CMS Physics Analysis Summary

Contact: cms-pag-conveners-top@cern.ch 2010/07/24

Selection of Top-Like Events in the Dilepton and Lepton-plus-Jets Channels in Early 7 TeV Data

The CMS Collaboration

Abstract

We present studies of the selection of events consistent with top quark pair production in the first 78 nb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV recorded by the CMS detector at the LHC. Results are presented for the lepton+jets and dilepton channels. In this initial small sample we do not expect a significant contribution from top-quark production; however, predicted background yields are compared to the overall yield of events collected in this early data. Kinematic comparisons demonstrate good agreement between simulated events and current data, putting the prospect of measurements of the top-quark pair-production cross section on firm ground.

1 Introduction

With the advent of the LHC era the CMS experiment [1] has the opportunity to observe the top quark for the first time away from the Tevatron. The top quark plays a unique role in our understanding of the fundamental world; consequently, studies of the top quark are a crucial component of the CMS physics program. The top quark is the most massive of the known standard model particles at $m_t = 173.1 \pm 0.6(\text{stat}) \pm 1.1(\text{syst})$ GeV/c^2 [2], its mass exceeding its closest fermionic relative, the b quark, by a factor of $\sim$35 and exceeding the most massive known boson, the Z, by nearly a factor of two. The top quark decays rapidly, long before having the chance to form a bound state hadron. Hence, the study of top-quark daughters affords a direct glimpse at the properties of the parent quark itself, allowing measurements of its mass, spin, charge and other properties. Additionally, since the Higgs boson of the standard model couples to fermions in strength proportional to the fermion's mass, the Higgs coupling to the top quark is large. Because of this, detailed study of the properties of the top quark can provide constraints on the yet-to-be observed Higgs [3]. Further, in light of its large mass, it is hypothesized that the top quark could play a role in electroweak symmetry breaking and the generation of particle masses in alternatives to the Higgs mechanism. Finally, several signatures of new physics accessible at the LHC either suffer from top-quark production as a significant background or contain top quarks themselves.

The planned CMS physics program contains an extensive collection of precision measurements of the properties of the top quark. However, in the early days of LHC operation, a primary goal is the first observation of top-quark pair production and the measurement of its production cross section. In this document we present initial results on the observation strategies for top-quark pair production in the first up to 78 ± 9 nb^{-1} [4] of pp collisions from the LHC at $\sqrt{s} = 7$ TeV.

At the LHC, the top quark is expected to be produced primarily via the strong interaction in $t\bar{t}$ pairs. Under the design conditions of the LHC, top-quark pair production proceeds predominantly through a gluon fusion process, which accounts for $\sim$90% of the produced top quark pairs. The remaining 10% come from quark-antiquark annihilation. A similar relationship exists for top-quark pair production at $\sqrt{s} = 7$ TeV, where the predicted NLO top quark pair production cross section is approximately 158 pb [5, 6]. Electroweak production of single top quarks is also possible; however, the expected single-top production cross section is significantly smaller than that of top-quark pairs and is not the focus of the early studies discussed here.

Within the standard model, the top quark is expected to decay via the weak process $t \to Wb$ nearly 100% of the time; subsequently top-quark pair events are characterized according to the decay of the W bosons. Two channels are the focus of the studies discussed below: the dilepton channel, in which both W bosons decay leptonically to either an electron or a muon and its associated neutrino; and the lepton+jets channel, in which one of the W bosons decays leptonically while the other decays to quarks which subsequently hadronize, resulting in jets of charged and neutral particles. A third channel, the all-hadronic channel in which both W bosons decay to quarks, will contribute to the suite of top-quark physics analyses in larger integrated luminosity data samples but is not discussed here.

In both the dilepton and lepton+jets channels, events are characterized by one or more electrons or muons with high transverse momentum p_T. These charged leptons provide the opportunity for efficient event triggering and the effective suppression of QCD multijet background. Given the large branching fraction of the tau lepton (τ) to hadronic final states, the decay $W \to \tau\nu$ requires unique treatment and is not explicitly considered as a separate channel here. Events

with leptonic decays of the τ will satisfy the event selection criteria and are counted as a small additional source of signal. Below we further distinguish the analyses in the dilepton and lepton+jets channels: the dilepton result has three modes, ee, $e\mu$, $\mu\mu$; and the lepton+jets channel has two modes, e+jets and μ+jets.

In addition to the charged leptons, the signature in both the dilepton and lepton+jets channels is distinguished by significant missing transverse energy $\not\!\!E_{\mathrm{T}}$, the manifestation of the weakly interacting neutrino in our event record. A requirement that an event possess a minimum amount of missing transverse energy is useful in reducing dilepton backgrounds from Z boson and Drell-Yan production, and further reduces the QCD multijet background in the lepton+jets channel.

Finally the $t\bar{t}$ signatures contain two or more particle jets with high transverse momentum, among which at least two originate from b quarks. A requirement on the minimum p_T of the jets reduces contributions to our selected sample from W/Z+jets production and further suppresses low-p_T multijet events. Furthermore, one can require that the selected jets be consistent with originating from b-quark production to increase the $t\bar{t}$ signal purity in the selected sample. Various techniques have been deployed at CMS for the identification of b-jets. These so-called b-tagging algorithms exploit the long-lifetime, heavy mass and unique decay properties of B hadrons to differentiate these jets from those that originate from other flavors.

By considering only the electronic and muonic decays of the W in the dilepton and lepton+jets channels, the expected $t\bar{t}$ production cross section of 158 pb must be reduced by a factor of 35% to take into account the branching ratios of the W boson. Consequently, in just 78 nb^{-1} of data, we do not expect to observe any top-quark pair events. We therefore focus on the study of the standard model backgrounds and use the data to validate our background estimation techniques.

This document is structured as follows. Section 2 describes the data and simulation samples and Section 3 the selection criteria. Background estimation techniques are presented in Sections 4 and 5, and the results are compared to the data in Section 6. We finish by presenting a few candidate events and summarizing the future outlook.

2 Data and Simulated Samples

Top-quark pair signal events in the data are collected using an online event trigger that requires the presence of a high p_T electron or muon. In the studies presented in this document, two types of triggers define our data sample: a muon trigger that accepts events containing any muon with $p_T > 9$ GeV/c; and a trigger designed to accept events with depositions in the electromagnetic calorimeter with $E_T > 15$ GeV. In both the μ+jets and dilepton analyses, reconstructed muon candidates are required to have $p_T >$20 GeV/c in the region where the trigger is fully efficient and the efficiency is stable. The electromagnetic deposition trigger was designed to collect photon events; however since no track veto is applied to the trigger photons, this trigger accepts events with electron candidates as well. We utilize this trigger in early data because its selection criteria simplifies the early trigger efficiency measurement.

The efficiency for these triggers will eventually be measured in data using leptonic decays of Z bosons. However, such a measurement requires a sufficiently large sample of Z bosons which we have not yet accumulated. Thus for initial studies, the trigger efficiency from the simulation is used as a temporary estimate.

In addition to the signal trigger samples, the data-driven estimate of the background from

non-W/Z leptons in the dilepton analysis utilizes jet triggers, with various jet E_T thresholds: 6, 10, 15 and 30 GeV. The different trigger samples are necessary to estimate the systematic uncertainty of the method.

Before being used in analysis, data events are required to come from periods in which the CMS detector was fully operational. Additionally, events are vetoed if they are identified as resulting from beam halo interactions or from beam scraping. Finally, events are required to possess at least one well-constructed primary vertex within $|z|$ <15 cm. For the results presented in this note, a data set corresponding to an integrated luminosity of 78 ± 9 nb^{-1} was analyzed in the dilepton and lepton+jets analyses.

Simulated samples of top-quark pair production events are made using the MADGRAPH event generator [7], subsequently processed with PYTHIA [8], and then processed with a full CMS detector simulation based on GEANT4 [9]. Events are generated with up to four additional hard partons.

Various background samples were produced for studies in the dilepton and lepton+jets channels. MADGRAPH is used for $W/Z/\gamma$+jets production and single top. Leptonic tau decays are included in the Drell-Yan samples. Pythia is used to generate QCD events used in the study of the multijet backgrounds.

The following cross sections have been used when normalizing the simulated samples to the luminosity corresponding to the analyzed dataset.

The NLO top-quark pair production cross section has been calculated as $\sigma_{t\bar{t}} = 157.5^{+23.2}_{-24.4}$ pb, using MCFM [5, 6]. The uncertainty in the cross section includes the scale uncertainties, determined by varying the factorization and renormalization scales by a factor 2 and 0.5 around the central scale choice of 172.5 GeV, and the uncertainties from the PDFs and the value of α_S, following the procedures from the MSTW2008 [10], CTEQ6.6 [11], and NNPDF2.0 [12] sets. The uncertainties are then combined according to the PDF4LHC prescriptions [13].

Similarly, the t-channel single top NLO cross-section, multiplied by the branching fraction BR($W \rightarrow l\nu$), has been determined as $\sigma_t = 21.0^{+1.1}_{-1.0}$ pb using MCFM [5, 14–16]. The uncertainty is defined similarly as for top-quark pair production. For both $t\bar{t}$ and single top-quark production, renormalization and factorization scales were set to $m_t = 172.5$ GeV/c^2. The single top-quark associated production (tW) cross section has been set to $\sigma_{tW} = 10.6 \pm 0.8$ pb [15].

The NNLO cross section of the production of W bosons decaying into leptons has been determined as $\sigma_{W\rightarrow l\nu} = 31314 \pm 1558$ pb using FEWZ [17], setting renormalization and factorization scales to $m_W = 80.398$ GeV/c^2. The uncertainty was determined in a similar way as for top-quark pair production. Finally, the Drell-Yan production cross section at NNLO has been calculated using FEWZ as $\sigma_{Z/\gamma^*\rightarrow ll}(m_{ll} > 20\,\mathrm{GeV/c^2}) = 4998 \pm 272$ pb and $\sigma_{Z/\gamma^*\rightarrow ll}(m_{ll} > 50\,\mathrm{GeV/c^2}) = 3048 \pm 132$ pb, respectively, where the scales were set to $m_Z = 91.1876$ GeV/c^2.

3 Event Selection

The selection criteria have been designed to select dilepton and lepton+jets $t\bar{t}$ events in early data, and optimized using simulated samples before the start of 7 TeV data-taking [18–20]. Although the needs of each mode are similar, the requirements vary slightly among the various modes due to the different signal and background conditions. The details of the dilepton and lepton+jets selections are presented below.

3.1 Dilepton Channel

The selection of dilepton events starts with the lepton trigger requirements mentioned above. Just one lepton is required to satisfy the trigger requirements for the event to be considered in the analysis; events in the $e\mu$ final state are required to pass either the muon or electromagnetic trigger requirement. Care is taken to remove duplicate $e\mu$ events that satisfy both triggers.

The muons used in the dilepton analysis are reconstructed [21] using two algorithms, one in which tracks in the silicon detector are matched to compatible signals in the calorimeters and muon systems, and another in which a simultaneous global fit is performed to hits in the silicon tracker and muon system. To be retained for analysis, muon candidates in this dilepton analysis are required to be successfully reconstructed by both algorithms and to have a transverse momentum p_{T} exceeding 20 GeV/c and pseudorapidity $|\eta| < 2.5$. Additionally, the track associated with the muon candidate is required to have a minimum number of hits in the tracker, must be consistent with originating from the beam spot, and have a high quality global fit including a minimum number of hits in the muon detector.

Leptons from the decay of W bosons are expected to be isolated from other event activity. For muon candidates with transverse momentum larger than 20 GeV/c, a cone of $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ is constructed around the muon track and the sum of the track momenta and electromagnetic and hadronic calorimeter energy deposits is calculated, excluding the contribution from the muon candidate itself. If this scalar sum exceeds 15% relative to the value of the muon candidate's momentum, the candidate is considered to be non-isolated and is rejected. The ratio of the scalar sum to the candidate's momentum is called relative isolation; we define it here for frequent later use.

Electron candidates are reconstructed [22] starting from a cluster of energy deposits in the crystal electromagnetic calorimeter, which is then matched to hits in the tracker. The resulting trajectory takes into account the significant energy loss of the electron through bremsstrahlung as it traverses the material of the tracker. Electron candidates are required to have $E_T > 20$ GeV and pseudorapidity $|\eta| < 2.4$. The electron candidate track is required to be consistent with originating from the beam spot. A simple cut-based selection in electron identification variables is applied to the reconstructed candidates that focuses on their shower shape and track-cluster matching; the criteria are optimized in the context of the inclusive $W \to e\nu$ measurement [23] and are designed to maximally reject fake electrons from QCD multijet production while maintaining 90% efficiency for electrons from the decay of W/Z bosons. Electron candidates within $\Delta R < 0.1$ of a tracker-based or globally-fit muon are rejected to remove fakes due to muon bremsstrahlung. In addition, electrons consistent with significant mismeasurement in the electromagnetic calorimeter are rejected. Electrons originating from photon conversions are removed by a dedicated algorithm. Finally, a similar isolation requirement, as applied to muons, is employed for electrons as well, rejecting candidates with more than 15% relative energy in the $\Delta R < 0.3$ isolation cone, excluding in the calculation the candidate's contribution itself.

To be retained for further analysis, a dilepton event is required to contain at least two charged lepton candidates of opposite sign passing all selection criteria. In cases where more than one such pair is found, the combination with the highest scalar sum of transverse momenta is chosen. Events are then assigned to one of the modes ee, $e\mu$ or $\mu\mu$. To reject a contribution from Z-boson decays in the ee and $\mu\mu$ modes, we exclude events with dilepton invariant mass, $M_{\ell\ell}$, near the Z boson mass by requiring $|M_{\ell\ell} - 91 \text{ GeV}/c^2| > 15 \text{ GeV}/c^2$.

The missing transverse energy $\not{E}_{\mathrm{T}}$ due to the escaping neutrinos in the W-boson decays is an

important distinguishing feature of $t\bar{t}$ events. At CMS there are several techniques for calculating $\not\!\!E_{\mathrm{T}}$ [24]; here raw $\not\!\!E_{\mathrm{T}}$ calculated from calorimeter towers is made more accurate through a series of corrections taking into account jet energy measurement scale adjustments, the contribution from the minimally interacting muons and, most importantly, a per-track correction that takes into account the inexact response of the calorimeter. This track correction to calorimeter response shows improved energy resolution especially for low energy charged particles. Neither the dominant background process, Drell-Yan $Z/\gamma^{\star} \rightarrow e^{+}e^{-}$ and $Z/\gamma^{\star} \rightarrow \mu^{+}\mu^{-}$, nor the difficult to model background from QCD multijet fakes, contain a natural source of $\not\!\!E_{\mathrm{T}}$. Hence in the ee and $\mu\mu$ modes, $\not\!\!E_{\mathrm{T}}$>30 GeV is required; in the $e\mu$ mode a looser requirement of $\not\!\!E_{\mathrm{T}}$>20 GeV is in place, relaxed to a lower threshold because of the significantly smaller contribution of Drell-Yan background.

Dilepton $t\bar{t}$ events will have at least two jets of particles from the top-quark decays. Jets are reconstructed using calorimeter information, corrected using input from the tracker [25, 26]. The anti-k_T clustering algorithm [27] with $R = 0.5$ is used for jet clustering. Corrections are applied to the raw jet momenta to establish relative uniform response of the calorimeter in η and absolute uniformity in the variation of calorimeter response as a function of p_{T}. The jet energy scale uncertainty for these track-corrected jets is expected to be ∼5%. Jet candidates are required to have p_{T} >30 GeV/c, $|\eta|$ <2.4 and must not overlap with any electron or muon candidate within $\Delta R < 0.4$. Although signal events contain at least two jets, we retain events with 0 or 1 jet to be used as background-dominated control samples to test the quality of the simulated event sample's representation of the data.

The two jets in dilepton $t\bar{t}$ events both originate from b-quark production. Several algorithms [28] exploit the long lifetime of the b-quark; once created, the resulting B hadron, significantly boosted, travels a macroscopic distance in the lab frame before decaying. The daughters in that decay will produce charged particle tracks inconsistent with originating from the primary pp interaction location. One b-jet identification algorithm in place at CMS examines the signed impact parameter significance ($IP_{sig} = \frac{IP}{\sigma_{IP}}$) of a subset of the high quality tracks in each jet. Tracks are ordered in decreasing IP_{sig} and the value of IP_{sig} of the second highest track is taken as the discriminant for this algorithm. A cut is placed in this discriminant; the working point in use here requires $IP_{sig} > 1.7$ for a jet to be tagged. This working point corresponds to an 81% efficiency for tagging b jets, as measured in simulated top-quark decays, at the expense of a 10% false positive rate, as measured in non-b jets from simulated QCD multijet events. Below we will report the results of the event selection in data both before and after applying the tag requirements.

3.2 Lepton+Jets Channel

Lepton+jets candidate events are triggered with the same muon and electron triggers described in the previous section.

Muon candidates are required to be reconstructed by both the tracker-driven and global fit algorithms [21]. In the μ+jets analysis, muon candidates are required to satisfy the kinematic cuts of $p_{\mathrm{T}} > 20$ GeV/c and $|\eta| < 2.1$. The isolation requirement employed in the μ+jets analysis is slightly tighter compared to the dilepton selection: the relative isolation is required to be smaller than 5 %. The requirements on the minimum number of silicon hits, the consistency with originating from the beam spot and the criteria on the global χ^2 fit are the same as the muon selection in the dilepton analysis.

For the μ+jets analysis, exactly one isolated muon candidate is allowed in a given event; if more than one is found, the event is vetoed. Additional loose muons are sought, with $p_{\mathrm{T}} > 10$ GeV/c

and $|\eta| < 2.5$ and up to a relative isolation of 20 %; if any such additional loose muons are found, the event is rejected. Furthermore, in events with an isolated high-momentum muon candidate, if additional electron candidates with $E_{\mathrm{T}} > 15$ GeV and similar isolation condition as the loose muons are present, the event is discarded as well. These criteria are in place to suppress Z-boson events and to construct a selected sample that is statistically independent from from the dilepton selection.

