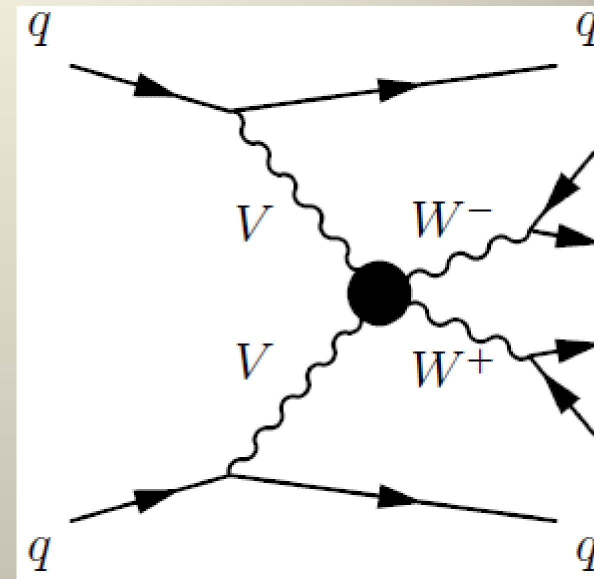
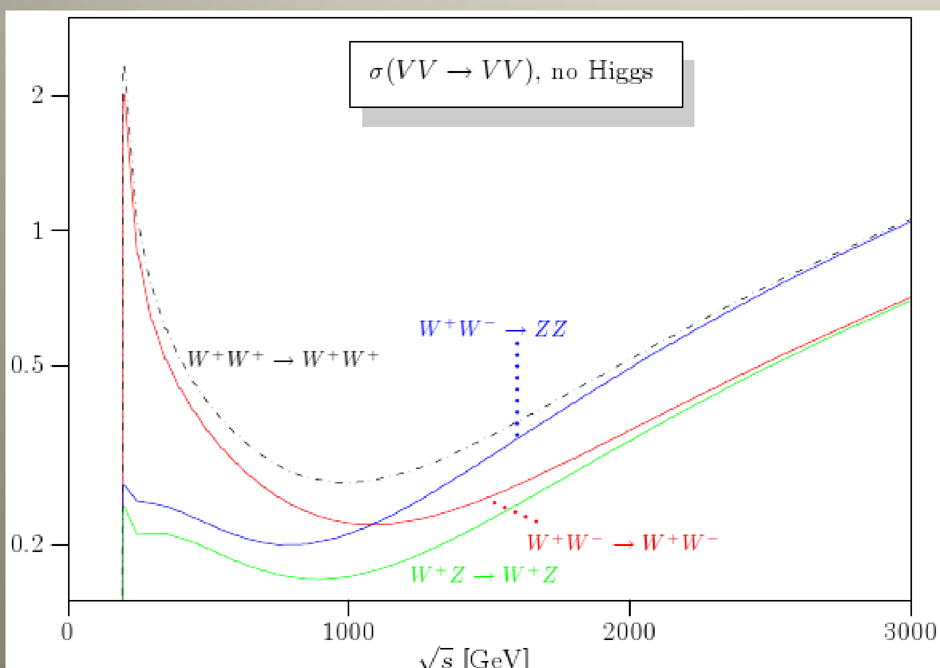
Vector Boson Scattering At High Mass

Hanzhe Liu (USTC)
Advisor:Junjie Zhu (UM)
REU final presentation
2012.8.10

Introduction

- What is VBS
  (vector boson scattering)
- Why VBS is important





Overview

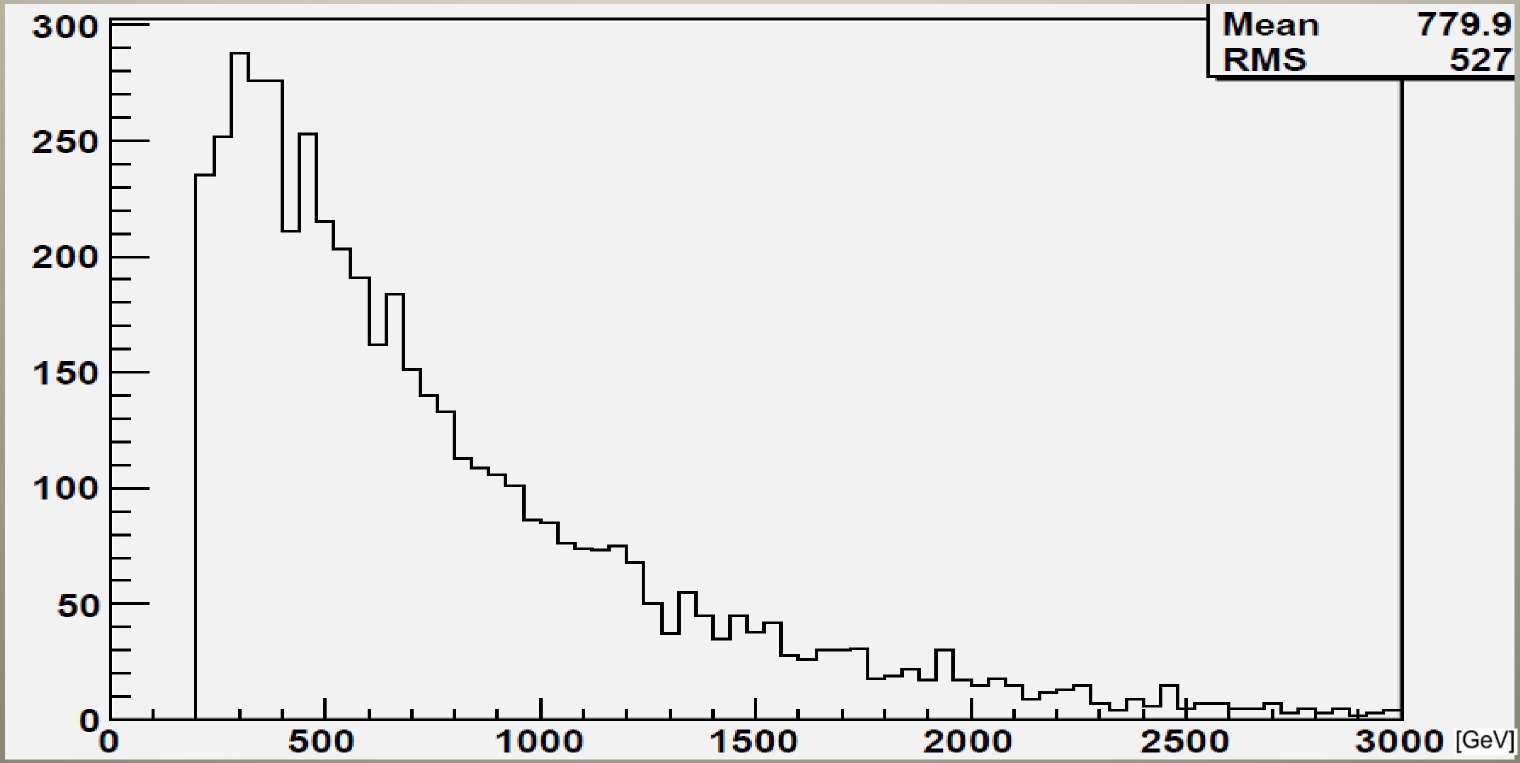
- Monte Carlo Generation of signal: Pythia
- Event kinematics studies
- Using Boosted Decision Tree (BDT) to separate signals from the Standard Model backgrounds

