Compressed EWK study(ISRC1N2)

Chengxin Liao

liaocx@ihep.ac.cn

Mar, Wed 19, 2025

Tasklist

- ~~Obtain (120,90), (140,90) in ML result~~

- ~~Significance map in HH&LH channel~~

- Bkg estimation(Top CR&VR on going)

- BSc thesis: https://www.overleaf.com/project/674e7119837a2580151a0868

SR definition

Pre-Selection

- lep-had channel: nTaus $\geq$ 1; nLeps $\geq$ 1
- had-had channel: nTaus $\geq$ 2; nLeps = 0
- MET $\geq$ 200; pass MET trigger
- $1 \leq nBaseJet \leq 8$
- b-veto
- OS

SR for HH channel

hyper parameter: NTrees=500, learning rate=0.05, max depth=12, MinNodeSize=1%(default)

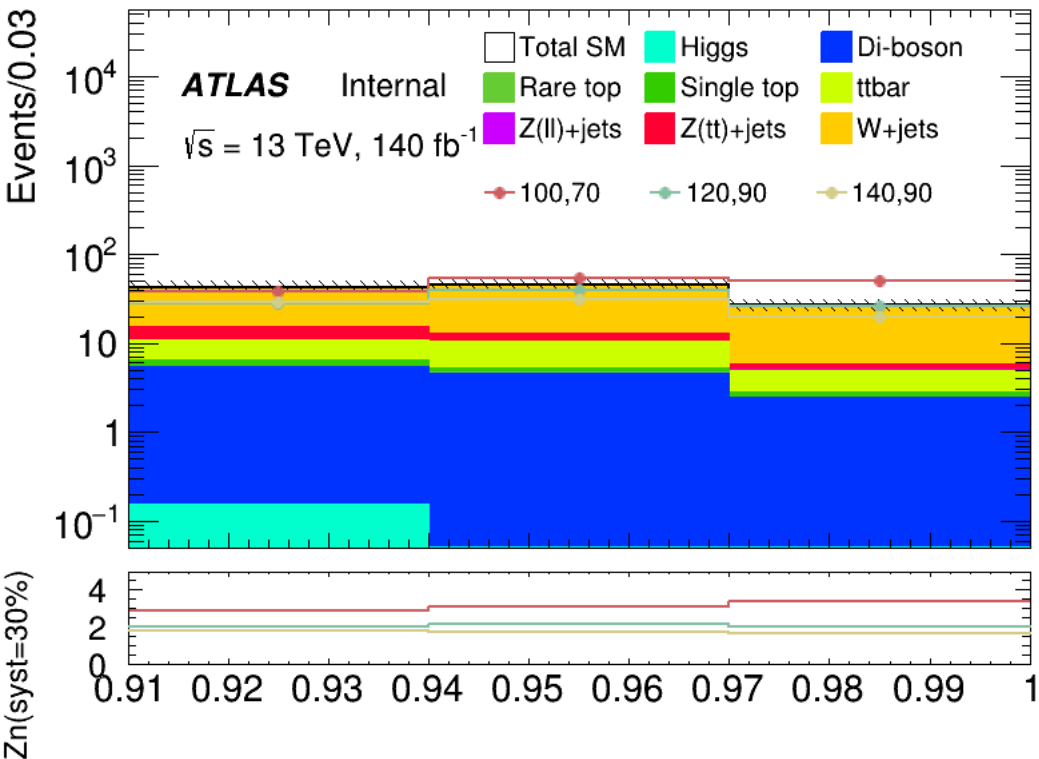**Pre-Selection + BDT score $\geq$ 0.91**

SR for LH channel

hyper parameter: NTrees=400, learning rate=0.01, max depth=10, MinNodeSize=1%(default)
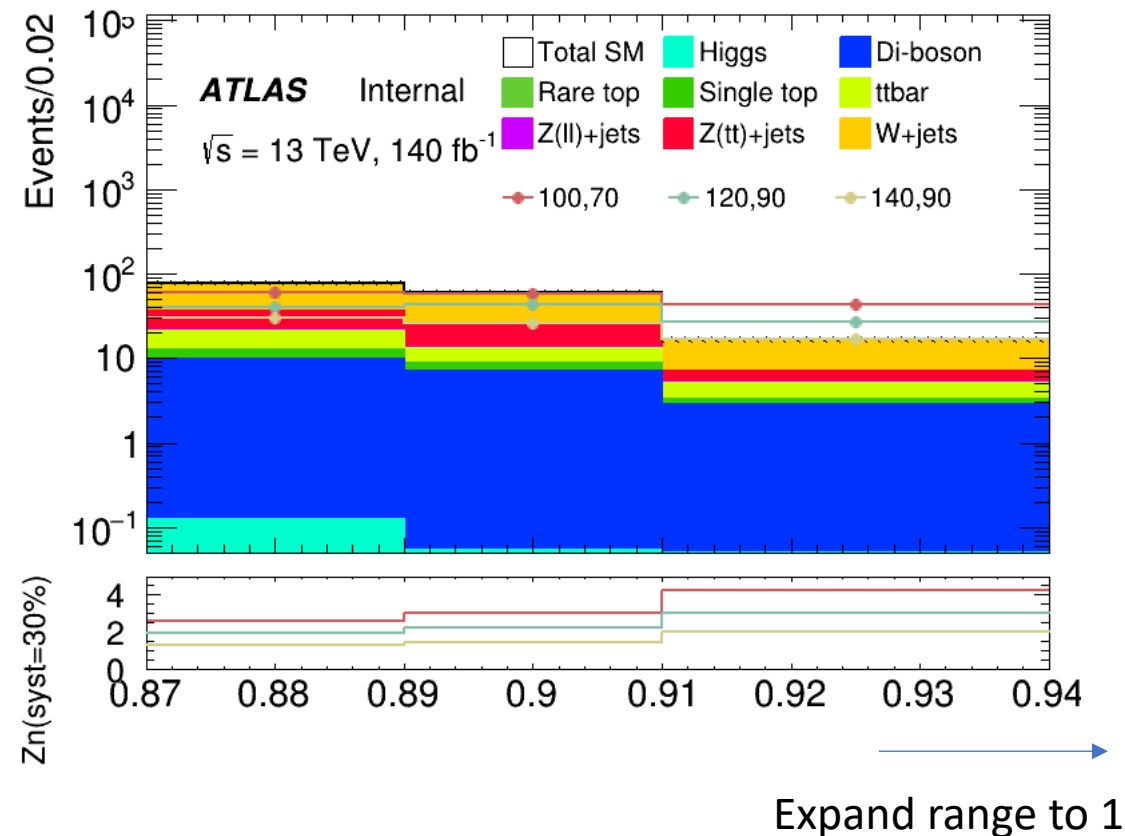
Pre-Selection + BDT score $\geq$ 0.87

