

---

# CMS Physics Analysis Summary

---

Contact: cms-pag-conveners-ewk@cern.ch

2009/06/29

Prospects for the first measurement of the  $WW$  production cross-section in  $pp$  collisions at  $\sqrt{s} = 10$  TeV center of mass energy with the CMS detector

The CMS Collaboration

## Abstract

This note describes an analysis strategy for measuring the  $WW$  production cross-section in  $100 \text{ pb}^{-1}$  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. We present the event yields and statistical and systematic uncertainties that we expect to achieve with this analysis strategy. We estimate a total uncertainty on the cross-section of 30% or better.



## 1 Introduction

Pair production of oppositely charged  $W$  bosons in the Standard Model proceeds via  $s$ - and  $t$ -channel diagrams at lowest order. The  $s$ -channel production probes the  $WWZ$  and  $WW\gamma$  triple gauge couplings. The process is sensitive to physics beyond the Standard Model via anomalous vector-boson couplings. The effect of the anomalous coupling on  $WW$  production is expected to grow with the larger invariant mass of the di-boson system available at the LHC compared with the LEP and Tevatron experiments [1]. In addition, the  $WW$  process represents one of the dominant irreducible backgrounds for the Standard Model Higgs boson searches [2, 3]. It is essential to establish its properties and cross-section to control it.

This note describes an analysis strategy for measuring the  $WW$  production cross-section in  $100 \text{ pb}^{-1}$  of  $pp$  collision data at  $\sqrt{s} = 10 \text{ TeV}$  using the CMS detector [4]. The strategy is to find two energetic isolated leptons (electron or muon) of opposite charge with large missing transverse energy and low hadronic activity established by a jet veto. Counting the number of events in the signal selection region, we estimate the  $WW$  yield by subtracting the expected contributions from various Standard Model background processes. A number of data-driven methods are developed to evaluate and cross-check the backgrounds.

## 2 Simulation

The Standard Model processes expected to contribute to the measurement of the  $WW$  cross-section include:  $WW$ , Drell-Yan ( $Z/\gamma^* \rightarrow \ell\ell$ ),  $t\bar{t}$ , single top ( $tW$  final state),  $W$ +jets, and the  $ZZ$ ,  $WZ$  and  $W\gamma$  diboson processes. The  $W\gamma$  process is modeled via ISR and FSR as implemented in the  $W$ +jets sample. The  $t\bar{t}$  and  $W$ +jets processes are generated using the Madgraph [5] Monte Carlo generator, while the remaining processes are generated using Pythia [6]. The Drell-Yan processes are generated with the invariant mass of the dilepton pair greater than  $20 \text{ GeV}/c^2$ . All processes are simulated using the full CMS detector simulation. The cross-section for each Monte Carlo sample is scaled to its next-to-leading order value. Table 1 gives details about the samples used for this study. The effect of multiple proton-proton interactions is not included.

Table 1: Monte Carlo data samples used for the analysis. The cross section numbers are taken from a reference set of next-to-leading order cross-section estimations for  $pp$  collisions at  $\sqrt{s} = 10 \text{ TeV}$ .

Sample name	Generator	cross-section, pb	Effective luminosity, $\text{pb}^{-1}$
$t\bar{t}$	Madgraph	414	2 500
$W$ +jets	Madgraph	45 000	190
$Z/\gamma^* \rightarrow ee$	Pythia	2 220	460
$Z/\gamma^* \rightarrow \mu\mu$	Pythia	2 220	580
$Z/\gamma^* \rightarrow \tau\tau$	Pythia	2 220	450
$WW$	Pythia	74	2 800
$ZZ$	Pythia	10.5	20 000
$WZ$	Pythia	32	6 700
$tW$	Pythia	29	4 750

