Draft version 1.3

ATLAS NOTE

October 13, 2014

Search for ttH in the multilepton final state: backgrounds and their estimation

S. Biondi[f], D. Boumediene[a], D. Calvet[a], D. DeMarco[h], E. Dubreuil[a], S. Gentille[g], D. Hohn[m], M. Kuna[g], C. Lester[b], F. Manghi[f], J. McFayden[l], S. Monzani[g], P. Onyisi[c], R. Ospanov[b], D. Paredes[a], M. Pitt[i], G. Salamanna[e], J. Schaarschmidt[i], F. Seifert[d], A. Sidoti[f], A. Sopczak[d], S. Valentinetti[f]

[a]*LPC Clermont*
[b]*Univ. Pennsylvania*
[c]*Univ. of Texas at Austin*
[d]*IEAP CTU in Prague*
[e]*Queen Mary Univ. London*
[f]*Univ. Bologna*
[g]*Univ. Roma I*
[h]*Univ. Toronto*
[i]*Weizmann Inst. of Sc.*
[l]*University College London*
[m]*University of Bonn*

Abstract

This note is the second in the series documenting the search for the associated production of a SM Higgs boson in association with a pair of top quarks, in the multi-lepton channel, in ATLAS using 20.3 fb^{-1} of pp collision data from the LHC collected in 2012. It is devoted to describing how the main backgrounds are estimated in the signal region. It focuses in particular on the reducible backgrounds consisting of fake leptons associated to real (i.e. primary) leptons, primarily $t\bar{t}$+jets and Z+jets . All 5 channels looked at (2ℓSS, 2ℓSS+$1\tau_{had}$, 3ℓ, 4ℓ, 1ℓ+1τ) are described. Their strategy to estimate the light and τ fake contributions in the SR are presented, attaining a relative uncertainty at the level of 30 to 50 %. The estimation of the probability for the electron charge to be mis-identified for electrons used in these analyses is also evaluated as a function of the electron kinematics. Finally, the expected number of $t\bar{t}V$ events in the signal region, together with theoretical systematics on the NLO cross-section derived for the phase space of these analyses are also shown.

Contents

1 Introduction

This is the second note documenting the search for the associated production of the Standard Model Higgs boson with a pair of top quarks ($t\bar{t}H$) in the multilepton final state, with 20.3 fb^{-1} of ATLAS data at $\sqrt{s}$=8 TeV. It is devoted to describing the strategy and providing the results of the estimation of the main backgrounds. As seen from Note 1 [1], the signal-to-background ratio at the optimal selections is less than 1 for all the channels included in this search. The dominant backgrounds are top quark pairs produced in association with jets or vector bosons; and di-boson production. The number of surviving background events after all selections is of the same order of magnitude as the expected number of signal events, between a factor 1 and 2.

Because they are comparable in size to the signal, it is imperative that the contributions of these backgrounds in the signal region (SR) be accurately estimated. In particular, Monte Carlo (MC) simulations are generally known not to model the reducible backgrounds from processes generating a 'fake' (reconstructed object not corresponding to a true particle) or a secondary (coming from a quark decay chain) lepton accurately. For this reason, a large number of studies is dedicated to evaluating the contribution of such reducible backgrounds in the SR using data as much as possible: yields and shapes are either taken directly from data (as it is for example the case of the Same Sign 2 lepton channel with no hadronically decaying tau leptons (τ_{had})); or validating the MC shapes used to extrapolate from control regions (CR) into the SR with data (as it is e.g. the case for the 3 lepton channel and Same Sign 2 lepton channel with 1 τ_{had}).

The other backgrounds are largely taken from Monte Carlo, either because they have a negligible impact with respect to the main ones; or because not enough statistics is available with the considered dataset to constrain them to an accuracy better than the theoretical one (as it is the case of $t\bar{t}V$). It follows that the main topic of this note is the treatment of the secondary and fake lepton backgrounds from the $t\bar{t}$ background (and Z+jets). For the channels where it is relevant, also a validation of the MC expectations for the diboson production with data is provided.

In the note the definition of CR used to extract the initial normalizations and the strategy to extract the background yields in the SR for each channel are described. Control plots used to identify MC-data differences, characterize the various backgrounds and assess the level of accuracy of MC are presented; and the final numbers of estimated background events in the SR are provided, together with the associated statistical and systematic uncertainty. The layout of the note is as follows: a general overview of the backgrounds is presented in the first sections; then a section each is dedicated to the various channels, in particular describing the data-driven background estimations. Each section also provides the effect of the remaining backgrounds (MC-driven) in the SR and the associated uncertainties, particularly discussing the impact of $t\bar{t}V$. A section devoted to the estimation of the probability of mis-identifying the sign of the electric charge of electrons is also included, given its relevance for the impact of W^+W^- and $W^\pm Z$ in the Same Sign 2 lepton channel.

2 Background processes

2.1 List of background processes, MC samples and weights

As already shown in Table 14 of [1], the background processes to $t\bar{t}H$ considered for these analyses are: ttV, VV, $t\bar{t}+X$ and $t+X$ (in the t-, s-, Wt channels), Z+jets, $t+Z$, $t\bar{t}WW$. It is intended $V = W, Z$ throughout the note.

The MC samples, corresponding Data Set ID (DSID) and cross-sections are the same as in Note 1 and are listed on Table 4 therein [1] and also provided for the reader in Appendix A.

The rate of all simulated processes is re-scaled to match the expected production cross-section for the integrated luminosity considered (20.3 fb^{-1}). Additionally, each simulated event is also re-weighted by the product of weights for: the distribution of the average number of interactions per bunch crossing, μ, divided by a factor 1.11 to match the primary vertex multiplicity in data [2], using the `PileUpReweightingTool`; the MV1 efficiency and mistag rate scale factors; the lepton trigger, reconstruction and identification scale factors. The b-tagging scale factors are those obtained from the `ptrel` calibrations for the full 2012 dataset; the lepton isolation scale factors are still those inherited from the $H \to WW$ analysis of HSG3, until a new round of ntuple production includes the values for the optimal isolation cuts obtained for these analysis.

Finally, the $t\bar{t}$ events (samples 117050, `Powheg+Pythia no-all-had`, and 181087, `Powheg+Pythia di-leptonic only, AF2`) undergo a further sequential re-weighting according to the values of the trasverse momenta of the $t\bar{t}$ system and then of the top quark. A need for this has been established in the ATLAS top quark measurements with additional jets [3] and a tool is provided ad-hoc and applied to all HSG8 analyses coherently [4].

2.2 Background properties

Background processes can be sorted into two categories:

- Events with a non prompt or a fake lepton selected as prompt lepton. These processes cannot lead to a final state compatible with thee signal signature without a mis-reconstructed object. In particular, for the $t\bar{t}H$ final state, events with a prompt lepton but with mis-reconstructed charge; and events in which a secondary muon from a (light or heavy flavour) quark decay is mistaken for a primary lepton from a Higgs boson or top quark decay are relevant.

  The main backgrounds of this sort are: $t\bar{t}$(+jets), Z+jets and W^+W^- . Data-driven techniques are used to control this category of events. Their importance varies depending on the channel.

- Events which can lead to the same final state as the signal. The main background of this category are: $t\bar{t}V$ (V=Z,W), tZ, $W^{\pm}Z$ and $W^{\pm}W^{\pm}$. In this analysis they are modelled using the Monte Carlo simulations and checked in data control regions to the level of accuracy possible with the accumulated statistics.

3 Top plus Vector Boson Production

Production of top quarks plus vector boson (including $t\bar{t}V$ and tZ) is an important background in all multilepton channels. In terms of jet multiplicity, $t\bar{t}Z$ perfectly mimics the $t\bar{t}H$, $H \to W^+W^-$ signal, and primarily affects channels where opposite sign leptons are selected. A large part of this component, arising from on-shell $Z \to \ell\ell$, can be removed via a Z mass veto on like-flavour, opposite sign leptons. However the $Z \to \tau\tau$ and γ^* components remain. The $t\bar{t}W$ and tZ processes generally require extra jets to reach the multiplicity of our signal regions. Shape uncertainties will be covered in a later note; here we discuss only overall cross section uncertainties.

The $t\bar{t}W$ and on-shell $t\bar{t}Z$ processes have had cross sections evaluated at NLO by different groups [5, 6], using different scale and SM parameter choices (leading to differences in predicted cross sections of 12–16%). We adopt the central values and scale uncertainties of Garzelli *et al.* [6]. The renormalization and factorization scales are set to $\mu_0 = m_t + m_V/2$, and the scale variations are to $2\mu_0$ and $\mu_0/2$. We evaluate the corresponding PDF uncertainties with aMC@NLO. This generator is validated by reproducing the cross sections and scale uncertainties of both [5, 6], as well as the PDF uncertainties determined in [5] for $t\bar{t}W$. The cross sections from the theoretical papers are confirmed within 2%, and the scale and systematic uncertainties reported by aMC@NLO also agree well with those from the papers.

Table 1: Cross sections and cross section uncertainties for $t\bar{t}V$ and tZ processes. Uncertainties are symmetrized.

Process	Cross section (fb)	Scale (%)	PDF (%)
$t\bar{t}W^+$	142.6	10	10
$t\bar{t}W^-$	60.5	11	8
$t\bar{t}Z$	205.7	12	9
tZ	160	4	7
$\bar{t}Z$	76	5	7

The tZ process is normalized to NLO based on the calculation in Ref. [7]. Here the scales are set to $\mu_0 = m_t$ and the scale variations are by a factor of four; the scale dependence is found to be quite small.

Monte Carlo events for these processes are generated with MadGraph 5 and showered with Pythia 6. For $t\bar{t}W$ events are generated with up to two extra partons at matrix element level, while for $t\bar{t}Z$ up to one extra parton at ME level is produced. The tZ process is simulated without extra partons. The NLO cross sections are implemented by applying a uniform k-factor to the LO events for each process. For $t\bar{t}Z$, there is a large component of off-shell production, and for the 3 and 4 ℓ channels low mass $\gamma^*/Z \to \ell\ell$ is an important background after on-shell production is removed with a Z veto. In this case the k-factor is determined by comparing LO and NLO cross sections for on-shell Z production only.

For now, we sum the cross sections for charge conjugate processes (i.e., $t\bar{t}W^+/t\bar{t}W^-$ and $tZ/\bar{t}Z$) and apply the resulting k-factor to the LO processes regardless of charge; the uncertainties are added linearly for the two charges. We intend to implement charge-dependent k-factors soon; this expected to be at most a 4% percent effect [5].

We have checked the effect of the $pp \to t\bar{t}WW$ process; it is small in all channels. At this point no explicit theory uncertainties have been assigned.

3.1 Control plots in $t\bar{t}V$ dominated regions

We use the 3ℓ events with a Z candidate to study the WZ and $t\bar{t}Z$ processes. (A Z candidate is defined as an opposite sign, like flavour lepton pair with invariant mass within 5 GeV of 91.2 GeV.) These events are unblinded regardless of the number of jets or b tags in the event.

We define $t\bar{t}Z$ control region A by removing the Z veto of the signal regions (≥ 4 jets, ≥ 1 b-tag and 3 jets, ≥ 2 b-tag) and requiring a Z candidate to be present in the event. The region defined by this is predicted to be 67% $t\bar{t}Z$, 17% WZ, and 13% tZ. We predict 19.3 ± 0.5 events from MC, and observe 28, giving a data/MC ratio of $1.45 \pm 0.27 \pm 0.03$ (where the errors are from data and MC statistics). Distributions of various variables are shown in Fig. 1. In these plots, the lepton ordering is defined in the usual manner of the 3ℓ analysis: lepton 0 is the lepton of opposite sign to the other two, lepton 1 is the same-sign lepton with smaller ΔR separation from lepton 0, and lepton 2 is the remaining same-sign lepton.

The hadronically decaying W and top masses are derived using a very simple reconstruction method. The pair of non-b-tagged jets in the event with smallest separation ΔR is chosen as the W decay candidate, and the invariant mass of the two jets is used to obtain the `mhadw` variable. This is therefore defined only in events where there are at least two light jets (in particular, it is not defined in the 3 jet, ≥ 2 b-tag region). The b-tagged jet with smallest ΔR to the reconstructed W four-momentum vector is then assumed to come from the same parent top quark, and is combined with the W to obtain `mhadtop`. In spite of being rather simple-minded this procedure produces recognizable peaks at the W and top quark masses which can be used to diagnose events.

To the extent that a localized excess is seen, it appears to be in the 4j, 1b category, which has the highest contamination from WZ. The excess also appears to arise from events with high-p_{T}jets that are close to each other (as seen in the `mhadw` spectrum). This will be studied in more detail.

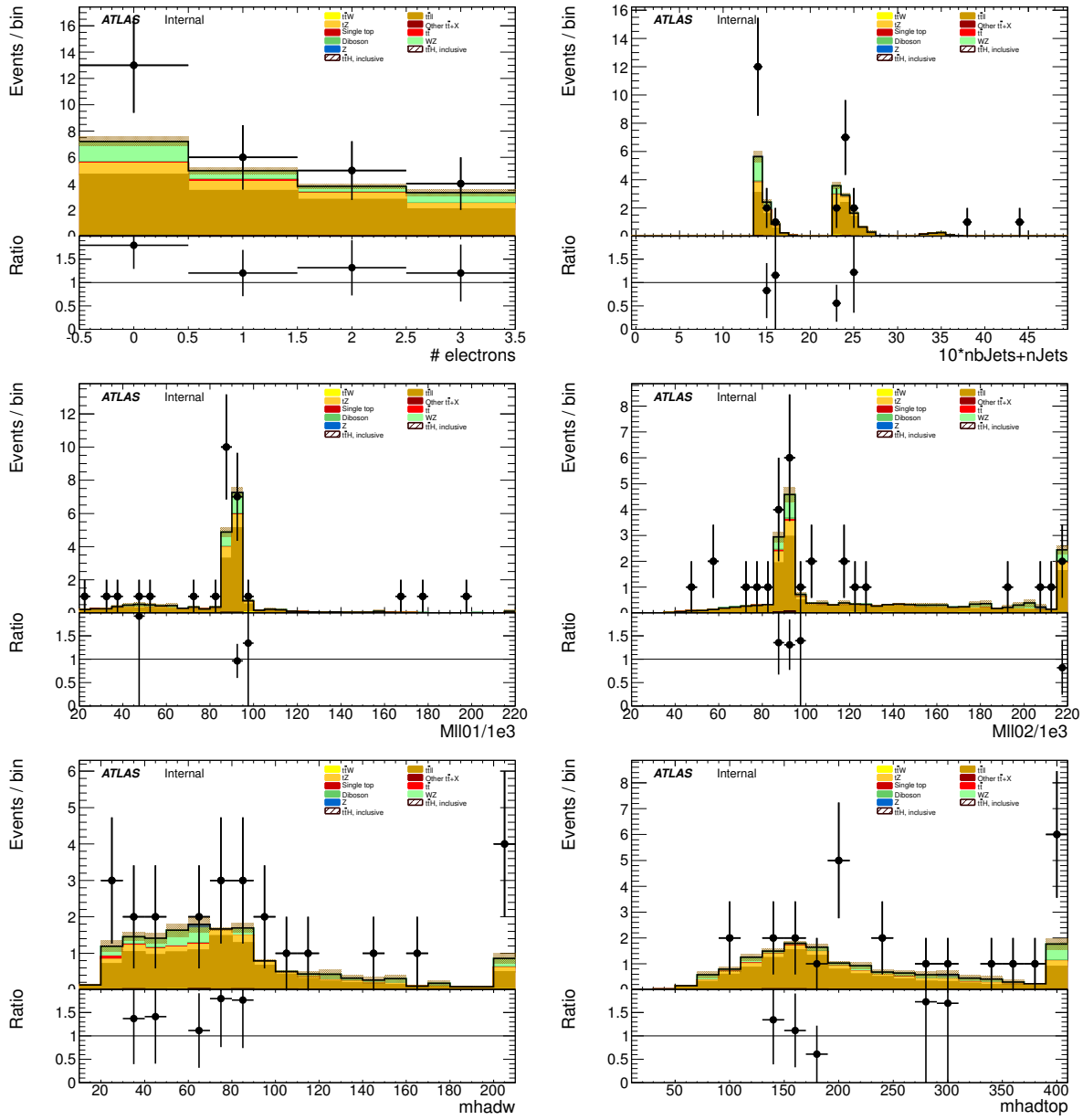


Figure 1: Data/MC comparison plots for $t\bar{t}Z$ control region A ($\geq$ 4 jets, $\geq$ 1 b-tag and 3 jets, $\geq$ 2 b-tag). In all plots, the rightmost bin contains any overflows. Top left: number of electrons. Top right: 10*the number of b-tags + the total number of jets. Middle left: the invariant mass of the (0,1) lepton pair (see the text for the definition of the lepton ordering). Middle right: the invariant mass of the (0,2) lepton pair. Bottom left: reconstructed hadronically decaying W mass (only defined for events with at least two non-b-tagged jets). Bottom right: reconstructed hadronically decaying top quark mass.

4 $W^{\pm}Z$ and ZZ Production

$W^{\pm}Z$ and ZZ di-boson production with additional and b-tagged jets constitute small contributions to the 3- and 4-lepton channels respectively. In the 3-lepton case $W^{\pm}Z$ comprises $\sim$ 1 event of $\sim$ 10 total background events while the ZZ contribution accounts for approximately 10% of the total background in the 4-lepton channel. Because of the small size of these contributions, each of the above processes can be assigned a non-aggressive uncertainty based on similar previous analyses with ATLAS and cross-checked with data validation regions and MC truth studies. We assign an overall 50% error on both the $W^{\pm}Z$ 3-lepton signal region contribution and the ZZ 4-lepton signal region contribution. The details of this error assignment are discussed below.

Both $W^{\pm}Z$ and ZZ production have been studied by ATLAS [8][**?**] but neither process has been investigated thoroughly in association with multiple jets and b-quark jets. However, both $W + b$ [9] and $Z + b$ [10] production in 7 TeV data have been shown to agree with MC models to within 20-30%. A single W produced in association with b-tagged jets possesses a similar topology to the $W^{\pm}Z + b$ process at a different energy scale and has been shown to be dominated by charm mis-tags and b-jets from gluon splitting and multiple parton interaction. The $W + b$ analysis unfortunately uses Alpgen MC with Herwig PS modeling and only provides results to 1 additional jet and therefore is not directly applicable to this $t\bar{t}H$ analysis (where $W^{\pm}Z$ is modeled using Sherpa with massive c and b quarks). $Z + b$ production originates from slightly different diagrams than $ZZ + b$ however the sources of the b-tags are similar and the analysis above provides results with Sherpa MC with an agreement of $\sim$ 30%.

Figure 2 shows the spectrum of the number of reconstructed and selected jets (NJet) in a Zb validation region, defined by 2 tight-isolated leptons within 10 GeV/$c^2$of the Z mass and with at least one b-tagged jet, using the $t\bar{t}H$ analysis definitions. The level of agreement in this region confirms at the 30% level seen in the 7 TeV analysis, discussed above.

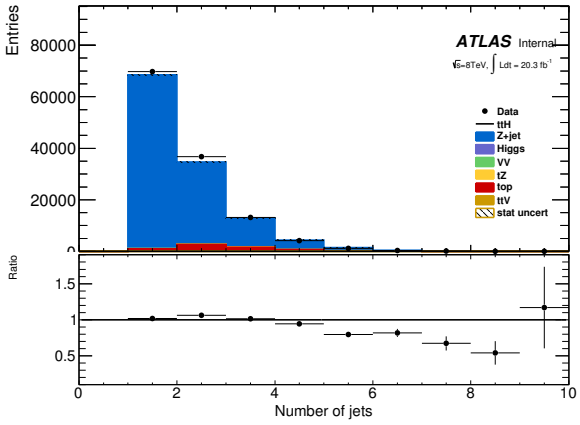


Figure 2: NJet spectrum for 2 tight-isolation leptons with 1 b-tagged jet (MV1_70)

In the following two sections the uncertainty assignments for each of these two di-boson processes will be reviewed in turn.

4.1 $W^{\pm}Z$ Uncertainty

The $t\bar{t}H$ analyses has two validation regions to test the Sherpa agreement with data for $W^{\pm}Z$: one inclusive 3 lepton region, using the three-lepton channel object and p_{T} cuts; and a $W^{\pm}Z + b$ region with 1 b-tagged jet, fewer than 4 jets (to remove $t\bar{t}V$), and a requirement that at least one same-flavor opposite sign pair have an invariant mass within 10 GeV/$c^2$of the Z mass. Figure 3 shows kinematic varibles for

the inclusive region [1]. The NJet spectrum shows good agreement within statistics across the full spectrum, giving confidence about the Sherpa high NJet SR extrapolation. Figure 5 shows NJet spectrum for the $W^{\pm}Z + b$ validation region with good agreement in the 1 and 2 jet bins, but a slight data-MC discrepancy in the 3 jet bin. The region has low stats and around $\sim 60\%$ purity.

We assign a conservative 50% systematic error to cover MC modeling based on these distributions and the agreement seen in similar $W + b$ and $Z + b$ analyses and use the MC central value for the final $W^{\pm}Z$ in the SR.

A cross-check is undertaken by examining the $W^{\pm}Z$ truth origins of the b-jet in the $W^{\pm}Z+b$ validation region (VR) and the signal region using the sherpa sample available. Table 2 shows these fractions. As expected the charm and b contributions dominate, though there is a small dependence on the number of jets. The composition of the VR is fairly similar to that of the signal region, especially in the 3-jet bin.

	Bottom	Charm	Light
$W^{\pm}Z + b$ VR 1 Jet	0.25 ± 0.03	0.054 ± 0.04	0.20 ± 0.03
$W^{\pm}Z + b$ VR 2 Jet	0.34 ± 0.04	0.052 ± 0.06	0.13 ± 0.03
$W^{\pm}Z + b$ VR 3 Jet	0.40 ± 0.07	0.041 ± 0.07	0.18 ± 0.04
$3l$ SR	0.43 ± 0.14	0.038 ± 0.17	0.18 ± 0.11

Table 2: Truth Origin of highest energy b-tagged jet in the $W^{\pm}Z + b$ VR and $3l$ SR

We fit the background-subtracted distribution of the $W^{\pm}Z + b$ using three templates based on the relative fractions of the bottom, charm and light components in Table 2. The subtraction of the $t\bar{t}V$ (plus tZ) and fake backgrounds come with a 20% and 30% error respectively. We allow only the b component to float because of the trend in the truth jet b-fraction and we obtain a preferred scale-factor of 1.87 ± 0.87 on the b component from the fit. When applied to the signal region b-component this would result in a re-normalization that is within the 50% error we assigned. The actual fit is show in Figure **??**.

4.2 ZZ Uncertainty

In order to investigate the MC agreement with data in the ZZ case, two validation regions similar to the $W^{\pm}Z$ case are defined. Firstly, a 4 lepton ZZ region is constructed using the object selections for the 4-lepton channel and requiring exactly two pairs of opposite sign-same flavour leptons with a di-lepton invariant mass within 10 GeV/$c^2$of the Z mass. Additionally, the $ZZ + b$ process is investigated directly using a similar validation region which again requires exactly two Z-candidate lepton pairs as well as at least 1 b-tagged jet. Some kinematic distributions are shown in Figure 6, and particular attention should be paid to the NJet spectrum which shows good data-MC agreement in the high-jet bins, with a slight discrepancy in the 1-jet bin. The agreement for the region with at least 2 jets yields confidence in the NJet MC modelling in this region which lies close to the 4-lepton signal region.

Recall that in the $W^{\pm}Z$ case an overall systematic uncertainty of 50% was assigned to cover the MC modeling. Based on the study of the ZZ and $ZZ + b$ validation regions and the overall agreement noted with the $Z + b$ analysis, we expect a similar error to be appropriate in the ZZ case. A similar truth origin study is undertaken in MC to demonstate a similar b-jet origin to the $W^{\pm}Z$ case. The true origin of the leading (highest energy) b-tagged jet is shown in Table 3 for the 4-lepton signal region as well as the $ZZ + b$ validation region described above divided into jet bins. It can be seen that in case, as it was in the $W^{\pm}Z$ case above, the true origin of the b-jet in $ZZ + b$ is dominated by c and b. Taking this study in tandem with the results from the $W^{\pm}Z$ investigation, it is appropriate to take the central value of the $ZZ + b$ background contribution in the 4-lepton SR from MC and to assign an overall systematic of 50% in order to account for the MC modeling limitations.

[1] the fakes are taken directly from MC

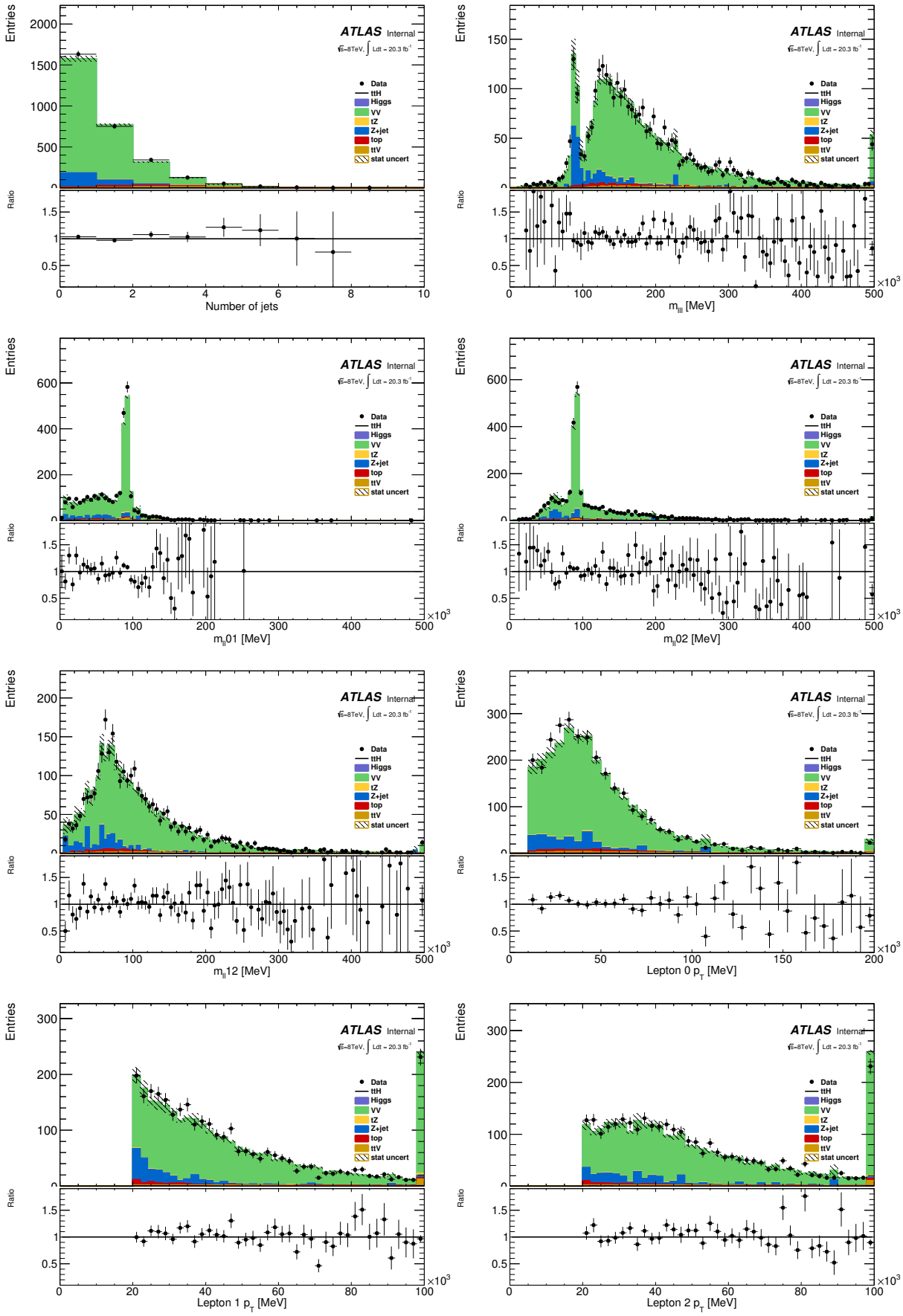


Figure 3: Inclusive 3 lepton $W^{\pm}Z$ validation region using the $t\bar{t}H$ lepton identification and momentum cuts: mass, number of jet and lepton momentum variables

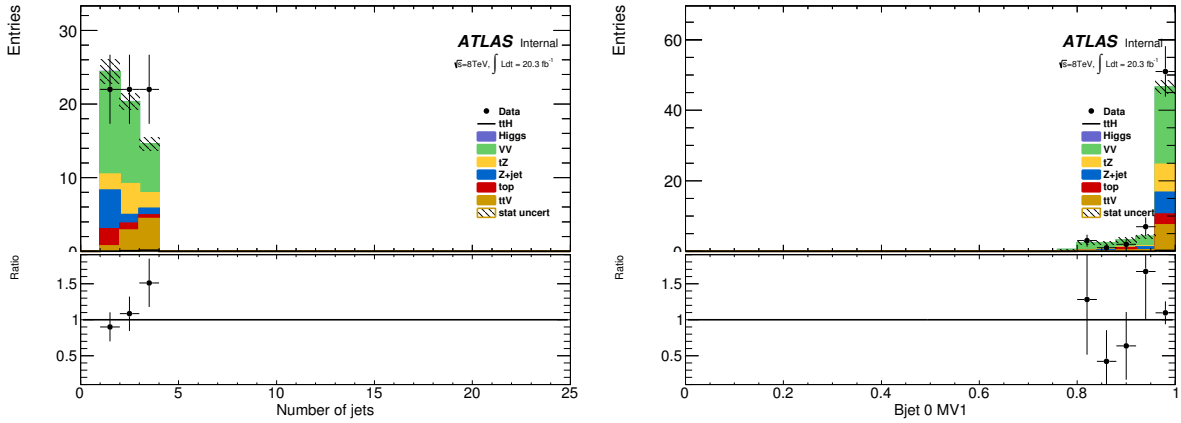


Figure 4: $W^{\pm}Z + b$ validation region: NJet and BJet MV1 variable

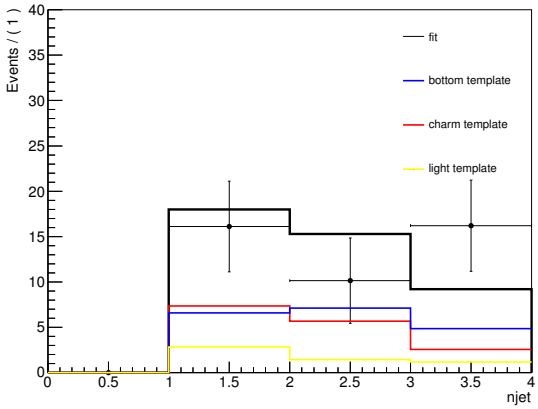
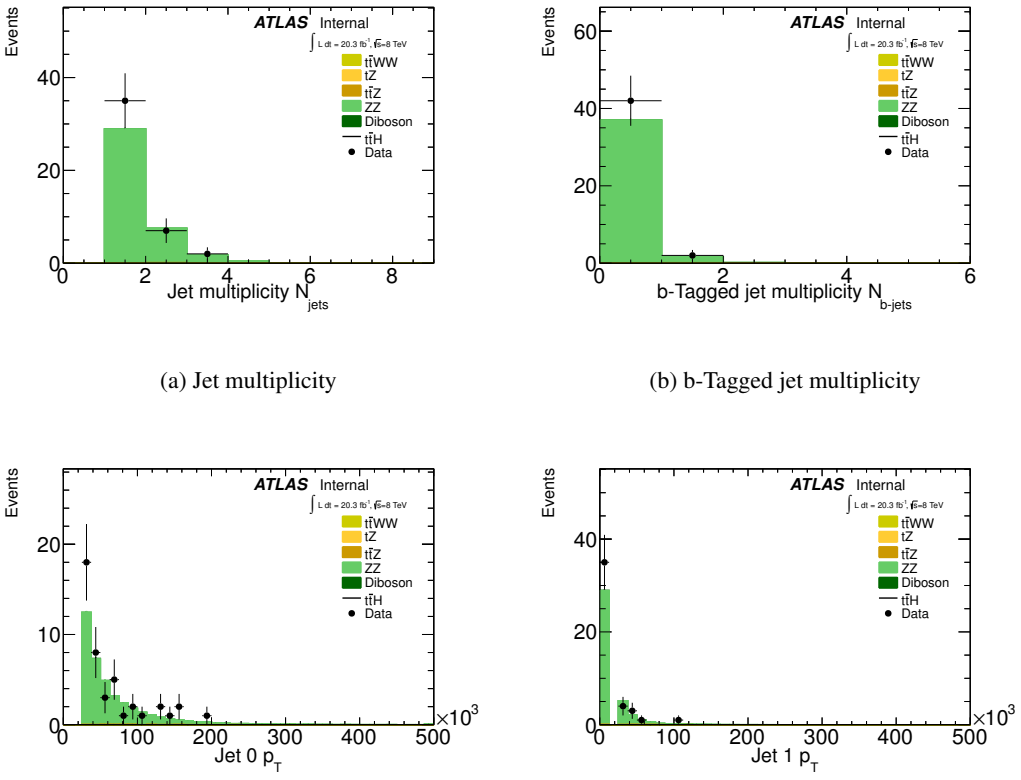


Figure 5: Template fit to background-subtracted data from the first three bins of the $W^{\pm}Z + b$ validation region. Charm (red), light (yellow) and bottom (blue) templates are used based on the numbers in Table 2. The background subtraction is done with a 20% error on the $t\bar{t}V$ and tZ components and 30% error on the fake components

	Bottom	Charm	Light
$ZZ + b$ VR 1 Jet	0.50 ± 0.02	0.21 ± 0.01	0.18 ± 0.01
$ZZ + b$ VR 2 Jet	0.25 ± 0.02	0.12 ± 0.01	0.11 ± 0.01
$ZZ + b$ VR 3 Jet	0.085 ± 0.014	0.040 ± 0.011	0.036 ± 0.011
$4l$ SR	0.020 ± 0.008	0.025 ± 0.008	0.014 ± 0.005

Table 3: Truth Origin of highest energy b-tagged jet in the $ZZ + b$ VR and $4l$ SR



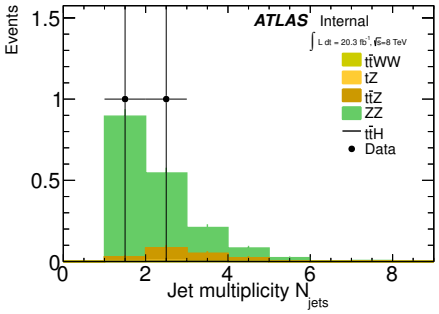
(a) Jet multiplicity

(b) b-Tagged jet multiplicity

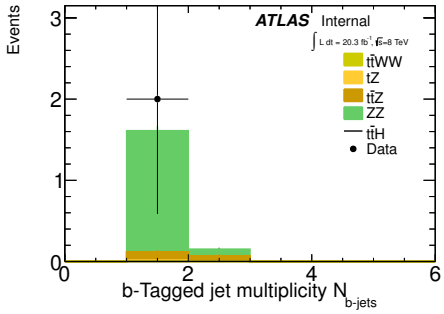
(c) Leading jet p_T

(d) Sub-leading jet p_T

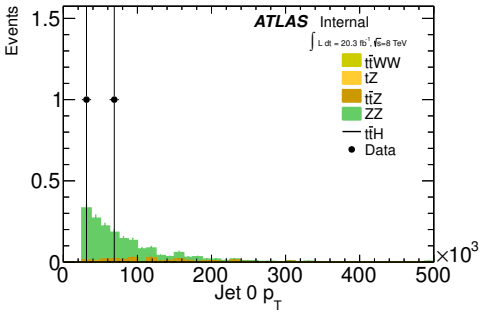
Figure 6: Jet-inclusive 4-lepton ZZ validation region using the $t\bar{t}H$ lepton identification and momentum cuts



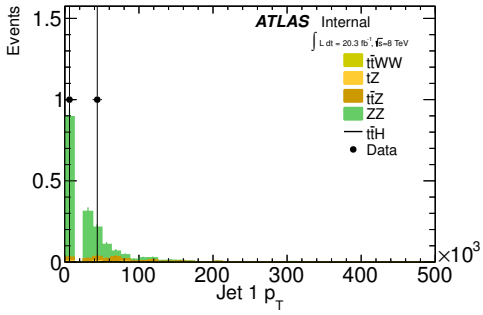
(a) Jet multiplicity



(b) b-Tagged jet multiplicity



(c) Leading jet p_T



(d) Sub-leading jet p_T

Figure 7: $ZZ + b$ validation region using the $t\bar{t}H$ lepton identification and momentum cuts

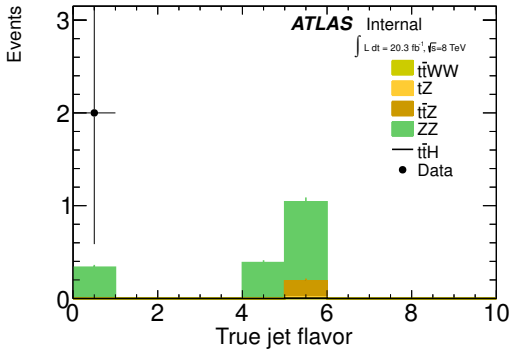


Figure 8: True flavour of leading (highest energy) b-tagged jet in $ZZ + b$ validation region

5 Source of fake leptons in $t\bar{t}$ events

$t\bar{t}$+jets events contain only 2 opposite prompt leptons. $t\bar{t}$+jets events can contribute to 2 same sign lepton channel and 3 lepton channel via production of additional non-prompt leptons. In order to understand sources of those non-prompt leptons, a special sample of $t\bar{t}$+jets MC simulation was produced which includes full MC truth records. This sample was used to carry following studies:

- Match reconstructed electrons and muons to stable generator leptons (truth leptons);
- Tabulate parent particle of the truth leptons matched to reconstructed leptons;
- Study sources of non-prompt leptons as a function of jet multiplicity;
- Study sources of non-prompt leptons for leptons with loosened selection criteria.

To simplify analysis and increase statistical precision, this section uses following simplified event categories:

- 2 reconstructed same sign leptons with $p_T > 10$ GeV and ≥ 1 reconstructed b-jet (2l);
- 3 reconstructed leptons with $p_T > 10$ GeV and ≥ 1 reconstructed b-jet (3l).

At least one of the leptons is required to match a triggered lepton. Tight reconstructed leptons are selected using criteria employed for 3 lepton analysis as described in Note 1. Loose reconstructed leptons serve as proxies for non-prompt (fake) leptons; they are selected by reversing the calorimeter isolation selection ($etcone20/p_{T,l} > 0.1$) and loosening the track isolation selection to $ptcone20/p_{T,l} < 1.0$.

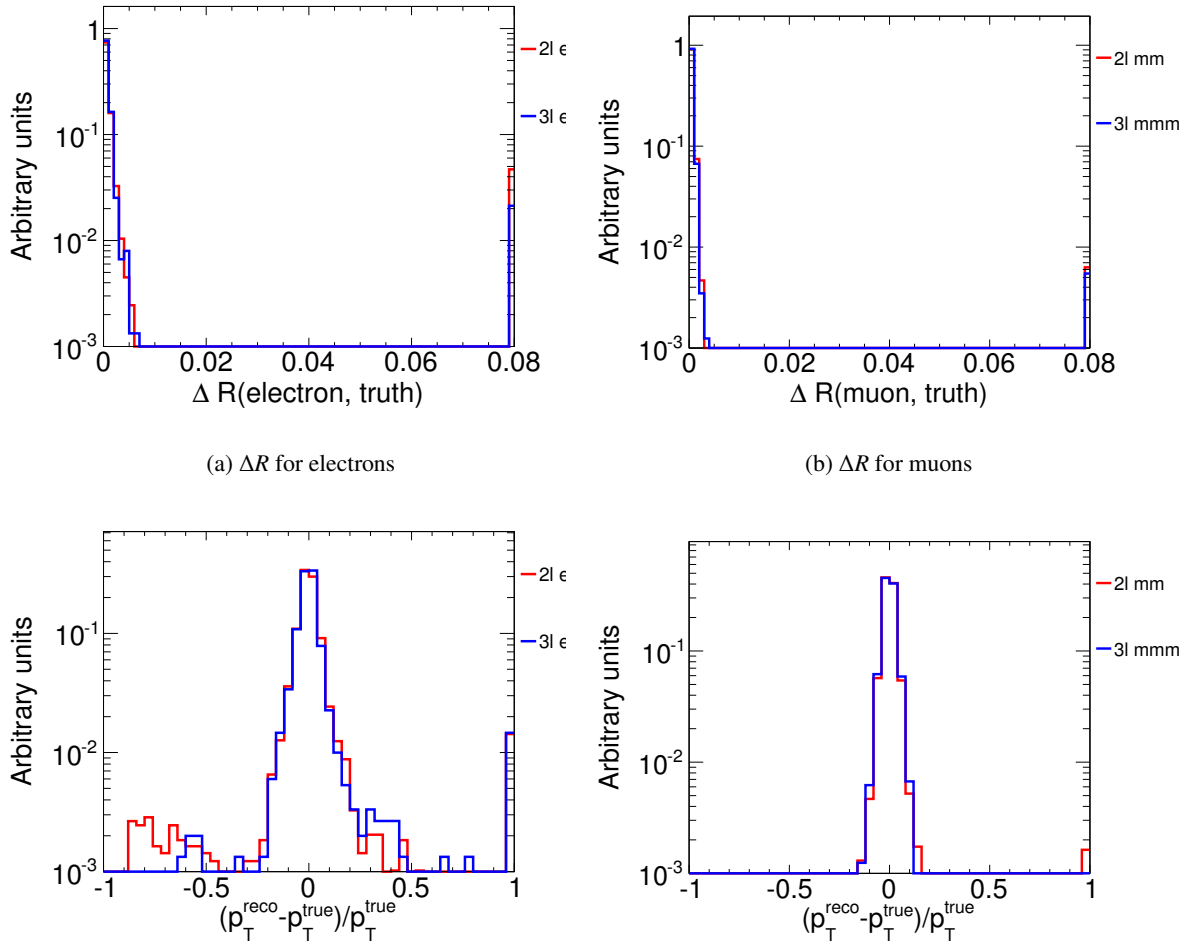
The above 2l and 3l categories are further subdivided into 2 cases: all leptons are selected with the tight selections; exactly one lepton is selected with the loose selections and the remaining leptons pass the tight selections. In plots shown in this section, the following convention is used:

- e - tight electron;
- d - loose electron;
- m - tight muon;
- p - loose muon.

The overall idea for estimating non-prompt lepton in this analysis is based on the following assumption: $t\bar{t}$ events with at exactly 1 loose leptons have similar jet multiplicity distribution as $t\bar{t}$ events with all tight leptons. Then the events with less than 3 jets can be used to measure a scale factor between loose non-prompt leptons and tight non-prompt leptons. This scale factor is then applied to events with high jet multiplicity and one loose lepton to estimate $t\bar{t}$ background in the signal region with high jet multiplicity and all tight leptons. The underlying principles of this procedure are tested with the MC simulation as described in this section.

5.1 Truth matching and lepton parents

The reconstructed electrons and muons are matched to stable (status code 1) generator leptons using a geometrical matching. Figure 9 (Figure 10) shows ΔR distance to and Δp_T for the closest generator lepton for the tight (loose) reconstructed leptons. For the both types of leptons, $\geq 99\%$ of reconstructed muons and $\geq 95\%$ of reconstructed electrons are successfully matched to the truth lepton.



(a) ΔR for electrons

(b) ΔR for muons

(c) Δp_T for electrons

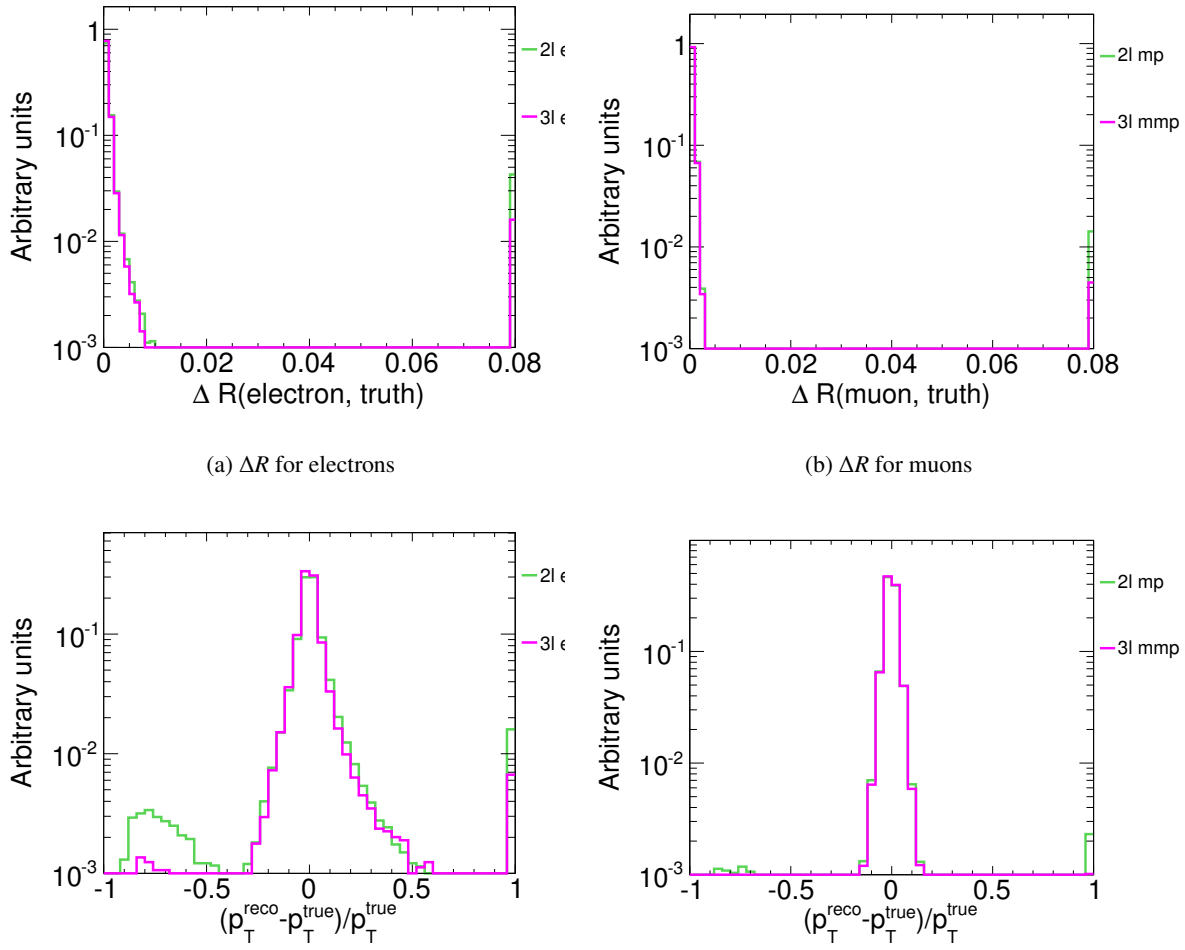
(d) Δp_T for electrons

Figure 9: Lepton truth matching for tight leptons

PDG ID	Parent	Leptons	%
24	$W^{\pm}$	2342	47.8
511	B^0	919	18.8
521	$B^{\pm}$	832	17.0
0	No match	230	4.7
531	B_s^0	158	3.2
5122	Λ_b^0	141	2.9
15	$\tau^{\pm}$	125	2.6
443	$J/\psi(1S)$	46	0.9
5132	$\Xi_b^{\pm}$	28	0.6
411	$D^{\pm}$	26	0.5
5232	Ξ_b^0	24	0.5
421	D^0	12	0.2
111	π_0	8	0.2
	Total	4891	99.9

PDG ID	Parent	Leptons	%
24	$W^{\pm}$	4401	47.8
511	B^0	1955	21.2
521	$B^{\pm}$	1699	18.4
531	B_s^0	309	3.4
5122	Λ_b^0	309	3.4
15	$\tau^{\pm}$	249	2.7
411	$D^{\pm}$	60	0.7
0	No match	58	0.6
443	$J/\psi(1S)$	53	0.6
5232	Ξ_b^0	47	0.5
5132	$\Xi_b^{\pm}$	39	0.4
421	D^0	15	0.2
	Total	9194	99.8

Table 4: Parents of 2 tight leptons: electrons (left) and muons (right).



(a) ΔR for electrons

(b) ΔR for muons

(c) Δp_T for electrons

(d) Δp_T for electrons

Figure 10: Lepton truth matching for loose leptons

PDG ID	Parent	Leptons	%
24	$W^{\pm}$	27072	47.7
511	B^0	9764	17.2
521	$B^{\pm}$	9683	17.1
0	No match	2436	4.3
531	B_s^0	1912	3.4
15	$\tau^{\pm}$	1805	3.2
5122	Λ_b^0	1489	2.6
421	D^0	774	1.4
411	$D^{\pm}$	646	1.1
111	π_0	352	0.6
443	$J/\psi(1S)$	234	0.4
5232	Ξ_b^0	202	0.4
5132	$\Xi_b^{\pm}$	180	0.3
431	$D_s^{\pm}$	155	0.3
	Total	56704	99.8

PDG ID	Parent	Leptons	%
24	$W^{\pm}$	57406	47.7
511	B^0	20735	17.2
521	$B^{\pm}$	20668	17.2
531	B_s^0	4143	3.4
15	$\tau^{\pm}$	4123	3.4
5122	Λ_b^0	3704	3.1
411	$D^{\pm}$	2724	2.3
421	D^0	2447	2.0
0	No match	1742	1.4
431	$D_s^{\pm}$	806	0.7
5232	Ξ_b^0	508	0.4
443	$J/\psi(1S)$	451	0.4
5132	$\Xi_b^{\pm}$	450	0.4
4122		286	0.2
	Total	120193	99.9

Table 5: Parents of tight plus loose leptons: electrons (left) and muons (right).

