Study of the HH Detection Sensitivity with BDT with HH→bbll+E_T^{miss}

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

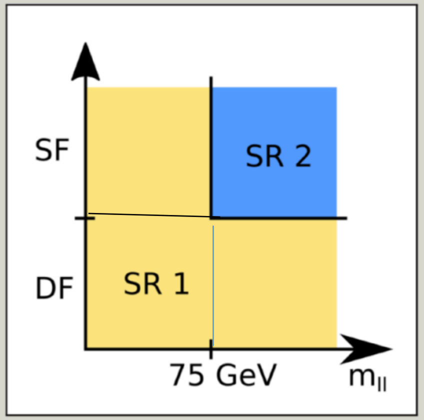
The Univ. of Michigan

9/14/2020

Event selection and expected yields

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

- **Signal region 2 (SR2)**
  - ee + μμ only
  - $75GeV < m_{ll} < 110GeV$
  - $40GeV < m_{bb} < 210GeV$
  - high m_{ll} bbZZ
- **Top control (Top CR)**
  - $m_{ll} > 110\ GeV$
- **DY control (Zll CR)**
  - $ee + \mu\mu$ only
  - $75GeV < m_{ll} < 110GeV$
  - $m_{bb} < 40GeV\ or\ m_{bb} > 210GeV$
  - also used by $bb\tau\tau$ analysis



