CDF/DOC/SEC_VTX/CDFR/8910
July 2007, v1.0

SecVtx Scale Factors and Mistag Matrices for the 2007 Summer Conferences

S. Grinstein and D. Sherman

Summary

In this note we present the SecVtx scale factors and mistag matrices to be used for the 2007 Summer Conferences, corresponding to the 1.8/fb dataset (up to and including period 11). Using the same Monte Carlo that was used for the 1.2/fb results, we conclude that adding the newer data does not affect the tight SecVtx scale factor within it's uncertainty. An additional systematic was added to the Mistag Matrix systematic error to cover the small discrepancy between the 1.2/fb matrix predictions and the tag rates observed in the new data.

Contents

1 SecVtx Scale Factors

The scale factor (SF) is defined as the ratio of the b-tagging efficiency in data to than in Monte Carlo (MC) simulation. A detailed discussion on the electron-method SF and it's the derivation and can be found in [1]. As to the writing of this note, no run dependent MC was generated corresponding to the data taking periods 9 to 11. In this note we simply add the new data and recompute the electron-method SF. A alternative approach (short of generation new MC) is to re-weight the current period 8 MC to cover the periods 8 to 11. This study was done for the tight SecVtx SF and found to be a small ($\sim 0.5\%$) correction [2].

In the following sections we present the validation of the 8 GeV electron data (used in the SF determination), comparisons between the data and the MC (*btopla*), and SF results.

1.1 Data Validation

Figure 1 shows the comparison between period 8 and periods 9, 10 and 11 for electron jets in the 8 GeV electron sample. Though most of the distributions agree very well, the instantaneous luminosity for periods 9 to 11 is larger and that is reflected in the number of quality 12 vertices reconstructed. Figure 2 shows that there are consistent results between the tag rates for the different periods. However, a small drop in the tag rates is observed in the newer data, see Table 3, note that this does not necessarily mean that the SF will drop.

blpc0d	blpc0h	blpc0i	blpc0i'(P8)	P9&P10	P9&P10&P11
5.75 ± 0.03	5.76 ± 0.03	5.82 ± 0.04	5.62 ± 0.03	5.60 ± 0.03	5.59 ± 0.03
19.0 ± 0.3	20.0 ± 0.3	19.5 ± 0.4	19.7 ± 0.3	18.6 ± 0.3	18.7 ± 0.3

Table 1: Electron jet fiducial tag rates (in %) for the different data taking periods. The second row requires a tight tag in the away jet.

If we compare the full 8 GeV electron dataset with the 1.2/fb MC sample (*btopla*), while there is good agreement on most variables, see Figure 3, there are significantly more reconstructed vertices in the data than the MC, this is not s surprise, given that the run-dependent MC only covers up to period 8.

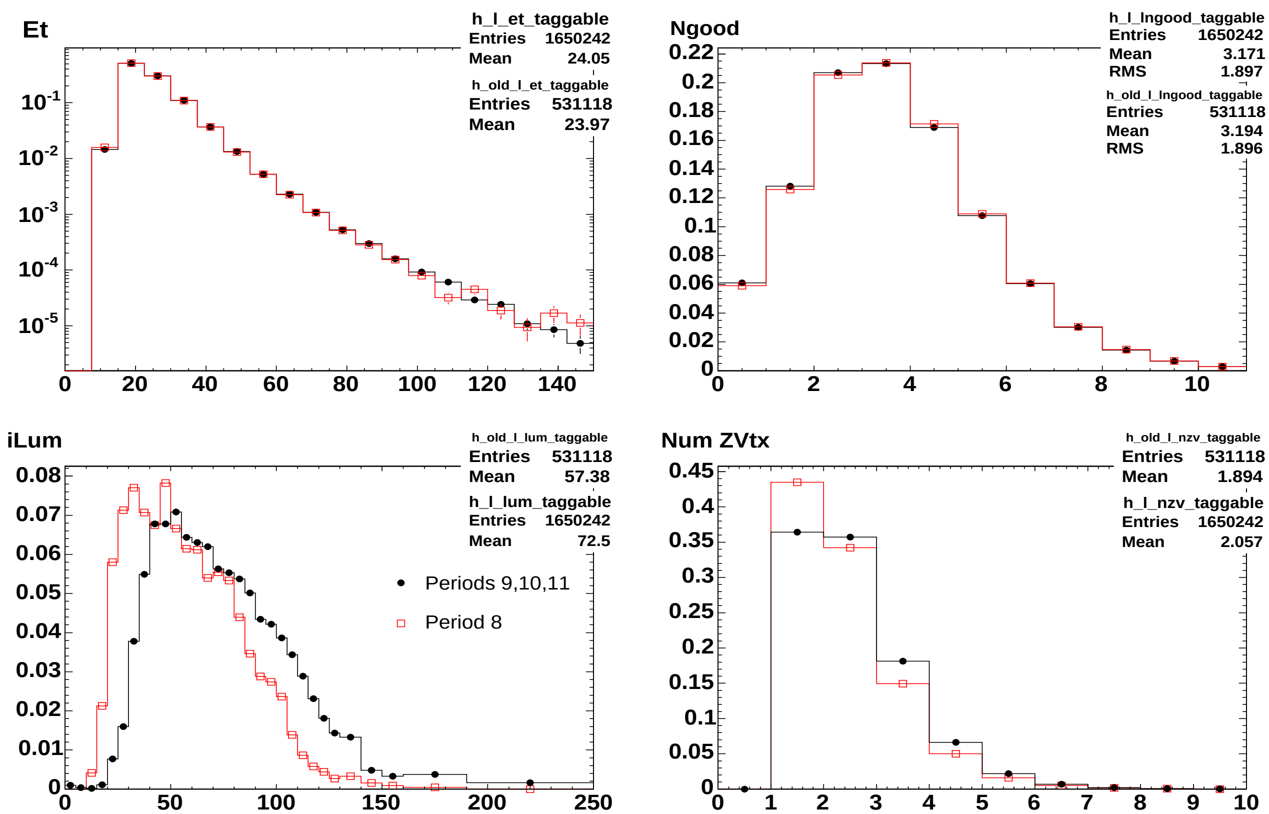

Figure 1: Comparisons between electron jets in period 8 (open squares) and periods 9, 10 and 11 (full circles). Due to the different instantaneous luminosities of these periods, the number of vertices distribution is significantly shifted towards larger values. No prescale correction applied.

