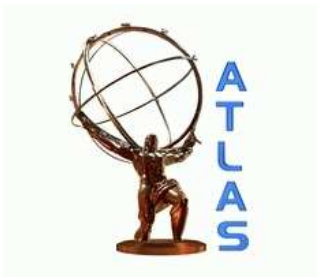
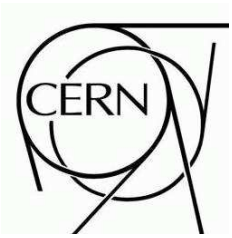


ATLAS NOTE



Data-Driven Determinations of W, Z and Top Backgrounds to Supersymmetry

The ATLAS Collaboration[1)]

This note is part of CERN-OPEN-2008-020. This version of the note should not be cited: all citations should be to CERN-OPEN-2008-020.

ATL-PHYS-PUB-2009-064
10 May 2009

Abstract

The Standard Model processes of W boson, Z boson and top quark production each in association with jets constitute major backgrounds to searches for Supersymmetry at the LHC. In this note, we estimate the contribution of these backgrounds for a basic SUSY selection, and discuss methods to derive them from the initial 1 fb^{-1} of integrated luminosity at ATLAS.

[1)]This note prepared by G. Akimoto, S. Asai, Y. Azuma, R. Bruneliere, P. Casado, C. Clement, P. de Jong, L. Fiorini, P. Hansson, N. Kanaya, Y. Kataoka, F. Koetsveld, A. Koutsman, H. Kroha, F. Legger, B.R. Mellado Garcia, L-M. Mir, H. Okawa, C. Osuna, Y. Pan, G. Redlinger, T. Sarangi, J. Sjolin, T. Suzuki, Y. Tomishima, D. Tovey, W. Verkerke, S.L. Wu, T. Yamazaki, Z. Yang, X-A. Zhuang and V. Zhuravlov.

1 Introduction

1.1 Motivation

The Large Hadron Collider (LHC) will provide excellent opportunities to search for new physics beyond the Standard Model, and the ATLAS detector [1] is a general purpose experiment to explore such new physics. Supersymmetry (SUSY) is a theoretically attractive model for new physics beyond the Standard Model, and searching for Supersymmetry is one of the main objectives of ATLAS. The actual search strategy is described elsewhere in this volume [2].

It is clear, however, that any discovery of new physics can only be claimed when the Standard Model backgrounds are understood and are under control. It is expected that at the LHC, Monte Carlo predictions will not be sufficient to achieve this: the backgrounds will have to be derived from the data themselves, possibly helped by Monte Carlo. The development and description of such data-driven background estimation is the topic of this note. We note that for a complete understanding of the backgrounds, multiple, independent methods are desired. Each of these may be sensitive to a specific background source, and affected by specific systematic effects. Only their consistency in combination allows for sufficient confidence in the control of the background to claim a discovery when a signal appears to be present.

1.2 Data-driven methods: scope of this note

The general aim of data-driven methods is to estimate from the data the Standard Model backgrounds and their uncertainties in a "signal" region, in which new physics may be present. Such a signal region is typically obtained after applying selection cuts, or multivariate methods, and the new physics is searched for as an excess in the number of selected events over background, or as an excess in certain regions of certain distributions.

The background estimation is performed by selection of "control samples", from which predictions in the signal region are derived. Good control samples should be as close as possible to the signal region, yet free of SUSY signal, give an unbiased estimate of background in the signal region, have sufficient statistics, and small theoretical uncertainties. This note intends to describe a number of ideas on selection of such control samples for SUSY searches. Good control of the composition of control samples is important for a correct extrapolation into the signal region.

The methods described in this note should not be regarded as the final word on these procedures, but rather present a number of ideas. Each of these ideas will have to be pursued further, and the effect of other systematic uncertainties will need to be studied. Furthermore, SUSY selection cuts will evolve, and so the methods will need to evolve too. We do believe, however, that a first indication of the uncertainties that can be expected can be given.

This note deals with top, W and Z backgrounds to SUSY searches with primary squark or gluino production, and assuming R-parity conservation. The initial priority is to simulate results for 1 fb^{-1} of integrated luminosity, and for the understanding of the detector expected. The other important QCD background of quark (other than top) and gluon jet production is treated elsewhere in this volume [3]. Backgrounds to alternative production models are also described elsewhere: direct gaugino production [4], and photonic and long-lived particle (such as R hadron) signatures [5].

1.3 SUSY contamination

If SUSY is discoverable, it is likely that SUSY events will creep into the control samples, thereby affecting the background estimates. In general, SUSY events, mistakenly regarded as Standard Model physics,

will lead to an overestimation of the background, and thus to a reduced SUSY event excess. The extent to which this happens will be analysis- and SUSY model-dependent.

Since we do not know whether SUSY exists, we quote the SUSY contamination effects separately from the other systematics. We will do so by running each data-driven estimation method not only over background samples, but also on a number of SUSY signal samples [6]. The samples represent various regions of mSUGRA parameter space, and together give an impression of the effects. The SU1 sample is a point in the stau coannihilation region, the SU2 sample in the focus point region, and the SU3 sample in the bulk region. The SU4 point is a low-mass point, just above the Tevatron limits. It has a very large cross-section, and kinematic distributions that are typically only slightly harder than the Standard Model background. As will be shown, this model has the largest SUSY contamination effect on the background estimates.

There are a number of ways that the data-driven methods can take the presence of SUSY into account:

1. Iteration. The Standard Model background is evaluated under the assumption that there is no SUSY. This will overestimate the background if there is SUSY, and reduce any excess. Nevertheless, if an excess is seen, the underlying assumption in the background estimation has been proven wrong, and a correction can be applied. This correction can be derived from the properties of the observed excess, and will lead to a new background estimate. An example of such a procedure is the "new MT method" described in section 3.3.3. However, other implementations are possible, and perhaps necessary, as well.

2. A combined fit determining the composition of the control sample, allowing for a possible SUSY contribution.

Both methods are investigated in this note. Nevertheless it is clear that these are preliminary ideas that require further investigation. Most likely, some form of iteration on the background determinations will be necessary.

1.4 Layout of this note

A number of important prerequisites for the studies presented here are described in an introductory note [6]:

- the physics processes that form the background to SUSY searches and how they are simulated, as well as a few SUSY event samples (SU1–SU8) that serve to estimate the effect of SUSY on our background estimates;

- the definition of objects like electrons, muon, taus, jets and missing transverse energy, and common variables like the effective mass M_{eff} ;

- the origin and common treatment of various systematic uncertainties, both from the simulation and from the performance of the detector.

Furthermore, the trigger menu that was used is described elsewhere [7]. In this note we then discuss the W, Z, and top-quark backgrounds and their data-driven estimation for two different SUSY search modes:

1. the mode with one isolated electron or muon (section 2);

2. the no-lepton mode, with a veto against isolated leptons (section 3).

2 One-lepton search mode

2.1 Selection

The one-lepton search mode is expected to play a major role in the SUSY search, since the requirement of an isolated lepton will be effective in suppressing QCD background. In this search mode, we require one isolated electron or muon, with a p_T of more than 20 GeV. We veto events with a second identified lepton with a p_T of more than 10 GeV, so that we have no overlap with the di-lepton search mode.

We demand at least four jets with $|\eta| < 2.5$ and $p_T > 50$ GeV, at least one of which must have $p_T > 100$ GeV. The transverse sphericity S_T should be larger than 0.2, and the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ should be larger than 100 GeV and larger than $0.2M_{\mathrm{eff}}$, where M_{eff} is the effective mass[2)]. The transverse mass M_T reconstructed from the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ should be larger than 100 GeV.

2.2 Backgrounds in Monte Carlo

In many SUSY models after the selection cuts have been applied clear excesses will be observed in the high $E_{\mathrm{T}}^{\mathrm{miss}}$ and high effective mass regions, as shown in Figure 1. The dominant background process for the one-lepton mode is $t\bar{t}$ (90%), with $W^{\pm}+$jets (10%) being the subdominant process. The neutrino emitted from the $W^{\pm}$ decays produces the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the both processes. Smaller contributions come from $Z+$jets, diboson and single top events and from QCD processes. It is interesting to note that the major $t\bar{t}$ background does not come from the semileptonic ($t\bar{t} \to b\bar{b}\ell\nu q\bar{q}'$) top pair events which are reduced by the M_T and $E_{\mathrm{T}}^{\mathrm{miss}}$ cuts, but rather from the double leptonic ($t\bar{t} \to b\bar{b}\ell\nu\ell\nu$) top decay where one lepton is not identified.

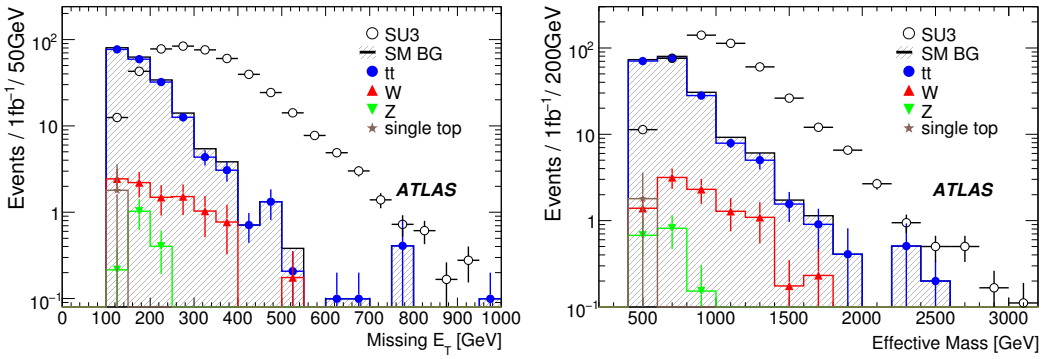


Figure 1: The $E_{\mathrm{T}}^{\mathrm{miss}}$ and effective mass distributions for the background processes and for an example SUSY benchmark point (SU3) in the one-lepton mode for an integrated luminosity of 1 fb^{-1}. The black circles show the SUSY signal. The hatched histogram show the sum of all Standard Model backgrounds; also shown in different colours are the various components of the background.

2.3 Data-driven estimation strategies

We discuss a variety of different methods to estimate the background from data. These methods differ in their approach and therefore are influenced by different systematic uncertainties, and they focus on different aspects of the background:

[2)] The variables S_T, M_{eff} and M_T are defined elsewhere in this volume [6].

1. estimation of W and $t\bar{t}$ background from a control sample formed by reversing one of the selection cuts (on M_T) (section 2.3.1);

2. estimation of the semileptonic $t\bar{t}$ background by explicit kinematic reconstruction and selection on top mass ("top box") (section 2.3.2);

3. estimation of the double leptonic $t\bar{t}$ background, where one lepton is missed, by explicit kinematic reconstruction of a control sample of the same process with both leptons identified (section 2.3.3);

4. estimation of that same double leptonic $t\bar{t}$ background from a control sample derived by a cut on a new variable HT2 (section 2.3.4);

5. estimation of $t\bar{t}$ background by Monte Carlo redecay methods (section 2.3.5);

6. estimation of W and $t\bar{t}$ background using a combined fit to control samples (section 2.3.6).

2.3.1 Creating a control sample by reversing the M_T cut

The transverse mass M_T is constructed from the identified lepton and the missing transverse energy. In the narrow-width limit M_T is constrained to be less than m_W for the semileptonic $t\bar{t}$ and the $W^{\pm}$ processes. Figure 2 shows that M_T is only weakly dependent on $E_{\mathrm{T}}^{\mathrm{miss}}$. This variable is therefore suitable for the estimation of the background distribution itself. Events with small M_T (< 100 GeV) are selected as the control sample, in which the $t\bar{t}$ ($\sim 84\%$) and $W^{\pm}$ ($\sim 16\%$) processes are enhanced over the SUSY and the other background processes. The large M_T (> 100 GeV) region is referred to as the signal region. Since, for the control sample, the other selection criteria are identical to those for events in the signal region, the same kinematic distributions including $E_{\mathrm{T}}^{\mathrm{miss}}$ can be obtained. The number of events for the various processes in signal region and control sample is summarized in the Table 1.

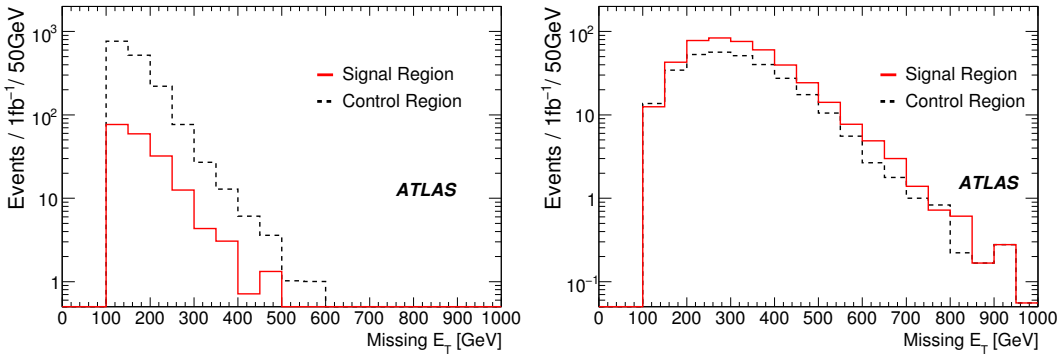


Figure 2: The $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for $t\bar{t}$ (left) and SUSY (SU3, right) signal. In both figures, the solid and dashed histograms show the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for $M_T > 100$ GeV and < 100 GeV, respectively. The numbers are normalized to 1 fb^{-1}.

The normalization factor is obtained from the event numbers of the signal region and the control sample ($100 < E_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV), in which the SUSY signal contribution is expected to be relatively small. Figure 3 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{eff} distributions which are obtained using this method to estimate the size of these backgrounds, and, for comparison, the true background distributions. The numbers of events with $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV and > 300 GeV are listed in Table 2. The prediction and the true values agree within the uncertainties, although somewhat less well for high $E_{\mathrm{T}}^{\mathrm{miss}}$.

Table 1: Number of background events and estimated numbers for $t\bar{t}$, $W^{\pm}$ and QCD processes without SUSY signal, normalized to 1 fb^{-1}.

	Signal Region	Control Sample
$t\bar{t}(\ell\nu q\bar{q})$	51 (25%)	1505 (77%)
$t\bar{t}(\ell\nu\ell\nu)$	140 (70%)	132 (7%)
$W^{\pm}(\ell\nu)$	10 (5%)	305 (16%)
SUSY(SU3)	450	317

The $t\bar{t}$ event composition of the control sample differs from that of the signal sample, since the M_T cut removes a much larger proportion of the semileptonic $t\bar{t}$ events. The control sample is still able to predict the background in the signal sample within statistical uncertainties. Nevertheless, the resulting systematic shift needs to be investigated, and would be desirable to obtain independent estimates of the fully-leptonic and semileptonic $t\bar{t}$ backgrounds separately.

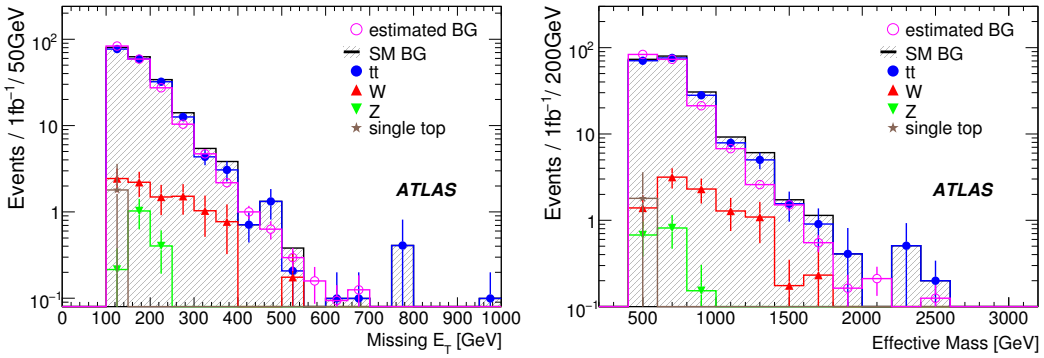


Figure 3: The $E_{\mathrm{T}}^{\mathrm{miss}}$ and effective mass distributions of the background processes for the one-lepton mode with an integrated luminosity of 1 fb^{-1}. The open circles show the estimated distributions with the M_T method. The hatched histogram shows the true sum of all Standard Model backgrounds; different symbols show the various contributions to the background.