High Mass Resonance Models

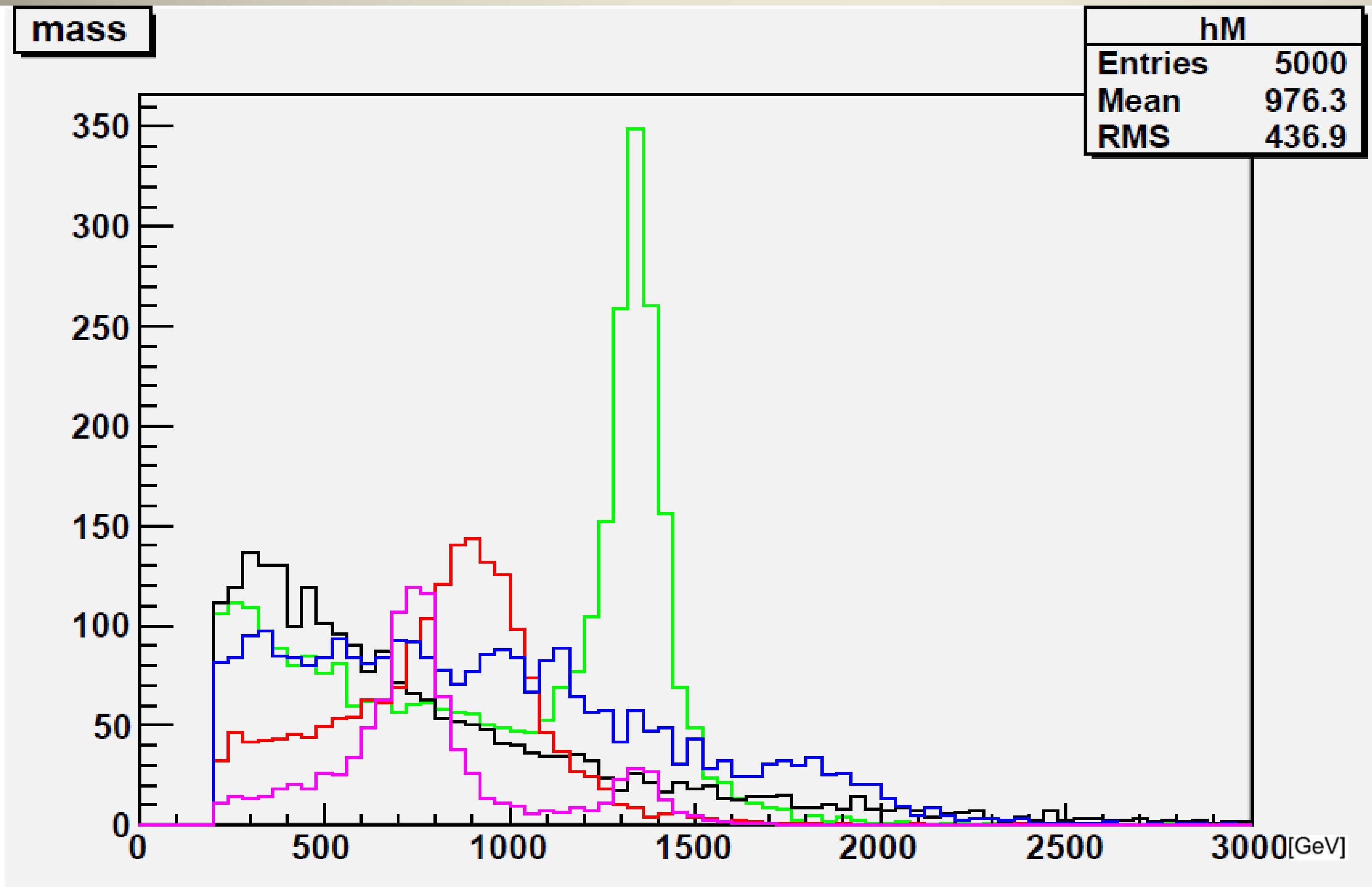
- exchange a SM Higgs boson
- 1 TeV scalar (used for BDT studies)
- 1.4TeV vector
- 1.9TeV vector
- a 800 GeV scalar and a 1.4 TeV vector