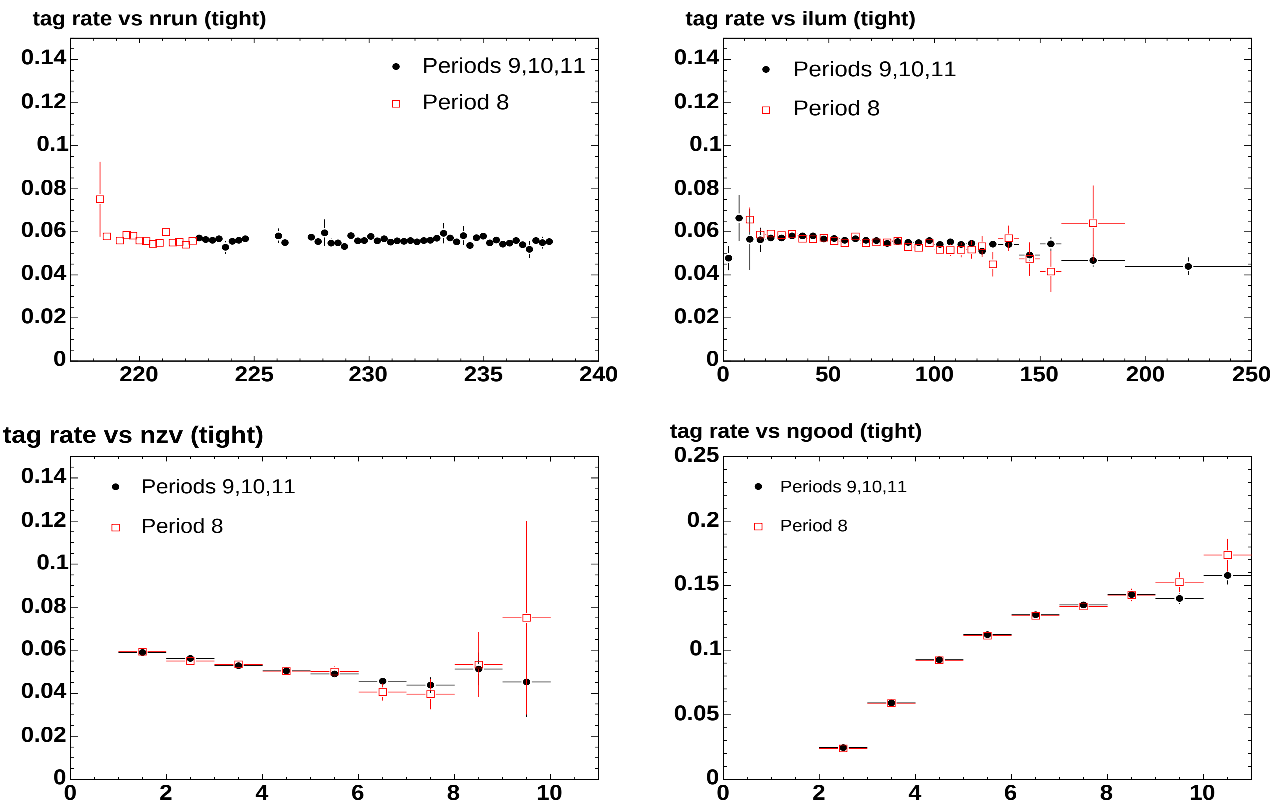

Figure 2: Electron jet tag rates in period 8 (open squares) and periods 9, 10 and 11. There is very good agreement between the two data periods.

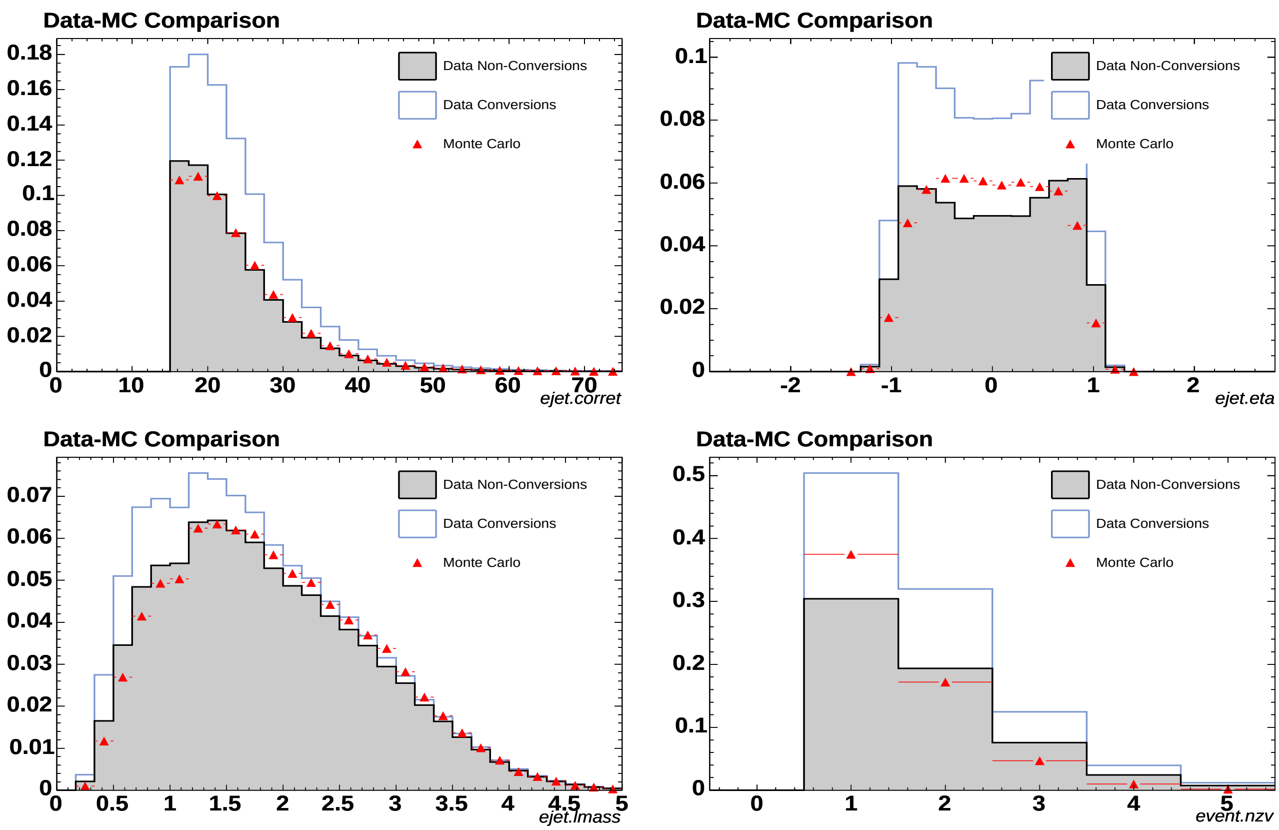

Figure 3: Data and MC distributions of electron jet E_T and η, electron tag mass (loose) and number of reconstructed vertices (quality 12). Conversions are also shown (not filled histogram) in the data. In the MC, the electron jet is required to be matched to a bottom or charm hadron.

	Data	Conversions	*btopla*
N	4491771	1549158	153147
N_+	144976	28694	15294
	Loose Tagger		
N^+	323280	35267	50123
N^-	33661	10226	1584
N_+^+	3287 5	2323	5595
N_+^-	1371	255	63
	Tight Tagger		
N^+	255363	22077	43024
N^-	13546	3845	774
N_+^+	27827	1802	4825
N_+^-	561	100	28

Table 2: Summary of the tag totals in data (including periods 8 to 11). The superscript refers to the electron-jet tag information, and the subscript refers to the away-jet tight tag. The number can be compared with those of Table 1 from reference [1].

1.2 Scale Factor Determination

This section presents the results for the tight SF, for details of the method please refer to [1]. Table 2 shows a summary of the data and Monte Carlos counts for the 8 GeV electron sample used in the scale factor measurement. The Monte Carlo number are unchanged with respect to the 1.2/fb result of [1], since no new run-dependent Monte Carlo has been generated (until this date).

The algebra used to determine the scale factor subtracts negative tags, but to account for the heavy flavor negative tags the $\alpha(N^+_{LF}/N^-_{all})$ is applied. We have not checked the α correction in the electron sample, but we found that in the multi-jet samples the correction did not change significantly (see section 2.2). Using the numbers in Table 2 and applying the same α corrections as in [1] we measure:

$$SF_{tight} = 0.975 \pm 0.008(stat) \pm 0.013(MC)$$

$$SF_{loose} = 0.979 \pm 0.007(stat) \pm 0.012(MC)$$

Figures 4 to 5 show the dependence of the scale factor and measured efficiency on various jet and event properties.

