002	0.071 ±0.007	2.99 ±0.07	1
Top veto	0.346 ±0.005	0.10 ±0.02	0.12 ±0.06	0.011 ±0.005	2.48 ±0.02	0.070 ±0.005	0.057 ±0.002	0.058 ±0.007	2.89 ±0.07	1
m_{ll}	0.331 ±0.005	0.04 ±0.01	0.09 ±0.05	0.008 ±0.005	1.48 ±0.02	0.044 ±0.004	0.035 ±0.002	0.029 ±0.005	1.73 ±0.06	1
p_{Tmax}	0.254 ±0.004	0.04 ±0.01	0.03 ±0.03	0.005 ±0.004	0.88 ±0.01	0.030 ±0.003	0.025 ±0.002	0.022 ±0.004	1.03 ±0.04	1
p_{Tmin}	0.195 ±0.004	0.03 ±0.01	0.03 ±0.03	0.005 ±0.004	0.60 ±0.01	0.020 ±0.003	0.018 ±0.001	0.015 ±0.003	0.72 ±0.03	1
$\Delta\phi$	0.162 ±0.003	0.02 ±0.01	0	0	0.368 ±0.009	0.012 ±0.002	0.012 ±0.001	0.010 ±0.003	0.43 ±0.01	0

Table 6.14: Expected number of events for 36.1 pb^{-1} in the $\mu\mu$ channel for the $m_H = 190\ GeV/c^2$ selection

	HWW190	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	0.831 ±0.008	28.5 ±0.4	9952 ±17	66.9 ±0.6	7.25 ±0.04	7.24 ±0.05	4.65 ±0.02	2.10 ±0.04	10069 ±17	9227
Z veto	0.629 ±0.007	22.2 ±0.3	872 ±5	40.6 ±0.5	4.95 ±0.03	0.65 ±0.01	0.443 ±0.007	1.66 ±0.04	942 ±5	932
projectedMET	0.497 ±0.006	13.2 ±0.3	1.2 ±0.2	0.8 ±0.2	2.57 ±0.02	0.159 ±0.007	0.066 ±0.003	0.97 ±0.03	18.9 ±0.4	16
CJV	0.238 ±0.004	0.12 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.046 ±0.004	0.034 ±0.002	0.057 ±0.007	2.2 ±0.2	1
Softmuon	0.237 ±0.004	0.10 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.043 ±0.004	0.033 ±0.002	0.051 ±0.006	2.2 ±0.2	1
Extralepton veto	0.237 ±0.004	0.10 ±0.02	0.06 ±0.04	0.3 ±0.1	1.55 ±0.02	0.034 ±0.003	0.033 ±0.002	0.051 ±0.006	2.2 ±0.2	1
Top veto	0.236 ±0.004	0.06 ±0.02	0.06 ±0.04	0.3 ±0.1	1.54 ±0.02	0.034 ±0.003	0.033 ±0.002	0.039 ±0.005	2.1 ±0.2	1
m_{ll}	0.227 ±0.004	0.015 ±0.008	0.06 ±0.04	0.2 ±0.1	0.88 ±0.01	0.019 ±0.002	0.021 ±0.001	0.015 ±0.003	1.3 ±0.1	1
p_{Tmax}	0.178 ±0.004	0.015 ±0.008	0.06 ±0.04	0.003 ±0.003	0.53 ±0.01	0.012 ±0.002	0.016 ±0.001	0.012 ±0.003	0.64 ±0.05	0
p_{Tmin}	0.136 ±0.003	0.015 ±0.008	0.03 ±0.03	0.003 ±0.003	0.364 ±0.009	0.008 ±0.002	0.011 ±0.001	0.008 ±0.002	0.44 ±0.03	0
$\Delta\phi$	0.111 ±0.003	0.005 ±0.005	0.03 ±0.03	0	0.229 ±0.007	0.004 ±0.001	0.0091 ±0.0010	0.008 ±0.002	0.28 ±0.03	0

Table 6.15: Expected number of events for 36.1 pb^{-1} in the ee channel for the $m_H = 190\ GeV/c^2$ selection

	HWW190	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	2.03 ±0.01	68.3 ±0.6	61 ±1	3.2 ±0.4	15.88 ±0.06	0.77 ±0.02	0.110 ±0.003	4.96 ±0.06	154 ±2	134
Z veto	2.03 ±0.01	68.3 ±0.6	61 ±1	3.2 ±0.4	15.88 ±0.06	0.77 ±0.02	0.110 ±0.003	4.96 ±0.06	154 ±2	134
projectedMET	1.80 ±0.01	52.0 ±0.5	1.4 ±0.2	1.9 ±0.3	11.43 ±0.05	0.52 ±0.01	0.033 ±0.002	3.85 ±0.05	71.1 ±0.6	64
CJV	0.850 ±0.008	0.50 ±0.05	0.12 ±0.06	1.0 ±0.2	7.10 ±0.04	0.183 ±0.008	0.0084 ±0.0009	0.25 ±0.01	9.1 ±0.3	9
Softmuon	0.847 ±0.008	0.44 ±0.05	0.12 ±0.06	1.0 ±0.2	7.07 ±0.04	0.158 ±0.007	0.0073 ±0.0009	0.22 ±0.01	9.0 ±0.3	9
Extralepton veto	0.847 ±0.008	0.44 ±0.05	0.12 ±0.06	1.0 ±0.2	7.07 ±0.04	0.135 ±0.007	0.0068 ±0.0008	0.22 ±0.01	9.0 ±0.3	9
Top veto	0.841 ±0.008	0.29 ±0.04	0.12 ±0.06	1.0 ±0.2	7.03 ±0.04	0.135 ±0.007	0.0067 ±0.0008	0.18 ±0.01	8.8 ±0.3	9
m_{ll}	0.651 ±0.007	0.09 ±0.02	0.09 ±0.05	0.7 ±0.2	3.16 ±0.03	0.061 ±0.004	0.0049 ±0.0007	0.065 ±0.007	4.2 ±0.2	4
p_{Tmax}	0.505 ±0.006	0.06 ±0.02	0	0.6 ±0.2	1.78 ±0.02	0.033 ±0.003	0.0023 ±0.0005	0.043 ±0.006	2.5 ±0.2	3
p_{Tmin}	0.388 ±0.005	0.03 ±0.01	0	0.2 ±0.1	1.25 ±0.02	0.022 ±0.003	0.0017 ±0.0004	0.033 ±0.005	1.6 ±0.1	0
$\Delta\phi$	0.295 ±0.005	0.015 ±0.008	0	0.011 ±0.005	0.61 ±0.01	0.011 ±0.002	0.0008 ±0.0003	0.017 ±0.004	0.67 ±0.02	0

Table 6.16: Expected number of events for 36.1 pb^{-1} in the $e\mu$ channel for the m_H= 190 GeV/c^2 selection

	HWW190	TTbar	ZJets	WJets	WW	WZ	ZZ	tW	Total bkg	Data
diLepton	4.07 ±0.02	138.2 ±0.8	25165 ±28	176.1 ±0.9	34.20 ±0.09	18.59 ±0.08	11.66 ±0.03	10.00 ±0.09	25554 ±28	24195
Z veto	3.57 ±0.02	122.4 ±0.8	2381 ±9	106.4 ±0.8	28.46 ±0.08	2.31 ±0.03	1.24 ±0.01	8.92 ±0.08	2651 ±9	2503
projectedMET	3.03 ±0.01	84.5 ±0.6	4.8 ±0.4	2.9 ±0.4	18.07 ±0.06	0.90 ±0.02	0.208 ±0.005	6.22 ±0.07	117.5 ±0.9	112
CJV	1.44 ±0.01	0.80 ±0.06	0.30 ±0.10	1.3 ±0.3	11.15 ±0.05	0.33 ±0.01	0.102 ±0.003	0.39 ±0.02	14.4 ±0.3	11
Softmuon	1.43 ±0.01	0.70 ±0.06	0.30 ±0.10	1.3 ±0.3	11.11 ±0.05	0.275 ±0.009	0.098 ±0.003	0.34 ±0.02	14.2 ±0.3	11
Extralepton veto	1.43 ±0.01	0.70 ±0.06	0.30 ±0.10	1.3 ±0.3	11.11 ±0.05	0.240 ±0.009	0.097 ±0.003	0.34 ±0.02	14.1 ±0.3	11
Top veto	1.42 ±0.01	0.45 ±0.05	0.30 ±0.10	1.3 ±0.3	11.04 ±0.05	0.238 ±0.009	0.096 ±0.003	0.27 ±0.01	13.7 ±0.3	11
m_{ll}	1.208 ±0.009	0.15 ±0.03	0.24 ±0.09	1.0 ±0.2	5.53 ±0.03	0.123 ±0.006	0.061 ±0.002	0.108 ±0.009	7.2 ±0.3	6
p_{Tmax}	0.937 ±0.008	0.11 ±0.02	0.09 ±0.05	0.6 ±0.2	3.18 ±0.03	0.075 ±0.005	0.043 ±0.002	0.077 ±0.008	4.1 ±0.2	4
p_{Tmin}	0.719 ±0.007	0.07 ±0.02	0.06 ±0.04	0.3 ±0.1	2.22 ±0.02	0.050 ±0.004	0.032 ±0.002	0.056 ±0.007	2.7 ±0.1	1
$\Delta\phi$	0.569 ±0.006	0.04 ±0.01	0.03 ±0.03	0.011 ±0.005	1.21 ±0.02	0.027 ±0.003	0.022 ±0.002	0.035 ±0.005	1.38 ±0.04	0

Table 6.17: Expected number of events for 36.1 pb^{-1} in the $\mu\mu + ee + e\mu$ channel for the m_H= 190 GeV/c^2 selection

6.8   BDT Analysis

When a small signal needs to be extracted from a large background, as is the case for the Higgs searches, more sophisticated methods than the usual cut-based approaches have to be explored and applied in order to gain sensitivity.

Multivariate techniques, such as Likelihoods, Neural Networks or Boosted Decision Trees have been widely used in High Energy Physics and in previous experiments, as in LEP and the Tevatron, to make maximal use of the information in data.

A multivariate analysis consists of two independent steps: a training phase, where the multivariate methods are trained, tested and evaluated, using an independent sample with a known signal and background composition, allowing the corresponding classifier to learn how to distinguish the properties of signal and background, and then an application phase, where the resulting function is applied for the classification of data samples with unknown composition.

The classifier used in this analysis is the so called Boosted Decision Trees (BDT), which is implemented in the TMVA (Toolkit for Multivariate Analysis) that provides a ROOT-integrated environment for the processing, parallel evaluation and application of multivariate classification techniques [109].

A Decision Tree is a binary classifier, in which successive decisions on a single variable at each level about wether a particular condition is satisfied or not are consecutively taken until a certain stop criterion is reached. The splitting criteria for each node is defined in the training process by finding the variable and corresponding cut value that provides the best separation between signal and background. Several of these decision trees are constructed and combined using the boosting to achieve a most robust classifier. A more detailed description of the BDT and the configuration of the parameters is shown in Appendix A.

The training phase involves the careful selection of the optimal variables that provide the best separation between signal and background, as well as the tuning of the different parameters of the BDT configuration.

The analysis will be described in detail in the following sections. Two different approaches have been followed, a three dimensional BDT (BDT 3D) which uses three different trainings against the main backgrounds and a one-dimensional BDT (BDT 1D).

6.8.1   BDT 3D

As described before, the analysis starts with the training of the BDT. Several parameters were tested, as well as different input variables and the size of the samples used for training. No significant improvement has been observed when changing the parameters from the default configuration. Computing time increases if more trees are added, but as no important gain was obtained, 400 trees have been taken for the training. The list of parameters used is summarized in Table 6.18. The same was used for all the trainings discussed in this section.

This analysis is based on the observation that the main backgrounds after an initial preselection is applied are the WW, $t\bar{t}$ and $Z/\gamma^* \to l^+l^-$, so the method has been optimized to reduce them as

Parameter	Value
N_{Trees}	400
Boost Type	AdaBoost
Separation Type	GiniIndex
nCuts	20
minEvents	$\max(20, \frac{N_{training}}{N^2_{variables}/10})$
Prune Method	CostComplexity
PruneStrenght	12

Table 6.18: BDT configuration parameters: Boost type, separation criteria for node splitting and granularity of the cut scanning, minimum number of events in the leaf node, prune method and amount of pruning.

much as possible.

The first step is to identify an adequate set of variables with some discrimination power. In addition to the main variables used in the cut-based analysis (p^{lep}_{Tmax}, p^{lep}_{Tmin}, $\Delta\phi_{ll}$, m_{ll} and $\not{E}_T$), some more have been identified to provide an additional discrimination against the backgrounds, although they are correlated with the other ones or the separation power is not so important, all together improve the BDT performance:

- The angle in the transverse plane between the $\not{E}_T$ and the leptons.This variable discriminates against events with no real $\not{E}_T$.

- The projected $\not{E}_T$

- $\Delta\phi$ between the leptons and the $\not{E}_T$, which helps to reduce the processes with non real $\not{E}_T$

- $\Delta\eta$ between the leptons, this provides some discrimination against WW background, as leptons from gg scattering are more central than those from $q\bar{q}$ as WW, resulting in a more central pseudorapidity distribution.

- p_T of the lepton pair divided by the $\not{E}_T$, this discriminates against $Z/\gamma^* \rightarrow l^+l^-$.

- Transverse mass of both lepton-$\not{E}_T$ pairs, which helps reducing non-W background

- The number of good tracks in the event, included to achieve a better $t\bar{t}$ reduction. This variable is constructed considering only the tracks passing some standard quality cuts: $IP < 0.1$cm, $dz < 0.2$cm, $p_T > 3$ GeV/c and at least 11 valid hits.

The BDT is not sensitive to the addition of variables with poor discrimination power, as they will be less used in the process of building the tree, so in principle any variable with some discrimination power can be added without weakening the performance of the BDT. A summary of the variables used in the BDT analysis is listed on Table 6.19.

The distributions of the variables used in the BDT training, comparing data and MC simulation are shown in Figures 6.9 to 6.12.

One of the main issues to deal with, is the fact that there are several backgrounds with different properties and how to treat them properly in the training step in order to obtain the best separation
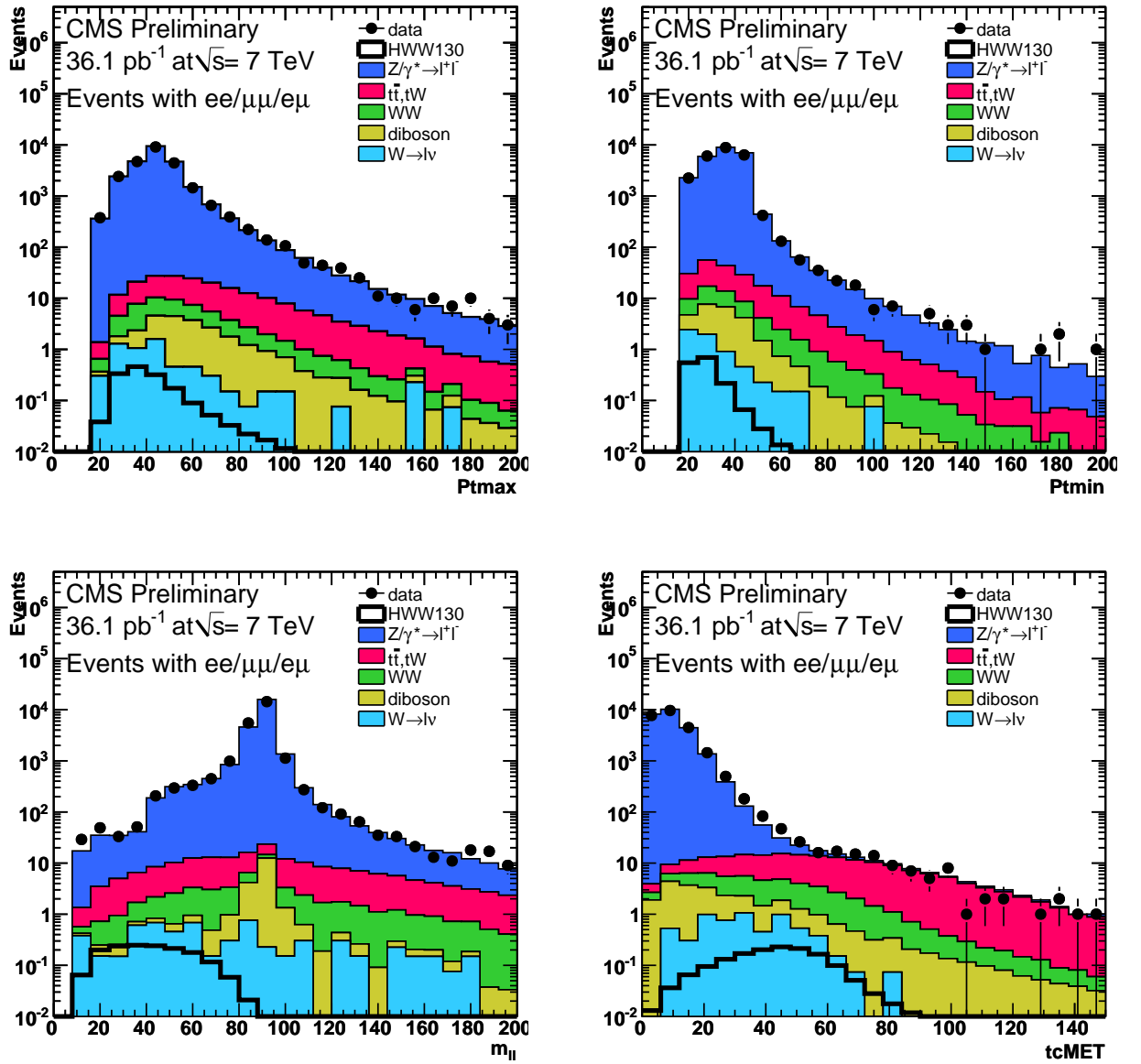
Figure 6.9: - Distributions of p_{Tmax}^{lep}, p_{Tmin}^{lep}, m_{ll} and $\not{E}_T$ for events passing the dilepton selection
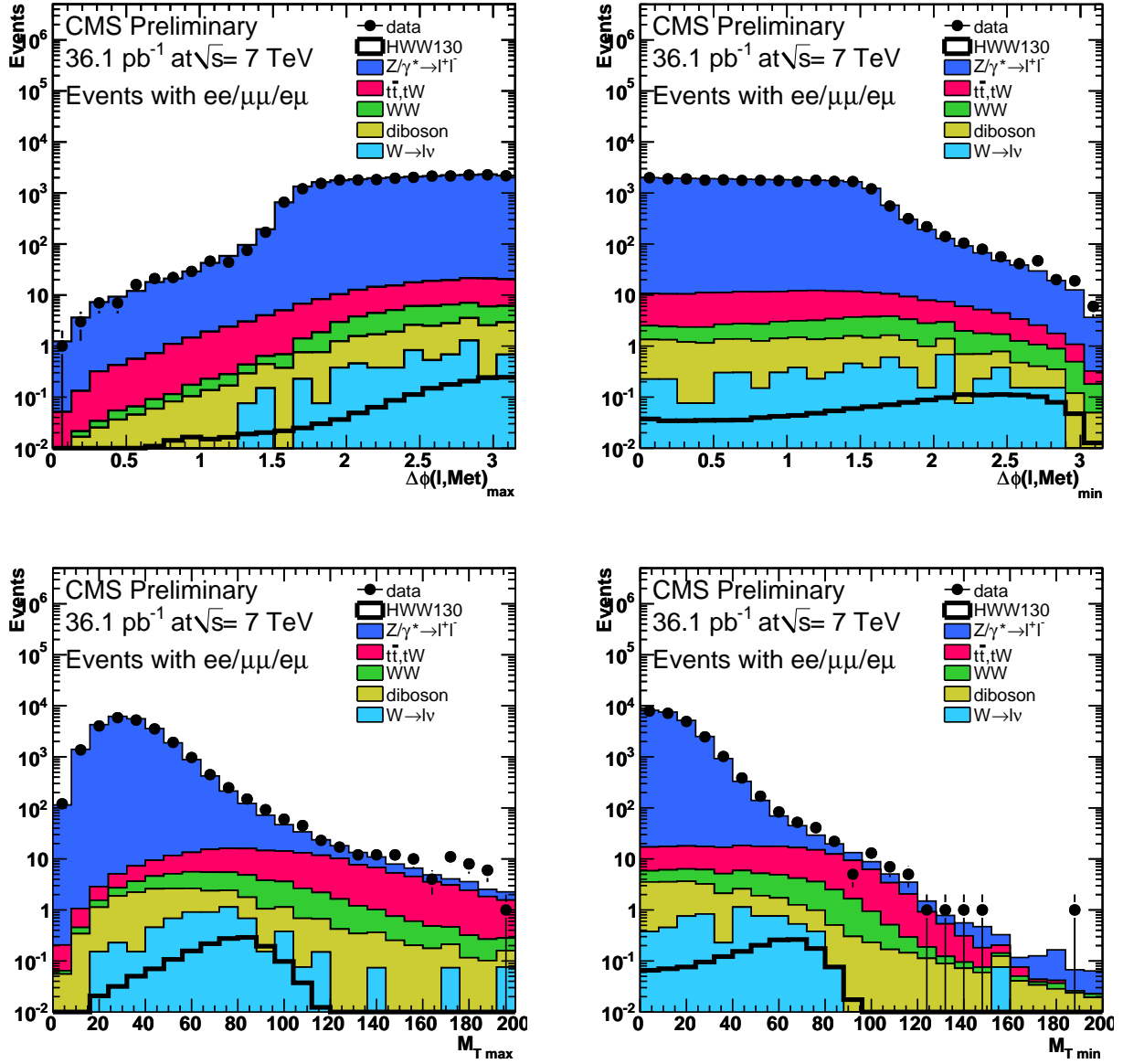
Figure 6.10: - Distributions of max $(\Delta\phi(l^1, \not{E}_T), \Delta\phi(l^2, \not{E}_T))$, min $(\Delta\phi(l^1, \not{E}_T))$, $m_T^{l_{max} - \not{E}_T}$ and $m_T^{l_{min} - \not{E}_T}$) for events passing the dilepton selection
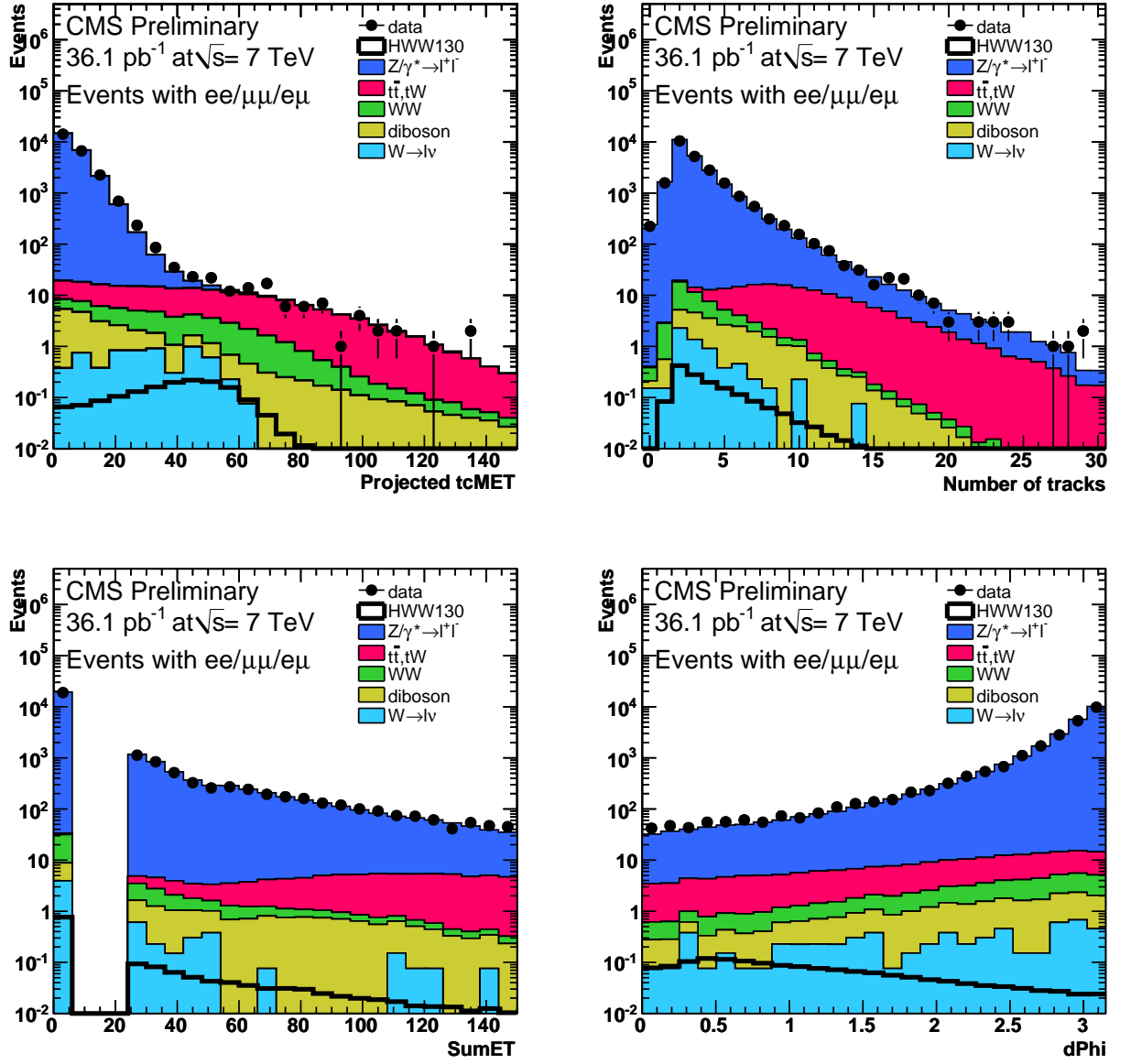
Figure 6.11: - Distributions of projected $\not{E}_T$, N_{tracks}, Sum of the energy of the selected jets and $\Delta\phi_{ll}$ for events passing the dilepton selection
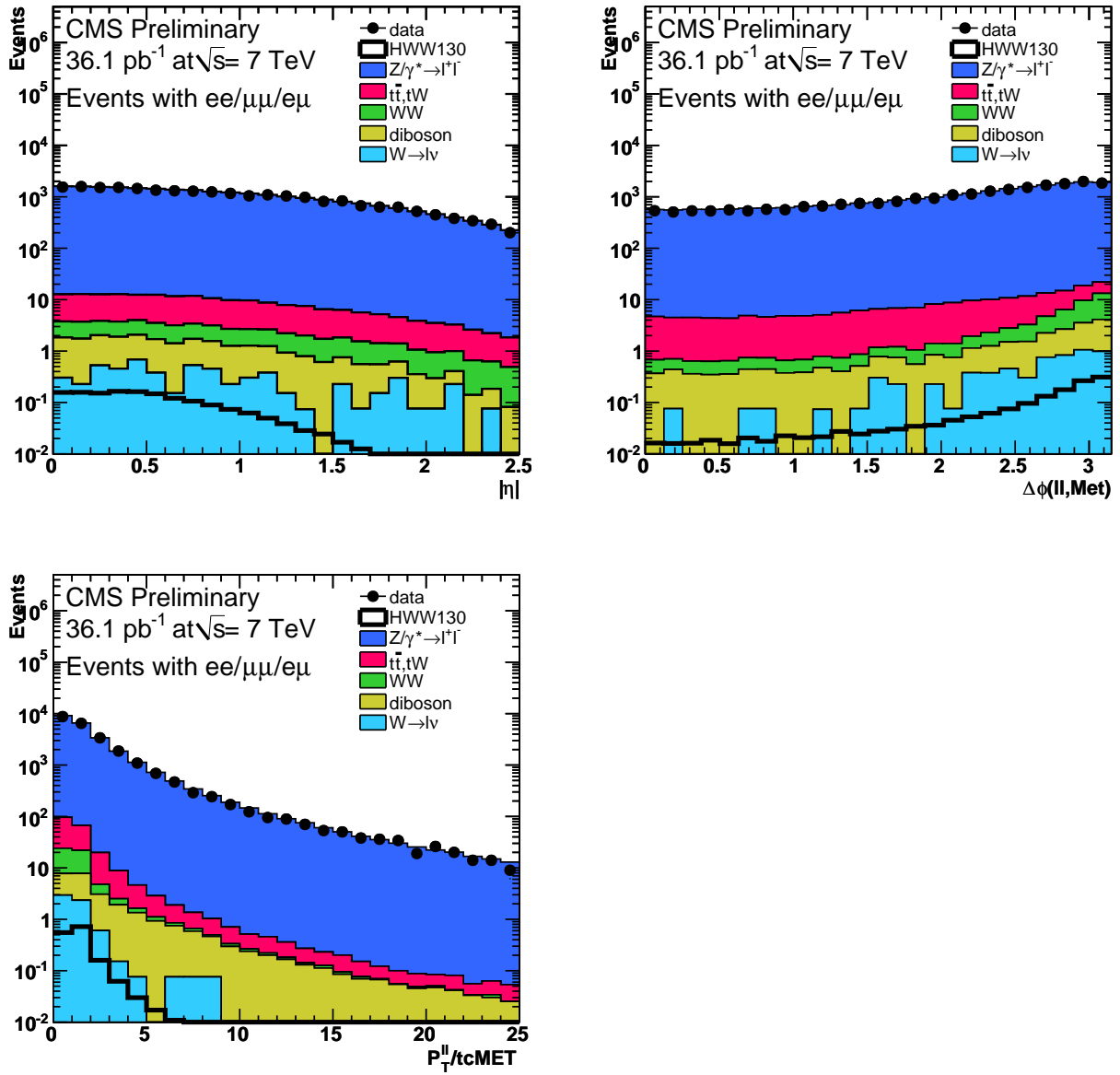
Figure 6.12: - Distributions of $\Delta\eta_{ll}$, $\Delta\phi(ll, E\!\!\!/_T)$ and $p_T^{ll}/E\!\!\!/_T$ for events passing the dilepton selection

