Preliminary 5.3.3 Mistag Matrix

Salvatore Rappoccio

September 10, 2004
High p_T b-Tag Meeting

- Preliminaries
  - Code status
  - Ntuple processing
- Mistag matrix
  - Cross checks
  - Comments/Observations
  - Plans for this round

Ntuples

- Data is processed up to Feb 13 (138425-179104)
- DQM Good Run list v6.0
- **Control samples:**
- `gjt[1-4]0d`: Jet20,50,70,100
- `gset0d`: Jet20,50,70,100

⇑ No strip, identifier is "BTag"

For locations, see

http://www-cdf.fnal.gov/internal/physics/top/gen5/gen5_topntuples.html

Mistag Matrix

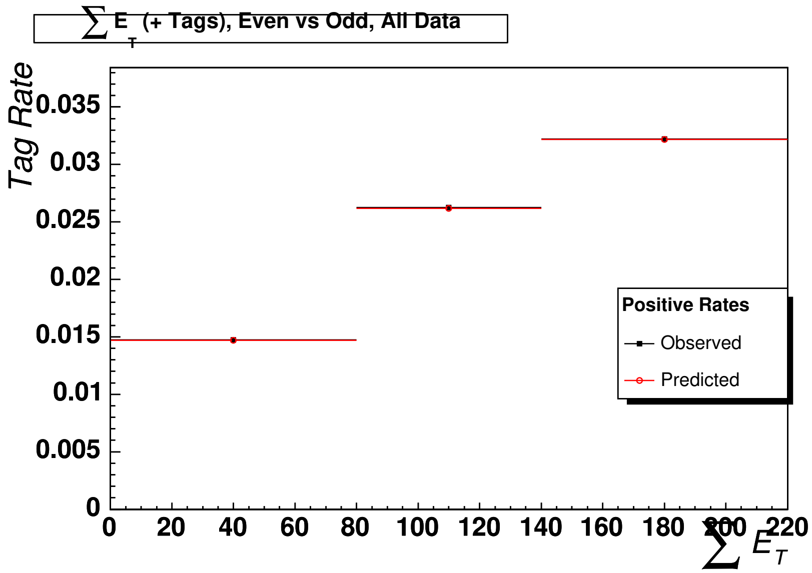
- Using same 5-d matrix as in the past
- Negative matrix is off by 1.3σ (20% probability of occurrance)
- Performing "even matrix on even events" gives exactly 1, so confident that there are no bugs, but a statistical fluctuation

Check	Pos Obs	Pos Pred	**Pos Ratio**
Even vs Odd, All Data	0.023338 ± 0.00005	0.02332 ± 0.00005	$\mathbf{1.001 \pm 0.003}$
$\sum E_T$ Sample	0.02641 ± 0.00008	0.02559 ± 0.00006	$\mathbf{1.032 \pm 0.004}$
Check	Neg Obs	Neg Pred	**Neg Ratio**
Even vs Odd, All Data	0.00555 ± 0.00002	0.00550 ± 0.00003	$\mathbf{1.009 \pm 0.007}$
$\sum E_T$ Sample	0.00564 ± 0.00004	0.00590 ± 0.00003	$\mathbf{0.955 \pm 0.008}$

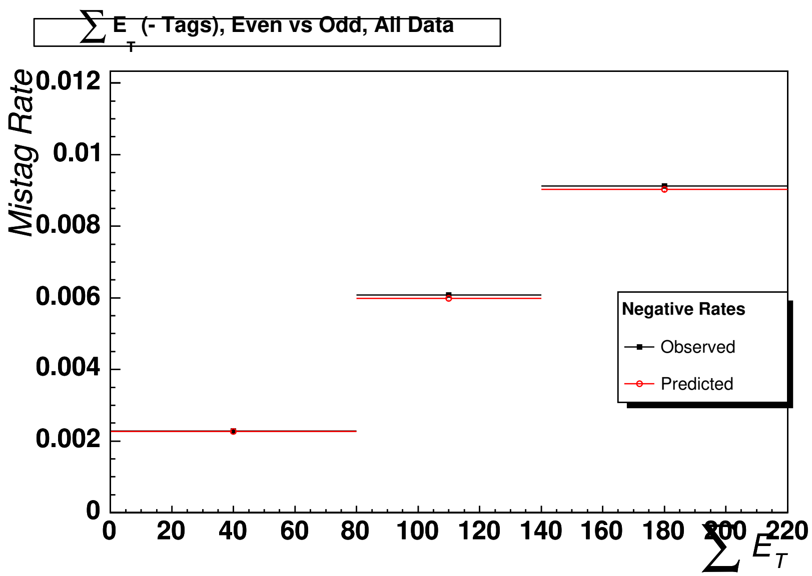
- $\sum E_T$ sample bias $\sim$ 5%, similar to other tagger

Next, examine matrix variables

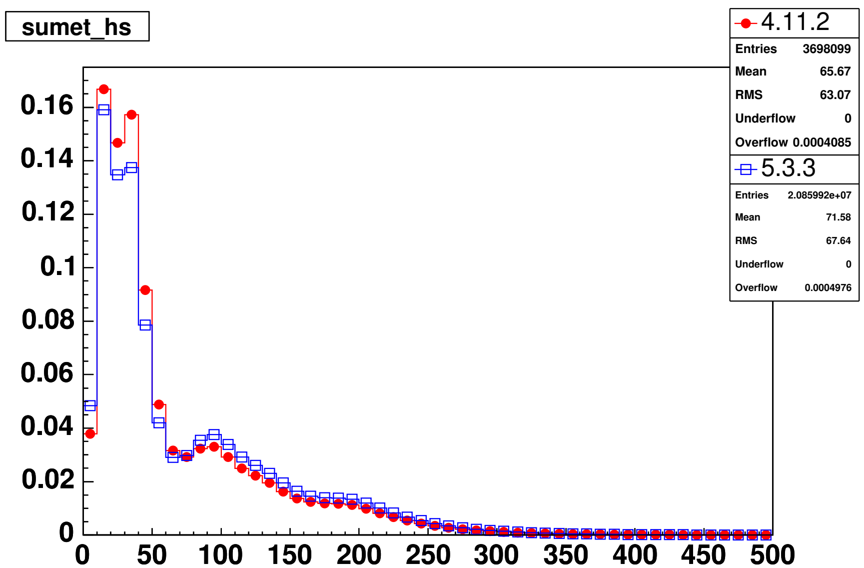
$\sum E_T$



$\sum E_T$ (+ Tags), Even vs Odd, All Data
Tag Rate
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0 20 40 60 80 100 120 140 160 180 200 220
$\sum E_T$
Positive Rates
Observed
Predicted

