

Department of Physics, Shandong University

Compressed EWK study(ISRC1N2)

Chengxin Liao
liaocx@ihep.ac.cn

Mar, Fri 7, 2025



Tasklist

- ~~Summary for the SR in C1N2ISR~~
- BSc thesis: <https://www.overleaf.com/project/674e7119837a2580151a0868>
- ABCD method note/bkg estimation: any note or paper I can follow/ref?

Outline

1. Hyperparameters optimization

1.1 LH channel

1.2 HH channel

2. Performance of Model

2.1 LH channel

2.2 HH channel

3. Backup

Hyperparameters Optimization



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

```
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
```

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.

Pre-Selection

lep-had channel: $nTaus \geq 1, nLeps \geq 1$

pass MET trigger; $MET \geq 200$

$1 \leq nBaseJet \leq 8$

b - Veto

OS

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_tau_min
pt_Vframe

High importance at shiyi's feature

Note:

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

Hyperparameters optimization(LH)

Grid Search:

Ntrees: 200, 300, 400, 500

Max Depth: 6, 8, 10, 12

MinNodeSize: 1%, 2%, 3%

Learning Rate: 0.01, 0.05, 0.1

Binned significance: $Z = \sqrt{2((s_i + b_i) \log\left(1 + \frac{s_i}{b_i}\right) - s_i)}$

Show top Zn

	Model Name	Binned Significance	Max Zn	Max Zn Bin
12	400_8_1_001	15.6795	4.31391	192
79	400_10_1_001	15.6755	4.26908	192
92	400_12_1_001	15.6890	4.21178	192
77	400_10_2_001	15.3196	4.11376	191
52	500_10_1_001	15.8304	4.11162	194
13	500_12_1_001	15.8210	4.05346	194
120	400_12_1_01	16.0665	4.02939	199
113	500_10_3_001	15.3232	4.02306	192
0	300_12_1_005	16.1734	4.01739	198
139	400_12_1_005	16.2126	4.00753	199
123	500_12_3_001	15.3067	4.00343	192
118	500_12_1_01	16.0441	4.00080	199
24	500_8_1_01	15.9307	3.99007	199
133	500_8_3_001	15.3061	3.97695	192
97	400_6_3_001	15.0010	3.97216	190
26	300_10_1_01	16.0095	3.96339	199
136	300_12_1_01	16.0204	3.94916	199
128	400_8_3_001	14.9962	3.93255	190
107	200_12_1_005	16.0375	3.93002	197
45	400_12_2_001	15.2724	3.92019	191
88	400_12_3_001	14.9991	3.91396	190

Shiyi's result of LH channel

Top Sig

	hy	sig	zn
400_10_2_0.05	15.3225	3.72536	
300_11_1_0.05	15.3127	3.87694	
500_10_2_0.05	15.3099	3.60778	
400_6_1_0.05	15.3075	3.91373	
500_8_1_0.05	15.2990	3.58389	
400_8_2_0.05	15.2980	3.74427	
300_6_1_0.05	15.2929	4.09837	
500_8_2_0.05	15.2891	3.63322	
200_11_1_0.05	15.2849	3.92924	
300_11_2_0.05	15.2804	3.85617	
400_11_2_0.05	15.2780	3.68484	
300_8_1_0.05	15.2753	3.82506	
300_10_1_0.05	15.2733	3.71921	
400_11_1_0.05	15.2701	3.60863	
500_6_1_0.05	15.2593	3.84429	
200_6_1_0.1	15.2559	3.90950	
400_12_1_0.05	15.2554	3.58328	
500_10_1_0.05	15.2493	3.49410	