Cut off some of the background tail without cut the signal but I will remove the upper cut later
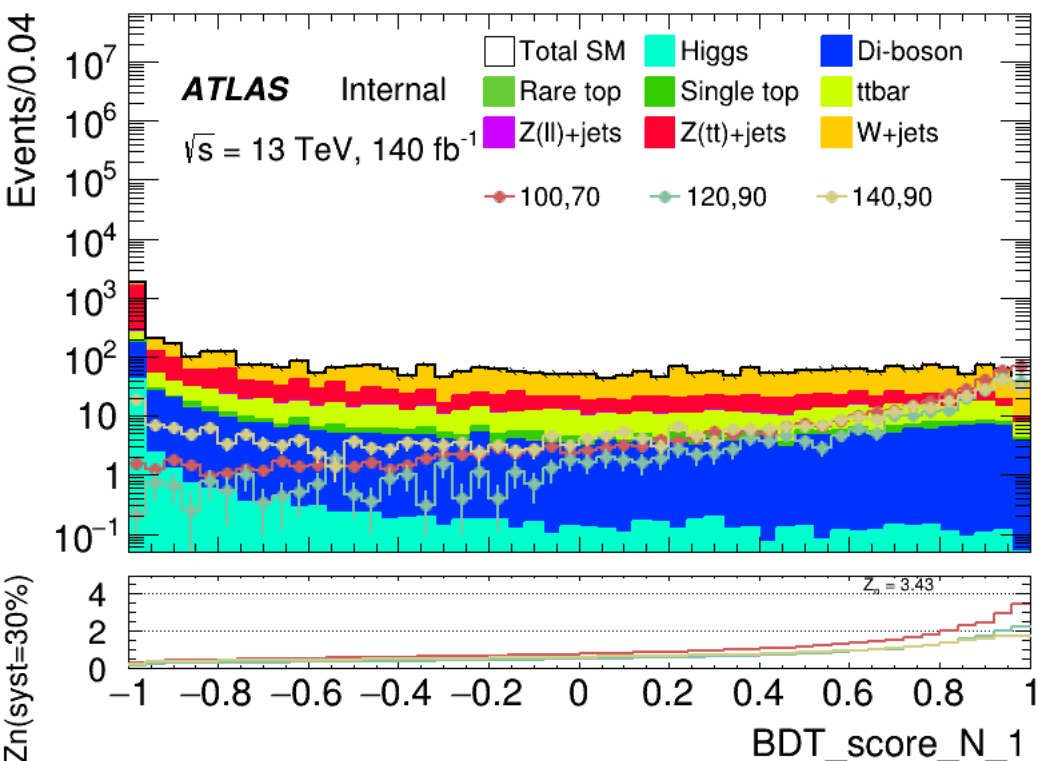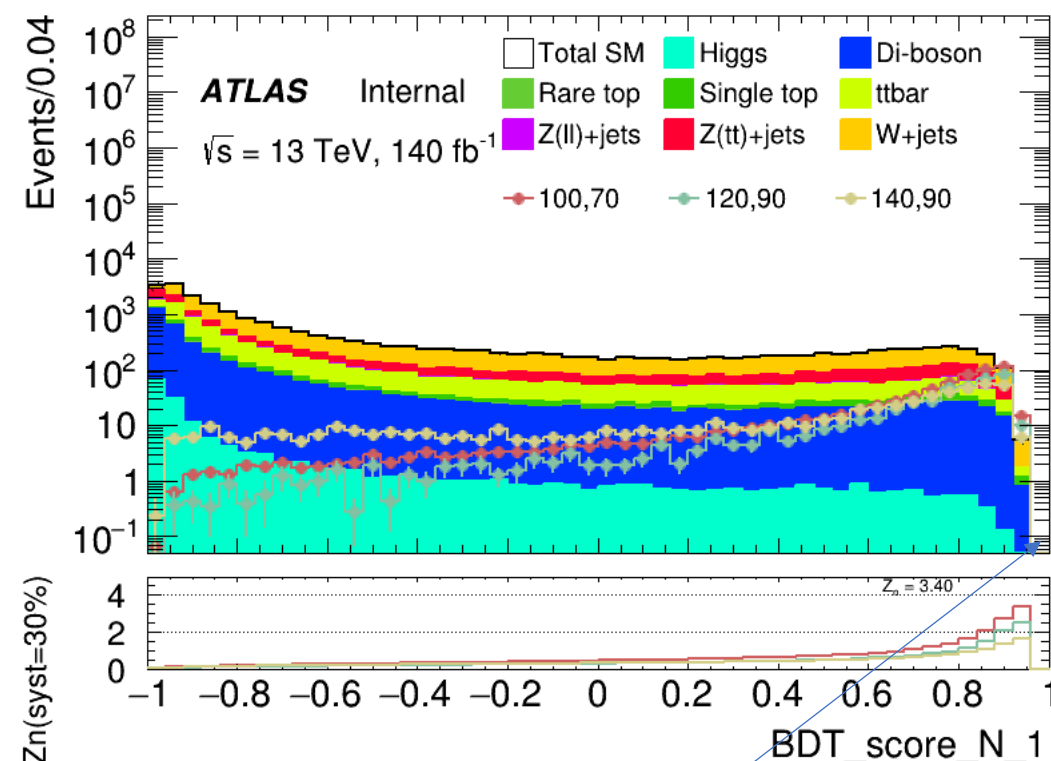
Chengxin Liao

HH channel

LH channel

Expand range to 1

The trend in these three signal are close

HH channel

LH channel

Merge the empty bin into last bin
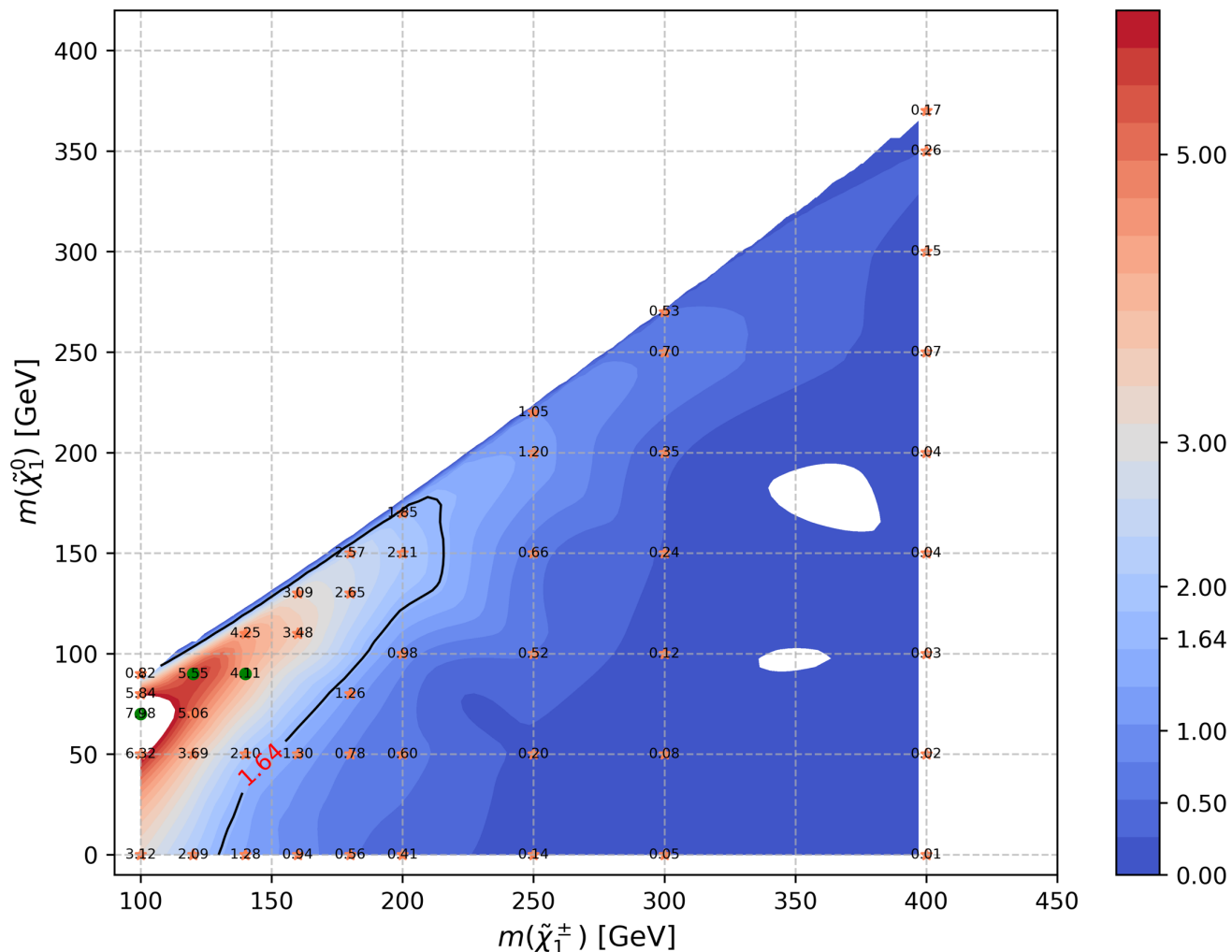
Chengxin Liao

