SecVtx Tag Rate Matrix : BLESSING

Speaker: Salvatore Rappoccio

Ayana Holloway, João Guimarães da Costa, Aaron Dominguez for the B-Tag Group

July 11, 2003

Joint Physics Mtg.

- Overview
- Issues Raised at Preblessing
- Cross-Checks
- Blessing Material

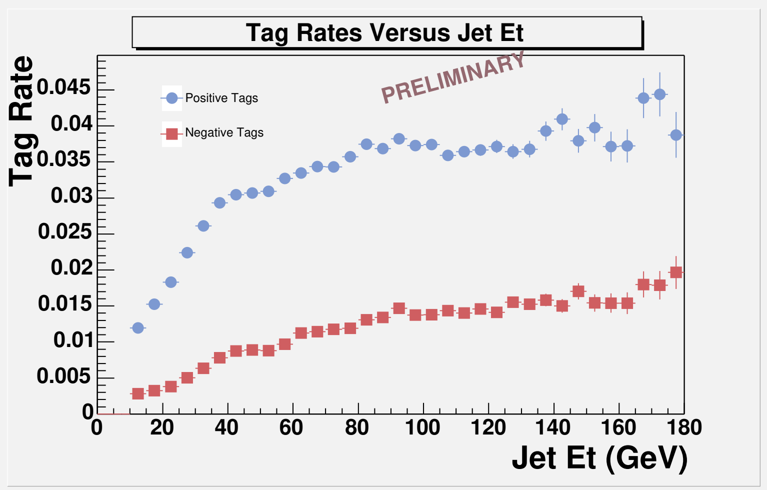
Overview

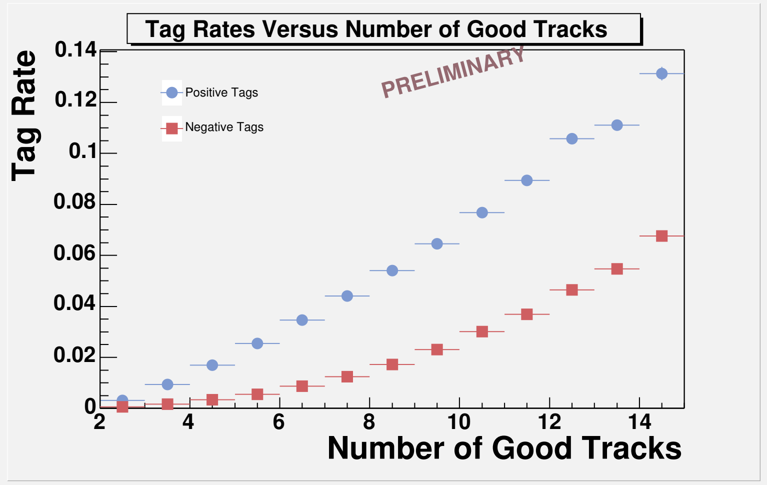
- SecVtx :
  - PrimeVtx with beam constraint, z-dependendence, time dependence of beamline
  - Using `4.11.1` with `TrackRefit`, latest alignment table (`100035 2 GOOD`)
  - Increases in (both) tag rates due to improved track refits in z and use of PrimeVtx .
- Mistag Outline
  - Used to estimate "Method 1" backgrounds for b-tagging analyses
  - 5-D histogram to represent parameterization of tagging rate as a function of E_T, N_{gd}, $\sum E_T$, η, ϕ
  - The rate matrix is defined for jets with ≥ 2 good tracks (i.e. *taggable* jets).
  - Tag is determined from $|L_{xy}/\sigma_{L_{xy}}| > 3.0$
  - Sign is determined from cosine of angle with jet axis
  - Rate is defined as:

$$\boxed{R_{+,-} \equiv \frac{N^{+,-}_{tagged\ jets}}{N_{taggable\ jets}}}$$

- Changes from preblessing: Now have 108 pb^{-1} of data (gjet08/gjet09)
  - Processed *old bad alignment data*: 152595-154012
  - Reprocessed crashed runs, some overlap in triggers
- Reduced number of bins

