MVA tools to chose the jet combination

- Introduction
  - What has been checked and what was the aim?
- Methods:
  - Likelihood ratio and neural network approach with and without decorrelated variables as input
- Plots and numbers
  - Summary of results and conclusions
- Next steps towards JEC estimation

Petra Van Mulders

IIHE-Top Quark Meeting
May 6th 2009

Aim

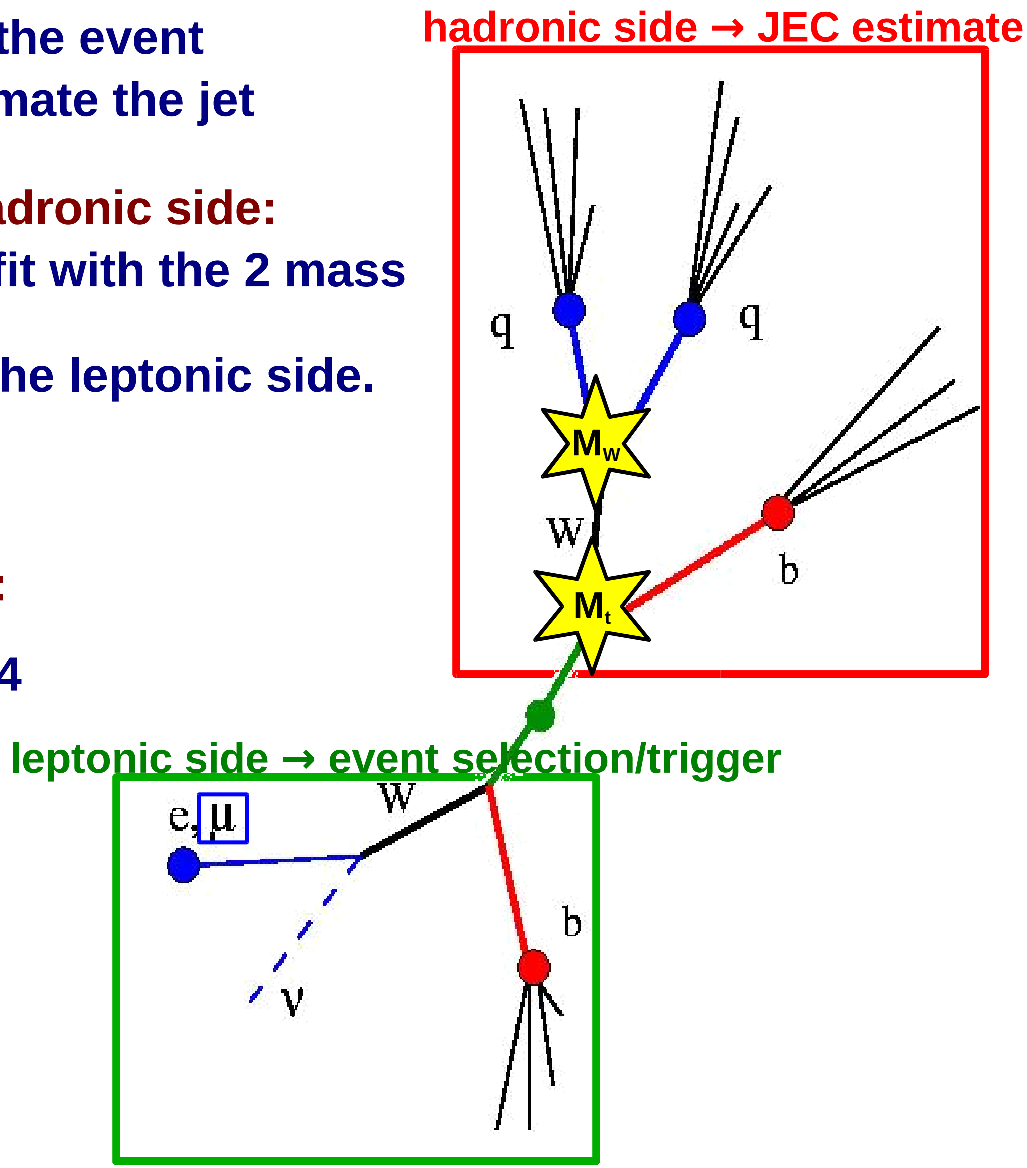
- **Chose variables:**
  - **High discriminating power**
  - **Stable ordering according to discriminating power after different event selections**
  - **Less then 40% correlated**
  - **No use of MET (not well known at startup)**
  - **< 20% correlation with hadronic W boson or top quark mass (to avoid bias)**
- **Study 4 methods to combine the selected variables:**
  - **Likelihood Ratio (LR) with and without decorrelated input variables**
  - **Neural Network (NN) with and without decorrelated input variables**

- **Aim:**
  - **Check purity versus efficiency when applying each of the 4 MVA methods (performance)**
  - **Check the effect of applying tight and less tight event selection:**
    - **On performance**
    - **On the jet combination with the highest MVA discriminant value: is the chosen jet combination the same in both cases? (overlap)**

⇒ is it possible to study the event selection afterwards?

Why is this study needed?

- **Semi-leptonic channel:**
  - **Leptonic side used to select the event**
  - **Hadronic side is used to estimate the jet energy calibration factors**
- **Need to identify the jets in the hadronic side:**
  - **Application of the kinematic fit with the 2 mass constraints**
  - **Don't care about the b-jet at the leptonic side.**
  - **Use 4 highest p_T jets as input**



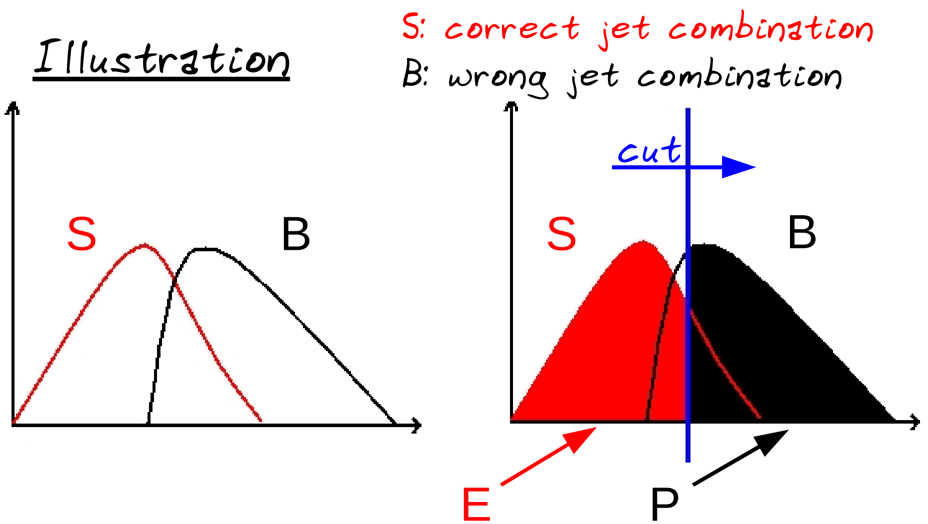
LOOSE versus TIGHT selection cuts:

