

# Identification of double $b$ -hadron jets from gluon-splitting with the ATLAS Detector

María Laura González Silva

Doctoral Thesis in Physics

Physics Department

University of Buenos Aires

November 2012



**UNIVERSIDAD DE BUENOS AIRES**

Facultad de Ciencias Exactas y Naturales

Departamento de Física

**Identificación de jets con hadrones  $b$  producidos por  
desdoblamiento de gluones con el detector ATLAS.**

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 Piegai

Lugar de Trabajo: Departamento de Física

Buenos Aires, Noviembre 2012

## Agradecimientos

Quiero agradecer a mi director, Ricardo Piegaia, por darme la oportunidad de trabajar en el proyecto ATLAS, por su dedicación y su enseñanza constante; y a mis compañeros de grupo, Gastón Romeo, Gustavo Otero y Garzón, Hernán Reisin y Sabrina Sacerdoti por el trabajo compartido y por brindarme su amistad a lo largo de estos años. Quiero agradecer a Ariel Schwartzman por darnos este análisis, por su caudal inagotable de ideas y por su generosidad y la de todo su equipo. Agradezco al Laboratorio CERN, al Experimento ATLAS, a los programas HELEN y e-Planet, al CONICET y al Fundación Exactas por hacer posible la realización de esta tesis.

Quiero agradecer el apoyo de mis compañeros de la carrera, especialmente a mis amigos Cecilia y Tomás. Quiero agradecer también a mis compañeros de grupo y oficina, Lean, Yann, Javier, Pablo, y Orel por estar siempre dispuestos a darme una mano. Quiero agradecer a mis colegas y amigos de la Universidad de La Plata, Fernando, Martín y Xabier por todos los momentos compartidos; y a los amigos que hice a lo largo de estos años en mis visitas al Laboratorio CERN, Dodo, Laura, Lucile, Bárbara, Teresa, Manouk, Alex, Olivier y Haris, por ser mi familia en la distancia.

Agradezco profundamente a mis amigos y a toda mi familia por su apoyo y aliento; y de manera especial a mamá y a Juan, por comprenderme y acompañarme en todo. A ellos les dedico esta tesis.

# Identificación de jets con hadrones $b$ producidos por desdoblamiento de gluones con el detector ATLAS.

## Resumen

En esta tesis se presenta un estudio de la subestructura de jets que contienen hadrones  $b$  con el propósito de distinguir entre jets- $b$  genuinos, donde el quark  $b$  se origina a nivel de elemento de matriz (por ejemplo, en decaimientos de top, W, o Higgs) y jets- $b$  producidos en la lluvia partónica de QCD, por el desdoblamiento de un gluón en un quark y un antiquark  $b$  cercanos entre sí. La posibilidad de rechazar jets- $b$  producidos por gluones es importante para reducir el fondo de QCD en análisis de física dentro del Modelo Estándar, y en la búsqueda de canales de nueva física que involucran quarks  $b$  en el estado final. A tal efecto, se diseñó una técnica de separación que explota las diferencias cinemáticas y topológicas entre ambos tipos de jets- $b$ . Esta se basa en observables sensibles a la estructura interna de los jets, contruidos a partir de trazas asociadas a éstos y combinados en un análisis de multivariable. En eventos simulados, el algoritmo rechaza 95% (50%) de jets con dos hadrones  $b$  mientras que retiene el 50% (90%) de los jets- $b$  genuinos, aunque los valores exactos dependen de  $p_T$ , el momento transversal del jet. El método desarrollado se aplica para medir la fracción de jets con dos hadrones  $b$  en función del  $p_T$  del jet, con 4,7 fb<sup>-1</sup> de datos de colisiones  $pp$  a  $\sqrt{s} = 7$  TeV, recogidos por el experimento ATLAS en el Gran Colisionador de Hadrones en 2011.

*Palabras clave:* Experimento ATLAS, Jets, Subestructura de Jets, QCD, Producción de jets  $b$ , Etiquetado de Jets  $b$ .

# Identification of double $b$ -hadron jets from gluon-splitting with the ATLAS Detector.

## Abstract

This thesis presents a study of the substructure of jets containing  $b$ -hadrons with the purpose of distinguishing between “single”  $b$ -jets, where the  $b$ -quark originates at the matrix-element level of a physical process (e.g. top,  $W$  or Higgs decay) and “merged”  $b$ -jets, produced in the parton shower QCD splitting of a gluon into a collimated  $b$  quark-antiquark pair. The ability to reject  $b$ -jets from gluon splitting is important to reduce the QCD background in Standard Model analyses and in new physics searches that rely on  $b$ -quarks in the final state. A separation technique has been designed that exploits the kinematic and topological differences between both kinds of  $b$ -jets using track-based jet shape and jet substructure variables combined in a multivariate likelihood analysis. In simulated events, the algorithm rejects 95% (50%) of merged  $b$ -jets while retaining 50% (90%) of the single  $b$ -jets, although the exact values depend on  $p_T$ , the jet transverse momentum. The method developed is applied to measure the fraction of double  $b$ -hadron jets as a function of jet  $p_T$ , using  $4.7 \text{ fb}^{-1}$  of  $pp$  collision data at  $\sqrt{s} = 7 \text{ TeV}$  collected by the ATLAS experiment at the Large Hadron Collider in 2011.

*Keywords:* ATLAS Experiment, Jets, Jet Substructure,  $b$ -jet Production, QCD, Gluon Splitting,  $b$ -tagging.

# Contents

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>The Multivariate Analysis</b>              | <b>2</b> |
| 1.1      | Multivariate methods . . . . .                | 2        |
| 1.2      | Likelihood training and performance . . . . . | 11       |
| 1.3      | Systematic uncertainties . . . . .            | 12       |
| 1.4      | Other Monte Carlo generators . . . . .        | 22       |

# Chapter 1

## The Multivariate Analysis

After the evaluation of the best discriminating variables, a tagging algorithm, capable of efficiently identifying single  $b$ -jets while rejecting merged  $b$ -jets, can be constructed using several different approaches. In this chapter we present the results of a multivariate likelihood method. Compared to single variable or 2-dimensional cuts analyses, with this technique the sensitivity is largely improved because several variables are combined to achieve the maximum separation power.

