Double B-hadron Jet Tagging and Identification of Gluon to $b\bar{b}$ jets with the ATLAS Detector

Lic. María Laura González Silva

Tesis Doctoral en Ciencias Físicas

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Noviembre 2012

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

Double B-hadron Jet Tagging and Identification of Gluon to $b\bar{b}$ jets with the ATLAS Detector

Trabajo de Tesis para optar por el título de

Doctor de la Universidad de Buenos Aires en el área Ciencias Físicas

por **María Laura González Silva**

Director de Tesis: Dr. Ricardo Piegaia

Consejero de estudios: Dr. Daniel Deflorian

Lugar de Trabajo: Departamento de Física (CONICET-UBA)

Buenos Aires, 2012

AGRADECIMIENTOS

Agradezco a...

Abstract

This thesis describes a method that allows the identification of double B-hadron jets originating from gluon-splitting. The technique exploits the kinematic differences between the so called "merged" jets and single B-hadron jets using track-based jet shape and jet substructure variables combined in a multivariate likelihood analysis. The ability to reject b-jets from gluon splitting is important to reduce and to improve the estimation of the b-tag background in Standard Model analyses and in new physics searches involving b-jets in the final state. In the simulation, the algorithm rejects 95% (50%) of merged B-hadron jets while retaining 50% (90%) of the tagged b-jets, although the exact values depend on the jet p_T.

Contents

Chapter 1

Double b-hadron jet identification

In this chapter we focus on the understanding of the internal structure of b-jets containing two b-hadrons by investigating the differences between these and single b-quark jets. These differences are expected to arise from the two-subjet (two b-hadrons) structure of double b-hadron or "merged" jets, which would tend to be wider and with a larger number of constituents. Based on these envisaged characteristics, simulated QCD samples of b-tagged jets were used to explore properties with potential discrimination power. The Monte Carlo distributions were compared to data from the 2011 run for validation. We present results from these studies and discuss the choice of the observables selected to build the multivariable tool presented in Chapter **??**.

1.1 Data Analysis

The tagging technique presented in this thesis relies on Monte Carlo predictions for the signal (single b) or background (merged b) hypotheses. The accuracy of the simulation is validated with data by comparing the distributions of the different variables studied.

The data samples employed correspond to proton-proton collisions at $\sqrt{s} = 7$ TeV delivered by the LHC and recorded by ATLAS between May and November 2011, with the LHC running with 50 ns bunch spacing, and bunches organized in bunch trains. Only data collected during stable beam periods in which all sub-detectors were fully operational are used. After the application of the data quality selection, the surviving data corresponds to an integrated luminosity of 4.7 fb^{-1}. The LHC instantaneous luminosity steadily increased during 2011. As a result, the average number of minimum-bias pile-up events, originating from collisions of additional protons in the same bunch as the signal collision, grew from from 3 to 20. This fact will be of importance when discussing the selection of discriminating variables.

The Monte Carlo event generators discused in Section **??** are used here. Samples of dijet events from proton-proton collision processes were simulated with Pythia version 6.423 [1], used both for the simulation of the hard $2 \to 2$ process as well as for the parton shower, underlying event, and hadronization models. The ATLAS AMBT2 tune of the soft model parameters was used [2]. In order to have sufficient statistics over the entire p_T spectrum, eight samples were generated with different thresholds of the hard-scattering partonic transverse momentum $\hat{p}_T$. Events from different samples were mixed taking into account their respective production cross sections. The simulated data sample used for the analysis gives an accurate description of the pile-up content and detector conditions for the full 2011 data-taking period.

1.1.1 Event selection

The event selection and quality criterion used to extract, from the data and Monte Carlo samples, the final set of jets for the analysis comprises different steps:

- **Trigger**. The event sample was collected using the ATLAS single jet triggers which select events with at least one jet with transverse energy above a given threshold. At the hardware Level 1 and local software Level 2 (see Section **??**), cluster-based jet triggers are used to select events with high-pt jets. The Event Filter, in turn, runs the offline anti-k_t jet finding algorithm with $R = 0.4$ on topological clusters over the complete calorimeter. At this stage, the transverse energy thresholds, expressed in GeV, are: 20, 30, 40, 55, 75, 100, 135, 180. These triggers reach an efficiency of 99% for events having the leading jet with an offline energy higher than the corresponding trigger thresholds by a factor ranging between 1.5 and 2. The jet triggers with the lowest p_T thresholds were prescaled by up to five orders of magnitude, and typically the same jet trigger is prescaled ten times more in the later data taking periods compared to the early ones.

- **Primary vertex**. The offline event selection requires at least one primary vertex candidate with 5 or more tracks. No requirements are placed on the longitudinal position (along the beam line) of the vertex as the beam spot is used as a constraint when fitting the vertex.

- **Primary jet algorithm**. The jet algorithm selected for the analysis was the ATLAS default anti-k_t algorithm [3], with a distance parameter $R = 0.4$, using calorimeter topological clusters [4] as input.

- **Jet calibration**. The EM+JES calibration scheme, described in Section**??**, was used to correct the jet energies for inhomogeneities and for the non-compensating nature of the calorimeter.

- **Jet quality**. Several quality criteria are applied to eliminate "fake" jets that are caused by noise bursts in the calorimeters and energy depositions belonging to a previous bunch crossing [5].

- **Jet tagging**. Only jets tagged as b-jets using the MV1 b-tagging algorithm at the 60% efficiency working point were considered.

- **Isolation**. b-tagged jets with close-by jets ($\Delta R < 0.8$) with p_T higher than 7 GeV at electromagnetic scale were not included in the analysis.

All jets, with transverse momentum between 40 and 480 GeV, the selected p_T range for the analysis, were required to be in a region with full tracking coverage, $|\eta_{jet}| < 2.1$, and they were classified in eight p_T bins chosen such as to match the jet trigger 99% efficiency thresholds (in GeV): 40, 60, 80, 110, 150, 200, 270, 360. An event is used if it satisfies the highest threshold trigger that is 99% efficient for the p_T bin that corresponds to the p_T of its leading jet.

In the case of MC, the reconstructed b-tagged jets were further classified into single and merged b-jets based on truth Monte Carlo information. A b-hadron is considered to be associated to a jet if the ΔR distance in $\eta - \phi$ space between the direction of the hadron and the jet axis is smaller than 0.4. Jets were labeled as merged (single) b-jets if they contain two (only one) b-hadron:

$$\text{single } b\text{-jets: } \Delta R(j, b_{1/2}) < 0.4 \tag{1.1}$$

$$\text{merged } b\text{-jets: } \Delta R(j, b_1) < 0.4 \ \& \ \Delta R(j, b_2) < 0.4 \tag{1.2}$$

where j is a jet in the event and $b_{1/2}$ are the b-hadrons in the event. In the case another size parameter is used for jet finding, the definitions in equations 1.1 and 1.2 change accordingly.

