Increase b-tagging acceptance for WH search

Barbara Alvarez, Florencia Canelli, Ricardo Eusebi, Chiara Ferrazza, Bruno Casal Larana, Javier Cuevas Maestro, Peter Dong, Craig Group, Enrique Palencia, Alberto Ruiz, Bernd Stelzer, Rocio Vilar, Rainer Wallny

September 1, 2009
CDF–Italy Meeting

- Introduction
- Loose Flavor Separator
- Input Variables
- output: signal & background
- Conclusions

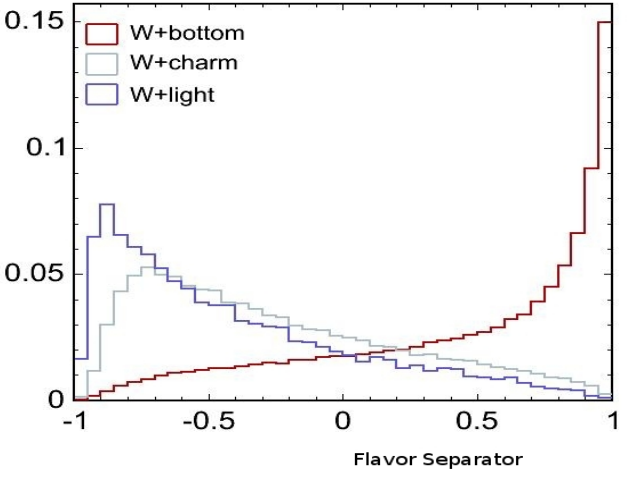
WH Matrix Element Analysis

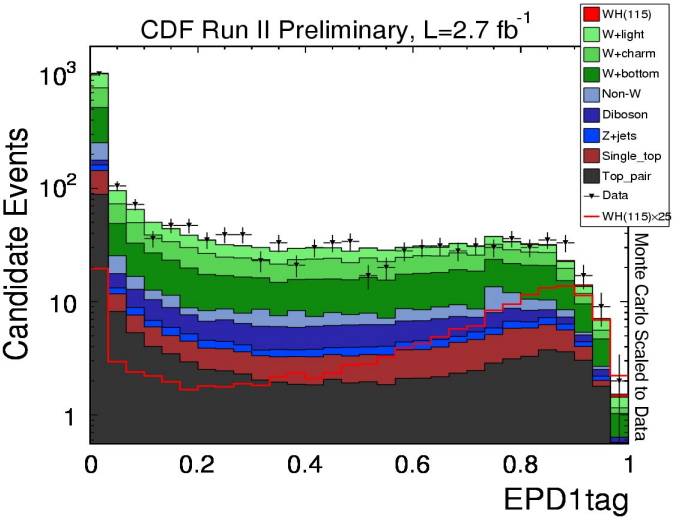
- ME is a technique to separate signal from background exploiting kinematic information
- we apply ME to $WH \rightarrow lvbb$ channel, where one or both b-jets are tagged w/ tight SecVtx
- we define a discriminant (Event Probability Discriminant, EPD):

$$EPD_{1tag} = \frac{b \cdot P_{WH}}{b(P_{WH}+P_{singletop}+P_{Wbb}+P_{tt})+(1-b)(P_{Wcc}+P_{Wcj}+P_{mistags}+P_{diboson})}$$

$$EPD_{2tag} = \frac{b_0 b_1 \cdot P_{WH}}{b_0 b_1 (P_{WH}+P_{schan}+P_{Wbb}+P_{tt})+b_0 \cdot (1-b_1) \cdot P_{tchan}+(1-b_0) \cdot (1-b_1) \cdot (P_{Wcc}+P_{Wcj}+P_{mistags}+P_{diboson})}$$

where P_{proc} is the event prob, and b is a function of KIT





- we want to extend to loose SecVtx $\rightarrow$ a loose KIT is needed

Background Estimation (Method2)

in the following: ST = tight SecVtx, SL = loose SecVtx, JP = JetProb

- there are two different scenarios: use events w/ a LS or with a JP tagged jet
- this table not updated: soon results for 4.3 fb^{-1}