M_{WW} distribution with a SM Higgs boson

- Exchange a SM Higgs boson
- minimum WW invariant mass -> 200GeV



Inclusive cross sections



Cross sections (fb):

SM:
2.116 (black)

1TeV Scalar:
2.65 (**red**)

1.4TeV vector:
1.528 (**green**)

1.9TeV vector:
1.746 (**blue**)

800GeV scalar & 1400TeV vector:
5.164 (**pink**)

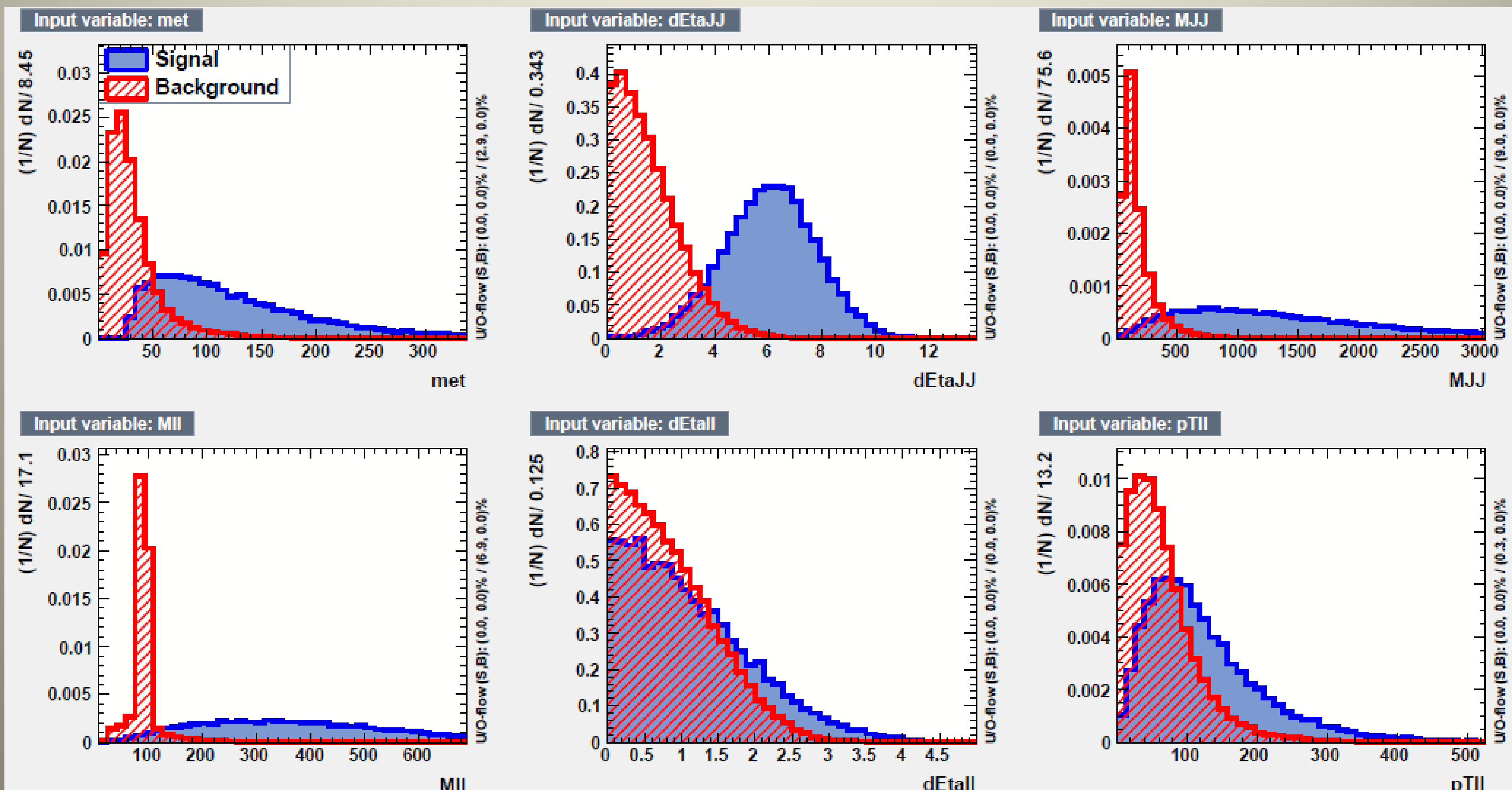
Event Pre-selection

- 2 isolated leptons with $pT>25$ GeV, $|\eta|<2.5$
- 2 forward jets with $pT>30$ GeV, $|\eta|<4.5$