Prescale Correction

The electron data sample used for the scale factor determination is collected with the ELECTRON_CENTRAL_8 trigger path, which has a dynamic prescale at Level 2. The prescale changes from 100 at the high luminosity region, to no prescale when the luminosity fall below $\sim 40E30$. However, when a run is started the pre-scales fall

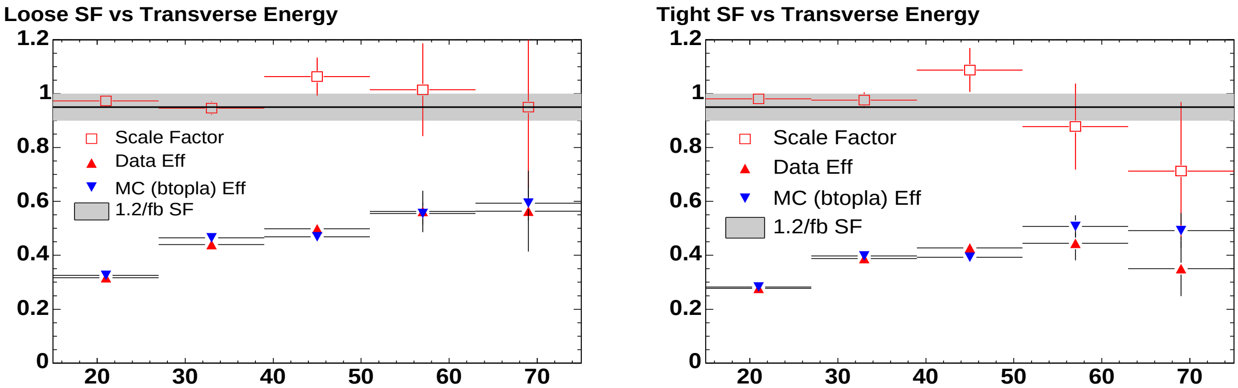

Figure 4: Dependence of the scale factor and efficiencies on the electron jet E_T for the loose (left) and tight taggers. The results include the full data sample (up to and including period 11). Errors are statistical only. The 1.2/fb result is also included.

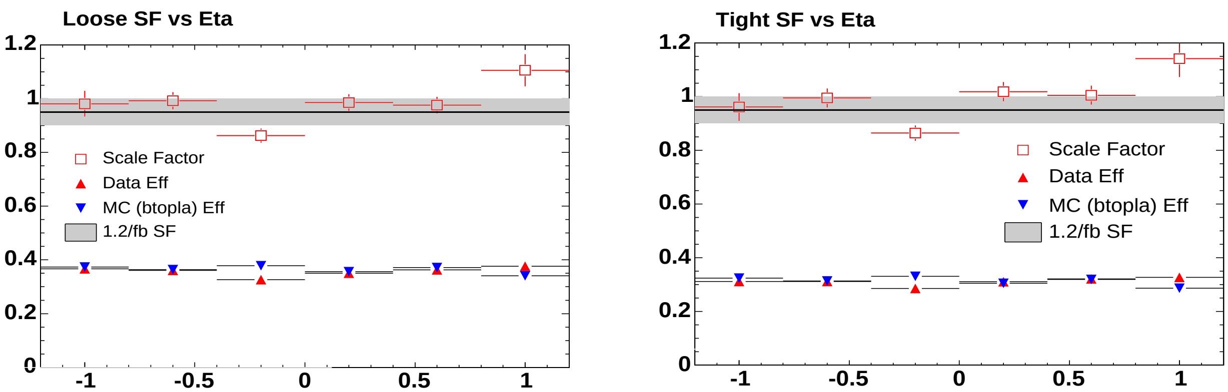

Figure 5: Dependence of the scale factor and efficiencies on the electron jet η for the loose (left) and tight taggers. The results include the full data sample (up to and including period 11). Errors are statistical only. The 1.2/fb result is also included.

to their highest default value until ScalerMon makes adjustments, this explains the fact that there are large pre-scales at low luminosity data (see Figure 11). The run-dependent Monte Carlo is generated based solely on the total integrated luminosity, so the event weighting will be inconsistent between data and simulation. To account for both run and luminosity dependent effects in the efficiency, we weight events by the trigger prescale at the instant the event was recorded - Tom Wright has provided the necessary mapping between pre-scales and timestamps to perform the weighting.

After re-weighting we find:

$$SF_{tight} = 0.956 \pm 0.006(stat) \pm 0.014(MC)$$

which differs in about 2% with the raw result.

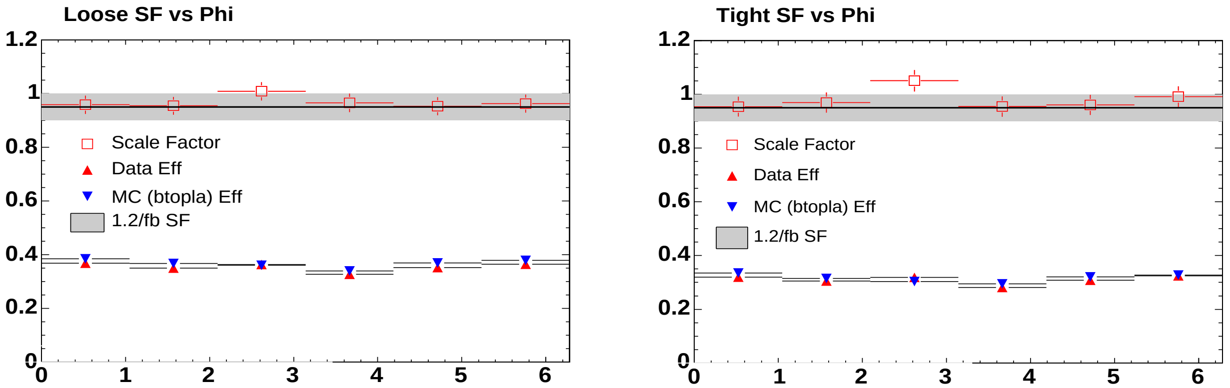

Figure 6: Dependence of the scale factor and efficiencies on the electron jet ϕ for the loose (left) and tight taggers. The results include the full data sample (up to and including period 11). Errors are statistical only. The 1.2/fb result is also included.

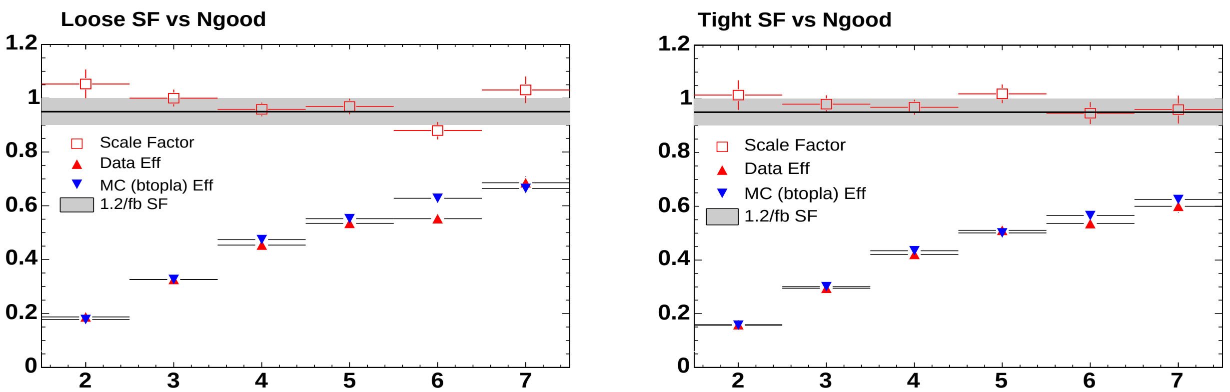

Figure 7: Dependence of the scale factor and efficiencies on the number of good tracks in the electron jet for the loose (left) and tight taggers. The results include the full data sample (up to and including period 11). Errors are statistical only. The 1.2/fb result is also included.