- **p_T(jets) > 30 versus 40 GeV/c, $|\eta|$ <2.4**
- **$p_T(\mu)$ > 20 versus 30 GeV/c, $|\eta|$<2.1**
- **μ isolated:**
  - → **RelIso > 0.90 versus 0.95**
  - → **VetoConeEM < 6 versus 4 GeV**
  - → **VetoConeHAD < 8 versus 6 GeV**
- **non-overlapping jets:** **ΔR(jet i,jet j) > 1.0**

Variables & discriminating power

- **136 variables are implemented which can be put in different categories:**
  - **cat. 1:** Mass, p_T, η, ϕ, θ of single candidates of the ttbar system
  - **cat. 2:** Variables which compare 2 candidates of the ttbar system: Δm, ΔR, $\Delta\phi$ or $\Delta\Omega$.
  - **cat. 3:** Special variables combining the p_T of the jets
  - **cat. 4:** Variables based on b-tag information with 6 different b-tag algorithms
    - Hadronic b, Leptonic b, or combination of both (sum or multiplication)

- **Discriminating power of the variables?**
  - Surface method: calculate the overlapping surface of the signal and background distributions: the smaller, the better!
  - PTE method: the higher, the better!
    - search maximum of P*E while moving cut from left to right
    - search maximum of P*E while moving cut from right to left
    - take maximum of the 2 calculated maxima as handle on the discriminating power



Ordering the variables

- **Table of variables ordered according to discriminating power (2 definitions)**
  - **Do ordering again starting from the variable with the highest discriminating power and after the following requirements:**
  - **Remove variables related to MET**
  - **< 40% correlated with each other and < 20 % with m_W, m_t**

LOOSE		TIGHT	
SURFACE	**PTE**	**SURFACE**	**PTE**
bTagProdImpactParaMVA	bTagProdImpactParaMVA	bTagProdImpactParaMVA	bTagProdImpactParaMVA
bTagHadBTrkCntHighEff	bTagHadBTrkCntHighEff	bTagHadBTrkCntHighEff	bTagHadBTrkCntHighEff
bTagLepBTrkCntHighEff	bTagLepBTrkCntHighEff	bTagLepBTrkCntHighEff	relativePtHadronicTop
relativePtHadronicTop	relativePtHadronicTop	relativePtHadronicTop	bTagLepBTrkCntHighEff
$\Delta\Omega$ (HadTop,HadW)	$\Delta\Omega$ (LepB,Muon)	$\Delta\Omega$ (HadTop,HadW)	$\Delta\Omega$ (LepB,Muon)
$\Delta\Omega$ (LepB,Muon)	$\Delta\Omega$ (HadTop,LepB)	$\Delta\Omega$ (LepB,Muon)	$\Delta\Omega$ (HadW,LepB)
$\Delta\Omega$ (HadTop,LepB)	$\Delta\Omega$ (HadTop,Muon)	$\Delta\Omega$ (HadTop,HadB)	$\Delta\Omega$ (HadTop,Muon)
$\Delta\Omega$ (HadTop,HadB)	ΔR(HadB,LepB)	$\Delta\Omega$ (HadW,LepB)	$\Delta\Omega$ (HadTop,HadW)
$\Delta\Omega$ (HadTop,Muon)	θHadW	$\Delta\Omega$ (HadTop,Muon)	$\Delta\Omega$ (HadB,LepB)
bTagSumSoftMuon	ΔR(HadB,Muon)	$\Delta\Omega$ (HadB,LepB)	$\Delta\phi$ (HadB,Muon)
ΔR(HadB,LepB)	$\Delta\Omega$ (HadTop,HadB)	$\Delta\phi$ (HadTop,Muon)	$\Delta\phi$ (HadTop,Muon)
ΔR(HadB,Muon)	$\Delta\phi$ (HadTop,Muon)	ΔR(HadB,Muon)	$\Delta\Omega$ (HadTop,HadB)

Ordering the variables

- **Table of variables ordered according to discriminating power (2 definitions)**
  - **Do ordering again starting from the variable with the highest discriminating power and after the following requirements:**
  - **Remove variables related to MET**
  - **< 40% correlated with each other and < 20 % with m_W, m_t**

Only 1 variable which makes use of b-tag information!

LOOSE		TIGHT	
SURFACE	**PTE**	**SURFACE**	**PTE**
bTagProdImpactParaMVA	bTagProdImpactParaMVA	bTagProdImpactParaMVA	bTagProdImpactParaMVA
~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~
~~bTagLepBTrkCntHighEff~~	~~bTagLepBTrkCntHighEff~~	~~bTagLepBTrkCntHighEff~~	relativePtHadronicTop
relativePtHadronicTop	relativePtHadronicTop	relativePtHadronicTop	~~bTagLepBTrkCntHighEff~~
$\Delta\Omega$ (HadTop,HadW)	$\Delta\Omega$ (LepB,Muon)	$\Delta\Omega$ (HadTop,HadW)	$\Delta\Omega$ (LepB,Muon)
$\Delta\Omega$ (LepB,Muon)	$\Delta\Omega$ (HadTop,LepB)	$\Delta\Omega$ (LepB,Muon)	$\Delta\Omega$ (HadW,LepB)
$\Delta\Omega$ (HadTop,LepB)	$\Delta\Omega$ (HadTop,Muon)	$\Delta\Omega$ (HadTop,HadB)	$\Delta\Omega$ (HadTop,Muon)
$\Delta\Omega$ (HadTop,HadB)	ΔR(HadB,LepB)	$\Delta\Omega$ (HadW,LepB)	$\Delta\Omega$ (HadTop,HadW)
$\Delta\Omega$ (HadTop,Muon)	θHadW	$\Delta\Omega$ (HadTop,Muon)	$\Delta\Omega$ (HadB,LepB)
~~bTagSumSoftMuon~~	ΔR(HadB,Muon)	$\Delta\Omega$ (HadB,LepB)	$\Delta\phi$ (HadB,Muon)
ΔR(HadB,LepB)	$\Delta\Omega$ (HadTop,HadB)	$\Delta\phi$ (HadTop,Muon)	$\Delta\phi$ (HadTop,Muon)
ΔR(HadB,Muon)	$\Delta\phi$ (HadTop,Muon)	ΔR(HadB,Muon)	$\Delta\Omega$ (HadTop,HadB)

