To the bottom of the stop: calibration of bottom-quark jets identification algorithms and search for scalar top-quarks and dark matter with the Run I ATLAS data

Pani, P.

Citation for published version (APA):
Pani, P. (2014). To the bottom of the stop: calibration of bottom-quark jets identification algorithms and search for scalar top-quarks and dark matter with the Run I ATLAS data
Analysis of $b$-jets properties in $5 \text{fb}^{-1}$ of data at $\sqrt{s} = 7$ TeV

$B$-jets properties and the performance of the tagging algorithms are studied in data and compared to the Monte Carlo simulation. The majority of the observables are found to be well described by the simulation, in particular the modelling of the hadronization track properties and their angular distribution with respect to the jet. A few variables have been found to be sensitive to contamination and not included in the comparison. Finally the $b$-tagging performance has been found to be in agreement to the Monte Carlo at the few percent level.

In this chapter, a number of properties of $b$-jets will be evaluated in the data sample collected with the ATLAS detector at $\sqrt{s} = 7$ TeV during the year 2011 and compared to the Monte Carlo simulation. In each of the Figures, the bottom panel indicates the ratio between the two distributions shown in the plot. For sake of clarity the investigated variables are divided into three categories: detector specific quantities, observables sensitive to the hadronization physics, and tagging algorithm performance. All distributions are reweighted according to the procedure in Section 4.4 and the background contamination has been subtracted according to the sideband subtraction method detailed in Chapter 5.
6.1 Evaluation of the detector reconstruction simulation

The detector specific variables describe a class of observables that are strongly influenced by the detector reconstruction. The accuracy of the description in the Monte Carlo simulation is evaluated in the following.

The number of innermost Pixel layer (the $B$-Layer) and other Pixel detector hits for hadronization and $B$ decay products are compared in Figure 6.1. They are of utmost importance for the tagger performance since they determine both the primary and the secondary vertices reconstruction efficiency and resolution. The distribution of the number of $B$-Layer hits and number of total Pixel detector hits is well reproduced by the Monte Carlo, with a small shift toward higher multiplicity values of approximately 10%. The number of hits associated to the SCT and TRT sub-detectors for the hadronization and the $B^+$ decay products have also been studied and show a similar behaviour, but with a smaller discrepancy of approximately 5%.

The second comparison for the detector simulation concerns the impact parameters of the hadronization tracks of the jet in the transverse plane ($d_0$) and along the beam axis ($z_0$), calculated with respect to the primary vertex of the interaction. They are shown in Figure 6.2 and they are characterised by a wider distribution in data compared to the Monte Carlo simulation, with a 20% discrepancy in the tails of the distribution. This underestimation of the spread of the distribution in Monte Carlo was confirmed by orthogonal studies [8] and it is likely to have an impact on the different performance of the $b$-taggers in data and Monte Carlo, since they exploit or rely on this information for the $b$-jets identification. It would be realistic to assume that this mismodelling contributes to the deviation from one of the $b$-tagging calibration scale factors. The distributions of the uncertainty associated to the impact parameters and the impact parameter distributions of the $B^+$ decay products have been also investigated and in this case they are found to be in very good agreement between data and Monte Carlo.
Analysis of $b$-jets properties  
Evaluation of the detector reconstruction simulation

Figure 6.1: The multiplicity distributions of $B$-Layer and other Pixel detector hits are presented. Tracks used to evaluate the multiplicity distributions are identified as hadronization products, which means that they have been associated to the jet and are not identified as $B^\pm$ decay products.

