DØNote 4890 - Draft v5.0

Performance of the NN b-tagging Tool on Pass 2 p14 Data

Tim Scanlon[1]

Miruna Anastasoaie[2]

1. Imperial College London – UK

2. Radboud University Nijmegen/NIKHEF – The Netherlands

November 30, 2005

Abstract

The performance of the neural net tagger has been measured and tested on the p14 pass 2 corrected data sample. The b-efficiency and fake rates have been calculated in 2D η and p_T parameterisations for six operating points with fake rates ranging from 0.3% to 4%. A considerable improvement in performance is seen over the input taggers. For a fixed fake rate improvements in signal efficiency are seen ranging from 50% for the tightest operating point to 15% for the loosest operating point, and fake rates are reduced to between a quarter and a third of their value for a fixed signal efficiency.

1 Introduction

A Neural Network (NN) b-tagging tool was developed and optimised on Monte Carlo (MC)[1]. In the MC studies the NN b-tagging tool demonstrated considerable improvement in its b-tagging performance compared to the constituent input b-taggers. This note covers the study of the NN tagger's performance on data following the b-id group's standard certification procedure[2] and using the techniques developed and employed in the JLIP[3] and SVT[4] pass 2 certifications.

The tracking in the p14 MC simulations of the DØ detector is not an entirely realistic simulation of real data. The tracking in the MC simulations is over optimistic in both the quality and number of tracks found. This results in the p14 MC b-tagging performance being over estimated compared to data. Therefore the tagging results found from running the b-tagging tools directly on the p14 MC samples can not be used as a realistic estimate of the actual b-tagging performance.

In order to use any track based b-tagging tool in p14 MC the b-tagging performance needs to be carefully measured in data and MC. Understanding the performance allows four functions to be defined which, when applied to MC, make the tagging more accurately describe the performance on data. These are:

Scale Factor (SF) - The factor by which the b and c MC tagging efficiency has to be multiplied to obtain a data tagging efficiency.

TRF_b **-** The efficiency to tag a b-jet in data.

TRF_c **-** The efficiency to tag a c-jet in data.

Fake Tag Rate (FTR) - The fake rate for a light jet (also sometimes referred to as the light jet efficiency).

These functions are parameterised in terms of the p_T and η of the jets. Future certifications will possibly also consider further parameterising the tagging rates in terms of luminosity, the z co-ordinate of the primary vertex and the number of primary vertices. These functions along with their errors are determined in the certification of the tagger.

The p14 MC/data difference could also lead to further problems for a multivariate technique such as the NN tagger. The NN tagger was trained on MC due to the difficulties of obtaining a pure enough sample of either b-jets or uds-jets. This could result in the performance of the NN tagger on data being intrinsically different to MC if the variables were not described well in MC. Therefore the NN input variables were also carefully examined in MC and data.

Section 2 covers the samples used in these studies. Section 3 covers the comparison between the NN input variables and NN output between data and MC. Section 4 covers the measurement of the b-efficiency on real data. Section 5 covers the measurement of the fake rate in real data.

1.1 Neural Network Tagger

The pass2 NN tagger took seven input variables from the JLIP, SVT and CSIP taggers as outlined in Table 1. The other NN specifications are outlined in Table 2.

Rank	Variable	Description
1	SVT_{SL} DLS	Decay Length Significance of the Secondary Vertex (SV)
2	CSIP Comb	Weighted combination of the tracks' IP Significances
3	JLIP Prob	Probability that the jet originates from the PV
4	SVT_{SL} χ^2_{dof}	Chi Square per degree of freedom of the SV
5	SVT_L N_{Tracks}	Number of tracks used to reconstruct the SV
6	SVT_{SL} Mass	Mass of the SV
7	SVT_{SL} Num	Number of SV found in the jet

Table 1: NN input variables ranked in order of power.

Parameter	Value
NN structure	7:24:1:1
Jet input selection cuts (failure results in NN output of 0)	SVT_{SL} DLS > 2.5 or JLIP Prob < 0.02 or CSIP Comb > 8
Number of training epochs	1000
Number of b-jets used in training (after jet selection cuts)	270,000 (220,000)
Number of fake jets used in training (after jet selection cuts)	470,000 (40,000)

Table 2: NN parameters.

1.2 Operating Points

The NN tagger had its performance evaluated for six operating points. The six operating points were chosen to give FTRs on data close to those outlined in Table 3.

Name	Tight	Medium	Loose	L2	L3	L4
NN Cut	> 0.775	> 0.65	> 0.5	> 0.325	> 0.25	> 0.2
Fake Rate	0.3	0.5	1.0	2.0	3.0	4.0

Table 3: The NN tagger's operating points with their corresponding design fake rates.

1.3 Negative and Positive Tags

The majority of fake tags attributed to *uds*-jets are due to resolution effects affecting the reconstruction of the PV and tracks. The impact parameters (IP) of such tracks should be roughly symmetrically distributed around zero (there will actually be slightly more with positive IP due to long lived *uds* particles), whereas tracks from b and c particles will predominantly have positive IPs due to their long lifetimes.

Taggers can take advantage of the symmetrical distribution of IPs of tracks from fake tags and the asymmetrical distributions of IPs from tracks from heavy flavour by constructing two types of tag. Tags constructed from the negative IP tracking component will have very little contribution from heavy flavour and will therefore be a good description of the fake rate in the positive tagging component. The positive IP tracking component will contain the vast majority of the heavy flavour signal with only approximately half the component of fake tags.

Each tagger provided negative tag (NT) and positive tag (PT) results for each jet. The definitions of what constituted a NT result varied for each tagger and are detailed below.

CSIP NT - Calculated from tracks which had a negative IP significance with respect to the primary vertex.

JLIP NT - Calculated from tracks which had a negative IP significance with respect to the primary vertex.

SVT NT - Secondary vertices (SV) which had a negative decay length and a $dR < 0.5$ with respect to the jet.

All jets had a NT and a PT result for each of the taggers. For the NN tagger a NT result was defined as the output from the NN when all of the NN inputs from the corresponding input taggers were from the NT results for the jet.

1.4 Away Tag

An away tag is an indirect method of tagging a jet without actually performing any cuts on the jet and thereby biasing the jet sample. A jet was defined as having an away tag if it was taggable and in an event with only one other taggable jet which had passed some tagging criteria. A jet which was required to have an away tag would have had its probability of being heavy flavour enhanced.

1.5 Calorimeter Regions

It is sometimes useful when undertaking performance studies to consider the DØ detector in three distinct η regions which have different detector responses and tracking capabilities: the Central Calorimeter (CC), Inter-Cryostat Region (ICR) and the End Cap (EC) as outlined in Table 4.

Calorimeter Region	**Physical η**
Central Calorimeter (CC)	$0 < \|\eta\| < 1.2$
Inter-Cryostat Region (ICR)	$1.2 < \|\eta\| < 1.8$
End Cap (EC)	$1.8 < \|\eta\| < 2.4$

Table 4: Calorimeter regions (which have different detector responses) in the DØ detector.

1.6 Statistical Errors

Statistical errors were accounted for in the certification by calculating +1 and -1 sigma curves which represented the 1 sigma deviation from the central fit value. The deviation of the fit from the central value were calculated at each point of the fit function by

$$\Delta f = \sum_{i}^{n} \sum_{j}^{n} E_{ij} \frac{\partial f}{\partial p_i} \frac{\partial f}{\partial p_j} \tag{1}$$

where Δf was the change in the central fit value, E_{ij} was the error matrix for the parameters, i and j relate to the fit parameters, n was the number of fit parameters and the derivatives were of the function with respect to the appropriate fit parameters.

2 Data and MC Samples

The data samples used were derived from the pass 2 corrected BID, EM1TRK (EM) and QCD skims consisting of data taken from July 2002 to August 2004[5]. The skimming criteria are outlined with the number of events in Table 5. The skims were processed excluding bad luminosity blocks and runs marked as bad in the run quality database. In addition to these cuts the jets from the EM1TRK (EM) skim had a missing E_T ($\not\!\!E_T$) cut of

$\not\!\!E_T < 10$ GeV placed on events and only events from the QCD skim which had passed the *JT_25TT_NG* trigger were selected. A larger skim combining (COMB) all the selected jets in the EM and QCD skims was also created. The p14 pass2 corrected MC samples used are outlined in Table 6.

Skim	**Skim Criteria**	**Events**
EM	1 Electron with $p_T > 4$ Gev	64 M
QCD	Jet Triggers	58 M
COMB	Combination of the QCD and EM skims	122 M
BID	One loose μ $p_T > 4$ GeV inside a 0.7 cone jet	91 M

Table 5: The pass 2 corrected data skims used in the certification of the NN tagger.

Sample	**Number of Events**
$t\bar{t}$	390,000
$Z \rightarrow b\bar{b}$	200,000
$Z \rightarrow b\bar{b}$ with μ	100,000
$Z \rightarrow c\bar{c}$	210,000
$Z \rightarrow c\bar{c}$ with μ	100,000
$Z \rightarrow q\bar{q}$	250,000
QCD $p_T = 20-40, 40-80, 80-160, 160-320$	1,160,000
QCD $b\bar{b}$ $p_T = 20-40, 40-80, 80-160, 160-320$	570,000
QCD $c\bar{c}$ $p_T = 20-40, 40-80, 80-160, 160-320$	620,000

Table 6: Pass 2 corrected MC samples used in certification of the NN tagger.

2.1 Data Processing

All the samples were processed using D0Correct v8.2 which corrected (with additional smearing for MC objects) all the electromagnetic, muon, jet and $\not\!\!E_T$ objects with the appropriate corrections and selection cuts[6].

When this work was undertaken the b-tagging values of interest were not easily accessible. To access the variables an altered version of *d0root_example*[7] was used with altered versions of the taggers which allowed all the variables of interest to be accessed and output into root-tuples. This has since been rectified and all the variables are now accessible via the *btags_cert*[8] package. The packages used in the certification are outlined in Table 7.

Package	Tag
bc_csiptagger	v00-00-12 (A)
btags_cert	v00-02-09
d0root_analysis	v00-09-68
d0root_btag	v00-09-80
d0root_csip	v00-00-12 (A)
d0root_jlip	v00-00-21
d0root_tmb	p16cert-br-02

Table 7: Package versions used in certification study. An (A) indicates that a package was altered from the CVS version.

2.2 Physics Objects Selection

The following physics objects were used in this analysis.

PV - A 2-pass probabilistic primary vertex reconstruction algorithm was used[9]. Events were required to have a primary vertex reconstructed from 4 or more tracks with $|z| < 60$ cm.

Track Jets - The standard p14 track jet algorithm was used[10] which formed a track jet from the tracks found within $dr < 0.5$ of a seed track. All tracks used in a track jet must have at least 2 SMT hits and the seed track must have $p_T > 1.0$ GeV. A track jet must consist of at least two tracks.

Jets - Calorimeter jets were reconstructed using a 0.5 cone jet algorithm (JCCB), with fake jets and electron candidates rejected. The standard jet energy scale (JES) correction (without muon corrections) was applied to the jets. Only taggable jets with a corrected $p_T > 15$ Gev were considered in this certification.

Flavoured Jets - In MC events jets were identified as being b-jets if they had a b-hadron within $dR < 0.5$ of the jet. c-Jets had a c-hadron but no b-hadron. uds-Jets or fake jets, which included gluon jets, had no heavy flavour at all present in the jet.

Taggable Jets - The standard p14 taggability definition was used, which defined a taggable jet as a good calorimeter jet matched to a track jet within a cone of $dR < 0.5$. The taggability definition was designed to select a sample of jets with a minimum tracking criteria which were independent from the underlying detector and data conditions. Only taggable jets were used, and only events with at least two taggable jets were considered in the certification.

As all the results were derived in relation to taggable jets the TRFs and SF should only be applied to a selection of taggable jets.

Muons - Medium muons were selected which had a $p_T > 4$ GeV, a global $\chi^2 < 100$ and were matched to a good jet within $dR < 0.5$.

2.3 Taggability Efficiencies

The taggability efficiencies are shown in Fig. 1 for the QCD/EM skims and the QCD MC sample. Differences between taggability in the MC and data samples existed due to the higher tracking efficiency in MC. The difference between the EM1TRK and QCD data samples were due to the requirement of a high p_T track in the EM skim.

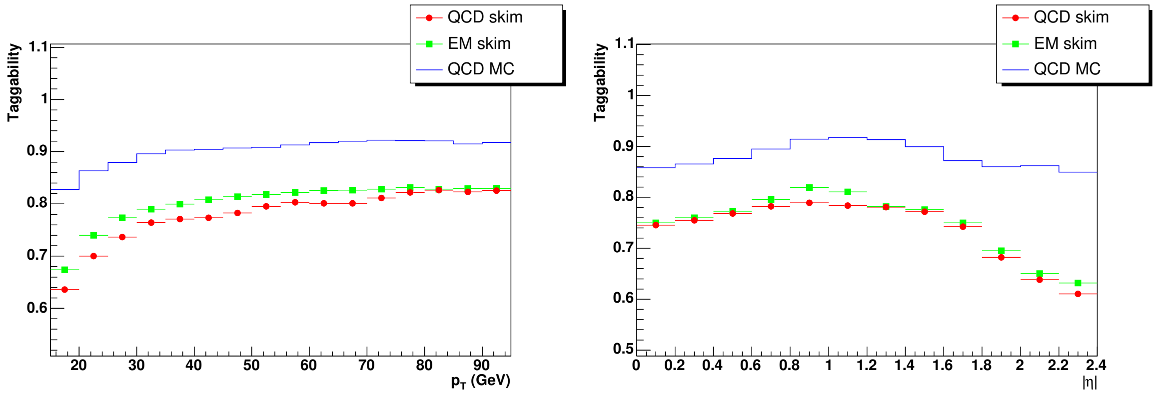

Figure 1: Taggability as a function of p_T (left) and η (right) for the QCD and EM skims and the QCD MC sample.

2.4 Jet Samples

The number of jets available in the MC samples after data processing and physics object and jet cut selections are outlined in Table 8. The different energy QCD samples were merged into a continuous p_T sample by weighting the different energy samples via the fall off in their jet p_T spectrums. Larger b and c-jets samples were created by combining all the appropriate jets from the different production channels.

The number of jets available in the data samples after data processing and physics object and jet cuts selections are outlined in Table 9.

Flavour	**Sample**	**Number of Jets**
b	$Z \to b\bar{b}$	240,000
	$t\bar{t} \to b$	600,000
	QCD $b\bar{b}$ (Merged)	270,000
	b (Combined)	1,800,000
$b \to \mu$	$Z \to b\bar{b} \to \mu$	54,000
	$t\bar{t} \to b \to \mu$	60,000
	$b \to \mu$ (Combined)	180,000
c	$Z \to c\bar{c}$	260,000
	$t\bar{t} \to c$	169,000
	QCD $c\bar{c}$ (Merged)	300,000
	c (Combined)	1,500,000
$c \to \mu$	$Z \to c\bar{c} \to \mu$	52,000
	$t\bar{t} \to c \to \mu$	7,100
	$c \to \mu$ (Combined)	90,000
uds	$Z \to q\bar{q}$	375,000
	QCD Fake (Merged)	500,000
All	QCD All	2,000,000

Table 8: Number of MC jets of each flavour available in each of the samples after data processing and jet selection.

Sample	**Number of Jets**
BID	24,000,000
EM	12,400,000
QCD	9,700,000
COMB	22,100,000

Table 9: Number of data jets available in each skim after data processing and jet selection.

Variable	Cut Value
JLIP Prob	< 0.002
JLIP $Prob_{Red}$	< 0.02
Tight SVT DLS	> 10
CSIP 3s	> 2
CSIP Mass	> 3.0

Table 10: Away tag cut applied on BID skim, designed to reduce both the *uds* and *c*-jet content in the *b*-enriched sample.

3 Data and MC Comparison

3.1 Variables

As the NN was trained on MC events, but will be applied to both data and MC events, a comparison study was carried out between the input variables in data and MC to ensure good agreement. This required two data samples to be produced, one enriched with *b*-jets, and one enriched with fake jets.

3.1.1 *b*-Jet Input Variables

The data *b* enriched jet sample was obtained from the muonic jets in the BID skim. The purity of the sample was increased without biasing the results by using a SLT [12] and an away tag.

The SLT tag was defined as a muon with a $p_{TRel} > 0.7$ with respect to the jet axis. A SLT tag should not bias the results as the SLT tagger has been shown to be uncorrelated with the other taggers[3][4][11]. The away tag used the cuts outlined in Table 10 which were designed by examining *b*, *c* and fake jet tagging variables in MC and selecting values which maximised the *b*-efficiency relative to the fake and *c*-efficiencies. The efficiency of the away tag was estimated from QCD MC by selecting events which had only two taggable jets present, and then measuring the probability of the fake, *c* and *b*-jets present in those events to have an away tag. This measurement would also take into account the effects of possible gluon splitting to co-linear heavy flavour in the away jet and QCD production of jets of different flavours. Although these efficiencies were measured entirely on MC they should prove sufficient as a rough estimate of the flavour content of the *b*-enriched sample. The away tag efficiencies along with the SLT tagging efficiencies and the estimated flavour content of the *b*-enriched sample are outlined in Table 11. The fact that the absolute *b*-efficiency value was very small was not a problem due to the large data sample available.

The *b*-enriched data sample was compared to the QCD $b\bar{b}$ MC sample with all the jets

Flavour	QCD Content	SLT Eff	Away Tag Eff	b-enriched Content
b	3%	23%	2.4%	96.5%
c	5%	8.4%	0.25%	3.2%
Fake	92%	0.05%	0.24%	0.3%

Table 11: The QCD content (estimated from QCD MC) and the SLT and away tag efficiencies with the estimated flavour content in the b-enriched sample estimated from multiplying the QCD flavour content by the away tag and SLT efficiencies.

required to have passed the NN jet input selection cuts (Table 2 in Sec. 1.1) and is shown in Fig. 2. Comparisons should be made between the shape of the input variables and not their absolute values which were set to be of equal area. Although the b-jet variables were similar the data b-jets clearly had worse tagging results. This was expected due to the MC's over optimistic tagging results and the data sample having *udsc* contamination which would degrade the tagging results seen in the data b-sample with respect to the MC results.

3.1.2 Fake Jet Input Variables

Obtaining a sample of fake jets without any or with minimal heavy flavour contamination was more difficult. Any cuts placed on the jet would bias the comparison and it was not possible to use an away tag method to enrich a sample with fake jets. However, the QCD skim NT results provided a good approximation to a sample of fake jet results. The QCD MC sample suggests heavy flavour fractions of approximately 5% c and 3% b when no tag was applied. Jets were also required to have passed the NN jet input selection cuts (Table 2 in Sec. 1.1) which increased the fraction of heavy jets in the sample to $\sim 10\%$ c and $\sim 10\%$ b.

The QCD NT data sample was compared to the NT results from the MC QCD All sample and is shown in Fig. 3. Comparisons should be made between the shape of the input variables and not their absolute values which were normalised to be of equal area. There was good agreement between the fake jets in MC and data. The main difference was seen for low JLIP Prob values. The disagreement between data and MC in this region was likely caused by the presence of b and c-jets in the NT samples which would be responsible for the majority of tags at low JLIP Prob. The over optimistic tagging results in MC b and c-jets would then produce the discrepancy seen between the data and MC samples.