For Blessing : Tag Rates Plots





Issues Raised At Preblessing

1. Histogram density (number of low statistics bins)
2. Further $\sum E_T$ Sample Cross Checks
3. Using a matrix derived from $\mathrm{Jet50}$ on other samples (`Jet20`)
4. MC comparison

Issue 1: Histogram Low Statistics Bins

- Finer binning for preblessing : 46080 bins

$$\begin{aligned}
\sum E_T &: (0, 80, 140, > 220) \\
N_{good\ tracks} &: (2, 3, 4, 5, 6, 8, 10, 13, > 13) \\
Jet\ E_T &: (0, 20, 35, 50, 65, 80, 100, 120, 150, 180, > 180) \\
Jet\ \phi &: (-15^\circ, 15^\circ, 45^\circ, 75^\circ, 105^\circ, 135^\circ, 165^\circ, 195^\circ, 225^\circ, 255^\circ, 285^\circ, 315^\circ, 345^\circ) \\
Jet\ \eta &: (-2.4, -1.1, -0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 1.1, 2.4)
\end{aligned}$$

- Reduce binning : 5376 bins

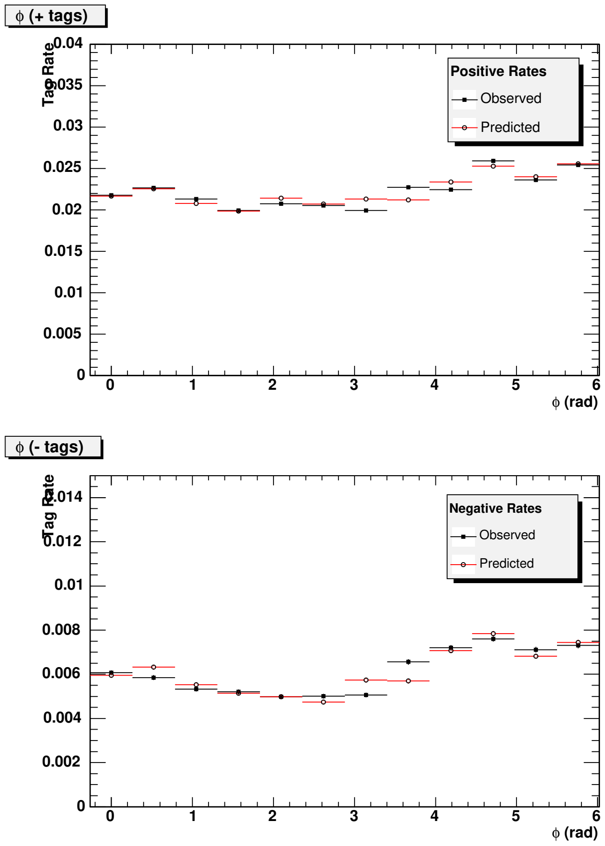
$$\begin{aligned}
\sum E_T &: (0, 80, 140, 220, > 220) \\
N_{good\ tracks} &: (2, 3, 4, 5, 6, 7, 8, 10, 13, > 13) \\
Jet\ E_T &: (10, 15, 22, 30, 40, 60, 90, 130, > 130) \\
Jet\ \phi &: (-15^\circ, 45^\circ, 105^\circ, 165^\circ, 225^\circ, 285^\circ, 345^\circ) \\
Jet\ \eta &: (0.0, 0.4, 0.8, 1.1, 2.4)
\end{aligned}$$

- Binned finer in low E_T region, added bin for $E_T = 10\ GeV$, $N_{good\ tracks}$, reduced binning in η and ϕ. Examine $|\eta|$, and every other wedge in ϕ
- Not all bins are physically possible

Issue 1: Histogram Low Statistics Bins

- Statistics are better, examine effects on predictability

ϕ, Observed, Predicted For New Matrix:



Still predicting tag rates well

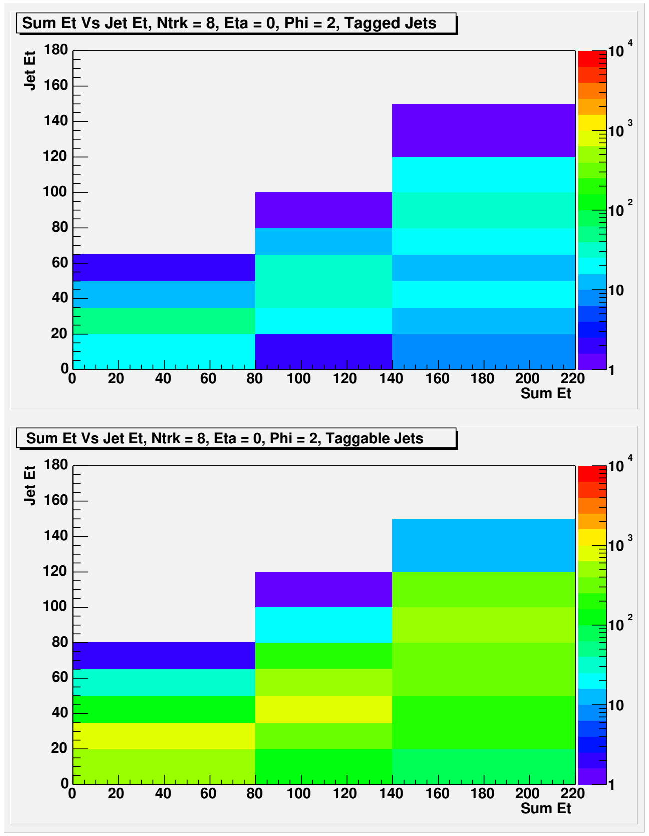
Issue 1: Histogram Low Statistics Bins

- To examine phase space coverage, select out three dimensions, and plot in the other two.
- Select $N_{good\ tracks} = 8$, $\eta = 0$, $\phi = 2.0$
- Plot $\sum E_T$ versus jet E_T
- Phase space in populated regions have good statistics in the tag rate histograms

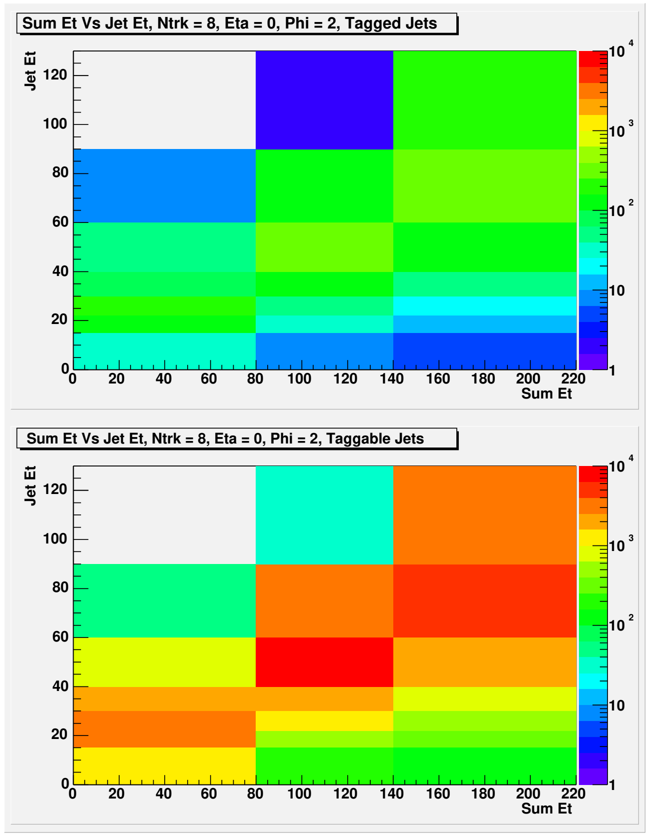
Issue 1 : Jet E_T (y-axis) Versus $\sum E_T$ (x-axis)

One Bin in matrix : $N_{good\ tracks} = 8$, $\eta = 0$, $\phi = 2.0$

More Bins : 46080



Fewer Bins : 6912



Issue 1: Histogram Low Statistics Bins

- Examine predictability in $\sum E_T$ sample using the entire jet sample matrix

New Matrix (Fewer Bins):

$$R^+_{Pred} = 0.02619 +/- 0.00011$$
$$R^+_{Obs} = 0.02757 +/- 0.00011$$
$$\frac{R^+_{Obs}}{R^+_{Pred}} = 1.0523 +/- 0.00590$$
$$R^-_{Pred} = 0.00726 +/- 0.00006$$
$$R^-_{Obs} = 0.00712 +/- 0.00005$$
$$\frac{R^-_{Obs}}{R^-_{Pred}} = 0.98073 +/- 0.01211$$

