Search for new physics in the 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, 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, I. Macneill, S. Jindariani

Fermi National Accelerator Laboratory, Batavia, Illinois

Abstract

We present the results of a search for new physics in the opposite-sign dileptons + jets + missing energy final state based on 204 pb^{-1} of 2011 CMS data.

Contents

1 Introduction

In this note we describe a search for new physics in the 2011 opposite sign isolated dilepton sample (ee, $e\mu$, and $\mu\mu$). 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 $Z \to ee$ and $Z \to \mu\mu$. 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 $P_T > 20$ GeV. In this analysis we lower the requirement on the second lepton to $P_T > 10$ 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 $P_T > 30$ GeV with $\mathrm{E_T^{miss}} > 30$ GeV (ee and $e\mu$) or $\mathrm{E_T^{miss}} > 20$ GeV ($e\mu$). We tighten the $\mathrm{E_T^{miss}}$cut to 50 GeV and we also require that the scalar sum of the P_T of all jets with $P_T > 30$ GeV be > 100 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 agreeement for the bulk of $t\bar{t}$ we move on to a search for new physics in the tail of the $t\bar{t}$.

- Our prejudice is that new physics would manifest itself in an excess of events with high $\mathrm{E_T^{miss}}$and significant hadronic activity. We define a a-priori search regions by tightening the $\mathrm{E_T^{miss}}$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 Section 9.1, 9.2 and 9.3).

2 Datasets, Triggers, Luminosity

We use the following datasets, and official golden json file (204 pb^{-1}):

Cert_160404-163869_7TeV_May10ReReco_Collisions11_JSON.txt

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. We use the high p_T dilepton trigger sample as the primary sample for this analysis, while the dilepton-H_T trigger sample is used as a crosscheck, and to search for a possible excess of events at low lepton p_T.

Currently we use data reconstructed in CMSSW 4_2_X and Spring11 MC reconstructed in CMSSW 3_11_X. We will update the MC samples 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`
  - Dilepton-H_T cross trigger sample
    - `ElectronHad_Run2011A-May10ReReco-v1_AOD`
    - `MuHad_Run2011A-May10ReReco-v1_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

3 Event Preselection

The purpose of the preselection is to define a data sample rich in $t\bar{t} \rightarrow$ 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 the trigger has large inefficiency below this threshold. Finally $\mathrm{E_T^{miss}} > 50$ GeV (we use pfmet). More details are given in the subsections below.

3.1 Event Cleanup

- Require at least one good deterministic annealing (DA) vertex
  - not fake
  - ndof > 4
  - $|\rho| < 2$ cm
  - $|z| < 24$ cm.

3.2 Muon Selection

Muon candidates are RECO muon objects passing the following requirements:

- $p_T > 5$ 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 $d_0 < 200$ μm and $d_z < 1$ cm
- $Iso \equiv E_T^{\mathrm{iso}}/p_T < 0.15$, 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 $\mathrm{E_T^{miss}}$
- **LIST FO DEFINITIONS HERE?**

3.3 Electron Selection

Electron candidates are RECO GSF electrons passing the following requirements:

- $p_T > 10$ 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:
  - $\sigma_{i\eta i\eta} < 0.01$ (EB), 0.03 (EE)
  - $\Delta\phi < 0.15$ (EB), 0.10 (EE)
  - $\Delta\eta < 0.007$ (EB), 0.009 (EE)
  - $H/E < 0.1$ (EB), 0.075 (EE)
- Impact parameter with respect to the first DA vertex $d_0 < 400$ μm and $d_z < 1$ cm.
- $Iso \equiv E_T^{\rm iso}/p_T < 0.15$. $E_T^{\rm iso}$ 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 $Dist < 0.02$ cm and $\Delta \cot\theta < 0.02$ is vetoed[6].
- Cleaning for ECAL spike (aka Swiss-Cross cleaning) has been applied at the reconstruction level (CMSSW 38x).
- **LIST FO DEFINITIONS HERE?**

3.4 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 GeV, since this kinematical region is not well reproduced in CMS Monte Carlo and to remove Upsilons.

In addition, we remove $Z \to \mu\mu\gamma$ 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.

3.5 Trigger Selection

We do not make any requirements on HLT bits in the Monte Carlo. Instead, as discussed in Section 4, a trigger efficiency weight is applied to each event, based on the trigger efficiencies measured on data (see Sec. 4).

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 cross trigger.

- High p_T dilepton trigger sample
  - `HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL`
  - `HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL`
  - `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

4 Trigger efficiency

The DoubleElectron triggers have an efficiency which is close to 100%. However, the efficiency for a single leg to pass the muon trigger is approximately 95% (**NEED A REF OR SOME PLOTS IN BACKUP FOR THIS**), causing the DoubleMu (El-Mu) triggers to have an efficiency of approximately 90% (95%). Unless otherwise specified, these trigger efficiencies are applied to the MC. **IS THIS SUFFICIENT? DO WE WANT A MORE DETAILED TRIG EFFICIENCY MODEL, OR MORE ACCURATE TRIG EFFICIENCIES?**

WILL ADD HT TURN ON CURVES HERE.