3.2 NN Output

The NN outputs for the EM/QCD NT skims and the b-enriched sample compared to their corresponding MC samples are shown in Fig. 4. Comparisons should be made between the shape of the NN output and not the absolute values which were normalised to the number of

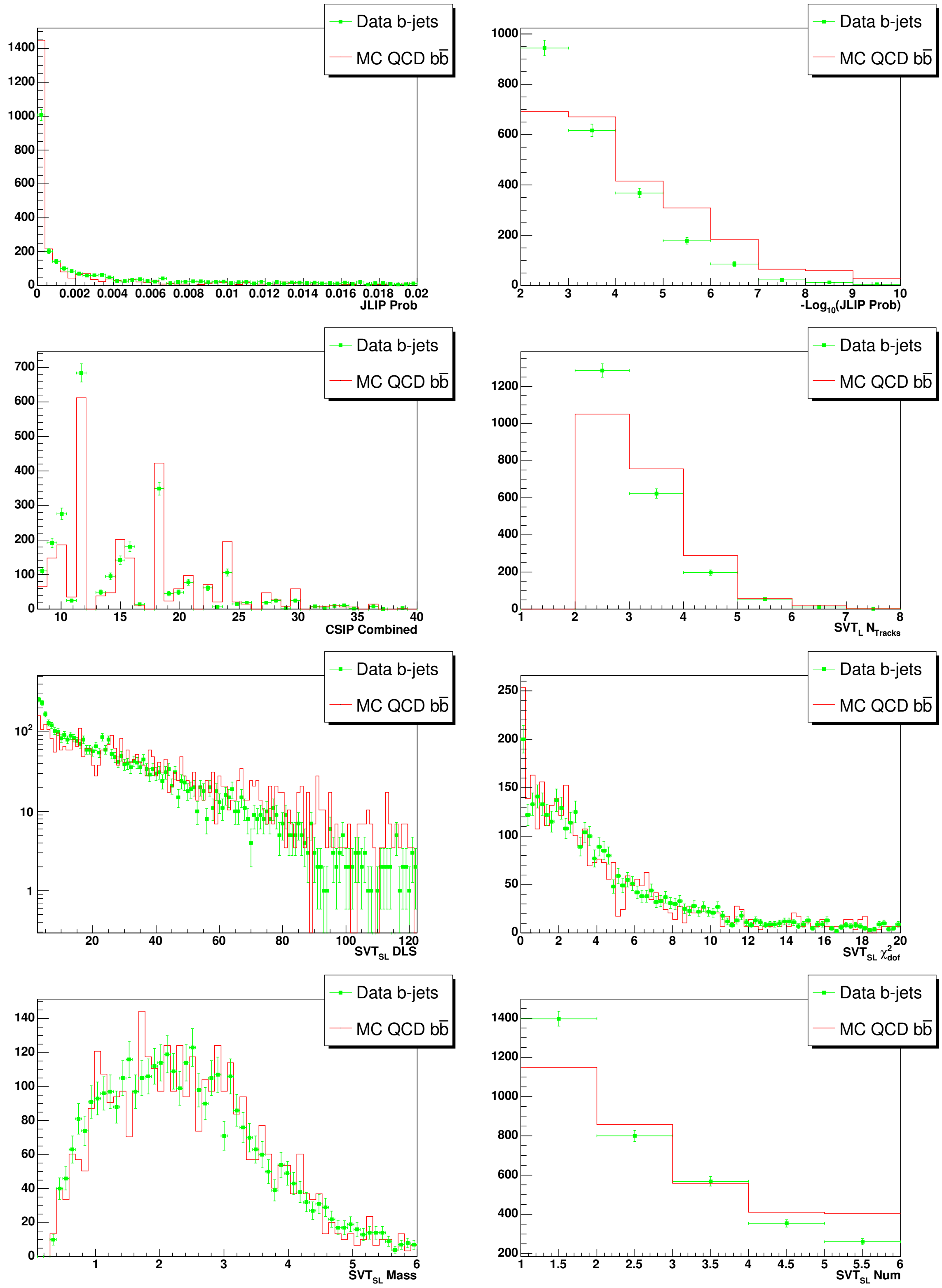

Figure 2: NN input variables in QCD $b\bar{b}$ MC and the b-jet enriched sample from the BID skim after the NN jet input cuts. The variables are JLIP Prob (top left), $-Log_{10}$ (JLIP Prob)(top right), CSIP Comb (2nd left), SVT_L N_{Tracks} (2nd right), SVT_{SL} DLS (3rd left), SVT_{SL} χ^2_{dof} (3rd right), SVT_{SL} Mass (bottom left) and SVT_{SL} (bottom right). Comparisons should be made between the shape of the input variables and not their absolute values which were normalised to have equal areas.

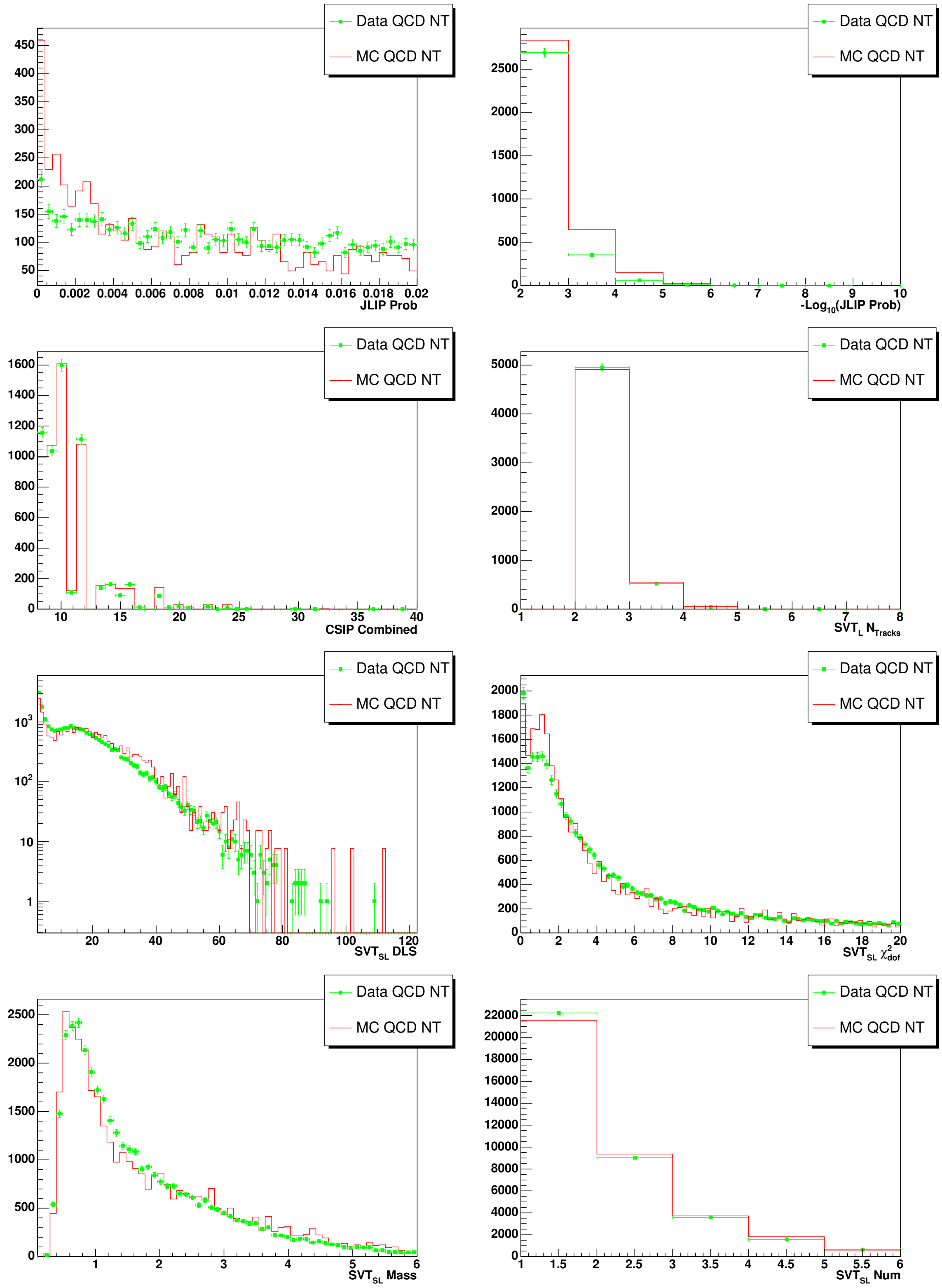

Figure 3: NN input NT variables in QCD MC and the QCD skim after the jet input cuts. The variables are JLIP Prob (top left), $-Log_{10}$ (JLIP Prob) (top right), CSIP Comb (2nd left), SVT_L N_{Tracks} (2nd right), SVT_{SL} DLS (3rd left), SVT_{SL} χ^2_{dof} (3rd right), SVT_{SL} Mass (bottom left) and SVT_{SL} (bottom right). Comparisons should be made between the shape of the input variables and not their absolute values which were normalised to have equal areas.

entries with NN probability > 0. There is good agreement between data and MC with any differences explained by the discrepancies seen in the input variable comparisons. The data b-jet NN output had a lower peak due to a combination of contamination from *udsc* jets and the lower tagging efficiency expected in data. The fake jets have very good agreement as would be expected from the good agreement between the data and MC variables. Figure 5 shows a comparative plot of the fake and b-jet NN output on data.

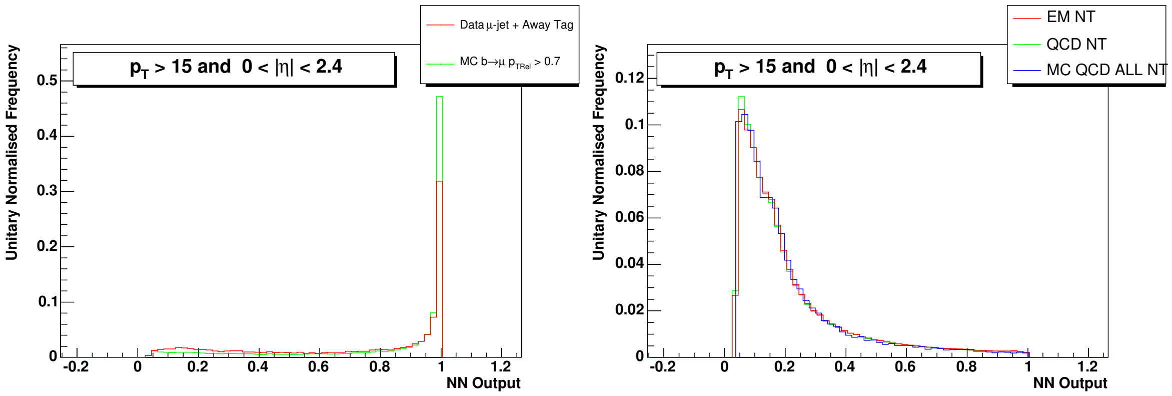

Figure 4: NN output data/MC comparison for the b-enriched skim and QCD $b\bar{b}$ MC (left) and the NT output from the QCD and EM skims compared to the QCD ALL MC sample (right). Comparisons should be made between the shape of the NN output and not the absolute value which was normalised to the number of entries with NN probability > 0.

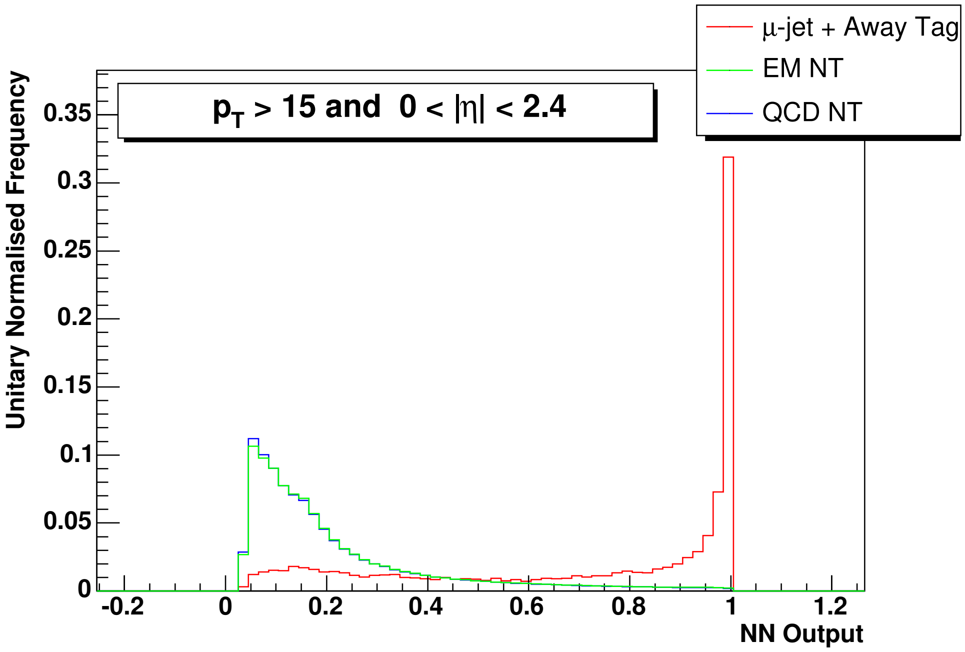

Figure 5: The comparative NN output from the b-enriched BID skim and the QCD and EM NT skims. Comparisons should be made between the shape of the NN output and not the absolute value which was normalised to the number of entries with NN probability > 0.

4 b-Efficiency Measurement

4.1 System8 Method

The b-jet efficiency on data was measured using the System8 (S8) formalism[13]. A system of 8 equations with 8 unknowns was constructed by equating the number of tags found by two different taggers on two different samples. The solutions to the equations included the efficiencies of the two taggers on b-jets.

The S8 equations were symmetrical in nature and in order that the equations were solvable the two taggers needed to be uncorrelated with different b-efficiencies and the two data samples needed to have different b-contents.

The two taggers used were the NN and SLT taggers. A SLT tag was defined as a muonic jet where the muon satisfied a p_{TRel} cut of 0.6 GeV with respect to the jet axis. In the certification of the JLIP and SVT taggers a value of 0.7 GeV was used. A lower SLT cut value was required for the NN tagger to ensure that the NN b-efficiency remained different from the SLT b-efficiency for all operating points.

The SLT efficiency was measured as approximately 70% for a p_{TRel} cut of 0.7 which was approximately the same efficiency as the looser NN operating points. The uncertainty in the S8 solutions for these points increased due to two of the symmetric solutions having similar answers. To reduce this affect a lower SLT operating point was chosen which ensured that the SLT efficiency was always larger than the NN efficiency.

The two data samples used were the BID muonic jet sample and a b-enriched subset consisting of muonic jets with an away tag of JLIP Prob < 0.01.

Correlation coefficients were introduced into the S8 formulation to account for the following effects:

α - Differences in the $udsc$-tagging efficiency between the two samples.

β - Differences in the b-tagging efficiency between the two samples.

κ_b - Correlations between the NN tagger and the SLT tagger on b-jets.

κ_{udsc} - Correlations between the NN tagger and the SLT tagger on $udsc$-jets.

p_{TRel} - Differences between the SLT tagging efficiency on c and uds-jets.

The values of the correlation coefficients were measured using MC samples. The statistical errors on the measurements were used to estimate systematic errors due to the uncertainty with which the correlation coefficients were known. The S8 equations were defined as

$$
\begin{aligned}
n &= n_b + n_{udsc} \\
p &= p_b + p_{udsc} \\
n^{SLT} &= \varepsilon_b^{SLT} n_b + \varepsilon_{udsc}^{SLT} n_{udsc} \\
p^{SLT} &= \varepsilon_b^{SLT} p_b + \varepsilon_{udsc}^{SLT} p_{udsc} \\
n^{NN} &= \varepsilon_b^{NN} n_b + \varepsilon_{udsc}^{NN} n_{udsc} \\
p^{NN} &= \beta \varepsilon_b^{NN} p_b + \alpha \varepsilon_{udsc}^{NN} p_{udsc} \\
n^{SLT,NN} &= \kappa_b \varepsilon_b^{SLT} \varepsilon_b^{NN} n_b + \kappa_{udsc} \varepsilon_{udsc}^{SLT} \varepsilon_{udsc}^{NN} n_{udsc} \\
p^{SLT,NN} &= \kappa_b \beta \varepsilon_b^{SLT} \varepsilon_b^{NN} p_b + \kappa_{udsc} \alpha \varepsilon_{udsc}^{SLT} \varepsilon_{udsc}^{NN} p_{udsc}
\end{aligned}
$$

where n was the number of jets in the muonic jet sample, p was the number in the b-enriched sample and ε was the efficiency of the tagger. The superscripts referred to the NN and SLT taggers and the subscripts referred to the flavour of the jets, b or $udsc$.

4.2 Measurement of the Correlation Coefficients

4.2.1 β

β was introduced to represent differences between the b-efficiencies in the two different samples. β was calculated from the ratio of the muonic b-jet to the muonic b-jet with an away tag tagging efficiencies. The value determined from the combined $b \rightarrow \mu$ MC sample is shown in Fig. 6. The β error was calculated taking into account the correlated numbers of events. The systematic error was estimated by varying the value of β within its statistical error.

4.2.2 κ

κ_b and κ_{udsc} which represented the correlations between the NN and SLT taggers on muonic jets were determined from the following relationship

$$\kappa_{b,udsc} = \frac{\varepsilon_{b,udsc}^{NN} \times \varepsilon_{b,udsc}^{SLT}}{\varepsilon_{b,udsc}^{NN\&SLT}} \tag{2}$$

where ε was the efficiency, the subscripts referred to the jet flavour, the superscripts referred to the taggers and the ampersand to both tags being present.

κ_b was measured in the combined $b \rightarrow \mu$ MC sample and is shown in Fig. 7. The κ error was calculated taking into account the correlated numbers of events. A linear dependency was observed in p_T and a straight line was used to model the coefficient. The error on the fit

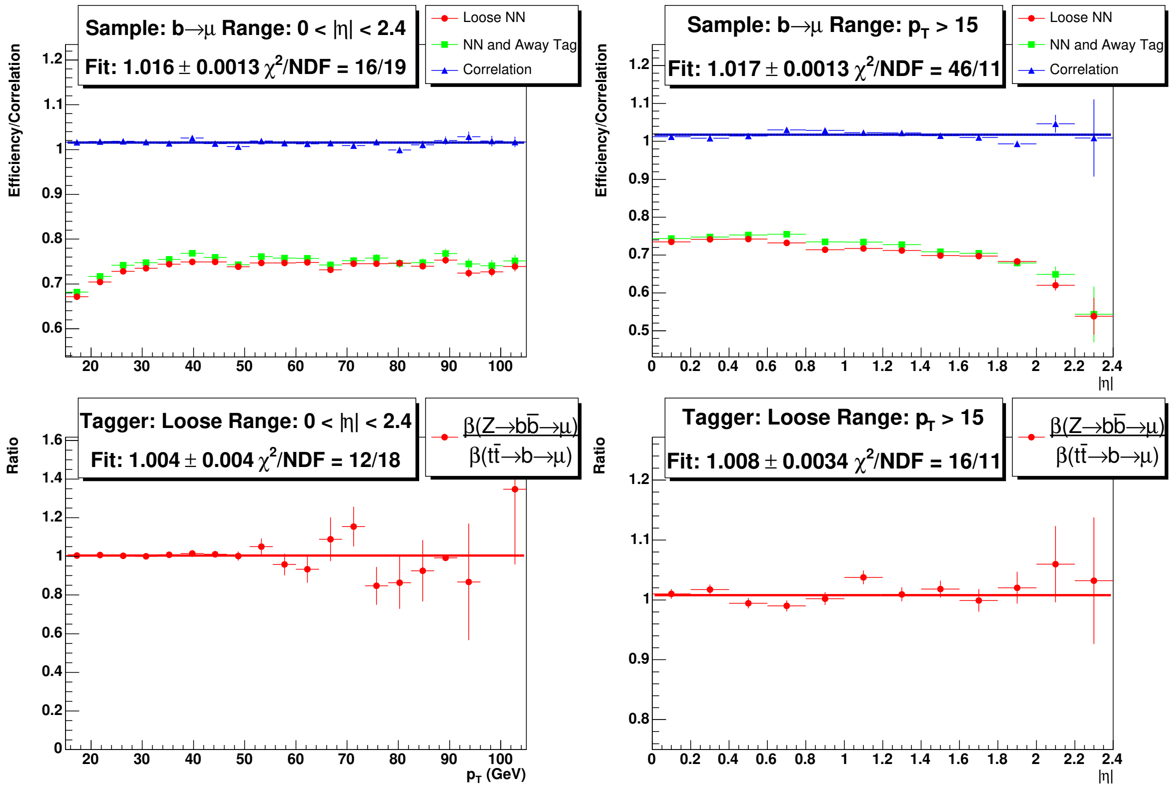

Figure 6: β in the $b \to \mu$ combined sample as a function of jet p_T (left) and η (right). The black dotted line represents the error on the straight line fit. The bottom plots show the ratio between β found in the $Z \to b\bar{b} \to \mu$ and $t\bar{t} \to b \to \mu$ samples in p_T (left) and η (right)

was determined from the full covariant error matrix and was therefore also a function of p_T. As the fit value and error were a function of p_T the systematic error due to the uncertainty in the measurement of κ_b also became a function of p_T.