Electron candidates are reconstructed in the same way as in the dilepton analysis. The electron candidates are subject to the tighter requirement $E_T > 30\,\mathrm{GeV}$ and are accepted within the range $|\eta| < 2.5$. The electron track is required to be consistent with the beam spot. The simple cut-based electron identification criteria is applied here as well; the criteria employed for the e+jets analysis is designed to provide optimal fake-electron rejection while maintaining a 70 % efficiency for electrons from the decays of W and Z bosons. An isolation requirement is applied to the electron candidate, requiring less than 10% relative energy in the candidate's $\Delta R < 0.3$ isolation cone. Finally, a dedicated photon conversion removal is applied to the electron candidates.

In the e+jets channel, exactly one isolated electron candidate is allowed in a given event, otherwise the event is rejected. Events with additional muon candidates with $p_{\mathrm{T}} > 10$ GeV/c and $|\eta| < 2.5$ and up to a relative 20% of the muon's p_{T} in a $\Delta R < 0.3$ isolation cone are vetoed. Finally, to guard against contributions from $Z \to ee$ background, additional loose electron candidates are sought, with $E_T > 20$ GeV and $|\eta| < 2.5$ and up to a relative 100% of the electron's energy in the isolation cone; if any such candidates are found, the di-electron invariant mass is calculated, and if $76 < M_{ee} < 106$ GeV/c^2, the event is rejected.

In these early studies in the lepton+jets signatures, the requirement of missing transverse energy is not applied. This choice will be revisited in future analyses, in which the benefits in multijet background rejection might motivate the implementation of significant $\not\!\!E_{\mathrm{T}}$.

Also in contrast to the dilepton analysis, jet candidates in the lepton+jets analyses are reconstructed using calorimeter information only. The clustering algorithm (anti-k_T, with $R = 0.5$), jet momenta correction scheme, and baseline kinematic selection criteria ($p_{\mathrm{T}} > 30$ GeV/c, $|\eta| < 2.4$) are the same as the dilepton analysis. An overlap veto is also in place as in the lepton+jets selection, requiring any qualifying jet to be well separated from an isolated muon (electron) in the μ+jets (e+jets) channel.

Top-quark pair events in the lepton+jets modes are expected to have four or more jets; in the studies described here we will examine the lower jet multiplicity sample to probe the validity of our background estimation techniques. In $t\bar{t}$ events, two of the jets originate from bottom quarks. We present results for the same b-tagging technique described earlier, here in the context of the selected jets in the lepton+jets channels. For the μ+jets analysis we also show the results of an algorithm looking for evidence of muonic decays of B hadrons, the so-called muon-in-jet algorithm. Its definition is very simple: if a jet has a reconstructed muon matched to within $\Delta R < 0.4$, it indicates the presence of a bottom-quark jet.

4 Background Estimation in the Dilepton Channel

Although simulated events can adequately represent many processes, there are some pathological backgrounds that are hard to model accurately. In such cases data-driven estimates of the yields of these events are preferred. In the dilepton analysis, two background sources must be handled this way: exceptional Drell-Yan events that evade the Z veto and are accompanied by

significant missing transverse energy ; and non-W lepton signatures from multijet and W+jets production. Techniques have been developed to extract the contributions from these processes from data.

4.1 Drell-Yan Background

In the dilepton event selection events are discarded if the invariant mass of the two isolated leptons is within 15 GeV/c^2 of the Z-boson mass. The rejected events are valuable to the analysis, however, as they can be used to estimate the residual contribution from $Z/\gamma^\star \to e^+e^-$ and $Z/\gamma^\star \to \mu^+\mu^-$ in events that survive the Z veto. These we label $N_{\text{out}}^{\ell\ell,\text{data}}$, "out" for outside the Z mass window.

A simulated Drell-Yan $Z/\gamma^\star \to e^+e^-$ sample is used to determine the ratio of events within and outside the Z window, $R_{\text{out/in}}^{ee} = N_{\text{out}}^{ee,\text{MC}}/N_{\text{in}}^{ee,\text{MC}}$. The ratio is then used to scale the number of events rejected in the Z mass window, $N_{\text{in}}^{ee,\text{data}}$. If the events rejected in data are purely from $Z/\gamma^\star$, then an estimate of $N_{\text{out}}^{ee,\text{data}}$ would simply be

$$N_{\text{out}}^{ee,\text{data}} = R_{\text{out/in}}^{ee}\left(N_{\text{in}}^{ee,\text{data}}\right) \tag{1}$$

and it would just be subject to the systematic uncertainty in the modeling of the $Z/\gamma^\star \to e^+e^-$ process. However in the data, the rejected events are not purely from $Z/\gamma^\star \to e^+e^-$ production; hence, a correction to $N_{\text{in}}^{ee,\text{data}}$ must be applied that takes into account the contributions from other sources:

$$N_{\text{out}}^{ee,\text{data}} = R_{\text{out/in}}^{ee}\left(N_{\text{in}}^{ee,\text{data}} - 0.5N_{\text{in}}^{e\mu,\text{data}}k_{ee}\right), \quad k_{ee} = \sqrt{\frac{N_{\text{in}}^{ee,\text{loose}}}{N_{\text{in}}^{\mu\mu,\text{loose}}}}, \tag{2}$$

where the number of events used to estimate the non-peaking dilepton background in the Z-mass window is derived from the events observed in $e\mu$ final state, $N_{\text{in}}^{e\mu,\text{data}}$, assuming lepton universality. $N_{\text{in}}^{e\mu,\text{data}}$ is scaled by the electron-to-muon efficiency ratio k_{ee}, which is derived from a $Z/\gamma^\star$-dominated sample, using events passing lepton identification, isolation criteria but other loosened event selections in ee ($N_{\text{in}}^{ee,\text{loose}}$) and $\mu\mu$ ($N_{\text{in}}^{\mu\mu,\text{loose}}$) final states. This calculation proceeds for the $\mu\mu$ mode in a similar fashion. The data-driven prediction for $Z/\gamma^\star$ is compared to the prediction from simulation in Table 1; the comparison is shown for three stages in the dilepton event selection to illustrate how contributions from this background evolve for various stages of the event selection. A conservative estimate of the systematic uncertainty of this method gives 50%, which is based on detector calibration effects and changes of the ratio $R_{\text{out/in}}$ with increasingly more stringent requirements (additional jets and missing transverse energy) as estimated in simulated data.

4.2 Non-W/Z Lepton Background

The background from leptons not originating from W- or Z-boson decays is particularly hard to simulate accurately, both in rate and kinematics. Such lepton candidates mostly arise from jet objects that are able to satisfy the lepton identification criteria. For example, a jet can satisfy the electron criteria by being particularly narrow and depositing a significant amount of its energy in the electromagnetic calorimeter; a jet can also satisfy the isolated muon criteria through semi-leptonic hadron decay and by having a small number of charged particle tracks, or through

Table 1: Data-driven prediction for Drell-Yan processes, compared to expectations from simulation assuming 78 nb^{-1} at three points in the evolution of the event selection. Selections for columns to the left are implied. Entries with zero mean values are assigned an uncertainty equal to one event weight. Uncertainties shown for simulated data are the same as used for Table 7. Data estimates are shown with statistical and systematic uncertainties quoted separately.

Sample	ID, ISO, Z-veto	with $N_{\text{jet}} \geq 1$	with $\not\!\!E_{\text{T}}$
ee			
DY in simulation	2.4 ± 0.7	0.41 ± 0.13	0.008 ± 0.002
DY estimate in data	$2.2 \pm 0.5 \pm 1.1$	$0.23 \pm 0.16 \pm 0.12$	$0.00^{+0.16}_{-0.00}\ ^{+0.08}_{-0.00}$
$\mu\mu$			
DY in simulation	2.9 ± 0.7	0.48 ± 0.12	0.019 ± 0.005
DY estimate in data	$2.9 \pm 0.6 \pm 1.5$	$0.36 \pm 0.21 \pm 0.18$	$0.00^{+0.25}_{-0.00}\ ^{+0.12}_{-0.00}$

narrow hadronic showers that punch through the calorimeter and give rise to signals in the muon chambers.

In the dilepton analysis a data-driven technique has been developed that is based on counts of lepton candidates passing criteria looser than the full identification and isolation requirements (fakeable objects). The number of fakeable objects is then weighted by the ratio of tight-to-loose number of lepton candidates (TL ratio) as a function of candidate transverse momentum and pseudorapidity to produce an estimate of non-W/Z leptons.

Fakeable object definitions are developed for electrons and muons that are, in effect, extremely loosened versions of the electron and muon criteria discussed in Section 3.1; these are:

Muon fakeable object:

- χ^2/ndof of global fit < 50 (as opposed to < 10).
- Transverse impact parameter with respect to the beam spot < 2 mm (as opposed to < 200 μm).
- Relative energy in isolation cone <40% (as opposed to <15%).

Electron fakeable object:

- Remove all of the 90%-efficient simple cut-based electron ID requirements.
- The cut on transverse impact parameter with respect to the beam spot was removed (used to be < 400 μm).

Fakeable objects used for the TL ratio measurement are collected using data from jet triggers, which are naturally depleted in real leptons from W- or Z-boson decays. Four jet triggers are used here, each designed to accept an event if it contains at least one jet above the E_{T} threshold 6, 10, 15, and 30 GeV. Samples with different jet trigger thresholds are necessary to estimate systematic effects of the method. With increasing instantaneous luminosity, a decreasing fraction of events satisfying lower jet E_{T} trigger requirements are selected for recording to allow all more important events to be collected. In the early data sample the trigger with a 15 GeV threshold, having the highest rates, is used to measure the central value of the TL ratio. All early data, where the jet triggers were operational and a higher fraction of jet-triggered events was written out, is used to measure the TL ratio. Data from more recent data-taking periods, during which only a small fraction of events with a 15 GeV trigger threshold were recorded, are not included. Figure 1 shows the TL ratio for electrons and muons in bins of p_{T}. The values

are stable to within a relative 50 % which is sufficient precision for an early $t\bar{t}$ dilepton cross section measurement. The systematic uncertainty on the estimates using this method is 50 % for contributions with one non-W/Z lepton and 100 % for contributions with two non-W/Z leptons.

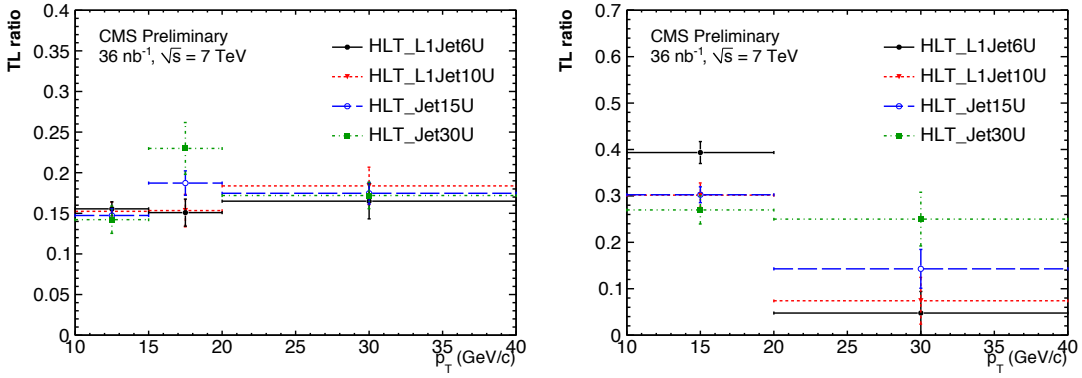

Figure 1: Electron (left) and muon (right) TL ratios as a function of fakeable object p_T.

5 Estimation of the QCD Multijet Background in the Lepton+Jets Channel

The event selection criteria for the lepton+jets channel discussed in Section 3.2 is designed to suppress backgrounds while maintaining good efficiency for $t\bar{t}$ events. In the end, the selected sample is not completely pure in signal events. A crucial component of the lepton+jets analyses is understanding the contribution in the selected sample from one particularly troublesome background source: QCD multijet events, in which jet objects conspire to satisfy the W-boson selection. As in the dilepton case, jets can fake the W-boson decay signature of a fake isolated and prompt electron or muon. If the energies of the remaining jet candidates in the event are poorly measured, the resulting spurious $\not\!\!E_T$, combined with the fake isolated and prompt lepton, can satisfy the requirements of leptonic W-boson decays. In the current incarnation of the lepton+jets event selection, no $\not\!\!E_T$ requirement is in place, making the multijet contributions in the selected sample more significant.

Three techniques, that provide a data-driven estimate of the contribution from QCD multijets to the selected lepton+jets sample, are employed. It is beneficial to perform more than one data-driven method in the analysis, in order to cross-check the results and to obtain a measure for the uncertainty in the true number of QCD events in the selected data-set. We briefly describe each method below.

5.1 ABCD Method in the μ+Jets Analysis

Muon candidates from W-boson decays have the following characteristics: They are isolated from other activity in the event, and they are consistent with originating at the pp interaction point. Hence, their relative isolation should favor small values, and their impact parameter with respect to the beam spot should be close to 0. This can be contrasted to muon candidates from multijet production, which generally have unrestricted relative isolation and impact parameter values. These characteristics are borne out in Fig. 2.

It has been verified using both collision as well as simulated events, that the impact parameter and the relative isolation for muon candidates from QCD multijet events are only weakly

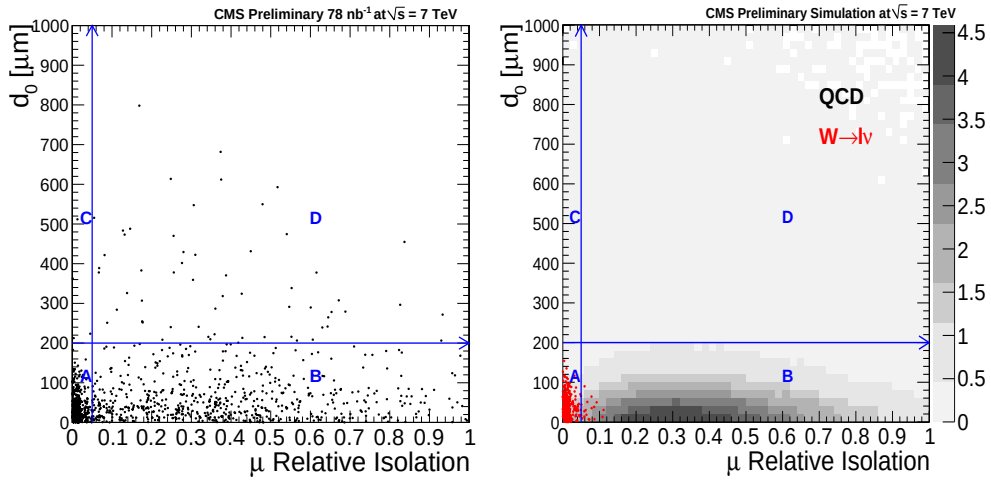

Figure 2: Distribution of the muon absolute impact parameter versus relative isolation for muons with $p_T > 20$ GeV/c from collision data (left) and from the simulation of QCD multijet and W+jets events (right), in $N_{\text{jets}} \geq 0$ events.

correlated. Under this condition, one can define the four distinct regions in the impact parameter versus relative isolation plane denoted as A, B, C and D in Fig. 2. If the two variables are completely uncorrelated, then the ratios of the populations of muon candidates obey the relationship $N_A/N_B = N_C/N_D$. If we take care to define region A to contain the signal muons from W decay, then N_A in the above expression gives us the number of muon candidates from multijet production faking our signal. N_A is then purely determined from the yield in data in the regions B, C and D, as these are background dominated.

Here we define signal region A as relative isolation < 0.05 and impact parameter < 0.02 cm. One can see from Fig. 2 the subsequent definitions of regions B, C and D. The results of this ABCD method in 78 nb^{-1} are reported in Table 2 for the two cases where at least zero or one jets are present in the μ+jets selected sample. A conservative 50% relative uncertainty is placed on the predicted N_{QCD} values and is driven by the limited statistics of the sample. The stability of the results has also been checked by varying the boundaries of regions B,C and D.

Table 2: Results of the ABCD method in the μ+jets sample for 78 nb^{-1} for events with at least zero or one reconstructed jet.

ABCD method (μ+jets)		
	$\geq$0-jet	$\geq$1-jet
Estimate $N^{est.}_{\text{QCD}}$	57$\pm$29	30$\pm$15
Prediction $N^{\text{MC}}_{\text{QCD}}$	21$\pm$2	7$\pm$2

5.2 Extrapolation Method using the Isolation Variable

The second technique for calculating the multijet background also uses the lepton relative isolation variable. We exploit the fact that non-isolated leptons come mostly from multijet events. Therefore, we can use events with large isolation values (i.e., non-isolated leptons) to define a control region dominated by multijet events. We fit the isolation distribution in the control region and extrapolate into the signal region, which, as discussed above, is characterized by small values of isolation, to obtain an estimate of the number of QCD multijet events in the signal region.