Ordering the variables

- **Table of variables ordered according to discriminating power (2 definitions)**
  - **Do ordering again starting from the variable with the highest discriminating power and after the following requirements:**
  - **Remove variables related to MET**
  - **< 40% correlated with each other and < 20 % with m_W, m_t**

LOOSE SURFACE	LOOSE PTE (crossed out)	TIGHT SURFACE	TIGHT PTE (crossed out)
bTagProdImpactParaMVA	bTagProdImpactParaMVA	bTagProdImpactParaMVA	bTagProdImpactParaMVA
~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~
~~bTagLepBTrkCntHighEff~~	~~bTagLepBTrkCntHighEff~~	~~bTagLepBTrkCntHighEff~~	relativePtHadronicTop
relativePtHadronicTop	relativePtHadronicTop	relativePtHadronicTop	~~bTagLepBTrkCntHighEff~~
ΔΩ (HadTop,HadW)	ΔΩ (LepB,Muon)	ΔΩ (HadTop,HadW)	ΔΩ (LepB,Muon)
ΔΩ (LepB,Muon)	ΔΩ (HadTop,LepB)	ΔΩ (LepB,Muon)	ΔΩ (HadW,LepB)
ΔΩ (HadTop,LepB)	ΔΩ (HadTop,Muon)	ΔΩ (HadTop,HadB)	ΔΩ (HadTop,Muon)
ΔΩ (HadTop,HadB)	ΔR(HadB,LepB)	ΔΩ (HadW,LepB)	ΔΩ (HadTop,HadW)
ΔΩ (HadTop,Muon)	θHadW	ΔΩ (HadTop,Muon)	ΔΩ (HadB,LepB)
~~bTagSumSoftMuon~~	ΔR(HadB,Muon)	ΔΩ (HadB,LepB)	Δϕ (HadB,Muon)
ΔR(HadB,LepB)	ΔΩ (HadTop,HadB)	Δϕ (HadTop,Muon)	Δϕ (HadTop,Muon)
ΔR(HadB,Muon)	Δϕ (HadTop,Muon)	ΔR(HadB,Muon)	ΔΩ (HadTop,HadB)

PTE is sensitive to the fact that the signal and background distributions cross each other several times (in any case result the same, but preference for surface method)

Ordering the variables

- **Table of variables ordered according to discriminating power (2 definitions)**
  - **Do ordering again starting from the variable with the highest discriminating power and after the following requirements:**
  - **Remove variables related to MET**
  - **< 40% correlated with each other and < 20 % with m_W, m_t**

use 7 first variables

LOOSE SURFACE	LOOSE PTE	TIGHT SURFACE	TIGHT PTE
bTagProdImpactParaMVA	bTagProdImpactParaMVA	bTagProdImpactParaMVA	bTagProdImpactParaMVA
~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~	~~bTagHadBTrkCntHighEff~~
~~bTagLepBTrkCntHighEff~~	~~bTagLepBTrkCntHighEff~~	~~bTagLepBTrkCntHighEff~~	relativePtHadronicTop
relativePtHadronicTop	relativePtHadronicTop	relativePtHadronicTop	~~bTagLepBTrkCntHighEff~~
ΔΩ (HadTop,HadW)	ΔΩ (LepB,Muon)	ΔΩ (HadTop,HadW)	ΔΩ (LepB,Muon)
ΔΩ (LepB,Muon)	ΔΩ (HadTop,LepB)	ΔΩ (LepB,Muon)	ΔΩ (HadW,LepB)
ΔΩ (HadTop,LepB)	ΔΩ (HadTop,Muon)	ΔΩ (HadTop,HadB)	ΔΩ (HadTop,Muon)
ΔΩ (HadTop,HadB)	ΔR(HadB,LepB)	ΔΩ (HadW,LepB)	ΔΩ (HadTop,HadW)
ΔΩ (HadTop,Muon)	θHadW	ΔΩ (HadTop,Muon)	ΔΩ (HadB,LepB)
~~bTagSumSoftMuon~~	ΔR(HadB,Muon)	ΔΩ (HadB,LepB)	Δφ (HadB,Muon)
ΔR(HadB,LepB)	ΔΩ (HadTop,HadB)	Δφ (HadTop,Muon)	Δφ (HadTop,Muon)
ΔR(HadB,Muon)	Δφ (HadTop,Muon)	ΔR(HadB,Muon)	ΔΩ (HadTop,HadB)

Correlation is 77%: so we can conclude that they contain the same information

Correlation table 7 variables

Var1: ΔΩ (HadTop,HadB)
Var2: ΔΩ (HadTop,LepB)
Var3: ΔΩ (HadTop,Muon)
Var4: ΔΩ (HadTop.HadW)
Var5: ΔΩ (LepB,Muon)
Var6: relativePtHadTop
Var7: bTagProdIPMVA

Check for the 7 chosen variables if the correlation is the same after applying the loose or tight event selection

loose/tight

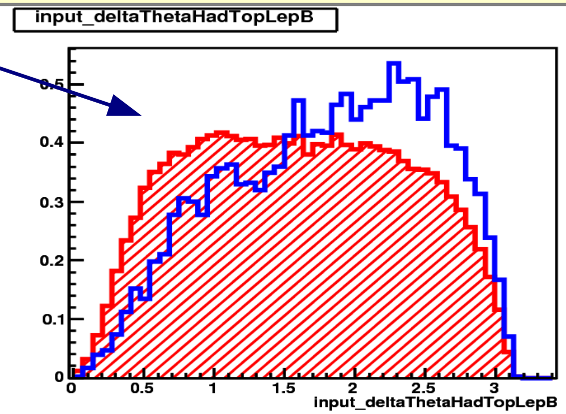
	Var1	Var2	Var3	Var4	Var5	Var6	Var7
Var1	**100**	8.9 / 9.6	16.0 / 17.5	13.8 / 14.7	3.2 / 0.6	13.0 / 15.3	0.4 / 0.5
Var2	8.9 / 9.6	**100**	9.6 / 9.5	-16.3 / -16.1	6.3 / 7.0	-11.1 / -11.4	1.2 / 3.9
Var3	16.0 / 17.5	9.6 / 9.5	**100**	14.3 / 13.8	11.8 / 10.9	23.8 / 24.1	2.5 / 1.1
Var4	13.8 / 14.7	-16.3 / -16.1	14.3 / 13.8	**100**	9.2 / 10.4	-16.5 / -16.9	0.9 / -3.9
Var5	3.2 / 0.6	6.3 / 7.0	11.8 / 10.9	9.2 / 10.4	**100**	-25.9 / -31.1	-4.2 / -6.5
Var6	13.0 / 15.3	-11.1 / -11.4	23.8 / 24.1	-16.5 / -16.9	-25.9 / -31.1	**100**	5.0 / 5.4
Var7	0.4 / 0.5	1.2 / 3.9	2.5 / 1.1	0.9 / -3.9	-4.2 / -6.5	5.0 / 5.4	**100**