Loose Secondary Vertex (2.7 fb^{-1})			
Process	1 tag (ST)	2 tags (ST-SL)	2 tags (ST-ST)
Total MC	364.2 ± 31.6	28.8 ± 4.0	61.2 ± 7.7
Total HF	993.7 ± 301.8	40.2 ± 13.0	77.9 ± 24.9
Mistags	448.3 ± 55.9	8.9 ± 2.2	2.1 ± 0.6
non-W	130.5 ± 52.2	6.0 ± 2.4	8.9 ± 3.6
Total Prediction	1936.7 ± 313.0	84.0 ± 14.0	150.2 ± 26.3
WH (115 GeV/c^2)	5.5 ± 0.4	0.8 ± 0.1	2.0 ± 0.2
Observed	1919	95	158

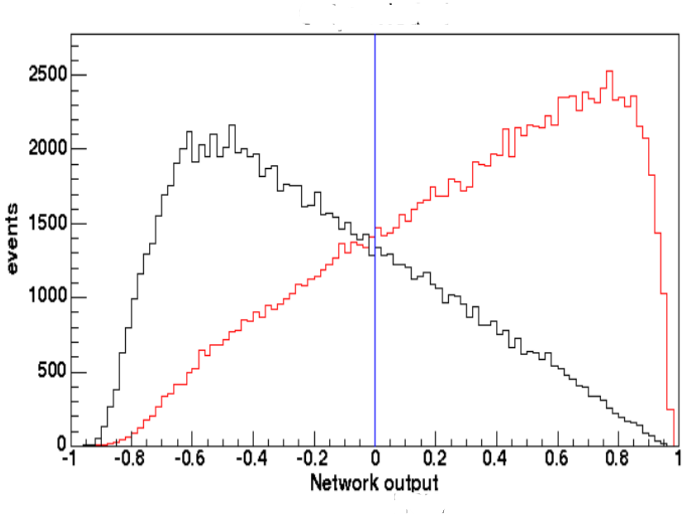
JP<5% (2.7 fb^{-1})			
Process	1 tag (ST)	2 tags (ST-JP)	2 tags (ST-ST)
Total MC	342.6 ± 29.0	51.6 ± 8.9	61.2 ± 7.7
Total HF	944.6 ± 286.8	87.6 ± 29.3	77.8 ± 24.9
Mistags	446.3 ± 54.8	10.7 ± 3.6	2.1 ± 0.6
non-W	129.0 ± 51.6	14.6 ± 5.8	8.9 ± 3.4
Total Prediction	1862.6 ± 297.9	164.5 ± 31.3	150.2 ± 26.3
WH (115 GeV/c^2)	4.9 ± 0.4	1.4 ± 0.2	2.0 ± 0.2
Observed	1855	161	158

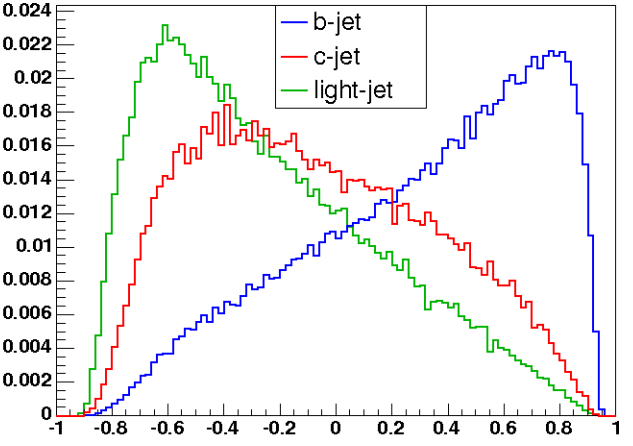
where Total MC includes WW, WZ, ZZ, Top-LJ, Top-Dil, single Top-T, single Top-S, Z+jets processes, and Total HF Wbb and Wcc/cj

- the ST-SL has similar S/B w.r.t. ST-JP but we will apply the flavor separator (for ST this led $\sim 15-20\%$ gain in sensitivity single-top analysis)