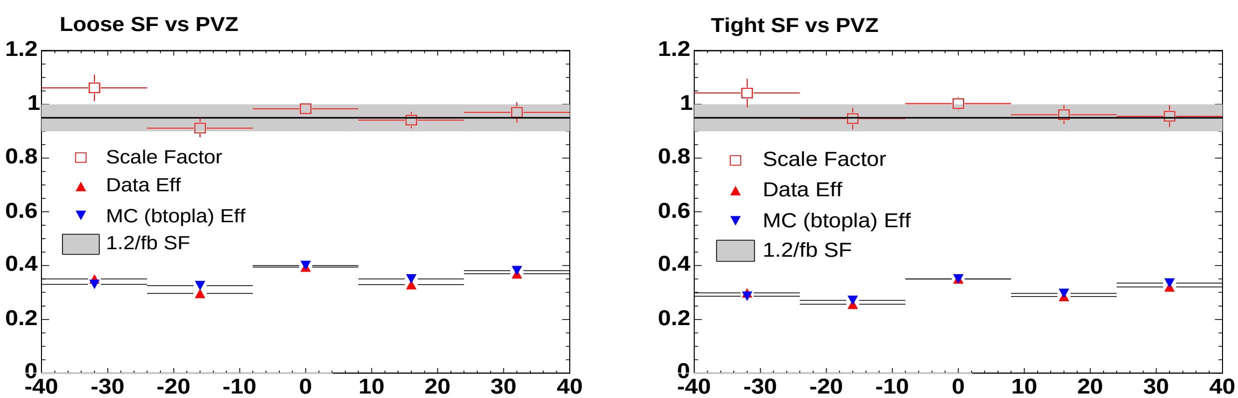

Figure 8: Dependence of the scale factor and efficiencies on the event primary vertex position for the loose (left) and tight taggers. The results include the full data sample (up to and including period 11). Errors are statistical only. The 1.2/fb result is also included.

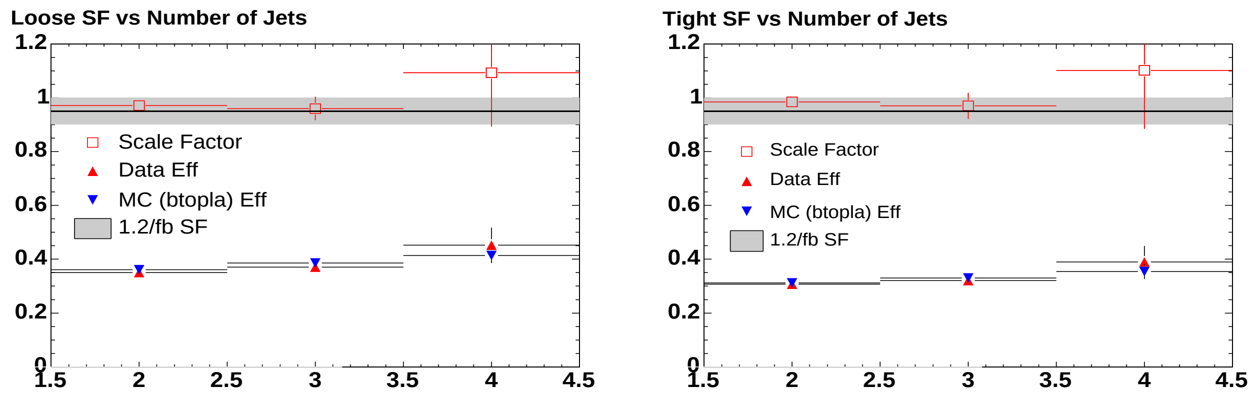

Figure 9: Dependence of the scale factor and efficiencies on the number of jets for the loose (left) and tight taggers. The results include the full data sample (up to and including period 11). Errors are statistical only. The 1.2/fb result is also included.

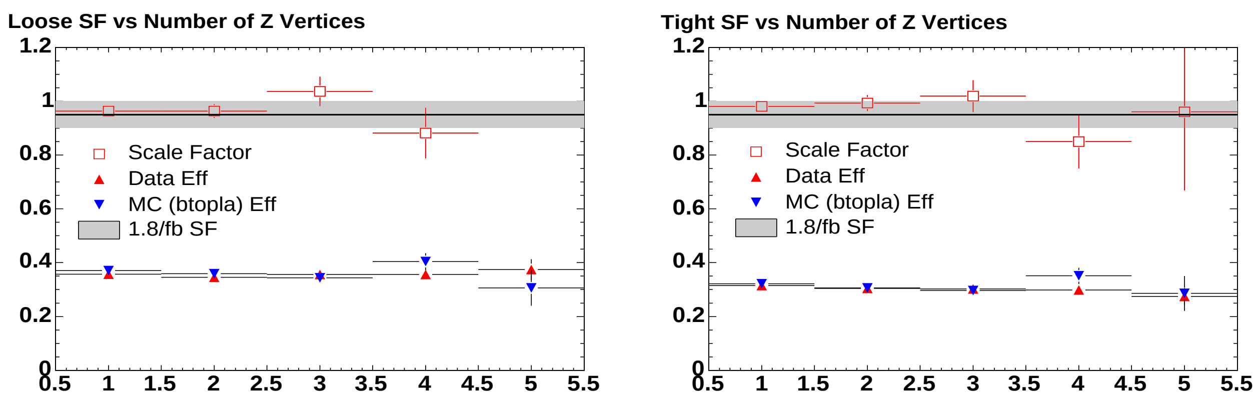

Figure 10: Dependence of the scale factor and efficiencies on the number of reconstructed vertices for the loose (left) and tight taggers. The results include the full data sample (up to and including period 11). Errors are statistical only. The 1.2/fb result is also included.