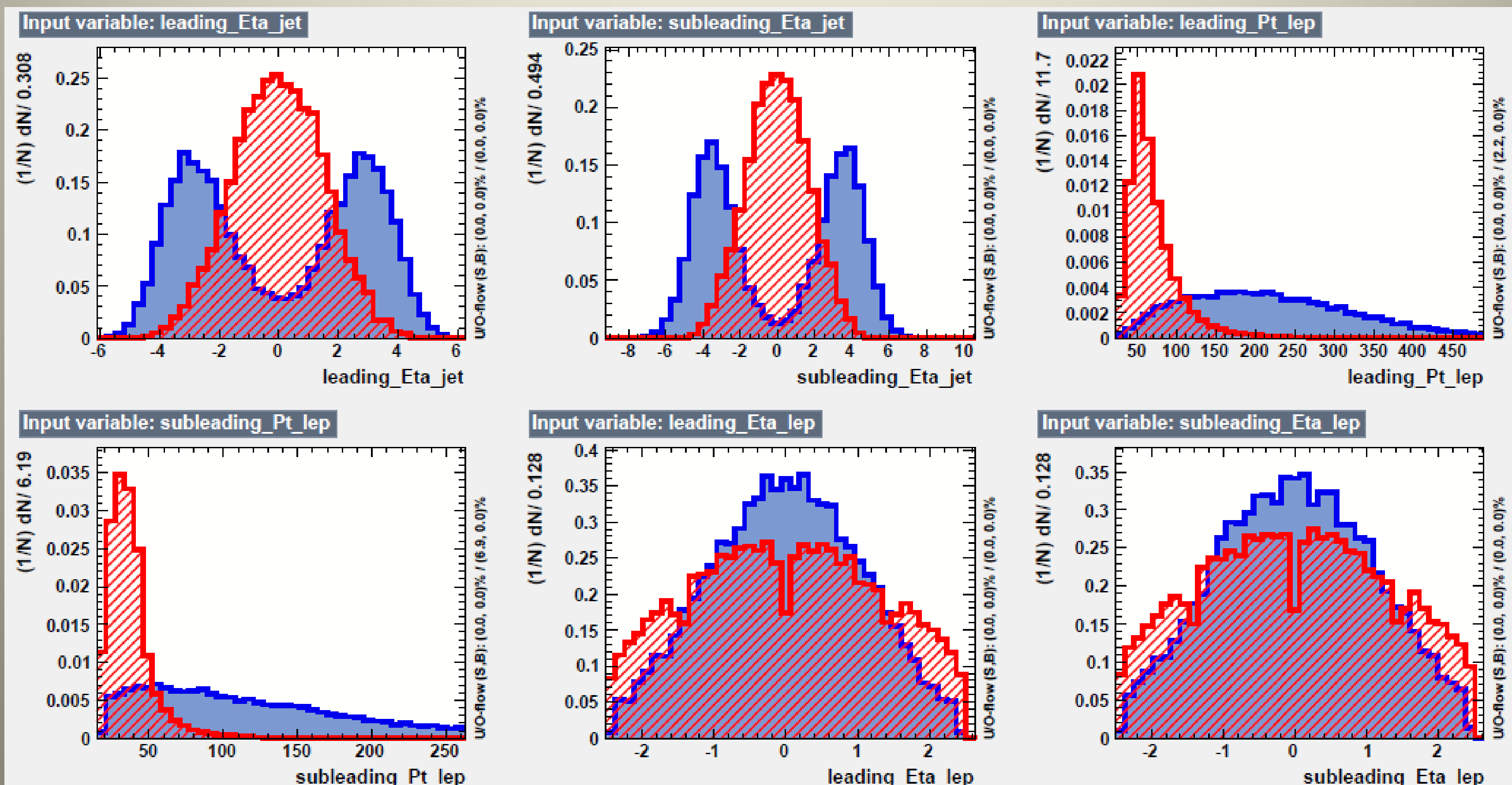
- Backgrounds: WZ, ZZ, Wγ, W+jets, Z+jets, WW(SM), ttbar, singletop