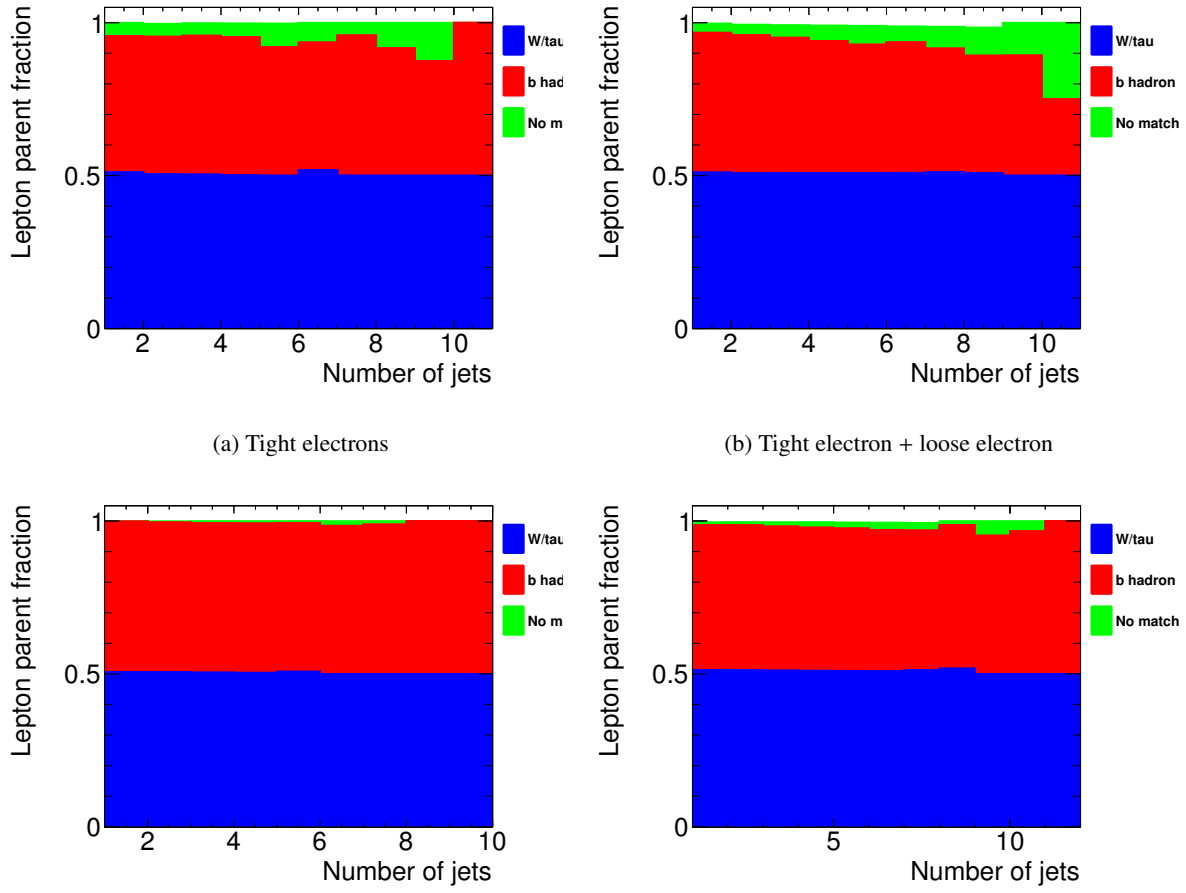
Tables 4 and 5 show generator parents of the truth leptons matched to the reconstructed leptons for 2l and 3l categories with all tight leptons or at least one loose lepton. In the 2l category, exactly one lepton is always from prompt decay of W or τ. The second lepton is produced in semi-leptonic decay of b-hadrons. Similar conclusions apply to the 3l category where 2 leptons are prompt and 3rd lepton is produced in semi-leptonic decay of b-hadrons. These conclusions apply both to tight and loose lepton categories which means that the loose leptons are good proxies for non-prompt tight leptons in $t\bar{t}$ events.

5.2 Sources of reconstructed leptons and jet multiplicity

For electrons, each reconstructed electron is always reconstructed as a jet because of clustered electromagnetic energy distributions in the calorimeters. Both for tight and loose electrons, the jet overlapping with the electron is removed from the analysis.

The tight non-prompt muons from semi-lepton b-hadron decay pass the isolation selection which means that the underlying b-jet is often not reconstructed as a jet, since otherwise the muon would not be isolated. In addition, the muon overlapping with the reconstructed jet us removed from the analysis which acts as an additional isolation requirement. The loose non-prompt muons from semi-lepton b-hadron decay fail the isolation selection which means that the underlying b-jet is often reconstructed as a jet. In order to allow direct correspondence between events with tight muons and events with one loose muon, the jet overlapping with the loose muon is removed from the analysis and the muon is kept.

Fraction of non-prompt leptons from b-hadrons is shown in Figures 11 and 12. The b-hadron fraction is constant for muons and changes by less than 20% for electrons.



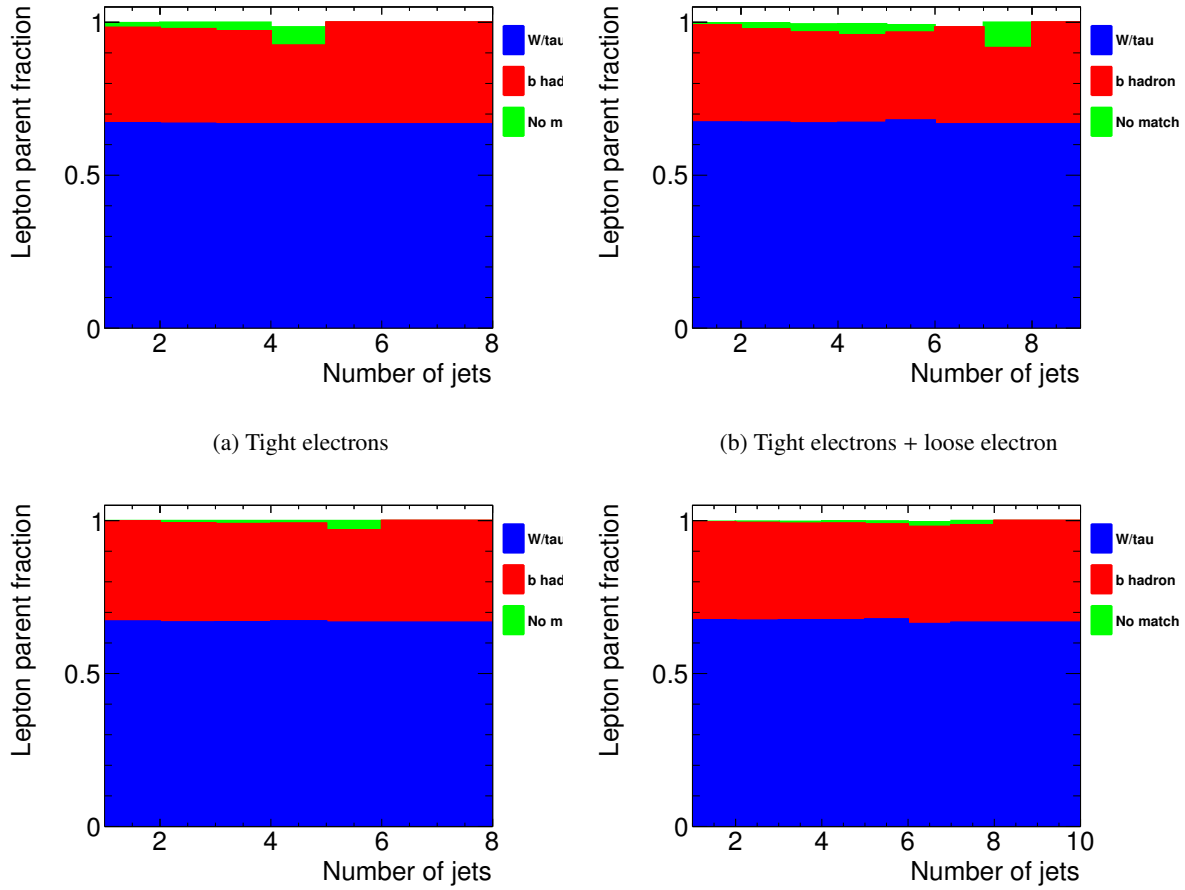
(a) Tight electrons

(b) Tight electron + loose electron

(c) Tight muons

(d) Tight muon + loose muon

Figure 11: Lepton sources a function of number of jets in 2-lepton channel



(a) Tight electrons

(b) Tight electrons + loose electron

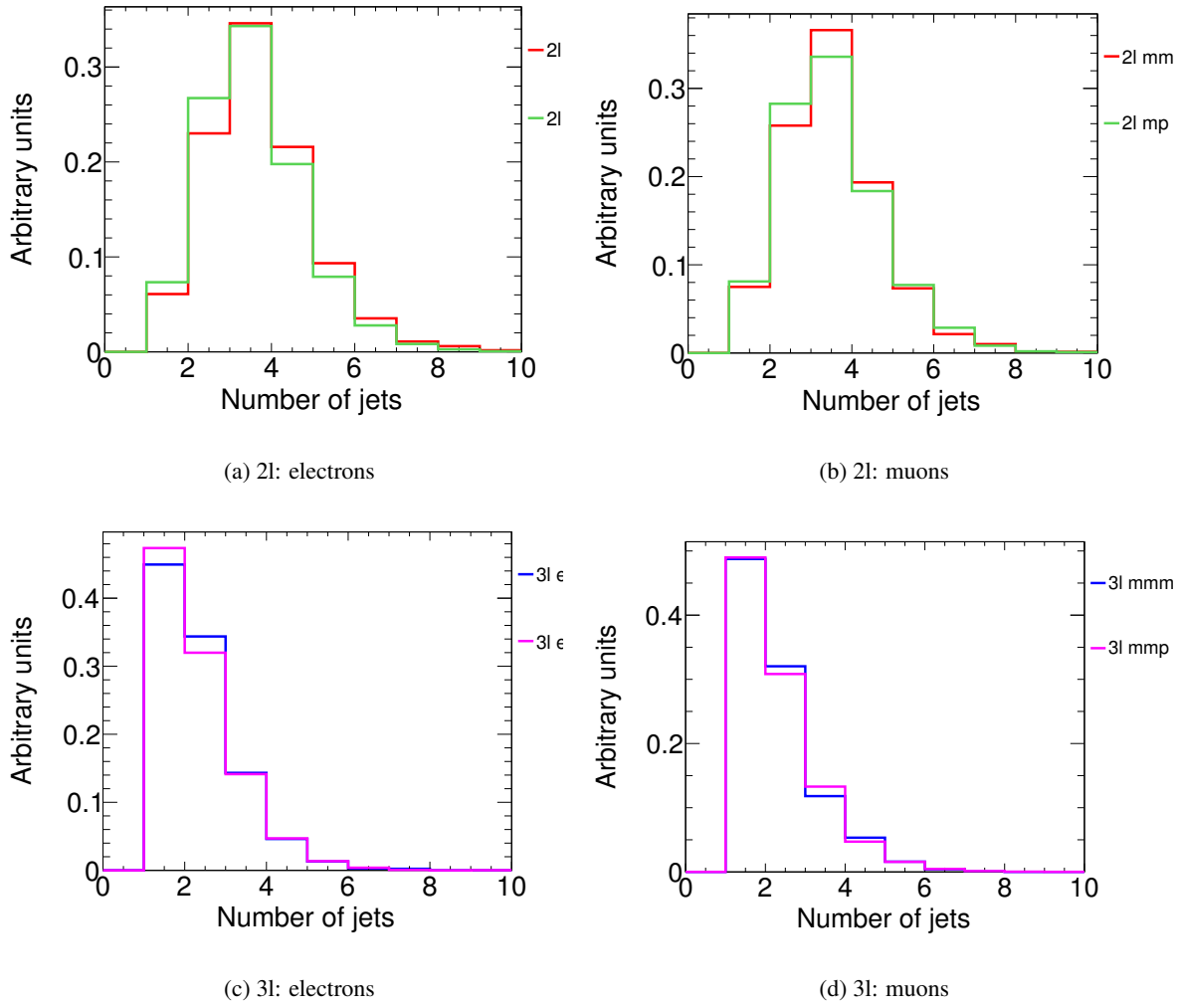
(c) Tight muons

(d) Tight muons + loose muon

Figure 12: Lepton sources a function of number of jets in 3-lepton channel

5.3 Jet multiplicity for tight and loose lepton regions

The analysis strategy is to use events with low jet multiplicity to measure correspondence between loose and tight leptons. Potential differences in jet multiplicity distributions between events with tight leptons and events with one loose leptons are evaluated using the MC simulation of $t\bar{t}$ events. Figure 13 shows shape of jet multiplicity distribution for the 2l and 3l events with all tight leptons and with one loose leptons. Ratios of corresponding distributions are shown in Figure 14. To within $\approx 20\%$, the jet multiplicity shape is the same for loose and tight leptons, and also for 2l and 3l events.



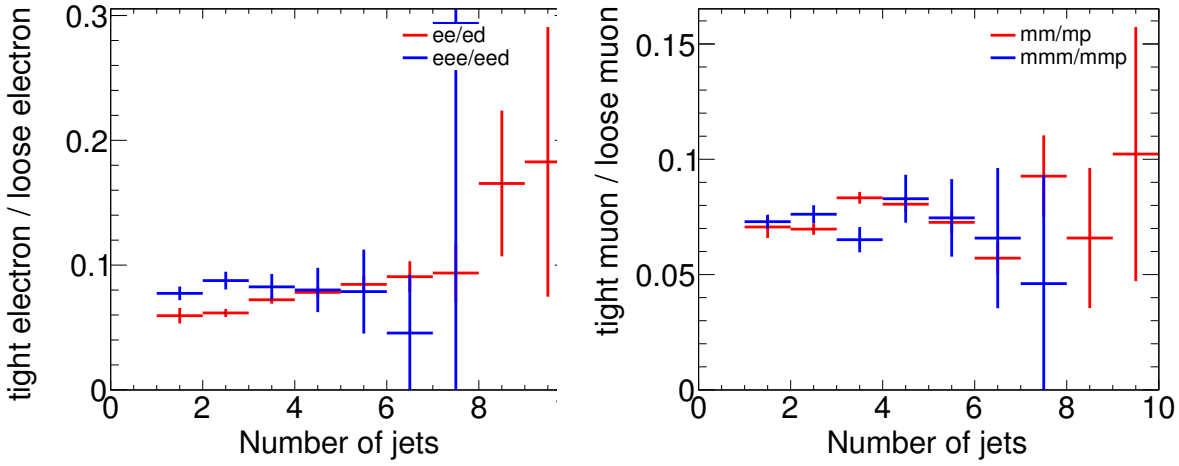
(a) 2l: electrons

(b) 2l: muons

(c) 3l: electrons

(d) 3l: muons

Figure 13: Jet multiplicity in 2-lepton (left) and 3-lepton (right) channels.



(a) Tight over loose electrons

(b) Tight over loose muons

Figure 14: Ratios of regions with tight and loose leptons in 2-lepton and 3-lepton channels

6 3 lepton channel

6.1 Estimation of $t\bar{t}$ background events with a secondary or fake lepton

6.1.1 Estimation strategy

Because the probability that 2 leptons are both fake or secondary is very small ($\approx$ 1 to 5% depending on the lepton p_T and η, compared to an expected uncertainty on the final fake yields at the level of 20-30% at least), we can assume that, in the case of the 3 lepton channel, the only relevant $t\bar{t}$ background contribution comes from di-leptonic decays (both W from top quarks decay leptonically). In this case the two real, primary leptons are accompanied by a third, fake, lepton. By definition, then, the lepton with an electric charge opposite in sign to the other two, is less likely to be a fake lepton (again, we are assuming that a double charge flip is a very unlikely event due to the small charge mis-identification probability). The fake lepton will be one of the two with same-sign charge. For this reason, a higher p_T cut (20 GeV instead of 10) is adopted for these leptons [1]. Lepton fakes are essentially only originated in the semileptonic decay of heavy flavour (HF) quarks, as shown in the MC truth-based study described in Section 5.

A method to estimate the contribution of $t\bar{t}$ background in the SR is developed stemming from these considerations. The background contribution is estimated *in-situ*: the overall normalization of the $t\bar{t}$ background background is taken from data, in a control region mostly populated by $t\bar{t}$ background events; while the shapes are taken from MC and validated with data in a different region.

More in detail, our strategy is as follows. The details of the implementation and associated results and plots are then provided in the following sections.

Initial normalization A starting number of $t\bar{t}$ background events is taken from data in an $t\bar{t}$ background -enriched control region with two tight leptons (i.e. leptons passing all the signal region selections listed in [1]), plus a third, *anti-tight* lepton. This is defined as a lepton passing most of the "tight" selections, but explicitly failing some specific cuts: this means that the CR is orthogonal to the SR. The event-based selections are the same as for the SR, so to have an as-small-as-possible extrapolation from CR to SR, to minimize the assumptions and associated uncertainties. The definitions of the CR are provided in the following for muons (Section6.1.3) and electrons respectively (Section 6.1.3).

Extrapolation into the SR A transfer factor θ for the events in the CR to end up in the SR is evaluated as the ratio of the number of $t\bar{t}$ background events with 3 leptons all tight, over the number of 2 tight + 1 anti-tight of the CR:

$$\theta = \frac{N_{\ell\ell\ell}}{N_{\ell\ell\not{\ell}}} \tag{1}$$

This represents the relative amount of $t\bar{t}$ background events in the SR over those in the CR and is estimated on $t\bar{t}$ background MC, where the numerator is accessible because it is not blinded. θ is a single number, computed inclusively for the same lepton kinematic regions of SR and CR. The accuracy of θ in MC is assessed by defining an auxiliary region, different from the CR, and also highly enhanced in $t\bar{t}$ background . The relevant variables (cuts on which are inverted to define the anti-tight lepton) are separately compared in shape between data and MC in a top quark-enriched auxiliary region to establish the accuracy of the simulation; and any difference results in a systematic uncertainty on θ. The definitions of the CR are provided in the following for muons (Section6.1.3) and electrons respectively (Section **??**).

SR yield extraction Finally, the estimated number of $t\bar{t}$ background events in SR is obtained putting together the above inputs, as summarized in the following expressions:

$$N_{data,SR} = N_{data,CR} \cdot \theta \tag{2}$$

Measuring the fakes in this way breaks the modeling into two pieces. The number of events in the CR depends on the $t\bar{t}$ production cross-section including additional jets and b-jets as well as the part of the object-level jet-to-anti-tight lepton fake rate. This piece is entirely obtained from data. The second piece, the θ extrapolation factor, is purely object-level and depends on the modeling of a few key variables of the fakes, namely the isolation and object id. This piece is obtained from MC directly, but the important distributions are verified in data.

6.1.2 Definition of signal region

The event selections defining the signal region for the 3 lepton channel, presented in Note1 [1], are listed here as a reminder to the reader:

- three tight leptons, the sum of the charges of which should be ± 1;
- the event must have passed one of the following triggers: `EF_e24vhi_medium1, EF_e60_medium1, EF_mu24i_tight, EF_mu36_tight` and one of the leptons has to be matched to an EF ROI;
- N(b-tags, MV1) ≥ 1 and N(jets, $p_T > 25$ GeV, $|\eta| < 2.5$) ≥ 4 **OR**
- N(b-tags, MV1) ≥ 2 and N(jets, $p_T > 25$ GeV, $|\eta| < 2.5$) $= 3$
- Invariant mass of two leptons, if of same flavor and opposite charge, should not be in the range [81, 101] GeV, to suppress the Z+jets contribution (Z veto).

In particular, it should be noticed that, with respect to the current version of Note1, an extended signal region is considered, which adds events with N(b-tags, MV1) ≥ 2 and N(jets, $p_T > 25$ GeV, $|\eta| < 2.5$) $= 3$. The benefit of this will be shown in Note 3, where the expected limit on $\frac{\sigma}{\sigma_{SM}}$, the ratio between the production cross-section of the signal corresponding to the current data set and the Standard Model expectation, is improved by $\approx 30\%$ when including this rather pure region. Estimations of background are presented here for both signal regions.

6.1.3 Muon fakes

$t\bar{t}$ background -enriched control region The control region considered is the one populated by events with 2 tight leptons of whichever flavor (e or μ, collectively labeled as "x") plus one anti-tight muon. The latter is a muon passing the analysis pre-selections, but whose isolation is large (ptcone20/$p_T > 0.1$ and Etcone20/$p_T > 0.1$) and for which the muon-jet overlap removal request ($\Delta R(\mu, \text{closest selected jet}) < 0.04 + 10\ GeV/p_T$) is not applied. This region is labeled "CR1" and the anti-tight muon is indicated as "p" in the following, following the convention of Section 5.

The rest of the selections are the following:

- electrons identified using the `VeryTightLikelihood` word, and passing the η, $d_0/\sigma(d_0)$, $z_0 \sin\theta$ cuts, isolation cuts of Table 7 in [1];
- the two tight leptons should have opposite electric charge;
- the event must have passed one of the following triggers: `EF_e24vhi_medium1, EF_e60_medium1, EF_mu24i_tight, EF_mu36_tight` and one of the leptons has to be matched to an EF ROI;
- N(b-tags, MV1) ≥ 1 and N(jets, $p_T > 25$ GeV, $|\eta| < 2.5$) ≥ 4 **OR**
- N(b-tags, MV1) ≥ 2 and N(jets, $p_T > 25$ GeV, $|\eta| < 2.5$) $= 3$

- Invariant mass of two leptons, if of same flavour and opposite charge, should not be in the range [81, 101] GeV, to suppress the Z+jets contribution (Z veto).

The latter are the same as in the signal region, defined in paragraph 6.1.2.

The expected number of $t\bar{t}$ background events in the control region from MC is presented in Table 6.

Control Region	Sample	N(events) in CR
CR1	Data	399.00 ± 20.02
	$t\bar{t}$	301.01 ± 1.87
	single top	11.12 ± 0.73
	other processes	34.41 (unc. negl.)
	data-other processes ("dt")	364.62 ± 20.02
	$t\bar{t}$ + single top ("mc")	312.13 ± 2.01
ratio	$\frac{\text{dt}}{\text{mc}}$	1.17 ± 0.08

Table 6: Expected number of $t\bar{t}$ background , single top and other MC events; and in data for the control region CR1. The uncertainty on the "other" processes is negligible with respect to the data statistical uncertainty.

The relative statistical uncertainty for data is 5%. The contribution of processes other than $t\bar{t}$ background and single top, estimated from MC, is about 10%.

While the only important quantity for the final estimation of fakes is the number of events in data (because MC is not used anywhere in the overall normalization), it is interesting to compare important kinematic and topological variables in CR1 between data and simulation. Comparison plots are presented in Figure 15 for global event quantities; in Figure 16 for anti-tight muon quantities for the case of CR1 before applying the Z mass veto (which allows to make comparisons with higher statistics without a large bias on the features of the fakes); and in Figure 17 after the Z mass veto.

Transfer factor into the signal region The second stage of the procedure is to extrapolate the number of fake muon events from the control region into the signal region. This corresponds to reverting the selections of the anti-tight muons into a tight muon, so to restore the two tight leptons + one tight muon configuration of SR. As mentioned in Section 6.1.1, this is done using simulated distributions. The resulting transfer factor θ_μ is simply the ratio of events in the SR over those in CR1:

$$\theta_\mu = \frac{N(x\mu\mu \text{ events})}{N(x\mu p \text{ events})} \tag{3}$$

In the above "x" indicates both electrons and muons inclusively. The above definition of θ_μ, in which the second and third lepton must be a muon, guarantees that events with an electron fake are not included in the estimation from the numerator, although reducing the sample size available for the determination of θ_μ. Because the anti-tight muon "p" can be either the second or the third lepton, this definition correctly account for all permutations of tight+fake muon events. When applied to the inclusive "xxp" population of CR1, it returns the estimate of all and only the events with a muon fake in the SR ($xx\mu_{fake} + x\mu_{fake}x$).

The level of accuracy of the MC prediction in terms of shapes of the variables used in the CR → SR extrapolation is checked in a third region, called *auxiliary*, highly enriched in $t\bar{t}$ background . This region must have enough statistics to allow a meaningful comparison and be sensitive to systematic differences between the model and the data. The auxiliary region is defined as follows:

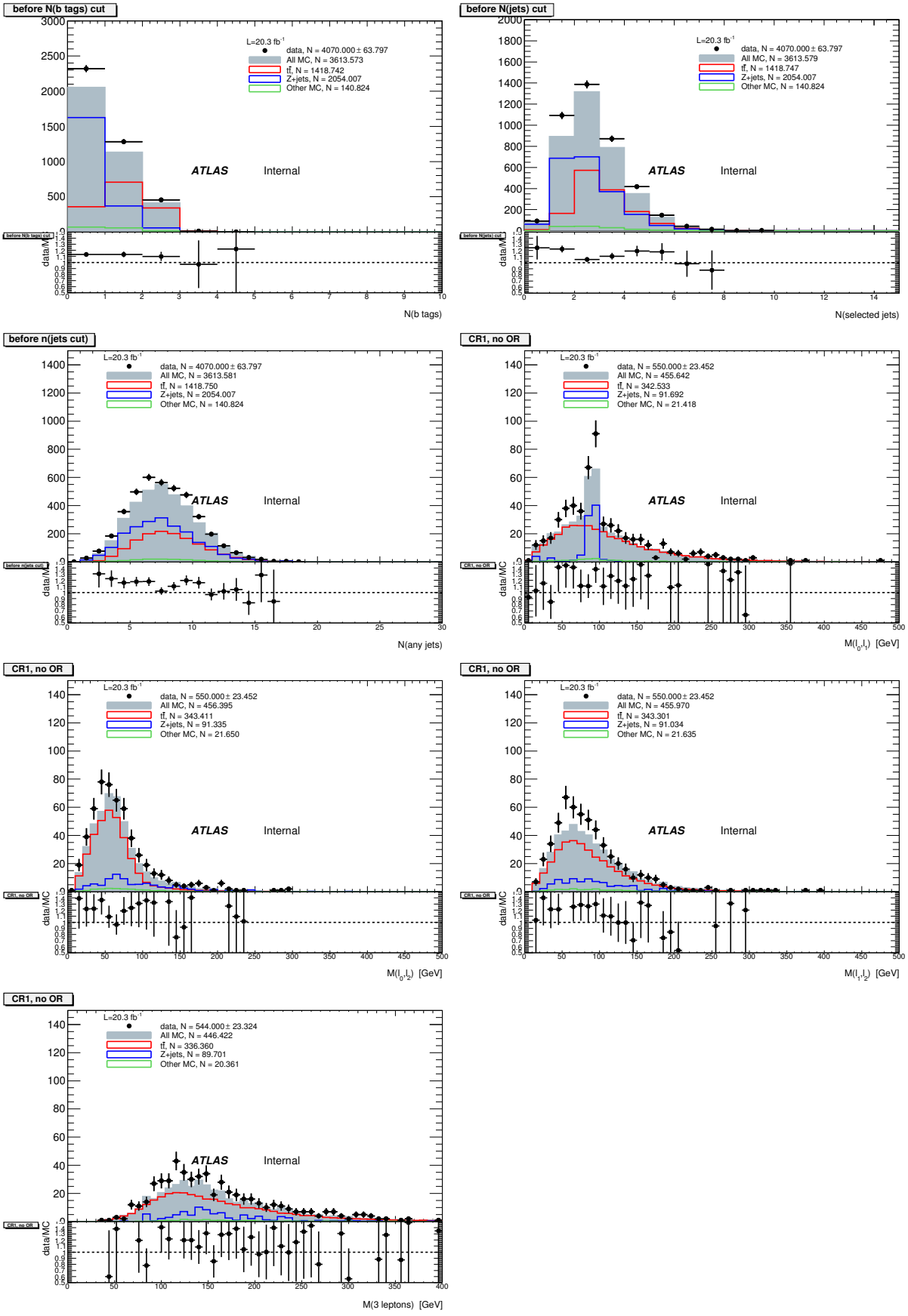


Figure 15: Top: Distributions of event quantities in CR1 in data (dots), compared with the total simulation (grey) and with overlaid the $t\bar{t}$ background distribution (red), Z+jets (blue) and all other processes (green). Each quantity is shown before a cut on the same quantity is applied. Since the b tagging requirement comes first in the cut flow, this is the most populated distribution. The uncertainty on the data distribution is statistical. The Data/MC ratio is also presented in each bottom pad. From top left to bottom right the quantities shown are: N(b tags), N(selected jets), N(any jets in the event), invariant mass of the two tight leptons, and of the two tight leptons with the anti-tight muon, and the invariant mass of the three leptons together.

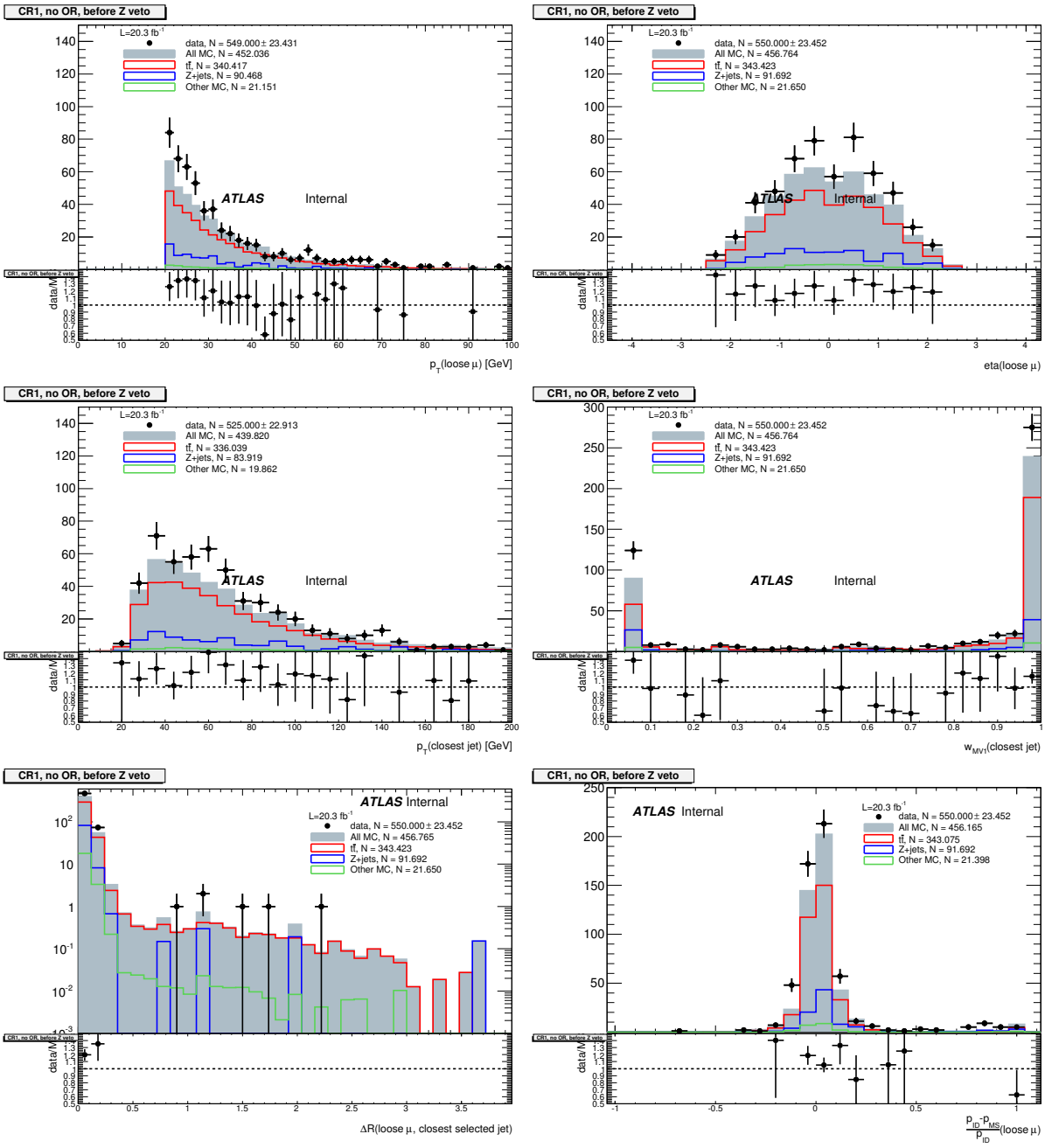


Figure 16: Top: Distribution of for the anti-tight muons in the control region CR1, before the application of the Z veto in data (dots), compared with the total simulation (grey) and with overlaid the $t\bar{t}$ background distribution (red), Z+jets (blue) and all other processes (green). The uncertainty on the data distribution is statistical. The number of events for each of them is also presented in the legend pad. Bottom: the Data/MC ratio is presented for the same distribution. The quantities presented are, from top left to bottom right, p_T, η, p_T(closest jet),w_{MV1}(closest jet), $\Delta R(\mu,,$closest selected jet), imbalance between the ID and MS momentum measurements.

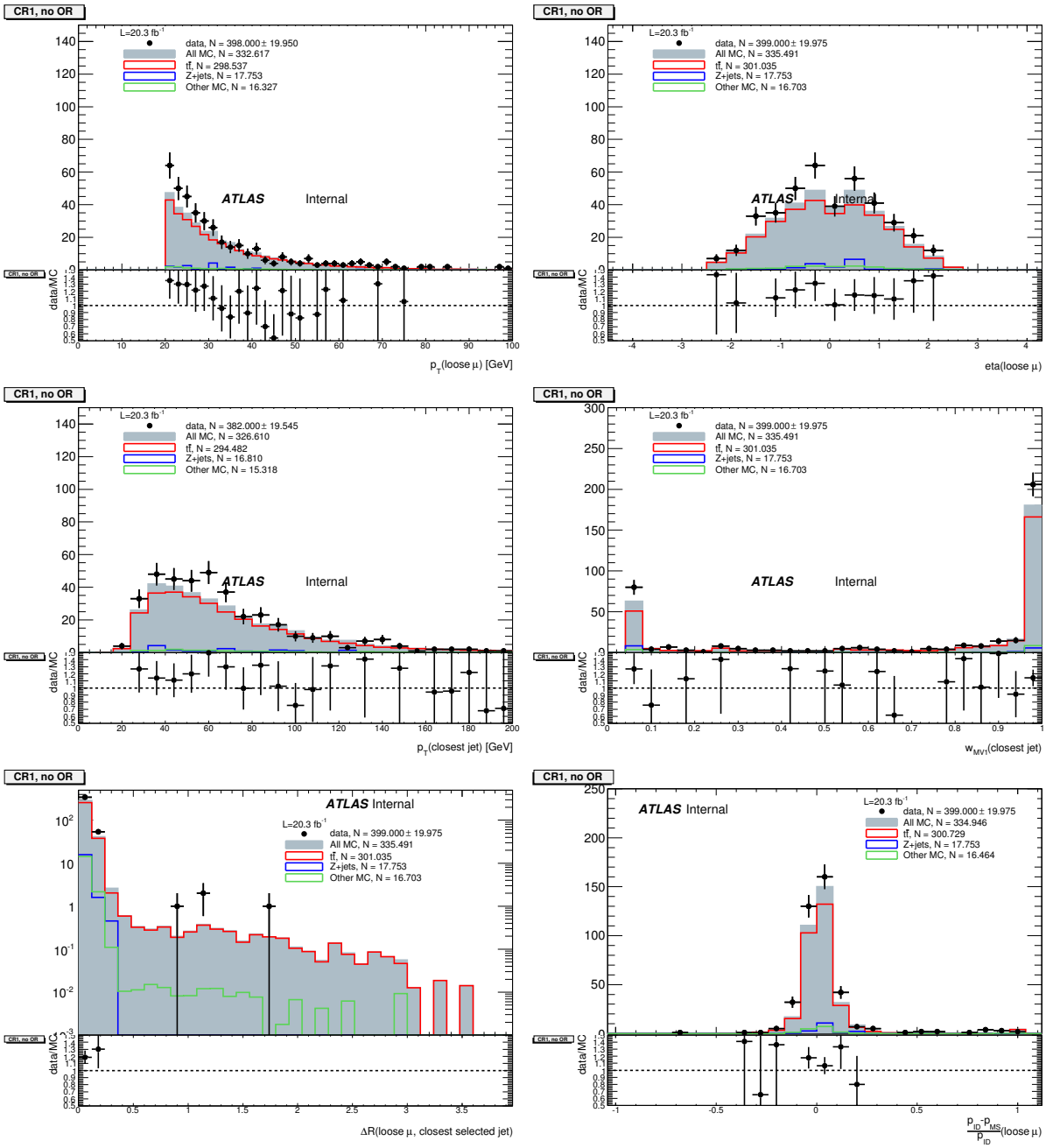


Figure 17: Top: Distribution of for the anti-tight muons in the control region CR1, after the application of the Z veto in data (dots), compared with the total simulation (grey) and with overlaid the $t\bar{t}$ background distribution (red), Z+jets (blue) and all other processes (green). The uncertainty on the data distribution is statistical. The number of events for each of them is also presented in the legend pad. Bottom: the Data/MC ratio is presented for the same distribution. The quantities presented are, from top left to bottom right, p_T, η, p_T(closest jet),w_{MV1}(closest jet), $\Delta R(\mu,,$closest selected jet), imbalance between the ID and MS momentum measurements.

- 2 x + 1 loose μ, with sum of charges $|Q_{tot}| = 1$. (Loose muons are those passing the identification, η, $d_0/\sigma(d_0)$ and $z_0 \sin\theta$ cuts, but for which the p_T, overlap removal with jets and isolation cuts are not imposed);
- triggers as SR;
- N(b tags) and N(selected jets) as SR;
- $E_T^{miss} > 50$ GeV;

The cut on the E_T^{miss} is aimed at suppressing especially the Z+jets background. `METRefFinal` is used. The distribution of `METRefFinal` for $t\bar{t}$ background and Z+jets in this region, before cutting, is shown in Figure 21, which also shows that the simulation models this quantity accurately in the top quark-dominated region. There is evidence of mismodelling especially at low METRefFinal, probably owing to an under-estimation of the Z+jets background.

The requests on the N(selected jets) and N(b tags) are the same as for the SR and CR1. The shape of the p_T(loose muon) and, especially, ΔR(loose muon, closest selected jet) are dependent on the jet and tag multiplicities: the busier the event, the higher the probability that the muon overlaps with a jet even if the two are not related. Furthermore, in a 2 b-tag event 2 b-jets pass all cuts (i.e. both energetic enough to pass the 25 GeV p_T cut): so the muon is more likely to show an overlap (and therefore be removed) in 2 b-tag events than in 1 b-tag events.

While the SR also includes the Z veto, as defined in Section 6.1.3, data-MC control plots are at pre-veto level. This is to maximize the statistics available. It is checked on the simulation that the shapes of the variables p_T(loose muon) and ΔR(loose muon, closest selected jet) do not depend on the application of the Z veto: this is shown in Figure 18.

The yields in this region are summarized in Table 7. The contamination from processes other than $t\bar{t}$

Sample	Yield in auxiliary region
Data	2046 ± 45
Total MC	1746
$t\bar{t}$	1583
Z+jets	65
Others(incl. signal)	99

Table 7: Expected number of $t\bar{t}$ background , Z+jets and other MC events; and of collision data in the auxiliary region defined in the text.

is expected to be 9%.

The control plots of data/MC agreement of the loose muon properties in the auxiliary region are provided both for the normalization (each MC normalized to the integrated luminosity) and for the shape (total MC normalized to the integral of the data distribution) of p_T(loose muon) and ΔR(loose muon, closest selected jet). The plots are displayed in Figures 19 and 20.

A trend is observed in the isolation variables, particularly ptcone20/p_T, whereby the fakes in data are more isolated than in the simulation. This suggests a different track multiplicity around the muon, possibly derived from a different modeling of the soft part of the hadronization of HF. This seems supported also by the difference in transverse momentum between data and MC. A 20% uncertainty band fully contains the level of disagreement. This is highlighted in Figure 20 with a bounding box around the ±20% agreement level in the ratio. An equivalent uncertainty is therefore applied to the θ_μ factor when extrapolating the $t\bar{t}$ background yield from CR1 into SR.

For reference, also the main event-based quantities are shown for this region in Figure 21

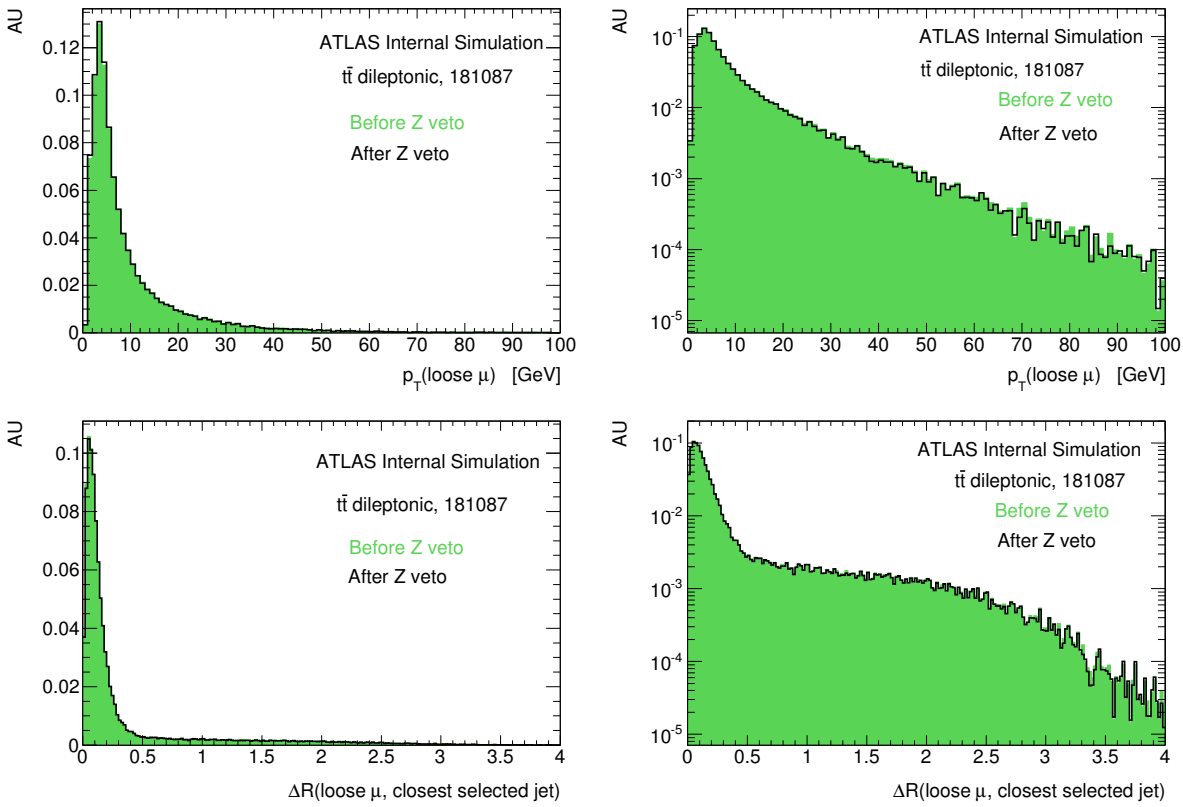


Figure 18: From top left to bottom right: p_T(loose muon) and ΔR(loose muon, closest selected jet) for a simulated sample of $t\bar{t}$ background di-leptonic decays, in linear and logarithmic scale (to check the tails). The full, light green, histogram is for the case before the Z veto is applied, while the black is for after the Z veto. Each distribution is separately normalized to unit area for a shape-only comparison. The events are from the auxiliary region.

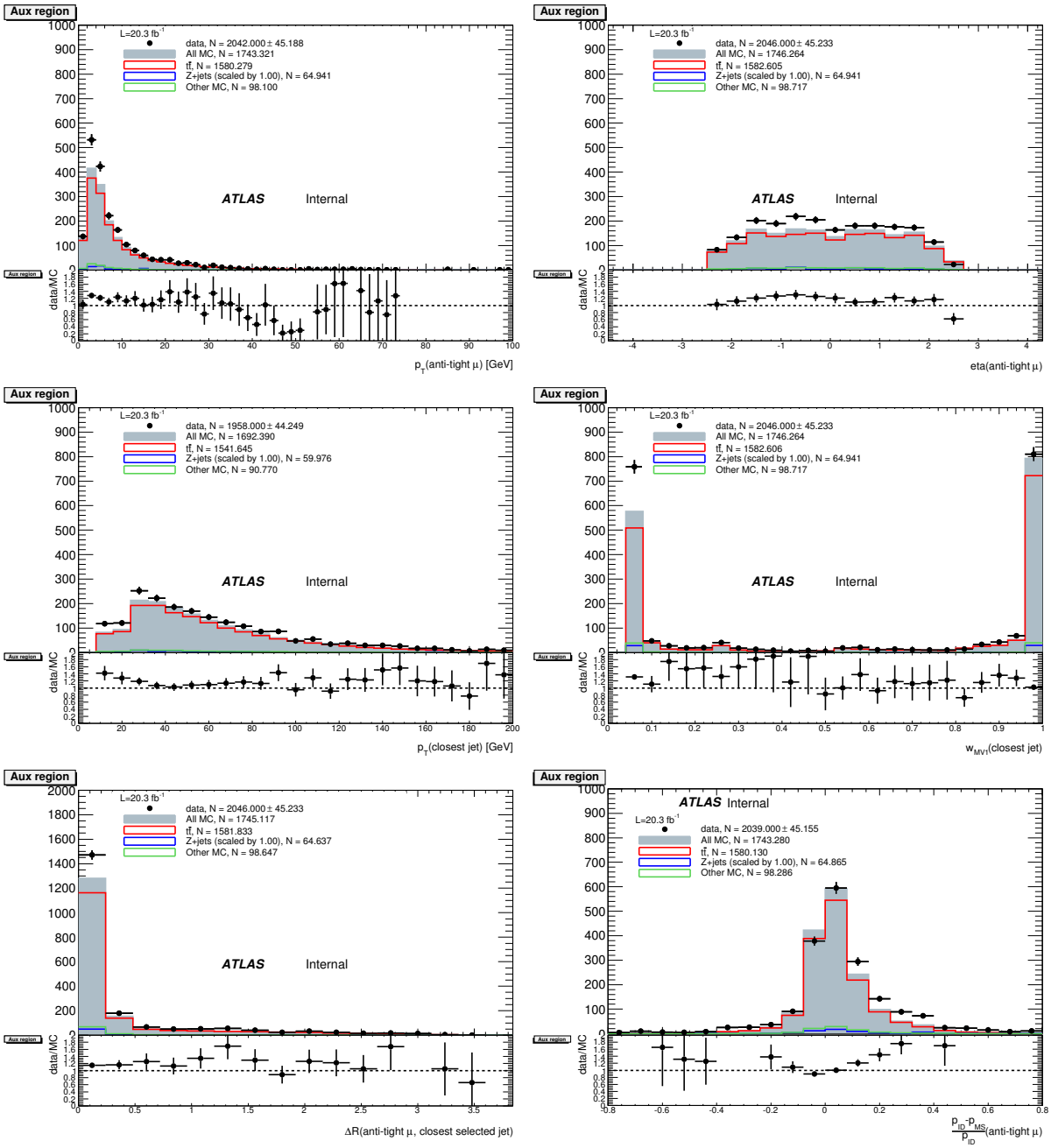


Figure 19: Top: Distributions of the properties of the loose muons in the auxiliary region, in data (dots), compared with the total simulation (grey) and with overlaid the $t\bar{t}$ background distribution (red), Z+jets (blue) and all other processes (green). The uncertainty on the data distribution is statistical. The number of events for each of them is also presented in the legend pad. Bottom: the Data/MC ratio is presented for the same distribution. The quantities presented are, from top left to bottom right, p_T, η, p_T(closest jet),w_{MV1}(closest jet), $\Delta R(\mu, ,$ closest selected jet), imbalance between the ID and MS momentum measurements.

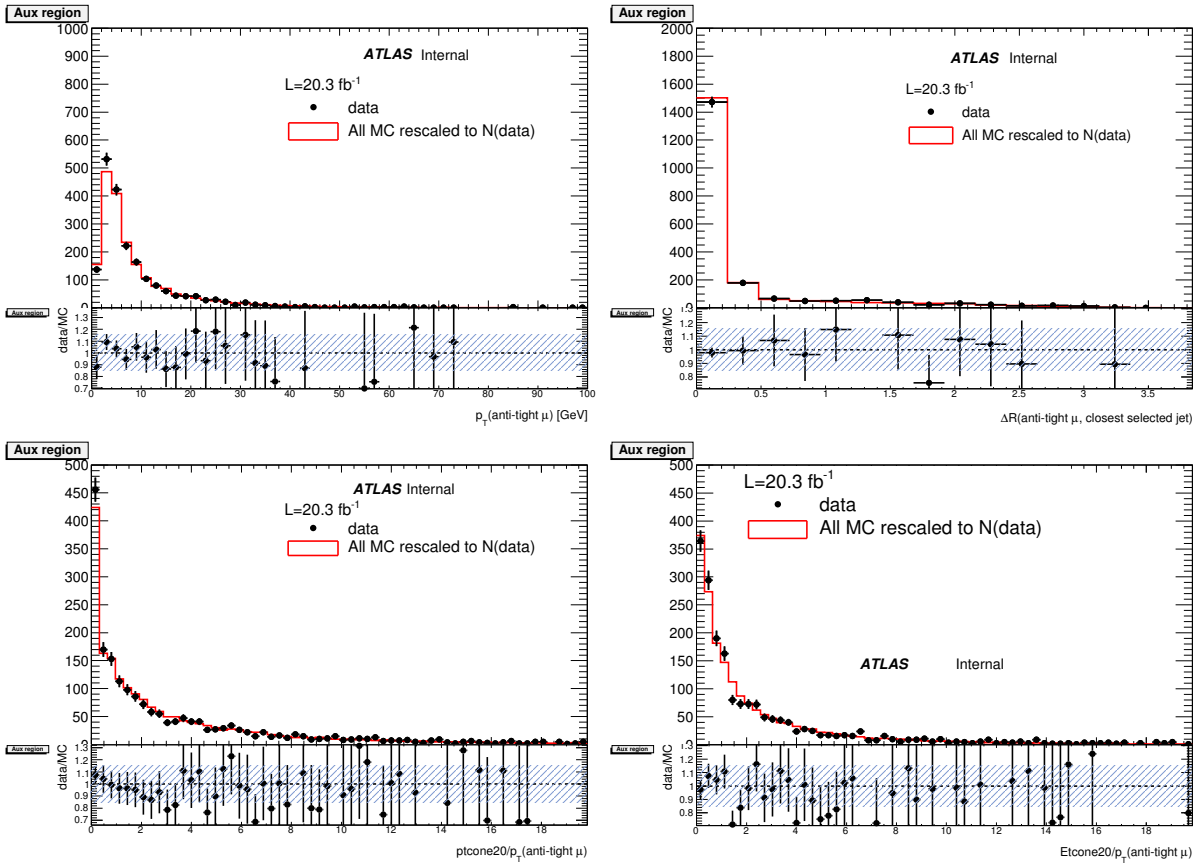


Figure 20: Distributions of the properties of the loose muons in the auxiliary region, in data (dots), compared with the total simulation (red line), rescaled to the integral of the data for a shape comparison. The uncertainty on the data distribution is statistical. The number of events for each of them is also presented in the legend pad. The variables probed are, top: p_T and $\Delta R(\mu, , \text{closest selected jet})$; bottom: ptcone20/$p_T$ and Etcone20/p_T. A ratio plot in the bottom pad, with a box containing the 20% area around 1, are displayed.

