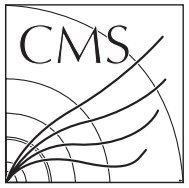
Available on CMS information server

CMS CR -2009/032



The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

16 January 2009 (v2, 27 January 2009)

Strategies for btagging calibration using collider data at CMS

Victor E. Bazterra Thomas Speer

Abstract

The CMS Collaboration is studying several algorithms to identify jets coming from the hadronization of bottom quarks (b-jets) which are present in a wide range of physics processes of interests such as in the decay of top quarks, Higgs bosons and several non-Standard Model particles. All of these b-tagging algorithms rely upon the reconstruction of lower level objects like tracks, vertices, and jets, which might make it difficult for the Monte Carlo simulation to exactly reproduce the performance observed in data. Reliable methods to measure performance of these algorithms directly from collider data have been developed. The CMS b-tagging group is working on several strategies which should yield reliable results already with $10\ \mathrm{pb}^{-1}$ of integrated luminosity.

Presented at *XII Advanced Computing and Analysis Techniques in Physics Research,November 3-7 2008,Erice,Italy,31/01/2009*

DRAFT
CMS Conference Report

2009/01/27
Archive Id: 1.3
Archive Date: 2009/01/16 22:01:28

Strategies for btagging calibration using collider data at CMS
2009 LaTeX/PdfLaTeX version

Victor E. Bazterra[1] and Thomas Speer[2]

[1] On behalf of the CMS Collaboration
University of Illinois at Chicago, 845 W Taylor St., Chicago, IL60607, USA
[2] On behalf of the CMS Collaboration
Brown University, Providence, Rhode Island, USA

Abstract

The CMS Collaboration is studying several algorithms to identify jets coming from the hadronization of bottom quarks (b-jets) which are present in a wide range of physics processes of interests such as in the decay of top quarks, Higgs bosons and several non-Standard Model particles. All of these b-tagging algorithms rely upon the reconstruction of lower level objects like tracks, vertices, and jets, which might make it difficult for the Monte Carlo simulation to exactly reproduce the performance observed in data. Reliable methods to measure performance of these algorithms directly from collider data have been developed. The CMS b-tagging group is working on several strategies which should yield reliable results already with 10 pb^{-1} of integrated luminosity.

1 Introduction

The ability to accurately identify b-jets is vital in reducing the otherwise overwhelming background from hadronization of light quarks and gluons (light-jets) and charm quark (c-jets). Several algorithms are available at the CMS experiment [1], and they take advantage of the properties of b-flavored hadrons to identify b-jets [2]. All of these algorithms rely upon the reconstruction of lower level objects like tracks, vertices and jets. While it was shown in simulation that b-tagging algorithms reach adequate performance in terms of b-jet efficiency and light jet rejection, it is not expected that that all observables on which the algorithms rely will be adequately modelled by the simulation. The simulation can therefore not be used to reliable estimate the performance of the algorithms, and methods to measure the performance directly from collider data are being developed.

The efficiency ϵ_q to tag jets of flavour q as a b-jet (b-tagging efficiency for b-jets and *mistag* rate for c+light-jets) is defined as:

$$\varepsilon_q = \frac{\text{Number of jets of flavor } q \text{ tagged as } b}{\text{Number of jets of flavor } q}. \tag{1}$$

Any b-tagging algorithm can be characterized by measuring their efficiency and mistag rate [2].

The Tevatron collider experiments (CDF [5] and D0 [6]) have developed methods to measure the performance of the tagging algorithms in collider data. Several of these methods are been implemented at CMS. Other methods, such as those based on decays of the top quark, can only be implemented at the Large Hadron Collider (LHC) where top quarks are going to be produced in large number.

This article is organized as follow. Sections 2.1 and 2.2 describe the *Ptrel* and *System8* methods. Section 2.3 shows how rejection rates from light quarks can be measure by counting the number of *negative tagged jets* to model the mistag rate due to detector effects like resolution, badly reconstructed tracks, etc. Top-quark based methods are then presented in sections 2.4 and 2.5. In section 3 shows the implementation details related to all methods. Section 4 shows the initial results for some methods. Finally, conclusion is presented in section 5.

