Available on CMS information server

CMS AN-2008/017

9 May 2008

Early measurements of top quark pair events in the $e + \mu$ channel with the first data of CMS

J. Caudron, A. Giammanco, D. Kcira, V. Lemaitre

Université Catholique de Louvain, Center for Particle Physics and Phenomenology, Louvain-La-Neuve, Belgium

Abstract

We report on a study aiming at an early observation of $t\bar{t}$ production in the di-leptonic decay channel with an electron and a muon in the final state, recorded by the CMS detector from pp collisions at a center-of-mass energy of $\sqrt{s} = 14$ TeV.

We propose two analysis strategies which make a minimal use of the HCAL information in order to be complementary to the standard approach based on calorimetric jets. These strategies rely on the assumption of a good lepton identification and are made more or less inclusive according to the stability and understanding of the different subdetectors. They are the “inclusive lepton method”, where a tight lepton selection is applied assuming a good control of the fake lepton rate, and the “track method”, where lepton selection cuts can be relaxed by means of using the Tracker information, including track-jets. One shows that, without using b-tagging, both methods can reveal the $t\bar{t}$ process with less than 10 pb^{-1}.

1 Introduction

Top quark pair production is the first complex process which we will observe at LHC, containing leptons, jets, missing energy and secondary vertices in the final state. Therefore the observation of these events in our data can be considered as a milestone in the physics commissioning of the experiment.

This note is focused on di-leptonic events in the $e + \mu$ channel. The aim of this note is to describe two alternative analyses designed for a start-up scenario, with the goal of "re-discovery" with the lowest possible integrated luminosity. In short, both analyses (described in detail in Sec. 4) make a minimal use of the HCAL information. The rationale behind this is to provide complementarity with respect to more traditional analyses which make use of calorimetric jets (see, for example, Refs. [1, 2]), since with very early data it may happen that misfunctioning of one specific sub-detector, or of part of it, obliges to exclude some runs from the analysis. One of the two analyses is even robust against a partial malfunctiong of the Tracker, since it only uses tracks for lepton reconstruction and selection. Having back-up strategies based on different sub-detectors might be crucial at start-up.

The use of HCAL is not completely excluded in this note: for example, we decided to make use of calorimetric isolation for leptons (although track-based isolation is also used, and can be used alone as a fall-back solution), and of the ratio between the HCAL and ECAL deposition for electron candidates (H/E), since these are local variables, as opposed to global variables like the number of calorimetric jets, or H_T, or the missing transverse energy. As an example, the presence of noisy calorimetric cells giving fake jets would have a small impact on our strategy, while it would affect any analysis based on the calorimetric variables listed above.

The two analyses presented in this note take into account the expected quality of the alignment and calibration at start-up (the "0 pb^{-1} scenario" [3]). The sensitivity of the analysis to mis-alignment and mis-calibration is discussed in Section 9.

During the start-up period of a new experiment it takes some time to understand the performance of the detector and the quality of event reconstruction. To reach a sufficient level of understanding as soon as possible, we employ simple analysis methods. We don't rely on b-tagging, nor missing energy, and we don't apply any kinematic fit.

This paper is organized as follows: Sec. 2 describes the samples used for this study, and discusses their eventual limitations; considerations about the trigger selection are presented in Sec. 3; the two analysis strategies are described in Sec. 4; the estimation from MC of the QCD multi-jet background is in Sec. 5, while a discussion on $ZQ\bar{Q}$ and what can be inferred about $WQ\bar{Q}$ is contained in Sec. 6; systematic uncertainties are discussed in Sec. 7; the final results are presented in Sec. 8; Sec. 9 discusses the robustness of the two analyses.

2 Data samples

These studies are designed for a data sample with an integrated luminosity of 10 pb^{-1}, in a "0 pb^{-1}" detector scenario.

The signal is $t\bar{t}$ with both the W bosons decaying into leptons, and the final state is one electron and one muon.

The backgrounds explicitely considered in this note are the following:

- W+jets,
- γ^*/Z+jets (the short-hand label "$Z + jets$" will be used in the following),
- dibosons (WW, WZ, ZZ),
- other $t\bar{t}$ decays,
- $ZQ\bar{Q}$ (see Sec. 6 where they will be also be used for a rough estimation of $WQ\bar{Q}$),
- QCD multi-jet background (see Sec. 5).

All these processes, with the exception of the diboson processes (WW, WZ, ZZ), are coming from samples contained in the CSA07 soups called "Chowder" and "Stew" [4], generated under "0 pb^{-1}" mis-alignment and mis-calibration conditions [3]. The diboson samples were produced as part of the "CSA07 Signal production", with "100 pb^{-1}" conditions. It has to be noted that these events mostly pass the selection through the presence of real leptons, so the use of conditions closer to ideal results in an over-estimation of this background component. In

Sec. 9 the sensitivity of the analysis to different detector conditions is explored by comparing the Chowder sample with an equivalent one in "10 pb^{-1}" conditions.

For the $t\bar{t}$, W/Z+jets and diboson samples the ALPGEN [12] generator was used, interfaced to PYTHIA [13] for showering and hadronization, while the QCD samples were entirely generated with PYTHIA (more details in Sec. 5). All samples were passed through the detector simulation, based on GEANT4 [14] ("full simulation").

The list of samples used in this study is summarized in Tab. 1. The cross section for the $t\bar{t}$ process is normalized to 833 pb [15], which corresponds to a NLO/LO K-factor of 1.85 with respect to the ALPGEN cross sections. An NLO/LO K-factor of 1.12 is uniformly applied on the W/Z+jets cross sections. Diboson cross sections are at LO.

No pile-up was included in these samples.

The DiLeptonMuonX skimming [5] is applied on the CSA07 soups samples. This skim is based on HLT1MuonIso, HLT2MuonNonIso and HLTXElectronMuonRelaxed triggers, on top of which it also requires at least two leptons with $p_T > 20$GeV.

3 Trigger aspects

In the DiLeptonMuonX skimming used for most samples (see Sec. 2) only events passing the HLT1MuonIso, HLT2MuonNonIso and HLTXElectronMuonRelaxed triggers are selected.

In this note it is assumed that the first ≈10 pb^{-1} will be recorded with a very low instantaneous luminosity in the range $10^{29} - 10^{31}$ $cm^{-2}s^{-1}$ (which also justifies our assumption of no pile-up), so all trigger thresholds will be relatively low (see, for example, the trigger tables proposed for 10^{30} $cm^{-2}s^{-1}$ and 10^{31} $cm^{-2}s^{-1}$ in Ref. [7]). In particular, our lepton p_T thresholds (defined in Sec. 4) are quite higher than those proposed for $L = 10^{32}$ $cm^{-2}s^{-1}$ in Ref. [8], so turn-on efficiency curves should not be an issue for the study presented here.

When real data will be available, this analysis will make use of the events selected by the single-electronic, single-muonic and electron plus muon HLT bits. An extensive study of the trigger efficiencies and turn-on curves for top events in single- and di-leptonic triggers may be found in Ref. [9], for the trigger table of Ref. [8].

4 Di-leptonic event selection

Two analysis strategies are proposed:

- A) **Inclusive lepton method.** This relies on tight lepton selection cuts. The expected number of signal events will be small, and the goal is to keep the background contamination to a minimum. Further selection can be based on the invariant mass of the di-leptonic system.
- B) **Track method.** The lepton selection cuts are relaxed, jets are only clusterized from charged tracks, and topological variables (aplanarity, sphericity, . . .) are built using the tracks in the event.

In the rest of the note they will be referred as "Method A" and "Method B" respectively.

4.1 Common preselection

We use the Top Quark Analysis Framework (TQAF) [10] in version TQAF_169_080212, with a preselection requiring at least one electron and one muon of opposite charge with $p_T > 18$ GeV/c, $|\eta| < 2.4$, Tracker isolation < 3 GeV/c and calorimeter isolation < 5 GeV for both leptons. It has to be reminded that in all the samples, with the exception of WW,WZ and ZZ, the DiLeptonMuonX skimming is applied, which implies that the two leptons have $p_T > 20$ GeV/c.

