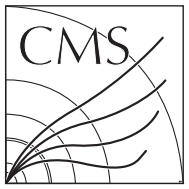
Available on CMS information server

CMS AN 2012/XXX



The Compact Muon Solenoid Experiment

Analysis Note

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

October 2, 2012

Search for Direct Stop Quark Pair Production in the Single Lepton Channel at 8 TeV

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

Fermilab National Accelerator Laboratory, Batavia, USA

D. Barge, C. Campagnari, A. George, F. Golf, J. Gran, D. Kovalskyi, V. Krutelyov

University of California, Santa Barbara, Santa Barbara, USA

W. Andrews, G. Cerati, D. Evans, R. Kelley, I. MacNeill, S. Padhi, Y. Tu, F. Würthwein, V. Welke, A. Yagil, J. Yoo

University of California, San Diego, San Diego, USA

Abstract

This note describes a search for direct stop quark pair production in the single lepton channel using 9.7 fb^{-1} of pp collision data at $\sqrt{s} = 8$ TeV taken with the CMS detector in 2012. A search for an excess of events over the Standard Model prediction is performed in a sample with a single isolated electron or muon, several jets, missing transverse energy and large transverse mass.

Contents

1 Introduction

This note presents a search for the production of supersymmetric (SUSY) stop quark pairs in events with a single isolated lepton, several jets, missing transverse energy, and large transverse mass. We use the full 2011 data sample, corresponding to an integrated luminosity of 9.7 fb^{-1}. This search is of theoretical interest because of the critical role played by the stop quark in solving the hierarchy problem in SUSY models. This solution requires that the stop quark be light, less than a few hundred GeV and hence within reach for direct pair production. We focus on two decay modes $\tilde{t} \to t\chi_1^0$ and $\tilde{t} \to b\chi_1^+$ which are expected to have large branching fractions if they are kinematically accessible, leading to:

- $pp \to \tilde{t}\bar{\tilde{t}} \to t\bar{t}\chi_1^0\chi_1^0$, and
- $pp \to \tilde{t}\bar{\tilde{t}} \to b\bar{b}\chi_1^+\chi_1^- \to b\bar{b}W^+W^-\chi_1^0\chi_1^0$.

Both of these signatures contain high transverse momentum (p_{T}) jets including two b-jets, and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) due to the invisible χ_1^0 lightest SUSY particles (LSP's). In addition, the presence of two W bosons leads to a large branching fraction to the single lepton final state. Hence we require the presence of exactly one isolated, high p_{T} electron or muon, which provides significant suppression of several backgrounds that are present in the all-hadronic channel. The largest backgrounds for this signature are semi-leptonic $\mathrm{t\bar{t}}$ and W+jets. These backgrounds contain a single leptonically-decaying W boson, and the transverse mass (M_{T}) of the lepton-neutrino system has a kinematic endpoint requiring $M_{\mathrm{T}} < M_W$. For signal stop quark events, the presence of additional LSP's in the final states allows the M_{T} to exceed M_W. Hence we search for an excess of events with large M_{T}. The dominant background in this kinematic region is dilepton $\mathrm{t\bar{t}}$ where one of the leptons is not identified, since the presence of two neutrinos from leptonically-decaying W bosons allows the M_{T} to exceed M_W. Backgrounds are estimated from Monte Carlo (MC) simulation, with careful validation and determination of scale factors and corresponding uncertainties based on data control samples.

The expected stop quark pair production cross section (see Fig. 1) varies between O(10) pb for $m_{\tilde{t}} =$ 200 GeV and O(0.01) pb for $m_{\tilde{t}} = 500$ GeV. The critical challenge of this analysis is due to the fact that for light stop quarks ($m_{\tilde{t}} \approx m_t$), the production cross section is large but the kinematic distributions, in particular M_{T}, are very similar to SM $\mathrm{t\bar{t}}$ production. In this regime it becomes very difficult to distinguish the signal and background. For large stop quark mass the kinematic distributions differ from those in SM $\mathrm{t\bar{t}}$ production, but the cross section decreases rapidly, reducing the signal-to-background ratio.

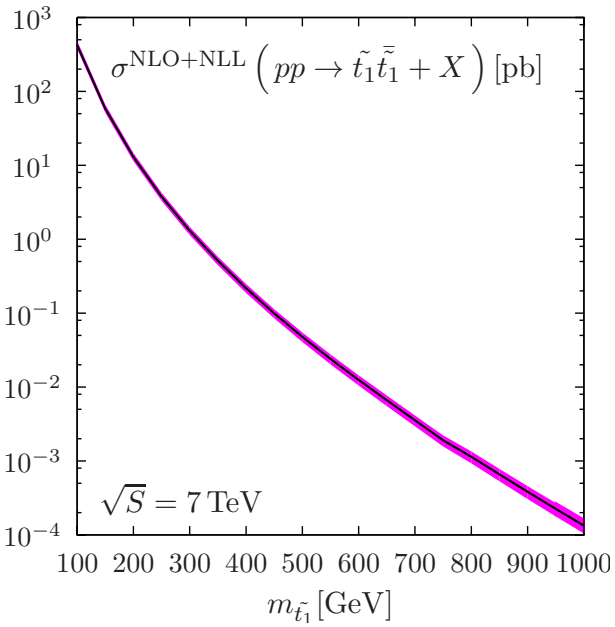


Figure 1: The stop quark pair production cross section in pb, as a function of the stop quark mass.

2 Overview and Analysis Strategy

[REVISE — MAKE SURE EVERYTHING IS CORRECT]

We are searching for a $t\bar{t}\chi^0\chi^0$ or $WbW\bar{b}\chi^0\chi^0$ final state (after top decay in the first mode, the final states are actually the same). So to first order this is "$t\bar{t}$+ extra $E_{\rm T}^{\rm miss}$".

We work in the ℓ+ jets final state, where the main background is $t\bar{t}$. We look for $E_{\rm T}^{\rm miss}$ inconsistent with $W \to \ell\nu$. We do this by concentrating on the $\ell\nu$ transverse mass (M_T), since except for resolution effects, $M_T < M_W$ for $W \to \ell\nu$. Thus, the initial analysis is simply a counting experiment in the tail of the M_T distribution.

The event selection is one-and-only-one high $p_{\rm T}$ isolated lepton, four or more jets, and some moderate $E_{\rm T}^{\rm miss}$ cut. At least one of the jets has to be btagged to reduce W+ jets. The event sample is then dominated by $t\bar{t}$, but there are also contributions from W+ jets, single top, dibosons, etc.

In order to be sensitive to $\widetilde{t\bar{t}}$ production, the background in the M_T tail has to be controlled at the level of 10% or better. So this is (almost) a precision measurement.

The $t\bar{t}$ events in the M_T tail can be broken up into two categories: (i) $t\bar{t} \to \ell$+ jets and (ii) $t\bar{t} \to \ell^+\ell^-$ where one of the two leptons is not found by the second-lepton-veto (here the second lepton can be a hadronically decaying τ). For a reasonable M_T cut, say $M_T > 150$ GeV, the dilepton background is of order 80% of the total. This is because in dileptons there are two neutrinos from W decay, thus M_T is not bounded by M_W. This is a very important point: while it is true that we are looking in the tail of M_T, the bulk of the background events end up there not because of some exotic $E_{\rm T}^{\rm miss}$ reconstruction failure, but because of well understood physics processes. This means that the background estimate can be taken from Monte Carlo (MC), after carefully accounting for possible data/MC differences. Sophisticated fully "data driven" techniques are not really needed.

Another important point is that in order to minimize systematic uncertainties, the MC background predictions are always normalized to the bulk of the $t\bar{t}$ data, ie, events passing all of the requirements but with $M_T \approx 80$ GeV. This removes uncertainties due to $\sigma(t\bar{t})$, lepton ID, trigger efficiency, luminosity, etc.

The ℓ+ jets background, which is dominated by $t\bar{t} \to \ell$+ jets, but also includes some W+ jets as well as single top, is estimated as follows:

1. We select a control sample of events passing all cuts, but anti-btagged, i.e. b-vetoed. This sample is now dominated by W+ jets. The sample is used to understand the M_T tail in ℓ+ jets processes.
2. In MC we measure the ratio of the number of ℓ+ jets events in the M_T tail to the number of events with $M_T \approx 80$ GeV. This ratio turns out to be pretty much the same for all sources of ℓ+ jets.
3. In data we measure the same ratio but after correcting for the $t\bar{t} \to$ dilepton contribution, as well as dibosons etc. The dilepton contribution is taken from MC after the correction described below.
4. We compare the two ratios, as well as the shapes of the data and MC M_T distributions. If they do not agree, we try to figure out why and fix it. If they agree well enough, we define a data-to-MC scale factor (SF) which is the ratio of the ratios defined in step 2 and 3, keeping track of the uncertainty.
5. We next perform the full selection in $t\bar{t} \to \ell$+ jets MC, and measure this ratio again (which should be the same as that in step 2).
6. We perform the full selection in data. We count the number of events with $M_T \approx 80$ GeV, after subtracting off the dilepton contribution, and multiply this count by the ratio from step 5 times the data/MC scale factor from step 4. The result is the prediction for the ℓ+ jets BG in the M_T tail.

Steps 1-4 above are all measurements on the b-vetoed samples in data and/or MC. Steps 5 and 6 are performed on the b-tagged sample.

To suppress dilepton backgrounds, we veto events with an isolated track of $p_{\rm T}>$ 10 GeV. Being the common feature for electron, muon, and one-prong tau decays, this veto is highly efficient for rejecting $t\bar{t}$ to dilepton events. The remaining dilepton background can be classified into the following categories:

- lepton is out of acceptance ($|\eta| > 2.50$)
- lepton has $p_{\rm T} < 10$ GeV, and is inside the acceptance
- lepton has $p_{\rm T} > 10$ GeV, is inside the acceptance, but survives the additional isolated track veto

The last category includes 3-prong tau decays as well as electrons and muons from W decay that fail the isolation requirement. Monte Carlo studies indicate that these three components populate the M_T tail in the proportions of roughly 6%, 47%, 47%. We note that at present we do not attempt to veto 3-prong tau decays as they are only 16% of the total dilepton background according to the MC.

The high M_T dilepton backgrounds come from MC, but their rate is normalized to the $M_T \approx 80$ GeV peak. In order to perform this normalization in data, the $W+$ jets events in the M_T peak have to be subtracted off. This introduces a systematic uncertainty.

There are two types of effects that can influence the MC dilepton prediction: physics effects and instrumental effects. We discuss these next, starting from physics.

First of all, many of our $t\bar{t}$ MC samples (eg: MadGraph) have BR($W \to \ell\nu$) = $\frac{1}{9} = 0.1111$. PDG says BR($W \to \ell\nu$) = 0.1080 ± 0.0009. This difference matters, so the $t\bar{t}$ MC must be corrected to account for this.

Second, our selection is $\ell+$ 4 or more jets. A dilepton event passes the selection only if there are two additional jets from ISR, or one jet from ISR and one jet which is reconstructed from the unidentified lepton, *e.g.*, a three-prong tau. Therefore, all MC dilepton $t\bar{t}$ samples used in the analysis must have their jet multiplicity corrected (if necessary) to agree with what is seen in $t\bar{t}$ data. We use a data control sample of well identified dilepton events with $E_{\rm T}^{\rm miss}$ and at least two jets as a template to "adjust" the N_{jet} distribution of the $t\bar{t} \to$ dileptons MC samples.

The final physics effect has to do with the modeling of $t\bar{t}$ production and decay. Different MC models could in principle result in different BG predictions. Therefore we use several different $t\bar{t}$ MC samples using different generators and different parameters, to test the stability of the dilepton BG prediction. All these predictions, **after** corrections for branching ratio and N_{jet} dependence, are compared to each other. The spread is a measure of the systematic uncertainty associated with the $t\bar{t}$ generator modeling.

The main instrumental effect is associated with the efficiency of the isolated track veto. We use tag-and-probe to compare the isolated track veto performance in $Z + 4$ jet data and MC, and we extract corrections if necessary. Note that the performance of the isolated track veto is not exactly the same on e/μ and on one prong hadronic tau decays. This is because the pions from one-prong taus are often accompanied by π^0's that can then result in extra tracks due to phton conversions. We let the simulation take care of that. Note that JES uncertainties are effectively "calibrated away" by the N_{jet} rescaling described above.

Finally, there are possible improvements to this basic analysis strategy that can be added in the future:

- Move from counting experiment to shape analysis. But first, we need to get the counting experiment under control.
- Add an explicit three prong tau veto
- Do something to require that three of the jets in the event be consistent with $t \to Wb, W \to q\bar{q}$. This could help reject some of the dilepton BG in the search for $\widetilde{t} \to t\chi^0$, but is not applicable to the $\widetilde{t} \to b\chi^+$ search.
- Consider the $M(\ell b)$ variable, which is not bounded by M_{top} in $\widetilde{t} \to b\chi^+$

3 Data Samples

[UPDATE]

The datasets used for this analysis are summarized in Tables 1 (data) and 2 (MC). The total integrated luminosity is 9.7 fb^{-1}after applying the official good run list. The main Monte Carlo samples are generated with Madgraph, though samples with alternative generators such as Powheg and MC@NLO are also used

for the derivation of systematic uncertainties in the $t\bar{t}$ background prediction. The triggers used to select both the signal and control samples are also summarized in Table. 3.

Dataset Name
Single Lepton Samples
/SingleElectron/Run2012A-13Jul2012-v1/AOD
/SingleMu/Run2012A-13Jul2012-v1/AOD
/SingleElectron/Run2012B-13Jul2012-v1/AOD
/SingleMu/Run2012B-13Jul2012-v1/AOD
/SingleElectron/Run2012C-PromptReco-v*/AOD
/SingleMu/Run2012C-PromptReco-v*/AOD
Dilepton Samples (only used for dilepton control region)
/DoubleElectron/Run2012A-13Jul2012-v1/AOD
/DoubleMu/Run2012A-13Jul2012-v1/AOD
/MuEG/Run2012A-13Jul2012-v1/AOD
/DoubleElectron/Run2012B-13Jul2012-v1/AOD
/DoubleMu/Run2012B-13Jul2012-v1/AOD
/MuEG/Run2012B-13Jul2012-v1/AOD
/DoubleElectron/Run2012C-PromptReco-v*/AOD
/DoubleMu/Run2012C-PromptReco-v*/AOD
/MuEG/Run2012C-PromptReco-v*/AOD

Table 1: Summary of data datasets used.