IHEP SUSY Meeting

Cubic method to interpolation
HH channel
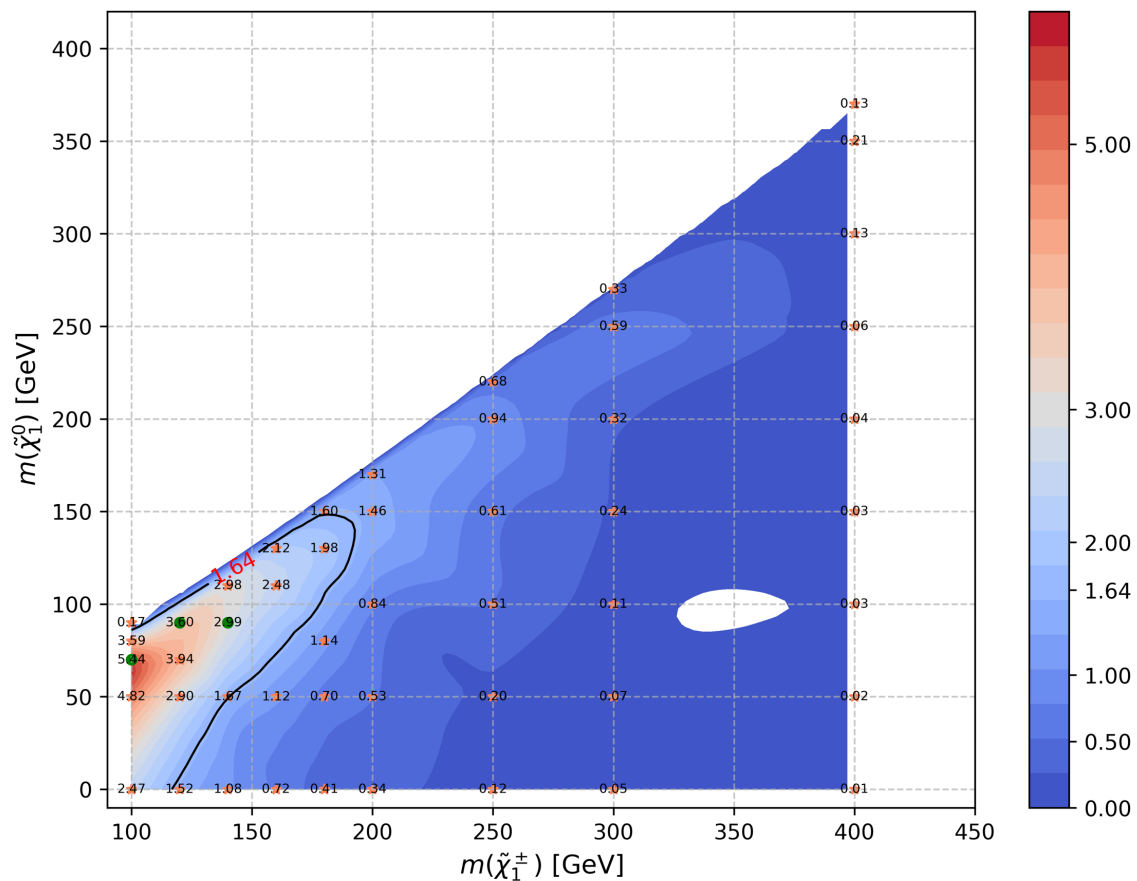
Cubic method to interpolation
LH channel
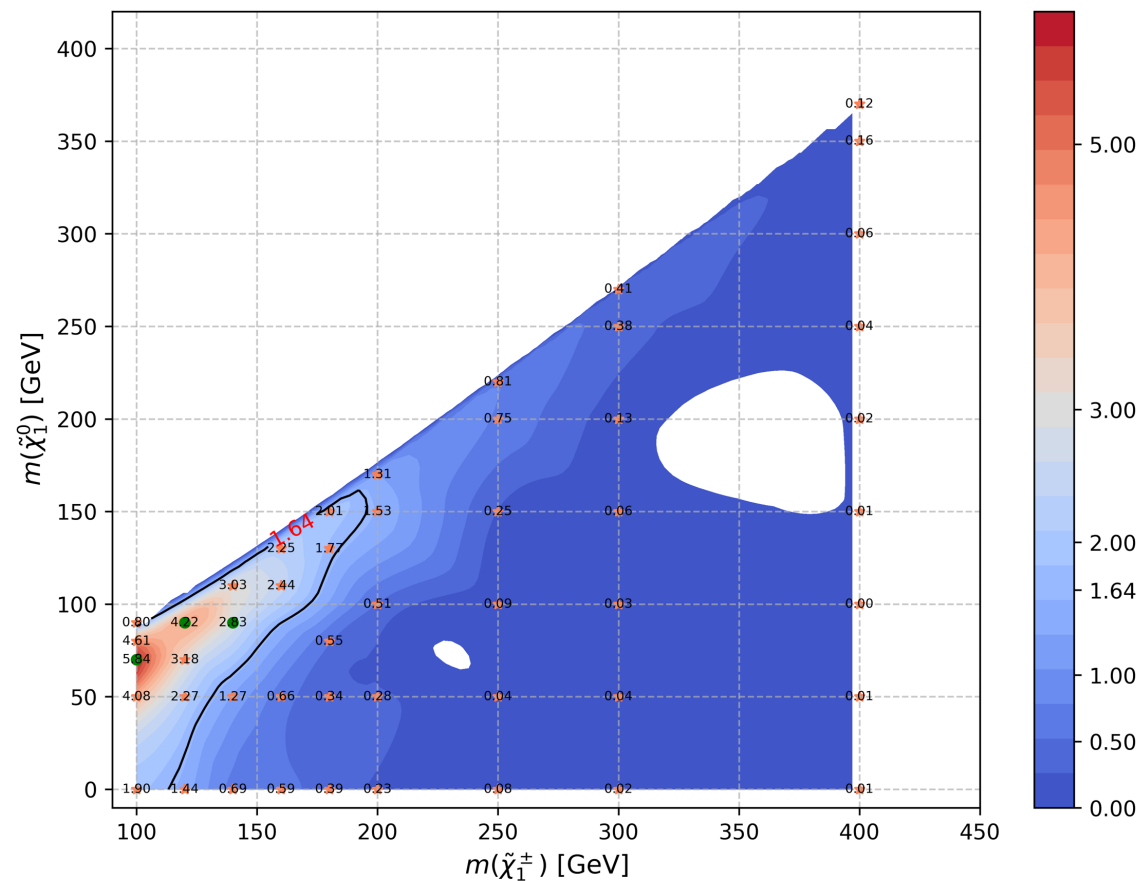
Cubic method to interpolation

