PROCEEDINGS OF SCIENCE

Data-driven methods for estimation of misreconstructed object background in lepton final states with ATLAS at $\sqrt{s} = 13$ TeV

Giulia Ucchielli* **on behalf of the ATLAS Collaboration**
Technische Universität Dortmund
E-mail: giulia.ucchielli@cern.ch

Neutrino mass generation is needed to explain neutrino oscillations. Several new physics models, such as left-right symmetric models or seesaw mechanisms, offer a solution and motivate searches for heavy bosons or heavy neutral leptons in either opposite-charge or same-charge leptons final states. This contribution discusses the challenging backgrounds for same-charge final states, either due to jets incorrectly identified as leptons or due to mismeasurements of the electron charge. The results presented here use the data collected by the ATLAS experiment at the LHC with a center-of-mass energy of $\sqrt{s} = 13$ TeV.

Sixth Annual Conference on Large Hadron Collider Physics (LHCP2018)
4-9 June 2018
Bologna, Italy

*Speaker.

PoS(LHCP2018)011

1. Introduction

Same-charge (SC) lepton final states provide a powerful signature towards new physics (NP) discoveries, thanks to the ATLAS detector's [1] excellent performance in lepton reconstruction, carrying low associated systematic uncertainties. Moreover, the majority of Standard Model (SM) processes produce two opposite-charge (OC) leptons, e.g. a $t\bar{t}$ event with subsequent dileptonic decay. A challenging background for SC final states arises from misreconstructed objects, as a consequence of particle-detector interaction.

2. Charge misidentification probability

Charge misidentification mainly occurs because of bremsstrahlung emission followed by photon conversion (the so-called *trident* event: $e^{\pm} \to e^{\pm}\gamma \to e^{\pm}e^{+}e^{-}$). The final state electron can be reconstructed with incorrect charge, if e.g. the information from the electromagnetic (EM) calorimeter is matched to the wrong electron inner detector (ID). To precisely quantify the probability for such an event to occur, simulation would need to accurately model particle-detector interaction and to provide a detailed description of the detector material. However, the probability based on simulation can be off by 10-20% and dedicated correcting procedures, based on data, are applied. First, pairs of OC and SC electrons are selected, if their $m(ee)$ invariant mass satisfies the following criteria: the Z peak region selects ee pairs with $|m(ee) - m_{\mathrm{OC}}(Z)| < 14$ GeV and $|m(ee) - m_{\mathrm{SS}}(Z)| < 15.8$ GeV. Two symmetric sideband regions are used to estimate and subtract the background from the Z peak, defined by 14 GeV $< |m(ee) - m_{\mathrm{OC}}(Z)| < 18$ GeV and 15.8 GeV $< |m(ee) - m_{\mathrm{SS}}(Z)| < 31.6$ GeV. The OC/SC peak regions are presented in Fig. 1, where the SC peak is clearly shifted, by approximately 2 GeV, to lower energies, and it is sligthly broader than the OC peak, due to bremsstrahlung energy loss. This motivates a few GeV difference when defining the Z peak region (as stated before). The total number of $Z \to ee$ events in the Z peak is[1] $N^{ij} = N^{ij}_{OC} + N^{ij}_{SC}$. Being ε_i the probability for an electron to *flip* its charge[2], the probability to observe N^{ij}_{SC} events is poissonian:

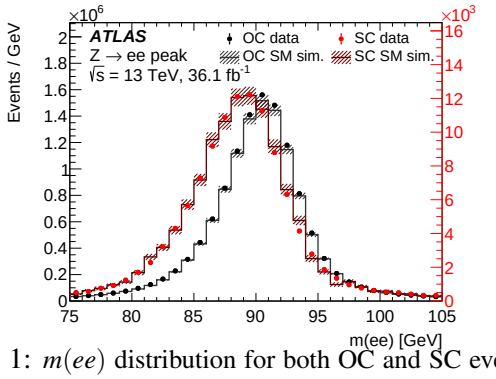


Figure 1: $m(ee)$ distribution for both OC and SC events in data and simulation for data corresponding to 36.1 fb^{-1} [2].