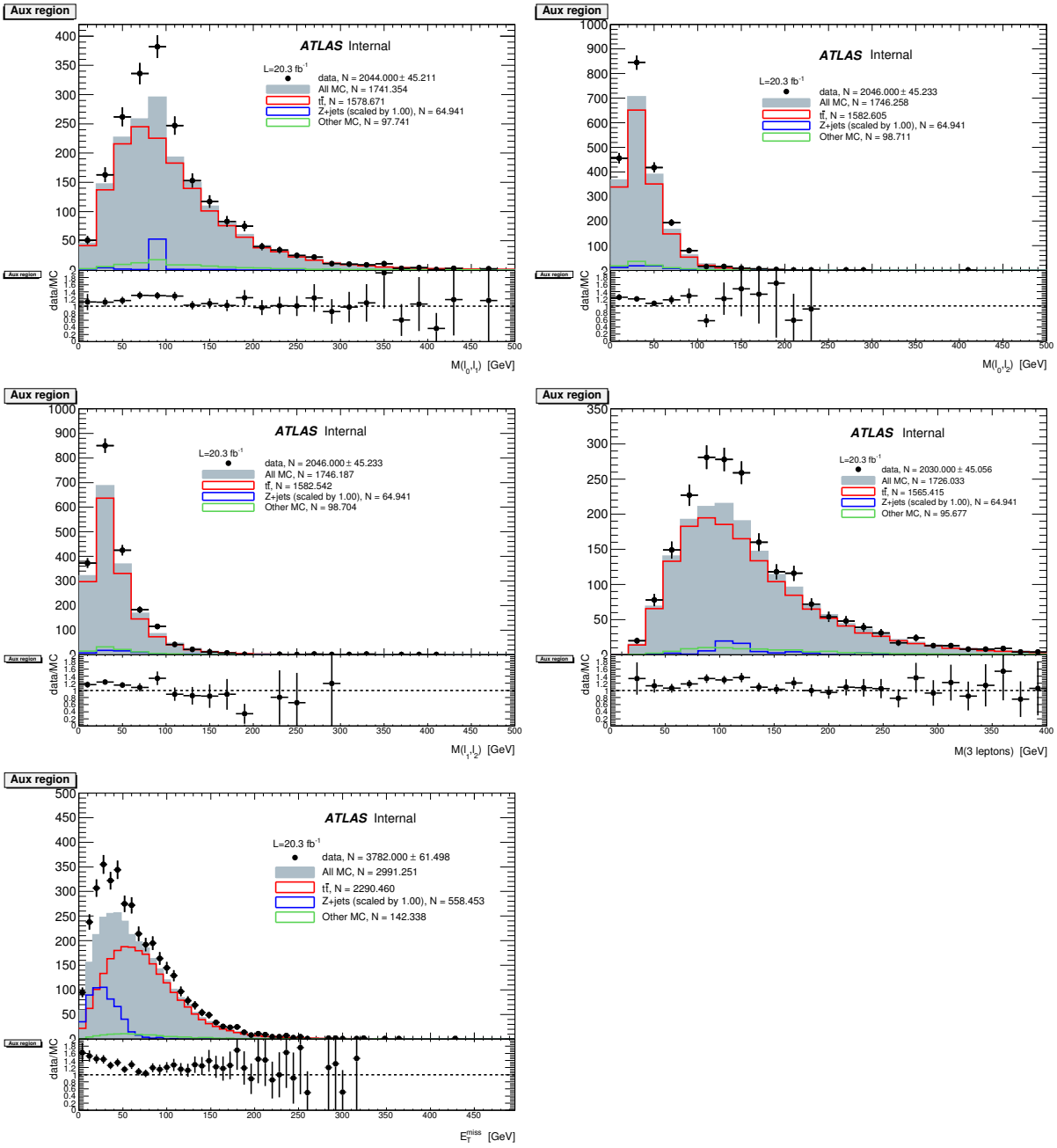


Figure 21: Top: Distributions of event quantities in the auxiliary region, in data (dots), compared with the total simulation (grey) and with overlaid the $t\bar{t}$ background distribution (red), Z+jets (blue) and all other processes (green). Each quantity is shown before a cut on the same quantity is applied. Since the b tagging requirement comes first in the cut flow, this is the most populated distribution. The uncertainty on the data distribution is statistical. The Data/MC ratio is also presented in each bottom pad. From top left to bottom right the quantities shown are: invariant mass of the two tight leptons, and of the two tight leptons with the anti-tight muon, and the invariant mass of the three leptons together; finally the E_T^{miss}.

The value of θ_μ extracted from the simulation ($t\bar{t}$ dileptonic sample), following the definition of Equation 3 is:

$$\theta_\mu = 0.0047 \pm 0.0011 \quad (4)$$

with a 23% relative overall uncertainty.

Results in the signal region Table 8 summarizes all the inputs for the final estimation of the number of $t\bar{t}$ background events with a fake muon in the SR; and the definition of the source sample/region.

Stage	Source (Sample/Region)	Value (and rel.unc.)
Overall starting normalization	CR1	364.62 ± 20.02 (5%)
Transfer factor θ_μ	SR/CR1, monitored with Aux sample	(0.47 ± 0.11)% (23%)

Table 8: Summary of regions and inputs to the extraction of the number of $t\bar{t}$ background events with a fake muon in the SR

For the $N_{\text{data},\mu,\text{SR}}$, the number of $t\bar{t}$ background events with a muon fake, we find:

$$N_{\text{data},\mu,\text{SR}} = 1.71 \pm 0.42 \quad (5)$$

with an overall relative uncertainty of 25%. The break down of the uncertainties is presented in Table 9

Source	Relative uncertainty on estimated N(fake muon events)
Sample size (stat) of data and "others" in CR1	5%
MC statistics for θ_μ	negligible
data/MC comparison of shapes in Aux	23%

Table 9: Break down of the uncertainties on the value of N(fake muon events).

6.1.4 Electron fakes

In general, the same approach is used for electron fakes in the 3 lepton channel as for muon fakes. In the following, the implementation of such strategy (already described above) and the differences with respect to the muon case are described.

$t\bar{t}$ background -enriched control region The sample composition of fake electrons contains one additional category to that of fake muons: γ conversions into e^+e^- pairs. To reflect this difference, and take into account also this source of fakes correctly, a dedicated choice of control regions, sensitive to conversions, is made. Namely, reverting or releasing the electron identification (`Very Tight Likelihood` (VTLH)) is used, because of the anti-conversion requests encompassed in this identification `word` [11]. The anti-tight electron is defined in the following way:

- inverting the VTLH word of the SR (i.e. vetoing electrons passing the VTLH word but requiring them to still pass the `Very Loose Likelihood` (VLLH) word) and releasing the electron isolation (both track and calorimeter).

The rest of the selections are the same as in the signal region and are listed again in Section 6.1.3. The expected number of $t\bar{t}$ background events in the control region from MC are presented in Table 10.

Also in the case of electrons, albeit not used in the normalization anywhere, it is interesting to validate the simulation against data in the CR, comparing important kinematic and topological variables. Such plots are presented in Figures 22 for global event quantities; and in Figure 23 for anti-tight electron quantities.

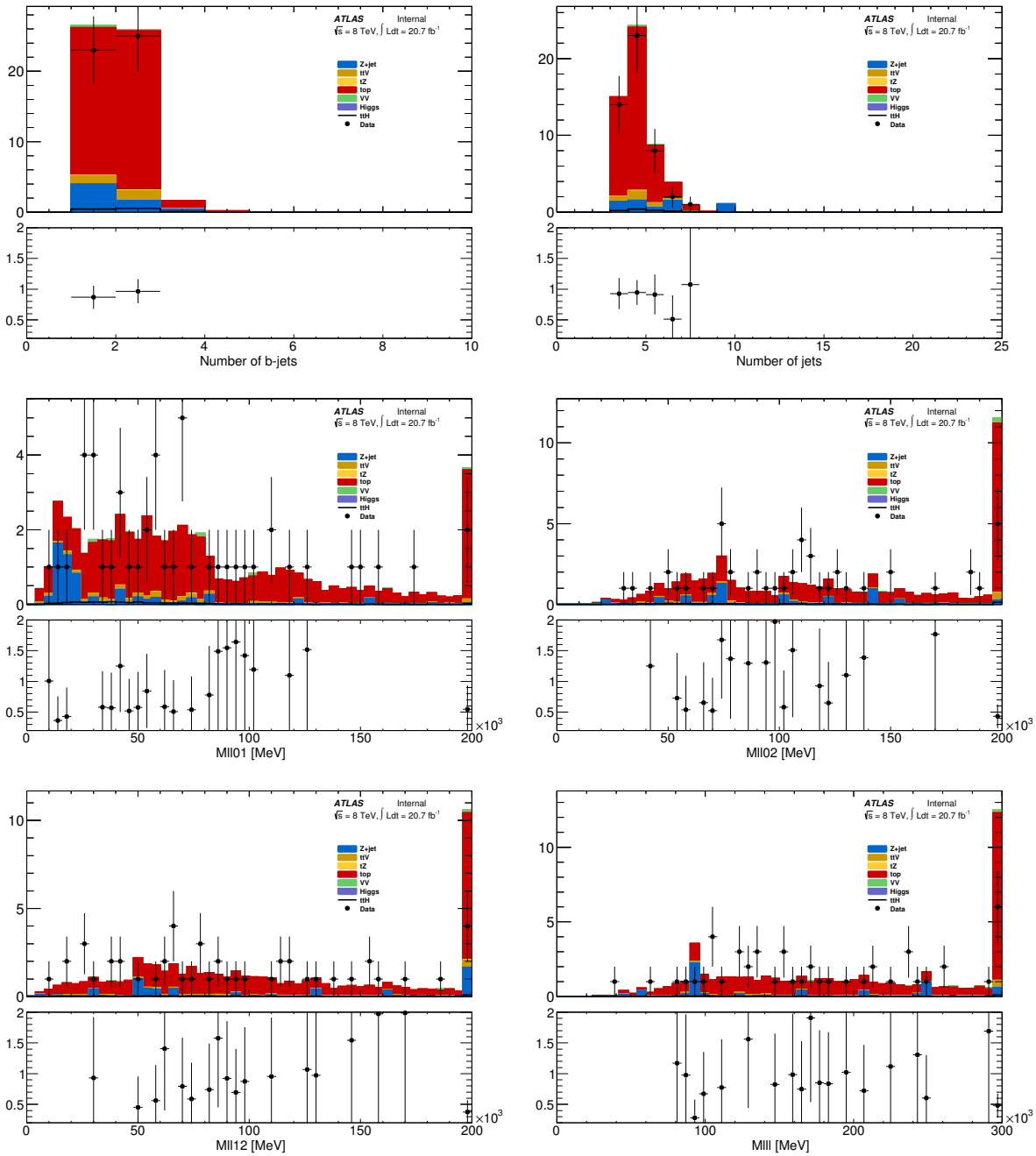


Figure 22: Top: Distributions of event quantities in the control region d1 (2 tight leptons + 1 anti-tight electron), in data (dots), compared with a stacked histogram of the various simulated samples: top in red, V+jets (blue), VV (purple) and ttV (yellow). The uncertainty on the data distribution is statistical. Bottom: the Data/MC ratio is presented for the same distribution. The quantities presented are, from top left to bottom right, N(b tags), N(selected jets), invariant mass of the two tight leptons, and of the two tight leptons with the anti-tight muon, and the invariant mass of the three leptons together.

Control Region	Sample	N(events) in CR
CR	Data	48 ± 5.29
	$t\bar{t}$	43.48 ± 0.77
	single top	1.22 ± 0.30
	other processes	9.47 ± 0.60
	data-other processes (”dt”)	38.55 ± 5.29
	$t\bar{t}$ + single top (”mc”)	44.7 ± 1.30
ratio	$\frac{dt}{mc}$	0.862 ± 0.12

Table 10: Expected number of $t\bar{t}$ background , single top other MC events; and in data for the anti-tight control region (CR)

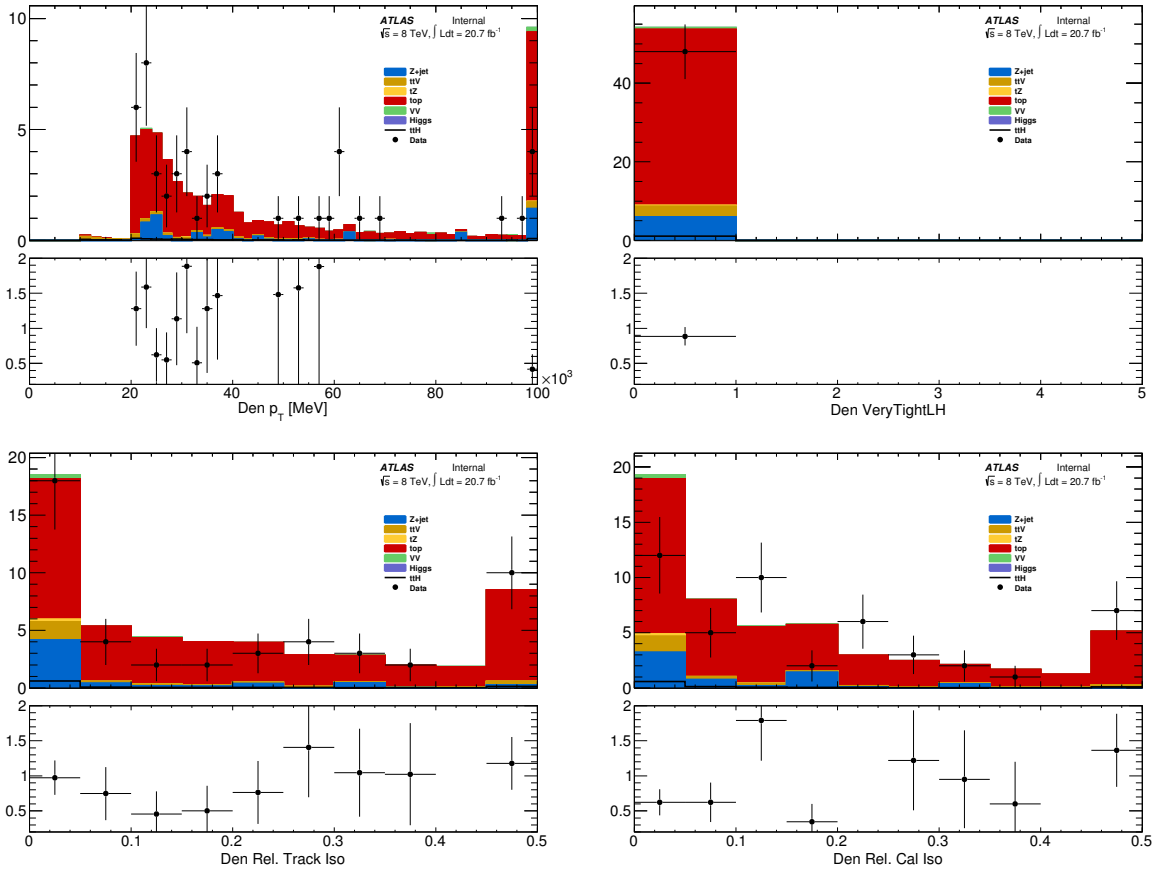


Figure 23: Top: Distributions of anti-tight electron quantities in the control region (2 tight leptons + 1 anti-tight electron), in data (dots), compared with a stacked histogram of the various simulated samples: top in red, V+jets (blue), VV (purple) and ttV (yellow). The uncertainty on the data distribution is statistical. Bottom: the Data/MC ratio is presented for the same distribution. The quantities presented are, from top left to bottom right, p_T, η, `Very Tight Likelihood word` value, ptcone20/p_T, Etcone20/p_T

Transfer factor into the signal region As with the muon fake estimate the θ_e transfer factor is directly extracted from MC from the ratio of events in the SR to the CR. In order to ensure that we are evaluating the transfer factor for electron fakes only, we use only events with the flavor structure shown in Equation 6. "d" refers to anti-tight electrons and "x" refers to either a tight electron or muon.

$$\theta_e = \frac{N(xee \text{ events})}{N(xed \text{ events})} \tag{6}$$

$$\theta_e = 0.0240 \pm 0.005(stat) \tag{7}$$

Similar to the muons we define an auxiliary region to test the data-MC agreement of the distributions (particle id and isolation) used in the θ_e extrapolation. The auxiliary region is defined in the same way for muons. The auxiliary region contains parts of the SR but is enlarged with lower NJet cuts to allow for a large statistics comparison. The θ_e factors from MC for the auxiliary cuts and the actual SR/CR cuts are well within statistical error, which suggests that the region is adequately similar for the data-MC agreement to be useful.

With the anti-tight electron selection (release isolation and fail VTLH), we are able to look at the data-MC agreement of the full isolation distributions (only for fail VTLH). In order to see the data-MC of VTLH, we define a second anti-tight region, where we require isolation to fail (Etcone20/p_{T}> 0.05 and ptcone20/p_{T}> 0.05) and release the VTLH idword. In the following, we refer to the original anti-tight selection as d1 and the new anti-tight selection as d2.

The yields in the auxiliary region are summarized in Table 11.

Sample	Yield in auxiliary region for d1	Yield in auxiliary region for d2
Data	172	134
Total MC	157.93 ±3.19	112.72 ± 2.17
$t\bar{t}$	133.39 ± 1.44 ± 1.71	102.38 ± 1.26
Z+jets	17.46 ± 2.81	7.58 ± 1.75
Others(incl. signal)	8.32 ± 0.34	3.34 ± 0.80

Table 11: Expected number of $t\bar{t}$ background , Z+jets and other MC events; and of collision data in the auxiliary region defined in the text; separate entries for the monitoring of the evolution of the isolation shape into the SR when the VTLH `word` is inverted (d1) and of the VTLH word when the isolation is inverted (d2) are presented. The yields are for events where the configuration (ℓ_0, ℓ_1, ℓ_2)=(ℓ_0, μ, e) is vetoed.

The contamination from processes other than $t\bar{t}$ is therefore expected to be of 11% (5%) for events in the auxiliary region corresponding to d1(to d2).

Data and MC are compared on various quantities, notably isolation and Identification `word`, to assess the accuracy of the description attained by the simulation. These are shown in Figures 24-25 for the auxiliary region corresponding to d1 and to d2 respectively.

The shapes of the variables relevant for the extrapolation (tracking isolation, calorimetric isolation and VTLH `word`), appear to be well modeled in MC. One must caution that the shapes of the isolation variables are plotted only when the particle id has failed and vice-versa. The assumption is that this agreement can be translated directly to the modeling of the isolation (particle id), when the electron passes the particle id (isolation). This requires that the correlation between the particle id and isolation variables is small and that residual correlation is well-modeled.

To encompass the level of data to MC agreement in these distributions in a systematic error on the transfer factor, we obtain pseudo-transfer factor (pseudo-θ_e) for data and MC in Equation 10, which

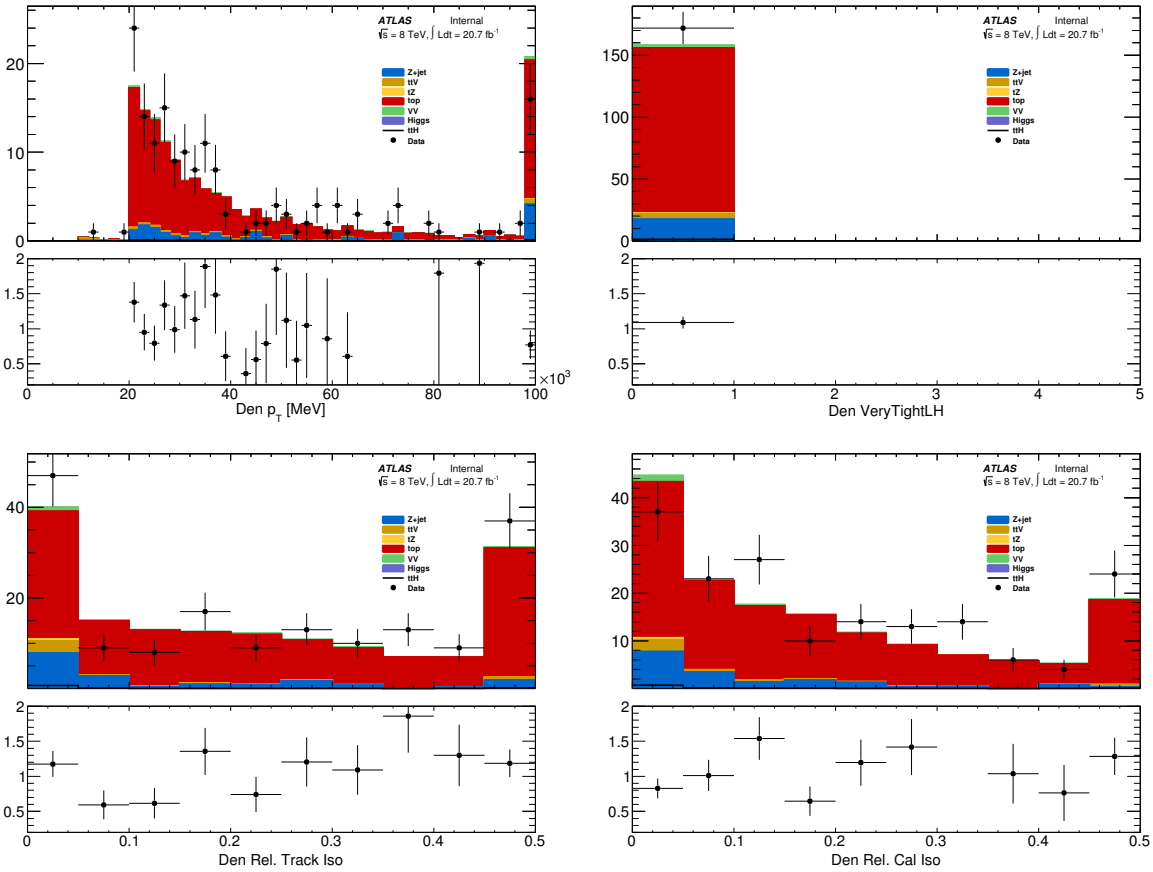


Figure 24: Top: Distributions of loose electron quantities in the auxiliary region corresponding to the CR d1. Data (dots) are compared with a stacked histogram of the various simulated samples: top in red, V+jets (blue), VV (purple) and ttV (yellow). The uncertainty on the data distribution is statistical. Bottom: the Data/MC ratio is presented for the same distribution. The quantities presented are, from top left to bottom right, p_T, η, `Very Tight Likelihood word` value, ptcone20/p_T, Etcone20/p_T

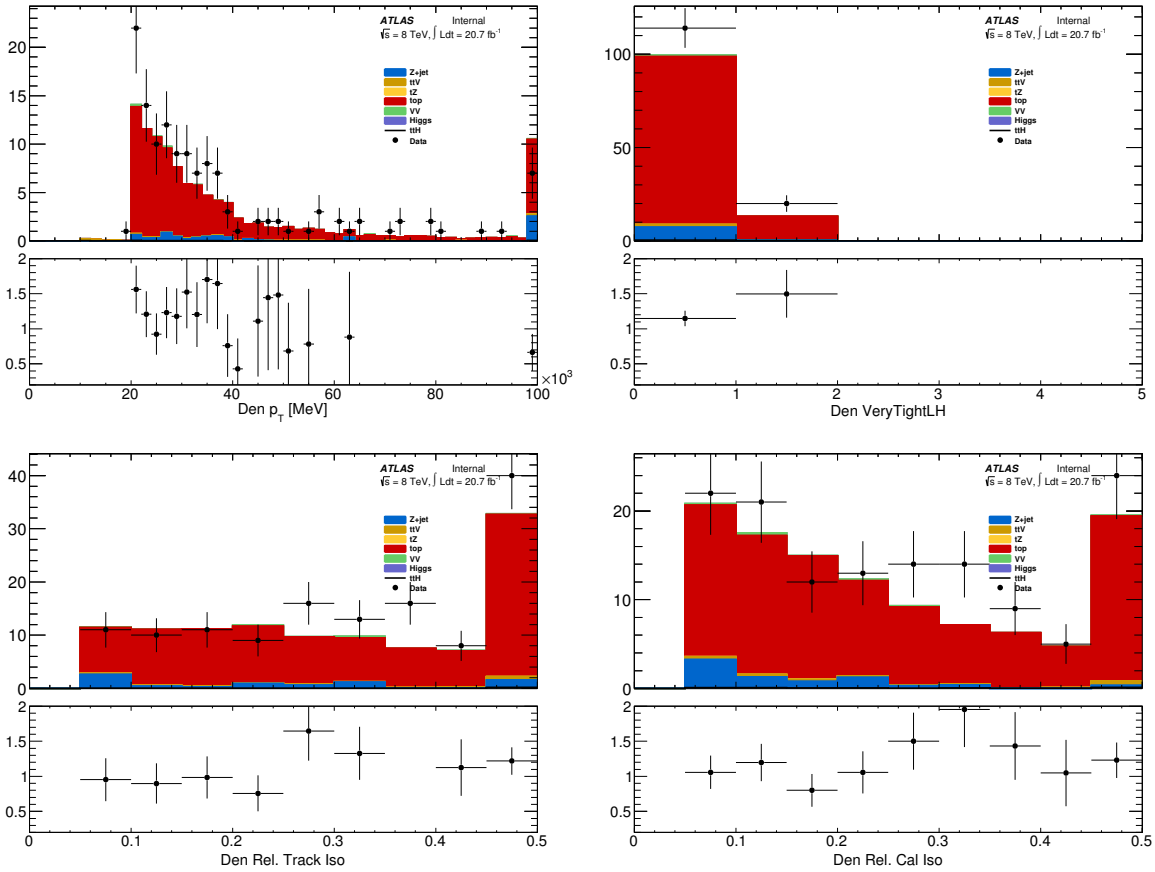


Figure 25: Top: Distributions of loose electron quantities in the auxiliary region corresponding to the CR d2. Data (dots) are compared with a stacked histogram of the various simulated samples: top in red, V+jets (blue), VV (purple) and ttV (yellow). The uncertainty on the data distribution is statistical. Bottom: the Data/MC ratio is presented for the same distribution. The quantities presented are, from top left to bottom right, p_T, η, `Very Tight Likelihood word` value, ptcone20/p_T, Etcone20/p_T

563 factorizes the extrapolation into two pieces, one for isolation and one for PID. If there is no correlation
564 between PID and iso, then pseudo-$\theta_e = \theta_e$.

$$f_{iso}(\text{pidfail}) = \frac{N_{passcut}}{N_{tot}} \tag{8}$$

565

$$f_{pid}(\text{isofail}) = \frac{N_{passcut}}{N_{failcut}} \tag{9}$$

$$\text{pseudo} - \theta_e = f_{iso}(\text{pidfail}) * f_{pid}(\text{isofail}) \tag{10}$$

566 The final values of pseudo-$\theta + e$ in MC and the corresponding values in data, with larger stat un-
567 certainty, are reported in Table 12. It is important to note two things. First, these are NOT to be used
568 as actual transfer factors, because they are obtained from distributions of the variables (particle id and
569 isolation) when other variables are reversed. Second, that the values of the pseudo-transfer factor ARE
570 indeed close to the real transfer factor suggests that the correlations between the two variable are small,
571 as was our assumption.

	pseudo-θ_e
Data	0.0265 ± 0.008
top MC	0.0219 ± 0.002
Data/MC	1.21 ± 0.36

Table 12: Values of the pseudo transfer factors for Data and top MC (single top and $t\bar{t}$ (181087)

572 The difference between data and MC is quoted as a systematic uncertainty (21%). This combines
573 with the statistical uncertainty from the MC to give an overall 29% uncertainty on the transfer factor, θ_e.

574 **Results in the signal region** The following Table 13 summarizes all the inputs for the final estimation
575 of the number of $t\bar{t}$ background events with a fake electron in the SR; and the definition of the source
sample/region.

Stage	Value (and rel.unc.)
Overall starting normalization (CR)	38.2± 6.9 (17%)
Transfer factor θ_e	$(2.40 \pm 0.65)\% \cdot 10^{-2}$ (29%)

Table 13: Summary of regions and inputs to the extraction of the number of $t\bar{t}$ background events with a fake electron in the SR

576

577 For the $N_{\text{data,e,SR}}$, the number of $t\bar{t}$ background events with an electron fake, we find:

$$N_{\text{data,e,SR}} \quad = \quad 0.91 \pm 0.29 \tag{11}$$

578 with an overall relative uncertainty of 34%.
579 The systematic uncertainties are the summed in quadrature errors resulting from the statistical error
580 of the control region (14% for ≥ 4 jets, 1 b-tag, 25% for 3 jets, ≥ 2-btag, the statistical error on the
581 transfer factor (21%), and the data-MC agreement in the extrapolating variables (21%).

6.1.5 Total number of fakes

Summing the numbers in the previous sections, and assuming they are uncorrelated, the following total number of $t\bar{t}$ background events in the SR is estimated:

$$N_{\text{data,SR}} = 2.62 \pm 0.51 \tag{12}$$

with an overall relative uncertainty of 20%.

The assumption that the two fake rates are uncorrelated and thus can simply be added linearly is motivated by the fact that the main uncertainties are related to closure tests and that the θ definition separates muon and electron results by construction. The compares well with the MC only estimate of 2.46 ± 0.19.

6.1.6 Cross-Check in Low Jet Data Region

To test the goodness of the method and the presence of possible biases, a validation on data is performed, whereby the estimate of top quark fakes is compared with the data yield in a top-quark enriched validation region.

To this end a validation region with 3 tight leptons, exactly 2 or 3 jets and exactly 1 b-tag is used. We call it simply "VR". This guarantees a large top quark contribution. The contamination from non-top-only processes (incl. ttV) is estimated from simulation and subtracted from the overall data yield. The difference is taken to be all originated from single and pair top production and compared with the estimation.

For the estimation, we use exactly the same procedure applied so far, but taking a new CR (called "VR-CR") defined with exactly 2 or 3 jets and exactly 1 b-tag; and a new auxiliary region for this jet and b-tag multiplicity. For the muon estimation: the number of *data* events of the *xxP* type in this VR-CR is 545.0 ± 23.3; the number of events from "other" processes from MC is 45.9 ± 1.3. Therefore, the resulting MC-subtracted data yield in VR-CR1 is 499.1 ± 23.5. For the electron estimation: the number of *data* events of the *xxd* type in this VR-CR is 176.0; the number of events from "other" processes from MC is 35.0 ± 5.2. Therefore, the resulting MC-subtracted data yield in VR-CR1 is 141.0 ± 13.2.

The value of θ is again extracted using *simulated* top quark events in the VR validation region over the new VR-CR1 following the same definition as in Equation 3. The corresponding values are: $\theta_\mu = 0.016 \pm 0.003$ and $\theta_e = 0.026 \pm 0.001$. Similar systematic uncertainties are assigned as in the estimate for the SR in both the muon and electron case.

We estimate 8.1 ± 1.7 events with a muon fake and 3.7 ± 1.5 events with an electron fake in this region. The total estimated fake events are, thus: 11.8 ± 2.3. This is to be compared with the background-subtracted data occurrence in VR of 9.8 ± 4.9 events . The $t\bar{t}$ + single top MC prediction in VR is 9.8 ± 0.4.

6.2 Rescaling of Z+jets events background

6.2.1 Strategy

The expected contribution of the Z+jets background in the 3 lepton SR is negligible ([1] and Table 33 in this note). With this in mind, for this channel a choice is made to simply evaluate the ratio of the Z+jets contribution in MC over data in a dedicated control region and use the same scaling factor to correct the sample contribution of this process at the beginning of the cutflows in the SR. The scaling factors are extracted in the context of the ttV analysis of the Top WG [12]. In that work a separate scaling factor for Z+HF ($b\bar{b}$ and $c\bar{c}$) than for Z+LF is obtained. The two are extracted from a region enhanced (at least 1 b-tag) and depleted (0 b-tags) in HF respectively. The scaling factors are obtained on a sample of opposite charge (OS) di-lepton events.

Table 14: Expected yields of $t\bar{t}$ + vector boson processes in the 3ℓ channel signal regions. The uncertainties given are due to Monte Carlo statistics and overall cross section, respectively.

Process	Yield
$t\bar{t}(Z/\gamma^*)$	$3.86 \pm 0.04 \pm 0.48$
$t\bar{t}W$	$2.49 \pm 0.11 \pm 0.25$
tZ	$0.55 \pm 0.04 \pm 0.03$

Z+HF and Z+LF are defined according to the Heavy Flavour Overlap Removal (HFOR) value of the event [13]. Events with HFOR = 0 or 1 are HF, while HFOR = 3 signal a LF event. The data and the MC yields are compared at the Z peak, that is to say for events with the dilepton invariant mass $M(2\ell) \in [81, 101]$, the same used for the Z veto in this analysis. Only same flavour events are considered. The multi-leg ALPGEN generator is used; and the events are showered with the HERWIG programme: this is the same setup as the one used in this analysis. It has been observed (e.g. [14]) that ALPGEN overestimates the $p_T(Z)$ at large values. A re-weighting of simulated events according to the transverse momentum of the di-lepton pair, $p_T(\ell\ell)$ is therefore also applied to match the data. The re-weighting factors are obtained on a sample with 0 b-tags and at least 2 jets; their consistency is also checked in a sample with at least 1 b-tag. The value of the normalization scale factors for Z+HF and Z+LF are 1.33 and 0.94 respectively.

The events are selected using the same object definitions and very close kinematic thresholds as in this analysis; the jet p_T threshold and b-tagging working point are the same. It follows that they can safely be applied to the 3 lepton channel in $t\bar{t}H$. The $p_T(\ell\ell)$ re-weighting and then the normalization rescaling are applied in sequence.

6.2.2 Control plots in 2 tight + 1 loose lepton region

The effect of the re-weighting and of the rescaling are checked in control regions where two tight leptons + 1 loose lepton (i.e. released isolation and muon-jet overlap removal cuts) with the same kinematic characteristics as the tight leptons, so to be as close as possible to the 3 lepton signal region and the original sample from where the scaling factors were obtained. The pair of tight leptons are used to compute the di-lepton quantities. The final effect of re-weighting the $p_T(\ell\ell)$ and re-scaling the normalization is shown in Figure 26 for a 1 b-tag sample with 1 or 2 selected jets, similar in flavour content to our SR.

6.3 Estimation of $t\bar{t}V$ and tZ events in the signal region

Production of top plus vector boson (including $t\bar{t}V$ and tZ) is the dominant background in the 3ℓ channel. In particular, the yield is somewhat larger than that for signal [1]. The $t\bar{t}(Z/\gamma^*) \rightarrow t\bar{t}\ell\ell$ process is the largest single background contributor in the 3l signal region overall, although $t\bar{t}W$ is slightly larger in the 3 jet exclusive, ≥ 2 b-tag region. We show expected yields and associated cross section systematic uncertainties with the nominal selection in Table 14.

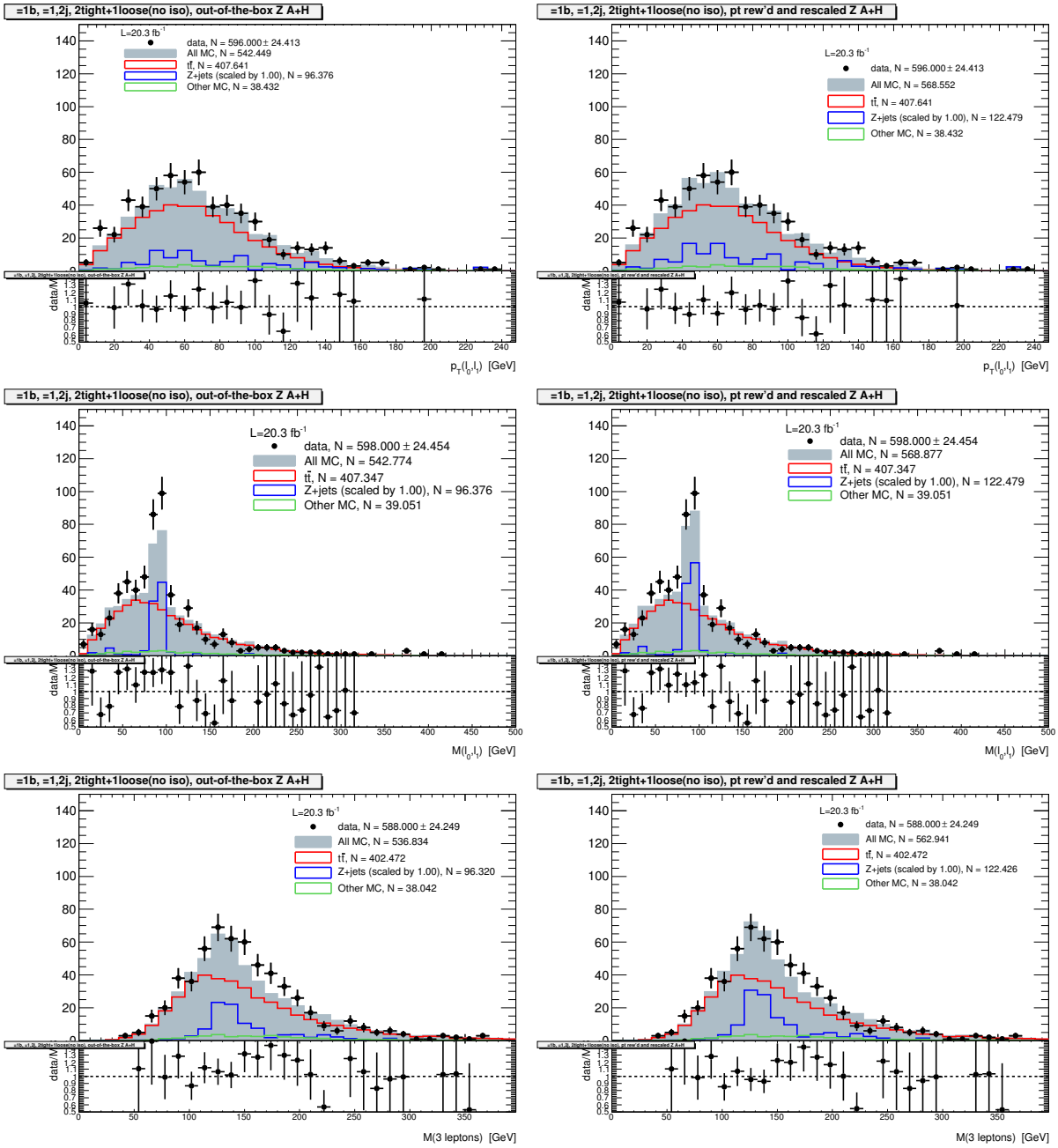


Figure 26: Z-related quantities in the 2 tight + 1 loose lepton sample, before (left) and after (right column) the application of the $p_T(\ell\ell)$ re-weighting and the normalization scale factors. The top plots show the transverse momentum of the di-tight-lepton system, $p_T(\ell_0\ell_1)$, the middle plots show their invariant mass $m(\ell_0\ell_1)$, while the bottom plots show the invariant mass of the 3-lepton system. Each plot has also the data/MC ratio displayed. The number of events is shown for each sample in the legend: data are black dots (+stat unc), $t\bar{t}$ background is red, Z+jets is black, while others are green.

7 2 lepton Same-Sign channel

7.1 Estimation of the mis-identification rate of the electron charge

In analyses where the final state of interest is defined by two leptons with the same electric charge, the background coming from Standard Model (SM) processes is very small. Therefore, it is crucial to consider the detector related backgrounds: objects which are mis-identified or mis-reconstructed such that they appear to have a same-sign dilepton final state.

Opposite-sign leptons from SM processes as Drell-Yan, W^+W^-, and mainly $t\bar{t}$, could contribute to the same-sign dilepton background if the charge of one of the leptons coming from the dileptonic decay of these processes is mis-measured. This process is called charge mis-identification. The contribution of this background to the same-sign dilepton signature is estimated by measuring the probability that the lepton charge is mis-reconstructed using a data-driven technique.

This section concentrates only on the measurement of the mis-identification rates of the electron charge[2]. There are two main sources of electron charge mis-identification:

- Hard Bremsstrahlung producing trident electrons ($e^{\pm} \rightarrow e^{\pm}\gamma^{*} \rightarrow e^{\pm}e^{+}e^{-}$) whose EM cluster is identified with the wrong electron's track, leading to a mis-identification of the charge. This source represents the main contribution to the background. The fraction of trident electrons depends on the amount of material that the electrons traverse. In the detector, the distribution of the material depends on $|\eta|$. Therefore, a strong dependence on $|\eta|$ is expected in the mis-identification rates.

- A slightly curved track that induces a measurement error. This effect is important at high transverse momentum. Thus, a small dependence on electron p_T is also expected in the mis-identification rates.

For muons this background is negligible compared to other background processes, and therefore is not considered in the analysis[3].

The rates are estimated in a $Z \rightarrow e^{+}e^{-} + jets$ sample for a specific electron selection. They are measured as a function of $|\eta|$ and p_T of the electron (to be more precise, the full $|\eta|$ and p_T ranges have been divided in regions, named bins, so that the rates are determined as a function of $|\eta|$ and p_T bins) in the low p_T region (< 140 GeV), and then extrapolated using a p_T dependent factor extracted from $t\bar{t}$ simulated events at higher p_T. These rates are then applied to scale signal events but having opposite-sign electrons, which provides the expected background contribution for the same-signal final state. In the low p_T region, the method is validated in Z simulated events by comparison with a truth-based algorithm. Comparison of the mis-identification rates obtained separately for electrons and positrons using the truth-matching method is also done.

The main strategy to estimate this background together with some preliminary concepts to compute the charge mis-identification rates are given in Section 7.1.1. The methods used to estimate these rates are described in Section 7.1.2. The data and simulated samples used in this study are listed in Note1 [1] as well as electron requirements. Event selection is presented in Section 7.1.4. This is followed by the measurement of these rates at $\sqrt{s} = 8$ TeV in Section 7.1.5. The closure test is presented in Section 7.1.6. The different systematic uncertainties considered are listed in Section 7.1.7. Overall, the summary and conclusions are presented in Section 7.1.8.

[2] In the description presented in this analysis positrons and electrons will both be called electrons. An exception is presented when the charge mis-identification rates are computed separately for both electrons and positrons.

[3] The rate of charge mis-identification for muons is only affected by the track curvature. Because of the long lever arm to the muon system and the fact that the charge is measured in both the inner detector and muon spectrometer the mis-identification rates of the muon charge are very low, making this background negligible compared to the other sources of background.

7.1.1 Main strategy and preliminary concepts

When a true opposite-sign event (for example $t\bar{t} \rightarrow bW^+\bar{b}W^- \rightarrow b\bar{b}e^+e^-\nu\bar{\nu}$) is produced, and assuming that ϵ is the rate of charge mis-identification for a single electron, there are three possibilities for this event to be reconstructed:

1. $e^+e^- + X$ without any charge mis-identification, with a probability of $(1-\epsilon)^2$,
2. $e^+e^- + X$ with the two electrons having a charge flip, with a probability of ϵ^2,
3. $e^\pm e^\pm + X$ when only one of the two electrons is mis-identified, with a probability of $2\epsilon(1-\epsilon)$.

Therefore, if there are N true opposite-sign events, the reconstructed events will be:

- $N^{os} = (1 - 2\epsilon + 2\epsilon^2)N$ opposite-sign events,
- $N^{ss} = 2\epsilon(1-\epsilon)N \simeq 2\epsilon N$ same-sign events,

where the last approximation for N^{ss} corresponds to the assumption that ϵ is very small.

Knowing this charge mis-identification rate ϵ, it is therefore possible to compute the estimated number of same-sign events N^{ss} from the measured number of opposite-sign events N^{os}, using the following expressions:

- $N^{ss} = \frac{\epsilon_i+\epsilon_j-2\epsilon_i\epsilon_j}{1-\epsilon_i-\epsilon_j+2\epsilon_i\epsilon_j}N^{os}$ for the ee channel,
- $N^{ss} = \frac{\epsilon}{1-\epsilon}N^{os}$ for the $e\mu$ channel,

where ϵ_i and ϵ_j are the charge mis-identification rates for the two different electrons.

Finally, the method to estimate the number of same-sign events produced by the Standard Model true opposite-sign processes has two steps:

1. the measurement of the charge mis-identification rate ϵ,
2. the application of the measured rates to measured opposite-sign events N^{os}.

The last one gives the estimation of N^{ss}.

7.1.2 Methods to estimate the electron charge mis-identification rates

Electrons coming from the leptonic decay of the Z boson are used to measure the charge mis-identification rates in data. In this study, a likelihood-based method has been used as a baseline to extract these rates. In simulated Z events, this method can be validated by comparison with a truth-based algorithm. Both methods are explained below.

Likelihood The likelihood method [15] assumes that the mis-identification rates of the electron charge are independent for different pseudorapidity regions. Therefore, the probability to have a number of same-sign events (N_{ss}^{ij}) with electrons in $|\eta|$ region i and j can be written as a function of the number of events N^{ij} as follows:

$$N_{ss}^{ij} = N^{ij}(\epsilon_i + \epsilon_j). \tag{13}$$

If all the same-sign events in the Z peak are produced by charge flip, then N_{ss}^{ij} is described by a Poisson distribution:

$$f(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \tag{14}$$

where k is the observed number of occurrences of the event, i.e. $k = N_{ss}^{ij}$, and λ is the expected number, i.e. $\lambda = N^{ij}(\epsilon_i + \epsilon_j)$. Thus, the probability for both electrons to produce a charge flip is expressed by:

$$P(\epsilon_i, \epsilon_j | N_{ss}^{ij}, N^{ij}) = \frac{[N^{ij}(\epsilon_i + \epsilon_j)]^{N_{ss}^{ij}} e^{-N^{ij}(\epsilon_i+\epsilon_j)}}{N_{ss}^{ij}!}. \tag{15}$$

The likelihood L for all the events is obtained by evaluating all the $|\eta|$ combinations:

$$L(\epsilon | N_{ss}, N) = \prod_{i,j} \frac{[N^{ij}(\epsilon_i + \epsilon_j)]^{N_{ss}^{ij}} e^{-N^{ij}(\epsilon_i+\epsilon_j)}}{N_{ss}^{ij}!}, \tag{16}$$

where the rates ϵ_i and ϵ_j can be obtained by minimizing the likelihood function. In this process, the $-\ln L$ is used in order to simplify and make easier the minimization. Terms which do not depend on the rates ϵ_i and ϵ_j are removed in this step. This way, the final function to minimize is given by the following expression:

$$-\ln L(\epsilon | N_{ss}, N) \approx \sum_{i,j} \ln[N^{ij}(\epsilon_i + \epsilon_j)] N_{ss}^{ij} - N^{ij}(\epsilon_i + \epsilon_j). \tag{17}$$

The events are selected within the Z peak and stored –with the electron order by $|\eta|$– in two triangular matrices: one for the same-sign events N_{ss}^{ij}, and the other one for all events N^{ij}. The likelihood method takes into account electron pairs with all $|\eta|$ combinations, which allows to use the full available statistics getting therefore lower statistical uncertainties. Moreover, it does not bias the kinematical properties of the electrons, compared to other methods like tag-and-probe.

The likelihood method can be easily extended to measure the charge mis-identification rates as a function of two parameters. In this study, the interest lies not only on the measurement of the rates as a function of the pseudorapidity, but also transverse momentum. Thus, the probability to find a same-sign event given the rates for each electron is $(\epsilon_{i,k} + \epsilon_{j,l})$, where the two indices represent binned $|\eta|$- and p_{T}-dependence. Thus, the Eq. 17 transforms into

$$-\ln L(\epsilon | N_{ss}, N) \approx \sum_{i,j,k,l} \ln[N^{ij,kl}(\epsilon_{i,k} + \epsilon_{j,l})] N_{ss}^{ij,kl} - N^{ij,kl}(\epsilon_{i,k} + \epsilon_{j,l}). \tag{18}$$

The likelihood method uses only Z *signal* events. Therefore, background coming from other processes where the dilepton invariant mass corresponds to the one of the Z boson needs to be subtracted. The background subtraction is done using a simple side-band method. This method consists in dividing the Z invariant mass in three regions, i.e. A, B and C, where B is the central region corresponding to the Z peak. The number of events is counted in the regions on the sides of the peak, i.e. n_A and n_C, and removed from the total number of events in the peak region B, n_B. This way, the number of signal events N_Z is given by

$$N_Z = n_B - \frac{n_A + n_C}{2}. \tag{19}$$

Once the background has been subtracted, the likelihood method can be applied. `MINUIT` is used for the minimization and `MIGRAD` to compute the uncertainty on these rates.

Truth-matching In simulated samples, the reconstructed electrons can be matched to the generated ones from the Z decay. A matching cone $\Delta R \leq 0.2$ is used to find reconstructed electrons that correspond to true electrons produced from the Z decay[4]. The matching is done only to the electrons produced directly from the Z decay (primary electrons). It does not match reconstructed electrons to electrons produced in the conversion of radiated photons from primary electrons, to electron-positron pairs, or

[4]It has been shown that for $\Delta R > 0.05$, the size of the cone has no incidence on the matching performance [15].

to electrons produced in other secondary processes. If trident electrons, the reconstructed electrons are matched to the primary electron if the electrons produced in the conversion are almost collinear to the primary electron.

7.1.3 Data and Monte Carlo samples

The data used correspond to the full 2012 ATLAS dataset taken in pp collisions at $\sqrt{s} = 8$ TeV. It corresponds to an integrated luminosity of 20.3 fb^{-1}. The simulated $Z+jets$ samples used for comparison are listed in Note (xxx ref to note1) and correspond to the inclusive ALPGEN samples. The $t\bar{t}$ simulated sample used to compute the extrapolation factor is the POWEG+PYTHIA sample.

7.1.4 Electron and event selection

Events are required to match the single electron trigger `EF_e24vhi_medium1`, which is `ORed` with `EF_e60_medium1` [16]. They are also required to have at least one primary vertex with at least 5 tracks, and to pass the Good Run List (data). Incomplete events or with noise bursts and data integrity errors in the LAr and Tile calorimeter are rejected. Bad jet rejection is also applied. Because the charge mis-identification rates are computed with $Z \rightarrow e^+e^- + jets$ events, exactly two electrons and no muons passing the selection described above are selected. At least one of the selected electrons must match the one that triggered the readout of the event.

Electrons are required to satisfy the same conditions as in given in [xxx ref to note1]. The study described in this section is based on `Very Tight Likelihood` electrons.

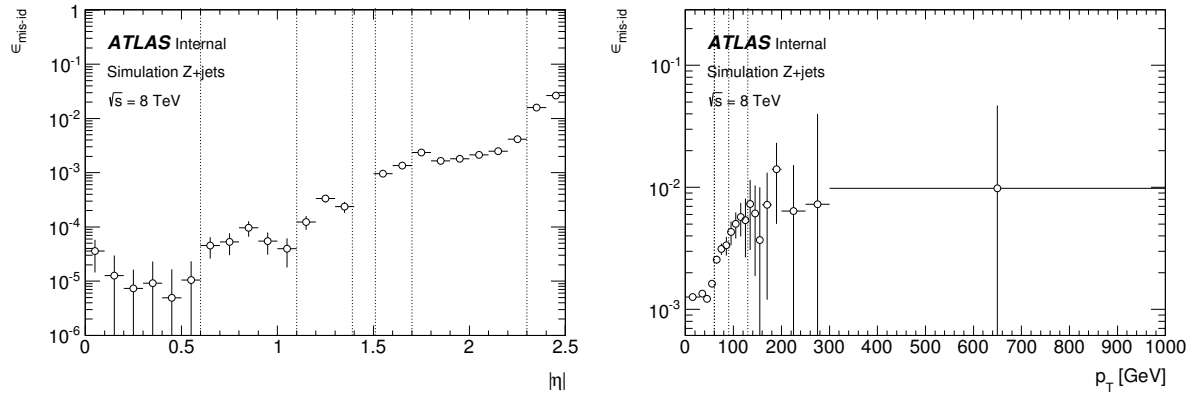
7.1.5 Estimation of the electron charge mis-identification rates at $\sqrt{s} = 8$ TeV as a function of $|\eta|$ and p_{T}

The procedure to estimate the charge mis-identification rates for electrons at $\sqrt{s} = 8$ TeV with the full 2012 ATLAS dataset has been divided in six different steps.

The first step consists in the determination of the $|\eta|$ and p_{T} binning to be used to extract the rates. This is followed by the determination of the regions (A, B and C) of the Z invariant mass to perform the background subtraction. In the analysis it is assumed that the charge mis-identification rates for electrons are the same as for positrons, therefore, a comparison between these rates is also done. This is followed by the validation of the likelihood method with the truth-based algorithm. After, the charge flip rates are computed in data using the two-dimensional likelihood method. Finally, because the previous rates are computed in the low p_{T} region (< 130 GeV), a p_{T} extrapolation factor is applied to these rates in order to take into account electrons with higher p_{T}. Each one of these points is developed in the following sections.

$|\eta|$ and p_{T} binning As stated before, the first step of this study consists in determining the $|\eta|$ and p_{T} binning to be used to extract the rates. Thus, the charge mis-identification rates are computed in Z simulated events with the truth-matching method for a very fine binning. In $|\eta|$, the bin width has been set to 0.1 as shown in Fig. 27(a). In p_{T}, it was set to 10 GeV until 160 GeV, and then increased until 300 GeV. Only one bin has been used above 300 GeV due to the low statistics in this region. The result is shown in Fig. 27(b). As can be seen, the rates are more or less constant in some regions. In the plot they are separated by the vertical lines. Therefore, seven different regions have been identified in $|\eta|$ and four in p_{T}. These regions have been used to define the $|\eta|$ and p_{T} binning shown in Tables 15 and 16, respectively.