With Pileup: Processed dataset name is (S3) Summer12_DR53X-PU_S10_START53_V7A-v*/AODSIM (S3) Summer11-PU_S3_START42_V11-v*/AODSIM		
Description	Primary Dataset Name	cross-section [pb]
$t\bar{t}$	/TTJets_MassiveBinDECAY_TuneZ2Star_8TeV-madgraph-tauola (S3)	225.2
$W \to \ell\nu$	/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tauola (S3)	31314.0
WW	/WW_TuneZ2Star_8TeV_pythia6_tauola (S3)	45.6
WZ	/WZ_TuneZ2Star_8TeV_pythia6_tauola (S3)	18.2
ZZ	/ZZ_TuneZ2Star_8TeV_pythia6_tauola (S3)	7.4
t (s-chan)	/T_TuneZ2Star_s-channel_8TeV-powheg-tauola (S3)	3.19
$\bar{t}$ (s-chan)	/Tbar_TuneZ2Star_s-channel_8TeV-powheg-tauola (S3)	1.44
t (t-chan)	/T_TuneZ2Star_t-channel_8TeV-powheg-tauola (S3)	41.92
$\bar{t}$ (t-chan)	/Tbar_TuneZ2Star_t-channel_8TeV-powheg-tauola (S3)	22.65
tW	/T_TuneZ2Star_tW-channel-DR_8TeV-powheg-tauola (S3)	7.87
$\bar{t}W$	/Tbar_TuneZ2Star_tW-channel-DR_8TeV-powheg-tauola (S3)	7.87
$Z/\gamma^* \to \ell\ell$	/DYJetsToLL_TuneZ2Star_M-50_8TeV-madgraph-tarball (S3)	3532.8
$t\bar{t}W$	/TTW_TuneZ2Star_8TeV-madgraph (S3)	0.1633
$t\bar{t}Z$	/TTZ_TuneZ2Star_8TeV-madgraph (S3)	0.139
$t\bar{t}\gamma$	/TTPhoton_TuneZ2Star_8TeV-madgraph (S3)	0.6545
WWγ	/WWPhoton_TuneZ2Star_8TeV-madgraph (S3)	0.177
WWZ	/WWZNoGstar_TuneZ2Star_8TeV-madgraph (S3)	0.0268
WWW	/WWW_TuneZ2Star_8TeV-madgraph (S3)	0.038
WZZ	/WZZNoGstar_TuneZ2Star_8TeV-madgraph (S3)	0.0088
ZZZ	/ZZZNoGstar_TuneZ2Star_8TeV-madgraph (S3)	0.00288
$\tilde{t}\bar{\tilde{t}} \to t\bar{t}\chi_1^0\chi_1^0$	/SMS-T2tt_Mstop-225to1200_mLSP-50to1025_8TeV-Pythia6Z (S2)	scan
$\tilde{t}\bar{\tilde{t}} \to b\bar{b}\chi_1^+\chi_1^-$	/SMS-T2bw_x-0p25to0p75_mStop-50to850_mLSP-50to800_8TeV-Pythia6Z (S2)	scan
$t\bar{t}$ ($Q^2 \times 2$)	/TTjets_TuneZ2Star_scaleup_8TeV-madgraph-tauola (S3)	225.2
$t\bar{t}$ ($Q^2 \times 0.5$)	/TTjets_TuneZ2Star_scaledown_8TeV-madgraph-tauola (S3)	225.2
$t\bar{t}$ ($x_q > 40$ GeV)	/TTjets_TuneZ2Star_matchingup_8TeV-madgraph-tauola (S3)	225.2
$t\bar{t}$ ($x_q > 10$ GeV)	/TTjets_TuneZ2Star_matchingdown_8TeV-madgraph-tauola (S3)	225.2
$t\bar{t}$ ($m_{\rm top} = 178.5$ GeV)	/TTJets_TuneZ2Star_mass178_5_8TeV-madgraph-tauola (S3)	225.2
$t\bar{t}$ ($m_{\rm top} = 166.5$ GeV)	/TTJets_TuneZ2Star_mass166_5_8TeV-madgraph-tauola (S3)	225.2
$t\bar{t}$	/TT_TuneZ2Star_8TeV-powheg-tauola (S3)	225.2

Table 2: Summary of Monte Carlo datasets used. TO BE UPDATED.

Triggers
Single Muon Sample
HLT_IsoMu17_v*
HLT_IsoMu24_v*
HLT_IsoMu30_eta2p1_v*
Single Electron Sample
HLT_Ele25_CaloIdVT_TrkIdT_CentralTriJet30_v*
HLT_Ele25_CaloIdVT_TrkIdT_TriCentralJet30_v*
HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_TriCentralJet30_v*
HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_TriCentralPFJet30_v*
Dimuon Sample (only used for dilepton control regions)
HLT_DoubleMu7_v*
HLT_Mu13_Mu7_v*
HLT_Mu13_Mu8_v*
HLT_Mu17_Mu8_v*
Electron-Muon Sample (only used for dilepton control regions)
HLT_Mu17_Ele8_CaloIdL_v*
HLT_Mu8_Ele17_CaloIdL_v*
HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_v*
HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_v*
Dielectron Sample (only used for dilepton control regions)
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v*
HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_v*
HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*

Table 3: Summary of triggers used. TO BE UPDATED

4 Event Selection

This analysis uses several different control regions in addition to the signal regions. All of these different regions are defined in this section.

4.1 Single Lepton Selection

[UPDATE SELECTION]

The single lepton preselection sample is based on the following criteria

- satisfy the trigger requirement (see Table. 1). Dilepton triggers are used only for the dilepton control region.
- select events with one high p_{T} electron or muon, requiring
  - $p_{\mathrm{T}} > 30$ GeV/c and $|\eta| < 2.1$
  - satisfy the identification and isolation requirements detailed in the same-sign SUSY analysis (SUS-11-010) for electrons and the opposite-sign SUSY analysis (SUS-11-011) for muons
- require at least 4 PF jets in the event with $p_{\mathrm{T}} > 30$ GeV within $|\eta| < 2.5$ out of which at least 1 satisfies the CSV medium working point b-tagging requirement
- require moderate $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV

Table 4 shows the yields in data and MC without any corrections for this preselection region.

Table 4: Raw Data and MC predictions without any corrections are shown after preselection.

4.2 Signal Region Selection

The signal regions (SRs) are selected to improve the sensitivity for the single lepton requirements and cover a range of scalar top scenarios. The M_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ variables are used to define the signal regions and the requirements are listed in Table 5.

Signal Region	Minimum M_{T} [GeV]	Minimum $E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]
SRA	150	100
SRB	120	150
SRC	120	200
SRD	120	250
SRE	120	300

Table 5: Signal region definitions based on M_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ requirements. These requirements are applied in addition to the baseline single lepton selection.

Table 6 shows the expected number of SM background yields for the SRs. A few stop signal yields for four values of the parameters are also shown for comparison. The signal regions with looser requirements are sensitive to lower stop masses M($\tilde{t}$), while those with tighter requirements are more sensitive to higher M($\tilde{t}$).

[1 PARAGRAPH BLURB ABOUT BACKGROUNDS AND INTRODUCE CONTROL REGIONS]

4.3 Control Region Selection

Control regions (CRs) are used to validate the background estimation procedure and derive systematic uncertainties for some contributions. The CRs are selected to have similar kinematics to the SRs, but have a different requirement in terms of number of b-tags and number of leptons, thus enhancing them in different SM contributions. The four CRs used in this analysis are summarized in Table 7.

Sample	SRA	SRB	SRC	SRD
$t\bar{t} \to \ell\ell$	700 ± 15	408 ± 12	134 ± 7	43 ± 4
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	111 ± 6	71 ± 5	15 ± 2	4 ± 1
W+jets	58 ± 35	57 ± 35	29 ± 26	26 ± 26
Rare	63 ± 3	40 ± 3	17 ± 2	7 ± 1
Total	932 ± 39	576 ± 38	195 ± 27	80 ± 26

Table 6: Expected SM background contributions, including both muon and electron channels. The uncertainties are statistical only. ADD SIGNAL POINTS.

4.4 MC Corrections

[UPDATE SECTION]

4.4.1 Corrections to Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$

The official recommendations from the Jet/MET group are used for the data and MC samples. In particular, the jet energy corrections (JEC) are updated using the official recipe. L1FastL2L3Residual (L1FastL2L3) corrections are applied for data (MC), based on the global tags GR_R_42_V23 (DESIGN42_V17) for data (MC). In addition, these jet energy corrections are propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, following the official prescription for deriving the Type I corrections.

Events with anomalous "rho" pile-up corrections are excluded from the sample since these correspond to events with unphysically large $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{T} tail signal region. In addition, the recommended MET filters are applied.

4.4.2 Branching Fraction Correction

The leptonic branching fraction used in some of the $t\bar{t}$ MC samples differs from the value listed in the PDG $(10.80 \pm 0.09)\%$. Table. 8 summarizes the branching fractions used in the generation of the various $t\bar{t}$ MC samples. For $t\bar{t}$ samples with the incorrect leptonic branching fraction, event weights are applied based on the number of true leptons and the ratio of the corrected and incorrect branching fractions.

4.4.3 Modeling of Additional Hard Jets in Top Dilepton Events

[SUMMARIZE, UPDATE]

Dilepton $t\bar{t}$ events have 2 jets from the top decays, so additional jets from radiation or higher order contributions are required to enter the signal sample. The modeling of addtional jets in $t\bar{t}$ events is checked in a $t\bar{t} \to \ell\ell$ control sample, selected by requiring

- exactly 2 selected electrons or muons with $p_{\mathrm{T}} > 20$ GeV
- $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV

Selection Criteria	exactly 1 lepton	exactly 2 leptons	1 lepton + isolated track
0 b-tags	CR1) W+Jets dominated: Validate W+Jets M_{T} tail	CR2) apply Z-mass constraint $\to$ Z+Jets dominated: Validate $t\bar{t} \to \ell +$ jets M_{T} tail comparing data vs. MC "pseudo-M_{T}"	CR3) not used
≥ 1 b-tags	SIGNAL REGION	CR4) Apply Z-mass veto $\to$ $t\bar{t} \to \ell\ell$ dominated: Validate "physics" modelling of $t\bar{t} \to \ell\ell$	CR5) $t\bar{t} \to \ell\ell$, $t\bar{t} \to \ell\tau$ and $t\bar{t} \to \ell$fake dominated: Validate τ and fake lepton modeling/ detector effects in $t\bar{t} \to \ell\ell$

Table 7: Summary of signal and control regions.

$t\bar{t}$ Sample - Event Generator	Leptonic Branching Fraction
Madgraph	0.111
MC@NLO	0.111
Pythia	0.108
Powheg	0.108

Table 8: Leptonic branching fractions for the various $t\bar{t}$ samples used in the analysis. The primary $t\bar{t}$ MC sample produced with Madgraph has a branching fraction that is almost 3% higher than the PDG value.

- ≥ 1 b-tagged jet
- Z-veto

Figure 2 shows a comparison of the jet multiplicity distribution in data and MC for this two-lepton control sample. After requiring at least 1 b-tagged jet, most of the events have 2 jets, as expected from the dominant process $t\bar{t} \to \ell\ell$. There is also a significant fraction of events with additional jets. The 3-jet sample is mainly comprised of $t\bar{t}$ events with 1 additional emission and similarly the $\geq$ 4-jet sample contains primarily $t\bar{t}+ \geq 2$ jet events. Even though the primary $t\bar{t}$ Madgraph sample used includes up to 3 additional partons at the Matrix Element level, which are intended to describe additional hard jets, Figure 2 shows a slight mis-modeling of the additional jets.

It should be noted that in the case of $t\bar{t} \to \ell\ell$ events with a single reconstructed lepton, the other lepton may be mis-reconstructed as a jet. For example, a hadronic tau may be mis-identified as a jet (since no τ identification is used). In this case only 1 additional jet from radiation may suffice for a $t\bar{t} \to \ell\ell$ event to enter the signal sample. As a result, both the samples with $t\bar{t} + 1$ jet and $t\bar{t}+ \geq 2$ jets are relevant for estimating the top dilepton bkg in the signal region.

Table 9 shows scale factors to correct the fraction of events with additional jets in MC to the observed fraction in data. These are applied to the $t\bar{t} \to \ell\ell$ MC throughout the entire analysis, i.e. whenever $t\bar{t} \to \ell\ell$ MC is used to estimate or subtract a yield or distribution. In order to do so, it is first necessary to count the number of additional jets from radiation and exclude leptons mis-identified as jets. A jet is considered a mis-identified lepton if it is matched to a generator-level second lepton with sufficient energy to satisfy the jet p_{T} requirement ($p_{\mathrm{T}} > 30$ GeV).

Jet Multiplicity Sample	Data/MC Scale Factor
N jets = 3 (sensitive to $t\bar{t} + 1$ extra jet from radiation)	0.97 ± 0.03
N jets $\geq$ 4 (sensitive to $t\bar{t}+ \geq 2$ extra jets from radiation)	0.91 ± 0.04

Table 9: Data/MC scale factors used to account for differences in the fraction of events with additional hard jets from radiation in $t\bar{t} \to \ell\ell$ events.

In the signal sample, leptons mis-identified as jets are not rare. Figure 3 shows the MC jet multiplicity distribution for $t\bar{t} \to \ell\ell$ events satisfying the full selection criteria before and after subtracting leptons mis-identified as jets. Approximately a quarter of the sample is comprised of 4-jet events that actually correspond to a 2-lepton + 3 jet event where the second lepton is counted as a jet. Lepton mis-identification depends strongly on the type of second lepton, occuring more frequently in the case of hadronic τs than leptonic objects. According to the $t\bar{t} \to \ell\ell$ MC, for hadronic τs, $\sim 85\%$ of multi-prong τs and about half the single-prong τ are mis-identified as jets. In the case of leptonic objects, the fractions are smaller, comprising about a third of e/μ from a W decay and $< 20\%$ for leptonic τs, mainly because of the softness of the decay products. The scale factors listed in Table. 9 are applied to the "cleaned" jet counts in the signal sample (shown in blue in Figure 3). The impact of applying the jet multiplicity scale factors on the $t\bar{t} \to \ell\ell$ is about a 10% reduction in the background prediction for the signal region.

Ultimately, the interesting quantity for reweighting is the number of additional hard jets from radiation and this information is accessed using the number of reconstructed jets. Figure 4 demonstrates in MC that the $t\bar{t} \to \ell\ell$ control sample, i.e. when both leptons are reconstructed, can indeed be used to reweight the $t\bar{t} \to \ell\ell$ signal sample, i.e. when one lepton is missed. The figure compares the number of additional

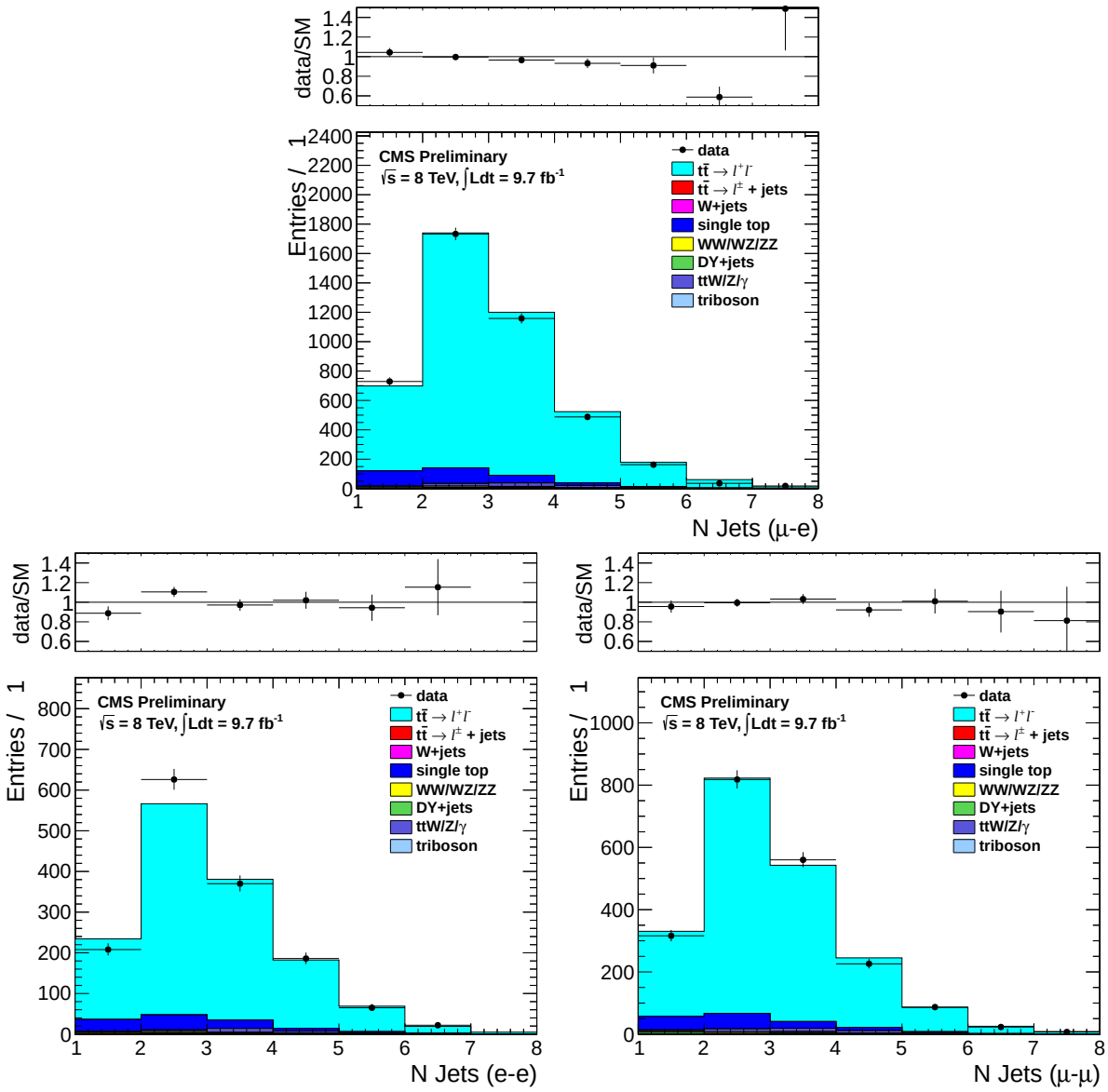


Figure 2: Comparison of the jet multiplicity distribution in data and MC for dilepton events in the e-μ (top), e-e (bottom left) and μ-μ (bottom right) channels.

jets from truth matching probed by N reconstructed jets, in this case 3 and ≥ 4 jets. In order to do so, jets that are truth-matched to the top decay products (the b-quarks and additional leptons) are removed. The 3-jet distribution shows excellent agreement and the differences in the $\geq$ 4-jet distribution are at most 5%. The impact of possible differences in the underlying distribution of extra jets between the signal and control $t\bar{t} \to \ell\ell$ samples are estimated by varying the scale factor contributions by 10% and calculating the change in the dilepton prediction. This effect is found to have a negligible impact on the prediction, well below 1%.

