

# Jet-Parton Matching Using MVA Techniques

Work Report  
6 April 2010

Maryam Zeinali

# Outline:

- Data sample & Software
- Ranking variables
  - Defining separation power
- Choice of variables for training
  - With & without event selection
- Comparing the two list of variables
- Results
  - Performance plots, mva output and top mass distribution
- Summary & to-do list
- Back-up
  - List of all variables used for monitoring
  - Distribution of input variables used for training
  - Table of linear correlation coefficients of the trained variables

# Data sample & Software:

## ■ Data sample:

- */TTbar/Summer09-MC\_31X\_V3-v1/GEN-SIM-RECO*, skimmed at gen level in order to select only semi electronic ttbar events.

## ■ Software:

- *CMSSW\_3\_1\_4* used with the *V06-02-07* tag.
- The study presented here has been done using MVA implementation within the CMSSW framework, no use of TMVA package based on ROOT!

# Ranking variables:

- Defining separation power:
  - Quoted from TMVA user guide

The **separation**  $\langle S^2 \rangle$  of a classifier  $y$ , defined by the integral [7]

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy, \quad (1)$$

where  $\hat{y}_S$  and  $\hat{y}_B$  are the signal and background PDFs of  $y$ , respectively (cf. Sec. 3.1.13). The separation is zero for identical signal and background shapes, and it is one for shapes with no overlap.

# Choice of variables for training:

## ■ Without event selection cuts

- Only events should have at least one electron and 4 jets to be accepted for training (no cuts on the objects)
- 10 best ranked variables w.r.t. separation power(\*):

<u>Variable Name</u>	<u>Separation Power</u>
1) BtagSumHadQHadQBarTrkCntHighEff	1.96004
2) BtagSumTrkCntHighEff	0.915039
3) DeltaRHadTopLepB	0.00515643
4) DeltaRLepBLEpton	0.00307777
5) DeltaRHadBLEpB	0.00252136
6) DeltaThetaHadWLepB	0.00237632
7) DeltaThetaHadTopHadW	0.000767957
8) DeltaRHadTopLepton	0.000677457
9) DeltaThetaHadTopHadB	0.000380204
10) DeltaRHadBLEpton	0.000350936

(\*): Selected vars. have less than 40% mutual correlations among each other and less than 20% with top & W masses.

# Choice of variables for training:

## ■ With event selection cuts

### ■ Events are asked to have

- 4 LEADING jets with  $et > 25$  GeV and  $|\eta| < 2.4$
- 1 LEADING electron with  $pt > 20$  GeV,  $|\eta| < 2.4$  (excluding the gap), `eidRobusTight`, `rellso < 0.1` and `d0 < 200` MicroMeter.

### ■ 10 best ranked variables w.r.t. separation power(\*):

#### Variable Name

#### Separation Power

1) BtagSumTrkCntHighPur	0.727449
2) BtagProdHadQHadQBarImpactParaMVA	0.0442964
3) BtagHadBImpactParaMVA	0.038125
4) DeltaRHadTopLepB	0.00593808
5) DeltaThetaLepBLEpton	0.00420551
6) DeltaRHadBLEpB	0.0027576
7) DeltaRHadWLepton	0.00141352
8) DeltaThetaHadTopHadW	0.00115513
9) DeltaRHadBLEpton	0.000620288
10) DeltaThetaHadTopHadB	0.000469168

(\*): Selected vars. have less than 40% mutual correlations among each other and less than 20% with top & W masses.

# Comparing the two list of variables:

without selection cuts

with selection cuts

1) BtagSumHadQHadQBarTrkCntHighEff

1) BtagSumTrkCntHighPur

2) BtagSumTrkCntHighEff

2) BtagProdHadQHadQBarImpactParaMVA

3) DeltaRHadTopLepB

3) BtagHadBImpactParaMVA

$\text{corr}(\text{DeltaRHadTopLepB}, \text{DeltaThetaHadWLepB}) < 40\%$

4) DeltaRLepBLEpton

4) DeltaRHadTopLepB

$\text{corr}(\text{DeltaRHadTopLepB}, \text{DeltaThetaHadWLepB}) > 40\%$

5) DeltaRHadBLEpB

5) DeltaThetaLepBLEpton

--> then var 6 can be kept in the list...

6) DeltaThetaHadWLepB

DeltaTheta instead of DletaR in this list!