Signal regions

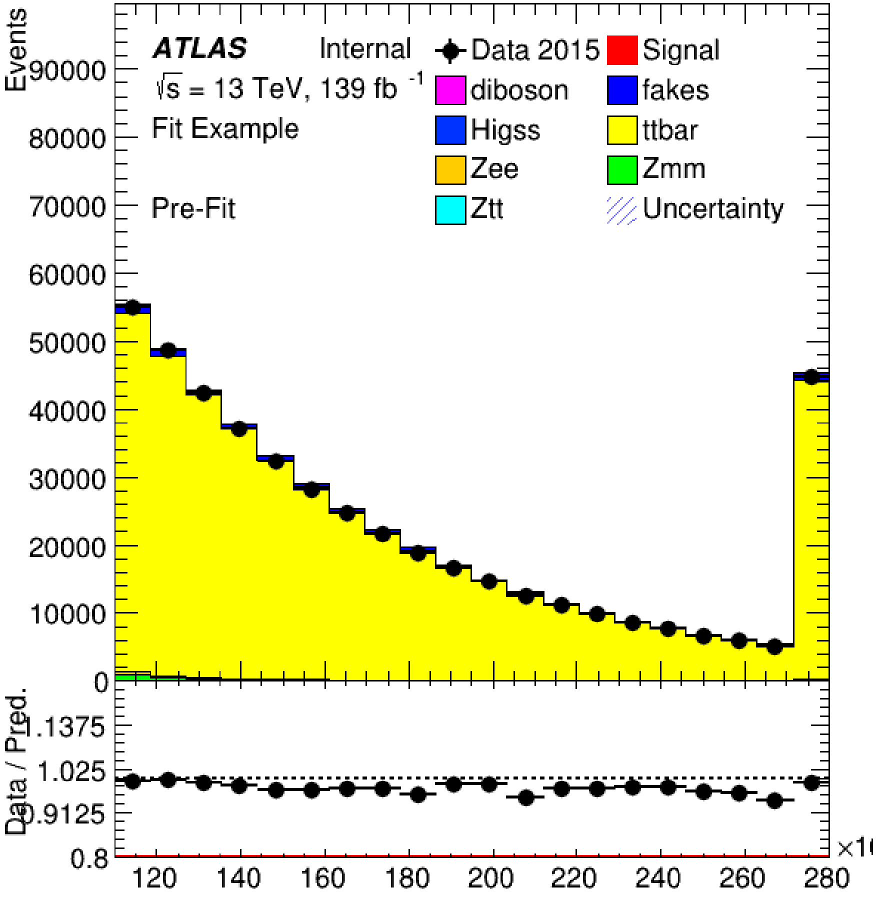
BDT training Used 4 SR as shown

Expected event yields in signal region 1

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

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

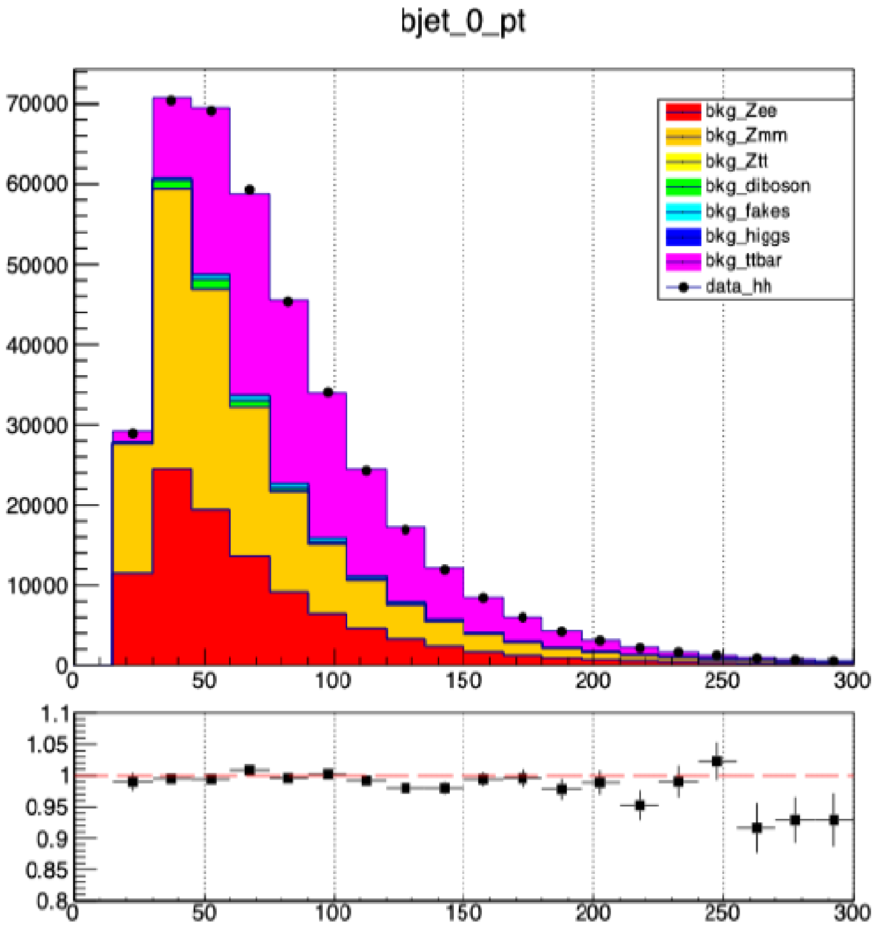
Control Regions



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

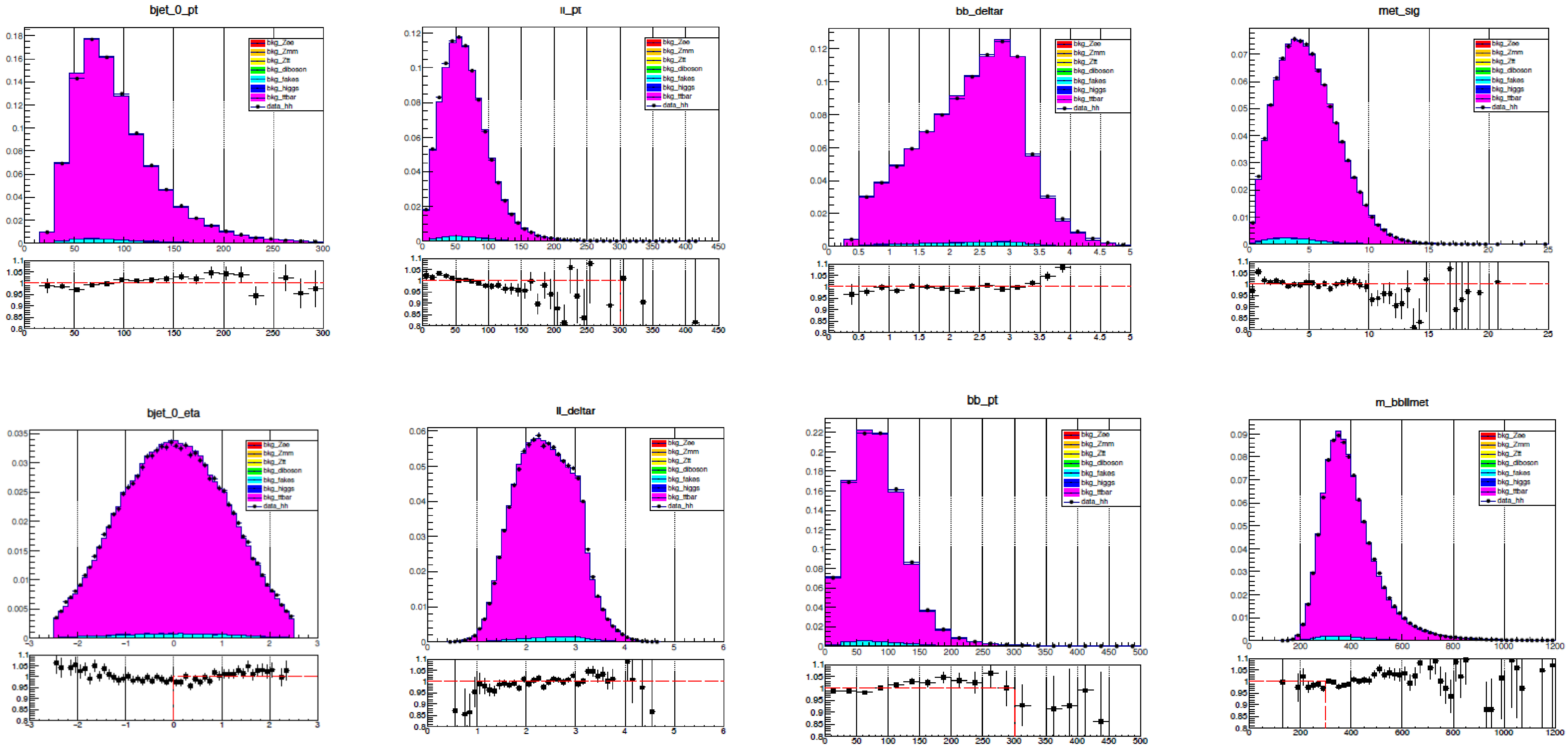
- R1:
- $75GeV < m_{ll} < 110GeV, ee + \mu\mu$
- R2:
- $75GeV < m_{ll} < 110GeV, e\mu + \mu e$
- R3:
- $15GeV < m_{ll} < 75GeV, ee + \mu\mu$
- R4:
- $15GeV < m_{ll} < 75GeV, e\mu + \mu e$

Multi-Class BDT Training

- Event pre-selection before training:

  $150\ GeV > m_{bb} > 80 GeV,\ |\Delta(R_{bb})| < 2.8$
- 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_T - mt_lep0_met, mt_lep1_met,
  - M_{bb} - m_bbll, m_bbllmet

Examples of input variable distributions (in R2)