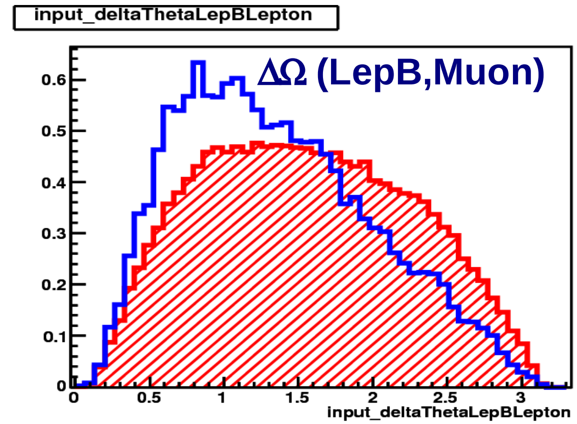
The correlations are the same within less then 5%.

numbers in %

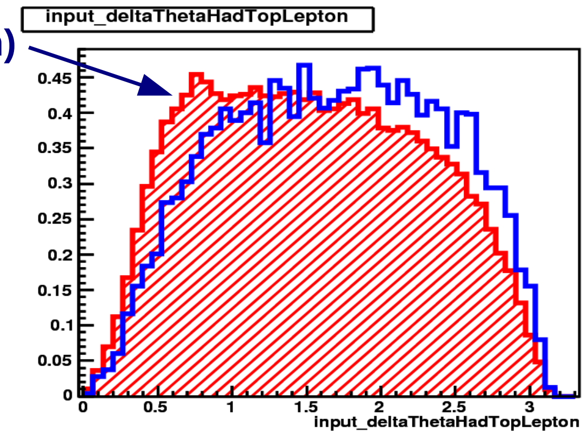
The 7 chosen variables

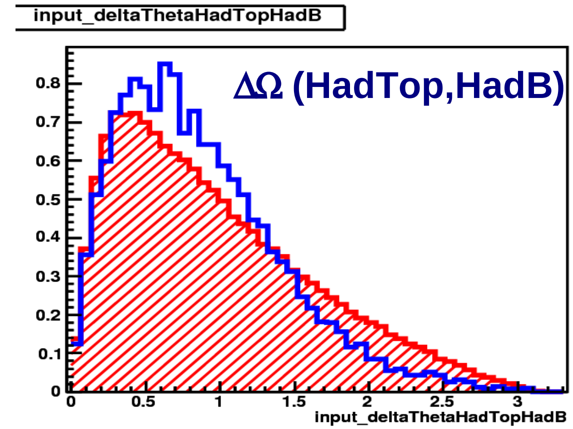
ΔΩ (HadTop,LepB)

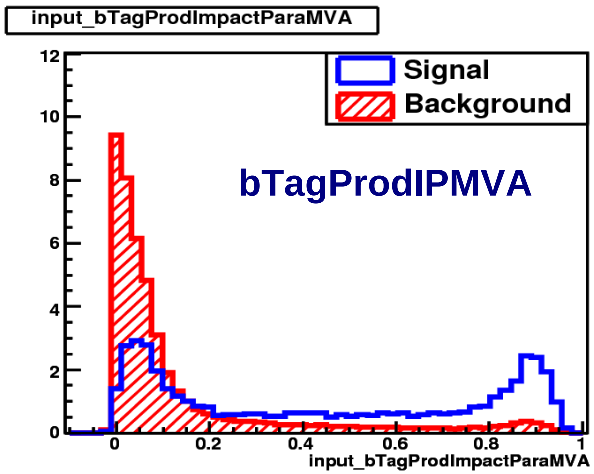




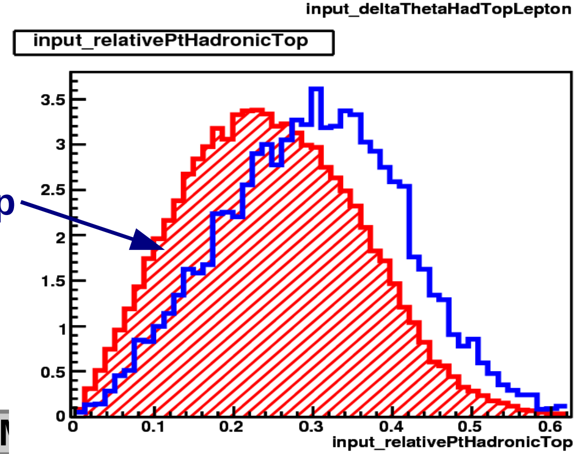
ΔΩ (HadTop,Muon)