### 1.1 Multivariate methods

Multivariate data analysis refers to a statistical technique used to analyze data that is composed of more than one variable. Classification is done through learning algorithms that make use of training events, for which the desired output is known, to determine the mapping function that describes a decision boundary. The following multivariate methods were explored:

- Likelihood ratio estimators (LLR)
- Neural Networks (NN)

- Boosted decision Trees (BDTs)

The method of LLR consists of building a model out of probability density functions (PDF) that reproduces the distributions of the input variables for signal and background. The likelihood ratio  $y_L(i)$  for event  $i$  is defined by:

$$y_L(i) = \frac{L_S(i)}{L_S(i) + L_B(i)}, \quad (1.1)$$

In the case of poorly correlated variables, the likelihood for being of signal type is obtained by multiplying the signal probability densities of all input variables and normalising this by the sum of the signal and background likelihoods,

$$L_{S(B)}(i) = \prod_{k=1}^{n_{var}} p_{S(B),k}(x_k(i)), \quad (1.2)$$

and where  $p_{S(B),k}(x_k(i))$  is the signal (background) PDF for the  $k$ th input variable  $x_k$ . All the PDFs are normalized to one.

The parametric form of the PDFs is generally unknown, however it is possible to empirically approximate its shape by nonparametric functions. Nonparametric models differ from parametric models in that the model structure is not specified a priori but is instead determined from the data sample used for training. A histogram is a simple example of a nonparametric estimate of a probability distribution. The nonparametric functions can be chosen individually for each variable and can be either polynomial splines of various degrees fitted to binned histograms<sup>1</sup> or unbinned kernel density estimators (KDE). The basic idea in the KDE approach is to model the distribution of a sample of events as the sum of individual equal-area Gaussian kernels,  $G$ ,

---

<sup>1</sup>A spline is a sufficiently smooth polynomial function that is piecewise-defined, and possesses a high degree of smoothness at the places where the polynomial pieces connect. It is often referred to as polynomial interpolation.



for each event. For a PDF  $p(x)$  of a variable  $x$ , one finds [1],

$$p(x) = \frac{1}{Nh} \sum_{i=1}^N G\left(\frac{x - x_i}{h}\right) = \frac{1}{N} \sum_{i=1}^N G_h(x - x_i), \quad (1.3)$$

where  $N$  is the number of training events,  $G_h(t) = G(t/h)h$  is the kernel function, and the extent of each event's contribution,  $h$ , can be kept constant for the entire training sample or be calculated adaptively as a function of the local density of events.  $h$  is termed the “bandwidth” of the kernel or, also, the “smoothing parameter”. The non-adaptive (NA) bandwidth, for a Gaussian kernel function, takes the form of

$$h_{\text{NA}} = \left(\frac{4}{3}\right)^{1/5} \sigma_x N^{-1/5} \quad (1.4)$$

where  $\sigma_x$  is the RMS of the variable  $x$ . The adaptive (A) approach uses as input the result of the non-adaptive KDE, but also takes into account the local event density. The adaptive bandwidth  $h_{\text{A}}$  then becomes a function of  $p(x)$

$$h_{\text{A}} = \frac{h_{\text{NA}}}{\sqrt{p(x)}}. \quad (1.5)$$

The adaptive approach improves the shape estimation in regions with low event density.

Instead of unbinned training data, the KDE approach we implemented uses a finely-binned histogram as input, which allows to significantly speed up the algorithm. The calculation of the optimal bandwidth is the first step. Subsequently, the smoothed high-binned histogram estimating the PDF shape is created by looping over the bins of the input histogram and summing up the corresponding kernel functions using  $h_{\text{A}}$  ( $h_{\text{NA}}$ ) in the case of the non-adaptive (adaptive) mode. This output is the final estimate for the PDF.

The smoothness of the kernel density estimate is evident compared to the discreteness of a histogram; kernel density estimates converge faster to the true underlying density for continuous random variables.

A generalization of the LLR estimator to  $n_{var}$  dimensions, where  $n_{var}$  is the number of input variables, exist in the theory today. If the multidimensional PDF for a signal and background were known, this classifier would exploit the full information contained in the input variables, and would hence be optimal. In practice however, huge training samples are necessary to sufficiently populate the phase space<sup>2</sup>. Kernel estimation methods may be used to approximate the shape of the PDF for finite training samples.

A simple probability density estimator denoted as Projective likelihood estimator *range search*, or PDE-RS, has been suggested in Ref. [2]. The PDE for a given test event is obtained by counting the number of training events that occur in the “vecinity” of the test event. The classification of the test event as being either signal or background type can be conducted by a local estimate of the probability density of it belonging to either class. The  $n_{var}$ -dimensional volume that encloses the “vecinity” is user-defined and adaptive: the volume is defined in each dimension with respect to the RMS of that dimension, estimated from the training sample. Although the adaptive volume adjustment is flexible and should perform better, it significantly increases the computing time of the PDE-RS discriminant.

One of the shortcomings of the original PDE-RS implementation is its sensitivity to the exact location of the sampling volume boundaries: an infinitesimal change in the boundary placement can include or exclude a training event. Kernel functions mitigate these problems by weighting each event within the volume as a function of its distance to the test event.

PDE-RS can yield competitive performance if the number of input variables is not too large and the statistics of the training sample is ample; on

---

<sup>2</sup>Due to correlations between the input variables, only a sub-space of the full phase space may be populated.

the other hand, it is a slowly responding classifier.

An artificial Neural Network (NN) is a nonlinear discriminant. It is, most generally speaking, a simulated collection of interconnected neurons, with each neuron producing a certain response at a given set of input signals. It can be viewed as a mapping from a space of input variables  $x_1, \dots, x_{n_{var}}$  onto, in the case of a signal-versus-background discrimination problem, a one-dimensional output variable. The behaviour of an artificial neural network is determined by the layout of the neurons, the weights of the inter-neuron connections, and by the response of the neurons to the input, described by the neuron response function. The neuron response function maps the neuron input (in  $R^n$ ) onto the neuron output ( $R$ ); often it can be separated into a synapse function ( $R^n \rightarrow R$ ) and a neuron activation function ( $R \rightarrow R$ ). The neuron activation function can be either a *linear*, *sigmoid*, *tanh*, or a *radial* function.

While in principle a neural network with  $n$  neurons can have  $n^2$  directional connections, the complexity can be reduced by organising the neurons in layers and only allowing direct connections from a given layer to the following one. This kind of neural network is termed multi-layer perceptron. The first layer of a multilayer perceptron is the input layer, the last one the output layer, and all others are hidden layers. For a classification problem with  $n_{var}$  input variables the input layer consists of  $n_{var}$  neurons that hold the input values,  $x_1, \dots, x_{n_{var}}$ , and one neuron in the output layer that holds the output variable, the neural net estimator  $y_{NN}$ .