5 Dilepton Yields

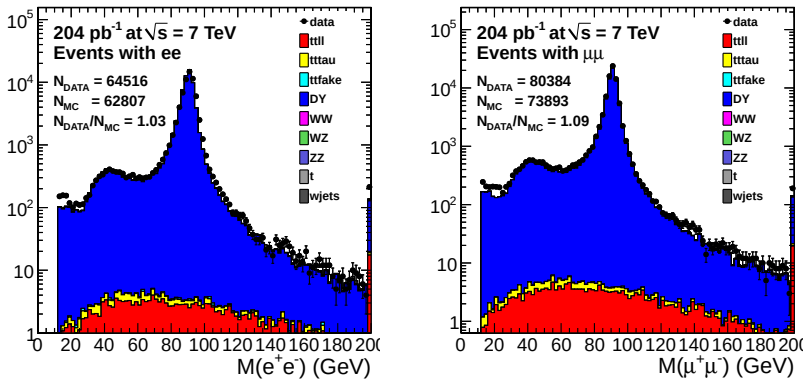

Figure 1: Distributions of dilepton mass in data and MC, in the ee channel (left) and $\mu\mu$ channel (right). The quoted yields refer to events inside the Z mass window taken as 76–106 GeV.

Here we show the data and MC yields with 2 selected leptons, displayed in Fig. 1. We observe a 3% excess in the ee channel and a 9% excess in the $\mu\mu$ channel. **not really sure whats going on here, or if we should comment further on the discrepancy...**

6 Preselection yields

The data yields and the MC predictions for the dilepton trigger sample are given in Table 1. We also look for an excess of low dilepton 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). Since we do not observe an excess of events at low lepton p_T (in fact we observe a slight deficit), we focus our attention on the dilepton trigger sample. Finally, we verify in Table 3 that the data samples collected with high p_T dilepton

triggers and lepton-H_T cross triggers give consistent yields, after including the requirements $H_T > 200$ GeV and lepton $p_T >$ (20,10) GeV.

The MC yields are normalized to 204 pb^{-1} using the cross-sections from Reference [7]. 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. Contributions for $t\bar{t} \rightarrow$ fake and $W +$ jets with 1 lepton not originating from W/Z decay (fake lepton) and the contributions from QCD multijet events with 2 fake leptons (single fakes and double fakes, respectively) are estimated with the data-driven fake rate method [4]. **NEED TO APPLY SPILLAGE CORRECTION HERE.** The DY contribution is dominated by DY$\rightarrow \tau^+\tau^-$, and we have verified with the data-driven $R_{out/in}$ method [4] that the contributions from DY$\rightarrow e^+e^-$ and 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.

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, which is most pronounced in the $e\mu$ channel.

Table 1: High p_T dilepton trigger data and Monte Carlo yields for the preselection (njets ≥ 2, H_T>100 GeV, $\mathrm{E_T^{miss}} > 50$ GeV, lepton $p_T >$ (20,10) GeV). $t\bar{t} \rightarrow \ell^+\ell^-$ corrresponds to dileptonic $t\bar{t}$ in which both W's decay via $W \rightarrow e\nu$ or $W \rightarrow \mu\nu$, $t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$corresponds to dileptonic $t\bar{t}$ in which both leptons come from W decay and at least 1 lepton is from $W \rightarrow \tau \rightarrow \ell$. **NO SPILLAGE CORRECTION APPLIED HERE. MC TOTAL FAKE PREDICTION IS 12 EVENTS, FR GIVES ~4X HIGHER...**

Sample	ee	$\mu\mu$	$e\mu$	tot
$t\bar{t} \rightarrow \ell^+\ell^-$	70.3 ± 1.7	77.8 ± 1.7	187.3 ± 2.7	335.4 ± 3.6
$t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$	16.1 ± 0.8	19.5 ± 0.9	42.0 ± 1.3	77.7 ± 1.8
single fakes	8.5 ± 2.2	17.0 ± 3.1	22.7 ± 3.7	48.3 ± 5.3
double fakes	0.0 ± 0.0	0.4 ± 0.3	1.0 ± 0.6	1.5 ± 0.6
DY	3.8 ± 1.0	5.7 ± 1.3	7.9 ± 1.5	17.4 ± 2.2
W^+W^-	0.8 ± 0.1	0.9 ± 0.1	2.0 ± 0.1	3.7 ± 0.2
$W^\pm Z^0$	0.2 ± 0.0	0.2 ± 0.0	0.4 ± 0.0	0.8 ± 0.0
Z^0Z^0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.2 ± 0.0
t	2.6 ± 0.1	2.9 ± 0.1	6.9 ± 0.2	12.5 ± 0.3
tot SM MC	102.5 ± 3.1	124.5 ± 3.9	270.4 ± 5.0	497.3 ± 7.1
data	113	122	313	548
LM1	13.1 ± 0.3	14.6 ± 0.3	7.5 ± 0.3	35.1 ± 0.6
LM3	3.0 ± 0.1	3.7 ± 0.1	5.3 ± 0.1	12.0 ± 0.2

Table 2: Dilepton-H_T trigger data and Monte Carlo yields for the preselection (njets ≥ 2, $H_T > 200$ GeV, $\mathrm{E_T^{miss}}$ > 50 GeV, lepton $p_T >$ (10,5) GeV and not lepton $p_T >$ (20,10) GeV). $t\bar{t} \rightarrow \ell^+\ell^-$ corrresponds to dileptonic $t\bar{t}$ in which both W's decay via $W \rightarrow e\nu$ or $W \rightarrow \mu\nu$, $t\bar{t} \rightarrow \ell^\pm\tau^\mp/\tau^+\tau^-$corresponds to dileptonic $t\bar{t}$ in which both leptons come from W decay and at least 1 lepton is from $W \rightarrow \tau \rightarrow \ell$. **NO SPILLAGE CORRECTION APPLIED HERE. MC TOTAL FAKE PREDICTION IS 1.7 EVENTS, FR GIVES ~4X HIGHER...**