This technique has been deployed for both the μ+jets and e+jets analyses. It was found that the

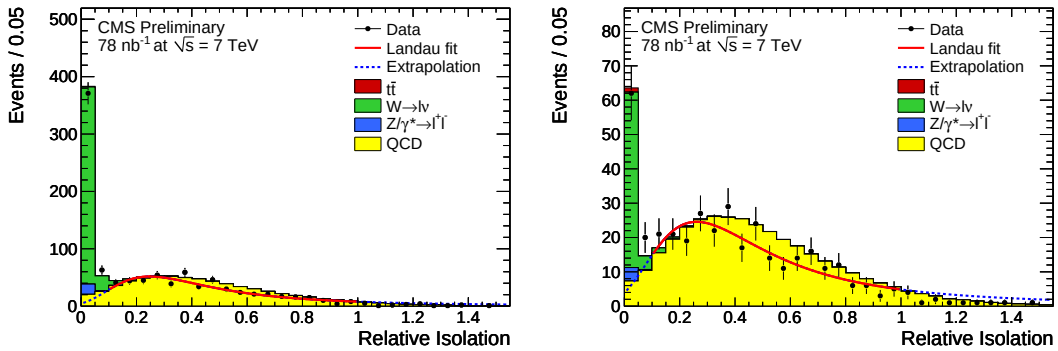

Figure 3: Fit and extrapolation to relative isolation in events with a high-p_T muon candidate and $\geq$0-jets (left) or $\geq$1-jet (right) in data using a Landau function. Shaded histograms denote expected signal and background processes based on simulation and are normalized to 78 nb^{-1}.

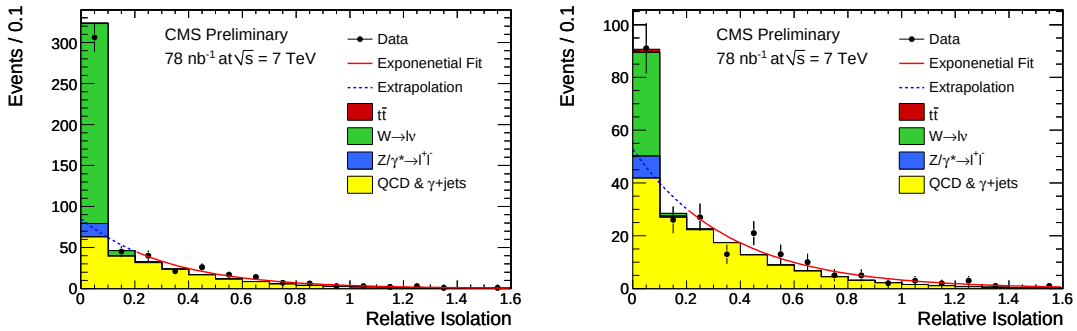

Figure 4: Fit and extrapolation to relative isolation in events with a high-E_T electron candidate and $\geq$0-jets (left) or $\geq$1-jet (right) in data using an exponential function. Shaded histograms denote expected signal and background processes based on simulation and are normalized to 78 nb^{-1}.

non-isolated μ+jet data is best fit using a Landau distribution, whereas for the e+jets channel, an exponential Ansatz is preferred. This can be explained by the fact that the QCD background composition is very different in the two channels: in μ+jets, the background is mostly due to muons originating from b/c-decays in jets, decays-in-flight of hadrons and hadronic showers reaching the muon system. In e+jets, it is also due to b/c-decays in jets, but there is in addition a significant contribution from electrons originating from photon conversions due to the large amount of silicon tracker material in front of the calorimeter.

Example fit results for events containing $\geq$0 and $\geq$1 jets in the μ+jets and e+jets analyses are shown in Figs. 3 and 4, respectively. Various fit intervals and binning schemes have been tried, some of which are summarized in Tables 3 and 4. We assign a 50 % uncertainty on the predicted number of QCD events in data.

5.3 Modeling the Kinematic Shapes for QCD Events in μ+Jets

In addition to determining the rate of multijet events in the selected sample, another feature of the relative isolation extrapolation method is the ability to construct a data-driven model for kinematic shape information from these events. By reversing the relative isolation cut defining our muon candidates (i.e. requiring the relative isolation to be greater than 5%) one has a pool of QCD-enriched events, from which kinematic distributions for this background source can

Table 3: Results of relative isolation extrapolation method in muon+jets events. Various fit ranges were pursued and the variation of the results is reported. A 50% systematic uncertainty is placed on the averaged result.

Isolation extrapolation method (μ+jets)			
Fit Range	$N_{\mathrm{QCD}}^{est.}(\geq$0-jet)	Fit Range	$N_{\mathrm{QCD}}^{est.}(\geq$1-jet)
0.13–0.8	11	0.13–0.9	8
0.15–0.8	6	0.15–0.9	5
0.10–1.0	9	0.10–1.0	7
0.15–1.0	3	0.15–1.0	2
Average $N_{\mathrm{QCD}}^{est.}$	7 ± 4		6 ± 3
Prediction $N_{\mathrm{QCD}}^{\mathrm{MC}}$	21 ± 2		7 ± 2

Table 4: Results of relative isolation extrapolation method in electron+jets events. Uncertainties on the estimate are obtained from the uncertainties on the fit parameters using error propagation. Various fit ranges were pursued and the variation of the results is reported. A 50% systematic uncertainty is placed on the averaged result.

Isolation extrapolation method (e+jets)		
Fit Range	$N_{\mathrm{QCD}}^{est.}(\geq$0-jet)	$N_{\mathrm{QCD}}^{est.}(\geq$1-jet)
0.1–1.6	67 ± 9	40 ± 6
0.2–1.6	73 ± 13	46 ± 9
0.3–1.6	71 ± 17	45 ± 12
Average $N_{\mathrm{QCD}}^{est.}$	70 ± 35	44 ± 22
Prediction $N_{\mathrm{QCD}}^{\mathrm{MC}}$	63 ± 7	42 ± 6

be obtained. Figure 5 has a comparison of the transverse mass distribution in μ+jets data compared to the background prediction from simulated W/Z+jets and the data-driven prediction for multijets, once we reverse the isolation requirement. Here we use the normalization for the QCD contribution from the ABCD method, demonstrating one way of synthesizing the results of these techniques. The agreement in Fig. 5 indicates this data-driven kinematic modeling strategy is reasonable even at this early stage.

However, comparing the results of the two QCD estimation methods in μ+jets presented in Tables 2 and 3, one observes a significant disagreement, even within the presently large uncertainties. In the comparisons of data and simulation, which will be presented in Section 6.2, the expected yields from QCD simulation will be used, which lie roughly in-between the results of the two data-driven approaches, and a 100% uncertainty will be assigned. The eventual measurement in the μ+jets mode can afford this level of uncertainty since the QCD contribution is relatively small and only an upper limit of its contribution is really necessary.

5.4 Template Fit in e+Jets Analysis

There are event-level kinematic variables that discriminate between QCD multijet and $t\bar{t}$ signal events: $\not\!\!E_{\mathrm{T}}$ and $H_{T,lep}$ both provide background discrimination, where $H_{T,lep}$ is the scalar sum of the missing transverse energy $\not\!\!E_{\mathrm{T}}$ and the lepton transverse energy E_T.

The third technique being explored for the e+jets analysis seeks to extract the QCD contribution to the signal sample from fits in these variables. Two models are considered in order to obtain template distributions for QCD multijet events: "background" electrons, in which the electron candidate very nearly satisfies the selection criteria but instead is a marginal failure; and

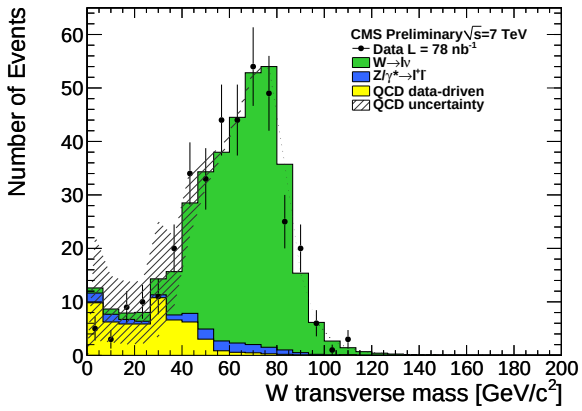

Figure 5: Transverse W-mass distribution for μ+jets events with ≥ 0 jets, corresponding to an integrated luminosity of 78 nb^{-1}. Also shown is the prediction for W/Z+jets from simulation, as well as a data-driven QCD template obtained from inverting the relative isolation requirement, and normalized according to the data-driven estimate from the ABCD method.

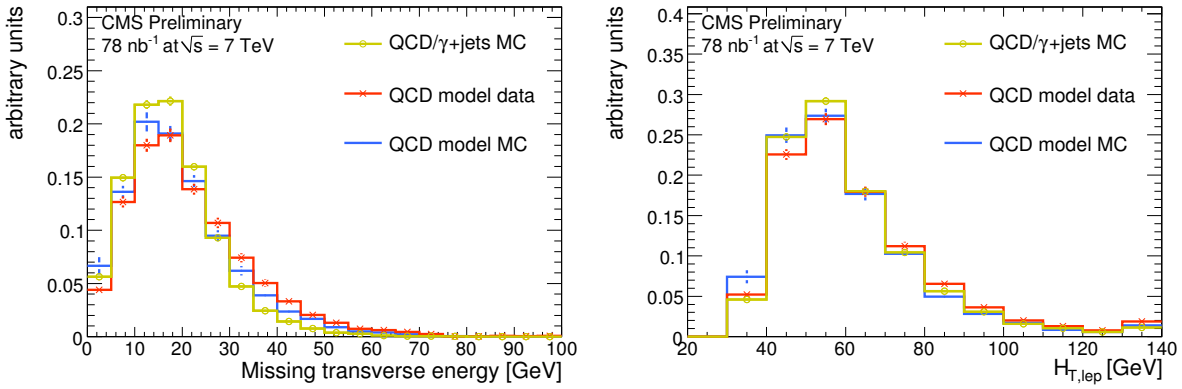

Figure 6: e+jets: Shape comparison for $\not\!\!E_{\text{T}}$ (left) and $\text{H}_{\text{T,lep}}$ (right) of the QCD simulation (labeled 'QCD/γ+jets MC') and the data-driven QCD template obtained from simulation (labeled 'QCD model MC'), as well as from data (labeled 'QCD model data').

jet-electrons, positively identified jet objects with large electromagnetic fraction, that closely resemble electron candidates. According to simulation, both selections yield a QCD purity of $\sim 99\%$. The QCD shapes obtained using the two models are similar, such that their normalized sum is used in the following.

Figure 6 shows a comparison of the QCD shape constructed from simulated events using the procedure described above with the predicted QCD shape from the simulation. Also shown is the QCD shape obtained from data. When comparing the template obtained from simulation with the true simulated distribution, a small bias on the shape of the distributions, due to the altered event selection, is visible.

The QCD contribution in the signal region, defined as either $\not\!\!E_{\text{T}} > 25$ GeV or $H_{T,lep} > 60$ GeV, is obtained from fitting a sum of templates to the data in the background region at small $\not\!\!E_{\text{T}}$ or $H_{T,lep}$: a QCD template obtained from data using the background model described above, and W/Z+jets templates from simulation.

A comparison of the results obtained using this method is presented in Table 5 for events without requirement on the number of reconstructed jets. Table 6 has results for events with at

least one jet. For the latter the signal region is defined as $\not\!\!E_T > 30$ GeV or $H_{T,lep} > 70$ GeV. The results are, within the statistical uncertainties reported here, well compatible between the fits using the two different variables. Systematic uncertainties due to the difference in shape between the QCD template models, and due to a bias introduced by QCD model, are being studied, but still comparable to or smaller than the statistical uncertainties.

By comparison of the results for the QCD estimation in *e*+jets using either the relative isolation extrapolation method or the template method with the respective prediction from QCD simulation, one can infer that the two methods are roughly consistent with each other, even though they yield predictions for slightly different phase space regions.

Table 5: *e*+jets: Results of the QCD estimation using templates for events without any jet requirement.

Variable	Template	QCD in bkg. region	QCD in sig. region	Whole dataset
$\not\!\!E_T$	QCD model	41±15	19±7	60±23
	prediction (sim.)	50.5±0.5	12.2±0.2	62.7±0.5
$H_{T,lep}$	QCD model	47±13	39±11	86±24
	prediction (sim.)	36.7±0.4	26.0±0.3	62.7±0.5

Table 6: *e*+jets: Results of the QCD estimation using templates for event with at least one jet.

Variable	Template	QCD in bkg. region	QCD in sig. region	Whole dataset
$\not\!\!E_T$	QCD model	28±17	8±5	37±21
	prediction (sim.)	36.3±0.4	5.3±0.1	41.6±0.4
$H_{T,lep}$	QCD model	26±10	10±4	36±14
	prediction (sim.)	29.2±0.4	12.4±0.2	41.6±0.4

6 Results

In the Sections above we have discussed the selection criteria that have been put in place to select $t\bar{t}$ events in early CMS data in both the dilepton and lepton+jets channels. We have also presented several methods that will help to provide an accurate assessment of the background contributions. This section presents the results of the selection and background estimates in the initial 78 ± 9 nb^{-1} of integrated luminosity. No evidence of $t\bar{t}$ signal is observed in this sample in agreement with expectations.

6.1 Dilepton Channel

Table 7 shows the comparison of the predicted and observed event yields in the dilepton channel for events containing one or more jets and required to pass lepton identification and isolation requirements as well as Z-veto. Also quoted are predictions from data-driven methods for fake leptons and for the Drell-Yan contributions, which are meant to replace the simulated contribution from $Z/\gamma^\star \to e^+e^-$, $Z/\gamma^\star \to \mu^+\mu^-$, QCD, W+jets, and non-dileptonic $t\bar{t}$.

The uncertainties on the signal and background yields from simulation are systematic. They are combined from estimates on the agreement of event selection efficiency between simulation and data (instrumental effects), uncertainties related to the cross section predictions from theory and an overall relative 11 % uncertainty on the measured integrated luminosity [4]. The systematic uncertainties are estimated assuming full signal event selections, including a number of jets and a missing transverse energy requirement, as described in [18].

The systematic uncertainty due to instrumental effects is comprised primarily of uncertainties on the lepton reconstruction, identification, and isolation criteria as well as uncertainties on the jet and missing transverse energy selection driven by the jet energy scale uncertainty. The combined instrumental uncertainty is estimated to be 15%, 20% and 25% in the dimuon, electron-muon, and dielectron final states respectively. This incorporates a 10% relative uncertainty per electron added to cover a discrepancy between simulation and data attributed to tracking detector to electromagnetic calorimeter alignment effects. This uncertainty is quoted separately for $t\bar{t}$ signal in Tables 7 and 8.

The uncertainty on the cross section predictions for the $t\bar{t}$ signal is about 16 %, and the corresponding uncertainty on the Drell-Yan process, assuming full signal event selections, is conservatively estimated to 15 %. Uncertainties on the remaining backgrounds are expected to be in the same range. Since these backgrounds are small, combined with the instrumental uncertainty, the total estimated uncertainty quoted in Table 7 is conservatively taken to be 50%.

The systematic uncertainty on the data-driven estimates, as discussed in Section 4, is 50% for Drell-Yan, 50% for W+jets, and 100% for QCD. This uncertainty is quoted separately from the statistical uncertainty of each method.

In just 78 nb^{-1} of data the event yields are vanishingly small. To better gauge the performance of the selection and background estimation techniques, a relaxed version of the event selection was executed. For this study just the trigger, lepton object identification and isolation, and dilepton requirements were put in place. The results are shown in Table 8; the background contributions therein are estimated from simulated samples and the quoted uncertainties are the same as in Table 7.

With the relaxed selection conditions we see a number of events in the ee and $\mu\mu$ modes. A comparison of a few interesting event-level variables are contained in the figures below. Figure 7 shows the jet multiplicity for events passing this criteria; we see one event in the $\mu\mu$ mode

with two jets, the jet multiplicity that is relevant for dilepton $t\bar{t}$ events.

Figure 8 contains the $\not\!\!E_{\mathrm{T}}$ distribution for the three dilepton modes and their cumulative sample; $t\bar{t}$ signal events will favor large $\not\!\!E_{\mathrm{T}}$. Once enough integrated luminosity is in hand, the $\not\!\!E_{\mathrm{T}}$ distribution will be crucial for $t\bar{t}$ signal extraction. Figure 9 shows the dilepton invariant mass and Fig. 10 the p_{T} sum of the two leptons for the events in the dilepton modes. Finally, the b-tag multiplicity for the ee and $\mu\mu$ modes is shown in Fig. 11; this will become another important variable in the early establishment of the top-quark pair signal.

Table 7: Events collected in data with at least one jet after Z-veto compared to expectations from simulated data normalized to 78 nb^{-1}. Also shown are estimates using data events for contributions with non-W/Z leptons and off-peak Drell-Yan, which are meant to replace non-dilepton $t\bar{t}$, W+jets, QCD and Drell-Yan expectations from simulation. Entries with zero mean values are assigned an uncertainty equal to one event weight. The uncertainties are described in the text.

Sample	ee	$\mu\mu$	$e\mu$
Dilepton $t\bar{t}$	$0.058 \pm 0.015 \pm 0.010$	$0.065 \pm 0.010 \pm 0.011$	$0.157 \pm 0.031 \pm 0.025$
Dibosons - VV	0.0045 ± 0.0023	0.0043 ± 0.0022	0.011 ± 0.006
Single top - tW	0.0040 ± 0.0020	0.0043 ± 0.0022	0.011 ± 0.005
Drell-Yan $\tau\tau$	0.008 ± 0.004	0.012 ± 0.006	0.016 ± 0.008
Drell-Yan $ee, \mu\mu$	0.41 ± 0.13	0.48 ± 0.12	0.003 ± 0.001
Non-dilepton $t\bar{t}$	0.002 ± 0.001	0.0003 ± 0.0001	0.003 ± 0.001
W+jets	0.003 ± 0.001	$0.000\ ^{+0.001}_{-0.000}$	0.005 ± 0.002
QCD multijets	$0.0\ ^{+1.2}_{-0.0}$	$0.0\ ^{+1.2}_{-0.0}$	$0.0\ ^{+1.2}_{-0.0}$
Total simulation	$0.45\ ^{+1.2}_{-0.11}$	$0.52\ ^{+1.2}_{-0.11}$	$0.21\ ^{+1.2}_{-0.03}$
QCD data-driven	$0.00\ ^{+0.06\ +0.06}_{-0.00\ -0.00}$	$0.0\ ^{+0.2\ +0.2}_{-0.0\ -0.0}$	$0.0\ ^{+0.1\ +0.1}_{-0.0\ -0.0}$
W+jets data-driven	$0.0\ ^{+0.2\ +0.1}_{-0.0\ -0.0}$	$0.0\ ^{+0.4\ +0.2}_{-0.0\ -0.0}$	$0.0\ ^{+0.4\ +0.2}_{-0.0\ -0.0}$
Drell-Yan data-driven	$0.23 \pm 0.16 \pm 0.12$	$0.35 \pm 0.21 \pm 0.18$	N/A
Data	0	1	0

Table 8: Expected signal and background yields for a relaxed dilepton selection, requiring just two oppositely charged leptons passing identification and isolation criteria compared to the yield in the initial 78 nb^{-1} of data. Background estimates are taken from simulation only. The uncertainties are assigned in the same way as in Table 7.