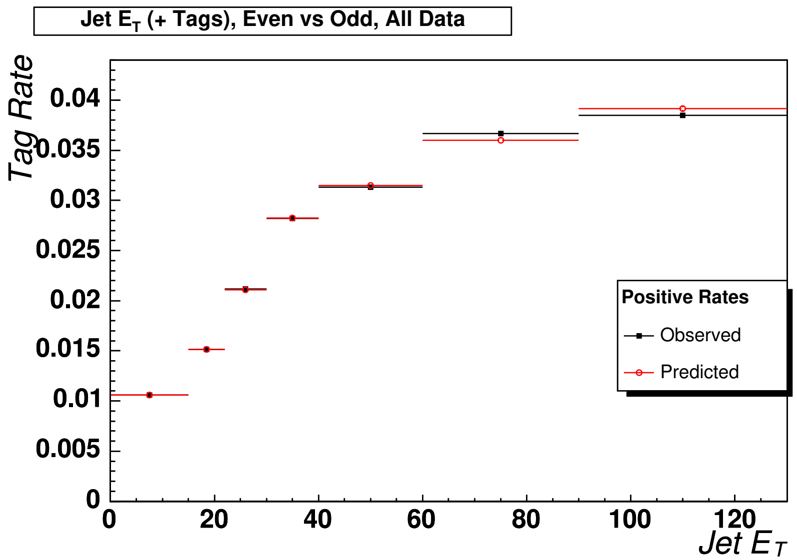


$\sum E_T$ (- Tags), Even vs Odd, All Data
Mistag Rate
0.012
0.01
0.008
0.006
0.004
0.002
0
0 20 40 60 80 100 120 140 160 180 200 220
$\sum E_T$
Negative Rates
Observed
Predicted

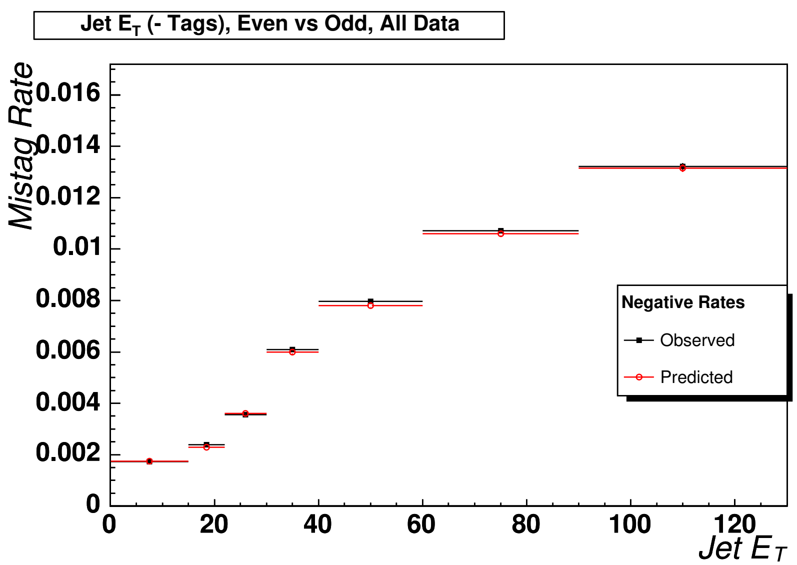


sumet_hs
4.11.2
Entries 3698099
Mean 65.67
RMS 63.07
Underflow 0
Overflow 0.0004085
5.3.3
Entries 2.085992e+07
Mean 71.58
RMS 67.64
Underflow 0
Overflow 0.0004976
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0 50 100 150 200 250 300 350 400 450 500

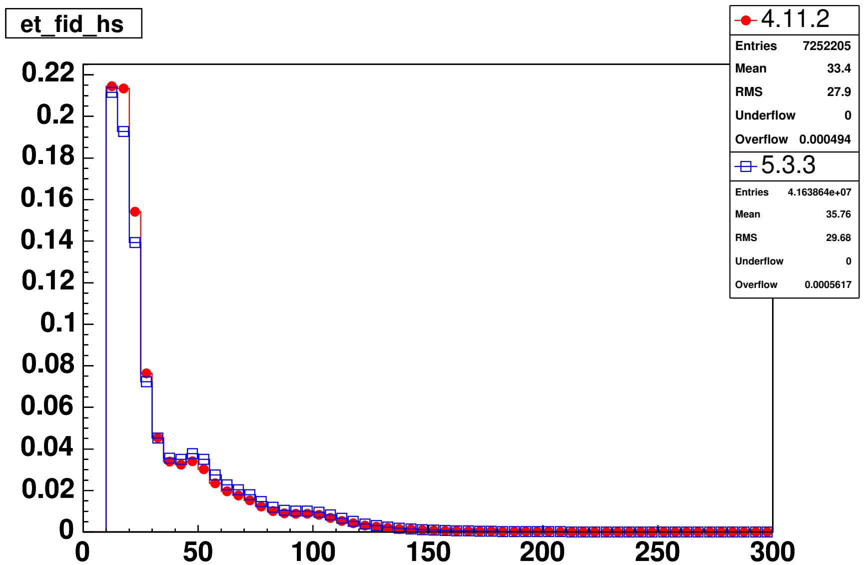
E_T



Jet E_T (+ Tags), Even vs Odd, All Data
Tag Rate
Jet E_T
Positive Rates
Observed
Predicted

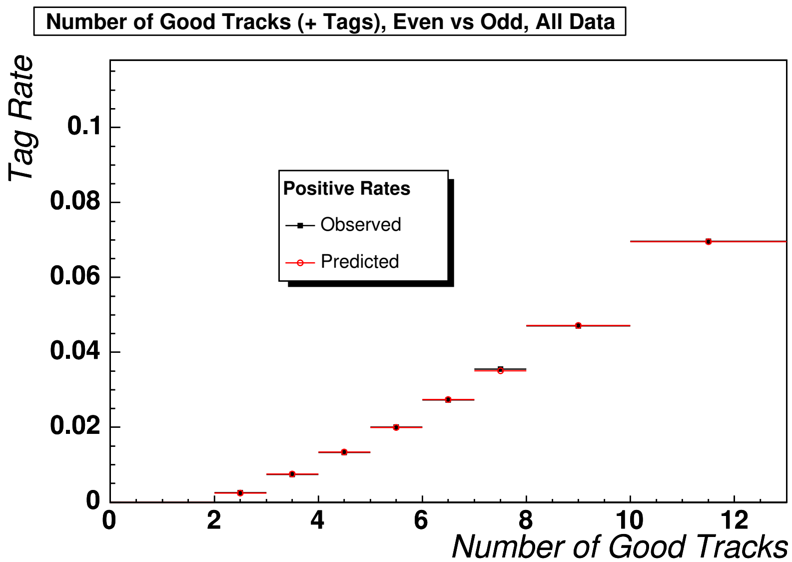


Jet E_T (- Tags), Even vs Odd, All Data
Mistag Rate
Jet E_T
Negative Rates
Observed
Predicted