A decision tree is a binary tree structured classifier similar to the one sketched in Fig. 1.1. The training, building or growing of a decision tree is the process that defines the splitting criteria for each node. The training starts with the root node, where an initial splitting criterion for the full train-

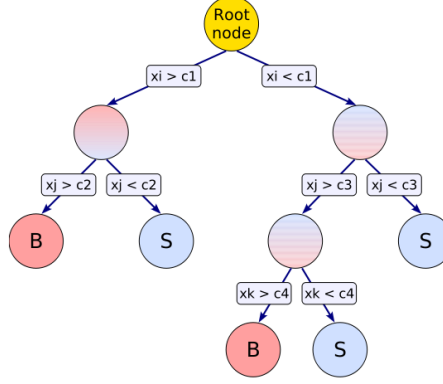


Figure 1.1: Schematic view of a decision tree. Starting from the *root node*, a sequence of binary splits using the discriminating variable  $x_i$  is applied to the data. The leaf nodes at the bottom end of the tree are labeled “S” for signal and “B” for background depending on the majority of events that end up in the respective nodes. Image taken from Reference [4].

ing sample is determined. The split results in two subsets of training events that each go through the same algorithm of determining the next splitting iteration. This procedure is repeated until the whole tree is built. At each node, the split is determined by finding the variable and corresponding cut value that provides the best separation between signal and background. The phase space is split this way into many regions that are eventually identified as “signal-like” or “background-like”, depending on the majority of training events that end up in the final *leaf* node. The boosting of a decision tree extends this concept from one tree to several trees which form a *forest*. Boosting increases the statistical stability of the classifier and typically also improves the separation performance compared to a single decision tree [3].

## Training and testing the MVA methods

A sub-set of the dijet Monte Carlo sample was used for training the three methods in the context of the Toolkit for Multivariate Data Analysis, TMVA [4], written in C++ language. After the event and jet selections, described in Section ??, were performed, the  $b$ -tagged jets were classified as signal (single  $b$ -jets) or background (merged  $b$ ).

Based on discrimination power, correlation and pile-up dependence, three variables were selected for the training: the jet track multiplicity, the track-jet width and the  $\Delta R$  between the axes of two  $k_t$  subjets in the jet. Signal and background datasets were given to the multivariate methods as input, containing a list of the  $b$ -tagged jets with the information of their  $p_T$  and values for the chosen tracking variables. Other variables such as  $\tau_2$  or  $\max\{\Delta R(trk, trk)\}$  were also tested leading to no gain in performance.

Two different likelihood configurations were evaluated: the simple likelihood ratio estimator, using an adaptive Gaussian KDE strategy for the estimation of the PDFs of the input variables; and the more sophisticated PDE-RS approach, adopting an adaptive mode for the volume search. The multidimensional PDE-RS classifier offers no gain in discrimination with respect to the more simpler likelihood method, with the further disadvantage of being more time consuming. The likelihood method shows good performance, and, given the low correlation of the input variables across  $p_T$ , constitutes an adequate method for single-merged discrimination with a fast training step.

An Multi-layer perceptron (MLP) neural net, with two hidden layers of  $n_{var}$  and  $n_{var} - 1$  neurons respectively, was also trained ( $n_{var} = 3$ ). Two different neuron activation functions were tested, *tanh* and *sigmoid*, with the latter showing better performance. The initial NN training was carried

out in 600 cycles, also termed “epochs”, for a fast implementation of the method. Although, faster, this configuration led to an irregularly shaped output. This was only fixed with a 3000-epoch training.

The Boosted Decision Tree (BDT) approach was implemented with 400 trees in the BDT forest. Due to the simplicity of the method where each training step (node splitting) involves only a one-dimensional cut optimisation, little tuning was required in order to obtain reasonably good results.

Examples of the distributions of the final output for these methods, evaluated in an orthogonal sample of simulated dijet events, are displayed in Fig. 1.3 for a medium  $p_T$  bin.

The outputs of the explored MVA discriminants are different in terms of shape and range, although the latter could be rearranged with a suitable variable transformation. In spite of these distinct features the performances of the different methods, expressed in rejection of merged  $b$ -jets as a function of single  $b$ -jet efficiency (see Section ??), agree within statistics, see Fig. 1.3. As opposed to Neural Network discriminants with large number of training cycles, the training and the application of the likelihood are very fast operations that are suitable for very large data sets and tuning of the training parameters. Although also very fast, a shortcoming of decision trees is their instability with respect to statistical fluctuations in the training sample from which the tree structure is derived. If two input variables exhibit similar separation power, a fluctuation in the training sample may cause the tree growing algorithm to decide to split on one variable, while the other variable could have been selected without that fluctuation. In such a case the whole tree structure is altered below this node, possibly resulting also in a substantially different classifier response [4]. It is for these reasons that the likelihood classifier is the selected method for our tagger.

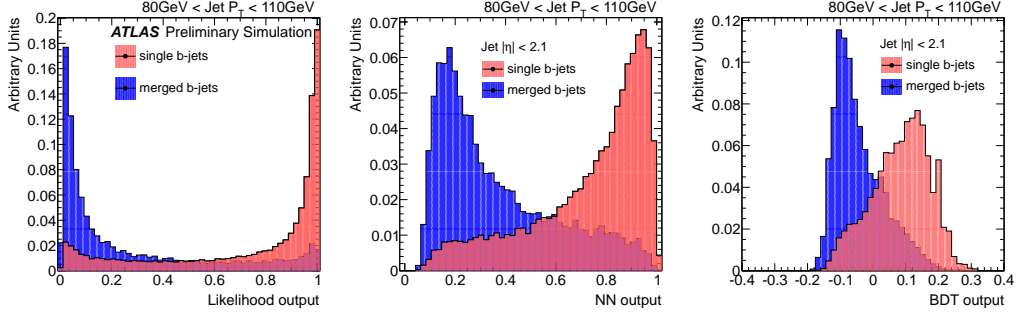


Figure 1.2: Distribution of the MVA discriminant outputs for the Likelihood (a), Neural Network (b) and Boosted Decision Trees (c) classifiers, for single and merged  $b$ -jets between 80 GeV and 110 GeV.

