# Search for Diboson Production in the lepton + MET + bb channel

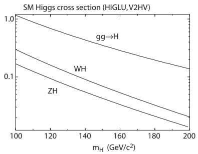
**M. Trovato**, G. Latino and C. Vernieri for the "3 jets" region

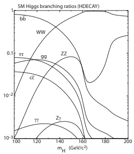
April 19, 2011

# Outline

# Motivations

- Heavy vector boson production measurements used to test the electroweak sector of the Standard Model (SM)
  $\Rightarrow$ SM allows for interactions between $W^{\pm}$, $Z^0$ ("triple gauge couplings")
- Preferred channel for Higgs search at Tevatron: $q\bar{q}' \to WH \to l\nu b\bar{b}, l \neq \tau$ (in $gg \to H \to b\bar{b}$ much more background)




- $WZ$ as preliminary step for $WH$ in $l\nu b\bar{b}$ final state
  - same final state signature
    $\Rightarrow$ $WZ$ data sample usable to optimize measurement techniques
  - easier: $\frac{\sigma_{WH} \cdot BR(W \to l\nu) \cdot BR(H \to b\bar{b})}{\sigma_{WZ} \cdot BR(W \to l\nu) \cdot BR(Z \to b\bar{b})} \simeq \frac{1}{5}$, for $m_H = 120$ GeV
  - $WZ$ is a background for the Higgs search
- $ZZ \to l^+ l^- b\bar{b}$ included to increase the statistics of the sample

## Why pretag?

The first mandatory step of this analysis it validate the MC Vs Data agreement where no *b*-tagging is applied

- not a large statistics to do that in the *b*-tagged sample!
- *b*-tagging adds additional complications (heavy flavour composition, etc...)

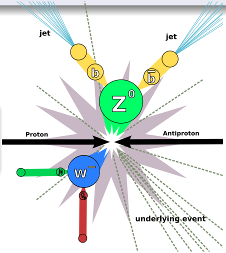
# Selection: $WZ \to l\nu q\bar{q}$ ($ZZ \to llq\bar{q}$ as signal)


## isolated high-$p_T$ triggered lepton (lep)

- Central electron (ele): $E_T > 20$ GeV, $|\eta| < 1.1$
- Central muon (muo): $p_T > 20$ GeV, $|\eta| < 1.0$

## $\nu$ candidate

- $\not{E}_T > 20$ GeV




## Vetoes: QCD rejection

- $M_T^W > 10(30)$ GeV for muo (ele)
- Cut on the probaby of fake Vs real $\not{E}_T$

## Two regions

- two jets with $E_T > 25, 25$ GeV, $|\eta| < 2$

Region 1: no additional jet with $E_T > 15$ GeV, $|\eta| < 2$
Region 2: one additional jet with $E_T > 15$ GeV, $|\eta| < 2$ (last part of the talk)

# SM processes composing the selected data sample

Following SM processes contributes to the selected sample:

- **Electroweak and top**: $WW$, $WZ$, $ZZ$, $Z$+jets, $t\bar{t}$, single-*top*
  - ⇒ rate normalized to the cross-section
  - ⇒ shapes from ALPGEN+Pythia, Pythia
- **QCD**: multi-jet production with a jet faking the lepton and fake $\not\!\!E_T$
  - ⇒ rate normalization and shapes from data
- **W($\rightarrow$ l$\nu$)+jets**
  - ⇒ rate normalization from data
  - ⇒ shapes from ALPGEN+Pythia