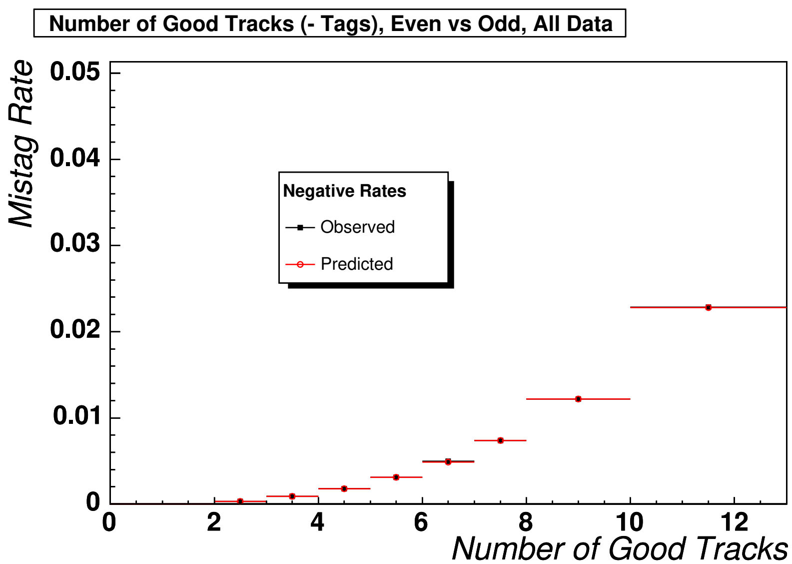


et_fid_hs
4.11.2
Entries 7252205
Mean 33.4
RMS 27.9
Underflow 0
Overflow 0.000494
5.3.3
Entries 4.163864e+07
Mean 35.76
RMS 29.68
Underflow 0
Overflow 0.0005617

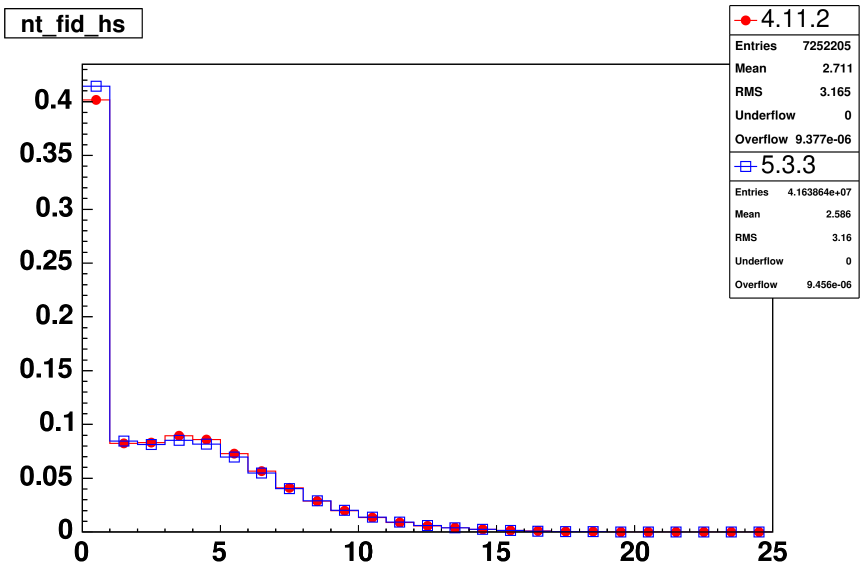
N_{trk}



Number of Good Tracks (+ Tags), Even vs Odd, All Data
Tag Rate
Positive Rates
Observed
Predicted
Number of Good Tracks

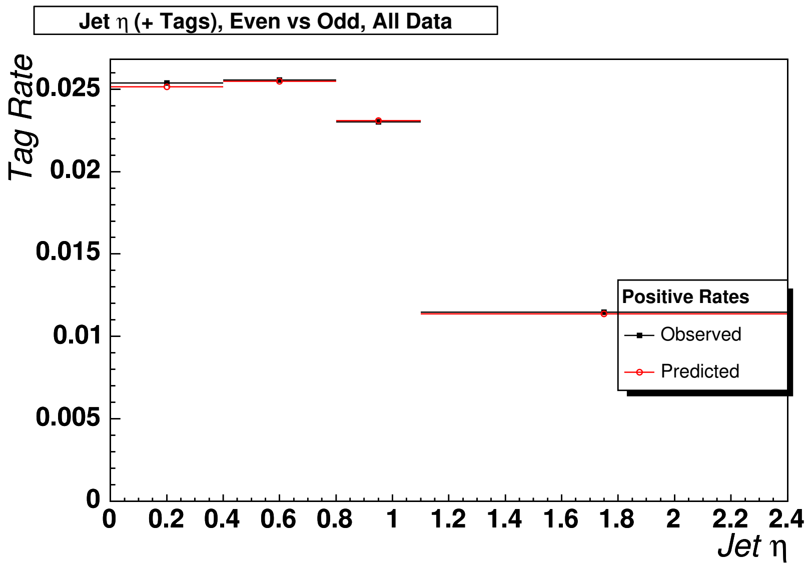


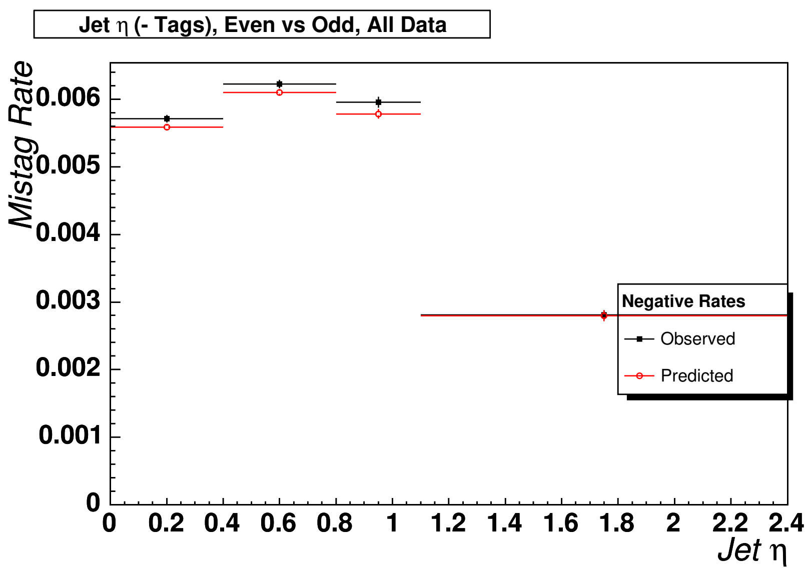
Number of Good Tracks (- Tags), Even vs Odd, All Data
Mistag Rate
Negative Rates
Observed
Predicted
Number of Good Tracks