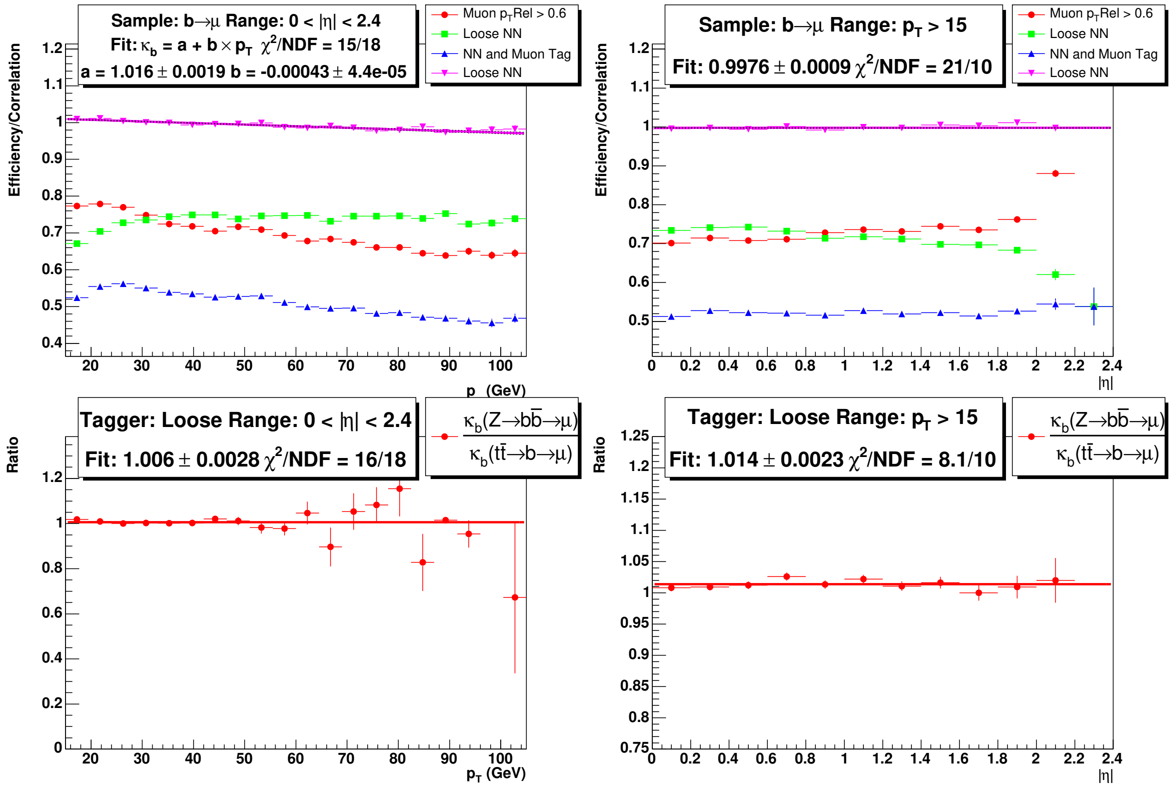

Figure 7: Correlation coefficient κ_b between the NN and the SLT tagger as measured in the combined $b \rightarrow \mu$ MC sample in jet p_T (left) and η (right) projections. The black dotted line represents the error on the fit calculated from the full covariant fit error matrix. The bottom plots show the ratio between κ_b as found in the $Z \rightarrow b\bar{b} \rightarrow \mu$ and $t\bar{t} \rightarrow b \rightarrow \mu$ MC samples in p_T (left) and η (right).

κ_{udsc} was measured using the combined $c \rightarrow \mu$ MC sample and is shown in Fig. 8. The κ error was calculated taking into account the correlated numbers of events. Compared to the measurement of κ_b the measurement of κ_{udsc} had a larger uncertainty associated with it due to the lower statistics available. To represent the increased uncertainty the statistical error on κ_{udsc} was increased to 3 σ.

4.2.3 α

The α correlation coefficient, which represented the differences between the tagging efficiency on $udsc$ jets in the two samples, was not calculated from MC. α was set equal to 1 and the error due to the uncertainty in the value was measured by varying the value from 0.2 to 1.8.

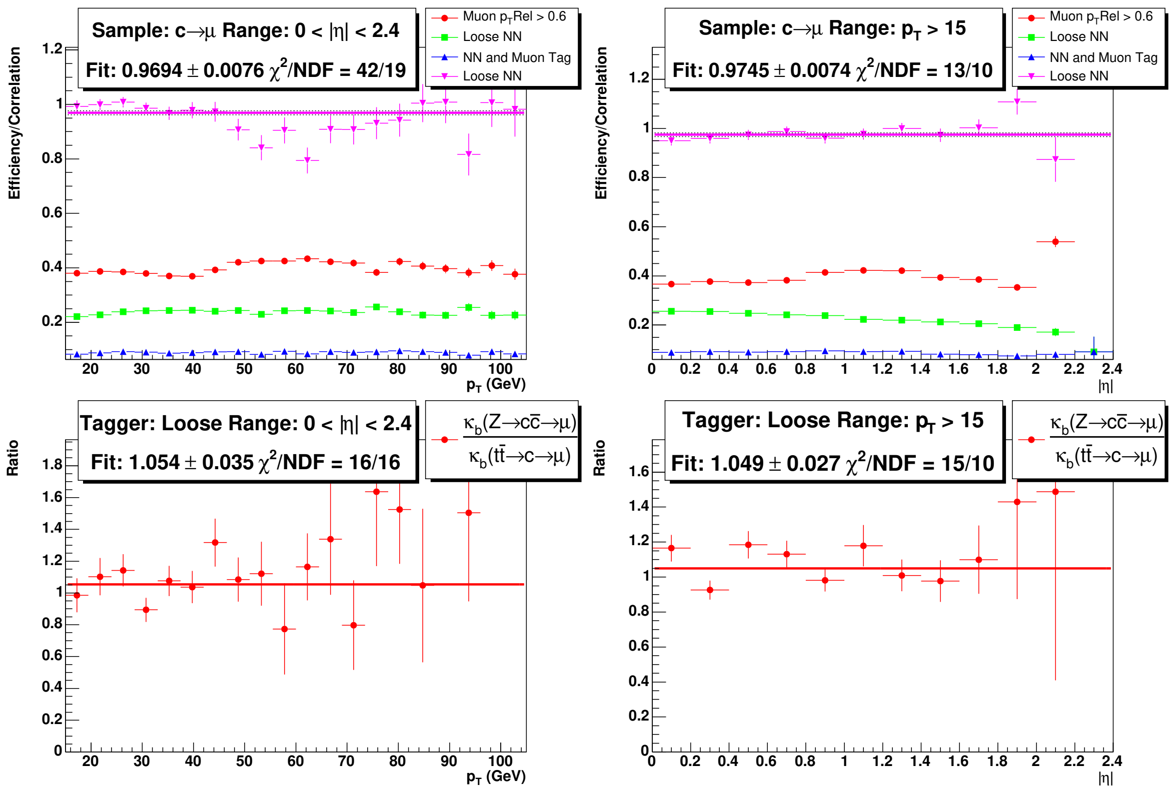

Figure 8: Correlation coefficient κ_{udsc} between the NN and the SLT tagger as measured in the combined $c \to \mu$ MC sample in jet p_T (left) and η (right) projections. The black dotted line represents the error on the fit. The bottom plots show the ratio between κ_{udsc} as found in the $Z \to c\bar{c} \to \mu$ and $t\bar{t} \to c \to \mu$ MC samples in p_T (left) and η (right).

4.2.4 p_{TRel}

The p_{TRel} correlation coefficient, due to the differences between SLT tagging efficiency on c and $udsc$-jets, was not represented in the S8 equations as a correlation coefficient. Instead the uncertainty caused by this assumption was evaluated by recalculating the S8 efficiency at different SLT operating points. The SLT p_{TRel} cut was varied from 0.4 to 0.8. The values of κ_b and κ_{cl} were dependent on the value of the SLT p_{TRel} cut therefore both these values were recalculated at the different SLT operating points.

4.2.5 Correlation Coefficient Values

The values of the correlation coefficients are given in Tables 12 and 13. The values and errors were consistent between the η and p_T projections. The κ_b value and error were p_T dependent so the value was evaluated at the average jet p_T value for the sample. When solving the S8 equations in various bins in p_T and η the S8 correlation coefficients were evaluated for the average jet p_T value in that bin.

	L4	**L3**	**L2**
α	1 ± 0.8	1 ± 0.8	1 ± 0.8
β	1.01 ± 0.00101	1.01 ± 0.00107	1.01 ± 0.00115
κ_b	0.996 ± 0.000695	0.997 ± 0.000735	0.997 ± 0.00079
κ_{cl}	0.98 ± 0.0161	0.98 ± 0.0172	0.981 ± 0.0188
p_{TRel} (GeV)	0.6 ± 0.2	0.6 ± 0.2	0.6 ± 0.2
	Loose	**Medium**	**Tight**
α	1 ± 0.8	1 ± 0.8	1 ± 0.8
β	1.02 ± 0.00129	1.02 ± 0.00143	1.02 ± 0.00156
κ_b	0.998 ± 0.000895	0.998 ± 0.000988	0.998 ± 0.00109
κ_{cl}	0.975 ± 0.0221	0.959 ± 0.0251	0.959 ± 0.0284
p_{TRel} (GeV)	0.6 ± 0.2	0.6 ± 0.2	0.6 ± 0.2

Table 12: The S8 correlation coefficients with associated errors as derived from the MC samples in the p_T projection evaluated at the point $p_T = 39$ GeV which was the average p_T of the sample.

4.3 System8 Efficiency and Systematic Errors

TMinuit[15] was used to solve the S8 equations, and the calculation was carried out on the entire sample and in several p_T and η regions to create a profile over the p_T and η phase space. The systematic uncertainties in the calculated b-efficiency, due to the errors in

	L4	**L3**	**L2**
α	1 ± 0.8	1 ± 0.8	1 ± 0.8
β	1.01 ± 0.000985	1.01 ± 0.00104	1.02 ± 0.00111
κ_b	0.998 ± 0.000702	0.998 ± 0.000742	0.999 ± 0.000797
κ_{cl}	0.978 ± 0.0166	0.978 ± 0.0178	0.979 ± 0.0195
p_{TRel} (GeV)	0.6 ± 0.2	0.6 ± 0.2	0.6 ± 0.2
	Loose	**Medium**	**Tight**
α	1 ± 0.8	1 ± 0.8	1 ± 0.8
β	1.02 ± 0.00125	1.02 ± 0.00138	1.02 ± 0.00151
κ_b	0.999 ± 0.000903	1 ± 0.000996	1 ± 0.00109
κ_{cl}	0.969 ± 0.0229	0.952 ± 0.0259	0.949 ± 0.0294
p_{TRel} (GeV)	0.6 ± 0.2	0.6 ± 0.2	0.6 ± 0.2

Table 13: The S8 correlation coefficients with associated errors as derived from the MC samples in the η projection.

the correlation coefficients, were determined by varying the coefficients by their aforementioned errors and resolving the equations. The relative differences in efficiencies between the solutions provided the systematic errors.

The S8 efficiencies along with the systematic errors for the whole BID muonic jet sample is shown in Table 14. The efficiency is seen to vary from 50% for the tightest operating point to 70% for the loosest operating point. The relative systematic errors, when calculated for the entire sample, vary from 1% for the tightest to 2% for the loosest operating point.

Error	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
Efficiency	71.4%	68.7%	65.3%	59.0%	54.0%	48.6%
α	1.72%	1.14%	0.817%	0.732%	0.854%	0.693%
β	0.128%	0.134%	0.144%	0.162%	0.182%	0.196%
κ_b	0.264%	0.278%	0.298%	0.339%	0.378%	0.412%
κ_{udsc}	0.326%	0.307%	0.286%	0.254%	0.237%	0.219%
p_{TRel}	0.708%	0.78%	0.691%	0.636%	0.655%	0.506%
Total	1.91%	1.45%	1.16%	1.07%	1.18%	0.992%

Table 14: Relative systematic uncertainties on the data b-jet efficiency calculated on the entire BID skim. The total systematic uncertainty was determined by adding the individual uncertainties in quadrature.

There are a couple of different ways of assigning an error due to the uncertainties in the correlation coefficients. As carried out in the JLIP and SVT certifications and above the systematic errors can be determined by solving the S8 equations for the entire sample. Another method is to measure the systematic error separately for each region of phase space in which the S8 equations were used.

In this certification the later method was employed. Primarily this was due to κ_b, as it's error varied as a function of p_T. Therefore the correct way to measure the systematic error due to the variation κ_b's error was to calculate the S8 systematic error in each p_T bin. This method was also deemed to be more accurate as the uncertainty in the S8 calculation was measured in each region of phase space. As each of the regions had different solutions there was no guarantee that the S8 errors would be the same in each region and would not be dependent on either p_T or η. Any such dependency would be overlooked if the systematic error was only evaluated for the entire sample.

The S8 errors were evaluated for each bin in the p_T and η projections and are shown in Fig. 9 for the L2 and Loose operating points. Evaluating the S8 errors in each bin produced overall larger systematic errors than calculating the S8 errors for the entire sample. The S8 errors appeared to have very little η dependence apart from at high η where the p_{TRel} uncertainty was substantially higher. The systematic errors demonstrated a downwards trend with increasing p_T. The constant fit for the total S8 systematic errors in both projections compared to the that for the entire sample also suggested that only using the S8 error calculated from the entire sample would have been an underestimate of the systematic uncertainty.

The S8 error calculated for each bin was combined quadratically with the Minuit fit error on the b-efficiency. The combined error was assigned as the total uncertainty with which the b-efficiency was known in each bin.

4.4 b-tagging and Scale Factor Parameterisation

The efficiencies in p_T were parameterised using

$$\varepsilon(p_T) = \frac{c}{1 + ae^{-bp_T}} \tag{3}$$

where ε was the efficiency and a, b, and c were constants to be determined. The efficiency was parameterised in η using

$$\varepsilon(\eta) = d + e\eta + f\eta^2 + g\eta^3 + h\eta^4 \tag{4}$$

where ε was the efficiency and d, e, f, g and h were constants to be determined. To create a 2D parameterisation it was assumed that the efficiency could be factorised into η and p_T components. A 2D parameterisation was given by the product of the two projections scaled by the total efficiency of the sample

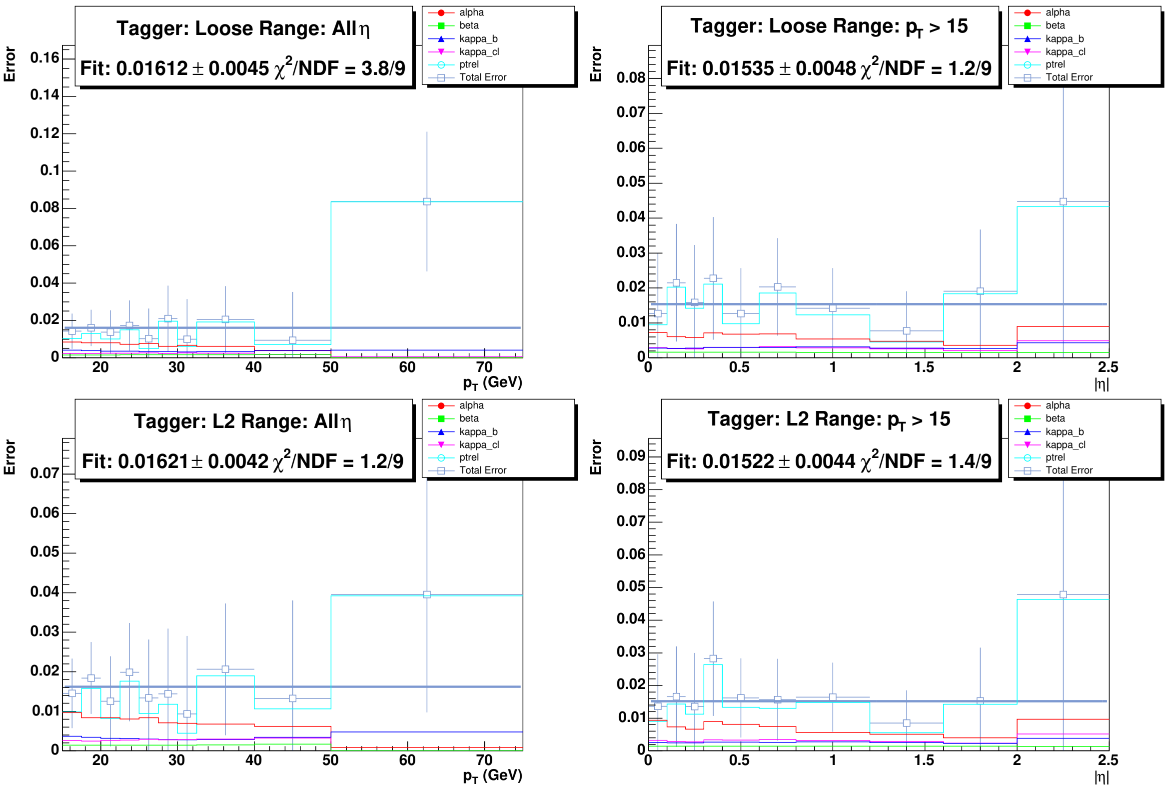

Figure 9: S8 systematic error calculated for each bin in p_T (left) and η (right) for the Loose operating point (top) and the L2 operating point (bottom). The other operating points show similar trends.

$$\varepsilon(p_T,\eta) = \frac{1}{\varepsilon^{All}}(\frac{c}{1+ae^{-bp_T}}) \times (d + e\eta + f\eta^2 + g\eta^3 + h\eta^4) \quad (5)$$

where ε^{All} was the efficiency on the entire sample. The efficiencies calculated from the S8 equations are shown in Figs. 10 and 11 along with the combined MC $b \to \mu$ sample efficiency and the data/MC SF produced by dividing the data efficiency by the MC efficiency.

4.5 Data/MC Scale Factor

The data/MC SF was measured in the previous section from the (Data muonic b-efficiency)/(MC muonic b-efficiency). The SF measures the effect on the tagging rate caused by the differences in tracking between data and MC. The profile of the SF in p_T and η allows tagging rates to be corrected over the full p_T and η phase space.

It was assumed that the SF could be applied to any MC tagging efficiency to correct from a MC to a data tagging efficiency. In making this assumption it was assumed that the SF, which was measured only on muonic jets, was applicable to all jets. The validity of this assumption was not tested and it was assumed that this affect was small in relation to the other errors. Further studies were planned into this systematic error in the next round of certification.

4.6 Tag Rate Functions

The inclusive b-jet data TRF was calculated by multiplying the inclusive b-jet MC TRF (produced from the combined inclusive b-jet MC sample) by the SF. The inclusive b-jet data and MC TRFs (TRF_b) are shown in Figs. 12 and 13.

