

Search for new physics in the 2011 opposite sign dilepton sample

D. Barge, C. Campagnari, P. Kalavase, D. Kovalskyi, V. Krutelyov, J. Ribnik

University of California, Santa Barbara

W. Andrews, G. Cerati, D. Evans, F. Golf, I. Macneill, S. Padhi, Y. Tu, F. Würthwein, A. Yagil, J. Yoo

University of California, San Diego

L. Bauerdick, I. Bloch, K. Burkett, I. Fisk, Y. Gao, O. Gutsche, B. Hooberman, S. Jindariani, J. Linacre

Fermi National Accelerator Laboratory, Batavia, Illinois

Version v2

Abstract

We present the results of a search for new physics in the opposite-sign leptons + jets + missing transverse energy final state based on 349 pb^{-1} of 2011 CMS data. We observe no evidence for an excess of events with large hadronic activity and missing transverse energy, and place upper limits on the non-SM contributions to the signal regions.

Contents

1	Changes since Last Version of AN	3
2	Introduction	3
3	Datasets, Triggers, Luminosity	3
4	Event Preselection	4
4.1	Event Cleanup	4
4.2	Muon Selection	5
4.3	Electron Selection	5
4.4	Invariant mass requirement	5
4.5	Trigger Selection	6
5	Trigger efficiency	6
6	Dilepton Yields	6
7	Preselection yields: Data/MC Comparison	7
8	Properties of data passing the preselection	7
9	Definition of the signal region	8
10	Data Driven Background Estimation Methods	13
10.1	ABCD method	13
10.2	ABCD' Method	14
10.3	Dilepton p_T method	15
10.4	Opposite-Flavor Subtraction	16
10.4.1	OF subtraction: Application to high p_T lepton sample	16
10.4.2	OF subtraction: Application to low p_T lepton sample	16
11	Results	18
11.1	Background estimate from the ABCD method	18
11.2	Background estimate from the $P_T(\ell\ell)$ method	18
11.3	Background estimate from OF subtraction	20
11.4	Summary of results	21
12	Acceptance and efficiency systematics	21
13	Derivation of Upper Limits	23
	Appendix A Fakeable Object Definitions	26

Appendix B The ABCD' Technique	27
Appendix C Data/MC Comparison: Preselection Region	30
Appendix D Data/MC Comparison: 2010 Signal Region	51

1 Changes since Last Version of AN

- Version v2
 - Update results to 349 pb^{-1}
 - Switch to signal regions in E_T^{miss} vs. H_T plane
- Version v1
 - Add OF subtraction studies for low lepton p_T sample
 - Add LM1, change MC normalization for data/MC plots in appendix
 - Several minor editorial comments

2 Introduction

In this note we describe a search for new physics in the opposite sign isolated dilepton sample (ee , $e\mu$, and $\mu\mu$) based on 349 pb^{-1} of 2011 data. The main source of isolated dileptons at CMS is Drell Yan and $t\bar{t}$. Here we concentrate on dileptons with invariant mass inconsistent with $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$, thus $t\bar{t}$ is the most important background. A separate search for new physics in the Z sample is described in a separate note [1]. This is an update of a (soon to be) published analysis performed on 2010 data [2, 3].

The search strategy is the following

- We start out with a pre-selection which is as close as possible to the published $t\bar{t}$ dilepton analysis [4] (same lepton ID, same jet definitions, etc.). We do make a couple of substantive modifications:
 1. The top analysis requires two leptons of $P_T > 20 \text{ GeV}$. In this analysis we search for new physics in 2 data samples. The first data sample is collected with high p_T dilepton triggers; for this sample we require leptons with $p_T > (20, 10) \text{ GeV}$ (leading lepton $p_T > 20 \text{ GeV}$, trailing lepton $p_T > 10 \text{ GeV}$). The second data sample is collected with dilepton- H_T cross triggers; for this sample we require leptons with $p_T > (10, 5) \text{ GeV}$. This is motivated by our desire to maintain sensitivity to possible SUSY signals with relatively low p_T leptons generated in the cascade decays of heavy objects.
 2. The top analysis requires at least two jets of $p_T > 30 \text{ GeV}$ with $E_T^{\text{miss}} > 30 \text{ GeV}$ (ee and $e\mu$) or $E_T^{\text{miss}} > 20 \text{ GeV}$ ($\mu\mu$). We tighten the E_T^{miss} cut to 50 GeV and we also require that the scalar sum of the p_T of all jets with $p_T > 30 \text{ GeV}$ be $> 100 \text{ GeV}$. These requirements considerably reduce backgrounds to the $t\bar{t}$ sample, *e.g.*, backgrounds from Drell Yan and W +jets.
- The pre-selection consists mostly of $t\bar{t}$ events. We perform data – Monte Carlo comparisons of kinematical distributions. Assuming reasonable agreement for the bulk of $t\bar{t}$ we move on to a search for new physics in the tails of the $t\bar{t}$ E_T^{miss} and H_T distributions.
- Our prejudice is that new physics would manifest itself in an excess of events with high E_T^{miss} and significant hadronic activity. We define a priori search regions by tightening the E_T^{miss} and hadronic activity requirements.
- We perform a counting experiment in the signal regions. We compare observed yields with expectations from Monte Carlo and with three independent data driven techniques (see Sections 10.1, 10.3 and 10.4).

3 Datasets, Triggers, Luminosity

We combine the official May10ReReco and June 3rd prompt reco json files, yielding a luminosity of 349 pb^{-1} .

We use two data samples, one collected with high p_T dilepton triggers and the other with dilepton- H_T cross triggers. These samples are complementary, since the dilepton- H_T trigger sample extends to lower lepton p_T , while the high p_T dilepton trigger sample does not include requirements on the hadronic activity in the event.

Currently we use data reconstructed in CMSSW 4.2.X and Spring11 MC reconstructed in CMSSW 3.11.X. We will update the MC when the Summer11 madgraph MC samples become available.

- Datasets

- High p_T dilepton trigger sample

- * DoubleElectron_Run2011A-May10ReReco-v1_AOD
- * DoubleMu_Run2011A-May10ReReco-v1_AOD
- * MuEG_Run2011A-May10ReReco-v1_AOD
- * DoubleElectron_Run2011A-PromptReco-v4_AOD
- * DoubleMu_Run2011A-PromptReco-v4_AOD
- * MuEG_Run2011A-PromptReco-v4_AOD

- Dilepton- H_T cross trigger sample

- * ElectronHad_Run2011A-May10ReReco-v1_AOD
- * MuHad_Run2011A-May10ReReco-v1_AOD
- * ElectronHad_Run2011A-PromptReco-v4_AOD
- * MuHad_Run2011A-PromptReco-v4_AOD

- Monte Carlo samples

- TTJets_TuneZ2_7TeV-madgraph-tauola_Spring11-PU_S1_START311_V1G1-v1
- DYtoEE_M-10To20_TuneZ2_7TeV-pythia6_Spring11-PU_S1_START311_V1G1-v1
- DYtoMuMu_M-10To20_TuneZ2_7TeV-pythia6_Spring11-PU_S1_START311_V1G1-v1
- DYtoTauTau_M-10To20_CT10_TuneZ2_7TeV-powheg-pythia-tauola_Spring11-PU_S1_START311_V1G1-v2
- DYtoEE_M-20_CT10_TuneZ2_7TeV-powheg-pythia_Spring11-PU_S1_START311_V1G1-v1
- DYtoMuMu_M-20_CT10_TuneZ2_7TeV-powheg-pythia_Spring11-PU_S1_START311_V1G1-v1
- DYtoTauTau_M-20_CT10_TuneZ2_7TeV-powheg-pythia-tauola_Spring11-PU_S1_START311_V1G1-v1
- DYJetsToLL_TuneD6T_M-50_7TeV-madgraph-tauola_Spring11-PU_S1_START311_V1G1-v1
- WWTo2L2Nu_TuneZ2_7TeV-pythia6_Spring11-PU_S1_START311_V1G1-v1
- WZtoAnything_TuneZ2_7TeV-pythia6-tauola_Spring11-PU_S1_START311_V1G1-v1
- ZZtoAnything_TuneZ2_7TeV-pythia6-tauola_Spring11-PU_S1_START311_V1G1-v1
- WJetsToLNu_TuneZ2_7TeV-madgraph-tauola_Spring11-PU_S1_START311_V1G1-v1
- TToBLNu_TuneZ2_s-channel_7TeV-madgraph_Spring11-PU_S1_START311_V1G1-v1
- TToBLNu_TuneZ2_t-channel_7TeV-madgraph_Spring11-PU_S1_START311_V1G1-v1
- TToBLNu_TuneZ2_tW-channel_7TeV-madgraph_Spring11-PU_S1_START311_V1G1-v1

4 Event Preselection

The purpose of the preselection is to reject backgrounds other than $t\bar{t} \rightarrow$ dileptons. We compare the kinematical properties of this sample with expectations from $t\bar{t}$ Monte Carlo.

The preselection is based on the $t\bar{t}$ analysis [4]. We select events with two opposite sign, well-identified and isolated leptons (ee , $e\mu$, or $\mu\mu$); one of the leptons must have $p_T > 20$ GeV, the other one must have $p_T > 10$ GeV. Events with dilepton mass consistent with $Z \rightarrow ee/\mu\mu$ are rejected. In case of events with more than two such leptons, we select the pair that maximizes the scalar sum of lepton p_T 's. There must be at least two pfjets of $p_T > 30$ GeV and $|\eta| < 3.0$; jets must pass loose `pfJetId` and be separated by $\Delta R > 0.4$ from any lepton with $p_T > 10$ GeV passing the selection. The scalar sum H_T of the p_T of all such jets must exceed 100 GeV, for the dilepton- H_T sample this requirement is increased to 200 GeV since these triggers have large inefficiency below this threshold. Finally $E_T^{\text{miss}} > 50$ GeV (we use `pfmet`). More details are given in the subsections below.

4.1 Event Cleanup

- Require at least one good deterministic annealing (DA) vertex

- not fake
- $\text{ndof} > 4$
- $|\rho| < 2$ cm
- $|z| < 24$ cm.

4.2 Muon Selection

Muon candidates are RECO muon objects passing the following requirements:

- $p_T > 5 \text{ GeV}$ and $|\eta| < 2.4$
- Global Muon and Tracker Muon
- χ^2/ndof of global fit < 10
- At least 11 hits in the tracker fit
- Impact parameter with respect to the first DA vertex $d_0 < 200 \mu\text{m}$ and $d_z < 1 \text{ cm}$
- $Iso \equiv E_T^{\text{iso}}/p_T < 0.15$, E_T^{iso} is defined as the sum of transverse energy/momentum deposits in ecal, hcal, and tracker, in a cone of 0.3
- At least one of the hits from the standalone muon must be used in the global fit
- Require tracker $\Delta p_T/p_T < 0.1$. This cut was not in the original top analysis. It is motivated by the observation of poorly measured muons in data with large relative p_T uncertainty, giving significant contributions to the E_T^{miss}

4.3 Electron Selection

Electron candidates are RECO GSF electrons passing the following requirements:

- $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$.
- Veto electrons with a supercluster in the transition region $1.4442 < |\eta| < 1.556$.
- VBTF90 identification[5] with requirements tightened to match the CaloIdT and TrkIdVL HLT requirements:
 - $\sigma_{i\eta i\eta} < 0.01 \text{ (EB)}, 0.03 \text{ (EE)}$
 - $\Delta\phi < 0.15 \text{ (EB)}, 0.10 \text{ (EE)}$
 - $\Delta\eta < 0.007 \text{ (EB)}, 0.009 \text{ (EE)}$
 - $H/E < 0.1 \text{ (EB)}, 0.075 \text{ (EE)}$
- Impact parameter with respect to the first DA vertex $d_0 < 400 \mu\text{m}$ and $d_z < 1 \text{ cm}$.
- $Iso \equiv E_T^{\text{iso}}/p_T < 0.15$. E_T^{iso} is defined as the sum of transverse energy/momentum deposits in ecal, hcal, and tracker, in a cone of 0.3. A 1 GeV pedestal is subtracted from the ecal energy deposition in the EB, however the ecal energy is never allowed to go negative.
- Electrons with a tracker or global muon within ΔR of 0.1 are vetoed.
- The number of missing expected inner hits must be less than two [6].
- Conversion removal via partner track finding: any electron where an additional GeneralTrack is found with $\text{dist} < 0.02 \text{ cm}$ and $\Delta \cot \theta < 0.02$ is vetoed [6].

We estimate the contributions from fake leptons using the data-driven fake rate (FR) method. The requirements defining the fakeable objects are listed in App. A.

4.4 Invariant mass requirement

We remove e^+e^- and $\mu^+\mu^-$ events with invariant mass between 76 and 106 GeV. We also remove events with invariant mass $< 12 \text{ GeV}$, since this kinematical region is not well reproduced in CMS Monte Carlo and to remove events with Upsilon.

In addition, we remove $Z \rightarrow \mu\mu\gamma$ candidates with the γ collinear with one of the muons. This is done as follows: if the ecal energy associated with one of the muons is greater than 6 GeV, we add this energy to the momentum of the initial muon, and we recompute the $\mu\mu$ mass. If this mass is between 76 and 106 GeV, the event is rejected.

4.5 Trigger Selection

We do not make any requirements on HLT bits in the Monte Carlo. Instead, as discussed in Section 5, a trigger efficiency weight is applied to each event, based on the trigger efficiencies measured on data (see Sec. 5).

We select data events using the following triggers. An event in the ee channel is required to pass a DoubleElectron trigger, an event in the $\mu\mu$ channel is required to pass a DoubleMu trigger, and an event in the $e\mu$ channel is required to pass a Ele-Mu trigger.

- High p_T dilepton trigger sample

- HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL
- HLT_DoubleMu7
- HLT_Mu13_Mu7
- HLT_Mu17_Ele8_CaloIdL
- HLT_Mu8_Ele17_CaloIdL

- Lepton H_T cross trigger sample

- HLT_DoubleMu3_HT150
- HLT_DoubleMu3_HT160
- HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT150
- HLT_Mu3_Ele8_CaloIdT_TrkIdVL_HT150
- HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT160
- HLT_Mu3_Ele8_CaloIdT_TrkIdVL_HT160
- HLT_DoubleEle8_CaloIdL_TrkIdVL_HT150
- HLT_DoubleEle8_CaloIdT_TrkIdVL_HT150
- HLT_DoubleEle8_CaloIdL_TrkIdVL_HT160
- HLT_DoubleEle8_CaloIdT_TrkIdVL_HT160

5 Trigger efficiency

For the high p_T dilepton triggers, the efficiencies have been measured to be approximately 100% (DoubleEle), 90% (DoubleMu), and 95% (Mu-Ele) [7]. In the following, unless otherwise specified we weight the ee , $\mu\mu$ and $e\mu$ MC events by these efficiencies. We do not apply any efficiency correction for the hadronic part of the dilepton- H_T triggers. We have verified that the efficiencies for these triggers with respect to an offline selection of $H_T > 200$ GeV is high (~ 90 – 95%), as shown in Fig. 1.

6 Dilepton Yields

Table 1: The data and MC yields in the ee and $\mu\mu$ final states for events with 2 selected leptons with invariant mass 76–106 GeV.

Sample	ee	$\mu\mu$
data	110402	140191
MC	107089	125998
data/MC	1.03	1.11

The data and MC dilepton mass distributions for events with 2 selected leptons are displayed in Fig. 2. The yields of Z events in the mass range 76–106 GeV are indicated in Table 1. In data we observe a slight excess with respect to MC expectations, of 3% (11%) in the ee ($\mu\mu$) channel, which we attribute to uncertainties in trigger efficiency, lepton selection efficiency, and integrated luminosity. We use the ratio of $Z \rightarrow \mu^+\mu^-$ to $Z \rightarrow e^+e^-$ yields in data to estimate the ratio of muon to electron selection efficiencies, and find $R_{\mu e} = \text{eff}(\mu)/\text{eff}(e) = 1.13$.

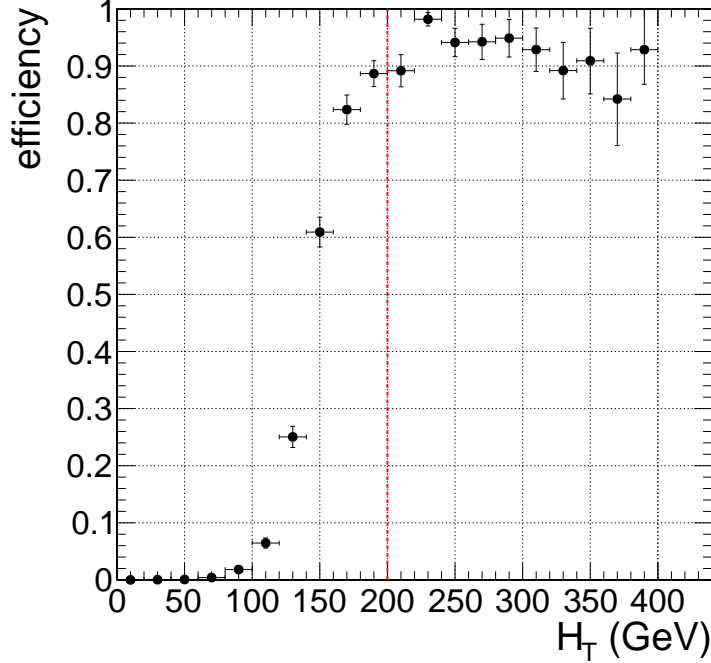


Figure 1: Efficiency for the dimuon- H_T trigger HLT_DoubleMu3_HT150 as a function of the offline H_T . Events are selected with the high p_T dilepton trigger HLT_DoubleMu7 and required to have 2 muons passing analysis selection. The vertical dashed line indicates the requirement $H_T > 200$ GeV, which is used in the preselection for the dilepton- H_T trigger sample.

7 Preselection yields: Data/MC Comparison

The data yields and the MC predictions for the dilepton trigger sample are given in Table 2. As anticipated, the MC predicts that the preselection is dominated by $t\bar{t}$. We observe a slight excess in data with respect to MC expectations. We also look for an excess of low lepton p_T events using the dilepton- H_T trigger sample, and requiring the leptons to pass lepton $p_T > (10,5)$ GeV but not pass lepton $p_T > (20,10)$ GeV (to remove overlap with the dilepton trigger sample), as summarized in Table 3. We observe reasonable agreement between the data and MC yields in this region. Finally, we compare in Table 4 the data samples collected with high p_T dilepton triggers and lepton- H_T cross triggers, including the requirements $H_T > 200$ GeV and lepton $p_T > (20,10)$ GeV. We find slightly fewer events in the dilepton- H_T trigger sample, consistent with the small inefficiency of the hadronic part of the trigger for an offline selection of $H_T > 200$ GeV.

The MC yields are normalized to 349 pb^{-1} using the cross-sections from Reference [8]. The MC is scaled by the approximate trigger efficiency (100% for ee , 95% for $\mu\mu$, and 95% for $e\mu$) and has been reweighted such that the distribution of reconstructed DA vertices matches that in data. The DY contribution is dominated by $\text{DY} \rightarrow \tau^+\tau^-$, and we have verified with the data-driven $R_{out/in}$ method [4] that the contributions from $\text{DY} \rightarrow e^+e^-$ and $\text{DY} \rightarrow \mu^+\mu^-$ are negligible. Also shown are the yields for LM1 and LM3, two of the LM points which are benchmarks for SUSY analyses at CMS. The LM yields are calculated at NLO using process-dependent k-factors computed from Prospino.

8 Properties of data passing the preselection

A number of kinematical distributions for events passing the preselection in data are compared with MC in Appendix C. Although we observe a slight overall excess of data, in general we find that the MC does a good job of reproducing the shapes of the kinematical distributions. Therefore we turn our attention to the tails of the $t\bar{t}$ events.

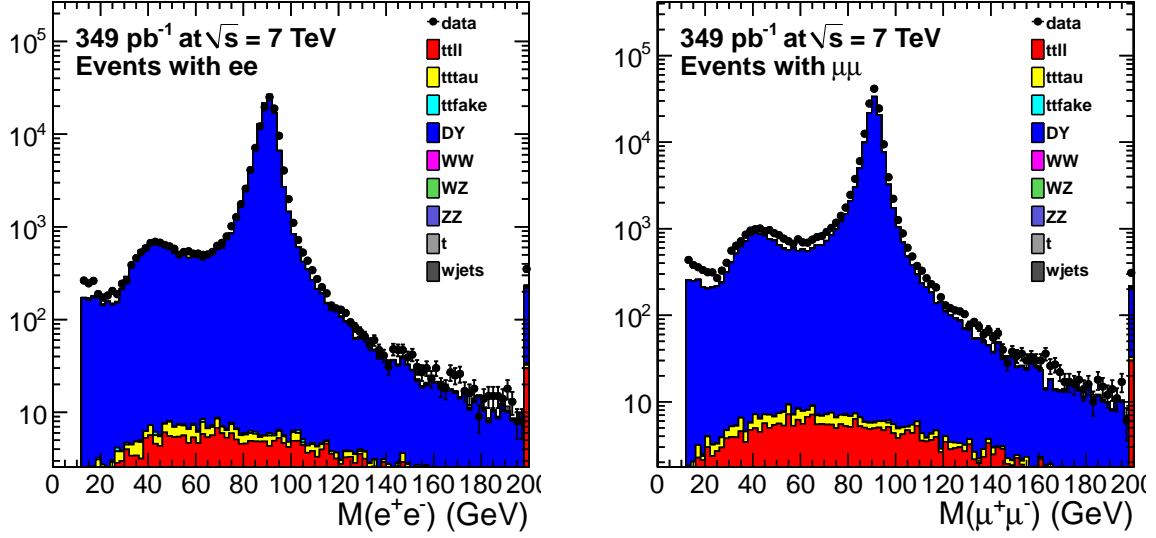


Figure 2: Distributions of dilepton mass in data and MC, in the ee channel (left) and $\mu\mu$ channel (right).

Table 2: High p_T dilepton trigger data and Monte Carlo yields for the preselection ($n_{\text{jets}} \geq 2$, $H_T > 100$ GeV, $E_T^{\text{miss}} > 50$ GeV, lepton $p_T > (20, 10)$ GeV).

Sample	ee	$\mu\mu$	$e\mu$	tot
$t\bar{t} \rightarrow \ell^+\ell^-$	120.1 ± 2.9	133.1 ± 2.9	320.3 ± 4.6	573.5 ± 6.1
$t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$	27.5 ± 1.4	33.3 ± 1.4	71.5 ± 2.2	132.3 ± 3.0
$t\bar{t} \rightarrow \text{fake}$	4.5 ± 0.6	1.3 ± 0.3	8.1 ± 0.7	13.9 ± 1.0
DY	6.6 ± 1.8	9.5 ± 2.1	13.4 ± 2.6	29.6 ± 3.8
W^+W^-	1.4 ± 0.2	1.5 ± 0.2	3.4 ± 0.2	6.3 ± 0.3
$W^\pm Z^0$	0.3 ± 0.0	0.4 ± 0.0	0.7 ± 0.1	1.3 ± 0.1
$Z^0 Z^0$	0.1 ± 0.0	0.1 ± 0.0	0.2 ± 0.0	0.4 ± 0.0
single top	4.5 ± 0.2	5.0 ± 0.2	11.9 ± 0.3	21.4 ± 0.5
$W + \text{jets}$	4.5 ± 1.9	0.0 ± 0.0	2.8 ± 1.7	7.3 ± 2.6
tot SM MC	169.7 ± 4.2	184.3 ± 3.9	432.2 ± 6.0	786.1 ± 8.3
data	193	201	485	879

9 Definition of the signal region

We define signal regions to look for possible new physics contributions in the opposite sign isolated dilepton sample. The choice of signal region is driven by three observations:

1. astrophysical evidence for dark matter suggests that we concentrate on the region of high E_T^{miss} ;
2. new physics signals should have high $\sqrt{\hat{s}}$;
3. observable high cross section new physics signals are likely to be produced strongly; thus, we expect significant hadronic activity in conjunction with the two leptons.

Following these observations, we define the following 3 signal regions by adding requirements of large hadronic activity and missing transverse energy to the preselection of Section 4.

- 2010 signal region: $H_T > 300$ GeV and $y > 8.5 \text{ GeV}^{1/2}$.
- high E_T^{miss} signal region: $H_T > 300$ GeV and $E_T^{\text{miss}} > 275$ GeV
- high H_T signal region: $H_T > 600$ GeV and $E_T^{\text{miss}} > 200$ GeV

Table 3: Dilepton- H_T trigger data and Monte Carlo yields for the preselection ($n_{\text{jets}} \geq 2$, $H_T > 200$ GeV, $E_T^{\text{miss}} > 50$ GeV, lepton $p_T > (10,5)$ GeV and not lepton $p_T > (20,10)$ GeV).

$t\bar{t} \rightarrow \ell^+ \ell^-$	0.6 ± 0.2	3.7 ± 0.5	4.1 ± 0.5	8.4 ± 0.7
$t\bar{t} \rightarrow \ell^\pm \tau^\mp / \tau^+ \tau^-$	0.6 ± 0.2	3.9 ± 0.5	3.5 ± 0.5	8.0 ± 0.7
$t\bar{t} \rightarrow \text{fake}$	0.1 ± 0.1	1.3 ± 0.3	1.5 ± 0.3	2.9 ± 0.4
DY	0.0 ± 0.0	3.7 ± 1.4	0.6 ± 0.6	4.3 ± 1.5
$W^+ W^-$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W^\pm Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Z^0 Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	0.0 ± 0.0	0.3 ± 0.1	0.3 ± 0.1	0.6 ± 0.1
$W + \text{jets}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	1.3 ± 0.3	12.9 ± 1.6	10.1 ± 1.0	24.3 ± 1.9
data	1	9	8	18

Table 4: Comparison of data yields in the high- p_T dilepton and dilepton- H_T trigger samples, passing the selection $n_{\text{jets}} \geq 2$, $E_T^{\text{miss}} > 50$ GeV, $H_T > 200$ GeV, lepton $p_T > (20,10)$ GeV.

Sample	ee	$\mu\mu$	$e\mu$	tot
high- p_T dilepton trigger	88	75	190	353
dilepton- H_T trigger	85	74	185	344

In our 2010 analysis, we cut on the quantity $y \equiv E_T^{\text{miss}} / \sqrt{H_T}$ rather than E_T^{miss} because the variables H_T and y are largely uncorrelated for the dominant $t\bar{t}$ background. This allows us to use a data-driven ABCD method to estimate the background (see Section 10.1). Here, we include 2 additional tighter signal regions using requirements on E_T^{miss} and H_T , since we observe that E_T^{miss} is a better discriminant between $t\bar{t}$ vs. SUSY. We have developed a novel technique, which is a variation of the ABCD method, to estimate the background in a signal region defined by E_T^{miss} and H_T requirements (see App. B).

The 2010 signal region is the same as the one used in the 2010 analysis, and was chosen to preserve about 1% of the $t\bar{t}$ sample. The additional signal regions (high E_T^{miss} and high H_T) have tightened requirements on E_T^{miss} and H_T , respectively, which reduce the expected background by roughly an order of magnitude.

For each signal region, we search in the high p_T dilepton trigger sample requiring lepton $p_T > (20,10)$ GeV, and separately in the dilepton- H_T trigger sample requiring lepton $p_T > (10,5)$ GeV and not passing lepton $p_T > (20,10)$ GeV, to remove overlap with the high p_T dilepton trigger sample. For the high p_T dilepton sample, for which we have prior experience with the 2010 analysis, we apply the ABCD, $p_T(\ell\ell)$, and OF subtraction background estimates. For the dilepton- H_T sample we currently apply only the OF subtraction background estimate.

We present the data and MC expected yields in the 3 signal regions in Tables 5-7 for the high p_T dilepton trigger sample and in Tables 8-10.