In addition, the rates are smaller in the central region of the detector where the material budget is also smaller and the track resolution has the best performance (see Fig. 28). Once moving to the outer part of



(a) (b)

Figure 27: Charge mis-identification rates for electrons as a function of $|\eta|$ (a) and p_{T} (b) measured in simulated $Z \to e^+e^- + jets$ events using the truth-matching method for a very fine binning. The vertical lines separate the regions where the rates are more or less constant.

Bin	0	1	2	3	4	5	6
$\|\eta\|$ region	[0,0.6]	[0.6,1.1]	[1.1,1.37]	[1.37,1.52]	[1.52,1.7]	[1.7,2.3]	[2.3,2.47]

Table 15: $|\eta|$ binning defined using the mis-identification rates computed with the truth-matching method in simulated Z events for a very fine binning.

Bin	0	1	2	3
p_{T} region	[15,60]	[60,90]	[90,130]	[130,1000]

Table 16: p_{T} binning defined using the mis-identification rates computed with the truth-matching method in simulated Z events for a very fine binning.

the inner detector, the amount of material the electrons go through increases, raising the number of trident electrons, and therefore the mis-measurement of the charge. The mis-identification rates as a function of $|\eta|$ are related to the material distribution of the pixel in the inner detector. About the p_{T} dependence, as explained in the introduction it was expected to find that the mis-identification rates increase as long as the electron p_{T} increases, since at high p_{T} a slightly curved track can induce a measurement error.

Regions for background subtraction The likelihood method uses measured same-sign and opposite-sign Z signal events in order to compute the charge flip rates. As shown in Fig. 29, the Z-peak for same-sign electrons is shifted to lower values with respect to opposite-sign electrons. It could introduce in some way a mis-measurement of the rates. Therefore, the three different regions used by the side-band method to extract the background have been defined differently in the opposite-sign and same-sign distributions. The definitions of these regions are shown in Fig. 30. Their specific values are shown in Table 17. These values have been obtained by means of a Gaussian fit in a reduced region around the Z-peak for both the distributions of same-sign and opposite-sign invariant mass. They have been determined using the mean value κ and the standard deviation σ given by the fit. Thus, the region B was defined as $\kappa \pm 4\sigma$. The specific value 4σ is chosen since with this peak width, the closure test gives the

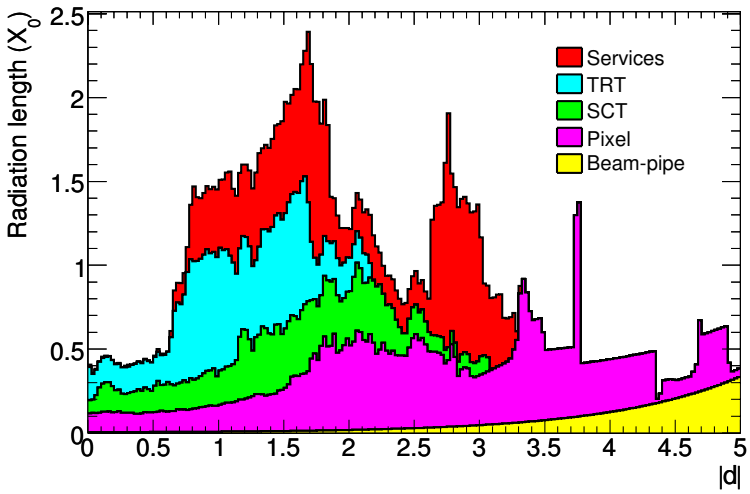


Figure 28: Material distribution at the exit of the inner detector as a function of $|\eta|$ [17].

best results. The regions A and C are defined so that they have the same width as the central region, i.e. 8σ. The variation of the rates due to the variation of the Z-peak region by 1σ definition will be taken into account as a systematic uncertainty.

Sample	A	B	C
Same-sign	[51.7,76.5]	[76.5,101.3]	[101.3,126.0]
Opposite-sign	[54.7,78.5]	[78.5,102.3]	[102.3,126.0]

Table 17: Ranges of the A, B and C to be used for the likelihood method to perform the background subtraction from the Z-peak.

Comparison of the charge mis-identification rates for electrons and positrons A crucial point of this study is to verify that the charge flip rates are the same for electrons as for positrons. These rates have been computed in simulated Z events using the truth-matching method as a function of ($|\eta|$,p_{T}). The ratio of these rates is shown in Fig. 31. As can be seen, in general there is a reasonable agreement within uncertainties. A systematic uncertainty will be assigned to these ($|\eta|$, p_{T}) bins where the ratio –taking into account the uncertainty– is not compatible with 1, as it will be shown later.

Validation of the likelihood method The likelihood method is used as a baseline to compute the mis-identification rates. Therefore, it needs to be validated. This is done by comparing the charge flip rates computed with both likelihood and truth-matching methods on simulated Z events. This comparison is shown in 1D as a function of $|\eta|$ and p_{T} in Figs. 32(a), and 32(b), respectively. As can be seen, the ratio of these rates (likelihood/truth-matching) shows that there is a good agreement, with the biggest difference being around 15%.

The two-dimensional version of the likelihood method was also validated in simulated Z events. Here, the likelihood function has 24 parameters (6×4 $|\eta|$-p_{T} bins) to be minimized. As there is not enough statistics in the high p_{T} region, it is not possible to find a minimization of the likelihood function for all these parameters. Therefore, the charge flip rates are measured in a p_{T} range where enough

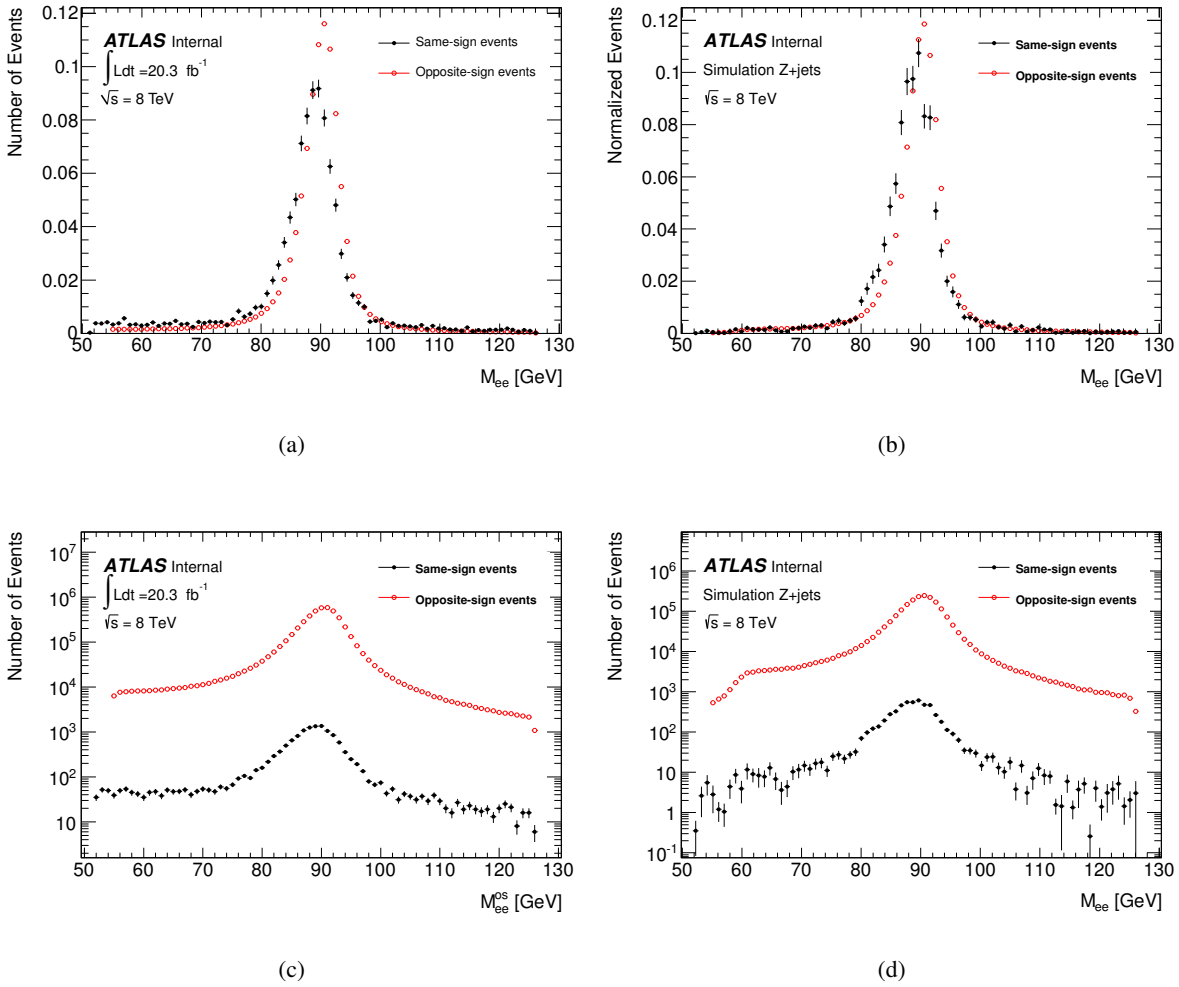


Figure 29: Comparison of the invariant mass distribution M_{ee} between same-sign and opposite-sign events around the Z-peak in data (a) and Monte Carlo (b). The distributions have been normalized. The Z-peak for same-sign electrons is shifted to lower values with respect to opposite-sign electrons because a larger momentum fraction has been radiated for the same-sign case, due to trident electrons. The same distributions without normalization are also shown for data (c) and Monte Carlo (d).

statistics is available –called here the low p_{T} region–, and defined to be the upper limit of the third p_{T} bin, i.e. below 130 GeV. The comparison with the truth-based algorithm is shown in Fig. 33, where the ratios of the charge flip rates are presented as a function of $|\eta|$ and parametrized in p_{T}. A good agreement within uncertainties can be observed. Differences between the rates extracted from both likelihood and truth-matching methods will be taken as a source of systematic uncertainty.

Estimation of the charge flip rates in data in the low p_{T} region Once the likelihood method has been validated, the charge flip rates in data can be computed. As for simulated samples, they are measured in $Z + jets$ events. The full 2012 ATLAS dataset taken at $\sqrt{s} = 8$ TeV has been used to estimate these rates. Again, there is not enough statistics in the high p_{T} region as to find a minimization of the likelihood function, and therefore to give an accurate measurement of the rates. As for Monte Carlo, the mis-identification rates are computed in a region where enough statistics is available, i.e. below 130 GeV.

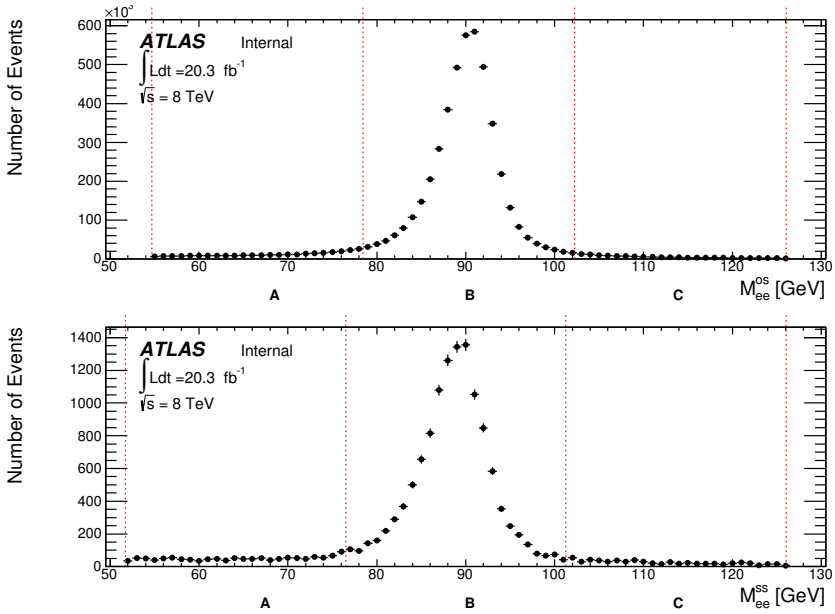


Figure 30: Dilepton mass spectra for the same-sign (top) and opposite-sign (bottom) electron pairs in data. The vertical lines represent the boundaries of the regions used for the background rejection.

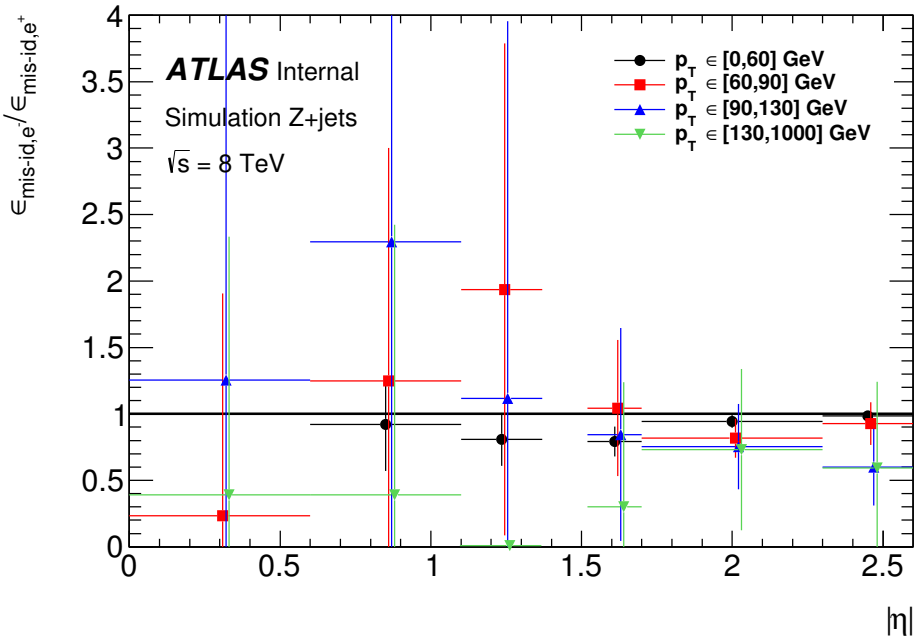
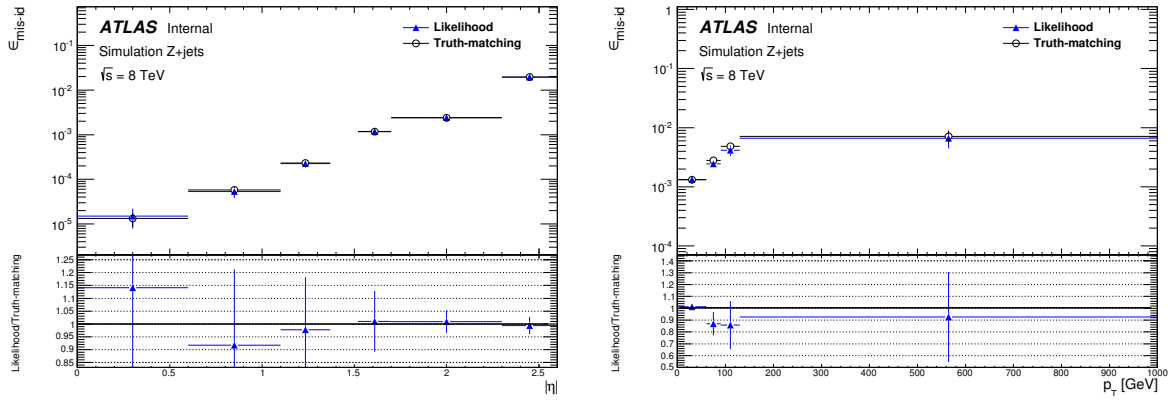


Figure 31: Ratio of the charge mis-identification rates for electrons and positrons as a function of $|\eta|$ and parametrized in p_{T} measured in simulated $Z \rightarrow e^{+}e^{-} + jets$ events using the truth-matching method. Big uncertainties are obtained due to the low statistics of the sample.

The result of this minimization can be seen in Fig. 34, where the charge flip rates are presented as a function of $|\eta|$ and parametrized in p_{T}. A comparison of these rates with the ones extracted from Monte Carlo using also the likelihood method is shown in Fig. 35. As can be seen, the agreement is reasonable within uncertainties.



(a) (b)

Figure 32: Charge mis-identification rates for electrons as a function of $|\eta|$ (a) and p_{T} (b) measured in simulated $Z \to e^+e^- + jets$ events with the truth-matching and likelihood methods.

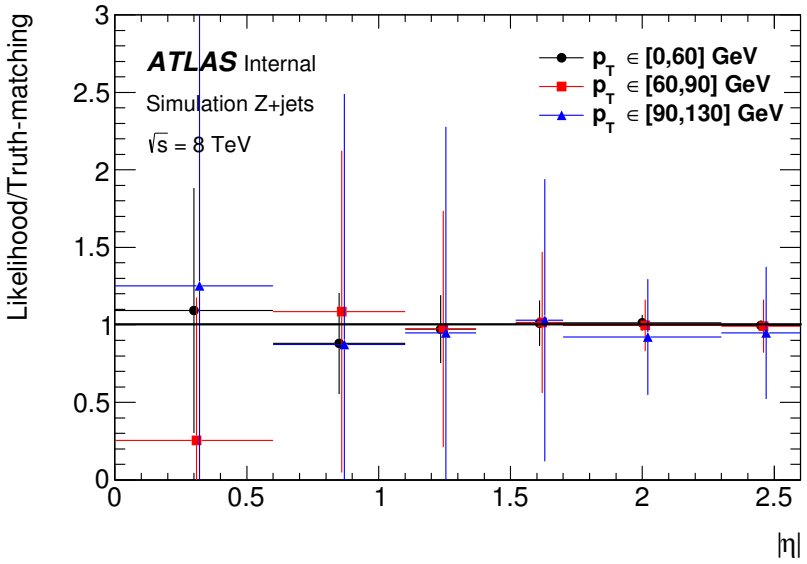


Figure 33: Ratio of the charge mis-identification rates for electrons as a function of ($|\eta|$, p_{T}) measured in simulated $Z \to e^+e^- + jets$ events with the truth-matching and the two-dimensional likelihood methods. Big uncertainties are obtained due to the low statistics of the sample.

Estimation of the charge flip rates in the high p_{T} region In the previous section, the mis-identification rates of the electron charge have been computed for Z events only in the low p_{T} region. Therefore, a new method has been developed in order to estimate the rates in the region with high p_{T}. This method consists basically in extrapolating the charge flip rates measured in Z events in the third p_{T} bin of the low p_{T} region, i.e. the rates where $p_{\mathrm{T}} \in [90, 130]$ GeV, with a p_{T} dependent factor extracted from simulated $t\bar{t}$ events. Thus, the charge flip rates in the high p_{T} region are defined as

$$\epsilon(|\eta|, p_{\mathrm{T}} > 130 \text{ GeV})_Z = \epsilon(|\eta|, p_{\mathrm{T}} \in [90, 130] \text{ GeV})_Z \times \alpha_{t\bar{t}}(|\eta|, p_{\mathrm{T}} > 130 \text{ GeV}), \quad (20)$$

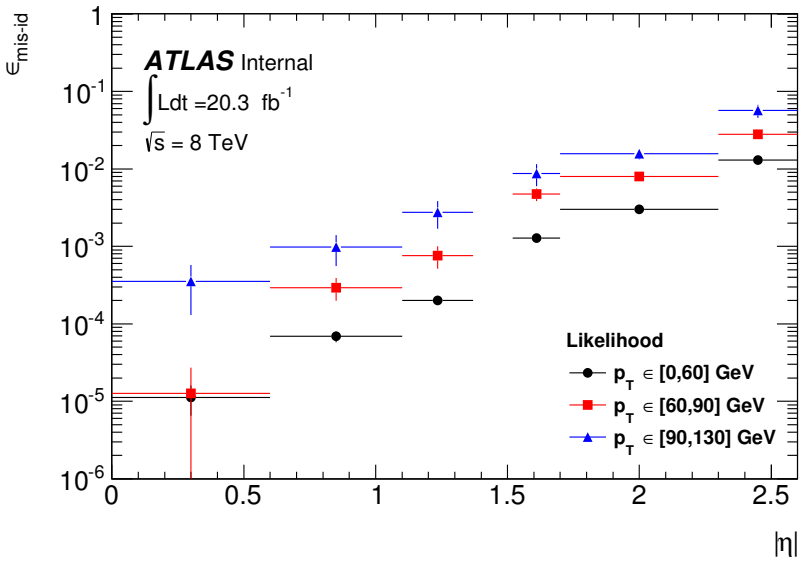


Figure 34: Electron charge mis-identification rates computed in data with the likelihood method in the low p_{T} region. The full 2012 dataset has been used to estimate the rates.

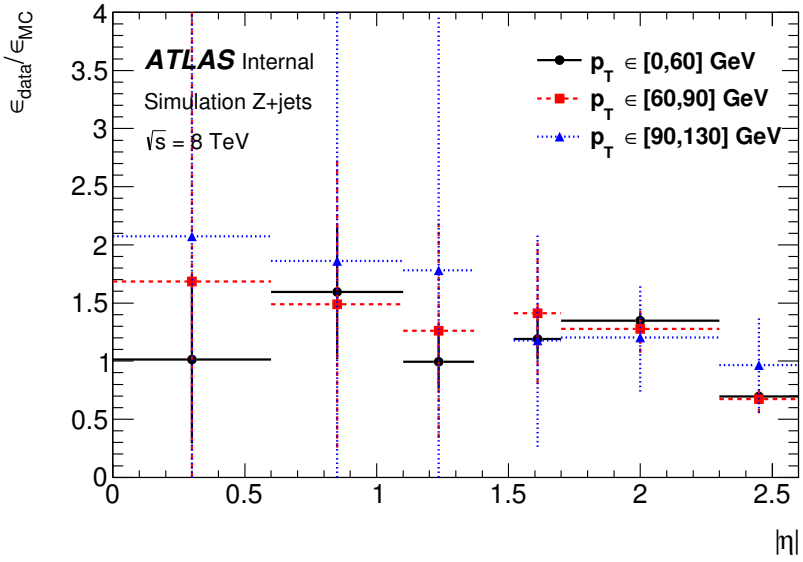


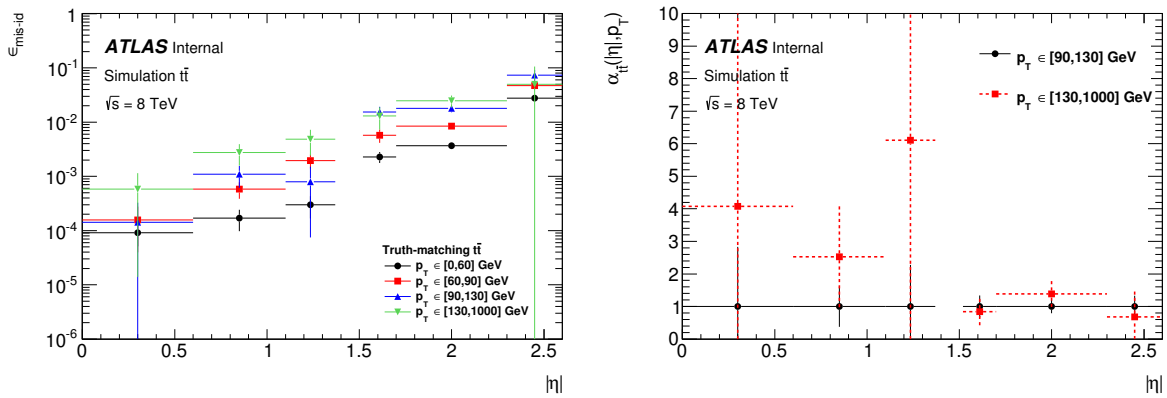
Figure 35: Comparison of the electron charge mis-identification rates computed in data and Monte Carlo with the likelihood method in the low p_{T} region.

with $\alpha_{t\bar{t}}$ being the p_{T} dependent correction factor extracted from $t\bar{t}$ and defined as

$$\alpha_{t\bar{t}}(|\eta|, p_{\mathrm{T}}) = \frac{\epsilon(|\eta|, p_{\mathrm{T}})_{t\bar{t}}}{\epsilon(|\eta|, p_{\mathrm{T}} \in [90, 130]\ \mathrm{GeV})_{t\bar{t}}}, \tag{21}$$

where $\epsilon(|\eta|, p_{\mathrm{T}})_{t\bar{t}}$ are the charge flip rates measured in $t\bar{t}$ events using the truth-matching method. These rates are shown in Fig. 36(a). The p_{T} dependent correction factor $\alpha_{t\bar{t}}$ is set to 1 in the low p_{T} region, i.e. below 130 GeV, and it is computed only for $p_{\mathrm{T}} > 130$ GeV; it is shown in Fig. 36(b). As can be seen, there is almost no dependence on $|\eta|$ at high $|\eta|$, while this is stronger in the low $|\eta|$ region. Finally,

857 the charge flip rates after correction for data are shown in Fig. 37, where both statistical and systematic
858 uncertainties are shown.



(a) (b)

Figure 36: a) Charge mis-identification rates for electrons as a function of $|\eta|$-p_{T} extracted from simulated $t\bar{t}$ events using the truth-matching method. b) p_{T} dependent factor extracted from $t\bar{t}$ Monte Carlo used to extrapolate the charge flip rates measured in Z events in the low p_{T} region. This extrapolation factor is set to 1 in the region where the electron p_{T} is below than 130 GeV.

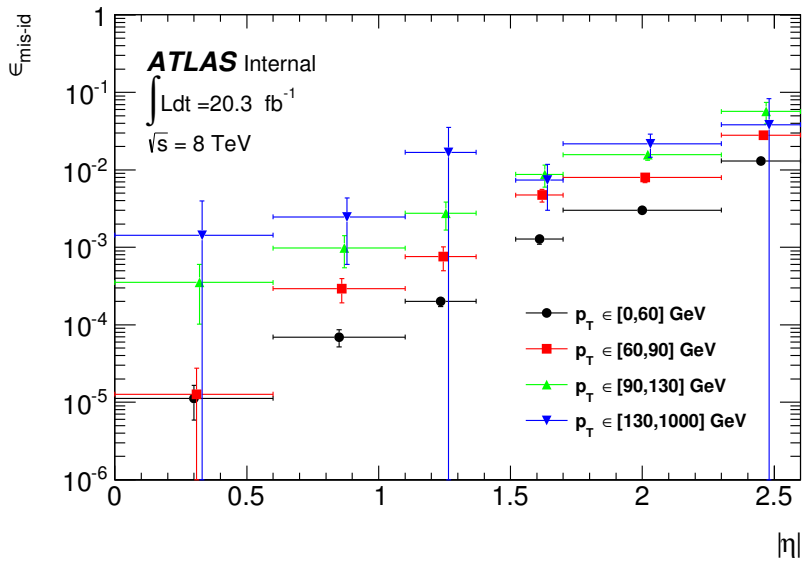
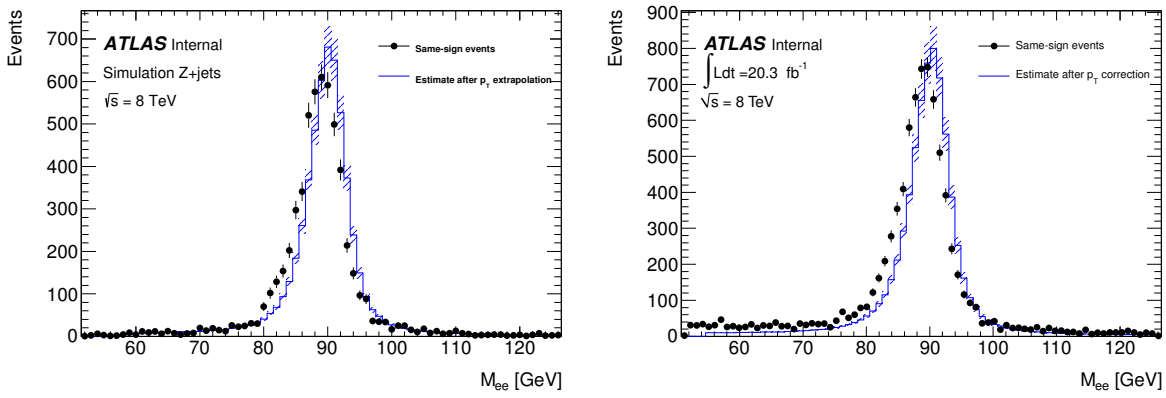


Figure 37: Electron charge mis-identification rates measured in data with the likelihood method on Z events (black points, red squares and blue triangles) as a function of $|\eta|$ and parametrized in p_{T}. The full 2012 dataset has been used to estimate the rates below 130 GeV. Above this value, the charge flip rates have been estimated by extrapolating the rates in the region where the $p_{\mathrm{T}} \in [90, 130]$ GeV with a p_{T} dependent factor extracted from simulated $t\bar{t}$ events (green triangles). Statistical and systematic uncertainties have been included in this plot.

7.1.6 Closure test

The rates after correction are finally validated by comparing the number of *measured* same-sign events with the *estimated* number of same-sign events. These last ones are computed by reweighting the measured number of opposite-sign events. This comparison was done in both, data and simulated $Z \rightarrow e^+e^- + jets$ samples. The invariant mass of the Z can be seen on Figure 38. In the simulated Z samples, the number of same-sign Z events is 5 049 while the estimation is $5\ 031^{+375}_{-365}$. In data, the number of same-sign Z events is 11 111 for an estimate of $11\ 094^{+882}_{-857}$. The uncertainties combine both statistical systematic uncertainties. In both cases, the closure test gives compatible results within uncertainties.



(a) (b)

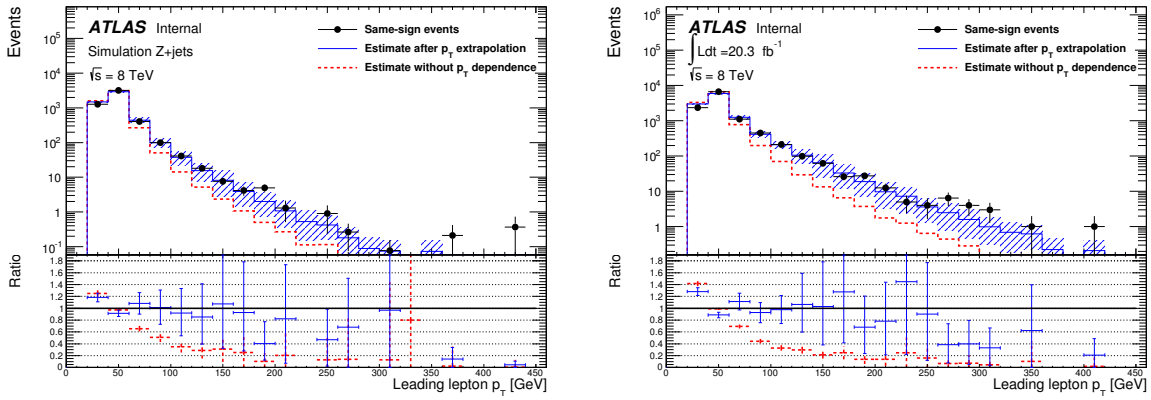
Figure 38: Closure test on simulated $Z \rightarrow e^+e^- + jets$ events (a) and data (b). The black circles show the distribution of same-sign events while the blue histograms show the distribution of the reweighted opposite-sign events together with the statistical and systematic uncertainties. The distributions are not expected to overlay exactly, due to the loss of energy of the trident electrons for the same-sign peak.

The effect of the p_{T} extrapolation factor can be investigated by comparing the distribution of the leading electron p_{T} of *measured* same-sign events with the *estimated* same-sign events using both, the charge flip rates after the p_{T} extrapolation, and the ones without taking into account the dependence on p_{T} (likelihood in 1D). This is shown in Fig. 39. In the high p_{T} region, i.e. above 130 GeV, the number of same-sign events in data is 253, while the estimation is 250^{+159}_{-136} for the two dimensional version of the likelihood method after the p_{T} extrapolation, and 16 ± 1 for the likelihood in 1D. In Monte Carlo, the number of same-sign events is 38, while the estimation is 32^{+25}_{-19} and 10 ± 1, for the two dimensional version of the likelihood method after the p_{T} extrapolation and for the likelihood in 1D, respectively. As can be seen, for both data and Monte Carlo, the estimate given by the rates with the p_{T} extrapolation is much better than the one given if the p_{T} dependence on the rates is not taken into account.

7.1.7 Systematic uncertainties

Different sources of systematic uncertainties have been taken into account in the charge mis-identication rates. The final uncertainty on the rates is the quadratic sum of five different contributions. They are enumerated below:

- The statistical uncertainty from the likelihood method $\sigma_{\epsilon}^{\mathrm{likelihood}}(|\eta|, p_{\mathrm{T}}) \times \alpha_{t\bar{t}}(|\eta|, p_{\mathrm{T}})$. It can be appreciated on Fig. 34.



(a) (b)

Figure 39: Distribution of the leading lepton p_{T} for the same-sign events (points) and reweighted opposite-sign events (histogram), in simulated $Z \rightarrow e^{+}e^{-} + jets$ events (a) and data (b). The uncertainty band includes statistical and systematic uncertainties. The dashed red line shows the estimate while using the weights without any dependence on p_{T}.

- The statistical uncertainty on the p_{T} dependent extrapolation factor $\epsilon(|\eta|, p_{\mathrm{T}})_{Z}^{\mathrm{likelihood}} \times \sigma_{\alpha_{t\bar{t}}}(|\eta|, p_{\mathrm{T}})$. This uncertainty only affects the region where the electron $p_{\mathrm{T}} > 130$ GeV. The value of $\sigma_{\alpha_{t\bar{t}}}(|\eta|, p_{\mathrm{T}})$ can be seen in Fig. 36(b).

- The difference between the rates measured with the two dimensional version of likelihood method and truth-matching on simulated $Z \rightarrow e^{+}e^{-}$ events. For a lepton $p_{\mathrm{T}} > 130$ GeV, the charge flip rates from the likelihood method after applying the p_{T} dependent extrapolation are used to compute this difference. It can be seen in Fig. 40. This uncertainty is taken into account only in the $(|\eta|, p_{\mathrm{T}})$ configurations where the ratio on these rates is not compatible with 1 within uncertainties.

- The half of the difference between true rates measured for electrons and positrons. This uncertainty is taken into account only in the $(|\eta|, p_{\mathrm{T}})$ configurations where the ratio on these rates is not compatible with 1 within uncertainties as shown in Fig. 31.

- The variation of the rates due to the variation of the Z-peak region definition when the background is extracted. In this case, the different regions have been variated by 1 σ. This uncertainty can vary from 0.5% to 33% the value of the charge flip rate depending on the $(|\eta|,p_{\mathrm{T}})$ region. Big uncertainties are obtained at low $|\eta|$ where the statistics is vey low.

The contribution of each one of these factors to the final uncertainty for the different $(|\eta|, p_{\mathrm{T}})$ configurations is shown in Fig. 41. The main contribution comes from the statistical uncertainty from the likelihood method and from the p_T extrapolation factor. An important source comes the difference between the measured and true rates on simulated Z events. The systematic uncertainty coming from the variation of the rates due to the variation of the Z-peak region definition becomes important a low $|\eta|$ and at low p_T. The systematic associated to the difference on the rates for electrons and positrons is negligible.

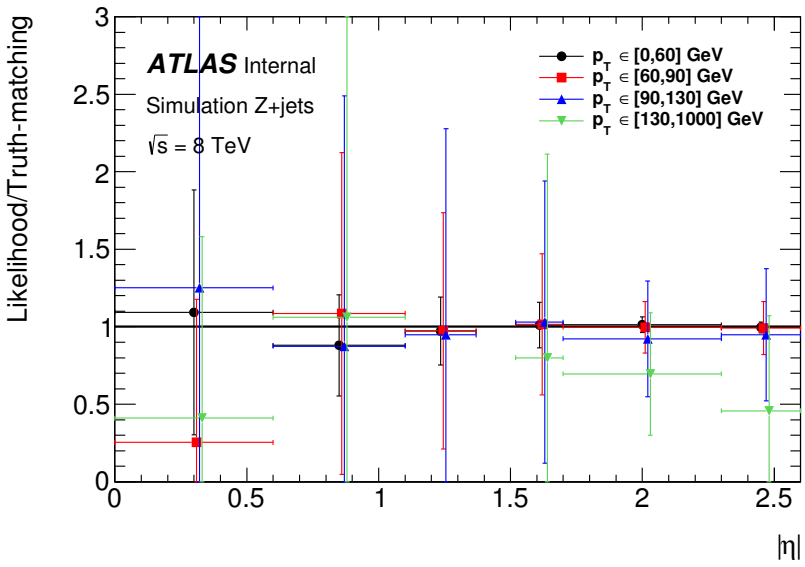


Figure 40: Ratio of the charge mis-identification rates for electrons as a function of ($|\eta|$, p_{T}) measured in simulated $Z \rightarrow e^{+}e^{-} + jets$ events with the truth-matching and the two-dimensional likelihood methods.

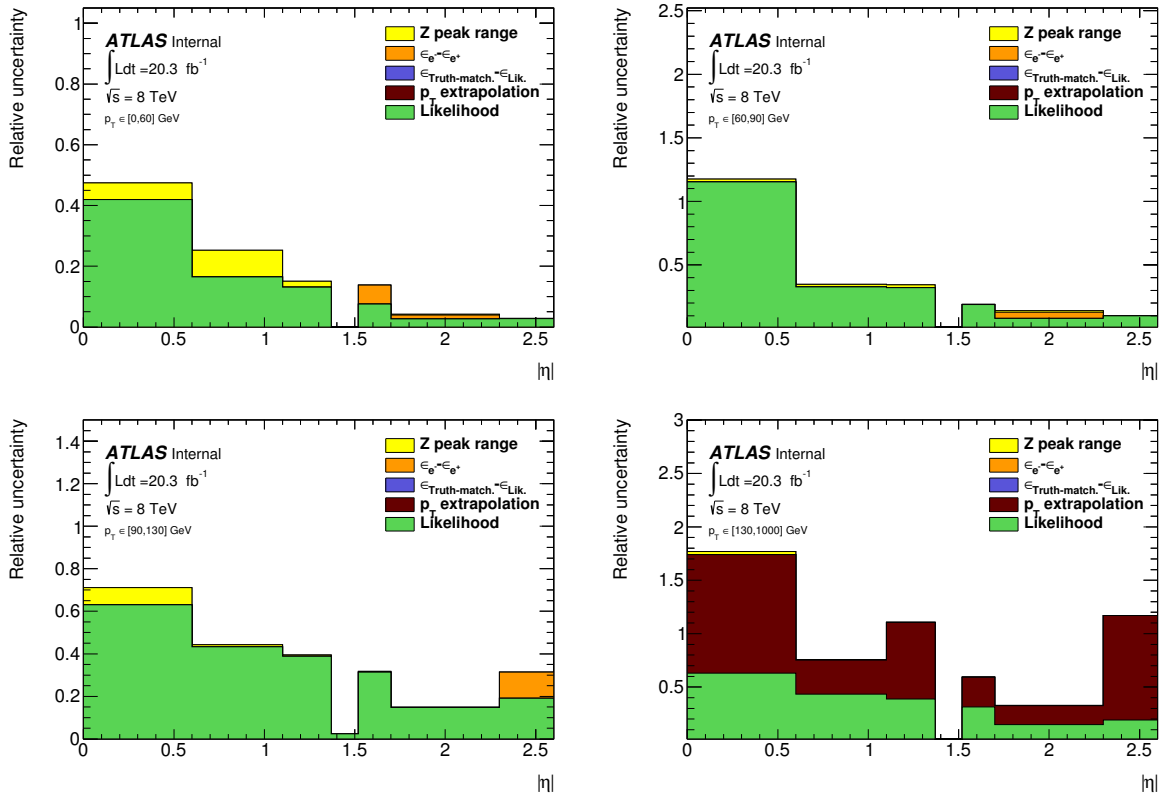


Figure 41: Relative systematic uncertainty contributions on the charge mis-identification rate, for different bins in p_{T} and $|\eta|$. Tight++ electrons have been used to produce this plot.

7.1.8 Summary and results

In this section has been presented the measurement of the mis-identification rates of the electron charge at $\sqrt{s} = 8$ TeV with the full 2012 ATLAS dataset. The two dimensional version of the likelihood method

has been used as a baseline to compute the rates for electrons with $p_{\mathrm{T}} < 130$ GeV. In this region, the likelihood method has been validated by comparison with a true based algorithm showing a reasonable agreement within uncertainties. Differences between these rates are taken into account as a systematic uncertainty. For the region where the lepton $p_{\mathrm{T}} > 130$ GeV, a new method has been developed in order to estimate the charge flip rates. It consists in applying a p_{T} dependent extrapolation factor extracted from $t\bar{t}$ simulated events. Different $t\bar{t}$ samples have been used to compute the correction factor, and a systematic uncertainty is assigned if the ratio of these extrapolation factors is not compatible with 1 within uncertainties. In addition, it has been checked that the charge flip rates estimated for electrons with the truth-matching method agree with the ones for positrons. The half of the difference between these rates has been considered as a systematic uncertainty if their ratio is not compatible with 1 within uncertainties. The variation of the rates due to the variation of the Z-peak region definition, the statistical uncertainty estimated from the likelihood method and the statistical uncertainty on the p_{T} dependent extrapolation factor, are also considered as sources of systematics. Closure tests done in both data and Monte Carlo show good results.

Applying these rates to opposite-sign events in the signal region (at least 5 jets, at least 1 b-tagged jet), the expected number of same-sign events with a charge mis-identification were estimated and are given in Table 33.

The predicted shapes for this background category are shown on Figures 42 and 43 in the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ channels respectively. The pseudorapidy shape of the electrons in the background with misidentified charge indicate a clear excess at $|\eta| > 1.5$ which is also described by the individual electron rates. An additional cut on the lepton pseudo-rapidity will be introduced in the signal region.

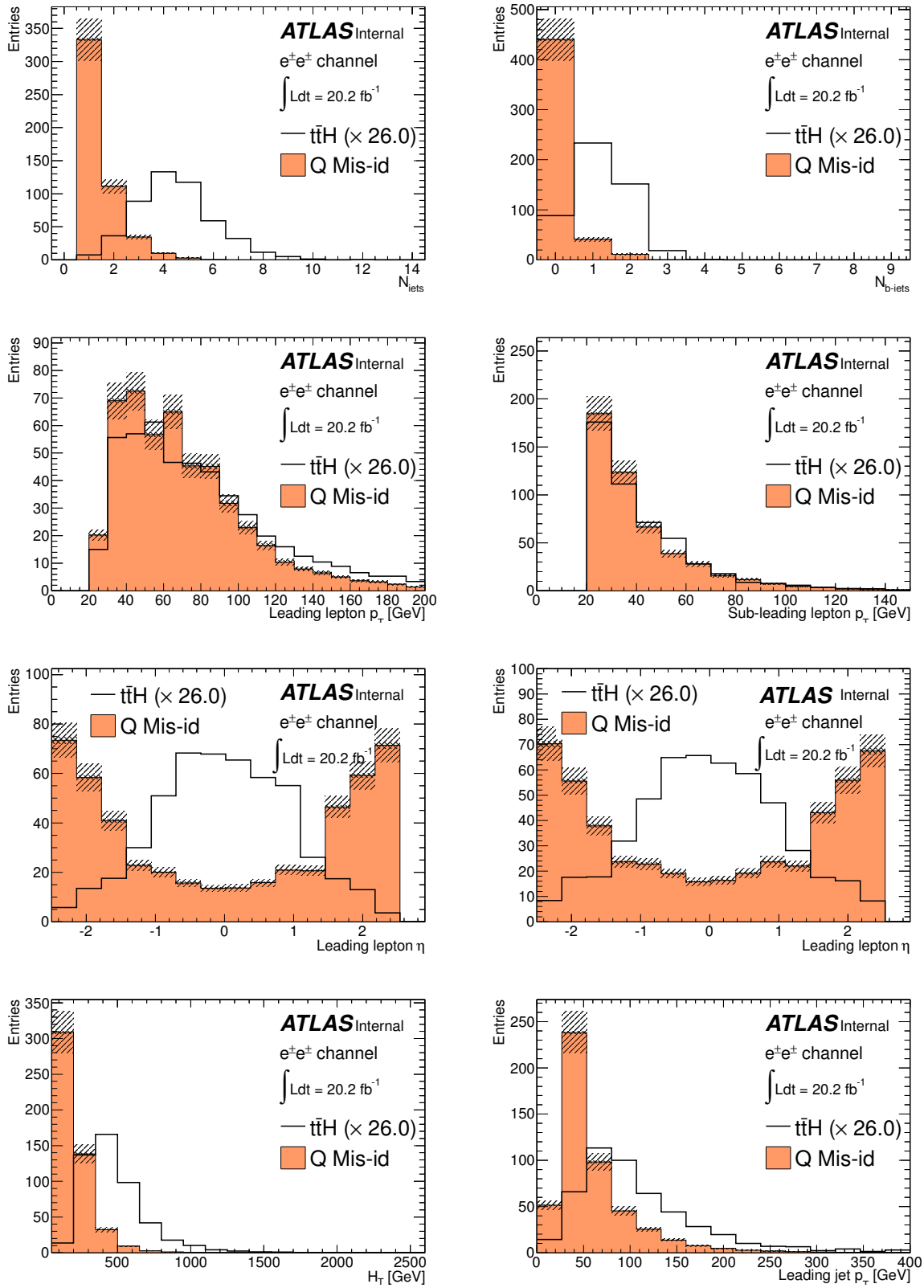


Figure 42: Expected shapes of the charge mis-identification background compared to the signal for the number of jets, number of b-tagged jets, p_{T} and η of leading and sub-leading leptons, H_T and p_{T} of the leading jet. Events shown have two same-sign electrons (Very Tight Likelihood) and at least one jet.

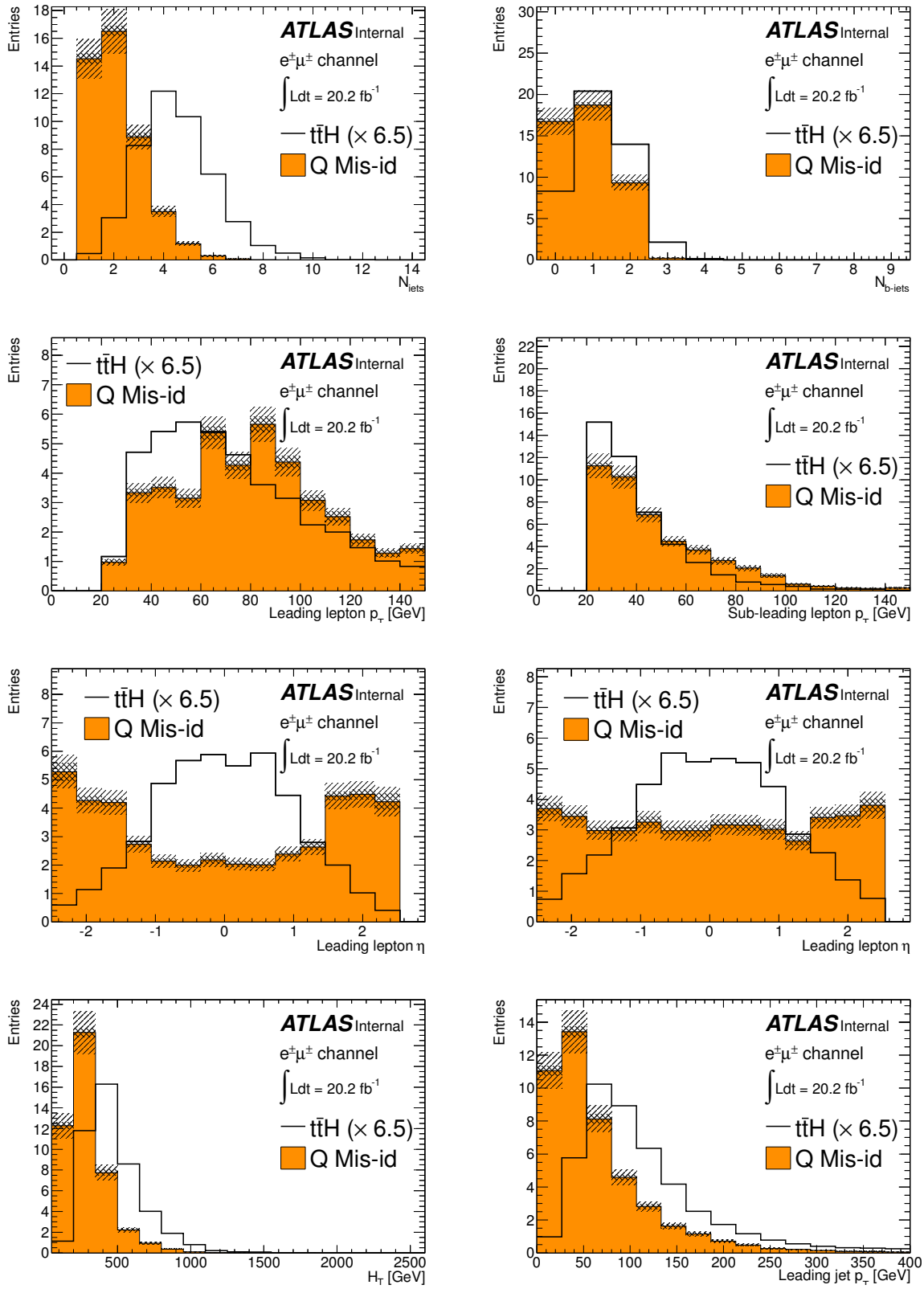


Figure 43: Expected shapes of the charge mis-identification background compared to the signal for the number of jets, number of b-tagged jets, p_{T} and η of leading and sub-leading leptons, H_T and p_{T} of the leading jet. Events shown have two same-sign $e^{\pm}\mu^{\pm}$ (Very Tight Likelihood electrons) and at least one jet.

7.2 Estimation of secondary or fake leptons

The estimate of fake or secondary leptons in the same-sign channel uses a baseline presented in Section 6 for the three lepton channel. It is however adapted to the properties of this channel where the same selection is applied on the two leptons (in $e^{\pm}e^{\pm}$ and $\mu^{\pm}\mu^{\pm}$), and where the control regions can more populated allowing the measurement of the fake factor from the data. In addition to this, the $e^{\pm}\mu^{\pm}$ channel can be used for the validation of the method.

The use of a fake factor method is motivated by the fact that the final sensitivity is extracted using a limited number of signal regions and where there is no need to predict the background shapes (counting experiment). If shapes were used, the Matrix Method would have been preferred.

7.2.1 Estimation strategy

This measurement is based on the assumption that the fake factor is stable with respect to the jet multiplicity. In the following, all the events contain at least one b-tagged jet, as in the signal region. The fake factor is defined as the ratio between the number of same-sign events with tight leptons only ($\ell\ell$) and events with one anti-tight lepton ($\ell\not{\ell}$):

$$\theta_{\ell} = \frac{N_{\ell\ell}}{N_{\ell\not{\ell}}} \tag{22}$$

where ℓ is e or μ. ℓ and $\not{\ell}$ are tight and anti-tight lepton selections. $\not{\ell}$ leptons are not used for the trigger matching.