$$Zn\_combine = \sqrt{Zn_{HH}^2 + Zn_{LH}^2}$$

TODO:
1. find a better way to interpolation

2. seems 1.64 can't cover sig with delta_mass = 10
maybe I should include more signal here

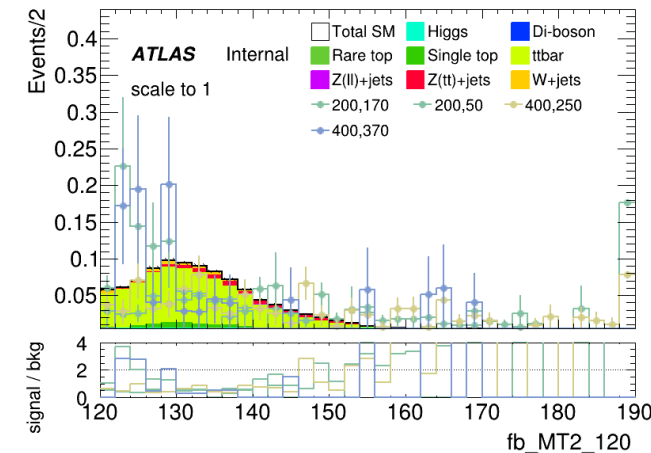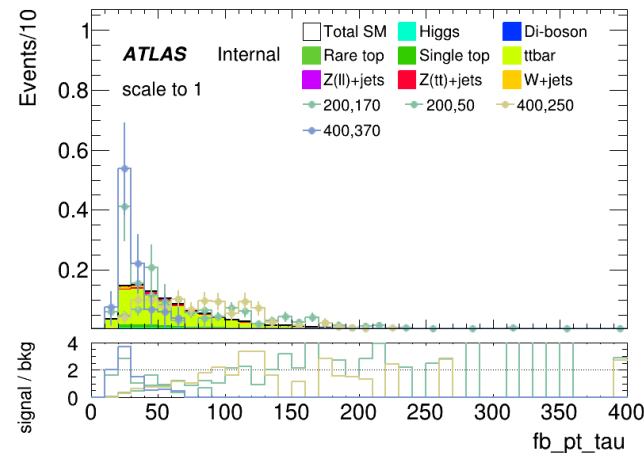## Top CR & VR

HH channel

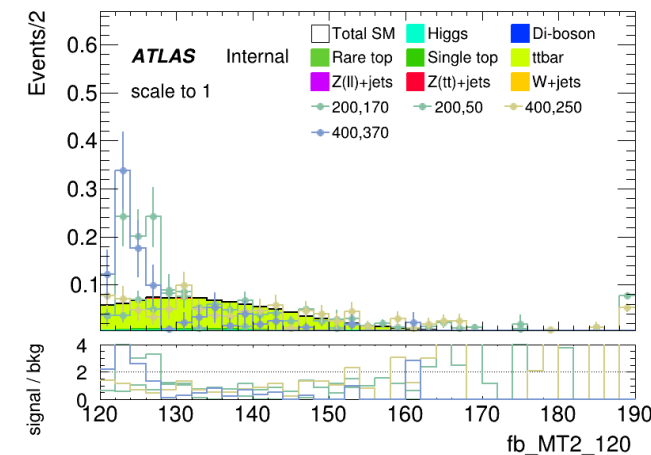### LH channel

Selection

- lep-had channel:nTaus≥1;nLeps≥1
- had-had channel:nTaus≥2;nLeps=0
- MET≥ 200; pass MET trigger
- $1 \leq nBaseJet \leq 8$
- OS
- $NbJets \geq 1$

Chengxin Liao

IHEP SUSY Meeting
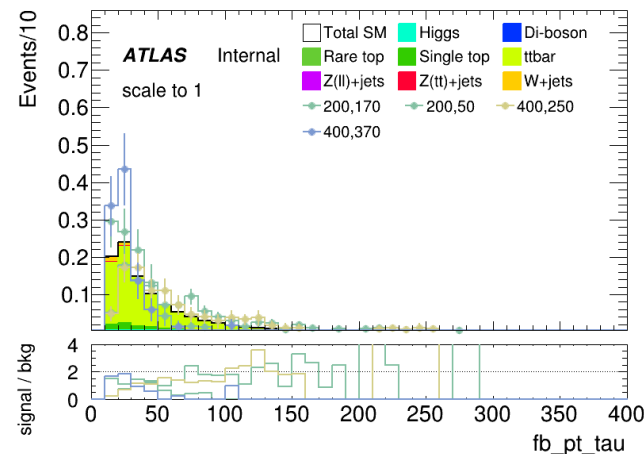
Backup

Hyperparameters optimization(LH)

Input(LH-Channel):

Sample:

Sig: ISRC1N2(mass_C1 = 100GeV, mass_N2 = 70GeV)->21225 entries

Bkg: 1703476 entries

All input data(C1N2_100_70 and Bkg) already passed pre-selection

Strategy:

method: BDTG

Separate sig(bkg) into five folders, one for test, the other three for train, and last one for validation set,then traverse all possibilities.

```
Signal     -- training events          : 12735
Signal     -- testing events           : 4245
Signal     -- training and testing events: 16980
Background -- training events          : 1022092
Background -- testing events           : 340692
Background -- training and testing events: 1362784
```

Pre-Selection

$lep\text{-}had\ channel\text{:}\ nTaus \geq 1, nLeps \geq 1$
$pass\ MET\ trigger;\ MET \geq 200$
$1 \leq nBaseJet \leq 8$
$b - Veto$
OS

Chengxin Liao

IHEP SUSY Meeting

Hyperparameters optimization(LH)

Variables(30):

Obj kinematics
nBase_Jet
mt_lep
e_lep(energy of tau2)

