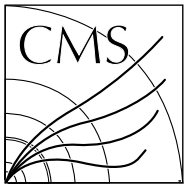
Available on CMS information server

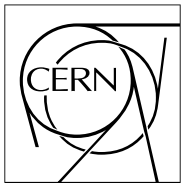
CMS NOTE 2009/XXX



The Compact Muon Solenoid Experiment

CMS Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland



December 6, 2009

Search for Single-Lepton SUSY Signatures (RA4): Background Determinations

List of everyone contributing to this document.

Abstract

An important event topology for SUSY searches is the signature with a single, isolated muon or electron; several jets above a p_T threshold; and a quantity such as missing transverse energy (MET) that signals the presence of one or more unobserved, high-momentum particles. This document describes several methods for data-driven background determinations, whose different approaches will provide essential crosschecks in the search for SUSY signals. The methods focus on the determination of the dominant backgrounds, which are due to $t\bar{t}$, W+jets, and QCD processes. For each method, we study the performance in predicting the background levels, the sensitivity to systematic uncertainties, and the effect of signal contamination in the control regions. We also present a plan for commissioning each method with early data. Finally, we consider how the different methods will be used in combination to obtain a comprehensive picture of the background composition.

Contents

List of Figures

List of Tables

1 Introduction

A key event topology for SUSY searches is the signature with a single, isolated muon or electron; several jets above a p_T threshold; and a quantity such as missing transverse momentum (MET) signaling at least one unobserved, high-momentum particle. This signature is expected in many different SUSY scenarios and is therefore one of great interest.

The single-lepton search is formulated as one part (RA4) of the SUSY Reference Analysis program, in which multiple methods are being developed to provide complementary approaches and extensive crosschecks. Another key aspect of the RA strategy is to ensure that the results from different searches can be compared in a straightforward way, providing a comprehensive and coherent set of signatures to aid in the interpretation of any new physics observations. Because the range of possible models is so large, there is considerable variation in their cross sections and kinematic features. For this reason, we adopt a simple, topological approach in which the analysis selection criteria are only loosely guided by SUSY models.

As for other SUSY signatures, the most basic issues to be addressed center around the relevant physics reconstruction objects: jets, missing transverse momentum (MET), electrons, and muons. We are particularly interested in the mechanisms and event sources that give rise to tails in MET distributions. Developing a deep understanding of detector and algorithm performance is an urgent task to which many groups in CMS are contributing, and this task will continue for the life of the experiment. The SUSY group will contribute to this work as part of its commissioning effort.

This document focuses on the development of robust techniques for data-driven background estimation. The dominant backgrounds for the single-lepton SUSY search at $\sqrt{s} = 10$ TeV are $t\bar{t}$ and W + jets, but other processes, such as QCD multijet production, also produce backgrounds that must be controlled and quantified. Monte Carlo predictions for backgrounds are not sufficiently reliable for (at least) two reasons: many of the relevant cross sections and kinematic distributions are not precisely known, and even if they were, our simulation of the detector and software performance has significant limitations. Thus, key goals for such data-driven techniques are minimal reliance on Monte-Carlo-simulated event samples and developing a full understanding of the strengths and weaknesses of each method. For each method, it is critical to devise a well defined procedure to determine the statistical and systematic uncertainties on the background estimate. In this document, we present a status report on data-driven background estimation methods for the RA4 analysis. Our philosophy is that for a potential scientific discovery of this magnitude, all potentially effective, complementary methods should be pursued.

Because the signal characteristics are poorly constrained, the optimization of selection cuts is a problematic issue. Although one could separately optimize cuts across a broad range of models, we have taken the approach that, for the most part, the definition of selection criteria should be guided primarily by considerations related to backgrounds and robustness of the reconstructed physics objects. This approach leads to a fairly simple event selection. Nevertheless, there can be cases in which a search in a particlar region of phase space, or with some added feature, can lead to a qualitatively different background composition and hence the possibility of a looser event selection. Thus, we envision the possibility of a small number number of branches within RA4, rather than a continuum of optimizations. The section on event selection outlines a possible strategy for multiple selections.

Typically, data-driven techniques rely on control samples that should be free of signal or at least dominated by background. Because the range of SUSY models is so extensive, however, guaranteeing that a given control sample is free of signal is not a trivial problem. Another generic problem is ensuring that the selection procedure for defining the control sample does not produce an uncontrolled bias in the background prediction derived from that sample. These considerations provide additional arguments for using as many complementary methods as possible.

For the purposes of this document, leptons refer to reconstructed muons and electrons; τ-leptons are not explicitly identified, although this possibility is under investigation. We note that the single-lepton analysis selection cuts include a veto on a second lepton; the goal is to avoid statistical overlap between the single-lepton and dilepton samples and to enable a simple interpretation of any observed signals. However, the p_T threshold and other requirements associated with the lepton veto, allow some dilepton and multilepton events to feed down into the nominal single-lepton sample.

The document is organized as follows. Section 2 presents an overview of the basic kinematic features of several SUSY models for the single-lepton signature. We show a set of plots that compares basic distributions of key quantities, to provide a physical picture of the kinds of events that are accessible in our searches.

Section 3 describes a nominal set of event selection cuts, along with a set of alternative cuts that provide other approaches to the analysis.

Section 4 describes the main backgrounds and their kinematic properties. We also show how the backgrounds behave as the selection cuts are applied, and we note which cuts are critical in selecting the event sample. The hierarchy of different backgrounds can vary significantly according to such cuts. The relative importance of $t\bar{t}$ and W+jets backgrounds depends strongly on the requirements on the number of jets, while the amount of QCD background is highly dependent on the minimum lepton p_T threshold.

Section 5 gives a brief synopsis of each method; these methods are then presented in detail in subsequent sections. In each such section, we describe one method in more detail, discuss the performance of the method without and with SUSY signal events mixed into the event sample, discuss systematic uncertainties, and present a commissioning plan for the method based on early data.

In Sec. 20 we present an overall discussion of the methods and discuss how they can be used together to understand the event sample and to gain confidence in a potential discovery. Finally, Sec. 21 presents our conclusions.

2 Signal Characteristics

In this section we describe general aspects of the single-lepton signature in SUSY models.

Section 2.1 gives a brief overview of the characteristics of the SUSY benchmark models most commonly used in CMS. We focus on the mass spectra, cross sections, production mechanisms, and the p_T spectra of the initial pair of SUSY particles.

Section 2.2 presents a physical picture of lepton production in these benchmark models. For a given SUSY mass spectrum, it is not too difficult to understand key features of the decay chains that lead to lepton production. We will see how a set of mass splittings control the relevant decay processes and how both two- and three-body decays can play a role.

In Sec. 2.3, we show a basic set of kinematic distributions for each of the benchmark models. These distributions allow us to characterize the event in terms of leptons, jets, and energy-scale quantities such as missing transverse momentum, focusing mainly on the quantities that are used in the initial selection cuts. (In Sec. 4, we will show the corresponding distributions for a variety background processes.) Many more kinematic distributions are shown in later sections, where the details of data-driven background estimates are described.

Finally, in Sec. 2.4 we depart from the original set of benchmark models and present results from a scan over mSUGRA parameters to give a broader picture of lepton production and measurement sensitivity.

2.1 SUSY benchmark models and their characteristics

For the CM energies we consider, the production of SUSY particles in proton-proton collisions is usually dominated by strong production of colored pairs of particles in four basic final states: gluino-gluino, gluino-squark, squark-squark, and squark-antisquark. These final states can be produced from multiple combinations initial-state partons: gluon-gluon, gluon-quark, quark-quark, quark-antiquark, and gluon-antiquark, as well as in multiple channels, indicated below as s, t, and u. The main strong-interaction processes are

$$\begin{aligned} gg(s,t,u) + q\bar{q}(s,t,u) &\rightarrow \tilde{g}\tilde{g} \\ gq(s,t_{\tilde{g}},t_{\tilde{q}}) &\rightarrow \tilde{g}\tilde{q} \\ qq(t,u) &\rightarrow \tilde{q}\tilde{q} \\ q\bar{q}(s,t) + gg(s,t,u) &\rightarrow \tilde{q}\bar{\tilde{q}} \end{aligned} \quad (1)$$

We indicate the virtual intermediate-state particle with a subscript in the cases where there is more than one possibility. For example, $t_{\tilde{g}}$ indicates a t-channel process in which the exchanged particle is a gluino.

In some situations, however, electroweak production of SUSY particles can yield a significant part of the SUSY production cross section. For example, chargino-neutralino pair production can occur via

$$d\bar{u}(s_{W^-}, t_{\tilde{d}_L}, t_{\tilde{u}_L}) \rightarrow \tilde{\chi}_i^0 \tilde{\chi}_j^- \quad (2)$$

Considering both strong and electroweak processes, as well as the multiple channels, we see that SUSY production is in general quite complicated. The relative importance of these processes depends on the masses of the produced SUSY particles relative to the available CM energy.

Table 2.1: Benchmark low-mass (LM) mSUGRA models used in the SUSY group: cross sections at $\sqrt{s} = 10$ TeV and parameters at the GUT scale.

Model	Cross Section (pb)	m_0 (GeV)	$m_{1/2}$ (GeV)	A_0	$\tan\beta$
LM0	110	200	160	-400	10
LM1	16.06	60	250	0	10
LM2	2.42	185	350	0	35
LM3	11.79	330	240	0	20
LM4	6.70	210	285	0	10
LM5	1.94	230	360	0	10
LM6	1.28	85	400	0	10
LM7	2.90	3000	230	0	10
LM8	2.86	500	300	-300	10
LM9	11.58	1450	175	0	50

Table 2.2: Benchmark low-mass (LM) mSUGRA models: particle masses

Model	$m(\tilde{g})$	$m(\tilde{\chi}_1^0)$	$m(\tilde{\chi}_2^0)$	$m(\tilde{\chi}_1^+)$	$m(\tilde{\chi}_2^+)$	$m(\tilde{t}_1)$	$m(\tilde{b}_1)$	$m(\tilde{u}_L)$	$m(\tilde{e}_L)$	$m(\tilde{\nu}_{eL})$
LM0	409	60	113	114	327	207	356	415	231	217
LM1	603	96	178	179	360	407	510	552	186	167
LM2	827	141	264	267	465	580	671	770	304	292
LM3	597	94	173	174	345	444	548	620	370	360
LM4	687	112	208	210	402	481	598	653	289	276
LM5	851	144	271	273	487	599	734	800	338	327
LM6	932	161	303	305	531	647	785	859	287	275
LM7	637	94	177	179	364	1790	2450	3000	2994	2992
LM8	738	120	228	230	467	544	710	811	539	533
LM9	488	65	110	107	223	882	1008	1474	1450	1447

The mSUGRA models can be specified by four parameters and a sign:

$$m_0, m_{1/2}, A_0, \tan\beta, \text{sign}(\mu) \tag{3}$$

Table 2.1 lists these parameters for a set of low mass (LM) benchmark models used in the CMS SUSY group. The cross sections for these models span the range 1 - 100 pb, but even smaller cross sections are possible, such as for the high-mass (HM) mSUGRA benchmark models, which are discussed in the CMS Physics Technical Design Report. In the full scope of the LHC physics program, we must be consider a very wide range of production cross sections, but for early running we will focus on the possibility of fairly large cross sections of tens of picobarns or more.

It is much easier to understand the phenomenology of SUSY models in terms of parameters at the electroweak scale. Table 2.2 lists particle masses for LM0–LM9, which are obtained by using renormalization group equations to run the particles down from the GUT-scale. We have chosen to focus on ten particles, the gluino ($\tilde{g}$,) the two lowest-mass neutralinos ($\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$), the two lowest-mass charginos ($\tilde{\chi}_1^+$, $\tilde{\chi}_2^+$), the lighter of each of the two t- and b-squark mass eigenstates ($\tilde{t}_1$, $\tilde{b}_1$), a representative squark ($\tilde{u}_L$), and a representative slepton and sneutrino ($\tilde{e}_L^-$, $\tilde{\nu}_{eL}$). From this set of particles and their masses, we can understand much of the basic physical behavior of the models. In particular, we would like to identify the dominant decay modes, as well as the dominant sources of leptons, using simple arguments (see the following section).

Figure 2.1 shows the mass spectrum for LM1. The gluino is the heaviest SUSY particle; it can decay via a large number (22) of two-body decay modes to squark-antiquark pairs (see below). The $\tilde{\chi}_1^0$, which is the LSP, has a mass around 100 GeV, and the lightest squark is the $\tilde{t}_1$, with a mass around 400 GeV. All of the squark are heavier than the heaviest gaugino, and the squarks generally decay into chargino + quark or neutralino + quark.

Figure 2.2 shows the distributions of pairs of SUSY particle types produced at $\sqrt{s} = 10$ TeV in the initial hard-scattering process for the benchmark models LM0, LM1, LM3, LM6, and LM9. Figure 2.3 shows the distributions of initially produced pairs of particle types for LM2, LM4, LM5, LM7, and LM8. From these plots, we can draw several conclusions:

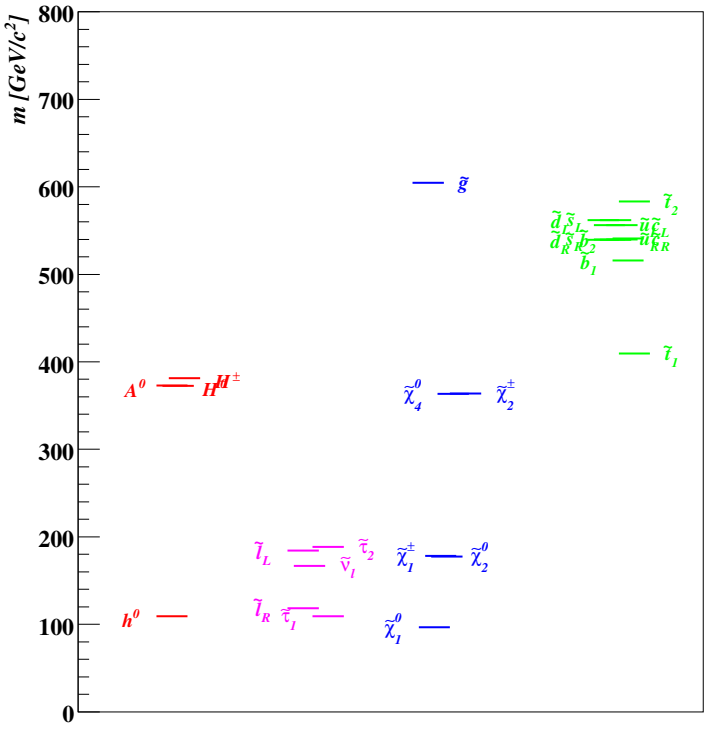


Figure 2.1: Mass spectrum of the LM1 model.

- At 10 TeV, a typical paradigm is for $\tilde{g}\tilde{q}$ to be the largest production process, followed by $\tilde{q}\tilde{q}$. This pattern occurs in LM1, LM2, LM3, LM4, LM5, LM6, and LM8.
- A different paradigm applies to LM7 and LM9, Here the largest production process is $\tilde{\chi}_i\tilde{\chi}_j$ (we do not distinguish among charge states), followed by $\tilde{g}\tilde{g}$. In both of these models the squark masses are very high compared with the gluino mass, and the masses of the neutralino and chargino are low compared with the gluino mass.
- In general, the cross section for $\tilde{q}\tilde{q}$ is larger than that for $\tilde{q}\tilde{\bar{q}}$. An exception is LM1, where the reverse is true. But in this case the two cross sections are still comparable.
- The cross section for $\tilde{g}\tilde{g}$ is typically smaller than that for $\tilde{q}\tilde{\bar{q}}$.

In summary several of the models have the hierarchy of production cross sections

$$\sigma(\tilde{g}\tilde{q}) > \sigma(\tilde{q}\tilde{q}) > \tilde{q}\tilde{\bar{q}} > \sigma(\tilde{g}\tilde{g}),$$

but this is by no means universal.

In contrast to the production cross sections, the decay patterns of the different models do not depend on the CM energy and will be discussed below. Note that the production cross section for each process can be obtained by multiplying the fractions from Figures 2.2 and 2.3 by the total cross sections listed in Table 2.1.

Figure 2.4 shows the p_T distributions of each particle in the initial pairs of SUSY particles pair produced in the benchmark models. The distribution are normalized to the same area to simplify comparison of their shapes. The p_T spectra for LM7 and LM9 models, which have substantial electroweak production, show peaks at much lower values than the other models.

2.2 Decay patterns and lepton production in SUSY models

In the SUSY models considered, most of the leptons are produced from a small number of types of parent particles: $W^{\pm}$, $\tilde{\chi}_i^{\pm}$, $\tilde{\chi}_j^0$, $\tilde{\ell}^{\pm}$, and Z. Figures 2.5 and 2.6 show the distribution of the types of particles whose decays produce muons. (We include only the particles listed in the Pythia documentation lines, so secondary muons from b and c semileptonic decays, such as $b \to c\mu^-\bar{\nu}$, are not included.) The models fall into two broad classes: those in which

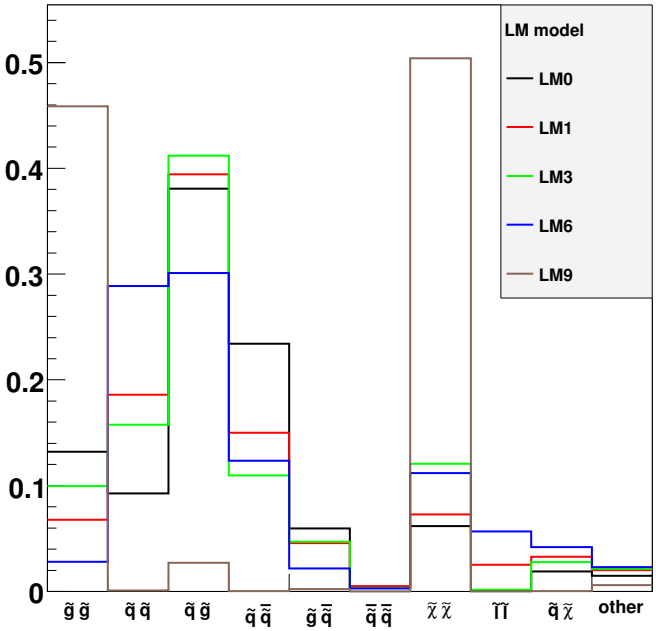


Figure 2.2: Distribution of SUSY particle types produced in the initial hard scatter in SUSY benchmark models LM0, LM1, LM3, LM6, and LM9.

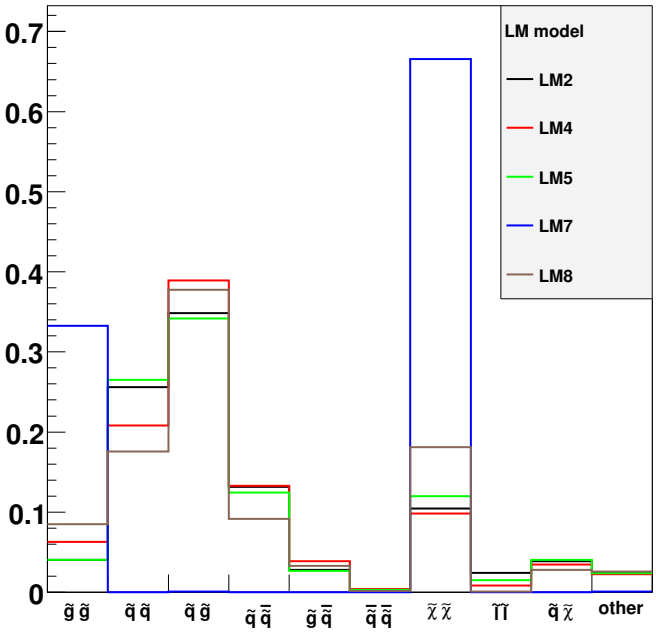


Figure 2.3: Distribution of SUSY particle types produced in the initial hard scatter in SUSY benchmark models LM2, LM4, LM5, LM7, and LM8.

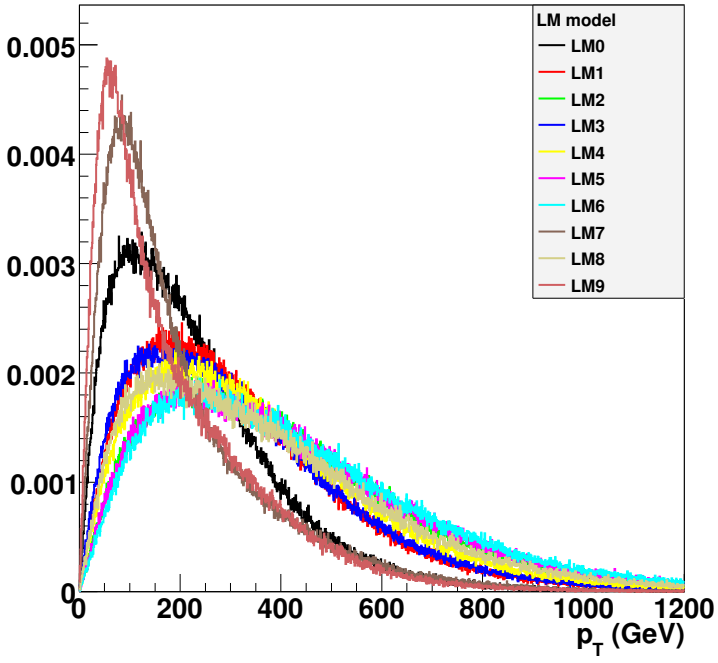


Figure 2.4: Generated p_T spectra of SUSY particles produced in the initial hard scatter. The properties of the LM benchmark models are described in the text.

the largest source of muons is $\chi_i^{\pm}$ (chargino) decay and those in which the largest source is W-decay (or in some cases Z-boson decay).

To understand how SUSY particle masses control the decay patterns, and to help understand the sources of lepton production, we consider the following questions:

- What are the dominant gluino decay modes? Which squarks, charginos, and neutralinos can be produced in the gluino decays?
- What are the dominant decays of the squarks that are most likely to be produced, either directly in the initial production process or via gluino decay?
- What are the dominant decays of the charginos and neutralinos that are produced in the decays of the gluinos and squarks?

In each case, we focus on the question of whether the dominant decays are two-body or three-body modes.

Table 2.3 lists the decays of spin-$1/2$ SUSY particles–gluinos, neutralinos, and charginos–that have the largest branching fractions. (Some of these branching fractions are nevertheless small.) Table 2.4 lists the decays of spin 0 (scalar) SUSY particles – squarks, sleptons, and sneutrinos – that have the largest branching fractions. We will see below how the mass hierarchy plays a central role in determining the decay patterns.

We start by examining gluino decays. The gluino decays strongly via the two-body modes $\tilde{g} \to \tilde{q}\bar{q}$ and $\tilde{g} \to \tilde{\bar{q}}q$, as long as

$$m(\tilde{g}) > m(\tilde{q}) + m(\bar{q}). \tag{4}$$

If this condition is not met, the decay becomes a three-body process in which a virtual intermediate-state squark decays weakly. This situation can lead to a long gluino lifetime in certain models, such as split SUSY, where the squarks are much heavier than the gluino.

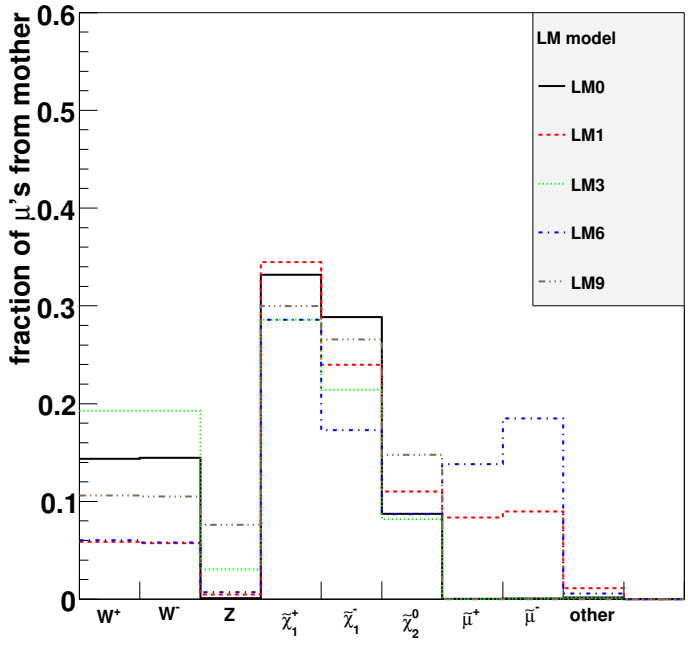


Figure 2.5: SUSY models: distribution of parent particles of muons in LM0, LM1, LM3, LM6, and LM9

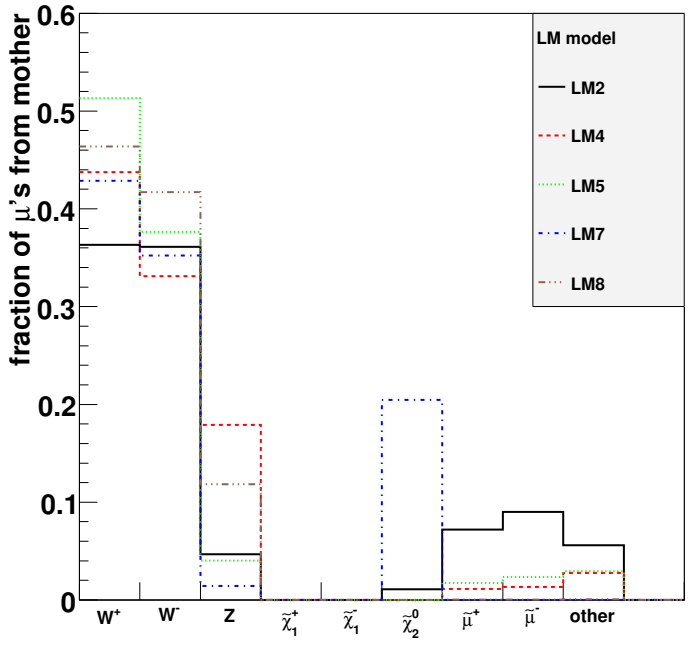


Figure 2.6: SUSY models: distribution of parent particles of muons in LM2, LM4, LM5, LM7, and LM8

Table 2.3: Spin-1/2 particles: decay modes with the largest branching fraction for SUSY LM benchmark models.

Model	$\tilde{g}$	$\tilde{\chi}_1^0$	$\tilde{\chi}_2^0$	$\tilde{\chi}_1^+$	$\tilde{\chi}_2^+$
LM0	$\tilde{b_1}\bar{b}$+c.c. (2×0.27)	-	$\tilde{\chi}_2^0\tau^+\tau^-$ (0.079)	$\tilde{\chi}_1^0 u\bar{d}$ (0.33)	$\tilde{t}_1\bar{b}$ (0.48)
LM1	$\tilde{b_1}\bar{b}$+c.c. (2×0.095)	-	$\tilde{\tau}_1^-\tau^+$+c.c. (2×0.26)	$\tilde{\tau}_1^+\nu_\tau$ (0.45)	$\tilde{\chi}_2^0 W^+$(0.27)
LM2	$\tilde{b_1}\bar{b}$+c.c. (2×0.13)	-	$\tilde{\tau}_1^-\tau^+$+ c.c. (2×0.48)	$\tilde{\tau}_1\nu_\tau$ (0.95)	$\tilde{\chi}_2^0 W^+$ (0.24)
LM3	$\tilde{b_1}\bar{b}$+c.c. (2×0.50)	-	$\tilde{\chi}_1^0 b\bar{b}$ (0.19)	$\tilde{\chi}_1^0 u\bar{d}$ (0.34)	$\tilde{\chi}_2^0 W^+$ (0.39)
LM4	$\tilde{b_1}\bar{b}$+c.c. (2×0.12)	-	$\tilde{\chi}_1^0 Z$ (1.0)	$\tilde{\chi}_1^0 W^+$ (1.0)	$\tilde{\chi}_1^0 W^+$ (0.32)
LM5	$\tilde{t_1}\bar{t}_1$+c.c. (2×0.11)	-	$\tilde{\chi}_1^0 h$ (0.87)	$\tilde{\chi}_1^0 W^+$ (0.98)	$\tilde{\chi}_2^0 W^+$ (0.31)
LM6	$\tilde{t_1}\bar{t}_1$+c.c. (2×0.093)	-	$\tilde{\nu}_{\tau 1}\bar{\nu}_\tau$+ c.c. ($2 \times 0.097$)	$\tilde{\nu}_{\tau 1}\tau^+$ (0.21)	$\tilde{\chi}_2^0 W^+$ (0.28)
LM7	$\chi_2^0 b\bar{b}$ (0.086)	-	$\tilde{\chi}_1^0 d\bar{d}$ (0.15)	$\tilde{\chi}_1^0 W^+$ (1.0)	$\tilde{\chi}_2^0 W^+$ (0.38)
LM8	$\tilde{t_1}\bar{t}$+c.c. (2×0.42)	-	$\tilde{\chi}_1^0 Z$ (1.0)	$\tilde{\chi}_1^0 W^+$ (1.0)	$\tilde{\chi}_2^0 W^+$ (0.34)
LM9	$\chi_2^0 b\bar{b}$ (0.16)	-	$\tilde{\chi}_1^0 d\bar{d}$ (0.15)	$\tilde{\chi}_1^0 u\bar{d}$ (0.33)	$\tilde{\chi}_2^0 W^+$ (0.65)

Table 2.4: Spin-0 particles: decay modes with the largest branching fraction for SUSY LM benchmark models.

Model	$\tilde{u}_L$	$\tilde{b}_1$	$\tilde{t}_1$	$\tilde{\mu}_L^-$	$\tilde{\nu}_\mu$
LM0	$\tilde{\chi}_1^+ d$ (0.66)	$\tilde{t}_1 W^-$ (0.43)	$\tilde{\chi}_1^+ b$ (1.0)	$\tilde{\chi}_1^- \nu_\mu$ (0.55)	$\tilde{\chi}_1^+ \mu^-$ (0.58)
LM1	$\tilde{\chi}_1^+ d$ (0.64)	$\tilde{\chi}_1^- t$ (0.49)	$\tilde{\chi}_1^+ b$ (0.63)	$\tilde{\chi}_1^0 \mu^-$ (0.87)	$\tilde{\chi}_1^0 \nu_\mu$ (1.0)
LM2	$\tilde{\chi}_1^+ d$ (0.64)	$\tilde{\chi}_1^- t$ (0.43)	$\tilde{\chi}_1^+ b$ (0.47)	$\tilde{\chi}_1^0 \mu^-$ (0.54)	$\tilde{\chi}_1^0 \nu + \mu$ (0.71)
LM3	$\tilde{\chi}_1^+ d$ (0.61)	$\tilde{\chi}_1^- t$ (0.36)	$\tilde{\chi}_1^+ b$ (0.51)	$\tilde{\chi}_1^- \nu_\mu$ (0.54)	$\tilde{\chi}_1^+ \mu^-$ (0.46)
LM4	$\tilde{\chi}_1^+ d$ (0.64)	$\tilde{\chi}_1^- t$ (0.38)	$\tilde{\chi}_1^+ b$ (0.53)	$\tilde{\chi}_1^- \nu_\mu$ (0.48)	$\tilde{\chi}_1^+ \mu^-$ (0.46)
LM5	$\tilde{\chi}_1^+ d$ (0.64)	$\tilde{\chi}_1^- t$ (0.36)	$\tilde{\chi}_1^+ b$ (0.45)	$\tilde{\chi}_1^- \nu_\mu$ (0.40)	$\tilde{\chi}_1^0 \nu_\mu$ (0.47)
LM6	$\tilde{\chi}_1^+ d$ (0.64)	$\tilde{\chi}_1^- t$ (0.37)	$\tilde{\chi}_1^+ b$ (0.43)	$\tilde{\chi}_1^- \nu_\mu$ (1.0)	$\tilde{\chi}_1^0 \nu_\mu$ (1.0)
LM7	$\tilde{g}u$ (0.84)	$\tilde{g}b$ (0.70)	$\tilde{g}t$ (0.67)	$\tilde{\chi}_1^- \nu_\mu$ (0.49)	$\tilde{\chi}^+ \mu^-$ (0.56)
LM8	$\tilde{\chi}_1^+ d$ (0.52)	$\tilde{\chi}_1^- t$ (0.34)	$\tilde{\chi}_1^+ b$ (0.45)	$\tilde{\chi}_1^- \nu_\mu$ (0.56)	$\tilde{\chi}_1^+ \mu^-$ (0.59)
LM9	$\tilde{g}u$ (0.83)	$\tilde{g}b$ (0.58)	$\tilde{g}t$ (0.58)	$\tilde{\chi}_2^- \nu_\mu$ (0.33)	$\tilde{\chi}_1^+ \mu^-$ (0.46)

Typically, the lightest squarks are $\tilde{t}_1$ and $\tilde{b}_1$, the lighter of the two top- and bottom-squark mass eigenstates. In LM0, two-body gluino decays are allowed, since

$$\begin{aligned} m(\tilde{g}) &> m(\tilde{t_1}) + m(t) \\ m(\tilde{g}) &> m(\tilde{b_1}) + m(b). \end{aligned} \tag{5}$$

However, $m(\tilde{g}) < m(\tilde{u}_L) + m(u)$, so not all of the quark flavors are allowed in two-body modes, and the processes

$$\begin{aligned} \tilde{g} &\to \tilde{t}_1\bar{t}, \qquad \bar{\tilde{t}}_1 t \\ \tilde{g} &\to \tilde{b}_1\bar{b}, \qquad \bar{\tilde{b}}_1 b \end{aligned} \tag{6}$$

have large branching fractions (23% each for each of the top+stop decays and 27% each for the decays to sbottom+bottom). These modes account for essentially the full gluino decay width in LM0.

In LM1 and LM4, the gluino can decay via similar modes as in LM0, but also via many additional two-body modes such as $\tilde{g} \to \tilde{u}_L\bar{u}$. In LM1 the gluino has over twenty possible decay modes with branching fractions above 2%. The number of decay modes (22) arises from

$$[6\ (\text{quarkflavors}) \times 2\ (\text{squark partners}) \times 2\ (\text{c.c.})] - 2, \tag{7}$$

where the two subtracted is due to the t_2 mass being too large to allow a two-body decay.

LM3 is interesting in that $\tilde{g} \to t_1\bar{t}$ is forbidden but $\tilde{g} \to b_1\bar{b}$ is allowed, even though $m(\tilde{t_1}) < m(\tilde{b_1})$. This is simply due to the large mass of the t quark. In fact, $B(\tilde{g} \to b_1\bar{b}) \approx 100\%$ in LM3. For most gluino decays, the mass of the daughter squark is negligible, so the relevant quantity for determining whether the process can proceed as a two-body mode is $m(\tilde{g} - m(\tilde{q})$ to a good approximation. But the case of $\tilde{g} \to \tilde{t}\bar{t}$ is a very important exception, since t_1 is the lightest squark in many models.

In both LM5 and LM6, the $\tilde{t}\bar{t}$+c.c. modes have the largest branching fractions, but many other two-body decays contribute at the few percent level, as for LM1.

LM7 and LM9 provide a sharp contrast. Here all of the two-body gluino decays are forbidden, because the squarks are very heavy. As a consequence, the gluino decays via three-body modes that proceed through an off-shell squark

in the intermediate state, for example, $\tilde{g} \to \tilde{b}^* \bar{b}$. The off-shell squark then decays weakly, producing a neutralino plus quark (*e.g.*, $\tilde{b}^* \to \tilde{\chi}_1^0 b$) or a chargino plus quark (*e.g.*, $\tilde{b}^* \to \tilde{\chi}_1^- t$) in the final-state.

We next consider the squark decays. In the SUSY models considered here, the squarks are heavier than the neutralinos and charginos, so the pattern of squark decays is different from that of u, d, s, c, and b quarks, which are lighter than the standard model gauge bosons through which they decay. In these SM decays, the gauge bosons ($W^{\pm}$, except in the case of loop processes) are therefore virtual, leading to three-body processes such as $b \to c\ell^- \bar{\nu}$.

The paradigm for squark decays (two-body decays) resembles to that for top quark decay, $t \to b\mathrm{W}^-$, with an on-shell W boson. In LM0, the $\tilde{t}_1$ or $\tilde{b}_1$ that are produced as daughter particles of the gluino can decay via two-body electroweak processes. By comparing the masses of the charginos and neutralinos to these squarks, we see that the decay modes

$$\begin{aligned} \tilde{t}_1 &\to \tilde{\chi}_1^+ b \,(100\%) \\ \tilde{b}_1 &\to \tilde{\chi}_2^0 b \,(29\%) \\ \tilde{b}_1 &\to \tilde{\chi}_1^- t \,(24\%) \\ \tilde{b}_1 &\to \tilde{t}_1 W^- \,(43\%) \end{aligned} \tag{8}$$

are kinematically allowed. The branching fractions are listed in parentheses.

Squarks are not only produced in gluino decay, but also in the initial hard scatter, as described in Eq. 1. Since these are typically $\tilde{u}$ or $\tilde{d}$ squarks, their masses are even higher than those of the $\tilde{t}_1$ and $\tilde{b}_1$, so they are also able to decay via two-body modes into neutralinos and charginos. However, in some cases, the $\tilde{u}$ and $\tilde{d}$ squark masses are sufficiently high than they can decay through two-body decays to a gluino and the same-flavor squark:

$$\tilde{u}_L \to \tilde{g} u \tag{9}$$

Such processes are not favored in LM0, where the gluino in nearly degenerate with the $\tilde{u}$ and $\tilde{d}$ squarks, but in LM7, for example, $B(\tilde{u}_L \to \tilde{g}u) \simeq 84\%$ and $B(\tilde{d}_R \to \tilde{g}d) \simeq 99\%$. This occurs because the gluino is lighter than the squarks, which also implies that the gluino decay is a three-body process.

In summary, two-body squark decays are the normal paradigm. Figures 2.5 and 2.6 confirm that squark decays are not themselves directly a significant source of leptons in these models. The production of W bosons in squark decays, however, is one important source of isolated leptons. In this case, we regard the W as the parent of the lepton rather than the squark, because the W is on shell.

We next investigate the decays of the charginos and neutralinos. From the masses (Table 2.2), we can see that in LM0 the natural two-body decays are kinematically forbidden:

$$\begin{aligned} \tilde{\chi}_1^-(114) &\to \tilde{\mu}(231)\bar{\nu} \\ \tilde{\chi}_1^-(114) &\to \tilde{\bar{\nu}}(217)\mu^- \\ \tilde{\chi}_1^-(114) &\to \tilde{\chi}_1^0(60) W^-(80) \\ \tilde{\chi}_2^0(113) &\to \tilde{\chi}_1^0(60) Z(91) \end{aligned} \tag{10}$$

(11)

However, the latter two processes can occur via off-shell W and Z bosons, that is, as three-body decays, leading to

$$\begin{aligned} \tilde{\chi}_1^+ &\to \tilde{\chi}_1^0 u\bar{d} \,(33\%) \\ \tilde{\chi}_1^+ &\to \tilde{\chi}_1^0 c\bar{s} \,(33\%) \\ \tilde{\chi}_1^+ &\to \tilde{\chi}_1^0 e^+ \nu_e \,(12\%) \\ \tilde{\chi}_1^+ &\to \tilde{\chi}_1^0 \mu^+ \nu_\mu \,(12\%) \\ \tilde{\chi}_1^+ &\to \tilde{\chi}_1^0 \tau^+ \nu_\tau \,(12\%) \end{aligned} \tag{12}$$

Similar processes occur for $\tilde{\chi}_2^0$ decay, for example,

$$\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 Z^* \to \tilde{\chi}_1^0 \mu^+ \mu^- \,(3.3\%). \tag{13}$$

In addition, the chargino and neutralino decays can proceed via two-body modes

$$\begin{aligned} \tilde{\chi}_2^0 &\rightarrow \tilde{\chi}_1^0 Z \\ \tilde{\chi}_2^+ &\rightarrow \tilde{\chi}_1^0 W^+ \\ \tilde{\chi}_2^+ &\rightarrow \tilde{\chi}_1^+ Z \end{aligned} \tag{14}$$

Thus, the parent particles are now appropriately considered to be the W and Z bosons, since they are on-shell.

LM6 provides an example where the two-body decays

$$\begin{aligned} \tilde{\chi}_2^0 &\rightarrow \tilde{\ell}^+ \ell^- \\ \tilde{\chi}_2^0 &\rightarrow \tilde{\nu}\bar{\nu} \\ \tilde{\chi}_1^+ &\rightarrow \tilde{e}_L^+ \nu_e \\ \tilde{\chi}_1^+ &\rightarrow \tilde{\nu}_{eL} \tilde{\nu}_e \end{aligned} \tag{15}$$

and similar modes are allowed, in addition to two-body modes with W and Z bosons.

In summary, we see that leptons can be produced in the two-body decays of on-shell W and Z bosons (especially in LM2, LM4, LM5, LM7, and LM8) or in the three-body decays of $\tilde{\chi}_i^0$ and $\tilde{\chi}_j^\pm$ (LM0, LM1, LM3, LM6, LM9). In the latter cases, the leptons are produced in the decays of virtual W and Z bosons). In some models, (LM1, LM2, LM6) leptons can be produced in slepton and to a lesser extent sneutrino decays.

2.3 Kinematic distributions for SUSY models

In this section we explore some of the kinematic distributions for the SUSY benchmark models, emphasizing basic quantities used in the event selection described in Sec. 3. These distributions provide an initial picture from an experimental perspective, as well as additional insights into the phenomenology of the models.

Our goal is not to optimize the event selection separately for each benchmark model, much less for each region of parameter space. As noted in the Introduction, the event selection is motivated by broader considerations, especially robustness of reconstruction objects and reliability of the background estimates. However, it is important that the selection criteria be sensible with respect to these models, and for this reason it is important to have basic kinematic distibutions in mind. Given that significant differences in the phenomenologies of certain modesl are evident, it is possible that the RA4 selection criteria could branch into a small number of approaches with different emphases, for example, an analysis with sensitivity to very low momentum leptons.

As discussed in the Sec. 1, the different SUSY Reference Analyses are designed to have non-overlapping event samples. In practice, we keep the overlap of the samples to a minimum. In the case of RA4, we exclude events with two, well-identified leptons because these events are part of the dilepton (opposite-sign or like-sign) or triplepton analyses. This approach is intended to simplify the interpretation, both statistical and physical, of the full set of results.

Thus, for the basic muon sample, events are required to satisfy the "loose" criteria discussed in Sec. 3, which are the following:

- Muon ID Requirements: muon candidates must satisfy the requirements
  1. $p_T > 10$ GeV and
  2. the CMS "V+jets" muon cuts listed in section 3.
- Single Muon Requirement: there must be one and only one muon candidate satisfying the Muon ID cuts listed above.
- Electron Veto: the events must not have any electrons satisfying $p_T > 10$ GeV and the V+jets electron cuts listed in section 3.

In some cases, as noted, we use an electron selection:

- Electron ID Requirements: electron candidates must satisfy the requirements

1. $p_T > 10$ GeV and
2. the CMS "V+jets" electron cuts based on robust tight ID, listed in section 3.

- Single Electron Requirement: There must be one and only one electron candidate satisfying the requirements listed above.
- Muon Veto: the events must not have any muons satisfying $p_T > 10$ GeV and the V+jets muon cuts listed in section 3.

Figure 2.7 shows distributions of the number of muons in each event satisfying the Muon ID Requirements listed above, before the veto on additional leptons is applied (upper left-hand plot). The upper-right-hand plots shows the number of electrons satisfying the Electron ID Requirements listed above, before the veto on additional leptons is applied. Roughly 10% of the events contain a single reconstructed muon (ignoring additional electrons), while only about 5% of the events contain a single reconstructed electron (ignoring additional muons). For muons, we also see that the ratio of dimuon to single-muon events is around 10%.

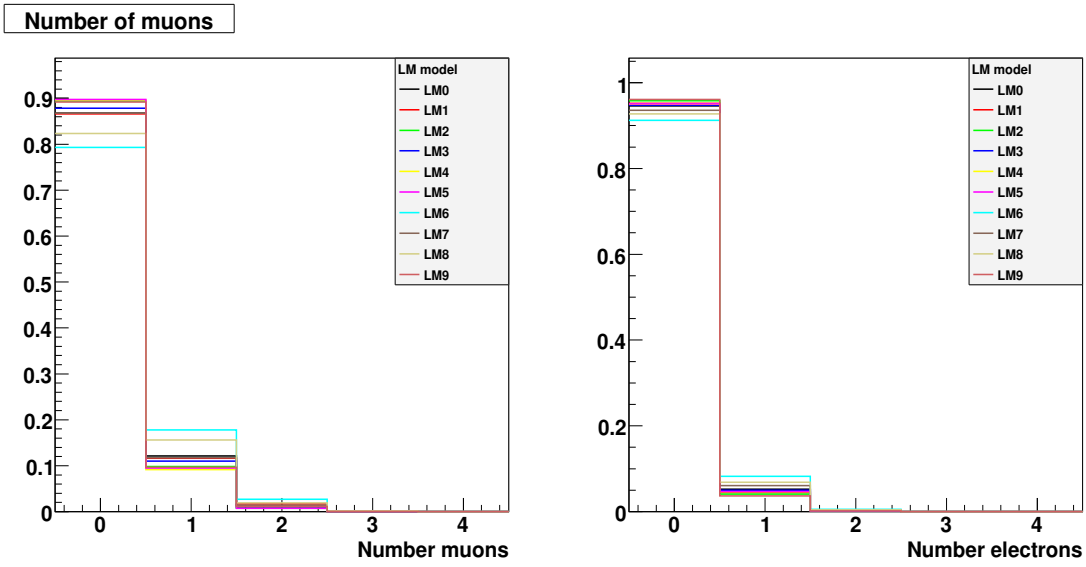


Figure 2.7: SUSY models: number of muons ($p_T > 10$ GeV and number of electrons ($p_T > 10$ GeV).

Figure 2.8 shows the momentum spectrum of the leading isolated muon (left) and the leading isolated electron (right). These plots are made with the Muon ID Requirements and the Electron ID Requirements, except the p_T cuts were removed so that the full reconstructed spectra can be seen. (Neither the Singe Muon/Electron Requirement nor the Electron/Muon Veto was applied.) We see that these observed muon spectra peak around 20 GeV for many models, while the poorer efficiency for electrons produces a spectrum that rises to a maximum that is well above 20 GeV for most models. We note that trigger requirements are not applied for these plots.

Figure 2.9 shows distributions of relative isolation for muons and electrons relative isolation. In these plots, we require all other Muon ID/Electron ID cuts including $p_T > 10$ GeV are applied, except for the isolation requirement. No Single Muon/Electron or Electron/Muon Veto cuts are applied. The relative isolation is defined by

$$\text{RelIso} = (\sum_{\Delta R<0.3} E_T^{\text{ecal}} + \sum_{\Delta R<0.3} E_T^{\text{hcal}} + \sum_{\Delta R<0.3} p_T^{\text{trk}})/p_T(\mu). \quad (16)$$

In other words, the RelIso quantity represents the scalar sum of calorimeter and tracker transverse energies within a cone of $\Delta R < 0.3$ around the muon, divided by the muon transverse momenta. The definition above omits one detail: the muon momentum itself, as well as its energy deposits in the calorimeters, are excluded from the sum. For both muons and electrons, we see clear peaks at low values of relative isolation in SUSY events. Although there is some degree of model dependence, isolated leptons dominate the distributions.

We next characterize the jet activity in SUSY events. We consider Calo jets stored in the allLayer1JetsIC5 collection (class pat::Jet), which are formed with the Iterative Cone (IC5) jet-clustering algorithm. We require the jets to

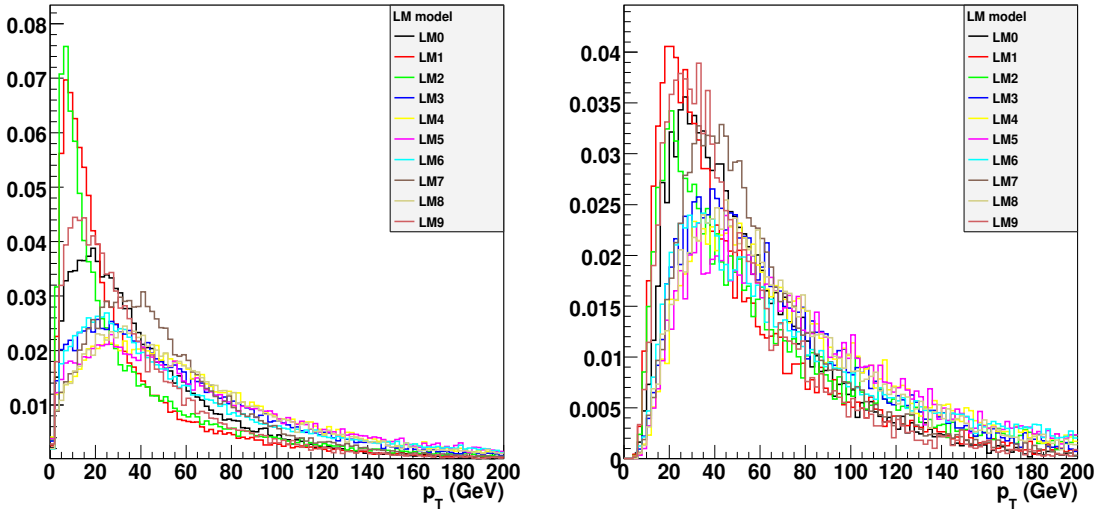


Figure 2.8: SUSY models: muon p_T spectra (left) and electron p_T spectra (right).

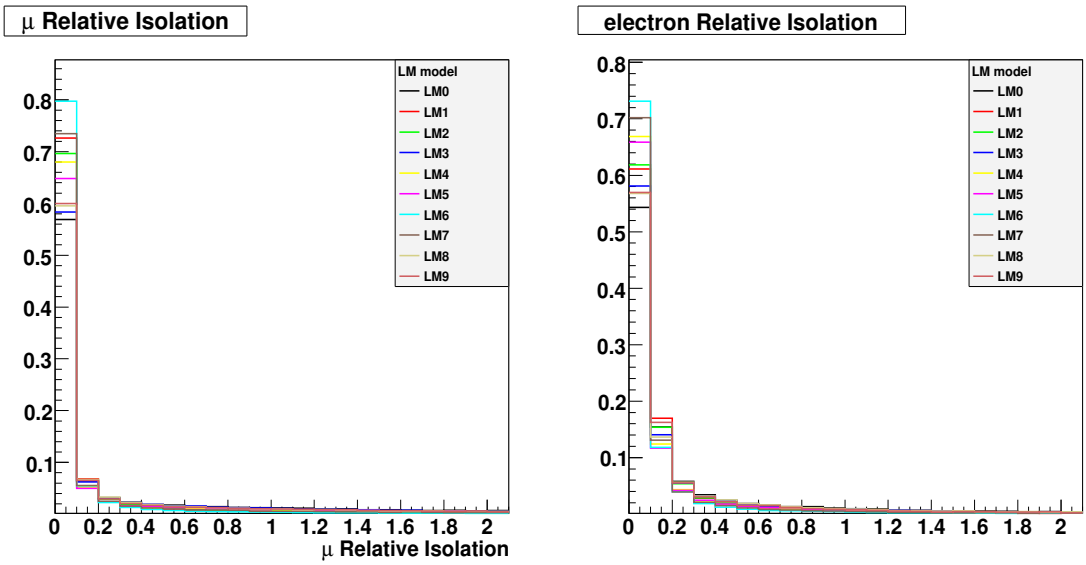


Figure 2.9: SUSY models: relative isolation distributions for muons and electrons.

satisfy $|\eta| \leq 2.5$ and to have electromagnetic energy fraction (EMF) less than 0.1. Figure 2.10 shows distributions of the numbers of jets, where a single muon is required (the vetos are applied on additional muons and electrons). The jet counting is performed for two different thresholds $p_T = 5$ GeV and $p_T = 100$ GeV. For most of the models, a requirement that there be at least one jet above 100 GeV and two additional jets above 50 GeV is reasonably efficient, with LM7 and LM9 representing special cases, as usual.

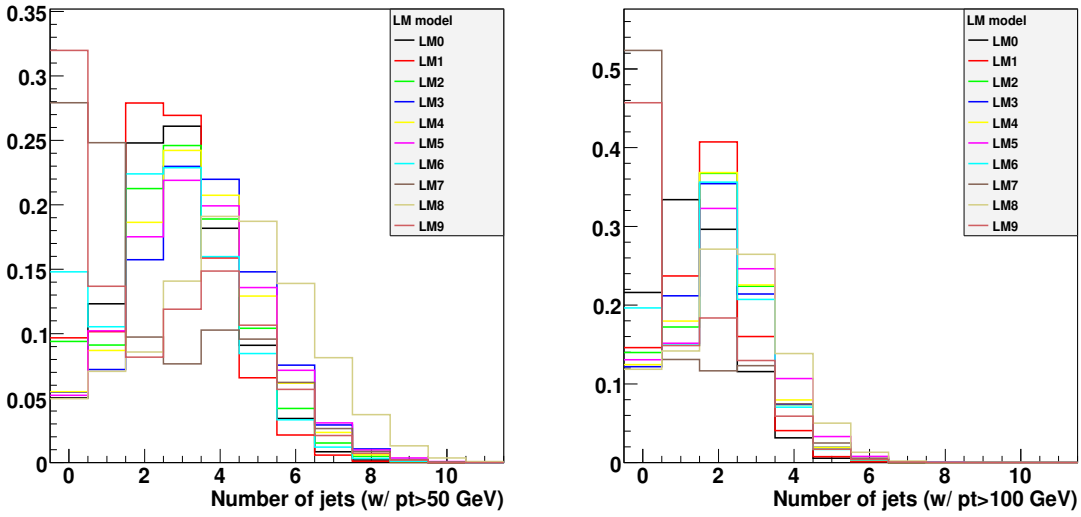


Figure 2.10: SUSY models: number of jets with $p_T > 50$ GeV (left) and number of jets with $p_T > 100$ GeV) (right), after a single isolated muon is required.

Figure 2.11 shows the reconstructed (L2L3 corrected) p_T distributions for each of the four highest p_T jets. In each of these plots, the distributions have been normalized to unity, so we are only comparing the shapes, not cross sections.

For the highest p_T jet, we see that each benchmark model has a broad component in the distribution with a peak in the range 100–350 GeV. Some of the models, however, also have prominent peak below 100 GeV, which is associated with electroweak production (see Eq. 2). In such events, there is no high p_T jet. This peak is particularly strong in models LM9, LM7, and LM6, with more modest effects in LM1 and LM2. LM0, a low-mass model with a large cross section, does not display a low-energy peak in this single-muon sample. From these distibutions, it is clear that in order for a generic jet p_T cut to be reasonably efficient on the upper (broad) peak, its value should be not higher than about 100 GeV. However, in certain models, a higher cut would still be efficient. If one were to observe an excess of events in the data beyond the SM expectation, a plot of the highest jet p_T distribution would be of great interest.

Model-by-model, the p_T distributions for the second highest p_T jet are similar to those for the highest p_T jet, but are softer. Note that the scale on this plot extends only to 500 GeV, whereas that for the highest p_T jet extends to 1 TeV. A cut at 100 GeV on the second highest p_T jet would result in a significant inefficiency in many models. The p_T distributions for the third and fourth highest p_T jets (plotted on scales extending only to 250 GeV), show that a threshold cut at 100 GeV on these jets would be quite inefficient.

We see that the jet p_T spectra across a broad range of models are not particularly hard. Because jets with p_T values below 30 GeV–50 GeV are often not physically meaningful, one can imagine the several approaches to setting the jet thresholds in SUSY events, such as

1. A staggered approach, such as $p_T^{J1} > 100$ GeV, $p_T^{J2} > 80$, $p_T^{J3} > 40$ GeV, and $p_T J4 > 40$ GeV.
2. A loose requirement that there be at least four jets above some low threshold, such as 50 GeV.
3. An approach with just one explicit jet p_T threshold, say $p_T^{J1} > 100$ and a requirement on the sum of the p_T values of the remaining jets be above some threshold, *e.g.*, $\sum_{i>1} p_T^{Ji} > 150 GeV$. This approach does not require more than two jets to be present.

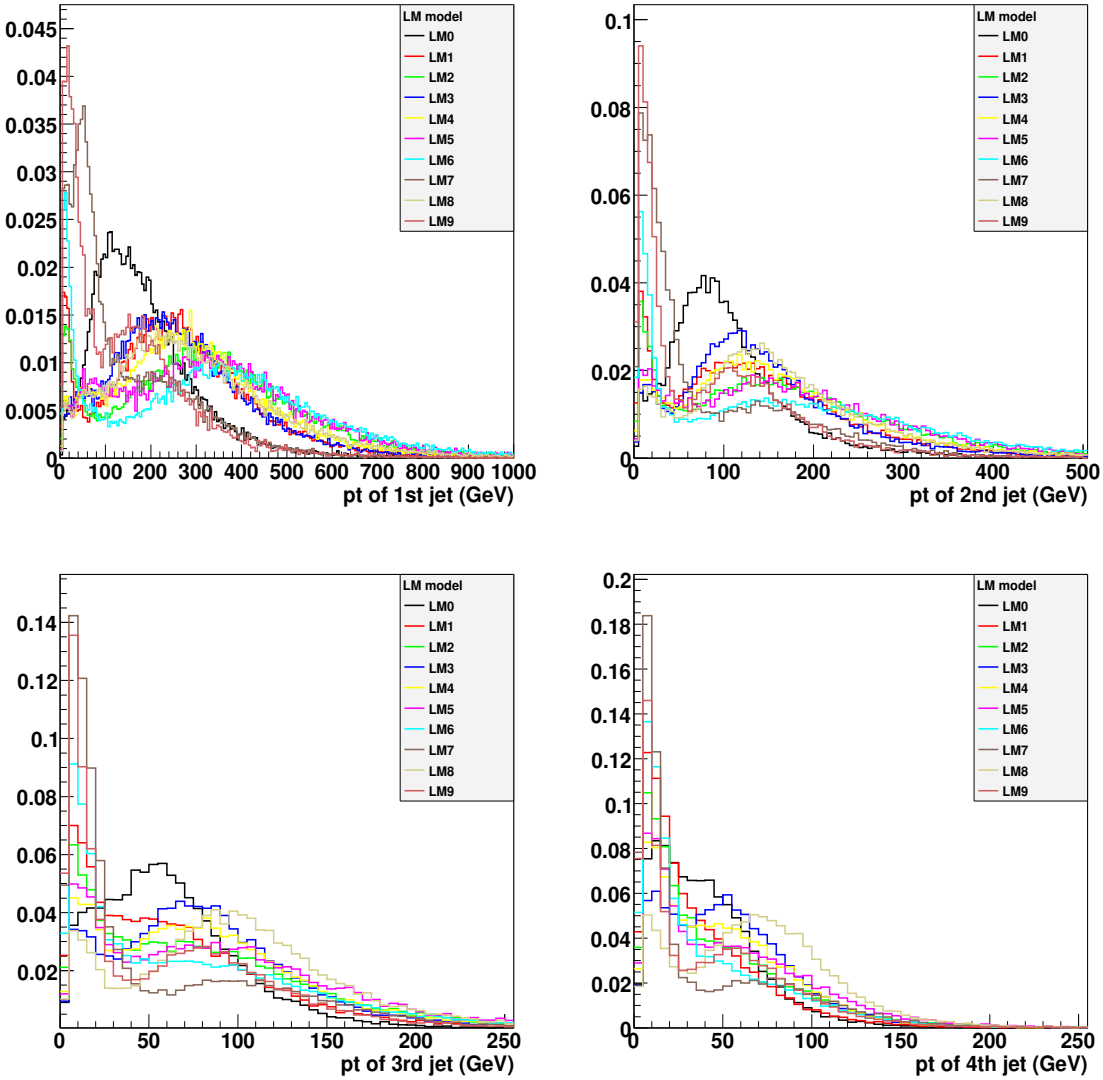


Figure 2.11: SUSY models: jet p_T spectra for each of the four highest p_T jets.

Figure 2.12 shows the distribution of several quantities that characterize the energy scales of the models. The MET ("missing transverse energy") is a complicated quantity intended to estimate the magnitude of the momentum carried by unobserved particles in the plane transverse to the beam direction. At the simplest level, it is based on a vector sum over the energy deposits in approxiately projective calorimeter towers. The energy is computed as an unweighted sum over the electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL) contributions in the tower. (In the barrel region, a single HCAL tower is associated with a 5×5 array of ECAL crystals, whereas in the endcap, a more complicated mapping is used to form the combined towers.)

$$\mathbf{p}_T^{\text{miss}} = - \sum_{n=\text{calo towers}} (E_n \sin\theta_n \cos\phi_n \hat{x} + E_n \sin\theta_n \sin\phi_n \hat{y}). \tag{17}$$

The simplest MET is the magnitude of this vector quantity:

$$\text{MET} = |\mathbf{p}_T^{\text{miss}}|. \tag{18}$$

A number of corrections are usually applied to the simple calorimeter-based MET. Jet Energy Scale (JES) corrections can be applied by first clustering the calorimeter towers into jets according to some algorithm. The jets formed in this way can be subjected to JES corrections (the unclustered energy is not corrected). For the plot shown in Fig. 2.12, the L2L3 jet corrections were applied to the Iterative Cone $\Delta R = 0.5$ (IC5) jets that constitute the clustered energy component. The MET shown here was obtained from collection allLayer1METsIC5 (class pat::MET). It uses the barrel (EB) and endcap (EE) electromagnetic calorimeter and the barrel (HB), endcap (HE), and forward (HF) parts of the hadronic calorimeter. The HO is not used.

A second correction is to account for muons, whose energy deposit in the calorimeter (typically a few GeV) does not accurately represent the muon momentum. To apply such a correction, muons must first be identified according to well defined criteria. The muon energy deposits in the calorimeter are then removed from the MET calculation and replaced by the true muon $\mathbf{p}_T$:

$$\mathbf{p}_T^{\text{miss}} = - \left(\sum_{n=\text{calo towers}} \mathbf{p}_T - \sum_{\text{muons}} \mathbf{p}_T^{\text{deposit towers}} + \sum_{\text{muons}} \mathbf{p}_T^{\text{tracker}} \right). \tag{19}$$

In the calculation used, the selected muons must have $p_T > 10$ GeV, $|\eta| < 2.5$, and must satisfy track-quality criteria.

The MET distributions in Fig. 2.12 typically show broad peaks with long tails extending to well above 100 GeV. Most of the distributions have peaks in the 100 GeV – 200 GeV range; the distributions for LM7 and LM9, however, peak well below 100 GeV and fall rapidly with MET. The LM0 MET distribution peaks somewhat below 100 GeV but has a substantial tail extending to over 300 GeV. In Sec. 4, we compare the signal and background MET distributions.

The quantity M_{eff} is the scalar sum of the p_T values for the four highest p_T jets and the p_T of the single lepton with $p_T > 10$ GeV in the event, together with the MET.

$$M_{\text{eff}} = \sum_{i=1,2,3,4} p_T^{Ji} + p_T^{\mu} + \text{MET}. \tag{20}$$

This quantity provides a measure of the overall energy scale of the event. The M_{eff} distributions typically show broad peaks with maxima in the range 600 GeV to 1.2 TeV. The distribution for LM0 peaks around 500 GeV, while LM7 and LM9 show additional lower peaks below 200 GeV.

The quantity H_T is simply the scalar sum of the p_T of each of the four highest p_T jets and the single muon. The MET is not included.

The quantity M_T is a measure of the invariant mass of the lepton and the missing momentum, computed using transverse quantities only. If the event contains a single $W \to \mu\bar{\nu}$ decay, then $p(W) = p(\mu) + p(\nu)$ (four-vector equation), and the quantity

$$p(W)^2 = (p(\mu) + p(\nu))^2 = \big(m(\mu)^2 + 2(E(\mu)E(\nu) - (p_x(\mu)p_x(\nu) + p_y(\mu)p_y(\nu) + p_z(\mu)_z(\nu)\big) \tag{21}$$

would correspond to M_W^2. The quantity M_T (which lacks a helpful label indicating which particles are included!) is defined as

$$M_T = \sqrt{2(p_T(\mu)E_T^{\text{miss}} - p_x(\mu)p_x(\nu) + p_y(\mu)p_y(\nu))}. \tag{22}$$

For a single $W \to \mu\bar{\nu}$ decays M_T shows a peak with a sharp falling edge close to the W mass. If a W decay were the source of both the muon and the MET, the requirement $M_T > M_W$ would remove most of the SM events. However, we can see from Fig. 2.12 that such a cut removes a significant fraction of the SUSY events as well. We will return to this point in the section on backgrounds (Sec. 4).