Number of events $N_{\ell\ell}$ and $N_{\ell\not{\ell}}$, are estimated from low jet multiplicity data ($\leq$ 3 jets) after the subtraction of prompt same-sign events. They are of two categories:

- $t\bar{t}V$, $W^{\pm}Z$, $W^{\pm}W^{\pm}$: they are estimated using the simulation ;
- prompt opposit-sign events with a charge mis-identification (for electrons only). Mainly due to $t\bar{t}$ events in high jet multiplicity and with V+jets contribution in low jet multiplicity. The $\ell\ell$ events of this sort are estimated using the data ($cf.$ Section 7.1) and noted N_{ee}^{QMisId}. The $\ell\not{\ell}$ events due to two real prompt leptons represent a small fraction (4% of the $\ell\not{\ell}(\leq 3$ jets) region) and are estimated with the simulation. Simulated event (from events with top quarks) passing the $\ell\not{\ell}$ selection and where the $\not{\ell}$ matches a real prompt lepton (using a $\Delta R < 0.2$ distance) are counted and noted $N_{e\not{e}}^{\text{QMisId MC}}$ since they are all due to events ($t\bar{t}$ or V+jets) with a mismeasured lepton charge.

The measured $\theta_{\ell\ell}$ can be written as:

$$\theta_e = \frac{N_{ee}}{N_{e\not{e}}}(\leq 3\text{jets}) = \frac{N_{ee}^{Data} - N_{ee}^{\text{Prompt SS}} - N_{ee}^{\text{QMisId}}}{N_{e\not{e}}^{Data} - N_{e\not{e}}^{\text{Prompt SS}} - N_{e\not{e}}^{\text{QMisId MC}}} \tag{23}$$

$$\theta_{\mu} = \frac{N_{\mu\mu}}{N_{\mu\not{\mu}}}(\leq 3\text{jets}) = \frac{N_{\mu\mu}^{Data} - N_{\mu\mu}^{\text{Prompt SS}}}{N_{\mu\not{\mu}}^{Data} - N_{\mu\not{\mu}}^{\text{Promt SS}}} \tag{24}$$

Thanks to these subtractions, no overlap between the fake and charge mis-identification estimates is expected.

The typical value of θ being 10^{-1}, the contribution of double fake leptons is at percent level and can be neglected.

The background events due to fake electrons or muons are estimated in the two signal regions, for n jets verifying $n = 4$ or $n \geq 5$, in the three channels $e^{\pm}e^{\pm}$, $\mu^{\pm}\mu^{\pm}$ and $e^{\pm}\mu^{\pm}$ and are noted N_{ee}^{fakes}, $N_{\mu\mu}^{\text{fakes}}$ and $N_{e\mu}^{\text{fakes}}$ respectively:

$$N_{ee}(n \text{ jets}) = N_{e\not{\ell}}(n \text{ jets}) \times \theta_e \tag{25}$$

$$N_{\mu\mu}(n \text{ jets}) = N_{\mu\not{\ell}}(n \text{ jets}) \times \theta_\mu \tag{26}$$

$$N_{e\mu}(n \text{ jets}) = N_{e\not{\ell}}(n \text{ jets}) \times \theta_\mu + N_{\mu\not{\ell}}(n \text{ jets}) \times \theta_e = N_{e\not{\ell}}(n \text{ jets}) \times \theta_\mu + N_{\mu\not{\ell}}(n \text{ jets}) \times \theta_e \tag{27}$$

In each estimate of the fake lepton background and in each channel, three regions are involved:

- $\ell\not{\ell}$, ≤ 3 jets
- $\ell\ell$, ≤ 3 jets
- $\ell\not{\ell}$, n jets ($n = 4$ or $n \geq 5$)

The two first regions are used to measure the θ factors which are applied to the third region. Their size govern the statistical precision while the validity of the extrapolation is described by a systematic uncertainty.

The low jet multiplicity data in the $e^{\pm}\mu^{\pm}$ channel, which are not used to measure the extrapolation factors, can be used to check the validity of the method in the unblinded low jet multiplicity sample. Uncertainties on θ_e and θ_μ are propagated to the fake estimate in $e^{\pm}\mu^{\pm}$ as well as the systematic uncertainties.

7.2.2 Electrons

Tight electrons are defined in [1]. Two definitions of the anti-tight electrons are tested:

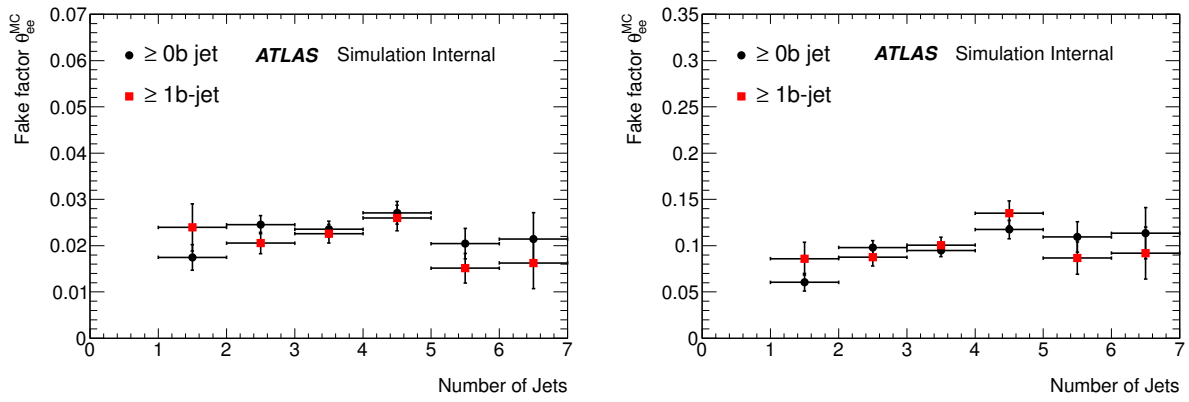
- Electron which passes all the selection criteria (including very tight likelihood Id.) except the isolation. This electron verifies $\text{E}_{\text{T}}^{\text{rel}} > 0.05$ and $\text{p}_{\text{T}}^{\text{rel}} > 0.05$. This method will be called *reverse isolation*.
- Electron which does not verify the very tight likelihood Id. and which is not isolated: $\text{E}_{\text{T}}^{\text{rel}} > 0.05$ and $\text{p}_{\text{T}}^{\text{rel}} > 0.05$. This method will be called *reverse Id.*.

The stability of the fraction ee and $e\not{e}$ with respect to the number of jets is shown on Figure 44 using these two anti-tight electron definitions.

The three regions used for the fake estimate are enriched in fake electrons, mainly due to $t\bar{t}$ events. The expected content of the $e^{\pm}e^{\pm}$ regions defined with these electrons is given in Table 21 and 22. A comparison between the expected and observed shapes is provided in Figures 46, 47, 48 and 49.

7.2.3 Muons

Tight muons are defined in [1]. Two approaches were investigated for anti-tight muons, the first one is presented in this note and relies on the fact that low p_{T} muons are enriched in fakes (cf. Section 5). The second definition of anti-tight muons is based on a reversed isolation, however, unlike the 3 lepton channel, this definition leads to a larger uncertainty in the same-sign channel.



(a) reverse Id.

(b) Reverse isolation

Figure 44: Stability of θ_e in $t\bar{t}$ events as a function of the number of jets. Events contain at least one fake electron. A fake electron is an electron which does not match a true lepton from W decay ($\Delta R < 0.2$). Anti-tight electrons are defined using a reverse Id and a reverse isolation 44(a) or a reverse isolation 44(b)

Anti-tight muons ($\not{\mu}$) are defined as muons with $p_{\mathrm{T}} < 20\ GeV$[5]. As shown on Table 21, region with one anti-tight muon are enriched in events with at least one fake muons. The stability of the fraction $\mu\not{\mu}$ over $\mu\mu$ with respect to the number of jets is tested and shown on Figure 45.

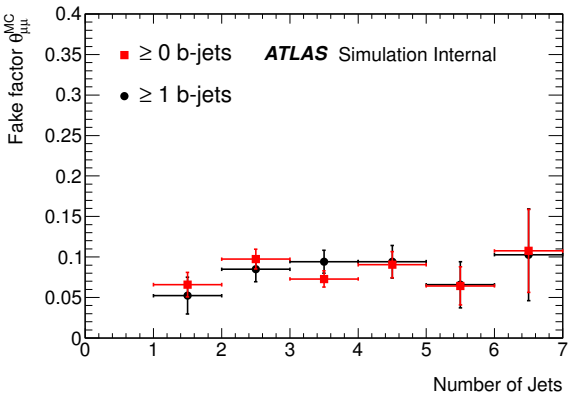


Figure 45: Stability of θ_μ in $t\bar{t}$ events as a function of the number of jets.

7.2.4 Statistical precisions

The precision on fake estimate depends on the data size in the three regions: $N_{\ell\ell}(\leq 3\text{jets})$, $N_{\ell\not{\ell}}(\leq 3\text{jets})$ and $N_{\ell\not{\ell}}(n\text{ jets})$

The number of events in these regions is used to estimate the statistical precision. Based on the number of events provided in Table 21, the expected statistical precisions on the fake prediction are summarized in Table 18.

For electrons, the method based on reverse isolation only suffers from a lack of events in the $\ell\not{\ell}(\geq 5$ jets) and $\ell\not{\ell}(4$ jets) regions. This leads to a higher statistical uncertainty on the final fake estimate compared

[5]because of the blinding only $p_{\mathrm{T}} < 10\ GeV$ are considered. Number of available $\not{\mu}$ is expected to increase after unblinding.

Channel	Statistical uncertainty [%]	
	4 jets	≥5 jets
$e^{\pm}e^{\pm}$ Rev. Id.	40.3	40.5
$\mu^{\pm}\mu^{\pm}$ Rev. p_{T}	33.3	39.7
$e^{\pm}\mu^{\pm}$	23.9	26.4

Table 18: Statistical uncertainties on fake estimate in the 3 same-sign channels and in the two signal regions (4 jets and ≥5 jets).

Factor	Expected (MC)	Measured (data)
θ_e Rev. Id.	0.0136 ± 0.0062	*0.0156 ± 0.0062*
θ_μ Rev. p_{T}	0.078 ± 0.012	*0.1156 ± 0.0288*

Table 19: Expected and measured values of the θ factors.

to the electron reverse Id method.

Since θ_e and θ_μ are used for the fake estimate in the $e^{\pm}\mu^{\pm}$ channel, the statistical uncertainty will be splitted in three terms to account for the correlation between the statistical uncertainties on N_{ee}^{fakes}, $N_{\mu\mu}^{\mathrm{fakes}}$ and $N_{e\mu}^{\mathrm{fakes}}$.

7.2.5 Fake factor from the data

θ_e and θ_μ are computed using the ratio from the ≤ 3 jets region and compared to the expected one derived from the simulation. They are compared to the measured one in Table 19.

Applying the full procedure, the estimated number of events due to fake or non-prompt lepton in the signal regions are given in Table 33.

7.2.6 Closure tests and systematic uncertainties ($e^{\pm}e^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ channels)

The validity of the method is tested with simulated $t\bar{t}$ events. The stability of the factors with respect to the jet multiplicity is checked and quantified with a closure test..

The closure test is performed by comparing the expected number of simulated $t\bar{t}$ events with fake leptons and passing the same-sign selection, $N_{\ell\ell}$, to the number of events estimated from the control regions: $N_{\ell\not{\ell}} \times \theta_\ell^{MC}$. $t\bar{t}$ events are used for this closure test if one of the two leptons does not match (within a distance $\Delta R < 0.2$) a true lepton from a W boson decay.

For the ≥ 5 jets signal region, the ratio of these two numbers is:

- $\frac{N_{e\not{e}}}{N_{ee}}(\geq 5\text{jets}) \times \theta_e^{MC} = 0.734 \pm 0.211$ (anti-Id and reverse isolation)
- $\frac{N_{e\not{e}}}{N_{ee}}(\geq 5\text{jets}) \times \theta_e^{MC} = 0.899 \pm 0.220$ (reverse isolation)

while for the 4 jet region, the closure tests result is:

- $\frac{N_{e\not{e}}}{N_{ee}}(4\text{ jets}) \times \theta_e^{MC} = 1.233 \pm 0.155$ (anti-Id and reverse isolation)
- $\frac{N_{e\not{e}}}{N_{ee}}(4\text{ jets}) \times \theta_e^{MC} = 1.448 \pm 0.163$ (reverse isolation)

The same procedure is applied to the $\mu^{\pm}\mu^{\pm}$ selection using $t\bar{t}$ events:

- $\frac{N_{\mu\not\mu}}{N_{\mu\mu}}(\geq 5\text{jets}) \times \theta_{\mu}^{MC} = 0.897 \pm 0.313$
- $\frac{N_{\mu\not\mu}}{N_{\mu\mu}}(4\text{ jets}) \times \theta_{\mu}^{MC} = 1.103 \pm 0.233$

The largest value between bias and precision on the test (due to the Monte Carlo size) is taken as systematic uncertainty on the method. The systematic effect is mainly due to the precision of the closure test on the simulation. Systematic uncertainties are summarize in Table 20 and are propagated to the $e^{\pm}\mu^{\pm}$ channel. Systematic uncertainties in the $e^{\pm}e^{\pm}$ channel being comparable between reverse id. and reverse isolation methods, the first one is adopted regarding its better statistical precision.

7.2.7 Closure tests in the $e^{\pm}\mu^{\pm}$ channel

The low jet multiplicity regions can be used to validate the extrapolation based on factor measured in $e^{\pm}e^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ channels.

Using $t\bar{t}$ MC sample with at least one fake lepton, the compatibility between predicted number of fake events and true one gives:

$$\frac{1}{N_{e\mu}^{\text{fakesMC}}} \times \left(N_{\mu\not e}^{MC} \times \theta_{e}^{MC} + N_{e\not\mu}^{MC} \times \theta_{\mu}^{MC} \right) = 1.186 \pm 0.163 \tag{28}$$

where the uncertainty is due the MC size only (uncertainty on θ not applied).

7.2.8 Validation region

Since data events with 2 or 3 jets and 1 b-tagged jet in the $e^{\pm}\mu^{\pm}$ channel are not used to to measure the extrapoplation factor they can be used as validation region. θ_e and θ_μ are applied to the number of $e\not\mu$ and $\mu\not e$ data events after subtraction of the prompt di-leptons in order to predict the fake contribution in the $e^{\pm}\mu^{\pm}$ sample: $N_{e\mu}^{\text{fakes}}$.

The comparison between this prediction and the observed number of events gives:

$$\frac{N_{e\mu}^{\text{fakes}}}{(N_{e\mu}^{\text{Data}} - N_{e\mu}^{\text{MCprompt}} - N_{e\mu}^{\text{QMisId}})} = 1.126 \pm 0.111 \tag{29}$$

where the uncertainty is statistical and due to the size of data in $e^{\pm}\mu^{\pm}$ (≤ 3 jets) sample.

Since a background normalisation is obtained with this fake estimate method, it has to be combined with a MC prediction of the distribution shapes. Using $t\bar{t}$ simulation scaled to the fake estimate in the $e^{\pm}\mu^{\pm}$ channel a background prediction can be compared to the data in this validation region. They are shown on Figures 54 and 55.

7.2.9 Systematic due to the charge mis-identification

The systematic on the charge mis-identification fraction which is subracted from the $\ell\ell$ selection is propagated to the fake estimate. This systematic is 100% correlated to the systematic associated to the charge mis-identification estimate in the signal region and will be kept separeted.

This systematic is identical for the fake estimate in 4 jets or ≥ 5 jets bins since it concerns the low jet multiplicity $\ell\ell$ region. In the $\ell\not\ell$ regions where the prompt fraction of same-sign is taken from the simulation, an uncertainty of 100% is applied. The impact of these uncertainties is given in Table 23.

Channel	Systematic uncertainty [%]	
	4 jets	≥5 jets
$e^{\pm}e^{\pm}$ Rev. Id.	21.8	26.7
$e^{\pm}e^{\pm}$ Rev. iso	44.8	22.0
$\mu^{\pm}\mu^{\pm}$ Rev. p_{T}	28.8	37.9
$e^{\pm}\mu^{\pm}$	18.4	22.0

Table 20: systematic uncertainties derived from closure tests for fake estimate in the 3 same-sign channels and in the two signal regions (4 jets and ≥5 jets). The precisions are given for two definitions of the anti-tight electrons.

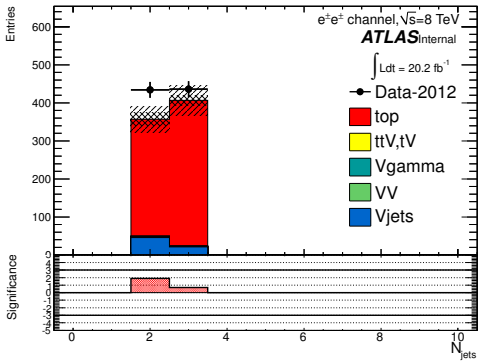
Process	N(events)
$e\not{e} \leq 3$ jets	
VV	7.13 ± 0.63
$V\gamma$	7.55 ± 1.27
$t\bar{t}V, tV$	6.68 ± 0.18
$V + jets$	59.4 ± 18.51
$t\bar{t}, t + X$	671.26 ± 12.76
$t\bar{t}$ prompts	32.97 ± 2.83
Total MC	752.0 ± 22.5
Data	967
Data / MC	1.28 ± 0.06
Data fakes (Data - prompts)	912.66

Process	N(events)
$\mu\not{\mu} \leq 3$ jets	
VV	4.34 ± 0.44
$V\gamma$	4.84 ± 2.05
$t\bar{t}V, tV$	0.74 ± 0.06
$V + jets$	21.77 ± 12.24
$t\bar{t}, t + X$	192.71 ± 6.59
Total MC	224.4 ± 14.1
Data	249
Data / MC	1.11 ± 0.09
Data fakes (Data - prompts)	239.07

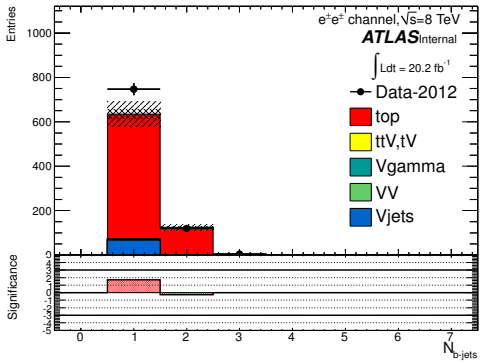
$ee \leq 3$ jets	
VV	3.30 ± 0.42
$V\gamma$	1.31 ± 0.65
$t\bar{t}V, tV$	3.96 ± 0.16
$V + jets$	8.3 ± 8.8
$t\bar{t}, t + X$	11.65 ± 1.67
Charge misID	8.54 ± 0.23
Total MC	28.52 ± 8.96
Data	32
Data / MC	1.12 ± 0.19
Data fakes (Data - prompts)	14.26

$\mu\mu \leq 3$ jets	
VV	6.27 ± 0.56
$V\gamma$	0.06 ± 0.25
$t\bar{t}V, tV$	10.00 ± 0.27
$V + jets$	1.22 ± 11.78
$t\bar{t}, t + X$	15.4 ± 2.1
Total MC	32.95 ± 2.21
Data	44
Data / MC	1.33 ± 0.17
Data fakes (Data - prompts)	27.65

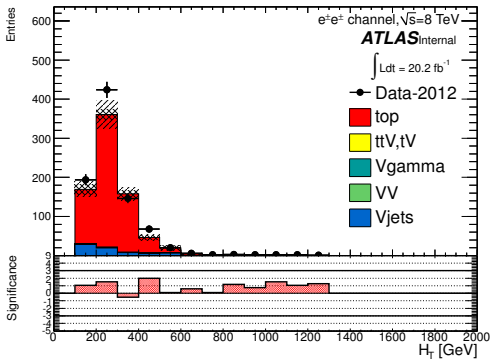
Table 21: Number of events of the main simulated background processes and of the data in the $e^{\pm}e^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ channels used for the measurement of θ_e and θ_μ. VV, $V\gamma$, $t\bar{t}V, tV$ and $t\bar{t}$ prompts (or charge misID) are the backgrounds which lead to prompt same-sign dileptons and are subtracted from the data to get a measured number of fakes. Uncertainties are statistical. The uncertainty on the $V + jets$ is neglected in the total when its contribution is negligible. The numbers labelled Data fakes are used the measure θ



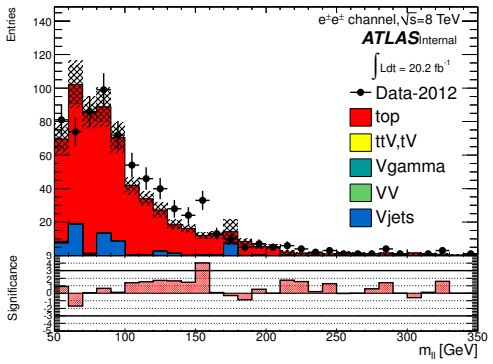
(a) Number of jets.



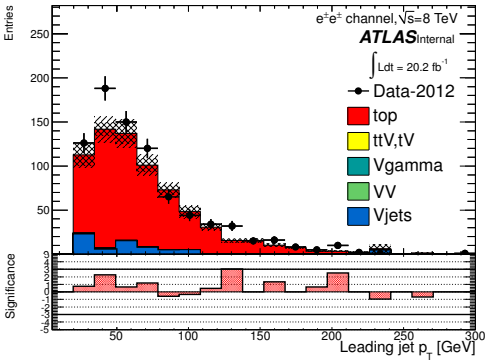
(b) Number of b-tagged jets.



(c) H_{T} (jets and leptons).

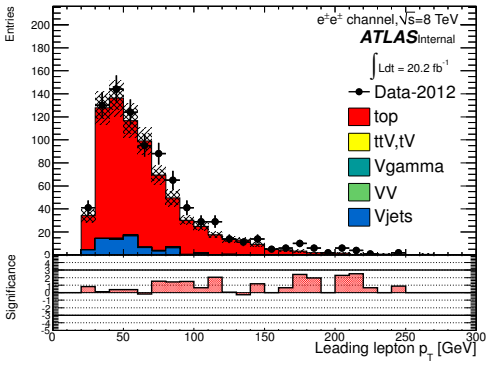


(d) Invariant mass of electron pair.

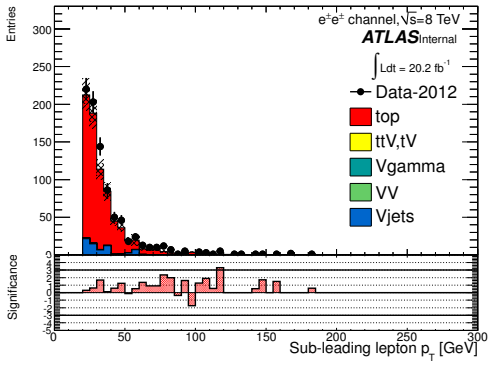


(e) Leading jet p_{T}.

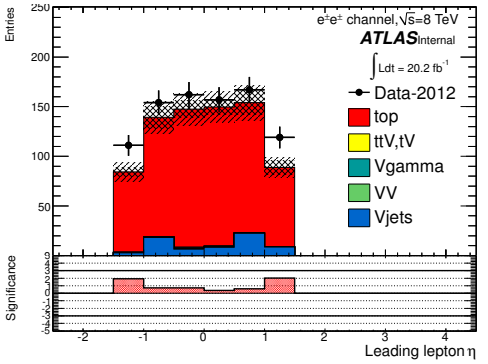
Figure 46: Comparison of the predictions for background (normalized to 20.3 fb^{-1}) and data in the same-sign $e^{\pm}e^{\pm}$ channel in the $e\not{p}$ (≤ 3 jets) region with at least one b-tagged jet.



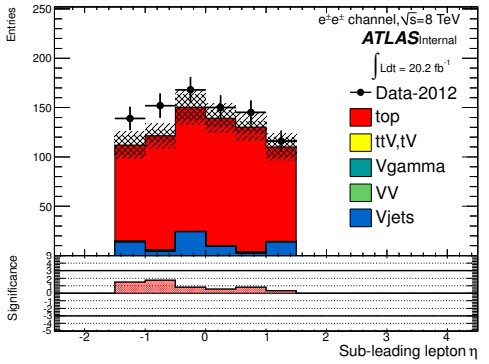
(a) Leading lepton p_{T}.



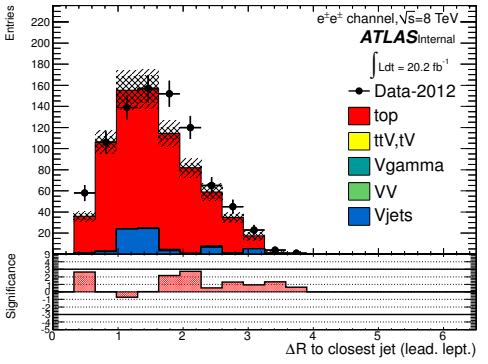
(b) Sub-leading lepton p_{T}.



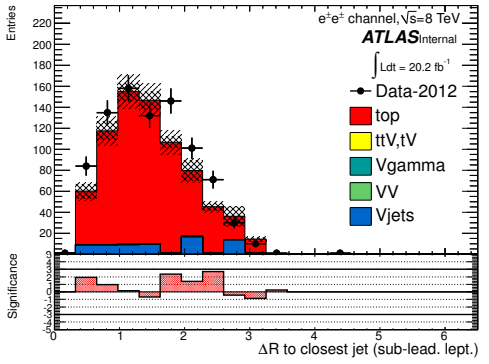
(c) Leading lepton η.



(d) Sub-leading lepton η.

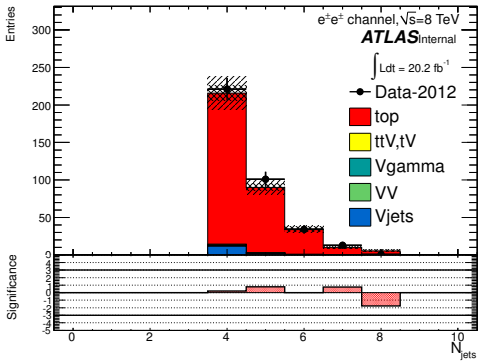


(e) Leading electron to closest jet distance.

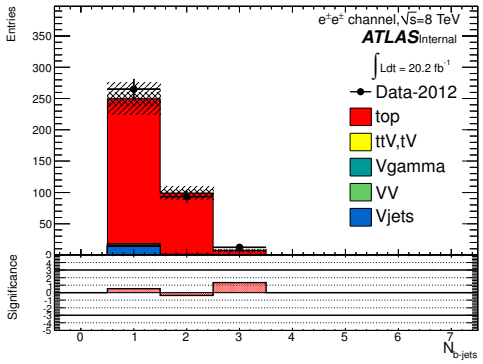


(f) Sub-leading electron to closest jet distance.

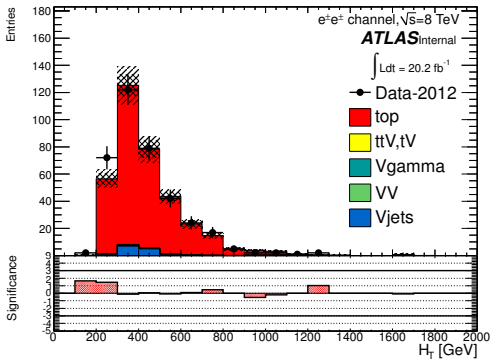
Figure 47: Comparison of the predictions for background (normalized to 20.3 fb^{-1}) and data in the same-sign $e^{\pm}e^{\pm}$ channel in the $e\not{e}$ (≤ 3 jets) region with at least one b-tagged jet.



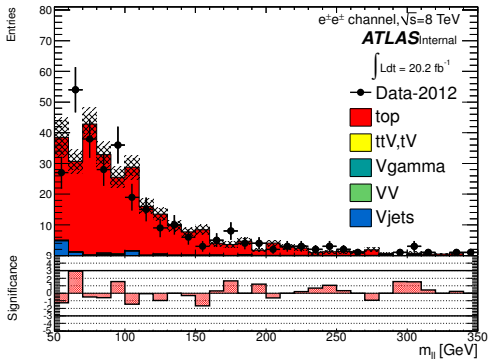
(a) Number of jets.



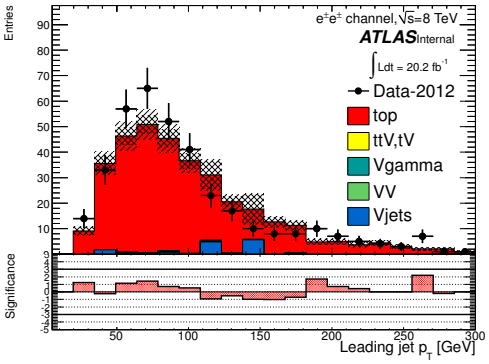
(b) Number of b-tagged jets.



(c) H_{T} (jets and leptons).

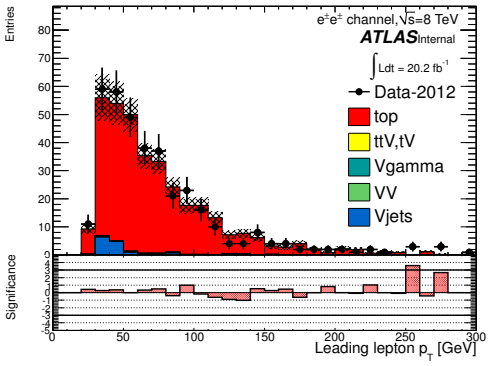


(d) Invariant mass of electron pair.

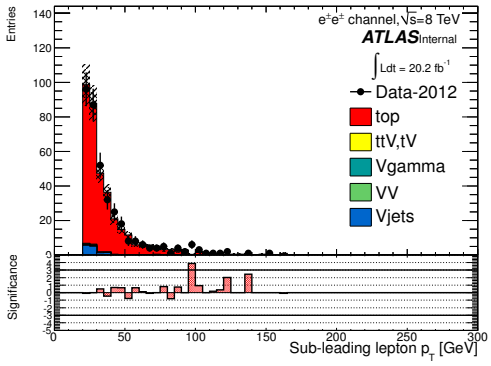


(e) Leading jet p_{T}.

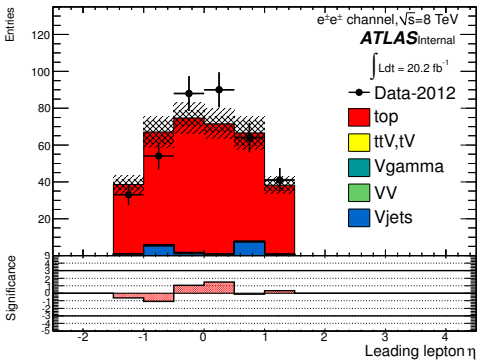
Figure 48: Comparison of the predictions for background (normalized to 20.3 fb^{-1}) and data in the same-sign $e^{\pm}e^{\pm}$ channel in the $e\not{p}$ (≥ 4 jets) region with at least one b-tagged jet.



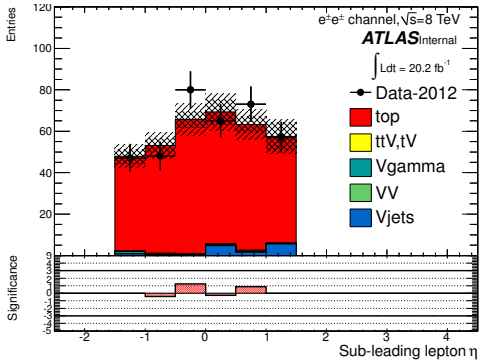
(a) Leading lepton p_{T}.



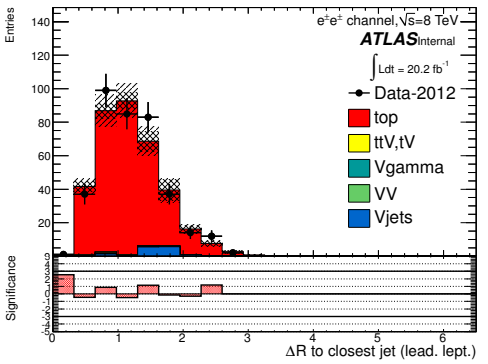
(b) Sub-leading lepton p_{T}.



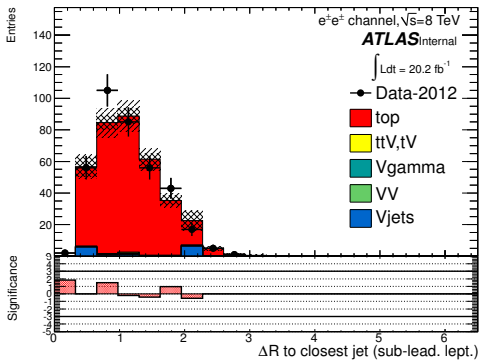
(c) Leading lepton η.



(d) Sub-leading lepton η.

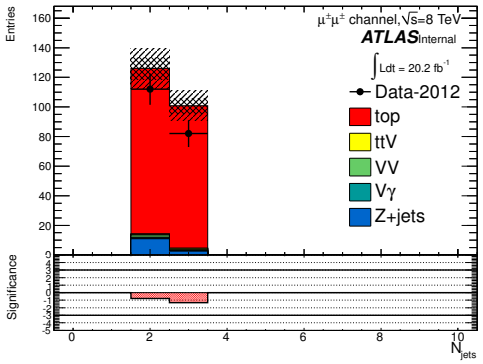


(e) Leading electron to closest jet distance.

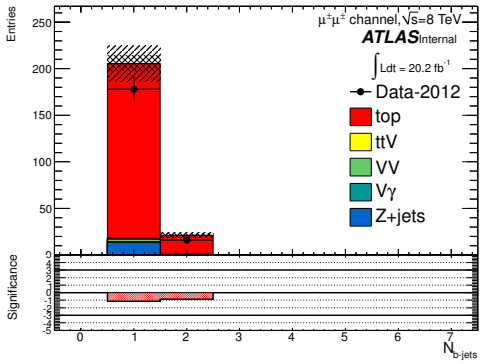


(f) Sub-leading electron to closest jet distance.

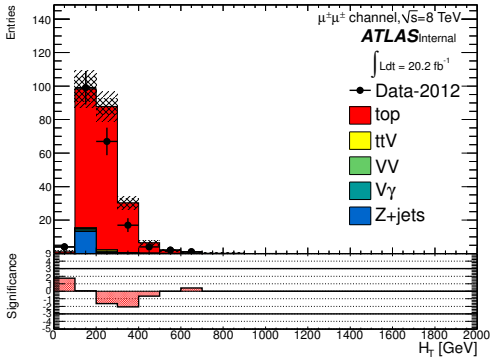
Figure 49: Comparison of the predictions for background (normalized to 20.3 fb^{-1}) and data in the same-sign $e^{\pm}e^{\pm}$ channel in the $e\not{p}$ (≥ 4 jets) region with at least one b-tagged jet.



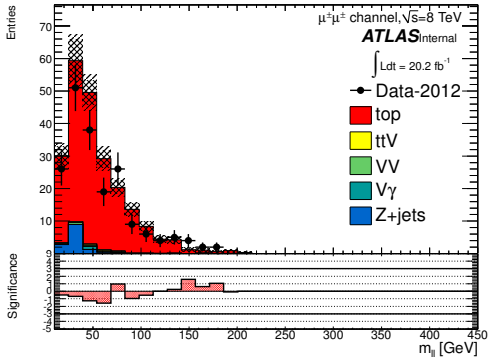
(a) Number of jets.



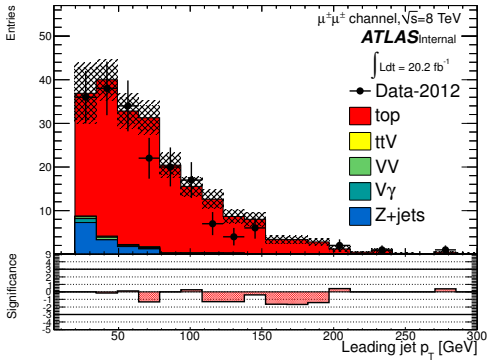
(b) Number of b-tagged jets.



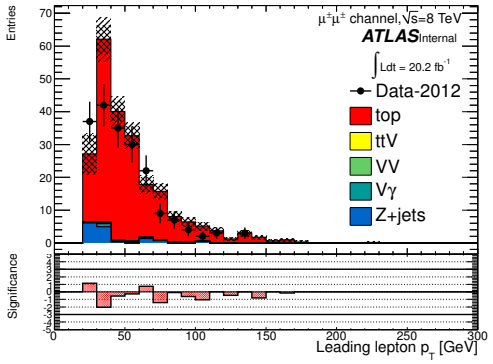
(c) H_{T} (jets and leptons).



(d) Invariant mass of muon pair.

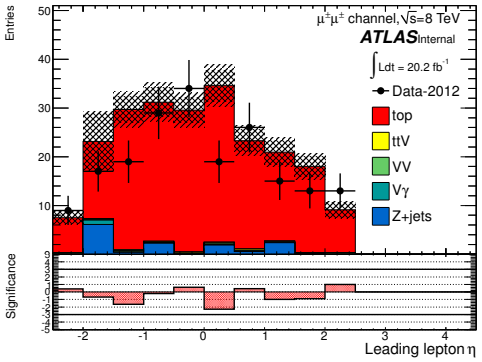


(e) Leading jet p_{T}.

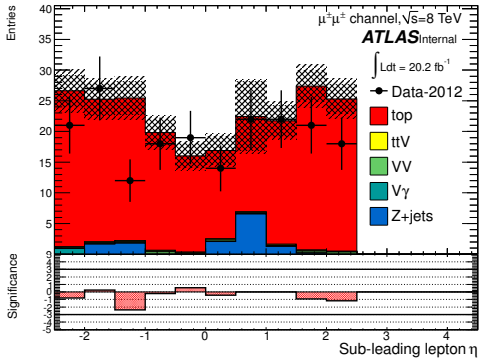


(f) Leading lepton p_{T}.

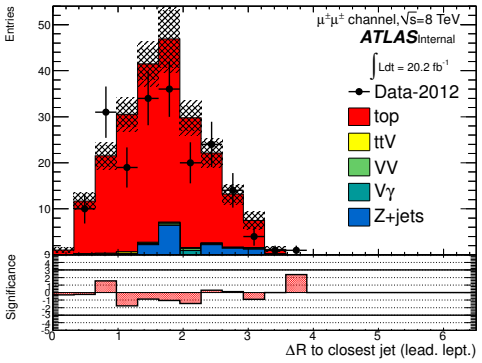
Figure 50: Comparison of the predictions for background (normalized to 20.3 fb^{-1}) and data in the same-sign $\mu^{\pm}\mu^{\pm}$ channel in the $\mu\not{\mu}$ ($\leq$ 3 jets) region with at least one b-tagged jet.



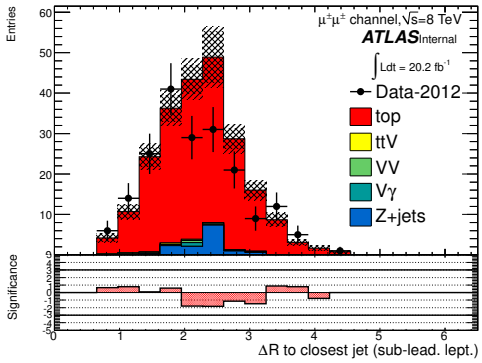
(a) Leading lepton η.



(b) Sub-leading lepton η.

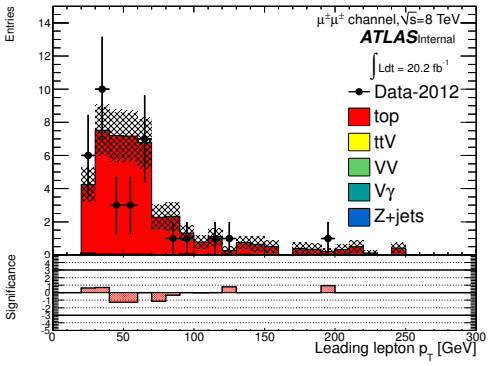


(c) Leading muon to closest jet distance.

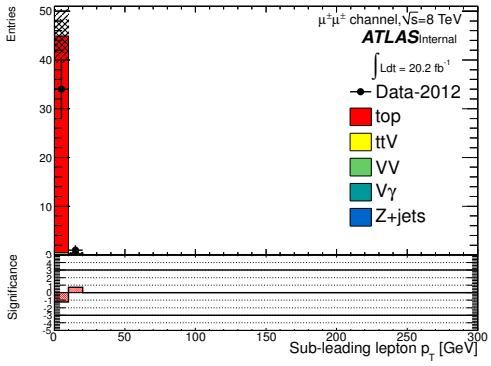


(d) Sub-leading muon to closest jet distance.

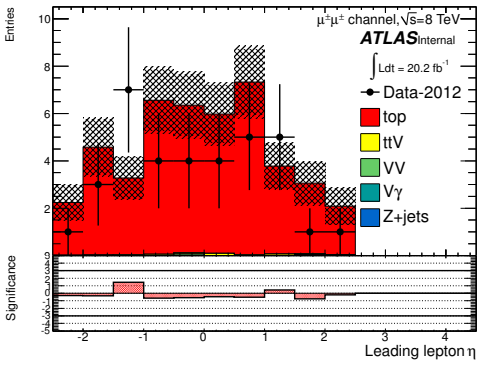
Figure 51: Comparison of the predictions for background (normalized to 20.3 fb^{-1}) and data in the same-sign $\mu^{\pm}\mu^{\pm}$ channel in the $\mu\mu\ell$ (≤ 3 jets) region with at least one b-tagged jet.



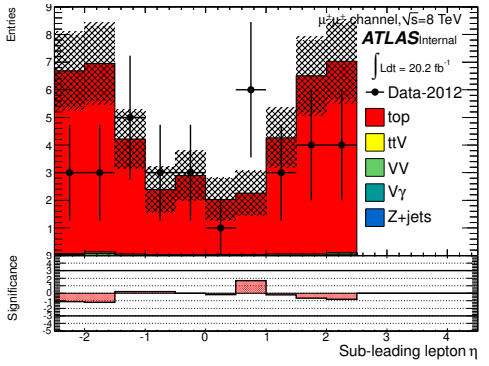
(a) Leading lepton p_{T}.



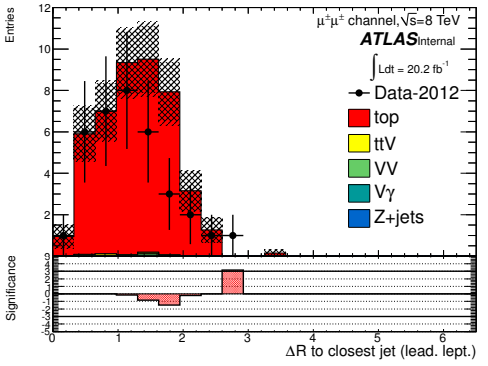
(b) Sub-leading lepton p_{T}.



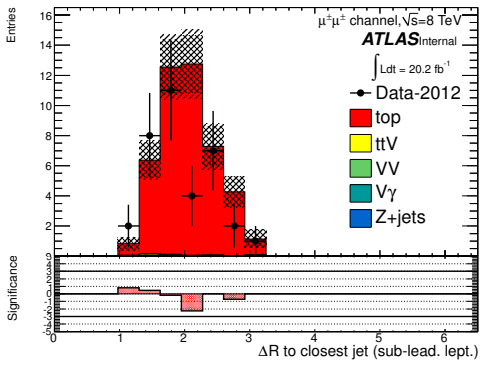
(c) Leading lepton η.



(d) Sub-leading lepton η.

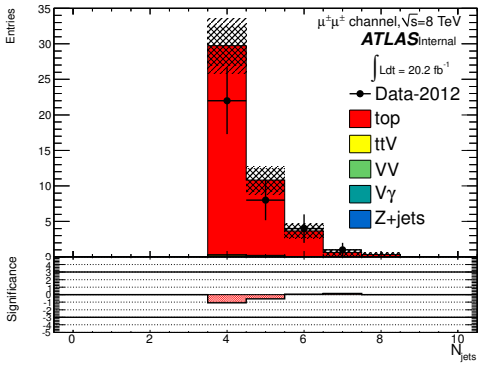


(e) Leading muon to closest jet distance.

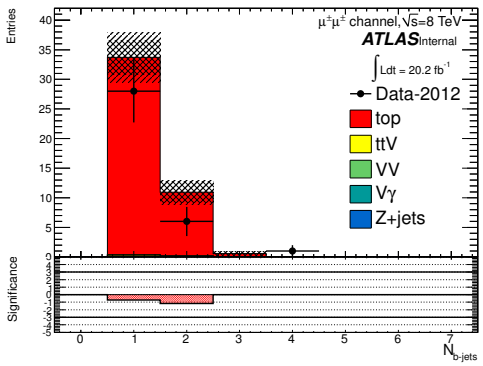


(f) Sub-leading muon to closest jet distance.

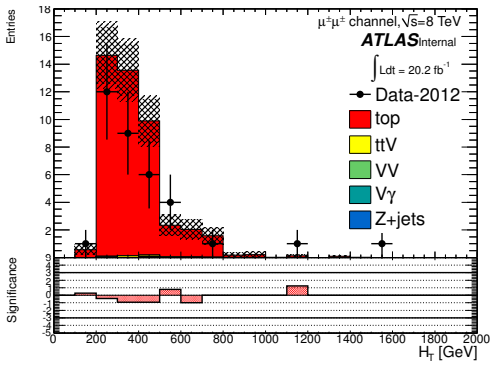
Figure 52: Comparison of the predictions for background (normalized to 20.3 fb^{-1}) and data in the same-sign $\mu^{\pm}\mu^{\pm}$ channel in the $\mu\not{\mu}$ (≥ 4 jets) region with at least one b-tagged jet.



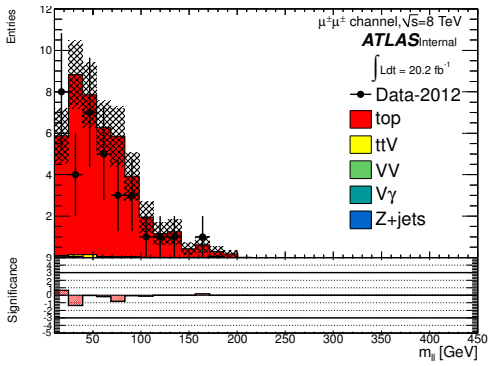
(a) Number of jets.



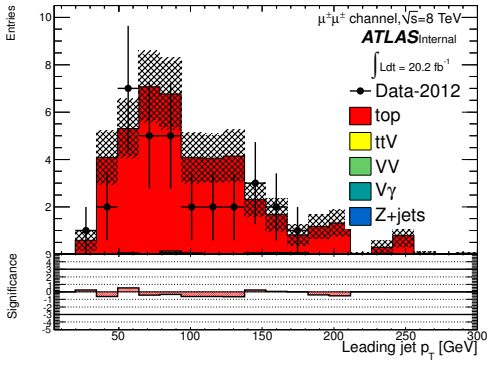
(b) Number of b-tagged jets.



(c) H_{T} (jets and leptons).



(d) Invariant mass of muon pair.



(e) Leading jet p_{T}.

Figure 53: Comparison of the predictions for background (normalized to 20.3 fb^{-1}) and data in the same-sign $\mu^{\pm}\mu^{\pm}$ channel in the $\mu\not{\mu}$ (≥ 4 jets) region with at least one b-tagged jet.

Process	N(events)
$e\not{e}$ 4 jets	
VV	0.16 ± 0.19
$V\gamma$	1.45 ± 0.73
$t\bar{t}V, tV$	3.07 ± 0.11
$V + jets$	11.61 ± 6.4
$t\bar{t}, t + X$	201.85 ± 6.83
$t\bar{t}$ prompts	3.68 ± 0.94
Total MC	218.1 ± 9.3
Data	228
Data / MC	1.05 ± 0.08
Data fakes (Data - prompts)	219.65

Process	N(events)
$e\not{e} \geq 5$ jets	
VV	0.12 ± 0.19
$V\gamma$	0.39 ± 0.46
$t\bar{t}V, tV$	3.03 ± 0.11
$V + jets$	5.77 ± 11.50
$t\bar{t}, t + X$	137.05 ± 5.52
$t\bar{t}$ prompts	2.15 ± 0.61
Total MC	146.4 ± 12.8
Data	154
Data / MC	1.05 ± 0.12
Data fakes (Data - prompts)	148.32

$\mu\not{\mu}$ 4 jets	
VV	0.11 ± 1.99
$V\gamma$	0.02 ± 0.02
$t\bar{t}V, tV$	0.25 ± 0.04
$V + jets$	0.107 ± 12.9
$t\bar{t}, t + X$	30.26 ± 2.65
Total MC	30.74 ± 3.32
Data	21
Data / MC	0.68 ± 0.16
Data fakes (Data - prompts)	20.61

$\mu\not{\mu} \geq 5$ jets	
VV	0.002 ± 0.10
$V\gamma$	0
$t\bar{t}V, tV$	0.29 ± 0.04
$V + jets$	0 ± 12.9
$t\bar{t}, t + X$	14.61 ± 2.00
Total MC	14.90 ± 2.01
Data	11
Data / MC	0.74 ± 0.32
Data fakes (Data - prompts)	10.70

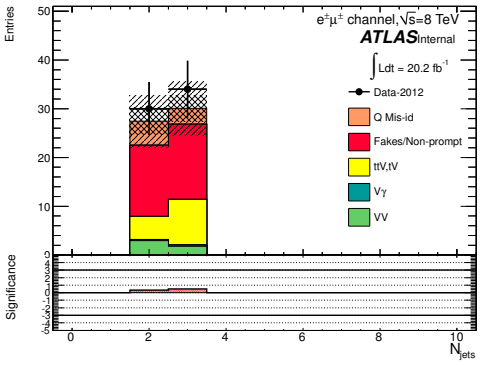
$\mu\not{e}$ 4 jets	
VV	0.13 ± 0.21
$V\gamma$	1.44 ± 0.63
$t\bar{t}V, tV$	1.49 ± 0.09
$V + jets$	1.26 ± 17.8
$t\bar{t}, t + X$	225.72 ± 7.32
$t\bar{t}$ prompts	4.75 ± 1.11
Total MC	235.37 ± 7.51
Data	180
Data / MC	0.77 ± 0.07
Data fakes (Data - prompts)	172.19

$\mu\not{e} \geq 5$ jets	
VV	0.08 ± 0.20
$V\gamma$	0.03 ± 0.40
$t\bar{t}V, tV$	1.97 ± 0.10
$V + jets$	0.35 ± 17.9
$t\bar{t}, t + X$	154.43 ± 5.89
$t\bar{t}$ prompts	1.89 ± 0.67
Total MC	158.52 ± 5.90
Data	131
Data / MC	0.83 ± 0.08
Data fakes (Data - prompts)	127.03

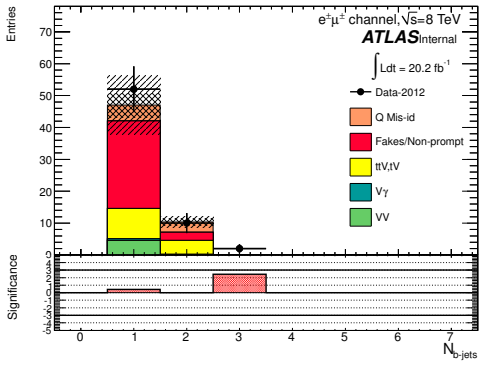
$e\not{\mu}$ 4 jets	
VV	0.89 ± 1.82
$V\gamma$	0
$t\bar{t}V, tV$	0.40 ± 0.05
$V + jets$	0.45 ± 11.7
$t\bar{t}, t + X$	22.11 ± 2.36
Total MC	24.93 ± 2.36
Data	39
Data / MC	1.56 ± 0.19
Data fakes (Data - prompts)	35.91

$e\not{\mu} \geq 5$ jets	
VV	0.002 ± 1.98
$V\gamma$	0
$t\bar{t}V, tV$	0.24 ± 0.04
$V + jets$	0.02 ± 17.8
$t\bar{t}, t + X$	11.84 ± 1.67
Total MC	12.65 ± 1.67
Data	18
Data / MC	1.42 ± 0.36
Data fakes (Data - prompts)	17.24

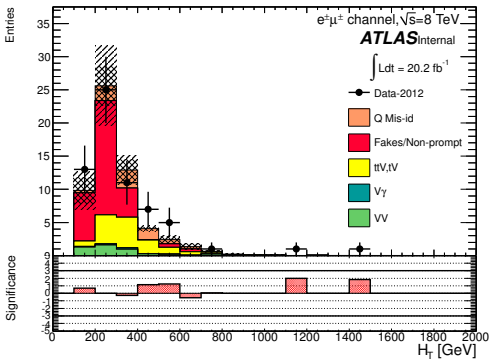
Table 22: Number of events of the main simulated background processes and of the data in the $e^{\pm}e^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ Control Regions at high jet multiplicity. $t\bar{t}$ prompts is a subsample of $t\bar{t}$ with two matching leptons (prompts). VV, $V\gamma$, ttV and $t\bar{t}$ prompts are the backgrounds which lead to prompt same-sign dileptons and are subtracted from the data to obtain the measured number of fakes. Uncertainties are statistical. The uncertainty on the $V + jets$ is neglected in the total when its contribution is negligible. The numbers labelled Data fakes are multiplied by θ_e or θ_μ in order to predict the number of fakes in the signal region



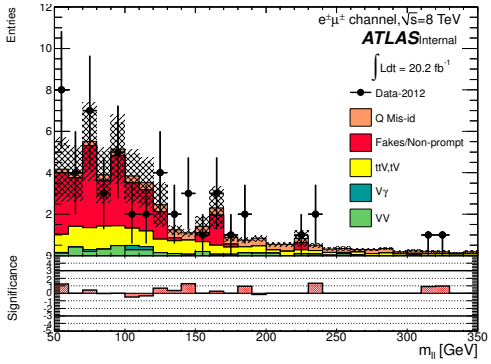
(a) Number of jets.