Table 2: Numbers of background events and estimated numbers for the sum of all background processes without SUSY signal, normalized to 1 fb^{-1}

	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV
True BG	203 ± 6	12.4 ± 1.6
Estimated BG	190 ± 8	9.4 ± 0.7
Ratio(Est./True)	0.93 ± 0.05	0.76 ± 0.11

SUSY signal contamination If supersymmetric particles are produced they are also likely to contribute to the control samples. The estimated $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution with the presence of a SUSY signal (SU3 point) is shown in Figure 4 (left), and the numbers are listed in Table 3. The background is overestimated due

to the SUSY contamination, and the inferred $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is biased towards larger values. However, the amount of the over-estimation is smaller than the SUSY signal itself, and a clear excess can still be observed, as shown in the figure. The same exercise was repeated for other SUSY signal points, as also shown in Table 3.

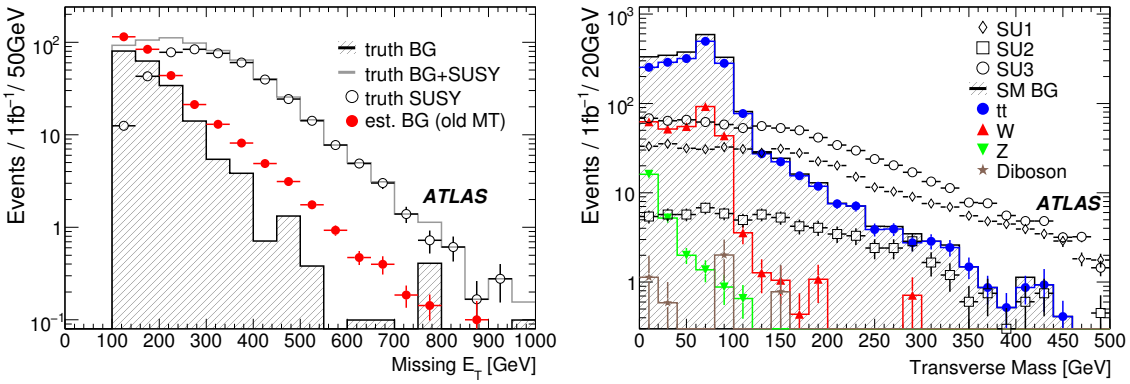


Figure 4: Left: the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution of the background processes for the one-lepton mode with an integrated luminosity of 1 fb^{-1}. The red dots show the estimated distributions with the M_T method, with SUSY signal (SU3) present. The hatched histogram shows the sum of all Standard Model backgrounds, and the OPEN histogram shows the SUSY signal (SU3). Right: the transverse mass distributions of the various SUSY signals (SU1, SU2 and SU3) with an integrated luminosity of 1 fb^{-1}. Background processes are superimposed for comparison. The hatched histogram shows the sum of all Standard Model backgrounds.

Table 3: Number of background events and estimated numbers for all background processes with SUSY signal, normalized to 1 fb^{-1}. Also the total number of events (SUSY + background) is shown.

	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV
True BG	203 ± 6	12.4 ± 1.6	203 ± 6	12.4 ± 1.6
	SU1		SU4	
Estimated BG	225 ± 9	21.6 ± 1.1	2366 ± 102	165 ± 12.7
True BG+SUSY	463 ± 7	194 ± 4	3177 ± 79	415 ± 29
	SU2		SU6	
Estimated BG	200 ± 9	10.9 ± 0.7	213 ± 9	16.3 ± 0.9
True BG+SUSY	249 ± 7	34 ± 2	365 ± 9	129 ± 5
	SU3		SU8	
Estimated BG	296 ± 10	33.3 ± 1.4	206 ± 9	13.7 ± 0.8
True BG+SUSY	653 ± 8	245 ± 4	354 ± 8	115 ± 5

Correcting for SUSY signal: "New MT method" If, even for overestimated backgrounds, the presence of a concrete SUSY excess is observed in data, we can try to correct the background estimates.

One possible procedure is described here, referred to as the "new MT method". More advanced implementations of such a correction procedure are possible and should be studied.

The new MT method makes use of the observation that in the one-lepton search mode, the M_T distribution of backgrounds falls off steeply beyond $\sim$ 100 GeV, whereas for many SUSY signal models this distribution falls only slowly. This is illustrated in Figure 4 (right). By making a general ansatz for the shape of the SUSY M_T distribution, and neglecting to first order the Standard Model background at high M_T, the SUSY contamination can be subtracted from the control sample. Obviously, remaining Standard Model background in the high M_T region and variations in the M_T shape for various SUSY signals are to be treated as systematic uncertainties on the method. Nevertheless, the data itself will tell what the M_T shape is.

In the simplest ansatz used here, the ratio of SUSY signal between the control sample $M_T < 100$ GeV and signal region $M_T > 100$ GeV is assumed to be constant for all SUSY signal samples. The normalization factor is obtained from the number of events in the signal region and the corrected control sample in the interval 100 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 150 GeV (instead of 100 – 200 GeV) to suppress the SUSY contribution in the normalization region. The statistical error becomes relatively larger when the narrow band is used for normalization, but the over-estimation of the normalization factor due to the SUSY signal can be suppressed. A lower $E_{\mathrm{T}}^{\mathrm{miss}}$ region, such as $E_{\mathrm{T}}^{\mathrm{miss}} = 70 - 100$ GeV, could be used for the normalization in future studies.

Figure 5 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the effective mass distributions of the estimated background processes. The true distributions of the background processes are also superimposed. The numbers in regions of $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV and 300 GeV are listed in Table 4. A reasonable agreement between the prediction and the true values is observed. For high values of $E_{\mathrm{T}}^{\mathrm{miss}}$, the method tends to subtract too much SUSY contamination and underestimates the background. More study is needed. The SU4 benchmark point is a special case because it has a particularly light SUSY particle spectrum.

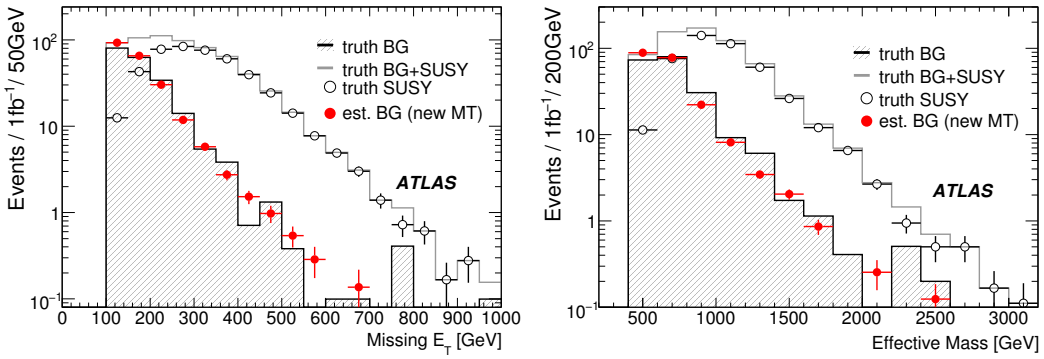


Figure 5: The $E_{\mathrm{T}}^{\mathrm{miss}}$ and effective mass distributions of the background processes for one lepton mode with an integrated luminosity of 1 fb^{-1}. The red dots show the estimated distributions with the "new M_T" method. The hatched histogram show the sum of all Standard Model backgrounds. The open circles indicate the SUSY (SU3) signal.

The systematic uncertainties[3)] for the MT method are summarized in Table 5. As well as variation of jet energy scale and lepton identification efficiency, the ALPGEN Monte Carlo was compared to MC@NLO, and parameters in ALPGEN (minimum p_T of partons and minimum ΔR between partons) were varied. This method is stable against these systematic uncertainties at the $\sim$ 15% level. More work is needed to estimate the SUSY contamination effects.

[3)] Throughout this note systematic uncertainties have been calculated according to the procedures outlined in the introduction to this chapter [6].

Table 4: Numbers of background events and estimated numbers for all background processes in the presence of various SUSY signals, using the new MT method. The numbers are normalized to 1 fb^{-1}.

	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV
True BG	203 ± 6	12.4 ± 1.6	203 ± 6	12.4 ± 1.6
	SU1		SU4	
Estimated BG	186 ± 11	8.9 ± 0.8	1382 ± 98	48.3 ± 12.7
True BG+SUSY	463 ± 7	194 ± 4	3177 ± 79	415 ± 29
	SU2		SU6	
Estimated BG	183 ± 11	8.8 ± 0.8	185 ± 11	8.1 ± 0.9
True BG+SUSY	249 ± 7	34 ± 2	365 ± 9	129 ± 5
	SU3		SU8	
Estimated BG	212 ± 11	12.3 ± 1.0	180 ± 11	6.6 ± 0.8
True BG+SUSY	653 ± 8	245 ± 4	354 ± 8	115 ± 5

Table 5: Systematic uncertainties of the one-lepton background estimations with the MT method, excluding those related to SUSY signal contamination. Numbers are normalized to 1 fb^{-1}

	Syst. error
Jet energy scale	$< 5\%$
Lepton ID efficiency	7%
MC@NLO vs ALPGEN	8%
Monte Carlo parameter variation (ALPGEN)	$< 5\%$

2.3.2 Topbox: a control sample for semileptonic top-pair background

Top mass reconstruction and "topbox" cuts This section describes a data-driven method, denoted the "topbox method", for estimating the $t\bar{t}$ background where one top decays leptonically, and the other hadronically.

For semileptonic $t\bar{t}$ events, the invariant mass of the leptonically decaying W boson can usually be reconstructed by assuming that the neutrino from the W decay is responsible for all missing energy. This is a fair assumption; after removal of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ (noisy/dead calorimeter cells etc.) in the event-cleaning procedure, the resolution on $E_{\mathrm{T}}^{\mathrm{miss}}$ is expected to be approximately equal to $0.55\sqrt{\sum E_T}$ [1], which is much smaller in a typical $t\bar{t}$ event than the $E_{\mathrm{T}}^{\mathrm{miss}}$ from the escaping neutrino. The fact that the mass of the leptonically decaying top can be reconstructed satisfactorily (see below) further justifies the assumption.

The core of the method is to construct both the semileptonic and the hadronic top decays in a $t\bar{t}$ event following the procedure below:

- The leptonic W is assumed to decay into the observed lepton and a neutrino which is responsible for all missing energy. The p_x and p_y components of the neutrino momentum are hence taken to be the x and y components of $E_{\mathrm{T}}^{\mathrm{miss}}$. The p_z component of the neutrino can be calculated using a W mass (m_W) constraint. The four-vector of the leptonic W is the sum of the four-vectors of the lepton and the reconstructed neutrino. For events with transverse mass M_T less than m_W , two solutions can be found. In the case of $M_T > m_W$ no real solution is possible and, in such cases, the momentum of the leptonic W is taken from the transverse components of the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$.

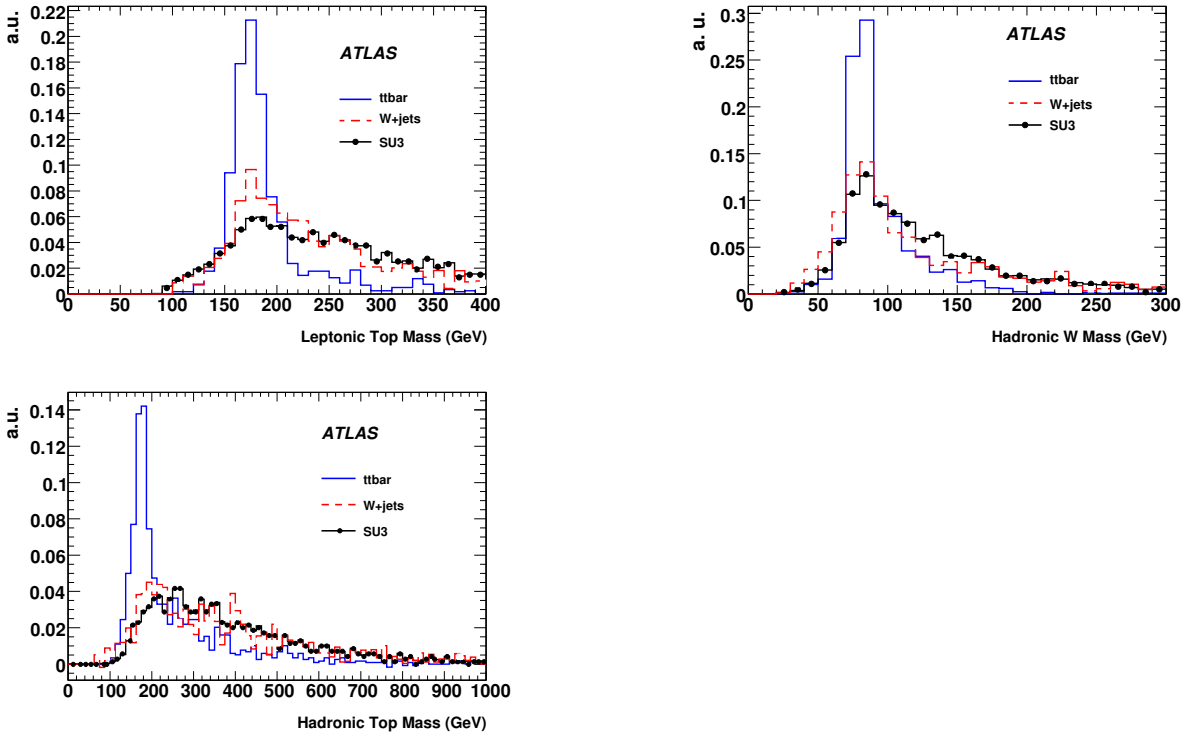


Figure 6: Normalized distributions for reconstructed $m_{\text{top-lep}}$, $m_{\text{W-had}}$, and $m_{\text{top-had}}$ for $t\bar{t}$, W + jets, and SU3 SUSY events, using the "topbox" method.

- The leptonic top is then reconstructed by taking the solution with the best reconstructed top mass ($m_{\text{top-lep}}$) from combinations of a jet and one of the above leptonic W solutions. The jet is taken from the pool of the four highest-p_T jets in the event. The best reconstructed top mass is defined to the one that is closest to the nominal top mass m_t .

- The hadronic W is then taken to be formed from the best reconstructed W mass ($m_{\text{W-had}}$) among the two-jet combinations from the remaining three jets in the pool. The best reconstructed W mass is defined to be the invariant closest to m_W .

- Finally, the hadronic top is taken to be the one with the best reconstructed top mass ($m_{\text{top-had}}$) among combinations of the hadronic W and one of the remaining jets.

The plots in Figure 6 show the distributions of the reconstructed masses $m_{\text{top-lep}}$, $m_{\text{W-had}}$, and $m_{\text{top-had}}$ after the mass reconstruction procedure described above. The distributions are made for $t\bar{t}$, SU3 and $W+$jets event samples with standard one-lepton cuts, except for a modified M_T requirement (see below in the control sample section). As expected, the topbox mass reconstruction procedure offers a very good separating power between $t\bar{t}$ and other processes.

The topbox cuts are then defined as follows: $|m_{\text{top-lep}} - m_t| < 25$ GeV, $|m_{\text{W-had}} - m_W| < 15$ GeV, and $|m_{\text{top-had}} - m_t| < 25$ GeV.

Topbox control sample To make the topbox control sample, events are selected with the standard SUSY search cuts in the one-lepton mode, with the exception that $M_T > 100$ GeV is replaced by $M_T < m_W$. In addition, the above topbox cuts are applied.

Table 6 shows the number of events of various processes in the topbox control sample. The $t\bar{t}+$jets process makes up more than 95% of the topbox control sample if no SUSY signal is present.

Table 6: Composition of the topbox control sample. Numbers shown correspond to an integrated luminosity of 1 fb^{-1} . The last five columns show the numer of SUSY events which would enter into the topbox control sample.