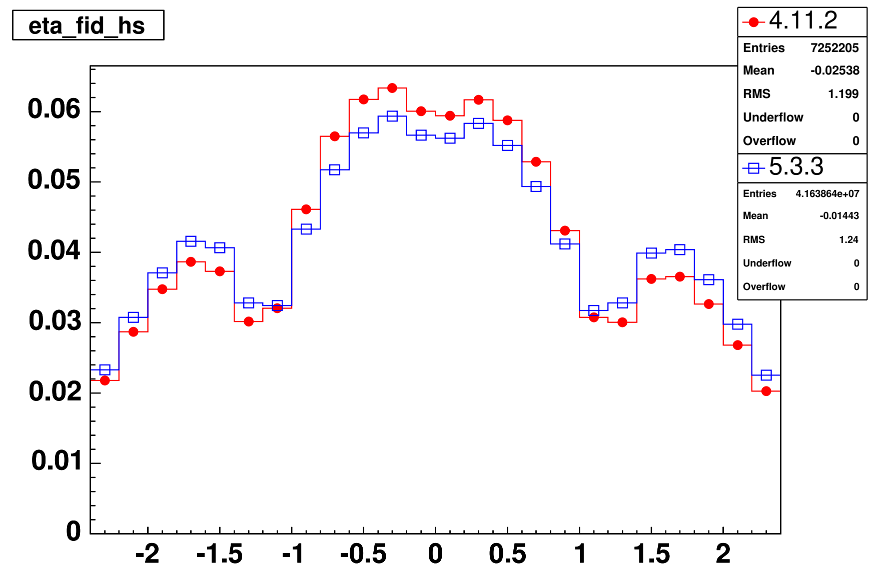


nt_fid_hs
4.11.2
Entries 7252205
Mean 2.711
RMS 3.165
Underflow 0
Overflow 9.377e-06
5.3.3
Entries 4.163864e+07
Mean 2.586
RMS 3.16
Underflow 0
Overflow 9.456e-06

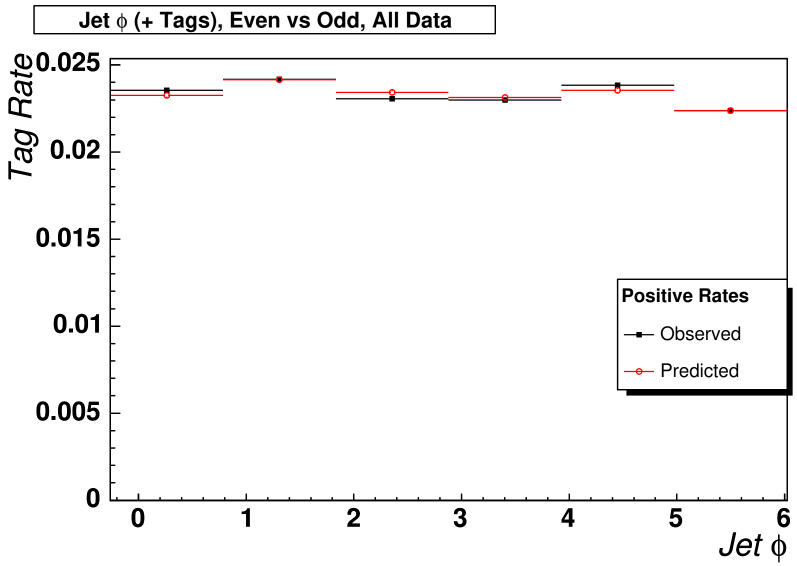
η

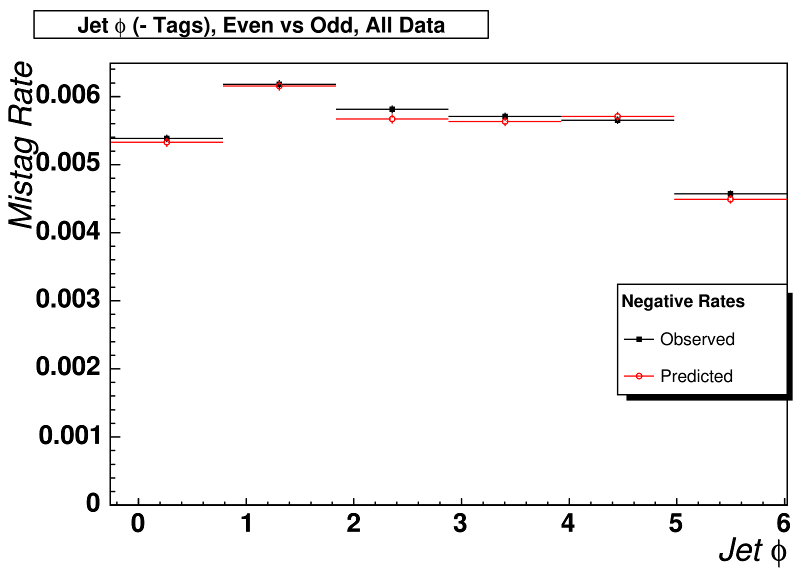

Jet η (+ Tags), Even vs Odd, All Data
Tag Rate
Jet η
Positive Rates
Observed
Predicted

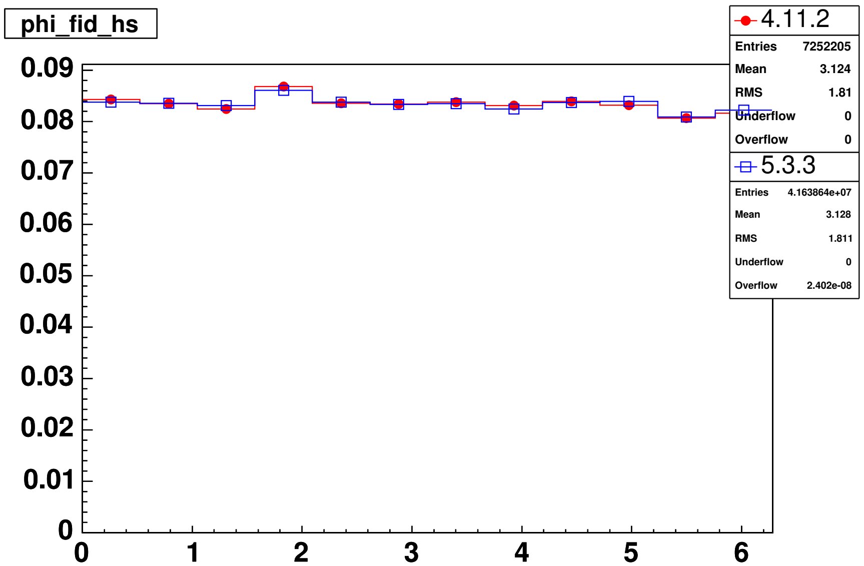

Jet η (- Tags), Even vs Odd, All Data
Mistag Rate
Jet η
Negative Rates
Observed
Predicted


eta_fid_hs
4.11.2
Entries 7252205
Mean -0.02538
RMS 1.199
Underflow 0
Overflow 0
5.3.3
Entries 4.163864e+07
Mean -0.01443
RMS 1.24
Underflow 0
Overflow 0

