

# Study of the HH Detection Sensitivity with BDT with $HH \rightarrow bbl + E_T^{miss}$

**Xiuyuan Zhang**, Shuzhou Zhang, Zhe Yang, Bing Li, Bing Zhou

The Univ. of Michigan

9/14/2020

# Event selection and expected yields

## • Event selection

- single & dilepton triggers
- exactly 2 light leptons
  - $p_T^l > 9\text{GeV}$
- opposite charge of lepton pair
- exactly 2 b-tagged jets
  - $p_T^{b-jet} > 20\text{GeV}$ , DL1r, 77%
- Signal region 1 (SR1)
  - $15\text{GeV} < m_{ll} < 75\text{GeV} (110\text{GeV})$ 
    - for  $ee + \mu\mu$  ( $e\mu + \mu e$ )
  - target  $bbWW$ ,  $bb\tau\tau$  and low  $m_{ll}$   $bbZZ$
  - can be further separated to 3 SR based on  $m_{ll}(> 75\text{GeV}$  or  $< 75\text{GeV})$ , and

lepton pair flavor type  
(same/different flavor),  
for BDT training

## • Signal region 2 (SR2)

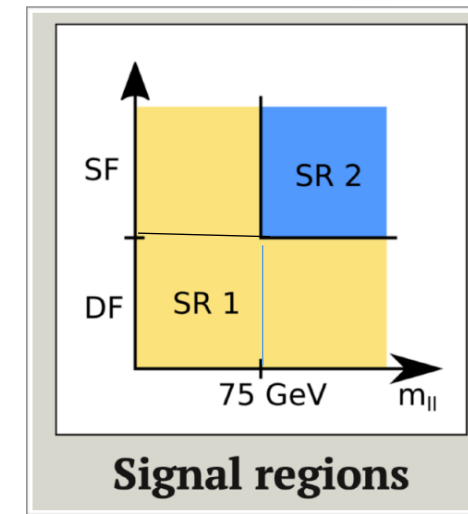
- $ee + \mu\mu$  only
- $75\text{GeV} < m_{ll} < 110\text{GeV}$
- $40\text{GeV} < m_{bb} < 210\text{GeV}$
- high  $m_{ll}$   $bbZZ$

## • Top control (Top CR)

- $m_{ll} > 110\text{GeV}$

## • DY control (Zll CR)

- $ee + \mu\mu$  only
- $75\text{GeV} < m_{ll} < 110\text{GeV}$
- $m_{bb} < 40\text{GeV}$  or  $m_{bb} > 210\text{GeV}$
- also used by  $bb\tau\tau$  analysis



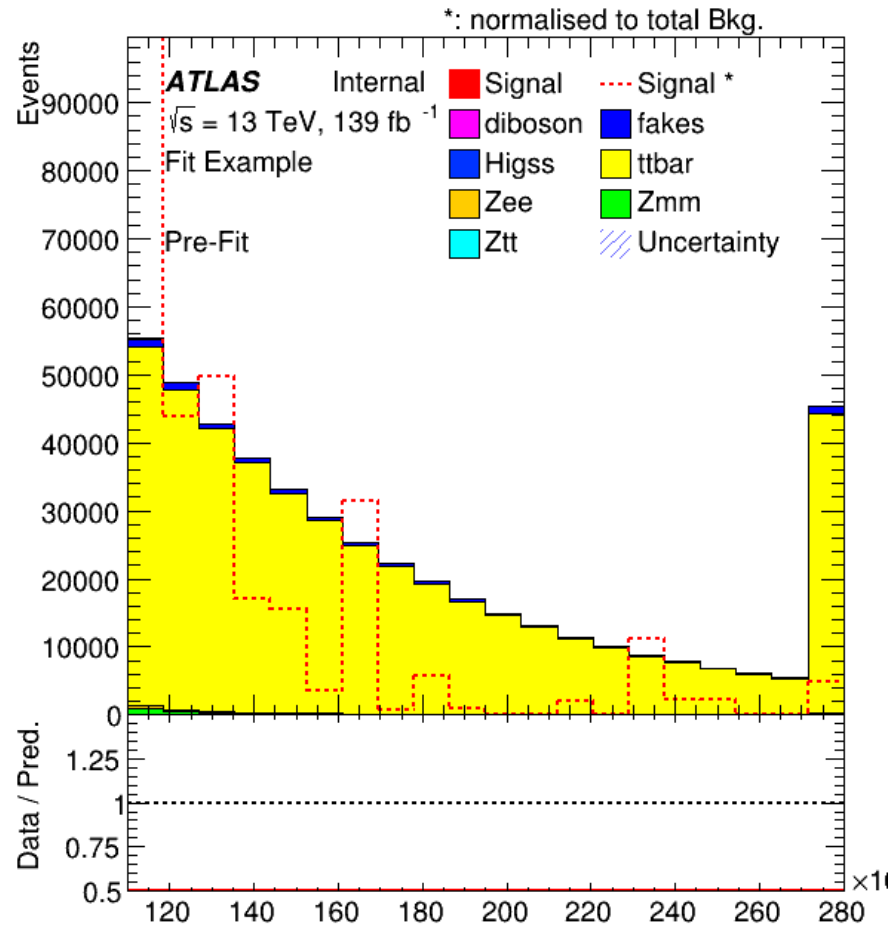
BDT  
training  
Used 4 SR  
as shown

## Expected event yields in signal region 1

►  $\int \mathcal{L} dt = 139\text{fb}^{-1}$  (2015-2018)

Process	$ee + \mu\mu$	$e\mu + \mu e$
$bbWW$	$4.40 \pm 0.07$	$4.49 \pm 0.07$
$bb\tau\tau$	$1.524 \pm 0.016$	$1.760 \pm 0.018$
$bbZZ$	$0.0744 \pm 0.0012$	$0.00168 \pm 0.00015$
$t\bar{t}/Wt/t\bar{t}V$	$240080 \pm 150$	$388090 \pm 190$
$Z \rightarrow \ell\ell$	$20760 \pm 250$	$110 \pm 11$
Fakes <sup>1</sup>	$8510 \pm 80$	$12940 \pm 70$
$Z \rightarrow \tau\tau$	$1370 \pm 50$	$1410 \pm 50$
Diboson	$308 \pm 4$	$142 \pm 2$
$H$	$207.5 \pm 0.7$	$271.8 \pm 0.8$
$s/b$	$22 \times 10^{-6}$	$15 \times 10^{-6}$