Angular correlations
dPhitt
dRtt
dRt1x
dPhiMin_xj
dPhiMax_tj

Event kinematics
Mll(Invariant Mass of tau1 and tau2)
METsig
MT2_50
Mwh(Invariant Mass of tau1 and MET)
Mwl(Invariant Mass of tau2 and MET)
MCT(Transverse Mass Squared)
Proj_j(Projection of pt jet on zeta)
Proj_tt(Projection of tau1+tau2 on zeta)
mtx_tau
Mtx_lep

ht_tau
mt_quad_sum
mt_sum
frac_MET_tau1
frac_MET_tau2
frac_MET_tt
frac_MET_sqrtHT_40
frac_jet_tau1
frac_jet_tau2
frac_jet_tt
MT_taumin
pt_Vframe

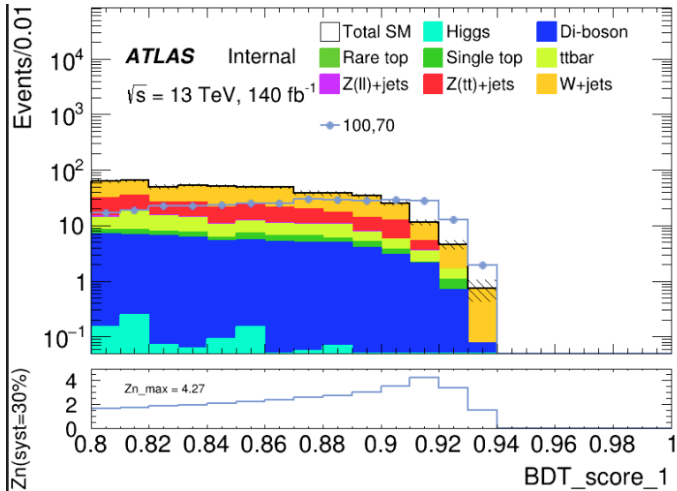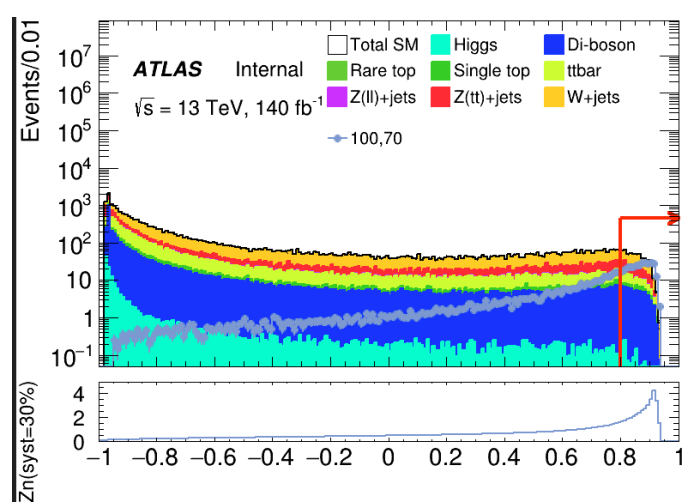**High importance at shiyi's feature**

Note:

zeta is bisector direction of tau1 and tau2[PhyUtils::bisector(tau1, tau2)]

Chengxin Liao
IHEP SUSY Meeting
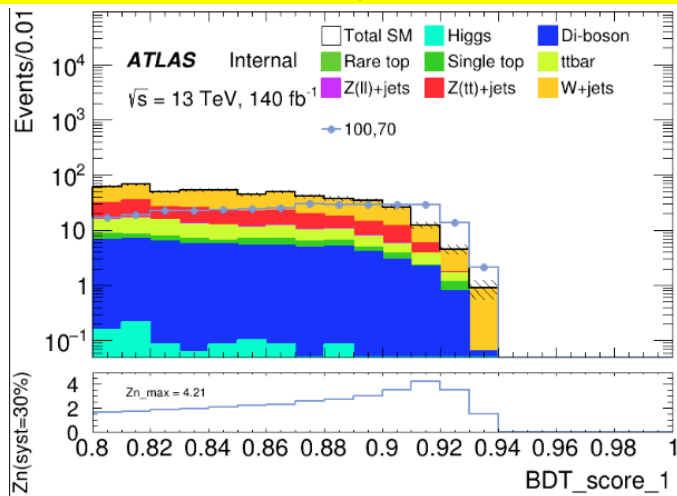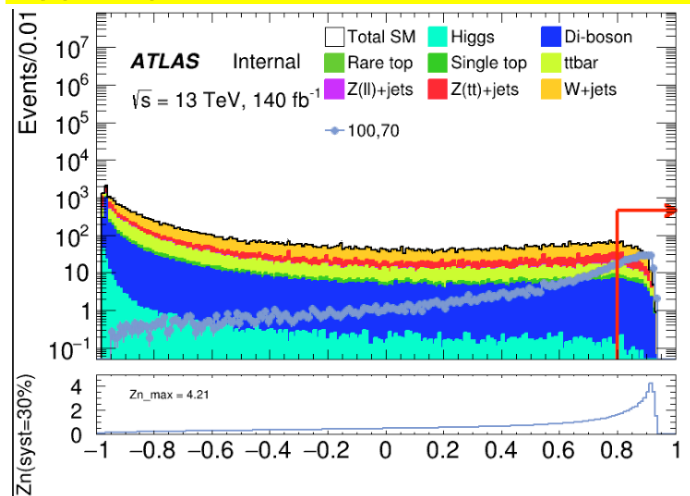
hyper parameter: NTrees=400, learning rate=0.01, max depth=10, MinNodeSize=1%(default)

Cut at BDT_score = 0.8

hyper parameter: NTrees=400, learning rate=0.01, max depth=12, MinNodeSize=1%(default)

Input(HH-Channel):

Sample:

Sig: ISRC1N2(mass_C1 = 100GeV, mass_N2 = 70GeV)->12180 entries

Bkg: 513850 entries

All input data(C1N2_100_70 and Bkg) already passed pre-selection

Strategy:

method: BDTG

Separate sig(bkg) into five folders, one for test, the other three for train, and last one for validation set,then traverse all possibilities.

```
Number of training and testing events
-----------------------------------------------------
Signal      -- training events              : 7311
Signal      -- testing events               : 2436
Signal      -- training and testing events: 9747
Background -- training events               : 308329
Background -- testing events                : 102770
Background -- training and testing events: 411099
```

Pre-Selection

$had\text{-}had\ \ channel: nTaus \geq 2, nLeps = 0$

$pass\ MET\ trigger; MET \geq 200$

$1 \leq nBaseJet \leq 8$

$b - Veto$

OS

Chengxin Liao

Hyperparameters optimization(HH)

Variables(26): <mark>Obj kinematics</mark>   <mark>Event kinematics</mark>
Pt_tt   Mll(Invariant Mass of tau1 and tau2)
                                              MIA
<mark>Angular correlations</mark>   MT2_150
dPhit1x   MET_Tau
dEtatt   Proj_tt
dPhiMax_xt   MstauA
dPhiztt   MCT
dPhitt   frac_MET_tt
dPhizxe   frac_MET_tau1
dPhiMin_xt   frac_MET_MeffInc_40
dPhit2x   frac_MET_Meff
dPhiMin_tj1
dRt2x
dRMax_xt
dRMin_tj   <mark>These vars are selected based on the importance</mark>
dRtt
sum_cos_dphi

