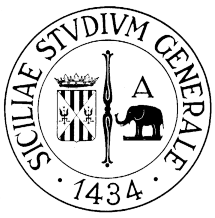
Search for supersymmetric particles in opposite sign dilepton events with the CMS detector



Gigi Cappello

Facoltà di Scienze Matematiche, Fisiche e Naturali

Università degli studi di Catania

A thesis submitted for the degree of

PhilosophiæDoctor (PhD) in Physics

2012

Supervisor: Prof. Alessia Tricomi

Reviewer: Dr. Giuseppe Salvatore Pappalardo
Ph.D. Coordinator : Prof. Francesco Riggi

Cover image:
di-muon event in CMS, sketch by Tom McCauley

Alla mia nipotina
Gaia

Contents

Introduction

The Standard Model of elementary particles and interactions (SM) is a successful theoretical apparatus, whose predictions have been experimentally proved with remarkable precision. Despite its impressive predictive power, the model has some theoretical and phenomenological limits, pointing to the existence of a more general theory, which the Standard Model can be thought of as a low-energy effective theory. The supersymmetric extension of the SM is among the most famous of these theories. Introducing a new symmetry to the model, known as supersymmetry, some theoretical issues (like the so called hierarchy problem) find an elegant solution. If supersymmetry exists (in the way it has been theorized), the supersymmetric partners of the SM particles should have masses not greater than $\mathcal{O}$(TeV). They could then be produced in $p-p$ collision at Large Hadron Collider (LHC).

The LHC is a collider installed at CERN in the ~ 27 km underground tunnel originally hosting the Large Electron-Positron collider (LEP). It started its operations in late 2009 with runs at center of mass energy of $\sqrt{s} = 2.36$ TeV. During the whole 2011 it operated at $\sqrt{s} = 7$ TeV delivering an integrated luminosity of $\sim 5 fb^{-1}$. At the present (2012) $p-p$ runs at $\sqrt{s} = 8$ TeV are ongoing.

Four big experiments are placed at four interaction points; two of them (ATLAS and CMS) are general purpose experiments, designed to accomplish many particle physics tasks, as the search for the Higgs boson and for supersymmetric particles.

In this thesis a search for supersymmetric particles in the data collected by CMS at $\sqrt{s} = 7$ TeV is presented. The search is focused on the selection of events characterized by:

- high missing transverse energy;
- high jet activity;
- two hard leptons of opposite charge.

This particular exclusive signature (which we refer to as *opposite sign dilepton* signature) has a very good background rejection power and should permit spectroscopic studies of the supersymmetric particles produced, once a sufficient amount of them is collected.

The bulk of the analysis consists on the isolation of the events of new physics, through the performance of some selection cuts on properly chosen discriminating variables. Even if, after this step, most part of the standard model background is actually rejected, an estimate of the background events still surviving (mainly top pairs production) is mandatory. The control of any residual background is made using data-driven techniques.

In this work, a particular data-driven method, involving many discriminating variables, is developed. An accurate setup of the method is absolutely necessary to obtain a correct data-driven estimate of the residual background. This delicate step requires the usage of validated Monte Carlo simulations of the Standard Model processes.

The thesis is organized as follows:

In chapter 1 a brief theoretical introduction to supersymmetry is exposed. Some details of the supersymmetry breaking mechanism and of different models are given. The cMSSM (constrained Minimal Supersymmetric Standard Model) is introduced.

In chapter 2 the LHC and the CMS experiment are described. The main features and performances of the sub-detectors of CMS are illustrated.

In chapter 3 after a brief introduction about physics at hadron colliders and simulation of hard scattering processes, a description of supersymmetric phenomenology at CMS is given. Different SUSY analyses are introduced; in particular, the opposite sign dilepton searches are described in more detail. Simplified Models (SMS) are also introduced as an alternative approach to the search with respect to the classical cMSSM scans.

In chapter 4 the definition of the selection criteria of electrons, muons, jets, MET (missing transverse energy), and H_T (scalar sum of jet's transverse momenta) are exposed and the analysis cuts are discussed. The second part of the chapter is dedicated to the validation of Monte Carlo simulations of the Standard Model background. In particular, the effect of pile-up reweighting and of systematic uncertainties on MC are studied.

In chapter 5 the Monte Carlo samples studied and validated in chapter 4 are used to develop and setup a data-driven method, called ABCD method.

In chapter 6 the ABCD method is applied on $\sqrt{s} = 7$ TeV data. The estimation obtained are used to set confidence limits in the cMSSM phase space.

1

Theoretical overview

A theoretical introduction to physics beyond the Standard Model is exposed. A brief analysis of the main issues related to the Standard Model permits to introduce a new space-time symmetry called Supersymmetry. Its main properties are discussed together with some details about its breaking mechanism.

1.1 Beyond the standard model

The Standard Model of elementary particles and interactions (SM) is a gauge theory based on the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ (color + weak isospin + weak hypercharge) spontaneously broken in $SU(3)_C \otimes U(1)_{em}$ (electroweak symmetry breaking) through a scalar Higgs filed. Although its extraordinary predictive power and phenomenological success, some theoretical issues and open questions strongly point to 'something else': a more general theory, including new classes of particles and interactions, of which the SM could be considered a low energy approximation.

	lepton	charge [e]	quark	charge[e]
1st family	e	-1	u	+2/3
	ν_e	0	d	-1/3
2nd family	μ	-1	c	+2/3
	ν_μ	0	s	-1/3
3rd family	τ	-1	t	+2/3
	ν_τ	0	b	-1/3

Table 1.1: Fermions of the Standard Model.

Particle	Field	Group	Mass [GeV]
gluons	G^a_μ	$\mathrm{SU(3)_C}$	0
weak bosons	$W^\pm$	$\mathrm{SU(2)_L} \otimes U(1)_Y$	80.4
	Z^0	$\mathrm{SU(2)_L} \otimes U(1)_Y$	91.2
photon	γ	$\mathrm{U(1)}_{em}$	0

Table 1.2: Bosons of the Standard Model.

1.1.1 The Standard Model of elementary particles and interactions

Matter constituents and fundamental interactions are described as excitations of quantum fields. Matter fields are fermionic (spin=1/2), while gauge fields, related to the fundamental forces, have spin=1, i.e. they are bosonic[1]. Three of the four fundamental interactions are well described by the model while gravity is not included.

The basic bricks of the theory are quarks and leptons: the last ones interact only weakly and (if charged) electromagnetically , while quarks also strongly. Both species exhibit a three-family classification, whose origin still remains unknown.

The theory describing the strong interaction is based on the $SU(3)_C$

(color-$SU(3)$) symmetry and is called *quantum chromo dynamics* (QCD). The carriers of the force (called *gluons*) are the eight generators of the symmetry group. Due to the non-abelian structure of the group, gluons interact not only with quarks but also with themselves (self interaction). This is the basis of distinctive features of the strong interactions, such as confinement and asymptotic freedom[2].

The complex phenomenology of weak interactions cannot be explained using a simple gauge group (as in QCD). Many experimental measurements show in fact a maximal C and P violation for weak interactions, suggesting that they can be described by a group for which left-handed fermions are doublets and right-handed ones are singlets ($SU(2)$-left). This however is strictly true only for charged currents (mediated by $W^{\pm}$ exchange) while neutral currents (mediated by Z^0) also couple with right-handed particles, although with a different coupling constant (i.e. with a different strength).

This particular behavior is successfully considered in account within a theoretical framework first introduced by Glashow, Weinberg and Salam in which weak and electromagnetic interactions are unified under the symmetry group $SU(2)_L \otimes U(1)_Y$. The hypercharge Y is different from the electromagnetic charge Q, although it is related to it by the Gell-Mann - Nishijima formula[3]:

$$Q = I_3 + \frac{1}{2}Y \tag{1.1}$$

where I_3 the third component of the weak isospin. The symmetry group of electromagnetic interactions $U(1)_{em}$ appears after the symmetry breaking of $SU(2)_L \otimes U(1)_Y$.

Since an explicit symmetry breaking would lead to non-renormalizable divergences, a spontaneous symmetry breaking mechanism (Higgs mecha-

nism) has to be introduced adding a new scalar isospin doublet

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{1.2}$$

The interaction with Φ provides masses to the $W^\pm$ and Z^0 bosons (but clearly not to photons) and to the fermions.

Gauge-invariance is preserved even adding to the electroweak Lagrangian density a term:

$$\mathcal{L}_s = (D_\mu \Phi)^\dagger D^\mu \Phi - \mu^2 \Phi^\dagger \Phi - h(\Phi^\dagger \Phi)^2 \tag{1.3}$$

with $h > 0$ and $\mu^2 < 0$, being D_μ the covariant derivative[1] for $SU(2)_L \otimes U(1)_Y$. A negative $\mu^2 < 0$ guarantees an infinity of degenerate states of minimum energy, not $SU(2)$-invariant, whose vacuum expectation value is:

$$\langle 0|\Phi^0|0\rangle = \sqrt{\frac{-\mu^2}{2h}} \equiv \frac{v}{\sqrt{2}} \tag{1.4}$$

After the spontaneous symmetry breaking the system is projected in a neutral vacuum state. The scalar particle associated to the Higgs field will then be a neutral boson of mass $m_H = \sqrt{-2\mu^2}$, called Higgs boson.

The masses of the fermions rise from *Yukawa* terms:

$$\mathcal{L}_{Yukawa} = \sum_f c_f \overline{\psi}^f_L \psi^f_R \Phi \tag{1.5}$$

The mass of any single fermion m_f will be proportional to v and to the Yukawa coupling c_f.

[1] The covariant derivative for a field A_μ with charge q is defined as $D_\mu = \partial_\mu - iqA_\mu$.

Looking at the whole SM Lagrangian, we enumerate 18 free parameters:

- **in gauge sector**: 3 coupling constants (α_s, g, g');
- **in Higgs sector**: Higgs mass m_H, Higgs parameter μ^2;
- **in Yukawa sector**: 4 mixing parameters ($\theta_{1,2,3}$, δ_{13}), 9 fermion masses m_f.

1.1.2 Limits of the Standard Model

The first experimental evidences of the Standard Model were the discoveries of neutral currents (Gargamelle, 1973 [4]) and of $W^{\pm}$ and Z^0 (UA1, UA2, 1983 [5]). During its data taking (1989 - 2000), the Large Electron-Positron collider (LEP) at CERN allowed extraordinary precise measurements of many SM parameters[6] (figure 1.1), all of them in perfect agreement with the theoretical predictions. Similar measurements have been performed by TeVatron at Fermilab[7] and Large Hadron Collider (LHC) at CERN. Recent efforts led the ATLAS and CMS collaborations at LHC to the discovery of a new boson with mass between 125GeV/c^2 and 127GeV/c^2 which seems to be consistent with a Higgs boson (figure 1.2) [8][9].

Despite these phenomenological successes, the SM suffers some formal inconsistencies and leaves many theoretical questions unanswered: the motivation of the three-families pattern, for example, or the origin of CP violation (which is closely linked to the matter-antimatter asymmetry) and hierarchy of neutrino masses do not find a solution within the SM. Will will however focus on other issues, more closely related to the introduction of supersymmetry.

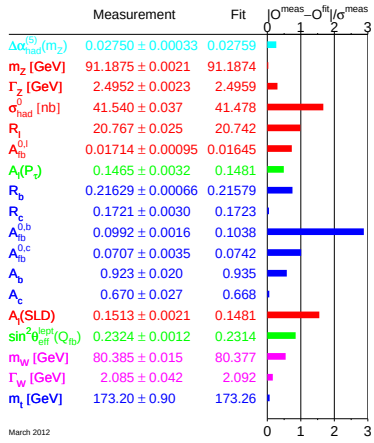


Figure 1.1: Pull between fitted and measured values of some SM parameters, performed by the LEP ElectroWeak Working Group (March 2012). From [6].

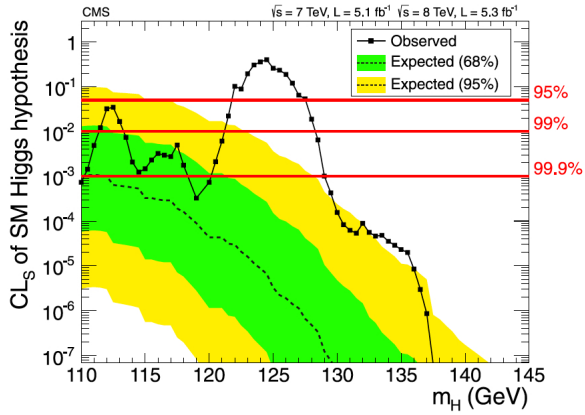


Figure 1.2: The CL_s values for the SM Higgs boson hypothesis as a function of the Higgs boson mass in the range 110145 GeV. From [8]

Gravity

Gravity is not included in the model. Looking at relative strengths of fundamental interactions it is clear that gravity is by far the most weak, so this lack in theory does not affects quantitative predictions at electroweak scale. Near the Plank scale (10^{19}GeV) however this is no more true, so including gravity in a consistent theoretical framework becomes mandatory. Unfortunately this framework can not be the SM. *Superstring* theories are elegant solutions to this problem, but they uppermost require a supersymmetric extension of the SM.

Dark matter

Evidence of the existence of dark matter rose from pioneering observations of clusters of galaxies by Zwicky (Zwicky,1933[10]) and from the studies on rotation curves of spiral galaxies, both pointing to mass distributions dramatically different with respect to the ones obtained considering visible objects only (stars, dusts). More recent studies on cosmic microwave background anisotropies (WMAP, 2010[11]) confirm that only 4% of the energy density of the universe is condensed in baryionic matter, while another 23% is carried by dark matter. It has to be made of weakly interactive non-baryionc particles. Relativistic neutrinos (the so called *hot dark matter* candidates) can be ruled out, since they could not ever lead to the formation of galaxies and clusters. Numerical solutions of Bolzmann equations for relic density of dark matter[12] point moreover to masses in the range $\mathcal{O}(10\text{GeV}) < m < \mathcal{O}(1000\text{GeV})$. They shouldn't be relativistic particles and are then called *cold dark matter*. The SM does not provide any good candidate for cold dark matter.

Unification of coupling constants

We can write down the Renormalization Group Equations (RGEs) for each of the free parameters of the SM, in particular for the three coupling constants[13]:

$$\frac{da_i}{dt} = b_i a_i^2 \tag{1.6}$$

where we used the notation

$$a_i \equiv \frac{\alpha_i}{4\pi} \equiv \frac{g_i^2}{16\pi^2}, \quad (g_i = g', g, g_s) \tag{1.7}$$

We should then study the evolution with scattering energy (Q^2) starting from an initial value (μ^2) as function of $t = \log(Q^2/\mu^2)$.
Coefficients b_i depend on the number of families of matter N_{fam} and of Higgs doublets N_{Higgs} in the theory. Equations 1.6 admit the simple solution

$$a_i(t) = \frac{a_i(0)}{1 - b_i t} \tag{1.8}$$

Considering, for the Standard Model, $N_{\text{fam}} = 3$ and $N_{\text{Higgs}} = 1$, we immediately get each b_i. As initial condition at $t = 0$ we can use the couplings at the Z-boson mass $\mu^2 = M_Z^2$, for which we have measurements from LEP[14].

$$\alpha_1(M_Z) = 0.017 \pm 0.001, \qquad \alpha_2(M_Z) = 0.034 \pm 0.001, \qquad \alpha_3(M_Z) = 0.118 \pm 0.003$$

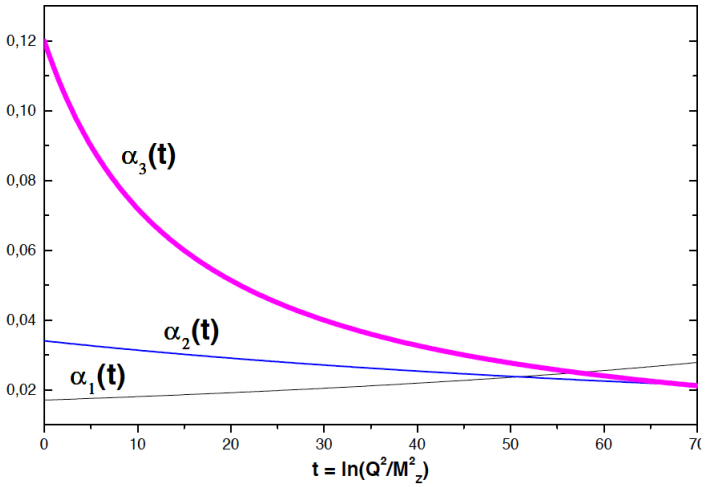


Figure 1.3: Evolution of the gauge couplings in the Standard Model.

The evolution of the gauge couplings are shown in Figure 1.3. We first focus on the difference in behavior for $\alpha_1 \sim \alpha_{QED}$ and $\alpha_3 = \alpha_s$. The first function increases at large momenta, while the second one has very large values at small momenta and approaches zero at large momenta (small distances).

Another important thing we have to notice is a clear unification pattern shown by the evolution of the constants. The three gauge couplings seem in fact to unify at an energy of the order of $10^{15} \div 10^{16}$GeV, which can be seen as a hint of the existence of a *great unified theory* (GUT). In this scheme, $M_{GUT} \sim 10^{16}$GeV should provide a natural and physically meaningful Λ_{cutoff} for renormalization scheme. However we have to warn about two issues:

- despite a qualitatively interesting unification pattern, the gauge couplings do not actually unify at exactly the same energy. The parameters of the equations 1.6 should be tuned in order to give an

exact unification: this can be done only introducing other fields in the model, that is, admitting the existence of a physics beyond the Standard Model manifesting between M_Z and M_{GUT};

- the unification scale pointed by the running couplings is very high compared to the electroweak symmetry breaking scale ($\sim 10^2$GeV). This creates a problem, called *hierarchy* problem, that will be now discussed.

Hierarchy problem

While the corrections to Higgs mass due to loops on SM fermions are maintained logarithmic by the chiral symmetry, the ones due to bosonic loops depend quadratically on the ultraviolet cutoff parameter. The one-loop corrected Higgs mass is than:

$$m_H^2(corr) \sim m_H^2 + c\Lambda_{cutoff}^2 \tag{1.9}$$

As just pointed, Λ_{cutoff} is of the same order of M_{GUT}. This would drive the Higgs mass very far from the bare value. The theory is still renormalizable, but the quantum corrections are several orders of magnitude larger than the the boson mass suggested by SM fits and measurements. This discrepancy is known as the *hierarchy problem*[15].
The only way to solve the problem without introducing any new physics beyond the SM is to imagine that every contribution to the correction at every loop level combines with the others in order to give acceptable mass values. Such a *fine tuning* however appears unnatural.

It should be noticed that bosonic and fermionic loops give rise to quadratic corrections whose c constants have opposite signs. We could

then imagine a fundamental boson-fermion symmetry, whose effect would be the cancellation of quadratic divergences at any level. If in fact there were equal numbers of fermions and bosons with the same couplings, equation 1.9 would have the form:

$$\delta m_H^2 \sim -|c|(\Lambda^2 + m_f^2) + |c|(\Lambda^2 + m_b^2) = |c||m_b^2 - m_f^2| \tag{1.10}$$

Such correction has a value not larger than the Higgs bare mass (hence naturally small), if

$$|m_b^2 - m_f^2| < 1\text{TeV}^2 \tag{1.11}$$

This naturalness argument is of fundamental importance, since is the only hint that such a symmetry, called *Supersymmetry*, could manifest at TeV scale, then could be reachable by the Large Hadron Collider.

1.2 Supersymmetry

Supersymmetry is a space-time symmetry first introduced by Wess and Zumino in the early '70s[16]. This symmetry transforms bosonic states into fermionic ones and vice-versa.

$$\mathcal{Q} \mid \mathcal{B}\rangle = \mid \mathcal{F}\rangle \tag{1.12}$$

$$\mathcal{Q}^\dagger \mid \mathcal{F}\rangle = \mid \mathcal{B}\rangle \tag{1.13}$$

The operator $\mathcal{Q}$ transforms scalar wave functions into spinors, which means it behaves like an anti-commutative spinor[17]. The commutation rules for $\mathcal{Q}$ are:

$$\{\mathcal{Q}, \mathcal{Q}^\dagger\} = P^\mu \tag{1.14}$$

$$\{\mathcal{Q}, \mathcal{Q}\} = \{\mathcal{Q}^\dagger, \mathcal{Q}^\dagger\} = 0 \tag{1.15}$$

$$[P^\mu, \mathcal{Q}] = \left[P^\mu, \mathcal{Q}^\dagger\right] = 0 \tag{1.16}$$

being P^μ the generator of Poincaré translations. From the last equation for $\mu = 0$, we deduce:

$$E_f \mid \mathcal{F}\rangle = P^0(\mathcal{Q} \mid \mathcal{B}\rangle) = \mathcal{Q}P^0 \mid \mathcal{B}\rangle = E_b\mathcal{Q} \mid \mathcal{B}\rangle = E_b \mid \mathcal{F}\rangle \tag{1.17}$$

then:

$$E_f = E_b \tag{1.18}$$

i.e. supersymmetric partners must have the same mass if the symmetry is conserved.

1.2.1 Minimal Supersymmetric Standard Model

The *Minimal Supersymmetric Standard Model* (MSSM) is the supersymmetric extension which adds the minimum number of new fields to the Standard Model. The minimal hypothesis consists of giving a single superpartner to every standard model particle, that is, more formally, replacing

Superfield	Fermions	Bosons	Particles
$\widehat{Q}$	(u_L, d_L)	$(\tilde{u_L}, \tilde{d_L})$	quarks/squarks
$\widehat{U}^C$	$\overline{u}_R$	$\tilde{u}^*_R$	
$\widehat{D}^C$	$\overline{d}_R$	$\tilde{d}^*_R$	
$\widehat{L}$	(ν_L, l_L)	$(\tilde{\nu_L}, \tilde{l_L})$	leptons/sleptons
$\widehat{E}^C$	$\bar{l}_R$	$\tilde{l}^*_R$	
$\widehat{\Phi}_u$	$(\tilde{h_u}^+, \tilde{h_u}^0)$	(ϕ^+_u, ϕ^0_u)	higgsinos/higgs
$\widehat{\Phi}_d$	$(\tilde{h_d}^0, \tilde{h_d}^-)$	(ϕ^0_d, ϕ^-_d)	
$\widehat{G}_i$	$\tilde{g}$	g	gluinos/gluons
$\widehat{W}_i$	$\tilde{W}$	W	winos/W's
$\widehat{B}$	$\tilde{B}$	B	bino/B

Table 1.3: List of the MSSM superfields.

every SM field with a single *superfield*. In particular we introduce a *chiral supermultiplet* for every fermionic and Higgs field in the Standard Model and a *gauge supermultiplet* for every gauge field in the Standard Model. The MSSM fields content is summarized in table 1.3.

The super-partners of fermions are spin=0 particles and are named by placing an *s* (from scalar) before the particle names (*squarks*, *sleptons*...). Bosons' super-partners are spin=1/2 fermions named *gauginos* and *higgsinos*. The super-partner of the hypothetical graviton is called *gravitino* and has an important role in many supsersymmetric models.

While within the SM only one Higgs doublet is necessary to give mass to all particles, for the MSSM things are more complicated. It can be demonstrated[17] that two different doublets are necessary. This lead to 5 different Higgs bosons

$$h, H, H^{\pm}, A$$

The first two are neutral and scalar (in particular the lightest one, h, would be very similar to a SM higgs boson), A is a pseudo-scalar, and $H^{\pm}$ are charged and scalar.

The spontaneous electroweak symmetry breaking projects the two Higgs doublets in two different vacua (v and $\overline{v}$) whose ratio

$$\tan\beta = \frac{\overline{v}}{v} \tag{1.19}$$

is a free parameter whose role will be discussed later.

Mixing of the neutral gauge particles (wino, bino, neutral higgsinos) results in four mass eigenstates, called neutralinos ($\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$, $\tilde{\chi}_4^0$) with the index increasing with mass. In the same way four charged gauginos, called charginos ($\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_2^{\pm}$), are obtained through mixing of the charged winos and charged higgsinos. Similar mixing happens between left and right squark and slepton states, although these mixing are often negligible, except for the pair ($\tilde{t}_R$, $\tilde{t}_L$), which mixes in ($\tilde{t}_1$, $\tilde{t}_2$).

We can write a supersymmetric Lagrangian density invariant under the standard model group[18]:

$$\begin{aligned}
\mathcal{L} = & \sum_i (D_\mu \mathcal{S}_i)^\dagger (D^\mu \mathcal{S}_i) + \frac{i}{2}\sum_i \overline{\psi}_i \not{D}\psi_i + \\
& + \frac{i}{2}\sum_A \overline{\lambda}_A \not{D}\lambda_A - \frac{1}{4}\sum_A F_{\mu\nu A}F^A_{\mu\nu} + \\
& - \sqrt{2}\sum_{i,A} g\left[\mathcal{S}_i^\dagger t_A \overline{\lambda}_A \frac{1-\gamma_5}{2}\psi_i + \text{h.c.}\right] + \\
& + \frac{1}{2}g\sum_A \left[t_A \sum_i \mathcal{S}_i^\dagger \mathcal{S}_i\right]^2 + \mathcal{L}(\mathcal{W})
\end{aligned} \tag{1.20}$$

where ψ_i and $\mathcal{S}_i$ are the spinor and scalar components of the superfield labeled by i, $F^A_{\mu\nu}$ is the kinetic tensor for the gauge field A and λ a corresponding quantity for the gaugino, while t_A are the generetors of the symmetry related to the gauge field A. The first two rows in (1.20) are kinetic terms for every superfield, the third row and the first term of the last row are trilinear and quadratic couplings of fields. The last term $\mathcal{L}(\mathcal{W})$ depends on the so-called textitsuperpotential $\mathcal{W}$. In the MSSM the superpotential has form:

$$\mathcal{W} = \mu\widehat{\phi}_u\widehat{\phi}_d + \lambda_u\widehat{\overline{U}}\widehat{Q}\Phi_u + \lambda_d\widehat{\overline{D}}\widehat{Q}\Phi_d + \lambda_e\widehat{\overline{E}}\widehat{L}\Phi_d \tag{1.21}$$

While this particular superpotential respects both lepton and baryon number conservation, the most general form of superpotential can also contain elements contributing to lepton and baryion number violating interactions. This terms however would have important phenomenological consequences, leading, for example, to a proton lifetime many order of magnitude shorter than the actual exclusion limits. In order to forbid such terms a new multiplicative quantum number, called *R-parity* is introduced, and its conservation is imposed. The R-parity is defined as

$$R = (-1)^{L+3B+2s} \tag{1.22}$$

with L and B the lepton and baryon number and s the spin of the particle. Clearly, according to this definition, SM particles have $R = +1$ while supersymmetric particles have $R = -1$. The conservation of R-parity has some important consequences:

- no mixing between SM and SUSY particles is permitted, which implies that baryon and lepton number violating interactions are forbidden;
- in a collider experiment, SUSY particles can only be produced in pairs;
- once produced, a SUSY particle will decay into lighter particles following a chain containing at least one SUSY particle per step;
- the *lightest supersymmetric particle* (LSP) must be stable.

LSP as cold dark matter candidate

The last item implies that the cosmological relic of LSP could be a good cold dark matter candidate. Since dark matter particles have to be only weakly interacting, good SUSY candidates are sneutrinos and lightest neutralinos ($\tilde{\chi}_0^1$). The firsts are ruled out by direct searches, so $\tilde{\chi}_0^1$, which indeed is the LSP in quite extended areas in the parameter spaces of many SUSY model, remains the only good candidate.
If the LSP is a neutralino, moreover, when produced in a collision experiment it could not be directly seen in detectors but, like a heavy neutrino, it should produce high amount of missing energy.

Evolution of gauge couplings in the MSSM

RGEs for gauge couplings have still form 1.6 with different b_i and $N_{\text{Higgs}} = 2$. Evolution of the coupling constants are shown in figure 1.4. Since supersymmetric particles cannot contribute to the evolution of gauge couplings at low energies (under their masses), the curves will show a change in slope at energy M_{SUSY}[19]

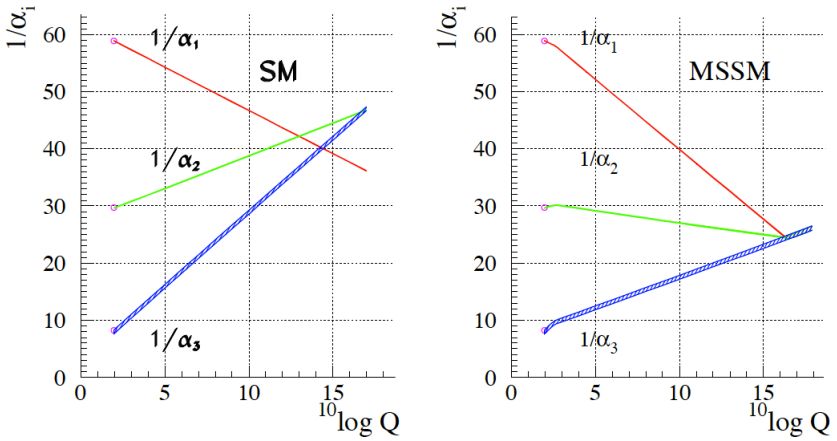


Figure 1.4: Evolution of the gauge couplings in the Standard Model.

$$M_{\text{SUSY}} = 10^{3.4\pm0.9\pm0.4}\text{GeV}$$

Moreover, as we can see in figure, in a MSSM scheme the three gauge couplings actually unify at the same point

$$M_{GUT} \sim 10^{16}\text{GeV}$$

that is compatible with many GUTs, like the Georgi-Glashow SU(5)[20]. It should be noticed that the unification of the three curves in a single point is absolutely not trivial because when introducing new particles all three curves are simultaneously influenced, leading to strong correlations between the slopes of the three lines. It can be shown, for example, that adding new generations or new Higgs doublets never yield unification[21].

1.2.2 Supersymmetry breaking and free parameters

Since the masses of the SM particles and their super-partners are not degenerate, equation 1.18 suggests that supersymmetry has to be a broken

symmetry. The solution of the hierarchy problem however points to mass differences not greater than the TeV scale. It is indeed the only theoretical limit we have about SUSY masses.

Up to now we do not know anything about the supersymmetry breaking mechanism. We then use to parametrize our ignorance adding an explicit soft breaking term to the MSSM Lagrangian

$$\begin{aligned}\mathcal{L}_{\text{soft}} = &-\sum_i \mathbf{m}_i^2 \mathcal{S}_i^* \mathcal{S}_i - \frac{1}{2}\sum_j M_j \tilde{\overline{G}}_j \tilde{G}_j + \\ &- \left(\mathbf{a}_u \tilde{\overline{U}} \tilde{Q} H_2 - \mathbf{a}_d \tilde{\overline{D}} \tilde{Q} H_1 - \mathbf{a}_e \tilde{\overline{E}} \tilde{L} H_1\right)\end{aligned} \tag{1.23}$$

containing

- scalar mass terms ($\mathbf{m}^2 \mathcal{S}\mathcal{S}$), in which $\mathbf{m}$ are 3x3 matrices;
- gaugino mass terms;
- trilinear scalar interactions ($\mathbf{a}\mathcal{S}\mathcal{S}\mathcal{S}$), being $\mathbf{a}$ 3x3 coupling matrices.

The full MSSM Lagrangian density contains then three terms:

$$\mathcal{L}_{\mathbf{MSSM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\mathcal{W}} + \mathcal{L}_{\text{soft}} \tag{1.24}$$

The first two terms do not introduce new free parameters with respect to the SM, since even SUSY particles interactions are ruled by the usual gauge couplings. The only new parameter is the additional VEV related to the second Higgs doublet (or, the more frequently used $\tan\beta$). The soft term instead contains many new parameters (every M, $\mathbf{m}$ and $\mathbf{a}$). Within the MSSM the total number of free parameters turns out to be 124[22]. It is a dramatically high number of degrees of freedom, since it introduce an unacceptable arbitrariness in the theory.

Constrained MSSM

A significant reduction of the number of free parameters can be performed introducing some constrains in the MSSM. A *constrained MSSM* (cMSSM) is a model in which is assumed that the symmetry breaking takes place in a "hidden sector" having no direct relation to the MSSM sector. The breaking is mediated to the observable sector by a particular heavy mediator X. The nature of the mediation is the distinctive feature of different models. Between the most studied we enumerate gauge-mediated (GMSB), anomaly-mediated (AMSB)and gravity-mediated.
In the last case the gravity is the mediation interaction, so that $M_X = M_{\text{Plank}}$. The minimal of gravity-mediated models is called *minimal supergravity* (mSUGRA).

In cMSSM, soft breaking parameters are assumed to be unified at GUT scales. In particular for mass terms and trilinear coupling:

$$\begin{aligned} M_1 = M_2 = M_3 &= m_{1/2} \\ \mathbf{m}_Q^2 = \mathbf{m}_U^2 = \mathbf{m}_D^2 = \mathbf{m}_E^2 = \mathbf{m}_L^2 &= m_0^2 \mathbf{1} \\ m_{H_1}^2 = m{H_2}^2 &= m_0^2 \\ \mathbf{a}_i &= A_0 \mathbf{y}_i \end{aligned} \tag{1.25}$$

Thus there are only five free parameters left:

- m_0: Universal scalar mass;
- $m_{1/2}$: Universal gaugino mass;
- A_0: Trilinear coupling at GUT scale;
- $\tan\beta$: ratio of the Higgs VEVs;

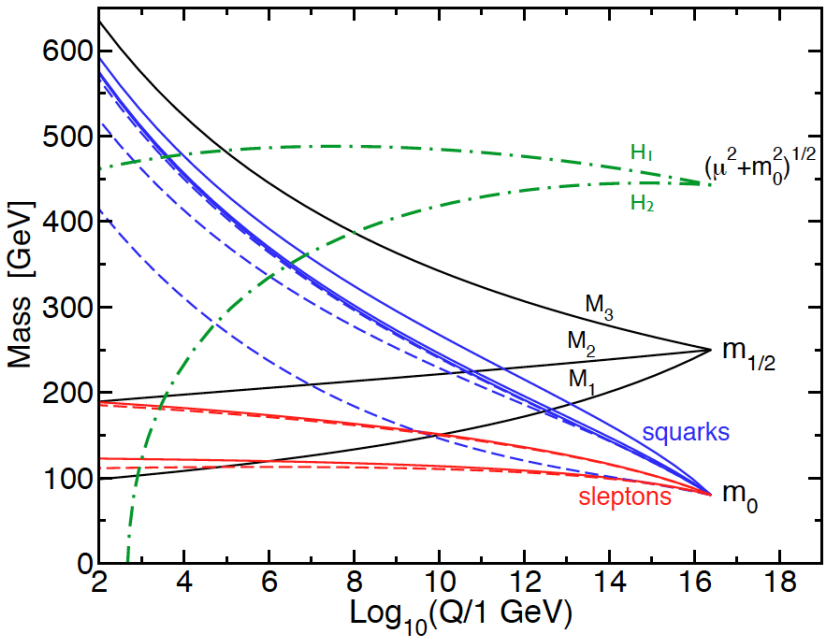


Figure 1.5: RGEs solutions for scalar and gaugino masses in a mSUGRA scenario.

- signμ: Sign of the higgsino mass parameter.

Driving the RGEs for **m** and **a** down to the TeV scale (fig. 1.5) permits to build SUSY mass spectra as function of these five parameters[23]. In the mSUGRA scenario:

Charginos and neutralinos

$$m_{\tilde{\chi}_2^0} \approx m_{\tilde{\chi}_1^\pm} \approx m_{\tilde{\chi}_1^0} \approx 0.8 m_{1/2} \tag{1.26}$$

$$m_{\tilde{\chi}_3^0} \approx m_{\tilde{\chi}_2^\pm} \approx m_{\tilde{\chi}_4^0} \approx \mathcal{O}(|\mu) \tag{1.27}$$

Gluino

$$m_{\tilde{g}} \approx 2.4 m_{1/2} \tag{1.28}$$

Sleptons

$$m_{\tilde{l}_L}^2 \approx m_0^2 + 0.54 m_{1/2} + \left(-\frac{1}{2} + \sin^2\theta_W\right) m_Z^2 \cos 2\beta \tag{1.29}$$

$$m_{\tilde{\nu}}^2 \approx m_0^2 + 0.54 m_{1/2} + \frac{1}{2} m_Z^2 \cos 2\beta \tag{1.30}$$

$$m_{\tilde{l}_R}^2 \approx m_0^2 + 0.15 m_{1/2} - m_Z^2 \sin^2\theta_W \cos 2\beta \tag{1.31}$$

Squarks

$$m_{\tilde{u}_L}^2 \approx m_0^2 + (c + 0.5) m_{1/2} + \left(\frac{1}{2} - \frac{2}{3}\sin^2\theta_W\right) m_Z^2 \cos 2\beta \tag{1.32}$$

$$m_{\tilde{d}_L}^2 \approx m_0^2 + (c + 0.5) m_{1/2} + \left(-\frac{1}{2} + \frac{1}{3}\sin^2\theta_W\right) m_Z^2 \cos 2\beta \tag{1.33}$$

$$m_{\tilde{u}_R}^2 \approx m_0^2 + (c + 0.07) m_{1/2} + \frac{2}{3}\sin^2\theta_W m_Z^2 \cos 2\beta \tag{1.34}$$

$$m_{\tilde{d}_R}^2 \approx m_0^2 + (c + 0.02) m_{1/2} - \frac{1}{3}\sin^2\theta_W m_Z^2 \cos 2\beta \tag{1.35}$$

where $4.5 < c < 6.5$.

The masses of gluino and lightest gauginos depend only on the universal gaugino mass $m_{1/2}$, while the mass scale for scalar supersymmetric particles are mainly ruled by m_0. $m_{1/2}$ also determines the mass splitting between right and left-handed scalars. These two parameters are then the most important in developing any search strategy and that is why exclusion limits are often shown in the $m_0 - m_{1/2}$ plane (also called cMSSM plane).

It has also to be noticed (from figure 1.5) that the parameter $\mu^2 + m_{H_2}^2$ runs to a negative value providing a natural way of generating a right shaped Higgs potential for the electroweak symmetry breaking.

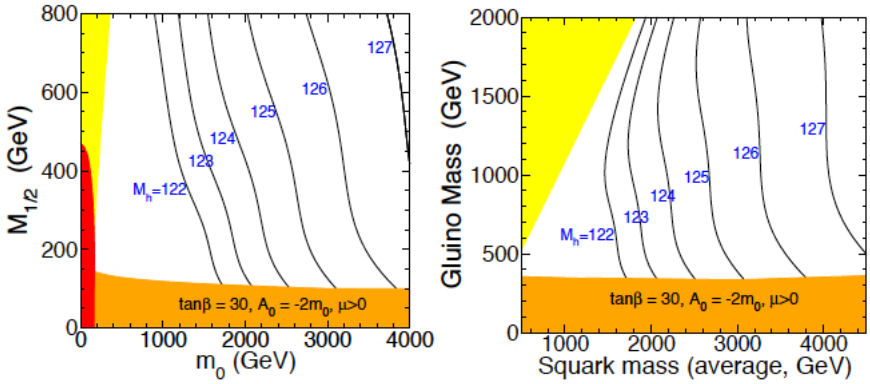


Figure 1.6: h mass value in a mSUGRA scenario represented in the mSUGRA plane (left) and in the squark average - gluino mass plane. In these plots $\tan\beta = 30$ and $A_0 = -2m_0$ are imposed.

Notes on SUSY lightest Higgs boson

If the boson recently discovered at CERN is actually a Higgs boson, under the hypothesis of the existence of supersymmetry it would have to be the lightest supersymmetric scalar neutral boson h. Although it is absolutely premature, it is very interesting to notice that this discovery should lead to strong constrains on the cMSSM parameter space (Figure 1.6) and on the SUSY breaking model itself (Figure 1.7). In fact since many supersymmetric models require a light h ($\lesssim 120\text{GeV}/c^2$), it seems that for this models the one just discovered should be a "border-line boson". Some theoretical studies[24] show that a h boson with mass $m_h = 126\text{GeV}$ virtually rules out many constrained MSSM models, leaving mSUGRA as the most promising candidate.

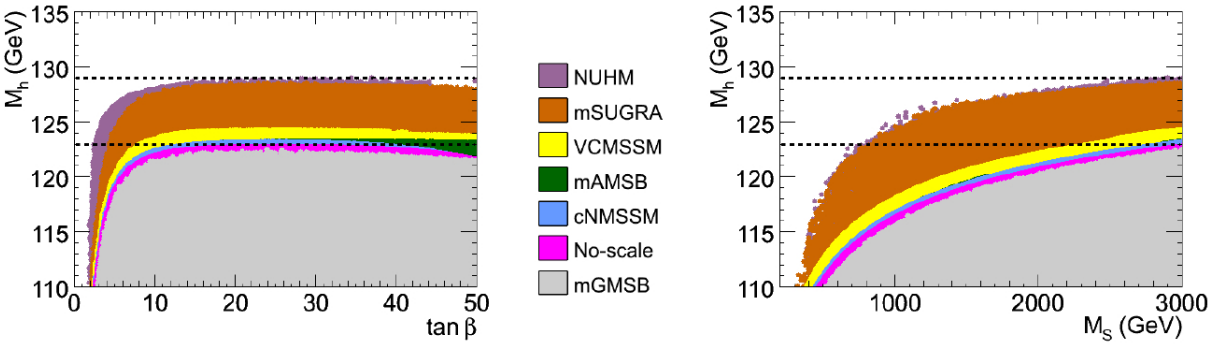


Figure 1.7: The maximal h mass value, as function of $\tan\beta$ and of the universal scalar mass (here indicated with M_S), for minimal Anomaly Mediated, Gauge Mediated, mSUGRA and some non minimal extensions.

2

The CMS detector at LHC

2.1 The Large Hadron Collider

The *Large Hadron Collider* (LHC)[25] is a two-ring collider installed at CERN (European Center for Nuclear Researches) in the $\sim 27km$ underground tunnel originally hosting the Large Electron-Positron collider (LEP) which operated there until 2000. The tunnel is situated under the French-Swiss border near the city of Geneva at a depth between 45m and 170m (figure 2.1).

The collider has been designed to provide $p-p$ collisions at a center-of-mass energy $\sqrt{s} = 14$ TeV with a peak luminosity of $L = 10^{34} cm^{-2} s^{-1}$, and $Pb-Pb$ collisions at $\sqrt{s} = 2.8$ TeV/A.

The proton beams are pre-accelerated in the Proton Sincrotron (PS) and in the Super Proton Sincrotron (SPS) to an energy of 450 GeV and then injected to the LHC for the final acceleration, performed using RF electric fields of maximum frequency of 400 MHz. Beams are maintained in orbit by 1232 superconducting magnets providing a peak dipole field of $8.3T$. The dipoles are operated at a temperature of $1.9K$. At design

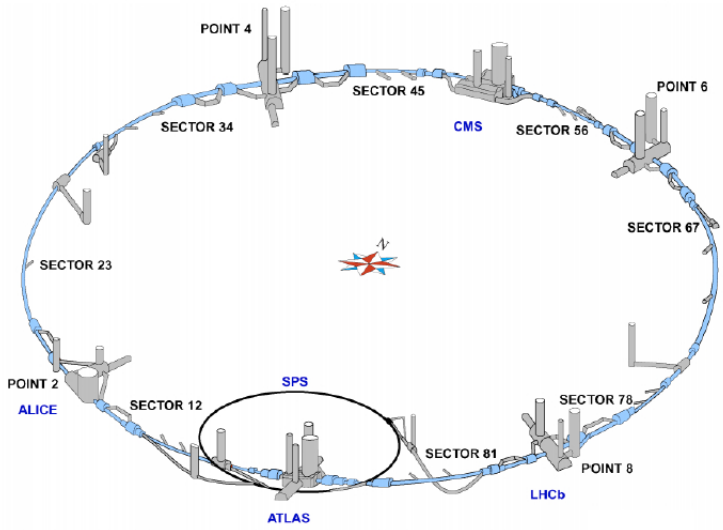


Figure 2.1: A scheme of LHC underground tunnel with interaction points and main experiments.

luminosity each beam will be made of 2835 bunches, of $\sim 10^{11}$ protons each, crossing every 25ns.

The luminosity delivered by a collider is based only on the beam parameters. For gaussian-shaped bunches, it is given by:

$$L = \frac{kN^2 f}{4\pi\sigma_x\sigma_y} F \tag{2.1}$$

where k is the number of colliding bunch pairs, N the number of protons per bunch, f is the revolution frequency, σs are the beam sizes at interaction point and F is a factor related to the crossing angle. Luminosity is a very important quantity since it is strongly related to the rate of a particular process:

	Design	**July 2012**
Beam Injection Energy (TeV)	0.45	0.45
Beam Energy (TeV)	7	4
Bunches per beam	2808	1374
Beam size (μm)	16	18
Protons per bunch	1.15×10^{11}	1.5×10^{11}
Stored beam energy (MJ)	362	110

Table 2.1: LHC operation conditions.

$$R = L \times \sigma. \tag{2.2}$$

This means that having a high luminosity permits to investigate low cross sections processes. The luminosity integrated over the operation time (*integrated luminosity*, $\mathcal{L}$) is usually measured as the inverse of a cross section.