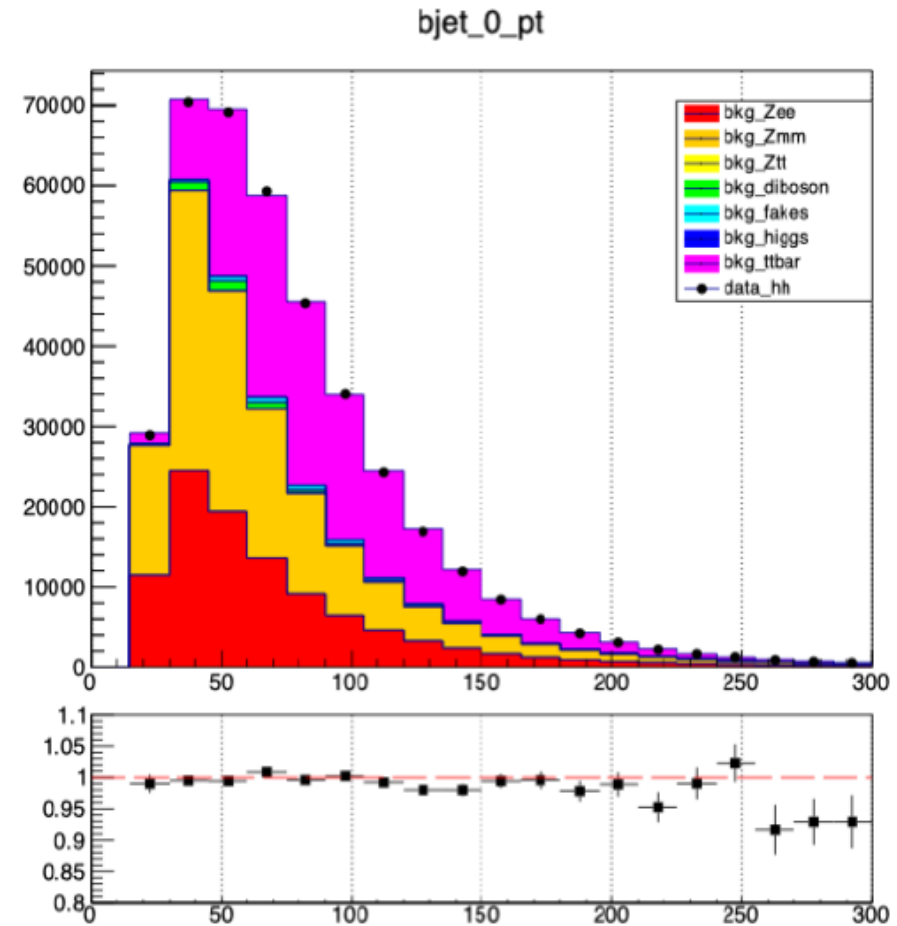
# Control Regions



Top:

Reweighed by a quadratic fitting function with an overall normalization factor 0.901

DY: An overall scale factor 1.245 is applied



# Four BDT training Regions

- R1:
  - $75\text{GeV} < m_{ll} < 110\text{GeV}, ee + \mu\mu$
- R2:
  - $75\text{GeV} < m_{ll} < 110\text{GeV}, e\mu + \mu e$
- R3:
  - $15\text{GeV} < m_{ll} < 75\text{GeV}, ee + \mu\mu$
- R4:
  - $15\text{GeV} < m_{ll} < 75\text{GeV}, e\mu + \mu e$

# Multi-Class BDT Training

- Event pre-selection before training:

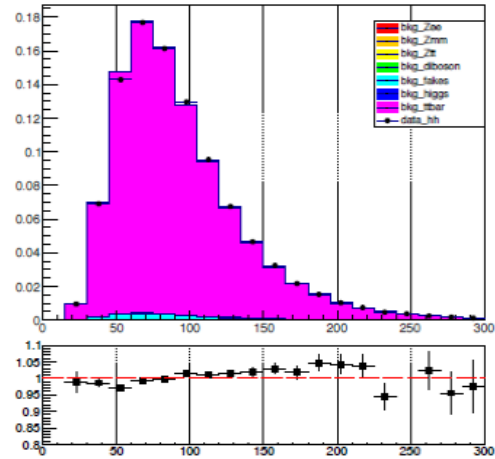
$$150 \text{ GeV} > m_{bb} > 80 \text{ GeV}, |\Delta(R_{bb})| < 2.8$$

- Input variables for BDT training:

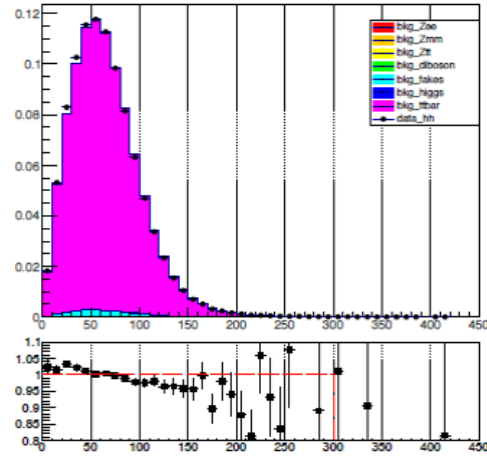
- **bjets** - bjet\_1\_pt, bjet\_1\_eta, bjet\_2\_pt, bjet\_2\_eta,
- **Dilepton** - ll\_m, ll\_pt, ll\_deltar, ll\_deltaeta, ll\_deltaphi,
- **bb** - bb\_m, bb\_pt, bb\_deltar, bb\_deltaeta, bb\_deltapt,
- **Met** - met\_met, met\_sig, met\_phi,
- **m<sub>T</sub>** - mt\_lep0\_met, mt\_lep1\_met,
- **M<sub>bb</sub>** - m\_bbll, m\_bbllmet