Process	ee	$\mu\mu$	$e\mu$
Dilepton $t\bar{t}$	0.077 $\pm$0.019$\pm$0.012	0.087 $\pm$0.013$\pm$0.014	0.161 $\pm$0.032$\pm$0.026
Dibosons - VV	0.023 $\pm$ 0.012	0.029 $\pm$ 0.015	0.034 $\pm$ 0.017
Single top - tW	0.0058 $\pm$ 0.0029	0.0063 $\pm$ 0.0032	0.012 $\pm$ 0.06
Drell-Yan $\tau\tau$	0.056 $\pm$ 0.028	0.06 $\pm$ 0.03	0.12 $\pm$ 0.06
Drell-Yan $ee, \mu\mu$	27 $\pm$ 8	32 $\pm$ 8	0.011 $\pm$ 0.005
Non-dilepton $t\bar{t}$	0.002 $\pm$ 0.001	0.0004 $\pm$ 0.0002	0.003 $\pm$ 0.002
W+jets	0.017 $\pm$ 0.009	0.0008 $\pm$ 0.0004	0.022 $\pm$ 0.011
QCD multijets	$0.0\ ^{+1.2}_{-0.0}$	$0.0\ ^{+1.2}_{-0.0}$	$0.0\ ^{+1.2}_{-0.0}$
Total simulation	28 $\pm$ 8	32 $\pm$ 8	$0.36\ ^{+1.2}_{-0.07}$
Data	25	32	0

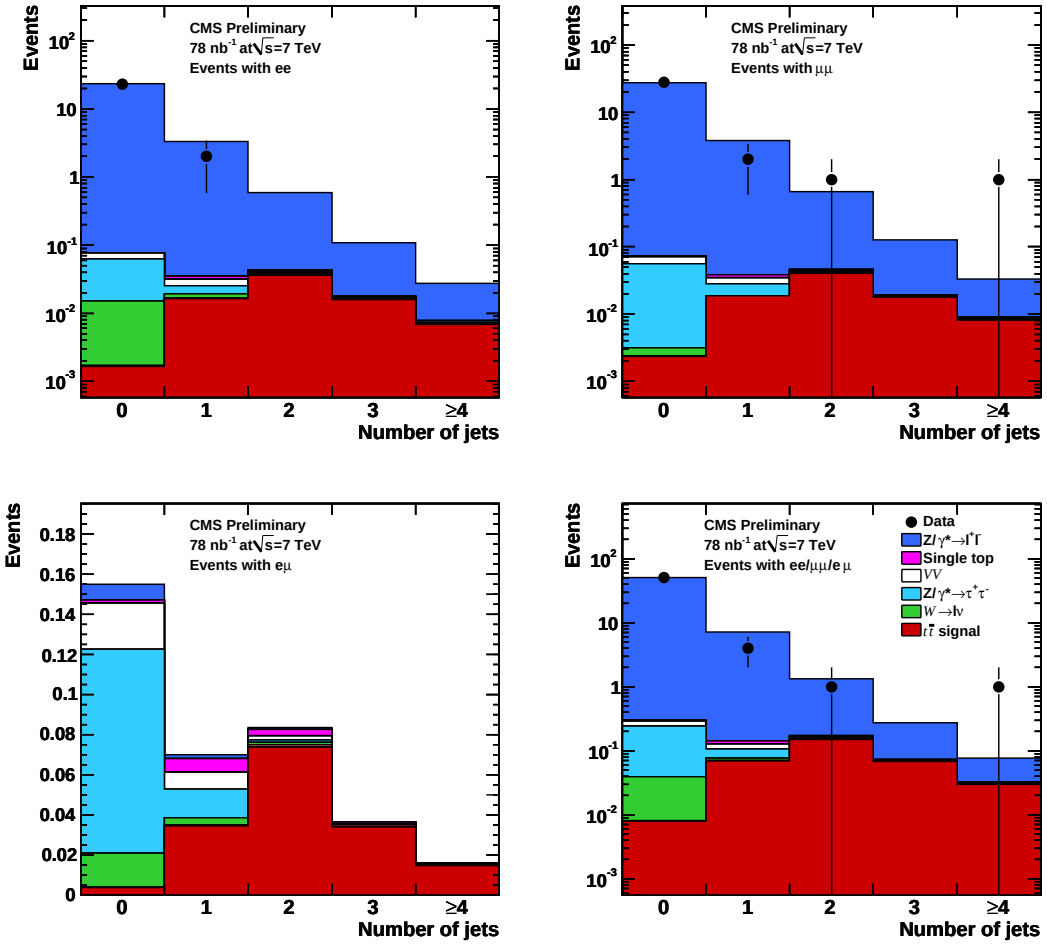

Figure 7: Jet multiplicity for the ee (top left), $\mu\mu$ (top right), $e\mu$ (bottom left) and all modes (bottom right) for events passing relaxed dilepton selection criteria in 78 nb^{-1} of integrated luminosity, compared to simulation-driven background prediction.

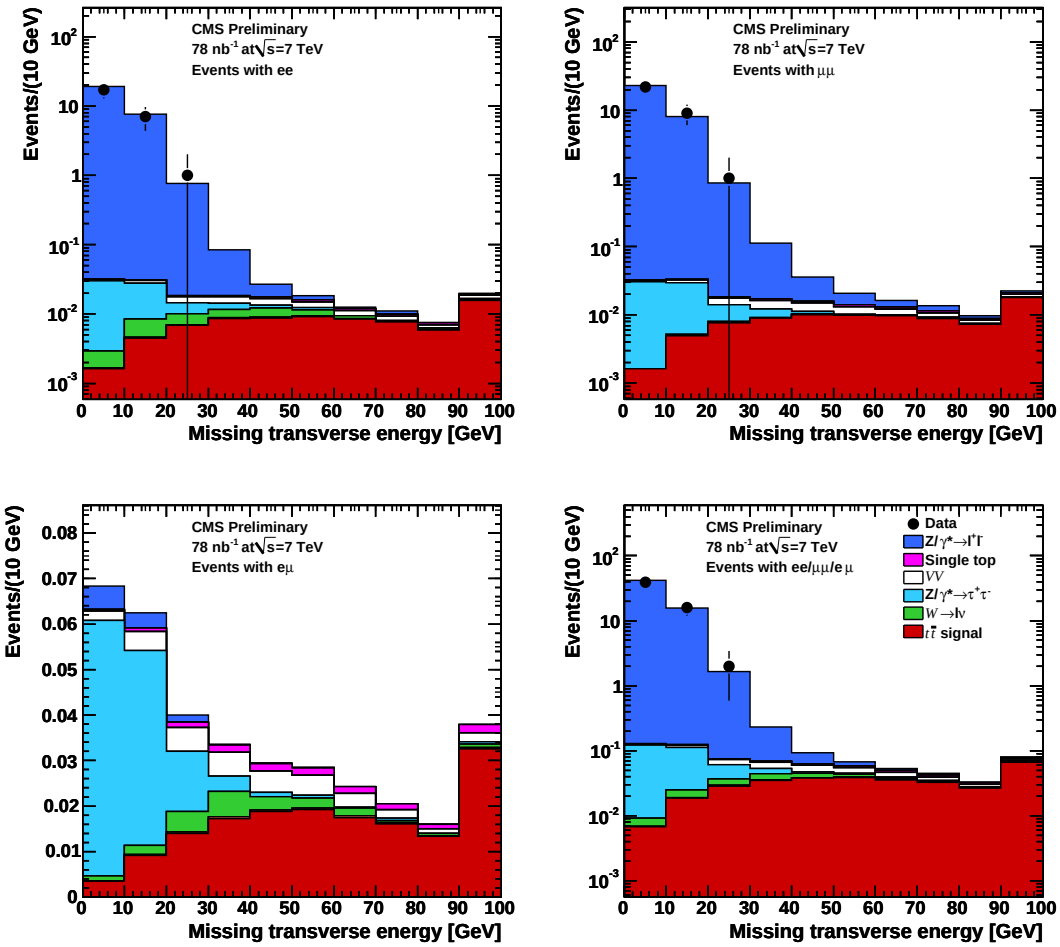

Figure 8: Missing transverse energy for the ee (top left), $\mu\mu$ (top right), $e\mu$ (bottom left) and all modes (bottom right) for events passing relaxed dilepton selection criteria in 78 nb^{-1} of integrated luminosity, compared to simulation-driven background prediction. The last bin includes overflows.

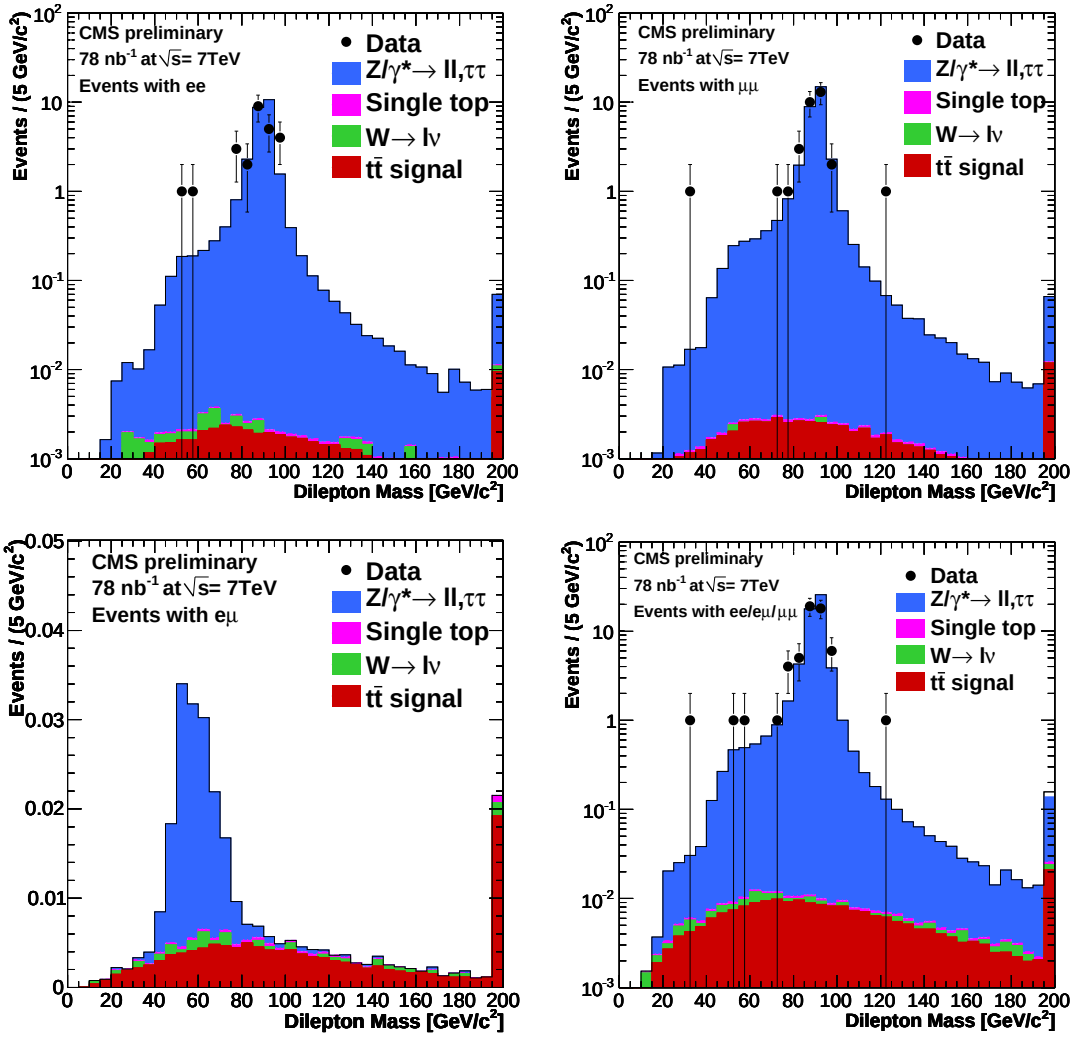

Figure 9: Dilepton invariant mass $M_{\ell\ell}$ for the ee (top left), $\mu\mu$ (top right), $e\mu$ (bottom left) and all modes (bottom right) in events passing relaxed dilepton selection criteria in 78 nb^{-1} of integrated luminosity, compared to simulation-driven background prediction. The last bin includes overflows.

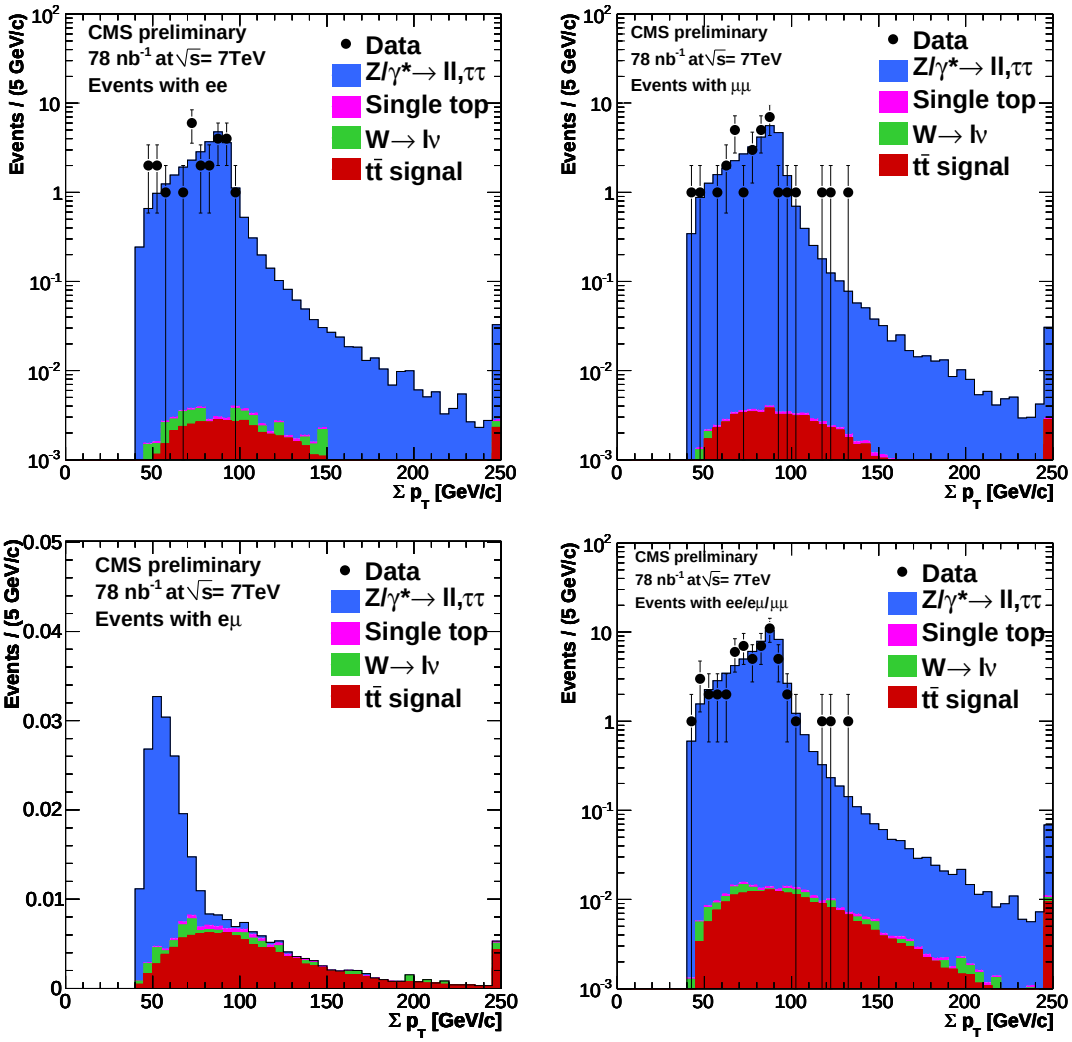

Figure 10: Sum of transverse momenta of two leptons for the ee (top left), $\mu\mu$ (top right), $e\mu$ (bottom left) and all modes (bottom right) in events passing relaxed dilepton selection criteria in 78 nb^{-1} of integrated luminosity, compared to simulation-driven background prediction. The last bin includes overflows.