Process	$t\bar{t}+$jets	$W+$Jets	SU1	SU2	SU3	SU4	SU6
Events	340.9	6.8	1.8	0.4	4.9	243.6	0.4

SUSY signal contamination Table 6 also shows the number of SUSY events, for various signal samples, in the topbox control sample, for 1 fb^{-1}. In this method, SUSY contamination is in general small. This fact makes the topbox method a good supplement to the other methods (e.g. the MT method). The exception is the SU4 benchmark point, which has a larger contribution because its light spectrum makes it rather similar to the $t\bar{t}$ background.

Estimation of the $t\bar{t}$ background in the signal region The $t\bar{t}$ contamination in the signal region is estimated by multiplying the number of events in the data topbox by a scaling factor R_{tt} . R_{tt} is defined as the ratio of the number of Monte Carlo $t\bar{t}$ events in the signal region (those that pass the one-lepton cuts) to that in the topbox control sample. The procedure is summarized by the following equations:

$$N_{t\bar{t}}^{\text{signal-region}}(\text{data}) = N_{t\bar{t}}^{\text{topbox}}(\text{data}) \cdot R_{tt} \tag{1}$$

$$R_{tt} \equiv N_{t\bar{t}}^{\text{signal-region}}(\text{MC}) / N_{t\bar{t}}^{\text{topbox}}(\text{MC}) \tag{2}$$

With fully simulated Monte Carlo samples, R_{tt} is determined to be 0.386 . The model dependence (variation of Monte Carlo generator and generator parameters) of this number is treated as a systematic uncertainty.

Systematics The systematic uncertainties of the topbox method are summarized in Table 7. The largest source of uncertainty is from the jet energy scale uncertainty; this is expected since the method relies heavily on the reconstruction of top and W masses. The Monte Carlo model dependency of R_{tt} is estimated by comparing MC@NLO and ALPGEN, and by variation of the ALPGEN parameters, and amounts to 8%. Finally, it is expected that extra jets due to event pile-up may affect the mass reconstruction resolution. However, this is relevant only in high luminosity scenarios, beyond the scope of this note. The statistical uncertainty on the topbox control sample normalization is estimated to be 5% for 1 fb^{-1} given that the effective cross-section of $t\bar{t}$ in the topbox is about 400 fb.

Table 7: Systematic uncertainties of the topbox method for 1 fb^{-1}.

Source	Contribution [%]
Jet energy scale	20
$E_{\text{T}}^{\text{miss}}$ scale	2
Monte Carlo Model dependence of R_{tt}	8
Total	22

2.3.3 Di-leptonic top with one lepton missed: kinematic reconstruction

Introduction Fully leptonic $t\bar{t}$ events may contribute to the one-lepton SUSY search sample if one of the two leptons originating from the W decay is not identified. Such events can be classified as: (1) events with one tau (51%); (2) events where one lepton is misidentified due to inefficiency of the lepton identification algorithms (20%); (3) events where one lepton is lost inside a jet (17%); (4) events where one lepton is not in the p_T or η acceptance (9%); and (5) events with two tau leptons (3%).

The method discussed here is based on the selection of a sample enhanced in $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events by requiring that the events satisfy a set of kinematic constraints particular to the $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ process. This sample, denoted as the control sample, with two isolated identified leptons, is used to estimate the contribution from the first two categories of events listed above. The contribution from category (1) is estimated by replacing one of the leptons in the control sample with a tau, and category (2) is estimated by removing one of the two leptons. The contribution from the categories (3)–(5) is not estimated from the control sample. Events were required to fire either the 4j50 multi-jet trigger or the j80_xE50 jet plus $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger [7].

Selection of the control sample The following requirements are imposed to select events in the control sample: two isolated oppositely-charged leptons (electron or muon), with $p_{\mathrm{T}} > 10$ GeV and at least one with $p_{\mathrm{T}} > 20$ GeV; at least three jets with $|\eta| < 2.5$ and $p_{\mathrm{T}} > 50$ GeV at least one of which must have $p_{\mathrm{T}} > 100$ GeV. Note that in contrast to the SUSY one-lepton search selection given in Sec. 2.1 only three jets are required, since the misidentified lepton or tau can produce the fourth jet.

For $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events the two leptons, two b jets and the x- and y-components of the $E_{\mathrm{T}}^{\mathrm{miss}}$ -vector satisfy the following kinematic constraints:

$$\begin{aligned}
(p_\nu + p_{\ell^+})^2 &= m_W^2, \\
(p_{\bar{\nu}} + p_{\ell^-})^2 &= m_W^2, \\
(p_\nu + p_{\ell^+} + p_b)^2 &= m_t^2, \\
(p_{\bar{\nu}} + p_{\ell^-} + p_{\bar{b}})^2 &= m_t^2, \\
p_{\nu_x} + p_{\bar{\nu}_x} &= E_{\mathrm{T,x}}^{\mathrm{miss}}, \\
p_{\nu_y} + p_{\bar{\nu}_y} &= E_{\mathrm{T,y}}^{\mathrm{miss}},
\end{aligned} \tag{3}$$

where $p_{\ell^\pm}$, $p_{\nu/\bar{\nu}}$, $p_{b/\bar{b}}$ are the lepton, neutrino and b-quark momenta respectively and m_W and m_t are the W boson and top quark masses. We assume that the only source of $E_{\mathrm{T}}^{\mathrm{miss}}$ is a pair of neutrinos, which is a fair assumption as shown in the previous section.

The final state contains two unknown neutrino momenta and the above system of equations has a two- or four-fold ambiguity, as the solution is given by a quartic equation which can be solved with standard analytical techniques [8]. Since there are at least three jets in each event, all possible combinations of jet pairs made from the three highest p_T jets are considered. Jet pairs for which the above system of equations has real solutions are denoted as b-jet pairs[4)]. Figure 7 (left) shows the number of b-jet pairs for the various processes contributing to the control sample.

Replacement procedure Each event in the control sample is used as a seed for producing a series of resimulated events. One of the two identified leptons in the seed event is replaced by tau lepton and a set of 1000 tau decays are simulated using the TAUOLA package [9]. The same procedure is repeated for the

[4)] Note that within this section only kinematical conditions have used to identify these b-jet pairs – no secondary-vertex requirement is used.

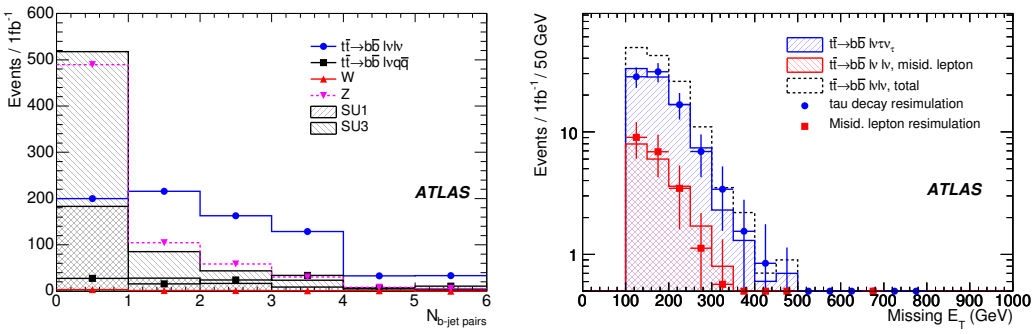


Figure 7: Left: distribution of number of b-jet pairs for events passing the control sample requirements in the kinematic reconstruction method. The fraction of $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events with no b-jet pairs is dominated by events with at least one b jet which is not among the three highest-p_T jets. Right: distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ for $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events with one tau lepton and events with a misidentified lepton compared to the estimation from resimulated events with an integrated luminosity of 1 fb^{-1}. The requirement on the number of b-jet pairs is not applied to the resimulated events. The distribution of all $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events is also shown.

second lepton in the seed event, yielding a total of 2000 events for every seed event. Each resimulated event is weighted by a factor of $1/\varepsilon$, where ε, the identification efficiency for the replaced lepton, is estimated from simulations.

The contribution of events where one lepton evades identification is estimated as follows. If the replaced lepton is an electron then a jet with the same momentum is substituted instead of it. If the lepton is a muon it is replaced by a so-called stand-alone muon (defined as a track in the muon spectrometer with no match to a track in the inner detector) justified by the fact that most muons not passing the muon definition are stand-alone muons. This procedure is applied to each of the two leptons in the seed events, resulting in two resimulated events for each seed event. The resimulated events are re-weighted with $\frac{1-\varepsilon}{\varepsilon}$.

For both kinds of resimulated events, the SUSY one-lepton search selection are subsequently applied.

As a closure test of the replacement procedures described above, the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for resimulated $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events passing the control sample selection apart from the requirement of b-jet pairs, is compared to the Monte Carlo prediction. The result is shown in Fig. 7 (right) and shows good agreement.

Normalization The number of $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events in the signal region is estimated by scaling of the sum of described above contributions with two scaling factors. The first factor takes into account the other categories of $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events that are not estimated by this method. This first factor is estimated from Monte Carlo to be $R^{\mathrm{MC}} = 1.4 \pm 0.1$. The second normalization factor, $R^{b\text{-jet pair}}$, takes into account the efficiency of $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events to pass the requirement on the number of b-jet pairs; it is defined as the ratio of resimulated events before and after the b-jet pair selection in a normalization region, $80 \leq E_{\mathrm{T}}^{\mathrm{miss}} \leq 120$ GeV, and found to have the value $R^{b\text{-jet pair}} = 1.4 \pm 0.1(\mathrm{stat}) \pm 0.1(\mathrm{syst})$.

Presence of SUSY A possible SUSY signal could have an effect on the background estimation in two ways: 1) by satisfying the kinematic constraints in Eq. 3 and therefore enter the control sample and 2) by entering the normalization region giving a systematic contribution to the scale factor $R^{b\text{-jet pair}}$. In Fig. 8 the estimated $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ background is shown with and without the contamination of a SUSY signal

Table 8: Estimated background corresponding to an integrated luminosity of 1 fb^{-1} for different mSUGRA benchmark points. The second column shows the relative increase of the estimated background with respect to the estimation without contamination from the SUSY signal. The third column shows the number of SUSY events. The Monte Carlo prediction of $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ background in the one lepton search mode is 136 events. The errors in the first column are statistical only.

SUSY point	Estimated Background	Relative change [%]	True Signal Events
No signal	120 ± 14		
SU1	137 ± 15	15	260
SU2	127 ± 15	5.9	45
SU3	176 ± 18	47	454
SU4	604 ± 38	405	2960
SU6	129 ± 16	7.8	162
SU8	124 ± 14	3.8	100

(SU3) while Tab. 8 gives the estimated number of $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ events in the presence of different SUSY signals.

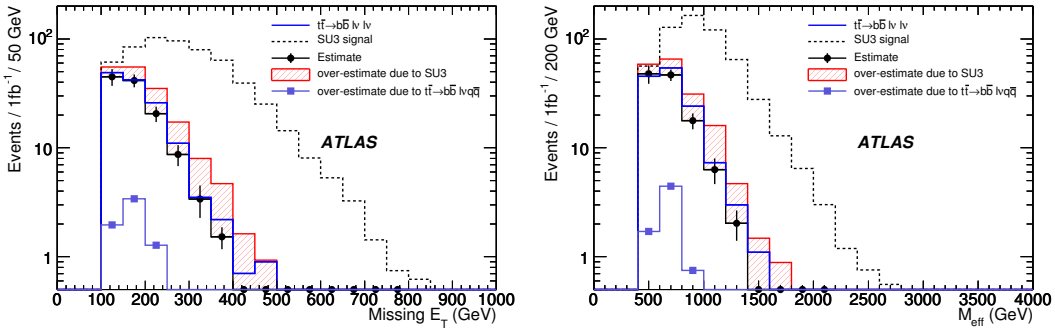


Figure 8: The $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and M_{eff} (right) distributions for the estimated and true $t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ contribution for the one-lepton SUSY search. Black points (red area) represent the estimation without (with) the presence of a signal from SUSY (SU3).

Systematic Uncertainties The systematic uncertainties for this method are summarized in Tab. 9. The uncertainty from the replacement procedure is estimated by comparing number of resimulated events to the Monte Carlo prediction, see Fig. 7(right). The uncertainty of R^{MC} is estimated by comparing MC@NLO and ALPGEN. The statistical uncertainty of $R^{b\text{-jet pair}}$ is calculated using binomial errors. The systematic uncertainty of this factor takes into account the difference in the shapes between $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution of the resimulated samples with and without applying the kinematical constraints in Eq. 3. The uncertainty due to background subtraction is dominated by the presence of $t\bar{t} \to b\bar{b}q\bar{q}\ell\nu$ events in the control sample. The systematic effects resulting from uncertainties in the lepton identification efficiency, the trigger efficiencies and the energy scale and resolution are expected to be much smaller.

Table 9: Breakdown of systematic uncertainties in the kinematic reconstruction method.

Source	Contribution [%]
Replacement	10
R^{MC}	10
Jet Energy Scale	9
$R^{b\text{-jet pair}}$ stat.	9
$R^{b\text{-jet pair}}$ syst.	8
Background Subtraction	3
Jet Energy Resolution	1
$E_{\mathrm{T}}^{\mathrm{miss}}$ scale	1
Total	21

2.3.4 Dileptonic top with one lepton missed: HT2

Introduction In this section we describe a method, denoted the "HT2 method", to estimate background from dileptonic $t\bar{t}$ production where one of the leptons is not identified. It relies on the (near) independence of $E_{\mathrm{T}}^{\mathrm{miss}}$ and the variable HT2. This variable is defined as:

$$\mathrm{HT2} \equiv \sum_{i=2}^{4} p_T^{\mathrm{jet}\,i} + p_T^{\mathrm{lepton}}. \tag{4}$$

In the HT2 method, the shape of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is estimated from dileptonic $t\bar{t}$ events with low HT2. This distribution is then normalized to the number of events at large HT2, but with low missing E_{T} , and can then be used to estimate the remaining backgrounds in the signal region of large HT2 and large $E_{\mathrm{T}}^{\mathrm{miss}}$.

For this method to work, the shape of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution needs to be independent of HT2. Note that in Equation 4, the leading jet p_{T} was excluded from the sum in order to reduce the correlation with $E_{\mathrm{T}}^{\mathrm{miss}}$. The correlation between the hightest-p_{T} jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ is likely to be due to simple kinematics, i.e. to first approximation, the rest of the event recoils against this leading jet. This is illustrated in Figure 9 which shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution (at Monte Carlo "truth level") in slices of leading and sub-leading jet p_{T} . The reduced dependence of the $E_{\mathrm{T}}^{\mathrm{miss}}$ shape on the jet p_{T} in the second-leading jet case is apparent, and will be further diminished by detector resolution effects.

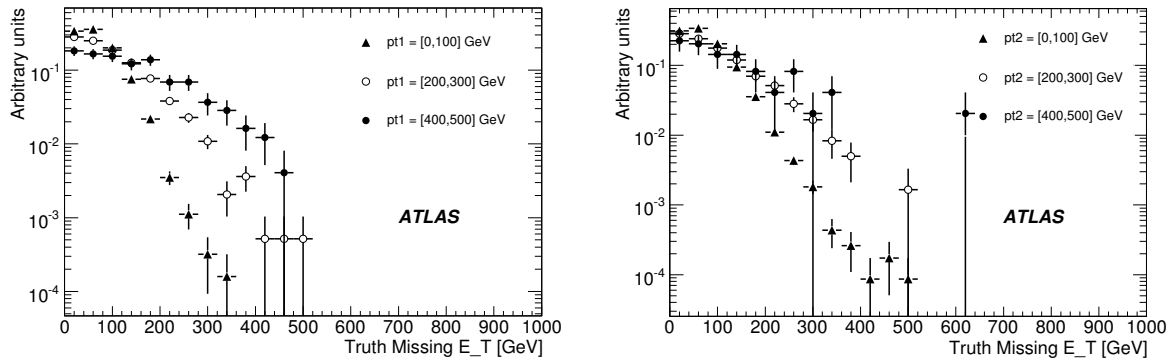


Figure 9: Missing E_{T} distribution in "lepton+jet" $t\bar{t}$ events with $M_T > 100$ GeV at Monte Carlo "truth" level. Left: as a function of truth leading-jet p_{T} . Right: as a function of truth second-leading jet p_{T} .