NN training

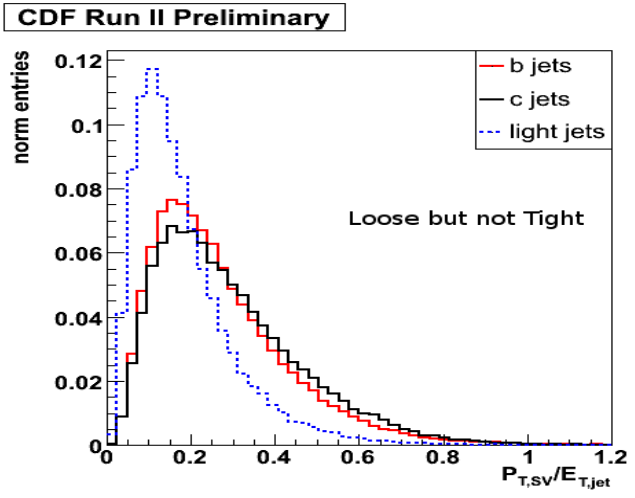
- we want to add the loose but not tight SecVtx events to the analysis and using a flavor separator would increase the sensitivity
- training sample: b-jets for signal (red histo) and c+mistags for background (black histo)
- we trained the NN over several variables related to the reconstructed secondary vertex, tracks associated and leptons inside the jet

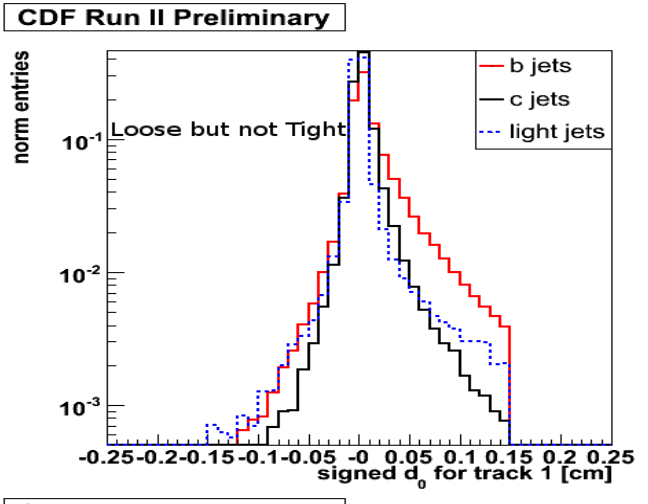


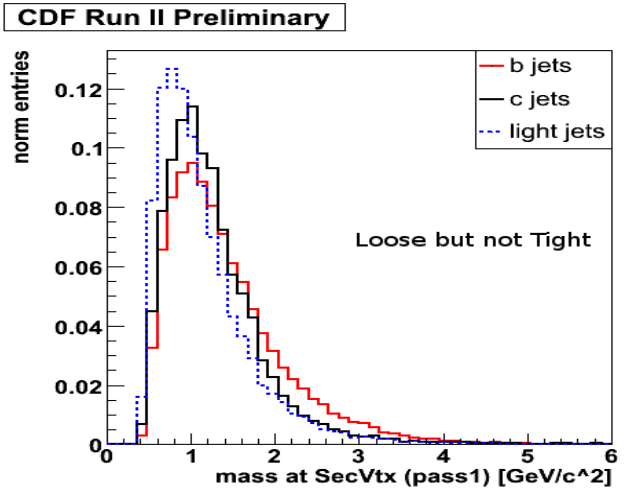


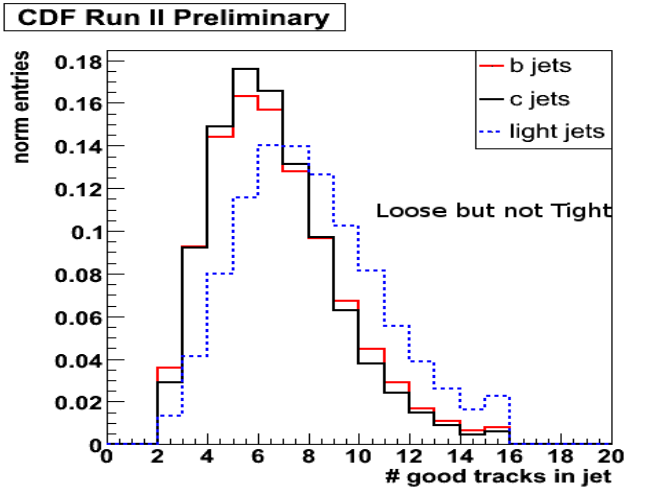
- good separation between b and light jets