Table 5: High p_T dilepton trigger data and MC yields in the 2010 signal region. MC errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
$t\bar{t} \rightarrow \ell^+\ell^-$	1.6 ± 0.3	2.5 ± 0.4	4.5 ± 0.5	8.5 ± 0.7
$t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$	0.7 ± 0.2	0.7 ± 0.2	2.0 ± 0.4	3.4 ± 0.5
$t\bar{t} \rightarrow \text{fake}$	0.2 ± 0.1	0.0 ± 0.0	0.3 ± 0.1	0.4 ± 0.2
DY	0.3 ± 0.3	0.7 ± 0.7	0.7 ± 0.7	1.7 ± 1.0
W^+W^-	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.3 ± 0.1
$W^\pm Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Z^0 Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0
$W + \text{jets}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	2.9 ± 0.5	4.0 ± 0.8	7.7 ± 1.0	14.5 ± 1.4
data	4	4	11	19
LM1	11.0 ± 0.4	13.0 ± 0.4	6.7 ± 0.3	30.7 ± 0.7
LM3	2.1 ± 0.1	2.5 ± 0.1	3.5 ± 0.2	8.1 ± 0.2

Table 6: High p_T dilepton trigger data and MC yields in the high E_T^{miss} signal region. MC errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
$t\bar{t} \rightarrow \ell^+\ell^-$	0.4 ± 0.2	0.4 ± 0.1	0.8 ± 0.2	1.6 ± 0.3
$t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$	0.1 ± 0.1	0.0 ± 0.0	0.1 ± 0.1	0.3 ± 0.1
$t\bar{t} \rightarrow \text{fake}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
DY	0.0 ± 0.0	0.0 ± 0.0	0.7 ± 0.7	0.7 ± 0.7
W^+W^-	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0
$W^\pm Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Z^0 Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W + \text{jets}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	0.6 ± 0.2	0.4 ± 0.1	1.6 ± 0.8	2.6 ± 0.8
data	2	0	2	4
LM1	6.2 ± 0.3	7.3 ± 0.3	3.9 ± 0.2	17.4 ± 0.5
LM3	1.2 ± 0.1	1.4 ± 0.1	1.8 ± 0.1	4.3 ± 0.2

Table 7: High p_T dilepton trigger data and MC yields in the high H_T signal region. MC errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
$t\bar{t} \rightarrow \ell^+\ell^-$	0.2 ± 0.1	0.3 ± 0.1	0.7 ± 0.2	1.1 ± 0.3
$t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$	0.1 ± 0.1	0.2 ± 0.1	0.4 ± 0.2	0.7 ± 0.2
$t\bar{t} \rightarrow \text{fake}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
DY	0.0 ± 0.0	0.0 ± 0.0	0.7 ± 0.7	0.7 ± 0.7
W^+W^-	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W^\pm Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Z^0 Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W + \text{jets}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	0.3 ± 0.1	0.5 ± 0.2	1.8 ± 0.8	2.5 ± 0.8
data	1	0	2	3
LM1	5.3 ± 0.3	5.7 ± 0.3	2.5 ± 0.2	13.5 ± 0.4
LM3	1.2 ± 0.1	1.4 ± 0.1	1.7 ± 0.1	4.3 ± 0.2

Table 8: Dilepton- H_T trigger data and MC yields in the 2010 signal region. MC errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
$t\bar{t} \rightarrow \ell^+\ell^-$	0.0 ± 0.0	0.2 ± 0.1	0.4 ± 0.2	0.5 ± 0.2
$t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$	0.0 ± 0.0	0.2 ± 0.1	0.1 ± 0.1	0.3 ± 0.1
$t\bar{t} \rightarrow \text{fake}$	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.1 ± 0.1
DY	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
W^+W^-	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W^\pm Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Z^0 Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W + \text{jets}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	0.0 ± 0.0	0.4 ± 0.1	0.6 ± 0.2	1.0 ± 0.2
data	0	1	0	1
LM1	0.3 ± 0.1	2.7 ± 0.2	2.1 ± 0.2	5.1 ± 0.3
LM3	0.0 ± 0.0	0.3 ± 0.0	0.2 ± 0.0	0.5 ± 0.1

Table 9: Dilepton- H_T trigger data and MC yields in the high y signal region. MC errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
$t\bar{t} \rightarrow \ell^+\ell^-$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.1 ± 0.1
$t\bar{t} \rightarrow \text{fake}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
DY	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
W^+W^-	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W^\pm Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Z^0 Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W + \text{jets}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.1 ± 0.1
data	0	0	0	0
LM1	0.2 ± 0.0	1.6 ± 0.1	1.3 ± 0.1	3.0 ± 0.2
LM3	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.3 ± 0.0

Table 10: Dilepton- H_T trigger data and MC yields in the high H_T signal region. MC errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
$t\bar{t} \rightarrow \ell^+\ell^-$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$t\bar{t} \rightarrow \text{fake}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
DY	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
W^+W^-	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W^\pm Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Z^0 Z^0$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$W + \text{jets}$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
data	0	0	0	0
LM1	0.2 ± 0.1	1.3 ± 0.1	0.7 ± 0.1	2.2 ± 0.2
LM3	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.3 ± 0.0

These results are summarized as:

- High p_T dilepton trigger sample

- 2010 signal region ($y > 8.5 \text{ GeV}^{1/2}$, $H_T > 300 \text{ GeV}$)
 - * observed yield : 19
 - * MC prediction : 14.5 ± 1.4
- high E_T^{miss} signal region ($E_T^{\text{miss}} > 275 \text{ GeV}$, $H_T > 300 \text{ GeV}$)
 - * observed yield : 4
 - * MC prediction : 2.6 ± 0.8
- high H_T signal region ($E_T^{\text{miss}} > 200 \text{ GeV}$, $H_T > 600 \text{ GeV}$)
 - * observed yield : 3
 - * MC prediction : 2.5 ± 0.8

- Dilepton- H_T trigger sample

- 2010 signal region ($y > 8.5 \text{ GeV}^{1/2}$, $H_T > 300 \text{ GeV}$)
 - * observed yield : 1
 - * MC prediction : 1.0 ± 0.2
- high E_T^{miss} signal region ($E_T^{\text{miss}} > 275 \text{ GeV}$, $H_T > 300 \text{ GeV}$)
 - * observed yield : 0
 - * MC prediction : 0.1 ± 0.1
- high H_T signal region ($E_T^{\text{miss}} > 200 \text{ GeV}$, $H_T > 600 \text{ GeV}$)
 - * observed yield : 0
 - * MC prediction : 0.0 ± 0.0

For all signal regions, we observe reasonable agreement between data and MC expectations. A data/MC comparison of the kinematic distributions of the 19 events passing the 2010 signal region selection is presented in App. D.

For the 2010 signal region, we observe 1 ee and 2 $e\mu$ events in the Z mass window, confirming the expectation from MC that the DY background is negligible. We also confirm the MC expectation that the contribution of fake leptons is small, using the data-driven fake rate method. The results are summarized as:

- High p_T dilepton trigger sample

- 2010 signal region ($y > 8.5 \text{ GeV}^{1/2}$, $H_T > 300 \text{ GeV}$)
 - * single fakes : 0.6 ± 0.6
 - * double fakes : 0.0 ± 0.0
- high E_T^{miss} signal region ($E_T^{\text{miss}} > 275 \text{ GeV}$, $H_T > 300 \text{ GeV}$)
 - * single fakes : 0.0 ± 0.0
 - * double fakes : 0.0 ± 0.0
- high H_T signal region ($E_T^{\text{miss}} > 200 \text{ GeV}$, $H_T > 600 \text{ GeV}$)
 - * single fakes : 0.5 ± 0.5
 - * double fakes : 0.0 ± 0.0

- Dilepton- H_T trigger sample

- 2010 signal region ($y > 8.5 \text{ GeV}^{1/2}$, $H_T > 300 \text{ GeV}$)
 - * single fakes : 0.2 ± 0.2
 - * double fakes : 0.1 ± 0.1
- high E_T^{miss} signal region ($E_T^{\text{miss}} > 275 \text{ GeV}$, $H_T > 300 \text{ GeV}$)

256 * single fakes : 0.0 ± 0.0
 257 * double fakes : 0.0 ± 0.0
 258 – high H_T signal region ($E_T^{\text{miss}} > 200$ GeV, $H_T > 600$ GeV)
 259 * single fakes : 0.0 ± 0.0
 260 * double fakes : 0.0 ± 0.0

261 Based on these results, we treat the contributions from fakes leptons in our signal regions as negligible.

262 10 Data Driven Background Estimation Methods

263 For the high p_T dilepton trigger sample, we use 3 data-driven methods to estimate the background in the signal
 264 region. The first one exploits the fact that H_T and y are nearly uncorrelated for the $t\bar{t}$ background. When predicting
 265 the backgrounds in the signal regions defined in the E_T^{miss} vs. H_T plane, we use a variation of the ABCD method
 266 which we refer to as the ABCD' method (see Appendix B). The second method is based on the fact that in $t\bar{t}$ the
 267 p_T of the dilepton pair is on average nearly the same as the p_T of the pair of neutrinos from W -decays, which
 268 is reconstructed as E_T^{miss} in the detector. The third method exploits the fact that in $t\bar{t}$ events the rates of same-
 269 flavor vs. opposite-flavor dilepton events are the same. For the low lepton p_T events collected by the dilepton- H_T
 270 triggers, we use the opposite-flavor subtraction technique.

271 We study the closure of these methods using our madgraph $t\bar{t}$ sample, as well as the powheg sample
 272 TTTTo2L2Nu2B_7TeV-powheg-pythia6_Spring11-PU_S1_START311_V1G1-v1 which has approximately 10 times
 273 more events in the dilepton channel than the madgraph sample. We use these samples to estimate correction factors
 274 and systematic uncertainties for the background predictions. However, the final choice of correction factors and
 275 uncertainties will be extracted from the Summer11 $t\bar{t}$ madgraph sample which will have 50 times as many events
 276 as the current madgraph sample. For the studies presented in this section, we do not apply trigger efficiency
 277 corrections or reweighting for number of reconstructed vertices since are not comparing MC to data.

278 10.1 ABCD method

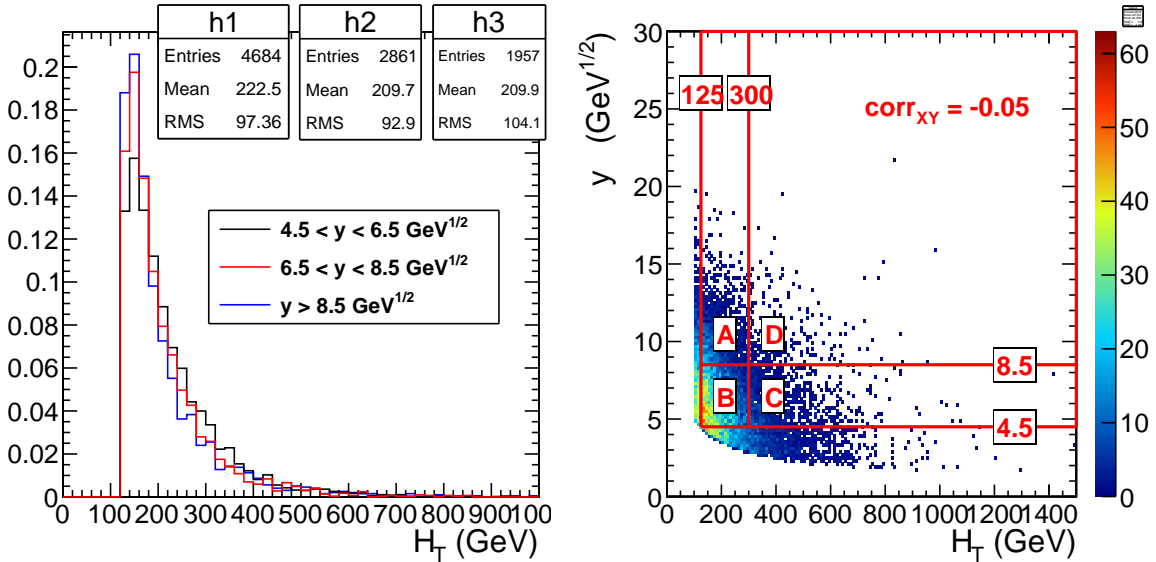


Figure 3: Left: distributions of H_T in MC $t\bar{t}$ events for different intervals of y . h1, h2, and h3 refer to the y intervals 4.5-6.5, 6.5-8.5 and >8.5 , respectively. Right: Distributions of y vs. H_T for $t\bar{t}$ MC. Here we also show our choice of ABCD regions. The correlation coefficient corr_{xy} is computed for events falling in the ABCD regions.

279 We find that in $t\bar{t}$ events H_T and y are nearly uncorrelated, as demonstrated in Fig. 3 (left). Thus, we can use an
 280 ABCD method in the y vs. H_T plane to estimate the background in a data driven way. We define 4 regions in the
 281 plane of y vs. H_T , as shown in Fig. 3 (right). The region D is the signal region, and the regions A, B and C are
 282 control regions. The predicted background in region D is given by $N_A \times N_C / N_B$.

In Table 11, we quote the $t\bar{t}$ MC expected results for 1 fb^{-1} . We find that the prediction agrees with the observed yield in the signal region within $\sim 20\%$, and we assign a corresponding systematic uncertainty. We also study the dependence of the ratio of observed to predicted signal yields, as a function of the y and H_T requirements used to define the signal region, shown in Fig. 4.

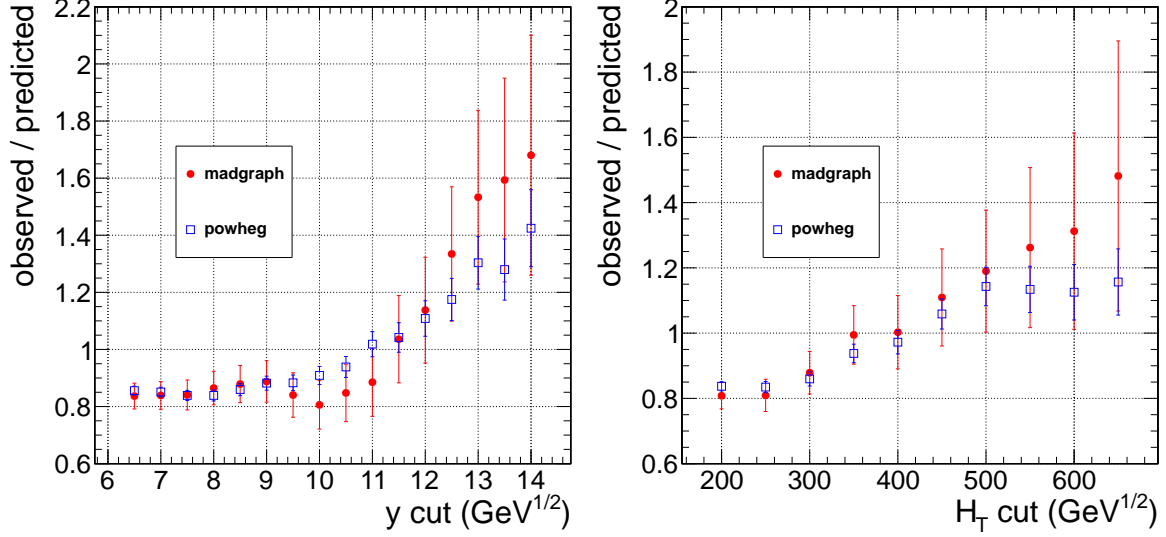


Figure 4: Variation of observed/predicted for the ABCD method as a function of the y and H_T cuts defining the signal region.

Table 11: Expected yields from $t\bar{t}$ MC in 1 fb^{-1} in the four ABCD regions for the signal region depicted in Figs. 7, as well as the predicted yield in region D given by $N_A \times N_C/N_B$ and the ratio of the observed signal yield to the prediction. The quoted uncertainties are statistical only.

signal region	sample	A	B	C	D	$N_A \times N_C/N_B$	obs/pred
2010 signal region	madgraph	251.3 ± 6.1	951.5 ± 11.9	165.2 ± 4.9	38.3 ± 2.4	43.6 ± 1.8	0.88 ± 0.07
	powheg	231.7 ± 2.0	850.6 ± 3.7	157.8 ± 1.6	37.0 ± 0.8	43.0 ± 0.6	0.86 ± 0.02

10.2 ABCD' Method

To estimate the background in a signal region defined in the E_T^{miss} vs. H_T plane, we perform a variant of the ABCD method which we refer to as the ABCD' method. This method is discussed in detail in App. B. In Table 12 we quote the expected results in 1 fb^{-1} from $t\bar{t}$ MC. Based on these results, we apply the scale factors and systematic uncertainties summarized in Table 15 to the background prediction of the ABCD' method.

Table 12: Expected results from $t\bar{t}$ MC in 1 fb^{-1} for the ABCD' method. The quoted uncertainties are statistical only.

signal region	sample	predicted	observed	obs/pred
2010 signal region	madgraph	44.0 ± 1.5	38.3 ± 2.4	0.87 ± 0.06
	powheg	43.3 ± 0.5	37.0 ± 0.8	0.85 ± 0.02
high E_T^{miss} signal region	madgraph	3.1 ± 0.4	4.6 ± 0.8	1.47 ± 0.31
	powheg	3.5 ± 0.1	4.5 ± 0.3	1.30 ± 0.10
high H_T signal region	madgraph	5.3 ± 0.8	5.5 ± 0.9	1.03 ± 0.23
	powheg	5.9 ± 0.3	6.2 ± 0.3	1.05 ± 0.07

10.3 Dilepton p_T method

This method is based on a suggestion by V. Pavlunin [9], and was investigated by our group in 2009 [10] and in our 2010 analysis [3]. The idea is that in dilepton $t\bar{t}$ events the lepton and neutrinos from W decays have the same p_T spectrum (modulo W polarization effects). One can then use the observed $p_T(\ell\ell)$ distribution to model the sum of neutrino p_T 's which is identified with the E_T^{miss} .

Then, in order to predict the $t\bar{t} \rightarrow$ dilepton contribution to a selection with $E_T^{\text{miss}} + X$, one applies a cut on $p_T(\ell\ell) + X$ instead. In practice one has to rescale the result of the $p_T(\ell\ell) + X$ selection to account for the fact that any dilepton selection must include a moderate E_T^{miss} cut in order to reduce Drell Yan backgrounds. This is discussed in Section 5.3 of Reference [10]; for a E_T^{miss} cut of 50 GeV, the rescaling factor is obtained from the MC as

$$K = \frac{\int_0^\infty \mathcal{N}(p_T(\ell\ell)) dp_T(\ell\ell)}{\int_{50}^\infty \mathcal{N}(p_T(\ell\ell)) dp_T(\ell\ell)} = 1.5.$$

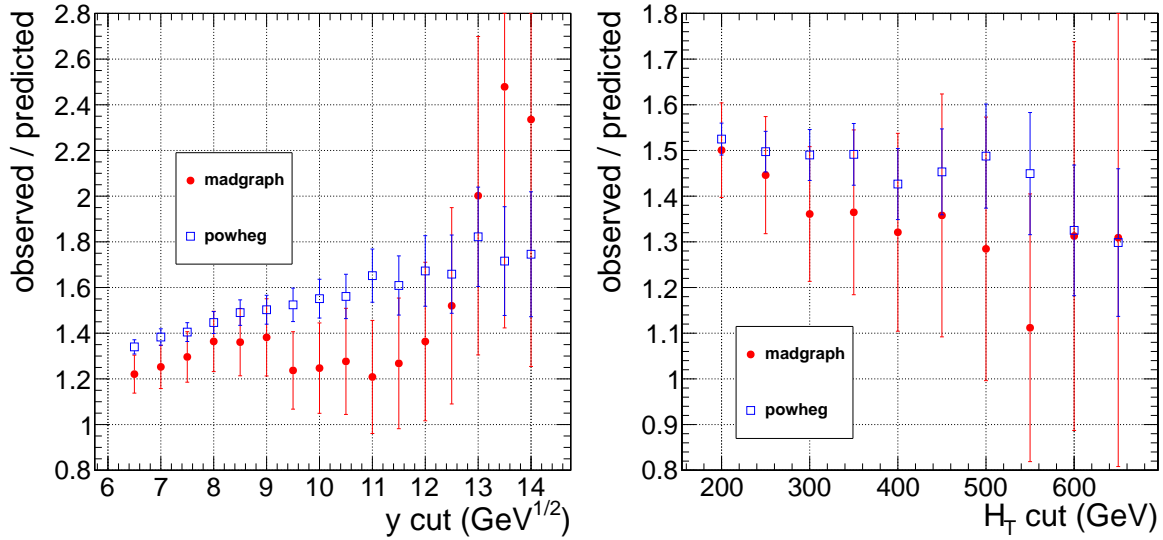


Figure 5: Variation of observed/predicted for the $p_T(\ell\ell)$ method as a function of the y and H_T cuts defining the signal region.

We summarize the expected results of the $p_T(\ell\ell)$ method in 1 fb $^{-1}$ $t\bar{t}$ MC in Table 13, and we show the dependence of observed/predicted vs. the signal region requirements in Fig. 5. Based on these results, we apply the scale factors and uncertainties summarized in Table 15 to the predicted background from the $p_T(\ell\ell)$ method. In [2], we have studied extensively the origin of the excess of observed vs. predicted events from this method. We found that it is due mostly to the W polarization, which results in a harder p_T distribution for the W neutrinos than charged leptons.

Table 13: Expected observed and predicted yields in 1 fb $^{-1}$ for $t\bar{t}$ MC for the $p_T(\ell\ell)$ method applied to the 3 signal regions, and the ratio of the observed signal yield to the prediction. The quoted uncertainties are statistical only.

signal region	sample	predicted	observed	obs/pred
2010 signal region	madgraph	28.2 ± 2.5	38.3 ± 2.4	1.36 ± 0.15
	powheg	24.8 ± 0.8	37.0 ± 0.8	1.49 ± 0.06
high E_T^{miss} signal region	madgraph	2.7 ± 0.8	4.6 ± 0.8	1.72 ± 0.59
	powheg	3.3 ± 0.3	4.5 ± 0.3	1.39 ± 0.15
high H_T signal region	madgraph	4.7 ± 1.0	5.5 ± 0.9	1.17 ± 0.31
	powheg	4.7 ± 0.3	6.2 ± 0.3	1.33 ± 0.11

As a validation of the $p_T(\ell\ell)$ method, we also apply the background estimate to the H_T sideband region 125–300 GeV, where we expect the events to be dominated by background and where the higher event yields allow us to test the method with higher statistical precision. The results of the $p_T(\ell\ell)$ method applied to $t\bar{t}$ MC in the H_T sideband region with the requirements corresponding to the 2010 signal region ($y > 8.5$ GeV $^{1/2}$) and the high E_T^{miss} region

Table 14: Expected observed and predicted yields in 1 fb^{-1} for $t\bar{t}$ MC for the $p_T(\ell\ell)$ method for the control regions in the H_T sideband region $125 < H_T < 300 \text{ GeV}$, and the ratio of the observed signal yield to the prediction. The quoted uncertainties are statistical only, assuming Gaussian errors.

control region	sample	predicted	observed	obs/pred
$125 < H_T < 300 \text{ GeV}, y > 8.5 \text{ GeV}^{1/2}$	madgraph	35.8 ± 1.3	51.3 ± 1.2	1.43 ± 0.06
	powheg	33.4 ± 0.4	47.3 ± 0.4	1.42 ± 0.02
$125 < H_T < 300 \text{ GeV}, E_T^{\text{miss}} > 200 \text{ GeV}^{1/2}$	madgraph	6.4 ± 1.3	9.9 ± 1.2	1.54 ± 0.36
	powheg	6.6 ± 0.4	8.4 ± 0.4	1.27 ± 0.10

($E_T^{\text{miss}} > 200 \text{ GeV}$) are given in Table 14. Based on these results we will apply a correction factor $K_C = 1.4 \pm 0.1$ ($K_C = 1.3 \pm 0.2$) to data for the $y > 8.5 \text{ GeV}^{1/2}$ ($E_T^{\text{miss}} > 200 \text{ GeV}$) requirements in the H_T sideband region.

Table 15: Summary of correction factors and systematic uncertainties for the ABCD' and $p_T(\ell\ell)$ methods in the 3 signal regions.

signal region	ABCD'	$p_T(\ell\ell)$
2010 signal region	1.0 ± 0.2	1.4 ± 0.2
high E_T^{miss} signal region	1.3 ± 0.2	1.5 ± 0.3
high H_T signal region	1.0 ± 0.2	1.3 ± 0.2

10.4 Opposite-Flavor Subtraction

The opposite-flavor subtraction technique exploits the fact that in $t\bar{t}$, the flavor of the 2 leptons from W decay are uncorrelated. Hence we expect equal rates of same-flavor (SF) ee or $\mu\mu$ vs. opposite-flavor (OF) $e\mu$ lepton pairs. In SUSY, the lepton flavors may be correlated, producing an excess of SF over OF events. We use the observed yield in the OF final state to predict the yields in the SF final state according to:

$$N(ee) = \frac{1}{2R_{\mu e}} N(e\mu) \text{ and } N(\mu\mu) = \frac{R_{\mu e}}{2} N(e\mu)$$

where $R_{\mu e}$ is the ratio of muon to electron selection efficiencies. This quantity is evaluated by taking the ratio of the number of observed $Z \rightarrow \mu^+\mu^-$ to $Z \rightarrow e^+e^-$ events, in the mass range 76-106 GeV with no jets or E_T^{miss} requirements (see Fig. 2). Alternatively, we can quantify the excess of SF vs. OF events with the quantity:

$$\Delta = R_{\mu e} N(ee) + \frac{1}{R_{\mu e}} N(\mu\mu) - N(e\mu), \quad (1)$$

which is predicted to be 0 for processes with uncorrelated lepton flavors. In order for this technique to work, the kinematic selection applied to events in all dilepton flavor channels must be the same, which is not the case for our default selection because the Z mass veto is applied only to same-flavor channels. Therefore when applying the OF subtraction technique we also apply the Z mass veto also to the $e\mu$ channel.

We will apply this technique to both the high p_T dilepton trigger and dilepton- H_T trigger data samples. In the following, we first apply the technique to $t\bar{t}$ MC with high p_T leptons, and then to $t\bar{t}$ MC with low p_T leptons.

10.4.1 OF subtraction: Application to high p_T lepton sample

We begin by applying the OF subtraction technique to $t\bar{t}$ MC with leptons passing $p_T > (20,10) \text{ GeV}$. Here we extract $R_{e\mu}$ by taking the ratio of $Z \rightarrow \mu^+\mu^-$ vs. $Z \rightarrow e^+e^-$ events in the window 76–106 GeV in DY MC. For data, we have verified that the contribution from fake leptons in the high lepton p_T signal regions is small using the data-driven fake rate method, hence we do not need to correct for fake leptons here. The results are summarized in Table 16, where we find values of Δ consistent with 0, as expected.

10.4.2 OF subtraction: Application to low p_T lepton sample

In this section, we apply the OF subtraction technique to $t\bar{t}$ MC with leptons passing $p_T > (10,5) \text{ GeV}$ but not passing $p_T > (20,10) \text{ GeV}$ (in order to remove overlap with the high lepton p_T sample). The OF subtraction in the low lepton p_T regime is complicated by 2 factors:

Table 16: Expected yields in $1 \text{ fb}^{-1} t\bar{t}$ MC for the OF subtraction method, and the quantity Δ , defined in Eq. 1. The quoted systematic uncertainty refers to that of $R_{\mu e}$.

region	sample	$N(ee)$	$N(\mu\mu)$	$N(e\mu)$	Δ
preselection region	madgraph	431.1 ± 8.0	531.3 ± 8.9	945.8 ± 11.8	$11.8 \pm 16.8 \text{ (stat)} \pm 1.1 \text{ (syst)}$
	powheg	383.1 ± 2.5	492.9 ± 2.9	876.2 ± 3.8	$-7.0 \pm 5.4 \text{ (stat)} \pm 0.8 \text{ (syst)}$
2010 signal region	madgraph	7.4 ± 1.0	10.7 ± 1.3	14.5 ± 1.5	$3.3 \pm 2.2 \text{ (stat)} \pm 0.04 \text{ (syst)}$
	powheg	7.2 ± 0.3	8.6 ± 0.4	16.8 ± 0.5	$-1.1 \pm 0.7 \text{ (stat)} \pm 0.03 \text{ (syst)}$

- The ratio of muon to electron selection efficiencies $R_{\mu e}$ increases significantly at low p_T , due to a drop in the electron selection efficiency.
- We reconstruct muons down to $p_T > 5 \text{ GeV}$ but electrons only to $p_T > 10 \text{ GeV}$.

