山东大学
SHANDONG UNIVERSITY

Department of Physics, Shandong University

Compressed EWK study(ISRC1N2)

Chengxin Liao
liaocx@ihep.ac.cn

Apr, Thu 17, 2025

Tasklist

- FF method var distribution check
- Multiclass result(failed)
- BSc thesis: https://www.overleaf.com/project/674e7119837a2580151a0868 (need to submit before the end of Apr)

FF method

Pre-Selection

SR(2ID):
nBaseTau == 2, medium tau == 2
nLeps == 0
nBaseJets >= 1
MET trigger & MET >= 200
OS
bVeto
Mtt_reco <=40 && Mtt_reco >= 130

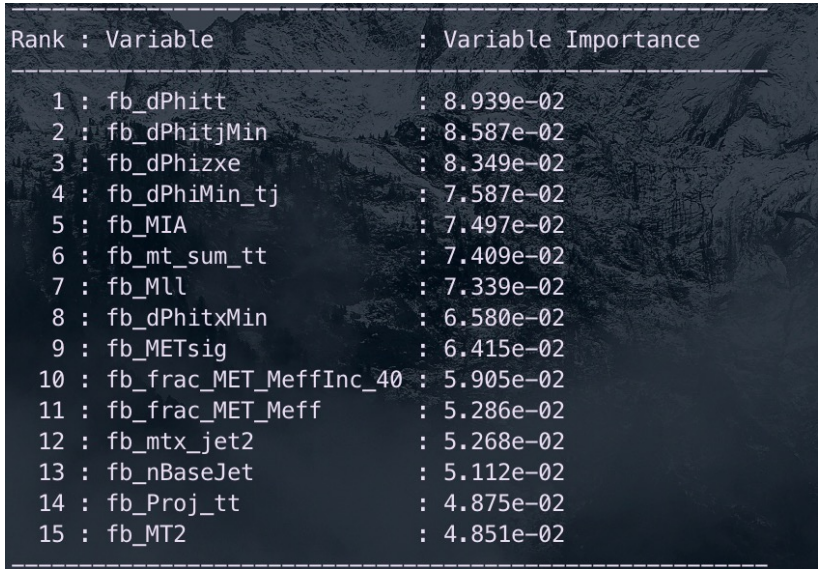
CR(ID):
nBaseTau == 1, medium tau == 1
nLeps >= 1
OS
MET trigger & MET >= 200
bVeto

Multiclass(HH)

Hyperparameters: Ntrees = 200, MaxDepth = 6, MinNodeSize = 2%, Learning rate = 0.03(initial setting)

Feature engineering:
Select a simple model and put all features into model, choose Top 30 vars based on importance list, drop high correlated vars

Final feature list:



Rank	Variable	Variable Importance
1	fb_dPhitt	8.939e-02
2	fb_dPhitjMin	8.587e-02
3	fb_dPhizxe	8.349e-02
4	fb_dPhiMin_tj	7.587e-02
5	fb_MIA	7.497e-02
6	fb_mt_sum_tt	7.409e-02
7	fb_Mll	7.339e-02
8	fb_dPhitxMin	6.580e-02
9	fb_METsig	6.415e-02
10	fb_frac_MET_MeffInc_40	5.905e-02
11	fb_frac_MET_Meff	5.286e-02
12	fb_mtx_jet2	5.268e-02
13	fb_nBaseJet	5.112e-02
14	fb_Proj_tt	4.875e-02
15	fb_MT2	4.851e-02

Split strategy: Separate entries by using mod 5, if separate follow sequence, all weighted entry will split into first fold

Multiclass(HH)

Class: C1N2, VV, Other_bkg

Test_auc = \sum{Test_auc_class}

Train_auc = \sum{Train_auc_class}

Hyperparameter tune:
use optuna to auto-optimze

constraint:

average of AUC need to >= 0.6

penalty function: score = test_auc – 0.3*auc_gap (auc_gap = abs(train_auc- test_auc))

maximum(score)

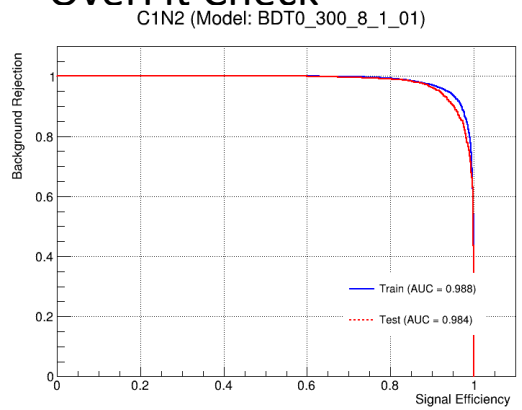
After check some models, find C1N2 result is great, so the constraint and AUC calculation only in VV and Other bkg

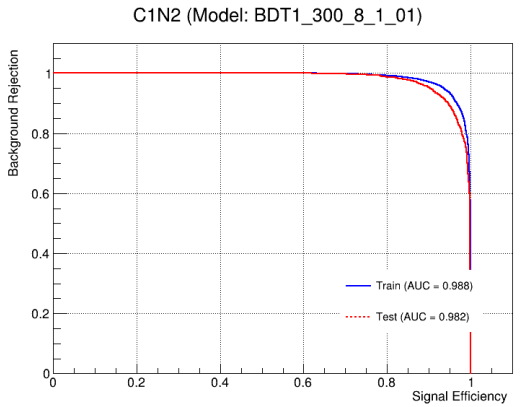
Grid Search
Ntrees: [200, 300, 400]
MaxDepth: [4, 6, 8, 10]
MinNode: [1, 3, 5, 7]
Learning rate: [0.001, 0.01, 0.1]