Variables used for training
1. p^l_{Tmax}
2. p^l_{Tmin}
3. $\not{E}_T$
4. m_{ll}
5. $\Delta\phi_{ll}$
6. $\Delta\eta_{ll}$
7. $m_T^{l_{max}-\not{E}_T}$
8. $m_T^{l_{min}-\not{E}_T}$
9. $p_T^{ll}/\not{E}_T$
10. $\Delta\phi(ll, \not{E}_T)$
11. $\max\left(\Delta\phi(l^1, \not{E}_T), \Delta\phi(l^2, \not{E}_T)\right)$
12. $\min\left(\Delta\phi(l^1, \not{E}_T), \Delta\phi(l^2, \not{E}_T)\right)$
13. projected $\not{E}_T$
14. N_{tracks}

Table 6.19: List of variables used for training the BDTs

for each. The variables discussed before have different shapes for each background, so the best variable to discriminate against one background might not be so good for others. In order to exploit their discriminating power, an approach that has been found to provide good results is to train independent BDTs for each of the main backgrounds: WW, $t\bar{t}$ and $Z/\gamma^* \to l^+l^-$. This way each function is optimized separately and combined afterwards to form the final discriminant variable.

The sample used for training has been obtained after the dilepton selection described in Section 6.5. The three final states together are considered in this analysis. A cut on Projected$\not{E}_T$ >20 GeV is also applied to the sample in order to reduce the $Z/\gamma^* \to l^+l^-$ contribution.

Once the three BDTs are trained, the functions are combined into a three-dimensional one, the signal and backgrounds are evaluated by both, such that WW-like backgrounds are mostly rejected by the first BDT, ttbar-like backgrounds by the second BDT and Z+Jets by the third one. An example of this can be observed in Figure 6.13, where the BDT 3D distributions are shown for the signal ($m_H = 130 GeV/c^2$) and main backgrounds: WW, $t\bar{t}$ and Z+jets. As can be observed, the signal clearly peaks in values around $\sim (0.9, 0.9, 0.9)$, while the backgrounds are more spread and extend toward lower values.

Finally, a spherical cut is applied around the region with highest signal content, the centre of the sphere and the radius are chosen optimizing the significance. This radius is the final discriminating variable and is defined as:

$$dR = \sqrt{(0.85 - BDT_1)^2 + (0.85 - BDT_2)^2 + (0.95 - BDT_3)^2} \tag{6.2}$$

As the kinematics change with the mass (see Figure 6.14), several trainings have been performed at different mass points: low mass, intermediate mass and high mass. The optimization process is repeated for each mass and examples for 130, 160 and 190 GeV/c^2 will be shown in the following sections.
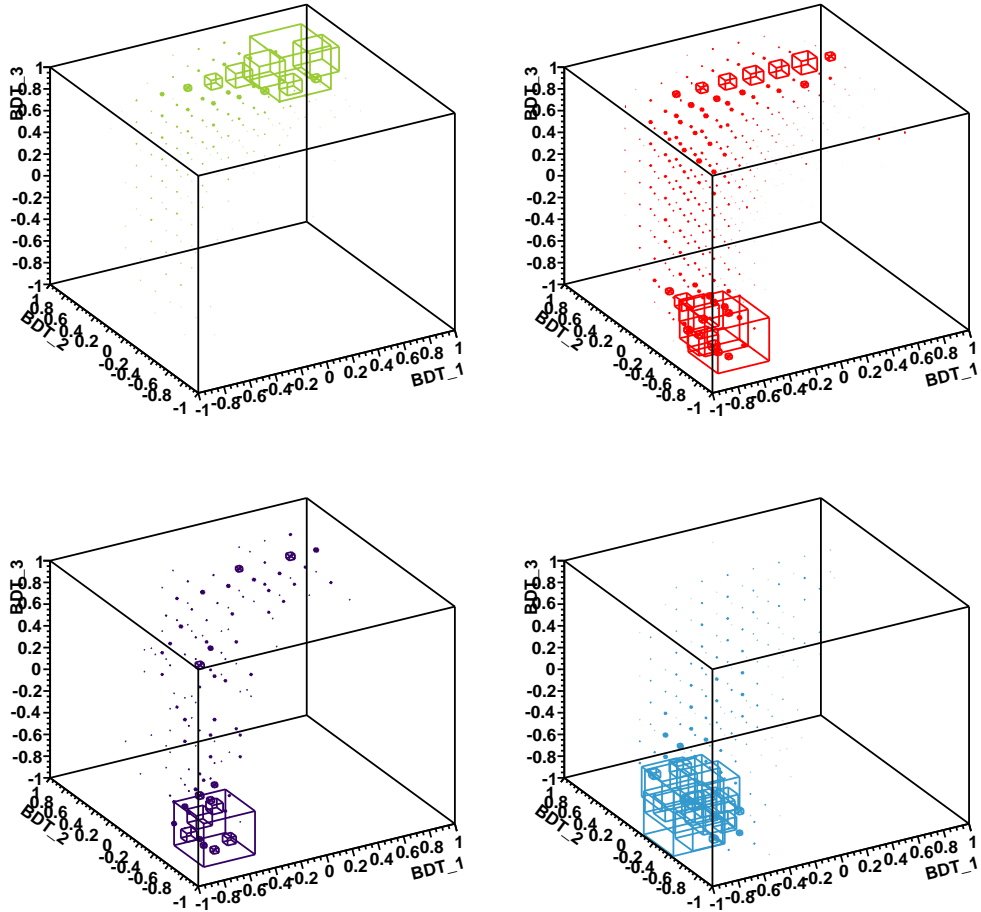
Figure 6.13: - BDT 3D distributions for signal (m_H= 130 GeV/c^2), WW, $t\bar{t}$ and Z+Jets (from top left, clockwise). The X axis shows the response of the BDT trained against WW, the Y axis the response of the BDT trained against $t\bar{t}$ and the Z axis the response of the BDT trained against Z+Jets. As can be observed, the signal tends to be concentrated around $\sim$(0.9,0.9,0.9), while the backgrounds are more spread and extend toward lower values
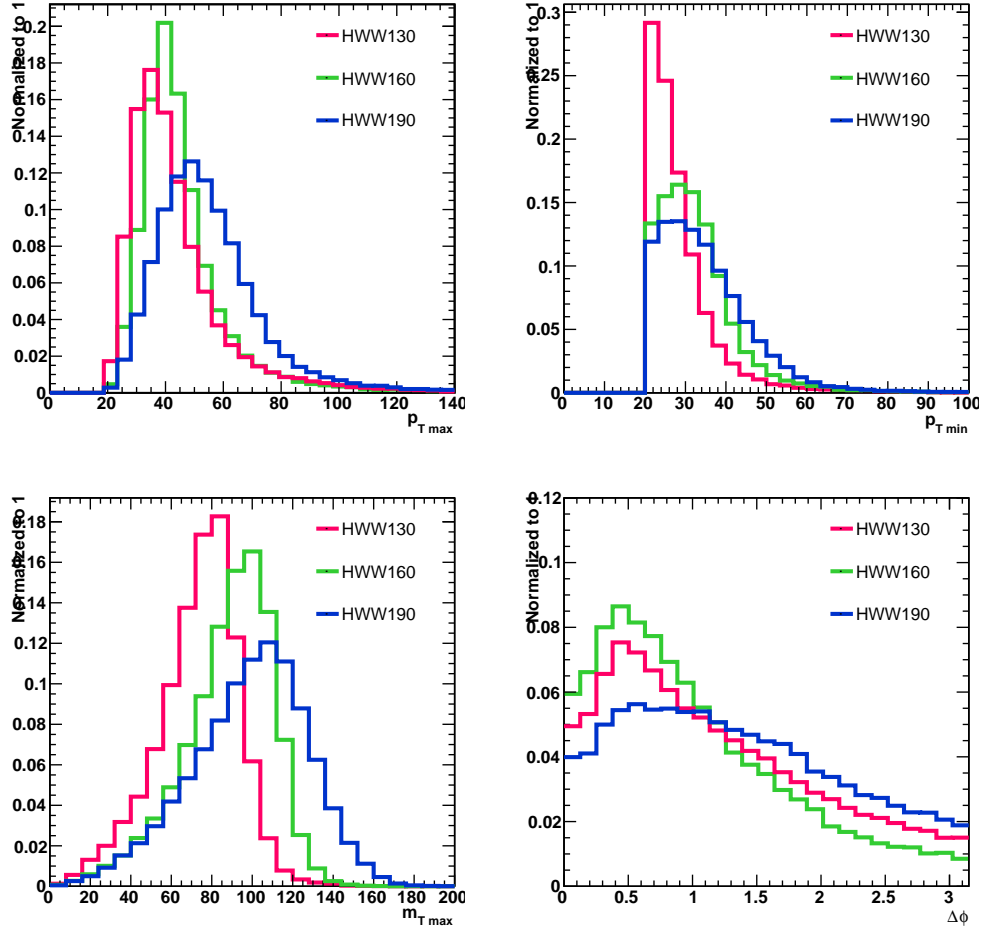
Figure 6.14: - Distributions of the p_T of the two leptons (p_{Tmax}^{lep} , p_{Tmin}^{lep}), $m_T^{l_{max} - \not{E}_T}$ and $\Delta\phi_{ll}$ for several Higgs mass hypotheses (130, 160 and 190 GeV/c^2)

The BDT 3D method described here is found to achieve a good discrimination power between signal and backgrounds if comparing with the cut-based analysis, specially against WW in those regions in which the $\Delta\phi$ variable is less discriminating.

More details on the training process and final results are given in the following sections.

6.8.1.1 Training at $m_H = $ 130 GeV/c^2

One of the aspects to be checked after the training is the importance of the variables used to construct the decision trees, the most they are used in the node splitting the best they are ranked, this provides an idea of which are the most discriminating variables to separate the signal from the background and will allow also to confirm that the classifier is working as expected. An example is shown in Table 6.20. As expected, in the training against WW background $\Delta\phi_{ll}$ variable is the one providing the best discrimination power, in the training against $t\bar{t}$ is N_{tracks} the most used and in the $Z/\gamma^* \to l^+l^-$ case is $\not{E}_T$ and the angles between leptons and $\not{E}_T$ the most discriminating ones. Distributions of some of these variables are presented in Figure 6.15.
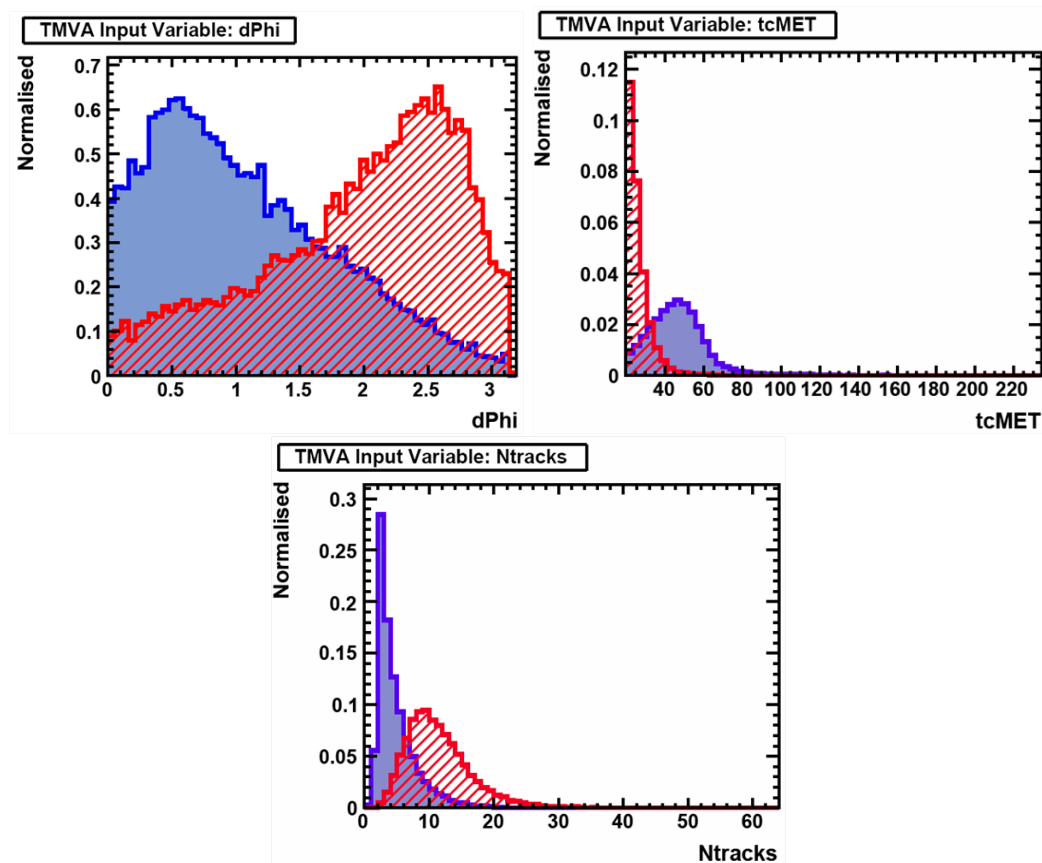
Figure 6.15: Examples of the distributions of some variables used in the BDT trainings: $\Delta\phi_{ll}$ for signal and WW, $\not{E}_T$ for signal and Z+jets and N_{tracks} for signal and $t\bar{t}$ events, in the $m_H = 130 \, GeV/c^2$ hypothesis.

Another important point is to check if there is overtraining risk. For this purpose, the TMVA program provides an estimation of the signal efficiencies at different background efficiency points, as well as the BDT function obtained with the independent training and test samples. Check this is well behaved, allow to be confident that the function is properly trained.

Figures 6.16 and 6.17 show the overtraining check performed for the three BDTs, where one can observe that the shape of the function is about the same for the training and test samples.

For the final analysis, in addition to the preselection cut described before, events are required to pass the jet veto and top-tagging cuts in order to better control the $t\bar{t}$ background.

The distribution of the BDT functions trained against WW (BDT1), $t\bar{t}$ (BDT2) and Z+Jets (BDT3) for events passing the final selection before the BDT cut is shown in Figure 6.18.

The final discriminant variable is represented in Figure 6.19. The signal tends to be concentrated at smaller values while the backgrounds tend to have larger values.

The number of signal and background events after applying various cuts on the classifier is represented in Figure 6.20.
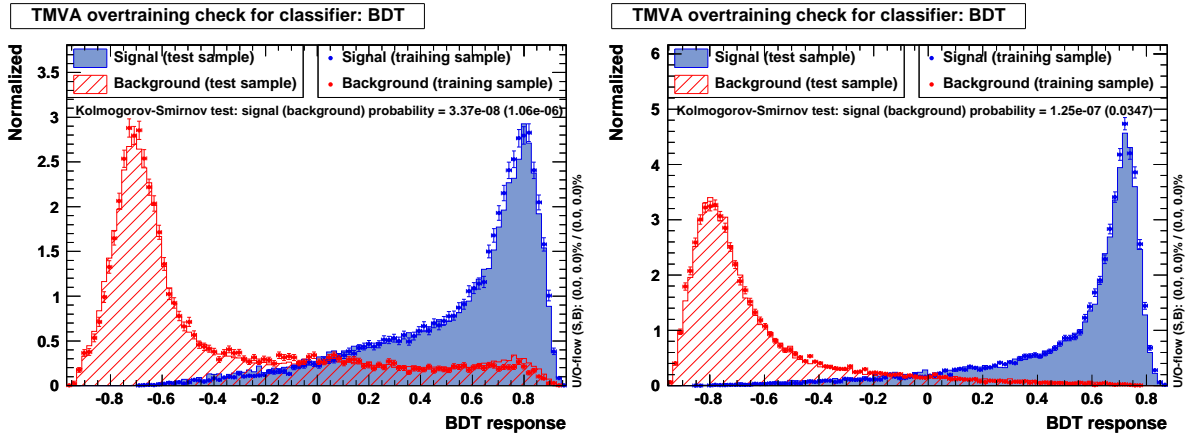
Figure 6.16: - Overtraining check for the BDT trained against WW (left) and against $t\bar{t}$ (right), in the $m_H = 130 \ GeV/c^2$ hypothesis.
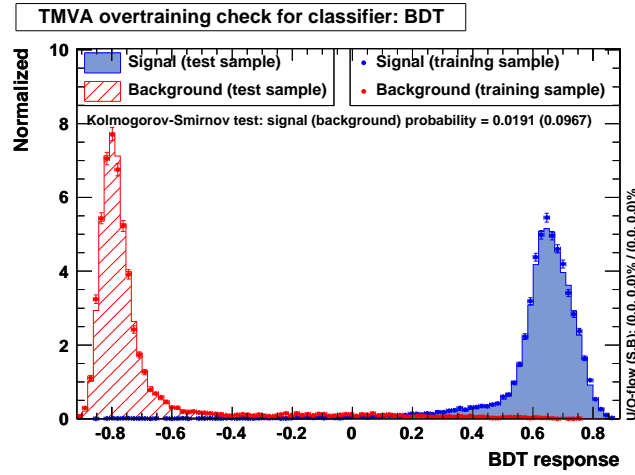
Figure 6.17: - Overtraining check for the BDT trained against Z+jets, in the $m_H = 130 \ GeV/c^2$ hypothesis.