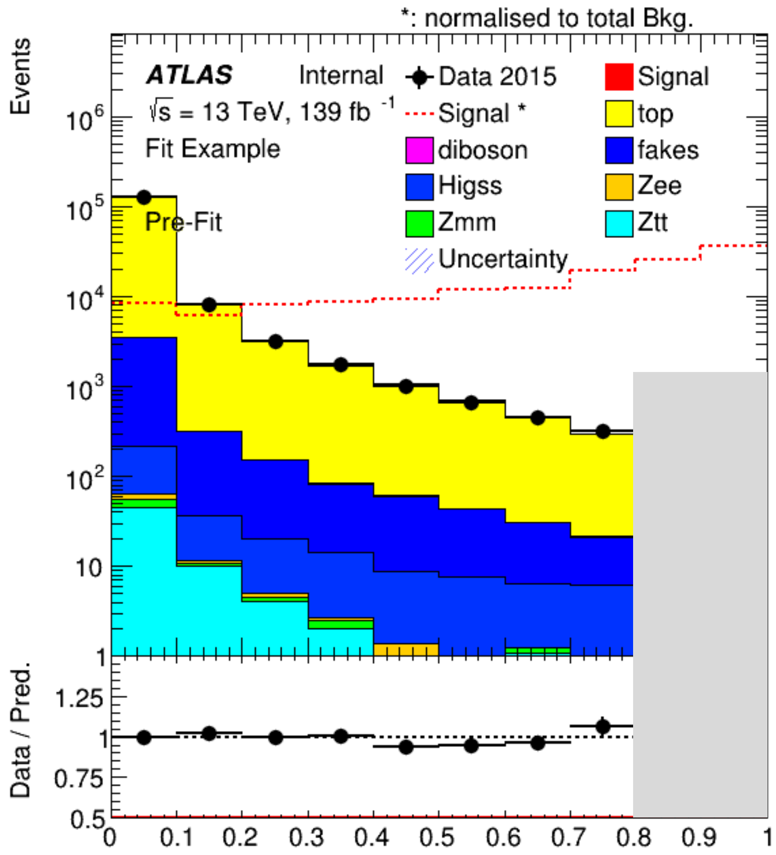


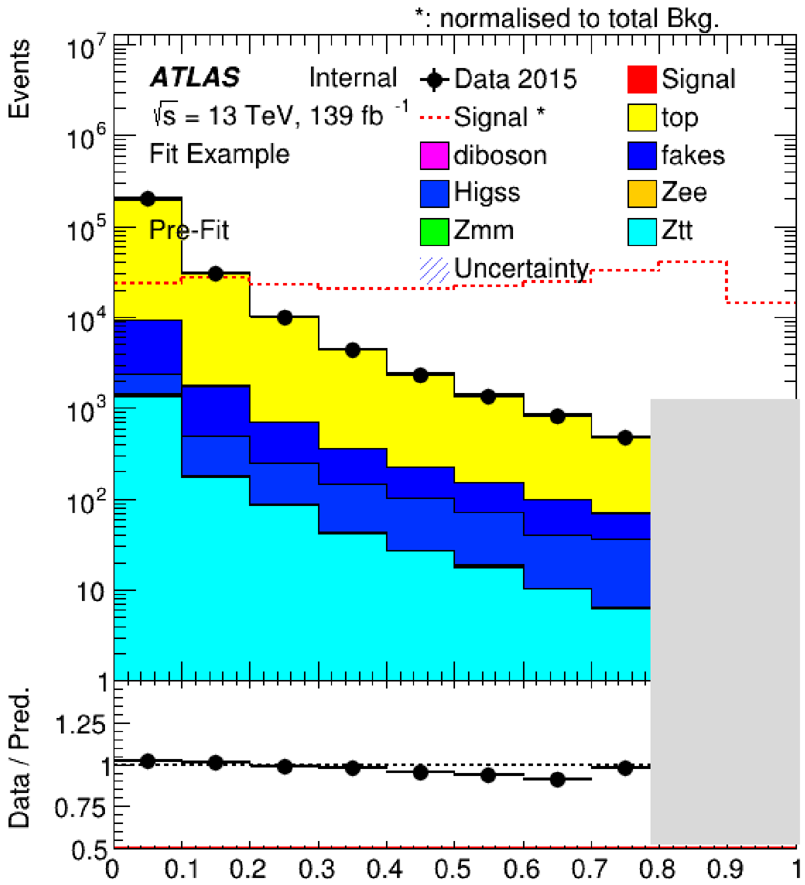
Multi-class BDT Training

Use multi-class training to train BDT in each region

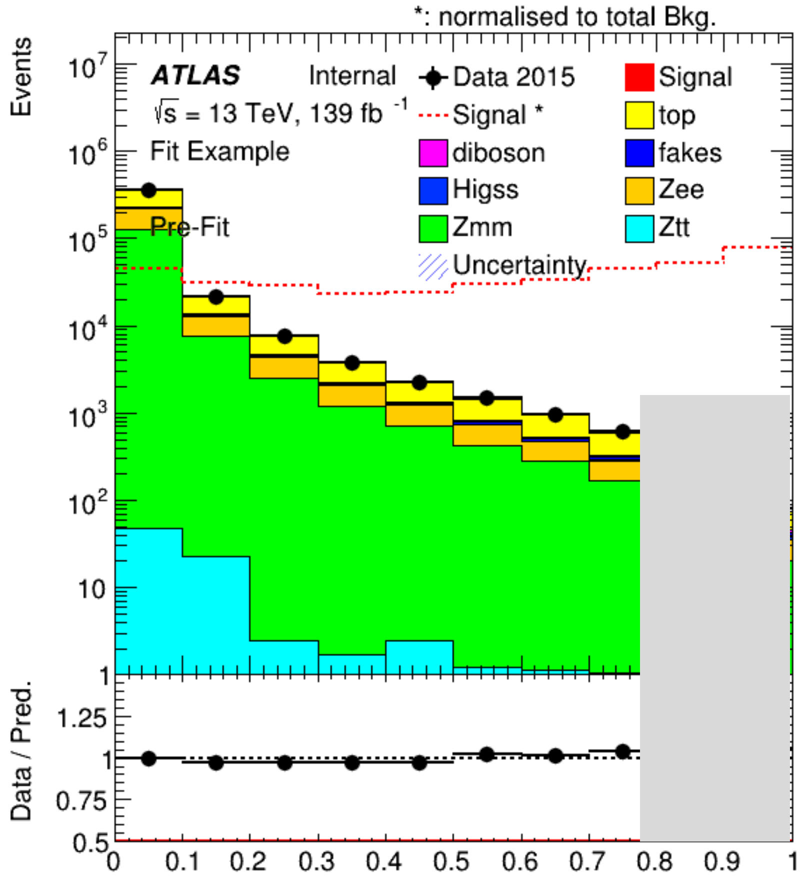
- In region 2 and region 4 (eμ), use 4 classes
  - Class 0: Signal
  - Class 1: Top
  - Class 2: Other(di-boson, single Higgs)
  - Class 3: fakes 6
- In region 1 and region 3(ee, $\mu\mu$), use 5 classes
  - Class 0:DY
  - Class 1:Signal
  - Class 2:Top
  - Class 3:Other(di-boson, single Higgs)
  - Class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