Old Matrix (More Bins):

$$R^+_{Pred} = 0.02623 +/- 0.00019$$
$$R^+_{Obs} = 0.02757 +/- 0.00011$$
$$\frac{R^+_{Obs}}{R^+_{Pred}} = 1.05089 +/- 0.00867$$
$$R^-_{Pred} = 0.00725 +/- 0.00015$$
$$R^-_{Obs} = 0.00712 +/- 0.00005$$
$$\frac{R^-_{Obs}}{R^-_{Pred}} = 0.98266 +/- 0.02138$$

- Errors are lower using the new matrix with fewer bins
- Predictability is the same

Issue 2: $\sum E_T$ Sample Cross Checks

- Predict tags in $\sum E_T$ sample to estimate systematic uncertainty using two different matrices: `Jet50` only, all events of all jet samples.
- Overall systematic difference is determined from the ratio

$$\alpha = \frac{R_{observed}}{R_{predicted}}$$

- `Jet50` Matrix Versus $\sum E_T$ Sample:

$$\begin{aligned} \alpha_+ &= 1.082 \pm 0.010 \\ \alpha_- &= 1.026 \pm 0.023 \end{aligned}$$

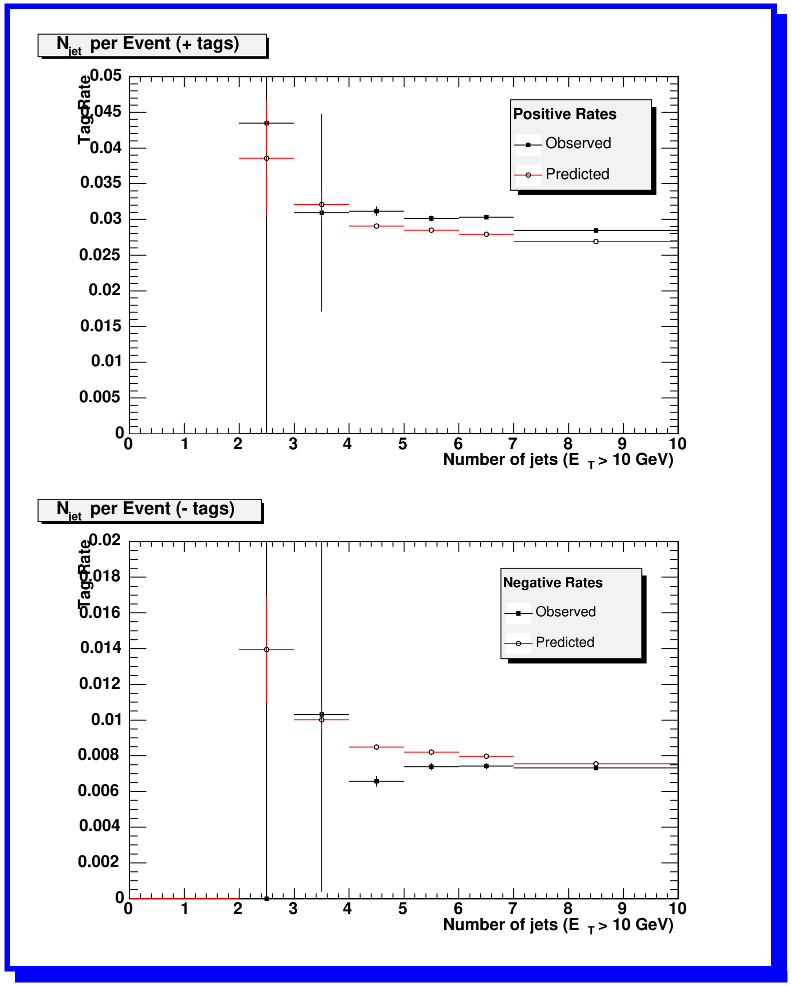
- All Jet Matrix Versus $\sum E_T$ Sample:

$$\begin{aligned} \alpha_+ &= 1.0524 \pm 0.0059 \\ \alpha_- &= 0.980 \pm 0.012 \end{aligned}$$