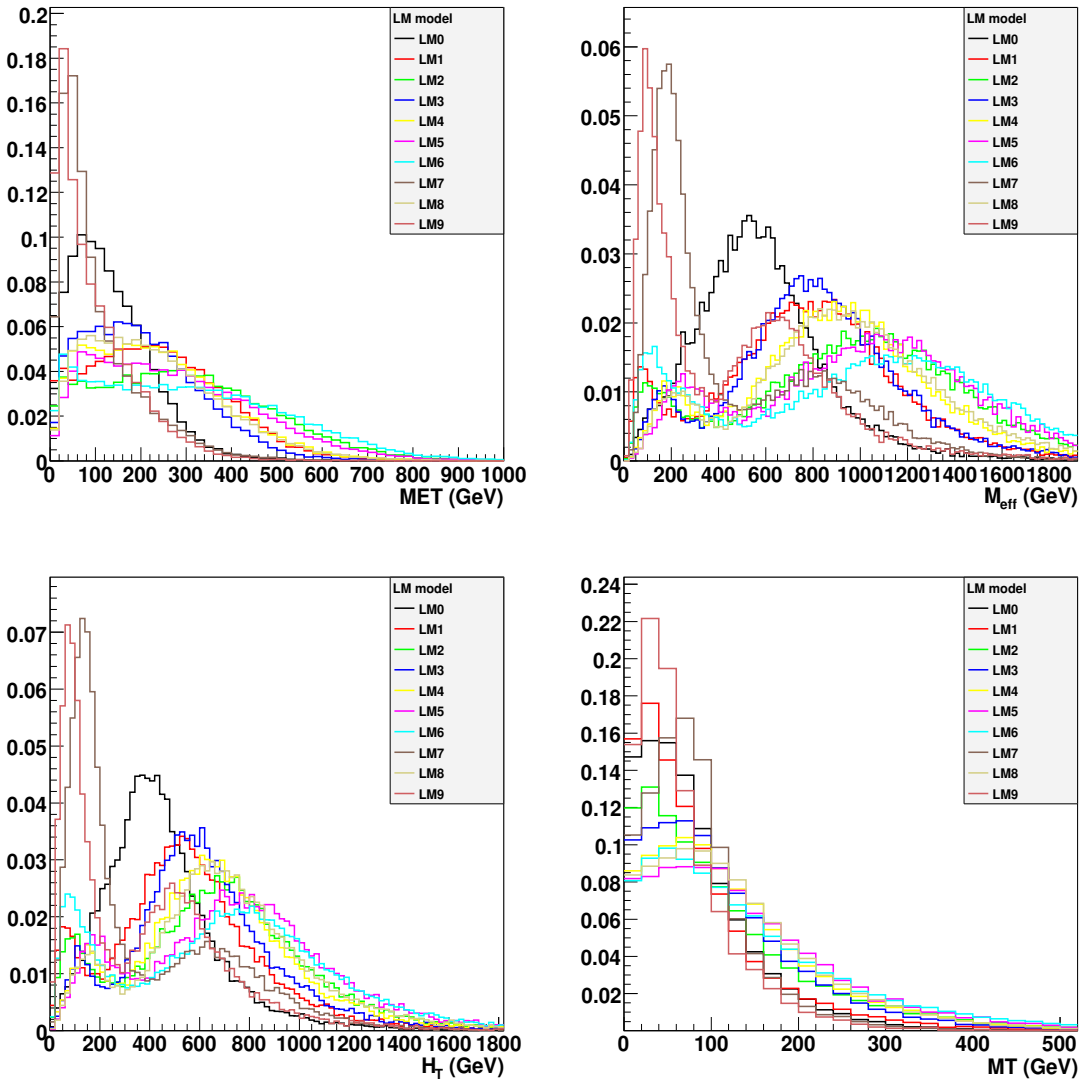


Figure 2.12: SUSY models: Distributions of MET, M_{eff}, H_T, and M_T.

2.4 Scan of SUSY model behavior across mSUGRA parameter space

To quantify SUSY yields more generally than is possible with the benchmark points discussed in the previous section, we scanned the mSUGRA parameter space at center-of-mass energies of 7 and 10 TeV. Since it is the $m_{1/2}$ and m_0 mass parameters that are primarily responsible for determining the production cross-sections, we simulated 323 samples (19 m_0 points in intervals of $\Delta m_0 = 50$ GeV and $\times$ 17 $m_{1/2}$ points in intervals of $\Delta M_{1/2} = 25$ GeV) at each center-of-mass energy. The other mSUGRA parameters were set as $\tan(\beta) = 10.0$, $A = 0$, and $\mu > 0$. Each sample contained 16000 events and was processed with FastSim in CMSSW_2_2_9.

Figure 2.13 shows the cross-section in fb as a function of $m_{1/2}$ and m_0 for the two center-of-mass energies, 7 TeV (left) and 10 TeV (right). As seen from Table 2.1, the cross sections for the LM points range from about 1–100 pb. Since the units in these plots are fb, a 10 pb cross section corresponds to 10^4 in the scale (green). In a sample with integrated luminosity of 100 pb^{-1}, 1000 events total (all processes, all decay modes) would be produced.

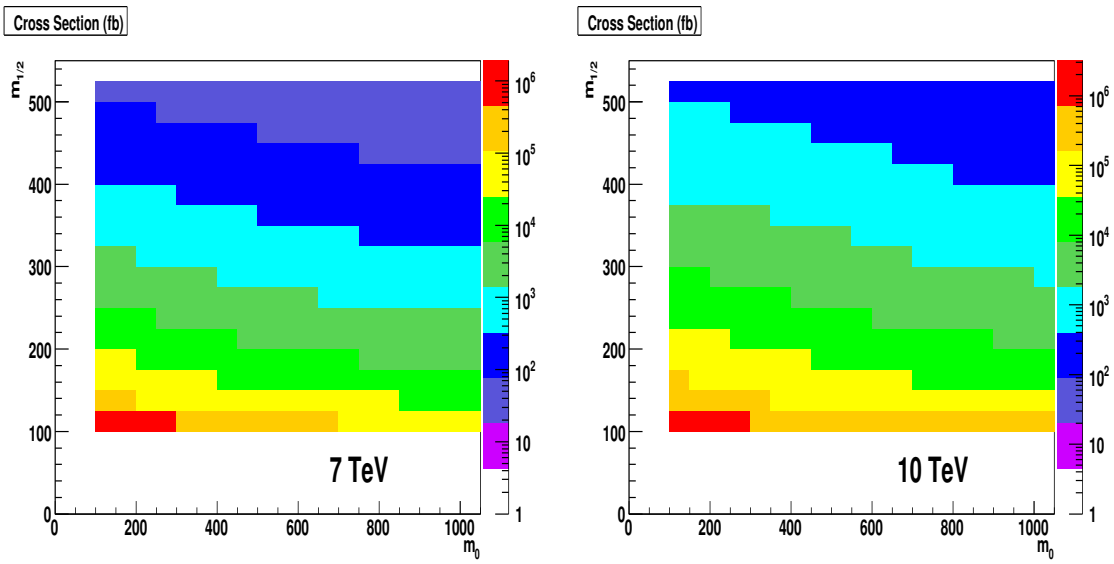


Figure 2.13: mSUGRA scan: total cross-section (fb) for $\sqrt{s}$ = 7 TeV (left) and 10 TeV (right) as a function of the parameters $m_{1/2}$ and m_0, with $\tan(\beta) = 10.0$, $A = 0$, and $\mu > 0$.

We next consider the yield of events passing the standard cuts defined in section 3. Figure 2.14 shows a scan of the number of events with one loosely selected muon ($p_T > 10$ GeV), MET > 200, and passing either the loose (left-hand plot) or tight (right-hand plot) jet selection. While these cuts are not necessarily precisely what one would use in an actual analysis, they provide a good indication of the event yield, since they use representative values for the three main ingredients: muon ID and momentum cut, jet selection, and a MET requirement.

Figure 2.15 shows the events yield for a tightly selected muon ($p_T > 20$ GeV) and either loose (left-hand plot) or tight (right-hand plot) jet selection requirements.

3 Event Selection

This section describes basic event selection requirements for the RA4 analysis. Because the background determination methods described later have potentially different requirements, it is not possible to define a unique selection. For example, some approaches may need a loose selection to obtain a sufficiently large control sample of background. Others may need a tighter selection to suppress QCD and W+jets contributions and obtain a $t\bar{t}$ dominated control sample.

Wherever possible, howevever, it is desirable to use a common selection to facilitate comparison of the different methods. We discuss here a set of standard selections that can be referenced in other sections. These should be applicable to most of the $t\bar{t}$ background prediction methods, but they will not be relevant for all analyses. Noteworthy exceptions are the low $N_{\rm jet}$ searches of sections 14 and 15, for which W+jets and QCD backgrounds will dominate over $t\bar{t}$, and the low p_T lepton searches of section 7, for which QCD backgrounds will dominate.

To cover the expected range of selection requirements, we define loose and tight cuts for muons as well as loose and tight cuts for jets. These are intended as signal selection cuts. Additionally, we describe a set of base jet cuts intended to select a large, background-dominated control sample. Pairing the different muon and jet cuts leads to several combinations of event selections, at least one of which should be appropriate for most background determination methods.

3.1 Muon selection

Carsten Magass

The basic muon identification cuts, which follow the V+jets group recommendations [**?**], are listed below.

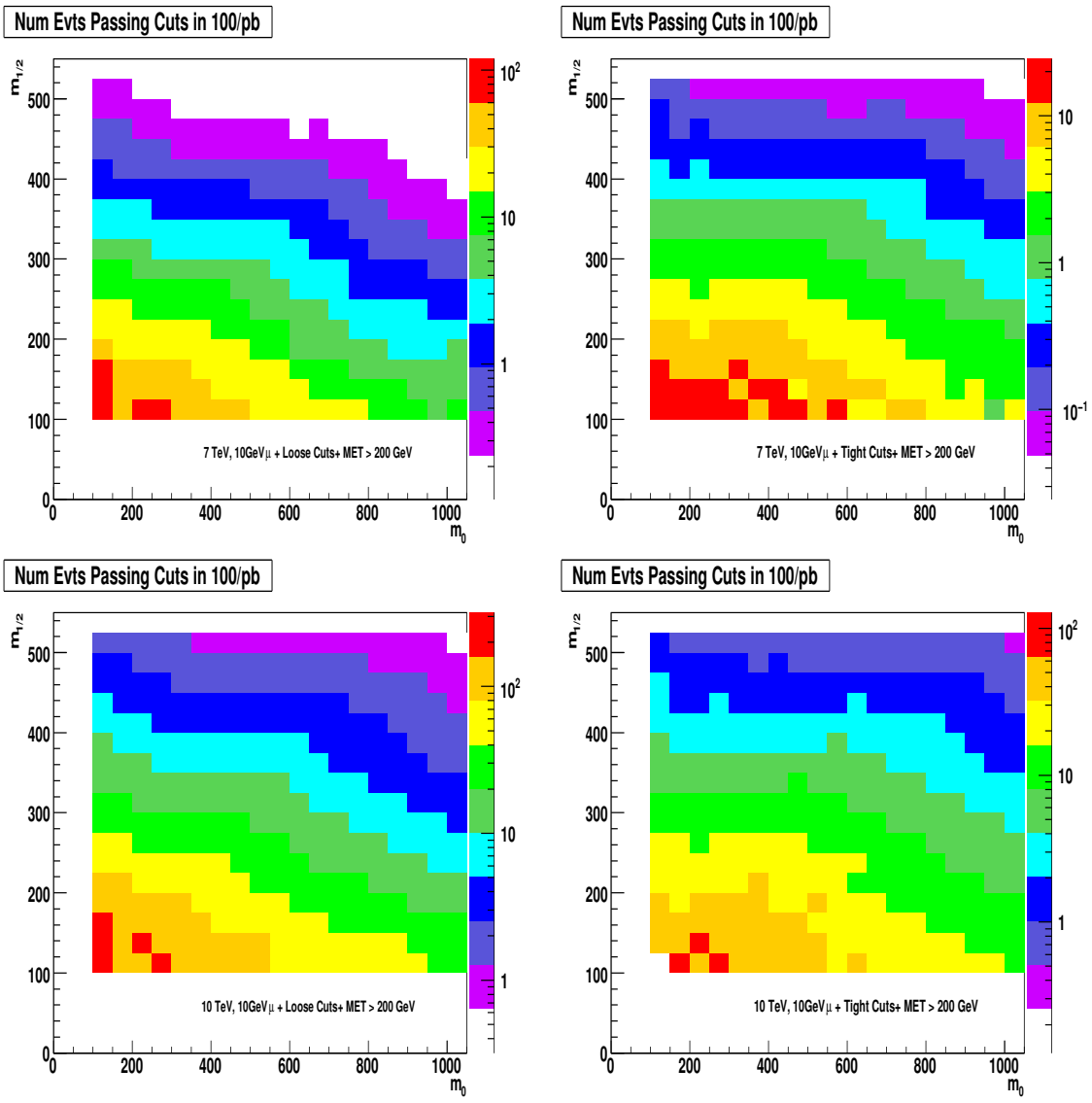


Figure 2.14: mSUGRA scan: number of events passing cuts requiring one loosely selected muon ($p_T > 10$ GeV), MET > 200, and either the loose or tight jet selection. The mSUGRA scan is performed as a function of $m_{1/2}$ and m_0 with $\tan(\beta) = 10.0$, $A = 0$, and $\mu > 0$.

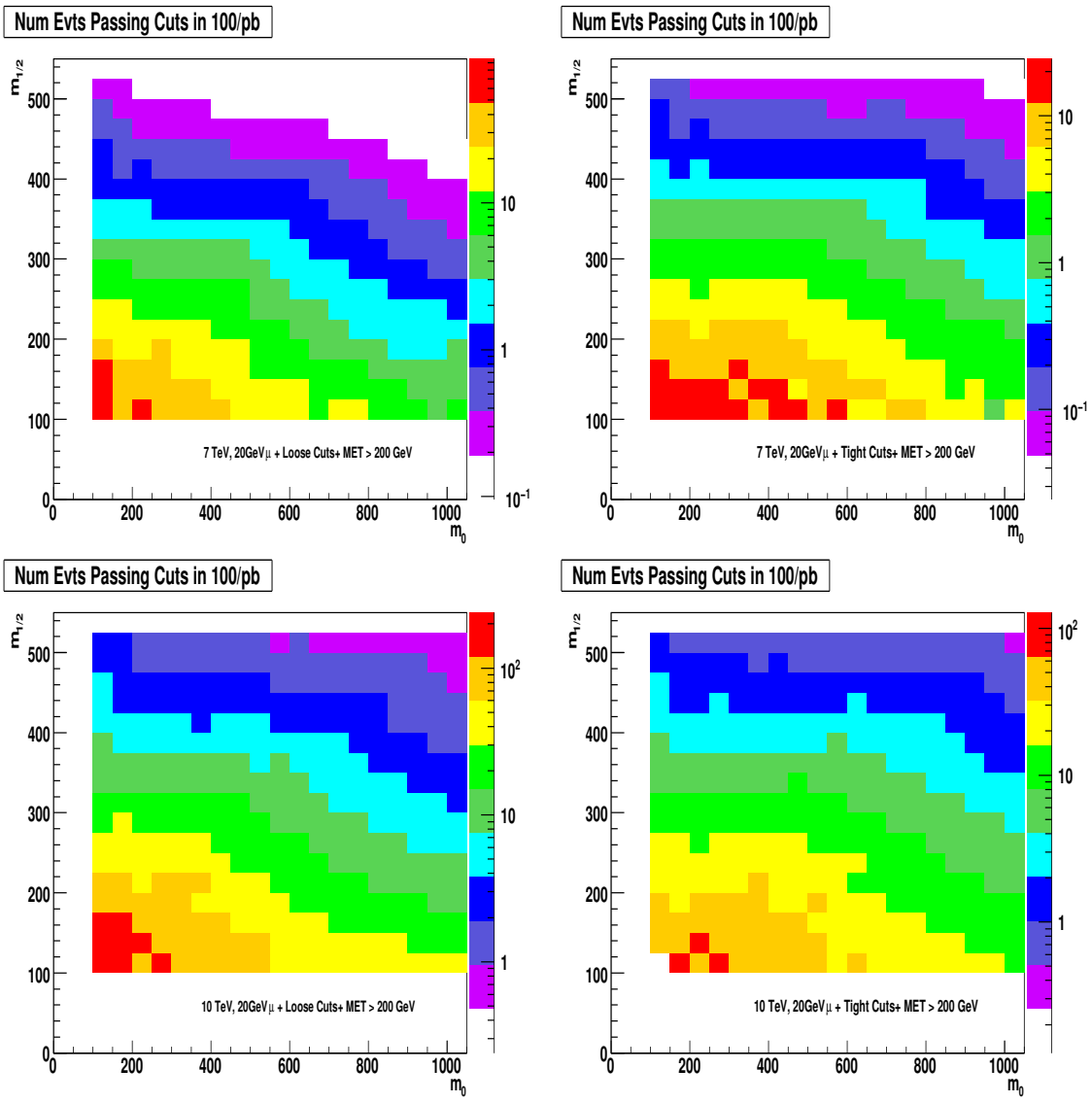


Figure 2.15: mSUGRA scan: number of events passing cuts requiring one tightly selected muon ($p_T > 20$ GeV), MET > 200, and either the loose or tight jet selection. The mSUGRA scan is performed as a function of $m_{1/2}$ and m_0 with $\tan(\beta) = 10.0$, $A = 0$, and $\mu > 0$.

Loose Muons	**Tight Muons**
• $p_T \geq 10$ GeV/c	• $p_T \geq 20$ GeV/c
• $\lvert\eta\rvert \leq 2.1$	• $\lvert\eta\rvert \leq 2.1$
• $\sum \mathrm{RelIso} < 0.1$	• $\sum \mathrm{RelIso} < 0.1$
• $\mathrm{ID} = \mathrm{GlobalMuonPromptTight}$	• $\mathrm{ID} = \mathrm{GlobalMuonPromptTight}$
• $\mathrm{NumValHits} \geq 11$	• $\mathrm{NumValHits} \geq 11$
• $\chi^2/\mathrm{ndof} < 10.0$	• $\chi^2/\mathrm{ndof} < 10.0$
• $\lvert d_0\rvert < 0.2$ cm	• $\lvert d_0\rvert < 0.2$ cm
• HCAL Veto Cone Energy < 6 GeV	• HCAL Veto Cone Energy < 6 GeV
• ECAL Veto Cone Energy < 4 GeV	• ECAL Veto Cone Energy < 4 GeV

The higher p_T cut in the tight category sacrifices the potentially large component of low p_T muons from SUSY decays for a simplified background composition.

Since the distribution of isolation energy is a useful tool for quantifying the QCD background, the isolation cuts are not intended to be used for the initial sample selection. Furthermore, the choice of final isolation cut remains under discussion due to significant inefficiencies for low p_T leptons, as described below.

3.1.1 Study of a muon ID selection with an absolute isolation requirement

The muon selection cuts defined in Sec. 3.1 may not be optimal for a SUSY search since

- the cuts have been developed for a clean environment in V+jets events and
- the muons in V+jets events tend to have higher transverse momenta than muons expected in SUSY events.

Because of this, detailed studies were performed [1] analyzing the isolation cuts. The following improved muon selection cuts are proposed:

- $p_T \geq 10$ GeV/c
- $\lvert\eta\rvert \leq 2.1$
- $\mathrm{ID} = \mathrm{GlobalMuonPromptTight}$
- $\mathrm{NumValHits} \geq 11$
- $\chi^2/\mathrm{ndof} < 10.0$
- $\lvert d_0\rvert < 0.2$ cm
- **TrkIso < 6 GeV**
- **ECalIso < 6 GeV**
- **HCalIso < 6 GeV**

In this selection the cut on the relative isolation is omitted and replaced by cuts on the individual isolation variables, namely the isolation in the tracker (TrkIso), in the electromagnetic (ECalIso) and hadronic calorimeter (HCalIso). The plots in Fig. 3.16 show the sum of all isolation quantities (SumIso = TrkIso + HCalIso + ECalIso) as a function of the muon momentum for the W+jets and the LM0 sample. The plots with red [blue] boxes display the distributions for prompt [non-prompt] muons. The cut $\sum \mathrm{RelIso} < 0.1$ is indicated by the green line. One can see that the current selection clearly removes (unwanted) non-prompt muons; but on the other hand many prompt muons are cut away – especially at lower muon momenta. This leads to a non-constant muon identification efficiency, which decreases at lower muon momenta as shown in Fig. 3.17.

The main advantages of the proposed muon selection are the flat and high efficiency over the entire muon momentum range and the ability to *individually* handle the isolation requirements. This might be useful in the case that problematic features like noise occur in parts of the detector (e. g. in the hadronic calorimeter) during data taking. Further, the selection relies on simpler cut definitions since the veto cuts are replaced by 'real' isolation requirements. A comparison of the performance, quantified here via the purity and the fake rate, is listed in Tab. 3.5. Summarizing, the proposed muon selection exhibits a performance comparable to the standard selection, but with a significantly improved signal reconstruction efficiency at lower muon momenta.

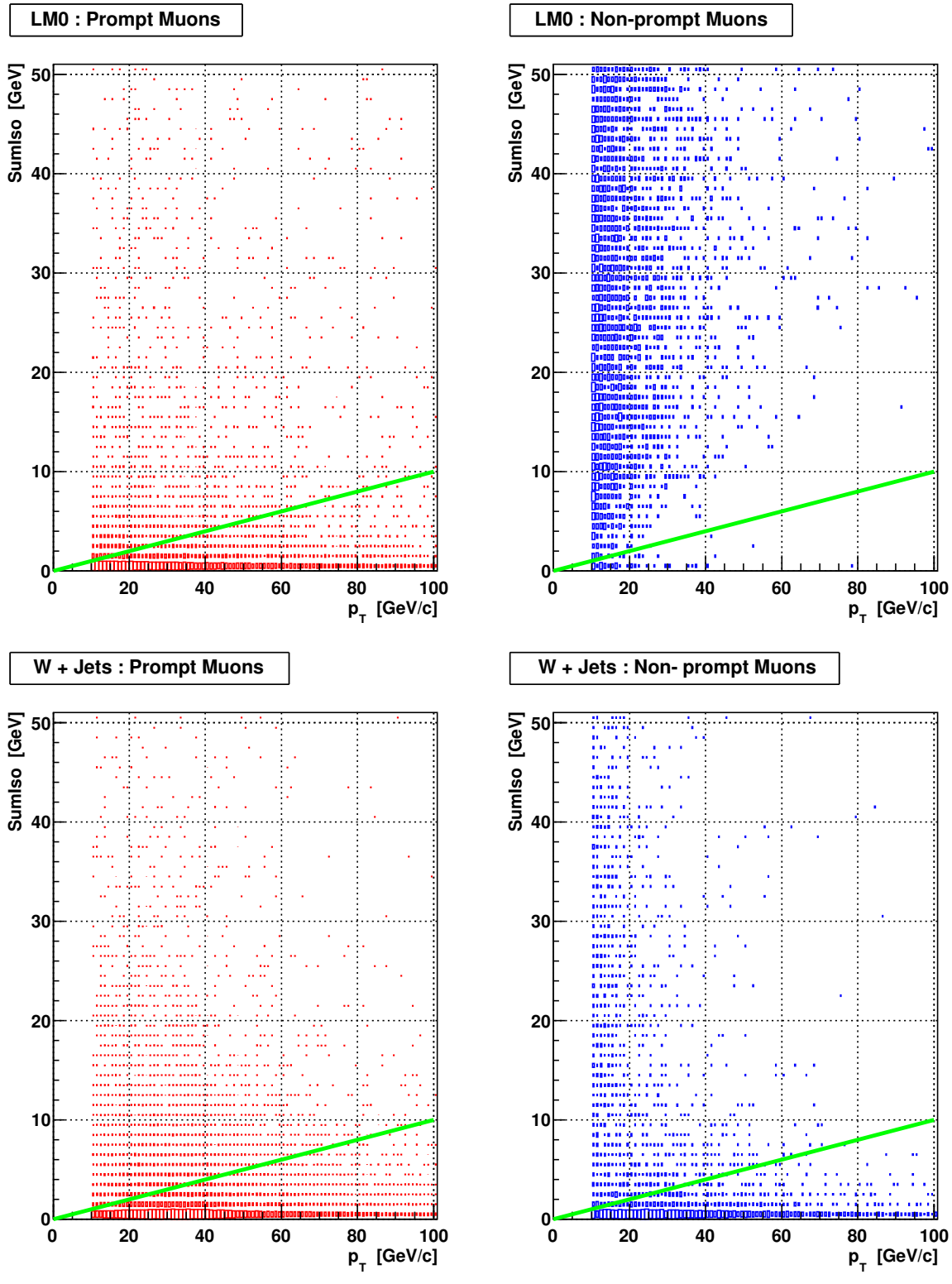


Figure 3.16: Overall isolation (SumIso) as a function of the muon transverse momentum for different samples and types of muons. The green line indicates the cut on the relative isolation of 0.1.

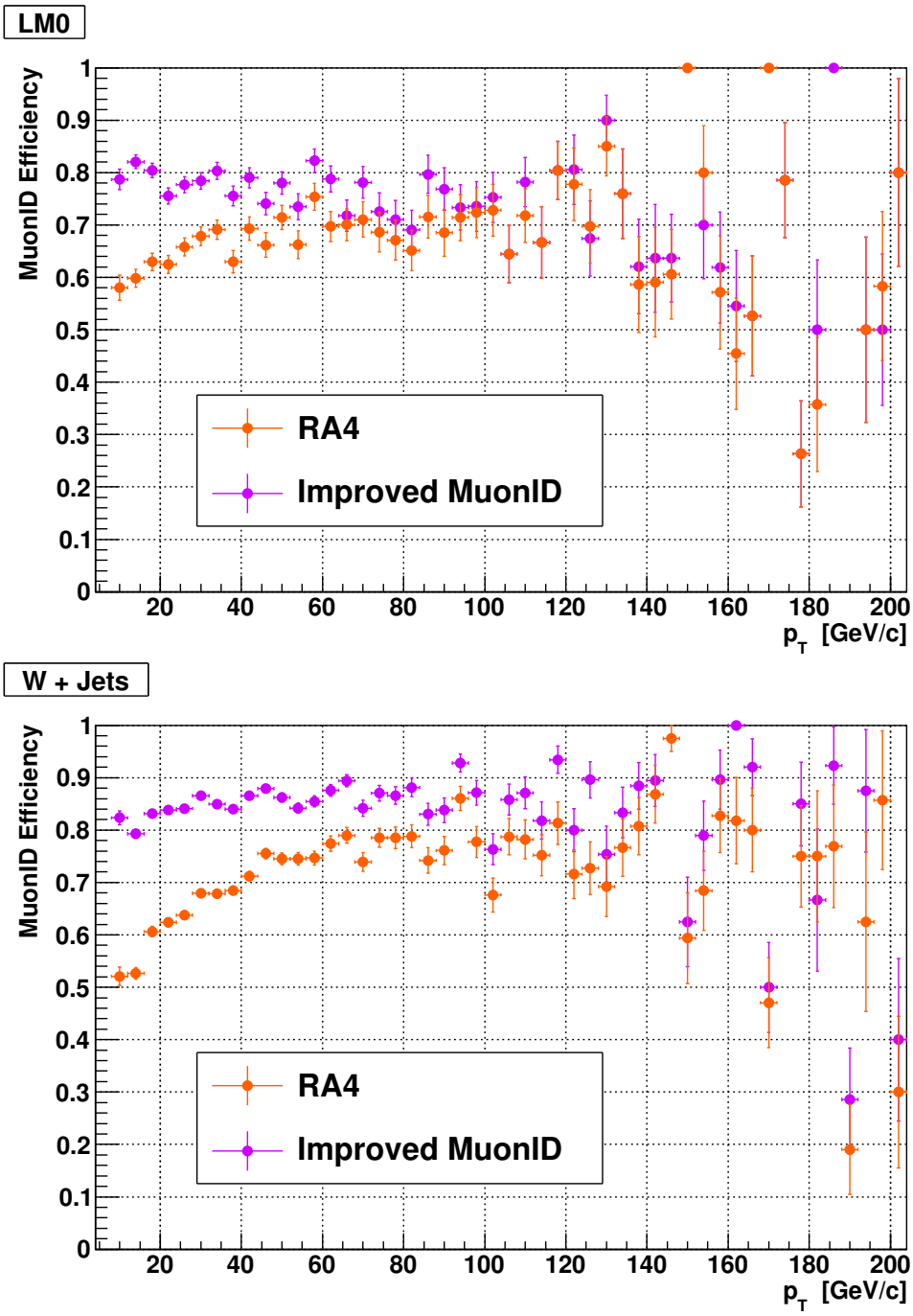


Figure 3.17: Muon reconstruction efficiency as a function of the muon transverse momentum for the current and the improved muon identification derived from the W+Jets and the LM0 sample.

Table 3.5: Performance of the current and the proposed improved muon selection (± 0.1 % statistical uncertainty).

	RA4		**Improved MuonID**	
	W + Jets	LM1	W + Jets	LM1
Purity (%)	94.2	99.4	94.1	98.3
Fake Rate (%)	5.9	0.6	5.8	1.7

3.1.2 Muon trigger

Daniel Teyssier

The online selection should use trigger bits adapted to each particular selection. Both Jet/MET or leptonic trigger bits could lead to select efficiently the data, depending on the cuts used. The preselection defined in Sec. 3 had been applied, and the relative efficiency calculated for the main Jet/MET and muonic trigger bits. The results using both the V+jets muon ID and the modified muon ID proposed in Sec. 3.1.1 are shown in Fig. 3.18.

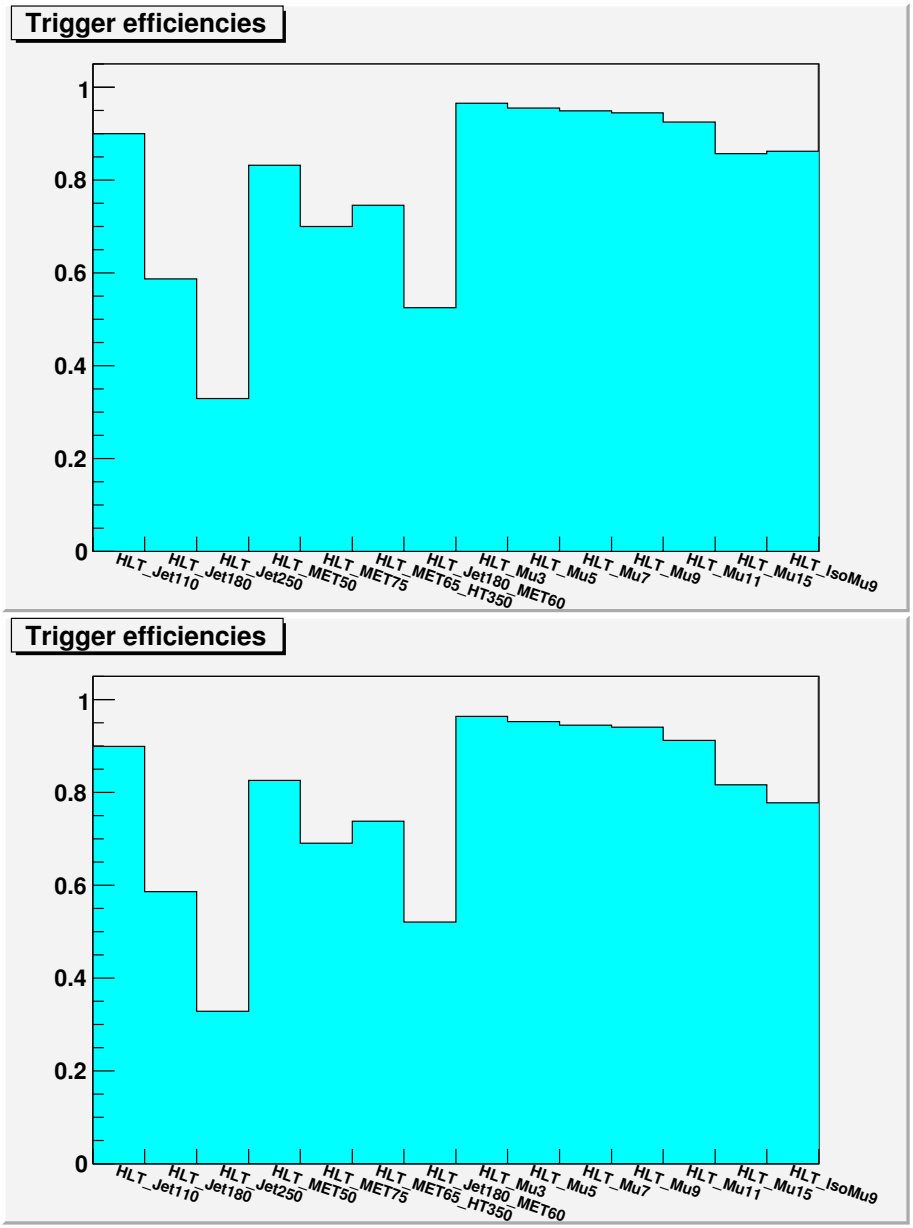


Figure 3.18: On the bottom, the relative trigger bits efficiencies are shown in the case of the V+jets muon ID selection, while on the top the modified muon ID was used.

The single muonic triggers, up to the HLT_mu9 bit, reach at least a 95% relative efficiency after the preselection, using both muon ID described previously. The apparent lower efficiencies concerning the Jet/MET trigger bits reflect the low cuts on the pt of the jets and the absence of MET cut at the preselection level.

Using a stronger selection on the jets momentum and the MET, the results are shown in 3.19. Then the Jet/MET triggers could reach relative efficiencies close to 100%, while the muonic triggers stay at the same level. This result is independant from the muon ID definition.

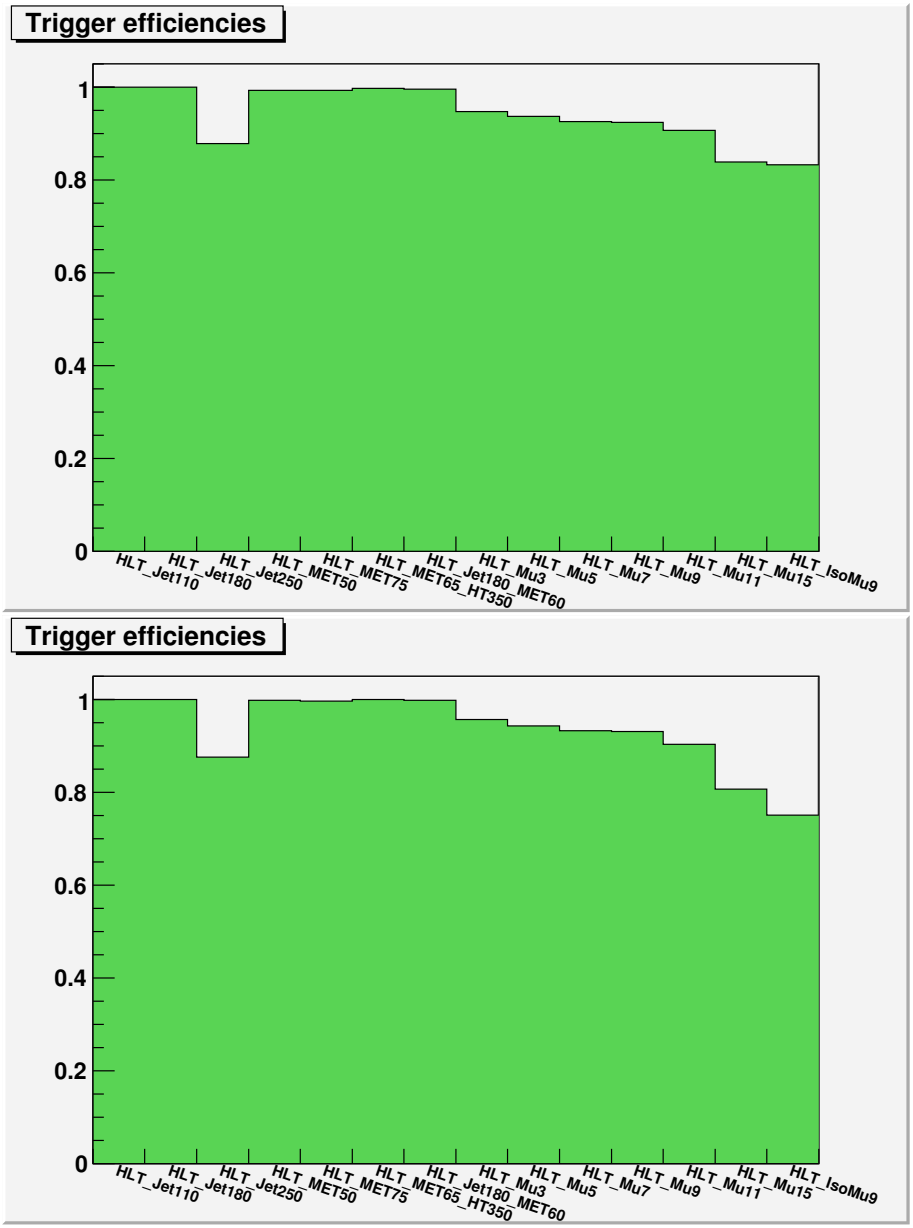


Figure 3.19: On the bottom, the relative trigger bits efficiencies are shown in the case of the V+jets muon ID selection, adding some cuts on jets momentum and MET, while on the top the modified muon ID was used.

3.2 Electron selection

Requirements are defined separately to select and to veto electrons. The veto is used for single muon event selection to maintain sample independence from dilepton searches.

The electron selection cuts are:

- GSF Electron
- $p_{\mathrm{T}} > 20$ GeV
- $|\eta| < 2.5$, and η not in ECAL gap ($1.47 < |\eta| < 1.567$)
- "robustLoose" identification [2]
- $\mathrm{RelIso} \equiv \frac{1}{p_T^{\mathrm{ele}}} \sum^{\mathrm{iso\ dep}} \left(E_T^{ECAL} + E_T^{HCAL} + p_T^{Track}\right) < 0.1$

The electron veto cuts are:

- $p_T \geq 10$ GeV/c
- $|\eta| \leq 2.0$
- $\sum \mathrm{RelIso} < 0.1$
- ElectronID == RobustTight
- NumValidHits ≥ 11
- $\chi^2/\mathrm{ndof} < 10.0$

3.3 Jet selection

The individual jet identification requirements are listed below.

- $p_T \geq 30$ GeV/c
- $|\eta| \leq 2.4$
- Hadronic Energy Fraction ≥ 0.1

Jets within $\Delta R \leq 0.4$ of an identified lepton are not counted. While this will have little effect on the final sample due to the lepton isolation requirements, it can substantially affect the jet counting in control samples that use non-isolated leptons.

The requirement that the hadronic energy fraction in jets exceeds 10% is of little relevance to Monte Carlo studies, but it will likely be important to avoid noise effects in data. The $p_T \geq 30$ requirement is quite loose and is tightened to 50 for most uses, except, e.g., for very loose jet counting.

As can be seen in Fig. 4.21, the W+jet and QCD processes dominate events with low N_{jet}, while the $t\bar{t}$ process becomes dominant after requiring three jets of $p_T > 50$ GeV. For the methods designed to predict $t\bar{t}$, it is important to suppress W+jets and QCD contributions to the control regions. To that end, we define three jet based event requirements.

- Base Jet Cuts: ≥ 3 jets with $p_T \geq 50$ GeV.
- Loose Jet Cuts: ≥ 3 jets with $p_T \geq 50$ GeV, at least one of which has $p_T \geq 75$ GeV.
- Tight Jet Cuts: ≥ 3 jets with $p_T \geq 75$ GeV and a fourth jet with $p_T \geq 50$ GeV.

The first provides some supression of non-$t\bar{t}$ backgrounds but leaves enough background for control sample measurements. The others provide loose and tight rejection of $t\bar{t}$, i.e., they are control regions with approximately signal-like kinematics, from which to measure the $t\bar{t}$ contribution to the final, high missing transverse energy signal region.

Sample	σ (pb)	$\geq$ 1 mu	dilepton veto	Base Jet Cuts	Loose Jet Cuts	Tight Jet Cuts	MET$\geq$200 GeV (Loose Jet Cuts)	MET$\geq$200 GeV (Tight Jet Cuts)
LM1	16.06	215.5	179.9	94.0	93.8	37.6	60.9	24.5
ttbar	414	6501.0	5937.4	1946.2	1794.0	311.9	40.0	14.2
W+jets	40000	$7.00 * 10^5$	$7.00 * 10^5$	788.1	731.9	60.2	19.8	2.2
Z+jets	3700	$9.77 * 10^4$	$6.08 * 10^4$	90.3	84.6	7.0	0	0
QCD (Inmu)	—	$7.49 * 10^4$	$7.48 * 10^4$	905.7	692.1	3.2	1.0	0.1
1 top (tW)	27.3	429.3	391.2	55.1	49.4	4.0	1.3	0.2
1 top (t)	63.6	1504.4	1503.3	62.9	58.5	7.1	0.8	0.2
1 top (s)	1.66	35.3	35.3	1.9	1.6	0.04	0	0
Diboson	11.8	474.7	345.0	1.4	1.3	0.1	0.3	0.03

Table 4.6: The cross section and number of events expected in 100 pb^{-1} after each cut are listed for different processes. The loose muon selection is used, and the results are shown for each of the three jet selections. The final two columns show the event yields in the high MET signal region.

3.4 Missing Transverse energy

Missing transverse energy is a primary observable for SUSY; indeed, prediction of the background's MET distribution is the goal for many of the methods described below. It is likely that a discovery will rely more on the MET *distribution* than on an event count above a single specific MET cut. However, it is useful to use such a MET cut to simplify presentation of each method's MET prediction. Usage of a single, reference cut of MET ≥ 200 allows comparison of the various methods' results.

4 Backgrounds

In this section, we consider the main backgrounds that are expected to appear in the single lepton analysis, and we examine their behavior as the selection cuts are applied.

We list in Table 4.6 the expected contributions from each of the different backgrounds, before and after each of the selection requirements. The results for the LM1 benchmark are included for comparison.

Figure 4.20 shows the MET distributions for various backgrounds after the requirement that there be a single, isolated muon satisfying the loose ($p_T > 10$ GeV) Muon ID requirements discussed in Sec. 3. No jet cuts have been applied. At this stage, the dominant background is W+jets, which is strongly suppressed by subsequent jet requirements.

The QCD background shown in Fig. 4.20 is already much smaller than the W+jets background. The sample used to study QCD was the so-called muon-enriched QCD sample, in which a muon is required at the generator level. This procedure has the advantage that it enables one to generate a larger number of events in which real muons from $b \rightarrow \mu$ or $c \rightarrow \mu$ decays occurs in jets, but it underestimates the total number of events passing the cuts because the full QCD sample is not generated, and even events without real muons can produce fake muons through punch-through and decays in flight. In addition, the QCD cross section is not a well known quantity, so the prediction here should only be regarded as a rough indication. A preliminary conclusion is that the lepton ID requirement is able to substantially suppress the QCD background. This issue should be studied with early data.

It is also apparent from Fig. 4.20 that the QCD background falls rapidly with increasing MET. At around 60 GeV, the $t\bar{t}$ background crosses over the QCD background, and becomes the second largest background after $t\bar{t}$. Smaller backgrounds are produced from Z+jets, single-top production, and VV (diboson) events. These sources contribute tails above 100 GeV in MET, but at a low level compared with W+jets and $t\bar{t}$.

Figure 4.21 shows the distributions of the number of jets above a threshold $p_T > 50$ GeV. From these plots it is clear that the W+jets background will be strongly suppressed by jet requirements, since most W bosons are accompanied but zero or one jets above this threshold. As for QCD, studies with early data will be important in verifying these Monte Carlo expectations.

Figure 4.22 shows the p_T distribution of the leading jet after only the muon requirement is applied. The jet spectrum in W+jets events falls steeply with jet p_T, and a large fraction of such events would be removed with a cut $p_T(\text{jet } 1) > 50$ GeV. The corresponding distribution for LM1 extends far above this value.

We next apply jet cuts. Figure 4.23 shows the MET distributions for the backgrounds after loose (upper plots) and tight (lower plots) jet requirements stated in Sec. 3. After the loose cuts, $t\bar{t}$ replaces W+jets as the dominant background for MET> 20 GeV. The LM1 signal starts to emerge for MET> 250 GeV. With the tighter jet cuts, the

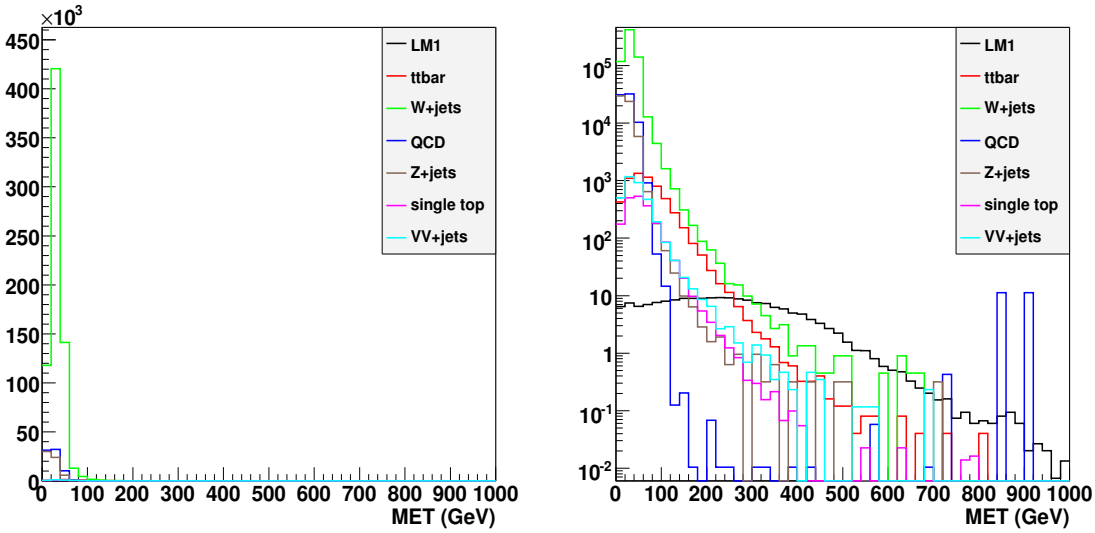


Figure 4.20: MET distributions for backgrounds shown on linear (left) and logarithmic scales (right).

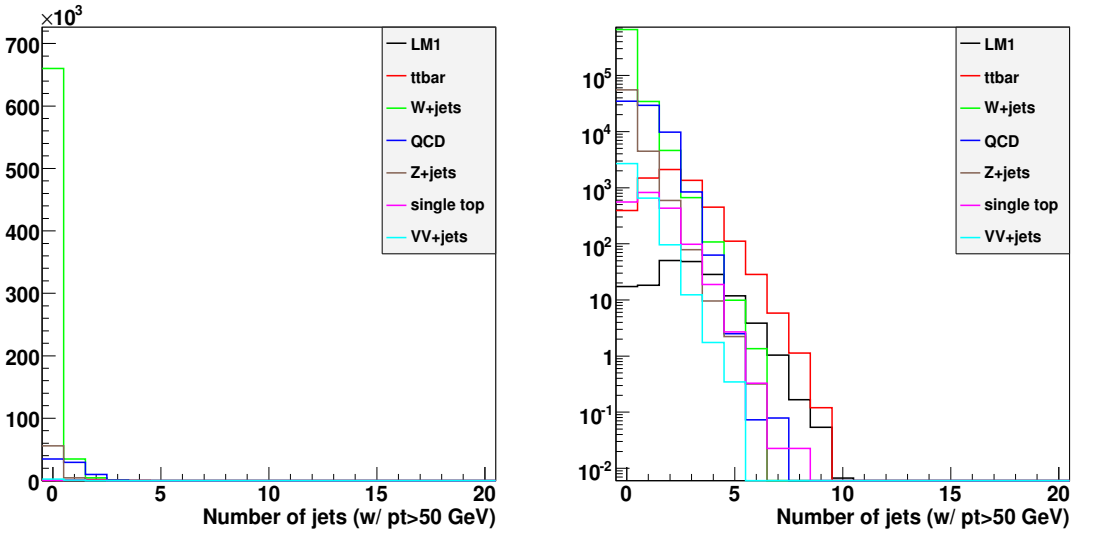


Figure 4.21: Number of jets with $p_T > 50$ GeV for backgrounds shown on linear (left) and logarithmic scales (right).

W+jets background is significantly suppressed, and $t\bar{t}$ becomes truly dominant, except again at very high values of MET.

Figure 4.24 shows the distribution of M_{eff} after loose and tight jet cuts. We observe behavior analogous to that seen in the MET distributions. The SUSY LM1 events produce a tail at very high values of M_{eff}, but this tail is comparable in size to that from the $t\bar{t}$ background. (Note that a MET cut has not been applied here.)

The M_T variable, which was described in Sec. 2, is shown for the backgrounds in Fig. 4.25. While the requirement $M_T > 80$ GeV would remove a substantial part of the SM backgrounds, the efficiency of such a cut for SUSY models is in general rather low as well.

We see that with rather simple cuts, we are able to dramatically reduce the SM backgrounds to the single-lepton SUSY search. However, the cross sections expected for much of the SUSY parameter space are such that it is desirable to further suppress the SM background. It is then mandatory that one find a method, or preferably multiple methods, to measure amount of remaining background. These tasks are described in the subsequent sections of this document.

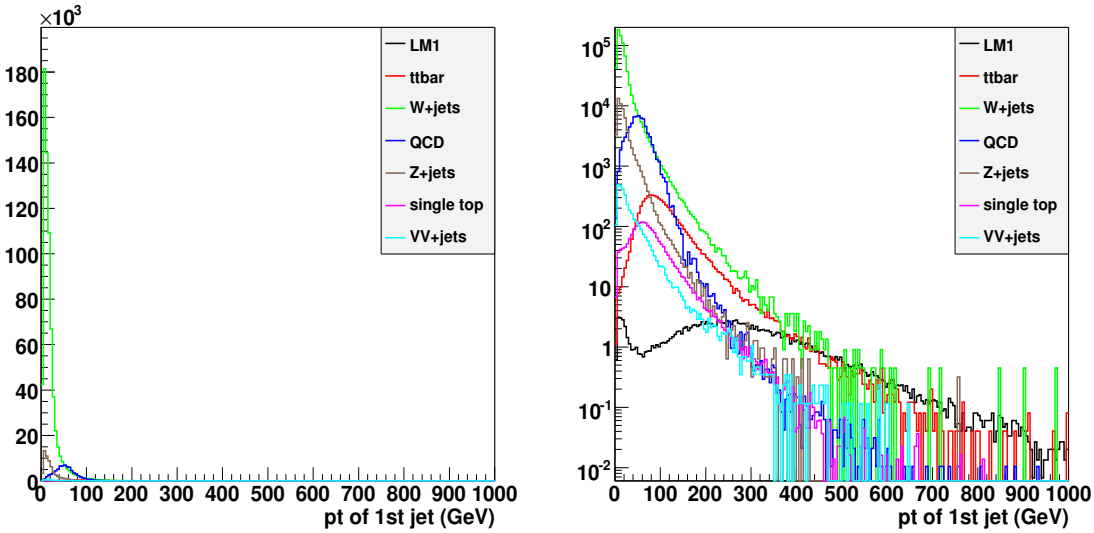


Figure 4.22: Leading jet p_T distributions with no jet cuts for backgrounds shown on linear (left) and logarithmic scales (right).

5 Background Determinations: Summary of Methods

In this section, we will have a very short description of each method and possibly a single figure illustrating the method.

5.1 Methods to suppress HCAL noise

Non-collisional sources of fake calorimetric MET can be minimized with the use of precise hit timing within the calorimeters. The HCAL has recently been demonstrated to achieve O(1ns) timing precision; however, the precision degrades at low energy. Nevertheless, a reduction factor of 5-10 in MET rates can be achieved with the use of HCAL timing. The key is to filter the reconstructed hits by their times with a filter shaped to match the degradation of timing precision as a function of energy. Hits with time outside the window calculated for the hit energy are rejected as "out-of-time." With the appropriately shaped filter the above quoted results have been demonstrated on detector noise (a source of significant fake MET particularly for HCAL) as well as cosmics and beam halo, while preserving signal MET distributions to the level of a few percent.

5.2 Standard Model MET Subtraction

To predict a $\not\!\!E_T$ distribution in the μ+jets+$\not\!\!E_T$ final state, one needs to model a) genuine $\not\!\!E_T$ from neutrinos produced in the decays of SM particles, and b) artificial $\not\!\!E_T$ due to instrumental effects and backgrounds. Since the charged lepton and neutrino $p_{\rm T}$ spectra in the SM $t\bar{t}$+jets and W+jets processes have similar shapes, the genuine $\not\!\!E_T$ from the neutrinos in SM events is modeled based the charged lepton $p_{\rm T}$ spectra. Effects producing artificial $\not\!\!E_T$ are sampled and modeled for each candidate event in-situ using multi-jet QCD events that have approximately the same configuration of jets as that of the candidate event. The genuine and artificial $\not\!\!E_T$ contributions are combined by smearing the charged lepton $\vec{p}_T$ measurement in each candidate event with a $\vec{\not\!\!E}_T$ resolution prediction from multi-jet QCD events by adding them at a random angle ϕ distributed uniformly from 0 to π in the transverse plane. The method is capable to predict SM backgrounds at high $\not\!\!E_T$ and a large number of jets to 20% or better without reliance on MC, and has discriminating power to reveal a new physics contribution in that region even in early data.

5.3 The α_T jet balancing method

Insert text here!

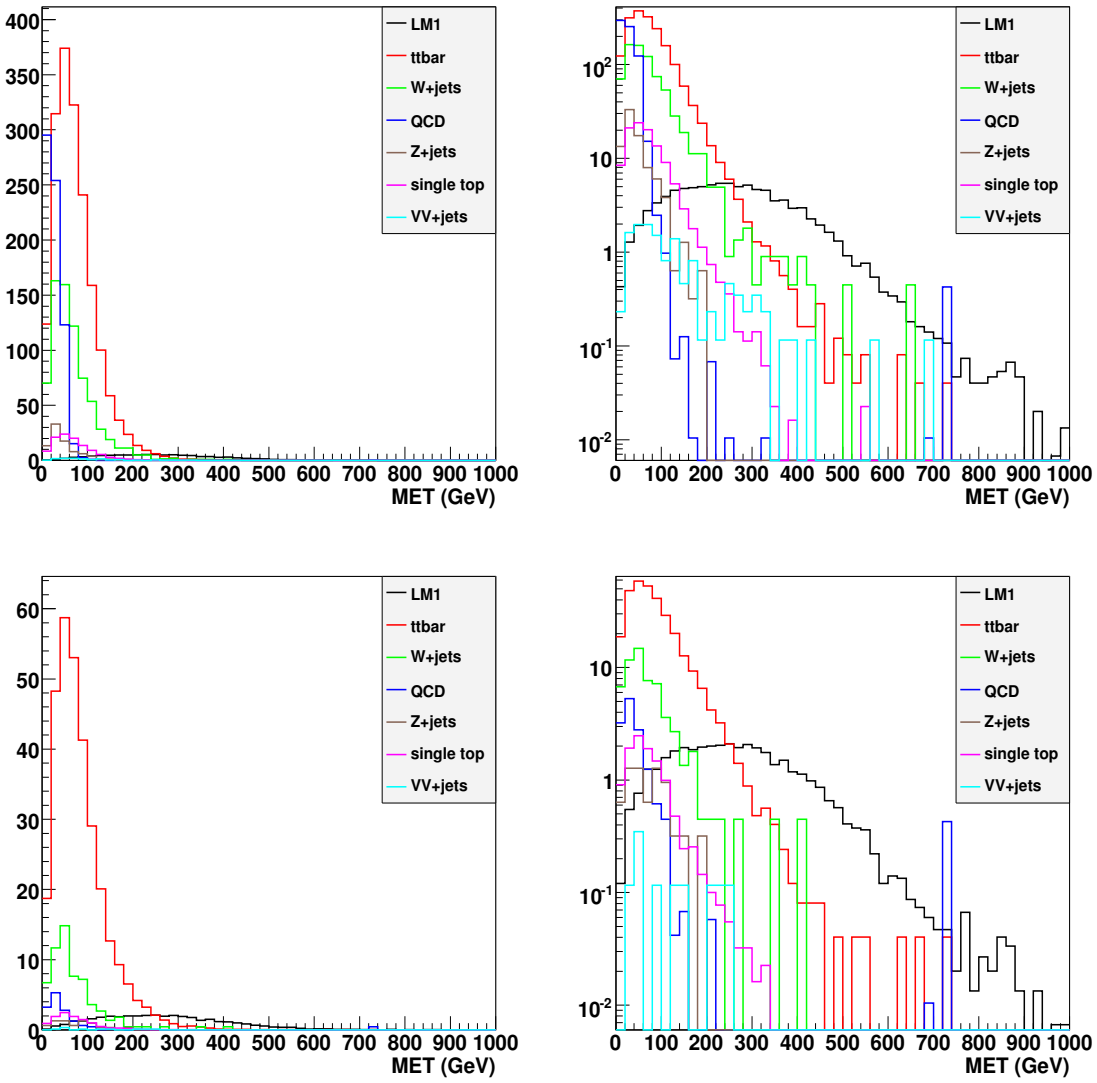


Figure 4.23: MET distributions for backgrounds shown after loose (upper plots) and tight jet cuts (lower plots) on linear (left) and logarithmic scales (right).

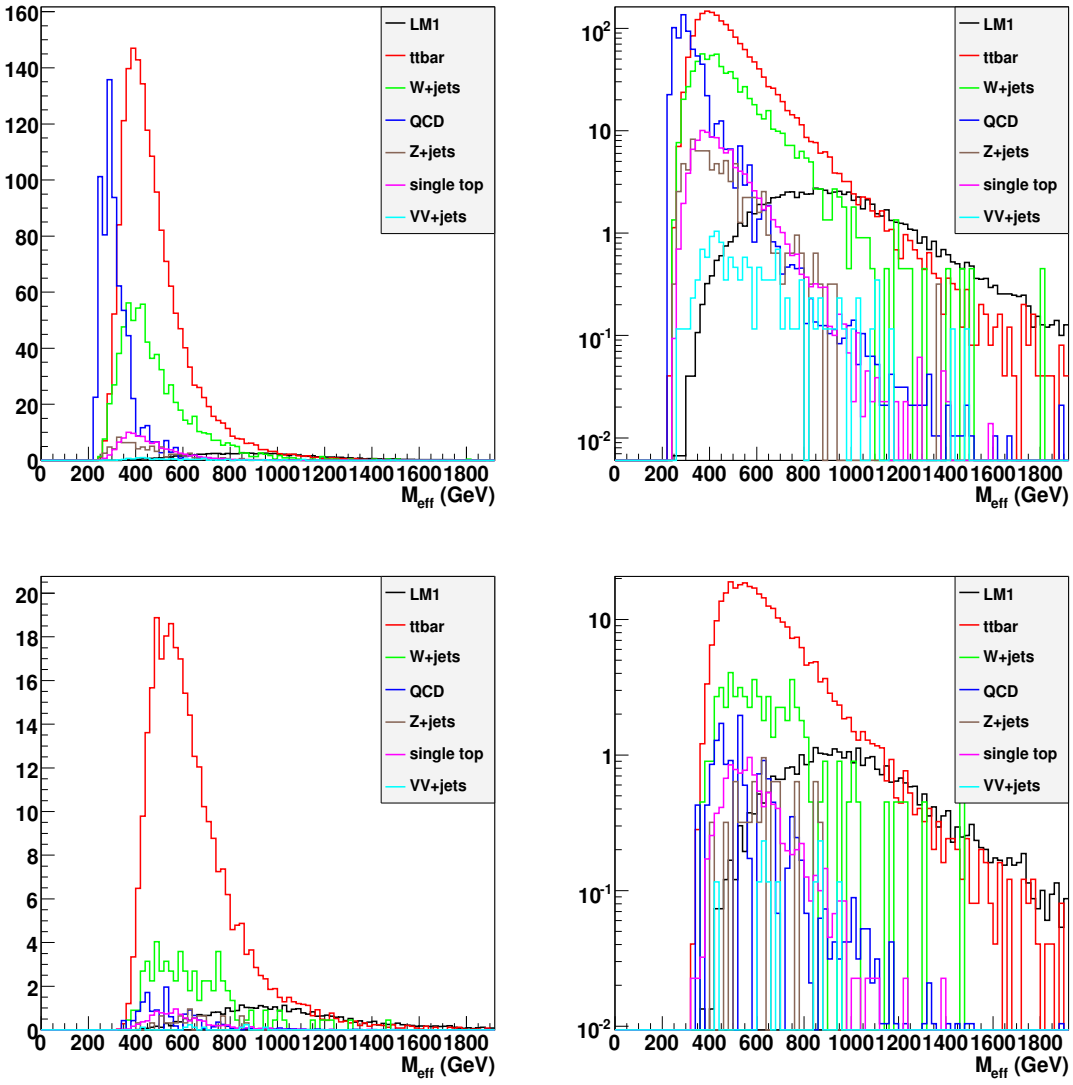


Figure 4.24: M_{eff} distributions for backgrounds shown after loose (upper plots) and tight jet cuts (lower plots) on linear (left) and logarithmic scales (right).

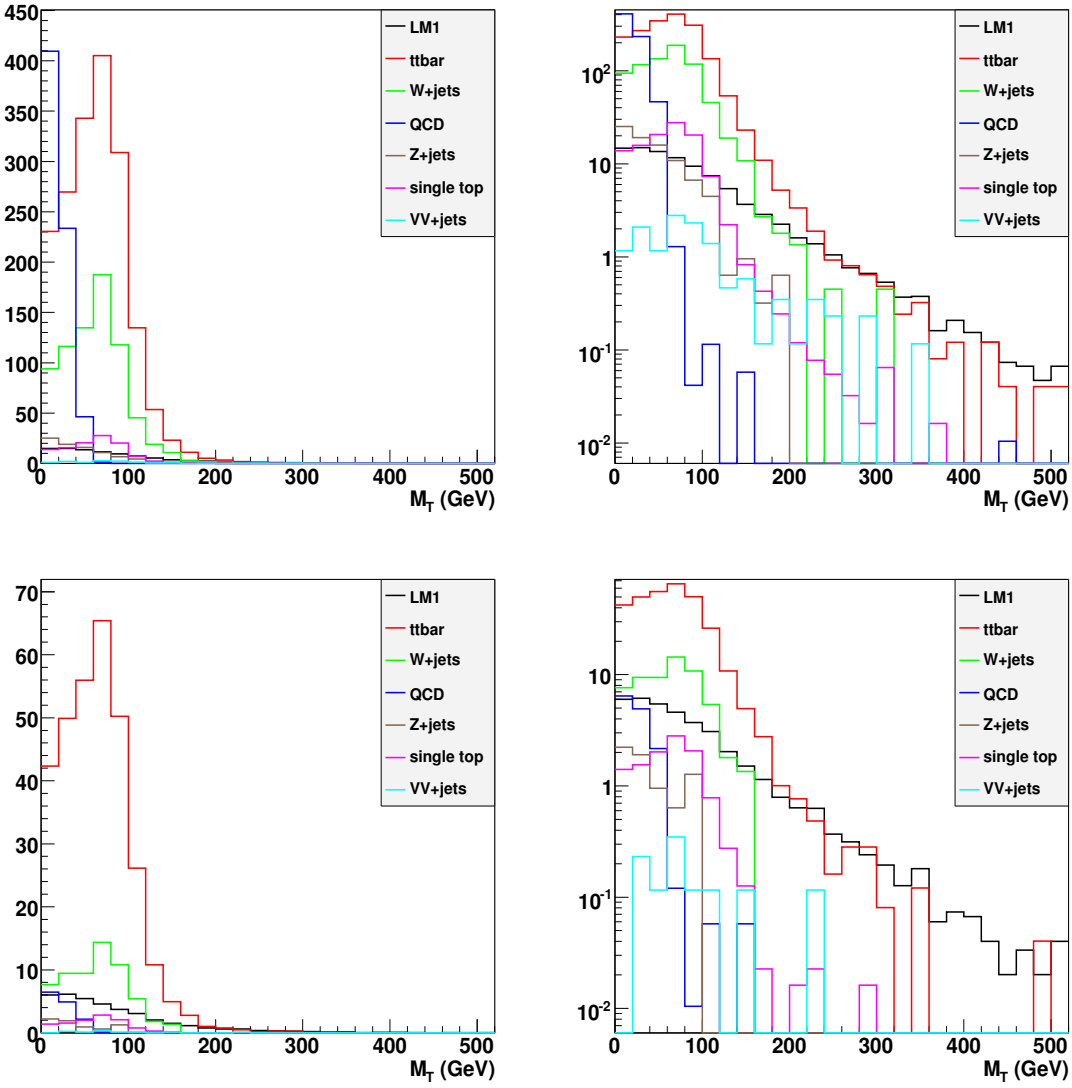


Figure 4.25: M_T distributions for backgrounds shown after loose (upper plots) and tight jet cuts (lower plots) on linear (left) and logarithmic scales (right).

5.4 Background estimation using $\chi^2(t\bar{t})$

In a SUSY leptonic search requiring exactly one muon, zero electrons, and modest jet cuts, the largest Standard Model background is from semi-leptonic $t\bar{t}$ events. To estimate the $t\bar{t}$ background, we construct a new variable, chi^2 , which uses the hypothesis of semi-leptonic $t\bar{t}$ for each event to reconstruct the leptonic top mass, hadronic W mass, and hadronic top mass. We choose the four highest pt jets in the event and take the lowest chi^2 permutation as the combination of jets assigned to the hadronic W, b-jet of the hadronic top, and b-jet of the leptonic top:

$$\chi^2(t\bar{t}) = \frac{(M_{j_1 j_2} - M_W)^2}{\sigma_{jj}^2} + \frac{(M_{j_1 j_2 j_3} - M_t)^2}{\sigma_{jjj}^2} + \frac{(M_{W_{\ell\nu} j_4} - M_t)^2}{\sigma_{\mu\nu j}^2} \tag{23}$$

where the weight factors are determined from Monte Carlo and take into account detector resolution eects for each term. In $t\bar{t}$ events a correct combination of jets gives a sharp peak at very low χ^2 , and the tail in the χ distribution comes mostly from the incorrect combination of jets. In SUSY events, there should not be a correct combination of jets to satisfy the semi-leptonic $t\bar{t}$ hypothesis, so the χ^2 distribution has a large tail and little or modest peak at a χ^2 of zero (the exact shape depends on the SUSY model). This shape dierence of the χ^2 distribution in $t\bar{t}$ and SUSY events can be used in conjunction with other variables to make a $t\bar{t}$ background prediction. The rst method we tried is an ABCD method with χ^2 vs. MET. This method underestimates the amount of top background by 50% due to the correlation between χ^2 and MET. To obtain a more precise estimate we then looked at the variable MET/ΣET , which is less correlated with χ than MET. An ABCD prediction with χ^2 vs. MET /ΣET gives an overestimate of the top background by 15-20pb[1] the χ^2 vs. MET /Σ ET ABCD method gives a statistical error of 30% and a systematic error of 5%-10%.

