IL NUOVO CIMENTO Vol. 35 C, N. 3 Maggio-Giugno 2012
DOI 10.1393/ncc/i2012-11216-7

Colloquia: TOP2011

Estimation techniques for instrumental backgrounds at the LHC

J. Sjölin on behalf of the ATLAS and CMS Collaborations
Stockholm University - Stockholm, Sweden

ricevuto l' 1 Marzo 2012
pubblicato online il 28 Maggio 2012

Summary. — Backgrounds induced by the detector and its environment in high energy hadron collisions are present in all measurements at the LHC, and they can in many cases be significant. Experimental techniques are presented that use data to measure these backgrounds, minimize their impact, and derive corrections that can be applied to top physics measurements.

PACS 14.65.Ha – Top quarks.

1. – What is meant by instrumental backgrounds

Backgrounds can be a mix of instrumental and physics backgrounds. An example of a physics background is hadron decays in flight. Some clear examples of instrumental backgrounds are electronic noise, miscalibrations, dead material, bremsstrahlung, and additional proton-proton collisions (pile-up). These instrumental backgrounds can lead to mistakes in measuring particles in the decays of top quarks. Missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) and jets are significantly affected by pile-up, electronic problems and beam-halo. Muons, electrons, taus and photons can be falsely identified in jet fragments.

One important and difficult source of instrumental background is pile-up. Additional in-time collisions in the current bunch crossing can produce extra leptons and add tower energy that does not come from the collision of interest. A good handle on this is the number of reconstructed primary vertices. Collisions from other bunch crossings that are out of time (OOT) are harder to get a direct handle on. A useful observable that includes all collisions is the mean number of interactions per crossing (μ). It is calculated as $\mu = L \times \sigma_{\mathrm{inel}}/(n_{\mathrm{bunch}} \times f_{\mathrm{r}})$, where L is the instantaneous luminosity, $\sigma_{\mathrm{inel}} = 71.5\,\mathrm{mb}$ is the inelastic cross section, n_{bunch} is the number of colliding bunches, and $f_{\mathrm{r}} = 11245.5\,\mathrm{Hz}$ is the LHC revolution frequency, see fig. 1 which shows μ for the currently accumulated ATLAS data. A typical example of how intricate the dependence between in-time and OOT pile-up can be is shown in fig. 1 [1]. The figure shows the average tower energy in the electro-magnetic calorimeter as a function of the number of reconstructed primary vertices for collisions taking place inside the bunch train, for different μ intervals. The calorimeter bi-polar pulse shape contains a long negative component to restore the baseline under pile-up. This will only work inside the bunch train, and large pile-up

© CERN on behalf of the ATLAS and CMS Collaborations under CC BY-NC 3.0

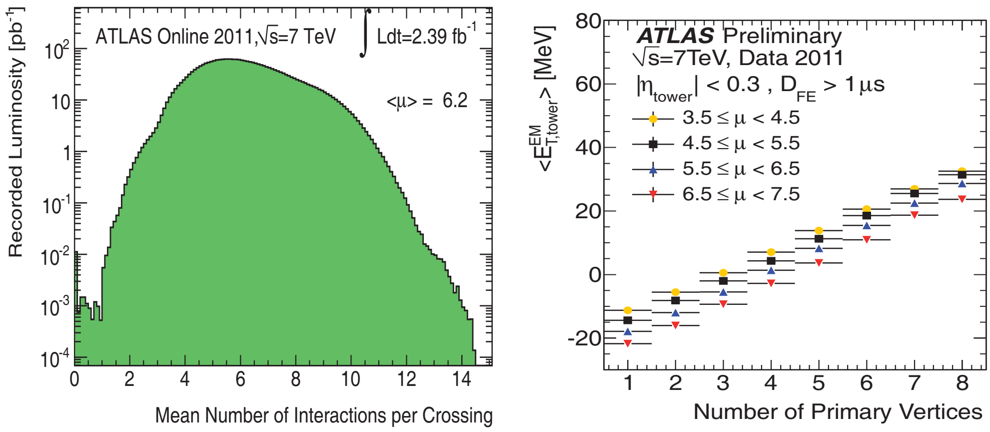

Fig. 1. – Mean number of interactions per bunch crossing in ATLAS (left), and the dependence of the mean tower energy in the electro-magnetic calorimeter *versus* the number of primary vertices, measured for collisions inside of the bunch train, shown for different μ intervals (right).

effects are expected at the edges due to the bunch train structure. However, even well inside the bunch train a dependence on the mean tower energy on μ is observed [2], which indicates OOT pile-up. At the moment ATLAS does not correct for these effects and it is added as a systematic uncertainty.