Other effects that have been examined include the impact of additional jets from pileup that may bias the jet multiplicity distribution, which is found to be a negligible effect in this dataset. The impact of the non-$t\bar{t} \to \ell\ell$ background on the jet fraction scale factors has also been studied. In particular, given the large uncertainty on the $Z/\gamma^* + HF$ MC prediction, this component has been varied by a factor of 2 and the resulting change on the dilepton prediction is $< 1\%$. As a result, the dominant source of uncertainty is the statistical uncertainty, primarily from the two-lepton control sample size, that corresponds to a 3% uncertainty on the dilepton prediction.

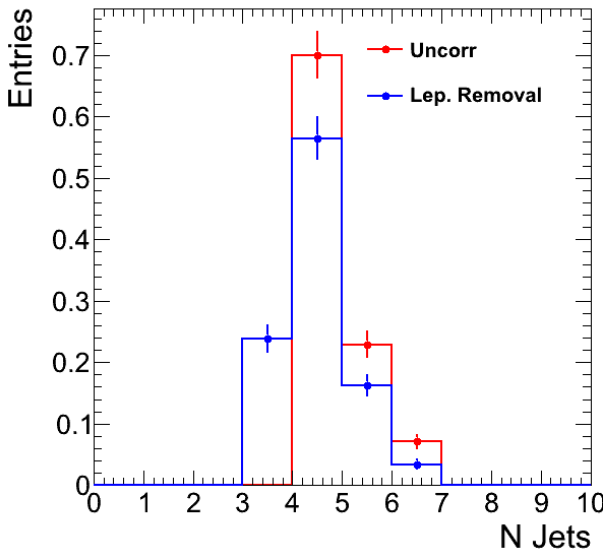


Figure 3: Comparison of the jet multiplicity distribution for $t\bar{t} \to \ell\ell$ events in MC in the signal sample before (red) and after (blue) applying the lepton-jet overlap removal. Note only the first 6 jets are shown.

The scale factors for the fraction of additional jets in the dilepton sample are applied throughout the analysis. It may be noted that this scaling is also performed consistently for the alternative $t\bar{t}$ samples, always reweighting the jet multiplicity distribution to the data in the $t\bar{t} \to \ell\ell$ control sample. In this way, effects truly arising from using different MC samples and settings can be examined, separately from issues related to the modeling of additional jets.

4.4.4 Efficiency Corrections

[TO BE UDPATED WITH T&P STUDIES ON ID, TRIGGER ETC]

4.4.5 Dilepton control regions

We define a dilepton control region requiring two isolated leptons, ee, $e\mu$, or $\mu\mu$ to study the jet multiplicity in data and MC, and derive scale factors based on their consistency. This study is documented in Section 4.4.3.

In this region we require:

- dilepton triggers
- two leptons with $p_{\mathrm{T}} > 20$GeV that pass our lepton id and isolation
- $E_{\mathrm{T}}^{\mathrm{miss}} > 50$GeV
- ≥ 1 b-tag, SSV medium

This sample is only partially overlapping with the single lepton preselection as it requires the dilepton rather than the single lepton triggers, and differs in the p_{T} requirement for the leading lepton. Table 10 shows the raw yields in data and MC prior to any corrections.

Table 10: Raw Data and MC predictions without any corrections are shown for the dilepton control region. This region is used for correcting the jet multiplicity seen in MC to that in data.

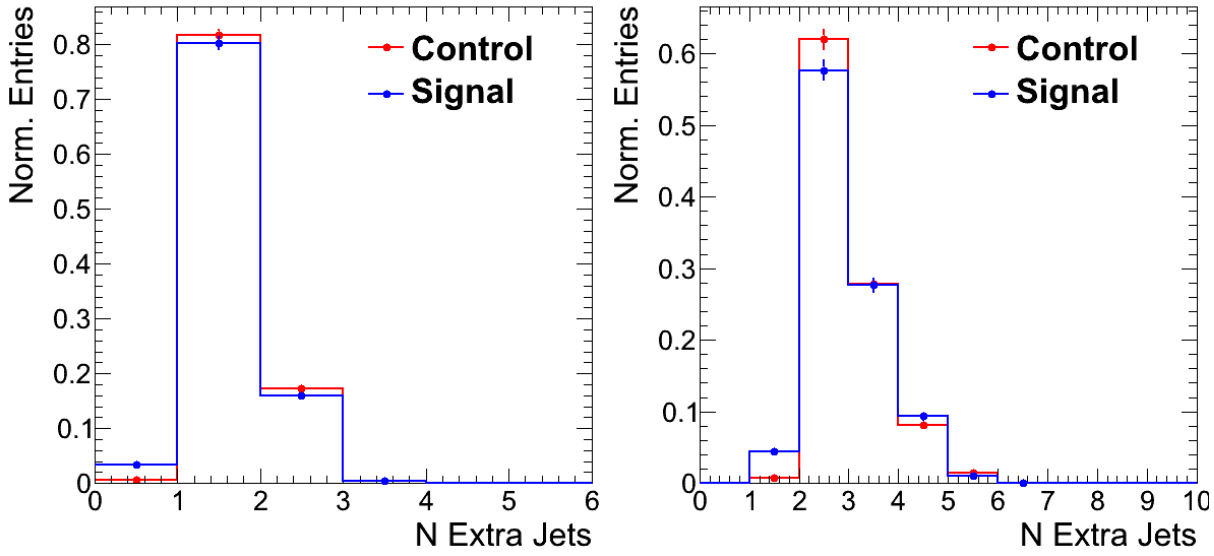


Figure 4: Comparison of the number of additional jets from radiation in the 3-jet (left) and $\geq$ 4-jet (right) bins for the control $t\bar{t} \to \ell\ell$ sample (with two reconstructed leptons) and the signal sample (with one reconstructed lepton). The distributions show good agreement, indicating that the usage of the reconstructed jet multiplicity in one sample to reweight the signal sample is indeed justified. **Fix me: Is this before or after the isolated track veto?**

5 Background Estimation

In order to search for a possible signal from stop decays giving rise to a signature of $t\bar{t}$ with additional $E_{\mathrm{T}}^{\mathrm{miss}}$ from the LSPs, it is necessary to determine the composition of the SM backgrounds in the signal region. This section details the methods pursued to estimate the background in the signal sample and describes the procedure to estimate the systematic uncertainties. The general strategy is to use the MC prediction for the backgrounds after applying corrections derived from data.

The most important background to a stop signal arises from SM $t\bar{t}$. The $t\bar{t}$ background may be separated into contributions containing a single lepton $t\bar{t} \to \ell +$ jets and two leptons $t\bar{t} \to \ell\ell$. As described in this section, the $t\bar{t} \to \ell\ell$ background is the dominant process in our signal region ($E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeVand $M_{\mathrm{T}} > 150$ GeV, ≥ 1 b-tags, isolated track veto), contributing $\sim 80\%$ of the background yield. This background has large true $E_{\mathrm{T}}^{\mathrm{miss}}$ and consequently larger M_{T} due to the presence of two neutrinos. Additional contributions to the single lepton sample arise from W+jets and single top. The combination of all single lepton backgrounds, $t\bar{t} \to \ell +$ jets, W+jets and single top, comprises $\sim 15\%$ of the signal sample. Finally, other background sources such as dibosons, $Z/\gamma^* +$ jets, in addition to rarer processes such as $t\bar{t}$ produced in association with a vector boson, dibosons and tribosons, provide a combined contribution to the signal sample at the level of $\sim 5\%$. Finally, the QCD background contribution is small, particularly in the signal sample, with a large $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement.

The total bkg in the signal region is estimated according to:

$$N_{bkg} = N_{1lep} + N_{2lep} + N_{rare}$$

$$N_{1lep} = N_{1lep}^{MC} \times \frac{(1-\epsilon_{fake})^{data}}{(1-\epsilon_{fake})^{MC}} \times \frac{N_{peak}^{data}}{N_{peak}^{MC}}$$

$$N_{2lep} = N_{2lep}^{MC} \times \frac{(1-\epsilon_{iso\ trk})^{data}}{(1-\epsilon_{iso\ trk})^{MC}} \times \frac{N_{peak}^{data}}{N_{peak}^{MC}}$$

All of these terms will be defined clearly in this section, including their corrections and sources of systematic errors.

5.1 Single Lepton Backgrounds

This category of backgrounds includes processes with a single leptonic W decay, giving rise to one lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ from a single neutrino. As a result, the M_{T} variable, constructed from the lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$, exhibits a kinematic edge at $M_{\mathrm{T}} \sim M_{\mathrm{W}}$. The main contributors to this background are $\mathrm{t\bar{t}} \to \ell + \mathrm{jets}$, W+jets and single top, though in the latter case there is a contribution from tW that can give rise to two leptons. The 2-lepton single top background is included in the rare backgrounds estimate.

5.1.1 W+Jets MC Modelling Validation

The estimate of the uncertainty on this background is based on CR1, defined by applying the full signal selection, but requiring 0 b-tags. The sample is dominanted by W+jets and is thus used to validate the MC modelling of this background.

Sample	CR1PRESEL	CR1A	CR1B	CR1C	CR1D
Muon M_{T}-SF	0.91 ± 0.03	0.96 ± 0.07	0.88 ± 0.11	1.05 ± 0.21	1.27 ± 0.41
Electron M_{T}-SF	0.82 ± 0.03	0.86 ± 0.06	1.09 ± 0.15	1.24 ± 0.30	1.11 ± 0.40

Table 11: M_{T} peak Data/MC scale factors applied to the single lepton samples and $\mathrm{t\bar{t}} \to \ell\ell$. The raw MC is used for backgrounds from rare processes. CR1PRESEL refers to a sample with $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV. The uncertainties are statistical only.

Sample	CR1PRESEL	CR1A	CR1B	CR1C	CR1D
Muon MC	456 ± 73	174 ± 44	51 ± 7	18 ± 2	10 ± 2
Muon Data	657	246	142	43	12
Muon Data/MC SF	1.44 ± 0.24	1.41 ± 0.37	2.80 ± 0.47	2.37 ± 0.46	1.23 ± 0.42
Electron MC	396 ± 64	147 ± 36	54 ± 4	19 ± 2	8 ± 2
Electron Data	702	223	144	50	23
Electron Data/MC SF	1.77 ± 0.29	1.52 ± 0.39	2.68 ± 0.30	2.57 ± 0.49	2.73 ± 0.76

Table 12: Yields in M_{T} tail comparing the MC prediction (after applying SFs) to data. CR1PRESEL refers to a sample with $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV and $M_{\mathrm{T}} > 150$ GeV. The uncertainties are statistical only.

5.1.2 Single Lepton Top MC Modelling Validation

The M_{T} tail for single-lepton top events ($\mathrm{t\bar{t}} \to \ell + \mathrm{jets}$ and single top) is dominated by jet resolution effects. The W cannot be far off-shell because $M_{\mathrm{W}} < M_{\mathrm{top}}$. The modeling of the M_{T} tail from jet resolution effects is studied using Z+jets data and MC samples. Z events are selection by requiring 2 good leptons (satisfying ID and isolation requirements) and requiring the $M_{\ell\ell}$ to be in the range $81-101$ GeV. The negative lepton is treated as a neutrino and so is added to the MET: $E_{\mathrm{T}}^{\mathrm{miss}} \to p_{\mathrm{T}}(\ell^-) + E_{\mathrm{T}}^{\mathrm{miss}}$, and the M_{T} is recalculated with the positive lepton $M_{\mathrm{T}}(\ell^+, E_{\mathrm{T}}^{\mathrm{miss}})$. The resulting "pseudo-$M_{\mathrm{T}}$" is dominated by jet resolution effects, since no off-shell Z production enters the sample due to the $M_{\ell\ell}$ requirement. This section describes how well the MC predicts the tail of "pseudo-M_{T}".

5.2 Top Dilepton Background

The dominant background to the signal sample comes from $\mathrm{t\bar{t}} \to \ell\ell$ events. Due to the presence of a second neutrino, $\mathrm{t\bar{t}} \to \ell\ell$ events do not have a kinematic edge at $M_{\mathrm{T}} \sim M_{\mathrm{W}}$. These events satisfy the selection criteria due to real $E_{\mathrm{T}}^{\mathrm{miss}}$ and do not depend on detector resolution or $E_{\mathrm{T}}^{\mathrm{miss}}$ mis-measurement effects. As a result, the $\mathrm{t\bar{t}} \to \ell\ell$ background is expected to be well modeled in the MC. The prediction for this background is thus derived from MC and normalized to the data in the M_{T} peak region. However, there are two aspects that require dedicated corrections and are detailed in this section:

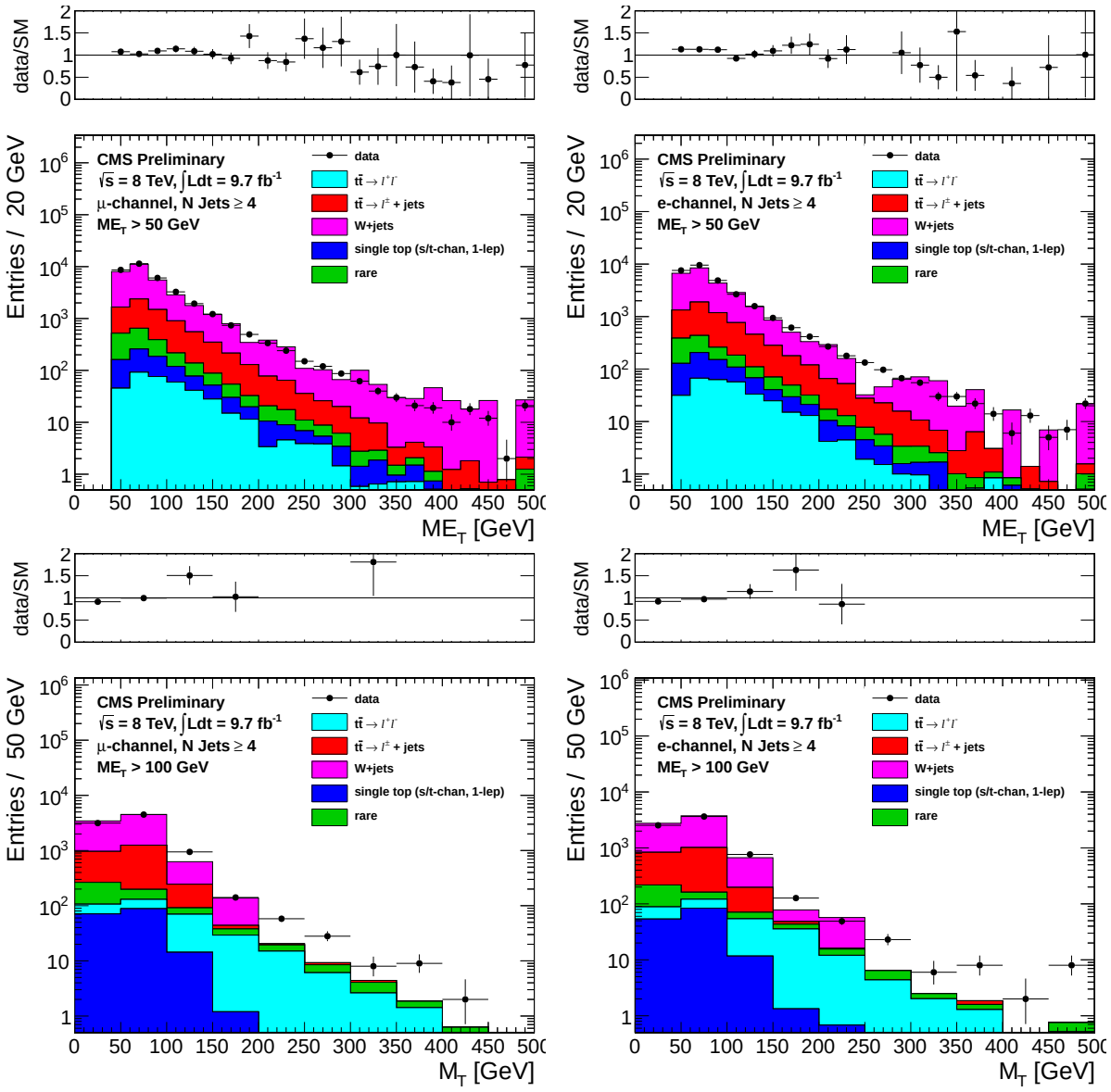


Figure 5: Comparison of the $E_{\mathrm{T}}^{\mathrm{miss}}$ (top) and M_{T} for $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ (bottom) distributions in data vs. MC for events with a leading muon (left) and leading electron (right) satisfying the requirements of CR1.