5.5 Top Box method to estimate $t\bar{t}$ background

The Topbox method estimates the $t\bar{t}$ contribution to SUSY signals such as missing transverse energy (MET) and total transverse energy (HT). We select events with one well isolated muon and 4 jets with transverse energy (ET) greater than 30 GeV with the leading jet $E_T > 80$ GeV. In semileptonic $t\bar{t}$, we should find a two jets combination with an invariant mass (M2) near the W boson mass, and a three jets combination with an invariant mass (M3) near the top quark mass. Event selections and detector effects may skew the invariant masses and cause them to deviate from their respective PDG values, but we can simulate these effects by using Monte Carlos (MC) matched jets to make the M2 and M3 Distribution in ttbar semileptonic events. Using these distributions a cut of $70 < M2 < 110$ and $150 < M3 < 210$ GeV was chosen. The Leptonic M3, which is calculated using the kinematics of the muon, the MET, and a jet, is not used since it could bias the MET prediction. The events passing the M2 and M3 cuts constitute the Topbox, and their kinematics are used to predict the contributions from $t\bar{t}$ in the full event sample.

The method can be applied to predict kinematic distributions, where the shape from the Topbox is scaled to the full sample's size, using the signal depleted region ($50 < \mathrm{MET} < 100$). In a pure $t\bar{t}$ sample, the method underpredicts the number of events with MET > 200 GeV by 13%. The prediction is 34.3 events while the actual full sample has 39.5 events. When the LM1 signal is added, the method predicts 62.6 events. This is a large overprediction due to signal contamination. However, the actual full sample has 96.5 events with MET > 200 GeV, so the method still shows an excess above the prediction. The Topbox method creates a ttbar enriched sample, which can be used to study and predict ttbar's contribution to SUSY signals such as high MET.

5.6 Dilepton Background

We expect new physics with a single muon final state to show up in events with large missing transverse energy (or MET) and large transverse mass, or MT. Therefore, an understanding of the MET tails in the background is essential. There have been many attempts to use MET and MT in an ABCD method to predict the amount of Standard Model background events, especially from $t\bar{t}$, in a $\mathrm{MET} > 200$ GeV, $M_T > 200$ GeV signal region. We have shown that the simplest approach to this method will not work because 80% of the events in the high MET, high M_T region are dilepton events and thus is completely different physics from the lower MET and M_T regions from which we are extrapolating. Of these dilepton events, 60% come from mutau events with the hadronic decay of the tau dominating, 25% come from muelectron events where the electron escapes the veto and 15% come from dimuon events where one of the muons is lost. Using this information, we estimate the MET tails in the high M_T region using reconstructed dimuon events where we fake the loss of one of the muons. The energy from the lost muon is added to the MET and M_T is recalculated. Then we scale the new MET distribution to the contribution to MET from dilepton events, as estimated from Monte Carlo. Our new MET distribution estimates the shape of the high MET tail but overestimates the contribution.

5.7 Combinatorial reconstruction of $t\bar{t}$ background / Topbox method

What denoted here as Topbox method is the combinatorial reconstruction of $t\bar{t}$ background. In this method the both top quarks become reconstructed step by step using informations of the myon, the jets and the transverse missing energy. Reconstructed this quantities, the method defines a control region with a high purity of top events and less signal contamination. When data is available, the topbox method calculates number of top events in the control region and extrapolates this number into the region with signal.

5.8 Estimation of $t\bar{t}$ background using b tagging

The aim of this method is to obtain the number of events coming from the standard model processes in the signal region from data, but also the shape of the MET distribution for these processes. We are mainly focusing on top pairs background as it is the dominant background for the leptonic SUSY searches but QCD and W+jets backgrounds can be be estimated and substracted in principle before applying the following method.

The first step is to obtain a template of the MET distribution for the $t\bar{t}$ contribution from data using a Control Region. Then, this template is normalized according to the number of events in a Normalization Region. And finally both number of events and tails of MET distribution can be estimated inside the Signal Region, where the search of SUSY is performed.

The definition of the Control Region have to fulfill the following requirements:

being ttbar enriched, being as background free as possible (standard processes like W+jets and QCD) and being unbiased with respect to the Signal Region. This last point means that the shape of the MET distribution coming from top pairs events should be same in the Control and Signal regions.

In order to obtain this unbiased template with the same proportion of semi-leptonic (dominant) and di-leptonic top pair events (important in the tail of MET), we use observables which are expected to be the same for both channels. The common signature is the presence of two b-quarks and a lepton.

That's why the Control Region is defined using b-tagging information and additional observables build with the two b-tagged jets required and one isolated lepton. These observables are the angle between the two b-tagged jets, the scalar sum of their transverse momenta and the mass of the system formed by the lepton and the closest b-tagged jet. The cuts on these variables are a compromise between the minimization of the statistical errors and the minimization of the signal contamination expected from some benchmark scenarios of SUSY.

5.9 W polarization method

Insert text here!

5.10 W background determination from Z control sample

To estimate the expected background contribution from the W boson decay channel in our signal region, we use a data-driven method with minimal reliance on Monte Carlo information. Our background estimation technique is divided into three parts. The first part uses the ABCD method to estimate the number of W decay events in the W control region (defined below). Part two involves estimating the efficiency (for a pure W sample) of the jet cuts required to be in our signal region (defined below). This is the only step in our method that is MC dependent. The final part deals with the fact that the signal region has a MET cut of 150 GeV, whereas the W control region has a MET cut of 35 GeV. Using a $Z \rightarrow e^+e^-$ sample extracted from data, we obtain a template for the W MET shape by dropping one of the two electrons and its associated calorimeter deposits and reconstructing the MET while making allowances for the kinematic differences between a Z and W event. This shape is then rescaled based on the estimated number of W events in the control region after applying the jet requirement efficiency. To predict the number of events in the signal region, we then consider the histogram bins with $\mathrm{MET} > 150$ GeV.

The signal and W control regions are defined as

- Signal region: (i) Single electron trigger fired, (ii) exactly one good electron, (iii) at least four jets with $Pt > 40$ GeV, (iv) MET> 150 GeV
- W control region: (i) Single electron trigger fired, (ii) exactly one good electron above 20 GeV, (iii) MET$>$ 35 GeV

5.11 Methods to suppress signal contamination in backround control regions

The ABCD-method predicts the background in the signal region by using two discriminating and independent variables to extrapolate from background dominated regions into the signal domain. Two of such promising variables for leptonic SUSY searches are the missing transverse energy (MET) and the transverse mass of the transverse reconstructed W-boson (M_T). The quality of the results obtained in this way is, however, strongly degraded by signal contamination of the background regions. An additional signal-suppressing cut is therefore applied in the problematic regions. This third variable is a combination of the total transverse energy (HT) and the transverse lepton momentum which strongly discriminates against the signal but effects the interdependence of the primary variables only weakly.

5.12 Determination of the correlation in ABCD

An alternative ABCD variable pair is MET in combination with a reconstructed hadronic top-quark mass ($M3$). These variables are more strongly dependent, which makes an estimation of the correlation paramount. This is achieved by measuring the correlation in four additional signal-depleted regions, obtained through subregions in mT. This procedure is termed the 8-fields method.

5.13 Fake Electrons

The isolation extrapolation method is a data-driven method that can be used to estimate the contribution of fake electrons to final states with a single electron, jets, and missing transverse energy (MET). The method works by extrapolating from a region dominated by the background to the signal region. Here, we use relative isolation as a discriminator between real and fake electrons. By fitting the relative isolation distribution above some threshold, which we assume is dominated by QCD, we can get the full shape of this distribution and use it to predict the contribution from this background.

5.14 QCD Background

Insert text here!

6 SM $\not{E}_T$ Subtraction

6.1 Description of Method

In the l+jets+$\not{E}_T$ channel, the dominant SM background is $t\bar{t}$+jets, in which one of the two top quarks decays semileptonically, $t \to bW^+(l^+\nu_l)$. The second largest SM background comes from the W+jets process with the W decaying leptonically, $W \to l\nu_l$. To predict the $\not{E}_T$ distribution for these SM processes, one needs to account for artificial $\not{E}_T$ due to instrumental effects, and genuine $\not{E}_T$ produced by the undetected neutrinos. I will first introduce a technique to predict artificial $\not{E}_T$, followed by a technique to predict genuine $\not{E}_T$, and then describe how these two predictions are combined to obtain a full SM $\not{E}_T$ prediction that can be compared with the measured $\not{E}_T$ [1].

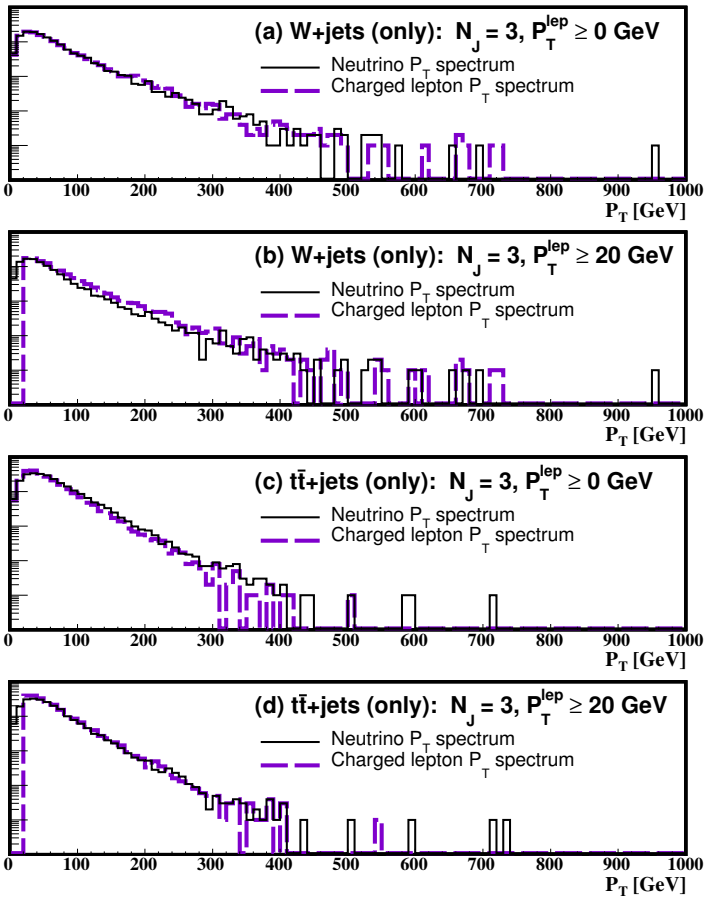
Since charged lepton momenta are well-measured, the artificial $\not{E}_T$ comes primarily from the system of jets, the detector, beam related backgrounds and non-collision effects. Let us consider an event with four jets and a W boson decaying to $\mu\nu_\mu$. The system of the four jets and other effects unrelated to the W are the main source of artificial $\not{E}_T$ in this event. These effects are difficult to predict or simulate. Instead, they are modeled *in-situ* using multi-jet QCD events with a kinematical configuration of jets similar to that in the four jet system of the W+jets event according to the following algorithm.

A pool of $\not{E}_T$ templates is constructed using multi-jet QCD events. QCD events for the templates are selected based on two variables: (1) N_J, number of jets above a high $p_{\rm T}$ threshold, 50 GeV or higher, and (2) $J_T \equiv \sum |\vec{p}_T^{\,\rm jet}|$ for jets above a low 20 GeV $p_{\rm T}$ threshold. A template is constructed from QCD events with the same N_J and in the same J_T bin as a $\not{E}_T$ distribution in that sample normalized to unity. The width of J_T bins is 10 GeV (100 GeV) below (above) 1 TeV. (This definition of a template can be modified without significant effect.) For each l+jets+$\not{E}_T$ event, N_J and J_T are measured and used to select the template from the pool with the same N_J in the corresponding J_T bin, which represents its artificial $\not{E}_T$ prediction.

The genuine $\not{E}_T$ in $t\bar{t}$+jets and W+jets comes from the undetected neutrinos produced in W decays. To avoid reliance on MC and theory, the neutrino $p_{\rm T}$ spectra are modeled using the charged lepton $p_{\rm T}$ spectra. If the W

bosons are not polarized in the transverse plane, the two p_{T} spectra should be the same. Event selection (mainly, the p_{T} threshold for charged leptons and the $|\eta_{\text{charged lepton}}|$ coverage), polarized W bosons produced in top quark decays [3], and the contamination from di-leptons in $t\bar{t}$+jets, lead to differences in the charged lepton and neutrino p_{T} spectra. However, these differences are small and can be reliably accounted for by corrections. Plots (a) and (c) in Figure 6.26 show charged lepton and neutrino p_{T} spectra at the generator level in W+jets ($l\nu_l$ and jets) and $t\bar{t}$+jets ($l\nu_l b\bar{b}$ and jets), respectively, for $|\eta_{\text{charged lepton}}| < 2.0$ from Ref. [1], where it is seen that the neutral and charged lepton spectra have very similar shapes. The effect of the 20 GeV charged lepton p_{T} threshold is seen in plots (b) and (d) of the same Figure that correspond to plots (a) and (c) after the application of the requirement $|\vec{p}_T^{\,\mu}| > 20$ GeV. I find that for modest integrated luminosities no corrections to the charged lepton p_{T} spectra may be required to achieve a prediction at high $\not\!\!E_T$ that is sufficiently precise for an early search since the statistical uncertainty dominates. A more detailed discussion of the genuine $\not\!\!E_T$ modeling is given in the next section.

Figure 6.26: **Gen. MC** samples: Comparisons of charged lepton and neutrino spectra in W+jets (a,b) and $t\bar{t}$+jets (c,d) without (a,c) and with (b,d) charged lepton p_{T} thresholds all for $N_J = 3$.



The artificial $\not\!\!E_T$ prediction modeled by templates and the charged lepton p_{T} spectra modeling the neutrino p_{T} spectra are combined into a full $\not\!\!E_T$ prediction. It is assumed that the neutrino (charged lepton) $\vec{p}_T$ and the artificial $\vec{\not\!\!E}_T$ interfere at a random angle ϕ distributed uniformly from 0 to π in the transverse plane. Under this assumption, for each l+jets+$\not\!\!E_T$ event, the charged lepton p_{T} is smeared with the artificial $\not\!\!E_T$ prediction from a template at an angle ϕ. The smeared charged lepton p_{T}'s are summed over all l+jets+$\not\!\!E_T$ events to obtain a $\not\!\!E_T$ prediction for the entire l+jets+$\not\!\!E_T$ sample.

Since higher sensitivity to new physics is expected in events with a large number of jets, the method is developed to model the high $\not\!\!E_T$ region in events with 3 or more jets. The method works best in that regime rather than in events with a small number of jets as demonstrated and explained in Ref [1]. Events with 2 jets are valuable as a validation and calibration sample. They are especially valuable for commissioning purposes.

This method is developed for searches in early data. It will work best if the LHC start-up is quick, new particles are strongly produced and not very heavy, such as squarks and gluinos in the low mass mSUGRA benchmarks [2]. In this scenario, a prediction of SM backgrounds in high $\not\!\!E_T$ tails in events with a large number of jets (3 jets or more) to about 20% may be sufficient to reveal new physics, which I use as an accuracy benchmark for the method. The method will also work with large integrated luminosities, but the role of the corrections described above becomes more important with larger event yields.

6.2 Performance of method without signal

There are three critical steps that need to be tested: the quality of the artificial $\not\!\!E_T$ modeling, the quality of the genuine $\not\!\!E_T$ modeling using the neutrino $p_{\rm T}$ spectra, and the validity of the assumptions used in combining the artificial and genuine $\not\!\!E_T$ predictions. With this in mind, I will first demonstrate how well the method performs in predicting only artificial $\not\!\!E_T$ in W+jets and $t\bar{t}$+jets. Second, I will smear the neutrino $p_{\rm T}$ spectra at the generator level in W+jets and $t\bar{t}$+jets to show how well the method models $\not\!\!E_T$ resolution effects in the full $\not\!\!E_T$ prediction when the genuine $\vec{\not\!\!E}_T$ is known. Third, I will show full tests where, in addition, the neutrino $p_{\rm T}$ spectra are modeled by the muon $p_{\rm T}$ spectra.

The **Winter09**, **Summer08** and **Gen. MC** samples in Table 6.7 are used to study the method. MC samples in Table 6.7 labeled as **Gen. MC** are self-produced. In these samples, the jet energy resolution is parameterized according to Ref [1]. They are used in Figure 6.26. For the **Winter09** and **Summer08** samples events are required to pass the `HLT_Mu9` trigger. No trigger requirement is used to select QCD events. It is assumed that prescaled small $p_{\rm T}$ jet triggers and unprescaled high $p_{\rm T}$ multi-jet triggers will be used to collect QCD events in the data. Muon and jet selection is close to the standard and can be modified. Muons are selected in the $|\eta| < 2.0$ range above the 20 GeV $p_{\rm T}$ threshold. Jets in the $|\eta| < 3.0$ range with a corrected $p_{\rm T}$ of 20 GeV or more are used to measure $\not\!\!E_T$ and J_T. Higher jet $p_{\rm T}$ thresholds of at least 50 GeV are used to measure the number of jets, N_J, in a robust manner. The vector of missing transverse energy, $\vec{\not\!\!E}_T$, is calculated as the vector opposite to the sum of $\vec{p}_T$ measurements of muons and jets in each event. $\not\!\!E_T$ calculated according to this prescription is often called missing H_T at CMS. It is required that the leading jet and $\vec{\not\!\!E}_T$ be not aligned in the transverse plane within 0.15 radians: $0.15 < |\Delta\phi^{\text{lead jet}-\not\!\!E_{\rm T}}| < (\pi - 0.15)$. (The jet with the highest $p_{\rm T}$ in an event is the leading jet of this event. Any other jet in this event is a non-leading jet.) These selection criteria are not optimized to any new physics model. Instead, they are chosen to ensure robust detector performance and maintain sensitivity to a wide range of new physics models at high $\not\!\!E_T$ and a large number of jets.

Process	$\sqrt{s}$ (TeV)	$\mathcal{L}$ (fb^{-1})	Data set path
Winter09:			
Z+jets	10	2.0	`/Zjets-madgraph/Winter09_IDEAL_V11_FastSim_v1/GEN-SIM-DIGI-RECO`
W+jets	10	1.0	`/Wjets-madgraph/Winter09_IDEAL_V11_FastSim_v1/GEN-SIM-DIGI-RECO`
$t\bar{t}$+jets	10	1.0	`/TTbar-madgraph/Winter09_IDEAL_V11_FastSim_v1/GEN-SIM-DIGI-RECO`
QCD	10	–	`/QCDpt30/Winter09_IDEAL_V11_FastSim_v1/GEN-SIM-DIGI-RECO`
		–	`/QCDpt80/Winter09_IDEAL_V11_FastSim_v1/GEN-SIM-DIGI-RECO`
Summer08:			
W+jets	10	0.2	`/WJets-madgraph/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO`
$t\bar{t}$+jets	10	0.2	`/TTJets-madgraph/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO`
QCD	10	–	`/QCDpt30/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO`
		–	`/QCDpt80/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO`
LM1	10	0.2	`/SUSY_LM1-sftsht/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO`
Gen. MC:			
W+jets	14	1.0	self-production
$t\bar{t}$+jets	14	1.0	self-production

Table 6.7: MC samples used to study the method.

6.2.1 Artificial $\not\!\!E_T$ prediction

Figure 6.27 shows comparisons of predicted and observed yields in the **Winter09** samples. The left columns show Z+jets. In the right columns W+jets events are used to emulate Z+jets events with a larger yield. The predictions (solid red) are consistent with the measured distributions (dashed blue) upon visual inspection in all plots of this Figure. Ratios of observed and predicted yields, $N_{\rm Observed}/N_{\rm Predicted}$, are shown in the first column of Figure 6.28, where the yields in each $\not\!\!E_T$ bin are integrals of the distributions shown in Figure 6.27 from that bin's $\not\!\!E_T$ value to infinity. Since my precision benchmark for predicting SM backgrounds at high $\not\!\!E_T$ and a large number of jets is 20%, I conclude that the methods performs reasonably well in predicting artificial $\not\!\!E_T$ in SM V+jets for an early search.

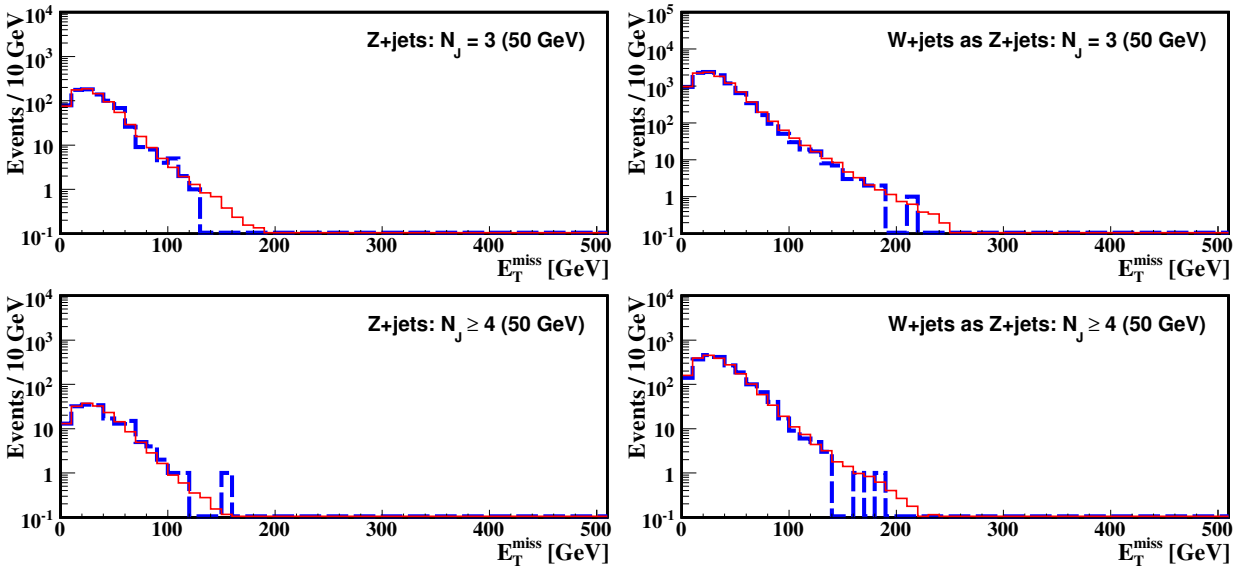


Figure 6.27: **Winter09** sample: Algorithm performance in Z+jets for $N_J = 3$ and ≥ 4 in the first and second rows, respectively. The left column is for Z+jets; the second column shows results for W+jets, where $W \rightarrow \mu\nu$ and the neutrino is treated as the second reconstructed muon using the generator $\vec{p}_\nu$. The observed $\not\!\!E_T$ distributions are shown in the dashed blue lines, their predictions obtained from multi-jet QCD events are the solid red lines.

SM $t\bar{t}$+jets events, where $t\bar{t} \rightarrow l\nu_l b\bar{b}q\bar{q}$, constitute a dominant background in the l+jets+$\not\!\!E_T$ signature for $N_J \geq 3$. To demonstrate this bias clearly, Figure 6.29 (left column) shows the artificial $\not\!\!E_T$ in $t\bar{t}$+jets, **Winter09** samples, where the neutrino four-momentum is assumed to be measured so that it is included in the $\not\!\!E_T$ calculation, and its prediction using QCD templates (solid red line). It is seen that the predictions are biased. Two dominant sources that lead to a bias in artificial $\not\!\!E_T$ predictions in $t\bar{t}$+jets are a) differences in N_J and J_T spectra between QCD and $t\bar{t}$+jets, and b) semileptonic bottom and charm decays in the two b-jets of each $t\bar{t}$+jets event. In the plots in the right columns of Figure 6.29, bottom and charm semileptonic decays are excluded so that one can make a judgment how significant the effect due to the differences in N_J and J_T is. By examining these and similar plots for other MC samples in Ref. [1], I conclude that, in general, these two effects are comparable in size. Depending on the jet energy resolution one or the other may dominate. As the jet energy resolution improves, the effect of bottom and charm semileptonic decays should become dominant. I note that the effect of semileptonic bottom and charm decays in $t\bar{t}$+jets events can be easily taken into account by modeling the undetected fraction of the b-jet energy carried by neutrinos and muons in QCD events.

6.2.2 Smearing of genuine $\not\!\!E_T$ by templates

The predictions of artificial $\not\!\!E_T$ in $t\bar{t}$+jets of Figure 6.29 is biased. It can be improved as described above but this is not necessary since at large $\not\!\!E_T$, the bias is an order of magnitude smaller compared to the genuine $\not\!\!E_T$ from the neutrinos having the $p_{\rm T}$ spectrum shown in Figure 6.26 for $N_J = 3$. When the neutrino $p_{\rm T}$ spectrum is combined with the artificial $\not\!\!E_T$ in the full $\not\!\!E_T$ prediction, the bias becomes insignificant as seen in Figure 6.30 (left column), where both $t\bar{t}$+jets and W+jets are included according to their expected cross sections. Ratios of observed

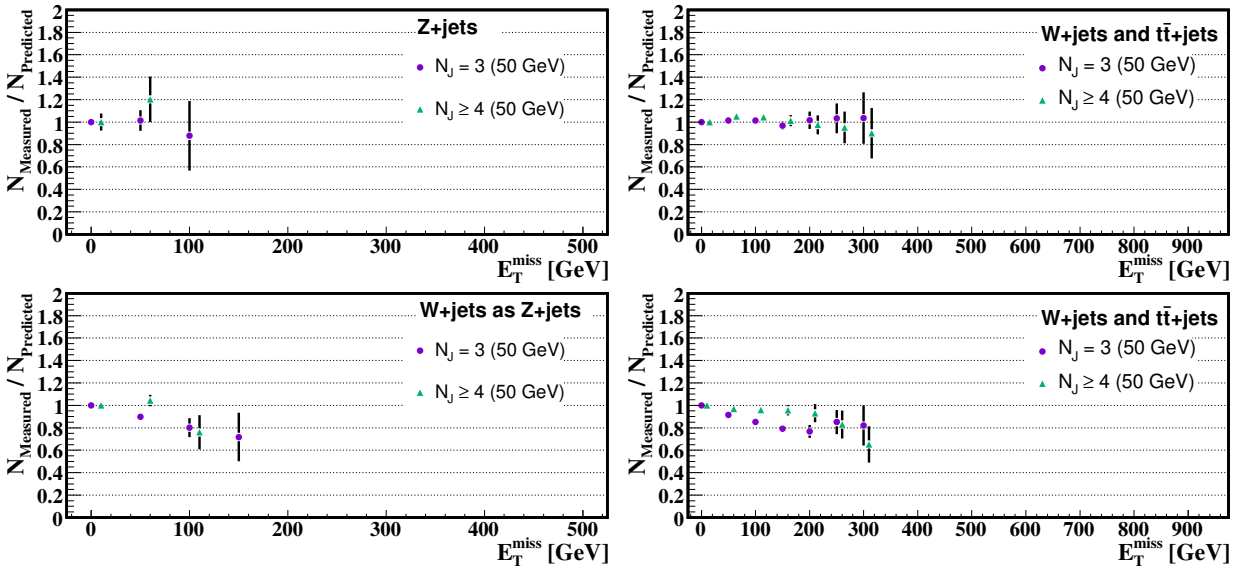


Figure 6.28: **Winter09** sample: Ratios of observed and estimated integrated yields. First column: Z+jets (top) and W+jets treated as Z+jets (bottom), second column: W+jets and $t\bar{t}$+jets combined based on the generator neutrino (top) and charged muon (bottom) p_{T} spectra. all obtained for the 50 GeV jet p_{T} threshold for N_J. In each plot two types of markers are shown for $N_J = 3$ (circles) and ≥ 4 (triangles). Note, the ratios are correlated since yields are integrated upwards.

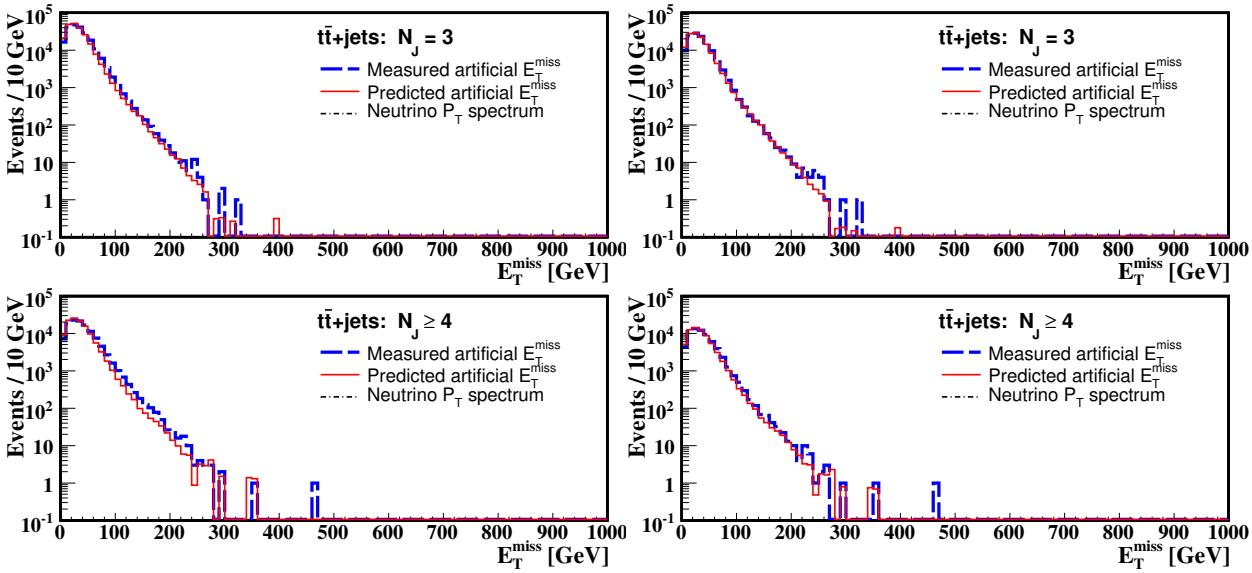


Figure 6.29: **Winter09** sample: Artificial $\not\!\!E_T$ in $t\bar{t}$+jets (dashed blue) and its prediction (solid red) for $N_J = 3$ and ≥ 4 in the first and second rows, respectively, for the 50 GeV jet p_{T} threshold for N_J. In the plots in the right column, events with semileptopic charm or bottom decays are excluded. The observed $\not\!\!E_T$ distributions are shown in the dashed blue lines, their predictions obtained from multi-jet QCD events are the solid red lines. These plots are made to reveal clearly the bias in predicting the artificial $\not\!\!E_T$ in $t\bar{t}$+jets. The integrated luminosity plotted here is larger than that used in Table 6.7.

and predicted yields, $N_{\mathrm{Observed}}/N_{\mathrm{Predicted}}$, for these plots are shown in the top plot of the second column in Figure 6.28.

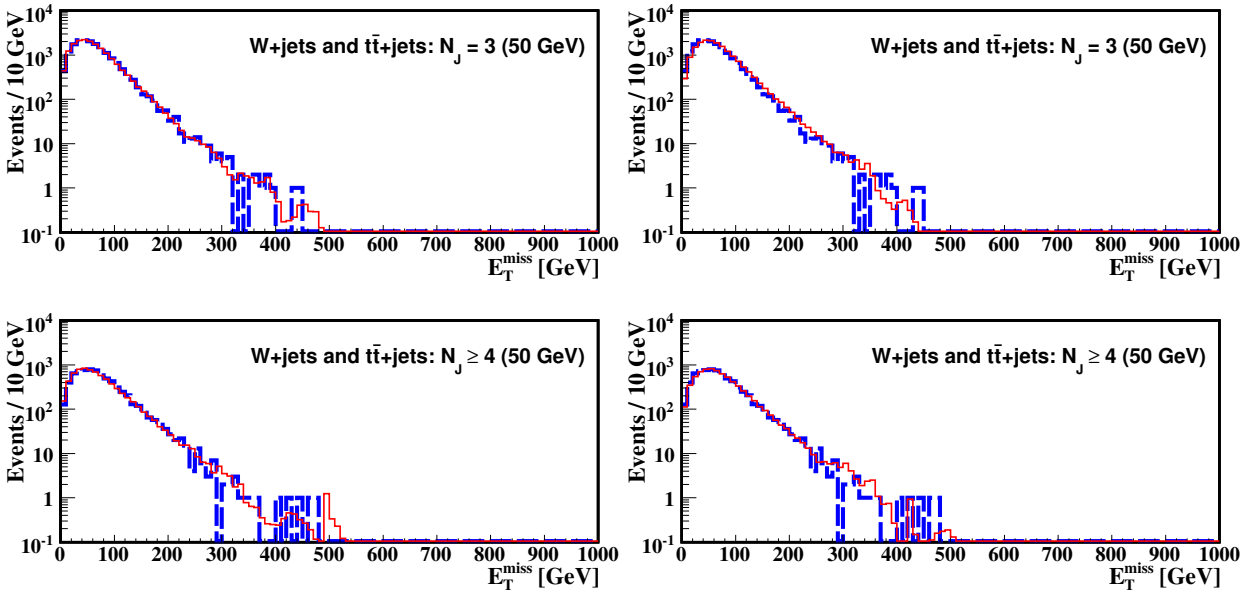


Figure 6.30: **Winter09** samples: Algorithm performance in W+jets and $t\bar{t}$+jets combined for $N_J = 3$ and ≥ 4 in the first and second rows, respectively. In the first and second columns the predictions are made based on the neutrino and charged lepton p_{T} spectra, respectively. The observed $\not\!\!E_T$ distributions are shown in the dashed blue lines, their predictions are the solid red lines.

The artificial $\not\!\!E_T$ is a dominant contributor in events with small genuine $\not\!\!E_T$. Figure 6.30 (left column) demonstrates that the accuracy of its prediction is sufficient to model the full $\not\!\!E_T$ distribution at small $\not\!\!E_T$. At high $\not\!\!E_T$, the missing momentum from the neutrino dominates over artificial $\not\!\!E_T$ so that the accuracy of the full $\not\!\!E_T$ prediction is highly dependent on how well the neutrino spectrum is modeled. This aspect of the method is critical in the l+jets+$\not\!\!E_T$ signature.

6.2.3 Full $\not\!\!E_T$ prediction

The neutrino p_{T} spectra can be modeled by the charged lepton p_{T} spectra. The solid and dashed lines in plot (a) of Figure 6.26 are the neutrino and charged lepton spectra in W+jets for $N_J = 3$, $W \to l\nu_l$, where the requirement on the charged lepton p_{T} of at least 20 GeV is removed. The W bosons are effectively unpolarized in the transverse plane in W+jets [3], so that the neutrino and charged lepton p_{T} spectra are consistent. The application of a p_{T} threshold on the charged lepton makes its spectrum harder, while the neutrino spectrum becomes softer, as seen in plot (b) of the same Figure. To model the neutrino spectrum using the reconstructed charged lepton spectrum in W+jets, the effect of the charged lepton p_{T} threshold needs to be corrected for. The corrections need to be obtained from MC simulation.

The solid and dashed lines in plot (c) of Figure 6.26 are the neutrino and charged lepton p_{T} spectra in $t\bar{t}$+jets for $N_J = 3$, $t\bar{t} \to l\nu_l b\bar{b}q\bar{q}$, without a threshold requirement on the charged lepton p_{T}. In the SM, 30% of W^+ bosons in top decays are produced in the transverse-minus helicity state (left-handed) and the rest are longitudinally polarized [4]. Left-handed W^+ bosons tend to produce charged leptons with a p_{T} spectrum that is softer compared to the neutrino p_{T} spectrum as seen in plot (c). Since the two spectra have similar shapes, it is possible to use the muon spectrum to model the neutrino spectrum in $t\bar{t}$+jets. Again, when a charged lepton p_{T} threshold is applied, the charged lepton spectrum becomes harder while the neutrino spectrum becomes softer, which leads to a higher consistency between the two spectra seen in plot (d). Nevertheless, the effects of the W polarization in top decays and the event selection, mainly due to the charged lepton p_{T} threshold, in $t\bar{t}$+jets, in general, need to be corrected.

In order to determine corrections to the charged lepton spectra for W+jets and $t\bar{t}$+jets from MC simulation, one needs to measure the shape of the p_{T} dependence of lepton reconstruction efficiencies and the relative fractions of W+jets and $t\bar{t}$+jets in the data sample. The former can be readily done via a standard technique based on $Z \to l^+l^-$ decays [5][6]. The latter should come from an independent measurement. With these two ingredients, corrections can be determined from MC simulation.

Since corrections to the charged lepton spectra are small, the reliance on details of MC simulation to determine the neutrino $p_{\rm T}$ spectra is minimal. In fact, for a 20 or 15 GeV threshold on charged lepton $p_{\rm T}$, no corrections are required to predict the $\not\!\!E_T$ distributions in $t\bar{t}$+jets in all N_J bins to 20% or better in the mock data samples. Corrections are needed for W+jets. Figure 6.30 (right column) shows the $\not\!\!E_T$ distribution and its prediction in l+jets+$\not\!\!E_T$, W+jets and $t\bar{t}$+jets combined, for $N_J = 3$ and $N_J \geq 4$ based on the charged lepton spectrum without corrections. Since $t\bar{t}$+jets dominates over W+jets in the $N_J = 3$ and $N_J \geq 4$ bins, the prediction is good to about 20% at high $\not\!\!E_T$ without corrections. Ratios of observed and predicted yields, $N_{\rm Observed}/N_{\rm Predicted}$, for these plots are shown in the lower plot of the second column in Figure 6.28. The $N_J \geq 4$ bin, where the prediction is the most robust, is likely to have the highest sensitivity to a new physics contribution compared to lower jet multiplicity events.

6.2.4 Concluding remarks

In the l+jets+$\not\!\!E_T$ signature, there is background from tauonic W decays in W+jets and $t\bar{t}$+jets. Tauonic W decays produce at least one additional neutrino that is a source of differences between the muon and neutrino spectra.

There are two types of tauonic W decays that contribute significant background: (1) W+jets and $t\bar{t}$+jets, where $W^- \rightarrow \tau\bar{\nu}_\tau$ with $\tau \rightarrow l\bar{\nu}_l\nu_\tau$, and (2) $t\bar{t}$ events, where $W^- \rightarrow l\bar{\nu}_l$ and $W^+ \rightarrow \bar{\tau}\nu_\tau$ with $\bar{\tau} \rightarrow (\text{hadrons}\ \bar{\nu}_\tau)$. The contribution from tauonic W decays is an order of magnitude smaller compared to that from $W \rightarrow l\nu_l$ decays. (The tauonic background of type 2 can be suppressed by vetoing events with isolated single hadronic tracks.) The τ branching fractions are well known. Therefore, the effects from $W \rightarrow \tau\nu_\tau$ on $\not\!\!E_T$ predictions can be well-modeled by an additional smooth correction to the muon $p_{\rm T}$ spectra that can be determined by MC simulation. Since the contribution from tauonic W decays is smaller compared to that from $W \rightarrow l\nu_l$ decays in l+jets+$\not\!\!E_T$, corrections for them are secondary in importance and will be ignored in this note. The same is done for di-lepton events where one of the two leptons is lost since they constitute an even smaller background contribution.

There are improvements that can be made, I will mention only two here. First, I am developing a procedure to reduce the single lepton background from $t\bar{t}$+jets and W+jets by subtracting the $\not\!\!E_T$ distribution measured in the low M_T region. This subtraction has a significant potential to increase the method's sensitivity to new physics. Second, the key metric of this search method is the consistency of the charged lepton $p_{\rm T}$ spectra with that of the neutrino, since artificial $\not\!\!E_T$ is less important in l+jets+$\not\!\!E_T$. In addition to $\not\!\!E_T$, other variables can be constructed from observables describing the charged lepton and neutrino that are either more sensitive to new physics or more convenient to use. Both statements about improvements are model dependent.

6.3 Performance of method with signal

The algorithm's performance with a new physics contribution is illustrated in Figure 6.31. The integrated luminosity of the mock data samples in this Figure is 200 pb^{-1} for $\sqrt{s} = 10$ TeV using the **Summer08** samples. New physics contribution in the Figure is the LM1 mSUGRA benchmark. The plots show SM backgrounds with new physics contributions (dashed) and their $\not\!\!E_T$ predictions (solid) from QCD templates. The dot-dashed lines represent SM backgrounds only to simplify comparisons.

New physics events tend to have large J_T and $\not\!\!E_T$, and can bias a) the artificial $\not\!\!E_T$ prediction based on templates and b) the genuine $\not\!\!E_T$ prediction based on charged leptons $p_{\rm T}$ spectra. I find that even under the most optimistic scenarios for new physics cross sections the contamination of $\not\!\!E_T$ templates does not lead to a significant bias in the artificial $\not\!\!E_T$ prediction. Furthermore, it is seen that the full $\not\!\!E_T$ prediction is not biased significantly at high $\not\!\!E_T$ in the case of LM1. An excess of signal events above the background prediction stands out clearly. Since in l+jets+$\not\!\!E_T$ the neutrino spectra are modeled based on the charged lepton spectra in each N_J bin, the method works best in this signature when the charged lepton spectrum in new physics events is soft compared to the combined $p_{\rm T}$ spectrum produced by new weakly interacting particles in the event.

6.4 Systematic uncertainties

Systematic uncertainties need to account for the statistical precision and biases of the method's background predictions at high $\not\!\!E_T$. Mechanisms by which biases in artificial $\not\!\!E_T$ may appear are studied in Ref. [1]. In that study, several mis-reconstruction effects are introduced to a degree where the method becomes biased to explore the boundaries of the domain where the method works. It is determined that despite the fact that the method is designed to model a variety of effects in-situ, its predictions may become biased in severely mis-reconstructed events. These biases appear in tests in which there are large differences in differential distributions between the hadronic

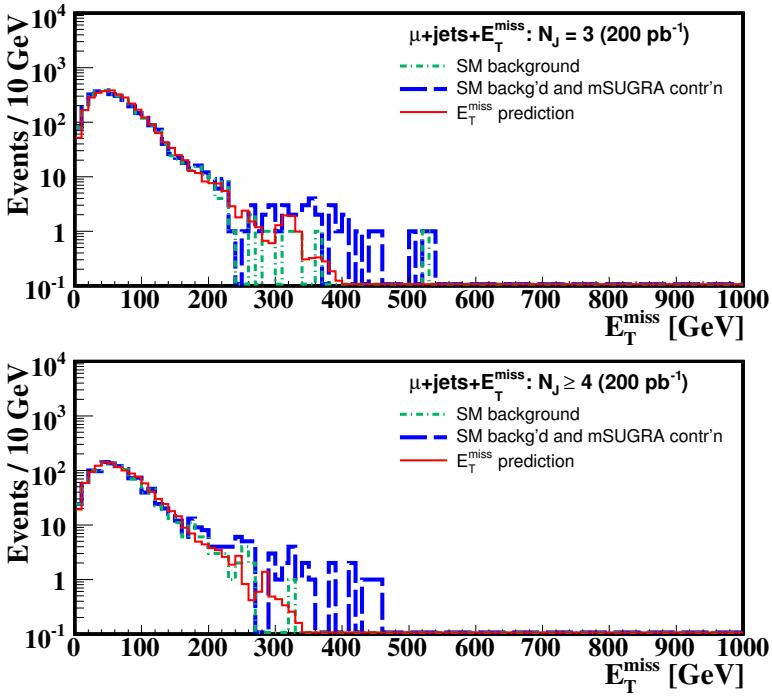


Figure 6.31: **Summer08** samples: Observed (dashed blue) and predicted (solid red) SM l+jets+$\not\!\!E_T$ for $N_J = 3$ (top) and $N_J \geq 4$ (bottom) with new physics contributions from mSUGRA benchmarks. The dot-dashed green lines highlight the SM contributions. The $p_{\rm T}$ threshold for N_J is 50 GeV. These plots correspond to 200 pb^{-1} at $\sqrt{s} = 10$ TeV.

system in V+jets and QCD events on the scale associated with the mis-reconstruction. These biases, however, can be avoided by imposing event quality selection criteria. It is also found that at higher N_J, jet $p_{\rm T}$ thresholds and J_T, the method tends to work better since a) there are fewer differences between in the hadronic systems of V+jets and QCD, b) averaging effects over V+jets and QCD events in the method predictions are stronger, and c) jet reconstruction tends to work better at higher jet $p_{\rm T}$ thresholds and J_T. Biases in the predictions of genuine $\not\!\!E_T$ may enter from the small residual differences between the charged lepton and neutrino $p_{\rm T}$ spectra. A jet energy scale offset would also lead to a bias since jet energy mis-measurements from the offset add up coherently along the V-direction in the transverse plane, while in QCD events, to first order, they cancel [1]. The method's susceptibility to these and other biases in artificial and genuine $\not\!\!E_T$ predictions can be studied in both data and MC. I will discuss briefly how this can be done next.

A sample of events with $N_J = 2$ for a small jet $p_{\rm T}$ threshold, such as 50 GeV, is more sensitive to biases, and the relative contribution from new physics can not be large in this sample. Therefore, these events can be used to validate the method's performance and place an upper bound on its biases at higher N_J and jet $p_{\rm T}$. Similarly, the application of event quality criteria are expected to reduce the number of severely mis-reconstructed events that may lead to biases in the prediction of artificial $\not\!\!E_T$. By varying the event quality selection criteria, one can determine if the method is subject to such biases or estimate their size. An excess due to a new physics contribution should be stable under variations of these criteria.

There are several sources of systematic uncertainty associated with the modeling of the neutrino $p_{\rm T}$ spectra by the charged lepton $p_{\rm T}$ spectra if MC is used to obtain corrections to the charged lepton $p_{\rm T}$ spectra. Because these corrections are small, uncertainties due to MC used to extract them enter only at second order. They can be estimated by varying the composition of the MC samples used to measure them and the reconstruction efficiencies of leptons and jets within their uncertainties. The uncertainties in the composition of the MC samples should come from an independent measurement of the relative W+jets and $t\bar{t}$+jets cross sections for different N_J. Note, in section 6.2 it is demonstrated that these corrections may be negligible for $N_J \geq 4$ at high $\not\!\!E_T$ in early data.

The QCD background to signal events with one or more fake leptons and cross-feeds among V+jets processes are not considered in this note. These backgrounds are expected to be small but need to be measured and accounted

for in the $\not\!\!E_T$ distributions and their predictions [6]. A large new physics contamination to QCD at large J_T, in general, may bias the prediction at large $\not\!\!E_T$ and hide a new physics contribution to V+jets. However, I find that even under the most optimistic scenarios for new physics cross sections such a contamination does not lead to a significant bias.

Even though the reliance on MC is much reduced in this method, MC can be used to validate the method and constrain its systematic biases as is done in Ref. [1]. Nonetheless, a study of control data samples is crucial for developing, optimizing and validating the final algorithm and quantifying its systematic uncertainties.

6.5 Early data commissioning plan

Since the QCD cross section is very large, only a small data sample is needed to construct $\not\!\!E_T$ templates. In early days of data taking, when the amount of integrated luminosity or the collision center-of-mass energy preclude a meaningful search, the templates can be used to help commission $\not\!\!E_T$. By comparing templates in the data with templates in MC, one can detect, identify and, perhaps, understand features and artifacts that lead to disagreement between the two sets of templates. These comparisons can be made for different N_J, J_T, jet $p_{\rm T}$ thresholds, $\eta_{\rm lead\ jet}$ and other variables. Different templates are likely to reveal different effects that lead to difference between the data and MC. By studying a large pool of templates, one may detect a variety of such effects, and obtain a comprehensive assessment of the $\not\!\!E_T$ performance and limitations. This study will also help remove severely mis-reconstructed events that can bias the method's predictions of artificial $\not\!\!E_T$.

The most critical aspect of the method in the l+jets+$\not\!\!E_T$ channel is the reliability and robustness of genuine $\not\!\!E_T$ modeling based on the charged lepton spectra. For reliable modeling, it is desirable to develop a set of selection criteria producing uniform muon reconstruction, identification and isolation efficiencies as a function of muon $p_{\rm T}$ to avoid correcting for non-uniform efficiencies. The muon efficiencies can be measured via a standard tag-and-probe technique with $Z \rightarrow \mu^+\mu^-$ decays [5][6], which is an essential tool for developing the muon selection. The $Z \rightarrow l^+l^-$+jets are also useful either to validate the jet energy scale corrections or to measure them independently.

The $N_J = 2$ sample in l+jets+$\not\!\!E_T$ is a good validation sample of the method since the relative amount of new physics contamination in this sample is small and the method is more susceptible to biases in events with small number of jets. In early data, when there is not enough data in high N_J bins, the method can be applied to events with $N_J = 2$. In particular, since W+jets dominates in that N_J bin, the procedure of applying corrections for the muon $p_{\rm T}$ threshold can be investigated. With these corrections, many aspects of the analysis can be tested and compared to MC simulation for $N_J = 2$.

References

[1] "Modeling missing transverse energy in V+jets at CERN LHC", arXiv::0906.5016; CMS AN 125/2009.

[2] G.L. Bayatian *et al.* (CMS Collaboration), Report No. CMS TDR 8.1, CERN/LHCC 2006-001; G. Aad *et al.* (ATLAS collaboration), arXiv:0901.0512.

[3] The W's (W^+ and W^- combined) in W+jets for $|\eta_{\rm charged\ lepton}| < 2$ are effectively unpolarized in the transverse plane. Effects from the residual W-polarization in the transverse plane are negligible in l+jets+$\not\!\!E_T$ for $N_J \geq 3$.

[4] G.L. Kane, G.A. Ladinsky, C.P. Yuan, Phys. Rev. D **45**, 124 (1992); M. Fischer, S. Groote, J.G. Korner and M.C. Mauser, Phys. Rev. D **63** 031501(R) (2001); V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **100**, 062004 (2008); T. Aaltonen *et al.* (CDF Collaboration), Phys. Lett. B **674**, 160 (2009).

[5] For example, G. Daskalakis *et al.* CMS AN-2007/019

[6] A. Abulencia *et al.* (CDF Collaboration), J. Phys. G: Nucl. Part. Phys. **34**, 2457 (2007); D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **94**, 091803 (2005).

7 The α_T jet-balancing method

Following the promising results of the α_T approach to all-hadronic SUSY searches, we present a generalization of the method for the search for SUSY in the exclusive single-lepton plus jets and missing energy channel. The

main motivation for the usage of the α_T variable remains the reduction of the QCD background and the associated reliable control over this background, especially in the early SUSY searches. This reduction and control of the QCD background is particularly important for the one-lepton analysis when the lepton transverse momentum is lowered to values as low as 5 GeV. This, in turn, would allow to cover more of the important soft lepton parameter space predicted by many SUSY models.

The present analysis is applied to the case of an integrated luminosity of 100pb^{-1} and demonstrates a possible way to establish a robust deviation of New Physics from the Standard Model expectations utilizing the centrality of the leading jet in the SUSY signal region.

7.1 Description of Method

7.1.1 Event selection

A "single-lepton" event is defined through a series of selection requirements on the basic physics objects, i.e. the electrons, muons, jets and the missing transverse energy. A concrete list of these requirements, including acceptance as well as identification (ID) and quality criteria, have been proposed by the Single-lepton Reference Analysis (RA4). In the scope of this analysis, we follow the guidelines proposed in RA4, while for some specific selection requirements we investigate possible changes which improve the analysis within the context of usage of the α_T variable to control the QCD background. One example is the usage of a cross cleaning tool which is used to correct the energy balancing in events with overlapping objects[1)].

However, the main modifications with respect to the RA4 guidelines are the lowering of the lepton momenta (down to 5 GeV threshold) as well as a redefinition of the lepton isolation - applicable to a soft lepton selection.

7.1.2 Lepton isolation

Lepton isolation is a key tool in reducing backgrounds from fake or heavy-flavor electrons and muons – collectively referred to as "QCD sources" in what follows. The standard recommendations for lepton isolation are provided by the V+jets Cross PAG group [1], and have been shown to work well for electrons and muons with momenta above ~ 30 GeV. For lower lepton p_T, however, the standard isolation selection has a significant impact on the efficiency of both electrons and muons. For this reason, a different isolation is considered for leptons in the region $p_T^\ell < 30$ GeV ("soft lepton" region). An optimization of the electron and muon isolation performance in the soft p_T region has been developed elsewhere [2]. The isolation variables and the cut values chosen are based on comparisons between SUSY signal efficiency and QCD background rejection.

In the high p_T region, electrons and muons are required to pass the combined relative isolation criterio according to the V+jets recommendations: $\text{CombIso}_{rel} = \sum_{\Delta R<0.3} p_T^{\text{track}}/p_T^\ell + \sum_{\Delta R<X} E_T^{\text{ECAL}}/E_T^\ell + \sum_{\Delta R<Y} E_T^{\text{HCAL}}/E_T^\ell < 0.1$, where $\sum_{\Delta R<0.3} p_T^{\text{track}}$ is the sum of the transverse momenta of the tracker-tracks in a cone ($\Delta R < 0.3$ for both electrons and muons) around the lepton direction. $\sum_{\Delta R<X} E_T^{\text{ECAL}}$ and $\sum_{\Delta R<Y} E_T^{\text{HCAL}}$ are the sums of the energy deposits in a cone ($\Delta R < 0.3$ for muons and $\Delta R < 0.4$ for electrons) around the lepton direction in the electromagnetic and hadronic calorimeter respectively.

In the soft lepton p_T region, we choose to use only the tracker based isolation since this is expected to perform more reliably at least for the early phase of the LHC operation. A cut on the absolute tracker isolation is set to 3 GeV for electrons and 5 GeV for muons:

$$\text{TrkIso}_{\text{abs}} = \sum_{\Delta R<0.3} p_T^{\text{track}} \begin{cases} < 3 \text{ GeV} & (\text{for electrons}) \\ < 5 \text{ GeV} & (\text{for muons}) \end{cases} \tag{24}$$

7.1.3 The α_T variable

In the N-jet all-hadronic analysis [3], the α_T variable has been redefined in such a way so as to reproduce, or "simulate", the kinematics of a di-jet system in a typical QCD event. The idea is to construct two "pseudo-jets" which balance each other in H_T, where the pseudo-jet H_T equals the scalar sum of the transverse momenta p_T

[1)] Electron-jet cleaning (within cone of 0.5): i) if the electron is isolated and the ratio of shared energy to electron energy is above 0.7, then the jet is removed, whereas ii) if the electron is non-isolated and the shared energy is above a threshold, then the electron is removed. Muon-Jet cleaning (within cone of 0.2): if the muon is not isolated and close to a jet, then the muon is dropped and its energy added vectorially to the jet.

of all the jets comprising the pseudo-jet. Jets are combined into pseudo-jets by minimizing the variable $\Delta H_T = |H_{T,1} - H_{T,2}|$. In this approach, the α_T variable is written as:

$$\alpha_T = \frac{1}{2}\frac{H_T - \Delta H_T}{M_T} = \frac{1}{2}\frac{H_T - \Delta H_T}{\sqrt{H_T^2 - MH_T^2}} \tag{25}$$

For a perfectly balanced system, we expect $\Delta H_T = 0$; in practice, mis-measurements of the jet energies as well as the exclusion of physics objects (in this case jets) due to acceptance or the quality cuts, can make $\Delta H_T > 0$. In this sense, ΔH_T is a measure of each kind of instrumental effect that distorts the momentum balance of the N-jet system.

The single-lepton analysis extends the definitions of the kinematic variables ΔH_T, MH_T, H_T and eventually α_T, to include the lepton object in addition to the jets.

The "leptonic" version of the α_T variable is intended to control the QCD background that survives the one-lepton selection due to fake leptons or leptons from heavy-flavor decays. It has been found to maintain the good performance in controling the QCD background as in the all-hadronic channel. This is illustrated in fig. 7.32, which shows the correlation between the ΔH_T and MH_T in the one-lepton channel, for the SUSY signal and the QCD N-jet background. One can notice that in the case of QCD (right plot), the correlation grows strong across the diagonal where the severe mismeasurements appear: large values of ΔH_T are produced along with large values of MH_T.

The functional form of the (leptonic) α_T is shown on the same figure for constant values equal to 0.55. It can be seen that the curve $\alpha_T > 0.55$ is able to reject all of the QCD events, as expected.

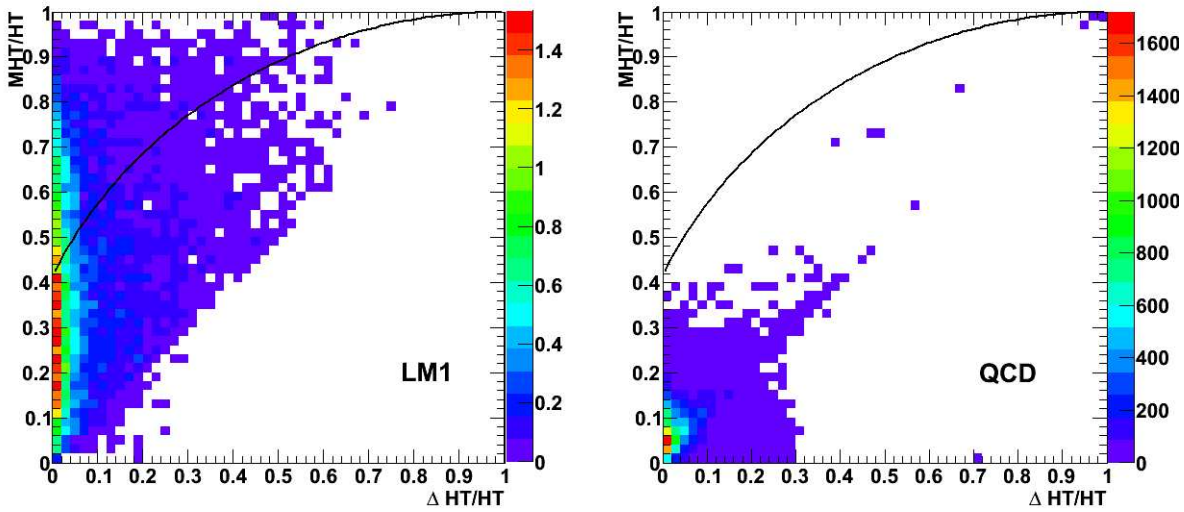


Figure 7.32: *The correlation of $\Delta H_T/H_T$ with MH_T/H_T in SUSY LM1 events (a) and QCD events (b), in the 1-lepton mode channel. An H_T cut of 350 GeV has been applied. The black line indicates constant values of $\alpha_T = 0.55$.*

7.2 Performance of method

In this section we describe a method for establishing a deviation from the Standard Model using the α_T variable.

7.2.1 Analysis cut flow

The single-lepton analysis cut-flow starts with the requirement of exactly one "good" lepton (electron or muon), with $p_T > 5$ GeV, in the event. The jet cuts are driven by the all-hadronic N-jet analysis and constist of requiring at least two jets with $p_T > 30$GeV and $|\eta| < 3$. The second jet must have $p_T > 100$GeV. An H_T cut at 350 GeV is also applied, in order to select events with a significant amount of hadronic-jet activity relevant to the SUSY environment. The final step in the selection consists of the requirement that the variable α_T should be above 0.55. This cut is expected to suppress (almost) all of the QCD N-jets (including $b\bar{b}$ + jets) events, as explained earlier.

For the sake of completeness, a comparison of the event yields between the α_T cut-flow and the RA4 selection, normalized for 100pb^{-1} of integrated luminosity, is presented next.

7.2.2 Comparison with RA4 results

The RA4 selection follows a more traditional approach of cutting on the missing transverse energy. Monte Carlo (MC) studies, have shown that the missing energy variable (whether it is MET or MHT) provides a fairly good separation between SUSY signal events[2)] and most of the SM background processes (W/Z + jets, $t\bar{t}$ + jets and QCD). The RA4 selection thus uses a direct cut on MET. Although that cut is initially rather loose (set to 100 GeV), we alternatively try to tighten the cut to 200 GeV, which seems to be an optimal cut value for numerous LM SUSY points, for the purpose of comparison with the α_T approach.

The event pre-selection for the two analysis paths (α_T versus RA4), is rather different in the two approaches:

a_T **approach:**
1. Exactly one muon or one electron with $p_T > 5$ GeV, while vetoing a second different-flavor lepton. Both electron and muon objects are required to satisfy the customly proposed isolation.
2. Veto on events with: i) a second lepton not passing the quality as well as isolation criteria, ii) at least one jet outside the eta acceptance.
3. At least two jets with $p_T > 30$ GeV, and the second leading jet with $p_T > 100$ GeV.
4. A scalar transverse energy sum of jet $H_T > 350$ GeV.

RA4 approach:
1. Exactly one muon or one electron with $p_T > 10$ GeV, while vetoing a second different-flavor lepton. The isolation imposed on the electron and muon objects is taken from the standard V+jets recommendation (relative combined isolation).
2. At least three jets with $p_T > 30$ GeV, and the third leading jet with $p_T > 50$ GeV.

The number of events expected with 100pb^{-1} of integrated luminosity, is calculated at each step in the cut flow, for all the SM backgrounds and the LM0 and LM1 SUSY signals. Tables 7.8 and 7.9 display the event yield in the one-muon channel, for the α_T and RA4 approaches respectively, whereas tables 7.10 and 7.11 show the corresponding numbers for the one-electron channel. As an alternative to the α_T variable, the tables also display the event yields using a direct cut on the MH_T/H_T variable. The latter shows an overall performance which is quite similar to that of the α_T-based selection. The final event yields in the different approaches are compared in terms of Signal-to-Background ratio (S/B) and the signal significance ($S/\sqrt{B}$). The SUSY signal used in "S" is that corresponding to the LM0 point.

It can be seen that, the input lepton selection differs significantly between the RA4 and α_T approaches as a result of the different lepton p_T thresholds and the isolation requirements. In both muon and electron channels, the dominant background contributions come from the W+jets and $t\bar{t}$+jets events. An interesting effect is that in the case of the α_T (or equivalently the MH_T/H_T) selection, the W background is higher (by almost a factor two) than the $t\bar{t}$ background. This is opposite to the RA4 selection result where $t\bar{t}$ is the dominant background. The effect is due the inclusion of the two-jet bin in the α_T selection, which therefore maintains a large amount of W events, as opposed to RA4 which has higher multiplicity requirement that favors $t\bar{t}$. In addition to this enhancement, a cut on the relative missing energy, i.e. MH_T/H_T or α_T, enhances the W component[3)]. For what concerns the QCD/$b\bar{b}$ backgrounds, these are drastically suppressed only after the final step in the selection (α_T or ME_T cut). The striking result is the strong suppression of the QCD using the α_T cut, leaving zero events after a cut on α_T.

Using LM0 as the SUSY signal, the S/B shows comparable performance between the α_T approach and RA4, whereas an improved significance (loosely measured via $S/\sqrt{B}$) is observed using the RA4 cut-flow. This is partly due to the reduced signal yield in the α_T approach – which is the main drawback of the method. Nevertheless, as it will be shown in the next sections, the α_T presents a major advantage in that the selection is very robust against jet energy mismeasurements (mainly affecting the copiously produced QCD jet events).

7.2.3 The jet-eta versus HT kinematic method

A significant fraction of the SUSY signal is expected to be present in high HT values. As shown in the previous section, the α_T variable is very powerful in separating the signal and background events in two regions. The

[2)] in R-parity conserving models

[3)] For a high H_T cut (350 GeV), MH_T/H_T falls more rapidly for the $t\bar{t}$ events than the W's in the tails of such a distribution.