2 Methods

2.1 The Ptrel method

The method is based on data samples that have at least two reconstructed jets and a non-isolated muon close to one of the jets. The jets are reconstructed with a simple iterative algorithm of cone $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.5$, where η and ϕ are the pseudorapidity and azimuthal angle, respectively [1]. Only those jets with $p_T > 20$ GeV and $|\eta| < 2.5$ are used by the method. From this data set, four samples can be defined:

- The *muon-in-jet* (n) sample contains at least two reconstructed jets and a non-isolated muon with $\Delta R(\mu, \text{jet}) < 0.4$, $p_T^\mu > 6.0$ GeV and $|\eta^\mu| < 2.5$. In the case that more than one muon is found, only the muon with the highest p_T is considered;
- The *muon-in-jet-away-jet-tagged* (p) sample is a subset of the *muon-in-jet* sample, where at least one of the remaining jet is loosely tagged as a b-jet;

[1] For more information about the CMS coordinate conventions see [7]

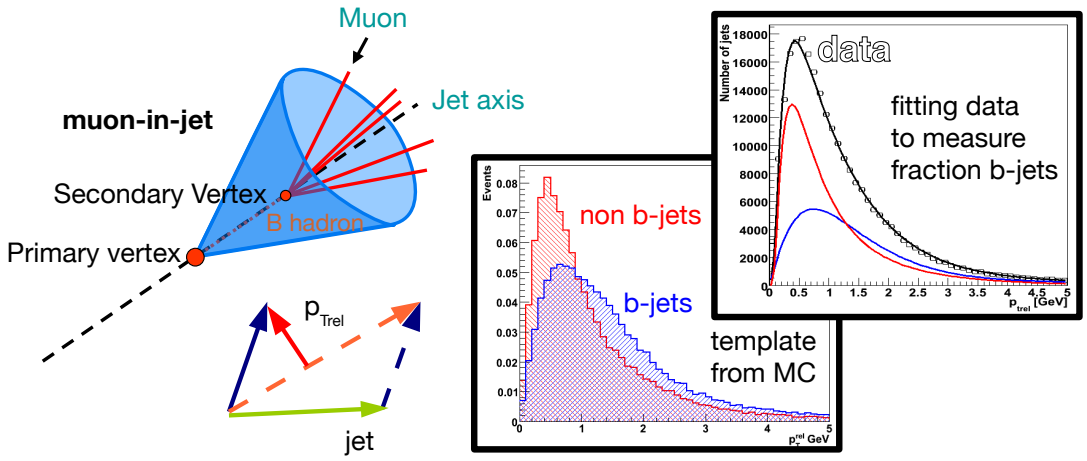


Figure 1: Schematic representation of moun-in-jet sample (left), p_{Trel} distributions for b- and c+light-jets (center) and b-fraction measurement from data by p_{Trel} fits (right).

- The *muon-in-jet-tagged* (n^{tag}) sample is a subset of the *muon-in-jet* sample, where the jet with the muon is tagged as a b-jet;
- The *muon-in-jet-double-tagged* (p^{tag}) sample is a subset of the *muon-in-jet-away-jet-tagged* sample, where the jet with the muon is tagged as a b-jet.

For all these cases, the p_{Trel} is defined as the transverse momentum of the muon relative to the direction of the total muon-jet momentum vector,

$$p_{Trel} = \frac{p^{\mu} \times p^{\mu+\text{jet}}}{|p^{\mu+\text{jet}}|}. \tag{2}$$

The basic idea of the Ptrel methods is to start with the *muon-in-jet* sample and measure the number of b-jets by fitting the p_{Trel} distribution of the muons to a linear combination of the b- and c+light-jet templates [3] (c.f. Figure 1). The process is repeated after tagging the jet with the muon by fitting the p_{Trel} distribution for *muon-in-jet-tagged* sample. The b-tagging efficiency is then calculated as the ratio between the number of b-jets after and before tagging, as determined by the p_{Trel} fits. The same procedures can be applied to the *muon-in-jet-away-jet-tagged* samples, but these will have much lower statistics.

The semileptonic $c\bar{c}$ and $b\bar{b}$ Monte Carlo (MC) simulation are used to create the p_{Trel} distributions of $c \to \mu X$ and $b \to \mu X$ [2] transitions, respectively. For light jets, the p_{Trel} distribution were similar in shape to those from c-jets.The ptrel distribution of c-jets is therefore taken as the template for c+light-jets. The templates were obtained for different ranges of jet p_T and $|\eta|$, and before and after tagging the *muon-in-jet* sample.