Rank	Training anti-WW	Training anti $t\bar{t}$	Training anti Z
1	$\Delta\phi_{ll}$	N_{tracks}	$\not{E}_T$
2	m_{ll}	$\Delta\phi_{max}(l-\not{E}_T)$	$\Delta\phi_{min}(l-\not{E}_T)$
3	$\Delta\phi_{max}(l-\not{E}_T)$	m_{ll}	$\Delta\phi(ll,\not{E}_T)$
4	$\Delta\eta_{ll}$	$\Delta\phi_{ll}$	$\Delta\phi_{max}(l-\not{E}_T)$
5	$m_{Tmax}(l-\not{E}_T)$	$\Delta\eta_{ll}$	$\Delta\phi_{ll}$
6	p_{Tmin}^{lep}	$\Delta\phi(ll,\not{E}_T)$	m_{ll}
7	$\Delta\phi(ll,\not{E}_T)$	$\Delta\phi_{min}(l-\not{E}_T)$	N_{tracks}
8	$m_{Tmin}(l-\not{E}_T)$	$m_{Tmax}(l-\not{E}_T)$	projected $\not{E}_T$
9	$\Delta\phi_{min}(l-\not{E}_T)$	$\not{E}_T$	$m_{Tmin}(l-\not{E}_T)$
10	projected $\not{E}_T$	$m_{Tmin}(l-\not{E}_T)$	$m_{Tmax}(l-\not{E}_T)$
11	$p_T^{ll}/\not{E}_T$	projected $\not{E}_T$	$\Delta\eta_{ll}$
12	p_{Tmax}^{lep}	p_{Tmax}^{lep}	$p_T^{ll}/\not{E}_T$
13	N_{tracks}	p_{Tmin}^{lep}	p_{Tmax}^{lep}
14	$\not{E}_T$	$p_T^{ll}/\not{E}_T$	p_{Tmin}^{lep}

Table 6.20: Ranking of the variables in the three BDTs
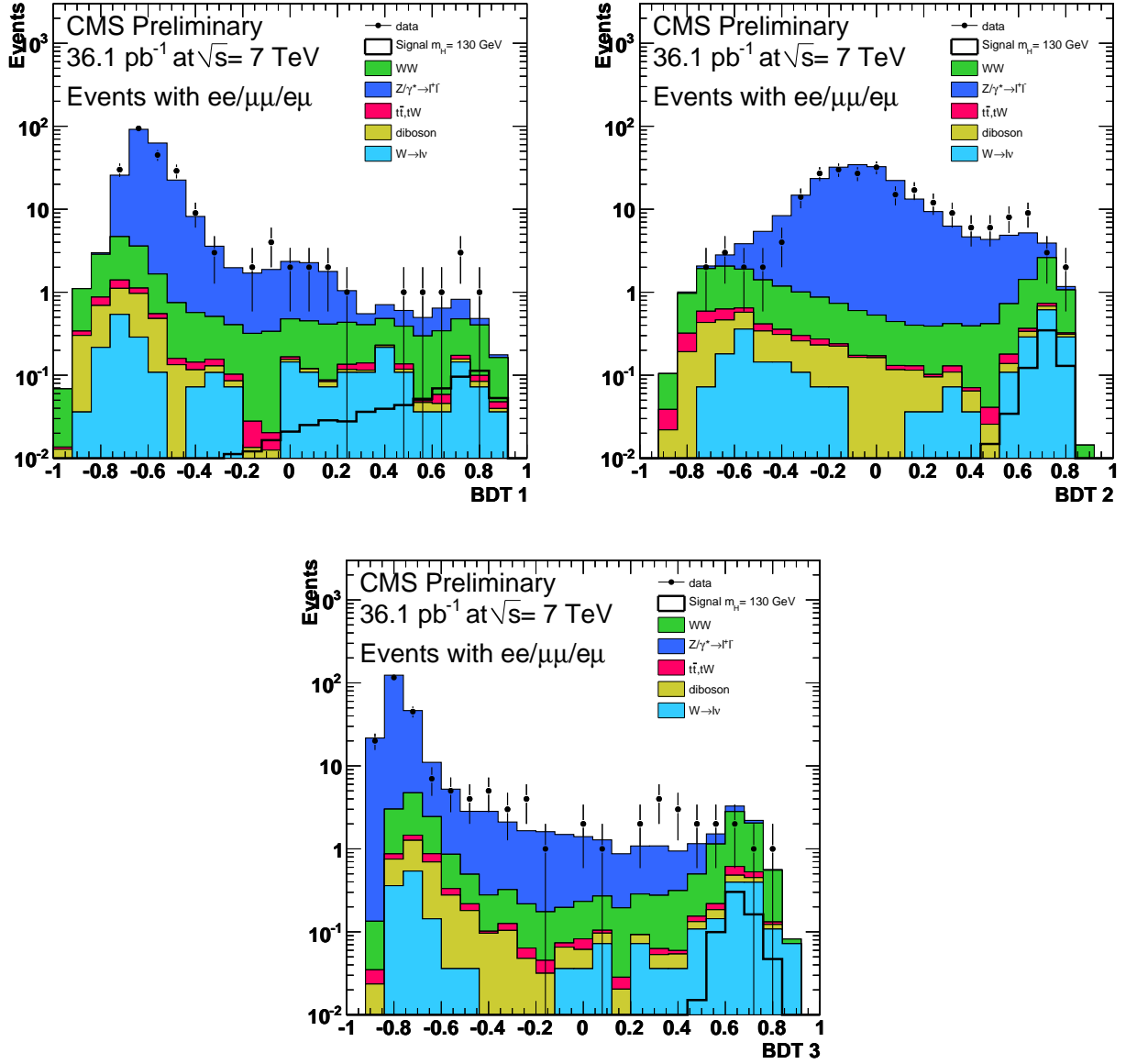
Figure 6.18: Distributions of the BDT functions trained against WW, $t\bar{t}$ and Z+jets, in the $m_H = 130$ GeV/c^2 hypothesis.
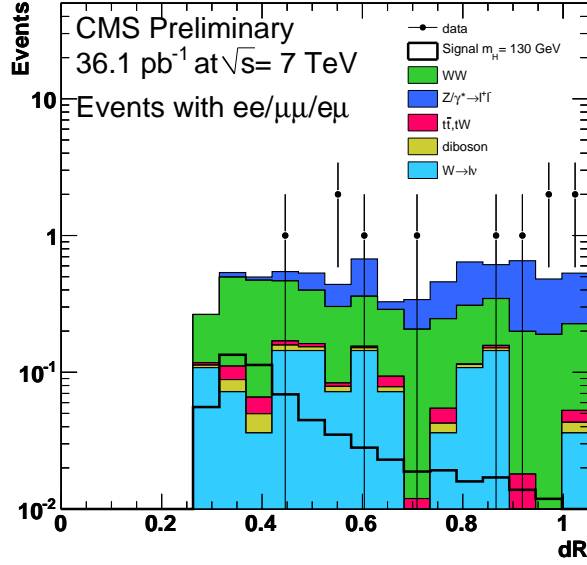
Figure 6.19: - Distribution of the final BDT variable (dR), optimized for m_H=130 GeV/c^2

The number of events after applying the optimal cuts on the BDT function is shown in Table 6.21.

	HWW130	WW	$t\bar{t}$	DY	VV	tW	Wjets	Total B	Data
R<0.45	0.350 ± 0.005	1.13 ± 0.02	0.03 ± 0.01	0.09 ± 0.04	0.044 ± 0.003	0.020 ± 0.004	0.3 ± 0.1	1.6 ± 0.1	1

Table 6.21: Expected number of events for $36pb^{-1}$, optimized for m_H= 130 GeV/c^2

Figure 6.21 shows the distribution of some variables after applying the optimal cut on the BDT.

6.8.1.2 Training at m_H= 160 GeV/c^2

The same training process explained before has been optimized for other mass hypothesis, the BDT distributions and final results are summarized here for m_H= 160 GeV/c^2

The distribution of the BDT function for the final sample before the BDT cut for m_H= 160 GeV/c^2 is shown in Figure 6.22.

The final discriminant variable is represented in Figure 6.23 and the Number of signal and background events after applying various cuts on this function is represented in Figure 6.24.

The number of events after applying the optimal cuts on the BDT function is shown in Table 6.22.

6.8.1.3 Training at m_H=190 GeV/c^2

The same training process explained before has been optimized in the same way for other mass hypothesis, the BDT distributions and final results are summarized here for m_H= 190 GeV/c^2.
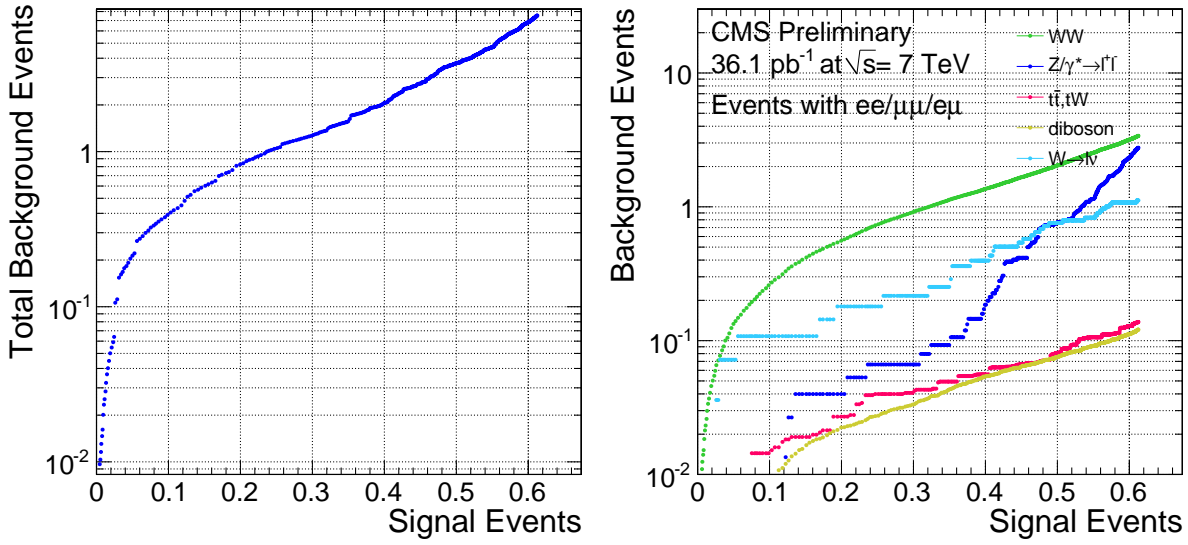
Figure 6.20: - Number of background vs signal events obtained after applying various cuts on the dR variable optimized for m_H=130 GeV/c^2
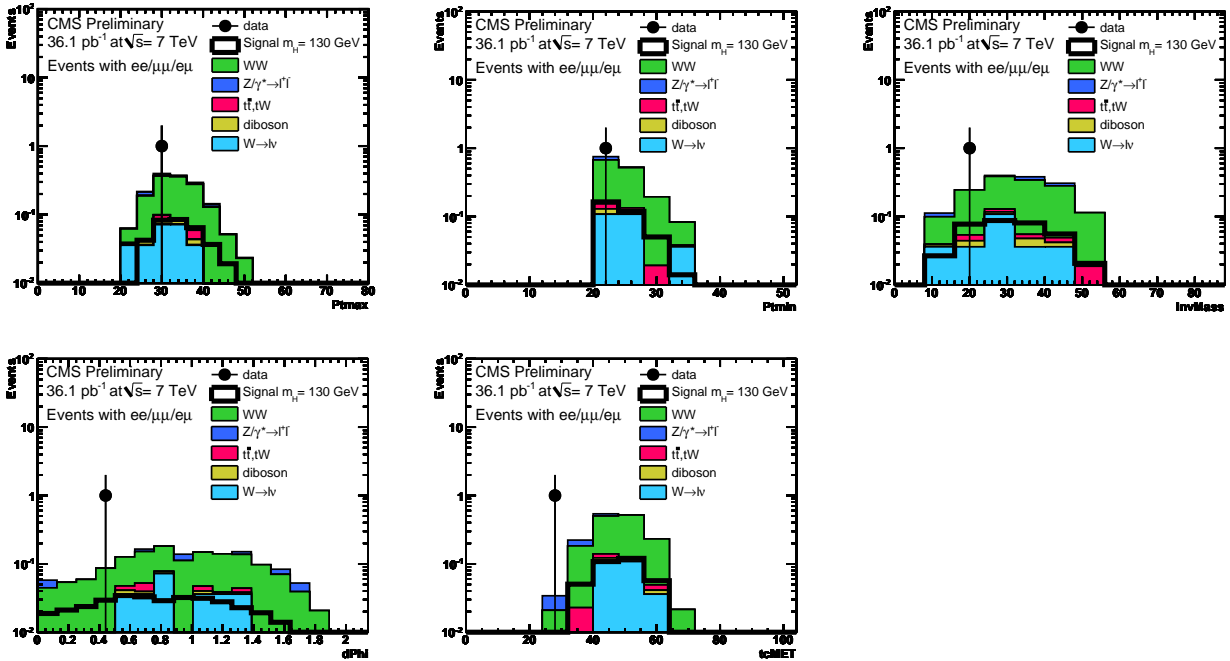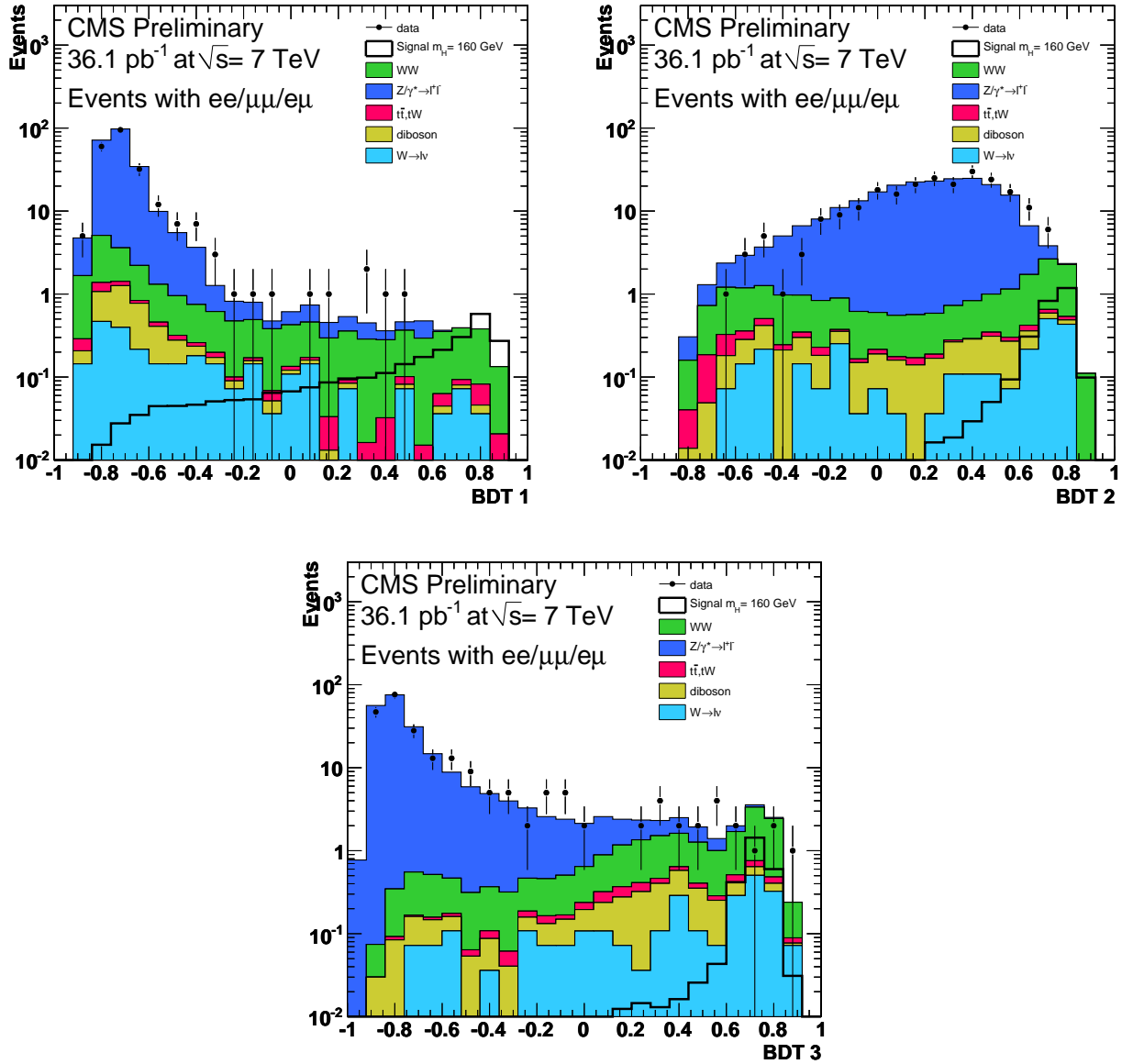

Figure 6.21: - Distributions of some of the main variables used in the MVA analysis after applying the optimal BDT cut, for a mass hypothesis of $m_H = 130 GeV/c^2$

	HWW160	WW	$t\bar{t}$	DY	VV	tW	Wjets	Total B	Data
R<0.36	1.30 ± 0.01	0.95 ± 0.01	0.04 ± 0.01	0	0.027 ± 0.002	0.027 ± 0.005	0.12 ± 0.06	1.16 ± 0.06	0

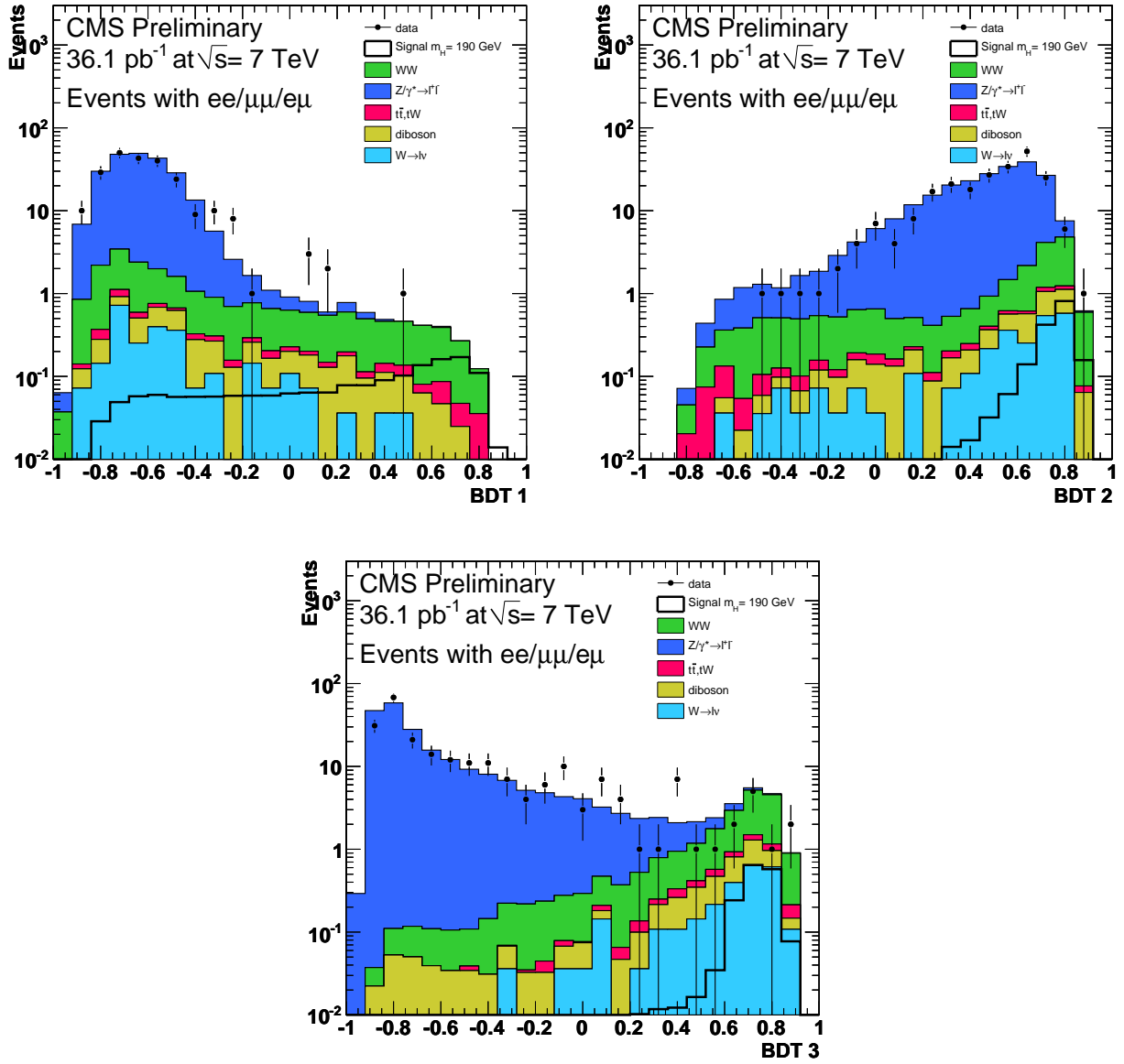**Table 6.22:** Expected number of events for $36pb^{-1}$, optimized for m_H= 160 GeV/c^2

164

Figure 6.22: Distributions of the BDT functions trained against WW, $t\bar{t}$ and Z+jets, in the m_H= 160 GeV/c^2 hypothesis.
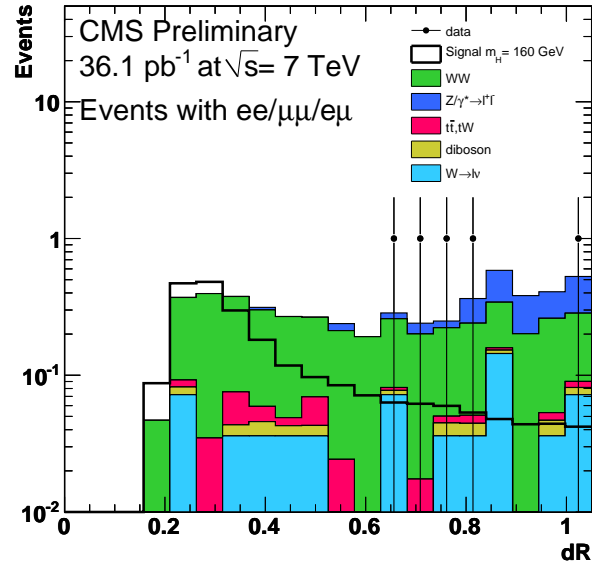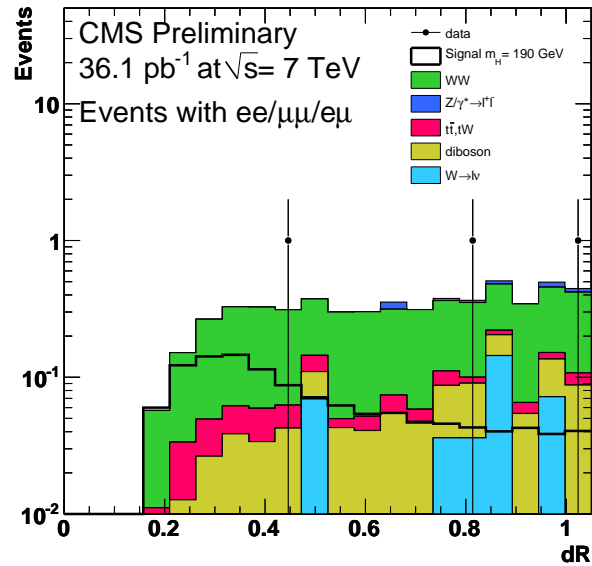
Figure 6.23: - Distribution of the final BDT variable (dR), optimized for m_H=160 GeV/c^2
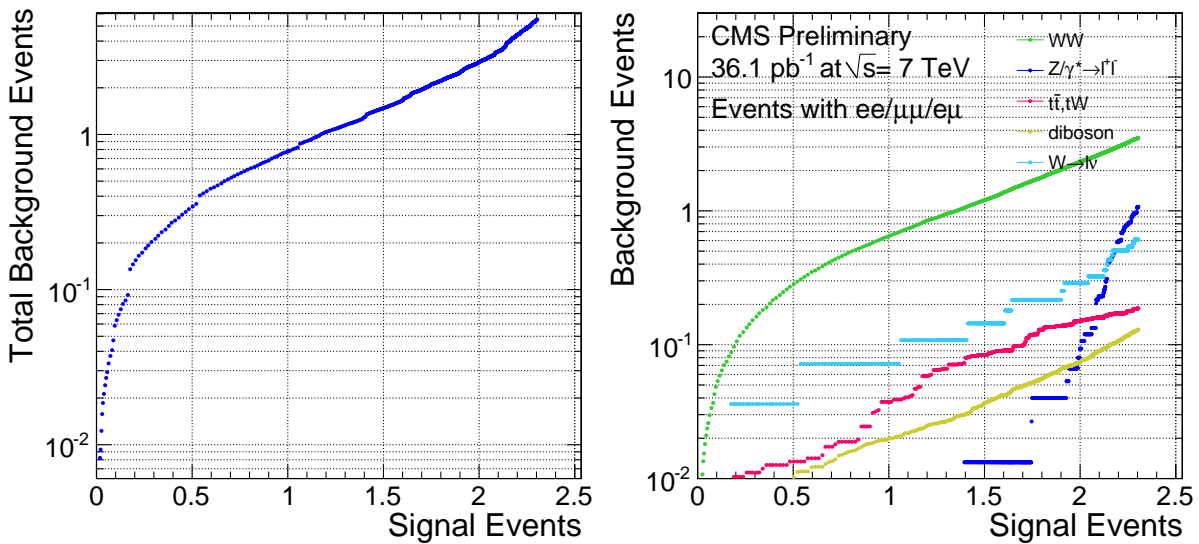
Figure 6.24: - Number of background vs signal events obtained after applying various cuts on the dR variable optimized for m_H=160 GeV/c^2

The distribution of the BDT function for the final sample before the BDT cut for $m_H = 190\ GeV/c^2$ is shown in Figure 6.25.