Our strategy is the following:

- Evaluate $R_{\mu e}$ from $t\bar{t}$ MC.
- Parameterize $R_{\mu e}$ as a function of lepton p_T . For now we split in 2 bins, $10 < p_T < 20 \text{ GeV}$ and $p_T > 20 \text{ GeV}$.
- For data, we will apply to $R_{\mu e}$ a trigger efficiency correction and subtract the expected contribution from fake leptons from the data-driven fake rate method (but neither correction is performed for the MC studies in this section).
- We first apply the OF subtraction to the preselection region, and then to the signal region.

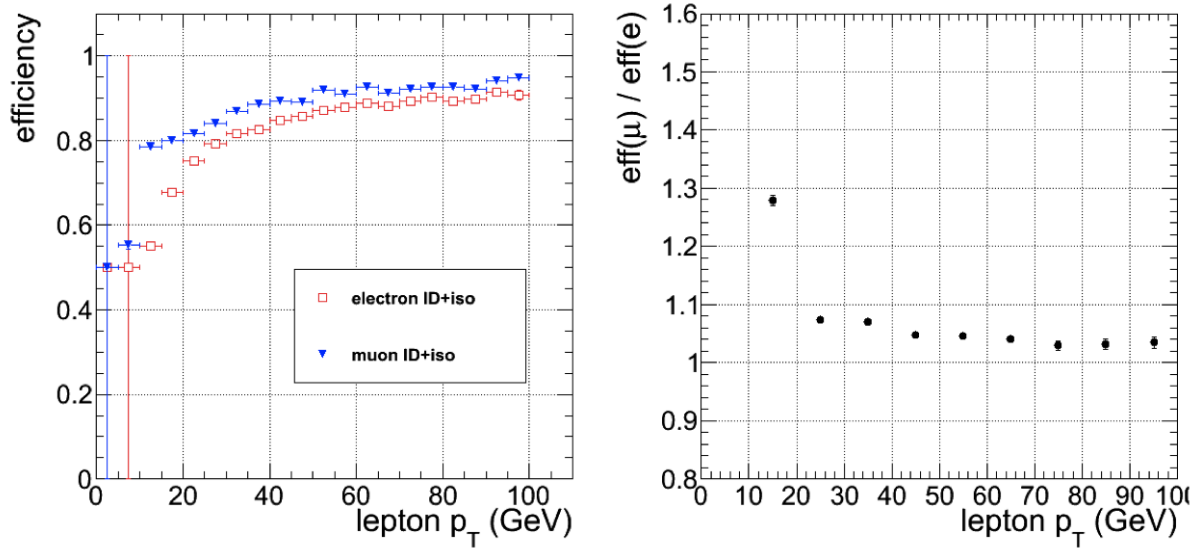


Figure 6: Left: the electron and muon selection efficiencies, as a function of lepton p_T , extracted from $t\bar{t}$ MC. Right: the ratio $R_{\mu e}$ of muon to electron selection efficiencies as a function of lepton p_T .

We begin by examining the dependence of $R_{\mu e}$ on lepton p_T , as shown in Fig. 6. We find that $R_{\mu e}$ increases in the region p_T 10–20 GeV, but is roughly constant for $p_T > 20 \text{ GeV}$. Hence we take $R_{\mu e}^{10-20} = 1.28$ for p_T 10–20 GeV and $R_{\mu e}^{>20} = 1.08$ for $p_T > 20 \text{ GeV}$, and assign a 5% systematic uncertainty.

Next, to take into account the fact the electrons and muons have a different p_T range, we split the low p_T sample into events with leptons passing $p_T > (10,10) \text{ GeV}$, denoted in the following as (10,10), and events with leptons passing $p_T > (10,5) \text{ GeV}$ but not passing $p_T > (10,10) \text{ GeV}$, denoted in the following as (10,5). We find the following relations, expected for backgrounds with uncorrelated lepton flavors:

- $N(ee)(10, 10) = 1/(2R_{\mu e}^{10-20})N(e\mu)(10, 10)$
- $N(\mu\mu)(10, 10) = (R_{\mu e}^{10-20}/2)N(e\mu)(10, 10)$
- $N(\mu\mu)(10, 5) = R_{\mu e}^{>20}N(e\mu)(10, 5)$

Note that for (10,10) events, both leptons are in the range p_T 10–20 GeV, hence the relevant efficiency ratio is $R_{\mu e}^{10-20}$. For (10,5) events, both ee and $\mu\mu$ events have a muon with p_T 5–10 GeV and an additional lepton with $p_T > 10$ GeV. In most cases the leading lepton has $p_T > 20$ GeV, hence we use $R_{\mu e}^{>20}$ for these events.

In this case we find the following expression for Δ , quantifying the excess of SF vs. OF yields:

$$\Delta = R_{\mu e}^{10-20}N(ee)(10, 10) + 1/R_{\mu e}^{10-20}N(\mu\mu)(10, 10) + 1/R_{\mu e}^{>20}N(\mu\mu)(10, 5) - N(e\mu)(10, 10) - N(e\mu)(10, 5) \quad (2)$$

In Table 17 we apply this technique to $t\bar{t}$ MC. As expected, we find Δ consistent with 0.

Table 17: Expected yields in 1 fb⁻¹ $t\bar{t}$ MC for the OF subtraction method in the low lepton p_T regime, and the quantity Δ , defined in Eq. 2. The quoted systematic uncertainty refers to that of $R_{\mu e}$.

region	sample	$N(ee)(10, 10)$	$N(\mu\mu)(10, 10)$	$N(e\mu)(10, 10)$	$N(\mu\mu)(10, 5)$	$N(e\mu)(10, 5)$	Δ
preselection region	madgraph	3.4 ± 0.7	4.9 ± 0.9	6.4 ± 1.0	22.9 ± 1.8	18.2 ± 1.6	4.8 ± 2.8 (stat) ± 1.0 (syst)
	powheg	2.5 ± 0.2	4.4 ± 0.3	7.0 ± 0.3	19.0 ± 0.6	16.6 ± 0.5	0.6 ± 0.9 (stat) ± 0.9 (syst)
2010 signal region	madgraph	0.0 ± 0.0	0.4 ± 0.3	0.6 ± 0.3	1.0 ± 0.4	0.7 ± 0.3	-0.0 ± 0.6 (stat) ± 0.1 (syst)
	powheg	0.1 ± 0.0	0.2 ± 0.1	0.4 ± 0.1	1.4 ± 0.2	1.1 ± 0.1	0.1 ± 0.2 (stat) ± 0.1 (syst)

11 Results

11.1 Background estimate from the ABCD method

We begin by applying the ABCD method to estimate the background in the 2010 signal region. The data yields in the four regions are summarized in Tables 19 and the y vs. H_T distributions are displayed in Fig. 7. The ABCD background prediction is $N_A \times N_C/N_B = 12.7 \pm 2.4$ (stat), in agreement with the MC expectation. We also apply the ABCD’ method to estimate the background in this region, following the procedure of App. B, and find a predicted background of 12.8 ± 2.9 (stat), in good agreement with the ABCD prediction.

Next, we apply the ABCD’ method to predict the yields in the high E_T^{miss} and high H_T signal regions. The y vs. H_T distributions for data are displayed in Fig. 8. The signal regions are indicated, as well as the control regions used to measure the $f(y)$ and $g(H_T)$ distributions. For the high E_T^{miss} signal region, we find a predicted yield of 1.0 ± 0.3 (stat), in reasonable agreement with the MC prediction. For the high H_T signal region, we do not find any events in the control region used to extract $g(H_T)$ with $H_T > 600$ GeV, and the ABCD’ background estimate is therefore 0. To assess the statistical uncertainty in this prediction, we add a single event “by hand” to the $g(H_T)$ distribution at $H_T = 600$ GeV, leading to a predicted yield of 0.6. **NEED TO THINK ABOUT THIS. THE PREDICTION DEPENDS ON WHERE IN HT YOU PUT THIS SINGLE EVENT. FOR EXAMPLE, PUTTING IT AT 1000 GIVES A PREDICTION OF 1.2 HOPEFULLY WITH 3X MORE STATS WE WON’T BE IN THIS SITUATION** These results are summarized in Table 18.

Table 18: Summary of results of the ABCD’ method, applied to the 3 signal regions. The correction factors are given in Table 15.

Signal Region	ABCD’ pred	correction factor	prediction
2010 signal region	12.8 ± 2.9	1.0 ± 0.2	12.8 ± 2.9 (stat) ± 2.6 (syst)
high E_T^{miss} signal region	1.0 ± 0.3	1.3 ± 0.2	1.3 ± 0.4 (stat) ± 0.2 (syst)
high H_T signal region	0.0 ± 0.6	1.0 ± 0.2	0.0 ± 0.6 (stat) ± 0.1 (syst)

11.2 Background estimate from the $P_T(\ell\ell)$ method

We begin by extracting the value of the E_T^{miss} acceptance scaling factor K from data, for the H_T control region 125–300 GeV and for the 2 signal regions $H_T > 300$ and $H_T > 600$ GeV. The quantity $1/K$ is the efficiency

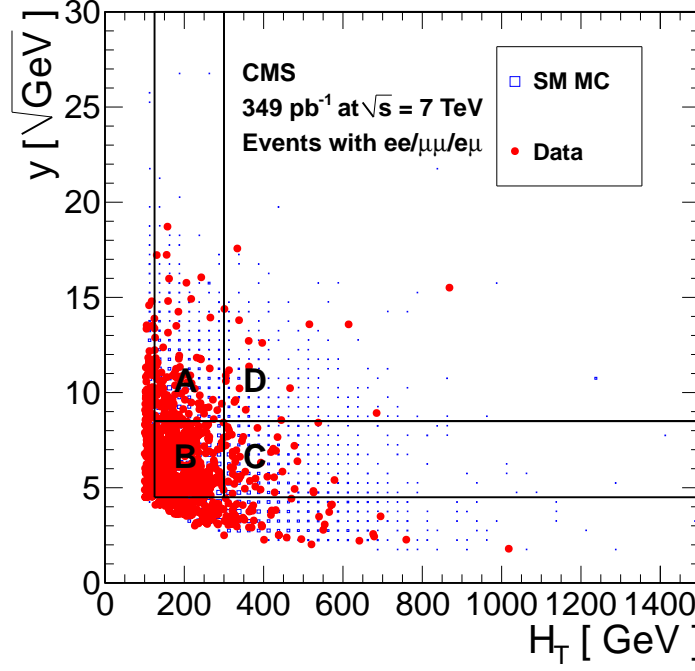


Figure 7: Distributions of y vs. H_T for SM Monte Carlo and data. The 2010 signal region boundaries are overlaid.

Table 19: Data yields in the four regions of Figure 7 for the 2010 signal region, as well as the predicted yield in region D given by $N_A \times N_C / N_B$. The quoted uncertainty on the prediction in data is statistical only, assuming Gaussian errors. We also show the SM Monte Carlo expectations with statistical errors only.

sample	N_A	N_B	N_C	N_D	$N_A \times N_C / N_B$
$t\bar{t} \rightarrow \ell^+ \ell^-$	63.9 ± 2.0	252.3 ± 4.1	43.3 ± 1.7	8.5 ± 0.7	11.0 ± 0.6
$t\bar{t} \rightarrow \ell^\pm \tau^\mp / \tau^+ \tau^-$	18.5 ± 1.1	55.9 ± 1.9	10.3 ± 0.8	3.4 ± 0.5	3.4 ± 0.4
$t\bar{t} \rightarrow \text{fake}$	1.0 ± 0.3	6.6 ± 0.7	1.1 ± 0.3	0.4 ± 0.2	0.2 ± 0.1
DY	0.9 ± 0.6	13.9 ± 2.6	1.3 ± 0.8	1.7 ± 1.0	0.1 ± 0.1
$W^+ W^-$	1.1 ± 0.1	2.7 ± 0.2	0.2 ± 0.1	0.3 ± 0.1	0.1 ± 0.0
$W^\pm Z^0$	0.1 ± 0.0	0.6 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Z^0 Z^0$	0.0 ± 0.0	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	3.3 ± 0.2	9.6 ± 0.3	0.4 ± 0.1	0.1 ± 0.0	0.1 ± 0.0
$W + \text{jets}$	1.0 ± 1.0	2.2 ± 1.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total SM MC	89.9 ± 2.6	344.0 ± 5.3	56.6 ± 2.1	14.5 ± 1.4	14.8 ± 0.7
data	110	381	44	19	12.7 ± 2.4

for events passing preselection and falling in the given H_T range to pass the requirement $p_T(\ell\ell) > 50$ GeV. The values of K extracted from data and $t\bar{t}$ MC are given in Table 20. For all 3 H_T regions, the value of K extracted from data agrees with the $t\bar{t}$ MC prediction, but for the $H_T > 600$ GeV region the statistical uncertainty in K from data is $\sim 100\%$. Therefore we use the value of K extracted from data for the control region 125–300 GeV and for the signal regions $H_T > 300$; for the $H_T > 600$ GeV region we use K from $t\bar{t}$ MC.

For each signal region D, we count the number of events falling in the region D', which is defined using the same requirements as D but switching the y requirement to a $p_T(\ell\ell)/\sqrt{H_T}$ requirement (2010 signal region) or switching the E_T^{miss} requirement to a $p_T(\ell\ell)$ requirement (high E_T^{miss} and high H_T signal regions). We subtract off the expected DY contribution using the data-driven $R_{\text{out/in}}$ technique, using $R_{\text{out/in}} = 0.13 \pm 0.07$. We then scale this yield by 2 corrections factors: K , the E_T^{miss} acceptance correction factor, and K_C , the correction factor determined in Sec. 10. Our final prediction N_P is given by:

$$N_P = (N(D') - N(DY)) \times K \times K_C,$$

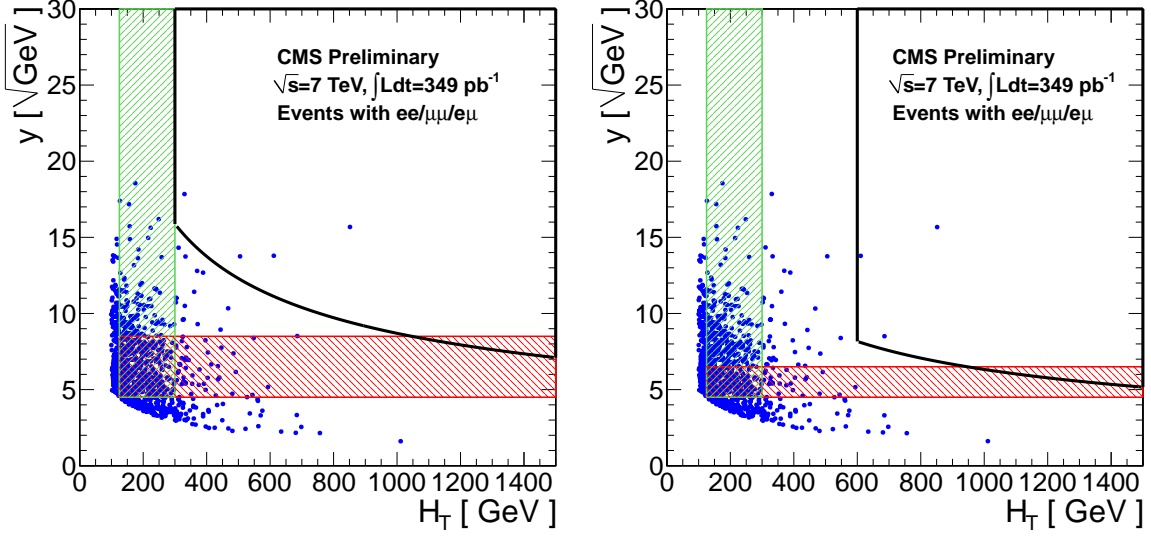


Figure 8: Distributions of y vs. H_T in data. The signal regions $E_T^{\text{miss}} > 275$ GeV, $H_T > 300$ GeV (left) and $E_T^{\text{miss}} > 200$ GeV, $H_T > 600$ GeV (right) are indicated with thick black lines. The $f(y)$ and $g(H_T)$ functions are measured using events in the green and red shaded areas, respectively.

Table 20: Summary of the E_T^{miss} acceptance scaling factor K , extracted from data and $t\bar{t}$ MC.

region	data	$t\bar{t}$ MC
control region: $125 < H_T < 300$ GeV	1.68 ± 0.14	1.67 ± 0.03
signal region: $H_T > 300$ GeV	1.45 ± 0.29	1.50 ± 0.06
signal region: $H_T > 600$ GeV	1.12 ± 1.19	1.32 ± 0.20

as summarized in Table 21, and displayed in Figs. 10-12. We also perform the $p_T(\ell\ell)$ method in the H_T sideband region 125–300 GeV, as a validation of the technique in a high statistics sample which is expected to be dominated by background. The results are summarized in Table 22 and displayed in Fig. 9. The prediction is extracted for the requirement $y > 8.5$ GeV $^{1/2}$ corresponding to the 2010 signal region, as well as for $E_T^{\text{miss}} > 200$ GeV corresponding to the high H_T signal region. In both cases, we observe good agreement between the predicted and observed yields.

11.3 Background estimate from OF subtraction

The results of the OF subtraction technique applied to the high p_T dilepton trigger sample are summarized in Table 23. We evaluate the quantity $\Delta = R_{\mu e} N(ee) + \frac{1}{R_{\mu e}} N(\mu\mu) - N(e\mu)$ with $R_{\mu e} = 1.12 \pm 0.05$ extracted from the ratio of $Z \rightarrow \mu^+\mu^-$ vs. $Z \rightarrow e^+e^-$ events in data. We perform the OF subtraction first in the preselection region, and find Δ consistent with 0, as expected. We then perform the OF subtraction in all 3 signal regions, and do not observe any excess of same-flavor vs. opposite-flavor events.

For the dilepton- H_T trigger sample, we observe only 1 event in the 2010 signal region, consistent with MC expectations, and no events in either the high y or high H_T signal regions. In the case of an excess of events at low lepton p_T , we will perform the OF subtraction technique of Sec. 10.4.2.

Table 21: Summary of results of the dilepton p_T template method applied to the 3 signal regions. The quantities indicated in the table are discussed in the text. The quoted statistical uncertainty in the prediction N_P is due to that of $N(D')$, the quoted systematic uncertainty includes that of $N(DY)$, K , and K_C .

Signal Region	$N(D')$	$N(DY)$	K	K_C	N_P
2010 signal region (UPDATE)	9	0.8 ± 0.4	1.5 ± 0.3	1.4 ± 0.2	16.7 ± 6.1 (stat) ± 4.2 (syst)
high E_T^{miss} signal region	3	0.5 ± 0.3	1.5 ± 0.3	1.5 ± 0.3	5.4 ± 3.8 (stat) ± 1.7 (syst)
high H_T signal region	1	0.0 ± 0.0	1.3 ± 0.2	1.3 ± 0.2	1.7 ± 1.7 (stat) ± 0.4 (syst)

Table 22: Summary of results of the dilepton p_T template method applied to the H_T sideband control region 125–300 GeV. The quantities indicated in the table are discussed in the text. The quoted statistical uncertainty in the prediction N_P is due to that of $N(D')$, the quoted systematic uncertainty includes that of $N(DY)$ and K_C . The predictions are compare with the observed yield N_O .

Control Region	$N(D')$	$N(DY)$	K	K_C	N_P	N_O
$125 < H_T < 300 \text{ GeV}, y > 8.5 \text{ GeV}^{1/2}$	54	2.6 ± 1.2	1.7 ± 0.1	1.4 ± 0.1	$120.9 \pm 17.3 \text{ (stat)} \pm 13.6 \text{ (syst)}$	110
$125 < H_T < 300 \text{ GeV}, E_T^{\text{miss}} > 200 \text{ GeV}$	4	1.0 ± 0.5	1.7 ± 0.1	1.3 ± 0.2	$6.5 \pm 4.4 \text{ (stat)} \pm 1.6 \text{ (syst)}$	6

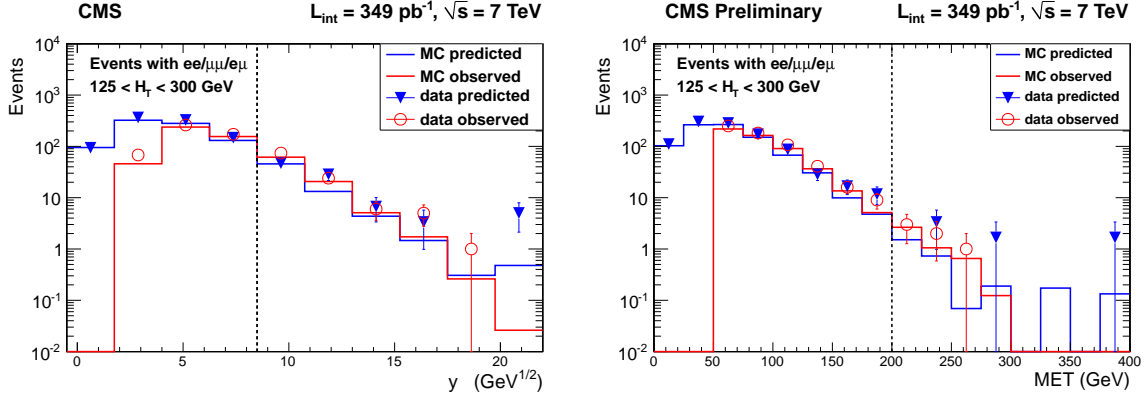


Figure 9: Results of the $p_T(\ell\ell)$ method in the H_T sideband region 125-300 GeV. Left: distributions of $p_T(\ell\ell)/\sqrt{H_T}$ (predicted) and y (observed) for SM MC and data. The vertical dashed lines indicate the requirement $y > 8.5 \text{ GeV}^{1/2}$, corresponding to the 2010 signal region. Right: distributions of $p_T(\ell\ell)$ (predicted) and E_T^{miss} (observed) for SM MC and data. The vertical dashed lines indicate the requirement $E_T^{\text{miss}} > 200 \text{ GeV}$, corresponding to the high H_T signal region.

11.4 Summary of results

A summary of our results is presented in Table 24. In all 3 signal regions, we observe reasonable agreement between the observed yields and the predictions from MC and data-driven background estimates. We therefore do not observe evidence for an excess of events above SM expectations. After assessing systematic uncertainties in Sec. 12, we proceed to set upper limits on the non-SM contributions to the signal regions in Sec. 13.

12 Acceptance and efficiency systematics

This is a search for new physics contributions to events with high E_T^{miss} and lots of jet activity. As seen in Section 11, there is no evidence for a contribution beyond SM expectations.

Strictly speaking it is impossible to talk about “acceptance and efficiency systematics” because these kinds of systematics only apply to a well defined final state. Nevertheless, we can make general statements about the systematic uncertainties, including quantitative estimates of the systematic uncertainties associated with a few specific processes.

The systematic uncertainty on the lepton acceptance consists of two parts: the trigger efficiency uncertainty and

Table 23: Summary of results for the OF subtraction technique. The quantity $\Delta = R_{\mu e}N(ee) + \frac{1}{R_{\mu e}}N(\mu\mu) - N(e\mu)$ is quoted with $R_{\mu e} = 1.12 \pm 0.05$. The quoted systematic uncertainty corresponds to that of $R_{\mu e}$. The $e\mu$ yields differ from those previously quoted because the Z mass veto is included here.

region	$N(ee)$	$N(\mu\mu)$	$N(e\mu)$	Δ
preselection region	193	201	394	$2.0 \pm 28 \text{ (stat)} \pm 1.8 \text{ (syst)}$
2010 signal region	4	4	9	$-0.9 \pm 4.2 \text{ (stat)} \pm 0.1 \text{ (syst)}$
high E_T^{miss} signal region	2	0	1	$1.3 \pm 1.9 \text{ (stat)} \pm 0.1 \text{ (syst)}$
high H_T signal region	1	0	1	$0.1 \pm 1.5 \text{ (stat)} \pm 0.0 \text{ (syst)}$

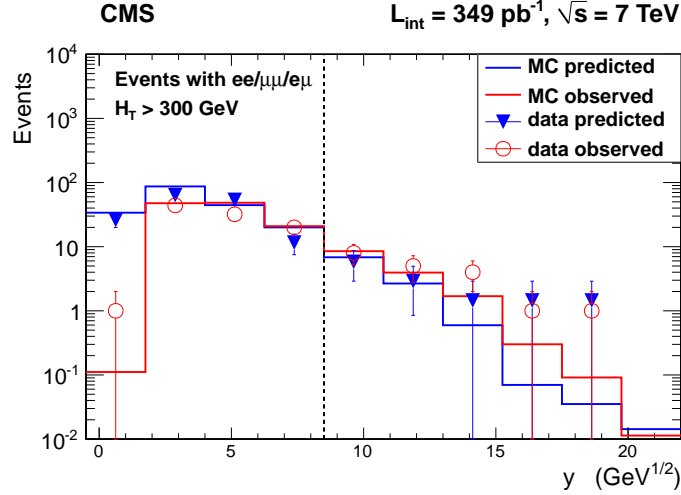


Figure 10: Distributions of $p_T(\ell\ell)/\sqrt{H_T}$ (predicted) and y (observed) for SM MC and data, for the $H_T > 300$ GeV. The vertical dashed lines indicate the requirement $y > 8.5 \text{ GeV}^{1/2}$ corresponding to the 2010 signal region.

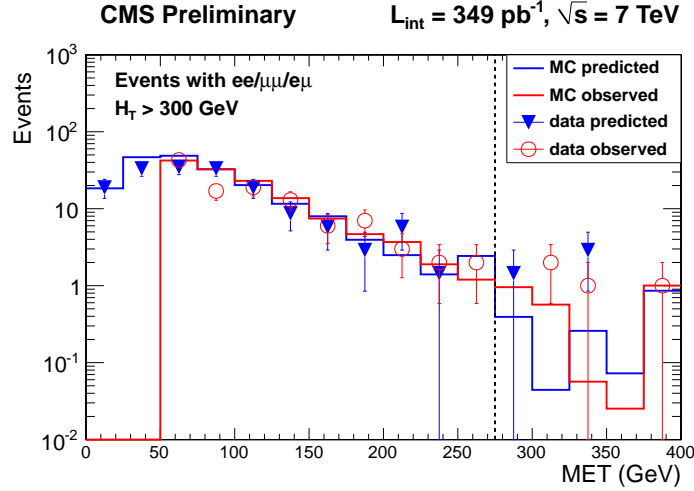


Figure 11: Distributions of $p_T(\ell\ell)$ (predicted) and E_T^{miss} (observed) for SM MC and data, for the region $H_T > 300$ GeV. The vertical dashed lines indicate the requirement $E_T^{\text{miss}} > 275 \text{ GeV}$, corresponding to the high E_T^{miss} signal region.

the ID and isolation uncertainty. We discuss these in turn.

The trigger efficiency for the high p_T dilepton triggers has been studied in [7] in detail. The efficiency is found to be approximately 100% for ee , 95% for $e\mu$, and 90% for $\mu\mu$. There is a small dependence of the efficiency on the lepton p_T , which we do not attempt to parameterize. Instead, we assign an uncertainty of 2%, based on the small variation of the efficiency on the lepton p_T .

To evaluate the uncertainty on the lepton ID and isolation efficiencies, we perform a tag and probe technique on Z data and MC. We calculate the efficiency for leptons passing isolation criteria to also pass the ID criteria, and the efficiency for leptons passing the ID criteria to also pass the isolation criteria, as summarized in Table. 25. We observe agreement between data and MC within about 2%, and assign a corresponding systematic uncertainty.

Another significant source of systematic uncertainty is associated with the jet and E_T^{miss} energy scale. The impact of this uncertainty is final-state dependent. Final states characterized by lots of hadronic activity and E_T^{miss} are less sensitive than final states where the E_T^{miss} and H_T are typically close to the requirement. To be more quantitative, we have used the method of Reference [4] to evaluate the systematic uncertainties on the acceptance for $t\bar{t}$ and two benchmark SUSY points. The uncertainties are calculated assuming a 5% uncertainty to the hadronic energy scale

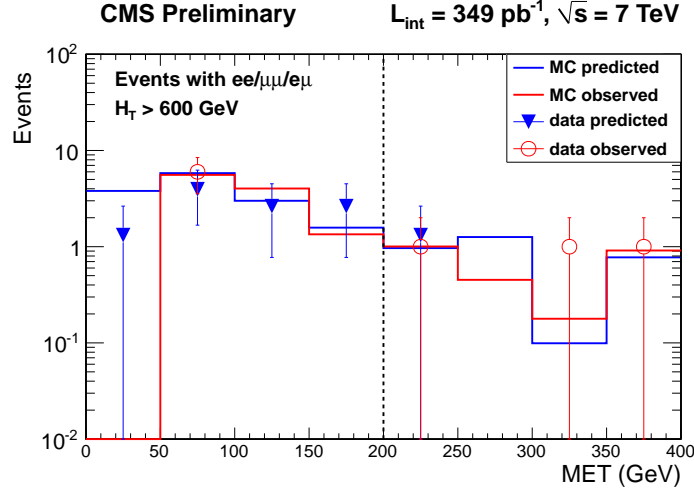


Figure 12: Distributions of $p_T(\ell\ell)$ (predicted) and E_T^{miss} (observed) for SM MC and data, for the region $H_T > 600$ GeV. The vertical dashed lines indicate the requirement $E_T^{\text{miss}} > 200$ GeV, corresponding to the high H_T signal region.