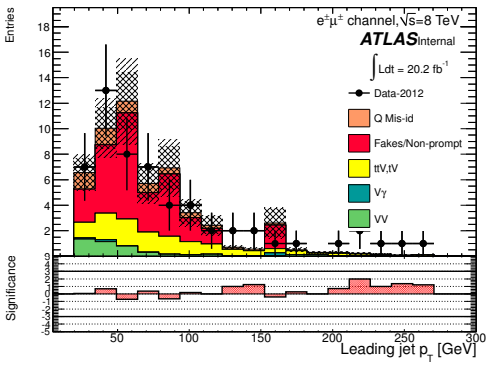
(b) Number of b-tagged jets.



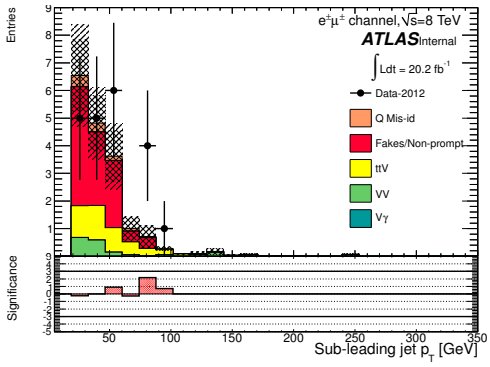
(c) H_{T} (jets and leptons).



(d) Invariant mass of electron pair.

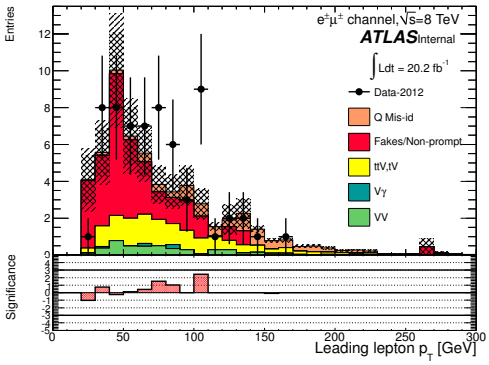


(e) Leading jet p_{T}.

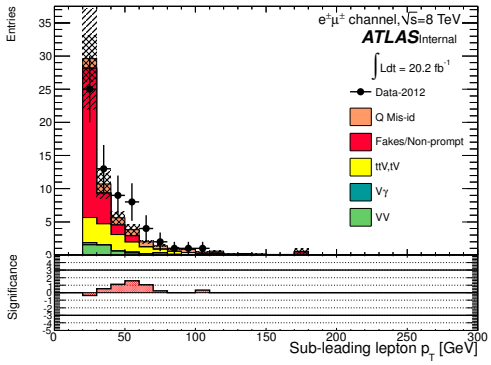


(f) Sub-leading jet p_{T}.

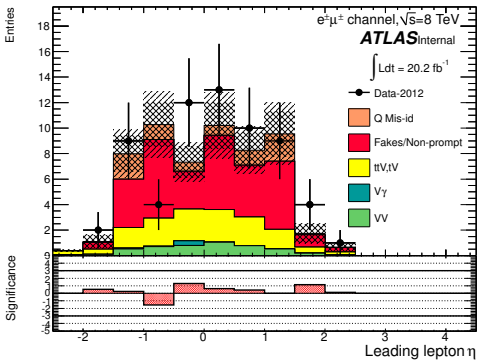
Figure 54: Comparison of the predictions for background ($t\bar{t}V$ and di-bosons normalized to 20.3 fb^{-1}, fakes/non-promts normalised with data-driven method prediction while the shape is taken simulated $t\bar{t}$) and data in the same-sign $e^{\pm}\mu^{\pm}$ channel with 2 or 3 jets and at least one b-tagged jet.



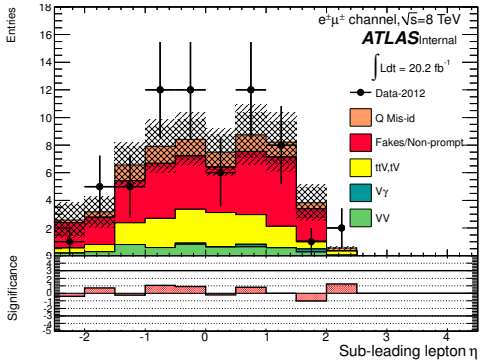
(a) Leading lepton p_{T}.



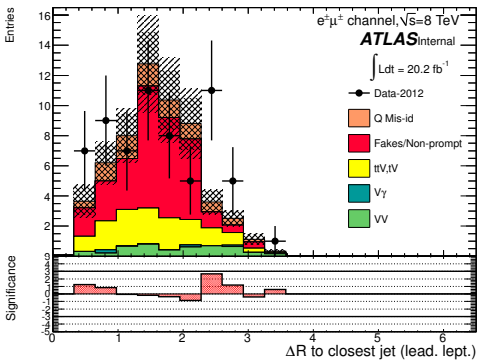
(b) Sub-leading lepton p_{T}.



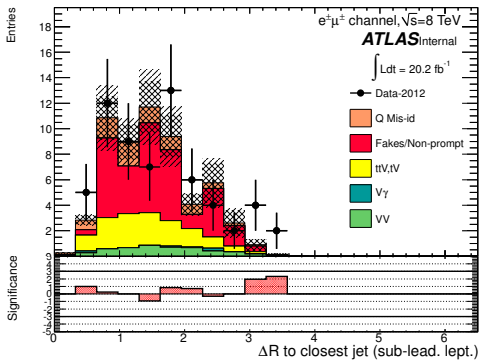
(c) Leading lepton η.



(d) Sub-leading lepton η.



(e) Leading electron to closest jet distance.



(f) Sub-leading electron to closest jet distance.

Figure 55: Comparison of the predictions for background ($t\bar{t}V$ and di-bosons normalized to 20.3 fb^{-1}, fakes/non-promts normalised with data-driven method prediction while the shape is taken simulated $t\bar{t}$) and data in the same-sign $e^{\pm}\mu^{\pm}$ channel with 2 or 3 jets and at least one b-tagged jet.

Uncertainties		Channels					
		4 jets			≥ 5 jets		
		$e^{\pm}e^{\pm}$	$\mu^{\pm}\mu^{\pm}$	$e^{\pm}\mu^{\pm}$	$e^{\pm}e^{\pm}$	$\mu^{\pm}\mu^{\pm}$	$e^{\pm}\mu^{\pm}$
Statistical	$\Delta\theta_e^{\text{stat}}$	39.6	–	14.2	39.6	–	18.5
	$\Delta\theta_\mu^{\text{stat}}$	–	24.7	15.8	–	24.7	13.1
	$\Delta N_{\ell\ell}(n \text{ jets})$(stat)	6.8	18.1	21.3	8.4	25.9	22.7
Systematics	$\Delta\theta_e^{\text{syst}}$ (closure)	21.8	–	7.8	26.7	–	12.5
	$\Delta\theta_\mu^{\text{syst}}$ (closure)	–	23.3	18.4	–	31.2	19.7
	Q Mis Id ($\ell\ell$)	2.2	–	1.5	2.4	–	1.5
Total		**45.7**	**38.5 (36.3)**	**35.7**	**48.5**	**47.8 (43.9)**	**39.6**
Systematic	Q Mis Id ($\ell\ell$)	24.0	–	8.6	24.0	–	11.3

Table 23: Summary of the uncertainties (in %) in $e^{\pm}e^{\pm}$ (reverse Id + reverse isolation method), $\mu^{\pm}\mu^{\pm}$ and $e^{\pm}\mu^{\pm}$. Statistical uncertainty is splitted into statistical uncertainties on θ_e and θ_μ and uncertainty due to the Control Region size ($\Delta N_{\ell\ell}(n \text{ jets})$). For channels with muons, uncertainties between parenthesis are obtained when anti-tight muons are defined such as $p_{\text{T}} < 20$ GeV (includes blinded data)

7.3 Estimation of $t\bar{t}V$ events in the signal region

Production of top plus vector boson (including $t\bar{t}V$ and tZ) is one of the two dominant background processes in the same-sign channel, together with the fake lepton background. It is estimated using Monte Carlo simulation and the predicted contribution is shown in Table 33.

7.4 Effect of electron charge mis-identification

The method described in Section 7.1 is applied to the same-sign channel with its nominal selection in the two signal regions (4 and ≥5 jets). Electrons pseudo-rapidity must verify $|\eta| < 1.5$ in order to reduce the impact od the charge mis-identification in the signal region. The predictions are given in Table 33.

The estimate of fake lepton background, charge mis-identification background, irreducible backgrounds ($t\bar{t}V$) presented in this section lead to a sensitivity which is compatible with the one targetted during the optimisation prodecure described in [1].

8 2 lepton Same-Sign plus one τ_{had}

The Main background contributions in the sub-channel with exactly two same-sign (SS) leptons (in the following lepton stands for either an electron or a muon) and exactly one hadronic tau candidate are production of $t\bar{t}$- and single-top events as well as $t\bar{t}V$. Smaller but non-negligible contributions are from events with diboson processes. In the following $t\bar{t}$, single-top and also four-top processes are summarized as top and give the most important background. Negligible backgrounds are due to Z+jets, W+jets and QCD multijet events.

8.1 Overview

Processes from top production are fake dominated. That means they include at least one object out of the two leptons and the hadronically decaying tau lepton (τ_{had}) which are misidentified from objects, like light quark- or gluon jets, heavy flavour jets or which have a mis-measured electric charge. The yield of these background processes are relatively small (in the order of one event) and the Monte-Carlo (MC) statistics is limited, especially for top involved processes. In addition, the object-misidentification rates include high uncertainties in MC simulations. Therefore, datadriven estimates using side-band regions are aimed for fake backgrounds in order to suppress MC generation based uncertainties. Processes from $t\bar{t}V$ and diboson production contribute by producing two real SS leptons and one τ_{had}, as the signal process does. Thus, sufficiently pure control regions are not available due to its signal-like event topology. Therefore, its an unreducible background and its yield is taken from MC, as well as that of diboson background events.

8.2 Estimation of fake backgrounds from top processes

8.2.1 Estimation strategy:

The signal region (SR) in that channel is defined by requiring two light leptons (electrons or muons) with the same reconstructed charge, one hadronic tau candidate, four reconstructed jets of which one is required to fulfil the b-tag criteria. The reducible main background contributions of $t\bar{t}$can produce two real isolated leptons if both Ws decay leptonically. As they would always have opposite charge, $t\bar{t}$events can leak in the signal region onlyif one of the leptons is a fake one. That means it is not produced by a W of the top decays but by heavy flavour (HF) decays or misidentified by light-quark or gluon jets. Thruth studies, as presented in Section 5, have shown that the fake origin of leptons are almost always B-meson or D-meson decays which produce one real electron or muon. There is also a probability that electrons are faked by light jets, but this was obtained to be less than 5%. Hence, the two SS leptons in the SR consist of one real, isolated one from a W decay and one real, non-isolated one from a HF decay which is SS to the W. There is also a small probability that both leptons are from real W decays and an electron has a misidentified charge. This case will be discussed later. The τ_{had} does not have to be fake, but can also be produced by the other W which is not decaying into a light lepton. Hence the τ_{had} contribution cannot be considered to be fake only but consists also of real τ_{had} objects. Truth studies have shown that the τ_{had} in $t\bar{t}$events in the SR consist to about 40% of true taus from W and to about 60% of fake taus from mostly jets and also B-meson to τ_{had} decays to a very small amount (compare Table 25).

To be able to account for the various fake sources and combinations of different fake contributions, it is aimed for a data-driven event-based estimation strategy which is inclusive in those fake possibilities. The data-driven estimate of top processes is based on an two-dimensional sideband method (or ABCD method [18]) which inverts two signal selection cuts to obtain three control regions (CR) orthogonal to

the SR. The signal region is hence labelled as A while the three CR are named as B, C and D. A high purity of the background process of top production has to be reached in the three CR. Then the contribution of this background in the SR can be estimated by the measured data yields in the CR B, C and D. Small remaining contributions of other background sources in the CR can be taken from MC and subtracted from data in the CR.

8.2.2 Definition of control regions

The three control regions of the ABCD method are constructed by inverting the cut on the number of jets compared to the signal region and by requiring only one tight light lepton and one loose one. The definitions of the loose electrons and muons is given by changing the requirements in Table 24 compared to the nominal tight lepton requirements.

Table 24: Defintions of loose leptons compared to tight ones.

Requirement	Electron	Muon
E_T-cone rel (inverted)	> 0.05	> 0.1
P_T-cone rel (relaxed)	< 1.0	< 1.0
ID (relaxed for electrons)	LooseLH	Tight

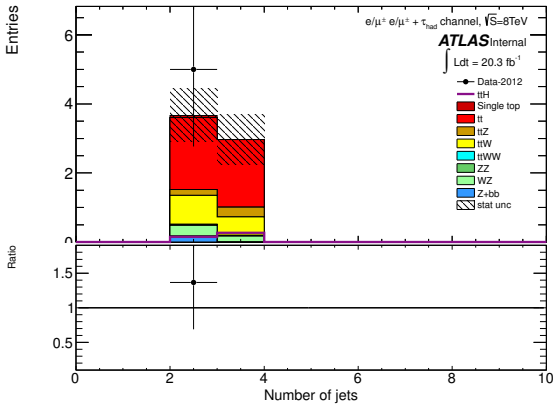
The aim of this loose lepton definitions is to construct control regions with one loose lepton instead of a tight one and obtaining about the same fraction of B-meson to lepton decays as in the SR with two tight leptons. Thus, the CR for the sideband method are defined as follows:

- A: SS, $N_{jets} \geq 4$, $N_{lep}(tight) = 2, N_{lep}(loose) = 0$
- B: SS, $1 < N_{jets} < 4$, $N_{lep}(tight) = 2, N_{lep}(loose) = 0$
- C: OS, $N_{jets} \geq 4$, $N_{lep}(tight) = 1, N_{lep}(loose) = 1$
- D: OS, $1 < N_{jets} < 4$, $N_{lep}(tight) = 1, N_{lep}(loose) = 1$

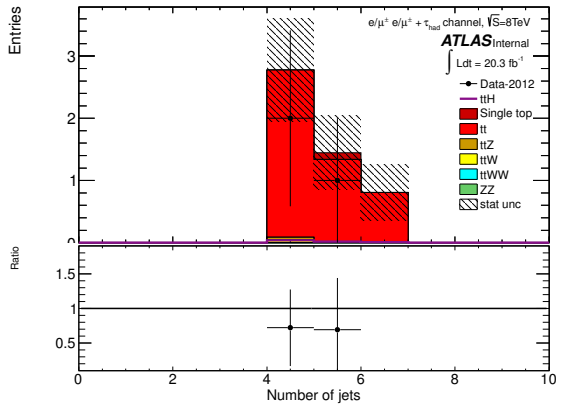
The purity of the CR B, C and D with regards to top processes are shown in the $N_j$et distribution in Figure 56. It can be seen that the two regions C and D with a loose electron or muon are very pure and enriched in $t\bar{t}$events while the tight CR B contains also a non-negligible fraction of $t\bar{t}V$ and Diboson events. These yields are subtracted from data for the fake estimate.

With this regions it is possible to measure the probability that a fake lepton from a B-meson decay passes the tight definition of the SR by using the event yields in regions D and B and apply this factor to region C. This gives the estimate of fake events in the SR A. An event based estimate taking the combined yields of loose electron and loose muon regions is chosen due to the limited statistics, especially in CR B of only 5 data events, limiting the staatistical precision of the method. Also the fraction of hadronic taus has to be considered in the method. As a large fraction is true τ_{had} from W decays, it is not possible to use a standalone tau fakerate method to account for the taus in the events. Instead, the true and fake tau contributions have to be obtained together. This is covered by the applied method as the number of fake events including the tau requirements is estimated from the three sideband regions. The truth compositions of the τ_{had} in the 4 defined regions A,B,C,D are summarized in Table 25.

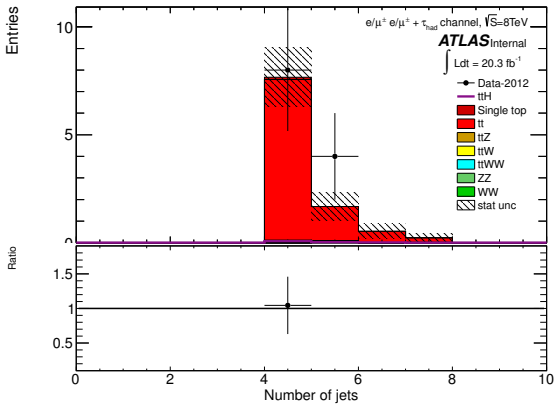
It can be seen that the truth compositions of τ_{had} are slightly different between the low-Njet regions B and D on the one hand and the high-Njet regions A and C on the other hand. As the event yield ratio of regions B and D is used to propagate the measured yield of region C into the SR A, the different fractions of true and fake taus could lead to a bias in the fake estimate. The difference of true tau fractions between



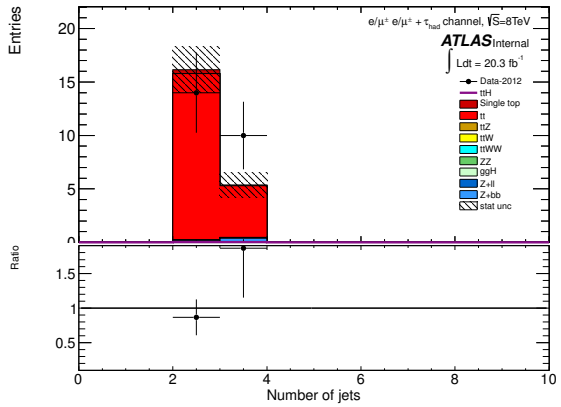
(a) N_{jets} in CR B



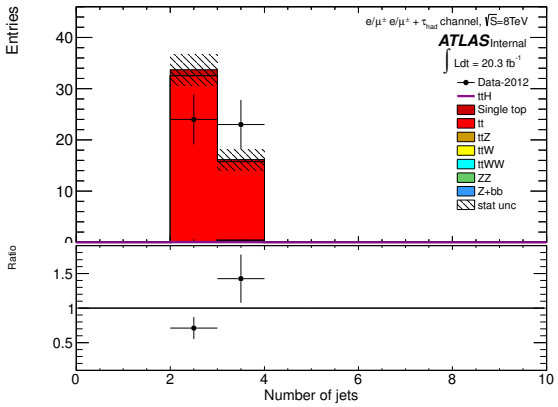
(b) N_{jets} in CR C with one loose electron



(c) N_{jets} in CR C with one loose muon



(d) N_{jets} in CR D with one loose electron



(e) N_{jets} in CR D with one loose muon

Figure 56: The distributions of the N_{jets} variable for CR B (a), CR C (b),(c) and CR D (d),(e).

Table 25: Composition of τ_{had} origins obtained from truth studies in the SR A and the three CR (B,C,D) used by the fake estimation method. This is performed with the fast-simulated version of (AF2) of the inclusive $t\bar{t}$ MC dataset (ID 117050). The light lepton p_T cut is lowered to 10 GeV for this study only to increase the statistics.

τ_{had} source	A [%]	B [%]
$W^{\pm} \to \tau_{had}$	**40.6**	**50.0**
Electron	4.3	0.6
Light Jet	53.5	47.3
$B \to \tau_{had}$	0.0	0.0
No truth match found	1.4	2.2
τ_{had} source	C [%]	D [%]
$W^{\pm} \to \tau_{had}$	**29.8**	**42.9**
Electron	0.4	0.5
Light Jet	64.1	52.1
$B \to \tau_{had}$	0.2	0.4
No truth match found	4.1	3.2

loose regions C and D is at 44% relative. This difference is added as additional systematic uncertainty to the fake estimate as this would give the maximal bias when propagating from regions B and D into C and A if the transfer factors for real and fake taus are maximal different.

Further, a MC closure test of this sideband method with top MC is performed using Equation 30:

$$N^A_{top} = \frac{N^C_{top}}{N^D_{top}} N^B_{top}. \tag{30}$$

This is done including the fully simulated (FULLSIM) inclusive $t\bar{t}$ MC dataset. The results of that closure tests are summarized in Table 26. The yields of the SR A and the estimation of SR A with the ABCD method are in agreement within the statistical uncertainties. The uncertainty of this closure test is included as a source of systematic uncertainty to the fake estimate.

Table 26: Summary of ABCD closure test from top MC (including single top and 4-top processes).

Region	FULLSIM
N^A_{top}	0.81±0.44
$\frac{N^C_{top}}{N^D_{top}} N^B_{top}$	0.86±0.25
$(\frac{N^C_{top}}{N^D_{top}} N^B_{top})/N^A_{top}$	1.07±0.67

8.2.3 Estimation of the fake background

The amount of top included events in the SR A is estimated from the data yields in regions B, C and D as:

$$N^A_{fake} = \frac{N^C_{Data-MC}}{N^D_{Data-MC}} N^B_{Data-MC}, \tag{31}$$

with MC being the remaining MC yields of $t\bar{t}V$, ttH and diboson processes in the corresponding region B, C or D. These contributions are subtracted from data using the obtained MC yields.

The result of the estimate yields to:

$$N^{A}_{fake} = \frac{N^{C}_{Data-MC}}{N^{D}_{Data-MC}} N^{B}_{Data-MC} = 0.46^{+0.59}_{-0.40}. \quad (32)$$

The uncertainty given here is the statistical only and includs the statistical uncertainty of the data yields in the three CR, as well as those of the subtracted MC yields. The low data event yield in CR B of only 5 events is dominating the statistical uncertainty and leads to the asymmetric Poisson error treatment. As discussed above, there is also a non-negligible probability that two real isolated light leptons of $t\bar{t}$events appear in the signal region in case one electron has a misidentified electric charge. As its an event-based estimation method, this case is covered by the method and included in the estimate and its uncertainties. The CR B with two tight leptons includes the fraction of charge mis-identified electrons as well. Therefore, it is also included in the estimate to SR A by applying the transfer factor from region D to B.

8.2.4 Systematic Uncertainties

One source of systematic uncertainty is the MC closure test as given in Table 26. The statistical precision of the closure test of 63%. A second source of systematic uncertainties is the different fraction of true and fake taus in CR B and D compared to CR C and SR A. The assigned systematic uncertainty is 44% which corresponds to the largest obtained variation of the true tau fraction between region C and D. The total relative systematic uncertainty of these two contributions is calculated as:

$$\Delta N^{A}_{sys}(relative) = \sqrt{(\Delta N^{closure})^2 + (\Delta N^{composition(taus)})^2} = \sqrt{(0.62)^2 + (0.44)^2} = 0.76. \quad (33)$$

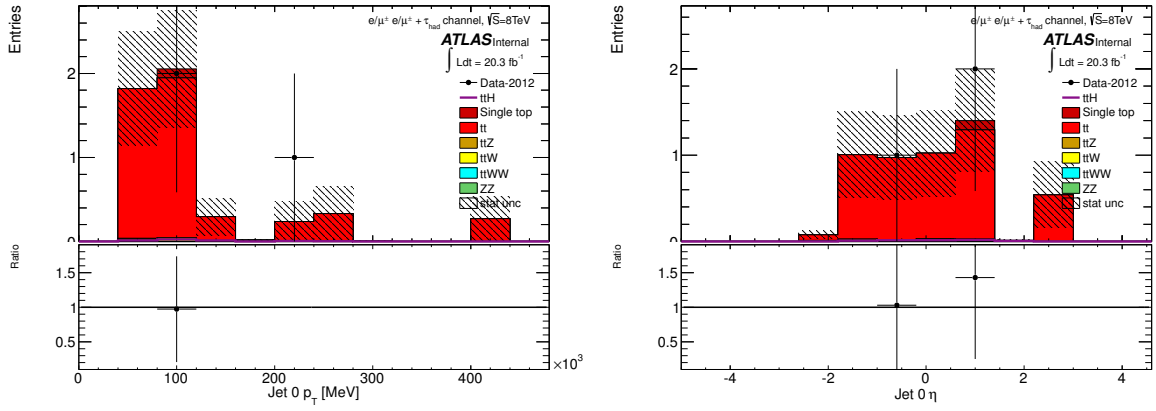
8.2.5 Results

The results obtained from Equations (32) and (33) lead to an estimation of top processes in the signal region of N^{A}_{top}(estimate) = $0.46^{+0.59}_{-0.40}$(stat) ± 0.35(sys) = $0.46^{+0.69}_{-0.53}$events.

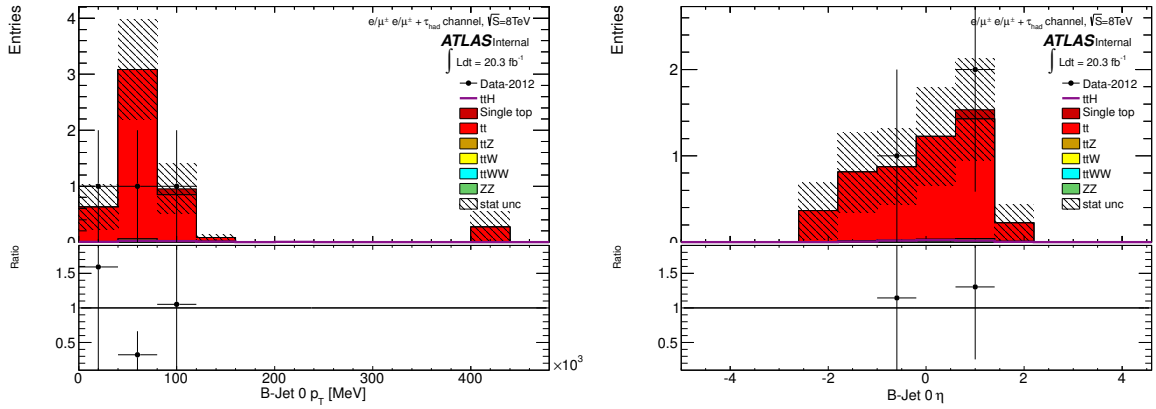
It can not be excluded that very small background contributions from Z+jets, W+jets or Multijet processes end up in the signal region. This can hardly be tested with the limited MC statistics and the passing MC events are at 0.0 (more details can be found in Section 8.3). However, these events would potentially end up in the control regions B, C and D as well. As the measured data in these regions is used for the estimate in the signal region A, it is assumed that small contributions of such processes are included in the estimate and the assigned uncertainties.

8.2.6 Cross-checks

Further properties of the object variables are shown in Figures 57 to 68. The control region with identical selection as the signal region, but with the requirement of one loose electron is shown in Figures 57-59. The control region with identical selection as the signal region, but with inverted jet multiplicity requirement and one loose electron are shown in Figures 60-62. The control region with identical selection as the signal region, but with the requirement of one loose muon is shown in Figures 63-65. The control region with identical selection as the signal region, but with inverted jet multiplicity requirement and one loose muon is shown in Figures 66-68. Overall a good data to MC modeling is observed within the statistical uncertainties.

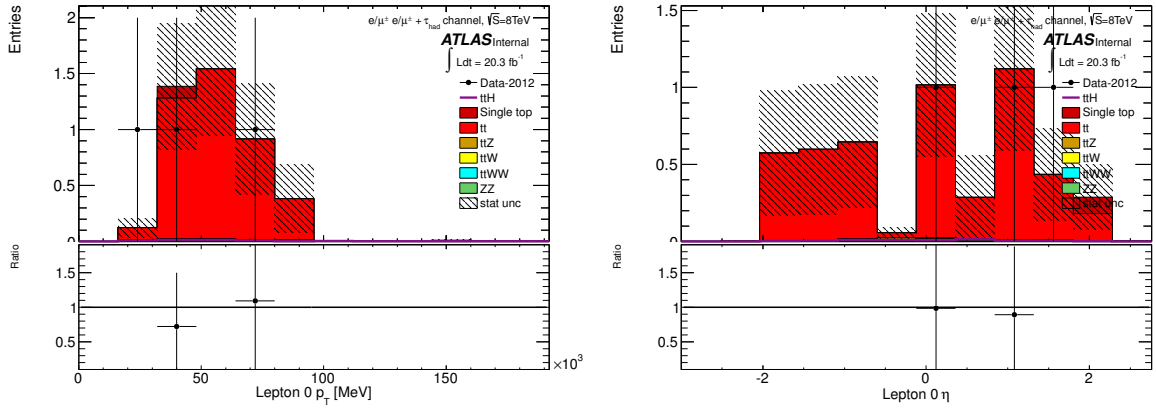


(a) $p_T(\text{jet}_{\text{leading}})$ (b) $\eta(\text{jet}_{\text{leading}})$

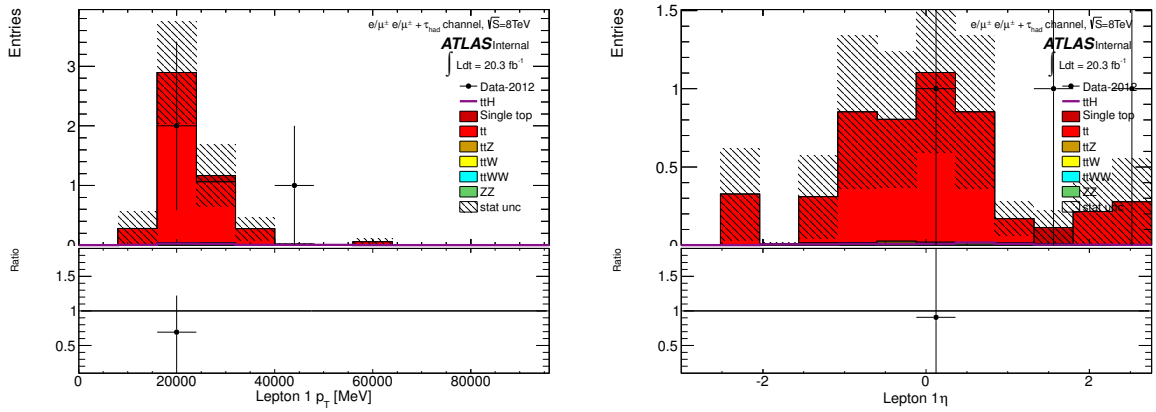


(c) $p_T(\text{Bjet}_{\text{leading}})$ (d) $\eta(\text{Bjet}_{\text{leading}})$

Figure 57: The distributions of jet kinematic variables in the SS high Njet control region with one loose electron.

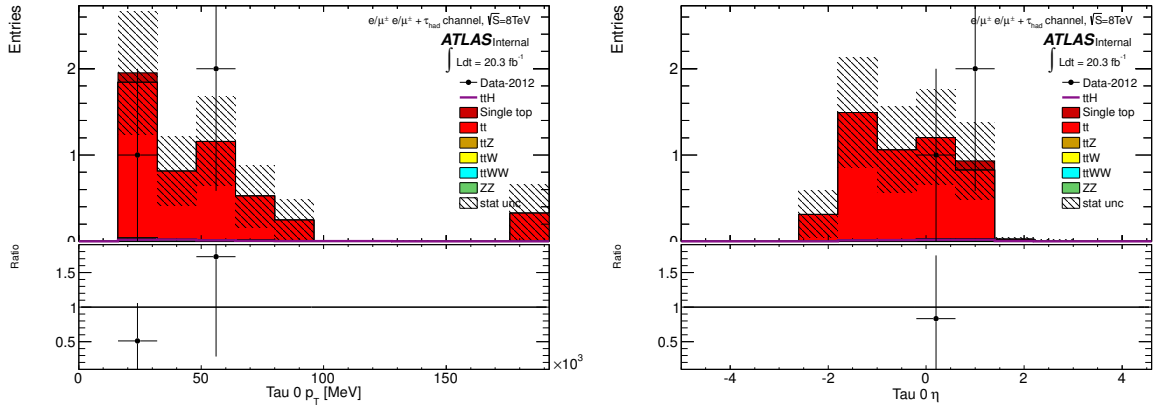


(a) $p_T(\text{lepton}_{\text{leading}})$ (b) $\eta(\text{lepton}_{\text{leading}})$

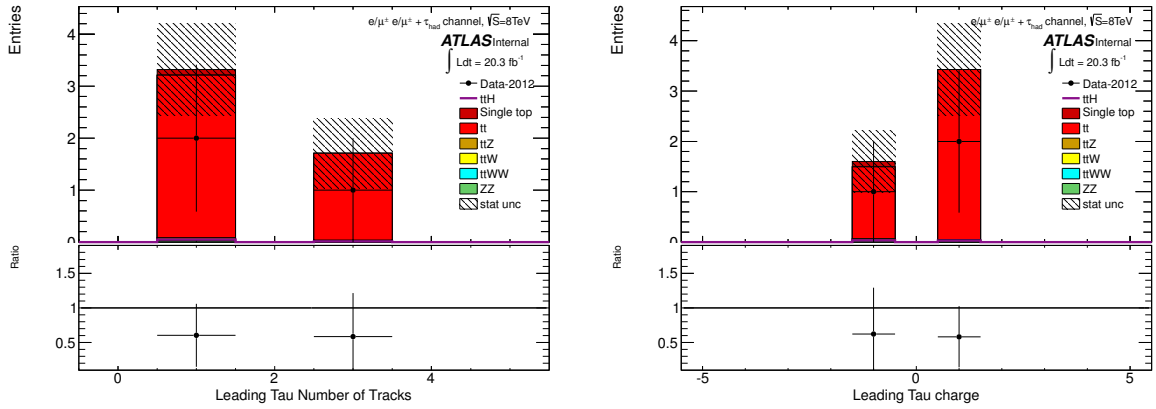


(c) $p_T(\text{lepton}_{\text{subleading}})$ (d) $\eta(\text{lepton}_{\text{subleading}})$

Figure 58: The distributions of lepton kinematic variables in the SS high Njet control region with one loose electron.

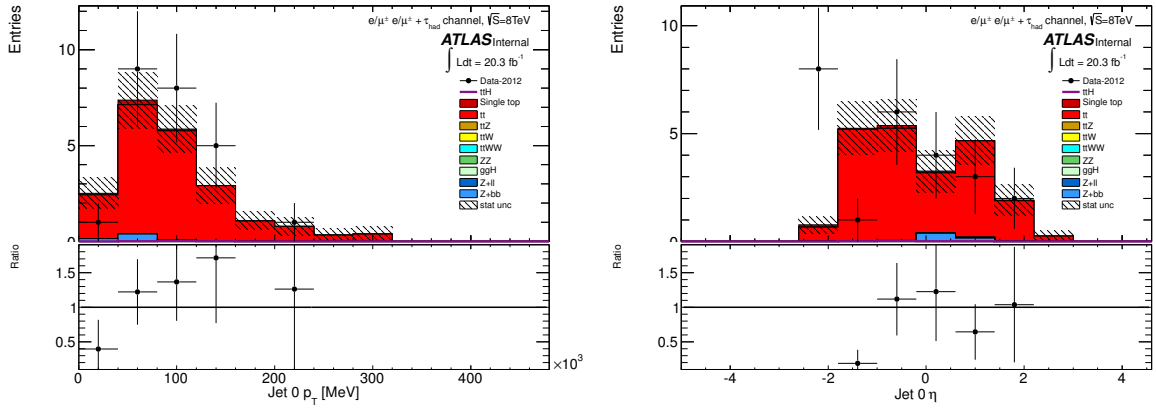


(a) $p_T(\tau)$ (b) $\eta(\tau)$

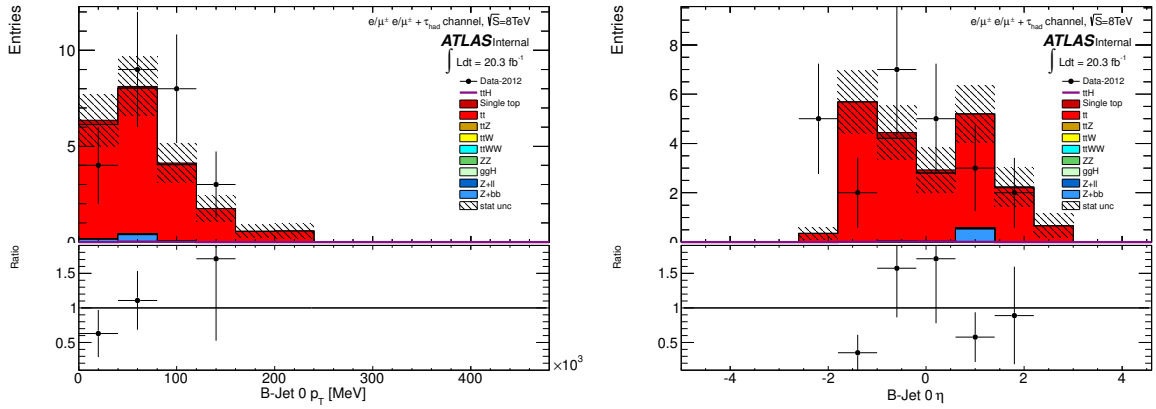


(c) $N_{track}(\tau)$ (d) charge(τ)

Figure 59: The distributions of hadronic tau variables in the SS high Njet control region with one loose electron.

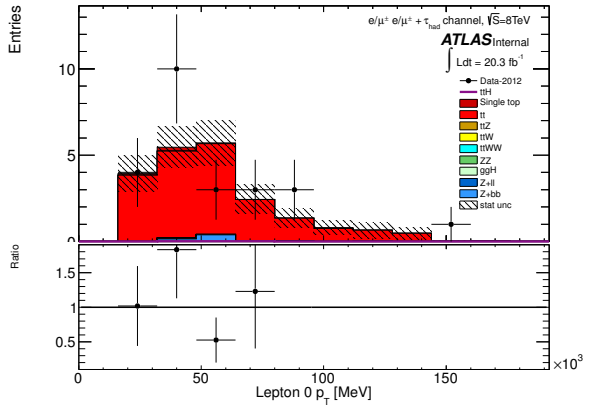


(a) $p_T(\text{jet}_{\text{leading}})$ (b) $\eta(\text{jet}_{\text{leading}})$

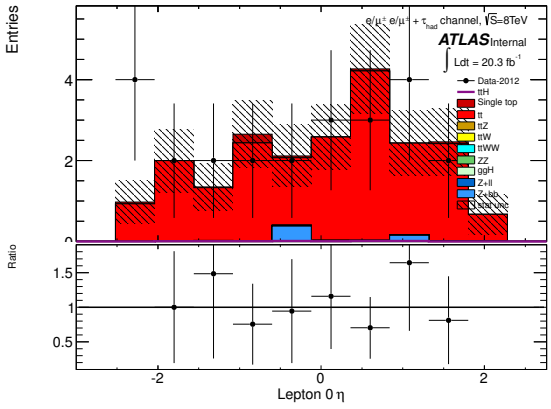


(c) $p_T(\text{Bjet}_{\text{leading}})$ (d) $\eta(\text{Bjet}_{\text{leading}})$

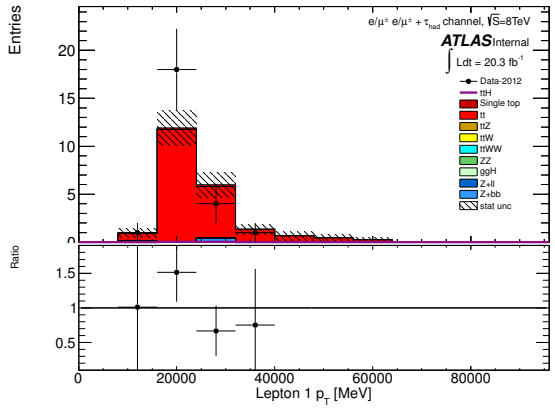
Figure 60: The distributions of jet kinematic variables in the SS low Njet control region with one loose electron.



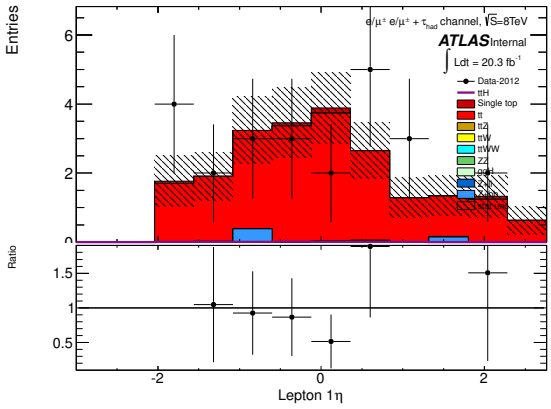
(a) $p_T(\text{lepton}_{\text{leading}})$



(b) $\eta(\text{lepton}_{\text{leading}})$

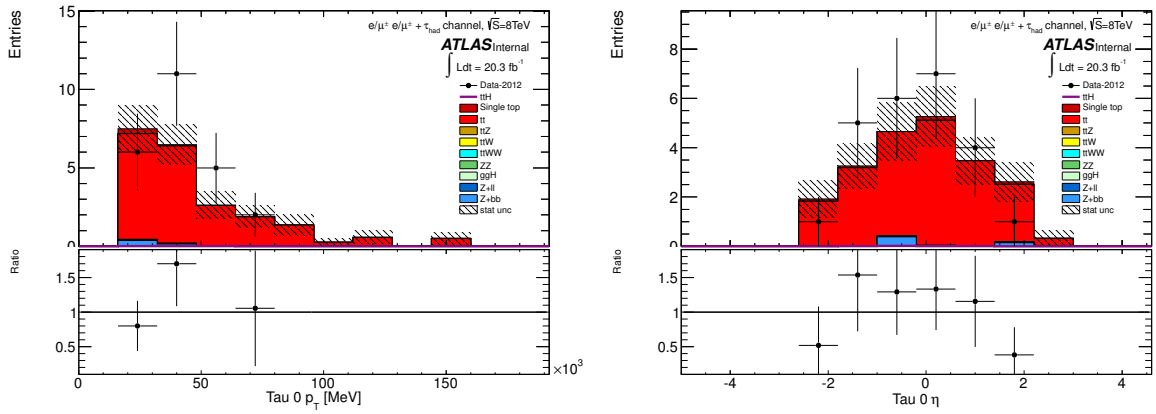


(c) $p_T(\text{lepton}_{\text{subleading}})$

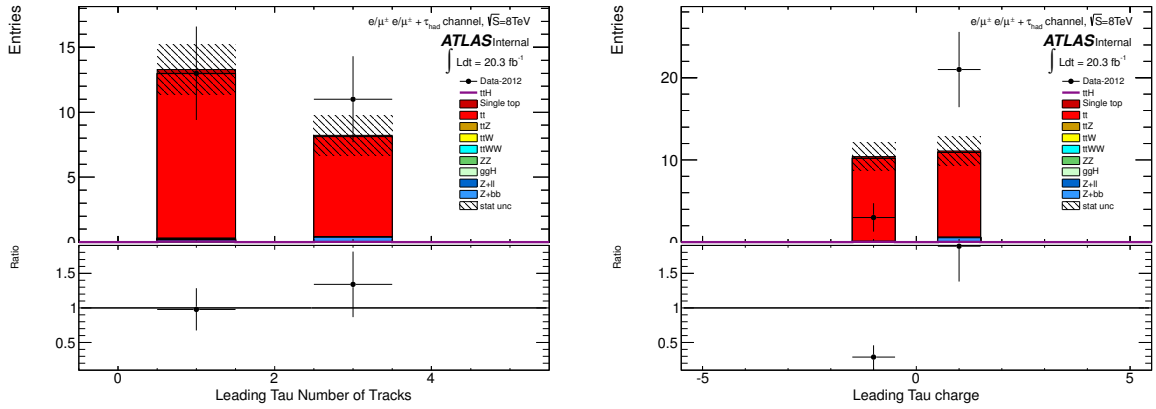


(d) $\eta(\text{lepton}_{\text{subleading}})$

Figure 61: The distributions of lepton kinematic variables in the SS low Njet control region with one loose electron.

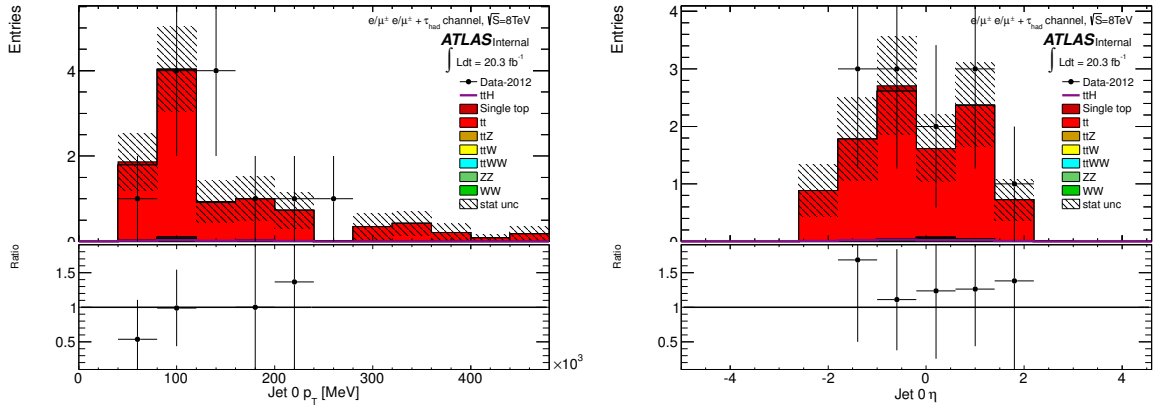


(a) $p_T(\tau)$ (b) $\eta(\tau)$

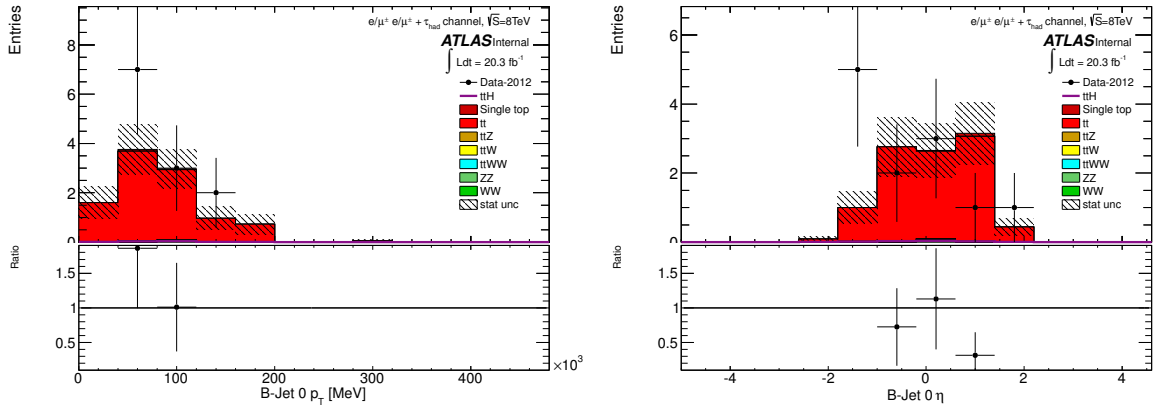


(c) $N_{track}(\tau)$ (d) charge(τ)

Figure 62: The distributions of hadronic tau variables in the SS low Njet control region with one loose electron.

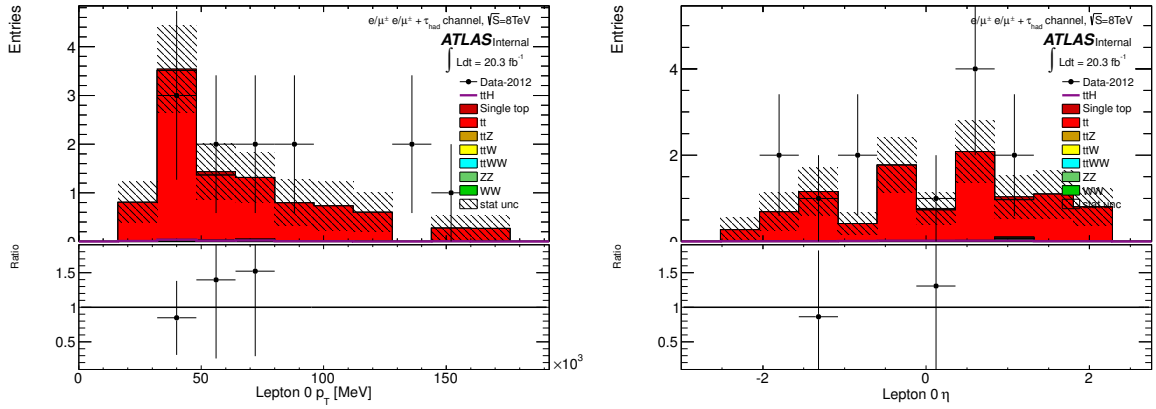


(a) $p_T(\text{jet}_{\text{leading}})$ (b) $\eta(\text{jet}_{\text{leading}})$

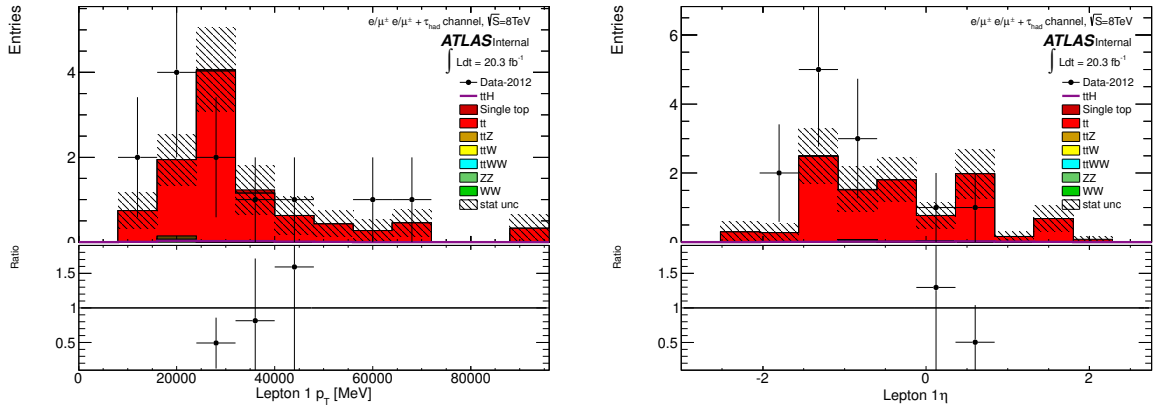


(c) $p_T(\text{Bjet}_{\text{leading}})$ (d) $\eta(\text{Bjet}_{\text{leading}})$

Figure 63: The distributions of jet kinematic variables in the SS high Njet control region with one loose muon.