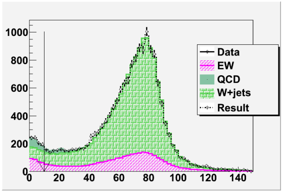
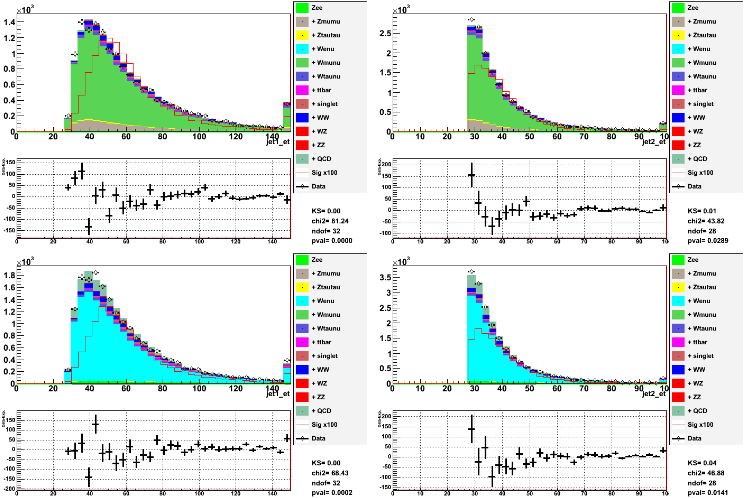


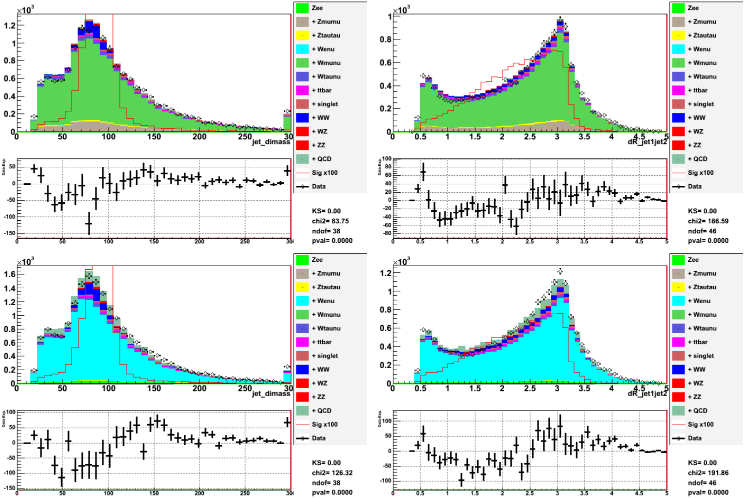

Figure: *Fit on $M_T^W$ for the QCD and W+jets rate in data*

# MC Vs data (pretag sample)

# first/second leading jet $E_T$ - Ele (up), Muo (bottom)



# Dijet Mass - Ele (up), Muo (bottom)



## Region 2: 3 jets are found...What do you do?

# Handling of the extra jet(s) at CDF: state of the art

**Definition**

jets are ordered in decreasing $E_T$: $j1, j2, j3$

CDF analyses by default either:

1. discard events
2. reconstruct $Z$-mass as $M_{j_1 j_2}$

→ but ∼ 25-40% of $WZ$ events are rejected or

→ but $M_{j_1 j_2}$ shape is degraded

# Handling of the extra jet(s) at CDF: state of the art

**Definition**

jets are ordered in decreasing $E_T$: $j1, j2, j3$

CDF analyses by default either:

1. discard events → but $\sim$ 25-40% of $WZ$ events are rejected or
2. reconstruct $Z$-mass as $M_{j_1 j_2}$ → but $M_{j_1 j_2}$ shape is degraded

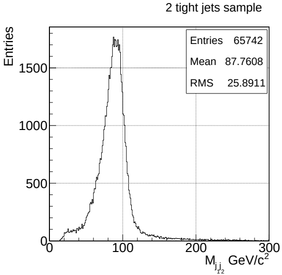
# Handling of the extra jet(s) at CDF: state of the art