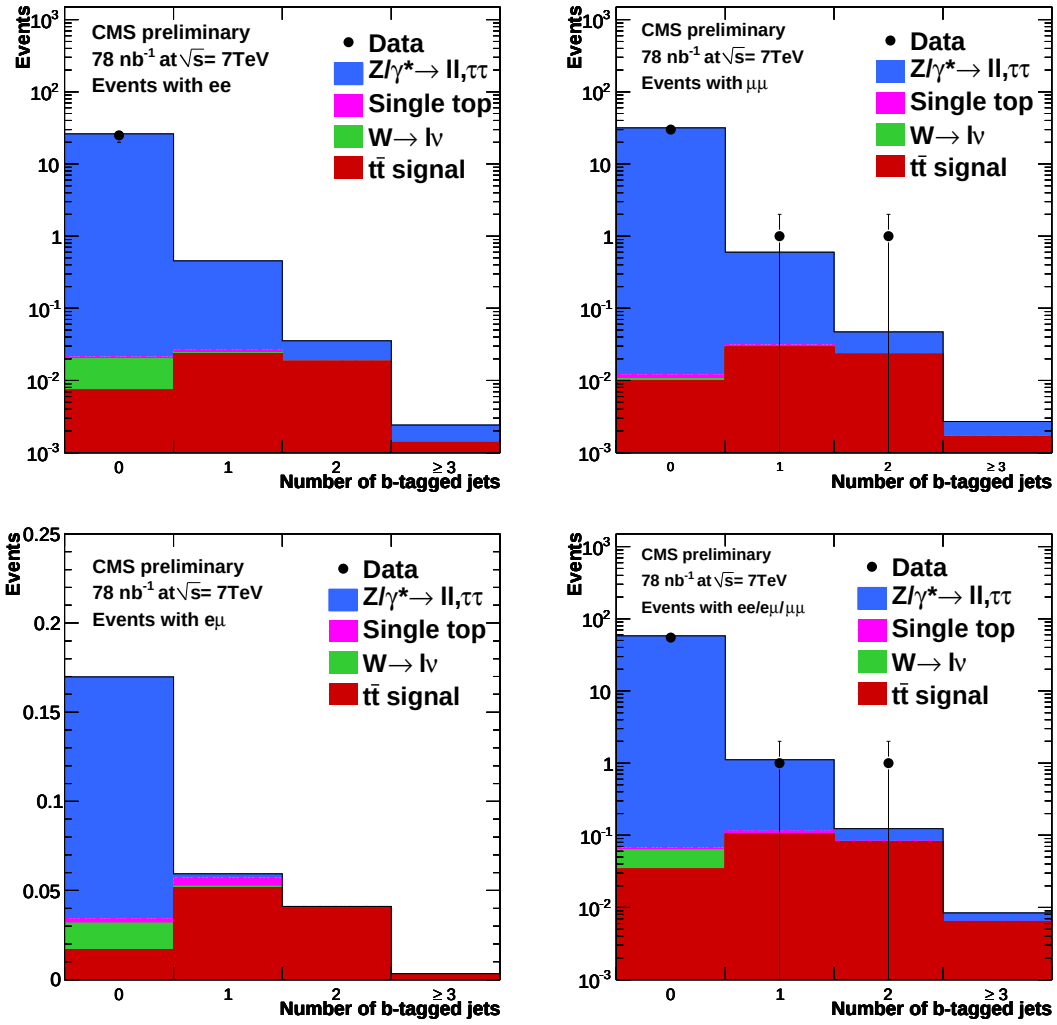

Figure 11: Number of b-tagged jets for the ee (top left), $\mu\mu$ (top right), $e\mu$ (bottom left) and all modes (bottom right) for the relaxed dilepton event selection of Table 8 in 78 nb^{-1} of data.

6.2 Lepton+Jets Channel

The yield of selected events in the dataset corresponding to an integrated luminosity of 78 nb^{-1} is presented for different requirements on the reconstructed number of jets for the e+jets and μ+jets channels in Tables 9 and 10, along with a comparison to the prediction from simulated event samples. The predictions for $t\bar{t}$, single top-quark production, as well as W/Z+jets are obtained using the MADGRAPH samples described in Section 2, normalized to theory cross sections also discussed in Section 2. The predicted number of QCD events is obtained from the PYTHIA simulation. In the case of e+jets, the QCD contribution includes γ+jet events.

Table 9: Predicted and measured event yields in the e+jets channel. The predicted yields, obtained from simulation, are normalized to the integrated luminosity of 78 nb^{-1} of the analyzed data sample. The uncertainties on the predicted yields are discussed in the text.

e+jets channel, L = 78 nb^{-1}							
Jets	$t\bar{t}$	single top	W+Jets	Z+Jets	QCD	Sum MC	Data
≥ 0	1.1 ± 0.2	0.32 ± 0.04	244 ± 29	17 ± 2	63 ± 7	325 ± 30	306
≥ 1	1.1 ± 0.2	0.29 ± 0.03	39 ± 7	8.4 ± 1.1	42 ± 6	91 ± 10	91
≥ 2	1.1 ± 0.2	0.17 ± 0.01	7.0 ± 1.7	1.7 ± 0.4	9.0 ± 2.3	19 ± 3	11
≥ 3	0.83 ± 0.17	0.07 ± 0.01	1.2 ± 0.4	0.30 ± 0.09	1.5 ± 0.5	3.9 ± 0.7	3
≥ 4	0.45 ± 0.11	0.02 ± 0.01	0.25 ± 0.10	0.05 ± 0.02	0.22 ± 0.10	0.99 ± 0.18	1

Table 10: Predicted and measured event yields in the μ+jets channel. The predicted yields, obtained from simulation, are normalized to the integrated luminosity of 78 nb^{-1} of the analyzed data sample. The uncertainties on the predicted yields are discussed in the text.

μ+jets channel, L = 78 nb^{-1}							
Jets	$t\bar{t}$	single top	W+Jets	Z+Jets	QCD	Sum MC	Data
≥ 0	1.3 ± 0.2	0.39 ± 0.04	344 ± 42	18 ± 3	21 ± 2	385 ± 42	371
≥ 1	1.2 ± 0.2	0.36 ± 0.04	51 ± 10	3.9 ± 1.1	7.3 ± 1.6	64 ± 10	62
≥ 2	1.2 ± 0.2	0.21 ± 0.02	8.5 ± 2.2	0.66 ± 0.41	0.92 ± 0.27	12 ± 2	10
≥ 3	0.95 ± 0.19	0.08 ± 0.01	1.5 ± 0.5	0.12 ± 0.08	0.12 ± 0.05	2.7 ± 0.6	4
≥ 4	0.52 ± 0.13	0.02 ± 0.01	0.29 ± 0.11	0.02 ± 0.02	0.01 ± 0.01	1.0 ± 0.1	0

The following uncertainties on the predicted yields were considered: the jet energy scale was varied within $\pm 10\%$ [25]; the luminosity was varied by $\pm 11\%$ [4]; the theory cross sections used to normalize the samples were varied within their uncertainties, see Section 2, which include variations of scales and parton density functions. For QCD, the statistical uncertainty of the sample was used instead of a theory uncertainty. Finally, all uncertainties were added in quadrature.

In the e+jets channel, 306 events pass the selection, of which three have more than three reconstructed jets. There is also one 4-jet event observed. In the μ+jets channel, 371 events are reconstructed, including four 3-jet events. As Tables 9 and 10 show, for this initial sample, the predicted yields are in good agreement with the observed yields in data.

Differential distributions are shown in Figures 12–18. Data are compared with $t\bar{t}$, single top and W/Z+jets simulations obtained as described above. For QCD multijet events, the distributions as well as their normalizations are obtained from the PYTHIA simulations. In addition, an error band reflects the estimated uncertainty on the amount of QCD of 50 % (100 %) in the e+jets (μ+jets) channel, as obtained from the data-driven QCD estimates discussed in Section 5. Within the uncertainties, good agreement is observed between data and simulation.

Figure 12 shows the transverse energy E_T^{el} and the pseudorapidity η^{el} for electron candidates.

The full event selection besides the jet requirements is applied. The E_T^{el} distribution is shown without the minimum $E_T^{el} > 30$ GeV requirement. The majority of electron candidates are at low transverse energy and would fail the $E_T^{el} > 30$ GeV threshold. W-boson events, evident only at large E_T^{el}, become dominating the event sample once the $E_T^{el} > 30$ GeV requirement is applied. The missing transverse energy $\not\!\!E_{\mathrm{T}}$, transverse mass $M_T(W)$ and $H_{T,lep} = E_T^{el} + \not\!\!E_{\mathrm{T}}$ distributions, also shown in Fig. 12, demonstrate that the sample is dominated by W events. Also the regions, in which the QCD contribution is expected to be important, are well described by the simulation.

Figure 13 shows that the present sample is dominated by 0- and 1-jet events, even though there are a few candidates at higher jet multiplicities, in which $t\bar{t}$ is expected to become more prominent. Not only the jet multiplicity, but also the transverse momentum and pseudorapidity of the leading jet are well described by simulation.

The jet multiplicity is shown in Figure 14 for events where at least one jet is identified as b-jet according to the loose working point of the track-counting algorithm. The purity of the $t\bar{t}$ signal increases quickly for b-tagged events and high jet multiplicity. The number of events with b-tags is still rather limited.

Similar conclusions can be drawn from the μ+jets results shown in Figures 15–17. These distributions also correspond to events passing the μ+jets selection without any requirement on jet multiplicity. The W-boson signal is clearly visible in Fig. 15, similar to the e+jets results. A slight excess of events is observed at low p_T^{μ}, $\not\!\!E_{\mathrm{T}}$, $M_T(W)$ and $H_{T,lep}$ values, where the contribution from QCD multijet events is expected to be important. It is, however, consistent with simulation within uncertainties. The jet distributions shown in Fig. 16 are well reproduced by the simulation as well.

Figure 17 shows the jet multiplicity for events containing at least one b-tagged jet using the the loose working point of the track-counting algorithm. In addition, the jet multiplicity is shown for events, in which at least one jet is associated with a muon. The number of observed tagged events is still comparatively small, as expected.

Finally, Fig. 18 shows the distribution of M3, the invariant mass of the three jets in the event that have the largest vectorially summed transverse momentum, for events with at least three jets and summing both the e+jets and μ+jets channels. M3 can be considered an estimator of the reconstructed top-quark mass in $t\bar{t}$ events. A few events are observed, for which the M3 value reconstructed is broadly in the range expected for $t\bar{t}$ production. However, it is difficult to draw any conclusions related to the production of top quarks at this time, given the very small dataset considered to date.

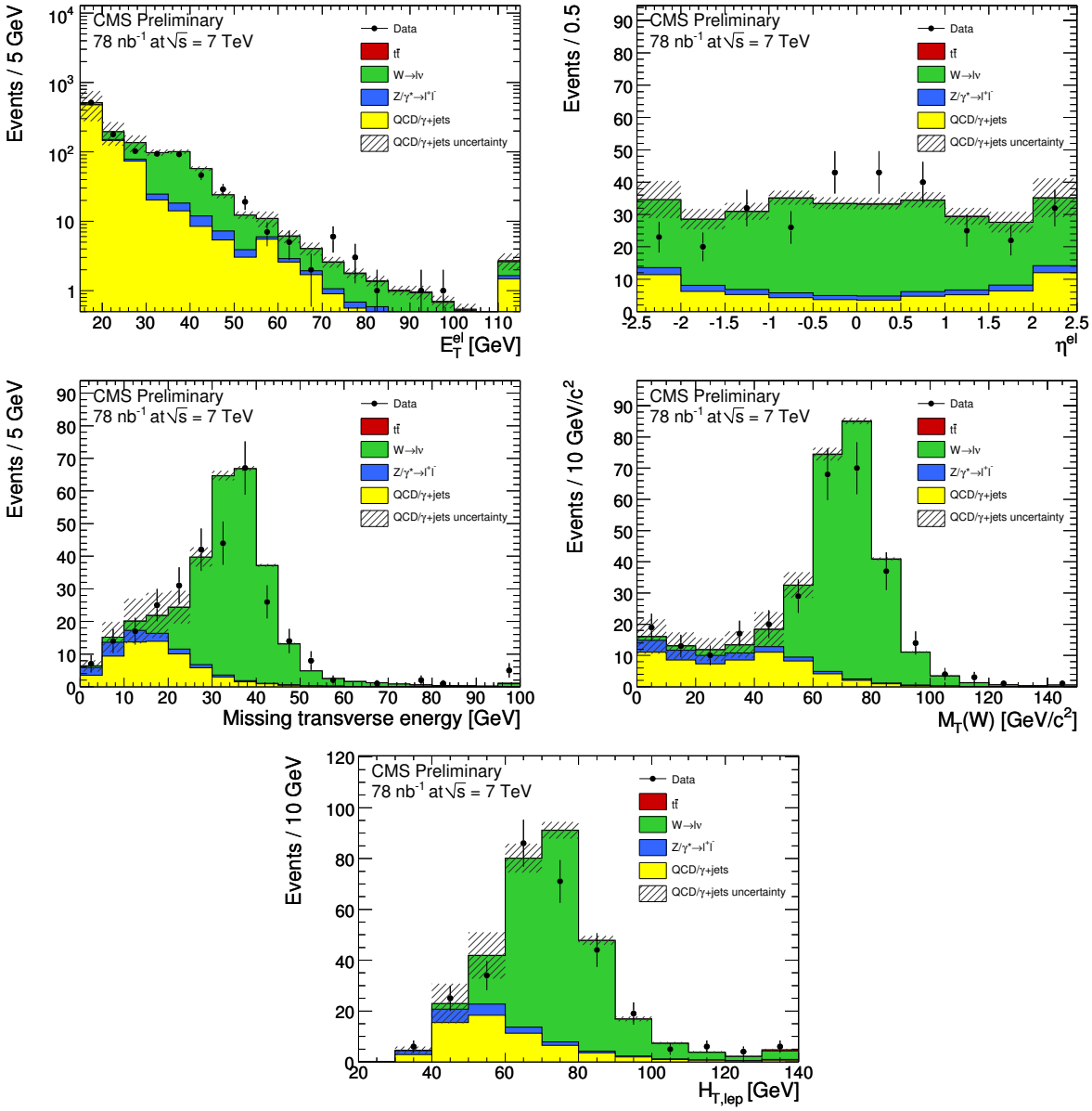

Figure 12: e+jets channel: Distributions are shown of the Electron transverse energy E_T^{el} and pseudorapidity η^{el} (top row), missing transverse energy and transverse mass $M_T(W)$ (middle row), as well as the $H_{T,lep} = E_T^{el} + \not{E}_T$ variable (bottom row), for any jet multiplicity and corresponding to 78 nb^{-1}. The E_T^{el} distribution is shown without the minimum $E_T^{el} > 30$ GeV requirement applied. Here and in Fig. 13 and 14, predictions from simulation are overlaid, as described in the text and normalized to the integrated luminosity of the data sample. The hatched error band corresponds to an uncertainty of 50 % on the number of expected QCD multijet events. The last bins include overflows.

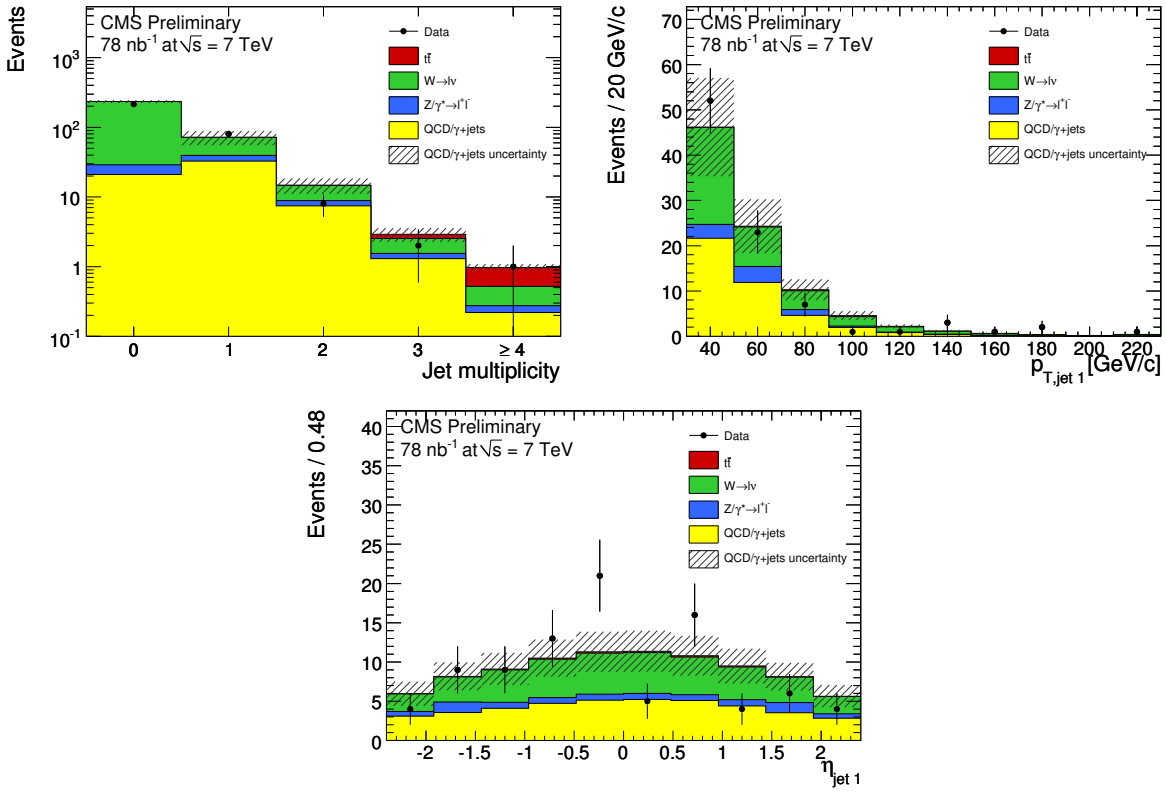

Figure 13: e+jets channel: Distributions are shown of the Jet multiplicity (top left), as well as transverse momentum $p_{\mathrm{T,jet\,1}}$ (top right) and pseudorapidity $\eta_{\mathrm{jet\,1}}$ (bottom) of the leading jet, corresponding to $78\,\mathrm{nb}^{-1}$.

