Measurement of the cross-section for b-jets produced in association with a Z boson at $\sqrt{s} = 7$ TeV with the ATLAS detector: analysis strategy and b-jet yield extraction

G. Chiodini [1] , N. Orlando [1,2] and S. Spagnolo [1,2] and the ATLAS Collaboration

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

[2]Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, Italy.

A measurement of the inclusive total cross-section for b-jet production in association with a Z boson in pp collisions at a centre-of-mass energy of $\sqrt{s} = 7$ TeV was published [1] by the ATLAS experiment with the 2010 data set, corresponding to an integrated luminosity of about 35 pb^{-1}. The production of Z bosons in association with jets at hadron colliders allows to test perturbative QCD (pQCD) calculations and modeling of the non perturbative features of events with heavy flavors. The relavance of this studies comes from the consideration that the production of one or more b-jets in association with a Z boson is a significant background to many important searches at the LHC, such as the Standard Model Higgs search, SUSY searches and searches for other physics beyond the Standard Model. With the large statistics of the 2011 ATLAS data set, corresponding to 5 fb^{-1}, the properties of events with b-jets and a Z boson can be studied in more details; in particular the differential production cross sections can be measured as a function of the b-jet and of the Z boson kinematics (transverse momentum and rapidity). In addition, the cross section for the production of Z and at least 2 b-jets can be measured with the available statistics. While the precision of the measurement based on the 2010 data was statistically limited, with the 2011 data a very good control of the systematics uncertainties is necessary in order to avoid spoiling the statistical precision.

The object definition and event selection criteria that are used in the Zb and Zbb analyses, are listed in Table 1. In addition the signal is defined by the requirement that one or more jets are identified as originating from a b-quark. A neutral network-based algorithm is used to tag a jet as a candidate b-jet. It is based on variables related to the secondary vertex identified in the jet, as the invariant mass of the system of tracks originating from it, the significance of the displacement with respect to the pp collision point, etc. It is used at the 75% efficiency operating point, corresponding to a cut $w > 0.404219$, where w is the weight in output from the neural network. The b-tagging algorithm retain a very small efficiency for light quark jets and a quite high efficiency for charm jets. Therefore, one of the most crucial steps of the data analysis is the extraction of the b-jet yield after the event and b-jet selection obtained with the criteria described so far, through its separation from the light and charm contamination in the sample of selected jets.

Event vertex and quality	≥ 3 tracks
e sel. cuts	author 1 or 3 $MediumPP$ quality passed $1.37 < \|\eta_e\| < 1.52$ excluded $\|\eta_e < 2.47$ e track $z_0 < 1\,\mathrm{mm}$, $\|d_0/\sigma\| < 10$
$Z \to ee$ sel.	2 from 2 sel. e of opp. charge no further e or μ in event $76 < M_{ee}(\text{ GeV}) < 106$
μ sel. cuts	STACO combined μ $z_0(\mu) < 1\,\mathrm{mm}$, $\|d_0/\sigma(d_0)\| < 3$ $\|\eta_\mu < 2.4$, $p_T(\mu) > 20$ GeV isolation: $\Sigma p_{\mathrm{T}}(ID)/p_{\mathrm{T}} < 0.1$
$Z \to \mu\mu$ sel.	2 sel. μ no sel. e with $\Delta R(e,\mu) > 0.1$ $76 < M_{\mu\mu}(\text{ GeV}) < 106$
Event sel. based on $E_{\mathrm{T}}^{\mathrm{miss}}$	pass "looser" $E_{\mathrm{T}}^{\mathrm{miss}}$ cleaning $E_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV
Jet sel. cuts	Anti-k_T jets, $R = 0.4$ $p_T(\text{jet}) > 20$ GeV $\|y(\text{jet}) < 2.4$ $\Delta R(\text{jet signal } \ell) > 0.5$ $JVF > 0.75$ pass "looser" cleaning cuts jet not in LAr hole
Event sel. based on jets	Zb: ≥ 1 sel. jet Zbb: ≥ 2 sel. jets

Table 1
Object and event selection criteria used for reconstructed events. For generator level selection, the same kinematic cuts are applied on truth variables, where applicable.

