New Developments in Data-driven Background Determinations for SUSY Searches in ATLAS

F. Legger on behalf of the ATLAS collaboration

Max-Planck-Instutut für Physik, München, Germany

Abstract. Any discovery of new physics relies on detailed understanding of the Standard Model background. At the LHC, we expect to extract the backgrounds from the data itself, with minimum reliance on Monte Carlo simulations. We describe new developments in ATLAS on such data-driven techniques, and prospects for their application on first data.

**Keywords:** LHC, Supersymmetry, ATLAS, data-driven estimation, inclusive searches
PACS: 11.30.Pb, 14.80.Ly, 29.85.Fj

INTRODUCTION

High-energy jets, missing transverse energy (E_T^{miss}) and possibly leptons are the typical signature of R-parity conserved SUSY events at the LHC. The observation of an excess of events with these signatures with respect to the Standard Model prediction may manifest the presence of SUSY. Due to the poor knowledge of Standard Model cross-sections, parton distribution functions, underlying event and parton showering at the LHC energy scale, as well as insufficient knowledge of the detector itself, a reliable prediction of the Standard Model backgrounds should be derived mainly from the experimental data with reduced reliance on Monte Carlo simulations. In ATLAS the searches for a SUSY signal in events with a multijet and E_T^{miss} signature are classified according to the number of reconstructed leptons (electrons or muons). In this note we present an overview of data-driven methods recently developed by the ATLAS collaboration to estimate the Standard Model background in SUSY searches in the one-lepton mode. The detailed description of the cuts applied in the one-lepton mode can be found in [1]. The most important Standard Model processes contributing to the background in this mode are $t\bar{t}$ and W plus jets production.

m_T-METHOD: REVERSING THE m_T CUT

In Standard Model processes the main source of the E_T^{miss} and isolated leptons is the decay of the W boson (produced directly in the pp collisions or in the top quark decays). Consequently the invariant mass constructed from the transverse component of the isolated lepton and E_T^{miss} (called transverse mass, m_T) should not exceed the mass of the W boson in an Standard Model event. In SUSY events, the E_T^{miss} is produced mainly by the pair of non-interacting lightest supersymmetric particles and the restriction on m_T does not hold. Following this, a signal region is defined by $m_T > 100$ GeV, while the
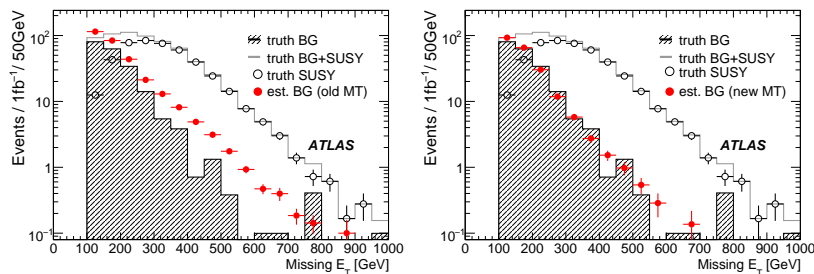
FIGURE 1. Missing transverse energy distribution for events passing one-lepton SUSY search cuts. Dashed histogram show the Standard Model background and circles show the result of data-driven estimation with m_T-method in the presence of SUSY with the mass scale of 600 GeV (SU3). Left (right) figure shows the estimate before (after) correcting for the SUSY contamination in the control sample.

events with $m_T < 100$ GeV constitute the control sample. The number of Standard Model events in the signal region and at high values of E_T^{miss} is derived from the control sample, applying the efficiency of the m_T cut. This efficiency is estimated in the low-E_T^{miss} region (normalization region), where the contribution of SUSY events is expected to be low.

The method is based on two assumptions. First, it requires low contribution of SUSY events both in the control sample and in the normalization region. Second, the method requires that E_T^{miss} and m_T are uncorrelated. The first assumption might not hold especially for low-mass SUSY. The second assumption is not true mainly because the fractions of semi-leptonic and di-leptonic top events depends on m_T. However the correlation between E_T^{miss} and m_T is not large, the linear correlation factor is about 4%. Due to the SUSY contamination in the control sample and in the normalization region, the estimate delivered by this method is above the true Standard Model background. The correlation between E_T^{miss} and m_T leads to background under-estimation for high-mass SUSY and to over-estimation for low-mass SUSY.

One can correct for the signal contamination in the control sample in the following way. First, the background in the signal region is estimated, and the remaining events are taken as the signal estimate. Then the signal estimate is translated back to the control region ($m_T < 100$ GeV) and subtracted from the background. Figure 1 compares the true Standard Model background to the estimated one in the presence of mSUGRA SUSY signal with mass scale of 600 GeV. The left (right) figure shows the comparison before (after) correcting the estimate for the SUSY contamination in the control sample. The correction procedure may not work for low-mass SUSY, as in this case the control sample contamination would be too large. Another disadvantage of the method is the need of an explicit assumption on the signal shape to translate it from the signal region to the control region.

