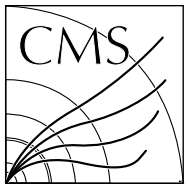
Available on CMS information server

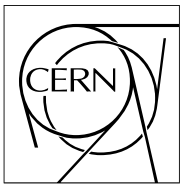
CMS AN -2009/042



The Compact Muon Solenoid Experiment

Analysis Note

The content of this note is intended for CMS internal use and distribution only



21 March 2009 (v5, 21 June 2009)

Prospects for measuring the WW production cross section in pp collisions at $\sqrt{s}$ =10 TeV

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

University of California, Santa Barbara

W. Andrews, D. Evans, F. Golf, J. Mülmenstädt, S. Padhi, Y. Tu, F. Würthwein, A. Yagil

University of California, San Diego

L. Bauerdick, I. Bloch, K. Burkett, I. Fisk, O. Gutsche

Fermi National Accelerator Laboratory, Batavia, Illinois

Abstract

This note describes an analysis strategy for measuring the WW production cross section in 100 pb^{-1} of pp collision data at $\sqrt{s} = 10$ TeV. Based on Monte Carlo simulations of WW and the major background processes, we explore suitable background reduction cuts and develop data-driven methods to estimate those backgrounds that we expect not to be reliably modeled by Monte Carlo. We present the expected event yields and statistical and systematic uncertainties that we expect to achieve with this analysis strategy. We estimate to have more than 50% chance to get 5σ signal significance, with the expected uncertainty on the cross section estimate of 30% or better.

Contents

1 Introduction **4**

2 Data samples **4**

3 Event selection **4**

3.1 Trigger . . . 6

3.2 Muon selection . . . 6

3.2.1 Primary muon selection . . . 6

3.2.2 Secondary muon selection . . . 7

3.3 Electron selection . . . 7

3.4 Isolation . . . 8

3.4.1 Muon isolation . . . 8

3.4.2 Electron isolation . . . 8

3.5 Missing E_T . . . 9

3.6 Jet veto . . . 9

3.7 Z veto . . . 12

3.8 Muon vetoes . . . 12

4 Expectations from Monte Carlo **12**

4.1 Selection efficiency . . . 12

4.2 Expected yields . . . 12

5 Background estimation **14**

5.1 W+jets background . . . 14

5.2 Top background . . . 15

5.3 Drell-Yan background . . . 16

6 Systematics **18**

6.1 Reconstruction efficiency uncertainties . . . 18

6.2 Background estimation uncertainties . . . 19

6.3 Luminosity uncertainties . . . 20

7 Cross section determination **20**

8 Conclusion **20**

Appendices **21**

A Anomalous triple gauge couplings **21**

B Electron selection details **21**

C Using the transverse mass of the W for $Z/\gamma^* \to \tau\tau$ suppression **24**

D Jet Veto at high η **24**

E Jet Veto using b tagging **24**

F Application of the Drell-Yan estimate method **26**

F.1 Method 26

F.2 Results 27

G Event yields in other samples **31**

H Estimating the sensitivity to observe WW in 100/pb **31**

H.1 In the Gaussian limit after a measurement is made 32

H.2 In the Poisson limit after a measurement is made 32

H.3 Predicting significance for a future measurement to be made 33

I QCD multi jet background estimation **33**

List of Tables

1 Datasets 5

2 Extra Tags 5

3 WW and top yields for different jet types and veto thresholds 12

4 Selection cuts and their effect on signal and background yields 13

5 Selection cuts and their effect on signal and background yields ($e\mu$ final state) 13

6 Expected event yields 13

7 WW fake estimates 14

8 Results of the Drell-Yan estimate 18

9 Efficiency Systematics 19

10 Expected event yields (data-driven) 20

11 Category-0 electron cuts 22

12 Category-1 electron cuts 22

13 Category-2 electron cuts 23

14 Forward-jet veto 25

15 Performance of b tagging to veto top background 26

16 Results of the non-peaking background estimate supporting the Drell-Yan estimate 30

17 Detailed results of the Drell-Yan estimate 31

18 Alternate samples of W events 31

19 Alternate samples of Drell-Yan events 32

20 Alternate Pythia samples of $V\gamma$ 32

List of Figures

1 Feynman diagrams for WW production 4
2 Soft muon reconstruction efficiency 7
3 $\not\!\!E_T$ performance for $ee/\mu\mu$ final state 10
4 Projected $\not\!\!E_T$ performance for $e\mu$ final state 10
5 Top jets E_T and η distributions 11
6 E_T distributions of the most energetic jets for WW and top samples 11
7 Absolute pseudo-rapidity distribution of the most central b quark from $t\bar{t}$ after jet veto . . 15
8 b-jet tagging efficiency 16
9 Light quark jet mis-tagging efficiency 17
10 Top tagging efficiency with soft muons on $t\bar{t}$ 17
11 Generator level di-lepton p_T and M_{ll} distributions for SM and aTGC with $\lambda_Z = \lambda_\gamma = 0.30$ 21
12 Efficiency of $Z/\gamma^* \to \tau\tau$ rejection using projected $\not\!\!E_T$ and transverse W mass 23
13 Jet E_T and η for all jets in $t\bar{t}$ Monte Carlo events after all other cuts have been applied . 24
14 Jet E_T and η for b jets in $t\bar{t}$ Monte Carlo events after all other cuts have been applied . . 25
15 Distribution of b-tagging discriminant in top and WW events 25
16 The reconstructed di-muon invariant mass for peaking processes 28
17 The reconstructed di-muon invariant mass for non-peaking processes 28
18 The reconstructed di-electron invariant mass for peaking processes 28
19 The reconstructed di-electron invariant mass for non-peaking processes 28
20 The reconstructed electron-muon invariant mass for peaking processes 29
21 The reconstructed electron-muon invariant mass for non-peaking processes 29
22 The ratio $R_{\mathrm{out/in}}$ as a function of the applied flat $\not\!\!E_T$ cut for Drell-Yan events. 30
23 The ratio $R_{\mathrm{out/in}}$ as a function of the applied flat $\not\!\!E_T$ cut for ZZ events. 30
24 QCD multi jet background estimation 34

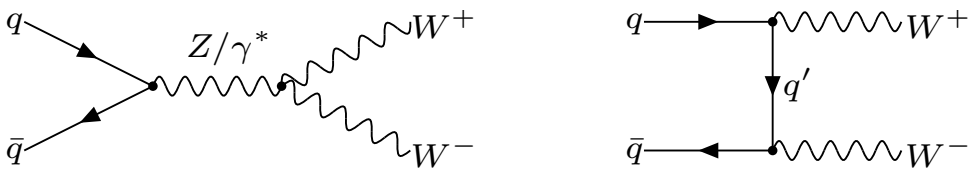


Figure 1: Lowest order Feynman diagram of the dominant production mechanisms for WW production.

1 Introduction

WW in the Standard Model proceeds via s- and t-channel diagrams as depicted in Figure 1 at lowest order. The s-channel production is a sensitive measure of the WWZ and $WW\gamma$ triple gauge couplings. This process is thus sensitive to physics beyond the Standard Model via anomalous vector-boson couplings, whose effect on WW production is expected to grow with larger invariant mass of the di-boson system available at LHC compared with LEP and Tevatron experiments. In addition, it represents one of the dominant irreducible backgrounds for the Standard Model Higgs boson searches. Therefore it is essential to establish properties and cross section of the WW process at the LHC energies with the CMS detector response.

This note describes an analysis strategy for measuring the WW production cross section in 100 pb^{-1} of pp collision data at $\sqrt{s}$ =10 TeV. The strategy is to find two isolated leptons (electron or muon) of opposite charge with large missing transverse energy and low hadronic activity established by a jet veto. The analysis uses the top dilepton analysis [1] as a baseline with additional selection requirements to suppress top and Drell-Yan production, and isolate WW events.

The note is organized as follows. First we discuss the data samples in use. Then we provide a detailed overview of all selection requirements followed by a summary of expected event yields for the dominant physics processes. We then establish data driven methods of background estimation. After that we discuss the expected systematic effects and associated uncertainties for the analysis with 100 pb^{-1} integrated luminosity. This is followed with an Appendix, where we discuss a few new methods to improve background suppression and background estimation in the future. The potential for measuring triple gauge couplings in this analysis is also briefly mentioned in the Appendix.

2 Data samples

This analysis is performed with a set of Summer/Fall08 Monte Carlo data samples. Table 1 summarizes the datasets used in the analysis. The analysis was done using the `CMSSW_2_2_3` software release with a number of additional changes listed in Table 2.

Most of the background samples are generated with Madgraph, whereas a majority of the individual physics processes are generated with Pythia. An exception is Drell-Yan, which is a Pythia sample since it has a better coverage of the low mass Drell-Yan going as low as 20 GeV. The Madgraph low mass sample has an additional requirement on minimal hadronic activity in the event and most of these events are vetoed by the jet veto. Appendix G shows estimated contributions for alternative samples.

3 Event selection

The WW signal cross section is several orders of magnitude lower than major background processes, such as $t\bar{t}$, W+jets and Drell-Yan. In this section we describe an event selection that allows us to suppress major backgrounds and control the remaining ones.

Our event selection consists of the following general requirements:

- We select events with two opposite sign isolated leptons (e or μ) of $p_T > 20$ GeV. We expect to collect these events with the single lepton triggers.
- We require $\not\!\!E_T$. Because of the Drell-Yan background, the $\not\!\!E_T$ requirement is stricter for $ee/\mu\mu$ than for $e\mu$.
- Events consistent with $Z \to ee$ and $Z \to \mu\mu$ are rejected.

Table 1: Datasets used for the analysis. The number of events corresponds to the number of successfully processed events. The cross section numbers are taken from a reference set of next to leading order cross section estimations maintained by the top physics analysis group [2].

Dataset name	Cross section, pb	Number of events	Effective luminosity, pb^{-1}
/TTJets-madgraph/Fall08_IDEAL_V9_v1/GEN-SIM-RECO	414	1 023 322	2 500
/WJets-madgraph/Fall08_IDEAL_V9_v1/GEN-SIM-RECO	45 000	8 799 192	190
/Zee_M20/Summer08_IDEAL_V9_reco-v3/GEN-SIM-RECO	2 220	1 008 888	460
/Zmumu_M20/Summer08_IDEAL_V9_reco-v2/GEN-SIM-RECO	2 220	1 275 840	580
/Ztautau_M20/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	2 220	994 800	450
/WW/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	74	203 591	2 800
/ZZ/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	10.5	200 564	20 000
/WZ_incl/Summer08_IDEAL_V9_v2/GEN-SIM-RECO	32	214 100	6 700
/SingleTop_tWChannel/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	29	139 048	4 750

Table 2: List of additional packages used on top of CMSSW_2_2_3 release.

Package	Tag
`TopQuarkAnalysis/TopObjectProducers`	V04-06-01
`PhysicsTools/PatAlgos`	V04-14-19
`PhysicsTools/PatUtils`	V03-05-02
`DataFormats/PatCandidates`	V03-18-04
`CondFormats/JetMETObjects`	V01-06-06
`JetMETCorrections/Configuration`	V01-08-11
`JetMETCorrections/Modules`	V02-09-00
`JetMETCorrections/Algorithms`	V01-07-11
`JetMETCorrections/JetPlusTrack`	V03-02-06
`RecoMET/METProducers`	V02-08-02-14
`RecoMET/METAlgorithms`	V02-05-00-16
`DataFormats/METReco`	V00-06-02-09
`RecoMET/Configuration`	V00-04-02-15

- We reject events with jet activity; this is because the main background for WW is $pp \to t\bar{t} \to WWb\bar{b}$.
- The jet veto is complemented by a veto on events with an additional muon. This allows to further reject $t\bar{t}$ events with $b \to \mu$; it also gives us a way to estimate the remaining $t\bar{t}$ background in a data-driven way.

These requirements are described in more detail in the remainder of this section. After final event selection with the samples used for this study we find no events with multiple dilepton candidates. In the future, if we need to select the best candidate among a few candidates, we will select a pair with the most energetic leptons.

3.1 Trigger

This is a dilepton analysis, so we could use the inclusive lepton triggers or the dilepton triggers. The analysis threshold on the leptons is 20 GeV. This is well above the anticipated thresholds for the single lepton triggers for luminosities up to 10^{31} cm^{-2} sec^{-1} [3].

Thus, our plan is to perform this analysis using the inclusive lepton triggers. As long as the analysis thresholds are well above the single lepton trigger thresholds, we gain almost nothing by adding the dilepton trigger. Clearly, if reality forces the lepton thresholds to be above 20 GeV, we will resort to lower threshold dilepton triggers.

In this Monte Carlo analysis, we use inclusive lepton triggers with no isolation, *i.e.*, the logical OR of `HLT_Ele15_SW_L1R` and `HLT_Mu9`. These triggers are present in the Summer08/Fall08 samples that we are using, as well as in the expected data taking trigger table[3].

To measure the efficiency of these two triggers in Monte Carlo, we select $WW \to e\mu$ events where both leptons are truth matched to $W \to e$ and $W \to \mu$ and where both leptons pass the requirements of Sections 3.2 and 3.3. We measure efficiencies of $96.2 \pm 0.6\%$ and $92.5 \pm 0.8\%$ for electrons and muons respectively. (Muons were required to have $|\eta| < 2.0$ to be within the single muon trigger acceptance).

In data we will measure the single lepton trigger efficiencies from tag-and-probe on $Z \to \mu\mu$ and $Z \to ee$ decays [14].

3.2 Muon selection

The analysis uses muons for two purposes. *Primary muons* have cuts chosen to select isolated, high-p_T muons coming from decays of W and Z bosons. *Secondary muons* are selected to be non-isolated, soft muons; these are used to tag b jets to estimate the top background yield.

3.2.1 Primary muon selection

The primary muon selection is the following:

1. Global Muon with $p_T > 20$ GeV/c.
2. Global fit $\chi^2/N_{\rm dof} \leq 10$.
3. Absolute value of the xy plane impact parameter from the silicon fit corrected for the beam spot: $|d_0| < 2$ mm.
4. Number of valid hits for the silicon fit ≥ 11.
5. Isolation as described in Section 3.4.