One critical aspect of these analyses is the reduction of leptons not coming from W decay. In fact, events of the category $W + jets$ in general pass the selection when one lepton comes from W decay and the second comes from one of the following cathegories: heavy quark decay, a misidentified hadron, a charged + neutral hadron coincidence (faking an electron), an electron from photon conversion, or a decay in flight of a light hadron. For the sake of simplicity, all these different lepton categories will be called "fakes" in the following. The amount of fakes, at the relatively high momenta considered here, is expected to be higher for electrons than for muons, so special care is given in this study to the former.

4.2 Method A

This method is based on tight leptonic cuts.

In order to reduce the probability to select a fake electron candidate, the electron-ID cut-based tool provided by the electron/gamma physics object package is used [11]. The cuts used in this study correspond to the "robust" option, expressely designed for start-up conditions, and are listed in Table 2.

The two selected leptons are the highest-p_T electron and the highest-p_T muon in the event, with $p_T(e) > 40$ GeV, $p_T(\mu) > 35$ GeV.The distributions of these variables are shown in Fig. 1.

The set of cuts discussed so far does not prevent the eventuality that a muon, radiating an almost collinear photon, passes the electron identification criteria. Even worse, it would be double counted both as a muon and as an electron. In this note the problem is dealed by just ignoring electron candidates within a cone of 0.3 from an identified muon. The efficacity of this cut can be appreciated from Fig. 2.

The selected events after preselection (including skimming), lepton selection and double-counting removal are listed in Table 3, and give the statistical significance shown in Fig. 3 as a function of integrated luminosity. The systematic uncertainties considered are described in Section 7 and are summarized for each process in Table 4.

4.3 Method B

4.3.1 Lepton identification and selection

The leptonic part of the selection differs from Method A (Sec. 4.2) only in the transverse momentum thresholds: $p_T(e) > 25$ GeV, $p_T(\mu) > 20$ GeV.

4.3.2 Track-based topological selection

The standard tracking [19] is used, with a selection tightened by requiring at least 8 valid hits per track.

Only tracks external to a 0.3 cone centered in the two selected leptons are taken into account, in order to be independent from the lepton selection.

The following kinematic variables are defined using track information only:

- aplanarity: $3/2 \times Q_1$, where Q_1 is the lowest eigenvalue of the 3D matrix $\mathcal{M} = \sum_{trk} M^{trk}$, where $M_{ij}^{trk} = p_i^{trk}.p_j^{trk}$ $(i, j = x, y, z)$, and p^{trk} is the track momentum;
- sphericity: $3/2 \times (Q_1 + Q_2)$, where Q_1 and Q_2 are the two lowest eigenvalues of $\mathcal{M}$;
- circularity: $3/2 \times (q_1 + q_2)$, where q_1 and q_2 are the two lowest eigenvalues of a 2D matrix defined as $\mathcal{M}$ but with $i, j = x, y$ only;
- centrality: the ratio between the scalar sum of the transverse momenta of tracks and the modulus of the vectorial sum of their momenta;
- number of tracks with p_T higher than 5 GeV/c (N_{trk}),
- scalar sum of the transverse momentum of tracks with p_T higher than 5 GeV/c ($\sum_{trk} p_T$).

The distributions of these variables after preselection are shown in Fig. 4, normalised to the expected number of signal and background events in 10 pb^{-1}.

A cut is applied on the number of tracks with p_T higher than 5 GeV/c ($N_{trk} \geq 2$). The remaining variables are strongly correlated, but they give a good S/B separation when combined in a multivariate analysis, as demonstrated in Fig. 5 which shows, after preselection, the output of a simple linear discriminant convoluted with a sigmoid (technically: a Neural Network with only one node). This selection is not applied in this analysis, since it gives no gain in significance after our previous cuts, and it is only shown here as an illustration of a possible alternative selection.

4.3.3 Tracker-based jet clustering

Jets are built by using the track collection as input. We will denote them as "track-jets" in order to emphasize the difference with respect to the usual calorimetric jets. In principle a refinement of the method can come by the restriction to only those tracks whose distance of closest approach to the primary vertex of the event is smaller than a certain cut; this is not considered necessary here, since this note deals with a low luminosity scenario where pile-up is considered negligible.

Both the k_T [20, 21] (in the FastJet [22] implementation) and IterativeCone algorithms are taken into consideration, with different parameters (D for k_T and R in the IterativeCone case, both scanned in the range $[0.4 - 0.9]$, with steps of 0.1). The event is rejected if the number of jets is less than two. Figure 6 shows the distribution of p_T for the leading and second-leading track-jets in the event.

The number of track-jets per event, for k_T with $D = 0.7$, with $p_T(j) > 20$ GeV, is shown in Fig. 7(right).

Jets within $\Delta R < 0.3$ from an identified lepton are excluded from the count. This excludes the possibility of double counting a lepton as a track-jet.

4.3.4 Results for Method B

The results, by applying the cuts in cascade, are shown in Table 5, and their uncertainties are listed in Table 6. The significance with and without systematic uncertainties, as a function of integrated luminosity, is shown in Fig. 8.

5 Estimation of the QCD background

In the "Stew" MC sample (filtered QCD) used in this note, only one event is selected by Method A and none survives the cuts of Method B (although the same event passes some steps of the latter selection). This event happens to contain bottomonium, while no generic QCD multi-jet event passes our preselection (see Sec. 4.1). Nevertheless, the limited MC statistics means that we cannot put very stringent limits on the expectation for this background.

The following subsections show different approaches used to improve our prediction of QCD contamination from Monte Carlo.

5.1 Cut factorization, QCD multi-jet

In order to cope with that, a factorized approach was applied using "PreCSA07" QCD samples (produced with CMSSW_1_5_2), see Table 7. Assuming that the two leptons are uncorrelated, the efficiency to have a $e\mu$ final state is given by the product of the efficiency the have one lepton. After lepton selection, without requiring opposite charge (which would unnecessarily reduce the available statistics by a factor of ≈ 0.5, depending on the amount of lepton pairs from $Q\bar{Q}$), the predicted number of events for Method A is found to be negligible (Table 8) while for Method B around three events survive after 10 pb^{-1} (Table 9).

In the estimation for Method B we didn't apply the cuts on event-shape variables and number of track-jets, which would probably provide a significant reduction. Nevertheless, to be conservative, we will use these numbers as an estimation of the error on this background component even after the complete list of cuts (see Sec. 7 for details).

5.2 Cut factorization, $b\bar{b}$

The samples used in the previous subsection are known to underestimate the heavy flavour component. For this reason, at the time of writing several $b\bar{b} + Nj$ samples are under production [6], but only one ($b\bar{b} + 1j$) is already available at RECO level. Both Methods A and B give no surviving event from this sample, so the same factorization of the leptonic cuts has been applied, and the result is in Table 10. It has to be noted that in this case the assumption of independence is less justified due to the presence of $b\bar{b} \to e\mu$ (directly from $b \to l$ or from the chain $b \to c \to l$), with a branching ratio of the order of percent. Anyway, the results seem to indicate that this background component is orders of magnitude away from being dangerous.

5.3 μ-enriched QCD sample

Some specific μ-enriched QCD samples have been generated as part of the CSA07 production. Table 11 shows the results obtained by running the Method A and B selections on dataset /ppMuPt20-15/CMSSW_1_6_7-CSA07-

1205151815/AODSIM, whose statistics (almost two million events) corresponds to 8.7 pb^{-1}, and is defined by the presence in each event of a real muon with $p_T^{gen} > 20$ GeV, and a $\hat{p_T} > 15$ GeV for the partonic interaction. Again, the amount of contamination seems not to pose a danger.

It has been estimated [24, 1] that real muons account for the vast majority of the selected muons in QCD events, especially after quality cuts which remove most fakes from punch-through and hadron decays. It is interesting to note that the visible cross section obtained from this sample is compatible with the estimate from the factorization of cuts discussed in Sec. 5.1.