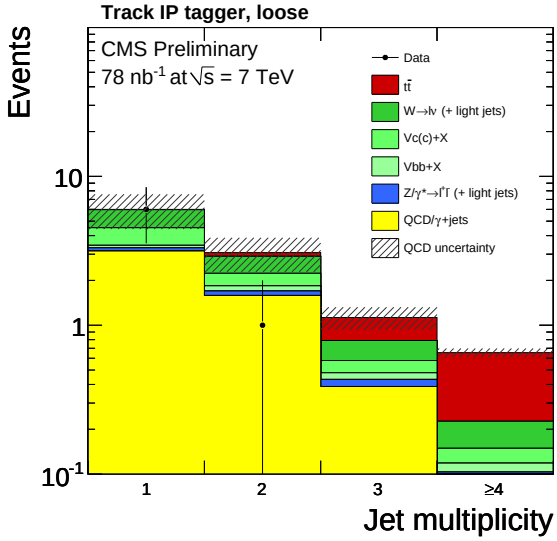

Figure 14: e+jets channel: The distribution of the jet multiplicity for events in which at least one jet is b-tagged using the loose working point of the track-counting algorithm is shown, corresponding to $78\,\mathrm{nb}^{-1}$.

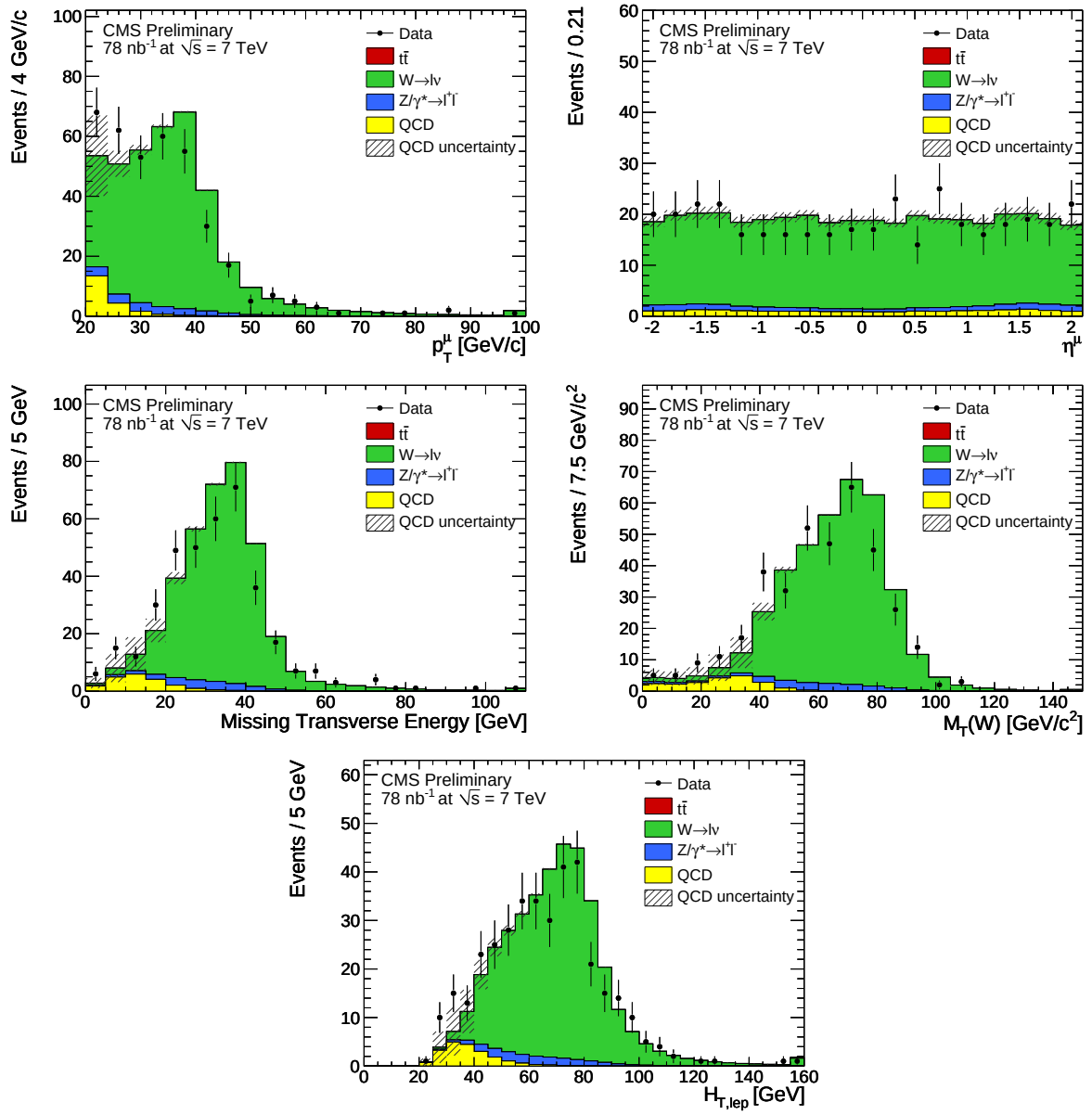

Figure 15: μ+jets channel: Distributions are shown of the muon transverse momentum p_T^{μ} and pseudorapidity η^{μ} (top row), missing transverse energy and transverse mass $M_T(W)$ (middle row), as well as the $H_{T,lep} = p_T^{\mu} + \not{E}_T$ variable (bottom row), for any jet multiplicity and corresponding to an integrated luminosity of 78 nb^{-1}. Here and in Fig. 16 and 17, predictions from simulation are overlaid, as described in the text and normalized to the integrated luminosity of the data sample. The hatched error band corresponds to an uncertainty of 100 % on the number of expected QCD multijet events. The last bins include overflows.

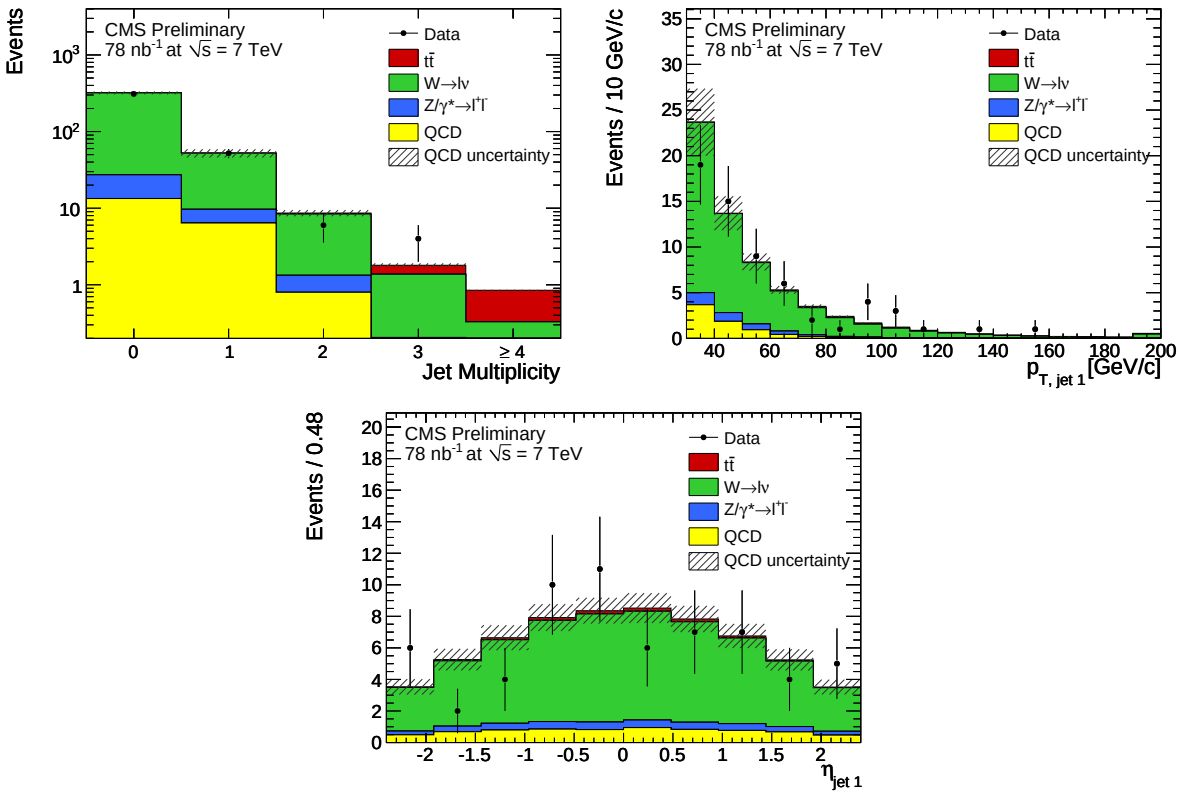

Figure 16: μ+jets channel: Distributions are shown of the jet multiplicity (top left), as well as transverse momentum $p_{\mathrm{T,jet\,1}}$ (top right) and pseudorapidity $\eta_{\mathrm{jet\,1}}$ (bottom) of the leading jet, corresponding to 78 nb^{-1} of data. The last bin of the $p_{\mathrm{T,jet\,1}}$ distribution shown on the top right includes overflows.

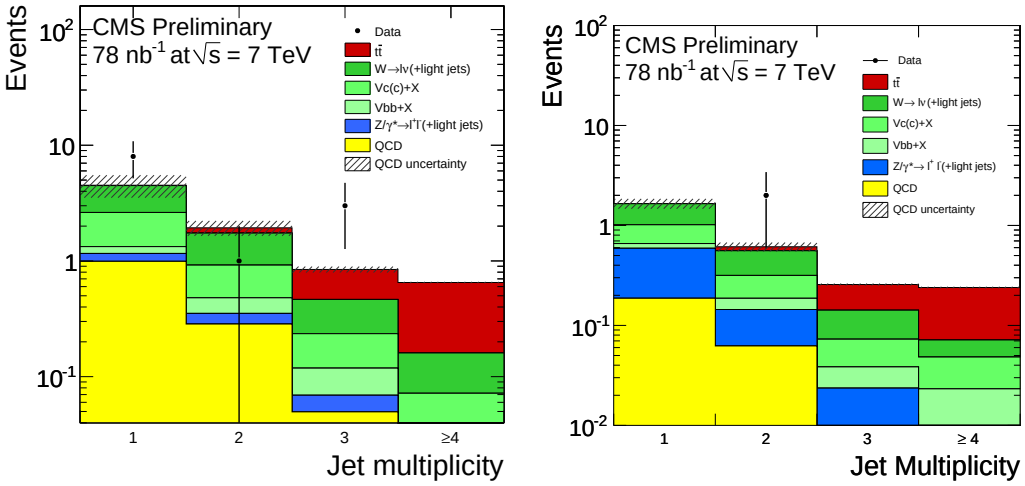

Figure 17: μ+jets channel: Distributions of the jet multiplicity for events containing at least one b-tagged jet, using the loose working point of the track-counting algorithm (left), as well as for events in which at least one jet is associated with a muon (right) are shown, corresponding to 78 nb^{-1}.

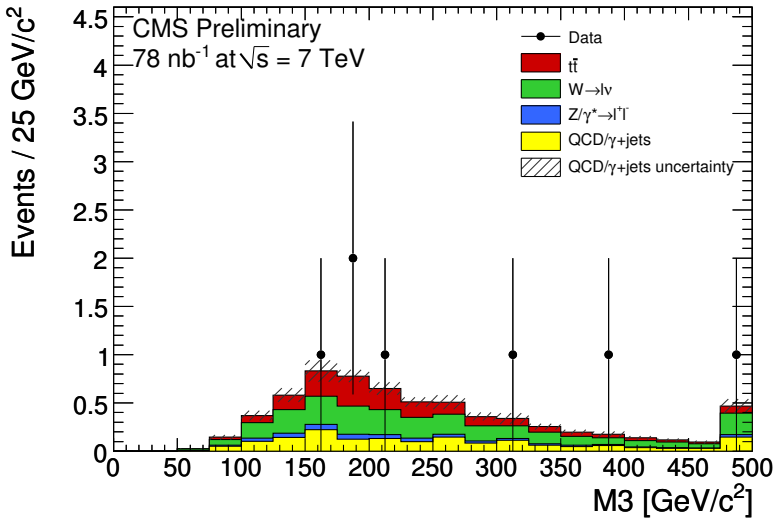

Figure 18: Distribution of the three-jet invariant mass M3 for events containing at least three jets, combined for both μ+jets and e+jets channels and corresponding to L $= 78$ nb^{-1}. Predictions from simulation are overlaid, as described in the text and normalized to the integrated luminosity of the data sample. The hatched error band corresponds to the uncertainty on the number of expected QCD multijet events. The last bin includes overflows.

7 Conclusions

The event selection criteria for early dilepton and lepton+jets $t\bar{t}$ analyses is designed to efficiently select signal events while suppressing contributions from background. The selected sample is not pure in signal. Therefore, techniques have been developed to accurately assess the constituents of the selected events.

The detailed study of event yields from the event selection as well their kinematic distributions are now possible in the data sample recorded so far at CMS. Substantive conclusions regarding the production of top quarks at the LHC can only be made with significantly more data, however, early indications are that the yields in data are reasonably represented by our predictions, and that both, the selection as well as the data-driven techniques developed over the past years based on simulated events, are performing as expected.

References

[1] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **0803** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[2] Tevatron Electroweak Working Group Collaboration, "Combination of CDF and D0 Results on the Mass of the Top Quark", arXiv:0903.2503.

[3] "LEP Electroweak Working Group", (2009). doi:http://lepewwg.web.cern.ch/LEPEWWG/.

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

[5] http://mcfm.fnal.gov/.

[6] R. Kleiss and W. J. Stirling, "Top quark production at hadron colliders: Some useful formulae", *Z. Phys.* **C40** (1988) 419–423. doi:10.1007/BF01548856.

[7] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph", *JHEP* **02** (2003) 027, arXiv:hep-ph/0208156.

[8] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[9] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270. doi:10.1109/TNS.2006.869826.

[10] A. D. Martin, W. J. Stirling, R. S. Thorne et al., "Uncertainties on $alpha_s$ in global PDF analyses and implications for predicted hadronic cross sections", *Eur. Phys. J.* **C64** (2009) 653–680, arXiv:0905.3531. doi:10.1140/epjc/s10052-009-1164-2.

[11] H.-L. Lai et al., "Uncertainty induced by QCD coupling in the CTEQ-TEA global analysis of parton distributions", arXiv:1004.4624.

[12] F. Demartin, S. Forte, E. Mariani et al., "The impact of PDF and α_s uncertainties on Higgs Production in gluon fusion at hadron colliders", *Phys. Rev.* **D82** (2010) 014002, arXiv:1004.0962. doi:10.1103/PhysRevD.82.014002.

[13] http://wwwhep.ucl.ac.uk/pdf4lhc/.

[14] J. M. Campbell, R. Frederix, F. Maltoni et al., "Next-to-Leading-Order Predictions for t-Channel Single-Top Production at Hadron Colliders", *Phys. Rev. Lett.* **102** (2009) 182003, arXiv:0903.0005. doi:10.1103/PhysRevLett.102.182003.

[15] J. M. Campbell and F. Tramontano, "Next-to-leading order corrections to W t production and decay", *Nucl. Phys.* **B726** (2005) 109–130, arXiv:hep-ph/0506289. doi:10.1016/j.nuclphysb.2005.08.015.

[16] J. M. Campbell, R. K. Ellis, and F. Tramontano, "Single top production and decay at next-to-leading order", *Phys. Rev.* **D70** (2004) 094012, arXiv:hep-ph/0408158. doi:10.1103/PhysRevD.70.094012.

[17] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through O(alpha(s)**2)", *Phys. Rev.* **D74** (2006) 114017, arXiv:hep-ph/0609070. doi:10.1103/PhysRevD.74.114017.

[18] CMS Collaboration, "Expectations for observation of top quark pair production in the dilepton final state with the early CMS data at $\sqrt{s} = 10$ TeV", *CMS PAS* **TOP-09-002** (2009).

[19] CMS Collaboration, "Prospects for the first Measurement of the ttbar Cross Section in the Muon-plus-Jets Channel at $\sqrt{s} = 10$ TeV with the CMS Detector", *CMS PAS* **TOP-09-003** (2009).

[20] CMS Collaboration, "Plans for an early Measurement of the ttbar Cross Section in the Electron-plus-Jets Channel at $\sqrt{s} = 10$ TeV", *CMS PAS* **TOP-09-004** (2009).

[21] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV", *CMS PAS* **MUO-10-002** (2010).

[22] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV", *CMS DP* **2010/032** (2010).

[23] CMS Collaboration, "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV", *CMS PAS* **EWK-10-002** (2010).

[24] CMS Collaboration, "MET Performance in Minimum-Bias and Jet Events from Proton-Proton Collisions at $\sqrt{s} = 7$ TeV", *CMS PAS* **JME-10-004** (2010).

[25] CMS Collaboration, "Jet Performance in pp Collisions at $\sqrt{s} = 7$ TeV", *CMS PAS* **JME-10-003** (2010).

[26] CMS Collaboration, "The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS", *CMS PAS* **JME-09-003** (2009).

[27] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063.

[28] CMS Collaboration, "Commissioning of b-jet identification with pp collisions at 7TeV", *CMS PAS* **BTV-10-001** (2010).

A Selected Candidate Events

This appendix contains the details of a few interesting event candidates which have been identified, not only in the initial dataset corresponding to 78 nb^{-1} of integrated luminosity, but also in more recent data.

A.1 Dimuon candidate event from 78 nb^{-1}

One candidate event in $\mu^+\mu^-$ final state, which is passing the event selections corresponding to Table 7 is discussed briefly below. This event has two opposite-sign muons (p_T = 53 and 40 GeV/c) with dimuon mass below the Z-veto region and one jet with p_T = 75 GeV/c). The event has missing transverse energy $\not{E}_T$ = 25 GeV. The jet passes the track-counting based loose b-tag with a mistag rate of about 10%, no secondary vertex tag is present. The invariant mass of the dimuon pair is approximately 34 GeV/c^2. Three views of this event are shown in Figs. 19, 20, and 21. This event was collected by CMS on June 29, 2010.

