Data-driven estimate of the multijet background to the selection of events with a Z boson and a b-jet

G. Chiodini [a] , N. Orlando [a,b] and S. Spagnolo [a,b] and the ATLAS Collaboration

[a]Istituto Nazionale di Fisica Nucleare sez. di Lecce, Italy.

[b]Dipartimento di Matematica e Fisica “Ennio De Giorgi”, Università del Salento, Italy.

The selection of events with a Z boson and b-jets doesn't suffer of large backgrounds; the main residual contributions are top pair-production, di-boson production, W+jets and multi-jet events. Multijet events fake signal events rarely. In order to mimic a $Z + b - jet$ event, a multijet event must contain two fake prompt leptons (either muon or electron from decays of heavy flavored hadrons, or muon from decays in fligh of π or K or jets or photons reconstructed as electrons) of the same flavor and opposite charge with invariant mass accidentally close the Z boson mass. In spite of the low probability of spill out of such events into the signal region, the extremely high rate of strong hard scattering in pp collisions, implies a non null contamination of the selection.

The fraction of multijet events in the inclusive Z+b selection is estimated with data driven methods to overcome the systematic uncertainties from the limited statistics of the available simulated background samples and from the modeling of the rate of misidentification of jets as leptons. The data driven procedure applied, both in the electron and muon channel, is based on the following assumptions. A given experimental distribution, receiving contributions from multijet background, in addition to other physics processes, can be modeled with:

$$F(x_i; \theta_j, N_{\overline{QCD}}, N_{QCD}) = N_{\overline{QCD}}\, g(x_i) + N_{QCD}\, f(x_i; \theta_j) \quad , \tag{1}$$

where $N_{\overline{QCD}}$ and N_{QCD} are the total number of non-multijet and multijet events passing the selection, g and f are the unitary-normalised distributions of the variable x (with value x_i in bin i or in event i, in case of an unbinned distribution) for non-multijet and multijet events. The non-multijet model $g(x_i)$ is estimated with MC simulations and the model for the multijet background $f(x_i; \theta_j)$ is extracted from data, in background enriched control regions. The distribution f is expressed as an analytic function of the variable x defined by the parameters θ_j, which are constrained by fitting the distribution to data in the control sample. If the shape of the distribution for multijet events is stable when moving from the background enriched sample to the selected sample, the amount of multijet background can be estimated by fitting equation 1 to the distribution observed in the signal region, with N_{QCD} and $N_{\overline{QCD}}$ are free parameters of the fit. The method relies on the assumption that the relative abundance and the shape of the various non-multijet processes is well described in simulation.

For both the electron and muon channels, the invariant mass distribution is used as the discriminating variable, modelled according to equation 1 where the shape of the non-multijet background $g_l(m_{ll})$ is a template obtained from simulation and the multijet contribution is parameterised by an exponentially decaying function, with slope α_l. The α_l parameters are obtained separately for electron and muon channels. Once the shape of the multi-jet background has been determined a background enriched control region for a given channel, the normalisation of the QCD background in the signal region is determined by fitting again the dilepton invariant mass distribution in the signal region with the same model used in the control region. Finally, the shape of the multi-jet background in the template fit of $\ln(p_b/p_c)$ is taken from data in a control region in which there is one jet tagged as originating from a b-quark.

As an example the procedure and results obtained in the case of the multijet background in the Z(ee)+b-jet channel are described in the following. A background enriched control region can be obtained by loosening or reversing the electron identification criteria adopted in the analysis. In addition, in order to increase the statistics, the selection of the control sample can be run up to the level of the Z selection or to the level of the Z+jet selection, in order to derive α_e. Several di-electron definitions, listed below, were used:

- 1st e: `Loose-NotMedium`; 2nd e: `LooseNotMedium` ($L\overline{M} - L\overline{M}$)
- 1st e: `Medium`; 2nd e: `NotMedium` ($M - \overline{M}$)

where `Medium` refers to the quality criteria applied to electron candidates in the signal selection, `NotMedium` is the reversed criteria and `Loose-NotMedium` refers to electrons passing loose identification criteria, but failing the `Medium` quality criteria.