6 Estimation of the $ZQ\bar{Q}$ background

In the context of the "CSA07 Signal production" [6] some $Zb\bar{b}{+}jets$ and $Zc\bar{c}{+}jets$ were produced with ALPGEN, divided by light jet multiplicities (see Table 12). The detector scenario is "100 pb^{-1}". A completely analogous production of $Wb\bar{b} + jets$ and $Wc\bar{c} + jets$ was foreseen but is not yet available at the time of writing, due to unsolved technical problems which makes the timescale of their availability completely unpredictable.

It has to be noted that a straightforward inclusion of these backgrounds in the analysis would lead to double counting, since a certain amount of heavy flavour content is already present in the $Z + jets$ and $W + jets$ samples. However, we opt for an explicitely conservative approach by ignoring this double counting since this background component is found to be negligible, as shown in Table 13.

By summing all the jet multiplicities, it can be seen that the probability of survival of a $Zb\bar{b}$ event is $\approx 10^{-6}$ in Method A and $\approx 2 \times 10^{-4}$ in Method B. Let's assume, for simplicity, that $\epsilon_{Wbb}/\epsilon_{Zbb} = \epsilon_{W+jets}/\epsilon_{Z+jets}$; from Tables 3 and 5 it can be seen that, for Method A, $\epsilon_{W+jets}/\epsilon_{Z+jets} \approx (3 \times 10^{-7})/(8 \times 10^{-6}) \approx 0.04$, while for Method B no $W + jets$ survives, so we take the efficiency of the next-to-last cut: $\epsilon_{W+jets}/\epsilon_{Z+jets} \approx (5 \times 10^{-7})/(5 \times 10^{-5}) \approx 0.01$.

These numbers, multiplied by a $Wb\bar{b} + X$ cross section of 19.9 pb [6] and by the survival probabilities for $Zb\bar{b}$, result in an expectation of 8×10^{-7} pb and 4×10^{-5} pb for Methods A and B respectively (i.e. 8×10^{-6} and 4×10^{-4} events after 10 pb^{-1}).

Thus we decide to neglect both $Wb\bar{b}$ and $Zb\bar{b}$ in the following.

7 Systematic Uncertainties

7.1 Monte Carlo statistics

The effect of the finite statistics of our MC samples is visible in the first row of Tables 4 and 6 (for Method A and B respectively).

7.2 Background normalization

We conservatively apply a 100% uncertainty on all the backgrounds (with the exception of "Other $t\bar{t}$", since its normalization is the same as the signal, being negligible the uncertainty on the branching ratios of the W) although the theoretical uncertainty on $W/Z + jets$ is much smaller [18]. But it has to be remarked that in the present analysis we are not properly considering, for lack of samples, the heavy flavour component of jets.

The effect can be seen in the second row of Tables 4 and 6 (for Method A and B respectively).

7.3 QCD contamination

In Sec. 5 we estimated the number of QCD events surviving the lepton selections of Method A and B by a factorization of the cuts. For the computation of the significance we decided to set N_{qcd} to 0, since no events survive in our Monte Carlo, but to conservatively consider ΔN_{qcd} equal to the sum of events in Tables 8 and 9 (for Method A and B respectively.)

8 Results

The final variables are the dileptonic invariant mass ($M_{e\mu}$) in the Method A and the number track-jets (k_T algorithm, D parameter equals to 0.7) with E_T higher than 20 GeV per event in the Method B. These variables, and the pseudo-experiment data for 10 pb^{-1}, are shown in Fig. 7.

No cut is applied on $M_{e\mu}$, while the requirement of at least two track-jets is part of the Method B selection. The events with 0 and 1 track-jets will be used for background control.

The final significances with and without systematic uncertainties, as a function of integrated luminosity, are shown in Figs. 3 and 8 for Methods A and B respectively.

9 Robustness tests

9.1 Sensitivity to misalignment and miscalibration

In order to test the robustness of our selections against different detector conditions, we run on the CSA07 Chowder soup with a 10 pb^{-1} detector scenario. Since a particular concern is the effect of misalignment on Method B, where tracks are of paramount importance, Fig. 9 and Fig. 10 compare, respectively, the number of tracks and the number of track-jets, i.e. the selection variables specific of Method B, between the two scenarios. Table 14 shows the difference in efficiency for both Methods. The conclusion is that in both Methods A and B an effect is visible, but not such to endanger the isolation of a clear signal.

9.2 Sensitivity to lepton and charged track momentum scale

We apply independently a 5% variation on the p_T of electrons, muons and other charged tracks. This number is expected to be much larger than the actual uncertainties at start-up.

The result of this exercise can be found in Table 15 for both Methods A and B. The lesson learnt is that these two analyses don't seem to depend very critically on the momentum scales of the physics objects.

10 Conclusions

In this note we explore two simple analysis methods that can be used to identify top pair events in the first CMS data.

Our preliminary results seem to indicate that an evidence at $\approx 4 - 5\ \sigma$ can be reached after 10 pb^{-1} with our Methods A and B (based, respectively, on lepton identification only, and on the use of Tracker-based variables), plateauing at around 7 σ and $8.5\ \sigma$ for Method A and B respectively. These numbers have to be seen just as rough indications, since many important elements are missing (see next subsection for a list of what we believe are the most important).

It has to be kept in mind that 1 pb^{-1} can be collected in roughly one week at a luminosity of $10^{30}\ cm^{-2}s^{-1}$, and in one day at $10^{31}\ cm^{-2}s^{-1}$.

10.1 What is missing

We consider this note simply as a step towards the preparation to a complete analysis. The most important missing elements, in our view, are:

- a proper simulation of the heavy flavour component in the backgrounds, namely the inclusion of $WQ\bar{Q}$ samples ($Q = b, c$), with of course a proper treatment of the double counting with respect to the $W + jets$ samples;
- a study of systematic effects of physical origin: most MC parameters come from fits to data at lower energy, and have very large uncertainty at 14 TeV at the start-up, before detailed studies are able to constrain them from LHC data; in particular we would like to estimate the sensititivity of our analysis strategies to the Underlying Event modeling;
- although a constant assumption throughout this note has been that pile-up will be negligible during this early phase, it is nevertheless possible that a relatively high instantaneous luminosity will be reached soon enough to have to deal with pile-up already in the analysis of the first inverse picobarns; we note that our Method B can be easily made robust against pile-up, by two means:

– the introduction of a primary vertex constraint in the track selection;
– the use of a feature of the FastJet algorithms, which permits to estimate and subtract pile-up (as well as contributions from Underlying Event processes) on an event-by-event basis (see Ref. [23]).

Acknowledgments

Many thanks, for several useful discussions and advices, to Jorgen D'Hondt, Claudio Campagnari and Martijn Mulders.

References

[1] **CMS Analysis Note 2008/015**, C.Campagnari et al., *" Expectations for $t\bar{t} \to$ dileptons in the early phase of CMS "*.

[2] **CMS Analysis Note 2008/016**, J.Andrea et al., *" Measurement of the Top Quark Pair Production Cross Section with $\mathcal{L} = 100\ \mathrm{pb}^{-1}$ in Dileptonic Final States using the CMS Detector"*.

[3] https://twiki.cern.ch/twiki/bin/view/CMS/DetectorPerformanceCSA07AlCaConstants

[4] https://twiki.cern.ch/twiki/bin/view/CMS/GeneratorProduction2007CSA07

[5] https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopQuarkCSA07

[6] https://twiki.cern.ch/twiki/bin/view/Main/AlpgenSummer07

[7] https://twiki.cern.ch/twiki/bin/view/CMS/TSG_17_III_08

[8] **CERN/LHCC 2007/021**, The CMS Collaboration, *"CMS High Level Trigger"*.

[9] **CMS Internal Note 2008/xxx**, J.Cuevas, M.Felcini et al., *"Trigger studies for top quark events ...."*. **FIXME: as soon as available, fix title and note number.**

[10] **CMS Internal Note 2007/068**, M.Hansen et al., *The Top Quark Analysis Framework*.

[11] https://twiki.cern.ch/twiki/bin/view/CMS/ElectronID

[12] **JHEP 07 (2003) 001**, M.L.Mangano et al., *"ALPGEN, a generator for hard multiparton processes in hadronic collisions"*.