ϕ

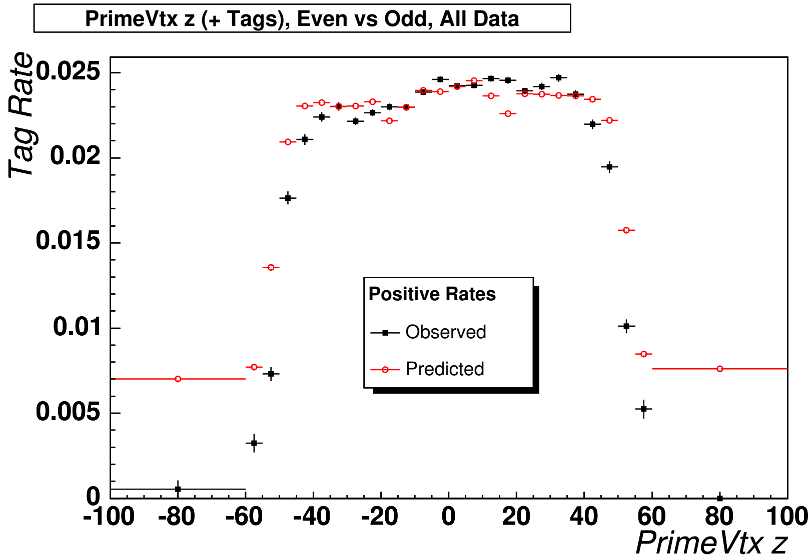


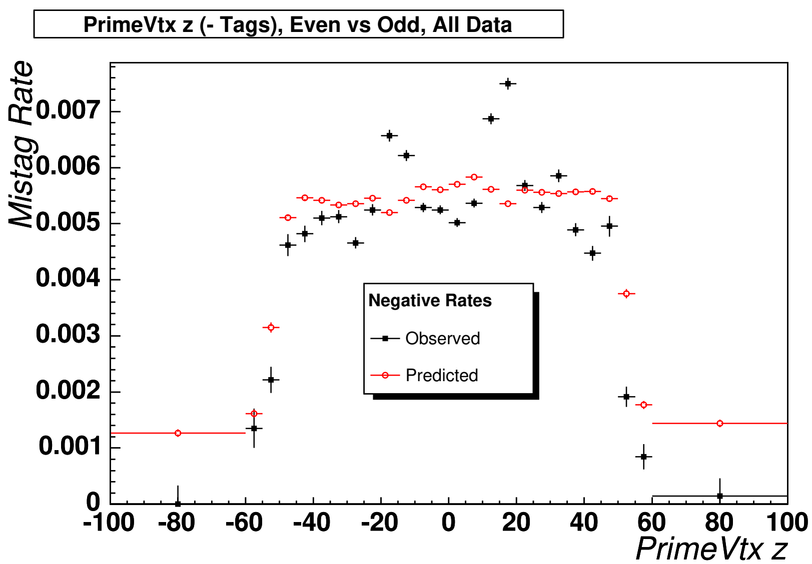


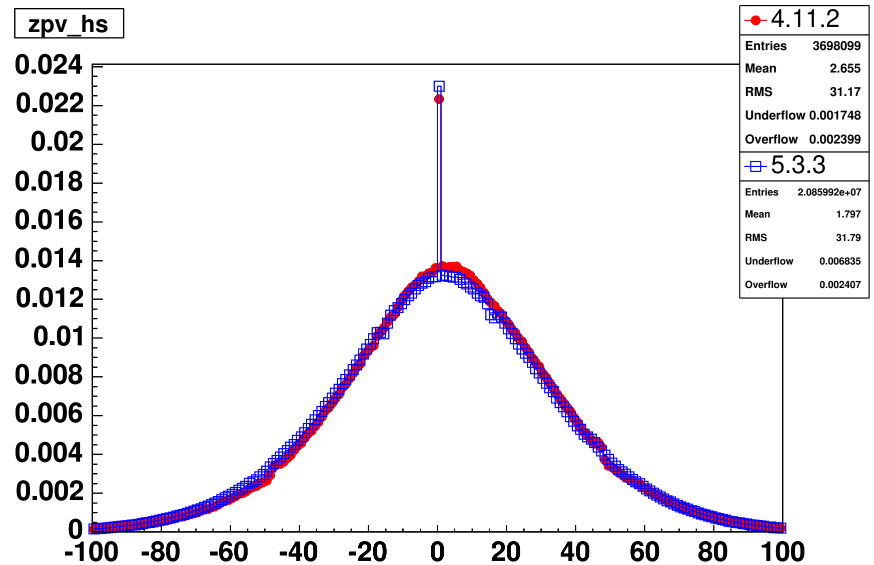


- Cross checks came out okay
- *Next, examine predicted variables*

Z_{PV}

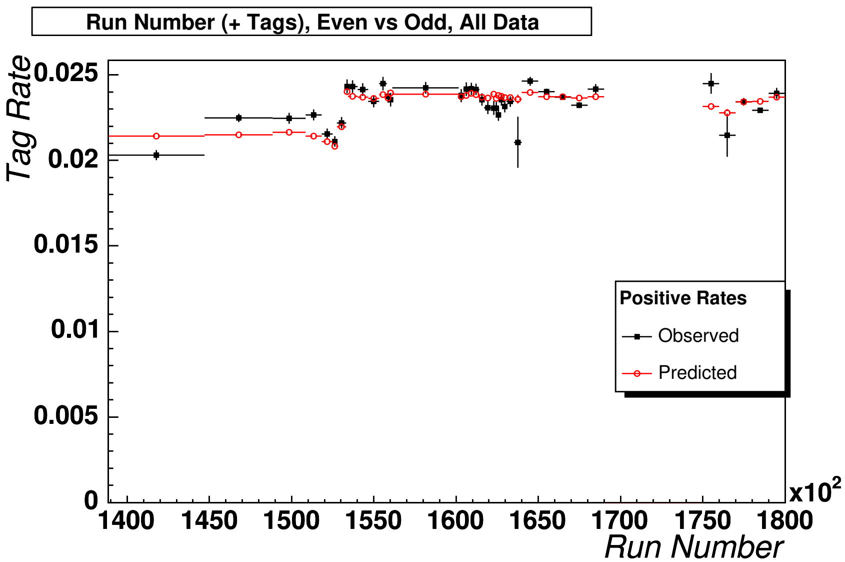


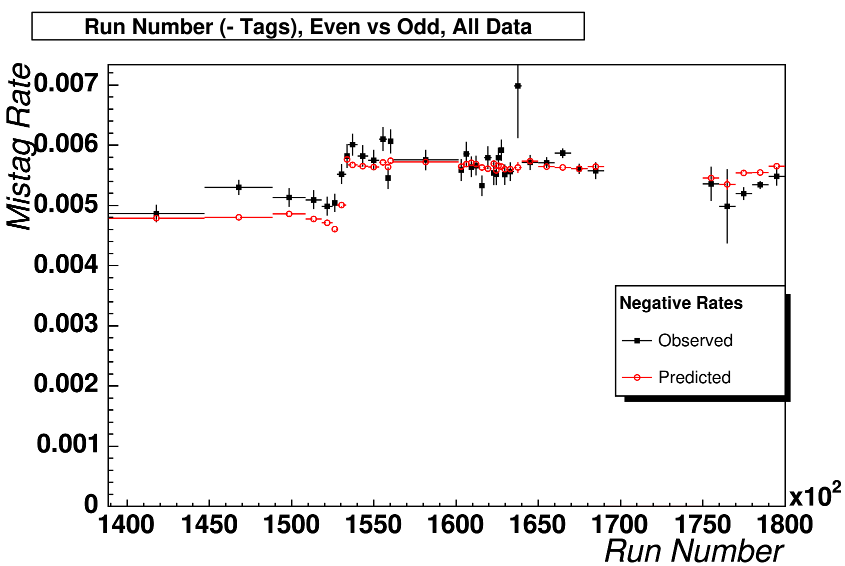


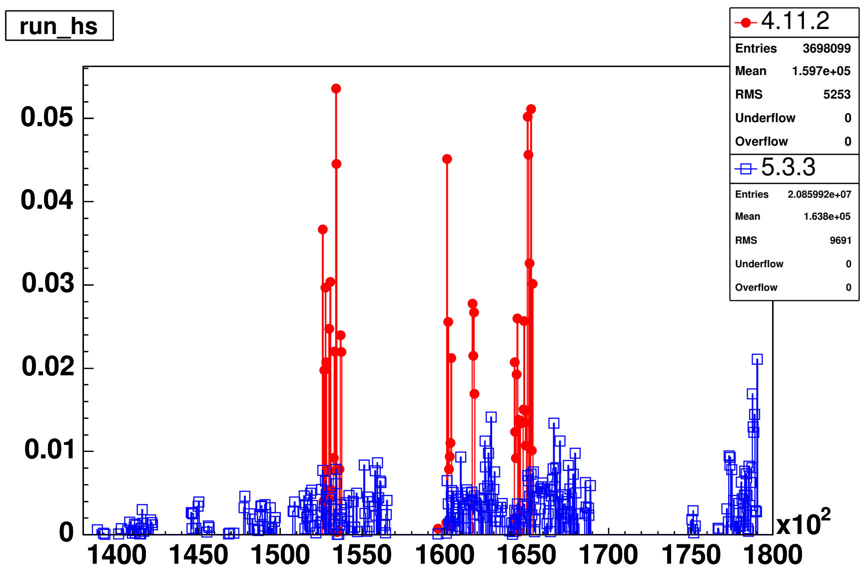


- Positive tags described very well
- Still seeing discrepancy in negative tags in bulkhead regions (even though we are using z-dependent pulls for the primary vertex)