Requirements 2, 3, and 4, are taken from the muon identification note from the muon POG [4]. The efficiency of the identification cuts is given in the muon ID note and is around 95%.

The efficiency of the isolation requirement on muons passing the other identification requirements listed above and truth matched to $W \to \mu$ in WW events is measured to be 93.1%. In data we will measure the muon ID and isolation efficiency from tag-and-probe on $Z \to \mu\mu$ decays [14].

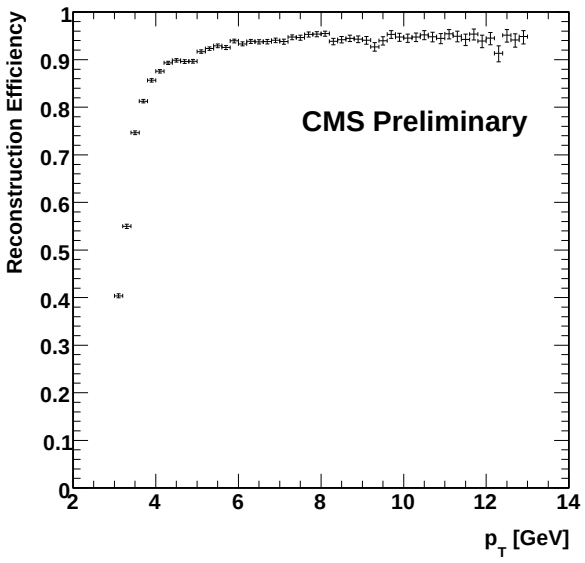


Figure 2: The reconstruction effiency of soft muons including the detector acceptance.

3.2.2 Secondary muon selection

The selection requirements for secondary muons differ from the selection requirements for the prompt muons used for the di-lepton event identification as follows:

1. $p_T > 3$ GeV
2. `TMLastStationOptimizedLowPtTight` selector
3. No isolation requirement

The `TMLastStationOptimizedLowPtTight` selector was chosen based on a balance of tagging efficiency and fake rate. This algorithm exploits the difference in the penetration depths between hadrons and muons, preserving high reconstruction efficiency for low momentum muons. Figure 2 shows the overall reconstruction efficiency for the soft muons including the detector acceptance.

3.3 Electron selection

Electrons were reconstructed using the pixel match GSF electron reconstruction algorithm. The following requirements were made to select well measured high-p_T primary electrons:

- $p_T > 20$ GeV/c
- $\Delta R > 0.1$ from any muon
- $|d_0| < 0.25$ mm in the xy plane corrected for the beamspot position
- Tight electron ID (Appendix B)
- Isolation as described in Section 3.4.

To reject fake electrons, a standard category-dependent electron ID is applied [8]. The expected performance of this electron ID is described in [9]. The POG-recommended set of thresholds was optimized using release `CMSSW_1_6_X` based samples. However, since the definition of two of the variables used was changed in the `CMSSW_2_2_X` release, the thresholds applied to these variables were updated in consultation with the E/gamma group. Details can be found in Appendix B.

In order to estimate the efficiency of the tight electron identification requirement and the isolation requirement, we select electrons that pass all other requirements listed above except the one that we estimate.

We also required the electron to be truth matched to $W \to e$. Looking at WW events we estimate the electron identification efficiency to be 96.7% and the electron isolation efficiency 80.0%. In data we will measure the electron identification and isolation efficiency from tag-and-probe on $Z \to ee$ decays [14].

3.4 Isolation

To reject leptons from hadronic jets, an isolation requirement is used. The details of the isolation differ for electrons and muons. In both cases we used relative isolation defined as

$$\text{Isolation} = \frac{p_T^{\text{lepton}}}{p_T^{\text{lepton}} + S} \tag{1}$$

where $S = \sum p_T^{\text{tracks}} + \sum E_T^{\text{HCAL}} + \sum E_T^{\text{ECAL}}$ is the sum of isolation contributions from tracker, ECAL and HCAL. All selected electrons and muons were required to be isolated such that Isolation > 0.92.

3.4.1 Muon isolation

For muons we use isolation information from the muon object for a $\Delta R < 0.3$ isolation cone. For details see [15]. Here is a brief summary of muon isolation requirements.

- Track isolation requirements:
  - input track collection: `generalTracks`
  - $|d_0| < 0.1$ cm (beamspot corrected)
  - $|z_0^{\mu} - z_0^{\text{track}}| < 0.2$ cm (beamspot corrected)
  - veto cone around muon track: $\Delta R_{\text{veto}} > 0.01$
- Calorimeter requirements:
  - input: `CaloTowers`
  - $E_T^{\text{ECAL}} > 200$ MeV
  - $E_T^{\text{HCAL}} > 500$ MeV
  - $E^{\text{ECAL-barrel}} > 120$ MeV (3 times noise level)
  - $E^{\text{ECAL-endcap}} > 450$ MeV (3 times noise level)
  - $E^{\text{HCAL}} > 600$ MeV (3 times noise level)
  - ECAL veto cone around muon track: $\Delta R_{\text{veto}} > 0.07$
  - HCAL veto cone around muon track: $\Delta R_{\text{veto}} > 0.10$

3.4.2 Electron isolation

We used the standard isolation requirements recommended by the E/gamma POG [16, 17] and accessed through the relevant PAT variables. The requirements for the tracker isolation sum are described below:

- Input track collection: `generalTracks`
- $p_T > 1$ GeV to select well reconstructed tracks.
- $|z_0 - z_0^{\text{track}}| < 0.2$ cm to select tracks from the primary vertex.
- Cone dimensions: $0.015 < \Delta R < 0.3$

The calorimeter isolations were computed within a cone of size $\Delta R < 0.4$. The standard ECAL isolation was computed by summing the E_T of ECAL of RecHits within the cone about each electron. In the HCAL the E_T of towers within a cone around each lepton was summed. The expected electron energy deposition was removed from the ECAL sum using the standard Jurassic method described below.

- Jurassic strip half width, barrel: 0.02; endcap: 0.02
- Jurassic veto cone radius, barrel: 0.045; endcap: 0.070
- RecHit noise cut, barrel: 0.08 GeV; endcap: 0.30 GeV

3.5 Missing E_T

Missing transverse energy ($\not\!\!E_T$) is a key discriminating variable between the Drell-Yan background and the $WW \to l^+\nu l^-\bar{\nu}$ signal, where the neutrinos carry off undetected energy. Given that Drell-Yan mostly contributes to the ee and $\mu\mu$ final states, these states are treated differently than $e\mu$, which has a much smaller Drell-Yan background.

The $\not\!\!E_T$ selection uses track-corrected $\not\!\!E_T$ (tcMET [18]). The tcMET algorithm applies calorimeter response corrections for muons and charged tracks on the raw calorimeter-based $\not\!\!E_T$. This approach allows one to take into account the non-linearity of the charged hadron response from the calorimeter and substitute the expected calorimeter deposition with the measured track momentum.

One of the main sources of fake $\not\!\!E_T$ in Drell-Yan events is poor measurement of the recoiling hadronic activity in the event. Even after the tcMET correction, the remaining fake $\not\!\!E_T$ can be significant and in a number of cases the $\not\!\!E_T$ direction is back to back with respect to the lepton pair momentum. (For details see [1]). We exploit this fact to lower the Drell-Yan background contribution for ee and $\mu\mu$ final states. Selection requirements for these final states are listed below.

- $\not\!\!E_T > 45$ GeV
- $\not\!\!E_T > 0.6 \cdot P_T^{\ell\ell}$ OR $\alpha > 0.25$
- *projected* $\not\!\!E_T > 20$ GeV

where $P_T^{\ell\ell}$ is the transverse momentum of the lepton pair and α is the acoplanarity angle defined as $\alpha = \phi_1 - \phi_2 - \pi$ with ϕ_1 and ϕ_2 representing the azimuthal angles of the first and second leptons respectively. Figure 3 shows the missing transverse energy performance for the $ee/\mu\mu$ final state.

The dominant Drell-Yan background contribution to the $e\mu$ final state comes from $Z/\gamma^* \to \tau\tau$ events with leptonic τ decays, since neutrinos from τ decays represent a natural source of $\not\!\!E_T$. To exploit the small opening angle between the lepton and the neutrino from the τ decay, we define a *projected* $\not\!\!E_T$ as the component of $\not\!\!E_T$ perpendicular to the momentum of the closest lepton if the $\Delta\phi$ between lepton and $\not\!\!E_T$ is smaller than 90°, otherwise the projected $\not\!\!E_T$ is equal to the $\not\!\!E_T$ used for all other selections. We require the projected $\not\!\!E_T$ to be at least 20 GeV. Figure 4 shows the *projected* $\not\!\!E_T$ performance for the e/μ final state.

We have also investigated the transverse mass of the W as an alternative variable to suppress $Z/\gamma^* \to \tau\tau$ and found its performance to be comparable. More details can be found in Appendix C.

3.6 Jet veto

The expected production cross section for WW at a center-of-mass energy of 10 TeV is roughly a factor of five smaller than $t\bar{t}$. As top quarks decay to Wb, both $t\bar{t}$ and WW have two W bosons in the final state. Additionally, $t\bar{t}$ events produce two b-quarks, observed as jets in the detector. In this section, we show how vetoing events with jets above some transverse energy threshold can be used to reduce the $t\bar{t}$ and single top backgrounds while maintaining high efficiency for WW.

Figure 5 shows the transverse energy and pseudorapidity distributions for generator level jets in $t\bar{t}$ and single-top Monte Carlo samples. A $t\bar{t}$ event will in general give more than one entry per plot, as there are two b-quarks per event to which generator level jets can be matched. We conclude that the vast majority of jets from b quarks are central, with a most likely transverse energy of about 60 GeV.

In this study we only consider jets in the central region, with $|\eta| < 3.0$. (We will investigate extending the jet veto to higher η when its behavior in the presence of beam is understood. In Appendix D we show what might be gained by extending the jet veto to higher η.) We have investigated various jet reconstruction techniques: calorimeter jets, track jets, and *jets plus tracks* (JPT) [13]. Table 3 compares the event yields for WW, $t\bar{t}$, and single top for a variety of different E_T thresholds, applied to different jets. For comparison we included generator level jets (GenJets). Figure 6 shows the E_T distributions of the highest-E_T jet per event for JPT.

Based on these studies, we choose JPT jets with an E_T threshold of 20 GeV. Events containing any JPT jets in the central region, $|\eta| < 3.0$, above this threshold, are vetoed.

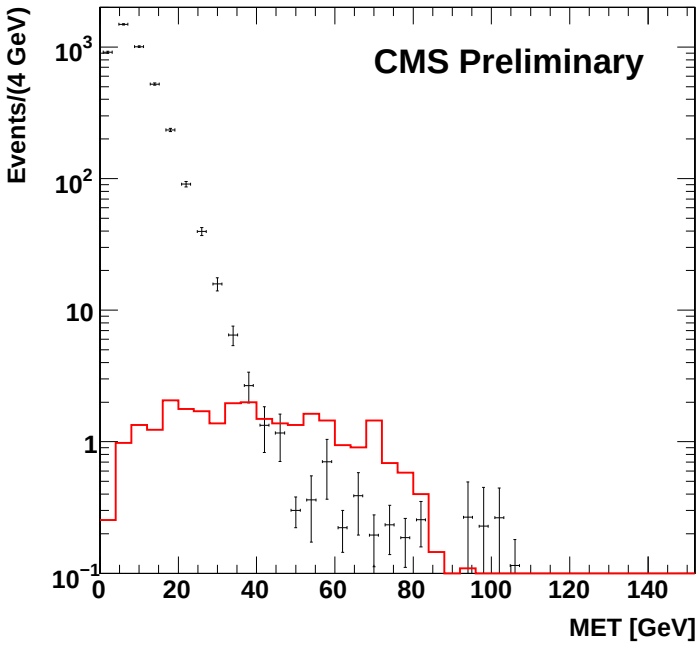


Figure 3: Missing transverse energy distributions for the $ee/\mu\mu$ final state. Signal (red) and all background (black) distributions are made using the standard selection excluding the $\not\!\!E_T$ cut itself. The distributions are normalized to an integrated luminosity of 100 pb^{-1}.

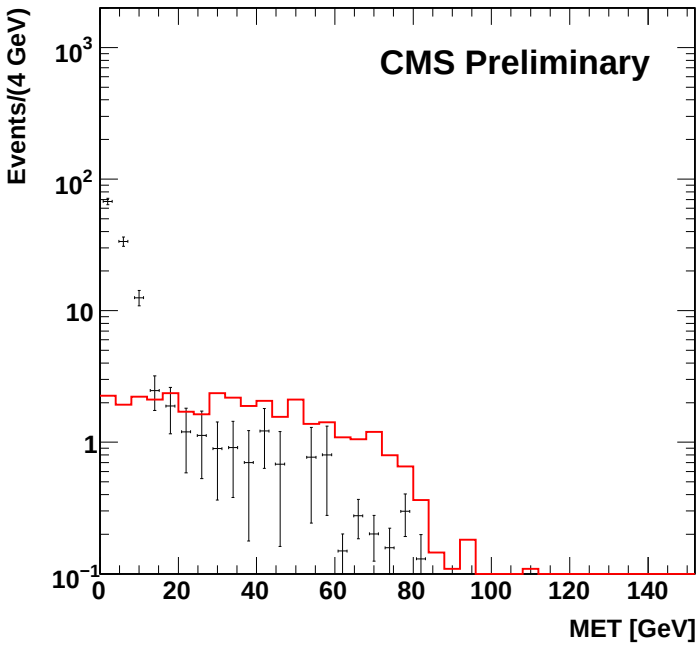


Figure 4: The projected $\not\!\!E_T$ distributions for the $e\mu$ final state. Signal (red) and all background (black) distributions are made using the standard selection excluding the $\not\!\!E_T$ cut itself. The distributions are normalized to an integrated luminosity of 100 pb^{-1}.