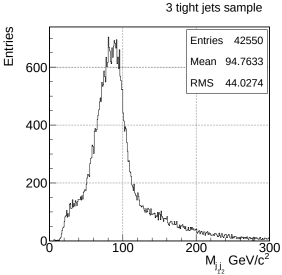
**Definition**

jets are ordered in decreasing $E_T$: $j1, j2, j3$

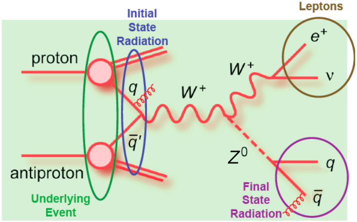
CDF analyses by default either:

1. discard events → but $\sim$ 25-40% of $WZ$ events are rejected or
2. reconstruct $Z$-mass as $M_{j_1 j_2}$ → but $M_{j_1 j_2}$ shape is degraded



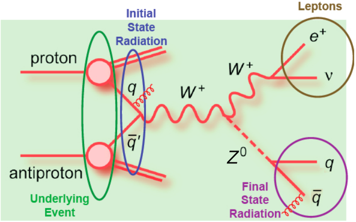
# Jet origin: where do the extra-jets comes from?




1. Radiation from interaction partons (ISR)
2. Radiation from Z-decay products (FSR)
3. lepton mis-identified as a jet (to be investigated)
4. Extra-activity produced by spectator partons/protons (probably negligible)

⇒ we'll be focusing on the first two effects

# Jet origin: where do the extra-jets comes from?

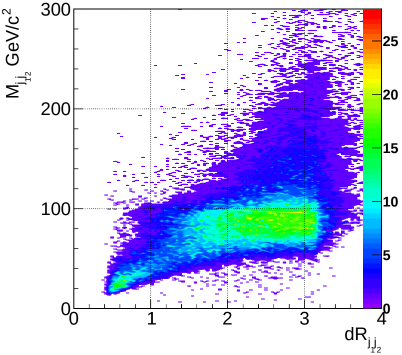



1. Radiation from interaction partons (ISR)
2. Radiation from Z-decay products (FSR)
3. lepton mis-identified as a jet (to be investigated)
4. Extra-activity produced by spectator partons/protons (probably negligible)

$\Rightarrow$ we'll be focusing on the first two effects

# History: $M_{j_1 j_2}$ Vs $dR_{j_1 j_2}$

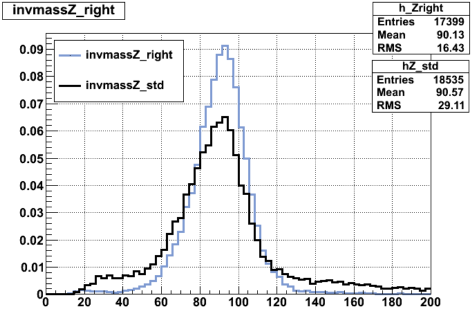
*ONCE UPON A TIME...*




- bulk of events at $M_{j_1 j_2} \sim M_Z^{pdg}$
- but large
  - low mass tails
  - high mass tails

# The best we can do

- If we knew the RIGHT jet combination from Z: $M_Z \rightarrow M_{Z_{right}}$ (blue) $\rightarrow$ RIGHT jet combination is
  1. correct jet pair if ISR
  2. jet triplet if FSR
- but we don't $\rightarrow M_Z \rightarrow M_{std} \equiv M_{j1j2}$ (black)



# How to understand the RIGHT jet combination (RJC)?

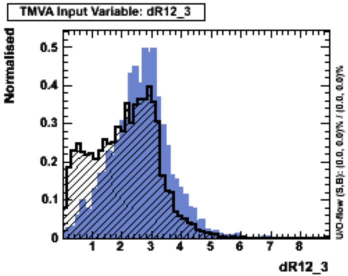
## Procedure

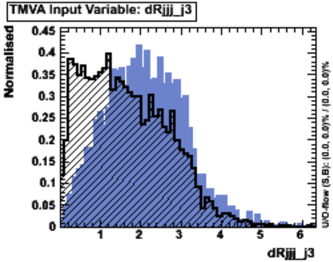
- Jets are matched in direction to quarks from Z decay
- Reject events ($\sim$ 40%) with # matches $\neq$ 2
- Investigate at generator level origin of the not-matched jet: ISR or FSR
  1. RJC = j1j2 in 40%
  2. RJC = j1j3 in 18%
  3. RJC = j2j3 in 9%
  4. RJC = j1j2j3 in 19%
- Isolate RJC = j1j2 ($\sim$ 40%). Coming next
- "bottle-neck" of the procedure...we'll play by hear on how to handle the rest 60%.