Chengxin Liao   IHEP SUSY Meeting
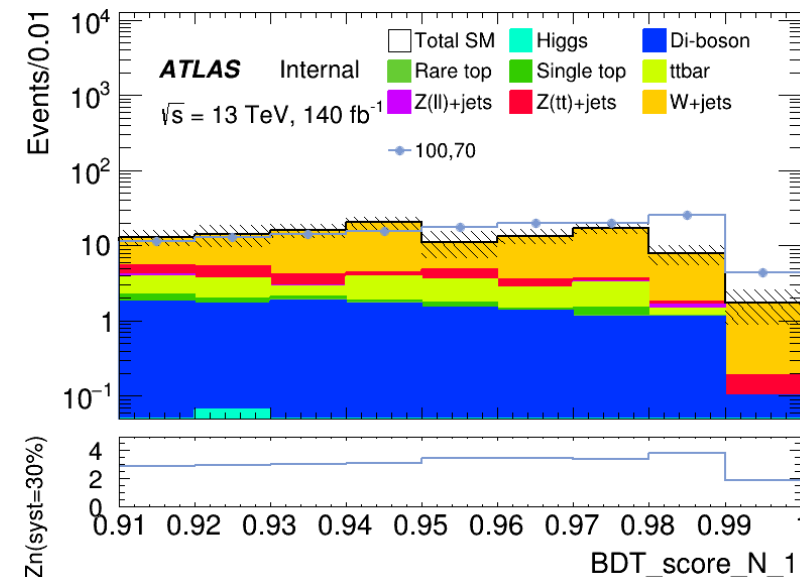
hyper parameter: NTrees=500, learning rate=0.05, max depth=12, MinNodeSize=1%(default)

Apply a rough cut at 0.80 to check the distribution   **It has a wider peak than LH signal region**
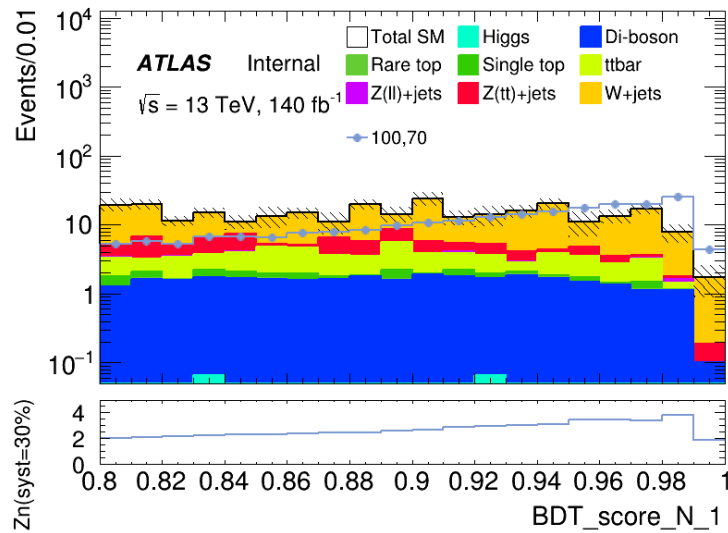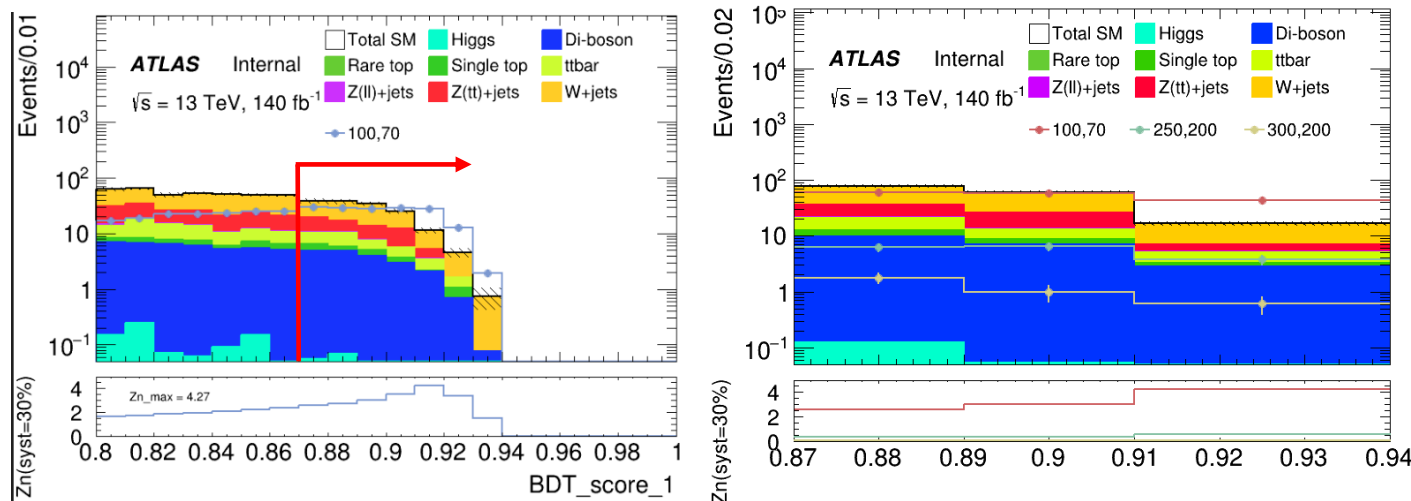

Precise cut at 0.91 to define signal region

Rebin to: [0.91, 0.94, 0.97, 1.00]

Performance of Model(LH)

Apply BDT score cut at 0.87

Root of square sum of Zn of each bin: 5.8479

Rebin to: [0.87, 0.89, 0.91, 0.94]

bin	max Zn	C1N2ISR (100,70)	bkg	Higgs	OtherTop	SingleTop	TopPair	VV	Wjets	Zlljets	Zttjets
(0.87-0.89)	2.59868	59.238+-1.484	76.648+-5.530(7.21%)	0.126+-0.034	0.033+-0.021	2.619+-0.548	8.311+-1.141	9.569+-0.520	39.971+-5.074	0.760+-0.130	15.259+-1.718
(0.89-0.91)	3.03656	57.663+-1.447	59.803+-3.946(6.59%)	0.053+-0.020	0.078+-0.030	1.761+-0.420	4.401+-0.823	6.851+-0.399	33.586+-3.367	0.453+-0.128	12.620+-1.792
(0.91-0.94)	4.26908	42.715+-1.251	16.632+-1.683(10.11%)	0.005+-0.004	0.006+-0.004	0.450+-0.202	1.819+-0.532	2.858+-0.249	9.733+-1.536	0.039+-0.020	1.722+-0.298