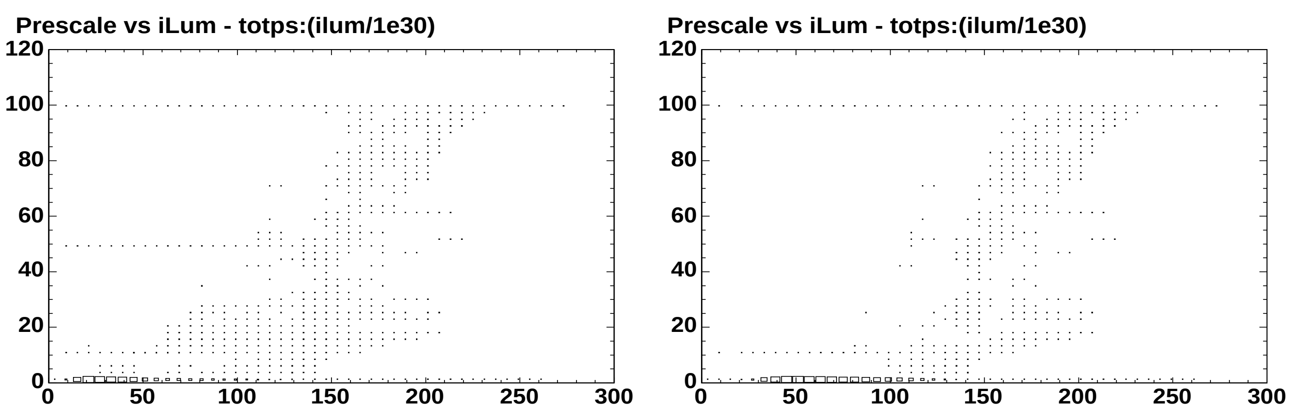

Figure 11: Scatter plot of trigger pre-scales versus instantaneous luminosity for the full data sample (left) and only for periods 9 to 11.

1.3 Re-weighting of Monte Carlo

Following a prescription from Kevin Lannon [3] we re-weighted the period 8 *btopla* MC to cover periods 8 through 10. While there is almost no effect in most of the kinematic variables (jet η, E_T, number of good tracks, primary vertex position, etc), there is clearly better agreement between the number of reconstructed vertices between the data and the simulation after re-weighting, as shown in Figure 12.

If we re-weight the Monte Carlo with the weights from [3] to cover the data periods 9 to 11, we find:

$$SF_{tight} = 0.968 \pm 0.008(stat) \pm 0.012(MC)$$

which differs in less than 1% with the raw result.

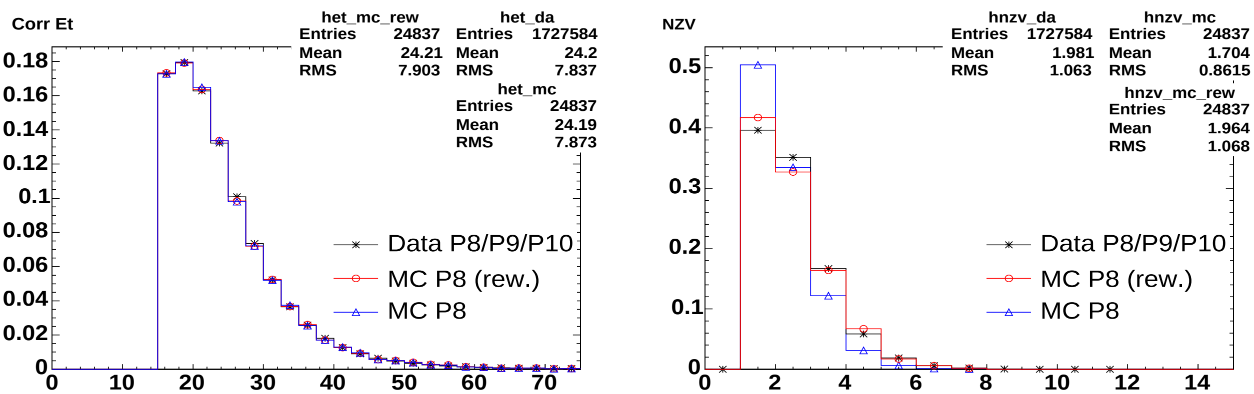

Figure 12: Monte Carlo re-weighting. While there is almost no effect in the E_T distribution, there is a notably improvement in the agreement between the number of vertices in the data and MC.

2 Mistag Matrices

In this section we validate the new jet data (periods 9 to 11), check the tag rates and compare the mistag predictions from the 1.2/fb mistag matrix and the observed rates in periods 9 to 11. The difference between the observed and predicted tag rates is added as a systematic to the 1.2/fb mistag matrix.

2.1 Mistag Matrix Studies in Periods 9 to 11

Figure 14 shows the effect of the different instantaneous luminosities in period 8 (included in the 1.2/fb matrix) and periods 9 to 11, while there is little effect in some variables, like jet E_T, the number of reconstructed vertices differ significantly. This is the expected major source of disagreement between the tag rates in jet data in 1.2/fb tag rates and periods 9 to 11.

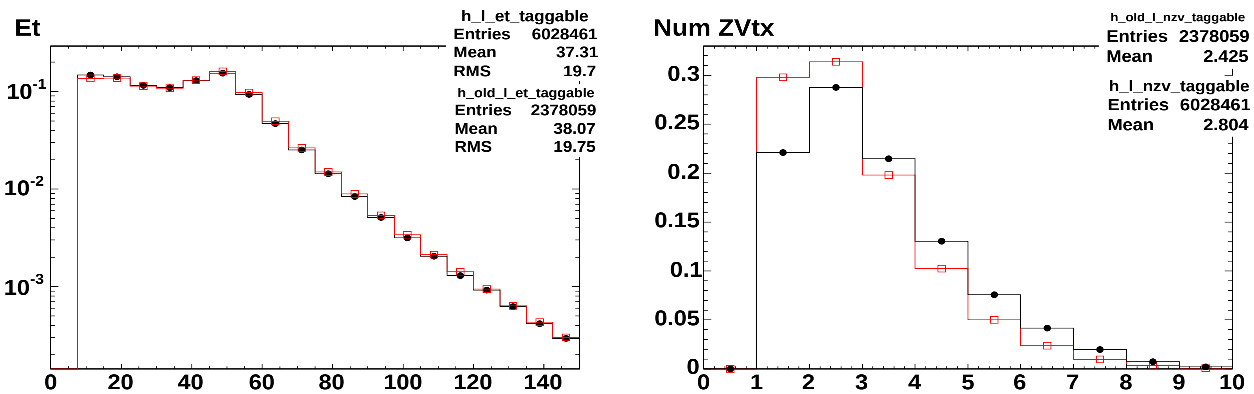

Figure 13: Monte Carlo re-weighting. While there is almost no effect in the E_T distribution, there is a notably improvement in the agreement between the number of vertices in the data and MC.

Table 3 presents the total tag rates in the 1.2/fb and the period 9 to 11 jet data samples. The 1.2/fb tag rates are very similar to the ones in CDF8264 [4], except for the extra data (8264 only uses the 1/fb data sample). Figures **??** and **??** show the negative tag rate in Jet 50 for periods 8 and 9 to 11. Though there is good agreement between period 8 and the recent data, the 1.2/fb matrix is expected to underestimate the tag rates for periods 9 to 11.