Sample	ee	$\mu\mu$	$e\mu$	tot
ttll	0.3 ± 0.1	2.2 ± 0.3	2.4 ± 0.3	4.9 ± 0.4
tttau	0.4 ± 0.1	2.3 ± 0.3	2.1 ± 0.3	4.7 ± 0.4
single fakes	0.0 ± 0.0	3.0 ± 1.2	3.0 ± 1.2	6.0 ± 1.6
double fakes	0.0 ± 0.0	0.6 ± 0.4	0.0 ± 0.0	0.6 ± 0.4
DY	0.0 ± 0.0	2.2 ± 0.8	0.4 ± 0.4	2.6 ± 0.9
WW	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
WZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
t	0.0 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.4 ± 0.1
tot SM MC	0.7 ± 0.2	10.4 ± 1.5	8.1 ± 1.3	19.3 ± 2.0
data	1	7	5	13
LM1	0.3 ± 0.1	2.9 ± 0.2	2.3 ± 0.1	5.5 ± 0.2
LM3	0.0 ± 0.0	0.4 ± 0.0	0.3 ± 0.0	0.7 ± 0.1

Table 3: Comparison of data yields in the high-p_T dilepton and dilepton-H_T trigger samples, passing the selection njets ≥ 2, $\mathrm{E_T^{miss}} > 50$ GeV, H_T>200 GeV, lepton p_T> (20,10) GeV.

Sample	ee	$\mu\mu$	$e\mu$	tot
high-p_T dilepton trigger	50	49	124	223
dilepton-H_T trigger	50	48	117	215

7 Properties of data passing the preselection

A number of kinematical distributions for events passing the preselection in data are compared with MC in Appendix **??**. 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.

8 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 $\mathrm{E_T^{miss}}$;
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 3.

- 2010 signal region: H_T> 300 GeV and $y > 8.5$ GeV$^{1/2}$.
- high y signal region: H_T> 300 GeV and $y > 13$ GeV$^{1/2}$.
- high $H_T$signal region: H_T> 600 GeV and $y > 8.5$ GeV$^{1/2}$.

We cut on the quantity $y \equiv \mathrm{E_T^{miss}}/\sqrt{\mathrm{H_T}}$ rather than $\mathrm{E_T^{miss}}$ 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 9.1). In the future, we plan to cut instead on $\mathrm{E_T^{miss}}$ and H_T, since we observe that $\mathrm{E_T^{miss}}$ 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 $\mathrm{E_T^{miss}}$ and H_T requirements (see App. A).

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 y and high H_T) have tightened requirements on y and H_T, respectively, which reduce the expected background by roughly an order of magnitude.

We present the data and MC expected yields in the 3 signal regions in Tables 4-6. Here we require lepton $p_T >$ (20,10) GeV and find the same yields (as well as the same individual events) in the high p_T dilepton trigger and lepton-H_T trigger samples.

Table 4: Data and MC yields in the 2010 signal region. The MC errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
ttll	0.9 ± 0.2	1.4 ± 0.2	2.6 ± 0.3	5.0 ± 0.4
tttau	0.4 ± 0.1	0.4 ± 0.1	1.2 ± 0.2	2.0 ± 0.3
ttfake	0.1 ± 0.1	0.0 ± 0.0	0.2 ± 0.1	0.3 ± 0.1
DY	0.2 ± 0.2	0.4 ± 0.4	0.4 ± 0.4	1.0 ± 0.6
WW	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.2 ± 0.0
WZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
t	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0
wjets	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	1.8 ± 0.3	2.3 ± 0.5	4.5 ± 0.6	8.6 ± 0.8
data	4	3	7	14
LM1	6.4 ± 0.2	7.6 ± 0.2	3.9 ± 0.2	18.0 ± 0.4
LM3	1.3 ± 0.1	1.5 ± 0.1	2.0 ± 0.1	4.8 ± 0.1

Table 5: Data and MC yields in the high y signal region. The MC errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
ttll	0.2 ± 0.1	0.3 ± 0.1	0.5 ± 0.1	0.9 ± 0.2
tttau	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.2 ± 0.1
ttfake	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
WW	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0
WZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
t	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
wjets	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	0.2 ± 0.1	0.4 ± 0.1	0.7 ± 0.2	1.3 ± 0.2
data	3	0	3	6
LM1	2.7 ± 0.2	3.2 ± 0.2	1.7 ± 0.1	7.6 ± 0.3
LM3	0.4 ± 0.0	0.5 ± 0.0	0.7 ± 0.1	1.7 ± 0.1

Table 6: Data and MC yields in the high H_T signal region. The errors are statistical only.