A similar procedure was used to determine the inclusive c-jet data TRF (TRF_c). The inclusive c-jet MC TRF (produced from the combined inclusive c MC sample) was multiplied by the SF to give the TRF_c as shown in Figs. 14 and 15.

4.7 Systematic Uncertainties

4.7.1 Parameterisation Systematic Errors

The parameterisation of the efficiency into η and p_T had a systematic error attributed to it due to the possible lack of factorisation and non-perfect parameterisations. To quantify this effect MC TRFs were produced for the combined b, $b \to \mu$ and c samples using the parameterisations described in the previous section. The TRFs were used to predict the number of tags in the samples in the three detector η regions outlined in Section 3, which

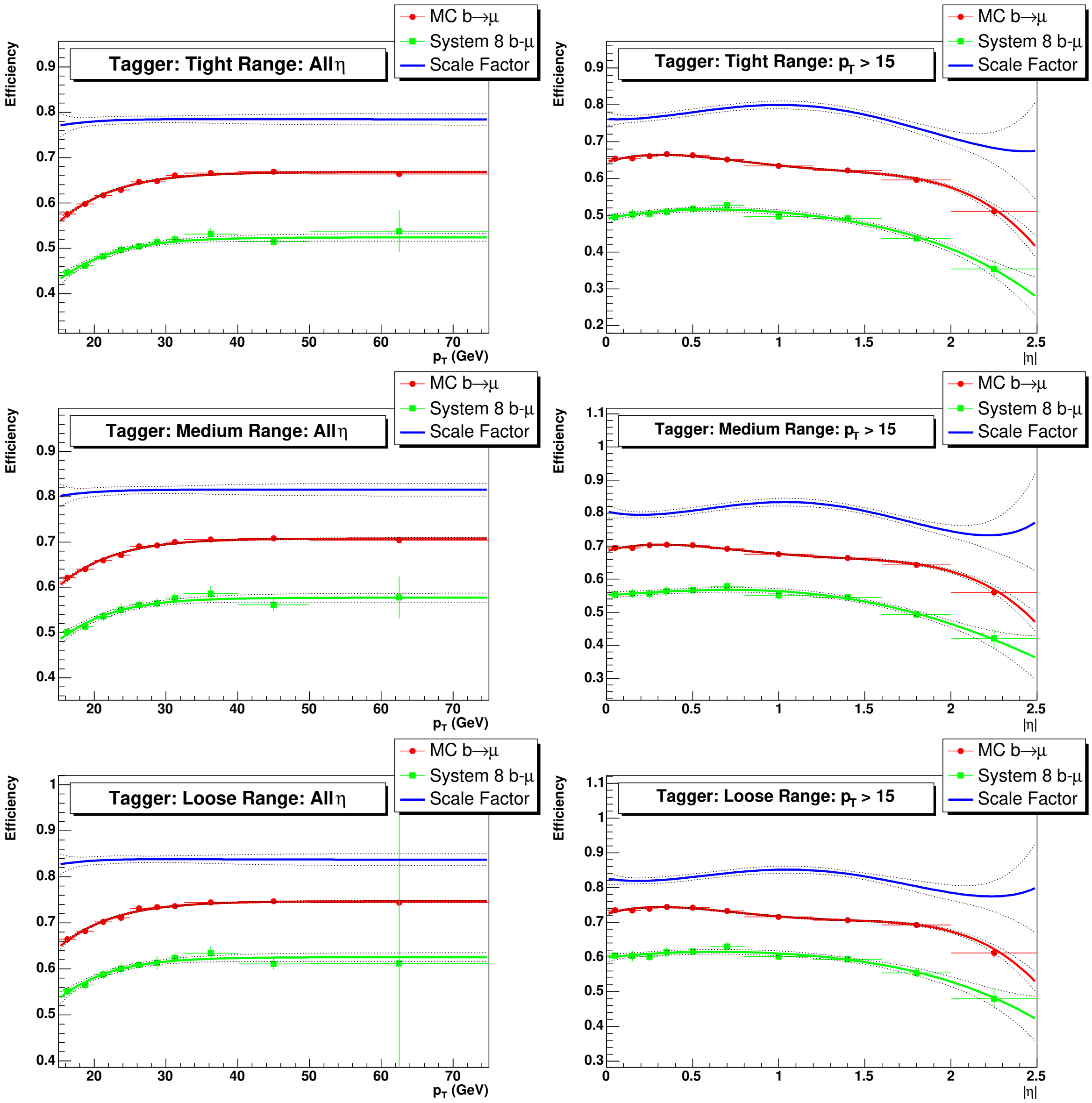

Figure 10: NN tagger muonic b-jet efficiency as a function of p_T (left) and η (right) in the BID μ-jet data sample for the Tight (top), Medium (middle) and Loose (bottom) operating points. The functions used for the parameterisation are outlined in the text and the black dotted curves represent the fit error as calculated from the full covariant error matrix.

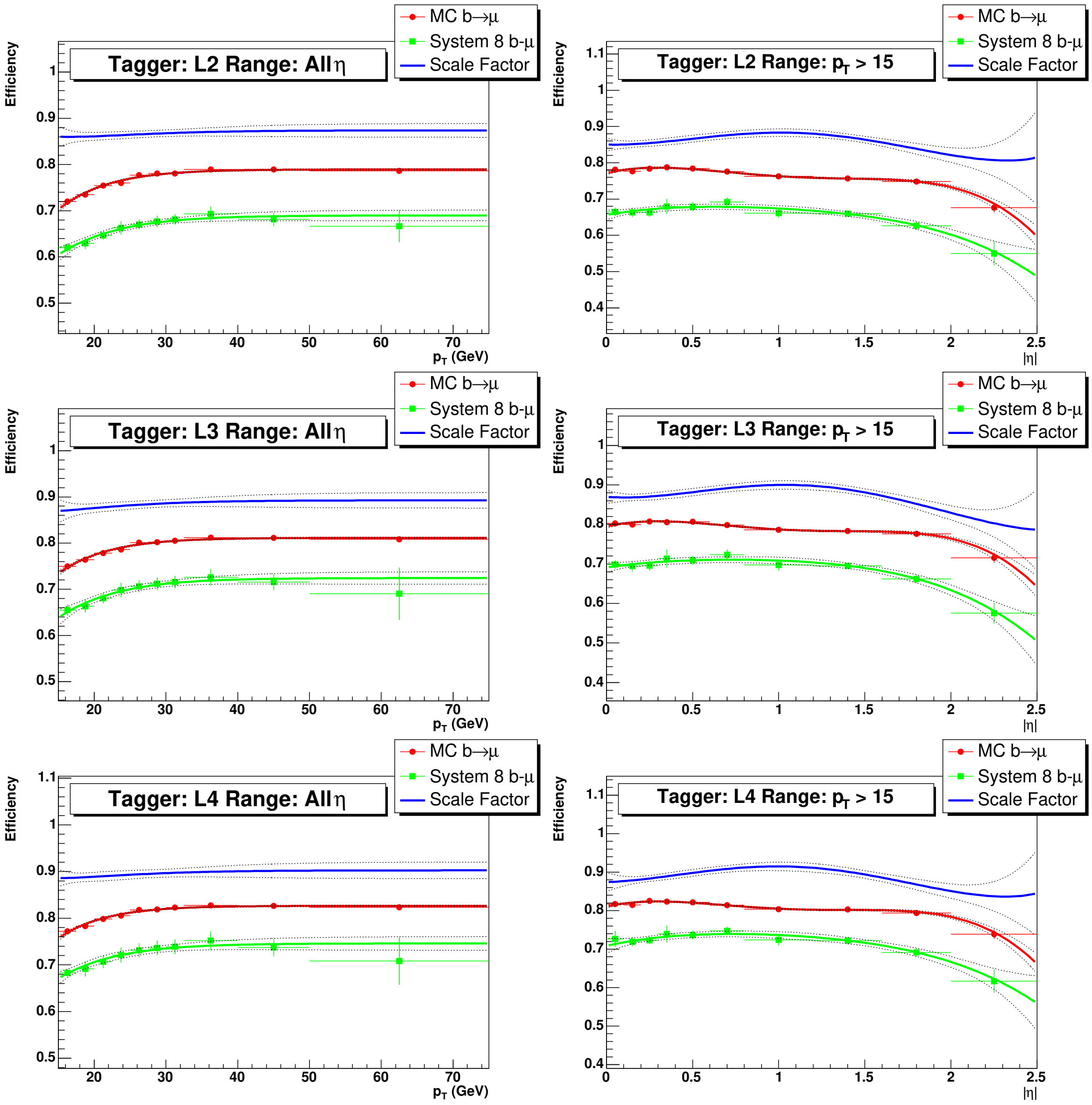

Figure 11: NN tagger muonic b-jet efficiency as a function of p_T (left) and η (right) in the BID μ-jet data sample for the L2 (top), L3 (middle) and L4 (bottom) operating points. The functions used for the parameterisation are outlined in the text and the black dotted curves represent the fit error as calculated from the full fit covariant error matrix.

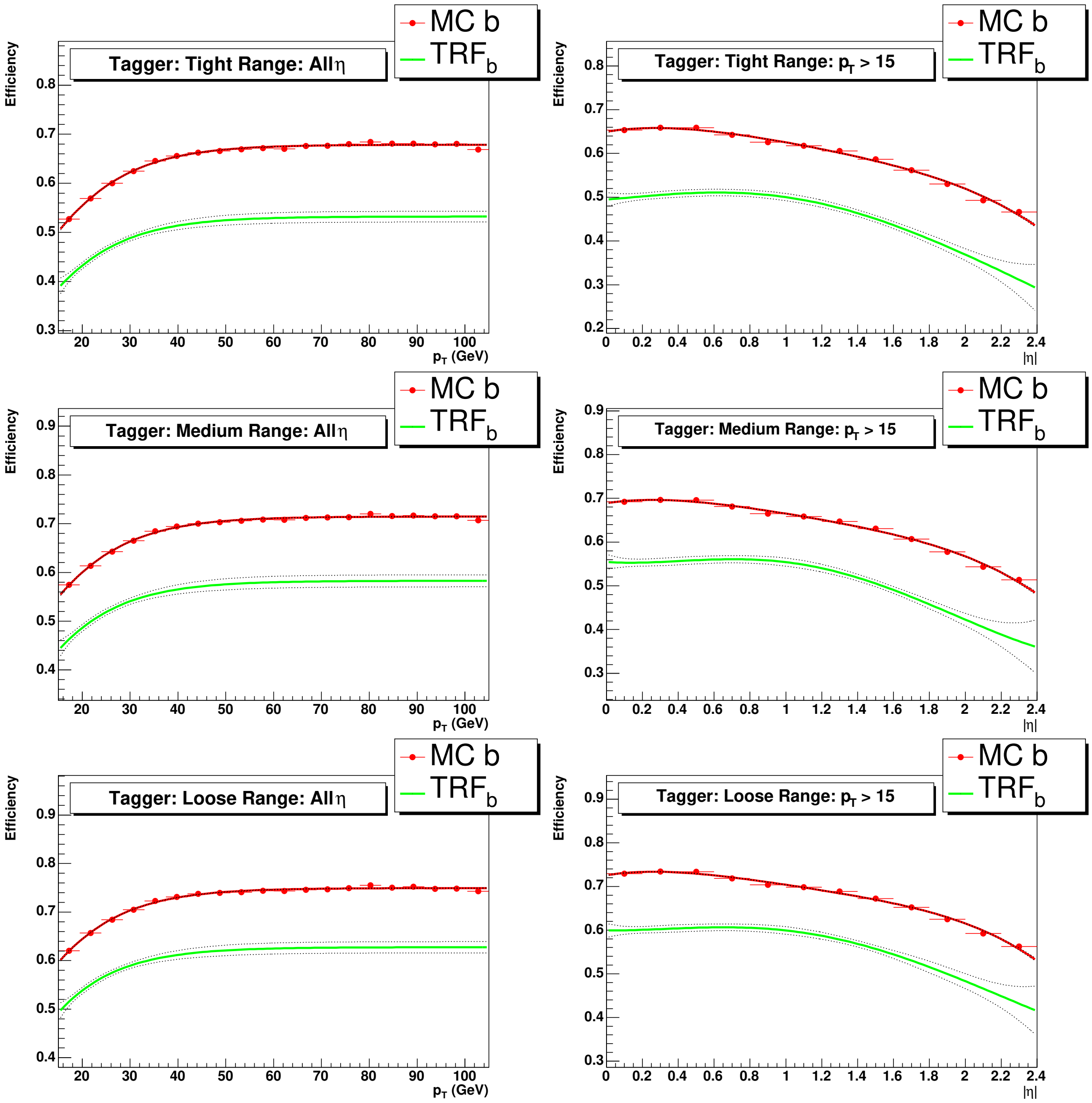

Figure 12: NN tagger b-jet efficiency as a function of p_T (left) and η (right) in the combined b-jet MC sample and the TRF_b derived from multiplying the MC b-jet efficiency by the data/MC SF for the Tight (top), Medium (middle) and Loose (bottom) operating points. The dotted black lines represent the error on the fit which were almost entirely inherited from the error on the scale factor. The functions used for the parameterisation are outlined in the text.

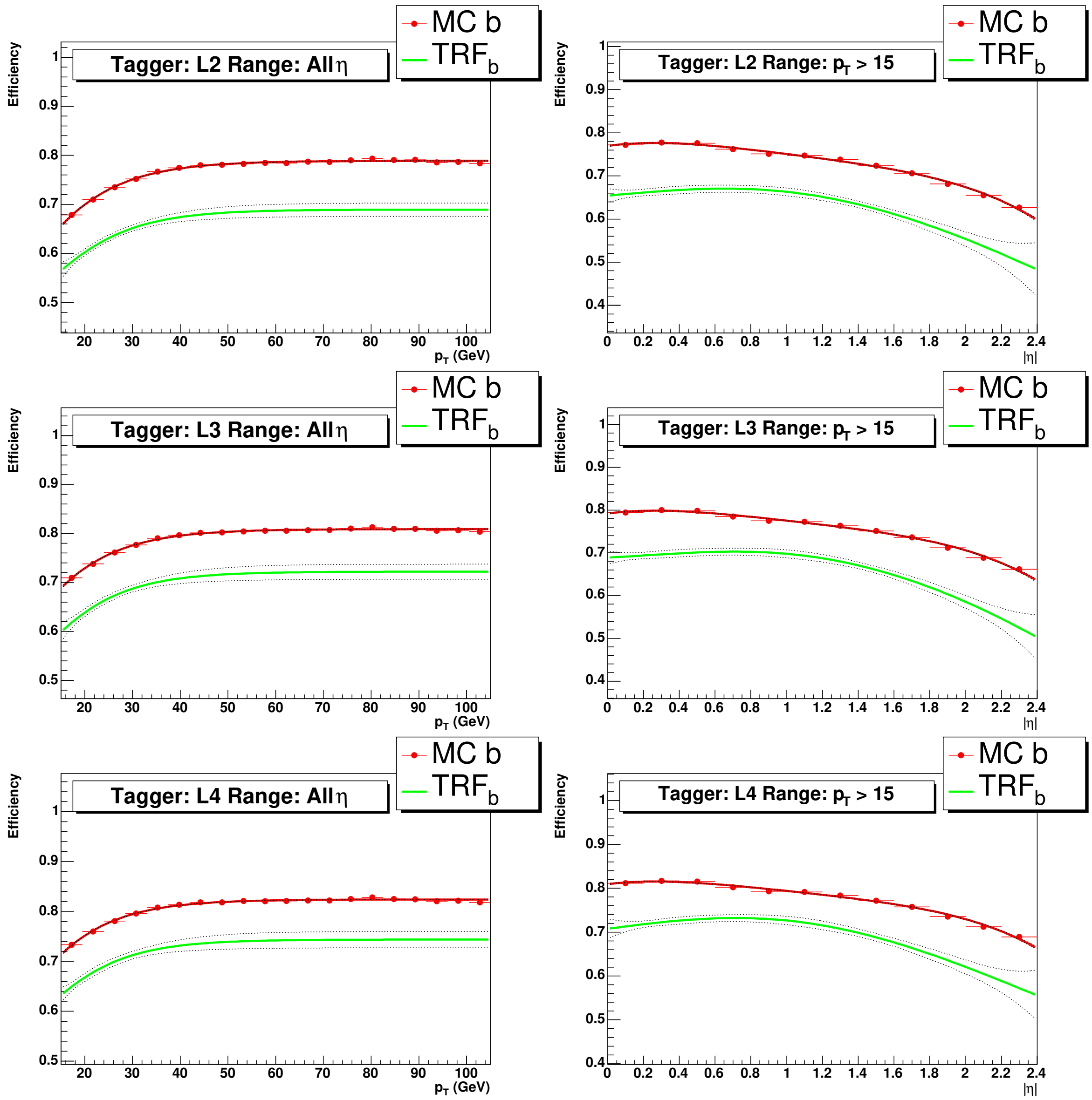

Figure 13: NN tagger b-jet efficiency as a function of p_T (left) and η (right) in the combined b-jet MC sample and the TRF_b derived from multiplying the MC b-jet efficiency by the data/MC SF for the L2 (top), L3 (middle) and L4 (bottom) operating points. The dotted black lines represent the error on the fit which were almost entirely inherited from the error on the scale factor. The functions used for the parameterisation are outlined in the text.

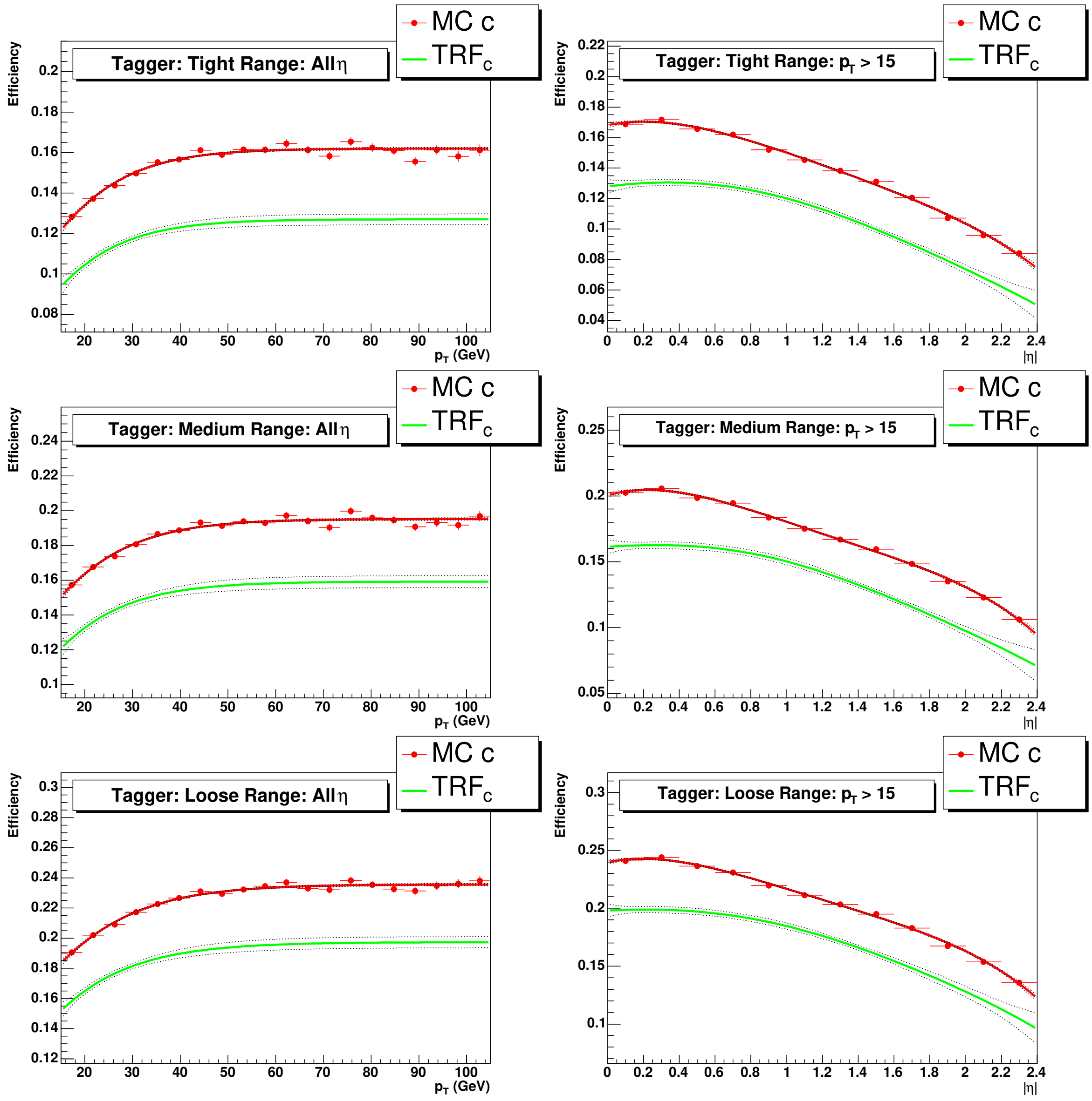

Figure 14: NN tagger c-jet efficiency as a function of p_T (left) and η (right) in the combined c-jet MC sample and the TRF_c derived from multiplying the MC c-jet efficiency by the data/MC SF for the Tight (top), Medium (middle) and Loose (bottom) operating points. The dotted black lines represent the error on the fit which were almost entirely inherited from the error on the scale factor. The functions used for the parameterisation are outlined in the text.