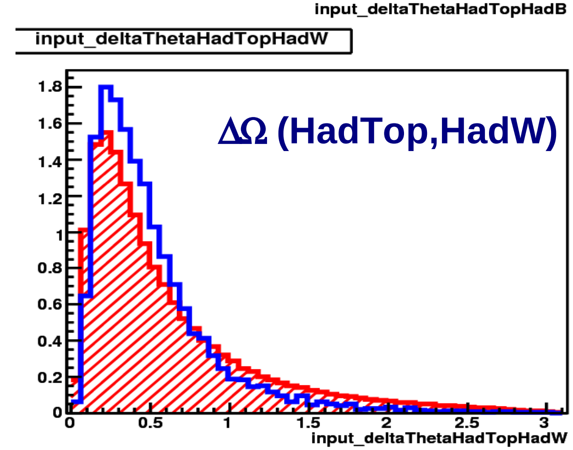






relativePtHadTop

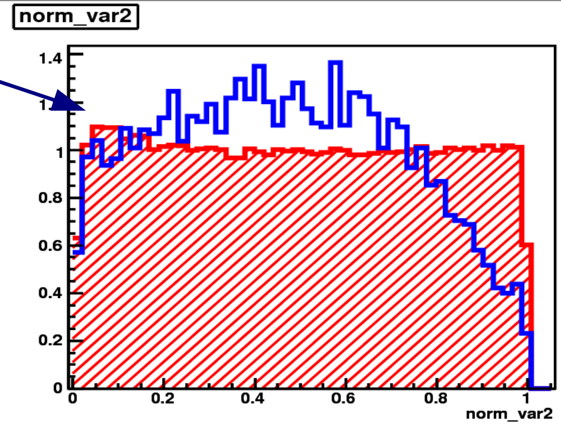


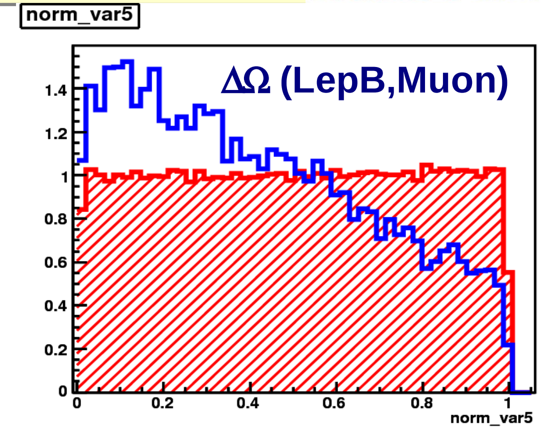


Normalized variables

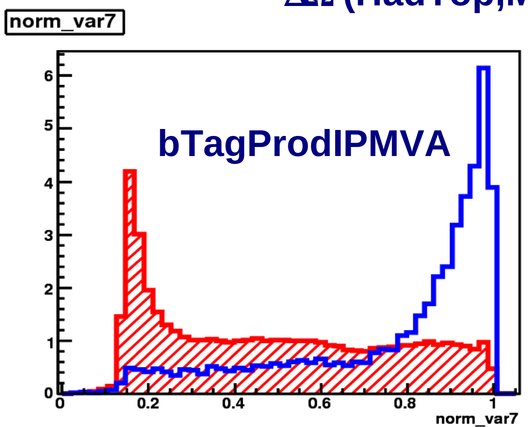
variables are normalized to 1 on y-axis
x-axis: between 0 and 1

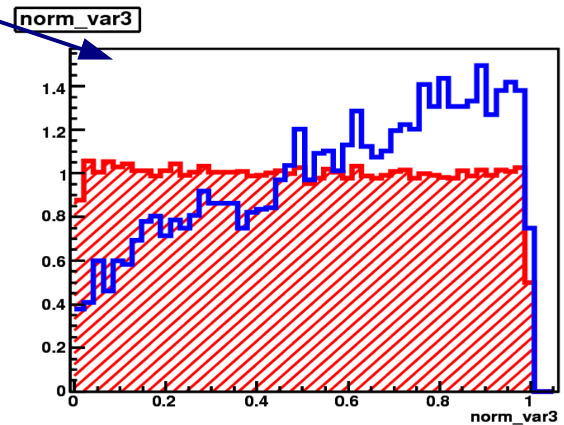
ΔΩ (HadTop,LepB)

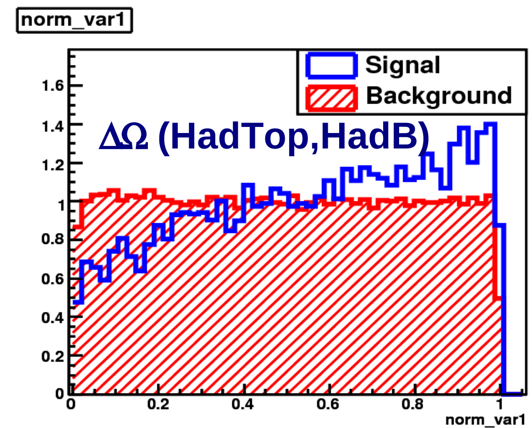




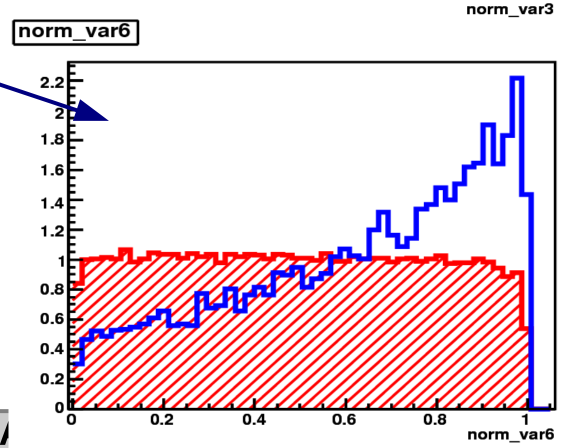
ΔΩ (HadTop,Muon)