Sample	ee	$\mu\mu$	$e\mu$	tot
ttll	0.1 ± 0.1	0.2 ± 0.1	0.2 ± 0.1	0.5 ± 0.1
tttau	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.1
ttfake	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	0.4 ± 0.4	0.4 ± 0.4
WW	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
WZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
t	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
wjets	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
tot SM MC	0.2 ± 0.1	0.3 ± 0.1	0.7 ± 0.4	1.2 ± 0.5
data	1	0	2	3
LM1	2.3 ± 0.1	2.7 ± 0.2	1.2 ± 0.1	6.2 ± 0.2
LM3	0.6 ± 0.0	0.6 ± 0.0	0.7 ± 0.1	1.9 ± 0.1

These results are summarized as:

- 2010 signal region ($y > 8.5$ GeV$^{1/2}$, $H_T >$ 300 GeV)
  - observed yield : 14
  - MC prediction : 8.6 ± 0.8
- high y signal region ($y > 13$ GeV$^{1/2}$, $H_T >$ 300 GeV)
  - observed yield : 6
  - MC prediction : 1.3 ± 0.2
- high H_T signal region ($y > 8.5$ GeV$^{1/2}$, $H_T >$ 600 GeV)
  - observed yield : 3
  - MC prediction : 1.2 ± 0.5

For all 3 signal regions, we have verified that we observe the same event yields (and same individual events) in the two data trigger samples. In the 2010 signal region, we observe 14 events, representing a slight excess with respect to the MC expectation. This excess is enhanced after moving to the high y signal region, where we find 6 events. In the high H_T region we observe 3 events, representing a slight excess.

For all signal regions, we observe 0 events in the Z mass window, confirming the expectation from MC that the DY background is negligible. We also confirm the MC prediction that the contribution of fake leptons is small, using the data-driven fake rate method. We find 1 lepton + fakeable object in the 2010 signal region, giving a predicted contribution from fake leptons of 0.6 ± 0.6, consistent with MC expectations. For the high y and high H_T signal regions, we do not observe any lepton + fakeable object events are our data-driven prediction for the fake contributions in these regions is zero.

9 Data Driven Background Estimation Methods

We use 3 data-driven methods to estimate the background in the signal region. The first one exploits the fact that H_T and y are nearly uncorrelated for the $t\bar{t}$ background (Section 9.1); the second one is based on the fact that in $t\bar{t}$ the 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 is reconstructed as $\mathrm{E_T^{miss}}$ in the detector.

We study the closure of these methods using our madgraph $t\bar{t}$ sample, as well as the powheg sample TTTo2L2Nu2B_7TeV-powheg-pythia6_Spring11-PU_S1_START311_V1G1-v1 which has approximately 10 times more events in the dilepton channel than the madgraph sample. We use these samples to estimate correction factors and systematic uncertainties for the background predictions. However, the final choice of correction factors and uncertainties will be extracted from the Summer11 $t\bar{t}$ madgraph sample which will have 50 times as many events as the current madgraph sample. For the studies presented in this section, we do not apply trigger efficiency corrections or reweighting for number of reconstructed vertices since are not comparing MC to data.

9.1 ABCD method

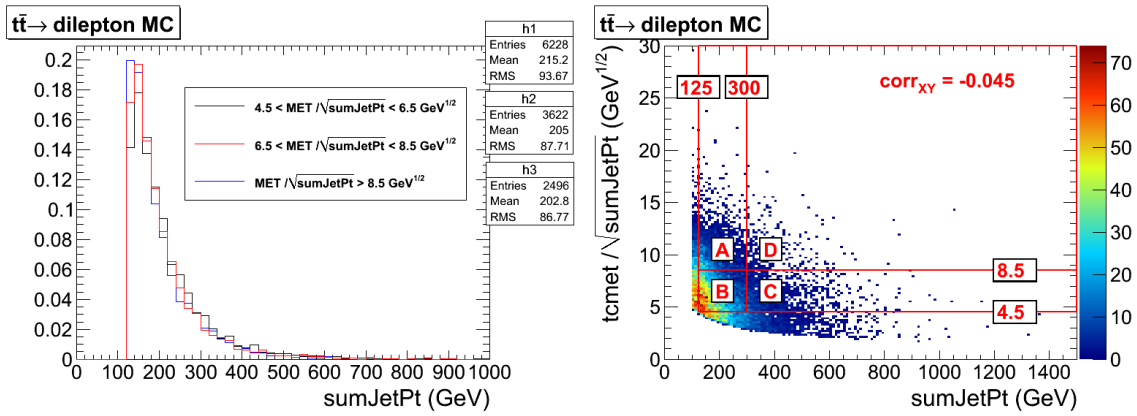

Figure 2: Left: distributions of SumJetPt in MC $t\bar{t}$ events for different intervals of MET/$\sqrt{\text{SumJetPt}}$. h1, h2, and h3 refer to the MET/$\sqrt{\text{SumJetPt}}$ intervals 4.5-6.5, 6.5-8.5 and >8.5, respectively. Right: Distributions of MET/$\sqrt{\text{SumJetPt}}$ vs. SumJetPt for SM Monte Carlo. Here we also show our choice of ABCD regions. The correlation coefficient corr_{XY} is computed for events falling in the ABCD regions. **THESE PLOTS ARE OLD MC, WILL UPDATE**

We find that in $t\bar{t}$ events H_T and y are nearly uncorrelated, as demonstrated in Fig. 2 (left). Thus, we can use an ABCD method in the y vs. $H_T$plane to estimate the background in a data driven way. We dfine 4 regions in the plane of y vs. H_T, as shown in Fig. 2. The region D is the signal region, and the regions A, B and C are control regions. The predicted background in region D is given by $N_A \times N_C/N_B$.