b-tagging

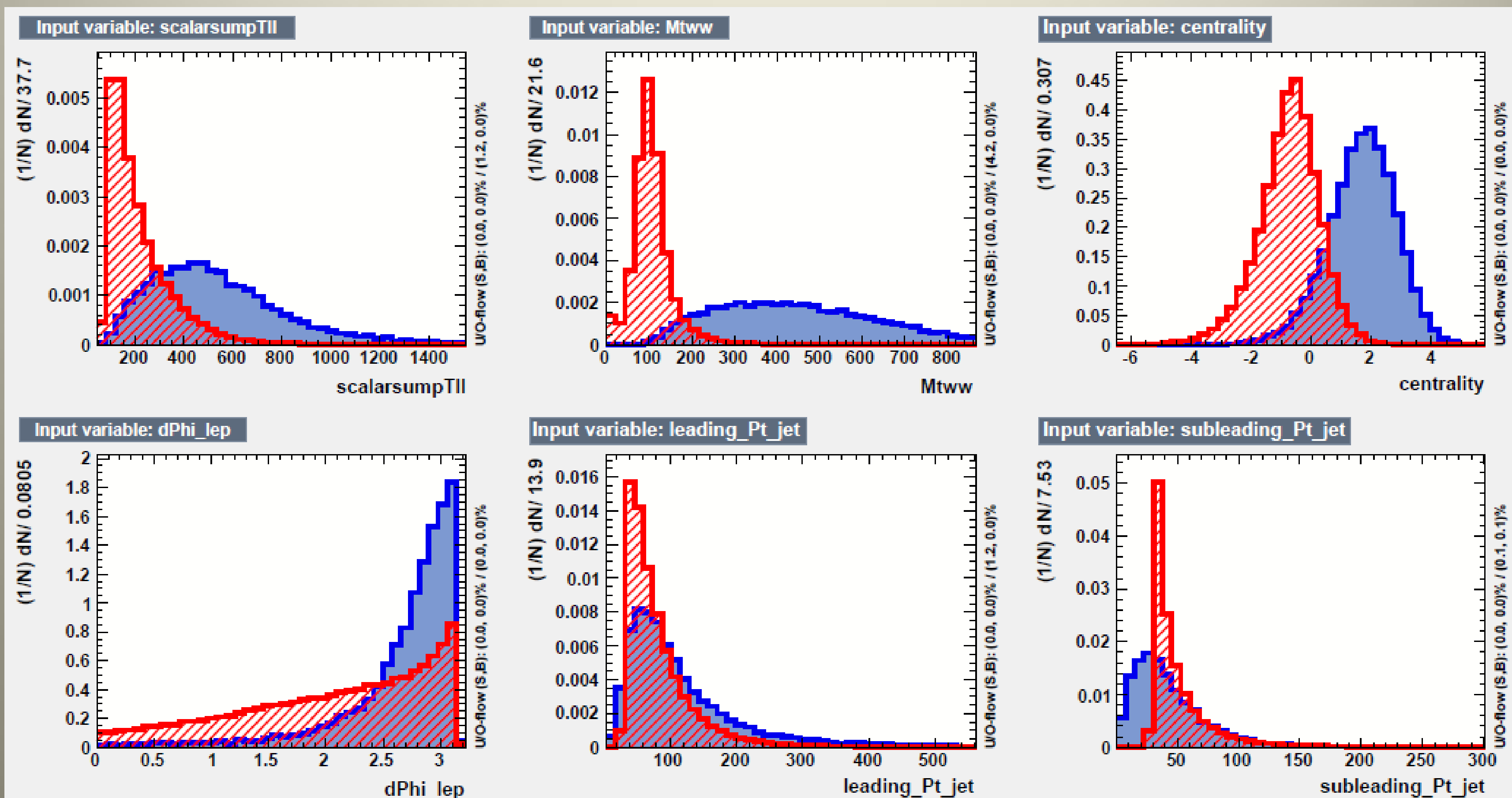
Kinematic Variable



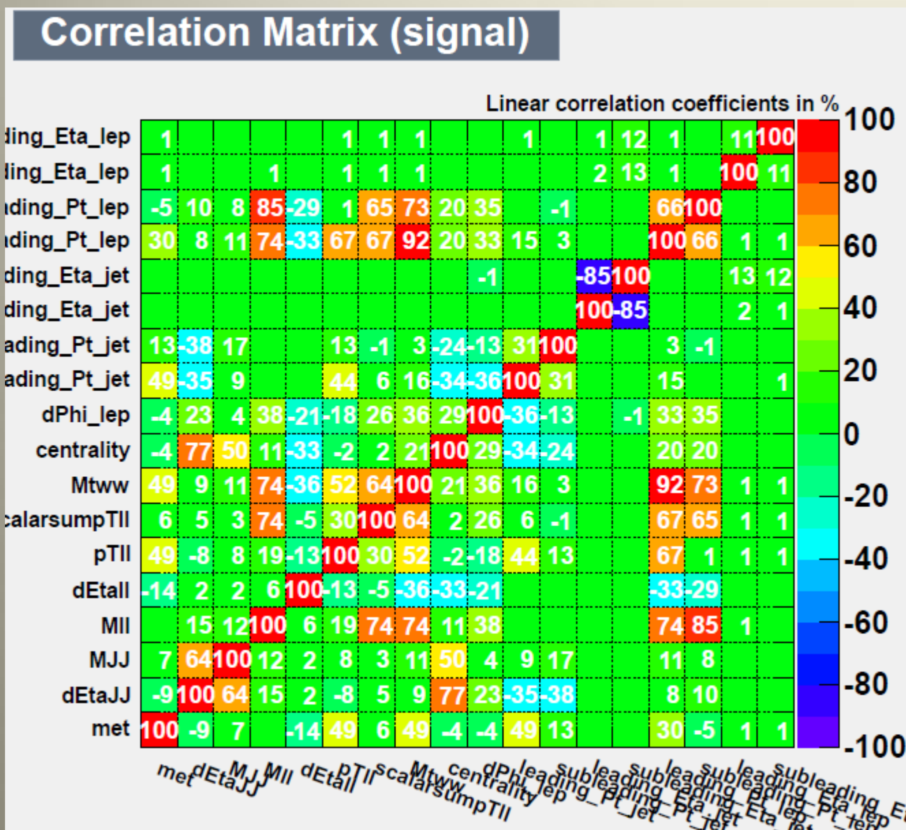
Kinematic Variables 2

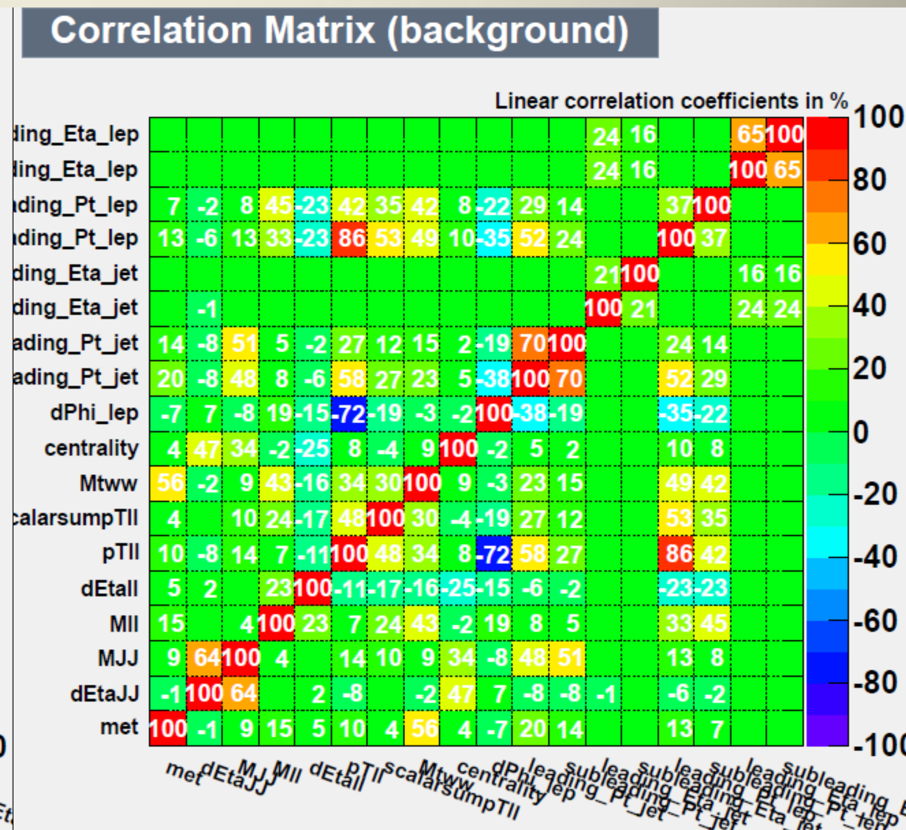


Kinematic Variable 3



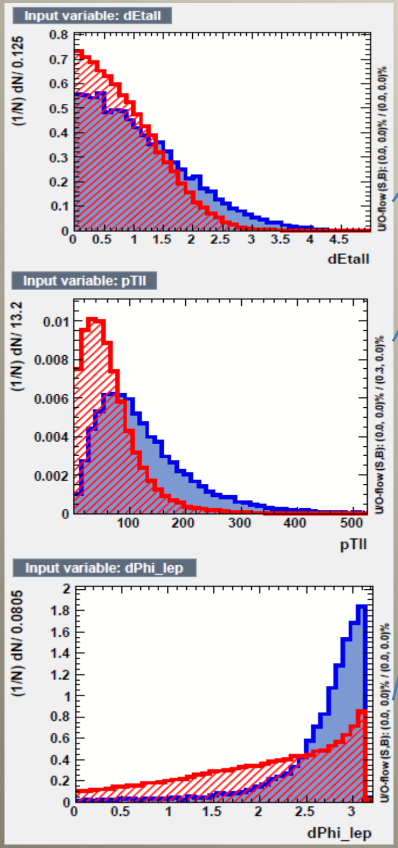
Input correlation matrix





Multivariate method

- In fact, compared to cut based methods, the advantage of BDT is to analyze the correlations of different variables
- Input data



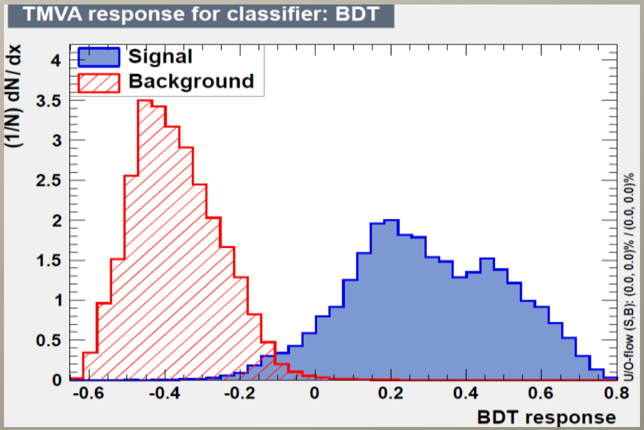
MVA (multivariate)

A large variety of multivariate classification algorithms