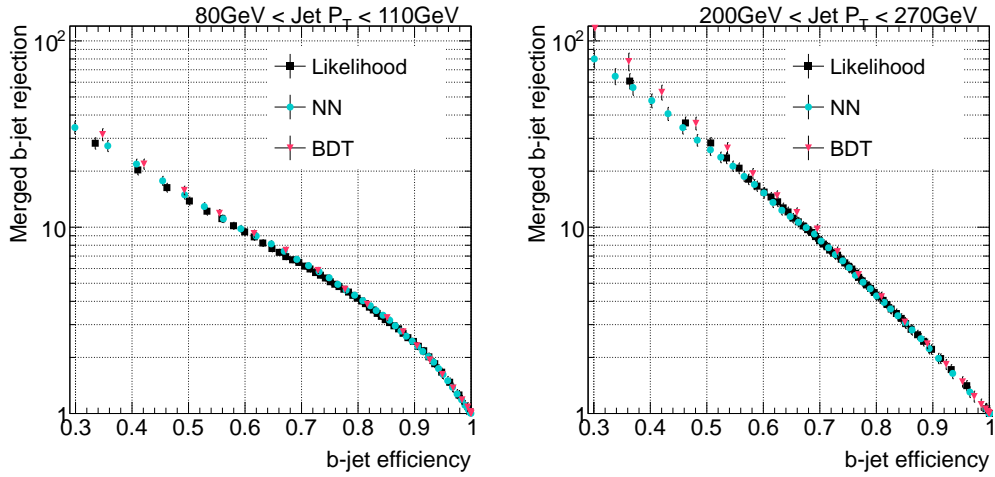


Figure 1.3: Rejection of merged  $b$ -jets as a function of single  $b$ -jet efficiency for the the different MVA methods evaluated for low and high jet  $p_T$ .

## 1.2 Likelihood training and performance

As indicated in section 1.1, a discriminant between single and merged  $b$ -jets was built by training a simple likelihood estimator, with the following three variables as input,

1. Jet track multiplicity
2. Track-jet width
3.  $\Delta R$  between the axes of 2  $k_t$  subjets within the jet

Given the correlation of the variables with the jet transverse momentum, the training sample was categorized in bins of calorimeter jet  $p_T$ , and independent likelihood classifiers were built for each category.

Due to the lack of statistics of merged jets in the low  $p_T$  bins, signal and background jets were not weighted by the dijet samples cross-sections to allow the contribution of subleading lower  $p_T$  jets from high  $p_T$  events. The gain in statistics in the first  $p_T$  bin was of more than 500%. It is important to stress that, although essential for data to MC comparisons (see Section ??), the weighting of the dijet Monte Carlo samples by their respective cross-sections is not necessary for studies performed at simulation level only. For the evaluation of the method the same procedure was followed.

The distribution of the likelihood output for single and merged  $b$ -jets is shown in Fig. 1.4 for low, medium and high transverse momentum jets.

The performance of the tagger in the simulation can be displayed in a plot of rejection of merged  $b$ -jets,  $(1/\epsilon_{bkg})$ , as a function of single  $b$ -jet efficiency,  $\epsilon_{sig}$ ; where  $\epsilon_{bkg}$  ( $\epsilon_{sig}$ ) is the probability that a double (single)  $b$ -hadron jet passes the single  $b$ -jet tagger. This is shown in Fig. 1.5 for the eight bins of jet  $p_T$  mentioned in section ??. The performance improves with  $p_T$ :

- $p_T > 40$  GeV: rejection above 8 at 50% eff.



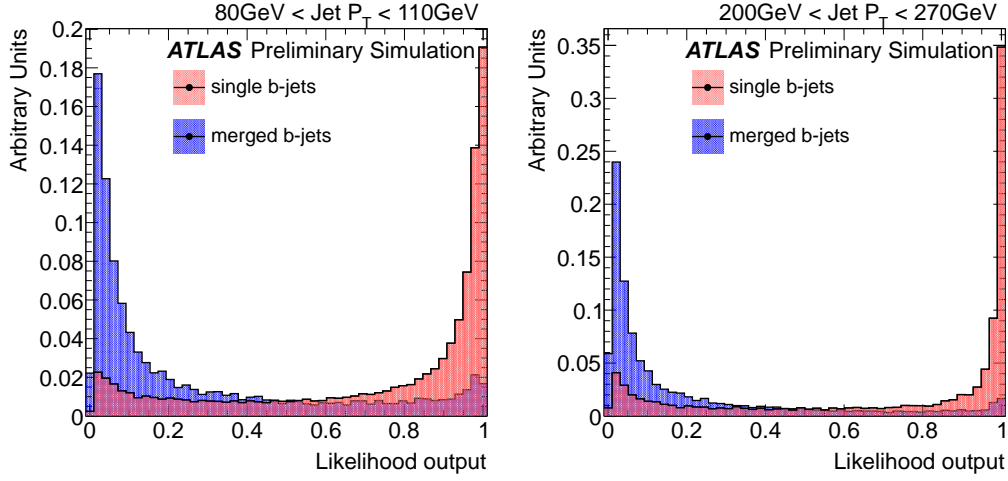


Figure 1.4: Distribution of the likelihood output for single and merged  $b$ -jets for medium and high  $p_T$  jets.

- $p_T > 60$  GeV: rejection above 10 at 50% eff.
- $p_T > 200$  GeV: rejection above 30 at 50% eff.

The rejection of merged jets attained as a function of  $p_T$  for the 50% and 60% single  $b$ -jet efficiency working points are summarized in Table 1.1, together with their relative statistical error. These are propagated from the Poisson fluctuations of the number of events in the merged and single  $b$  distributions. The error is slightly lower for the 60% efficiency working point because a higher efficiency allows for a greater number of Monte Carlo events to measure the performance.

### 1.3 Systematic uncertainties

The development, training and performance determination of the tagger is based on simulated events. Although the agreement between simulation and data explored in section ?? is a necessary validation condition, it is also

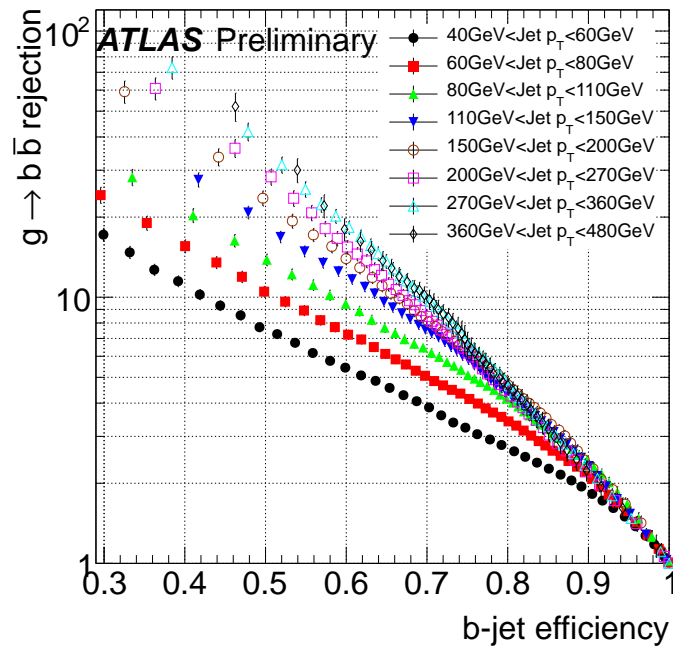


Figure 1.5: Rejection of merged  $b$ -jets as a function of single  $b$ -jet efficiency for dijet events in 8 jet  $p_T$  bins.