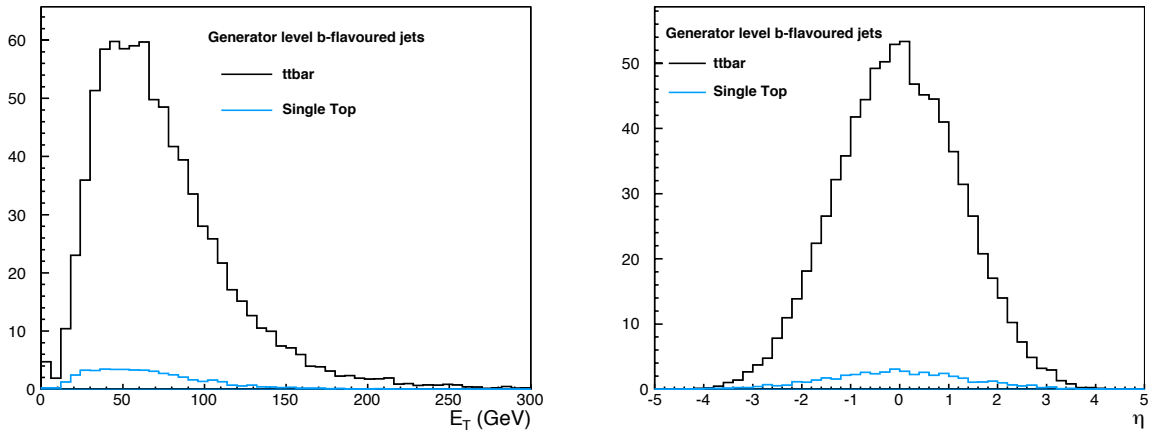


Figure 5: Transverse energy (left) and pseudorapidity (right) distributions of generator level jets matched to b-quarks in $t\bar{t}$ (black) and single-top (blue) Monte Carlo samples. Distributions are normalized to an integrated luminosity of 100 pb^{-1}.

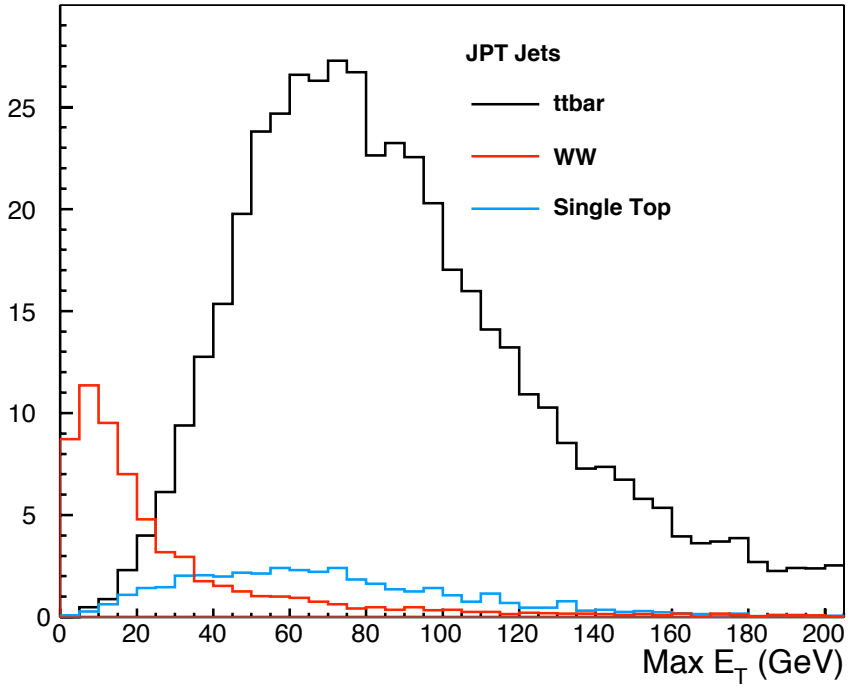


Figure 6: Distribution of the transverse energy of the highest-E_T jet per event using JPT jets with $|\eta| < 3.0$ for WW, $t\bar{t}$, and single top. All samples are normalized to an integrated luminosity of 100 pb^{-1}.

Table 3: Expected event yields normalized to an integrated luminosity of 100 pb^{-1} at a center of mass energy of 10 TeV. All cuts except the soft-muon veto are applied. *Trk-Calo* is combined tracker jet and calo jet veto.

Jet Veto	$t\bar{t}$	tW	WW
JPT @ 25 GeV	8.7 ± 0.6	3.6 ± 0.3	39.8 ± 1.2
JPT @ 20 GeV	4.2 ± 0.4	2.2 ± 0.2	35.1 ± 1.1
JPT @ 15 GeV	1.6 ± 0.3	1.2 ± 0.2	28.5 ± 1.0
Calo Jet @ 15 GeV	18.4 ± 0.9	5.8 ± 0.3	42.2 ± 1.2
Calo Jet @ 10 GeV	6.7 ± 0.5	3.0 ± 0.2	35.9 ± 1.1
Trk-Calo Jet @ 15 GeV	9.9 ± 0.7	3.7 ± 0.3	39.2 ± 1.2
GenJet @ 25 GeV	7.1 ± 0.6	2.9 ± 0.2	38.9 ± 1.2
GenJet @ 20 GeV	2.9 ± 0.4	1.7 ± 0.2	33.5 ± 1.2

3.7 Z veto

To further reduce the Drell-Yan contribution to the selected sample, we veto ee and $\mu\mu$ candidates with a dilepton invariant mass consistent with a Z, $76 < m_{\ell\ell} < 106$ GeV. We remove 80% of the WZ, $> 90\%$ of the ZZ, and all remaining $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ by applying the Z veto in addition to the other selection cuts. The WW efficiency of the veto is approximately 80% in the same-flavor final states.

3.8 Muon vetoes

Among the various types of electron fakes, we identified a special case when a muon fakes an electron, due to significant energy deposition in the calorimeter caused by final state radiation. Therefore, we require that there are no muons within a cone of $\Delta R = 0.1$ around the electron candidate. The muon cuts used are the secondary muon selections described in Section 3.2.

We also veto all events that have an extra muon passing the soft muon selection requirements after all selection cuts are applied. This suppresses some of the remaining $t\bar{t}$, WZ and ZZ background events.

4 Expectations from Monte Carlo

4.1 Selection efficiency

Table 4 shows the effect of each selection cut on the signal and background yields. Compared to the triggered sample, the final selected sample contains four orders of magnitude less background. Table 5 contains the same information for the $e\mu$ final state only.

4.2 Expected yields

As shown in Table 6, we expect a total of about 35 WW events. A total of ≈ 12 background events pass all our cuts. Note that this is not the estimated number of background events in this analysis, but rather the Monte Carlo truth number of background events that pass all cuts. We developed techniques to estimate the dominant backgrounds from data as described in detail in the following sections.

We also considered contributions from other physics processes besides those listed in Table 6. A vector boson associated production with heavy quarks was found to be a small effect, so for simplicity we relied on the Madgraph W+jets sample to cover this physics process contribution. See Appendix G for details.

The multi jet QCD processes is another potential source of background. The requirements to have two energetic well isolated leptons and significant missing transverse energy in the event makes it very unlikely for any QCD events to pass the final event selection.

Table 4: Selection cuts and their effect on signal and background yields. The cut in the n-th column is applied in addition to the cuts in the preceding $n-1$ columns. The "base selection" requires opposite-sign leptons with $p_T > 20$ GeV. Event yields are normalized to 100 pb^{-1}. In addition to the default inclusive WW sample, an exclusive $WW \to l^+\nu l^-\bar{\nu}$ sample has been used. The background total is the sum of the yields of all the non-WW samples.

Process	Base	Trigger	Lepton ID	Isolation	Jet veto	$\not\!\!E_T$	Z veto	μ veto
WW	314.0	288.2	156.6	115.5	71.4	38.0	35.1	35.0
$WW \to l^+\nu l^-\bar{\nu}$	193.9	189.7	157.0	121.0	72.5	39.4	36.6	36.6
WZ	157.2	146.6	101.7	76.4	12.2	3.7	1.6	1.0
ZZ	85.6	81.5	67.0	48.8	8.8	3.1	0.3	0.3
W+jets	10351.6	9327.6	427.5	9.8	7.3	4.7	4.7	4.7
$Z/\gamma^* \to ee$	52172.3	51863.5	45228.1	29756.5	21256.6	1.1	0.2	0.2
$Z/\gamma^* \to \mu\mu$	58132.2	55733.5	50969.9	44741.3	31026.1	3.5	1.0	0.9
$Z/\gamma^* \to \tau\tau$	1304.1	1198.5	537.8	350.0	238.1	0.7	0.7	0.7
$t\bar{t}$	4832.8	4481.1	1605.4	780.7	6.6	4.6	4.2	3.3
tW	277.3	258.5	106.6	65.2	3.3	2.4	2.2	1.9
Background total	127313.0	123090.8	99044.0	75828.8	52559.1	23.7	15.0	12.9

Table 5: Selection cuts and their effect on signal and background yields in the $e\mu$ final state. All comments from Table 4 apply.

Process	Base	Trigger	Lepton ID	Isolation	Jet veto	$\not\!\!E_T$	Z veto	μ veto
WW	149.7	139.6	79.5	56.6	35.3	24.4	24.4	24.4
$WW \to l^+\nu l^-\bar{\nu}$	98.7	97.1	79.0	59.9	36.3	25.6	25.6	25.6
WZ	34.9	32.0	11.2	7.9	1.2	1.0	1.0	0.7
ZZ	10.7	9.9	2.7	1.7	0.1	0.1	0.1	0.1
W+jets	5163.6	4611.2	201.6	7.3	5.7	4.7	4.7	4.7
$Z/\gamma^* \to ee$	129.6	126.3	20.0	0.7	0.4	0.0	0.0	0.0
$Z/\gamma^* \to \mu\mu$	1485.8	1438.1	49.9	7.3	4.2	0.7	0.7	0.5
$Z/\gamma^* \to \tau\tau$	594.1	555.2	272.2	169.5	115.2	0.7	0.7	0.7
$t\bar{t}$	2428.2	2305.8	815.2	387.6	3.1	2.5	2.5	2.1
tW	138.8	132.2	54.8	32.9	1.7	1.5	1.5	1.3
Background total	9985.9	9210.6	1427.6	614.8	131.7	11.1	11.1	10.0

Table 6: Expected event yields for dominant processes at 10 TeV with 100 pb^{-1} of integrated luminosity. Errors represent statistical uncertainties due to limited number of events in corresponding Monte Carlo samples. Note: $W\gamma$ event yield is estimated from the full Madgraph W+jets by looking at the Monte Carlo truth information.

	$Z/\gamma^* \to ee$	$Z/\gamma^* \to \mu\mu$	$Z/\gamma^* \to \tau\tau$	$t\bar{t}$	tW
ee	0.22 ± 0.22	0.00 ± 0.00	0.00 ± 0.00	0.44 ± 0.14	0.27 ± 0.08
$\mu\mu$	0.00 ± 0.00	0.35 ± 0.25	0.00 ± 0.00	0.75 ± 0.18	0.37 ± 0.09
$e\mu$	0.00 ± 0.00	0.52 ± 0.30	0.67 ± 0.39	2.08 ± 0.30	1.29 ± 0.16
Total	0.22 ± 0.22	0.87 ± 0.39	0.67 ± 0.39	3.27 ± 0.38	1.94 ± 0.20

	W+jets(excluding $W\gamma$)	$W\gamma$	WZ	ZZ	WW
ee	0.00 ± 0.00	0.00 ± 0.00	0.09 ± 0.04	0.12 ± 0.02	3.88 ± 0.38
$\mu\mu$	0.00 ± 0.00	0.00 ± 0.00	0.20 ± 0.05	0.15 ± 0.03	6.75 ± 0.50
$e\mu$	3.11 ± 1.27	1.55 ± 0.89	0.68 ± 0.10	0.05 ± 0.02	24.40 ± 0.94
Total	3.11 ± 1.27	1.55 ± 0.89	0.97 ± 0.12	0.32 ± 0.04	35.04 ± 1.13

Table 7: WW fake estimates. All other processes contribute fewer than 0.05 events.

	WW	W+jets
all	0.5 ± 0.0	1.9 ± 0.2
$\mu\mu$	0.1 ± 0.0	0.1 ± 0.1
$e\mu$	0.4 ± 0.0	1.6 ± 0.2
ee	0.1 ± 0.0	0.3 ± 0.1

5 Background estimation

The precision to which the WW cross section can be measured is determined by two factors: the precision with which we know the signal efficiency, and the precision with which we know the background contributions remaining after the selection requirements have been applied.

Some backgrounds are expected to be well modeled by Monte Carlo; WZ and ZZ fall in this category. Other backgrounds are expected not to be well modeled by Monte Carlo: W+jets, where we are sensitive to the details of lepton fakes; Drell-Yan, where we are sensitive to the extreme tails of the $\not\!\!E_T$ distribution; and $t\bar{t}$, where we are sensitive to the low-energy tail of b jets, fall in this category. The data-driven methods required to estimate these backgrounds are described in this section.

5.1 W+jets background

In the context of this analysis, only primary W, Z, and top decay leptons are considered to be true leptons. All other reconstructed leptons are considered fake. For example, in addition to instrumental lepton fakes, leptonic bottom and charm decays and leptons from conversions are also considered fakes.

The W+jets process is the dominant source of lepton-fake background. The data-driven method described in detail in a dedicated note [6] provides a prediction for this background in both shape and normalization. The typical fake di-lepton signature from this process consists of one real lepton from the W decay and one fake lepton from a jet mis-identified as either an electron or a muon.

Applying the fake rates to the full MC sample yields a total number of predicted fake events of 2.4 ± 0.2 in reference to the full number of selected events of 47.9 ± 2.1. The total prediction contains a contribution from real leptons, mostly from WW, that fail the lepton selection cuts. Table 7 shows the size of this contribution. Since the electron selection efficiency will be known when we perform this analysis on data, we can correct for the real-lepton contribution. The error here is statistics only. The procedure for deriving the systematics once data is available is described in [6]. Past experience [12] indicates that a systematic error of 20% might be expected.