# Examples of input variable distributions (in R2)

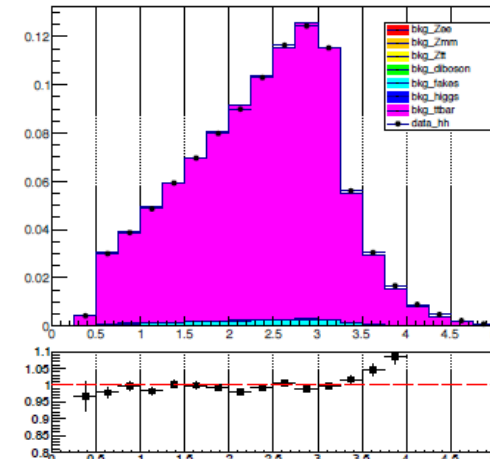
bjet\_0\_pt



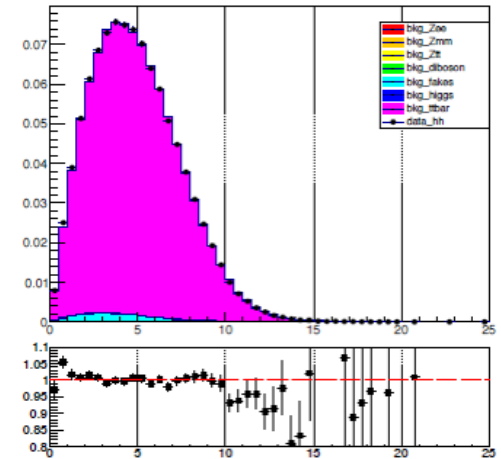
ll\_pt



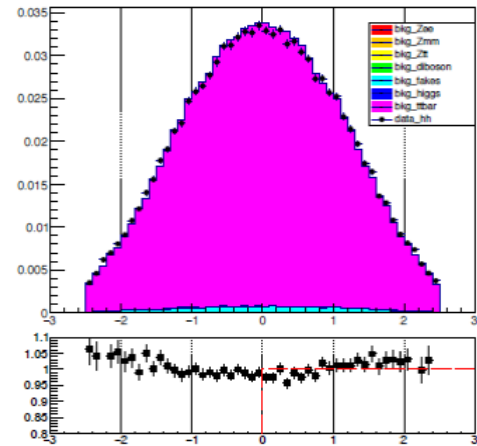
bb\_deltar



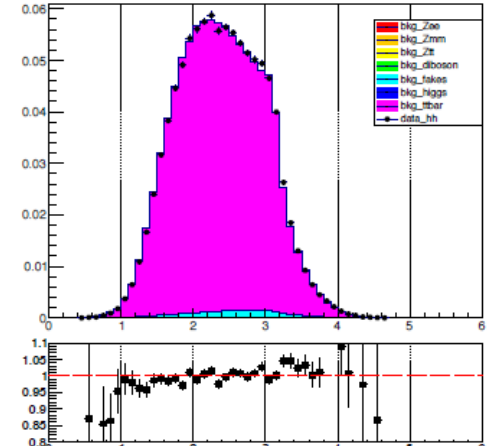
met\_sig



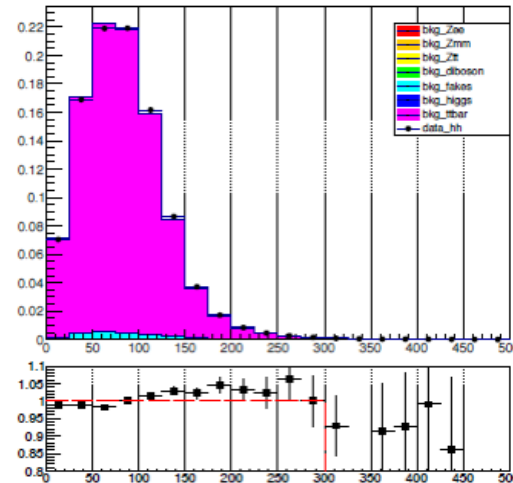
bjet\_0\_eta



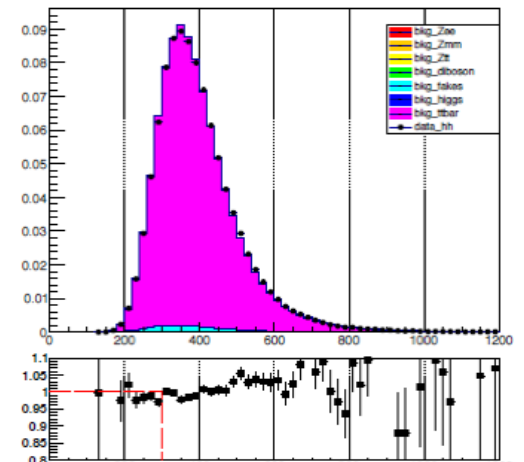
ll\_deltar



bb\_pt



m\_bblimet

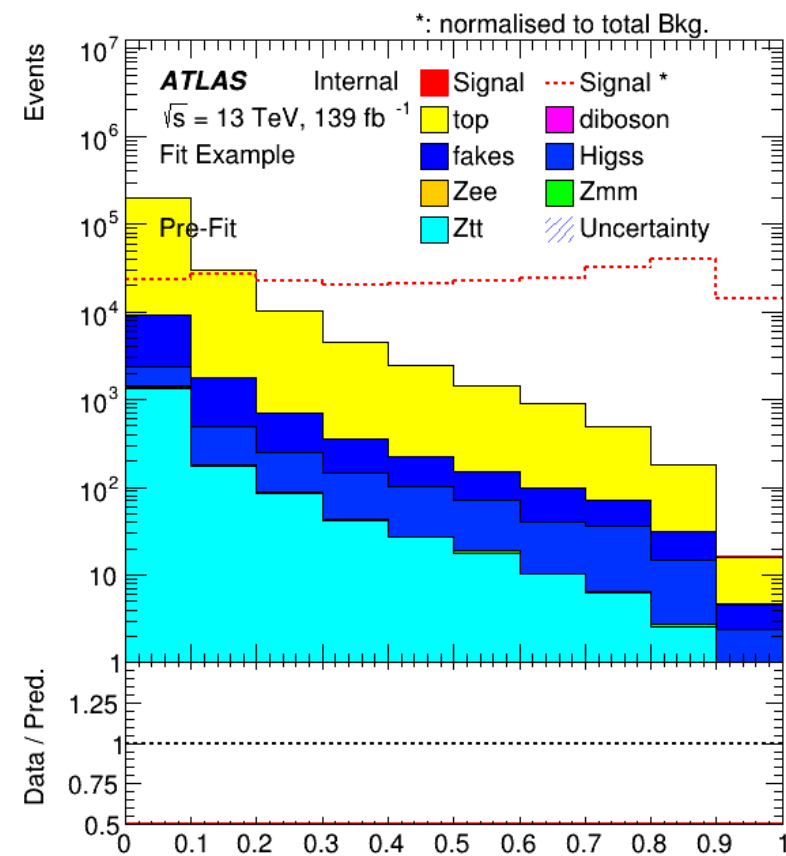
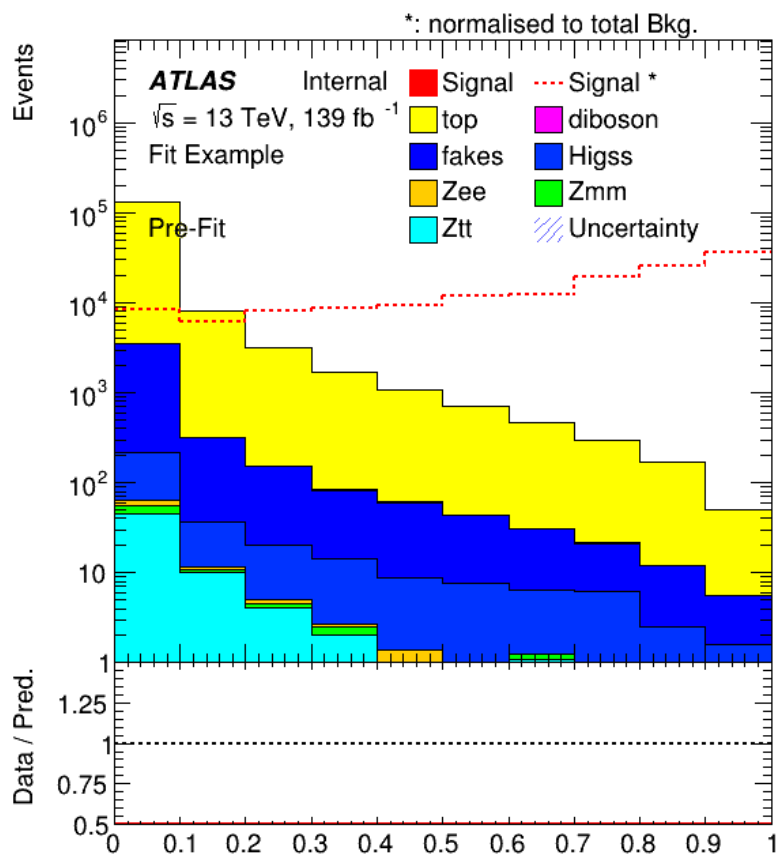