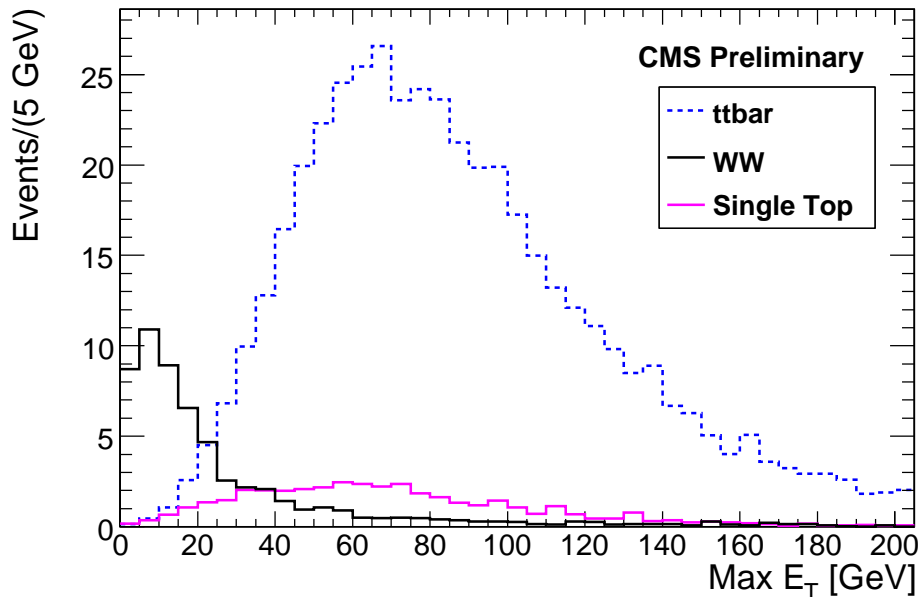


Figure 1: Distribution of the transverse energy of the jet with the largest transverse energy for the  $WW$ ,  $t\bar{t}$  and single top processes. All samples are normalized to an integrated luminosity of  $100 \text{ pb}^{-1}$ .

### 3 Event Selection

The  $WW$  events are reconstructed in leptonic  $W$  boson decay channels, leading to two leptons (muons and electrons only) in the final state and large missing transverse energy due to the corresponding neutrinos. We select events that pass a single-lepton trigger and have two oppositely charged isolated leptons with  $p_T > 20 \text{ GeV}/c$ . We define an isolation variable  $\text{iso} = p_T / (p_T + S)$ , where  $S$  is the sum of the transverse momenta of tracks (excluding the lepton track) and the transverse energy of the electromagnetic and hadronic calorimeters, within an isolation cone of size  $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.3$ . For electron candidates we remove from the sum the electromagnetic energy associated with the electron track, taking into account possible radiation effects. The isolation requirement is  $\text{iso} > 0.92$ .

The  $WW$  cross-section is several orders of magnitude lower than that of the major background processes, such as  $W$ +jets,  $t\bar{t}$  and Drell-Yan. The  $W$ +jets background is dominated by fake leptons originating from jets and it is effectively reduced by the lepton identification and isolation requirements. Details on the lepton identification and isolation can be found in the CMS physics technical design report [4].

We use a jet veto to reduce top backgrounds (Fig. 1). We reject events with at least one jet with pseudo-rapidity  $|\eta| < 3.0$  and transverse energy larger than  $20 \text{ GeV}$ . Jets are reconstructed from combinations of energy depositions in the electromagnetic and hadronic calorimeters, and corrected for the average energy response of tracks in the calorimeters [7].

We reduce the Drell-Yan background by removing events with a dilepton invariant mass consistent with a  $Z$  mass ( $76 < m_{\ell\ell} < 106 \text{ GeV}/c^2$ ), and imposing tight missing transverse energy requirements, which depend on the dilepton final state. The missing transverse energy ( $\cancel{E}_T$ ) is calculated as the opposite vector sum of all calorimeter energy depositions corrected for muons and for the average track response in the calorimeters [8]. The track response correction allows to suppress the Drell-Yan background by a factor of 2.4 in the  $ee/\mu\mu$  final states, which do not have a natural source of the missing energy. Figure 2 shows a comparison between the standard