Sample	CR2PRESEL0	CR2PRESEL1	CR2A	CR2B	CR2C	CR2D
DY MC	35 ± 2	30 ± 2	18 ± 2	32 ± 3	12 ± 2	5 ± 1
Data - non-DY MC	65 ± 9	50 ± 8	36 ± 6	49 ± 7	25 ± 5	14 ± 4
Data/MC SF	1.88 ± 0.29	1.68 ± 0.30	1.94 ± 0.40	1.54 ± 0.29	2.12 ± 0.58	2.96 ± 1.22

Table 13: Yields in M_{T} tail comparing the MC prediction (after applying SFs) to data. CR2PRESEL refers to a sample with $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV and $M_{\mathrm{T}} > 150$ GeV. The uncertainties are statistical only.

- the modeling of additional jets from radiation, required to satisfy the 4-jet selection criteria.
- the modeling of the isolated track veto efficiency, which is applied to explicitly reject leptons from W and W $\to \tau$ decays and single prong τ decays.

The systematic uncertainty associated with the MC prediction then has two components. One that is derived by comparing various generators, and a second from the uncertainties on the various correction

factors used. These are described at the end of this section.

5.2.1 Normalization of the Top Prediction

In this section we discuss the factor $\frac{N_{peak}^{data}}{N_{peak}^{MC}}$ in Equation XX. The same factor is applied to both the single and dilepton estimates.

The overall normalization of the $t\bar{t}$ sample is determined by scaling to the $M_{\rm T}$ peak control region, following a procedure similar to that described in Section. 5.1. This control region is dominated by $t\bar{t}$ albeit in its single lepton decay mode. The basic idea is that after adjusting the modeling of additional jets from radiation in the $t\bar{t} \to \ell\ell$ sample and correcting the leptonic branching fractions in the $t\bar{t}$ sample, the MC prediction for the $t\bar{t} \to \ell +$jets and $t\bar{t} \to \ell\ell$ samples is subject to the same sources of uncertainty: the $t\bar{t}$ cross section, the luminosity, the selection efficiencies, etc... The exception is the veto on events containing an isolated track, since this last requirement has a different impact on the $t\bar{t} \to \ell +$ jets and $t\bar{t} \to \ell\ell$ samples. The impact of this requirement is addressed separately in Section. 5.2.2.

The $M_{\rm T}$-peak scale factor is thus determined after applying the full analysis selection with the exception of the isolated track veto. Specifically, the pre-veto sample is defined by the following requirements

- At least 1 selected electron (muon) with $p_{\rm T}>$ 30 GeV and $|\eta| < 2.5$ ($|\eta| < 2.1$)
- At least 4 selected jets, of which at least 1 is b-tagged
- $E_{\rm T}^{\rm miss} > 100$ GeV

Scaling the overall normalization to the $M_{\rm T}$ peak largely reduces the dependence on the $t\bar{t}$ cross section and cancels systematic uncertainties associated with effects such as the luminosity, selection efficiencies, etc... However, the $M_{\rm T}$ peak control sample is contaminated by non-$t\bar{t}$ processes, particularly W+jets that contributes at the $5-10\%$ level, even after requiring a b-tagged jet. The W+jets+HF process is a particular concern given the large theoretical uncertainties associated with their production. Therefore a systematic uncertainty is derived to account for the uncertainty in this background component. The normalization of the W+jets sample is scaled up and down by 50% and the full background estimate recomputed.

It should be noted that the background prediction obtained using the same normalization scale factor based on the $M_{\rm T}$ peak for the single lepton and dilepton samples is subject to smaller uncertainties than the prediction obtained by normalizing the $t\bar{t} \to \ell\ell$ component alone to the overall yield in the two-lepton control sample. The reason is that the normalization of the two-lepton control yield depends on effects that do not impact the $t\bar{t} \to \ell\ell$ sample with one reconstructed lepton (i.e. the signal sample of interest) in the same way. Example of these effects include the dilepton trigger and second lepton reconstruction efficiencies.

In conclusion, the pre-veto sample is used to define an overall data over MC scale factor ($SF^{\rm all}$) in the $M_{\rm T}$ peak control region, that is applied to all background predictions and is simply defined as

- $N_{\rm peak}^{\rm all}$ = data yield in the peak region $60 < M_{\rm T} < 100$ GeV
- $M_{\rm peak}^{\rm all}$ = MC yield in the peak region $60 < M_{\rm T} < 100$ GeV
- $SF^{\rm all} = N_{\rm peak}^{\rm all}/M_{\rm peak}^{\rm all}$

For all subsequent steps, the scale factor $SF^{\rm all}$ is applied to all MC contributions.

5.2.2 The Isolated Track Veto

In this section we discuss the factors $\frac{(1-\epsilon_{fake})^{data}}{(1-\epsilon_{fake})^{MC}}$ and $\frac{(1-\epsilon_{iso\ trk})^{data}}{(1-\epsilon_{iso\ trk})^{MC}}$ in Equation XXX.

The $t\bar{t} \to \ell\ell$ background is further suppressed after the 4-jet requirement by removing events with any isolated track with $p_{\rm T} > 10$ GeV. As isolation definition we use relative track isolation $\sum p_{\rm T}/p_{\rm T}(trk)$ in a cone of size $R = 0.3 < 0.1$. This isolated track veto rejects events with an e or a μ, as well as single-prong

τ-decays. This veto is very effective at reducing the dilepton background. In particular, according to the $t\bar{t} \to \ell\ell$ MC, the veto removes about three-quarters of events with an e or μ from the W decay and almost half the leptonic and single prong τ decays. The veto has no impact on multi-prong τs, though this is a smaller component overall. Since the $t\bar{t} \to \ell\ell$ background includes different types of processes, it is useful to first characterize the composition of this background.

5.2.3 Top Dilepton Sample Composition

The $t\bar{t} \to \ell\ell$ background may be categorized based on the type of second lepton, as shown in Table. 14. The main component is from electrons and muons from a W decay or through an intermediate τ decay. The second largest component arises from single-prong hadronic τ decays, followed by multi-prong τs. Finally an additional contribution arises from leptons falling in the forward region, outside the Tracker acceptance $|\eta| > 2.5$ (refered to as 'lost').

Sample	Yield	Fraction [%]
$t\bar{t} \to l^+l^-$(lost)	7 ± 1	6
$t\bar{t} \to l^+l^-(e/\mu)$	30 ± 3	26
$t\bar{t} \to l^+l^-(\tau_{\rm lep})$	21 ± 2	18
$t\bar{t} \to l^+l^-(\tau_{\rm had} \to 1-{\rm prong})$	39 ± 3	34
$t\bar{t} \to l^+l^-(\tau_{\rm had} \to 3-{\rm prong})$	19 ± 2	16
total $t\bar{t} \to l^+l^-$	117 ± 5	100

Table 14: Dilepton events satisfying the full selection criteria and $E_{\rm T}^{\rm miss} > 100$ GeV, $M_{\rm T} > 150$ GeV, separated by decay modes. Recall that $t\bar{t} \to \ell\ell$ accounts for $\approx 80\%$ of the total background.**Fix me: Is this before or after the isolated track veto?**

The isolated track veto does not apply to the components where the second lepton falls outside the acceptance or where it decays to a hadronic tau that is not explicitly rejected. For the cases where the second lepton includes an electron or muon or a charged π/K, it is possible to further distinguish cases when the relevant particle targeted by the veto is below the $p_{\rm T}$ threshold. Matching the truth-level particle to reconstructed tracks shows that in $t\bar{t} \to \ell\ell$ MC

- for $t\bar{t} \to l^+l^-(e/\mu)$, about a third of the sample falls below the $p_{\rm T}$ threshold of the track veto, and the remaining two thirds fail the isolation
- for $t\bar{t} \to l^+l^-(\tau_{\rm lep})$, about 80% are soft $p_{\rm T} < 10$ GeV and about 20% are non-isolated
- for $t\bar{t} \to l^+l^-(\tau_{\rm had} \to 1-{\rm prong})$, about 70% are soft, as expected from a τ decay product and the rest fail the isolation criteria.

In summary, the combination of these fractions with the relative sample composition listed in Table. 14 shows that only about a third of the $t\bar{t} \to \ell\ell$ background sample is from 2nd leptons (e, μ, or $\tau \to$1-prong) which satisfy $p_{\rm T} > 10$ GeV but fail the track isolation criterion veto[1)]. The performance of the isolation used in the track veto requirement is the subject of the next section.

It should also be noted that according to the MC, track reconstruction inefficiencies affect a few percent ($\sim 1-2\%$) of the leptonic and single prong τ events. The tracking efficiency in this analysis is taken from MC, which is expected to provide good modeling of isolated tracks with $p_{\rm T} > 10$ GeV. The impact of possible differences between data and MC is found to be negligible. In particular, the case of single-prong taus is the most challenging to model due to the effect of nuclear interactions in the tracker material. Past studies of the tracking efficiency for pions [6] provide a data/MC uncertainty in the tracking efficiency of 3.9%[2)]. Propagating this uncertainty to the total background estimate yields a total uncertainty of

[1)] Explicitly, the fraction of events that give rise to a sufficiently energetic lepton or single prong τ is: 70% of e-μ events which are 26% of the sample, 20% of leptonic tau events which are 18% of the sample and 30% of single prong τ events which are 10% of the sample.

[2)] This tracking efficiency uncertainty estimate is conservative for this analysis since it includes tracks of $p_{\rm T}$ down to 250 MeV, where material effects are larger and so are the corresponding uncertainties.

$< 0.5\%$. The reason is that the tracking efficiency uncertainty only applies to single prong τ decays with $p_{\mathrm{T}} > 10$ GeV, which are under 30% of the dilepton component, which in turn is $\sim 80\%$ of the total sample.

To conclude, the $\mathrm{t\bar{t}} \to \ell\ell$ background arises from events where the second lepton falls outside the acceptance (both in η and p_{T}), because the event contains a hadronic tau that is not explicitly rejected or because the second lepton fails the isolation requirement. Even though the $\mathrm{t\bar{t}} \to \ell\ell$ sample is quite heterogenous and comprises multiple types of second lepton events, there are two main sources of uncertainty in this estimate:

- Acceptance effects, which are estimated by using alternative MC samples. Here acceptance refers to the combination of η and $p_{\mathrm{T}}$of the leptons.
- Detector effects, mainly arising from understanding the performance of the isolated track veto, which impacts only about a third of the total $\mathrm{t\bar{t}} \to \ell\ell$ sample.

5.2.4 Summary of the $\mathrm{t\bar{t}} \to \ell\ell$ Background Estimation Procedure

The SM background in the signal region, defined by requirements of large $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{T}, is estimated using MC. The MC is validated using data control samples, which are used to derive data-to-MC scale factors and corresponding uncertainties.

The procedure to estimate the background prediction may be summarized as

- Apply the state-of-the-art corrections to the MC, reflecting the best knowlege of the detector performance, in order to improve the agreement with the data. This includes effects such as the modeling of the pileup, the jet energy scale, $E_{\mathrm{T}}^{\mathrm{miss}}$ corrections, etc...
- Correct the leptonic branching fractions in the $\mathrm{t\bar{t}}$ MC
- Use the dilepton sample with two selected leptons to reweight the N_{jets} distribution in $\mathrm{t\bar{t}} \to \ell\ell$ MC, which is not necessarily well-modeled due to the presence of additional jets from radiation.
- Use the pre-veto sample (i.e. applying the full analysis selection with the exception of the isolated track veto) to define a scale factor in the M_{T} peak region. This scale factor corrects for effects of integrated luminosity, $\mathrm{t\bar{t}}$ cross section, lepton selection and trigger efficiencies.
- Correct the $\mathrm{t\bar{t}} \to \ell\ell$ sample for differences between data and MC in the isolation for events with a second lepton. This correction is derived using $\mathrm{Z} + 4$ jet events and applied to the $\mathrm{t\bar{t}} \to \ell\ell$ sample.
- In the signal sample, after applying the full selection including the isolated track veto, derive a scale factor to account for possible data vs. MC discrepancies in the isolated track fake rate for backgrounds which have a single genuine lepton. This scale factor is applied to the single lepton backgrounds only.

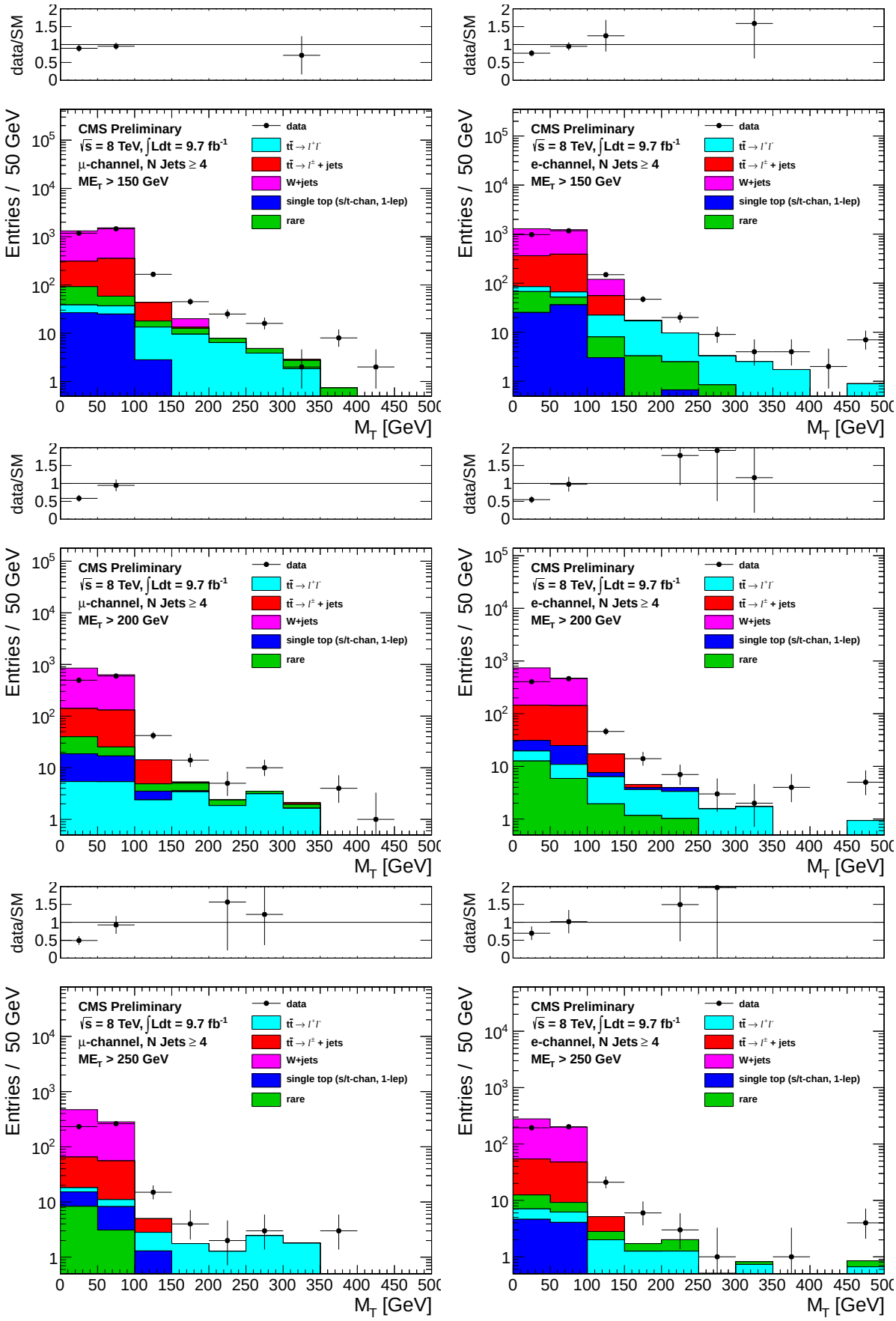


Figure 6: Comparison of the M_{T} distribution in data vs. MC for events with a leading muon (left) and leading electron (right) satisfying the requirements of CR1. The $E_{\mathrm{T}}^{\mathrm{miss}}$ requirements used are 150 GeV (top), 200 GeV (middle) and 250 GeV (bottom).