$$f(N^{ij}_{SC};\lambda) = \frac{\lambda^{N^{ij}_{SC}} e^{-\lambda}}{N^{ij}_{SC}!} \tag{2.1}$$

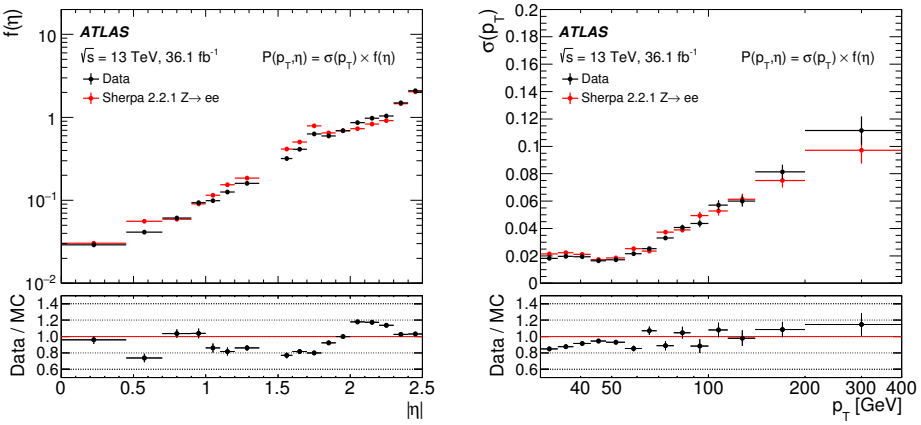
with $\lambda = (\varepsilon_i(1-\varepsilon_j) + \varepsilon_j(1-\varepsilon_i))N^{ij}$. The minimization of $-\ln\mathscr{L}$, where $\mathscr{L}$ is:

$$\mathscr{L}(\varepsilon|N_{SC},N) = \prod_{i,j} \frac{[N^{ij}(\varepsilon_i+\varepsilon_j)]^{N^{ij}_{SC}} e^{-N^{ij}(\varepsilon_i+\varepsilon_j)}}{N^{ij}_{SC}!} \tag{2.2}$$

[1] Here i and j indicate the kinematic configuration, in terms of p_{T} and $|\eta|$, of the electrons in the pair.

[2] Neglecting $\mathscr{O}(2)$ terms in ε where both electrons flipped their charged.

returns the measured ε_i and ε_j, which depend on both the electron $|\eta|$ and p_{T}, as shown in Fig. 2. The data over simulation ratio, displayed in Fig. 2, is used as a *scale-factor*[3] (SF) to correct the simulation. The truth origin of the electron is checked: the SF is applied to charged-flipped electrons while the anti-SF is applied to electrons with correct charge.



(a) (b)

Figure 2: Charge misidentification probability $\varepsilon(p_{\mathrm{T}},\eta) = \sigma(p_{\mathrm{T}}) \times f(\eta)$, applied as a 1D×1D parametrisation of electron $|\eta|$ (a) and p_{T} (b) [2].

3. Fake lepton background

Fake and *non-prompt*[4] leptons are objects wrongly reconstructed as leptons originating from the interaction point (named *prompt* leptons). A non-prompt lepton arises e.g. from hadron decays inside a jet, and is a real lepton *faking* its actual origin. Additionally, jets where the electrically charged component provides a signal in the ID and a calorimetric energy deposit, can fake electrons. A simulated sample with a large number of events would be needed to properly describe such events. Therefore, dedicated data-driven techniques are used. One of the most frequently used methods for fake estimation is the *fake-factor* method [2]. This method exploits two different lepton definitions:

- a *tight* sample containing leptons passing stringent identification and isolation criteria, which are used to define the analysis regions;
- a *loose* sample required to fail i.e. the tight identification or isolation requirements, since fake leptons are usually less isolated.