# Multi-class BDT Training

## Use multi-class training to train BDT in each region

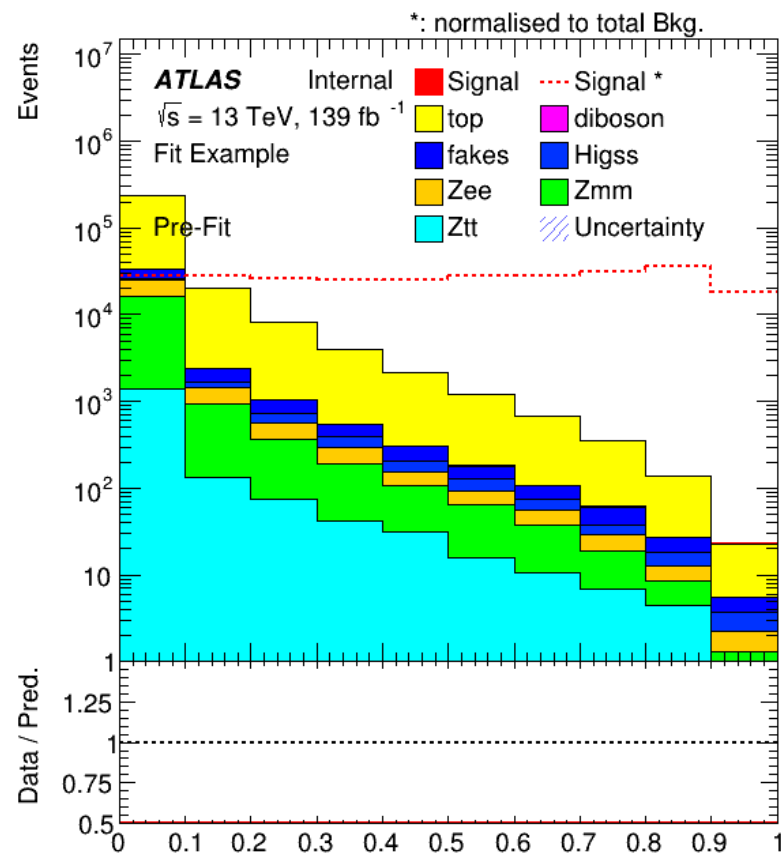
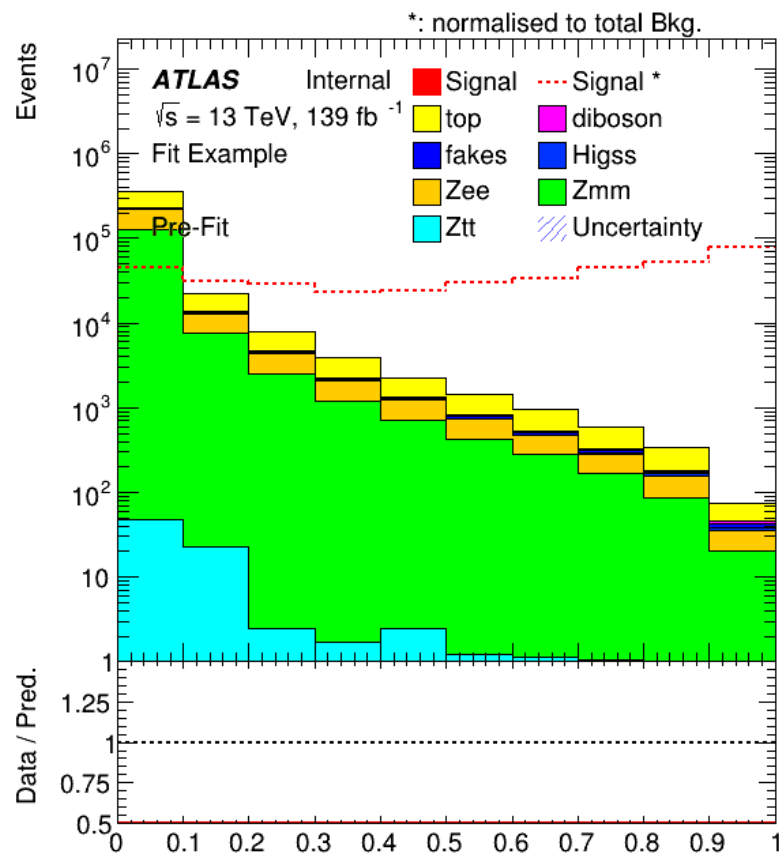
- In region 2 and region 4 ( $e\mu$ ), use 4 classes
  - Class 0: **Signal**
  - Class 1: Top
  - Class 2: Other(di-boson, single Higgs)
  - Class 3: fakes 6
- In region 1 and region 3( $ee, \mu\mu$ ), use 5 classes
  - Class 0:DY
  - Class 1:**Signal**
  - Class 2:Top
  - Class 3:Other(di-boson, single Higgs)
  - Class 4:fakes 5

# BDT Spectra in R2 and R4 ( $e\mu$ )





# BDT Spectra in R1 and R3 (ee, $\mu\mu$ )



# Sensitivity Fitting Program Setup (1)

Using Asimov data to obtain the expected sensitivity; Statistic ONLY

- POI: SigXsecOverSM(norm factor of signal), 1, -10, 10
- Nuisance parameter:
- mutop(norm factor of top background), 1, 0, 3
- FitBlind:TRUE
- SignificanceBlind:TRUE
- POIAsimov: 1
- Top control region:
- Variable: ll\_m(20 bins, 110GeV, 300GeV)

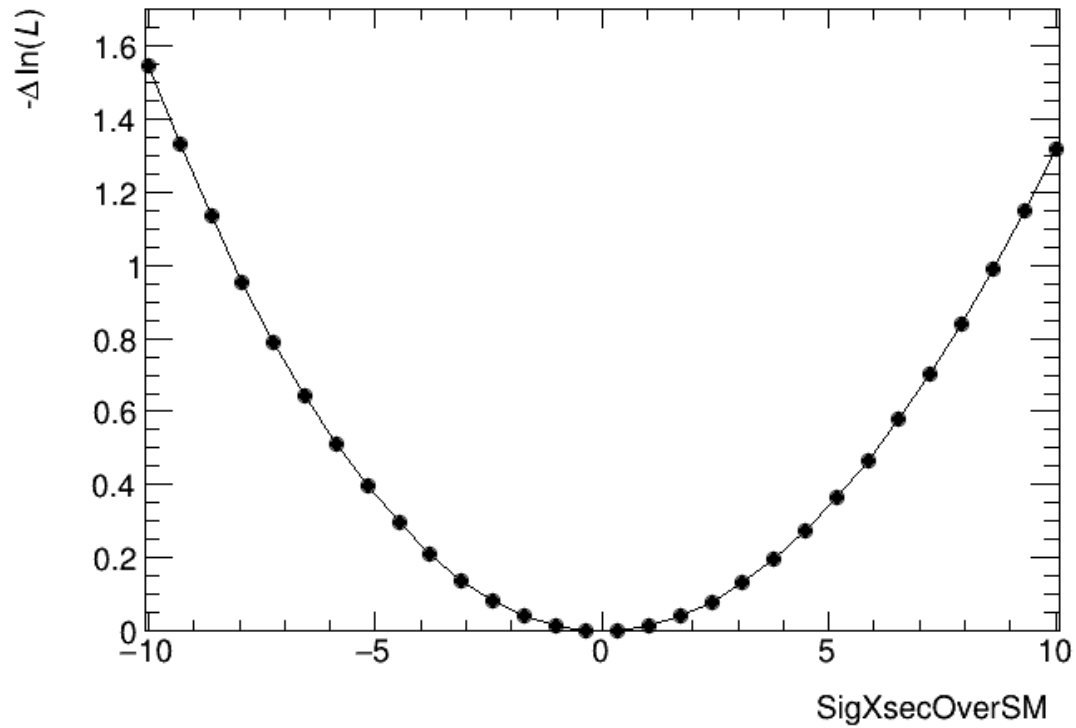
# Sensitivity Fitting Program Setup (2)