LHC operation started in late 2009 with a pilot run at the injection beam energy of 450 GeV. In 2010 the center of mass energy was increased from 900 GeV to 2.36 TeV and then to half the design value, 7 TeV, reaching an integrated luminosity of $\mathcal{L} \sim 50pb^{-1}$ (Fig.2.2-a). During the whole 2011, LHC operated ad $\sqrt{s}$ =7 TeV. The $\mathcal{L}$ recorded by CMS experiment was 6.1 fb^{-1} (Fig.2.2-b). The 2012 $p-p$ run is being performed at $\sqrt{s}$ =8 TeV and will last till December. Up to August 2012 an integrated luminosity $\mathcal{L} \sim 7fb^{-1}$ has been delivered, that is expected to be about one third of the whole 2012 luminosity (Fig.2.2-c). After a technical stop in 2013, LHC will be driven at its design energy. A comparison between mid-2012 and design operation conditions is summarized in table 2.1[26].

Seven experiments are put along the tunnel at four interaction points

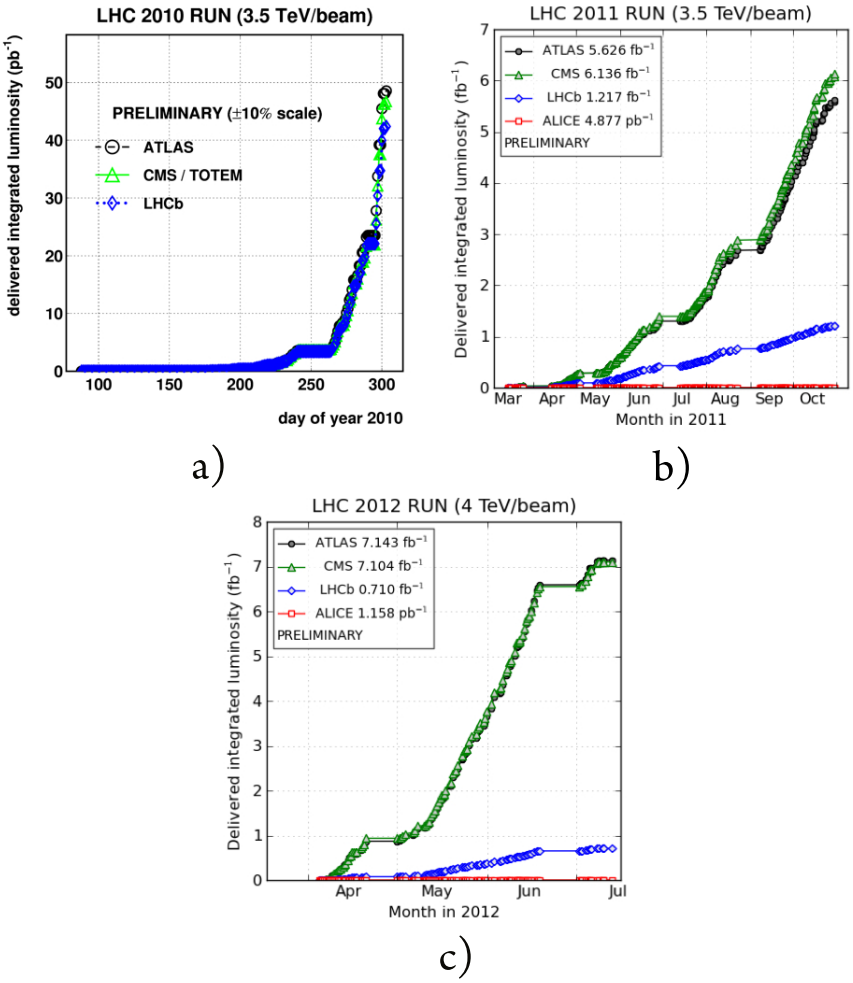


Figure 2.2: LHC recorded luminosities in 2010 (a), 2011 (b) and first half of 2012 (c).

- **ALICE** (A Large Ion Collider Experiment): dedicated to studies of very high density strong interacting matter produced in heavy ion collisions[27];
- **ATLAS** (A Toroidal Lhc ApparatuS) and **CMS** (Compact Muon Solenoid): two general-purpose experiments with similar particle physics tasks, but adopting different experimental philosophies[28];
- **LHCb**: dedicated to studies of CP violating interaction in heavy quarks sector[29];
- **LHCf** (LHC-forward): two small electromagnetic calorimeters in the very forward regions of ATLAS interaction point, dedicated to π^0 cross section measurements for cosmic rays physics[30];
- **MoEDAL** (Monopole and Exotic Detector At Lhc): specific purpose experiment for magnetic monopoles detection[31];
- **TOTEM** (TOTal Elastic and diffractive cross section Measurements): it shares the CMS interaction point. The detector is dedicated to measurement of total cross section and elastic scattering[32].

2.2 Compact Muon Solenoid

CMS (*Compact Muon Solenoid*) is one of the four large experiments at LHC, dedicated, together with the ATLAS detector, to multi-purpose particle physics measurements. It is located at Interaction Point 5, in France, 100 m below ground near the village of Cessy.
The detector (Figure 2.3) has a cylindrical shape with length of 21m and 15m of diameter. Its dimensions are more compact than the ATLAS ones but it is heavier (it weights approximately 12500t). The inner part of the

detector is put inside a solenoid producing a 4T magnetic field parallel to the beam lines. Outside the solenoid, the muon detectors are installed in the iron return joke frame. The experiment has almost a full solid angle coverage and it can be divided in three sectors: the central *barrel* and the two *end-caps*, installed at both sides of the barrel.

The design of the detector is motivated both by physics tasks and the specifications of the LHC. In order to accomplish the goals required by new physics, the experiment has been designed having:

- good inner tracking and pixel detection, with good τ and b-jet tagging;
- a very good, fast and redundant muon detector;
- high resolution and high granularity calorimeters;
- good missing transverse energy resolution.

Moreover, the challenging working conditions of the LHC make some characteristics mandatory:

- at design luminosity about 10^9 $p-p$ collisions per second are expected, thus a fast readout system and an efficient trigger are required;
- high granularity and time resolution to handle with underlying events and pile-up;
- high resistance to radiation damage for sensors and readout electronics.

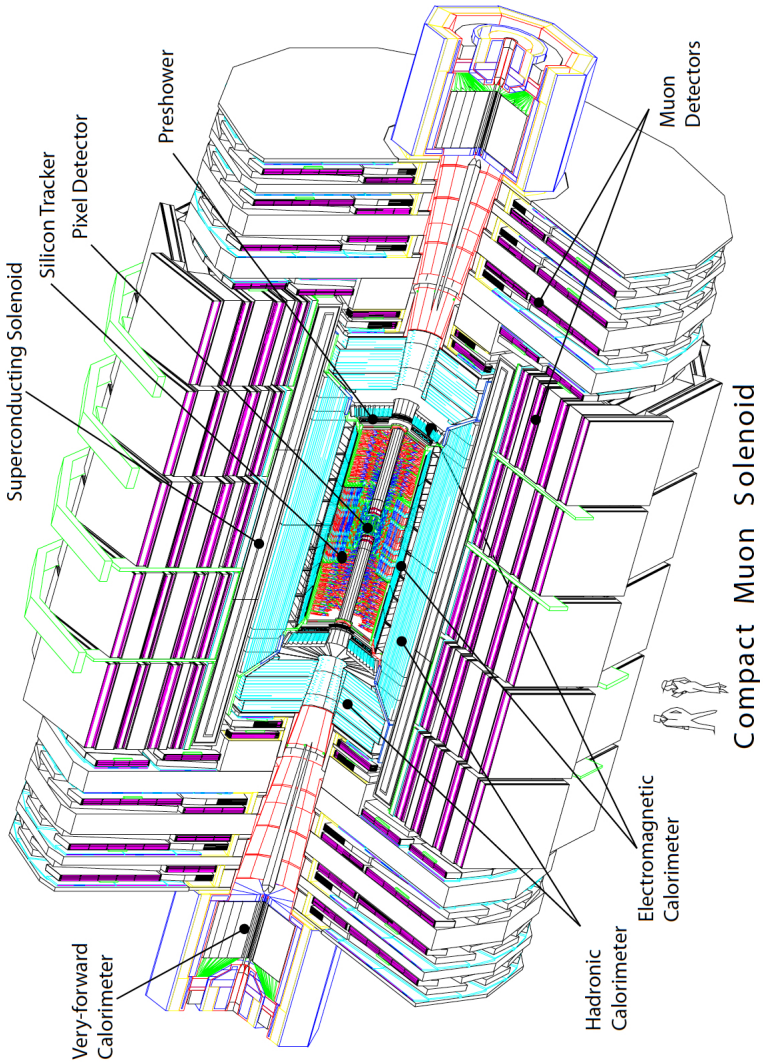


Figure 2.3: A sketch of the CMS detector showing the inner components.

The CMS sub-detectors are deeply described in ref. [33]. We now introduce just the main features of each sub-detector, from the interaction point to outer layers. We start with the tracking system (Pixel detector and Silicon Strip Tracker). We then describe the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). Outside the superconducting solenoid coil there is the muon system, made of Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Drift Tubes (DT).

CMS conventions and important variables

A right-handed coordinate system is introduced with origin in the geometrical center of the apparatus. The x-axis is oriented horizontally and points to LHC center. The y-axis is oriented vertically pointing upwards. The z-axis is then directed along the counterclockwise direction of the LHC ring. With this frame definition the azimuthal angle $-\pi < \phi \leq \pi$ is determined in the $x-y$-plane while the polar angle $0 \leq \theta \leq \pi$ is defined with respect to the positive z-axis. Once defined ϕ and θ we can build the *transverse momentum* and *transverse energy*:

$$p_T = |\mathbf{p}| \sin\theta$$

$$E_T = E \sin\theta$$

The *missing transverse energy* (MET) is defined starting from the measured energy of all visible particles:

$$\text{MET} = \sqrt{\left(\sum_i E_{i,x}\right)^2 + \left(\sum_i E_{i,y}\right)^2}$$

A very useful variable in particle physics is the pseudo-rapidity η defined as

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

A pseudo-solid angle ΔR is defined as a distance in the $\eta-\phi$-plane:

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

2.2.1 Tracking system

The innermost sub-detector is the tracking system. It consists of a pixel detector, whose main task is to define the position of the vertices, and a silicon strip tracker which performs precise measurements of momentum of charged particles, starting from their bending due to a homogeneous 3.8T magnetic field.
The system is totally based on solid state technology. It represents indeed one of the most massive ever usage of silicon detectors[34].

Near the interaction vertices a huge track multiplicity is recorded due to $\gtrsim 1000$ minimum bias charged particles produced per crossing. High granularity, large hit redundancy and powerful track reconstruction algorithms are then necessary, together with materials resistant to high radiation doses.
The material budget moreover has to be as as thin as possible, in order to avoid photon conversions and bremsstrahlung that could affect the ECAL performances. The thickness encountered by the particle obviously depends on η, as shown in figure 2.4. This last requirement leads to constrains on material budget and on the number of active layers.

The design adopted is shown in figure 2.5. The inner part is occupied by the pixel detector, while the Silicon Strip detector is divided in *inner* and *outer* regions along the radial coordinate and in *barrel* and *end-caps* (or disks) along η.

Pixel Detector

A vertex detector gives the position of primary and secondary vertices and the first seeds for track reconstruction. The technology adopted in CMS is that of pixels. Every pixel has a surface of $100 \times 150 \mu\text{m}^2$ and gives a

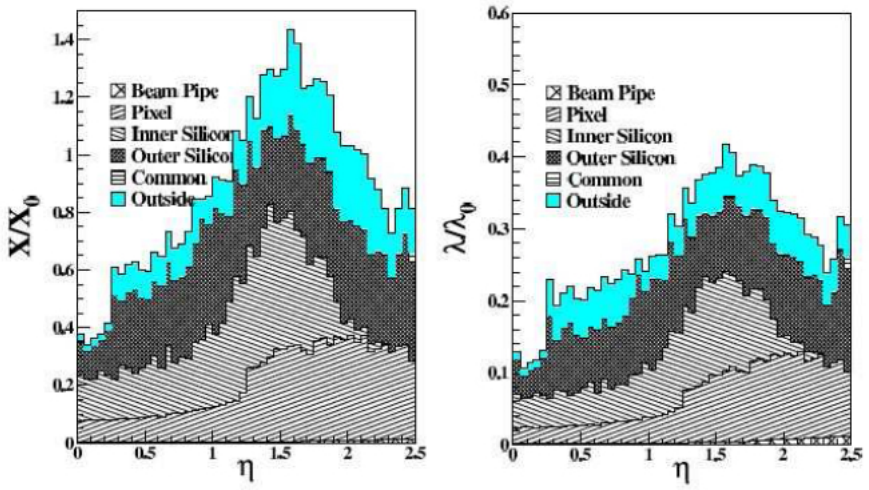


Figure 2.4: The tracker's material thickness in radiation lengths as function of pseudo-rapidity. Contributions from active materials and other materials are shown.

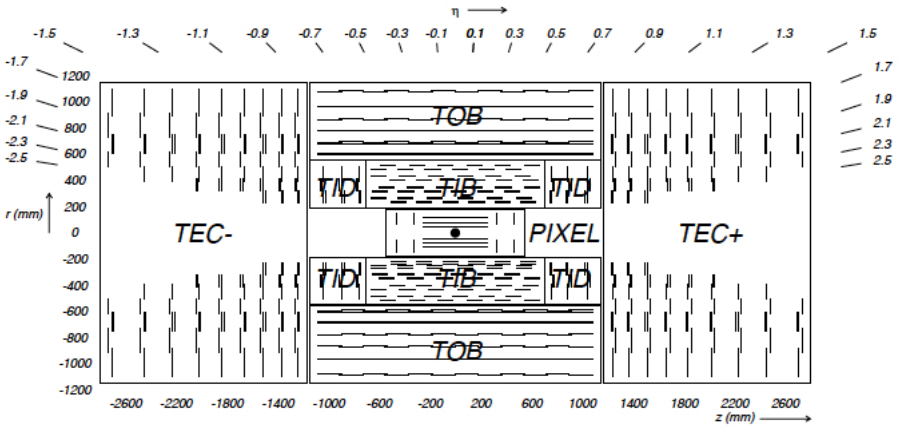


Figure 2.5: A scheme of the tracking system.

digital signal related to the hit of a particle leaving an amount of energy over a threshold, fixed to a signal-to-noise ratio of 5. The total active area of 1m^2 has in total 66 millions of pixels arranged in 1440 modules distributed over three barrel layers ($|\eta| < 1.5$) with radii 4cm, 7cm and 10cm and two disks per side at $z = \pm 34.5\text{cm}$ and $z = \pm 46.5\text{cm}$. At least 3 hits are guaranteed for particles with $|\eta| < 2.4$.

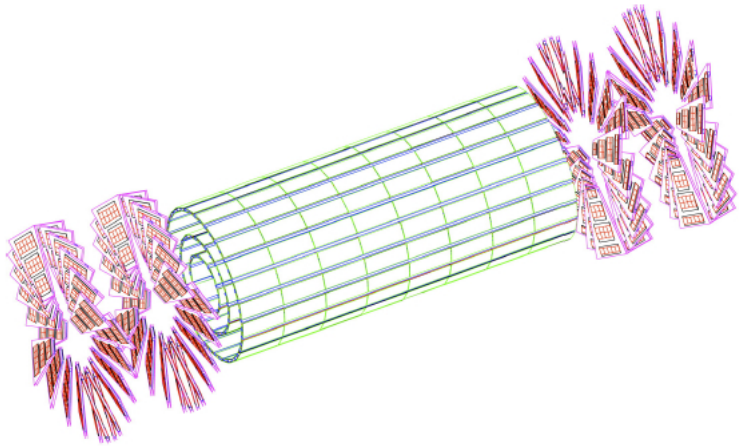


Figure 2.6: A 3D view of the pixel detector.

The space resolution in the barrel region is 10μm along the $r-\phi$-plane and 15μm along z. In the end-caps regions is 15μm along the $r-\phi$-plane and 20μm along z.

Silicon Strip Detector

Outside the pixel detector a silicon strip detector covers a length of 5.6m and a radius between 0.2m and 1.2m, with a total surface of 225m^2. The barrel region is equipped with ten layers of microstrips (four of them are

Resolution	ϕ	r(double-sided)
TIB	23-34μm	230μm
TOB	35-52μm	530μm
TID/TEC	15μm (low-r)	
	50μm (high-r)	

Table 2.2: Resolution of the SSD.

double-sided), while every end-cap region ($1.5 < |\eta| < 2.5$) has three inner mini-disks and nine outer disks. The detector is divided in four parts:

- **Tracker Inner Barrel (TIB)**, containing the four innermost layers of the barrel region. Strips are 300μm tick with a length of $7 \div 12.5$cm and are parallel to the beam pipe;

- **Tracker Outer Darrel (TOB)**, with the outer six barrel layers. Strips are ~ 20 cm long and 500μm tick;

- **Tracker Inner Disks (TID)**, containing the three mini-disks. Each disk has radial strips with thickness 300μm;

- **Tracker End-Caps (TEC)**, with nine external disks, each divided in seven rings. The first three rings are equipped with 300μm tick strips while the last four has 500μm strips.

The space resolutions for each region are summarized in table 2.2. The two faces of the double-side detectors are mounted back-to-back at a stereo angle of 100mrad, providing a 3D view of the impact point.

The minimum number of hits for a almost straight track is always > 9. The tracking efficiency has been measured[35] to be $> 98\%$ for muons from

J/ψ decays. The transverse momentum resolution depends on p_T and η.

$$\frac{\delta p_T}{p_T} = \begin{cases} (1.5 p_T/\text{GeV} + 0.5)\%, & \text{for } |\eta| < 1.6 \\ (3.0 p_T/\text{GeV} + 0.5)\%, & \text{for } |\eta| > 1.6 \end{cases} \tag{2.3}$$

2.2.2 Calorimetry

A calorimeter gives a signal proportional to the energy released by a particles passing through it. For a complete energy detection the particle must be completely stopped in the active material. The energy loss by photon or electrons involves other processes than that for hadrons. Thus different kind of detectors are necessary: *electromagnetic calorimeters* (ECAL) for photons or electrons showers, *hadronic calorimeters* (HCAL) for baryons and mesons.

Between the tracker and the coil in CMS an homogeneous ECAL is placed, followed by a sandwich HCAL.

Electromagnetic calorimeter

The CMS ECAL[36] is an homogeneous calorimeter made of 61200 lead tungstenate ($PbWO_4$) crystals in the barrel and 7324 crystals in the end-caps. Each crystal has a projective geometry: in the barrel region, they are arranged in towers with internal surface of 22x22mm^2, external surface of 26x26mm^2 and depth of 230mm, corresponding to $25.8X_0$ (radiation lengths). In the end-caps, towers are 28.6x28.6mm^2 inside, 30x30mm^2 outside and 220mm deep ($24.7X_0$). Towers are moreover arranged in 5x5 clusters.

In the region between $1.65 < |\eta| < 2.6$ a 20cm thick preshower detector is placed between the tracker and the end-caps' ECAL. It is made of two lead radiators interleaved with two silicon strip detectors. The preshower

add some radiation length to the ECAL and enhances the $\gamma - \pi_0$ discrimination.

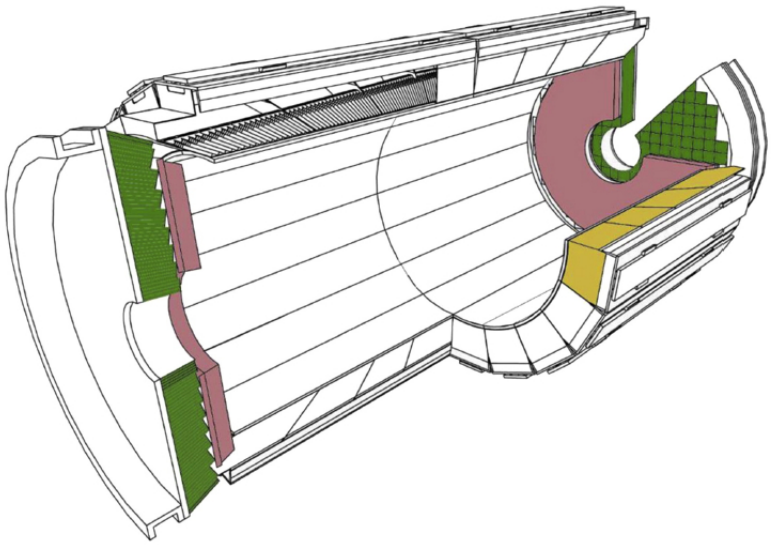


Figure 2.7: A scheme of the ECAL.

The lead tungstenate has been chosen because of its fast scintillation response and of its hardness to radiation damage. It has moreover a small Moliere radius.

The energy resolution follows the function:

$$\left(\frac{\delta E}{E}\right)^2 \approx \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \tag{2.4}$$

where the energy is expressed in GeV. The first term is a Poisson term, with $a = 2.8\%$ (barrel), 5% (end-caps). The second is a noise contribution with $b = 125$ MeV (barrel), 500 MeV (end-caps). The constant term

$c = 0.18$ is obtained from early measurements[37], and is lower than the expected value of 0.3.

Hadronic calorimeter

Between the ECAL and the solenoid (at 1.77m $< r <$ 2.95m), the CMS HCAL is a sandwich calorimeter with layers of copper radiators and plastic scintillators[38]. Copper is a high radiation length material without ferromagnetic properties (so suitable for working in a strong magnetic field). In the barrel region the HCAL is divided in 16(η)x18(ϕ) towers of thickness 5.4λ (absorption lengths). This depth is not sufficient to fully contain an hadronic shower, so two other layers are placed outside the coil at $|\eta| < 1.26$, the so called *Outer Hadron Calorimeter* (HO).

A similar pattern (without HO) can be observed in the end-caps ($1.3 < |\eta| < 3$). The granularity is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for $|\eta| > 1.6$. The total λ in this region is 10.

The energy resolution can be parametrized as:

$$\left(\frac{\delta E}{E}\right)^2 = \left(\frac{85\%}{E}\right)^2 + (7.4\%)^2 \tag{2.5}$$

for energies between 30GeV and 1TeV.

At $\pm$11.15 m from the interaction point the Hadron Forward Calorimeter (HF) covers a pseudo-rapidity up to 5. It is made of steel absorbers and quartz fibers as active material. Such a robust design is mandatory since in the forward region the hadron rate is very high.

2.2.3 Muon detectors

The region outside the coil is completely dedicated to muon identification. Muons are the less stopped in material between all charged particles and are the only which can trespass the calorimeters. CMS has a very challenging and redundant muon system[39]. It consists of three different kind of detectors: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). They are arranged as shown in figure 2.8. The overall active surface is 25000m^2.

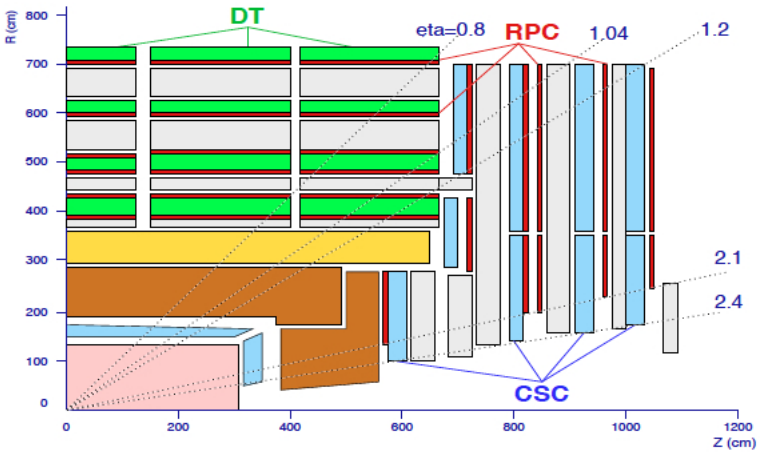


Figure 2.8: A sketch of the muon system in CMS.

At $|\eta| < 1.2$ four layers of DT are alternated to the iron wheels of the return yoke. The muon system is then still inside a return magnetic field of $\sim$ 2T, oriented backwards with respect to the inner field. Muons in this region are bended in the reverse direction drawing a characteristic S-shaped trajectory. The space resolution in the central region is $\sim 100\mu$m.

In the end-caps ($0.9 < |\eta| < 2.4$) CSCs are mainly used because of

their better performances in higher multiplicity environments. As in the central region, we count four layers of chambers perpendicular to the beam line. The space resolution is $\sim 200\mu$m.

RPCs are used both in the barrel and in the end-caps. Their very fast response (time resolution is 3ns) makes them a perfect trigger system for the CSCs and DTs.

A combined determination of the muon momentum using the tracker and the muon system leads to a very good resolution:

$$\frac{\delta p_T}{p_T} = a\sqrt{p_T/\mathrm{TeV}} \tag{2.6}$$

with a depending on pseudo-rapidity. In the barrel, for $p_T = 1$ TeV a resolution of 4% is obtained, degrading to 10% in the end-caps. For $p_T = 10$GeV, the central resolution is 0.5%. First data from cosmic rays and from collisions confirm that the muon system is working at design resolution[40][41].

2.2.4 Trigger and Computing

At design luminosity ($10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$) about 10^9 collisions per second are expected, corresponding to a data-stream of $\sim$ 100TByte/s. By now, a computing system capable to manage such amount of data does not exist. Moreover not every event contains interesting information from a physical point of view, so the main task of the trigger system is to perform a rapid event selection reducing the event rate to 100Hz (a reduction of a factor 10^6). The CMS trigger consists of two steps: a hardware trigger (*Level 1* - L1) and a software based *high level trigger* (HLT)[42].

Level 1 trigger

A first fast selection is preformed via hardware, analyzing only few topical information from muon system and calorimeters with reduced granularity. As result of this raw but very fast analysis some so called primitives (for example muons or calorimetry deposit over a threshold) are required for a positive trigger response. The level 1 trigger conditions are common to almost every analysis, so having a static hardware trigger does not represent an unacceptable limit. On the other hand the gain in speed with respect to a more ductile software system is huge.
The trigger is made of a chain of electronic steps, each performing a particular selection with rates compatible with the bunch-crossing time. The whole maximum decision time is 3.2μs and is limited by the maximum latency time of data in the pipeline memories integrated in the front-end electronics. The L1 output rate is limited to 30kHz, leaving a safety margin from the 100kHz design value.

High level trigger

At such a lower rate, a flexible software trigger can be used. The high level trigger runs on a filter farm with over 1000 CPUs each processing a single event. At this level the trigger uses information from all the sub-detectors at full resolution (the full event information is built after L1 decision by a *builder network* with a maximum data flow of 100 GBytes/s). A HLT decision takes less then 40ms and the final event rate is 100Hz, which is sufficiently small to be recorded for the off-line analysis[43].

Computing

In CMS $\sim$ 1PByte of data is produced every year. The HLT output events are reconstructed at a CERN facility called *Tier0*. Data are then distributed to several computing servers all over the world (*Tiers1* and *Tiers2*) for direct access. Every tier is connected to each other through the computing *Grid*. The access to any data in the grid can be done using the CRAB tool[44].

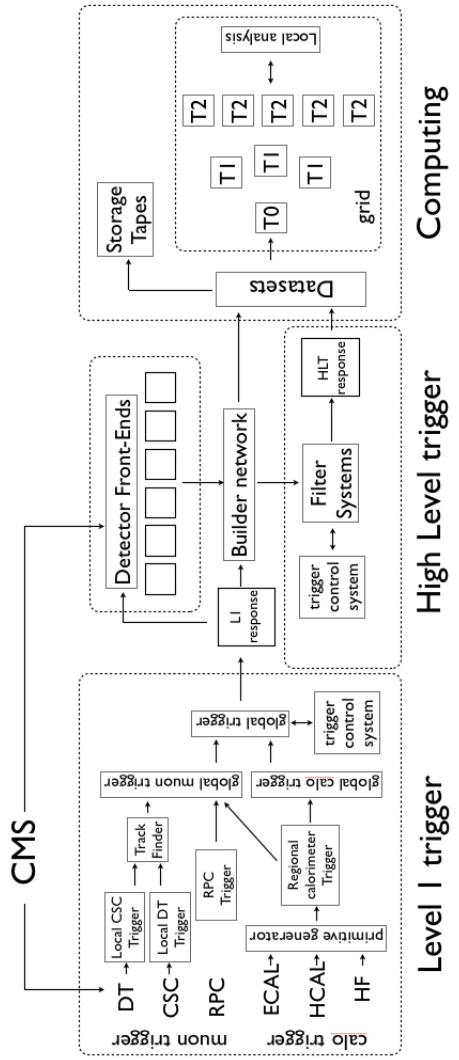


Figure 2.9: Diagram of CMS trigger and computing system.

3

Supersymmetry at CMS

As discussed in Chapter 1, the hierarchy problem points to the existence of some new physics beyond the Standard Model at TeV scale, i.e. at energies reachable by the LHC. Moreover, at design luminosity, the production rate for supersymmetric particles could be high enough to permit a discovery[1].

In this chapter a description of supersymmetry phenomenology at CMS is exposed, after a brief introduction about physics at hadron colliders and simulation of hard scattering processes. Different SUSY analyses are introduced together with their discovery power. A particular attention is dedicated to the opposite sign lepton searches, which are the main subject of this thesis.

3.1 Physics at Hadron Colliders

Hadron collisions are much more complex than electron ones, since protons are not elementary objects. A proton is made of gluons, valence quarks

[1]For example, assuming an inclusive cross section of $\mathcal{O}(1)$fb for a SUSY process, we expect $\mathcal{O}(10^2)$ events of new physics per year at design luminosity.

and sea quarks, each one carrying a fraction x of proton's momentum. Given an energy scale Q^2 the probability for a certain constituent i of carrying a momentum fraction x is given by a *parton density function* (PDF) $f_i^P(x, Q^2)$.

Proton PDFs can be obtained from experiments at other colliders, like HERA[45], and extrapolated to LHC energies[46] (Figure 3.1 a) and b)).

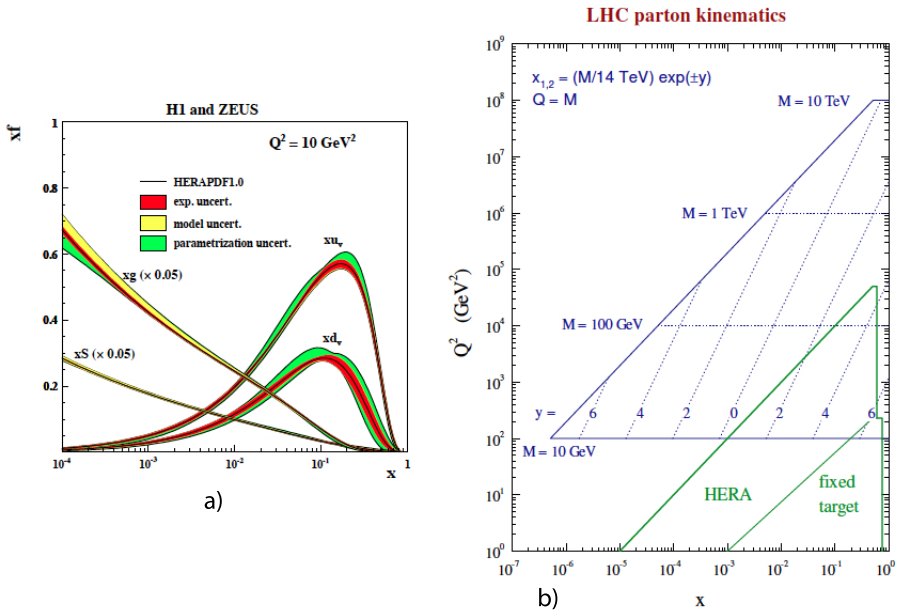


Figure 3.1: a) Gluon and parton PDFs measured by HERA at $Q^2 = 10\text{GeV}^2$. b) Scatter plot x vs. Q^2 for production of an object of mass M and rapidity y at LHC design energy. Current experimental constrains are also shown.

Gluon and parton PDFs have to be taken into account when calculating the cross section of a hard process. We consider for example the squark-gluino production:

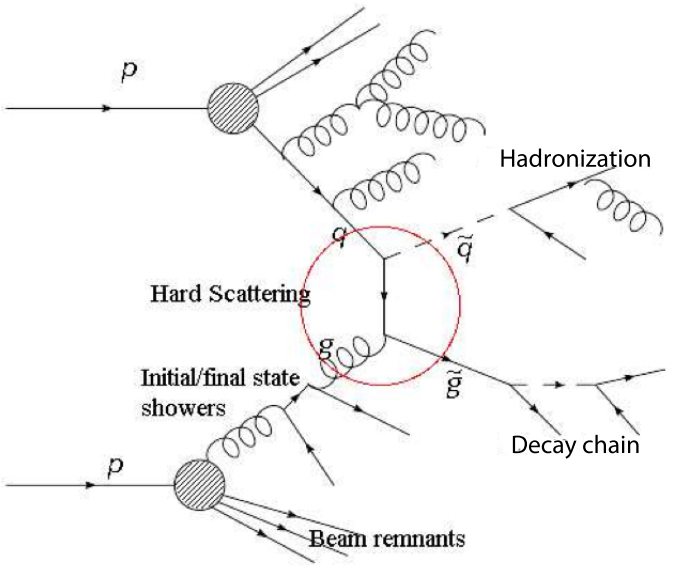


Figure 3.2: Full $p-p$ scattering including the hard SUSY process $q+g \rightarrow \tilde{q}+\tilde{g}$.

$$q + g \rightarrow \tilde{q} + \tilde{g} \tag{3.1}$$

Its cross section in a collision between proton A and proton B at a certain Q^2 can be expressed as function of PDFs $f_i^P(x, Q^2)$ and of the cross section of the elementary process ($d\hat{\sigma}_{(q+g\rightarrow\tilde{q}+\tilde{g})}$):

$$d\sigma_{(A+B\rightarrow\tilde{q}+\tilde{g}+X)} = \sum_{q,g} \int_0^1 dx_q \int_0^1 dx_g f_i^P(x_q, Q^2) f_i^P(x_g, Q^2) d\hat{\sigma}_{(q+g\rightarrow\tilde{q}+\tilde{g})} \tag{3.2}$$

In figure 3.2 a typical $p-p$ process is schematized, including the hard scattering 3.1, that is the circled part. All the other processes occurring are called *underlying event* and their knowledge is fundamental to correctly understand and simulate the detector's response.
The underlying event consists of

- **Initial and Final State Radiation** (ISR/FSR), also called *parton showering*. It is the radiation of gluons or quark pairs from any colored particle involved in the event. It is often studied developing perturbative QCD calculations.

- **Beam remnants**: partons that do not take part to the hard process produce color singlets giving an important contribution to the particle flux especially in the high-η/low-p_T regions.

- **Hadronization**. Every colored particle produced in decay chains, in ISR/FSR, or in beam remnants will arrange with others, producing hadrons. The study of hadronization processes requires non-perturbative QCD models.

3.2 Monte Carlo simulation

A full simulation of hard collisions consists of three steps: event generation, simulation of detectors and event reconstruction. The whole simulation chain together with reconstruction of events in CMS is performed within a framework called CMSSW (CMS-SoftWare)[47], containing several thousands of packages, including generators, a GEANT4 simulation of the experiment[48], and several analysis kits.

Event generation

At generation level, hard scattering process, parton showering and hadronization are simulated. In CMSSW two different generators are used: PYTHIA6 [49] and MADGRAPH[50].
PYTHIA6 calculates the hard scattering process without additional partons. Radiated gluons or quark pairs are modeled as soft partons. This produces, under certain condition, simulated sub-leading jets softer than actually observed. Event generators like MADGRAPH solve this problem using a different approach called *matrix element calculation*, including radiated partons (up to 9) to the hard process simulation.

Hadronization is simulated in PYTHIA6 only, using a phenomenological model called *string fragmentation*, where interactions between partons are modeled by colored string, whose breaking lead to the formation of hadrons.

Detector simulation

A detailed model of the CMS experiment (geometry and materials) is implemented in GEANT4. First of all every possible interaction of visible particles in a particular event is simulated, and the hits at sub-detectors are reproduced. In the second step, called *digitalisation*, the read-out electronic response to the hits is simulated. The final output has then the same form of any actual CMS data event and can be analyzed with the same tools.

For a complete simulation, two other effects have to be included: the trigger effect and the pile up interactions. They will be discussed in detail in the next chapter.

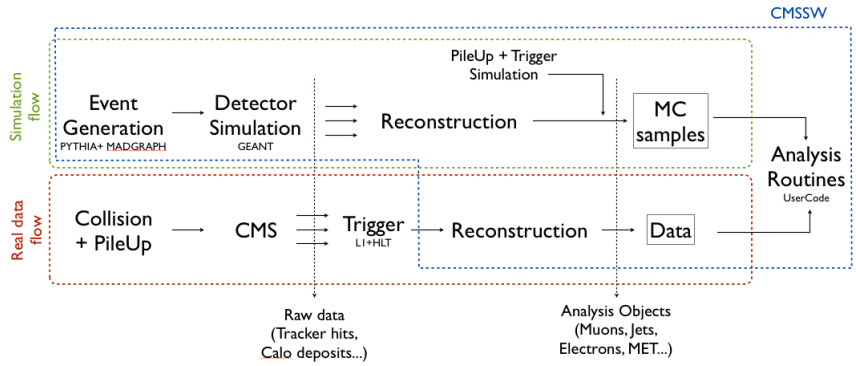


Figure 3.3: Simulation and data flow. The steps managed by CMSSW are underlined in blue.

3.3 Supersymmetry searches at CMS

3.3.1 Production

In a $p-p$ collision at LHC, squark, gluinos, neutralinos and charginos can be directly produced. Under the hypothesis of R-parity conservation these particles would be produced in pairs. Since strong production are favored with respect to electroweak productions, if they are kinematically permitted, we mainly expect production of squarks and gluinos pairs. Three cases are possible (Figure 3.4):

- $m_{\tilde{g}} \gg m_{\tilde{q}}$. The production of two squarks is dominant;
- $m_{\tilde{g}} \ll m_{\tilde{q}}$. The production of two gluinos is dominant;
- $m_{\tilde{g}} \sim m_{\tilde{q}}$. Diagrams involving the production of squark pairs, gluino pairs and squark-gluino pairs give similar contributions.

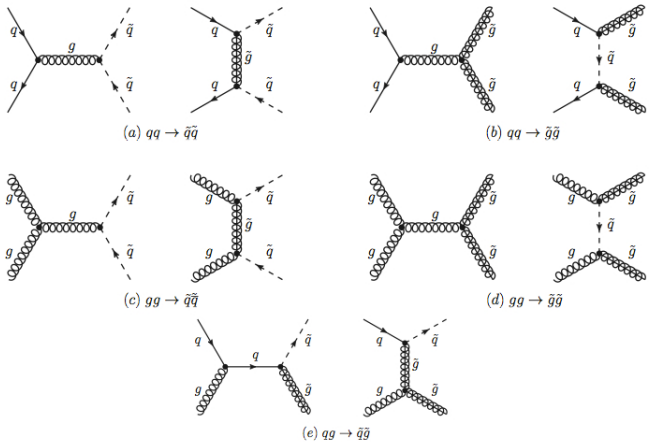


Figure 3.4: Diagrams of the production of squark and gluino pairs.

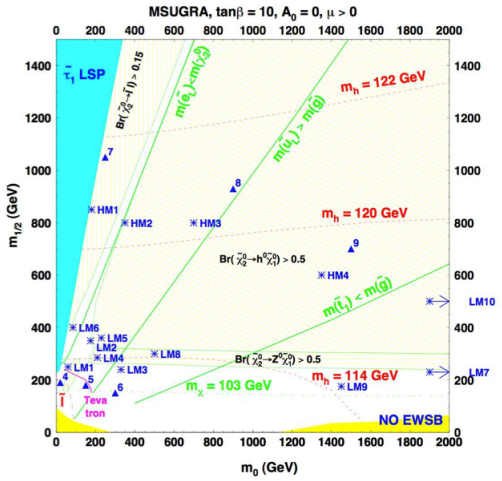


Figure 3.5: cMSSM plane with contour lines defining regions in which squarks are lighter than gluino, squarks are heavier than gluino, and $m_{\tilde{g}} \sim m_{\tilde{q}}$.

These possibilities define different regions in the cMSSM plane, as shown in figure 3.5.

If squark and gluinos are too heavy to be produced at LHC (Figure 3.6), electroweak production of neutralinos and charginos become important. The study of these processes is more complicated since neautralinos and charginos are mixed states of gauginos and higgsinos. In a scan the cMSSM space then, even the matrix elements change, modifying the weights of different diagrams.

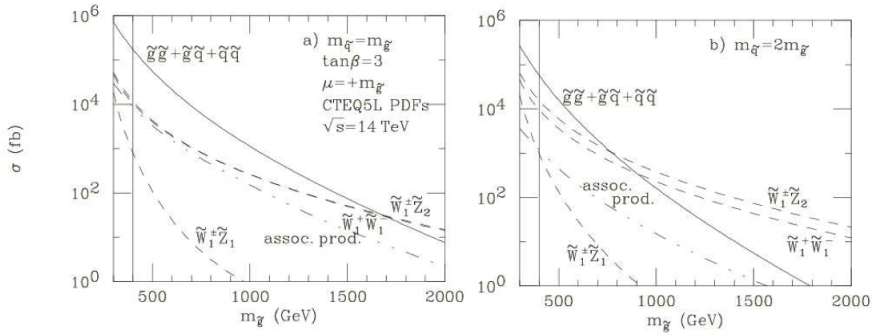


Figure 3.6: Production cross section for SUSY particles as function of the gluino mass. Are analyzed the conditions $\tan\beta = 3$, sign$\mu = +$, and degeneration of squark masses. In the left side $m_{\tilde{g}} = m_{\tilde{q}}$ is supposed, while the in the right plot $m_{\tilde{q}} = 2m_{\tilde{g}}$ is considered.

3.3.2 Decay Chains

Under R-parity conservation, once produced, squarks and gluinos decay following chains with a SUSY particle at every step and ending with the LSP ($\tilde{\chi}_1^0$ in the most area of cMSSM plane).

If squarks and gluinos are between the lightest SUSY particles, these chains are quite short. If instead they are heavier, cascades are topo-

logically more complex (figure 3.7). However, a generic SUSY signature involves:

- high jet multiplicity, especially in long chains, with high p_T, due to the high mass of the parent particle. This leads to high scalar sum of the transverse momentum of the jets (high H_T signature);
- large missing transverse energy, related to the escape of $\tilde{\chi}_1^0$ (high MET signature);
- possibly, one or more hard leptons can be produced during the cascade (leptonic signature).

The last item is very important for our analysis and will be discussed in more detail in the next sections of this chapter.

A typical decay chain, involving all the signatures presented is shown in figure 3.8.

3.4 SUSY studies

3.4.1 mSUGRA Benchmark points

In order to develop search strategies covering different phenomenological scenarios, some *benchmark points* (i.e. some particular points in mSUGRA parameters space) have been chosen by the CMS collaboration. Focusing on one point (or on a small set of points), Monte Carlo analyses based on a detailed and precise detector simulation (*fullsim*) ca be performed. We count 14 *low mass* points (LM0-LM13), characterized by relatively low mass spectra (then suitable for an early discovery), and 4 *high mass* points (HM1-HM4). The parameters of the LM points are defined in table 3.1 and their position in the cMSSM plane is shown in figure 3.5.

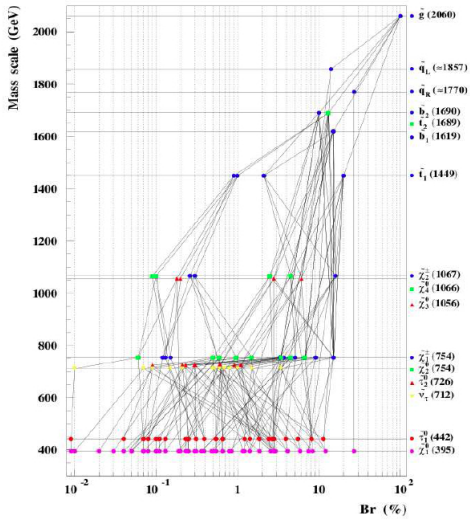


Figure 3.7: Possible decay modes and branching ratios of a gluino in a high-mass test point of mSUGRA parameters space.

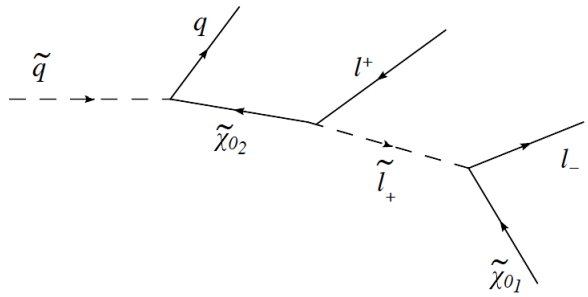


Figure 3.8: Feynman diagram of a chain starting from a squark, leading to the production of jets, leptons and high MET.

Benchmark	$m_0(\mathrm{GeV}/c^2)$	$m_{1/2}(\mathrm{GeV}/c^2)$	$A_0(\mathrm{GeV})$	$\tan\beta$	signμ
LM0	200	160	-400	10	+
LM1	**60**	**250**	**0**	**10**	**+**
LM2	185	350	0	35	+
LM3	330	240	0	20	+
LM4	**210**	**285**	**0**	**10**	**+**
LM5	230	360	0	10	+
LM6	85	400	0	10	+
LM7	3000	230	0	10	+
LM8	500	300	-300	10	+
LM9	1450	175	0	50	+
LM10	3000	500	0	10	+
LM11	250	325	0	35	+
LM12	2544.58	246.56	-856.75	47.59	+
LM13	**270**	**218**	**-553**	**40**	**+**

Table 3.1: mSUGRA low mass benchmark points. In bold: the three points analyzed in this work.

Although these points cover a wide range of different mass spectra (and then production and decay scenarios), they are all characterized by having $\tilde{\chi}_1^0$ as LSP and being compatible with cosmological measurements performed by WMAP, thus they all admit a SUSY particle as Cold Dark Matter candidate.