Run Number

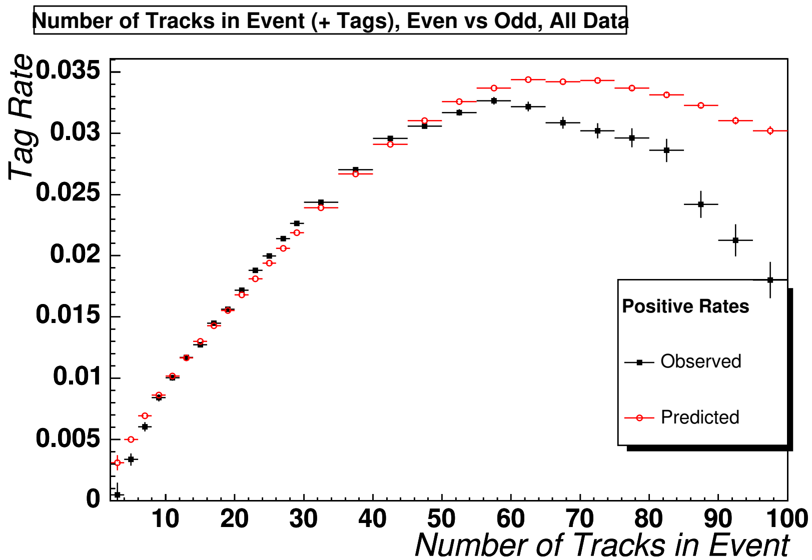


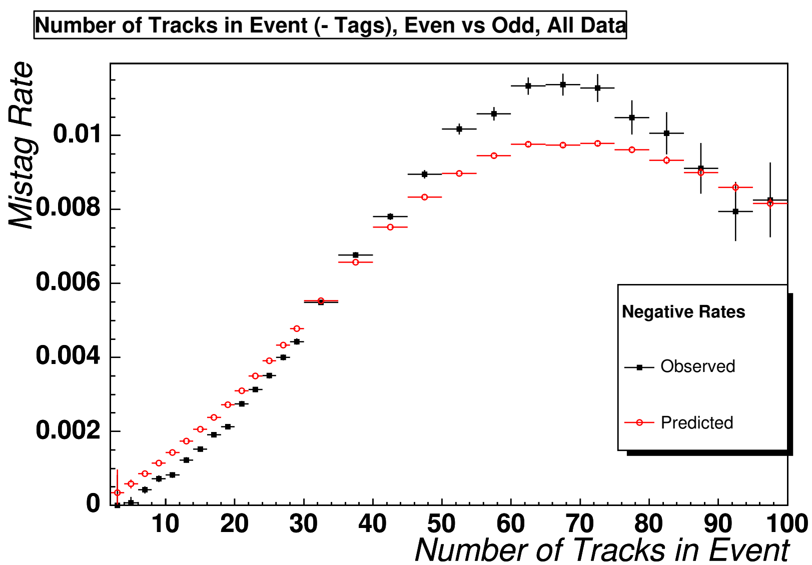


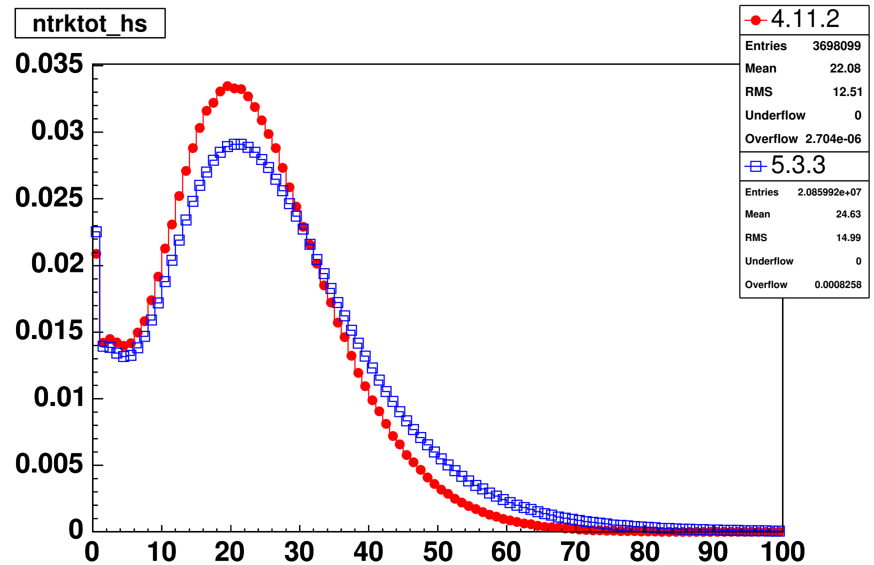


- Predicting rates versus run number adequately, although the actual rates jump around a bit

$N_{Total-Trks}$

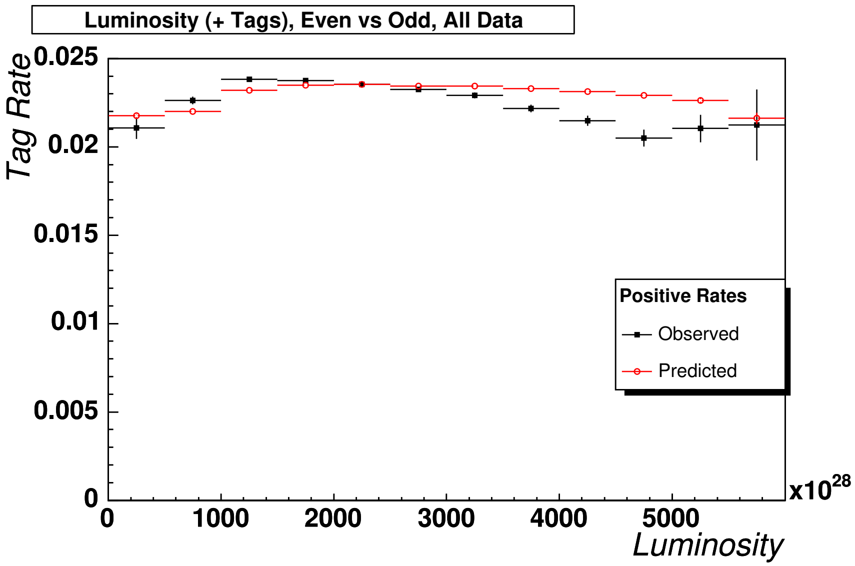