| Jet $p_T$<br>(GeV ) | single $b$ -jet efficiency 50% |           | single $b$ -jet efficiency 60% |           |
|---------------------|--------------------------------|-----------|--------------------------------|-----------|
|                     | Rejection                      | stat.err. | Rejection                      | stat.err. |
| 40 - 60             | 8                              | 4%        | 5                              | 3%        |
| 60 - 80             | 10                             | 4%        | 7                              | 4%        |
| 80 - 110            | 14                             | 5%        | 9                              | 4%        |
| 110 - 150           | 19                             | 5%        | 12                             | 4%        |
| 150 - 200           | 23                             | 5%        | 14                             | 5%        |
| 200 - 270           | 30                             | 7%        | 16                             | 6%        |
| 270 - 360           | 36                             | 7%        | 19                             | 6%        |
| 360 - 480           | 41                             | 8%        | 18                             | 8%        |

Table 1.1: The merged  $b$ -jet rejection for the 50% and 60% efficiency working points in bins of  $p_T$ .

important to investigate how the tagger performance depends on the systematic precision with which the MC simulates the data. In particular we have considered:

- presence of additional interactions (pile-up);
- uncertainty in the  $b$ -jet tagging efficiency;
- uncertainty in the track reconstruction efficiency;
- uncertainty in the track transverse momentum resolution;
- uncertainty in the jet transverse momentum resolution;
- uncertainty in the jet energy scale.

#### *I. Pile-up*

The size of this effect was studied by comparing the performance of the

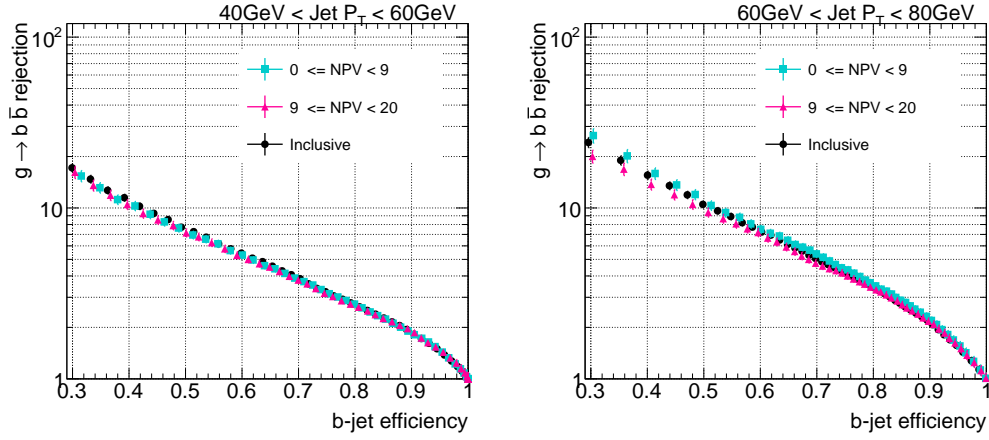


Figure 1.6: Rejection of merged  $b$ -jets as a function of single  $b$ -jet efficiency in bins of NPV for two low jet  $p_T$  bins.

likelihood discriminant with  $b$ -jets in events with small (1-9) and large (9-20) number of primary vertices. A comparison of the performance in these two sub-samples relative to the inclusive sample is shown in Fig. 1.6 for the two lowest  $p_T$  bins, where the effect of pile-up is more important. As expected from the use of tracking (as opposed to calorimeter) variables, no significant dependence with pile-up is observed. Performance differences between high and low number of primary vertices events are  $\leq 2\%$ . The impact of pile-up might be larger in 2012 data.

## II. $b$ -tagging efficiency

The performance of heavy-flavor tagging in Monte Carlo events is calibrated to experimental data by means of the scale factors (SFs). The SFs are defined as the ratio of the heavy-flavor tagging efficiency in data over that in Monte Carlo for the different jet flavors. They are measured by the ATLAS Flavour Tagging Working group, and their measurement carries a systematic uncertainty.

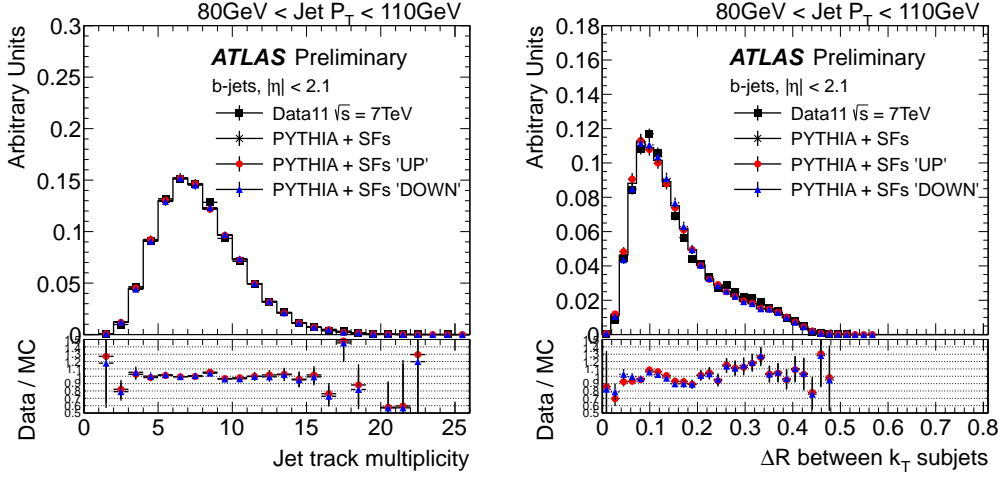


Figure 1.7: The effect of a variation in the  $b$ -tagging Scale Factors on the tracking variables distributions. Scale Factors were varied up (down) by 1-sigma to evaluate the systematic uncertainty from this source. The ratio data over MC is shown for MC PYTHIA with SFs varied up (circles) and down (triangles).

To estimate the impact of this uncertainty a conservative approach is followed: the SFs are varied in all the  $p_T$  bins simultaneously by one standard deviation both in the up and down directions. The MC distributions weighted by the varied SFs show no major deviations from the nominal, see Fig. 1.7. In the same manner, the effect of the  $b$ -tagging calibration uncertainty on the likelihood performance, shown in Fig. 1.8, is  $< 1\%$ , negligible with respect to the statistical uncertainty. This was indeed expected. The scale factors depend on the true flavor of the jet and on its  $p_T$ , but these are basically constant in the performance determination, which is based on single flavor (true  $b$ -) jets classified in  $p_T$ -bins.