Selection cut	QCD	bb	Z	W	$t\bar{t}$	LM1	LM0	S/B	$S/\sqrt{B}$
μ-selection	2414.2	619.	59095.5	716527.	4502.9	175.7	1247.6		
3-jet cut	649.6	160.9	106.6	987.9	1639.	95.3	752.8		
$ME_T > 100$	1.3	1.9	8.8	176.5	356.1	85.6	498.3	0.6	15.2
$ME_T > 180$	$7 \cdot 10^{-2}$	0	0.9	37.3	51.5	66.7	232.2	2.6	24.5

Table 7.8: *Event-yield out of the RA4 cut-flow, normalized to* $100pb^{-1}$*, in the one-muon channel.*

Selection cut	QCD	bb	Z	W	$t\bar{t}$	LM1	LM0	S/B	$S/\sqrt{B}$
μ-selection	196044.2	38303.2	66993.5	784266.	5504.	244.6	1665.6		
veto cuts	123817.	23276.3	57322.6	752253	2898.8	150.3	832.3		
2-jet cut	36574.	6418.7	72.1	908.7	552.2	94.9	404.		
$H_T > 350$ GeV	14653.6	2564.	43.1	594.3	475.3	92.9	389.		
$\alpha_T > 0.55$	0	0	0.6	13.9	6.6	23.4	38.2	1.8	8.3
$MH_T/H_T > 0.4$	0.3	0.7	0.6	33.2	15.2	44.1	76.6	1.5	10.8

Table 7.9: *Event-yield out of the* α_T *cut-flow with leptons of* $p_T > 5$ *GeV, normalized to* $100pb^{-1}$*, in the one-muon channel.*

Selection cut	QCD	bb	Z	W	$t\bar{t}$	LM1	LM0	S/B	$S/\sqrt{B}$
e-selection	6405.1	258.1	55822.2	624817.	3732.5	123.6	909.6		
3-jet cut	1219.5	52.8	125.7	895.2	1309.3	63.9	532.		
$ME_T > 100$	1.6	0.2	2.6	147.8	262.8	57.3	352.4	0.8	17.3
$ME_T > 180$	0	$4 \cdot 10^{-3}$	0	28.3	38.8	45.	160.5	2.4	19.6

Table 7.10: *Event-yield out of the RA4 cut-flow, normalized to* $100pb^{-1}$*, in the one-electron channel.*

Selection cut	QCD	bb	Z	W	$t\bar{t}$	LM1	LM0	S/B	$S/\sqrt{B}$
e-selection	27905.4	1751.1	65216.2	755410.	4068.27	164.0	1057.8		
veto cuts	18155.2	1064.5	31401.	729891.	2191.8	100.5	537.2		
2-jet cut	6609.1	397.6	75.	891.9	396.0	62.4	252.2		
$H_T > 350$ GeV	2694.8	142.4	52.2	599.2	344.8	61.0	244.2		
$\alpha_T > 0.55$	0	0	0	9.0	5.2	16.1	25.1	1.8	6.7
$MH_T/H_T > 0.4$	1.3	0.3	0	26.3	11.4	30.0	50.8	1.3	8.1

Table 7.11: *Event-yield out of the* α_T *cut-flow with leptons of* $p_T > 5$ *GeV, normalized to* $100pb^{-1}$*, in the one-electron channel.*

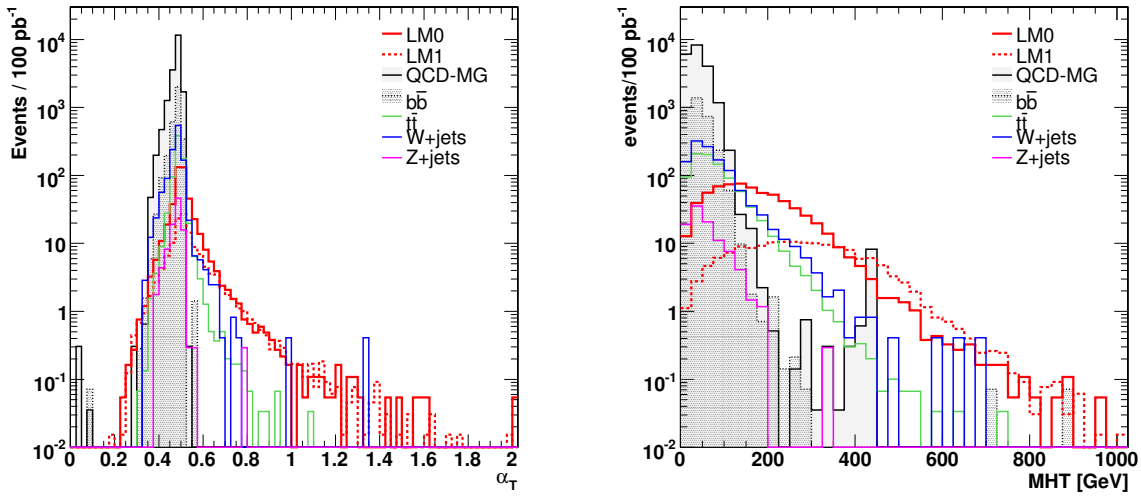


Figure 7.33: *The* α_T *(a) and* MH_T *(b) distributions for the LM0 and LM1 SUSY signal and all the SM backgrounds superimposed, for an integrated luminosity of* $100pb^{-1}$*.*

QCD (and $b\bar{b}$) backgrounds can be controlled and, in fact, can be totally rejected by requesting that events satisfy $\alpha_T > 0.55$. The above, plus the different eta dependence between signal and SM background samples[4)], lead to establish the following analysis strategy.

Fisrt, we introduce the variable R_{α_T} which is defined as the ratio of the number of events passing the α_T cut over the number of events failing it:

$$R_{\alpha T} = \frac{N(\alpha_T > 0.55)}{N(\alpha_T < 0.55)} \tag{26}$$

We then study the behavior of $R_{\alpha T}$ as a function of the leading jet $|\eta|$. This is done in different regions of H_T: it is expected that at low values of H_T the ratio will be dominated by Standard Model processes, whereas at high values the SUSY signal will be relatively more prominent. The basic idea is, therefore, to establish a different behavior of $R_{\alpha T}$ vs $|\eta|$ as we move from the background-dominated region (low-H_T) to the potentially signal-rich region (high H_T). Figure 7.34 illustrates this effect with the $R_{\alpha T}$ ratios plotted as a function of the leading jet eta, for three different regions in H_T, assuming a background-only hypothesis (left plot) and a signal-plus-background hypothesis (right plot). A quantitative procedure for establishing such deviation from the SM expectations is currently under development[5)].

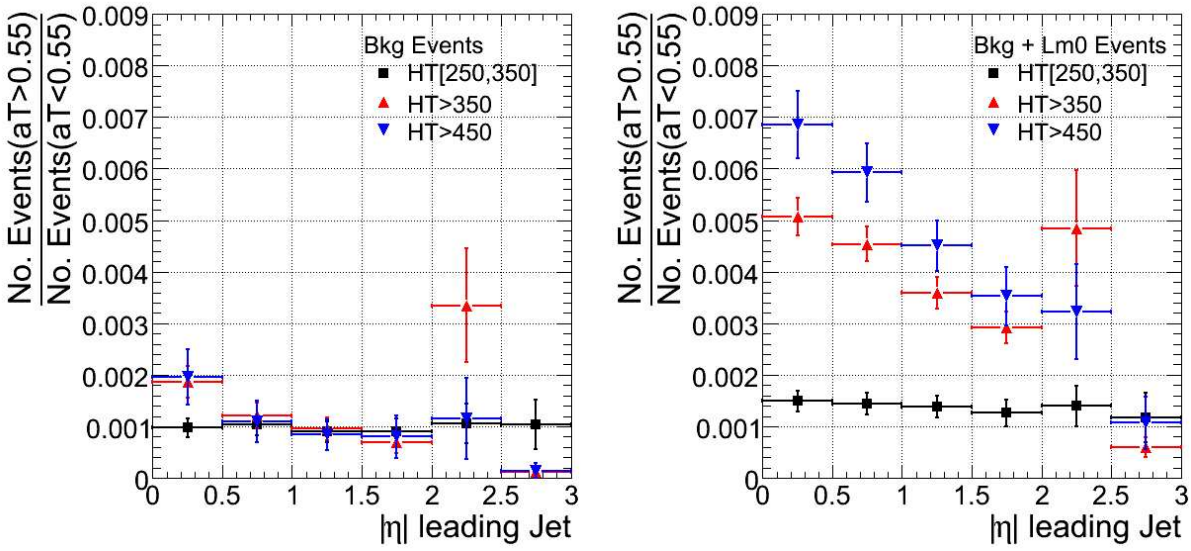


Figure 7.34: *The $R_{\alpha T}$ versus the leading jet $|\eta|$ for the SM background-only hypothesis (left) and the SUSY signal plus SM background hypothesis (right) - in three H_T bins [250, 350], [350, inf], [450, inf].*

7.3 Systematic uncertainties

Two differenct aspects of systematic uncertainties have been considered, both related to the jet reconstruction. The first one addresses moderate systematic variations in the event yield due to imperfect modeling of the jet energy resolution in MC and uncertainties in the jet calibration. This effect has been studied in detail elsewhere [4]. The second is an attempt to quantify the robustness of α_T under drastic jet mismeasurements. This is done by estimating the frequency with which jet mis-measurements have to occur but would still allow a SUSY discovery significance of at least 5 sigma. This study is presented in the following subsection.

7.3.1 α_T stress test

Figure 7.35 shows the effect of different jet mismeasurements on the distributions of α_T and MH_T, for the QCD and $b\bar{b}$ samples. To an extreme scenario, one jet with transverse momentum above 30 GeV is randomly selected per event and scaled by a constant factor F. The factors tested are 0.1, 0.3, 0.5, 2 and 3. In all cases, the tails of

[4)] SUSY processes are expected to show more central behavior with the leading jet η, than the QCD and EWK background processes which appear rather flatish.

[5)] First studies on the issue can be found here [4].

the MH_T distribution are dramatically affected by the jet mismeasurements, while there is relatively little effect on α_T.

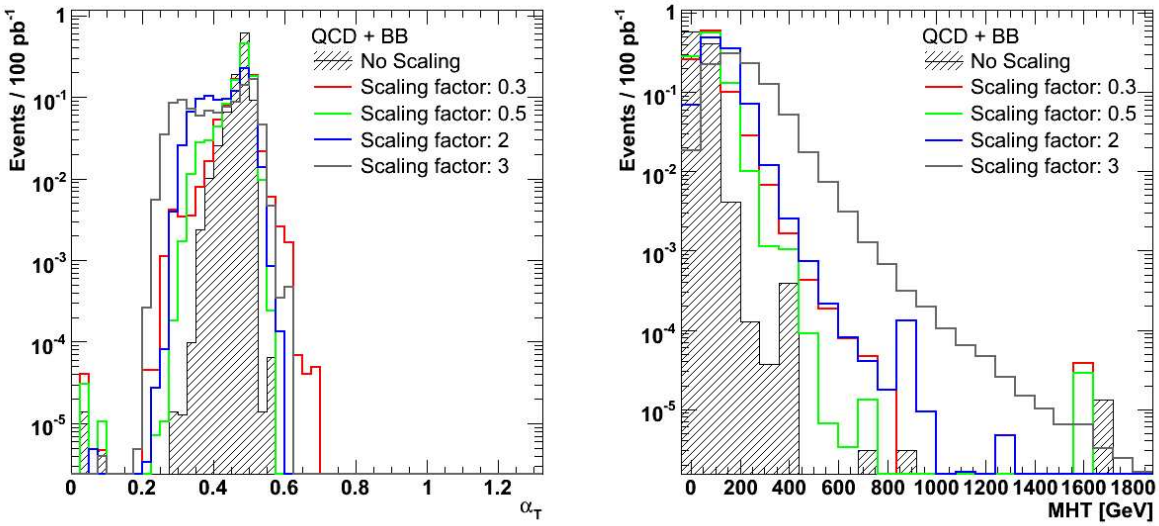


Figure 7.35: *The α_T (a) and MH_T (b) distributions after a drastic jet rescaling by different factors in QCD events. The results correspond to the extreme scenario of rescaling one jet per event.*

In order to obtain a rough measure of the performance of the α_T versus the MH_T approach under drastic mismeasurements of the jet energies, the following stress test was used: assuming a zero uncertainty on the measurement of the EWK background, a systematic uncertainty, ΔB, assigned to the QCD component, would change the significance, $S/\sqrt{B}$, by $S/\sqrt{B+\Delta B^2}$. A maximum uncertainty on the QCD background (ΔB) which would still allow a SUSY discovery with $\sim 5\sigma$, can then be estimated for each final cut - an α_T or an MH_T cut. In practise, one needs to calculate the ΔB uncertainty on QCD as a function of the probability of mis-measurement, as shown on fig. 7.36. For each given mis-measurement fraction (scaling factor F) considered, the maximum rate for drastic jet mis-measurements, that would still allow a SUSY discovery, is provided in table 7.12. Obviously, the higher the frequency by which the jet mis-measurements need to occur, the more robust the cut variable is. This favors α_T performance in almost all the cases. Rescalings by 0.1 and 0.3 factors show comparable performance between the high MH_T cut and the α_T one, which can be understood by the indirect effect of lack of statistics (a downward scaling of the jet energies results in rejecting events due to the H_T cut).

Therefore, the above results have demonstrated that the α_T cut performs more resilient under possibly large QCD background uncertainties.

cut	ΔB	**F=0.1**	**F=0.3**	**F=0.5**	**F=2**	**F=3**
$MH_T > 100$ **GeV**	167.5	$< 1/10000$	$< 1/10000$	$< 1/10000$	$< 1/10000$	$< 1/10000$
$MH_T > 200$ **GeV**	77.0	1/25	1/13	1/6	1/350	1/5000
$\alpha_T > 0.55$	11.5	1/50	1/15	$> 1/5$	1/20	1/77

Table 7.12: *Maximum rate of drastic jet mis-measurements, that would still allow the "SUSY significance" to remain > 5, for different jet p_T scaling factors:0.1, 0.3, 0.5, 2 and 3.*

Overall, despite the reduced event-yield of the α_T approach, the analysis favors in terms of robustness and reliability to control the most challenging background for LHC (QCD). In parallel, this approach avoids questions of ME_T/MH_T-based analyses, like where to place the cut and how to control the QCD tails and therefore can be considered as a reliable alternative method to the traditional RA4-like approaches.

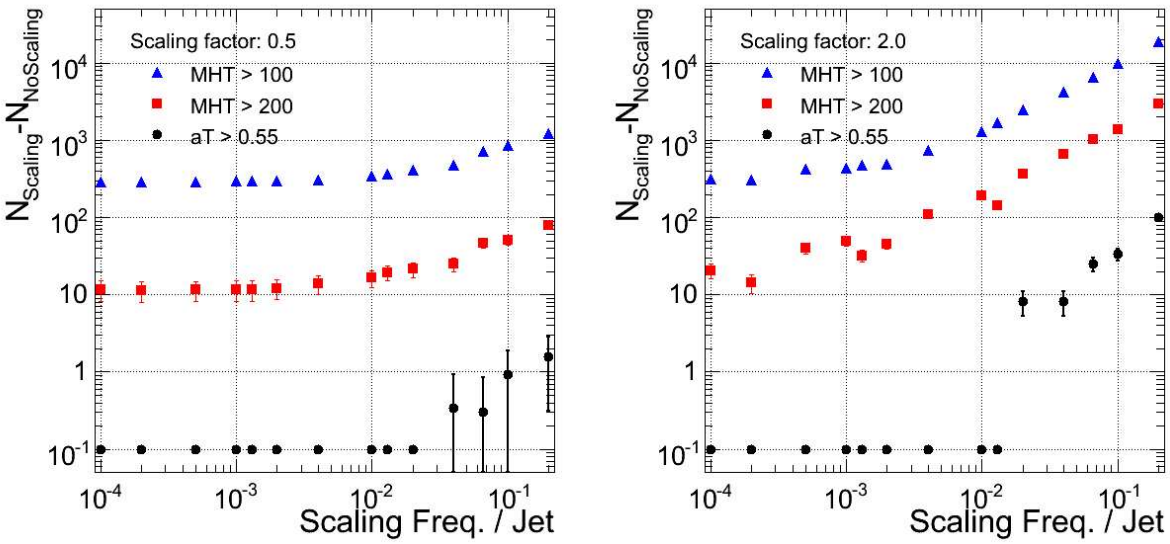


Figure 7.36: *Uncertainty on QCD background (ΔB) as a function of the probability of mis-measurement, for two different scaling factors of the jet energy: 0.5 (left) and 2 (right).*

References

[1] https://twiki.cern.ch/twiki/bin/view/CMS/VplusJets

[2] **CMS AN-2009/167,** ***"Study of isolation properties of SUSY low-p_T leptons."***

[3] **CMS AN-2008/082**, *"Search for SUSY with exclusive n-jet events"*

[4] **Note in preparation,** ***An application of the α_T jet-balancing method to the one-lepton mode SUSY searches.***

8 $t\bar{t}$ background using $\chi^2(t\bar{t})$

8.1 Introduction

In a SUSY leptonic search requiring exactly one muon (w/ $p_T > 10$ GeV), zero electrons, and the tight jet cuts described in section 3.3, the largest Standard Model background is from $t\bar{t}$ events, as can be seen in Table 4.6. This study focuses on determining the $t\bar{t}$ background using a new variable, χ^2, which has large discriminating power between $t\bar{t}$ and SUSY and is explained in detail in section 9.2. Section 9.3 explains how the χ^2 variable is used along with the variable $\not{E}_T / \Sigma E_T$ in an ABCD method to predict the number of expected $t\bar{t}$ background events. Section 9.3 also describes the results of the prediction in the presence of the other major background, W+jets (QCD is left to section 25), and signal contamination. Section 9.4 then describes the statistical and systematic uncertainties in the method using a very large $t\bar{t}$ sample and running many pseudoexperiments. Section 9.5 goes on to describe an early data commissioning plan.

The muon cuts and electron veto cuts used in this study are the same as the V+jets group recommendation cuts listed in Sections 3.1 and 3.2. The muon is required to have a p_T greater than 10 GeV, and an electron considered for veto purposes must have a p_T greater than 10 GeV. The jet cuts used in this study are slightly different than the proposed tight jet cuts described in Section 3.3. In this study we require three jets of $p_T \geq 80$ GeV and a fourth jet of $p_T \geq 30$ GeV, where a jet is defined as:

- $p_T \geq 30$ GeV
- $|\eta| \leq 2.5$
- Hadronic Energy Fraction ≥ 0.1

The samples used in Sections 9.2 and 9.3 are from the Summer08 and Fall08 Monte Carlo production. The large statistics FastSim $t\bar{t}$ sample used in Section 9.4 for statistical and systematic studies was produced in CMSSW 2.2.6. The cross sections used for each sample are listed in Table 4.3. The Physics Analysis Toolbox (PAT) was run in CMSSW 2.2.6 with the additional tags:

- PhysicsTools/PatAlgos V04-14-24
- PhysicsTools/PatUtils V03-05-02
- DataFormats/PatCandidates V03-18-07
- CondFormats/JetMETObjects V01-08-02

and then private ntuples were made from the PAT collections.

8.2 Description of χ^2 variable

To estimate the $t\bar{t}$ background, we construct a new variable, χ^2 , which uses the hypothesis of semi-leptonic $t\bar{t}$ for each event to reconstruct the leptonic top mass, hadronic W mass, and hadronic top mass. The χ^2 variable has been used before in a top cross section analysis at CMS, Ref. [1], and a different form of the χ^2 was used for exotic paricle searches at CDF, Ref. [2]. In our case we construct the χ^2 by choosing the four highest p_T jets in the event and taking the lowest χ^2 permutation as the combination of jets assigned to the hadronic W, b-jet of the hadronic top, and b-jet of the leptonic top:

$$\chi^2(t\bar{t}) = \frac{(M_{j_1 j_2} - M_W)^2}{\sigma_{jj}^2} + \frac{(M_{j_1 j_2 j_3} - M_t)^2}{\sigma_{jjj}^2} + \frac{(M_{W_{\ell\nu} j_4} - M_t)^2}{\sigma_{\mu\nu j}^2} \tag{27}$$

where the weight factors are: σ_{jj}^2 = 10.5, σ_{jjj}^2 = 19.3, $\sigma_{\mu\nu j}^2$ = 21.2. For simplicity the weight factors are determined from Monte Carlo and take into account detector resolution effects for each term.

Each piece of the χ^2 describes a different part of the semi-leptonic $t\bar{t}$ decay and is formed from reconstructed objects. The leptonic top mass is determined by using the $\not{E}_T$, and the jet and muon four-vectors. The $\not{E}_T$ is taken to be the neutrino p_T, and the W mass constraint is used to determine the neutrino p_z. Solving for the neutrino p_z using the muon four-vector and W mass gives a quadratic solution with two solutions. The solution with p_z closest to the lepton is chosen and in the case where the answer is complex (33% of the time in semi-leptonic $t\bar{t}$ and 50% of the time in di-leptonic $t\bar{t}$ events passing the second lepton veto) due to reconstruction effects or a non-semileptonic $t\bar{t}$ event, the real part of the solution is chosen. After determining the W four-vector, it is used along with the jet four-vector to determine the leptonic top mass. The hadronic W mass is determined from the four-vectors of two of the jets in the event. The hadronic top mass is determined from the two jets used for the W hadronic mass and a third jet. The leptonic and hadronic top masses, and hadronic W mass are simulataneously calculated in the χ^2 for each of the different jet permutations. When using 4 jets, there is 12 different combinations of jets in the χ^2. When using 5 or 6 jets in the χ^2 the number of different jet permuatations increases to 60, 180, respectively. We have investigated using more than 4 jets in the χ^2; however, the large number of combinatorics with 5 or 6 jets dilutes the discriminating power of χ^2 between $t\bar{t}$ and SUSY. For the rest of this study we only use the 4 highest jets to calculate the χ^2.

Looking at each piece in the χ^2 there is a clear shape difference betwen $t\bar{t}$ and LM1. Fig. 8.37 shows the hadronic W mass, hadronic top mass, and the leptonic top mass for $t\bar{t}$ and LM1, all plotted for the lowest χ^2 permutation in the event. In the leptonic top mass distibution $t\bar{t}$ has a sharp peak around 175-200 GeV, whereas LM1 peaks much higher and has a much larger tail. LM1 also has a much larger tail in the hadronic W mass distribution, where $t\bar{t}$ peaks at around the W mass. The hadronic top mass distribution also shows LM1 with a much larger tail than $t\bar{t}$ and $t\bar{t}$ also has the interesting feature of a double peak. The first peak is at around the top mass and comes from cases where you get the correct combination of jets for the hadronic top mass. The second peak around 250 GeV is from events where the hadronic top mass is computed with the incorrect combination of jets.

When the hadronic W, hadronic top, and leptonic top pieces are combined with their corresponding weight factors into the χ^2 the discrimination between $t\bar{t}$ and SUSY is even larger than for a single piece of the χ^2. Fig. 8.38 shows the χ^2 distribution for $t\bar{t}$, W+jets, and LM0 through LM9, where all distributions are normalized to one. $t\bar{t}$ has a strong peak in χ^2 at approximately zero and a small tail, whereas the SUSY points have modest or small peak at zero and a large tail. The W+jets χ^2 distribution is somewhere between $t\bar{t}$ and SUSY; W+jets has a smaller peak at low χ^2 than $t\bar{t}$, but larger than any SUSY point. Approximately 50% of $t\bar{t}$ events have a $\chi^2 < 20$, approximately 30% of W+jets events have a $\chi^2 < 20$, and between 2-24% of SUSY events have a $\chi^2 < 20$, depending on the model. The broader SUSY distributions are due to the fact that in SUSY events there should not be a correct combination of jets to satify the semi-leptonic $t\bar{t}$ hypothesis. The two SUSY models with the largest peak at low χ^2 are LM0 and LM9. We have investigated whether these peaks are due to these two models having more top

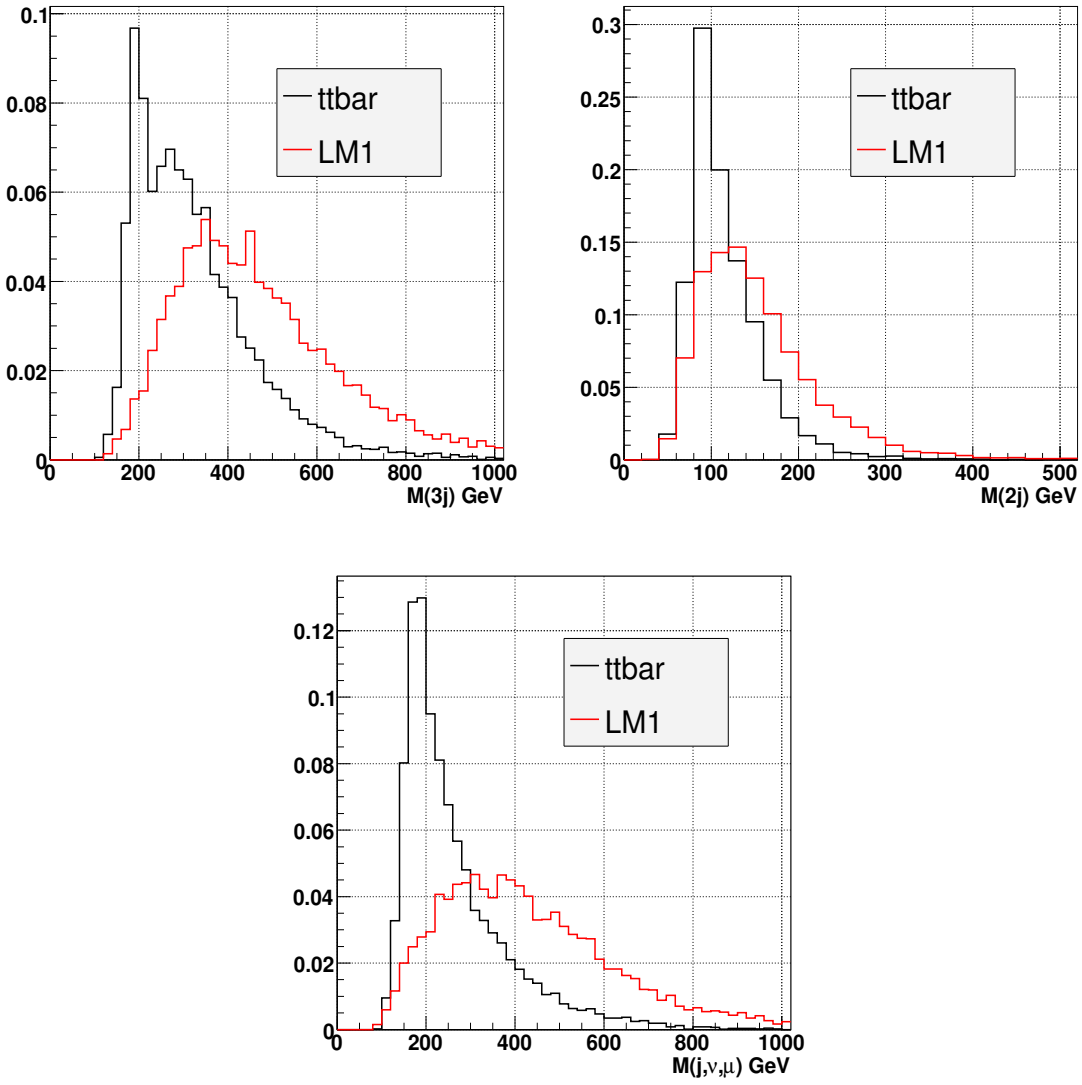


Figure 8.37: Hadronic W invariant mass from two jets, hadronic top invariant mass from 3 jets, leptonic top mass from muon, $\not\!\!E_T$, and a jet. All distributions normalized to one and shown for lowest χ^2 in the event.

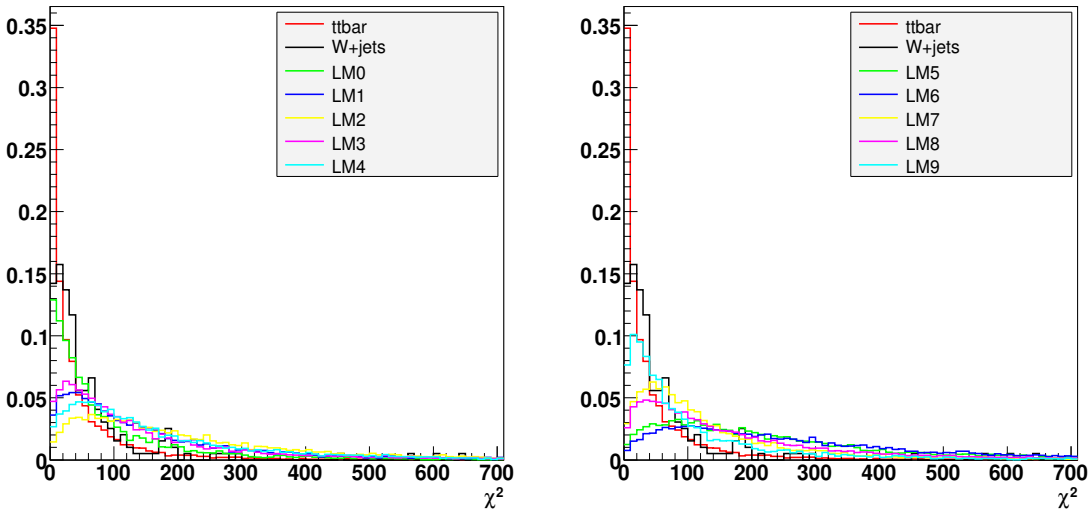


Figure 8.38: χ^2 distribution for ttbar, W+jets, and LM0-LM9. Distributions normalized to 1 and only the best χ^2 for each event is plotted.

quarks, and thus possibly satisfying the χ^2 hypothesis more often. However, Fig. 8.39 shows that LM8 has the most top quarks, not LM9 or LM0. LM8 has a much smaller peak at low χ^2 than either LM0 or LM9; thus, more top quarks in a SUSY event does not necessarily mean smaller χ^2.

Figure 8.39: The number of top quarks per event in SUSY models LM0-9, normalized to 1.

The main contribution to the tail in the $t\bar{t}$ χ^2 distribution is the incorrect assignment of jets in the χ^2. Fig. 8.40 shows that the $t\bar{t}$ χ^2 distribution has a much sharper peak at zero in events where the correct combination of jets is used. Fig. 8.40 also shows that the $t\bar{t}$ χ^2 distribution has a very large tail in events where zero jets are used correctly. Thus, most of of the tail from the $t\bar{t}$ χ^2 comes from events where the jets are not used properly in the χ^2.

Though the incorrect combination of jets can lead to a larger χ^2 in $t\bar{t}$ events, it does not necessarily diminish the discriminating power between $t\bar{t}$ and SUSY. Fig. 8.41 shows the mean of the five lowest χ^2 combinations in each event, for $t\bar{t}$ and LM1. The $t\bar{t}$ distribution has a large peak at low χ^2 while the LM1 distribution has a large tail and little peak at low χ^2. Fig. 8.41 shows that even when you consider the five lowest χ^2 combinations (where at least four combinations in $t\bar{t}$ are inherently incorrect) you still get discriminating power with respect to SUSY LM1. This is because the non-lowest χ^2 combinations in $t\bar{t}$ tend to be close to the lowest χ^2.

The variable χ^2 can clearly be used to discriminate between $t\bar{t}$ and SUSY, and in the next section we will describe the method to determine the number of $t\bar{t}$ background events using the χ^2 variable in combination with other variables. Only the lowest χ^2 combination per event will be considered in the rest of this study.

8.3 Performance of χ^2 vs. $\not{E}_T / \Sigma E_T$ ABCD method

We would like to perform an ABCD method to predict the number of background events in a SUSY signal region. Since the largest background is $t\bar{t}$, we will first look at how the prediction works in pure $t\bar{t}$ events, and then give an estimate for all backgrounds. The obvious first choice for a variable to use along with the χ^2 in an ABCD method is $\not{E}_T$, since it has been shown to have large discriminating power between Standard Model and SUSY. However, if you perform a χ^2 vs. $\not{E}_T$ ABCD method, you run into problems due to correlations between the two variables in $t\bar{t}$ events. Fig. 8.42 shows the $\not{E}_T$ plotted in slices of χ^2 for $t\bar{t}$. This plot shows that the high χ^2 region has a much larger $\not{E}_T$ tail than the low χ^2 regions. This leads to a large underprediction of the background in the signal region (for $t\bar{t}$ with no SUSY the method predicts about half of the $t\bar{t}$ events you actually have in the signal region). We have tried to measure this correlation, but the focus in the rest of this paper will be on the investigation of

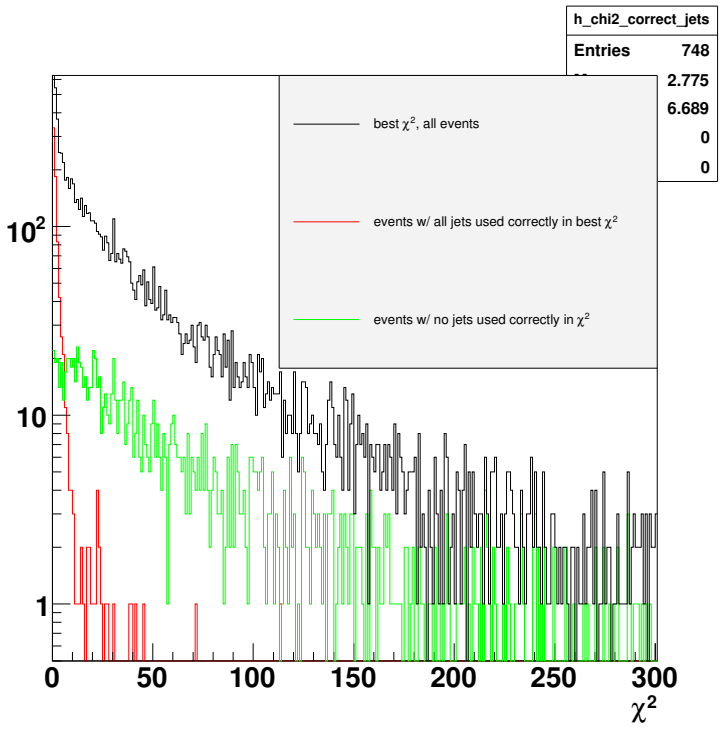


Figure 8.40: χ^2 distribution for all events, events where the best χ^2 gets all the jets correct, and events where the best χ^2 gets all the jets incorrect.

Figure 8.41: The mean of the 5 lowest χ^2 combinations in each event for LM1 and $t\bar{t}$, normalized to one.

another variable that is less correlated with χ^2. We found that $\not\!\!E_T/\Sigma E_T$ has less correlation with χ^2, where ΣE_T is defined to be the sum of $\not\!\!E_T$ and jets in the event. Fig. 8.42 shows the $\not\!\!E_T/\Sigma E_T$ distribution in slices of χ^2 and though the slices differ slightly the high χ^2 slice does not have a much larger tail, as is the case in $\not\!\!E_T$. The correlation factor for $\not\!\!E_T/\Sigma E_T$ and χ^2, -0.07, is 3 times smaller than the correlation factor for $\not\!\!E_T$ and χ^2, 0.20. The only disadvantage to using $\not\!\!E_T/\Sigma E_T$ instead of $\not\!\!E_T$ in an ABCD method with χ^2 is that $\not\!\!E_T/\Sigma E_T$ has a little less discriminating power between SUSY and the background. Fig. 8.43 shows the $\not\!\!E_T$ and $\not\!\!E_T/\Sigma E_T$ distributions after the jet cuts.

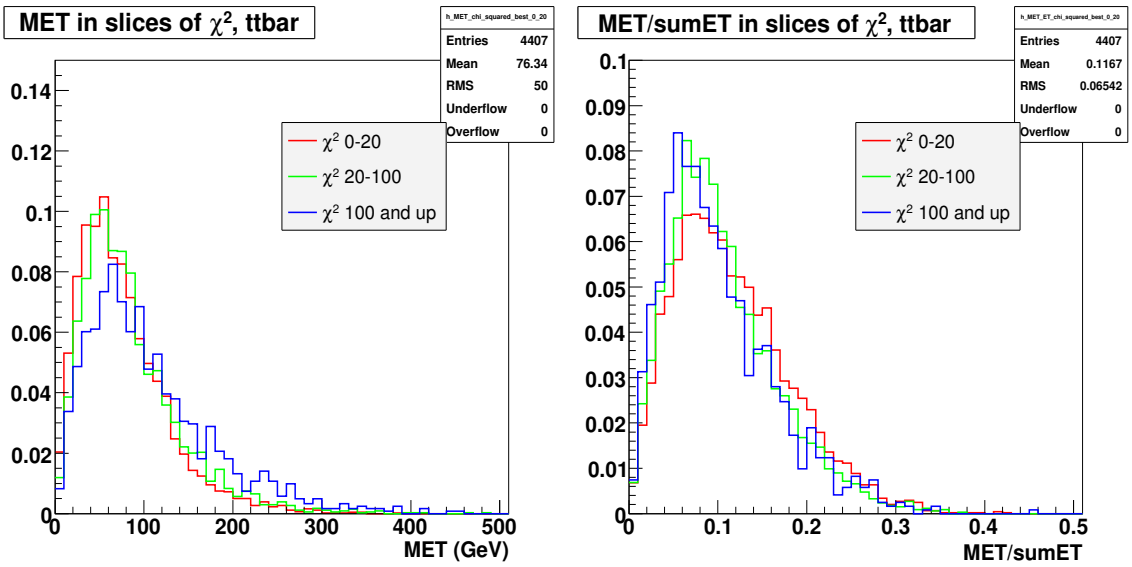


Figure 8.42: $\not\!\!E_T$ and $\not\!\!E_T/\Sigma E_T$ distributions in slices of χ^2 for $t\bar{t}$. Distributions normalized to one.

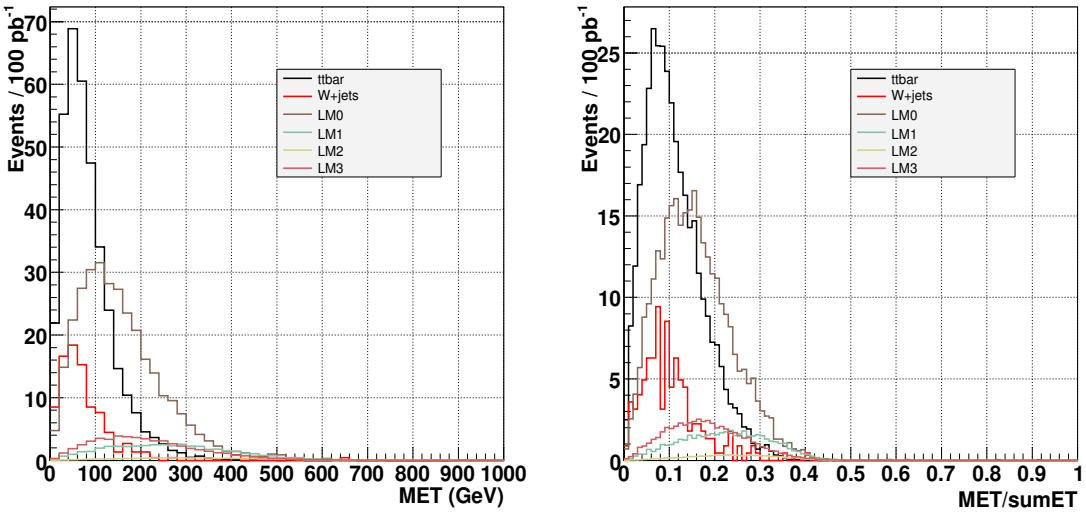


Figure 8.43: $\not\!\!E_T$ and $\not\!\!E_T/\Sigma E_T$ distributions for ttbar, W+jets, and LM0-LM3. Distributions normalized to $100pb^{-1}$.

In addition to studying the correlation between χ^2 and $\not\!\!E_T/\Sigma E_T$ in $t\bar{t}$, we also investigated the composition of the $t\bar{t}$ background as the $\not\!\!E_T/\Sigma E_T$ or χ^2 is increased and found it didn't change significantly. If this where the case, then we could be using single lepton events in the control region to predict dilepton events in the signal region, a problem that plagues the $\not\!\!E_T$ vs. MT ABCD method. Fig. 8.44 shows that the dilepton fraction is mostly flat as you go up in χ^2 and $\not\!\!E_T/\Sigma E_T$. In the signal region 20-25% of $t\bar{t}$ events are dilepton, which is slightly higher than the 10% of events passing the tight jet cuts that are dilepton, but should make a small effect on the prediction.

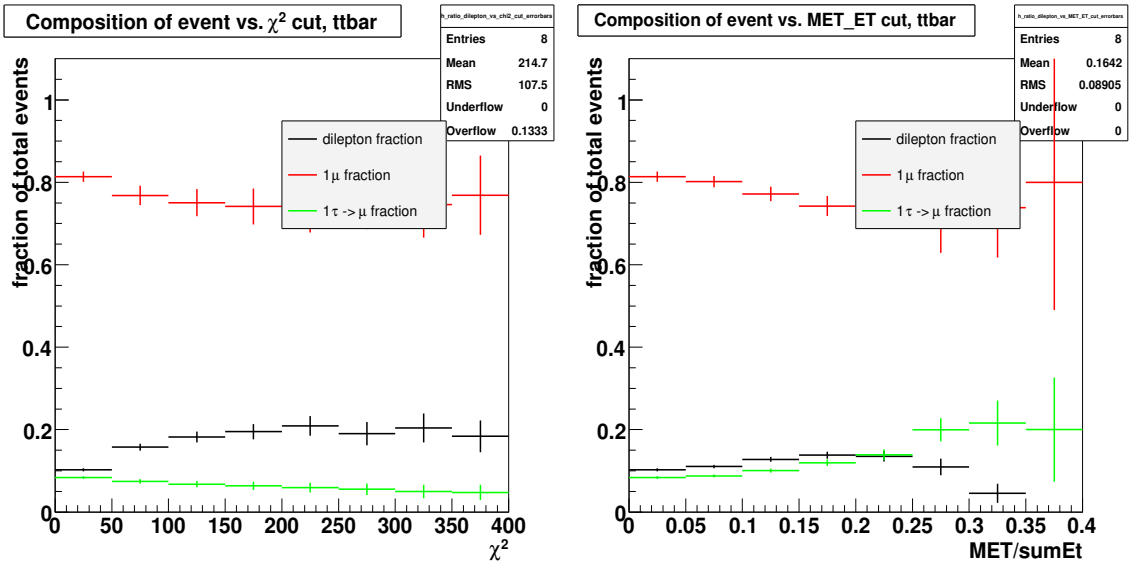


Figure 8.44: Dilepton fraction vs. χ^2 cut, $\not\!\!E_T / \Sigma E_T$ cut.

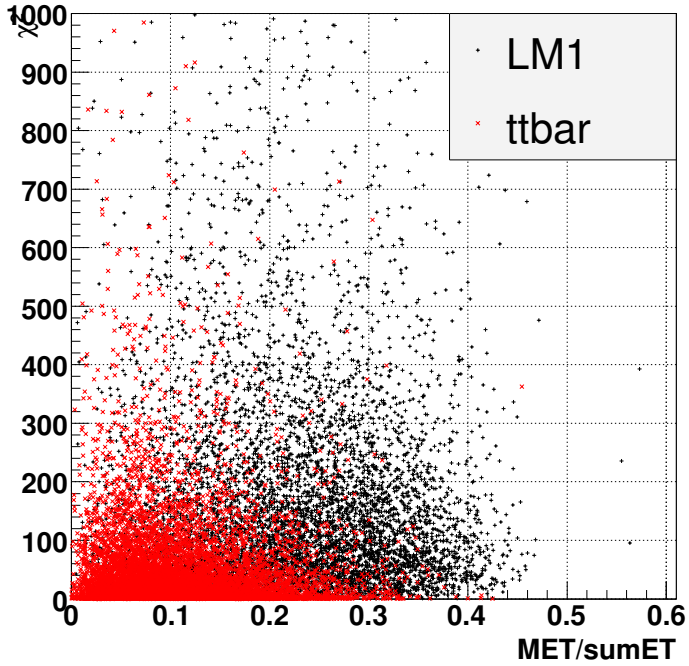


Figure 8.45: χ^2 vs $\not\!\!E_T / \Sigma E_T$ for $t\bar{t}$, LM1.

Table 8.13: The number of background and LM1 events in different ABCD regions. The numbers are all for 100 pb^{-1}.

Cut	LM1	ttbar	W+jets	Z+jets	QCD (Inmu)	single top (tW,t,s)	VV+Jets
$\chi^2 < 20, 0.1 < \not{E}_T/\Sigma E_T < 0.2$	1.2	75.0	15.3	0.6	0.6	2.1	0
$\chi^2 < 20, \not{E}_T/\Sigma E_T \geq 0.2$	2.1	20.5	2.7	0.3	1.46	0.8	0.01
$\chi^2 \geq 20, 0.1 < \not{E}_T/\Sigma E_T < 0.2$	12.5	67.0	22.5	2.5	0	3.7	0.01
$\chi^2 \geq 20, \not{E}_T/\Sigma E_T \geq 0.2$	24.9	15.5	4.5	0.6	0.5	0.7	0.03

Table 8.14: Number of predicted background events in the signal region. The number of SUSY LM1 events in the signal region is 24.9 events. All numbers are normalized to $100pb^{-1}$.

	predicted	actual
$t\bar{t}$ only	18.3	15.5
$t\bar{t}$ w/ LM1 contamination	23.6	15.5
$t\bar{t}$ and W+jets	25.0	19.8
$t\bar{t}$, W+jets, w/ LM1 contamination	30.6	19.8

In applying the χ^2 vs. $\not{E}_T/\Sigma E_T$ method, the signal region is defined to be at high χ^2 and high $\not{E}_T/\Sigma E_T$, where Fig. 8.45 shows that there is more SUSY LM1 than $t\bar{t}$. The different ABCD regions are defined to be:

$$\begin{aligned} A: &\quad \chi^2 < 20, 0.1 < MET/\Sigma E_T < 0.2 \\ B: &\quad \chi^2 > 20, 0.1 < MET/\Sigma E_T < 0.2 \\ C: &\quad \chi^2 < 20, MET/\Sigma E_T > 0.2 \\ D: &\quad \chi^2 > 20, MET/\Sigma E_T > 0.2 \end{aligned}$$

where D is the signal region. We then predict the number of background events in the signal region by using the equation N(D) = N(B)/N(A) * N(C), assuming that the correlation between the variables in negligible and N(D)/N(C) = N(A)/N(C). Table 8.13 shows the number of background and LM1 events in each of the ABCD regions for 100 pb^{-1}. In the signal region, LM1 dominates over all the backgrounds. In each of the control regions $t\bar{t}$ is the dominant background as expected, though W+jets and LM1 also have some number of events in these regions. The amount of Z+jets, and VV+Jets, and single top is negligible and will not be considered further. The amount of QCD in each region is also negligible; however, since we do not trust the Monte Carlo for QCD predictions, Section 25 of this document describes data-driven techniques for determining the amount of QCD in each region.

In the case where $t\bar{t}$ is the only background considered and there is no signal contamination the ABCD prediction works well. Table 8.14 shows the predicted vs. actual number of events for different scenarios. The small overestimate (15-20%) in the case of pure $t\bar{t}$ and no signal contamination comes from the small negative correlation between χ^2 and $\not{E}_T/\Sigma E_T$ in $t\bar{t}$. If you add SUSY LM1 contamination to this estimate, the prediction becomes a bit worse, giving an overestimate of 35%. These signal contamination effects come from the handful of LM1 events that appear in the high χ^2, low $\not{E}_T/\Sigma E_T$ region, as seen in Table 8.13. If $t\bar{t}$ and W+jets are both considered for the ABCD method, and no signal contamination is applied, the prediction becomes slightly worse than when only considering $t\bar{t}$. Table 8.14 shows that in this case the prediction gives an overestimate of approximately 25%. When adding the LM1 contamination to the $t\bar{t}$ and W+jets prediction, the overestimate in the prediction is approximately 50%. The error analysis of the method for the $t\bar{t}$ background is discussed in the next section.

8.4 Statistical and systematic uncertainties

In the previous section we took a small sample of $t\bar{t}$ events (1 M) and made one prediction for the number of $t\bar{t}$ in the signal region. However, with only one small sample it is hard to know how statistically significant the prediction is. For a more thorough investigation of the statistical and systematic errors of the χ^2 vs. $\not{E}_T/\Sigma E_T$ method we ran many pseudoexperiments to get a distribution of actual and predicted background numbers, with residuals and pulls. To run these pseudoexperiments we produced a 14 M event FastSim $t\bar{t}$ sample in CMSSW 2.2.6. Validation plots for this large FastSim $t\bar{t}$ sample are shown in Figure 11.72. With this sample we were able

to run 335 pseudoexperiments of 100 pb^{-1} (each pseudoexperiment was 41400 events). The same ABCD regions from Table 8.13 were used for the pseudo-experiment predictions.

Figure 8.46 shows the distribution of the actual number of $t\bar{t}$ events in the signal region for the 335 pseudoexperients. The distribution has a mean of 20.4 events and a RMS of 4.2. The mean can be compared with the FullSim result shown in the previous section of 15.5 $t\bar{t}$ events in the signal region. The FullSim result is less than the mean of the FastSim distribution, but the FullSim result does fit nicely into the low tail of the FastSim distribution. Figure 8.46 shows that the mean number of predicted $t\bar{t}$ events in the signal region is 23.8, about 15% higher than the mean of the actual prediction.

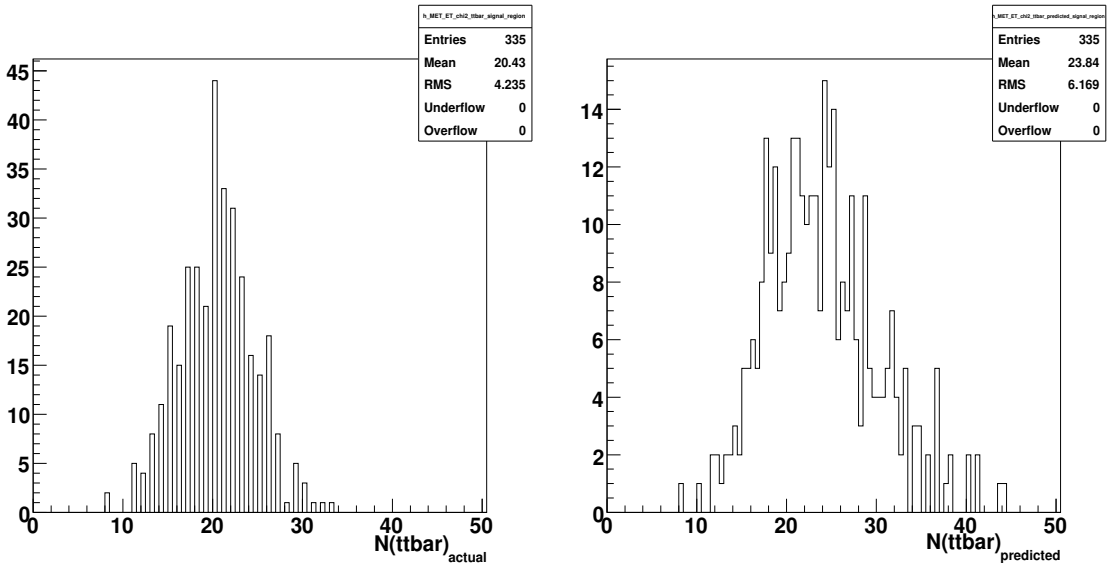


Figure 8.46: Number of actual, predicted $t\bar{t}$ in signal region for 335 pseudoexperiments of $100pb^{-1}$.

The residual plot in Figure 8.47 shows that we tend to overestimate the number of $t\bar{t}$ in the signal region by 3.7 events (16%) for each pseudoexperiment. This systematic bias is consistent with what was seen in the FullSim $t\bar{t}$ sample in the previous section. The RMS of the residual distribution is 7.7, which gives a statistical uncertainty on the prediction of 32%.

The pull distribution in Figure 8.47 has a RMS of 0.90, which means that we tend to slightly overestimate the statistical error on N(actual)- N(predicted) for each pseudo-experiment. Both the pull and residual plots are somewhat gaussian and it is nice to see that there is not a large tail in either distribution at negative values, which would signify a drastic underestimate of the predicted number of background events.

The method seems to work well when only statistical errors are considered; however, we would like to see what happens when systematic uncertainties are applied. The systematic uncertainties investigated for this method are varying the jet spectrum, the jet resolution, and the jet energy scale. For the following discussion jets have the following requirements: $p_T > 30$ GeV, $\eta \geq 2.5$, and hadronic energy fraction greater than or equal to 0.1.

The reason the jet spectrum is important for the ABCD method is that the χ^2 tends to work better in events with only 4 jets as opposed to events with 5 or more jets. In events with 5 or more jets we only use the four highest p_T jets when computing the χ^2. In this case, if one of the correct jets is not in the four highest p_T jets then it is not possible to get all the jets correctly in the χ^2. In $t\bar{t}$ events passing the jet cuts, approximately 45% of events have exactly 4 jets, while 55% of events have 5 or more jets. We change the jet spectrum so only 39% of events have exactly 4 jets, and 61% have 5 or more jets. The exact numbers for this change in jet spectrum are shown in Table 8.15.

To look at changes in the jet resolution, each jet's p_T in the event is smeared according to a gaussian of width 10% or 30% (i.e., in the case of smearing by 10% if there was a jet of p_T = 30 GeV, we smeared the jet according to a gaussian of width 3 GeV). The $\not\!\!E_T$ in the event is scaled to take into account the changed jet p_T's. To look at changes in the jet energy scale each jet's p_T in the events is multiplied by a factor of 0.9 or 1.1, to simulate decreasing or increasing, respectively, the jet energy scale. Just like in the case of the jet resolution, the $\not\!\!E_T$ is scaled to take into account the different jet p_T's.

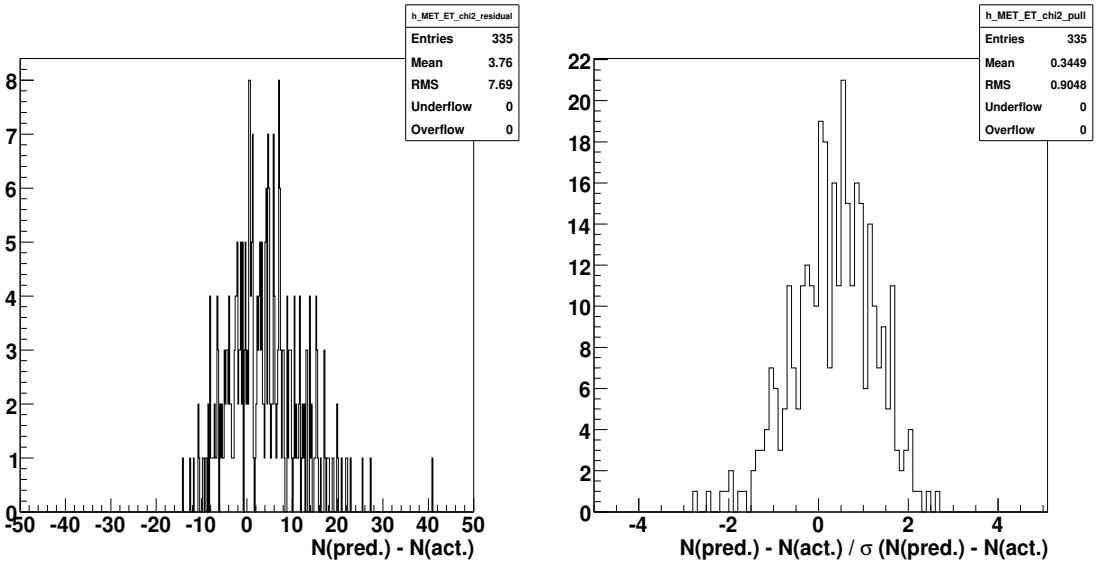


Figure 8.47: Residual, pull for $t\bar{t}$ in signal region for 335 pseudoexperiments of $100pb^{-1}$.

Table 8.15: Change in number of jet spectrum.

Num Jets (Passing Cuts)	Percent of All Jets	Percent of all jets, reducing 4 jet piece
4	45%	39%
5	33%	36%
6	15%	17%
7 or more	7%	8%

For each of the different systematic changes, 335 pseudoexperiments are run and the number of actual and predicted $t\bar{t}$ in the signal region is computed. Figure 8.48 shows the mean number of actual and predicted $t\bar{t}$ events in the signal region for variations in the systematics. The first bin in Figure 8.48 shows the mean number of actual and predicted $t\bar{t}$ with no systematics applied. The means from this first bin come from the distributions in Figure 8.46. As Figure 8.48 shows, the number of actual $t\bar{t}$ in the signal region varies drastically as different systematics are applied. However, even as the different systematics are changed and the actual number of $t\bar{t}$ in the signal region fluctuates, the prediction stays good.

We then assign a percent error for each systematic uncertainty. The error for each systematic uncertainty is computed in the following way: (1) the percent difference between the actual and predicted number of events with no systematics applied is calculated, (2) the percent difference between the actual and predicted number of events with a particular systematic applied is calculated, (3) the absolute difference between (1) and (2) is computed and used as the percent error for that particular systematic. The different systematic uncertainties are shown in Table 8.16 along with the statistical error and systematic bias. The statistical error and systematic bias dominate over the different jet uncertainties, which vary between 1-5%. Overall, the method tends to perform well when the systematics are changed. In the future we should study the systematic error from pdf uncertainties as well.

8.5 Early data commissioning plan

The LHC will initially run at 7 TeV for some amount of time before ramping up to 10 TeV. At 7 TeV it will be very hard to see SUSY in the four jet case because of the limited amount of statistics. At 7 TeV the $t\bar{t}$ cross section is 90 pb. With an integrated luminosity between 20-100 pb^{-1} (depending on how long the run is) there will be 1800-9000 $t\bar{t}$ events. With this number of events it will be possible to make some loose jet cuts on four jets (for instance four jets of $p_T > 30$ GeV), reducing the number of events to the order of 100, and then start looking at the χ^2 distribution. This includes looking at the individual distributions of each piece of the χ^2 and how they compare to what we have seen in the Monte Carlo. In addition, we will be following closely the work in the top group, where the top quark pair cross section measurement in the muon plus jets channel uses the χ^2, Ref. [1], albeit in a slightly different way.

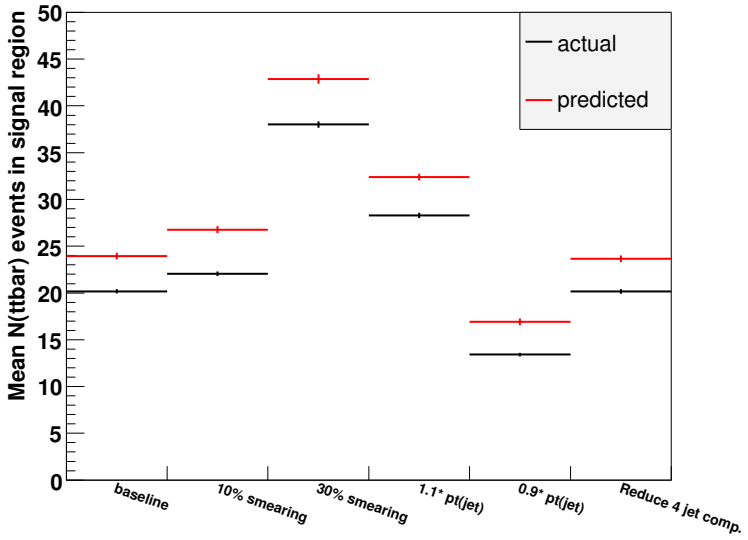


Figure 8.48: Robustness of χ^2 vs. $\not\!\!E_T/\Sigma E_T$ method for different systematics.

Table 8.16: Sources of uncertainty and their estimated error on the $t\bar{t}$ prediction.

Source of Uncertainty	Error(%)
Statistical	32
Systematic Bias	16
Jet Smearing (10%)	3
Jet Smearing (30%)	5
Jet Pt * 0.9	5
Jet Pt * 1.1	4
Reduce 4 jet component	1

References

[1] D. Green, Gail Hanson, Gen-Yuan Jeng, Francisco Yumiceva "Early Observation of Top Quark Pair Production in the Muon plus Jets Channel", **CMS AN-2009/084**.

[2] D. Acosta *el al.* (CDF Collaboration), Phys. Rev. Lett. **93**, 061802 (2004).

9 $t\bar{t}$ background from the TopBox method

Wing To

The topbox method utilizes a sample of partially reconstructed semileptonic $t\bar{t}$ events to predict $t\bar{t}$'s contribution to the high MET and high HT tails. The method utilizes an event selection of a single muon of 10 GeV and 4 or more jets with p_T greater than 75, 50, 50, and 30 GeV. All histograms and results are shown with 100 pb^{-1} of simulated data, unless otherwise noted. In subsection 9.1, the method is described and demonstrated in a pure $t\bar{t}$ sample. In subsection 9.2, a SUSY signal is added to show how well the method works in the presence of signal contamination. In subsection 9.3, the effect of a W+jets contribution are studied. Then subsection 9.4 discusses the effect of a QCD contribution, and some methods are discussed to predict and subtract that contribution. Finally, in subsection 9.5 all these contributions are put together, to study the final performance.

9.1 Description of Method

The topbox method reconstructs the invariant masses of 2-jet (M2) and 3-jet (M3) combinations to find $t\bar{t}$-like events. One expects in the semileptonic $t\bar{t}$ case to find one and only one 2-jet combination with an invariant mass near the W-boson mass and one 3-jet combination near the top-quark mass. However, the M2 and M3 distributions

will have a spread due to resolution effects. Some jets will also be removed by the selection cuts, misreconstructed or not reconstructed at all. The selection bias will also skew M2 and M3, which will be discussed in Section 9.1.1. We simulate these effects by using Monte Carlo (MC) matched M2 and M3 distributions. Utilizing this MC-based information, we establish M2 and M3 cuts as: $75 < M2 < 115$ and $160 < M3 < 200$ GeV. Events passing these cuts define the "topbox" sample, i.e., a $t\bar{t}$ enriched sample.

Kinematic distributions in the topbox sample are used to predict the same distributions in the full sample. For example, Fig. 9.49 and 9.50 compare several kinematic distributions between the topbox and full samples, with equal normalization. The shapes are reasonably well matched, making the topbox sample a useful predictor of semileptonic $t\bar{t}$ kinematics. In data, we will need to determine this normalization with signal, and other types of background presents, which will be discussed in subsection 9.2, 9.3, and 9.4.

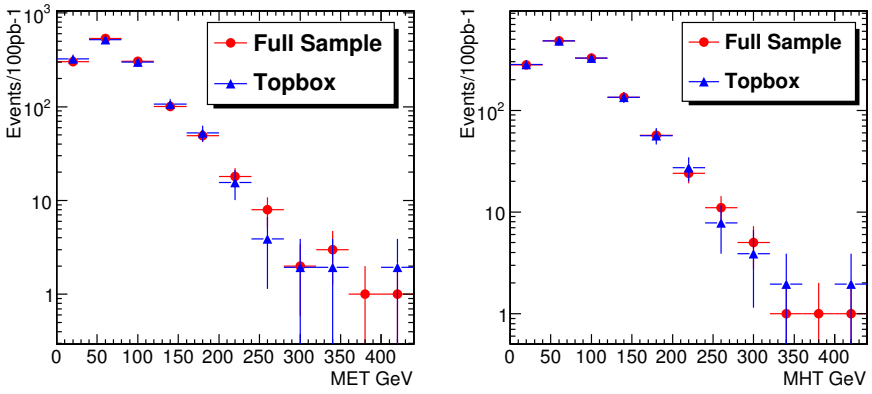


Figure 9.49: $\not\!\!E_T$ and MHT distributions of all selected $t\bar{t}$ and Topbox events. The topbox events are scaled up to the same normalization as selected events.

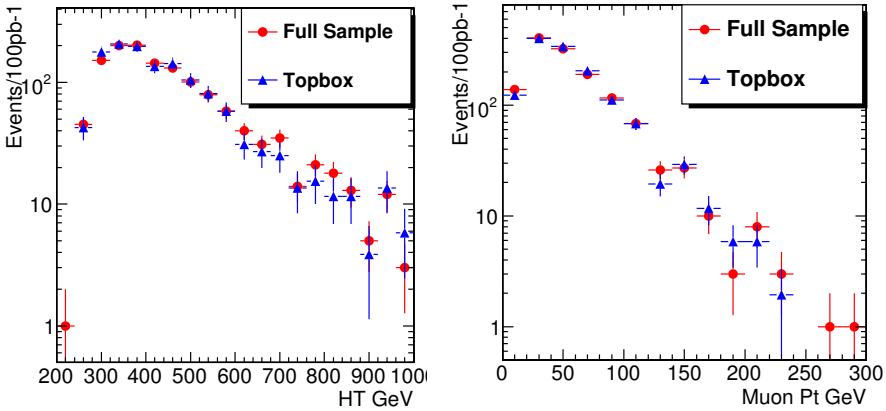


Figure 9.50: HT and μ_{p_T} distributions of all selected $t\bar{t}$ and Topbox events. The Topbox is scaled up to the same normalization as selected events.

9.1.1 W and Top mass skewing

The M2 and M3 distributions are shown in Fig. 9.51, using jet assignments from MC truth information. Notice that the mean of these distributions are actually at 88 GeV and 182 GeV, respectively. This bias could result from the selection, specifically the minimum jet p_T requirement. To confirm this, we reduce the jet p_T cut to only 5 GeV, for all jets, and see a shift back toward the nominal W-boson and top-quark masses, as shown in Fig. 9.52. However, since we must use the tigher jet p_T cuts, the M2 and M3 cuts are, henceforth, placed around an *effective* W-mass of 90 GeV and an *effective* top-mass of 180 GeV.

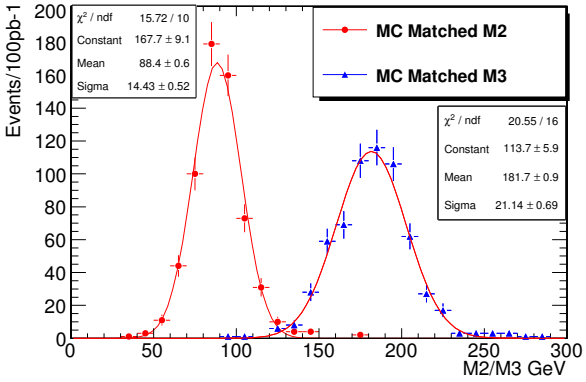


Figure 9.51: M2 and M3 using MC Matching.

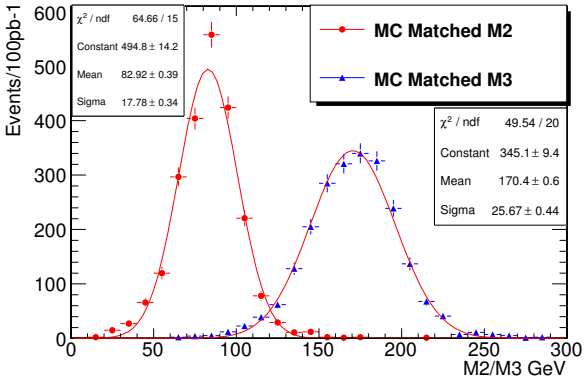


Figure 9.52: MC Matched M2 and M3 with Jets $P_T > 5$ GeV cuts.

9.1.2 Leptonic M3 MET bias

The mass of the leptonically decaying top could also be used to select the topbox sample. If we assume the $\not\!\!E_T$ is entirely due to the neutrino from semileptonic $t\bar{t}$ decay, we could use the W-boson's mass and the muon's four-momentum to calculate 2 possible solutions for the neutrino's p_z. Then we could loop over all remaining jets in the event to find the jet, muon, and neutrino four momenta that reconstruct to a leptonic invariant mass (M3Lep) nearest to the top-quark mass. This would be an effective way to reduce the other SM backgrounds since it will suppress non-$t\bar{t}$ events. However, using a high statistics sample of $t\bar{t}$, corresponding to 1 fb^{-1}, clearly shows that the leptonic M3 biases the $\not\!\!E_T$ distribution, by about a factor of two. So, we leave out the leptonic M3 until further studies of its effects are understood.