To further reduce the correlation between HT2 and $E_{\mathrm{T}}^{\mathrm{miss}}$, the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance was used. This is

to remove the correlation which arises from the fact that the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution depends on $\sum E_{\mathrm{T}}$, where $\sum E_{\mathrm{T}}$ is clearly related to HT2. A simple form of $E_{\mathrm{T}}^{\mathrm{miss}}$ significance was used here, defined as $E_{\mathrm{T}}^{\mathrm{miss}}$ significance = $E_{\mathrm{T}}^{\mathrm{miss}}$ /$[0.49 \cdot \sqrt{\sum E_{\mathrm{T}}}]$.

The results shown here are from a data sample consisting of the sum of $t\bar{t}$ (semi-leptonic and dileptonic decay modes) plus W($l\nu$)+jets (where $l = e, \mu, \tau$). The trigger used in this analysis was the logical OR of the 4j50 multi-jet, the e22i single electron and the mu20 single muon triggers [7].

A control sample defined by HT2 < 300 GeV was used to estimate the shape of the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance. The assumption is that this shape is independent of HT2 so it can be used to predict the shape of the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance in the signal "band" defined by HT2 > 300 GeV. The normalization of the prediction in the signal band was obtainined by the number of events with HT2 > 300 GeV, but at low $E_{\mathrm{T}}^{\mathrm{miss}}$, specifically $8 <$ $E_{\mathrm{T}}^{\mathrm{miss}}$ significance < 14. A comparison of this predicted background with the correct background is shown in Figure 10 (left). The agreement between the predicted background and the actual background in Fig. 10 is reasonable, indicating that the correlation between HT2 and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is small. A numerical comparison of predicted and actual background levels can be seen in Table 10. For each value of the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance cut, a rough equivalent in $E_{\mathrm{T}}^{\mathrm{miss}}$ is listed as a guide, but it should be emphasized that the cut in $E_{\mathrm{T}}^{\mathrm{miss}}$ is not sharp. The number of events is for HT2 > 300 GeV, which corresponds approximately to a cut on the effective mass of $M_{\mathrm{eff}} > 600$ GeV.

Table 10: Predicted and actual background levels as a function of $E_{\mathrm{T}}^{\mathrm{miss}}$ significance cut for an integrated luminosity of 1 fb^{-1} in the HT2 analysis. A rough equivalent $E_{\mathrm{T}}^{\mathrm{miss}}$ cut is listed, but the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut is not sharp.

$E_{\mathrm{T}}^{\mathrm{miss}}$ sig. cut	Rough equivalent $E_{\mathrm{T}}^{\mathrm{miss}}$ cut [GeV]	Predicted BG	Actual BG
14	180	57.3 ± 5.5	60.6 ± 3.2
16	200	34.8 ± 4.5	39.2 ± 2.6
18	220	19.1 ± 3.1	23.6 ± 2.0
20	240	10.1 ± 2.1	15.1 ± 1.5
22	260	6.2 ± 1.8	9.8 ± 1.2
24	280	3.8 ± 1.5	6.2 ± 0.9
26	300	1.3 ± 0.7	3.5 ± 0.6

The ratio of observed to predicted backgrounds for a $E_{\mathrm{T}}^{\mathrm{miss}}$ significance cut of 14 is 1.06 ± 0.12; while the ratio is consistent with unity, we take the uncertainty on the ratio (12%) as a systematic uncertainty due to possible correlations between HT2 and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance. Monte Carlo samples with larger numbers of events would provide one possible way to further study the potential for correlations.

The distribution of the "orthogonal" variable, namely HT2, was predicted in a similar way. The HT2 distribution was measured in a control region defined by $8 <$ $E_{\mathrm{T}}^{\mathrm{miss}}$ significance < 14. This distribution was then normalized to the number of events at large $E_{\mathrm{T}}^{\mathrm{miss}}$ significance and low HT2, specifically, $E_{\mathrm{T}}^{\mathrm{miss}}$ significance > 14, and 150 GeV $<$ HT2 $<$ 300 GeV. The results are shown in Fig. 10 (right).

The near independence of HT2 and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance should provide an important tool in understanding jet energy and $E_{\mathrm{T}}^{\mathrm{miss}}$ performance in the complex events that make up the background to SUSY searches. After all the SUSY selection cuts have been applied, the jet energy performance can be studied by looking at the HT2 distribution for low $E_{\mathrm{T}}^{\mathrm{miss}}$ events; conversely, the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution can be studied by selecting events with low HT2. Events in the tails of these distributions can be examined for signs of detector problems.

Systematic uncertainties due to detector miscalibrations The results of systematic uncertainties due to detector performance are summarized in Table 11. The energy scale variations change the background

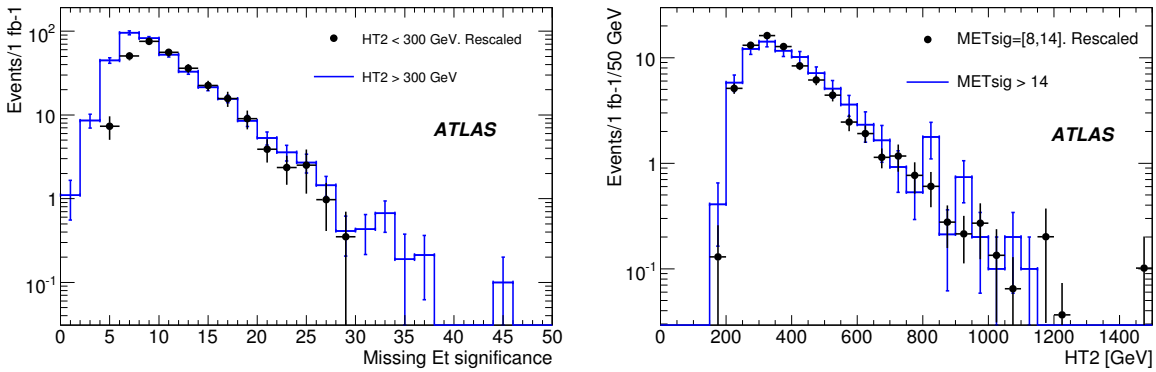


Figure 10: Left: Points: Predicted $E_{\mathrm{T}}^{\mathrm{miss}}$ significance distribution in a $t\bar{t}$ plus $W+$jets sample. Histogram: actual $E_{\mathrm{T}}^{\mathrm{miss}}$ significance distribution. Right: Predicted HT2 distribution in the same sample. Histogram: actual HT2 distribution.

Table 11: Predicted and actual background levels (for 1 fb^{-1}, HT2 method) for $E_{\mathrm{T}}^{\mathrm{miss}}$ significance > 14 as a function of systematic effects applied to the reconstructed objects.

Modification	Predicted BG	Actual BG	Actual/predicted
Baseline	57.3 ± 5.5	60.6 ± 3.2	1.05 ± 0.12
Energy scaled up	64.1 ± 5.5	79.3 ± 3.7	1.24 ± 0.12
Energy scaled down	45.5 ± 4.5	47.3 ± 2.7	1.04 ± 0.12
Jet resolution smearing	55.5 ± 5.1	65.3 ± 3.4	1.18 ± 0.12

level by about 30% while the worsening jet energy resolution results in about a 10% increase in background. However the predictions tend to change in the same direction as the actual backgrounds, and generally continue to provide reasonable determinations. We assign a 20% systematic uncertainty due to detector effects.

Systematic uncertainties due to event generation parameters The systematic uncertainties in the method due to changes in Monte Carlo event generation parameters were studied with ALPGEN. The parton p_T cut in ALPGEN was changed from 40 to 15 GeV and the renormalization scale was reduced by a factor of 2. The results of the studies are summarized in Table 12. We assign a 20% systematic uncertainty due to event generation uncertainties.

Table 12: Predicted and actual background levels (for 1 fb^{-1}, HT2 method) for $E_{\mathrm{T}}^{\mathrm{miss}}$ significance > 14 as a function of changes in the Monte Carlo generation parameters.

Modification		Predicted BG	Actual BG	Actual/predicted
$t\bar{t}$	$W+$jets			
PT40, scale 1.0	PT40, scale 0.5	73.3 ± 5.8	63.9 ± 3.2	0.87 ± 0.11
PT40, scale 0.5	PT40, scale 0.5	133.8 ± 7.2	109.2 ± 3.6	0.82 ± 0.05
PT15, scale 1.0	PT40, scale 0.5	91.1 ± 12.6	72.5 ± 6.0	0.80 ± 0.13

Background estimation in the presence of SUSY In this section, we repeat the background estimation in the presence of SUSY signal. Figure 11 (left) shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance distributions for the true background, true signal, and the estimated background, as well as the observed distribution of signal plus background. The SUSY signal here is the so-called "1 TeV SUSY" point ($m_0 = m_{\frac{1}{2}} = 400$ GeV, $\tan\beta$ = 10, A=0, $\mu > 0$).

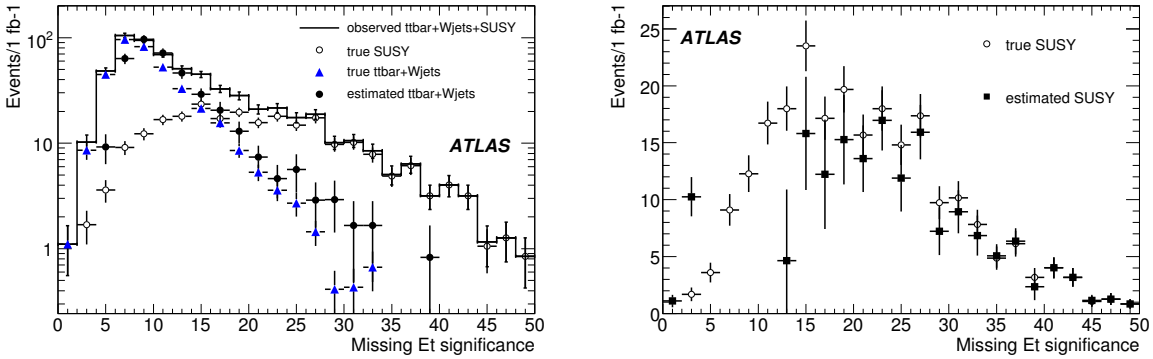


Figure 11: Left: Histogram: observed $E_{\mathrm{T}}^{\mathrm{miss}}$ significance distribution for the sum of $t\bar{t}$ plus $W+$jets background plus SUSY signal. Open circles: SUSY signal. Blue triangles: true $t\bar{t}$ plus $W+$jets background. Black filled circles: estimated background. The SUSY signal shown here is the "1 TeV SUSY" point (see text). Right: Open circles: true SUSY signal as a function of $E_{\mathrm{T}}^{\mathrm{miss}}$ significance. Black: estimated SUSY yield, obtained from the difference of the observed $E_{\mathrm{T}}^{\mathrm{miss}}$ significance distribution minus the estimated background distribution.

Because of the signal contamination in the control region, the background level is overestimated, leading to an underestimation in the excess of signal over background. Nevertheless, it is clear that by cutting harder on $E_{\mathrm{T}}^{\mathrm{miss}}$ significance, for example, the signal can still be clearly seen over the estimated background. A comparison of the estimated signal yield to the true signal is shown in Figure 11 (right).

The results for all the tested SUSY points are summarized in Table 13.

2.3.5 Top background estimation with top redecay simulation

Introduction It is possible to isolate a pure biased sample of fully-leptonic $t\bar{t}$ events by selecting low $E_{\mathrm{T}}^{\mathrm{miss}}$ (to reduce SUSY signal) opposite sign dilepton events where one and only one pair of invariant mass combinations $m_{\ell j}$ between the two leptons and two hardest jets (b jets if tagging available) gives values below the expected endpoint from $t \to Wb \to \ell\nu b$ decays: $m_{\ell j}^{\max} = \sqrt{m_t^2 - m_W^2}$ (neglecting m_b).

A possible use of such a sample is to estimate the background of fully-leptonic $t\bar{t}$ events to SUSY searches. One can reconstruct the kinematics of the decaying particles (W's or top quarks), remove their inferred decay products from the reconstructed event (including the event $E_{\mathrm{T}}^{\mathrm{miss}}$), redecay the reconstructed W's or top quarks using an event generator (e.g. PYTHIA) and then merge the simulated re-decay products back into the parent ('seed') event. By redecaying particles earlier in the decay chain (i.e. the top rather than the W) the kinematic bias obtained from the event selection can be minimised. This technique has a number of advantages over conventional Monte Carlo techniques. In particular the event generator is used purely for modelling relatively well-understood decay and hadronisation processes – initially poorly understood aspects of process generation, such as parton distributions and the underlying event model, are effectively obtained from the data. In principle this technique is applicable also to other background processes such as $Z \to \tau^+\tau^-$, which could be modelled by replacing identified electrons or muons in $Z \to \ell^+\ell^-$ control sample events with redecayed taus.

Table 13: True and estimated background and signal, using the HT2 method, when the background estimation is performed in the presence of SUSY signal. The numbers are for an integrated luminosity of 1 fb^{-1} except for the SU4 point where 100 pb^{-1} is used.

SUSY point	$E_{\mathrm{T}}^{\mathrm{miss}}$ sig. cut	True BG	Est. BG	True signal	Est. signal	True $S/\sqrt{B}$	Estimated $S/\sqrt{B}$
SU1	16	39.2 ± 2.6	100.5 ± 10.4	219.7 ± 8.7	158.4 ± 13.8	35.1	15.8
	20	15.1 ± 1.5	53.1 ± 7.8	167.0 ± 7.6	128.9 ± 11.0	43.0	17.7
	24	6.2 ± 0.92	33.1 ± 6.5	120.8 ± 6.4	93.8 ± 9.2	48.6	16.3
SU2	14	60.6 ± 3.2	69.1 ± 6.4	30.4 ± 2.3	21.9 ± 7.5	3.9	2.6
	16	39.2 ± 2.6	43.1 ± 5.3	24.0 ± 2.1	20.2 ± 6.2	3.8	3.1
	18	23.6 ± 2.0	24.1 ± 3.7	18.3 ± 1.8	17.9 ± 4.6	3.8	3.6
	20	15.1 ± 1.5	13.9 ± 2.7	13.5 ± 1.6	14.7 ± 3.5	3.5	3.9
SU3	16	39.2 ± 2.6	198.1 ± 22.5	328.1 ± 14.9	169.2 ± 27.2	52.4	12.0
	20	15.1 ± 1.5	119.9 ± 18.5	228.9 ± 12.5	124.1 ± 22.4	59.0	11.3
	24	6.2 ± 0.92	62.9 ± 13.7	144.7 ± 9.9	88.0 ± 16.9	58.3	11.1
SU4	16	3.92 ± 0.26	120.7 ± 8.7	76.4 ± 4.0	-40.4 ± 9.6	38.6	-3.7
	20	1.51 ± 0.15	47.4 ± 5.5	37.4 ± 2.8	-8.5 ± 6.1	30.4	-1.2
	24	0.62 ± 0.09	17.8 ± 3.3	18.8 ± 2.0	1.6 ± 3.9	23.9	0.4
SU6	16	39.2 ± 2.6	71.5 ± 7.2	140.5 ± 5.3	108.2 ± 9.3	22.4	12.8
	20	15.1 ± 1.5	36.5 ± 5.0	108.8 ± 4.7	87.4 ± 7.0	28.0	14.5
	24	6.2 ± 0.92	25.1 ± 4.3	79.3 ± 4.0	60.3 ± 6.0	31.9	12.0
1 TeV	16	39.2 ± 2.6	61.1 ± 6.8	155.0 ± 5.7	133.1 ± 9.2	24.7	17.0
	20	15.1 ± 1.5	27.6 ± 4.4	118.1 ± 5.0	105.6 ± 6.8	30.4	20.1
	24	6.2 ± 0.92	15.6 ± 3.5	84.5 ± 4.2	75.1 ± 5.6	34.0	19.0

It should be noted that this technique is at best an approximation, assuming as it does the factorisation of each $t\bar{t}$ event into two independent tops, and hence neglecting effects such as colour connection and spin correlations between the tops and other partons in the event. It is therefore unlikely to be competitive with a detailed Monte Carlo study using a fully tuned generator and validated parton distribution functions. In the early days of data-taking however it potentially provides a route to a rapid direct estimate of $t\bar{t}$ background from data complementary to, and independent from, more conventional estimates.