The tight-to-loose ratio is used to compute the fake-factor (F):

$$F = \frac{N_{\mathrm{tight}}}{N_{\mathrm{loose}}} \tag{3.1}$$

which is parametrised in bins of p_{T} and $|\eta|$ and measured in regions designed to be enriched in fake leptons, as reported in Table 1. Once F is measured, it can be used to estimate the number of

[3]The SF is defined as SF $= \varepsilon(\mathrm{data})/\varepsilon(\mathrm{MC})$ while the anti-SF is anti-SF $= (1-\varepsilon(\mathrm{data}))/(1-\varepsilon(\mathrm{MC}))$.

[4]Collectively called *fakes* in the following.

events containing *at-least* one fake lepton in any analysis region[5] with a similar composition of fakes. Analysis regions are identified by the presence of tight leptons and in a simple two-lepton case, the contribution from the fake background follows from[6]:

$$N^{\text{fake}} = \left[F(N_{TL}+N_{LT})-F^2N_{LL}\right]_{\text{data}} - \left[F(N_{TL}+N_{LT})-F^2N_{LL}\right]^{\text{prompt}}_{\text{MC}} \quad (3.2)$$

Selection for fake-enriched regions	
Muon channel	Electron channel
Single-muon trigger	Single-electron trigger
b-jet veto	b-jet veto
One muon and one jet	One electron
$p_{\text{T}}(\text{jet}) > 35$ GeV	Number of tight electrons < 2
$\Delta\phi(\mu,\text{jet}) > 2.7$	$m(ee) \notin [71.2, 111.2]$ GeV
$E_{\text{T}}^{\text{miss}} < 40$ GeV	$E_{\text{T}}^{\text{miss}} < 25$ GeV

Table 1: Selection criteria defining the fake-enriched regions used to measure F for the μ and e channels [2]. Both regions are dominated by dijet events.

where N_{TL}, N_{LT} and N_{LL} are the events containing at least one loose lepton. It is worth noticing that when the size of N_{TL}, N_{LT} and N_{LL} is zero, the fake-factor computation leads to a 0 ± 0 estimated fakes. The problem of insufficient data statistics can be solved by computing the poissonian probability to observe at least one event at 68% CL. Solving $P(0|\lambda) = 32\%$ for λ gives 1.14, equally divided into $N_{TL} = N_{LT} = N_{LL} = 0.38$. By choosing typical values for F, such as $F_e = 0.5$ and $F_\mu = 0.9$, and inserting them into Eq.(3.2):

$$N_e^{\text{fakeup}} = 2\times0.38\times0.5 - 0.38\times0.5^2 = 0^{+0.29}_{-0.0} \quad (3.3)$$

$$N_\mu^{\text{fakeup}} = 2\times0.38\times0.9 - 0.38\times0.9^2 = 0^{+0.38}_{-0.0} \quad (3.4)$$

upper statistical uncertainties on a zero background prediction are derived.

4. Systematic uncertainties

The uncertainty associated to the measurement of the charge misidentification probability arises from the statistical uncertainty of both data and simulated $Z \to ee$ events; it ranges between 10% and 20% as a function of the electron p_{T} and η. The uncertainty on F arises from the statistical uncertainty and from the unknown fake composition. The latter is assessed by varying the nominal F selection criteria (summarized in Table 1). Moreover, the uncertainty on the yield of prompt leptons from W/Z boson decays is estimated by varying the total yield of simulated samples by $\pm10\%$, corresponding to the size of the QCD scale, α_s and PDF uncertainties. The total uncertainty on F varies between 10% and 50% across p_{T} and η bins.