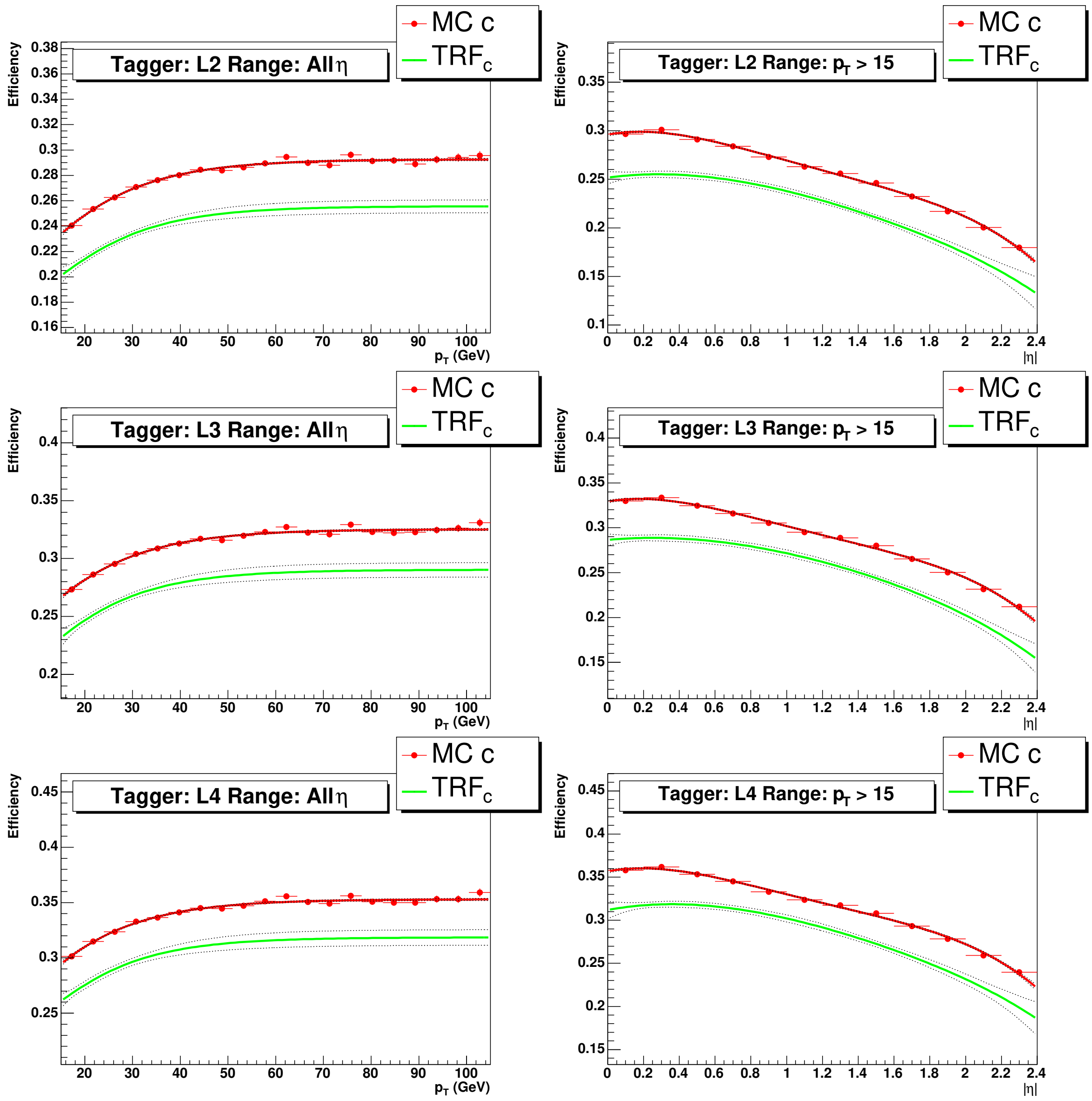

Figure 15: NN tagger c-jet efficiency as a function of p_T (left) and η (right) in the combined c-jet MC sample and the TRF_c derived from multiplying the MC c-jet efficiency by the data/MC SF for the L2 (top), L3 (middle) and L4 (bottom) operating points. The dotted black lines represent the error on the fit which were almost entirely inherited from the error on the scale factor. The functions used for the parameterisation are outlined in the text.

were then compared to the actual number of tags found. Figures 16, 17 and 18 show the comparison for the b, $b \rightarrow \mu$ and c TRFs for the Loose operating point.

A systematic error was calculated by comparing the ratio of the actual and predicted tags in the three η regions. A constant fit was used for the ratio in each of the η regions. These fits were combined into one fit covering all η by weighting the different η fits by the number of jets in that region. The error was taken as the deviation of the ratio from 1, or if the error was consistent with 1 within the error scaled by $\sqrt{\chi^2/NDF}$, the error was taken to be the scaled fit error. The parameterisation which gave the largest error, in this case the combined b sample, was used as the systematic error for the factorisation and parameterisation error on each of the 2D parameterisations. Table 15 shows the parameterisation errors.

The systematic error may appear to underestimate the error as individually several bins show differences of a few percent, and in the EC at high p_T the difference was 5 %. However, when integrating over a reasonable wide spectrum of jets the discrepancies will tend to even out to zero as there are overestimates in some regions (low p_T) and underestimates in others (high p_T). So when using a wide spectrum of jets, as most b-tagging based analysis do, the errors are reasonable values.

Error	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
b	0.415%	0.486%	0.58%	0.747%	0.866%	1.01%
$b \rightarrow \mu$	0.366%	0.368%	0.371%	0.378%	0.384%	0.391%
c	0.375%	0.435%	0.527%	0.41%	0.399%	0.435%

Table 15: Parameterisation error on the b, $b \rightarrow \mu$ and c parameterisations.

4.7.2 Monte Carlo Sample Dependencies

Another systematic error existed due to the use of generic combined samples of b, c and mounic b jets to derive the TRFs. Differences existed between the tagging efficiencies in the different samples and therefore a systematic error needed assigning due to these differences.

To test the agreement between the jets produced from the different processes, and to assign a systematic error due to any differences, the ratio of the efficiencies of each of the MC samples were compared to the combined sample in three η regions. Figures 19, 20 and 21 show comparisons between the tagging efficiencies in the combined b, $b \rightarrow \mu$ and c MC samples and the individual MC sample with which the largest discrepancy was seen.

An overall systematic error was assigned by combining the ratios in each of the regions to an all η ratio by weighting the ratios by the number of jets in each of the regions. The error was taken as the deviation of the ratio from 1, or if the error was consistent with 1 within the error scaled by $\sqrt{\chi^2/NDF}$, the error was taken to be the scaled fit error. The sample which gave the largest error with the combined sample was used as the systematic error.

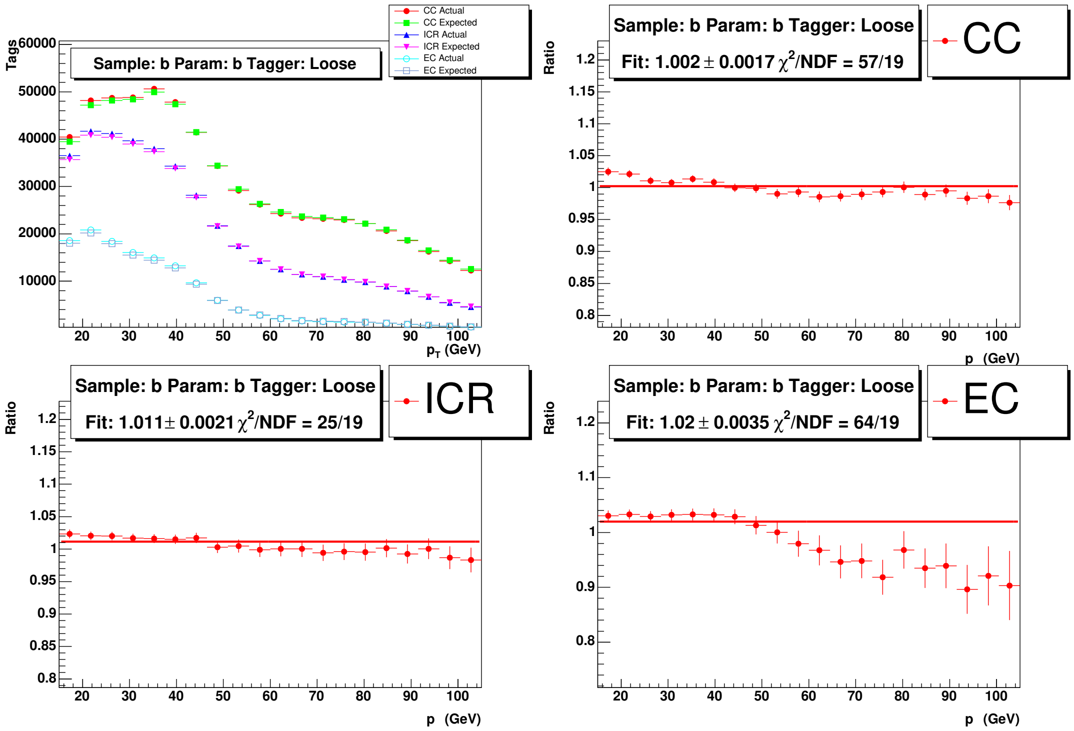

Figure 16: The actual and the predicted number of tags in the combined b sample using a 2D TRF derived from the same sample (top right) and the ratio of the number of tags in the CC (top left), ICR (top right) and EC (bottom) η regions.

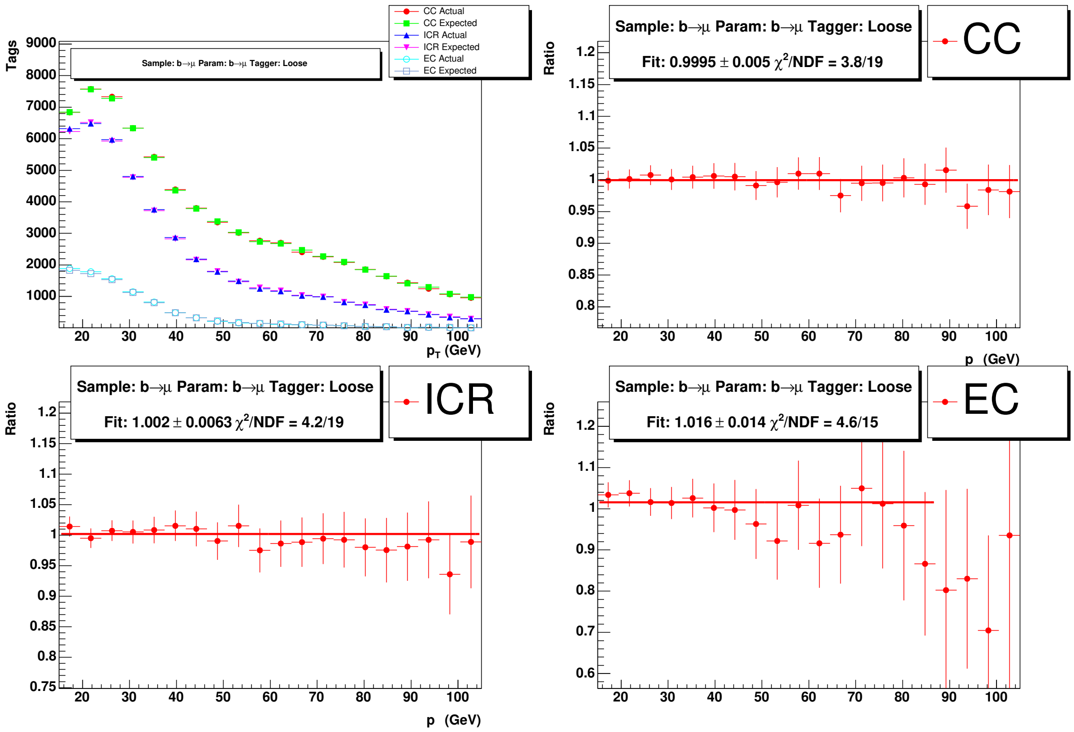

Figure 17: The the actual and the predicted number of tags in the combined $b \to \mu$ sample using a 2D TRF derived from the same sample (top right) and the ratio of the number of tags in the CC (top left), ICR (top right) and EC (bottom) η regions.

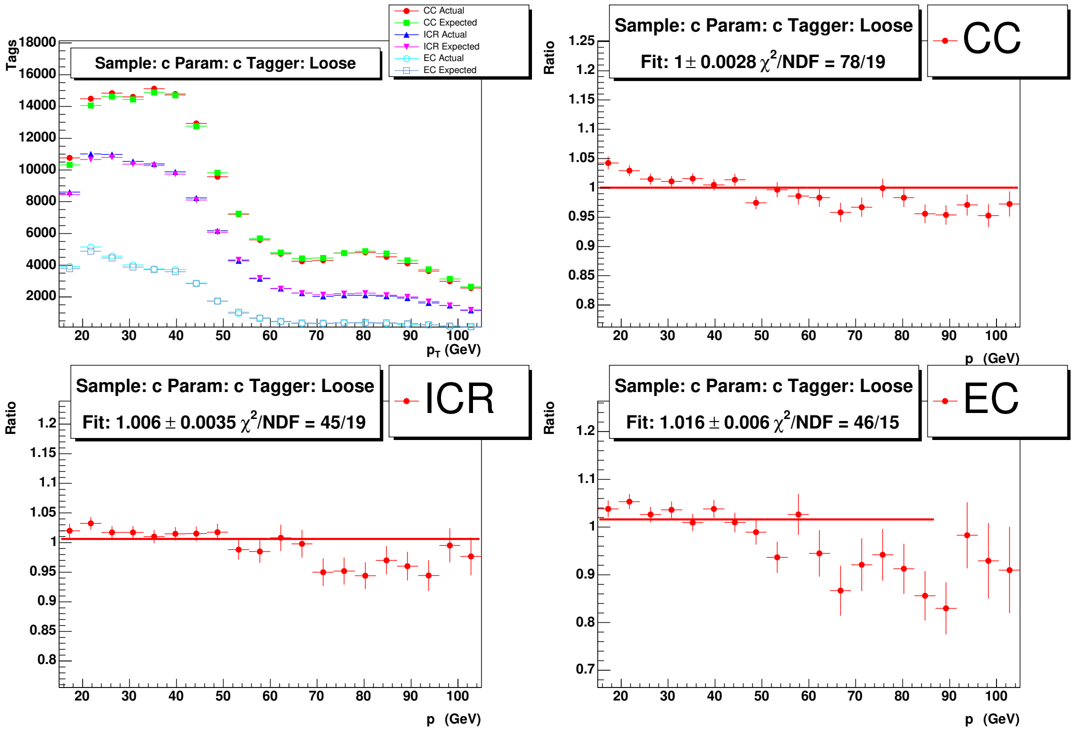

Figure 18: The the actual and the predicted number of tags in the combined c sample using a 2D TRF derived from the same sample (top right) and the ratio of the number of tags in the CC (top left), ICR (top right) and EC (bottom) η regions.

Table 16 shows the errors for each of the constituent samples compared to the combined sample.

4.7.3 System8 Systematic Errors

It should be noted at this point that in previous certifications the systematic error calculated due to the uncertainty in the S8 correlation coefficients was combined with the other systematic errors to give an overall systematic error. In this certification the S8 systematic error was calculated in each bin in the projections and it was concluded (see section 4.3 for discussion) that the individual bin error was a more accurate representation of the uncertainty on the data b-efficiency than a one size fits all S8 systematic error. As such for each bin the S8 systematic error for that bin was added in quadrature with the fit error from Minuit. This gave an overall uncertainty with which the efficiency was known for each bin. This was repeated for each bin in p_T, η and for the S8 efficiency of the entire sample. The combined errors were then used to fit the parameterised curves in p_T and η, with the S8 error folded into the statistical errors.

Error	L4	L3	L2	Loose	Medium	Tight
directbb/b	0.268%	0.267%	0.289%	0.33%	0.342%	0.346%
zbb/b	0.435%	0.481%	0.544%	0.62%	0.625%	0.571%
tt/b	0.477%	0.496%	0.528%	0.623%	0.666%	0.726%
directcc/c	2.24%	2.26%	2.04%	1.62%	0.947%	0.55%
zcc/c	2.34%	2.4%	2.49%	2.22%	2.06%	2.05%
ttc/c	0.454%	0.458%	0.653%	1.39%	2.17%	2.52%
ttmu/bmu	0.291%	0.255%	0.274%	0.31%	0.342%	0.375%
zbbmu/bmu	0.256%	0.271%	0.292%	0.331%	0.364%	0.4%
directbbmu/bmu	0.256%	0.241%	0.259%	0.293%	0.323%	0.355%

Table 16: Errors due to the differences between the combined b, c and muonic b-jet samples and their constituent samples.

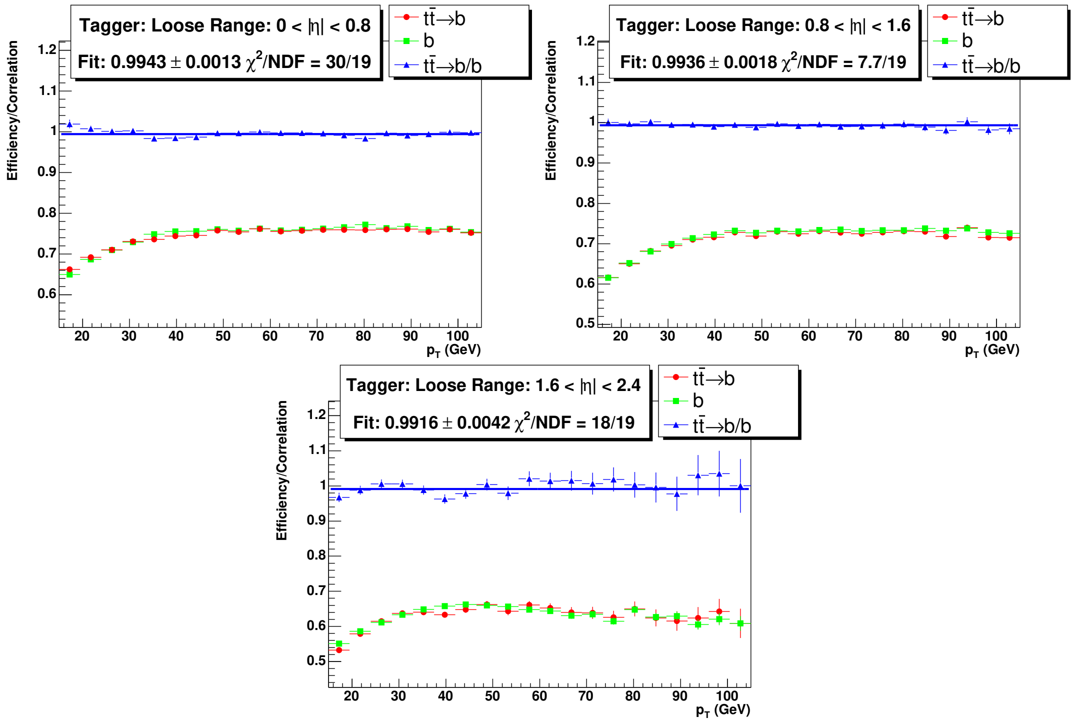

Figure 19: The ratio of the b-tagging efficiencies in the $t\bar{t}$ and combined b samples in the CC (top left), ICR (top right) and EC (bottom) η regions.