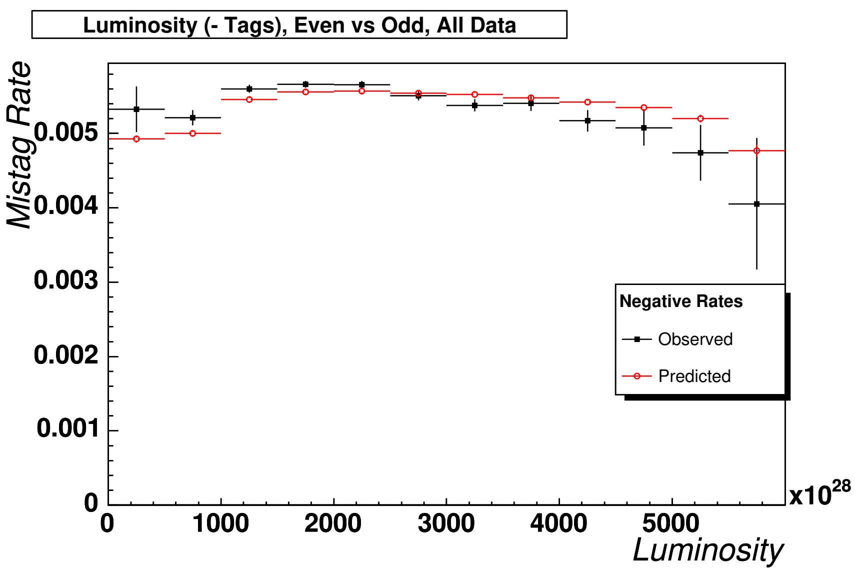


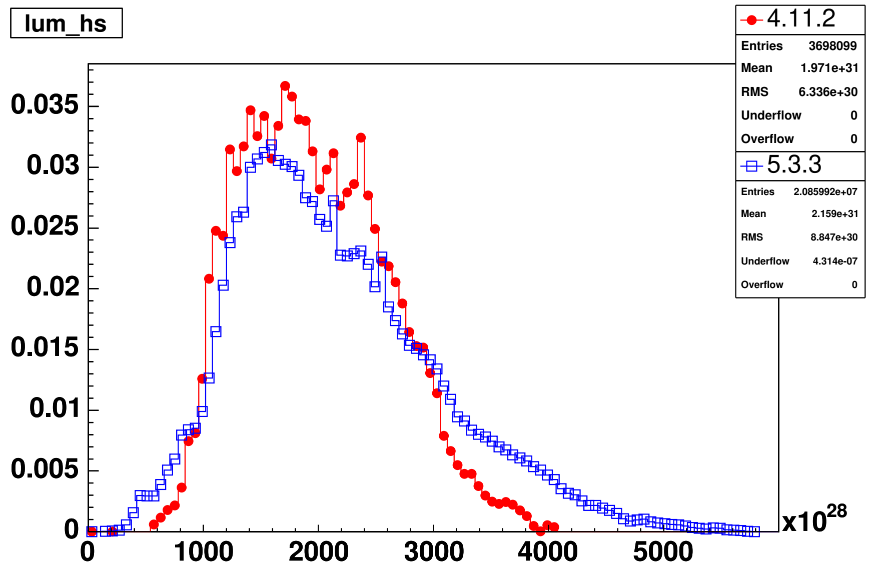


- Prediction is good for low numbers of tracks
- Discrepancy occurs at large numbers of tracks
- Due to increased luminosity, populate discrepant region more
- **Related to the increased instantaneous luminosity**