In Table 8, we quote the MC expected yields for 1 fb^{-1}. In general we find that the prediction agrees with the observed yield in the signal region within ∼30-50% for all signal regions. 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. Based on these results, we apply the scale factors and uncertainties summarized in Table 9 to the predicted background from the ABCD method.

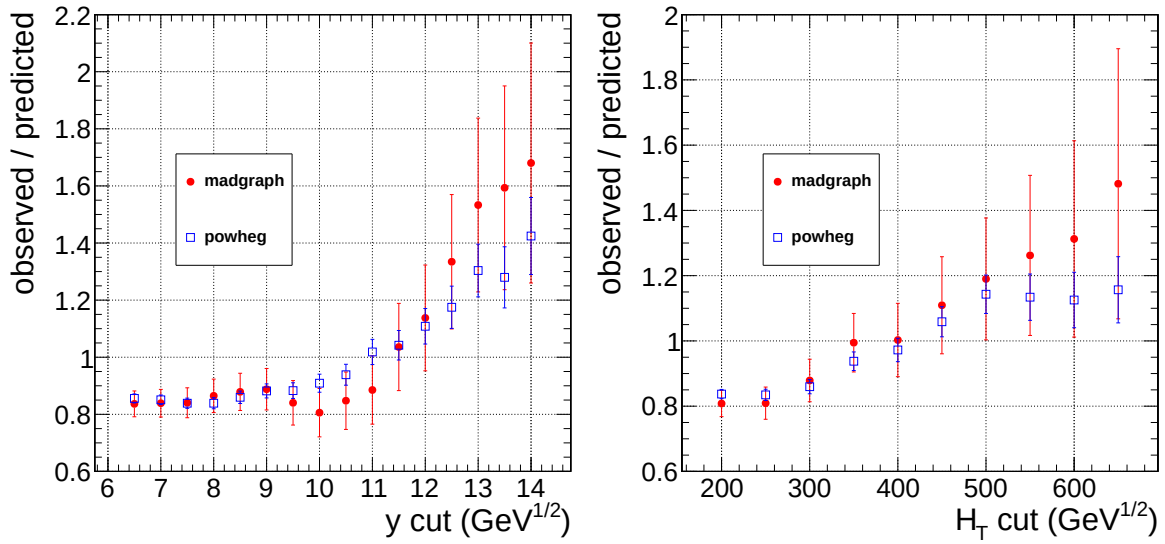

Figure 3: Variation of observed/predicted for the ABCD method as a function of the y and H_T cuts defining the signal region.

Table 7: Expected yields from $t\bar{t}$ MC in 1 fb^{-1} in the four ABCD regions (depicted in Figs. 5-7, as well as the predicted yield in region D given by A $\times$ C / B and the ratio of the observed signal yield to the prediction. The quoted uncertainties are statistical only, assuming Gaussian errors.

signal region	sample	A	B	C	D	A $\times$ B / C	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
high y signal region	madgraph	18.4 ± 1.6	951.5 ± 11.9	165.2 ± 4.9	4.9 ± 0.9	3.2 ± 0.3	1.53 ± 0.30
	powheg	17.3 ± 0.5	850.6 ± 3.7	157.8 ± 1.6	4.2 ± 0.3	3.2 ± 0.1	1.30 ± 0.09
high H_T signal region	madgraph	251.3 ± 6.1	951.5 ± 11.9	11.1 ± 1.3	3.8 ± 0.8	2.9 ± 0.3	1.31 ± 0.30
	powheg	231.7 ± 2.0	850.6 ± 3.7	12.5 ± 0.5	3.8 ± 0.3	3.4 ± 0.1	1.13 ± 0.08

9.2 Dilepton P_T method

This method is based on a suggestion by V. Pavlunin[8], and was investigated by our group in 2009[9]. 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 $\mathrm{E_T^{miss}}$.

Then, in order to predict the $t\bar{t} \rightarrow$ dilepton contribution to a selection with $\mathrm{E_T^{miss}}$+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 $\mathrm{E_T^{miss}}$cut in order to reduce Drell Yan backgrounds. This is discussed in Section 5.3 of Reference [9]; for a $\mathrm{E_T^{miss}}$cut of 50 GeV, the rescaling factor is obtained from the MC as