Figures 15 and 17 show the 1.2/fb matrix predictions and observed tag rates in periods 9 to 11 for the tight tagger, for variables which are part of the mistag matrix parameterization, we also show the predicted/observed rate ratios. Figures 18 and 19 show the same comparisons but for the number of jets and the instantaneous luminosity, which are not part of the matrix variables. At high luminosity the discrepancy between the predicted and observed rates is quite large (about 25% above $200E30$), this is caused by the fact that the number of reconstructed vertices (which was included in the matrix recently to account for luminosity effects, replacing the

Trigger	Periods 9 to 11	1.2/fb result
Loose Positive		
All	0.02906 ± 0.00002	0.03138 ± 0.00002
Jet20	0.00850 ± 0.00002	0.01118 ± 0.00003
Jet50	0.02432 ± 0.00005	0.02875 ± 0.00005
Jet70	0.03650 ± 0.00004	0.03867 ± 0.00005
Jet100	0.04968 ± 0.00005	0.05060 ± 0.00006
Loose Negative		
All	0.01420 ± 0.00001	0.01397 ± 0.00002
Jet20	0.00336 ± 0.00001	0.00370 ± 0.00001
Jet50	0.01135 ± 0.00004	0.01193 ± 0.00003
Jet70	0.01789 ± 0.00002	0.01748 ± 0.00004
Jet100	0.02556 ± 0.00004	0.02443 ± 0.00004
Tight Positive		
All	0.01627 ± 0.00001	0.01796 ± 0.00002
Jet20	0.00511 ± 0.00001	0.00700 ± 0.00002
Jet50	0.01394 ± 0.00004	0.01688 ± 0.00004
Jet70	0.02044 ± 0.00003	0.02205 ± 0.00004
Jet100	0.02717 ± 0.00004	0.02799 ± 0.00004
Tight Negative		
All	0.00586 ± 0.00001	0.00568 ± 0.00001
Jet20	0.00129 ± 0.00001	0.00138 ± 0.00001
Jet50	0.00455 ± 0.00002	0.00475 ± 0.00002
Jet70	0.00735 ± 0.00002	0.00709 ± 0.00002
Jet100	0.01079 ± 0.00002	0.01018 ± 0.00003

Table 3: Summary of tight and loose tag rates for the various jet triggers used in the mistag matrix. The left column corresponds to data from periods 9 to 11 only, while the right is the 1.2/fb rate (up to and including period 8).

jet ϕ parameter) is not longer linear with luminosity in this regime, this is shown in Figure 20.

Finally, Figure 21 shows the ratio of the 1.2/fb matrix predicted tag rates and the observed rates in periods 9 to 11 for the tight and loose taggers. We take the average of all the jet triggers and weight the effect by the amount of data in periods 9 to 11 compared to the 1.2/fb mistag matrix dataset. This results in an additional systematic of 3% for the loose and tight negative tag rate errors. We have included this additional error to **BTagObjects**, tag *btag_1700invpb_v1*.

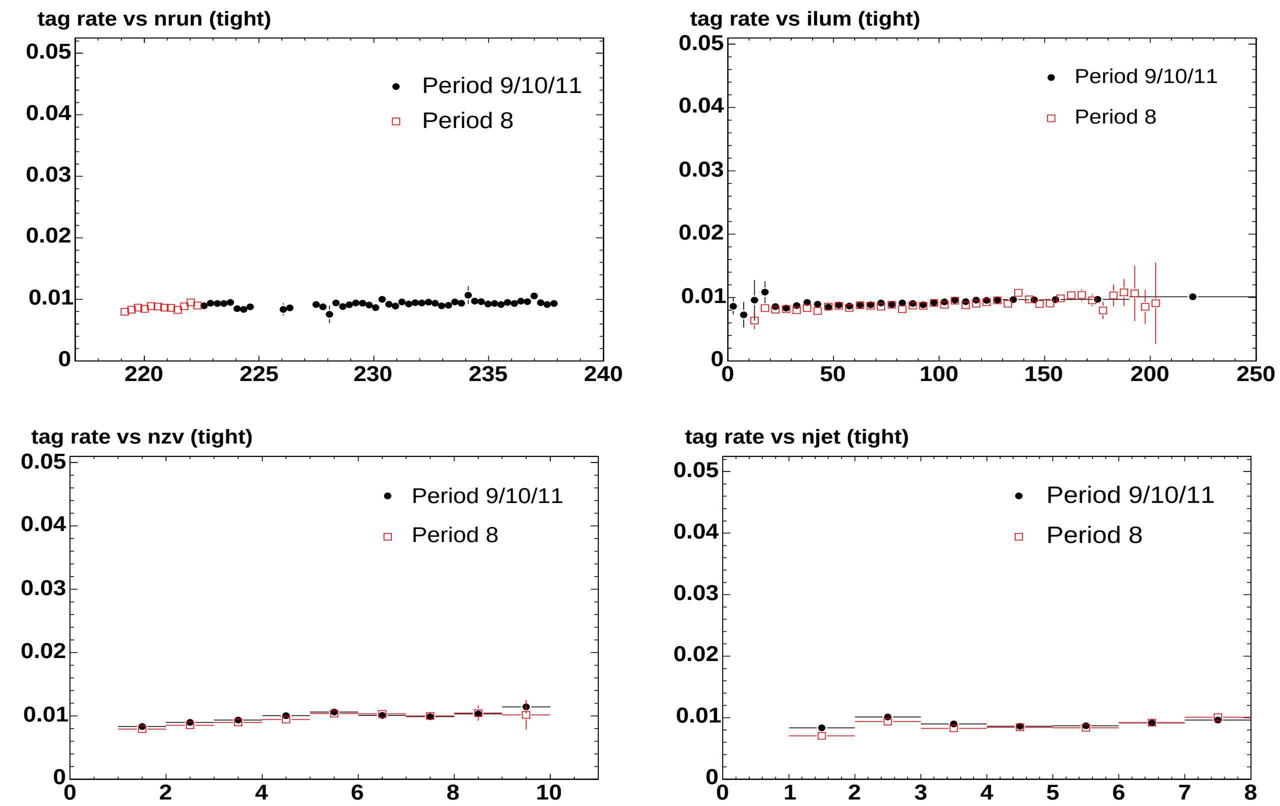

Figure 14: Negative tag rates in period 8 and periods 9 to 11 versus run number, instantaneous luminosity, number of reconstructed vertices and number of jets.

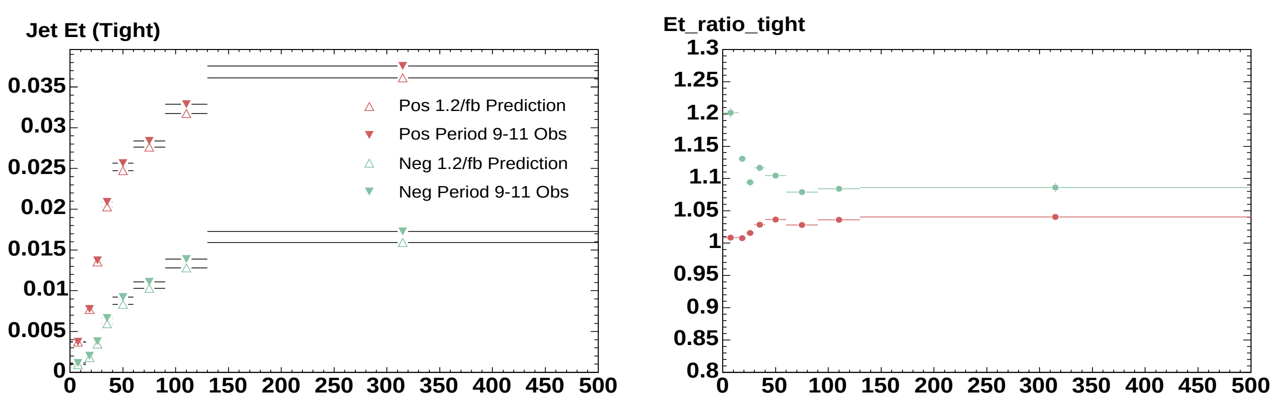

Figure 15: Predicted (with the 1.2/fb matrix) and observed tag rates in periods 9 to 11.