2.2 The System8 method

The System8 method is based on the same set of samples (defined in the section 2.1), only adding a cut on the p_{Trel} of the muon ($p_{Trel} > 0.8$ GeV) for both the muon-in-jet (n^{mu}) and muon-in-jet-away-jet-tagged (p^{mu}) samples. All samples are related by the following system of

[2] $b \to \mu X$ includes both direct $b \to \mu$ and cascade $b \to c \to \mu$ decays.

equations

$$\begin{aligned} n &= n_b + n_{cl} \\ p &= p_b + p_{cl} \\ n^{\text{tag}} &= \varepsilon_b^{\text{tag}} n_b + \varepsilon_{cl}^{\text{tag}} n_{cl} \\ p^{\text{tag}} &= \beta\, \varepsilon_b^{\text{tag}} p_b + \alpha\, \varepsilon_{cl}^{\text{tag}} p_{cl} \\ n^{\text{mu}} &= \varepsilon_b^{\text{mu}} n_b + \varepsilon_{cl}^{\text{mu}} n_{cl} \\ p^{\text{mu}} &= \varepsilon_b^{\text{mu}} p_b + \varepsilon_{cl}^{\text{mu}} p_{cl} \\ n^{\text{tag,mu}} &= \kappa_b\, \varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}} n_b + \kappa_{cl}\, \varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}} n_{cl} \\ p^{\text{tag,mu}} &= \kappa_b\, \beta\, \varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}} p_b + \kappa_{cl}\, \alpha\, \varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}} p_{cl}\ . \end{aligned} \tag{3}$$

The terms on the left hand side represent the total number of muon-jets in each sample before tagging $\{n,\, p,\, n^{\text{mu}},\, p^{\text{mu}}\}$ and after tagging $\{n^{\text{tag}}, p^{\text{tag}},\, n^{\text{tag,mu}}, p^{\text{tag,mu}}\}$. The right hand side of the equations consist of the number of b- and c+light-jets in each sample $\{n_b,\, n_{cl},\, p_b,\, p_{cl}\}$, and the tagging efficiencies for b- and c+light-jets $\{\varepsilon_b^{\text{tag}}, \varepsilon_{cl}^{\text{tag}}\}$. The scale factors β and α represent the ratio of the tagging efficiencies for b and c+light-jets respectively, corresponding to the two samples n and p. They are defined as

$$\begin{aligned} \beta &= \frac{\varepsilon_b^{\text{tag}}\text{from muon-in-jet-away-jet-tagged sample}}{\varepsilon_b^{\text{tag}}\text{from muon-in-jet sample}}, \\ \alpha &= \frac{\varepsilon_{cl}^{\text{tag}}\text{from muon-in-jet-away-jet-tagged sample}}{\varepsilon_{cl}^{\text{tag}}\text{from muon-in-jet sample}}. \end{aligned} \tag{4}$$

The $\varepsilon_b^{\text{mu}}$ and $\varepsilon_{cl}^{\text{mu}}$ are the efficiency of applying a cut on the p_{Trel} of the muon. These equations assume a weak correlation between the p_{Trel} cut and the b-tagging algorithm that is used, and it is only corrected by the correlation factors κ_b and κ_{cl}, defined as

$$\kappa_b = \frac{\varepsilon_b^{\text{tag,mu}}}{\varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}}} \quad \text{and} \quad \kappa_{cl} = \frac{\varepsilon_{cl}^{\text{tag,mu}}}{\varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}}}\,. \tag{5}$$

Equations 3 defines a non-linear system with 8 equations and 8 unknowns. All parameters on the left hand side are measured on the different samples, and the 8 unknown efficiencies are on the right hand side. This assumes that the correlation factors α, β, κ_b and κ_{cl} can be safely derived from MC simulations, see figure 2. The efficiencies and mistag rate can therefore be determined by solving this system of equations.

2.3 Mistag rate measurement using negative tags

The negative tag is a concept usable for lifetime-based b-tagging algorithms. It is based on the possibility of defining positive (respectively negative) discriminant if the associated tracks are reconstructed downstream (upstream) with respect to the primary interaction vertex [4]. Negative and positive discriminators used by the b-tagging algorithms should be approximately symmetric for light-jets. This is because light-jets are produced at the primary vertex and therefore any deviation from it is due to resolution effects or mis-reconstruction. However, because of the long lifetime of B hadrons, the discriminator of b-jets will have a very different distribution with large positive values (c.f. Figure 3). The detailed definition for positive and negative discriminators depend on the b-tagging algorithms [4].