### III. Track reconstruction efficiency

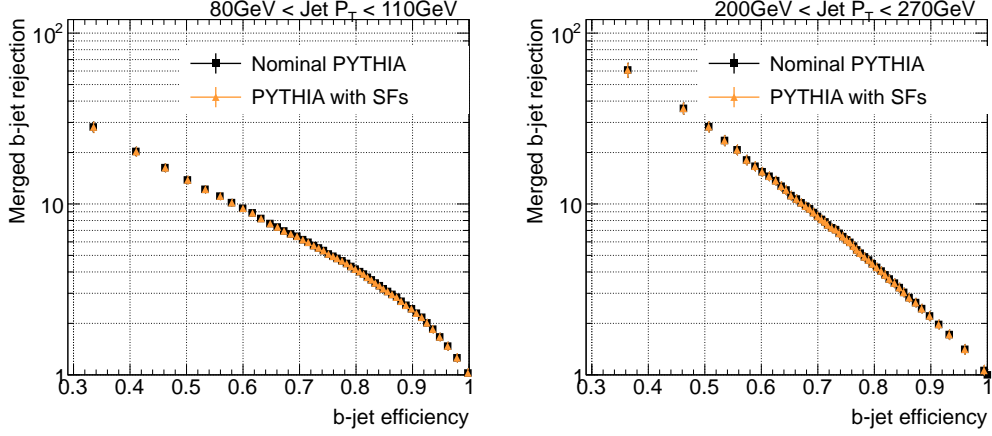


Figure 1.8: Rejection of merged  $b$ -jets as a function of single  $b$ -jet efficiency with and without scale factors as weights.

The track reconstruction efficiency,  $\epsilon_{trk}$ , parametrised in bins of  $p_T$  and  $\eta$ , is defined as:

$$\epsilon_{trk} = \frac{N_{rec}^{matched}(p_T, \eta)}{N_{gen}(p_T, \eta)} \quad (1.6)$$

where  $N_{rec}^{matched}$  is the number of reconstructed tracks matched to a generated charged particle, and  $N_{gen}(p_T, \eta)$  is the number of generated charged particles in that bin<sup>3</sup>. As the track reconstruction efficiency is determined from MC, the main systematic uncertainty results from the level of agreement between data and MC. Since charged hadrons are known to suffer from hadronic interactions with the material in the detector, a good description of the material in MC is needed to get a good description of the track reconstruction efficiency. The uncertain knowledge of the material in the inner detector is the main source of systematic uncertainties in the tracking efficiency [5]. An

---

<sup>3</sup>The matching between a generated particle and a reconstructed track uses a cone-matching algorithm, associating the particle to the track with the smallest  $\Delta R$  within a cone of radius 0.15.

increase (decrease) in material leads to an increase (decrease) in the number of hadronic interactions, hence to a decrease (increase) in the reconstruction efficiency.

The tracking efficiency systematics are given in bins of track  $\eta$ . For tracks with  $p_T^{\text{track}} > 500$  MeV the uncertainties are independent of  $p_T$ : 2% for  $|\eta^{\text{track}}| < 1.3$ , 3% for  $1.3 < |\eta^{\text{track}}| < 1.9$ , 4% for  $1.9 < |\eta^{\text{track}}| < 2.1$ , 4% for  $2.1 < |\eta^{\text{track}}| < 2.3$  and 7% for  $2.3 < |\eta^{\text{track}}| < 2.5$  [5]. All numbers are relative to the corresponding tracking efficiencies.

To test the impact of these uncertainties, a fraction of tracks determined from the track efficiency uncertainty was randomly removed. The tracking variables were re-calculated and the performance of the nominal likelihood was evaluated in the new sample with worse tracking efficiency. The rejection-efficiency curves show a small degradation of the performance which is comparable to the statistical uncertainty. The effect is however systematically present over all 16  $p_T$  bin/working points, without a clear  $p_T$  dependence. We have thus taken the average over  $p_T$ , and obtained a global systematic uncertainty of 4% both for the 50% and 60% efficiency working points. The performance comparison is shown in Fig. 1.9 for two  $p_T$  bins.

#### *IV. Track momentum resolution*

The knowledge of the track momentum resolution is limited by the precision both in the material description of the Inner Detector and in the mapping of the magnetic field. Its uncertainty propagates to the kinematic variables used in the double  $b$ -hadron jet tagger. In order to study this effect, track momenta are over-smearing according to the measured resolution uncertainties, before the track selection cuts are applied. The actual smearing is done in  $1/p_T$ , with an upper bound to the resolution uncertainty given by  $\sigma(1/p_T) = 0.02/p_T$  [6].

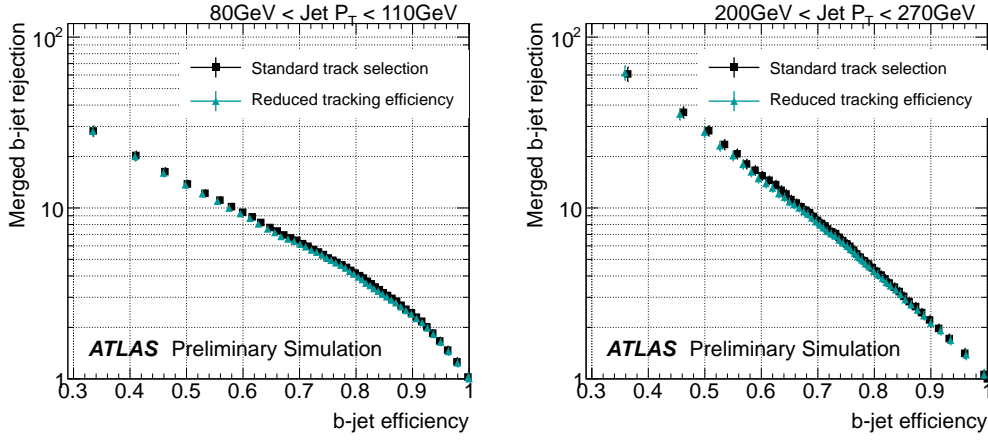


Figure 1.9: Rejection of merged  $b$ -jets as a function of the single  $b$ -jet efficiency showing shift in likelihood performance caused by a reduction in the tracking efficiency.

The effect is found to be negligible, see Fig 1.10.

#### V. Jet energy scale and momentum resolution

The jet energy scale (JES) uncertainty for light jets reconstructed with the anti- $k_t$  algorithm with distance parameter  $R = 0.4$  and calibrated to the EM+JES scale is between  $\sim 4\%$  at low  $p_T$  and  $\sim 2.5\%$  for jets with  $p_T > 60$  GeV in the central region [7]. In the case of  $b$ -jets, and additional uncertainty arising from the modelling of the  $b$ -quark production mechanism and the  $b$ -quark fragmentation was determined from systematic variations of the Monte Carlo simulation. The resulting fractional additional JES uncertainty for  $b$ -jets has an upper bound of 2% for jets with  $p_T \leq 100$  GeV and it is below 1% for higher  $p_T$  jets. To obtain the overall  $b$ -jet uncertainty this needs to be added in quadrature to the light JES uncertainty.