Table 24: Summary of the observed and predicted yields in the 3 signal regions. MC errors are statistical only. The systematic uncertainty on the ABCD and $p_T(\ell\ell)$ method is from the scaling factors from MC closure only. For the OF subtraction, the quantity $\Delta = R_{\mu e}N(ee) + \frac{1}{R_{\mu e}}N(\mu\mu) - N(e\mu)$ is quoted; the systematic uncertainty here is from the ratio of muon to electron selection efficiencies.

	2010 signal region	high E_T^{miss} signal region	high H_T signal region
Observed yield	19	4	3
MC prediction	14.5 ± 1.4	2.6 ± 0.8	2.5 ± 0.8
ABCD prediction	$12.7 \pm 2.4 \text{ (stat)} \pm 2.5 \text{ (syst)}$		
ABCD' prediction	$12.8 \pm 2.9 \text{ (stat)} \pm 2.6 \text{ (syst)}$	$1.3 \pm 0.4 \text{ (stat)} \pm 0.2 \text{ (syst)}$	$0.0 \pm 0.6 \text{ (stat)} \pm 0.1 \text{ (syst)}$
$p_T(\ell\ell)$ prediction	$16.7 \pm 6.1 \text{ (stat)} \pm 4.2 \text{ (syst)}$	$5.4 \pm 3.8 \text{ (stat)} \pm 1.7 \text{ (syst)}$	$1.7 \pm 1.7 \text{ (stat)} \pm 0.4 \text{ (syst)}$
OF subtraction (Δ)	$-0.9 \pm 4.2 \text{ (stat)} \pm 0.1 \text{ (syst)}$	$1.3 \pm 1.9 \text{ (stat)} \pm 0.1 \text{ (syst)}$	$0.1 \pm 1.5 \text{ (stat)} \pm 0.0 \text{ (syst)}$

in CMS.

13 Derivation of Upper Limits

Here is the plan:

- Finalize systematic uncertainty for ABCD' and $p_T(\ell\ell)$ methods.
 - For $p_T(\ell\ell)$ the missing part is jet/met scale uncertainty on K_C ,
 - For ABCD', will assess uncertainties using obs/pred for Derek's toy studies
- Decide how to assess the bkg estimate. Take weighted average of ABCD' and $p_T(\ell\ell)$ methods?

Table 25: Tag and probe results on $Z \rightarrow \ell\ell$ on data and MC. We quote ID efficiency given isolation and the isolation efficiency given ID. **UPDATE TABLE WITH 2011 DATA/MC**

	Data T&P	MC T&P
$\epsilon(id iso)$ electrons	0.925 ± 0.007	0.934 ± 0.004
$\epsilon(iso id)$ electrons	0.991 ± 0.002	0.987 ± 0.002
$\epsilon(id iso)$ muons	0.962 ± 0.005	0.984 ± 0.002
$\epsilon(iso id)$ muons	0.987 ± 0.003	0.982 ± 0.002

Table 26: Summary of efficiency uncertainties due to the 5% uncertainty in the hadronic energy scale.

Signal Region	$t\bar{t}$	LM1	LM3
2010 signal region	0.28	0.06	0.07
high E_T^{miss} signal region	0.35	0.14	0.18
high H_T signal region	0.30	0.20	0.21

- 446 • Derive model-indepenent yield upper limits
- 447 • Exclude LM1 and/or LM3, if possible
- 448 • For approval timescale, add the scan

References

- [1] ADD REF TO MET TEMPLATES NOTE, WHEN AVAILABLE
- [2] CMS AN-2010/370
- [3] arXiv:1103.1348v1 [hep-ex], “Search for Physics Beyond the Standard Model in Opposite-Sign Dilepton Events at $\sqrt{s} = 7$ TeV.”
- [4] Phys.Lett.B695:424-443,2011
- [5] <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SimpleCutBasedEleID>
- [6] D. Barge *et al.*, AN-CMS2009/159.
- [7] CMS AN-2011/155
- [8] <https://twiki.cern.ch/twiki/bin/view/CMS/ProductionReProcessingSpring10>
- [9] V. Pavlunin, Phys. Rev. **D81**, 035005 (2010).
- [10] D. Barge *et al.*, AN-CMS2009/130.
- [11] W. Andrews *et al.*, AN-CMS2009/023.
- [12] D. Barge *et al.*, AN-CMS2010/257.
- [13] W. Andrews *et al.*, AN-CMS2010/274.
- [14] J. Conway, <http://www-cdf.fnal.gov/physics/statistics/code/bayes.f>.
- [15] G. Landsberg, <https://twiki.cern.ch/twiki/pub/CMS/EXOTICA/cl95cms.c>
- [16] <https://hypernews.cern.ch/HyperNews/CMS/get/susy/617/2/1.html>
- [17] <https://twiki.cern.ch/twiki/bin/view/CMS/SUSY38XSUSYScan>
- [18] arXiv:hep-ph/0605240v2
- [19] CleanExclusion.cc available at <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SUSYLimitTools>
- [20] R. Cousins, http://www.physics.ucla.edu/~cousins/stats/cousins_lognormal_prior.pdf
- [21] S. Harper, private communication (relayed to us by M. Chiorboli.).
- [22] A. Barr *et al.*, J.Phys.G29:2343-2363,2003; Cheng, H.C., Han, arXiv:hep-ph/0810.5178v2.
<http://indico.cern.ch/contributionDisplay.py?contribId=3&confId=66410>
- [23] <http://indico.cern.ch/contributionDisplay.py?contribId=5&confId=93837>
- [24] M. Narain *et al.*, CMS AN-2010/259; we thank the Brown group for providing their code to us.

Appendix A Fakeable Object Definitions

We estimate the contributions from leptons not originating from W/Z decay (fake leptons) using the data-driven fake rate method [4]. We define the following ‘fakeable object’ selections, by taking the electron and muon requirements listed in Sec. 4 and loosening the following requirements:

- electrons

- $d_0 < 0.2 \text{ cm}$

- $Iso \equiv E_T^{\text{iso}}/p_T < 0.4$, E_T^{iso} is defined as the sum of transverse energy/momentum deposits in ecal, hcal, and tracker, in a cone of 0.3 A 1 GeV pedestal is subtracted from the ecal energy deposition in the EB, however the ecal energy is never allowed to go negative.

- muons

- $d_0 < 0.2 \text{ cm}$

- $\chi^2/\text{ndof of global fit} < 50$

- $Iso \equiv E_T^{\text{iso}}/p_T < 0.4$, E_T^{iso} is defined as the sum of transverse energy/momentum deposits in ecal, hcal, and tracker, in a cone of 0.3

Appendix B The ABCD' Technique

We have developed a novel technique, based on the ABCD technique which we perform in the plane of y vs. H_T , which allows us to estimate the background in a signal region defined in the E_T^{miss} vs. H_T plane. We refer to this as the ABCD' technique.

First, we extract the y and H_T distributions $f(y)$ and $g(H_T)$ from data, using events from control regions which we expect to be dominated by background. Under the assumption that y and H_T are weakly-correlated (the same assumption which allows the use of the ABCD technique), we can predict the distribution of events in the y vs. H_T plane as:

$$\frac{\partial^2 N}{\partial y \partial H_T} = f(y)g(H_T) \quad (3)$$

Once the distribution of events in the y vs. H_T plane is known, the number of events falling in any region of this plane can be deduced. In particular, we can deduce the number of events falling in a region defined by requirements on E_T^{miss} and H_T .

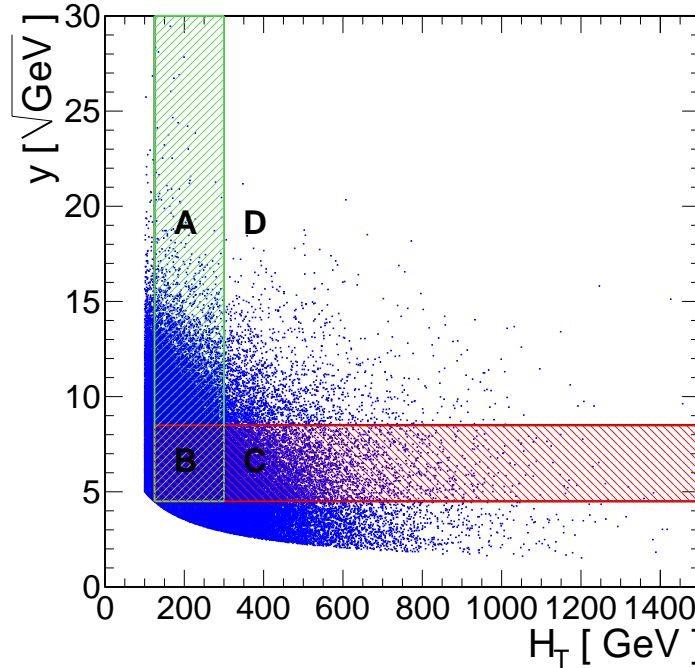


Figure 13: Distributions of y vs. H_T for $t\bar{t}$ MC. The $f(y)$ and $g(H_T)$ functions are measured using events in the green and red shaded areas, respectively.

In this section, we study the ABCD' technique by applying it to $t\bar{t}$ MC (we use the powheg sample which has 10 times the number of dilepton events as the madgraph sample). As a cross-check, we first use the ABCD' technique to estimate the background in the 2010 signal region defined as $y > 8.5 \text{ GeV}^{1/2}$ and $H_T > 300 \text{ GeV}$, which may be compared to the prediction of the standard ABCD technique. We measure the $f(y)$ and $g(H_T)$ distributions using events in the regions indicated in Fig. 13, yielding the distributions shown in Fig. 14. Next, we randomly sample values of y and H_T from these distributions; each pair of y and H_T values is a pseudo-event. We generate 1 million pseudo-events, and find the ratio $R_{S/C}$, the ratio of the number of pseudo-events falling in the 2010 signal region (ie. region D) to the number of pseudo-events falling in a control region, defined as the OR of the A, B and C regions. We then multiply this ratio by the number of $t\bar{t}$ MC events which fall in the control region to get the predicted yield, ie. $N_{\text{pred}} = R_{S/C} \times N(\text{control})$. To estimate the statistical uncertainty in the predicted background, we smear the bin contents of $f(y)$ and $g(H_T)$ according to their uncertainties. We repeat the prediction 20 times with these smeared distributions, and take the RMS of the deviation from the nominal prediction as the statistical uncertainty. The results are summarized in Table 27, which show that the prediction of the ABCD' method is consistent with that of the ABCD method within the statistical uncertainty.

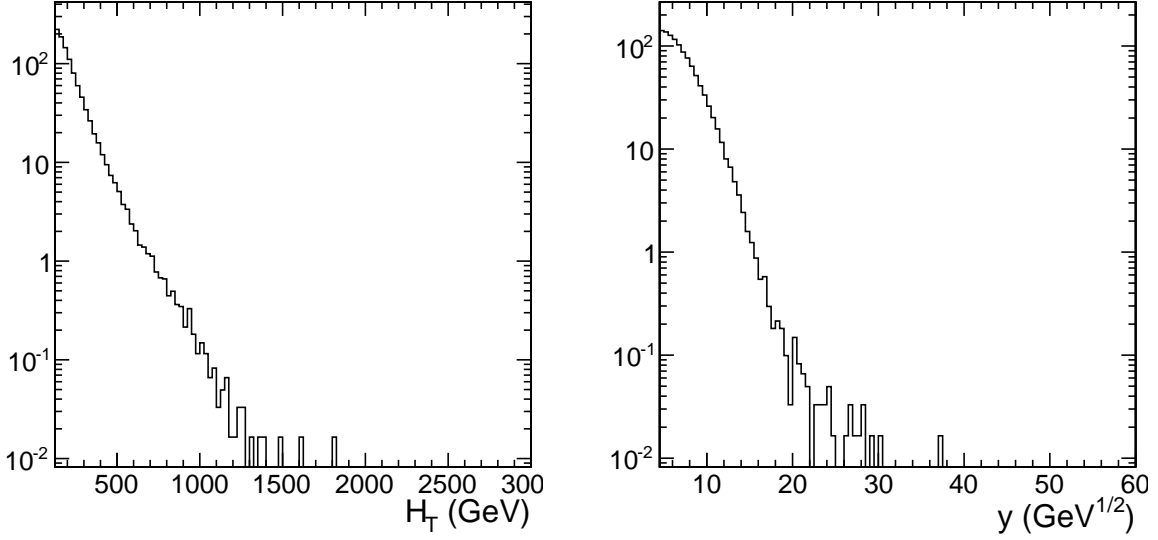


Figure 14: The distributions $f(y)$ and $g(H_T)$ extracted from $t\bar{t}$ MC.

We next apply the ABCD' technique to estimate the background in the regions:

- SR1: $E_T^{\text{miss}} > 275 \text{ GeV}$, $H_T > 300 \text{ GeV}$
- SR2: $E_T^{\text{miss}} > 200 \text{ GeV}$, $H_T > 600 \text{ GeV}$.

Note that we are unable to predict the yield in this region using the standard ABCD technique, since E_T^{miss} and H_T are not weakly correlated. The signal regions are shown in Fig. 15. There is a sizable overlap between SR2 and region C, hence when estimating the background for this signal region we restrict the control region used to measure $g(H_T)$ and to determine $N(\text{control})$ to $4.5 < y < 6.5 \text{ GeV}^{1/2}$. As summarized in Table 18, we find agreement between the observed and predicted yields in these signal regions within $(30 \pm 10)\%$ and $(5 \pm 7)\%$ for SR1 and SR2, respectively.

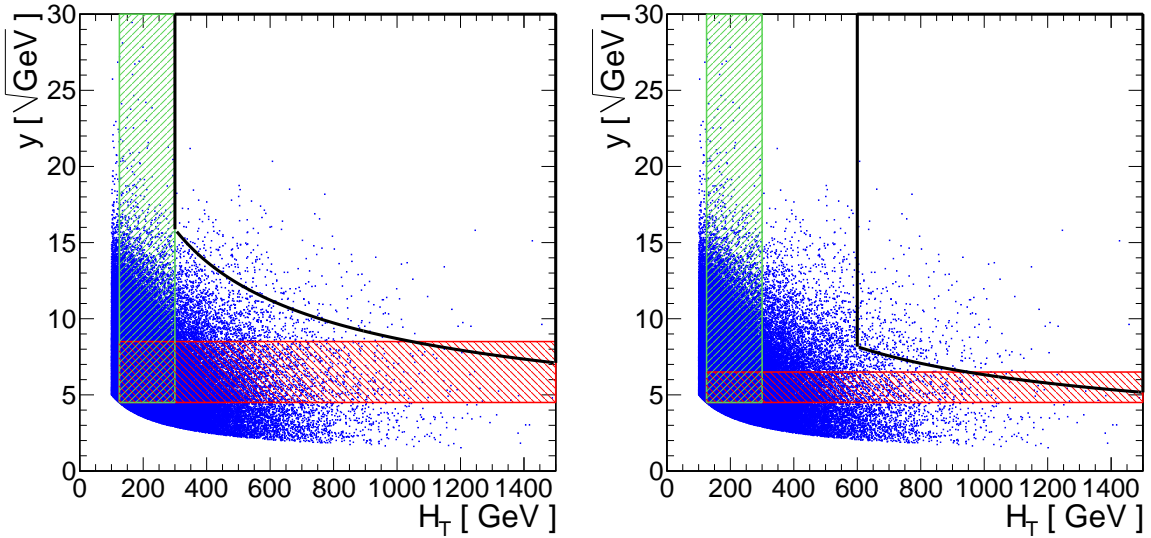


Figure 15: Distributions of y vs. H_T for $t\bar{t}$ MC. The signal regions $E_T^{\text{miss}} > 275 \text{ GeV}$, $H_T > 300 \text{ GeV}$ (left) and $E_T^{\text{miss}} > 200 \text{ GeV}$, $H_T > 600 \text{ GeV}$ (right) are indicated with thick black lines. The $f(y)$ and $g(H_T)$ functions are measured using events in the green and red shaded areas, respectively.

Table 27: Summary of results expected results of the ABCD' method for 1 fb^{-1} using $t\bar{t}$ MC. The quantities $R_{S/C}$ and $N(\text{control})$ are discussed in the text. The prediction of the ABCD' technique (ABCD' pred) is compared with the observed signal yield (observed) and where applicable, with the prediction of the ABCD technique (ABCD pred). The errors in ABCD' pred and observed/ABCD' pred include only the statistical uncertainty in the observed yield.

	2010 signal region	$E_T^{\text{miss}} > 275 \text{ GeV}, H_T > 300 \text{ GeV}$	$E_T^{\text{miss}} > 200 \text{ GeV}, H_T > 600 \text{ GeV}$
$R_{S/C}$	3.5×10^{-2}	2.8×10^{-3}	4.9×10^{-3}
$N(\text{control})$	1239	1239	1187
ABCD' pred	43.4	3.5 ± 0.1	5.9 ± 0.3
observed	37.0 ± 0.8	4.5 ± 0.3	6.2 ± 0.3
observed/ABCD' pred	0.85 ± 0.02	1.30 ± 0.10	1.05 ± 0.07
ABCD pred	43.0 ± 0.6	N/A	N/A

WILL ADD HERE DESCRIPTION OF DEREK'S TOY STUDIES, AND DERIVE SYST UNCERTAINTIES BASED ON THEM

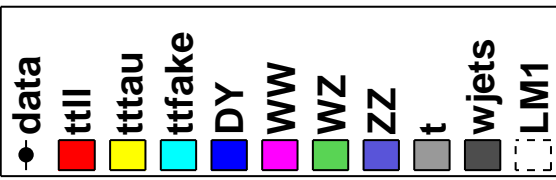
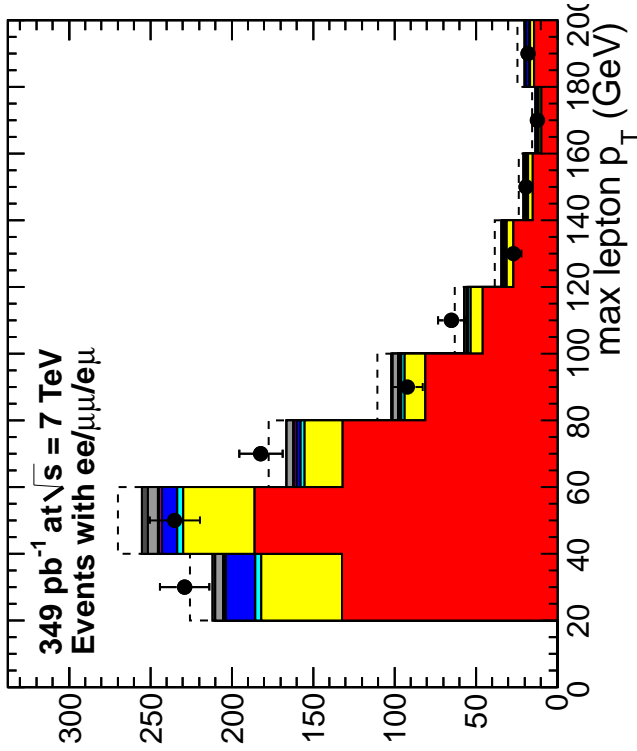
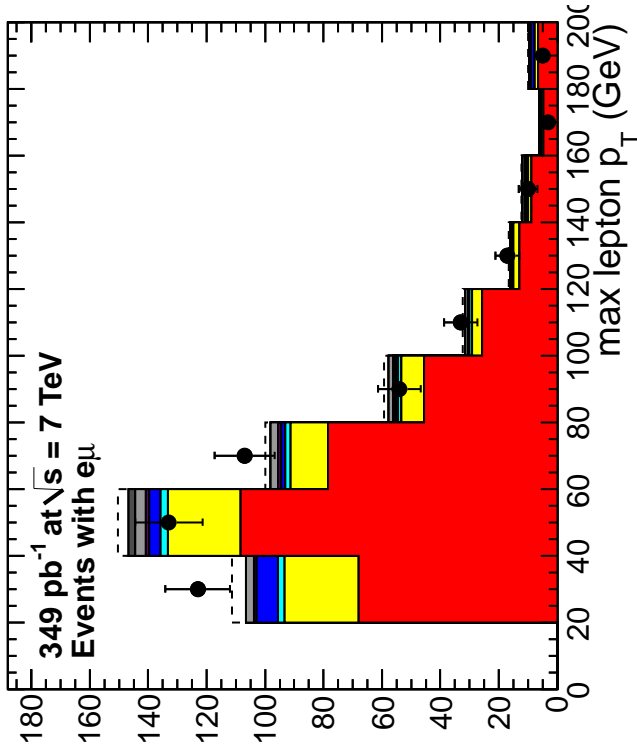
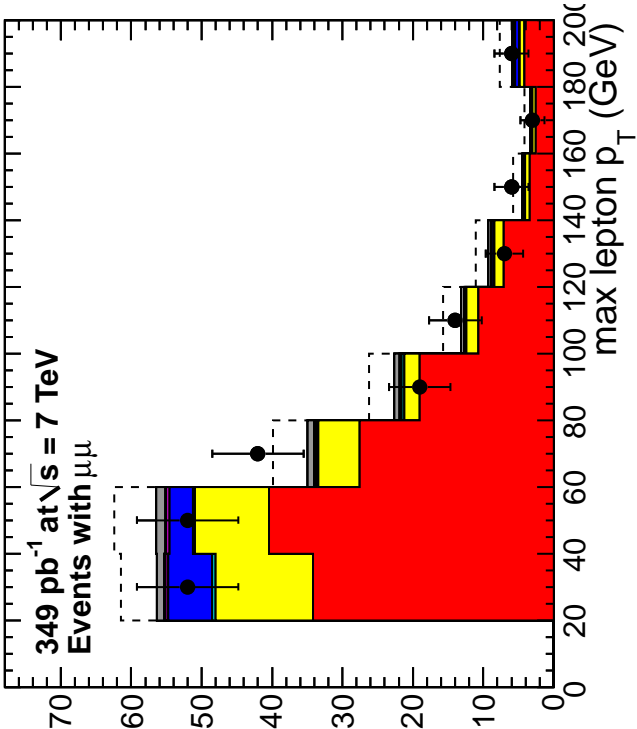
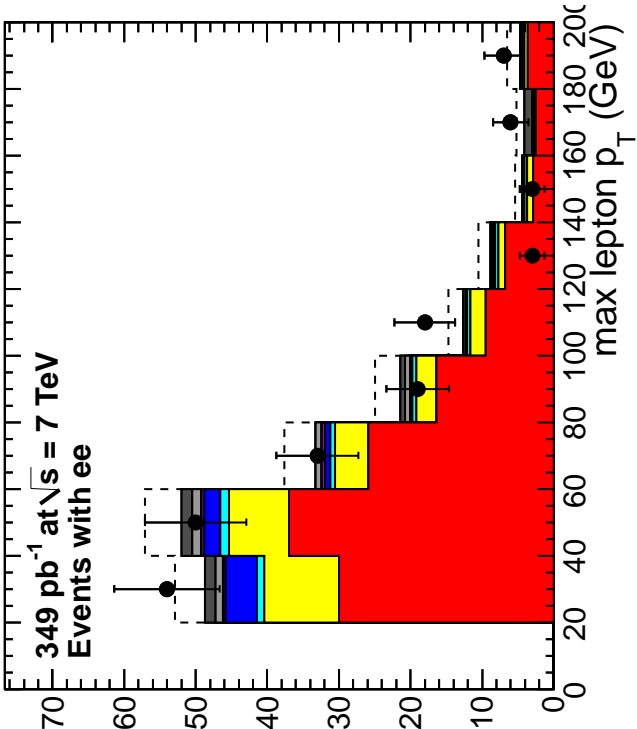
Appendix C Data/MC Comparison: Preselection Region

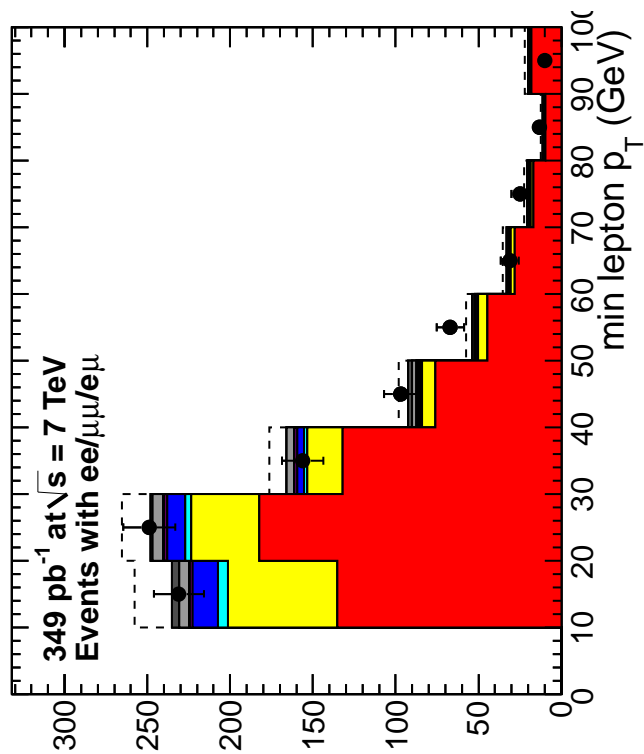
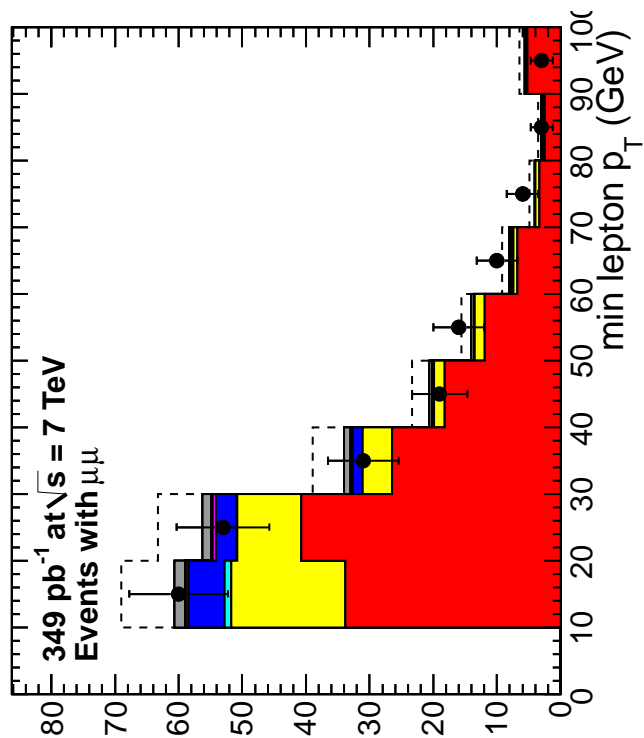
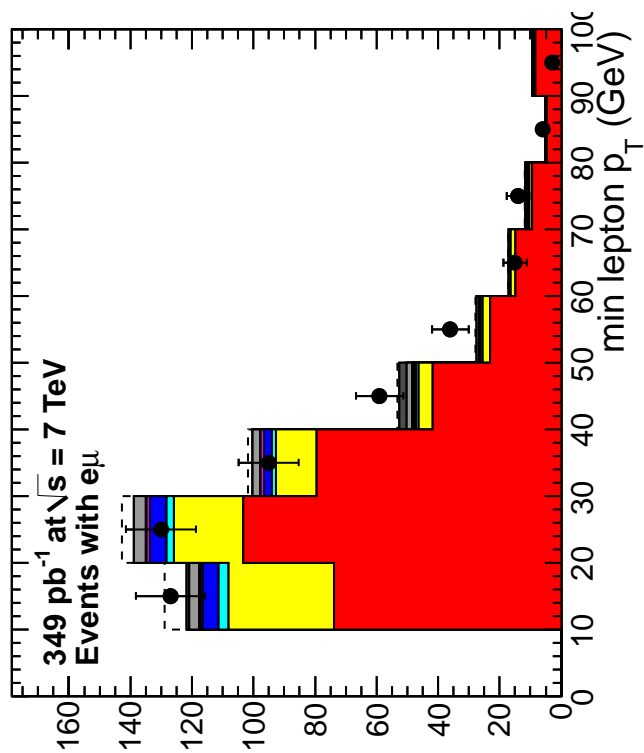
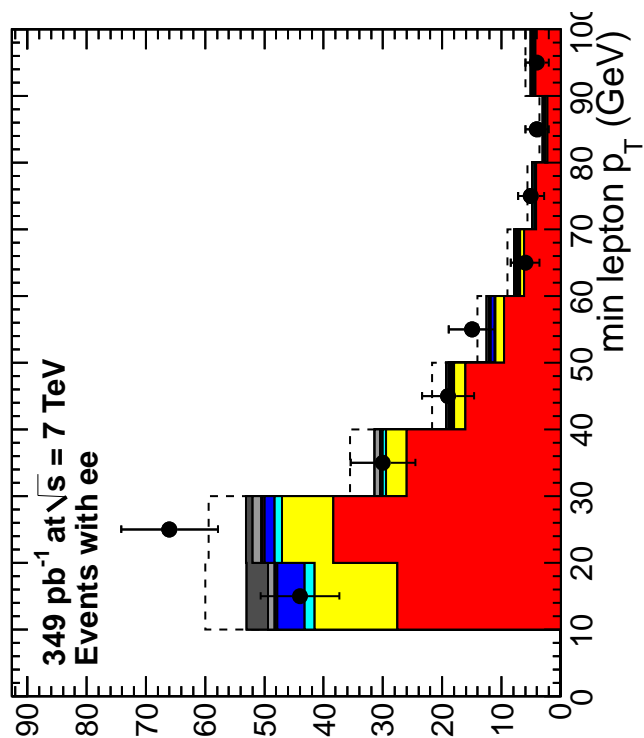
Here we compare data and MC distributions for data passing the preselection requirements. The high p_T dilepton trigger data is used.