Seed event selection Seed events were selected from the 'data' with cuts designed to maximise the number of fully leptonic $t\bar{t}$ ('2ℓ-$t\bar{t}$') events while minimising the number of Standard Model backgrounds or SUSY signal events. Events were required to pass the j45_xE50 jet + $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger [7]. Single and dilepton triggers were not included in this study, but are planned to be added in future analyses. Subsequently, the following criteria were applied: $N_{\mathrm{jet}} \geq 2$, $p_T(\mathrm{jet}_2) > 20$ GeV, two Opposite Sign (OS) isolated leptons should be present, $p_T(\ell_2) > 10$ GeV, if the two leptons are of the same flavour $|m_{\ell\ell} - m_Z| > 15$ GeV and $m_{\ell\ell} > 10$ GeV is required, $|m_{\tau\tau} - m_Z| > 15$ GeV, where $m_{\tau\tau}$ is calculated assuming the neutrinos travel parallel to their parents, and $E_{\mathrm{T}}^{\mathrm{miss}} < \frac{1}{2}(p_T(\ell_1) + p_T(\ell_2))$. The upper limit on $E_{\mathrm{T}}^{\mathrm{miss}}$ as a function of lepton p_T rejects SUSY signal events. For the purposes of this early-data study b-tagging was assumed to be either not available or not well-understood.

Kinematic Reconstruction and Redecay Simulation The $m_{\ell j}$ distribution of selected events (the histogram in Figure 12) contains a prominent edge at the expected position $m_{\ell j}^{\mathrm{max}} = 155.4$ GeV ($m_t = 175$ GeV). Events were further selected in which one and only one of the two possible pairs of ℓj combinations obtained from the two leptons and two hardest jets gave $m_{\ell j}$ values which were both less than $m_{\ell j}^{\mathrm{max}}$.

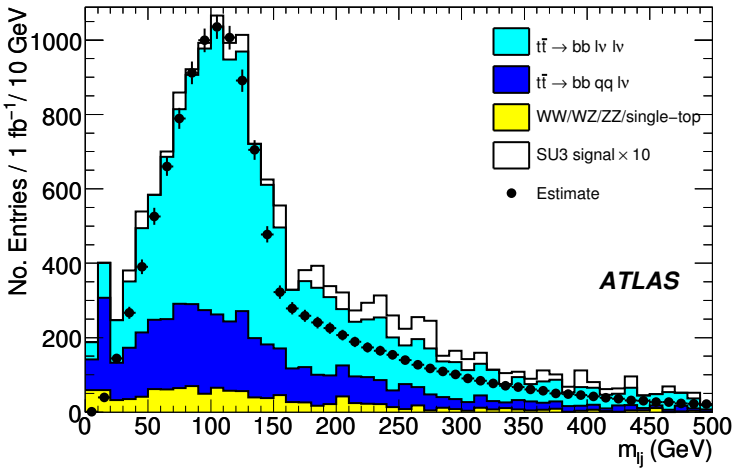


Figure 12: Distributions of $m_{\ell j}$ values for various different Standard Model backgrounds and SUSY signal. The histograms show distributions of 'data' events selected with the 2ℓ-$t\bar{t}$ selection. The data-points show the equivalent distribution estimated with redecay simulation, normalised to the peak.

For 2ℓ-$t\bar{t}$ signal events the two (b) jets, two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ components satisfy the constraints of Eq. 3 given in section 2.3.3. These constraints, assuming massless neutrinos, leptons and jets, may be solved for the 8 unknown 4-momentum components of the neutrinos. The constraints together give a quartic equation which can be solved with standard analytical techniques. If no solution was obtained then to maximise statistics the real part of the least imaginary solution was taken. If multiple solutions were obtained the solution with the smallest mean $|p_z|$ of the reconstructed top was used.

This selection results in 2207 dileptonic $t\bar{t}$ events for 1 fb^{-1}. The contamination by other Standard Model processes and SUSY signal events was 912 events, dominated by semileptonic $t\bar{t}$ events, as shown in Figure 12. The SUSY contamination in the sample is small, due to the tight selection cuts.

Four-vectors of the two reconstructed top quarks from each event were passed to a modified version of PYTHIA 6.4 [10]. 1000 redecayed tops were produced from each reconstructed seed top, with each W forced to decay to e, μ or τ. This 'recycling' of seed events increases the statistics of decay resimulated events for the final $E_{\mathrm{T}}^{\mathrm{miss}}$ estimation process but leads to correlations between resimulated events derived from the same seed event. These correlations were taken into account in the final uncertainties quoted below. Decay products were passed to the ATLAS fast simulation program, and then merged back into their parent seed events.

As a cross-check of the estimation procedure redecayed events were passed through the same selection as their parent seed events and the distribution of $m_{\ell j}$ constructed and normalised to the seed $m_{\ell j}$ distribution. This is shown in Fig. 12 and indicates good agreement below $m_{\ell j}^{\mathrm{max}}$.

Use in one-lepton search background estimate Decay resimulated events were subjected to the standard one-lepton SUSY search selection described in section 2.1, with one modification: $M_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) >$ 150 GeV.

The remaining background at high $E_{\mathrm{T}}^{\mathrm{miss}}$ following such cuts is dominated by semi-leptonic $t\bar{t}$ events, and to a lesser extent leptonic $W+$jets, $Z+$jets, single-top and di-boson events. For all these backgrounds one expects primarily a Jacobian peak in the event $M_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ distribution near M_W (M_Z for $Z+$jets). This 1ℓ-Jacobian background was estimated with the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution of events selected with the same cuts, with the exception of the $M_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ cut, which was reversed to require $M_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) < 100$ GeV.

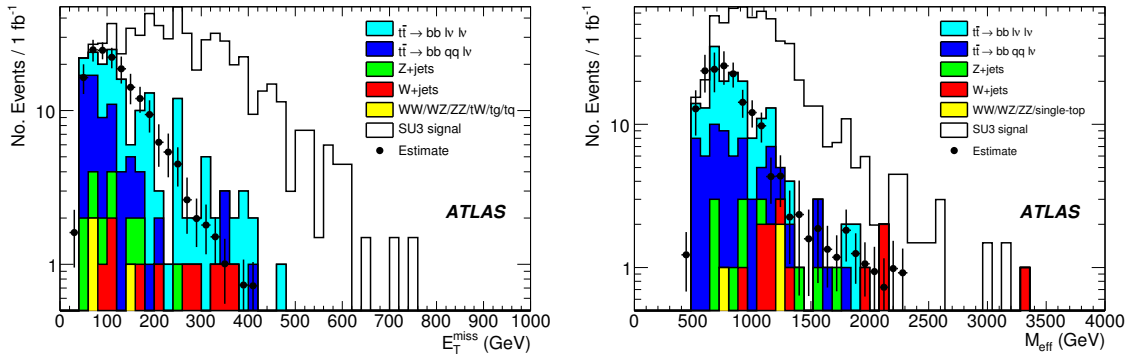


Figure 13: $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and M_{eff} (right) distributions of events passing the basic one-lepton SUSY selection cuts described in the text. Note that $Z+$jet and $W+$jet backgrounds are under-represented in these plots for $E_{\mathrm{T}}^{\mathrm{miss}} < 80$ GeV or $M_{\mathrm{eff}} < 350$ GeV due to filter requirements applied to the respective Monte Carlo samples.

The 1ℓ-Jacobian and the 2ℓ-$t\bar{t}$ background estimates were compared to 'data' events subjected to the one-lepton selection criteria described above. The two estimates were simultaneously normalised to the 'data' $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in two bins: $40 < E_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV and $100 < E_{\mathrm{T}}^{\mathrm{miss}} < 140$ GeV. The total (2ℓ-$t\bar{t}$ $+ 1\ell$-Jacobian) normalised estimate is plotted in Fig. 13(left) together with the 'data' $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. For $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV the agreement between the estimate and the 'data' is good: 30.7 ± 9.8 (30%) estimated, versus 39 'observed'.

In Fig. 13(right), the M_{eff} distribution is shown, with the estimate normalised with the same factors as used in Fig. 13(left). The semi-leptonic $t\bar{t}$ forms a larger fraction of the background at large M_{eff} compared to at large $E_{\mathrm{T}}^{\mathrm{miss}}$ because it produces a larger number of jets than fully-leptonic $t\bar{t}$.

SUSY contamination The shape of the estimated distribution is effectively insensitive to the presence of SUSY signal, primarily due to the low $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement in the 2ℓ-$t\bar{t}$ selection.

However, there may be SUSY signal in the normalisation region. The bias in the estimate will be proportional to the amount of SUSY in the normalization region, which is largest for the samples with the highest SUSY cross-section: SU3 and SU4. The effect of admixture of SU3 or SU4 signal events is shown in Table 14. In case of SU3, the background estimate is 60% higher, in case of SU4 as large as a factor 15. For both samples, however, the excess of signal events is still significant.

In principle the contamination effect could be reduced by normalization to more signal-free regions. The 2ℓ-$t\bar{t}$ and 1ℓ-Jacobian estimates could be normalized for example to the tail and peak of the $M_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ distribution at low $E_{\mathrm{T}}^{\mathrm{miss}}$.

Table 14: Estimated and true background with the redecay method, in case of no SUSY signal, or presence of SU3 or SU4 SUSY signals. For the latter, also the amount of observed signal plus background events is shown, proving that although the background is overestimated, the excess is still present.

Sample	Estimated BG	True BG	Observed signal + BG
No SUSY	30.7 ± 9.8	39	39
SU3	50.8 ± 13.6	39	392
SU4	456 ± 102	39	1230

2.3.6 Combined fit method

The fit-based method for measuring the background, as described in this section, aims to improve upon the plain M_T sideband subtraction method for the one-lepton SUSY search mode. By analysing data in a L-shaped region at both low-$E_{\mathrm{T}}^{\mathrm{miss}}$ in the full M_T range and at low M_T in the full $E_{\mathrm{T}}^{\mathrm{miss}}$ range, and performing a two-dimensional extrapolation into the SUSY signal region we hope to enhance the background estimation. In a fit, correlations between $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_T can be taken into account. Furthermore, an explicit assumption can be put in that there is a finite SUSY contamination in the control sample.

For our purposes, we define in this analysis a sideband (SB) region and a SUSY signal (SIG) region. For both regions, we apply the standard SUSY one-lepton selection cuts defined in section 2.1, with the exception of the cut on M_T. The SB region is defined by the following additional cut: $E_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV *or* $M_T < 150$ GeV. In this analysis, the SIG region is defined by: $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV *and* $M_T > 150$ GeV. This classification assumes that the mass scale of SUSY is much higher than 200 GeV, which is true for all SUSY points considered in this note except SU4.

The analysis described here has been applied to electron + jets events only, but we expect the muon + jets analysis to be completely analogous.

The main Standard Model backgrounds we expect are single-leptonic $t\bar{t}$ decays, double-leptonic $t\bar{t}$ decays, and $W+$jets events. Other backgrounds, such as $Z+$jets, diboson production or single top are negligible for 1 fb^{-1}. The trigger used was an OR of the e22i single electron trigger, and the 4j50 multi-jet trigger [7].

Shape of the backgrounds We will try to fit the contributions of the backgrounds in three observables: M_T, $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{top}. Here m_{top} is the invariant mass of the three jets in the event with the largest vector-summed p_T [11].

We construct probability density functions (p.d.f.'s) that model the major contributing processes in the three observables after the event selection. We have done this both with and without explicitly taking into account correlations between the three observables. Most of these correlations are in any case consistent with zero.

Figure 14 shows the distribution of the three observables for each type of background we consider, as well as for one SUSY signal sample (SU3). The empirical model that we use to describe each background type is overlaid on the data. This comparison of shapes demonstrates that there is sufficient information in these three observables to be able to measure each of them in a combined fit.

We perform the procedure of constructing an empirical model from Monte Carlo for multiple SUSY data points. A striking feature of a comparison of $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_T distributions of these SUSY points *in the SB region* is that, with the exception of lower mass SUSY point SU4, they are all quite similar in shape. We can thus construct a model-independent 'Ansatz' shape to describe the SUSY contamination at low energy.

To validate this procedure, we perform fits to a "data" sample consisting of either background only, or background plus SU3 SUSY signal. The yields we find in a fit to 1 fb^{-1} are listed in Table 15 and are in agreement within errors with the truth values of the fitted event mix. If, in contrast, a SU3 SUSY signal would be present in data, but the fit would not allow for SUSY (see table), the fit would overestimate dileptonic $t\bar{t}$ and $W+$jets, and underestimate semileptonic $t\bar{t}$. This is as expected, as the SUSY signal has a long tail in M_T, but no substantial peak in the top mass.

The same table also lists the yields obtained from a fit in which all components are taken as a simple uncorrelated product of three one-dimensional p.d.f.'s. The yields and their uncertainties are very similar to those from the fit with models that include correlations and indicates that the effect of correlations in the description of the background components is minor.

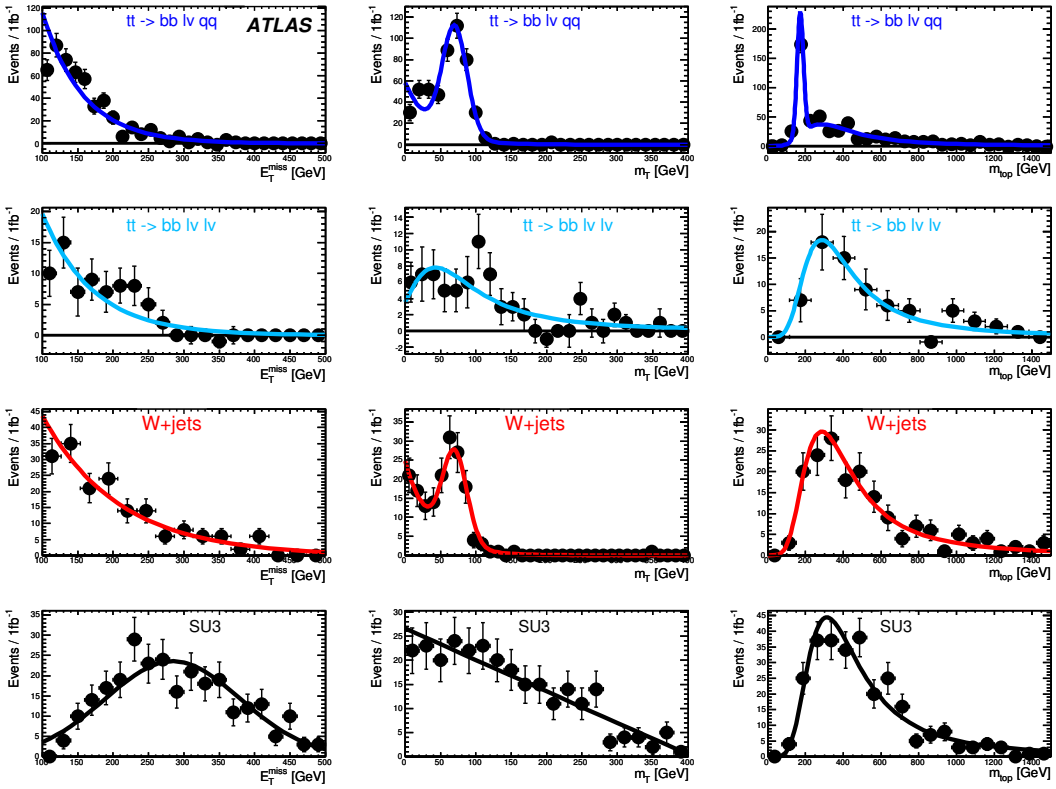


Figure 14: Distributions in missing E_T (left), M_T (middle) and m_{top} (right) of single lepton $t\bar{t}$ (top row), double lepton $t\bar{t}$ (second row), W + jet events (third row) and SUSY SU3 (last row). Each distribution is overlaid with a projection of the three-dimensional model that is fitted to that sample.