Extract a maximum of the available information from data

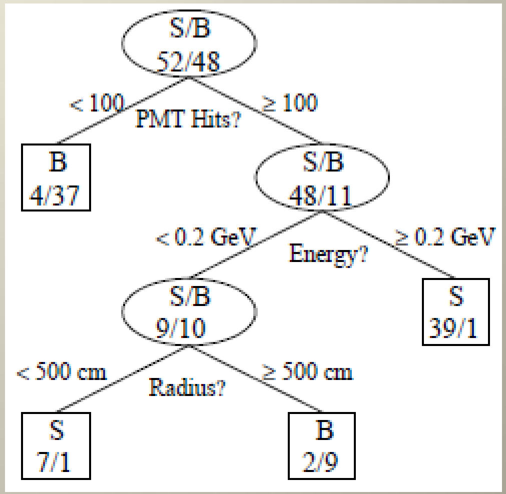
Output: each events has a score

Output distribution looks like this:



Boosted Decision Tree

- A decision tree



- Define the purity in a node:

$$P = \frac{\sum_s W_s}{\sum_s W_s + \sum_b W_b}$$

- Define: $Gini = (\sum_{i=1}^{n} W_i) P(1-P)$
- When a node is split into two nodes, one minimizes:

$$Gini_{left\ child} + Gini_{right\ child}.$$

- The resulting tree is a decision tree
- One decision tree is not enough!