  $\rightarrow$ Caterina Vernieri will return on this during the Open discussion

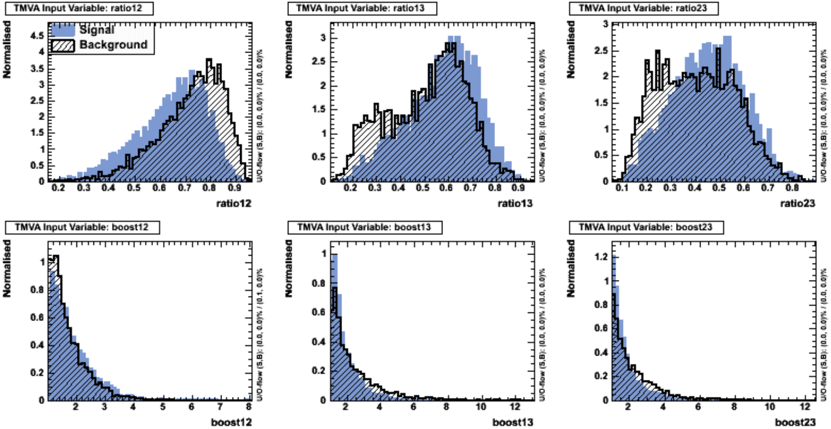
# Isolating RJC = j1j2

- A neural-network (NN) is trained on WZ events to distinguish RJC = j1j2 (signal) from the rest (background)
  $\rightarrow$ background weighted to have the same dijet mass shape
- Several kinematical variables (full list and distributions in backup)
- $dR_{(j1j2)j3}$,$dR_{(j1j2j3)j3}$



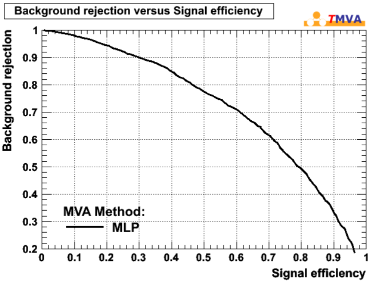



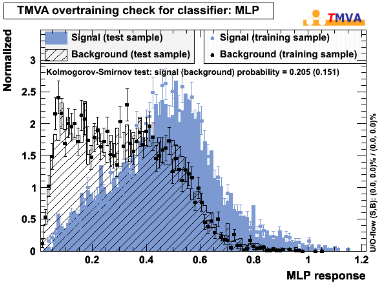
- $M_{j1j2}/M_{j1j2j3}$, $M_{j1j3}/M_{j1j2j3}$, $M_{j2j3}/M_{j1j2j3}$
- $(E_{j1}+E_{j2})/m_{j1j2}$, $(E_{j1}+E_{j3})/m_{j1j3}$, $(E_{j2}+E_{j3})/m_{j2j3}$



# NN output

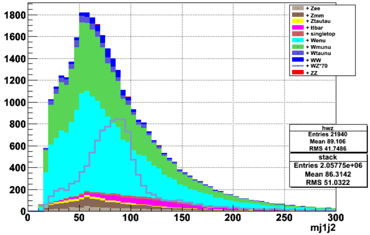
- Background rejection Vs signal efficiency
- Signal Vs Background separation