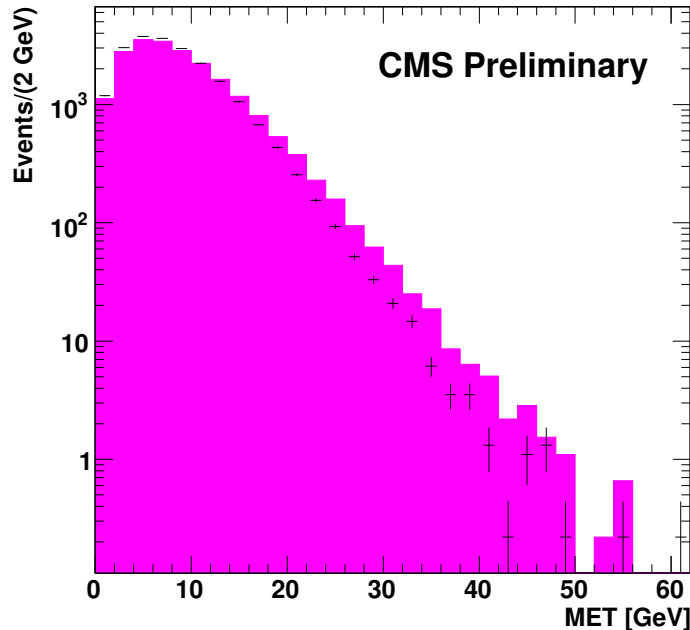


Figure 2: The missing transverse energy distributions for the Drell-Yan events using the calorimeter based  $\cancel{E}_T$  with (points) and without (histogram) the average track response corrections. Both algorithms correct for the muon response in the calorimeter.

calorimeter based only  $\cancel{E}_T$  and the track corrected one used in this analysis.

For the  $ee$  and  $\mu\mu$  final states we require  $\cancel{E}_T > 45 \text{ GeV}$ , and if  $\cancel{E}_T/P_T^{\ell\ell} < 0.6$ , where  $P_T^{\ell\ell}$  is the transverse momentum of the lepton pair, we also require the angle between the opposite direction of  $\cancel{E}_T$  and the dilepton transverse direction to be less than  $0.25$ . The latter requirement removes Drell-Yan events with large fake  $\cancel{E}_T$  due to poor response of the calorimeter to the recoiling hadronic activity in the event. Figure 3 shows the two dimensional distributions of  $\cancel{E}_T/P_T^{\ell\ell}$  and the acoplanarity angle for Drell-Yan and  $WW$  events.

For the  $e\mu$  final state we use a projected  $\cancel{E}_T$ , which is defined as the transverse component of  $\cancel{E}_T$  with respect to the closest lepton, if the  $\Delta\phi$  between the lepton and  $\cancel{E}_T$  is smaller than  $90^\circ$ . We require the projected  $\cancel{E}_T$  to be at least  $20 \text{ GeV}$ . This significantly reduces the  $Z/\gamma^* \rightarrow \tau\tau$  background, where both tau leptons decay leptonically. This selection requirement is also applied to the  $ee$  and  $\mu\mu$  final states.

In order to suppress poorly reconstructed Drell-Yan events as well as  $WZ$  and the remaining  $t\bar{t}$  background events, we veto events that have an extra identified muon in the event.

## 4 Expected event yield

Table 2 shows the expected contributions of the various Standard Model processes listed in Table 1 after all selection cuts are applied. Figure 4 and Figure 5 show the dilepton mass and the missing transverse energy distributions, respectively, after the final event selection. The dominant remaining background processes are top and  $W$ +jets, with  $5.21 \pm 0.43$  and  $3.11 \pm 1.27$  events, respectively. After the final selection is applied, there remain  $35.0 \pm 1.1$  signal  $WW$  events, and  $12.9 \pm 1.8$  total background events. Note that this is not the estimated number of background events in this analysis, but rather the number of generated background events that