The systematic uncertainty originating from the jet energy scale is obtained by scaling the  $p_T$  of each jet in the simulation up and down by one



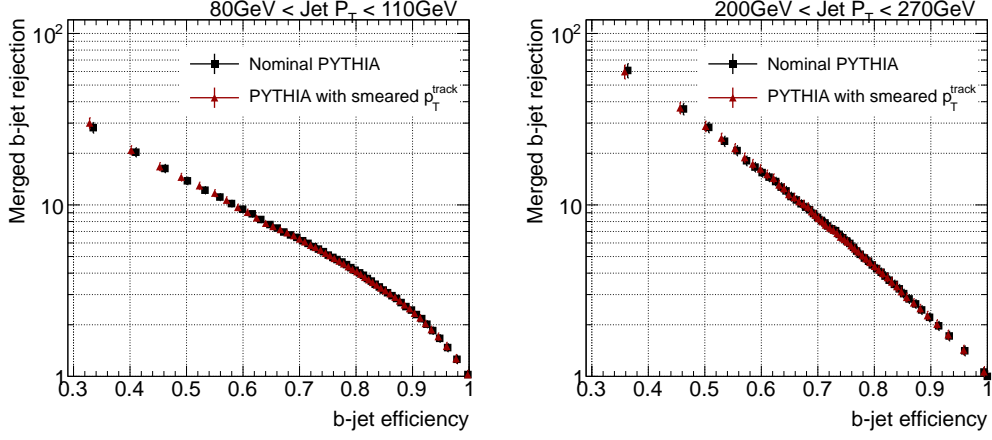


Figure 1.10: Rejection of merged  $b$ -jets as a function of the single  $b$ -jet efficiency showing the effect of the track momentum resolution uncertainty. It is found to be negligible with respect to the statistical uncertainty.

standard deviation according to the uncertainty of the JES. The result is shown in Fig. 1.11a for a medium  $p_T$  bin. The effect on the likelihood performance is an average variation of 5% for the 50% and 60% efficiency working points.

The jet momentum resolution was measured for 2011 data and found to be in agreement with the predictions from the PYTHIA-based simulation [8]. The precision of this measurement, determined in  $p_T$  and  $\eta$  bins, is typically 10%. The systematic uncertainty due to the calorimeter jet  $p_T$  resolution was estimated by over-smearing the jet 4-momentum in the simulated data, without changing jet  $\eta$  or  $\phi$  angles. The performance, shown in Fig. 1.11b, is found to globally decrease by 5%, without a particular  $p_T$  dependence.

The different contributions to the systematic uncertainty on the merged  $b$ -jet rejection are summarized in Table 1.2.

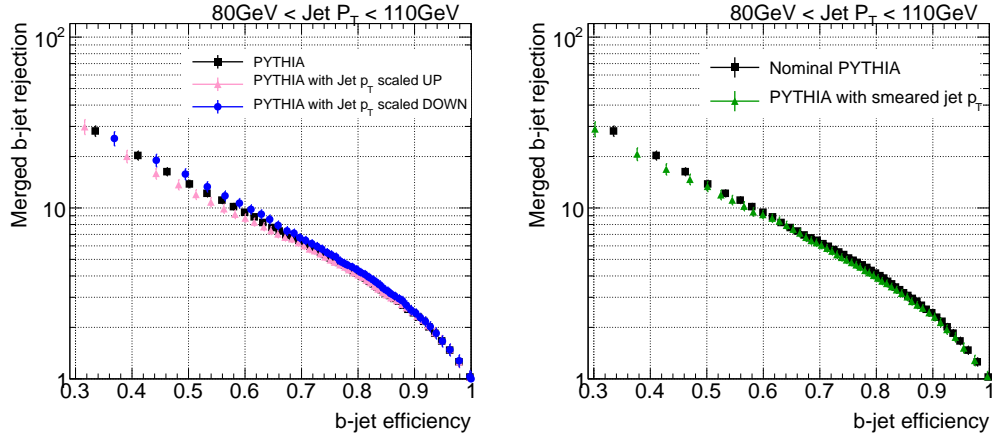


Figure 1.11: Rejection of merged  $b$ -jets as a function of single  $b$ -jet efficiency for (a) jets with smeared  $p_T$  and (b) for jets with varied energy scale compared to nominal.

| Systematic source               | Uncertainty |
|---------------------------------|-------------|
| pile-up                         | 2%          |
| $b$ -tagging efficiency         | negligible  |
| track reconstruction efficiency | 4%          |
| track $p_T$ resolution          | negligible  |
| jet $p_T$ resolution            | 5%          |
| jet energy scale                | 5%          |

Table 1.2: Systematic uncertainties in the merged  $b$ -jet rejection (common to both the 50% and the 60% efficiency working points).

Although the likelihood training was performed in EM+JES calibrated jets, the performance of the tagger was also evaluated in jets calibrated with the LC+JES scheme, described in Section ???. A small degradation of the performance is observed, but comparable with the statistical uncertainties.

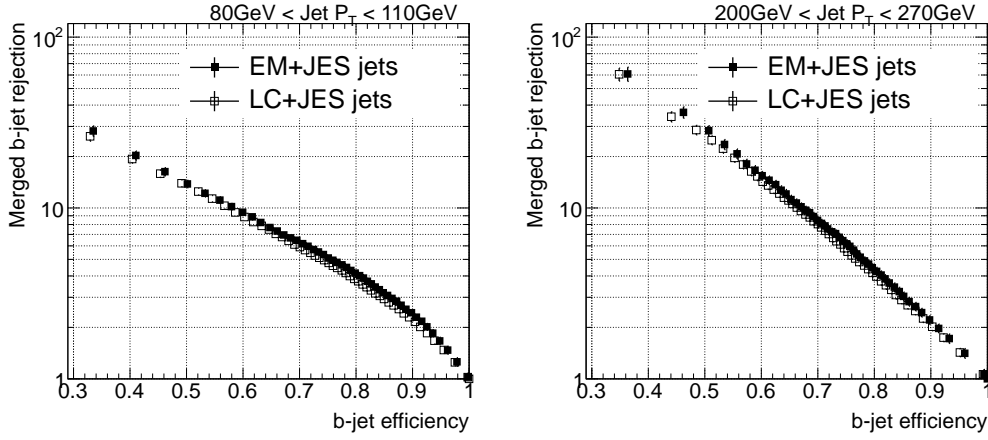


Figure 1.12: Rejection of merged  $b$ -jets as a function of single  $b$ -jet efficiency for jets calibrated to the EM+JES (LC+JES) scale, between 80 GeV and 110 GeV and 200 GeV and 270 GeV.