Fitting the data The fit with fixed shapes relies on simulated events to determine the shapes of the various background components, while all yields are fitted from the data. The next step is to release as many of the shape parameters in the fit to the data; in total there are 15 of these parameters. In the limit that all parameters can be floated, with the exception of the SUSY ansatz shape, the method becomes almost independent on simulation input and fully data driven. It turns out that on a 1 fb^{-1} sample we can float all but two of the $W+$jets and 1-lepton $t\bar{t}$ shape parameters. These two are the fraction of events that give the correct top quark mass in m_{top} for single lepton $t\bar{t}$ and the fraction of events with a correctly constructed W boson in M_T in single lepton $t\bar{t}$.

Floating the 2-lepton $t\bar{t}$ shape parameters in addition causes the fit to become unstable because the shapes of the dilepton component and that of the SUSY ansatz model are very similar. We are currently investigating the possibility of introducing additional constraints on the shape of the di-lepton $t\bar{t}$ events in order to solve this.

The result of the 1 fb^{-1} fit with floating parameters is shown in Figure 15.

The final step in the analysis is to extrapolate the yields of the standard model background components from the sideband region to the signal region. Table 16 shows the yields from the combined fit with floating shapes extrapolated to the signal region while propagating all (correlated) parameter uncertainties; for comparison the same table also shows the results with shapes kept fixed. The fits describe the data within the statistical uncertainties.

Table 15: Yields from the combined fit with fixed shapes for the various modeled physics processes. The leftmost column lists the yield fitted with models that include correlations in the model structure. The middle column shows these yields for models in which correlation terms have been disabled. The rightmost column gives the true yields in the fitted event mix.

Component	Fitted Yield	Fitted Yield w/o correlations	True Yield
no SUSY in data, no SUSY component in fit			
$W+$jets	186 ± 33	200 ± 27	173
1-lepton $t\bar{t}$	507 ± 32	482 ± 32	502
2-lepton $t\bar{t}$	53 ± 13	55 ± 16	70
SUSY	-	-	0
no SUSY in data, SUSY component in fit			
$W+$jets	185 ± 33	201 ± 27	173
1-lepton $t\bar{t}$	507 ± 32	482 ± 32	502
2-lepton $t\bar{t}$	52 ± 15	56 ± 16	70
SUSY	4.4 ± 3.0	-7.1 ± 16.0	0
SU3 in data, no SUSY component in fit			
$W+$jets	292 ± 35	261 ± 34	173
1-lepton $t\bar{t}$	386 ± 30	356 ± 31	502
2-lepton $t\bar{t}$	338 ± 25	389 ± 34	70
SUSY	-	-	271
SU3 in data, SUSY component in fit			
$W+$jets	181 ± 40	194 ± 35	173
1-lepton $t\bar{t}$	521 ± 36	509 ± 35	502
2-lepton $t\bar{t}$	35 ± 23	15 ± 30	70
SUSY	280 ± 22	293 ± 24	271

Table 16: Yields from the combined fit with either fixed or floating shapes in the sideband region extrapolated to the full parameter space, the truth yields in full parameter space, the extrapolated yields into the signal region and the truth yields in the signal region.

Component	Extrap. Yield in FULL		True	Extrap. Yield in SIG		True
	Shape Fixed	Shape Floating	FULL	Shape Fixed	Shape Floating	SIG
$W+$jets	205 ± 45	227 ± 68	173	0.5 ± 0.4	-1.2 ± 2.7	2
1-lepton $t\bar{t}$	476 ± 35	485 ± 59	502	0.4 ± 0.2	-1.1 ± 3.9	0
2-lepton $t\bar{t}$	62 ± 38	17 ± 54	70	4.5 ± 2.9	4.7 ± 7.9	5
SUSY SU3	273 ± 33	287 ± 38	271	92.7 ± 2.8	95.6 ± 4.0	91

Effect of SUSY signal While table 16 shows the results of the fit for the various backgrounds and the SUSY signal in the case of the SU3 sample, Table 17 shows results, for fits with floating shapes, for other SUSY signal samples.

As has been noted earlier, SU4 is a special case. It is too close to the background for the Ansatz of the shape to be valid, and it can not be fitted very well.

Systematics Table 18 summarizes the results of a series of systematic studies. These studies are quantified in terms of relative variations of the measured SUSY cross-section, which we define as the counted number of events in the data in the SIG region minus the fitted and extrapolated number of SM model

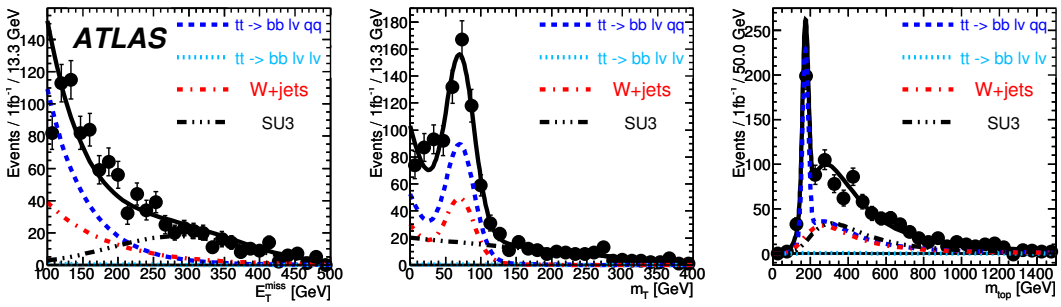


Figure 15: Distribution of missing E_T (left), M_T (center) and m_{top} (right) of a 1 fb^{-1} mix of $t\bar{t}$, $W+$ jets standard model events and SUSY SU3 events overlaid with projections of the combined model on these observables that was fitted to this mix of events with floating yield parameters and floating shaped parameters. For each projection the contributions of the 1-lepton $t\bar{t}$ ttbar contribution (dark blue), 2-lepton $t\bar{t}$ ttbar contribution (light blue), $W+$jet contribution (red) and ansatz SUSY constribution (black) are shown.

Table 17: Yields from the combined fit with floating shapes in the sideband region extrapolated to the full parameter space, the truth yields in full parameter space, the extrapolated yields into the signal region and the truth yields in the signal region, for various SUSY samples.

Component	Extrap. Yield in FULL	True in FULL	Extrapolated Yield in SIG	True in SIG
SU1				
$W+$jets	215 ± 56	173	1.8 ± 2.0	2
1-lepton $t\bar{t}$	486 ± 54	502	0.6 ± 0.6	0
2-lepton $t\bar{t}$	19 ± 35	70	0.5 ± 1.7	5
SUSY SU1	154 ± 30	129	50.1 ± 2.6	46
SU2				
$W+$jets	226 ± 51	173	1.3 ± 1.0	2
1-lepton $t\bar{t}$	452 ± 49	502	0.7 ± 0.5	0
2-lepton $t\bar{t}$	81 ± 32	70	8.0 ± 5.0	5
SUSY SU2	0 ± 10	14	1.0 ± 6.1	4
SU6				
$W+$jets	215 ± 54	173	0.5 ± 0.8	2
1-lepton $t\bar{t}$	469 ± 52	502	0.2 ± 0.6	0
2-lepton $t\bar{t}$	29 ± 34	70	2.6 ± 2.7	5
SUSY SU6	117 ± 29	86	38.8 ± 2.5	35
SU8				
$W+$jets	239 ± 53	173	2.5 ± 2.4	2
1-lepton $t\bar{t}$	485 ± 51	502	1.0 ± 1.2	0
2-lepton $t\bar{t}$	66 ± 45	70	15.1 ± 12.0	5
SUSY SU8	34 ± 26	79	34.5 ± 13.0	46

events expected in that same region, and divided by the efficiency for the selection of SUSY events, including the SIG region cuts, as measured from a pure sample of simulated SU3 SUSY events.

The second column of Table 18 lists the relative variation under the influence of each systematic variation if no extrapolation is applied in the fit and the analysis is performed without the SIG region cut. This column thus quantifies the effect on the shape fit stability only. The third columns shows the effect

of each systematic study on the fitted yield in the signal region using extrapolation. Most effects are here of the order of 5-7%. The uncertainties due to Monte Carlo statistics are explicitly listed in the table.

Table 18: List of studied systematic uncertainties for the combined-fit method.

Systematic variation	w/o extrapolation $\pm$ MC-error [%]	in SIG $\pm$ MC-error [%]
Jet energy scale	1.9 ± 0.7	3.1 ± 1.1
Jet energy resolution	3.7 ± 0.5	0.5 ± 0.1
Electron energy scale	4.8 ± 0.6	5.6 ± 1.4
Electron energy resolution	9.0 ± 1.2	7.2 ± 1.4
Electron identification efficiency	0.5 ± 0.2	3.6 ± 2.3
Soft $E_{\mathrm{T}}^{\mathrm{miss}}$ scale	8.1 ± 1.8	7.4 ± 3.4

3 No-lepton search mode

3.1 Selection

In the no-lepton search mode, we veto all events with an identified electron or muon with a p_T of more than 20 GeV. We demand at least 4 jets with $|\eta| < 2.5$ and $p_T > 50$ GeV, one of which must have $p_T > 100$ GeV. The transverse sphericity S_T should be larger than 0.2, and the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ should be larger than 100 GeV and larger than 0.2 M_{eff} , where M_{eff} is the effective mass. We add one more cut against the QCD background: the minimum value of the difference in azimuthal angle between the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector and the three highest-p_T jets should be larger than 0.2. This cut is futher discussed in a dedicated note on QCD background estimation [3].

3.2 Backgrounds in Monte Carlo

Figure 16 shows the distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{eff} after all the selections are applied.

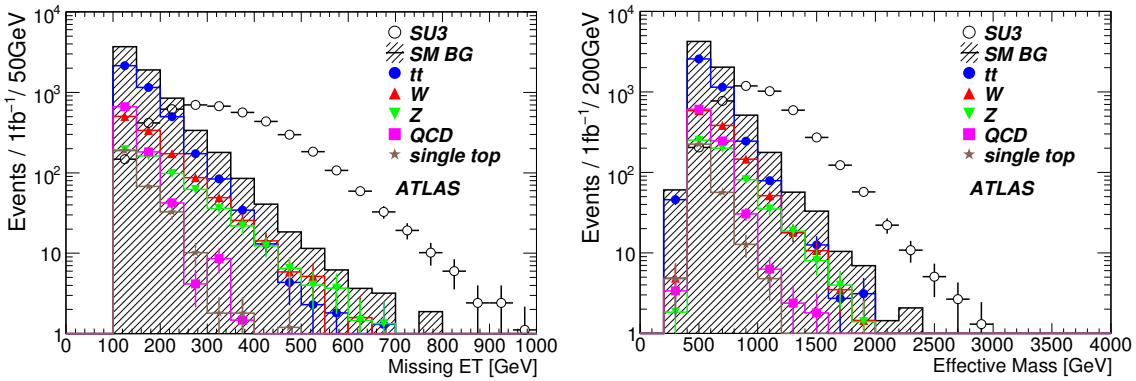


Figure 16: The $E_{\mathrm{T}}^{\mathrm{miss}}$ and effective mass distributions of the SUSY signal and background processes for the no-lepton mode with an integrated luminosity of 1 fb^{-1}. The open circles show the SUSY signal (SU3 point). The shaded histogram shows the sum of all Standard Model backgrounds; different symbols show the various components.

3.3 Data-driven estimation strategies

In this section we discuss data-driven estimation strategies for the no-lepton search mode. The strategies we have studied are:

1. estimation of Z ($\rightarrow \nu\bar{\nu}$) plus jets from Z ($\rightarrow \ell^+\ell^-$) plus jets, purely from data (replace method, section 3.3.1);

2. estimation of Z and W in an analogous way, but also helped by Monte Carlo (section 3.3.2);

3. estimation of W and $t\bar{t}$ background from a one-lepton control sample derived by reversing one of the selection cuts (on M_T) (section 3.3.3);

4. estimation of the cross-section of the $t\bar{t} \rightarrow b\bar{b}q\bar{q}'\tau\nu$ process with hadronic tau decay (section 3.3.4).

3.3.1 Replace method: $Z \rightarrow \nu\bar{\nu}$ from $Z \rightarrow \ell^+\ell^-$

Introduction The $Z \rightarrow \nu\bar{\nu}$ background is one of the main background process in the no-lepton channel. In order to estimate and reproduce the number of expected background events, as well as the shape of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{eff} distributions, $Z \rightarrow \ell^+\ell^-$ events are selected, and the charged leptons are replaced by neutrinos. However, as the ratio of branching-ratios $\mathrm{Br}(Z \rightarrow \ell^+\ell^-)/\mathrm{Br}(Z \rightarrow \nu\bar{\nu})$ is small, statistical uncertainties will tend to be relatively large. Two solutions are proposed :

1. Taking the distribution shape from $Z \rightarrow \ell^+\ell^-$ data but constraining it via a fit plus the assumption of a smooth evolution of the fitting parameters when relaxing the cuts. This is the method described in this section.

2. Taking the distribution shape from Monte Carlo simulation as described in the next section (Section 3.3.2).

The Monte Carlo method is more sensitive to generator-level and detector systematic uncertainties, but does not suffer from the larger statistical uncertainties, whereas the replace method precision is limited by the number of events in the control sample, but less sensitive to systematic uncertainties from the detector. Both methods have to account for the fact that the detected charged lepton pairs will not cover the full phase space of the neutrinos.

Control Sample Selection The control sample selection is identical to the no-lepton SUSY search selection, except that two electrons or two muons are required, and that the missing E_T ($E_{\mathrm{T}}^{\mathrm{miss}}$) is replaced by $p_T(\ell^+\ell^-) \simeq p_T(Z)$. Thus it is assumed that neutrinos are the main contribution to $E_{\mathrm{T}}^{\mathrm{miss}}$ when the Z boson decays into two neutrinos, such that $E_{\mathrm{T}}^{\mathrm{miss}}$ is roughly equivalent to $p_T(Z)$ for this physics process. The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution of ATLAS is sufficient for this to be a good approximation. In addition to pairs of isolated charged leptons, a sample composed of $Z \rightarrow e^{\pm}X$ is added, where X is a non-isolated electron or an electron-like object with very loose cuts. This additional sample is used to increase the statistics and measure the electron identification efficiency via the "tag-and-probe" method. The goal of the tag-and-probe method is to select on one side a good electron (tag) and look at the other side to the nature of the object (the probe) which matches the constraint on the Z mass. Two cuts are added to reject the remaining backgrounds : $81 < M_Z(\ell^+\ell^-) < 101$ GeV, and Missing $E_T < 30$ GeV. After all cuts the number of selected events is summarized in Table 19. In particular, Table 19 shows the effect of an upper cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ in order to reject $t\bar{t}$ background in the $Z \rightarrow e^{\pm}X$ channel. It has been verified that the efficiencies measured with the tag-and-probe method agree with efficiencies obtained directly from Monte Carlo, confirming that X is indeed dominated by electrons rather than hadronic jets.

Table 19: Number of selected $Z \to \ell^+\ell^- + \geq 4$ jets events in 1 fb^{-1} in the replace-method analysis.

Process	No $E_{\mathrm{T}}^{\mathrm{miss}}$ cut	$E_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV
$Z \to e^+e^- +$ n jets	18.2	14.1
$Z \to e^{\pm}X +$ n jets	25.6	19.6
$Z \to \mu^+\mu^- +$ n jets	33.2	26.1
$Z \to \tau^+\tau^- +$ n jets	1.6	0.
$t\bar{t} \to bb\ell\nu\ell\nu +$ n jets	56.2	0.3
$t\bar{t} \to bb\ell\nu qq +$ n jets	506.6	2.5

Lepton identification and acceptance corrections A number of correlations must be applied in order to derive $Z \to \nu\bar{\nu}$ distributions from $Z \to \ell^+\ell^-$: (1) a fiducial correction, since we cannot detect e and μ leptons beyond $|\eta| = 2.5$; (2) a kinematics correction for the additional cuts used to select $Z \to \ell^+\ell^-$, including the Z invariant-mass window, the p_T cut on the leptons, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut; and (3) a correction for the lepton identification efficiency. The first two effects have to be computed from simulation whereas the lepton identification efficiency can be measured from collision data using the tag-and-probe method.