Obtain the upper limits by fitting four signal regions simultaneously

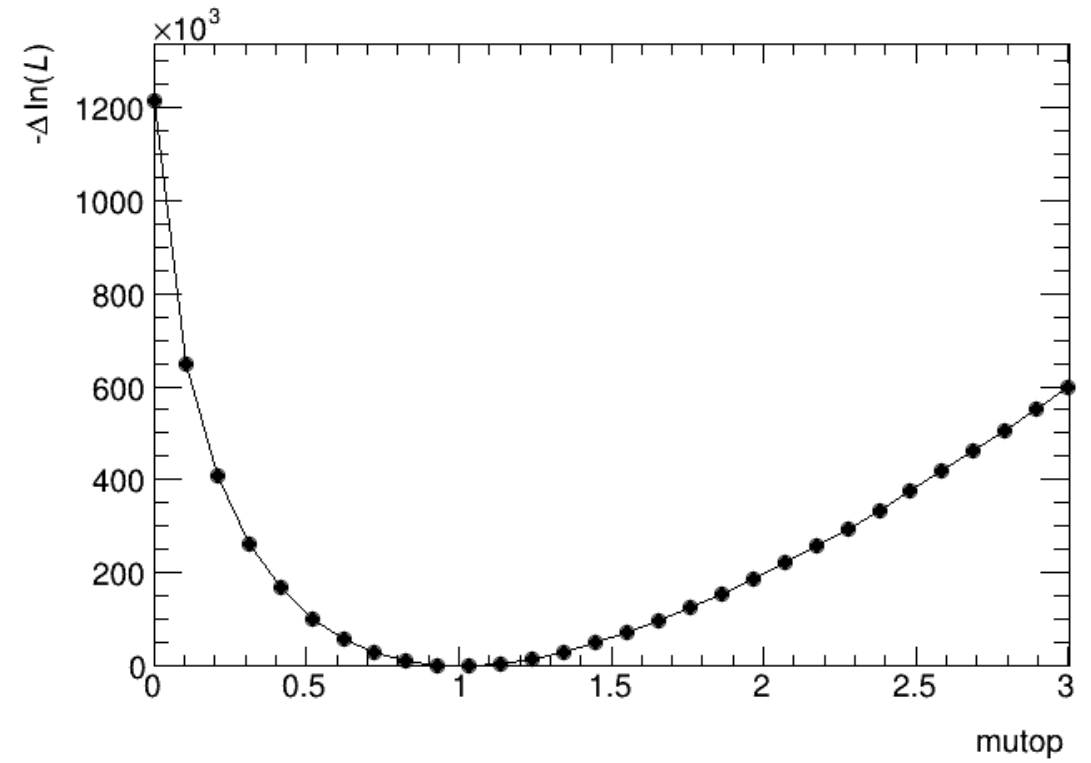
- Region 1(HM/SF):
  - Variable: BDT\_R1\_5\_class\_sig(10 bins, 0,1)
- Region 2(HM/DF):
  - Variable: BDT\_R2\_4\_class\_sig(10 bins, 0,1)
- Region 3(LM/SF):
  - Variable: BDT\_R3\_5\_class\_sig(10 bins, 0,1)
- Region 4(LM/DF):
  - Variable: BDT\_R4\_4\_class\_sig(10 bins, 0,1)

# Likelihood Plots

Signal



Top



# Expected Upper Limits on signal strength $\mu$

$-2\sigma$	$-\sigma$	Median	$\sigma$	$2\sigma$
6.511	8.74	12.13	17.03	23.35

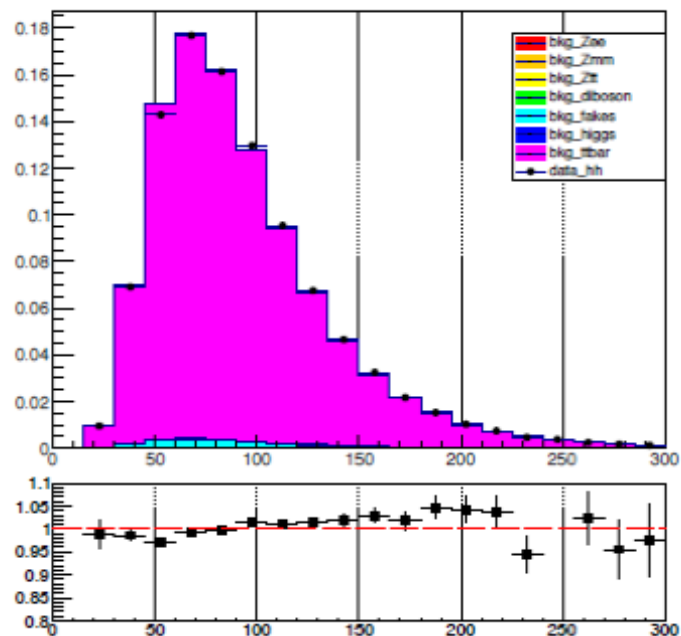
The limit is slightly worse than the training using DNN  
Will further study the systematic uncertainty impact

# Summary

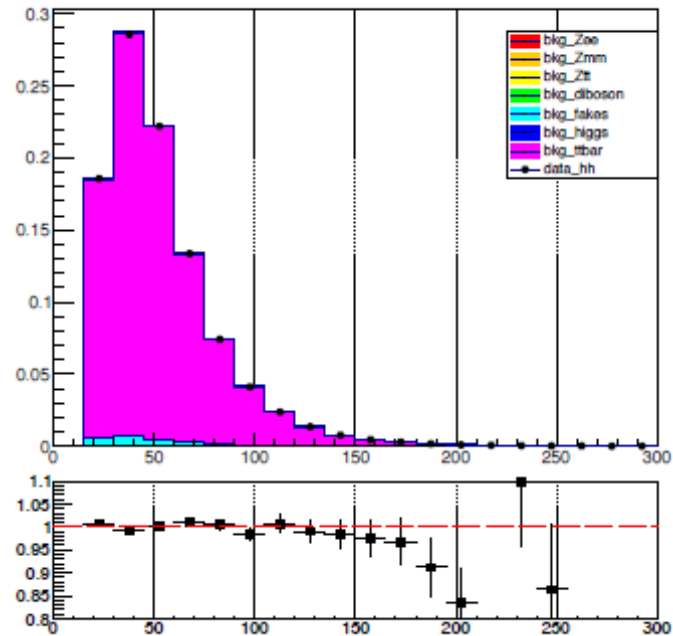
- MVA analysis performed with BDT to cross check the DNN method
- Slightly worse expected upper limits from BDT, but still quite close
- Will have further check with systematic uncertainties included
- Would like to document to supporting note