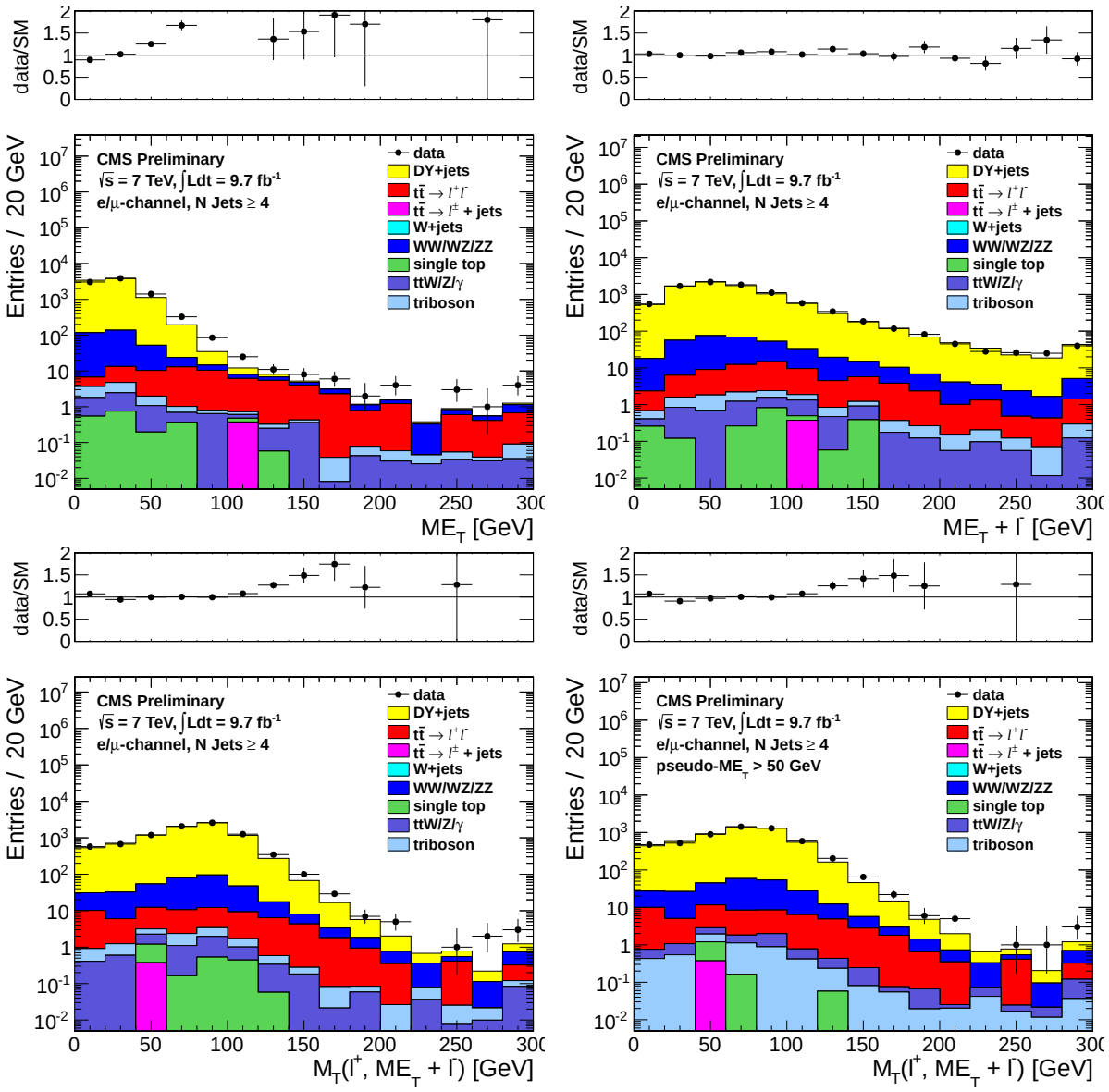


Figure 7: Comparison of the $E_{\mathrm{T}}^{\mathrm{miss}}$ (top, left), pseudo-$E_{\mathrm{T}}^{\mathrm{miss}}$ (top, right) and pseudo-M_{T} (bottom) distributions in data vs. MC for events satisfying the requirements of CR2, combining both the muon and electron channels. The pseudo-M_{T} distributions are shown before any additional requirements (bottom, left) and after requiring pseudo-$E_{\mathrm{T}}^{\mathrm{miss}}$¿50 GeV (bottom, right).

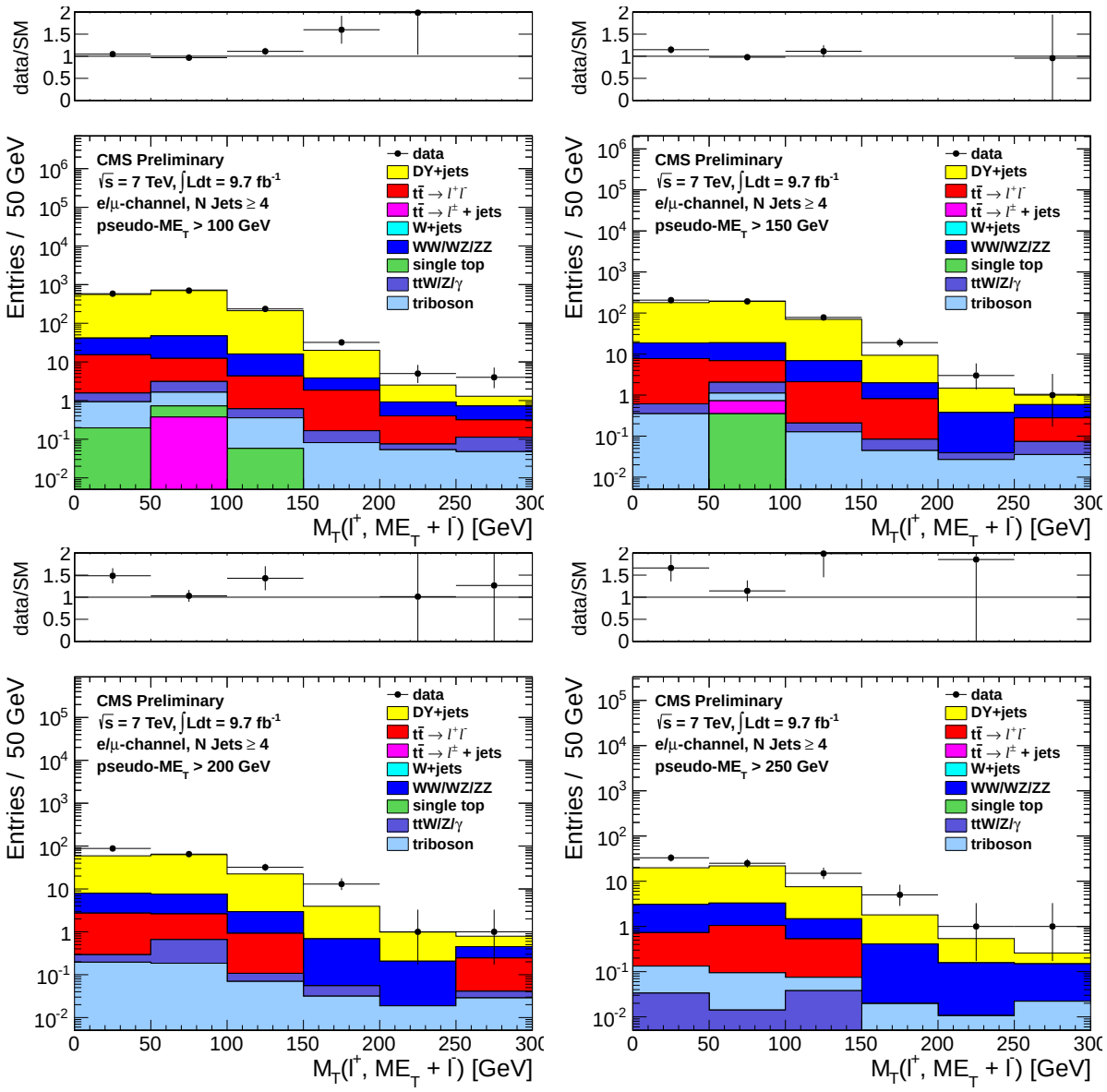


Figure 8: Comparison of the M_{T} distribution in data vs. MC for events satisfying the requirements of CR2, combining both the muon and electron channels. The pseudo-$E_{\mathrm{T}}^{\mathrm{miss}}$ requirements used are 100 GeV (top, left), 150 GeV (top, right), 200 GeV (bottom, left) and 250 GeV (bottom, right).

5.3 Check of MC modelling of $t\bar{t} \to \ell\ell$

[EXPLAIN THE CROSS CHECKS DONE TO VALIDATE THE MC MODELLING OF $t\bar{t} \to \ell\ell$]

5.3.1 Validation of the "Physics" Modelling of the $t\bar{t} \to \ell\ell$ MC

[EXPLAIN ALL THE CR4 2-lepton $t\bar{t} \to \ell\ell$ SAMPLE]

Sample	CR4A	CR4B	CR4C	CR4D
Muon Data/MC-SF	0.91 ± 0.04	0.94 ± 0.07	1.06 ± 0.13	1.03 ± 0.22
Electron Data/MC-SF	0.95 ± 0.04	1.00 ± 0.08	0.85 ± 0.12	0.83 ± 0.19

Table 15: Data/MC scale factors for total yields, applied to compare the shapes of the distributions. The uncertainties are statistical only.

Sample	CR4A	CR4B	CR4C	CR4D
Muon MC	199 ± 7	102 ± 6	29 ± 3	8 ± 1
Muon Data	187	108	34	9
Muon Data/MC SF	0.94 ± 0.08	1.06 ± 0.12	1.17 ± 0.23	1.09 ± 0.40
Electron MC	203 ± 8	97 ± 5	26 ± 2	8 ± 1
Electron Data	201	102	25	5
Electron Data/MC SF	0.99 ± 0.08	1.06 ± 0.12	0.97 ± 0.21	0.60 ± 0.29

Table 16: Yields in M_{T} tail comparing the MC prediction (after applying SFs) to data. The uncertainties are statistical only.

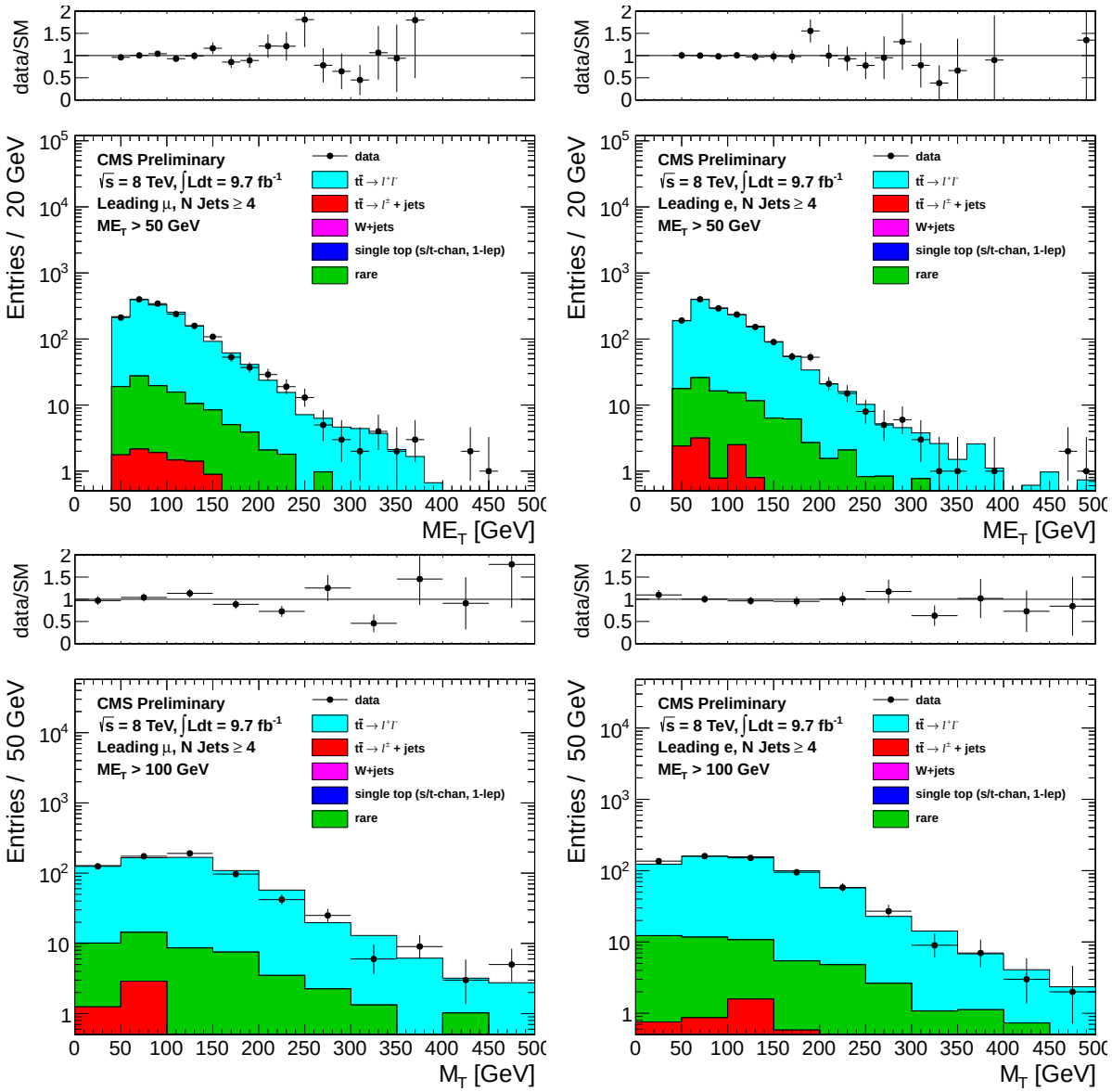


Figure 9: Comparison of the $E_{\mathrm{T}}^{\mathrm{miss}}$ (top) and M_{T} for $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ (bottom) distributions in data vs. MC for events with a leading muon (left) and leading electron (right) satisfying the requirements of CR4.