After all corrections, the distribution can be summarized by the following formula:

$$N_{Z\to\nu\bar{\nu}}(E_{\mathrm{T}}^{\mathrm{miss}}) = N_{Z\to\ell^+\ell^-}(p_T(\ell^+\ell^-)) \times c_{\mathrm{Kin}}(p_T(Z)) \times c_{\mathrm{Fidu}}(p_T(Z)) \times \frac{\mathrm{Br}(Z\to\nu\bar{\nu})}{\mathrm{Br}(Z\to\ell^+\ell^-)}, \qquad (5)$$

where $N_{Z\to\nu\bar{\nu}}(E_{\mathrm{T}}^{\mathrm{miss}})$ is the corrected number of events per bin of missing E_T, $N_{Z\to\ell^+\ell^-}(p_T(\ell^+\ell^-))$ is the raw number of control sample events as a function of $p_T(Z)$, c_{Kin} and c_{Fidu} are the kinematic and fiducial corrections. The $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{eff} distributions of $Z \to e^+e^- + e^{\pm}X$ and $Z \to \mu^+\mu^-$ events after all corrections are compared to $Z \to \nu\bar{\nu}$ distributions in Figure 17.

For very high values of $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{eff} , statistics is low. In order to present a smooth prediction of the background, for example as a function of M_{eff} , a fit of the shape has been performed. Of course, the fit is also affected by the low statistics in the tail, but by relaxing the jet p_T and $E_{\mathrm{T}}^{\mathrm{miss}}$ cuts and observing how the fit parameters evolve with the cuts, a smooth prediction can be made.

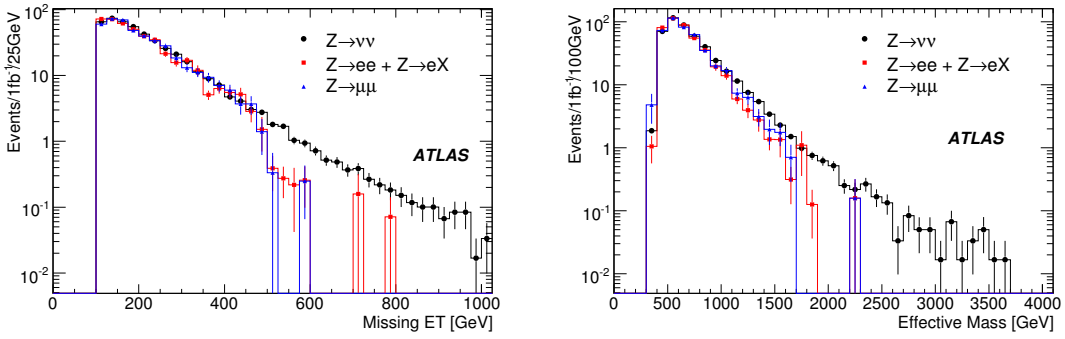


Figure 17: (Left) $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution after all corrections for $Z \to \nu\bar{\nu}$, $Z \to e^+e^- + e^{\pm}X$ and $Z \to \mu^+\mu^-$ processes. The number of events corresponds to an integrated luminosity of 1 fb^{-1}. (Right) M_{eff} distribution for the same physics processes.

Systematic uncertainties The effects of various Monte Carlo generator and detector systematic uncertainties are summarized in Table 20. The main generator systematic uncertainty is due to the variation of

the renormalization scale and affects the acceptance correction, whereas the principal detector systematic uncertainty is related to the soft part of the missing transverse energy which is not taken into account when replacing neutrinos by charged leptons.

Table 20: Summary of systematic uncertainties for the replace method.

Description	Relative uncertainty $\Delta N/N$ [%]
MC generator systematics	
ALPGEN parameter variation	6.3
Detector systematics	
Electron energy scale	0.05
Electron energy resolution	0.03
Electron id efficiency	0.50
Muon energy scale	0.30
Muon energy resolution	0.39
Muon id efficiency	1.00
$E_{\mathrm{T}}^{\mathrm{miss}}$ scale (soft part)	4.5
Total systematics (quadratic sum)	~ 8
Total statistics	~ 13
Total uncertainties	~ 15

SUSY contamination in the control sample Due to the tight control sample selection cuts, in particular $E_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, the SUSY contamination in the control sample in 1 fb^{-1} is negligible: 0.1 events for SU1, 0.4 events for SU2, 0.9 events for SU3, and < 0.07 events for SU6.

3.3.2 Z and W background estimates from $Z \to \ell^+\ell^-$ plus Monte Carlo shape

A modification of the method of the previous section is described here. Denoted the "MC method", it uses only the number of events in the $Z(\to \ell^+\ell^-)+$jets control sample for normalization, but otherwise relies on Monte Carlo simulation of kinematical distributions of events. The same normalization factor is used for both the $Z(\to \nu\bar{\nu})$ background and the $W(\to \ell\nu)$ background since the production mechanism is very simliar.

Control sample The event selection of the control sample demands: two opposite-sign same-flavour leptons with $p_T > 20$ GeV; $E_{\mathrm{T}}^{\mathrm{miss}} < 40$ GeV; a di-lepton mass $M_{\ell\ell}$ within $\pm$ 10 GeV of m_Z . Subsequently, the standard no-lepton SUSY cuts are applied after replacing $\ell^+\ell^-$ with $\nu\bar{\nu}$.

The number of events selected with 1 fb^{-1} is summarized in Table 21. The contamination from $t\bar{t}$ events is about 10^{-2} and therefore not significant. The number of events estimated by a full simulation sample (with event filter $p_T^Z > 80$ GeV) is 72 ± 3 for 1 fb^{-1}, which leads to a statistical uncertainty on the estimation of 12%.

The method has been tested with a pseudo data sample prepared with Monte Carlo parameters set differently from the standard Monte Carlo sample[5)]. In this pseudo data sample, the shape of $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{eff} distributions is not affected; only the normalization changes significantly. The MC method is able to recover such a change and predict the background correctly.

[5)] renormalisation and factorisation scale 0.8 times the default, a minimum parton p_T of 30 GeV (rather than 40 GeV), and separation between partons $\Delta R_{jj} = 0.6$ (rather than 0.7).

Table 21: Number of events in the MC method control sample (for 1 fb^{-1}).

process	μ mode	e mode	μ + e (sum)
$Z(\rightarrow \ell^+\ell^-)$	51 ± 1	34 ± 1	85 ± 1
$Z(\rightarrow \tau^+\tau^-)$	< 0.1	< 0.1	< 0.1
$t\bar{t}$	1.0 ± 0.3	0.5 ± 0.2	1.5 ± 0.4

Table 22: Systematic uncertainty from variation of Monte Carlo generator parameters (ALPGEN) for the MC method. The pseudo data sample is discussed in the text.

$Z(\rightarrow \nu\bar{\nu})$	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV	$M_{\mathrm{eff}} > 800$ GeV
pseudo data sample	5%	3%
half renormalization scale	2%	0%
lower parton p_T	9%	4%
$W(\rightarrow \ell\nu)$	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV	$M_{\mathrm{eff}} > 800$ GeV
pseudo data sample	16%	7%
half renormalization scale	4%	2%
lower parton p_T	12%	15%

Table 23: Systematic uncertainty from detector performance in the MC method.

	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV	$M_{\mathrm{eff}} > 800$ GeV
Jet energy scale	6%	6%
Jet energy resolution	1%	1%
$E_{\mathrm{T}}^{\mathrm{miss}}$ soft component scale	1%	$< 1\%$
Lepton energy scale	$< 1\%$	$< 1\%$
Lepton identification efficiency	2%	2%

Systematic uncertainty The MC method relies on Monte Carlo for the shapes of the background distributions. The systematic uncertainty from the Monte Carlo is estimated by using samples with different Monte Carlo parameters (pseudo real data, half renormalization scale sample, lower parton threshold sample). The results are summarized in Table 22; the deviation of the samples is 16% at most.

Other potential sources of systematic error in the background estimation from uncertainties in detector performance are summarized in Table 23. Since Monte Carlo is used, an uncertainty in the experimental jet energy scale affects the MC method prediction significantly more than for the replace method (Section 3.3.1).

SUSY contamination of the control sample The tight selection cuts of the control sample makes any SUSY contamination negligible (as was also the case for the replace method discussed in Section 3.3.1).

3.3.3 Control sample constructed from an M_T-selected one-lepton sample

***W* and top backgrounds** The $t\bar{t}$ and $W^{\pm}+$jets processes contribute to the background in the no-lepton mode, when the lepton emitted in the $W^{\pm} \rightarrow \ell\nu$ process is not identified. The reasons why the emitted lepton is missed are summarized in Table 24. Hadronic decay of τ's and the lepton being out

of acceptance (the p_T of the lepton is required to be larger than 20 GeV) are the dominant reasons, and similar reasons/ratio's are observed in both $t\bar{t}$ and $W^{\pm}+$jets processes.

Table 24: Numbers of events in which leptons are missed for various reasons in the semi-leptonic and leptonic decays of top and $W^{\pm}$. The numbers are normalized to 1 fb^{-1} and listed after the SUSY no-lepton selection is applied.

	$t\bar{t}$	$W^{\pm}+$jets
$W^{\pm} \to \tau \to$ hadron	1993 (43%)	773 (42%)
Out of acceptance	1805 (39%)	762 (41%)
Isolation (close to jet)	807 (18%)	322 (17%)

The production processes $t\bar{t}$ and $W^{\pm}$ with $W^{\pm} \to \ell\nu$ where the lepton is identified constitute good control samples from which to estimate these background processes in the no-lepton mode, since similar kinematic distributions are expected except for the presence of the lepton. For the control sample, the same kinematic selections as for the signal in the no-lepton mode are applied. In addition, exactly one isolated lepton (e or μ with p_T larger than 20 GeV) is required and the transverse mass between this lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is required to be smaller than 100 GeV to enhance the $t\bar{t}$ and $W^{\pm}$ processes. After these selections, this identified lepton is treated as if had been missed and all kinematic variables are recalculated.

The distributions of $t\bar{t}$ and $W^{\pm}+$jets events differ after the SUSY selections are applied. The $E_{\mathrm{T}}^{\mathrm{miss}}$ for $W^{\pm}$ tends to be larger than for $t\bar{t}$ events, since the boost factor of the $W^{\pm}$ is larger for $W^{\pm}+$jets after several high p_T jets are required. Therefore the reproduced distributions are sensitive to the mixture of the $t\bar{t}$ and $W^{\pm}+$jets events in the control sample. The fractions of the $t\bar{t}$ and $W^{\pm}+$jets processes in the control sample are 81% and 19%, respectively, close to the actual backgrounds in the signal region, which are 73% and 27% respectively. The systematic errors due to the uncertainties in the cross-section of $t\bar{t}$ and $W^{\pm}+$jets will be discussed later.

The estimated distribution is normalized with the data at 100 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 200 GeV, where the contribution of the SUSY signal is expected to be small. The estimated distributions are slightly harder than the true distributions of the background processes, but similar distributions are obtained. The number of estimated background events is summarized in the two top rows in Table 25. Reasonable agreement is observed at high $E_{\mathrm{T}}^{\mathrm{miss}}$.

QCD, W and top background without and with SUSY signal In a dedicated note within this volume [3] various methods for the estimation of QCD background from data are discussed.

In this section, we include QCD background, and estimate it as follows. It has been shown that after the removal of events with mismeasured $E_{\mathrm{T}}^{\mathrm{miss}}$ (e.g. noisy or dead calorimeter cells), as discussed elsewhere in this volume [12], semi-leptonic heavy quark (b,c) decays are the dominant contribution to large $E_{\mathrm{T}}^{\mathrm{miss}}$ in the QCD background. A function is derived to represent the momentum fraction taken by the neutrino in b and c quark decays. This function is then applied to a control sample taken from data (at least four jets with p_T larger than 50 GeV, p_T of the leading jet larger than 100 GeV, $E_{\mathrm{T}}^{\mathrm{miss}}$ smaller than 100 GeV) dominated by light-quark QCD events. The resulting distributions are normalized to data in the region of $\Delta\phi_{\mathrm{min}} < 0.2$, where $\Delta\phi_{\mathrm{min}}$ is the minimum value of the difference in azimuthal angle between the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector and any of the three highest p_T jets, as discussed in section 3.1 and the dedicated note [3].

The $W^{\pm}$ and $t\bar{t}$ background processes can also be estimated from the data, as discussed in the previous paragraph. Since all these processes are present in the data simultaneously, the background estimations

should be done together. The same holds true for the presence of SUSY signal, which would contaminate the control samples.

Figure 18 (top row) shows the estimated and true distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ and the effective mass for the combined background processes, without SUSY signal. The background distributions are reproduced well and the correct normalizations are obtained for all the background processes. The numbers of events are also summarized in Table 25.

Table 25: Number of true and estimated (MT method) background events in the no-lepton mode, for $t\bar{t}$, $W^{\pm}$ and QCD processes without SUSY signal, normalized to 1 fb^{-1}.

	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV
$t\bar{t}$ and $W^{\pm}$ only		
True BG (top and $W^{\pm}$)	6894 ± 83	276 ± 17
estimated BG (top and $W^{\pm}$)	7018 ± 269	311 ± 28
QCD, $t\bar{t}$ and $W^{\pm}$		
True BG (QCD, top and $W^{\pm}$)	8077 ± 90	300 ± 17
Estimated BG	8158 ± 273	327 ± 28
Ratio(Est./True)	1.01 ± 0.04	1.09 ± 0.11

If SUSY exists, SUSY signal can contaminate the background estimations. This is illustrated in Figure 18 (middle row) and Table 26. In the figure, the SUSY SU3 signal point is used; the table shows the variation over a number of SUSY samples. The figure and the table show that SUSY contamination causes a decrease of SUSY event excess by typically 30%, but that the considered points, with the exception of SU2, are still observable with 1 fb^{-1}.

Table 26: Number of background events and estimated (MT method) numbers for $t\bar{t}$, $W^{\pm}$ and QCD processes, as well as various SUSY signals, in the no-lepton mode, normalized to 1 fb^{-1}.