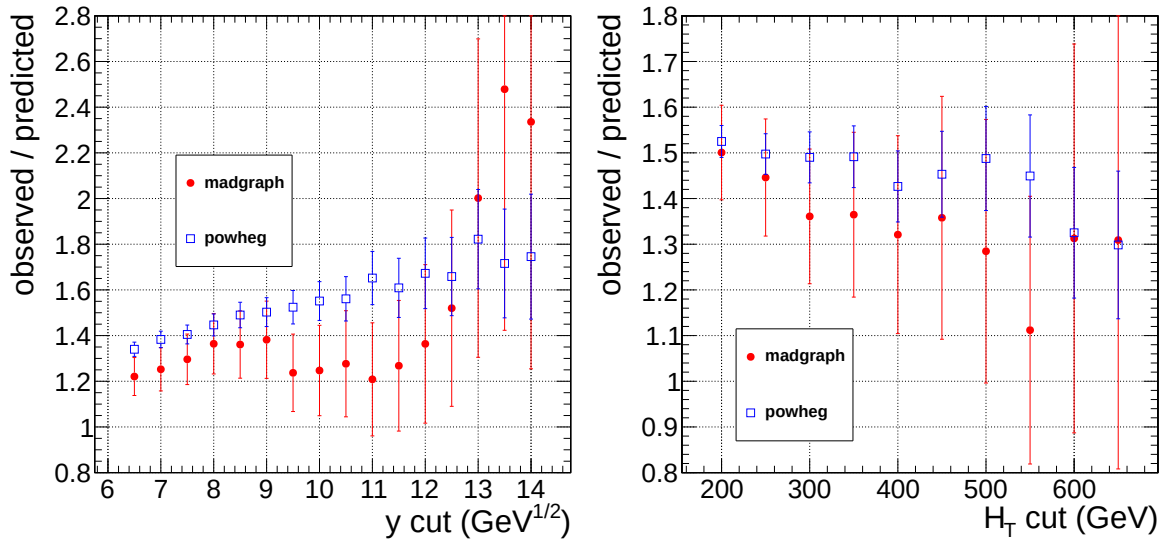

Figure 4: Variation of observed/predicted for the $p_T(\ell\ell)$ method as a function of the y and H_T cuts defining the signal region.

Table 8: Expected observed and predicted yields in 1 fb^{-1} for the $p_T(\ell\ell)$ method, and the ratio of the observed signal yield to the prediction. The quoted uncertainties are statistical only, assuming Gaussian errors.

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 y signal region	madgraph	2.4 ± 0.7	4.9 ± 0.9	2.00 ± 0.70
	powheg	2.3 ± 0.2	4.2 ± 0.3	1.82 ± 0.22
high H_T signal region	madgraph	2.9 ± 0.8	3.8 ± 0.8	1.31 ± 0.43
	powheg	2.9 ± 0.2	3.8 ± 0.3	1.33 ± 0.14

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

CURRENTLY TAKING K FROM MC, SHOULD WE TAKE IT FROM DATA?

Table 9: 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 y signal region	1.3 ± 0.3	1.7 ± 0.3
high H_T signal region	1.2 ± 0.2	1.3 ± 0.2

9.3 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 same-flavor over opposite-flavor events. We use the observed yield in the OF final state to predict the yields in the SF final state according to:

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

where $R_{\mu e} = X \pm Y$ is the ratio of muon to electron selection efficiencies. This quantity has been evaluated by taking the ratio of the number of observed $Z \to \mu^+\mu^-$ to $Z \to e^+e^-$ events in data, in the mass range 76-106 GeV with no jets or $\mathrm{E_T^{miss}}$ requirements.

10 Results

10.1 Background estimate from the ABCD method

The data yields in the four regions are summarized in Tables 10-Table 12 for the 3 signal regions. The ABCD background prediction $N_A \times N_C/N_B$ is scaled by the correction factors determined in Sec. 9, as summarized in Table 9.

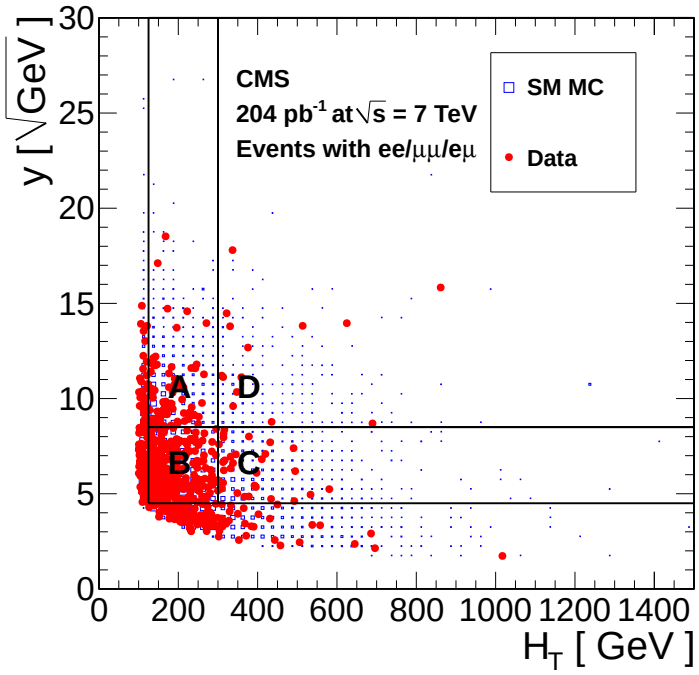

Figure 5: Distributions of y vs. H_T for SM Monte Carlo and data. The 2010 signal region boundaries are overlayed.