Figure 6.2: Transverse and longitudinal impact parameters of the hadronization tracks calculated with respect to the primary vertex of the interaction.
6.2 Comparison of hadronization physics observables

The reliability of the description of the hadronization process in simulated events is verified by means of the second group of quantities. These so called physics variables are connected directly with the physics involved in the simulation, such as the spatial distribution of hadronization products, the track multiplicity and the structure of the $b$-jet as well as the topological distribution of the $B$ decay products and their orientation with respect to the jet axis. A number of these variables were proven in Chapter 5 to have a dependence on the $B$ candidate invariant mass, such that the background subtraction procedure would introduce a bias. Among these are quantities like the probability of the $B$ decay products to be associated to the jet by angular matching and the angular separation ($\Delta R$) between the kaon track and either the jet axis or the $B$ direction of flight. For this reason, the angular distributions of tracks with respect to the jet presented in this section have been evaluated only on muon and hadronization tracks. In addition, all distributions aiming to evaluate the $b$-tagging algorithms performance in the following section were obtained by requiring that all the three decay products of the $B^\pm$ were associated to the jet, in order to remove possible disagreements raising from events in which not all the $B$ decay products were among the sample of tracks used by the tagging algorithms.

Figure 6.3: Fraction of momentum carried by the $B^\pm$ with respect to the jet.
6.2.1 B hadron and jet correlations

The first comparison that is presented aims to investigate the correlations between the B hadron and the associated jet. Figure 6.3 shows the fraction of momentum carried by the B* with respect to the jet, an observable that is correlated with the B fragmentation function of the jet showering process. The bulk of the distribution is well described by the simulation. However, a larger tail in data can be observed for events in which the B hadron momentum is more than one and a half times larger than the one of the associated jet, which hints to a deficiency of the simulation in modelling soft hadronization processes. The discrepancy seems to be more prominent, although with lower statistical significance, in the Barrel region and at low hadron momentum.

In order to complete the overview on the correlations between the B hadron and its associated jet, the distribution of the angular distance (ΔR) of the B hadron to the jet axis is shown in Figure 6.4. It shows a good agreement with data that reflects the quality of the Monte Carlo generator tuning and the good description the b-quark and its hadronization process.
6.2.2 Jet topology

The general topology of the $b$-jet is studied through the multiplicity of the tracks in the jet and the angular distance between these tracks and the jet axis.

A first evaluation of the reliability of the Monte Carlo simulation is obtained by studying the multiplicity of the hadronization tracks of the jet. A different density of tracks in the jet environment might highlight a simulation inaccuracy that causes different performance of the $b$-taggers in data and in Monte Carlo. The plots in Figure 6.5 show a good agreement of the hadronization track multiplicity and its $p_T$ profile, with a very small tendency of the simulation to underestimate the multiplicity.

As the jet-track association is mostly based on angular matching, the $\Delta R$ of the hadronization tracks with respect to the jet axis has also been considered. As shown in Figure 6.6(a) it is extremely well described by the simulation. Furthermore, the distribution of the maximum of this $\Delta R$ for each jet, representing the most "far-away" hadronization product, is presented in Figure 6.6(b). Also this distribution is well modelled by the simulation, including the bin at zero that represent those events in which no other charged particles in addition to the $B^\pm$ decay products have been associated to the matched jet.

The muon angular distance with respect to the $B$ direction of flight is shown in Figure 6.7. There is a reasonable agreement, up to $\Delta R = 0.4$, between the data and the simulation. Above $\Delta R = 0.4$, the hypothesis of the background subtraction was proven not to hold (see Chapter 5) and a bias from the procedure was expected. However, the $\Delta R$ between the muon tracks and the jet axis, for which no bias was expected from the background subtraction, shows a consistent disagreement between the data and the simulation. It was not possible to unambiguously determine whether this effect is a real mismodelling of the simulation or it is caused by a not understood correlation effect between the $B$ invariant mass and the $\Delta R$ between the muon tracks and the jet axis.

In summary it is safe to state that the simulation of the hadronization physics, once disentangled from the rest of the event (background contamination effects, $B$ decay, etc.), is in very good agreement with the experimental data.
Figure 6.5: Number of hadronization tracks in the $b$-jet. The hadronization tracks are identified as all the tracks associated to the jet except the $B^\pm$ decay products. The multiplicity is also shown as a function of the jet $p_T$. 