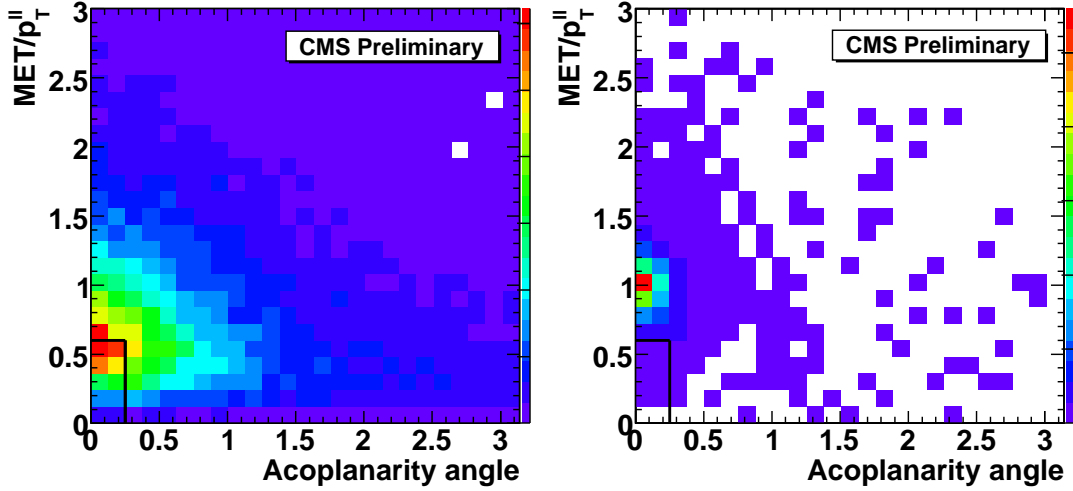


Figure 3: Distributions of  $E_T/P_T^{\ell\ell}$  as a function of the acoplanarity angle between the dilepton pair azimuthal angle and the opposite of the  $E_T$  azimuthal angle. The left plot shows the distribution for the Drell-Yan events and the right one shows the  $WW$  signal events. The black box outlines the area that is removed in the final event selection.

Table 2: Expected event yields for dominant processes at 10 TeV with  $100 \text{ pb}^{-1}$  of integrated luminosity. The errors represent the statistical uncertainties due to the limited number of events in the corresponding Monte Carlo samples. Note: the  $W\gamma$  event yield is estimated from the full Madgraph  $W$ +jets sample by looking at the generator-level information.

	$ee$	$\mu\mu$	$e\mu$	all
$t\bar{t}$	$0.44 \pm 0.14$	$0.75 \pm 0.18$	$2.08 \pm 0.30$	$3.27 \pm 0.38$
$tW$	$0.27 \pm 0.08$	$0.37 \pm 0.09$	$1.29 \pm 0.16$	$1.94 \pm 0.20$
$W$ +jets (excluding $W\gamma$ )	$0.00 \pm 0.00$	$0.00 \pm 0.00$	$3.11 \pm 1.27$	$3.11 \pm 1.27$
$W\gamma$	$0.00 \pm 0.00$	$0.00 \pm 0.00$	$1.55 \pm 0.89$	$1.55 \pm 0.89$
$WZ$	$0.09 \pm 0.04$	$0.20 \pm 0.05$	$0.68 \pm 0.10$	$0.97 \pm 0.12$
$Z/\gamma^* \rightarrow ee$	$0.22 \pm 0.22$	$0.00 \pm 0.00$	$0.00 \pm 0.00$	$0.22 \pm 0.22$
$Z/\gamma^* \rightarrow \mu\mu$	$0.00 \pm 0.00$	$0.35 \pm 0.25$	$0.52 \pm 0.30$	$0.87 \pm 0.39$
$Z/\gamma^* \rightarrow \tau\tau$	$0.00 \pm 0.00$	$0.00 \pm 0.00$	$0.67 \pm 0.39$	$0.67 \pm 0.39$
$ZZ$	$0.12 \pm 0.02$	$0.15 \pm 0.03$	$0.05 \pm 0.02$	$0.32 \pm 0.04$
Background total	$1.14 \pm 0.27$	$1.82 \pm 0.32$	$9.96 \pm 1.67$	$12.91 \pm 1.72$
$WW$ signal	$3.88 \pm 0.38$	$6.75 \pm 0.50$	$24.40 \pm 0.94$	$35.04 \pm 1.13$

pass all cuts.