2. – Procedures before estimation and corrections

Before starting to estimate and correct for instrumental backgrounds the data must be cleaned by removing events with some indication of misreconstruction or significant beam-halo interactions. A typical procedure is to first require at least one good primary vertex —ATLAS requires at least five tracks— then to accept events with only clean jets. A clean jet has good timing in order to remove beam-halo, it passes the good quality criteria calculated from the energies measured in the calorimeter cells, and it has consistent levels of electro-magnetic and hadronic energy fractions. An example of how efficient the cleaning is can be seen in fig. 2, which shows $E_{\mathrm{T}}^{\mathrm{miss}}$ using particle flow (PF) in CMS before and after cleaning [3]. The plots show the result of standard event cleaning, in which the sum of track transverse momentum over H_T is greater than 0.1, and jets must be within $|\eta| < 2.4$. The particle flow technique improves the robustness against instrumental backgrounds compared to the traditional calorimeter vector sum. However, not even PF is safe against pile-up, since both in-time and out-of-time (OOT) pile-up affect the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. An example from CMS is shown in fig. 2, which shows the resolution measured in photon-jet balancing. The origin of the additional OOT pile-up in 2011 data is under investigation.

3. – Data-driven $E_{\mathrm{T}}^{\mathrm{miss}}$ estimation

Missing transverse energy has large instrumental uncertainties which are very difficult to model in Monte Carlo. In analyses that reject physics backgrounds based on the absence of real $E_{\mathrm{T}}^{\mathrm{miss}}$, this easily becomes a large systematic uncertainty. Two such

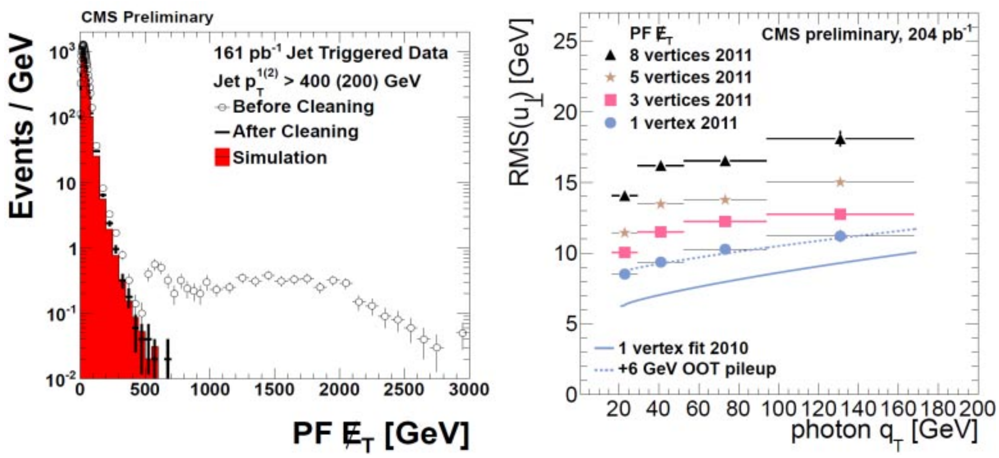

Fig. 2. – Missing transverse energy using particle flow (PF) before and after cleaning (left), and resolution measured in photon-jet balancing during different pile-up conditions (right).

examples are top analyses in the dielectron and dimuon channels. To get a robust estimate of the background yield both ATLAS and CMS use data-driven methods based on extrapolations from sidebands in data. The extrapolation needs an observable independent of $E_{\mathrm{T}}^{\mathrm{miss}}$, and for dileptons the common choice is the invariant mass of the the two leptons (m_{ll}). The sideband is defined as $|m_{ll} - m_Z| < \Delta$, and $E_{\mathrm{T}}^{\mathrm{miss}} > 30\,\mathrm{GeV}$. ATLAS uses $\Delta = 10\,\mathrm{GeV}$ and CMS uses $\Delta = 15\,\mathrm{GeV}$. The extrapolation using events from inside the sideband ($N^{\mathrm{in}}_{Z/\gamma^*}$) to the signal region ($R_{\mathrm{out/in}}$) is based on Monte Carlo. The estimate of the Drell-Yan ($N^{\mathrm{out}}_{Z/\gamma^*}$) background in the signal region then becomes $N^{\mathrm{out}}_{Z/\gamma^*} = R_{\mathrm{out/in}} N^{\mathrm{in}}_{Z/\gamma^*}$. The events in the sideband ($N^{\mathrm{in}}_{Z/\gamma^*}$) must be corrected for back-