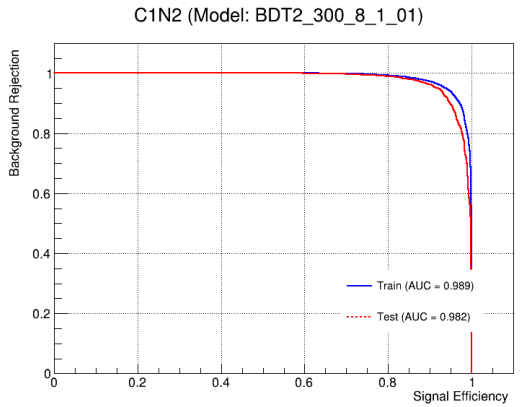
Best one: Ntree=300, MaxDepth=8, MinNode=1%, Learning Rate=0.1

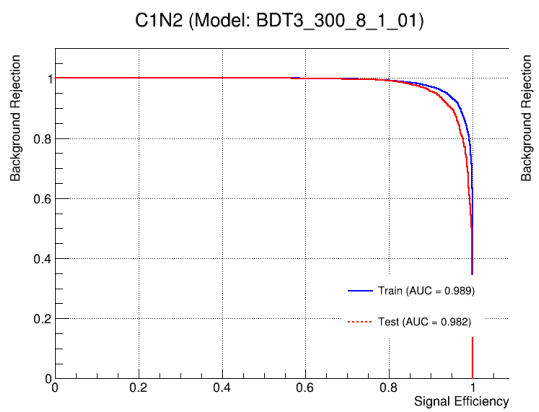
Multiclass(HH)