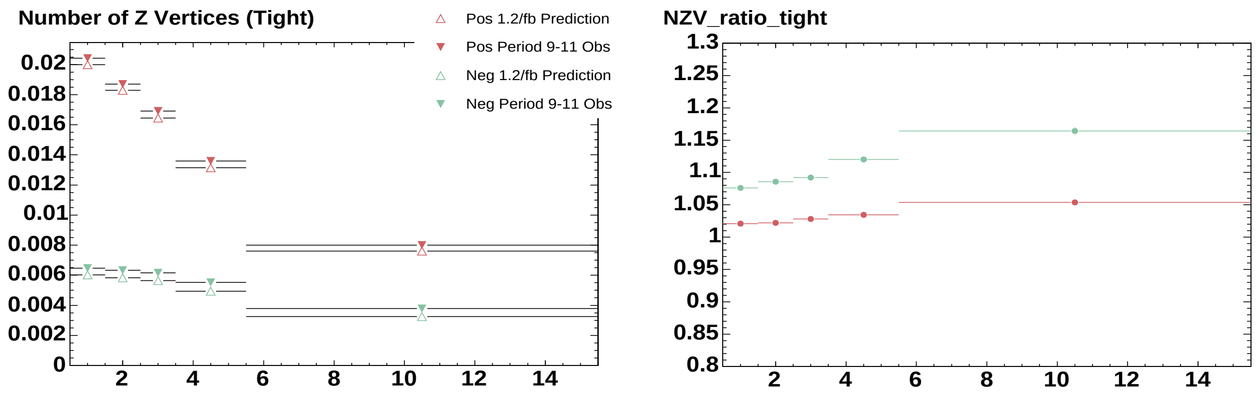

Figure 16: Predicted (with the 1.2/fb matrix) and observed tag rates in periods 9 to 11.

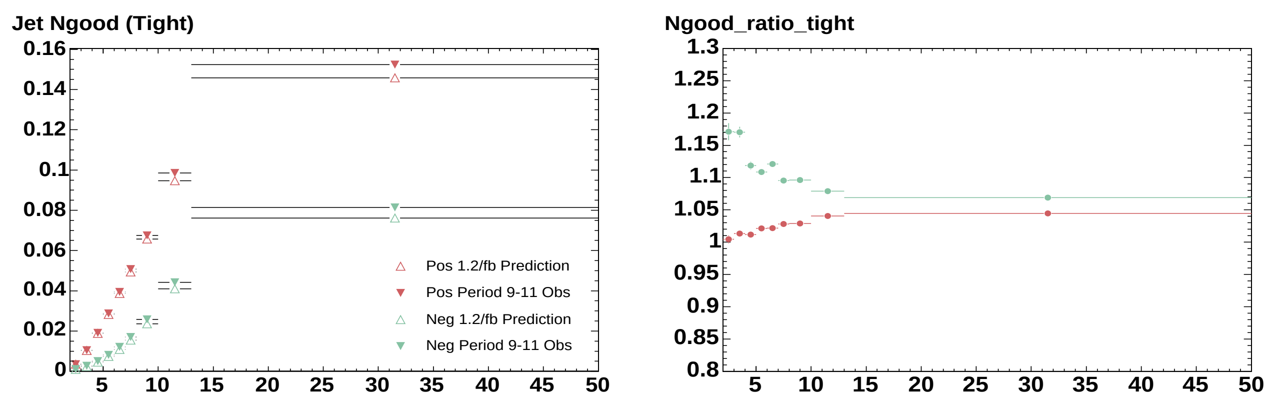

Figure 17: Predicted (with the 1.2/fb matrix) and observed tag rates in periods 9 to 11.

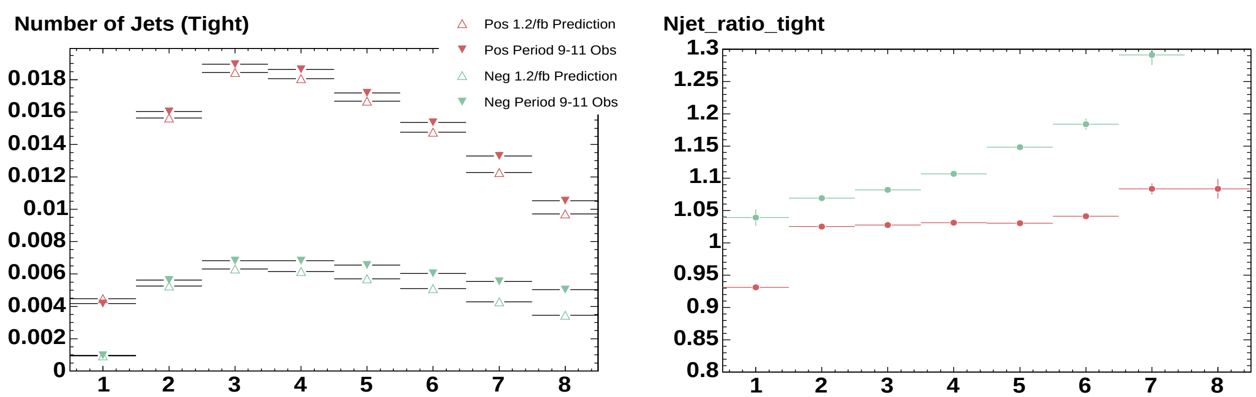

Figure 18: Predicted (with the 1.2/fb matrix) and observed tag rates in periods 9 to 11.

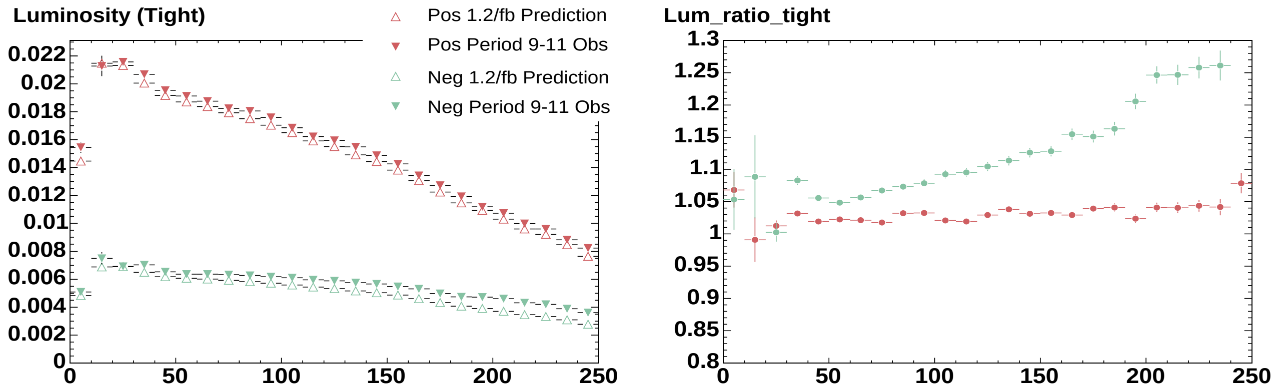

Figure 19: Predicted (with the 1.2/fb matrix) and observed tag rates in periods 9 to 11.