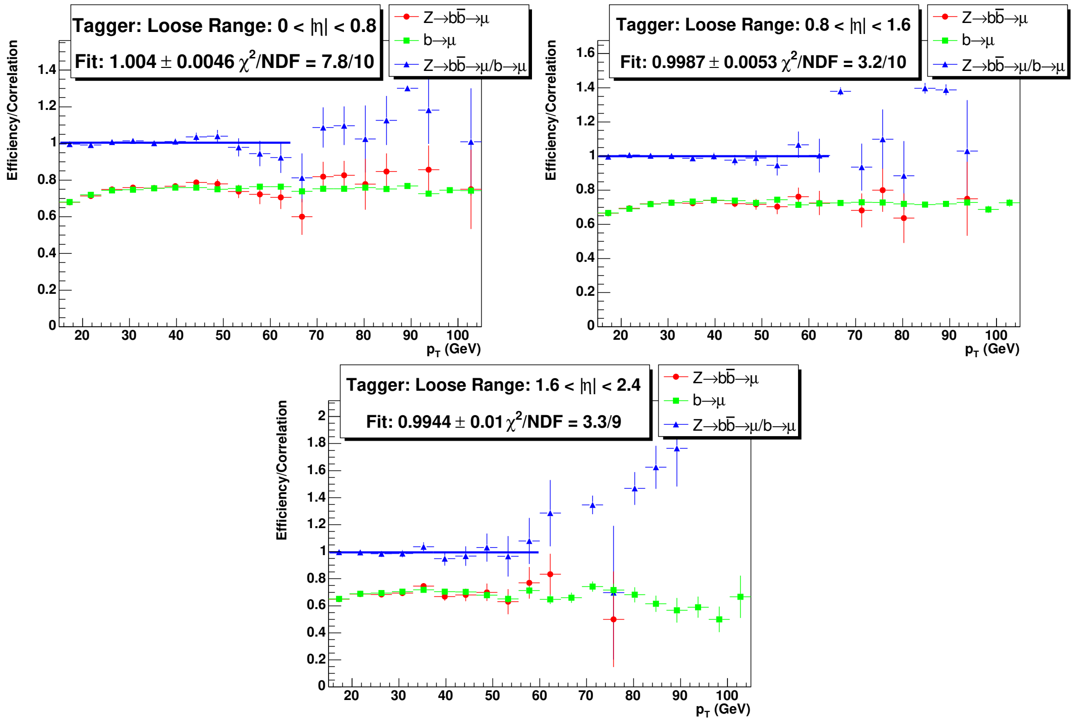

Figure 20: The ratio of the b-tagging efficiencies in the $Z \to b\bar{b} \to \mu$ and combined $b \to \mu$ samples in the CC (top left), ICR (top right) and EC (bottom) η regions.

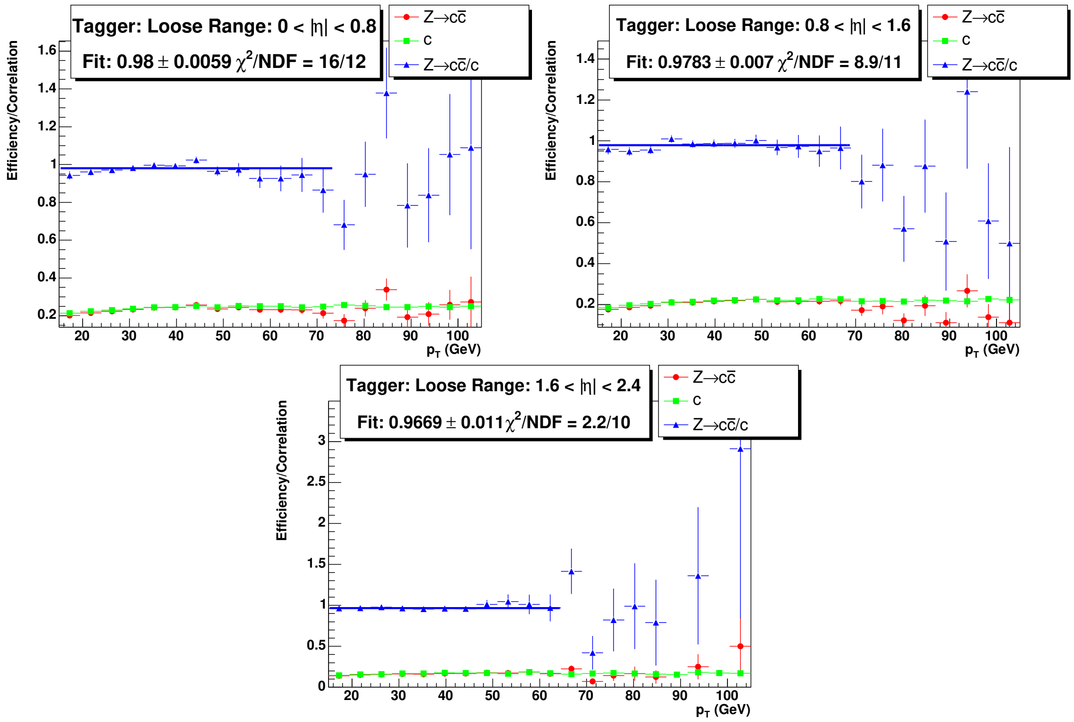

Figure 21: The ratio of the c-tagging efficiencies in the QCD $Z \to c\bar{c}$ and combined c samples in the CC (top left), ICR (top right) and EC (bottom) η regions.

4.7.4 Total Systematic Errors

Each of the TRFs and the SF had an individual total systematic error constructed from different components as detailed below. The total systematic errors are shown in Table 17 and are added in quadrature to the statistical error to give the total error as detailed in the next sections.

SF - The factorisation error and the systematic error due to the MC sample differences in $b \rightarrow \mu$ MC, added in quadrature.

TRF_b **-** The SF systematic error with an additional term due to the MC sample differences in the b-jet MC, added in quadrature.

TRF_c **-** The SF systematic error with an additional term due to the MC sample differences in the c-jet MC, added in quadrature.

Error	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
SF	0.487%	0.556%	0.649%	0.817%	0.939%	1.08%
TRF_b	0.682%	0.745%	0.837%	1.03%	1.15%	1.3%
TRF_c	2.39%	2.46%	2.58%	2.37%	2.26%	2.32%

Table 17: Total systematic errors on the SF and TRFs.

4.7.5 Statistical Error

The overall statistical errors were calculated by evaluating

$$\sigma_{stat} = \frac{f^{+1\sigma}(p_T) \times f^{+1\sigma}(\eta)}{f^{+1\sigma}_{All}} - \frac{f(p_T) \times f(\eta)}{f_{All}} \tag{6}$$

where σ_{stat} is the statistical error, f was the parameterised fit in p_T, η and for the entire sample (All) and $+1\sigma$ were the +1 sigma fits from the central function value. This was also repeated with the -1 sigma curves with the largest deviation taken as the statistical error.

4.7.6 Total Error

The total errors, given by the statistical and systematic errors combined in quadrature, for the SF, TRF_b and TRF_c are shown in Fig. 22. The statistical errors were the dominant error and the dominant component of the statistical error was the uncertainty with which the b-efficiency was known in data.

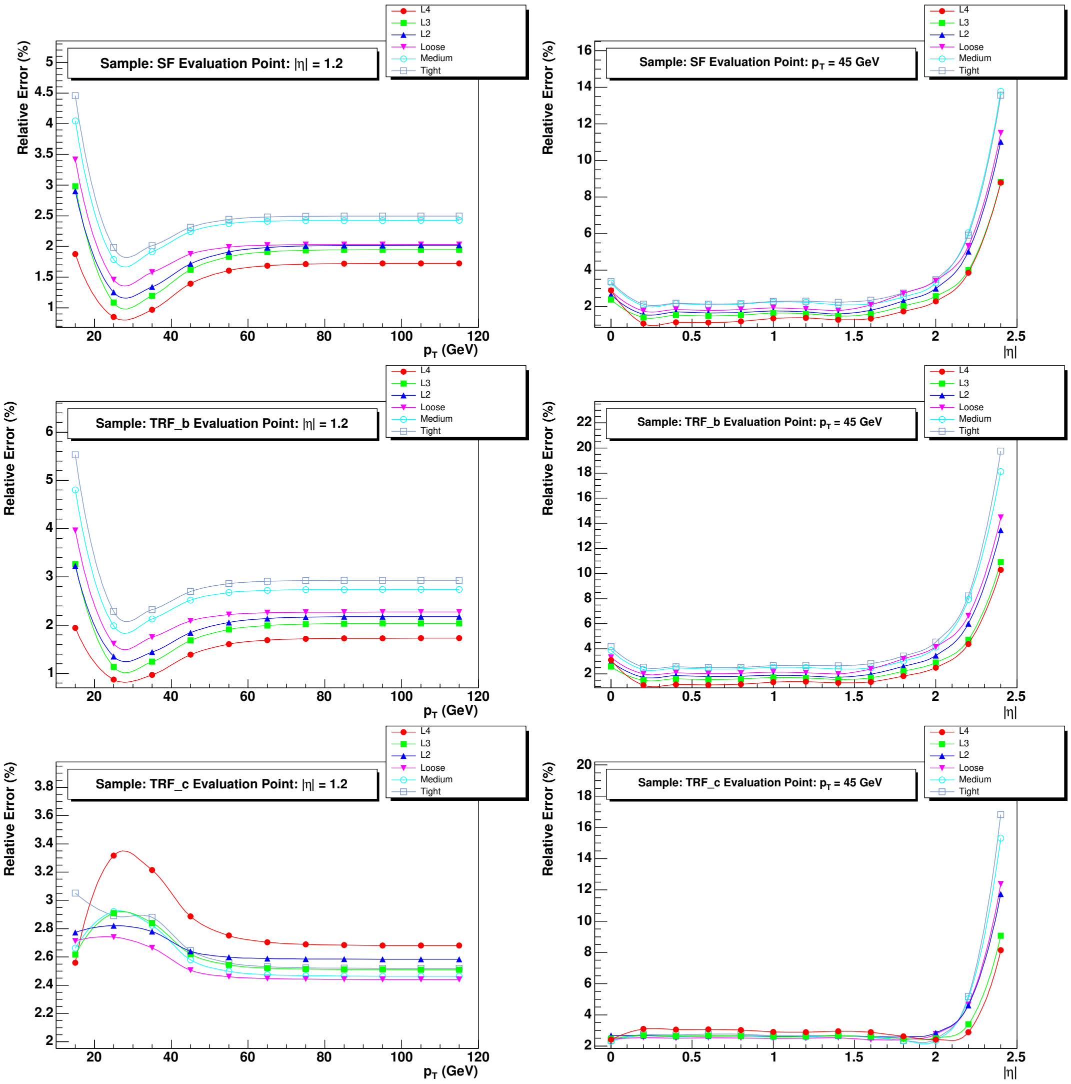

Figure 22: The total relative error (combined systematic and statistical) for the Scale Factor (SF) (top), TRF_b (middle) and TRF_c (bottom) in terms of p_T (left) when $\eta = 1.2$ and η (right) when $p_T = 45$ GeV.

5 Fake Rate

5.1 Negative Tag Rate

The negative tag rates (NTRs) produced from the combined EM and QCD skims were used to estimate the fake tag rates (FTRs) on data. The shape of the p_T distributions were different in the forward and central calorimeter regions and so it was therefore not deemed valid to assume that the FTR factorised into p_T and η when parameterising the NTR. Thus, instead of factorising the NTR parameterisation in η and p_T, as for the S8 efficiencies (there was insufficient statistics to repeat this procedure in the S8 analysis), the NTR was parameterised in p_T in the three η regions outlined in Section 1.5.

The NTRs for the NN tagger are outlined in Table 18 for the COMB skim and are shown for the EM, QCD and COMB skims for the Loose tagger in Fig. 23 along with the NT ratios between the QCD and EM skims.

Region	L4	L3	L2	Loose	Medium	Tight
CC	5.36%	4.13%	2.96%	1.54%	0.884%	0.501%
ICR	4.92%	3.75%	2.65%	1.34%	0.75%	0.415%
EC	4.74%	3.58%	2.48%	1.18%	0.613%	0.304%

Table 18: NTR for the COMB skim.

A small discrepancy was observed between the NTRs in the QCD and EM skims, although the observed differences were much smaller than those seen for the JLIP and SVT taggers. In the CC and ICR regions the differences were less than 1 %, although in the EC region a difference of 4 % was seen. Numerous studies have been conducted into the NTR differences between the EM and QCD skims [3][4]. Several possible explanations for the differences between the NTRs have been suggested although no one reason has been shown to be the cause. The two most important seem to be the trigger conditions and the skimming criteria, which could cause biases towards certain types of jets. The SVT and JLIP taggers got much better agreement between the EM and QCD skims by only using jets with EMF < 0.8 (Low EMF). However, repeating this cut for the NN tagger made the agreement a lot worse. Figure 24 shows the NT rates and the agreement between the QCD and EM skims for low EMF jets. The agreement between the NT rates was a lot worse and the reason for this difference was not properly understood. For the JLIP and SVT taggers the opposite effect was seen, where a good agreement between was seen between low EMF jets with a bad agreement seen when using all the jets in the skim.

The COMB skim was used to predict the FTR with a systematic error assigned due to the difference between the constituent QCD and EM skims. The systematic error was calculated from a constant fit to the EM and QCD ratio. Half the difference between the fit value and unity was taken as the systematic error, or if the ratio was consistent with unity within the fit error scaled by $\sqrt{\chi^2/NDF}$, then the scaled fit error was taken as the error. The

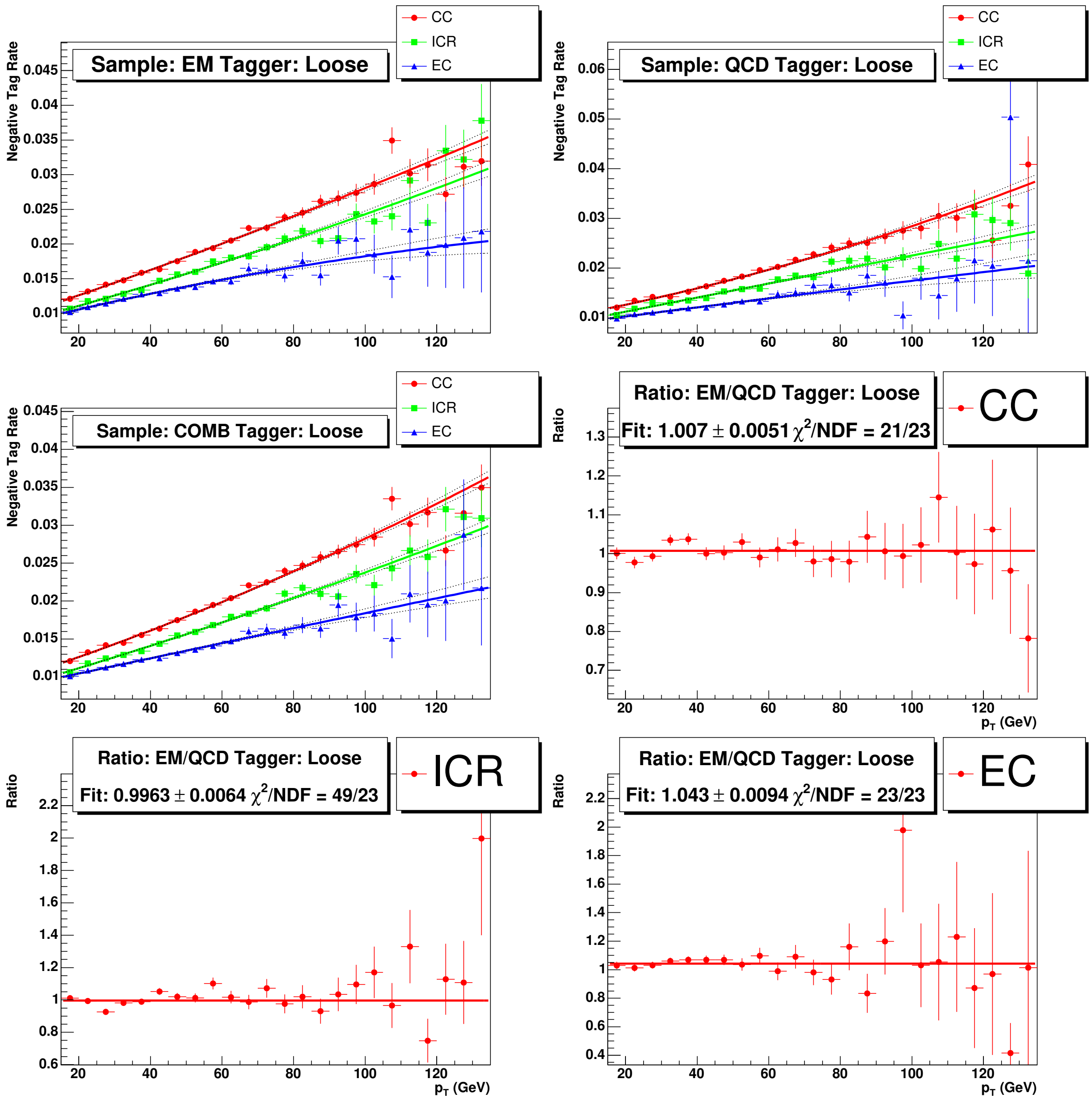

Figure 23: The NTR parameterisation for the EM skim (top left), QCD skim (top right), the COMB skim (middle left) and the ratio of the EM and QCD NTR in the CC (middle right), ICR (bottom left) and EC (bottom right) η regions. The NTR was parameterised with a second order polynomial and the dotted lines show the fit error calculated using the full fit covariant error matrix.