## 5 Background estimation methods

In order to measure correctly the  $WW$  cross-section, contributions from the dominant remaining backgrounds need to be reliably estimated. For the data driven methods described below we use a single sample that is a mixture of all Monte Carlo events with proper weights, including signal  $WW$  events.

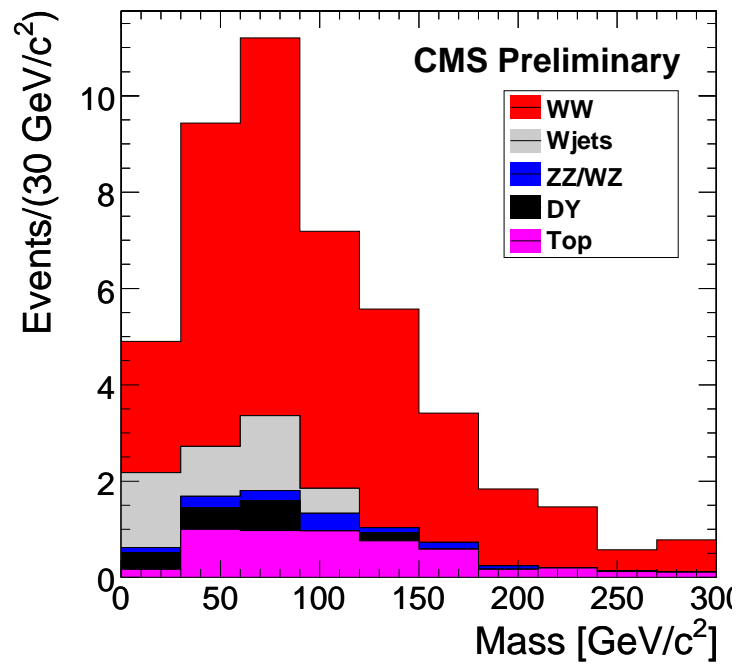


Figure 4: Dilepton mass distribution for the final event selection scaled to  $100 \text{ pb}^{-1}$  of integrated luminosity

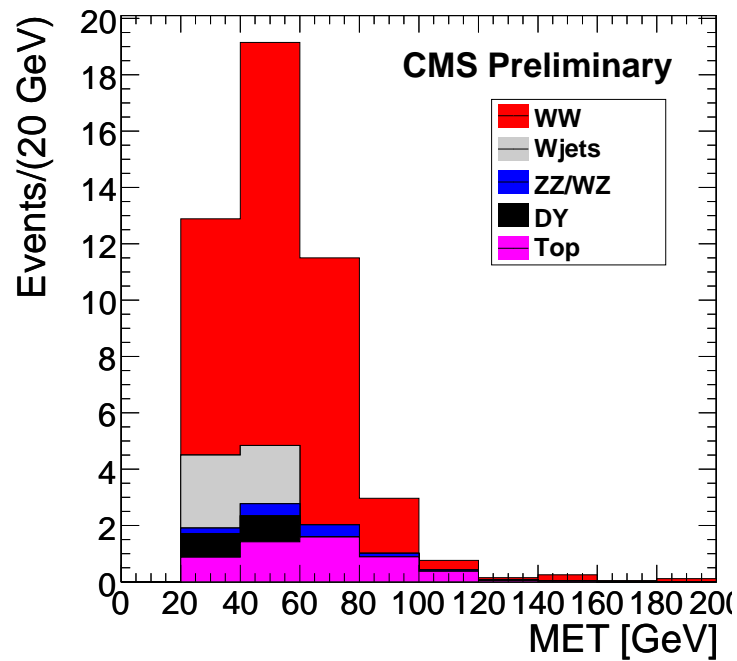


Figure 5: Missing transverse energy distribution for the final event selection scaled to  $100 \text{ pb}^{-1}$  of integrated luminosity