1.1.2 Track selection

It is important to select genuine tracks belonging to jets. Only tracks located within a cone of radius $\Delta R(j, \text{track}) \leq 0.4$ around the jet axis were considered. Cuts on ${p_T}^{\text{trk}} > 1.0$ GeV and the χ^2 of the track fit, $\chi^2/ndf < 3$, are applied. In addition, tracks are required to have a total of at least seven precision hits (pixel or micro-strip) in order to guarantee at least 3 z-measurements. Tracks are also required to fulfill cuts on the transverse and longitudinal impact parameters at the perigee to ensure that they arise from the primary vertex. As cutting on impact parameter (IP) significance might be detrimental for b-jets, where large IP values are expected, relaxed cuts were used, $|IP_{xy}| < 2$ mm, and $|IP_z \sin\theta| < 2$ mm, with θ being the polar angle measured with respect to the beam axis. The track quality cuts are summarized in table 1.1.

Track parameter	Selection
p_T	> 1 GeV
d_0^{PV}	< 2 mm
$z_0^{PV} \sin\theta$	< 2 mm
$\chi^2/ndof$	< 3
Number of Pixel hits	≥ 2
Number of SCT hits	≥ 4
Number of Pixel+SCT hits	≥ 7

Table 1.1: Track selection criteria used for double b-hadron jet tagging, where d_0^{PV} and z_0^{PV} denote the transverse and longitudinal impact parameters derived with respect to the primary vertex. The $chi^2/ndof$ is that of the track fit.

1.2 Kinematic differences between single and double b-hadron jets

The differences between genuine b-quark jets and double b-hadron jets, that in QCD originate mainly from gluon splitting, are expected to arise from the two-subjet structure of merged jets. In this section we present the study of a set of jet shape and substructure variables for the discrimination between single and merged b-jets. These variables are built from jet constituents either at calorimeter level (topological clusters) or tracks associated to the jet.

I. Jet track multiplicity

The jet track multiplicity is a variable simple to calculate that carries important information of the jet inner structure. It is defined as the number of tracks with p_T above 1 GeV, satisfying the quality cuts described in section 1.1.2, and contained within a cone of radius $R = 0.4$ around the jet axis. Figure 1.1 shows its distribution for two p_T bins, representative of the range covered in this study. It is observed that merged b-jets contain on average around two more tracks than single b-jets at low jet p_T, with a larger difference at higher p_T values. The jet track multiplicity corresponds to tracks with p_T above 1 GeV, satisfying the quality cuts described in section 1.1.1. The effect of the minimum track p_T requirement was examined by lowering the selection cut to $p_T > 0.5$ GeV. On the one hand this could lead to an improvement in discrimination if it captured more information about the fragmentation process; on the other hand, a lower minimum track p_T can make the method more sensitive to pile-up with the addition of soft tracks incorrectly associated to the jets. It was observed that reducing the p_T cut

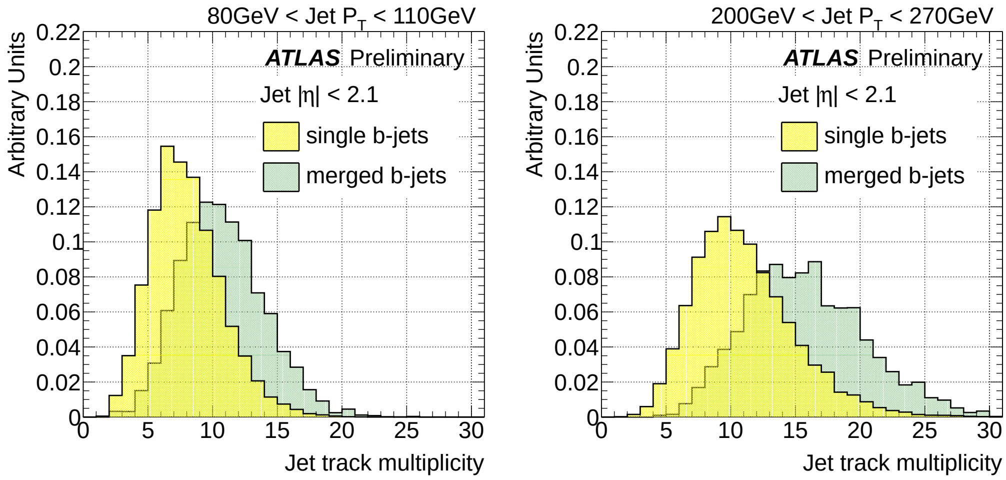

Figure 1.1: Distribution of the track multiplicity in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

of the tracks degrades the discrimination because it widens the distributions without increasing the separation between single and merged jets. No improvement was obtained either by circumscribing the selection to tracks with large impact parameter significance. The potential sensitivity achieved by enriching the sample in tracks associated to the b-hadron is counterbalanced by the lower number of tracks associated.

II. Jet width

The jet width is part of a set of continuous variables that try to distinguish individual particles/subjets within the jet by means of smooth funcion of $(\delta\eta, \delta\phi)$ away from the jet axis, in order to form combinations like geometric moments. This particular combination is a linear moment which sums the distances between the jet constituents and its axes, weighted by the constituent p_T, and then normalized to the total p_T of the jet. Its definition is,

$$Jet\ width = \frac{\sum_{i=1}^{N} {p_T}^{const_i}\, \Delta R(const_i, jet)}{\sum_{i=1}^{N} {p_T}^{const_i}} \tag{1.3}$$

where N is the total number of calorimeter or track constituents. This observable is also highly correlated to the mass of the jet.

Figure 1.2 shows the distributions for the track-jet width for which the sum in equation 1.3 runs over the N tracks associated to the jet, using the same criteria as for the jet track multiplicity. As expected, merged b-jets are wider than single b-jets.

The jet width can also be measured in terms of calorimeter variables, replacing tracks by topological clusters in the sum. Although it offers good separation, this variable is more sensitive to the amount of pile-up in the event than its track-based counterpart. This is illustrated in Fig. **??**, which shows the distribution of calorimeter width and track-jet width for single b-jets in events with low and high number of primary vertices (NPV) in a low p_T region where the effect of pile-up is more important.

In general, all the studied calorimeter-based jet variables show similar dependences with NPV. For this reason the track-based versions are preferred as more robust discriminators.

III. Maximum ΔR between track pairs

Figure 1.4 shows the distribution of the maximum ΔR between track pairs in the jets (Max$\{\Delta R(trk, trk)\}$). Merged b-jets show significantly higher values for this variable over a broad range of jet p_T. The distinct characteristic of this variable is that the separation between single b-jets and merged does not depend on jet p_T. In spite of its good discrimination power, we have looked