COMBINED FIT

This method [1] estimates separately the contribution of the various Standard Model processes. Each Standard Model process, semi-leptonic $t\bar{t}$, di-leptonic $t\bar{t}$ and W (+ jets)
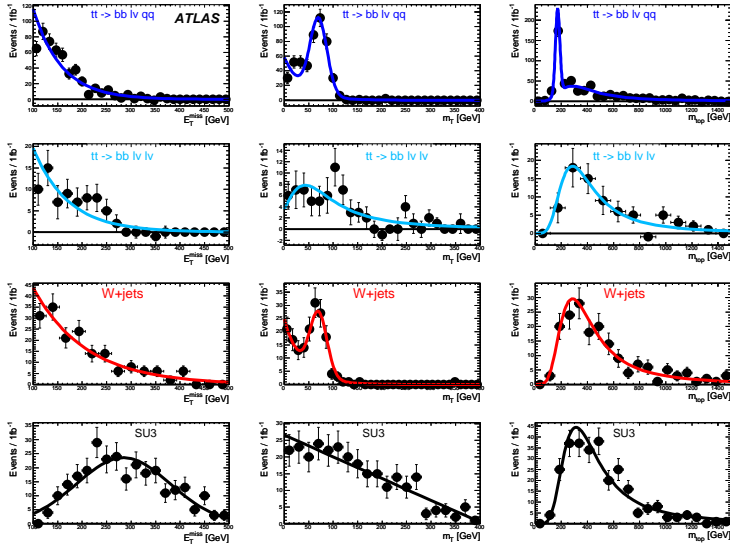
FIGURE 2. Distributions of E_T^{miss}, m_T and m_{top} in semi-leptonic and di-leptonic $t\bar{t}$ events, W (+ jets) events and SUSY events for mSUGRA point with SUSY scale of 600 GeV. Superimposed lines are the parametrizations used in combined fit method.

is parametrized in terms of the distributions of three variables, E_T^{miss}, m_T and m_{top}[1]. The parametrization is based on the Monte Carlo prediction for the distribution of the variables, as well as on the analysis of dedicated control samples. The contribution of SUSY signal is described with a "generic shape", derived from the shapes of the ATLAS benchmark SUSY points [2]. The simultaneous fit of the distributions of E_T^{miss}, m_T and m_{top} is performed for the sum of background components and the "generic shape" of the SUSY contribution. The fit parameters include the normalizations of each background and SUSY contributions as well as parameters of the backgrounds parametrization and the SUSY "generic shape". Figure 2 shows the distributions of the variables E_T^{miss}, m_T and m_{top} compared to the chosen parametrization for the background components and the ATLAS mSUGRA benchmark point with a mass scale of 600 GeV. The advantage of this method with respect to the m_T-method is the explicit accounting of the correlation between the variables.

TILES METHOD

In the Tiles Method [3] the m_{eff}–m_T plane[2] is divided in quadrants, called tiles. Each tile has contributions from SUSY signal and from the background. The fractions of the Standard Model events entering each tile is estimated from the Monte Carlo. The yields

---

[1] m_{top} is defined as an invariant mass of three jets with largest vector-summed transverse momentum

[2] Effective mass m_{eff} is defined as a scalar sum of transverse momenta of the four leading jets, transverse momenta of all identified leptons and E_T^{miss}

of SUSY and background events is estimated by solving the system of equations

$$
\begin{aligned}
n_A &= f_A \cdot n_{SM} + SUSY_A \\
n_B &= f_B \cdot n_{SM} + SUSY_B \\
n_C &= f_C \cdot n_{SM} + SUSY_C \\
n_D &= f_D \cdot n_{SM} + SUSY_D \\
SUSY_A \cdot SUSY_D &= SUSY_B \cdot SUSY_C
\end{aligned}
\tag{1}
$$

Here A, B, C and D denote four tiles, n_X is the measured number of events in the tile X, f_X is the fraction of Standard Model events in the tile X, n_{SM} is the unknown number of Standard Model events in all tiles and $SUSY_X$ is the unknown number of SUSY events in the tile X. The last equation comes from the assumption of no correlation between m_{eff} and m_T for signal events. The correlation of the variables for the Standard Model contribution is fully taken into account by the values of f_X. The method can be generalized to a $N \times N$ tiles configuration. In such a configuration the system of equations (1) is over-constrained and a solution can be found by minimizing the extended log-likelihood estimator constructed from the measured number of events in the tile and the expected number of events calculated as the sum of Standard Model and SUSY contributions.

The statistical error of the method depends on the tiles configuration and can be optimized for a given luminosity. With a small number of tiles the system only has a few degrees of freedom and the statistical fluctuations of the number of measured events in the tiles have a large influence on the estimated yields of signal and Standard Model events. In the configuration with many tiles, statistics in each tile is low, thus the statistical error of the estimate is high. The optimal number of tiles for a luminosity of 1 fb^{-1} is found to be 8×8. A configuration with many tiles has the advantage of the possibility to add new parameters to the fit, for example, a linear correlation factor, or separation of the various background contributions.

CONCLUSION

Several data-driven methods for the estimation of Standard Model background to the one-lepton SUSY searches are presented. The conventional m_T-method has been studied and several developments have been proposed. The over-estimate due to the presence of SUSY events in the control sample can be corrected by an iterative procedure. The combined fit technique and the Tiles Method allow the explicit accounting of variable correlation and presence of SUSY.

REFERENCES

1. G. Aad, et al., "Data-Driven Determinations of W, Z and Top Backgrounds to Supersymmetry," in *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512, CERN-OPEN-2008-020, 2009, pp. 1525–1561.
2. G. Aad, et al., "Supersymmetry Searches," in *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512, CERN-OPEN-2008-020, 2009, pp. 1514–1524.
3. G. Aad, et al., *Background Estimation for Inclusive SUSY Searches - The Tiles Method* (2009), aTLAS note ATL-PHYS-PUB-2009-077.