[13] **JHEP 05 (2006) 026**, T.Sjostrand et al., *"PYTHIA 6.4 physics and manual"*.

[14] **IEEE Transactions on Nuclear Science 53 No. 1 (2006) 270-278**, J.Allison et al., *"Geant4 developments and applications"*

[15] **Nucl.Phys. B529 (1998) 424-450**, R.Bonciani, S.Catani, M.L.Mangano, *"NLL resummation of the heavy-quark hadroproduction cross-section"*.

[16] **CMS Analysis Note 2007/026**, N.Adam et al., *"Towards a measurement of the inclusive $W \to e\nu$ and $Z \to e^+e^-$ cross sections in pp collisions at $\sqrt{s} = 14$ TeV"*.

[17] **CMS Analysis Note 2007/031**, N.Adam et al., *"Towards a measurement of the inclusive $W \to \mu\nu$ and $Z \to \mu^+\mu^-$ cross sections in pp collisions at $\sqrt{s} = 14$ TeV"*.

[18] **Phys. Rev. D 68 (2003) 094021 [arXiv:hep-ph/0308195]**, J.Campbell, R.K.Ellis and D.L.Rainwater, *"Next-to-leading order QCD predictions for $W + 2j$ and $Z + 2j$ production at the CERN LHC"*.

[19] **CMS NOTE 2006/041**, W. Adam, B. Mangano, Th. Speer, T. Todorov, *Track Reconstruction in the CMS tracker*.

[20] **Nucl. Phys. B 406 (1993) 187**, S.Catani, Y.L.Dokshitzer, M.H.Seymour and B.R.Webber, *Longitudinally invariant K_T clustering algorithms for hadron hadron collisions*.

[21] **Phys. Rev. D 48 (1993) 3160 [arXiv:hep-ph/9305266]**, S.D.Ellis and D.E.Soper, *Successive combination jet algorithm for hadron collisions*.

[22] **Phys. Lett. B 641 (2006) 57 [hep-ph/0512210]**, M.Cacciari and G.P.Salam, *"Dispelling the N^3 myth for the k_T jet-finder"*.

[23] **Phys.Lett.B 659 (2008) 119 [arXiv:0707.1378]**, M.Cacciari, G.P.Salam and G.Soyez, *"Pileup subtraction using jet areas"*.

[24] http://indico.cern.ch/getFile.py/access?contribId=0&resId=0&materialId=slides&confId=27655

Table 1: Monte Carlo samples used in this note.

Process	Dataset	Scenario	N_{evt}
$W + jets$ $Z + jets$ $t\bar{t} + jets$	/CSA07Muon/CMSSW_1_6_7-CSA07-Chowder-A1-PDMuon-ReReco-startup-Skims4-topDiLeptonMuonX/USER	0 pb^{-1}	71715
$B/B \rightarrow J/\psi$ QCD Bottomonium Charmonium e/μ enriched	/CSA07Muon/CMSSW_1_6_7-CSA07-Stew-A1-PDMuon-ReReco-startup-Skims4-topDiLeptonMuonX/USER	0 pb^{-1}	16594
WW incl.	/WW_incl/CMSSW_1_6_7-CSA07-1198096665/AODSIM	100 pb^{-1}	744261
WZ incl.	/WZ_incl/CMSSW_1_6_7-CSA07-1198096684/AODSIM	100 pb^{-1}	362291
ZZ incl.	/ZZ_incl/CMSSW_1_6_7-CSA07-1198096655/AODSIM	100 pb^{-1}	143113
$W + jets$ $Z + jets$ $t\bar{t} + jets$	/CSA07Muon/CMSSW_1_6_7-CSA07-Chowder-P1-PDMuon-Skims4-topDiLeptonMuonX/USER	10 pb^{-1}	725296

Table 2: Definition of the "robust" cuts for electron identification for electrons in barrel ($|\eta| < 1.479$) and in the endcap ($|\eta| > 1.479$) of the ECAL, from Ref. [11].

Cut	ECAL barrel	ECAL endcaps
H/E	< 0.115	< 0.150
$\sigma_{\eta\eta}$	< 0.0140	< 0.0275
$\|\Delta\eta_{in}\|$	< 0.090	< 0.092
$\|\Delta\phi_{in}\|$	< 0.0090	<0.0105

Table 3: Visible cross sections (in pb) after preselection and lepton selection in Method A. The uncertainties are only statistical. The HLT bits considered are listed in Sec. 3, and are already applied in the skim for Chowder samples.

	Dilept. $t\bar{t}$	Other $t\bar{t}$	WW	WZ	ZZ	$W + jets$	$Z + jets$	"Stew"
No selection:	10.28 (NLO)	822.7 (NLO)	114.3 (LO)	49.9 (LO)	16.1 (LO)	61.4×10^3 (NLO)	5.3×10^3 (NLO)	-
skim & preselection:	7.53 ± 0.18	0.07 ± 0.02	0.99 ± 0.04	0.14 ± 0.01	0.032 ± 0.006	0.88 ± 0.10	6.95 ± 0.31	0.013 ± 0.007
HLT:	-	-	0.84 ± 0.03	0.13 ± 0.01	0.029 ± 0.006	-	-	-
tight lepton selection:	3.168 ± 0.117	0.004 ± 0.004	0.31 ± 0.02	0.041 ± 0.006	0.011 ± 0.003	0.029 ± 0.017	0.106 ± 0.034	0.004 ± 0.004
double-counting removal:	3.160 ± 0.117	0	0.31 ± 0.023	0.039 ± 0.006	0.011 ± 0.003	0.020 ± 0.014	0.044 ± 0.023	0.004 ± 0.004

Table 4: Statistical error (MC) and background normalization error after Method A selection.

	signal	$W + jets$	$Z + jets$	WW	WZ	ZZ	Other $t\bar{t}$	"Stew"
$\Delta N_{\text{MC}}/N$	3.7%	85%	77%	7.4%	15%	29%	-	-
$\Delta N_{\text{norm}}/N$	0	100%	100%	100%	100%	100%	0	100%

Table 5: Visible cross-section (in pb) after the cuts in Method B. The uncertainties are only statistical.

	Dilept. $t\bar{t}$	Other $t\bar{t}$	WW	WZ	ZZ	$W + jets$	$Z + jets$	“Stew”
lepton selection:	6.520±0.168	0.022±0.010	0.71±0.03	0.097±0.010	0.020±0.005	0.312±0.061	2.608±0.190	0.004±0.004
double-counting removal:	6.508±0.168	0.017±0.009	0.70±0.03	0.094±0.010	0.020±0.005	0.303±0.060	1.614±0.151	0.004±0.004
$N_{trk} \geq 2$	5.982±0.161	0.017±0.009	0.18±0.02	0.040±0.006	0.019±0.005	0.033±0.019	0.255±0.052	0.004±0.004
$N_{trk-jets} \geq 2$	4.833±0.145	0.013±0.008	0.043±0.008	0.024±0.005	0.011±0.003	0	0.095±0.030	0

Table 6: Statistical error (MC) and background normalization error after Method B selection.

	signal	$W+jets$	$Z+jets$	WW	WZ	ZZ	Other $t\bar{t}$	"Stew"
$\Delta N_{\mathrm{MC}}/N$	1.6%	-	26%	17%	15%	31%	48%	-
$\Delta N_{\mathrm{norm}}/N$	0	100%	100%	100%	100%	100%	0	100%

Table 7: Samples used in the QCD factorization study.