The final discriminant variable is represented in Figure 6.26.

The number of signal and background events after applying various cuts on the classifier is represented in Figure 6.27.

The number of events after applying the optimal cuts on the BDT function is shown in Table 6.23.

	HWW190	WW	$t\bar{t}$	DY	VV	tW	Wjets	Total B	Data
R<0.45	0.648 ± 0.007	1.07 ± 0.02	0.06 ± 0.02	0	0.146 ± 0.006	0.043 ± 0.006	0.003 ± 0.003	1.33 ± 0.03	0

Table 6.23: Expected number of events for $36pb^{-1}$, optimized for $m_H = 190\ GeV/c^2$

Figure 6.25: Distributions of the BDT functions trained against WW, $t\bar{t}$ and Z+jets, in the $m_H = 190$ GeV/c^2 hypothesis

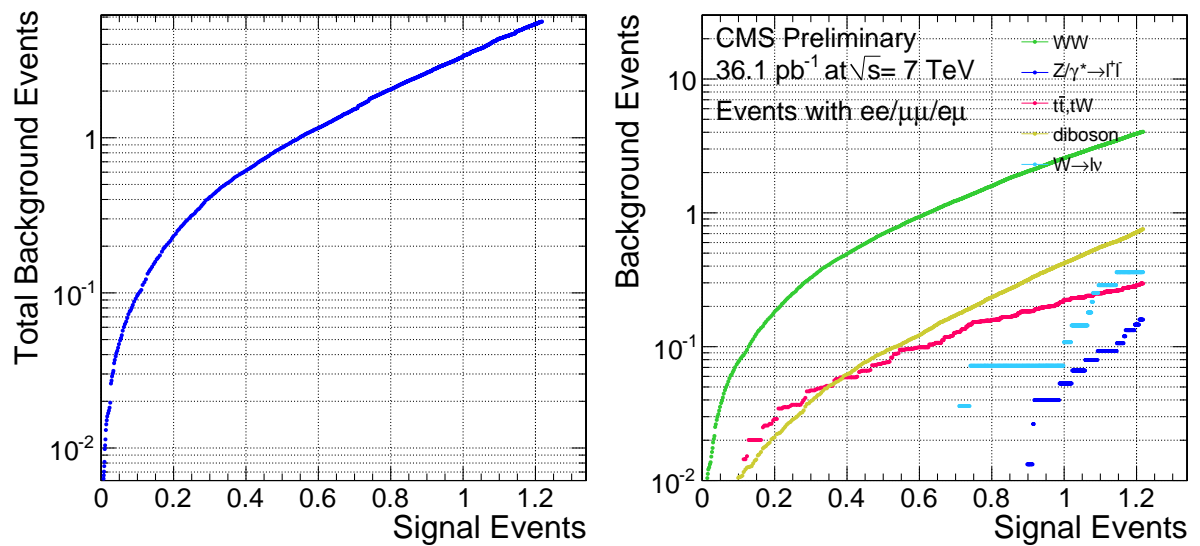**Figure 6.26:** - Distribution of the final BDT variable (dR), optimized for m_H=190 GeV/c^2

Figure 6.27: - Number of background vs signal events obtained after applying various cuts on the dR variable optimized for m_H=190 GeV/c^2

6.8.2 BDT 1D

A BDT analysis more simple has been used for this low luminosity scenario, oriented to exclude a SM Higgs boson in the context of SM with a fourth generation of fermions. In this approach, the training sample is obtained after the WW selection and an upper loose cut on m_{ll} is applied to reduce the background contamination [110].

Figure 6.28 shows the BDT function for different Higgs masses for the SM Higgs scenario and for the fourth generation fermion family scenario, respectively. The cut on the classifier output has been chosen so to have the same level of background as in the cut based analysis.
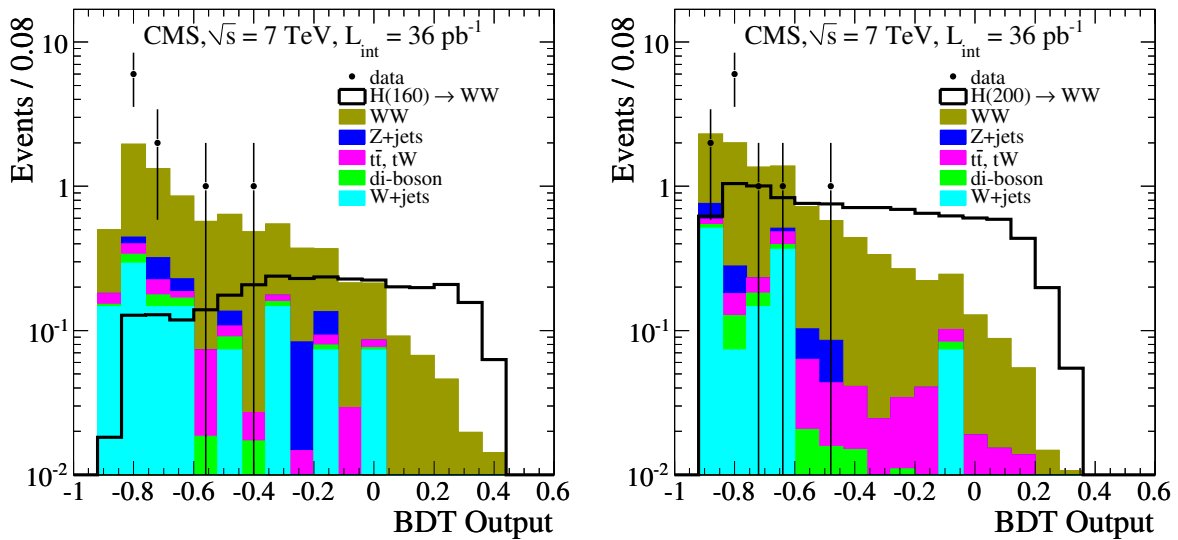

Figure 6.28: - BDT outputs for signal and backgrounds for m_H=160 (left) and m_H=200 (right). The Higgs boson event yield is normalized to the SM expectation (left) and to the fourth family scenario (right). [107]

Table 6.24 shows the number of signal and total background events for several mass hypothesis obtained in the BDT 1D analysis.

$m_H(GeV/c^2)$	Data	SM $H \to WW^*$	SM with 4th gen $H \to WW^*$	Total background
Cut-based analysis				
130	1	0.30± 0.01	1.73± 0.04	1.67± 0.10
160	0	1.23± 0.02	10.35± 0.16	0.91± 0.05
200	0	0.47± 0.01	3.94± 0.07	1.47± 0.09
210	0	0.34± 0.01	2.81± 0.07	1.49± 0.05
400	0	0.19± 0.01	0.84± 0.01	1.06± 0.03
Multivariate analysis				
130	1	0.34± 0.01	1.98± 0.04	1.32± 0.18
160	0	1.47± 0.02	12.31± 0.17	0.92± 0.10
200	0	0.57± 0.01	4.76± 0.07	1.47± 0.07
210	0	0.42± 0.01	3.47± 0.07	1.44± 0.07
400	0	0.20± 0.01	0.90± 0.01	1.09± 0.07

Table 6.24: Number of events observed in 36 pb^{-1} of data, with the signal and background predictions after the full selection, for both cut-based and BDT approaches. Uncertainties are statistical only. Correction factors and background data-driven estimation are included in these estimates, as described in Section 6.10 and 6.9 respectively

6.9 Background estimation

This is an important aspect in the $H \rightarrow WW^*$ analysis, where no mass peak can be reconstructed and thus having to rely on a counting experiment. Therefore the control of the background and the estimation from data whenever possible is essential.

This section summarizes the methods that are used to estimate the main backgrounds, most of them are similar to the ones already discussed in the $t\bar{t}$ cross section measurement (see Sections 5.10 and 5.10.2) and therefore they are only briefly described here.

6.9.1 Fake leptons from W+jets and QCD

The $W + jets$ background becomes a dangerous background to the $H \rightarrow WW^*$ signal when at least one jet or track is misidentified as a lepton, leading to the presence of two leptons in the event, one prompt lepton from the W boson decay and a fake non-prompt lepton from the hadronic decay of short lived mesons (D, B) or the decay in the detector of long lived particles(pions, kaons), or from misidentified hadrons due to, for example, electron conversions or $\pi^{\pm}\pi^0$ overlap.

These events are difficult to model accurately in simulation, so they should be measured from data. The data-driven approach followed here is the same as described for the $t\bar{t}$ cross section measurement, adapted to the characteristics of this analysis.

A set of loose lepton-like objects (fakeable objects) is defined and then the efficiency for those objects to pass the full lepton selection ("fake rate") is calculated and parameterized as functions of p_T and η . Assuming that it is sample independent, the fake rate is computed in a background-enriched dataset and then applied on the sample of interest.

The systematic uncertainty on this estimation is obtained by applying the same method to another control sample with different selection criteria. A value of 50% is derived from a closure test, where a tight-to-loose rate derived from QCD simulated events is applied to a W+ jets simulated sample to predict the rate of events with one real and one misidentified lepton.

For the WW selection, the total fake electron and fake muon contributions is found to be 1.2 $\pm$ 0.3 (stat.) $\pm$ 0.6 (syst.) and 0.5 $\pm$ 0.3 (stat.)$\pm$ 0.3 (syst.), respectively [111].

6.9.2 Drell Yan estimation

To estimate this background, the same method as described for the top cross section measurement is used. The observed number of $\mu^+\mu^-$and e^+e^-events inside the Z mass window in data, after the subtraction of the non-Z contributions, is scaled by the ratio of the number of events outside and inside the Z mass window obtained from simulation.

The ratio $R_{out/in}$ is estimated in a sample of events without the $\not{E}_T$ cut applied, Table 6.26 shows number of $Z/\gamma^* \rightarrow l^+l^-$ events measured from data compared to the events expected from simulation after the full WW selection.

There are 4 $\mu^+\mu^-$events in data observed inside the control region but no e^+e^-events.The $R_{out/in}$ value is quite flat and similar for electrons and muons, so the final $Z/\gamma^* \rightarrow l^+l^-$ contribution can be estimated in $e^+e^-$and $\mu^+\mu^-$channels together and it is found to be 0.2 $\pm$ 0.2(stats.) $\pm$ 0.2(syst.) for

e^+e^-	No MET cut	MET cut
DY MC	631 ± 4	0.06 ± 0.04
DY Data driven estimate	578 ± 8	-0.06 ± 0.06
$R_{out/in}$	0.0893 ± 0.0006	0.08 ± 0.06
$\mu^+\mu^-$		
DY MC	1053 ± 6	0.121 ± 0.06
DY Data driven estimate	1027 ± 12	0.364 ± 0.35
$R_{out/in}$	0.0985 ± 0.0006	0.133 ± 0.07

Table 6.25: Drell Yan background data driven estimation in the $\mu^+\mu^-$ and e^+e^- channels compared to simulation for $36.1pb^{-1}$ for events passing all WW selection.

e^+e^-	MET cut
DY MC	0.06 ± 0.04
DY Data driven estimate	-0.07 ± 0.07
$R_{out/in}$	0.0893 ± 0.06
$\mu^+\mu^-$	
DY MC	0.121 ± 0.06
DY Data driven estimate	0.27 ± 0.26
$R_{out/in}$	0.0985 ± 0.07

Table 6.26: Drell Yan background data driven estimation in the $\mu^+\mu^-$ and e^+e^- channels compared to simulation for $36.1pb^{-1}$ for events passing all WW selection. R taken as the value obtained without MET cut

the WW selection.

The systematic uncertainty of this method arises primarily from the dependence of $R_{out/in}$ on the $\not{E}_T$ cut and to stay conservative it is assigned the maximum variation.

6.9.3 Top background estimation

Most of the top background is rejected by the jet veto requirement. $t\bar{t}$ and tW events differ from the signal in the presence of one or two b-quarks, so to suppress and estimate the remaining top background it is needed to identify the missing b-quarks in the event.

The top background can be estimated from data by counting events with either an additional soft muon or at least one b-tagged jet with p_T below the jet veto threshold.

There are no events (in the control region) rejected by the top-veto in data after applying the full WW selection so it is only possible to put an upper limit in the contribution from top production and it is taken directly from simulation, which predicts $0.73 \pm 0.05(\text{stat.}) \pm 0.73(\text{syst.})$ events, where a 100% systematic uncertainty is assigned as a conservative estimate of the difference between data and simulation.

6.9.4 WW background

The WW production is the dominant irreducible background for the search, so it is essential to establish its properties and cross-section to control it. The cross section has been measured [107] but it cannot be directly used since the Higgs signal is also part of the measurement.

A normalization region in the phase-space is defined to measure the background rate. This region has to be close to the signal region, but still with a minimum presence of signal events. Then, the background measured in that region is used as the reference to estimate the magnitude of the background in the signal region. For that, the measured events in the background region, N_{bkg}^N, are multiplied by the ratio of the efficiencies:

$$N_{bkg}^S = \frac{\varepsilon_{bkg}^S}{\varepsilon_{bkg}^N} N_{bkg}^N \tag{6.3}$$

If the normalization region is close enough to the signal region, most of the systematic errors will cancel in the ratio of the efficiencies, giving as a result a reliable estimation of the background contamination in the signal region. The precise definition of a control region depends of the definition of the signal region.

To reject the signal contamination, a cut on the dilepton invariant mass is applied. Two different regions are defined depending on the Higgs mass hypothesis. For the low Higgs ($m_H \leq 200 \; GeV/c^2$) events are selected with $m_{ll} > 100 \; GeV/c^2$, while for the high mass Higgs ($m_H > 200 \; GeV/c^2$) events with $m_{ll} < 100 \; GeV/c^2$. With this, the statistical precision of the method is 55% and 50%, respectively.

In addition, there is a small contribution to the uncertainty from varying the PDFs, QCD renormalization and scales, since there are some differences between the low and high mass dilepton regions on all other distributions. The overall effect is about 2.8%, which is negligible in comparison with

	WW	$t\bar{t}$	tW	$Z/\gamma^* \to l^+l^-$	$W/\gamma + jets$	WZ +ZZ
$\mu^+\mu^-$	0.96± 0.02	0.05± 0.02	0.029± 0.005	0.03± 0.03	0.003± 0.003	0.048± 0.004
e^+e^-	0.63± 0.01	0.06± 0.02	0.024± 0.004	0	0.081± 0.07	0.027± 0.002
$e^\pm\mu^\mp$	2.71± 0.02	0.19± 0.03	0.085± 0.008	0.03± 0.03	0.12± 0.07	0.056± 0.004

Table 6.27: Number of events estimated in the WW control region defined for the low higgs mass region ($m_H \le 200 \; GeV/c^2$)

the statistical uncertainty at this moment.[111].

The $gg \to WW$ background contribution is taken from simulation, as there is not enough sensitivity in data to measure it.

6.9.4.1   Other backgrounds

The contribution from WZ, ZZ, W + γ, $Z/\gamma^* \to \tau^+\tau^-$ have to be estimated from Monte Carlo simulation, after applying the proper data corrections for lepton, trigger and jet veto efficiencies.

WZ and ZZ backgrounds are partially estimated from data when the two selected leptons are coming from the same Z boson pair, together with the Z+ jets process, but if the leptons are coming from different bosons, it is necessary to use the Monte Carlo expectations to estimate the contribution as it is not included. In any case, those contributions are expected to be small. WZ is rejected by requiring only two high p_T isolated leptons in the event, ZZ is also rejected by that requirement and by the $\not{E}_T$ selection which is small for this events, so this background is almost negligible after all analysis requirements.

The $W + \gamma$ background, where the γ decays into an electron-positron pair, is difficult to estimate from data, apart from several cross-checks that can be performed. For instance, applying the same standard selection, but requiring two same-sign leptons, gives a sample dominated by W + jets and W + g events. Again, we should emphasize that the expected contribution is very small, thanks to the stringent γ conversion requirements applied in this analysis.

The $Z/\gamma^* \to \tau^+\tau^-$ background, where both tau leptons decay into the $l\nu$ final state, is difficult to estimate directly from data since it does not show a resonant mass peak distribution. Nevertheless, both leptonic branching ratios and cross-section are well-known; and the process is relatively clean with two leptons in the final state. Hence, we can rely on the Monte Carlo expectations to estimate this background. In addition, the expected contribution is rather small, thanks to the $\not{E}_T$ requirements and the relatively soft leptons from the tau decays.

6.10   Signal efficiencies and systematic uncertainties

To compute the signal efficiency, several factors have to be taken into account, such as the lepton identification and isolation, jet veto and kinematic requirements, in addition to the acceptance of having both leptons in the fiducial region. Simulation is used to predict those efficiency, but applying the proper Monte Carlo to data corrections.

For the lepton selection and jet veto efficiencies, $Z/\gamma^* \to l^+l^-$ events are used as a control sample

to study the difference between data and simulation, and to derive data to simulation scale factors. These scale factors are later used to either correct the signal efficiency in simulation or to provide the systematic uncertainties.

6.10.1 Lepton identification and trigger efficiencies

The lepton and trigger selection efficiencies are measured using the Tag and Probe method, in the same way as done for the $t\bar{t}$ cross section measurement.

For electrons with $p_\mathrm{T} > 20$ GeV/c, the efficiency scale factors obtained are 0.969 ± 0.019 and 0.992 ± 0.026, for the barrel and endcap, respectively. Also a 1.4% uncertainty is assigned on the electron reconstruction efficiency.

For muons with $p_\mathrm{T} > 20$ GeV/c, the total efficiency scale factor obtained is consistent with 1 with an uncertainty of 1 % . The trigger efficiency is assumed to be 100% , with an uncertainty of 1.5% .

6.10.2 Jet Veto selection efficiency

The jet veto efficiency and its systematic uncertainties is estimated in a data-driven way using $Z/\gamma^* \to l^+l^-$ events.

Simulated events are used to predict the $H \to WW^*$ signal efficiency and $Z/\gamma^* \to l^+l^-$ events are used to study the data-to simulation efficiency scale factors of the lepton selection and jet veto requirement.

The $H \to WW^*$ jet veto efficiency in data is estimated as the value obtained from simulation, multiplied by a data over simulation scale factor, which is derived from $Z/\gamma^* \to l^+l^-$ events.

$$\varepsilon^{data}_{H\to WW^*} = \varepsilon^{data}_{Z} \cdot \left(\frac{\varepsilon^{MC}_{H\to WW^*}}{\varepsilon^{MC}_{Z}}\right) \tag{6.4}$$

The jet veto efficiency uncertainty of $H \to WW^*$ can be factorized into the Z efficiency uncertainty in data and HWW/Z efficiency ratio uncertainty in simulation. The former is statistically dominated, while theoretical uncertainties due to higher order corrections contribute most to the HWW/Z efficiency ratio uncertainties.

Due to details in the implementation of the POWHEG[90] calculation, the resulting Higgs boson p_T spectrum is harder than the most precise spectrum calculated to NNLO with resummation to next-to-next-leading-log (NNLL) order. Therefore, the Higgs boson p_T distribution is reweighted in POWHEG to match the NNLO+NNLL prediction. The signal efficiency, estimated after this reweighting, is 14% larger than that from uncorrected POWHEG calculations, and it is independent of the Higgs boson mass. This effect is expected since harder p_T spectrum of Higgs is associated with more initial state radiation, which makes the jet veto efficiency lower.

6.10.3 Systematic uncertainties

The main sources of systematic error considered in the analysis are the following:

- Luminosity. The uncertainty coming from the luminosity measurement is one of the largest sources of systematic uncertainty, and it is 11% .

Source	$H \to WW^*$	qqWW	ggWW	WZ/ZZ non Z resonant	top	$Z/\gamma^* \to l^+l^-$	W+ jets	$V(W/Z) + \gamma$
Luminosity	11	–	–	11	–	–	–	11
Trigger efficiency	1.5	1.5	1.5	1.5	–	–	–	1.5
Muon efficiency	0.7	0.7	0.7	0.7	–	–	–	0.7
Elec. id efficiency	2.4	2.4	2.4	2.4	–	–	–	2.4
Reco. efficiency	1.4	1.4	1.4	1.4	–	–	–	1.4
momentum scale	1.3	1.3	1.3	1.5	–	–	–	2
Pile-up	0.5	0.5	0.5	0.5	–	–	–	0.5
$\not{E}_T$ resolution	1.0	1.0	1.0	1.0	1.0	3.0	–	1.0
Jet veto	6.9	–	5.4	5.4	–	–	–	5.4
PDF uncertainties	3.0	2.6	–	2	–	–	–	4
NLO effects	2.0	1.1	–	3.5	–	–	–	10
Fakes	–	–	–	–	–	–	50	–
WZ/ZZ cross-section	–	–	–	3.0	–	–	–	–
qq WW norm.	–	55	–	–	–	–	–	–
gg WW norm.	–	–	50	–	–	–	–	–
tX norm.	–	–	–	–	100	–	–	–
DY norm.	–	–	–	–	–	100	–	–
statistics	1	1	1	4	6	50	30	12

Table 6.28: Summary of all sources of systematic uncertainties considered in the analysis

- Lepton scale. Due to several factors, the energy scale for electrons and the momentum scale for muons have a large uncertainty for the current data reconstruction version, so an additional systematic uncertainty have been estimated by varying the transverse momentum of the muons by 1% and 2% and 4% for electrons in the barrel and the endcap, respectively. The contribution to the uncertainty on the signal efficiency is only about 1.3% .