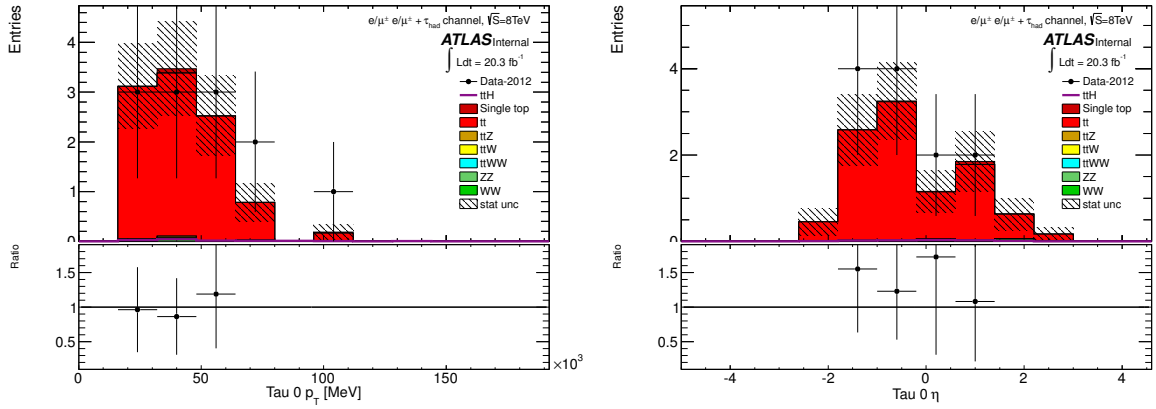


(a) $p_T(\text{lepton}_{\text{leading}})$ (b) $\eta(\text{lepton}_{\text{leading}})$

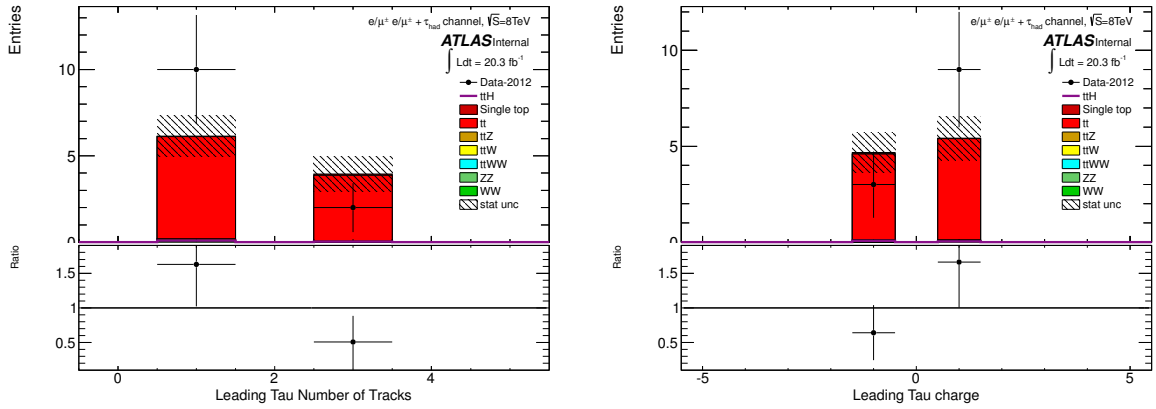


(c) $p_T(\text{lepton}_{\text{subleading}})$ (d) $\eta(\text{lepton}_{\text{subleading}})$

Figure 64: The distributions of lepton kinematic variables in the SS high Njet control region with one loose muon.

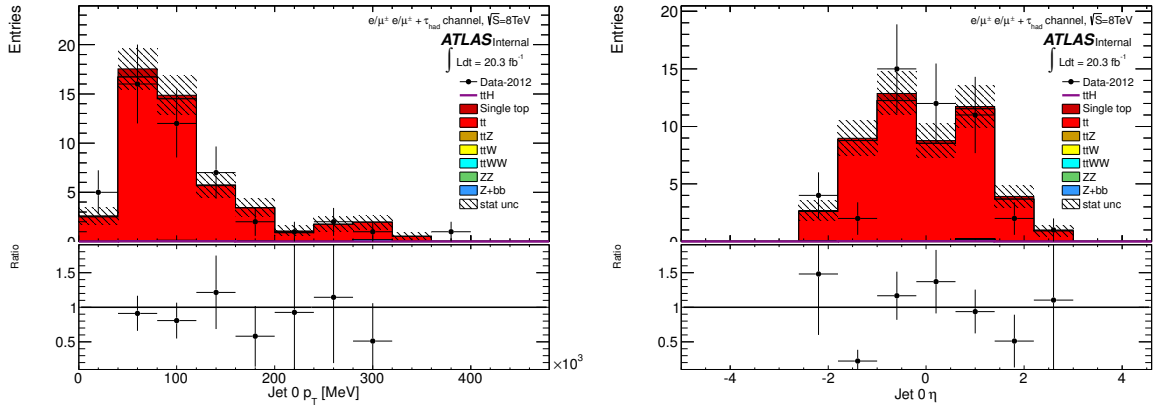


(a) $p_T(\tau)$ (b) $\eta(\tau)$

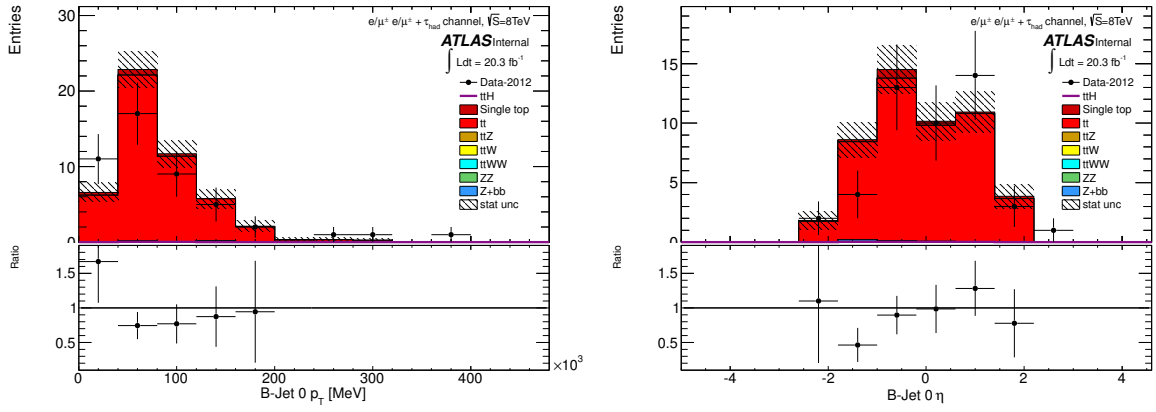


(c) $N_{track}(\tau)$ (d) charge(τ)

Figure 65: The distributions of hadronic tau variables in the SS high Njet control region with one loose muon.

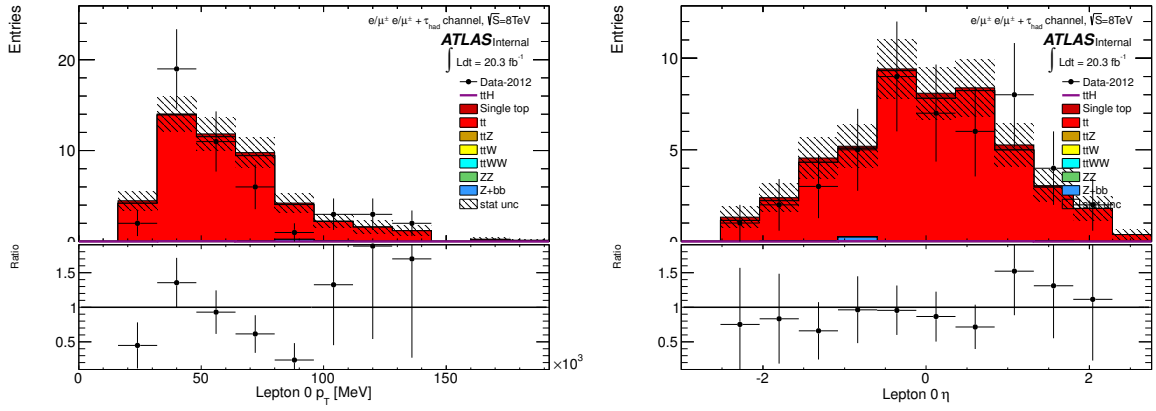


(a) $p_T(\text{jet}_{\text{leading}})$ (b) $\eta(\text{jet}_{\text{leading}})$

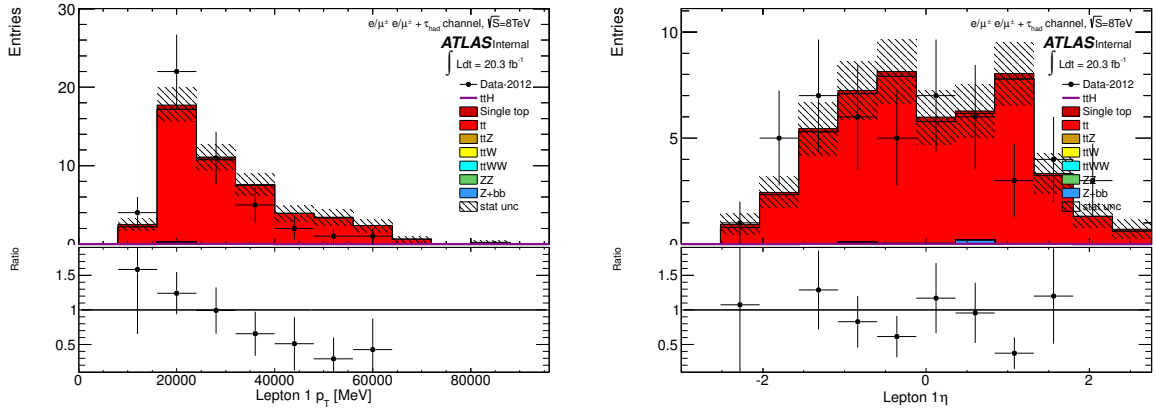


(c) $p_T(\text{Bjet}_{\text{leading}})$ (d) $\eta(\text{Bjet}_{\text{leading}})$

Figure 66: The distributions of jet kinematic variables in the SS low Njet control region with one loose muon.

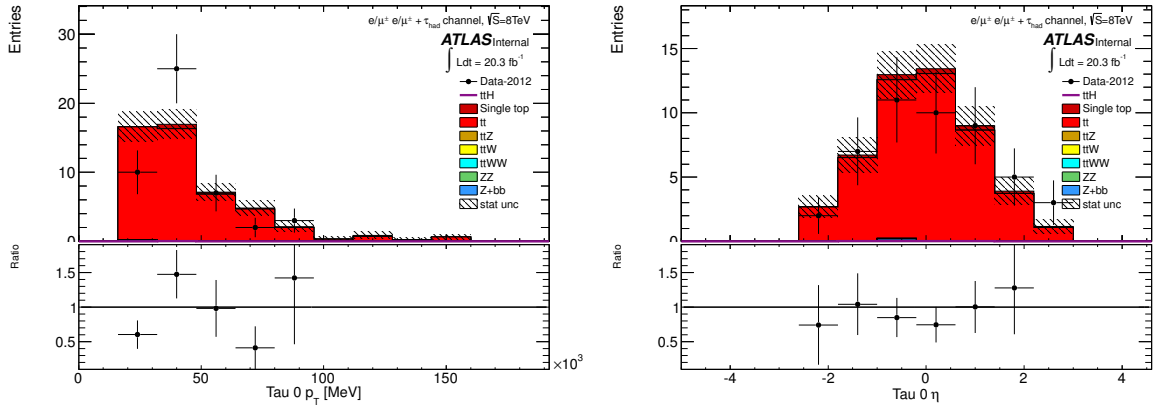


(a) $p_T(\text{lepton}_{\text{leading}})$ (b) $\eta(\text{lepton}_{\text{leading}})$

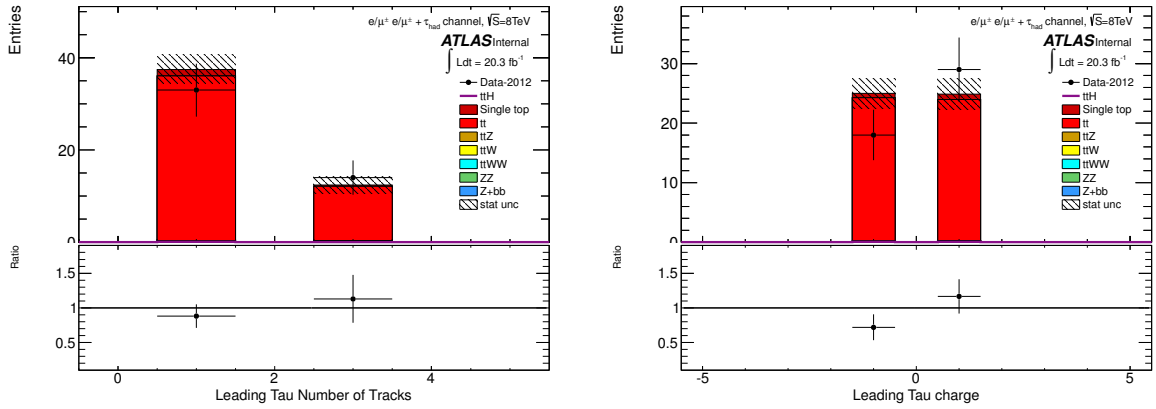


(c) $p_T(\text{lepton}_{\text{subleading}})$ (d) $\eta(\text{lepton}_{\text{subleading}})$

Figure 67: The distributions of lepton kinematic variables in the SS low Njet control region with one loose muon.



(a) $p_T(\tau)$ (b) $\eta(\tau)$



(c) $N_{track}(\tau)$ (d) charge(τ)

Figure 68: The distributions of hadronic tau variables in the SS low Njet control region with one loose muon.

8.3 Estimation of fake backgrounds from Z+jets processes

8.3.1 Estimation strategy

Events from Z+jets processes have a very small contribution in the signal region. A veto of events with two OS electrons having their invariant mass in a window of 10 GeV around the Z-mass is required for the SR to suppress Z→ee events. As the probability for muons to be reconstructed with the opposite charge or to be misidentified as a hadronic tau is much smaller than for electrons, the number of Z+jets events which can leak in the SR is very small. There are only those two possibilities for this to happen:

- Z+jets background where the hadronic tau is faked by an extra jet and one lepton is reconstruced with charge-flip or also faked by an additional jet.
- Z+jets background where one electron is misidentified as a hadronic tau candidate while a second light lepton is faked by an additional jet. The contribution of muons faking hadronic tau candidates can be neglected.

However, with applying the tau electron veto BDT medium, also the contribution of electrons faking hadronic taus can be neglected. Hence, the dominant contribution is the first one with the hadronic tau faked by an extra jet.

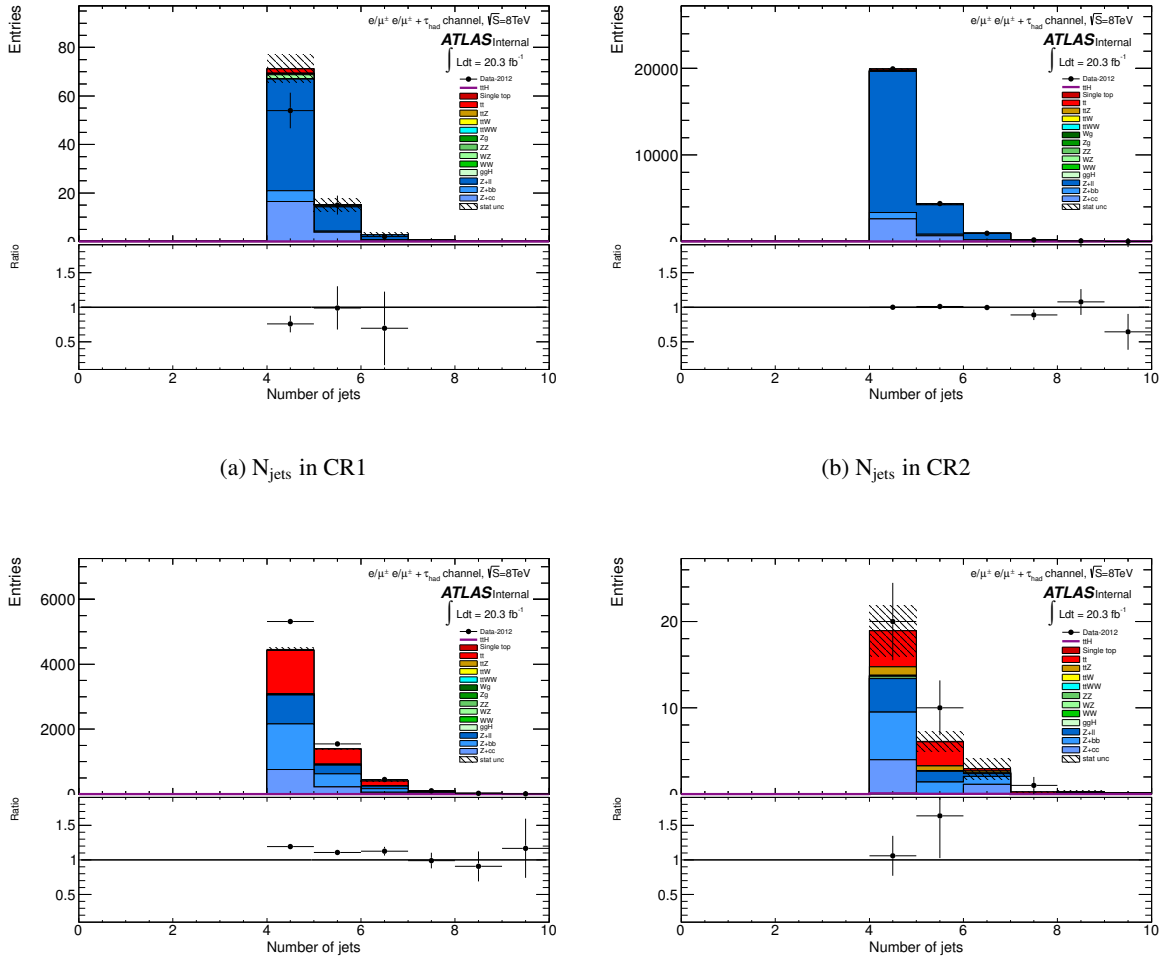
The pure MC estimate leads to zero events after the requirement of the hadronic tau candidate and possible very small remainings in data are covered by the ABCD method. However, to prove that the MC modelling is reliable and with it the subtracted yields in regions B,C and D, it is aimed to prove the normalization of Z+jets processes from data by constructing a Z+jets pure control region. Those CR include requirements of opposite charge of the two leptons, a Z mass requirement instead of the Z mass veto (for dielectron events) and a b-veto. Thus, a high yield of Z+jets events is reached with a simultaneous suppression of top events.

8.3.2 Definition of control regions

The Z+jets background with one additional jet faking the hadronic tau candidate is estimated from MC with correcting possible mis-modeling of fake rates with a Z-pure CR from data. This control region is labelled as CR1, while further CR for cross-checks are labelled as CR2, CR3 and CR4. Their definitions are summarized as follows:

- CR1: OS, $N_{jets} \geq 4$, Z-mass requirement of the two leptons (10 GeV around the Z mass), b-veto, $N_{tau} \geq 1$
- CR2: OS, $N_{jets} \geq 4$, Z-mass requirement of the two leptons (10 GeV around the Z mass), b-veto, no tau requirement
- CR3: OS, $N_{jets} \geq 4$, Z-mass requirement of the two leptons (10 GeV around the Z mass), $N_{b-tag} \geq 1$, no tau requirement
- CR4: OS, $N_{jets} \geq 4$, Z-mass requirement of the two leptons (10 GeV around the Z mass), $N_{b-tag} \geq 1$, $N_{tau} \geq 1$

The distributions of N_{jet} of the four CR can be seen in Figure 69. The Z+jets normalization is obtained from CR1 which includes the tau requirements as in the signal region. CR2 gives the normalization with the same requirements but without the hadronic tau identification. The regions CR3 and CR4 give those distributions but with at least one b-tag instead of b-veto as in the signal ragion for cross-checks. It



(a) N_{jets} in CR1

(b) N_{jets} in CR2

(c) N_{jets} in CR3

(d) N_{jets} in CR4

Figure 69: The distributions of the N_{jets} variable for CR1 (a), CR2 (b), CR3 (c) and CR4 (d).

has to be noted that the very good agreement without tau requirement in CR2 changes to a overestimation of MC when requiring at least one identified hadronic tau. This indicates an well known overestimation of the jet to tau fake-rate in MC.

8.3.3 Results:

From CR1 a MC scale factor of $k_Z = 0.63 \pm 0.09$ is obtained. The normalization factor and its uncertainty have no relevant influence on the very small MC yields. Possible small contributions from Z+jets as well as W+jets, which are hardly to estimate from MC are treated to be covered by the data-driven method, as they are not subtracted from the data yields in the control regions.

8.3.4 Cross-checks

To check that the normalization is still conserved when requiring a b-tag (as in the SR), the CR3 and CR4, as shown on Subfigures (c) and (d) of Figure 69, are used. That leads to less pure CR with a

significant top contribution, but within the uncertainties, the overall normalization is still conserved.

8.4 Remark on event selection optimization

In order to optimize the object- and event selection as described in Note 1 [1], pure MC based event yields are used in the optimization procedure. Hence, the numbers for the background estimation obtained here, especially top-induced backgrounds, are compared with the MC ones in Table 27.

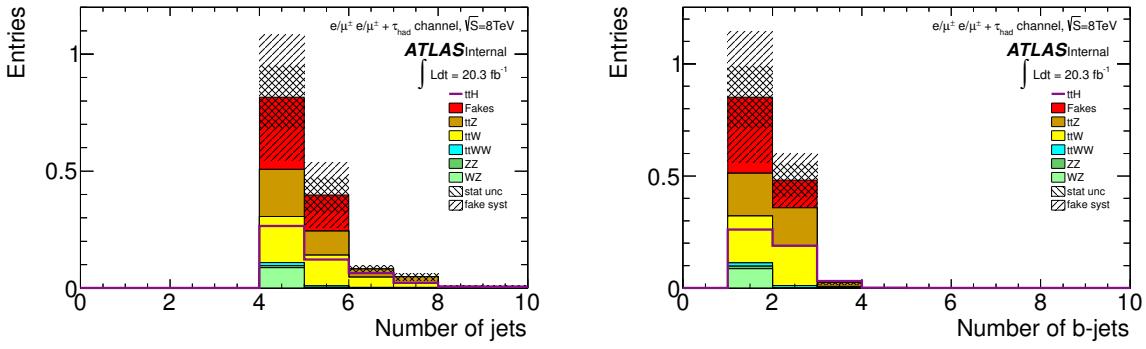
Table 27: Resulting event numbers used in optimization procedure of Note 1 and obtained here for the top-related (fake) background sources.

Background process	MC only yield	Obtained in fake estimation
$t\bar{t}$/ single-top / 4-top	0.81±0.44	$0.46^{+0.69}_{-0.53}$

It can be seen that the data-driven estimate for top related fake-backgrounds yields to a lower yield of top related fake-backgrounds due to the lower observed data yields in regions B and C compared to MC. However, the data-driven estimate and the MC only one are in agreement within their uncertainties. The upper limit on the top fake background is reduced with the data-driven estimate.

8.5 Signal region distributions of estimated fake events

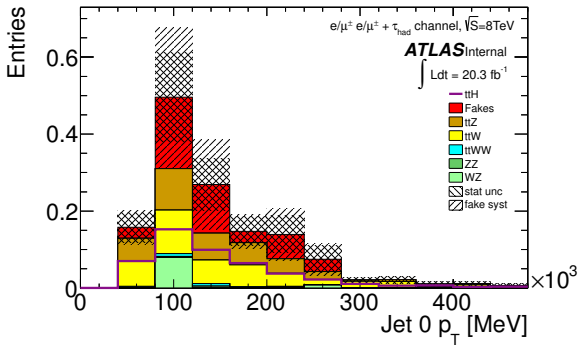
Here, distributions of relevant object and event variables are shown for the signal region including fake background estimation in Figures 70-73. This is done by applying the data-driven ABCD method to each bin of the distributions. However, due to lack of statistics, especially in CR B, not the full shapes of all three sideband regions are used. Instead, the integrated ratio of regions B and D are used while the shapes are taken from region C only. The shaded errors include the statistical uncertainty of MC as well as of the fake estimate and the systematic uncertainty of the fake estimate as described above.



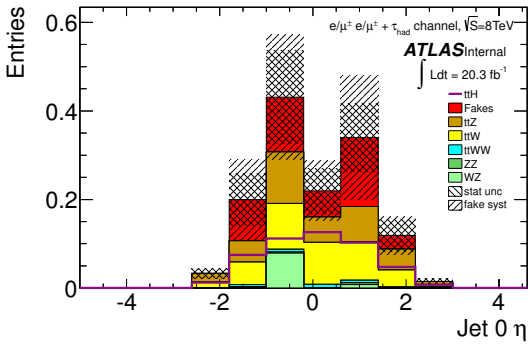
(a) N_{jets} (b) N_{B-jets}

(c) N_{lepton} (d) $N_{electron}$

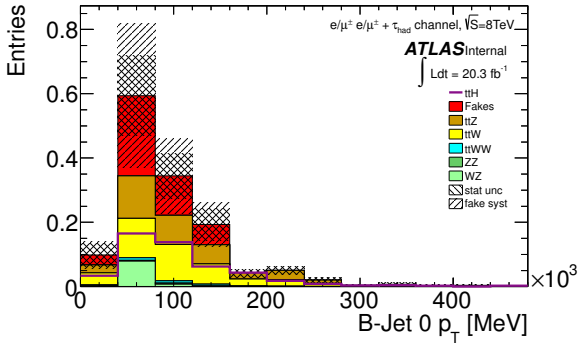
Figure 70: The distributions of jet- and lepton multiplicity variables in the signal region.



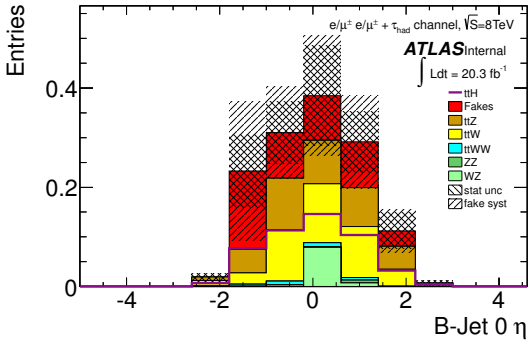
(a) $p_T(\text{jet}_{\text{leading}})$



(b) $\eta(\text{jet}_{\text{leading}})$

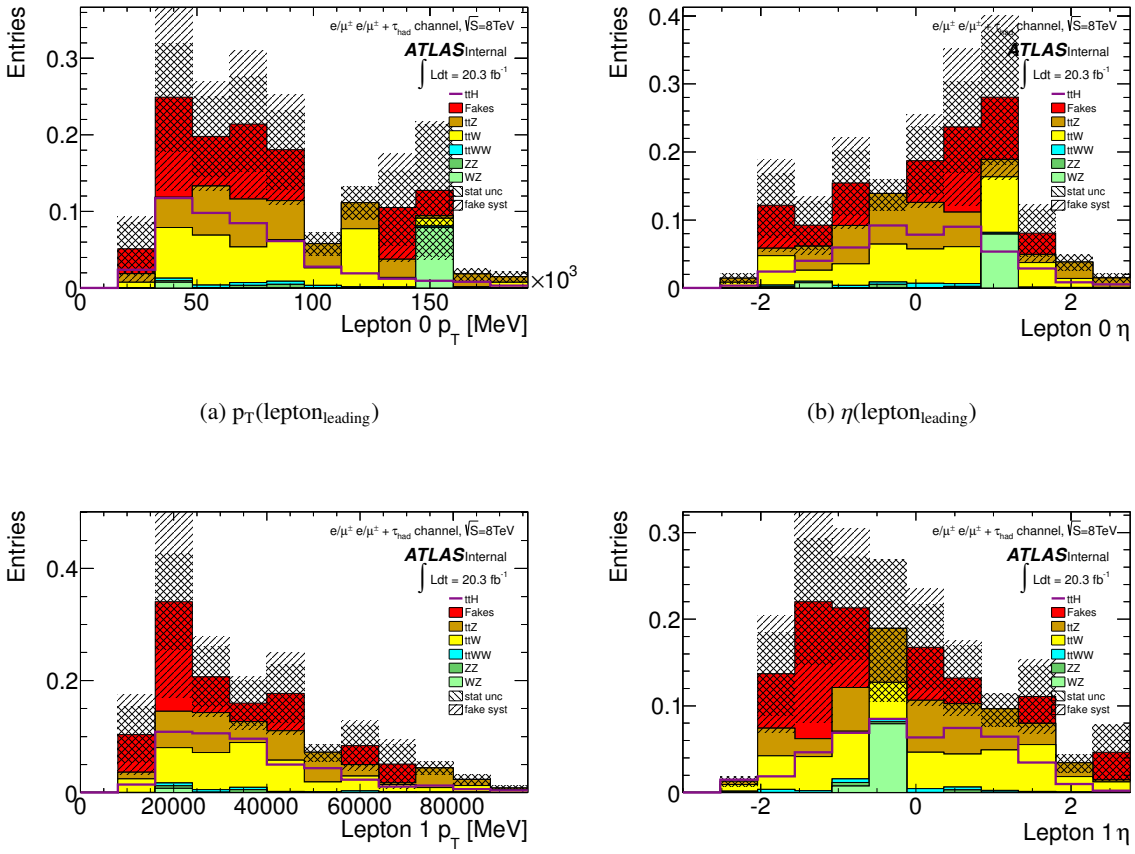


(c) $p_T(\text{Bjet}_{\text{leading}})$



(d) $\eta(\text{Bjet}_{\text{leading}})$

Figure 71: The distributions of jet kinematic variables in the signal region.



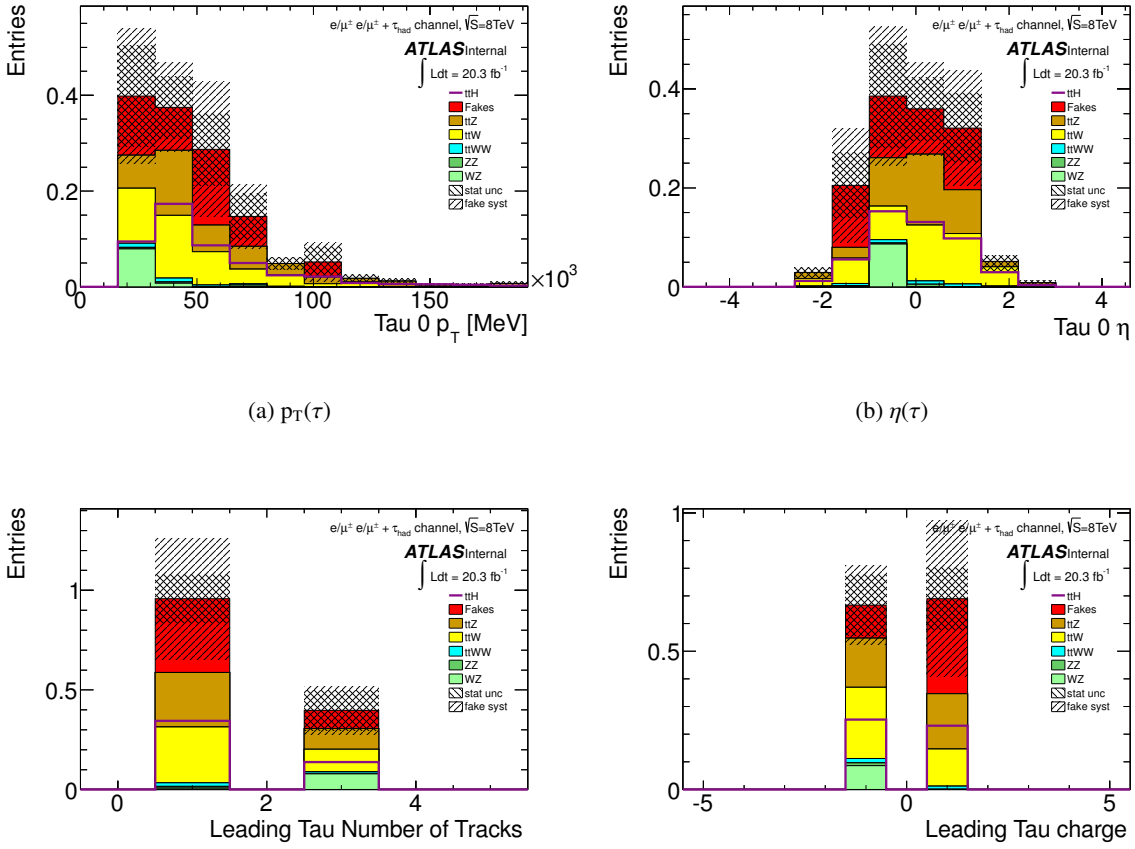
(a) $p_T(\text{lepton}_{\text{leading}})$

(b) $\eta(\text{lepton}_{\text{leading}})$

(c) $p_T(\text{lepton}_{\text{subleading}})$

(d) $\eta(\text{lepton}_{\text{subleading}})$

Figure 72: The distributions of lepton kinematic variables in the signal region.



(a) $p_T(\tau)$

(b) $\eta(\tau)$

(c) $N_{track}(\tau)$

(d) charge(τ)

Figure 73: The distributions of hadronic tau variables in the signal region.

9 4 lepton channel

9.1 Estimate of Z+jets and $t\bar{t}$ background contributions from secondary or fake leptons

Due to the high lepton multiplicity in the four lepton channel and the tight lepton requirements which were optimized and outlined in Note 1, the expected contribution of background from fake leptons to the four lepton signal region is very small. A thorough in-situ method of estimating this background solely using four lepton events is difficult because of the low statistics involved. The strategy for estimating this contribution thus relies upon input from the three lepton analysis, in particular the extrapolation factors for estimating yields in the signal region, as well as a dedicated extrapolation factor derived from two lepton auxiliary regions. The procedure for this estimate as well as the various cross-checks and uncertainty evaluations will be outlined in this section.

9.1.1 Estimation strategy

The data-driven procedure for the fake estimate used in the four lepton analysis proceeds as a two-part data extrapolation. First, control regions enriched in $t\bar{t}$ and Z+jets are constructed with 2 tight leptons (those passing all selection criteria outline in Note 1) plus 2 *anti-tight* leptons (which explicitly fail one or a set of selection criteria). It is important to note that these CRs are orthogonal to the signal region due to the explicit failure of selection criteria of one or more of the leptons. Identical event-level selections to the signal region are applied, up to and including the Z-veto (i.e. jet-inclusive).

From these CRs, which are mainly dominated by $t\bar{t}$ with small contributions from single $t + X$ and Z+jets events, the MC estimate from prompt backgrounds (primarily ZZ and $t\bar{t}Z$) is subtracted from the data yield and the result is taken as the fake contribution. This fake contribution in each CR is extrapolated into the jet-inclusive signal region (comprised of 4 tight leptons plus identical cuts to the CRs) by the application of extrapolation factors related to the anti-tight leptons passing the tight lepton selection criteria. The extrapolation factors, different for electrons and muons, are those used for the three lepton $t\bar{t}$ background estimate and are derived from three lepton control regions. A description of this extrapolation factor derivation is included in section 6 of this note. Truth information studies and closure tests on $t\bar{t}$ MC were performed in order to verify the transferability of the three lepton factors to the four lepton procedure, and an associated non-closure systematic uncertainty is included in the final fake estimate.

Finally, a second extrapolation is done from the jet-inclusive SR to the full selection by a separate extrapolation factor which is derived from the jet spectra of two lepton events with at least two b-tagged jets and represents the efficiency of $t\bar{t}$ and Z+jets events passing the SR N_{jet} requirement. A detailed description of the motivation, results and validation for this entire procedure is included in this section.

9.1.2 Lepton and control region definitions

In order to simplify the use of the three lepton-derived extrapolation factors for fake leptons, the anti-tight lepton definitions are common between both the three and four lepton channels. For both electrons and muons, reversed or relaxed isolation requirements are used to define the objects while for electrons the particle identification requirement is also loosened.

Anti-tight muons are required to have a relative calorimeter isolation of $E_T\text{cone20}/p_T \geq 0.1$ and a relative track isolation of $p_T\text{cone20}/p_T < 1.0$ while passing all other tight muon selections. Anti-tight electrons are similarly required to have $E_T\text{cone20}/p_T \geq 0.05$ and $p_T\text{cone20}/p_T < 1.0$ and must pass the VeryLooseLH PID working point along with all other tight electron selections.

For this estimate procedure, three control regions are constructed using these anti-tight lepton definitions, as follows:

- CR1: 2 tight leptons + 2 anti-tight electrons
- CR2: 2 tight leptons + 2 anti-tight muons
- CR3: 2 tight leptons + 1 anti-tight electron + 1 anti-tight muon

All signal region event-level selections are applied up to and including the Z-veto. N_{jet} and N_{b-jet} requirements are not applied. Figure 74 displays the jet and electron multiplicities for each of the three CRs in order to illustrate the background prediction composition as well as the Data/MC agreement.

9.1.3 Jet-inclusive SR extrapolation

The data yields along with the MC predictions for each fake estimate control region are listed in Table 28. The fake contribution is taken as the data in each region less the MC prediction for prompt 4ℓ events. It can be seen that these regions are largely dominated by $t\bar{t}$ with some contribution from Z+jets . Both contributions are treated together in a single data-driven estimate.

Table 28: MC predictions and data yields in each 2T+2L control region. Selections include all signal region cuts up to and including the Z-veto, jet-inclusive. The prompt MC prediction is dominated by diboson production and also includes small contributions from the $t\bar{t}H$ signal, ggF $H \to 4\ell$, $t\bar{t}V$ and tZ backgrounds.

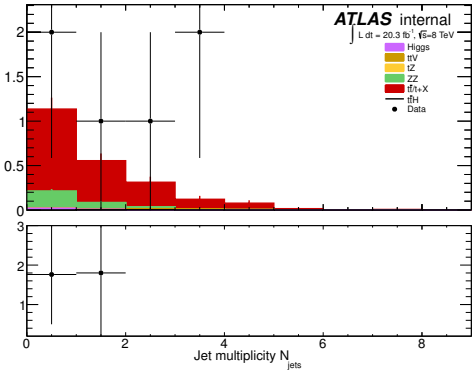
CR	prompt	WZ	$t\bar{t}/t + X$	Z+jet	Sum Bkg.	Data
2T+2L(ee)	0.52 ± 0.27	0.00 ± 0.00	1.87 ± 0.17	0.00 ± 0.00	2.39 ± 0.21	7
2T+2L($\mu\mu$)	0.25 ± 0.25	0.00 ± 0.00	6.62 ± 0.31	0.64 ± 0.49	7.45 ± 0.58	7
2T+2L($e\mu$)	0.57 ± 0.57	0.00 ± 0.00	8.23 ± 0.38	0.30 ± 0.30	9.03 ± 0.49	15

In each of these CRs, a transfer factor is applied to extrapolate each estimate from the 2 tight leptons + 2 anti-tight leptons selection to the jet-inclusive SR with 4 tight leptons. The transfer factor is taken as the appropriate product of θ factors from the three lepton analysis associated with the flavour of the loose leptons. For example, in order to extrapolate the 2T+2L(ee) value into the jet-inclusive SR, the CR value is scaled by θ_e^2 where θ_e is the single loose electron extrapolation factor derived with three lepton events. Table 29 list the results of this first extrapolation from the 2T+2L CRs to the jet inclusive SR.

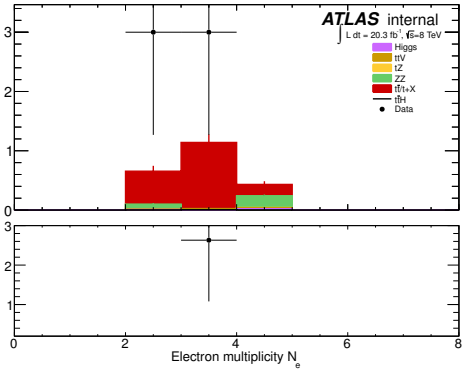
Table 29: Results for the first data extrapolation from 2T+2L lepton control regions to 4T lepton SR, jet-inclusive.

Control region	2T+2L(ee)	2T+2L($\mu\mu$)	2T+2L($e\mu$)
Data	7	7	15
Prompt MC	0.52 ± 0.27	0.25 ± 0.25	0.57 ± 0.57
Fake contribution, data	6.5 ± 0.3	6.8 ± 0.3	14.4 ± 0.6
θ-factor (3ℓ, MC)	$\theta_e^2 = (0.58 \pm 0.17) \cdot 10^{-3}$	$\theta_\mu^2 = (0.022 \pm 0.007) \cdot 10^{-3}$	$\theta_e\theta_\mu = (0.13 \pm 0.04) \cdot 10^{-3}$
SR extrapolation, jet-inclusive	0.0037 ± 0.0011	0.00015 ± 0.00005	0.0016 ± 0.0005

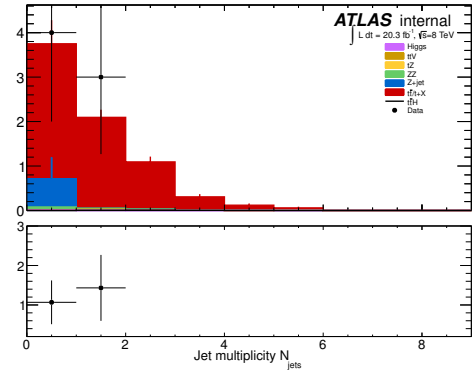
In total the data estimate from fakes at this stage is 0.0055 ± 0.0030(stat.). The second extrapolation takes this value into the SR selection with an additional requirement on N_{jets}.



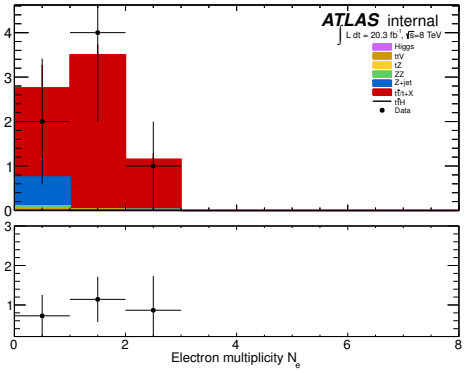
(a) Jet multiplicity, CR1 (2T+2L(ee))



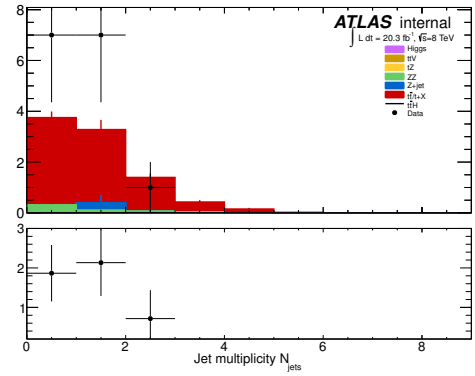
(b) Electron multiplcity, CR1 (2T+2L(ee))



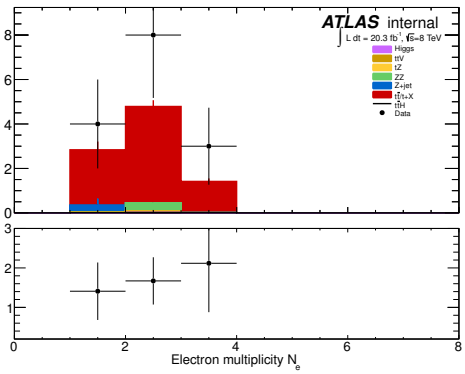
(c) Jet multiplicity, CR2 (2T+2L($\mu\mu$))



(d) Electron multiplcity, CR2 (2T+2L($\mu\mu$))



(e) Jet multiplicity, CR3 (2T+2L($e\mu$))



(f) Electron multiplcity, CR3 (2T+2L($e\mu$))

Figure 74: Jet and electron multiplicities for the CRs used for the first extrapolation in leptons

9.1.4 Motivation and derivation of jet-selection extrapolation factor ϵ_{jets}

It was found that the majority of fake leptons (of either flavour) originate from a b-hadron. In the case of $t\bar{t}$ then, the events which pass the four lepton SR selection are those with 2 tight leptons, 2 fake leptons (from lepton-faking b-jets) and at least 2 additional jets, where at least one of these jets is b-tagged. This topology comes from $t\bar{t}$ +b or $t\bar{t}$ +c/lq where the c- or light-quark jet is mistagged as heavy-flavour. The event topology for the Z+jets contribution in the SR is similar, with additional considerations taken for the small number of fake leptons originating from light-flavour jets.

With this in mind, a second extrapolation factor is derived from auxiliary regions with exactly 2 tight leptons and at least 2 b-jets. The two auxiliary regions (AR) constructed are:

- AR1: 2 tight leptons + at least 2 b-jets + at least 4 total jets
- AR2: 2 tight leptons + at least 2 b-jets, with any number of other jets

Recall that the jet multiplicity selections in the four lepton signal region are requirements of at least 2 jets with at least 1 b-tagged jet. AR1 then corresponds to the 2T lepton events with sufficient jet multiplicity to provide 2 additional fake leptons as well as to pass the N_{jet} requirement of the SR. Conversely, AR2 corresponds to 2T lepton events which have a sufficient number of jets to provide a four lepton signature from the 2 additional fake leptons however would not pass the SR jet multiplicity requirement. The ratio of these two regions, AR1/AR2, thus gives a measure of the efficiency of the jet multiplicity requirement of the SR on the contributions from fake leptons. This ratio, ϵ_{jets} can be applied as an extrapolation factor to the data-driven fake estimate in the jet-inclusive SR which was computed in the previous section to yield the total estimate for fake background from $t\bar{t}$ and Z+jets in the SR.

Table 30: MC prediction and data yields in the 2T leptons auxiliary regions used for the derivation of the jet-selection efficiency extrapolation factor ϵ_{jets}. The MC prediction for non-$t\bar{t}$ or Z+jets is subtracted from the data and the remaining data estimate is used in the calculation of ϵ_{jets}

AR	Non-fake	$t\bar{t}/t + X$	Z+jet	Sum Bkg.	Data
AR1: $2\ell, \geq 4j(\geq 2b)$	12.0 ± 0.4	5239.58 ± 8.86	193.24 ± 13.59	5444.88 ± 16.24	5546
AR2: $2\ell, \geq 2j(\geq 2b)$	154.9 ± 1.6	25879.41 ± 19.81	880.24 ± 24.55	26882.15 ± 31.59	26603

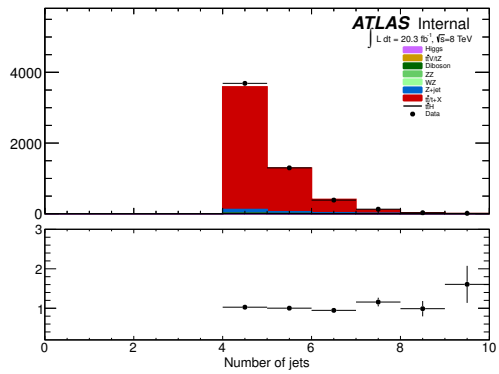
The MC predictions and data yields for the 2T ARs are listed in Table 30. Figure 75 shows the jet and b-tagged jet multiplicities for each of these ARs. To compute the extrapolation factor, the data yield in each region is taken and the MC estimates for the backgrounds not associated with the extrapolation are removed. The factor is then computed as

$$\epsilon_{\text{jets}} = \frac{(\text{Data} - \text{non-fake MC})_{\text{AR1}}}{(\text{Data} - \text{non-fake MC})_{\text{AR2}}} \tag{34}$$

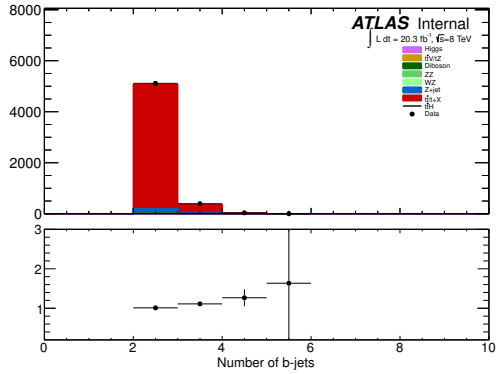
The jet-selection efficiency factor is found to be: $\epsilon_{\text{jets}} = \frac{(5534.0 \pm 0.4)}{(26448.0 \pm 1.6)} = 0.20 \pm 0.02$.

9.1.5 Results

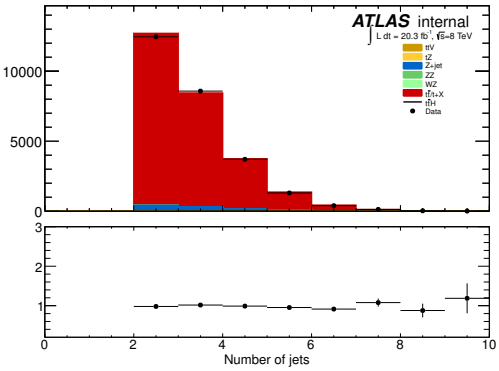
The total results of the full double-extrapolation for the estimate of $t\bar{t}$ and Z+jets background contributions are listed in Table 9.1.5 along with the extrapolation factors used at each stage. Results are shown for the inclusive 4-lepton signal region as well as the two signal region categories used in the final result.



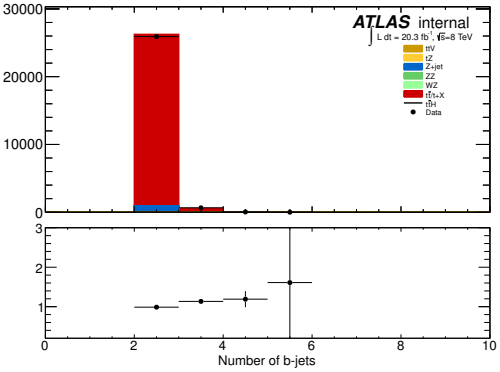
(a) Jet multiplicity, AR1



(b) b-Tagged jet multiplcity, AR1



(c) Jet multiplicity, AR2



(d) b-Tagged jet multiplcity, AR2

Figure 75: Jet and b-tagging jet multiplicities for the ARs used in the derivation of the jet-selection efficiency extrapolation factor ϵ_{jets}

Table 31: Full results for the double-extrapolation used to estimate the fake background contribution from $t\bar{t}$ and Z+jets events in the four lepton SR from data. The data-driven estimate is shown at each stage of the extrapolation procedure, from the 2T+2L CRs to the full SR selection estimate

Extrapolation stage	Inclusive SR	Z-depleted SR	Z-enriched SR
CR1 $\overset{\theta_e^2}{\Longrightarrow}$ SR, jet-inclusive	0.0037 ± 0.0011	0	0.0037 ± 0.0011
CR2 $\overset{\theta_\mu^2}{\Longrightarrow}$ SR, jet-inclusive	0.00015 ± 0.00005	0	0.00015 ± 0.00005
CR3 $\overset{\theta_e\theta_\mu}{\Longrightarrow}$ SR, jet-inclusive	0.0016 ± 0.0005	$(0.45 \pm 0.14) \cdot 10^{-3}$	0.0012 ± 0.0004
Total SR, jet-inclusive	0.0055 ± 0.0030	$(0.45 \pm 0.14) \cdot 10^{-3}$	0.0050 ± 0.0027
SR, jet-inclusive $\overset{\epsilon_{\text{jets}}}{\Longrightarrow}$ SR	$(1.1 \pm 0.6) \cdot 10^{-3}$	$(0.09 \pm 0.03) \cdot 10^{-3}$	$(1.0 \pm 0.5) \cdot 10^{-3}$

9.1.6 Closure tests and uncertainties

In order to test the data-derived extrapolation factors for both stages of the extrapolation, a set of closure tests was performed. Using an AFII $t\bar{t}$MC sample, the agreement between the extrapolation estimate and the MC prediction is compared for both the θ lepton extrapolation factors (from 3ℓ) and the ϵ_{jets} jet efficiency factor. Table 9.1.6 shows the results for each θ-factor in their respective 2T+2L CRs, while Table 9.1.6 shows the results for the ϵ_{jets} closure.