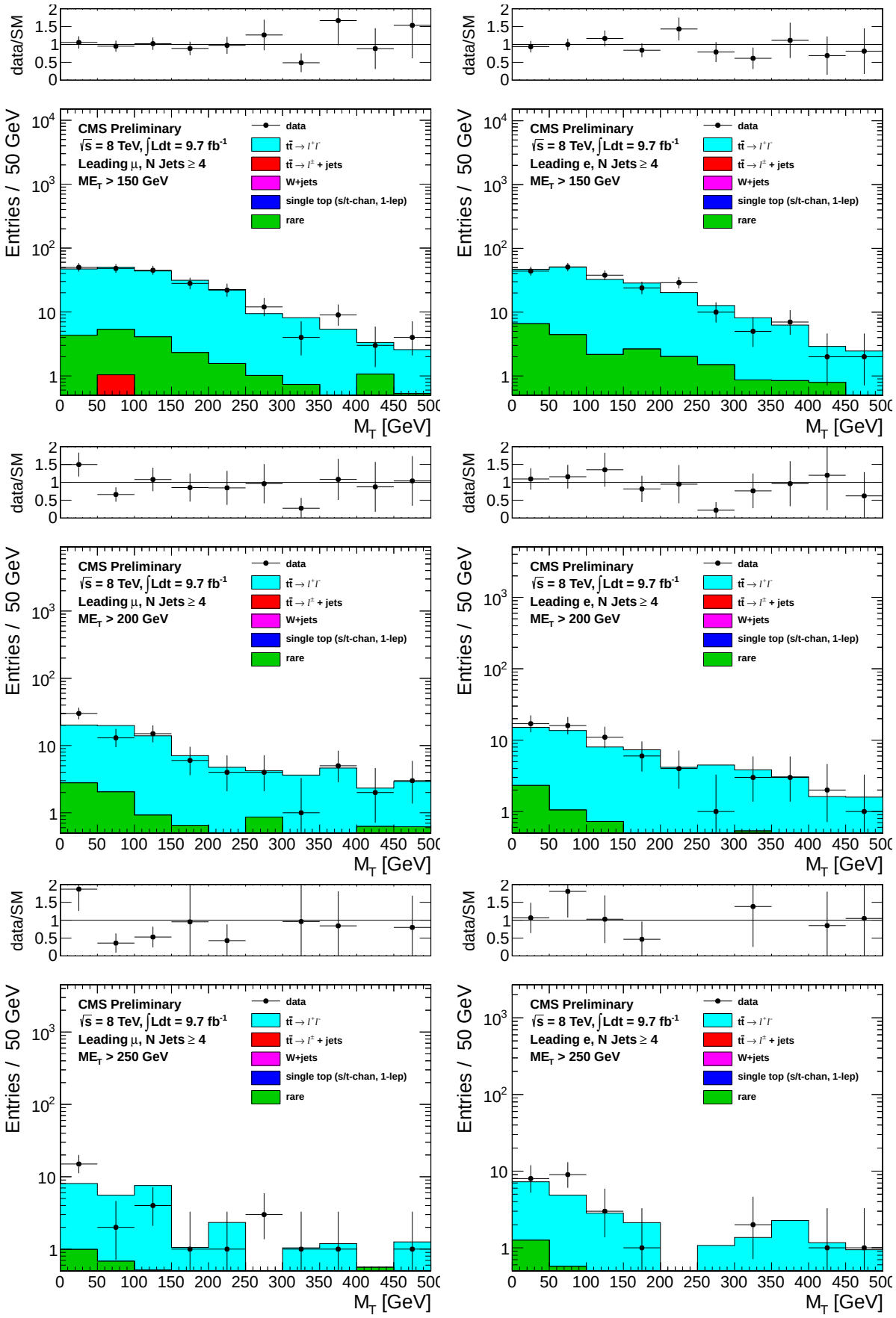


Figure 10: Comparison of the M_{T} distribution in data vs. MC for events with a leading muon (left) and leading electron (right) satisfying the requirements of CR4. The $E_{\mathrm{T}}^{\mathrm{miss}}$ requirements used are 150 GeV (top), 200 GeV (middle) and 250 GeV (bottom).

5.3.2 Validation of the lepton + isolated track Sample Prediction

[EXPLAIN ALL THE CHECKS FOR CR5: LEPTON + ISOLATED TRACK SAMPLE]

[ALSO NEED BETTER TITLE FOR THIS SECTION!]

Sample	CR5A	CR5B	CR5C	CR5D
Muon pre-veto M_{T}-SF	0.98 ± 0.02	0.95 ± 0.04	0.99 ± 0.08	0.89 ± 0.15
Muon post-veto M_{T}-SF	1.28 ± 0.07	1.20 ± 0.13	1.22 ± 0.24	1.25 ± 0.43
Electron pre-veto M_{T}-SF	0.83 ± 0.02	0.75 ± 0.04	0.64 ± 0.07	0.63 ± 0.12
Electron post-veto M_{T}-SF	1.10 ± 0.08	1.02 ± 0.11	0.89 ± 0.19	1.27 ± 0.41

Table 17: M_{T} peak Data/MC scale factors. The pre-veto SFs are applied to the $\mathrm{t\bar{t}} \to \ell\ell$ sample, while the post-veto SFs are applied to the single lepton samples. The raw MC is used for backgrounds from rare processes. The uncertainties are statistical only.

Sample	CR5A	CR5B	CR5C	CR5D
Muon MC	293 ± 9	161 ± 7	51 ± 4	16 ± 2
Muon Data	315	165	62	13
Muon Data/MC SF	1.07 ± 0.07	1.03 ± 0.09	1.21 ± 0.18	0.82 ± 0.25
Electron MC	253 ± 8	126 ± 5	37 ± 3	12 ± 2
Electron Data	286	135	39	15
Electron Data/MC SF	1.13 ± 0.08	1.07 ± 0.10	1.07 ± 0.19	1.21 ± 0.35

Table 18: Yields in M_{T} tail comparing the MC prediction (after applying SFs) to data. The uncertainties are statistical only.

5.4 Other Backgrounds

Additional background contributions from rare processes include

- $\mathrm{t\bar{t}}$ in association with a boson $\mathrm{t\bar{t}}$ + WZ/γ^*
- Z/γ^*+Jets
- diboson WW/WZ/ZZ
- triboson WWW/WWZ/WZZ/ZZZ
- dilepton single top tW.

These backgrounds are small, contributing at the $\sim 5\%$ level and their predictions are taken from MC, normalized to the corresponding cross sections. A 50% systematic uncertainty is assigned for all these backgrounds.

Backgrounds from QCD are expected to be small in the signal regions with large M_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$

5.5 Background Prediction

6 Background Estimate Derivation

[THIS IS WHERE THE NUBMERS TO DERIVE THE BACKGROUND ESTIMATES ARE DUMPED AND THE CALCULATION EXPLAINED]

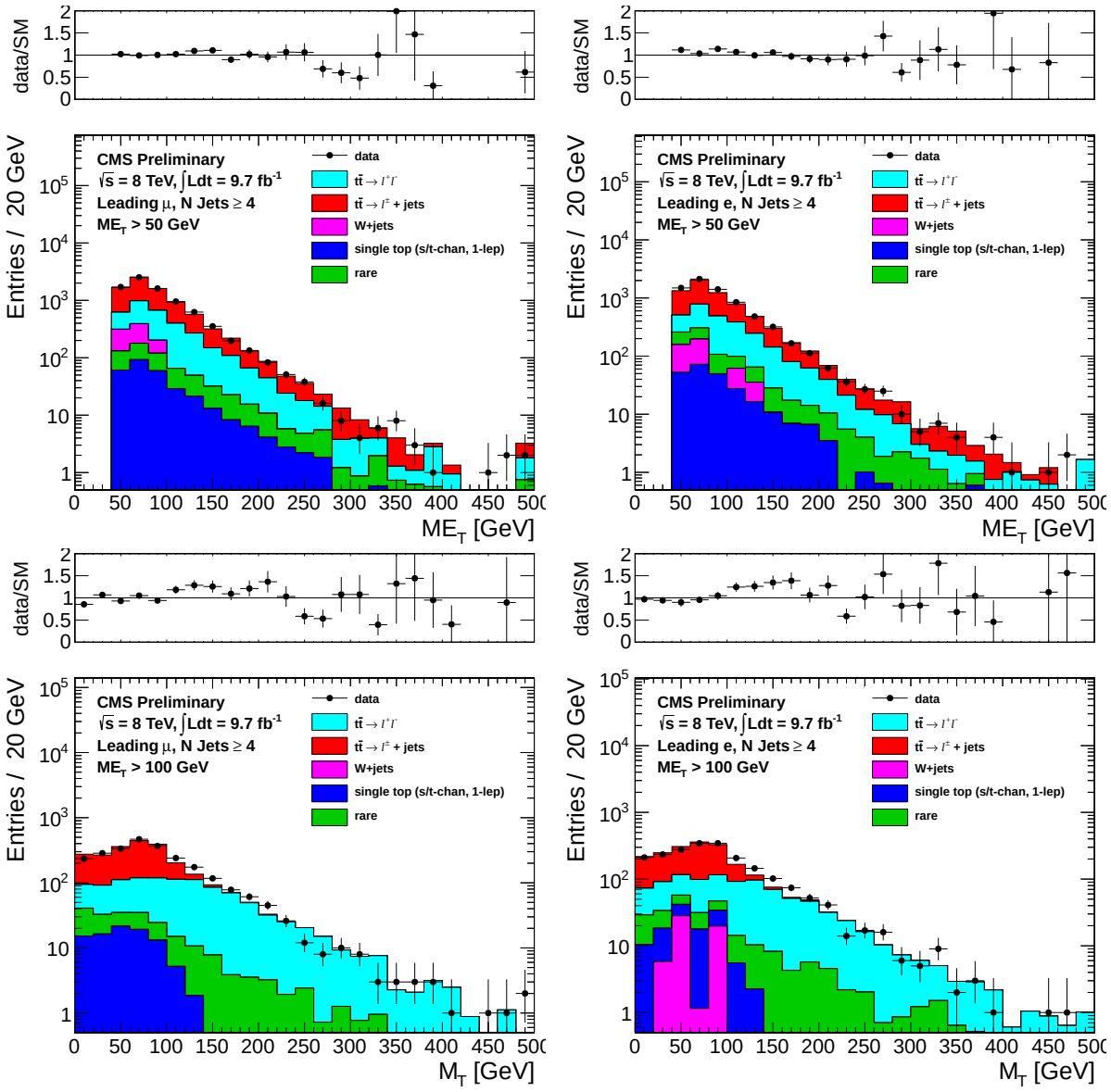


Figure 11: Comparison of the $E_{\mathrm{T}}^{\mathrm{miss}}$ (top) and M_{T} for $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ (bottom) distributions in data vs. MC for events with a leading muon (left) and leading electron (right) satisfying the requirements of CR5.

7 Systematic Uncertainties

7.1 Uncertainty on the $\mathrm{t\bar{t}} \to \ell\ell$ Acceptance

The $\mathrm{t\bar{t}}$ background prediction is obtained from MC, with corrections derived from control samples in data. The uncertainty associated with the theoretical modeling of the $\mathrm{t\bar{t}}$ production and decay is estimated by comparing the background predictions obtained using alternative MC samples. It should be noted that the full analysis is performed with the alternative samples under consideration, including the derivation of the various data-to-MC scale factors. The variations considered are

- Top mass: The alternative values for the top mass differ from the central value by 5 GeV: $m_{\mathrm{top}} = 178.5$ GeV and $m_{\mathrm{top}} = 166.5$ GeV.
- Jet-parton matching scale: This corresponds to variations in the scale at which the Matrix Element partons from Madgraph are matched to Parton Shower partons from Pythia. The nominal value is $x_q > 20$ GeV. The alternative values used are $x_q > 10$ GeV and $x_q > 40$ GeV.

Sample	SRA	SRB	SRC	SRD
Muon				
$t\bar{t} \to \ell\ell$	536 ± 13	145 ± 7	42 ± 4	16 ± 2
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	4953 ± 41	1408 ± 22	429 ± 12	149 ± 7
W+jets	332 ± 88	74 ± 38	42 ± 30	42 ± 30
Rare	152 ± 6	55 ± 4	19 ± 2	8 ± 1
Total	5973 ± 99	1682 ± 45	532 ± 32	214 ± 31
Data	5861	1608	526	192
Electron				
$t\bar{t} \to \ell\ell$	498 ± 12	146 ± 7	49 ± 4	17 ± 2
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	4581 ± 39	1307 ± 21	407 ± 12	134 ± 7
W+jets	560 ± 111	224 ± 72	140 ± 58	46 ± 33
Rare	135 ± 6	51 ± 4	21 ± 2	9 ± 1
Total	5774 ± 119	1728 ± 76	617 ± 59	207 ± 34
Data	4822	1314	405	134
Muon+Electron Combined				
$t\bar{t} \to \ell\ell$	1034 ± 18	291 ± 9	92 ± 5	33 ± 3
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	9534 ± 57	2715 ± 30	836 ± 17	283 ± 10
W+jets	891 ± 142	298 ± 82	182 ± 65	88 ± 44
Rare	287 ± 9	106 ± 5	41 ± 3	17 ± 2
Total	11747 ± 154	3410 ± 88	1149 ± 67	421 ± 45
Data	10683	2922	931	326

Table 19: Preveto MC and data yields in M_{T} peak region. The n-jets k-factors have been applied to the $t\bar{t} \to \ell\ell$. The uncertainties are statistical only.

Sample	SRA	SRB	SRC	SRD
Muon				
$t\bar{t} \to \ell\ell$	196 ± 8	53 ± 4	16 ± 2	7 ± 1
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	4459 ± 39	1275 ± 21	389 ± 11	136 ± 7
W+jets	332 ± 88	74 ± 38	42 ± 30	42 ± 30
Rare	109 ± 5	38 ± 3	13 ± 2	5 ± 1
Total	5096 ± 97	1440 ± 44	460 ± 32	189 ± 31
Data	4842	1343	447	164
Electron				
$t\bar{t} \to \ell\ell$	200 ± 8	62 ± 4	24 ± 3	8 ± 2
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	4117 ± 37	1163 ± 20	367 ± 11	121 ± 6
W+jets	541 ± 110	224 ± 72	140 ± 58	46 ± 33
Rare	96 ± 5	37 ± 3	15 ± 2	7 ± 1
Total	4954 ± 117	1486 ± 75	546 ± 59	182 ± 33
Data	3980	1086	345	107
Muon+Electron Combined				
$t\bar{t} \to \ell\ell$	396 ± 11	115 ± 6	40 ± 4	14 ± 2
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	8576 ± 54	2438 ± 29	756 ± 16	256 ± 9
W+jets	872 ± 141	298 ± 82	182 ± 65	88 ± 44
Rare	205 ± 7	75 ± 4	28 ± 3	12 ± 2
Total	10050 ± 152	2926 ± 87	1006 ± 67	371 ± 45
Data	8822	2429	792	271

Table 20: MC and data yields in M_{T} peak region after full selection. The n-jets k-factors have been applied to the $t\bar{t} \to \ell\ell$. The uncertainties are statistical only.

Sample	SRA	SRB	SRC	SRD
Muon				
$t\bar{t} \to \ell\ell$	337 ± 10	198 ± 8	71 ± 5	23 ± 3
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	52 ± 4	39 ± 3	9 ± 2	3 ± 1
W+jets	29 ± 26	29 ± 26	29 ± 26	26 ± 26
Rare	33 ± 3	22 ± 2	9 ± 1	4 ± 1
Total	452 ± 28	288 ± 27	118 ± 26	56 ± 26
Electron				
$t\bar{t} \to \ell\ell$	312 ± 10	180 ± 7	53 ± 4	17 ± 2
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	59 ± 4	32 ± 3	6 ± 1	1 ± 1
W+jets	29 ± 24	28 ± 24	0 ± 0	0 ± 0
Rare	30 ± 2	18 ± 2	8 ± 1	3 ± 1
Total	429 ± 27	258 ± 25	67 ± 4	21 ± 2
Muon+Electron Combined				
$t\bar{t} \to \ell\ell$	649 ± 14	378 ± 11	125 ± 6	40 ± 4
$t\bar{t} \to \ell +$ jets & single top (1ℓ)	111 ± 6	71 ± 5	15 ± 2	4 ± 1
W+jets	58 ± 35	57 ± 35	29 ± 26	26 ± 26
Rare	63 ± 3	40 ± 3	17 ± 2	7 ± 1
Total	881 ± 39	545 ± 37	185 ± 27	77 ± 26

Table 21: MC yields in M_{T} tail region after full selection. The n-jets k-factors have been applied to the $t\bar{t} \to \ell\ell$. The uncertainties are statistical only. Note these values are only used for the rare backgrounds prediction.