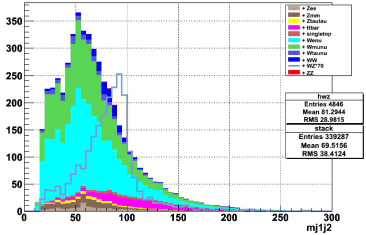


# Putting all together

- We cut on NN>0.5
  $\rightarrow \sim$ 63% RJC=12
  $\rightarrow \sim$ 14% RJC=13, 9% RJC=23, 3% RJC=123 after the cut




- Signal peak is appreciably narrower after the NN cut
  $\rightarrow$ Resolution improved by $\sim$25%
  $\rightarrow$ Signal over background improved by a factor of 2
- more work to do to improve NN

# Conclusions and future plans

## Region1: $= 2$ jets

- The agreement data Vs MC in the pretag shapes looks not so bad
  1. however, still some mis-modeling
  2. investigating other generators (SHERPA) for W+jets
  3. need to do more validation on the data-driven QCD models
- how will the shapes look like in the tag sample?

## Region2: $= 3$ jets

- The technique seems promising

  $\rightarrow$ if successfull, it will drastically increase the signal acceptance of several important analyses (Diboson, Higgs, etc.)

## Plans

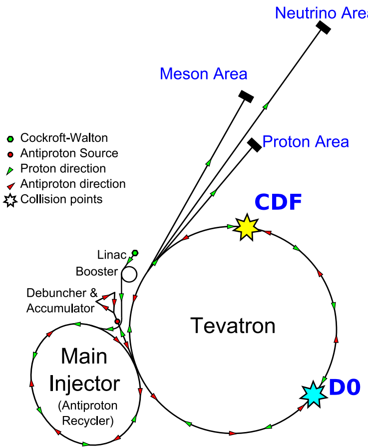
- 3 jet technique to be applied to the $b$-tag region

  $\rightarrow$ however, the behavior in $b$-tag region may be different
  1. different background composition (mostly $t\bar{t}$)
  2. $b$-quarks fragment in a different way compared to the light quarks

# Part I

# Backups

# Tevatron Collider (Run II)

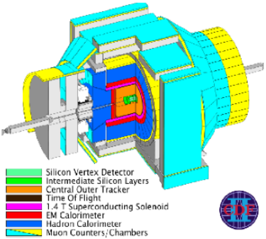



- Run II: current data taking period (2001 → ...)
- $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV
- 2 high-luminosity regions, 2 experiments: CDF, DØ
- Maximum instantaneous luminosity $\sim 3 \cdot 10^{32}$ cm$^{-2}$s$^{-1}$
- Integrated delivered luminosity $\sim 10$ fb$^{-1}$ (December 2010)

# CDFII (Collider Detector at Fermilab in Run II)

- Tracking System composed of:
  - silicon microstrip trackers for precise vertex identification
  - open-cell drift chamber

  in a magnetic field (1.4 T)
- Sampling calorimeters split into projective towers
- Planar drift chambers backed by scintillation counters for muon identification




**Coordinates and definitions**

- $r$, $\phi$, $\eta \equiv -\ln(\theta/2)$
- $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$
- $E_T = E \cdot \sin\theta$

# Weight

We normalized at the same distribution MJ1J2 in both subsamples.

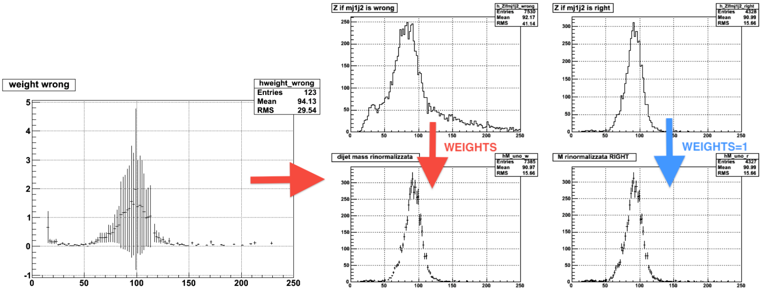



Figure: Right: Mj1j2 distribution in both subsamples before and after normalization. Left: Weights used for normalizing.