Top Zn

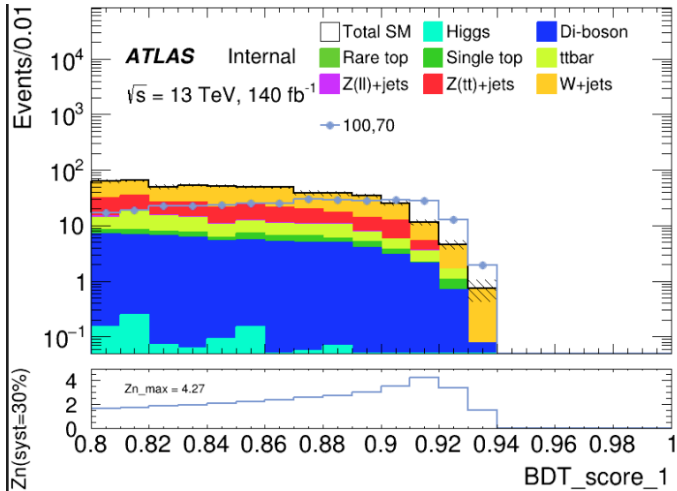
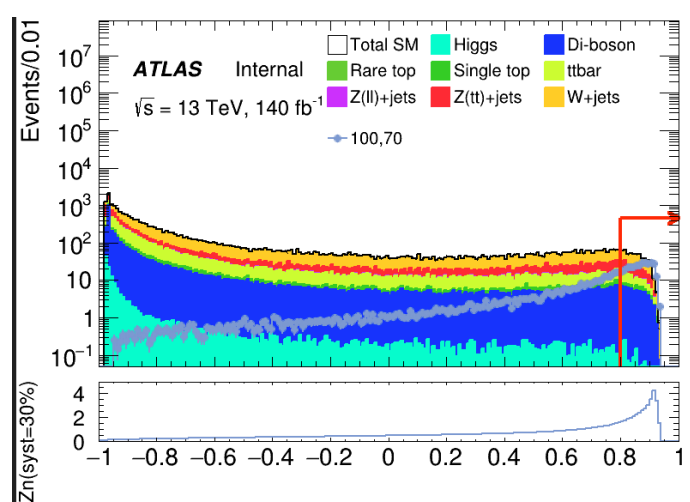
	hy	sig	zn
200_6_3_0.05	15.0164	4.29022	
200_6_1_0.05	15.0755	4.10077	
300_6_1_0.05	15.2929	4.09837	
200_10_2_0.05	15.1606	4.09228	
200_12_2_0.05	15.1803	4.04800	
200_8_2_0.05	15.0857	4.01373	
200_8_3_0.05	14.9662	4.01324	
200_6_2_0.05	14.9743	3.94396	
200_11_1_0.05	15.2849	3.92924	
300_6_2_0.05	15.1858	3.91508	
400_6_1_0.05	15.3075	3.91373	
200_6_1_0.1	15.2559	3.90950	
200_12_1_0.05	15.2279	3.90380	
400_8_1_0.01	14.6829	3.90189	
300_11_1_0.05	15.3127	3.87694	
200_8_1_0.05	15.1285	3.85623	
300_11_2_0.05	15.2804	3.85617	
500_6_1_0.05	15.2593	3.84429	

400_10_1_001, 15.6755, 4.26908, 192, 200
 400_10_1_001, 15.4762, 3.53693, 96, 100
 400_10_1_001, 15.1985, 3.40439, 49, 50
 400_10_1_001, 15.3013, 3.53693, 39, 40
 400_10_1_001, 14.8172, 3.40439, 25, 25
 400_10_1_001, 15.06, 3.53693, 20, 20
 400_10_1_001, 13.9532, 1.6563, 10, 10

400_12_1_001, 15.689, 4.21178, 192, 200
 400_12_1_001, 15.4949, 3.52564, 97, 100
 400_12_1_001, 15.2434, 3.52564, 49, 50
 400_12_1_001, 15.3089, 3.52196, 39, 40
 400_12_1_001, 14.8653, 3.52564, 25, 25
 400_12_1_001, 15.0506, 3.52196, 20, 20
 400_12_1_001, 13.9276, 1.6473, 10, 10

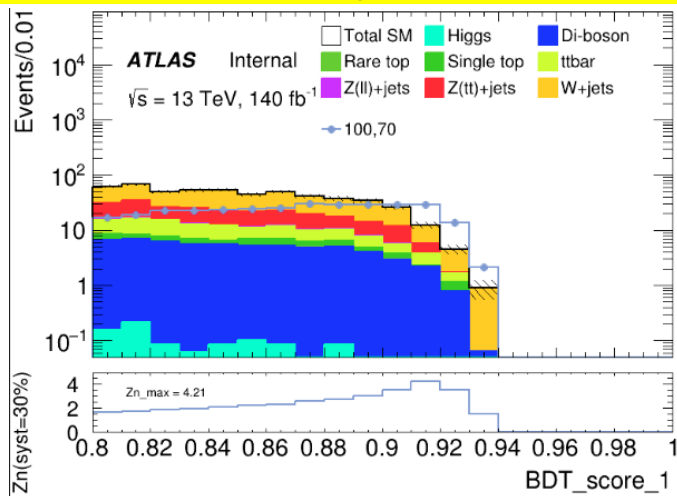
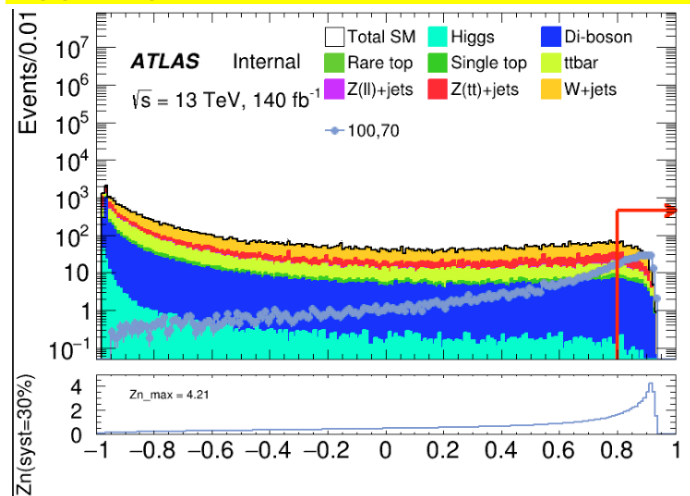
Hyperparameters optimization(LH)

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)



Hyperparameters optimization(HH)

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

pass MET trigger; $MET \geq 200$

$1 \leq nBaseJet \leq 8$

b - Veto

OS

Hyperparameters optimization(HH)

Variables(26): **Obj kinematics**

Pt_tt

Angular correlations

dPhit1x

dEtatt

dPhiMax_xt

dPhiztt

dPhitt

dPhizxe

dPhiMin_xt

dPhit2x

dPhiMin_tj1

dRt2x

dRMax_xt

dRMin_tj

dRtt

sum_cos_dphi

Event kinematics

Mll(Invariant Mass of tau1 and tau2)

MIA

MT2_150

MET_Tau

Proj_tt

MstauA

MCT

frac_MET_tt

frac_MET_tau1

frac_MET_MeffInc_40

frac_MET_Meff

These vars are selected based on the importance

Hyperparameters optimization(HH)

Grid Search:

Ntrees: 200, 300, 400, 500

Max Depth: 6, 8, 10, 12

MinNodeSize: 1%, 2%, 3%

Learning Rate: 0.01, 0.05, 0.1

Show top Zn

Model Name	Binned Significance	Max Zn	Max Zn Bin
500_12_1_005	14.2770	3.83857	199
300_10_1_01	13.9648	3.76965	198
200_6_1_01	13.9250	3.74940	198
500_6_3_01	14.2740	3.72616	199
400_10_1_01	13.9553	3.70167	199
300_6_2_01	13.9366	3.69620	199
300_10_2_01	14.0094	3.67743	199
300_8_1_01	14.0434	3.67624	198
200_8_1_01	14.1925	3.67005	198
400_12_1_005	14.1384	3.66529	199
200_6_2_01	14.2209	3.65978	199
200_6_3_01	13.7197	3.64427	199
500_10_1_01	13.8227	3.63722	198
500_8_1_01	13.8369	3.61405	198
400_10_2_01	14.2001	3.60950	199
500_6_1_005	14.0399	3.60132	197

Rebin result

Model Name	Binned Significance	Max Zn	Max Zn Bin	bin num
500_12_1_005	16.0862	3.8635	198	200
500_12_1_005	15.9967	3.62563	99	100
500_12_1_005	15.9318	3.62563	50	50
500_12_1_005	15.6612	3.07372	40	40
500_12_1_005	15.3086	2.45396	25	25
500_12_1_005	15.0825	2.20391	20	20

$$\text{Binned significance: } Z = \sqrt{2((s_i + b_i) \log\left(1 + \frac{s_i}{b_i}\right) - s_i)}$$

Compared with form result, there has a significant improvement in Zn

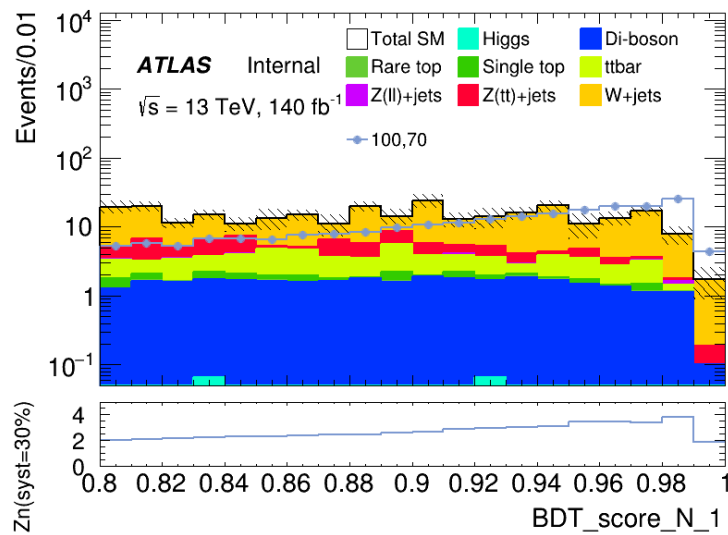
	Model Name	Binned Significance	Max Zn	Max Zn Bin
137	100_8_3_005	12.1380	3.27179	48
128	100_6_1_01	12.9663	3.24539	49
15	100_12_3_005	12.1254	3.22901	48
85	100_10_3_005	12.1150	3.20722	48
105	200_10_1_01	13.1608	3.19603	50
104	100_8_1_01	12.8853	3.18398	50
63	200_6_1_005	12.7673	3.17520	49
2	200_6_1_01	12.9052	3.17138	50
38	100_6_2_01	12.8248	3.16297	49
131	300_8_2_01	13.1256	3.16255	50
93	300_6_1_005	12.9703	3.14200	50
73	100_6_1_005	12.4457	3.14142	48
69	400_6_1_01	12.9285	3.14074	50
54	200_8_3_01	12.8685	3.13397	50
12	200_6_2_005	12.7035	3.12582	49
33	100_6_2_005	12.2453	3.11746	48
66	400_6_1_005	12.9369	3.10400	50
45	100_10_3_01	12.7388	3.10074	49
7	100_12_3_01	12.6318	3.10071	49
48	400_12_2_01	12.9393	3.09236	50
72	400_12_2_005	12.9179	3.06882	50
62	300_6_1_01	12.8501	3.06869	50
133	400_10_1_01	12.9846	3.06413	50
91	100_6_3_01	12.6291	3.06320	49
9	400_8_3_01	12.9337	3.06226	50
58	300_10_3_005	12.8854	3.05992	49
36	300_10_1_01	12.9906	3.05755	50
89	100_10_2_01	12.8406	3.05210	49
43	200_8_2_01	13.2400	3.04952	50
19	400_6_2_005	12.8263	3.04206	50

Hyperparameters optimization(HH)

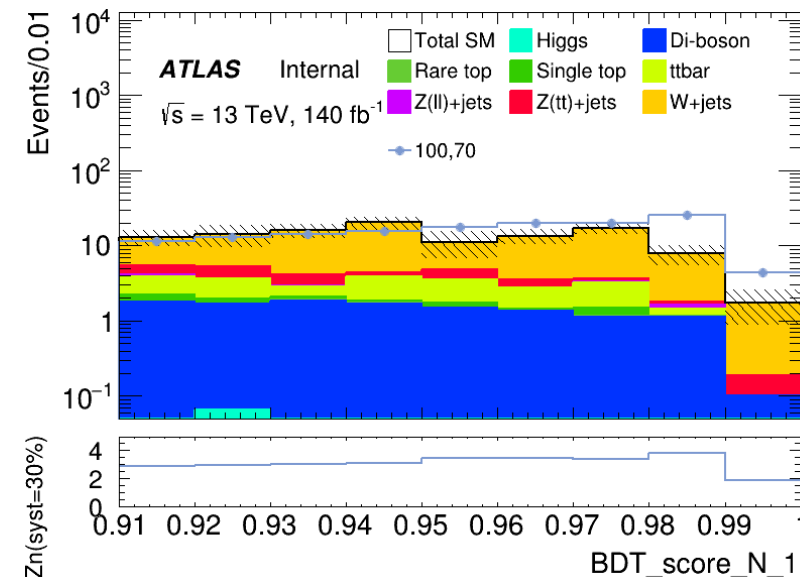
hyper parameter: NTrees=500, learning rate=0.05, max depth=12, MinNodeSize=1%(default)

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



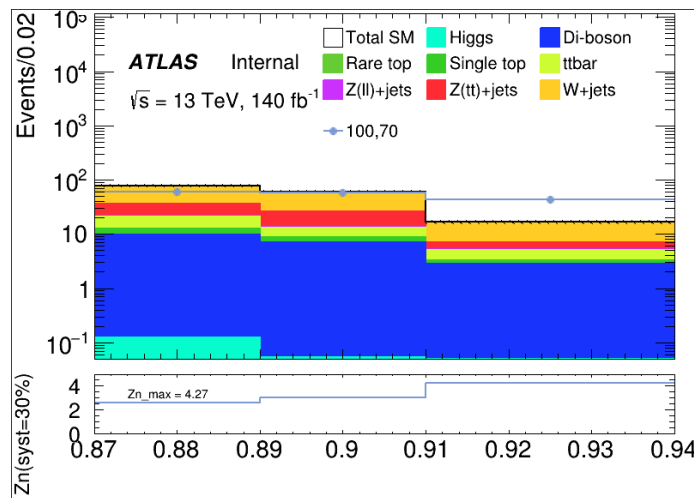
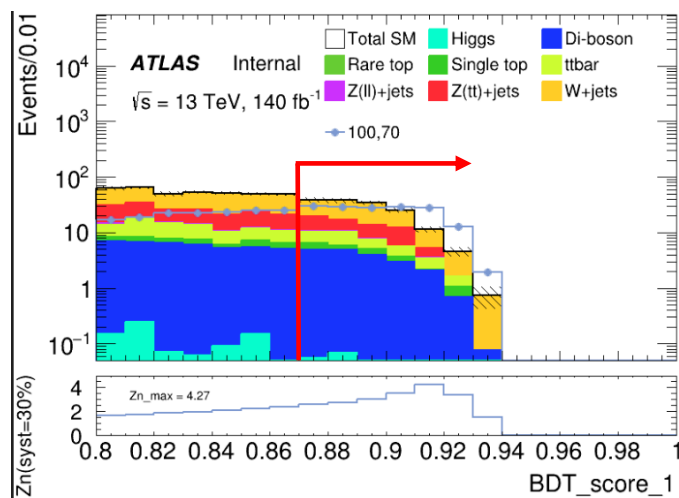
Rebin to: [0.91, 0.94, 0.97, 1.00]

Performance of Model



Performance of Model(LH)

hyper parameter: NTrees=400, learning rate=0.01, max depth=10, MinNodeSize=1%(default)



Apply BDT score cut at 0.87

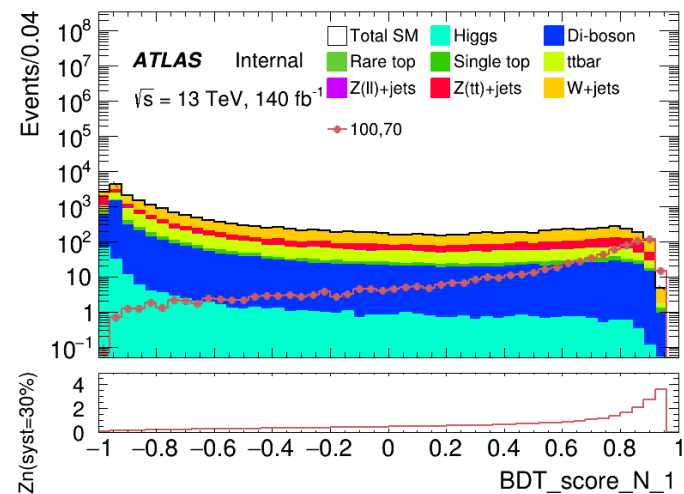
Root of square sum of Z_n of each bin: 5.8479

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

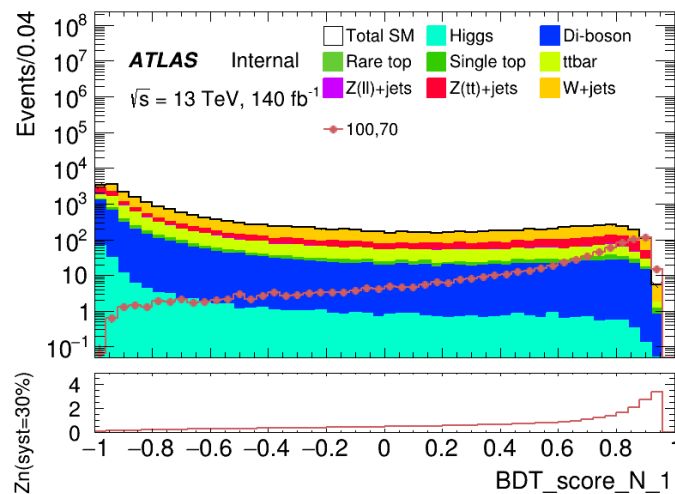
bin	max Z_n	C1N2ISR (100,70)	bkg	Higgs	OtherTop	SingleTop	TopPair	VV	Wjets	Zlljets	Zttjets
(0.87-0.89)	2.59868	59.238+-1.484	76.648+-5.530(7.21%)	0.126+-0.034	0.033+-0.021	2.619+-0.548	8.311+-1.141	9.569+-0.520	39.971+-5.074	0.760+-0.130	15.259+-1.718
(0.89-0.91)	3.03656	57.663+-1.447	59.803+-3.946(6.59%)	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

Performance of Model(LH)

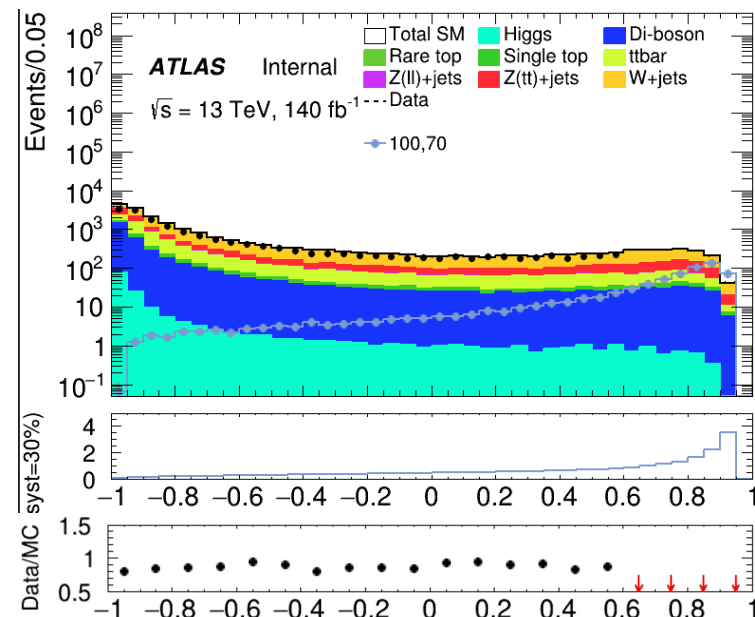
hyper parameter: NTrees=400, learning rate=0.01, max depth=10, MinNodeSize=1%(default)



BDT score distribution of Validation set



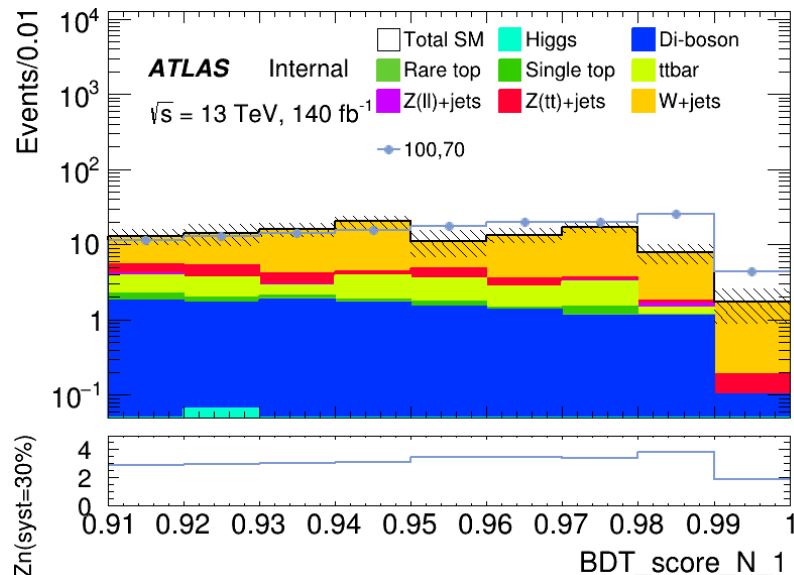
BDT score distribution of Test set



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

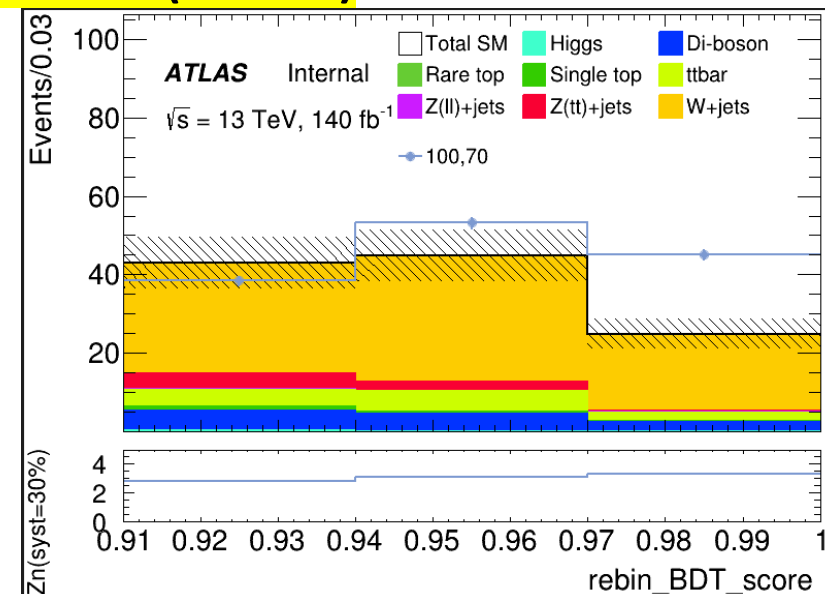
Performance of Model(HH)

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



Root of quadratic sum of $Z_n = 5.3163$

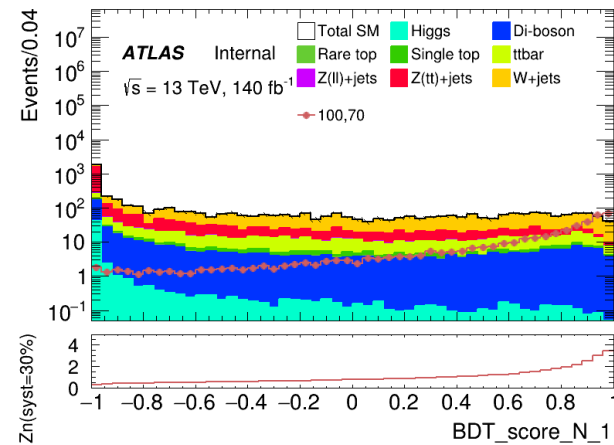
Rebin to: [0.91, 0.94, 0.97, 1.00]



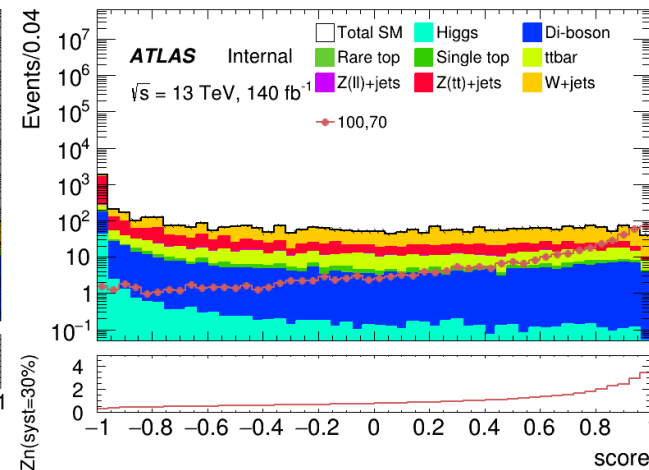
bin	Zn	C1N2ISR (100,70)	bkg	SingleTop	Zttjets	Wjets	OtherTop	VV	Zlljets	TopPair	Higgs
bin1	2.8678146	38.472+- 1.192	42.986+- 6.490(15.09%)	0.958+- 0.303	4.090+- 0.675	28.198+- 6.388	0.026+- 0.020	5.144+- 0.289	0.265+- 0.185	4.156+- 0.802	0.150+- 0.041
bin2	3.1193828	53.206+- 1.399	44.788+- 6.647(14.86%)	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

Performance of Model(LH)

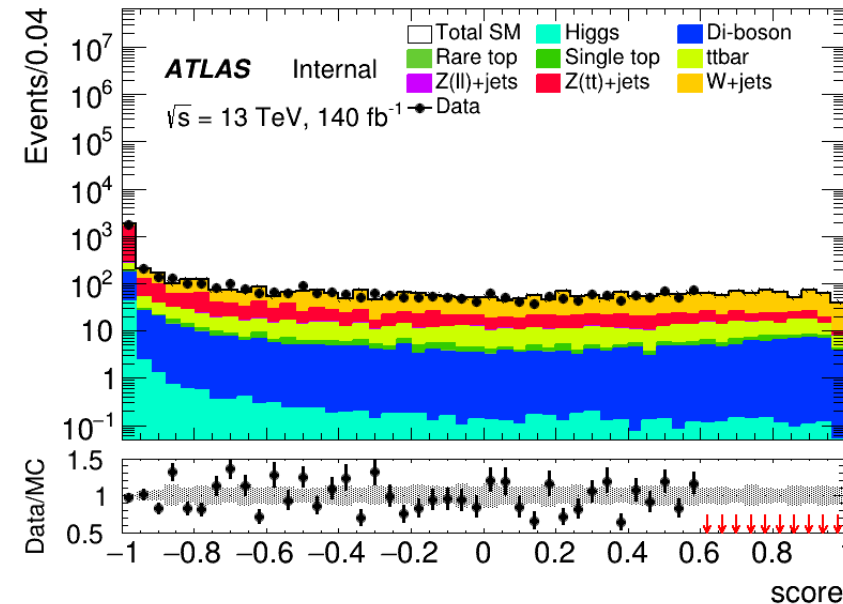
hyper parameter: NTrees=500, learning rate=0.05, max depth=12, MinNodeSize=1%(default)



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)

TODO

1. Reading notes about ABCD method
- 2.

Backup



BackUp

An Interesting Method to Rebin

It can be proved that there are 2^{n-1} ways to rebin if histogram have n bins except 2 bins

2 bins

1. [0] [1] (separate)
2. [0+1] (merged)

Ordered the method and trun into binary number

3 bins

1. [0] [1] [2]
2. [0+1] [2]
3. [0] [1+2]
4. [0+1+2]

Based on Mathmatical Induction



2^{n-1} ways to rebin

4 bins

1. [0] [1] [2] [3]
2. [0+1] [2] [3]
3. [0] [1+2] [3]
4. [0] [1] [2+3]
5. [0+1+2] [3]
6. [0] [1+2+3]
7. [0+1] [2+3]
8. [0+1+2+3]

Example in 4 bins

000 (no walls): [0+1+2+3]
001 (wall at 2-3): [0+1+2] [3]
010 (wall at 1-2): [0+1] [2+3]
011 (walls at 1-2, 2-3): [0+1] [2] [3]
100 (wall at 0-1): [0] [1+2+3]
101 (walls at 0-1, 2-3): [0] [1+2] [3]
110 (walls at 0-1, 1-2): [0] [1] [2+3]
111 (all walls): [0] [1] [2] [3]
That's $2^3 = 8$ options (3 gaps).

Feature_distribution