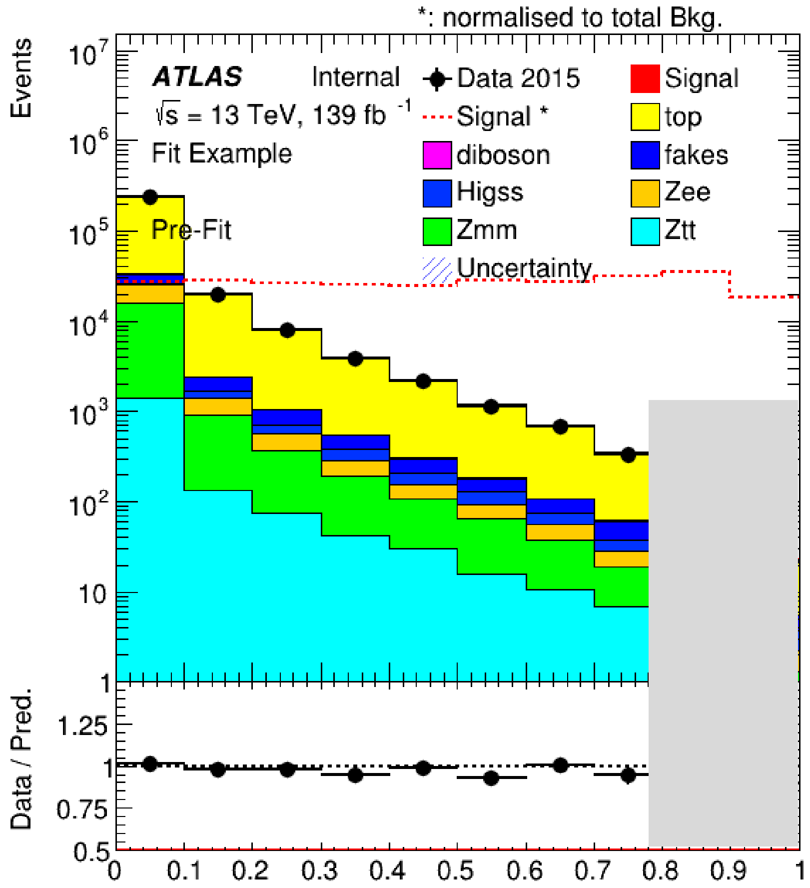
BDT Spectra in R2 and R4 (eμ)





BDT Spectra in R1 and R3 (ee, μμ)





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)

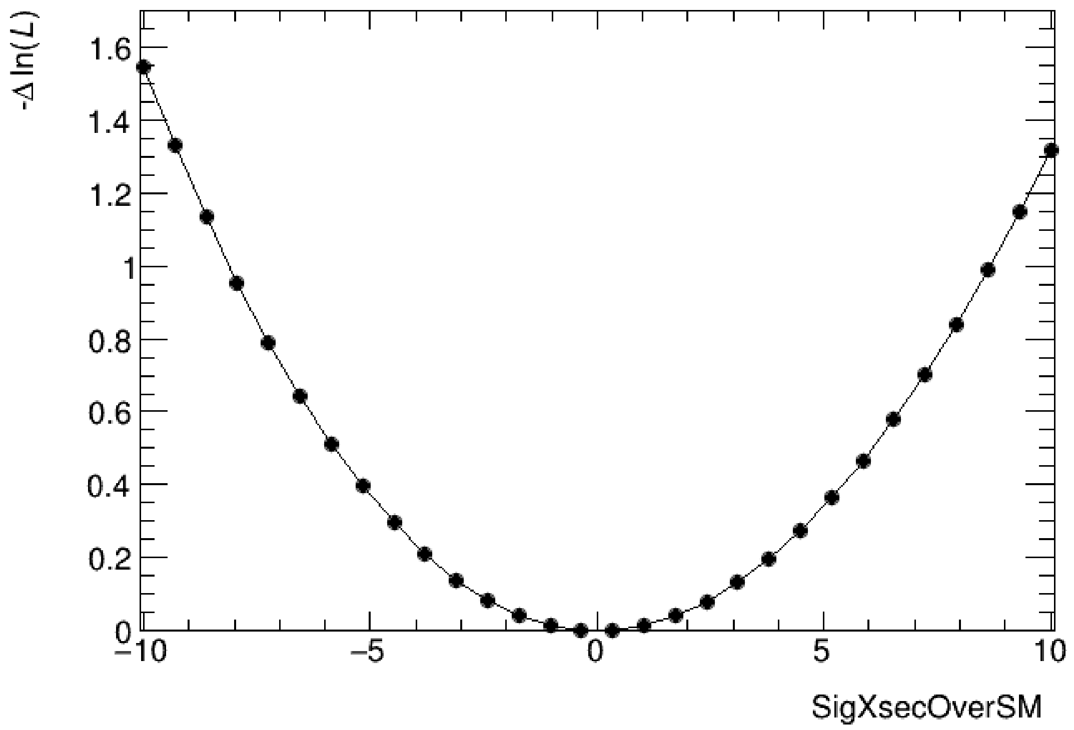
Sensitivity Fitting Program Setup (2)