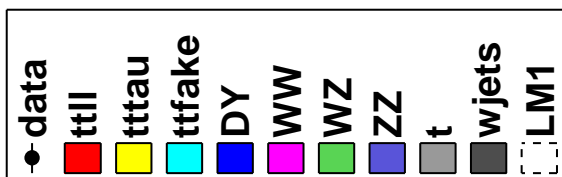
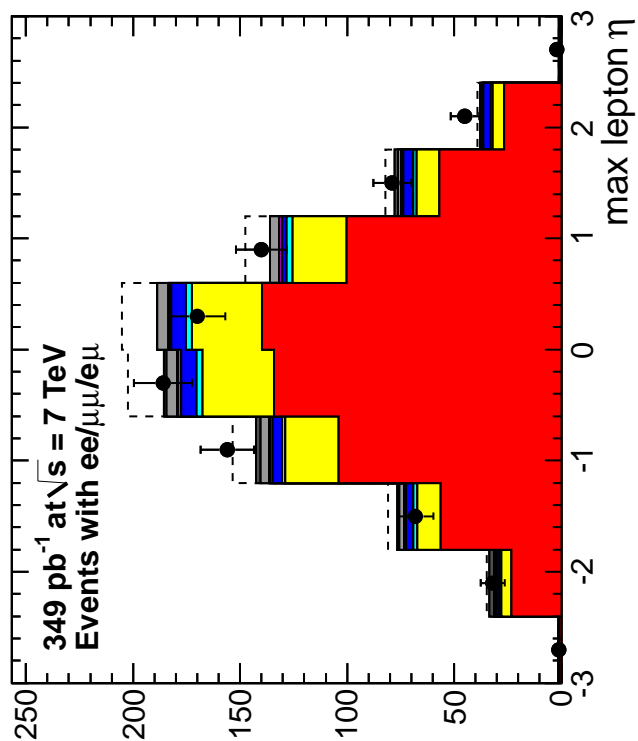
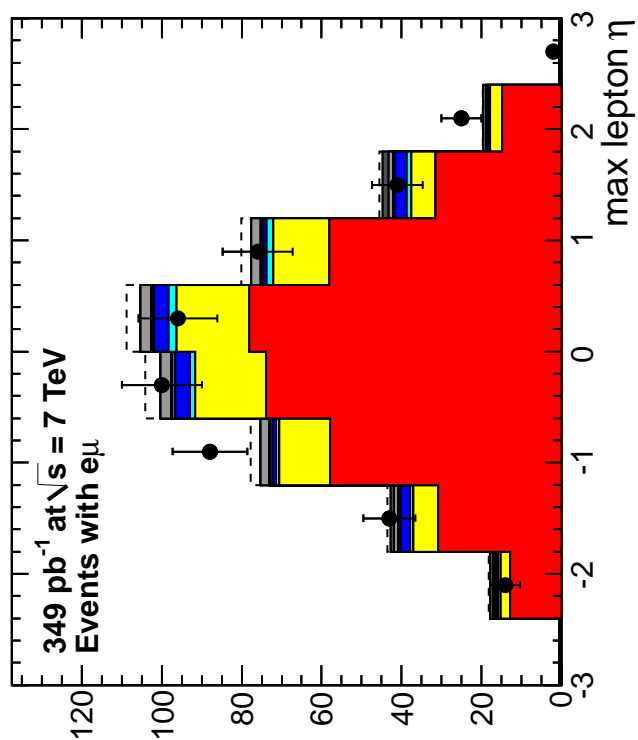
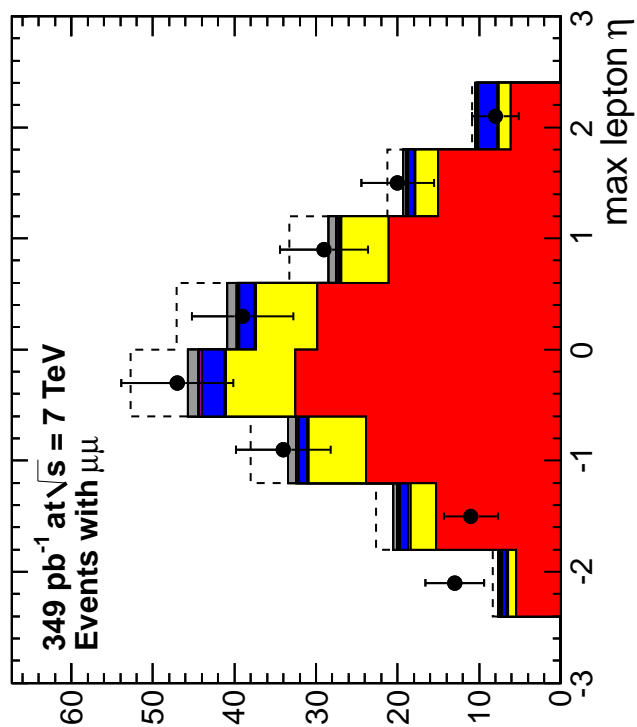
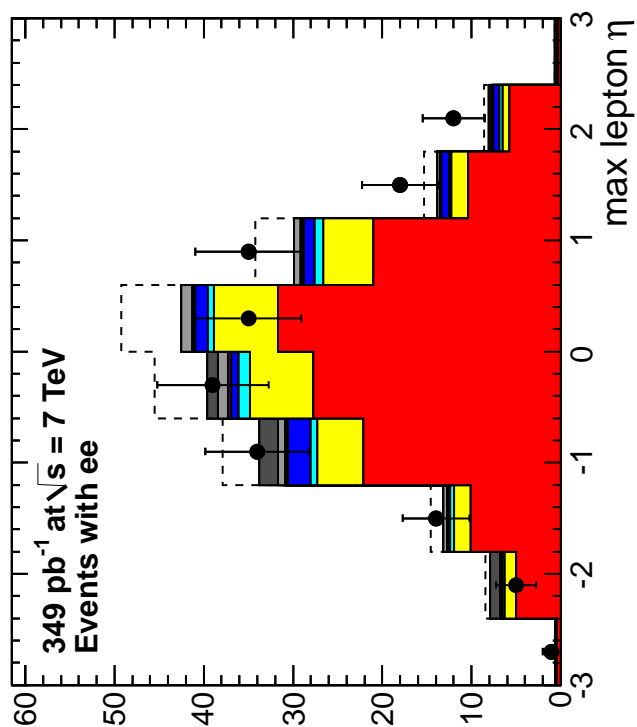
All MC samples are scaled by an overall factor 1.12, the ratio of the observed data yield in the preselection region to the prediction from MC. For illustration purposes, we overlay the distributions from the LM1 SUSY benchmark point.

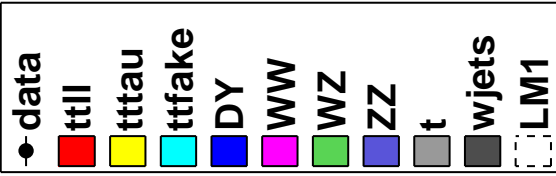
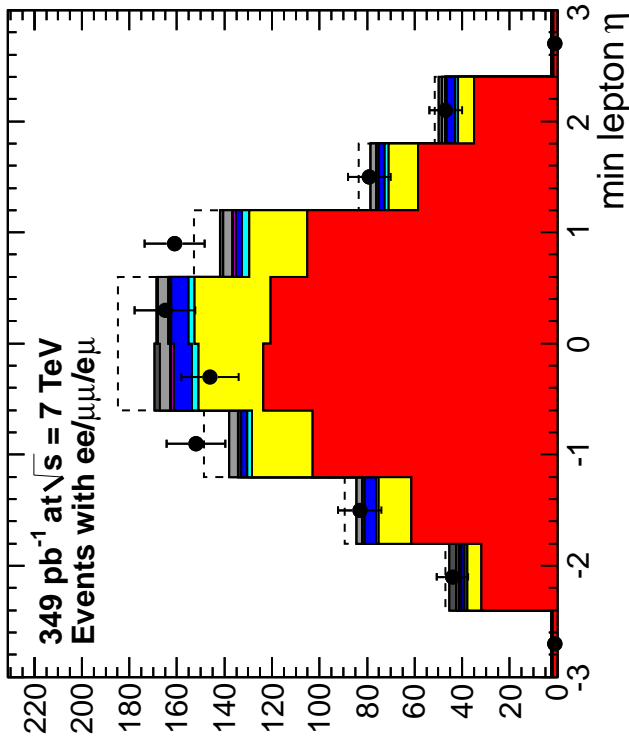
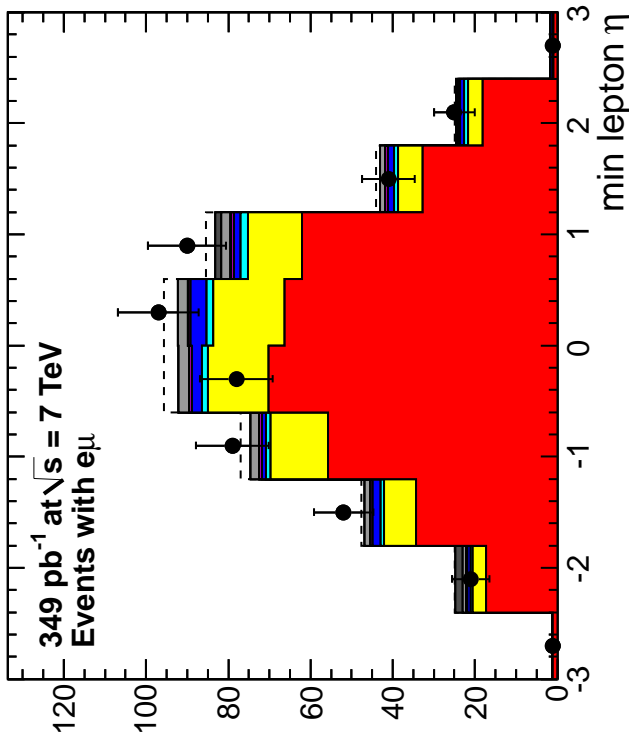
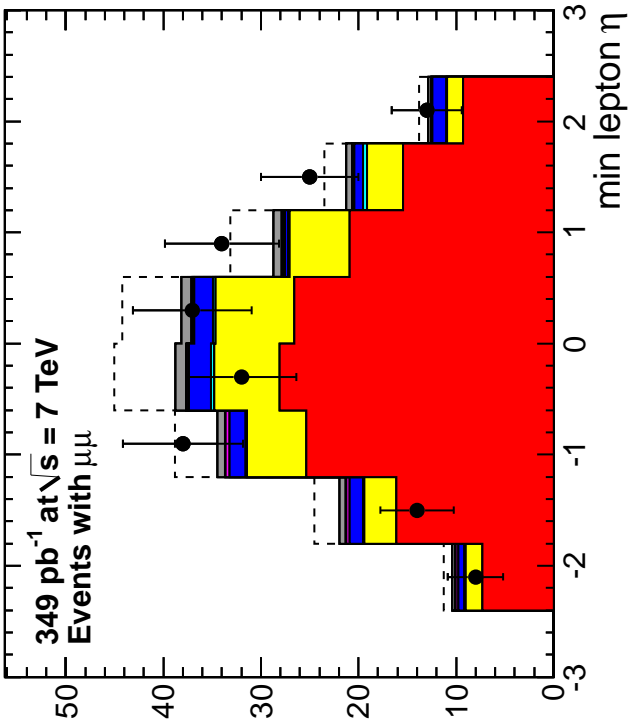
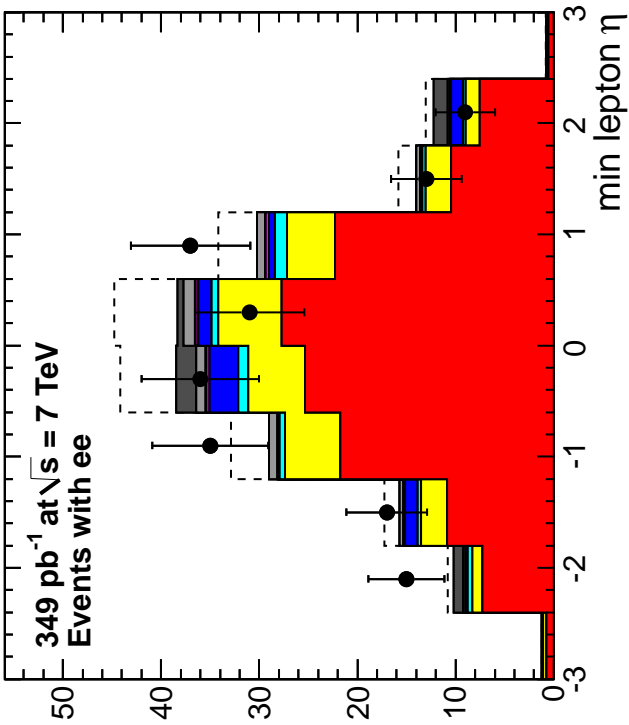
The meaning of most of the variables plotted in the following figures should be obvious. There are some exceptions that we explain below:

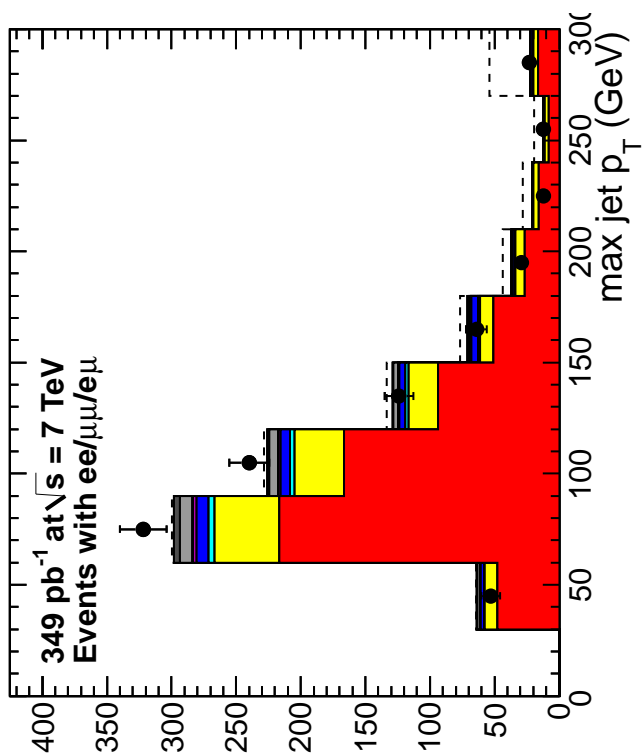
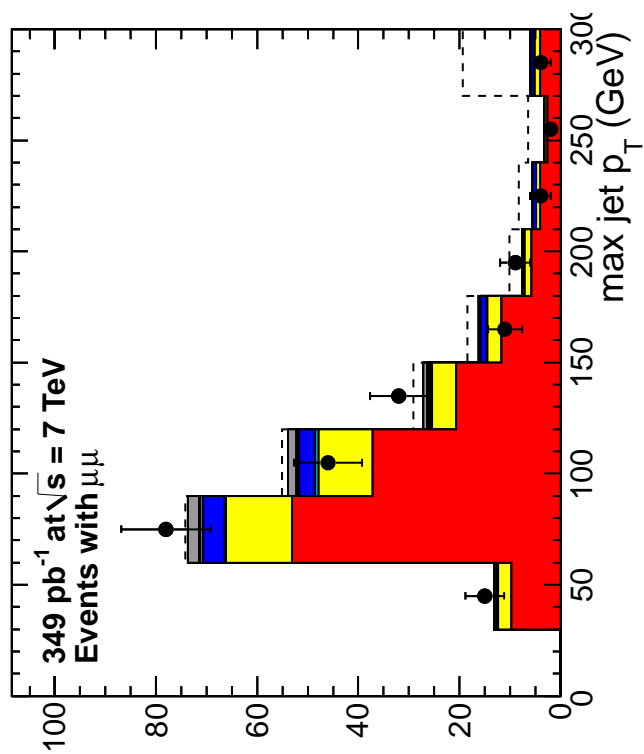
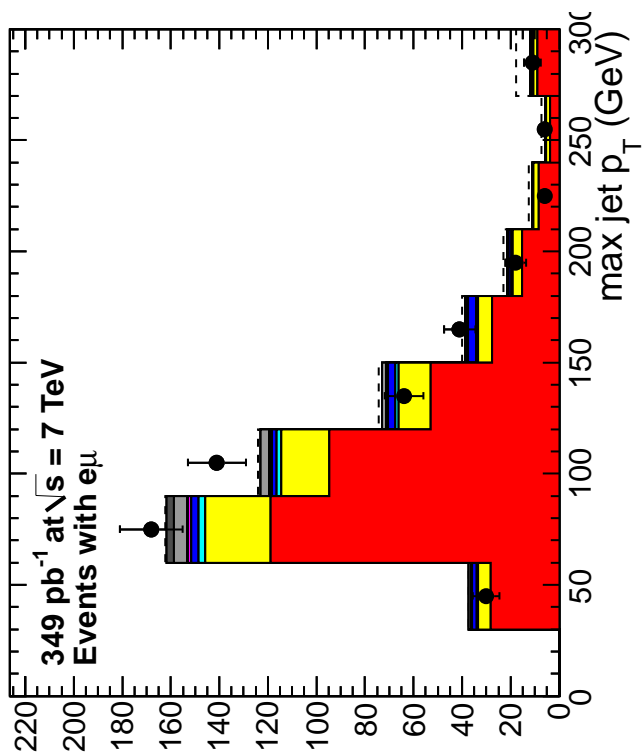
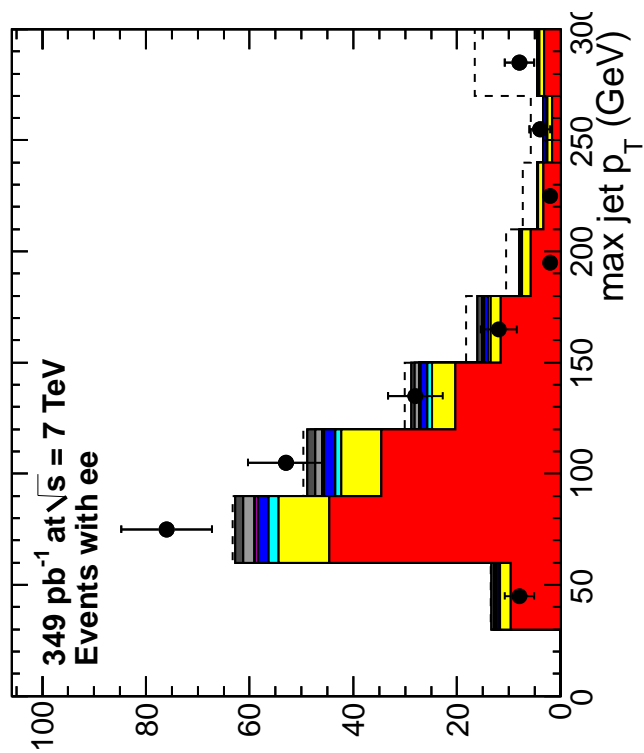
- $MT2$ is a kinematical quantity built from the two leptons and the E_T^{miss} . For events with two $W \rightarrow \ell$ decays it should have a sharp kinematical cutoff at W mass. For more details, see Reference [22].
- $MT2J$ is very much like $MT2$ but it is built out of the leptons, the E_T^{miss} and the two jets. For $t\bar{t}$ events it has a kinematical cutoff at M_{top} , with tails due to the fact that occasionally one of the b -jets is not found and is replaced by a gluon jet from ISR or FSR. For more details, see Reference [23].

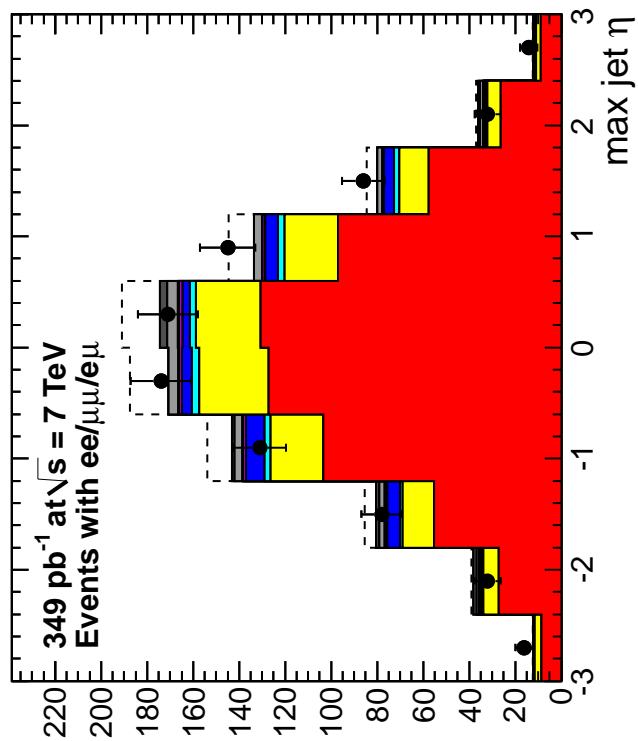
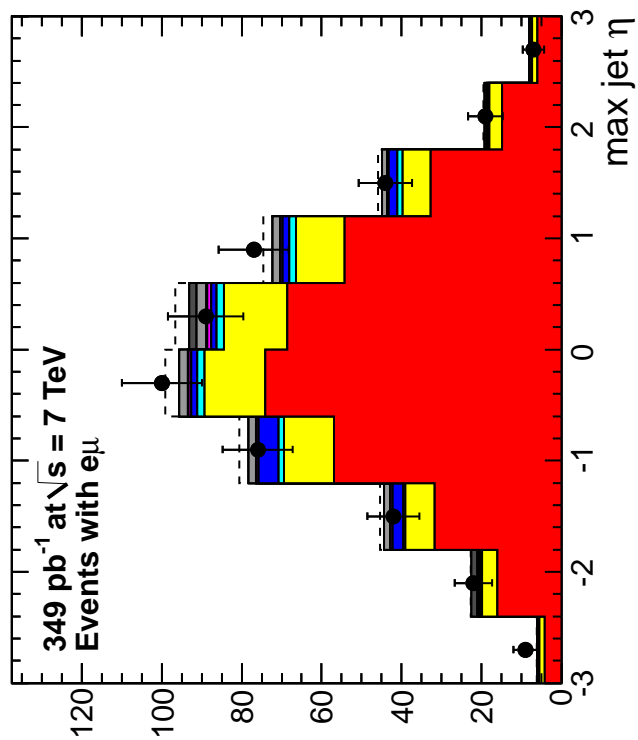
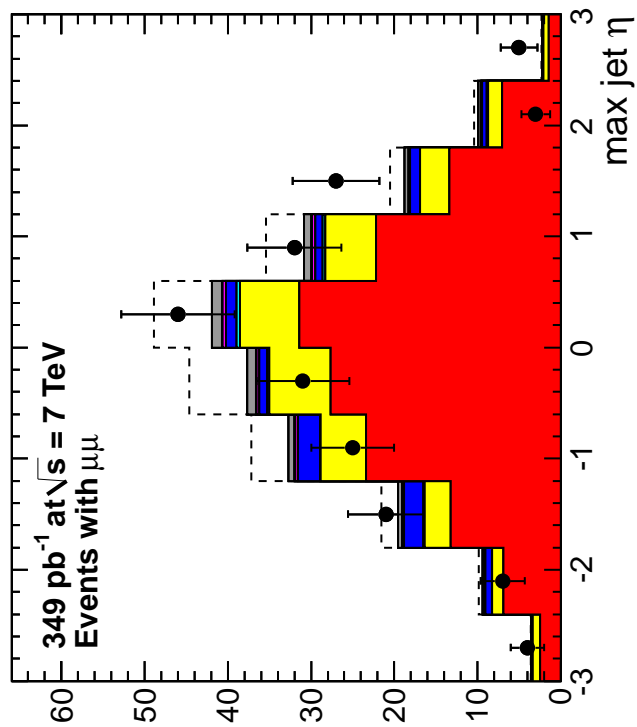
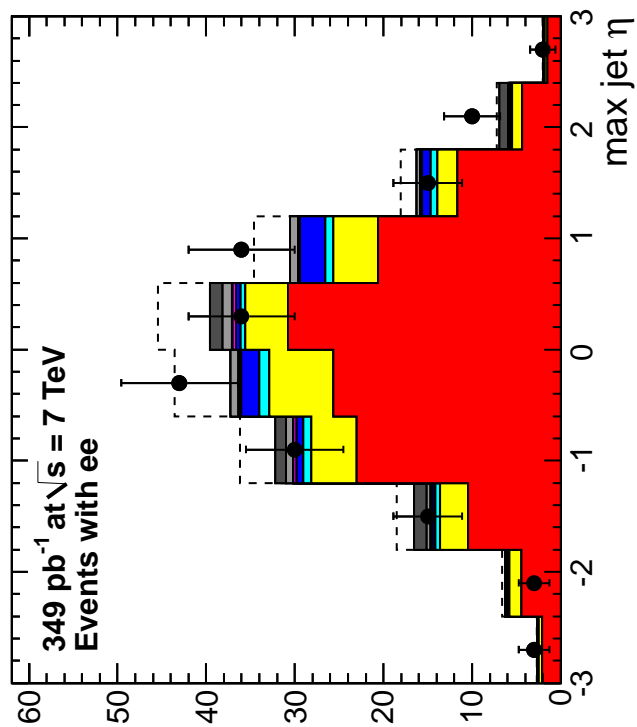