We also developed an alternative W+jets background estimation method using the sideband of the isolation distribution of the leptons. The shape for isolated leptons is extracted from $Z \to ll$ data and the shape for non-isolated leptons is extracted from QCD samples. Details of this method and its application can be found in [6]. The method predicts 1.8 ± 0.4 events, which is consistent with the fake rate method estimation. The error here includes systematics. Since this method accounts for the full shape of the isolated lepton distribution, the prediction is not biased by a real-lepton contamination.

Combining the two prediction methods, we take the average as an estimate of the W+jets background and the difference as an estimate of the error. Therefore, we predict the W+jets background contribution to be 2.1 ± 0.6 events.

In order to assess the consistency of this prediction with the Monte Carlo estimation we excluded $W\gamma$ events (1.6 ± 0.9) from the full W+jets Madgraph sample (4.7 ± 1.6). A detailed examination of the nine events in the full sample showed that three of them are $W \to \mu\nu\gamma$. In these events the γ is emitted at a large angle, converts, and gives a reconstructed electron. The W+jets background estimation methods described above are not expected to capture these events. Thus, the prediction (2.1 ± 0.6 events) should only be compared with 3.1 ± 1.3 W+jets events where one of the leptons is from hadronic activity in the event.

The remaining 1.6 ± 0.9 W+jets events, corresponding to $W\gamma$ final state radiation, need to be estimated

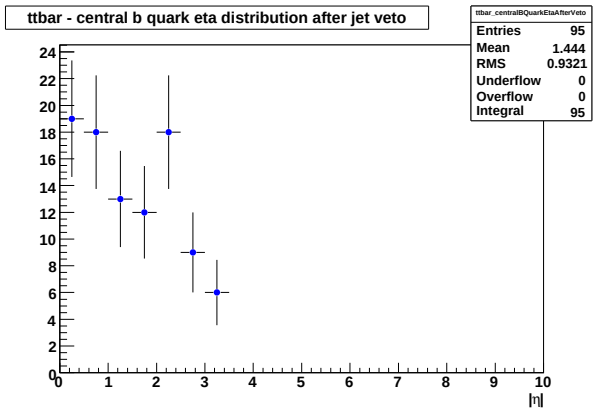


Figure 7: Absolute pseudo-rapidity distribution of the most central b quark from $t\bar{t}$ after the jet veto.

in a different way, *e.g.*, using a dedicated radiative Monte Carlo such as that of Baur, which is at present not available in CMS. For the moment we take the Pythia based estimate and assign a 100% systematic uncertainty to it. We note that application of a dedicated conversion veto would significantly reduce the impact of this background. A prototype of such a veto has been developed by us [7], but is not yet used in this analysis.

It is worth to point out that these methods applied on data would estimate the multi jet QCD contributions as well as W+jets contributions. See Appendix I for more details.

5.2 Top background

Top processes, $t\bar{t}$ and tW, represent one of the major backgrounds for this analysis. Estimation of their contribution is complicated by the unknown jet veto efficiency, which is hard to simulate reliably. Topologically $t\bar{t}$ is $W^+W^-b\bar{b}$ and tW is W^+W^-b, which differ from WW only by presence of extra b jets.

In order to estimate the remaining top background we may count a number of additional soft muons in the event coming from b jets and then estimate the total top background by dividing this number by the soft-muon tagging efficiency. The method works because most of the remaining top background events have b jets within the acceptance of the muon detector, but their measured energy is lower than the veto threshold (Fig. 7). Since we limit jet veto to the $|\eta| < 3.0$ region, the relative fraction of b quarks outside of the acceptance region is increased. In this analysis the effect is small and can be ignored. If better control of this systematic effect is necessary, we can extend the jet veto to the forward calorimeter region..

The method relies on the purity of the muon selection and the fact that most extra muons in data are from top decays. To avoid counting muons from vector boson decays as muons from the b quarks, we require the top tagging muon to either be non-isolated (relative isolation < 0.9) or have low momentum ($p_T < 20$ GeV).

The overall top tagging efficiency after all cuts, and before the extra muon veto is:

- $t\bar{t}$: 22.1%
- tW: 13.9%
- all top combined: 19.2%

The strategy to apply the method on data is the following. First we estimate the top tagging efficiency in top enriched samples (1 or 2 jets) on Monte Carlo simulated events. Next we compare the results with data and verify that they are consistent. After that we extract the tagging efficiency from Monte Carlo samples with the jet veto applied and use it for the remaining top background estimation. This approach allows for better control of systematics effects due to changes in the kinematic distributions of top events with the application of the jet veto.

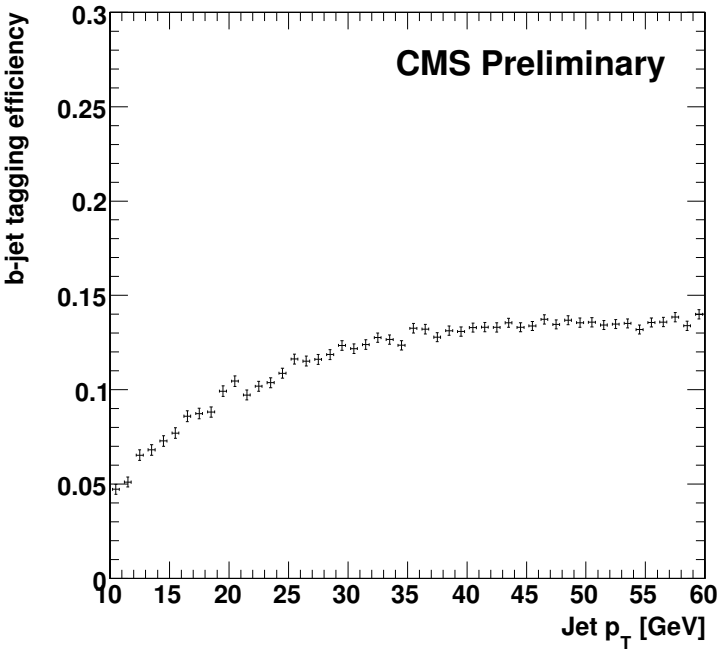


Figure 8: The b-jet tagging efficiency as a function of the jet transverse momentum counting the number of soft muons in the jet.

Given the small number of expected top events with 100 pb^{-1} integrated luminosity, we will have at most one or two tagged events with large statistical uncertainty. If we see no tagged events, we can still estimate the expected uncertainty on the top background contribution. Taking current expected central values, the tagged number of events will most likely be 1 ± 1, which corresponds to 4 ± 4 top background events (assuming 20% top tagging efficiency and the fact that we veto tagged events).

It is worth noting that there can be a correlation between the jet veto and the top tagging with soft muons. Figure 10 shows that with the current choice of jet veto the effect is negligible. Nevertheless, the procedure may need to be modified depending on what kind of jets are used for the veto. If the jet reconstruction algorithm uses muons to compute jet energy, most b jets at low energy will receive a significant increase in energy and more $t\bar{t}$ events will fail the jet veto, reducing the number of events counted with the top tag. This does not change the final selection in any way, because all these events are vetoed anyway, either by the jet veto or the extra muon veto, but it may lower the top tagging efficiency leading to a large statistical uncertainty on the remaining $t\bar{t}$ background. In order to restore the tagging efficiency and improve the statistical uncertainty in such cases, we can subtract the muon momentum contribution from the jet energy and then apply the jet veto requirement.

Figure 8 and Figure 9 show the b-jet tagging efficiency and the light quark mistagging efficiency counting number of muons in the jet. The distributions were made using the $t\bar{t}$ inclusive sample of events without any cuts applied. This sample has large statistics with many jets of different types. The jet flavor is estimated by looking for a hard scatter parton (status 3) in the cone dR=0.3 around the jet direction. For each jet we apply the soft muon counting technique. This method can also be used on data by properly selecting a set of di-jets allowing for a better understanding of the jet tagging effiency.

5.3 Drell-Yan background

The method used to estimate the Drell-Yan contribution to the selected data sample is presented in detail in a dedicated note [5]. The aim of the method is to predict, in a data-driven way, the Drell-Yan yield after applying an $\not\!\!E_T$ cut and a Z-mass veto. This is accomplished by using the near-Z mass region ($76 < m_{\ell\ell} < 106$ GeV) to normalize Monte Carlo to data. The method also provides an estimate for the *peaking component* of WZ and ZZ, *i.e.*, those candidates where the leptons come from the same Z decay.

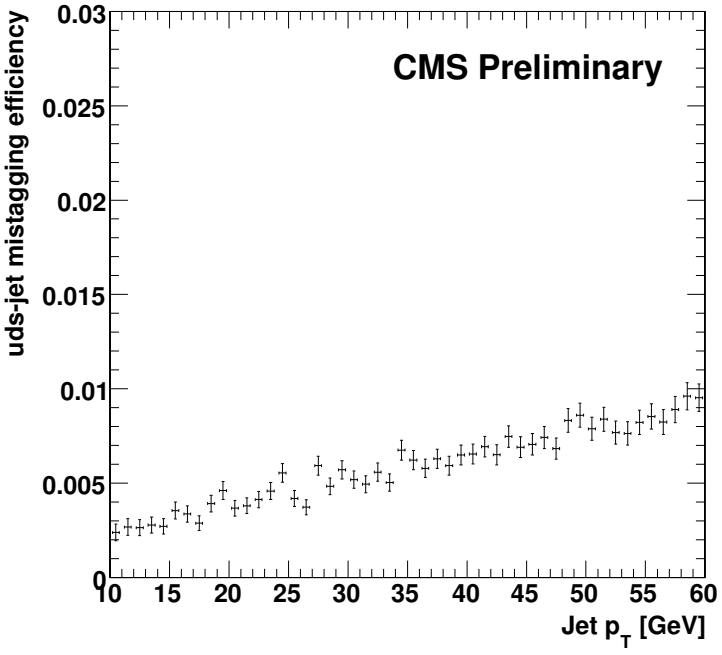


Figure 9: The light quark jet mis-tagging efficiency as a function of the jet transverse momentum counting the number of soft muons in the jet.

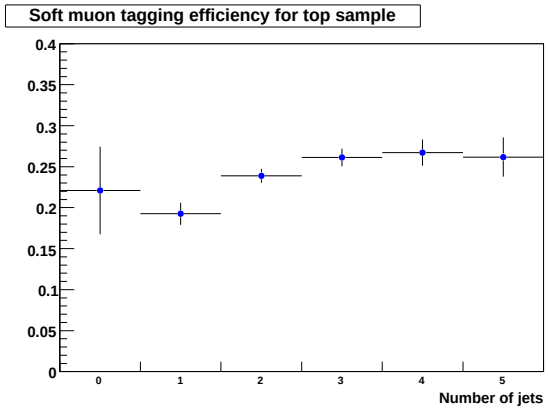


Figure 10: Top tagging efficiency with soft muons on $t\bar{t}$ sample with the default event selection. The efficiency is defined on a sample of top events that passed final selection excluding the extra muon veto. It is the ratio of the number of events that were tagged with soft muons to the total number of events in the sample.

Table 8: Results of the Drell-Yan estimate including the peaking components of the ZZ and WZ backgrounds after all selection requirements computed for an integrated luminosity of 100 pb^{-1}. Results agree with the expected values.

	Predicted DY/ZZ/WZ yield	True DY/ZZ/WZ yield
ee	0.2 ± 0.3	0.4 ± 0.2
$\mu\mu$	0.4 ± 0.4	0.6 ± 0.3

The results of the method are shown in Table 8. The prediction of 0.6 ± 0.5 events in the ee and $\mu\mu$ final states is statistically consistent with the expected yield from the Monte Carlo truth of 1.0 ± 0.4.

A comparison based on higher-statistics samples can be found in [5]. The WW analysis uses Summer/Fall08 samples, while the dedicated note is based on CSA07 samples; Appendix F gives the detailed results of the application to the Summer/Fall08 samples.

6 Systematics

The systematic uncertainties on the expected cross-section measurement of WW can be separated in to three major categories: uncertainties on reconstruction efficiency, uncertainties on background estimations, and uncertainty on luminosity.

6.1 Reconstruction efficiency uncertainties

Most of the selection requirements for the WW event selection, such as triggers, lepton identification and lepton isolation, can be well estimated in data using the tag-and-probe method with a Z sample. Given the large amount of data expected for this sample, the corresponding systematic uncertainties should be well controlled. We assign 4% systematic uncertainty, which is consistent with other analyses [11].

Among the remaining selection requirements, Z and muon vetoes can be safely ignored since they have negligible effect on the reconstruction efficiency. The remaining two selection requirements, $\not\!\!E_T$ and jet veto, account for the dominant systematic effects in the reconstruction efficiency estimation. In both cases we intend to use data to determine the systematics.

In order to estimate the jet veto efficiency for WW event selection, we may look at Z events, which represent another case of a physics process that does not have jets at the leading order. By comparing efficiency extracted from data with the Monte Carlo efficiency, we will derive a correction factor which will define our systematic uncertainty. Given that we do not have any data at the moment, we tried to compare Madgraph Monte Carlo samples with Pythia ones. In these samples, the efficiency of the jet veto was 75% for the Madgraph sample and 70% for the Pythia sample. So we assign 7% systematic uncertainty on the jet veto efficiency. A comparison of WW Madgraph and Pythia samples shows a similar discrepancy of 7% (jet veto efficiency is 57.2% and 61.4% respectively).

The signal efficiency of the $\not\!\!E_T$ cuts depends on the core of the $\not\!\!E_T$ resolution. We can measure this resolution, *i.e.* the x and y components of $\not\!\!E_T$ for a sample that has zero true $\not\!\!E_T$, in data using again the Z peak. The procedure here would be to compare the x and y component of $\not\!\!E_T$ between data and Monte Carlo for the Z sample, and derive from this a systematic variation with which the $\not\!\!E_T$ in WW in Monte Carlo can be convoluted in order to estimate the corresponding change in efficiency. In the absence of data, we assign 5% systematic uncertainty for this, which is likely an overestimate since the corresponding systematics in CDF [12] amounted to only 1% .