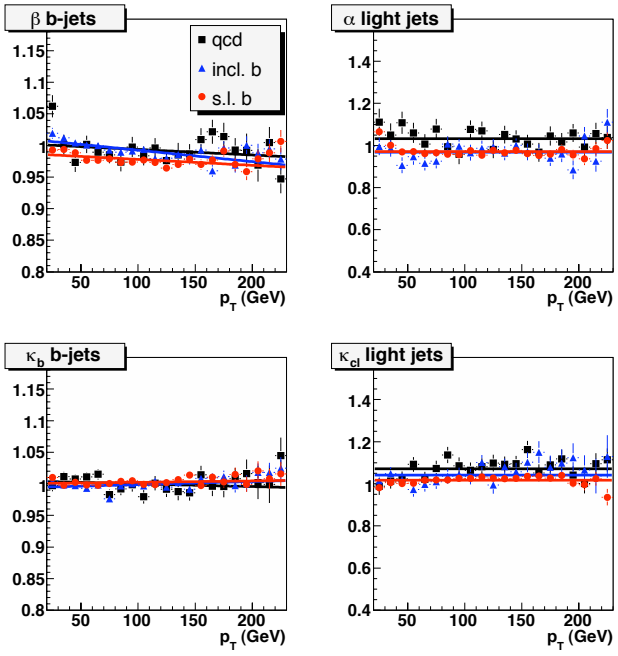


Figure 2: System8 scale and correlation factors for TCL operating point (see the section with results) for different Monte Carlo samples. The factors are fitted to a linear function for each of the samples considered.

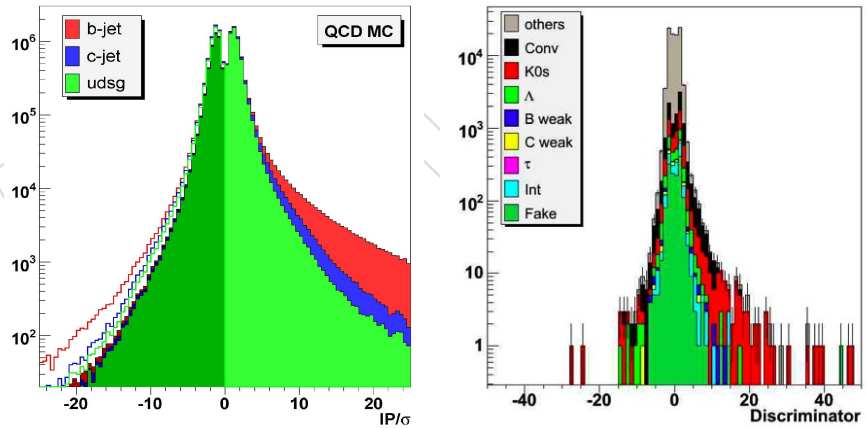


Figure 3: Impact parameter significance of the second highest $|IP/\sigma_{IP}|$ track (used for the *loose* and *medium* TrackCounting operating points) in QCD Monte Carlo jets (left). Contributions for the different jet flavours are shown as colored filled areas. The negative impact parameters are presented darker than the positive ones. A positive tag veto is applied to jets with a negative impact parameter track: this jet is rejected if it has any track with $|IP/\sigma_{IP}| > 4$. Negative $|IP/\sigma_{IP}|$ distributions without requiring the positive tag veto are presented as colored histograms (green for light-jets, blue for light+c-jets, red for light+c+b-jets). Various factors that affect the computation of the light mistag rate from the overall light+c+b negative tag rate (right).

Due to the symmetry of the distribution for light jets, the mistag rate due to light-jets is expected to be proportional to the negative tagging efficiency. The mistage rate can therefore be measure from multi-jet data, so mistag rate can be estimated as

$$\varepsilon_{\text{light}} = \varepsilon_{-} \cdot R_{\text{light}} \,. \tag{6}$$

where ε_{-} is the negative tagging efficiency derived from data. The proportionality constant $R_{light} = \varepsilon^{\text{MC}}_{\text{light}}/\varepsilon^{MC}_{-}$ is the ratio between the tagging efficiencies of light-jets and the negative tagging efficiency of all light+c+b-jets extracted from simulation. The value of R_{light} is close to one in case the negative and the positive discriminats are exactly symmetric. It increases due to the presence of long-lived particles, conversions and interactions with material that bias light-jets towards positive discriminator values (right of figure 3). It may nevertheless decrease due to small number of c- and b-jets that are negatively tagged (left of figure 3).

2.4 Top-quark based method: Likelihood ratio technique

This method measures the b-jet performance from $t\bar{t}$ events by isolating a jet sample with highly enriched b-jet content using a likelihood ratio. Semi-leptonic and fully leptonic decays are considered.

The likelihood ratio is used both to select top quark decays, and to discriminate between correct and wrong associations between the final state jets and the initial partons in the different combinatorial solutions. The purity of the jet sample can be increased by imposing a selection threshold on the likelihood ratio and the correct jet association is assumed to be the one with the highest likelihood ratio value among the different combinatiorial solutions.