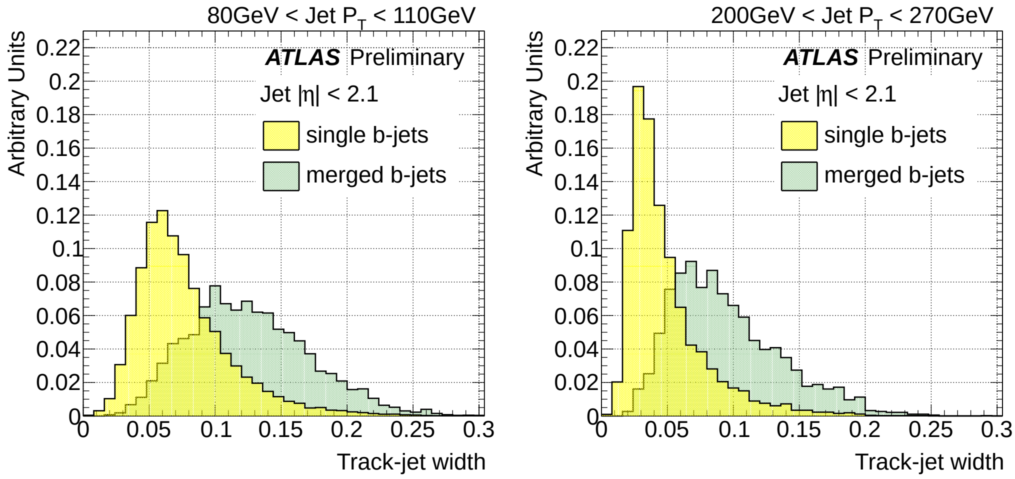

Figure 1.2: Distribution of track-jet width in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

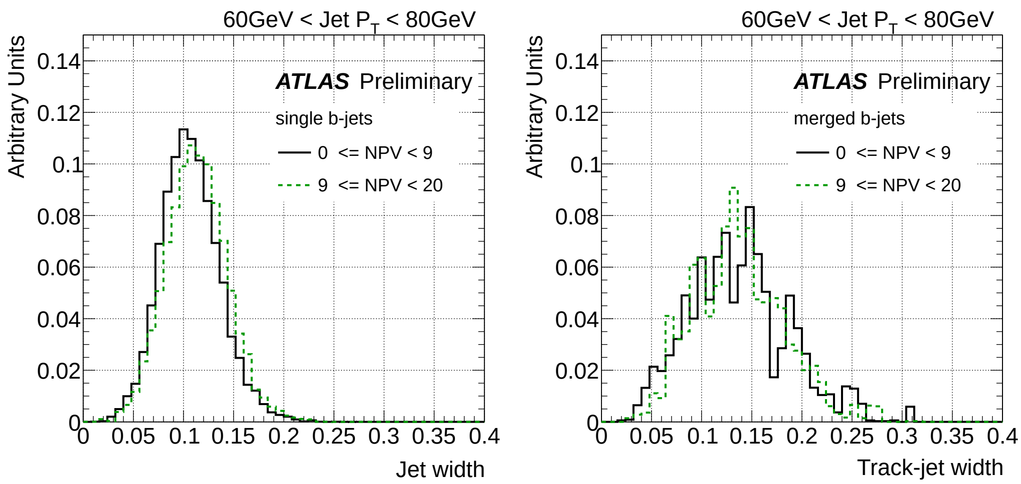

Figure 1.3: Distribution of jet width using topological clusters (left) and tracks (right) for single b-jets in two bins of number of primary vertices (NPV) for jets between 60 GeV to 80 GeV.

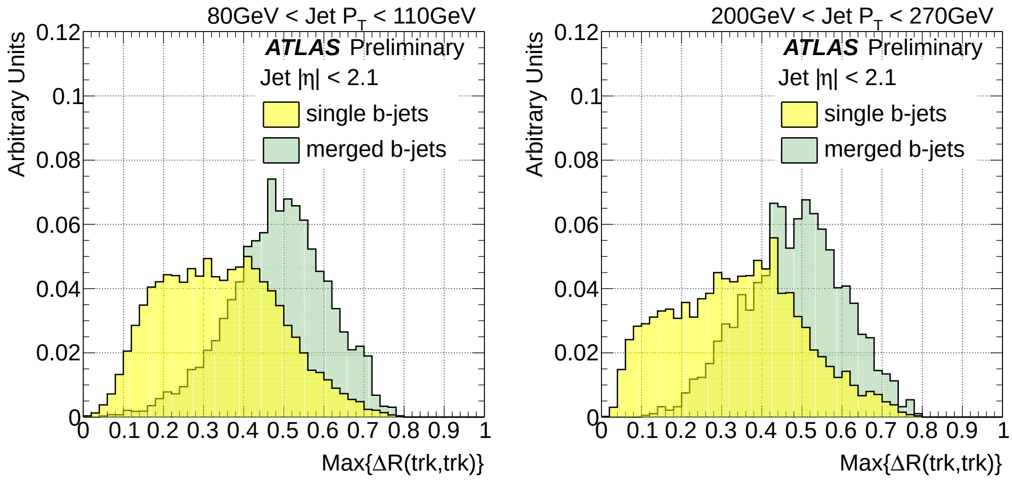

Figure 1.4: Distribution of the maximum ΔR between pairs of tracks in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

for alternatives to $\text{Max}\{\Delta R(trk, trk)\}$ as it is not an infrared safe observable and is sensitive to soft tracks originating from pile-up.

IV. ΔR between the axes of two k_t subjets

The distribution of the ΔR between the axes of the two exclusive k_t subjets in the jet is shown in Fig. 1.5 for single and merged b-jets. In order to build this variable the k_t algorithm [6] is applied to all the tracks associated to the jet using a large k_t distance parameter to ensure that all of them get clustered. The clustering is stopped once it reaches exactly two jets. We observe that this variable also provides good separation, with the advantage of infrared safeness and insensitivity to pile-up.

V. N-subjettiness variables

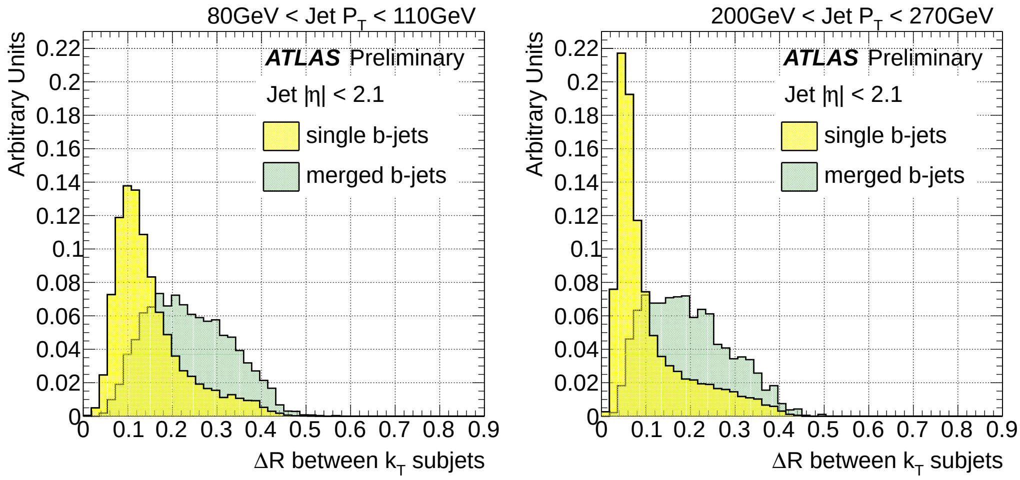

Figure 1.5: Distribution of the ΔR between the axes of the two k_t subjets in the jet for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

As mentioned above, the N-subjettiness [7] is a jet shape that describes the energy flow within a jet. It quantifies the degree to which radiation is aligned along specified subjet axes. This jet shape was adapted from the event shape N-jettiness [8].