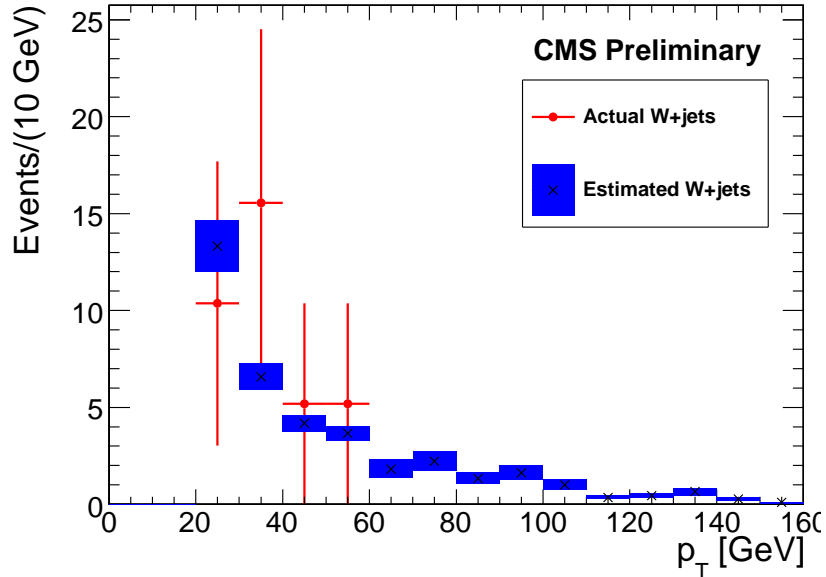


Figure 6: Illustration of the fake-rate method for the  $W$ +jets background. The plot shows the observed and predicted distributions of the fake electron  $p_T$  in  $W$ +jets Monte Carlo dilepton events in the  $e\mu$  final state. The fake rates for each type of lepton were extracted from independent QCD samples. The distributions are scaled to  $100 \text{ pb}^{-1}$ .

### 5.1 $W$ +jets background

The  $W$ +jets background process has one real isolated lepton, a natural source of missing transverse energy from the neutrino, and one fake lepton. We have developed two methods for estimation of this background from data. The first method is called the “fake-rate” method. The selection criteria are relaxed for one of the leptons (a fakeable object) so that the background dominates. The fakeable object has looser lepton identification, isolation and impact parameter requirements. Then using an independent QCD sample, we estimate the probability of the fakeable object to pass the signal lepton selection. The probability is parameterized as a function of lepton momentum and pseudo-rapidity. Figure 6 shows the actual distribution of the  $p_T$  of the fake electrons in  $W$ +jets events compared to estimations derived using the fake-rate method just described.

The second method for estimating the  $W$ +jets background uses the sideband of the isolation distribution to evaluate the remaining background in the signal region. The shape of the isolation distribution is parameterized using QCD samples, and the variation in predictions is used as an estimate of the systematic uncertainty. Figure 7 illustrates the method. Combining the predictions of the two methods, we take the average as an estimate of the  $W$ +jets background and the difference as an estimate of the error.

Both methods mentioned above can estimate backgrounds due to fakes originating from the hadronic activity in the event for all final states, i.e.  $ee$ ,  $\mu\mu$  and  $e\mu$ . They do not cover the  $W\gamma$  processes, where a photon from the final state radiation is radiated at a large angle with respect to the lepton and converts to an electron-positron pair with most of the energy going to either the electron or the positron. Such events produce isolated fake electrons and need to be estimated differently. We rely on Monte Carlo simulations to evaluate this background and assign a 100% systematic uncertainty on the prediction.