Each decay topology will use observables which exploit the relevant kinematic properties of the events. For each observable, two distributions are derived, one for signal events featuring the correct association (denoted S), and one for background events or signal with wrong associations (denoted B). Once the signal and background distributions are made, the $S/(S+B)$ distribution is derived bin by bin way. The $S/(S+B)$ distribution is then parameterized with a function $f_i(x_i)$ for each observable x_i, and the likelihood is defined as

$$\mathcal{L} = \prod_i f_i(x_i) \quad \text{or} \quad \mathcal{L} = \prod_i \frac{f_i(x_i)}{1 - f_i(x_i)} \tag{7}$$

When a jet sample with sufficient purity has been isolated, the efficiency is obtained by measuring the fraction x_{tag} of the tagged jets on this sample. This fraction is related to the b-tagging efficiency as

$$x_{tag} = \varepsilon_b x_b + \varepsilon_{cl}(1 - x_b), \tag{8}$$

where x_b is fractions b-jets in the sample before tagging and $\{\varepsilon_b, \varepsilon_{cl}\}$ correspond to the b-tagging efficiency and mistag rate respectively. Therefore b-tagging efficiency can thus be written as

$$\varepsilon_b = \frac{1}{x_b}\left[x_{tag} - \varepsilon_{cl}(1 - x_b)\right] \,. \tag{9}$$

From equation 9 it is clear that for a highly pure samples $x_b \to 1$ it is true that $x_{tag} \to \varepsilon_b$. In practice $x_b < 1$ and therefore the fraction of b-jets x_b and the mistag rate ε_{cl} need to be estimated from MC simulations. The appropriate cut on the likelihood ratio is then chosen to minimize the total uncertainty of the measurement. A cut at larger values will reduce the contamination from non-b jets, reducing thus the systematic uncertainty, albeit at the price of a higher statistical uncertainty. Each quantity has thus to be evaluated as a function of the cut on the likelihood ratio.

2.5 Top-quark based method: Flavor-tag consistency likelihood method

Within the SM, top quarks are expected to decays almost 100% of the times to a W boson accompanied by a b-quark. In the semi-leptonic channel, given the b, c-jet identification efficiencies and light quark mistag rates, the number of events with n_b tagged b-jets, n_c tagged c-jets, and n_l tagged light-jets can be predicted. By enforcing a consistency between the predicted number of events with one, two or more tagged jets to the actual number of observed events with that particular combination, the b- and c-tagging efficiencies can be measured.

The following log-likelihood function is minimized to measure the tagging efficiencies and the $t\bar{t}$ cross section

$$\mathcal{L} = -2 \log \prod_n P(N_n, \bar{N_n}), \tag{10}$$

where N_n is the measured number of events with n tagged jets, $\bar{N_n}$ is the expected number of events with n tagged jets and $P(N_n, \bar{N_n})$ is the Poisson distribution.

The number of events $\bar{N_n}$ with $n = 0, 1, 2$ tagged jets is predicted using statistical information about the event jet flavor structure as obtained from MC. This information from the simulated data sets provides the fractions of events $\{f_{ijk}\}$ with i, j, k of b-, c- and light jets, respectively. The fractions are used together with the three tagging efficiencies $\{\varepsilon_b, \varepsilon_c, \varepsilon_l\}$, the acceptance ϵ and the $t\bar{t}$ cross section $\sigma_{t\bar{t}}$ to estimate the expected number of events in the different tagging categories

$$\bar{N_n} = L \cdot \sigma_{t\bar{t}} \cdot \epsilon \cdot \sum_{i,j,k} f_{ijk} \sum_{i'+j'+k'=n}^{i' \le i, j' \le j, k' \le k} \left[C_i^{i'} \epsilon_b^{i'} (1-\epsilon_b)^{(i-i')} C_j^{j'} \epsilon_c^{j'} (1-\epsilon_c)^{(j-j')} C_k^{k'} \epsilon_l^{k'} (1-\epsilon_l)^{(k-k')} \right] \tag{11}$$

where C_j^i are the binomial coefficients and L the luminosity. Equation 11 implies that the contents of the experimental dataset has negligible amount of background events. In case it is impractical to reject all background, it is possible to generalize this equation to