Given candidate subjets directions determined by an external algorithm such as the exclusive k_t procedure [9], the variables is defined as,

$$\tau_N^{(\beta)} = \frac{1}{\sum_k p_{Tk}\,(R_0)^\beta} \sum_k p_{Tk} (\min\{\Delta R_{j1,k},\ \Delta R_{j2,k}, ...,\ \Delta R_{jN,k}\})^\beta \qquad (1.4)$$

The sum runs over the k constituent particles in a given jet where $p_{T,k}$ are their transverse momenta, and $\Delta R_{j1,k}$ is the distance between the candidate subjet $j1$ and a constituent particle k. R_0 is the characteristic jet radius used in the original jet clustering algorithm. The exponential weight, β, can optionally be applied to the angular distance computed between the subjets and the jet constituents.

This jet shape was designed to identify boosted N-prong hadronic decays. With $\beta = 1$, the definition above indicates that jets with $\tau_N \approx 0$ have all their radiation aligned with the candidate subjet directions and therefore have N (or fewer) subjets. Jets with $\tau_N \gg 0$ have a large fraction of their energy distributed away from the candidate subjet direction and therefore have at least $N + 1$ subjets.

To separate boosted hadronic objects from the QCD jet background, one could use the complete set of τ_N (with different values of β) in a multivariate analysis. However, [7] showed that a simple cut on the ratio τ_N/τ_{N-1} provides excellent discrimination power for N-prong hadronic objects. In particular, τ_2/τ_1 can identify boosted W/Z and Higgs bosons, with the angular weighting exponent $\beta = 1$ providing the best discrimination.

Since eq. 1.4 is linear in each of the constituent particle momenta, this variable is an infrared- and colliner-safe observable. In subsequent work [10], Thaler and van Tilburg showed that the initial step of choosing candidate subjet axes is in fact unnecessary. In particular, the quantity in equation 1.4 can be minimised over the candidate subjet directions, further improving boosted object discrimination.

The definition of N-subjettiness is not unique, and different choises can be sued to give different weights to the emissions within a jet. There generalizations of N-subjettiness are similar to different “angularities” [11] used in $e^+e^- \rightarrow$hadrons measurements.

N-subjettiness variables, as described in Ref. [7], were originally designed to identify boosted objects, like electroweak bosons and top quarks, decaying into collimated shower of hadrons which a standard jet algorithm would reconstruct as single jets. It is defined as:

$$\tau_N = \frac{1}{\sum_k p_{Tk}\, R_0} \sum_k p_{Tk} \min\{\Delta R_{S_1,k},\, \Delta R_{S_2,k}, ..., \, \Delta R_{S_N,k}\} \qquad (1.5)$$

where R_0 is the jet radius used in the jet clustering algorithm and the sum runs over the constituents of the jet. To avoid dependence on pile-up we consider the track-based n-subjettiness, where the sum is over the tracks in the b-tagged jet. $\Delta R_{S_j,k}$ is the distance in the rapidity-azimuth plane between the axis of subjet j and constituent track k. This jet shape variable quantifies to what degree a jet can be regarded as composed of N subjets. For instance, a jet with a two pronged structure, with all tracks clustered along two directions, is expected to have a smaller τ_2 value than a jet with tracks uniformly distributed in $\eta - \phi$ space.

Plots of τ_2 are shown in Fig. 1.6. In spite of its expected 2-prong substructure, merged b-jets have higher values of τ_2 than single b-jets. The explanation of this behavior can be found in Fig. 1.7, where its correlation with track-jet width ($\sim \tau_1$) is shown for single and merged b-jets. The two variables are highly correlated and for this reason wider jets have a larger τ_2. This suggests to switch from an absolute to a width-normalized τ_2. Fig. 1.8 thus shows the distributions of τ_2/τ_1. This ratio is often used but, although as expected somewhat larger values are obtained for single than for merged b-jets, specially at high p_T, we decided not to use this variable as it offers only marginal discrimination.

VI. Jet Mass

The jet mass, like the linear radial moment, also depends on the radiation pattern of the event. It is the most basic observable for disinguishing massive

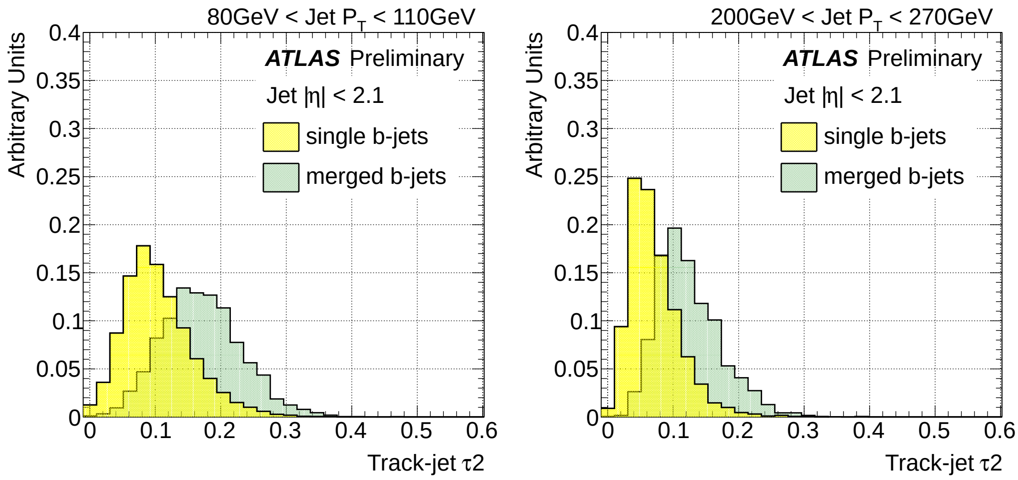

Figure 1.6: Distribution of τ_2 in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

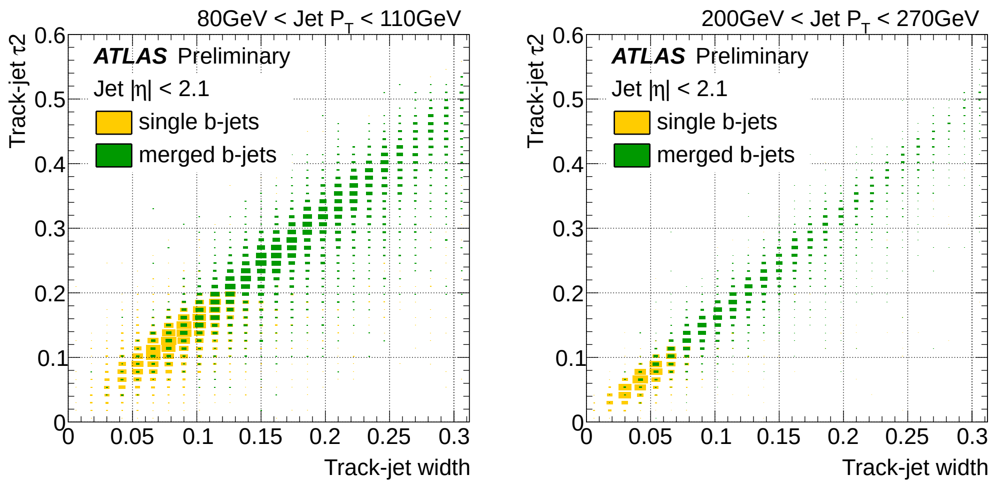

Figure 1.7: Correlation between τ_2 and track-jet width for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