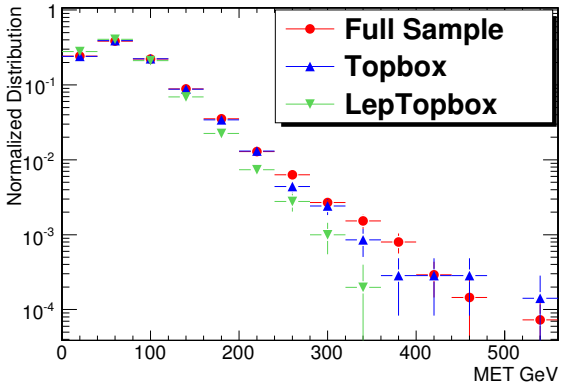


Figure 9.53: $\not\!\!E_T$ Distribution for Hadronic and Leptonic M3 selected events.

9.2 Signal Contamination

In actual data, the selection cuts will allow both $t\bar{t}$, SUSY, and other SM backgrounds in the sample. In this section we show how adding SUSY LM1 into the $t\bar{t}$ sample affects the predictions. The normalization scaling will not work in this case since we expect LM1 to have a small contribution in low $\not\!\!E_T$ bins and a larger contribution to the high $\not\!\!E_T$ bins. So, we use the integral of the $\not\!\!E_T$ region between 50 and 100 GeV as a scale factor from the topbox sample to the full sample. Figue 9.54 shows the $\not\!\!E_T$ distribution in the full sample, the $t\bar{t}$ only sample and the topbox plus SUSY prediction, with the scale factor.

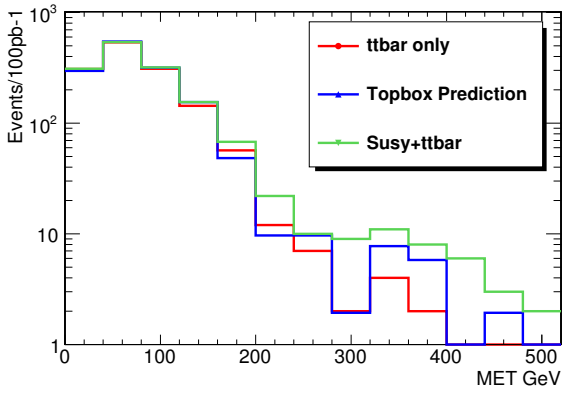


Figure 9.54: $\not\!\!E_T$ distribution $t\bar{t}$ sample, Topbox Predicted, and $t\bar{t}$+SUSY sample

The method's prediction agrees well with the $t\bar{t}$ only sample. We also added the LM1 into the $t\bar{t}$+SUSY distribution. At only an integrated luminosity of 100 pb^{-1}, we only expect very few events to pass our selection cuts. In this case only 82 events passes the SUSY standard selections. The topbox method predicts 42 events with $\not\!\!E_T > 200GeV$ of which 33 events are accountable to $t\bar{t}$. The combined $t\bar{t}$ and LM1 sample has 79 events with $\not\!\!E_T > 200GeV$, which is a significant excess above the prediction. The disagreement between the prediction and actual ttbar contribution is caused by signal contamination. This systematic error is expected since some SUSY events will look like $t\bar{t}$ due to its multi-lepton and multi-jet characteristics. Many SUSY models have top-quarks within the decay chain, thus the method will not be able to reject those types of events.

9.3 W+jets Contribution

The previous plots were shown with only $t\bar{t}$ and SUSY events, however in data the 1 muon plus 4 jets sample will also have contributions from W+jets events, where the W decays into a prompt muon. In this section we will examine how much W+jets effects the topbox method. Using a sample of 100 pb^{-1} of W+Jets, our μ + multijets selection cut reduces the W+jets contribution to 95 events. The hadronic topbox selection further reduces this to 17 events. In Fig. 9.55, the $\not\!\!E_T$ distribution for the full sample, with and without, W+jets is shown. We can see that the W contribution is small and does not affect the shape of the distribution. This is expected since both semileptonic $t\bar{t}$ and W+Jets events produce a prompt neutrino. Thus the $\not\!\!E_T$ distribution should be very similiar. The same can be seen with the topbox selection where the effect is even smaller, a single event difference. Since these deviations are insignificant, on the order of a few percent, we do not attempt to correct for this effect.

9.4 QCD Contribution

We expect the tight selection to reject most QCD contributions to the high MET tail, but fake and/or secondary leptons can cause non-trivial changes to the topbox based prediction. Using 1 pb^{-1} (1.8M events) of QCD p_T80+ events, we studied the effect of QCD. We found only 24 events passing selection. In order to obtain better statistics, to simulate a 100 pb^{-1} sample, the muon selection cuts of d_0 and combined relative isolation were loosened to 20 times their normal values for the QCD sample. To demonstrate that the $\not\!\!E_T$ distribution does not change due to the change in the selection cuts, the MET distributions of the normal, 5 times, 10 times, 20 times and 30 times the normal isolation cuts were made and the distribution did not change significantly among the different isolation cuts. The cuts of 20 times the normal isolation was choosen because it has approximately 2400 events, which is the projected number of QCD events with muons passing our selection cut with 100 pb^{-1} of integrated luminosity. Figure 9.56, shows the MET distribution for the full sample selection, with and without, QCD in red and the same for the Topbox selection in blue.

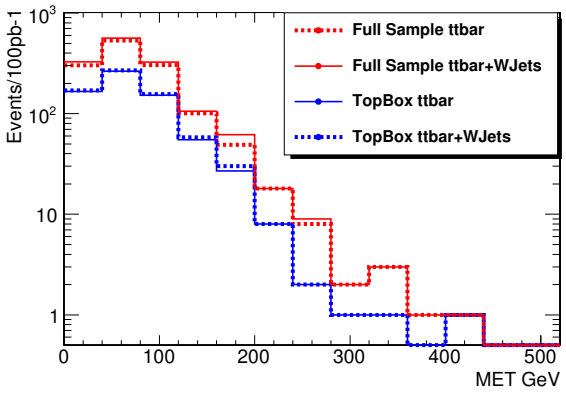


Figure 9.55: $\not\!\!E_T$ Distribution Selection and Topbox events with and without W+Jets contamination.

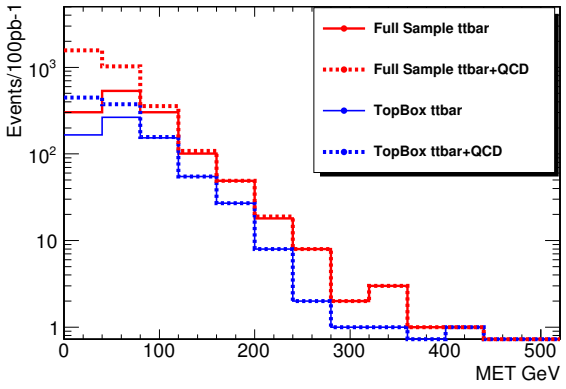


Figure 9.56: QCD $\not\!\!E_T$ distribution Selection and Topbox events with and without QCD contamination.

The distributions show that QCD has a large contribution for $\not\!\!E_T < 100GeV$. We don't want to place a $\not\!\!E_T < 100$ cut on all events because that is an important part of the $t\bar{t}$ background prediction. So, we make a prediction of the MET contribution from QCD and subtract it. To reproduce the QCD MET shape we attempt to use two methods. First, we take events with zero muons, second, we look for events with non-isolated muons by inverting the isolation cut. The $\not\!\!E_T$ shapes of these control samples are compared to that of the QCD in the selected sample in Fig. 9.57.

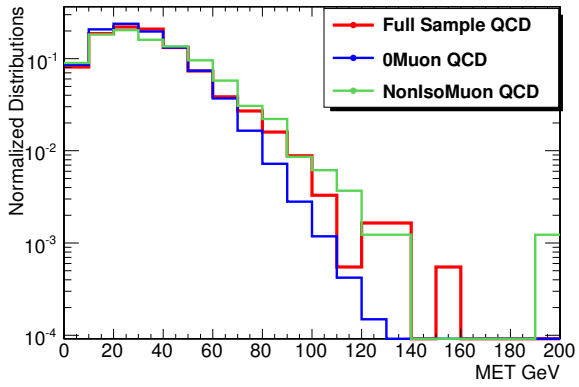


Figure 9.57: $\not\!\!E_T$ distribution of QCD-only, 0 Muon, and Non-Isolated Muon Selections

Both shapes compare reasonably well to the normal selection's QCD shape. However, the zero muon selection seems to under estimate the $\not\!\!E_T$ tail. This could be explained by the fact that most QCD events, which pass the single muon selection, have a real muon, and hence a neutrino, from secondary decays such those from b-mesons

and taus. The zero muon sample cannot account for this, although the non-isolated muon sample will. It is therefore more accurate.

Next, we have to find the normalization of the distribution. From Figure 9.47, we see that QCD seems to dominate over $t\bar{t}$ with $\not\!\!E_T$ bins lower than 30 GeV. We can take this as the normalization region and scale the QCD prediction acccordingly. We demonstrate this using the QCD+$t\bar{t}$ single muon selection and QCD+$t\bar{t}$ non-isolated muon selection, then scaling the latter using the integral of the two distributions with $\not\!\!E_T < 30GeV$, as shown in Fig 9.58. The QCD-only $\not\!\!E_T$ distribution is also shown for comparison with the predicted QCD $\not\!\!E_T$ distribution. Since there are some $t\bar{t}$ events in the low $\not\!\!E_T$ bin, the method overestimates the normalization factor by 23.4%. This is a very rough estimator of the amount of QCD in our full sample and assumes QCD dominates over $t\bar{t}$ at low $\not\!\!E_T$, which may or may not be the case in real data. More robust and detailed methods for estimating QCD are described elsewhere in this document.

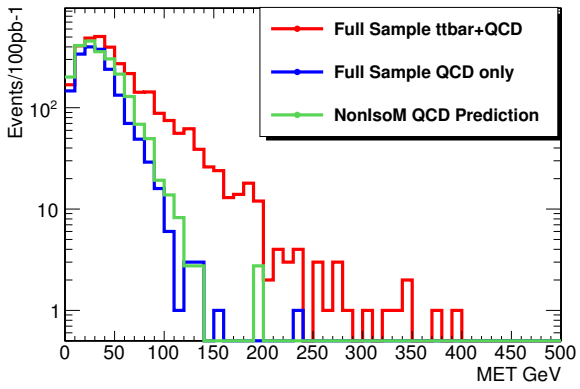


Figure 9.58: $\not\!\!E_T$ distribution of QCD+$t\bar{t}$,QCD only and Non Isolated Muon Prediction using Low MET Scaling.

9.5 Subsection 5: Predictions with all backgrounds

Now that we have an idea of how each of the large contributors affect the topbox method's prediction. All these samples are combined together to create a mixed sample of $t\bar{t}$, LM1, W+jets, and QCD. We work backward by first applying $\not\!\!E_T$ subtraction, found in subsection 9.4, out of the total $\not\!\!E_T$ distribution in the combined sample. The W+Jets contribution are ignored at this point since the effect is on a few percent level. Then we follow what we did in the signal contamination section but switch to a background region of $80 < \not\!\!E_T < 140$ GeV instead of $50 < \not\!\!E_T < 100$ GeV. This is necessary because our QCD subtraction method creates a large underestimate of $t\bar{t}$ below 80 GeV. In Fig. 9.59, the $\not\!\!E_T$ prediction is shown in blue and the true $t\bar{t}$ $\not\!\!E_T$ is shown in green. The red distribution shows the MET distribution in the full sample. Table {table:topboxpredictions tabulates the prediction and actual value of $\not\!\!E_T > 200$ with various contamination added into the sample.

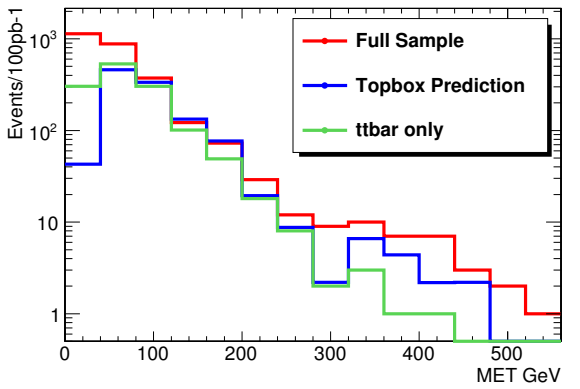


Figure 9.59: MET distribution of Predicted using non-isolated QCD subtraction (blue), full sample (red), and $t\bar{t}$ only sample (green).

Table 9.17: MET Prediction and Actual Full Sample with all contaminations

MET > 200	No Contam	LM1	W+Jets	QCD NIso	QCD 0Muon
Full Sample	$33.0 \pm 5.74_{stat}$	$79.0 \pm 8.89_{stat}$	$80.0 \pm 8.94_{stat}$	$80.0 \pm 8.94_{stat}$	$80.0 \pm 8.94_{stat}$
Prediction	$25.7 \pm 5.07_{stat}$	$41.9 \pm 6.47_{stat}$	$42.6 \pm 6.53_{stat}$	$49.1 \pm 7.01_{stat}$	$49.4 \pm 7,03_{stat}$

9.6 Systematic uncertainties

Systematic uncertainties for the topbox method should be classified into two main categories. First category of errors are those common throughout the RA4 analysis. These are effects of jet energy corrections which tends to increase energy of reconstructed jets in quarky events. We can reproduce the effects of uncertainty of JES correction for example to 10%. For simplicity we only study the effects of LM1 contamination which is tabulate in Table 9.18. JES correction uncertainty creates a large error in our prediction because the M2 and M3 variable relies on the jets to have been reconstructed with the correct energy. A 10% deviation on each jet will lead to 10% to 20% deviation in M2 and M3 variable. The topbox method relies on MC to predict the peaks of M2 and M3 distribution and makes a tight cut around the effect W-boson and top-quark masses. This deviation will cause the method to select non-toplike events. Thus the purity will suffer in cases of top events, while more signal contamination will come from LM1 events. Further investigation will be needed to validate the peaks and width of M2 and M3 to validate the method's predictions.

Table 9.18: MET Prediction with various JES corrections.

MET > 200	LM1	LM1 JES+10%	LM1 JES-10%
Full Sample	79.0	92.0	67.0
Prediction	41.9	75.3	47.5
$t\bar{t}$-only	33.0	36.0	28.0

The second category belongs to those specific to this methods. From Table 9.17, we can see that the method actually under predicts the amount of $t\bar{t}$ in a pure $t\bar{t}$ sample. High MET events tends to have more jets this will lead to more combinatoric in the Topbox selections, which gives a greater than of mismatching a jet. The second column shows that signal contamination greatly cause an over prediction of the about $t\bar{t}$ in the method. This is due to that fact that some LM1 event will contain top quark and its decay product. This is irreducible since the method specifically look for events with hadronic top decays. The next biggest cause of error is QCD's contribution. The QCD subtraction method tends to over estimate QCD in the low MET region. Thus, when the QCD part of the distribution is subtracted out, the method takes away $t\bar{t}$ events also. The method relies on the low MET region for scaling which will lead to a scale which is too large.

9.7 Early data commissioning plan

Early data will consist mostly of QCD jets derived from light quarks. The method will requires a good estimator of QCD's contribution to the single lepton selection. Thus a study of muon fake rate will be the most sensible during early data taking. The topbox method utilizes semileptonic $t\bar{t}$ decay to predict $t\bar{t}$'s contribution to single lepton SUSY searches. QCD and W+Jets contribution do not affect the prediction greatly but it must be validated in real data. The biggest problem seems to be signal contamination which creates an over prediction of $t\bar{t}$ if the SUSY signal contains top quarks. Jet energy correction mismeasurement can also create a large overestimate of the number of $t\bar{t}$ events in the sample. The method still requires careful validation using real data.

10 $t\bar{t}$ background using χ^2-sorting and sideband subtraction

Lorenzo Agostino

10.1 Introduction

Physics beyond the Standard Model can manifest itself in several ways. Within the SUSY framework the signal is often characterized by large jet multiplicities and large E_t^{miss}. The lepton + multijet + E_t^{miss} final state is considered one of the best candidates for early discovery because it provides a good trade-off between a relatively large cross section and a contained background. Because of its topology, understanding the $t\bar{t}$ background is of crucial importance before it is foreseeable to claim observation of new physics in the lepton + multijet + E_t^{miss} final state. An extensive use of Monte Carlo simulation to extract this background should be avoided in the initial

Table 10.19: Monte Carlo sample information.

Sample	σ (pb)	Int. Luminosity (pb^{-1})	N. Events	Weight
LM1	16.06	6525	104800	0.01532567
$t\bar{t}$	317	2986	946644	0.03348962
W+Jets	40000	244	9745661	0.40983606
Z+Jets	3700	341	1262816	0.29325513
QCD100-250	15000000	0.844	12662923	118.483412
QCD250-500	400000	12.7	5064539	7.87401574
QCD500-1000	14000	355	4690718	0.28169014
QCD1000-Inf	370	2883	1066863	0.03468609

period of data taking. For this reason a data driven approach would be desirable when estimating this background in data. In addition, a non negligible fraction of the background comes from W events where the W decays to an electron and neutrino. While the final goal of this analysis is to extract the $t\bar{t}$ background component alone, in this analysis the W background is implicitly assumed to be part of the $t\bar{t}$ sample. A method for disentangling $t\bar{t}$ from the W background is described in Section **??** and relies on a sample of ZJets events with the Z decaying in two electrons to model the W background.

The paper is organized as follow: in section 10.3 the event selection is described as well as the reconstruction of top and W masses for $t\bar{t}$ candidate events. The sideband subtraction technique and the extraction of the $t\bar{t}$ contribution will be described in Section 10.4. In Section 10.6 we will present the results and some remarks about the systematics studies which are being carried out.

10.2 Monte Carlo Samples

The main backgrounds in this analysis come from electroweak processes (Z^0, $W^{\pm}$ and $t\bar{t}$) and from electron misidentification of jets from QCD events. In particular, semileptonic $t\bar{t}$ events, where one of the two tops decays semileptonically (with a neutrino remaining undetected) and the other hadronically, are characterized by similar signature as expected from SUSY. Furthermore, at the LHC a large cross section is expected for $t\bar{t}$ [**?**]. These conditions make this background the main challenge of this analysis. The list of Monte Carlo samples used is reported in Table 10.19. The LM1 and the QCD samples have been generated using PYTHIA while the electroweak samples Z^0, $W^{\pm}$ and $t\bar{t}$ have been generated using MadGraph.

10.3 Event Selection and Mass reconstruction

This analysis is based on simulated data rescaled to 100 pb^{-1} of integrated luminosity collected at the center of mass energy of $\sqrt{s} = 10$ TeV. To extract the $t\bar{t}$ background contribution from data, a clean sample of $t\bar{t}$ events must be selected. Therefore the event selection must be chosen in such way to preserve an adequate statistics of top events as well as the highest possible efficiency for SUSY candidates. In addition, other sources of backgrounds should be reduced as much as possible.

Events are required to fulfill a single electron trigger requirement. Besides the trigger, an event is accepted in the final sample if it contains only one electron satisfying the following criteria: $E_t > 20$ GeV, a relative calorimetric isolation less than 0.1 and is within the electromagnetic calorimeter acceptance defined by $|\eta| < 1.47$ and $1.56 < |\eta| < 2.5$.

Since we are interested in high jet multiplicity events and in semileptonic top decays, at least four jets with $E_t > 40$ GeV and $|\eta| < 3.0$ must be found. In addition the jets must deposit no more that 10% of the energy in the electromagnetic calorimeter. Finally an $E_t^{miss} > 35$ GeV is required. The kinematic distributions for the leading electron in the backgrounds sample are shown in Fig. 10.60 whereas distributions for jets and $E_t^{miss} > 35$ are shown in Fig. 10.61. A summary of the analysis cuts is reported in Table 10.3

After the event selection has been performed, the sample of $t\bar{t}$ events is obtained by simultaneously reconstructing the hadronic W mass, the leptonic top mass and hadronic top mass. Since we have decided not to use B-tagging information to maximize the efficiency, combinatorial background is significant. To achieve the best resolution while minimizing the number of combinations, a χ^2-minimization algorithm [**?**] is used to sort among the top four

Variable	Cut
Num. electrons	=1
E_t	$>$20 GeV
$\|\eta\|$	[0, 1.47],[1.55,2.5]
Ecal Isolation/ E_t	< 0.1
Number of Jets Jets	$>= 4$
E_t	$>$40 for all jets.
$\|\eta\|$	[0,3.0]
EM Fraction	$<$0.1
E_t^{miss}	$>$35. GeV

Table 10.20: Table of the analysis selection cuts.

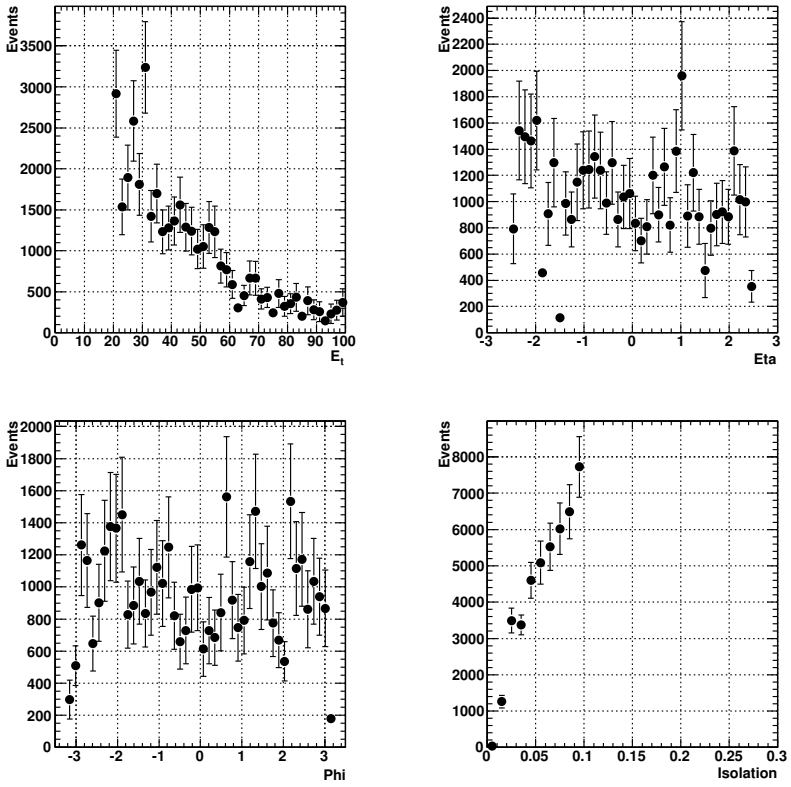


Figure 10.60: Kinematic distributions of the leading electron for all the backgrounds combined.

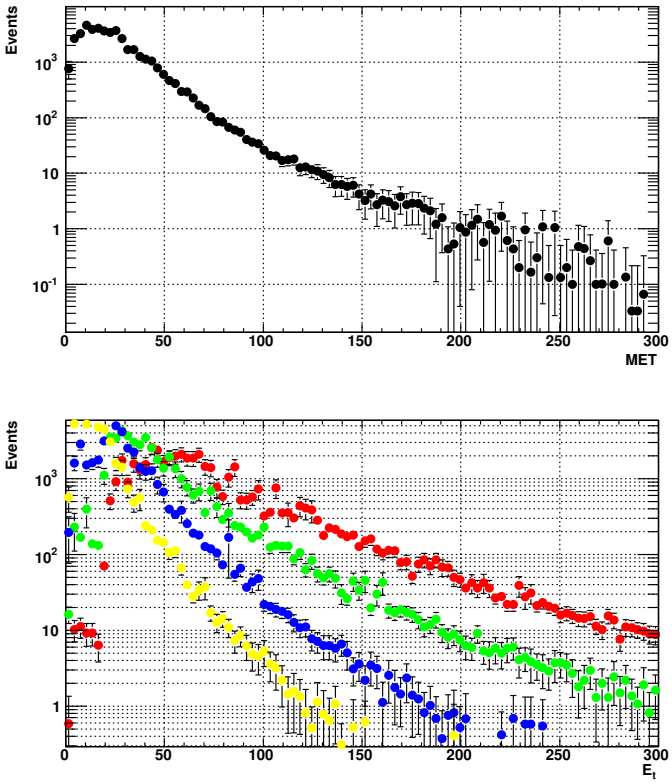


Figure 10.61: Distribution of the missing transverse energy (top) and of the four leading jets E_t (bottom) for all the backgrounds combined.

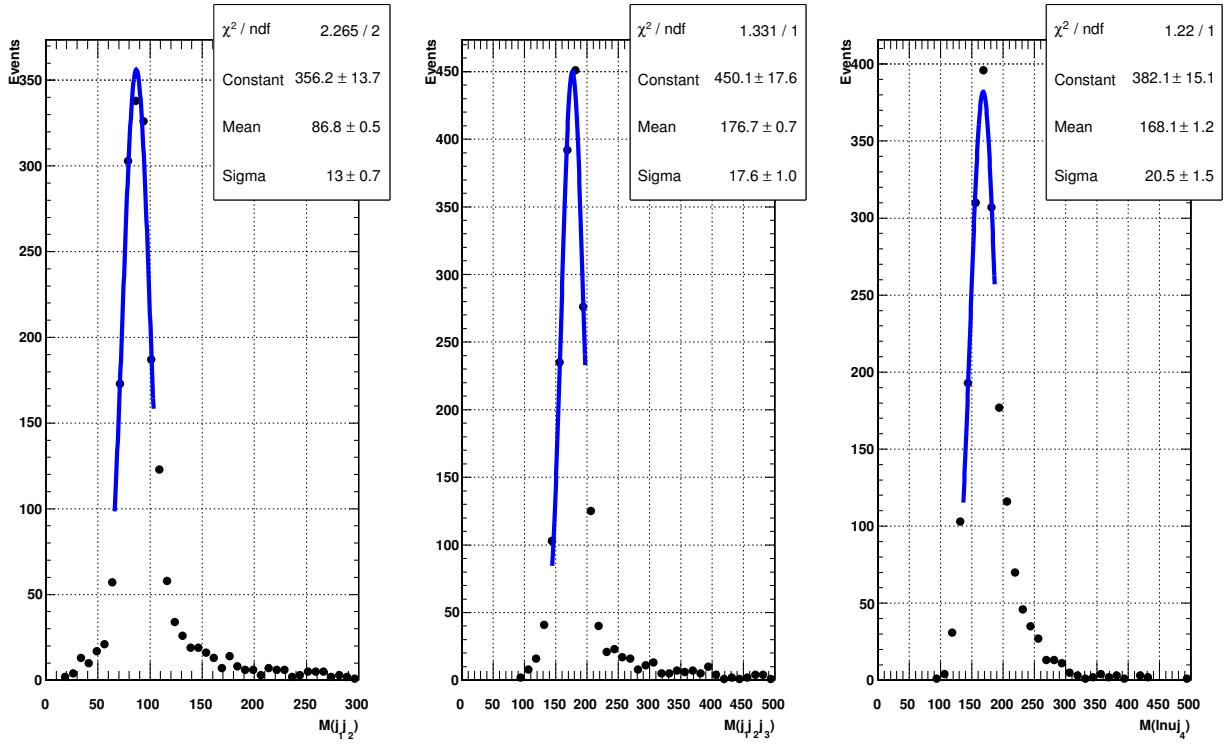


Figure 10.62: From the left to the right the hadronic W and top mass and the leptonic top mass are shown with the corresponding resolution from MC matching are shown. The resolutions are used in the χ^2-sorting of the four jets.

jets ranked by E_t. If j_1 and j_2 are the two jets from the hadronic W decay and j_3 and j_4 are the two b-jets from the top decays, the χ^2 which needs to be minimized can be defined as follows:

$$\chi^2 = \frac{(M_{top} - M(l\nu j_4))^2}{\sigma^2_{l\nu j_4}} + \frac{(M_{top} - M(j_1 j_2 j_3))^2}{\sigma^2_{j_1 j_2 j_3}} \frac{(M_W - M(j_1 j_2))^2}{\sigma^2_{j_1 j_2}}$$

where $\sigma_{l\nu j_4}$, $\sigma_{j_1 j_2 j_3}$ and $\sigma_{j_1 j_2}$ are obtained from Monte Carlo after matching reconstructed and generated particles. In Fig. 10.62 the three resolutions from Monte Carlo are shown. The masses are obtained by matching reconstructed and generated particles, the distributions are then fit to a single Gaussian which returns a resolution of 13 GeV, 17.6 GeV and 20.5 GeV for the hadronic W, the hadronic top and the leptonic top respectively.

The jet combination which produces the smallest χ^2 is used in the analysis. In the computation of the leptonic top decay, the z-component of the neutrino is calculated by solving the quadratic equation where the W mass is constrained to have the nominal value of 80.4 GeV. Solving this equations returns up to two real solutions of which the one which delivers a leptonic top mass closer to the nominal value of 175 GeV is chosen. If no solution is returned, the z-component of the neutrino is taken to be the electron z-component. In Fig. 10.63 the three masses contributing to the χ^2 are shown while the χ^2 distribution for signal and backgrounds is shown in Fig. 10.64 .

10.4 Side-band subtraction

Once the masses are calculated using the χ^2-sorting method, the hadronic W mass is used for a sideband subtraction which aims at reducing the SUSY signal contribution in the $t\bar{t}$ sample. The region between 60 GeV and 120 GeV is defined as the signal region while the region between 120 GeV and 180 GeV is defined as the sideband region. The underlying assumption is that the SUSY signal is spread over the entire mass range and that it is relatively flat, that is, it does not peak significantly in the W mass region. It is understood that, in a scenario where SUSY is present, the signal could exhibit different behavior than what we observe in our Monte Carlo simulation. The general statement is that the less peaking is the signal in the hadronic W mass, the more sensitive the method will be to the presence of new physics.

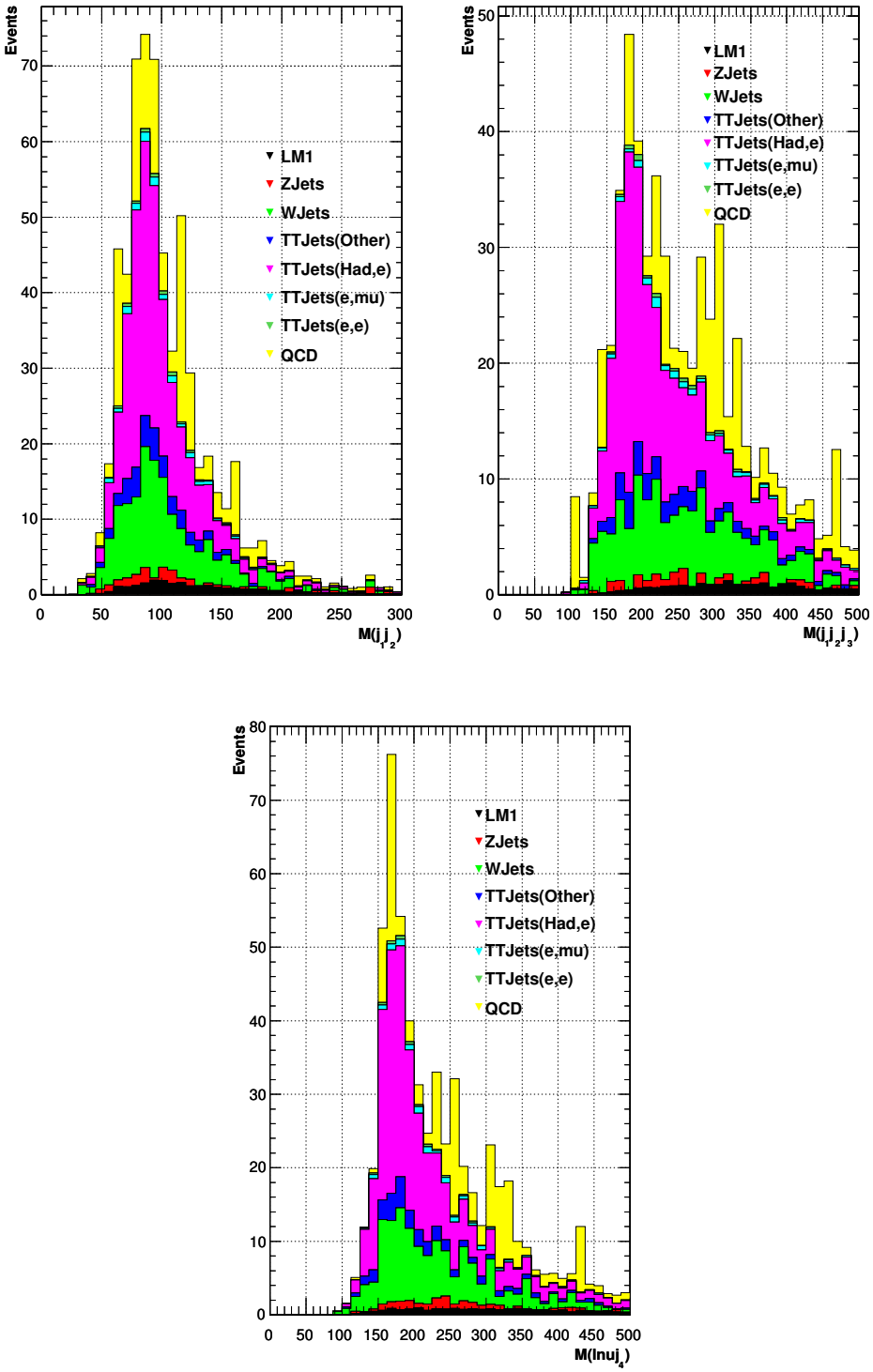


Figure 10.63: Stacked plots of the masses obtained with χ^2-sorting method where backgrounds and signal contributions are visible. The signal (black) component has a much smaller bias towards the W or top masses.

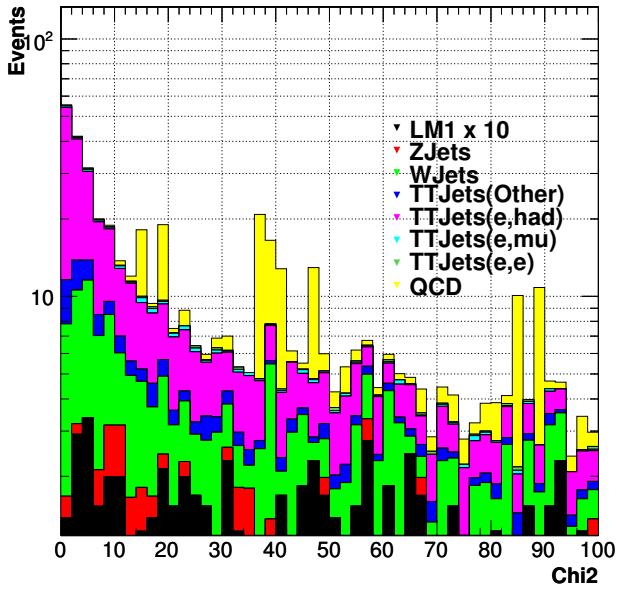


Figure 10.64: Stacked plot of the jet sorting χ^2 distribution for signal and backgrounds. The χ^2 minimization performs well for $t\bar{t}$ and poorly for LM1, meaning that the LM1 does not peak strongly in the top and W mass distributions, justifying the sideband subtraction technique to reduce signal contamination in the top sample.

Because a non negligible part of the $t\bar{t}$ background tail is also contained in the sideband region, the subtraction will affect the overall number of $t\bar{t}$ events in the sample. For this reason, only the shape obtained by this method is considered while the overall normalization must be recomputed by other means. This is achieved by normalizing the sideband subtracted E_t^{miss} shape in the low E_t^{miss} region where the $t\bar{t}$ (and WJets) backgrounds dominates with respect to the residual contribution from both signal and QCD background. The normalization region is chosen to be the E_t^{miss} between 50 and 100 GeV. In Fig. 10.65 a stacked histogram of the E_t^{miss} for all the backgrounds and signal components is shown. In Fig. 10.66 the E_t^{miss} distributions for signal and sideband regions as well as a comparison between sideband subtracted and total E_t^{miss} after normalization are shown.

10.5 Systematics

Several systematics studies are in consideration. The first systematic effect that is being considered is the bias introduced in the E_t^{miss} shape when the sideband subtraction. The hadronic W mass as opposite to leptonic W mass has been chosen on purpose to minimize the correlation with the E_t^{miss} due to the neutrino from the leptonic W. Nevertheless a small correlation which translates in a overestimate of signal in the signal region is observed. The study is performed removing the signal from the Monte Carlo sample and applying the full analysis to this sample. In Fig. 10.67 the plot of the sideband subtracted E_t^{miss} shape normalized to the total E_t^{miss} is shown as well as a plot of the difference between these two distributions in the signal region. Without SUSY signal we expect an excess in the signal region ($E_t^{miss} > 150$ GeV) of roughly 6 events (30% of the observed background) when all the backgrounds are combined. Of these about half of the events come from the semileptonic $t\bar{t}$. A detailed analysis of these bias and of how it affects different background components, as well as of possible ways to reduce it are underway.

10.6 Conclusions

Once the $t\bar{t}$ background shape and normalization has been determined, the presence of new physics can be inferred by looking for an excess of events in the high end tail of the E_t^{miss} distribution, where SUSY is supposed to be roughly of the same size as the standard model background. In Fig. 10.68, the total E_t^{miss} for events passing the selection is overlaid to the normalized background shape. An excess is observed in the region above 100 GeV.

In Fig. 10.68 a plot showing the excess obtained by subtracting the two distributions is compared to the expected number of LM1 signal events from Monte Carlo simulation. Integrating the distributions above E_t^{miss} of 150 GeV we obtain 18.6 ± 7.1 events of which 6 ± 2.4 are due to the change in shape due to the sideband subtraction. By a

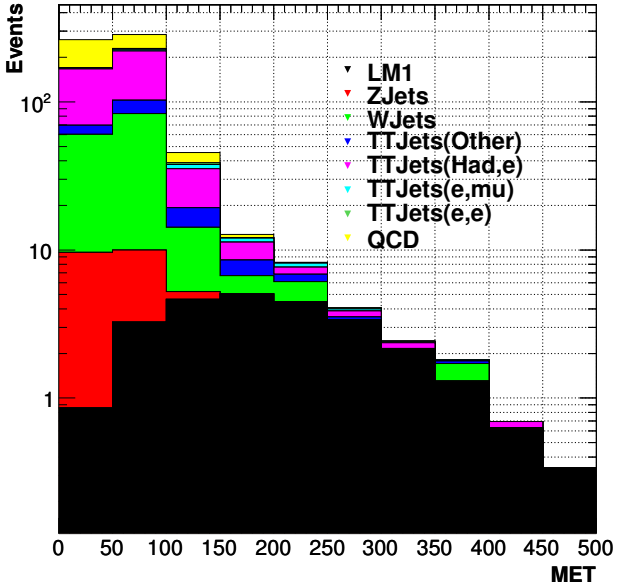


Figure 10.65: Stack histogram of the background and signal contributions to the E_t^{miss} distribution. In the $50 < E_t^{miss} < 100$ bin the $t\bar{t}$ and W components dominate.

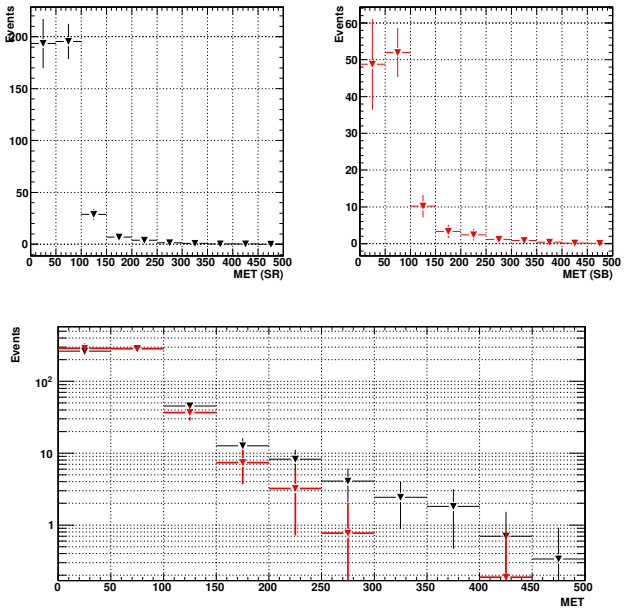


Figure 10.66: E_t^{miss} distribution for the signal region (top left), sideband region (top right) and comparison between the background distribution after sideband subtraction and the total E_t^{miss}. An excess of events is visible in the high-end tail of the E_t^{miss} distribution.

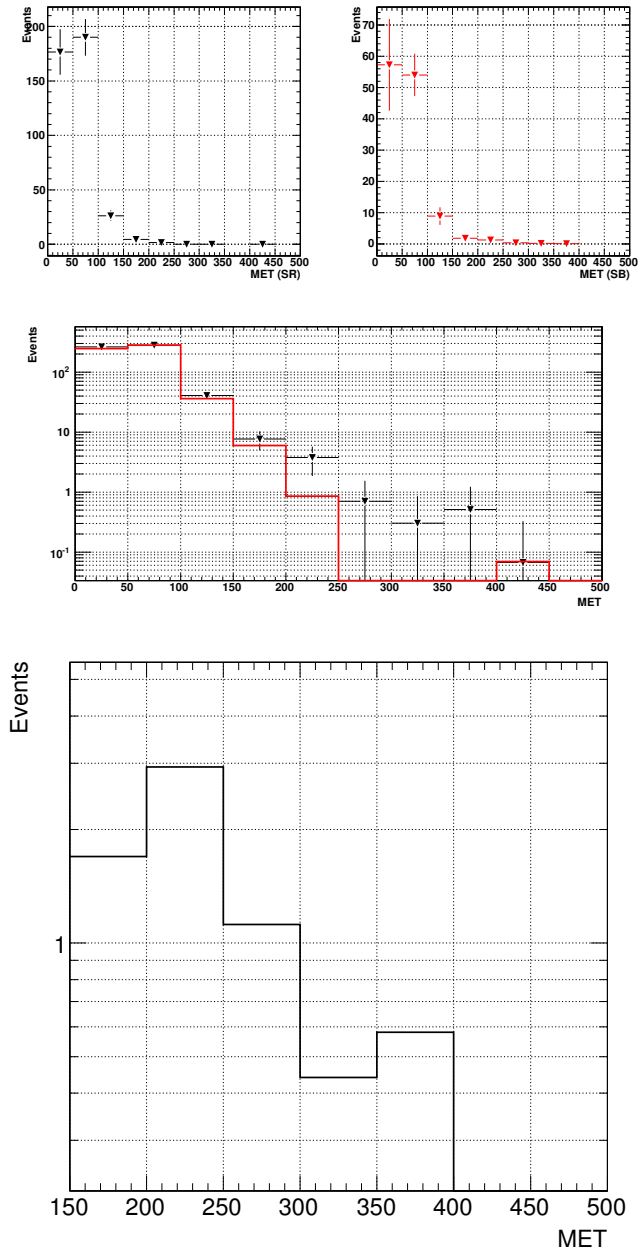

Events
MET (SR)
Events
MET (SB)
Events
MET
Events
1
150 200 250 300 350 400 450 500
MET

Figure 10.67:

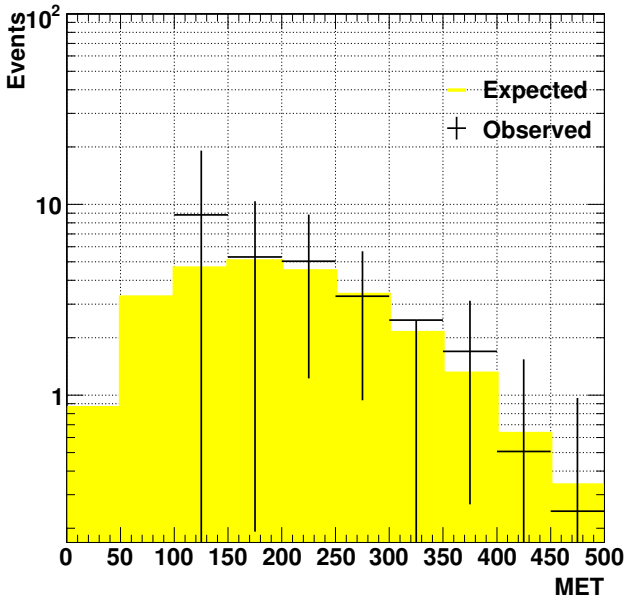


Figure 10.68:

simple correction of this bias we observe an excess of 12.6 ± 7.5 events, which should be compared to 11.8 ± 4.7 events expected.

11 Dilepton Background from $t\bar{t}$

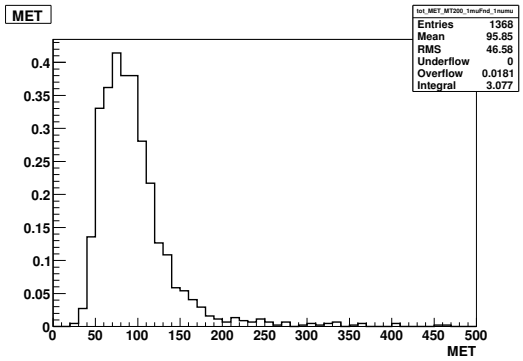
Methods which use M_T and MET to predict the contribution from $t\bar{t}$ background in the signal region, such as the ABCD method, require there be little correlation between the variables, but more so that the physics which dominates the background region, $MET < 200$ GeV and/or $M_T < 200$ GeV, is the same as the physics with dominates in the signal region, $\{MET, M_T\} > 200$ GeV. This turns out not to be the case. In the high M_T bins, figure 11.69 shows that the MET tails are not dominated by single muon events but instead come primarily from dilepton events; nearly 77% of events in these bins are dilepton events. Of these, nearly 60% come from the decay: $t\bar{t} \rightarrow \mu + \nu_\mu + \tau + \nu_\tau \rightarrow \mu + \nu_\mu + 2\nu_\tau + jets$. There is not only a correlation between MET and M_T, but the physics with contributes in the signal region is completely different from that in the background dominated regions. We investigated a method to predict this $t\bar{t}$ MET tail in data, which uses reconstructed dimuon events to simulate the dilepton contribution to MET in single muon events by "throwing away" one of the muons and recalculating MET and M_T.

11.1 Description of Method

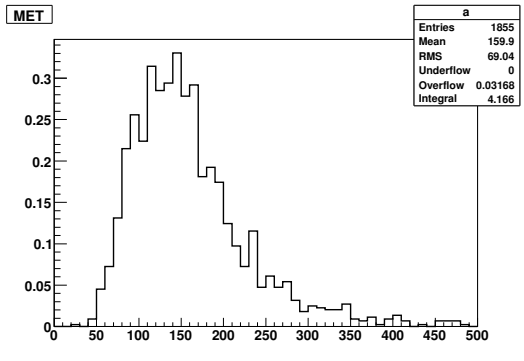
Since there are two different types of MET to be discussed, true MET in an event and recalculated MET which is the true MET plus the energy from a "thrown-away" muon, it is prudent to distinguish between the two. MET^X_{True} refers to the true MET in an event, where the X refers to the source of the MET, i.e. the neutrino content of the event from MC truth information. MET^Y_{Recalc} is the MET which is recalculated by adding in the energy from a discarded muon in an attempt to model the dilepton contribution in single muon events. The Y refers to the method by which the MET was recaluclated, i.e. $2\mu \rightarrow 1\mu + 1 \not{\mu}$ or $2\mu \rightarrow 1\mu 1\tau$. These methods are discussed below.

Given a sample of reconstructed dimuon events, we can simulate the dilepton contribution to MET in the single muon channel. There are two types of dilepton events which can show up in single muon events. The first are lost muon/electron events, where the second muon or the electron is not detected and thus a dilepton event is demoted to a single muon event with the lost lepton adding to the missing energy. The second category is the mu + tau events with an hadronic decay of the tau. There is a very small contribution from the tau to leptonic events but since these are suppressed by both branching fraction and event vetoes they are negligible.

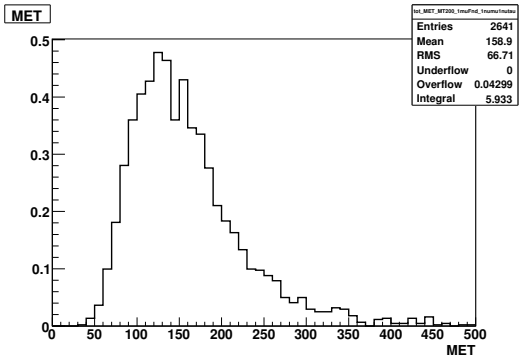
In the first case, we run over a sample of reconstructed dimuon events and for each event, we take one of the muons and add it's energy to the MET_{True} and recalculate M_T. If this new "event" passes our event selection and M_T



(a) MET from events with $1\nu_\mu$: $MET_{True}^{1\nu_\mu}$



(b) MET from events with $2\nu_\mu$ and $1\nu_\mu 1\nu_e$: $MET_{True}^{2\nu_\mu+1\nu_{mu}1\nu_e}$



(c) MET from events with $1\nu_\mu 1\nu_\tau$: $MET_{True}^{1\nu_\mu 1\nu_\tau}$

Figure 11.69: Contributions to MET in single muon events from: true single muon events (top-left), dilepton = $\mu\mu$ or μe (top-right), and $\mu + \tau \rightarrow$ hadronic (bottom).

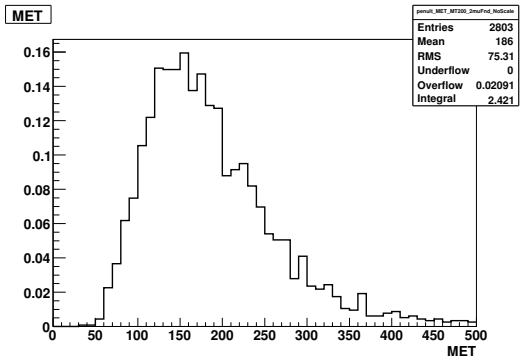
cuts then it is kept and its MET_{Recalc} recorded. We do this for each muon in the event, therefore, each event gives us two new calculations of MET_{Recalc} and M_T. The new $MET_{Recalc}^{2\mu\rightarrow 1\mu+1\mu'}$ distribution can be seen in figure 70(a).

In the second case, we again run over each of the dimuon events but instead of using the entire energy from the "thrown-out" muon, we add only half of its energy to the MET_{True}, simulating the tau's decay to a neutrino and jet and assuming that the neutrino gets, on average, half of the tau's energy. Again this is done for each muon in the event. We have not taken into account the new "jets" in our jet counting, but we have found that on average these "jets" have p_T less than 50 GeV and would fall below our jet cuts anyway. The $MET_{Recalc}^{2\mu\rightarrow 1\mu 1\tau}$ distribution from this method is given in figure 70(b).

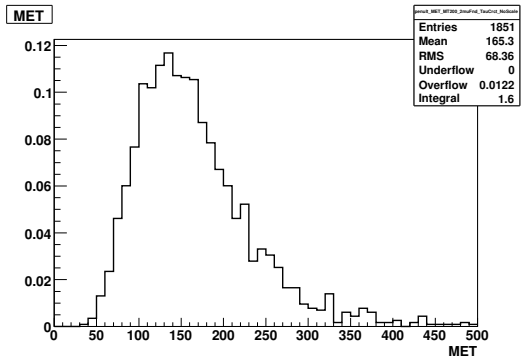
There are now two separate MET_{Recalc} distributions, one for lost muon/electrons 70(a) and one for mu-tau events 70(b), but we don't have the normalizations. In order to get the normalizations we use Monte Carlo truth information and scale the two distributions to their individual, MC determined contributions. The two normalized distributions are then added together to give a final estimate of the MET distribution as seen in figure 70(c). In other words:

$$\begin{aligned} MET_{Recalc}^{Total} &= \alpha \cdot MET_{Recalc}^{2\mu\rightarrow 1\mu+1\mu'} + \beta \cdot MET_{Recalc}^{2\mu\rightarrow 1\mu+1\tau} \\ \text{Figure } 70(c) &= \alpha \cdot \text{Figure } 70(a) + \beta \cdot \text{Figure } 70(b), \text{ where} \end{aligned} \tag{28}$$

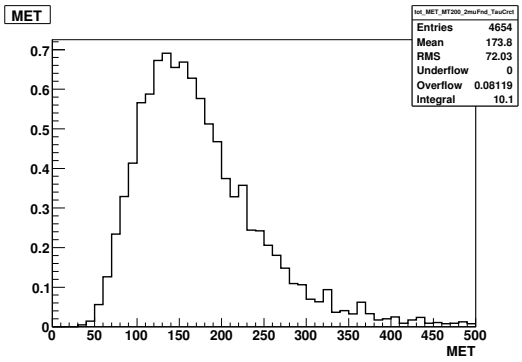
$$\alpha = \frac{\int_{MC} MET_{True}^{2\nu_\mu+1\nu_\mu 1\nu_e}}{\int_{MC} MET_{Recalc}^{2\mu\rightarrow 1\mu+1\mu'}} = \frac{\int_{MC} \text{Figure } 69(b)}{\int_{MC} \text{Figure } 70(a)} \tag{29}$$



(a) $2\mu \rightarrow 1\mu$ + 1"$\not{\mu}$"



(b) $2\mu \rightarrow 1\mu$ + 1"τ"$\rightarrow$ jets



(c) $MET_{Recalc} = \alpha\times$Figure 70(a) + $\beta\times$Figure 70(b)

Figure 11.70: Figure 70(c) is the linear combination of figures 70(a) and 70(b) for $M_T > 200$ GeV. The coefficients, α and β, are calculated from Monte Carlo.

$$\beta = \frac{\int_{MC} MET_{True}^{1\nu_\mu 1\nu_\tau}}{\int_{MC} MET_{Recalc}^{2\mu\rightarrow 1\mu+1\tau}} = \frac{\int_{MC} \text{Figure } 69(c)}{\int_{MC} \text{Figure } 70(b)} \tag{30}$$

Note that the coefficients, α and β, are completely taken from Monte Carlo, whereas the shapes will be taken from reconstructed dimuon data. For the coefficients to properly normalize the MET_{Recalc} distributions, our dimuon sample from which they were calculated must be as purely $t\bar{t}$ as possible, i.e., with minimal SUSY signal contamination. To accomplish this we only take dimuon events with a SumPt, defined as SumPt = $\sum_{\mu s} p_T^\mu + \sum_{\text{jets}} p_T^{jets} + MET$, between 200 GeV and 900 GeV, as shown in figure 11.71. Although there is still some signal contamination, signal events which pass the high MET and M_T cuts tend to fall in the high tail of the SumPt plot and so this cut greatly reduced signal contamination. However, the same can be said for the $t\bar{t}$ distribution and so it also biases our prediction on the very ends of the MET_{Recalc} tail and leads to an underestimate. With this cut on SumPt we retain 89% of the $t\bar{t}$ events while excluding 60% of the LM1 events, specifically events which would most bias our prediction.

11.2 Performance of method without signal

We use cuts that match the loose jet and loose muon selection specified in Sec. 3, except that we require $p_T^\mu > 15$ GeV. For our dimuon sample we use the same event selection cuts as above but require two muons with $p_T > 10$ GeV as well as 200 GeV $<$ SumPt $<$ 900 GeV as mentioned above.

To get a large sample of $t\bar{t}$ we generated and simulated our own sample of 14 million $t\bar{t}$ events using PYTHIA along with CMSSW FastSim simulation software in CMSSW 2.2.6. It is then scaled to 100 pb^{-1} luminosity using the given the NLO calculation of the cross section for $t\bar{t}$, 414 pb [2]. To validate this sample we compared it to the official, Fall 08 MadGraph $t\bar{t}$ sample (/TTJets-madgraph-Fall08-IDEAL-V9-v2/GEN-SIM-RECO). Figure

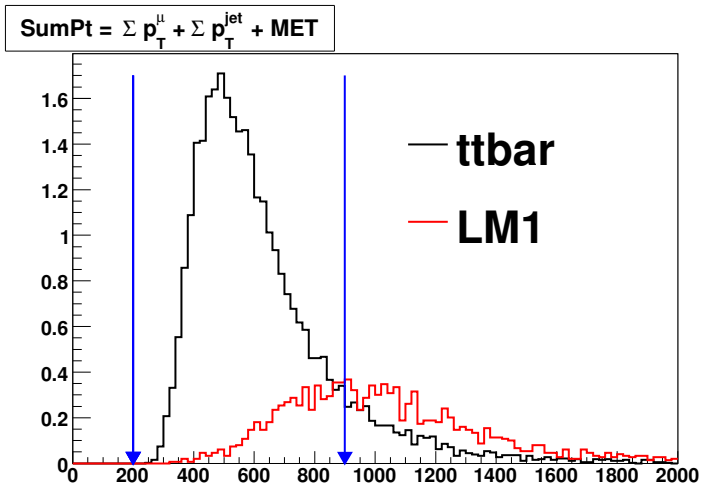


Figure 11.71: SumPt in dimuon events. A cut is placed at 200 GeV $<$ SumPt $<$ 900 GeV to maximize the amount of $t\bar{t}$ events while minimizing the signal contamination from SUSY.

11.72 compares several, high level kinematic distributions. While there is a small difference in yield and shape, the qualitative shapes are good enough for our needs.

Other backgrounds, listed below, had nearly no contributions to either the MET_{True} or the MET_{Recalc}. The contributions from single top, tW-channel, and dibosons had the next largest contribution after $t\bar{t}$ but these amounted to the order of 0.01 events/100 pb^{-1}. This is not surprising given the jet cuts and the high MET and M_T cuts. To what extent the Monte Carlo can be trusted is another matter; even if the true cross section for these events is ten times higher, we'd still have a small contribution compared to $t\bar{t}$.

- /WJets-madgraph-Fall08-IDEAL-V9-v1-/GEN-SIM-RECO
- /ZJets-madgraph-Fall08-IDEAL-V9-reco-v2-/GEN-SIM-RECO
- /SingleTop-tChannel-Summer08-IDEAL-V9-v1-/GEN-SIM-RECO
- /SingleTop-tWChannel-Summer08-IDEAL-V9-v1-/GEN-SIM-RECO
- /SingleTop-sChannel-Summer08-IDEAL-V9-v1-/GEN-SIM-RECO
- /VVJets-madgraph-Fall08-IDEAL-V9-v2-/GEN-SIM-RECO

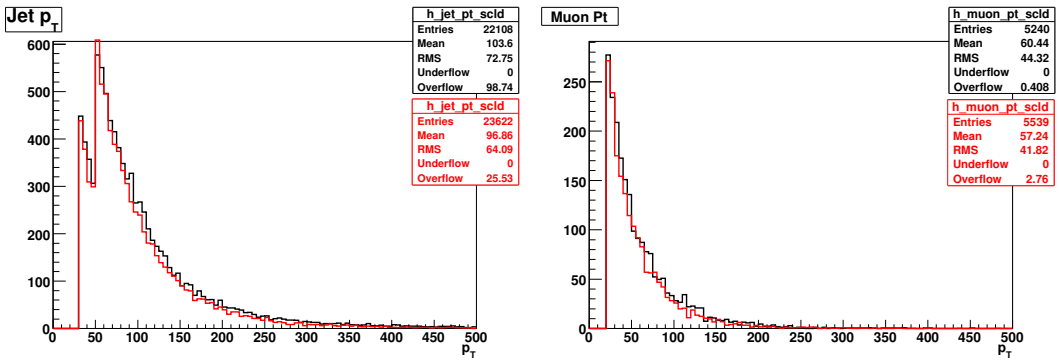
11.2.1 Performance

Figure 11.73 shows how well the procedure estimates the shape of the high MET tail, defined as $MET > 200$ GeV. The underestimate due to the SumPt cut is evident at the very highest end of the tail and amounts to about 20%. The method does not attempt to estimate the entire MET distribution, but only the high MET tail. At low MET, the physics is dominated by true single muon events, which is the reason for the stark difference between the estimated and actual values in this region. A simple ABCD method could be used to predict the single muon contribution and added to the dilepton prediction to give a full prediction for MET in single muon events.

11.2.2 Error Analysis

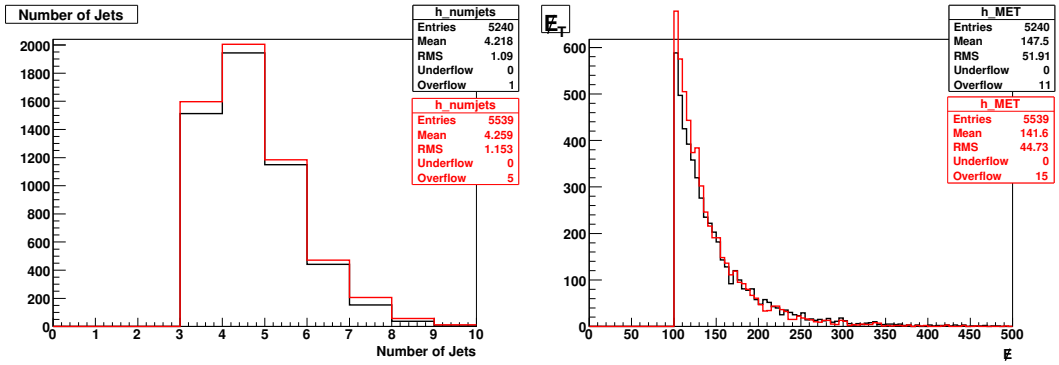
In order to study the consistency of the performance the large, 14 million event $t\bar{t}$ sample was split into 68 separate samples each containing 500 pb^{-1} worth of events. The procedure was rerun on all the samples and residual and pull distributions were made. Before we could make the pull distributions we needed to find our sources of error. If we calculate a residual, T:

$$T = N_{Predicted} - N_{Observed} \tag{31}$$



(a) Jet p_T

(b) Muon p_T



(c) Number of Jets

(d) MET

Figure 11.72: Validation plots of high level quantities. Our FastSim sample is in black while the MadGraph sample is in red. Both samples are scaled to 1 fb^{-1}.

$$\text{Pull} = \frac{T}{\delta T} \text{ where,} \tag{32}$$

$$\delta T = \sqrt{\alpha^2 N_{\mu\mu+\mu e} + \beta^2 N_{\mu\tau\to\text{jets}} + N_{Predicted}} \tag{33}$$

Since we're using 500 pb^{-1} pseudoexperiments we can assume our statistics are large enough to use Gaussian statistics and that $N_{\mu\mu+\mu e}$ and $N_{\mu\tau\to\text{jets}}$ are sufficiently large. The statistical errors on α and β are small due to the large sample size from which they are calculated.

Figure 11.74 shows the residual and pull distributions for these 68 pseudoexperiments. The consistent underestimate is evident in the residual plot. There is an average difference between prediction and observed of 1.5 events. In 500 pb^{-1} there are, on average, 12 events which would imply a 12.5% underestimate.

11.3 Performance of method with signal

To investigate the performance of the method with signal the same procedure was applied to an LM1 sample (/SUSY-LM1-sftsht-Summer08-IDEAL-V9-v1/GEN-SIM-RECO) which was generated with a cross section of 16.06 pb and simulated in CMSSW 2.2.6 using FullSim. After scaling, the LM1 and $t\bar{t}$ histograms were added and the MET estimation was compared to the true MET. In figure 11.75 there is good separation between signal and background, even in 100 pb^{-1} although there is clear contamination from signal in the estimate. Again, we're interested in the high MET tail, specifically where $MET > 200$ GeV, and it is here where the separation is greatest.

Figure 11.76 shows the results of the method against other LM points. In figure 76(a) we compare the MET_{True} in single muon events from LM0 and $t\bar{t}$ in black to our prediction in red. LM0 and $t\bar{t}$ have similar SumPt distributions which implies there is a large contamination in our reconstructed dimuon sample from LM0. This turns out to be

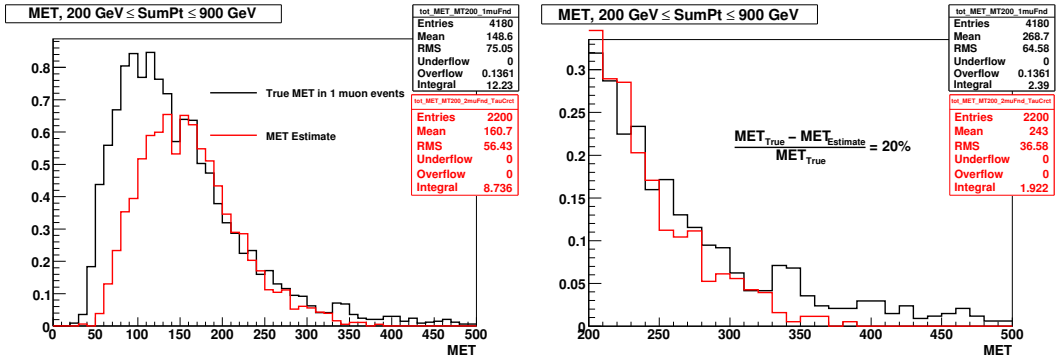


Figure 11.73: True single μ MET in black along with the prediction from dimuons in red (left). (Right) Same plot but zoomed into the tail region $MET > 200$ GeV

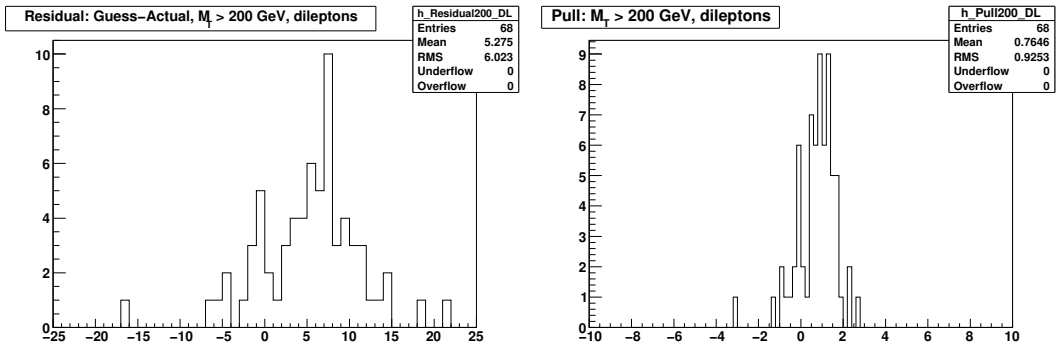


Figure 11.74: Residual (left) and pull (right) distributions for 68 pseudoexperiments of 500 pb^{-1} each

unimportant given the large cross section for LM0, 110 pb. The MET_{True} overshadows any prediction made by the method. So in a case such as this, it is not that the method works or does not work, but more that it doesn't matter to begin with.