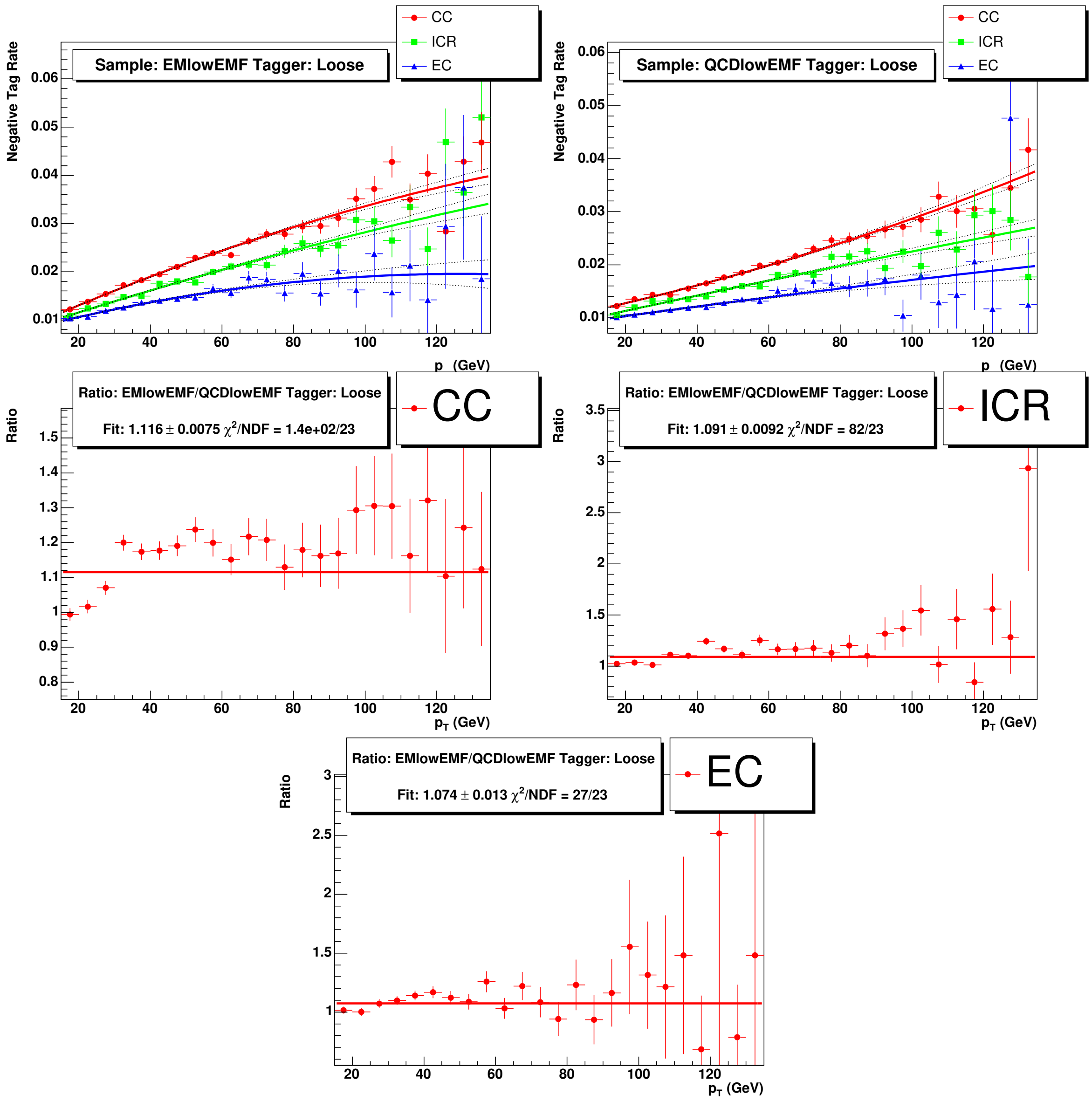

Figure 24: The NTR parameterisation for the EM skim (top left), QCD skim (top right) for jets with EMF < 0.8 and the ratio of the EM and QCD NTR in the CC (middle right), ICR (bottom left) and EC (bottom right) η regions. The NTR was parameterised with a second order polynomial and the dotted lines show the fit error calculated using the full fit covariant error matrix.

systematic errors are shown in Table 19 for the difference when using all jets and Table 20 when comparing only low EMF jets. The comparison which showed the largest discrepancy, in this case the low EMF comparison, was used to assign the error.

Region	L4	L3	L2	Loose	Medium	Tight
CC	0.461%	0.42%	0.513%	0.363%	0.34%	0.576%
ICR	0.344%	0.347%	0.385%	0.471%	0.523%	0.721%
EC	1.81%	1.87%	1.96%	2.13%	1.77%	1.2%

Table 19: Error assigned due to the EM and QCD skim NTR difference.

Region	L4	L3	L2	Loose	Medium	Tight
CC	3.9%	4.23%	4.92%	5.78%	6.7%	8%
ICR	3.14%	3.53%	3.93%	4.53%	5.58%	6.84%
EC	2.89%	2.92%	3.16%	3.68%	3.34%	3.27%

Table 20: Error assigned due to the EM and QCD skim NTR difference for jets with EMF < 0.8.

5.2 Parameterisation Systematic

To test the parameterisation of the NT in the three η regions a comparison was made between the number of tags found by the tagger and the number predicted from the NTR. Figure 25 shows the expected number of tags and the predicted number of tags as a function of p_T in each of the η regions. A systematic error due to the parameterisation of the NT was calculated from the straight line fit to the ratio of the actual and predicted number of tags. The error was taken as the deviation of the ratio from 1, or if the ratio was consistent with 1 within the error scaled by $\sqrt{\chi^2/NDF}$, the error was taken to be the scaled fit error. The systematic errors are shown in Table 21.

As a crash check the actual and predicted number of tags from the COMB parameterisation were also compared on the QCD and EM skims and are shown in Figs. 26 and 27 with the errors outlined in Tables 22 and 23. These results should be compared to the errors assigned due to the difference between the EM and QCD skims in Table 19. The differences are small, $< 1\%$ except for the EC, and are well within the the larger low EMF ratio assigned to account for the discrepancies between the skims.

5.3 MC Negative Tag Scale Factors

The NTR was not a perfect approximation for the PT fake rate as a contribution existed in the NTR from c and b-quarks and an asymmetry existed between the NTs and PTs due to

Region	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
CC	0.14%	0.158%	0.186%	0.256%	0.338%	0.448%
ICR	0.17%	0.194%	0.229%	0.321%	0.428%	0.575%
EC	0.228%	0.261%	0.312%	0.0641%	0.621%	0.881%

Table 21: The systematic uncertainties on the COMB skim parameterisation.

Region	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
CC	0.593%	0.558%	0.651%	0.5%	0.5%	0.912%
ICR	0.267%	0.305%	0.361%	0.503%	0.67%	0.9%
EC	1.95%	2.05%	2.17%	2.34%	2.3%	4.02%

Table 22: The systematic uncertainties on the QCD skim parameterisation.

Region	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
CC	0.546%	0.529%	0.626%	0.541%	0.545%	0.907%
ICR	0.22%	0.251%	0.297%	0.415%	0.554%	0.744%
EC	1.76%	1.85%	1.96%	2.11%	1.99%	2.02%

Table 23: The systematic uncertainties on the EM skim parameterisation.

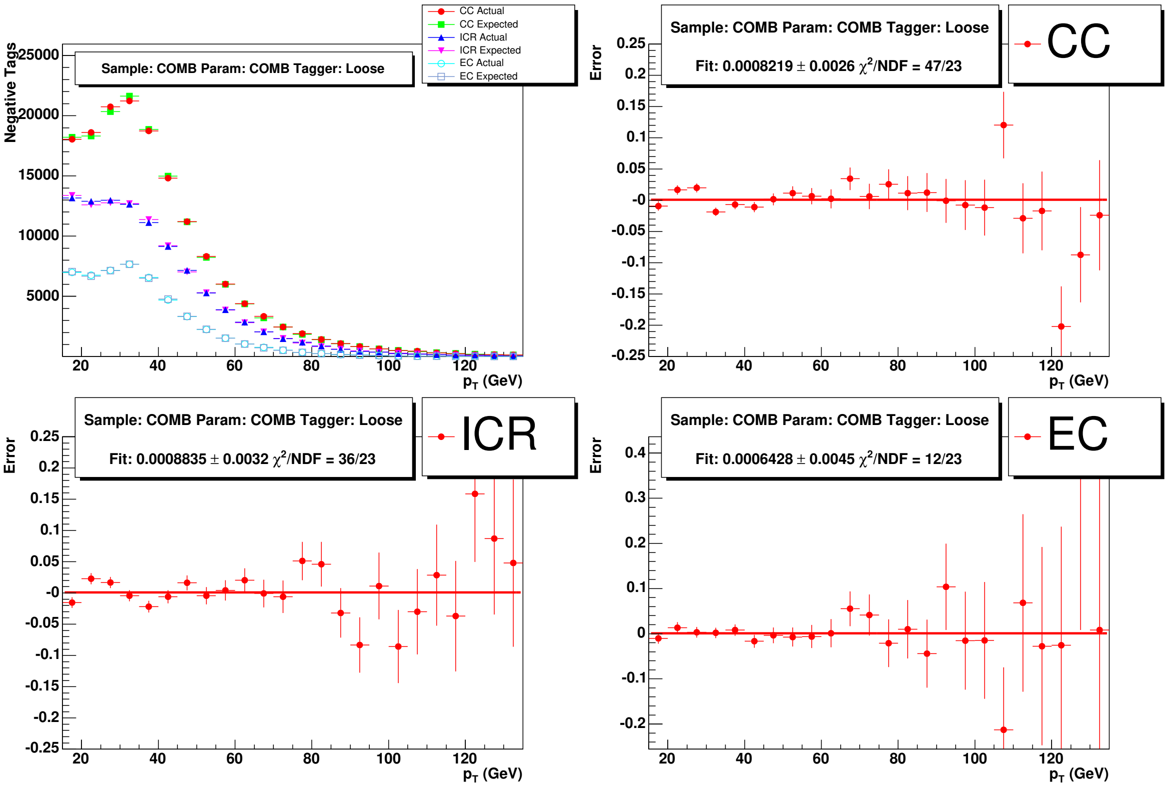

Figure 25: The predicted and actual number of NT in the three eta regions for the loose tagger on the COMB skim (top left) and the relative difference between the actual and observed number of NTs in the CC (top right), ICR (bottom left) and EC (bottom right) η regions.

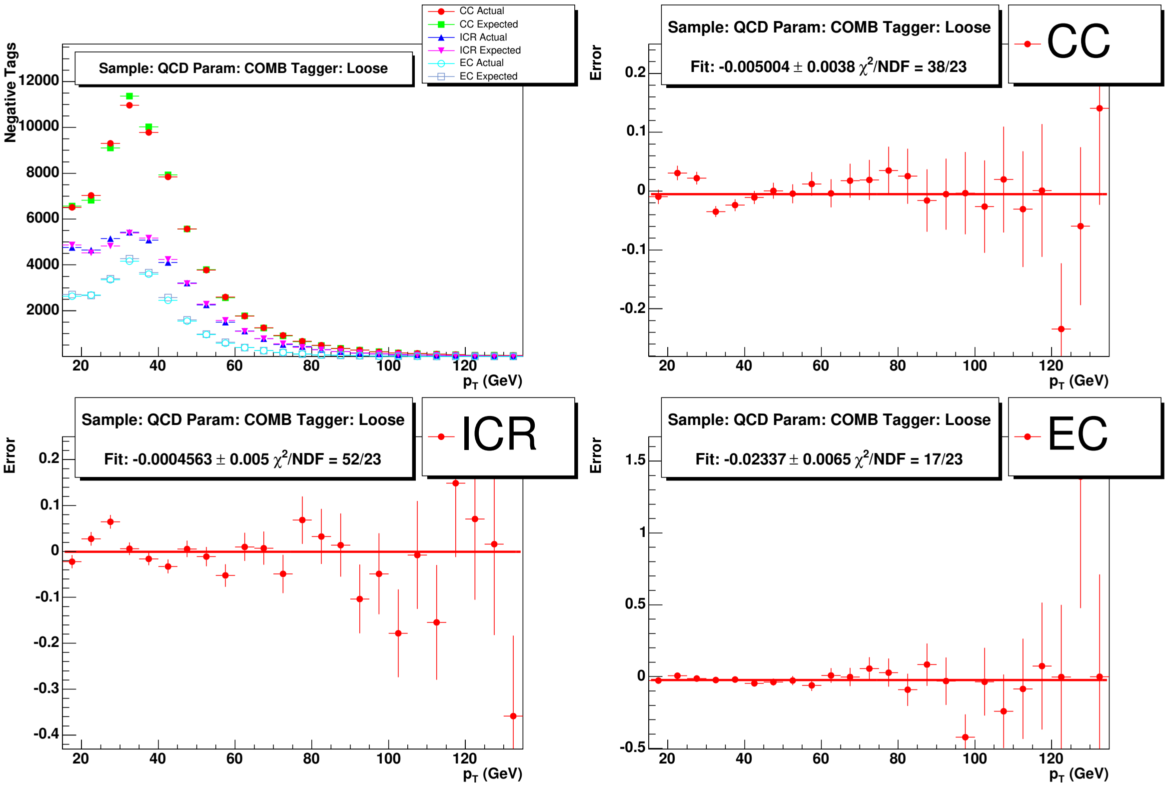

Figure 26: The predicted and actual number of NT in the three eta regions for the loose tagger on the QCD skim (top left) and the relative difference between the actual and observed number of NTs in the CC (top right), ICR (bottom left) and EC (bottom right) η regions.

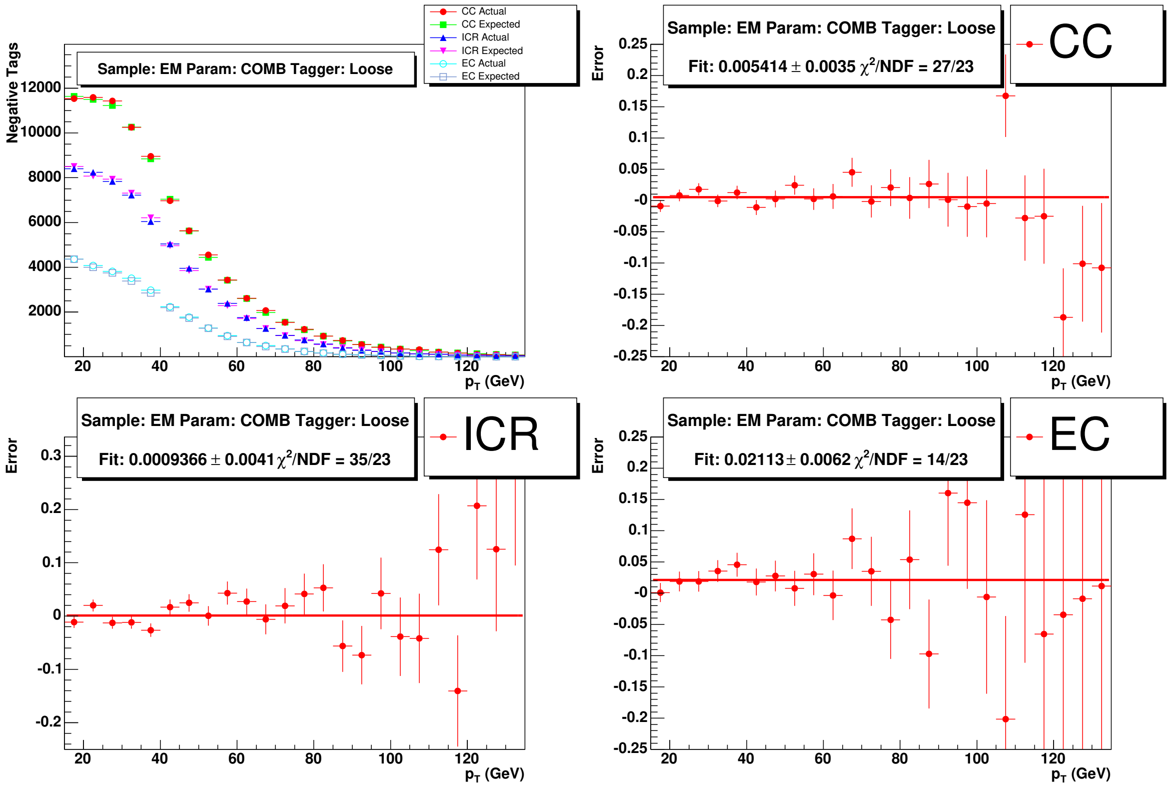

Figure 27: The predicted and actual number of NT in the three eta regions for the loose tagger on the EM skim (top left) and the relative difference between the actual and observed number of NTs in the CC (top right), ICR (bottom left) and EC (bottom right) η regions.

some long lifetime uds quarks. To correct the NTR to the PT fake rate two NT correction scale factors (NT SF) were applied, both of which were measured in QCD MC.

SF_{hf} **-** The ratio of the light jet NTR to the total NTR.

SF_{ll} **-** The ratio between the light jet PTR and the light jet NTR.

Applying these two correction factors to the NTR yielded the data FTR.

$$FTR = SF_{hf} \times SF_{ll} \times NTR \tag{7}$$

The NT SFs as a function of p_T in the three η regions are shown in Fig. 28 for the L4 operating point. The fit error on the NT SFs were scaled up by the $\sqrt{\chi^2/NDF}$ of the fit to account for the uncertainty in the fit parameterisations.

In the EC a constant fit was used. However, in the ICR and CC regions the NT SF was observed to be constant at high p_T but with a distinct dip at low P_T. Several fit function were tested to try and model this function however none were successful. Fitting a straight line fit at high p_T and a parabola at low p_T was also tested however this produced discontinuities in both the value of the NT SF and the gradient. To overcome these problems a fourth order polynomial was used between $15 < p_T < 95$ in the CC region and $15 < p_T < 135$ in the ICR region. These ranges included the beginning of the constant region at high p_T so that the fit would reproduce the dip and find a maxima at the beginning of the constant region. The fit value and error at the maxima were then used to continue the curve as a constant value. This method ensured a FTR which was a continuous function in both value and gradient and which correctly described the constant at high p_T.

5.4 Fake Rate

The NN tagger FTRs are shown in Fig. 29 and the rates for the COMB skim are shown in Table 24.

Region	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
CC	4.13%	3.09%	2.14%	1.03%	0.565%	0.289%
ICR	4.22%	3.14%	2.17%	1.06%	0.555%	0.288%
EC	4.86%	3.62%	2.49%	1.19%	0.63%	0.299%

Table 24: The data FTR calculated from the COMB NTR corrected for the presence of heavy flavour and light quark PT/NT asymmetry.

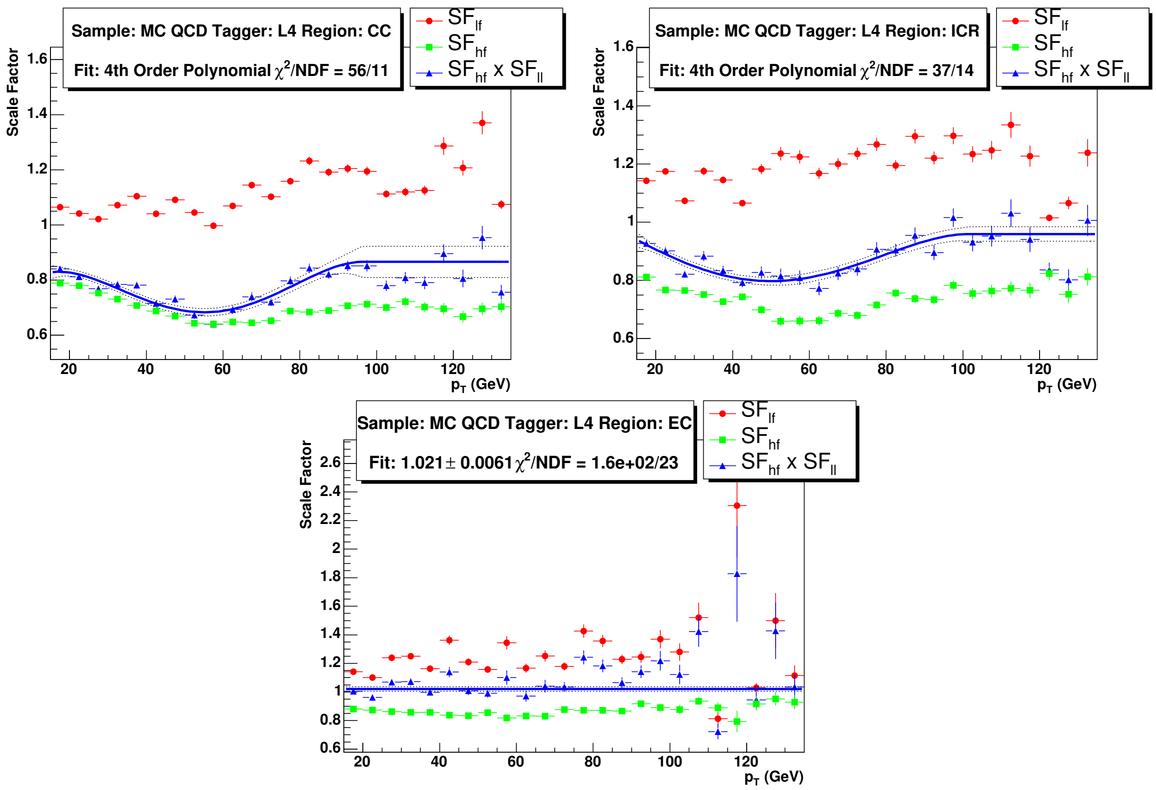

Figure 28: The heavy flavour correction (SF_{hf}), light jet asymmetry correction (SF_{ll}) and total NT SF correction in the CC (top left), ICR (top right) and EC regions (bottom) for the L4 operating point. The fit to the SF in the ICR and CC was a fourth order polynomial continued as a constant at the fits maxima and a straight line in the EC region. The black dotted line is the error on the fit calculated from the full fit covariant error matrix. The error on the fit was scaled up by $\sqrt{\chi^2/NDF}$ due to the poor agreement.