$\hat{p_T}$	Dataset	cross-section (pb)	N_{evt}
15-20	/QCD_Pt_15_20/CMSSW_1_5_2-CSA07-2026/GEN-SIM-DIGI-RECO	1.46 10^9	1.31 10^6
20-30	/QCD_Pt_20_30/CMSSW_1_5_2-CSA07-2162/GEN-SIM-DIGI-RECO	630 10^6	2.055 10^6
30-50	/QCD_Pt_30_50/CMSSW_1_5_2-CSA07-2048/GEN-SIM-DIGI-RECO	163 10^6	1.315 10^6
50-80	/QCD_Pt_50_80/CMSSW_1_5_2-CSA07-2049/GEN-SIM-DIGI-RECO	21.6 10^6	0.95 10^6
80-120	/QCD_Pt_80_120/CMSSW_1_5_2-CSA07-2027/GEN-SIM-DIGI-RECO	3.08 10^6	0.725 10^6
120-170	/QCD_Pt_120_170/CMSSW_1_5_2-CSA07-2171/GEN-SIM-DIGI-RECO	494 10^3	1.27 10^6
170-230	/QCD_Pt_170_230/CMSSW_1_5_2-CSA07-2069/GEN-SIM-DIGI-RECO	101 10^3	1.04 10^6
230-300	/QCD_Pt_230_300/CMSSW_1_5_2-CSA07-2050/GEN-SIM-DIGI-RECO	24.5 10^3	0.82 10^6
300-380	/QCD_Pt_300_380/CMSSW_1_5_2-CSA07-2061/GEN-SIM-DIGI-RECO	6.24 10^3	1.29 10^6
380-470	/QCD_Pt_380_470/CMSSW_1_5_2-CSA07-2172/GEN-SIM-DIGI-RECO	1.78 10^3	1 10^6
470-600	/QCD_Pt_470_600/CMSSW_1_5_2-CSA07-2096/GEN-SIM-DIGI-RECO	683	1.025 10^6
600-800	/QCD_Pt_600_800/CMSSW_1_5_2-CSA07-2097/GEN-SIM-DIGI-RECO	204	0.7 10^6
800-1000	/QCD_Pt_800_1000/CMSSW_1_5_2-CSA07-2028/GEN-SIM-DIGI-RECO	35.1	0.885 10^6
-	/bb1j-alpgen/CMSSW_1_6_7-CSA07-1205618198/RECO	280 10^3	1.024 10^6

Table 8: Results of the factorization method on QCD, for Method A lepton selection. Results are expressed in terms of visible cross section (pb).

$\hat{p_T}$	ϵ_e	ϵ_μ	$\epsilon_e \times \epsilon_\mu$	$\epsilon_{e\mu}$	$\sigma_{evt}(e \times \mu)$	$\sigma_{evt}(e\mu)$
15-20	$<7.63\times10^{-7}$	7.63×10^{-7}	$<8.24\times10^{-13}$	$<7.63\times10^{-7}$	$<1.2\times10^{-3}$	$<1.11\times10^{3}$
		$\pm7.63\times10^{-7}$				
20-30	$<4.87\times10^{-7}$	3.41×10^{-6}	$<1.77\times10^{-12}$	$<4.87\times10^{-7}$	$<1.12\times10^{-3}$	$<3.07\times10^{2}$
		$\pm1.29\times10^{-6}$				
30-50	3.8×10^{-6}	4.56×10^{-6}	1.73×10^{-11}	$<7.6\times10^{-7}$	2.83×10^{-3}	$<1.24\times10^{2}$
	$\pm1.7\times10^{-6}$	$\pm1.86\times10^{-6}$	$\pm1.05\times10^{-11}$		$\pm1.71\times10^{-3}$	
50-80	4.74×10^{-5}	6.32×10^{-6}	2.99×10^{-10}	$<1.05\times10^{-6}$	6.46×10^{-3}	<22.7
	$\pm7.06\times10^{-6}$	$\pm2.58\times10^{-6}$	$\pm1.3\times10^{-10}$		$\pm2.81\times10^{-3}$	
80-120	4.55×10^{-5}	1.52×10^{-5}	6.91×10^{-10}	$<1.38\times10^{-6}$	2.13×10^{-3}	<4.25
	$\pm7.92\times10^{-6}$	$\pm4.57\times10^{-6}$	$\pm2.4\times10^{-10}$		$\pm0.74\times10^{-3}$	
120-170	4.65×10^{-5}	1.1×10^{-5}	5.12×10^{-10}	$<7.87\times10^{-7}$	2.53×10^{-4}	<0.39
	$\pm6.05\times10^{-6}$	$\pm2.95\times10^{-6}$	$\pm1.52\times10^{-10}$		$\pm0.75\times10^{-4}$	
170-230	4.81×10^{-5}	1.54×10^{-5}	7.4×10^{-10}	$<9.62\times10^{-7}$	7.47×10^{-5}	$<9.7\times10^{-2}$
	$\pm6.8\times10^{-6}$	$\pm3.85\times10^{-6}$	$\pm2.12\times10^{-10}$		$\pm2.15\times10^{-5}$	
230-300	5.0×10^{-5}	7.32×10^{-6}	3.66×10^{-10}	$<1.22\times10^{-6}$	8.96×10^{-6}	$<3.0\times10^{-2}$
	$\pm7.81\times10^{-6}$	$\pm2.99\times10^{-6}$	$\pm1.6\times10^{-10}$		$\pm3.92\times10^{-6}$	
300-380	6.74×10^{-5}	1.09×10^{-5}	7.32×10^{-10}	$<7.75\times10^{-7}$	4.57×10^{-6}	$<4.84\times10^{-3}$
	$\pm7.23\times10^{-6}$	$\pm2.9\times10^{-6}$	$\pm2.11\times10^{-10}$		$\pm1.32\times10^{-6}$	

Table 9: Results of the factorization method on QCD, for Method B lepton selection. Results are expressed in terms of visible cross section (pb).

$\hat{p_T}$	ϵ_e	ϵ_μ	$\epsilon_e \times \epsilon_\mu$	$\epsilon_{e\mu}$	$\sigma_{evt}(e \times \mu)$	$\sigma_{evt}(e\mu)$
15-20	3.05×10^{-6} $\pm1.53\times10^{-6}$	1.53×10^{-6} $\pm1.08\times10^{-6}$	4.66×10^{-12} $\pm4.04\times10^{-12}$	$<7.63\times10^{-7}$	6.81×10^{-3} $\pm5.89\times10^{-3}$	$<1.11\times10^{3}$
20-30	6.81×10^{-6} $\pm1.82\times10^{-6}$	4.38×10^{-6} $\pm1.46\times10^{-6}$	2.98×10^{-11} $\pm1.27\times10^{-11}$	$<4.87\times10^{-7}$	1.88×10^{-2} $\pm0.80\times10^{-2}$	$<3.07\times10^{2}$
30-50	6.31×10^{-5} $\pm6.93\times10^{-6}$	1.9×10^{-5} $\pm3.8\times10^{-6}$	1.2×10^{-9} $\pm2.74\times10^{-10}$	$<7.6\times10^{-7}$	0.20 ±0.04	$<1.24\times10^{2}$
50-80	1.6×10^{-4} $\pm1.3\times10^{-5}$	3.26×10^{-5} $\pm5.86\times10^{-6}$	5.22×10^{-9} $\pm1.03\times10^{-9}$	$<1.05\times10^{-6}$	0.11 ±0.02	<22.7
80-120	1.27×10^{-4} $\pm1.32\times10^{-5}$	3.86×10^{-5} $\pm7.3\times10^{-6}$	4.9×10^{-9} $\pm1.06\times10^{-9}$	$<1.38\times10^{-6}$	1.51×10^{-2} $\pm0.33\times10^{-2}$	<4.25
120-170	1.17×10^{-4} $\pm9.58\times10^{-6}$	2.6×10^{-5} $\pm4.52\times10^{-6}$	3.03×10^{-9} $\pm5.83\times10^{-10}$	7.87×10^{-7} $\pm7.87\times10^{-7}$	1.5×10^{-3} $\pm0.29\times10^{-3}$	0.39 ±0.39
170-230	1.17×10^{-4} $\pm1.06\times10^{-5}$	4.42×10^{-5} $\pm6.52\times10^{-6}$	5.19×10^{-9} $\pm8.98\times10^{-10}$	$<9.62\times10^{-7}$	5.24×10^{-4} $\pm0.91\times10^{-4}$	$<9.71\times10^{-2}$
230-300	1.3×10^{-4} $\pm1.26\times10^{-5}$	2.44×10^{-5} $\pm5.45\times10^{-6}$	3.18×10^{-9} $\pm7.75\times10^{-10}$	$<1.22\times10^{-6}$	7.8×10^{-5} $\pm1.9\times10^{-5}$	$<2.99\times10^{-2}$
300-380	1.29×10^{-4} $\pm9.99\times10^{-6}$	3.26×10^{-5} $\pm5.02\times10^{-6}$	4.19×10^{-9} $\pm7.24\times10^{-10}$	$<7.75\times10^{-7}$	2.61×10^{-5} $\pm0.45\times10^{-5}$	$<4.84\times10^{-3}$

Table 10: Results of the factorization method on $b\bar{b} + 1jet$, for Method A and Method B lepton selections. Results are expressed in terms of visible cross section (pb).