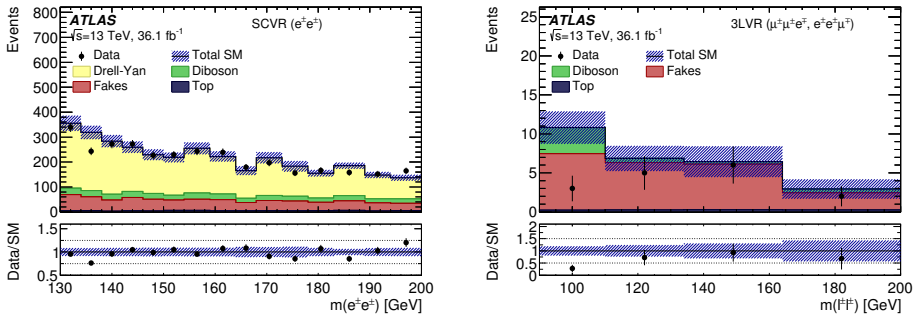
5. Methods used in new physics searches

The methods discussed in Sections 2 and 3 are applied to two NP searches, proving their adaptability over several final states with different lepton/jet multiplicities. The search for doubly charged Higgs boson pair production ($pp \to H^{\pm\pm}H^{\mp\mp} \to \ell^\pm\ell^\pm\ell^\mp\ell^\mp$) [2] has 2/3/4 leptons in the final state and is inclusive in jet multiplicity. Depending on lepton multiplicity and flavour, the major backgrounds are Drell–Yan, diboson production and fake leptons. The search for heavy

[5] Analysis regions are divided into *control regions* (CRs), where background normalizations are constrained, *validation regions* (VRs) used to validate the background prediction and *signal regions* (SRs).

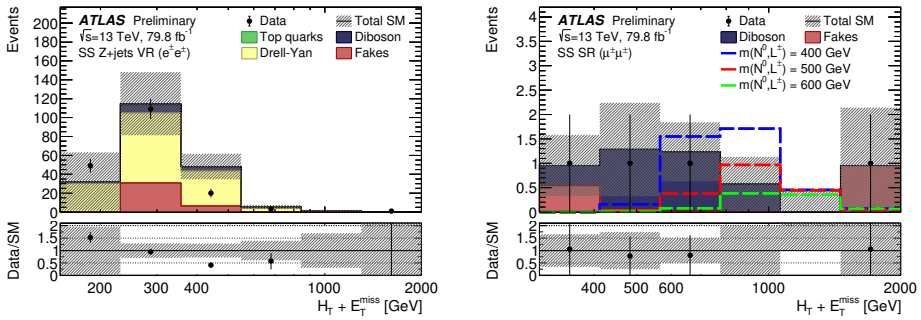
[6] The residual prompt lepton component, from W/Z boson decays, is subtracted using simulation.

lepton multiplets [3] is motivated in type-III seesaw models ($pp \to N^0 L^\pm$, with $N^0 \to W^\pm \ell^\mp$ and $L^\pm \to W^\pm \nu$), where the final signature contains two OC or SC leptons, two jets and missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$. The major backgrounds are Drell–Yan, diboson, top and fakes.



(a) (b)

Figure 3: Invariant mass distribution of the SC pair in a two SC electrons (a) and in a three-lepton (b) VR in the search for $H^{\pm\pm}$ production [2].



(a) (b)

Figure 4: Scalar sum of objects p_{T} (H_{T}) plus $E_{\mathrm{T}}^{\mathrm{miss}}$ in a SC leptons (a) VR and in the $\mu^\pm\mu^\pm$ SR (b) in the type-III seesaw search [3].

In both analyses, a binned maximum-likelihood fit is performed. Normalization factors for simulated backgrounds are extracted from CRs and extrapolated to VRs and SRs. The prediction for fake events is left free to vary within its total uncertainty. Figures 3 and 4 show the validity of the backgrounds estimation in regions dominated by Drell–Yan production and fake leptons. The total predicted backgrounds agree with the observed data within the total uncertainty.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[2] ATLAS Collaboration, *Search for doubly charged Higgs boson production in multi-lepton final states with the ATLAS detector using proton-proton collisions at* $\sqrt{s} = 13$ TeV, Eur.Phys.J. **C78** (2018) 199.

[3] ATLAS Collaboration, *Search for type-III seesaw heavy leptons in proton-proton collisions at* $\sqrt{s} = 13$ TeV *with the ATLAS detector.* ATLAS-CONF-2018-020. https://cds.cern.ch/record/2621484