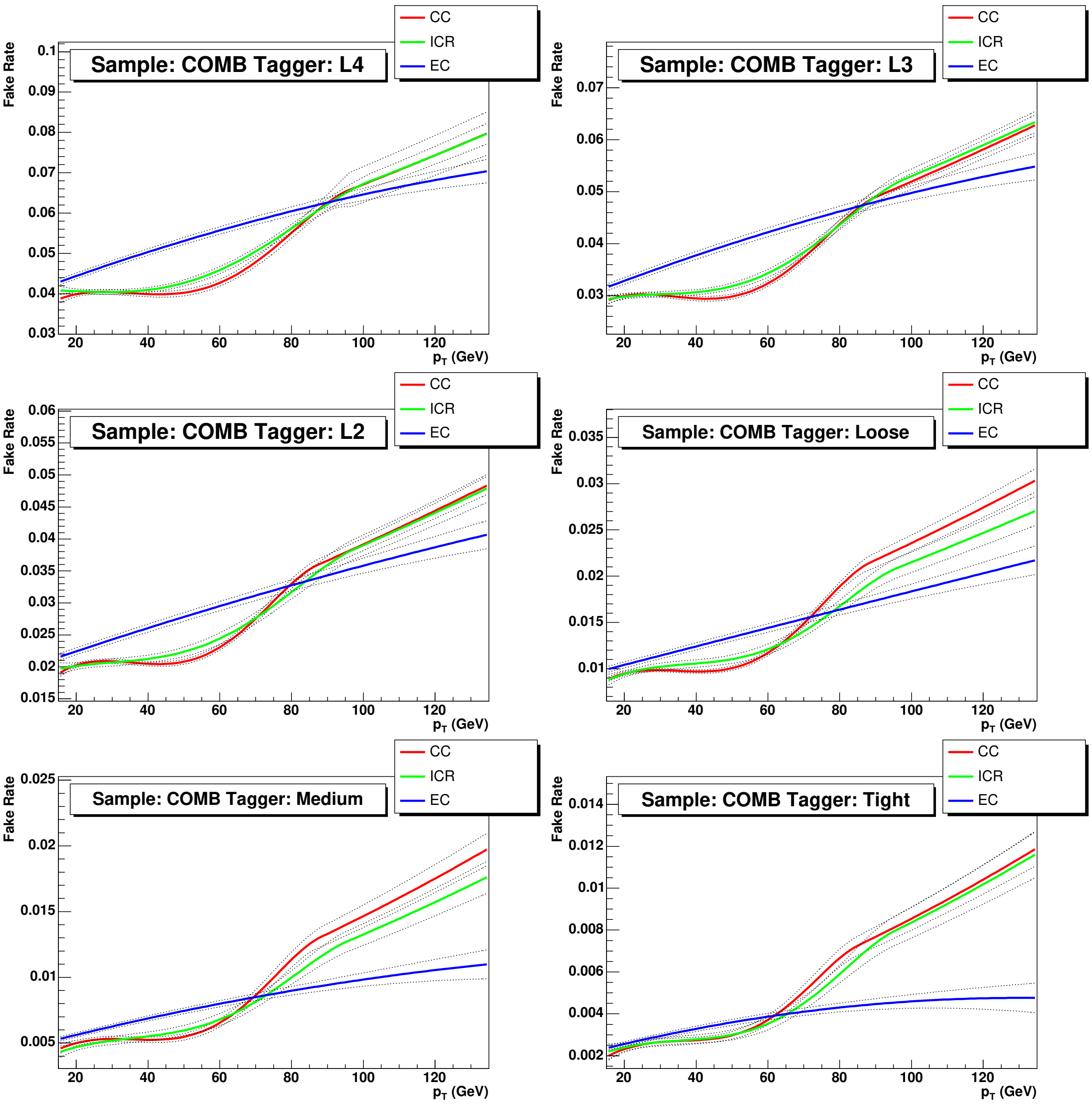

Figure 29: The FTRs parameterised in the three η regions for the L4 (top left), L3 (top right), L2 (middle left), Loose (middle right), Medium (bottom left) and Tight (bottom left) operating points. The fake rate was parameterised with a second order polynomial, and the dotted black lines represent the fit error calculated using the full error matrix.

Correcting the NTR with the MC SF correction introduced another systematic error due to the uncertainty in the proportions of b and c in the simulation. To estimate a systematic error the number of b and c quarks present in the MC was varied by $\pm 20\%$ and the fake rate recalculated with the largest discrepancy taken as the systematic error on the b/c content. The errors are outlined in Tables 25 and 26.

Region	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
CC	3.11%	3.49%	3.97%	5.01%	5.93%	7.21%
ICR	2.88%	3.26%	3.75%	4.84%	5.69%	6.89%
EC	1.32%	1.51%	1.59%	2.13%	2.78%	3.41%

Table 25: The systematic error in the FTR due to the uncertainty in the b-content in the MC.

Region	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
CC	2.48%	2.64%	2.83%	3%	3.1%	3.08%
ICR	2.28%	2.38%	2.56%	2.8%	2.99%	3.08%
EC	1.3%	1.4%	1.49%	1.67%	1.98%	2.18%

Table 26: The systematic error in the FTR due to the uncertainty in the c-content in the MC.

5.5 Total Systematics

‘ The total systematic error for the FTR were calculated by adding the systematic uncertainties on the b and c content (as they were likely to be highly correlated) and then adding this in quadrature with the parameterisation and EM/QCD ratio systematic errors. The combined systematics are shown in Table 27.

The total error on the FTR were given by adding in quadrature the total systematic error for the appropriate region to the statistical error given by the difference between the FTR central value and the +1/-1 sigma fit curves. The systematic error was much larger than the statistical error and so the errors in Table 27 are a good indication of the total error.

Region	**L4**	**L3**	**L2**	**Loose**	**Medium**	**Tight**
CC	6.81%	7.43%	8.37%	9.83%	11.1%	12.9%
ICR	6.02%	6.64%	7.38%	8.87%	10.3%	11.9%
EC	3.88%	4.08%	4.3%	5.16%	5.74%	6.24%

Table 27: Total systematic errors on the FTR.

6 Data Performance

The performance of the NN and JLIP taggers as measured on data including full statistical and systematic errors on the measurement is shown in Fig. 30 for all jets and also for jets in the with $\eta < 0.8$ and $p_T > 30$ GeV. The NN tagger demonstrates considerable improvement over the JLIP tagger for all operating points.

The data performance of the NN tagger was also evaluated on two MC samples with the tagging performance and error calculated jet by jet using the derived TRFs. This demonstrates the overall performance and error of the NN tagger when applied jet by jet to a MC sample. The performance is shown for all jets in the $Z \rightarrow bb$ and $Z \rightarrow qq$ MC and also for jets in the with $|\eta| < 1.2$ and with $p_T > 30$ GeV for QCD b and uds-jets in Fig. 31. A clear improvement is seen in both the MC samples compared to the JLIP tagger, the uncertainty with which the b-efficiency is known is about half the value of the JLIP tagger although the fake rate is seen to have a larger uncertainty for the looser operating points.

7 Conclusion

The performance of the NN tagger on pass 2 corrected data was measured and parameterised into 2D tag rate functions for six operating points with fake rates ranging from 0.3% to 4.0%. The performance of the NN tagger on data compared to the JLIP tagger showed considerable improvement, although the improvement was smaller than that seen in MC. For a fixed fake rate improvements in signal efficiency were seen ranging from 50% for the tightest operating to 15% for the loosest operating point. Fake rates were typically reduced to between a quarter and a third of their value for a fixed signal efficiency.

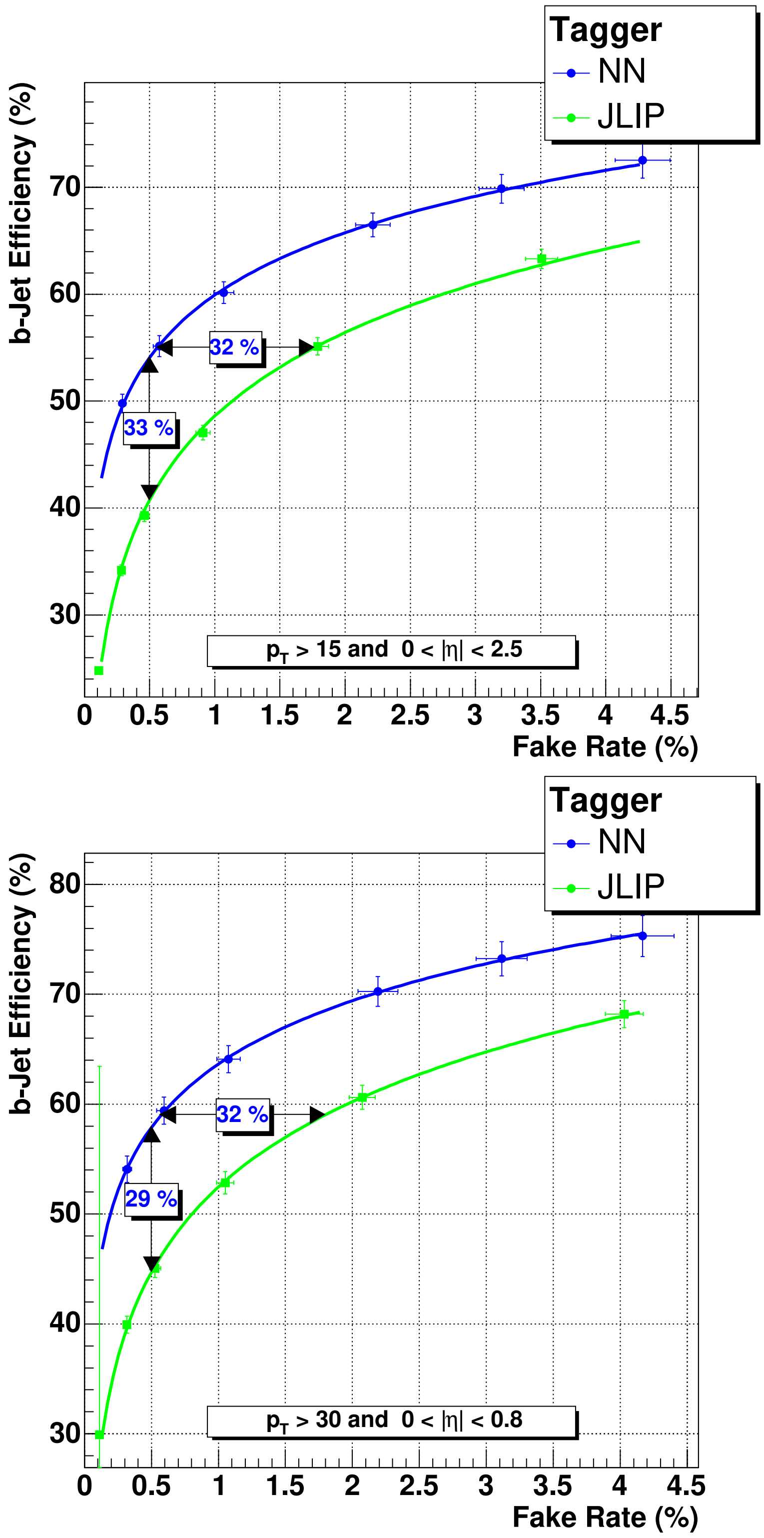

Figure 30: Performance profile of NN and JLIP taggers on the BID and COMB skim. The top plot is for all jets and the bottom plot for jets with $\eta < 0.8$ and $p_T > 30$ GeV. The error on the plots represents the total uncertainty, statistical and systematic, on the performance measurements. The NN tagger demonstrates large performance gains over the JLIP tagger, with increases in efficiency of up to 50% for a fixed fake rate. Fake rates are typically reduced to between a quarter and a third of their value for a fixed signal efficiency.

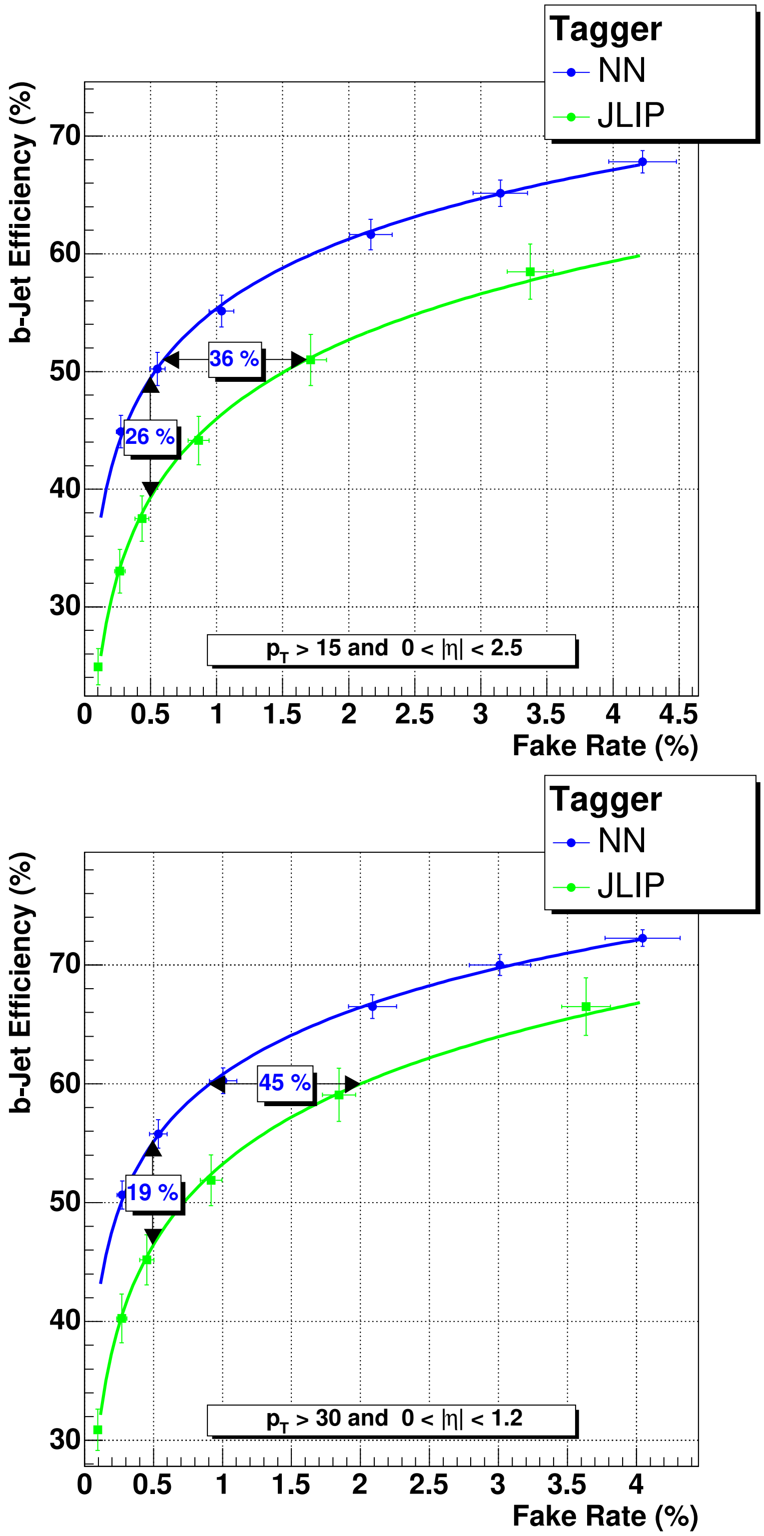

Figure 31: Data performance profile of NN and JLIP taggers on the $Z \rightarrow b\bar{b}$ and $Z \rightarrow q\bar{q}$ MC samples (top) and the QCD $b\bar{b}$ and the QCD $q\bar{q}$ MC samples for $\eta < 1.2$ and $p_T > 30$ GeV (bottom). The performance is calculated from the NN and JLIP TRFs on a jet by jet basis. The error represents the full statistical and systematic error due to the uncertainty with which we know the data efficiency and fake rate for the entire sample when calculated per jet. The NN tagger is seen to have a smaller uncertainty on the b-efficiency measurement but a larger uncertainty on the fake rate for the looser operating points.

References

[1] T. Scanlon "A Neural Network b-tagging Tool", DØNote 4889.

[2] *b*-id Certification Guidelines
http://www-d0.fnal.gov/phys_id/bid/d0_private/certification/p14Pass2/togetcert.html

[3] D. Bloch, B. Clément "Update of the JLIP b-tagger Performance in p14/pass2 with JES 5.3", DØNote 4824.

[4] D. Boline, L. Feligioni, M. Narain "Update on b-quark Jet Identification with Secondary Vertex Reconstruction using DØReco version p14-Pass2 Samples", DØNote 4796.

[5] Pass 2 Skimmed Datasets
http://www-d0.fnal.gov/Run2Physics/cs/skimming/pass2.html

[6] E. Busato, F. Deliot, R. Hauser, J. Stark, R. Stroehmer, P. Verdier, M. Verzocchi, "D0Correct v8", DØNote 4646

[7] *d0root_example* Instructions
http://www-d0.fnal.gov/d0dist/dist/releases/development/d0root_example/doc/

[8] *btags_cert* Instructions
http://www-d0.fnal.gov/d0dist/dist/packages/btags_cert/devel/doc/

[9] Minimum Bias Probability selection of Primary Vertex
http://www-clued0.fnal.gov/∼aran/d0work/mbpsel/how_to_use_PVSelect_MinBiasProb.html

[10] Jets with Tracking
http://www-d0.fnal.gov/∼sch/d0_private/jetWorkshop2002.ps.gz

[11] R. Demina, A. Khanov, F. Rizatdinova, E. Shabalina "Measurement of *b*-tagging efficiency and mis-tagging rates with CSIP method", DØNote

[12] K. Hanagaki, J. Kasper Soft Lefton (muon) tagging method
http://www-d0.fnal.gov/phys_id/bid/d0_private/certification/p14Pass2/SLT/SLT_v1_1.html

[13] B. Clément, D. Bloch, D. Gelé, S. Greder, A.C. Le Bihan, I. Ripp-Baudot, "SystemD or how to get signal, backgrounds and their efficiencies with real data", DØNote 4159.

[14] TMultiLayerPerceptron
http://root.cern.ch/root/html/TMultiLayerPerceptron.html

[15] TMinuit
http://root.cern.ch/root/html/TMinuit.html