- Pile-up. The effect of the pile-up has also been evaluated by reweighting the Monte Carlo simulation to match the number of reconstructed vertices found in data. The effect for events with real $\not{E}_T$ is relatively small, just about 0.5 % .

- Uncertainties on PDFs are considered and contribute to the signal efficiency uncertainty on about 3 % . We have also computed the change of the acceptance of having two lepton in the fiducial region when varying renormalization and factorization QCD scales. This uncertainty is found to be about 2% .

- $\not{E}_T$ resolution. Possible resolution effects in the $\not{E}_T$ related variables are studied, giving a contribution to the signal efficiency uncertainty of about 1% .

- The limited Monte Carlo statistics is also considered.

A summary of all systematic uncertainties taken into account is shown in Table 6.28. The total uncertainty depends on the Higgs mass case, it is about 40% on the background estimation while it is about 14% on the signal efficiency.

6.11 Results

A large variety of methods exists to express the statistical significance of a possible signal in the presence of background. Two statistical methods have been used to estimate an upper limit on the Higgs boson production cross section times branching ratio. Both methods use the same likelihood function: the expected value of the observed number of events is modeled as a Poisson random variable, whose expected value is the sum of contributions from the various signal and background processes.

The first method is based on Bayesian inference [112], while the second method, known as CLs, is based on the hybrid Frequentist-Bayesian approach [113]. Both methods account for systematic uncertainties, and provide similar values for the upper limits.

No excess compatible with Higgs signal production is found in the current data sample. Therefore, 95% confidence level cross-section upper limits are obtained in the 120-600 GeV/c^2 mass region.

The $\sigma_H \times BR(H \rightarrow WW^* \rightarrow 2l2\nu)$ upper limits are about three times larger than the SM expectation for $m_H = 160\ GeV/c^2$, with the current amount of data, the observed limits have no sensitivity to the SM Higgs boson. When compared to the recent theoretical calculations performed in the context of a SM extension by a sequential fourth family of fermions with very high masses [114], the results of BDT analysis exclude at 95% C.L. a Higgs boson with mass in the range from 144 to 207 GeV, where the mean expected exclusion at 95% C.L. is in the range from 147 to 193 GeV/c^2. Similar results are achieved using the cut-based approach.

Figure 6.29 shows the 95% mean expected and observed C.L. upper limits on the $\sigma_H \times BR(H \rightarrow WW^* \rightarrow 2l2\nu)$ for Higgs boson masses between 120 and 600 GeV/c^2 using the cut-based and BDT analysis.
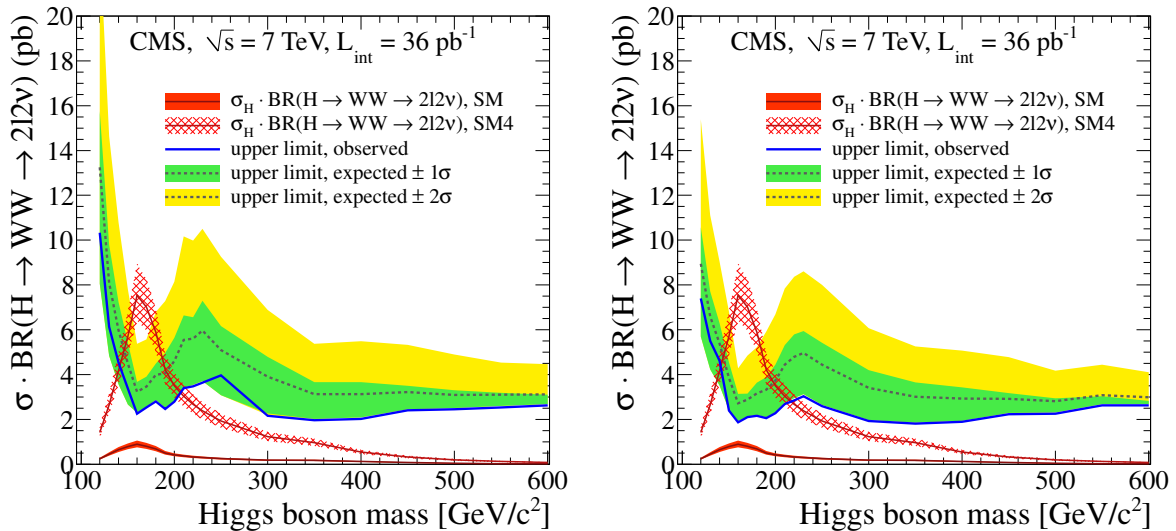

Figure 6.29: - 95% mean expected and observed C.L. upper limits on the cross section times BR for masses in the range 120-600 GeV/c^2 using (a) cut-based and (b) multivariate BDT event selections. [107]. Results are obtained using a Bayesian approach.

$m_H(GeV/c^2)$	Lim.obs cut-based	Lim.exp cut-based	Lim. obs BDT-based	Lim.exp. BDT-based
130	6.30	8.07	5.66	6.57
160	2.29	3.22	1.93	2.72
200	2.80	4.59	2.32	3.72
210	3.41	5.53	2.76	4.43
400	2.08	3.12	1.94	2.93

Table 6.29: 95% observed and mean expected C.L. upper limits on the cross section $\sigma_H \times BR(H \rightarrow WW^* \rightarrow 2l2\nu)$ for four Higgs masses. The results of the cut-based and the multivariate based event selections are obtained using a Bayesian approach.

6.12 Perspectives

After the Winter shutdown, the LHC has re-started its operations in March 2011 and is performing really well, it is expected that the experiments can collect about 1 fb^{-1} of data before the Summer. The sensitivity to Higgs boson for this luminosity has been evaluated [115].

Figure 6.30 shows the combined exclusion limits for different center of mass energies and integrated luminosities, using the analysis presented in this work. From the plot, one can see that with 1 fb^{-1} at 7 TeV, a SM Higgs boson could be excluded with a mass from 135 to 450 GeV/c^2. With 5 fb^{-1} of data, CMS should reach an exclusion sensitivity for the SM Higgs from the current LEP limit ($m_H > 114 \ GeV/c^2$) to approximately 600 GeV/c^2.

Figure 6.31 shows the expected statistical significance in presence of the Higgs bosons for different center of mass energies and integrated luminosities. As can be observed, with 1 fb^{-1} of fata at 7 TeV, an excess of events with more than 3σ significance in the mass range between 145 and 195 GeV/c^2 could be seen if the SM Higgs boson mass is in that window.
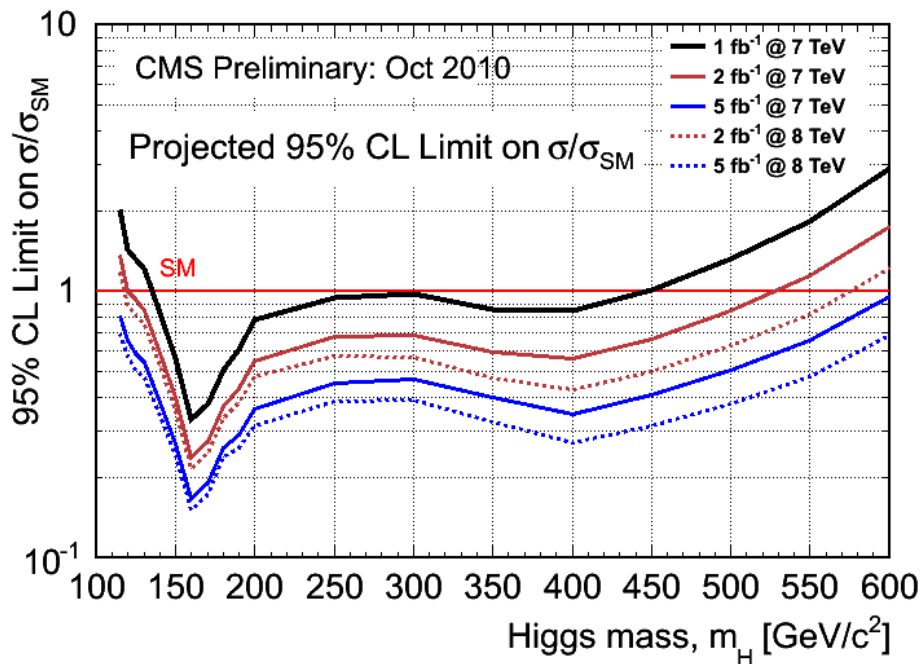
Figure 6.30: - Projected exclusion limits at 7 and 8 TeV and a few benchmark luminosities
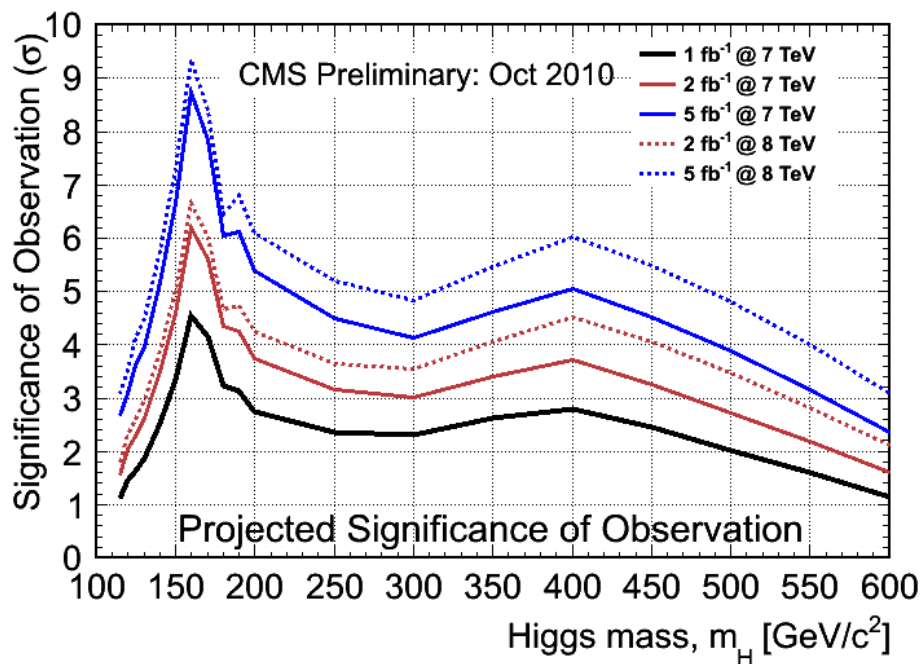

Figure 6.31: - Projected significance at 7 and 8 TeV and a few benchmark luminosities

7

Conclusions

The measurement of the $t\bar{t}$ cross section, as well as the search for the SM Higgs boson in $H \rightarrow WW^*$ decay mode, in the dilepton final states, for the 36 pb^{-1} of data collected during the 2010 run of the LHC has been presented. This amount of data, allowed the improvement of the level of understanding of the detector and the development of several SM measurements and searches for many new particles and phenomena.

The observation of the top quark was a major goal of the CMS experiment, as it requires a good understanding and reconstruction of almost all the relevant observables in an event produced in pp collisions: leptons (e,μ), $\not{E}_T$ and b-tagged jets. The presence of two high p_T leptons, significant amount of $\not{E}_T$ and at least two hight p_T jets coming from the hadronization of b-quarks allowed us to achieve a quite pure selection of events. Already with the first 3.1 $\pm$ 0.3 pb^{-1} of data, collected up to August 2010, it has been possible not only to observe the top quark but to measure for the first time at the LHC the $t\bar{t}$ cross section in the dilepton channel, applying the analysis selection and all the data-driven techniques that were developed using simulated samples during the previous years of the work presented in this thesis. After applying the full selection, 11 events were observed in data (3 e^+e^-, 3 $\mu^+\mu^-$ and 5 $e^\pm\mu^\mp$), with 2.0$\pm$ 1.0 expected background events. Backgrounds from Drell-Yan and non-W/Z boson production were estimated from data, while the remaining backgrounds were taken from simulation. Several sources of systematic uncertainty were studied, obtaining an overall uncertainty of 6.4% for the signal and 11% for the background. The measured cross section is 194 $\pm$ 72 (stat.) $\pm$ 24 (syst.) $\pm$ 21 (lumi.) pb [79], which is consistent with the next-to-leading order theoretical predictions. To further confirm that the observed events were really from top quark production, the property that the jets from $t\bar{t}$ events come from the hadronization of b-quarks was exploited using the b-tagging technique. The number of jets satisfying this b-tagging criteria in events passing the analysis selection is shown in Figure 7.1, providing additional confidence on the top-likeliness of the selected candidates.

The full 2010 dataset was further analyzed in order to perform a more precise measurement of this $t\bar{t}$ cross-section. Given that the b-tagging was understood since the beginning and it showed excellent performance already with such a low luminosity, surpassing the expectations, this tool was used to obtain a pure sample of top events.

After applying similar analysis requirements as in the low luminosity measurement [79] and asking in addition for at least one of the two jets to be b-tagged, 90 candidates were observed in data (15 e^+e^-, 24 $\mu^+\mu^-$ and 51 $e^\pm\mu^\mp$), with 8 $\pm$ 2 expected background events, estimated using updated data-driven techniques (see Table 7.1 and Figure 7.2).
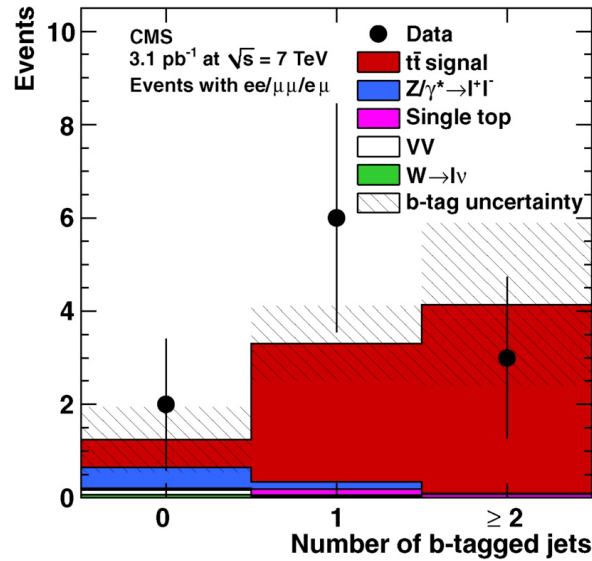
7. CONCLUSIONS

Figure 7.1: - Number of b-tagged jets in events passing all dilepton selection criteria for all three dilepton modes combined in 3.1 pb^{-1}, compared to signal and background predictions. The hatched bands reflect the expected uncertainties on the b-tag efficiency for signal events.

Final state	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$
Events in data	15	24	51
All backgrounds	1.55 ± 0.7	3.56 ± 1.8	3.2 ± 0.7
Total acceptance, %	0.235 ± 0.021	0.304 ± 0.026	0.858 ± 0.071
Cross section, pb	$159.4 \pm 46 \pm 17 \pm 6$	$187.3 \pm 45 \pm 23 \pm 8$	$155.2 \pm 23 \pm 13 \pm 6$

Table 7.1: Summary of $t\bar{t}$ signal production cross section measured in separate event selection channels. The number of events observed in data, the total background expectations, the total signal acceptance (includes geometric detector acceptance and event selection efficiency) with systematic uncertainties, and the cross section measurements are shown separately. The uncertainties on the cross section include statistical, systematic, and luminosity normalization uncertainties.

All possible sources of systematic uncertainties were studied: modeling of the lepton reconstruction and selection, jet energy scale and jet energy resolution, imperfect understanding of b-tag and mistag efficiencies, uncertainties from the methods used for the background estimation and theoretical uncertainties coming from the generator parameters that described the process kinematics and hadronization. The largest systematic uncertainty on the signal selection was determined to come from the b-tagging and was estimated to be 5% .

For the final cross-section measurement, several analysis applying slightly different selections were combined: at least 2 jets with at least one b-tagged jet (described in this thesis), at least 2 jets without any b-tagging requirement and the 1-jet analysis. This led to the value : $\sigma_{t\bar{t}} = 168 \pm 18$(stats.)$\pm 14$(sys.) ± 7(lumi.) [82], consistent with the NLO cross section prediction of 158^{+23}_{-24} pb or the approximate NNLO calculation of $163^{+7}_{-5} \pm 9$ pb [116].

Figure 7.3 shows the $t\bar{t}$ cross section as a function of the center-of-mass energy $\sqrt{s}$ of the collision, comparing measurements made at the Tevatron $p\bar{p}$ collider [101; 102],[103; 104] with the LHC measurements[98][82][100].
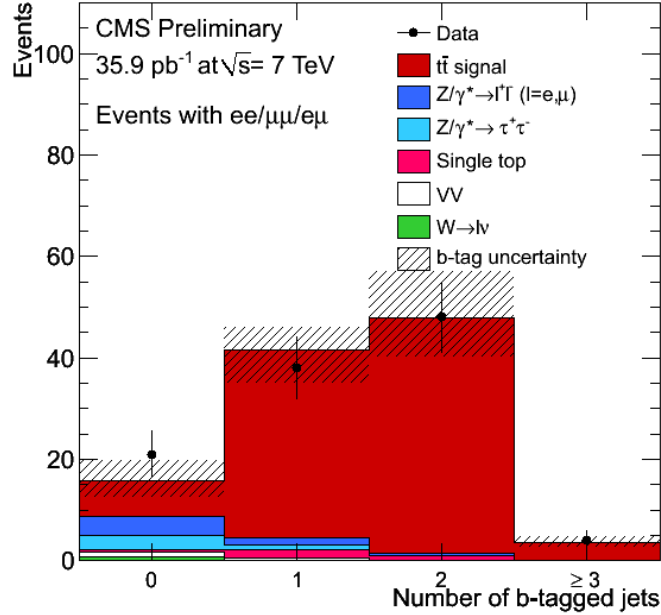

Figure 7.2: - Number of b-tagged jets in events passing all dilepton selection criteria for all three dilepton modes combined in 35.9 pb^{-1}, compared to signal and background predictions. The hatched bands reflect the expected uncertainties on the b-tag efficiency for signal events.

Once this important process predicted within the SM was re-established, the searches for evidence of new physics constitute the main goal of the CMS physics program, and the search for the SM Higgs boson is one of the major goals.

In this work we searched for the $H \rightarrow WW^* \rightarrow 2l2\nu$ final state that is the most promising in a wide range of Higgs boson masses around 160 GeV/c^2, where it decays almost exclusively to two W's and the cross section is high. The search was performed using the full 2010 dataset, applying the analysis previously developed in simulation. The integrated luminosity was not yet large enough to be sensitive to the SM Higgs boson but it was possible to set limits in the context of a SM extension with a fourth generation of fermions [114].

We analyzed the full 2010 data and, as in the case of the $t\bar{t}$ cross-section measurement, we developed methods to estimate the contribution of the different background processes WW, $t\bar{t}$, $Z/\gamma^* \rightarrow l^+l^-$ and $W + jets$. In the context of the WW electroweak cross-section measurement, we have calculated the following values for the most important backgrounds using data-driven techniques: $1.7 \pm 0.4 \pm 0.7$ W+jets events and $0.2 \pm 0.2 \pm 0.3$ $Z/\gamma^* \rightarrow l^+l^-$ events, which are then scaled according to the $H \rightarrow WW^* \rightarrow 2l2\nu$ analysis selection efficiencies, as there are no events remaining in the control regions after the full selection is applied.

In addition to the analysis based on sequential cuts applied on the main observables, a multivariate analysis, based on the use of Boosted Decision Trees, was developed to improve further the sensitivity. Additional variables were used to improve the classifier performance, such as the trans-
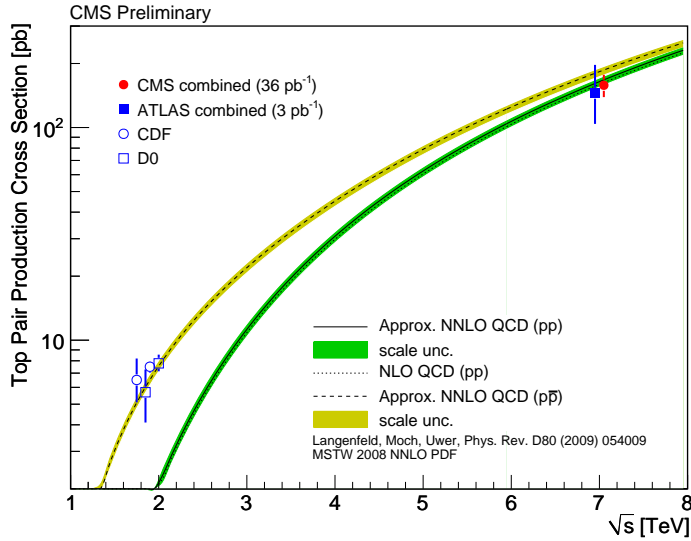
7. CONCLUSIONS

Figure 7.3: - Top pair production cross section as a function of $\sqrt{s}$, for both $p\bar{p}$ and pp collisions [98]. Tevatron measurements made at $\sqrt{s} = 1.8$ TeV are taken from [101; 102], while those made at $\sqrt{s} = 1.96$ TeV are taken from [103; 104]. The CMS combined measurement is shown, as well as the ATLAS measurement from [100]. Data points are slightly displaced horizontally for better visibility. Theory predictions at approximate NNLO are obtained using [105; 106]. The error band of the prediction corresponds to the scale uncertainty.

verse mass or the angle between the leptons and the $\not{E}_T$. Both approaches were optimized for the Higgs boson mass hypotheses considered, to properly account for the different kinematics of the signal.

Using either of the two analysis, no excess compatible with Higgs signal production was found in the data sample and 95% confidence level cross-section upper limits were derived in the 120-600 GeV/c^2 mass region (see Table 7.2 and Figure 7.4) [107].

$m_H(GeV/c^2)$	Lim.obs cut-based	Lim.exp cut-based	Lim. obs BDT-based	Lim.exp. BDT-based
130	6.30	8.07	5.66	6.57
160	2.29	3.22	1.93	2.72
200	2.80	4.59	2.32	3.72
210	3.41	5.53	2.76	4.43
400	2.08	3.12	1.94	2.93

Table 7.2: 95% observed and mean expected C.L. upper limits on the cross section $\sigma_H \times BR(H \to WW^* \to 2l2\nu)$ for four Higgs masses. The results of the cut-based and the multivariate based event selections are obtained using a Bayesian approach.
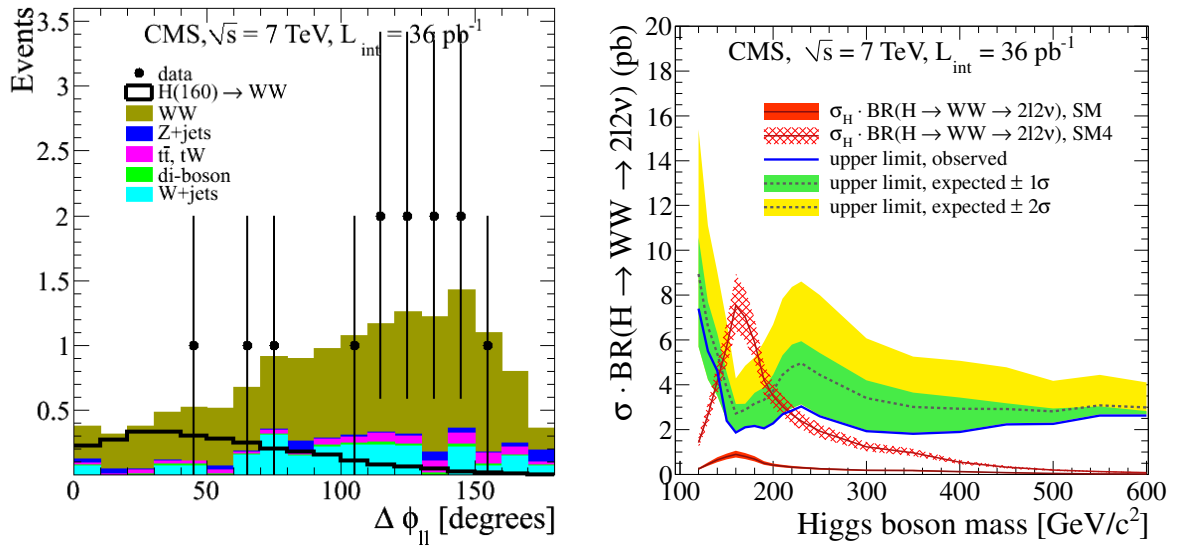
Figure 7.4: - Left:Azimuthal angular separation between the two selected leptons after WW selection, for $m_H = 160 \ GeV/c^2$ SM Higgs signal and for backgrounds. Right: 95% mean expected and observed C.L. upper limits on the cross section times BR for masses in the range 120-600 GeV/c^2 using multivariate BDT event selections. Results are obtained using a Bayesian approach. [107]

7. CONCLUSIONS

8

Resumen del trabajo realizado

El Modelo Estándar(SM) describe el comportamiento de todas las partículas subatómicas conocidas, explicando con éxito todos los datos de colisiones de altas energías. Sin embargo, todavía quedan cuestiones por resolver en este modelo, como la naturaleza del mecanismo responsable de la ruptura de la simetría electrodébil y el orígen de la masa.