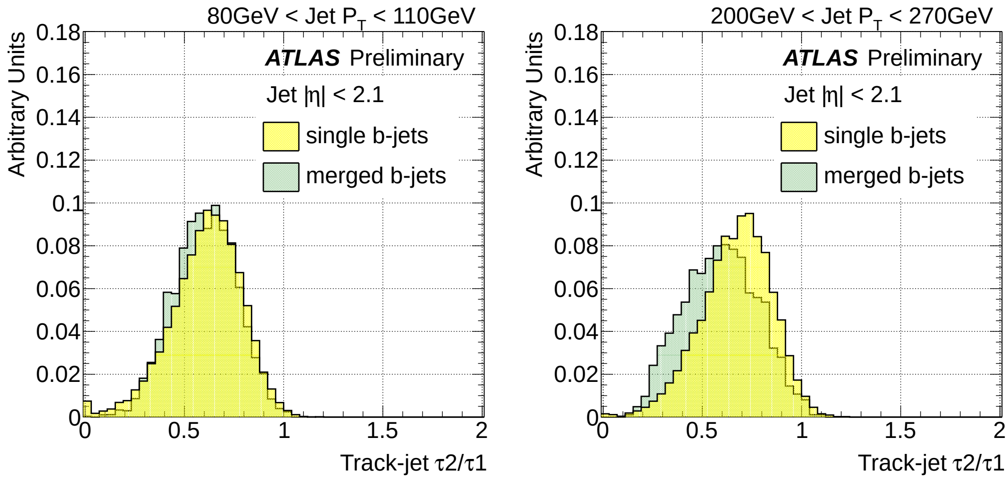

Figure 1.8: Distribution of τ_2/τ_1 in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

boosted objects from jets originating from quarks or gluons. The latter are expected to be dominated by wide-angle emissions, with increase probability to see high mass jets initiated from gluons as opposed to quarks [12].

Figure 1.9 shows the distribution of the jet mass for single and merged b-jets.

VII. Number of k_t subjets

With the development of the k_t algorithm, subjets were first used in the description of the hadronic final state in e^+e^- annihilation, such as the study of the jet multiplicity at different energy scales [13]. By using the sequential recombination algorithms introduced in the previus section, it is straightforward to define a "subjet algorithm" in which the structure of the jet's constituents is resolved using either the same jet finder algorithm or a

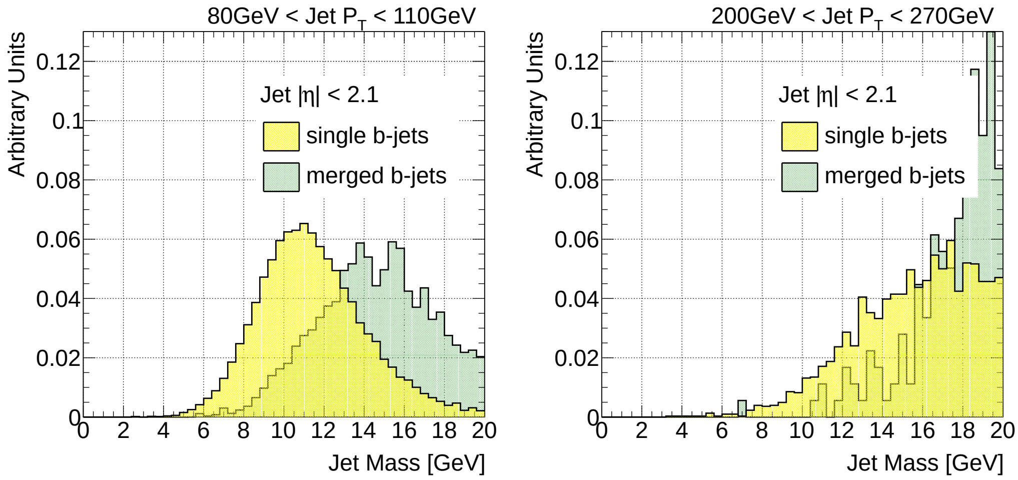

Figure 1.9: Distribution of jet mass in GeV for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

new one with a fixed (smaller) distance parameter.

The subjet multiplicity – the number of subjets within a jet – provides information on the distribution of energy and multiplicity of particles within a jet. For instance, in [14] the result of meassuring this "radiation variable" on quark- and gluon-initiated jets indicates that gluon-initiated jets tend to have on average higher subjet multiplicity. This result is consistent with the QCD prediction that gluons radiate more than quarks. In the case of this and different other analyses the k_t algorithm is rerun for subjet finding.

As an alternative to fixed distance parameter subjets, it is also possible to undo the last step in the recombination sequence [15] in order to identify the decay products of an object. This approach is used in seveal jet grooming procedures[1], see for instance [17].

It is also possible to extend the use of individual subjets in conjunction

[1]Jet grooming comprises dedicated techniques to remove uncorrelated radiation within a jet. A review of these procedures can be found in [16].

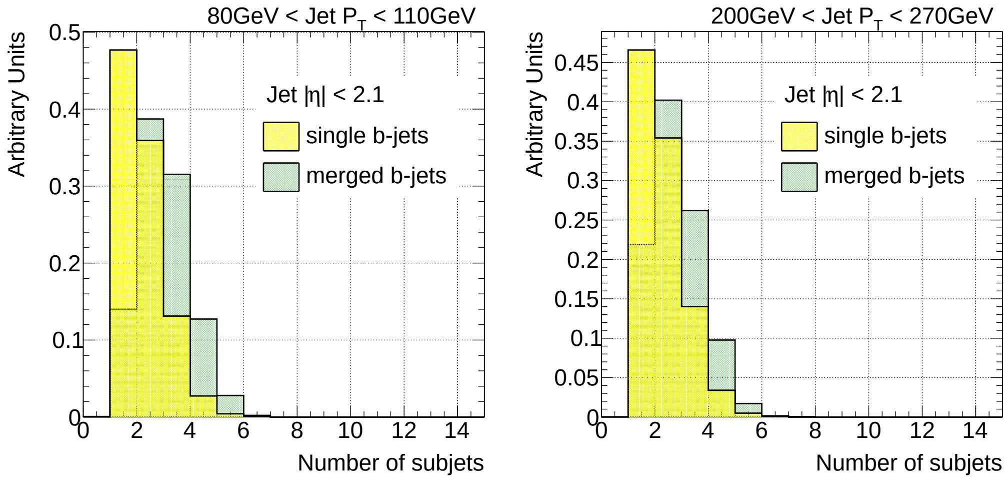

Figure 1.10: Distribution of the number of k_t sub-track-jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

with more traditional jet shape variables. Using these tools, an inclusive jet shape based on the substructure topology of a single jet, "N-subjettiness" [7] is defined.