	2T+2L(ee)	2T+2L($\mu\mu$)	2T+2L($e\mu$)
MC after Z-veto	1.807 ± 0.157	6.436 ± 0.294	7.714 ± 0.323
θ_ℓ	$\theta_e^2 = (0.069223)^2$	$\theta_\mu^2 = (0.071962)^2$	$\theta_e\theta_\mu$
SR extrapolation, jet inclusive	0.0087 ± 0.0008	0.033 ± 0.002	0.038 ± 0.002
	↘	↓	↙
Total SR extrapolation, jet inclusive		0.0797 ± 0.0029	
MC prediction in SR, jet inclusive		0.08 ± 0.03	

	SR (4T)	2T+2L(ee)	2T+2L($\mu\mu$)	2T+2L($e\mu$)
MC after Z-veto	0.089 ± 0.034	1.807 ± 0.157	6.436 ± 0.294	7.714 ± 0.323
ϵ_{jets}	(0.20 ± 0.02)	–	–	–
Jet extrapolation	0.018 ± 0.007	0.374 ± 0.033	1.331 ± 0.064	1.595 ± 0.071
c.f. MC after jet selection	0.041 ± 0.024	0.463 ± 0.080	1.489 ± 0.141	1.784 ± 0.155

It can be seen that the θ lepton extrapolation factors close very well. This result verifies the usability of the 3ℓ-derived lepton extrapolation factors in the four lepton extrapolation procedure. Additionally, there is good agreement in the closure results for the jet efficiency extrapolation. This procedure allows an upper limit to be set on the contribution from fake lepton backgrounds in the 4-lepton signal region.

10 1 lepton and 2 τs

10.1 Strategy to estimate fake τ's

The 1 lepton + 2 hadronic τs channel is targeting the $ttH \rightarrow \tau_{\text{had}}\tau_{\text{had}}$, where the pair of top quarks decays semi-leptonically. The final event selection therefore considers a pair of reconstructed hadronic τs (which will be denoted in this chapter as τ_{had}) and one isolated lepton from the leptonic top decay. The reconstruction of τ_{had} is based on the MVA approach [19]. Since the reconstruction of τ_{had} is based on several discriminating variables, a data-driven estimation of the τ_{had} fake rates is essential for reducing potential systematic uncertainties. Uncertainties arise from imperfect modelling of the detector response to τ-like jets and due to the fact that a large fraction of (pile-up) jets are reconstructed as τ_{had}. In this chapter we will describe a data-driven method for the estimation of the fake τ_{had} contribution to our signal region based on the matrix-method approach. In section 10.1.1 we will describe the Matrix-method approach used in this channel, section 10.1.3 describes the procedure of estimating τ_{had} fake rates and motivation for parametrization of the related efficiencies, in section 10.1.4 we will show results on the closure tests with MC and extraction of appropriate normalization factors and systematic uncertainties; and in section 10.1.5 we will show several control plots of various kinematic variables used in the event selection of this channel.

10.1.1 Matrix-method

The general idea of the Matrix-Method (MM) approach [20] is to estimate the contamination of fake objects in the signal region. The method is based on the measured efficiencies for both real and fake objects, which are in our case the hadronicaly reconstructed τs (τ_{had}). First, we consider two different selection criteria of reconstructed hadronic τs - *Loose* and *Medium* tau ID. The signal region (SR) is defined as containing events with pairs of τ_{had} that pass *Medium* ID selection. Then, using the matrix-method, we can estimate the number of fake events in our SR based on the efficiencies measured in control regions (CR). In the control region at least one τ_{had} fails the *Medium* ID, this region is therefore orthogonal to our SR. In the following text we will describe the method first with one reconstructed object, and then describe the method with two reconstructed objects emphasizing the simplifications used with this method.

Using a similar notation as in [20], we denote by N^t the number of reconstructed *Medium* τ_{had} and by N^l the number of reconstructed *Loose* τ_{had}. To illustrate the method, we first consider the simple case of events with a single reconstructed hadronic τ. We can express the number of reconstructed *Medium* and *Loose* τs in the following way:

$$N^t = N^r_{pass} + N^f_{pass} \tag{35}$$

$$N^l = N^r + N^f, \tag{36}$$

where $N^r_{pass}(N^f_{pass})$ and $N^r(N^f)$ are real(fake) reconstructed τ_{had} that pass *Medium* and *Loose* selection ID, respectively.

Denoting $r = \frac{N^r_{pass}}{N^r}$ to be the efficiency of real τ to pass *Medium* selection and $f = \frac{N^f_{pass}}{N^f}$ to be the efficiency of a fake τ to pass the *Medium* selection, we can express the number of reconstructed events that pass *Loose* or *Medium* τ ID using the efficiency matrix:

$$\begin{pmatrix} N^t \\ N^l \end{pmatrix} = \begin{pmatrix} r & f \\ 1 & 1 \end{pmatrix} \begin{pmatrix} N^r \\ N^f \end{pmatrix} \tag{37}$$

By inverting the efficiency matrix, the number of fakes passing our *Medium* selection can be expressed as a function of the measured quantities in the following way:

$$N^t_{fakes} = fN^f = \frac{f}{r-f}\left(rN^l - N^t\right) = \frac{f}{r-f}\left(rN^j - (1-r)N^t\right), \tag{38}$$

where $N^j = N^l - N^t$ is the number of reconstructed *Loose* τ_{had} that fail the *Medium* selection criteria. In case of no real τ contamination in this control region ("anti-τ" control region - N^j) from the selection or by subtraction, ie. if $r = 1$, then Equation 38 reduces to a simpler form:

$$N^t_f = FF \times N^j, \tag{39}$$

where $FF = \frac{f}{1-f}$ is called the fake factor. This approach is commonly called the "fake-factor" method [21].

In case of two τ candidates, we define four orthogonal regions with τs that pass (t) or fail (j) *Medium* ID selection. Then one can use a 4×4 efficiency matrix to obtain the number of reconstructed events:

$$\begin{pmatrix} N^{tt} \\ N^{tj} \\ N^{jt} \\ N^{jj} \end{pmatrix} = \begin{pmatrix} r_1 r_2 & r_1 f_2 & f_1 r_2 & f_1 f_2 \\ r_1(1-r_2) & r_1(1-f_2) & f_1(1-r_2) & f_1(1-f_2) \\ (1-r_1)r_2 & (1-r_1)f_2 & (1-f_1)r_2 & (1-f_1)f_2 \\ (1-r_1)(1-r_2) & (1-r_1)(1-f_2) & (1-f_1)(1-r_2) & (1-f_1)(1-f_2) \end{pmatrix} \begin{pmatrix} N^{rr} \\ N^{rf} \\ N^{fr} \\ N^{ff} \end{pmatrix}, \tag{40}$$

where the first (second) index is referred to a leading (sub-leading) reconstructed τ_{had}. Again, by inverting the efficiency matrix, one can express the number of entries originating from events with at least one non-prompt τ_{had} in the following way:

$$N^{tt}_{fakes} = \left(1 - \frac{\alpha_1}{FF_1}\frac{\alpha_2}{FF_2}\right)N^{tt} + \frac{\alpha_1}{FF_1}FF_2N^{tj} + FF_1\frac{\alpha_2}{FF_2}N^{jt} - \alpha_1\alpha_2N^{jj}, \tag{41}$$

where $\alpha_i = \frac{r_i f_i}{r_i - f_i}$. Similarly for a case of a single fake tau, by selecting control sample with no N^{rr} events, then $\lim_{r\to 1}\alpha = FF$, and Equation 41 reduces to a simpler form:

$$N^{tt}_{fakes} = FF_2 \times N^{tj} + FF_1 \times N^{jt} - FF_1 \times FF_2 \times N^{jj} \tag{42}$$

In case of $N^{fr} = 0$, Equation 41 reduces to an even simpler form:

$$N^{tt}_{fakes} = \theta \times N^{tj}, \tag{43}$$

with $\theta = N^{jt}/N^{jj}$, and this approach is commonly called the "ABCD" method.

10.1.2 Fake τ sources

The dominant background in the $1\ell 2\tau$ channel is $t\bar{t}$ pair production, where one or two τ_{had} are fakes from various sources: b-jets, hadronic W boson decays, or jets that are not associated with the $t\bar{t}$ decay. Since the reconstructed lepton is almost always originating from leptonic top decays (in $\sim 98\%$ of the cases), the two τ_{had} can be dissociated with respect to a charge correlation with the lepton. Fig. 76 shows the breakdown of different τ_{had} sources in the SR separately for τ_{had} with a same/opposite charge sign with respect to the selected lepton.

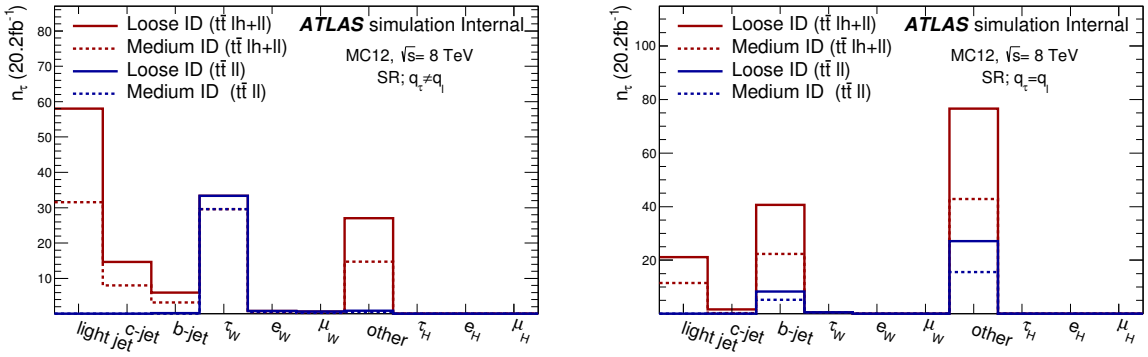


Figure 76: Sources of τ_{had}, calculated from $t\bar{t}$ MC truth information, for events passing all selection cuts. The plots show the expected number of τ_{had}s that pass *Loose* (solid line) or *Medium* (dashed line) identification criteria. The plots also show the contribution of the fully-leptonic $t\bar{t}$ decay mode (blue line) with respect to the semi-leptonic and fully-leptonic $t\bar{t}$ decays modes. Left: τ_{had} charge is opposite to the lepton charge; Right: τ_{had} charge is the same as the lepton charge. "other" stands for a fake τ_{had} that is not originating from the $t\bar{t}$ decay.

Table 32 shows the fraction of different sources for τ_{had} candidates in the SR with same/opposite charge as the reconstructed lepton.

The different origin of τ_{had} fakes leads to different fake rates. In order to avoid systematic biases on the matrix-method, the τ_{had} are categorized with respect to a charge correlation with the reconstructed lepton.

Table 32: Fraction of τ_{had} sources for two different charge selection obtained after applying all selection cuts, measured from the truth information on the $t\bar{t}$ MC sample.

Source	$Q(\tau_{\text{had}}) = Q(\ell)$		$Q(\tau_{\text{had}}) \neq Q(\ell)$	
	Loose	Medium	Loose	Medium
W+jets	16.23%	15.16%	51.59%	45.03%
real τ_{had}	0.38%	0.69%	23.75%	33.70%
b-jets	28.94%	28.83%	4.27%	3.64%
"other"	54.44%	55.30%	19.24%	16.75%

10.1.3 Fake rate estimation

The scaling factor θ can be measured in the dedicated fake control region. The fake-CR is defined by requiring 1 lepton, 2 τ_{had} with opposite charge, $n_{jet} \leq 2$ or by applying b-veto. The fakes scaling factors extracted from τ_{had} with same sign to a reconstructed lepton. This region contains ~97% of fake τ_{had}, and the real τ_{had} contribution is shown in Fig. 77. The small contribution from the real τ_{had}, is taken as

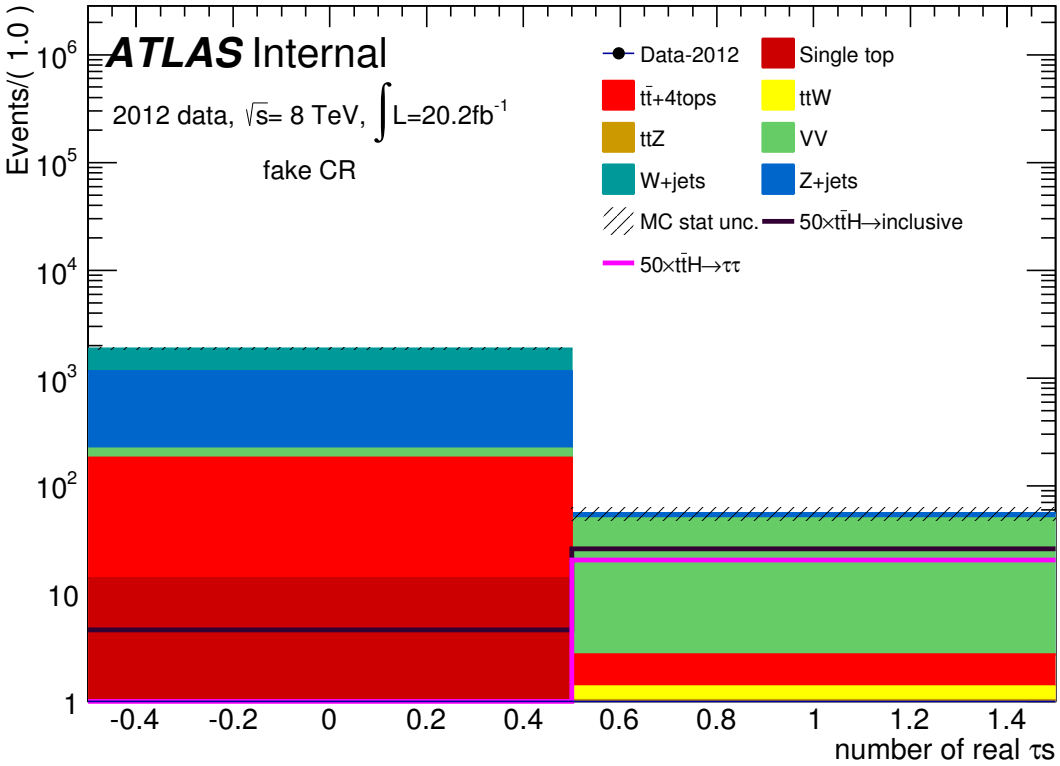


Figure 77: Number of real τ_{had} with the same charge as the reconstructed lepton in the fake CR, estimated from MC simulation.

a systematic uncertainty of an order of 3%.

The scaling factors were obtained from the data as a function of τp_T and η for 1 and 3 track τ_{had} separately. Jets originating from different processes might have different efficiencies due to intrinsic jet

sub-structure. B-jets that are misidentified as hadronic τs have a different reconstruction efficiency as compared to light jets. In order to access the b-jet "like" topologies, a τ MV1 weight of the associated jet is used to distinguish between different types of jets. For consistency with the proper b-tagging in our analysis, the MV1 cut corresponding to 70% wp. is used. The mean life-time of a tau lepton is only $\approx$5 times smaller than the life-time of a B-mesons arising from b-jets. Therefore, reconstructed hadronic τs (mainly 3 track candidates) can have displaced secondary vertices, resulting in high b-tag weights. This is used to categorized τ candidates in our fake factor method. However, no cut on the τ candidate MV1 weight is used in the actual event selection.

The fake-scaling factors for different τ_{had} selections and their fraction in the dedicated control region are shown in Fig. 78 parametrized as a function of a τ_{had} transverse momentum and in Fig. 79 as a function of a $\tau_{\text{had}}\eta$.

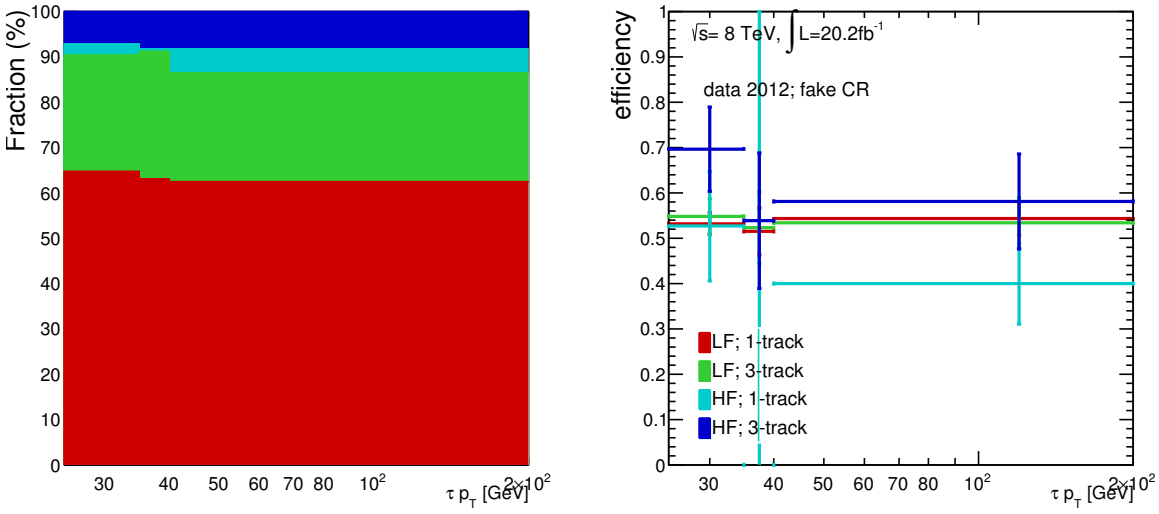


Figure 78: Scale factors for fake τ_{had} calculated as a function of τ_{had} transverse momentum for different tau selection (right) and the fraction of the selected τ_{had} in this control region (right). LF denotes τ candidates that hasv a small MV1 weight (light flavor), and HF denotes the τ with large MV1 weight (heavy flavor).

10.1.4 Closure tests

The fake scaling factors can have a systematic uncertainty due to the different jet composition between the dedicated fake CR and the signal region. To overcome this systematic effect, the method was validated with the fake CR when the τ_{had} has an opposite charge sign as the reconstructed lepton, and the τ passes or fails *Medium* identification criteria. The estimated fake scale factors (θ) were tested on a $t\bar{t}$ MC sample for different cuts and also in the SR. The difference between the measured fake factors and those estimated from MC is a sign of non-closure of the method and this difference is therefore taken as a systematic uncertainty of our method. However, the non-closure is within the statistical uncertainties of the samples used. Fig. 80 shows the closure tests with different sets of cuts.

The dominant uncertainty of the method is the statistical uncertainty. When the fake scale factor is about $\theta \sim 0.5$, the statistical uncertainty in the SR will be approximately the square root of the number of expected events in the SR.

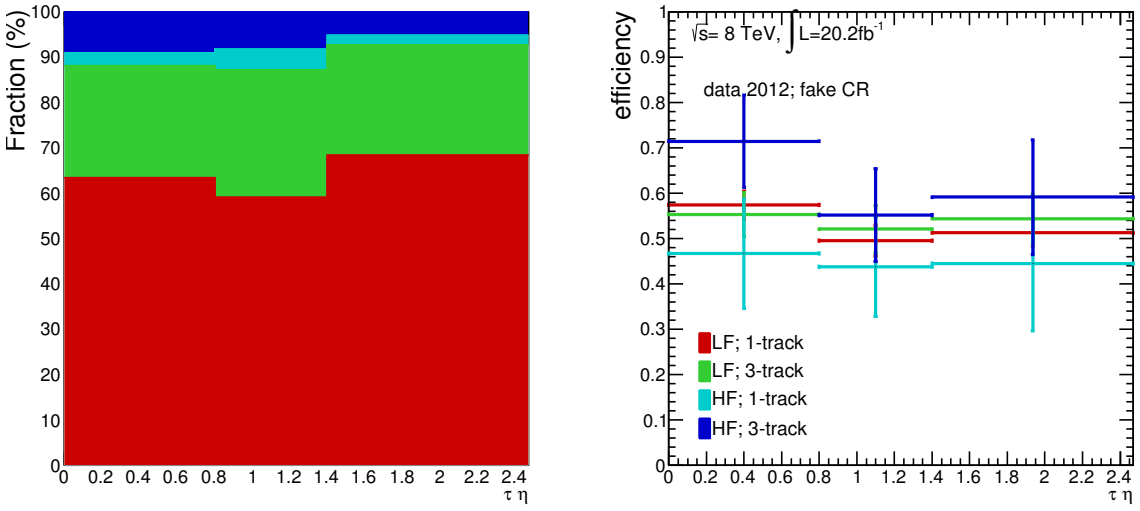


Figure 79: Scale factors for fake τ_{had} calculated as a function of $\tau_{\text{had}}\eta$ for different tau selection (right) and the fraction of the selected τ_{had} in the control region (right).

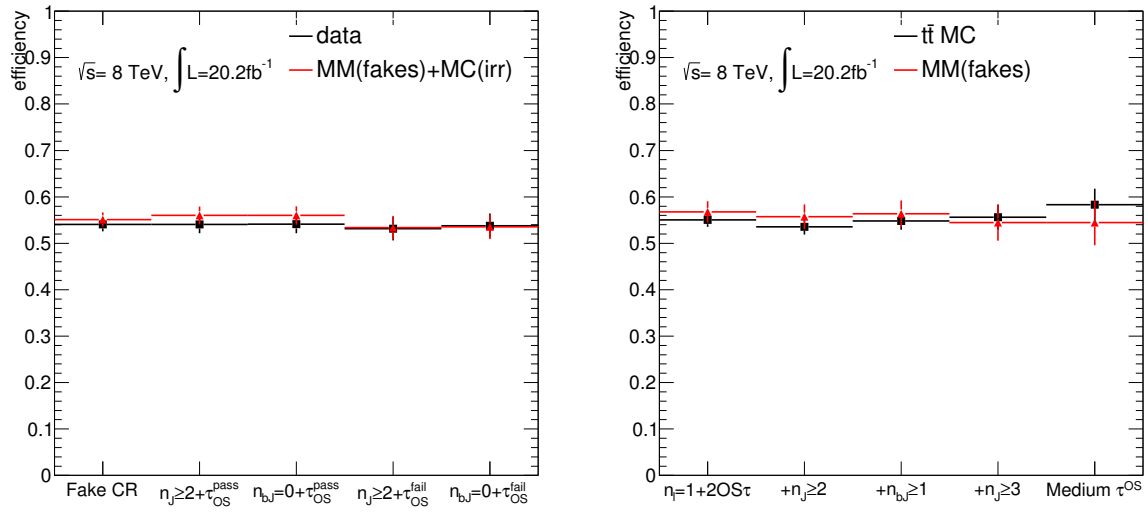


Figure 80: Scale factors for fake τ_{had} calculated as a function of τ_{had} transverse momentum for different tau selection (right) and the fraction of the selected τ_{had} in this control region (right).

10.1.5 Control plots

The results of estimating the fake background using the fake scale factors are shown in this section. Fig. 81 show various kinematical variables obtained in the fake CR, while Fig. 82 shows several kinematical variables obtained at the preselection level.

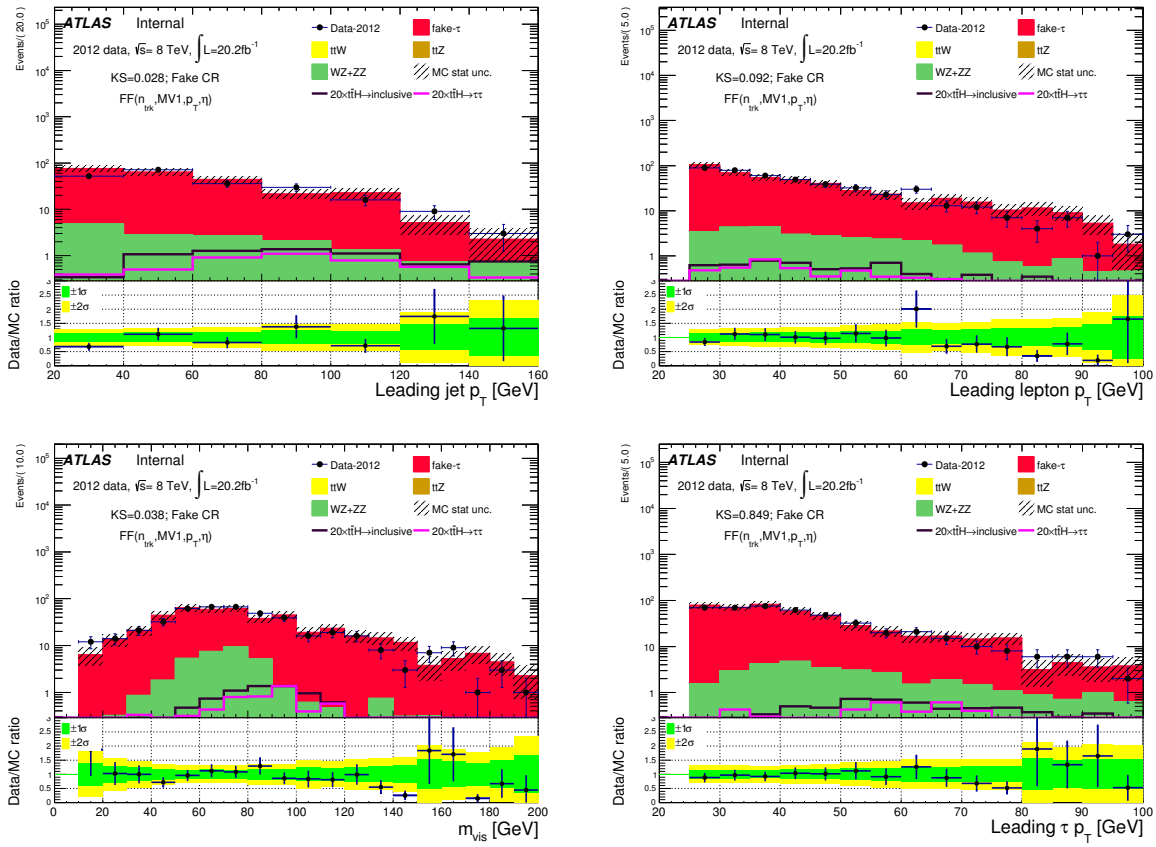


Figure 81: Distribution of several kinematical variables in the fake CR, where the fake τ contribution is estimated from the data.

A reasonable agreement between data and the estimated backgrounds is achieved with our data-driven fake estimation method.

11 Conclusions

We have evaluated the backgrounds in the search for the associated production of a Higgs boson with a top quark pair in the multilepton final state. In particular, the reducible backgrounds in which one of the leptons is a fake have been extracted with data driven techniques. Another important background, $t\bar{t}V$, is estimated from MC, and the uncertainty on the overall normalization is obtained by varying the theoretical inputs within an agreed range. Table 33 summarizes the results in the same fashion as the same table in Note 1 [1].

A Appendix A: Monte Carlo samples used in this analysis.

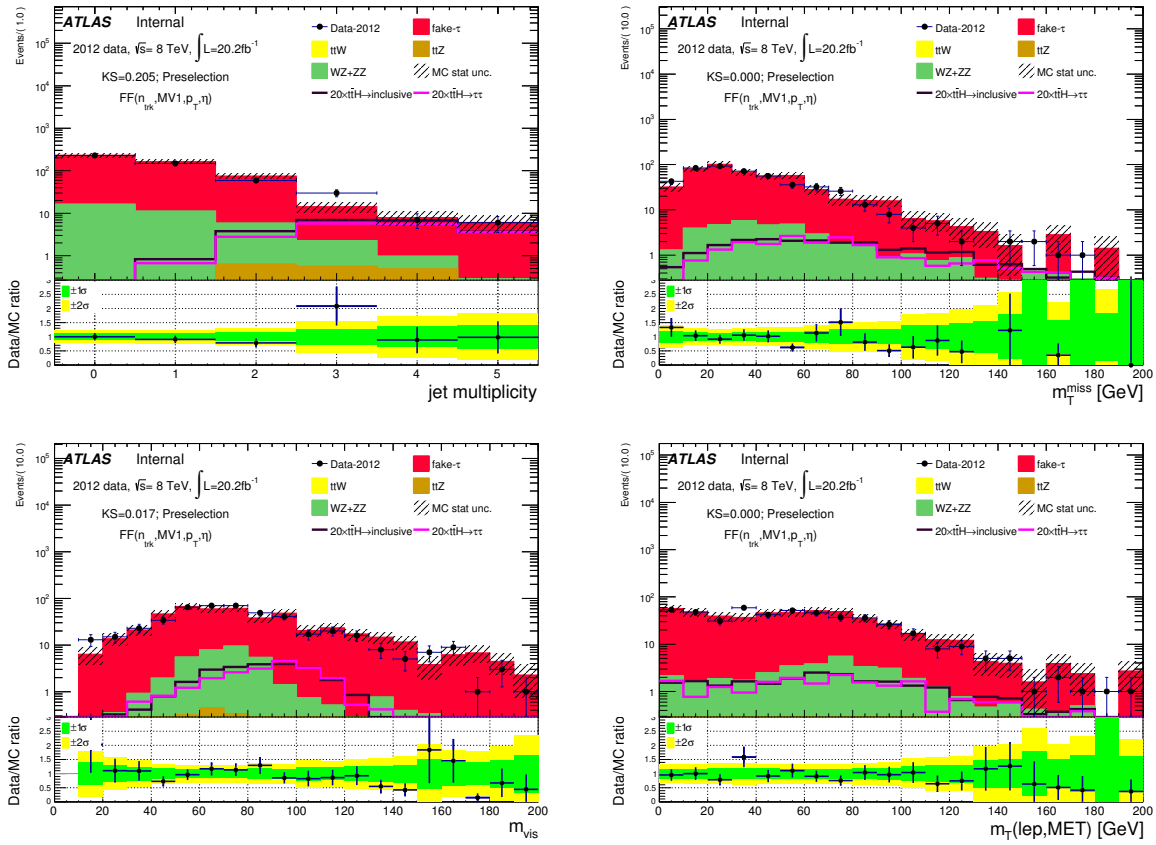


Figure 82: Distribution of several kinematical variables after preselection level, where the fake τ contribution is estimated from the data.

Table 33: Expected number of events in the 5 channels. For the processes estimated from data, backgrounds from fake leptons and charge mis-identification are shown separately and identified as *DD*. The latter is only applicable to the Same-sign channels. The lines with "*MC*" refer to the yields of the two source processes (top quark and Z+jets) as separately expected in simulation. The total and significance lines marked as "fake DD" are the result of replacing the top quark and Z+jets MC yields with the data-driven ones for 2lSS; and the top quark yield with the data-driven one for 3lep. "–" means that the entry is not applicable. Uncertainties are statistical only. *The 1lep+2Tau will be added to this table when ready for review*

	Same-sign						3 leptons	1τ	4 leptons	
	≥ 5 jets			4 jets					Z enriched	Z depleted
	$e^{\pm}e^{\pm}$	$e^{\pm}\mu^{\pm}$	$\mu^{\pm}\mu^{\pm}$	$e^{\pm}e^{\pm}$	$e^{\pm}\mu^{\pm}$	$\mu^{\pm}\mu^{\pm}$				
$t\bar{t}H$	0.73 ± 0.03	2.13 ± 0.05	1.41 ± 0.04	0.44 ± 0.02	1.16 ± 0.03	0.74 ± 0.03	2.34 ± 0.04	0.47 ± 0.02	0.19 ± 0.01	0.03 ± 0.00
$t\bar{t}V$	2.60 ± 0.13	7.42 ± 0.17	5.01 ± 0.16	3.05 ± 0.13	8.39 ± 0.24	5.79 ± 0.20	7.21 ± 0.24	0.85 ± 0.06	0.74 ± 0.05	0.00 ± 0.00
tZ							0.71 ± 0.03		*incl. in $t\bar{t}V$*	*incl. in $t\bar{t}V$*
VV	0.48 ± 0.25	0.37 ± 0.23	0.68 ± 0.30	0.77 ± 0.27	1.93 ± 0.80	0.54 ± 0.30	0.89 ± 0.25	0.11 ± 0.10	0.08 ± 0.01	0.00 ± 0.00
$t\bar{t}$, tX (MC)	1.31 ± 0.67	2.55 ± 0.84	1.76 ± 0.67	4.99 ± 1.19	8.19 ± 1.41	3.70 ± 1.03	2.46 ± 0.19	0.70 ± 0.38	0.00 ± 0.00	0.00 ± 0.00
Z+jets (MC)	0.16 ± 0.16	0.28 ± 0.20	0.12 ± 0.12	1.37 ± 0.78	0	0.23 ± 0.23	0	0	0.00 ± 0.00	0.00 ± 0.00
fake leptons (DD)	2.31 ± 0.97	3.87 ± 1.01	1.24 ± 0.41	3.43 ± 1.38	6.82 ± 1.63	2.38 ± 0.78	2.62 ± 0.51	$0.46^{+0.69}_{-0.53}$	$(1.1 \pm 0.6) \cdot 10^{-3}$	$(0.09 \pm 0.03) \cdot 10^{-3}$
Q misid (DD)	1.10 ± 0.09	0.85 ± 0.08	–	1.82 ± 0.11	1.39 ± 0.08	–	–	*incl. in fakes*	–	–
Tot Background (fake MC)	4.56 ± 1.17	10.62 ± 1.54	7.57 ± 1.31	10.18 ± 2.43	18.51 ± 2.54	10.26 ± 1.82	11.27 ± 0.40	1.66 ± 0.40	0.83 ± 0.07	0.01 ± 0.00
Tot Background (fake DD)	6.49 ± 1.04	12.51 ± 1.04	6.93 ± 0.52	9.07 ± 1.42	18.53 ± 1.83	8.71 ± 0.88	11.43 ± 0.62	$1.42^{+0.70}_{-0.54}$	0.831 ± 0.075	0.0110 ± 0.0003
$s/\sqrt{b}$ (fake MC)	0.34	0.65	0.51	0.14	0.27	0.23	0.70	0.36	0.21	0.30
$s/\sqrt{b}$ (fake DD)	0.29	0.60	0.54	0.15	0.27	0.25	0.69	0.39	0.21	0.29
$s/\sqrt{b \oplus 0.3\text{fake(MC)} \oplus 0.2\text{ttV}}$	0.33	0.58	0.47	0.12	0.22	0.21	0.63	0.33	0.207	0.30
$s/\sqrt{b \oplus 0.3\text{fake(DD)} \oplus 0.2\text{ttV}}$	0.27	0.53	0.50	0.14	0.23	0.23	0.62	0.36	0.207	0.286

Table 34: Monte Carlo samples used for background description. Unless otherwise specified MadGraph samples use Pythia 6 for showering and Alpgen samples use Herwig+Jimmy.

Sample	Process	Generator	Cross-section [fb]	$\mathcal{L}$ [fb^{-1}]	Detector simulation
119353	ttbarW+Np0	MadGraph	122.84	2633.2	Full
174830	ttbarW+Np1	MadGraph	62.98	6347.7	Full
174831	ttbarW+Np2	MadGraph	48.95	7450.8	Full
119355	ttbarZ+Np0	MadGraph	90.70	4409.9	Full
174832	ttbarZ+Np1	MadGraph	60.78	6581.1	Full
174833	ttbarZ+Np2	MadGraph	53.29	5580.1	Full
158344	4top	MadGraph	0.69	291441.1	Full
158813	ttbarWW	Madgraph+Pythia8	1.90	104890.0	AF2
117050	ttbar(noFullHad)	Powheg+Pythia6	137320.39	109.1	Full/AF2
181087	ttbar(dilepton)	Powheg+Pythia6	26554.16	1334.1	AF2
110101	single top tchan	AcerMC+Pythia6	28433.15	271.0	Full
110119	single top schan→l	Powheg+Pythia6	1817.64	660.1	Full
110141	Wt	Powheg+Pythia6	2337.67	424.8	Full
146890	Wγ*→lνee(mll<7GeV)	MadGraph	11256.00	35.5	Full
146892	Wγ*→l$\nu\mu\mu$(mll<7GeV)	MadGraph	2769.18	108.2	Full
146894	Wγ*→l$\nu\tau\tau$(mll<7GeV)	MadGraph	295.81	101.4	Full
126892	WW→lνlν	Sherpa	5679.00	475.4	Full
179974	WZ→lllν	Sherpa	10231.79	263.8	Full
161987	WZ→lllν+jj(6EWcoupling)	Sherpa	12.56	1592.5	Full
161988	ZZ→llll+jj(6EWcoupling)	Sherpa	0.74	679628.6	Full
126937	ZZ→4e(mll>4GeV)	Powheg+Pythia8	66.71	16488.7	Full
126938	ZZ→2e2μ(mll>4GeV)	Powheg+Pythia8	141.29	11319.4	Full
126939	ZZ→2e2τ(mll>4GeV)	Powheg+Pythia8	99.54	11048.7	Full
126940	ZZ→4μ(mll>4GeV)	Powheg+Pythia8	67.06	16398.8	Full
126941	ZZ→2μ2τ(mll>4GeV)	Powheg+Pythia8	100.30	10960.8	Full
126942	ZZ→4τ(mll>4GeV)	Powheg+Pythia8	7.79	38491.4	Full
126949	ZZ→ee$\nu\nu$(mll>4GeV)	Powheg+Pythia8	168.00	1787.1	Full
126950	ZZ→ $\mu\mu\nu\nu$(mll>4GeV)	Powheg+Pythia8	168.00	1785.7	Full
126951	ZZ→ $\tau\tau\nu\nu$(mll>4GeV)	Powheg+Pythia8	168.00	1785.7	Full
116601	gg→ZZ→4e	gg2ZZ+Herwig	0.67	134328.4	Full
116602	gg→ZZ→4μ	gg2ZZ+Herwig	0.67	134328.4	Full
116603	gg→ZZ→2e2μ	gg2ZZ+Herwig	1.35	66666.7	Full
146434	Wγ+4p	Alpgen	2440.76	149.5	Full
146435	Wγ+5p	Alpgen	536.04	111.9	Full
146436	Wγ+0p_LepPho_EF	Alpgen	82935.65	172.4	Full
146437	Wγ+1p_LepPho_EF	Alpgen	30712.27	174.1	Full
146438	Wγ+2p_LepPho_EF	Alpgen	13396.59	216.4	Full
146439	Wγ+3p_LepPho_EF	Alpgen	5156.53	166.7	Full
158818	W+-W+-jj	Madgraph+Pythia8	344.42	551.7	AF2
107650	Zee+0p(60<M<2000GeV)	Alpgen	875477.10	7.3	Full
107651	Zee+1p(60<M<2000GeV)	Alpgen	190859.10	7.0	Full
107652	Zee+2p(60<M<2000GeV)	Alpgen	59956.35	6.8	Full

Table 34: Monte Carlo samples used for background description. Unless otherwise specified MadGraph samples use Pythia 6 for showering and Alpgen samples use Herwig+Jimmy.

Sample	Process	Generator	Cross-section [fb]	$\mathcal{L}$ [fb^{-1}]	Detector simulation
107653	Zee+3p(60<M<2000GeV)	Alpgen	17496.75	6.3	Full
107654	Zee+4p(60<M<2000GeV)	Alpgen	4624.18	6.5	Full
147078	Zee+5p(60<M<2000GeV)	Alpgen	1393.71	222.0	Full
107660	Z$\mu\mu$+0p(60<M<2000GeV)	Alpgen	875895.30	7.5	Full
107661	Z$\mu\mu$+1p(60<M<2000GeV)	Alpgen	190367.10	6.7	Full
107662	Z$\mu\mu$+2p(60<M<2000GeV)	Alpgen	60161.76	6.7	Full
107663	Z$\mu\mu$+3p(60<M<2000GeV)	Alpgen	17497.98	6.3	Full
107664	Z$\mu\mu$+4p(60<M<2000GeV)	Alpgen	4654.07	6.4	Full
147086	Z$\mu\mu$+5p(60<M<2000GeV)	Alpgen	1395.07	220.3	Full
146930	Z$\tau\tau$+0p(60<M<2000GeV)_2l_EF	Alpgen	34707.74	103.2	Full
146931	Z$\tau\tau$+1p(60<M<2000GeV)_2l_EF	Alpgen	9009.69	270.5	Full
146932	Z$\tau\tau$+2p(60<M<2000GeV)_2l_EF	Alpgen	3175.34	156.3	Full
146933	Z$\tau\tau$+3p(60<M<2000GeV)_2l_EF	Alpgen	1028.86	173.4	Full
146934	Z$\tau\tau$+4p(60<M<2000GeV)_2l_EF	Alpgen	317.18	106.1	Full
147094	Z$\tau\tau$+5p(60<M<2000GeV)	Alpgen	1398.02	221.6	Full
146860	Zee+0p(10<M<60GeV)_2l_EF	Alpgen	43274.26	159.2	Full
146861	Zee+1p(10<M<60GeV)_2l_EF	Alpgen	26370.87	170.6	Full
146862	Zee+2p(10<M<60GeV)_2l_EF	Alpgen	8702.67	151.4	Full
146863	Zee+3p(10<M<60GeV)_2l_EF	Alpgen	2795.55	156.5	Full
146864	Zee+4p(10<M<60GeV)_2l_EF	Alpgen	777.16	140.7	Full
146835	Zee+5p(10<M<60GeV)	Alpgen	824.28	89.1	Full
146870	Z$\mu\mu$+0p(10<M<60GeV)_2l_EF	Alpgen	44947.85	259.8	Full
146871	Z$\mu\mu$+1p(10<M<60GeV)_2l_EF	Alpgen	27298.35	271.0	Full
146872	Z$\mu\mu$+2p(10<M<60GeV)_2l_EF	Alpgen	8957.82	243.2	Full
146873	Z$\mu\mu$+3p(10<M<60GeV)_2l_EF	Alpgen	2875.39	254.4	Full
146874	Z$\mu\mu$+4p(10<M<60GeV)_2l_EF	Alpgen	794.01	137.7	Full
146845	Z$\mu\mu$+5p(10<M<60GeV)	Alpgen	825.54	96.3	Full
146880	Z$\tau\tau$+0p(10<M<60GeV)_2l_EF	Alpgen	100.77	287.5	Full
146881	Z$\tau\tau$+1p(10<M<60GeV)_2l_EF	Alpgen	176.04	170.4	Full
146882	Z$\tau\tau$+2p(10<M<60GeV)_2l_EF	Alpgen	109.41	263.7	Full
146883	Z$\tau\tau$+3p(10<M<60GeV)_2l_EF	Alpgen	52.13	396.0	Full
146854	Z$\tau\tau$+4p(10<M<60GeV)	Alpgen	3084.48	70.0	Full
146855	Z$\tau\tau$+5p(10<M<60GeV)	Alpgen	824.55	141.6	Full
109300	Zeebb+0p(40<M<2000)	Alpgen	10304.57	14.6	Full
109301	Zeebb+1p(40<M<2000)	Alpgen	4001.07	20.0	Full
109302	Zeebb+2p(40<M<2000)	Alpgen	1463.95	30.7	Full
109303	Zeebb+3p(40<M<2000)	Alpgen	618.42	8.1	Full
109305	Z$\mu\mu$bb+0p(40<M<2000)	Alpgen	10300.27	14.5	Full
109306	Z$\mu\mu$bb+1p(40<M<2000)	Alpgen	4002.42	20.0	Full
109307	Z$\mu\mu$bb+2p(40<M<2000)	Alpgen	1452.63	31.0	Full
109308	Z$\mu\mu$bb+3p(40<M<2000)	Alpgen	623.23	8.0	Full
109310	Z$\tau\tau$bb+0p(40<M<2000)	Alpgen	10302.11	14.5	Full

Table 34: Monte Carlo samples used for background description. Unless otherwise specified MadGraph samples use Pythia 6 for showering and Alpgen samples use Herwig+Jimmy.

Sample	Process	Generator	Cross-section [fb]	$\mathcal{L}$ [fb^{-1}]	Detector simulation
109311	Z$\tau\tau$bb+1p(40<M<2000)	Alpgen	3988.52	20.1	Full
109312	Z$\tau\tau$bb+2p(40<M<2000)	Alpgen	1468.37	30.6	Full
109313	Z$\tau\tau$bb+3p(40<M<2000)	Alpgen	612.43	8.2	Full
126414	Zeecc+0p(40<M<2000)	Alpgen	19254.42	31.4	Full
126415	Zeecc+1p(40<M<2000)	Alpgen	8480.36	30.7	Full
126416	Zeecc+2p(40<M<2000)	Alpgen	3592.09	30.6	Full
126417	Zeecc+3p(40<M<2000)	Alpgen	1403.55	28.5	Full
126418	Z$\mu\mu$cc+0p(40<M<2000)	Alpgen	19248.27	31.2	Full
126419	Z$\mu\mu$cc+1p(40<M<2000)	Alpgen	8478.39	31.3	Full
126420	Z$\mu\mu$cc+2p(40<M<2000)	Alpgen	3588.65	32.0	Full
126421	Z$\mu\mu$cc+3p(40<M<2000)	Alpgen	1399.37	28.6	Full
117706	Z$\tau\tau$cc+0p(40<M<2000)	Alpgen	19251.96	31.2	Full
117707	Z$\tau\tau$cc+1p(40<M<2000)	Alpgen	8484.42	31.2	Full
117708	Z$\tau\tau$cc+2p(40<M<2000)	Alpgen	3579.30	26.2	Full
117709	Z$\tau\tau$cc+3p(40<M<2000)	Alpgen	1394.82	28.7	Full
181471	Zγ*$\rightarrow$4e(mll1/2>4/<4GeV)+1p	Sherpa	2828.60	282.8	Full
181472	Zγ*$\rightarrow$2μ2e(mll1/2>4/<4GeV)+1p	Sherpa	2345.00	255.4	Full
181473	Zγ*$\rightarrow$2τ2e(mll1/2>4/<4GeV)+1p	Sherpa	1593.30	31.9	Full
181474	Zγ*$\rightarrow$2e2μ(mll1/2>4/<4GeV)+1p	Sherpa	486.10	308.6	Full
181475	Zγ*$\rightarrow$4μ(mll1/2>4/<4GeV)+1p	Sherpa	508.40	294.8	Full
181476	Zγ*$\rightarrow$2τ2μ(mll1/2>4/<4GeV)+1p	Sherpa	422.90	354.7	Full
181477	Zγ*$\rightarrow$2e2τ(mll1/2>4/<4GeV)+1p	Sherpa	4.00	5000.0	Full
181478	Zγ*$\rightarrow$2μ2τ(mll1/2>4/<4GeV)+1p	Sherpa	4.00	5000.0	Full
181479	Zγ*$\rightarrow$4τ(mll1/2>4/<4GeV)+1p	Sherpa	4.10	4878.0	Full
160155	ggF_H(125)$\rightarrow$ZZ$\rightarrow$4l	Powheg+Pythia8	5.57	35913.1	Full

References

[1] Boumediene, D et al. *Search for Standard Model Higgs boson production in association with a top quark pair in multilepton signatures with the ATLAS detector : Object and event selection*. ATL-COM-PHYS-2014-053, 2014.

[2] InDetTracking CP. InDetTrackingPerformanceGuidelines twiki. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/InDetTrackingPerformanceGuidelines#Pile_up_rescaling, 2013.

[3] Bell, W et al. Measurement of top quark pair differential cross section with ATLAS in pp collisions at $\sqrt{s}$ = 7 TeV : Measurement of top quark pair differential cross section with ATLAS using 5/fb of data collected in 2011 and the rel.17 version of the Atlas software. ATL-COM-PHYS-2012-1137, 2013.

[4] ttH, H-¿bb analysis team. ttbar Powheg+Pythia reweghting. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TTH2bbWinter2013#ttbar_Powheg_Pythia_reweghting, 2013.

[5] John M. Campbell and R. Keith Ellis. $t\bar{t}W^{\pm}$ production and decay at NLO. *JHEP*, 1207:052, 2012.

[6] M.V. Garzelli, A. Kardos, C.G. Papadopoulos, and Z. Trocsanyi. $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects. *JHEP*, 1211:056, 2012.

[7] Campbell, John and Ellis, R. Keith and Röntsch, Raoul. Single top production in association with a Z boson at the LHC. *Phys.Rev.*, D87:114006, 2013.

[8] Georges Aad et al. Measurement of WZ production in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector. *Eur.Phys.J.*, C72:2173, 2012.

[9] Georges Aad et al. Measurement of the cross-section for W boson production in association with b-jets in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector. *JHEP*, 1306:084, 2013.

[10] Georges Aad et al. Measurement of differential production cross-sections for a Z boson in association with b-jets in 7 TeV proton-proton collisions with the ATLAS detector. 2014.

[11] Egamma CP. Electron Likelihood PID twiki. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ElectronLikeli 2013.

[12] Dao, V et al. Measurement of the associated production of a vector boson (W, Z) and top quark pair in the opposite sign dilepton channel with pp collisions at $\sqrt{s}$=8 TeV with the ATLAS detector at the LHC. https://cds.cern.ch/record/1698058/, 2014.

[13] Top WG. Heavy Flavor Overlap Removal Tool. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HforTool, 2014.

[14] ATLAS Collaboration. *Measurement of the production cross section of jets in association with a Z boson in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*. JHEP07(2013), 032, 2013.

[15] A Alonso and B Meirose. New data-driven methods for lepton charge mis-identification. Technical Report ATL-COM-PHYS-2012-264, CERN, Geneva, Mar 2012. This note details the charge mis-identification data driven methods developed for arXiv: 1202.5520 [hep-ph], ATL-COM-PHYS-2011-1442.

[16] B Acharya et al. Object selection and calibration, background estimations and MC samples for top quark analyses using the full 2012 data set. Technical Report ATL-COM-PHYS-2013-1016, CERN, Geneva, Jul 2013.

[17] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[18] ATLAS Collaboration. *ABCD method in searches*. https://twiki.cern.ch/twiki/pub/AtlasProtected/ATLASStatisticsFAQ/ABCD.pdf, 2012.

[19] ATLAS Collaboration. Identification of the Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector, Jul 2013.

[20] K Becker, T Cornelissen, F Derue, A Henrichs, D Hirschbhl, X Lei, O Nackenhorst, F O'Grady, D Pelikan, M Pinamonti, S Pires, J Sjlin, and P Tepel. Estimation of Fake Lepton Background for Top Analyses Using the $\sqrt{s} = 8$ TeV Dataset. Technical Report ATL-COM-PHYS-2013-1100, CERN, Geneva, Aug 2013.

[21] ATLAS Collaboration. Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector, Nov 2013.