Sample	SRA	SRB	SRC	SRD
Muon pre-veto M_{T}-SF	0.98 ± 0.02	0.96 ± 0.04	0.99 ± 0.08	0.90 ± 0.15
Muon post-veto M_{T}-SF	0.95 ± 0.01	0.93 ± 0.03	0.97 ± 0.05	0.86 ± 0.07
Muon veto M_{T}-SF	0.97 ± 0.01	0.97 ± 0.01	0.98 ± 0.02	0.96 ± 0.04
Electron pre-veto M_{T}-SF	0.83 ± 0.02	0.75 ± 0.04	0.64 ± 0.07	0.63 ± 0.12
Electron post-veto M_{T}-SF	0.80 ± 0.01	0.72 ± 0.02	0.62 ± 0.04	0.57 ± 0.06
Electron veto M_{T}-SF	0.96 ± 0.01	0.96 ± 0.02	0.96 ± 0.03	0.90 ± 0.05

Table 22: M_{T} peak Data/MC scale factors. The pre-veto SFs are applied to the $t\bar{t} \to \ell\ell$ sample, while the post-veto SFs are applied to the single lepton samples. The veto SF is shown for comparison across channels. The raw MC is used for backgrounds from rare processes. The uncertainties are statistical only.

Sample	SRA	SRB	SRC	SRD
Muon TTP SL Top	0.011	0.025	0.020	0.015
Muon TTP W+Jets	0.039	0.021	0.038	0.085
Electron TTP SL Top	0.012	0.025	0.018	0.011
Electron TTP W+Jets	0.035	0.020	0.000	0.000

Table 23: Ratio of MC events in the M_{T}-tail over events in the M_{T}-peak for $t\bar{t} \to \ell +$ jets (also used for 1-lepton single top) and W+jets. These are derived before applying the b-tagging requirement. ADD STAT. UNCERTAINTIES

- Renormalization and factorization scale: The alternative samples correspond to variations in the scale $\times 2$ and $\times 0.5$. The nominal value for the scale used is $Q^2 = m^2_{\mathrm{top}} + \sum_{\mathrm{jets}} p^2_{\mathrm{T}}$.
- Alternative generators: Samples produced with different generators include MC@NLO and Powheg (NLO generators) and Pythia (LO). It may also be noted that MC@NLO uses Herwig6 for the hadronisation, while POWHEG uses Pythia6.

- Modeling of taus: The alternative sample does not include Tauola and is otherwise identical to the Powheg sample.

- The PDF uncertainty is estimated following the PDF4LHC recommendations[CITE]. The events are reweighted using alternative PDF sets for CT10 and MSTW2008 and the uncertainties for each are derived using the alternative eigenvector variations and the "master equation". In addition, the NNPDF2.1 set with 100 replicas. The central value is determined from the mean and the uncertainty is derived from the 1σ range. The overall uncertainty is derived from the envelope of the alternative predictions and their uncertainties.

7.1.1 Isolated Track Veto: Tag and Probe Studies

In this section we compare the performance of the isolated track veto in data and MC using tag-and-probe studies with samples of Z→ee and Z→ $\mu\mu$. The purpose of these studies is to demonstrate that the efficiency to satisfy the isolated track veto requirements is well-reproduced in the MC, since if this were not the case we would need to apply a data-to-MC scale factor in order to correctly predict the $t\bar{t} \to \ell\ell$ background. This study addresses possible data vs. MC discrepancies for the **efficiency** to identify (and reject) events with a second **genuine** lepton (e, μ, or τ →1-prong). It does not address possible data vs. MC discrepancies in the fake rate for rejecting events without a second genuine lepton; this is handled separately in the top normalization procedure by scaling the $t\bar{t} \to \ell +$jets contribution to match the data in the $M_{\rm T}$ peak after applying the isolated track veto. Furthermore, we test the data and MC isolated track veto efficiencies for electrons and muons since we are using a Z tag-and-probe technique, but we do not directly test the performance for hadronic tracks from τ decays. The performance for hadronic τ decay products may differ from that of electrons and muons for two reasons. First, the τ may decay to a hadronic track plus one or two π^0's, which may decay to $\gamma\gamma$ followed by a photon conversion. As shown in Figure 17, the isolation distribution for charged tracks from τ decays that are not produced in association with π^0s are consistent with that from es and μs. Since events from single prong τ decays produced in association with π^0s comprise a small fraction of the total sample, and since the kinematics of τ, π^0 and $\gamma \to e^+e^-$ decays are well-understood, we currently demonstrate that the isolation is well-reproduced for electrons and muons only. Second, hadronic tracks may undergo nuclear interactions and hence their tracks may not be reconstructed. As discussed above, independent studies show that the MC reproduces the hadronic tracking efficiency within 4%, leading to a total background uncertainty of less than 0.5% (after taking into account the fraction of the total background due to hadronic τ decays with $p_{\rm T} > 10$ GeV tracks), and we hence regard this effect as neglgigible.

The tag-and-probe studies are performed in the full 2011 data sample, and compared with the DYJets madgraph sample. All events must contain a tag-probe pair (details below) with opposite-sign and satisfying the Z mass requirement 76–106 GeV. We compare the distributions of absolute track isolation for probe electrons/muons in data vs. MC. The contributions to this isolation sum are from ambient energy in the event from underlying event, pile-up and jet activitiy, and hence do not depend on the $p_{\rm T}$ of the probe lepton. We therefore restrict the probe $p_{\rm T}$ to be > 30 GeV in order to suppress fake backgrounds with steeply-falling $p_{\rm T}$ spectra. To suppress non-Z backgrounds (in particular $t\bar{t}$) we require $E_{\rm T}^{\rm miss} < 30$ GeV and 0 b-tagged events. The specific criteria for tags and probes for electrons and muons are:

- Electrons
  - Tag criteria
    - Electron passes full analysis ID/iso selection
    - $p_{\rm T} > 30$ GeV, $|\eta| < 2.5$
    - Matched to 1 of the 2 electron tag-and-probe triggers
      - `HLT_Ele17_CaloIdVT_CaloIsoVT_TrkIdT_TrkIsoVT_SC8_Mass30_v*`
      - `HLT_Ele17_CaloIdVT_CaloIsoVT_TrkIdT_TrkIsoVT_Ele8_Mass30_v*`
  - Probe criteria
    - Electron passes full analysis ID selection
    - $p_{\rm T} > 30$ GeV

- Muons
  - Tag criteria
    - Muon passes full analysis ID/iso selection
    - $p_{\mathrm{T}} > 30$ GeV, $|\eta| < 2.1$
    - Matched to 1 of the 2 electron tag-and-probe triggers
      - `HLT_IsoMu30_v*`
      - `HLT_IsoMu30_eta2p1_v*`
  - Probe criteria
    - Muon passes full analysis ID selection
    - $p_{\mathrm{T}} > 30$ GeV

The absolute track isolation distributions for passing probes are displayed in Fig. 14. In general we observe good agreement between data and MC. To be more quantitative, we compare the data vs. MC efficiencies to satisfy absolute track isolation requirements varying from > 1 GeV to > 5 GeV, as summarized in Table 24. In the $\geq$0 and $\geq$1 jet bins where the efficiencies can be tested with statistical precision, the data and MC efficiencies agree within 7%, and we apply this as a systematic uncertainty on the isolated track veto efficiency. For the higher jet multiplicity bins the statistical precision decreases, but we do not observe any evidence for a data vs. MC discrepancy in the isolated track veto efficiency.

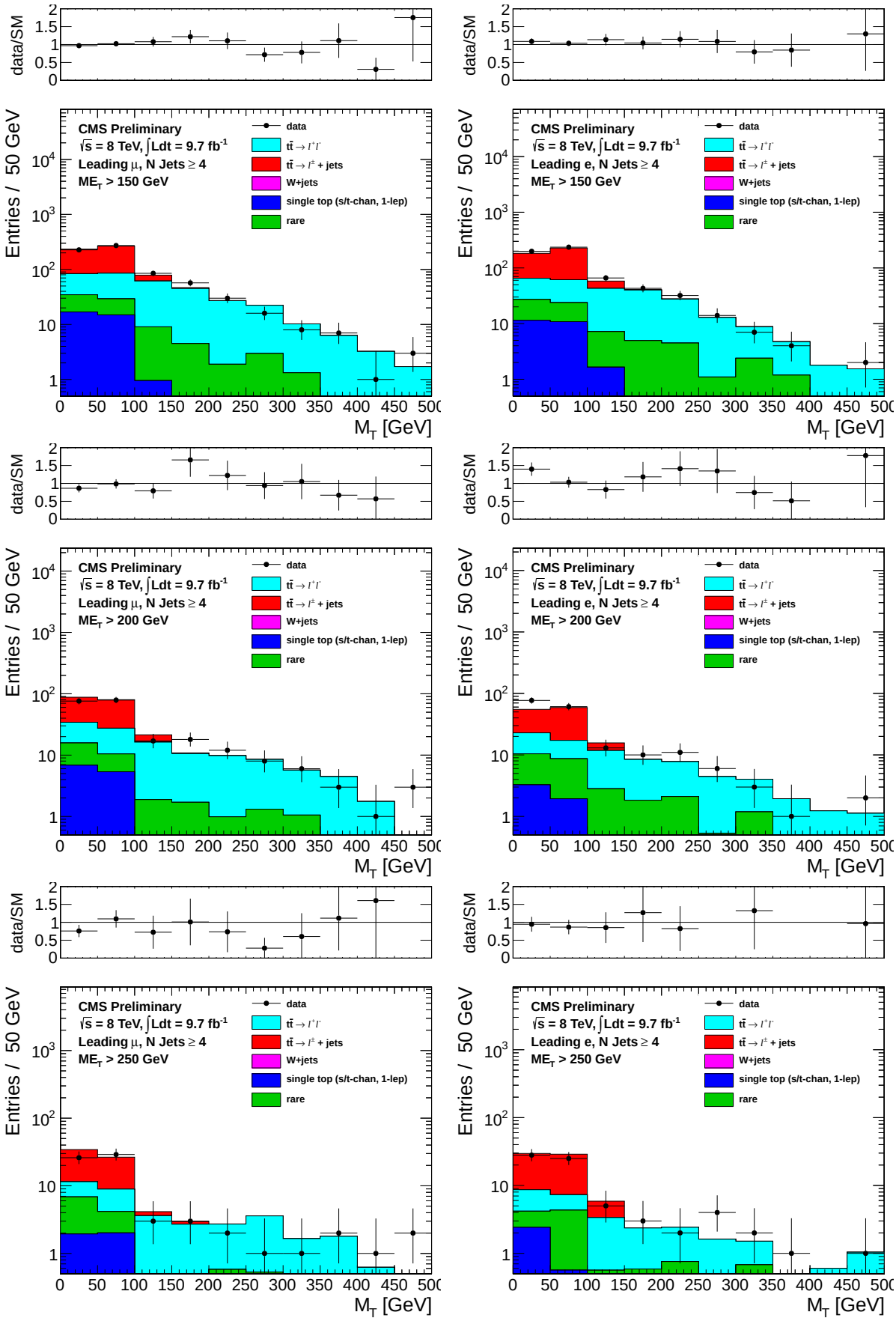


Figure 12: Comparison of the M_{T} distribution in data vs. MC for events with a leading muon (left) and leading electron (right) satisfying the requirements of CR5. The $E_{\mathrm{T}}^{\mathrm{miss}}$ requirements used are 150 GeV (top), 200 GeV (middle) and 250 GeV (bottom).

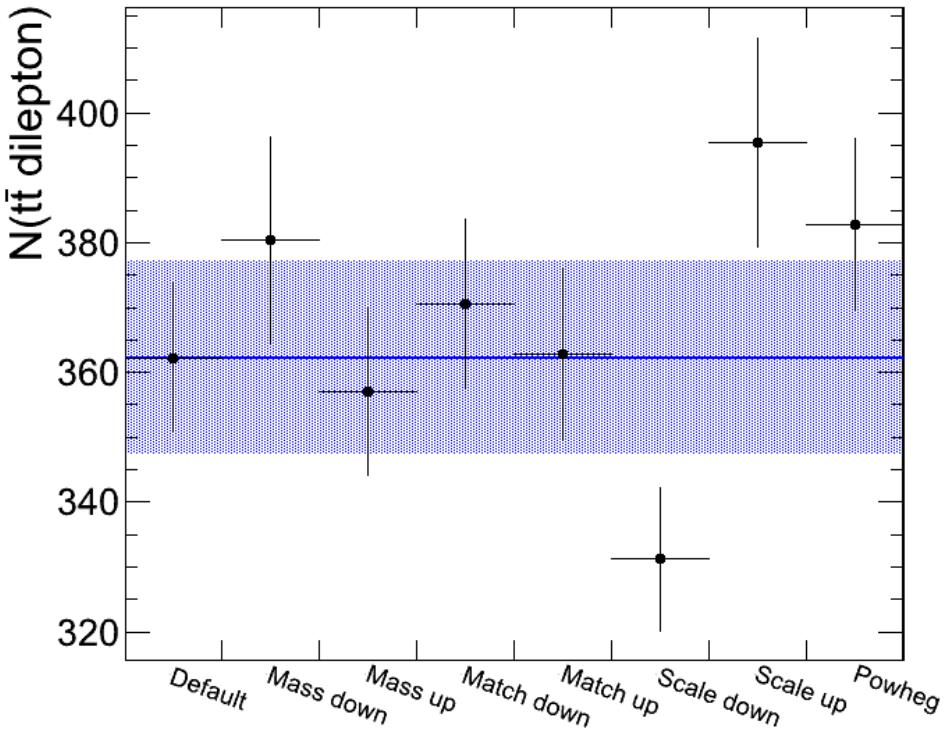


Figure 13: Comparison of the $t\bar{t} \to \ell\ell$ central prediction with those using alternative MC samples. The blue band corresponds to the total statistical error for all data and MC samples. The alternative sample predictions are indicated by the datapoints. The uncertainties on the alternative predictions correspond to the uncorrelated statistical uncertainty from the size of the alternative sample only.

Figure 14: Comparison of the absolute track isolation in data vs. MC for electrons (left) and muons (right) for events with the N_{jets} requirement varied from $N_{\text{jets}} \geq 0$ to $N_{\text{jets}} \geq 4$.