Backup: Kinematic distributions

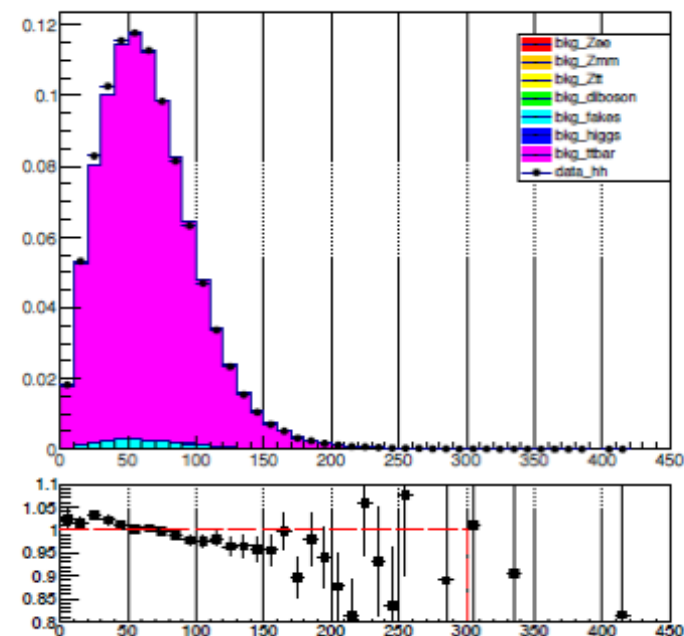
bjet\_0\_pt



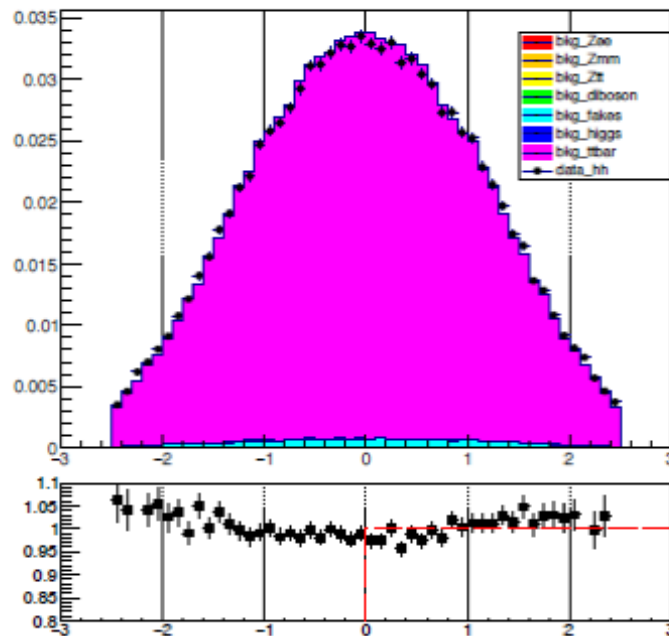
bjet\_1\_pt



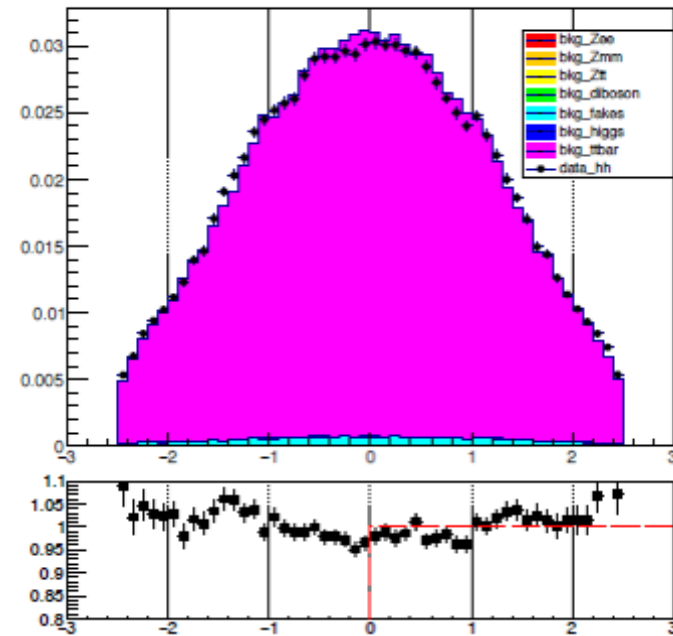
l\_l\_pt



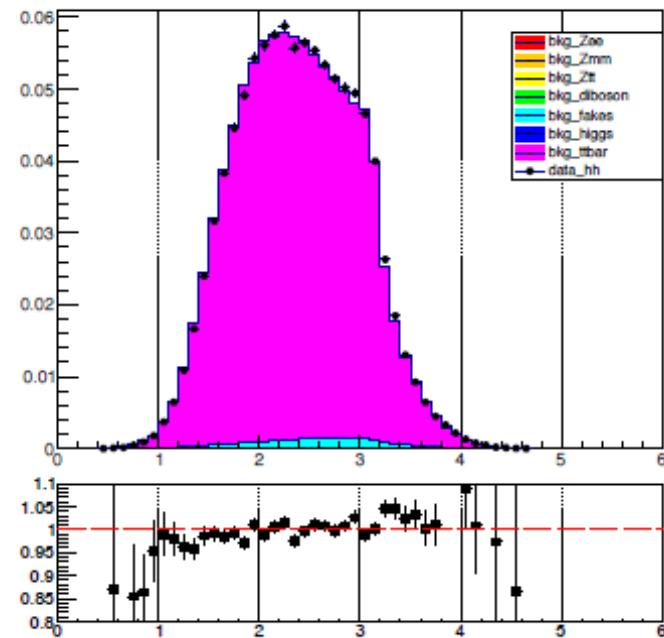
bjet\_0\_eta



bjet\_1\_eta

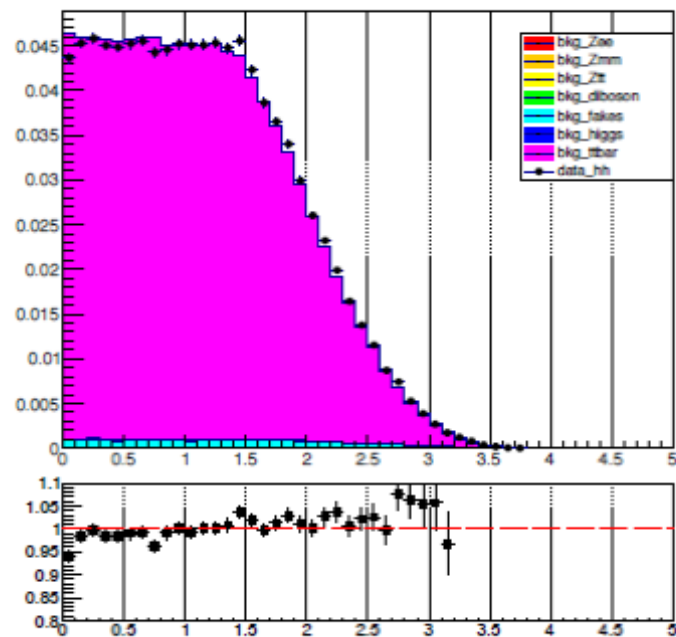


l\_l\_deltar