There may also be differences in the WW efficiency between nature and our Monte Carlo because the LO physics model in the Monte Carlo does not fully describe nature. We propose to study this in the future by comparing NLO and LO Monte Carlo, *i.e.* mc@NLO and Pythia or Madgraph. As we do not have an mc@NLO sample available at this point, we compare $\not\!\!E_T$ cut efficiency for Madgraph and Pythia samples. Based on this comparison, we estimate the systematic uncertainty to be of the order of 5% ($\not\!\!E_T$ selection efficiency is 58.4% and 55.6% respectively). We note that this is likely to be an underestimate given that CDF observed a 12% difference between mc@NLO and Pythia in the same final state [12].

Table 9: Reconstruction efficiency systematic uncertainty.

Two lepton selection	4%
Jet veto	7%
Missing transverse energy (resolution function)	5%
Missing transverse energy (next to leading order effects)	10%
Parton distribution function	2%
Total	14%

We thus double the systematics in this case to 10% until we will have evaluated it with mc@NLO Monte Carlo once CMS will have generated such a Monte Carlo.

Finally, we take into account the systematic uncertainty due to our knowledge of the parton density functions (PDFs) used in the simulation. To assess the uncertainty, we used the PDF weights method [19]. We count the weighted total number of events generated and passing our selection. The weighting is performed on an event-by-event basis for each of the 40 error sets included in CTEQ6.1. Using the modified tolerance method on the resulting set of acceptance times efficiency values we conclude an uncertainty of 2%.

These efficiency systematics are summarized in Table **??**. The total efficiency systematics add up to 14%.

6.2 Background estimation uncertainties

Table 10 shows the expected contribution from various backgrounds as we estimate it on the combined Monte Carlo sample. Since the purities of the final selection for ee, $\mu\mu$ and $e\mu$ final states are comparable, we combine events and count total the number of events for each process.

Contributions of $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ to the ee and $\mu\mu$ final states are estimated with the method described in section 5.3. This method also estimates ZZ and WZ contributions when both leptons from the Z decay are correctly reconstructed. The total for these contributions including systematic errors is 0.6 ± 0.5.

For $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ contributions to the $e\mu$ final state as well as $Z/\gamma^* \to \tau\tau$ and $W\gamma$ we estimate the background from Monte Carlo, and assign 100% systematic uncertainty because in all these cases, we are sensitive to the tails of distributions that may not be well modeled in the Monte Carlo, and we have not come up with a strategy to measure them from data. This 100% systematic error is added in quadrature to the error from Monte Carlo statistics to arrive at 0.7 ± 0.8 for $Z/\gamma^* \to \tau\tau$ contribution and 0.5 ± 0.6 for $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ contributions to the $e\mu$ final state. Therefore the total contribution of the Drell-Yan processes estimated from Monte Carlo is 1.2 ± 1.0. $W\gamma$ contribution is estimated to be 1.6 ± 1.8.

As it was mentioned earlier, ZZ and WZ yields from off-peak $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ contributions to ee and $\mu\mu$ final states are estimated with the Drell-Yan background estimation method from data. The remaining ZZ and WZ backgrounds, most importantly in the $e\mu$ channel, are estimated from Monte Carlo. We assign half of the next-to-leading order k factor as the uncertainty for these latter contributions. The systematic error is added in quadrature with the error from Monte Carlo statistics to arrive at 0.9 ± 0.5.

The top contribution was estimated in section 5.2 as 4.0 ± 4.0.

The W+jets background is estimated as the average of the two methods presented in section 5.1. The uncertainty is estimated as the full difference between the predictions from the two methods, or $\sim 30\%$. We thus have chosen a systematics larger than the systematics of either of the two techniques individually. We believe this to be rather conservative.

The total estimated background including systematic uncertainties is 10.4 ± 4.6.

Table 10: Expected event yields for dominant processes at 10 TeV with 100 pb^{-1} of integrated luminosity extracted from combined Monte Carlo sample, which is treated as "data" sample. Multi-boson contributions are extracted from exclusive Monte Carlo samples. The errors represent statistical and systematic uncertainties combined.

data-derived			Monte Carlo based		
DY/$WZ/ZZ \to ee/\mu\mu$	$t\bar{t} + tW$	W+jets	DY (other)	WZ not from Z	$W\gamma$
0.6 ± 0.5	4.0 ± 4.0	2.1 ± 0.6	1.2 ± 1.0	0.9 ± 0.5	1.6 ± 1.8

6.3 Luminosity uncertainties

Luminosity uncertainty will be determined when data is available. For now all Monte Carlo analyses were recommended to use an estimate of 10%.

7 Cross section determination

Based on the results of the previous sections we may estimate the expected cross-section and its uncertainty. The total number of events after final selections is 35.0 ± 1.1 for WW plus 12.9 ± 1.8 for the sum of all backgrounds. We thus may expect around 48 events passing all cuts at 100 pb^{-1} of data.

The signal significance estimation depends on the question that we want to ask. In data we estimate the signal significance by running a set of Monte Carlo pseudo-experiments assuming no signal and finding the probability to observe a given number of events or more. Since we do not have data yet, instead of assuming that we observed 48 events total, we may want to estimate our chances to observe more than 5σ deviation assuming that total number of observed events can fluctuate (see Appendix H). Performing a set of Monte Carlo pseudo-experiments we found that we have more than 60% chance to get 5σ signal significance with 100 pb^{-1} at 10 TeV center of mass energy.

The expected signal WW contribution can be estimated based on the total event yield by subtracting the background estimation as 37.5 ± 8.3 events, where the uncertainty is added in quadrature to the uncertainty on the total event count (48) and the background estimation.

In order to estimate the WW cross section, we need to take into account the uncertainty on the signal reconstruction efficiency, which is 14%, and the luminosity systematic uncertainty which is expected to be of the order of 10%. Adding these uncertainties in quadrature with the uncertainty on the signal yield after the background subtraction (22%), we estimate overall uncertainty on the cross section measurement to be of the order of 28%.

8 Conclusion

We have presented expectations of observing W boson pair production in the dilepton final state with $\not\!\!E_T$ and no jets using 100 pb^{-1} of CMS data at 10 TeV center of mass energy. We estimate an uncertainty on the cross section estimate of 30% or better.

We developed a number of data-driven methods to control all major background contributions. The largest contribution to the uncertainty on the background estimation is found to be from the top background, which we estimate from "data" using the top tagging technique with soft muons. The precision of the method is statistics limited.

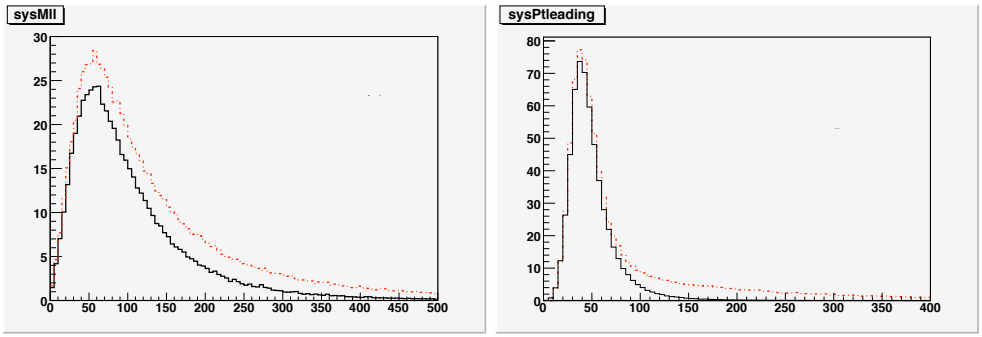


Figure 11: Distributions of generator level di-lepton invariant mass (left) and leading lepton p_T (right) for the Standard Model (solid black), as well as aTGC (dash-dotted red) with $\lambda_Z = \lambda_\gamma = 0.30$. The scale is for an arbitrary integrated luminosity.

Appendices

A Anomalous triple gauge couplings

At this point, we have not studied the sensitivity of CMS for measuring triple gauge couplings in WW, however, a simple quantitative argument can be used to set the scale of what may be measurable with 100 pb^{-1}. This could be done by choosing a specific set of anomalous triple gauge couplings (aTGC) and estimating the affect on the predicted event yield.

The specific couplings chosen are $\lambda_Z = \lambda_\gamma = 0.30$ and a cut-off scale of 2 TeV. These particular values where chosen because they represent roughly the published 95% C.L. limits [10]. The total cross section for this choice of couplings is $\sim 40\%$ higher than the Standard Model, with a significant fraction of the excess at large p_T for the leading lepton.

The generator level di-lepton invariant mass and leading lepton p_T distributions for the Standard Model and this aTGC choice are shown in Figure 11. The integrated luminosity scale of these figures is arbitrary. The leading p_T figure has 18% of the total cross section for the chosen coupling beyond p_T of 150 GeV. If we simply scale this by our expected yield, implicitly assuming the efficiency to be flat in leading lepton p_T, then we would predict 9 events above 150 GeV for these couplings, compared to 0.4 events expected in the standard model.

This simple argument appears to indicate that 100 pb^{-1} of CMS data should be sufficient to provide an interesting result on anomalous couplings in WW.

B Electron selection details

To reject fake electrons, a standard category-dependent electron ID is applied [8]. The expected performance of this electron ID is described in [9]. The POG-recommended set of thresholds was optimized using `CMSSW_1_6_X`. In the `CMSSW_2_2_X` some variables used in the selector have changed: H/E was changed to compute the HCAL E_T using the nearest HCAL tower to the super cluster and the log weighting factor used in $\sigma_{\eta\eta}$ was changed. Therefore the thresholds applied to these variables were updated in consultation with the E/gamma group [20].

The electron signature in the detector greatly depends on how much energy it radiates. We define the fractional energy loss to bremsstrahlung as determined by the track momentum p_{in} estimated at the interaction point and the track momentum p_{out} when the track exits the tracker.

$$f_{\text{brem}} = \frac{p_{\text{in}} - p_{\text{out}}}{p_{\text{in}}} \tag{2}$$

We require $E/p > \max[0.8, 0.9(1 - f_{\text{brem}})]$, where E is the supercluster energy, p is the track momentum at the vertex.

Electrons are defined as being in the endcap if $|\eta| > 1.479$ and categorized according to their E/P and $fBrem$ as follows.

- Category-0: fBrem > 0.06 barrel (0.10 endcap) and $0.8 < E/P < 1.2$.
- Category-1: fBrem < 0.06 barrel (0.10 endcap).
- Category-2 otherwise.

By cutting more tightly according to the expected signal to background ratio in each category, extra rejection power is possible compared to a simpler technique without categorization. The variables used to distinguish between real and fake electrons are now described.

- H/E is the ratio of the E_T of the closest HCAL tower to the E_T of the electron reconstructed in the ECAL.
- $\sigma_{\eta\eta}$ is the covariance defined to be $\sum_i w_i(\eta_i - \bar{\eta}_{5\times5})^2/\sum_i w_i$ where $w_i = 4.7 + \ln(E_i/E_{5\times5})$.
- $\Delta\phi_{in}$ is the difference between the ϕ co-ordinate of the electron measured in the ECAL and the projection of the track as measured at the primary vertex and projected through the magnetic field to the ECAL.
- $\Delta\eta_{in}$ is the difference between the η co-ordinate of the electron measured in the ECAL and the projection of the track as measured at the primary vertex and projected to the ECAL.
- E_{seed}/P_{in} is the ratio of the energy of the seed cluster that initiated the bremsstrahlung recovery process to the momentum of the track at the vertex.

The variable, $\sigma_{\eta\eta}$ was corrected for the η dependence of the crystal size in the endcap in the usual way by subtracting $0.2(|\eta| - 2.3)$ from the computed value. If E/P is greater than 1.5 then the $\Delta\phi_{in}$ threshold is loosened to 0.09 in the barrel region and 0.092 in the endcap region. The category dependent thresholds are tabulated in Tables 11, 12 and 13.

Table 11: Electron cuts for category-0, Bremming.

Parameter	Threshold (barrel)	Threshold (endcap)
H/E	< 0.041	< 0.034
$\sigma_{\eta\eta}$	< 0.0118	< 0.0271
$\Delta\phi_{in}$	< 0.032	< 0.025
$\Delta\eta_{in}$	< 0.0055	< 0.0060
E_{seed}/P_{in}	> 0.24	> 0.32

Table 12: Electron cuts for category-1, Low brem.

Parameter	Threshold (barrel)	Threshold (endcap)
H/E	< 0.028	< 0.012
$\sigma_{\eta\eta}$	< 0.0112	< 0.0263
$\Delta\phi_{in}$	< 0.016	< 0.035
$\Delta\eta_{in}$	< 0.0030	< 0.0055
E_{seed}/P_{in}	> 0.94	> 0.83

Table 13: Electron cuts for category-2, Bad track.

Parameter	Threshold (barrel)	Threshold (endcap)
H/E	< 0.025	< 0.016
$\sigma_{\eta\eta}$	< 0.0104	< 0.0256
$\Delta\phi_{in}$	< 0.0525	< 0.065
$\Delta\eta_{in}$	< 0.0065	< 0.0075
E_{seed}/P_{in}	> 0.11	N/A

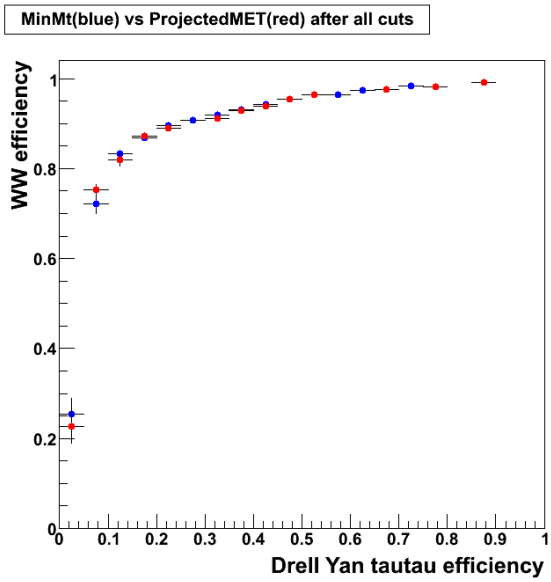


Figure 12: Efficiency of WW vs $Z/\gamma^* \to \tau\tau$ selection for the projected $\not\!\!E_T$ and the minimal transverse mass of the W bosons, with the default selection. Current selection requirements correspond to 85% WW efficiency.