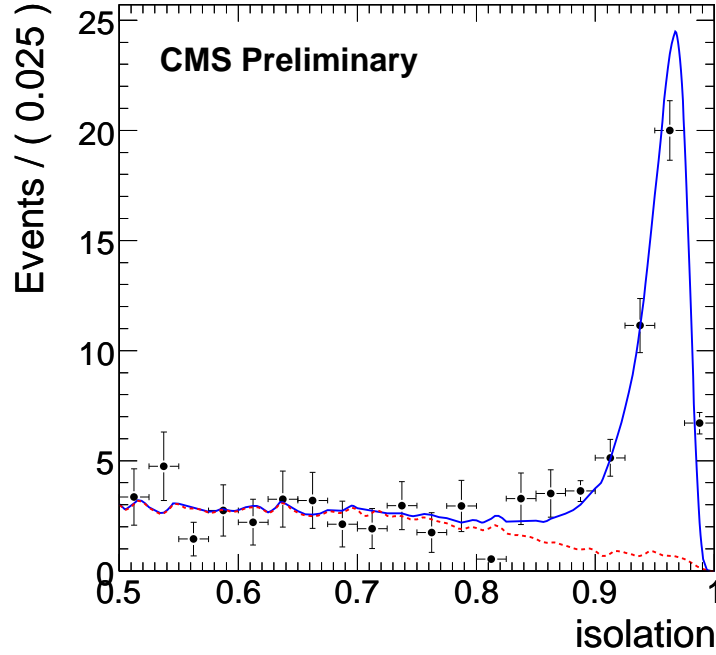


Figure 7: Illustration of the isolation sideband  $W$ +jets background estimation method. The plot shows an unbinned maximum likelihood fit on the full sample of events. The signal probability distribution function (pdf) was extracted from  $Z \rightarrow ee$  events. The background pdf was extracted from an independent QCD sample. The dashed curve represents the background distribution. The distributions are scaled to  $100 \text{ pb}^{-1}$ .

## 5.2 Top background

In order to estimate the top background we count the number of events with additional soft muons coming from  $b$  jets and then estimate the total top background by dividing this number by the probability to have a muon originating from jets in top decays. Most of the remaining top background events have  $b$ -jets within the acceptance of the muon system, but fail the jet veto minimum transverse energy requirement.

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 ( $p_T < 20 \text{ GeV}$ ).

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 various systematic effects. Overall  $t\bar{t}$  tagging efficiency is shown on Figure 8.

## 5.3 Other backgrounds

Monte Carlo simulations cannot predict reliably the fake missing transverse energy distribution for Drell-Yan events. Therefore we have developed a data-driven method to estimate the off-peak  $Z/\gamma^*$  contribution of  $Z/\gamma^* \rightarrow ee$ ,  $Z/\gamma^* \rightarrow \mu\mu$ ,  $WZ$  and  $ZZ$  processes, by normalizing the expected dilepton mass distribution extracted from Monte Carlo simulations to the event yield

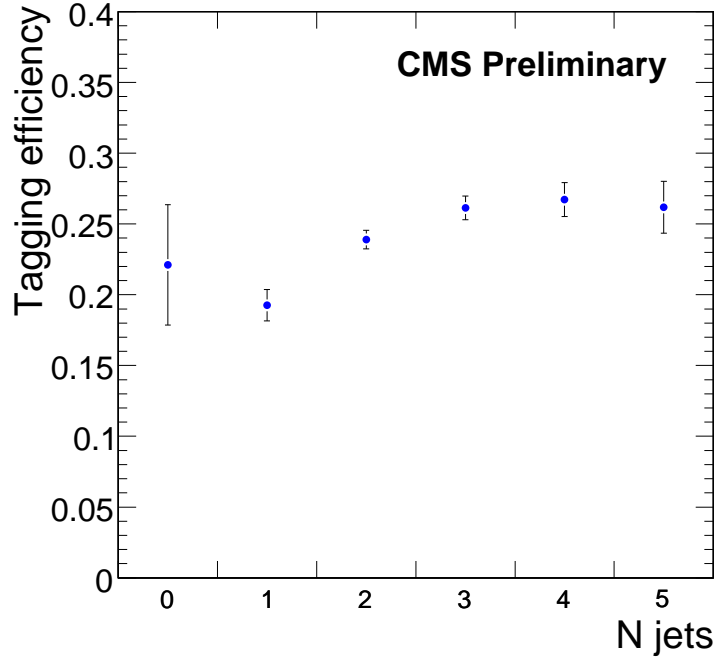


Figure 8: The  $t\bar{t}$  tagging efficiency as a function of the number of jets counting the number of soft muons in the jet.