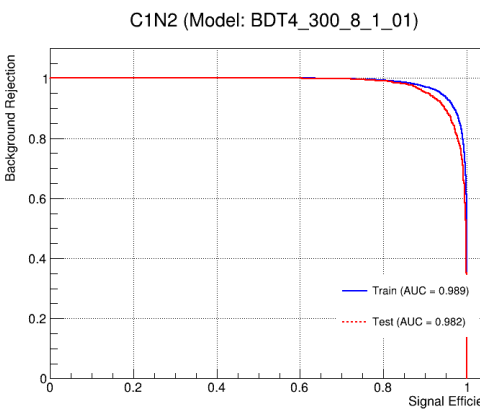
OverFit Check

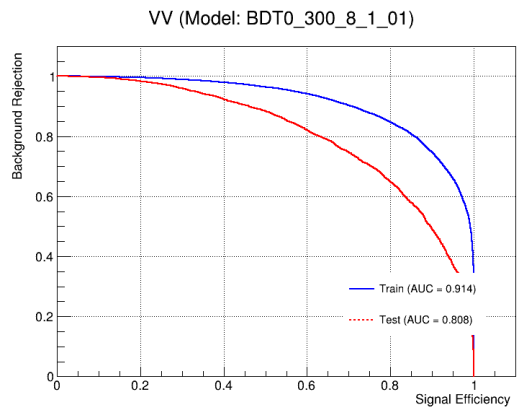


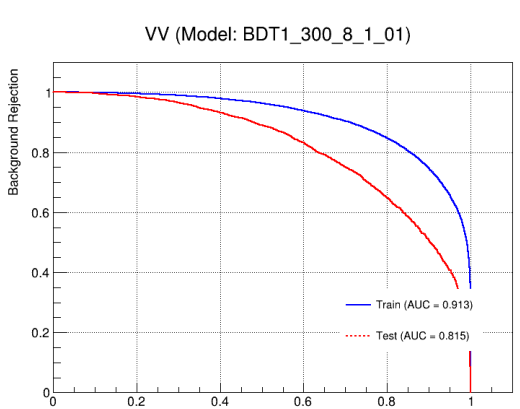


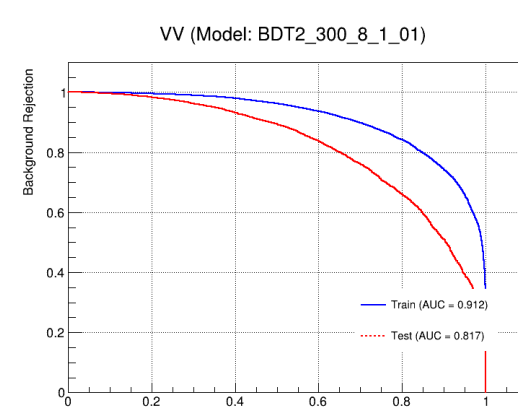


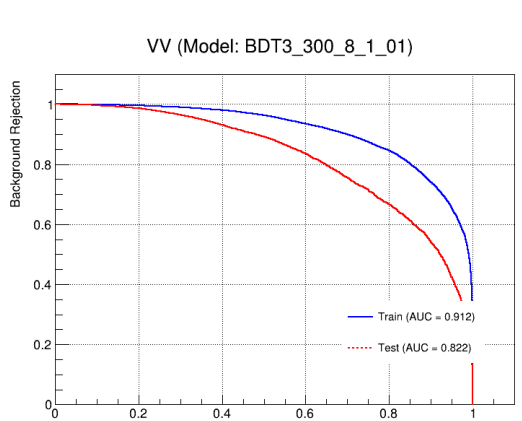


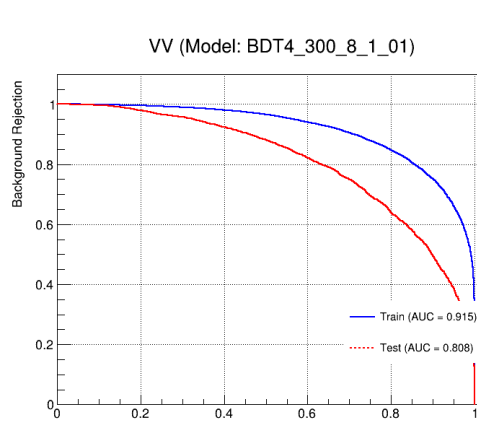


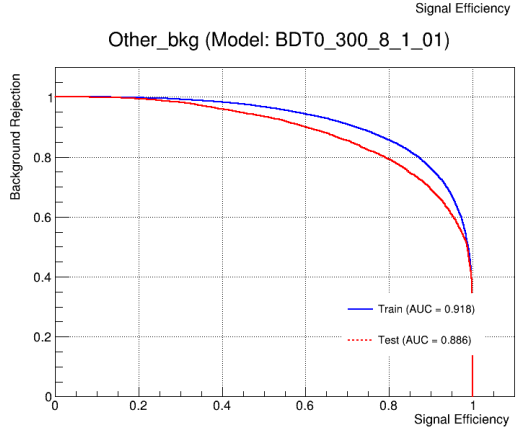


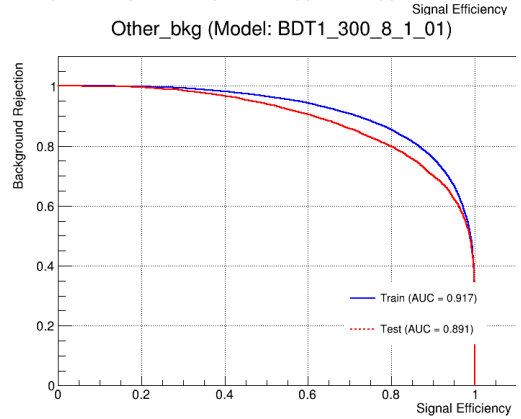


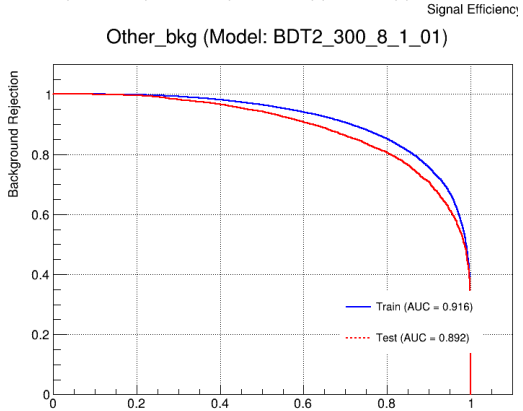


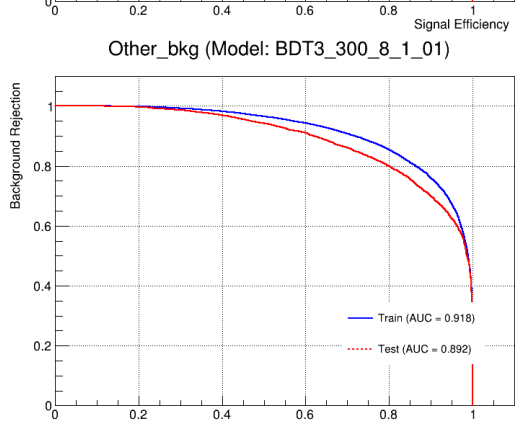


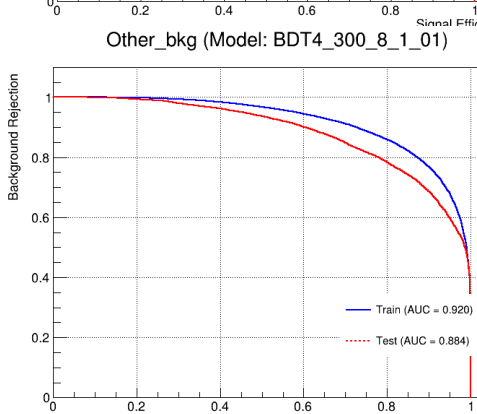




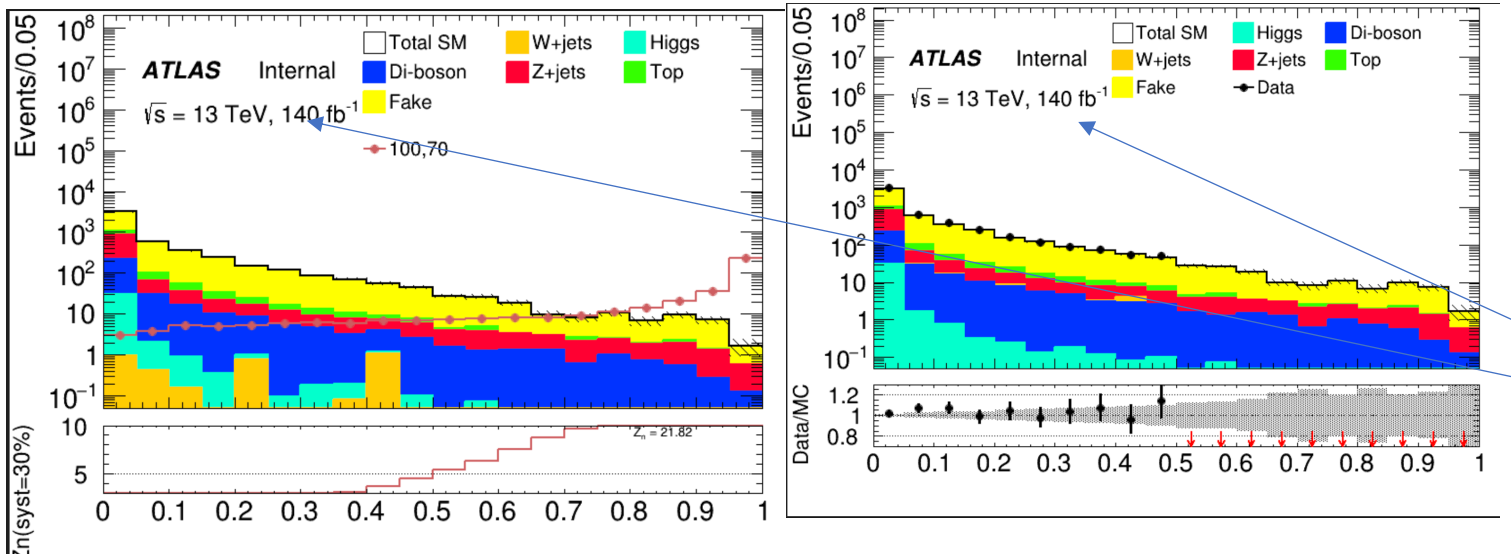








SR



Blind data >= 0.5

Lumi label is wrong, it should be 193.4

Region	VV	Top	Fake	Higgs	Zjets	Wjets	C1N2ISR_100p0_70p0_run2	TotalBkg
[0.0, 0.1]	217.04 ± 2.32	250.77 ± 5.44	2604.03 ± 37.13	32.09 ± 0.50	677.66 ± 8.77	1.35 ± 0.58	6.93 ± 0.51	3782.94 ± 38.61
[0.1, 0.2]	25.53 ± 0.88	29.49 ± 1.88	517.46 ± 15.62	1.08 ± 0.09	31.85 ± 2.03	0.19 ± 0.14	10.35 ± 0.62	605.60 ± 15.89