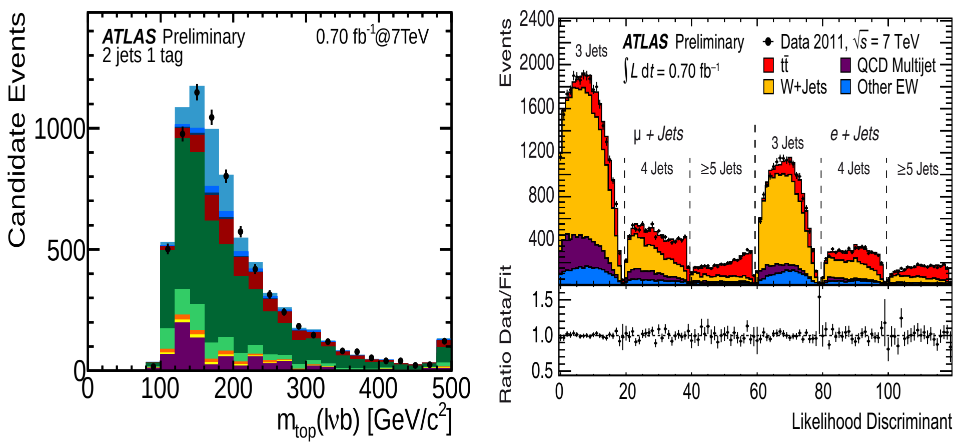

Fig. 3. – Single-top t-channel example of the jet-electron method for QCD estimation *versus* the top mass (left), and a top pair cross-section example where the discriminant QCD background is estimated with the matrix method (right).

grounds. ATLAS does this using Monte Carlo, while CMS uses efficiency-corrected events in the electron-muon channel. The final systematic uncertainty for both experiments is around 50% [4, 5].

4. – Data-driven estimation of fake leptons

Another very important instrumental background is fake leptons. In top physics fake leptons are defined as reconstructed leptons not originating from a real W or Z/γ^*. Typical sources for fake electrons are photons or neutral pions overlapping with a random track, unidentified conversions, semi-leptonic b-jet decays, and hadron decays in flight. Most fake leptons originate from jets, which make them both difficult to model and computationally heavy to generate. To overcome these two problems fake leptons are estimated using data. Fake muons are dominated by real muons and originate mostly from semi-leptonic b-jet decays and hadron decays in flight.

The most common methods for fake lepton estimation are all inherited from the Tevatron. They are called jet-electron, anti-electron, and matrix methods [6]. The jet-electron method uses the jet triggers and selects jets that are more likely to pass as an electron, *e.g.* have high electro-magnetic fraction and at least four tracks. The jet-electron events are then passed through the analysis. A shape for QCD is derived and the normalization is determined by a fit to the $E_{\mathrm{T}}^{\mathrm{miss}}$ spectrum. The same technique can be used also for muons using reversed isolation. The jet-electron method has been used by ATLAS for QCD estimation in the single-top t-channel analysis [7], yielding a top mass spectrum shown in fig. 3. The anti-electron model is similar to the jet-electron model but instead uses the electron trigger to reduce trigger bias. The anti-electrons are defined by reversing some of the electron identification cuts. The matrix method can be seen an extension of the anti-electron model, where the anti-electrons (or anti-muons) are used together with the signal leptons to form a system of equations. Since the matrix method is used to solve for the number of fake leptons, there no need for fitting, and the solved number of true leptons removes any issue of signal contamination in the fake lepton estimate. The fake lepton solution in the single-lepton channel reads

$$N_{\mathrm{fake}}^{\mathrm{tight}} = \frac{\epsilon_{\mathrm{fake}}}{\epsilon_{\mathrm{real}} - \epsilon_{\mathrm{fake}}}(N^{\mathrm{loose}}\epsilon_{\mathrm{real}} - N^{\mathrm{tight}}). \tag{1}$$

An example from ATLAS which uses the matrix method for the $t\bar{t}$ cross-section in the single-lepton untagged channel [8] is shown in fig. 3. Uncertainties in the untagged sample are around 50%.

REFERENCES

[1] `http://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults?rev=12`.
[2] `http://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2011Towerunderpileup`.
[3] CMS-DP-2011/010, `http://cdsweb.cern.ch/record/1369231`.
[4] The CMS Collaboration, CMS-TOP-11-002, `http://cdsweb.cern.ch/record/1354223`.
[5] ATLAS-CONF-2011-100, `http://cdsweb.cern.ch/record/1369215`.
[6] The ATLAS Collaboration, *Eur. Phys. J. C*, **71** (2011) 1577, arXiv:1012.1792 [hep-ex].
[7] ATLAS-CONF-2011-101, `http://cdsweb.cern.ch/record/1369217`.
[8] ATLAS-CONF-2011-121, `http://cdsweb.cern.ch/record/1376413`.