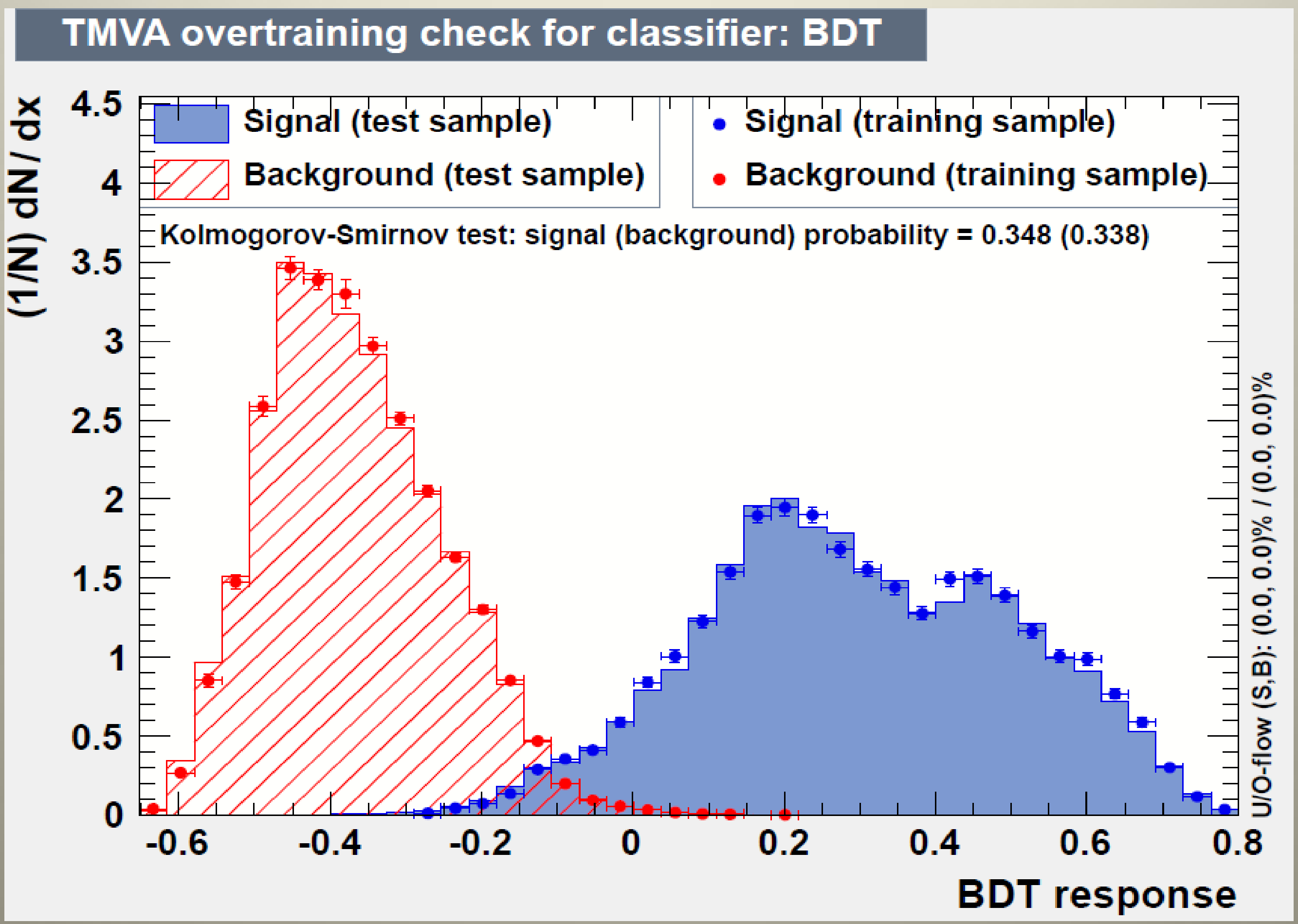
Boosted Decision Tree

- notations:
  - x_i = the set of PID variables for the ith event.
  - $y_i = 1$ if the ith event is a signal event and $y_i = -1$ if the event is a background event.
  - w_i = the weight of the ith event.
  - $T_m(x_i) = 1$ if the set of variables for the ith event lands that event on a signal leaf and $T_m(x_i) = -1$ if the set of variables for that event lands it on a background leaf.
  - $I(y_i \neq T_m(x_i)) = 1$ if $y_i \neq T_m(x_i)$ and 0 if $y_i = T_m(x_i)$.
- adaptive boost (adaboost):
- Define for the mth tree: $err_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq T_m(x_i))}{\sum_{i=1}^{N} w_i}$ $\alpha_m = \beta \times \ln((1 - err_m)/err_m)$
- Change the weight: $w_i \rightarrow w_i \times e^{\alpha_m I(y_i \neq T_m(x_i))}$ renormalisation: $w_i \rightarrow \frac{w_i}{\sum_{i=1}^{N} w_i}$

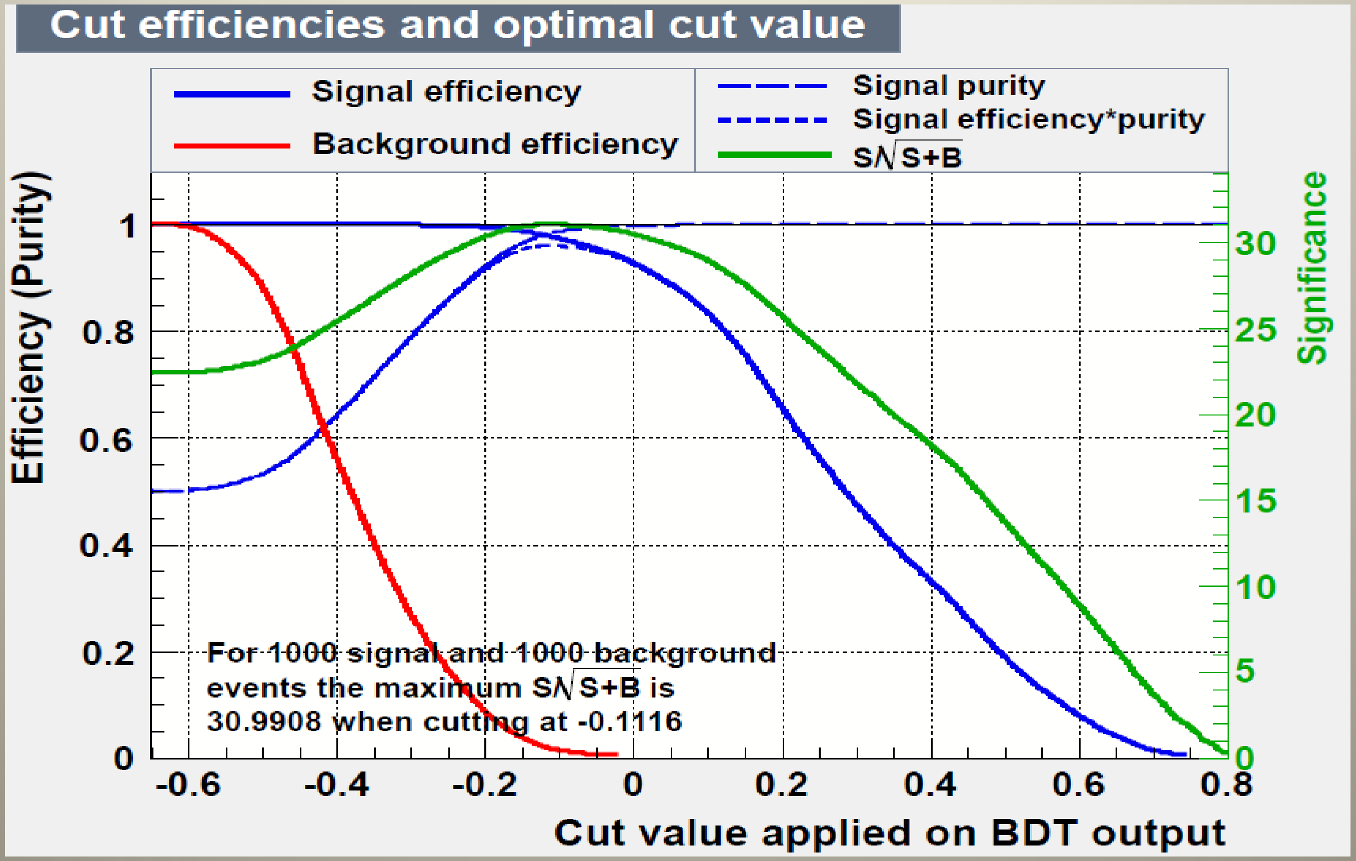
  training events……(takes a long time)