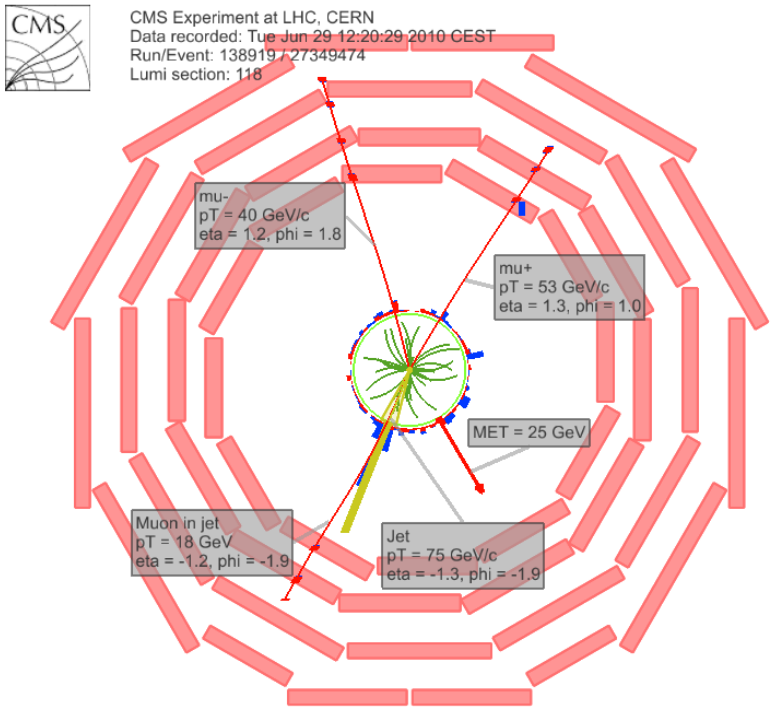

Figure 19: Event display of a candidate event with two muons and one jet in r-ϕ view. The jet is b-tagged with a loose track counting algorithm with about 10% mistag rate.

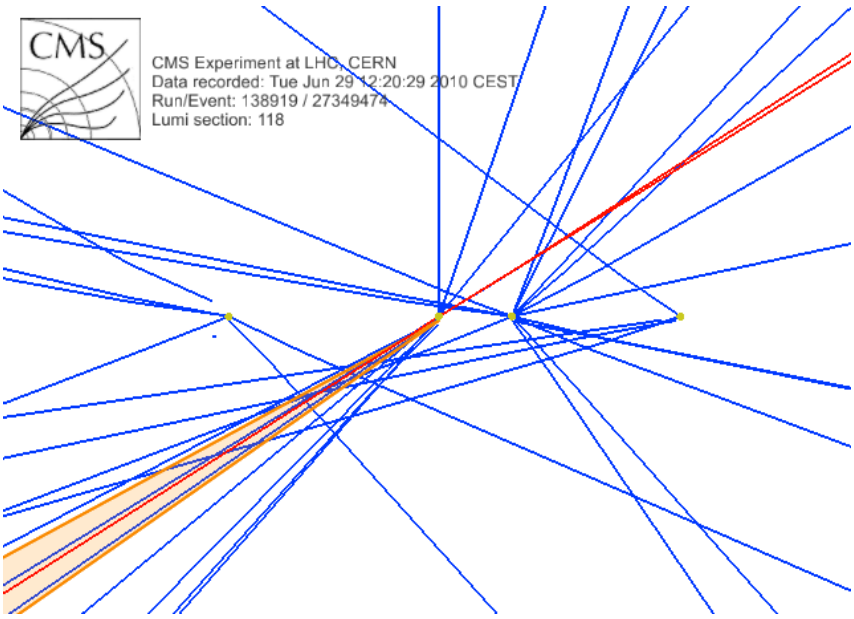

Figure 20: Event display of a candidate event with two muons and one jet in ρ-z view. Four primary vertices are clearly visible. The muon candidates and the jet all originate from the same primary vertex. Note that except for the right-most vertex, all tracks are required to have $p_{\mathrm{T}} > 1$ GeV/c to be able to clearly see the vertices.

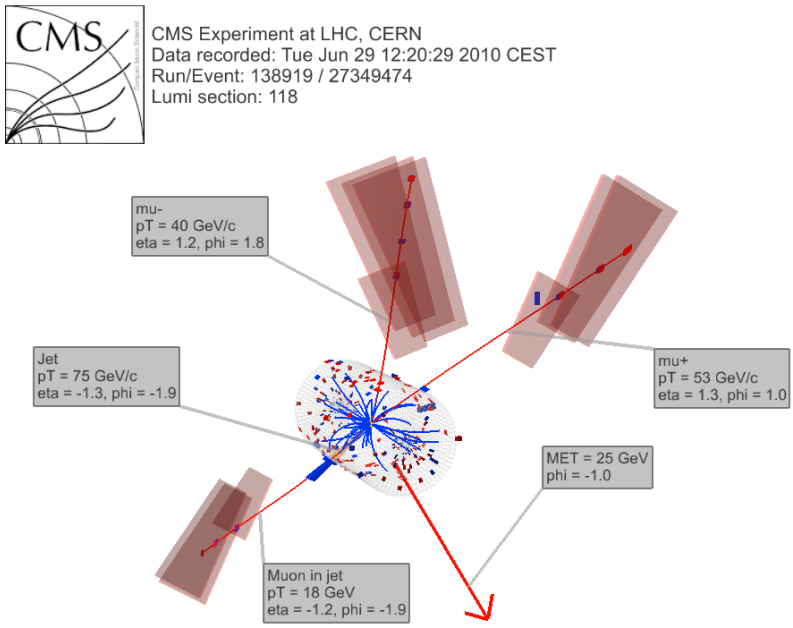

Figure 21: Event display of a candidate event with two muons and one jet in 3-dimensional view. Muon detectors with hits from the three muons in the event are highlighted. The view point is slightly from above, from the positive rapidity side.

A.2 Dimuon candidate with two b-tagged jets and missing transverse energy

One candidate event in $\mu^+\mu^-$ final state, which is passing full event selections[1] is discussed briefly below. This event has two opposite-sign muons (p_T = 57 and 27 GeV/c) with dimuon mass of 26 GeV/c, two b-tagged jets (p_T = 56 and 45 GeV/c), and $\not\!\!E_T$ of 57 GeV. The jets pass the track-counting and simple secondary vertex b-tag requirements. Preliminary calculations indicate that the reconstructed top quark masses found in this event lie in the region [160,220] GeV/c^2. Several views of this event are shown in Figures 22, 23, 24, 25, and 26. This event was collected by CMS on July 18, 2010.

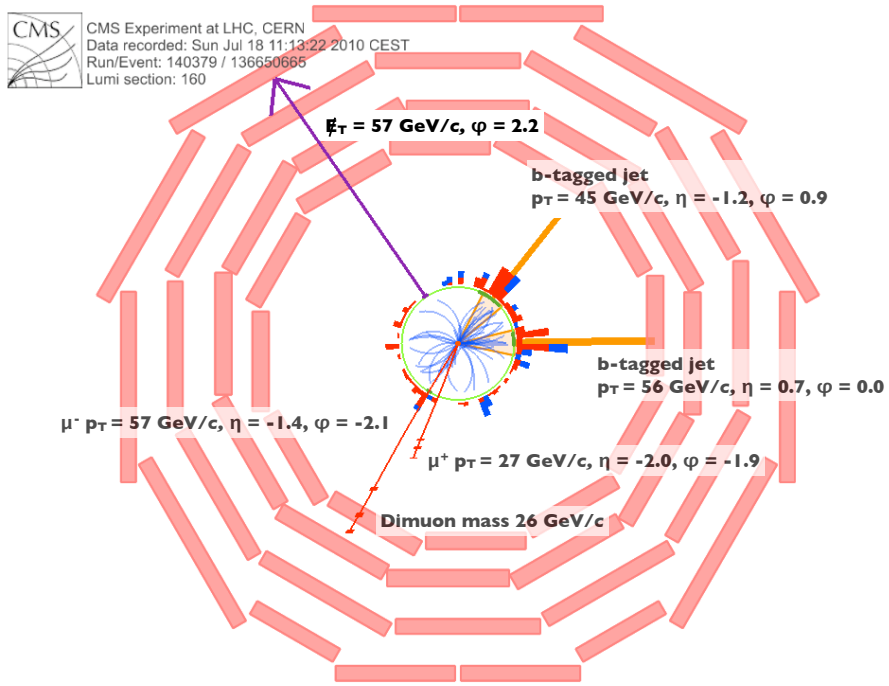

Figure 22: Event display of a candidate event with two muons, two jets, and missing transverse energy in r-ϕ view. Both jets are b-tagged. The event passes full event selections with a large margin.

[1]This event is coming in data sample collected after the dataset used for results presented in this note with 78 nb^{-1}.

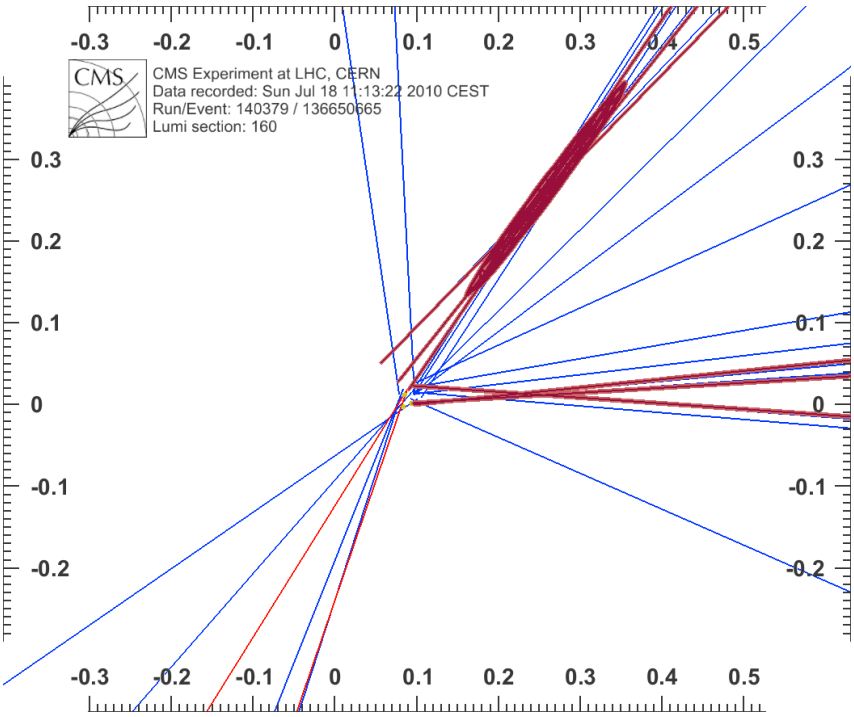

Figure 23: Event display of a candidate event with two muons, two jets, and missing transverse energy in r-ϕ view zoomed on the primary vertex. The horizontal and vertical scales are in centimeters. Both secondary vertices are highlighted and are clearly visible: the error ellipse corresponds to six standard deviations. Only tracks belonging to the candidate event primary vertex are displayed. The event passes full event selections with a large margin.

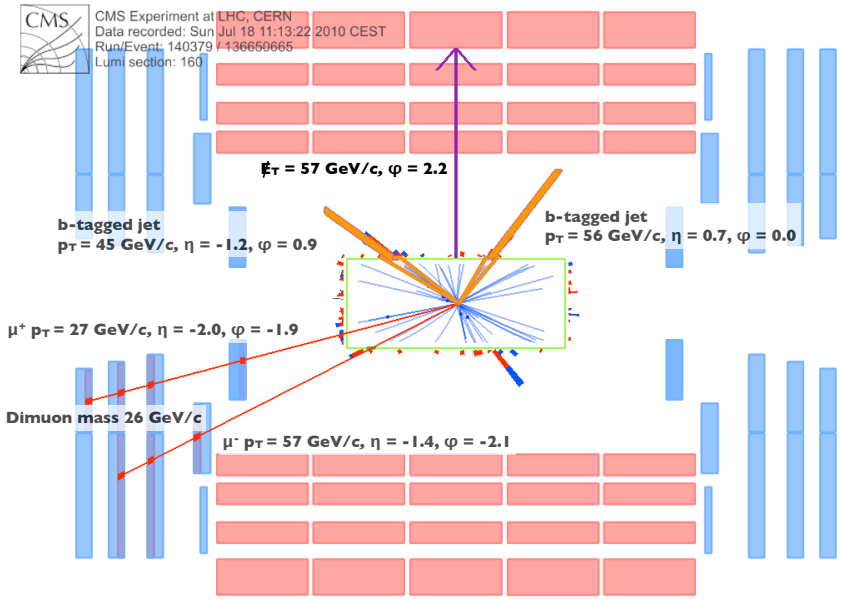

Figure 24: Event display of a candidate event with two muons, two jets, and missing transverse energy in ρ-z view. Both jets are b-tagged. The event passes full event selections with a large margin.

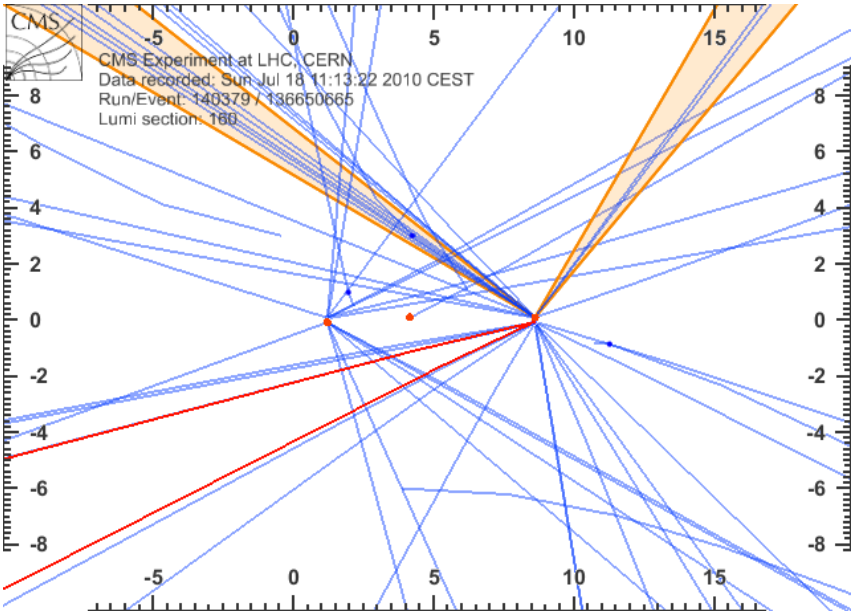

Figure 25: Event display of a candidate event with two muons, two jets, and missing transverse energy in ρ-z view zoomed to show three primary vertices in the event. Charged particle tracks with $p_{\mathrm{T}} > 0.75$ GeV/c are shown. The horizontal and vertical scales are in centimeters. Both jets belong to the same vertex as the muons. The event passes full event selections with a large margin. The multiple primary vertices are indicative of pileup, defined as multiple pp interactions in a given bunch crossing.

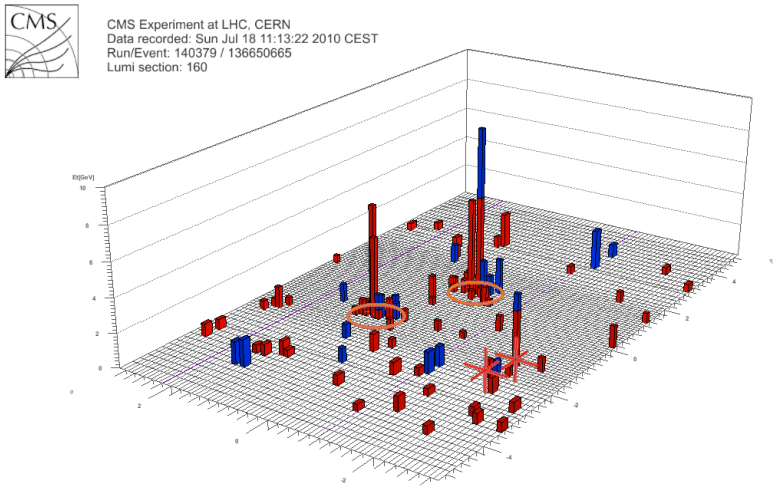

Figure 26: Event display of a candidate event with two muons, two jets, and missing transverse energy in calorimeter lego view. The jets passing event selection are highlighted with ellipses. The points of muons impact on the calorimeter surface are highlighted with crosses. The event passes full event selections with a large margin.

A.3 Electron+jets candidate event

In the e+jets analysis, we find one event with four jets, that passes the full event selection. Figure 27 shows the event display of this event in three different views. The event contains one isolated electron with $E_T = 44$ GeV, four jets with transverse momenta of 163, 70, 73 and 35 GeV/c, respectively. None of the jets is b-tagged. The missing transverse energy is $\not\!\!E_{\mathrm{T}}$=105 GeV, the transverse mass is 33 GeV/c^2, and the three-jet invariant mass $M3$=208 GeV/c^2. The event was recorded by CMS on July 2, 2010.