6) DeltaRHadBLEpB

7) DeltaThetaHadTopHadW

7) DeltaRHadWLepton

8) DeltaRHadTopLepton

8) DeltaThetaHadTopHadW

9) DeltaThetaHadTopHadB

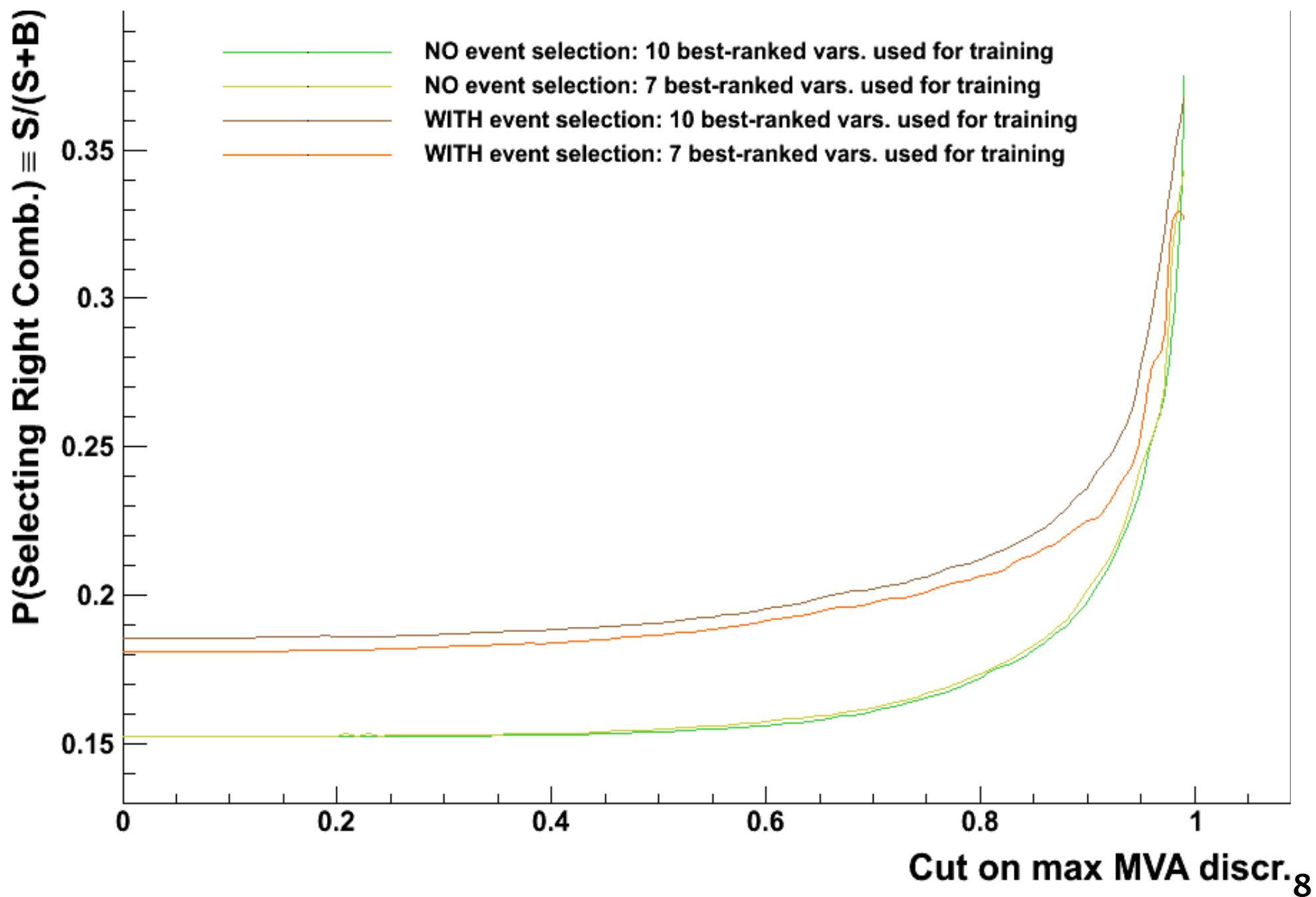
9) DeltaRHadBLEpton

10) DeltaRHadBLEpton

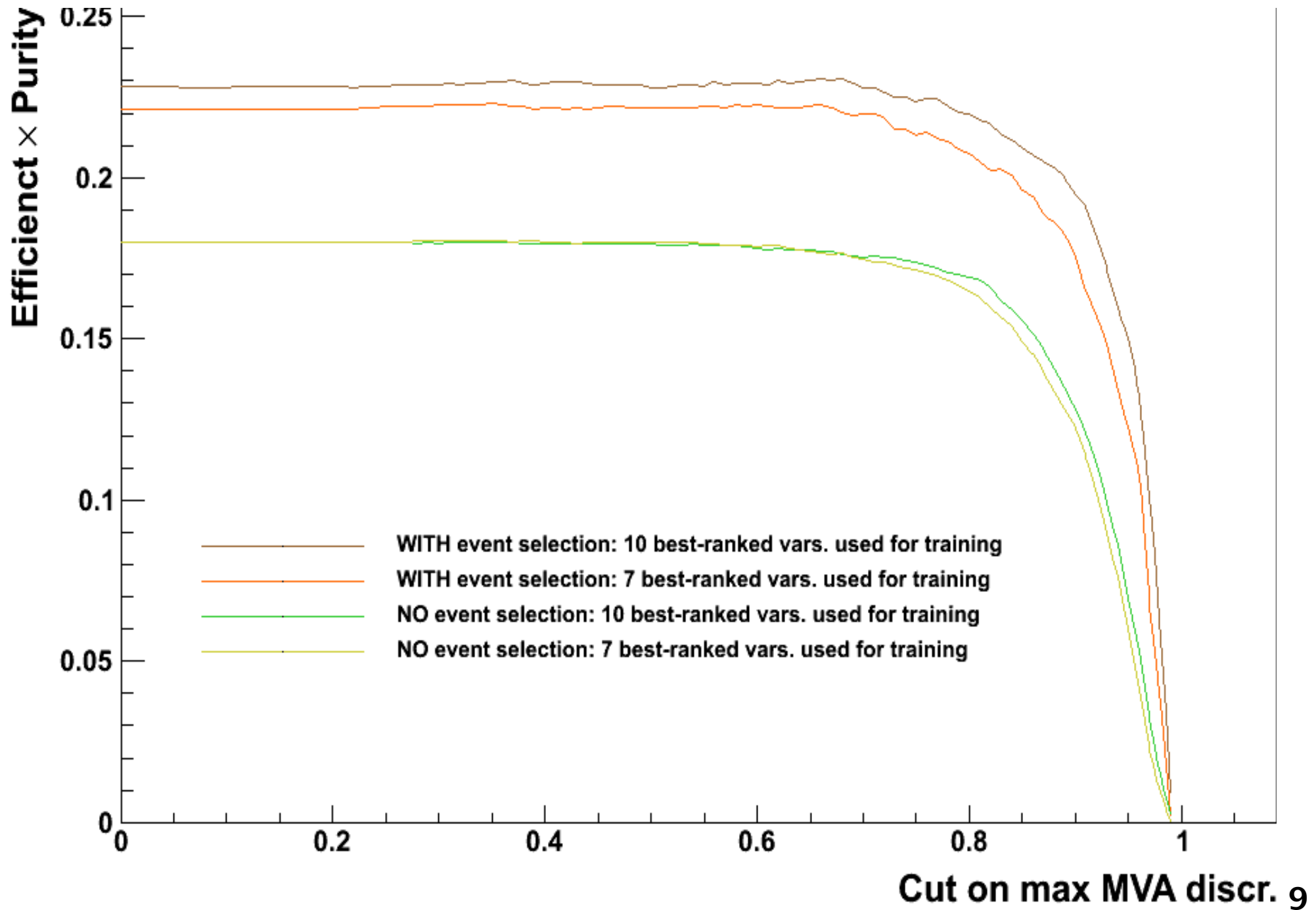
10) DeltaThetaHadTopHadB

Correlation  
> 50%,  
location in  
the list  
has been  
changed!

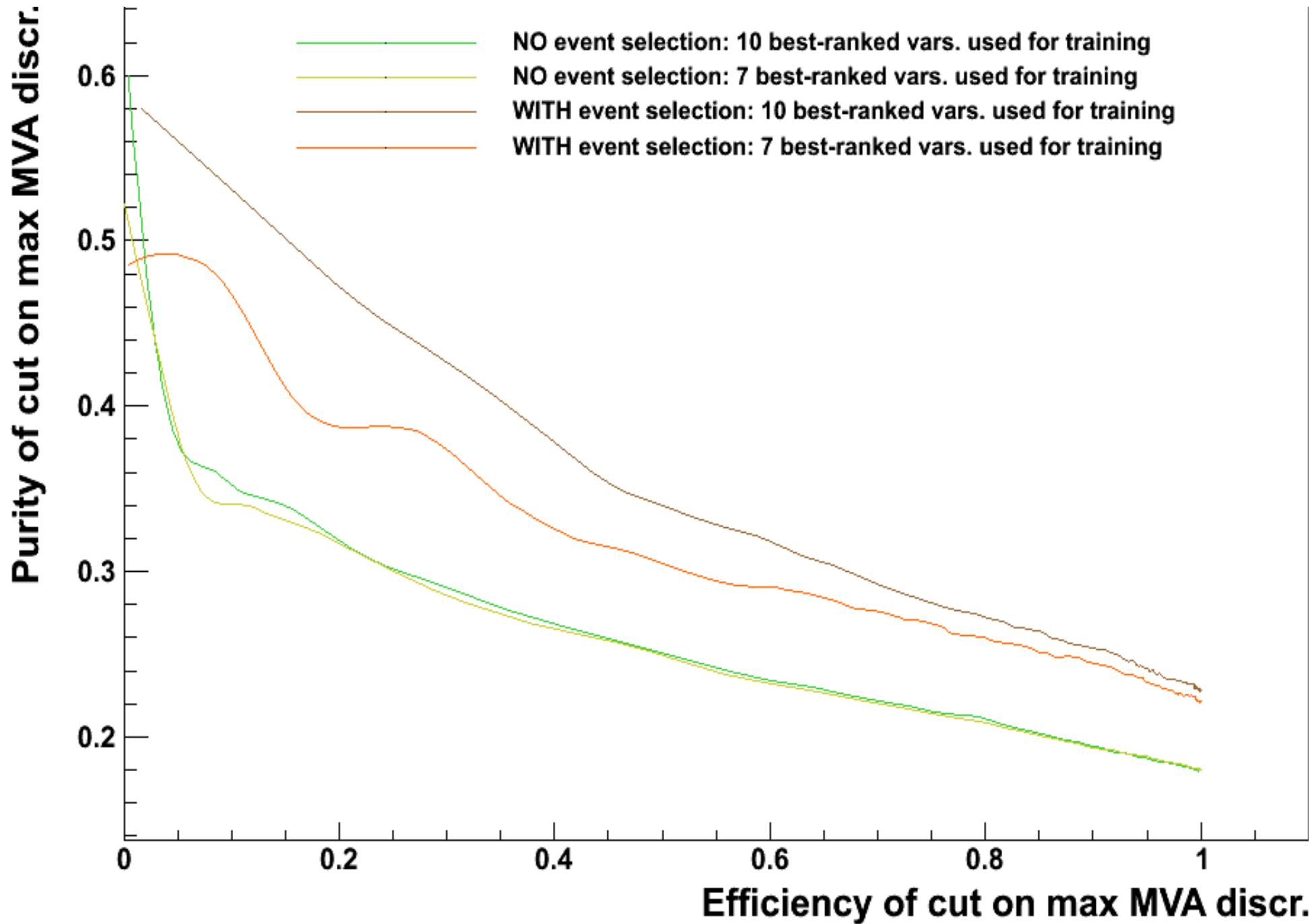
# Probability of selecting right jet combination vs. the cut on max MVA discriminator:



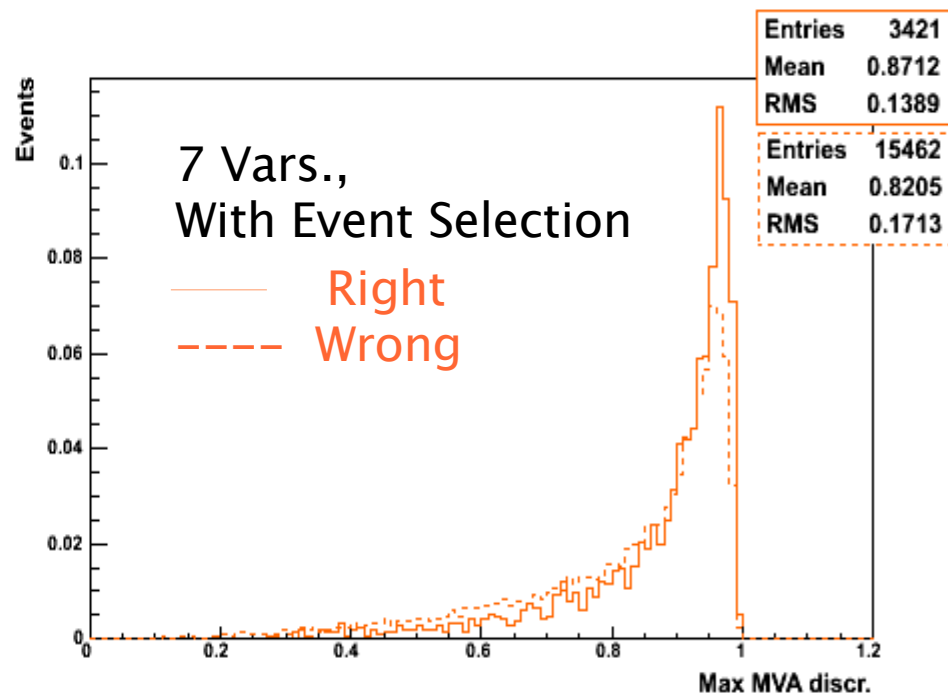
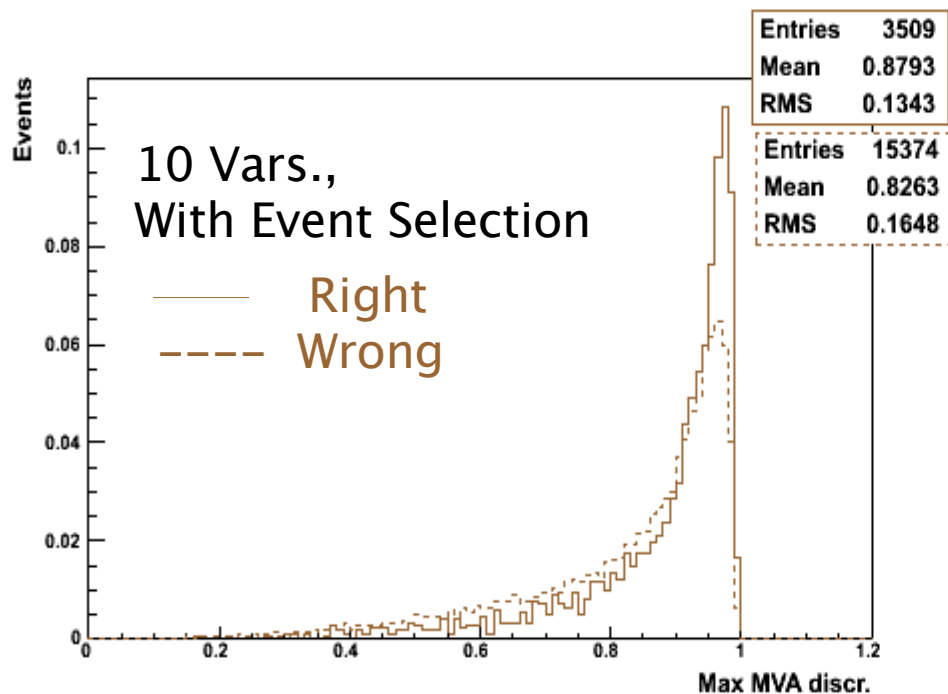
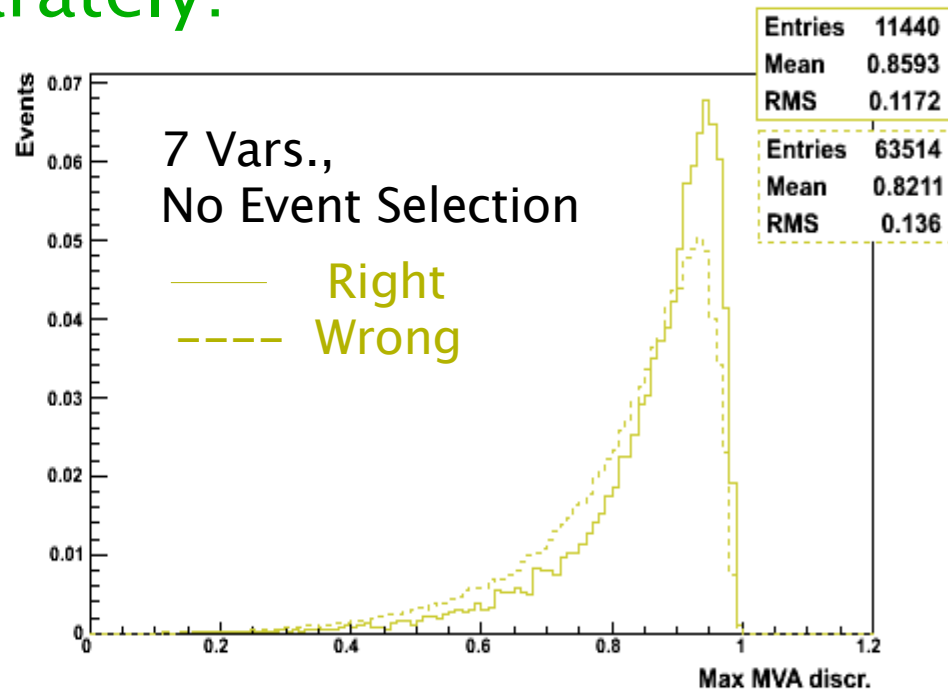
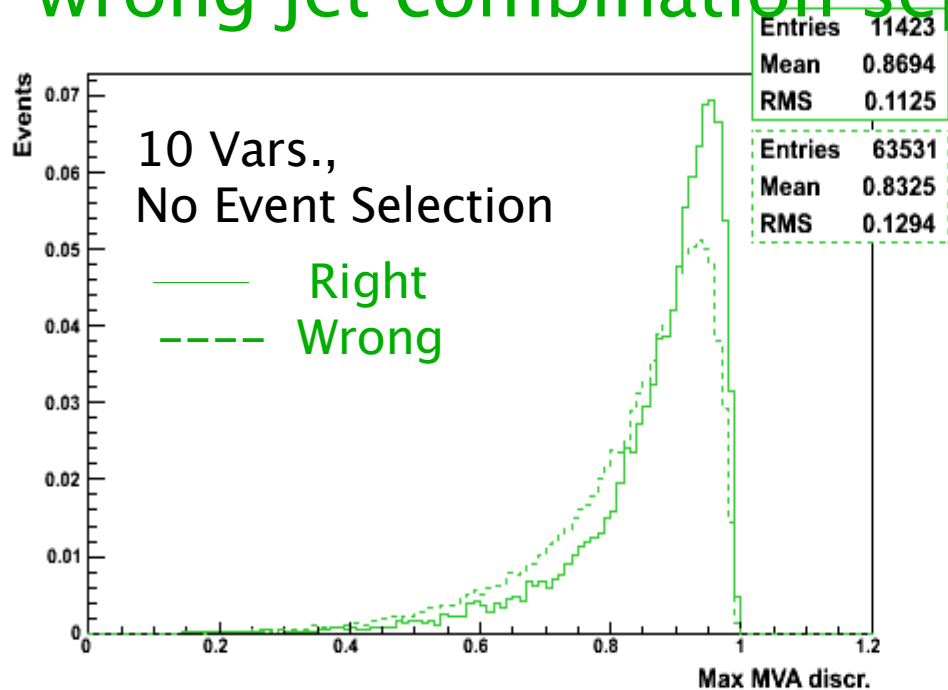
# Efficiency \* Purity of cut on max MVA discriminator vs. the cut on max MVA discriminator:



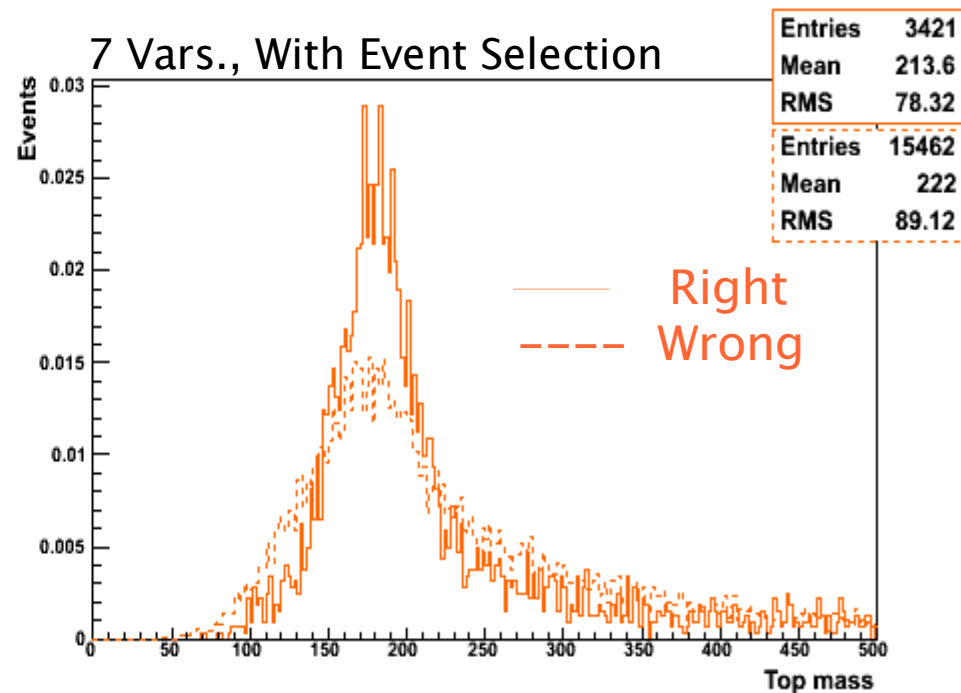
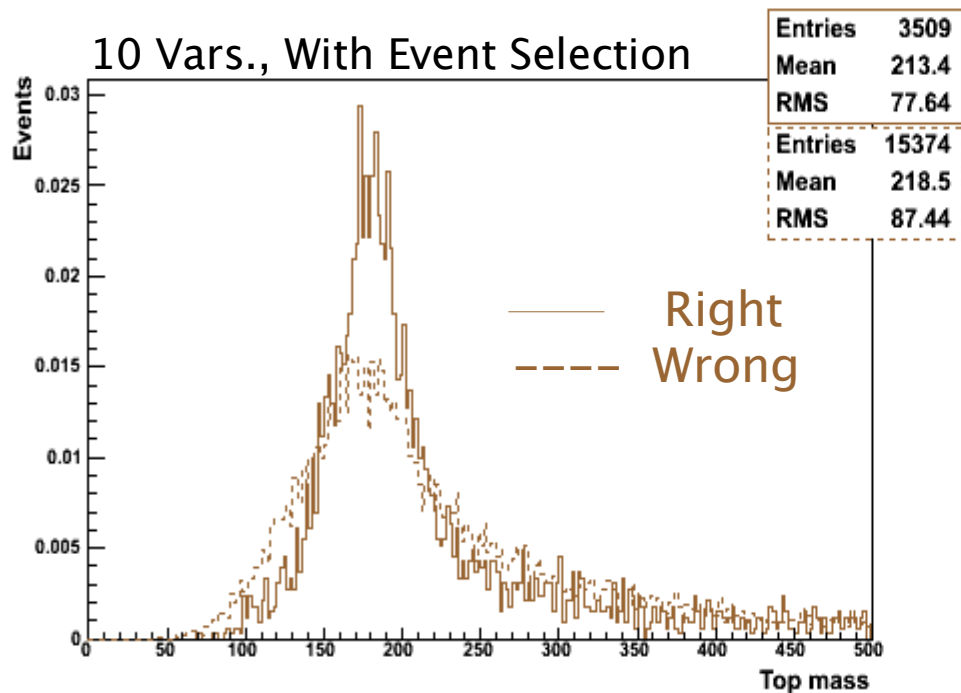
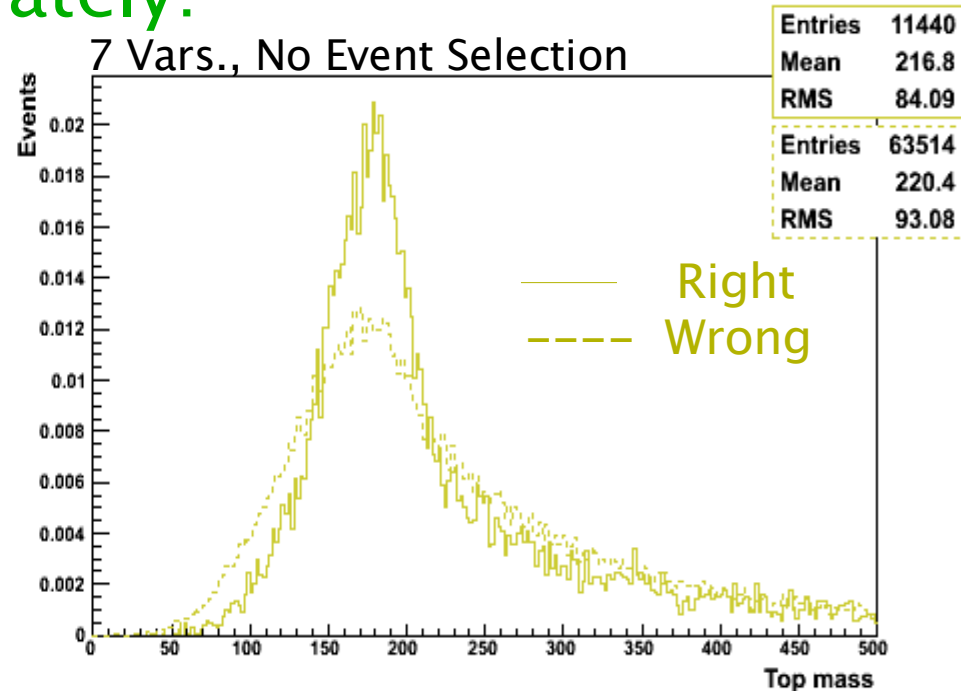
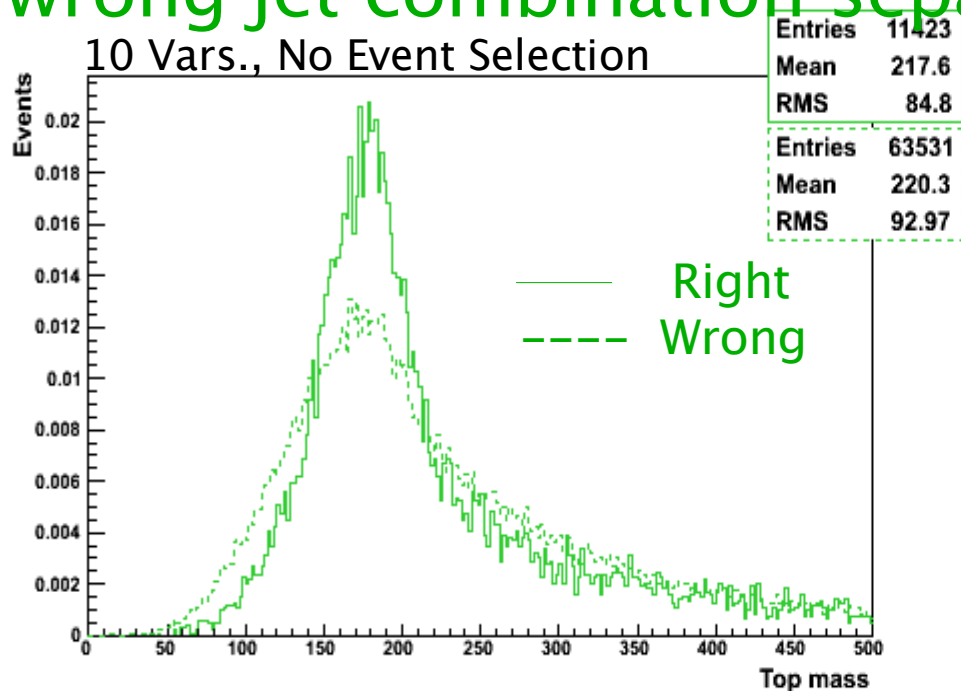
# Purity of cut on max MVA discriminator vs. Efficiency of cut on max MVA discriminator:



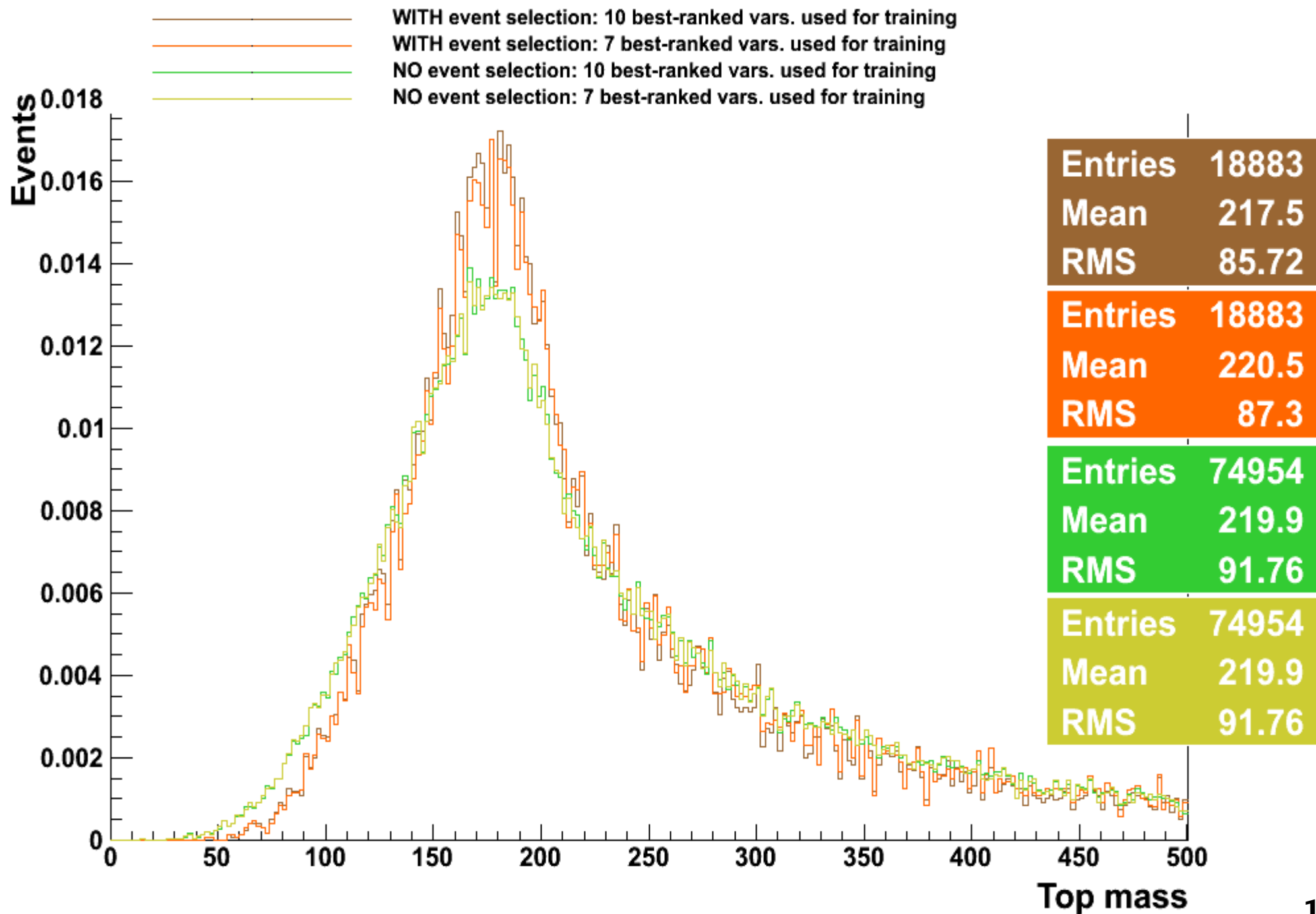
# Max MVA discriminator per event shown for right & wrong jet combination separately:



# Top mass returned by MVA per event shown for right & wrong jet combination separately:



# Top mass returned by MVA method per event:



# Summary & to-do list:

### DONE #####

- The study of jet parton association using MVA method in the semi electronic decay channel of  $t\bar{t}$  events were presented.
- Two categories of best-ranked variables according to their separation power based on whether applying event selection cuts or not were introduced.
- The performance plots on how this method works in order to select the right jet combination were shown.

### TO BE DONE #####

- Check if the distribution of b-tagging variables fed to the MVA trainer, were correct.
- Check the performance of the MVA method using another set of variables for training (This time the ranking could be done w.r.t. the correlation coefficients: the smaller the better!). We've already seen that the separation power of almost all variables are less than 0.1: none of them is prior to the other one!
- Apply kinematic fit on the selected combination of jet parton from MVA and analyze the results.

**BACK-UP**

# List of all vars. used for monitoring:

## b-tagging related vars.

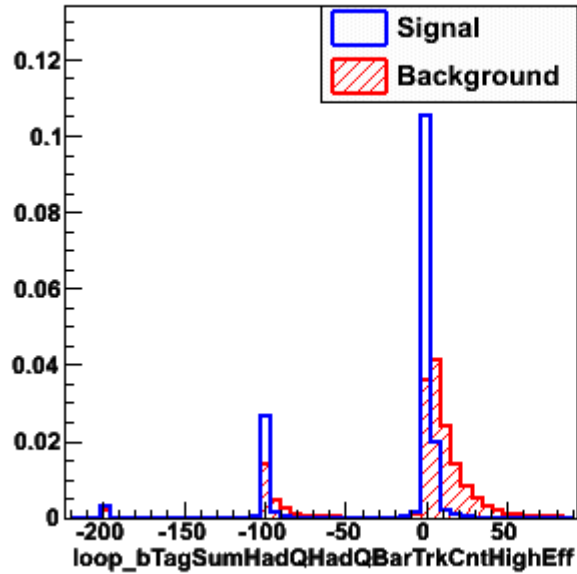
'bTagHadBImpactParaMVA'  
'bTagHadBTrkCntHighEff'  
'bTagHadBTrkCntHighPur'  
'bTagLepBImpactParaMVA'  
'bTagLepBTrkCntHighEff'  
'bTagLepBTrkCntHighPur'  
'bTagProdHadQHadQBarImpactParaMVA'  
'bTagProdImpactParaMVA'  
'bTagSumHadQHadQBarImpactParaMVA'  
'bTagSumHadQHadQBarTrkCntHighEff'  
'bTagSumHadQHadQBarTrkCntHighPur'  
'bTagSumImpactParaMVA'  
'bTagSumTrkCntHighEff'  
'bTagSumTrkCntHighPur'

## non b-tagging related vars.

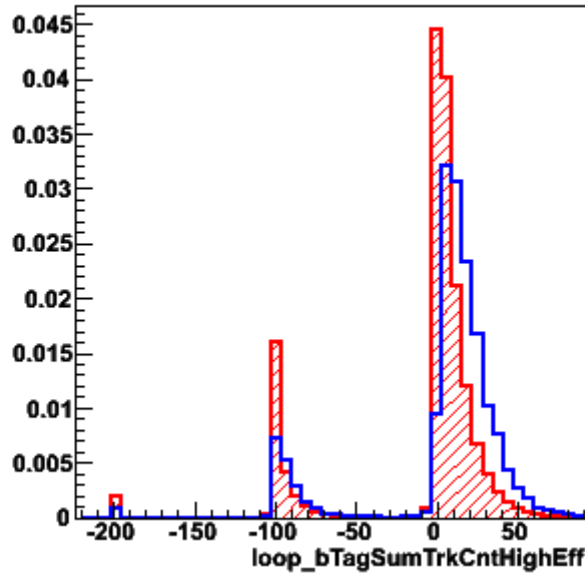
'deltaRHadBLepB'  
'deltaRHadBLEpton'  
'deltaRHadQHadQBar'  
'deltaRHadTopHadB'  
'deltaRHadTopHadW'  
'deltaRHadTopLepB'  
'deltaRHadTopLepton'  
'deltaRHadWHadB'  
'deltaRHadWLepB'  
'deltaRHadWLepton'  
'deltaRLepBLEpton'  
'deltaThetaHadBLepB'  
'deltaThetaHadBLEpton'  
'deltaThetaHadQHadQBar'  
'deltaThetaHadTopHadB'  
'deltaThetaHadTopHadW'  
'deltaThetaHadTopLepB'  
'deltaThetaHadTopLepton'  
'deltaThetaHadWHadB'  
'deltaThetaHadWLepB'  
'deltaThetaHadWLepton'  
'deltaThetaLepBLEpton'  
'relativePtHadronicTop'