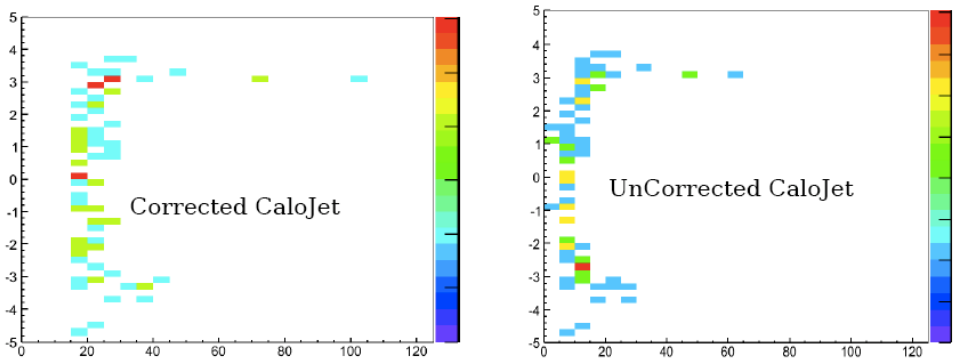


Figure 13: Scatter plot showing jet E_T against η for $t\bar{t}$ Monte Carlo events after all other cuts including the central jet veto are applied. A significant fraction of jets are found at large η. Both corrected (left) and uncorrected (right) jet distributions show this behavior.

C Using the transverse mass of the W for $Z/\gamma^* \rightarrow \tau\tau$ suppression

This analysis rejects $Z/\gamma^* \rightarrow \tau\tau$ mainly through the projected $\not\!\!E_T$ defined in Section 3.5. Another popular selection variable that helps to suppress $Z/\gamma^* \rightarrow \tau\tau$ background events is the transverse mass of the W, which is defined as the invariant mass of the lepton and $\not\!\!E_T$ projected on the transverse plane, which can be expressed in a convenient form as:

$$M_T = \sqrt{2\not\!\!E_T P_T^{\ell}(1 - \cos\Delta\phi)} \tag{3}$$

Since we have two W bosons in the final state, the smaller transverse mass of the two is used.

Both the projected $\not\!\!E_T$ and M_T use identical inputs and have similar asymptotic behavior, i.e. only when the angle between one of the leptons and $\not\!\!E_T$ is small are the projected $\not\!\!E_T$ and M_T small. Figure 12 shows Drell-Yan background suppression performance for both variables, which is identical at the current level of required background suppression.

D Jet Veto at high η

The present study ignores jets with $|\eta| > 3.0$. It is thus possible that $t\bar{t}$ events can pass the jet veto discussed in section 3.6, not only because the b-quarks are soft, but also due limited pseudo-rapidity acceptance that we use. Figure 13 shows corrected and uncorrected jets in $t\bar{t}$ after all other cuts are applied. Figure 14 shows the same distribution for uncorrected jets which are required to be matched to b-quarks at the generator level.

Table 14 shows the expected number of events that could be removed with a further veto on uncorrected forward calo jets above a given transverse energy threshold. We conclude that approximately 10% of the remaining $t\bar{t}$ events after the central jet veto could be removed at a cost in efficiency for WW of approximately 1%.

While the present analysis does not use such a veto, we expect to evaluate it in more detail once collision data from the forward calorimeter is available.

E Jet Veto using b tagging

One might argue that jets from b-quarks in $t\bar{t}$ events differ from jets in WW not only by their transverse energy, but also by the fact that B-hadrons have significant lifetime. It is thus reasonable to ask if the jet veto discussed in Section 3.6 can be improved by including b-tagging information.

Figure 15 shows the highest b-discriminant found in events that survive after the JPT based jet veto. The Impact Parameter MVA discriminant was used. We investigated several of the tagging algorithms

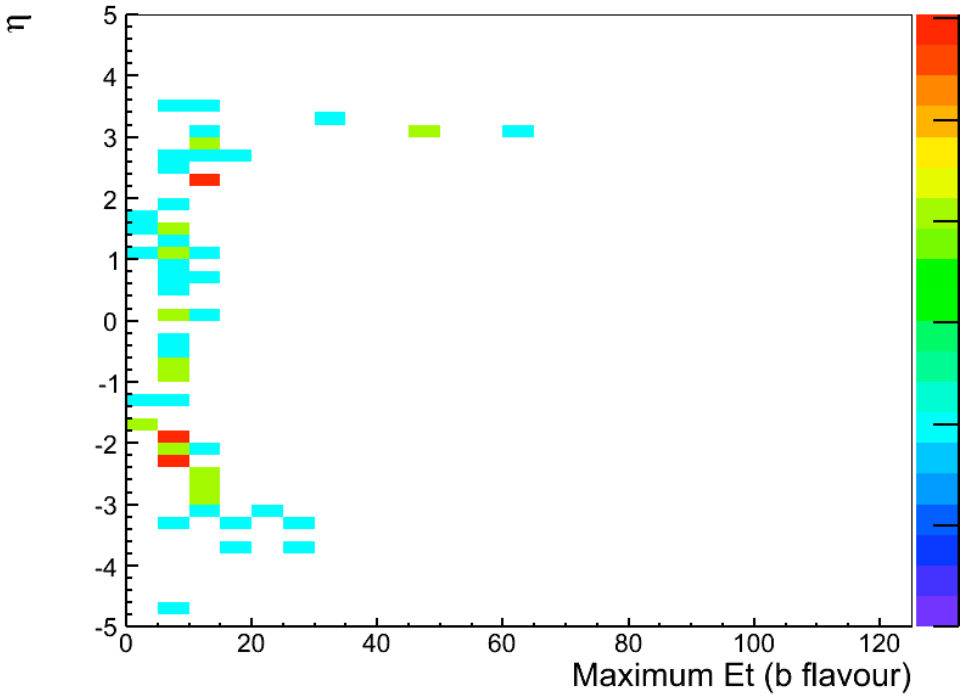


Figure 14: Scatter plot showing jet E_T against η for $t\bar{t}$ Monte Carlo events after all other cuts including the central jet veto are applied. Jets are uncorrected and matched to b-quarks at the generator level. Comparing with Figure 13 indicates that the large transverse energy jets at large η are from b-quarks.

Table 14: Expected number of events vetoed by forward jet veto normalized to an integrated luminosity of 100 pb^{-1} at a center of mass energy of 10 TeV. Without forward jet veto approximately 40 events from WW, 4 from $t\bar{t}$, and 2 from tW survive.

E_T Threshold	WW	$t\bar{t}$	tW
15 GeV	0.80	0.65	0.46
20 GeV	0.49	0.41	0.20
25 GeV	0.20	0.24	0.12

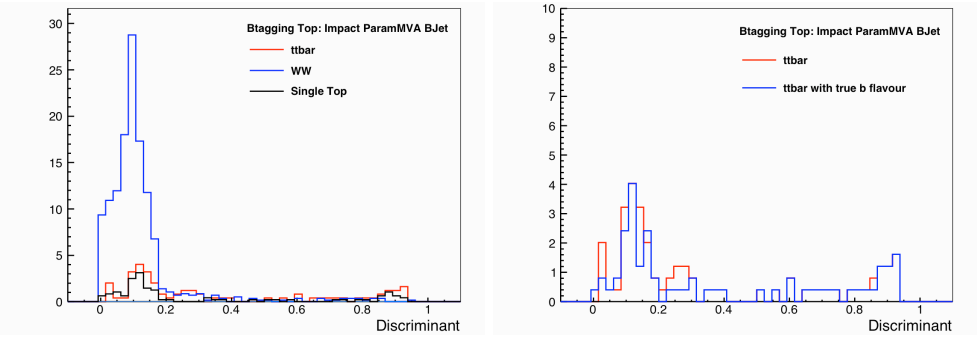


Figure 15: Maximum b-discriminant distribution per event (Impact parameter MVA) after the JPT based jet veto. The left plot compares WW, $t\bar{t}$ and tW, while the right plot overlays reconstructed discriminants for $t\bar{t}$ (red) and $t\bar{t}$ with generator level matched b-quarks (blue). The b-tagging algorithm is applied to the corrected PAT jets that survived the JPT based jet veto. Both plots are normalized to an integrated luminosity of 100 pb^{-1} at a center of mass energy of 10 TeV.

Jet Veto Definition	$t\bar{t}$	WW	tW
JPT @ 25GeV	8.05 ± 0.57	41.66 ± 0.85	3.64 ± 0.28
ImpactParamMVA > 0.2 plus JPT @ 25GeV	3.83 ± 0.39	40.10 ± 0.83	2.21 ± 0.21
JPT @ 20	4.07 ± 0.40	37.09 ± 0.80	2.27 ± 0.22

Table 15: Expected event yields normalized to 100 pb^{-1} of luminosity at a center of mass of 10TeV. All standard analysis cuts except for the soft muon veto are applied.

used in CMS and found some of them to be of comparable quality to this algorithm. It is clear that there is significant information left after the jet veto that could be exploited. One might argue that a more optimal event selection would include a looser JPT veto augmented with b-tagging.

Further investigation is needed to fully understand the use of b-tagging for soft transverse momentum jets, such as those that remain after the JPT jet veto. However, Table 15 provides a preliminary indication of the possible improvement that could be gained.

F Application of the Drell-Yan estimate method

If both leptons have the same flavor, then a background to the di-lepton + MET selection may be introduced from the Drell-Yan process where the MET arises due to instrumental mis-measurement. This background is reduced by ignoring events with a reconstructed di-lepton invariant mass close to the mass of the Z, where the Drell-Yan cross section is enhanced.

To estimate any remaining background, a data driven method has been investigated and studied in detail with the CSA07 samples [5]. This method treats the ignored events as a control region within which the Drell-Yan mass spectrum predicted by simulation can be normalized. The method is re-iterated and the results of its application in this analysis of the Summer/Fall08 samples are presented.

F.1 Method

A scale factor between the number of Drell-Yan events predicted by simulation and measured in data may be computed in the near-Z control region. This is expressed in Equation 4, where the total number of Drell-Yan events selected in data or simulation, N_{DY}, is the sum of the number of events inside and outside the near-Z control region, $N^{in}_{DY} + N^{out}_{DY}$.

$$N^{out\ (est)}_{DY} = \frac{N^{in}_{DY\ DATA}}{N^{in}_{DY\ MC}} \cdot N^{out}_{DY\ MC} \tag{4}$$

For convenience and better factorization of the relevant uncertainties, the two terms which come from simulation can be compressed into a single term, $R_{out/in} = N^{out}_{DY\ MC}/N^{in}_{DY\ MC}$. Thus $R_{out/in}$ is computed and the estimate of the Drell-Yan background outside of the near-Z control region comes from applying the ratio to the number of Drell-Yan events measured inside the control region in data.

Since the event selection requires a large MET, the number of Drell-Yan events in the control region may not be sufficiently larger than other sources of events to assume that the number counted in this region is Drell-Yan dominated. Non Drell-Yan processes which may contribute to the near-Z control region are split into two categories:

- **Peaking backgrounds**, such as WZ and ZZ give a peak in the reconstructed di-lepton invariant mass distribution at the Z mass if both selected leptons come from the Z in the case of WZ, or the same Z in the case of ZZ.
- **Non Peaking backgrounds**, such as WW, $t\bar{t}$, tW and $W+jets$ will give a continuum contribution in the reconstructed di-lepton invariant mass. The shape of this continuum is the weighted sum of the unknown shapes of each background, where the weighting of each contribution is unknown.

If simulation of the peaking backgrounds predicts a value of the ratio, $R_{out/in}$, similar to Drell-Yan then they may be included together with the Drell-Yan itself. It was previously found that the ZZ contribution can be expected to be predominantly due to the $l^+l^-\nu_l\bar{\nu}_l$. Because the neutrinos provide real MET, the ZZ contribution may be large.

The non peaking backgrounds must be estimated from data and subtracted. This may be achieved by measuring the number of events in the control region in the $e-\mu$ final state, $N^{in}_{e\mu\ DATA}$. This number, scaled by a factor taking into account the combinatorics and efficiency to reconstruct the different flavor final state relative to each same flavor final state may be used as an estimate of the non peaking background in the same flavor final state. Thus, the estimate of the number of events outside of the near-Z control region due to the Drell-Yan and ZZ processes can be expressed as in Equation 5.

$$N^{out\ (est)}_{DY/ZZ} = (N^{in}_{ll\ DATA} - k \cdot N^{in}_{e\mu\ DATA}) \cdot R_{out/in} \tag{5}$$

The constant, k, is equal to 0.5 for combinatorics between the $e\mu$ and ll final states, multiplied by a correction due to the difference in efficiency to reconstruct and select a muon compared to an electron. This correction can be determined from the number of ee and $\mu\mu$ events in the control region with no MET requirement applied. This is expressed in Equation 6, where it is assumed that the true number of $q\bar{q} \to e^+e^-$ is equal to the true number of $q\bar{q} \to \mu^+\mu^-$, and that the purely geometric acceptances, A_{ee} and $A_{\mu\mu}$, are the same for electrons and muons.

$$\frac{n^{obs}_{\mu\mu}}{n^{obs}_{ee}} = \frac{N^{true}_{\mu\mu} \cdot A_{\mu\mu} \cdot \varepsilon^2_\mu}{N^{true}_{ee} \cdot A_{ee} \cdot \varepsilon^2_e} = \frac{\varepsilon^2_\mu}{\varepsilon^2_e} \tag{6}$$

$$\therefore k = \frac{1}{2}\sqrt{\frac{n^{obs}_{\mu\mu}}{n^{obs}_{ee}}} \tag{7}$$

Since the control region can be expected to be dominated by Drell-Yan events when no MET requirement is applied, the correction factor can be easily measured from data. The measurement of the correction from data can be compared with simulated Drell-Yan events.

F.2 Results

After applying the same flavor event selection to both the same flavor and $e-\mu$ final states without the Z-veto, the number of events inside and outside the veto region was counted for each of the datasets used. These are shown in Figures 16 and 17 for the di-muon peaking and non-peaking contributions, Figures 18 and 19 for di-electrons and Figures 20 and 21 for the $e-\mu$ final state.