# Jets at CDF (JETCLU)

- Cluster of calorimeter towers

$\rightsquigarrow$ a list of seed towers ($E_T > 1\ GeV$) is sorted in decreasing $E_T$

$\rightsquigarrow$ the seed towers contained in a in a 49-towers square centered around the hardest seed are grouped into a precluster (PC)

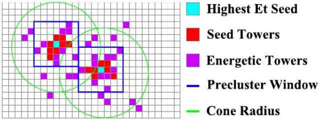
$\rightsquigarrow$ $\theta^{PC}, \phi^{PC}$ are calculated according to "Original Snowmass Scheme" (REFERENZA)

$\rightsquigarrow$ a cone of radius R(=0.4,0.7) is drawn about $\theta^{PC}, \phi^{PC}$ and towers ($E_T > .1\ GeV$) within the cone used to calculate $\theta^{PC}, \phi^{PC}, E_T^{PC}$

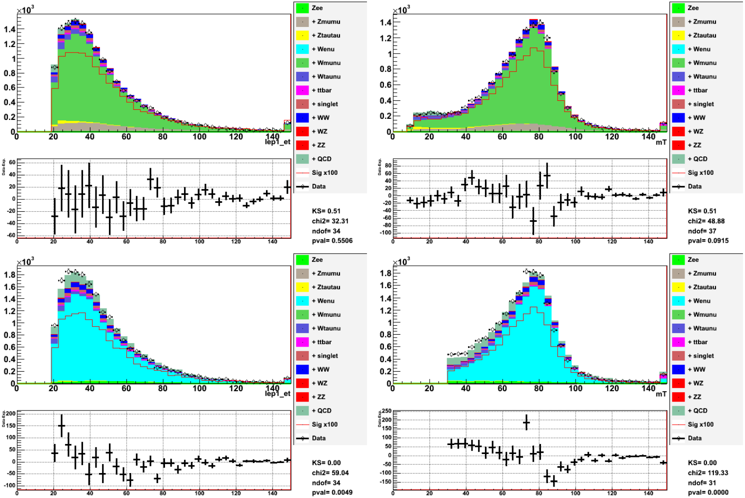
$\rightsquigarrow$ loop over the step above until a stable cone is found

$\rightsquigarrow$ apply the split/merge procedure to have the final jet list

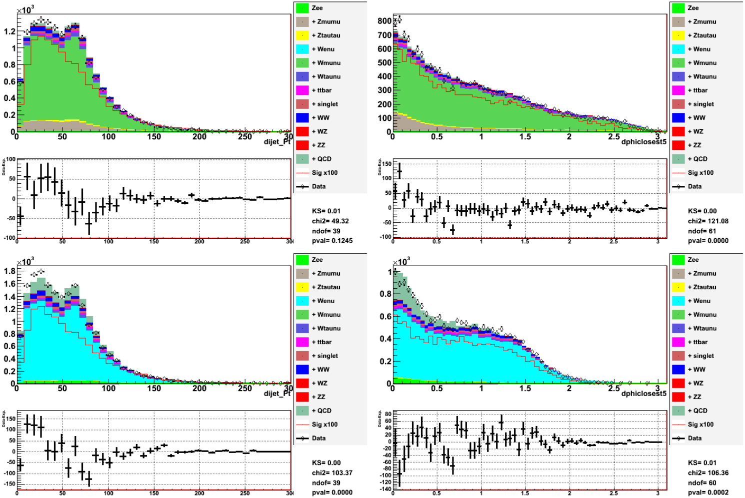
$\rightarrow$ $E_T^{PC}$ is corrected for known instrumental and physics effects