El mecanismo de Higgs proporciona una de las soluciones más simples al problema, requiere la existencia de una nueva partícula escalar, el bosón de Higgs, que todavía no se ha detectado. Todas las propiedades de esta partícula se pueden predecir dentro del Modelo Estándard excepto su masa, que es un parámetro libre de la teoría. Después del descubrimiento del quark top en 1995, el bosón de Higgs es la única pieza que falta en este modelo.

La búsqueda del bosón de Higgs se ha llevado a cabo en diferentes experimentos,y, hasta el momento, se han puesto límites a su masa. Las búsquedas en el colisionador e^+e^- LEP proporcionaron un límite inferior a la masa de 114.4 GeV/c^2 a un nivel de confianza del 95% [1]. Actualmente, los experimentos CDF y D0 en el colisionador $p\bar{p}$ Tevatron siguen realizando búsquedas de Higgs y han alcanzado suficiente sensibilidad para excluir un bosón de Higgs del SM con una masa comprendida entre aproximadamente 158 and 175 GeV/c^2 a un nivel de confianza del 95% C.L [2]. Otras medidas indirectas, como las derivadas de medidas electrodébiles de precisión ponen un líimite superior en 185 GeV/c^2 al 95% C.L. [3]. Por tanto, todos los límites experimentales y teóricos obtenidos hasta el momento favorecen un bosón de Higgs ligero. Todas estas medidas contribuyen, además, a orientar las búsquedas en el LHC.

El LHC es un colisionador pp, diseñado para producir colisiones a una energía en centro de masas de $\sqrt{s} = 14$ TeV, siete veces superior a la del Tevatron, lo que permitirá estudiar un rango amplio de procesos, tanto dentro del Modelo Estándar como evidencias de física más allá de este Modelo. Por tanto, si el mecanismo de Higgs es válido los experimentos del LHC tienen el potencial para encontrar el bosón de Higgs, permitiendo concluir la búsqueda llevada a cabo durante muchos años en diferentes colisionadores.

El trabajo presentado en esta tesis doctoral se centra en la medida de la sección eficaz de producción de pares de quarks top-antitop y en la búsqueda del bosón de Higgs del SM en el modo de desintegración WW, ambos procesos en el estado final dileptónico, con los datos recogidos durante el año 2010 en colisiones pp a $\sqrt{s} = 7$ TeV, lo que corresponde a una luminosidad integrada de 36 pb^{-1}. Todo esto ha sido posible gracias a las excelentes prestaciones y funcionamiento del detector CMS durante el año 2010. Muy rápidamente después del comienzo de la toma de datos ha sido posible re-

alizar diferentes estudios de procesos predichos por el Modelo Estándard y ,a medida que aumentaba la luminosidad, realizar medidas más detalladas. La Figura 8.1 muestra la reconstrucción realizada por CMS de la mayoría de las resonancias conocidas en los canales e^+e^- y $\mu^+\mu^-$.
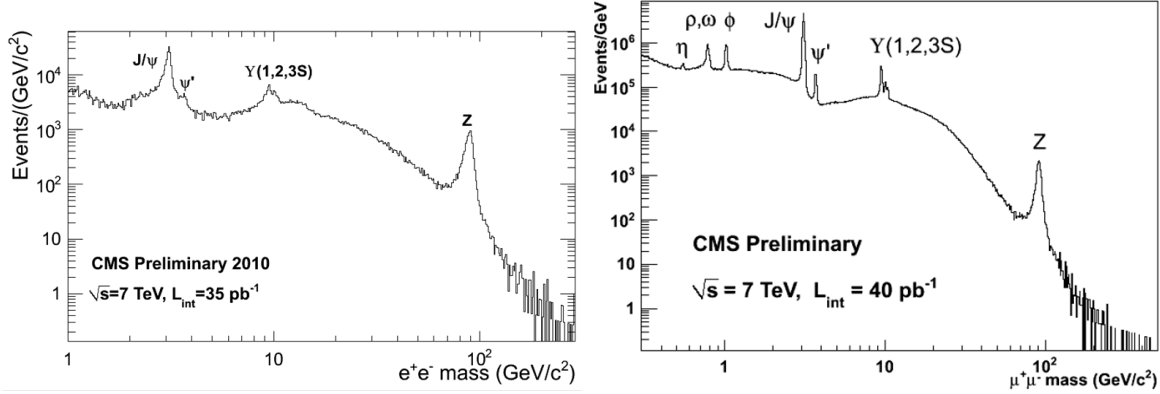

Figure 8.1: - Reconstrucción de las resonancias conocidas en los canales e^+e^- (izquierda) y $\mu^+\mu^-$ (derecha)

La observación del quark top fue uno de los objetivos principales del experimento CMS, pues requiere entender adecuadamente la reconstrucción de la mayoría de los observables físicos que aparecen en sucesos producidos en colisiones pp, como la reconstrucción de leptones(e,μ), la medida de la energía faltante ($E\!\!\!/_T$) y el etiquetado de quarks b. En el LHC, los pares de quarks top se producen principalmente mediante fusión de gluones, mientras que en el Tevatron, lo hacen mediante procesos de aniquilación de de quarks $q\bar{q}$, por lo tanto, su estudio en el LHC contribuye también al entendimiento de su mecanismo de producción.

La medida de la sección eficaz de pares de quarks top-antitop fue realizada en el canal dileptónico, un estado final muy limpio, que permite observar la señal en una fase más temprana. La presencia de dos leptones de alto p_{T} , alto valor de $E\!\!\!/_T$ y de dos jets de alto p_{T} , que la mayor parte de las veces provienen de la hadronización de quarks b, ha permitido obtener una selección muy pura de sucesos con quarks top.

Utilizando incluso los primeros 3 pb^{-1} de datos recogidos hasta Agosto de 2010, ha sido posible con el trabajo desarrollado en esta tesis, no solamente observar el quark top sino medir la sección eficaz de producción de pares de quark top-antitop por primera vez en el acelerador LHC. Tras aplicar la selección completa, se observaron 11 sucesos en los datos (3 e^+e^-, 3 $\mu^+\mu^-$ and 5 $e^{\pm}\mu^{\mp}$) y se esperaban aproximadamente 2.0± 1.0 sucesos de fondo. Los fondos principales que provienen de procesos de Drell-Yan, así como aquellos en los que no hay una producción de W y Z se estimaron directamente a partir de los datos, mientras que el resto de los fondos, provientes de la producción de dibosones, *single-top* y $Z/\gamma^* \to \tau^+\tau^-$ se estimaron mediante simulación.

Se estudiaron diferentes fuentes de incertidumbre sistemática, obteniendo un valor global de 6.4% para la señal y 11% para el fondo total. La sección eficaz de este proceso se ha determinado, obteniendo un valor de 194 ± 72 (stat.) ± 24 (syst.) ± 21 (lumi.) pb [79], que resulta consistente con el valor predicho por el SM hasta NLO. Para poder dar una confirmación adicional respecto a que los sucesos observados procedían de la desintegración de un quark top, se ha utilizado la propiedad de que en principio los jets que provienen de este quark se producen tras la hadronización de quarks b. Esta propiedad ha sido explotada utilizando técnicas de etiquetado de quarks b desarrolladas en CMS.

El número de jets que satisfacen este criterio de etiquetado de quarks b en sucesos que pasan la selección completa se muestra en Figura 8.2 y proporciona confianza añadida en el hecho de los sucesos considerados provienen de la desintegración de quarks top.
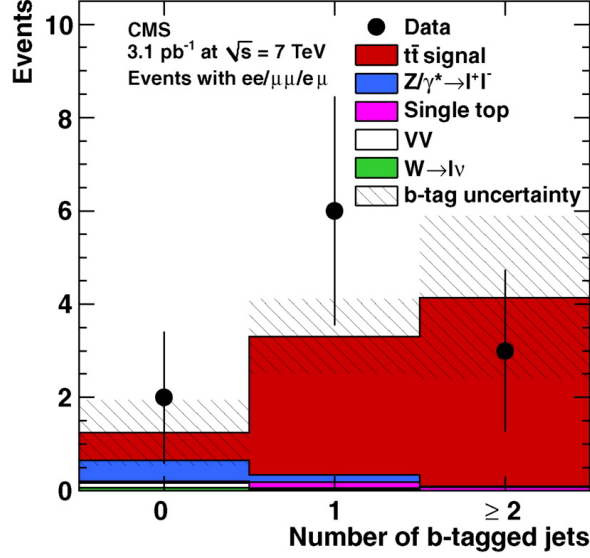

Figure 8.2: - Número de jets etiquetados como b-jets en sucesos que pasan la selección de diletpones para los tres estados finales combinados recogidos para una luminosidad de 3.1 pb^{-1}, comparados con las predicciones de señal y fondo. Las bandas rayadas reflejan las incertidumbres esperadas en la eficiencia de etiquetado de quarks b para los sucesos de señal.

El conjunto completo de datos recogidos en 2010 ha sido utilizado para realizar una medida más precisa de la sección eficaz. En este caso, fue posible utilizar el etiquetado de quarks b en la obtención de la muestra final de candidatos, pues este complejo proceso de etiquetado fue adecuadamente entendido con esta luminosidad.

Tras aplicar un análisis de caracteristicas similares al realizado en la publicación [79], requiriendo la presencia de al menos uno de los jets como b-jets, se observaron 90 candidatos (15 e^+e^-, 24 $\mu^+\mu^-$ and 51 $e^\pm\mu^\mp$) , esperando 8 $\pm$ 2 sucesos de fondo que fueron estimados utilizando igualmente técnicas basadas en datos. (Ver Tabla 8.1 y Figura 8.3).

Final state	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$
Events in data	15	24	51
All backgrounds	1.55 ± 0.7	3.56 ± 1.8	3.2 ± 0.7
Total acceptance, %	0.235 ± 0.021	0.304 ± 0.026	0.858 ± 0.071
Cross section, pb	$159.4 \pm 46 \pm 17 \pm 6$	$187.3 \pm 45 \pm 23 \pm 8$	$155.2 \pm 23 \pm 13 \pm 6$

Table 8.1: Resumen de la sección eficaz medida en los tres canales $t\bar{t}$. Se muestran separadamente el número de sucesos observados, el fondo esperado, la eficiencia para la señal con su correspondiente incertidumbre sistemática. Se consideran incertidumbres sistematicas, estadisticas y las derivadas de la luminosidad.

Se estudiaron todas las posibles fuentes de incertidumbre sistemática, modelado de la recon-

strucción de leptones, escala y resolución de energia de los jets, etiquetado de quarks b, y *mistags*, incertidumbres en la estimación del fondo asi como incertidumbres teóricas provinientes de parámetros de los generadores utilizados y de los derivados del proceso de hadronización. El valor mayor proviene de la incertidumbre en la selección de la señal debida al etiquetado de quarks b, y ha sido estimada en torno al 5% .

En este caso, para la determinación de la sección eficaz final, se utilizaron diferentes estados finales: presencia de al menos 2 jets y al menos uno etiquetado como b-jet, que es el descrito en esta tesis, al menos dos jets, y sucesos con 1 jet, lo que produjo un valor de $\sigma_{t\bar{t}} = 168 \pm 18(stats.) \pm 14(sys.) \pm 7(lumi)$ [82], consistente con la predicción a NLO de 158^{+23}_{-24} pb o con el cálculo NNLO de $163^{+7}_{-5} \pm 9$ pb [116].

La Figura 8.4 muestra la sección eficaz de producción de pares top-antitop en función de la energía en centro de masas de la colisión, comparando medidas realizadas en el colisionador $p\bar{p}$ Tevatron [101; 102],[103; 104] con las medidas realizadas con colisiones pp en el LHC [98][82][100].
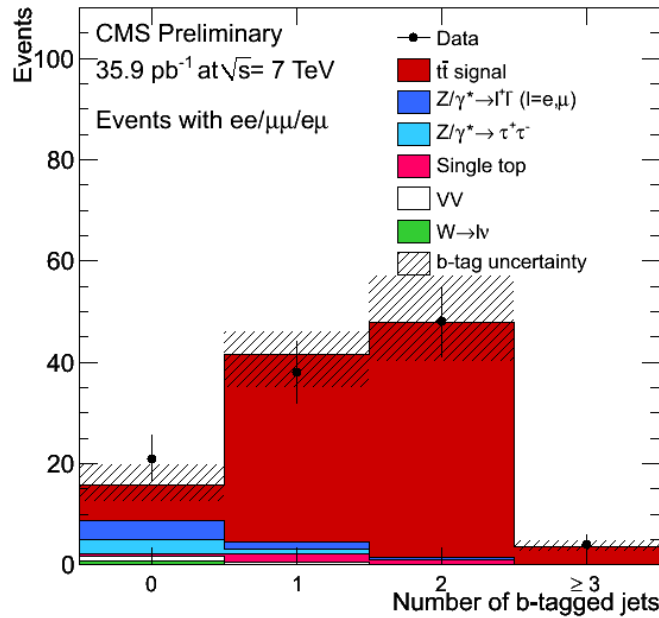

Figure 8.3:   - Número de b-jets en sucesos que pasan toda la selección de dileptones para los tres modos combinados, 35.9 pb^{-1}, comparada con las predicciones de señal y fondo. La zona rayada refleja las incertidumbres esperadas en la eficiencia de etiquetado de quarks b para los sucesos de señal.

Una vez confirmadas las predicciones del SM en CMS, el objetivo fundamental del programa de Fisica del experimento lo constituye la búsqueda de nuevas partículas y más concretamente la del bosón de Higgs del SM.

En este trabajo, se ha realizado la búsqueda del bosón de Higgs en el canal $gg \rightarrow H \rightarrow WW^* \rightarrow 2l2\nu$. El proceso de fusión de gluones es el mecanismo de producción predominante en el LHC y el canal WW es considerado como uno de los más sensibles a la búsqueda del bosón de Higgs en el rango de masas intermedias entre 120 y 200 GeV/c^2.
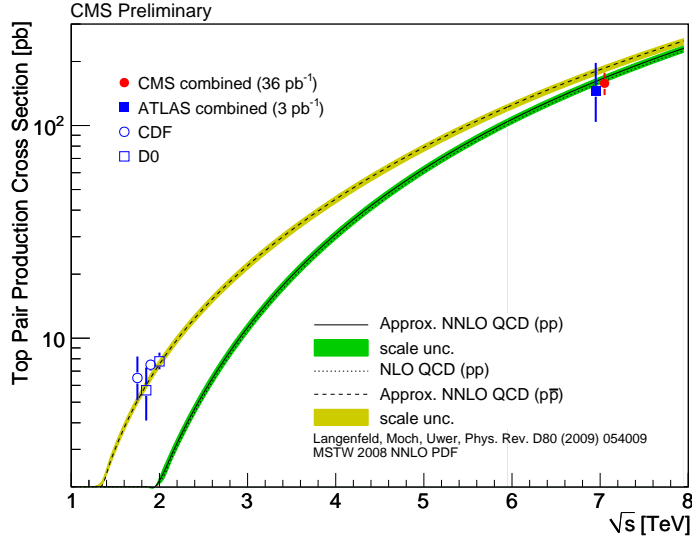
Figure 8.4: - Sección eficaz de producción de pares de quarks top en función de $\sqrt{s}$, para colisiones $p\bar{p}$ y pp [98]

Este canal tiene una signatura limpia caracterizada por la presencia de dos leptones de alto p_T , una cantidad significativa de $\not{E}_T$ debida a los neutrinos no detectados, así como poca actividad hadrónica. Las principales fuentes de fondo vienen de diferentes procesos de SM que pueden tener estados finales similares a los de la seal, principalmente $W + jets$, $Z + jets$, $t\bar{t}$ y el casi irreducible WW. Otras fuentes de fondo son los procesos WZ, ZZ y tW.

Dada la presencia de neutrinos en el estado final, no se puede reconstruir un pico de masa en este canal, y el hecho de que la sección eficaz de producción de la seal es varios órdenes de magnitud menor que la de los fondos, la búsqueda depende de un control lo más preciso posible de las fuentes de fondo dominantes.

Una reconstrucción e identificación de leptones eficientes, así como un criterio robusto de aislamiento son importantes para seleccionar los leptones del W y es la clave para reducir el fondo de $W + jets$ cuando uno de los W se desintegra en un lepton. La producción de pares top-antitop es uno de los fondos principales, se diferencia de la señal por la actividad hadrónica y por tanto para reducirlo se vetan los sucesos que presenten jets energéticos, además la herramienta de etiquetado de quarks b también permite reducirlo. Una estimación correcta de la $\not{E}_T$ en sucesos que no tienen $\not{E}_T$ genuina es crucial para reducir el fondo de $Z + jets$ a un nivel aceptable. La variable principal para discriminar la señal del fondo WW es el ángulo entre los leptones en el plano transverso al haz $\Delta\phi_{ll}$. Los leptones emitidos por los bosones W que proceden de la desintegraciòn del osòn de Higgs son más cercanos en el plano transverso debido a las correlaciones de spin, mientras que la separaciòn angular de los leptones que provienen directamente de la desintegración del WW tienden a ser mayores. Esta es una de las características distintivas principales entre la señal y el fondo WW, especialmente en el rango de masas pròximo a 160 GeV/c^2.

Se analizaron los datos de 2010, aplicando las técnicas de análisis desarrolladas previamente con simulación. La luminosidad acumulada no es suficiente para tener sensibilidad al bosón de Higgs pero fue posible poner límites en el contexto de una extensión del Modelo Estándar con una cuarta

generación de fermiones.[114].

Se desarrollaron métodos para estimar la contribución de los principales fondos WW, $t\bar{t}$, $Z/\gamma^* \to l^+l^-$ and $W + jets$, al igual que se realizó para la medida de la sección eficaz del top.

En el contexto de la medida de la sección eficaz de producción del proceso electrodébil WW, se obtuvieron los siguientes valores para los fondos más importantes a partir de los datos: $1.7 \pm 0.4 \pm 0.7$ y sucesos de $W + jets$, $0.2\pm 0.2\pm 0.3$ sucesos de $Z/\gamma^* \to l^+l^-$, los cuales se escalan de acuerdo a las eficiencias de selección en el análisis $H \to WW^* \to 2l2\nu$, ya que no hay sucesos en la región de control al final de la secuencia de análisis.

Además del análisis basado en cortes secuenciales aplicados en los principales observables, un análisis multidimensional, basado en el uso de *Boosted Decision Trees* se desarrolló para mejorar más la sensibilidad. Variables adicionales se usaron para mejorar la discriminación, como la masa transversa o el ángulo entre los leptones y la energía transversa faltante. Ambos métodos se optimizaron para todas las hipótesis de masa consideradas, para tener en cuenta las diferencias cinemáticas de la señal.

Usando cualquiera de los dos análisis, no se ha observado ningún exceso compatible con la señal de Higgs en la muestra de datos, por tanto, se han establecido límites de exclusión a un nivel de confianza del 95% en la región de masa 120-600 GeV/c^2 (Ver Tabla 8.2 y Figura 8.5) [107].

$m_H(GeV/c^2)$	Lim.obs cut-based	Lim.exp cut-based	Lim. obs BDT-based	Lim.exp. BDT-based
130	6.30	8.07	5.66	6.57
160	2.29	3.22	1.93	2.72
200	2.80	4.59	2.32	3.72
210	3.41	5.53	2.76	4.43
400	2.08	3.12	1.94	2.93

Table 8.2: Límites de exclusión esperados y observados a un nivel de confianza en la sección eficaz de producción de $\sigma_H \times BR(H \to WW^* \to 2l2\nu)$ para cuatro hipótesis de masa del bosón de Higgs.
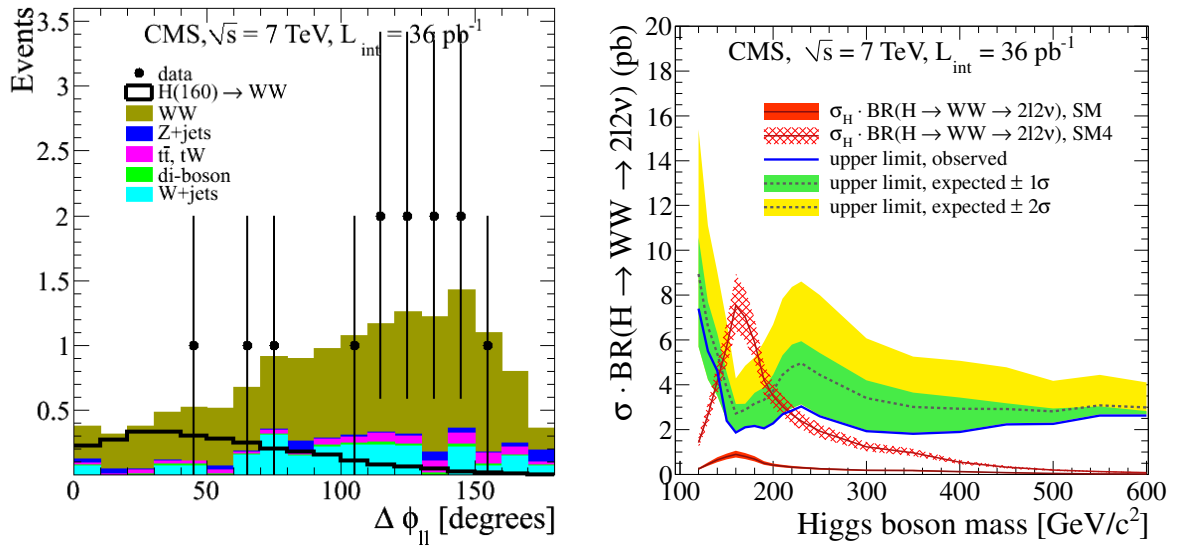
Figure 8.5: - Izquierda: ángulo entre los leptones en el plano transverso después de la selección WW para una bosón de Higgs del SM de masa $m_H = 160 \ GeV/c^2$ y para los fondos, Derecha: Límites de exclusión esperados y observados a un nivel de confianza del 95% en un rango de masa 120-600 GeV/c^2 usando un aálisis secuencial y un análisis multidimensional. [107]

8. RESUMEN DEL TRABAJO REALIZADO

9

Appendix A

9.1 Multivariate Analysis

9.1.1 Boosted Decision Trees

A Decision Tree [117] is a binary classifier with a structure similar to what is represented in Figures 9.1 and 9.4. Successive decisions on a single variable at each level about wether a particular condition is satisfied or not are consecutively taken until a certain stop criterion is reached. This forms a tree-like structure in which an event is finally classified as signal or background according to the class of the final node (called "leaf") [109].
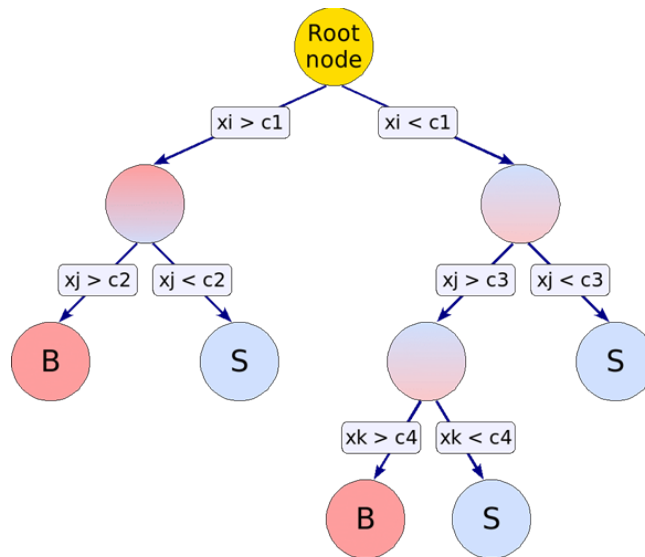
Figure 9.1: - Scheme of a decision tree

The training of a decision tree is the process that defines the splitting criteria for each node. The training starts with the root node, in which all the events are present and where an initial splitting criterion for the full training sample is determined. The split results in two subsets of training events that each go through the same algorithm of determining the next splitting iteration. This procedure is repeated until the whole tree is built. At each node, the split is determined by finding the variable and corresponding cut value that provides the best separation between signal and background. The node splitting stops once it has reached the minimum number of events which is specified in the BDT configuration. The leaf nodes are classified as signal or background according to the class the majority

of events belongs to.