(a) multiplicity distribution

(b) $p_T$ profile
Comparison of hadronization physics observables

Figure 6.6: Angular distance between the hadronization tracks of the jet and the jet axis and distribution of the maximum value per each jet.
### Figure 6.7: Angular distance between the muon tracks from the B decay and the B hadron direction of flight or the distance between the muon tracks and the jet axis.
6.3 Intermezzo: detector effects on hadronization physics

The tracks used by the $b$-taggers have to fulfil stricter quality requirements than the tracks investigated in the previous sections. These requirements have been introduced in Chapter 4.3, Table 4.1 and involve higher cuts on the number of hits on tracks in the Inner Detector and a minimum track transverse momentum of 1 GeV. Consequently, once the good description of the detector is verified, it is interesting to investigate the ability of the Monte Carlo simulation to describe the jet physics if only these "high quality" (Btag-Quality) tracks are considered. The number of inner detector hits and the impact parameters have been studied for this additional track category. As before, these show a good agreement in the shapes with the same tendency of the Monte Carlo simulation of overestimating the number of hits on a track.

Secondly, the angular distribution of the hadronization tracks that fulfil Btag-Quality requirements with respect to the jet axis was investigated. Both the distribution of $\Delta R$ between each track and the jet axis and the maximum $\Delta R$ for each jet are presented in Figures 6.8(a) and 6.8(b) and found to be in excellent agreement with the data.

![Figure 6.8: Angular distance between the hadronization tracks of the jet and the jet axis and the maximum value for each jet for tracks of Btag-Quality.](image-url)
Analysis of $b$-jets properties  
Intermezzo: detector effects on hadronization physics

Figure 6.9: Distribution of the number of $B$ decay products associated to the matched jet that fulfil Btag-Quality requirement.

Figure 6.10: Distribution of the number of hadronization tracks in a jet that fulfil Btag-Quality requirement.
Subsequently, the distribution of the number of B decay products in the jet that fulfil the Btag-Quality criteria is presented in Figure 6.9 and it is also found to be well modelled.

Finally, the multiplicity of hadronization tracks that fulfil these Btag-Quality requirements is presented in Figure 6.10, with the purpose of probing the response to these selection criteria of the jet hadronization products. There is a good agreement between the data and the Monte Carlo simulation, with a small tendency from the latter in underestimating the reconstruction performance by 5%, estimating a lower number of Btag-Quality tracks with respect to the data. In order to further investigate this effect, an additional set of requirements has been defined, each obtained by removing one of the requirements that define the Btag-Quality tracks criteria. Each of these sets of requirements was separately investigated, in the attempt to identify which of the Btag-Quality requirement(s) is mainly responsible for the disagreement.

All distributions have shown a very good agreement between the Monte Carlo simulation and the data, which leads us to the conclusion that the 5% disagreement on the inclusive distribution is originating from the combination of many small effects (<1%) in each of the variables used to define the Btag-Quality requirements and not from a single mismodelled distribution.

### 6.4 Tagging algorithms performance studies

The last comparison aims to evaluate the simulation of the performance of the b-tagging algorithms. Given that the $B^\pm$ decay is completely reconstructed, it is possible to investigate the performance of the taggers under study evaluating their ability to correctly identify the tracks from the decay. In order to disentangle the jet association efficiency from the performance evaluation, only events in which the $B^\pm$ decay products have been associated to the matched jet are considered.

In such a clear topology as the $B^\pm$ decay we consider, it would prove particularly interesting to identify differences between data and Monte Carlo. This decay is characterised by little impact of track reconstruction efficiency effects, due to the fact that in all events it has already been possible to reconstruct the $B^\pm$. Therefore this study allows to disentangle this effect from any other possible source of disagreement. On the other hand, it is important to highlight that any conclusions about a good modelling of the performance in these events cannot be easily extrapolated to an inclusive b-jets sample.