The three points underlined in table 3.1 (LM1, LM4, LM13) have been studied in this work. Given the benchmark point particle masses are calculated by a routine called SOFTSUSY[51]. In figure 3.9 the mass spectra for each point are shown. LM1 and LM4 are characterized by gluinos heavier than squarks, thus squark pair production is favored in these cases. In LM13 squark and gluino masses are similar so both productions occur with the same probability.

A particular feature of LM1 is that sleptons have masses intermediate between those of the two lightest neutralino. This makes possible a particular sequence of two-body decays:

$$\tilde{\chi}_2^0 \to l + \tilde{l} \to l^+ + l^- + \tilde{\chi}_1^0$$

leading to an opposite sign lepton pair whose invariant mass shows a characteristic end-point (Figure 3.8).

Is important to notice that two of these points (LM1, and LM4) have been already excluded by direct searches in CMS [1]. This fact however has no importance for our work, since we do not aim to prove if SUSY exists at one of these benchmark points. They are actually used only as a set of different phenomenological possibilities on which our methods can be tested.

[1] See e.g. [52] [53] [54] [55] .

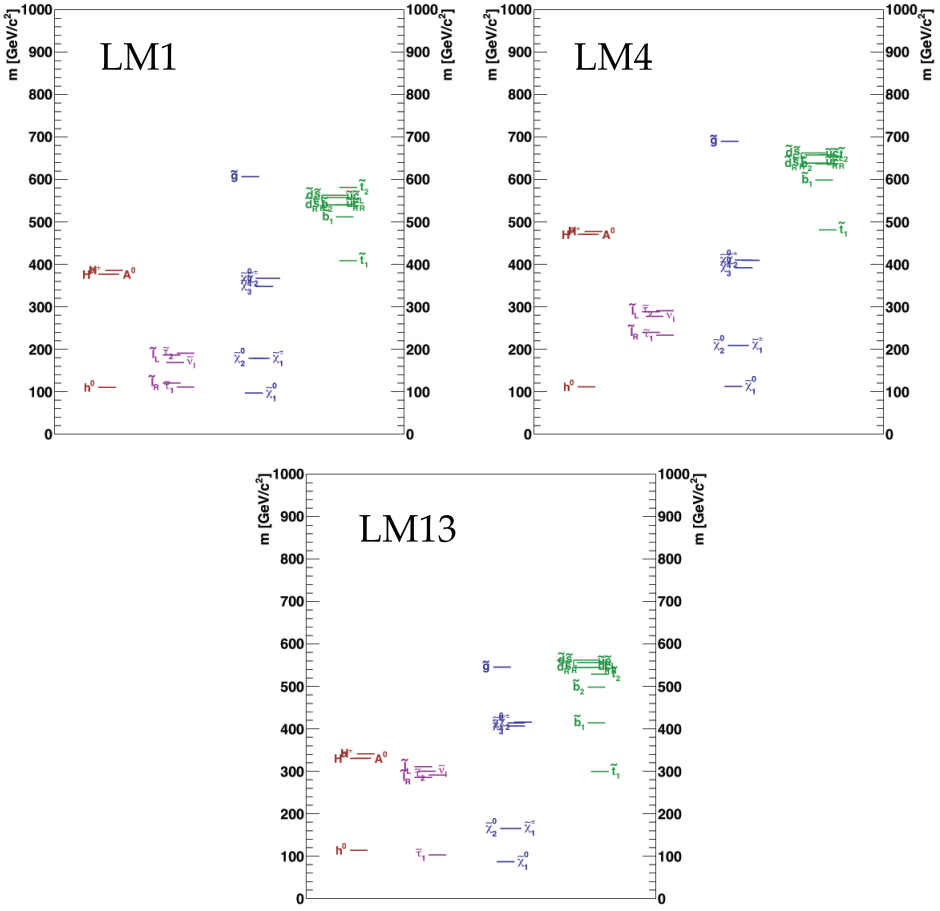


Figure 3.9: mSUGRA mass spectra for three benchmark points (LM1, LM4, LM13).

3.4.2 cMSSM scans

A scan of a large part of cMSSM plane is of fundamental importance in computing exclusion limits or evaluating the discovery potential of a certain analysis. It however requires the production of a very large num-

ber of samples. They cannot be produced with a full detector simulation (which takes on average some minutes per event for a complete simulation and reconstruction). For this purpose, a faster but less detailed simulation, called *fastsim*, can be made[56]. A fastsim, uses a simplified geometry model and some parametrization obtained from fullsims. The output samples are similar to those obtained with fullsim, hence the same analysis tools can be used. A typical SUSY event requires ~ 1 second to be completely fast-simulated.

3.4.3 Simplified models

SUSY searches have been historically focused on particular constrained models. cMSSM however covers only a limited set of possible mass spectra and mass splitting, due to the strong relations between masses within these models. Moreover many production and decay channels contribute to each benchmark point in a strongly model-dependent way. There are two possible ways to go beyond the cMSSM. The first is to consider a more generalized model (e.g. pMSSM[57]), the second is to describe discovery potentials and exclusion limits in term of *simplified models*. The interpretation of CMS results using these more flexible models is an useful alternative and complementary approach to the traditional cMSSM analyses.

In a simplified model a limited set of hypothetical particles and decay chains is considered, matching to the experimental results of a specific search channel. The free parameters of these models are then the particle masses and the branching ratios. Each simplified model spectrum (SMS) consists of a small list of new hypothetical particles involving one or few more topologies. They can be seen as blocks of more complete models

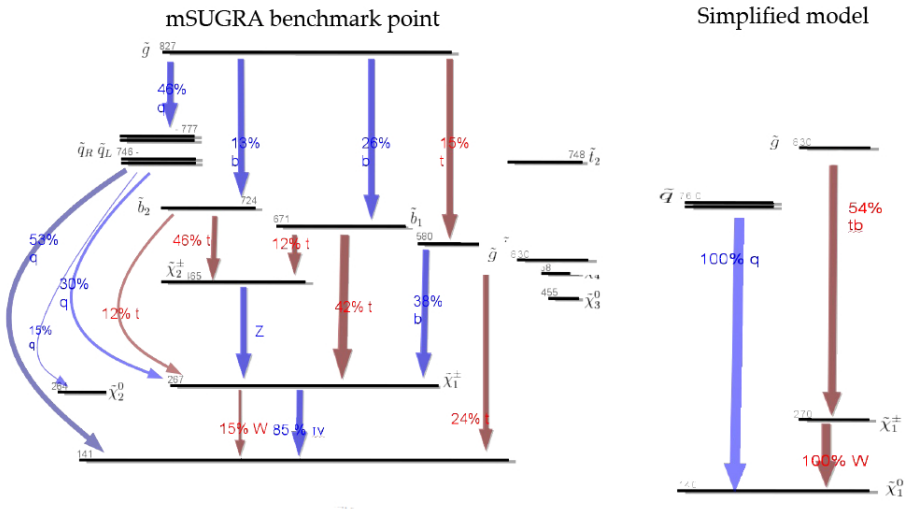


Figure 3.10: (Left) Mass spectrum and branching ratios for a specific mSUGRA benchmark point. (Right) The SMS starting from the same particles and involving the production of the same intermediate gaugino.

(Figure 3.10). Every topology has a codename which summarizes information about parent particles, chain length and, optionally, final states. For example, the topology:

T1bbbb

is characterized by the production of gluino pairs, directly decaying to LSP with two b-jets per leg in the final state (Figure 3.11).

In general, the first letter is always a "T" (for topology), followed by:

- 1,3,5 (odd numbers) when gluino-gluino are produced;
- 2,4,6 (even numbers) when squark pairs are produced;
- GQ for gluino-squark production.

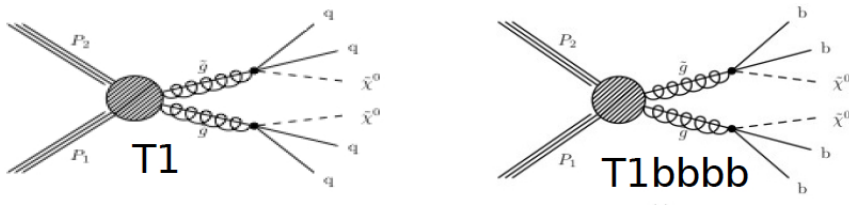


Figure 3.11: Diagrams of a T1 simplified topology. In the right side the final states are specified (topology T1bbbb).

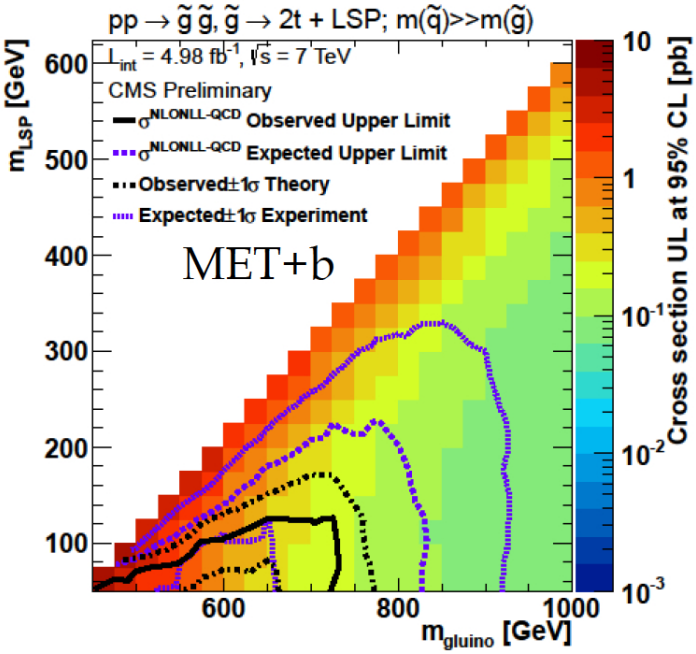


Figure 3.12: Example of a simplified model exclusion plot. From [58]

Other productions are not considered. Defined a production, the numbers also specify intermediate states:

- 1,2: direct decay to LSP;
- 3,4: intermediate neutralino or chargino on one leg;
- 5,6: intermediate neutralino or chargino on both legs.

The CMS simplified models were chosen to cover a large part of the kinematic phase space of all considered final states. Limits are presented exclusively for each topology and the results are shown as a function of the masses of the particles involved (figure 3.12).

3.5 Dilepton analyses

3.5.1 Reference Analyses

Many different analysis strategies have been developed by the CMS SUSY group. Some of them look at events with jets and MET (*hadronic searches*), other make the more exclusive requirement of one or more leptons in the final state (*leptonic searches*). Depending on the final states involved, analyses are divided into groups called *Reference Analyses* (RA), organized as follows

Hadronic Working group:

- **RA1** - Inclusive jet searches;
- **RA2** - Exclusive jet searches (as subgroups: RA2b, RA2τ);
- **Razor** - This Analysis works with both hadronic and leptonic final states and is based on the construction of a dimensionless variable called *razor* (R)[59].

Photon Working group:

- **RA3** - Search for events with γ+MET.

Leptonic Working group:

- **RA4** - Single lepton searches;
- **RA5** - Same Sign (SS) dilepton analysis (two leptons of the same charge);
- **RA6** - Opposite Sign (OS) dilepton analysis;
- **RA7** - Events with more than two leptons.

3^{rd} **generation Working group**:

Focused on SUSY channels involving the production of 3^{rd} generation particles ($\tilde{\tau}$, $\tilde{b}$) thus leading to final states with taus and b-jets. These analyses are necessary to enhance the sensitivity of SUSY searches in regions of the cMSSM plane difficult to reach.

3.5.2 Opposite Sign dilepton events

A decay chain generated by a squark or a gluino can involve the production of one or more charged leptons. In the analysis here presented we intend by leptons only electrons or muons and not taus, whose identification require more complex techniques[60].

In a SUSY event, two sparticles (and hence two chains) are generated, leading to the possibility to produce lepton pairs of the same charge (or *same sign*, SS) or of *opposite sign* (OS). In some cases, OS pairs can be generated inside the same chain. This is a very interesting possibility, since a systematic study of dilepton invariant-mass spectra might allow a

spectroscopy of the masses involved in the chain[61].
OS dilepton events are studied with particular attention not only for this possibility but also (and this is a crucial point during the discovery phase) because they are a cleaner channel with respect to hadronic searches, for which high cross-section processes like QCD productions, represent a dramatic background to deal with.

In this work, opposite sign electrons and muons are considered. Thus three different channels are studied: ee, μe, and $\mu\mu$. In addition to the leptons, high MET and large hadronic activity (at least two jets with p_T over a threshold and high H_T) are required.

3.6 Main sources of background for OS dilepton analysis

Looking at the MET + jets + 2 OS leptons signature permits a strong reduction of SM backgrounds (in particular the huge contribution of QCD processes). However some backgrounds still survive the selection cuts. The main backgrounds analyzed in this work are listed in the following.

$t\bar{t}$+jets

The top pair production represents the main source of physical background for MET + jets + 2 OS leptons analysis. Both tops decay in a $W + b$ pair (BR = 99.8%). Depending on the decay mode of each W we distinguish fully hadronic decays, semi-leptonic decays, and fully leptonic decays. This last mode (figure 3.13) has a BR = 11% and is characterized by the production of a opposite sign pair of leptons (one from every W, thus of any flavor), a physical MET generated by neutrinos, and jets (two b-jets plus

other jets coming from ISR and beam remnants).

However SUSY chains are generated by particles which are expected to be more massive than a top quark, thus leading to harder jets on average (higher H_T). Moreover MET in SUSY events would be generated by a massive neutralino and would then be higher on average than the one generated by neutrinos.

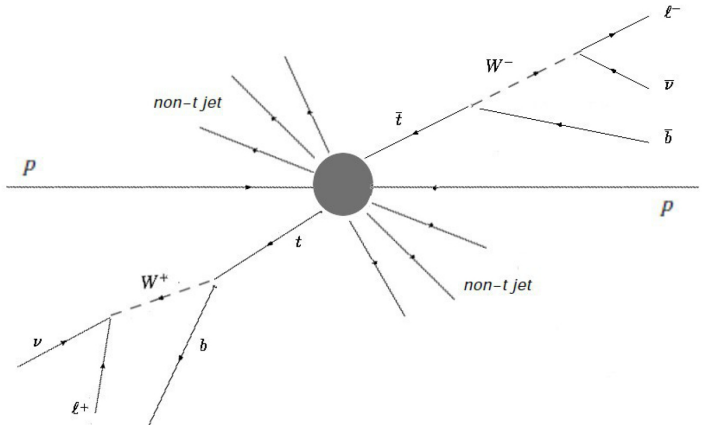


Figure 3.13: A $t\bar{t}$+jets process.

Drell-Yan + jets

A Drell-Yan process occurs when a quark from one proton annihilate with a sea-antiquark from the other, creating a virtual photon or Z decaying into a pair of opposite sign leptons (obviously of the same flavor). At the same time, jets can be produced by ISR and beam remnants. In this case there is no physical MET. Nevertheless an un-physical missing transverse energy could rise from an imperfect jet reconstruction.

Single top production

Single top productions occur mainly via three electroweak processes:

- **s-channel**. A virtual W is produced from a $q - \overline{q}$ annihilation and decays in a $t - \overline{b}$ pair. The top can then decay leptonically. A second lepton may come from the decay of the b quark of from a jet mis-identification (*fake lepton*).

- **t-channel**. A quark scatters with a sea b from the other proton leading to the production of a top (figure 3.14-b). The final states are similar to those seen for the s-channel.

- **tW-channel**. A top and a W boson are produced via electroweak interaction of a gluon with a sea b. If both t and W decay leptonically, a final state with two leptons, MET from neutrinos and jets (from b or ISR/beam remnants) occurs.

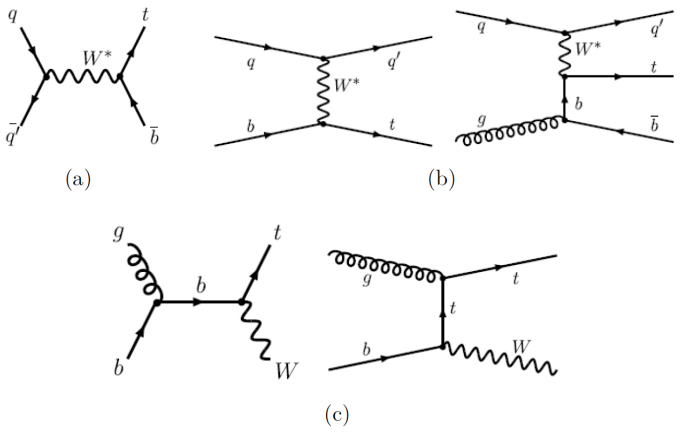


Figure 3.14: Single top processes occurring in s-channel (a), t-channel (b), tW-channel (c).

W+jets

This process represent a background only when a jet is mis-identified as a lepton. It has however to be taken into account since its cross section is significantly higher then that of the other processes here considered (table 3.2).

Vector boson pairs

A double leptonic decay of WW+jets or a ZZ+jets with one Z decaying in a opposite sign leptonic pair and the other in two neutrinos leads to final states with two OS leptons, MET and jets, thus represents another physical background to our searches.

The cross sections at $\sqrt{s} = 7$TeV (2010 and 2011 data taking) are listed in table 3.2, together (for comparison) with the cross section expected for SUSY at cMSSM benchmark points LM1, LM4, and LM13. The cross sections for SUSY processes are evaluated at leading order. For them a *k-factor* is considered as a correcting term to transform the LO cross section into a NLO one.

Process	Cross section (pb) ($\sqrt{s} = 7$TeV)
$t - \bar{t}$+jets	157.5
Drell-Yan + jets:	
$10 < M_{ll} < 50$ GeV/c^2	400
$M_{ll} > 50$ GeV/c^2	3048 (NNLO)
Single top:	
s-channel	1.5
t-channel	20.5
tW-channel	10.6
W+jets	31314 (NNLO)
Vector boson pairs	4.8
LM1	4.88 (LO, k = 1.34)
LM4	1.87 (LO, k = 1.35)
LM13	6.89 (LO, k = 1.42)

Table 3.2: Cross sections of background processes at $\sqrt{s} = 7$TeV (2010 and 2011 data taking) and (for comparison) of SUSY production at LM1, LM4 and LM13. If not specified, cross sections are calculated at Next to Leading Order (NLO). Calculations are made using the package PROSPINO[62]. k-factors are also specified for LM points.

4

Event selection, data and Monte Carlo simulation

The details of the analysis presented in this thesis are introduced in this chapter. After the definition of the selection criteria of physics objects (electrons, muons, jets, MET, and HT), the analysis cuts are discussed, together with datasets and triggers used. The second part of the chapter is dedicated to studies on Monte Carlo simulations of the Standard Model background. In particular, the effects of pile-up reweighting and of systematic uncertainties are studied. Finally a data/MC comparisons are performed.

4.1 Analysis objects

From an operative point of view, analysis objects (e.g. electrons, muons or jets) are defined starting from the signature they leave interacting with sub-detectors, or, more likely, from the steps necessary to reconstruct them from raw information (energy deposits, hits in the tracker...).

A complete event reconstruction is a very complex task since the definition of any object is often ambiguous: the attribution of one track to a certain energy deposit, for example, is not always immediate as well as the assignment of a calorimetric cluster to one of two overlapping jets. Therefore a quality selection, based on cuts on appropriate variables, is the basis of a correct object reconstruction.

Standard criteria have been developed by the CMS collaboration to maximize the efficiency and purity of object selection. In the following sections the definition of the objects used in this work will be discussed.

4.1.1 Electrons

Electrons are charged particles interacting mainly electromagnetically with matter. Their signature in CMS is then characterized by some hits in the tracker and by the production of an electromagnetic shower in the ECAL (made by photons from bremsstrahlung and electron-positron pairs). Since the ECAL has a high granularity, an electromagnetic shower is collected by many cells. Neighbor cells containing energy deposits form a *cluster*, while groups of clusters are arranged in *super clusters* (SC). The reconstruction of an electron consists in matching a SC with a track obtained fitting different hits in the tracker. Within the reconstruction algorithms developed by the CMS collaboration[63], the *Gaussian Sum Filter* (GSF) has been used. Electrons reconstructed in this way are called GSF-electrons.

The charge of an electron can be measured in different ways. It can be inferred from the curvature of the track associated (GSF-track charge) or studying the energy distributions in the shower (SC charge). A redundant approach is used: an electron charge is considered well measured if both methods are in agreement.

Many other variables are considered in electron identification in order to avoid, for example, the occurrence of electron from photon conversion, non-prompt electrons (not originated in the primary interaction) or jets mis-identified as electrons.

Prompt electrons

The impact parameter with respect to the vertex ($|d_{0,pv}|$) has not to exceed a certain threshold, to ensure that an electron is originated in the primary vertex.

Conversion rejection

Electrons not originated from converted photons have to satisfy these requirements:

- they have to be far enough from an opposite charged electron[1] (photon conversion produces pairs of electrons). The distance is evaluated using an impact parameter (d_0) and an angular distance ($\Delta \cot \theta$);
- no missing hits in the tracker. If a photon is produced in the interaction vertex rather than an electron, this neutral particle would not leave any hit in the tracker until the conversion, and would then produce missing hits in the tracker.

[1]The object "electron" includes both e^- and e^+.

Jets rejection

Some jets could produce showers in the ECAL, simulating the behavior of an electron. However the shower produced have different shapes. A set of shape variables is used in CMS to avoid jet mis-identifications. The cuts on these variables can be set to different levels of identification efficiency. In the studies here presented the *working point 90* (WP90) IDs are used, meaning an efficiency of 90%.

4.1.2 Muons

Muons are characterized by signals in the tracking system and in the muon chambers. Two different reconstruction algorithms exist, depending on which information is used as seed[64]. If the reconstruction starts from the muon system the muon will be called *global muon*. In this case a track in the muon system is extrapolated from a seed in the CSC or in the DT, and then matched with a track in the inner tracking system. If instead an inner track is matched with some hits in the muon chambers, a *tracker muon* is reconstructed. To be tagged as good muon both global and tracker identifications are often required.

Muons reconstructed only in the muon chambers are called *standalone muons*. It is, for example, the case of muons coming from cosmic rays.

Prompt muons

In addition to the tagging, some requirements on the number of hits both in the tracker and in the muon chambers are imposed as well as a good quality track-fit (parametrized by the variable χ^2/N_{dof}). However these conditions do not ensure that a muon is originated in the primary vertex.

An additional condition, as for electrons, on the impact parameter with respect to the vertex ($|d_{0,pv}|$) is then imposed.

Calorimetric deposits

Muons are, in a wide range of momenta, minimum ionizing particles (MIP). This means that they leave only a small fraction of their energy in calorimeters. Calorimetric deposits have to be consistent with a MIP. This is ensured studying both ECAL and HCAL deposits.

4.1.3 Isolation

A fundamental quantity in electrons and muons reconstruction is the *isolation*, a measurement of the activity (tracks and deposits) near a particle (figure 4.1). It is used to verify if a lepton is produced inside a jet. There

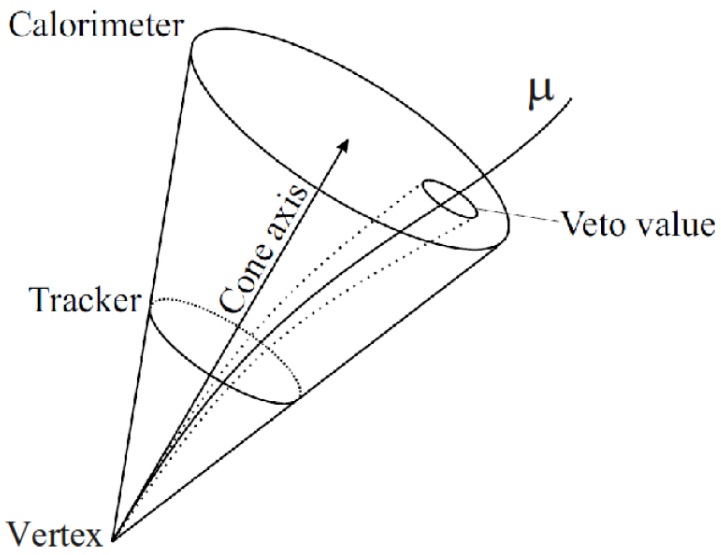


Figure 4.1: Isolation of a muon.

are three different kind of isolation:

Tracker Isolation (TkIso): defined as the sum of the transverse momenta of all the tracks within a cone of aperture $\Delta R < 0.3$ centered on the lepton direction:

$$\text{tkIso} = \sum_{\Delta R<0.3} p_T. \tag{4.1}$$

Electromagnetic Isolation (ECALIso) : the sum of the transverse energies deposited in the ECAL inside a cone defined as above:

$$\text{ECALIso} = \sum_{\Delta R<0.3} E_T^{\text{ECAL}}. \tag{4.2}$$

Hadronic Isolation (HCALIso): the sum of the transverse energies deposited in the HCAL inside a cone defined as above:

$$\text{HCALIso} = \sum_{\Delta R<0.3} E_T^{\text{HCAL}}. \tag{4.3}$$

A dimensionless quantity called *relative isolation* (RelIso) is introduced to simultaneously account for the three isolations. It is defined as:

$$\text{RelIso} = \frac{1}{p_T} \left(\text{TkIso} + \text{ECALIso} + \text{HCALIso}\right). \tag{4.4}$$

The lepton itself is obviously excluded by the computation of the activity. This is made defining a small veto-cone around the electron or the muon.

The selection criteria applied for this analysis are summarized in table 4.1 for electrons and in table 4.2 for muons.

Electrons	
Acceptance cuts	
$p_T > 10$ GeV $\|\eta\| \leq 2.5$ $\|\eta\| \leq 1.442$ or $\|\eta\| \geq 1.560$	 To account for the barrel/endcaps gap.
Prompt electron	
$\|d_{0,pv}\| < 0.04$ $\|d_{Z,pv}\| < 1$	Impact parameter. z distance from PV.
Conversion rejection	
$\|d_0\| > 0.02$ $\|\Delta \cot\theta\| > 0.02$ $N_{\text{missinghits}} \leq 1$	From any opposite charged electron.
Jet rejection (WP 90)	
Barrel	Endcap
$\sigma_{i\eta,i\eta} < 0.01$ $\|\Delta\phi_{In}\| < 0.15$ $\|\Delta\eta_{In}\| < 0.007$ $H/E < 0.01$	$\sigma_{i\eta,i\eta} < 0.03$ $\|\Delta\phi_{In}\| < 0.1$ $\|\Delta\eta_{In}\| < 0.009$ $H/E < 0.0075$
Isolation	
$RelIso < 0.15$	
Muon-Electron cleaning	
$\Delta R(Ele, \mu) > 0.1$	From any global or tracker muon.

Table 4.1: Selection criteria for electrons. According to the CMS convection, distances are expressed in cm.

Muons	
Acceptance cuts	
$p_T > 5$ GeV $\|\eta\| \leq 2.4$	
Prompt muon	
$\|d_{0,pv}\| < 0.02$ $\|d_{Z,pv}\| < 1$	Impact parameter. z distance from PV.
Muon Id	
IsGlobalMuon && IsTrackerMuon	
Good track reconstruction	
$Ntk_{hits} > 10$ $N\chi^2 < 10$ $Nmu_{hits} > 0$	Hits in the inner tracker. Good track fit at tracker. Hits in the muon system.
Isolation	
$RelIso < 0.15$	
Minimum Ionizing particle	
Sig$p_T < 0.1$	Sig$p_T = \frac{p_T^{\mu}(ECAL)}{p_T^{\mu}}$

Table 4.2: Selection criteria for muons. Distances are in cm.

4.1.4 Jets

Jets are products of quarks' hadronization. They are collimated bunches of hadrons whose axis is close to the direction of the original quark. Jets reconstruction at hadron collider experiment is a very challenging task, due to the intense hadron activity: pile-up and the underlying event can introduce an energy bias to the jets. Moreover the assignment of a particle to one jet rather than another often leads to ambiguities. Finally a loss of some particles in the jet (due for example to a fall out of the detector acceptance) could lead to an incomplete reconstruction of the jet itself.

Many complex reconstruction algorithms exist[65], for example *cone algorithms* (as SIS-cone or Iterative-cone) or *clustering algorithms* (like anti-k_T). The algorithm used for this analysis is the so called *Particle Flow* algorithm (PF).

In the PF[66], information from all the sub-detectors are combined in a coherent way to perform a reconstruction of the whole event. Tracks and clusters (and not only calorimetric deposits) are used as seeds for the jet reconstruction. The PF algorithm is safer than the other algorithms with respect to particle duplications or erroneous assignments to one jet.

Jet ID

The CMS collaboration has defined some Jet Identification criteria. In our work the jet LooseID is used, with criteria regarding the following variables:

- energy fractions: fractions of energy released by charged or neutral particles in the ECAL or in the HCAL;

Jets	
Acceptance cuts	
$p_T > 30$ GeV $\|\eta\| \leq 3$	
Jet Id	
Pf LooseID	
EMF > 0.1 $n_{const} > 1$ $n_{90}^{\text{hits}} \geq 2$	Electromagnetic fraction. Number of constituents.
Jet-Lepton cleaning	
$\Delta R(\text{jet}, \text{lep}) > 0.4$	From any selected electron or μ.

Table 4.3: Selection criteria for jets. Distances are in cm.

- n_{hits}^{90} : the minimum number of hits which contributes to the 90% of energy deposit (this quantity is studied for example to discard unphysically "hot" towers);
- number of constituents inside a jets.

Lepton cleaning

In order to avoid erroneous contamination from leptons, a lepton-jet cross cleaning is performed. Any jet whose axis is within a cone of $\Delta R < 0.4$ around a lepton is removed.

The selection criteria applied for jets are listed in table 4.3.

4.1.5 MET and H_T

The missing transverse energy has been introduced in chapter 2. It is a measurement of the energy unbalance in an event due, for example to

neutrinos or other weakly interacting neutral particles. It can be evaluated from a computation of transverse energy of reconstructed jets, leptons and of unclustered calorimetric deposits as well.

$$\text{MET} = |-\sum_i \overrightarrow{p_{Ti}}|. \tag{4.5}$$

Many reconstruction algorithms can be used to compute the MET. In this work *Particle Flow MET* (pfMET) is used. As the name suggests, in this case, PF objects are used as input for MET calculation.

Another key quantity in our analysis is the scalar sum of the transverse momenta of jets, H_T:

$$H_T = \sum_i p_{T,i}^{jet}. \tag{4.6}$$

For the calculation, only jets passing the selection criteria are used. A p_T threshold (30 GeV, for this work) is imposed for jets to be included in H_T calculations.

Both MET and H_T play a fundamental role in selecting events of physics beyond the standard model. Thus cuts on these variables are imposed in our analysis, as will be discussed in the next sections.

4.2 Pre-selection cuts

In the pre-selection stage, an event cleaning is performed requiring high purity tracks, good primary vertex, and low noise in the calorimeters. These criteria are quite similar for almost every CMS analysis. After this

stage, more specific cuts are defined, requiring, for example, the presence of jets or lepton, defined as in the previous section.

The pre-selection in this work is made with a local analysis tool called *OSDLAnalyzer*[67] (see Appendix A for more details) working within the CMSSW framework, which produces skimmed n-tuples from Monte Carlo or Data samples. These n-tuples are filled with the events passing the pre-selection. Many variables useful for analysis are stored for every event. Events are stored in different trees depending on the dilepton channel: DoubleMu ($\mu\mu$), MuEG (μe), and DoubleEle (ee).

Good vertex selection

At least one good primary vertex is required. Primary vertices are reconstructed from some properly chosen tracks [68], fitted with an adaptive vertex-fit method[69]. A good primary vertex has to be produced not too far from the interaction point, hence conditions on d_z and d_ρ (longitudinal and transverse distances) with respect to this point are imposed. A condition on the number of degrees of freedom of the reconstructed vertex (N_{dof}) is imposed too.

Trigger

Events have to pass double-lepton triggers. The trigger menus used for this analysis will be presented in the next sections. The trigger is applied only on data. For the Monte Carlo samples, only a correction factor to account for trigger efficiency is used.

Event pre-selection	
Primary vertex	
Not Fake flag $d_z < 24$ $d_\rho < 2$ $N_{\text{dof}} > 4$	With respect to the interaction point.
Trigger For data only	
Leptons	
$n_{lep} = 2$ $charge_1 \neq charge_2$ $m_{ll} > 12 GeV/c^2$	No low mass resonances.
Jets	
$n_{jet} \geq 2$	

Table 4.4: pre-selection criteria. Distances are in cm.

Pre-selection for OSDL analysis

For our analysis we require events with exactly two leptons (muons or electrons) of opposite charge. In order to remove background from low mass resonances, a minimum value of 12 GeV/c^2 for the invariant mass of the two leptons is required. Furthermore, we require $p_T > 20$GeV for the first lepton and $p_T > 10$GeV for the second.

Finally at least two good particle flow jets are required with $p_T > 30$ GeV, having $\Delta R > 0.4$ with both the selected leptons.

Pre-selection criteria are summarized in table 4.4.

4.3 Event selection

The event selection is aimed to select with high efficiency candidate events with the topology expected for new physics and at the same time to reject events with topologies typical of Standard Model background process. The requirement of two leptons and of two or more hard jets sensibly reduces some physical backgrounds, like QCD. Moreover, as already discussed, SUSY events are characterized by large MET and high hadronic activity (and hence high H_T). A cut on $MET < 50GeV$ and $H_T < 100GeV$ is imposed to reduce part of the background.

In figure 4.2 the MET and HT distributions for $t\bar{t}$ events and LM1-SUSY events, after pre-selection cuts are applied, are shown[1].

Another important variable o disentangle background from candidate new physics events is the invariant mass of the lepton pair (m(ll)). Most of the Drell-Yan background surviving the pre-selection (figure 4.3) is mediated by the production of a Z boson, hence for these events m(ll) $\sim$ m$_Z$ (91 GeV). A sizable reduction of the background can be obtained imposing a veto on events with dilepton invariant mass within a window around the Z mass ($76GeV < m(ll) < 106GeV$). This veto is used only for ee and $\mu\mu$ events, since an $e\mu$ pair can not be generated in a Drell-Yan process.

It has moreover to be noticed that even if any new physics manifests inside the Z-window, it would be completely covered by the huge Standard Model background. Hence the Z-window is used as a background-control region in many analyses.

[1] As can be seen from picture 4.2 the cuts imposed on MET and H_T are quite loose for many beyond-the-Standard Model phenomenologies. Anyway this perfectly suit our analysis, since (as will be discussed in the next chapter) many other discriminating variables are involved, hence thigh cuts would dramatically reduce the statistics.

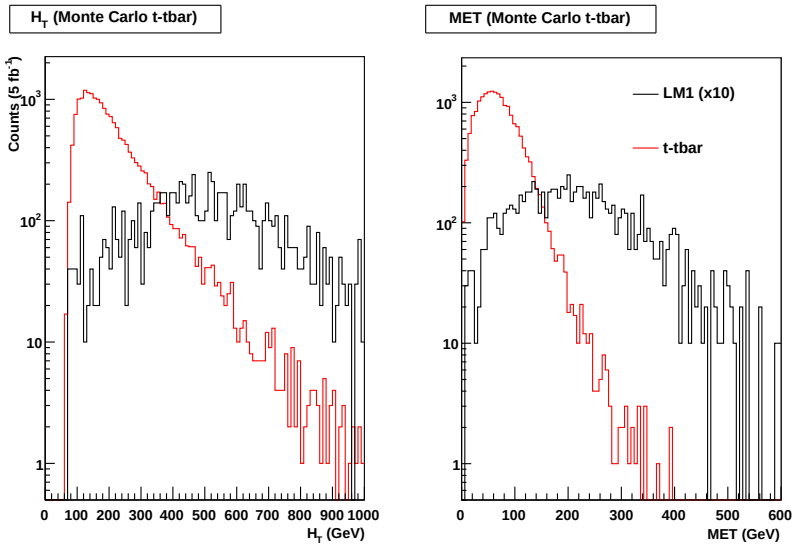


Figure 4.2: Distribution of H_T and MET for preselected sample ($ee + e\mu + \mu\mu$) of Monte Carlo $t\bar{t}$ and SUSY at LM1.

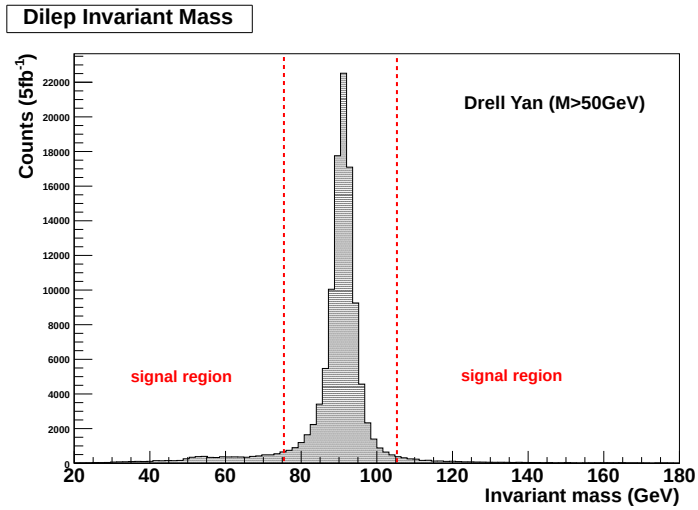


Figure 4.3: Distribution of the dilepton invariant mass for a preselected sample of Drell-Yan events. The Z-window is also highlighted.

Event selection
MET > 50GeV
$H_T > 100$GeV
$\lvert m_{ll} - m_z \rvert > 15$ GeV (ee and $\mu\mu$ only)

Table 4.5: Selection cuts.

4.4 Recorded datasets

A total integrated luminosity of $\mathcal{L} = 4.98 fb^{-1}$ of data (corresponding almost to the whole $\sqrt{s} = 7$ TeV 2011 data taking) has been analyzed. The data samples are divided into three different skims, based on some properties of the final states:

DoubleMu skim
(Events with two high p_T muon candidates.)
* /DoubleMu/Run2011B − PromptReco − v1/AOD
* /DoubleMu/Run2011A − PromptReco − v6/AOD
* /DoubleMu/Run2011A − 05Aug2011 − v1/AOD
* /DoubleMu/Run2011A − PromptReco − v4/AOD

MuEG skim
(Events with high p_T muon or electron candidates.)
* /MuEG/Run2011B − PromptReco − v1/AOD
* /MuEG/Run2011A − PromptReco − v6/AOD
* /MuEG/Run2011A − 05Aug2011 − v1/AOD
* /MuEG/Run2011A − PromptReco − v4/AOD

DoubleEle skim
(Events with two high p_T electron candidates.)
* /DoubleElectron/Run2011B − PromptReco − v1/AOD
* /DoubleElectron/Run2011A − PromptReco − v6/AOD
* /DoubleElectron/Run2011A − 05Aug2011 − v1/AOD
* /DoubleElectron/Run2011A − PromptReco − v4/AOD
* /DoubleElectron/Run2011A − 03Oct2011 − v1/AOD

4.4.1 Trigger menus

Double lepton triggers are applied to data, in order to skim events with at least 2 leptons with high p_T. The p_T thresholds of the triggers used are smaller than the requirements for dilepton in the selection cuts ($p_T^{1,2}$ =(20,10) GeV) thus they do not introduce any bias in the analysis. Since data samples are divided in DoubleMu, MuEG, and DoubleEle, the first sample is required to pass a Double Muon trigger, the second a Mu-Ele trigger and the last a Double Electron trigger. Any trigger is a logical OR of different triggers bits:

Double Electron
* `HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL`
* `HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT...`
* `HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT...`

Double Muon
* `HLT_DoubleMu7`
* `HLT_Mu13_Mu7`
* `HLT_Mu13_Mu8`
* `HLT_Mu17_Mu8`

Mu-Ele
* `HLT_Mu17_Ele8_CaloIdL`
* `HLT_Mu8_Ele17_CaloIdL`
* `HLT_Mu17_Ele8_CaloIdT_CaloIsoVL`
* `HLT_Mu8_Ele17_CaloIdT_CaloIsoVL`

The numbers in the tags refer to the minimum p_T of the triggered leptons. The efficiency of the triggers (i.e. the fraction of events actually triggered over the events fulfilling the trigger requirements) can be studied using a modified Tag and Probe method described in [70]. This efficiency depend on the p_T and η of the leptons. Anyway, an overall efficiency of:

- 90% for Double Muon trigger;
- 95% for Mu-Ele trigger;

- 100% for Double Electron trigger

can be assumed[70]. Triggers are not applied to Monte Carlo samples. Instead, they are scaled for the above efficiencies.

4.5 Monte Carlo validation

The estimate of the residual Standard Model background (i.e. those SM events surviving the analysis cuts) is a crucial element in any search for new physics. As will be more deeply discussed later, this estimate is performed using a *data-driven method*: a quantitative estimate (and a measurement of any eventual excess in data) is not obtained from a simulation, but directly from data. Monte Carlo simulations are only used to build and test data-driven techniques, to define signal and background regions and for a first evaluation of the power of the method in the background prediction. In this context it is of extreme importance to establish if the Monte Carlo is able to reproduce the trend of data, in other words if the simulated background behaves exactly like the real one in terms of:

- selection yield (fraction of events passing the selection criteria);
- shape of the distribution.

A validation of the Monte Carlo background simulation is the subject of the remaining part of this chapter.

4.6 MC Samples

The Monte Carlo samples used belong to the *Summer11* and *Fall11* MC productions and are produced on purpose to faithfully reproduce the working condition of 2011 data taking. They have been produced and analyzed

with the 4_2_X version of CMSSW (the same version used to analyze the data).

The samples used to simulate the Standard Model background (see section 3.6) are listed below:

$t\bar{t}$+jets
* /TTJets_TuneZ2_7TeV − madgraph − tauola/Summer11 − PU_S4
* /TTJets_TuneZ2_7TeV − madgraph − tauola/Fall11 − PU_S6

DY+jets
* /DYJetsToLL_TuneD6_M − 10To50_7TeV − madgraph − tauola
* /DYJetsToLL_TuneZ2_M − 50_7TeV − madgraph − tauola

Single Top
* TToBLNu_TuneZ2_s − channel_7TeV − madgraph
* TToBLNu_TuneZ2_t − channel_7TeV − madgraph
* TToBLNu_TuneZ2_tW − channel_7TeV − madgraph

W+jets
* WJetsToLNu_TuneZ2_7TeV − madgraph − tauola/Summer11 − PU

Vector boson pair
* VVJetsTo4L_TuneD6T_7TeV − madgraph − tauola

They are produced using MADGRAPH (section 3.2). For the SUSY signal at benchmark points LM1, LM4, and LM13 the following PYTHIA samples have been used:

* /LM1_SUSY_sftsht_7TeV − pythia6/Summer11 − PU_S4
* /LM4_SUSY_sftsht_7TeV − pythia6/Summer11 − PU_S4
* /LM13_SUSY_sftsht_7TeV − pythia6/Summer11 − PU_S4

Before going on with Data/MC comparison, in order to validate the simulations to be used for studies of data driven methods, two fundamental issues have to be studied. The first is related to the simulation of the pile-up (i.e. of the minimum-bias events overlapping to the hard scattering), which will be discussed in the next section, the second is the study of the systematic errors of the MC, arising from uncertainties on theoretical models and reconstruction effects.

4.7 Pile-up simulation

Every proton bunch crossing contains $\sim 10^{11}$ protons with a very short bunch spacing (25ns at design luminosity). Thus multiple minimum-bias collisions (events not characterized by hard scattering) overlap the hard scattering process generating, for example, a bias in energy deposits that has to be taken into account. This overlap is called *pile-up* (PU) and can be generated by minimum-bias events from a bunch crossing that take place during the hard scattering (in-time pile-up) or just before/after (out-of-time pile-up).

The number of pile-up interactions has a distribution whose mean value increase linearly with the instantaneous luminosity, thus the more it rises, the more important the pile-up effects become.

4.7.1 Event reweight

When a data/simulation comparison is performed, the Monte Carlo samples have to be adapted to the pile-up distribution in data. This can be done making an event-by-event reweighting of the simulations.

A good parametrization of the pile-up can be made looking at the number of primary vertices recorded for every event. In the data, the distribution of primary vertices is obtained from a measurement of the instantaneous luminosity (which is a source of systematic uncertainties, as will be discussed in section 4.8). In Monte Carlo sets, instead, these distribution are generated in such a way to cover (but not to exactly reproduce) the conditions expected for each data taking period[71].

The unweighted distributions of vertices for data and MC are plotted together in figure 4.4.

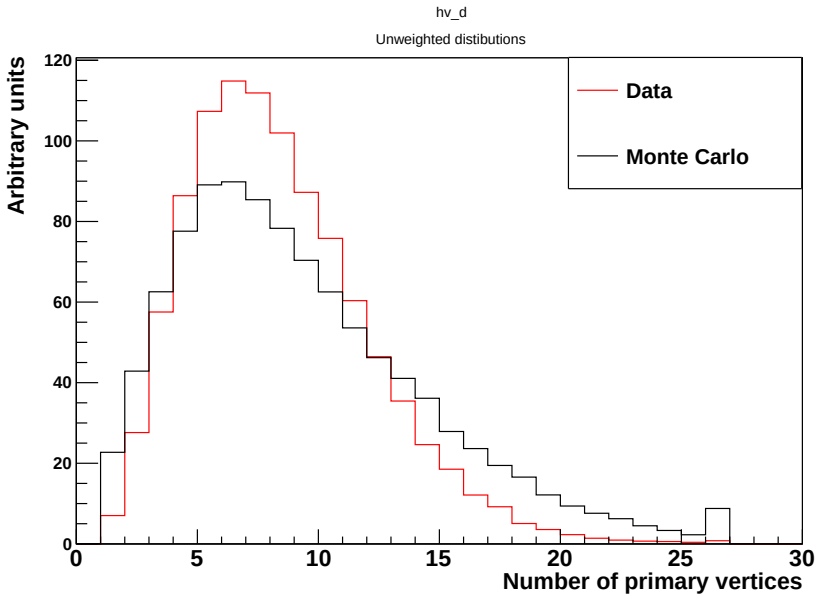


Figure 4.4: Distribution of number of primary vertices in data and Monte Carlo ($\mu\mu$ sample). The MC distribution is unweighted.