- Will summarize these numbers later

Issue 2: $\sum E_T$ Sample Cross Checks

Number of Jets, $\sum E_T$ Sample and Jet Samples



Issue 3: Check Trigger Sample Bias

- Check tag rates in `Jet20` using `Jet50`-only matrix
- Check tag rates in odd events of `Jet20` using matrix from even events of all-jet samples.
- `Jet50` Matrix vs `Jet20`:

$$\begin{aligned} \alpha_+ &= 0.818 \pm 0.013 \\ \alpha_- &= 0.798 \pm 0.031 \end{aligned}$$

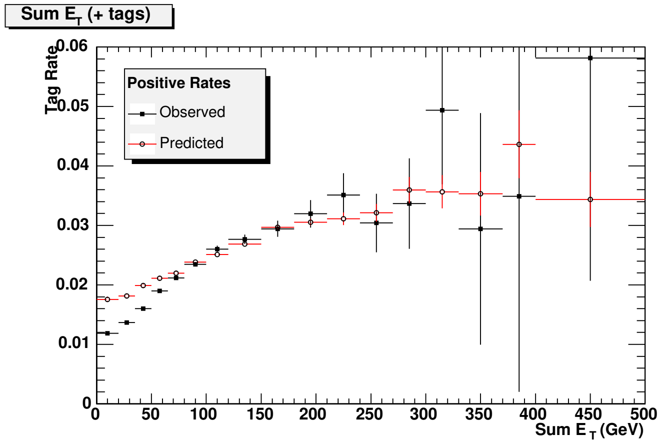
- All Jet Even Matrix vs Odd, `Jet20`:

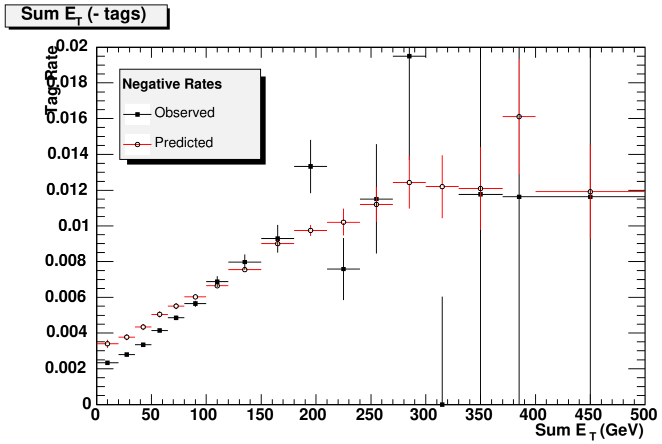
$$\begin{aligned} \alpha_+ &= 1.0039 \pm 0.0055 \\ \alpha_- &= 0.990 \pm 0.012 \end{aligned}$$

- `Jet20` is not a typical "physics" sample for top analysis : *Very gross overestimate of systematic*
- Will summarize these numbers later