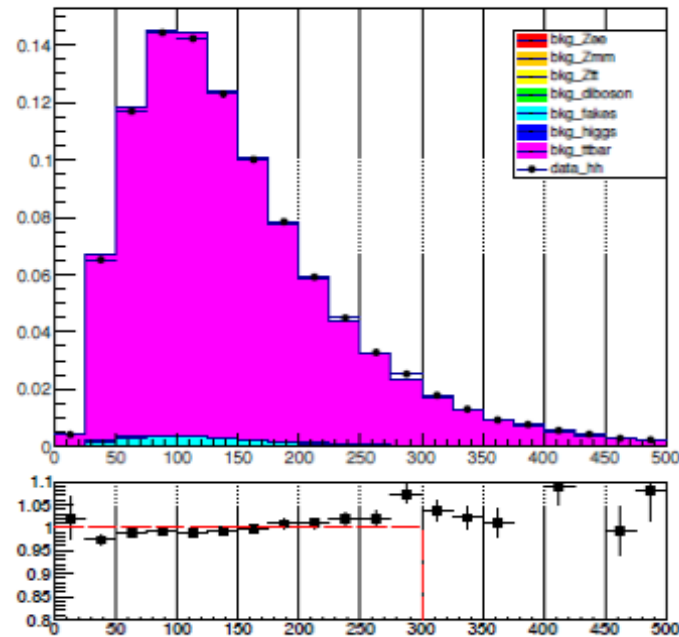




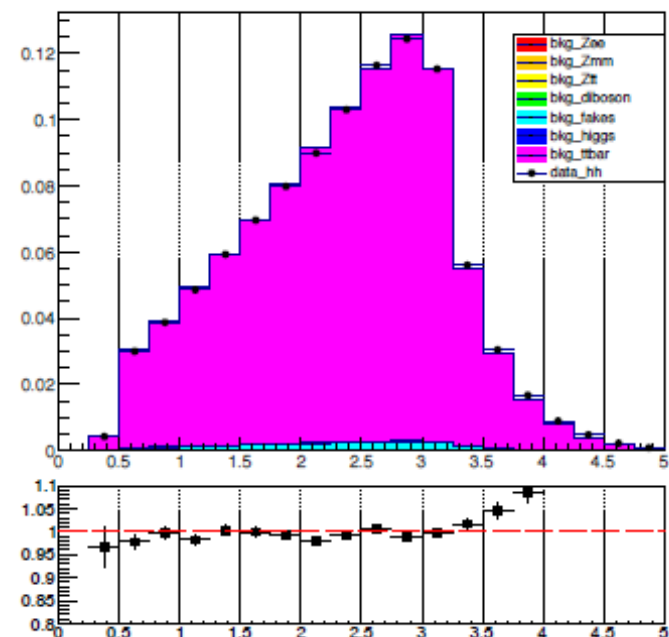
ll\_deltaeta



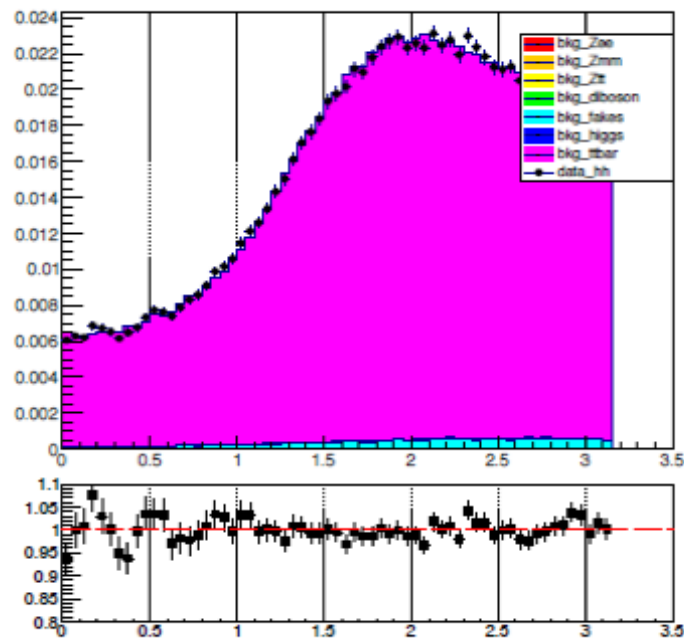
bb\_m



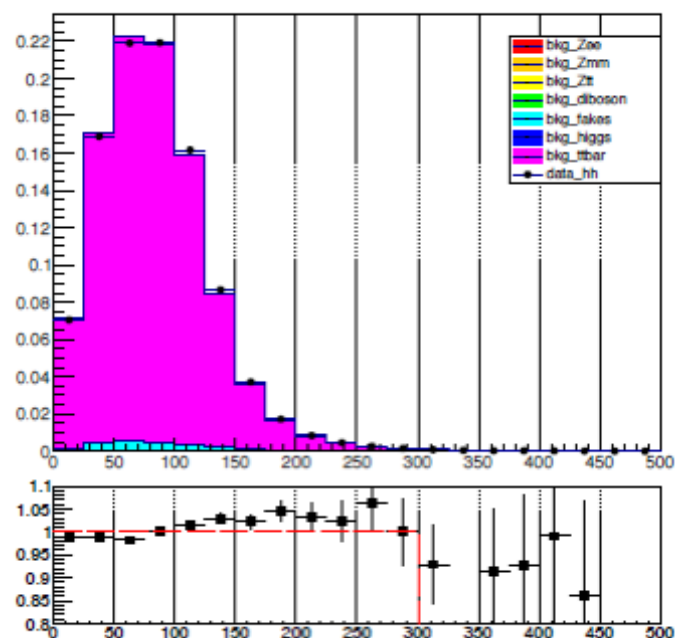
bb\_deltaeta



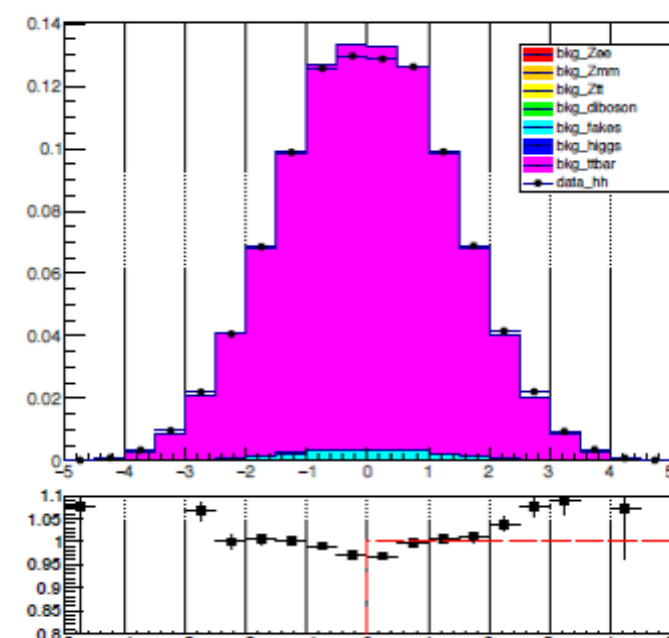
ll\_deltaphi



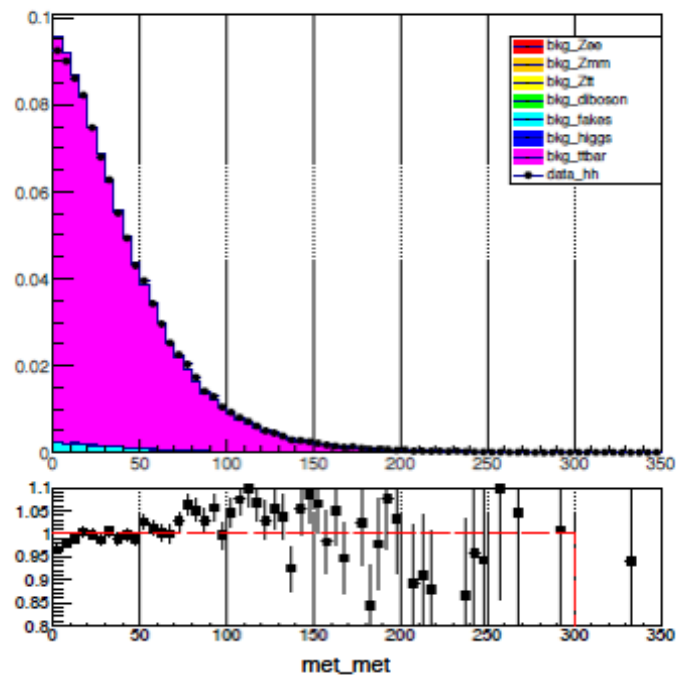
bb\_pt