Table 24: Comparison of the data vs. MC efficiencies to satisfy the indicated requirements on the absolute track isolation, and the ratio of these two efficiencies. Results are indicated separately for electrons and muons and for various jet multiplicity requirements.

e + ≥0 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.088 ± 0.0003	0.030 ± 0.0002	0.013 ± 0.0001	0.007 ± 0.0001	0.005 ± 0.0001
mc	0.087 ± 0.0001	0.030 ± 0.0001	0.014 ± 0.0001	0.008 ± 0.0000	0.005 ± 0.0000
data/mc	1.01 ± 0.00	0.99 ± 0.01	0.97 ± 0.01	0.95 ± 0.01	0.93 ± 0.01
μ + ≥0 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.087 ± 0.0002	0.031 ± 0.0001	0.015 ± 0.0001	0.008 ± 0.0001	0.005 ± 0.0001
mc	0.085 ± 0.0001	0.030 ± 0.0001	0.014 ± 0.0000	0.008 ± 0.0000	0.005 ± 0.0000
data/mc	1.02 ± 0.00	1.06 ± 0.00	1.06 ± 0.01	1.03 ± 0.01	1.02 ± 0.01
e + ≥1 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.099 ± 0.0008	0.038 ± 0.0005	0.019 ± 0.0004	0.011 ± 0.0003	0.008 ± 0.0002
mc	0.100 ± 0.0004	0.038 ± 0.0003	0.019 ± 0.0002	0.012 ± 0.0002	0.008 ± 0.0001
data/mc	0.99 ± 0.01	1.00 ± 0.02	0.99 ± 0.02	0.98 ± 0.03	0.97 ± 0.03
μ + ≥1 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.100 ± 0.0006	0.041 ± 0.0004	0.022 ± 0.0003	0.014 ± 0.0002	0.010 ± 0.0002
mc	0.099 ± 0.0004	0.039 ± 0.0002	0.020 ± 0.0002	0.013 ± 0.0001	0.009 ± 0.0001
data/mc	1.01 ± 0.01	1.05 ± 0.01	1.05 ± 0.02	1.06 ± 0.02	1.06 ± 0.03
e + ≥2 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.105 ± 0.0020	0.042 ± 0.0013	0.021 ± 0.0009	0.013 ± 0.0007	0.009 ± 0.0006
mc	0.109 ± 0.0011	0.043 ± 0.0007	0.021 ± 0.0005	0.013 ± 0.0004	0.009 ± 0.0003
data/mc	0.96 ± 0.02	0.97 ± 0.03	1.00 ± 0.05	1.01 ± 0.06	0.97 ± 0.08
μ + ≥2 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.106 ± 0.0016	0.045 ± 0.0011	0.025 ± 0.0008	0.016 ± 0.0007	0.012 ± 0.0006
mc	0.108 ± 0.0009	0.044 ± 0.0006	0.024 ± 0.0004	0.016 ± 0.0004	0.011 ± 0.0003
data/mc	0.98 ± 0.02	1.04 ± 0.03	1.04 ± 0.04	1.04 ± 0.05	1.06 ± 0.06
e + ≥3 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.117 ± 0.0055	0.051 ± 0.0038	0.029 ± 0.0029	0.018 ± 0.0023	0.012 ± 0.0019
mc	0.120 ± 0.0031	0.052 ± 0.0021	0.027 ± 0.0015	0.018 ± 0.0012	0.013 ± 0.0011
data/mc	0.97 ± 0.05	0.99 ± 0.08	1.10 ± 0.13	1.03 ± 0.15	0.91 ± 0.16
μ + ≥3 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.111 ± 0.0044	0.050 ± 0.0030	0.029 ± 0.0024	0.019 ± 0.0019	0.014 ± 0.0017
mc	0.115 ± 0.0025	0.051 ± 0.0017	0.030 ± 0.0013	0.020 ± 0.0011	0.015 ± 0.0009
data/mc	0.97 ± 0.04	0.97 ± 0.07	0.95 ± 0.09	0.97 ± 0.11	0.99 ± 0.13
e + ≥4 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.113 ± 0.0148	0.048 ± 0.0100	0.033 ± 0.0083	0.020 ± 0.0065	0.017 ± 0.0062
mc	0.146 ± 0.0092	0.064 ± 0.0064	0.034 ± 0.0048	0.024 ± 0.0040	0.021 ± 0.0037
data/mc	0.78 ± 0.11	0.74 ± 0.17	0.96 ± 0.28	0.82 ± 0.30	0.85 ± 0.34
μ + ≥4 jets	> 1 GeV	> 2 GeV	> 3 GeV	> 4 GeV	> 5 GeV
data	0.130 ± 0.0128	0.052 ± 0.0085	0.028 ± 0.0063	0.019 ± 0.0052	0.019 ± 0.0052
mc	0.105 ± 0.0064	0.045 ± 0.0043	0.027 ± 0.0034	0.019 ± 0.0028	0.014 ± 0.0024
data/mc	1.23 ± 0.14	1.18 ± 0.22	1.03 ± 0.27	1.01 ± 0.32	1.37 ± 0.45

fix me: What you have written in the next paragraph does not explain how ϵ_{fake} is measured. Why not measure ϵ_{fake} in the b-veto region?

A measurement of the ϵ_{fake} in data is non-trivial. However, it is possible to correct for differences in the ϵ_{fake} between data and MC by applying an additional scale factor for the single lepton background alone, using the sample in the M_{T} peak region. This scale factor is determined after applying the isolated track veto and after subtracting the $\mathrm{t\bar{t}} \to \ell\ell$ component, corrected for the isolation efficiency derived previously. As shown in Figure 15, the efficiency for selecting an isolated track in single lepton events is independent of M_{T} so the use of an overall scale factor is justified to estimate the contribution in the M_{T} tail.

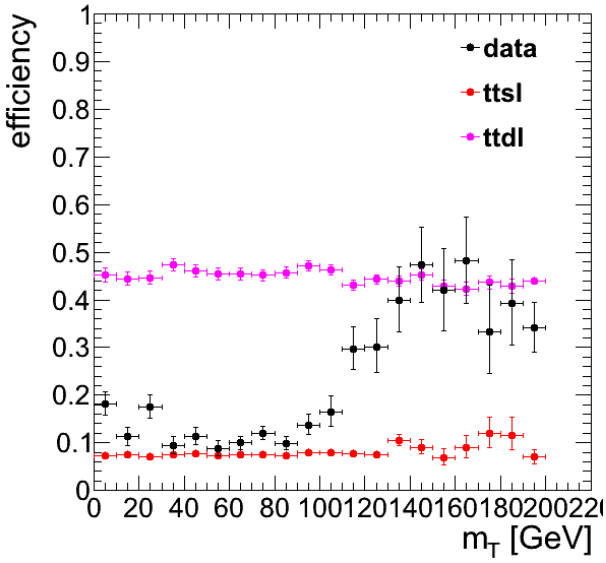


Figure 15: Efficiency for selecting an isolated track comparing single lepton $t\bar{t} \to \ell + \text{jets}$ and dilepton $t\bar{t} \to \ell\ell$ events in MC and data as a function of M_{T}. The efficiencies in $t\bar{t} \to \ell + \text{jets}$ and $t\bar{t} \to \ell\ell$ exhibit no dependence on M_{T} while the data ranges between the two. This behavior is expected since the low M_{T} region is predominantly $t\bar{t} \to \ell + \text{jets}$, while the high M_{T} region contains mostly $t\bar{t} \to \ell\ell$ events.

8 Results

9 Conclusion

References

[1] https://hypernews.cern.ch/HyperNews/CMS/get/SUS-12-007/32.html

[2] arXiv:1204.3774v1 [hep-ex]

[3] D. Barge, CMS AN-2011/464

[4] CMS Collaboration, *Search for supersymmetry in events with a Z boson, jets and momentum imbalance* PAS SUS-11-021

[5] CMS Collaboration, *"Measurement of the b-tagging efficiency using* $\mathrm{t\bar{t}}$ *events"*, PAS BTV-11-003, in preparation.

[6] CMS Collaboration, *Measurement of Tracking Efciency*, PAS TRK-10-00

[7] CMS Collaboration, *Performance of b-jet identification in CMS*, PAS BTV-11-001

[8] M. Narain for BTV POG, https://indico.cern.ch/getFile.py/access?contribId=0&resId=1&materialId=slides&confId=163892

[9] arXiv:1103.1348v1, D. Barge *at al.*, CMS AN-CMS2011/269.

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

[11] V. Pavlunin, CMS AN-2009/125

[12] A reference to the top paper, once it is submitted. Also D. Barge *et al.*, AN-CMS2010/258.

[13] Changes to the selection for the 38x CMSSW release are given in https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopDileptonRefAnalysis2010Pass5.

[14] https://twiki.cern.ch/twiki/bin/viewauth/CMS/SimpleCutBasedEleID

[15] https://twiki.cern.ch/twiki/bin/viewauth/CMS/EgammaWorkingPointsv3

[16] D. Barge *at al.*, AN-CMS2009/159.

[17] B. Mangano *et al.*, AN-CMS2010/283.

[18] https://twiki.cern.ch/twiki/bin/viewauth/CMS/CrossSections_3XSeries, https://twiki.cern.ch/twiki/bin/view/CMS/ProductionSpring2011

[19] CMS Collaboration, "Measurement of CMS luminosity", *CMS-PAS* **EWK-10-004** (2010).

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

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

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

[23] L. Bauerdick *at al.*, AN-CMS2011/155.

[24] CMS-PAS-JME-10-010.

[25] arXiv:1103.6083v1, J. T. Ruderman, D. Shih

[26] H. Haber, G. Kane, Phys. Reports 117, Nos. 2-4 (1985) 75-263.

[27] http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/UserCode/SusyAnalysis/SLHAFILES/TChiwz/

[28] http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/UserCode/SusyAnalysis/SLHAFILES/TChizz/

A Performance of the Isolation Requirement

The last requirement used in the analysis is an isolated track veto. This selection criteria rejects events containing a track of $p_{\mathrm{T}} > 10$ GeV with relative track isolation $\sum p_{\mathrm{T}}/p_{\mathrm{T}}(trk)$ in a cone of size $R = 0.3 < 0.1$. It may be noted that only tracks consistent with the vertex with highest $\sum p_{\mathrm{T}}^2$ are considered in order to reduce the impact of spurious tracks, for example from pileup interactions. This requirement has very good performance. Figure 16 shows the efficiency for rejecting dilepton events compared to the efficiency for selecting single lepton events for various cone sizes and cut values. The chosen working point provides a signal efficiency of $\epsilon(sig) = 92\%$ for a background rejection of $\epsilon(bkg) = 53\%$ in MC. With "signal" ("background") we are referring to $\mathrm{t\bar{t}} \to \ell + \mathrm{jets}$ ($\mathrm{t\bar{t}} \to \ell\ell$).

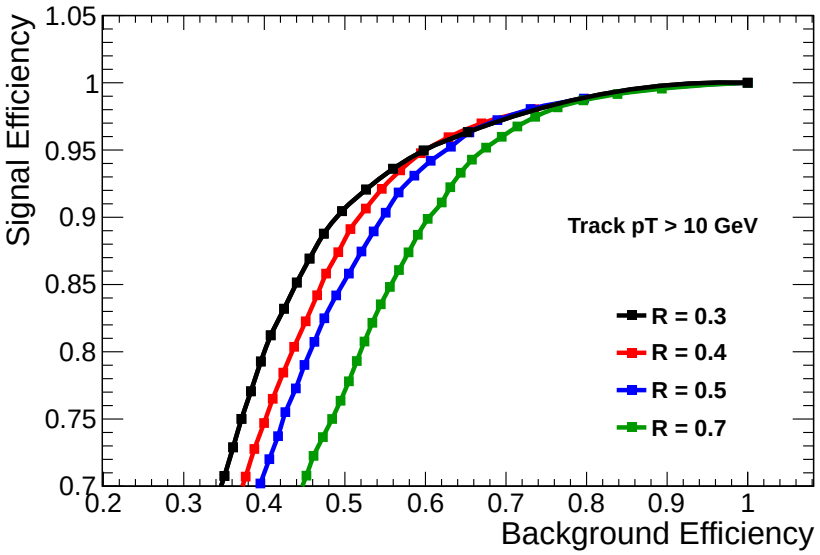


Figure 16: Comparison of the performance in terms of signal (single lepton events) efficiency and background (dilepton events) rejection for various cone sizes and cut values. The current isolation requirement uses a cone of size $\Delta R = 0.3$ and a cut value of 0.1, corresponding to $\epsilon(sig) = 92\%$ for $\epsilon(bkg) = 53\%$. ADD ARROW OR LINE TO INDICATE WORKING POINT.

It should be emphasized that the isolated track veto has a different impact on the samples with a single lepton (mainly $\mathrm{t\bar{t}} \to \ell + \mathrm{jets}$ and W+jets) and that with two leptons (mainly $\mathrm{t\bar{t}} \to \ell\ell$). For the dilepton background, the veto rejects events which have a genuine second lepton. Thus the performance may be understood as an efficiency $\epsilon_{iso\ trk}$ to identify the isolated track. In the case of the single lepton background, the veto rejects events which do not have a genuine second lepton, but rather which contain a "fake" isolated track. The isolated track veto thus effectively scales the single lepton sample by (1-ϵ_{fake}), where ϵ_{fake} is the probability to identify an isolated track with $p_{\mathrm{T}} > 10$ GeVin events which contain no genuine second lepton. It is thus necessary to study the isolated track efficiency $\epsilon(trk)$ and ϵ_{fake} in order to fully characterize the veto performance.

The veto efficiency for dilepton events is calculated using the tag and probe method in Z events. A good lepton satisfying the full ID and isolation criteria and matched to a trigger object serves as the tag. The probe is defined as a track with $p_{\mathrm{T}} > 10$ GeV that has opposite charge to the tag and has an invariant mass with the probe consistent with the Z mass.

Fix me: fkw does not understand why you refer to $p_{\mathrm{T}} > 10$ GeVhere, given that in the very next paragraph you state that this is measured via the absolute track isolation, implying, but not explicitly stating, that a much higher p_{T} threshold is used to get a clean Z signal. ???

The variable used to study the performance of the veto is the absolute track isolation, since it removes the dependence of the isolation variable on the p_{T} of the object under consideration. This is particularly useful because the underlying p_{T} distribution is different for second leptons in $\mathrm{t\bar{t}} \to \ell\ell$ events compared to Z events, particularly due to the presence of τs that have softer decay products. As shown in Figure 17,

the absolute isolation is consistent between $Z+4$ jet events and $t\bar{t} \to \ell\ell$ events, including leptons from W and τ decays. This supports the notion that the isolation, defined as the energy surrounding the object under consideration, depends only on the environment of the object and not on the object itself. The isolation is thus sensitive to the ambient pileup and jet activity in the event, which is uncorrelated with the lepton p_{T}. It is thus justified to use tag and probe in $Z+4$ jet events, where the jet activity is similar to $t\bar{t} \to \ell\ell$ events in our $N_{\mathrm{jets}} > 4$ signal region, in order to estimate the performance of the isolation requirement for the various leptonic categories of $t\bar{t} \to \ell\ell$ events.

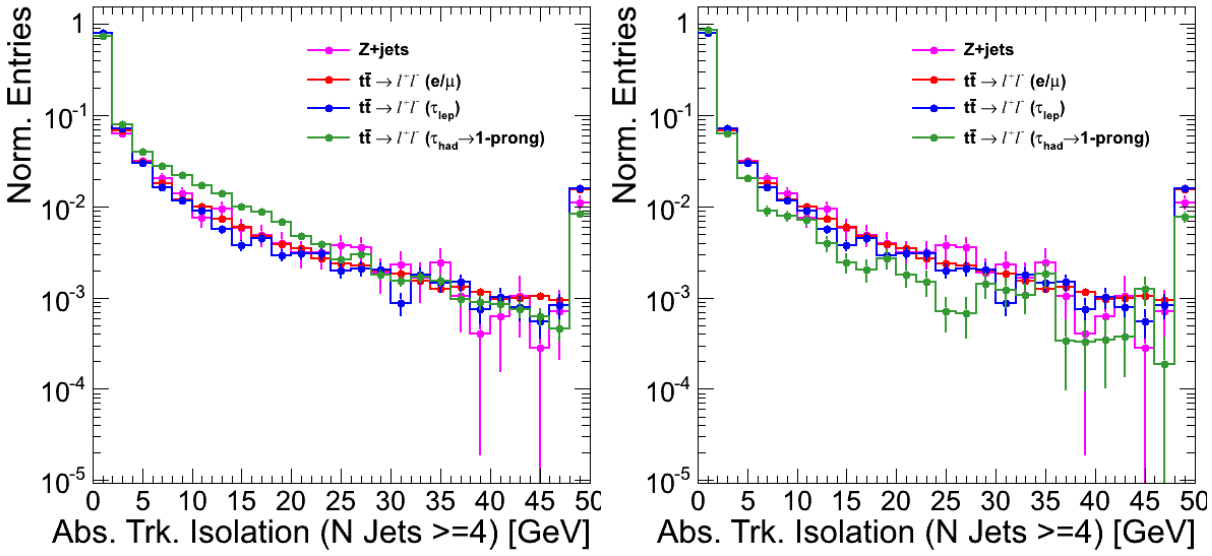


Figure 17: Comparison of absolute track isolation for track probes in $Z+4$ jet and $t\bar{t} \to \ell\ell$ events for different lepton types. The isolation variables agree across samples, except for single prong τs, that tend to be slightly less isolated (left). The agreement across isolation distributions is recovered after removing single prong τ events produced in association with π^0s from the sample (right).