Figure 1.10 shows the distribution of the number of sub-track-jets single and merged b-jets.

VIII. ΔR between leading constituents

An alternative approach to measuring the width is to use the angular separation of the two hardest constituents inside jets. This has the advantage of effectively removing any dependence on the shower development within the calorimeter and focuses on the hard component of the jet. Figure 1.11 shows the distribution of the ΔR between leading tracks in the jet for single and merged b-jets.

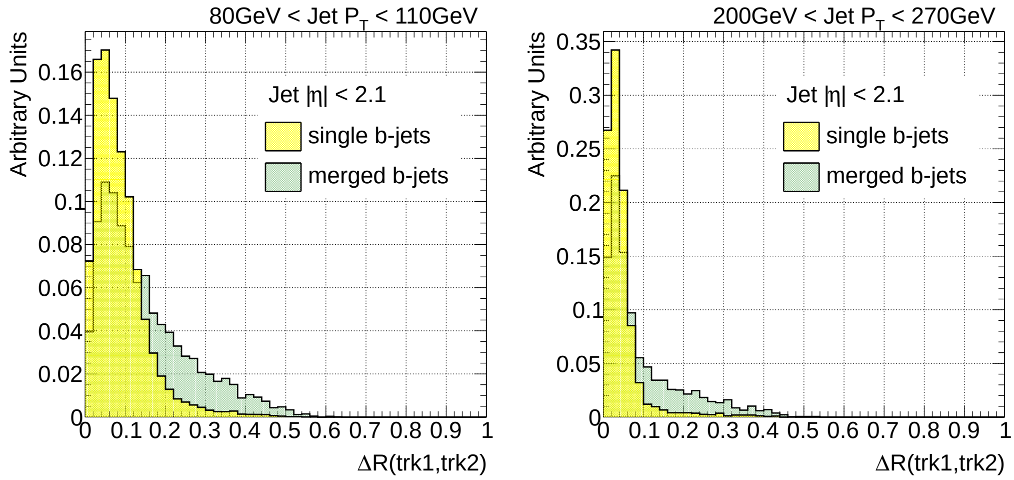

Figure 1.11: Distribution of ΔR between leading tracks for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

IX. Jet eccentricity

In defining a jet moment there are several ways to weight the momentum and define the center of the jet. We have defined the jet width as the first moment of the transverse energy with respect to the jet axis; another example of useful combination is the jet pull [18]. But it is also natural to look at higher moments, such as those contained in the covariance tensor,

$$C = \sum_{i \in jet} \frac{p_T^i |r_i|}{p_T^{jet}} \begin{pmatrix} \Delta y_i^2 & \Delta y_i \Delta \phi_i \\ \Delta \phi_{ii} & \Delta \phi_i^2 \end{pmatrix}.$$

Here, $r_i = (\Delta y_i, \Delta \phi_i) = c_i - J$, where $J = (y_J, \phi_J)$ is the location of the jet and c_i is the position of a cell or particle with transverse momentum p_T^i. The eigenvalues $a \geq b$ of this tensor are similar to the semimajor and semiminor axes of an elliptical jet. The jet eccentricity, defined below, is a combination of these eigenvalues, and it is a measure of how elongated is the

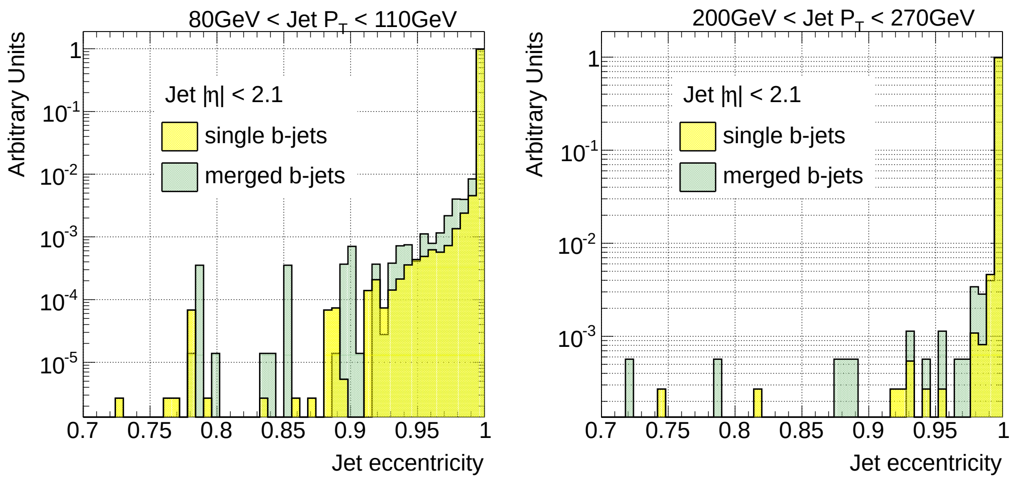

Figure 1.12: Distribution of the jet eccentricity for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

area of a jet.

$$e = \sqrt{\frac{(a^2 - b^2)}{a}} \tag{1.6}$$

Figure 1.12 shows the distribution of the jet track-eccentricity for single and merged b-jets.

We also explored the potential improvement of constructing kinematic variables with only displaced tracks, as these are the ones expected to arise from the decay of B-hadrons. Cuts of 2, 2.5 and 3 on the track transverse impact parameter significance were investigated leading however to no gain in discrimation power.

In Figures 1.13 and 1.14 two examples are shown.

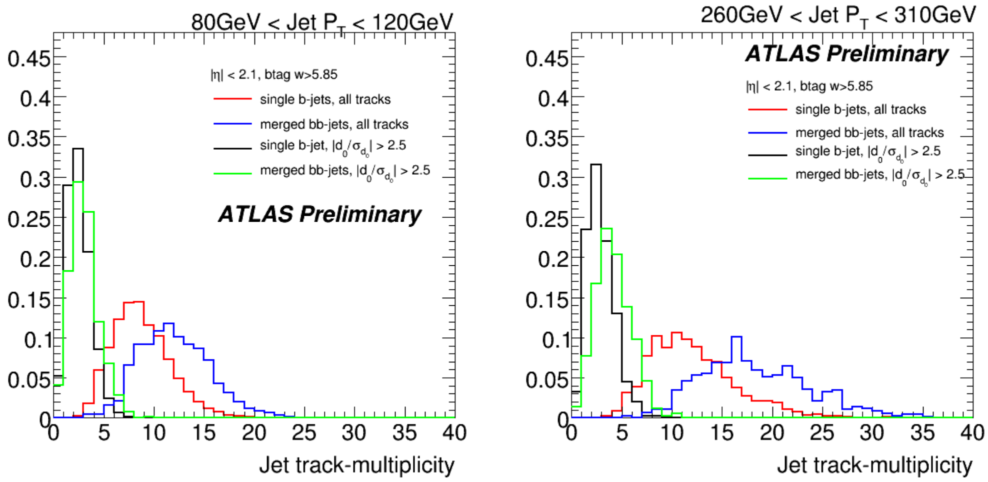

Figure 1.13: Distribution of the jet track multiplicity single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right), for all and displaced tracks only.