$$\begin{aligned} \bar{N_n} &= \bar{N_n}^{t\bar{t}} + \bar{N_n}^{\texttt{background}} \\ &= L \cdot \sigma_{t\bar{t}} \cdot \epsilon_{t\bar{t}} \cdot \left[\sum_{i,j,k} f_{ijk}^{t\bar{t}} \sum_{i'+j'+k'=n}^{i' \le i, j' \le j, k' \le k} (...) + \right. \\ &+ \left. \frac{\sigma_{\texttt{background}}}{\sigma_{t\bar{t}}} \cdot \frac{\epsilon_{\texttt{background}}}{\epsilon_{t\bar{t}}} \cdot \sum_{i,j,k} f_{ijk}^{\texttt{background}} \sum_{i'+j'+k'=n}^{i' \le i, j' \le j, k' \le k} (...) \right], \end{aligned} \tag{12}$$

where $(...)$ stand for the expression in square brackets from (11). The $\{f_{ijk}^{\texttt{background}}\}$, $\epsilon_{\texttt{background}}$ and $\sigma_{\texttt{background}}$ are similar to those defined in the previous paragraph for processes other than $t\bar{t}$.

3 Implementation design

For most of the algorithm, pre- and post-conditions are very similar. This is why all of them share a basic design defined by the following four points:

- A procedure for extracting essential information for performing measurements while reducing file size and decoupling ourselves from the CMS reconstruction framework (CMSSW) [8].
- This procedure produces files that will be used only by the CMS b-tagging group.

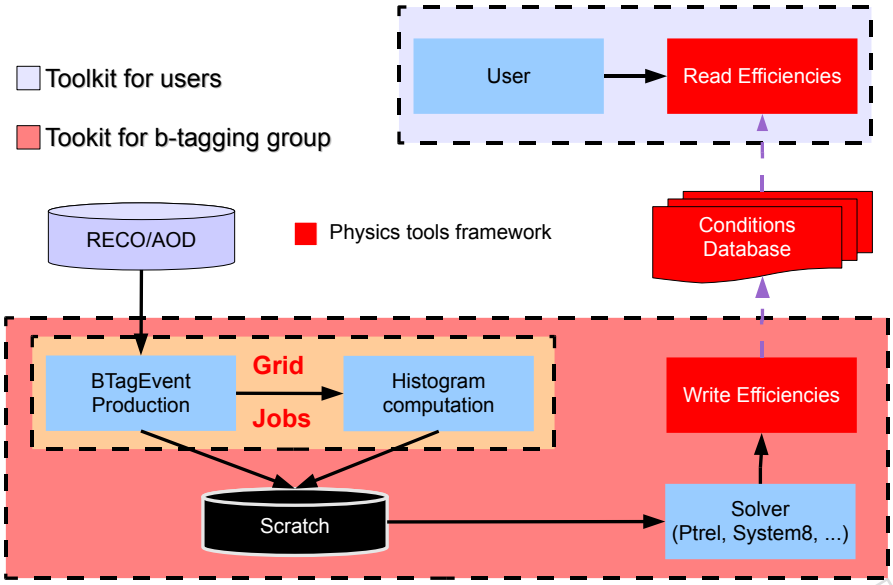


Figure 4:

Algorithm	Operating point	b-efficiency	c-mistag	light-mistag
Track Counting	Loose (TCL)	70.49 ± 0.20	32.33 ± 0.16	9.98 ± 0.02
	Medium (TCM)	50.30 ± 0.21	10.77 ± 0.10	0.96 ± 0.01
	Tight (TCT)	31.94 ± 0.20	2.93 ± 0.06	0.10 ± 0.01

Table 1: Operating points and average tagging efficiencies for the TrackCounting algorithm, determined from MC truth, for jets with $p_T > 20$ GeV and $|\eta| < 2.5$.

- The analysis of the extracted data is performed.
- The results of the measurement will be available to the collaboration by using the condition database [8].

The extensive use of the CMSSW framework will guaranty the use of correct calibrations needed for the reconstruction consistent with the detector conditions. Moreover the distribution of the efficiencies as part of CMS condition database will enforce the use of the measured efficiencies consistent with the reconstruction conditions.

4 Initial results

Only initial results for all methods are presented. More studies are underway for analyzing systematic and statistical uncertainties of each method.

For simplicity, three operating points, *loose*, *medium* and *tight*, are defined in order to select an average fraction of 10%, 1% and 0.1%, respectively, of light-jets obtained from a QCD MC samples with 80 GeV $< \hat{p}_T <$ 120 GeV, (c.f. Table 1). As an example, only the results from TrackCounting b-tagging algorithm are shown.