	ϵ_e	ϵ_μ	$\epsilon_e \times \epsilon_\mu$	$\epsilon_{e\mu}$	$N_{evt}(e \times \mu)$	$N_{evt}(e\mu)$
Method A	1.056×10^{-5} $\pm3.18\times10^{-6}$	4.416×10^{-5} $\pm6.51\times10^{-6}$	4.662×10^{-10} $\pm2.09\times10^{-10}$	$<2.21\times10^{-6}$	1.31×10^{-4} $\pm5.86\times10^{-5}$	<0.618
Method B	1.583×10^{-4} $\pm1.23\times10^{-5}$	5.789×10^{-4} $\pm2.36\times10^{-5}$	9.168×10^{-8} $\pm1.09\times10^{-8}$	$<2.21\times10^{-6}$	0.026 $\pm3.04\times10^{-3}$	<0.618

Table 11: Visible cross section (in pb) for the μ-enriched QCD sample ($p_T^{\mu,gen} > 20$ GeV, $\hat{p_T} > 15$ GeV).

Preselection:	20.0±1.5
Method A	
tight lepton selection:	0 ($<$ 0.26 95% CL)
double-counting removal:	0 ($<$ 0.26 95% CL)
Method B	
lepton selection:	1.5± 0.4
double-counting removal:	1.4±0.4
$N_{trk} \geq 2$	1.3±0.4
$N_{trk-jets} \geq 2$	0.23±0.16

Table 12: Monte Carlo samples for $ZQ\bar{Q}$ processes.

Process	Dataset	cross-section (pb)	N_{evt}
$Zb\bar{b}0j$	/zbb0j-alpgen/CMSSW_1_6_7-CSA07-1204273917/RECO	1.66	5123
$Zb\bar{b}1j$	/zbb1j-alpgen/CMSSW_1_6_7-CSA07-1204133697/RECO	0.29	5339
$Zb\bar{b}2j$	/zbb2j-alpgen/CMSSW_1_6_7-CSA07-1204133588/RECO	0.05	4767
$Zb\bar{b}3j$	/zbb3j-alpgen/CMSSW_1_6_7-CSA07-1204525182/RECO	0.01	9194
$Zc\bar{c}0j$	/zcc0j-alpgen/CMSSW_1_6_7-CSA07-1204525123/RECO	1.43	5280
$Zc\bar{c}1j$	/zcc1j-alpgen/CMSSW_1_6_7-CSA07-1204273971/RECO	0.24	5583
$Zc\bar{c}2j$	/zcc2j-alpgen/CMSSW_1_6_7-CSA07-1204274023/RECO	0.04	4223
$Zc\bar{c}3j$	/zcc3j-alpgen/CMSSW_1_6_7-CSA07-1205410444/RECO	0.01	5993

Table 13: Visible cross section (in pb) for the $ZQ\bar{Q} + jets$ samples.

	$Zb\bar{b}0j$	$Zb\bar{b}1j$	$Zb\bar{b}2j$	$Zb\bar{b}3j$
No selection:	1.66	0.29	0.05	0.01
Method A:	0	0	0	$(1\pm1)\times10^{-6}$
Method B:	$(3\pm3)\times10^{-4}$	$(5\pm5)\times10^{-5}$	$(1\pm1)\times10^{-5}$	$(1\pm1)\times10^{-6}$
	$Zc\bar{c}0j$	$Zc\bar{c}1j$	$Zc\bar{c}2j$	$Zc\bar{c}3j$
No selection:	1.43	0.24	0.04	0.01
Method A:	0	0	0	$(1.7\pm1.7)\times10^{-6}$
Method B:	0	$(4\pm4)\times10^{-5}$	$(1.9\pm1.3)\times10^{-5}$	$(3.3\pm2.4)\times10^{-6}$

Table 14: Visible cross section (in pb) for the Chowder sample with 10 pb^{-1} detector conditions, compared to 0 pb^{-1}.

Method A	10 pb^{-1} scenario	0 pb^{-1} scenario
Signal	3.292 ± 0.038	3.160 ± 0.117
$W + jets$	0.050 ± 0.008	0.020 ± 0.014
$Z + jets$	0.046 ± 0.007	0.044 ± 0.023
Other $t\bar{t}$	0.005 ± 0.002	0
Method B	**10 pb^{-1} scenario**	**0 pb^{-1} scenario**
Signal	4.334 ± 0.043	4.833 ± 0.145
$W + jets$	0.022 ± 0.005	0
$Z + jets$	0.103 ± 0.010	0.095 ± 0.030
Other $t\bar{t}$	0.026 ± 0.003	0.013 ± 0.145

Table 15: Sensitivity of the selection efficiencies to a $\pm5\%$ momentum rescaling for electrons, muons and generic charged tracks.

method A	DiLept. $t\bar{t}$	Other $t\bar{t}$	WW	WZ	ZZ	$W + jets$	$Z + jets$	QCD	Other "Stew"
electron:	2.5%	0.0%	3.3%	1.2%	0.0%	28.3%	17.5%	0.0%	0.0%
muon:	1.6%	0.0%	3.1%	6.1%	5.0%	0.0%	17.5%	0.0%	50.0%
track:	0.3%	0.0%	0.2%	0.0%	**FIXME**	0.0%	0.0%	0.0%	0.0%
method A	DiLept. $t\bar{t}$	Other $t\bar{t}$	WW	WZ	ZZ	$W + jets$	$Z + jets$	QCD	Other "Stew"
electron:	1.3%	17.6%	1.3%	0.0%	5.0%	0.0%	5.5%	0.0%	0.0%
muon:	1.0%	17.6%	0.0%	6.2%	0.0%	0.0%	5.5%	0.0%	0.0%
track:	1.8%	0.0%	9.2%	6.2%	**FIXME**	0.0%	4.2%	0.0%	0.0%

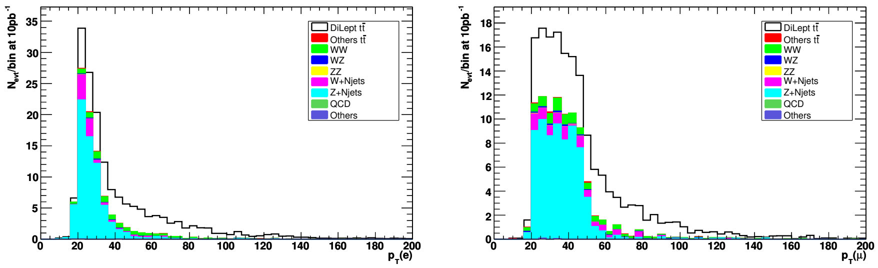


Figure 1: Distribution of transverse momentum for the electron candidate (left) and the muon candidate (right), after preselection. The distributions are normalised to 10 pb^{-1}.