Obtain the upper limits by fitting four signal regions simultaneously

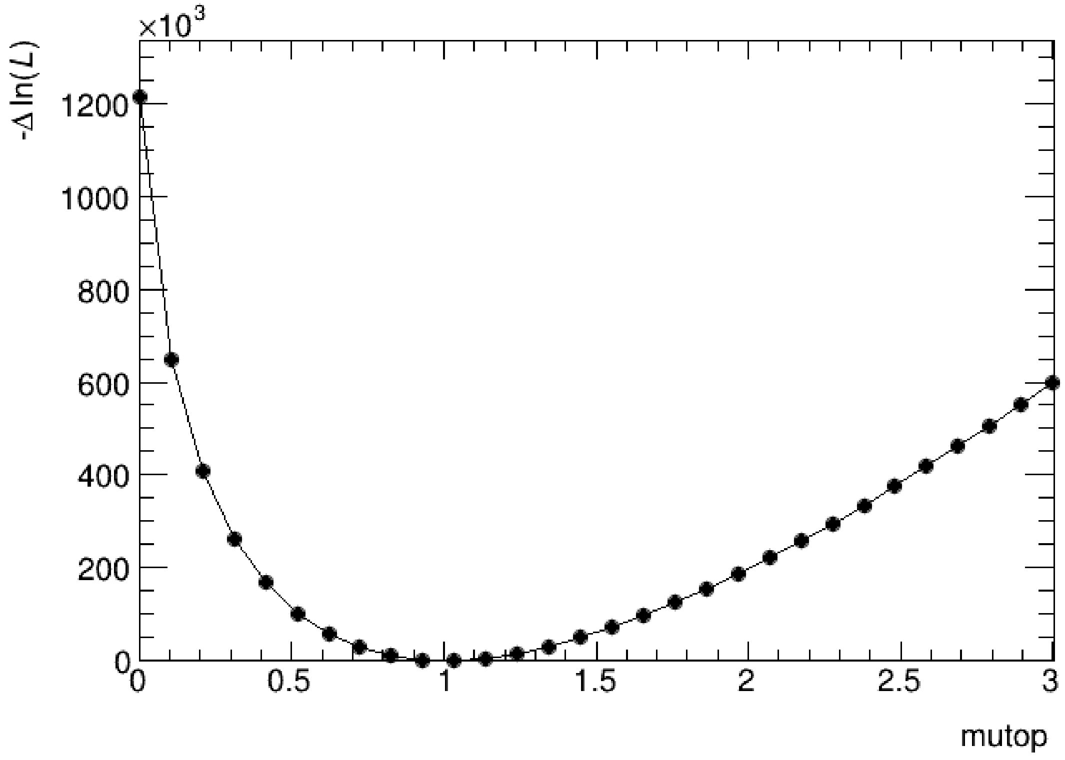
- Region 1(HM/SF):
  - Variable: BDT_R1_5_class_sig(10 bins, 0,1)
- Region 2(HM/DF):
  - Variable: BDT_R2_4_class_sig(10 bins, 0,1)
- Region 3(LM/SF):
  - Variable: BDT_R3_5_class_sig(10 bins, 0,1)
- Region 4(LM/DF):
  - Variable: BDT_R4_4_class_sig(10 bins, 0,1)

Likelihood Plots

Signal



Top



Expected Upper Limits on signal strength μ

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

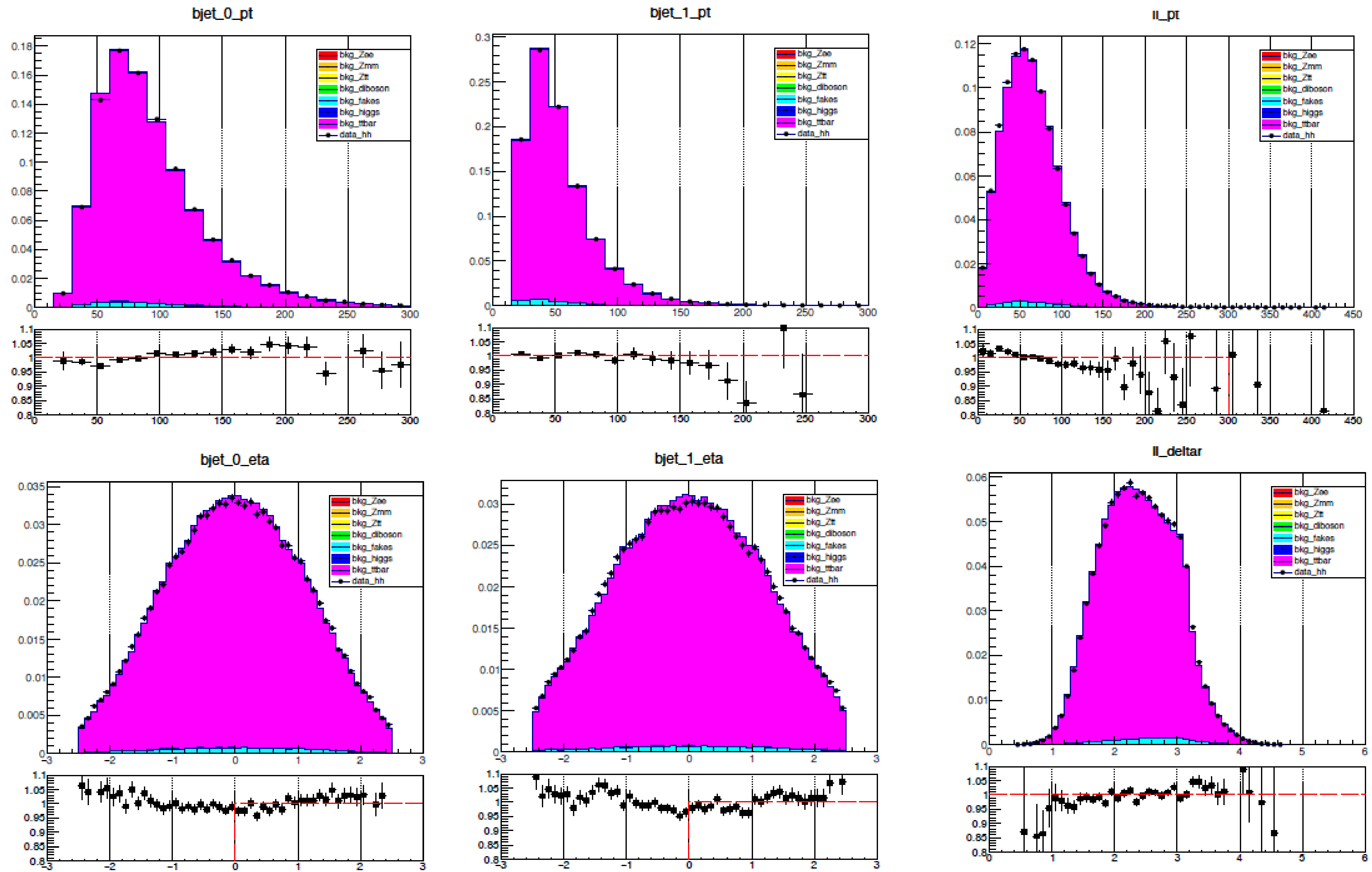
The limit is slightly worse than the training using DNN
Will further study the systematic uncertainty impact