4.7.2 Jet and HT correction

The calorimetric deposits from minimum-bias pile-up events generate an overall bias in energy measurement that is particularly important for jet reconstruction and thus for the computation of H_T.

This pile-up effect can be accounted for both data and Monte Carlo performing the so-called *L1 Fastjet subtraction*[1][72].

[1]The Level 1 jet correction is actually a subtraction of the energy not directly associated to the hard scattering. Thus it accounts not only for the pile-up but also, for example, for the energy background raising from the electronic noise. It has to be said, for completeness, that the Level 1 correction (as the name itself suggests) is not the only possible jet correction. Higher level correction are usually less important and take into account non-linear effects in calorimetry response or simulation calibration effects.

An average background energy deposit (E_{bkg}) is computed and its surface density

$$\rho = \frac{E_{bkg}}{\Sigma_{calo}} \tag{4.7}$$

is stored for every event (figure 4.5). The corrected energy of a jet is given by:

$$E_{corr} = E_{jet} - \rho \times \Sigma_{jet} \tag{4.8}$$

where E_{jet} is the measured jet energy and Σ_{jet} is the base surface of the jet cone.

In a similar way the corrected H_T is obtained summing up the transverse momenta of the L1 fastjet corrected jets

$$H_T^{corr} = \sum_i (p_T i - \rho \times \Sigma_i). \tag{4.9}$$

A comparison between corrected and uncorrected distribution for the leading jet p_T and H_T in $t\bar{t}$ Monte Carlo sample and in data is shown in figure 4.6.

4.7.3 Correction of leptons isolation

The energy bias parametrized by the average density ρ also affects the isolation of electrons and muons. If an energy background is introduced, in fact, leptons result less isolated than they actually are. Every isolation term (TkIso, EcalIso, HcalIso) can be corrected using a formula similar to the one for the L1 fastjet subtraction (equation 4.9):

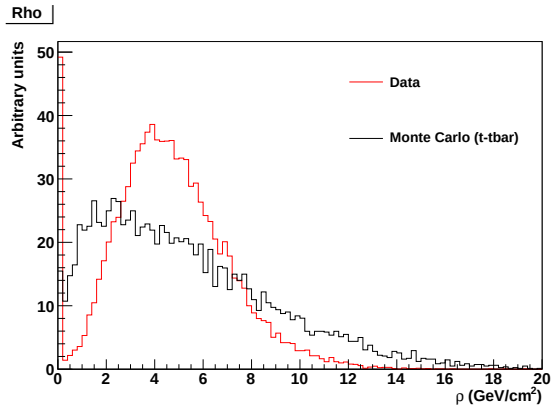


Figure 4.5: Distribution of the average density of energy deposit (ρ) in data and Monte Carlo.

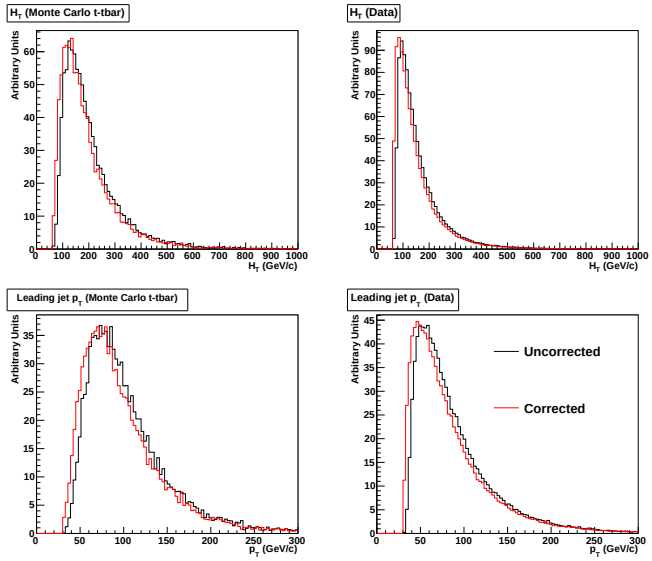


Figure 4.6: Comparison between corrected and uncorrected variables for data and $t\bar{t}$ Monte Carlo. Distributions of H_T and of p_T of the leading jet are plotted.

$$Iso^{corr} = Iso - \rho \times \Sigma_{\text{effective}} \tag{4.10}$$

where $\Sigma_{\text{effective}}$ is an effective area obtained by imposing the isolation criteria to be independent on the number of primary vertices (hence on the pile-up).

In figure 4.7 a distribution of the isolation vs. number of primary vertices is shown for $\mu\mu$ data. A fit on these distributions shows a linear dependence between isolations and pile-up. It can be shown[73] that this dependence can be removed using the following definition for $\Sigma_{\text{effective}}$:

$$\rho \times \Sigma_{\text{effective}} = \frac{\text{slope}(Iso)}{\text{slope}(\rho)} \tag{4.11}$$

that is using the ratio of the slopes of the linear fits of isolations and ρ distributions.

4.8 Systematics in Monte Carlo simulations

In every counting analysis, Poisson statistics is used. Given a number N of events, the statistical uncertainty is taken into account using the standard deviation of a Poisson distribution:

$$\sigma = \sqrt{N} \tag{4.12}$$

Sources of systematic errors have also to be accounted. They originate from uncertainties on theoretical models or from biases of the measurements. The estimate of the amount of these errors is often not trivial and

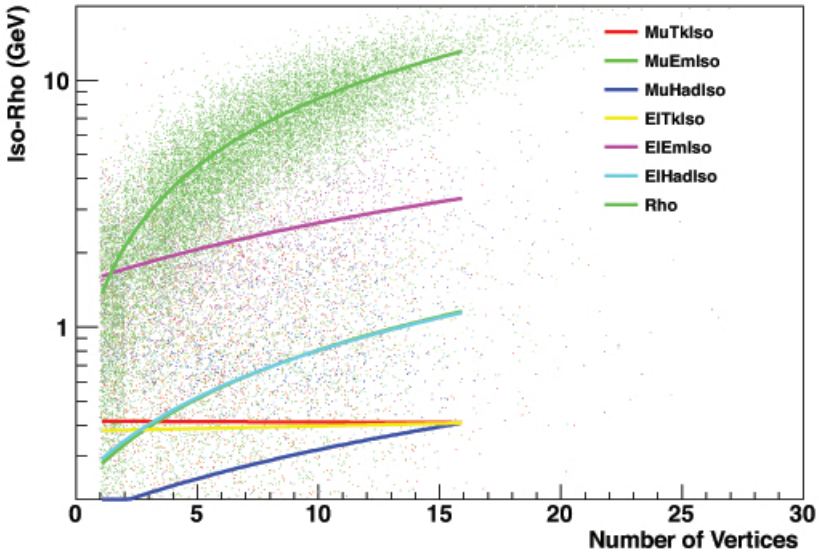


Figure 4.7: Distribution of muons and electrons isolations (and of energy density ρ) versus number of primary vertices in $\mu\mu$ data. Linear fits of every distribution is also shown.

is generally evaluated varying a certain quantity related to the source of uncertainty and observing how it influences the final yield.

In order to perform a correct comparison between data and Monte Carlo, many sources of systematics have to be considered. The effects of these uncertainties are summarized in table 4.6, and are described in more details in the following.

A total systematic error of 9% is obtained. This is a conservative estimate, since for every systematic source the worst case has been considered. This has been done to avoid underestimations and to cover other secondary sources.

Source of systematics	Error
Jet Energy Scale	±6.7%
Trigger Efficiency	±2%
Pile-up	±1.1%
Luminosity	±4.5%
Lepton reconstruction	±2%
Lepton isolation	±2%
Total (conservative)	±9%

Table 4.6: Systematic uncertainties for the MC simulations of Standard Model processes.

For a faster analysis, some of the errors (Jet Energy Scale) are evaluated for the $t\bar{t}$ background, which dominates after all the analysis cuts. Moreover the results shown in the plots are relative to the $\mu\mu$ channel. Similar results are obtained for the other two channels.

Jet Energy Scale

In CMS, energy is measured by charge collecting detectors. The amount of charge collected has to be transformed into an energy deposit. The conversion factor is called *Jet Energy Scale* (JES) and its estimation requires precise calibrations. The uncertainty related to them has to be accounted as a source of systematics.

It has been shown[74] that JES uncertainty for every jet is always smaller than 5% for jets with a $p_T > 30$GeV. The effect of JES uncertainty on our data/MC comparison can be then evaluated by varying JES correction by ±5% and estimating how this modify the total number of Monte Carlo ($t\bar{t}$) events selected (table 4.7).

Channel	Selected	JES +5%	JES −5%	Variation (symm.)
$\mu\mu$	2540	2666	2394	±5.7%
$e\mu$	5792	6077	5471	±5.5%
ee	1998	2107	1864	±6.7%
Total Error (conservative)				±6.7%

Table 4.7: Systematic effects in $t\bar{t}$ yields due to modification on Jet Energy Scale. Even if positive and negative shifts in JES generate asymmetric variations for yields, the error is considered symmetric (for excess).

In figure 4.8 the effects due to the variations in JES on the p_T of the leading jet and on the H_T distributions for the $\mu\mu$ channel are shown.

A maximum uncertainty of ±6.7% is found, and is applied to all the channels. It has been calculated on the $t\bar{t}$ sample only, in order to perform a faster simulation. However it should be stressed that the $t\bar{t}$ background is the most important one for our analysis and is almost the only one which survives after the selection cuts are applied, hence the results obtained can be extended to the whole Monte Carlo sample without significant error.

Trigger efficiency

It has been shown for opposite sign dilepton analyses[75] that uncertainty for the efficiency of the dilepton triggers is always smaller than 2%. This value will be taken for our estimate.

Pile-up

The number of pile-up interactions in data are evaluated from the measurement of instantaneous luminosity multiplied by the total inelastic cross section. Both cross section and luminosity are affected by errors that have to be accounted in the PU reweighting process.

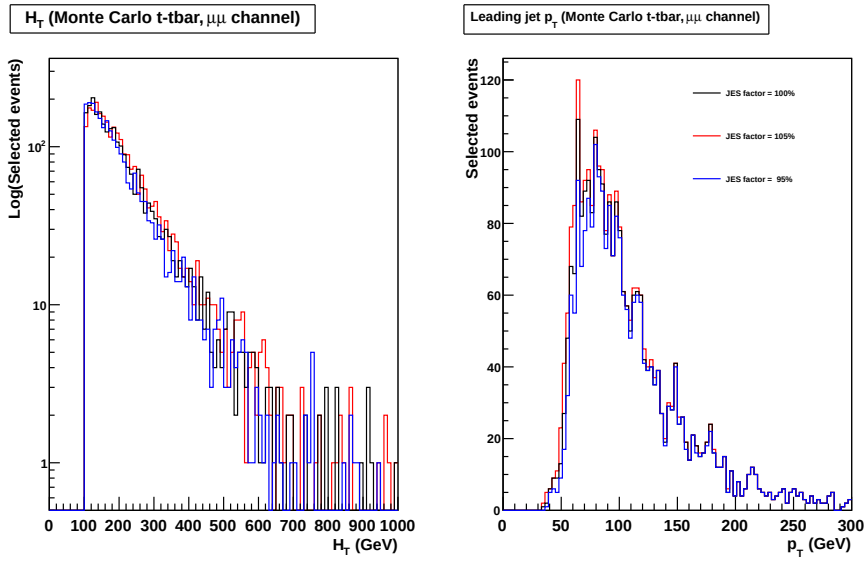


Figure 4.8: Effect of ±5% JES variations in H_T and leading jet p_T ($\mu\mu$ channel , $t\bar{t}$ sample).

The PU systematics are studied by shifting the mean of the distribution of the number of vertices by ±0.6 (about 1/10 of the mean value, figure 4.9). The variation of event yield is then analyzed (table 4.8).

A total uncertainty of ±1.1% is found.

Luminosity

Monte Carlo samples are normalized to the recorded luminosity of data ($\mathcal{L} = 4.98 fb^{-1}$). The recorded instantaneous luminosity is measured in CMS with different methods[76]. The uncertainty related to these measurements is estimated to be ±4.5%.

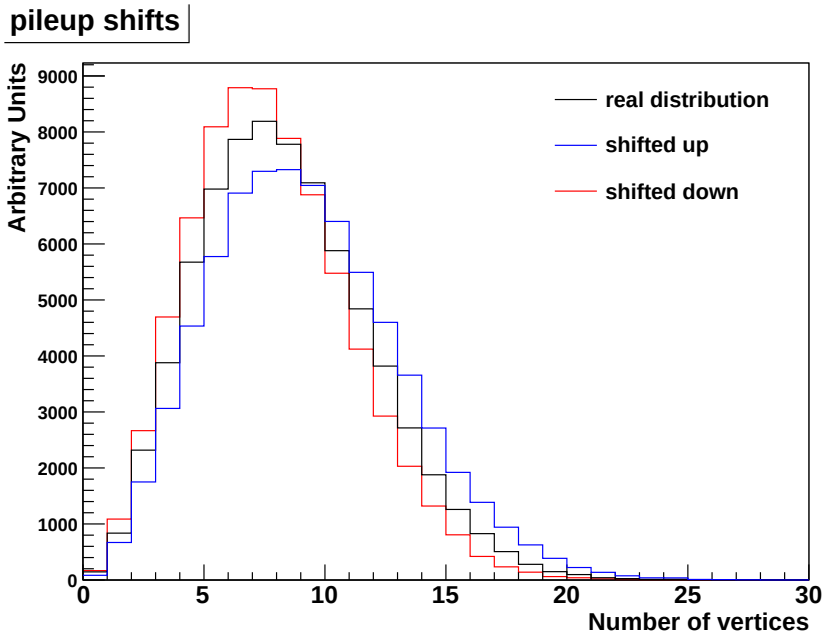


Figure 4.9: Distribution of primary vertices in data ($\mu\mu$ sample) and shifts of the mean value.

Lepton reconstruction

The probability for muons and electrons to be reconstructed, identified and isolated have been studied in detail[70] using a Tag and Probe method based on $Z \rightarrow l^{+}l^{-}$ events. A total uncertainty of 2% for both muons and electrons has been calculated. In addition another 2% of error is taken into account to cover a potential mis-modelling in lepton isolation evaluation[75].

Channel	Selected	+0.6 shift	-0.6 shift	Variation (symm.)
$\mu\mu$	2697	2727	2669	$\pm$1.1%
$e\mu$	5827	5855	5804	$\pm$0.8%
ee	2713	2734	2693	$\pm$0.5%
Total Error (conservative)				$\pm$1.1%

Table 4.8: Systematic shifts due to modification on primary vertex distribution. Even if positive and negative shifts generate asymmetric variations, the error is considered symmetric (for excess).

4.9 Data/MC comparison

The data and Monte Carlo yields for selected events are shown in table 4.9, for every channel. The contributions from each background sample is highlighted. The Monte Carlo yields have been first normalized to data luminosity ($\mathcal{L} = 4.98 fb^{-1}$) and scaled for the trigger efficiencies (1 for ee, 0.95 for $e\mu$, 0.9 for $\mu\mu$). Finally MC distributions have been reweighted such that the distributions of primary vertices matches that in data.

A substantial agreement in selection yields for data and Monte Carlo is found. The relative differences (table 4.10) are at most of $\sim 6\%$, which can completely be attributed to the systematic uncertainties discussed above.

Sample	ee	$\mu\mu$	$e\mu$	total
$t\bar{t}$	2299.66 ± 77.22	2302.11 ± 79.30	5263.47 ± 192.23	9865.24 ± 221.82
DY - M10To50	11.09 ± 0.11	6.99 ± 0.09	14.02 ± 0.21	32.10 ± 0.25
DY - M50	224.25 ± 7.53	268.34 ± 9.24	276.42 ± 10.09	769.01 ± 15.62
top - s-ch	0 ± 0	0 ± 0	0 ± 0	0 ± 0
top - t-ch	1.31 ± 0.11	0.93 ± 0.09	0.65 ± 0.21	2.89 ± 0.25
top - tW-ch	84.84 ± 2.85	80.70 ± 2.78	138.22 ± 5.05	303.76 ± 6.43
VV	42.05 ± 1.41	24.99 ± 0.86	65.93 ± 2.41	132.97 ± 2.92
W	50.14 ± 0.61	13.46 ± 0.19	68.80 ± 0.93	132.40 ± 1.13
MC	2713.34 ± 77.65	2697.52 ± 79.88	5827.51 ± 192.57	11238.40 ± 222.47
(syst)	(± 244.2)	(± 242.8)	(± 524.5)	(± 1011.5)
Data	2820	2842	5684	11346
LM1	257	351	156	764
LM4	28	41	64	133
LM13	111	108	272	491

Table 4.9: Event yields for Monte Carlo (corrected for PU effects) and Data. Monte Carlo are normalized to data luminosity and scaled for trigger efficiency. Systematic uncertainties are shown in parentheses for the sum of all Monte Carlo samples. SUSY yields at LM1, LM4, and LM13 are also shown for comparison.

Sample	(Data - MC)/Data
$\mu\mu$	+5.8%
$e\mu$	−5%
ee	+3.8%

Table 4.10: Relative differences between Monte Carlo and data yields.

Even more important for our studies is to check whether there is consistency in the shapes of the variables, that is if Monte Carlo can adequately model data distributions. In figures 4.10-4.13 data and Monte Carlo distributions are compared for some key variables of our analysis:

- $m(ll)$, invariant mass of the dilepton pair;
- $pfMET$, transverse missing energy calculated with the particle flow algorithm;
- H_T;
- $NJets$, number of good jets per event.

Comparisons of other variables (including those used for data-driven background estimate) are shown in appendix B[1].

After the pile-up reweighting and the energy corrections, the distribution of selected data is well modeled by Monte Carlo simulation. They can therefore be used to study data-driven methods of background control, as will be explained in details in the next chapter.

[1] In these plots only statistical uncertainties are shown.

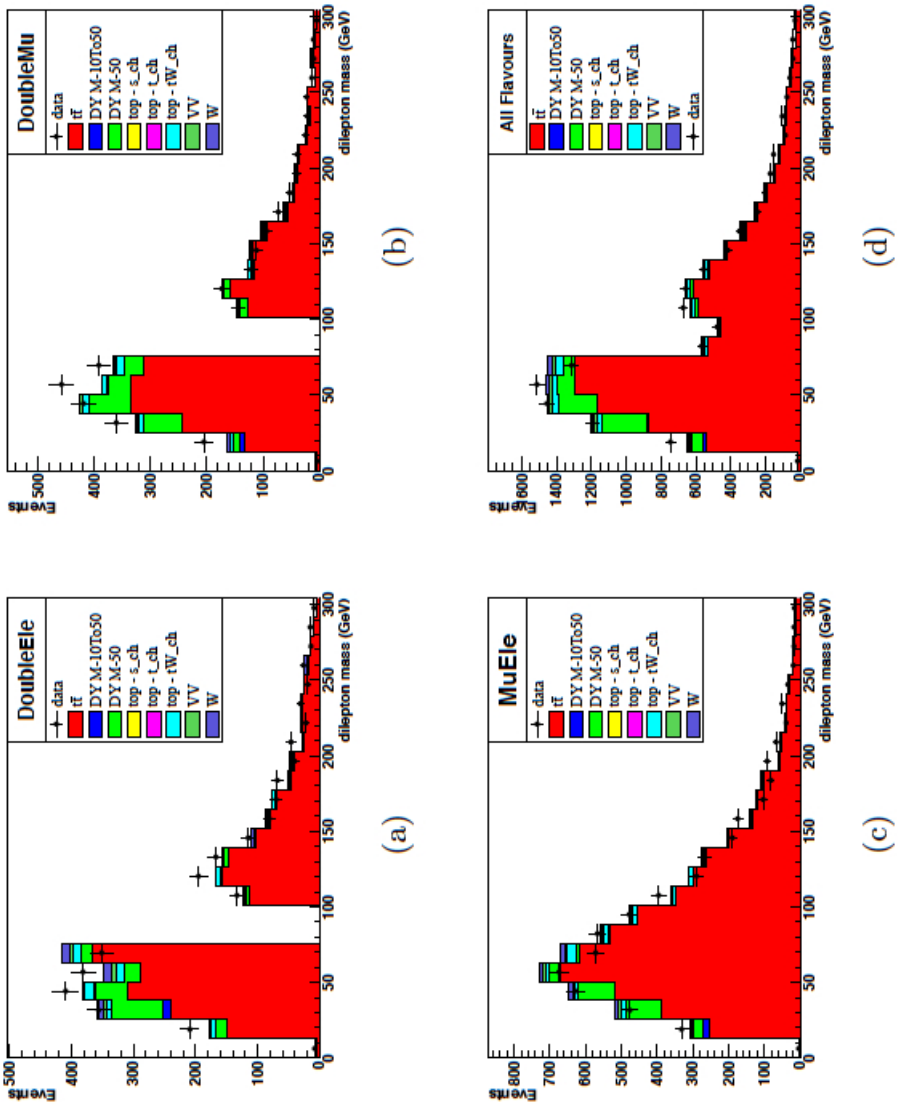


Figure 4.10: Distribution of dilepton invariant mass for data and Monte Carlo in (a) ee channel, (b) $e\mu$ channel, (c) $\mu\mu$ channel, and (d) all flavors.

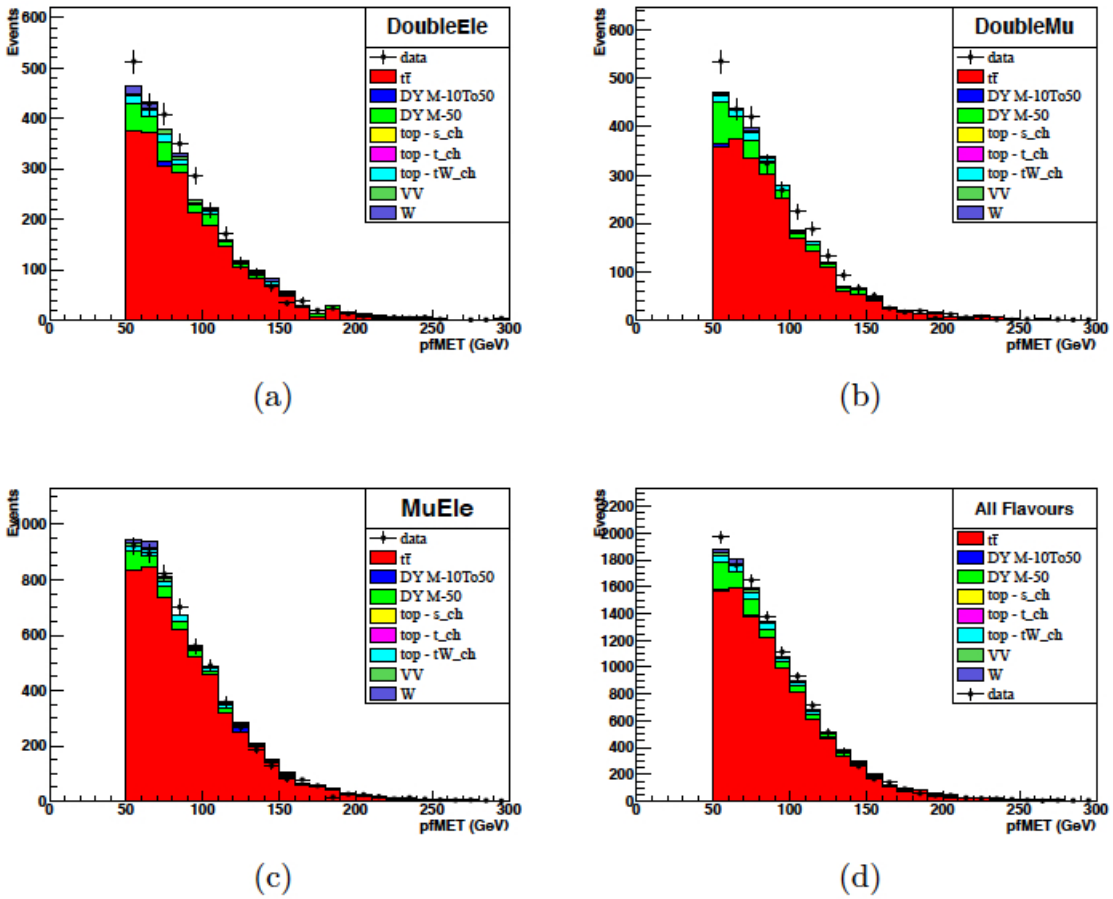


Figure 4.11: Distribution of MET for data and Monte Carlo in (a) ee channel, (b) $e\mu$ channel, (c) $\mu\mu$ channel, and (d) all flavors.

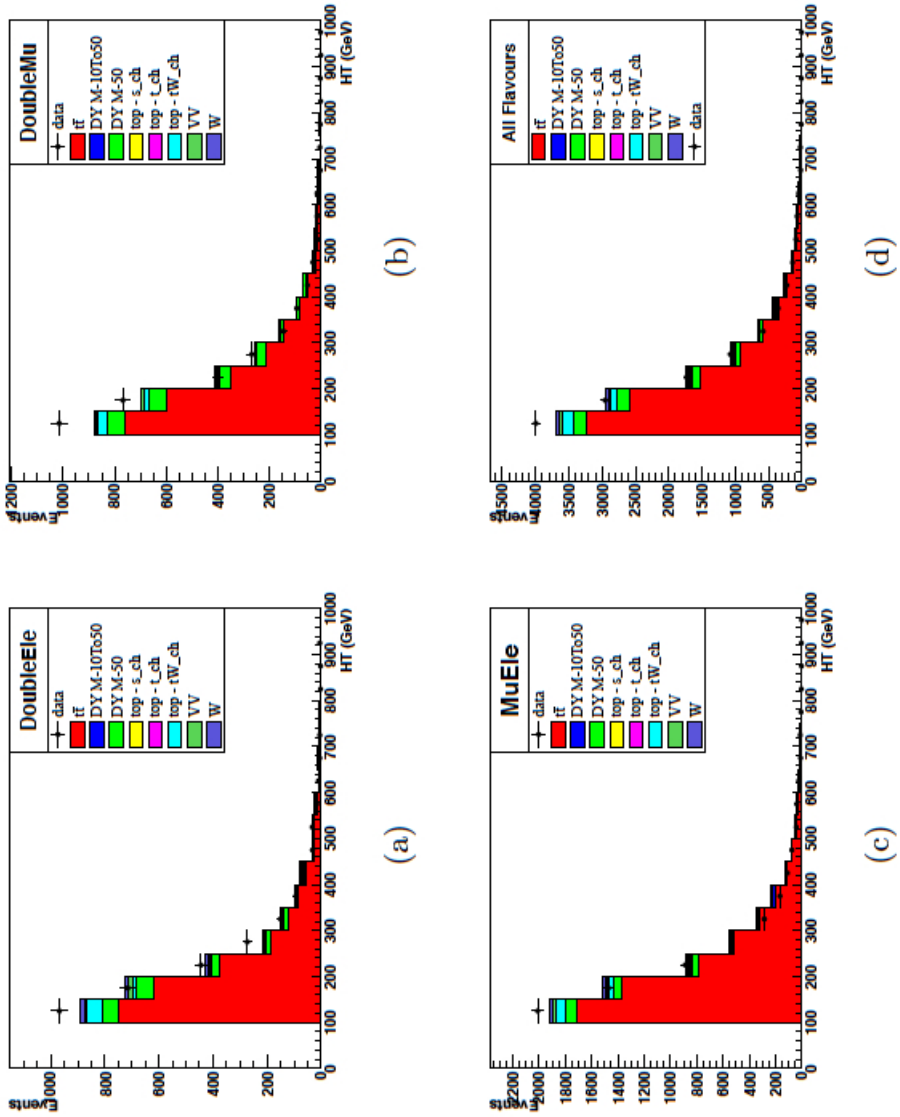


Figure 4.12: Distribution of H_T for data and Monte Carlo in (a) ee channel, (b) $e\mu$ channel, (c) $\mu\mu$ channel, and (d) all flavors.

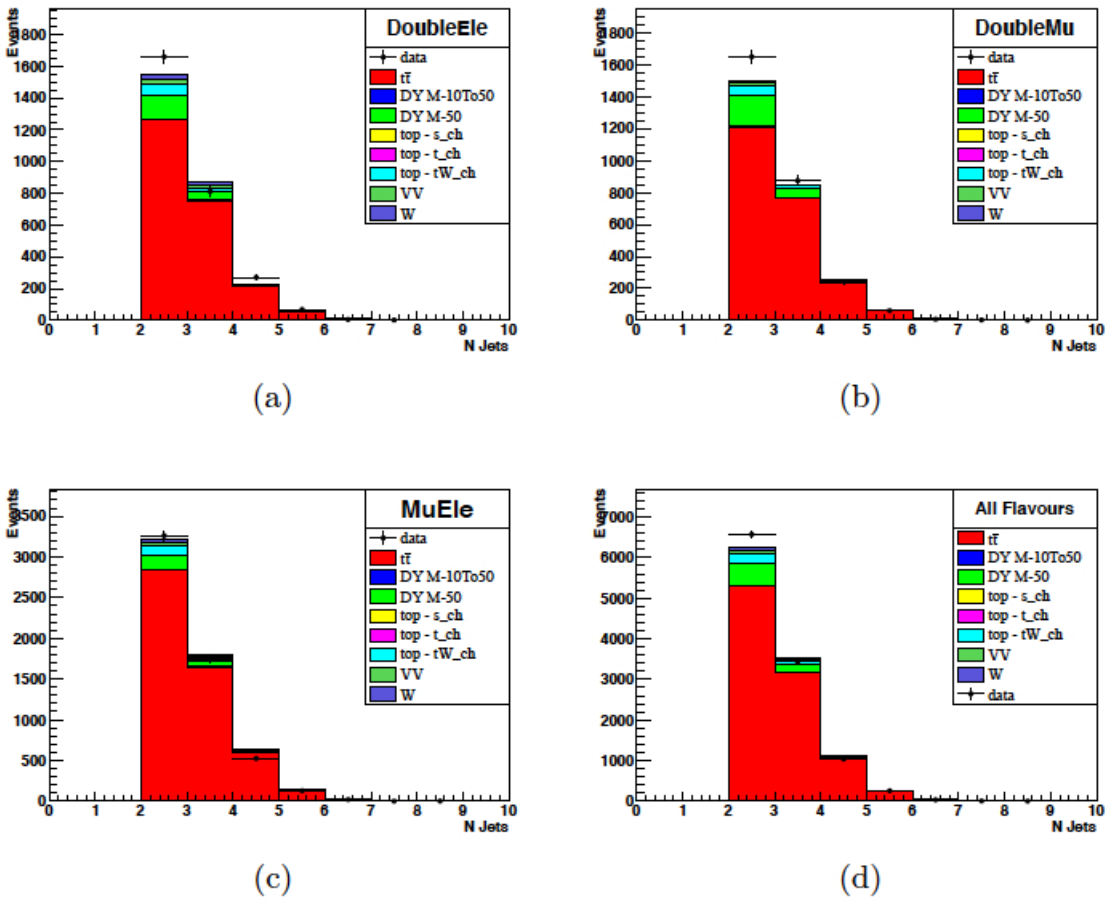


Figure 4.13: Distribution of number of good jets for data and Monte Carlo in (a) ee channel, (b) $e\mu$ channel, (c) $\mu\mu$ channel, and (d) all flavors.

5

ABCD estimate

The Monte Carlo samples studied and validated in the previous chapter are here used to develop and setup a data-driven method, called ABCD method, for the estimate of residual Standard Model background in a region where the occurrence of signal is expected.

After a discussion about the importance of data-driven methods in counting experiments and once introduced the details about the ABCD method, a particular variant of it, which uses many discriminating variables for independent estimates, is setup. Seven variables are analyzed and their discriminating power is investigated. These variables are then coupled to each other and the ABCD method is applied to each pair. Those pairs which show a good and robust estimation power will be used to obtain a background estimate in data (which will be the subject of the next chapter).

5.1 Data-driven methods

A discovery experiment is based on the count of events exceeding the Standard Model background after the application of properly defined analysis cuts. Keeping the background under control is then the fundamental task of the whole analysis, and this could be done in several ways. The first and more immediate, could be a *quantitative* usage of Monte Carlo backgrounds, that is, a search for an excess with respect to what predicted by simulations. However this method has severe limits. Despite the accuracy of our simulations (from the generation of the hard scattering to the simulations of the detectors), biases on the underlying theoretical models or on the extrapolations (for example in the cross sections) can lead to hardly predictable under- or over-estimates.

For this reason it is necessary to develop methods more directly linked to real data and less to the predictions of Monte Carlo simulations. These are called $data-driven$ methods, since the control of the residual background is based on the data.

Many data-driven techniques have been developed by the CMS collaboration for several analyses. Some of them are based on the extrapolation of the shape of some variable to a tail region, where the signal is expected. Other techniques use more than one variable or try to simulate a distribution from a certain template. Although very different in practice, all these methods are based on the existence of a *control zone*, a region where any signal is expected to be dominated by the Standard Model background; a study of this zone should permit to infer the behavior of the background where instead a significant amount of signal could rise.

The method developed in this thesis belongs to the class of the so called $ABCD$ techniques.

5.1.1 The ABCD method

In the ABCD method a pair of properly chosen discriminating variables is studied. The plane defined by the pair is divided in four different areas, called A, B, C, and D (figure 5.1). One of these, usually labeled with "C", is treated as a *signal region*, i.e. the plane is to be divided in such a way that any signal of new physics is expected to be found in the C region while its contamination in the other three regions could be neglected. A, B, and D regions are called *background or control regions*, since the play a key role in the data-driven control of the residual background in the signal region. This first step needs to be driven by Monte Carlo simulations.

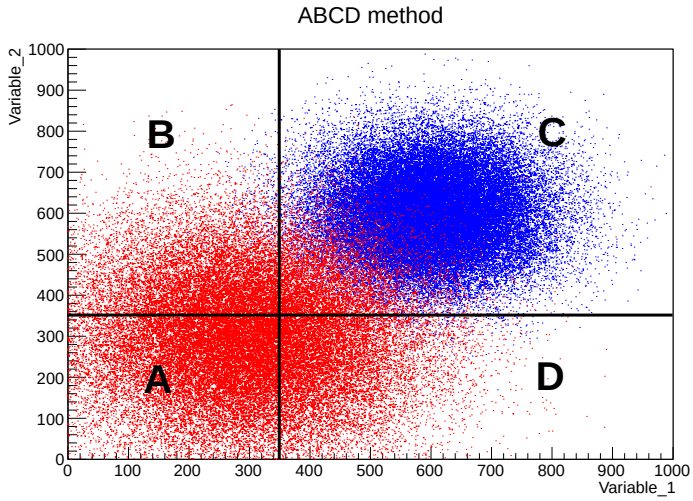


Figure 5.1: ABCD division of the plane defined by two uncorrelated variables. In blue an hypothetical signal distribution, in red an hypothetical background. This color convention will be maintained in the following plots.

A suitable pair of variables has to fulfill two fundamental requirements, to be verified via Monte Carlo:

- they must have good discriminating power between signal and background;
- they need to be uncorrelated.

If the last condition is fulfilled, it is easy to prove[1] that regardless the choice of regions, the number of events in them satisfies the relation

$$\frac{N_B}{N_A} = \frac{N_C}{N_D}. \tag{5.1}$$

So, the number of events in the region C can be inferred from that in the control regions:

$$N_C^{\text{est}} = N_D \times \frac{N_B}{N_A} \tag{5.2}$$

If the signal contamination in the control regions is actually negligible (i.e. its presence does not modify A, B, and D), the N_C^{est}, evaluated using the 5.2, is a data-driven estimate of the background contamination in the signal region.

Obviously, what just described represents an ideal condition: it is impossible to find two variables totally uncorrelated or with a perfect separation between signal and background. It is then important to perform a

[1] Let us consider a distribution of two variables v_1 and v_2 with density $\rho(v_1, v_2)$. The total number of events in one of the four regions is then:

$$N_i = \int dv_1 \int dv_2 \rho(v_1, v_2).$$

If v_1 and v_2 are uncorrelated, the density can be factorized as

$$\rho(v_1, v_2) = \rho_x(v_1) \times \rho_y(v_2),$$

hence the 2-dimensional integral become just a product of two 1-dimensional integrals. This liearization explains why equation 5.1 works.

validation of the method and closure tests, in order to define well-working pairs. The variables that provide a good separation power between signal and background (in particular $t\bar{t}$) are few, and are never completely uncorrelated. The following paragraphs are dedicated to the setup of the ABCD method, starting from the choice of some promising variables.

It is clear that any data-driven analysis is not completely independent from the Monte Carlo simulations. Its importance still resides in the definition of the regions and the choice of pairs. There partially remains, therefore, a dependency on theoretical models, which, however, is far less dramatic than in the case in which the estimate totally relies upon simulation.

5.2 Discriminating variables

Unlike other ABCD analyses developed by the CMS collaboration in the past[70], the method described in this work is not focused on the study of one single pair of discriminating variables, but is rather based on a parallel study of several, virtually well discriminating, variables. For each variable, regions of signal and background are defined using Monte Carlo simulations. Each variable is then coupled with the other and systematic studies, including correlation analyses and evaluation of the estimate power, are performed for each pair. The most promising pairs are selected for a data-driven estimate of the residual background. Using multiple pairs of variables at the same time would allow to combine independent estimates, which would lead to statistically robust results.

Seven discriminating variables have been studied, which implies 21 different pairs. Three of them have been chosen among the variables involved in the analysis (see section 4.3). They are:

- the invariant mass of the lepton pair, m(ll);
- the missing transverse energy (MET);
- the scalar sum of the jet transverse momenta (H_T).

We will refer to these variable as the *in-analysis variables*. When they are used in an ABCD estimate, the standard selection cut on them is not applied (indeed, the region below the cut value is part of the background region). The other four variables are called the *out-of-the analysis variables*, since they do not take part to the standard selection. When they are not used as ABCD variables, no cut on them is applied. The usage of the variables in our analysis is illustrated in figure 5.2.

The out-of-the analysis variables are more complex than the in-analysis ones, and they will be defined in the next paragraphs. These variables are:

- the invariant mass of the most massive hemisphere (or HemiMax);
- the invariant mass of a particular lepton-jet pair (m(lj));
- the lepton-MET transverse mass (m_T(l-MET));
- $Y = MET/\sqrt{H_T}$.

5.2.1 In-analysis variables

The in-analysis variables have been already introduced in section 4.3. Hence we will only focus our discussion on their discriminating power between signal and background.

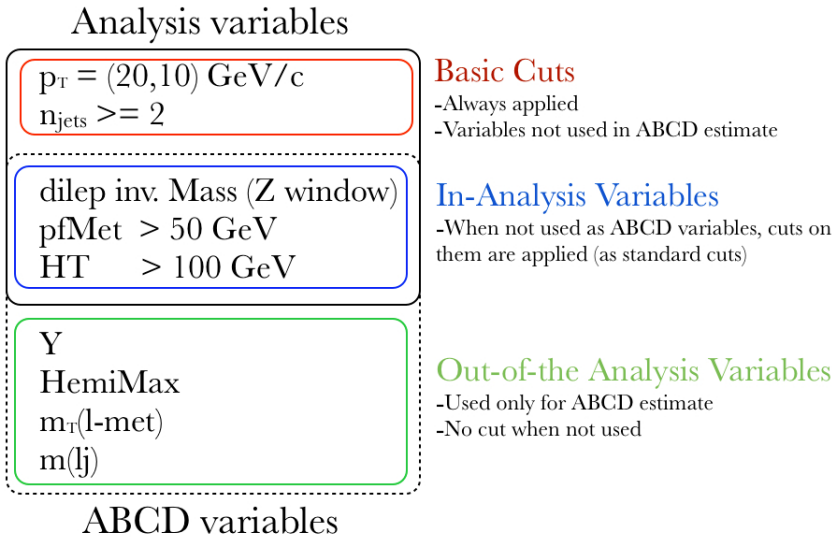


Figure 5.2: Variables used in this analysis.

m(ll)

The invariant mass of the dilepton pair (m(ll)) is a combination of the energies and the momenta of the two selected leptons:

$$m(ll)^2 = (E_{l1} + E_{l2})^2 - (\overrightarrow{p_{l1}} + \overrightarrow{p_{l2}})^2 \tag{5.3}$$

If the two leptons are the products of the same decay, this quantity represents the rest mass of the parent particle. As mentioned in section 3.5, a systematic study of this quantity would permit a spectroscopy of the particles involved in particular chains. Unfortunately, this means that the shape of the m(ll) distribution for a signal sample strongly depends on the mass spectrum, hence we have no hints on where SUSY would manifest. In figure 5.3 the invariant mass distributions for SUSY at LM1, LM4, and LM13 are compared, and they show a clearly different behavior.

Therefore, if on one hand it is difficult to uniquely define a signal region, on the other, the Z-window, where most of the Standard Model background concentrates (figure 4.3), represents a perfect control region. For this reason, m(ll) will be used as ABCD variable.

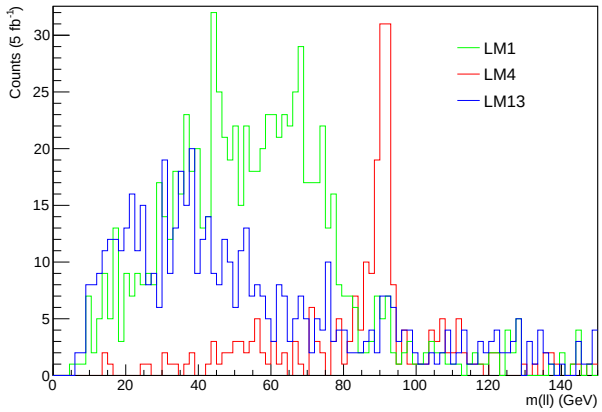


Figure 5.3: m(ll) for SUSY samples at LM1, LM4, and LM13.

MET and H_T

Unlike m(ll), whose distribution is strongly mass-dependent, high missing transverse energy and high jet activity (hence, high H_T) are expected to be common signatures of many models beyond the SM. Therefore, high MET and low H_T regions will be used as signal regions, while low MET and H_T will be used as control regions.

5.2.2 Out-of-the-analysis variables

The out-of-the analysis variables are particular combinations of basic variables (like leptons or jets momenta and energies) or more complex global

variables built to optimize the separation between signal and background. This can be achieved in two ways:

- by properly combining variables with good discriminating power, like MET and H_T;
- by combining variables in such a way that the background manifests a characteristic behavior.

Y and m_T(l-MET) belong to the first of these classes.

Y

Y is defined as

$$Y = \frac{\text{MET}}{\sqrt{H_T}} \tag{5.4}$$

It is therefore a combination of two in-analysis variables. Previous works[70] demonstrated that it has a good separation power in a wide area of the cMSSM plane.

$\mathbf{m_T}$(l-MET)

m_T(l-MET) is the transverse mass of the leading lepton with the MET:

$$m_T(\text{l} - \text{MET}) = \sqrt{(E_T^l + \text{MET})^2 - (p_T^l + \text{MET})^2}\,. \tag{5.5}$$

m(lj)

m(lj) is the invariant mass of one of the four possible combination of the two leptons with the first two jets in the event. In particular it is defined

as:

$$m(\mathrm{lj}) = \min\{\max[\, m(l1, j1)\,,\, m(l2, j2)\,]\,,\, \max[\, m(l1, j2)\,,\, m(l2, j1)\,]\,\} \tag{5.6}$$

where $j1$ and $l1$ are the leading jet and lepton, $l2$ is the second lepton and $j2$ is the sub-leading jet. It can be shown that the distribution of such combination has an end-point for $t\bar{t}$ at

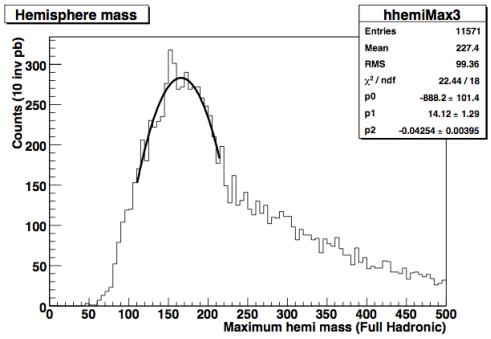
$$\sqrt{m_t^2 - m_W^2} \sim 153\,\mathrm{GeV}\,, \tag{5.7}$$

if both the lepton and the jets come from the same top. Therefore the overall distribution for the background surviving the selection cuts presents a peak at $\sim$ 150GeV, followed by a fairly steep fall. Hence the region between 100GeV and 150GeV represents a natural background region.

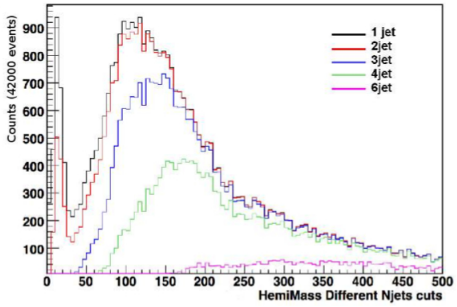
HemiMax

HemiMax stands for the invariant mass of the most massive hemisphere in the event. The *hemisphere separation* [77] is a class of recursive algorithms which associate every object of an event to one of two groups called hemispheres. The hemisphere separation is particularly useful when we need to separate the objects coming from the decay of two primary particles. If the separation is well performed, and all the events is completely reconstructed, the invariant mass distributions of both the hemispheres is centered on the mass of the parent particle. This is for example the case of $t\bar{t}$ production (figure 5.4 (a)).