# Distribution of input vars. used for training: NO event selection applied!

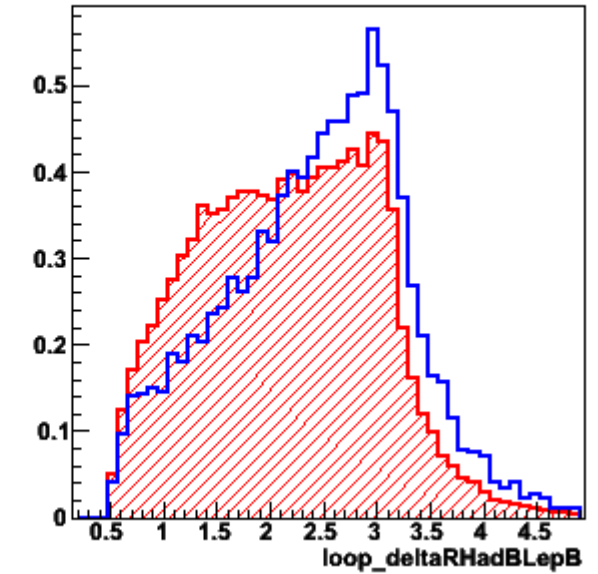
loop\_bTagSumHadQHadQBarTrkCntHighEff



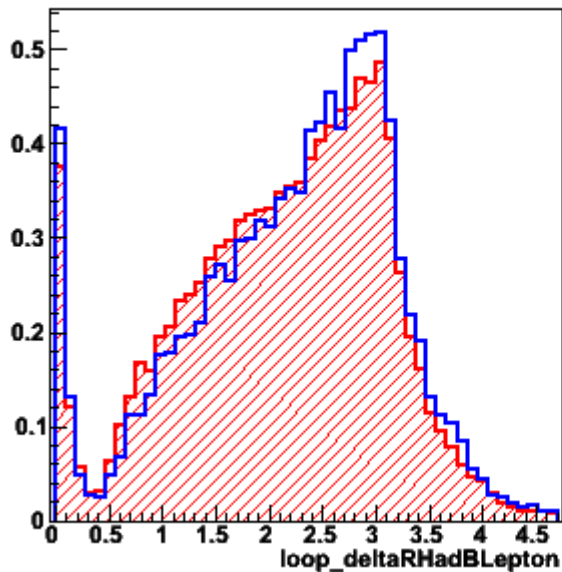
loop\_bTagSumTrkCntHighEff



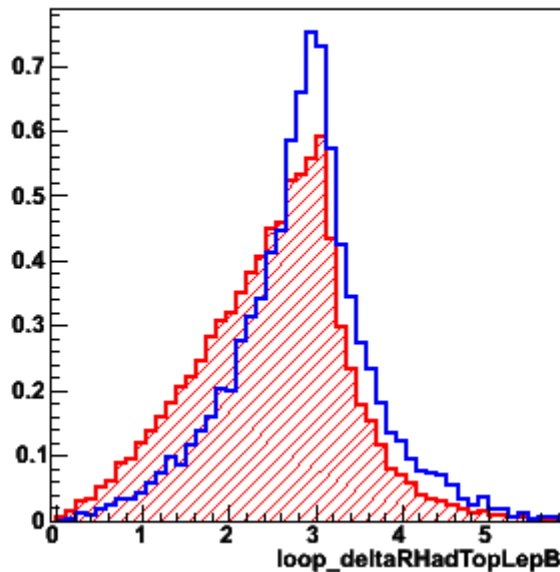
loop\_deltaRHadBLepB



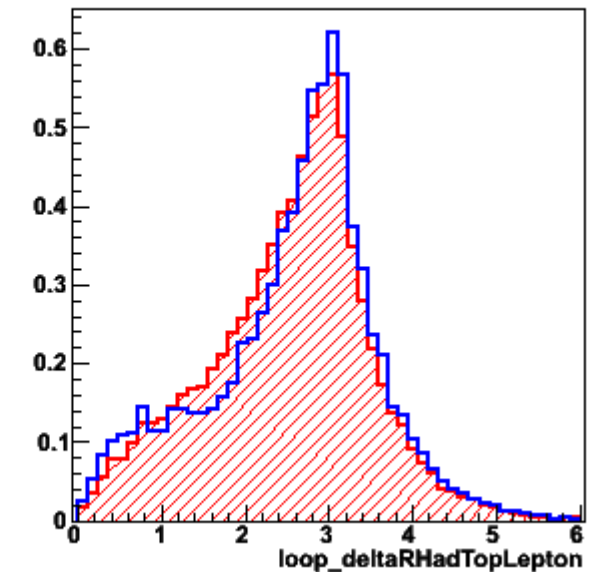
loop\_deltaRHadBLEpton



loop\_deltaRHadTopLepB

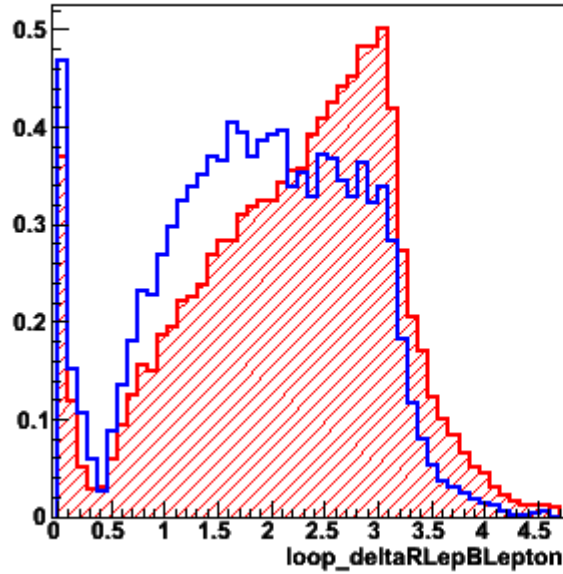


loop\_deltaRHadTopLepton

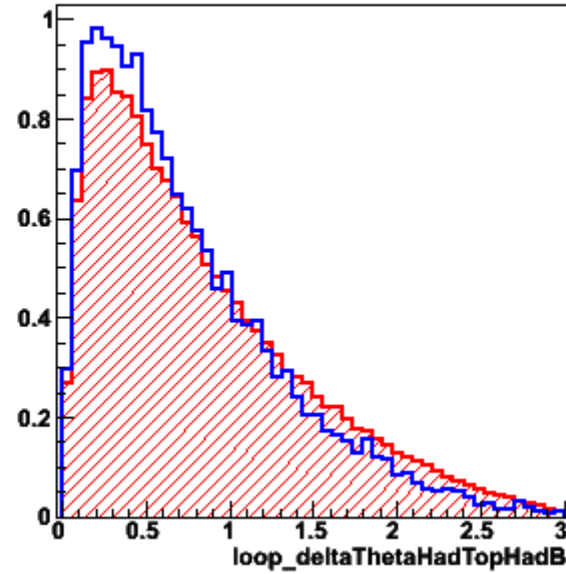


# Distribution of input vars. used for training: NO event selection applied!

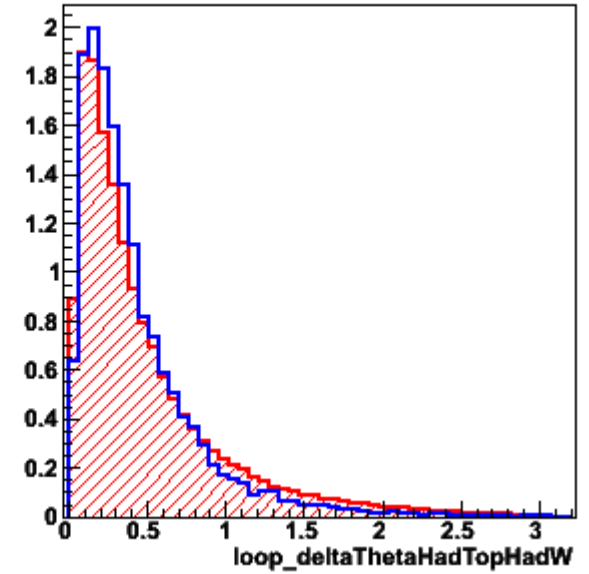
loop\_deltaRLepBLEpton



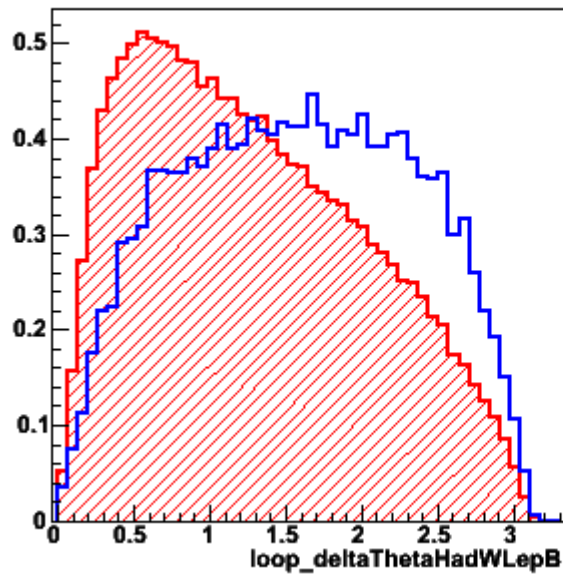
loop\_deltaThetaHadTopHadB



loop\_deltaThetaHadTopHadW



loop\_deltaThetaHadWLepB



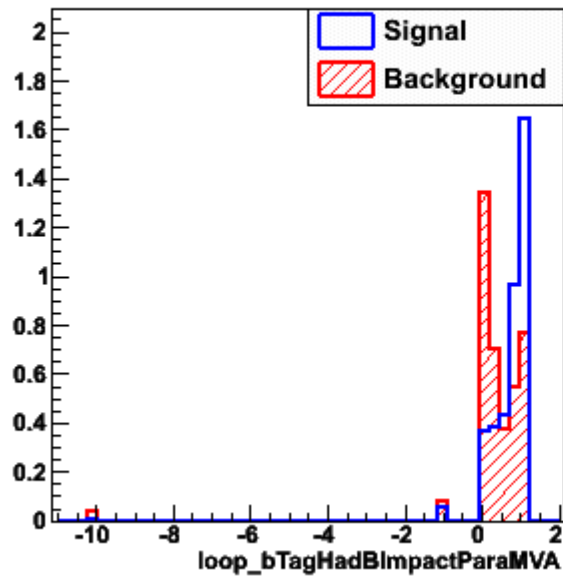
# Table of linear correlation coefficients of the trained variables: NO event selection applied!

**correlation matrix (signal + background)**

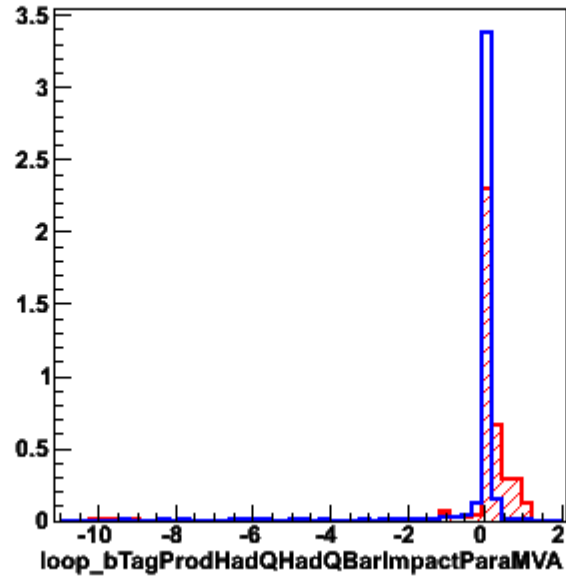
norm_bTagSumHadQHadQ	-0.109715	0.173314	0.100111	-0.0897981	0.145473	-0.0801976	0.00962459	0.00962459	-0.0549713	1
norm_bTagSumTrkCntHighEff	-0.00237572	0.155616	0.147594	0.0333637	0.168988	-0.0823141	-0.0861582	-0.0861582	1	-0.0549713
norm_deltaRLepton	-0.0806726	0.0685908	0.0663281	-0.165919	-0.126763	-0.102922	-0.0621619	1	-0.0861582	0.00962459
norm_deltaRHadLepton	-0.0806726	-0.0324077	-0.1629	-0.000226709	0.111075	0.361187	1	-0.0621619	-0.0861582	0.00962459
norm_deltaRHadTopLepton	0.00235587	-0.155873	-0.126082	-0.0262873	-0.119966	1	0.361187	-0.102922	-0.0823141	-0.0801976
norm_deltaThetaHadWLepton	-0.0841753	0.320823	0.191938	0.369026	1	-0.119966	0.111075	-0.126763	0.168988	0.145473
norm_deltaRHadTopLeptonB	0.308575	-0.121205	-0.0824359	1	0.369026	-0.0262873	-0.000226709	-0.165919	0.0333637	-0.0897981
norm_deltaThetaHadTopHadB	-0.115534	0.120677	1	-0.0824359	0.191938	-0.126082	-0.1629	0.0663281	0.147594	0.100111
norm_deltaThetaHadTopHadW	-0.0368351	1	0.120677	-0.121205	0.320823	-0.155873	-0.0324077	0.0685908	0.155616	0.173314
norm_deltaRHadBLepB	1	-0.0368351	-0.115534	0.308575	-0.0841753	0.00235587	-0.0806726	-0.0806726	-0.00237572	-0.109715