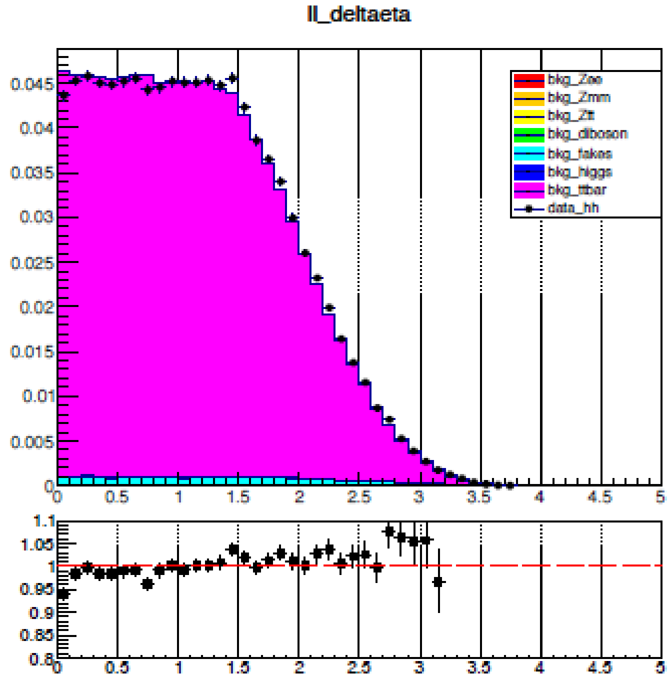
Summary

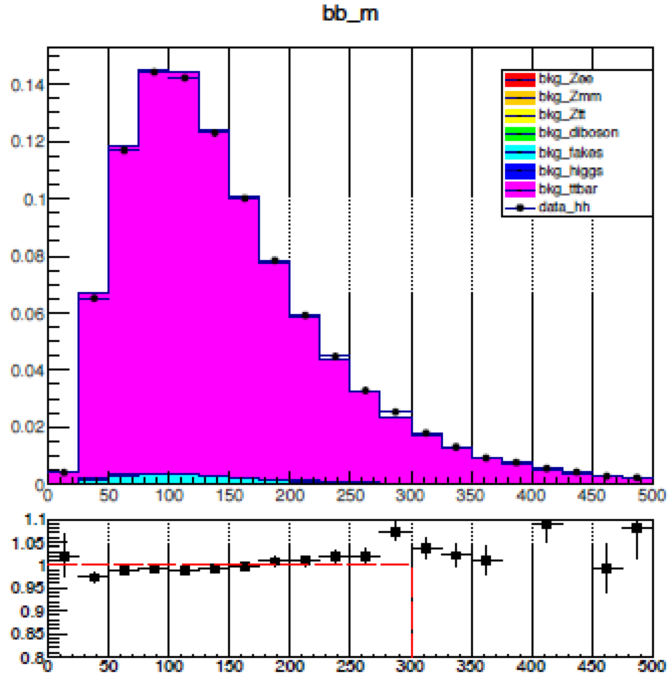
- MVA analysis performed with BDT to cross check the DNN method
- Slightly worse expected upper limits from BDT, but still quite close
- Will have further check with systematic uncertainties included

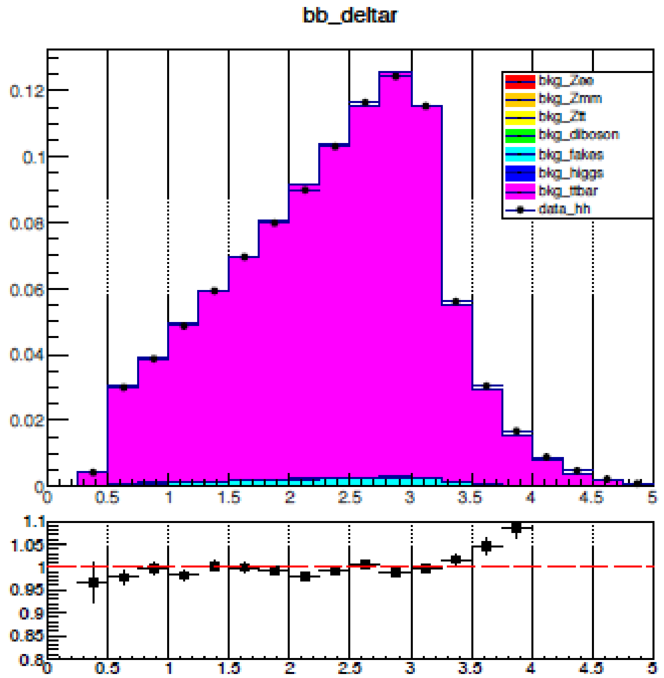
- Would like to document to supporting note