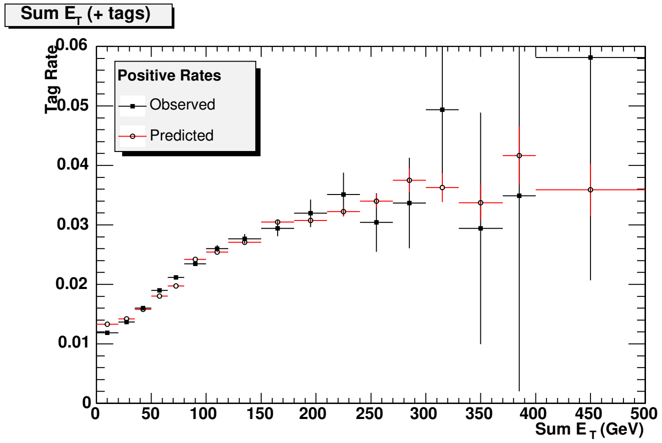
Issue 3: Check Trigger Sample Bias

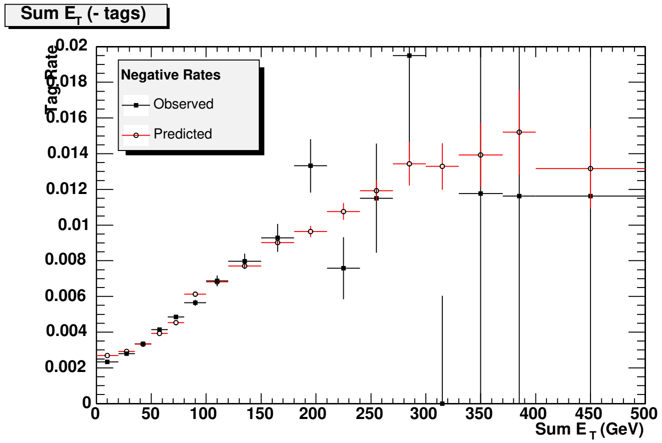
Jet50 Versus Jet20





All Jet Samples Versus Jet20

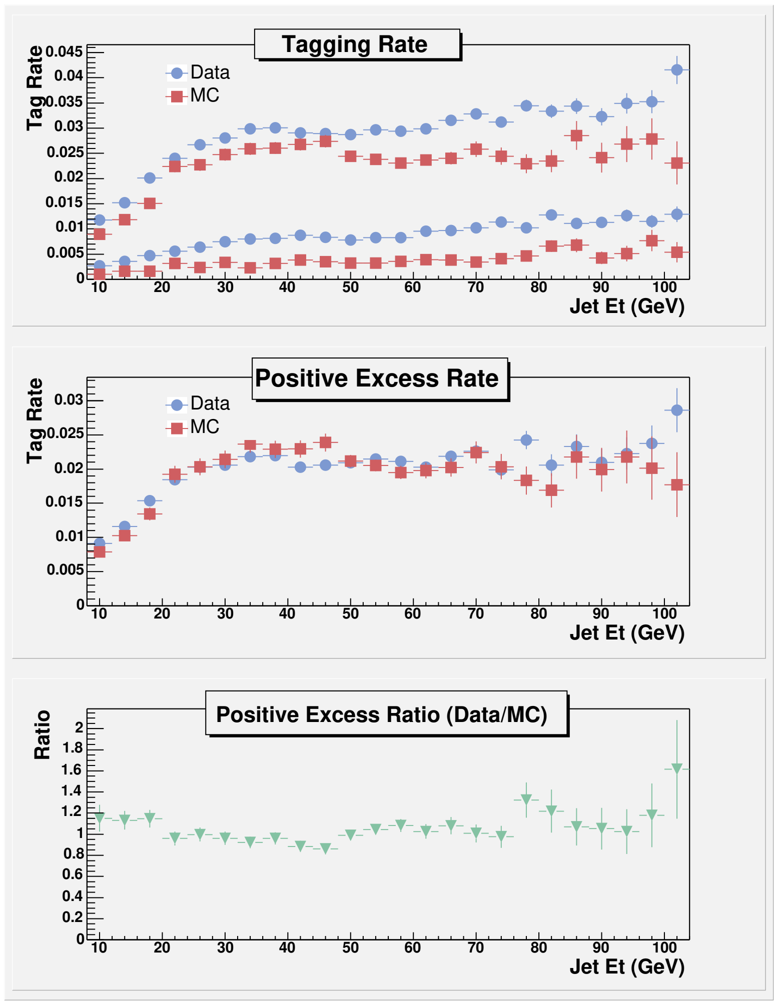




Issue 4: Monte Carlo Comparisons

- Compared Jet50 data with realistic PYTHIA MC (jtop4e, 500k events), run simulated is 151435
- Examine tag rates as a function of Jet E_T
- Positive and negative rates are both higher in data
- Positive minus negative rate is consistent with Monte Caro