[0.2, 0.3]	12.93 ± 0.60	12.13 ± 1.22	232.77 ± 10.77	0.38 ± 0.05	15.10 ± 0.73	0.72 ± 0.91	11.39 ± 0.66	274.02 ± 10.92
[0.3, 0.4]	7.49 ± 0.49	6.96 ± 0.92	133.71 ± 8.16	0.30 ± 0.05	9.29 ± 0.57	0.08 ± 0.08	12.13 ± 0.67	157.84 ± 8.24
[0.4, 0.5]	5.24 ± 0.48	3.92 ± 0.70	85.17 ± 6.51	0.18 ± 0.03	6.48 ± 0.41	1.09 ± 0.98	13.76 ± 0.70	102.08 ± 6.65
[0.5, 0.6]	2.76 ± 0.28	1.80 ± 0.47	44.43 ± 4.77	0.12 ± 0.03	4.88 ± 0.46	0.00 ± 0.00	15.21 ± 0.75	53.98 ± 4.82
[0.6, 0.7]	2.70 ± 0.32	0.17 ± 0.06	22.07 ± 3.58	0.07 ± 0.02	3.74 ± 0.31	-0.11 ± 0.11	16.06 ± 0.77	28.64 ± 3.61
[0.7, 0.8]	1.61 ± 0.26	0.70 ± 0.28	13.46 ± 2.88	0.04 ± 0.02	2.96 ± 0.25	0.00 ± 0.00	20.77 ± 0.87	18.77 ± 2.91
[0.8, 0.9]	1.28 ± 0.22	0.38 ± 0.22	12.02 ± 2.56	0.01 ± 0.01	2.58 ± 0.27	0.00 ± 0.00	34.61 ± 1.13	16.27 ± 2.59
[0.9, 1.0]	0.39 ± 0.10	0.08 ± 0.07	6.93 ± 1.79	0.01 ± 0.01	1.51 ± 0.17	0.00 ± 0.00	268.10 ± 3.12	8.92 ± 1.80

fake

TotalBkg	Fake	purity	Data	Data/Bkg
2891.87+-34.3311	2878.83+-34.3195	0.995491	2916	1.00864

Fake CR/VR:

nBaseTau == 1, medium tau == 1

nLeps >= 1

SS

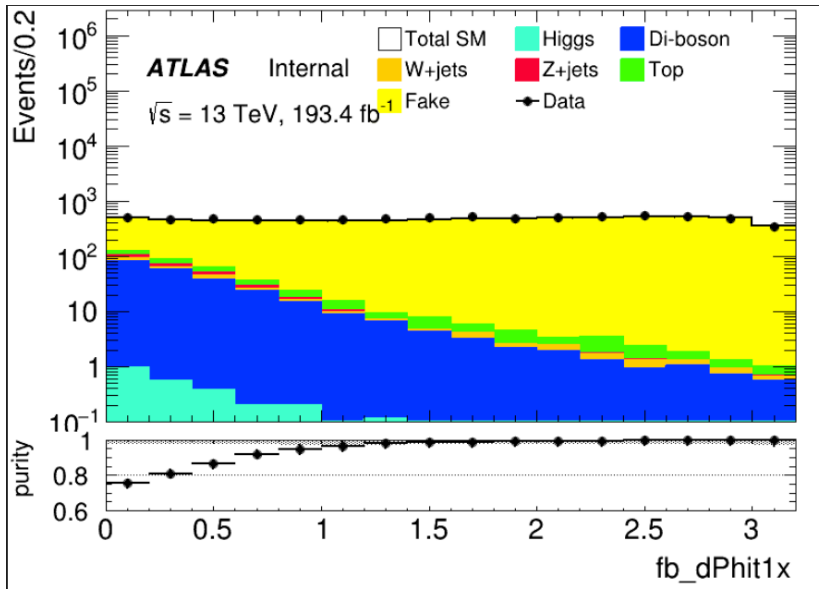
MET trigger & MET >= 200

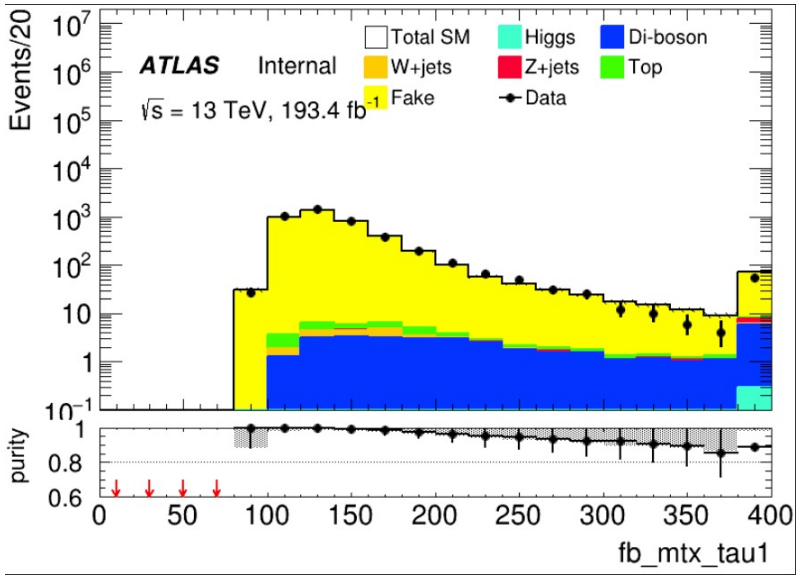
bVeto

$\Delta\Phi(tau, MET) > 2$

$\boldsymbol{M_T(tau, MET) < 150}$

← Orthogonal with LH SR



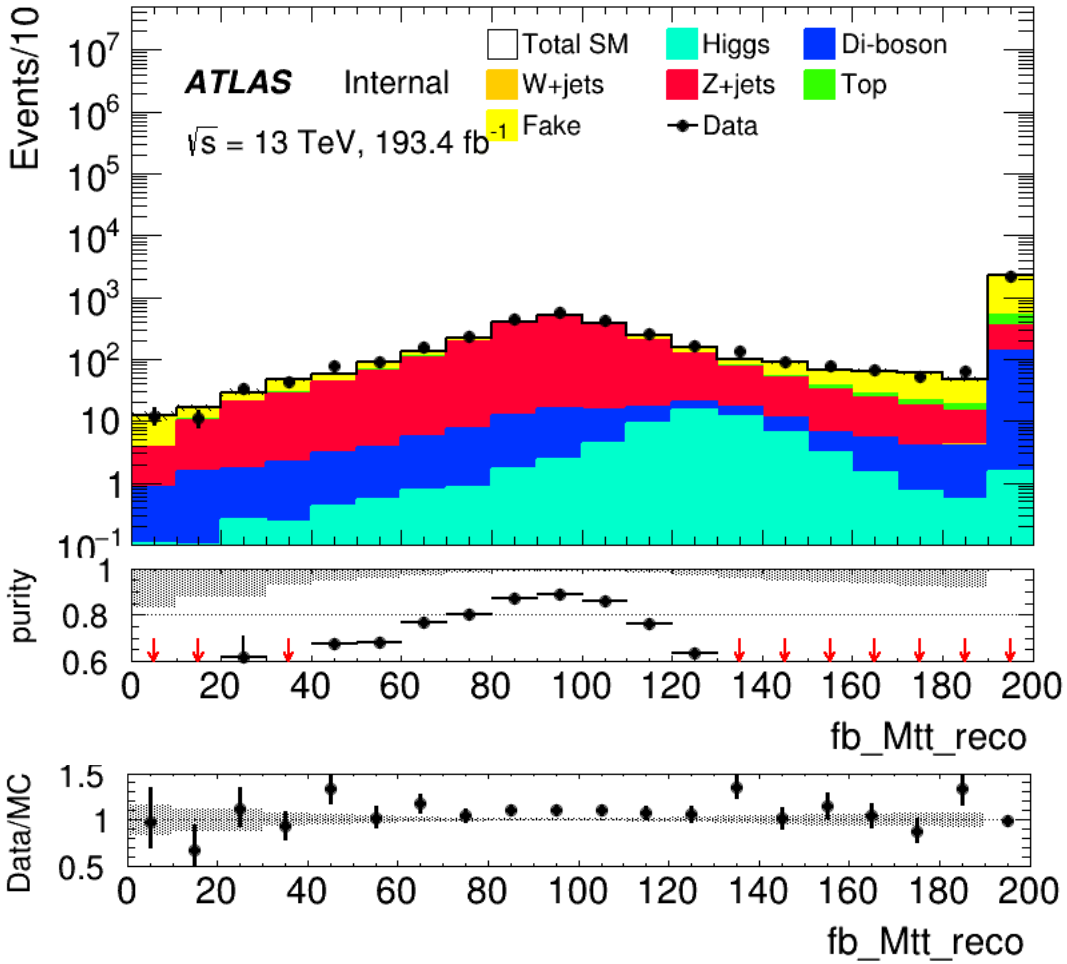


Zjets

>= 2 medium tau
== 0 lepton
METtrig && MET >= 200
OS
b Veto
$M_T(MET, \tau) + M_T(MET, l) < 80$