Table 10: Data yields in the four regions of Figure 5 for the 2010 signal region, as well as the predicted yield in region D given by A $\times$ C / 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	A	B	C	D	A $\times$ B / C
$t\bar{t}$	48.8 ± 1.4	184.1 ± 2.7	31.9 ± 1.1	7.3 ± 0.5	8.5 ± 0.4
DY	0.5 ± 0.4	8.2 ± 1.5	0.7 ± 0.5	1.0 ± 0.6	0.0 ± 0.0
W^+W^-	0.6 ± 0.1	1.6 ± 0.1	0.1 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
$W^\pm Z^0$	0.1 ± 0.0	0.3 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Z^0Z^0	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	1.9 ± 0.1	5.6 ± 0.2	0.2 ± 0.0	0.1 ± 0.0	0.1 ± 0.0
W + jets	0.6 ± 0.6	1.2 ± 0.6	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total SM MC	52.6 ± 1.6	201.2 ± 3.2	33.1 ± 1.2	8.6 ± 0.8	8.6 ± 0.4
data	72	238	29	14	8.8 ± 2.0

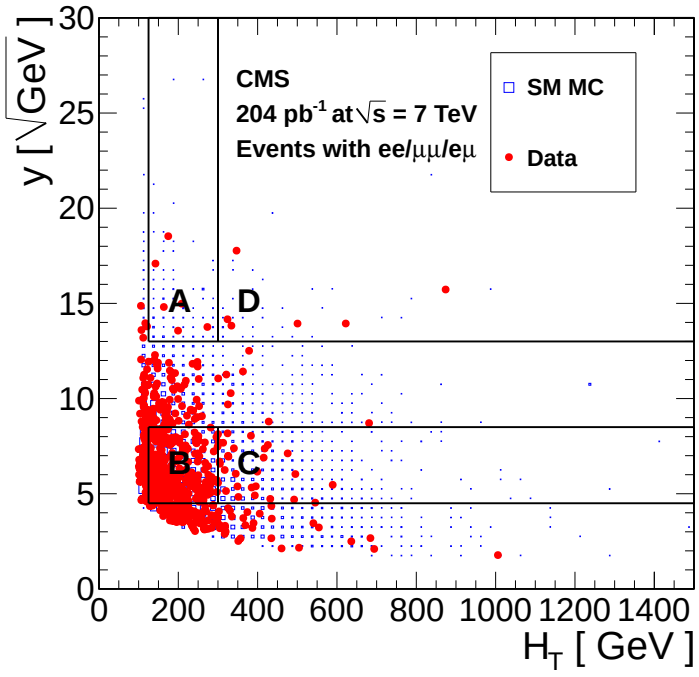

Figure 6: Distributions of y vs. H_T for SM Monte Carlo and data. The high y signal region boundaries are overlayed.

Table 11: Data yields in the four regions of Figure 6 for the high y signal region, as well as the predicted yield in region D given by A $\times$ C / 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 uncertainties.

sample	A	B	C	D	A $\times$ B / C
$t\bar{t}$	3.6 ± 0.4	184.1 ± 2.7	31.9 ± 1.1	1.2 ± 0.2	0.6 ± 0.1
DY	0.3 ± 0.3	8.2 ± 1.5	0.7 ± 0.5	0.0 ± 0.0	0.0 ± 0.0
W^+W^-	0.1 ± 0.0	1.6 ± 0.1	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
$W^\pm Z^0$	0.0 ± 0.0	0.3 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Z^0Z^0	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
single top	0.2 ± 0.0	5.6 ± 0.2	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
W + jets	0.0 ± 0.0	1.2 ± 0.6	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total SM MC	4.1 ± 0.5	201.2 ± 3.2	33.1 ± 1.2	1.3 ± 0.2	0.7 ± 0.1
data	6	238	29	6	0.7 ± 0.3

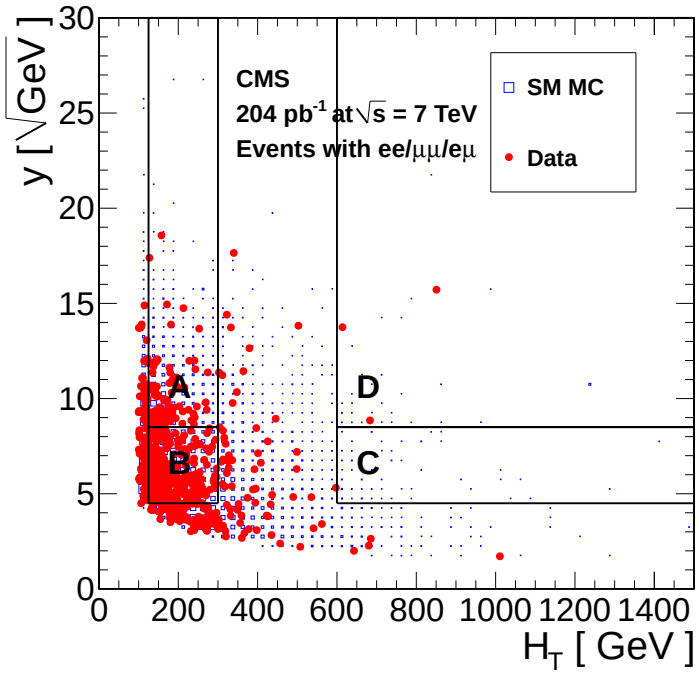

Figure 7: Distributions of y vs. H_T for SM Monte Carlo and data. The high H_T signal region boundaries are overlayed.