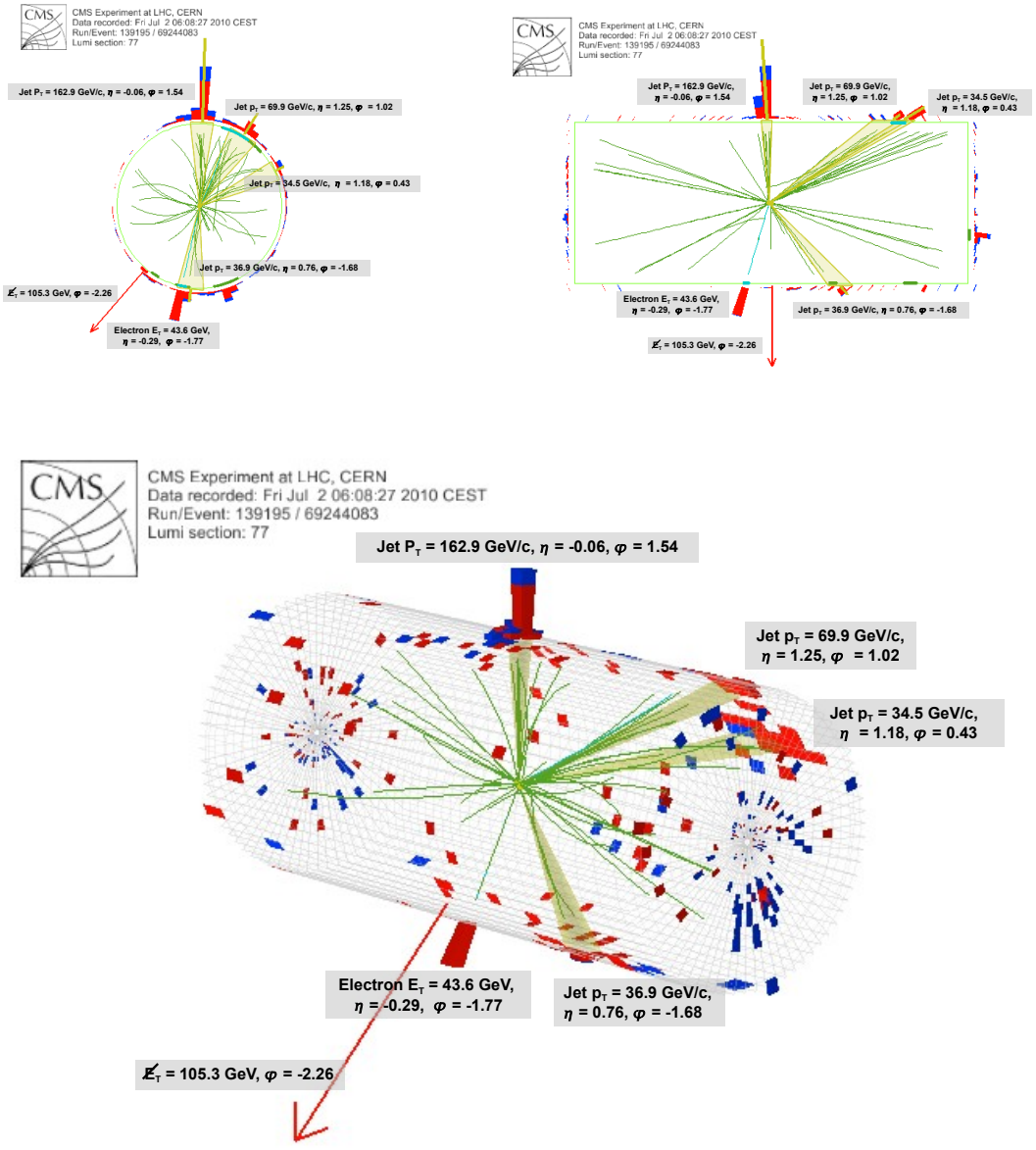

Figure 27: Event displays of the electron plus four jets event, shown in r-ϕ (top left), ρ-z (top right) and 3D view (bottom). Depicted are the four jets passing all quality criteria, the isolated electron and the missing transverse energy.

A.4 Electron+jets candidate with two b-tagged jets and missing transverse energy

Another candidate event in the e+jets mode passing the full event selection has been found[2] and is discussed briefly below. This event has one isolated electron with E_T =41 GeV, $\not\!\!E_T$=44 GeV, and four high p_{T} jets, with p_{T} = 109, 73, 68 and 61 GeV/c, among which two are b-tagged. The reconstructed transverse W mass is 77 GeV/c^2, the invariant mass of the untagged jets is 102 GeV/c^2, and the two possible hadronic top combinations, the 3-jet system comprised of the two untagged jet and either the highest or second highest p_{T} tagged jets, have masses 232 and 208 GeV/c^2, respectively. Four views of this event are shown in Figures 28, 29, 30, and 31. This event was recorded by CMS on July 18, 2010.

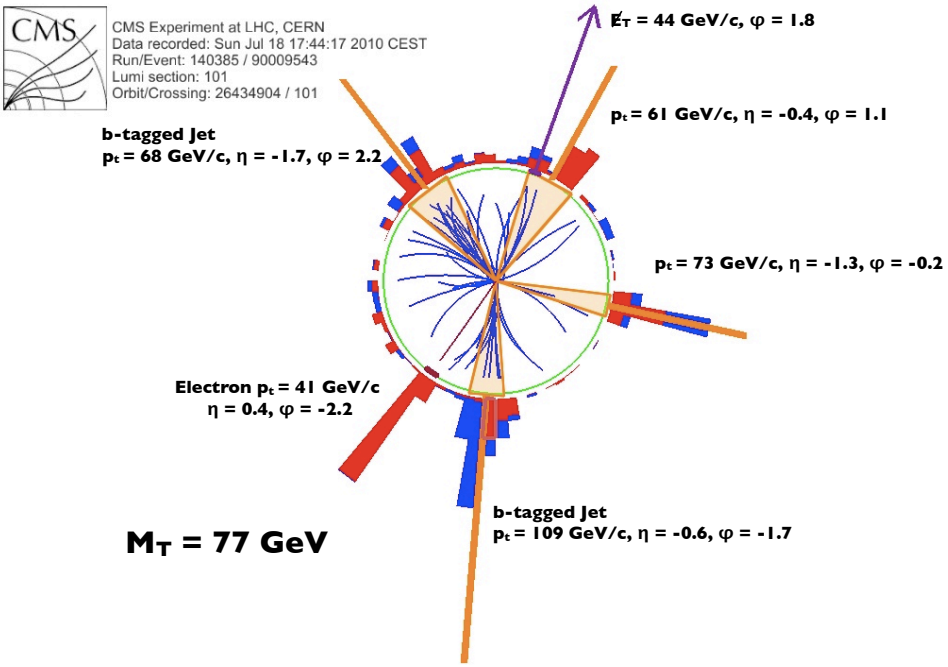

Figure 28: Event display of the r-ϕ view of a candidate event with one electron, four jets, and missing transverse energy. Two of the jets are b-tagged. The event passes the full e+jets event selection criteria.

[2]This event is coming in data sample collected after the dataset used for results presented in this note with 78 nb^{-1}.

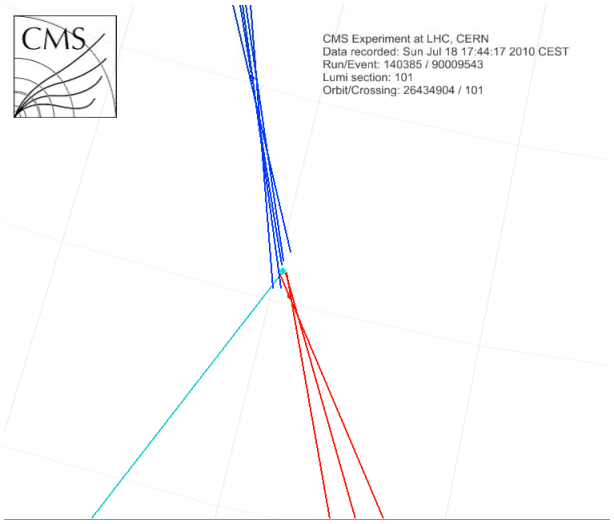

Figure 29: Event display closeup of secondary vertices of a candidate event with one electron, four jets, and missing transverse energy. The view is rotated in 3D to best depict the found vertices.

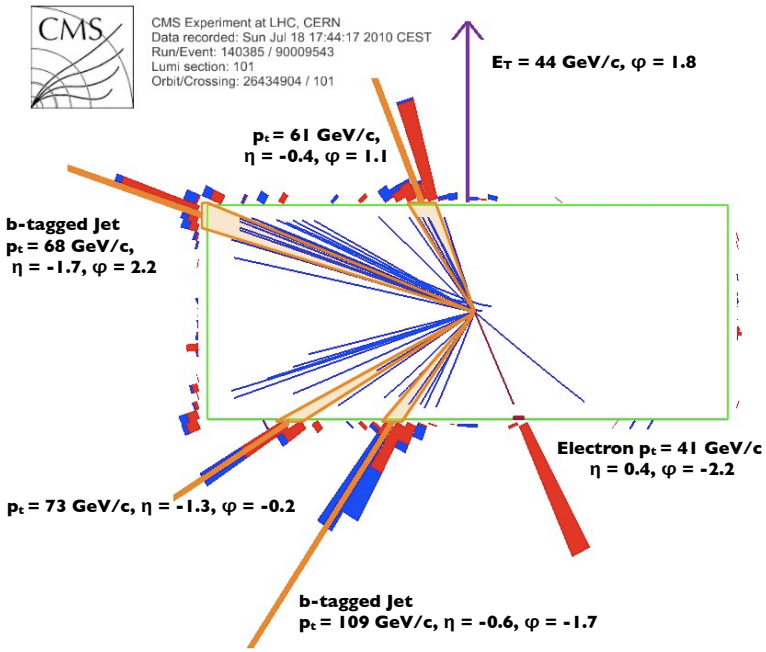

Figure 30: Event display of the ρ-z view of a candidate event with one electron, four jets, and missing transverse energy. Two of the jets are b-tagged. The event passes the full e+jets event selection criteria.

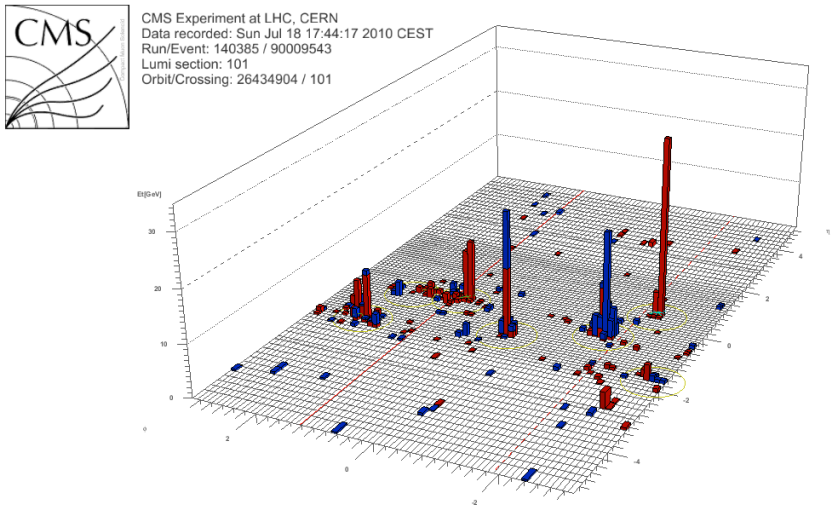

Figure 31: Event display of a candidate event with one electron, four jets, and missing transverse energy in calorimeter lego view.

A.5 Muon+jets candidate with one b-tagged jet and missing transverse energy

A candidate event in the μ+jets mode passing the full event selection has been found[3] and is discussed briefly below. This event has one isolated muon with $p_{\mathrm{T}} = 31$ GeV/c, large $\not\!\!E_T = 119$ GeV, and four high p_{T} jets, with p_{T}'s of 152, 82, 57 and 43 GeV/c among which the second highest p_{T} jet is b-tagged. The transverse mass, calculated from the lepton and neutrino measurements in the transverse plane, is 104 GeV/c^2, consistent with originating from a leptonically decaying W. Among the three untagged jets, three combinations for the hadronic W are possible. These three candidates have mass values of 104, 105, and 151 GeV/c^2. The hadronic top decay can be constructed from combining the various hadronic W candidates and the remaining jet; these combinations yield a most-likely hadronic-side top mass value of 210 GeV/c^2. Three views of this event are shown in Figures 32,33, and 34. This event was recorded by CMS on July 14, 2010.

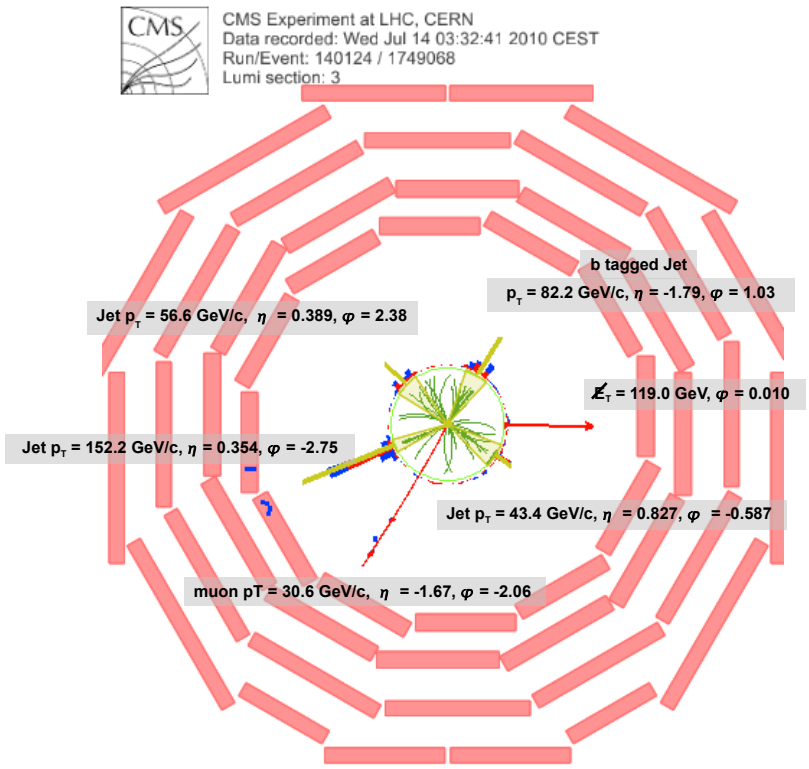

Figure 32: Event display of the r-ϕ view of a candidate event with one muon, four jets, and missing transverse energy. One of the jets is b-tagged. The event passes the full μ+jets event selection criteria.

[3]This event is coming in data sample collected after the dataset used for results presented in this note with 78 nb^{-1}.

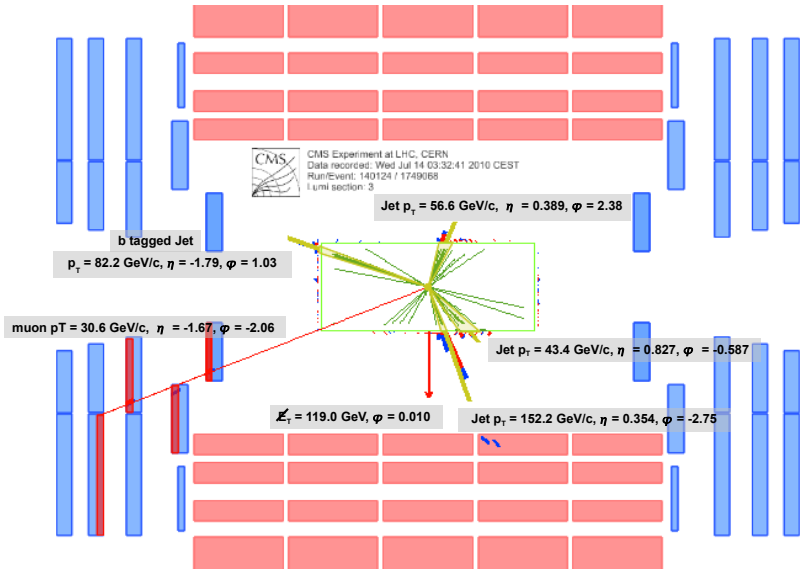

Figure 33: Event display of the ρ-z view of a candidate event with one muon, four jets, and missing transverse energy. One of the jets is b-tagged. The event passes the full μ+jets event selection criteria.

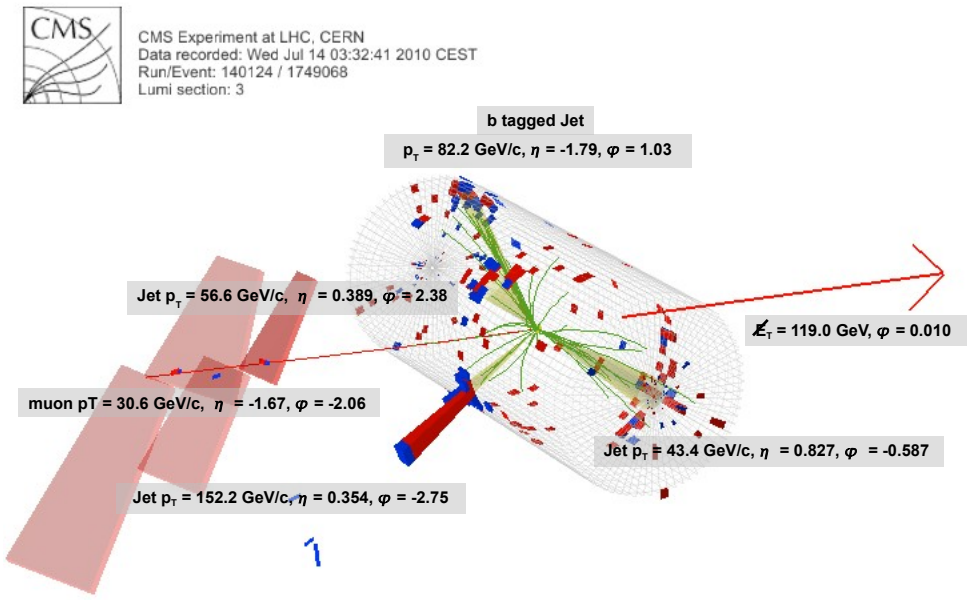

Figure 34: Event display of a candidate event with one muon, four jets, and missing transverse energy in 3D.