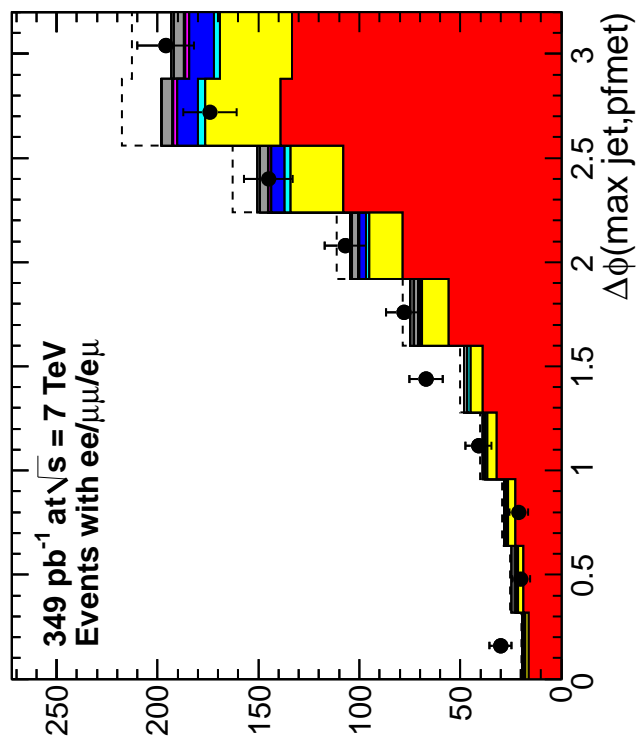
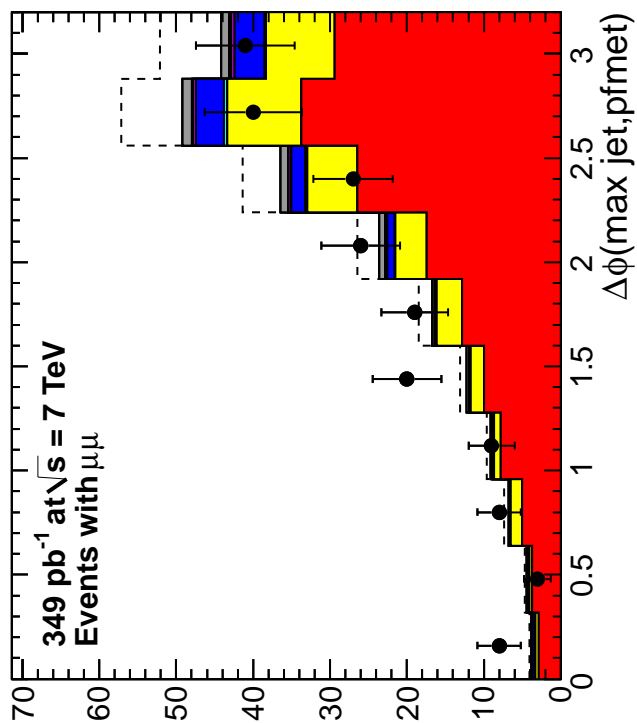
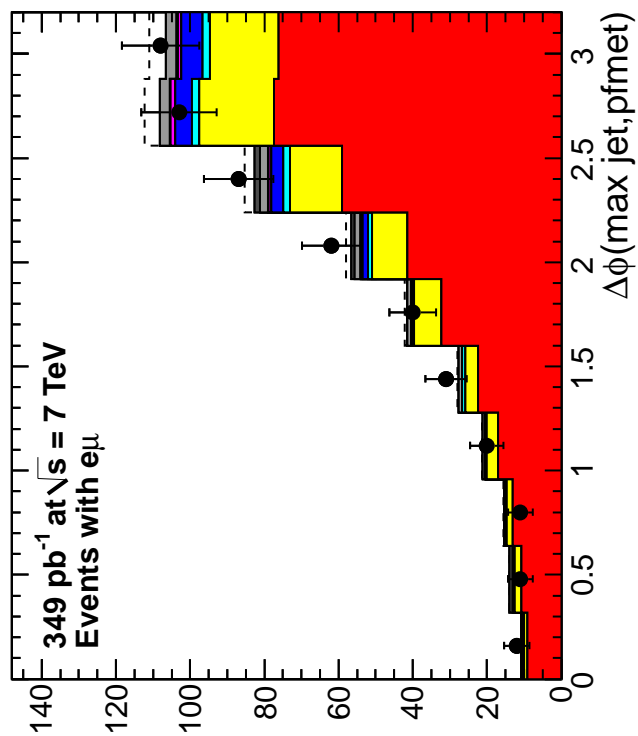
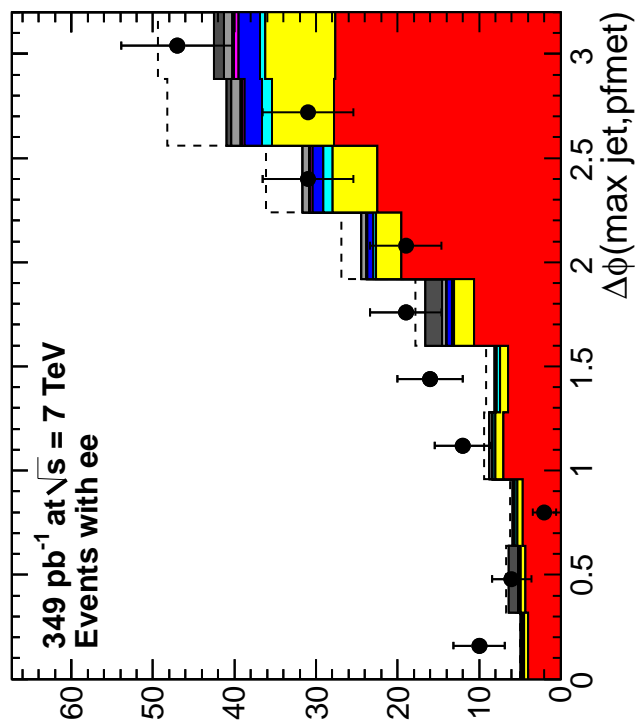


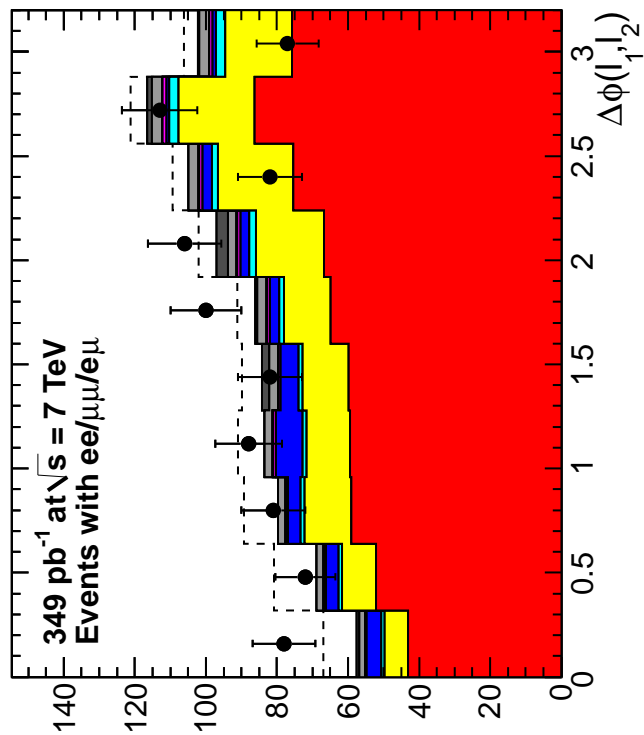
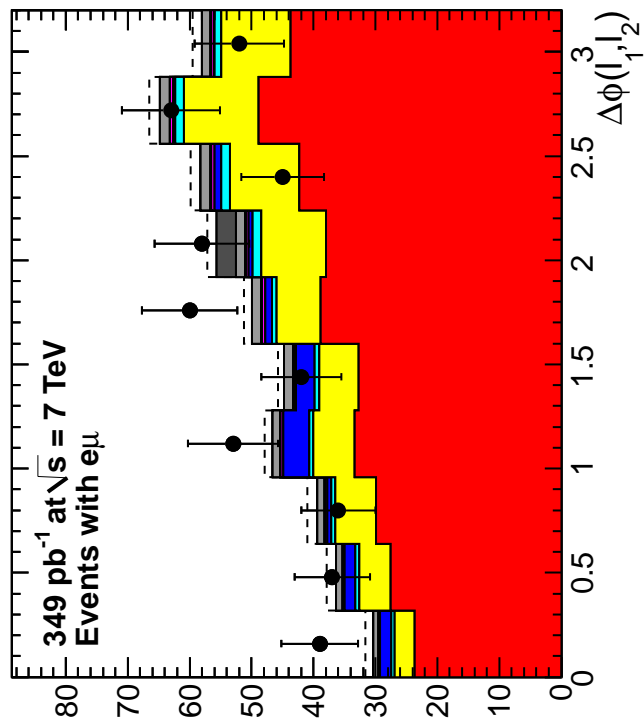
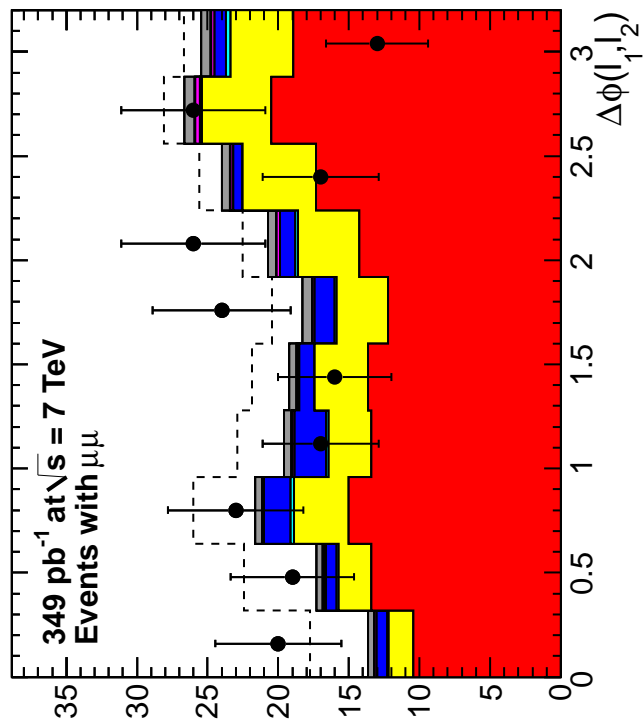
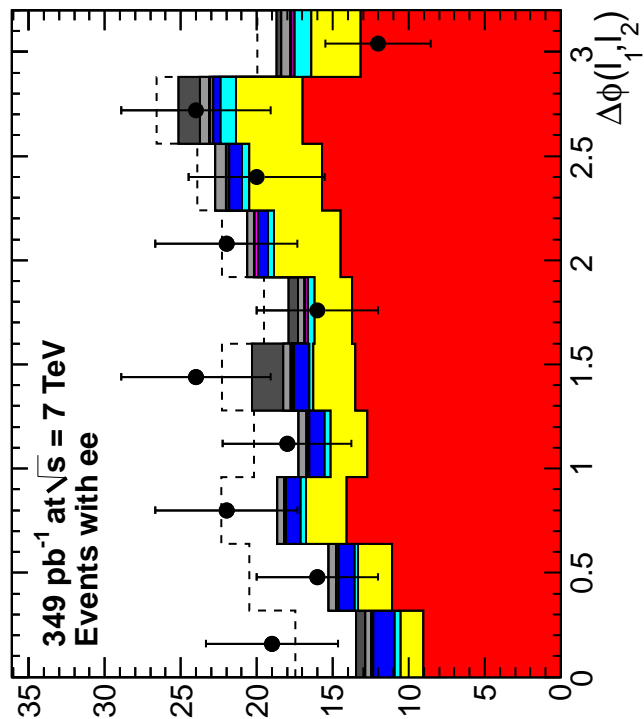


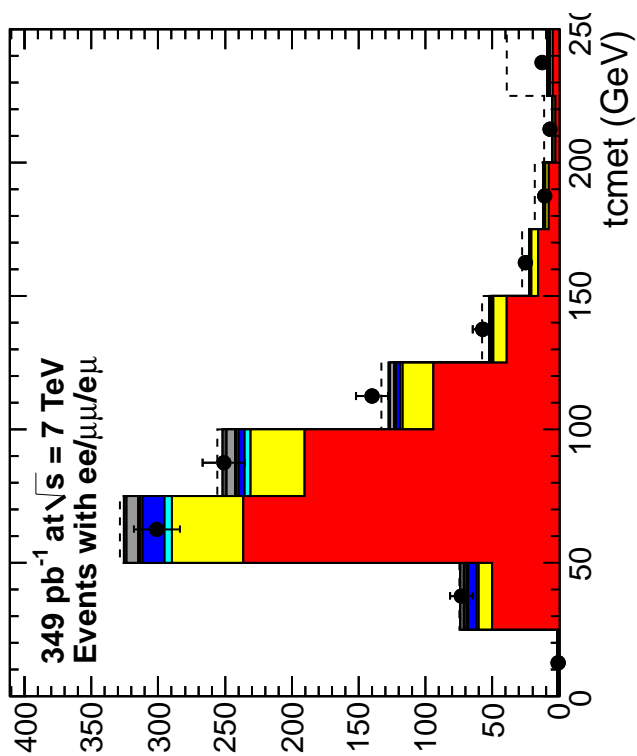
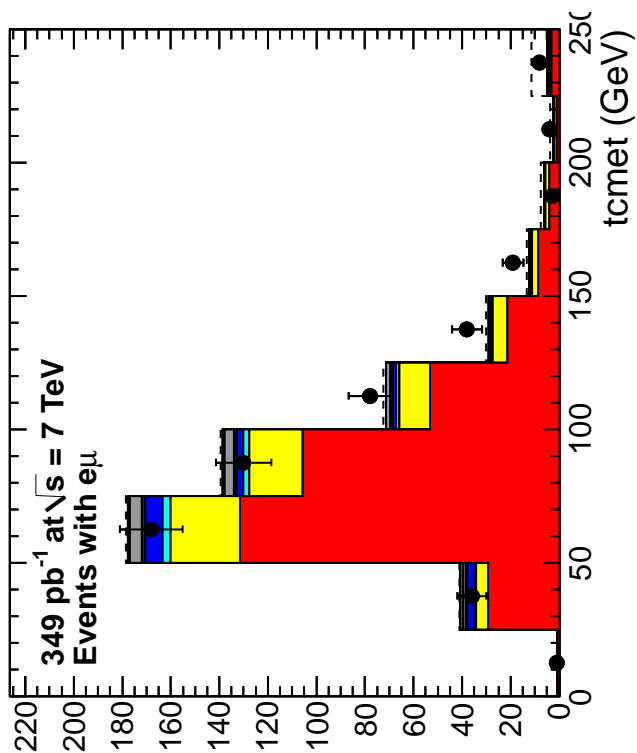
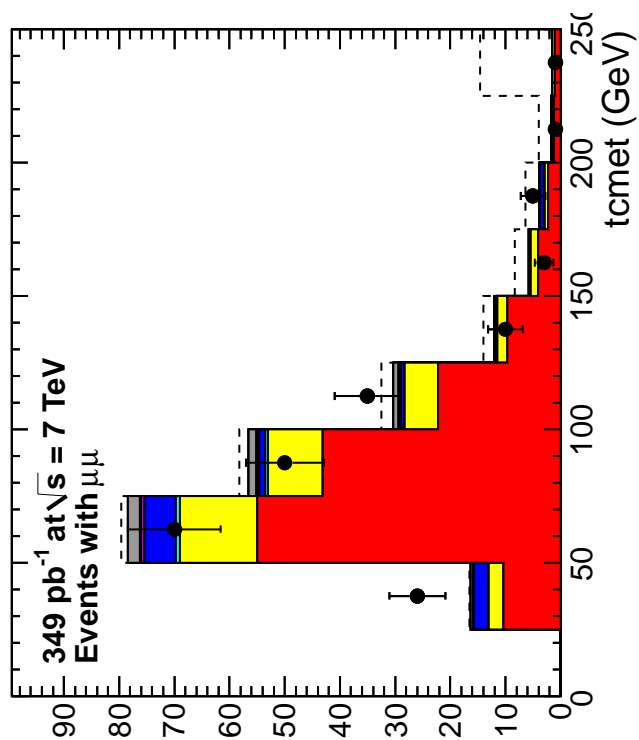
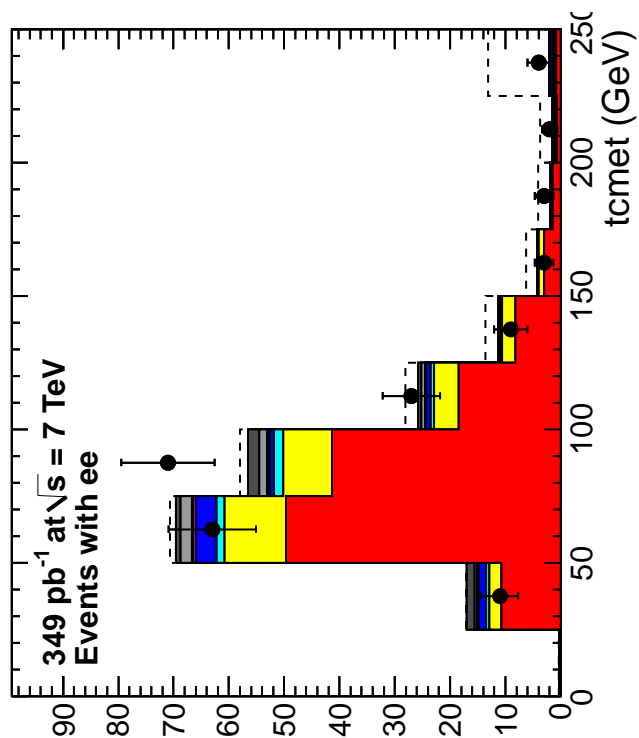


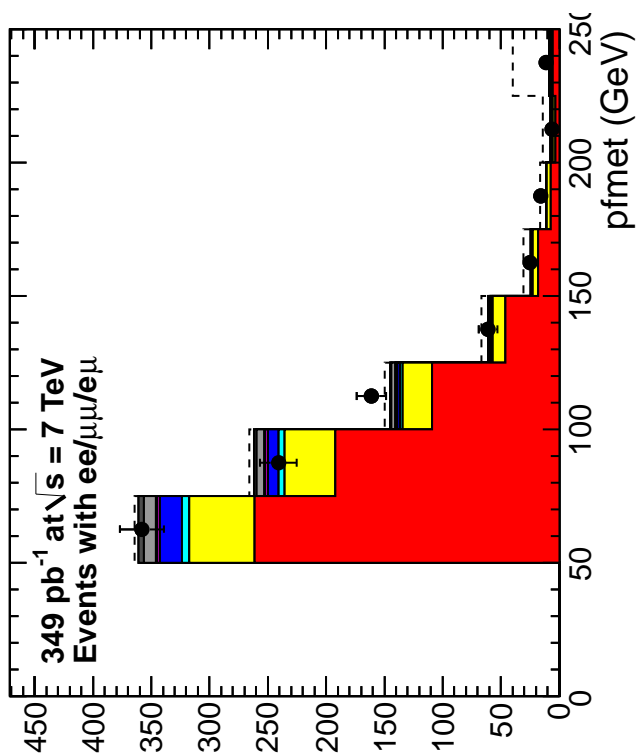
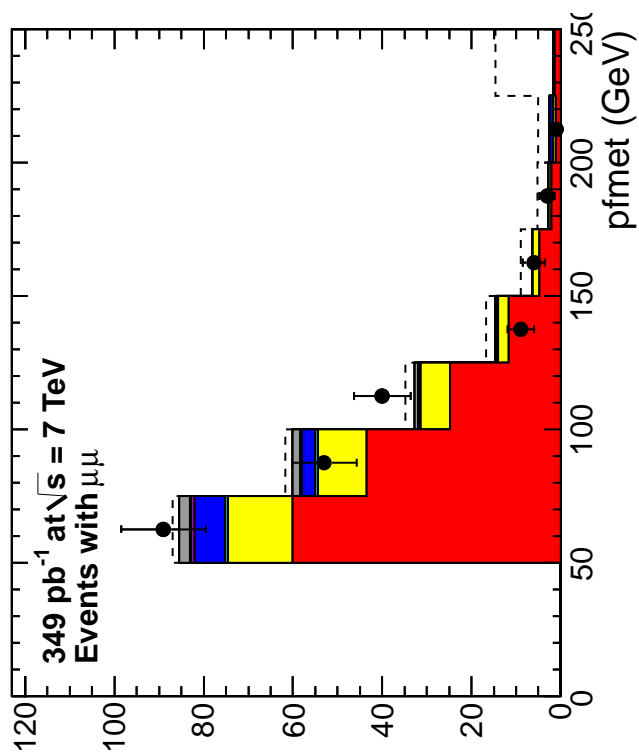
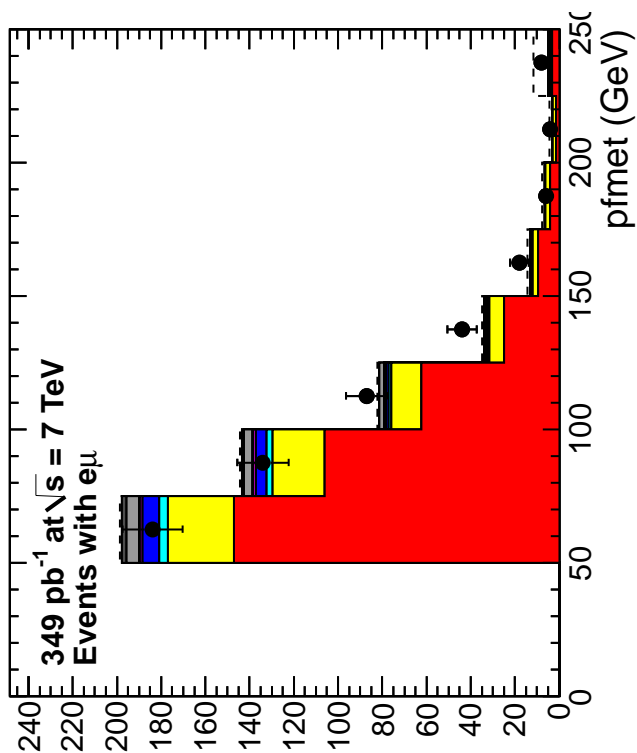
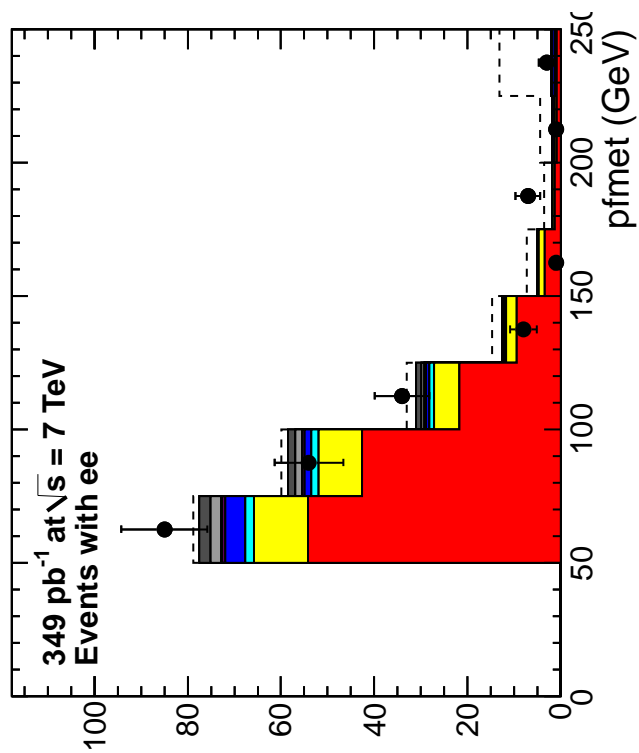


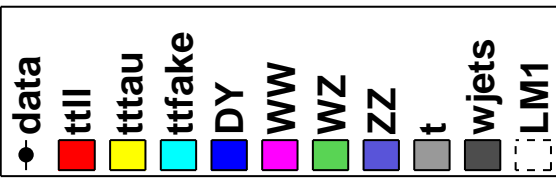
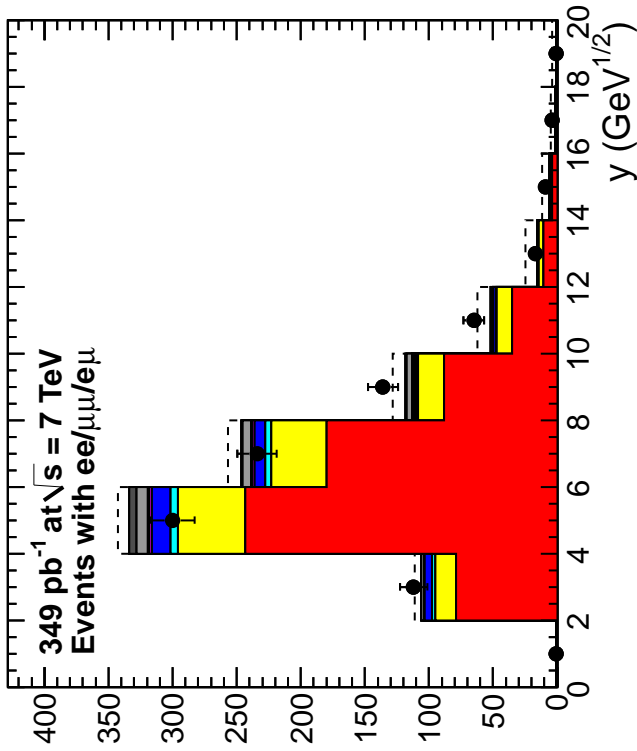
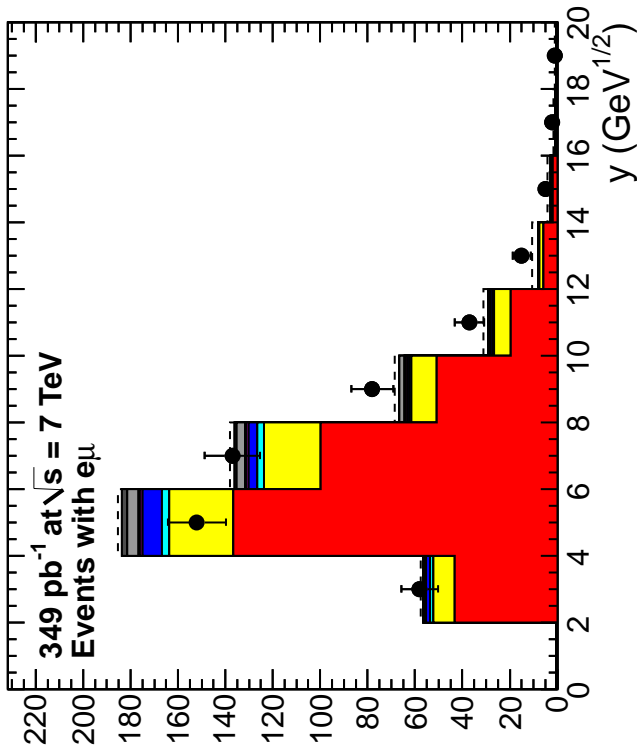
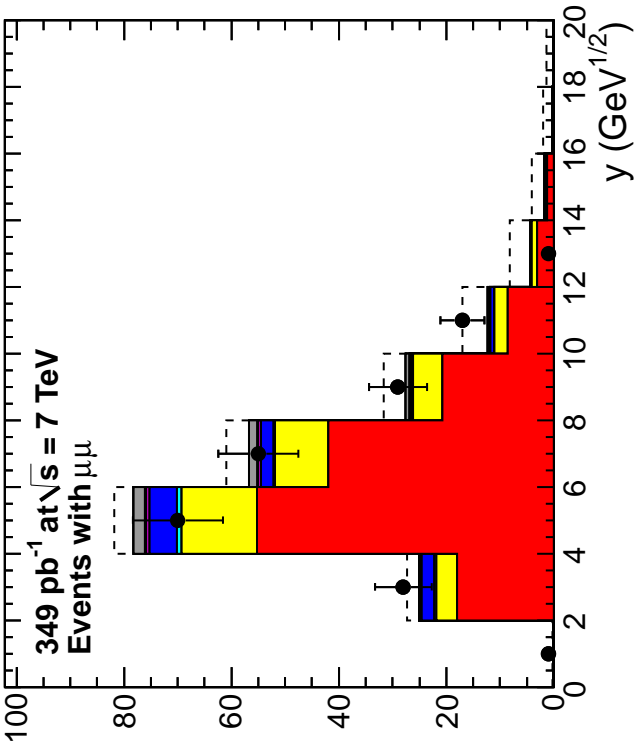
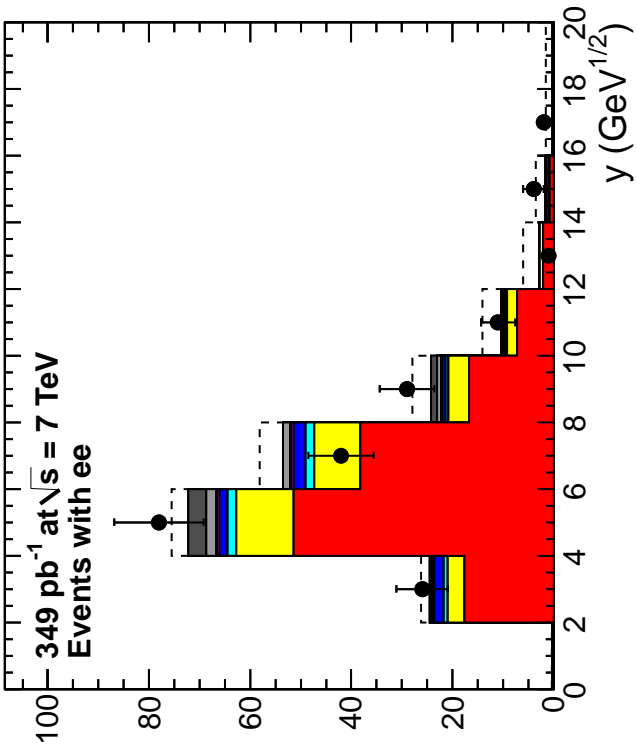