Examples of input variables



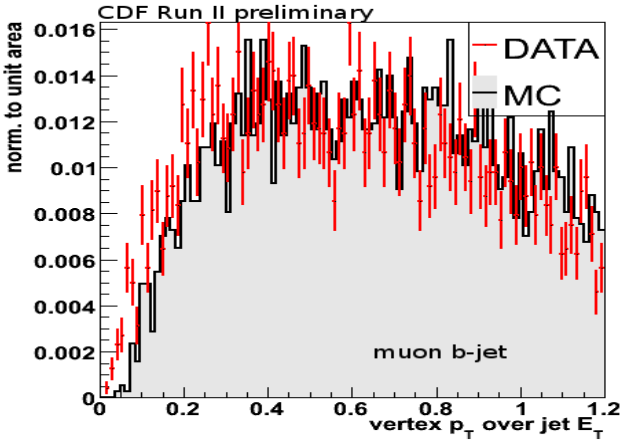


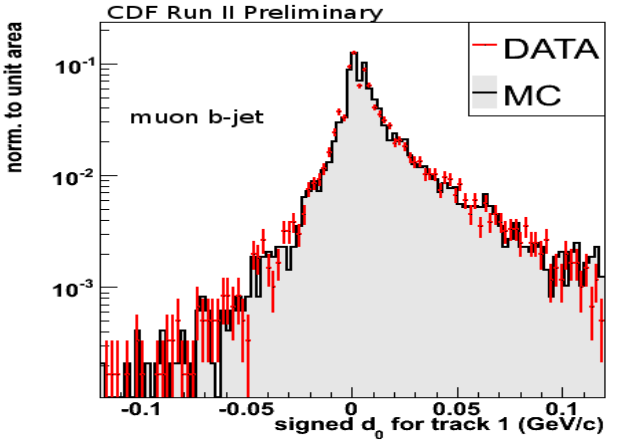


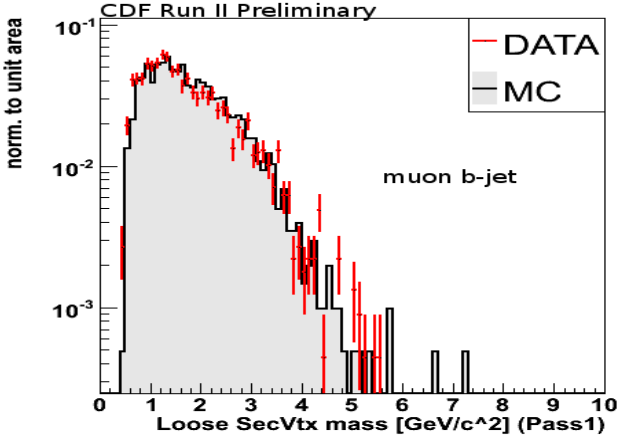


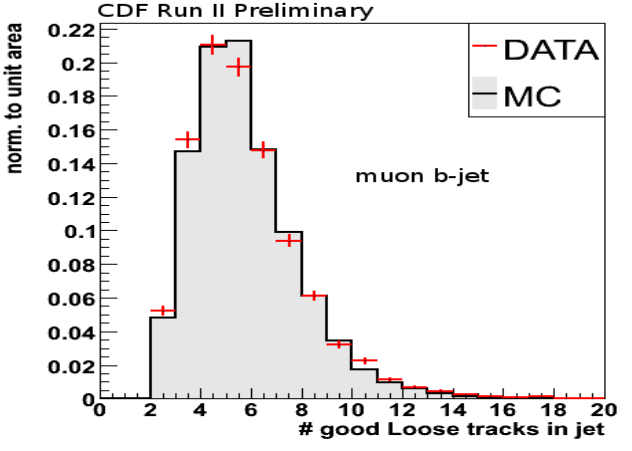
Examples of validation plots (muons)