Anyway, due to mis-reconstructions and mainly to the presence of neutrinos in leptonic and semi leptonic decays, the hemisphere mass distribution for $t\bar{t}$ events is centered to smaller values and is broader than expected (figure 5.4 (b)).



(a)



(b)

Figure 5.4: Studies on the hemispheres reconstruction (from[77]). (a) Hemisphere mass for a fully hadronic $t\bar{t}$ sample. (b) Hemisphere mass for $t\bar{t}$ samples with different number of jets. In leptonic and semi leptonic decays, the hemisphere mass distribution is broader and centered to smaller values than the fully hadronic decays.

The most massive hemisphere is on average the better reconstructed, hence the distribution of its invariant mass will be more similar to the theoretical one. In this case the region between 100 GeV and 180 GeV represents a natural background region.

5.3 Setup of the method

5.3.1 Definition of signal and control regions

The background and signal regions for every variable have been defined looking at their 1-dimensional distributions, and comparing the simulated Standard Model background with SUSY signal at LM1, LM4, and LM13. These distributions (compared with LM1) are shown in figures 5.6 - 5.12. In table 5.1 the boundaries of signal and background regions defined in this way are listed. They are used to outline the A,B,C,and D zones in the planes defined by every pair (figure 5.5). The 2-dimensional distributions of all the 21 pairs with the corresponding ABCD divisions are shown in appendix C.

Note that signal region's boundaries of MET and H_T are actually higher than the standard selection cuts introduced in the previous chapter (MET>50 GeV, H_T >100 GeV). This has been made to enhance the background rejection power. However, when these two variables are not used in any ABCD pair, the standard cuts are applied.

The studies presented in the next sections have been conducted separately in the channels ee, $e\mu$, $\mu\mu$, and in the inclusive channel $ee+e\mu+\mu\mu$ as well. For the sake of simplicity, only the results obtained for this last channel will be shown in the following. The plots regarding the other three can be found in appendix C.

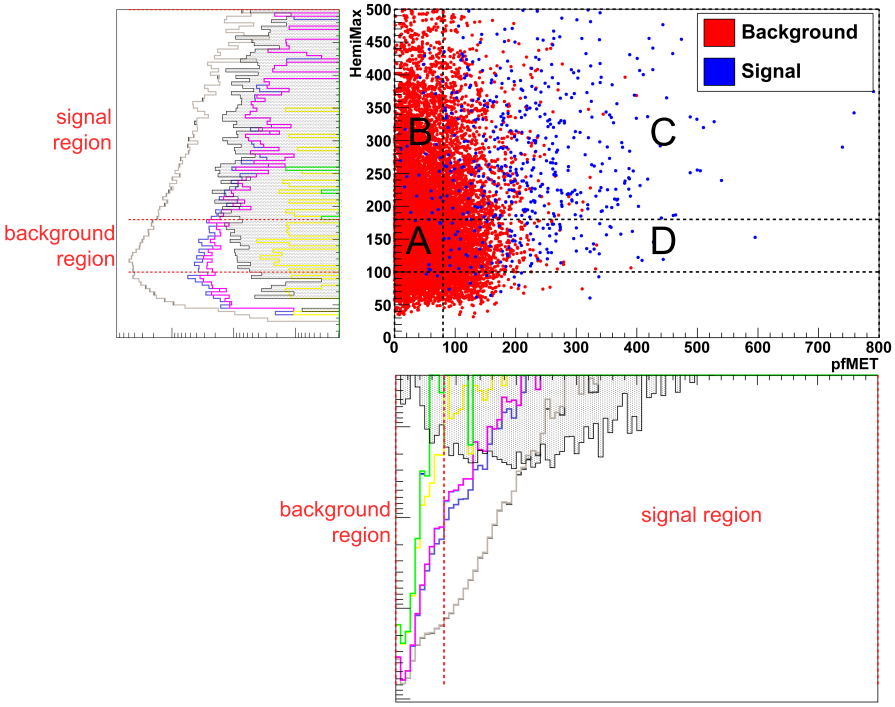


Figure 5.5: The background and signal regions outlined for the one-dimensional distribution of each variable define uniquely the A,B,C,D regions of the two-dimensional distributions. As an example, the ABCD division is shown for MET Vs. HemiMax pair.

Variable	x_0^{bkg}	x_1^{bkg}	x_0^{sig}	x_1^{sig}
m(ll) (GeV)	76	106	0	300
pfMET (GeV)	0	80	80	800
H_T (GeV)	120	150	150	1400
Y ($\sqrt{\text{GeV}}$)	6	7.5	7.5	30
m(lj) (GeV)	80	150	150	500
m_T(l-MET) (GeV)	80	150	150	500
HemiMax (GeV)	100	180	180	500

Table 5.1: Boundaries of the signal and background regions for the variables under study. The background region is the one between x_0^{bkg} and x_1^{bkg} (same for the signal region). The background region of m(ll) is the one outside the Z-window.

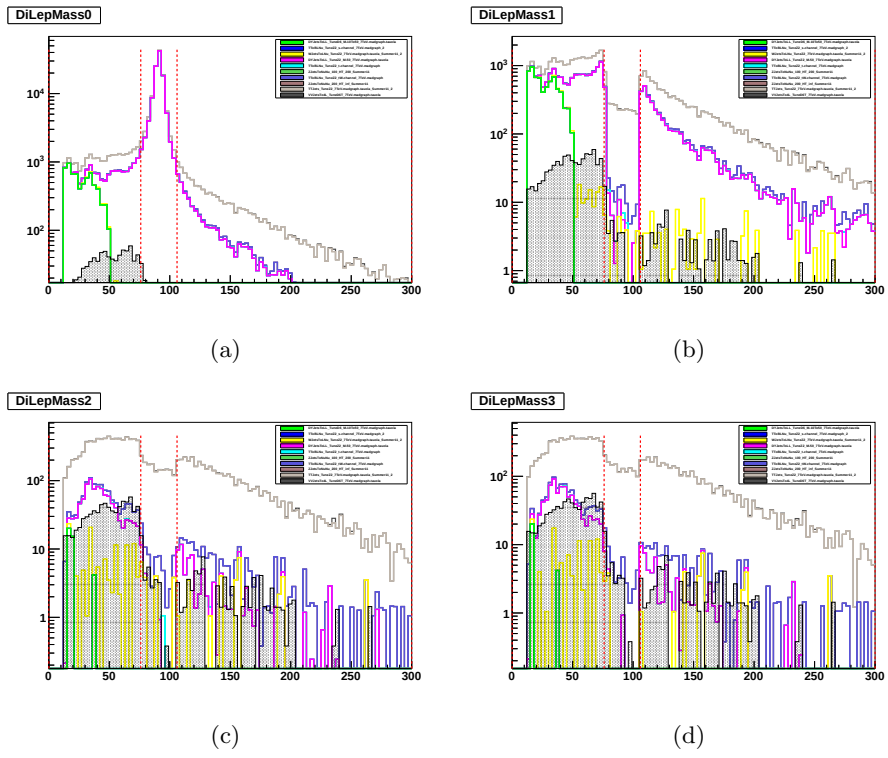


Figure 5.6: Variations in the distribution of dilepton invariant mass m(ll) under the flow of selection cuts: (a) preselected events, (b) Veto on Z window, (c) Veto on Z window + MET>50 GeV, (d) Veto on Z window + MET>50 GeV + H_T >100 GeV. Backgrounds are shown in stack, while the dot-filled histograms show the distribution of signal at LM1. The red dashed lines are the boundaries of the signal region.

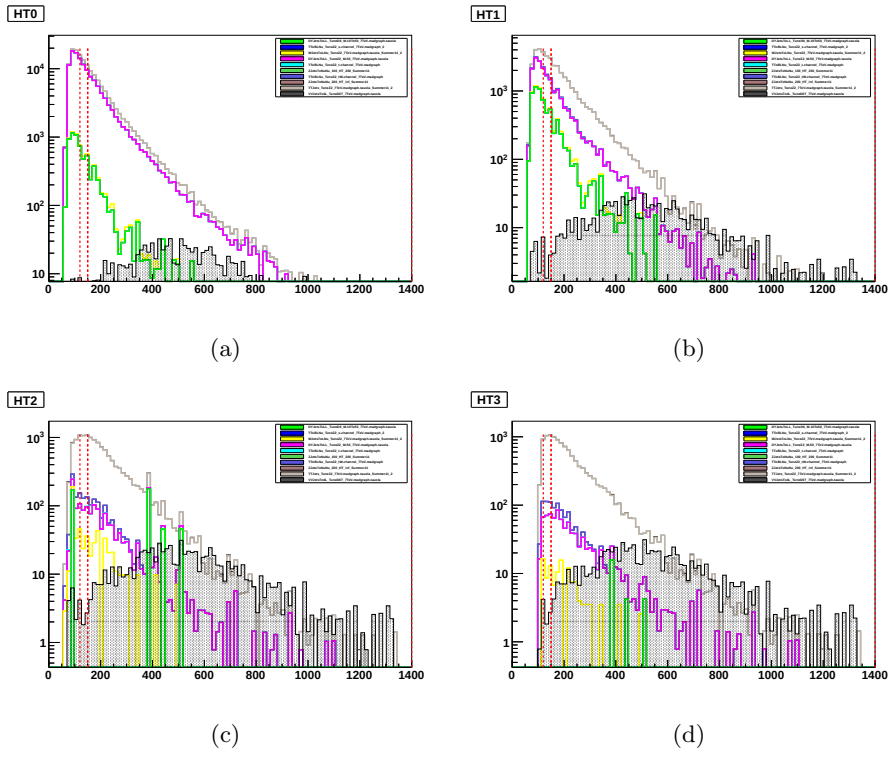


Figure 5.7: Variations of the H_T distribution under the flow of selection cuts: (a) preselected events, (b) Veto on Z window, (c) Veto on Z window + MET>50 GeV, (d) Veto on Z window + MET>50 GeV + H_T >100 GeV. Backgrounds are shown in stack, while the dot-filled histograms show the distribution of signal at LM1. The red dashed lines are the boundaries of the signal region.

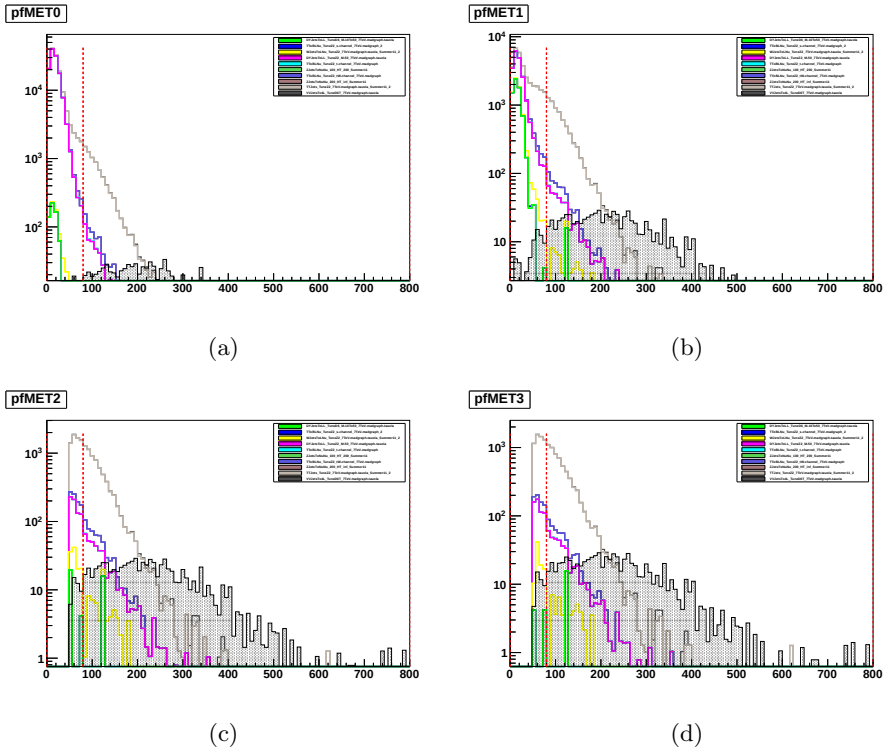
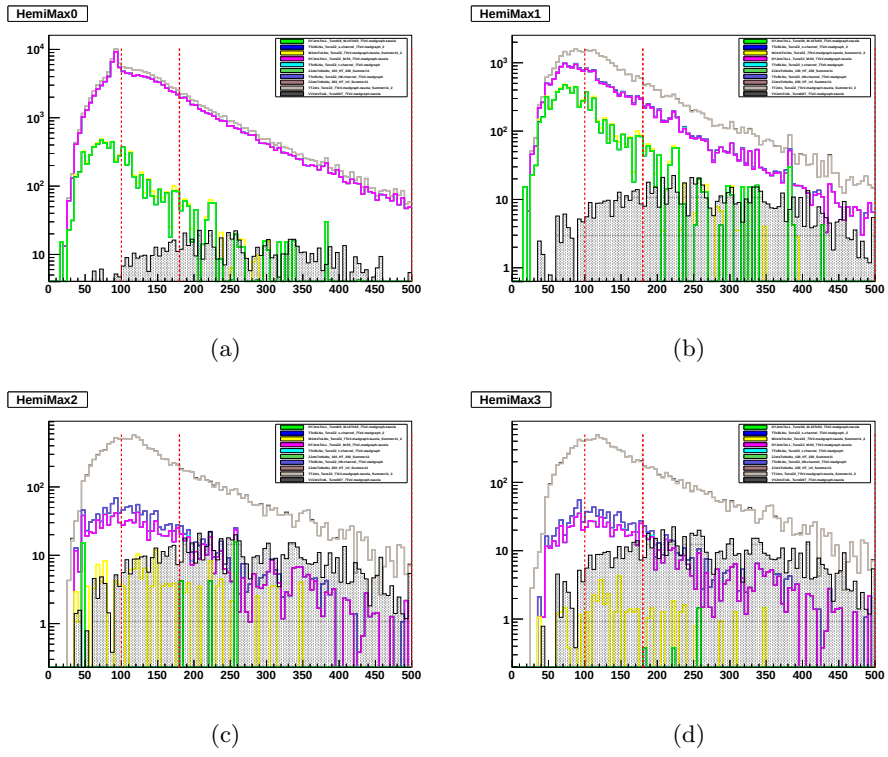


Figure 5.8: Variations of the MET distribution under the flow of selection cuts: (a) preselected events, (b) Veto on Z window, (c) Veto on Z window + MET>50 GeV, (d) Veto on Z window + MET>50 GeV + H_T >100 GeV. Backgrounds are shown in stack, while the dot-filled histograms show the distribution of signal at LM1. The red dashed lines are the boundaries of the signal region.



(a) (b) (c) (d)

Figure 5.9: Variations in the distribution of invariant mass of the most massive hemisphere under the flow of selection cuts: (a) preselected events, (b) Veto on Z window, (c) Veto on Z window + MET>50 GeV, (d) Veto on Z window + MET>50 GeV + H_T >100 GeV. Backgrounds are shown in stack, while the dot-filled histograms show the distribution of signal at LM1. The red dashed lines are the boundaries of the signal region.

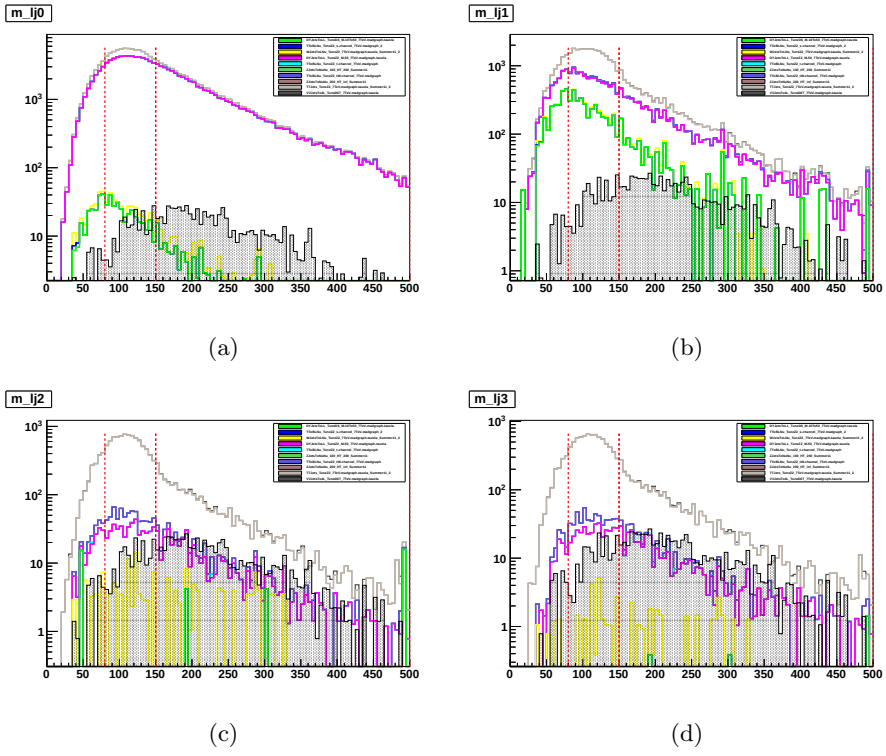
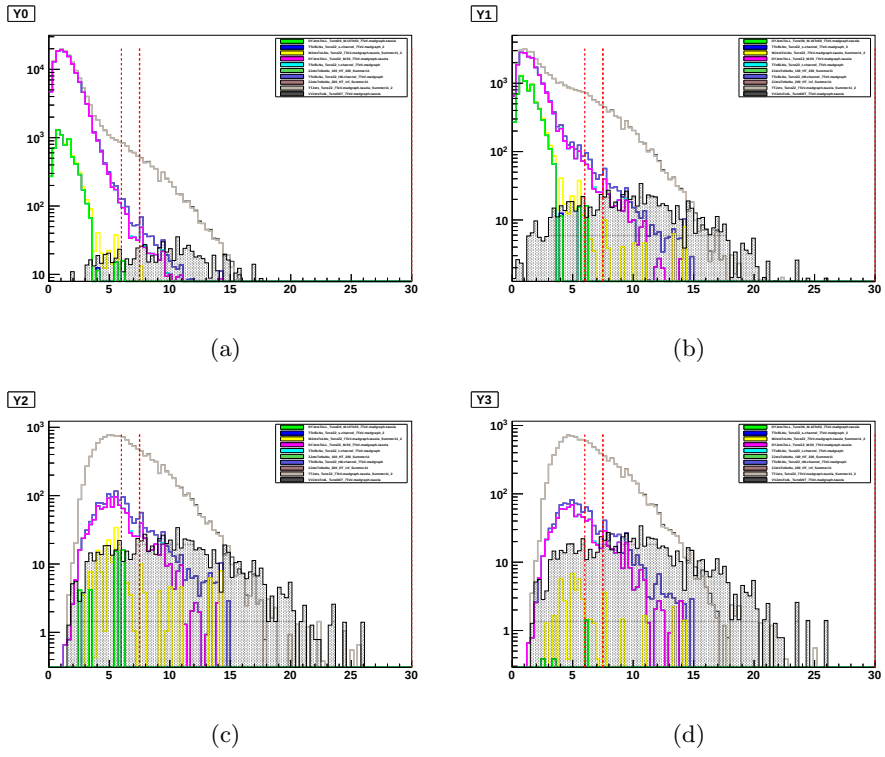
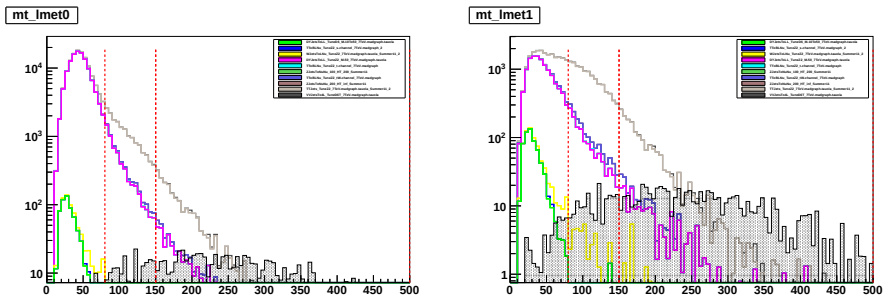


Figure 5.10: Variations in the distribution of m(lj) under the flow of selection cuts: (a) preselected events, (b) Veto on Z window, (c) Veto on Z window + MET>50 GeV, (d) Veto on Z window + MET>50 GeV + H_T >100 GeV. Backgrounds are shown in stack, while the dot-filled histograms show the distribution of signal at LM1. The red dashed lines are the boundaries of the signal region.



(a) (b) (c) (d)

Figure 5.11: Variations of the Y distribution under the flow of selection cuts: (a) preselected events, (b) Veto on Z window, (c) Veto on Z window + MET>50 GeV, (d) Veto on Z window + MET>50 GeV + H_T >100 GeV. Backgrounds are shown in stack, while the dot-filled histograms show the distribution of signal at LM1. The red dashed lines are the boundaries of the signal region.



(a) (b)

(c) (d)

Figure 5.12: Variations in the distribution of m_T(l-MET) under the flow of selection cuts: (a) preselected events, (b) Veto on Z window, (c) Veto on Z window + MET>50 GeV, (d) Veto on Z window + MET>50 GeV + H_T >100 GeV. Backgrounds are shown in stack, while the dot-filled histograms show the distribution of signal at LM1. The red dashed lines are the boundaries of the signal region.

5.3.2 Correlation of the pairs

As mentioned, 21 different pairs are generated with the seven variables under study. Their code-names (whose interpretation is trivial) are listed in the following:

pfMET Vs. HT	HT Vs. HemiMax	DiLepMass Vs. mt_lmet
pfMET Vs. DiLepMass	HT Vs. Y	HemiMax Vs. Y
pfMET Vs. HemiMax	HT Vs. m_lj	HemiMax Vs. m_lj
pfMET Vs. Y	HT Vs. mt_lmet	HemiMax Vs. mt_lmet
pfMET Vs. m_lj	DiLepMass Vs. HemiMax	Y Vs. m_lj
pfMET Vs. mt_lmet	DiLepMass Vs. Y	Y Vs. mt_lmet
HT Vs. DiLepMass	DiLepMass Vs. m_lj	m_lj Vs. mt_lmet

The distributions for all the pairs in $ee + e\mu + \mu\mu$ channel can be found in appendix C.

A necessary condition for the ABCD method to be applied to one pair is that the two variables are not correlated. A first selection is then based on the evaluation of the correlation factor of every pair. An estimator based on the Pearson coefficient[78] ρ has been used. ρ is defined as the ratio between the covariance of the two variables and the product of their variances:

$$\rho = \frac{\mathrm{cov}(V_1, V_2)}{\sigma_{V1}\sigma_{V2}}. \tag{5.8}$$

The *uncorrelation factor* we used in this analysis is defined as:

$$f_{\mathrm{unc}} = 100 \times (1 - |\rho|). \tag{5.9}$$

f_{unc} varies between 0% (for completely correlated variables) and 100% (for totally uncorrelated pairs). The uncorrelation factors of all the pairs

have been evaluated from Monte Carlo samples and are shown in the matrix in figure 5.13.

A threshold of $f_{\text{unc}} > 65\%$ has been imposed to the pairs to be considered sufficiently uncorrelated. 14 out of the 21 pairs passed through the uncorrelation selection (figure 5.14).

5.3.3 ABCD estimate with Monte Carlo

Up to now we used Monte Carlo simulations to

- define signal and background regions for some theoretically well discriminating variables. The discrimination power has been checked comparing background distributions with SUSY signal at three benchmark points (LM1, LM4, and LM13);
- select only those pairs of variables which result uncorrelated (over a certain threshold).

In other words, we proved the two fundamental hypotheses for the application of the ABCD method. Since however we are far from ideal conditions, these hypotheses have to be considered only necessary but not sufficient. The actual estimate power of the residual background using the selected pairs has still to be tested. This requires three steps:

1. prove that we are able to make a correct estimate under the background-only hypothesis (i.e. using Monte Carlo Standard Model background samples);
2. verify that adding some signal do not alter the estimate, which means that the signal contamination of the control regions is maintained within acceptable levels;

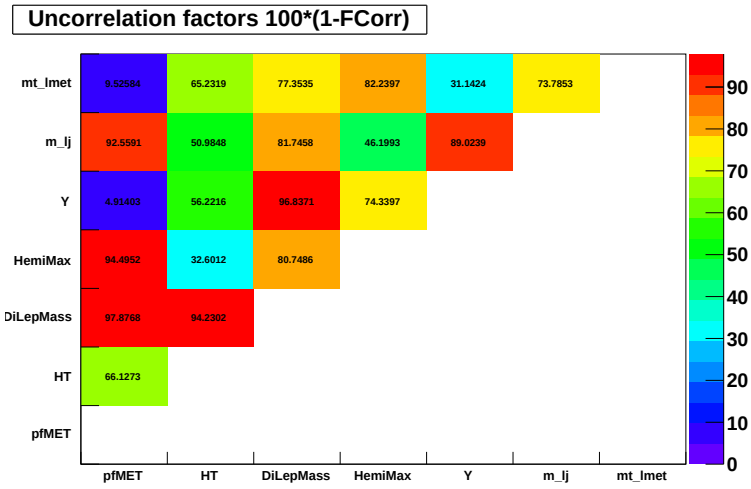


Figure 5.13: Matrix of the uncorrelation factors of the pairs generated by the seven variables under study.

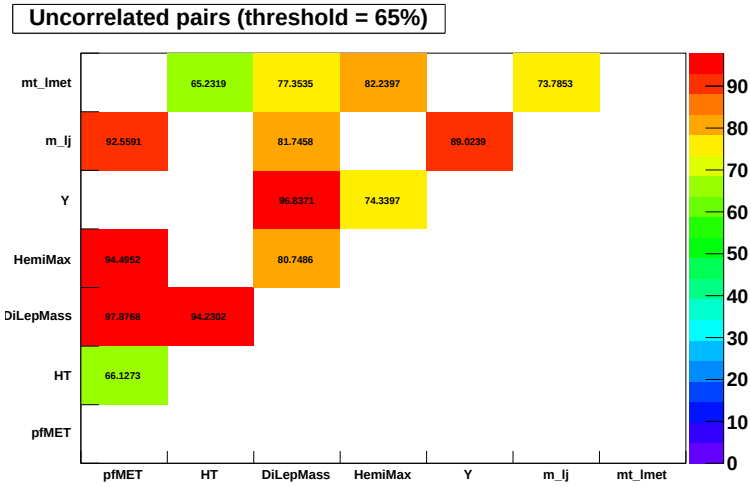


Figure 5.14: Matrix of the uncorrelation factors of the pairs generated by the seven variables under study.

3. prove the possibility to discover or exclude some signal, i.e. prove that the signal could produce significant excesses in C region.

To investigate the first point we count the number of Monte Carlo background selected events in the four regions of the plane defined by each of the 21 pairs. Let us focus on one pair, and call these numbers:

$$N_A, N_B, N_C, N_D;$$

Then, we can estimate the number of background events in the region C using the formula 5.2

$$N_C^{\text{est}}[b] = N_D \times \frac{N_B}{N_A}, \tag{5.10}$$

where b stands for evaluation under the background-only hypothesis. Since the actual number of background events in region C, N_C, is known, we can evaluate the ratio

$$r[b] = \frac{N_C^{\text{est}}[b] - N_C}{N_C} \tag{5.11}$$

to measure the goodness of our estimate. The more $r[b]$ approaches to zero, the better the estimate is. If instead $N_C^{\text{est}}[b] = 0$, then $r[b] = -1$. The ratio has then a minimum, but not a maximum; hence positive values (which indicate over-estimations) and negative values (meaning under-estimations) are not symmetric, and this is to be taken into account for a correct interpretation of the results.

A similar procedure is made to evaluate the goodness of the estimate in the signal+background hypothesis. In this case, a Monte Carlo signal sample is added to the background and, as before, the number of events in the four regions is counted. The Monte Carlo permits to separate in the region C, the events coming from the background and from the signal:

$$N_C(\text{sig}), N_C(\text{bkg}).$$

The ABCD estimate of background in signal region is still

$$N_C^{\text{est}}(\text{bkg})[s+b] = N_D \times \frac{N_B}{N_A}. \tag{5.12}$$

The goodness of the estimate can be evaluated using the ratio

$$r[s+b] = \frac{N_C^{\text{est}}(\text{bkg})[s+b] - N_C(\text{bkg})}{N_C(\text{bkg})} \tag{5.13}$$

which has the same characteristics of $r[b]$. The values of $r[b]$ and of $r[s+b]$ (evaluated using signal at LM1, LM4, and LM13) for all the pairs are summarized in table 5.2

Uncertainties in the table are statistical only[1]. The systematics related to the ABCD method will be studied in the next chapter, since they become important when applying it to data. In figures 5.15 - 5.17, $r[b]$ and $r[s+b]$ are drawn for the three signal samples.

It is useful to comment the results we reported. For one of the pairs (pfMET Vs. Y) an estimate of the background is not possible, because the current choice of signal and background regions leaves the B region empty. This is due to the fact (see figure in appendix C) that MET and Y are actually strongly correlated, since Y is defined to be linearly dependent on MET. So, even if the estimate would have led to any result, this pair would have been discarded. Even Y and H_T are correlated, although less than MET and Y. The correlation in this case is stronger in the low-Y/low-H_T

[1]The errors in table are calculated propagating the Poisson uncertainties. Formally it is not completely correct, since the counts in the ABCD regions are not uncorrelated (their sum is linked to the total number of events). Nevertheless it can be demonstrated, in this case, that considering these values uncorrelated leads to a small over-estimate.

Pair	$r[b]$	$r[s+b]$		
		LM1	LM4	LM13
pfMET Vs. HT	-0.4262 ± 0.0198	-0.4212 ± 0.0199	-0.4250 ± 0.0199	-0.4210 ± 0.0199
pfMET Vs. DiLepMass	-0.9309 ± 0.0024	-0.9262 ± 0.00247	-0.9233 ± 0.00253	-0.9272 ± 0.0024
pfMET Vs. HemiMax	-0.0265 ± 0.0341	0.0397 ± 0.0360	-0.0212 ± 0.0342	0.0270 ± 0.0356
pfMET Vs. Y	—	—	—	—
pfMET Vs. mt_lmet	-0.9280 ± 0.0075	-0.9260 ± 0.0077	-0.9278 ± 0.0075	-0.9263 ± 0.0076
pfMET Vs. m_lj	0.4068 ± 0.0492	0.4961 ± 0.0519	0.4165 ± 0.0495	0.5015 ± 0.0520
HT Vs. DiLepMass	0.1274 ± 0.0517	0.2042 ± 0.0545	0.2525 ± 0.0563	0.1880 ± 0.0539
HT Vs. HemiMax	-0.8933 ± 0.0170	-0.8876 ± 0.0179	-0.8931 ± 0.0170	-0.8892 ± 0.0176
HT Vs. Y	2.9063 ± 0.1840	3.1016 ± 0.1918	2.9696 ± 0.1866	3.1026 ± 0.1919
HT Vs. mt_lmet	-0.5757 ± 0.036	-0.559 ± 0.0372	-0.5741 ± 0.0360	-0.5605 ± 0.0370
HT Vs. m_lj	-0.7899 ± 0.0235	-0.7759 ± 0.0249	-0.7891 ± 0.0236	-0.7773 ± 0.0249
DiLepMass Vs. HemiMax	0.0492 ± 0.0599	0.2095 ± 0.0665	0.2356 ± 0.0667	0.1720 ± 0.0650
DiLepMass Vs. Y	0.0122 ± 0.0599	0.1734 ± 0.0666	0.1613 ± 0.0651	0.1280 ± 0.0647
DiLepMass Vs. mt_lmet	-0.1691 ± 0.0562	0.0365 ± 0.0653	0.1274 ± 0.0682	-0.0158 ± 0.0631
DiLepMass Vs. m_lj	0.0501 ± 0.0642	0.2184 ± 0.0713	0.3576 ± 0.0761	0.2260 ± 0.0721
HemiMax Vs. Y	0.1960 ± 0.0814	0.4392 ± 0.0951	0.2266 ± 0.0830	0.3710 ± 0.0910
HemiMax Vs. mt_lmet	-0.3826 ± 0.0402	-0.2351 ± 0.0478	-0.3751 ± 0.0406	-0.2729 ± 0.0459
HemiMax Vs. m_lj	-0.8389 ± 0.0114	-0.7974 ± 0.0135	-0.8355 ± 0.0116	-0.8097 ± 0.01297
Y Vs. mt_lmet	-0.8751 ± 0.0104	-0.8397 ± 0.0123	-0.8631 ± 0.011	-0.8474 ± 0.0118
Y Vs. m_lj	-0.0717 ± 0.0648	0.1424 ± 0.0769	-0.0185 ± 0.0677	0.0673 ± 0.0726
mt_lmet Vs. m_lj	-0.5930 ± 0.0266	-0.4855 ± 0.0323	-0.5829 ± 0.0272	-0.4935 ± 0.0318

Table 5.2: Evaluated values of $r[b]$ (estimated goodness in background-only hypothesis) and $r[s+b]$ (estimate goodness in signal+background hypothesis. Error listed are statistical only).

zone (figure in appendix C), where the control regions are usually placed. This results into a dramatic cut-dependency of the estimate, as has been proved varying the boundaries [79]. Finally, it seems that $r[b]$ and $r[s+b]$ are closer for those pairs for which a strong under-estimation is found. This however does not mean that the estimates are really closer than those obtained for the other pairs, it is rather an effect of the asymmetry discussed above.

Pair	f_{unc} (%)
pfMET Vs. HemiMax	94.49
HT Vs. DiLepMass	94.23
DiLepMas Vs. HemiMax	80.74
DiLepMas Vs. Y	96.84
DiLepMas Vs. m_lj	81.74
DiLepMas Vs. mT_lMET	77.35
Y Vs. m_lj	89.02

Table 5.3: The seven selected pairs and their uncorrelation factors.

Seven pairs have been selected fulfilling these requirements:

- $|r[b]| < 0.2$;
- $r[s+b] - r[b] < 0.2$ for at least two of the three signal samples;
- uncorrelation of the variables (as defined in the previous section).

These pairs are listed together with their uncorrelation factor (table 5.3).

5.3.4 Separation of the signal

The last point to be checked is the possibility to discover or to exclude the presence of some signal. For this purpose, the ratio

$$r\star = \frac{N_C(\text{sig} + \text{bkg}) - N_C(\text{bkg})}{N_C(\text{bkg})} \tag{5.14}$$

is introduced. It is the relative difference between the total number of events (signal and backround) in the signal region and the number of background events in the same region. Obviously these quantities can be evaluated only for the simulations and not for data. Formula 5.14 is

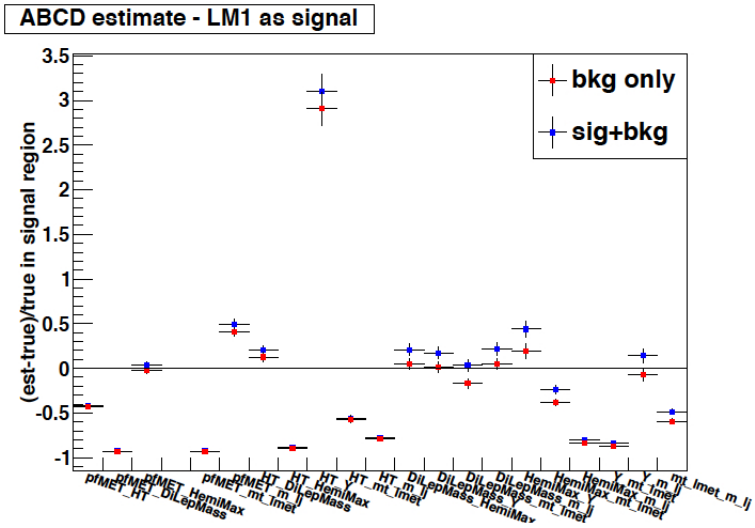


Figure 5.15: $r[b]$ (red markers) and $r[s+b]$ (blue markers) for every pair under study. SUSY signal at LM1.

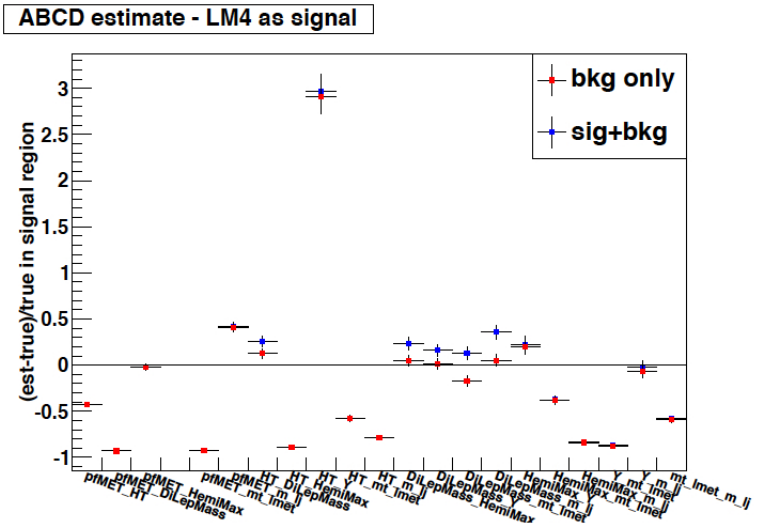


Figure 5.16: $r[b]$ (red markers) and $r[s+b]$ (blue markers) for every pair under study. SUSY signal at LM4.

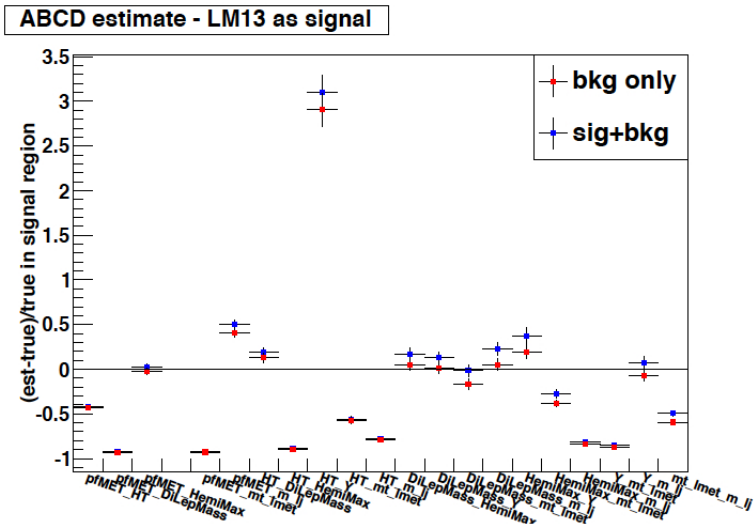


Figure 5.17: $r[b]$ (red markers) and $r[s+b]$ (blue markers) for every pair under study. SUSY signal at LM13.

similar to $r[b]$ and $r[s+b]$ introduced in the previous section; in this case, however, the more $r\star$ approaches to zero, the more the signal+background hypothesis is indistinguishable from the background-only one, hence we will not be able to exclude or claim the presence of signal. Therefore, unlike $r[b]$ and $r[s+b]$, we require $r\star$ to be as much greater than zero as possible. The values of $r\star$, compared to $r[b]$ and $r[s+b]$ for LM1, LM4, and LM13 are shown in figures 5.18 - 5.20 for the seven pairs till now selected.

Regarding the signal at LM4 (figure 5.19), it is clearly evident that for almost every pair, especially those involving the dilepton invariant mass, the signal is not able to produce an appreciable excess. This is due to the unfortunate shape of the m(ll) distribution at that benchmark point (see figure 5.3), with a peak centered at a value just above the Z mass. For LM1 and LM13, instead, we can identify five pairs for which $r\star$ has values sufficiently greater than zero.

These pairs are:

pfMET Vs. HemiMax

DiLepMas Vs. HemiMax

DiLepMas Vs. Y

DiLepMas Vs. m_lj

Y Vs. m_lj

Their 2-dimensional distributions are shown in figures 5.21 - 5.25.

A solution to obtain a better separation between background-only and signal+background hypotheses (i.e. to increase the value of $r\star$) should be to redefine the boundaries of the signal region, in order to exclude much more background. If however the number of background events in the signal region $N_C(bkg)$ become to small, the method is likely to become inherently unstable, since it would rely on a too low statistics. That is why we preferred to simply discard some pairs, rather than redefine their signal regions.

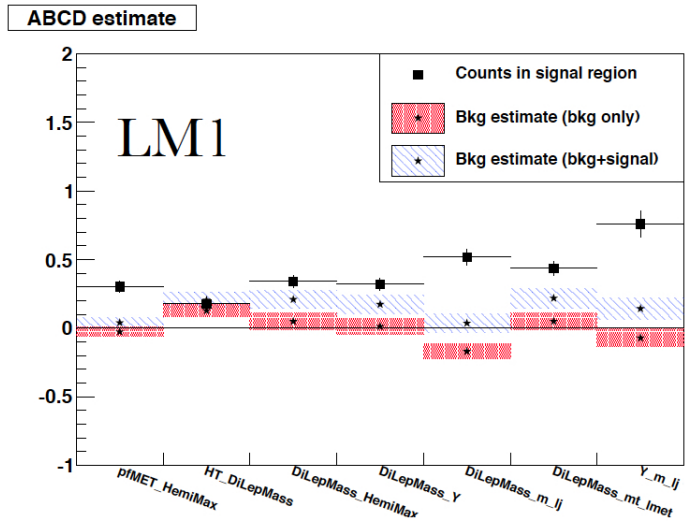


Figure 5.18: $r\star$ (simple markers) and $r[b]$ (red boxes) and $r[s+b]$ (blue boxes) for the seven selected pairs. SUSY signal at LM1.

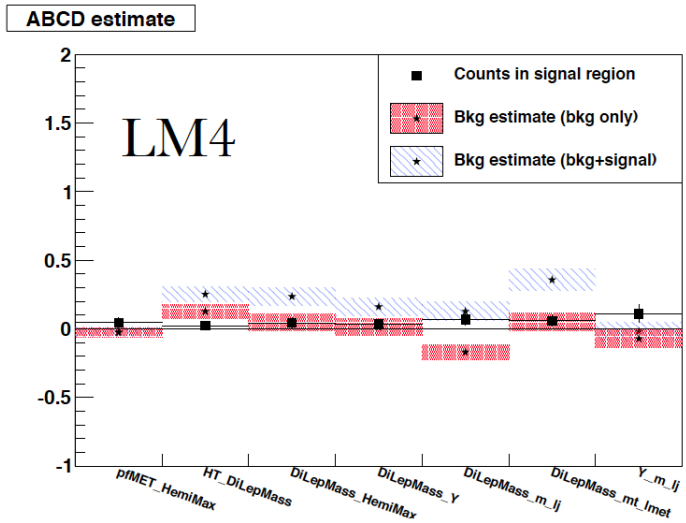


Figure 5.19: $r\star$ (simple markers) and $r[b]$ (red boxes) and $r[s+b]$ (blue boxes) for the seven selected pairs. SUSY signal at LM4.

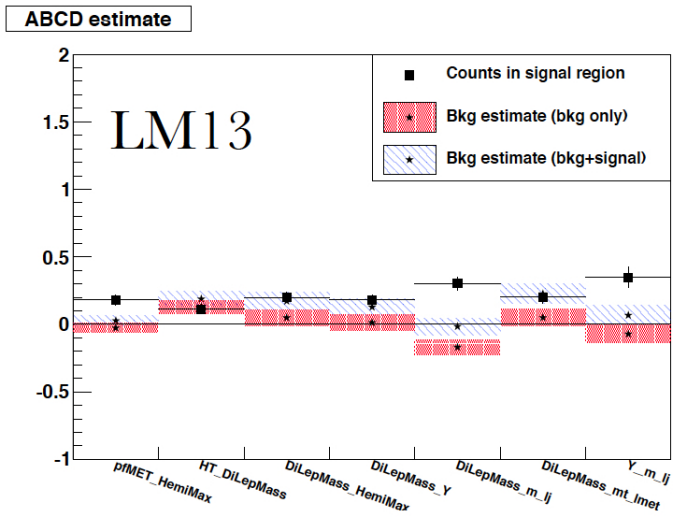


Figure 5.20: $r\star$ (simple markers) and $r[b]$ (red boxes) and $r[s+b]$ (blue boxes) for the seven selected pairs. SUSY signal at LM13.

5.3.5 Studies of robustness

The *quantitative estimates* (N_C^{est}) we obtain for the five selected pairs obviously depend on the choice of the signal and control regions that we made at the beginning of the ABCD method setup. The method itself, however, is based on the fact that the *goodness of the estimate* (in other words, the value of r) does not depend on the boundaries of the regions, if they are chosen in such a way that a possible signal does not contaminate the control regions in an appreciable way. This means that a well-working pair must give an estimate which is robust with respect to fairly small variations of the boundaries.