A comparison of the performances is shown in Fig. 1.12 for two  $p_T$  bins, representative of the jet momentum range covered.

## 1.4 Other Monte Carlo generators

The development, training and performance determination of the tagger has been done using Monte Carlo events generated with the PYTHIA event simulator, interfaced to the GEANT4 based simulation of the ATLAS detector. An immediate question is what the performance would be if studied with a different simulation. In this section we investigate this question for the PYTHIA Perugia tune and the HERWIG++ event generators (Section ??).

Fig. 1.13 shows a comparison of the likelihood rejection, at the 50% efficiency working point, between nominal PYTHIA and the alternative simulations as a function of the jet  $p_T$ . The larger errors are due to the reduced statistics available, which are even lower for the Perugia case than for HER-

WIG.

The performance in HERWIG shows a systematic trend, with agreement at low  $p_T$  and increasingly poorer performances compared to PYTHIA as  $p_T$  grows. For the Perugia tune, on the other hand, there is no definite behavior, with the performance fluctuating above or below the nominal simulation for different  $p_T$  bins consistently with the statistical uncertainties.

The reason for the systematic difference observed between the performances of PYTHIA and HERWIG can be traced to the extent with which jets are accurately modelled. Fig. 1.14 compares the measured jet track multiplicity distributions in  $b$ -tagged jets and the prediction from both simulations, for low and high  $p_T$  jets. It is observed that indeed HERWIG++ does not correctly reproduce the data, particularly at high  $p_T$ . The level of agreement is found to be better for track-jet width and the  $\Delta R$  between the axes of the two  $k_t$  subjects in the jet, the two other variables used for discrimination.

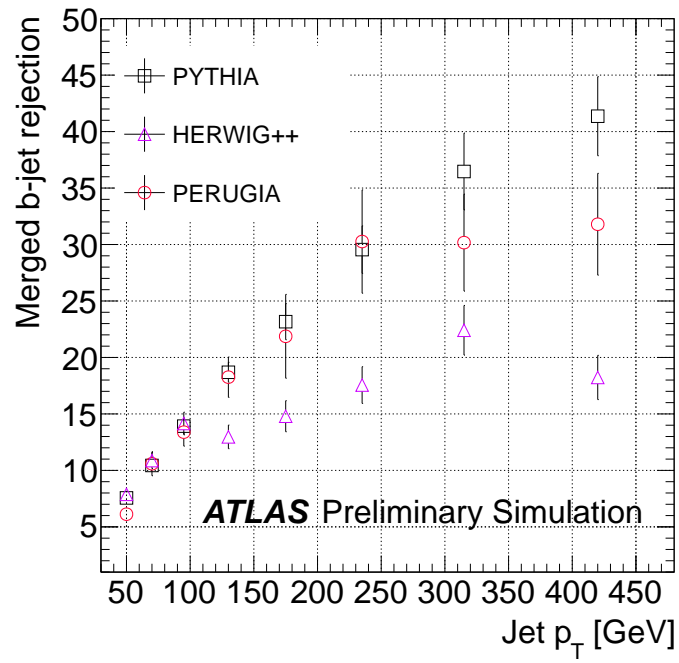


Figure 1.13: Rejection of merged  $b$ -jets as a function of jet  $p_T$  for different Monte Carlo generators, at the 50% efficiency working point.

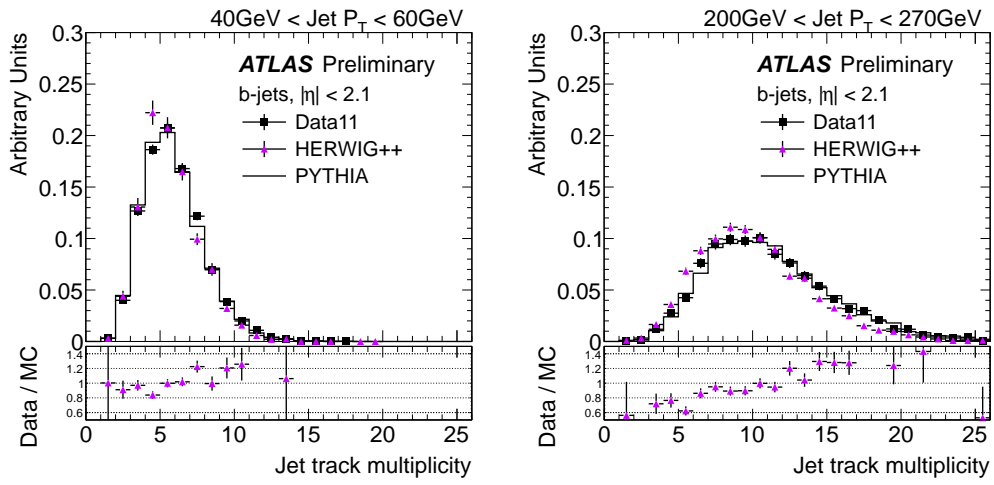


Figure 1.14: Distribution of the jet track multiplicity in 2 different jet  $p_T$  bins, for experimental data collected during 2011 (solid black points) and HERWIG++ events (solid violet triangles). The ratio data over HERWIG++ simulation is shown at the bottom of the plot. PYTHIA distribution is also shown for reference.

# Bibliography

- [1] Scott, D.W. Multivariate Density Estimation: Theory, Practice, and Visualization. *John Wiley and Sons, Inc. United States*, 1992.
- [2] T. Carli and B. Koblitz. A Multivariate discrimination technique based on range searching. *Nucl.Instrum.Meth.*, A501:576–588, 2003.
- [3] Quinlan, J.R. Simplifying decision trees. *International Journal of Human-Computer Studies*, 51(2):497 – 510, 1999.
- [4] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. Von Toerne, and H. Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.
- [5] G. Aad et al. Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC. *New J.Phys.*, 13:053033, 2011.
- [6] ATLAS Collaboration. Estimating Track Momentum Resolution in Minimum Bias Events using Simulation and  $K_s$  in  $\sqrt{s} = 900$  GeV collision data. *ATLAS-CONF-2010-009*, 2010.
- [7] Aad, G. and others. Jet energy measurement with the ATLAS detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV. 2011.

- [8] G. Romeo, A. Schwartzman, R. Piegaia, T. Carli, and R. Teuscher. Jet energy resolution from in-situ techniques with the atlas detector using proton-proton collisions at a center of mass energy  $\sqrt{s} = 7$  tev. *ATL-COM-PHYS-2011-240*, 2011.