There is good separation between SumPt in $t\bar{t}$ events and LM4, and thus our SumPt cut ensures a very pure sample of dimuon $t\bar{t}$ events. Even though the cross section for LM4 is just under half that of LM1, 6.7 pb, the clean dimuon sample from $t\bar{t}$ ensures good separation between MET_{True} and the prediction, even with 100 pb^{-1}, as seen in Figure 76(b).

11.4 Systematic uncertainties

Since MET measurements are closely related to jet energy measurements we expect jet energy measurement errors to play the most significant role in how well we can measure MET. To see the effects of jet energy mismeasurements on our MET_{Recalc} prediction we simulated five jet energy mismeasurement scenarios and compared the performance of the method to a baseline performance without excess jet energy miscaluclations. We model jet

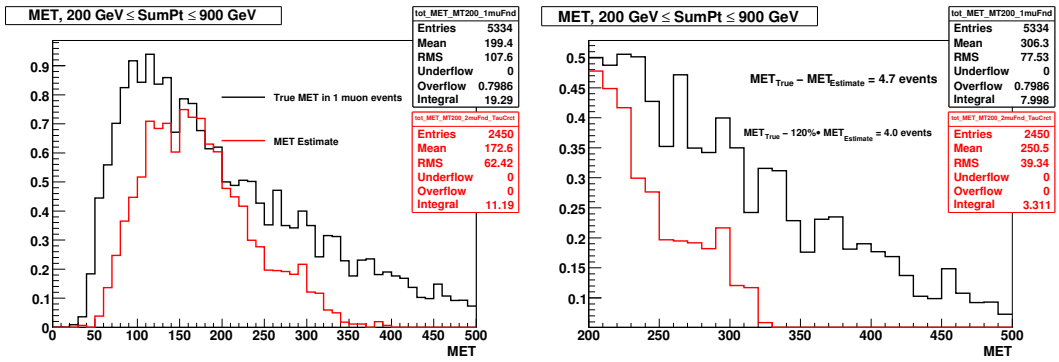
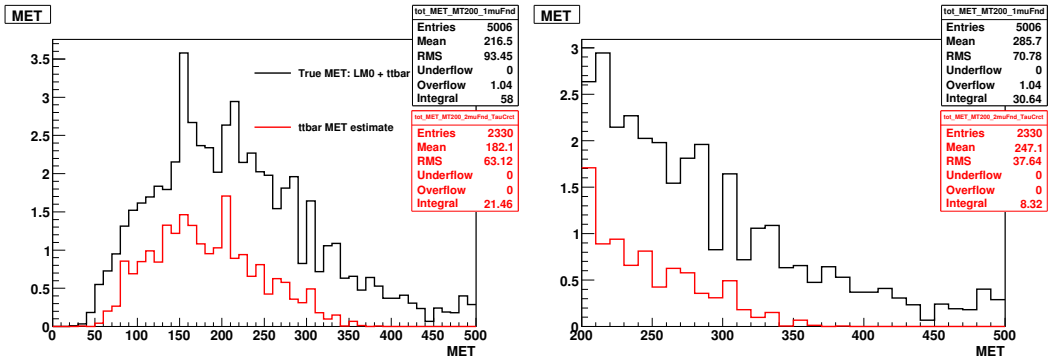
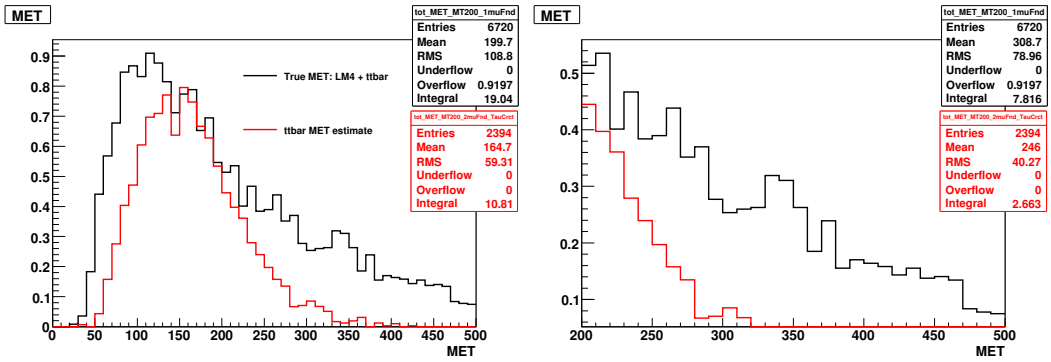


Figure 11.75: True single mu MET in LM1 + $t\bar{t}$ (black) compared to the prediction (red). The right plot shows the same thing zoomed into the region of interest.



(a) MET_{True} in single muon LM0 + $t\bar{t}$ events in black compared to $t\bar{t}$ prediction in red.



(b) MET_{True} in single muon LM4 + $t\bar{t}$ events in black compared to $t\bar{t}$ prediction in red.

Figure 11.76: A comparison of different LM points.

energy resolution effects by smearing the p_T of the jet using a random number chosen from a gaussian distribution centered at the p_T of the jet, with a width of 10, 20 and 30% of that p_T. We also changed the scale of the jet energy by 10% higher and lower for all jets. In both cases, the new jet energies were taken into account and the MET was adjusted accordingly.

Figure 11.77 shows that 10% differences in jet energy, whether it be total scale or resolution, produce little change between MET_{True}, in black, and the prediction, in red. The prediction consistently underestimates by 20% or less. However, the method breaks down if the jet energy resolution gets too large. This effect is primarily due to the SumPt cut on our dimuon sample, since the cut excludes the high SumPt tail. This same tail is enhanced in the single muon sample due to the smearing. With large jet energy smearing the dimuon sample is no longer a good representation of the high MET tails in single muon events. If we remove the SumPt cut, we can see from figure 11.78 that the prediction again returns to within 20% of the true value. However, energy resolution is not a constant but a function of the jet p_T. To more properly take that into account, we smear the jet energy by the function [1]:

$$\frac{\sigma_{p_T}}{p_T} = \sqrt{\left(\frac{4.35}{p_T}\right)^2 + \left(\frac{1.34}{\sqrt{p_T}}\right)^2 + (0.03)^2}\text{ , for } |\eta| \leq 1.4$$

$$\frac{\sigma_{p_T}}{p_T} = \sqrt{\left(\frac{5.03}{p_T}\right)^2 + \left(\frac{0.96}{\sqrt{p_T}}\right)^2 + (0.04)^2}\text{ , for } 1.4 \leq |\eta| \leq 2.4$$

In conjunction with the jet energy resolution effects already taken into account by the simulation, this effectively doubles the expected jet energy resolution. We see in figure 11.77 that the prediction agrees to within 20% with this doubling of the standard jet energy resolution.

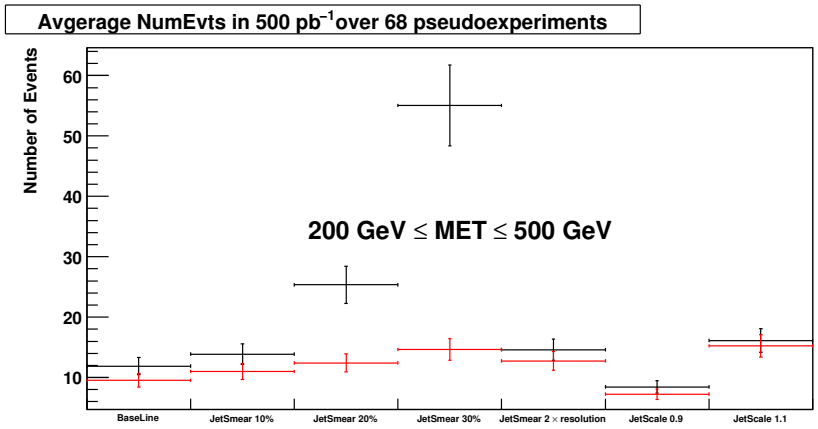


Figure 11.77: Average number of $t\bar{t}$ events with 200 GeV $\leq MET \leq$ 900 GeV over 68 pseudoexperiments, with 500 pb^{-1} of data each. Black is the true number of events while red is the prediction.

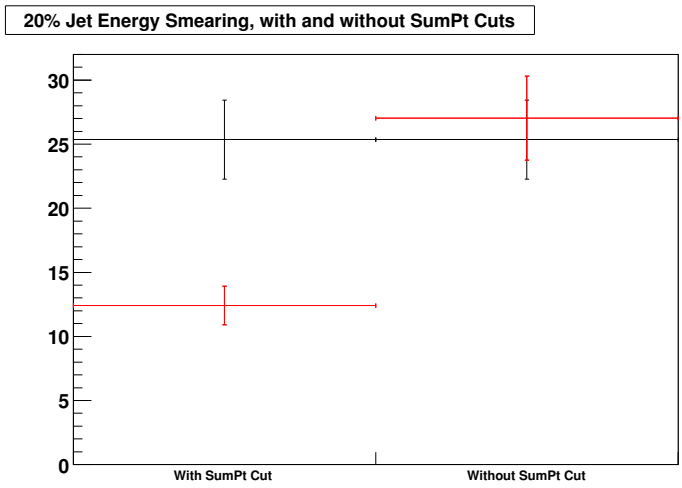


Figure 11.78: Average number of $t\bar{t}$ events with 20% jet energy smearing with and without a SumPt cut. Jet energy resolution has a large effect on the outcome of the prediction when a SumPt cut is applied.

11.5 Early data commissioning plan

With early data we'll want to test the method to verify the predictions. In order to do the verification we'll need a very pure sample of $t\bar{t}$. One way to get such a sample would be to use regions we expect to be dominated by background, i.e. the forward region. If the method works as expected in these forward regions then we can move to the central region.

Since the method fails if the resolution is too large we hope to get the jet energy resolution to within 15% of ideal using early data. This will require an in depth study of jets, perhaps looking and Z + 1j events.

11.6 Conclusion

Our method attempts to predict the high MET tails in single muon $t\bar{t}$ events in high M_T bins. In these regions, MET is dominated by two dilepton scenarios. In the first scenario, one of the two leptons, e or μ, is lost or escapes the veto. The other scenario involves top decays to taus, where the tau decays hadronically. We predict this effect by using reconstructed dimuon events and treating one of the muons as a lost lepton or a tau and adding its contribution to the MET. This is scaled to the expected contribution from Monte Carlo. The method works to within 20% and is robust against most types of realistic sources of systematic error. The method has been tested against three LM points and there is consistent separation between predicted MET from $t\bar{t}$ and expected MET due to signal. With early data we hope to test the prediction using forward regions as a way to select mostly $t\bar{t}$ events and measure the jet energy resolution to a significant level.

References

[1] The CMS Collaboration, "CMS Physics Analysis Summary", **CMS PAS TOP-09-005**

[2] V. Chetluru, *et al.*, "Jet Reconstruction Performance at CMS", **CMS AN-2009/067**

12 Combinatorial semileptonic top-pair reconstruction / Topbox method

W.Bender

12.1 Description of method

The combinatorial top reconstruction also called Topbox method is a method to estimate $t\bar{t}$ background. One of the tops decays leptonically while the other one decay hadronically. It is a basic method to estimate top background in a semileptonic SUSY scenario.

Under the assumption that the neutrino of the leptonically decaying W boson is responsibly for all missing tranverse energy the invariant mass of all participating particles can be reconstructed. After several corrections to the MET and the reconstructed jets, both tops can be reconstructed in 4 four steps:

- Searching for semileptonic top events, one leptonic W decays into a lepton and a neutrino. The Neutrino is responsible for missing energy. As this missing energy can only be messured by its transverse components $E_x$and E_y this compoments can be taken as the p_x and the p_y components of the neutrino. As there is no measured information about the p_z component of the neutrino, this component is calculated under the theoretical assumption of a W mass contrain. This means that p_z is calculated with an invariant mass of the neutrino and the lepton which equals the W mass m_w. As this is a quadratic equation, there is a possibility of having two solutions or no solution if $m_T > m_W$. In the second case the value of p_z is set to 0.
- The recontructed leptontic W is then used to reconstruct a leptonic top searching for an additional (b-)jet. As there is no b-tagging information available in the first data, all combinations from the pool of the for highest jets (in p_t) are probed to have a invariant mass as close to the top mass as possible. Together with the ambiguities to the p_z of the neutrino, the jet with the leptonic top mass closed to m_t is taken to be the leptonic (b-)jet. The invariant mass is labeled to be the leptonic top mass $m_{top,lep}$.
- The other hadronical top consists of three remaining jets. Two out of three belong to the same hadronically decaying W boson. The two jets with an invariant mass closest to mass of the W m_w are taken to be the hadronical W boson distribution $m_{W,lep}$.
- Finally there is a distribution of the hadronical recontructed invariant top mass $m_{top,had}$ of which a valid range on the allowed top mass can be choosen.

This additional variables can be used to cut on and to define a top box control region (see figure 12.79).

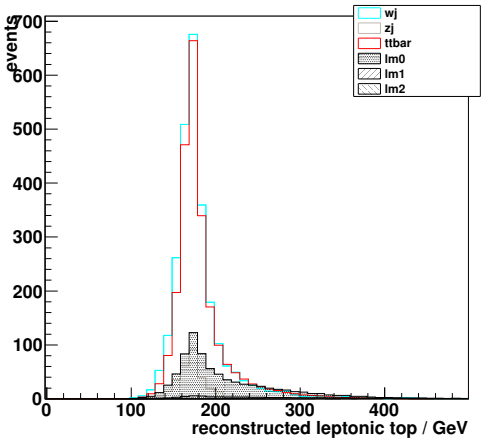
12.2 Topbox control sample

With the addional reconstructed variables of the topbox method a control region with enriched top events can be defined. Cuts on the topbox variables of $|m_{top,lep} - m_t| < 25GeV$, $|m_{W,lep} - m_t| < 15GeV$ and $|m_{top,had} - m_t| < 25GeV$ are applied. Additional cuts apart from the preselection cuts and the one on the topbox variables can be also included to reduce the QCD background which might contribute to low values of m_t.

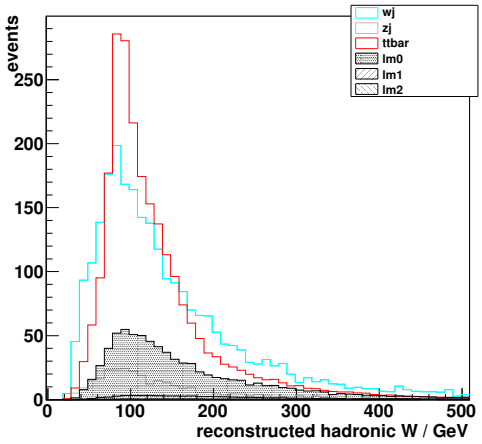
process	$t\bar{t} + jets$	Z+jets	W+jets	LM0	LM1	LM2
events	144 ± 2.1	5.7 ± 1.4	38 ± 4	15 ± 1	0.306 ± 0.06	0.033 ± 0.007

Table 12.21: *Expected number of events and uncertainty on MC statistics after the preselection and topbox cuts for the main background processes ($t\bar{t}$+jets, Z + jets, W + jets) and SUSY signals LM0,LM1 and LM2 scaled to a luminosity of 100pb^{-1}.*

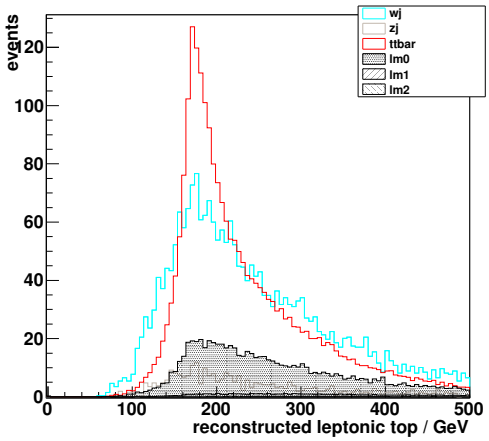
The number of selected events are displayed in table 12.21 for the myon channel.



(a) reconstructed leptonic top



(b) reconstructed hadronic W



(c) reconstrcted hadronic top

Figure 12.79: Reconstructed particles in $t\bar{t}$ events. This variables are used to cut on and define a control region.

12.3 Exprapolation of $t\bar{t}$ background in the signal region.

After defining a control region enriched with $t\bar{t}$ events, a scaling factor $R_t\bar{t}$ can be defined to extrapolate the number of top events into the signal region. This region is defined to have at least 3 jets above a threshold of $p_t > 50GeV$ and a missing transverse energy above $MET > 100GeV$. An overlap of the two regions is avioded by an inverted cut on the topbox variables in the signal region. The scaling factor $R_{t\bar{t}}$ is defined by the ratio of events in control and the signal regio.

$$R_{tt} = N_{tt}^{signal}(MC)/N_{tt}^{control}(MC)$$

When later first data is available, this factor can be used to scale the measured events in the control region to the number of events in the signal region.

$$N_{tt}^{signal}(data) = R_{tt} \cdot N_{tt}^{control}(data)$$

Of cause other events like the ones from W background has to be substracted to get $N_{tt}^{control}(data)$. The value of $N_{tt}^{signal}(data)$ is the amount of estimated top events relevant as SUSY background.

12.4 Impact of systematic uncertainties to the estimation

Systematic uncertainties are tue to the fact that there is only a limited knowlegde of the theoretical modeling and description in the reconstruction of particles. The biggest impact on the number of events in the control is exspected to be the jet energy scale which is asummed to be know at a precision of 5%. This leads to a systematic uncertainty of 15% of the number of top events in the control region. Other uncertainties are caused by additional backgrounds (like W+Jets) conminating the control region. As this samples has be be manually substracted from the measured data in the control region, the total aquatically added MC uncertainties on all samples of about 5% and the error on the background cross sections of about 3% (assumed to be 10% on W+Jets and Z+Jets each) has to be considered in the top estimation.

12.5 Performance of method with signal

In table 12.21 there is shown the number of events also for some SUSY signals. As these number of events are not very huge, this basically means that signal contamination seems to be in the order of 10% (for LM0) in worst case.

13 $t\bar{t}$ background using b tagging

13.1 Description of Method

A method to estimate the $t\bar{t}$ background in the $E_{\mathrm{T}}^{\mathrm{miss}}$ tail for the muon+jets channel using b-tagging will be presented in this section. The top pairs production is the dominant background in the leptonic SUSY searches while requiring for many high Pt_t jets, that's why the estimation of this background with data-driven techniques is a necessary step before claiming any discovery. The method described here provide an estimation of the number of the top pair events in the data but also the shape of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. In principle, this method could be extended to other distributions as long as they are not too correlated with the b-tagging information.

13.1.1 Selection

As the method rely on the requirement of 2 b-jets out of at least 4 jets expected for the $t\bar{t}$ events, the selection of the events is based on a 4 jets topology. In this section 2 different selections will be used:

- Selection 1: "loose selection" (S1)

  1 muon with P_t >20 GeV - 0 electron - 4 jets with $P_t > 50$ GeV

- Selection 2: "hard selection" (S2)

  1 muon with P_t >20 GeV - 0 electron - 3 jets with $P_t > 75$ GeV and a fourth one with $P_t > 50$ GeV

The selection of the muons, electrons and jets is described at the begining of this note in the section 3. By default, the selection 2 is applied to obtain the following results presented and will be called "Signal Region" (SR). Additionaly to the selection, a cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ is applied with a default threashold of 100 GeV in order to search for SUSY signal.

13.1.2 Principle

To estimated both number of $t\bar{t}$ events and $E_{\mathrm{T}}^{\mathrm{miss}}$ shape in the Signal Region, a template of $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution from data is obtained from data using a Control Region. The template obtained is then normalized to the number of events observed in the data in a given range of $E_{\mathrm{T}}^{\mathrm{miss}}$, below the $E_{\mathrm{T}}^{\mathrm{miss}}$ threashold applied for the search of SUSY. This range is called the Normalisation Region. By default this range is between 0 and the threashold applied on $E_{\mathrm{T}}^{\mathrm{miss}}$: [0,100] GeV. The principle of this method is pretty simple but the robutness mainly depends on the definition of the Control Region.

13.1.3 Definition of the Control Region

The defintion of the Control Region is the key ingredient of this method. It has to fulfill a list of requirements:

- being $E_{\mathrm{T}}^{\mathrm{miss}}$-shape invariant compare to the Signal Region

- being $t\bar{t}$ enriched which implies
    - being as "Standart Model background free" as possible
    - being as "SUSY free" as possible

The $E_{\mathrm{T}}^{\mathrm{miss}}$-shape invariance is the most important requirement in order to not biased the estimation. The Standard Model backgrounds contamination could in first approximation degrade the $t\bar{t}$ estimation but a use of other specific data-driven methods to estimate those backgrounds is still possible in case of important contamination and otherwise then can be obtained from the Monte Carlo prediction. The SUSY contamination of the Control Region will degrade the $t\bar{t}$ estimation and diminuish the observability of the SUSY signals as this could tend to an overestimation in case of important contamination. The last point which was not yet mention concerns the statistic of the Control Region sample. If the size of the control sample is too small, the statistical errors will be important and especially for an estimation of $E_{\mathrm{T}}^{\mathrm{miss}}$ shape which imply the use of various bins. A balance as to be found in order to fulfill all this requirements which could dependant on different parameters, mainly the luminosity, the selection and the SUSY parameters to exclude.

In this method, the definition of the Control Region is mainly based on the requirement of 2 b-tagged jets. The choice of the b-tagging algorithm and the associated working point is therefore important and can have an impact on the measurement's precision. It will be discuss in the following subsection.

In the addition to the requirement of the b-tagging, different variables are built and used to enrich the Control Region sample. As the SM backgrounds are already reduced by the high cuts selection, the main aim of these cuts is to reduce the contamination of the new physics in this control sample. In order to respect the $E_{\mathrm{T}}^{\mathrm{miss}}$-shape invariance, the variables used has to be uncorrelated with $E_{\mathrm{T}}^{\mathrm{miss}}$. It has to be mentionned that both semi-leptonic and di-leptonic $t\bar{t}$ channels contribute to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution with varying ratio. In order to estimate both components correctly, the variables are built using common kinematics properties. The presence of the 2 b-quarks and at least one muon is the common base of both channels. With this, many variables where built using angles, masses, energies but only three of them where retained:

- $H_T(b,b)$: the scalar sum of the P_t of the 2 b-tagged jets
- $\Delta R(b,b)$: the angle between the 2 b-tagged jets
- $Mass(b,l)$: the invariant mass of the muon and the closest b-tagged jet

13.2 Performance of method without signal

In this subsection we will focus on the performance of the method obtained without signal. First we will show the results on $t\bar{t}$ sample only and then with the addition of the SM backgrounds namely W+jets, Z+jets and QCD multi-jets.

13.2.1 Event yield

The two sets of selection where applied on the different dominant Standart Model processes. The number of events expected for a luminosity of 100 pb^{-1} are shown is the table 13.22. It has to be mentionned that for convenience reasons, a preselection was applying on the differents datasets used before the selection itself. This preselection was the requirement of at least 4 jets with $P_t > 15$ GeV $\&$ η <2.4 and at least one muon with P_t >15 GeV $\&$ η <2.1 without any additional quality cuts.

For both selection, the number of QCD is negligible and the Z+jets contamination does not exceed few percents. In principle, these numbers could be obtained safely from Monte Carlo in the future. Concerning the W+jets contamination, it is closed to 18% for both selection in the Signal Region and closed to 12% after $E_{\mathrm{T}}^{\mathrm{miss}}$ cut. This suggests that a method to estimate the W+jets background should be used before the application of the estimation of $t\bar{t}$. For the next estimations, it will be assumed that this numbers come from the Monte Carlo and we will see in the subsection dedicated to systematic, the impact of the incertitude on the W+jets prediction on the $t\bar{t}$ estimation.

13.2.2 Performance with different b-tagging algorithms

The core of the method imply the use of b-tagging, that's why various algorithms and working points where tried. Three categories of b-tagging where applied based on track counting method, impact parameters or secondary

cuts	Ttbar	Z+jets	W+jets	QCD
preselection	9552	4628	27080	2100000
1 muon	4028	2199	15780	6417
0 electron	4028	2199	15700	6417
4 jets[50-50-50-50 GeV]	473	15	107	8
$E_{\mathrm{T}}^{\mathrm{miss}}$ >100 GeV	109	2	17	0
2 b-jets	298	8	56	0
Mass(b,l)<160 GeV	265	2	17	2
$H_T(b,b) < 500$ GeV	164	0	0	0
$\Delta R(b,b) > 2.3$	158	0	0	0
$E_{\mathrm{T}}^{\mathrm{miss}}$ >100 GeV	34	0	0	0
4 jets[75-75-75-50 GeV]	245	8	56	2
$E_{\mathrm{T}}^{\mathrm{miss}}$ >100GeV	65	2	9	0
2 b-jets	163	8	56	0
Mass(b,l)<160 GeV	140	1	11	0
$H_T(b,b) < 500$ GeV	91	0	0	0
$\Delta R(b,b) > 2.3$	86	0	0	0
$E_{\mathrm{T}}^{\mathrm{miss}}$ >100 GeV	21	0	0	0

Table 13.22: This table shows the numbers of expected events for a luminosity of 100 pb^{-1} for differents processes, namely $t\bar{t}$, Z+jets, W+jets and QCD multi-jets. A preselection of 4 jets with a P_t >15 GeV & η <2.4 and at least one muon with a P_t > 15 GeV & η <2.1 was applying before the selection itself. One can see in this table the numbers obtained for the loose (S1) and hard selection (S2) used to defined the Signal Region and the Control Region.

vertex. For each of the algorithms, three working points called "loose", "medium" and "tight" are defined by the b-tagging POG to have an light quarks mistag rate respectively equal to 0.1, 0.01 and 0.001. [Ref to note BTV-09-002] For a mistag rate, the efficiency can vary, for example the b-quark effiency vary between 72% and 85% for the loose working point among the algorithms studied.

On the figure 13.80 is shown the distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ obtained while required 2 b-tagged jets using this differents of b-tagging conditions. These distributions can be compared to the reference distribution which is the expected distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ for $t\bar{t}$ in the Signal Region. All the curves are in agreement within their error bars and the comparison demonstrates the shape invariance of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution under the application of b-tagging.

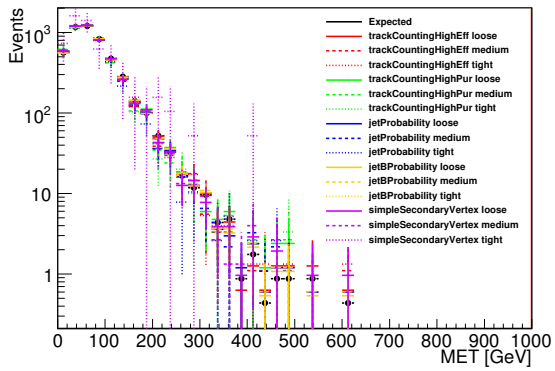


Figure 13.80: This plot presents different $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution obtained on $t\bar{t}$ sample. The black histogram corresponds to the expected distribution while the colored histograms are the shape obtained after applying various b-tagging algorithms and working points (loose, medium, tight) and then renormalized in the range [0-100] GeV.

The table 13.23 presents the number of $t\bar{t}$ events expected while using the full method with the use of several b-tagging algorithms. The results obtained for a luminosity of 100 pb^{-1} with the selection S2 show an agreement among the algorithms and no bias is observed. By example, a dispersion of about 5% is observed while the statistical precision is on the order of 20-30%. Harder is the cut on the b-tagging variable, smaller is the size of the control sample, higher are the statistical errors. To minimize these errors, it is proposed to use the "loose" b-tagging working point. A special case is the use of the jetBProbability algorithm where relative errors are closed

to be constant. This is due to the fact that the efficiency remains high while diminuising the mistag rate (0.85, 0.80, 0.79 for the 3 working points). This algorithm seems to be promising but we should not focus on it for the startup. The loose working point of the trackCountingHighEff will be the reference for the rest of this section and could be a good starting point for data-taking.

b-tagging algorithm	loose	medium	hard
trackCountingHighEff	57±14 (23%)	57±17 (30%)	56±19 (33%)
trackCountingHighPur	60±18 (30%)	53±24 (46%)	53±27 (52%)
jetProbability	58±14 (23%)	54±17 (31%)	50±23 (47%)
jetBProbability	60±13 (22%)	60±14 (23%)	59±14 (23%)
simpleSecondaryVertex	61±17 (29%)	59±20 (33%)	41±88 (214%)

Table 13.23: Number of $t\bar{t}$ events estimated using various b-tagging algorithm and working points. This numbers are obtained using the hard selection (S2), the default defining of the Control Region and a luminosity of 100 pb^{-1}.

13.2.3 Control Region variables

The three variables used to define the Control Region are already presented in the first subsection, we will now discuss about the cuts applied on them. Two of the variables are build with the 2 b-tagged jets required. We will first discuss about the ΔR between both b-jetst which tends to be high for the top pair production. As we can see in the figure 13.81, the distribution peacks close to π. In such process, the two top quarks tend to be back to back and for high P_t tops the decay products tend to be collimated. By consequence, the two b quarks tend also to be back to back. Higher is the P_t theashold applied on jets, more visible is this effect. Tests on other angular variables like $Delta\eta$ or $Delta\phi$ where performed but the most powerful variable remains ΔR.

Cutting on such variable can also affects the shape of $E_{\mathrm{T}}^{\mathrm{miss}}$ as both variables are not completely uncorrelated. To estimate the distorsion introduced while cutting on the variables of the Control Region, a estimator called χ^2 is built and defined as following:

$$\chi^2 = 1/Nbins \sum_{bins} (n0 - n1)^2 / max(\sigma_{n0}, \sigma_{n1})^2 \tag{34}$$

where Nbins is the number of bins for the two histograms, n0 and n1 bins values, σ_{n0} and σ_{n1} their associated defined for both histograms.
In order to choose a value for the cut apply on this variable, the efficiency plot versus the cut is presented on the figure 13.87 as well as the χ^2 curve. A cut of $\Delta R >$2.3 is retained and gives an efficiency of about 67% for a still small χ^2 equal to 3.10^{-5}.

The second variable only based on the 2 b-tagged jets is $H_T(b,b)$. The distribution of this variable is presented on the figure 13.81. The threashold is determined as 2 times the P_t threashold applied on jets, then the distribution peaks around 200 GeV and the distribution before falling down. As the SUSY can appear with high P_t objects, it is envisageable to cut on the tail of this variable. A cut of $H_T(b,b) <$ 500 GeV is applied later on and corresponds to an efficiency of 95% for a chi^2 of 2.10^{-4}. This variable is more correlated to $E_{\mathrm{T}}^{\mathrm{miss}}$ and less powerfull than the previous, but the performance to reject SUSY has to still to be shown.

The last variable used is the invariant mass of the lepton and the closest b-tagged jet. For $t\bar{t}$ events, we expect to identify a b-quark and a lepton coming from the desintegration of the same top which imply that the mass of both objects should be smaller than the mass of top. The distribution presented in the figure 13.81 confirms this and the maximum observed is closed to 100 GeV. The tail of the distribution for $Mass(b,l) > Mass(top)$ is mainly explained by the resolution of the objects and the misidentification of the objects. A cut of $Mass(b,l) <$160 GeV is proposed which corresponds to an efficiency of 86% for a χ^2 of 9.10^{-5}

The default definition of the Control Region used later on can be summarized as following:

- Two "loose" b-tagged jets among the 4 selected jets using the trackCounting algorithm
- $\Delta R(b,b) > 2.3$
- $H_T(b,b,) <$ 500 GeV
- $Mass(b,l) <$ 160 GeV

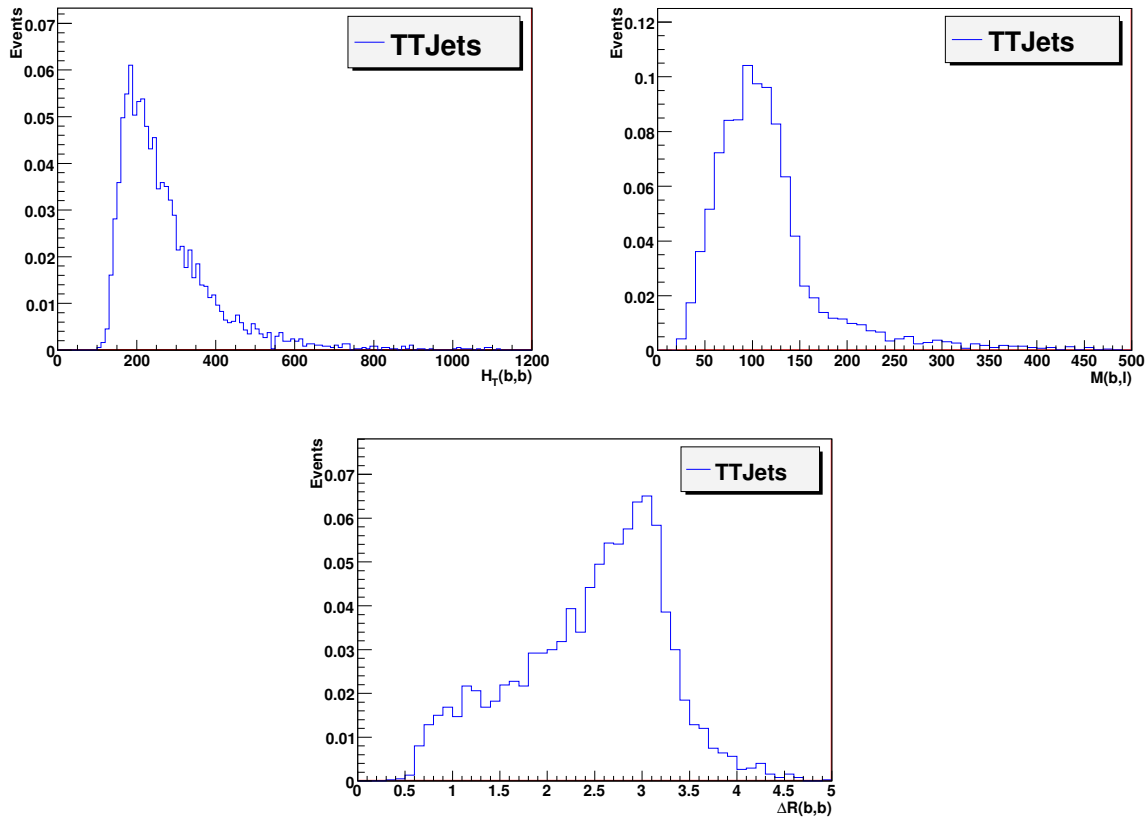


Figure 13.81: Distributions obtained for the three variables used to define the Control Region, namely the scalar sum of the P_t of 2 b-tagged jets $H_T(b,b)$, the invariant mass of the selected muon and the closest b-tagged jet $Mass(b,l)$ and the angle between those 2 particles $\Delta R(b,l)$. The distributions are computed on $t\bar{t}$ events after the hard selection (S2) and the requirement of at least 2 loose b-tagged jets using TrackCountingHighEff algorithm.

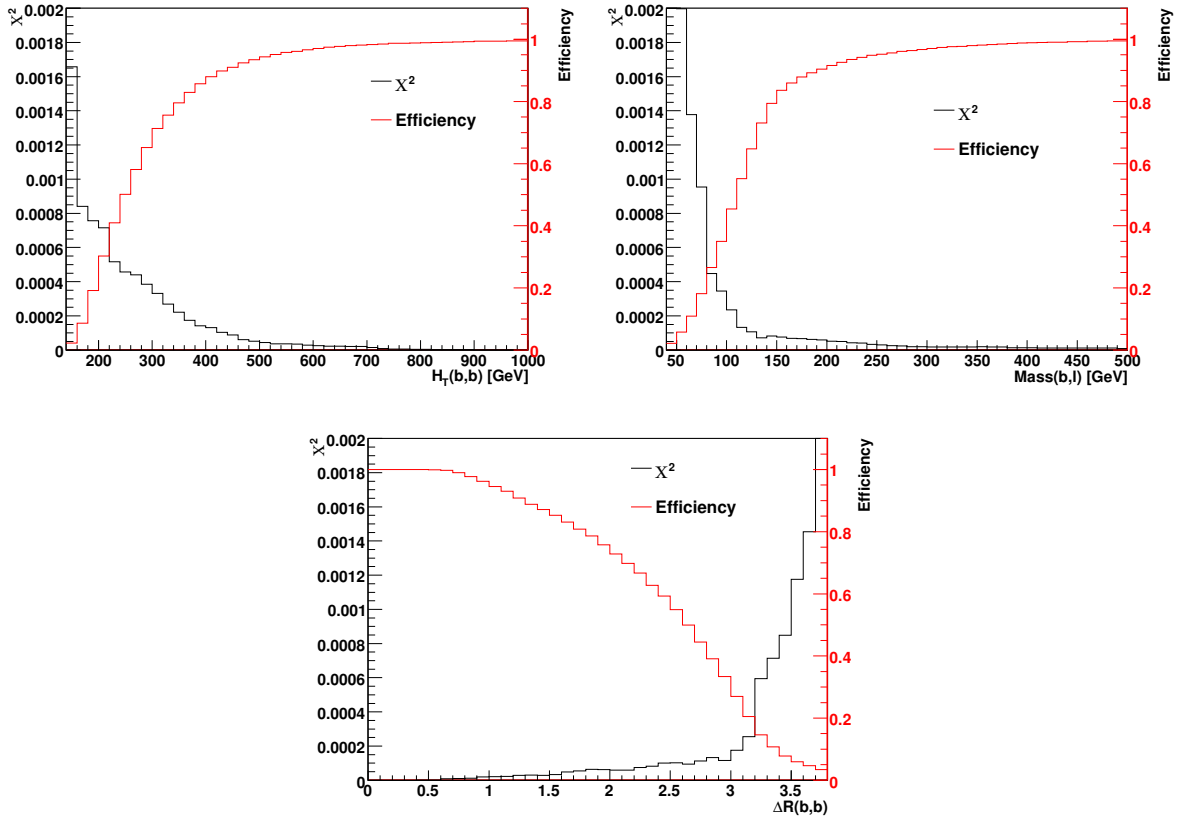


Figure 13.82: Performances obtained while cutting over the variables shown in the figure 13.81. The red curve show the efficiency versus the cut apply on the variable while the the black curve show the value of an estimator called χ^2 which evaluate the distorsion of the resulting distribution after cutting compared to the original one.

We will now focus on the results obtained within this conditions. The figure 13.83 show the distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ obtained whiled applying these different cuts. On the left plot, the cuts are applied independantly while on the right plot they are sequentially applied. Both plots demonstrate that the cuts applied doesn't affect the shape of $E_{\mathrm{T}}^{\mathrm{miss}}$ more than a few percents effect on a large range of $E_{\mathrm{T}}^{\mathrm{miss}}$. This means that the Control Region defined here fulfill the first requirement described previously: the $E_{\mathrm{T}}^{\mathrm{miss}}$-shape invariance.

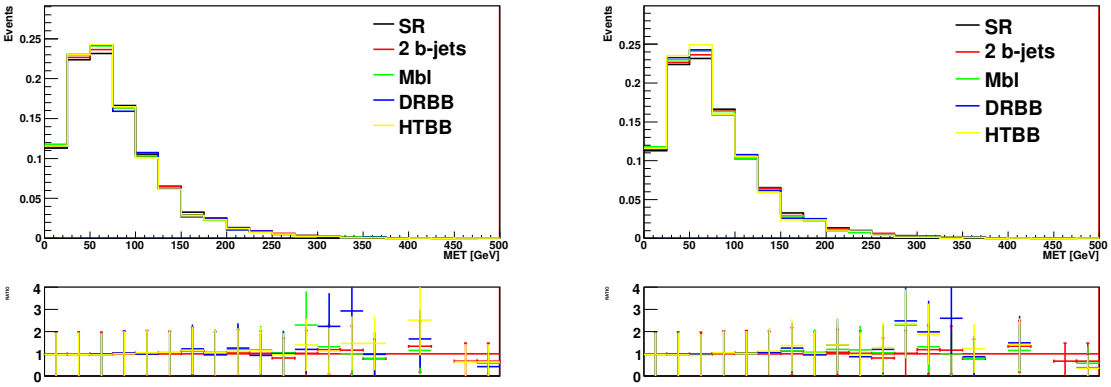


Figure 13.83: These plots show the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution shape obtained on $t\bar{t}$ sample with different conditions. On the top right, the distribution are obtained for different cuts applied independently while on the top left the distribution are obtained adding sequentially the cuts. On the bottom are show the ratio of the distributions compared to the expected $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution.

The figure 13.84 presents the results obtained with this method on the $t\bar{t}$ sample. The left plot compares the expected and estimated $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions and their ratio: no bias is observed. The right plot presents the expected and estimated distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ normalized to unity for the inclusif sample, the semi-leptonic events and the di-leptonic events. The semi-leptonic distributions are closed to the inclusive distributions as it is the domininant component with the selection applied. The important point is that both expected and estimated distributions for di-leptonic component are closed to each other. This demonstrates that the method doesn't bias the estimation of this component which start to be relatively important at high $E_{\mathrm{T}}^{\mathrm{miss}}$.

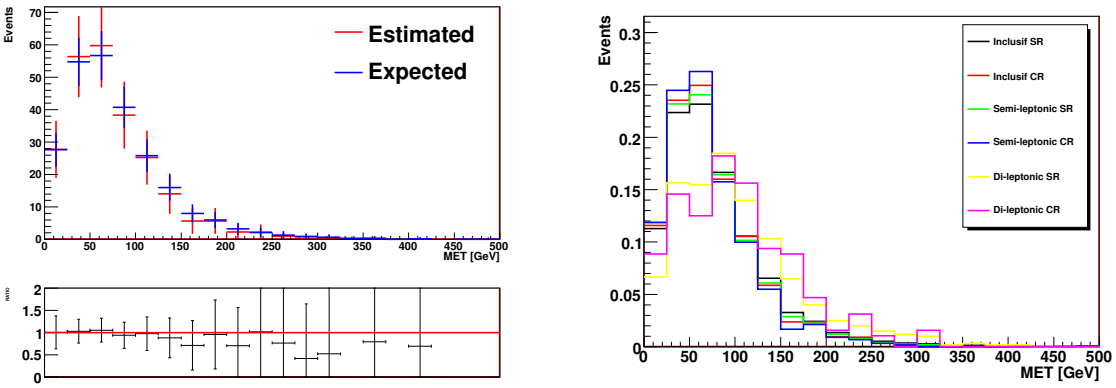


Figure 13.84: The left plot shows the expected (in blue) and estimated (in red) distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$. The estimation is performed using the full definition of the Control Region described in this section. The ration between curves is shown below. The right plot shows the normalized (to unity) distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution obtained in different conditions: in the Control Region compare to the Signal Region for the inclusif sample, the semi-leptonic $t\bar{t}$ events and the di-leptonic $t\bar{t}$ events.

The final result obtained on the $t\bar{t}$ sample is the plot of the statistical errors versus the luminosity. It is presented in the figure 13.85 for the different b-tagging conditions described previously. The spread is important and can be explained by the performances of the b-tagging algorithms in term of efficiency. The important numbers derived form this curve are the more precised measurements expected for 10, 20, 50, 100 and 500 pb^{-1} which are respectively 66%, 50%, 32%, %22 and 10%. This numbers are obtained with the default definition of the Control Region

but for low luminosity scenario it is possible to adapt the definition of the Control Region in order to diminuish the statistical errors which will be the dominant errors at startup.

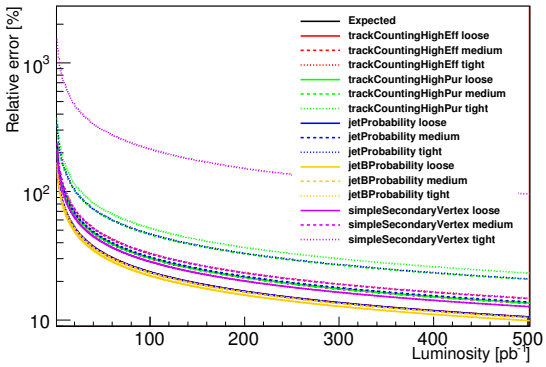


Figure 13.85: Statistical errors versus the luminosity computed for the use of different b-tagging algorithms and differents working points (loose, medium, tight).

13.3 Performance of method with signal

13.3.1 Control Region variables

The results presented in this subsection are obtained on $t\bar{t}$ sample with the addition of SUSY samples obtained with specific set of parameters and called LMX. The figure 13.86 presents the distribution of the variables used to define the Control Region for all these samples. The $H_T(b,b)$ distribution has a longer tails for the SUSY processes compare to $t\bar{t}$. This means that the b-quarks are mainly produced with higher P_t than the ones coming from the top decay. LM0 is a particular case because the distribution is closer to the $t\bar{t}$ distribution than the other LMX points.

The $Mass(b,l)$ distribution tends also to have a longer tail than the $t\bar{t}$ distribution as expected. The $\Delta R(b,b)$ distribution is less peacked at π than the $t\bar{t}$ distribution and the shape of these distributions are closed to each other for the different LMX samples.

One of the aim of the Control Region is to enriched the sample in $t\bar{t}$ events and reject the SUSY events. In order to estimate the discrimination power of this variable as function of the cut used we defined a variable as the production of the efficency cut on $t\bar{t}$ events times the rejection power on SUSY sample. Maximising this variable is equivalent to increasing the significance of $t\bar{t}$ events in the Control sample where SUSY is seen as a background and $t\bar{t}$ as the signal. The results are shown in the figure 13.87. For $H_T(b,b)$ the curves are really spread for the different LMX samples and this is due to the relative branching ratio of the processes involved. The maximum is found between 300 GeV and 450 GeV depending on the LMX point. A cut at 500 GeV was previously defined mainly for correlation reason between $E_{\mathrm{T}}^{\mathrm{miss}}$ and $H_T(b,b)$, and this cut can remains the same but an optimisation is obviously envisageable. For $\Delta R(b,b,)$ the curves are spread as well and the maximum vary between 120 GeV and 170 GeV which means that the 160 GeV cut previously defined is contained in this interval. It seems to be a good compromise between the SUSY rejection, the maintenance of the shape invariance and the minimization of the statistical errors. For $Mass(bl,)$, the spread of the curves is really less important which makes the application of such cut more less parameter sensitive. The maximum is between 2.2 and 2.4 which means that the cut of 2.3 applied does make sense as it will also maximize the SUSY rejection.

13.3.2 Event yield

The number of events expected for the various SUSY scenarii computed for a luminosity of 100 pb^{-1} can be found in the table 13.24. This numbers are presented for the 2 selection discribed and for both Signal and Control Region. The rejection factor is between 4 and 10 between both regions, as a comparison this number was closed to 3 for $t\bar{t}$ (cf table 13.22). For the majority of the LMX scenarii, the rejection factor is high (around 10) and this should have an impact on the significance esimated. But for the LM0, LM3, LM8, LM9 the rejection factor is not so important (around 4-5) and this will diminuish the observability of the signal as the estimation will be enhanced by the SUSY contamination.

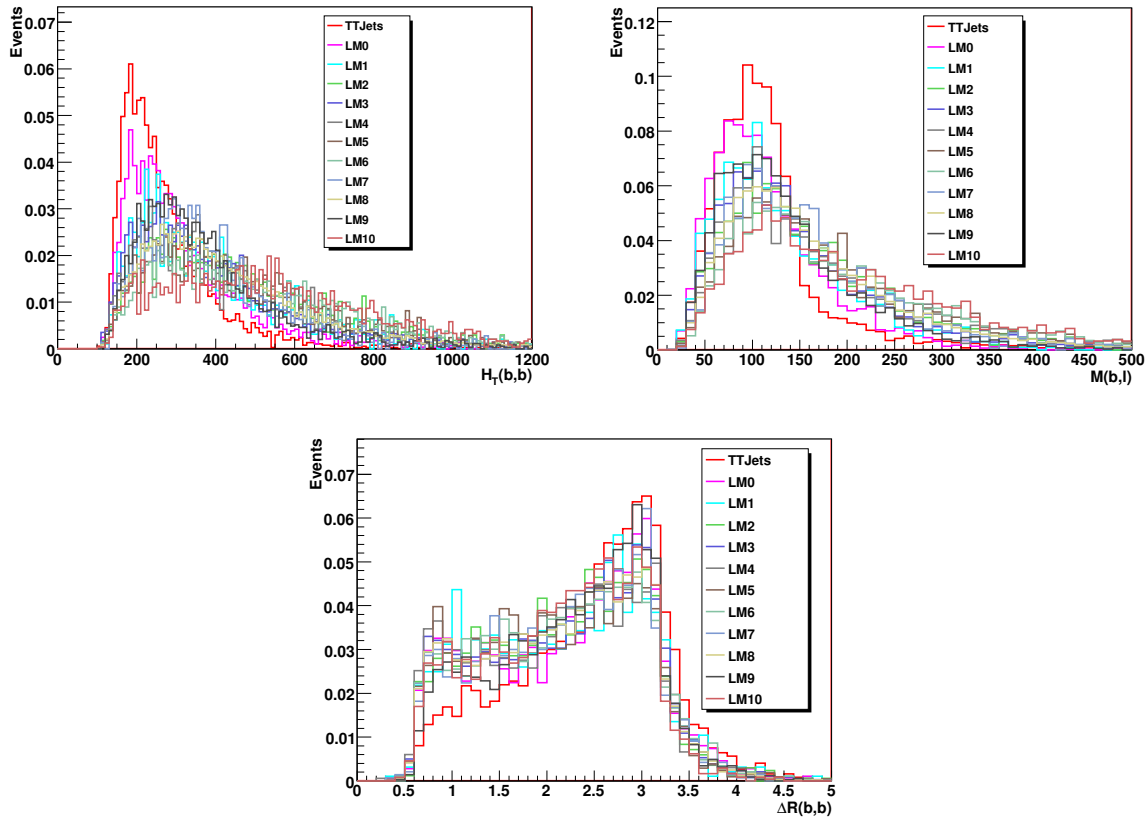


Figure 13.86: Distributions of the three variables used to define the Control Region as described in the figure 13.81, for $t\bar{t}$ and the different SUSY samples used: LM0 to LM10. The distributions are computed after applying the hard selection (S2) and the requirement of at least 2 loose b-tagged jets using TrackCountingHighEff algorithm.

cuts	LM0	LM1s	LM2	LM3	LM4	LM5	LM6	LM7	LM8	LM9	LM10
1 muon	1079	125	17	59	56	17	21	31	45	86	1
0 electron	1079	125	17	59	56	17	21	31	45	86	1
4 jets[50-50-50-50 GeV]	320	29	6	27	22	7	5	9	28	30	0.1
$E_{\mathrm{T}}^{\mathrm{miss}}$ >100 GeV	208	26	5	22	19	7	5	7	24	20	0.1
2 b-jets	216	15	4	18	13	5	3	5	22	23	0.08
Mass(b,l)<160 GeV	176	11	2	12	8	3	2	3	13	17	0.04
$H_T(b,b) < 500$ GeV	89	5	1	6	3	1	0.7	2	6	9	0.02
$\Delta R(b,b) > 2.3$	80	4	0.7	5	3	0.8	0.5	1	5	8	0.01
$E_{\mathrm{T}}^{\mathrm{miss}}$ >100 GeV	45	3	0.6	4	2	0.7	0.5	1	5	8	0.01
4 jets[75-75-75-50 GeV]	220	24	5	22	19	6	5	8	25	26	0.1
$E_{\mathrm{T}}^{\mathrm{miss}}$ >100GeV	148	21	5	18	16	6	5	6	21	17	0.08
2 b-jets	155	13	3	15	11	4	3	5	20	20	0.08
Mass(b,l)>160 GeV	123	9	2	10	6	2	1	3	12	14	0.04
$H_T(b,b) < 500$ GeV	64	4	1	5	3	1	0.7	1	6	8	0.01
$\Delta R(b,b) > 2.3$	56	3	0.6	4	2	0.7	0.4	1	4	6	0.01
$E_{\mathrm{T}}^{\mathrm{miss}}$ >100 GeV	33	3	0.5	3	2	0.6	0.4	0.8	3	4	0.008

Table 13.24: This table shows the numbers of expected events for a luminosity of 100 pb^{-1} for differents SUSY samples: LM0 to LM10. The numbers are obtained for both loose (S1) and hard selection (S2) used to defined the Signal Region and the Control Region. Contrary to the table 13.22, no preselection was applied.

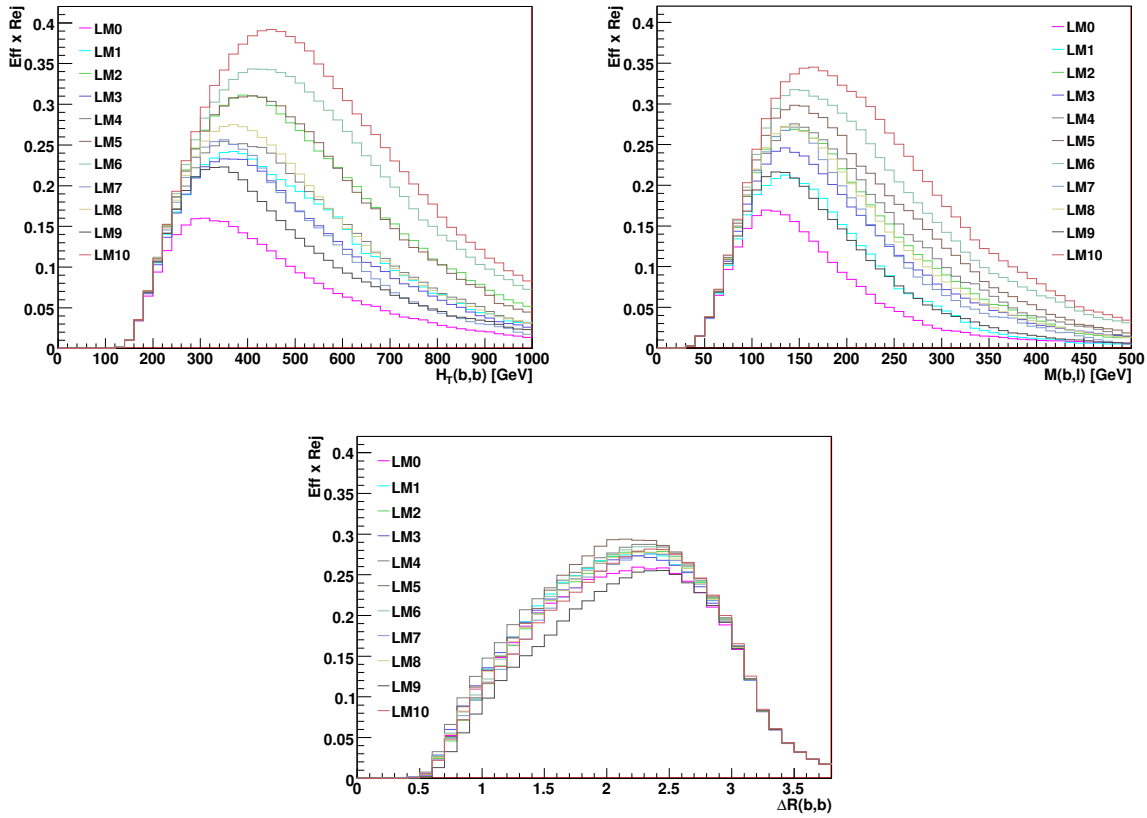


Figure 13.87: Performance obtained while cutting over the variables shown in the figure 13.86. The curves correspond to the value of an estimator which is defined as the efficiency of applying a cut on the variable for $t\bar{t}$ sample multiplied by the rejection of applying this cut on the corresponding SUSY sample (Eff x Rej).

13.3.3 $t\bar{t}$ estimation with SUSY contamination

The contamination of SUSY can appear in two different ways. There is first a possible contamination of the Control sample as discussed previously. The main way to reduce it would be by tuning the choice of the b-tagging algorithm and working point as well as the choice of the variables used in Control Region definition and their cuts. The second possible contamination appears in the normalisation procedure. The figure 13.88 presents the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for $t\bar{t}$ and for the SUSY scenarii. As excpected, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is higher in the SUSY events compared to $t\bar{t}$ events, netherthless SUSY events are also present in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ range which means that there is no possible normalisation region SUSY free. To define this normalisation region which is a range in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, a balance as to be found between the minimisation of the SUSY contamination and the minimization of the statistical errors which tends to opposite direction.

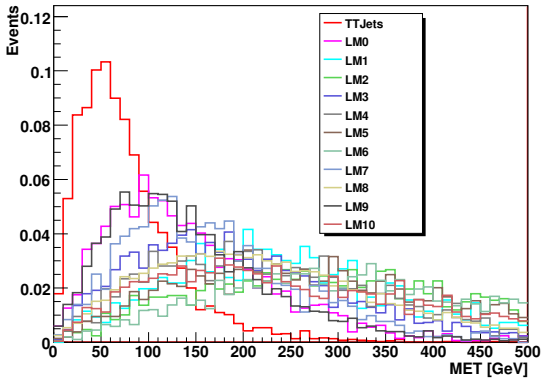


Figure 13.88: $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution of $t\bar{t}$ and the various SUSY samples obtained with the hard selection (S2).

In order to quantify this effects, a scan on the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut (X) was performed in such way that the normalisation of the template obtained from the Control Region was done in the $E_{\mathrm{T}}^{\mathrm{miss}}$ range [0,X] and the computation of the significance for a given SUSY scenario in the $E_{\mathrm{T}}^{\mathrm{miss}}$ range [X,∞]. The expected and estimated significance computed as $S/\sqrt{(}S+B)$ and their ratio are presented in the figure 13.89. The expected significance have a maximum for a $E_{\mathrm{T}}^{\mathrm{miss}}$ cut which is between 125 GeV and 175 GeV. The ration between expected and estimated significance shown in the right plot is not flat and not egal to unity for all scenarii. This means first that the estimation could differ from the expectation. Three categories of estimation can be dinstinguished. For the five SUSY scenarii for which the ratio is closed to unity LM1, LM3, LM5, LM8 and LM9 which means that there is no bias on the estimation. The case where the ratio is lower that one. By example for LM2, LM5, LM6 and LM7, the ratio is close to 0.5 at low MET cut, which means that the estimated significance is divided by a factor and the observability is reduced. The extrem case is the LM10 point where the expected significance is so small that the discovery could not be envisage via the current technique. The last case is when the ratio is higher than one, which is the case for the LM0 point. This effect is more complex because it implies a high SUSY contamination in both Normalisation and Control Region. This situation as to be avoid in the sense that the estimated significance is high than the expectation. But one way to observe this is to study the stability of the $t\bar{t}$ sample varying the $E_{\mathrm{T}}^{\mathrm{miss}}$-cut. In principle, the estimation should remains the same and a huge variation as observed here, could be a sign of high cross-section new physics signature.

13.4 Systematic uncertainties

There is two categories of systematic uncertainties, the ones which came from theory and the others which came from experiment. Among the experimental systematics the jet energy scale (JES) is expected to be the most important one. The tables 13.25 and 13.26 present the main results for different jet energy scale varying from -30% to +30% respectively for the 2 selection used S1 and S2. The main conclusion is that a incertainty of 10% on the JES can be translated by an incertainty of 30% of the number of $t\bar{t}$ events expected. This as to be compared with the 30% of statistical errors achievable for 50 pb^{-1}.

The uncertainty coming from knowledge of the b-tagging effiency can be derived from results obtained while varying the algorithms and the working point. This can be neglected compare to JES as it will mainly change the statistic of the Control sample by about 20% for 10% on the b-tagging efficiency which will change the errors by only few pourcents (4% for L=100pb^{-1} with trackCounting algorithm).

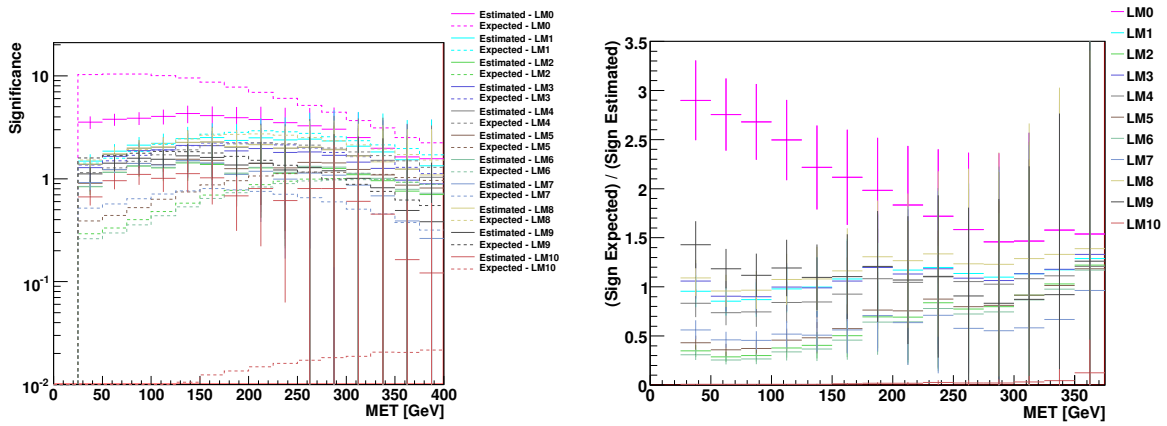


Figure 13.89: The left plot present the expected and estimated significance computed on the various SUSY samples for different cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ (x-axis). The Normalisation Region is defined as [0-$E_{\mathrm{T}}^{\mathrm{miss}}$-cut] while the Signal Region is defined for $E_{\mathrm{T}}^{\mathrm{miss}}$ >$E_{\mathrm{T}}^{\mathrm{miss}}$-cut. The right plot shows the ratio of the expected and estimated significance obtained in the same conditions. The significance is defined here as S/$\sqrt{S+B}$.

	Events				Signifcance	
	observed	estimated	observed	estimated	expected	observed
JES	$t\bar{t}$		$t\bar{t}$ + LM0		$t\bar{t}$ + LM0	
-30%	34±6	53±7	120±11	53±43	7.8	6.2
-20%	53±7	79±9	179±13	79±52	9.4	7.5
-10%	78±9	112±11	244±16	112±60	10.6	8.5
0%	109±10	150±12	316±18	150±68	11.6	9.4
+10%	137±12	178±13	380±19	178±73	12.5	10.4
+20%	168±13	210±14	445±21	210±78	13.1	11.1
+30%	197±14	240±15	502±22	240±83	13.6	11.7

Table 13.25: Impact of the jet energy scale on the number of events expected and estimated for different scenarii: from -30% to +30% in presence of LM0 signature or without SUSY signal. This results are obtained with the loose selection (S1) and with the use of the default definition of the Control Region. The numbers are obtained for a luminosity of 100 pb^{-1}.

	Events				Signifcance	
	observed	estimated	observed	estimated	expected	observed
JES	$t\bar{t}$		$t\bar{t}$ + LM0		$t\bar{t}$ + LM0	
-30%	16±3	14±18	28±5	66±8	6.2	4.7
-20%	28±5	26±24	48±7	107±10	7.7	5.7
-10%	45±7	41±29	71±8	158±13	9.0	7.0
0%	65±8	57±38	94±10	213±15	10.1	8.2
+10%	89±9	80±40	124±11	273±17	11.1	9.0
+20%	113±10	103±46	152±12	335±18	12.1	10.0
+30%	142±12	129±52	186±14	396±20	12.7	10.5

Table 13.26: Impact of the jet energy scale on the number of events expected and estimated for different scenarii: from -30% to +30% in presence of LM0 signature or without SUSY signal. This results are obtained with the loose selection (S2) and with the use of the default definition of the Control Region. The numbers are obtained for a luminosity of 100 pb^{-1}.

13.5 Early data commissioning plan

The method presented in this section was shown to predict the number of $t\bar{t}$ events with a statistical precision of about 20% and systematic precision of about 30% mainly coming from JES for a luminosity of 100 pb^{-1}. This analysis cannot be applied directly to the early data but this is a proposal of studies which could be performed before performing the $t\bar{t}$ estimation with the first data.

- Estimation of QCD multi-jets.
  It was discussed according to the selection table obtained, that the QCD background could be neglected in this method and that the number could be derived from Monte Carlo. As the uncertainty on this process are high, one could think about estimating this background while relaxing the selection criteria in order to control this assumption. If the background appears to be more important than expected, the selection could be tuned hardering the cuts or an estimation of QCD before the $t\bar{t}$ estimation could be envisaged.

- Estimation of W+jets.
  The W+jets background is not negligeable in the signal region sample. As the cross-section of this background is higher than the $t\bar{t}$ cross-section, there is the possibility to estimate W+jets with the selection scenarii described here before applying the $t\bar{t}$ estimation procedure.