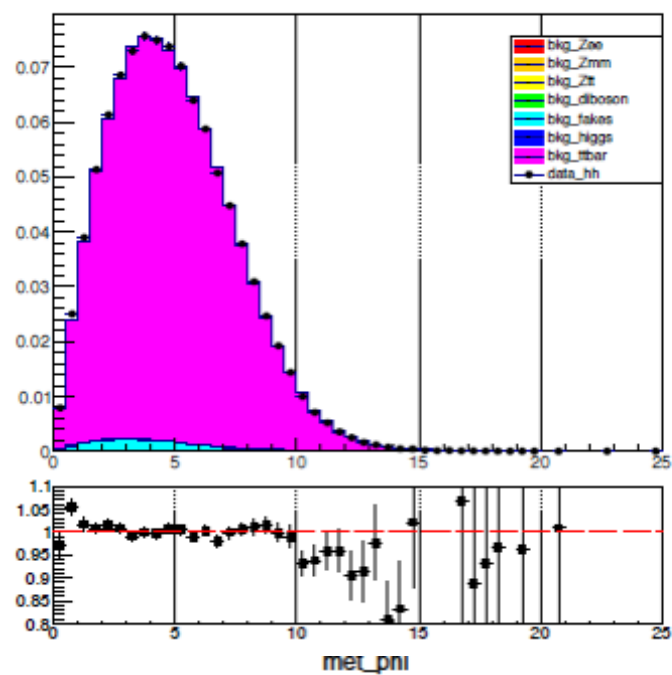
bb\_deltaeta



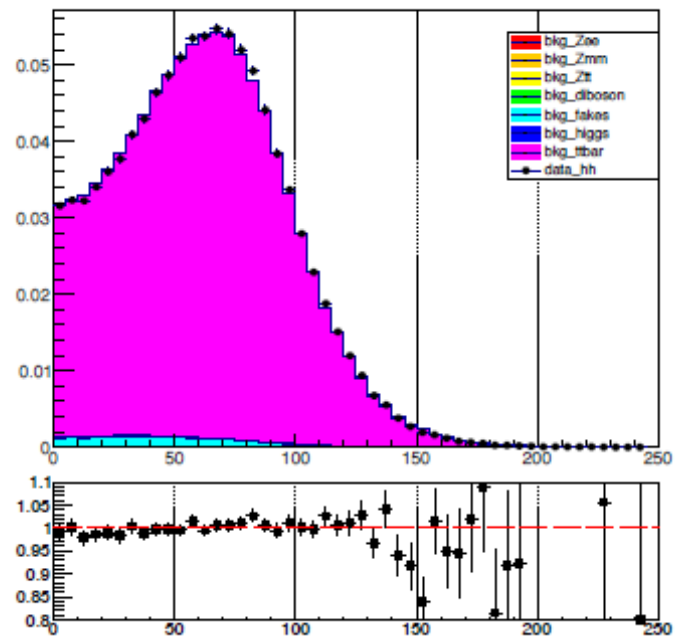
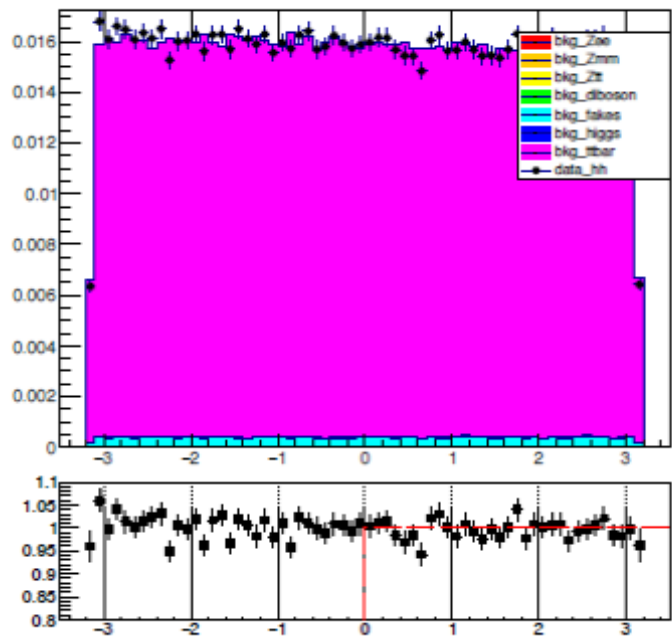
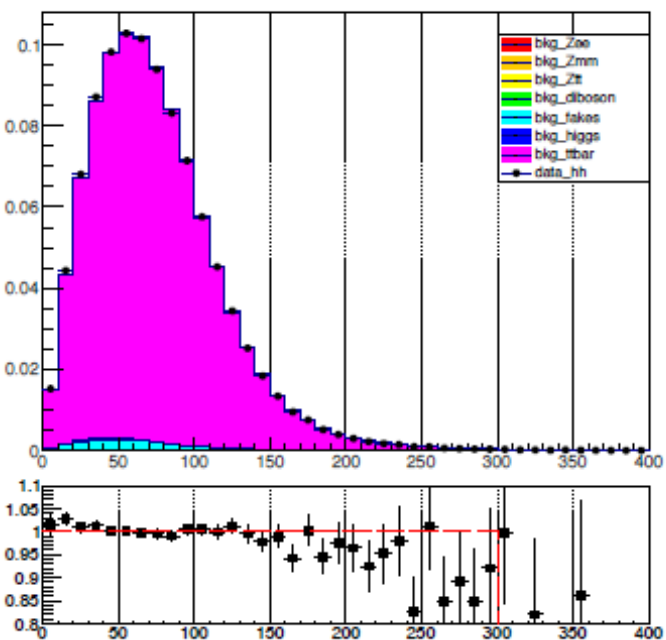
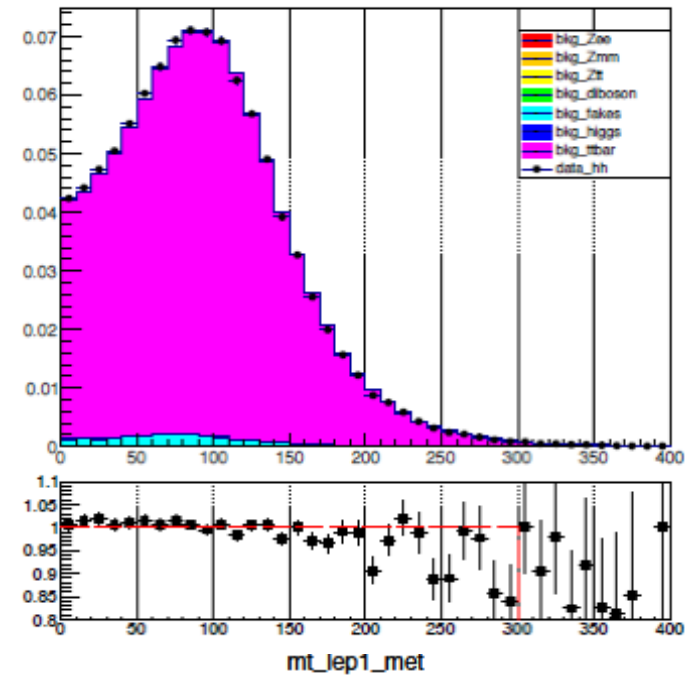
bb\_deltapt



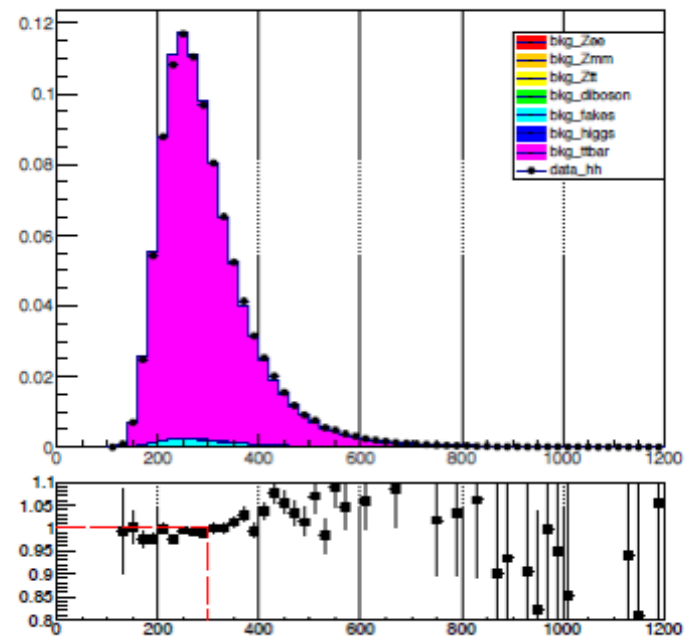
met\_sig



mt\_lep0\_met



m\_bbll



m\_bbllmet