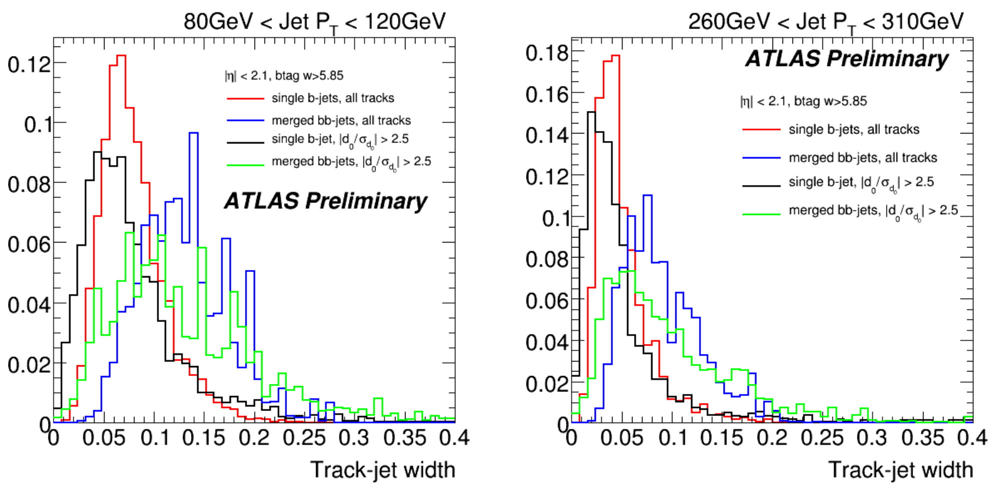

Figure 1.14: Distribution of the track-jet width for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right), for all and displaced tracks only.

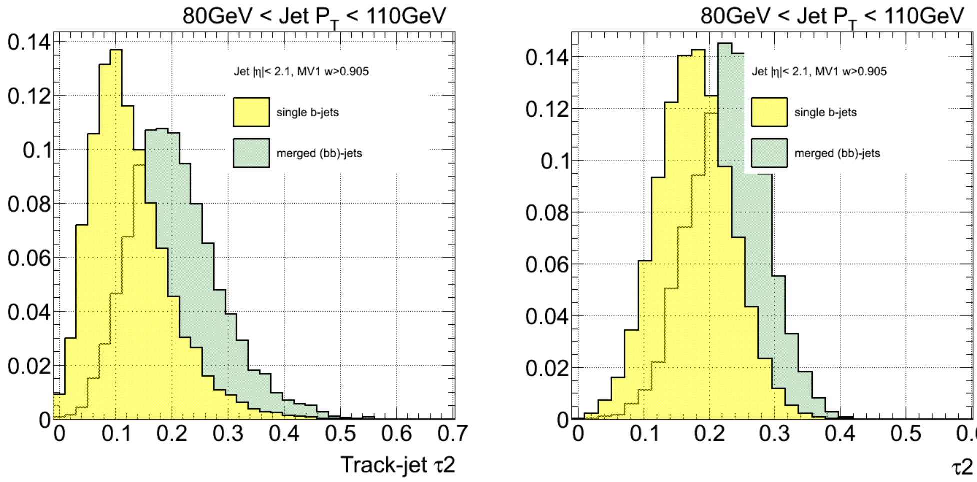

Figure 1.15: Distribution of τ_2 for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

1.2.1 Further studies using "ghost-association" and bigger cone jets

In order to better understand the behavior observed for τ_2, ΔR between the axes of k_T subjets and jet eccentricity in anti-k_T 0.4 jets, these variables were studied for other two different scenarios,

- using the active area of jets (with clusters used as input to jet reconstruction).
- using bigger 0.6 anti-k_T jets

in order to enhance the efficiency to capture the decay products in gluon to $b\bar{b}$-jets.

Figures 1.15 to 1.17 show distributions of variables mentioned above for single and merged b-jets between 80 GeV to 110 GeV.

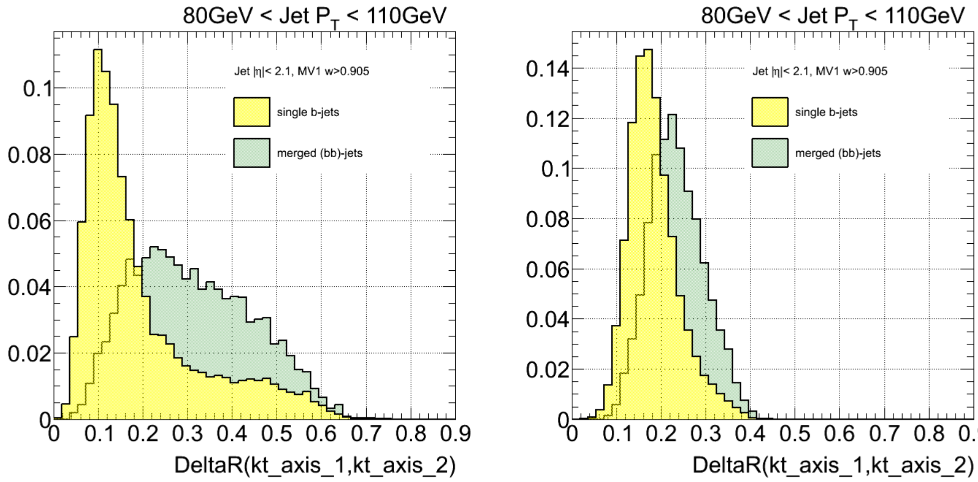

Figure 1.16: Distribution of ΔR between k_T subjets for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

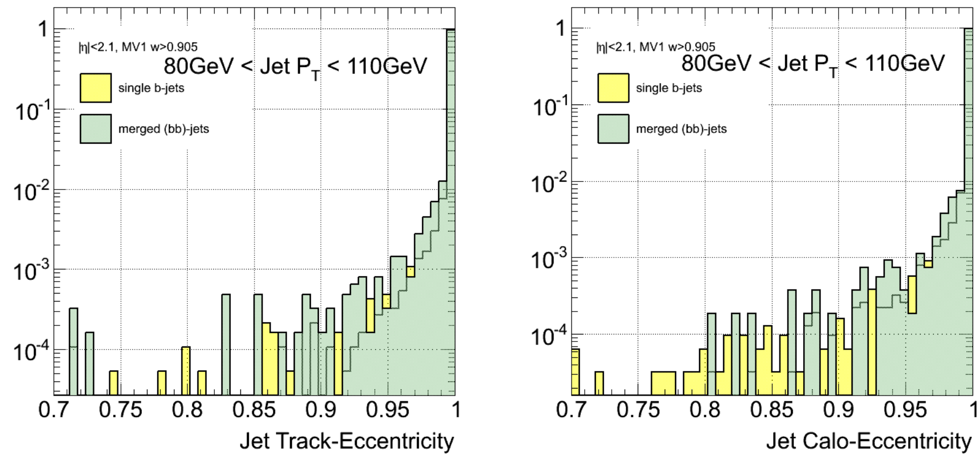

Figure 1.17: Distribution of the jet eccentricity for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

1.3 Validation of the jet variables in data

In order to study the extent to which the simulation reproduces the distributions observed in data for the different variables explored a set of comparison plots is presented. Fig. 1.18 shows the distributions of jet track multiplicity, track-jet width and ΔR between the axes of the two k_t subjets, in two different jet p_T bins in dijet Monte Carlo and data events collected by ATLAS during 2011. The distributions are normalized to unit area to allow for shape comparisons. There is a good agreement between data and simulation. It should be remarked that the observed agreement is actually not a direct validation of the description in the MC of the relevant variables, but its convolution with the simulated relative fractions of light-, c-, b- and bb-jets in the b-tagged generated jet sample. To some extent, some level of compensation can take place between these two effects.