Chengxin Liao   IHEP SUSY Meeting
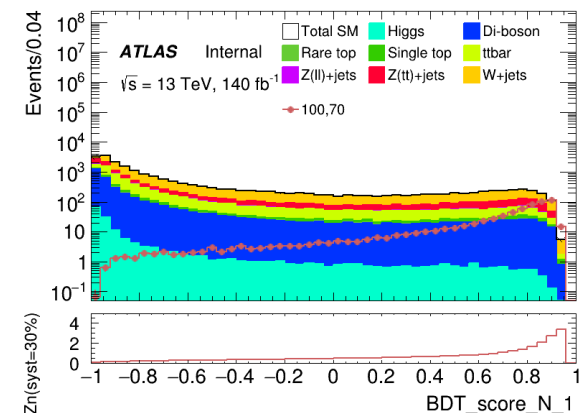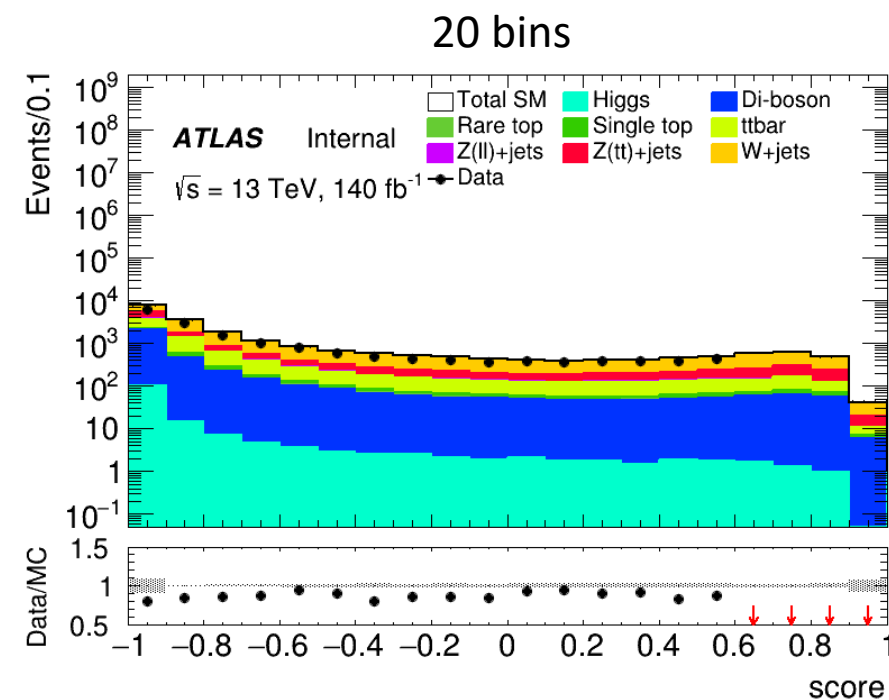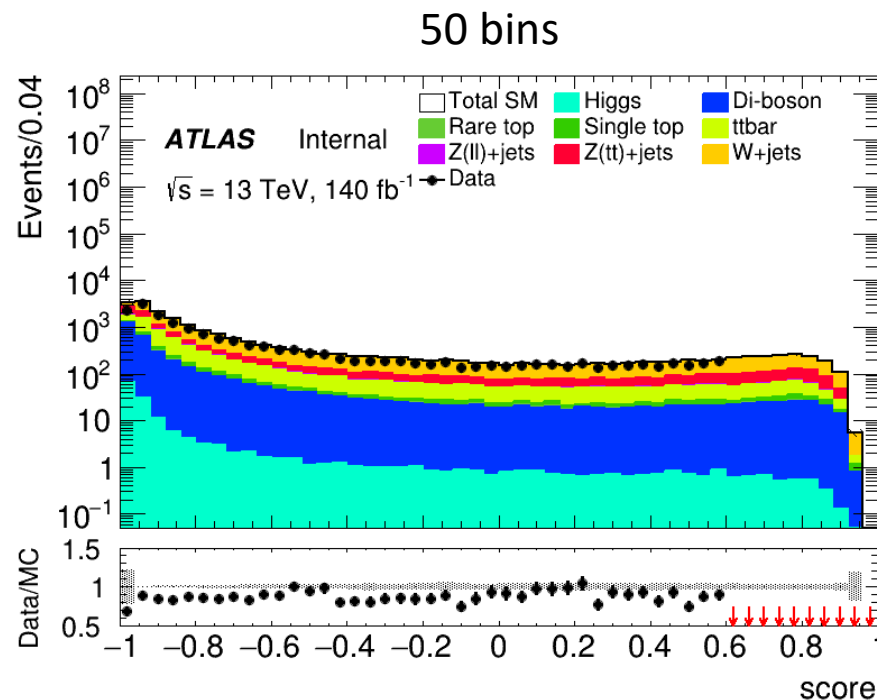
hyper parameter: NTrees=400, learning rate=0.01, max depth=10, MinNodeSize=1%(default)

BDT score distribution of Validation set

50 bins

20 bins
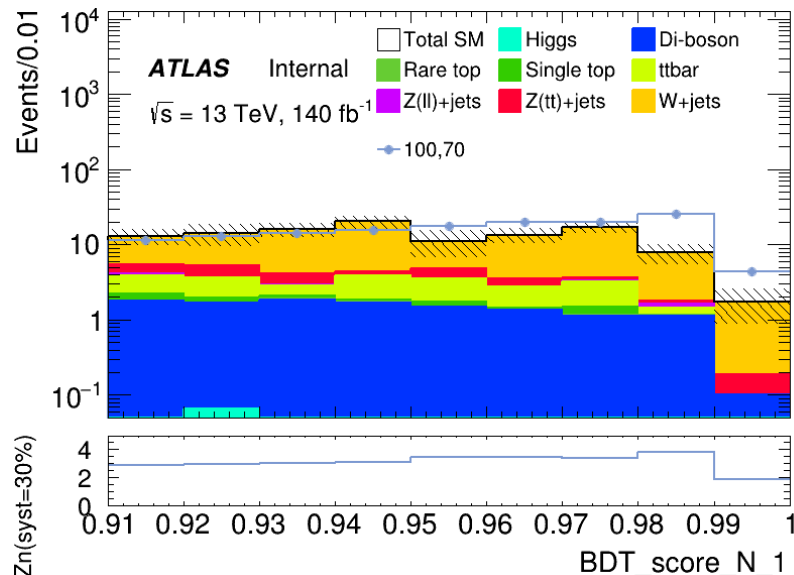

BDT score distribution of test set and data
(Blind with events with score > 0.6)

BDT score distribution of Test set
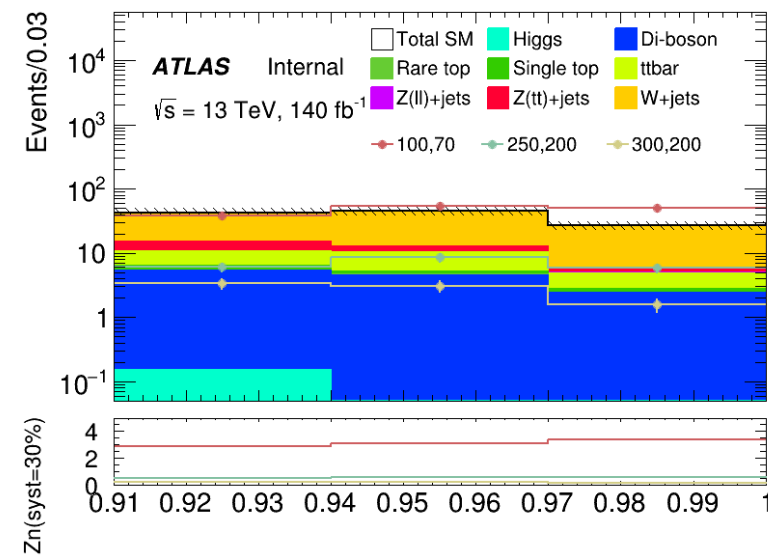
hyper parameter: NTrees=500, learning rate=0.05, max depth=12, MinNodeSize=1%(default)