A robustness test has been conducted varying the boundaries between the background and the signal region of the variables involved in the five selected pairs, except the dilepton invariant mass (we preferred not to

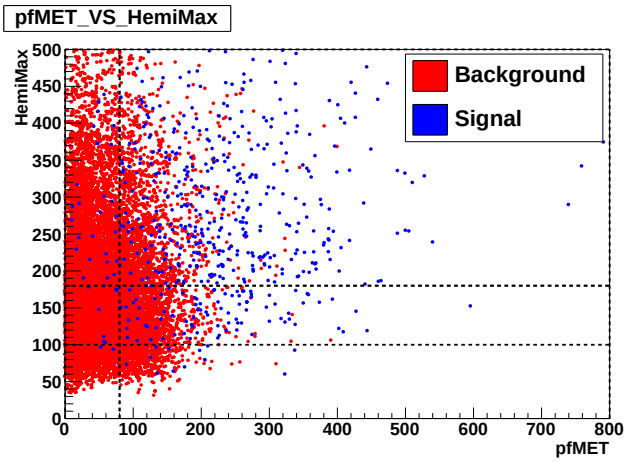


Figure 5.21: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by pfMET Vs. HemiMax. Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).

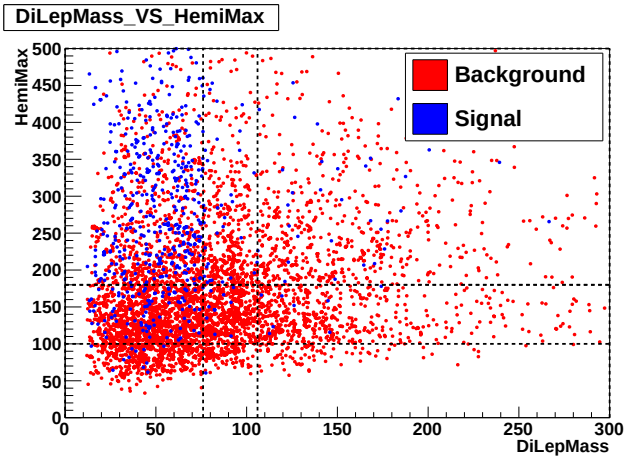


Figure 5.22: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by DiLepMass Vs. HemiMax. Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).

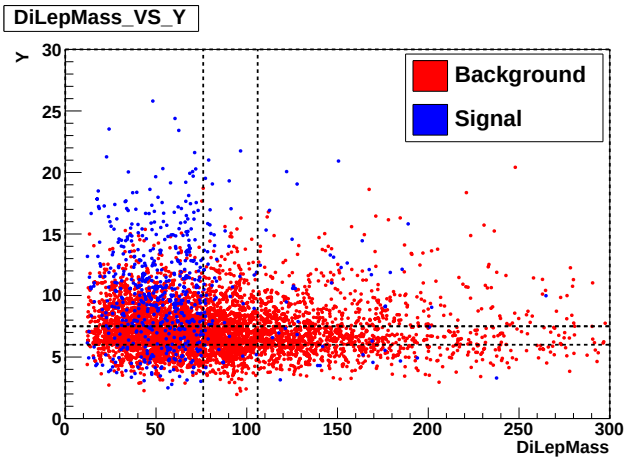


Figure 5.23: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by DiLepMass Vs. Y. Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).

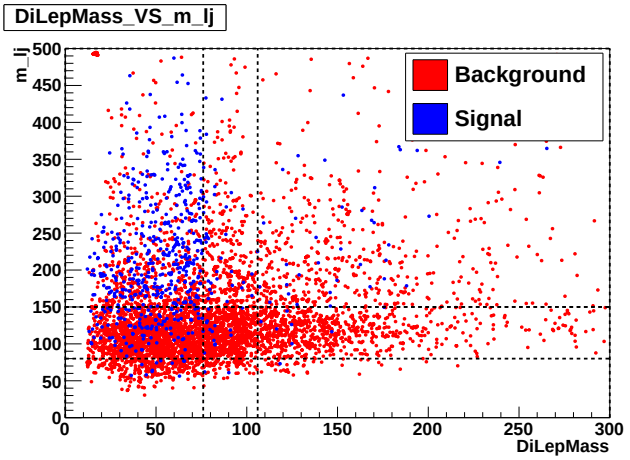


Figure 5.24: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by DiLepMass Vs. m_lj. Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).

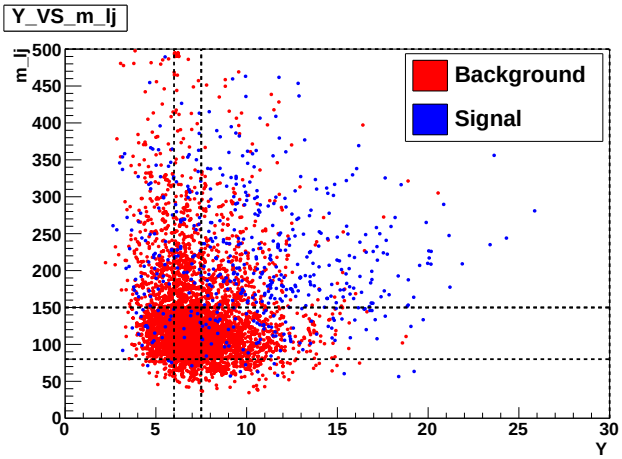


Figure 5.25: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by Y Vs. m_lj. Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).

modify the definition of the Z-window). Four variations with respect to a central value (the standard boundary) have been studied for each variable (table 5.4).

The values of $r[b]$ obtained for the five pairs are shown in figures 5.26 - 5.28. A total variation of $r[b] \leq \pm 0.15$ has been found for almost all the boundary shifts studied. The method has then proved to be quite robust with respect to boundary variations of $\sim 20\%$.

Looking at the $r[b]$ distributions, moreover, it appears that the central values are those for which there is the best estimate. It can be seen as an interesting proof of the goodness of the choice of control and signal regions performed simply looking at the 1-dimensional distributions.

Variable	Negative variations		central value	Positive variations	
MET (GeV)	60	70	80	90	110
H_T (GeV)	130	140	150	170	200
Y ($\sqrt{\text{GeV}}$)	6.5	7	7.5	8.5	9
m(lj) (GeV)	120	140	150	160	180
HemiMax (GeV)	150	170	180	190	210

Table 5.4: Variation of the position of the boundary between background and signal region for robustness test.

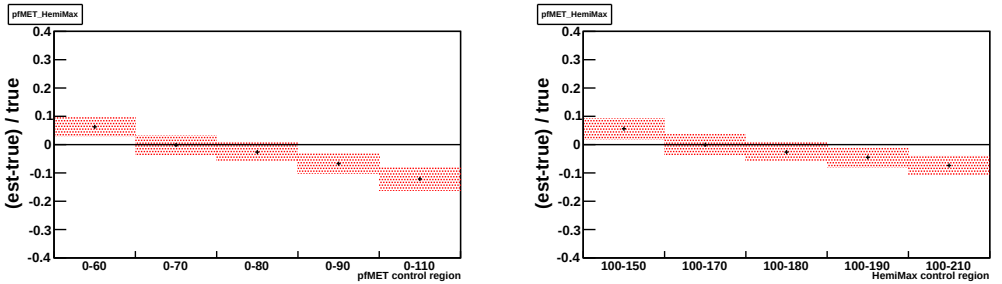


Figure 5.26: Variation of $r[b]$ of the pair pfMET Vs. HemiMax after modification of the boundary between background and signal region of MET (left) and HemiMax (right).

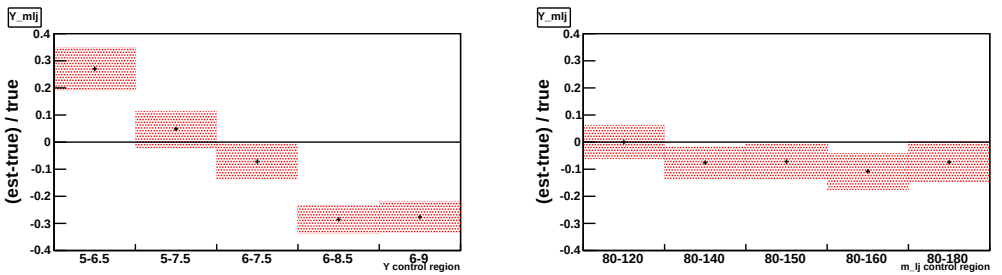
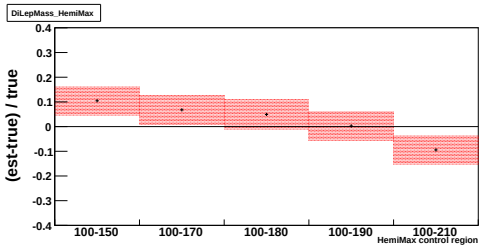
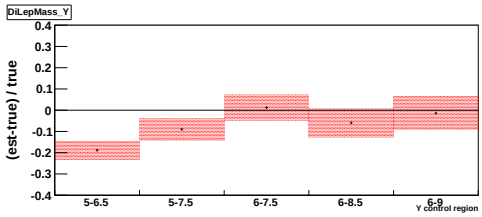


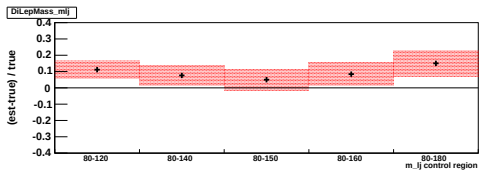
Figure 5.27: Variation of $r[b]$ of the pair Y Vs. m_lj after modification of the boundary between background and signal region of Y (left) and m(lj) (right).



(a)



(b)



(c)

Figure 5.28: Variation of $r[b]$ of the pair (a) DiLepMass Vs. HemiMax, (b) DiLepMass Vs. Y, (c) DiLepMass Vs. m_lj after modification of the boundary between background and signal regions.

6

Discussion of results

The analysis developed in the previous chapters is now applied to CMS data. An integrated luminosity of 5fb^{-1}, corresponding to the whole 2011 data-taking, has been analyzed using the five selected ABCD pairs.

First of all an estimate of the systematic uncertainties of the ABCD method is performed using two different approaches, giving consistent results.

The last part of the chapter is devoted to the definition of exclusion limits in the cMSSM frame, performed using a Bayesian calculation of confidence levels.

6.1 Systematics of the ABCD method

An estimate of the systematic uncertainty related to the ABCD method is extremely important to perform a correct and unbiased interpretation of the results. Possible systematic errors of the method could arise from the use of finite size Monte Carlo statistics, and from the choice of the boundaries of signal and control regions. This last point takes also into

account the uncertainties on energy scales and on pile-up, since these can affect the shapes of the distributions, and hence can change the number of events inside the signal region.

The systematic effects on our ABCD method have been studies in two independent ways:

1. using a *toy Monte Carlo method*, i.e. several small Monte Carlo samples are generated and the whole analysis is applied to each of them. We use the difference of the estimates obtained for the toy samples to get the systematic error.

2. *Varying the boundaries* of the signal region and studying how this affects the goodness of the estimate ($r[b]$). It is something similar to what has been done studying the robustness of the method (section 5.3.5).

The two estimates of the systematic errors have then been compared.

6.1.1 Toy Monte Carlo method

The first way we estimated the systematic uncertainty related to our data-driven method consists in generating some relatively small Monte Carlo background samples (or *toy samples*) with the same number of events (i.e. simulate the same integrated luminosity), to whom the method is applied, after the whole pre-selection and selection chain is performed.

Seven different toy samples containing each 130000 $t\bar{t}$ events (corresponding to $\sim 160pb^{-1}$ of integrated luminosity) have been generated. For a faster simulation we focused on $t\bar{t}$ only, in the $\mu\mu + e\mu + ee$ channel, since anyway it is the most important background for our analysis and almost the only one wich survives the selection cuts. The effects of the

Sample (All flavors)	Run on:	Pre-selected	Selected
TTJets_Toy_1	130000	3311	1680
TTJets_Toy_2	130000	3389	1767
TTJets_Toy_3	130000	3350	1710
TTJets_Toy_4	130000	3307	1720
TTJets_Toy_5	130000	3263	1710
TTJets_Toy_6	130000	3285	1686
TTJets_Toy_7	130000	3350	1722
Variation	0%	3.8%	5.1%

Table 6.1: Pre-selected and selected events in every toy MC sample.

pre-selection and selection cuts on each toy sample are shown in table 6.1: starting with the same number of events, the pre-selection generate a maximum difference of 3.8%, while after the selection cuts, the maximum difference in the number of events between the samples is $\sim 5\%$.

Then, the ABCD method has been applied to the toy MC samples. The estimates of the number of events in signal regions (N_C^{est}) are summarized in table 6.2 and in figure 6.1. Instead of taking the difference between the maximum and the minimum estimates for each pair as error, we preferred to use the standard deviation of the distributions of the estimates. As can be seen in figure 6.1, in fact, in some cases (for example toy sample 4 in MET Vs. HemiMax or toy sample 3 in DiLepMass Vs. HemiMax) the estimates obtained differ too much from the other, which would lead to an over-estimation of the error, if one simply considers the difference. The relative values of systematic uncertainties for each pair, obtained in this way, are listed in table 6.3. They range between 9.4% (MET Vs. HemiMax) and 22.5% (DiLepMass Vs. Y) with an average value of $\sim 15\%$, in agreement with values obtained in similar studies[70].

Pair	Toy_1	Toy_2	Toy_3	Toy_4	Toy_5	Toy_6	Toy_7
MET Vs. HemiMax	209	213.3	198.54	269.8	205.43	223.8	210.09
DiLepMass Vs. HemiMax	175.6	195.72	267.1	187.15	198.5	225.9	182.78
DiLepMass Vs. Y	228.8	212.65	246.3	186.2	322.77	279.82	150.8
DiLepMass Vs. m_lj	150.17	141.59	159.6	110.1	154.64	152.7	134.3
Y Vs. m_lj	60.1	80.7	72.1	53.94	94.17	75.13	61.7

Table 6.2: Estimate in the region C of the five selected pairs for every toy MC sample.

Pair	Mean	St.Dev.	**Rel. Error** (%)
MET Vs. HemiMax	234.1	22.1	**9.4**
DiLepMass Vs. HemiMax	221.7	29.5	**13.3**
DiLepMass Vs. Y	236.8	53.1	**22.5**
DiLepMass Vs. m_lj	134.8	15.6	**11.6**
Y Vs. m_lj	74.05	12.8	**17.3**

Table 6.3: Estimate of the systematic error for the five selected pairs obtained with toy MC technique.

These uncertainties can be used to calculate also the systematic error on the relative differences between estimated and true background events in regions C, $r[b]$. They are shown in table 6.4 for the total Monte Carlo background ($5fb^{-1}$) used for the setup of the ABCD method[1].

[1]For the studies of systematics, the Monte Carlo used are corrected for pile-up effects, but no trigger efficiency or luminosity factors are applied.

Figure 6.1: Estimate in the region C of the five selected pairs for every toy MC sample. Bars represent statistical errors.

Pair	N_C^{est}	N_C^{true}	$r[b]$	syst. on $r[b]$ (symm.)
MET Vs. HemiMax	1776 .06	1824.44	-0.026	**0.09**
DiLepMass Vs. HemiMax	1660.02	1582.12	0.0491	**0.14**
DiLepMass Vs. Y	1732.6	1711.6	0.012	**0.23**
DiLepMass Vs. m_lj	1382.05	1316.4	0.05	**0.13**
Y Vs. m_lj	520	560.2	-0.07	**0.16**

Table 6.4: Estimate of the systematic error of $r[b]$, evaluated using $5fb^{-1}$ of Monte Carlo background.

Variable	x_0^{bkg}	x_1^{bkg}	x_0^{sig}	x_1^{sig}
m(ll)	76	106	0	300
Enlarged	77	105		
Scaled	75	107		
MET	0	80	80	600
Enlarged		70	70	
Scaled		90	90	
HemiMax	100	180	180	500
Enlarged		165	165	
Scaled		195	195	
Y	6	7.5	7.5	30
Enlarged		7	7	
Scaled		8	8	
m(lj)	80	150	150	500
Enlarged		135	135	
Scaled		165	165	

Table 6.5: Boundaries of enlarged and scaled signal regions.

6.1.2 Variation of the boundaries

A different estimate of the systematic uncertainty on $r[b]$ has been performed investigating the effects of an enlargement or scaling of $\sim 10\%$ of the signal region. It can be done by varying the values of the boundary between the signal and the background region of both the variables in each pair (an example is shown in figure 6.2). The boundaries of enlarged and scaled signal regions, together with their original values, are listed in table 6.5.

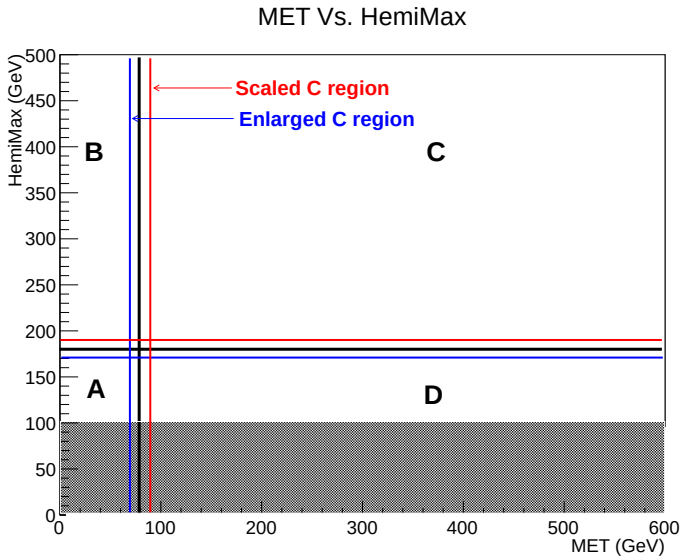


Figure 6.2: Example of variation of the boundaries for the pair MET Vs. HemiMax. The blue region C is the enlarged one, while the red lines limit the scaled region.

It could be argued that the efficiency of this method would be compromised by the arbitrariness of the choice of the enlarged and scaled boundaries. It is indeed a limit of this technique. Nevertheless, this would

not affect too much the evaluation of the uncertainty, since, as demonstrated in the previous chapter (section 5.3.3), the goodness of the ABCD estimate is robust with respect to small variations of the regions.

The values of $r[b]$ obtained for the enlarged and scaled signal regions are shown in table 6.6. The difference between the $r[b]$ evaluated in the enlarged and true region C ($|\Delta r[b]|$ (enlarged-true)) and that between the $r[b]$ evaluated in the scaled and true region C ($|\Delta r[b]|$ (scaled-true)) are listed in table 6.7. As a conservative estimate of the systematic uncertainty related to $r[b]$, a symmetric extension of the greater value for each pair is taken (last column of table 6.7). The errors obtained range between 0.07 and 0.15.

Pair	True boundaries			Enlarged signal region			Scaled signal region		
	N_C^{est}	N_C^{true}	$r[b]$	N_C^{est}	N_C^{true}	$r[b]$	N_C^{est}	N_C^{true}	$r[b]$
MET Vs. HemiMax	1776.06	1824.44	-0.026	2734.9	2618.9	0.044	1146.49	1253.5	-0.085
DiLepMass Vs. HemiMax	1660.02	1582.12	0.0491	2755	2294.8	0.201	980.97	1061.2	-0.075
DiLepMass Vs. Y	1732.6	1711.6	0.012	1877.7	2178	-0.14	1325.05	1377.8	-0.038
DiLepMass Vs. m_lj	1382.05	1316.14	0.05	2454.7	2138	0.15	879	846.7	0.038
Y Vs. m_lj	520	560.2	-0.07	853	893.6	-0.045	370.3	377.9	-0.02

Table 6.6: Estimate of the background in signal regions and calculation of $r[b]$ values for the five selected pairs using the true boundary definition, and the enlarged and scaled signal region.

Pair	$\|\Delta r[b]\|$ (enlarged-true)	$\|\Delta r[b]\|$ (scaled-true)	syst. on $r[b]$ (symm.)
MET Vs. HemiMax	0.071	0.059	**0.07**
DiLepMass Vs. HemiMax	0.151	0.125	**0.15**
DiLepMass Vs. Y	0.150	0.051	**0.15**
DiLepMass Vs. m_lj	0.098	0.012	**0.10**
Y Vs. m_lj	0.026	0.052	**0.05**

Table 6.7: Systematic uncertainty of $r[b]$ for every pair.

Pair	Syst. on $r[b]$ Toy	Bound	Relative difference (%)	**Definitive syst. on r[b]**
MET Vs. HemiMax	0.09	0.07	22	**0.09**
DiLepMass Vs. HemiMax	0.14	0.15	7	**0.15**
DiLepMass Vs. Y	0.23	0.15	34	**0.20**
DiLepMass Vs. m_lj	0.13	0.10	23	**0.13**
Y Vs. m_lj	0.16	0.05	68	**0.15**

Table 6.8: Systematic uncertainty of $r[b]$ for every pair using the Toy MC method and the variation of boundaries method. In the last column the final systematic uncertainty are listed.

These uncertainty can be compared with those obtained using the toy MC method (last column of table 6.4) for a cross check between the two techniques. The results of the comparison are summarized in table 6.8. As can be seen, a substantial agreement in the orders of magnitude is found (except for the last pair, the average relative difference between the errors evaluated with the two methods is $\sim 20\%$). However a perfect agreement for every pair was not expected, due to the effects of the arbitrariness for the second method and the problems related to the statistics of the toy model method (few toy samples, each simulating a relatively low integrated luminosity).

The definitive values of the systematic uncertainty that aree used for our data-driven method are listed in the last column of table 6.8. They have been chosen looking at the results of both the estimations, again favoring conservative values (i.e. taking the maximum between the two estimates, or a value close to it if the estimates differ too much), in order to ensure that the uncertainty is not under-estimated.

6.2 ABCD with $\sqrt{s}$=7 TeV data

We now apply the method studied in this work to the CMS data at $\sqrt{s} = 7$ TeV (see section 4.4 for a list of the datasets used).

As already explained, the method consists in counting the number N_A, N_B, and N_D of events in the control regions defined in the plane of two uncorrelated variables, and using them to infer the number N_C^{est} of residual background events in the signal region. This algorithm can be applied to more than one pair of discriminating variables. In particular we were able to find five of these pairs.

The estimates obtained are listed in table 6.9 (second column, N_C^{est}). In the third column of the same table, the actual numbers of data recorded in the signal regions (N_C^{data}) is shown.

Since we deal with data, we obviously do not have any information about the origin of the events in region C (if they are actually Standard Model background or if they are some possible signal of new physics) and variables like $r[b]$, $r[s+b]$ or $r\star$ can not be defined. Instead we study the relative difference of N_C^{data} and N_C^{est}:

$$r = \frac{N_C^{\text{data}} - N_C^{\text{est}}}{N_C^{\text{data}}}. \tag{6.1}$$

If $r \sim 0$ within the uncertainties, then the estimate is compatible with the background-only hypothesis. If, instead, r is significantly greater than zero, an excess in data, with respect to the estimated background, is found.

The values of r obtained for the five pairs are listed in the last column of table 6.9. The systematic uncertainty is estimated sccordingly to the prescription described in the previous section. In figure 6.3 a comparison between the values of r, the background only hypothesis ($r = 0$, black

horizontal line), and the values of $r[b]$ evaluated with the background Monte Carlo simulations (red boxes) is shown.

Pair	N_C^{est} ($\pm$ stat.)	N_C^{data} ($\pm$ stat.)	r ($\pm$ stat. $\pm$ syst.)
pfMET Vs. HemiMax	1614.38 $\pm$ 42.03	1743 $\pm$ 41.75	0.074 $\pm$ 0.033 $\pm$ 0.090
DiLepMass Vs. HemiMax	1491.68 $\pm$ 78.34	1526 $\pm$ 39.06	0.023 $\pm$ 0.057 $\pm$ 0.150
DiLepMass Vs. Y	1657.29 $\pm$ 90.11	1708 $\pm$ 41.33	0.030 $\pm$ 0.058 $\pm$ 0.200
DiLepMass Vs. m_lj	1715.94 $\pm$ 87.87	1311 $\pm$ 36.21	-0.309 $\pm$ 0.076 $\pm$ 0.130
Y Vs. m_lj	544.69 $\pm$ 30.72	563 $\pm$ 23.73	0.032 $\pm$ 0.068 $\pm$ 0.150

Table 6.9: ABCD estimate on $\sqrt{s} = 7$TeV data in $\mu\mu + e\mu + ee$ channel. The value of r with both statistical and systematic errors is also shown.

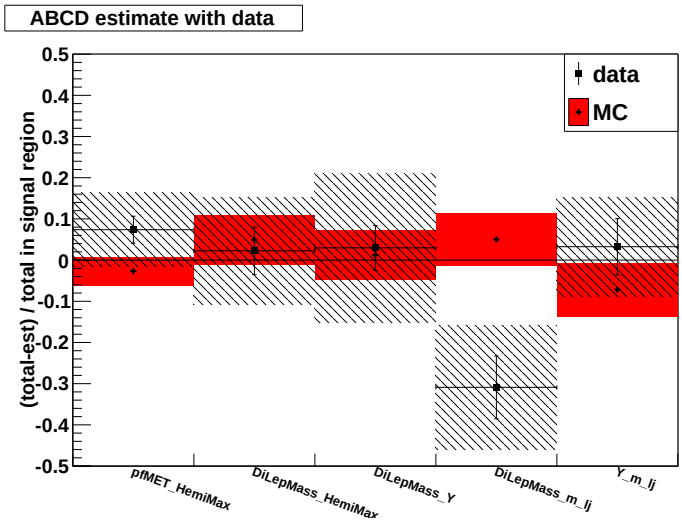


Figure 6.3: Evaluation of r from data (black dots) for each pair. The total error (a squared sum of statistical and systematic) is represented by the diagonal-lines filled boxes. The background only hypothesis ($r = 0$, black horizontal line) is also visible. The red boxes show the values of $r[b]$ from Monte Carlo simulations.

For four out of the five pairs the r-value results compatible with zero

within the errors. In one pair (DiLepMass Vs. m_lj) a discrepancy in data is found (r = -0.309 ± 0.076 [stat.] ± 0.130 [syst.]). This difference has a significance of 2σ (significance values for all the pairs are reported in table 6.10),and has to be investigated in deeper detail.

Pair	r (± total err.)	N_σ
pfMET Vs. HemiMax	0.074 ± 0.095	0.78
DiLepMass Vs. HemiMax	0.023 ± 0.160	0.13
DiLepMass Vs. Y	0.030 ± 0.210	0.14
DiLepMass Vs. m_lj	-0.309 ± 0.150	2.06
Y Vs. m_lj	0.032 ± 0.160	0.19
Average	-0.011 ± 0.065	0.17

Table 6.10: Significance (i.e. number of sigmas, N_σ) of r on $\sqrt{s}$ = 7TeV data in $\mu\mu + e\mu + ee$ channel. The error used is the squared sum of statistical and systematic uncertainties. The average value is shown in the last row.

While the actual numbers of events in the signal regions and their estimate (i.e. N_C^{data} and N_C^{est}) strongly depend on the choice of the pair and of the boundaries, their relative differences, r, are independent on these details. Therefore, the estimates for the five pairs can not be directly compared to each other, while an average value of r can be calculated and used as a coherent information from all the pairs. Using the numbers listed in table 6.10, an average value of

$$r = -0.011 \pm 0.065$$

is found. It indicates a difference with a significance of 0.17σ, hence totally indistinguishable from a background-only hypothesis (figure 6.4).

If we try to exclude the pair DiLepMass Vs. m_lj, which shows a not well explained behavior, an average value of

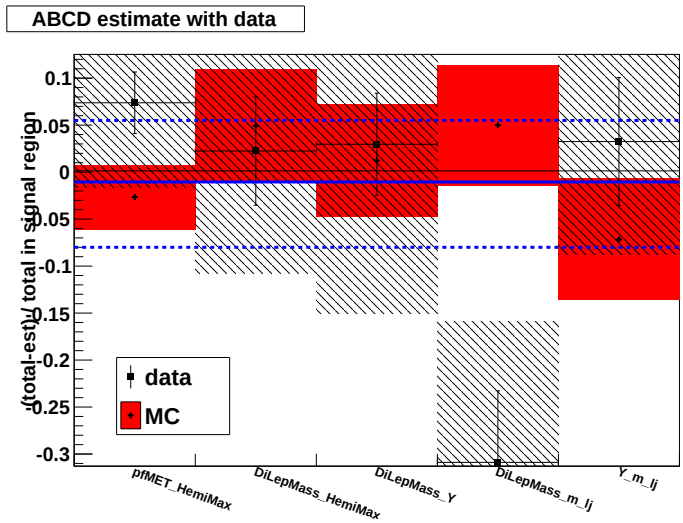


Figure 6.4: Zoom of figure 6.3 in which the average value of r with its error bars (blue lines) is shown. The consistency with the background only hypothesis is clearly evident.

$$r = 0.052 \pm 0.069$$

is obtained. This weak excess in data has a significance of 0.76σ, and again indicates a full consistency with the background only hypothesis.

6.3 Exclusion limits

No excess in data has been found using the analysis developed in this thesis. This is in agreement with what obtained from other parallel analyses in CMS. The absence of excess in the data can be interpreted in terms of exclusion limits in the cMSSM frame, using a Bayesian approach. The boundaries obtained can be compared to those from similar CMS searches.

6.3.1 Confidence level

The expected number of events in the signal region (μ), under a certain signal hypothesis, is the sum of the estimated number of background events in that region and the number of signal events:

$$\mu = b_{\text{est}} + s, \tag{6.2}$$

where s can be factorized as the product of the recorded integrated luminosity times the cross section times the selection efficiency:

$$s = \mathcal{L}\sigma\epsilon. \tag{6.3}$$

Since counting experiments follow Poisson's statistics, the probability of counting N_{obs} events, when μ are expected is given by:

$$P(N_{\text{obs}}|\mu) = \frac{\exp^{-\mu}\,\mu^{N_{\text{obs}}}}{N_{\text{obs}}!} \tag{6.4}$$

We are indeed interested in turning this probability, i.e. evaluating the probability that the number of observed events is compatible with a Poisson distribution centered in μ or, finally, with a particular signal cross section. It can be performed using the Bayes' formula:

$$P(\mu|N_{\text{obs}}) = \frac{P(N_{\text{obs}}B|\mu)P(\mu)}{P(N_{\text{obs}})}. \tag{6.5}$$

$P(\mu)$ and $P(N_{\text{obs}})$ are prior probabilities to be properly assigned (it is the main feature of Bayes approach). In our analysis we are interested in cross sections, hence we can manage the other terms in 6.3 as *nuisance parameters*. It is then useful to factorize the prior terms such that[80]:

$$P(\mu|N_{\text{obs}}) = P(\sigma; \mathcal{L}, \epsilon, b|N_{\text{obs}}) = P(N_{\text{obs}}|\mu)P(\sigma)P(\mathcal{L}, \epsilon, b|N_{\text{obs}}). \tag{6.6}$$

A flat prior for σ is used, corresponding to the hypothesis that the only a-priori knowledge about the cross section is that it is non-negative. Integrating over the nuisance parameters we get the probability of a certain signal cross section after having observed N_{obs} events:

$$P(\sigma|N_{\text{obs}}) = \int d\mathcal{L}\,d\epsilon\,db \frac{\exp^{-(\sigma\mathcal{L}\epsilon+b)}(\sigma\mathcal{L}\epsilon+b)^{N_{\text{obs}}}}{N_{\text{obs}}!} P(\mathcal{L},\epsilon,b), \tag{6.7}$$

where we used equations 6.3 and 6.4 for $P(N_{\text{obs}}|\mu)$. We can factorize the prior distribution for the nuisance parameters into:

$$P(\mathcal{L},\epsilon,b) = \pi(\mathcal{L})\pi(\epsilon)\pi(b). \tag{6.8}$$

A Gaussian distribution with expectation value equal to the measured luminosity is used for $\pi(\mathcal{L})$, while a Poisson function centered in b_{est} is used for $\pi(b)$. To describe the signal efficiency prior $\pi(\epsilon)$ a Log-normal distribution is often chosen, since it well approximates the product of independent random variables. The probability 6.7 can be translated into an upper bound on the signal cross section integrating over σ up to a confidence level of 95%:

$$0.95 = \int_0^{\sigma_{95}} P(\sigma|N_{\text{obs}})d\sigma. \tag{6.9}$$

Equation 6.9 has to be solved for σ_{95} in every point of the cMSSM or the simplified model plane. A point is excluded with $CL \geq 95\%$ if

$$\sigma_{point} \geq \sigma_{95}. \tag{6.10}$$

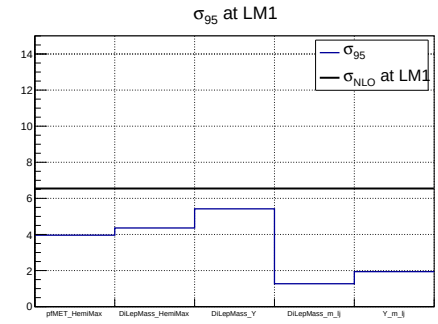
The evaluation of σ_{95} can be performed at the cMSSM benchmark points used to setup our analysis (LM1, LM4, and LM13). In table 6.11 the meaningful parameters are summarized for each of the ABCD pairs (for the confidence levels' calculation, the estimate with any pair has to be considered as a standalone analysis).

Pair	$\mathcal{L}$	$\Delta\mathcal{L}$	ϵ	$\Delta\epsilon$	b	Δb	N_{obs}
LM1							
pfMET Vs. HemiMax	4980	225	0.017	0.002	1614	113	1743
DiLepMass Vs. HemiMax	4980	225	0.017	0.002	1491	198	1526
DiLepMass Vs. Y	4980	225	0.017	0.002	1657	248	1708
DiLepMass Vs. m_lj	4980	225	0.018	0.002	1715	170	1311
Y Vs. m_lj	4980	225	0.013	0.002	544	54	563
LM4							
pfMET Vs. HemiMax	4980	225	0.007	0.0007	1614	113	1743
DiLepMass Vs. HemiMax	4980	225	0.006	0.0006	1491	198	1526
DiLepMass Vs. Y	4980	225	0.005	0.0005	1657	248	1708
DiLepMass Vs. m_lj	4980	225	0.006	0.0006	1715	170	1311
Y Vs. m_lj	4980	225	0.006	0.0006	544	54	563
LM13							
pfMET Vs. HemiMax	4980	225	0.003	0.0003	1614	113	1743
DiLepMass Vs. HemiMax	4980	225	0.003	0.0003	1491	198	1526
DiLepMass Vs. Y	4980	225	0.003	0.0003	1657	248	1708
DiLepMass Vs. m_lj	4980	225	0.003	0.0003	1715	170	1311
Y Vs. m_lj	4980	225	0.003	0.0003	544	54	563

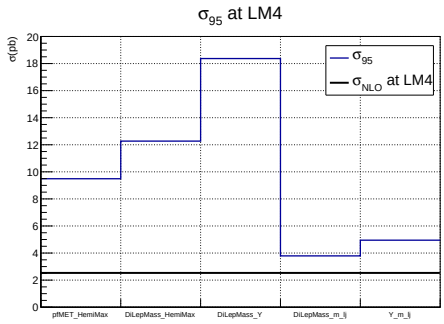
Table 6.11: Input parameters for Bayesian confidence level calculation.

The acceptance times efficiency ϵ is the ratio between the number of signal events in every C region and the number of simulated signal events. A conservative value of 10% has been used for $\Delta\epsilon$, obtained from the systematic error in MC simulations (see section 4.8). The computation of the integral 6.9 has been performed using a dedicated ROOT toolkit, called *RooStats_cl95* [81]. The resulting limit cross sections for LM1, LM4, and LM13 and for the five ABCD pairs are shown in figure 6.5.

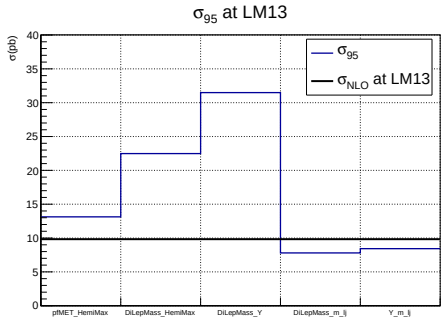
As can be clearly seen, our analysis is able to exclude LM1 at confidence level greater than 95% using all the pairs. With Y Vs. m_lj we are also able to exclude LM13. Nothing can be said for LM4: with the analyzed integrated luminosity, signal and background-only hypotheses are statistically indistinguishable, as had already been suggested in section 5.3.4. From this point of view, the pair DiLepMass Vs. m_lj seems to give the most promising results. It has however to be noticed that this calculation is deeply biased by the unexplained overestimate in Monte Carlo for this pair, hence its interpretation requires an extraordinary caution.



(a)



(b)



(c)

Figure 6.5: Values of σ_{95} for the five ABCD pairs at (a) LM1, (b) LM4, and (c) LM13. The SUSY cross sections at each point are indicated by a black line.

6.3.2 Scan in cMSSM plane

To set exclusion limits in the cMSSM space, the calculation described above has to be performed for every point of a properly chosen lattice covering the whole $m_0 - m_{1/2}$ plane. In every point an evaluation of:

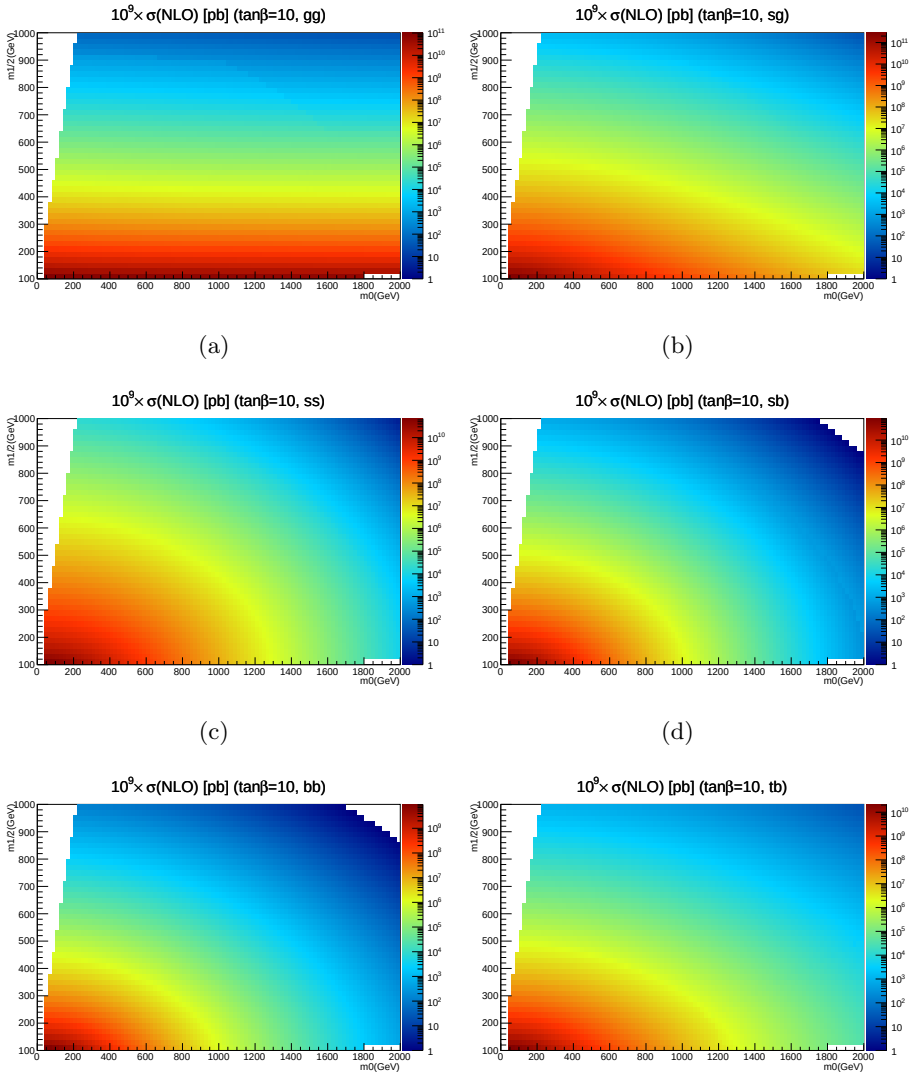
- cross section at Next to Leading Order;
- signal acceptance times efficiency (ϵ) with its uncertainty,

is needed. As suggested by the CMS guidelines for 2011 analyses[82] the remaining cMSSM parameters have been set to the following values:

- $\tan\beta = 10$;
- $A_0 = 0$;
- $\text{sign}\mu = +$.

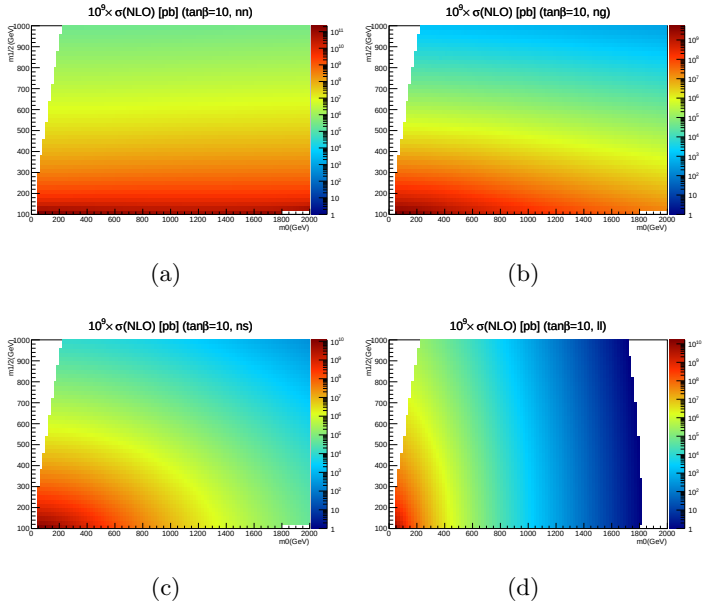
The mass spectra for every lattice point have been generated with *SoftSUSY* package and given as input to PYTHIA for cross section calculation at leading order. Finally the k-factors have been evaluated with PROSPINO. Figures 6.6 and 6.7 show the cross sections at NLO for any possible SUSY final state, while the total cross sections, obtained summing all the channels, are shown in figure 6.8. The lattice points are placed in steps of 20GeV in both m_0 and $m_{1/2}$.

The computation of the acceptance times efficiency (ϵ) for each pair is an extremely long task, since it requires the simulation of $\sim$20000 signal events in every point of the cMSSM plane and to perform the whole analysis on them. In order to allow the whole process to be completed in a reasonable time, a fast simulation rather than a full simulation has been made.



(a) (b) (c) (d) (e) (f)

Figure 6.6: SUSY cross section related to the production of gluino or squark states (including 3^{rd} generation ones).



(a) (b)

(c) (d)

Figure 6.7: SUSY cross section related to the production of gaugino or slepton states.

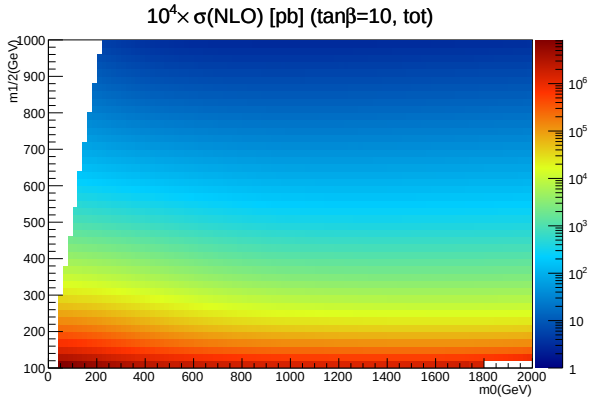


Figure 6.8: Total cross section of SUSY production.

Moreover a wider lattice spacing has been used:

- $0 < m_0 < 3000$GeV, divided in 50 bins (step size 60GeV);
- $0 < m_{1/2} < 1000$GeV, divided in 20 bins (step size 50GeV).

Hence a total number of 1000 cMSSM point have been simulated. A dilepton filter has been applied to the signal sample, in order to provide an higher statistics and to prevent the occurrence of too few selected events per point. The filter efficiency (around $10\% - 20\%$ in the cMSSM plane) has to be taken into account in the evaluation of ϵ.

The values of ϵ obtained for the five pairs are shown in figures 6.9-6.13.

Finally the comparisons of cross sections (equation 6.10 at every lattice point permits to draw the exclusion limits for the five ABCD pairs. These limits are shown in figure 6.14. The plot shows also the regions excluded by previous searches at LEP and TeVatron as well as regions theoretically excluded.

The position of LM1 and LM4 are indicated by black dots. Point LM13 is not shown since it does not belong to the $\tan\beta = 10$ plane (it has $\tan\beta = 40$). As already argued, LM1 is clearly excluded by our analysis, using every ABCD pair. It would appear the same for LM4 too, despite we showed it is not. It is indeed due to the low resolution of the scan lattice: LM4 is actually between the first not totally excluded points.

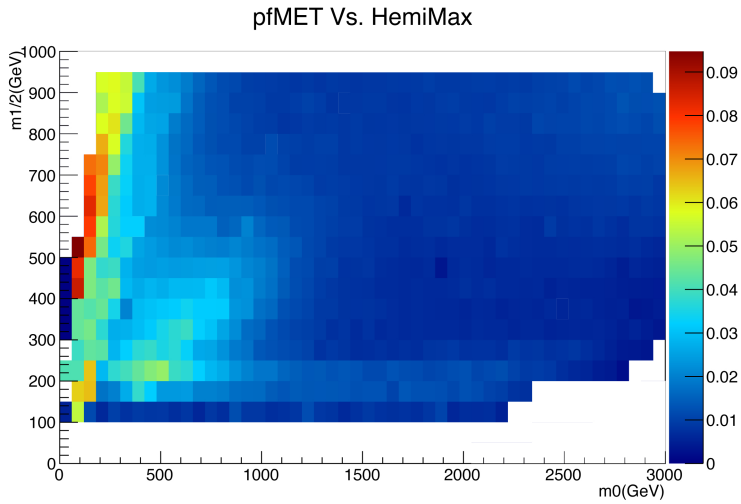


Figure 6.9: Acceptance times efficiency for pfMET Vs. HemiMax in the cMSSM plane.