jet with semi-leptonic b decay which enhances b-fraction in data



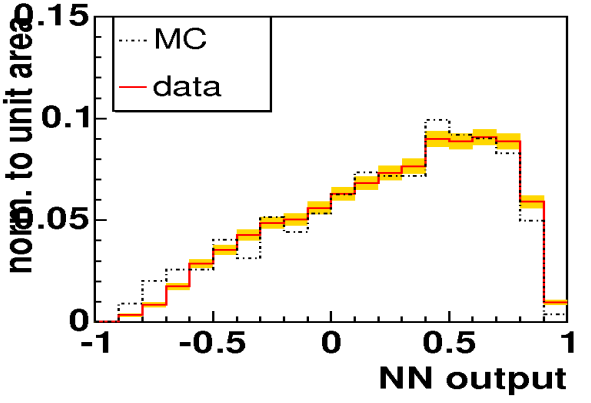


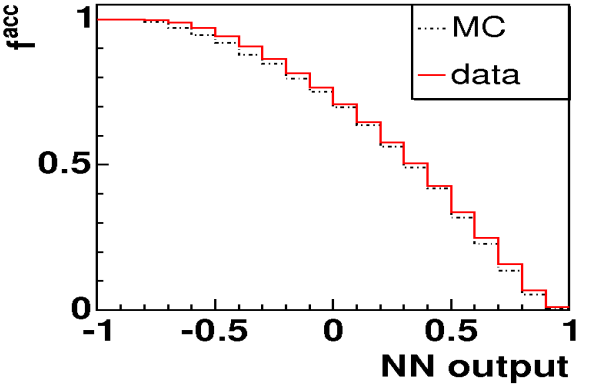


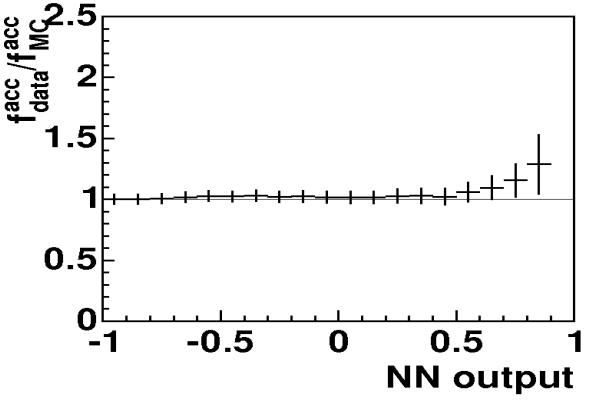


Comparison of signal output

- same selection as for SevVtx scale factor
- 2 jets: lepton jet ST, away jet SL but not ST (b content $\sim 73\%$)



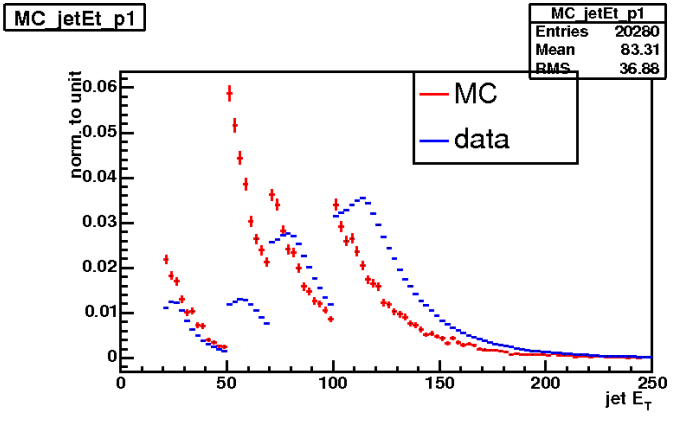




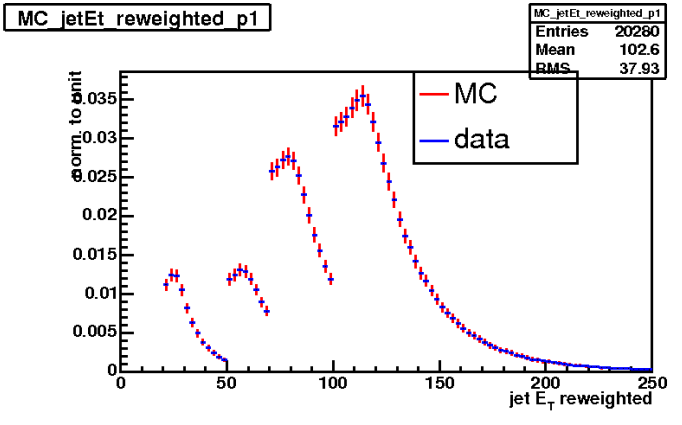
- within the given statistics, looks fine: no correction needed