Root of quadratic sum of Zn = 5.3163

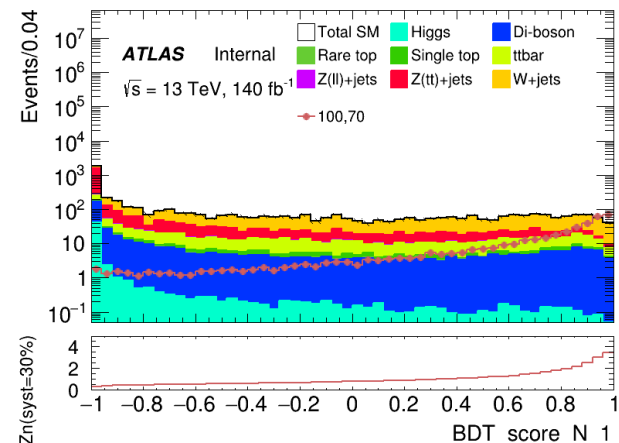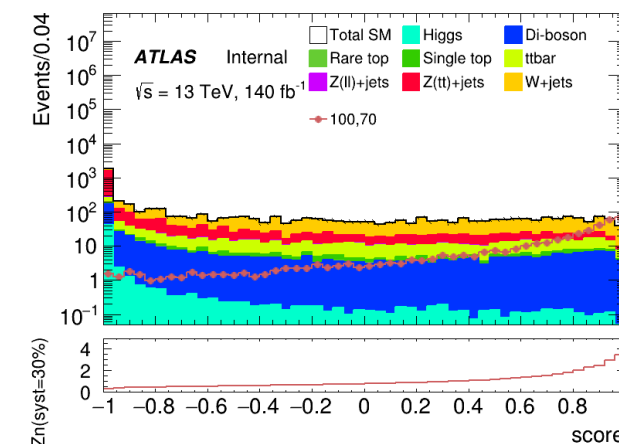Rebin to: [0.91, 0.94, 0.97, 1.00]

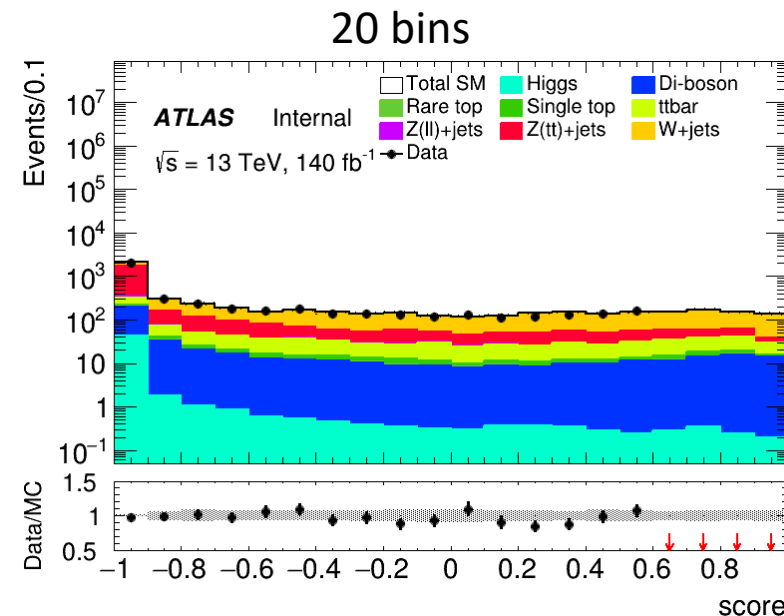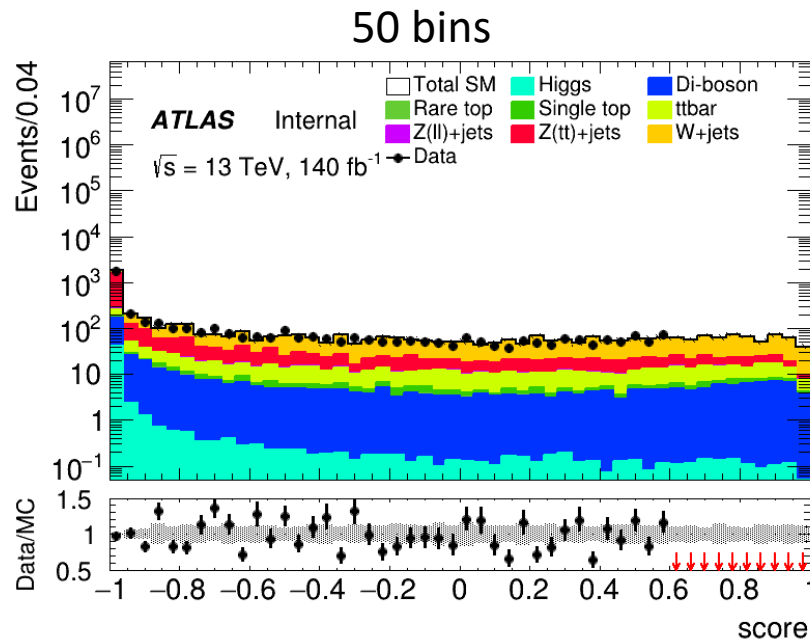| bin	Zn	C1N2ISR (100,70)	bkg	SingleTop	Zttjets	Wjets	OtherTop	VV	Zlljets	TopPair	Higgs
bin1	2.8678146	38.472+-1.192	42.986+-6.490(15.09%)	0.958+-0.303	4.090+-0.675	28.198+-6.388	0.026+-0.020	5.144+-0.289	0.265+-0.185	4.156+-0.802	0.150+-0.041
bin2	3.1193828	53.206+-1.399	44.788+-6.647(14.86%)	0.473+-0.239	2.435+-0.612	32.118+-6.547	0.033+-0.022	4.512+-0.303	0.040+-0.026	5.158+-0.888	0.019+-0.014
bin3	3.4088980	49.550+-1.350	26.640+-3.905(14.65%)	0.348+-0.184	0.600+-0.129	21.041+-3.850	0.034+-0.017	2.363+-0.189	0.196+-0.123	2.050+-0.570	0.007+-0.007

hyper parameter: NTrees=500, learning rate=0.05, max depth=12, MinNodeSize=1%(default)

50 bins

20 bins

BDT score distribution of Validation set

BDT score distribution of Test set

BDT score distribution of test set and data
(Blind with events with score > 0.6)