Table 3: Expected event yields for dominant processes at 10 TeV with  $100 \text{ pb}^{-1}$  of integrated luminosity extracted from the combined Monte Carlo samples, which are treated as the “data” sample. Multi-boson contributions are extracted from exclusive Monte Carlo samples. The errors represent a combination of the statistical and systematic uncertainties.

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

in the Z veto mass region. In this way we can control the Drell-Yan contribution even if the  $\cancel{E}_T$  performance in data is worse than in Monte Carlo.

All other backgrounds are estimated from Monte Carlo samples. Table 3 summarizes the Standard Model backgrounds as we would derive them from data. The residual multi-jet QCD background was found to be negligible. The total background including systematic uncertainties is  $10.4 \pm 4.6$  events, which is consistent with the true Monte Carlo prediction of  $12.9 \pm 1.8$  events.

## 6 Results

The expected signal  $WW$  contribution can be estimated from the total event yield by subtracting the background yield. The result is  $37.5 \pm 8.3$   $WW$  events, where the uncertainty is the combined uncertainty on the total event count (6.9) and the background estimation (4.6).

In order to estimate the  $WW$  cross-section, we need to take into account the uncertainty on the signal reconstruction efficiency, which is 14% (Table 4), and the luminosity systematic uncer-

Table 4: List of systematic uncertainties.

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%

tainty which is expected to be on the order of 10%. Adding these uncertainties in quadrature with the uncertainty on the signal yield after the background subtraction (22%), we estimate a total uncertainty on the cross-section measurement to be on the order of 28%.

## 7 Conclusion

We have presented expectations of measuring the  $W$  boson pair production cross-section in the dilepton final state using  $100 \text{ pb}^{-1}$  of CMS data at 10 TeV center of mass energy. We estimate a total uncertainty on the cross-section to be about 30% or better.

We have developed a number of data-driven methods to control all the major background contributions. The largest contribution to the systematic uncertainty is found to come from the top background, which we estimate from “data” using a soft-muon tagging technique. The precision of this method to estimate the top background is statistics limited.

## References

- [1] L. J. Dixon, Z. Kunszt, and A. Signer, “Vector boson pair production in hadronic collisions at order  $\alpha_s$ : Lepton correlations and anomalous couplings,” *Phys. Rev.* **D60** (1999) 114037, arXiv:hep-ph/9907305. doi:10.1103/PhysRevD.60.114037.
- [2] CMS Collaboration, “Search Strategy for a Standard Model Higgs Boson Decaying to Two  $W$  Bosons in the Fully Leptonic Final State,”. CMS PAS HIG-09-006.
- [3] CDF and D0 Collaboration, G. Hesketh, “Searches for the standard model Higgs boson at the Tevatron,” *Nucl. Phys. Proc. Suppl.* **177-178** (2008) 219–223. doi:10.1016/j.nuclphysbps.2007.11.112.
- [4] CMS Collaboration, G. L. Bayatian et al., “CMS physics: Technical design report, Volume I,”. CERN-LHCC-2006-001.
- [5] F. Maltoni and T. Stelzer, “MadEvent: Automatic event generation with MadGraph,” *JHEP* **02** (2003) 027, arXiv:hep-ph/0208156.
- [6] T. Sjostrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 physics and manual,” *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.
- [7] CMS Collaboration, “Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS,”. CMS PAS JME-09-002.
- [8] CMS Collaboration, “Performance of Track Corrected Missing ET in CMS,”. CMS PAS JME-09-010.