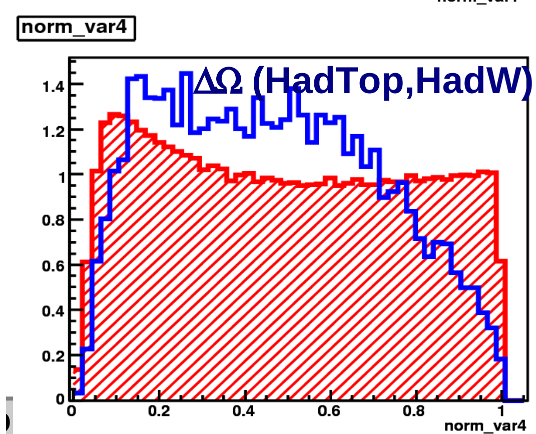






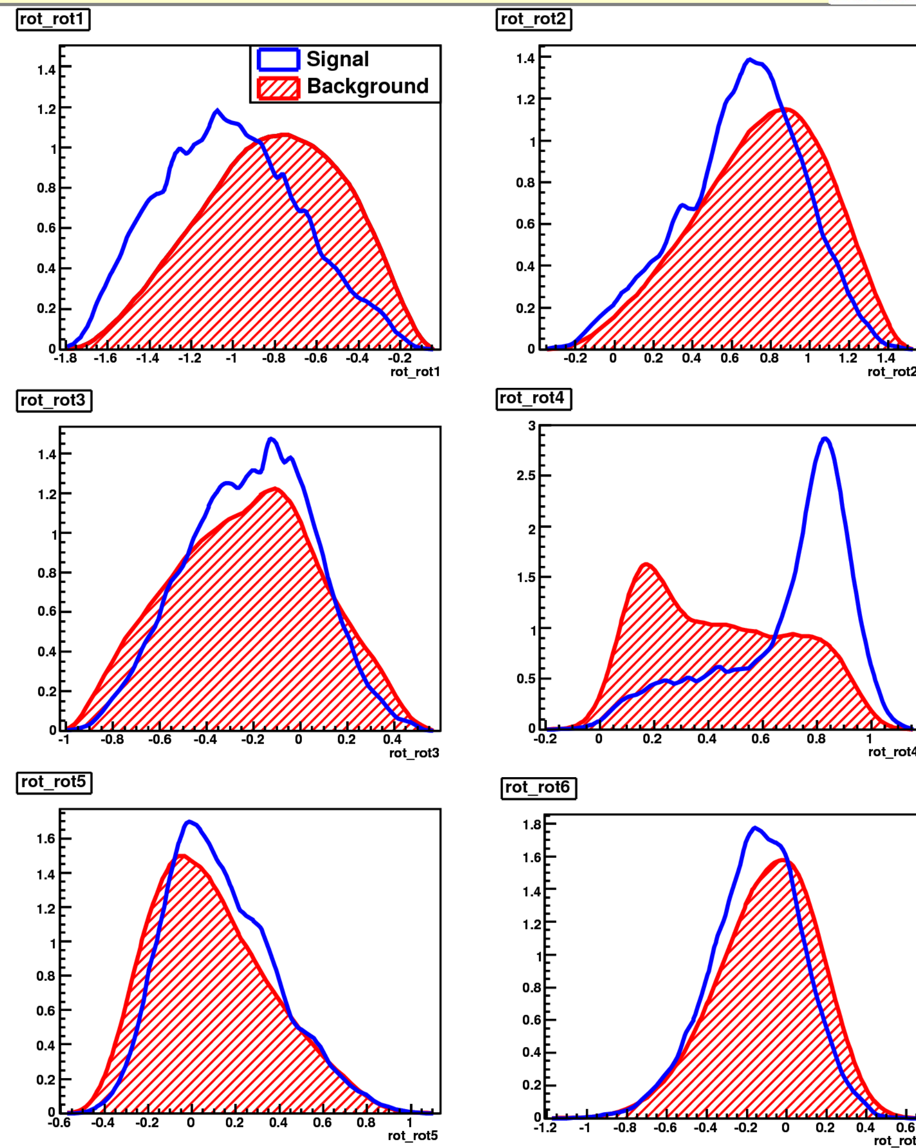
relativePtHadTop



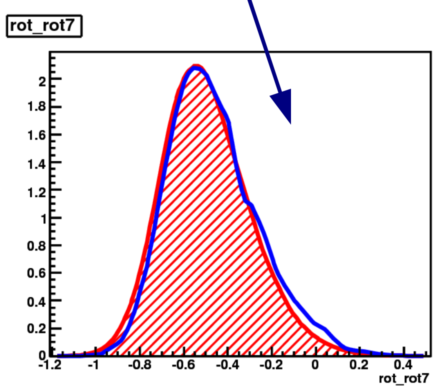


Variables after rotation

7 decorrelated variables:
no clear correspondence
between the 7 initial variables

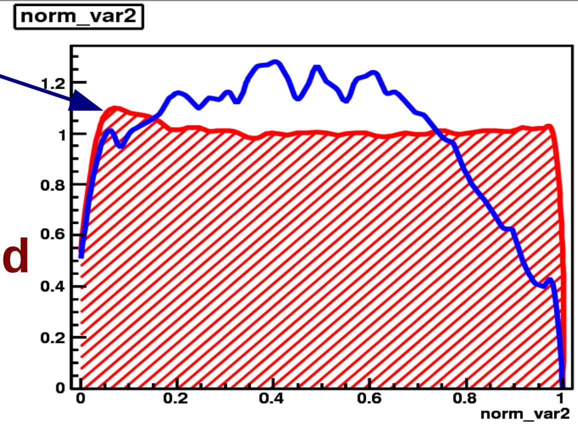


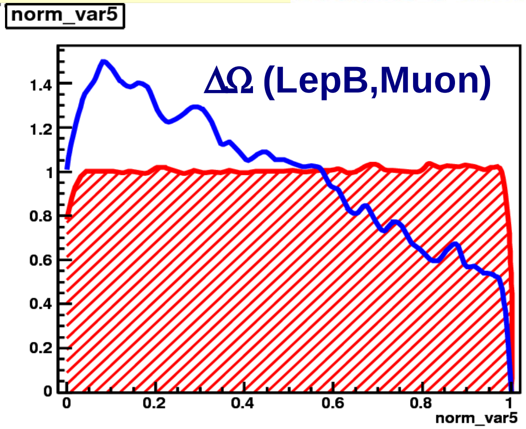
no extra information:
can be dropped (not done)



Variables as input for LR or NN

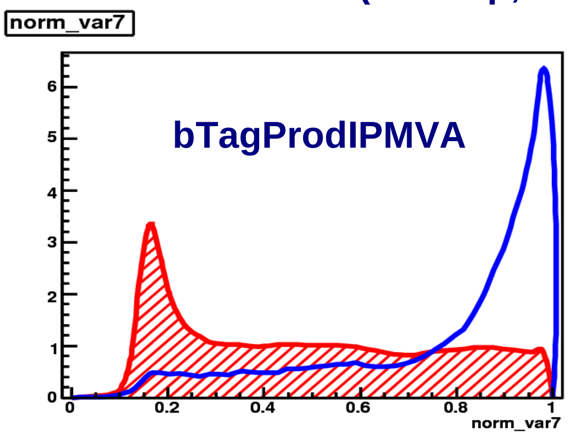
$\Delta\Omega$ (HadTop,LepB)