There are several methods available to select the optimal variable and specific cut at each node. As the idea is to separate signal from background, a cut that selects predominantly background is as valuable as selecting signal, so the criteria are symmetric with respect to the event classes. The separation criteria is maximum when the samples are fully mixed (purity P=0.5) and fall off to zero when the sample consist of one event class only. One of the most common methods and the one used here is the so called GiniIndex algorithm.The Gini Index is defined as:

$$Gini = (\sum_{i=1}^{n} W_i) \cdot P(1 - P))$$ (9.1)

with N being the total number of events in the node.

The idea is to minimize $Gini_{node1} + Gini_{node2}$ and to determine the increase in quality when a node is split into two branches, one maximizes $V = Gini_{father} - (Gini_{son1} + Gini_{son2}$.

Since the splitting criterion is always a cut on a single variable, the training procedure selects the variable and cut value that optimizes the increase in the separation index between the parent node and the sum of the indices of the two daughter nodes, weighted by their relative fraction of events.

The cut values are optimized by scanning over the variable range with a granularity that is set via the option nCuts in the BDT configuration. The default value of nCuts=20 proved to be a good compromise between computing time and step size. Finer stepping values did not increase noticeably the performance of the BDTs.

In principle, the splitting could continue until each leaf node contains only signal or only background events, which could suggest that perfect discrimination is achievable. However, such a decision tree would be strongly overtrained. To avoid overtraining a decision tree must be pruned.

Pruning is the process of cutting back a tree from the bottom up after it has been built to its maximum size. Its purpose is to remove statistically insignificant nodes and thus reduce the overtraining of the tree. One method used to achieve this is called Cost complexity pruning [118], it relates the number of nodes in a subtree below a node to the gain in terms of misclassified training events by the subtree compared to the node itself with no further splitting. Figure 9.2 shows an example of the number of nodes present in a decision tree before and after pruning is applied.

The cost estimate R chosen for the misclassification of training events is given by the misclassification rate 1 - max(P, 1-P) in a node. The cost complexity for this node is then defined by:

$$\rho = \frac{R(node) - R(subtreebelowthatnode)}{Nnodes(subtreebelowthatnode) - 1}$$ (9.2)

The node with the smallest ρ value in the tree is recursively pruned away as long as $\rho <$ PruneStrength, which is defined in the parameter configuration of the BDT.

A ranking of the BDT input variables is derived by counting how often the variables are used to split decision tree nodes, and by weighting each split occurrence by the separation gain-squared it has achieved and by the number of events in the node. This measure of the variable importance can be used for a single decision tree as well as for a forest.
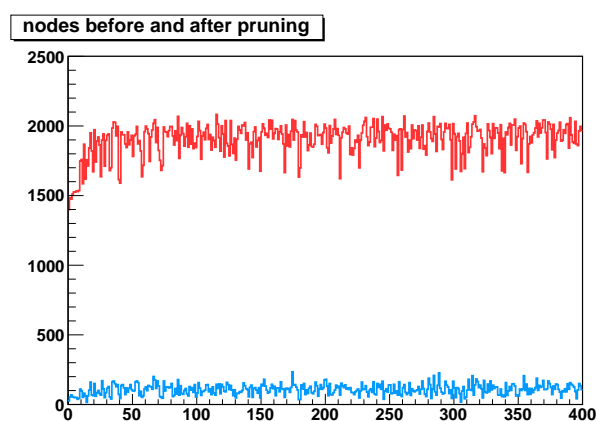
Figure 9.2: - Example of the number of nodes before and after pruning vs the tree number

A Decision Tree could be sensitive to possible fluctuations in the training sample from which the structure is derived. Statistical stability can be improved using the so-called boosting. Several decision trees are derived from the same training ensemble by reweighting events and finally combined into a single classifier constructed by a weighted average of the responses of the individual decision trees.

The most popular boosting algorithm is the so-called AdaBoost (Adaptive Boost) [119]. In a classification problem, events that were misclassified during the training of a decision tree are given a higher event weight in the training of the following tree. Starting with the original event weights when training the first decision tree, the subsequent tree is trained using a modified event sample where the weights of previously misclassified events are multiplied by a common boost weight α. The boost weight is derived from the misclassification rate, err, of the previous tree:

$$\alpha = \frac{1 - err}{err} \tag{9.3}$$

By construction, the error rate is err $<= 0.5$ as the same training events used to classify the output nodes of the previous tree are used for the calculation of the error rate. The weights of the entire event sample are then renormalized such that the sum of weights remains constant. We define the result of an individual classifier as h(x), with (x being the tuple of input variables) encoded for signal and background as h(x) = +1 and -1, respectively. The boosted event classification yBoost(x) is then given by:

$$y_{Boost}(x) = \frac{1}{N_{trees}} \cdot \sum_{i=1}^{N_{trees}} \ln(\alpha_i) h_i(x) \tag{9.4}$$

where the sum is over all classifiers in the collection. Small (large) values for yBoost(x) indicate a background-like (signal-like) event. Figure 9.3 shows the error fraction and AdaBoost weights for each of the trees in the "forest".

Decision Trees are also insensitive to the addition of poorly discriminant variables, since they are basically ignored, as at each level only the best discriminant are used, therefore the same variable can be used in several nodes or even other variables can never be used. One of the most important parameters in the configuration of a decision tree is the minimum number of events that a final node must have. If this is large, it may be that some details can be not well modeled and if this is very
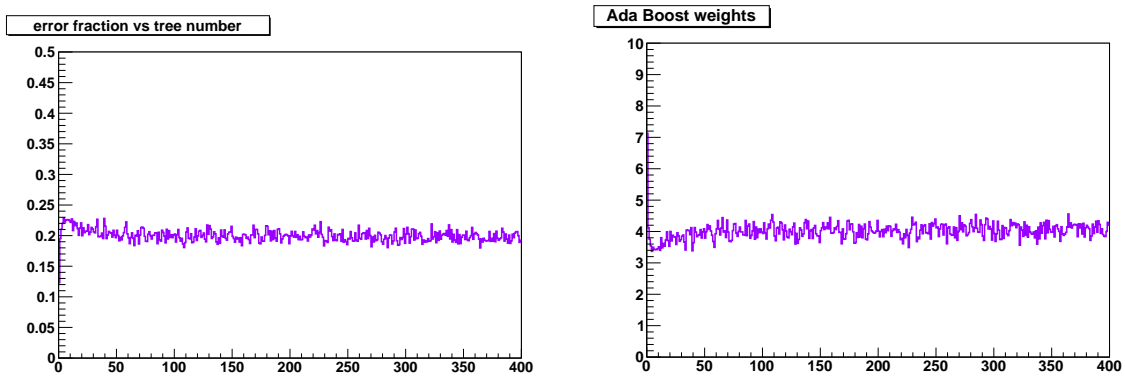
Figure 9.3: - Error fraction and AdaBoost weights vs the tree number

small the overtraining risk grows.

Overtraining leads to a seeming increase in the classification or regression performance over the objectively achievable one, if measured on the training sample, and to an effective performance decrease when measured with an independent test sample. A convenient way to detect overtraining and to measure its impact is therefore to compare the performance results between training and test samples. Such a test is performed by TMVA with the results printed to standard output.
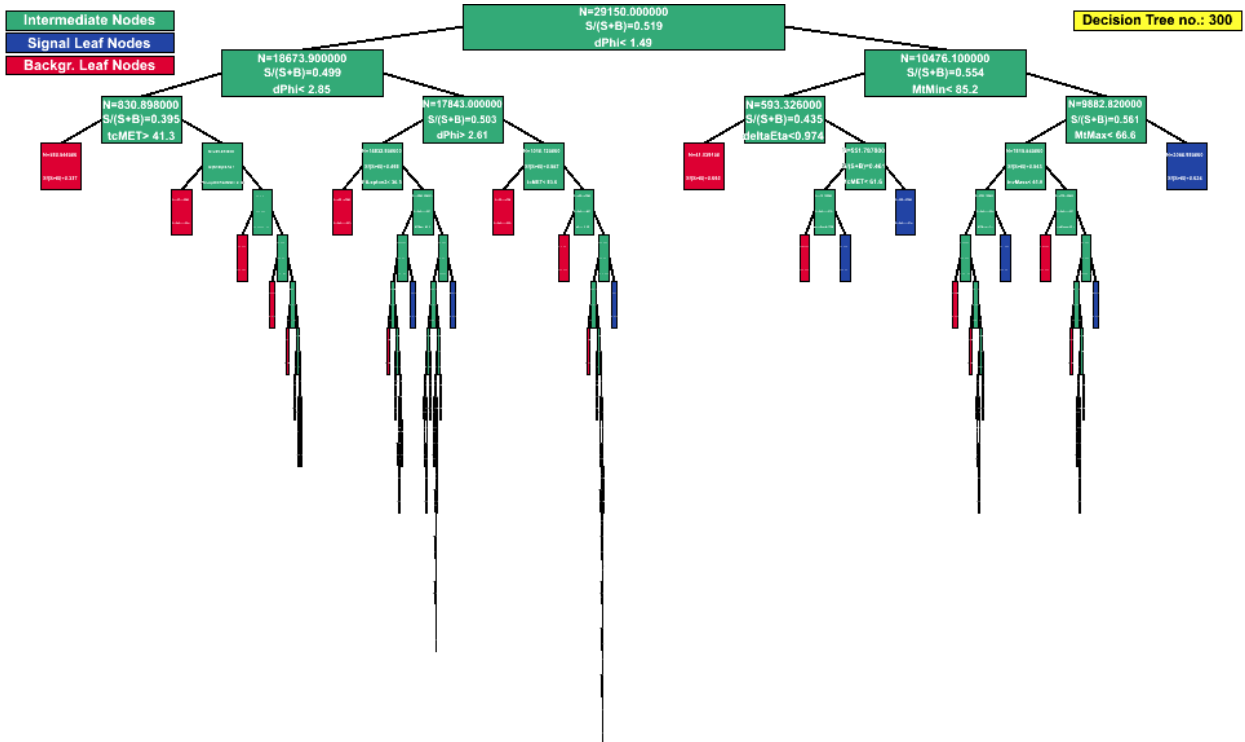

Figure 9.4: - Example of a decision tree applied to the Higgs analysis described in this thesis

9.2 SM Higgs Monte Carlo based studies before the data taking

Higgs search strategies have been defined in the past based on MC, detailed studies were performed to optimize each part of the analysis in order to improve sensitivity, adapting it to the different proposed running conditions of the LHC: first to the nominal center of mass energy of 14 TeV, then to 10 TeV and finally to 7 TeV and assuming different integrated luminosities according to the plans.

Before the beginning of the data taking, a multivariate analysis based on Boosted Decision Trees was studied to improve sensitivity, and was also improved and adapted to the needs. Several methods have been tested in the past: from a BDT trained against the sum of all backgrounds to the combination of 2 or 3 independent BDTs.

A brief summary of the results obtained for the 14 TeV [120] and 10 TeV [121] scenario is given here.

9.2.1 Analysis in the 14 TeV centre of mass energy scenario

The analysis proposed for this energy was assumed for an integrated luminosity of 1 fb^{-1}. A summary of the results is shown in Table 9.1 and in Figures 9.5 and 9.6.

m_H [GeV]	120	130	140	150	160	170	180	190	200
N_S	5.5± 0.4	6.6± 0.2	7.7± 0.9	14± 1	18± 1	12.7± 0.9	16± 1	10.7± 0.9	7.9± 0.8
N_B	40± 3	40± 3	25± 2	25± 2	6± 1	6± 1	23± 3	25± 3	25± 3

Table 9.1: Expected number of events for $1fb^{-1}$ as a function of the Higgs mass at 14 TeV
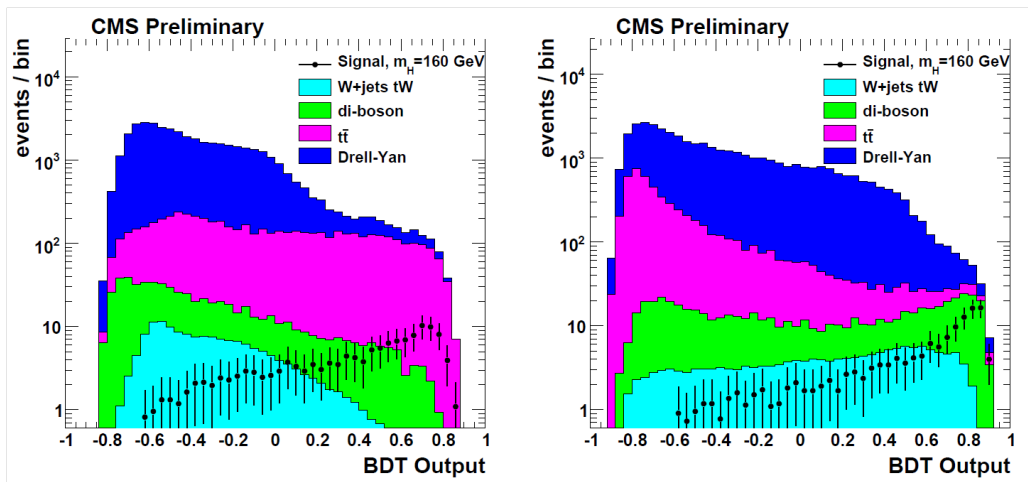

Figure 9.5: - Distributions of the BDT function of a BDT trained against WW (left) and against $t\bar{t}$ (right) normalized to $1fb^{-1}$ at 14 TeV

9.2.2 Analysis in the 10 TeV centre of mass energy scenario

The analysis proposed for this energy assumed an integrated luminosity of 200 pb^{-1} and the corresponding improvements on the knowledge of the detector at that time were incorporated to the
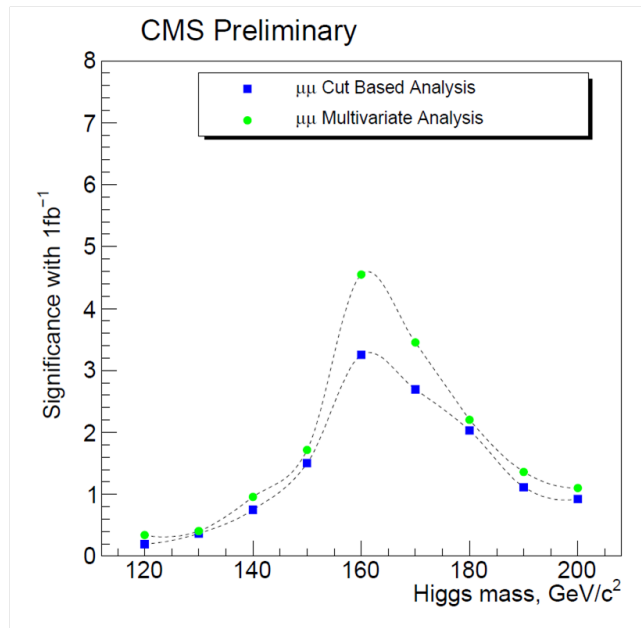
Figure 9.6: - Significance as a function of the Higgs mass at 14 TeV

study.
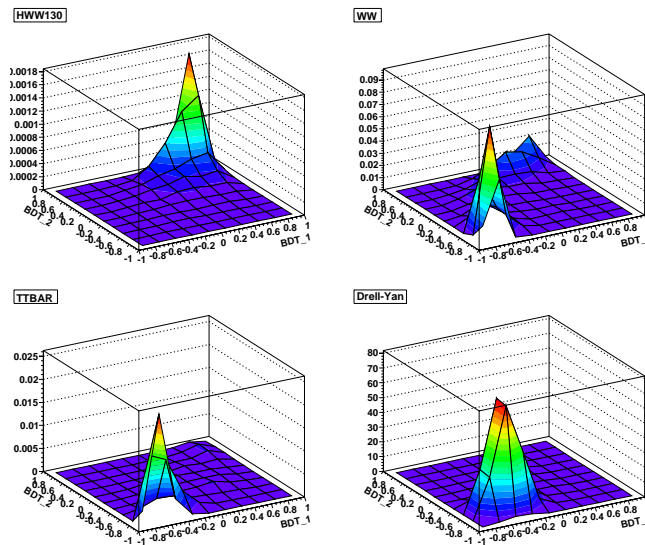A summary of the results is shown in Tables 9.2 to 9.6 and Figures 9.7, 9.8, 9.9 and 9.10.

Figure 9.7: BDT 2D Distributions for the signal(m_H= 130 GeV/c^2) and main backgrounds WW, $t\bar{t}$ and Drell-Yan. The X axis shows the response of a BDT trained against WW, while the Y axis represents the response of a BDT trained against $t\bar{t}$, a clear discrimination between signal and background is observed.

	HWW130	WW	WZ	ZZ	$t\bar{t}$	tW	$Z + jets$
$\mu\mu$ final state	7.8	56.9	25.7	9.24	457.8	37.5	1213.2
Jet Veto	4.65	38.4	7.9	6.10	12.7	3.7	346.9
N_{total}	2.44 ± 0.06	4.4 ± 0.2	0.13 ± 0.06	0.14 ± 0.01	0.43 ± 0.06	0.07 ± 0.05	0.2 ± 0.1

Table 9.2: Expected number of events for $200pb^{-1}$ at 10 TeV, optimized for $m_H = 130 \ GeV/c^2$

	HWW150	WW	WZ	ZZ	$t\bar{t}$	tW	$Z + jets$
$\mu\mu$ final state	17.9	56.9	25.7	9.24	457.8	37.5	1213.2
Jet Veto	10.4	38.4	7.9	6.10	12.7	3.7	346.9
N_{total}	3.1 ± 0.1	2.4 ± 0.2	0.05 ± 0.04	0.035 ± 0.005	0.24 ± 0.04	0.07 ± 0.05	0

Table 9.3: Expected number of events for $200pb^{-1}$ at 10 TeV, optimized for $m_H = 150 \ GeV/c^2$

	HWW160	WW	WZ	ZZ	$t\bar{t}$	tW	$Z + jets$
$\mu\mu$ final state	22.9	56.9	25.7	9.24	457.8	37.5	1213.2
Jet Veto	13.1	38.4	7.9	6.10	12.7	3.7	346.9
N_{total}	4.2 ± 0.1	1.5 ± 0.1	0.05 ± 0.04	0.023 ± 0.004	0.14 ± 0.03	0.03 ± 0.03	0

Table 9.4: Expected number of events for $200pb^{-1}$ at 10 TeV, optimized for $m_H = 160 \ GeV/c^2$

	HWW170	WW	WZ	ZZ	$t\bar{t}$	tW	$Z + jets$
$\mu\mu$ final state	23.2	56.9	25.7	9.24	457.8	37.5	1213.2
Jet Veto	12.75	38.4	7.9	6.10	12.7	3.7	346.9
N_{total}	4.1 ± 0.1	2.4 ± 0.2	0.05 ± 0.04	0.035 ± 0.005	0.24 ± 0.04	0.07 ± 0.05	0

Table 9.5: Expected number of events for $200pb^{-1}$ at 10 TeV, optimized for $m_H = 170 \ GeV/c^2$

	HWW200	WW	WZ	ZZ	$t\bar{t}$	tW	$Z + jets$
$\mu\mu$ final state	13.4	56.9	25.7	9.24	457.8	37.5	1213.2
Jet Veto	6.8	38.4	7.9	6.10	12.7	3.7	346.9
N_{total}	1.24 ± 0.05	1.4 ± 0.1	0.21 ± 0.07	0.23 ± 0.01	0.29 ± 0.05	0.3 ± 0.1	0.2 ± 0.1

Table 9.6: Expected number of events for $200pb^{-1}$ at 10 TeV, optimized for $m_H = 200 \ GeV/c^2$

9. APPENDIX A

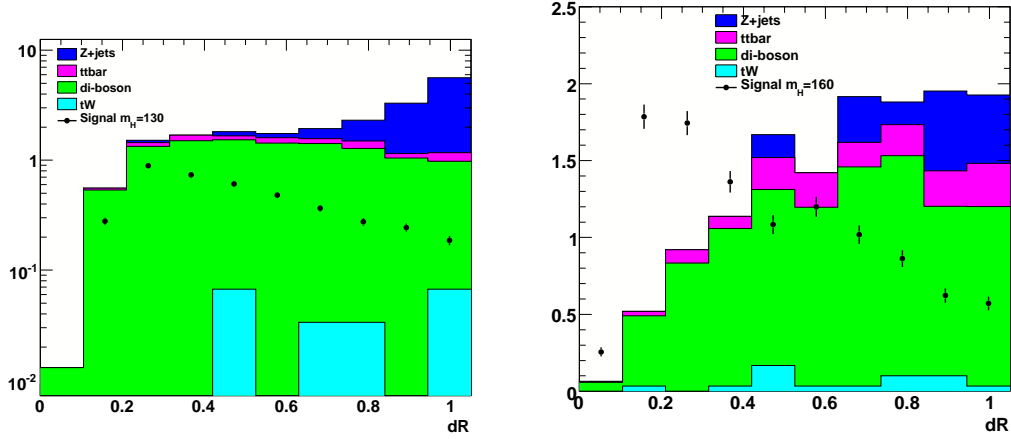**Figure 9.8:** Distribution of the final BDT 2D discriminating variable for m_H=130 GeV/c^2 (left) and m_H=160 GeV/c^2 (right), for an integrated luminosity of $200pb^{-1}$ at 10 TeV
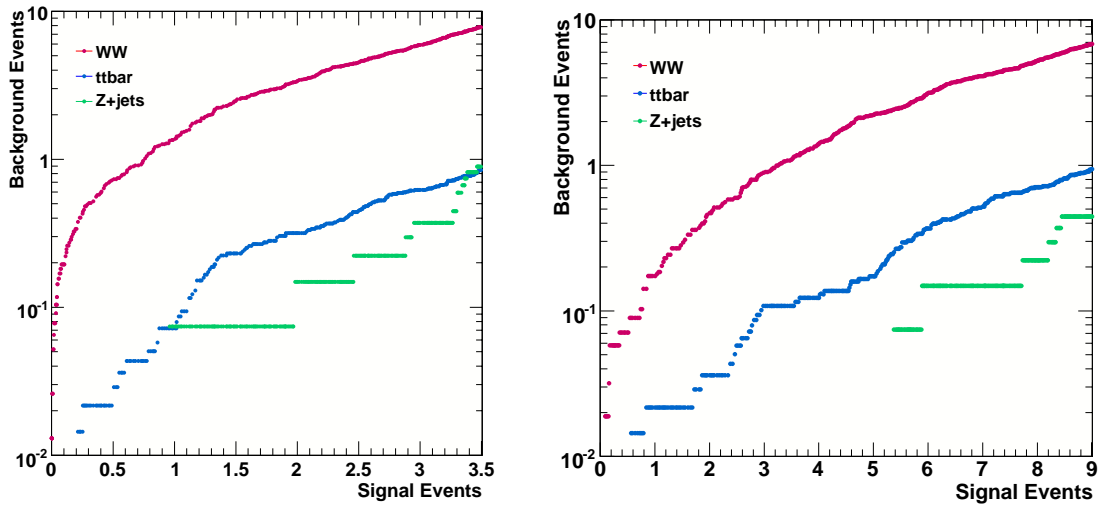