Mistags

- we use negatively tagged jets to represent light positively tagged jets
- MC samples: btopqb, btoprb, btopsb, btoptb
- Data samples: gjt1XX, gjt2XX, gjet3XX, gje4XX (XX=kd,kh,ki,mi,mj,mk,mm)
- we apply following cuts:
  - btoprb/gjt2XX $\rightarrow E_T \geq 50$ GeV
  - btopsb/gjt3XX $\rightarrow E_T \geq 70$ GeV
  - btoptb/gjt4XX $\rightarrow E_T \geq 100$ GeV
- MC jet E_T distributions reweighted to data



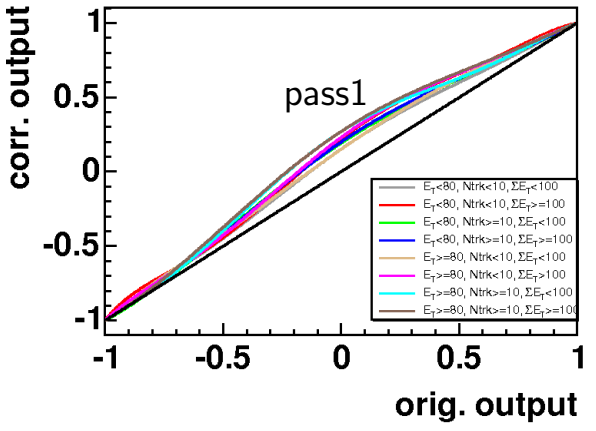
before

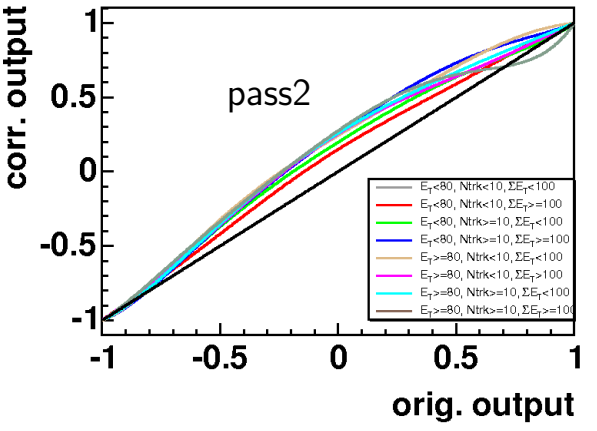


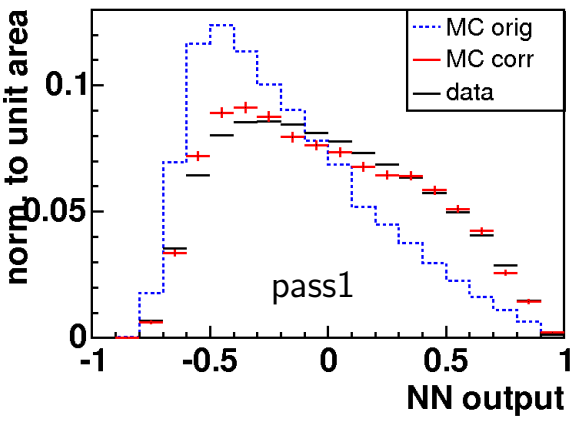
after

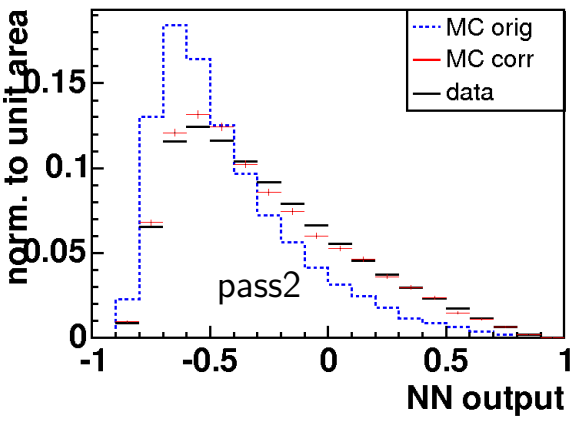
Correction Functions & Corrected Outputs