The measured b-tagging efficiency are shown in Figure 5 for *Ptrel* (left and center) and *System8* (right) in function of the jet p_T and $|\eta|$ (requiring $p_T > 30$ GeV) for the TCL operating point. The measured efficiencies are compared with the true efficiencies obtain from MC. It is clear that

both methods reproduce efficiencies and their dependencies in p_t and $|\eta|$. An initial estimate of the uncertainties expected with both methods are shown in table 2.

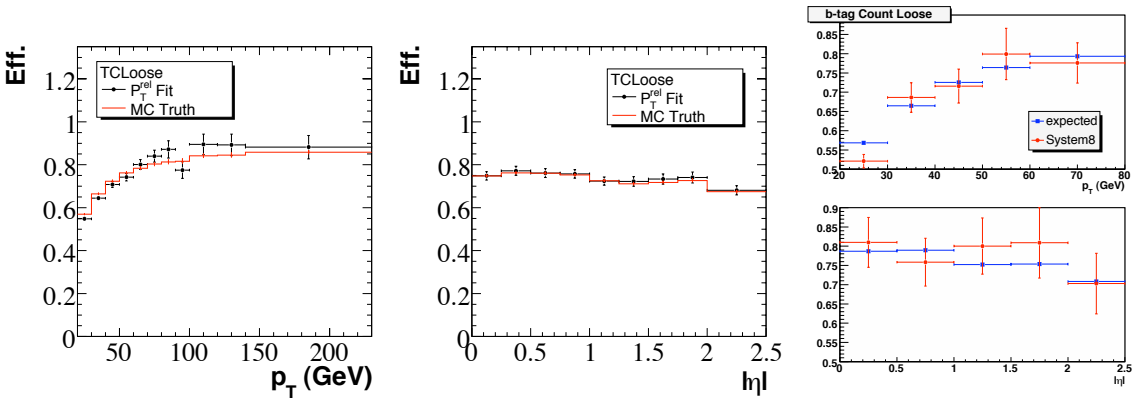


Figure 5: Measured *b*-efficiency from *Ptrel* (left and center) and *System8* (right) in function of the jet p_T and $|\eta|$ (requiring $p_T > 30$ GeV) for the TCL operating point, compared to the MC true (expected) efficiency. The points are shown with statistical uncertainty only.

Operating point	Loose			Medium			Tight		
Luminosity (pb^{-1})	10	100	1000	10	100	1000	10	100	1000
pTrel (n)									
statistics data	1.7	0.5	0.2	2.4	0.8	0.3	2.9	0.9	0.3
Template	15	10	5	15	10	5	15	10	5
Total error (%)	**15**	**10**	**5**	**15**	**10**	**5**	**15**	**10**	**5**
System8									
β	5.8	5.8	2.9	6.3	6.3	3.2	5.7	5.7	2.9
α	0.4	0.4	0.2	0.4	0.4	0.2	0.4	0.4	0.2
κ_b	3.4	3.4	1.7	3.6	3.6	1.8	3.3	3.3	1.7
κ_{cl}	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.1
p_{Trel}	2.8	2.8	2.8	2.9	2.9	2.9	3.0	3.0	3.0
statistics MC	2.3	2.3	2.3	2.6	2.6	2.6	2.7	2.7	2.7
statistics data	7.2	2.3	0.7	8.4	2.6	0.8	8.7	2.7	0.9
Total error (%)	**10.5**	**8.0**	**6.4**	**11.8**	**8.6**	**5.4**	**11.6**	**8.2**	**5.3**

Table 2: Summary of uncertainties expected for *b*-tagging efficiencies measured with the Ptrel and System8 methods for different luminosity scenarios for the TrackCounting b-tagging algorithm.

In the case of the *negative tag method* to measure mistag rate, $\varepsilon^{\text{MC}}_{\text{light}}$ and ε^{MC}_{-} are presented in figure 6 as a function of the jet p_T and $|\eta|$ for the TrackCounting medium operating point. A positive tag veto is defined as a negative tag jet that is rejected if it has any track with $|IP/\sigma_{IP}| > 4$. The increase with p_T is correlated with the increase of track multiplicity in jets. The decrease at high $|\eta|$ is related to the reduced tracker acceptance and smaller tracking efficiency in the forward region. The R_{light} value is about 2.1 for jets with $p_T > 50$ GeV. One can also notice that in this p_T range, the ratio of the mistag efficiencies between uds and gluon jets is about 0.6 ± 0.1 for all operating points. A summary of uncertainties expected for light mistag rate measured with this method is shown in Table 3.