Luminosity

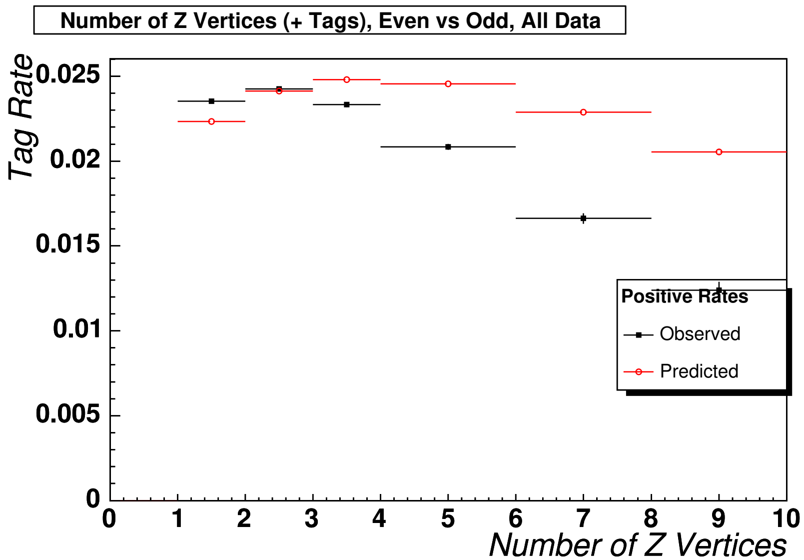


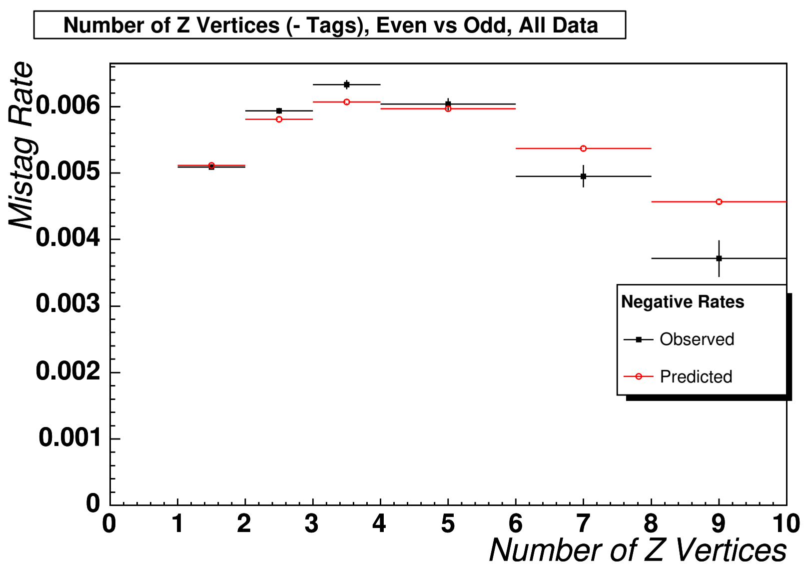


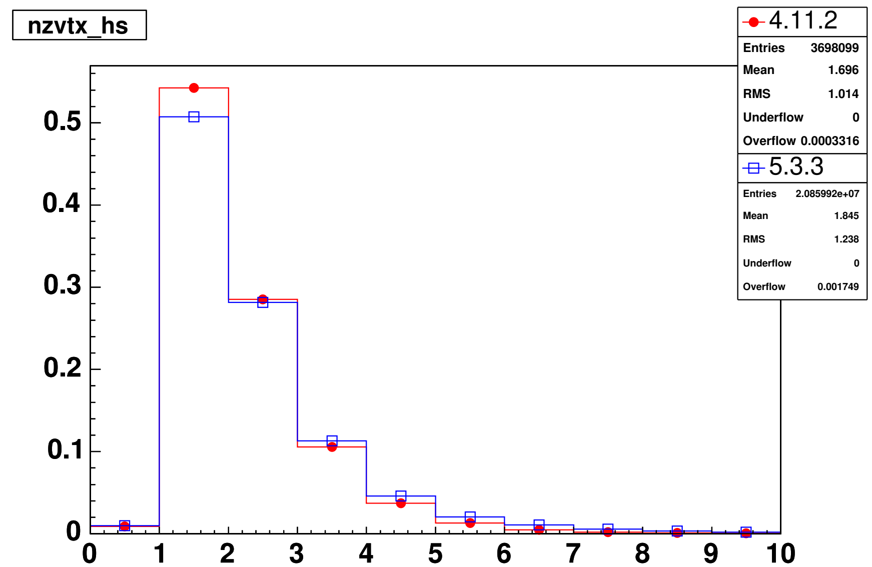


- For low luminosity, prediction and observation agree
- For higher luminosity, prediction and observation don't agree so well

N_{z-vtx}







- Another way of looking at the same problem
- Discrepant regions are now more populated

Matrix Cross Check Summary

- Good predictivity across many variables
- Discrepant variables relate to **busy environment!**
- **Need to evaluate effects of multiple interactions on tag rates**
- Run I: took care of this by examining $\sum p_T$ of tracks in sub-leading z-vertices
- Examine this further

Heavy Flavor Contamination

- Almost all sources of systematic error come from *heavy flavor bias* in the matrix
  - Create matrix with one heavy flavor fraction
  - Apply to data with anotehr heavy flavor fraction
  - Use $\sum E_T$ of the jets in the event to "account" for this (although that is a strong word)
- **GOAL:** make an actual "mistag matrix" that has no heavy flavor contamination
  - Remove $\sum E_T$ from the matrix
  - Remove η and ϕ to make "multiplier matrices" that take into account detector effects, plus different underlying η distributions

Summary And Plans

- Summary
  - Matrix looks okay in the current incarnation
  - Cross checks are there
  - Need to put machinery in place for tight and loose taggers in same ntuple
- Plans
  - Next big step is to remove heavy flavor from matrix
  - Majority of work will be devoted to this