Finally, the score for a given event is: $T(x) = \sum_{m=1}^{M} \alpha_m T_m(x)$

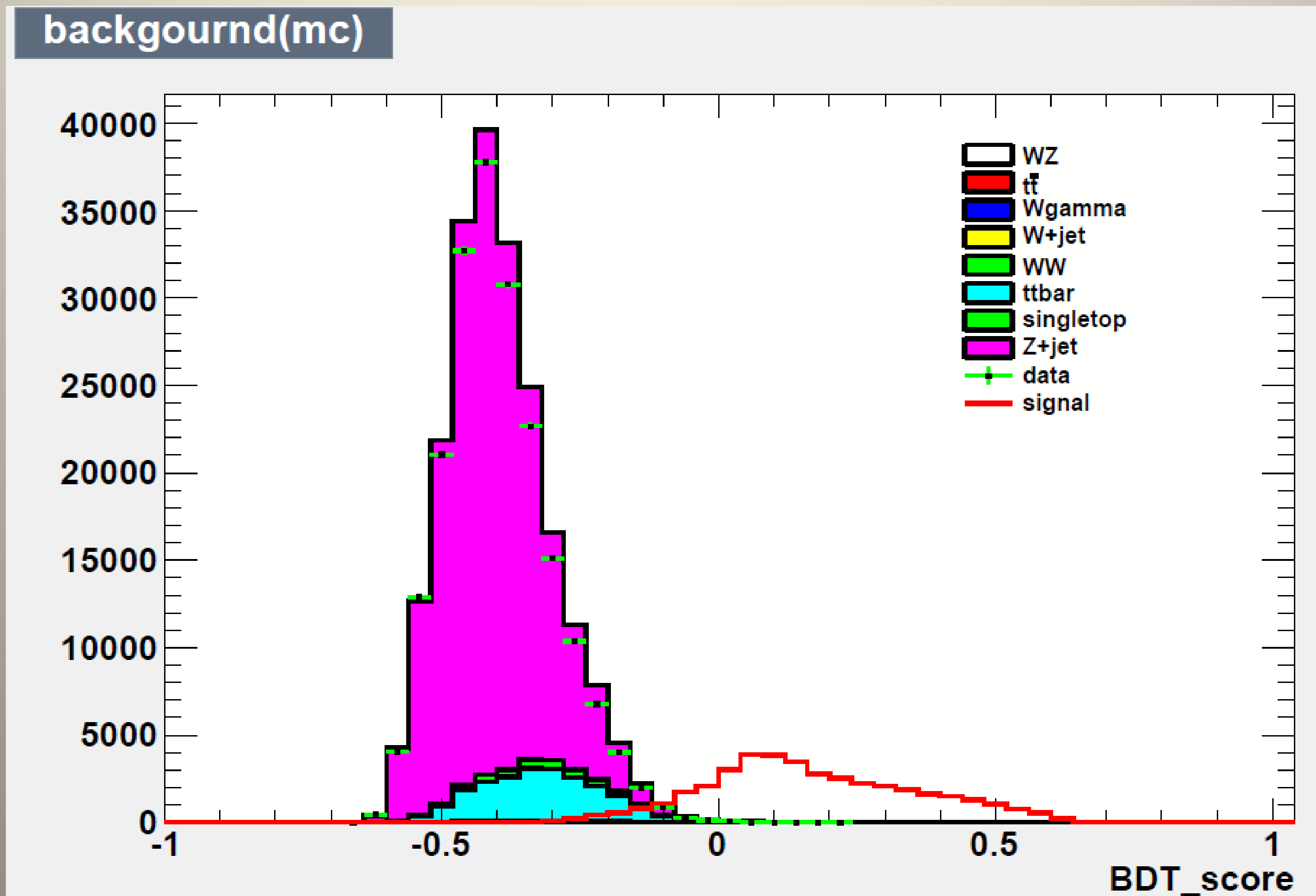
BDT training



Cut efficiency



BDT application



Reference

[1] von Jan W. Schumacher PhD-CERN-THESIS-2010-140_VBS

[2] Hai-Jun Yang , Bryon P.Roe , Ji Zhu
Studies of Boosted Decision Trees for MiniBooNE Particle Identification

Thank you !