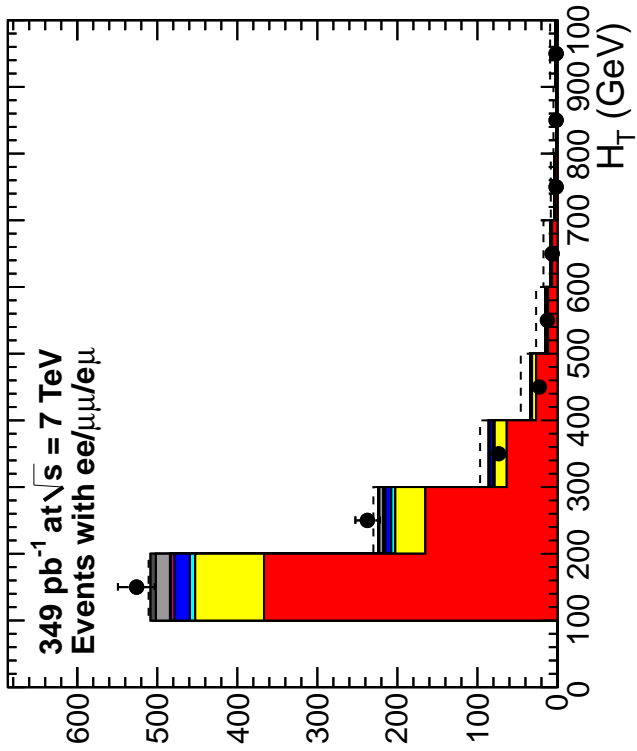
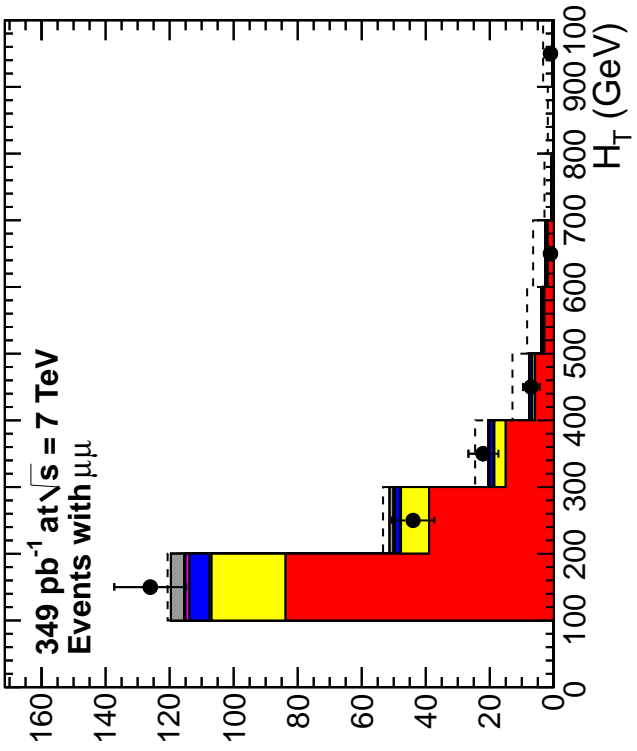
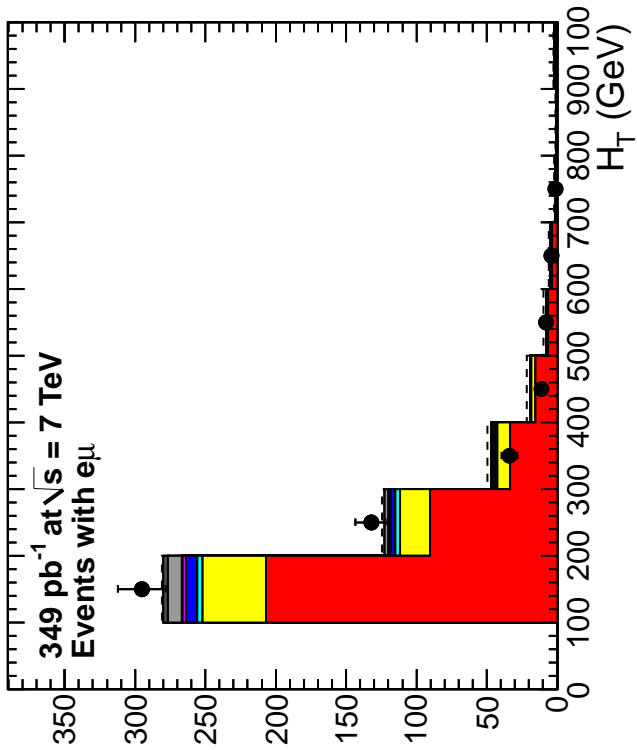
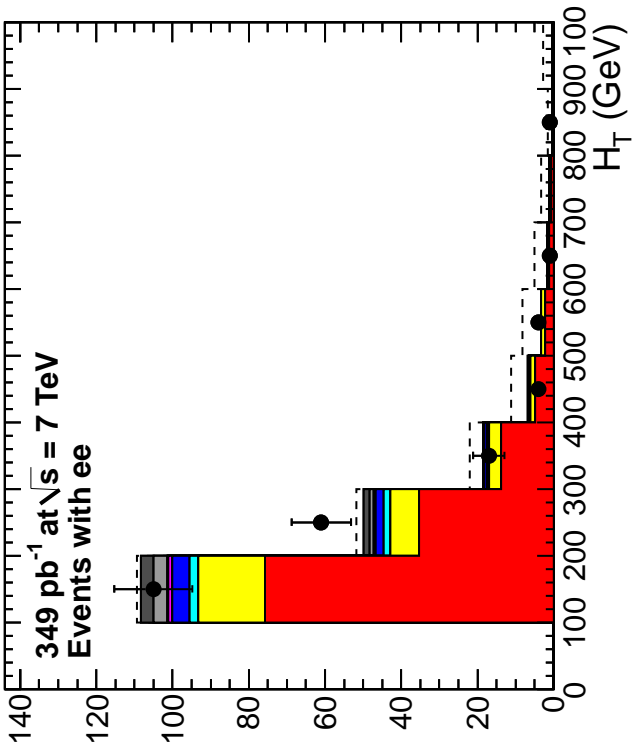


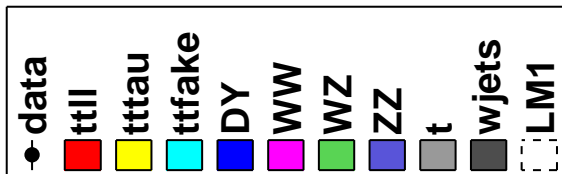
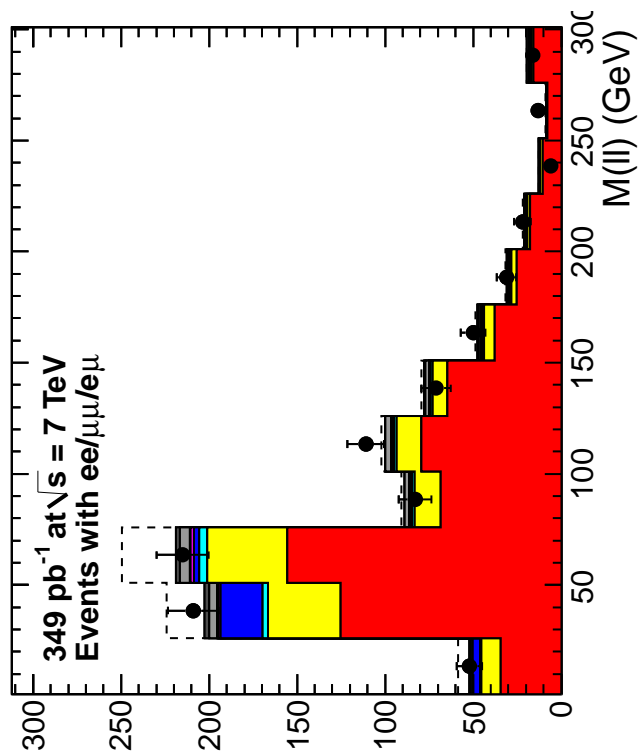
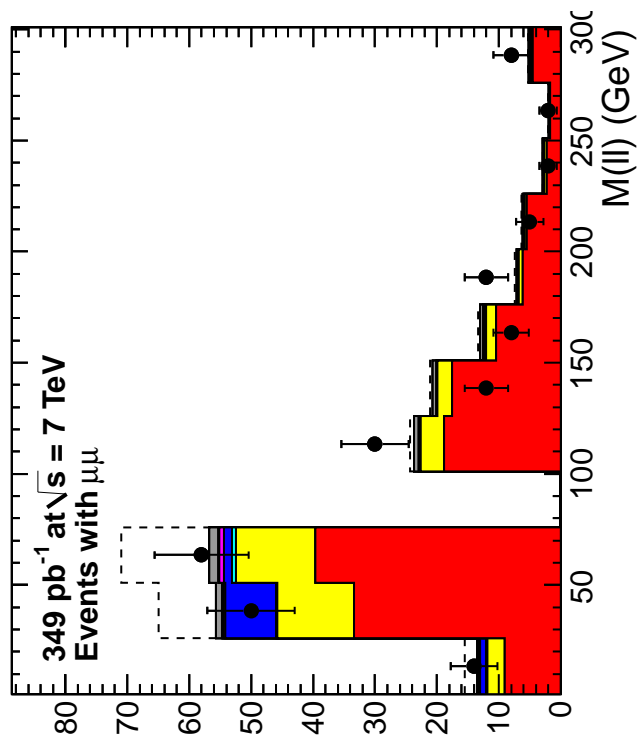
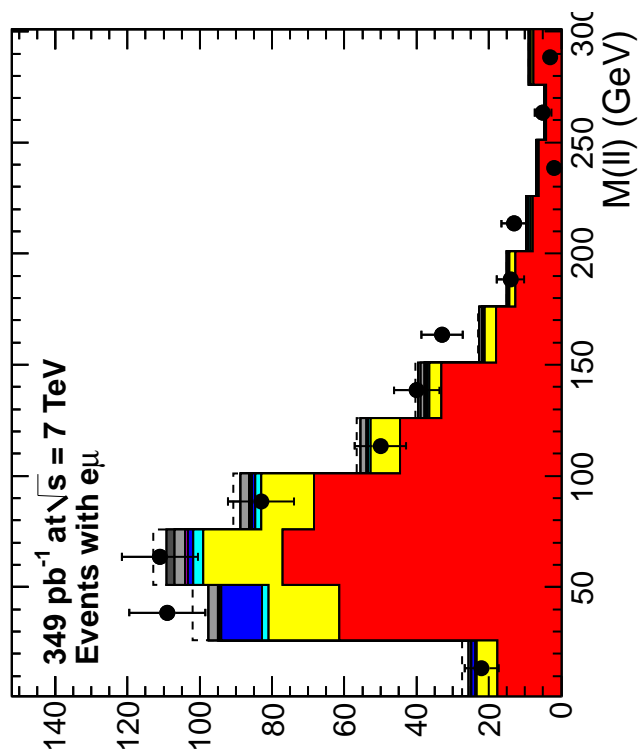
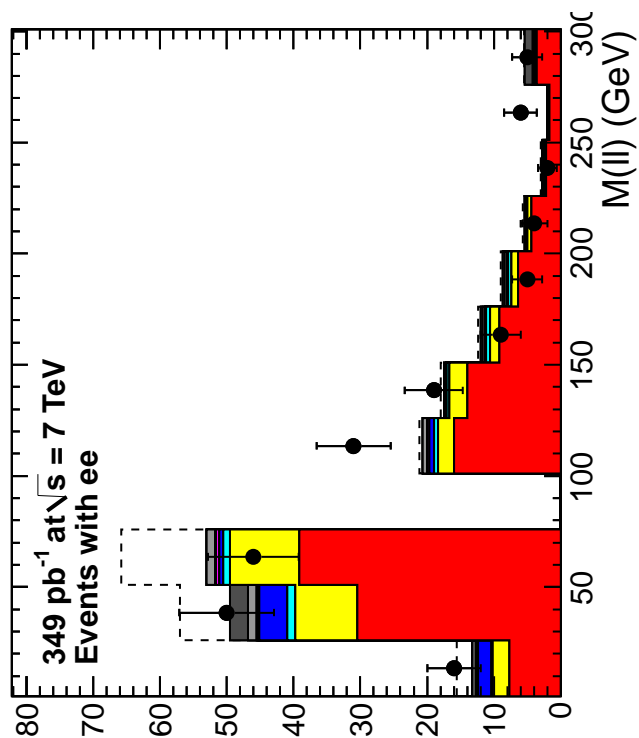


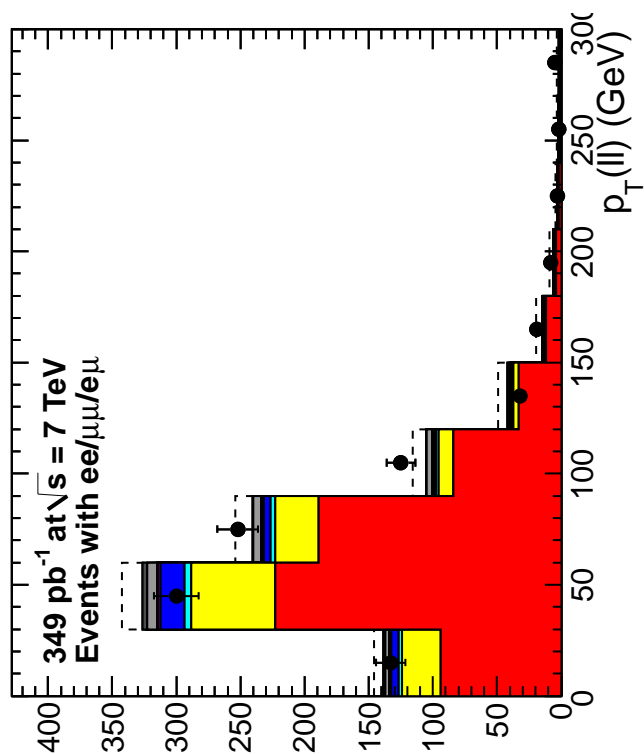
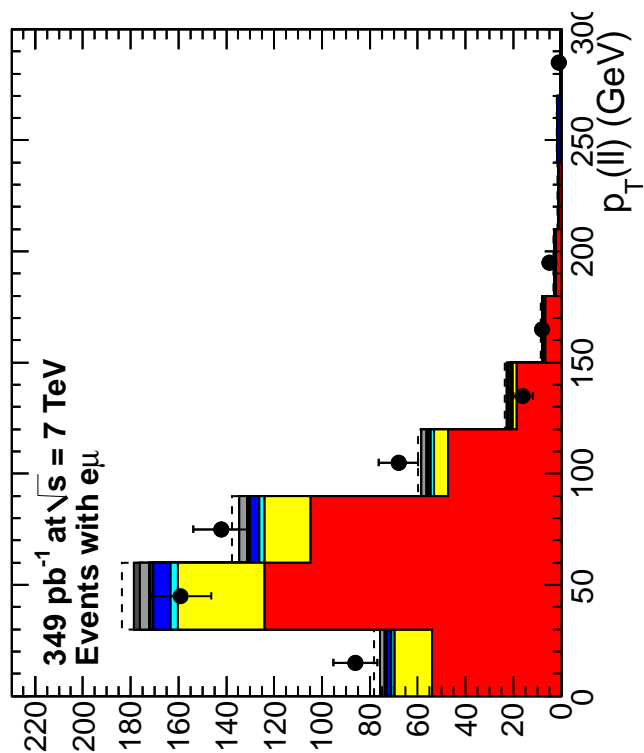
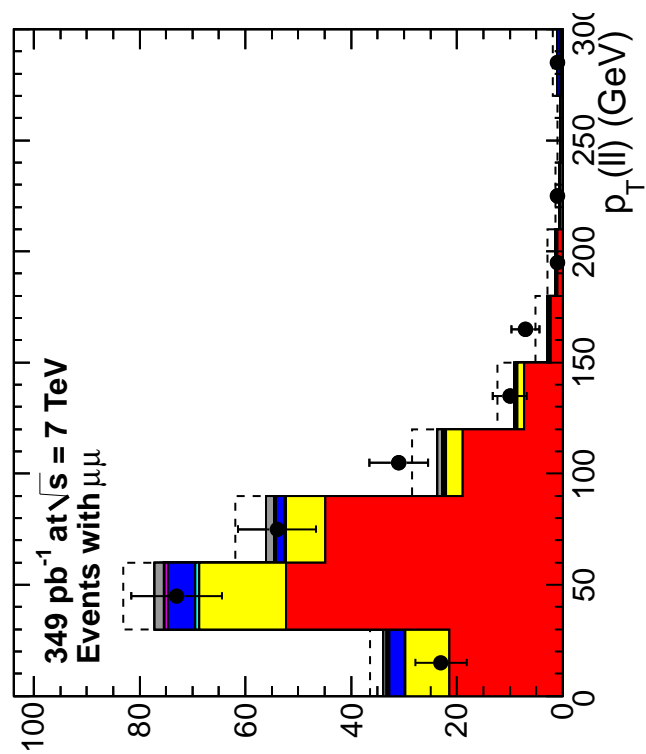
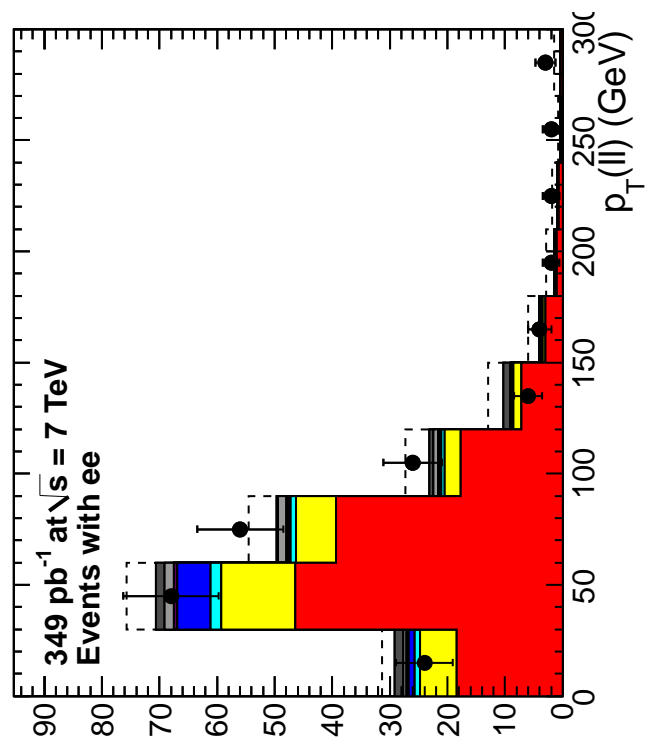


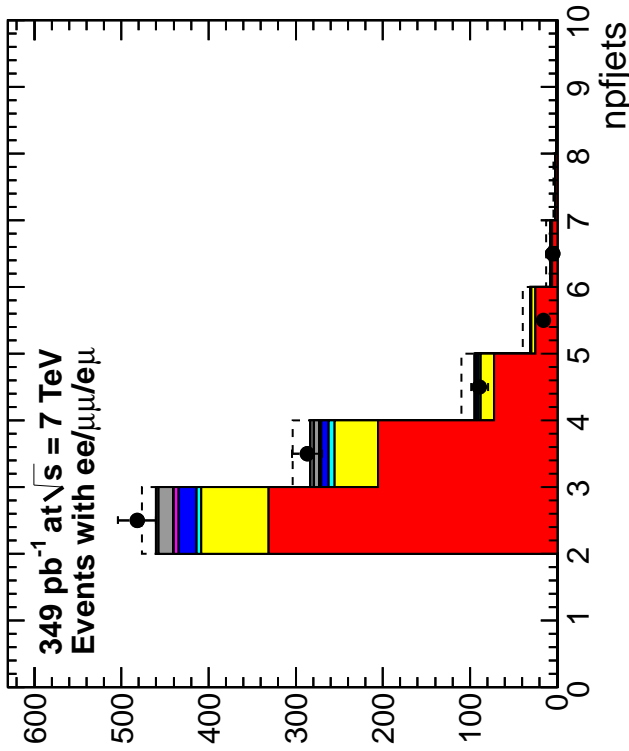
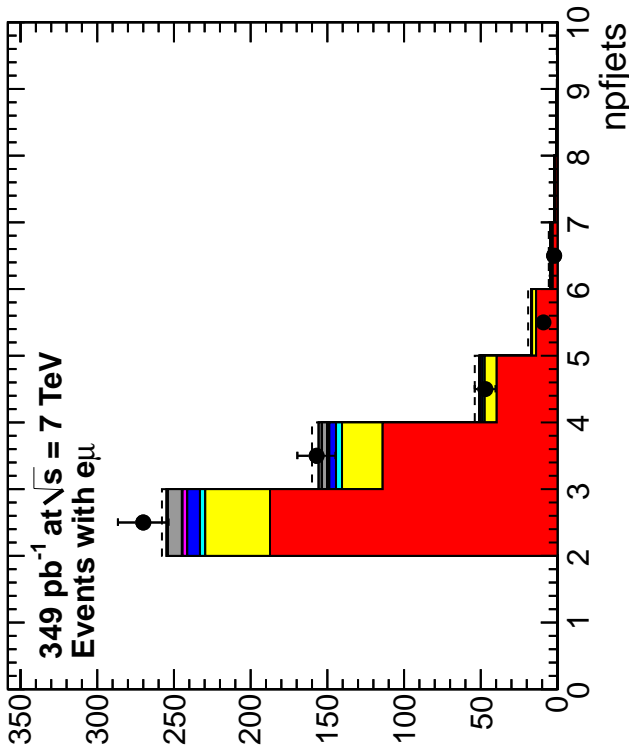
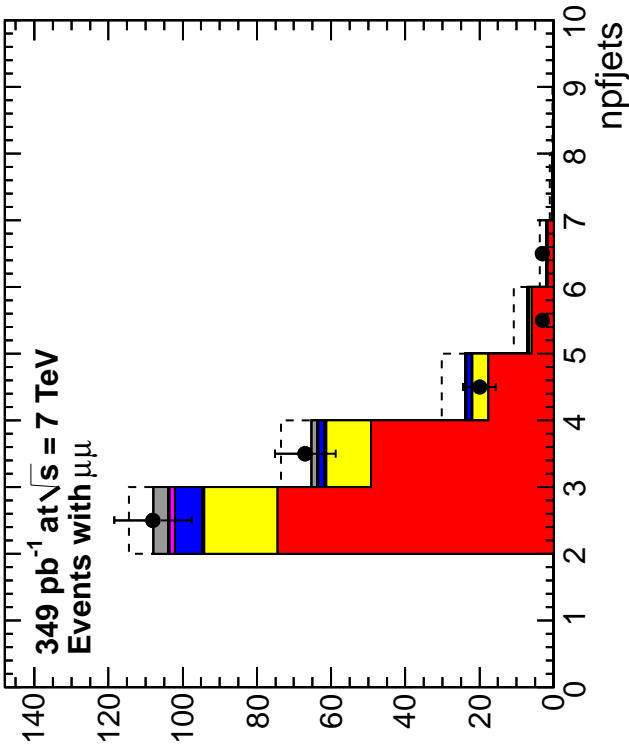
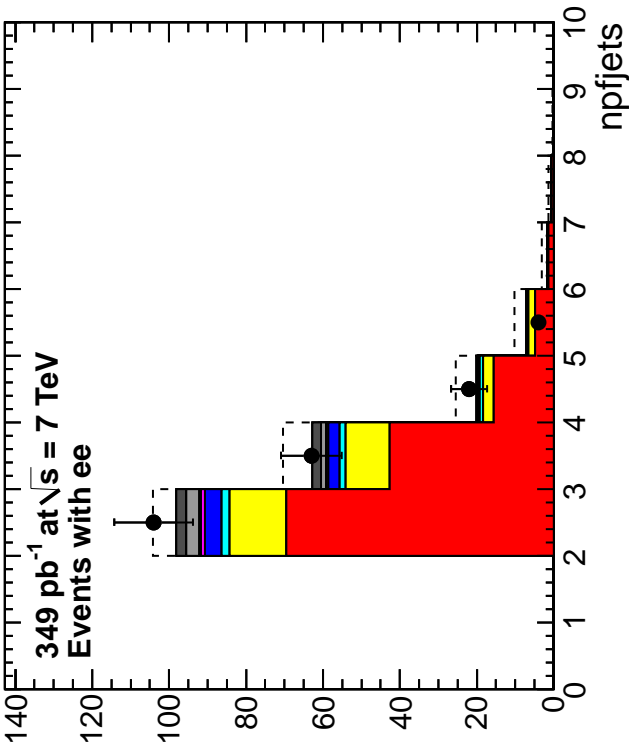