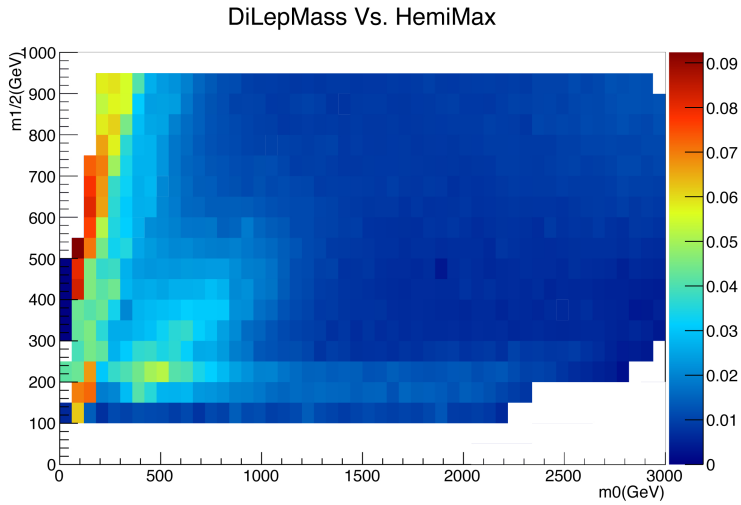


Figure 6.10: Acceptance times efficiency for m(ll) Vs. HemiMax in the cMSSM plane.

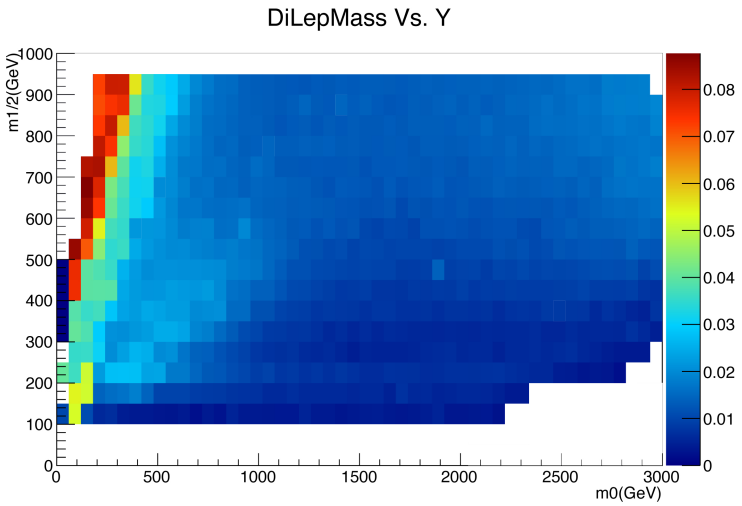


Figure 6.11: Acceptance times efficiency for m(ll) Vs. Y in the cMSSM plane.

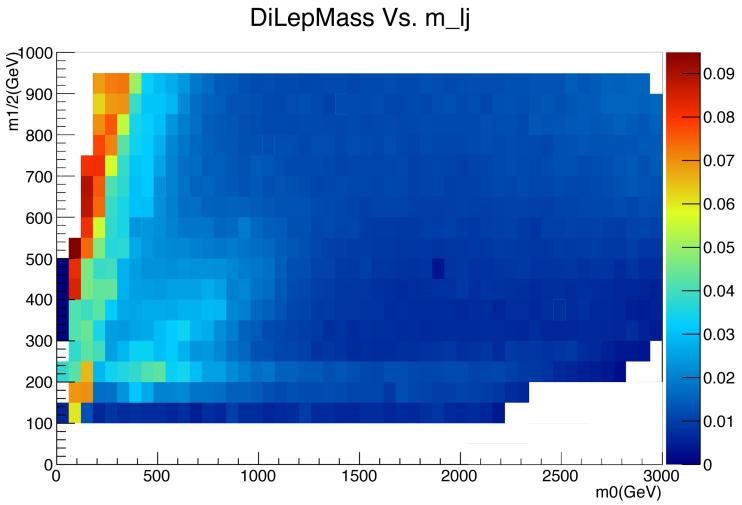


Figure 6.12: Acceptance times efficiency for m(ll) Vs. m(lj) in the cMSSM plane.

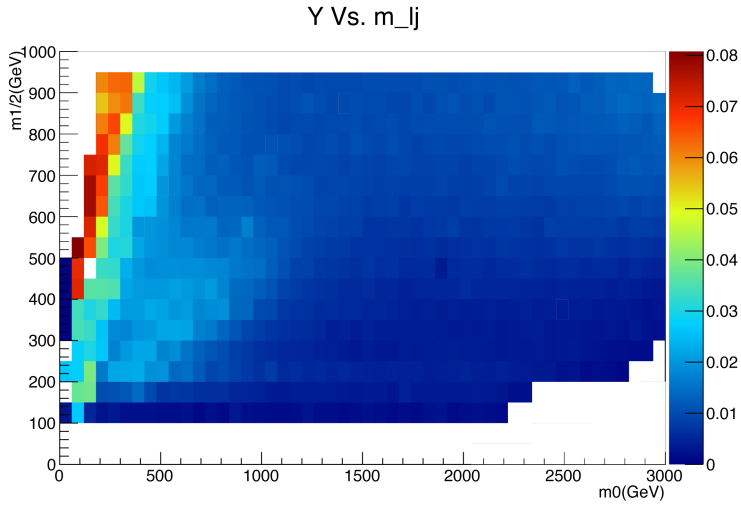


Figure 6.13: Acceptance times efficiency for Y Vs. m(lj) in the cMSSM plane.

As can be seen from figure 6.15 the exclusion boundary of the analysis here presented is virtually superimposable to that of the whole CMS OS dilepton searches at $\mathcal{L} = 1fb^{-1}$ (early 2011 data taking)[82]. These searches anyway include both Z-vetoed and not Z-vetoed dilepton analyses, which have a stronger exclusion power. Moreover we have to stress that in general, dilepton searches are obviously less powerful than hadronic analyses in setting exclusion boundaries. They are indeed more pure and hence more suitable for systematic measurements of the properties of SUSY particles, once they are discovered. Even if, up to now, any significant excess has been found, many studies involving the searches for mass edges in dilepton spectra (see section 3.5.2) are already ongoing.

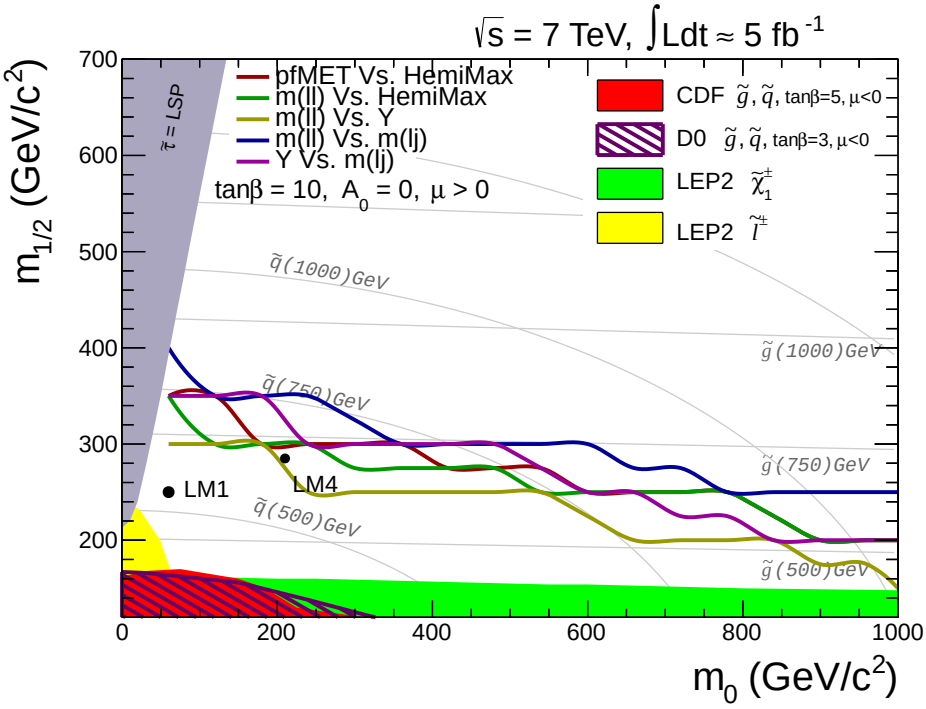


Figure 6.14: Exclusion limits at 95% CL, using the five ABCD pairs studied. Experimentally and theoretically excluded regions are draw. The black dots represent LM1 and LM4 benchmark point (LM13 does not belong to $\tan\beta = 10$ plane).

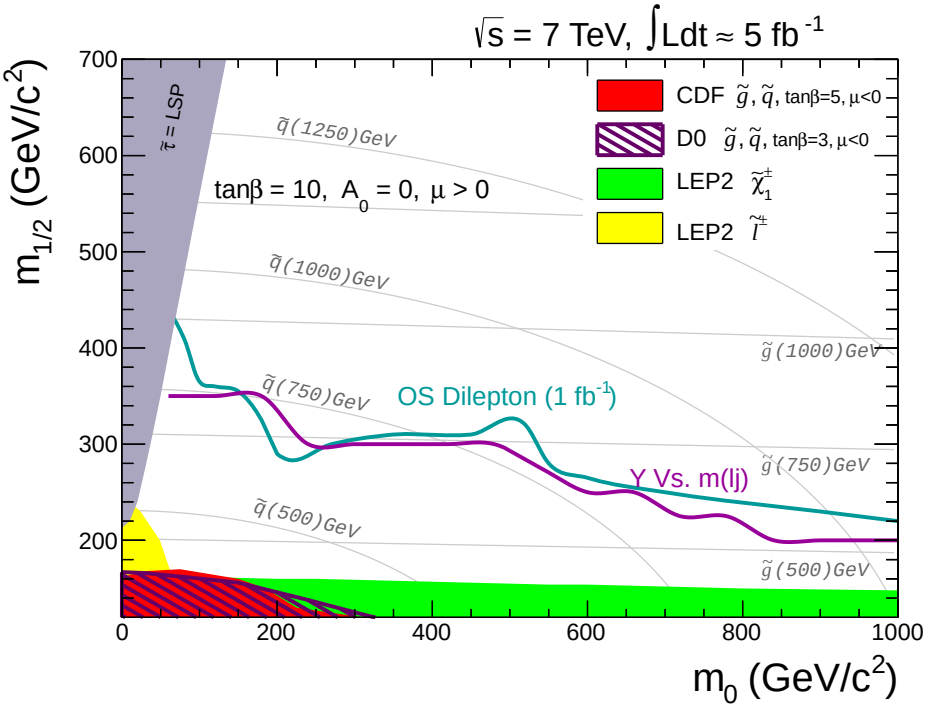


Figure 6.15: Comparison between the exclusion power of our analysis (in this plot Y Vs. m(lj) limit is shown) and the boundary from all the CMS OS dilepton searches at 1fb^{-1}.

Summary

A search for physics beyond the standard model, within a supersymmetric scenario, have been developed in this thesis. The analysis has been performed on $\sim 5fb^{-1}$ of CMS data, corresponding to the whole 2011 data taking at $\sqrt{s} = 7$ TeV. The analysis strategy is based on the search of an excess in events with two leptons of opposite sign, large hadron activity and high missing transverse energy, with respect to the Standard Model predictions.

The cuts are imposed to reject most of the Standard Model background, mainly arising from QCD processes, Drell Yan and top production. While the requirement of two leptons rules out the QCD events, the DY is dramatically reduced imposing an additional veto in the dilepton invariant mass (Z-vetoed analysis). After the analysis cuts, then, the residual background is that arising from $t\bar{t}$ production.

An estimate of the residual SM background has been performed in a data-driven way, using a variant of the *ABCD technique*. The strength of the method developed in this work is the usage of many discriminating variables, and hence of more than one ABCD pair. The selection of the best-working pairs has been driven by Monte Carlo studies. The MC samples have been deeply studied and validated before being used in the

setup of the data-driven method. Finally five pairs have been chosen to be used with CMS data. They show:

- good background estimate power;
- robustness with respect to slight variations of the ABCD regions and of statistics;
- good discriminating power between signal and background hypotheses, which is necessary for discovery or to set exclusion limits.

The last condition has been verified using simulated SUSY signal at three different mSUGRA benchmark points: LM1, LM4, and LM13.

Moving to data, no significant excess has been found using any of the pairs. This result is in agreement with other parallel analysis developed by the CMS collaboration. In particular, the values of the relative differences between data and background estimates, obtained with all the five pairs are consistent with a background-only hypothesis within the uncertainties.

Since no excess has been found, exclusion limits at 95% of confidence level have been set, using a Bayesian approach. The CLs calculations for signal at LM1, LM4, and LM13 have been made. Our analysis excludes the benchmark point LM1, while nothing can be said for LM4. It is indeed a not surprising results, since points like LM4 are not much sensitive to Z-vetoed analyses. A scan in cMSSM plan (with $\tan\beta = 10$) has then been performed and exclusion limits have been set.

The analysis developed in this thesis has shown to work properly and to provide interesting results. However several improvements and additional studies are under investigation:

- the information from the discriminating variables studied in this work can be used as input for a neural network analysis. The NN output itself could be treated as a pair variable for an ABCD estimate;
- the exclusion limits need to be interpreted within the simplified models' frame. An analysis based on several discriminating variables, like that here developed, could be particularly efficient, for example, in scanning regions of the SMS plane characterized by compressed mass spectra. These regions have still not been deeply explored;
- the analysis will be soon moved to 2012 data at $\sqrt{s} = 8$ TeV. This migration is not trivial since it involves new trigger paths, different pile-up reweightings and a new definition of physics objects.

Appendix A

Analysis software

A schematic review of the softweare used for the analysis presented in this thesis is shown in this appendix. In the first section a scheme of the action and the interplay of the different software modules is presented, while in the second the content of local analysis ROOT NTuples is listed.

Software modules

Figure A.1 presents a diagram of the interplay of the software packages used in our work. The first column contains the input or output files: they are mainly ROOT NTuples (.root files), the second one shows the tasks to be performed, while in the last the names of the executables used to accomplish them are listed. Our analysis starts from the information included in the data and Monte Carlo NTuples produced by the ETH SUSY group. These files are copied in local areas in Catania and used as input for our preselection code, a CMSSW module called OSDLAnalyzer. This software module produces local NTuples (CT NTuples), whose content is shown in the next section. The selection and the data-driven analy-

sis are performed on these NTuples, using some ROOT macros included in a package called OSDmacro. The final output include 1-dimensional and 2-dimensional histograms, correlation matrices as well as log files in which the results of data/MC comparison or data-driven estimates are summarized.

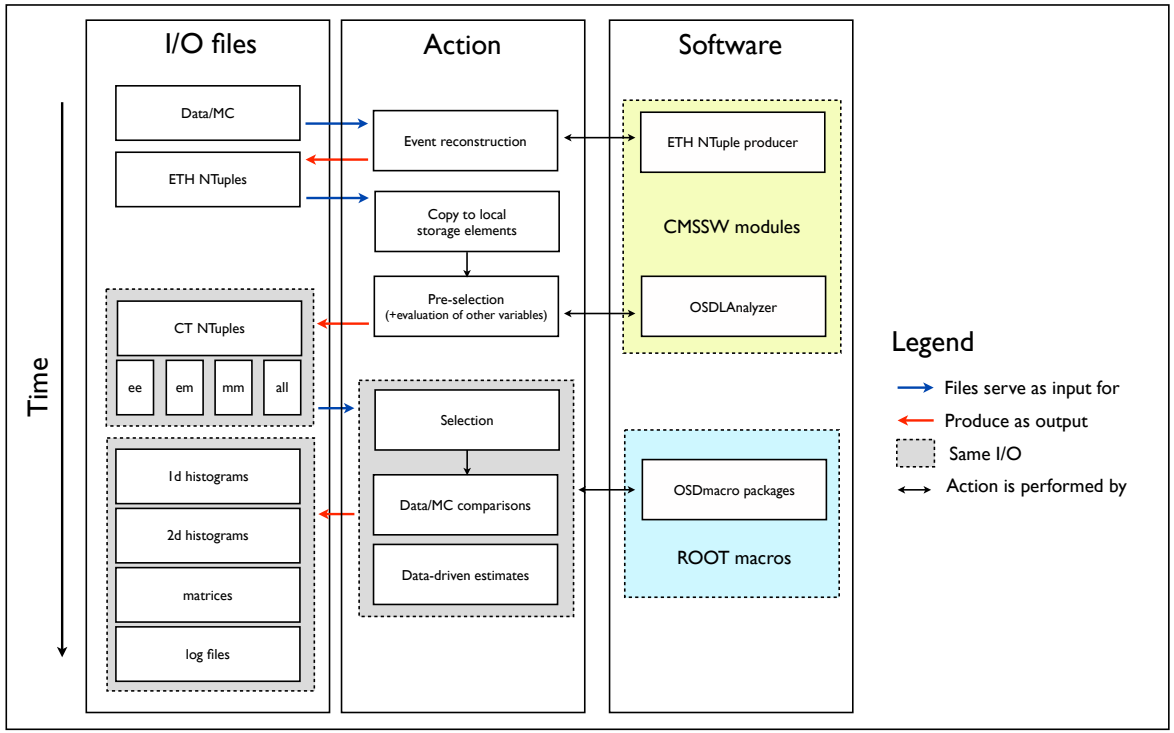


Figure A.1: Scheme of the software used in this analysis.

CT NTuples

In the following pages the content of the ROOT NTuples stored in the local computing farm (gr1farm1.ct.infn.it) is shown. CT NTuples ROOT files are made of a single TTree[83], with many *branches*, each containing variables and information of interest for our analysis. Since CT NTuples contain only a subset of the variables included in ETH NTuples (plus some other variable, like Hemisphere information, and ABCD variables), and is filled only with pre-selected events, they are sensibly light files, which is an undeniable advantage.

Variable name	C++ Type	Description
Event Infos		
Run	int	Run ID number
Event	int	Event ID number
LumiSec	int	Luminosity section number
ExtXSecLO	float	External cross section
IntXSec	float	Internal cross section
PUnumInteractions	int	Number of pile-ups
NVrtx	int	Number of primary vertices
HLT Infos		
HLT_DoubleMu7_v1	bool	Is this HLT bit
HLT_DoubleMu7_v2	bool	fired?
HLT_DoubleMu7_v3	bool	(double lepton bits only)
HLT_DoubleMu3_HT150_v1	bool	
HLT_DoubleMu3_HT150_v2	bool	
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_v1	bool	
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_v2	bool	
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_v3	bool	
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_v4	bool	
HLT_DoubleEle8_CaloIdL_TrkIdVL_HT150_v1	bool	
HLT_DoubleEle8_CaloIdL_TrkIdVL_HT150_v2	bool	
HLT_Mu17_Ele8_CaloIdL_v1	bool	
HLT_Mu17_Ele8_CaloIdL_v2	bool	
HLT_Mu17_Ele8_CaloIdL_v3	bool	

Variable name	**C++ Type**	**Description**
HLT_Mu17_Ele8_CaloIdL_v4	bool	
HLT_Mu8_Ele17_CaloIdL_v1	bool	
HLT_Mu8_Ele17_CaloIdL_v2	bool	
HLT_Mu8_Ele17_CaloIdL_v3	bool	
HLT_Mu8_Ele17_CaloIdL_v4	bool	
HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT150_v1	bool	
HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT150_v2	bool	
HTL_GoodElEvent	bool	Electron event flags
HTL_GoodElFakesEvent	bool	
HTL_GoodElEvent_RA5	bool	
HTL_GoodElEvent_TDL	bool	
HTL_GoodMuEvent	bool	Muon event flag
HTL_GoodHadronicEvent	bool	Jet event flag
Muon Infos		
NMus	int	Number of good muons
MuPt	std::vector<float>	p_T of selected muons
MuEta	std::vector<float>	muon η
MuPhi	std::vector<float>	muon ϕ
MuCharge	std::vector<int>	muon charge
MuTight	std::vector<bool>	Tight muon ID flag
MuIso	std::vector<float>	Relative isolation
MucIso	std::vector<float>	Corrected relative iso
MuDRJet	std::vector<float>	ΔR w.r.t. the jets

Variable name	**C++ Type**	**Description**
MuDRHardestJet	std::vector<float>	ΔR w.r.t. the leading jet
MuD0	std::vector<float>	muon impact parameter
MuDz	std::vector<float>	z distance from primary vertex
MuD0BS	std::vector<float>	beam spot impact parameter
MuDzBS	std::vector<float>	z distance from beam spot
MuPtE	std::vector<float>	transverse momentum error
MuGenID	std::vector<int>	Generator infos
MuGenMoID	std::vector<int>	(MC only)
MuGenGMoID	std::vector<int>	
MuGenType	std::vector<int>	
MuGenMoType	std::vector<int>	
MuGenGMoType	std::vector<int>	
MuMT	float	di-muon transverse mass
MuMinv	float	di-muon invariant mass
Electron Infos		
NEls	int	Number of good electrons
ElCh	std::vector<int>	electron charge
ElChIsCons	std::vector<int>	consistency between charge meausrements
ElPt	std::vector<float>	p_T of selected electrons
ElEta	std::vector<float>	electron η
ElPhi	std::vector<float>	electron ϕ
ElD0	std::vector<float>	electron impact parameter

Variable name	C++ Type	Description
ElD0Err	std::vector<float>	error on impact parameter
ElEoverP	std::vector<float>	eleID variables
ElHoverE	std::vector<float>	
ElSigmaIetaIeta	std::vector<float>	
ElDeltaPhiSuperClusterAtVtx	std::vector<float>	
ElDeltaEtaSuperClusterAtVtx	std::vector<float>	
ElDR04TkSumPt	std::vector<float>	Isolations
ElDR04EcalRecHitSumEt	std::vector<float>	
ElDR04HcalTowerSumEt	std::vector<float>	
ElConvPartnerTrkDist	std::vector<float>	
ElConvPartnerTrkDCot	std::vector<float>	
ElChargeMisIDProb	std::vector<bool>	mis-identification probability
ElDRjet	std::vector<float>	ΔR w.r.t. the jets
ElDRhardestjet	std::vector<float>	ΔR w.r.t. the leading jet
ElGenID	std::vector<int>	Generator infos
ElGenMID	std::vector<int>	(MC only)
ElGenGMID	std::vector<int>	
ElGenType	std::vector<int>	
ElGenMType	std::vector<int>	
ElGenGMType	std::vector<int>	
ElRelIso	std::vector<float>	relative isolation
ElcRelIso	std::vector<float>	corrected relative isolation
ElMInv	float	di-electron invariant mass

Variable name	**C++ Type**	**Description**
ElMTInv	float	di-electron transverse mass
HT/MET		
HT	float	H_T
cHT	float	corrected H_T
SumEt	float	$\sum E_T$
cSumEt	float	corrected $\sum E_T$
tcMET	float	Track corrected MET
pfMET	float	Particle Flow MET
MuCorrMET	float	muon corrected MET
Jet Infos		
NJets	int	Number of good jets
JetPt	std::vector<float>	p_T of selected jets
cJetPt	std::vector<float>	corrected p_T of selected jets
JetEta	std::vector<float>	jets η
JetPhi	std::vector<float>	jets ϕ
dPhiMJ1	float	
dPhiMJ2	float	
R12	float	
R21	float	
R12plusR21	float	
SumPtJets	float	Sum of jet p_Ts
JBtag	std::vector<float>	b-tagging probability
NBJets	int	number of b-tagged jets

Variable name	**C++ Type**	**Description**
Event Shape, Hemispheres and global variables		
AlphaT_h	float	
AlphaCT_h	float	
AlphaT	float	
AlphaCT	float	
AlphaT_new	float	
AlphaCT_new	float	
GTThrust	float	
GThrustM	float	
Axis1	std::vector<float>	4-direction of the first hemisphere's axis
Axis2	std::vector<float>	4-direction of the second hemisphere's axis
Mass1	float	Invariant mass of the first hemisphere
Mass2	float	Invariant mass of the second hemisphere
Phi1	float	ϕ of the first hemisphere's axis
Phi2	float	ϕ of the second hemisphere's axis
Eta1	float	η of the first hemisphere's axis
Eta2	float	η of the second hemisphere's axis
HemidR	float	ΔR between the hemispheres
HemiAco	float	Hemisphere's acoplanarity

Variable name	**C++ Type**	**Description**
		$(= \pi - \Delta\phi)$
HemiMax	float	HemiMax
N_Hemi1Obj	int	Number of constituents of the first hemisphere (N_{o1})
N_Hemi2Obj	int	Number of constituents of the second hemisphere (N_{o2})
HemiAsimmetry	float	$(N_{o2} - N_{o1})/(N_{o2} + N_{o1})$
HemiMassAsimmetry	float	(Mass2-Mass1)/(Mass2+Mass1)
HT2	float	$p_T(l_1) + p_T(l_2) + p_T(j_1) + p_T(j_2)$
m_lj	float	m(lj)
mt_lmet	float	m_T(lMET)
Y	float	Y
Dilepton variables		
lDPhi	float	$\Delta\phi$ between the two leptons
lDR	float	ΔR between the two leptons
DiLepMass	float	m(ll)
IsInZWindow	bool	Z-Veto
IsSelected	bool	Is good OSDL event (always = 1)
DiLepIdx	std::vector<int>	Indices of the two selected leptons
DiLepFlavour	int	(=0 $\mu\mu$, =1 μe, =2 ee)

Appendix B

Data/MC comparisons

In this appendix, data and Monte Carlo comparison are shown, after the pile up, jet energy, and isolation correction have been performed (see chapter 4). Many different variables are investigated, related to key informations for our anaysis. Distributions for all the three channel (ee, $e\mu$, $\mu\mu$) plus the whole sample ($ee + e\mu + \mu\mu$) are shown. As mentioned in section 4.9, a good agreement is found for all the studied distributions.

PU reweight

To check the goodness of the rewithing algorithm the first variable, whose distibution for data and Monte Carlo is compared, is the

- number of primary vertices (figure B.1).

ABCD variables

The distributions of the variables used in the ABCD analysis (chapter 5) are then investigated. First the *in-analysis variables* (section 5.2.1) are shown:

- invariant mass of the lepton pair (figure B.2);
- MET (figure B.4);
- H_T (figure B.3).

The *out-of-the-analysis variables* (introduced in section 5.2.1) are then presented:

- m(lj) (figure B.5);
- m_T(l-MET) (figure B.6);
- Invariant mass of the most massive hemisphere (figure B.7);
- Y (figure B.8).

A good agreement in the yelds and shapes between data and Monte Carlo for these variables is mandatory, since the ABCD method is set-up using Monte Carlo informations (section 5.3).

Lepton and jet veriables

Some acceptance cuts for leptons and jets are imposed (as discussed in section 4.1). In particular, electrons, muons, and jets are required to fulfill some conditions about transverse momenta (p_T) and angular emission (η). A comparison between data and simulations for the distributions of these variables has been investigated, and is shown in the following plots.

For leptons:

- p_T and η of the first (hardest) lepton (figures B.9, B.10);
- p_T and η of the second lepton (figures B.11, B.12).

For jets:

- p_T and η of the first (hardest) jet (figures B.14, B.15);
- number of good jets per event (figure B.13).

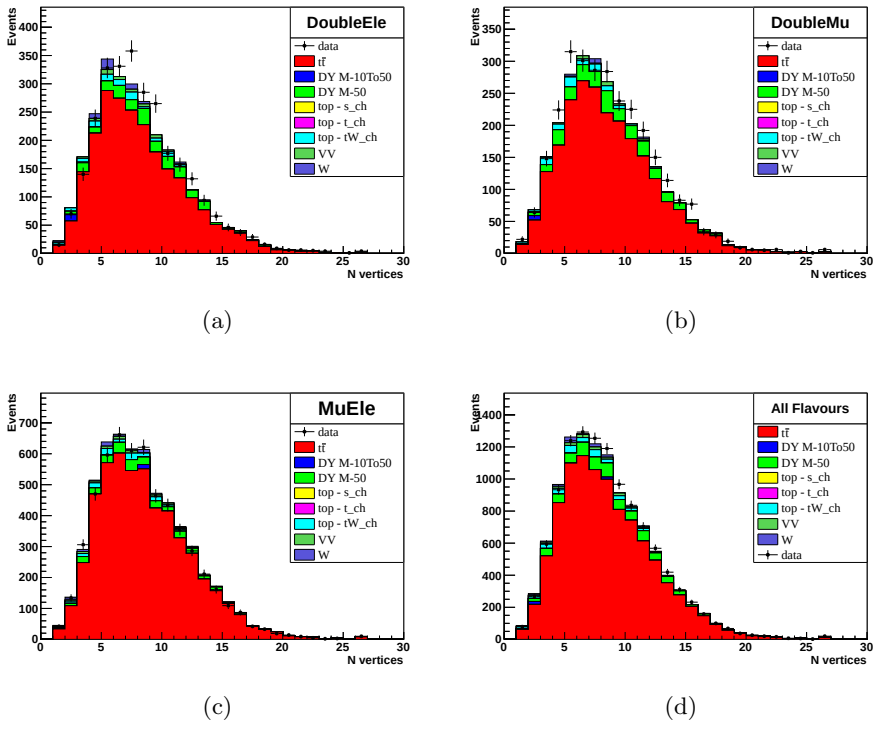
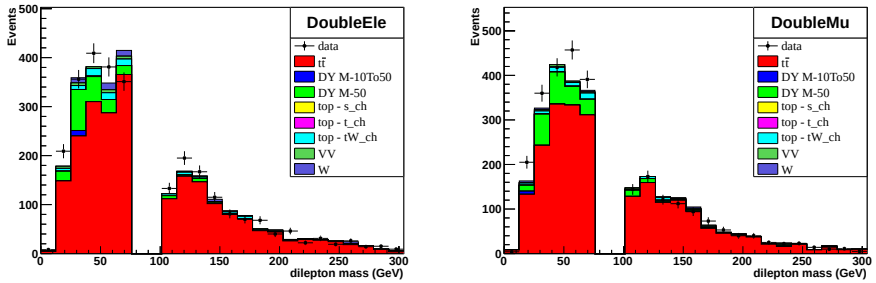
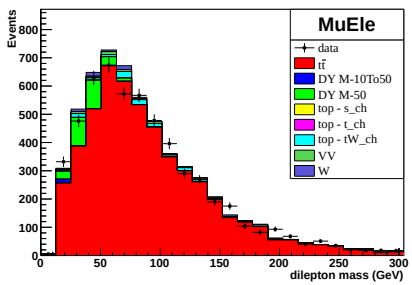


Figure B.1: Number of primary vertices per event for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background (after the PU reweight) in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee+e\mu+\mu\mu$.

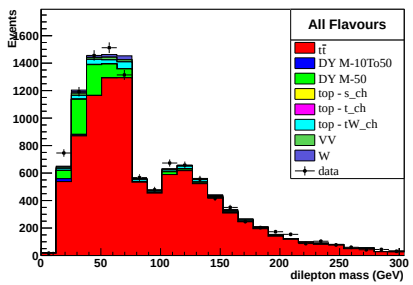


(a)

(b)

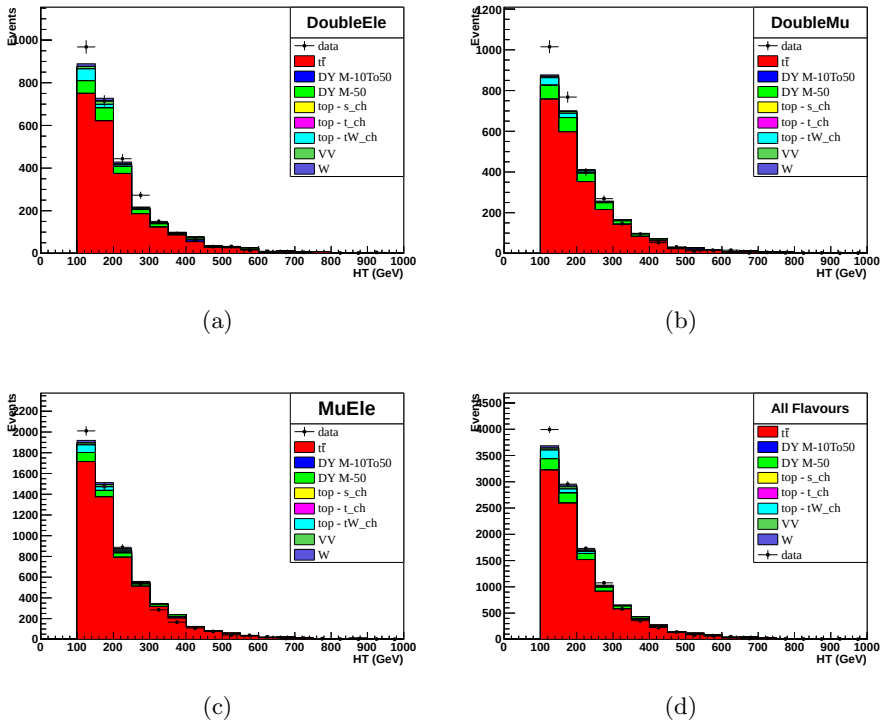


(c)



(d)

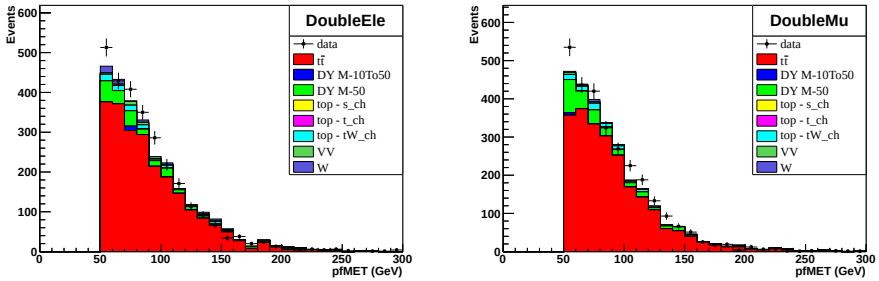
Figure B.2: Dilepton invariant mass distribution for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.



(a) (b)

(c) (d)

Figure B.3: Dilepton invariant mass distribution for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.



(a) (b)

(c) (d)

Figure B.4: MET distribution for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

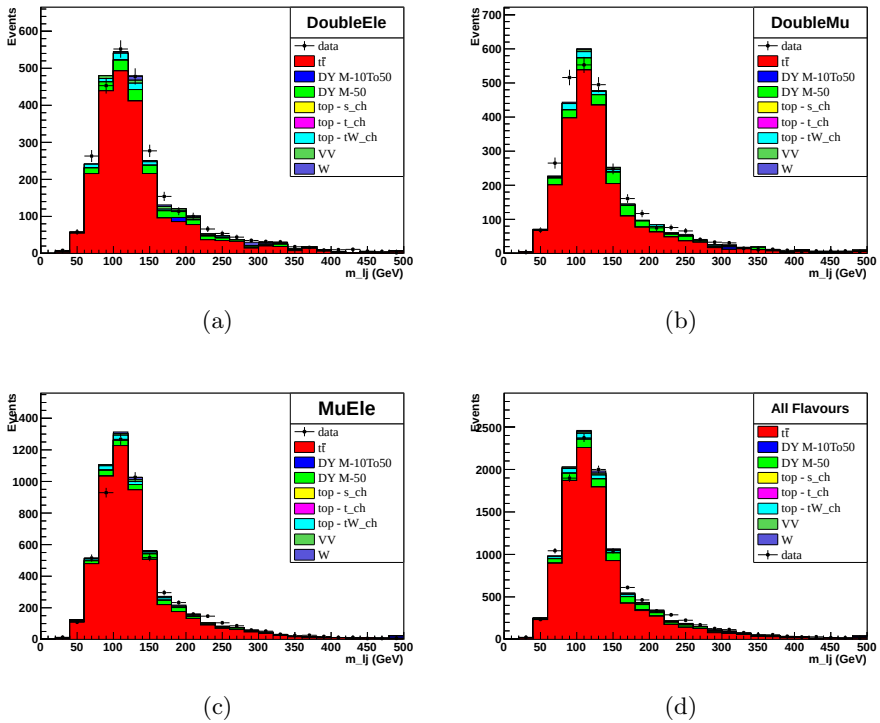
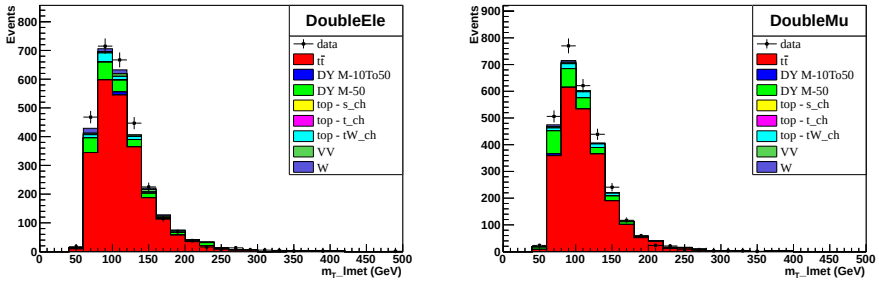
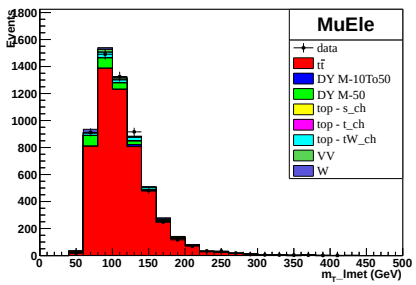


Figure B.5: m(lj) distribution for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

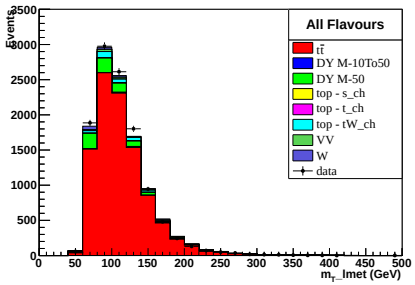


(a)

(b)

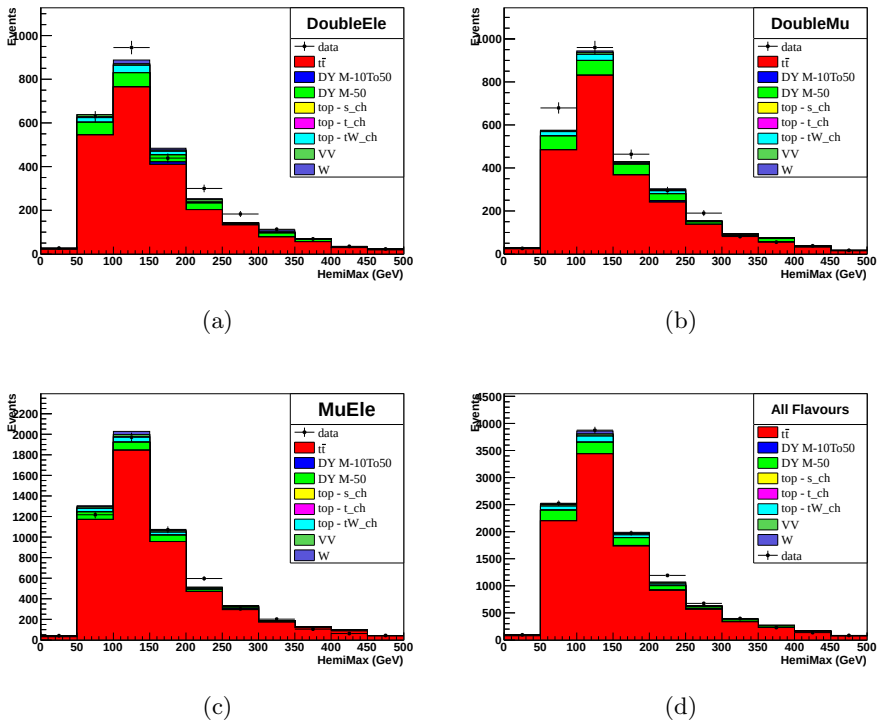


(c)



(d)

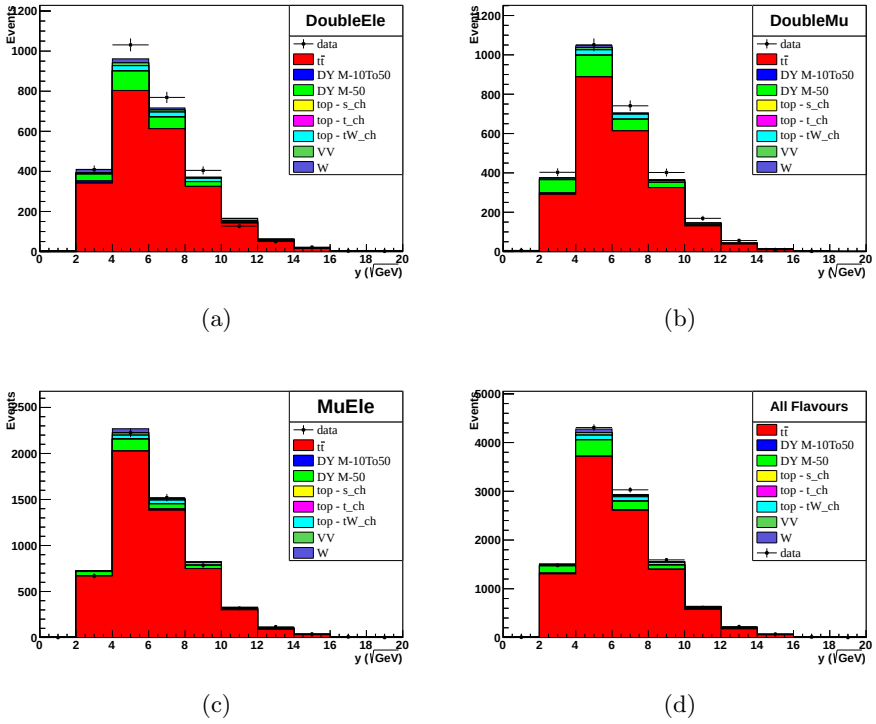
Figure B.6: m_T(l-MET) distribution for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.



(a) (b)

(c) (d)

Figure B.7: Distribution of the invariant mass of the most massive hemisphere for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.



(a) (b)

(c) (d)

Figure B.8: Y distribution for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

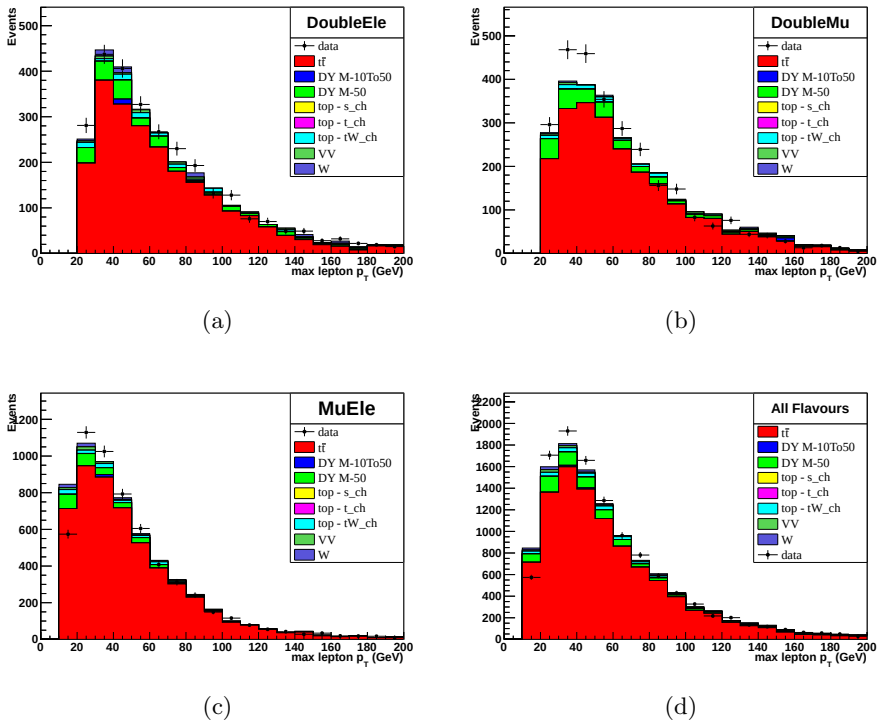
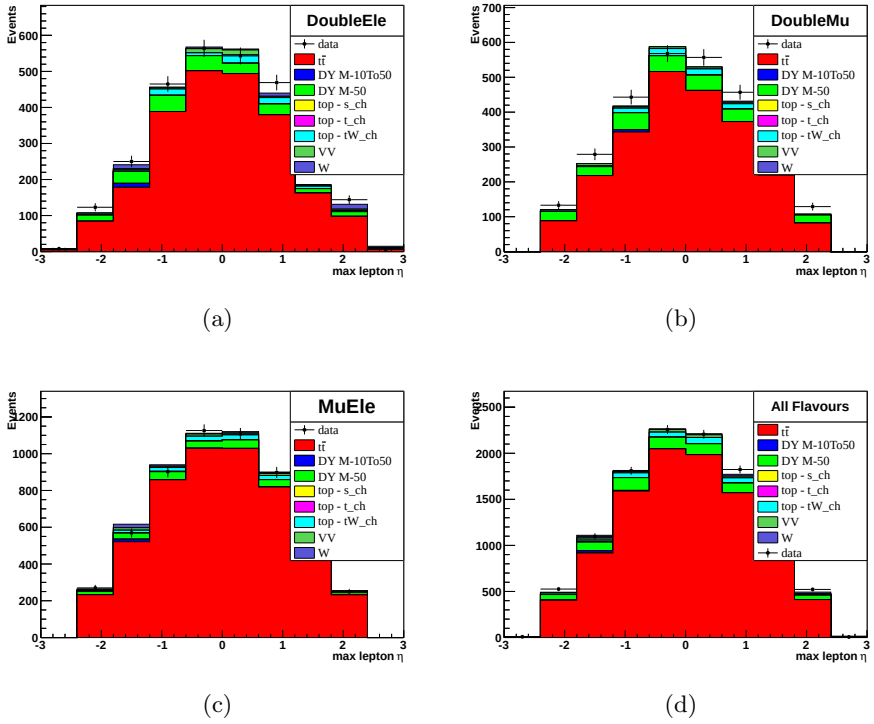


Figure B.9: Transverse momentum distribution of the first lepton for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.