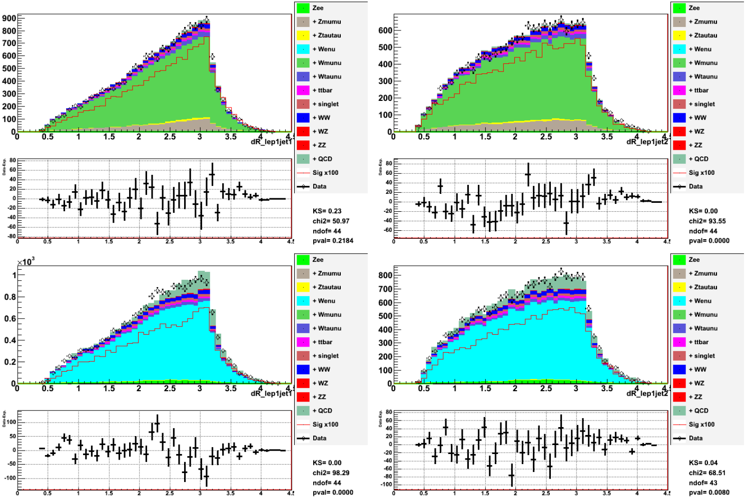
# lep $E_T$, $M_T^W$ - Ele (up), Muo (bottom)



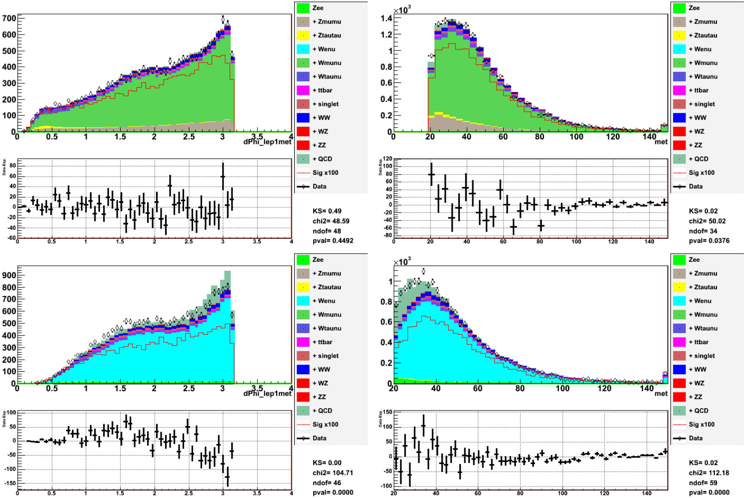
# dijet Pt, dphiclosest5 - Ele (up), Muo (bottom)



# $dR_{lep1jet}$ - Ele (up), Muo (bottom)



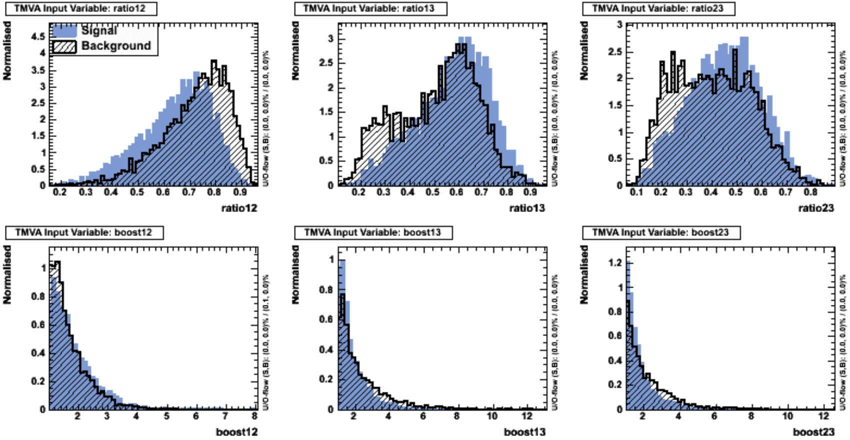
# $dphi_{lep1met}$, met - Ele (up), Muo (bottom)