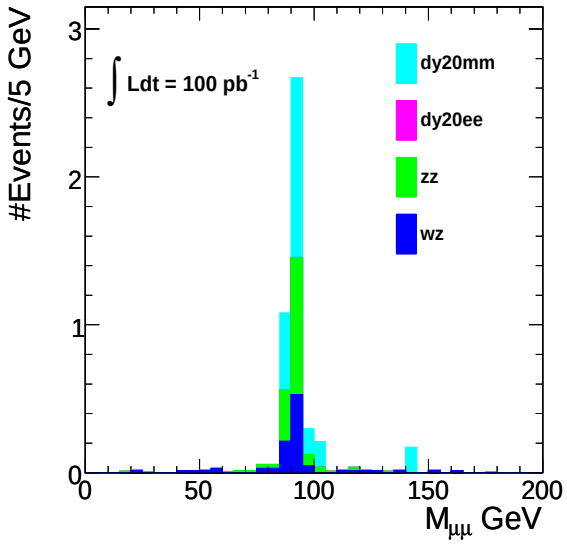


Figure 16: The reconstructed di-muon invariant mass for processes that peak at the Z mass (peaking).

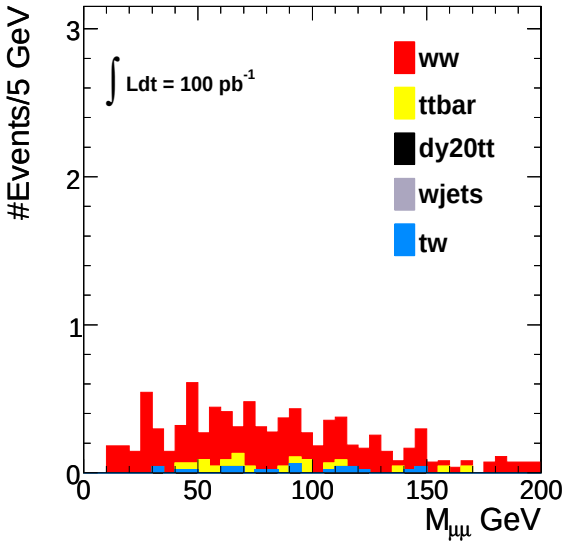


Figure 17: The reconstructed di-muon invariant mass for processes that do not peak at the Z mass (non-peaking).

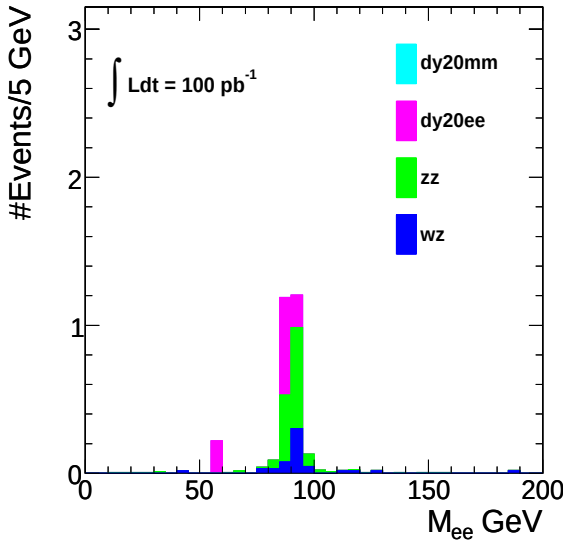


Figure 18: The reconstructed di-electron invariant mass for processes that peak at the Z mass (peaking).

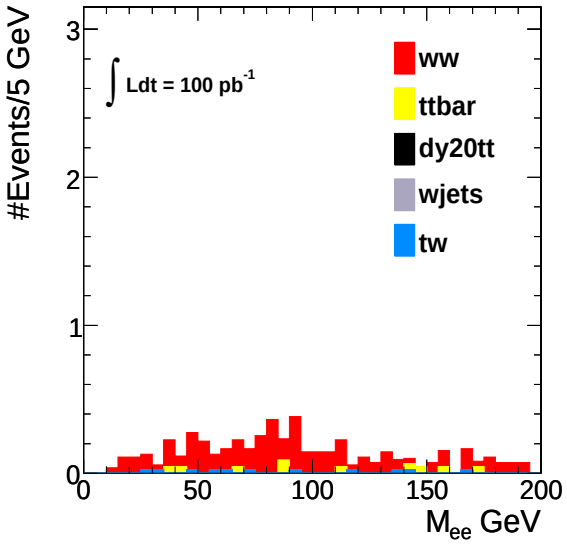


Figure 19: The reconstructed di-electron invariant mass for processes that do not peak at the Z mass (non-peaking).

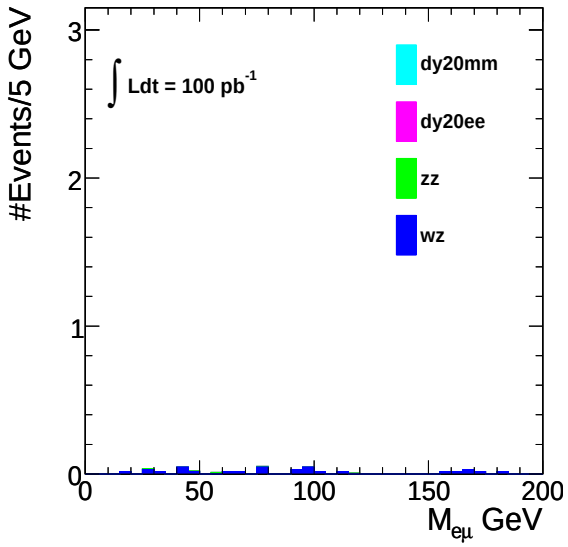


Figure 20: The reconstructed electron-muon invariant mass for processes that peak at the Z mass (peaking).

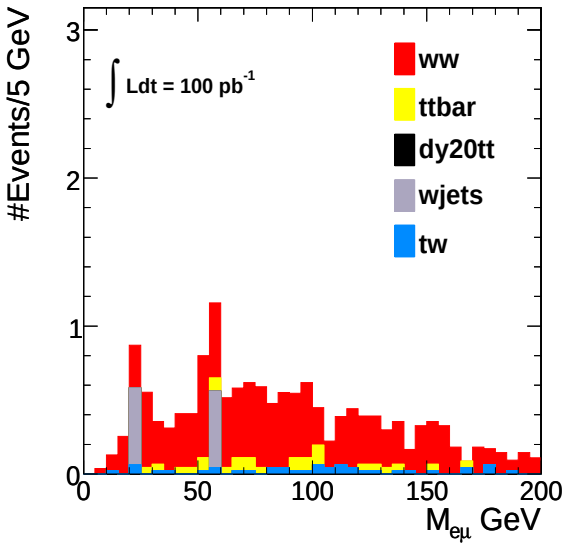


Figure 21: The reconstructed electron-muon invariant mass for processes that do not peak at the Z mass (non-peaking).

The ratio, $R_{out/in}$ was plotted as a function of the MET cut applied for Drell-Yan and ZZ events. These were found to be the main contributions to the near-Z control region. The results shown in Figures 22 and 23 were found to be compatible with those previously found with CSA07 samples [5].

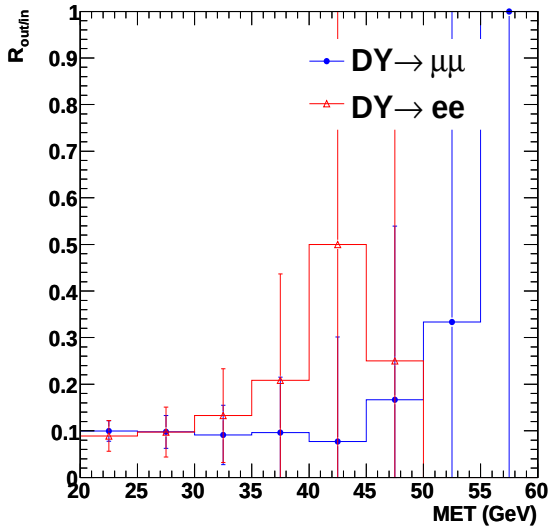


Figure 22: The ratio $R_{\mathrm{out/in}}$ as a function of the applied flat $\not\!\!E_T$ cut for Drell-Yan events.

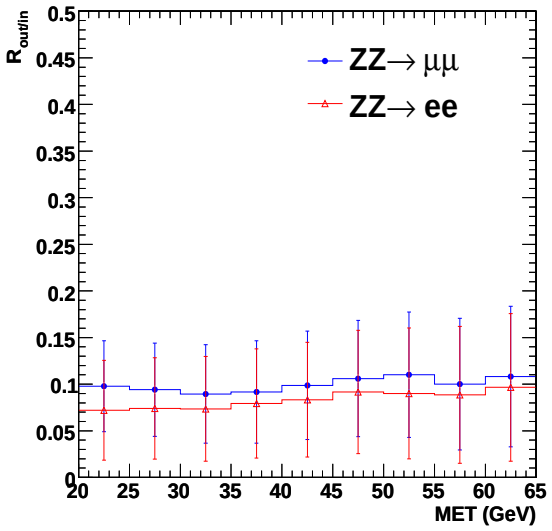


Figure 23: The ratio $R_{\mathrm{out/in}}$ as a function of the applied flat $\not\!\!E_T$ cut for ZZ events.

Because an insufficient number of high MET simulated Drell-Yan events was available, the ratio $R_{out/in}$ was taken with the MET requirement loosened to 20 GeV. The ratio was computed using simulated Drell-Yan events to be 0.09 ± 0.09 for di-electrons, and 0.10 ± 0.10 for di-muons with a 100% systematic error applied. This is consistent with previous studies and the ZZ events shown in Figure 23.

The ratio of the electron and muon reconstruction and selection efficiencies was computed to be 0.83 ± 0.01. To estimate the non-peaking contribution to the di-muon final state, this ratio is inverted to be 1.20 ± 0.01. Multiplying by a factor of 0.5 for combinatorics between the $e\mu$ and ll final states, the non-peaking background to the near-Z control region was estimated. The results shown in Table 16 demonstrate a good agreement between the estimates and the true event content from the Monte Carlo truth. Because the $e\mu$ final state contains approximately twice as many events as each ll final state, the estimate benefits from an improved uncertainty over other techniques that only use same flavor events.

Table 16: Results of the non-peaking background estimate in the near-Z control region for the Drell-Yan background estimate. Results and uncertainties computed for an integrated luminosity of 100 pb^{-1}.

State	k	$N^{in}_{e\mu}$	Est. $N^{in}_{ll\ non-peaking}$	True $N^{in}_{ll\ non-peaking}$
ee	0.42 ± 0.0041	3.3 ± 1.8	1.4 ± 0.8	1.5 ± 0.2
$\mu\mu$	0.60 ± 0.0066	3.3 ± 1.8	1.9 ± 1.1	1.8 ± 0.3

The estimated non-peaking backgrounds were subtracted from the near-Z control region and the remaining number of events multiplied by the expected ratio, $R_{out/in}$. These results are tabulated for the di-electron and di-muon final states in Table 17. Reasonable agreement is found, bearing in mind the small expected number of DY, ZZ and WZ events outside of the Z-veto region. It is concluded that these backgrounds can be estimated with sufficient accuracy and precision for this analysis.

Table 17: Detailed results of the Drell-Yan estimate including the peaking components of the ZZ and WZ backgrounds. Results and uncertainties computed for an integrated luminosity of 100 pb^{-1}.

State	$R_{out/in}$	N_{ll}^{in}	Est. $N_{peaking}^{in}$	Est. $N_{peaking}^{out}$	True $N_{peaking}^{out}$
ee	0.09 ± 0.09	4.2 ± 0.5	2.8 ± 0.9	0.2 ± 0.3	0.4 ± 0.2
$\mu\mu$	0.10 ± 0.10	6.1 ± 0.7	4.2 ± 1.3	0.4 ± 0.4	0.6 ± 0.3

G Event yields in other samples

Several alternative samples were used to cross check that we are not missing some important backgrounds.

Table 18 shows the expected yield for Wc and $Vq\bar{q}$ samples, where V refers to any vector boson and q represents heavy flavor quark (b or c). It is expected that these processes are simulated in the W+jets Madgraph sample that we used for the analysis. The numbers confirm that the contribution from W with heavy quarks is small.

Table 19 shows comparison between Madgraph and Pythia Drell-Yan samples. The results are statistically consistent.

Table 20 shows contribution of alternative Pythia $W\gamma$ and Madgraph $Z\gamma$ samples. The yield from $W\gamma$ is a bit higher than that of the Madgraph W+jets sample. It is known that for $W\gamma$ Pythia does not properly simulate the final state radiation as well as the photon spectrum. Since the results of $W\gamma$ extracted from Madgraph W+jets sample are statistically consistent with the Pythia prediction, we decided to use the Madgraph sample till we have $W\gamma$ properly generated with a proper generator like Bauer. The systematic error that we assigned on $W\gamma$ contribution covers the difference between Madgraph and Pythia predictions.

H Estimating the sensitivity to observe WW in 100/pb

In the following section we explain in some detail how we propose to calculate the significance of any WW signal we may observe in 100/pb, as well as how we calculated the expected significance of the signal excess for this note.

Let's start with a simplified situation where we have an expected yield total, N_Y, and an expected background N_{bkg}. The latter is estimated via some independent bkg data sample as:

$$N_{bkg} = N_{bkg\ sample} \times sf$$

Here, sf stands for scale factor, and $N_{bkg\ sample}$ is an event count in the independent background sample. The scale factor has an error that is generally truely systematic in nature, while the error on $N_{bkg\ sample}$ is statistics only. Both, N_Y and $N_{bkg\ sample}$ are event counts, and thus distributed according to a Poisson distribution.

If $N_{bkg\ sample}$ is sufficiently large, then we can approximate it as a Gaussian distributed variable, and

Table 18: Alternate samples of W events

Alternate samples of W events. W+jets is the reference Madgraph sample used in the analysis. All results are normalized to 100 pb^{-1} of data.