(a) (b)

(c) (d)

Figure B.10: η distribution of the first lepton for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

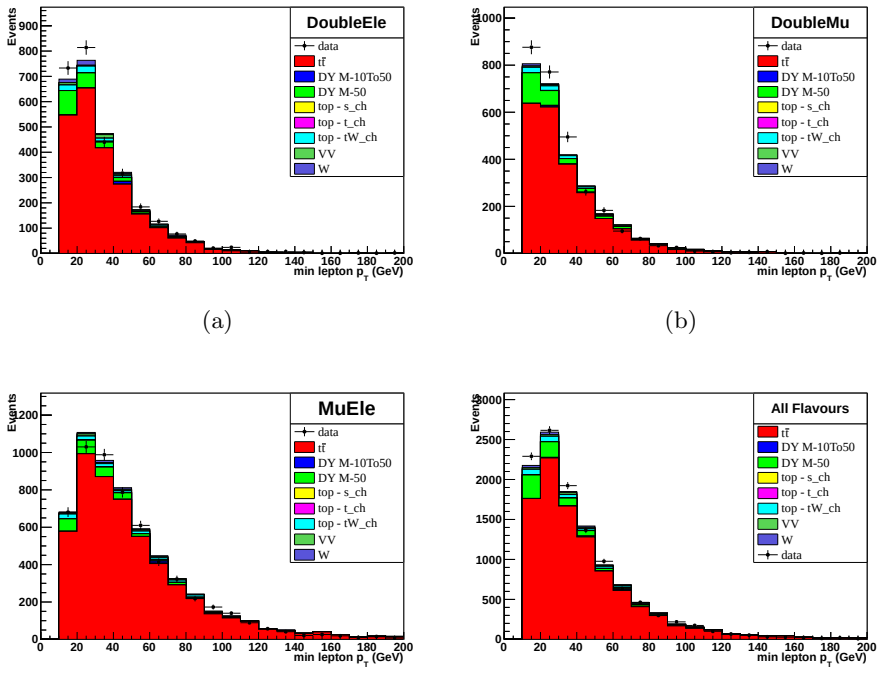
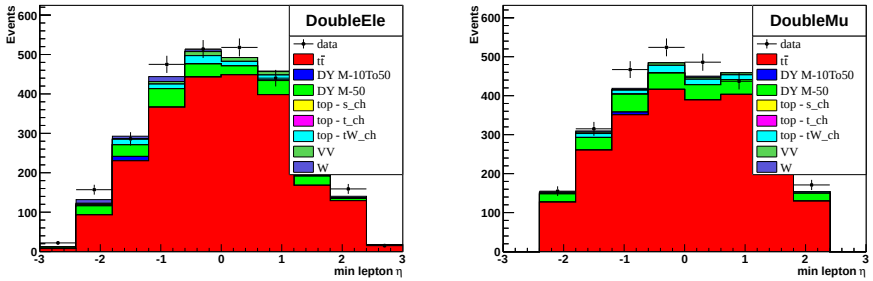
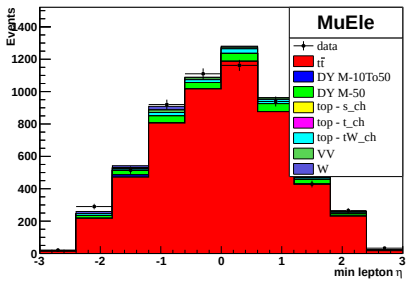


Figure B.11: Transverse momentum distribution of the second lepton for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

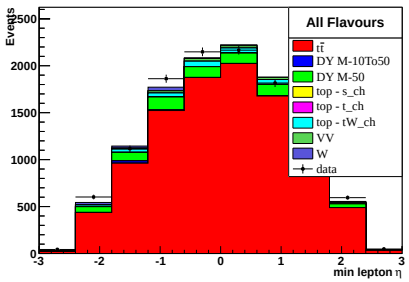


(a)

(b)



(c)



(d)

Figure B.12: η distribution of the second lepton for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

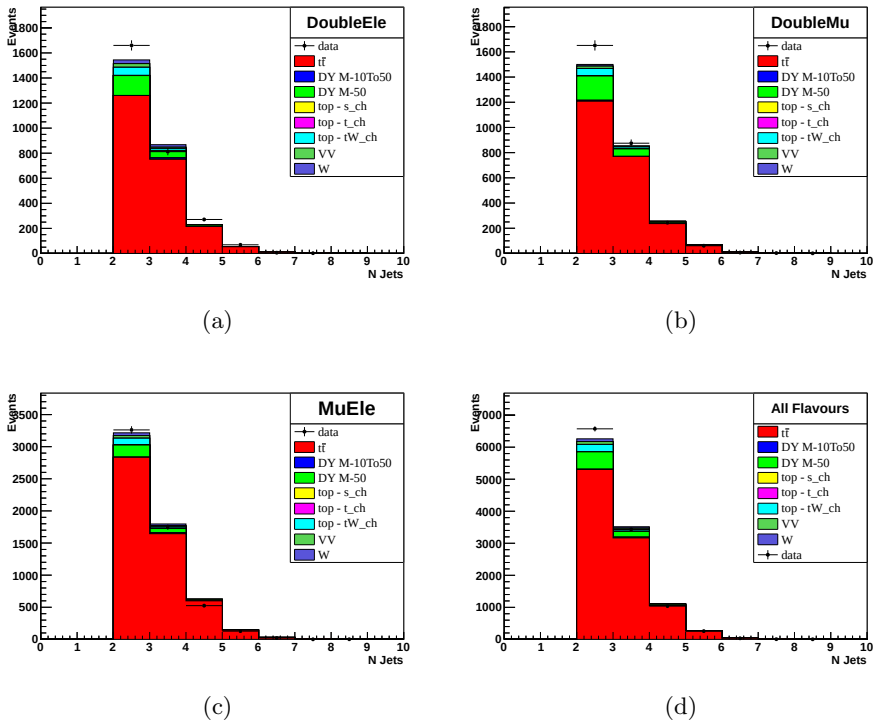


Figure B.13: Distribution of number og good jets per event for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

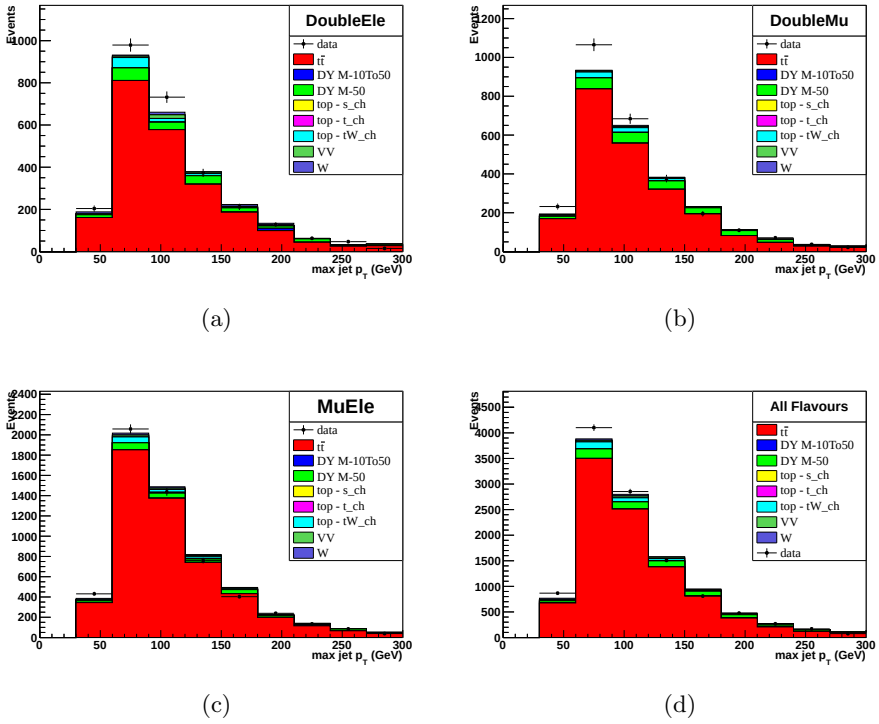


Figure B.14: Transverse momentum distribution of the first jet for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

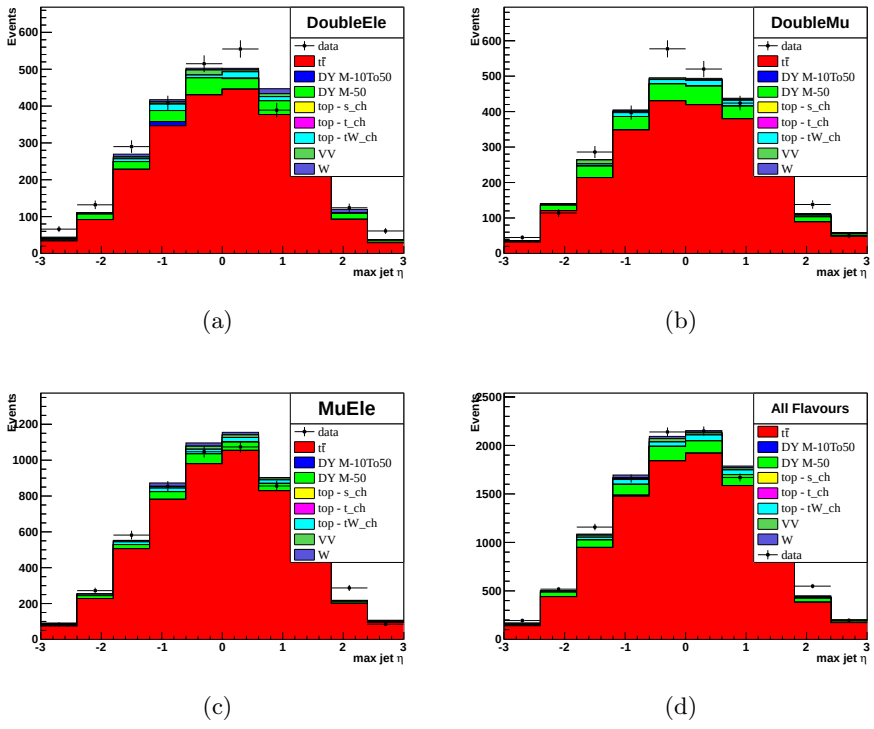


Figure B.15: η distribution of the first jet for $\mathcal{L} = 5fb^{-1}$ data at $\sqrt{s} = 7$ TeV, compared to Monte Carlo simulation of SM background in channel: (a) ee; (b) $\mu\mu$; (c) $e\mu$; (d) $ee + e\mu + \mu\mu$.

Appendix C

ABCD pairs

This appendix contains some plots about the setup of the ABCD method. The 2-dimensional distribution of all the 21 pairs are shown in figures C.1 - C.4. Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$). The signal sample used is LM1.

Figures C.5 - C.8 show the goodness of the estimate ($r[b]$ and $r[s+b]$ ratios) at LM1, LM4, and LM13 independently evaluated for the channels $\mu\mu$, $e\mu$, and ee, and in the sum of the three channels.

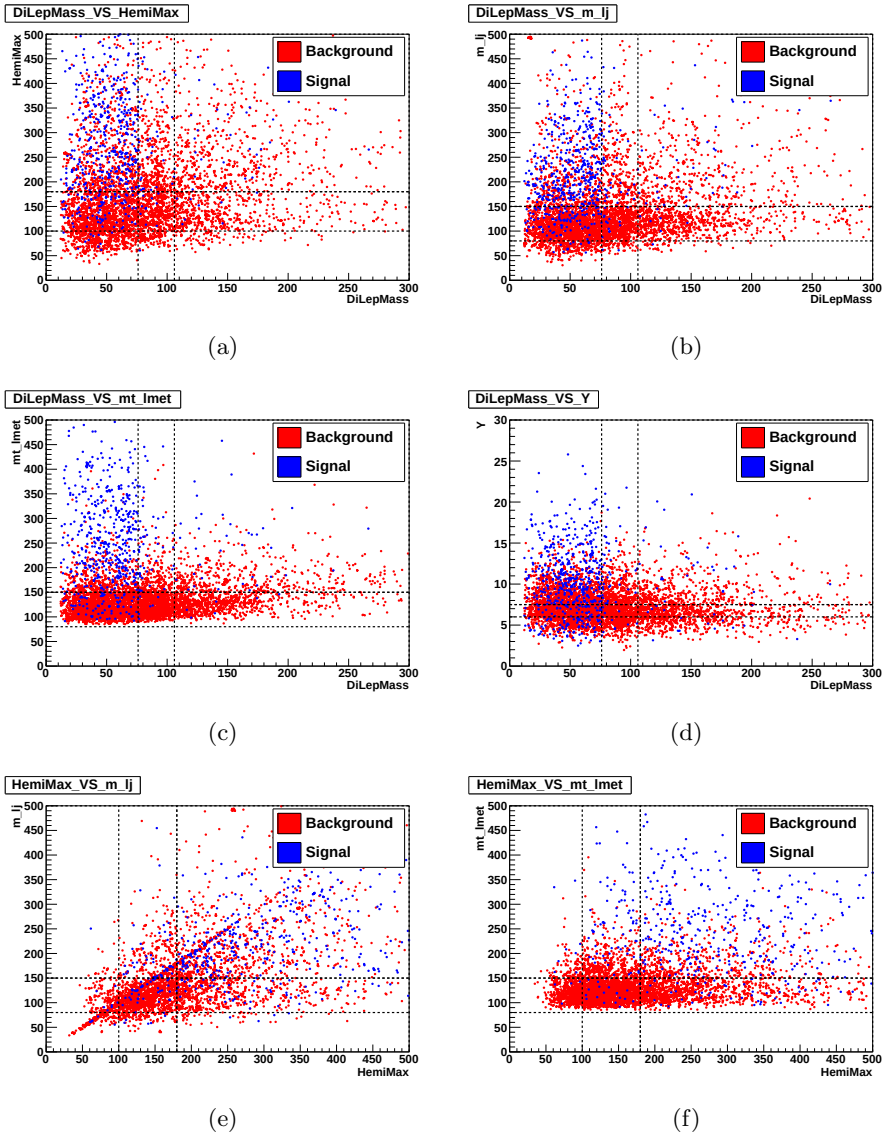


Figure C.1: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by: (a) m(ll) vs. HemiMax, (b) m(ll) vs. m(lj), (c) m(ll) vs. m_T(l-MET), (d) m(ll) vs. Y, (e) HemiMax vs. m(lj), (f) HemiMax vs. m_T(l-MET). Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).

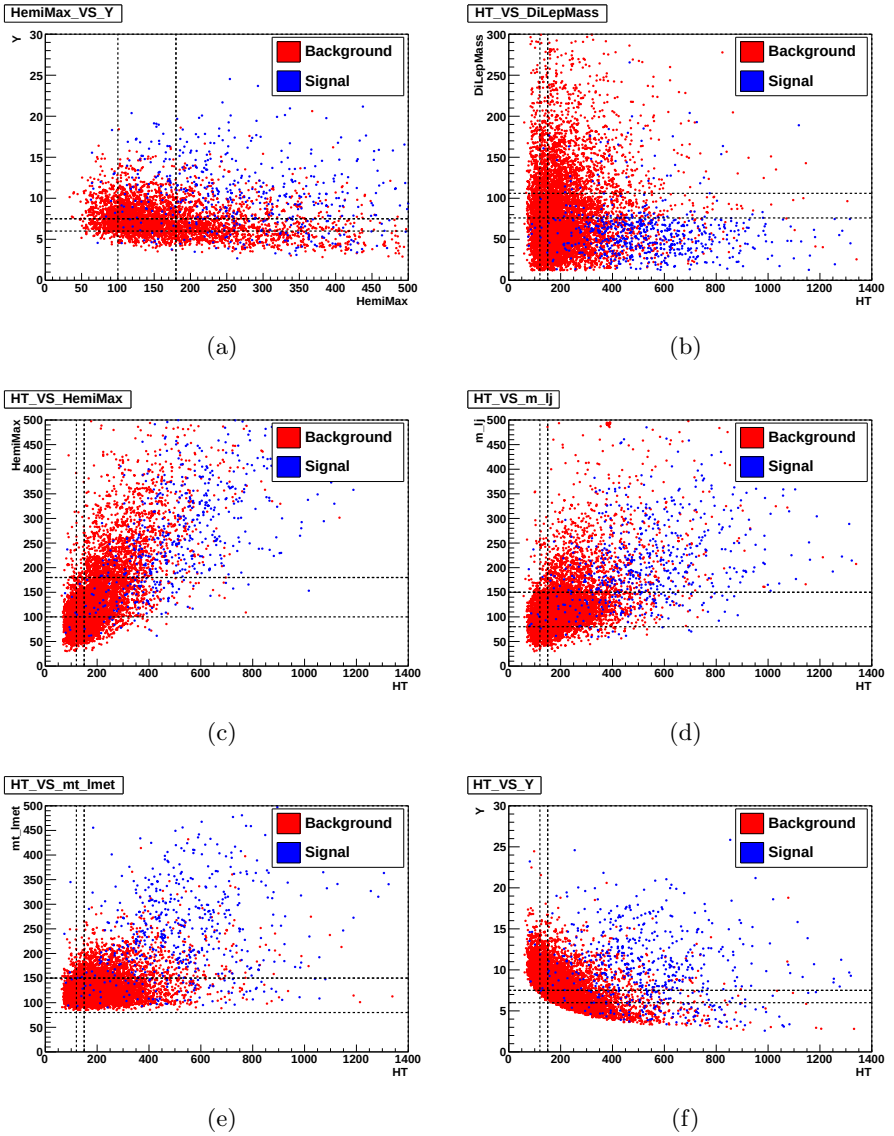


Figure C.2: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by: (a) HemiMax vs. Y, (b) H_T vs. m(ll), (c) H_T vs. HemiMax, (d) H_T vs. m(lj), (e) H_T vs. m_T(l-MET), (f) H_T vs. Y. Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).

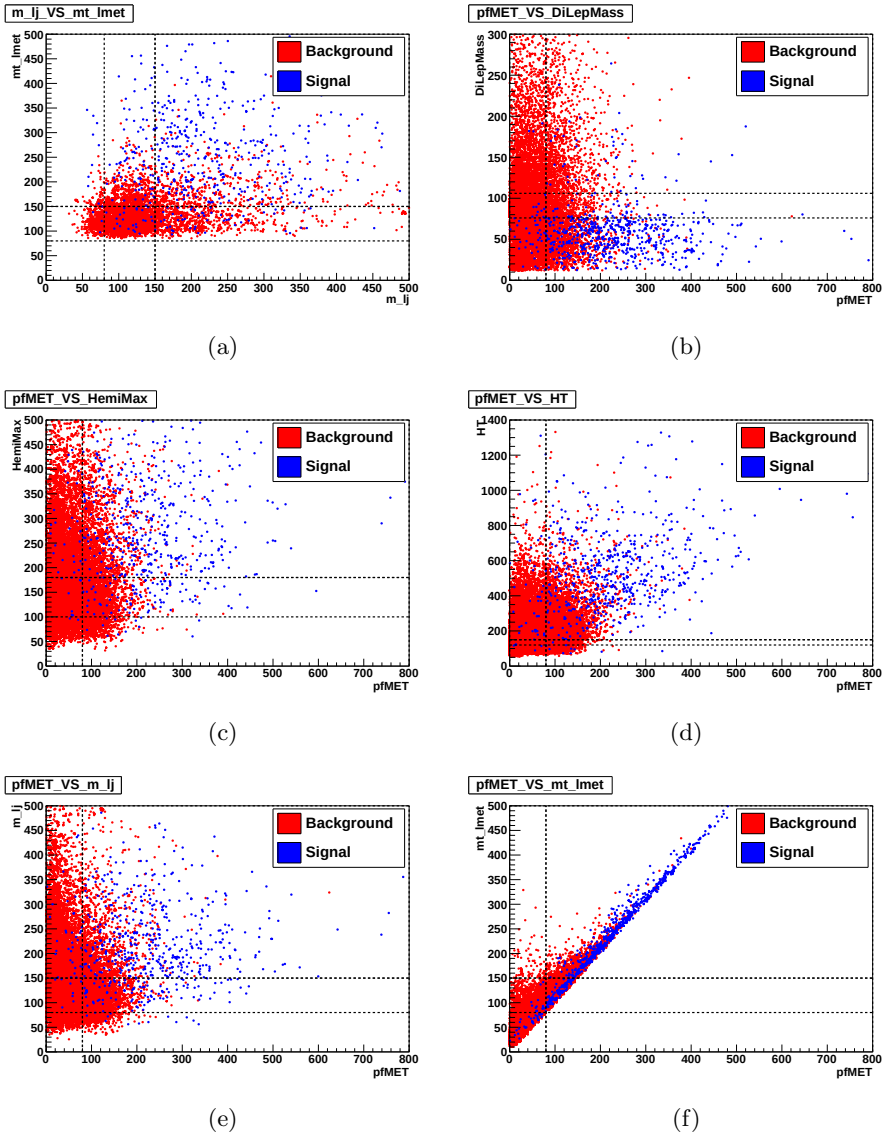
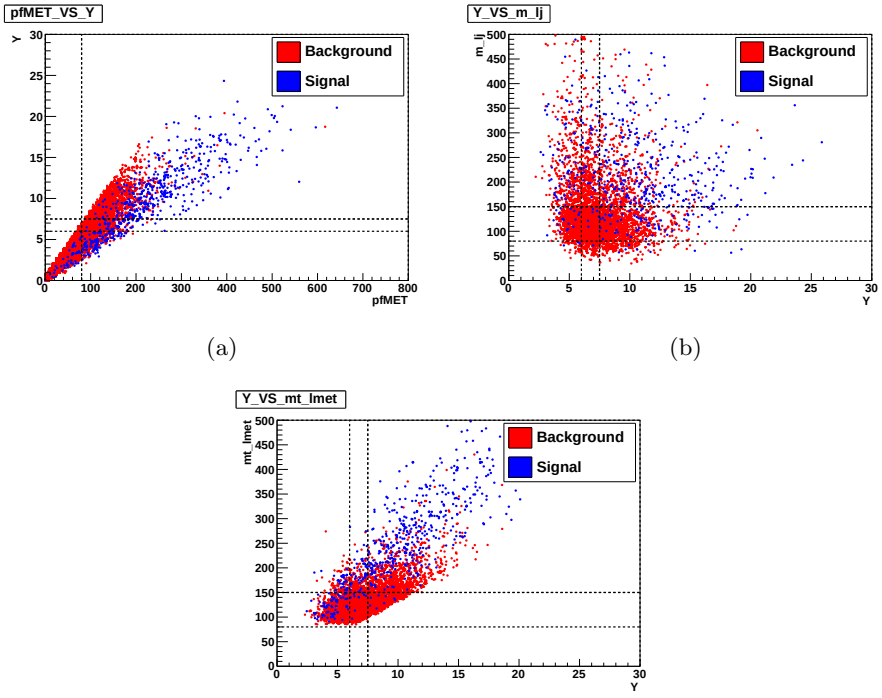


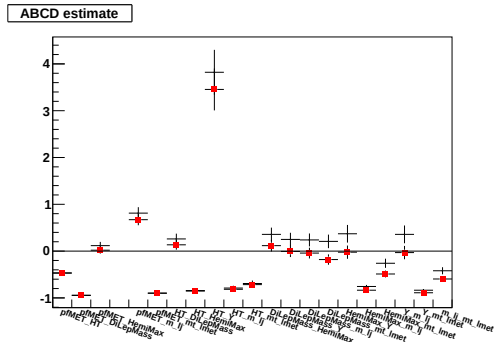
Figure C.3: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by: (a) m(lj) vs. m_T(l-MET), (b) MET vs. m(ll), (c) MET vs. HemiMax, (d) MET vs. H_T, (e) MET vs. m(lj), (f) MET vs. m_T(l-MET). Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).



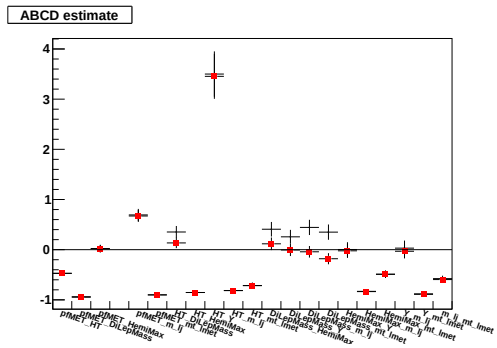
(a) (b)

(c)

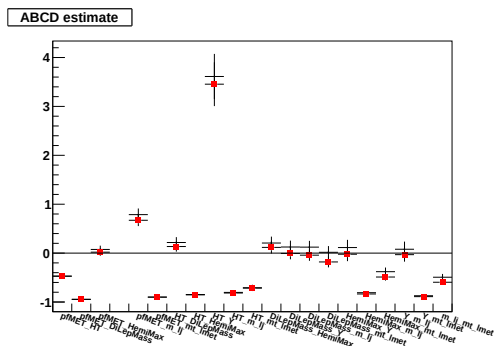
Figure C.4: Distribution of Monte Carlo background (red) and signal at LM1 (blue) in the plane defined by: (a) MET vs. Y, (b) Y vs. m(lj) , (c) Y vs. m_T(l-MET). Distributions concern the sum of all three channels (ee+eμ+$\mu\mu$).



(a)

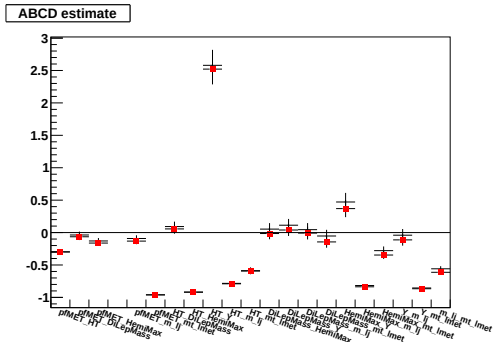


(b)

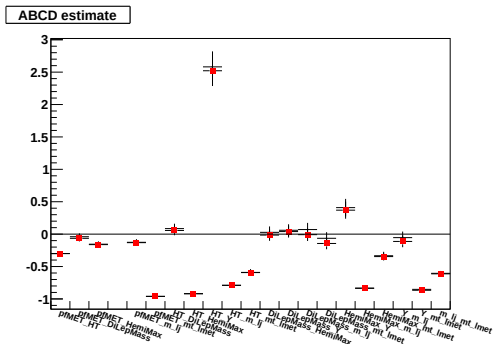


(c)

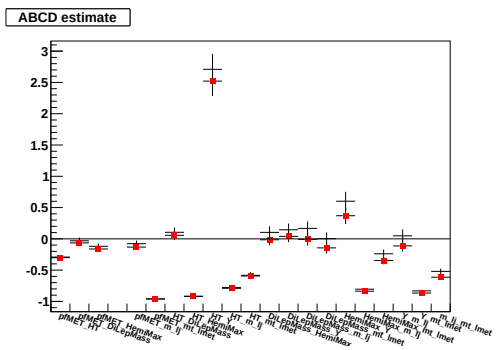
Figure C.5: $r[b]$ (red) and $r[s+b]$ (black) evaluated for Monte Carlo in $\mu\mu$ channel. Signal at (a) LM1, (b) LM4, (c) LM13.



(a)

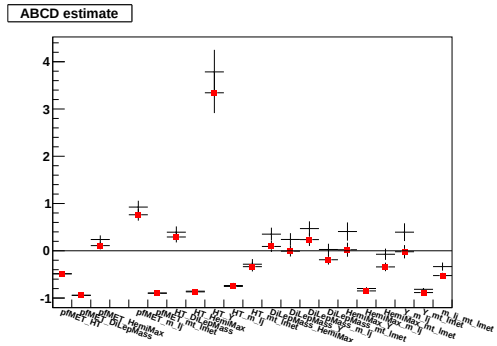


(b)

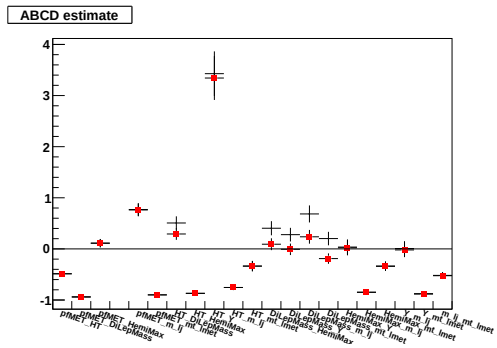


(c)

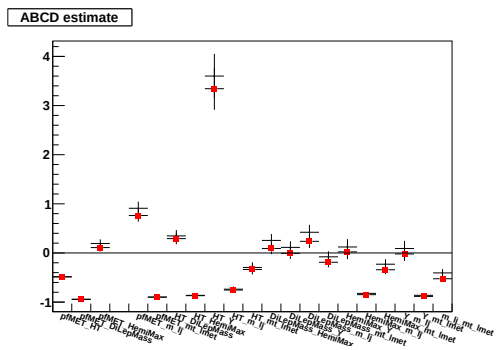
Figure C.6: $r[b]$ (red) and $r[s+b]$ (black) evaluated for Monte Carlo in $e\mu$ channel. Signal at (a) LM1, (b) LM4, (c) LM13.



(a)

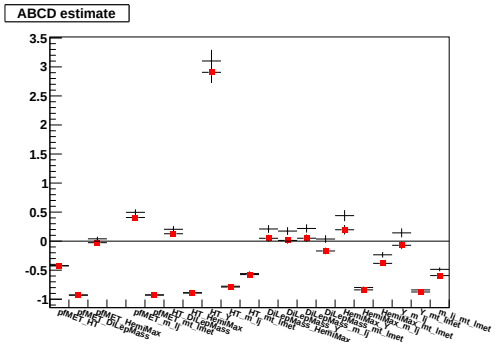


(b)

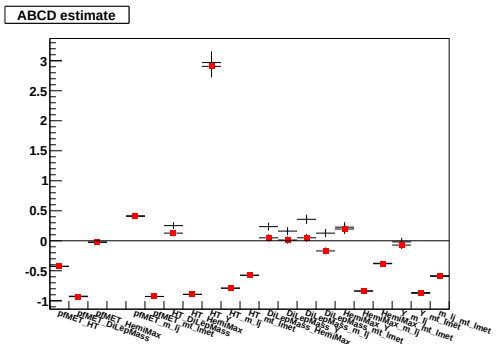


(c)

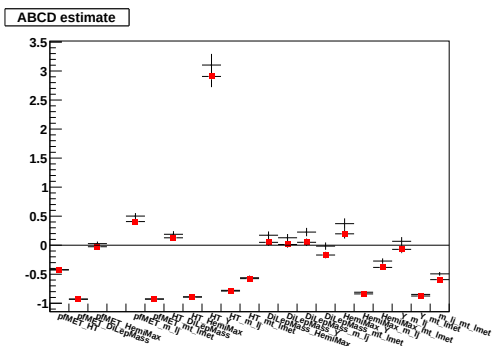
Figure C.7: $r[b]$ (red) and $r[s+b]$ (black) evaluated for Monte Carlo in ee channel. Signal at (a) LM1, (b) LM4, (c) LM13.



(a)



(b)



(c)

Figure C.8: $r[b]$ (red) and $r[s+b]$ (black) evaluated for Monte Carlo in $\mu\mu + e\mu + ee$ channel. Signal at (a) LM1, (b) LM4, (c) LM13.

References

[1] Francis. Halzen and Alan D. Martin. **Quarks and Leptons**. Wiley, 1985. 6

[2] David J. Gross and Frank Wilczek. **Ultraviolet Behavior of Non-Abelian Gauge Theories**. *Phys. Rev. Lett.*, **30**:1343–1346, Jun 1973. 7

[3] M. Gell-Mann. **The interpretation of the new particles as displaced charge multiplets**. *Il Nuovo Cimento (1955-1965)*, **4**:848–866, 1956. 10.1007/BF02748000. 7

[4] F.J. Hasert et al. **Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment**. *Nuclear Physics B*, **73**(1):1 – 22, 1974. 9

[5] Daniel Denegri. % bf The discovery of the W and Z. *Physics Reports*, **403-404**:107 – 145, 2004. 9

[6] LEP Electroweak Working Group. **http://lepewwg.web.cern.ch/LEPEWWG/**. 9, 10, 235

[7] TeVatron Electroweak Working Group. **http://tevewwg.fnal.gov**. 9

[8] CMS collaboration. % bf Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, **716**(1):30 – 61, 2012. 9, 10, 235

[9] ATLAS collaboration. % bf Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, **716**(1):1 – 29, 2012. 9

[10] F. Zwicky. **On the Masses of Nebulae and of Clusters of Nebulae**. *Journal of Astrophysics*, **86**:217, 1937. 11

[11] E. Komatsu et al. **Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation**. *The Astrophysical Journal Supplement Series*, **192**(2):18, 2011. 11

[12] Joakim Edsjö and Paolo Gondolo. **Neutralino relic density including coannihilations**. *Phys. Rev. D*, **56**:1879–1894, Aug 1997. 11

[13] D.I. Kazakov. **Supersymmetry in particle physics: the renormalization group viewpoint**. *Physics Reports*, **344**(4-6):309 – 353, 2001. 12

[14] **Electroweak parameters of the Z0 resonance and the standard model**. *Physics Letters B*, **276**(1-2):247 – 253, 1992. 12

[15] Steven Weinberg. **Implications of dynamical symmetry breaking**. *Phys. Rev. D*, **13**:974–996, Feb 1976. 14

[16] J. Wess and B. Zumino. **Supergauge transformations in four dimensions**. *Nuclear Physics B*, **70**(1):39 – 50, 1974. 15

[17] Stephen P. Martin. **A Supersymmetry primer**. *arXiv:hep-ph/9709356*, 1997. 16, 17

[18] Xerxes Tata Howard Baer. *Weak Scale Supersymmetry: From Superfields to Scattering Events*. 18

[19] H. Farstenau U. Amaldi, W. deBoer. **Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP**. *Physics Letters B*, **260**(3-4):447 – 455, 1991. 20

[20] Howard Georgi and S. L. Glashow. **Unity of All Elementary-Particle Forces**. *Phys. Rev. Lett.*, **32**:438–441, 1974. 21

[21] W. Boer, R. Ehret, and D.I. Kazakov. **Predictions of SUSY masses in the minimal supersymmetric grand unified theory**. *Zeitschrift fr Physik C Particles and Fields*, **67**:647–664, 1995. 21

[22] I.N. McArthur. **Gauging of nonlinearly realized symmetries**. *Nuclear Physics B*, **452**(12):456 – 468, 1995. 22

[23] G.G. Ross and R.G. Roberts. **Minimal supersymmetric unification predictions**. *Nuclear Physics B*, **377**(3):571 – 592, 1992. 24

[24] Stephen P. Martin. **Supersymmetry: Theoretical implications**, July 2012. Review of Discovery Physics Results from ICHEP. 27

[25] Lyndon Evans and Philip Bryant. **LHC Machine**. *Journal of Instrumentation*, **3**(08), 2008. 29

[26] G. Papotti for the LHC team. **LHC: status and prospects**, July 2012. LPCC. 31

[27] The ALICE Collaboration. **The ALICE experiment at the CERN LHC**. *Journal of Instrumentation*, **3**(08):S08002, 2008. 33

[28] ATLAS Collaboration. **The ATLAS Experiment at the CERN Large Hadron Collider**. *Journal of Instrumentation*, **3**(08):S08003, 2008. 33

[29] The LHCb Collaboration. **The LHCb Detector at the LHC**. *Journal of Instrumentation*, **3**(08):S08005, 2008. 33

[30] LHCf Collaboration. **The LHCf detector at the CERN Large Hadron Collider**. *Journal of Instrumentation*, **3**(08):S08006, 2008. 33

[31] MoEDAL Collaboration. **MoEDAL technical design report**. *CERN/LHCC 2009-006*, 2010. 33

[32] TOTEM Collaboration. **The TOTEM Experiment at the CERN Large Hadron Collider**. *Journal of Instrumentation*, **3**(08):S08007, 2008. 33

[33] CMS collaboration. **Physics Technical Design Report, Volume I**. *CERN/LHCC 2006-021*, 2006. 36

[34] CMS collaboration. **Physics Technical Design Report, The Tracker project**. *CERN/LHCC 98-6*, 1998. 38

[35] CMS collaboration. **Measurement of Tracking Efficiency**. *CMS PAS TRK- 10-002*, 2010. 41

[36] CMS collaboration. **The ECAL Project, Technical Design Report**. *CERN-LHCC 33*, 1997. 42

[37] CMS collaboration. **Electromagnetic calorimeter commissioning and performance with 7 TeV data**. *CMS PAS EGM-10-002*, 2010. 44

[38] CMS collaboration. **The HCAL Project, Technical Design Report**. *CERN-LHCC 31*, 1997. 44

[39] CMS collaboration. **The Muon Project, Technical Design Report**. *CERN-LHCC 32*, 1997. 45

[40] CMS Collaboration. **Performance of the CMS drift tube chambers with cosmic rays**. *Journal of Instrumentation*, **5**(03):T03015, 2010. 46

[41] CMS Collaboration. **Performance of the CMS cathode strip chambers with cosmic rays**. *Journal of Instrumentation*, **5**(03):T03018, 2010. 46

[42] CMS collaboration. **The Trigger System, Technical Design Report**. *CERN-LHCC 38*, 2000. 46

[43] CMS collaboration. **CMS High Level Trigger, Technical Design Report**. *CERN-LHCC 21*, 2007. 47

[44] D.Spiga et al. **The CMS remote analysis builder (CRAB)**. *HiPC '07*, 2007. 48

[45] H1 and ZEUS collaborations. **Combined Measurement and QCD Analysis of the Inclusive $e^+ - p$ Scattering Cross Sections at HERA**. *DESY 09-158*, 2009. 52

[46] S. Alekhin et. al. **HERA and the LHC - A workshop on the implications of HERA for LHC physics: Proceedings - Part A**. *CERN-2005-014*, 2005. 52

[47] **CMSSW Application Framework**. 54

[48] S. Agostinelli et al. % bf Geant4: a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **506**(3):250 – 303, 2003. 54

[49] S. Mrenna et. al. **PYTHIA 6.4 Physics and Manual**. *JHEP 0605*, 2006. 55

[50] Johan Alwall, Pavel Demin, Simon de Visscher, Rikkert Frederix, Michel Herquet, Fabio Maltoni, Tilman Plehn, David L. Rainwater, and Tim Stelzer. **MadGraph/MadEvent v4: the new web generation**. *Journal of High Energy Physics*, **2007**(09):028, 2007. 55

[51] B.C. Allanach. % bf SOFTSUSY: A program for calculating supersymmetric spectra. *Computer Physics Communications*, **143**(3):305 – 331, 2002. 62

[52] CMS collaboration. **Search for new physics in events with opposite-sign leptons, jets, and missing transverse energy in pp collisions at $\sqrt{s} = 7$ TeV**. *arXiv:1206.3949*, 2012. 62

[53] CMS collaboration. % bf Search for physics beyond the standard model in events with a Z boson, jets, and missing transverse energy in pp collisions at $\sqrt{s} = 7$ TeV. *Physics Letters B*, **716**(2):260 – 284, 2012. 62

[54] CMS collaboration. **Search for New Physics with Same-Sign Isolated Dilepton Events with Jets and Missing Transverse Energy**. *Phys. Rev. Lett.*, **109**:071803, 2012. 62

[55] CMS collaboration. **Search for New Physics in the Multijet and Missing Transverse Momentum Final State in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV**. *Phys. Rev. Lett.*, **109**, 2012. 62

[56] % bf twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideFastSimulation. 64

[57] Carola F. Berger et al. **Supersymmetry without prejudice**. *Journal of High Energy Physics*, **2009**(02):023, 2009. 64

[58] Josh Thompson on behalf of the CMS collaboration. **CMS searches for 3rd generation SUSY**, July 2012. 66, 237

[59] CMS collaboraiton. **Search for supersymmetry with the razor variables**. *CMS PAS SUS-11-008*, 2011. 67

[60] CMS collaboration. **Performance of tau reconstruction algorithms in 2010 data collected with CMS**. *CMS PAS TAU-11-001*, 2011. 68

[61] Massimiliano Chiorboli. *Supersymmetric particle reconstruction with the CMS detector at LHC*. PhD thesis, Università di Catania - Dottorato di ricerca in Fisica, XV ciclo. 69

[62] M.Spira W.Beenakker, R.Hoepker. **PROSPINO: A Program for the Production of Supersymmetric Particles in**

Next-to-leading Order QCD. *arXiv:hep-ph/2611232*, 1996. 73, 249

[63] CMS Collaboration. **Electron reconstruction and identification at sqrt(s) = 7 TeV**. *CMS-PAS-EGM-10-004*. 76

[64] CMS Collaboration. **Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV**. *CMS-PAS-MUO-10-002*. 78

[65] CMS Collaboration. **Performance of Jet Algorithms in CMS**. *CMS-PAS-JME-07-003*, 2009. 84

[66] **Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET**. *CMS-PAS-PFT-09-001*, 2009. 84

[67] Catania Susy Group TWiki Space. **https://twiki.cern.ch/twiki/bin/viewauth/CMS/CTSusyAnalysis**. 87

[68] CMS Vertex Reconstruction Offline Guide. **https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideVertexReco**. 87

[69] R Fruhwirth, Wolfgang Waltenberger, and Pascal Vanlaer. **Adaptive Vertex Fitting**. (CMS-NOTE-2007-008), Mar 2007. 87

[70] **Search for physics beyond the standard model in opposite sign dilepton events in pp collisions at $\sqrt{s} = 7$ TeV**. 92, 93, 104, 117, 121, 155

[71] CMS PileUp twiki.
https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupMCReweightingUtilities. 95

[72] Matteo Cacciari and Gavin P. Salam. % bf Pileup subtraction using jet areas. *Physics Letters B*, **659**:119 – 126, 2008. 96

[73] Yeng-Ming Tzeng. **Search for Vector-like Quarks $T \to tZ(B \to bZ)$ at $\sqrt{s} = 7$ TeV in CMS**, 2012. 99

[74] CMS collaboration. **Determination of jet energy calibration and transverse momentum resolution in CMS**. *Journal of Instrumentation*, **6**(11):P11002, 2011. 101

[75] CMS Collaboration. **Search for New Physics with same sign dileptons, Jets, and Missing Transverse Energy with 4.7 fb^1 of Data**. *CMS-AN-2011/437*, 2011. 102, 104

[76] CMS Collaboration. **Measurement of CMS Luminosity**. *CMS-PAS-EWK-10-004*, 2010. 103

[77] Gigi Cappello. *Studio del fondo $t\bar{t}$ nella ricerca di particelle supersimmetriche con l'esperimento CMS*. PhD thesis, Università di Catania - Tesi di laurea specialistica in Fisica, 2009. 122, 123, 239

[78] from Wikipedia Pearson coefficient.
http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient. 134

[79] A.Tricomi G.Cappello, M.Scinta. **OS dileptons with ABCD Technique**, June 2012. 139

[80] I. Bertram et al. **A Recipe for the construction of confidence limits**. 2000. 167

[81] RooStats Web Home. **https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome**. 170

[82] SUSY Tools and guidelines for limits (2011 data publications). **https://twiki.cern.ch/twiki/bin/view/CMS/SUSYLimitTools2011?skin=drupal**. 172, 178

[83] TTree from Root online resource. **http://root.cern.ch/root/html/TTree.html**. 188

List of Figures

List of Tables

Acknowledgements

This thesis is the systesis not only of my three years work as a Ph.D student, but of a whole path inside the world of High energy Physics, started since my bachelor's degree in 2006. I owe thanks to many people encountred during this journey.

First of all to Prof. Alessia Tricomi, my Ph.D. supervisor and my extraordinary guide in Particle Physics. Then to Prof. Sebastiano Albergo, who always believed in me and always gave me the possibility to express my skills.

Many thanks to the CMS group of Catania. To Prof. Salvatore Costa, Prof. Cristina Tuvé, and Prof. Renato Potenza for their kind support. A very special acknowledgement to Dr. Massimiliano Chiorboli, who has been an irreplaceable example of scientist.

I also wish to give my gratitude to the whole CMS SUSY group, in particular to Luc Pape, Frederic Ronga, Filip Moortgat, Benjamin Stieger, and Marco Peruzzi, who helped me in the developement of my analysis.

And [in Italian from now on...] grazie ai miei genitori, a mia sorella Elisa e a Marco, che credono in me più di quanto non lo faccia io.

Ai colleghi (il termine non rende affatto onore al nostro rapporto): Davide, il nostro inviato all'estero, Salvo, Dario e Adriano, che più che supportarmi mi sopporta...

Grazie anche a Manuel, il "mio" tesista, per la compagnia e soprattutto per l'essenziale aiuto.

Infine grazie ai miei amici (immaginate una lunga lista), senza i quali questa tesi avrebbe visto la luce parecchio tempo prima, no, a ben pensarci non sarebbe esistita affatto.