An extended maximum likelihood binned fit to the distribution of $\ln(p_b/p_c)$ is used as the baseline fit approach to extract the b-jet yield in the signal region. p_b/p_c is the ratio of the probability that a given jet is a b-jet and the probability that it is a c-jet; the probabilities are computed based on the neural network output. Fitting this distribution was found to give the best precision as compared to e.g. SV0 mass (the invariant mass of the tracks associated to the secondary vertex) used in the 2010 analysis. This precision is due to the low correlation, induced by this discriminating variable, of the *beauty* template with the *light* and *charm* templates. However, a feature of the $\ln(p_b/p_c)$ discriminant is that the *light* and *charm* templates are almost entirely correlated. Although fitting them individually gives consistent results, the central fit combines the *charm* and *light* templates to have a well constrained fit. Thus the two variables floated free in the fit are the *beauty* and the *charm* + *light* template normalisations. Two dimensional combinations of discriminating variables were also studied, but did not yield superior results in terms of precision on the b-jet yield and tended to be slightly less stable especially when looking at differential fits in e.g. b-jet p_{T}.

The QCD multi-jet background is subtracted from each distribution using the estimatation provided by a data driven method.

For the central results, the electron and muon channels are first combined for each distribution by simply adding the two sets of data, MC templates and backgrounds together. The procedure is shown to improve the statistical precision and is robust against systematic effects related to flavour discriminant shapes and unfolding. A cross-check of the individual electron and muon channels is also performed and gives consistent yields.

The combined data are compared to MC after the template fit results have been used to define the normalisation of the *beauty* and *charm*+*light* templates, for all of the differential bins of the analysis; an example of the result in one analysis bin is given in figure 1

The fit bias in the number of b-jet estimation is checked with a toy Monte-Carlo ensemble test. A set of pseudo-data is generated according to the flavour composition estimated in data; then a flavour fit is performed to each pseudo-dataset. The procedure is repeated separately in each analysis bin. The results of the ensemble test can be summarised in terms of reduced χ^2, see Fig. 2 for an example, and the corresponding fit probabilities. The fit quality has been studied for all bins of the analysis, separating events according to the b-jet $\mathrm{p_T}$, rapidity, the Z $\mathrm{p_T}$ and the separation in ϕ and η between the Z and the leading b-jet.

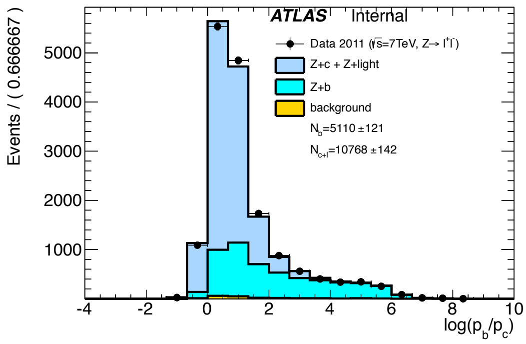

Figure 1. Flavor fit in the bins 20 $\leq$b-jet p_{T}[GeV]$<$ 30.

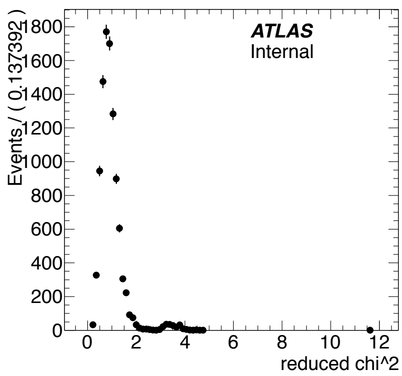

Figure 2. Reduced χ^2 distribution in the bins 20 $\leq$b-jet p_{T}[GeV]$<$ 30.

The derivation of the differential cross section proceed through the correction of the measured distributions of the kinematic variables for the detector effects such as efficiency and acceptance, which imply a bin dependent scaling of the observed rate of events, and resolution which is responsible for bin to bin migration of the events. Finally the distributions are converted in differential cross sections taking into account the integrated luminosity recorded by the experiment in exactly the same detector quality conditions as required by the analysis.

Finally, the measurements are compared to theory predictions after a carefull matching of the operative definition of the observables in the experimental data and in the calculation.

REFERENCES

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