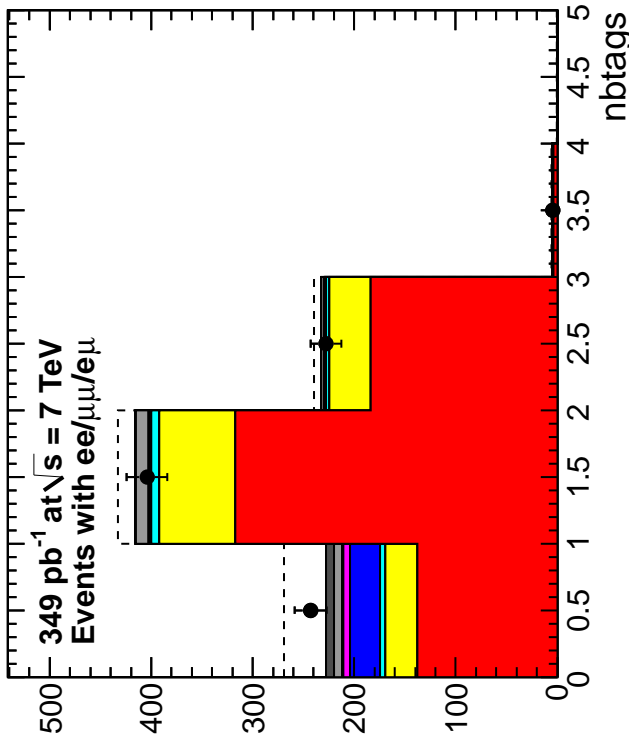
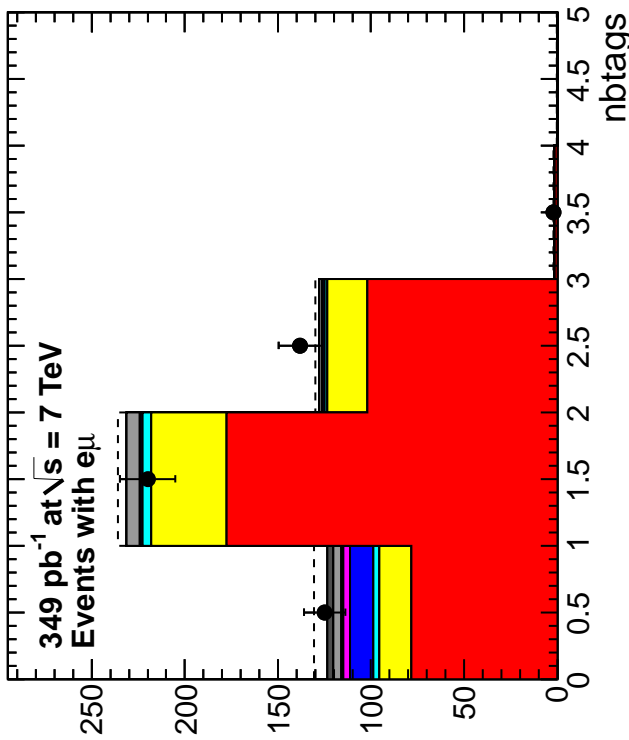
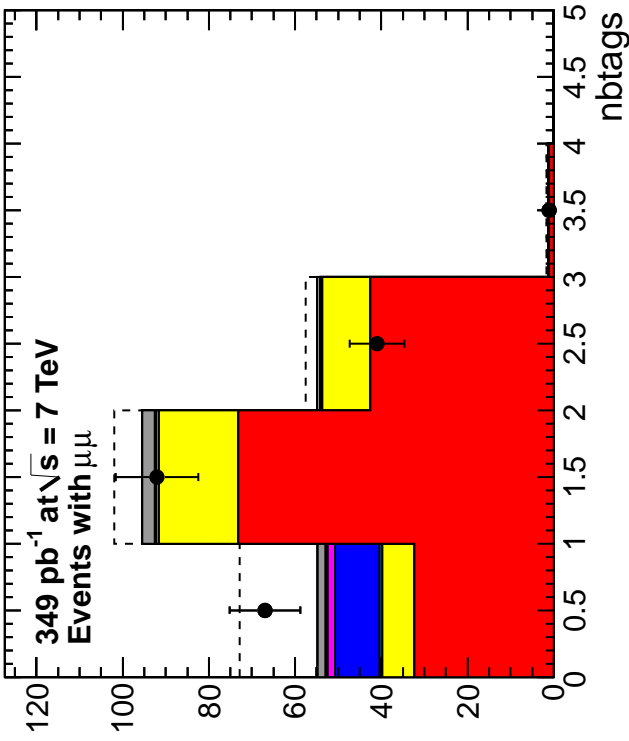
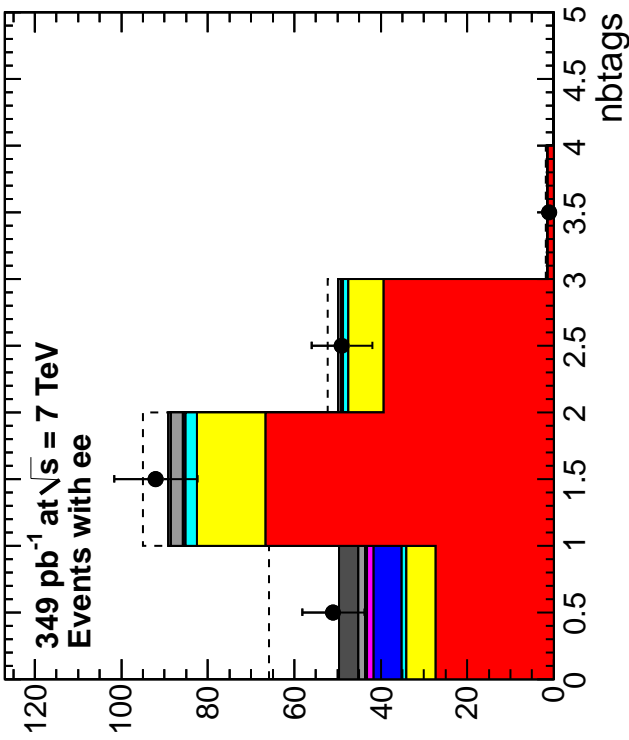


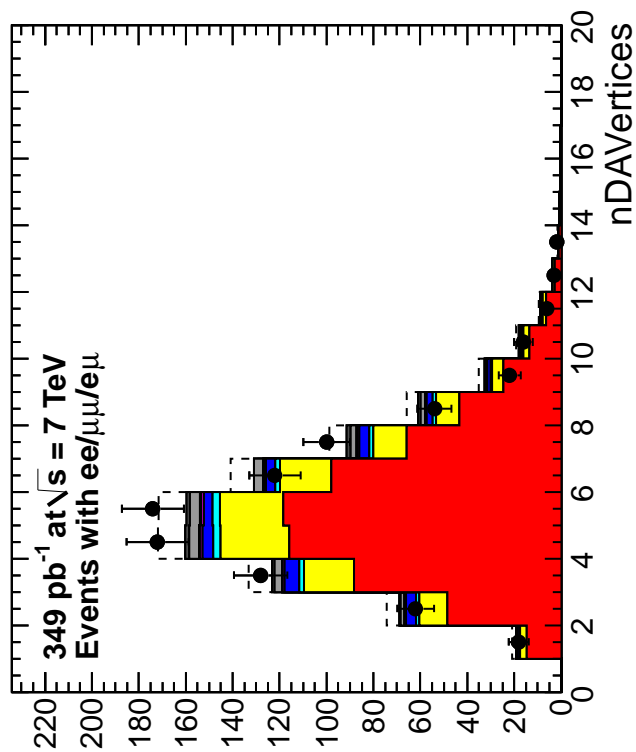
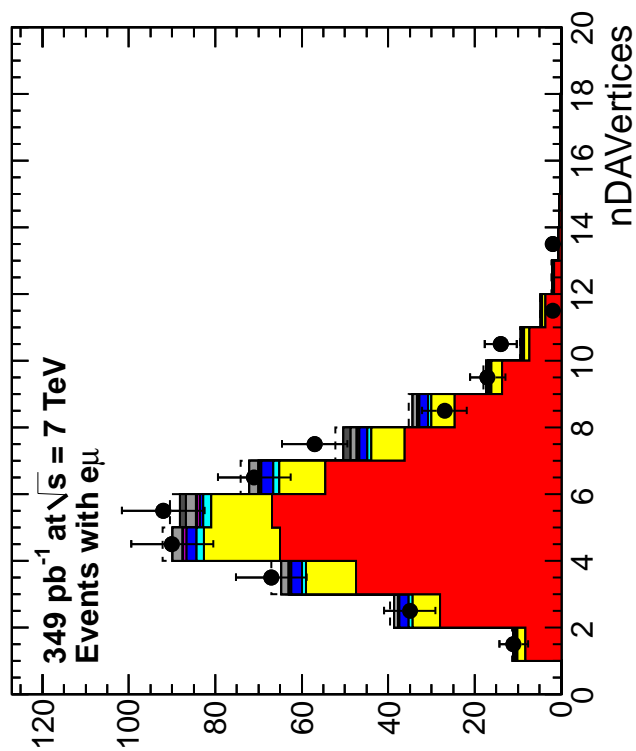
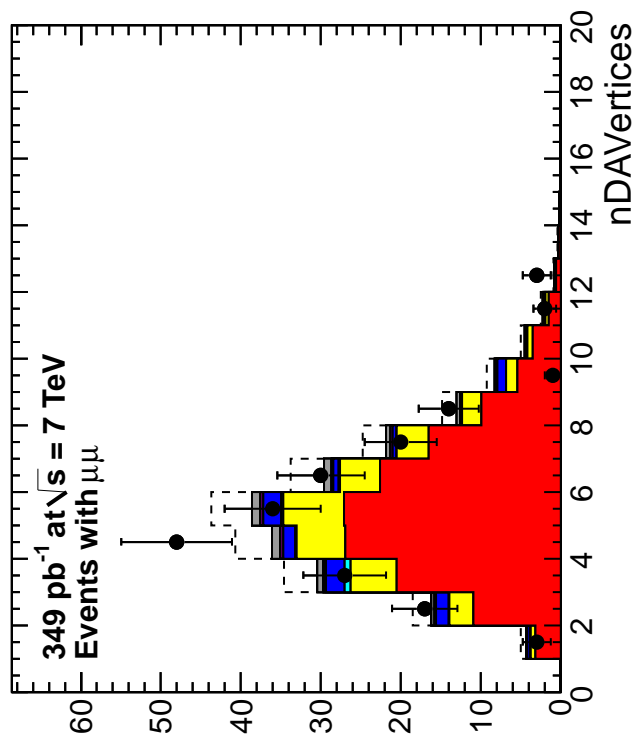
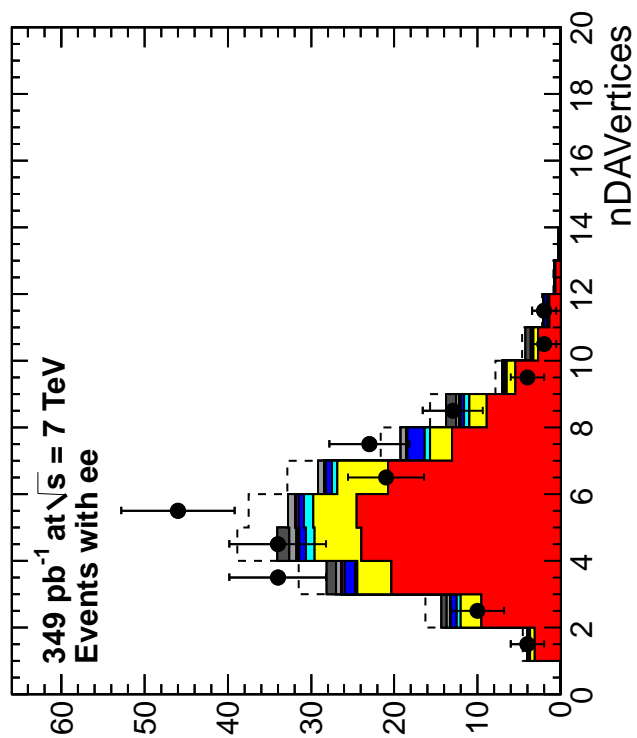


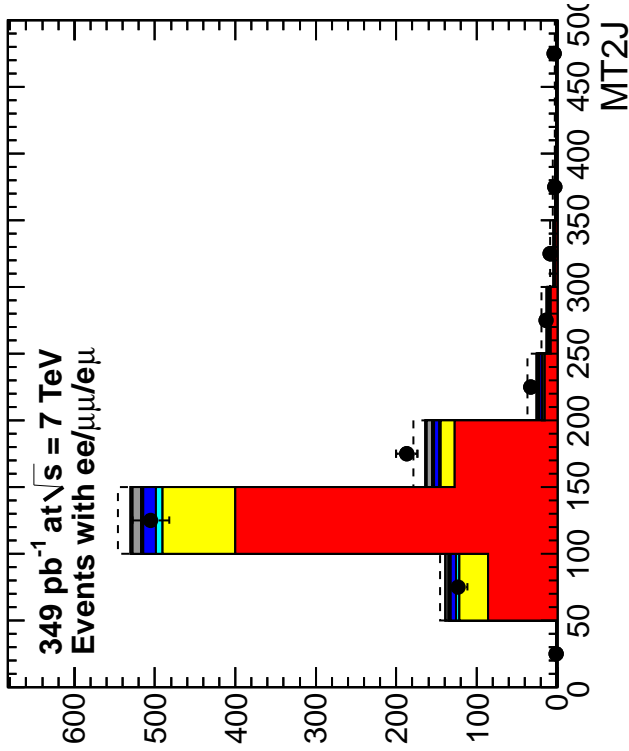
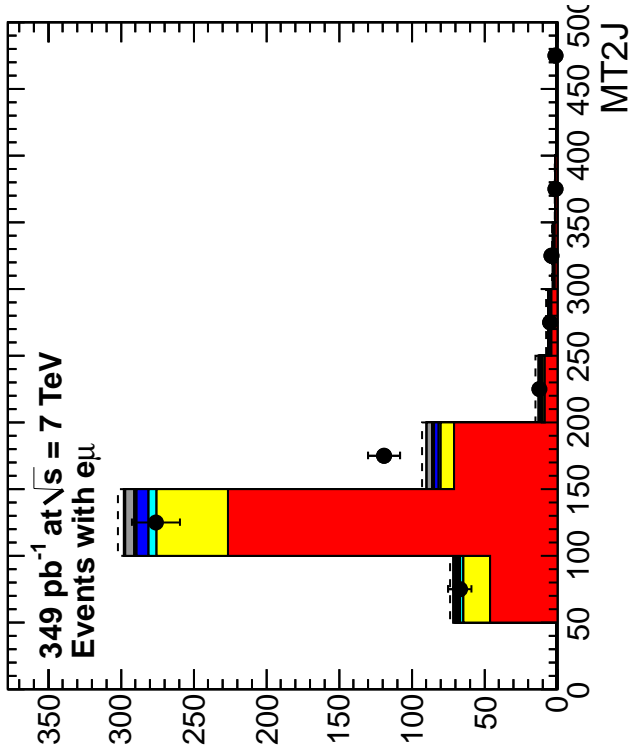
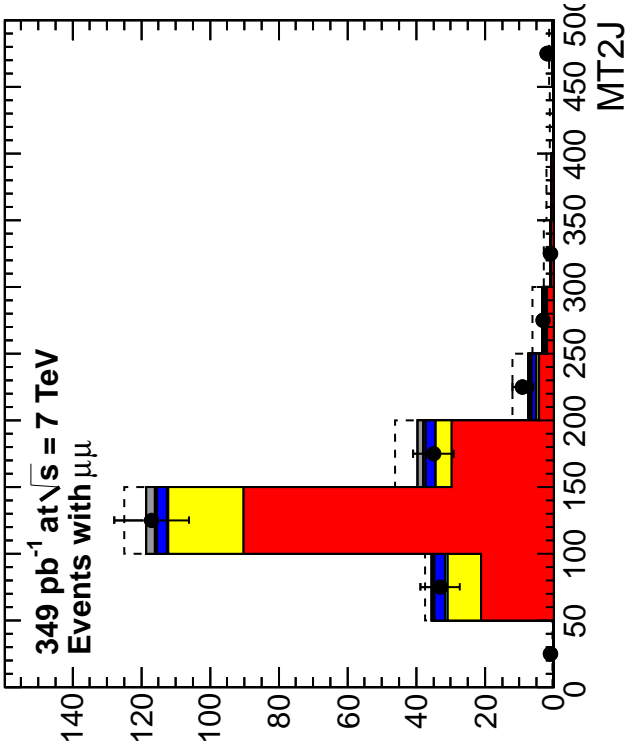
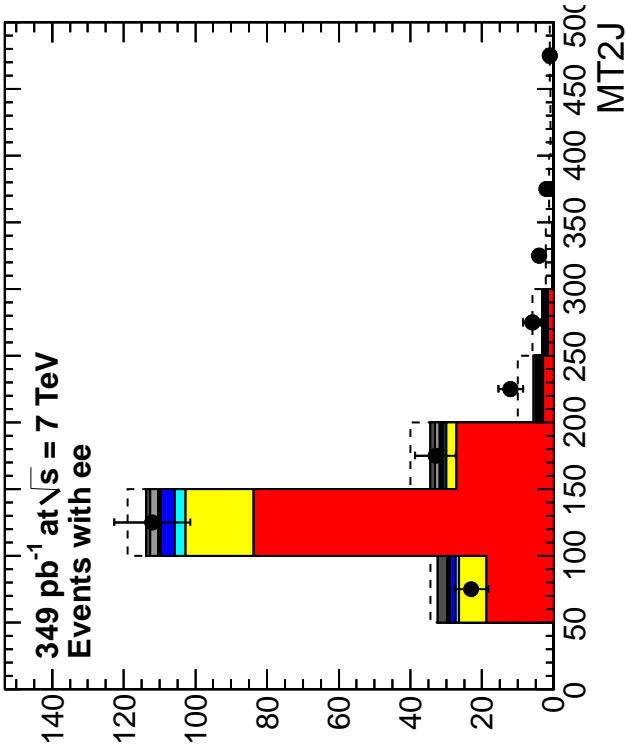


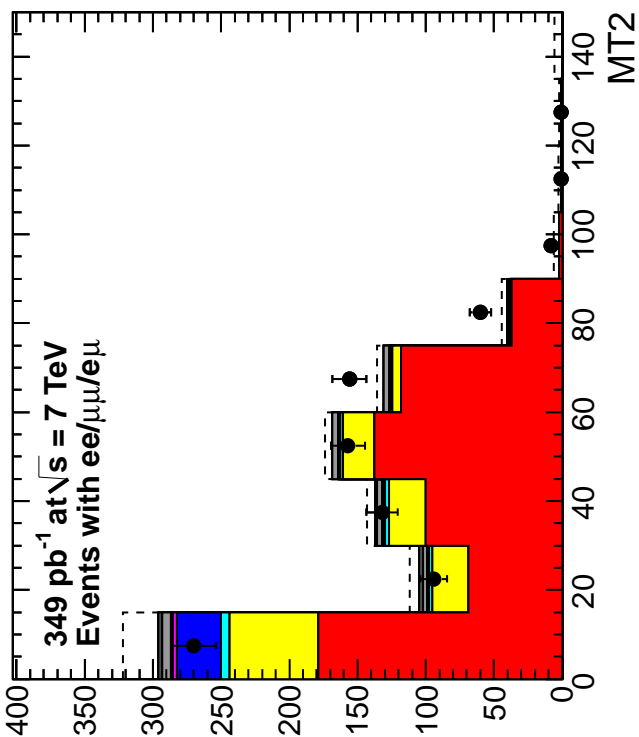
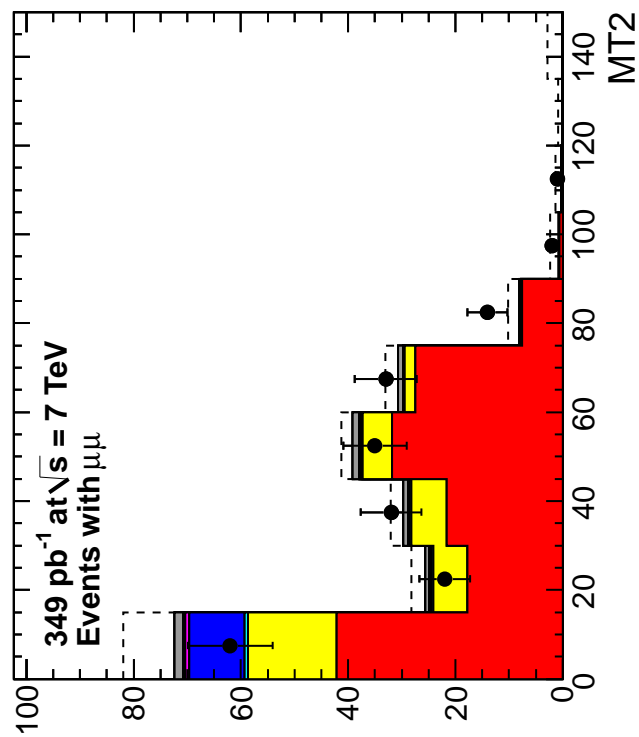
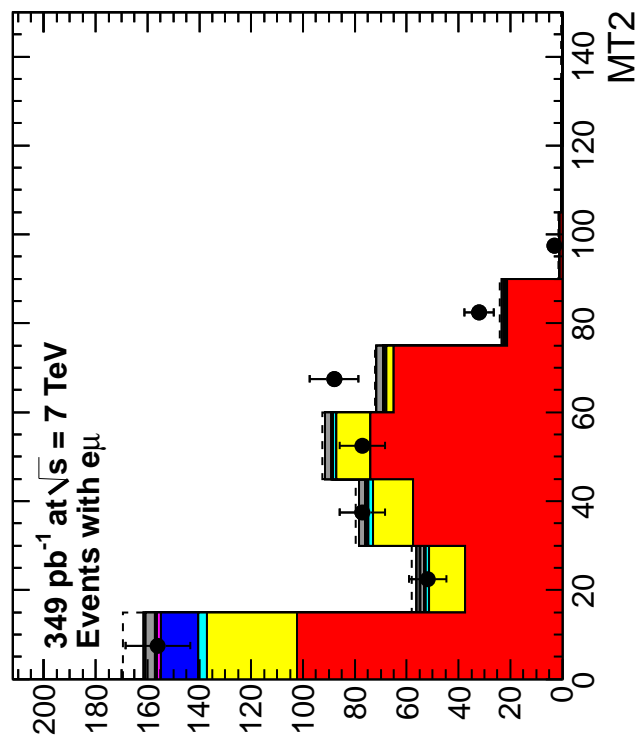
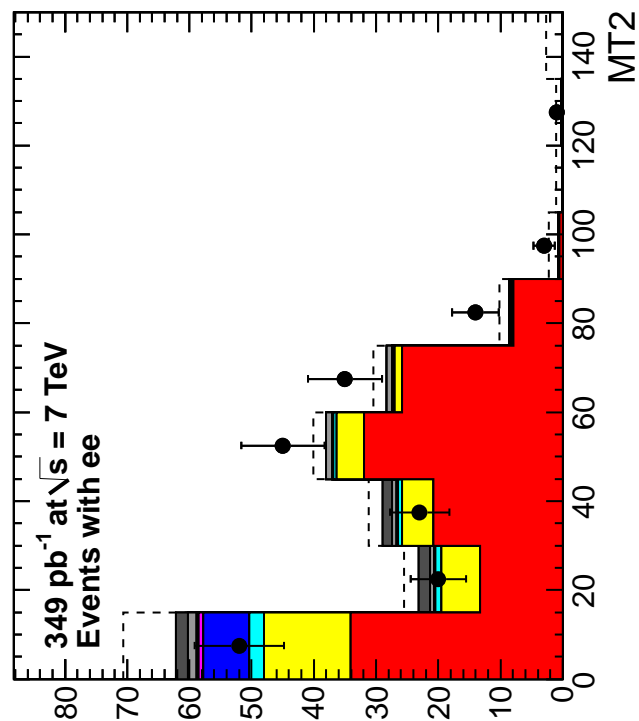


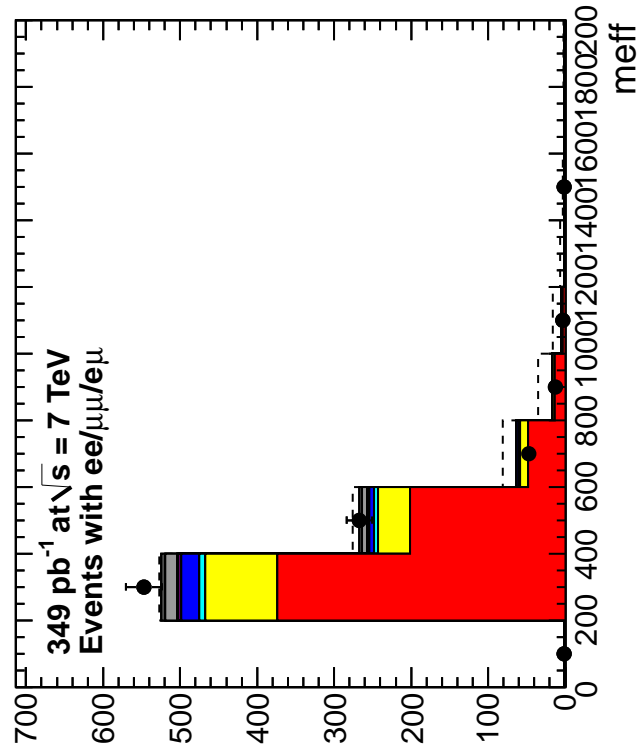
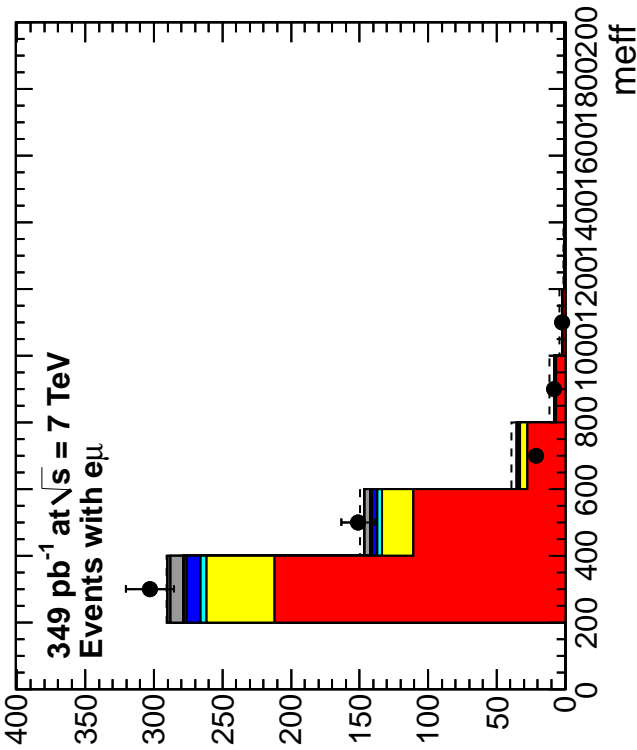
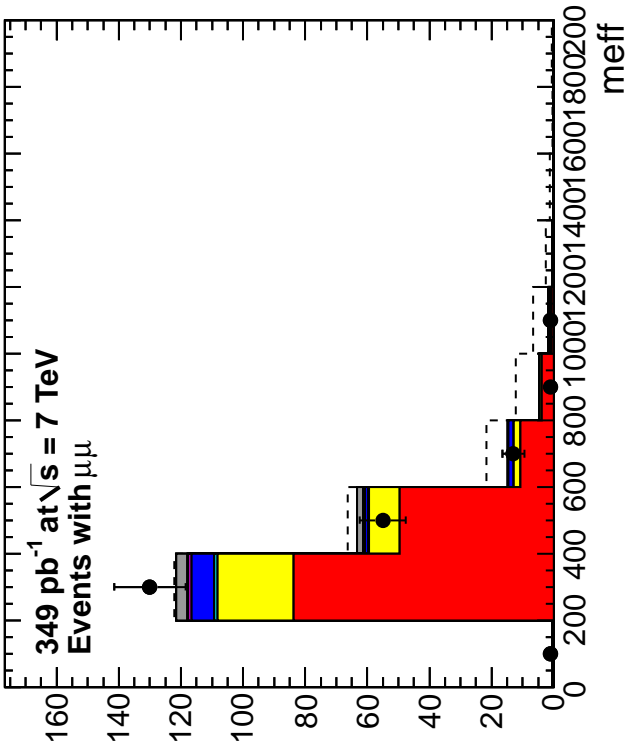
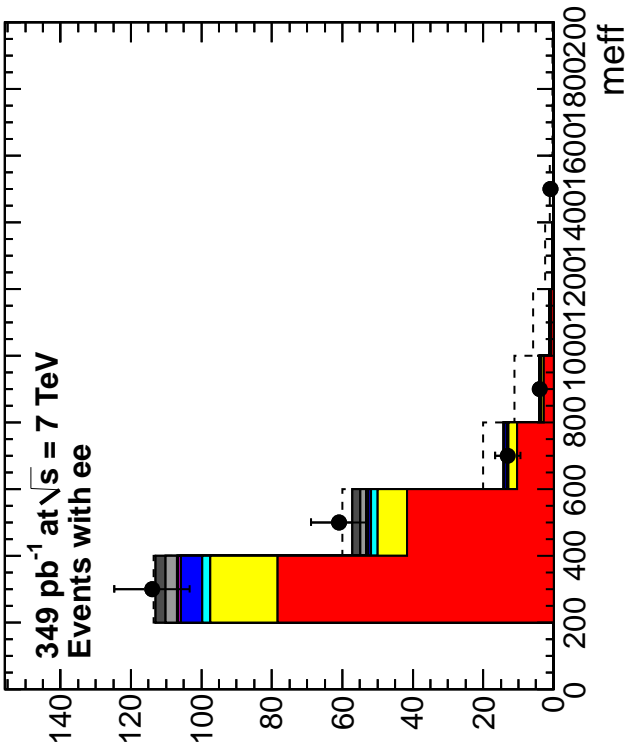












Appendix D Data/MC Comparison: 2010 Signal Region

Here we compare data and MC distributions for data passing the 2010 signal region requirements (preselection + $y > 8.5 \text{ GeV}^{1/2} + H_T > 300 \text{ GeV}$. The high p_T dilepton trigger data is used. We observe 19 events in this region.

All MC samples are scaled by an overall factor 1.12, the ratio of the observed data yield in the preselection region to the prediction from MC. For illustration purposes, we overlay the distributions from the LM1 SUSY benchmark point.