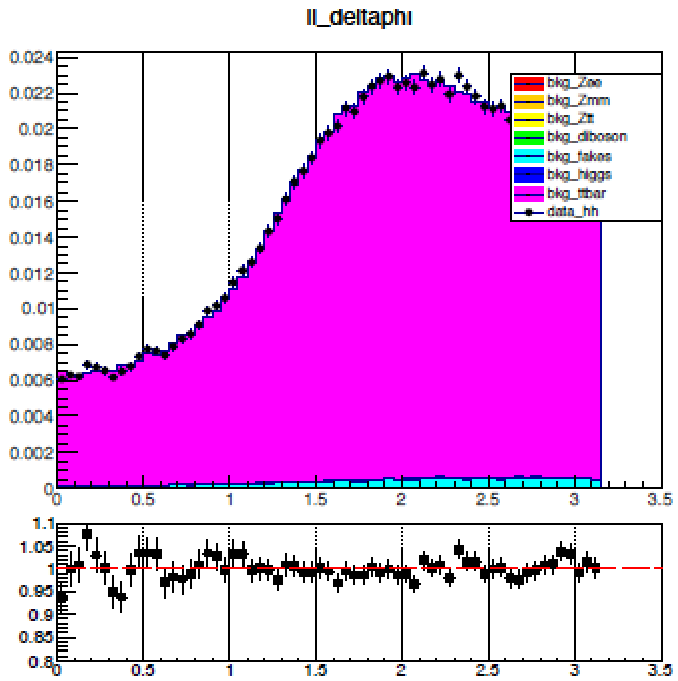
Backup: Kinematic distributions

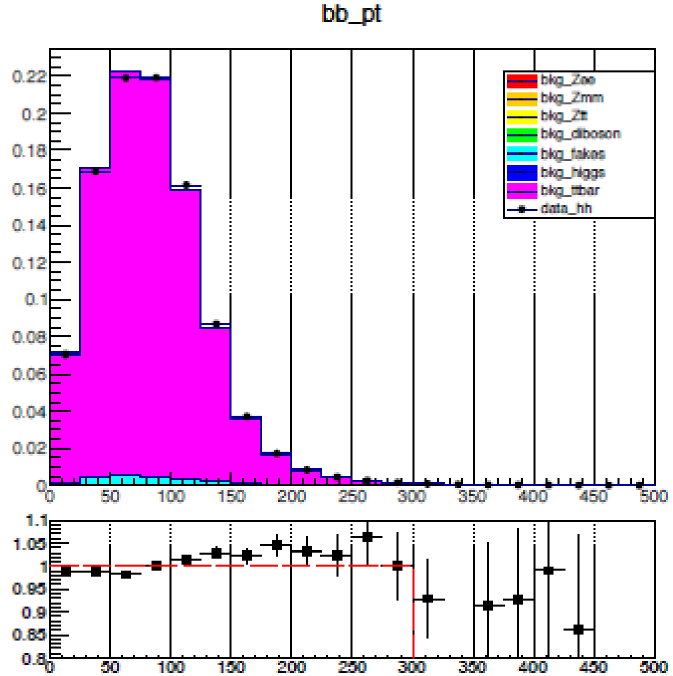

bjet_0_pt
bjet_1_pt
ll_pt
bjet_0_eta
bjet_1_eta
ll_deltar
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh

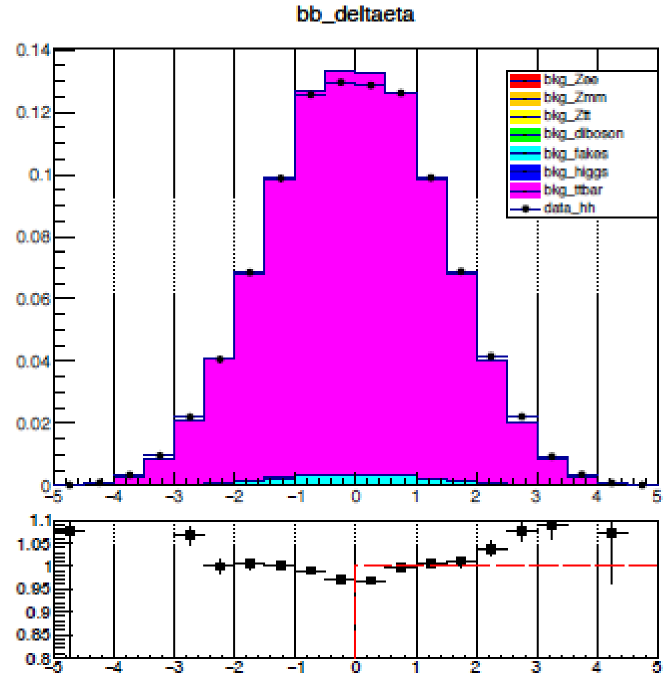

ll_deltaeta
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh

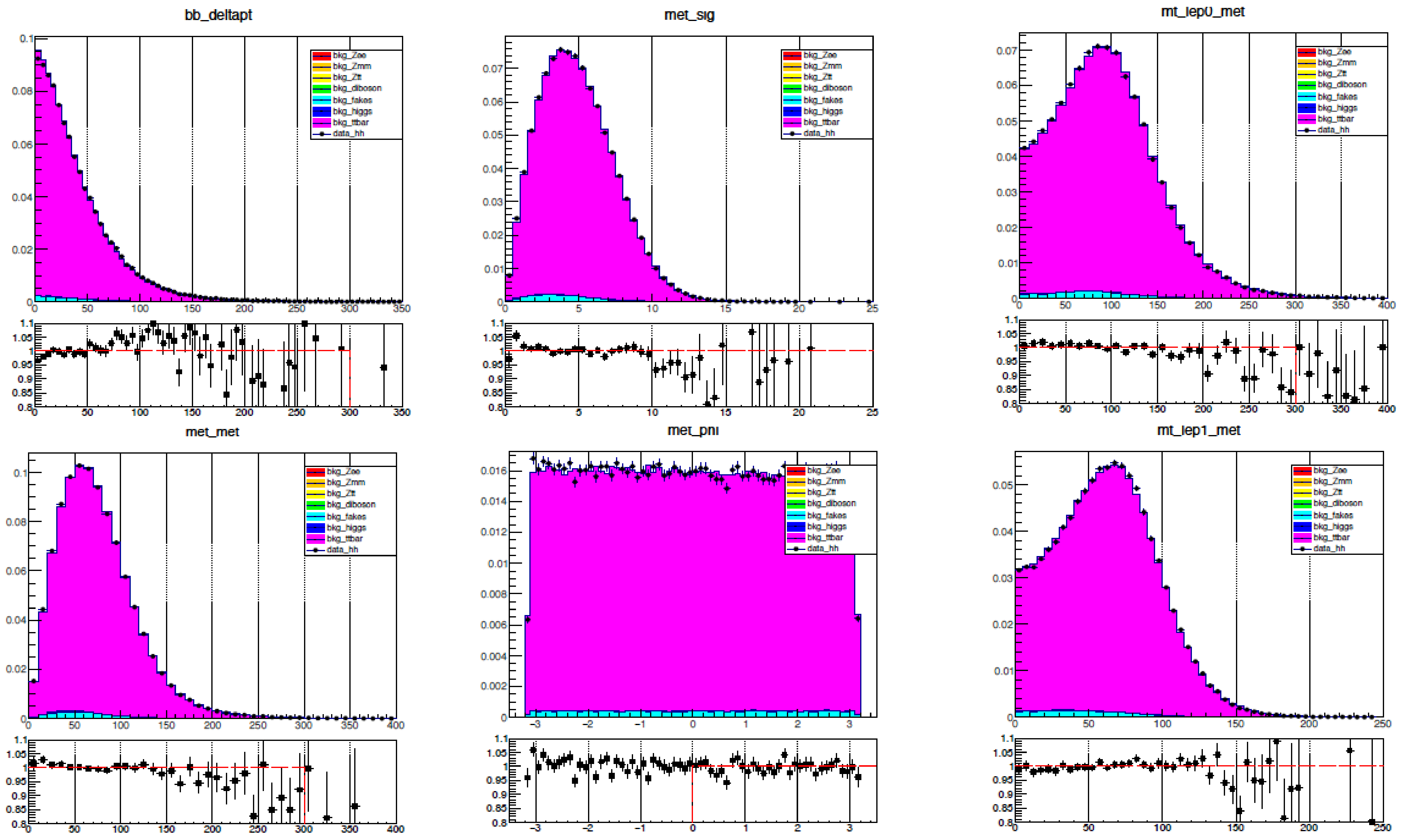

bb_m
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh

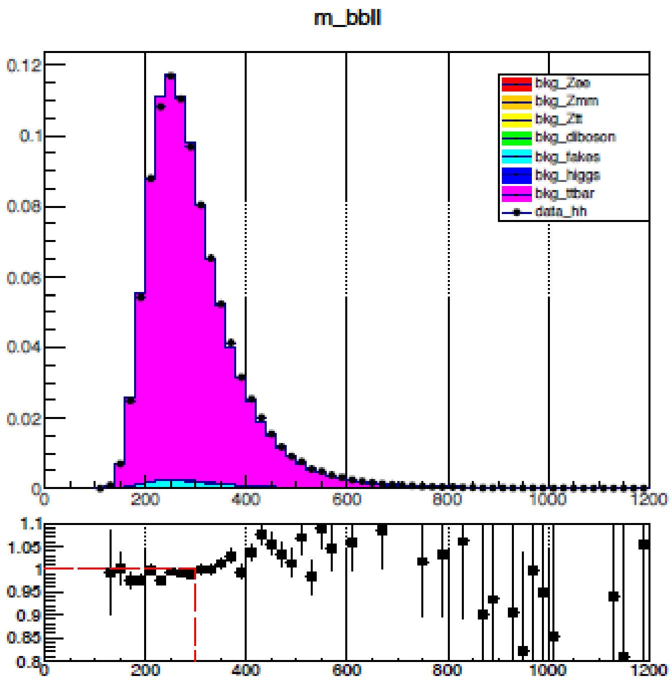

bb_deltar
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh

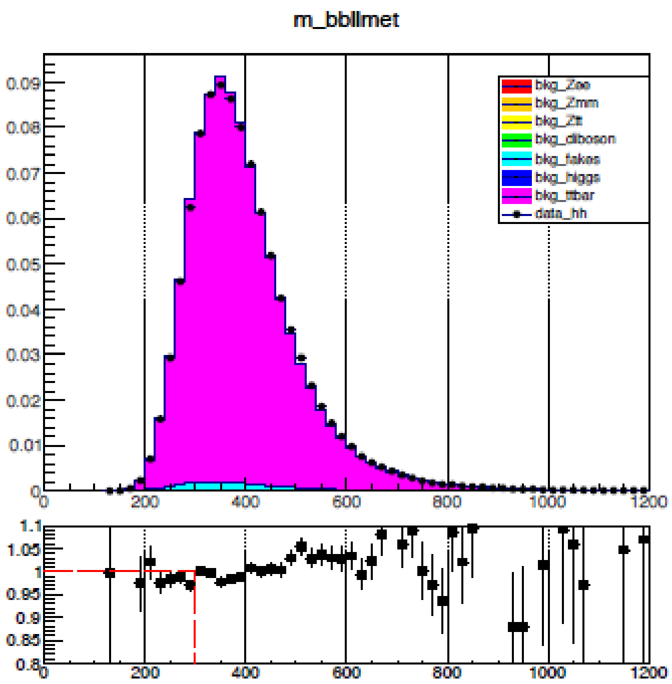

ll_deltaphi
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh


bb_pt
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh


bb_deltaeta
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh


bb_deltapt
met_sig
mt_lep0_met
met_met
met_phi
mt_lep1_met
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh


m_bbll
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh


m_bbllmet
bkg_Zee
bkg_Zmm
bkg_Ztt
bkg_diboson
bkg_fakes
bkg_higgs
bkg_ttbar
data_hh