	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV
True BG (QCD,top and $W^{\pm}$)	8077 ± 90	300 ± 17	8077 ± 90	300 ± 17
	SU1		SU4	
Estimated BG	8493 ± 283	510 ± 39	27527 ± 588	1409 ± 83
True BG + SUSY signal	9152 ± 96	1078 ± 33	34209 ± 185	3535 ± 59
	SU2		SU6	
Estimated BG	8198 ± 274	329 ± 30	8362 ± 279	431 ± 35
True BG + SUSY signal	8193 ± 91	351 ± 19	8930 ± 95	924 ± 30
	SU3			
Estimated BG	9188 ± 299	633 ± 44		
True BG + SUSY signal	11333 ± 106	2113 ± 46		

New MT method The "new MT method", discussed in section 2.3.1, provides a first rough method to correct for the presence of SUSY signal in the control sample, once a SUSY excess has been observed in data. With this method, the distributions shown in the bottom row of Figure 18 are obtained. The figure shows the estimated background distributions with the new MT method, compared to the true

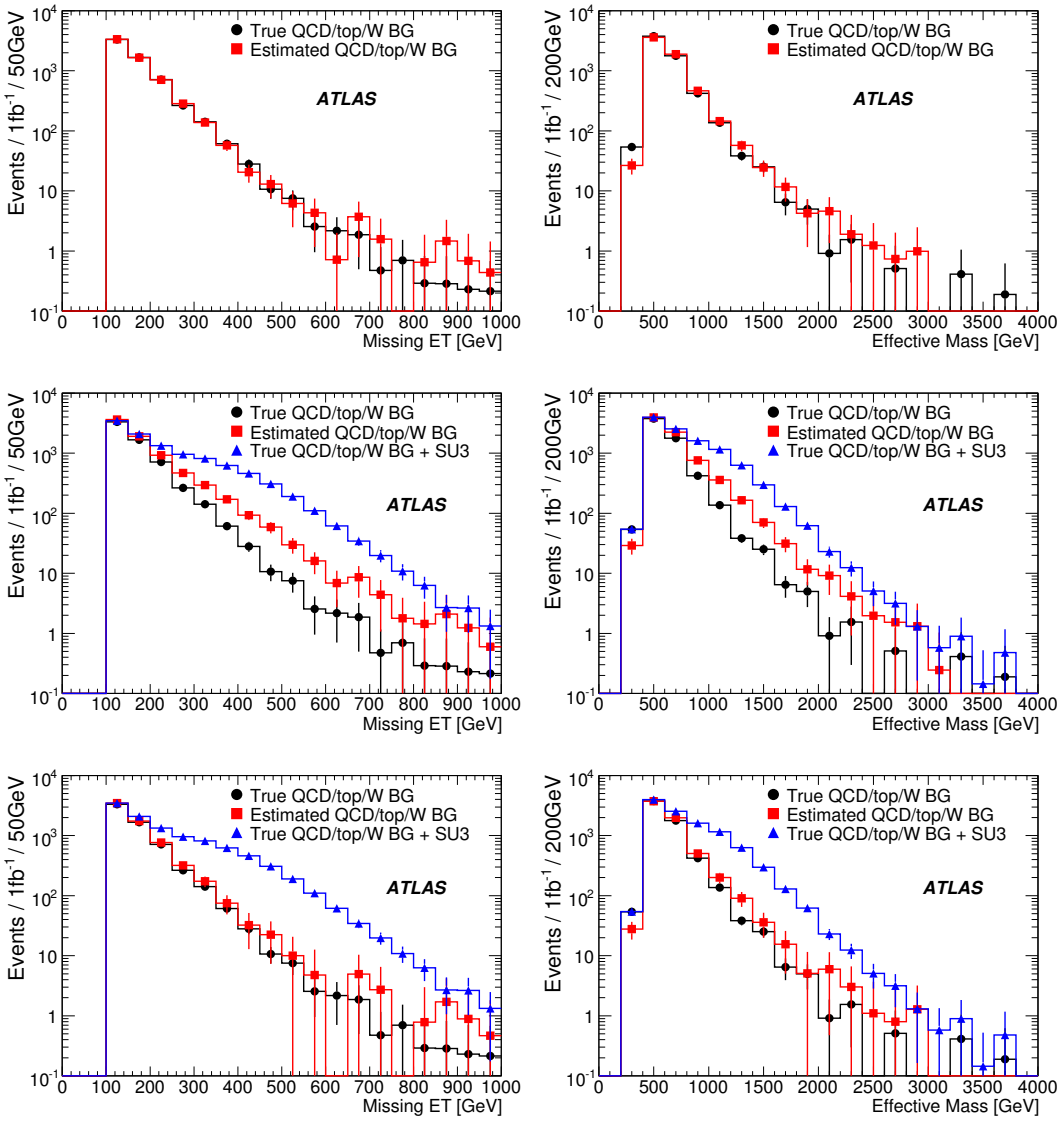


Figure 18: The estimated distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ and the effective mass of the $t\bar{t}$, $W^{\pm}$ and QCD backgrounds in the no-lepton mode with a luminosity of 1 fb^{-1}. In the plots in the top row, no SUSY signal is present in the data, and black/red histograms show the true/estimated (MT method) background distributions. In the plots in the middle and bottom rows, a SUSY SU3 signal is present in the data, and the blue histograms show the background plus the SUSY signal. In the two plots in the middle row, no correction was applied. In the two bottom plots, a correction with the "new MT method" was performed.

background distributions, when SU3 SUSY signal is present in data.

Systematic uncertainties The systematic errors of this method are summarized in Table 27. The ALPGEN parameter variation includes a variation of the relative fraction of $t\bar{t}$ and $W+$jets in the sample. Although the actual number of background events varies with jet energy scale and Monte Carlo generator parameters by some 25%, the data-driven MT method is able to predict the background in the signal region to within an assigned systematic error of 15%.

Table 27: Systematic uncertainties of the MT-method background estimations in the no-lepton mode without SUSY signal. Also changes of the absolute background numbers are listed. Numbers are normalized to 1 fb^{-1}.

	Syst. error	Change in background level
Jet energy scale	$< 5\%$	25%
Lepton energy scale	$< 5\%$	0%
Lepton Efficiency	$< 5\%$	$< 1\%$
MC@NLO vs ALPGEN	$< 5\%$	16%
ALPGEN parameter variation	$< 5\%$	8%

3.3.4 Top pairs with τ decay

Introduction The precise determination of the cross-section for the process of top-pair production with one tau that decays hadronically, $t\bar{t} \to W(q\bar{q}')W(\tau_{\rm had}\nu_\tau)b\bar{b}$, is relevant because it constitutes an important background to SUSY searches with no leptons and significant $E_{\rm T}^{\rm miss}$. In fact, if no tau veto is applied, about 65% of the total $t\bar{t}$ background in the no-lepton mode corresponds to events containing one tau.

Event reconstruction and selection The topology of $t\bar{t} \to W(qq')W(\tau_{\rm had}\nu)b\bar{b}$ events consists of two light–quark jets, one tau, and two b jets.

The control of the tau fake rate is very important in a busy environment like $t\bar{t}$ where the purity of the reconstructed tau sample is low due to the large jet multiplicity. A high tau purity is needed in order to reduce the internal combinatorial background and the background from the semileptonic ($t\bar{t} \to W(\ell\nu_\ell)W(q\bar{q}')b\bar{b}$ where $\ell \in \{e,\mu\}$) decays of $t\bar{t}$. In this analysis, we use a calorimeter-based tau reconstruction algorithm [1], and require a minimum visible p_T of the identified tau of 25 GeV.

The event is built independently on the hadronic side and the leptonic side, and topology variables useful to identify $t\bar{t}$ events and reject QCD and $W+$jet background are extracted. On the hadronic side, a hadronic W invariant mass is built choosing, among all the combinatorial possibilities of di-jets, the pair of jets with closest invariant mass to its PDG value. The hadronic top is built combining this hadronic W with the closest identified b jet in ΔR. The b-jet identification is loose, with a 75% efficiency for b jets from top decay.

The $E_{\rm T}^{\rm miss}$ is combined with the identified tau in order to build a leptonic W transverse mass. We assume a collinear approximation for the decay of the tau (the visible products of the hadronic decay of the tau and the associated $\bar{\nu}_\tau$ are collinear), and determine the invariant transverse mass of the leptonic W. The resulting leptonic W is then combined with the closest (in ΔR) b jet to constitute the leptonic top.

For each event, a reconstructed tau will build a leptonic W (and top) in combination with $E_{\rm T}^{\rm miss}$, and will have an associated hadronic W (and top). If there is more than one reconstructed tau we select the one that is associated with the jet pair that gives the hadronic W invariant mass closest to its PDG value.

Once the event is built, topology variables suitable for $t\bar{t}$ selection are computed and selection cuts are applied: $E_{\rm T}^{\rm miss} > 35$ GeV, no identified electron or muon with $p_T > 15$ GeV should be present in the event, the angle $\Delta\phi$ between the two reconstructed top quarks should be larger than 2.5, the ratio between the transverse momentum of the two reconstructed top quarks should be smaller than 2, the angular distance (ΔR) between the reconstructed b jets should be larger than 1, and the angle $\Delta\phi$ between the missing momentum vector and the hadronic b jet should be larger than 0.5.

With the loose b tagging used here, 2910 $t\bar{t}(q\bar{q}', \tau_{\rm had})$ events are selected for 1 fb^{-1}. The background consists of QCD and $W+$jets, for which 110 and 100 events respectively are selected. Therefore, with loose b tagging a good signal-to-background ratio can already be reached, although the uncertainty on

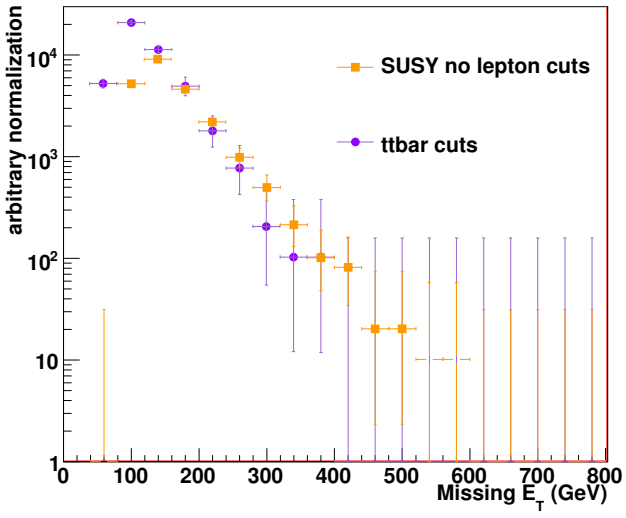


Figure 19: Missing E_T of $t\bar{t}$ events with $(q\bar{q}', \tau_{\rm had})$ for (circles) $t\bar{t}$ selection as in this analysis (control sample) and (squares) for the SUSY no-lepton mode selection.

the QCD background number is large. If one were to use tighter b tagging (60% efficiency), 1650 $t\bar{t}(q\bar{q}', \tau_{\rm had})$ events would be selected, against 2 QCD events and no $W+$jet events.

In the presence of SUSY, the numbers of SUSY signal events that would pass the event selection with the loose b tag for 1 fb^{-1} are given in Table 28. They are generally small, with the exception of the SU4 point.

Table 28: The number of SUSY events remaining after $t\bar{t}(q\bar{q}', \tau_{\rm had})$ selection for different SUSY points, for 1 fb^{-1}.

	SU1	SU2	SU3	SU4	SU6	SU8
N_{evt}	22 ± 1	5 ± 1	155 ± 5	1700 ± 60	70 ± 4	45 ± 3

Estimation of the $t\bar{t}$ with $(q\bar{q}', \tau_{\rm had})$ background to SUSY. Figure 19 shows that the selection applied in order to identify the $t\bar{t}$ events with $(q\bar{q}', \tau_{\rm had})$ introduces little bias in the $E_{\rm T}^{\rm miss}$ distribution, as compared to the SUSY no-lepton mode selection.

Applying the SUSY no-lepton mode cuts, we estimate 210 $t\bar{t}$ $(q\bar{q}', \tau_{\rm had})$ events as remaining background to the SUSY no-lepton mode for 1 fb^{-1}.

Systematic uncertainties In the data-driven analysis of $t\bar{t}$ decays in the $(q\bar{q}', \tau_{\rm had})$ final state, the most relevant contributions to the detector uncertainties are the jet and $E_{\rm T}^{\rm miss}$ energy scale, b-tagging efficiency and τ identification efficiency.

The b-tagging efficiency plays an important role in the $t\bar{t}(q\bar{q}', \tau_{\rm had})$ reconstruction, since one of the selection criteria is that the two b jets expected in the final state should be reconstructed and correctly identified[6)].

[6)] This is in fact the only analysis in this note where b tagging is used.

Table 29: Systematic variation of the $t\bar{t}(q\bar{q}', \tau_{\rm had})$ cross-section due to detector-related uncertainties.

Systematic variation	Cross-section variation [%]
Jet energy scale	2.5
b-tagging efficiency	7.5
Light quark rejection in b-tag	1.3
τ-identification efficiency	3.4
Light quark rejection in τ-identification	4.5

The systematic contribution to the measurement of the $t\bar{t}(q\bar{q}', \tau_{\rm had})$ cross-section due to the τ identification efficiency has been estimated by varying the τ identification efficiency and the light quark rejection factor by 10%.

The uncertainty on the QCD background in the $t\bar{t}(q\bar{q}, \tau_{\rm had})$ sample is large due to the limited number of Monte Carlo events which could be generated. Probably tight b tagging is required in this analysis. Further study is needed on this topic.

4 Multi-lepton and tau search modes

As well as the one-lepton and no-lepton search modes described earlier, there is considerable SUSY discovery potential in the multi-lepton and tau search modes, as discussed elsewhere in this volume [2,4].

The data-driven estimation of backgrounds in these modes, particularly the opposite-sign dilepton and the tau modes, can use the methods that have been described earlier in the context of the one-lepton search mode have also proven to be useful. These include the MT method, the HT2 method, the kinematic reconstruction method and the redecay method. Furthermore, for the same-sign dilepton mode, a technique based on lepton isolation, as described in the note on QCD backgrounds [3] could be further developed.

5 Discussion

The methods presented in this note represent a number of ideas on how top, W and Z backgrounds to SUSY searches can be extracted from the data, with appropriately chosen control samples. The results indicate that we expect, with 1 fb^{-1}, to be able to measure in the no-lepton mode:

- the $Z \rightarrow \nu\bar{\nu}$ background with two different methods to 8–13% stat. error, 10–15% syst. error;
- the $t\bar{t}$ background with hadronic tau decay to $< 6\%$ stat. error, 10–15% syst. error (but with a caveat for the QCD background);
- and the sum of top, W and QCD backgrounds with the MT method to 4–8% stat. error, and 15% syst. error.

In the one-lepton mode:

- the MT method gives the sum of $t\bar{t}$ and W background to 4–8% stat. error, 15% syst. error;
- the semileptonic $t\bar{t}$ background can be estimated to 5% stat. error, 22% syst. error;
- we can determine the fully leptonic $t\bar{t}$ background to 10% stat. error, 20% syst. error in at least three independent ways;
- and we have a combined fit method to extract all components.

These methods can also be applied to the multi-lepton and tau search modes.

The results obtained with the MT method in this note are stable with respect to systematic variations in detector performance and Monte Carlo parameters and cross-sections to the 15% level. However, the MT method measures a sum of semileptonic and fully leptonic $t\bar{t}$ and $W/Z+$jets background; it relies on a control sample with different composition than the signal sample, and there is a subtle interplay between the $t\bar{t}$ and $W/Z+$jets components of the background. More work is needed to understand possible systematic effects. It is desirable to understand the individual components of the backgrounds as well, and the various other methods discussed in this note appear to succeed in this.

The presence of SUSY signal would affect the background estimates, at a level that depends on the SUSY signal properties, as well as on the method. Methods with very tight control samples (replace method, topbox method) see almost no effect. For the other methods, the background is overestimated by typically 20–30% for samples like SU1, SU2, SU3 and SU6. If a SUSY excess is nevertheless observed (which is possible with 1 fb^{-1}), a correction for the background overestimation can be applied. First ideas have been presented in this note, using the MT method, and the combined fit method. More work is needed in this area. The SU4 benchmark point is a special case because of its light spectrum. It produces events with kinematics which are similar to the Standard Model backgrounds and its cross-section is high, so many methods would struggle to provide background predictions. It would, however, not be missed [11].

References

[1] ATLAS Collaboration, *The ATLAS experiment at the Large Hadron Collider*, paper submitted to JINST.

[2] ATLAS Collaboration, *Prospects for SUSY discovery based on inclusive searches*, this volume.

[3] ATLAS Collaboration, *Estimation of QCD backgrounds to searches for Supersymmetry*, this volume.

[4] ATLAS Collaboration, *Multi-lepton Supersymmetry searches*, this volume.

[5] ATLAS Collaboration, *Supersymmetry signatures with high p_T photons or long-lived heavy particles*, this volume.

[6] ATLAS Collaboration, *Supersymmetry searches with ATLAS*, this volume.

[7] ATLAS Collaboration, *The ATLAS trigger for early running*, this volume.

[8] J. Sjolin, J. Phys. G : Nucl. Part. Phys. **29** (2003) 543–560.

[9] S. Jadach, Z. Was, R. Decker and J.H. Kuhn, Comput. Phys. Commun. **76** (1993) 361–380.

[10] T. Sjostrand, S. Mrenna and P. Skands, JHEP **05** (2006) 026.

[11] ATLAS Collaboration, *Determination of top pair production cross-section in ATLAS*, this volume.

[12] ATLAS Collaboration, *Measurement of missing transverse energy in ATLAS*, this volume.