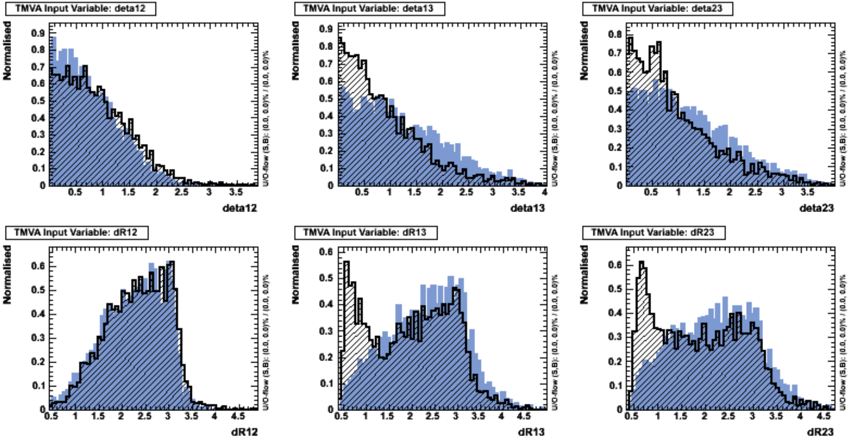


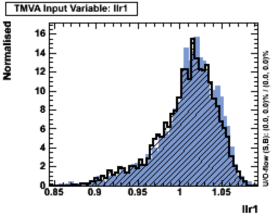
# Isolating RJC = j1j2

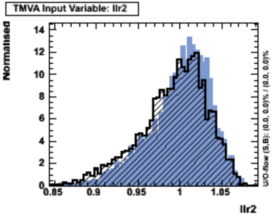
- A neural-network is trained to distinguish RJC = j1j2 from the rest
- following variables are used (Caterina puoi controllare?)
  1. $m_{jj'}/m_{j1j2j3}$ *
  2. $\gamma_{jj'} \equiv (E_j + E_{j'})/m_{jj'}$ *
  3. $dR_{j\ell}$ *
  4. $dR_{jj',j''}$ *
  5. $\mathrm{d}\eta_{jj'}$ *
  6. llr : Q prob/G prob *
  7. 'W-boost' $\equiv \mathbf{P_T}_{\ell} + \not{E}_{\mathbf{T}}$
  8. 'pt-imbalance' $\equiv \mathbf{P_T}_{J1} + \mathbf{P_T}_{J2}$ - $\mathbf{P_T}_{\ell} + \not{E}_{\mathbf{T}}$

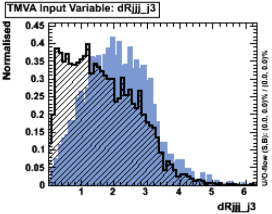
*: 3 combinations

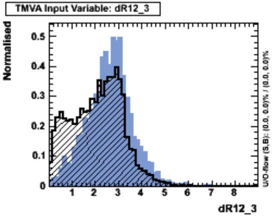

TMVA Input Variable: ratio12
Signal
Background
Normalised
ratio12
TMVA Input Variable: ratio13
Normalised
ratio13
TMVA Input Variable: ratio23
Normalised
ratio23
TMVA Input Variable: boost12
Normalised
boost12
TMVA Input Variable: boost13
Normalised
boost13
TMVA Input Variable: boost23
Normalised
boost23


TMVA Input Variable: deta12
Normalised
deta12
TMVA Input Variable: deta13
Normalised
deta13
TMVA Input Variable: deta23
Normalised
deta23
TMVA Input Variable: dR12
Normalised
dR12
TMVA Input Variable: dR13
Normalised
dR13
TMVA Input Variable: dR23
Normalised
dR23


TMVA Input Variable: llr1
Normalised
llr1
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


TMVA Input Variable: llr2
Normalised
llr2
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


TMVA Input Variable: dRjjj_j3
Normalised
dRjjj_j3
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


TMVA Input Variable: dR12_3
Normalised
dR12_3
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%