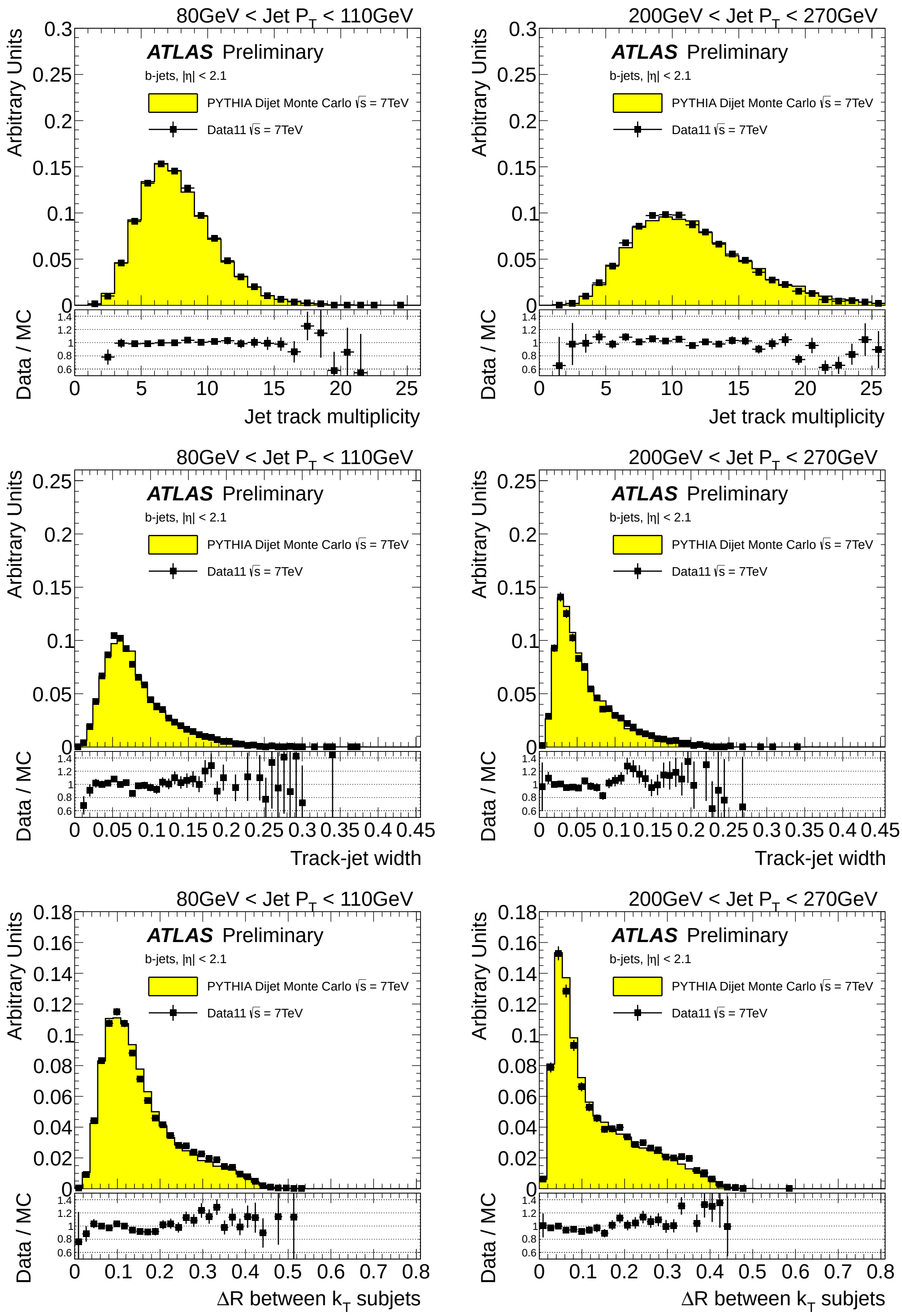

Figure 1.18: Distribution of three tracking variables in 2 different jet p_T bins, for experimental data collected by ATLAS during 2011 (solid black points), and simulated data (filled histograms). The ratio data over simulation is shown at the bottom of each plot.

Bibliography

[1] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[2] Atlas tunes of pythia 6 and pythia 8 for mc11. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.

[3] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[4] W Lampl et al. Calorimeter Clustering Algorithms: Description and Performance. (ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003), Apr 2008.

[5] ATLAS Collaboration. Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data. *ATLAS-CONF-2012-020*, 2012.

[6] S. Catani, Y.L. Dokshitzer, H. Seymour, and B.R. Webber. Longitudinally invariant K(t) clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, B406:187, 1993.

[7] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 1103:015:026, 2011.

[8] Iain W. Stewart, Frank J. Tackmann, and Wouter J. Waalewijn. *n* jettiness: An inclusive event shape to veto jets. *Phys. Rev. Lett.*, 105:092002, Aug 2010.

[9] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, Oct 1993.

[10] Jesse Thaler and Ken Van Tilburg. Maximizing boosted top identification by minimizing n-subjettiness. *Journal of High Energy Physics*, 2012:1–33, 2012. 10.1007/JHEP02(2012)093.

[11] Carola F. Berger, Tibor Kúcs, and George Sterman. Event shape–energy flow correlations. *Phys. Rev. D*, 68:014012, Jul 2003.

[12] Leandro G. Almeida, Seung J. Lee, Gilad Perez, Ilmo Sung, and Joseph Virzi. Top quark jets at the lhc. *Phys. Rev. D*, 79:074012, Apr 2009.

[13] S. Catani, Yu.L. Dokshitzer, F. Fiorani, and B.R. Webber. Average number of jets in e+eâĹŠ annihilation. *Nuclear Physics B*, 377(3):445 – 460, 1992.

[14] R. Snihur. Subjet multiplicity in quark and gluon jets at d0. *Nuclear Physics B - Proceedings Supplements*, 79(1âĂŞ3):494 – 496, 1999. <ce:title>Proceedings of the 7th International Workshop on Deep Inelastic Scattering and QCD</ce:title>.

[15] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.

[16] A. Abdesselam, E. Bergeaas Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, et al. Boosted objects: A Probe of beyond the Standard Model physics. *Eur.Phys.J.*, C71:1661, 2011.

[17] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Techniques for improved heavy particle searches with jet substructure. *Phys. Rev. D*, 80:051501, Sep 2009.

[18] Jason Gallicchio and Matthew D. Schwartz. Seeing in color: Jet superstructure. *Phys. Rev. Lett.*, 105:022001, Jul 2010.