Control region	Data	Z	$t\bar{t}$	Di-boson	single-top	selection
$M-\overline{M}$	338094.0	120606.3	754.2	860.5	169.7	Z
$L\overline{M}-L\overline{M}$	16742.0	2804.9	1.6	3.5	0.3	Z
$M-\overline{M}$	113589.0	24946.6	698.4	787.3	142.3	Z+jet
$L\overline{M}-L\overline{M}$	4984.0	545.3	1.5	2.9	0.2	Z+jet

Table 1
Number of events selected in data for each on the control regions defined for the extraction of the shape of the multijet background. The expected number of events in simulation for the various non-multijet processed contributing are also reported.

Control region	α	$N_{\overline{mj}}$	N_{mj}	selection
$M-\overline{M}$	-0.022 ± 0.001	126760.1 ± 593.1	211331.1 ± 660.5	Z
$L\overline{M}-L\overline{M}$	-0.026 ± 0.001	2374.6 ± 115.1	14367.3 ± 158.8	Z
$M-\overline{M}$	-0.020 ± 0.001	27400.9 ± 346.7	86187.5 ± 423.1	Z+jet
$L\overline{M}-L\overline{M}$	-0.022 ± 0.001	404.3 ± 60.1	4579.5 ± 88.2	Z+jet

Table 2
The α_e parameter obtained in the various control regions and the normalization of multijet and non-multijet processes extracted from the fit of the invariant mass spectrum. The sum of the number of events for non-multijet processes is consistent with the MC expectations reported in table 1.

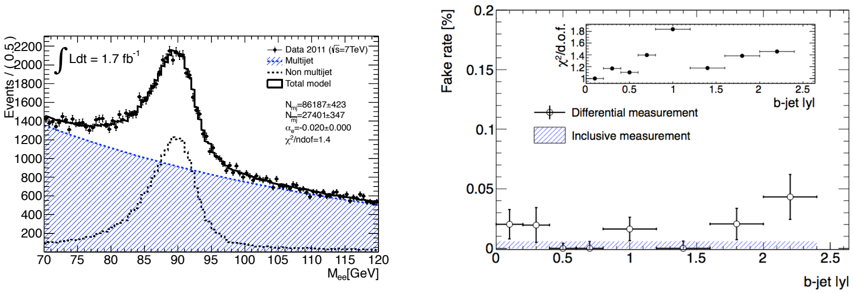

Figure 1. Left: Fit to the invariant mass distribution of a control region for the Z+jet selection. Right: QCD contamination of the Z+b-jet selection as a function of the b-jet rapidity.

To measure α_e, the model in equation 1, with a floating exponential slope in addition to signal and background normalisation, was fit to the invariant mass distributions corresponding to the two different object definitions in a wide mass range (70-140 GeV) around the Z mass. The fit results for the parameter α_e is shown in Fig 1, for $M-\overline{M}$ pairs satisfying the Z+jet selection. In table 1 the total statistics selected in data for the various control regions is reported along with the MC prediction for the electroweak processes.

From table 2, the value of the parameter α_e measured in the variuos control region is consistent and the normalization of sum of the electroweak processes, determined from the fit to the data, is also found in reasonable agreement with the prediction reported in table 1.

Using the shape of the multi-jet background defined by the value of α_e obtained from the control region, the invariant mass distribution of the di-electron pairs in teh signal region is fit to the sum of the electroweak processes (with total normalization free in the fit) and to the QCD background, with normalization free in the fit. The result, i.e. the estimated fractional contamination of the selection from multi-jet events is reported in figure 1 as a function of the b-jet rapidity.

REFERENCES

1. ATLAS Coll.(G. Aad et al.), Phys. Lett. B706 (2012) 295-313.