- b-tagging study.
  As the $t\bar{t}$ estimation is based on b-tagging, the impact of the b-tagging use in the selection could be study. If the Standard Model backgrounds are well described by the Monte Carlo or well estimated by data-driven method as suggested previously, it's possible to study the shape invariance while applying the b-tagging. While apply sequentially the requirement of 1 and then 2 b-tagged jets after SM background substraction, one should expect a shape invariance of $E_{\mathrm{T}}^{\mathrm{miss}}$. It this is not the case, three interpretations are possible. The first one is that the SM background are not well described by Monte Carlo or the data-driven methods are biased. The second would be a sign of an important variation of the b-tagging efficiency as function of the P_t of the jets which could explain such behaviour. And the last but not least, would be a first sign of an important new physics contamination in the distributions.

- Variables of the Control Region.
  The comparison of the distribution of the three variables used in the Control Region from data and from Monte Carlo is a necessity. The study of their shape, their correlation and their sensitivity systematic effects are important to study the robustness of the method with data.

- Cut tuning.
  There is the possibility to tune first the selection and then the cuts applying on the variable of the Control Region. On can think about simplifying the definition of the Control Region using only one variable instead of three. In order to put exclusion limits on specific SUSY parameters, it's also possible to tune the cuts to thess parameters while maximising the discrimination power for this particular point.

- Estimation.
  For the low luminosity, the priority will be on the estimation of the number of events itself without considering the shape of MET. While increasing the statistic, a first shape could be obtain with a binning adapted to the precision desired.

14 W helicity and charge asymmetry in W+jets events and applications

In contrast to LEP and TEVATRON the initial CP states and the according CP inverted initial states are not produced in equal amounts at LHC. Furthermore, the energies and luminosities where not sufficient to produce large amounts W with large transverse momentum P_T (>100 GeV). These facts and different W production mechanisms make the following discussions unique to LHC. At LHC the W-boson spin is not aligned with the beam-axis as has predominantly been the case at TEVATRON and LEP. Thus charge asymmetries due to polarization do also effect observables in the transverse plane.

The production mechanism dominating [**?**] includes a stiff (valence) quark, hence the production of W^+ (from u-quark) is increasingly enhanced with increasing W P_T, as can be seen in Fig 14.90. The polarization mechanisms as discussed in [**?**] lead to a predominately left-handed W helicity (Table 14.28). Solely the W helicity determines how on average the momentum of the W is divided to the leptonic decay products of a certain helicity. Given

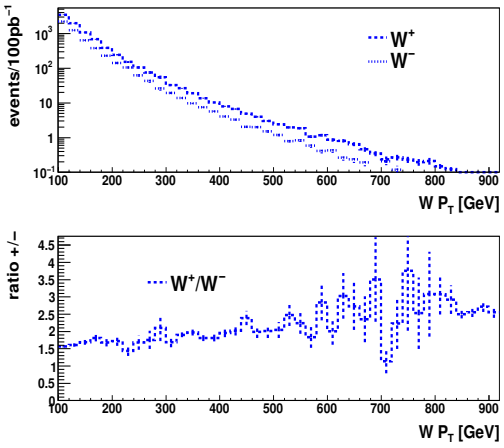


Figure 14.90: The W P_T distributions for the different charges and their ratio.

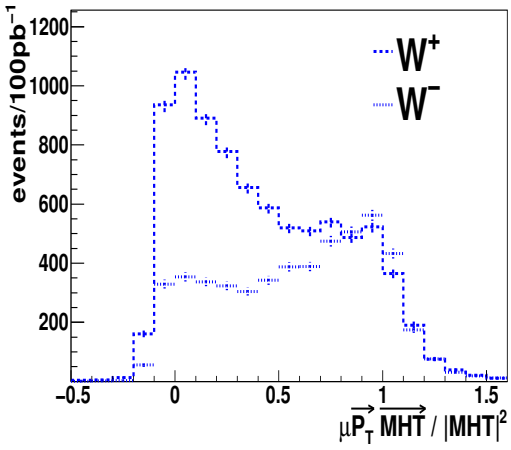


Figure 14.91: Neutrino P_T distribution for the different W-boson charges and their ratio.

Table 14.27: Helicity parameters of the W-boson for the ALPGEN generator, with $P_T > 100$ GeV.

sample	W-boson charge	left-handed	right-handed	longitudinal
ALPGEN	+	0.599 ± 0.010	0.274 ± 0.008	0.126 ± 0.013
ALPGEN	-	0.572 ± 0.013	0.300 ± 0.010	0.126 ± 0.017

that the charged leptons from the W^+ decay are right-handed and the charged leptons from the W^- decay are left-handed the effects are opposite for the different charges. Meaning that the positively charged lepton typically is given less of the momentum from W than it is the case for the negatively charged lepton. Figure 14.91 shows the projection of the charged lepton momentum to the normalized W P_T over the W P_T. This clearly shows that W^- give more of their transverse momentum to the charged muons than the W^+. So W^+ are not only produced in larger numbers, but they also tend to give more transverse momentum to the neutral (hence missing energy) lepton, than their negatively charged counterpart. Furthermore, because of the fact that positively charged leptons tend to take only a small fraction of the W momentum, a transverse momentum cut (*e.g.* $P_T < 10$ GeV) on positively charged leptons also cuts out many hight P_T neutrinos. The high transverse momentum part of the neutrino distribution from W^- is less effected by acceptance cuts on the charged lepton, since typically the negatively charged (left-handed) lepton is anyhow the leading lepton.
In contrast to the W helicity in W+jets the helicity of Ws in $t\bar{t}$ events is well understood, pdf independent and already tested in data.
There are several ways to utilize the W helicity and charge asymmetry.

14.1 W helicity and charge asymmetry in RA4

A way of application of W helicity and charge asymmetry is to use the relation between charged lepton and neutrino spectra, which is solely determined by the W helicity and charge asymmetry. Then by measuring the lepton spectra the neutrino spectra (missing energy) can be estimated. Corrections between the two spectra must be taken from MC. Muon-neutrino replacement methods are commonly used **??** and have also been proposed in Section **??**.
The charge asymmetry can also be directly utilized. The W+jet helicity and charge asymmetric production, leads to the effect that much more neutrinos of high transverse momentum stem from W^+, then from W^-. This effect leads to a much smaller W+jet background contribution in the negative charge channel than in the positive charge channel. The search could be splitted into charge charge symmetric part, and a charge asymmetric part (positive charge - negative charge). To present the effect a simple pres-election as is the standard RA4 for muons was applied. However, the lepton isolation modified: The relative combined isolation is <0.1 for muons above 30 GeV transverse momentum and for muons between 10-30 GeV a e GeV tracker isolation of 5 GeV is required. The

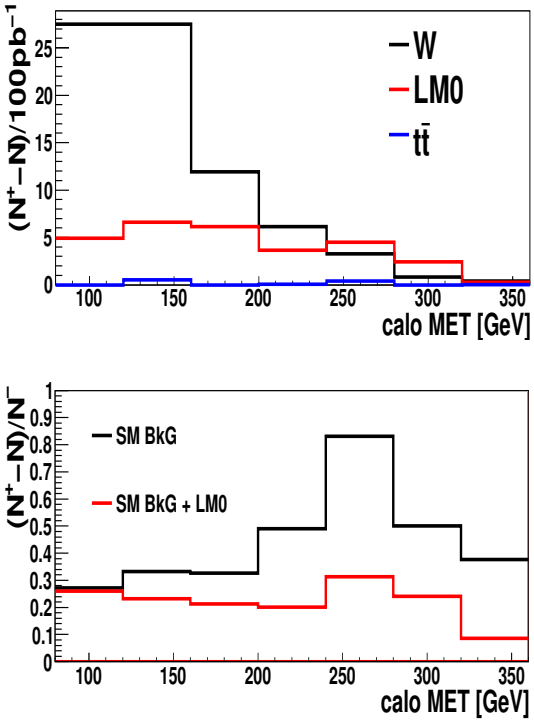


Figure 14.92: Calorimetric MET for different charges of leptons and datasets. The upper distribution shows the difference of the distribution of the different charges. The lower distribution is the ratio of charge asymmetric part and negative charge.

size of the different parts after the RA4 cut-flow are listed in table 14.28. It can be seen that in the case of LM0 SUSY, the excess is even slightly more significant in the symmetric part (all events - asymmetric part) on its own, compared to a search were both charges are combined.
The separation of charge has further interesting aspects:

- In the asymmetric part the W+jet events would be directly be compared to SUSY, since QCD and $t\bar{t}$ are symmetric in charge.
- The charge asymmetry is SUSY stems from gluino-squark production, where the squark comes from an valence quark. The squark from the valence quarks will be predominately u-squarks and in their decay chain produce positively charged leptons. If the lepton stems from the decay chain of the gluino, no charge preference is expected. So a charge division allows to learn something on the production mechanism of SUSY.
- In the symmetric part the W+jet background is largely reduced and the BSM signal can be better compared to predominately $t\bar{t}$ and QCD events.

The charge asymmetric part can be illustrated by subtracting the distribution from event with different charges. As example asymmetric part of the calorimetric MET is shown in Fig. 14.92. It can be seen that the asymmetric part is dominated by W+Jet events. The lower distribution in Fig. 14.92 represents the ratio of the charge asymmetric to the charge symmetric part. For SM only, the asymmetry part relatively increases w.r.t. the symmetric part with increasing MET. This is expected as explained in []. In the case of SUSY signal, for large values of calorimetric MET the ratio distribution in completely dominated by the ratio of the SUSY-benchmark point and the asymmetry is hence not increasing with MET.

Within the RA4 cut-flow, charge asymmetry tests are useful to test W+jets events directly against SUSY. If the W+jets background can be estimated independently, *e.g.* from Z+jets or γ+jets, then the charge separation is especially interesting. For the the negative charges the W contribution to the background is reduced by a factor of about $\sim$2.5 compared to the positive charges. So the negative charges are ideal to test a pure $t\bar{t}$ background

Table 14.28: Events per 100 pb^{-1} after the RA4 cut-flow with one muon (P$_T$ $>$ 10 GeV).

charges	W+jets	$t\bar{t}$	SUSY LM0	S/$\sqrt{B}$
all	176	356	498	21.6
positive - negative	78	-3	28	3.2
2$\times$negative	98	353	467	22.0

estimation against data. The asymmetry would further be enhanced in lower jet multiplicities or at lower proton-proton collision energies, where W+jet background becomes more important.

14.2 RA 4 variation based on W helicity

The RA 4 cut-flow could be varied in order to test if the polarization effects in data correspond to the SM prediction. Here a variable can be introduced which is sensitive to all W helicity degrees of freedom, which are two. One could take,*e.g.*, as free parameters the ratio of left-handed and right-handed helicity and the ratio of the longitudinal helicity to other helicity states. The relation between the charged lepton spectra and neutrino spectra is only sensitive to the first degree of freedom. A variable sensitive to both degrees of freedom is:

$$\vec{P_T} \cdot M\vec{H}T/MHT^2 \tag{35}$$

$\vec{P_T}$ is the charged lepton momentum and $M\vec{H}T \sim W$ P$_T$ the recoil of the jets. *E.g.* left-handed W^+ helicity will lead values close to 0 since most WP_T is given to the neutral particle. Longitudinal helicity leads to a peak at 0.5, while left-handed W^+ will lead values close to 1. The shape of this distribution depends predominately on the W helicity. The lower jet multiplicities can be used to verify quantitatively this effect in data in the kinematic region defined below:

- 110 GeV Jet trigger.
- $MHT(\sim WP_T) > 200$ GeV.
- Isolated global muon P$_T$ $>$ 15 GeV and $|\eta| < 2.1$.
- ΔR (Muon,jet) $>$ 0.5 (QCD rejection).
- less than three jets ($t\bar{t}$ rejection).

Fig. 14.93, shows the variable (Eq. 35) for negatively charged muon with more than 15 GeV transverse momentum. Fig. 14.93 shows the variable for positively charged muons. The dominating left-handed polarization is clearly visible in the distributions.

In order to utilize the polarization for a search, first the new physics expectation needs to be considered. In this note the chosen new physics is R-parity conserving (RPC) mSugra. The conservation of the lepton number for RPC leads to the fact that a single charged lepton is produced with an neutrino. Furthermore two LSP are produced which also add to the missing momentum. Hence the charged lepton direction is typically not very correlated to the missing momentum and its momentum amplitude typically smaller than the missing energy.
The above arguments lead to the expectation, that RPC mSugra events have typically values around zero or even smaller in the above defined variable. The one jet bin is typically RPC mSugra free, since the RPC leads to the production of at least two sparticles, so more than one jet can be required. However, this is supposed to be an example of the application and is by no means optimized towards low mass RPC mSugra benchmark points.
The same selection as in the previous sections was used, apart from the requirement that there need to be more than one jet. Fig 14.95 shows the sum of both charges for the variable (Eq. 35) and as expected the mSugra parameter points tend towards small values. Given that the distribution is dominated by the understood helicity of the W, its shape can taken from MC. For this study the shape was taken from generator level, with lepton acceptance cuts as in the reconstruction, but no efficiency corrections. The shape of this generator level distribution was used to estimate the cut efficiency for an cut at 0.3, which is in a region were the efficiency for $t\bar{t}$ and W+jets are similar. This efficiency was then used to predict the number of events below the cut value from the number of events above the value for the fully reconstructed events. Table 14.29 shows the predicted and the actually counted events in the signal region (below 0.3). The prediction does work well and excesses can be seen for benchmark points LM1 and LM0. Fig. 14.96 shows the difference between the charges of variable (35) and it can be seen that also in this phase space the asymmetric part is dominated by W+jet events.

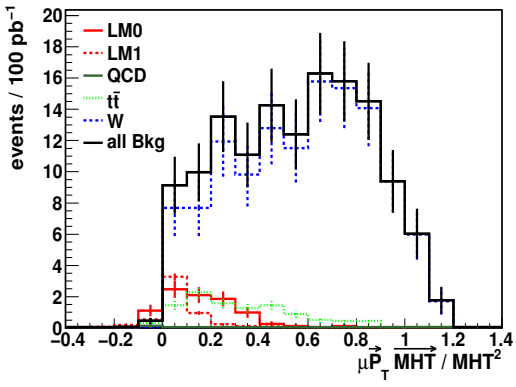


Figure 14.93: Lepton $\vec{P}_T$ projection on $\vec{M}HT$ over MHT^2 for negative lepton charge.

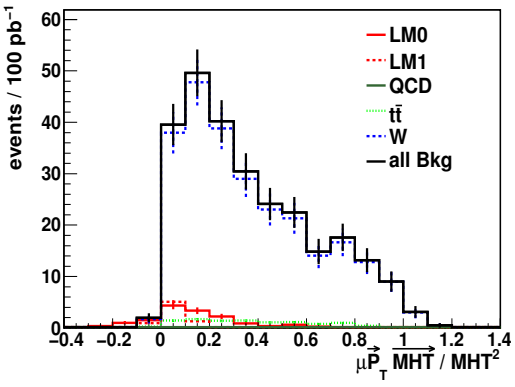


Figure 14.94: Lepton $\vec{P}_T$ projection on $\vec{M}HT$ over MHT^2 for positive lepton charge.

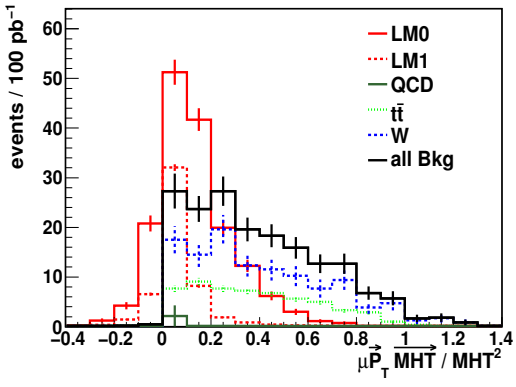


Figure 14.95: Lepton $\vec{P}_T$ projection on $\vec{M}HT$ over MHT^2 for both lepton charges combined.

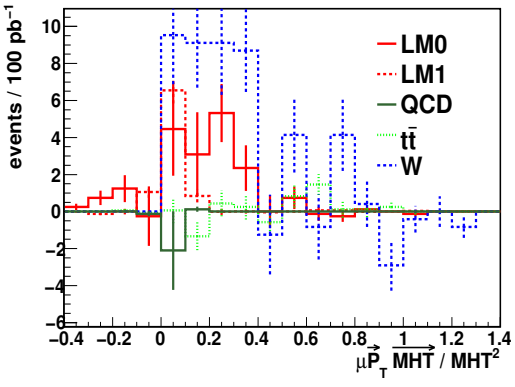


Figure 14.96: Difference of lepton $\vec{P}_T$ projection on $\vec{M}HT$ over MHT^2 for the lepton charges.

Table 14.29: Predicted and measured events for 100 pb^{-1} at 10 TeV.

datasets	predicted W+jets and $t\bar{t}$ events	measured events
SM only	70	79
LM0+SM	86	219
LM1+SM	70	130

14.3 Systematic uncertainties

From the theoretical [] side and by utilizing the one jet bin or 7 TeV data the W helicity in W+jets events can and need to be understood. The initial data at 7 TeV will presumably be SUSY signal free and allow to test the W helicity and charge asymmetry effects in W+jet events. The same will apply for example also the case in muon replacement methods, which also relies MC predictions on W helicity and charge asymmetry. For $t\bar{t}$ events, the W helicity and the charge symmetry are well known.
The overall lepton reconstruction efficiency is not required to be known, since only the polarization of the accepted events is tested. However, the change of the reconstruction efficiency as function of P_T and $|\eta|$ will be required to some extend. Typically Z events decaying to muons are used for such studies.
Also the resolution of MET does effect the proposed variable as well as the hight P_T muon resolution.

14.4 Early data commissioning plan

A plan to establish the W helicity in W+jets is layed out in [].

15 W background determination from Z control sample

From purely MC studies, we know that the background from the W decay channel is significant in a signal region with one electron, jets and large $\not{E}_T$. Hence, it is crucial to develop a data driven method to accurately estimate this contribution. Our goal is to predict how many W decay events we expect to see in our signal region for 100 pb^{-1} of data.

In Sec. 15.1, we state the various definitions pertinent to our analysis. Our background estimation technique is divided into three parts (Sections 15.2, 15.3, 15.4). Section 15.2 introduces the ABCD method, which we use to estimate the number of W decay events in the W control region. The control region contains a subset of the cuts that comprise our signal region. The purpose of the control region is to provide an accurate estimate of the number of W decay events that one could use to calculate the W cross section as a sanity check. Section 15.3 involves estimating the efficiency (for a pure W sample) of the jet cuts used in the definition of our signal region. This is the only step in our method that is MC dependent. We also discuss why we apply a MC-based efficiency requirement rather than applying the jet cuts prior to employing the ABCD method. Section 15.4 deals with the fact that the signal region has a $\not{E}_T$ cut of 150 GeV, whereas the W control region has a $\not{E}_T$ cut of 35 GeV. A data-driven method of projecting from the control region to the signal region is discussed. Section 15.5 contains the results for our method in the presence of all relevant SM backgrounds. Section 15.6 examines how our method behaves in the presence of all relevant SM backgrounds, as well as the SUSY LM1 point. Section 15.7 contains discussions about the sources of systematic error associated with the components discussed in Sections 15.2, 15.3, 15.4, and their estimated impact on the predictions. Section 15.8 looks at how we can utilize early data to prepare for our analysis.

15.1 Event selection

Before looking at a description of the background estimation method, it is useful to note certain definitions relevant to our analysis:

- Single electron trigger: HLT_Ele15_LW_L1R (15 GeV trigger with large pixel-matching windows and relaxed L1 conditions)
- Good electron:
  - $p_T > 20$ GeV
  - $0 < |\eta| < 1.47$ or $1.567 < |\eta| < 2.5$
  - RobustLooseElectronID (Egamma POG) passed
  - Fractional calorimeter isolation $\equiv$ (ECAL Iso + HCAL Iso) / Electron $p_T < 0.1$
- Jet:
  - L2L3 corrected RECO sisCone5 Jet
  - $|\eta| < 3$
  - Electromagnetic fraction (EMF) < 0.9
  - $\Delta R > 0.3$ w.r.t. all good electrons
- BSM signal region:
  - Single electron trigger fired,
  - exactly one good electron,
  - at least four jets with $p_T > 50$ GeV,
  - $\not{E}_T >$ 150 GeV
- W control region:
  - Single electron trigger fired,
  - exactly one good electron,
  - $\not{E}_T >$ 35 GeV

15.2 W control region and the ABCD method

The W control region is a region in phase space where W decay events dominate over all other types of events. For our control region, QCD events are the biggest source of contamination. The ABCD method is a suitable choice for estimating the QCD background in the control region, because as explained in the next paragraph, the requirements that need to be satisfied for it to be applicable are met. Before applying the ABCD method, we impose all the event selection cuts that form our control region, with the exception of two, which are the two variables used by the ABCD method. These are:

1. the cut on fractional calorimeter isolation for the electron, and
2. the cut on uncorrected $\not\!\!E_T$.

The reason for choosing uncorrected (Calo) $\not\!\!E_T$ instead of jet and muon corrected $\not\!\!E_T$ is discussed in Sec. 15.2.1. These two variables are chosen for their efficacy in distinguishing between signal and background. We make a 2D plot with these two variables as the two axes. On the 2D plot, we define regions A, B, C and D. It must be the case that in regions A, B and D, the QCD background dominates the W signal. For region C (the control region), we additionally require that W dominates over QCD, because as previously explained, we want a precise estimate of the number of W events for a cross section calculation. The region definitions used by us are shown in Table

	A	B	C	D
Lower MET Bound	0	0	35	35
Upper MET Bound	20	20	–	–
Lower FCI Bound	0.2	0	0	0.2
Upper FCI Bound	–	0.1	0.1	–

Table 15.30: Definitions for regions A, B, C and D. FCI stands for fractional calorimeter isolation.

15.30.

The idea behind the ABCD method is that for zero correlation i.e. a featureless 2D distribution, N_A/N_B should equal N_D/N_C. So we also require that the correlation between the two variables is close to 0. The number of background events in region C is given by:

$$N_C = N_D \times N_B/N_A$$

Fig. 15.97 shows the 2D plot for the QCD sample. As we can see, the correlation is indeed small. In Table 15.31, we see the number of events for QCD, Z + jets, $t\bar{t}$ and W + jets samples for regions A, B, C and D.

	A	B	C	D
W	$(2.78 \pm 0.03) \times 10^3$	$(4.08 \pm 0.01) \times 10^4$	$(2.40 \pm 0.01) \times 10^5$	$(8.49 \pm 0.01) \times 10^3$
QCD	$(3.17 \pm 0.36) \times 10^7$	$(1.80 \pm 0.14) \times 10^6$	$(2.72 \pm 0.56) \times 10^4$	$(7.80 \pm 0.26) \times 10^5$
Z	$(1.51 \pm 0.02) \times 10^3$	$(3.24 \pm 0.01) \times 10^4$	$(1.04 \pm 0.02) \times 10^3$	$(2.54 \pm 0.09) \times 10^2$
$t\bar{t}$	$(7.15 \pm 0.05) \times 10^2$	$(4.12 \pm 0.04) \times 10^2$	$(2.34 \pm 0.01) \times 10^3$	$(1.22 \pm 0.01) \times 10^3$
Total	$(3.17 \pm 0.36) \times 10^7$	$(1.88 \pm 0.14) \times 10^6$	$(2.70 \pm 0.06) \times 10^5$	$(7.90 \pm 0.26) \times 10^5$

Table 15.31: Counting events in each ABCD region from the major background samples.

These values show that QCD dominates in regions A, B, and D, and W is substantial in region C. Thus, Table 15.31, in addition to Fig. 15.97, indicates that the ABCD method is a suitable choice for our purposes.

15.2.1 Problems with corrected $\not\!\!E_T$

For this discussion, corrected $\not\!\!E_T$ means that the $\not\!\!E_T$ takes into account the corrected jet and muon energies, and uncorrected $\not\!\!E_T$ refers to the calotower deposit based quantity. In Fig. 98(a), we evaluate the performance of corrected and uncorrected $\not\!\!E_T$ for a W sample by comparing both to generated $\not\!\!E_T$, and in Fig.98(b), we do the same for a QCD sample. As we can see, for events with true $\not\!\!E_T$ (like in a W sample), corrected $\not\!\!E_T$ performs better (the mean of the distribution corresponding to corrected $\not\!\!E_T$ is -0.02, as opposed to -1.36 for the uncorrected $\not\!\!E_T$), but the reverse holds for events with no true $\not\!\!E_T$ (like in a QCD event). There is a bump in the QCD plot around 20 GeV for corrected $\not\!\!E_T$, which artificially inflates the number of QCD events with high $\not\!\!E_T$. Our hypothesis for this bump is related to jet corrections applied to $\not\!\!E_T$. For a jet to contribute to $\not\!\!E_T$ corrections, it must have a uncorrected p_T above 20 GeV and EMF < 0.9. So, for a QCD event with one or more jets, it is possible for one jet to be below this threshold, which results in an asymmetric correction that produces the observed bump.

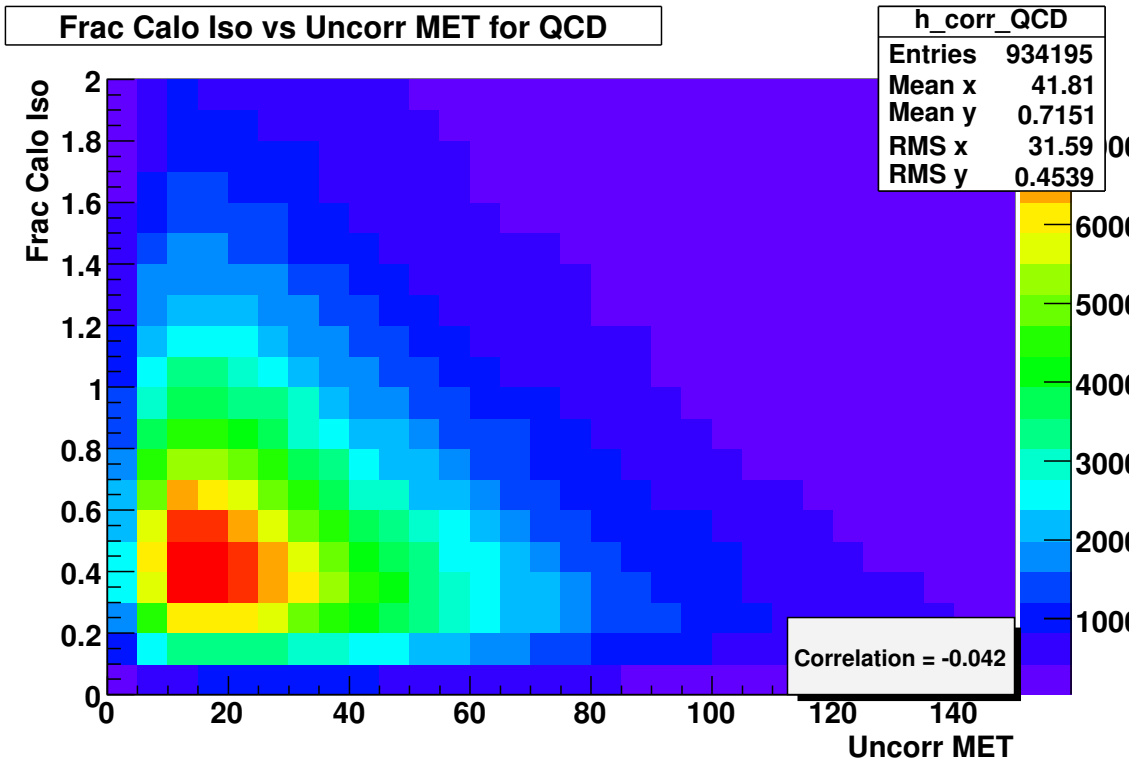
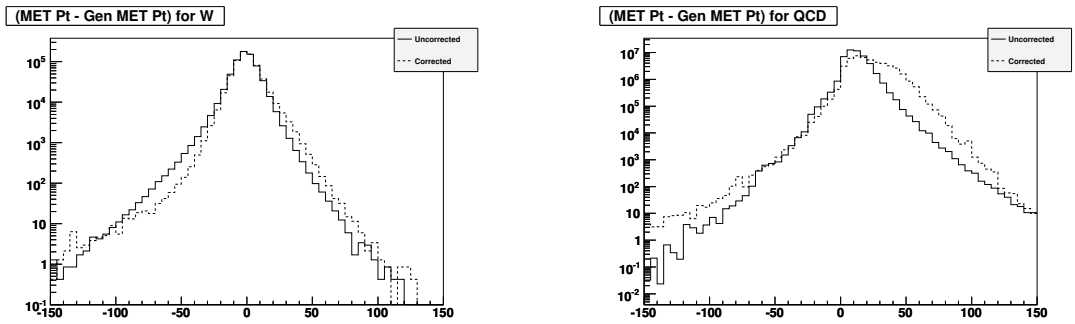


Figure 15.97: 2D ABCD plot for QCD sample; note the low correlation.



(a) W + jets

(b) QCD

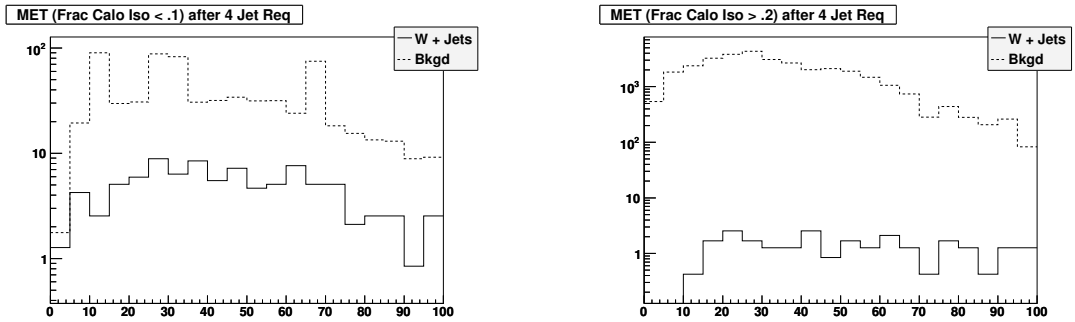
Figure 15.98: Comparing the performances of corrected and uncorrected $\not{E}_T$.

15.3 MC-based efficiency of the 4 jet requirement for W events in the control region

If we require that an event in the W control region have four or more jets, W events are suppressed relative to QCD events. As a result, the ABCD method cannot provide a precise W estimate. In Fig. 99(a), we see the $\not{E}_T$ distribution for events in the well isolated regions (B and C) after requiring at least 4 jets above 50 GeV, and in Fig. 99(b), we see the same thing for poorly isolated regions (A and D). Fig. 99(b) tells us that in regions A and D, the background dominates the signal as required by the ABCD method. But Fig. 99(a) tells us that it is impossible to select a suitable region C, where W dominates over the background. It is this inability to select a suitable control region that prevents us from applying the jet cuts prior to the ABCD method.

To circumvent this problem, we employ a MC-based estimate of the efficiency of the requirement that a W event present in the control region has at least 4 jets above 50 GeV. The reason for obtaining this efficiency for only W decay events rather than all events in the control region is that we expect contamination from QCD events in the control region, and the efficiency for QCD events would be higher, which would lead to an overestimate. The result is quoted below.

$$\text{Jet requirement efficiency} = (3.56 \pm 0.25) \times 10^{-4} \tag{36}$$



(a) Well Isolated Region

(b) Poorly Isolated Region

Figure 15.99: Understanding why jet cuts cannot be applied prior to ABCD method.

15.4 Using a pure Z sample from data to derive the $\not\!\!E_T$ shape for W

We begin by extracting a pure Z sample from data by making a cut on the reconstructed Z mass (81 GeV $< Z$ mass $<$ 101 GeV). To do this, we require at least one good electron, and then choose a second electron (which has a relaxed p_T cut of 5 GeV, and a relaxed isolation cut of 0.2) which gives the best possible reconstructed Z mass. Using this pure $Z \rightarrow ee$ sample, we obtain a $\not\!\!E_T$ template by dropping one of the two electrons (which acts as an analog to the neutrino in the W event) and recalculating the $\not\!\!E_T$, while correcting for the kinematic differences between a Z and W event. For each Z event, there are two values of the $\not\!\!E_T$ template, one for each of the electrons. This $\not\!\!E_T$ template shape is then rescaled based on the estimated number of W events in the control region after applying the jet requirement efficiency. Using this $\not\!\!E_T$ distribution, we count all events with $\not\!\!E_T > 150$ GeV to predict the number of events in the signal region.

Ideally, we would like to apply the condition that each event in our pure $Z \rightarrow ee$ sample has four or more jets above 50 GeV. However, due to the statistics associated with 100 pb^{-1} of data, this condition does not leave us with a sufficient number of events. So, in order to compare the W $\not\!\!E_T$ shape to the $\not\!\!E_T$ template shape obtained from Z, we need to relax the jet requirements. We choose to have at least one jet above 50 GeV, and at least three additional jets above 20 GeV. However, it is expected that the W $\not\!\!E_T$ shape when requiring four or more jets above 50 GeV will not be the same as the W $\not\!\!E_T$ shape when requiring at least one jet above 50 GeV, and at least three additional jets above 20 GeV. To correct for this, we intend to use a very high statistics W sample to calculate these two shapes, and then apply a MC-based correction that is obtained from this study to the Z MET template. In the interim, we apply a simpler correction factor as explained below.

For the W sample:

- Number of entries with $\not\!\!E_T > 35$ GeV & 4 jets above 50 GeV = $N_{W,35,1}$ = 224
- Number of entries with $\not\!\!E_T > 35$ GeV & 1 jet above 50 GeV & 3 more jets above 20 GeV = $N_{W,35,2}$ = 3063
- Number of entries with $\not\!\!E_T > 150$ GeV & 4 jets above 50 GeV = $N_{W,150,1}$ = 11
- Number of entries with $\not\!\!E_T > 150$ GeV & 1 jet above 50 GeV & 3 more jets above 20 GeV = $N_{W,150,2}$ = 48

$$\text{Correction factor} = \frac{N_{W,150,1} \times N_{W,35,2}}{N_{W,150,2} \times N_{W,35,1}} = 3.13 \pm 1.07 \tag{37}$$

The predicted number of W events must be scaled by this correction factor to get the final estimate.

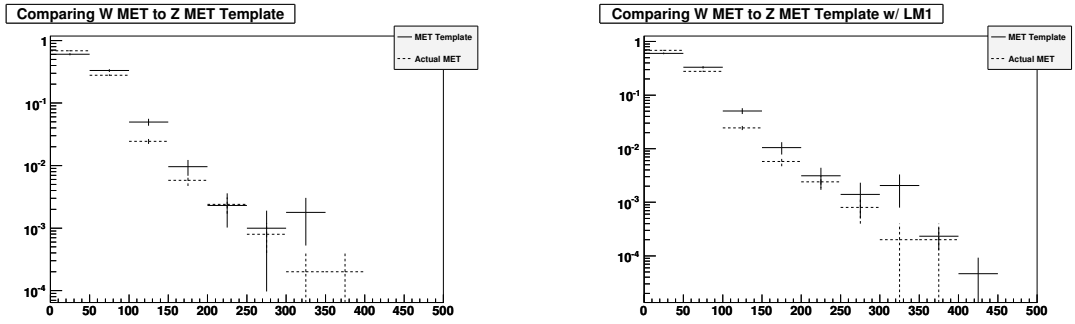
15.5 Performance of method without signal

In Table 15.32, we apply the ABCD on our SM MC samples (Z, W, $t\bar{t}$, QCD) to check how well it performs.

The next step in our method is to multiply the predicted number of W events by the efficiency of our jet requirement (at least 4 jets above 50 GeV) for W events in our control region (cf. Eqn. 36). The final step is to get the ratio of the number of events with $\not\!\!E_T > 150$ GeV to the number of events with $\not\!\!E_T > 35$ GeV. Fig. 100(a) compares the

	Predicted	Actual
QCD	$(4.68 \pm 0.38) \times 10^4$	$(2.72 \pm 0.56) \times 10^4$
W	$(2.24 \pm 0.07) \times 10^5$	$(2.40 \pm 0.01) \times 10^5$

Table 15.32: Testing performance of ABCD method in predicting number of events in region C using MC samples. Quoted errors are statistical only.



(a) SM Only

(b) LM1 Included

Figure 15.100: Comparing the W $\not\!\!E_T$ shape to the Z $\not\!\!E_T$ template shape.

W $\not\!\!E_T$ shape to the Z $\not\!\!E_T$ template shape after requiring at least one jet above 50 GeV, and at least three additional jets above 20 GeV. The agreement assures us that our approach is justifiable.

The numbers associated with the $\not\!\!E_T$ template are:

- Number of events above 35 GeV = 221 $\pm$ 8
- Number of events above 150 GeV = 4.81 $\pm$ 1.11
- Ratio = 0.0218 $\pm$ 0.0051
- Correction factor (cf. Eqn. 37) = 3.13 $\pm$ 1.07

Thus, the final results are:

- Actual number of W events in signal region = 4.7 $\pm$ 1.4
- Predicted number of W events in signal region = 5.5 $\pm$ 2.3

15.6 Performance of method with signal

Here, in addition to the SM MC samples, we also include the SUSY LM1 point to check how our method is affected by its presence. The first step of our process, the ABCD method, yields a W prediction that is indistinguishable from the one seen in Sec. 15.5. Since the jet requirement efficiency is obtained by running on just the W sample, it is also identical to the result quoted in Sec. 15.5. The only point of divergence is when we obtain the Z $\not\!\!E_T$ template. Fig. 100(b) is analogous to Fig. 100(a), except that the Z $\not\!\!E_T$ template also includes LM1. There is still good agreement between the W and Z shapes.

The numbers associated with the $\not\!\!E_T$ template are:

- Number of events above 35 GeV = 223 $\pm$ 8
- Number of events above 150 GeV = 5.75 $\pm$ 1.11
- Ratio = 0.0258 $\pm$ 0.0051
- Correction factor (cf. Eqn. 37) = 3.13 $\pm$ 1.07

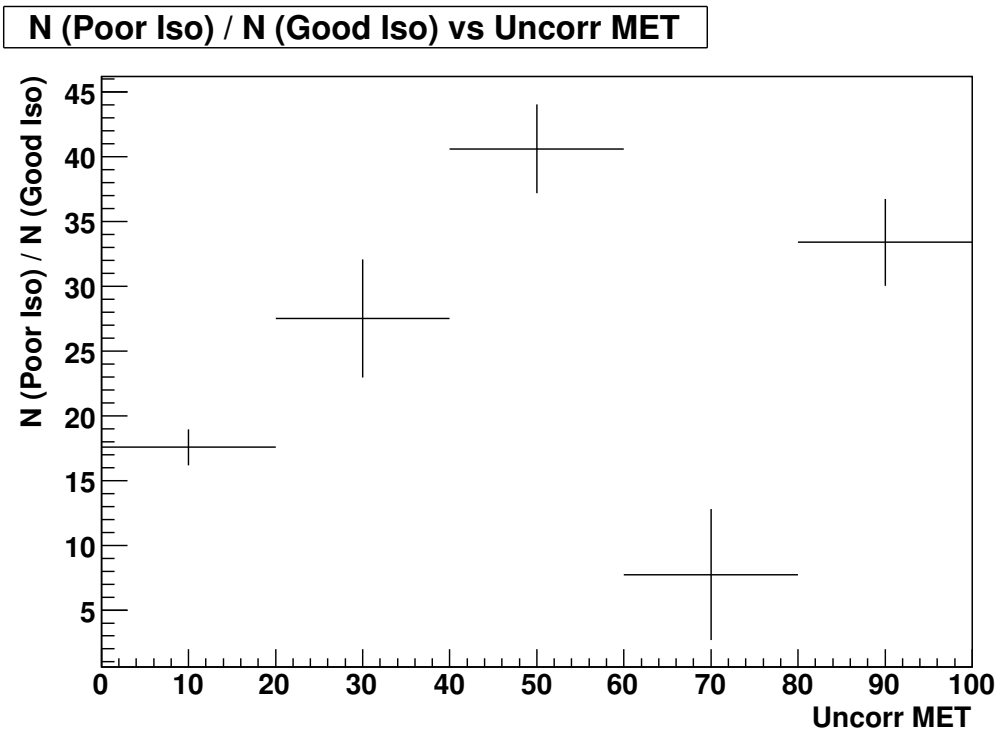


Figure 15.101: Effect of non-zero correlation on ABCD method.

Thus, the final results are:

- Actual number of W events in signal region = 4.7 $\pm$ 1.4
- Predicted number of W events in signal region = 6.5 $\pm$ 2.6

15.7 Systematic uncertainties

Here we explore the systematic uncertainties associated with the three steps that comprise our estimate of W events in the BSM signal region.

15.7.1 Systematic error in ABCD method for non-zero correlation

The ABCD method assumes zero correlation between the two variables. For us, the correlation is small, but it is non-zero. Hence we need to understand how the estimate is dependent on the amount of correlation.

In Fig. 15.101, we see the ratio of the number of events in the poorly isolated region (fractional CaloIso > 0.2) to the number of events in the well isolated region (fractional CaloIso < 0.1) as a function of $\not\!\!E_T$ for QCD events involved in the ABCD method. If the correlation were zero, this would be flat. We see that the value of this ratio can vary by as much as a factor of 2 compared to the value in the $\not\!\!E_T < 20$ bin, which is the high statistics region we use for normalization. In other words, the true value of the number of background events in region C can vary by a factor of 2 from the predicted value using the normalization factor from the $\not\!\!E_T < 20$ region. Hence, a conservative estimate of the error on the predicted number of background events is the predicted number itself.

15.7.2 Robustness of the ABCD method

In early data, it is expected that the $\not\!\!E_T$ resolution will not be optimal, so we need to ensure that our method is not sensitive to small amounts of misalignment and miscalibration. There are two $\not\!\!E_T$ boundaries we employ: one at 20 GeV, and another at 35 GeV. In Table 15.33, we see the effect of sliding each of these boundaries to the right or left by 5 GeV. This is an arbitrarily chosen number meant to reflect the uncertainty in $\not\!\!E_T$ for early data. In all cases, the estimates are consistent to the actual value within error bars, indicating that the method is indeed robust.

	Predicted	Actual
A, B: MET < 20, C, D: MET > 30	$(3.21 \pm 0.11 \pm 0.94) \times 10^5$	$(3.69 \pm 0.01) \times 10^5$
A, B: MET < 20, C, D: MET > 40	$(9.87 \pm 0.35 \pm 2.48) \times 10^4$	$(1.07 \pm 0.01) \times 10^5$
A, B: MET < 15, C, D: MET > 35	$(2.30 \pm 0.07 \pm 0.40) \times 10^5$	$(2.40 \pm 0.01) \times 10^5$
A, B: MET < 25, C, D: MET > 35	$(2.24 \pm 0.07 \pm 0.46) \times 10^5$	$(2.40 \pm 0.01) \times 10^5$

Table 15.33: Comparing number of W events to check robustness of ABCD method. For the predicted values, the first error is statistical, and the second, systematic.

15.7.3 Systematic error on jet requirement efficiency (incomplete)

Current idea is to use two different MC generators (Madgraph & Pythia), and see how the efficiency is affected as a result.

15.7.4 Systematic error estimates for using $\not\!\!E_T$ shape from Z sample (incomplete)

Since we use a MC-based correction factor to account for the fact that W $\not\!\!E_T$ shape when requiring four or more jets above 50 GeV will not be the same as the W $\not\!\!E_T$ shape when requiring at least one jet above 50 GeV, and at least three additional jets above 20 GeV, we also need to establish the generator dependence of this.

15.8 Early data commissioning plan

16 Signal Suppression in Background Control Regions

The presence of signal contamination in the control region poses a great problem in the classical ABCD method. This holds particularly true for our "C" region, which is characterized by a large $\not{E}_T$ and a small m_T value (Fig. 16.102). No additional correlations are introduced if one can remove the signal contamination in the "A" and "C" regions synchronously. This section will present such a signal suppression cut.

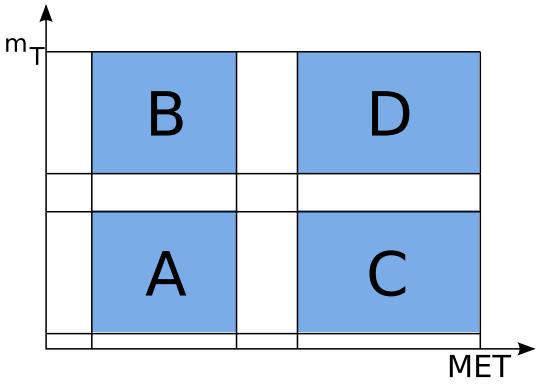


Figure 16.102: This section's definition of the ABCD regions.

16.1 Description of Method

As explained in great detail in Sec. 14, the ratio $\frac{p_T^{\text{muon}}}{MHT}$ is a suitable discriminator between signal and background. Fig. 16.104 visualises the discriminative power of this variable. In order to increase the flexibility of the discrimination, we decided to discriminate with an affine rather than a linear function in the plane that is defined by the p_T^{muon} and the MHT variables.

Our signal suppression cut reads:

$$HT > x \cdot p_T^{\text{muon}} + y \tag{38}$$

We substituted MHT for HT since our experimental findings show that HT has higher discrimination power. Since this fact is in stark disagreement with the theoretical picture, further investigation is needed.

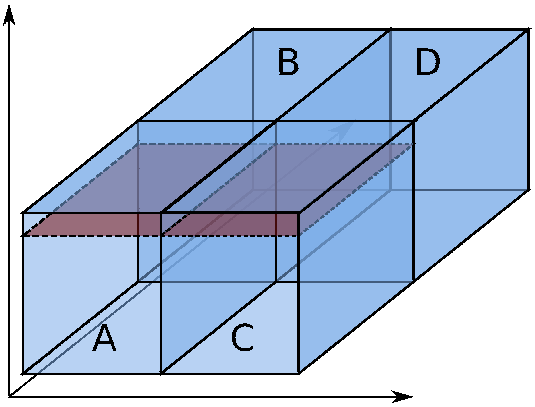


Figure 16.103: Adding additional cuts in A and C does not introduce correlations into the ABCD method.

16.2 Technical setting

For a detailed description of the signals see Sec. 2, for a discussion on the details of the backgrounds see Sec. 4. In this chapter we focus on the low mass (LM) points 1, 2, and 8. The considered background consists of W+jets, $t\bar{t}$, and InclusiveMu5Pt50.

The loosest of the standard RA4 preselections have been applied: $p_T^{Muon} > 10$ GeV (`vJetMuons`), 3 jets $>$ (75, 50, 50) GeV (`allLayer1Jets`). Additionally, a met isolation cut $\Delta\phi(\not{E}_T, j_i) > 0.3$ was employed. Our four

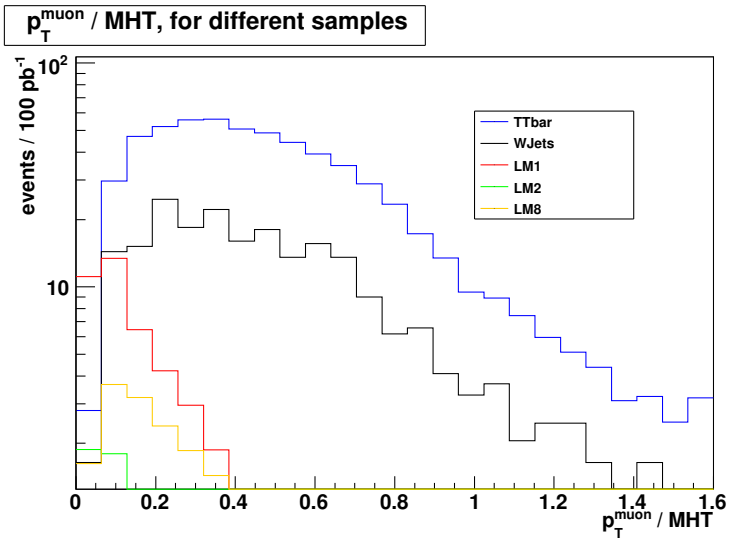


Figure 16.104: p_T^μ / MHT, for various samples. Our preselection (Sec. 16.2) has been applied.

ABCD regions are such that $\not{E}_T > 50$ GeV and $m_T > 50$ holds true for all of them. These lower cuts are applied prior to performing the ABCD method and are in this section considered part of the preselection.

Determination of a suitable working point

The cut condition (Eq. 38) has two free parameters, denoted by x and y.

The choice of x and y must meet the following requirements:

(a) the signal / background ratio should be sufficiently low,

(b) the correlation factor ϵ should be close to zero,

(c) (a) and (b) must be robust, i.e. resilient against small changes in the choice of x and y.

Fig. 16.106 depicts the x and y dependence of the signal over background ratio and of ϵ.

$(x, y) = (10, 400)$ has been chosen as the default working point; it seems to the authors to be a fair compromise between the requirements (a), (b), and (c). Fig. 16.105 shows the signal suppression cut in the p_T^{muon} versus HT plane; the top left triangle will be cut off.

16.3 Performance of method without signal

In order to compare performances, the method of *bootstrapping statistics* [**?**] has been performed. In this approach, the numbers of background and signal events in the regions A,B,C,D are taken from a large Monte Carlo production. These numbers are in the subsequent analysis treated as if they were the "true" Poisson parameters of the fundamental distribution. That way an infinite amount of pseudo-experiments with a certain integrated luminosity can be simulated by drawing from the four Poisson distributions. With this "meta-statistical" approach, the statistical performance of a method can be studied in great detail.

The performance of the method without in the absence of a signal has been compared against the "classical" ABCD method without the signal suppressing cut. Since the method explicitly targets signal suppression, the performance of the method must degrade slightly. More specifically, the statistical error increases by a small amount – as can be seen in Fig. 16.107.

16.4 Performance of method with signal

In the presence of signal contamination the method is expected to reduce the systematic bias at the expense of slightly increased statistical errors. Fig. 16.109 shows the performance of the method for LM1, LM2, and LM8, as

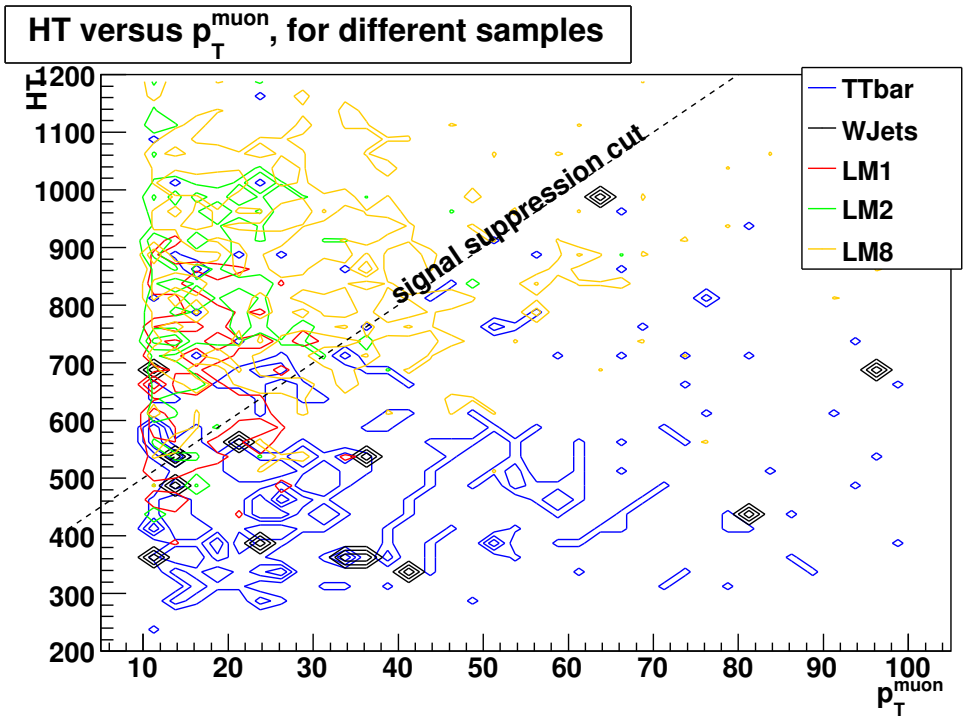


Figure 16.105: The p_T^{muon} versus HT plane.

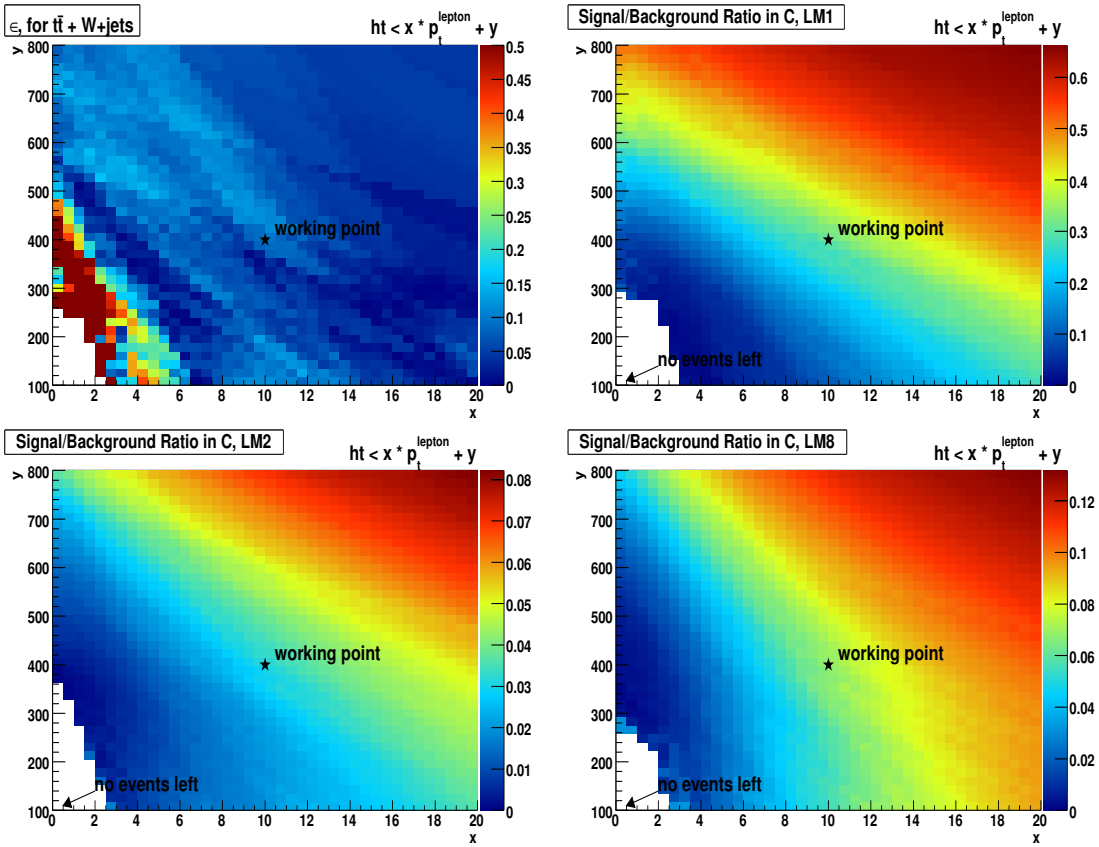


Figure 16.106: ϵ as a function of x and y (top left). Signal over background ratios r of $t\bar{t}$ in the C regions, for various signals (all other three plots).

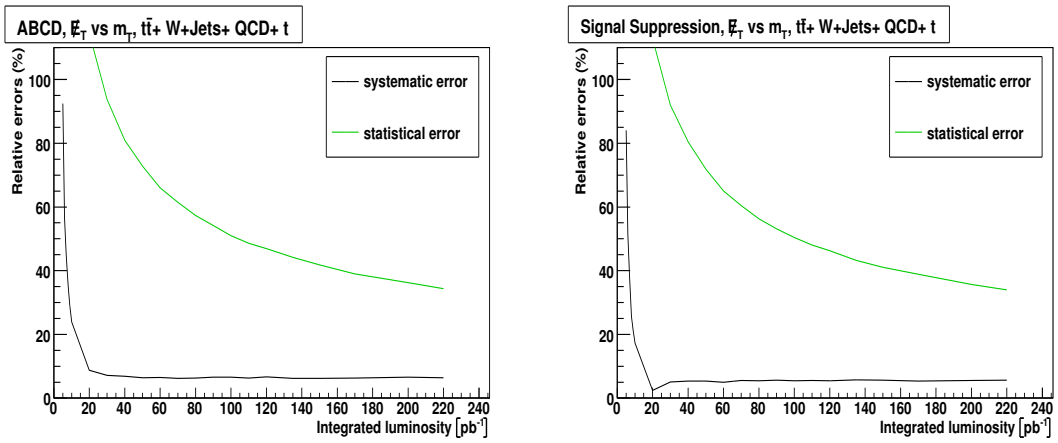


Figure 16.107: Relative errors, as a function of integrated luminosities

signal	signal events	bkg events	estimated bkg
no signal	0	5.06	$4.84 \pm 1.22^{\mathrm{stat}}$
LM1	17.85		$6.51 \pm 1.45^{\mathrm{stat}}$
LM2	4.43		$5.01 \pm 1.24^{\mathrm{stat}}$
LM8	9.80		$5.28 \pm 1.28^{\mathrm{stat}}$

Figure 16.108: Results of the signal suppressing method, for different mSUGRA points.

a function of the integrated luminosities. The cut works as expected. Fig. 16.110 translates the asset of the method into gains in signal significances, which we define as

$$S = \left\langle \frac{n_D^{est} + sys}{\sigma^{est}} \right\rangle \tag{39}$$

Here, n_D^{est} equals the estimated number of signal events in D, σ^{est} is its error estimate. sys refers to the systematic bias from the ABCD method. The brackets $\langle \dots \rangle$ denote averaging over N pseudo experiments. The signal suppressing method constantly scores better than the ABCD method. Tab. 16.108 summarizes the results.

16.5 Systematic uncertainties

Correlations between the two discriminating variables and non-suppressed signals in the control regions induce systematic uncertainties that have been subsumed in the "systematic biases" discussed in the previous subsection. We now wish to quantify the changes in the systematic biases in terms of an unknown jet energy scale, or, similarly, in small deviations in the definitions of the ABCD regions.

16.5.1 Systematic bias as a function of jet energy scale

The following table shows how a mis-measured jet energy scale affects the estimate, for all backgrounds plus LM1. It can be seen that the event counts differ quite dramatically with the jet energy scale. The quality of the estimate, though, is robust against mis-measured jet energy scales.

jet energy scale	true signal in D	true bkg in D	estimated bkg in D
1.2	25.21	19.0	$21.47 \pm 2.8^{\mathrm{stat}}$
1.1	21.76	10.97	$12.83 \pm 2.09^{\mathrm{stat}}$
1.0	17.85	5.06	$6.51 \pm 1.45^{\mathrm{stat}}$
0.9	14.05	2.67	$4.22 \pm 1.49^{\mathrm{stat}}$
0.8	10.28	1.74	$2.91 \pm 1.12^{\mathrm{stat}}$

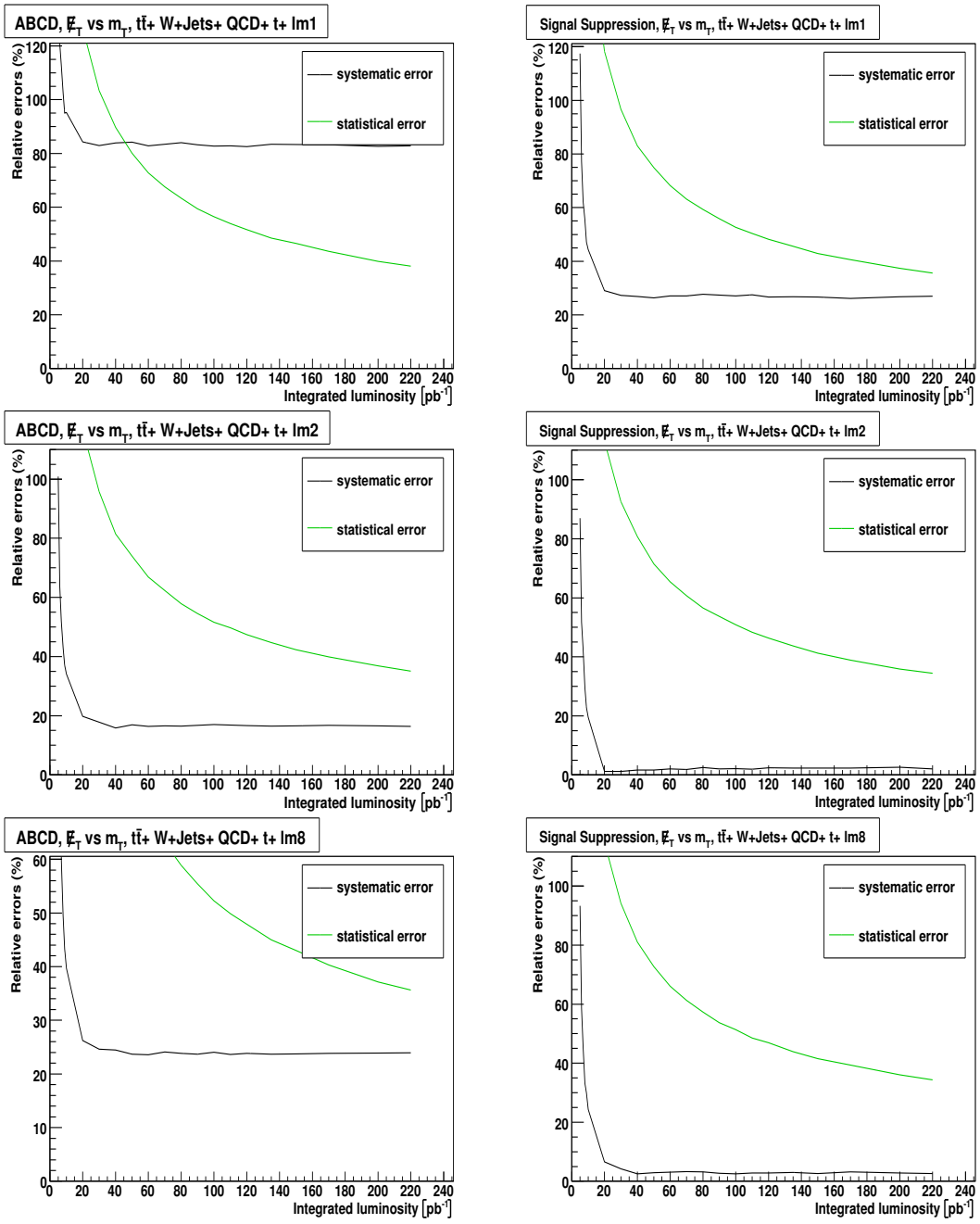


Figure 16.109: Relative errors, as a function of integrated luminosities, for different signals. ABCD method versus signal suppression method.

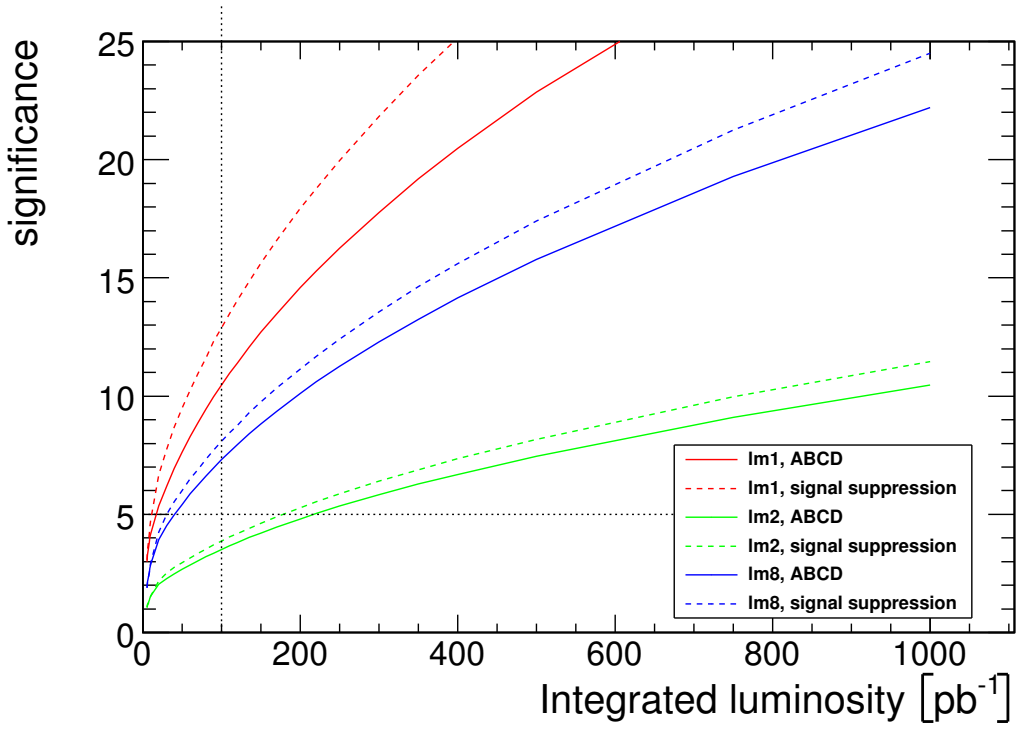


Figure 16.110: Signal significances, as a function of integrated luminosities, with the ABCD and the signal suppression methods, for various LM points.

16.5.2 Systematics due to small deviations in the ABCD regions

Similar to Sec. 16.5.1, the influence of deviations in the choice of the ABCD regions has been studied. The following table summarizes the results for all backgrounds + LM1. Additionally, Fig. 16.111 shows how the true and the estimated numbers of events change with the definition of the ABCD regions. We conclude that the method is very insensitive with respect to the $\not\!\!E_T$variable; the choice of the m_T cut has a greater role. This fact is easily understood: $m_T \approx 80$GeV corresponds to the kinematic endpoint of the W mass.