$\Rightarrow$ Mistags are larger in data



Cross-Checks Summary

Check	Pos Obs	Pos Pred	**Pos Ratio**
Even-Odd	0.022280 ± 0.000063	0.022320 ± 0.000068	$\mathbf{0.9982 \pm 0.0042}$
JET50-JET20	0.015946 ± 0.000052	0.01950 ± 0.00031	$\mathbf{0.818 \pm 0.013}$
JET50-JET70	0.03039 ± 0.00015	0.03009 ± 0.00020	$\mathbf{1.0098 \pm 0.0084}$
JET50-JET100	0.03455 ± 0.00014	0.03362 ± 0.00058	$\mathbf{1.028 \pm 0.018}$
JET50-Sum Et	0.027824 ± 0.000091	0.02573 ± 0.00022	$\mathbf{1.082 \pm 0.010}$
Even-Odd JET20	0.015946 ± 0.000052	0.015884 ± 0.000070	$\mathbf{1.0039 \pm 0.0055}$
Trig	0.021263 ± 0.000090	0.021783 ± 0.000075	$\mathbf{0.9761 \pm 0.0053}$
Non-Trig	0.023354 ± 0.000091	0.022930 ± 0.000082	$\mathbf{1.0184 \pm 0.0053}$
Sum Et	0.02757 ± 0.00011	0.02619 ± 0.00011	$\mathbf{1.0524 \pm 0.0059}$

Check	Neg Obs	Neg Pred	**Neg Ratio**
Even-Odd	0.006116 ± 0.000034	0.006123 ± 0.000044	$\mathbf{0.9989 \pm 0.0091}$
JET50-JET20	0.003413 ± 0.000024	0.00428 ± 0.00016	$\mathbf{0.798 \pm 0.031}$
JET50-JET70	0.009582 ± 0.000085	0.00930 ± 0.00013	$\mathbf{1.030 \pm 0.017}$
JET50-JET100	0.012190 ± 0.000085	0.01151 ± 0.00044	$\mathbf{1.060 \pm 0.041}$
JET50-Sum Et	0.007301 ± 0.000047	0.00712 ± 0.00015	$\mathbf{1.026 \pm 0.023}$
Even-Odd JET20	0.003413 ± 0.000024	0.003446 ± 0.000036	$\mathbf{0.990 \pm 0.012}$
Trig	0.005654 ± 0.000047	0.005871 ± 0.000046	$\mathbf{0.963 \pm 0.011}$
Non-Trig	0.006588 ± 0.000049	0.006399 ± 0.000057	$\mathbf{1.029 \pm 0.011}$
Sum Et	0.007123 ± 0.000055	0.007263 ± 0.000070	$\mathbf{0.980 \pm 0.012}$

For Blessing : Systematics

- Systematic uncertainties based on differences between `Jet50` and `Jet20`, and differences between $\sum E_T$ sample and full matrix
- Saw difference between `Jet50` and `Jet20` in $\mathrm{Run\ I}$, winter analysis did not examine this. Divide difference by two for systematic uncertainty.
- Averaging the two systematics together to get full systematic uncertainty

Systematic effect	Positive Tags	Negative Tags
Trigger jet bias	3%	4%
Sample bias($\sum E_T$)	6%	6%
Jet threshold bias	9%	9%
Sample Statistics	1%	1%
Total	11%	12%

Summary For Blessing:

- **Tag rate matrix** from full dataset
- **Plot of tag rates** as function of Jet E_T and number of good tracks
- **Systematics** estimated from sample biases

Generating the Statistical Errors

$\Rightarrow$ The statistical errors stored with the mistag matrices are essentially binomial, where, as in the winter, we use the upper confidence bound for bins with low tagged-jet statistics (N_i^{tagged} less than 10).

Calculating the Errors for a Distribution

When calculating the uncertainty on the predicted number of tags in a given bin k of your distribution, you *must account for correlations* when you combine the errors $\sigma(r^i)$, for all jets in bin i of the matrix.

- The errors associated with jets from the *same* bin in the mistag matrix must be treated as entirely correlated, and directly added, while
- The predicted probabilities to tag jets which fall into *different* bins in the mistag matrix are derived from uncorrelated measurements, and should be added in quadrature.

$$\sigma(N_k^{\text{tagged}}) = \sqrt{\sum_i (N_{k,\text{bin}i} \times \sigma(r^i))^2}$$

- This is cumbersome because you must keep track of the number of jets in each matrix bin *in each bin of your histogram* – but for one-dimensional histograms we provide a class (for use in `ROOT` macros) `PredictedDistribution` that makes it easier.