- small deviations between data & MC in the mistag sample
- better agreement after applying correction functions

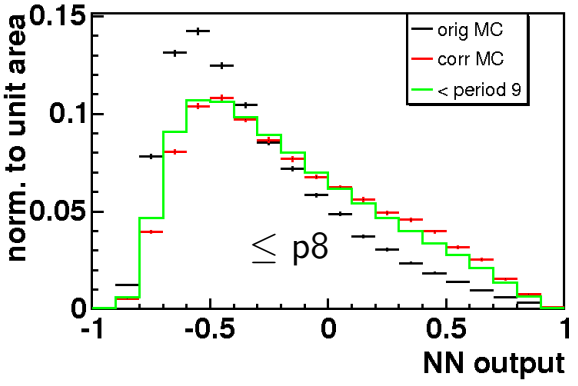


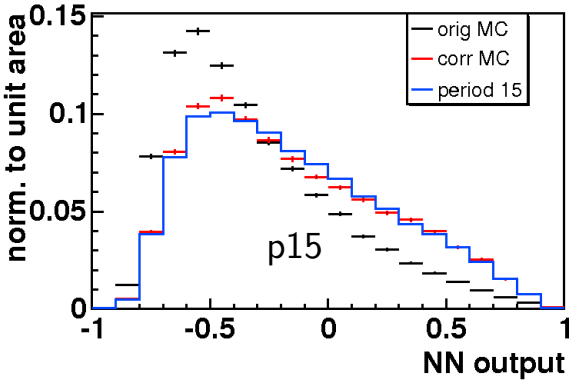


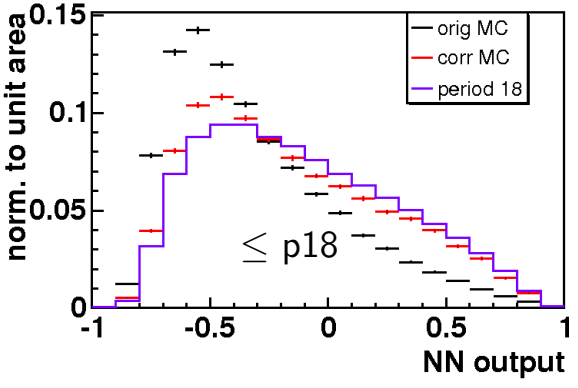


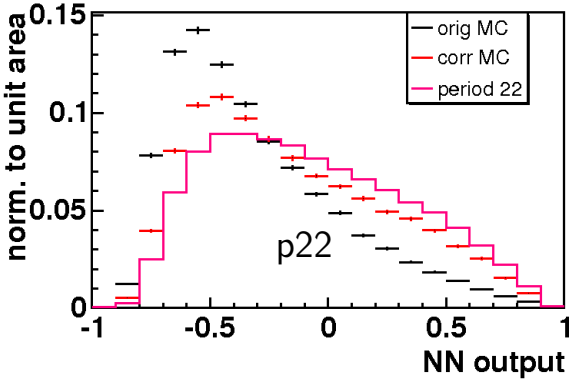


Mistags vs. data period









- good agreement for the first 2.7 fb^{-1}
- more deviations for latest data periods $\rightarrow$ corrrection functions are calculated w/ a low lum MC

Conclusions & Plans

- conclusions:
  - we want to increase b-tagging acceptance my using loose SecVtx tagged jets
  - we develop a loose KIT for loose but not tight jets, in the same way as for the tight
  - we validate input variables and the NN output in a b-enriched sample
  - mistag validation needs some correction functions to be applied
  - residual discrepances in mistags are related to high luminosity data
- plans:
  - calculate Method2 tables for 4.3 fb^{-1}
  - hopefully obtain limits for ST-SL events
  - conclude my thesis for October