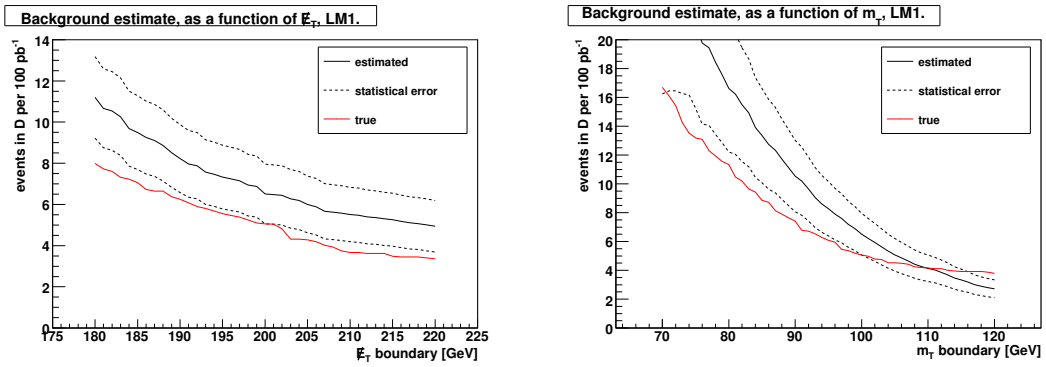


Figure 16.111: Estimated versus true number of events in D, as a function of the region definition.

contribution	percentage
single μ	87 %
di-μ	1 %
$\mu + \tau$	5 %
$\mu + e$	2 %
other	5 %

Table 16.34: Anatomy of $t\bar{t}$, in the control regions

contribution	percentage
single μ	26 %
di-μ	16 %
$\mu + \tau$	40 %
$\mu + e$	16 %
other	2 %

Table 16.35: Anatomy of $t\bar{t}$, in the signal region

region boundaries	true bkg in D	estimated bkg in D
$\not\!\!E_T > 200, m_T > 100$	5.06	$6.51 \pm 1.45^{\mathrm{stat}}$
$\not\!\!E_T > 210, m_T > 100$	3.67	$5.51 \pm 1.32^{\mathrm{stat}}$
$\not\!\!E_T > 205, m_T > 100$	4.28	$6.00 \pm 1.38^{\mathrm{stat}}$
$\not\!\!E_T > 195, m_T > 100$	5.55	$7.33 \pm 1.55^{\mathrm{stat}}$
$\not\!\!E_T > 190, m_T > 100$	6.25	$8.23 \pm 1.65^{\mathrm{stat}}$
$\not\!\!E_T > 200, m_T > 110$	4.16	$4.14 \pm 0.92^{\mathrm{stat}}$
$\not\!\!E_T > 200, m_T > 105$	4.53	$5.08 \pm 1.13^{\mathrm{stat}}$
$\not\!\!E_T > 200, m_T > 95$	7.42	$10.54 \pm 2.48^{\mathrm{stat}}$
$\not\!\!E_T > 200, m_T > 90$	6.08	$8.28 \pm 1.88^{\mathrm{stat}}$

16.6 On di-leptonic contributions

The signal suppressing method treats the backgrounds *inclusively*, i.e. it does not discriminate between the various physics channels that constitute the background in the signal region. The fact, that the physics channels in the control regions are very different from the ones in signal region, is also ignored. In order to overcome this obstacle, the various $t\bar{t}$ contributions to the control regions and the signal regions have been dissected similar to what has been done in Sec. 11.

It is the final aim of this subsection to estimate the di-leptonic contributions to the single-leptonic estimate. These contributions will be called the "estimees", or, if referred to as the sum of the individual contributions, the "estimee". The "estimees" will be estimated from the di-leptonic background, where no lepton is lost, i.e. where two leptons appear at the reconstruction level. These estimated shapes of $\not\!\!E_T$ and m_T will be called "estimates".

Tab. 16.35 shows the different compositions of the signal and control regions. In order to fully control the various contributions, it seems evident that they be estimated individually, rather than inclusively. To this end, the contributions are estimated from the di-muon $t\bar{t}$ region, "loosing" one of the muons to model the non-reconstructed lepton in the single-lepton channel. For each case, the scale of the model is taken individually from Monte Carlo. Also, depending on the case, a certain fraction of the lost muon's energy is put into the $\not\!\!E_T$. The fractions of energy to be put into the $\not\!\!E_T$ have been determined by comparing the $\not\!\!E_T$ and m_T shapes of the "estimees" with the "estimates", and are summarized in Tab. 16.36.

Additionally, the di-muon case is split up into two subcases. The first corresponds to a muon that has never been reconstructed. The second case includes two-fully reconstructed muons one of which was considered for the final muon selection. The two subcases will be referred to as "lost muon" and "not selected muon".

Fig. 16.112 shows the sum of all di-leptonic contributions and the total estimate. An integrated luminosity of 100 pb^{-1} amounts a total of expected 4.1 $t\bar{t}$ events, 3.7 of which are from di-leptonic contributions. Our procedure estimates 3.8. For discussions on the errors and the systematics of such a method, see Sec. 14. This section intends to focus on the remaining estimate: the signal suppressing method can now be applied to the remaining

contribution	fraction
μ, lost μ	50 %
μ, not selected μ	0 %
$\mu + \tau$	40 %
$\mu + e$	30 %

Table 16.36: Fraction of "lost" muon energy to be put into the $\not\!\!E_T$.

signal	signal events	total bkg	semi-lept bkg	estimated semi-lep
no signal	0	5.06	2.22	$3.99 \pm 1.04^{\mathrm{stat}}$
LM1	17.85			$5.49 \pm 1.25^{\mathrm{stat}}$
LM2	4.43			$4.15 \pm 1.06^{\mathrm{stat}}$
LM8	9.80			$4.39 \pm 1.11^{\mathrm{stat}}$

Table 16.37: Results of the signal suppression method after subtraction of di-leptonic contributions, for 100 pb^{-1}.

single-leptonic $t\bar{t}$ and W+jets background. Tab. 16.37 shows the performance of the estimates after the estimated di-leptonic contributions have been subtracted.

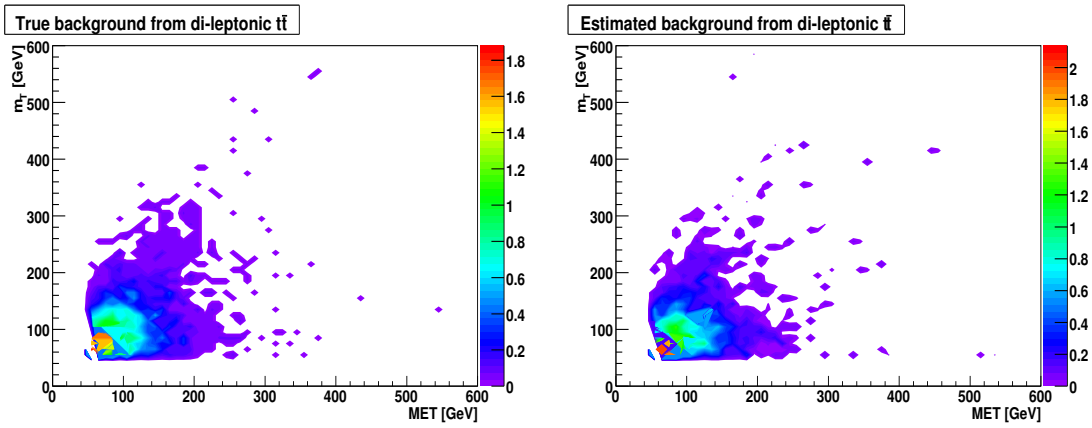


Figure 16.112: True(left) and estimated(right) di-leptonic contributions to the single-lepton case

The results of the estimate after subtraction of the di-leptonic contributions is consistent with the results of the all-inclusive estimate, although it must be stated that the errors of the di-leptonic estimates have not yet been taken into account.

16.7 Early data commissioning plan

It is our aim to try to reproduce the cut flow plot (Fig. 16.113) with data (integrating over the different samples). Also, we aim to validate the Monte Carlo QCD prediction via a simple ABCD method, using $\not\!\!E_T$ and the lepton isolation as the ABCD variables. Having done that, it will be interesting to see what fraction of events is affected by the signal suppression cut, which can be compared with Monte Carlo. Finally, the number of events in the ABCD regions themselves must be compared with our Monte Carlo simulations.

17 Eight Fields Method

17.1 Description of Method

In order for the ABCD method to work, it is crucial that the two ABCD variables be independent from each other. Only then does the defining equation of ABCD hold true:

$$\frac{\lambda_A}{\lambda_B} = \frac{\lambda_C}{\lambda_D} \tag{40}$$

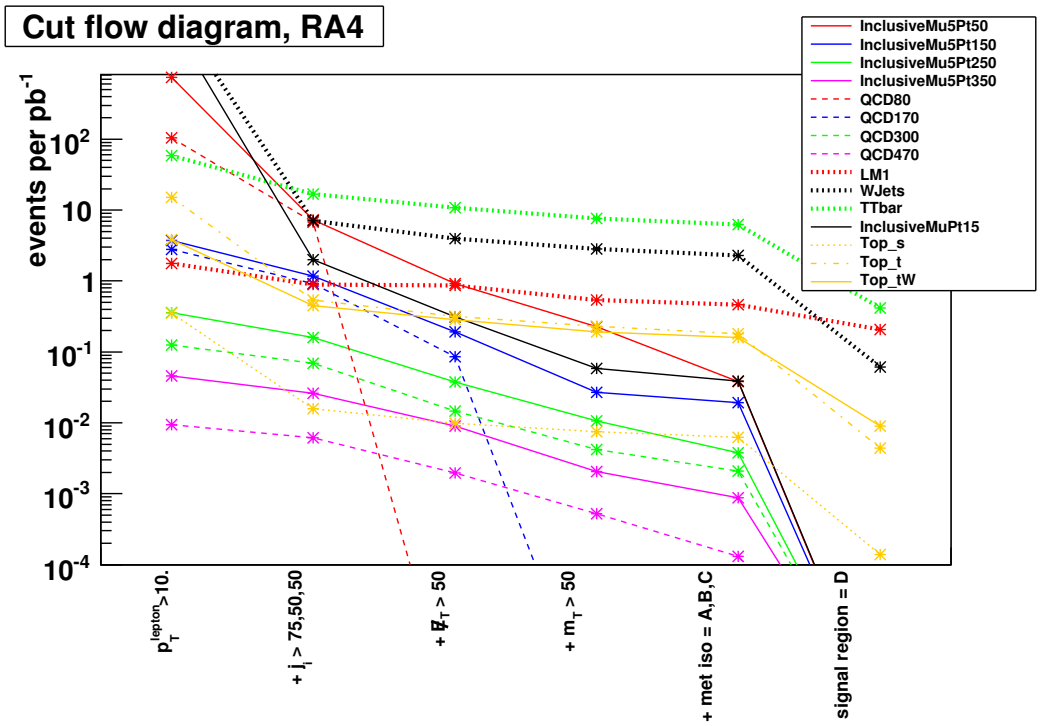


Figure 16.113: Cut flow for the signal suppression method.

Here we have used just the same definition of ABCD as in Sec. 16, see Fig. 17.114. λ_i in the equation above refers to the Poissonian parameter (a.k.a. the average number of events) in region i.

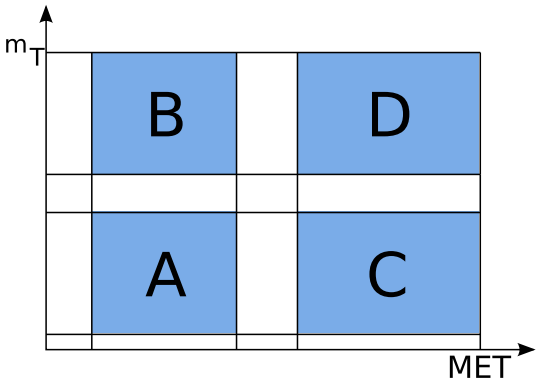


Figure 17.114: This section's definition of the ABCD regions.

In realistic scenarios, Eq. 40 does not hold. We generalize ABCD to

$$\frac{\lambda_A}{\lambda_B} = \kappa \frac{\lambda_C}{\lambda_D} \tag{41}$$

where κ is to be measured in a signal-free region. The eight fields method works as follows: We perform ABCD in the $\not{E}_T$ and the $M3$ variable, where $M3$ is the invariant mass of those three jets that maximize their total p_T. Additionally to the base selection (Sec. 16.2), $m_T > 50$ is required. κ' is measured from A'B'C'D', which differ from the unprimed regions in that $50 < m_T < 100$ is demanded. κ' is then propagated to the unprimed regions. Fig. 17.115 attempts to capture the method in a simple plot.

An important cross check is the comparison of κ and κ', i.e. the κ obtained in the primed and in the unprimed regions. Tab. **??** lists the κ values for different signals. It seems clear that the method will not work perfectly; still, from the table we expect it to improve upon the classical ABCD method.

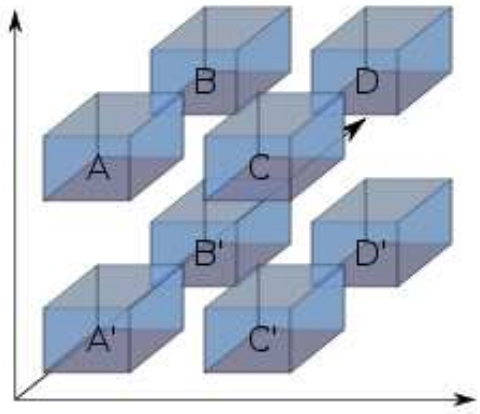


Figure 17.115: ABCD in eight fields

signal	κ in unprimed regions	κ' in primed regions
no signal	1.14	1.09
LM1	2.14	1.75
LM2	1.43	1.20
LM8	1.56	1.23

17.2 Performance of method without signal

The method performs well in the case of signal absence: Fig. 17.116 shows the plots obtained with the bootstrapping approach already discussed in Sec. 16.3. It must be noted, however, that introducing the additional $M3$ variable comes at the price of a lower statistics than the ABCD method.

signal events	bkg events	estimated bkg
0	21.62	$20.57 \pm 11.08^{\mathrm{stat}}$

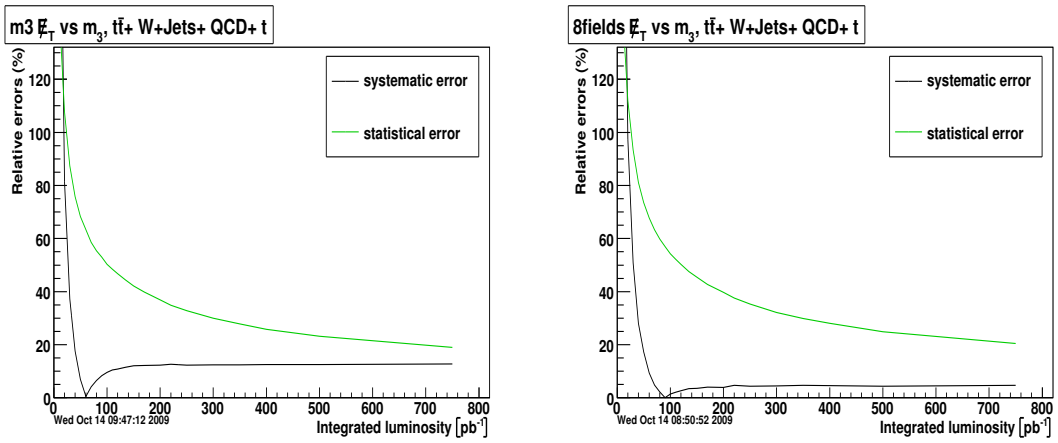


Figure 17.116: Relative errors of $\not{E}_T$ versus $M3$ (left) and eight fields method (right), as a function of integrated luminosity.

17.3 Performance of method with signal

In the presence of a signal, the eight fields becomes virtually unusable:

signal	signal events	bkg events	estimated bkg
no signal	0	21.62	$20.57 \pm 11.08^{\mathrm{stat}}$
LM1	30.65	21.62	$42.63 \pm 20.43^{\mathrm{stat}}$
LM2	6.17	21.62	$23.36 \pm 12.41^{\mathrm{stat}}$
LM8	12.11	21.62	$26.55 \pm 13.43^{\mathrm{stat}}$

Still, for higher integrated luminosities it remains a powerful idea for dealing with ABCD variable correlations. Combining this strategy with methods that can cope with signal contamination (e.g. the signal suppression cut, Sec. 16) is an interesting future topic.

17.4 Systematic uncertainties

Since the method is unable to cope with signal contamination we restrict ourselves to study systematic uncertainties in the absence of a signal:

For the dependency on the jet energy scale:

jet energy scale	true bkg in D	estimated bkg in D
1.2	81.90	$80.86 \pm 28.40^{\mathrm{stat}}$
1.1	47.09	$46.88 \pm 17.87^{\mathrm{stat}}$
1.0	21.62	$20.57 \pm 11.08^{\mathrm{stat}}$
0.9	12.07	$10.38 \pm 10.62^{\mathrm{stat}}$
0.8	7.04	$4.43 \pm 8.03^{\mathrm{stat}}$

The estimate is stable with respect to fluctuations in the jet energy scales.

17.5 Early data commissioning plan

Additionally to the commissioning plans presented in Sec. 16, the measured κ can be compared with the Monte Carlo truth.

18 Determination of Fake Electron Background

In this section, we investigate a data-driven method that estimates the background contribution from events with fake electrons: the relative isolation extrapolation method. We define fake electrons to be reconstructed electrons that are not from the decay of W bosons, Z bosons, τ leptons, or new physics particles. From this definition, the two major sources of electron fakes that our method attempts to estimate are:

1. **A jet faking an electron**: Here, a jet may deposit a large amount of energy in the ECAL that is matched to a single high-p_{T} track that is reconstructed as an electron.
2. **A heavy quark decaying to an electron**: Although these electrons are usually not isolated, the kick from the quark decay might knock the electron out of the jet enough for it to appear isolated.

The methods presented here has been used in previous experiments as well as by other groups within CMS. For example, see [3] for a detailed explanation of their extrapolation of isolation method. In that Note, the methods are targeted to predict the fake contribution to the prompt electron and muon signatures of WW and $t\bar{t}$ events.

For this study, we employ the following object selection criteria:

- **Electron**:
  - GSF Electron
  - $p_{\mathrm{T}} > 20$ GeV
  - $|\eta| < 2.5$, and η not in ECAL gap ($1.47 < |\eta| < 1.567$)
  - "robustLoose" identification [2]
  - $\mathrm{RelIso} \equiv \frac{1}{p_T^{\mathrm{ele}}} \sum^{\mathrm{iso\ dep}} \left(E_T^{ECAL} + E_T^{HCAL} + p_T^{Track}\right) < 0.1$
    - Note that the calorimeter isolation deposits in this sum do not include the energy deposited by the electron
- **Jet**:
  - L2L3 corrected sisCone5 Jet
  - $p_T > 50$ GeV
  - $|\eta| < 3.0$
  - Electromagnetic Fraction (EMF) < 0.9
  - Jets within $\Delta R < 0.3$ of an identified electron are not counted

Our event selection for the single electron final state is:

- **Trigger**: passes single electron trigger HLT_Ele15_LW_L1R,
- **Muon Veto**: no muons passing the tight muon selection in Sec. 3,
- **Electron**: exactly one electron,
- **Jets**: four or more jets,
- $\not\!\!E_T$: SC5 jet and muon corrected$\not\!\!E_T > 150$ GeV.

18.1 Description of method

The isolation extrapolation method works by extrapolating from a region dominated by the background to the signal region. Here, we use relative isolation (RelIso), given by

$$\mathrm{RelIso} = \frac{1}{p_T^{\mathrm{ele}}} \sum^{\mathrm{iso\ dep}} \left(E_T^{ECAL} + E_T^{HCAL} + p_T^{Track}\right) , \tag{42}$$

as a discriminator between electrons and fakes. By fitting the relative isolation distribution above some threshold, which we assume is dominated by QCD (fakes), we can get the full shape of this distribution and use it to predict the contribution from the QCD background.

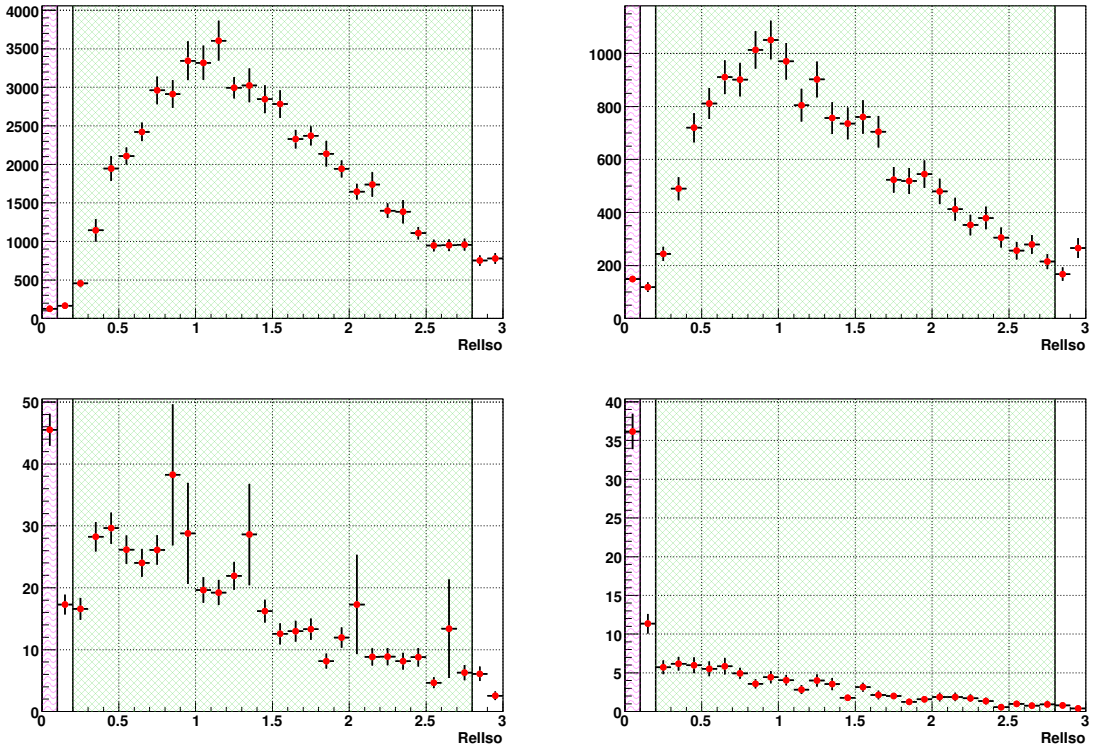


Figure 18.117: RelIso distribution for all combined MC sample after requiring 4 good jets, at most one electron passing our full selecion, and a $\not\!\!E_T$ cut of $0 < \not\!\!E_T < 25$ GeV (top left), $50 < \not\!\!E_T < 75$ GeV (top right), $100 < \not\!\!E_T < 125$ GeV (bottom left), and $\not\!\!E_T > 150$ GeV (bottom right). From this, we choose the signal region (purple wavy line region) to be $0 <$ RelIso < 0.1 and the fit region (green cross-hatched region) to be $0.2 <$ RelIso < 2.8.

18.1.1 Choosing signal and fit regions

Since the shape of RelIso for electrons will not necessarily be the same for our MC samples and data (due to hard to model detector effects, changes in isolation definitions, etc.), we describe the method for determing signal and fit regions for this variable. For our combined MC sample (MadGraph QCD, W+Jets, Z+Jets, and TT+Jets), we expect RelIso to be the combination of shapes from two different sources: electrons and fakes. Real electrons are expected to peak at low values of RelIso (since they are well isolated). Conversely, fake electrons are expected to peak away from a RelIso of zero with a long tail out to large isolation values. Some of these fake electrons will have a RelIso in the signal region, and this is the fake contribution we are trying to estimate.

In this way, we can determine an ideal signal region by seeing how far the real electron RelIso distribution extends. Once we have the signal region, we choose the fit region to start at a slightly higher RelIso value (to provide some separation) and then we cut off the fit region at some reasonable value. To determine the RelIso signal region, we first require that the event have 4 or more good jets (the jet cut for the signal region). In addition, we require that there be no muons and at most one electron that passes our selection criteria. Then, we plot the RelIso of all electrons passing our selection with the isolation requirement removed. To distinguish between the real and fake electron RelIso shape, we do not initially apply the $\not\!\!E_T$ requirement for the signal region. Here, we expect that the dominant shape will be that for fake electrons (as QCD will dominate without any $\not\!\!E_T$ cut). We then apply more and more strict $\not\!\!E_T$ requirements and note how the real electron RelIso shape becomes more prominent with the harder $\not\!\!E_T$ cuts. From Fig. 18.117, we can see just this behavior, and we choose the signal region to be $0 <$ RelIso < 0.1 and the fit region to be $0.2 <$ RelIso < 2.8.

18.2 Performance of method without signal

Before applying our isolation extrapolation method to all MC samples (to mimic how the procedure will be used for data), we first tested the method when running only over our QCD samples, which only contain fake electrons. This was done as a consistency check. We found that the function that best fit the RelIso distribution was

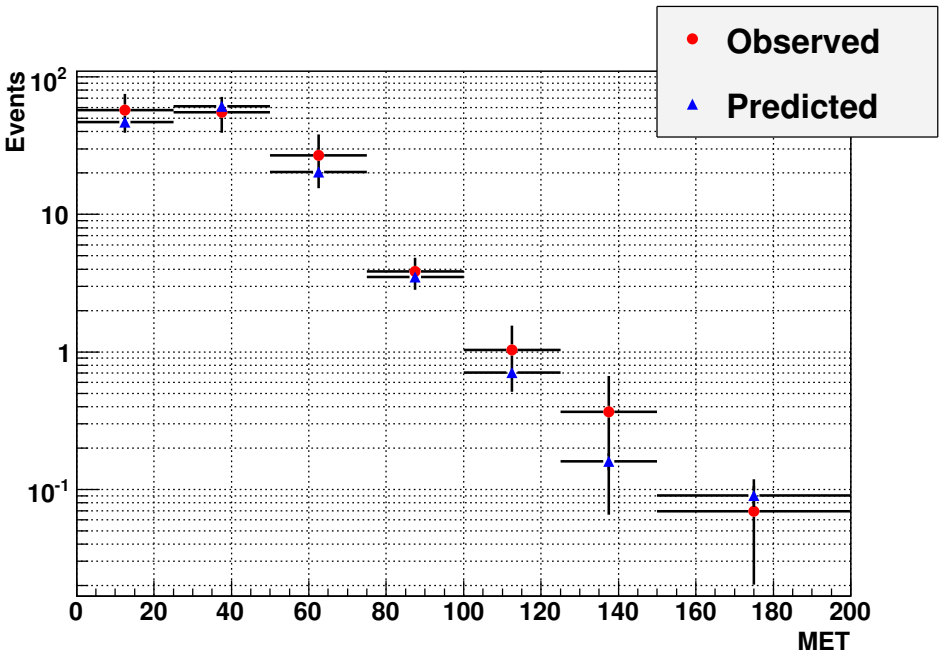


Figure 18.118: Comparing observed (from summing the histogram) and predicted (from the fit) number of events with no muons, 1 fake electron, and at least 4 good jets, as a function of $\not\!\!E_T$ for our QCD sample. The last $\not\!\!E_T$ bin includes all events with $\not\!\!E_T > 150$ GeV. Overall, the method performs very well.

$$f(x; p_0, p_1, p_2) = p_0 \cdot x^2 \cdot \mathrm{Exp}\left[\frac{-(x-p_1)^2}{p_2^2}\right] \tag{43}$$

where x is the RelIso variable and (p_0, p_1, p_2) are the three parameters of the function.

After fitting the RelIso shape for the QCD sample using eq. (43) in the fit region ($0.2 <$ RelIso < 2.8), we extrapolate this fit to the signal region ($0 <$ RelIso < 0.1). Then, we compare the observed number of events in the signal region to the predicted number of events found using our fit. In order to increase our statistics and to test the methods for different $\not\!\!E_T$ cuts, we compare observed and predicted number of events for seven different $\not\!\!E_T$ regions in Fig. 18.118. The error on the prediction was found by propagating the error on each of the fit parameters to the error on the fit function in our signal region. We see that the prediction agrees very well with the observed number of events for all $\not\!\!E_T$ bins.

18.3 Performance of method with signal

Figure 18.119 shows the RelIso distribution and its fit, using eq. (43), for all MC samples after requiring at least 4 good jets and $\not\!\!E_T > 150$ GeV. The shape of the RelIso distribution for real electrons should be described by subtracting the fake background shape (our fit) from the total RelIso distribution. To test this, we compared our background-subtracted RelIso distribution to the MadGraph W+Jets RelIso distribution after requiring at least 4 good jets (but no $\not\!\!E_T$ requirement). Since the W+Jets sample has much more statistics than our combined sample after the jet and $\not\!\!E_T$ cuts, we normalized both distributions to have unit area. This comparison is shown in Fig. 18.120, and we see that the two shapes agree very well.

18.4 Systematics

The two largest sources of systematic error for this method come from the choice of fit region and fit function. To estimate the systematic error, we begin by varying the fit region. As described in Sec. 18.1.1, the fit region was chosen to be $0.2 <$ RelIso < 2.8. The start and the end boundary of the fit region (0.2 and 2.8) were either held fixed or varied by $\pm$ 0.1 to give 9 different fit regions. For a given fit region, several different fit functions were applied (e.g. variants of eq. (43), a Landau distribution, etc.), and we took the prediction from the best fit function (determined by the lowest fit $\chi^2/ndof$.)

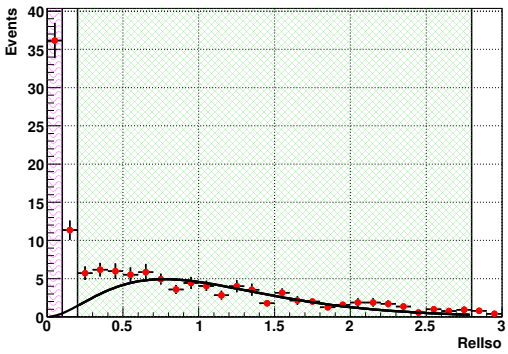


Figure 18.119: RelIso distribution for all MC samples and the resulting fit after requiring 4 good jets and $\not\!\!E_T > 150$ GeV.

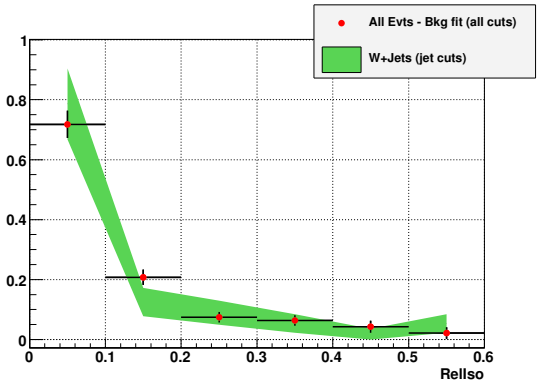


Figure 18.120: Comparing the RelIso shape in the signal region for all MC samples after subtracting the 1 fake background (red dots) to MadGraph W+Jets events without the $\not\!\!E_T$ cut (green shaded region). Both were normalized to unit area, and they agree within error.

In this way, we sample several fit regions and fit functions, and we can observe our sensitivities to these choices. As a conservative estimate, our systematic error is given by the range of predictions using the variations described above. We note that this estimate can be performed in data. Using this procedure, we predict the number of events with 1 fake electron in the signal region to be

$$N_{1\text{fake}} = 0.16 \pm 0.02(\text{fit}) \pm 0.32(\text{sys})$$

where the first error comes from propagating the errors on the fit parameters and the second error is found using the procedure described in this section.

18.5 Early data commissioning plan

With the early data, the emphasis will be placed on commissioning and understanding our detector. So, while we will not be able to perform our analysis during this time, it is important that we use the early data to understand the various quantities that will eventually enter into our analysis. For this data-driven method and understanding the behavior of fake electrons, it will be essential that we study electrons in great detail. Therefore, we will study two separate skims of data: one believed to be enriched in electrons and another believed to be enriched in fakes. The former can achieved by running on the electron secondary dataset SD_Ele15_LW_L1R with additional event selection cuts to reduce QCD. The skim to study fake electrons can be obtained by running on the SD_Jet50U or CS_DiJetAve30 secondary datasets with additional event selection cuts to enhance QCD. Once we have these two skims, we can investigate electron isolation in both cases. In addition, we can investigate how the electron isolation shape changes depending on the event topology (e.g. additional jets, $\not\!\!E_T$, etc.)

19 QCD determination in $\chi^2(t\bar{t})$ method

19.1 Description of Method

Various techniques for estimating the QCD background in the $\chi^2(t\bar{t})$ vs $\not\!\!E_T/\Sigma E_T$ method were studied. The $\chi^2(t\bar{t})$ vs $\not\!\!E_T/\Sigma E_T$ method relies on the lack of correlation between the variables in $t\bar{t}$ events to separate them from the SUSY signal. However, QCD background in the data could interfere with the estimation by effecting the balance of events between the regions. Although MC simulation predicts that the effect will be very small, this study prepares for the possibility of a much larger QCD contamination in the data by estimating it in each $\chi^2(t\bar{t})$ vs $\not\!\!E_T/\Sigma E_T$ region.

The QCD samples used here were generated by Herwig and are in the form of configurable analysis ntuples called InclusiveMu5. Originally in four pieces of $\hat{p}_T$, the samples were stitched together by cutting on the leading MC parton p_T for each event. The $t\bar{t}$ ntuples used were obtained from Madgraph.

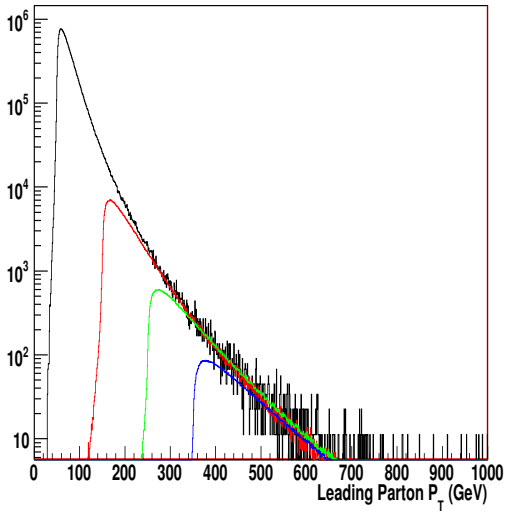


Figure 19.121: Leading MC parton p_T for the four inclusive muon samples.

The typical RA4 cuts for identifying jets and leptons were used for this study. In addition, it required 3 jets with $p_T > 80$ GeV and a 4th jet with $p_T > 30$ GeV. No relative isolation cut was put on the single lepton being studied.

A major distinguishing variable in QCD, which can be much higher than t$\bar{t}$ events, is the relative isolation of the lepton.

$$RelIso = \frac{E_{iso}^{calo} + p_{T,iso}^{tracker}}{p_{T,lepton}} \tag{44}$$

(45)

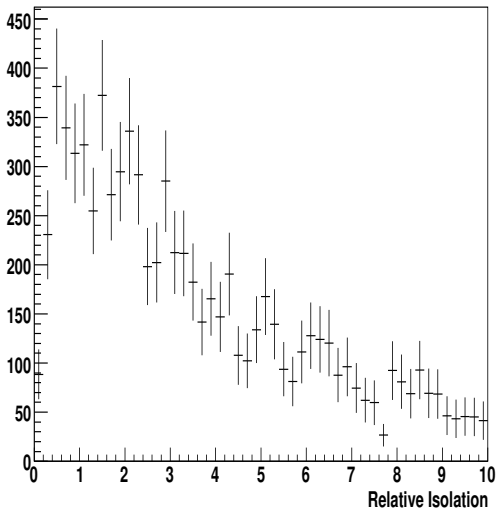


Figure 19.122: Relative isolation for QCD muons.

Without this cut, many more QCD events pass the selection, and the amount of isolated background can be estimated for each region by its behavior at higher relative isolation. Figures 19.125 and 19.125 show the events in $\chi^2(t\bar{t})$ and $\not\!\!E_T/\Sigma E_T$ for the QCD sample alone and the QCD sample with $t\bar{t}$ added, respectively.

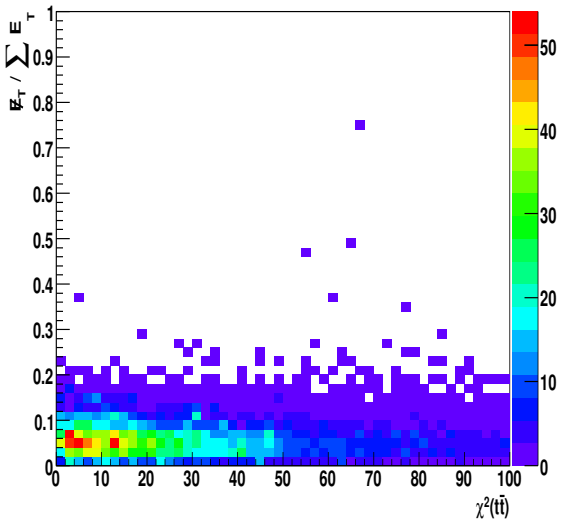


Figure 19.123: $\chi^2(t\bar{t})$ vs $\not\!\!E_T/\Sigma E_T$ for QCD events.

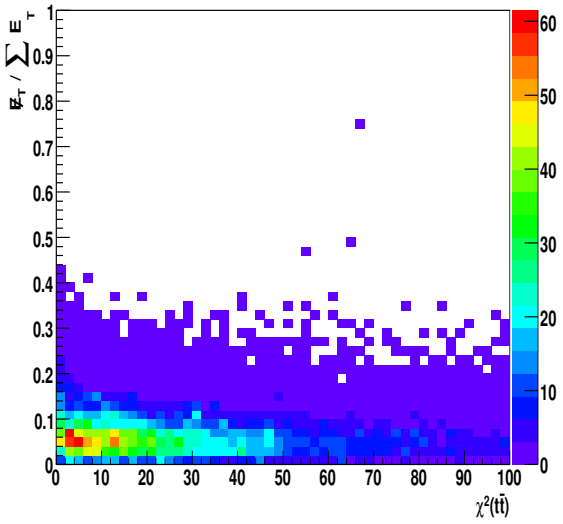


Figure 19.124: $\chi^2(t\bar{t})$ and $\not\!\!E_T/\Sigma E_T$ for QCD and $t\bar{t}$ events.

Prompt muons in the $t\bar{t}$ sample leak beyond a relative isolation of 0.1, so to extrapolate the distribution for QCD and $t\bar{t}$ secondary muons, we should only consider the shape above a relative isolation of 0.3 or higher. An attempt was made to make a linear fit on the relative isolation of the QCD sample over the region between 0.3 and 1.0. However, upon observing the shape at higher statistics by loosening jet cuts, this fit was shown to overestimate the isolated background muons, due to the non-linear shape of the distribution below a relative isolation of 0.3.

19.2 Performance of method without signal

A way of estimating the number of QCD events falling below a relative isolation of 0.1 involves fitting the distribution to a different set of events having the same shape. Unlike a linear approximation, this method is more accurate and can be used over a larger fit region, improving the quality of the fit. Leptons used to predict the shape of our distribution have been chosen from events with two jets having $p_T > 30$ and no others with $p_T < 30$ GeV.

The shape of relative isolation for events passing the analysis cuts, along with the template fit to it, is shown in figure 19.125.

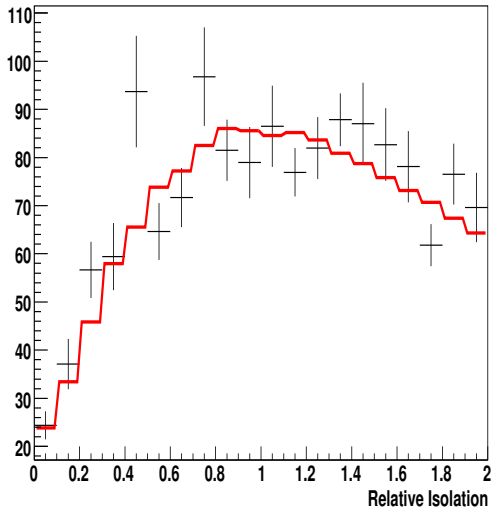


Figure 19.125: Relative isolation for QCD leptons with the dijet template fit.

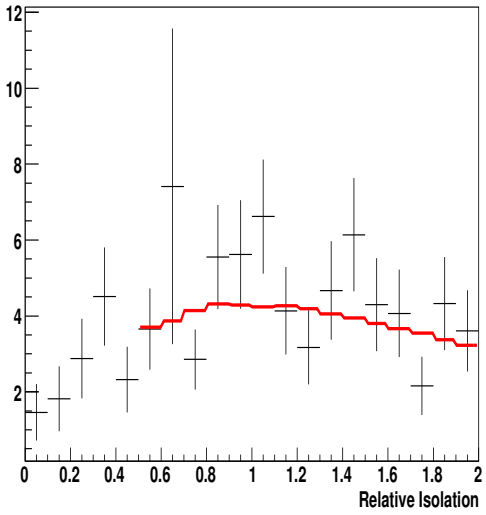
This fit was performed for each region in χ^2(t$\bar{t}$) vs$\not\!\!E_T/\Sigma E_T$, over several different ranges. The choice of template cuts and range effects the resulting estimation, and therefore contributes to statistical uncertainty. Figure 19.126 demonstrates the how the template works on the QCD sample only, for each region in χ^2 and $\not\!\!E_T/\Sigma E_T$. The t$\bar{t}$ sample is added for figure 19.127. Starting from the upper left and moving clockwis for both figurese, each region represents χ^2 >20 and$\not\!\!E_T/\Sigma E_T$ <0.1, χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2, χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2, and χ^2 <20 and$\not\!\!E_T/\Sigma E_T$ <0.1.

Region	Template Estimate Fit from 0.4 to 1.0	Template Estimate Fit from 0.5 to 2.0	Actual QCD and Secondary Muons
χ^2 <20 and$\not\!\!E_T/\Sigma E_T$ <0.1	2.10 $\pm$ 0.04	2.09 $\pm$ 0.04	0.6 $\pm$ 0.4
χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0	0.095 $\pm$ 0.002	0.080 $\pm$ 0.001	0
χ^2 >20 and$\not\!\!E_T/\Sigma E_T$ <0.1	1.87 $\pm$ 0.03	1.79 $\pm$ 0.03	1.5 $\pm$ 0.7
χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2	0.176 $\pm$ 0.003	0.059 $\pm$ 0.001	0.5 $\pm$ 0.4

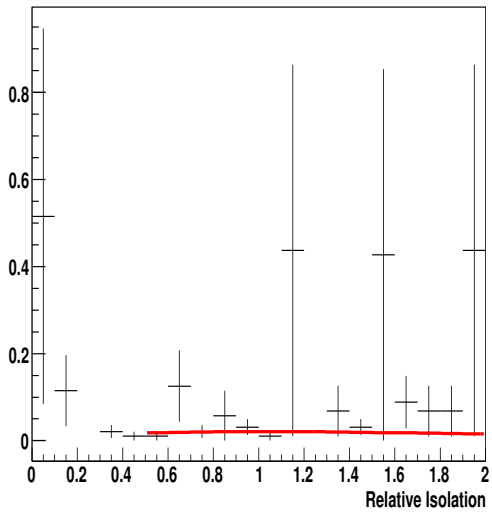
A known technique for distinguishing QCD from $t\bar{t}$ events involves an ABCD estimation in relative isolation and impact parameter significance. Muons from QCD events are uncorrelated in these variables, while $t\bar{t}$ prompt muons are concentrated at low relative isolation and impact parameter significance.

A major benefit of this technique is that it does not rely on any presumption of a shape in the distribution of relative isolation for QCD in the data. The regions are divided into relative isolation of < 0.1 and > 0.5, and impact parameter significance of <3 and >5. However, due to very low statistics, the estimation of background events in the $t\bar{t}$ region has a very large error. Starting from the upper left and moving clockwise, each region represents χ^2 >20 and$\not\!\!E_T/\Sigma E_T$ <0.1, χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2, χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2, and χ^2 <20 and$\not\!\!E_T/\Sigma E_T$ <0.1.

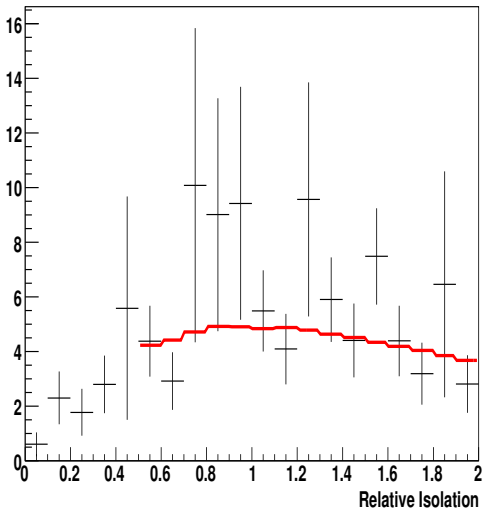
Region	2nd ABCD Method	Actual QCD
χ^2 <20 and$\not\!\!E_T/\Sigma E_T$ <0.1	2 $\pm$ 3	0.6 $\pm$ 0.4
χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2	0.23 $\pm$ 0.05	0
χ^2 >20 and$\not\!\!E_T/\Sigma E_T$ < 0.1	0.8 $\pm$ 0.9	1.5 $\pm$ 0.7
χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2	0.5 $\pm$ 0.1	0.5 $\pm$ 0.4



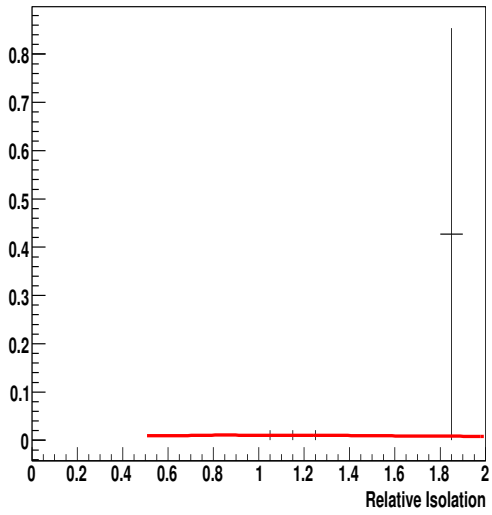
(a) χ^2 >20 and $\not\!\!E_T/\Sigma E_T$ <0.1



(b) χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2

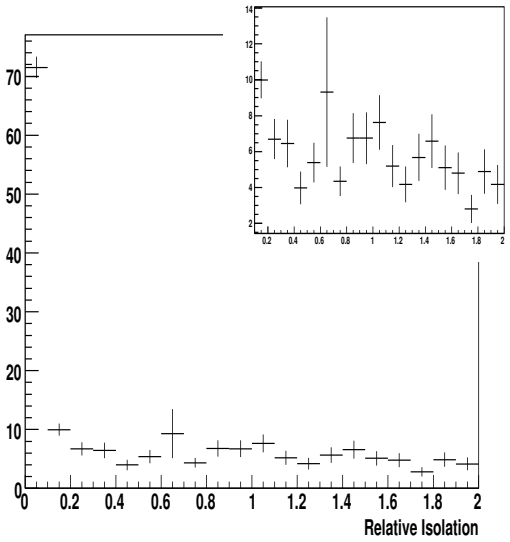


(c) χ^2 <20 and $\not\!\!E_T/\Sigma E_T$ <0.1

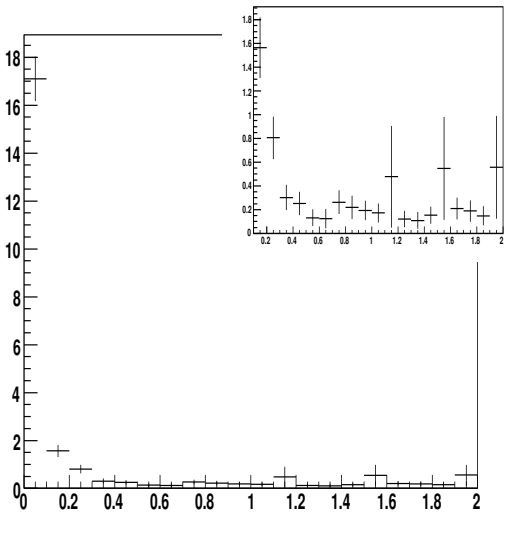


(d) χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2

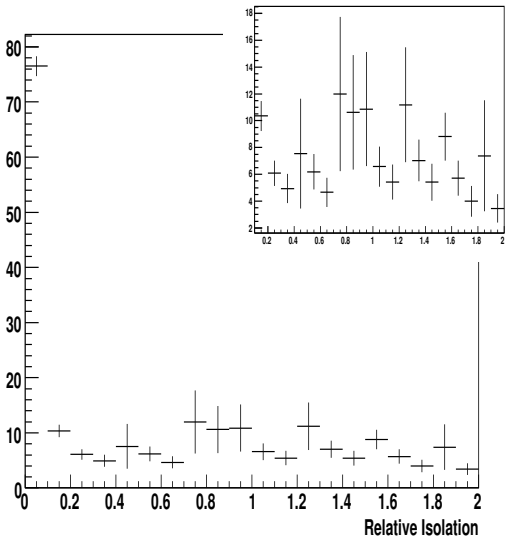
Figure 19.126: Template fits for QCD, by regions of χ^2 and $\not\!\!E_T/\Sigma E_T$



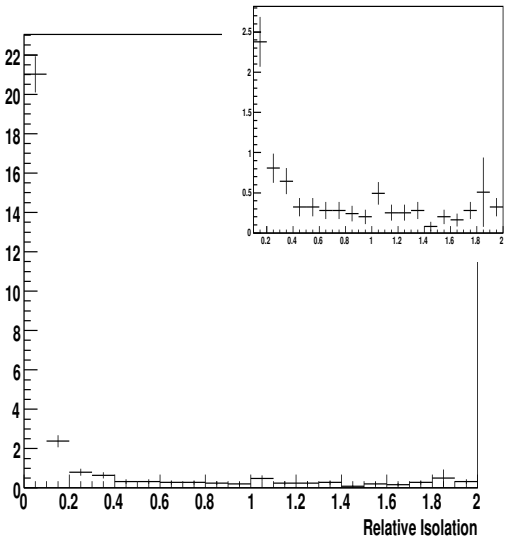
(a) χ^2 >20 and$\not\!\!E_T/\Sigma E_T$ <0.1



(b) χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2

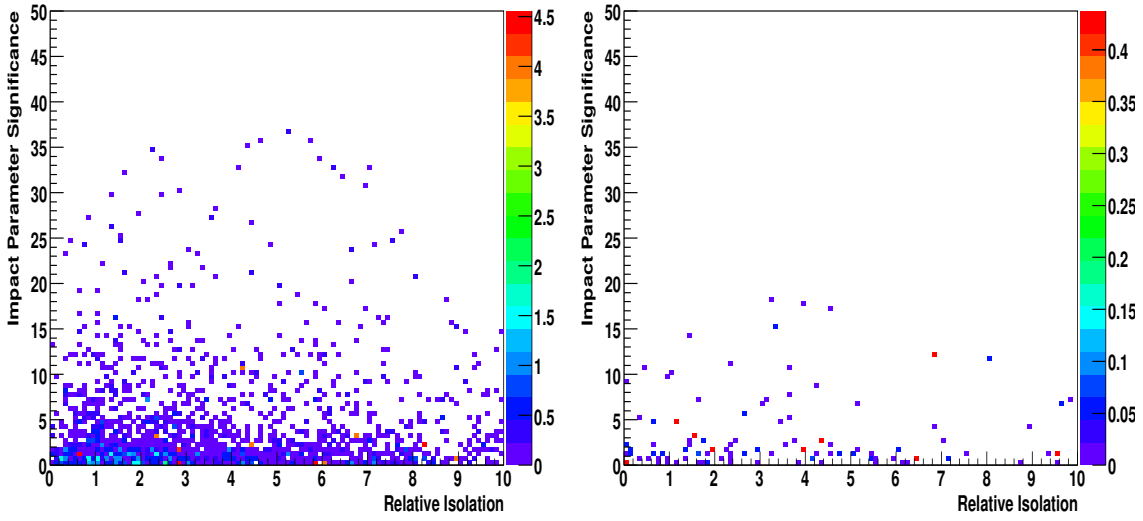


(c) χ^2 <20 and$\not\!\!E_T/\Sigma E_T$ <0.1



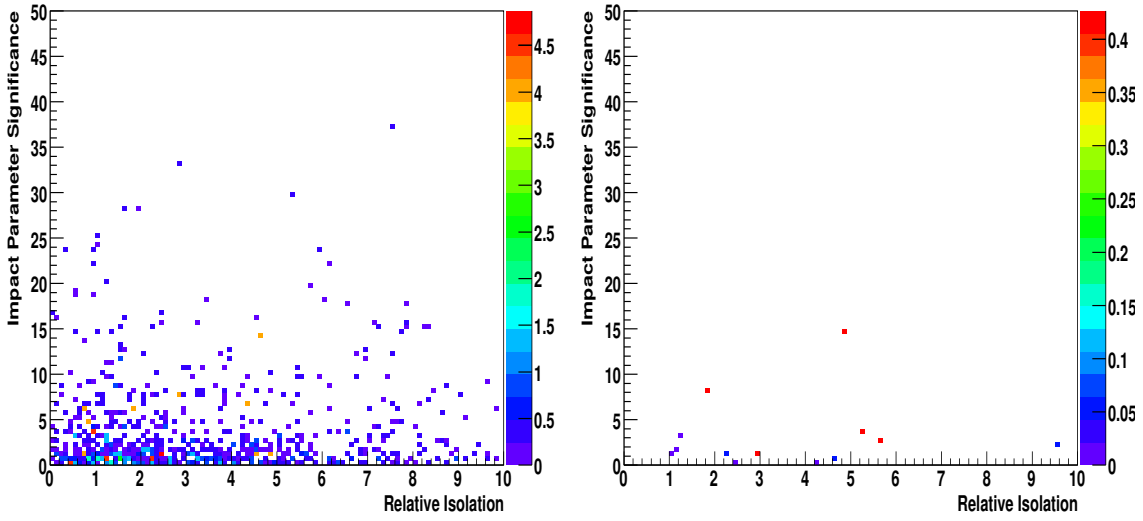
(d) χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2

Figure 19.127: Template fits for QCD and t$\bar{t}$, by regions of χ^2 and$\not\!\!E_T/\Sigma E_T$



(a) χ^2 >20 and $\not\!\!E_T/\Sigma E_T$ <0.1

(b) χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2



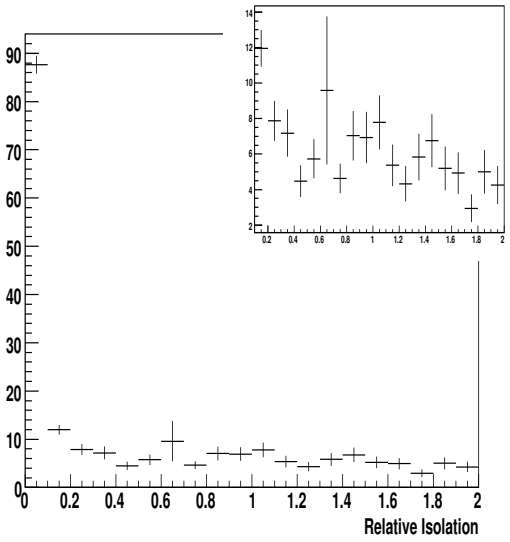
(c) χ^2 <20 and $\not\!\!E_T/\Sigma E_T$ <0.1

(d) χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2

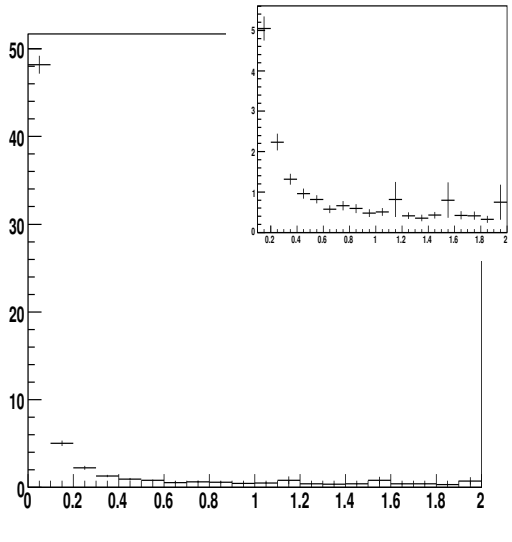
Figure 19.128: Relative isolation vs impact parameter significance, for the QCD events only.

19.3 Performance of method with signal

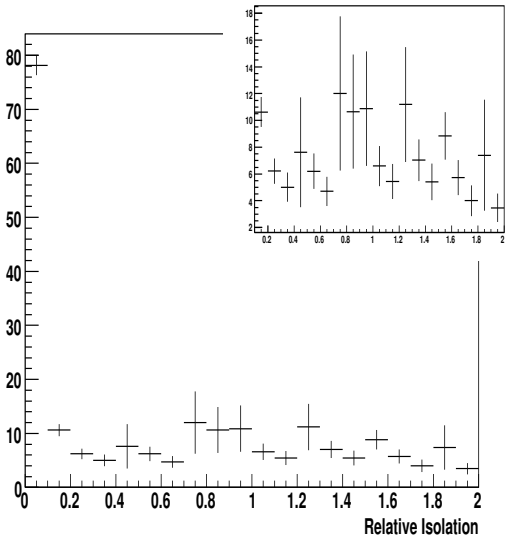
The same procedures were carried out with the addition of an LM1 MC sample. The following figures illustrate the results of performing the template fit and ABCD method estimations. Figure 19.129 shows a template fit for all three samples, and figure 19.130 shows the second ABCD method applied to the same.



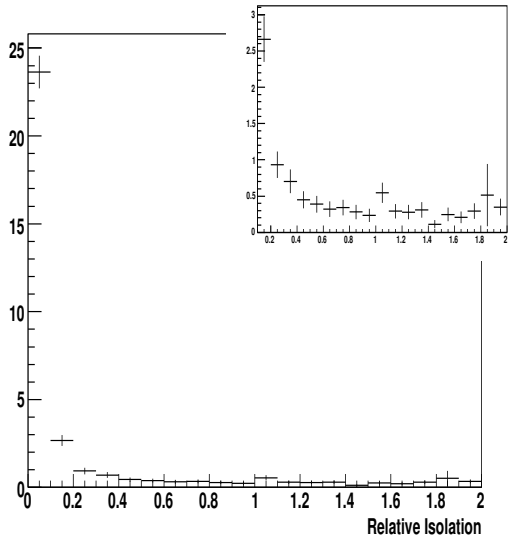
(a) χ^2 >20 and $\not\!\!E_T/\Sigma E_T$ <0.1



(b) χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2

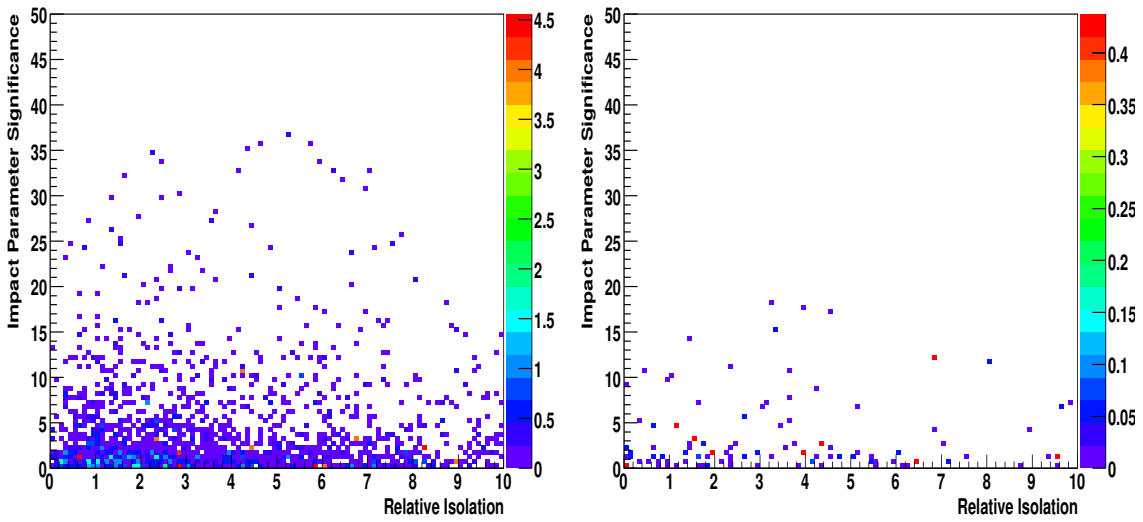


(c) χ^2 <20 and $\not\!\!E_T/\Sigma E_T$ <0.1



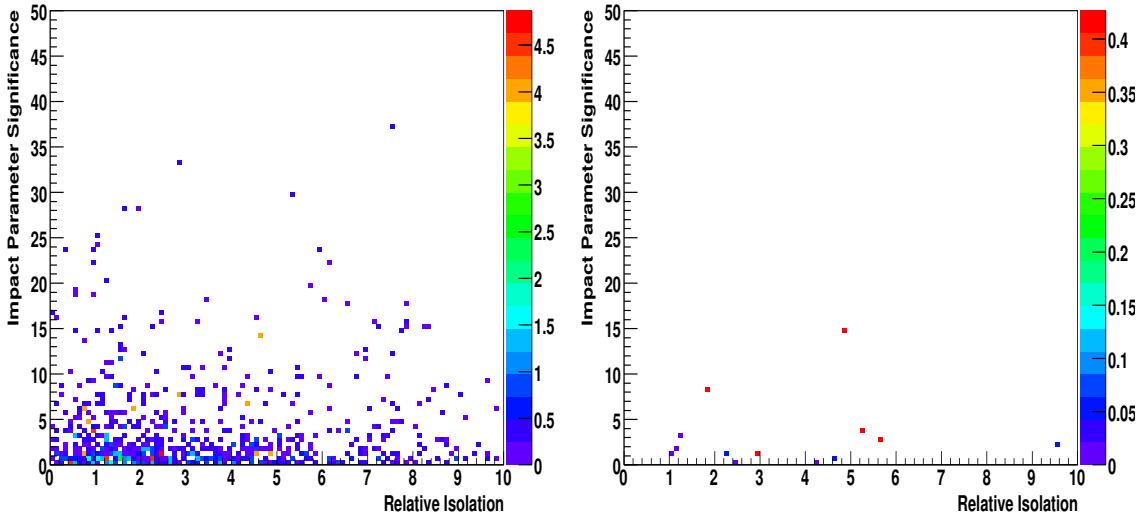
(d) χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2

Figure 19.129: Template fits for QCD, t$\bar{t}$, and LM1, by regions of χ^2(t$\bar{t}$) and $\not\!\!E_T/\Sigma E_T$



(a) χ^2 >20 and $\not\!\!E_T/\Sigma E_T$ <0.1

(b) χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2



(c) χ^2 <20 and $\not\!\!E_T/\Sigma E_T$ <0.1

(d) χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2

Figure 19.130: Relative isolation vs impact parameter significance for QCD, $t\bar{t}$, and LM1, by regions of $\chi^2(t\bar{t})$ and $\not\!\!E_T/\Sigma E_T$

Region	Template Estimate Fit from 0.4 to 1.0	Template Estimate Fit from 0.5 to 2.0	2nd ABCD Method	Actual QCD
χ^2 <20 and $\not\!\!E_T/\Sigma E_T$ <0.1	2.20 ± 0.04	2.09 ± 0.04	2 ± 3	0.6 ± 0.4
χ^2 <20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2	0.113 ± 0.002	0.080 ± 0.001	0.32 ± 0.06	0
χ^2 >20 and $\not\!\!E_T/\Sigma E_T$ <0.1	1.99 ± 0.03	1.80 ± 0.03	1 ± 1	1.5 ± 0.7
χ^2 >20 and 0.1<$\not\!\!E_T/\Sigma E_T$ <0.2	0.234 ± 0.004	0.586 ± 0.001	1.4 ± 0.2	0.5 ± 0.4

We have studied two possible techniques for estimating QCD contamination within the $\chi^2(t\bar{t})$ vs$\not\!\!E_T/\Sigma E_T$ method. Both show reasonable results, but with high errors due to low statistics. Having both the template fit and second ABCD method will increase our understanding of QCD the data such that the $\chi^2(t\bar{t})$ vs $\not\!\!E_T/\Sigma E_T$ analysis is not adversely effected.

19.4 Commissioning

The first few inverse picobarns of data will be very important for the commissioning of this method. The QCD MC samples used in these studies do not include fake muons, which could potentially double this background. Very soon after the start of data taking, we will obtain enough QCD events to properly characterize its behavior with respect to relative isolation. From there, we can optimize the appropriate template fit or ABCD regions for the sideband subtraction.

20 Combining information from different methods

This section makes a comparison of the various approaches to the background determination in the single-lepton SUSY search. We discuss how to combine the information from different methods.

21 Conclusions

This section summarizes the results of this study and presents a set of conclusions and directions for future work.

References

[1] C. Magass, "Study of Muon ID for RA4," *Talks given in Leptonic SUSY Meetings May 7 & July 2, 2009* .

[2] J. Branson, M. Gallinaro, P. Ribeiro, R. Salerno, and M. Sani, "A cut based method for electron identification in CMS," *CMS Note* **2008/082** (2008) .

[3] W. Andrews, D. Evans, F. Golf, J. Mulmenstadt, S. Padhi, Y. Tu, F. Wurthwein, A. Yagil, C. Campagnari, P. Kalavase, D. Kovalskyi, V. Krutelyov, J. Ribnik, L. Bauerdick, I. Bloch, K. Burkett, I. Fisk, and O. Gutsche, "Data-driven methods to estimate the electron and muon fake contributions to lepton analyses," *CMS Note* **2009/041** (2009) .