CR: $80 < M_{tt}reco < 110$
VR: $40 < M_{tt}reco < 80$ || $110 < M_{tt}reco < 130$

Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	1296.91	1132.62+-5.33999	0.873322	1438	1.04934
VR	895.804	659.573+-4.16375	0.736292	978	1.05085

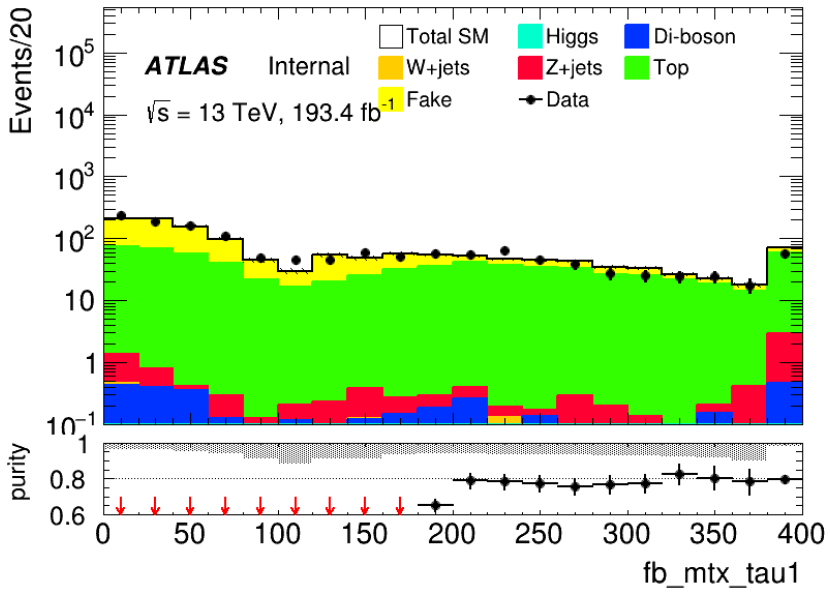


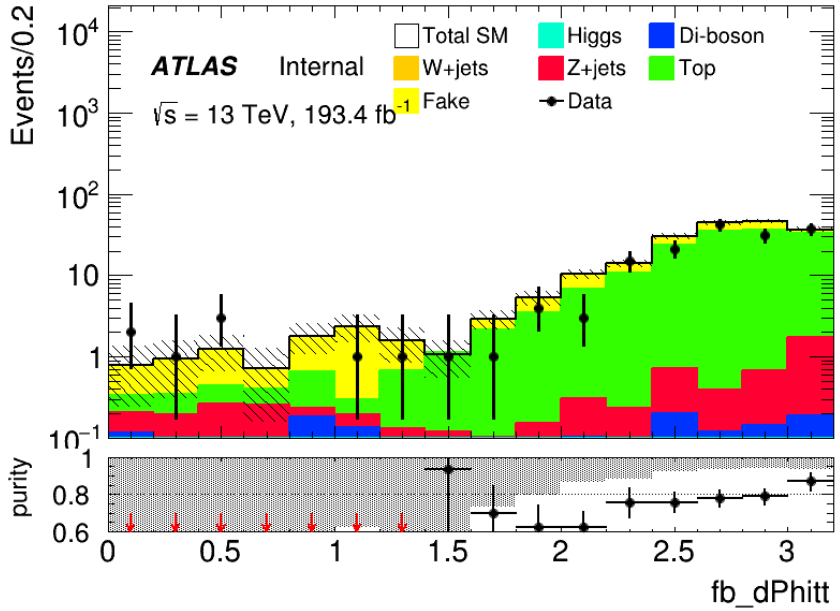
Top

>= 2 medium tau
== 0 lepton
METtrig && MET >= 200
OS
>= 1 bTag

CR: $M_T(l, MET) > 300$

VR: $200 < M_T(l, MET) < 300$

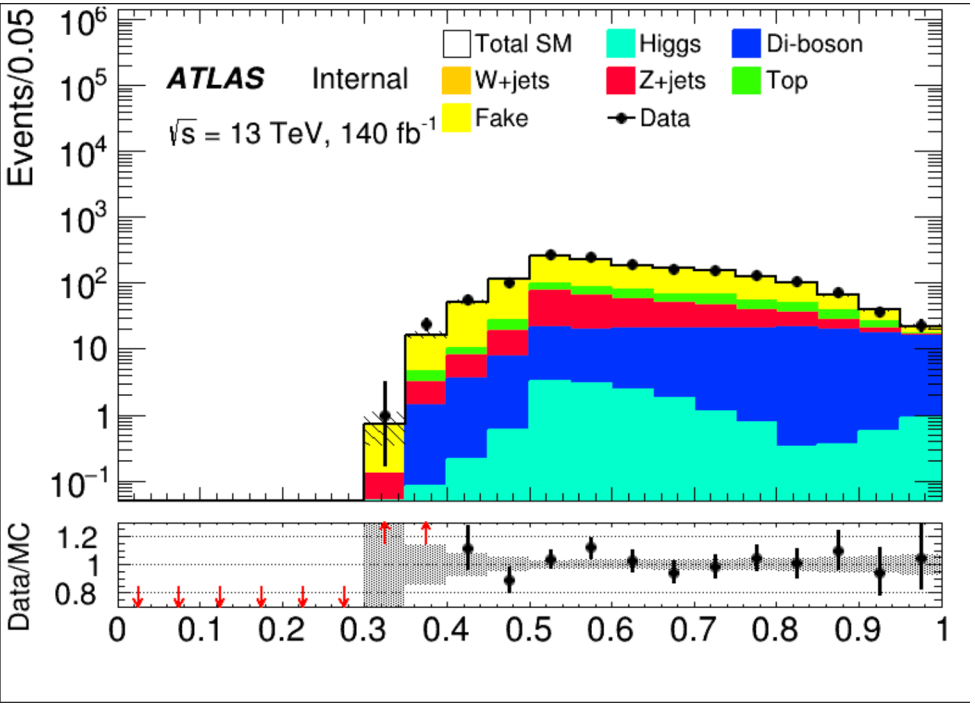


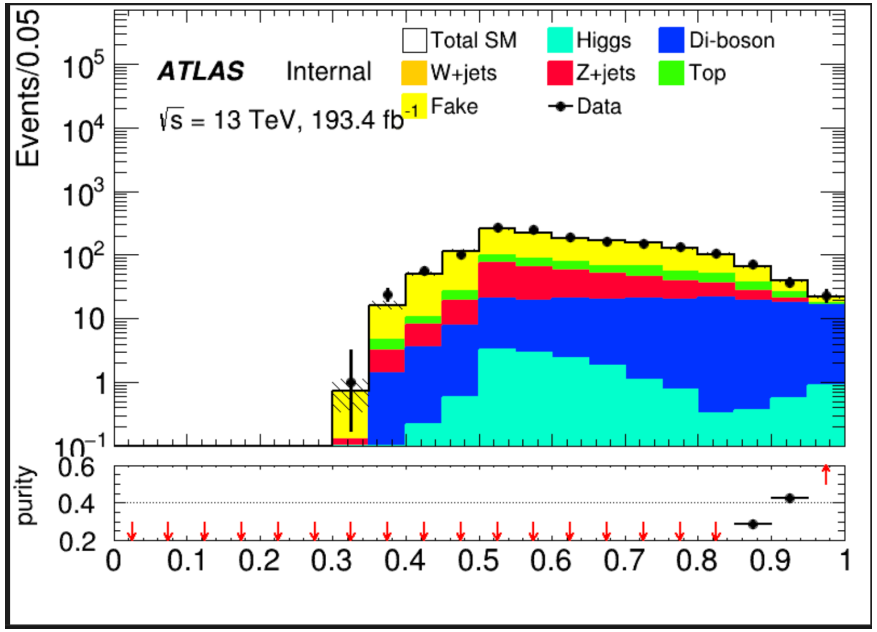


Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	171.478+-5.3865	136.178+-3.88968	0.794142	1438	0.8538
VR	313.202+-7.84499	171.478+-5.3865	0.731	314	1.00255