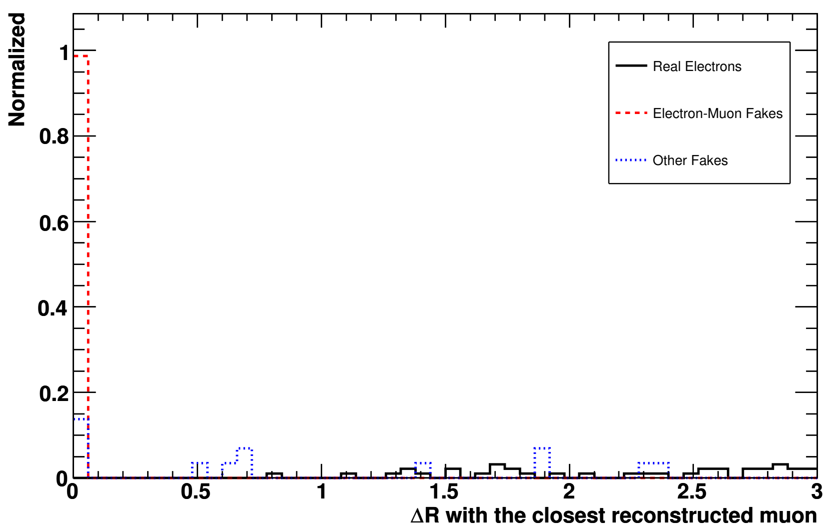


Figure 2: Distance in (η, ϕ) between electron candidates and closest muon candidate, for $Z + jets$ events, according to their Monte Carlo truth. Normalization is arbitrary.

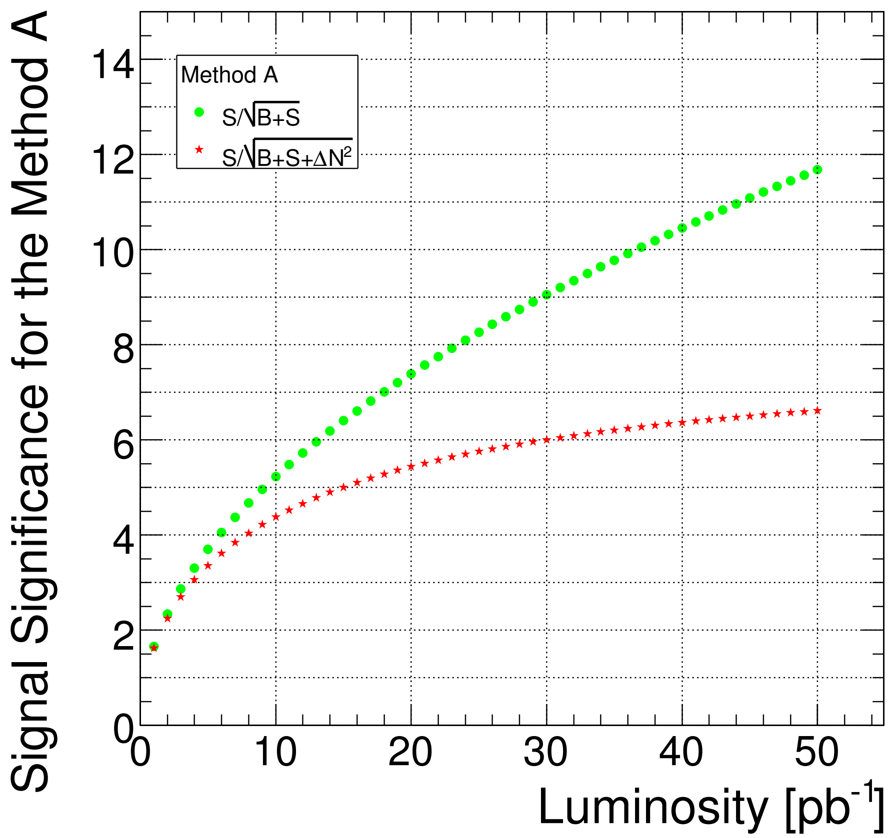


Figure 3: Signal significance for Method A, as a function of integrated luminosity, with and without systematic uncertainties.

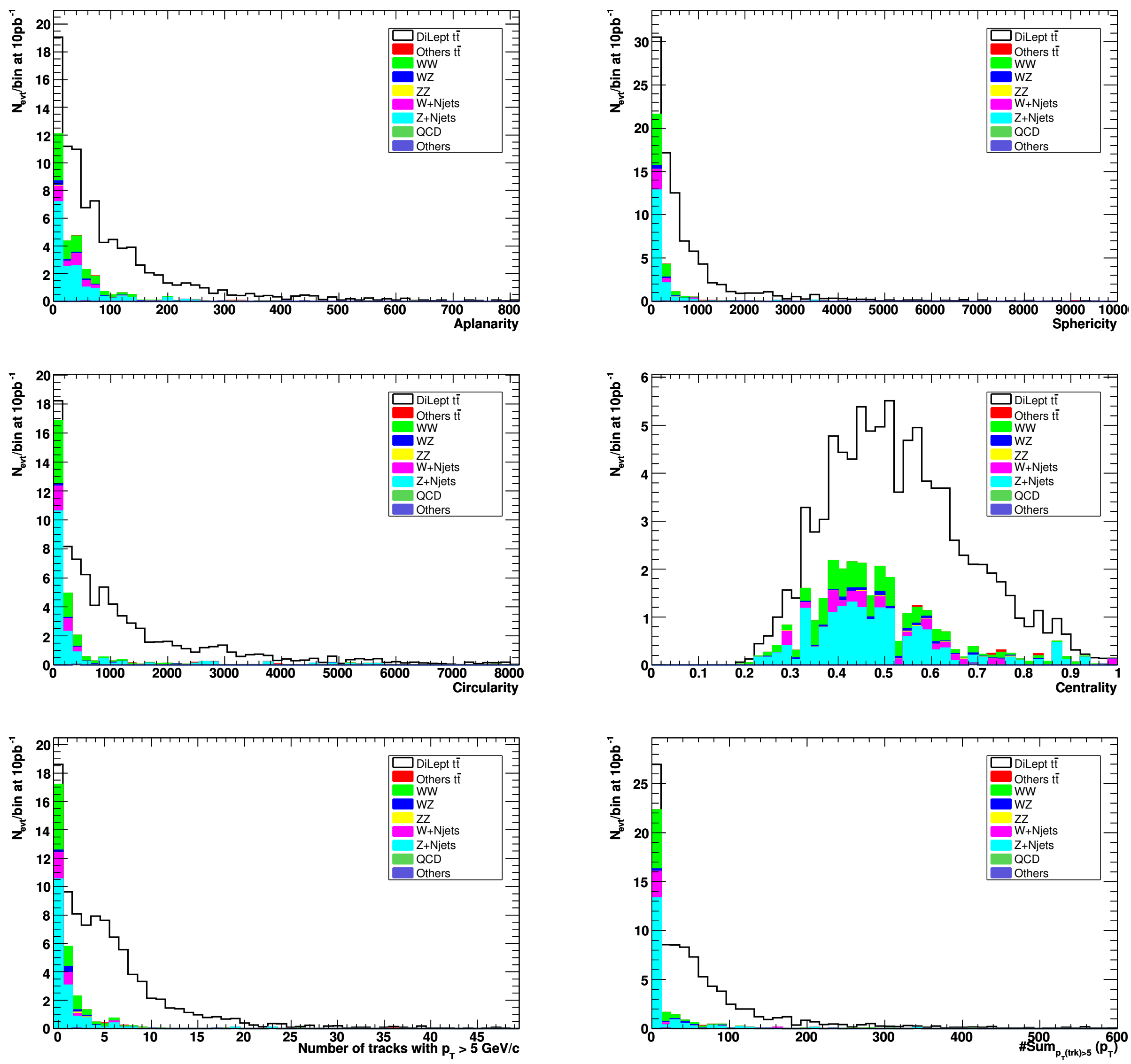


Figure 4: Aplanarity, sphericity, circularity, centrality, number of tracks with $p_T > 5$ GeV, $\sum_{trk} p_T$, after the leptonic selection of Method B. All distributions are normalised to 10 pb^{-1}.

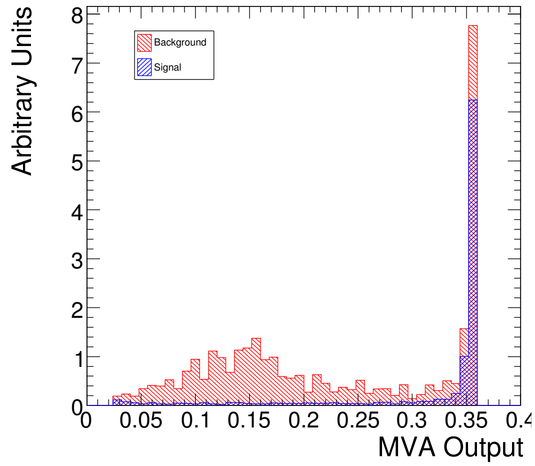


Figure 5: Distribution of a linear discriminant combining aplanarity, sphericity and circularity, after the leptonic selection of Method B. The normalization is arbitrary.