First of all, Figures 6.11(a) and 6.11(b) show the number of tracks associated by
Figure 6.11: Number of tracks associated with the displaced vertices reconstructed by the SV1 (a) and JetFitter (b) algorithms. Zero associated tracks means that no displaced vertex was reconstructed. (c) Number of reconstructed vertices for JetFitter. (d) Number of tracks used by the IP3D tagger for the weight evaluation. Only events where all the B decay products have been associated to the jet are considered.
Figure 6.12: SV1 and JetFitter performance in associating the $B$ decay products to the displaced vertex (top and middle left) and frequency of kaons, muons and hadronization tracks association to the displaced vertex (vertices) (top and middle right). Cross check for evaluating whether the $B$ decay products are included in the set of tracks used by the IP3D tagging algorithm for the weight evaluation (bottom left and right).
the vertex finding taggers SV1 and JetFitter to the displaced vertex. Due to the B decay topology under investigation the expected average number of tracks is three. As visible in the figure, the Monte Carlo simulation describes the outliers of this distribution with an agreement better than 10%. In Figure 6.11(c) the number of vertices reconstructed by the JetFitter tagger is shown. Also in this case the average value is compatible with the decay topology and the outliers are described by the simulation with an accuracy between 10 and 20%. Finally, Figure 6.11(d) shows the total number of tracks used by the Impact Parameter (IP) tagger. This tagger performs the b-jet identification based on the impact parameter significance of all Btag-Quality tracks in the jet, so in this case, the expected average number of tracks is higher because the total number of tracks used in the evaluation is shown. The Monte Carlo simulation is characterised by a narrower distribution than the distribution in data, with a discrepancy up to 25% for the very particular case in which no tracks associated to the jet fulfil Btag-Quality requirements.

The performance of the b-taggers are evaluated by the efficiency with which they associate the B decay products to the reconstructed, displaced vertex (vertices) as shown in Figure 6.12. As a cross check, also the IP tagger is analysed (Figures 6.12(e)-6.12(f)), in order to verify whether the B decay products are selected among the tracks used for the weight evaluation. The same information is shown in the plots.
Conclusions Analysis of b-jets properties

in the right column of these two Figures, where the probability with which one of the muons, the kaon or an hadronization track are associated to the displaced vertex is shown. Again, In the case of the IP tagger, a higher number of hadronization tracks is expected.

Finally the tagging efficiency is studied as a function of the working points for the JetFitter tagger (Figure 6.13). Each bin of the distribution corresponds to a working point of the algorithm and the efficiency of each working point in selecting b-jets from an inclusive sample is also indicated. The efficiency measured in this sample is higher than the nominal one for all working points due to the clear signature of the B decay used to select the b-jets. The agreement between the experimental data and the Monte Carlo simulation is excellent (⩽ 0.5%), also expected for this clean B decay topology.

6.5 Conclusions

The study presented in this part of the thesis gives an overview of a number of observables that characterise the properties of b-jets and their modelling by the available Monte Carlo simulation.

Three different categories of quantities have been investigated: the ones related to detector reconstruction effects, the ones related to the hadronization physics of the b-jets and the performance of the b-taggers. Except for a limited number of cases for which no clear conclusion could be drawn, the Monte Carlo simulation showed a reliable modelling of all these quantities and of the response of the b-taggers to the b-jets selected by this analysis.

In the short term, this study provides confidence in the Monte Carlo tunings and modelling of the fragmentation function of B hadrons, of the hadronization process in the jet and of the jet universality assumption on which the modern Monte Carlo tuning methods rely. In fact, tunings 113 rely on information based on many different experiments and assume for example that particles fragment in the same way at a hadron collider as they did at the Large Electron-Positron (LEP) collider. Since the infrared environment in hadron collisions is characterised by a different (hadronic) initial-state vacuum by a larger final-state gluon component, and also by simply having a lot more colour flowing around in general, it is important to check to what precision this assumption holds explicitly, e.g., by measuring multiplicity and \( p_T \) spectra of identified particles, particle-particle correlations, and particle production ratios in situ at hadron colliders. The hadronization track investigation in this
Analysis of $b$-jets properties

Conclusions

analysis represents a first, preliminary attempt to validate the modelling of the behaviour of particles in the specific environment of $b$-jets and showed very positive results.

In the long term, it would be interesting to improve the analysis strategy in order to allow the investigation of the properties of a more inclusive sample of B decays, considering also $B^0$ and $B_s$ hadrons and considering charm decays as well. Exploiting a more representative sample of the majority of $b$-jets in ATLAS it will be possible to draw more impelling conclusions on the modelling of the taggers performance and improve even more their simulation.