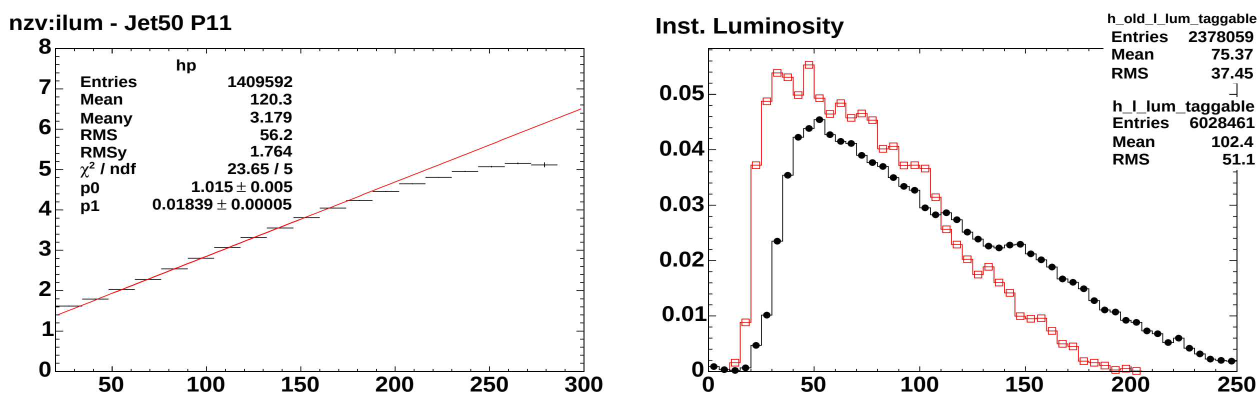

Figure 20: The plot on the left shows that the linearity between luminosity and number of reconstructed vertices break at about $200E30$, however, only a small fraction of the data is collected at such high luminosities (no prescale correction applied to these plots).

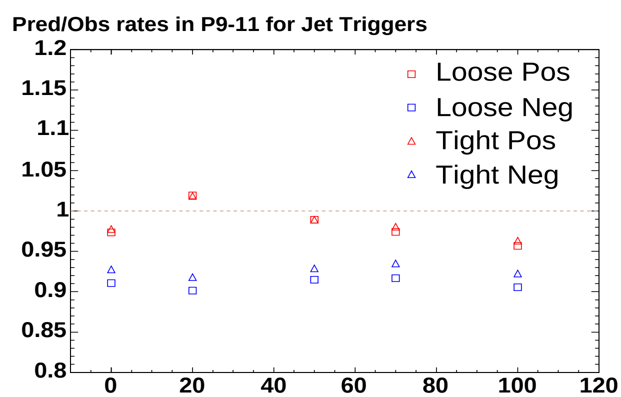

Figure 21: The ratio of the 1.2/fb matrix predicted tag rates and the observed rates in periods 9 to 11. The x-axis corresponds to the jet trigger, while the entry at zero is the average of all the jet triggers.

2.2 Asymmetry Correction

The negative tag rate extracted from the generic jet data underestimates the positive tag light flavor content. Though resolution effects are expected to by symmetric, other sources bias the light flavor tags positive. Therefore an asymmetry correction is applied to the negative mistag rate extracted from the generic jet data so that:

$$\alpha\beta R^{-}_{mistag} = N^{+}_{LF}/N^{pretag}_{LF}$$

Where $\alpha = N^{+}_{LF}/(N^{-}_{all})$ is the largest correction which accounts for the heavy flavor contribution to the negative tagged sample. For more details please refer to [5].

To determine α tagged distributions are fitted with Monte Carlo templates. Since the L_{2D} tail is poorly modeled in the simulation, the tag excess (positive - negative tags) is fitted, thus the light flavor template is dominated by material interactions and K_s/Λ decays. From the vertex mass distributions fits the heavy flavor fractions are extracted and used to calculate α. Such a fit for the tight tagger using the Jet 20 sample is shown in Figure 22. The result of the fit is presented in Table 4. We find little difference in the α corrections measured with the 1.2/fb data sample (from [5]) and the corrections determined using the 1.5/fb sample, which includes up to period 10. Thought we have not explicitly check the α jet E_T dependence and have not included period 11, we feel that there is no need to introduce any new systematic given the excellent agreement between the 1.2/fb and 1.5/fb α numbers shown in Table 5.

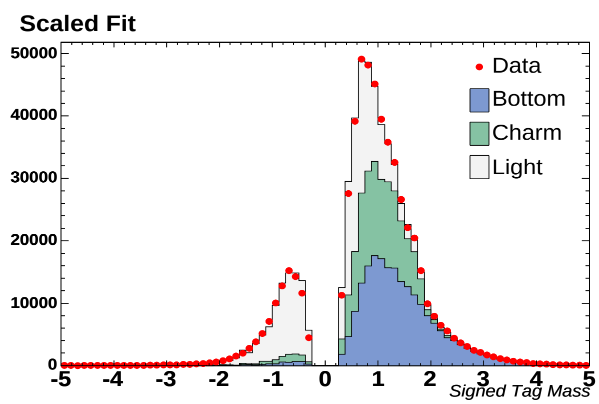

Figure 22: The vertex mass fit for Jet 20 and the tight tagger. The plot shows the unfolded result (the fit is performed on the excess tags).

3 Conclusions

We presented the SecVtx scale factors and mistag matrices to be used for the 2007 Summer Conferences, corresponding to the 1.8/fb dataset (up to and including period 11). We recommend using the same scale factor as for the 1.2/fb results

(kJets)	Positive	Negative	Excess
Data	466.4	95.2	371.1
Bottom	204 ± 5	4.2 ± 0.1	200 ± 5
Charm	133 ± 6	7.1 ± 0.3	127 ± 6
Light	129 ± 9	84 ± 6	448 ± 3
	$\chi^2/ndof = 0.91$		

Table 4: Fit vertex mass results for the tight tagger using Jet 20 data up to and including period 10. The corresponding value of α is 1.35 ± 0.09.

Trigger	Tagger	$\alpha(1.2/fb)$	$\alpha(1.5/fb)$
Jet20	Tight	1.36 ± 0.09	1.35 ± 0.09
Jet20	Loose	1.30 ± 0.05	1.29 ± 0.05
Jet50	Tight	1.34 ± 0.05	1.34 ± 0.05
Jet50	Loose	1.27 ± 0.03	1.26 ± 0.03
Jet70	Tight	1.44 ± 0.09	1.43 ± 0.07
Jet70	Loose	1.26 ± 0.05	1.24 ± 0.04
Jet100	Tight	1.41 ± 0.06	1.41 ± 0.05
Jet100	Loose	1.36 ± 0.04	1.39 ± 0.05

Table 5: Results of the α asymmetry correction for the 1.2/fb and the 1.5/fb (including period 10) data samples.

(0.95 ± 0.05 for the tight and loose taggers, and 0.88 ± 0.05 for the ultra-tight tagger [6]). An additional systematic was added to the Mistag Matrix systematic error to cover the small discrepancy between the 1.2/fb matrix predictions and the tag rates observed in the new data. This additional error is included in **BTagObjects** *cvs* tag *btag_1700invpb_v1*.

References

[1] Grinstein, Guimaraes da Costa, Sherman, "The SecVtx electron-method scale factor for Winter 2007", CDF8625.

[2] Grinstein, Joint Physics Meeting talk, 18th July 07.

[3] Kevin Lannon, http://glast2.mps.ohio-state.edu/lannon/81

[4] J. Guimaraes da Costa and D. Sherman, "Loose and Tight SecVtx Tag Matrices with 1/fb", CDF8264.

[5] S. Grinstein, J. Guimaraes da Costa and D. Sherman, "SecVtx Mistag Asymmetry for Winter 2007", CDF8626.

[6] See http://www-cdf.fnal.gov/internal/physics/top/RunIIBtag/.