Figure 9.9: Number of signal events vs number of background events after several cuts on the BDT function shown in Figure 9.8, for m_H=130 GeV/c^2 (left) and m_H=160 GeV/c^2 (right).
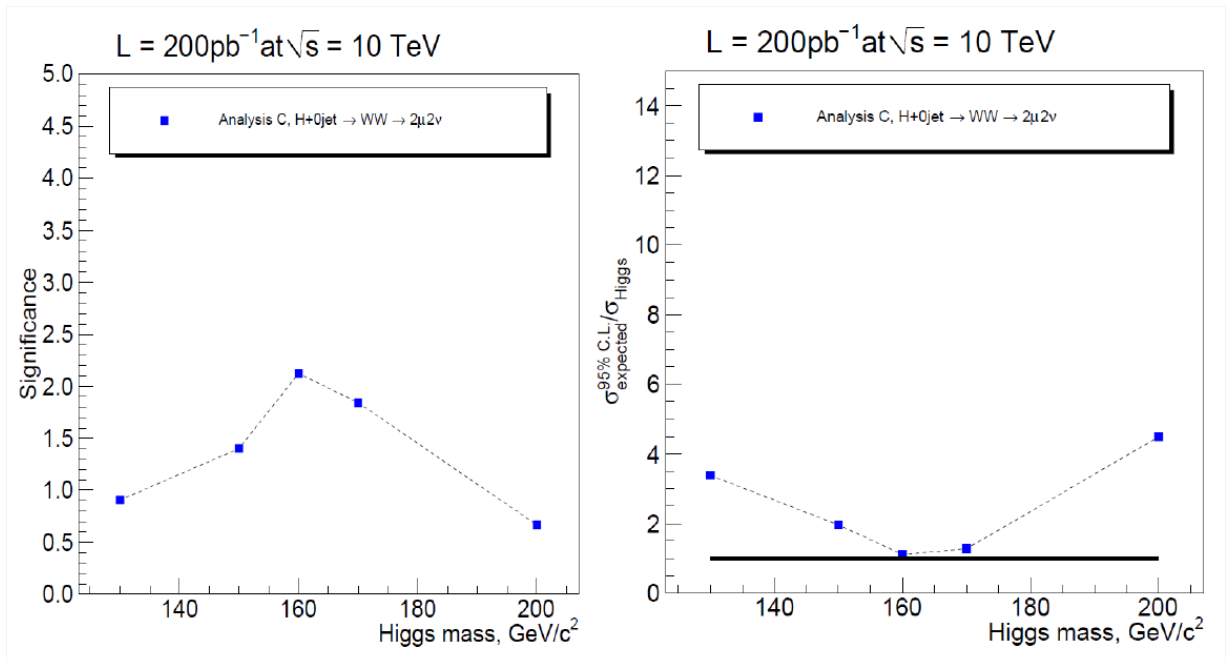
Figure 9.10: - Significance (left) and 95% C.L. exclusion limits (right) as a function of the Higgs mass for 200 pb^{-1} at 10 TeV

9. APPENDIX A

References

[1] **LEP Working Group for Higgs boson searches** Collaboration, R. Barate *et. al.*, *Search for the standard model Higgs boson at LEP*, *Phys. Lett.* **B565** (2003) 61–75, [`hep-ex/0306033`]. 1, 16, 18, 187

[2] **CDF and D0** Collaboration, T. Aaltonen *et. al.*, *Combined CDF and D0 Upper Limits on SM Higgs Boson Production with up to 8.2 fb-1 of Data*, `arXiv:1103.3233`. 1, 19, 20, 187

[3] J. Alcaraz, *Precision Electroweak Measurements and Constraints on the SM*, `arXiv:0911.2604`. 1, 14, 187

[4] D. Perkins, *Introduction to High Energy Physics*. Addison-Wesley, USA, 1987. 5

[5] S.L.Glashow, *Partial Symmetries of weak interactions*, *Nucl.Phys.* **22** (1961) 579–588. 5

[6] S.Weinberg, *A model of leptons*, *Phys.Rev.Lett.* **19** (1967) 1264–1266. 5

[7] A.Salam, *Weak and electromagnetic interactions*, in *Elementary Particle Theory, Proceedings of the Nobel Symposium held 1968 at Lerum, Sweden* (N.Svartholm, ed.), pp. 367–377. Almqvist & Wiksell, 1968. 5

[8] **Particle Data Group** Collaboration, K. Nakamura *et. al.*, *Review of particle physics*, *J. Phys.* **G37** (2010) 075021. 7, 79, 113

[9] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*. Addison-Wesley, USA, 1995. 7

[10] R. Feynman, *QED: the strange theory of light and matter*. Princeton University Press, 1985. 8

[11] E. Noether, *M. A. Tavel's English translation of Noether's Theorems (1918), reproduced by Frank Y. Wang*, `physics/0503066v1`. 8

[12] C. R. H. G. S. Guralnik and T. W. B. Kibble, *Global Conservation Laws and Massless Particless*, *Phys. Rev. Lett.* **13** (1964) 585?587. 11

[13] P. Higgs, *Broken Symmetries and the Mass of the Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508–509. 11

[14] P. Higgs, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Lett.* **12** (1964) 132–133. 11

[15] F. Englert and R. Brout, *Global Conservation Laws and Massless Particless*, *Phys. Rev. Lett.* **13** (1964) 321–323. 11

REFERENCES

[16] **CDF** Collaboration, F. A. et al, *Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74** (1995) 2626 – 2631. 19

[17] **D0** Collaboration, S. A. et al., *Observation of the Top Quark*, Phys. Rev. Lett. **74** (1995) 2632 – 2637. 19

[18] M. E. M. H. M. Georgi, S. L. Glashow and D. V. Nanopoulos, *Higgs Bosons from Two-Gluon Annihilation in Proton-Proton Collisions*, Phys. Rev. Lett. **40** (1978) 692 – 694. 22

[19] M. G. P. N. S. Catani S, D. de Florian, *Soft-gluon resummation for Higgs boson production at hadron colliders*, JHEP **0307** (2003) 028, [`hep-ph/0306211`]. 22

[20] **CMS** Collaboration, "CMS Higgs public results." 24

[21] **CMS** Collaboration, G. L. Bayatian *et. al.*, *CMS technical design report, volume II: Physics performance*, J. Phys. **G34** (2007) 995–1579. 25

[22] J. Erler, *The Mass of the Higgs Boson in the Standard Electroweak Model*, Phys.Rev.D81:051301 (2010) [`hep-ph/1002.1320v1`]. 25

[23] M. Dittmar and H. K. Dreiner, *How to find a Higgs boson with a mass between 155-GeV - 180-GeV at the LHC*, Phys. Rev. **D55** (1997) 167–172, [`hep-ph/9608317`]. 25

[24] S. Baffioni *et. al.*, *Search for the higgs boson in the zz* decay channel with the cms experiment*, . CMS-PAS-HIG-08-003 / CMS AN-2008/050. 25

[25] M. V. Acosta *et. al.*, *Towards the search for the sm higgs boson produced in vector boson fusion and decaying into a τ pair in cms with 1 fb^{-1}: τ identifications studies*, . 25

[26] M. V. Acosta *et. al.*, *Search for the sm higgs boson produced in vector boson fusion and decaying into a tau pair in cms with 1 inverse femtobarn*, . CMS-PAS-HIG-08-008 /CMS AN-2007/035. 25

[27] **LHCb** Collaboration, *LHCb technical design report: Reoptimized detector design and performance*. CERN-LHCC-2003-030. 26, 31

[28] **CDF** Collaboration, F. Abe *et. al.*, *Observation of top quark production in p$\bar{p}$ collisions with the collider detector at fermilab*, Phys. Rev. Lett. **74** (Apr, 1995) 2626–2631. 26

[29] **D0** Collaboration, S. Abachi *et. al.*, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632–2637, [`hep-ex/9503003`]. 26

[30] **CDF and D0 Collaboration** Collaboration, *Combination of CDF and D0 Results on the Mass of the Top Quark using up to 5.6 fb^{-1} of data*, `arXiv:1007.3178`. Long author list - awaiting processing. 26

[31] E. . Bruning *et. al.*, *LHC design report. Vol. I: The LHC main ring*, . CERN-2004-003-V-1. 29

[32] **ATLAS** Collaboration, G. Aad *et. al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003. 31

[33] **CMS** Collaboration, R. Adolphi *et. al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004. 31, 32

[34] **ALICE** Collaboration, *ALICE: Technical proposal for a large ion collider experiment at the CERN LHC.* CERN-LHCC-95-71. 31

[35] **CMS** Collaboration, *The CMS magnet project : Technical Design Report*, 1997. 34

[36] **CMS** Collaboration, G. L. Bayatian *et. al.*, *CMS physics: Technical design report.* CERN-LHCC-2006-001. 34

[37] **CMS** Collaboration, V. Khachatryan *et. al.*, *CMS Tracking Performance Results from early LHC Operation*, Eur. Phys. J. **C70** (2010) 1165–1192, [`arXiv:1007.1988`]. 35

[38] **CMS** Collaboration, *The CMS electromagnetic calorimeter project : Technical Design Report* , 1997. 36

[39] **CMS** Collaboration, *Changes to CMS ECAL electronics : addendum to the Technical Design Report*, 2002. 36

[40] **CMS** Collaboration, *The CMS hadron calorimeter project : Technical Design Report*, 1997. 37

[41] **CMS** Collaboration, *The CMS muon project : Technical Design Report* , 1997. 39

[42] **CMS** Collaboration, *CMS TriDAS project : Technical Design Report; 1, the trigger systems*, 2000. 42

[43] **CMS** Collaboration, *CMS trigger and data-acquisition project : Technical Design Report*, 2002. 43

[44] **CMS** Collaboration, *The CMS Computing Model*, 2004. 46

[45] **GEANT4** Collaboration, S. Agostinelli *et. al.*, *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303. 46

[46] G. Codispoti *et. al.*, *CRAB: A CMS application for distributed analysis*, IEEE Trans. Nucl. Sci. **56** (2009) 2850–2858. 47

[47] R. Brun and F. Rademakers, *Root - an object oriented data analysis framework*, Inst. Meth. in Phys. Res. A **389, 81** (1997). 48

[48] **CMS** Collaboration, G. Abbiendi *et. al.*, *Muon Reconstruction in the CMS Detector*, . CMS AN 2008/097. 51

[49] CMS collaboration, *Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV*, PAS MUO-10-002 (2010). 51

[50] **CMS** Collaboration, M. Mulders *et. al.*, *Muon Identification in CMS*, . CMS AN 2008/098. 53

[51] S. Baffioni *et. al.*, *Electron reconstruction in CMS*, Eur. Phys. J. **C49** (2007) 1099–1116. 55

[52] R. Frühwirth, *Track fitting with non-gaussian noise*, Computer Physics Communications **100** (1997) 1–16. 56

[53] W. Adam *et. al.*, *Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc*, . CMS NOTE-2005/001. 56

[54] CMS collaboration, *Electron reconstruction and identification at at $\sqrt{s} = 7$ TeV* , PAS EGM-10-004. 57, 92

REFERENCES

[55] G. P. Salam, *Towards Jetography*, `hep-ph/0906.1833v2`. 58

[56] CMS collaboration, *Determination of the Jet Energy Scale in CMS with pp Collisions at* $\sqrt{s} = 7$ *TeV*, *PAS JME-10-010* (2010). 58, 61, 62, 113

[57] **CMS** Collaboration, R. Bainbridge *et. al.*, *The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS*, . CMS-PAS-JME-09-002 / CMS AN 2008/084. 59

[58] **CMS** Collaboration, *Particle Flow Reconstruction of Jets, Taus, and MET*, . CMS PAS PFT-09-001; CMS AN-2009/039. 59

[59] **CMS** Collaboration, *Performance of Jet Reconstruction with Charged Tracks only*, . CMS PAS JME-08-001; CMS AN-2008/041. 59

[60] **CMS** Collaboration, *Performance of Jet Algorithms in CMS*, . CMS PAS JME-07-003; CMS AN-2008/001. 60

[61] G. S. G. P. Salam, *A practical Seedless Infrared-Safe Cone jet algorithm*, *JHEP* **0705:086** (2007) [`hep-ph/0704.0292`]. 60

[62] S. Catani *et. al.*, *New clustering algorithm for multi - jet cross-sections in e+ e- annihilation*, *Phys.Lett. B* **269** (1991) 432–438. 61

[63] S. Catani *et. al.*, *Longitudinally-invariant k-clustering algorithms for hadron-hadron collisions*, *Nucl. Phys. B* **406** (1993) 187–224. 61

[64] D. E. S. S. D. Ellis, *Successive Combination Jet Algorithm For Hadron Collisions*, *Phys.Rev. D* **48** (1993) 3160–3166, [`hep-ph/9305266`]. 61

[65] S. M. B. W. Yu.L. Dokshitzer, G.D. Leder, *Better Jet Clustering Algorithms*, *JHEP* **9708:001** (1997) [`hep-ph/9707323`]. 61

[66] T. W. M. Wobisch, *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*, `hep-ph/9907280`. 61

[67] G. S. M. Cacciari, G. P. Salam, *The anti-k_t jet clustering algorithm*, *JHEP* **0804:063** (2008) [`hep-ph/0802.1189`]. 61, 85, 138

[68] CMS collaboration, *CMS MET Performance in Events Containing Electroweak Bosons from pp Collisions at* $\sqrt{s} = 7$ *TeV*, *PAS JME-10-005* (2010). 62

[69] CMS collaboration, *Missing E_T performance in CMS*, *PAS JME-07-001* (2007). 64

[70] **CMS** Collaboration, F. Golf *et. al.*, *Performance of Track-Corrected Missing Transverse Energy in CMS*, . CMS-PAS-JME-09-010/CMS AN-2009/022. 64

[71] CMS collaboration, *Particle-Flow Event Reconstruction in CMS*, *PAS PFT-09-001* (2010). 64

[72] CMS collaboration, *Algorithms for b jet identification in cms*, *PAS BTV-09-001* (2009). 65

[73] CMS collaboration, *Commissioning of b-jet identification with pp collisions at sqrts= 7 tev*, *PAS BTV-10-001* (2010). 66

[74] P. Nason, S. Dawson, and R. K. Ellis, *The total cross section for the production of heavy quarks in hadronic collisions*, *Nuclear Physics B* **303** (1988), no. 4 607 – 633. 69

[75] W. Beenakker, H. Kuijf, W. L. van Neerven, and J. Smith, *QCD Corrections to Heavy Quark Production in p anti-p Collisions*, Phys. Rev. **D40** (1989) 54–82. 69

[76] J. M. Campbell and R. K. Ellis, *MCFM for the Tevatron and the LHC*, Nucl. Phys. Proc. Suppl. **205-206** (2010) 10–15, [`arXiv:1007.3492`]. 69

[77] R. K. Ellis and S. Parke, *Top-quark production by w-gluon fusion*, Phys. Rev. D **46** (Nov, 1992) 3785–3788. 69

[78] J. et al, *Measurement of the $t\bar{t}$ cross section in the dilepton channels with L= 100 pb^{-1} using the CMS detector at $\sqrt{s} = 10$ TeV*, CMS AN-2009/047 (2009). 72

[79] **CMS** Collaboration, V. Khachatryan *et. al.*, *First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV*, Phys. Lett. **B695** (2011) 424–443, [`arXiv:1010.5994`]. 79, 181, 188, 189

[80] N.Adam *et. al.*, *Generic Tag and Probe Tool for Measuring Efficiency at CMS with Early Data*, CMS AN-2009/111 (2009). 86

[81] L.Bauerdick *et. al.*, *A measurement of top quark pair production cross section in dilepton final states in pp collisions at 7 TeV*, CMS AN-2010/410 (2010). 92, 93, 116, 129

[82] CMS collaboration, *Measurement of the top-quark pair-production cross section in the dilepton channel at $\sqrt{s} = 7$ TeV*, PAS TOP-10-005 (2010). 92, 93, 182, 183, 190

[83] J.Andrea *et. al.*, *Measurement of the top dilepton cross section using b-tagging at $\sqrt{s} = 7$ TeV with 36.1 pb^{-1} in pp collisions*, CMS AN-2010/389 (2010). 92, 93

[84] W.Andrews *et. al.*, *A method to measure the contribution of $DY \to l^+ l^-$ to a di-lepton + MET selection* , CMS AN-2009/023 (2009). 104

[85] H. Bakhshian *et. al.*, *Computing the contamination from fakes in leptonic final states*, CMS AN-2010/261 (2010). 108

[86] H. Bakhshian *et. al.*, *Lepton fake rates in dilepton final states*, CMS AN-2010/397 (2010). 108

[87] J. Alwall *et. al.*, *MadGraph/MadEvent v4: The New Web Generation*, JHEP **0709:028** (2007) [`hep-ph/0706.2334`]. 113, 125

[88] CMS collaboration, *Jet Energy Resolution in CMS at $\sqrt{s} = 7$ TeV*, PAS JME-10-014 (2010). 115

[89] F. P. R. P. A. P. M.L. Mangano, M. Moretti, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307 001** (2003) [`hep-ph/0206293`]. 125

[90] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, [`arXiv:0709.2092`]. 125, 176

[91] P. S. T. Sjöstrand, S. Mrenna, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026, [`hep-ph/0603175`]. 126

[92] G. Corcella *et. al.*, *HERWIG 6.5 release note*, `hep-ph/0210213`. 126

[93] CMS collaboration, *Luminosity estimation for 2010 data - update*, PAS EWK-11-001 (2011). 126

REFERENCES

[94] P.Bargassa *et. al.*, *Measurement of the $t\bar{t}$ production cross section in the dilepton channel at $\sqrt{s} = 7$ TeV* , *CMS AN-2010/414* (2010). 129

[95] L. Lyons, D. Gibaut, and P. Clifford, *How to combine correlated estimates of a single physical quantity*, *Nucl. Instrum. Meth.* **A270** (1988) 110. 129

[96] T. D. W. Group, *Combination of results and summary of measurement of the top pair production cross section at 7 TeV in 2010 data*, *CMS AN-2011/018* (2011). 130

[97] CMS collaboration, *Measurement of the $t\bar{t}$ production cross section at $\sqrt{s} = 7$ TeV using b-quark jet identification techniques in lepton+jets events*, *PAS TOP-11-001* (2011). 130

[98] CMS collaboration, *Combination of top pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV and comparisons with theory*, *PAS TOP-11-001* (2011). 130, 131, 132, 183, 184, 190, 191

[99] CMS collaboration, *Measurement of the $t\bar{t}$ production cross section at $\sqrt{s} = 7$ TeV using the kinematic properties of lepton+jets events*, *PAS TOP-11-001* (2011). 130

[100] **Atlas** Collaboration, G. Aad *et. al.*, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7 TeV$*, *Eur. Phys. J.* **C71** (2011) 1577, [`arXiv:1012.1792`]. 130, 132, 183, 184, 190

[101] **CDF** Collaboration, A. A. Affolder *et. al.*, *Measurement of the t anti-t production cross-section in p anti-p collisions at $\sqrt{s} = 1.8 TeV$*, *Phys. Rev.* **D64** (2001) 032002, [`hep-ex/0101036`]. 132, 183, 184, 190

[102] **D0** Collaboration, V. M. Abazov *et. al.*, *t anti-t production cross-section in p anti-p collisions at $\sqrt{s} = 1.8 TeV$*, *Phys. Rev.* **D67** (2003) 012004, [`hep-ex/0205019`]. 132, 183, 184, 190

[103] C. Collaboration, *Combination of CDF top quark pair production measurements with up to 4.6 fb^{-1}*, *CDF Note 9913* (2009). 132, 183, 184, 190

[104] **D0** Collaboration, V. M. Abazov *et. al.*, *Measurement of the top quark pair production cross section in the lepton+jets channel in proton-antiproton collisions at $\sqrt{s}=1.96$ TeV*, `arXiv:1101.0124`. 132, 183, 184, 190

[105] M. Aliev *et. al.*, *– HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR*, *Comput. Phys. Commun.* **182** (2011) 1034–1046, [`arXiv:1007.1327`]. 132, 184

[106] U. Langenfeld, S. Moch, and P. Uwer, *Measuring the running top-quark mass*, *Phys. Rev.* **D80** (2009) 054009, [`arXiv:0906.5273`]. 132, 184

[107] **CMS** Collaboration, S. Chatrchyan *et. al.*, *Measurement of WW Production and Search for the Higgs Boson in pp Collisions at sqrt(s) = 7 TeV*, *Phys. Lett.* **B699** (2011) 25–47, [`arXiv:1102.5429`]. 135, 170, 174, 178, 184, 185, 192, 193

[108] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, *CERN-2011-002* (CERN, Geneva, 2011) [`arXiv:1101.0593`]. 136

[109] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA: Toolkit for Multivariate Data Analysis*, *PoS* **ACAT** (2007) 040, [`physics/0703039`]. 151, 195

[110] W. et al, *Search for Higgs Boson Decays to Two W Bosons in the Fully Leptonic Final State at* $\sqrt{s} = 7$ *TeV, CMS AN-2010/411* (2010). 170

[111] W. et al, *First Measurement of* $pp \rightarrow WW$ *Production Cross-Section at* $\sqrt{s} = 7$ *TeV, CMS AN-2010/344* (2010). 172, 175

[112] J. F. A. OHagan. 178

[113] R. D. Cousins and V. L. Highland, *Incorporating systematic uncertainties into an upper limit, Nucl. Instrum. Meth.* **A320** (1992) 331–335. 178

[114] Q. Li, M. Spira, J. Gao, and C. S. Li, *Higgs Boson Production via Gluon Fusion in the Standard Model with four Generations, Phys. Rev.* **D83** (2011) 094018, [arXiv:1011.4484]. 178, 183, 192

[115] C. collaboration, *The cms physics reach for searches at 7 tev,* . CMS NOTE 2010/008. 179

[116] 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]. 182, 190

[117] Y. Quinlan, *Simplifying Decision Trees Regression Trees,* . 195

[118] R. O. L. Breiman, J. Friedman and C. Stone, *Classification and Regression Trees,* . 196

[119] Y. Freund and R. Schapire. 197

[120] **CMS** Collaboration, C. C. et al, *Search Strategy for a SM Higgs Boson Decaying to Two W Bosons in the Fully Leptonic Final State,* . PAS HIG 08 006. 199

[121] G. B. et al, *Search Strategy for a SM Higgs Boson Decaying to Two W Bosons in the Fully Leptonic Final State at* $\sqrt{s} = 10$ *TeV,* . CMS AN 2009/139. 199

Acknowledgements