	W+jets	$Wc\bar{c}$	Vqq
all	4.7 ± 1.6	0.2 ± 0.1	0.2 ± 0.1
$\mu\mu$	0.0 ± 0.0	0.1 ± 0.1	0.1 ± 0.0
$e\mu$	4.7 ± 1.6	0.1 ± 0.1	0.1 ± 0.1
ee	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Table 19: Alternate samples of Drell-Yan events. The left three columns represent the Madgraph samples including the lower mass region. The rest is Pythia. The predictions are normalized to 100 pb^{-1} of data.

	$Z/\gamma^* \to ee$	$Z/\gamma^* \to \mu\mu$	$Z/\gamma^* \to \tau\tau$	DY($m > 20$) $\to \tau\tau$	DY($m > 20$) $\to \mu\mu$
all	0.0 ± 0.0	0.0 ± 0.0	0.7 ± 0.5	0.7 ± 0.4	0.9 ± 0.4
$\mu\mu$	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.3 ± 0.2
$e\mu$	0.0 ± 0.0	0.0 ± 0.0	0.7 ± 0.5	0.7 ± 0.4	0.5 ± 0.3
ee	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

propagate the error on the background estimate as follows:

$$\frac{\sigma(N_{bkg})}{N_{bkg}} = \sqrt{\frac{N_{bkg\ sample}}{N^2_{bkg\ sample}} + (\frac{\sigma_{sf}}{sf})^2}$$

As an aside, if $N_{bkg\ sample}$ is not large enough to be approximately Gaussian, then the above equation makes no sense.

In either case, the resulting N_{bkg} is a prediction for the mean of a Poisson distributed variable. I.e. it is the predicted event count for background in the signal region.

So let's deal with these two cases separately.

H.1 In the Gaussian limit after a measurement is made

Once we have made an actual measurement determining the event counts for $N_{bkg\ sample}$ and N_Y, it is straightforward how to assess the statistical significance of our result. To phrase it clearly, we need to ask:

What is the probability that a fluctuation of the background leads to N_Y or more yield observed?

In the Gaussian limit, this question is answered by the following procedure:

- Draw a random number from a Gaussian distribution with mean N_{bkg} and standard deviation $\sigma(N_{bkg})$. The resulting float is the expected mean of the Poisson distribution for the background in the signal region.
- Draw a random number from a Poisson distribution with mean being the random number from the previous step. The resulting integer is the actual bkg drawn for this pseudo experiment. Record whether or not this integer is larger than the actual count in the signal region.
- Repeat the procedure roughly 10^9 times, and determine the fraction of times bkg fluctuates above the yield seen.

H.2 In the Poisson limit after a measurement is made

Once we have made an actual measurement determining the event counts for $N_{bkg\ sample}$ and N_Y, we then need to run a toy Monte Carlo in order to convolute the Poisson distribution of $N_{bkg\ sample}$ with the scale factor, sf.

Table 20: Alternate Pythia samples of $V\gamma$ normalized to 100 pb^{-1} of data.

	$W\gamma$	$Z\gamma$
all	2.6 ± 0.5	0.3 ± 0.1
$\mu\mu$	0.0 ± 0.0	0.0 ± 0.0
$e\mu$	1.5 ± 0.4	0.3 ± 0.1
ee	1.1 ± 0.3	0.0 ± 0.0

- Draw a random number from a Poisson distribution with mean $N_{bkg\ sample}$.
- Draw a random number from a Gaussian distribution with mean sf and standard deviation σ_{sf}.
- Multiply the two numbers from above to arrive at a float for the expected background in the signal region for this pseudo experiment. The resulting float is the expected mean of the Poisson distribution for the background in the signal region.
- Draw a random number from a Poisson distribution with mean being the random number from the previous step. The resulting integer is the actual bkg drawn for this pseudo experiment. Record whether or not this integer is larger than the actual count in the signal region.
- Repeat the procedure roughly 10^9 times, and determine the fraction of times bkg fluctuates above the yield seen.

H.3 Predicting significance for a future measurement to be made

In this case, the question is entirely different. We can at best know what "significance" we expect "on average". However, in most cases, this is a meaningless question. The much more relevant question is the following:

What is the probability that my future experiment will provide me with a "5σ" observation?

The only way to do this is to devise a toy Monte Carlo. This toy Monte Carlo generates many experiments, evaluates for each experiment the significance as described above. We then plot the significances, and integrate the area above a given level, and divide by the total area, to get the desired probability.

In this case, N_Y is drawn from its Poisson distribution in addition to the previously described procedure.

For this note we did this latter procedure, using $N_{bkg} = 10.4 \pm 4.6$ as a Gaussian distributed variable and its standard deviation. We thus pretended to be in the Gaussian regime. For N_Y we furthermore used 48, and allowed that to fluctuate according to a Poisson distribution, as described above.

We found that we have around 45-50% chance to get 5σ observation for this analysis in real data.

I QCD multi jet background estimation

The data driven background estimation methods described in section 5.1 allow one to estimate not only the W+jets background contribution, but also multi jet QCD background when both leptons originate from hadronic activity in the event, either as true leptons from heavy quark decays or pure fakes.

In general it is hard to estimate the multi jet QCD contribution from Monte Carlo samples since the probability for a QCD event to fake two well isolated energetic leptons in an event with large missing transverse energy is very small and therefore it requires a lot of statistics to simulate such events. The generic QCD Monte Carlo samples available in the Summer/Fall 2008 production represent just a small fraction of the analysis target integrated luminosity of 100 pb^{-1}. Even the QCD enriched samples do not have enough statistics to estimate the multi jet QCD contribution from Monte Carlo.

Figure 24 shows an application of the isolation sideband background method on the following samples:

- `WW_Summer08_IDEAL_V9_v1`
- `InclusiveMuPt15`
- `InclusiveMu5Pt50`
- `QCD_EMenriched_Pt20to30`
- `QCD_EMenriched_Pt30to80`
- `QCD_EMenriched_Pt80to170`
- `QCD_BCtoE_Pt20to30`

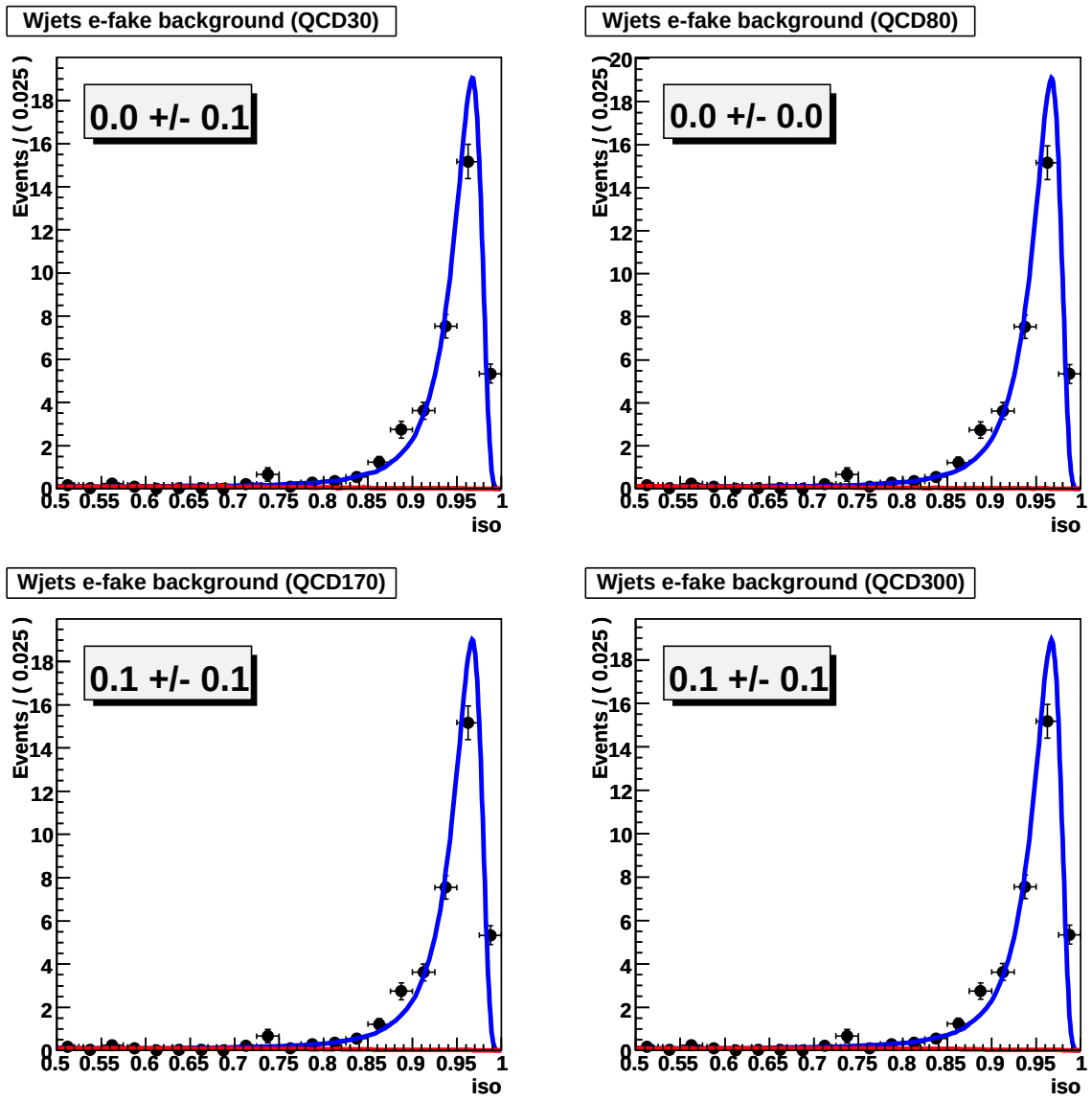


Figure 24: Unbinned maximum likelihood fits on a set of Monte Carlo samples including the QCD enriched samples, Drell-Yan sample (used to determine isolated lepton probability distribution function) and WW signal events. The background pdf was extracted from independent generic QCD samples. The red curve represents the background distribution. The QCD multi jet background estimates due to fake electrons in the region $[0.92, 1.00]$ are shown in the boxes. The event yield is normalized to 100 pb^{-1} at 10 TeV.

- `QCD_BCtoE_Pt30to80`
- `QCD_BCtoE_Pt80to170`
- `Zee_M20_Summer08_IDEAL_V9_reco-v3`
- `Zmumu_M20_Summer08_IDEAL_V9_reco-v2`

The probability distribution functions used to fit the background contribution are taken from independent generic QCD samples. The plots clearly show very little if any at all QCD background. The background contribution is expected to rise as we loosen the isolation requirements (lower values of relative isolation) and we see no sign of this behavior. This gives us confidence that in the signal region, where both leptons have tight isolation requirements, the QCD contribution is negligible. It is worthwhile also to note that

even though some of the QCD enriched samples have isolation requirements in the sample definitions, we still can use these samples for the test since we require two isolated leptons in the event, so the isolation requirements can be relaxed for one of the leptons.

References

[1] **CMS AN-2008/015**, *"Expectations for $t\bar{t} \rightarrow$ dileptons in the early phase of CMS"*
PAS TOP-08-01, *"Di-lepton ttbar cross section with 10pb^{-1}"*

[2] http://www.ge.infn.it/~tosi/cms/topMC.html, *"Reference Page for Top-Quark Analyses: Samples and Cross-Sections"*

[3] https://twiki.cern.ch/twiki/bin/view/CMS/TSG_18_II_09_1E31, *"HLT paths for the 1E31 core menu (v0.3, 18-FEB-09)"*

[4] **CMS AN-2008/098**, *"Muon Identification in CMS."*

[5] **CMS AN-2009/023**, *"A Method to Measure the Contribution of DY$\rightarrow l^+l^-$ to a di-lepton + MET Selection"*

[6] **CMS AN-2009/041**, *"Data-driven methods to estimate the electron and muon fake contribution to lepton analyses"*,

[7] http://indico.cern.ch/contributionDisplay.py?contribId=5&confId=49595, P. Kalavase, *"Electron Backgrounds from Photon Conversions"*

[8] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideCutBasedElectronID, *"Cut Based Electron ID"*

[9] http://indico.cern.ch/getFile.py/access?contribId=3&resId=0&materialId=slides&confId=22460, M. Sani, *"Cut Based Electron ID"*

[10] V.M.Abazov *et al.* (D0 Collaboration), *"Limits on anomalous trilinear gauge couplings from WW$\rightarrow$ e^+e^-, WW$\rightarrow$ $e^{\pm}\mu^{\mp}$, and WW$\rightarrow$ $\mu^+\mu^-$ events from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV"*, Phys. Rev. D **74**, 057101 (2006)

[11] **CMS AN-2008/051**, *"Measurement of the ZW cross section at the LHC"*
PAS EWK-08-003, *"Observation of WZ production"*

[12] S. C. Hsu, "A Study of The Standard Model Higgs, WW and ZZ Production in Dilepton Plus Missing Transverse Energy Final State at CDF Run II," **FERMILAB-THESIS-2008-61**, Chapter 7.2

[13] **CMS AN-2009/031**, *"Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS"*

[14] **EWK-07-002**, *"Measurement of the W and Z cross section with muons"*,
EWK-08-005, *"Measurement of the W and Z cross section with electrons"*

[15] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIsolation, *"Muon Isolation"*

[16] http://indico.cern.ch/contributionDisplay.py?contribId=4&confId=43146, M. LeBourgeois, *"Electron Isolation Status and Recommendations from e/gamma POG"*

[17] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideEgammaIsolationIn2_1And2_2AndPlansFor3_0, *"Standard e/gamma isolation definitions"*

[18] **CMS AN-2009/022**, *"Correcting Missing Transverse Energy Using Tracks"*

[19] D. Bourilkov, R. C. Group and M. R. Whalley, *"LHAPDF: PDF use from the Tevatron to the LHC"*, arXiv:hep-ph/0605240.

[20] http://indico.cern.ch/contributionDisplay.py?contribId=0&confId=49643, M. Sani, *"Retuning electron ID"*