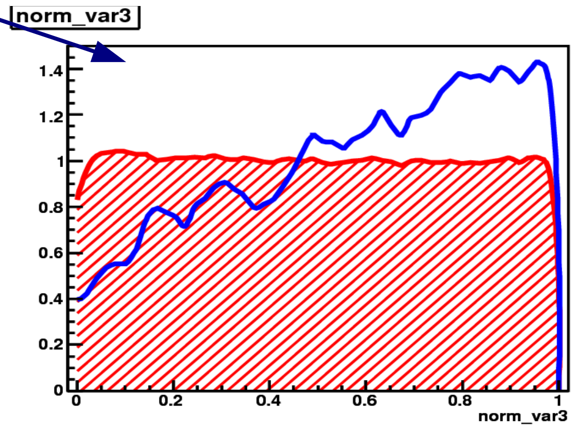


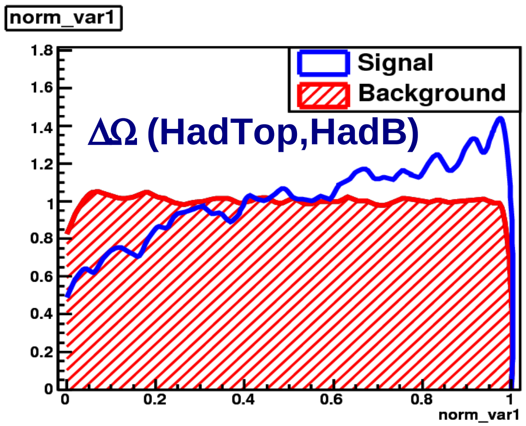


The (rotated) distributions are smoothened and combined by a LR or NN

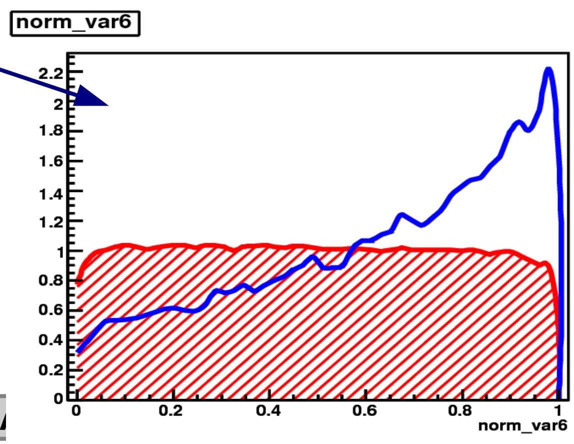
$\Delta\Omega$ (HadTop,Muon)

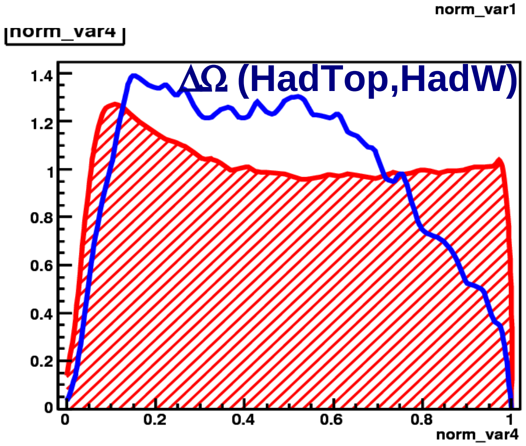






relativePtHadTop





Overlap best jet combination

- Among the selected events using a loose event selection the subsample of events which were selected after a tight event selection is found.
- The discriminator for each jet combination and each event in the subsample has been calculated with 2 different mva-files according to the loose or tight event selection.
- We want to know in how many of the events the best jet combination (i.e. the jet combination with the highest discriminator) is the same after using a different input mva-file (i.e. using a different event selection)

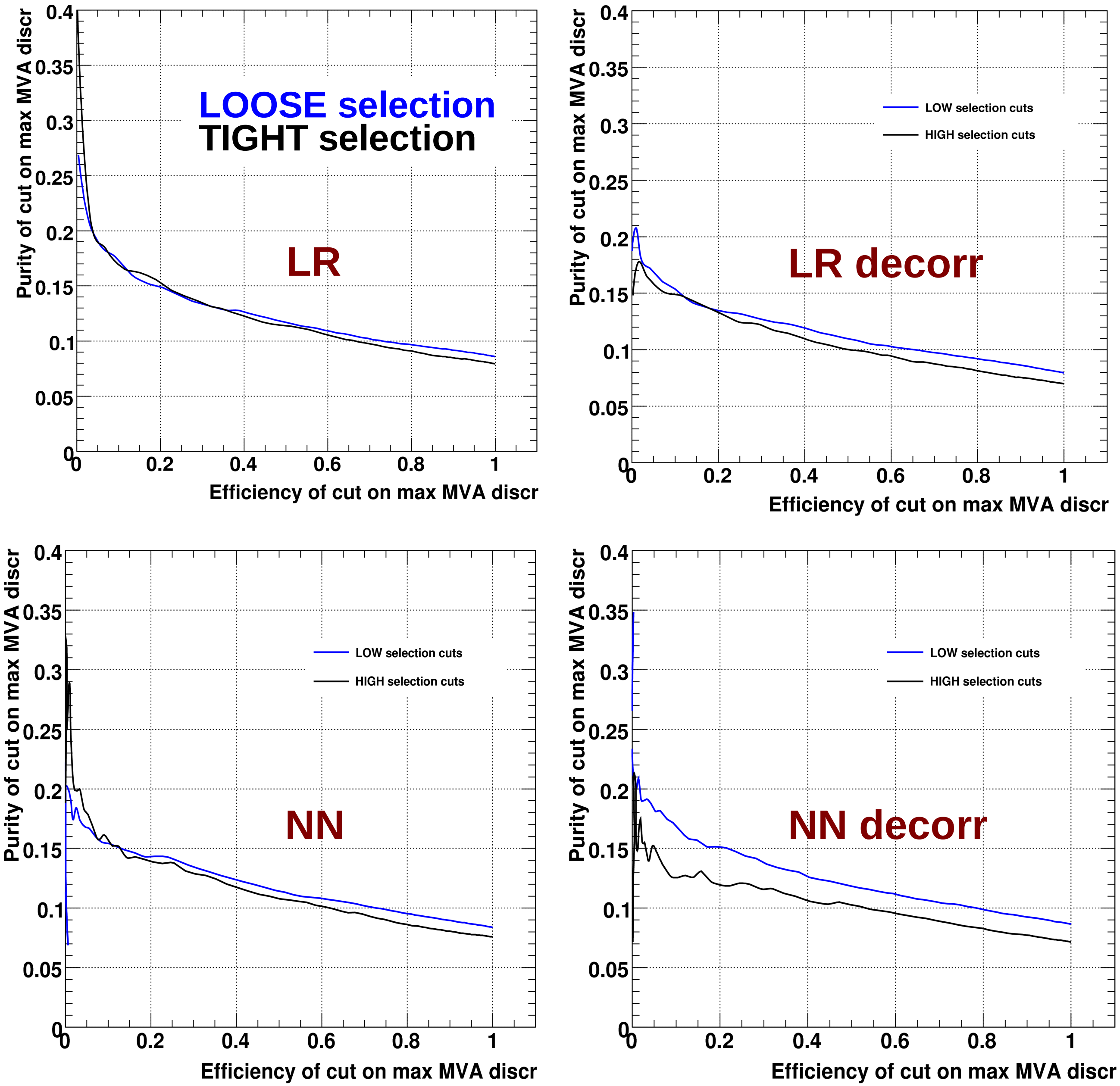
	LR	**LR decorr**	**NN**	**NN decorr**
no cut	87.2	78.2	84.3	81.5
0.2	87.3	78.3	84.3	81.6
0.4	87.6	79.3	85.2	81.0
0.6	86.9	79.8	83.8	79.4
0.7	85.7	78.6	83.8	80.4
0.8	85.7	75.9	81.4	78.8
0.9	84.5	55.8	70.8	71.3

numbers in %

- **The LR method has the highest percentage of events between the sample with high and low selection cuts in which the same jet combination is chosen.**
- **Check the performance of the different MVA tools:**