Table 12: Data yields in the four regions of Figure 7 for the high H_T signal region, as well as the predicted yield in region D given by A $\times$ C / B. The quoted uncertainty on the prediction in data is statistical only, assuming Gaussian errors. Since the yield in region C is 0, we assess as the uncertainty the prediction corresponding to 1 observed event in 1. We also show the SM Monte Carlo expectations with statistical uncertainties.

sample	A	B	C	D	A $\times$ B / C
ttall	48.8 ± 1.4	184.1 ± 2.7	2.2 ± 0.3	0.7 ± 0.2	0.6 ± 0.1
DY	0.5 ± 0.4	8.2 ± 1.5	0.0 ± 0.0	0.4 ± 0.4	0.0 ± 0.0
WW	0.6 ± 0.1	1.6 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
WZ	0.1 ± 0.0	0.3 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
ZZ	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
t	1.9 ± 0.1	5.6 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
wjets	0.6 ± 0.6	1.2 ± 0.6	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total SM MC	52.6 ± 1.6	201.2 ± 3.2	2.2 ± 0.3	1.2 ± 0.5	0.6 ± 0.1
data	72	238	0	3	0.0 ± 0.3

Table 13: Summary of results of the ABCD method, applied to the 3 signal regions.

Signal Region	$N_A \times N_C/N_B$	correction factor	prediction
2010 signal region	8.8 ± 2.0	1.0 ± 0.2	8.8 ± 2.0 (stat) $\pm$ 1.8 (syst)
high y signal region	0.7 ± 0.3	1.3 ± 0.2	0.9 ± 0.4 (stat) $\pm$ 0.2 (syst)
high H_T signal region	0.0 ± 0.3	1.2 ± 0.2	0.0 ± 0.4 (stat) $\pm$ 0.1 (syst)

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

Table 14: Summary of results of the dilepton p_T template method. The number of events in the given signal region with the y requirement replaced by a $p_T(\ell\ell/\sqrt{H_T}$requirement is $N(D')$, the estimated DY contribution from the $R_{out/in}$ technique is $N(DY)$, the $\mathrm{E_T^{miss}}$ acceptance scaling factor is K, the residual correction factor is K_C, and the the prediction from the $p_T(\ell\ell)$ method is N_P. The quoted statistical uncertainty is due to that of $N(D')$, the quoted systematic uncertainty includes that of $N(DY)$ and K_C. **K and KC taken from MC, do we want to take K and/or KC from data?**

Signal Region	$N(D')$	$N(DY)$	K	K_C	N_P
2010 signal region	3	0.4 ± 0.3	1.5	1.4 ± 0.2	X $\pm$ X (stat) $\pm$ X (syst)
high y signal region	2	0.1 ± 0.1	1.5	1.7 ± 0.3	4.8 ± 3.4 (stat) $\pm$ X (syst)
high H_T signal region	0	0.0 ± 0.1	1.3	1.3 ± 0.2	0.0 ± 1.7 (stat) $\pm$ 0.3 (syst)

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. We subtract off the expected DY contribution using the data-driven $R_{out/in}$ technique, using $R_{out/in} = 0.13 \pm 0.07$. **add plot justifying this value**. We then scale this yield by 2 corrections factors: K, the $\mathrm{E_T^{miss}}$ acceptance correction factor, and K_C, the correction factor determined in Sec. 9. Our final prediction N_P is given by the following, as summarized in Table 15:

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

I will add here the results of victory method applied to the H_T sideband region as a validation of the method

10.3 Summary of results

Table 15: Summary of the observed and predicted yields in the 3 signal regions. **currently just taking ee+mm-em for OF subtraction, need to update with efficiency/trig correction**.

Signal Region	2010 signal region	high y signal region	high H_T signal region
Observed yield	14	6	3
MC prediction ABCD prediction	$p_T(\ell\ell)$ prediction	OF subtraction ($ee + \mu\mu - e\mu$)	

In summary, we observe a slight excess of events with respect to MC and data-driven predictions in our loose 2010 signal region. The excess with respect to MC and ABCD predictions is enhanced after tightening the y requirement, but the observed yield shows no significant excess with respect to the $p_T(\ell\ell)$ method prediction. There is also a slight excess of events with respect to MC and data-driven predictions for the high H_T signal region.

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 *at al.*, AN-CMS2009/159.

[7] https://twiki.cern.ch/twiki/bin/view/CMS/ProductionReProcessingSpring10

[8] V. Pavlunin, Phys. Rev. **D81**, 035005 (2010).

[9] D. Barge *at al.*, AN-CMS2009/130.

[10] W. Andrews *et al.*, AN-CMS2009/023.

[11] D. Barge *at al.*, AN-CMS2010/257.

[12] W. Andrews *et al.*, AN-CMS2010/274.

[13] J. Conway, `http://www-cdf.fnal.gov/physics/statistics/code/bayes.f`.

[14] G. Landsberg, `https://twiki.cern.ch/twiki/pub/CMS/EXOTICA/cl95cms.c`

[15] `https://hypernews.cern.ch/HyperNews/CMS/get/susy/617/2/1.html`

[16] `https://twiki.cern.ch/twiki/bin/view/CMS/SUSY38XSUSYScan`

[17] arXiv:hep-ph/0605240v2

[18] `CleanExclusion.cc` available at `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SUSYLimitTools`

[19] R. Cousins, `http://www.physics.ucla.edu/c̃ousins/stats/cousins_lognormal_prior.pdf`

[20] S. Harper, private communication (relayed to us by M. Chiorboli.).

[21] A. Barr *at 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`

[22] `http://indico.cern.ch/contributionDisplay.py?contribId=5&confId=93837`

[23] M. Narain *et al.*, CMS AN-2010/259; we thank the Brown group for providing their code to us.

Appendix A The ABCD' Method