# Distribution of input vars. used for training: event selection applied!

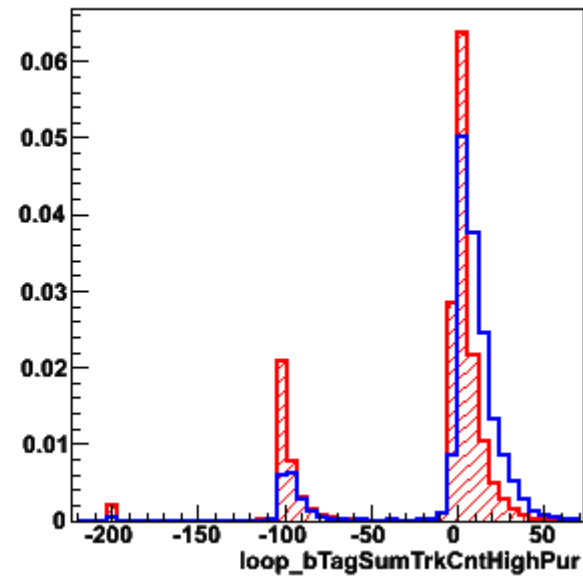
loop\_bTagHadBImpactParaMVA



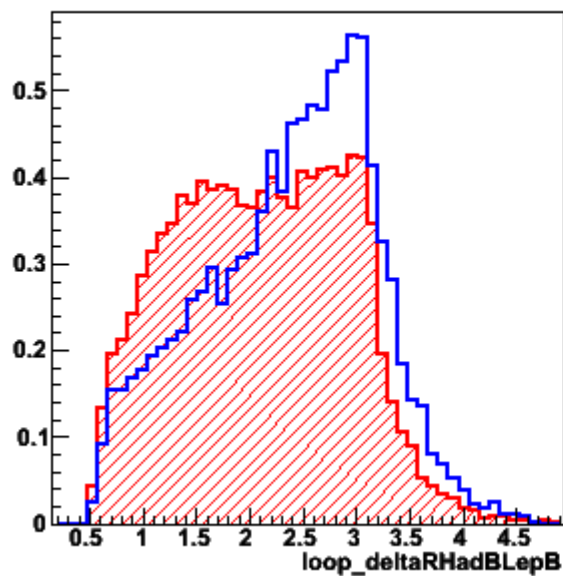
loop\_bTagProdHadQHadQBarImpactParaMVA



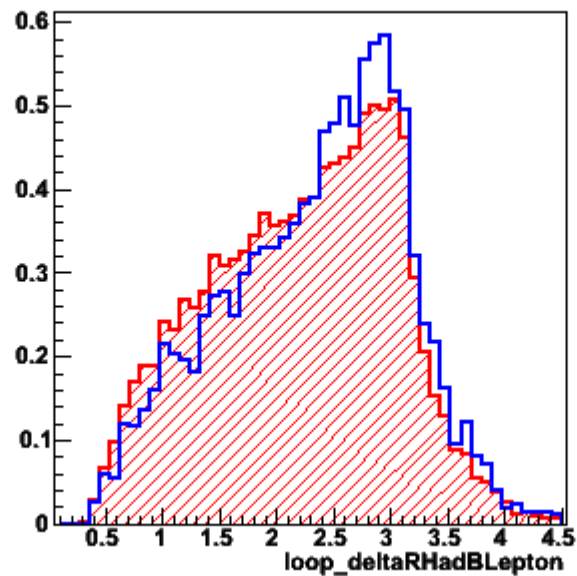
loop\_bTagSumTrkCntHighPur



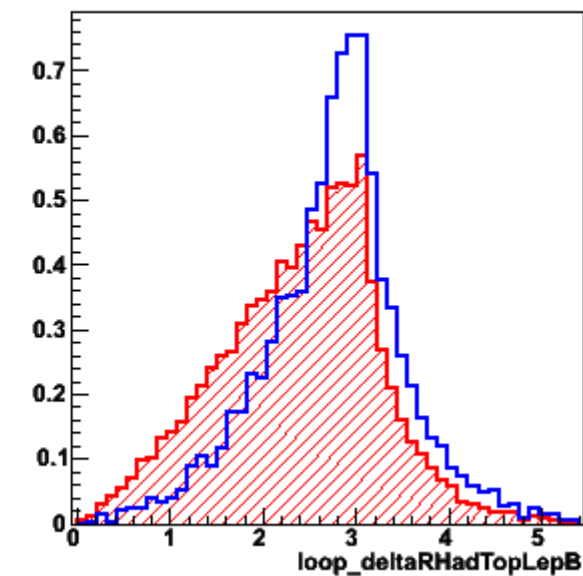
loop\_deltaRHadBLepB



loop\_deltaRHadBLEpton

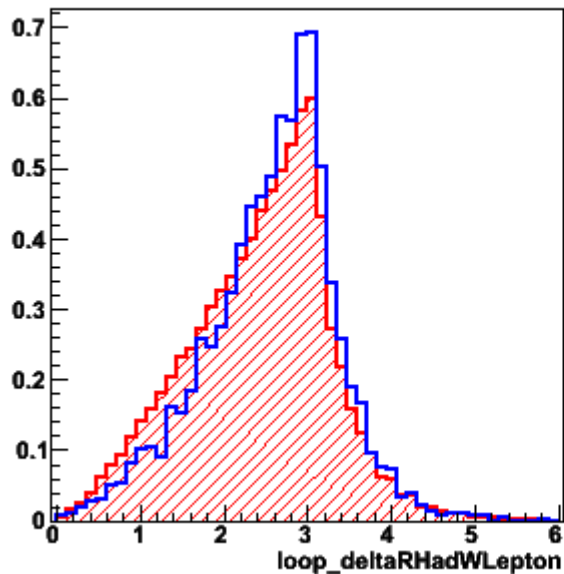


loop\_deltaRHadTopLepB

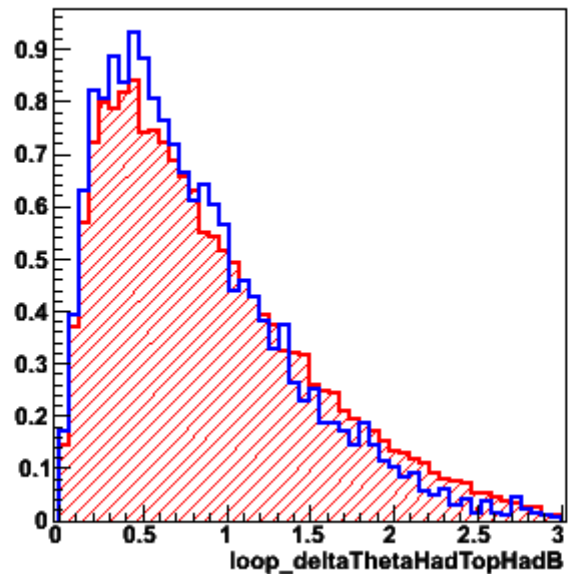


# Distribution of input vars. used for training: event selection applied!

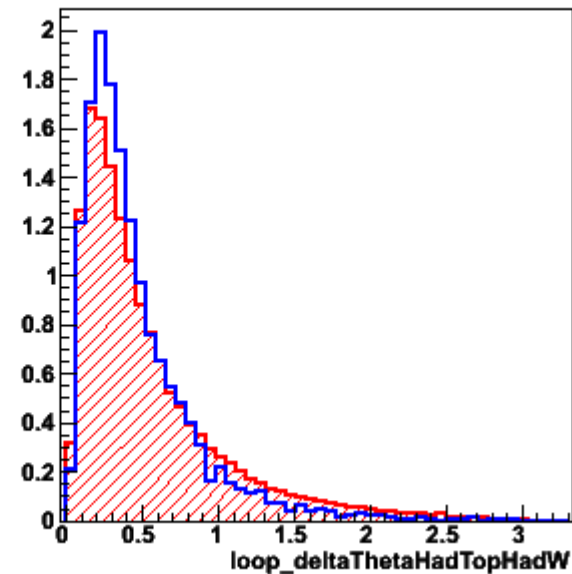
loop\_deltaRHadWLepton



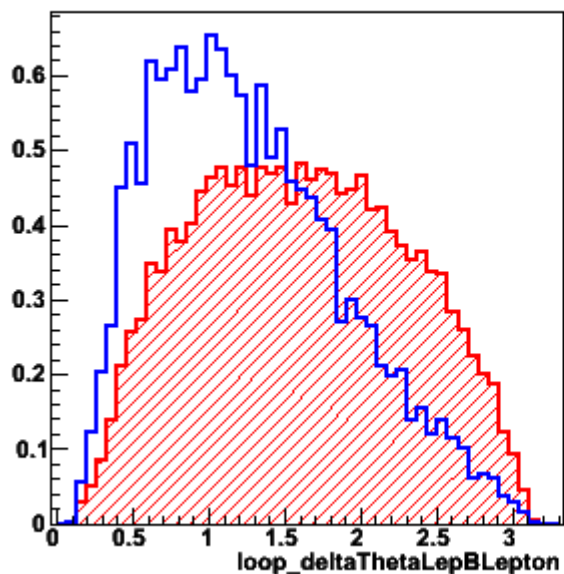
loop\_deltaThetaHadTopHadB



loop\_deltaThetaHadTopHadW



loop\_deltaThetaLepBLEpton



# Table of linear correlation coefficients of the trained variables: event selection applied!

**correlation matrix (signal + background)**

ImpactParaMVA	-0.14985	-0.0256545	0.0167175	-0.0528278	-0.0511414	0.057707	0.0491577	-0.206856	-0.183751	1
norm_bTagProdHadQHadC	0.0492103	0.0592063	0.00740097	0.0288427	0.00561279	-0.0424709	-0.00140092	0.318866	1	-0.183751
ImpactParaMVA	0.105814	0.132337	0.064335	-0.0424399	0.0420419	-0.148801	0.0431408	1	0.318866	-0.206856
norm_bTagSumTrkCntHighPur	-0.157558	0.236778	0.180812	-0.180412	-0.0783688	-0.0823772	1	0.0431408	-0.00140092	0.0491577
norm_deltaThetaHadTopHadB	-0.195091	-0.0235794	-0.253528	-0.00647692	-0.149366	1	-0.0823772	-0.148801	-0.0424709	0.057707
norm_deltaThetaHadTopHadW	0.30865	-0.241741	-0.0423746	-0.0625188	1	-0.149366	-0.0783688	0.0420419	0.00561279	-0.0511414
norm_deltaThetaHadTopHadR	0.330717	-0.116557	-0.0739365	1	-0.0625188	-0.00647692	-0.180412	-0.0424399	0.0288427	-0.0528278
norm_deltaRHadWLepton	-0.109077	0.0328233	1	-0.0739365	-0.0423746	-0.253528	0.180812	0.064335	0.00740097	0.0167175
norm_deltaRHadTopLepB	-0.0374178	1	0.0328233	-0.116557	-0.241741	-0.0235794	0.236778	0.132337	0.0592063	-0.0256545
norm_deltaRHadBLEpton	1	-0.0374178	-0.109077	0.330717	0.30865	-0.195091	-0.157558	0.105814	0.0492103	-0.14985