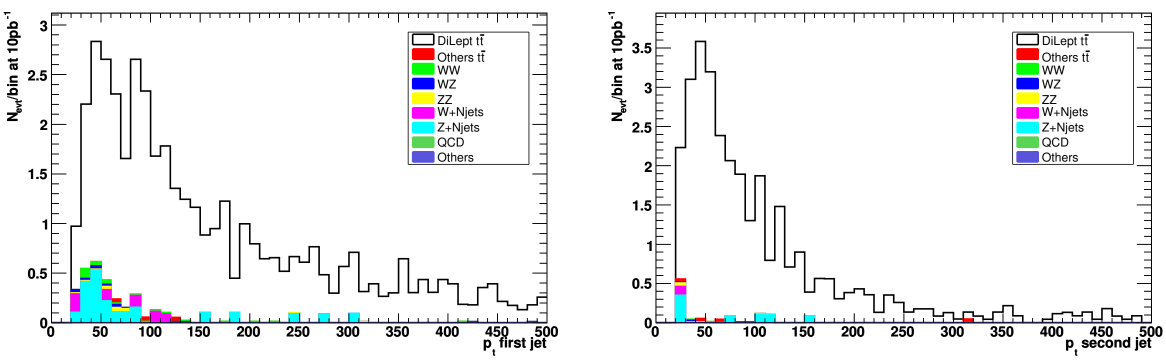
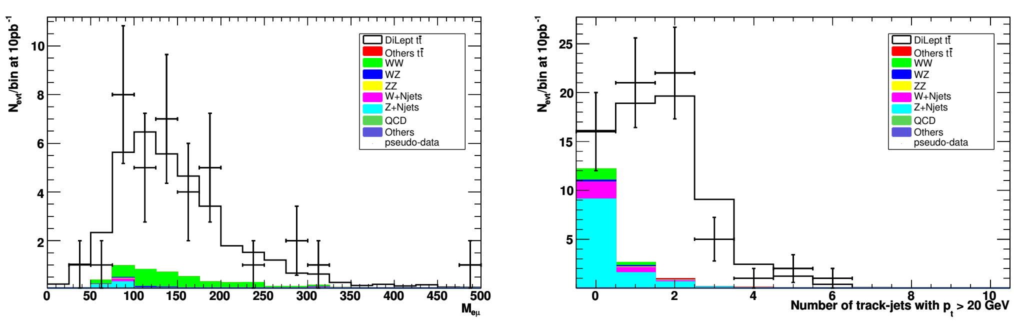


Figure 6: Distribution of p_T for the first (left) and second (right) track-jets, after leptonic and topological selection, for the k_T (FastJet) algorithm, with $D = 0.7$. Distributions are normalised to the expected number of signal and background events after 10 pb^{-1}.



a) b)

Figure 7: Expected number of events after an accumulated statistics of 10 pb^{-1} for the "control variable" of each method, with pseudo-experiment data, superimposed to MC expectation for signal and background. Method A (left): the reconstructed dileptonic mass. Method B (right): the number of track-jets with $p_T > 20$ in the k_T (fastjet) algorithm, with the $D = 0.7$.

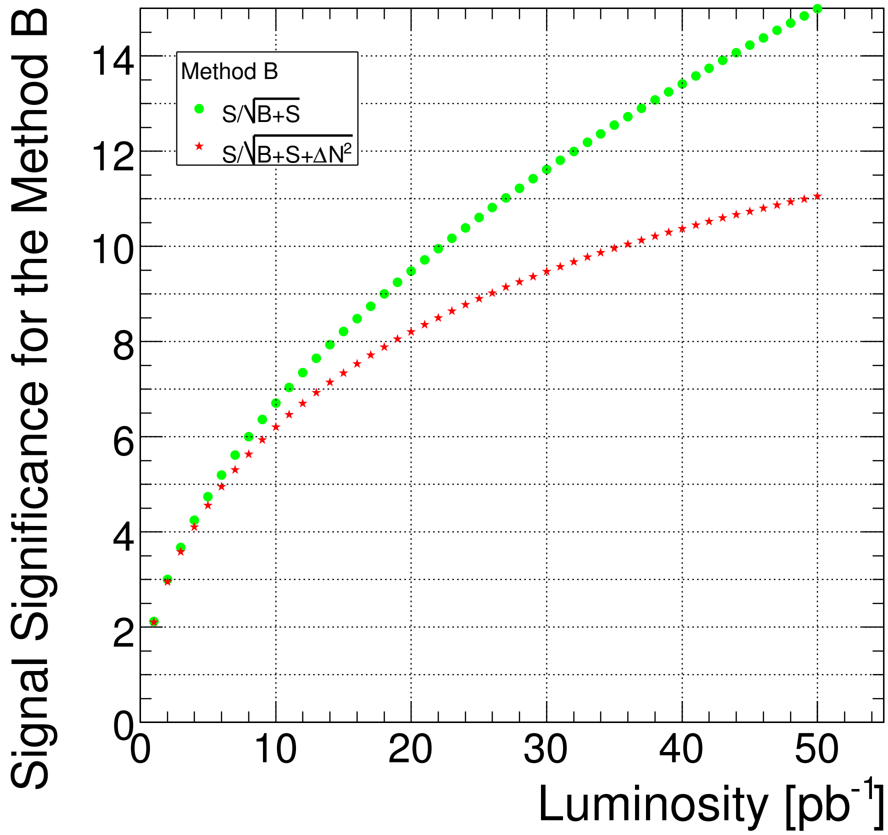


Figure 8: Signal significance for Method B, as a function of integrated luminosity, with and without systematic uncertainties.

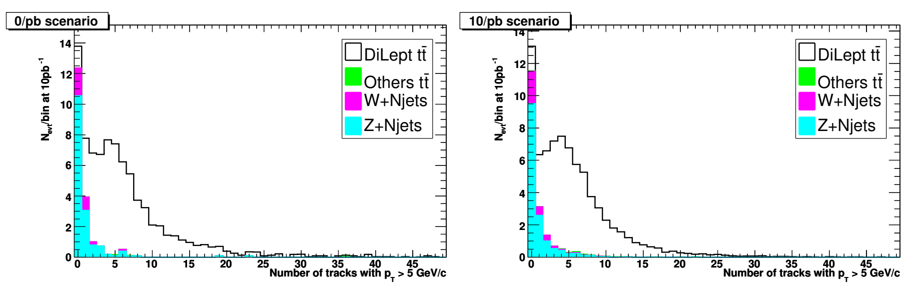


Figure 9: Number of tracks with $p_T > 5$ GeV for Chowder samples in the “0 pb^{-1}” (left) and “10 pb^{-1}” (right) scenarios.

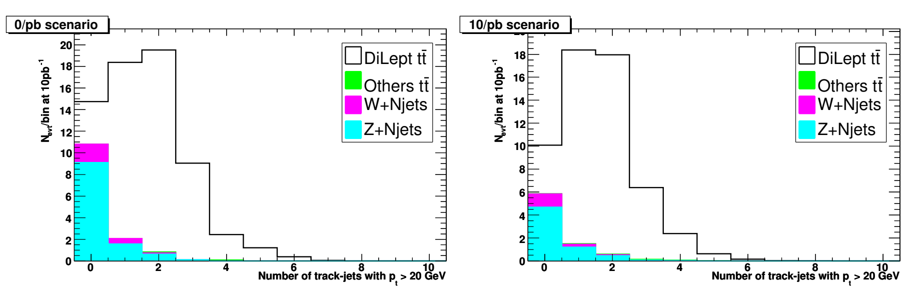


Figure 10: Number of track-jets with $p_T > 20$ GeV for Chowder samples in the “0 pb^{-1}” (left) and “10 pb^{-1}” (right) scenarios.