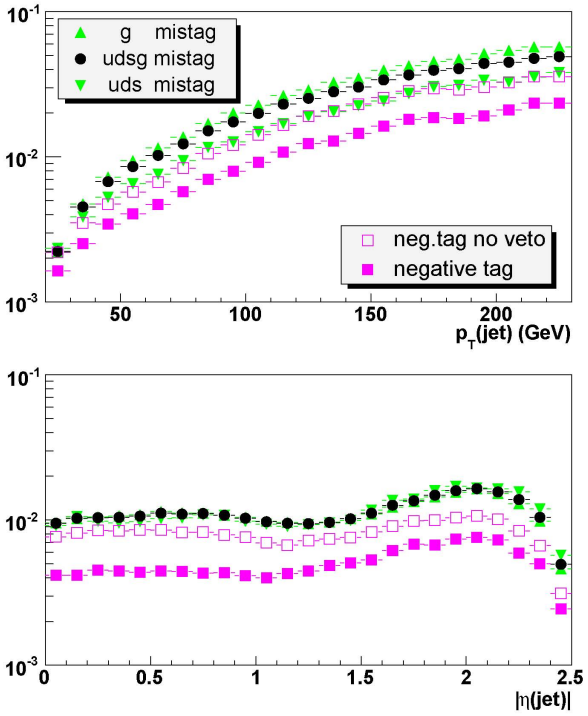


Figure 6: Mistag efficiency and negative tag rate as a function of the jet(upper plot) p_T and (lower plot) $|\eta|$: (full dots) udsg mistag efficiency and (full squares) udsg+c+b negative tag rate, also shown are (triangles) the tagging efficiencies for uds and g-jets separately and (open squares) the negative tag rate if no postive tag veto is applied. Jets from the QCD MC are tagged with the TrackCounting medium operating point.

Operating point	Loose			Medium			Tight		
Luminosity (pb^{-1})	10	100	1000	10	100	1000	10	100	1000
b fraction	1.4	1.4	0.6	0.8	0.8	0.3	1.2	1.2	0.5
c fraction	0.8	0.8	0.3	0.7	0.7	0.3	1.3	1.3	0.5
g fraction	0.8	0.8	0.4	1.4	1.4	0.7	2.3	2.3	1.2
V0 fraction	1.4	1.4	0.7	3.6	3.6	1.8	4.6	4.6	2.3
other displaced processes	1.4	1.4	0.7	3.6	3.6	1.8	4.6	4.6	2.3
IP sign flip	0.7	0.3	0.2	4.5	1.9	1.4	24.0	10.2	7.6
statistics MC	0.1	0.1	0.1	0.4	0.4	0.4	1.2	1.2	1.2
statistics data	0.4	0.1	-	1.6	0.5	0.2	5.5	1.7	0.6
sampling	2.0	2.0	2.0	5.0	5.0	5.0	13.0	13.0	13.0
Total error (%)	**3.4**	**3.4**	**2.4**	**8.8**	**7.6**	**5.9**	**28.7**	**18.1**	**15.5**

Table 3: Summary of uncertainties expected for light mistag rate measured with the negative tag method for different luminosity scenarios for the TrackCounting b-tagging algorithm.

5 Conclusions

Five methods were presented to measure b-tagging efficiency and mistag rate from collider data. However, all methods rely to some level on MC information.

The robustness of each method depends on their sensitivity to the amount of data and the way simulated information is used by them. This is why it is important to develop several strategies to take advantage of their complementary features.

In the case of *Ptrel*, *System8* and *Negative tag* methods the studies are advanced enough to provide initial estimations of their systematic and statistical uncertainties. However for $t\bar{t}$ based methods the initial proof of principle is underway. More work is needed before to provide a reliable set of initial estimates.

References

[1] The CMS collaboration. The CMS experiment at the CERN LHC. JINST 3:S08004,2008.

[2] The CMS collaboration. CMS physics Technical Design Report: Volume I (PTDR1), Detector Performace and Software CERN-LHCC-2006-001, CMS-TDR-008-1, section 12.2.

[3] The CMS collaboration. Performance Measurement of b-tagging Algorithms Using Data containing Muons within Jets. CMS PAS BTV-07-001.

[4] The CMS collaboration. Measuring uds mistag rate of b tag using negative tags. CMS PAS BTV-07-002.

[5] The CDF experiment at Fermilab, http://www-cdf.fnal.gov.

[6] The D0 experiment at Fermilab, http://www-d0.fnal.gov.

[7] The CMS collaboration. CMS physics Technical Design Report: Volume I (PTDR1), Detector Performace and Software CERN-LHCC-2006-001, CMS-TDR-008-1, section 1.2.

[8] The CMS collaboration. CMS Computing TDR (CTDR). CERN-LHCC-2005-023.