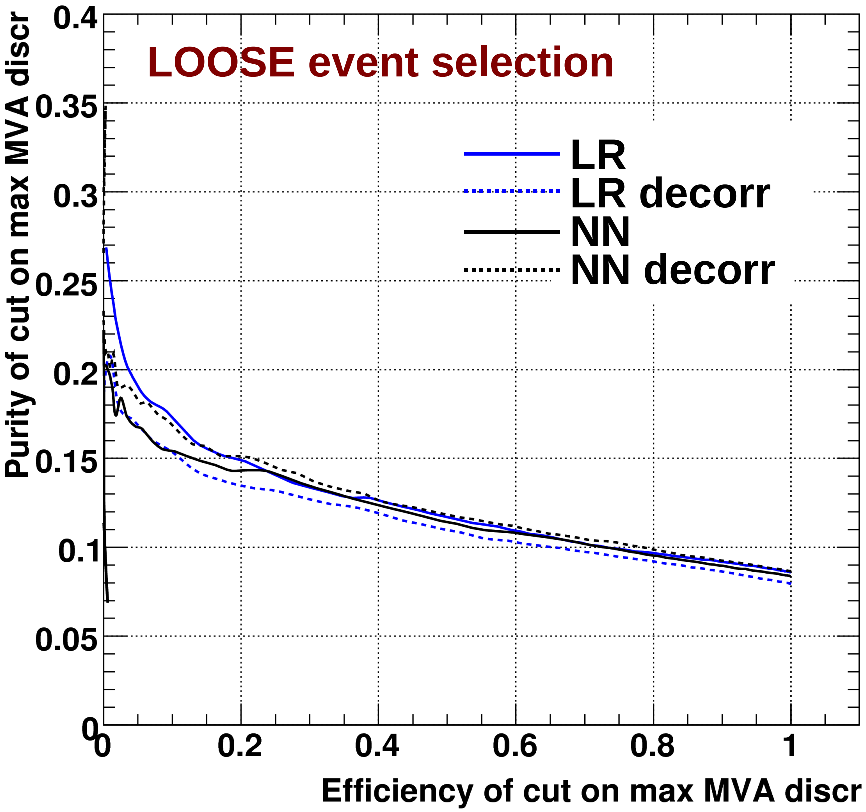
  purity of the cut on the MVA discriminant versus the efficiency of this cut

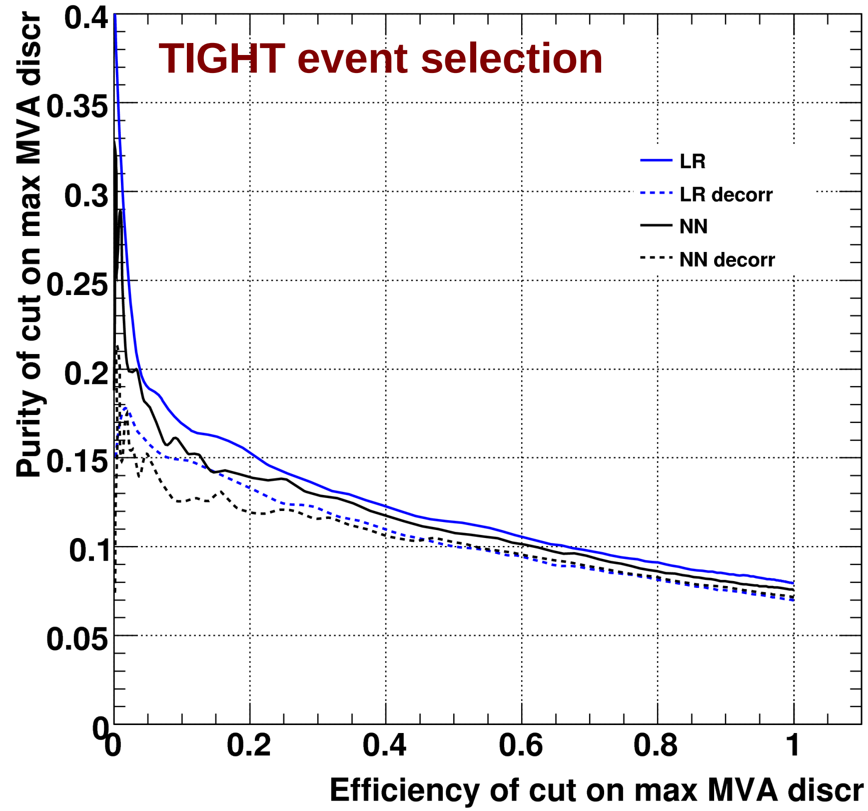
Performance w.r.t event selection



Within 1-3% (NN decorr) the performance of a MVA method after the loose or tight event selection is the same

Comparison of MVA methods





- **All the methods show a similar performance within 1-2% at high efficiencies and 5% at low efficiencies**
- **We see that the difference in performance is a bit bigger for the sample where the tight event selection has been applied: here the LR is favored**
- **Another reason to chose for the LR is that it has the biggest overlap (slide 14)!**
- **Previous slide: LR has the least difference in performance w.r.t event selection**

Conclusions & Outlook

- **7 variables with good discriminating power were selected to be used as input for the MVA tools**
- **All MVA methods yield a similar performance in purity versus efficiency plots**
- **The methods give a similar performance after applying loose or tight event selection cuts**
- **LR highest overlap of choosing same jet combination (87%), hence I chose this method for the continuation of my analysis**

Given the previous bullets we can answer the question:

Is it possible to study the event selection later?

- **It's safe to apply the loose selection cuts and refine the cuts after an event selection study has been done.**
- **Because of this conclusion I started the production of the ROOT-files needed for my analysis. This production is in the meantime almost done for signal and background samples.**

Next steps:

- **Study event selection**
- **Perform estimation of the jet energy correction factors with and without differentiation as a function of p_T and η**
- **... many other things ...**

Very fresh bonus material