VV

C1N2_score <= 0.7 && Max(score) == VV_score && VV_score >= 0.90



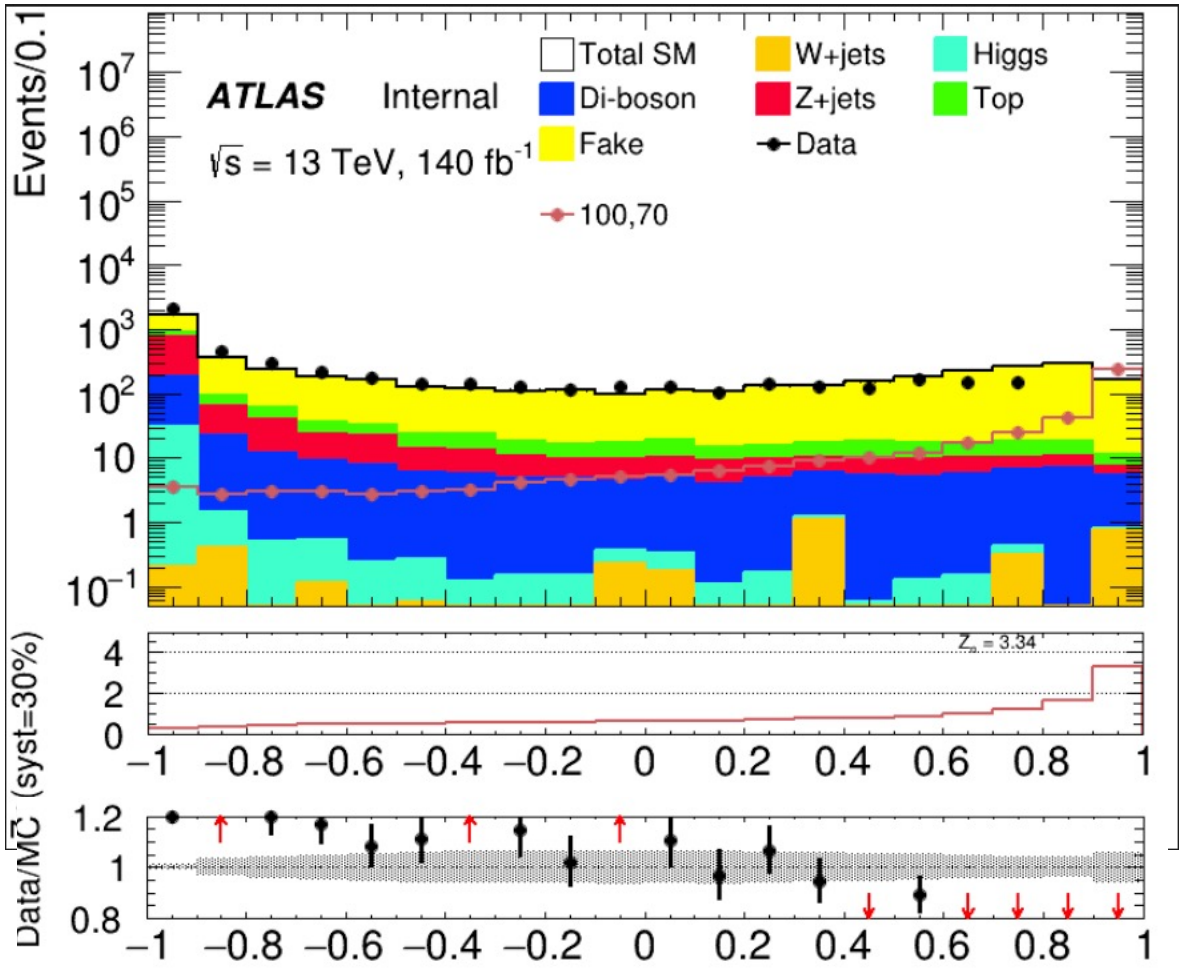


Region	TotalBkg	VV	purity	Data	Data/Bkg
VR	61.3683+-2.9461	31.7024	0.516593	60	0.9777

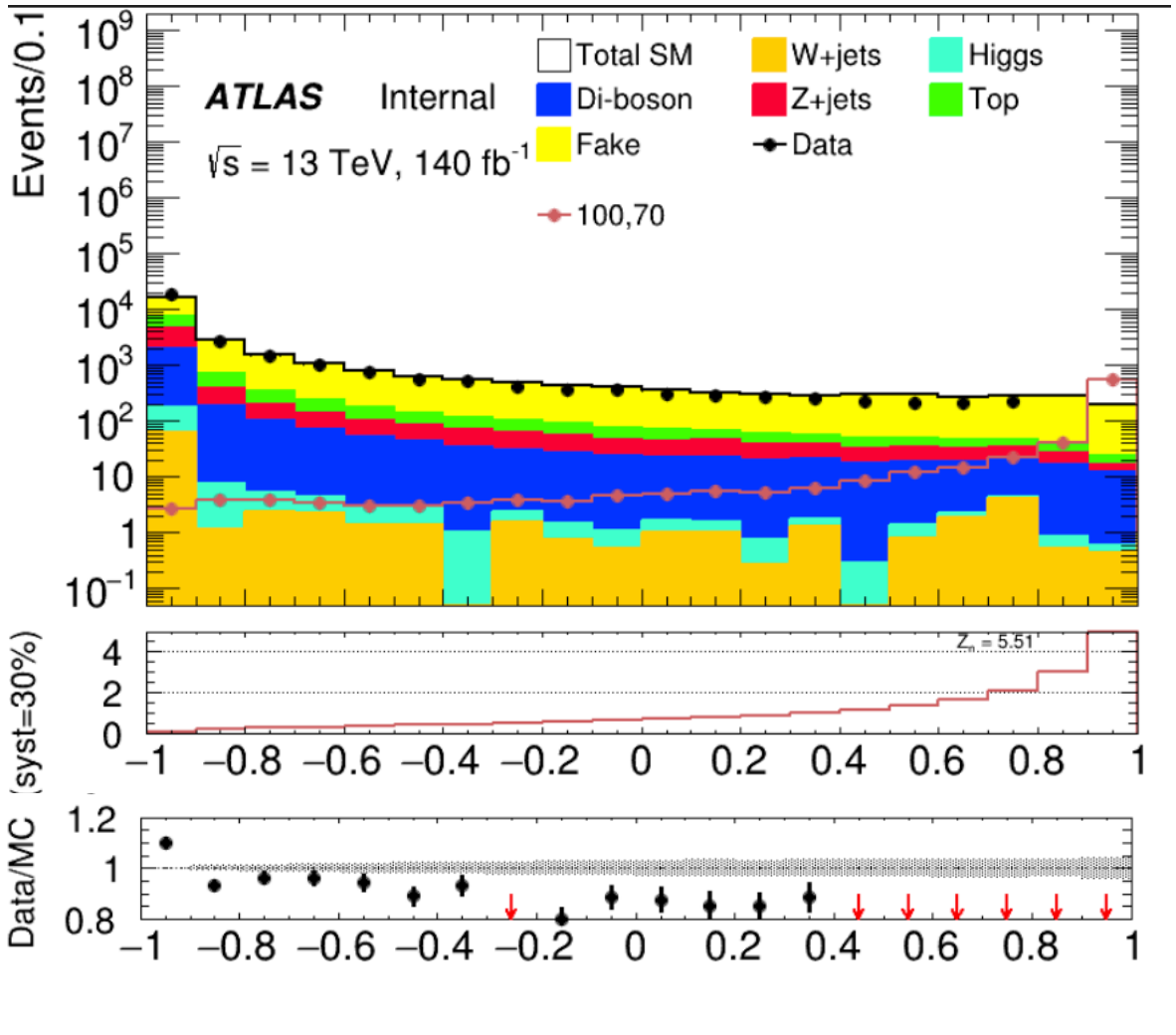
Backup

SR

HH channel



LH channel



Bkg decay mode

Wjets: W->e/muon + nu
W->tau+nu(can contribute true tau_had)
jet misidentified to a fake tau

Zjets: Z->ll/tautau
jet misindentified to fake tau

Top: top->W+b, W can contribute a true tau_had
b-quark is a source of fake

VV: W/Z

LH channel: $\geq 1tau,\ \geq 1lep$
Wjets: W contribute lep, jets misidentified to fake
Zjets:
SingleTop: W contribute lep, b-quark misidentified to fake
VV:

HH channel: $\geq 2tau,\ == 0lep$
Wjets: W contribute tau_had, plus a fake tau
Zjets: Z->tautau(had) or 2 fake tau
SingleTop: W contribute a tau_had, plus a fake tau
VV: