



Search

Chengxin Liao
On behalf of the EWK-2tau team
liaocx@ihep.ac.cn

May, Fri 16, 2025



Target Signal

- C1C1, C1N2 via stau with $\geq 2\tau + E_T^{miss} + ISR$

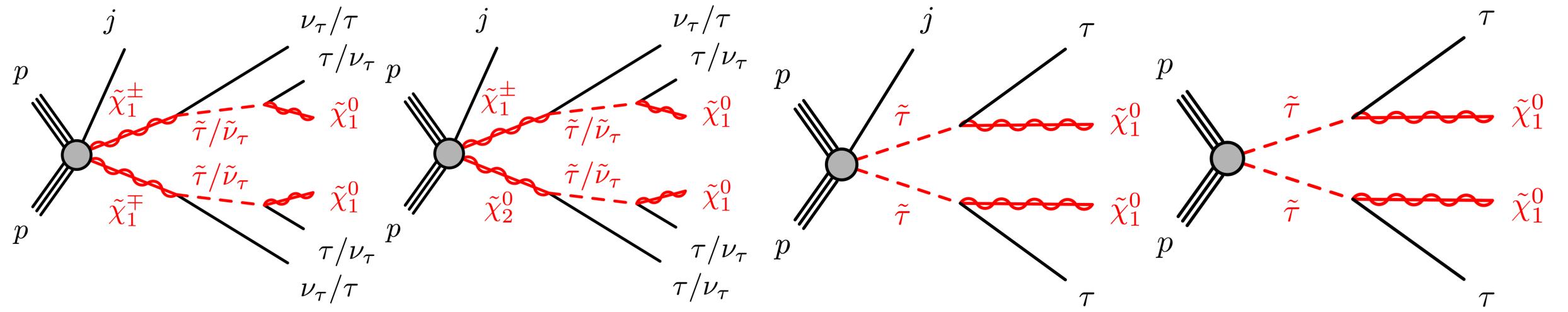
- Direct stau with $2\tau + E_T^{miss}$

ISR channel and inclusive channel

- Previous paper: [JHEP 05 \(2024\) 150](#)

HadHad channel: SR optimization using neuron network/BDT for full Run-2 + partial Run-3 data

LepHad channel: using single lepton trigger, MET trigger, new final state targeting at compressed region



Ntuple setup

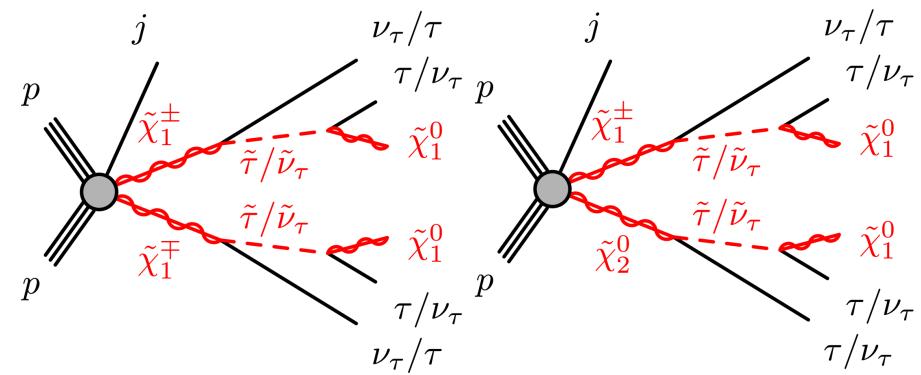
- [MelAnalysis framework](#)
- DAOD PHYS
- ptag: signal p6244, bkg p6490/p6491, data p6479/p6269
- background list: [mc20](#), [mc23](#)

Type	dsid	sample
$t\bar{t}$	410470	PhPy8EG_A14_ttbar_hdamp258p75_nonallhad
	410471	PhPy8EG_A14_ttbar_hdamp258p75_allhad
single top	410644,410645	PowhegPythia8EvtGen_A14_singletop_schan_lept
	410658,410659	PhPy8EG_A14_tchan_BW50_lept
	601352,601355	PhPy8EG_tW_dyn_DR_incl
Rare top	304014	MadGraphPythia8EvtGen_A14NNPDF23_3top_SM
	410276	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttee_mll_1_5
	410277	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttmumu_mll_1_5
	410278	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttautau_mll_1_5
$t\bar{t}X$	410081	MadGraphPythia8EvtGen_A14NNPDF23_ttbarWW
	504330	aMCPy8EG_NNPDF30NLO_A14N23LO_ttee
	504334	aMCPy8EG_NNPDF30NLO_A14N23LO_ttmumu
	504338	aMCPy8EG_NNPDF30NLO_A14N23LO_ttZqq
	504342	aMCPy8EG_NNPDF30NLO_A14N23LO_ttautau
	504346	aMCPy8EG_NNPDF30NLO_A14N23LO_ttZnumu

Type	dsid	sample
Higgs	346343	PhPy8EG_A14NNPDF23_NNPDF30ME_ttH125.allhad
	346344	PhPy8EG_A14NNPDF23_NNPDF30ME_ttH125.semilep
	346345	PhPy8EG_A14NNPDF23_NNPDF30ME_ttH125.dilep
	345097	PowhegPythia8EvtGen_NNLOPS_mnlo_30_ggh125_mumu
	345121	PowhegPythia8EvtGen_NNLOPS_mnlo_30_ggh125_tautau137
	345121	PowhegPythia8EvtGen_NNLOPS_mnlo_30_ggh125_tautau15hp20
	345122	PowhegPythia8EvtGen_NNLOPS_mnlo_30_ggh125_tautau15hp20
	345123	PowhegPythia8EvtGen_NNLOPS_mnlo_30_ggh125_tautau30h20
	345149	PowhegPythia8EvtGen_NNPDF30_AZNLOCTEQ6L1_VBFH125_bb
	346191	PowhegPythia8EvtGen_NNPDF30_AZNLOCTEQ6L1_VBFH125_tautau137
	346191	PowhegPythia8EvtGen_NNPDF30_AZNLOCTEQ6L1_VBFH125_tautau15hp20
	346192	PowhegPythia8EvtGen_NNPDF30_AZNLOCTEQ6L1_VBFH125_tautau15hp20
	346193	PowhegPythia8EvtGen_NNPDF30_AZNLOCTEQ6L1_VBFH125_tautau30h20
	345053	PowhegPythia8EvtGen_NNPDF3_AZNLO_Wph125j_MINLO_lvbb_VpT
	345054	PowhegPythia8EvtGen_NNPDF3_AZNLO_Wph125j_MINLO_lvbb_VpT
	345055	PowhegPythia8EvtGen_NNPDF3_AZNLO_ZH125j_MINLO_llbb_VpT
	345056	PowhegPythia8EvtGen_NNPDF3_AZNLO_ZH125j_MINLO_lvbb_VpT
	345057	PowhegPythia8EvtGen_NNPDF3_AZNLO_ggZH125_llbb
	345098	PowhegPythia8EvtGen_NNPDF3_AZNLO_ggZH125_Hmmnu_Zinc
	345103	PowhegPythia8EvtGen_NNPDF30_AZNLO_ZH125j_Hmmnu_Zinc_MINLO
	345104	PowhegPythia8EvtGen_NNPDF30_AZNLO_Wph125j_Hmmnu_Winc_MINLO
	345105	PowhegPythia8EvtGen_NNPDF30_AZNLO_Wph125j_Hmmnu_Wind_MINLO
	345109	PowhegPythia8EvtGen_NNPDF3_AZNLO_Wph125j_MINLO_lvcc_VpT
	345110	PowhegPythia8EvtGen_NNPDF3_AZNLO_Wph125j_MINLO_lvcc_VpT
	345111	PowhegPythia8EvtGen_NNPDF3_AZNLO_ZH125j_MINLO_llcc_VpT
	345112	PowhegPythia8EvtGen_NNPDF3_AZNLO_ZH125j_MINLO_lvcc_VpT
	345113	PowhegPythia8EvtGen_NNPDF3_AZNLO_ggZH125_llcc
	345211	PowhegPyEG_NNPDF30_AZNLO_Wph125j_Winc_MINLO_tautau
	345212	PowhegPyEG_NNPDF30_AZNLO_Wph125j_Winc_MINLO_tautau
	345217	PowhegPyEG_NNPDF30_AZNLO_ZH125j_Zinc_MINLO_tautau
	346329	PowhegPyEG_NNPDF30_AZNLO_ggZH125_lltautau_file

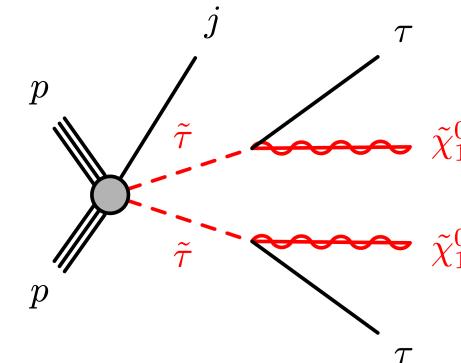
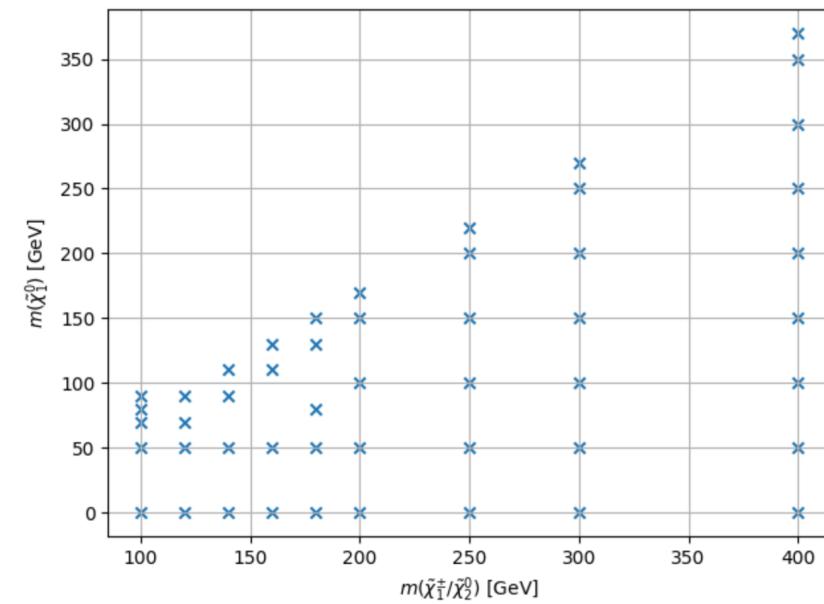
Type	dsid	sample
Z+jets	700320-700322, 700467-700469	Sh_2211_Zee_maxHTpTV2
	700323-700325, 700470-700472	Sh_2211_Zmumu_maxHTpTV2
	700792-700794, 700901-700903	Sh_2214_Ztautau_maxHTpTV2
	700358	Sh_2211_Zee2jets_Min_N_TChannel
	700359	Sh_2211_Zmm2jets_Min_N_TChannel
	700360	Sh_2211_Ztt2jets_Min_N_TChannel
W+jets	700338-700340	Sh_2211_Wenu_maxHTpTV2
	700341-700343	Sh_2211_Wmunu_maxHTpTV2
	700344-700349	Sh_2211_Wtaunu_*_maxHTpTV2
	700362	Sh_2211_Wenu2jets_Min_N_TChannel
	700363	Sh_2211_Wmunu2jets_Min_N_TChannel
	700364	Sh_2211_Wtaunu2jets_Min_N_TChannel
Diboson	345705,345706	Sherpa_222_NNPDF30NNLO_ggllll
	345718	Sherpa_222_NNPDF30NNLO_ggllvvWW
	345723	Sherpa_222_NNPDF30NNLO_ggllvvZZ
	364288	Sherpa_222_NNPDF30NNLO_llll_lowMllPtComplement
	364289	Sherpa_222_NNPDF30NNLO_lllv_lowMllPtComplement
	364290	Sherpa_222_NNPDF30NNLO_llvv_lowMllPtComplement
	364302,364303	Sherpa_222_NNPDF30NNLO_ggZ*Zqq
	364304,364305	Sherpa_222_NNPDF30NNLO_ggW*lvW*qq
	700600	Sh_2212_llll
	700601	Sh_2212_lllv
	700602,700603	Sh_2212_llvv_*
	700604	Sh_2212_lvvv
	700605	Sh_2212_vvvv
	701000	Sh_2214_lllljj
	701005	Sh_2214_lllvjj
	701010,701015	Sh_2214_llvvjj_*
	701020	Sh_2214_lllljj_Int
Triboson	701025	Sh_2214_lllvjj_Int
	701030,701035	Sh_2214_llvvjj_*
	701085	Sh_2214_ZqqZll
	701090	Sh_2214_ZbbZll
	701095	Sh_2214_ZqqZvv
	701100	Sh_2214_ZbbZvv
	701105	Sh_2214_WqqZll
	701110	Sh_2214_WqqZvv
	701115	Sh_2214_WlVZqq
	701120	Sh_2214_WlVZbb
	701125	Sh_2214_WlVWqq
	364242	Sherpa_222_NNPDF30NNLO_WWW_3l3v_EW6
	364243-364244	Sherpa_222_NNPDF30NNLO_WWZ
	364245-364246	Sherpa_222_NNPDF30NNLO_WZZ
	364247-364249	Sherpa_222_NNPDF30NNLO_ZZZ

Signal samples

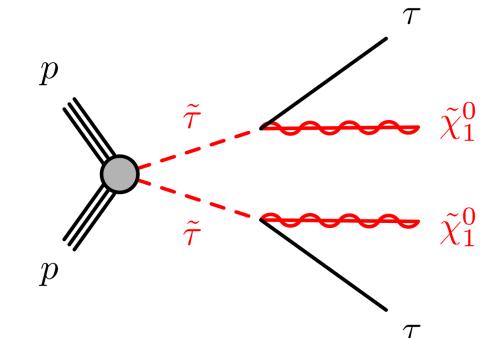


$\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ mass degenerate and pure wino

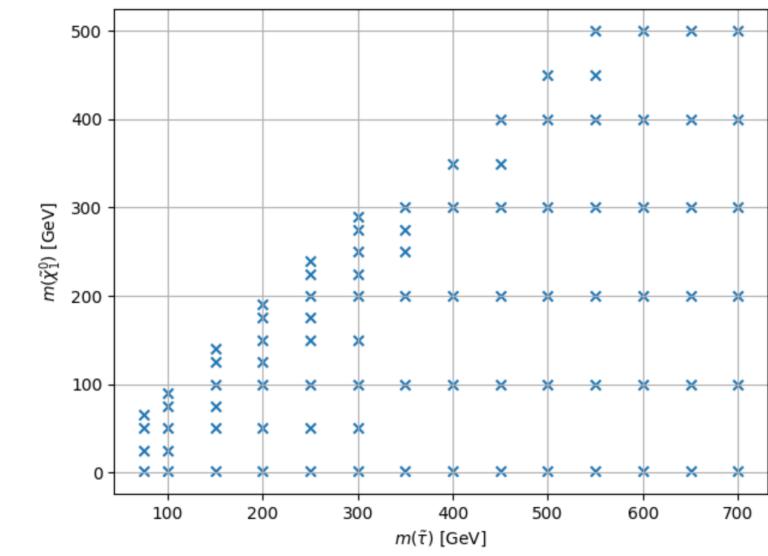
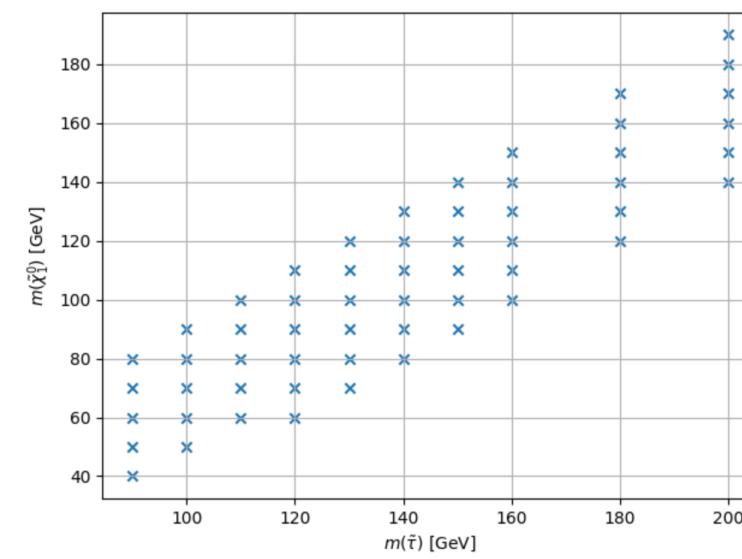
$\tilde{\tau}/\tilde{\nu}$ mass degenerate and lie midway between $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$



$\tilde{\tau}_L/\tilde{\tau}_R$ mass degenerate
Bino-like $\tilde{\chi}_1^0$



$\tilde{\tau}_L/\tilde{\tau}_R$ mass degenerate
Bino-like $\tilde{\chi}_1^0$



Object Definition

- Full ST config: [Run2](#), [Run3](#)

Selection	Electron	Muon	Tau	Jet
Baseline	$p_T > 4.5 \text{ GeV}$ $ \eta < 2.47$ $ z_0 \cdot \sin \theta < 0.5 \text{ mm}$ ID: LooseAndBLayerLLH	$p_T > 5 \text{ GeV}$ $ \eta < 2.5$ $ z_0 \cdot \sin \theta < 0.5 \text{ mm}$ ID: Medium	$p_T > 15 \text{ GeV}$ $ \eta < 1.37 \text{ or } 1.52 < \eta < 2.5$ $ Q = 1$ 1 or 3 tracks Ele ID: RNN Loose Tau ID: VeryLoose	AntiKt4EMPFlow $p_T > 20 \text{ GeV}$ $ \eta < 4.5$ JVT: FixedEffPt if $p_T < 60 \text{ GeV}$ bTag: GN2v01, 77% WP
Signal	ID: TightLLH Run2 ISO: Loose_VarRad, HighPtCaloOnly Run3 ISO: Tight_VarRad $ d_0/\sigma_{d_0} < 5$	ISO: PflowLoose_VarRad	Tau ID: Medium	

Overlap Removal

- ST default overlap removal
- Do Tau OLR

Reject	Against	Criteria
electron	electron	shared track, $p_{T1} < p_{T2}$
tau	electron	$\Delta R < 0.2$
tau	muon	$\Delta R < 0.2$
muon	electron	is calo-muon & shared ID track
electron	muon	shared ID track
photon	electron	$\Delta R < 0.4$
photon	muon	$\Delta R < 0.4$
jet	electron	$\Delta R < 0.2$
electron	jet	$\Delta R < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^{\text{ele}})$
jet	muon	NumTrack < 3 & (ghost-associated or $\Delta R < 0.2$)
muon	jet	$\Delta R < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^{\mu})$
jet	tau	$\Delta R < 0.2$

Trigger Strategy

Stau:

MET trigger for high MET region,
Single lepton trigger for low MET region

C1C1/C1N2: MET trigger

Trigger	Trigger name	Year	HLT	Offline
MET trigger	HLT_xe70_mht	2015	70	200
	HLT_xe90_mht_L1XE50	2016	90	
	HLT_xe100_mht_L1XE50	2016	100	
	HLT_xe110_pufit_L1XE55	2017	110	
	HLT_xe110_pufit_L1XE50	2017	110	
	HLT_xe110_pufit_xe70_L1XE50	2018	70	
	HLT_xe110_pufit_xe65_L1XE50	2018	65	
	HLT_xe65_cell_xe90_pfopufit_L1XE50	2022,2023	90	
single electron trigger	HLT_e24_lhmedium_L1EM20VH	2015	24	25
	HLT_e60_lhmedium	2015	60	61
	HLT_e120_lhloose	2015	120	121
	HLT_e26_lhtight_nod0_ivarloose	2016-2018	26	27
	HLT_e60_lhmedium_nod0	2016-2018	60	61
	HLT_e140_lhloose_nod0	2016-2018	140	141
	HLT_e26_lhtight_ivarloose_L1EM22VHI	2022-2023	26	27
	HLT_e60_lhmedium_L1EM22VHI	2022-2023	60	61
	HLT_e140_lhloose_L1EM22VHI	2022-2023	140	141
single muon trigger	HLT_mu20_iloose_L1MU15	2015	20	21
	HLT_mu26_ivarmedium	2016-2018	26	27.3
	HLT_mu50	2015-2018	50	52.5
	HLT_mu24_ivarmedium_L1MU14FCH	2022-2023	24	25.2
	HLT_mu50_L1MU14FCH	2022-2023	50	52.5

Direct Stau ISR signal region definition(HH)

- Preselection
- SR selection using neuron network score
 - Fake tau estimation using fake factor method
 - DNN signal score > 0.7
- Background composition

HH Pre-selection

≥ 2 medium taus (OS)

0 base lepton

bveto

MET trigger

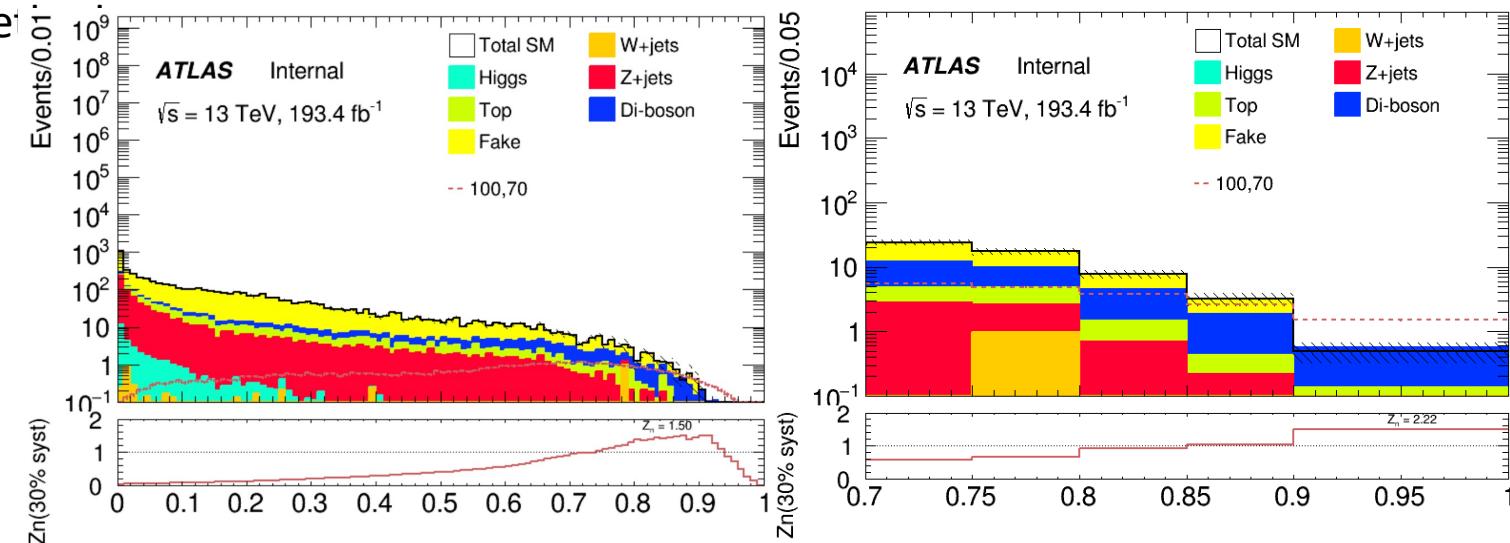
$\text{MET} \geq 200 \text{ GeV}$

$M_{\tau\tau}^{\text{reco}} < 40 \text{ GeV or } M_{\tau\tau}^{\text{reco}} > 130 \text{ GeV}$

SR

Pre-selection

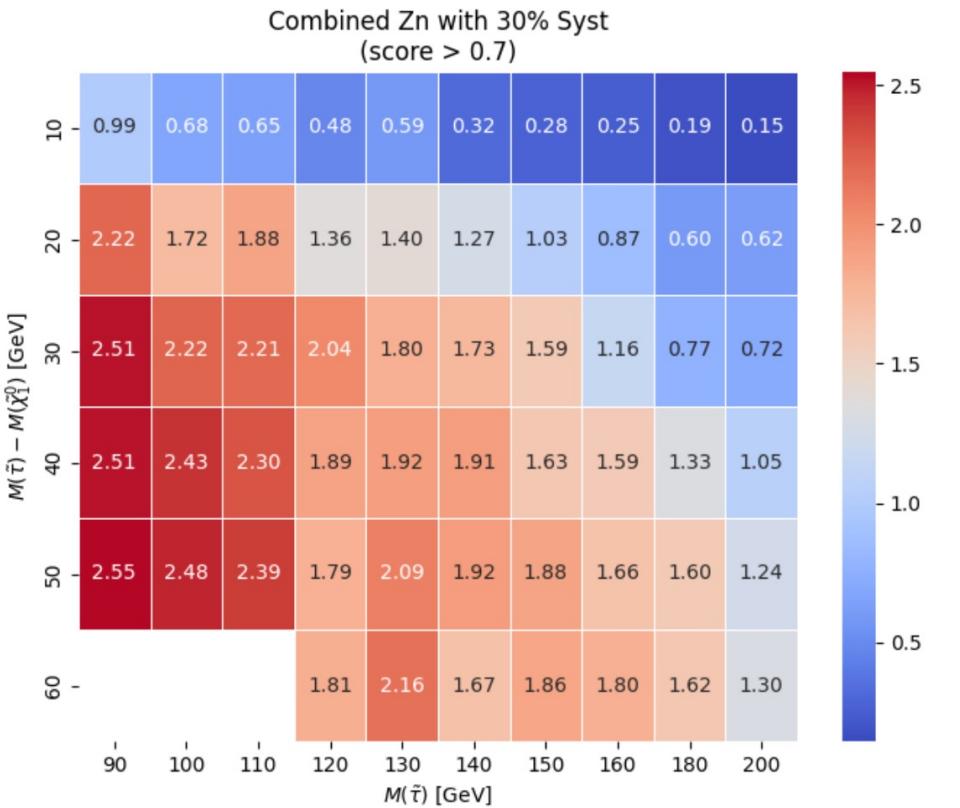
DNN score > 0.7



Process	[0.70,0.75]	[0.75,0.80]	[0.80,0.85]	[0.85,0.90]	[0.90,1.00]	Combined
TotalBkg	24.17 ± 2.35	17.35 ± 2.07	7.69 ± 1.13	3.20 ± 0.77	0.49 ± 0.17	52.90 ± 4.84
Fake	11.95 ± 2.23	7.35 ± 1.66	3.24 ± 1.04	1.33 ± 0.72	-0.07 ± 0.09	23.80 ± 3.05
VV	7.32 ± 0.52	5.18 ± 0.52	2.97 ± 0.27	1.45 ± 0.22	0.42 ± 0.08	17.33 ± 0.82
Top	2.12 ± 0.49	2.23 ± 0.52	0.80 ± 0.31	0.21 ± 0.14	0.11 ± 0.10	5.47 ± 0.80
Zjets	2.69 ± 0.23	1.61 ± 0.21	0.66 ± 0.17	0.20 ± 0.08	0.02 ± 0.07	5.18 ± 0.38
Wjets	0.00 ± 0.00	0.98 ± 0.98	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.98 ± 0.98
Higgs	0.09 ± 0.03	0.00 ± 0.00	0.02 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.14 ± 0.03
StauStauISR-100-70	5.61 ± 0.10	4.90 ± 0.09	3.84 ± 0.08	2.59 ± 0.07	1.53 ± 0.05	18.48 ± 0.18
ZnSignificance	0.58	0.65	0.91	1.03	1.50	2.22

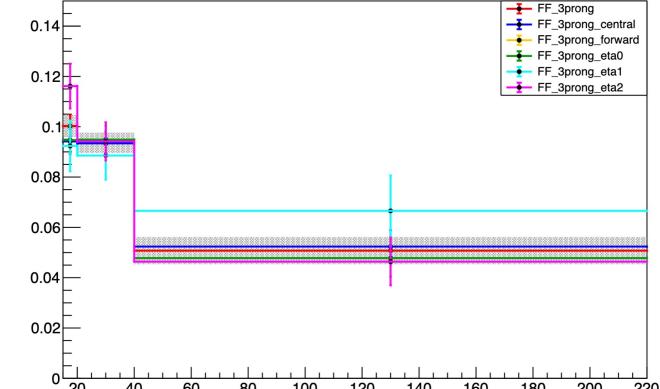
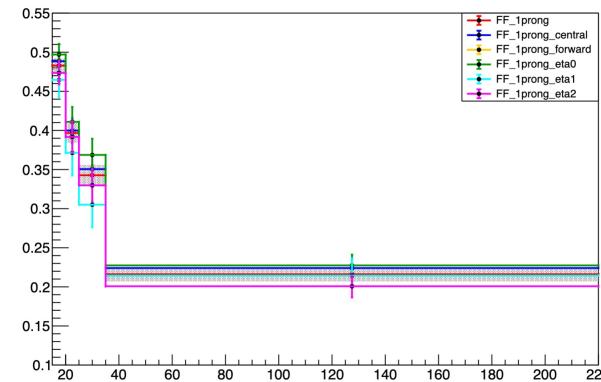
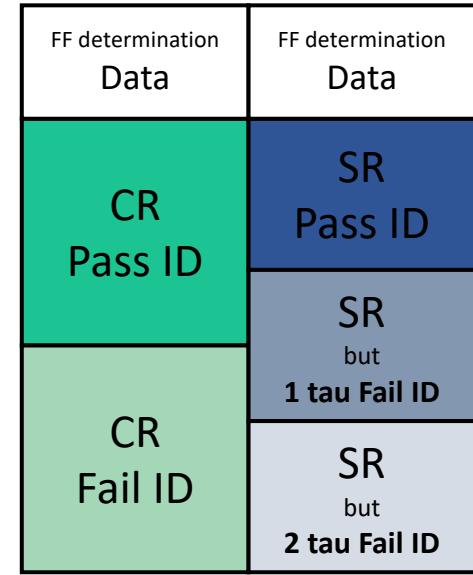
Direct Stau ISR signal region definition(HH)

- Expected sensitivity
 - 30% flat systematic uncertainty



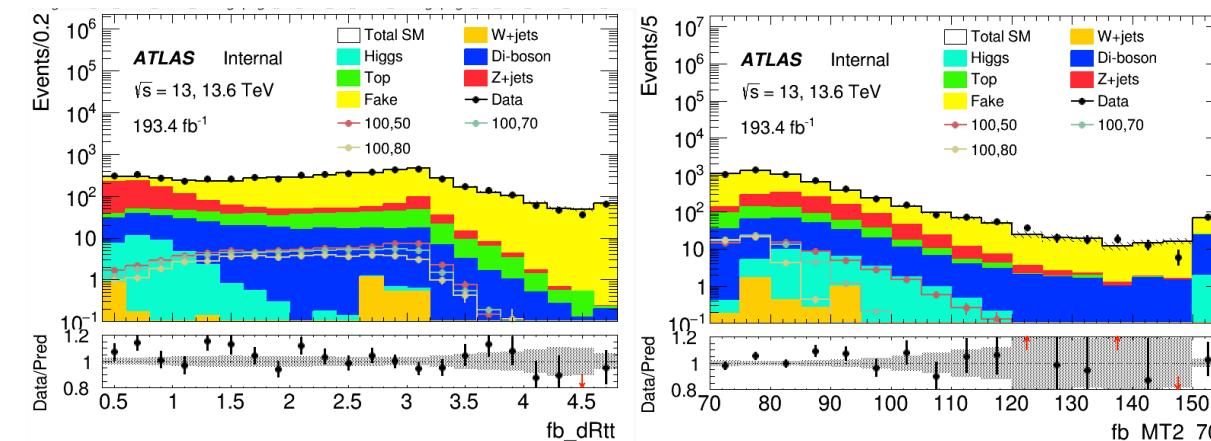
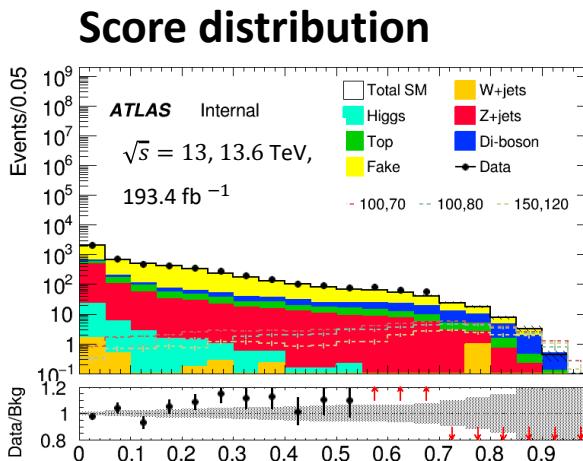
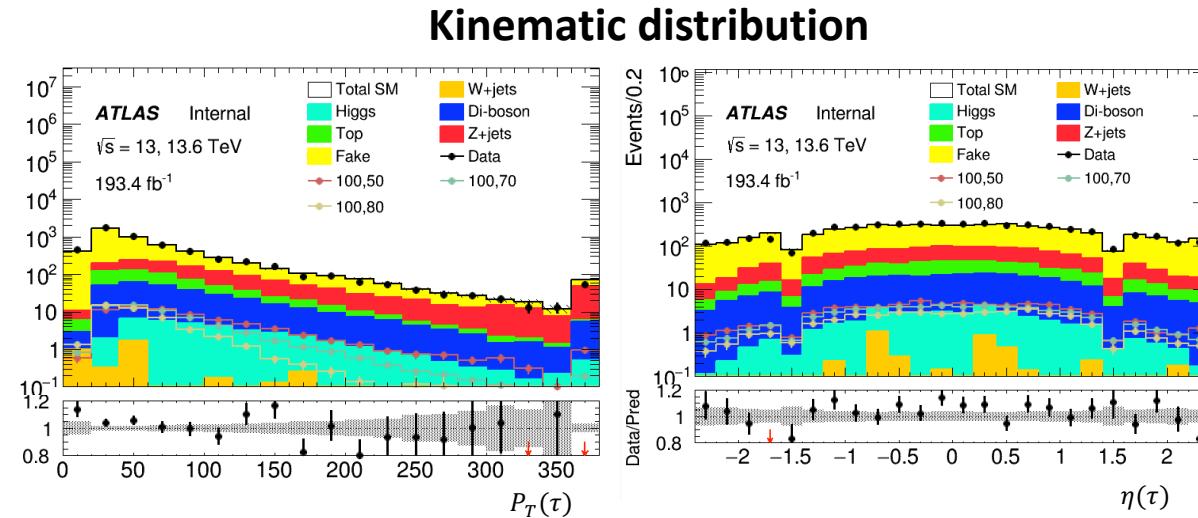
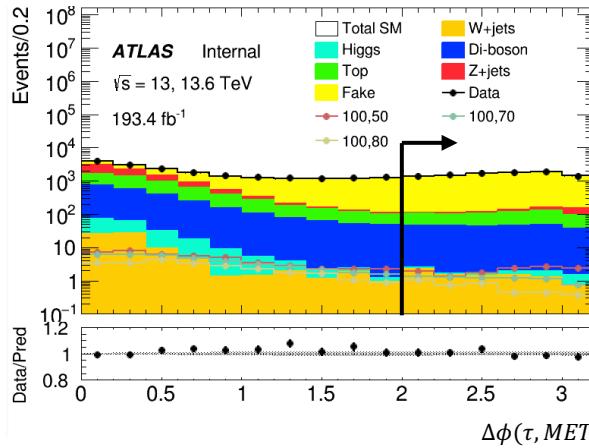
Direct Stau ISR fake estimation

- CRs (fake factor computation)
 - METtrig
 - MET ≥ 200
 - bveto
 - 1 signal lepton
 - $\Delta\phi(\tau, \text{MET}) > 2$
 - **ID: ≥ 1 medium tau**
 - **antiID: ≥ 1 VeryLoose tau, 0 medium tau**
- SRs
 - preselection
 - 2ID: ≥ 2 medium tau
 - 1ID1antiID: ≥ 2 VeryLoose tau , 1 medium tau
 - 2antiID: ≥ 2 VeryLoose tau , 0 medium tau
- **Binned in prongness, tau eta, tau pT**
 - Eta bins
 - 2 bins: central [0,1.37], forward [1.52,2.5]
 - 3 bins: eta0,1,2 for [0,1), [1, 1.37], [1.52,2.5]
- **Auto binning:**
 - $> 10\%$ of events in nominator and denominator
 - Add bins to bin i until it is not consistent anymore with bin i - 1
 - Relative stat uncertainty on ratio smaller than 50%
 - $>10\%$ events in nominator and denominator



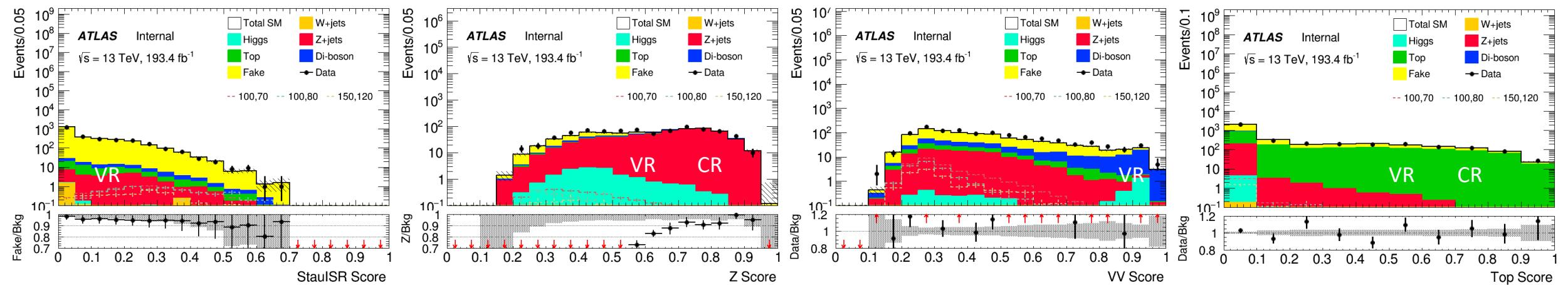
Direct Stau ISR fake estimation validation(HH)

- $\Delta\phi(\tau, \text{MET})$ N-1 plots
- Data-driven fakes in preselection region



Direct Stau ISR background estimation (HH)

Region	Selections	Total Bkg	Dominant Bkg	Purity	Data	Data/Bkg	
Fake VR	signal score < 0.7, max bkg score: Fake	2870 ± 40	2770 ± 40	0.96	2794	0.972	
Z CR	signal score < 0.7, max bkg score: Z	Z Score > 0.7	276 ± 7	257 ± 6	0.93	295	1.07
Z VR		Z Score $\in (0.5, 0.7)$	251 ± 7	198 ± 5	0.79	264	1.05
VV VR	signal score < 0.7, max bkg score: VV	VV Score > 0.85	48 ± 2.8	30 ± 1.1	0.62	55	1.15
Top CR	HH bTag	Top Score > 0.6	374 ± 9	271 ± 5	0.72	375	1
Top VR		Top Score $\in (0.5, 0.6)$	178 ± 6	178 ± 6	0.64	195	1.09



Direct Stau ISR signal region definition(LH)

- Preselection
- SR selection using neuron network score
 - Fake tau estimation using fake factor method
 - DNN signal score > 0.7
- Background composition

LH Pre-selection

≥ 1 medium taus

1base lepton, 1 signal lepton

Opposite-sign

bveto

MET trigger

$\text{MET} \geq 200$

$\Delta\phi(\tau, \text{MET}) < 2$

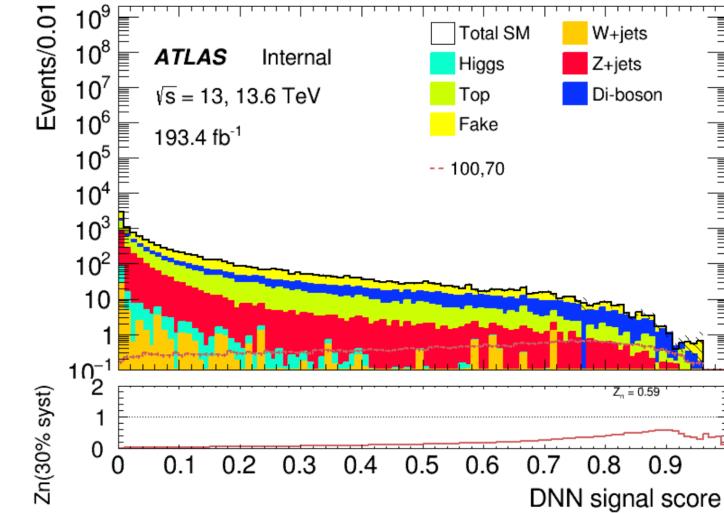
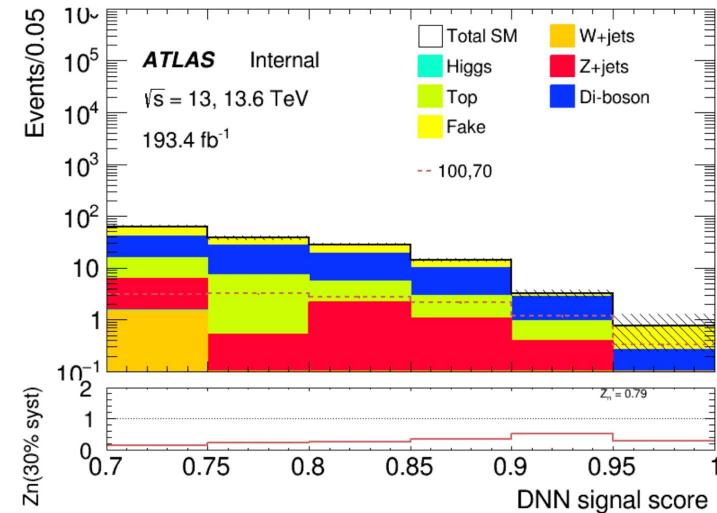
$M_{T2,70} < 100$

$\Delta R_{\tau\ell} \in (0.6, 3.6)$

SR

Pre-selection

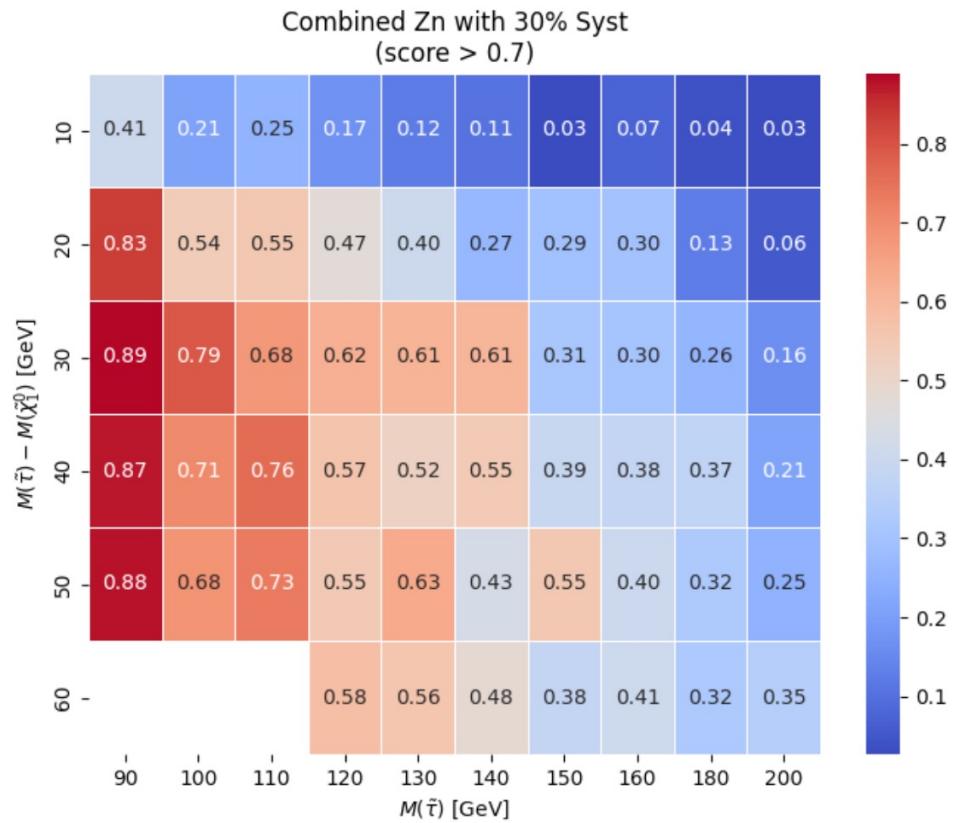
DNN score > 0.7



Process	[0.70,0.75]	[0.75,0.80]	[0.80,0.85]	[0.85,0.90]	[0.90,0.95]	[0.95,1.00]	Combined
TotalBkg	63.19 ± 3.25	38.56 ± 3.70	28.54 ± 2.06	14.29 ± 1.34	3.17 ± 0.54	0.75 ± 0.51	148.50 ± 5.56
VV	25.27 ± 0.86	19.64 ± 0.81	12.90 ± 0.63	6.97 ± 0.45	1.77 ± 0.24	0.21 ± 0.04	66.76 ± 1.43
Fake	22.41 ± 2.65	11.87 ± 1.87	10.13 ± 1.85	4.40 ± 1.16	0.45 ± 0.40	0.49 ± 0.51	49.75 ± 3.96
Top	9.28 ± 1.05	6.53 ± 0.86	3.32 ± 0.60	1.86 ± 0.48	0.56 ± 0.26	0.00 ± 0.00	21.56 ± 1.58
Zjets	4.74 ± 0.29	3.33 ± 0.25	2.16 ± 0.24	1.03 ± 0.16	0.37 ± 0.10	0.05 ± 0.02	11.68 ± 0.49
Higgs	0.06 ± 0.03	0.11 ± 0.04	0.03 ± 0.01	0.03 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.22 ± 0.06
Wjets	1.44 ± 1.28	-2.92 ± 2.95	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.00	-1.47 ± 3.22
StauStauISR-100-70	3.13 ± 0.07	3.17 ± 0.07	2.75 ± 0.07	2.17 ± 0.06	1.18 ± 0.04	0.32 ± 0.02	12.71 ± 0.15
ZnSignificance	0.15	0.23	0.26	0.36	0.53	0.29	0.79

Direct Stau ISR signal region definition(LH)

- Expected sensitivity
 - 30% flat systematic uncertainty



C1C1 ISR signal region optimization(HH)

- Input:

- data: MC events passing pre-selection
 - signal: ISRC1C1(with C1 mass=100GeV, N1 mass=70 GeV)(39382 events)
 - bkg(1089081 events).

Pre-selection	
lep-had channel:	$n\text{Taus} \geq 2; n\text{Leps} = 0$
MET	≥ 150 ; pass MET trigger
	$1 \leq n\text{BaseJet} \leq 8$
	Opposite-sign lepton-hadronic tau pair
	bveto
	jet $p\text{t} > 100$ GeV

feature:Pzetaj1,Pzetae,MCTtt,m_jet,dPhiSRV,dPhiVI,nS_tau,PtV,dPhit2x,maxdPhit1j,maxdPhit2j,METOOPtau2,METOHTtt,METOHTjet,Ptje
t1OPtau1,MT2tt_110,mtx_jet,mtx_lep,mt_lep,mt_C1C1,MTtau2met,MTtaumin,MTsum,Mtsum,Mtsumj,MII,MT2,MET,METsig

- Strategy

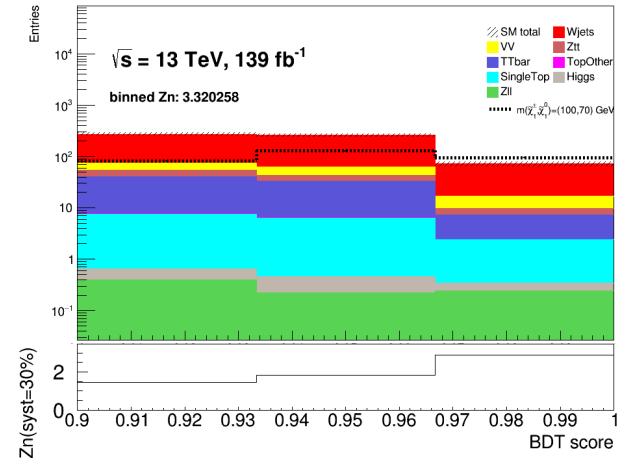
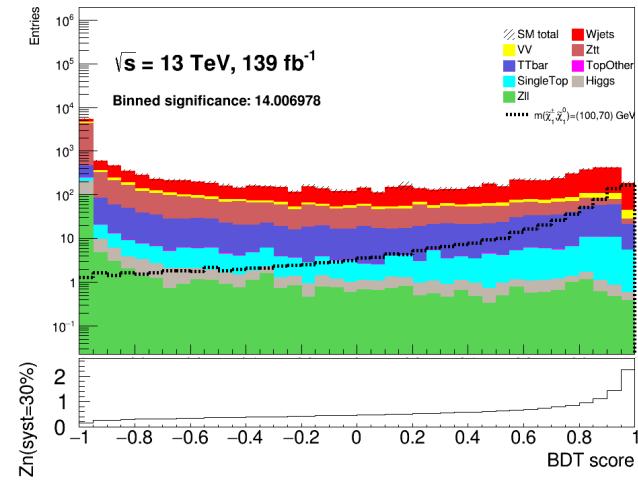
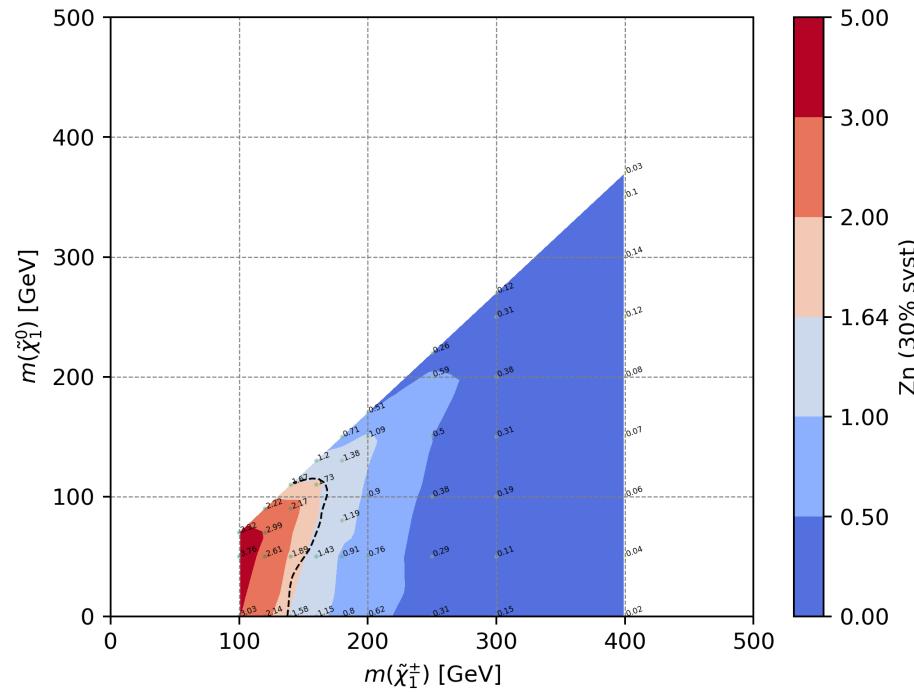
- TMVA.Types.kBDT
- 5-fold train
- hyper parameter: NTrees=400,learning rate=0.1,max depth=8, MinNodeSize=1%

other set-up: nCuts=20:BoostType=Grad:UseBaggedBoost=true:BaggedSampleFraction=0.5:NegWeightTreatment=Pray

C1C1 ISR signal region definition(HH)

SR: pre-selection+BDT score>0.9, binned with 3 bins

Sensitive mapping



SM process	SR2	SR2-Bin1	SR2-Bin2	SR2-Bin3
Wjets	448 ± 28	197 ± 18	55 ± 9	197 ± 19
VV	49.1 ± 1.2	20.7 ± 0.9	7.4 ± 0.4	21.0 ± 0.9
Ztt	25.3 ± 2.9	9.8 ± 1.6	2.3 ± 1.1	13.1 ± 2.1
TTbar	64.5 ± 3.1	26.7 ± 2.0	4.9 ± 0.9	32.9 ± 2.2
TopOther	0.38 ± 0.07	0.15 ± 0.04	0.031 ± 0.028	0.21 ± 0.05
SingleTop	14.7 ± 1.3	5.8 ± 0.8	2.1 ± 0.5	6.8 ± 0.9
Higgs	0.60 ± 0.06	0.24 ± 0.04	0.100 ± 0.026	0.26 ± 0.04
Zll	0.86 ± 0.29	0.22 ± 0.07	0.24 ± 0.18	0.40 ± 0.20
SM total	603 ± 28	271 ± 19	260 ± 19	72 ± 9
$m(\tilde{\chi}_1^\pm, \tilde{\chi}_1^0) = (100, 70)$	307.3 ± 2.4	82.9 ± 1.2	129.1 ± 1.5	95.3 ± 1.3

C1C1 ISR signal region optimization(LH)

- Input:
 - data: MC events passing pre-selection
 - signal: ISRC1C1(with C1 mass=100GeV, N1 mass=70 GeV)(39336 events)
 - /publicfs/atlas/atlasnew/SUSY/users/liangsy/compressISRC1C1/sample/ntuple/
 - bkg(1559557 events).
 - /publicfs/atlas/atlasnew/SUSY/users/yuanjiarong/ISR2024/ntuple/V04_June
- Pre-selection

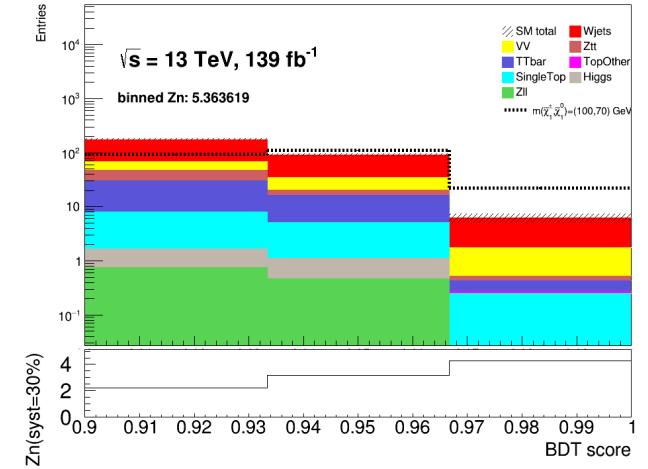
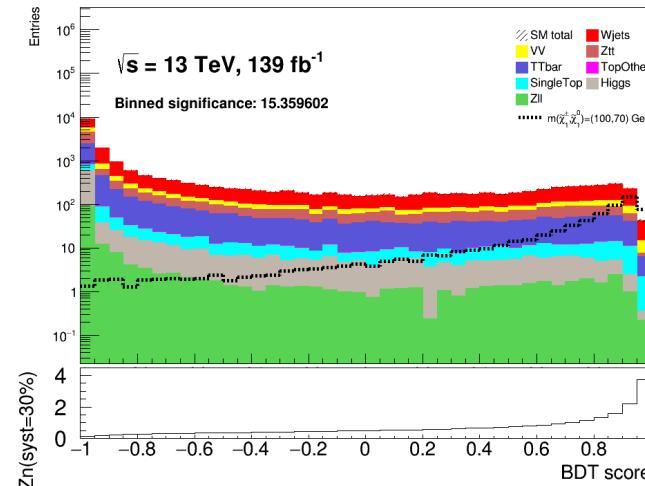
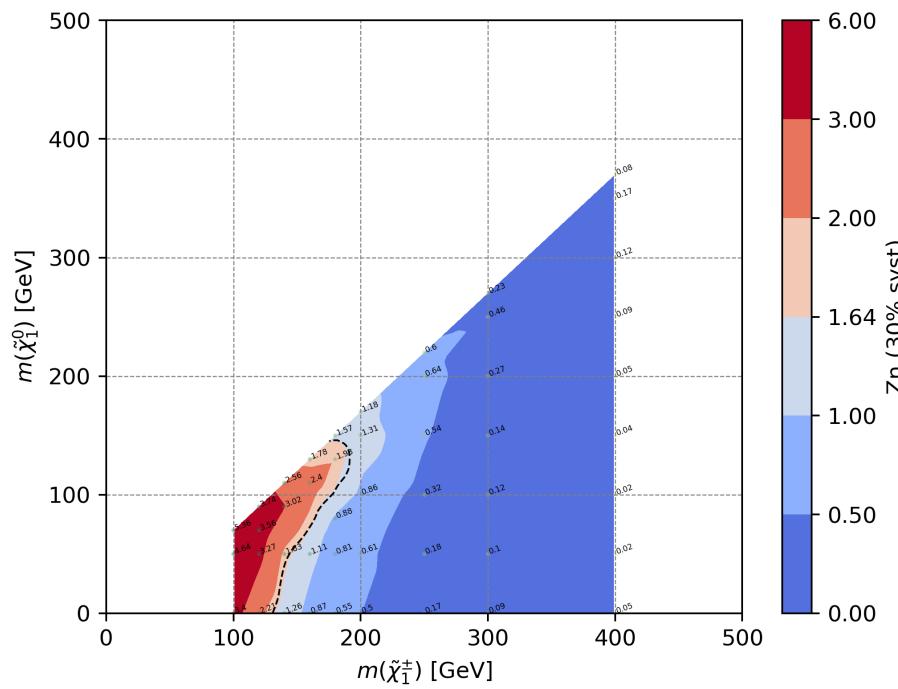
 - lep-had channel: $n\text{Taus} \geq 1$; $n\text{Leps} = 1$
 - $\text{MET} \geq 200$; pass MET trigger
 - $1 \leq n\text{BaseJet} \leq 8$
 - Opposite-sign lepton-hadronic tau pair
 - bveto
 - jet $\text{pt} > 100$ GeV

- feature: dRt1x,dRtt,dPhitt,e_lep,mt_lep,ht_tau,METsig,nBaseJet,mindPhiJe,maxdPhiit2j,MTtau1met,MTtau2met,MTtaumin,
MTtot,MTsum,Minvtt,Minvt2j,MT2tt_110,MCTtt,Pzetaj1,Pzetatt,METOPtau1,METOPtau2,METOHTtt,METOHTjet,Ptjet1OPtau1,Ptjet1OPtau2,Ptjet1OHTtt,PtV
- Strategy
 - TMVA.Types.kBDT
 - 5-fold train
 - hyper paramter: NTrees=300,learning rate=0.05,max depth=6, MinNodeSize=1%
- other set-up: nCuts=20:BoostType=Grad:UseBaggedBoost=true:BaggedSampleFraction=0.5:NegWeightTreatment=Pray

C1C1 ISR signal region definition(LH)

SR: pre-selection+BDT score>0.9, binned with 3 bins

Sensitive mapping



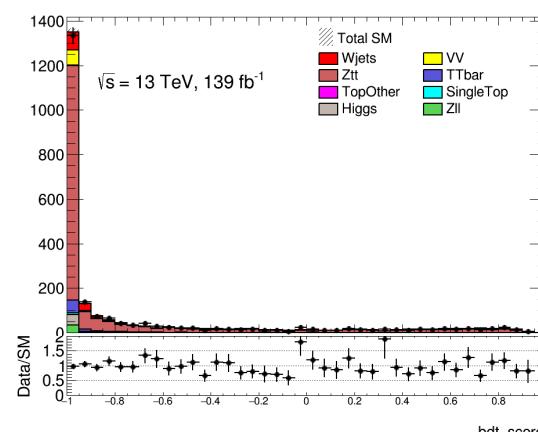
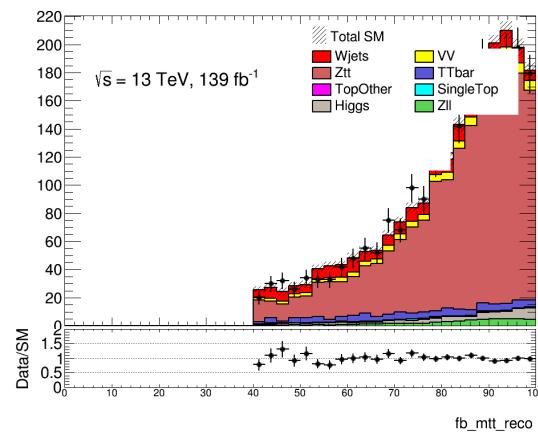
SM process	SR1	SR1-Bin1	SR1-Bin2	SR1-Bin3
Wjets	167 ± 8	56 ± 4	4.4 ± 1.1	106 ± 7
VV	35.90 ± 0.95	13.81 ± 0.54	1.25 ± 0.17	20.8 ± 0.8
Ztt	22.4 ± 2.6	4.5 ± 0.7	0.10 ± 0.04	17.8 ± 2.6
TTbar	33.2 ± 2.3	11.0 ± 1.3	0.15 ± 0.15	22.0 ± 1.8
TopOther	0.27 ± 0.05	0.061 ± 0.025	0.022 ± 0.019	0.18 ± 0.04
SingleTop	10.5 ± 1.0	4.0 ± 0.7	0.24 ± 0.11	6.3 ± 0.8
Higgs	1.59 ± 0.29	0.66 ± 0.19	0.008 ± 0.005	0.92 ± 0.22
Zll	1.22 ± 0.20	0.46 ± 0.13	0.007 ± 0.005	0.75 ± 0.16
SM total	272 ± 9	175 ± 7	91 ± 5	6.2 ± 1.2
$m(\tilde{\chi}_1^\pm, \tilde{\chi}_1^0) = (100, 70)$	226.0 ± 2.0	92.9 ± 1.3	110.6 ± 1.4	22.5 ± 0.6

C1N2 ISR Z+jets estimation(LH)

selections highlighted in green make SR and CR/VR orthogonal

	selection	purity	Total Bkg	Dominant Bkg	Data	Data/Bkg
ZCR	Pre-selection 40 GeV < $M_{\tau\tau}^{\text{reco}}$ < 100 GeV BDT score < -0.8	0.787792	1615.5+- 16.8115	1272.68+- 15.0305	1611	0.997217
ZVR	Pre-selection 40 GeV < $M_{\tau\tau}^{\text{reco}}$ < 100 GeV -0.8 < BDT score < 0.9	0.758041	623.808+- 12.4111	472.872+- 10.2423	619	0.992292

Distributions after pre-selection and 40 GeV < $M_{\tau\tau}^{\text{reco}}$ < 100 GeV (Z include Ztt and Zll)

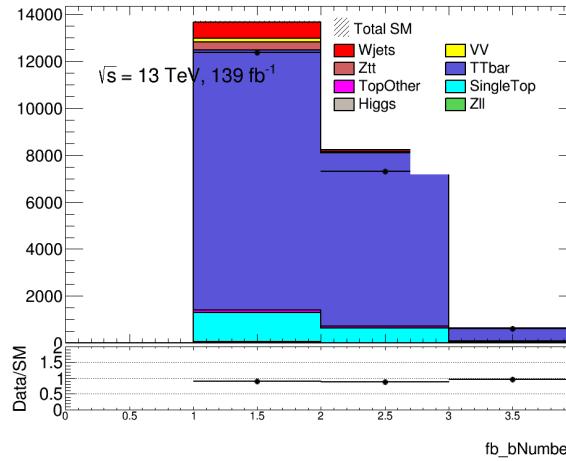
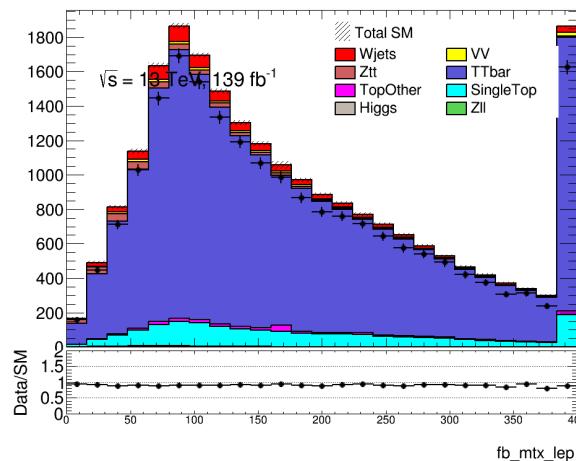


C1N2 ISR Top estimation(LH)

selections highlighted in green make SR and CR/VR orthogonal

	selection	purity	Total Bkg	Dominant Bkg	Data	Data/Bkg
topCR	Pre-selection without bVeto Number of B jets > 0 $M_{int}(lepton, met) > 200 \text{ GeV}$	0.96145	8367.19+- 34.542	8043.72+- 34.0691	7404	0.894123
topVR	Pre-selection without bVeto Number of B jets > 0 $120 \text{ GeV} < M_{int}(lepton, met) < 200 \text{ GeV}$	0.942112	5674.03+- 35.218	5345.57+- 34.4953	5141	0.906058

Distributions after pre-selection without bVeto (top include TopOther,TTbar,Single top)



C1N2 ISR signal region optimization

- BDT method for signal region optimization
- Preselection based on final state($\tau\tau$ or τl)
- Orthogonal had-had and lep-had channel
- Figure of merit: Z_n
- 30% flat systemic uncertainty
- 5fold way to Train(Cross-Validation)

LH channel	HH channel
≥ 1 BaseJet,	Leading Jet $p_t \geq 100$ (ISR)
MET trigger,	MET $\geq 200 GeV$
	b-jet veto
	$\tau\tau(\ell)$ opposite sign
	$m(\tau_1, \tau_2) \leq 40 GeV$ or $m(\tau_1, \tau_2) \geq 130 GeV$
$\geq 1\tau_{had}$	$\geq 2\tau_{had}$
$\geq 1lep$	$= 0lep$

Grid Search for the best model

Hyperparameter	Scan Range
NTrees	200, 300, 400
MaxDepth	4, 6, 8
MinNode	1, 3, 5
Learning rate	0.01, 0.03, 0.05, 0.08, 0.1

Penalty function to balance the AUC and overfit

$$\mathcal{F} = AUC_{validation} - 0.3 \times AUC_{gap}$$

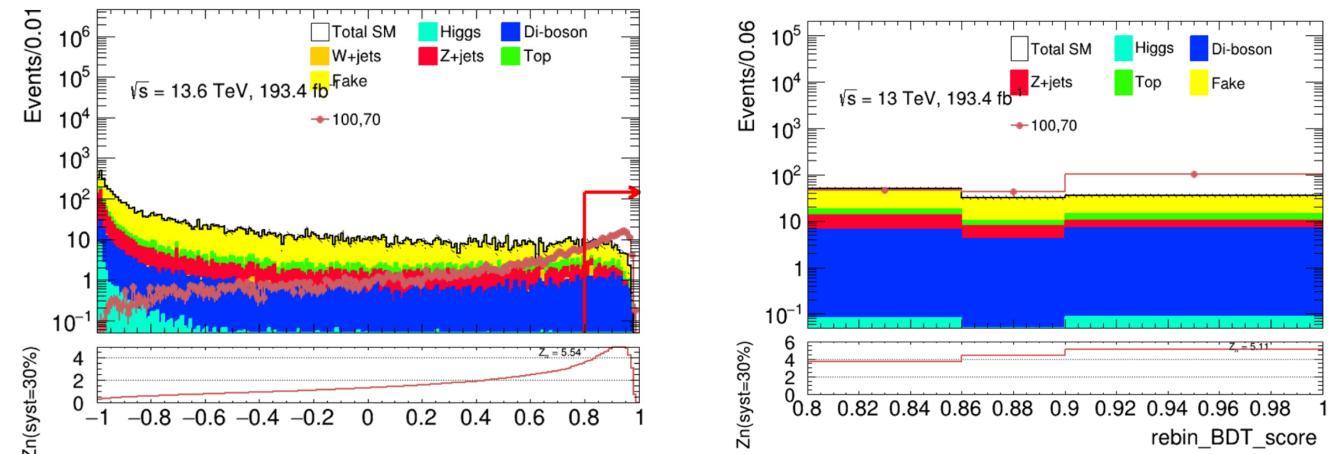
$$AUC_{gap} = |AUC_{train} - AUC_{validation}|$$

C1N2 ISR signal region definition(HH)

SR definition using BDT score(HH)

- Fake tau estimation using fake factor method
- BDT score > 0.8
- Rebin to three bins
 $[0.80, 0.86], [0.86, 0.90], [0.90, 1.00]$

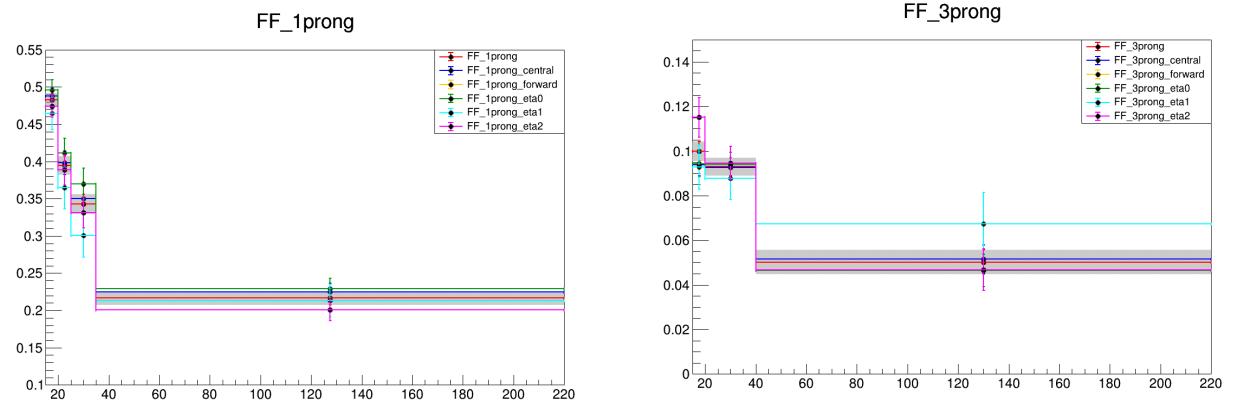
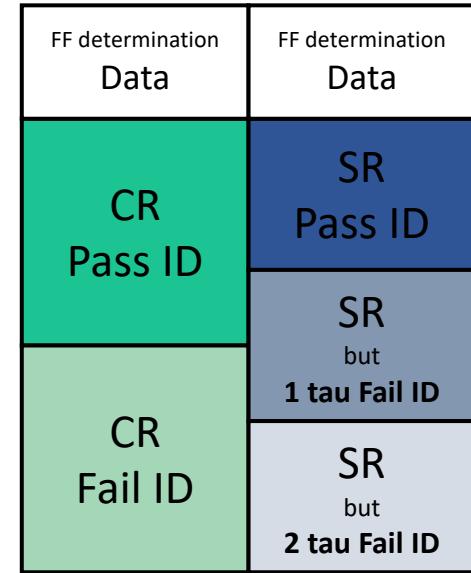
LH channel	HH channel
b-jet veto	
$\tau\tau(\ell)$ opposite sign	
MET trigger, MET $\geq 200 GeV$	
$m(\tau_1, \tau_2) \leq 40 GeV$ or $m(\tau_1, \tau_2) \geq 130 GeV$	
≥ 1 BaseJet, Leading Jet $p_t \geq 100$ (ISR)	
$\geq 1\tau_{had}$	$\geq 2\tau_{had}$
$\geq 1lep$	$= 0 lep$
$BDTscore \geq 0.80$	$BDTscore \geq 0.91$



SM Process	[0.80, 0.86]	[0.86, 0.90]	[0.90, 1.00]
VV	6.508 ± 0.368	4.124 ± 0.341	6.855 ± 0.398
Top	4.719 ± 0.686	2.663 ± 0.546	4.331 ± 0.731
Fake	31.456 ± 3.898	21.525 ± 3.292	20.438 ± 2.976
Higgs	0.080 ± 0.023	0.045 ± 0.019	0.086 ± 0.031
Z+jets	6.659 ± 0.416	3.564 ± 0.286	3.207 ± 0.250
W+jets	0.978 ± 0.978	0.000 ± 0.000	0.000 ± 0.000
Total Bkg	50.402 ± 4.042	31.921 ± 3.354	34.917 ± 3.118
$m(\tilde{\chi}_1^\pm, \tilde{\chi}_2^0) = (100, 70)$	46.569 ± 1.327	43.128 ± 1.283	101.059 ± 1.961
Z_n	3.65	4.37	5.11

C1N2 ISR fake estimation(HH)

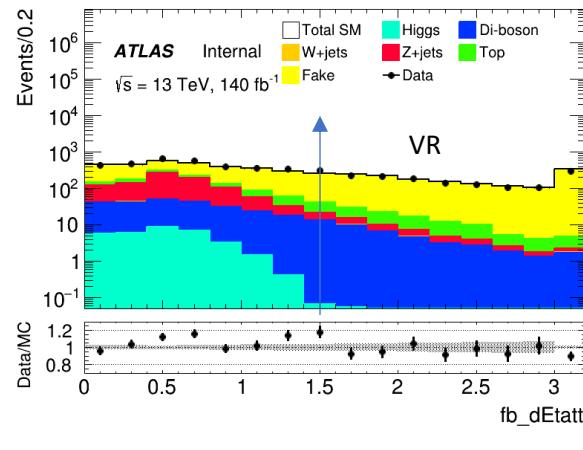
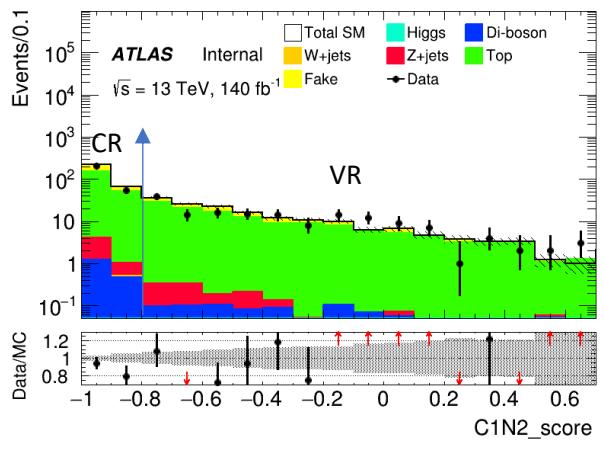
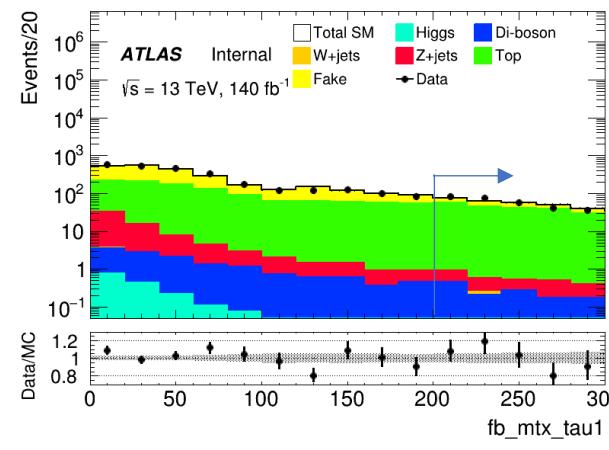
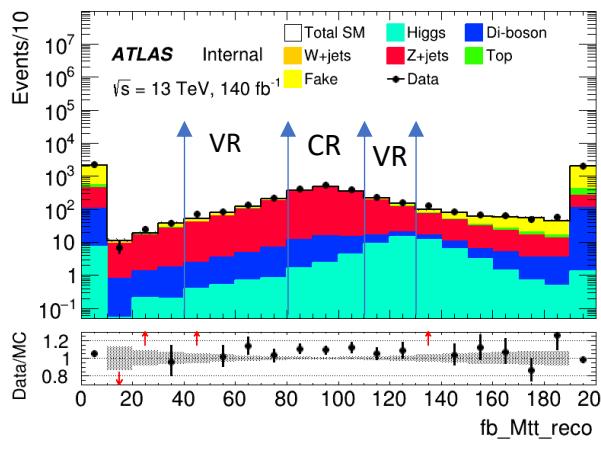
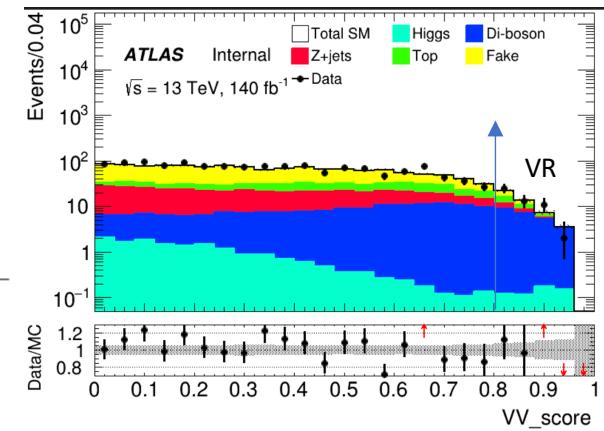
- CRs (fake factor computation)
 - METtrig
 - MET ≥ 200
 - bveto
 - at least 1 signal lepton
 - $\Delta\phi(\tau, \text{MET}) > 2$
 - **ID: ≥ 1 medium tau**
 - **antiID: ≥ 1 VeryLoose tau, 0 medium tau**
- SRs
 - preselection
 - 2ID: ≥ 2 medium tau
 - 1ID1antiID: ≥ 2 VeryLoose tau , 1 medium tau
 - 2antiID: ≥ 2 VeryLoose tau , 0 medium tau
- **Binned in prongness, tau eta, tau pT**
 - Eta bins
 - 2 bins: central [0,1.37], forward [1.52,2.5]
 - 3 bins: eta0,1,2 for [0,1), [1, 1.37], [1.52,2.5]
- **Auto binning:**
 - $> 10\%$ of events in nominator and denominator
 - Add bins to bin i until it is not consistent anymore with bin i - 1
 - Relative stat uncertainty on ratio smaller than 50%
 - $>10\%$ events in nominator and denominator



C1N2 ISR background estimation(HH)

Selection for control region and validation region

Process	Top		Z+jets		Multi-bosons	Fake		
	TCR	TVR	ZCR	ZVR				
Charge combination								
Trigger								
N medium τ			OS MET trigger, $E_T^{\text{miss}} \geq 200\text{GeV}$					
N lep					≥ 2			
n_{BaseJet}					$= 0$			
Jet p_T [GeV]					≥ 1			
N b-jets					≥ 100			
$m(\tau_1, \tau_2)$ [GeV]			≥ 1					
$M_T(\tau_1, E_T^{\text{miss}})$	≤ 40 or ≥ 130		[80, 110]	$[40, 80]$ or $[110, 130]$		$= 0$		
$d\eta(\tau_1, \tau_2)$	≥ 200					≤ 40 or ≥ 130		
C1N2 score	[-1, -0.8]	[-0.8, 0.7]				≤ 0.7		
VV score	-	-				≥ 0.80	-	
Total bkg	290+7	162+5	1420+10	1286+11	46+2	2000+28		
Dominant Bkg	206+4	131+3	1221+5	993+4	24+1	1851+28		
Purity	0.71	0.80	0.86	0.77	0.52	0.93		
Data	264	161	1559	1427	51	1950		
Data/MC	0.91	0.99	1.09	1.10	1.10	0.98		

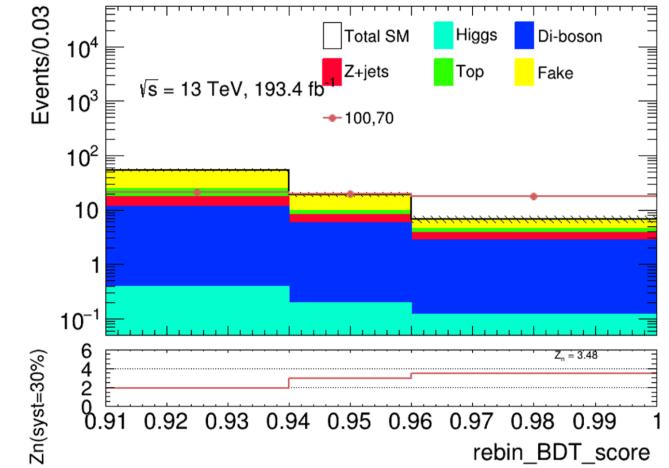
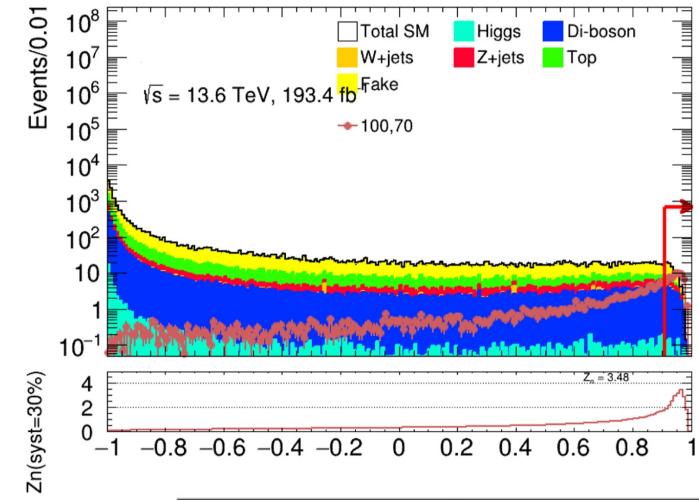


C1N2 ISR signal region definition(LH)

SR definition using BDT score(LH)

- Fake tau estimation using fake factor method
- BDT score > 0.91
- Rebin to three bins
[0.91, 0.94], [0.94, 0.96], [0.96, 1.00]

LH channel	HH channel
b-jet veto	
$\tau\tau(\ell)$ opposite sign	
MET trigger, MET $\geq 200 GeV$	
$m(\tau_1, \tau_2) \leq 40 GeV$ or $m(\tau_1, \tau_2) \geq 130 GeV$	
≥ 1 BaseJet, Leading Jet $p_t \geq 100$ (ISR)	
$\geq 1\tau_{had}$	$\geq 2\tau_{had}$
$\geq 1lep$	$= 0lep$
$BDTscore \geq 0.80$	$BDTscore \geq 0.91$



Process	[0.91, 0.94]	[0.94, 0.96]	[0.96, 1.00]
VV	11.105 ± 0.565	5.550 ± 0.367	2.666 ± 0.263
Top	7.432 ± 0.956	1.524 ± 0.441	0.578 ± 0.237
Fake	28.625 ± 3.295	9.404 ± 1.929	2.400 ± 1.009
Higgs	0.376 ± 0.057	0.189 ± 0.052	0.118 ± 0.040
Z+jets	5.515 ± 0.357	2.251 ± 0.222	1.026 ± 0.118
W+jets	0.173 ± 0.142	0.000 ± 0.000	0.000 ± 0.000
Total Bkg	53.226 ± 3.497	18.918 ± 2.018	6.788 ± 1.067
C1N2 (100,70)	21.447 ± 0.901	19.678 ± 0.874	17.543 ± 0.820
Z_n	1.90	2.94	3.47

C1N2 ISR background estimation(LH)

Selection for control region and validation region

Process	Top		Z+jets		Multi-bosons	Fake
	TCR	TVR	ZCR	ZVR	MBVR	FakeVR
Charge combination						
Trigger			OS MET trigger, $E_T^{miss} \geq 200\text{GeV}$			
N medium τ				≥ 1		
N lep				≥ 1		
$n_{BaseJet}$				≥ 1		
Jet p_T [GeV]				≥ 100		
N b-jets	≥ 1			$= 0$		
$m(\tau_1, l)$ [GeV]	≤ 40 or ≥ 130		[80,110]	[40,80] or [110,130]	≤ 40 or ≥ 130	
$M_{inv}(l, MET)$	≥ 300		—	—	—	—
$\Delta\phi(MET, \tau_1)$	—	—	—	—	—	≥ 2
$dR(\tau, l)$	—	—	≤ 0.6	≤ 0.7	—	—
C1N2 score	[-1, -0.95]	[-0.95, 0.7]				
VV score	—	—	—	—	≥ 0.80	—
Total bkg	4151+-11	1212+-13	917+-6	823+-4	78+-2	4983+-40
Dominant Bkg	3781+-20	1102+-11	823+-4	606+-3	53+-1	4390+-40
Purity	0.91	0.90	0.90	0.84	0.68	0.88
Data	3626	1119	908	745	72	5043
Data/MC	0.87	0.91	0.99	1.03	0.92	1.01



山东大学
SHANDONG UNIVERSITY

Backup



Top(HH) Binary

== 2 medium tau
== 0 lepton
METtrig && MET \geq 200
OS
nBaseJet \geq 1
Jet pt $>$ 100GeV
Mtt_reco $<$ 40GeV && Mtt_reco $>$ 130GeV

\geq 1 bTag(improve top events)

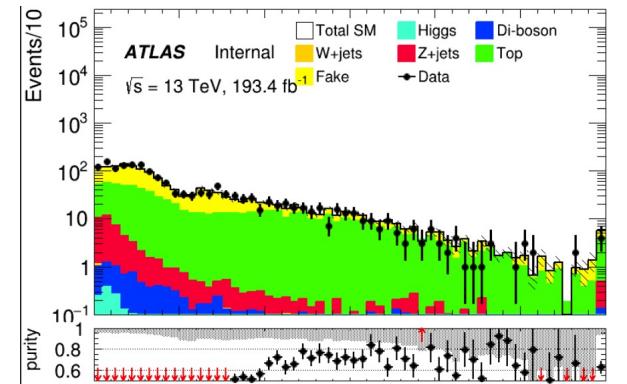
mtx_tau1 \geq 200(to improve top purity)

CR: C1N2 score \leq -0.8

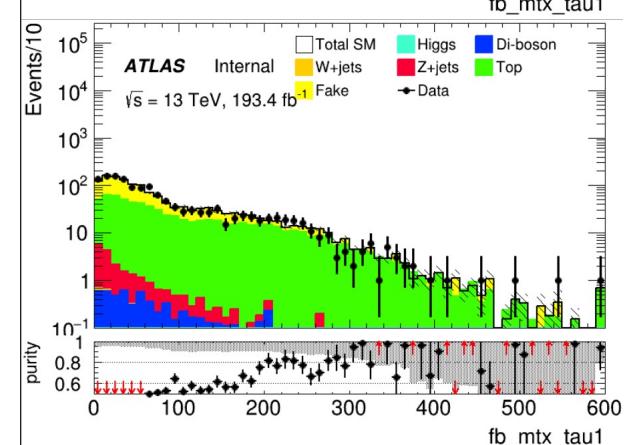
VR: -0.8 \leq C1N2 score \leq 0.7

Same with pre-selection

CR after score cut



VR after score cut



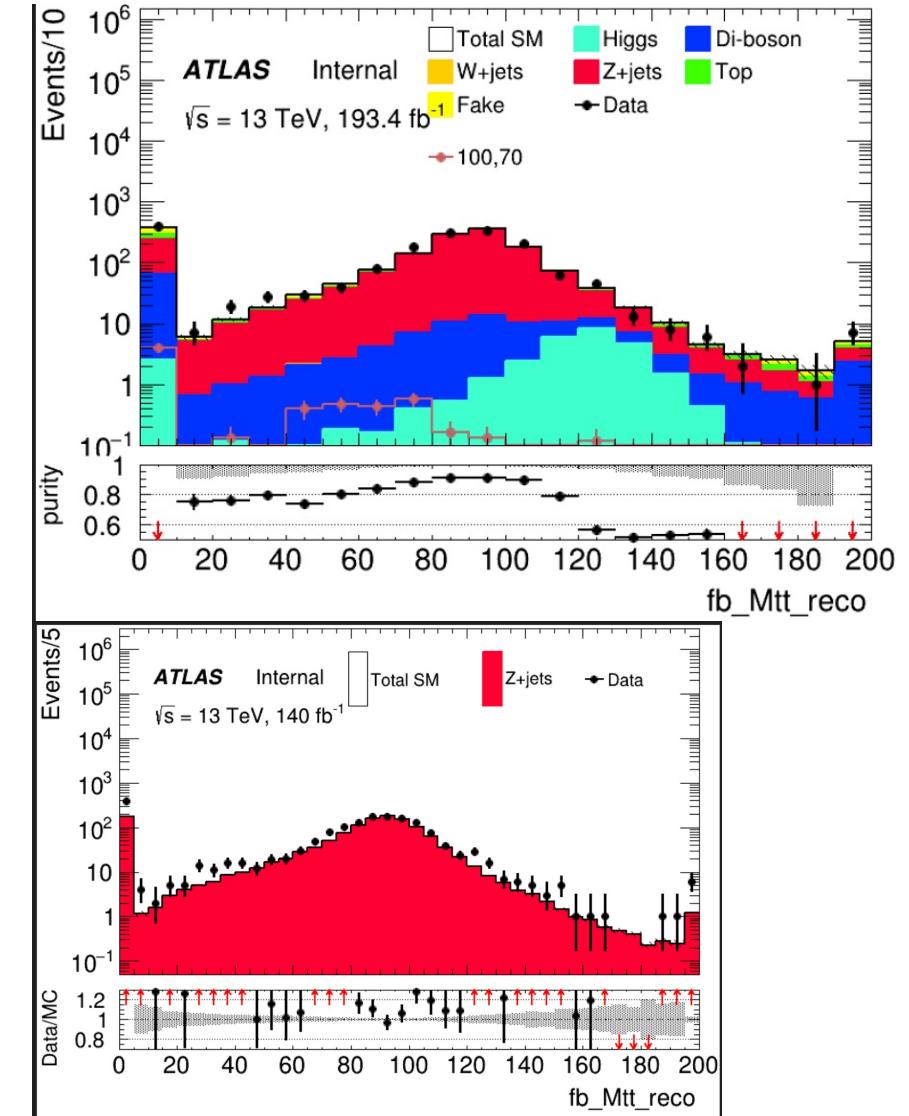
Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	290.841+-7.459	206.202+-4.807	0.709	264	0.91
VR	162.58+-5.438	131.125+-3.869	0.806	161	0.99

Zjets(LH) Binary

== 1 medium tau
 >= 1 lepton
 METtrig && MET >= 200
 OS
 bVeto
 nBaseJet >= 1
 Jet pt > 100GeV
 dRtt <= 0.6(improve purity of Zjets)
 C1N2 score <= 0.7(Othogonal with SR)
 CR: Mlt reco >= 80 && Mlt reco <= 110
 VR: (Mlt reco >= 40 && Mlt reco < 80) || (Mlt reco > 110 && Mlt reco < 130)

Same with pre-selection

Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	917.505+-6.668	823.435+-4.028	0.897	908	0.99
VR1	592.091+-5.617	518.684+-3.244	0.876	635	1.07
VR2	128.617+-2.384	87.882+-1.217	0.683	120	0.93
VR	720.708+-6.102	606.566+-3.464	0.842	745	1.034



Top(LH) Binary

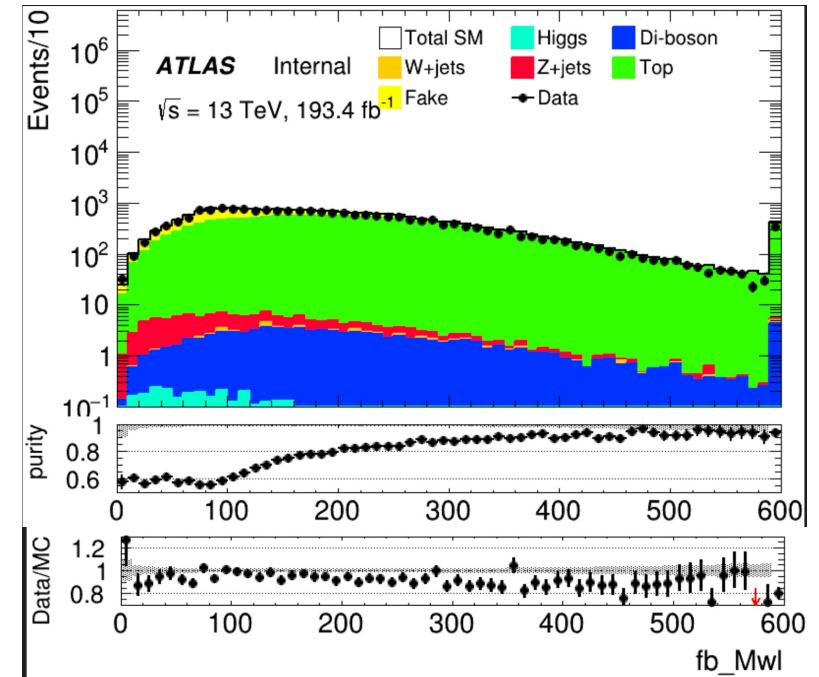
== 1 medium tau
>= 0 lepton
METtrig && MET >= 200
OS
nBaseJet >= 1
Jet pt > 100GeV
Mtt_reco < 40GeV && Mtt_reco > 130GeV

>= 1 bTag(improve top events)
C1N2 score <= 0.7(Othogonal with SR)

CR: $450 < M_{inv}(l, MET) < 560$

VR: $400 < M_{inv}(l, MET) < 450$

Same with pre-selection



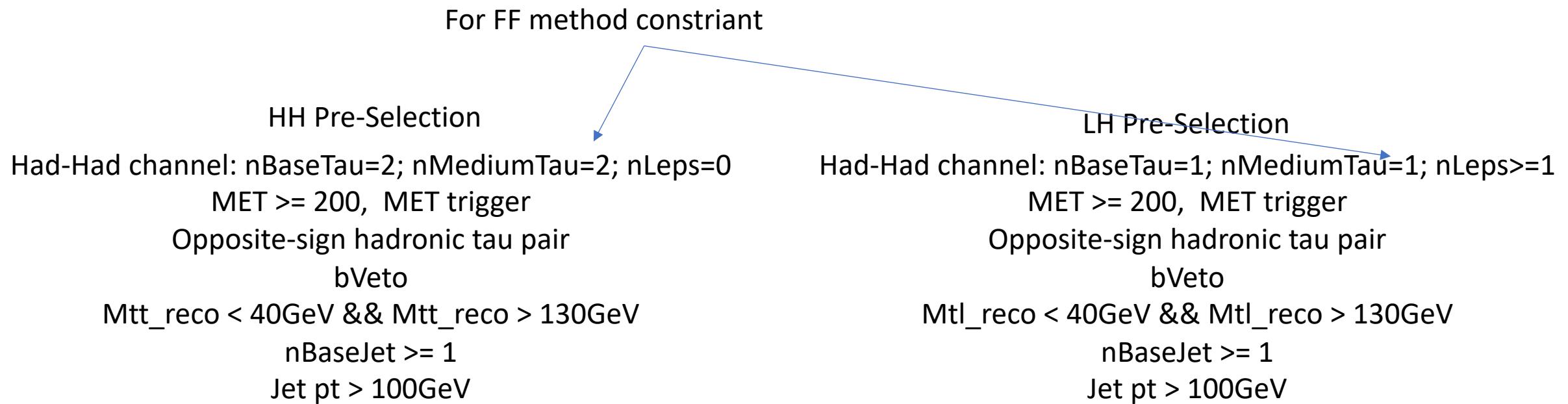
Try to define by using C1N2 score

Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	888.156+-11.505	830.914+-9.676	0.935	785	0.884
VR	902.627+-11.837	828.424+-9.679	0.912	792	0.878

Introduction

Reference point: C1 mass = 100GeV, N2 mass = 70 GeV

SR optimization: BDT method, optuna to auto-optimize hps



Binary class(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_dEtatt	: 5.153e-02
: 2	: fb_dRtt	: 4.318e-02
: 3	: fb_dRMax_xt	: 4.248e-02
: 4	: fb_dPhitt	: 4.228e-02
: 5	: fb_MIA	: 4.205e-02
: 6	: fb_METsig	: 3.979e-02
: 7	: fb_dPhizxe	: 3.972e-02
: 8	: fb_dPhiztt	: 3.942e-02
: 9	: fb_frac_MET_tau1	: 3.735e-02
: 10	: fb_dPhiMin_xt	: 3.513e-02
: 11	: fb_dPhiMin_tj1	: 3.512e-02
: 12	: fb_MT2_150	: 3.494e-02
: 13	: fb_frac_MET_MeffInc_40	: 3.474e-02
: 14	: fb_dRMin_tj	: 3.467e-02
: 15	: fb_eta_tau2	: 3.454e-02
: 16	: fb_frac_MET_tt	: 3.452e-02
: 17	: fb_frac_MET_Meff	: 3.408e-02
: 18	: fb_dPhit2x	: 3.277e-02
: 19	: fb_dPhiMax_xt	: 3.207e-02
: 20	: fb_dRt2x	: 3.131e-02
: 21	: fb_dPhit1x	: 3.089e-02
: 22	: fb_frac_MET_tau2	: 3.085e-02
: 23	: fb_Mll	: 2.960e-02
: 24	: fb_MET_Jet	: 2.734e-02
: 25	: fb_sum_cos_dphi	: 2.530e-02
: 26	: fb_pt_Vframe	: 2.272e-02
: 27	: fb_Pt_tt	: 1.912e-02
: 28	: fb_MstauA	: 1.881e-02
: 29	: fb_Proj_t1	: 1.594e-02
: 30	: fb_Proj_tt	: 1.427e-02
: 31	: fb_MCT	: 1.345e-02

Weight choose: no weight, abs(weight)

No weight have better performance
but abs(weight) fit our analysis requirement

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

Binary class(HH)

Hyperparameter tune:
use optuna to auto-optmize

constraint:

average of AUC need to ≥ 0.6

penalty function: $\text{score} = \text{test_auc} - 0.3 * \text{auc_gap}$ ($\text{auc_gap} = \text{abs}(\text{train_auc} - \text{test_auc})$)
 $\text{maximum}(\text{score})$

Class: C1N2, bkg

$\text{Test_auc} = \sum \{\text{Test_auc_class}\}$
 $\text{Train_auc} = \sum \{\text{Train_auc_class}\}$

Grid Search

Ntrees: [200, 300, 400]

MaxDepth: [4, 6, 8, 10]

MinNode: [1, 3, 5, 7]

Learning rate: [0.001, 0.005, 0.01, 0.05, 0.1]



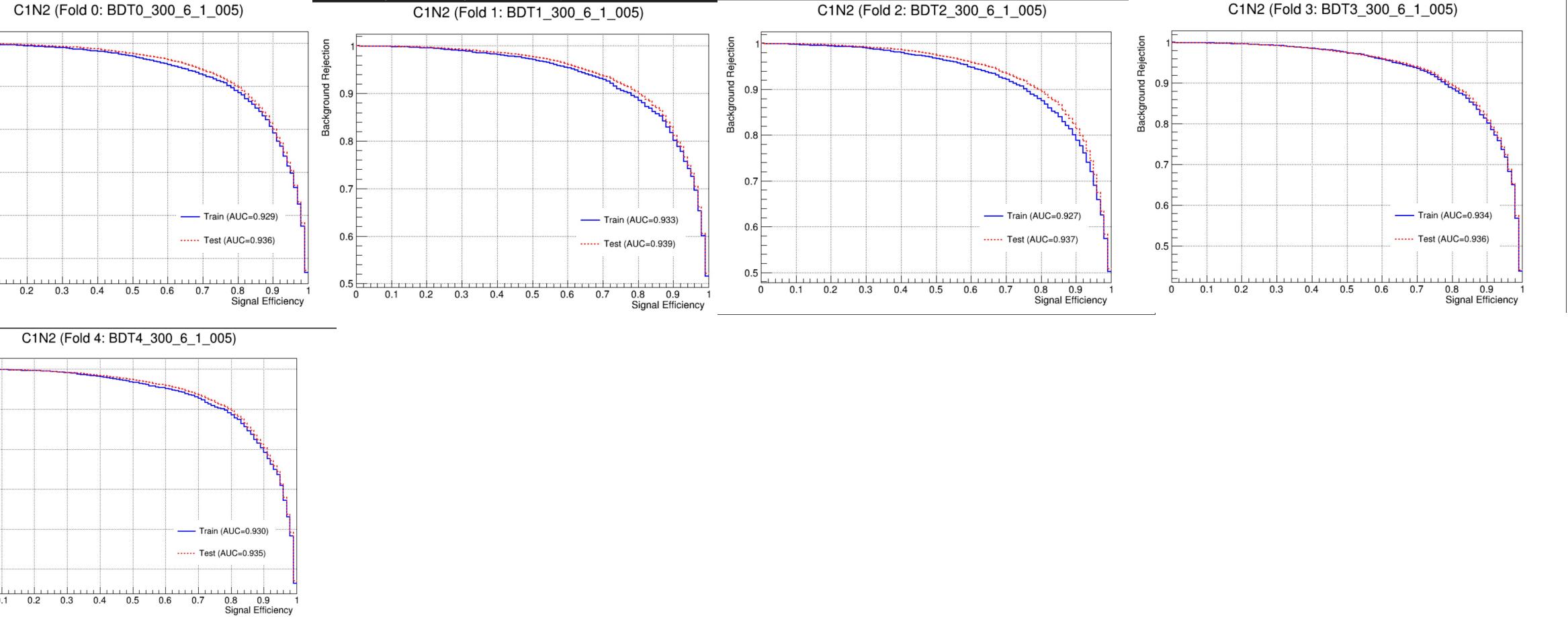
Best one: Ntree=300, MaxDepth=6, MinNode=1%, Learning Rate=0.05



There still have rooms to optimize for lr

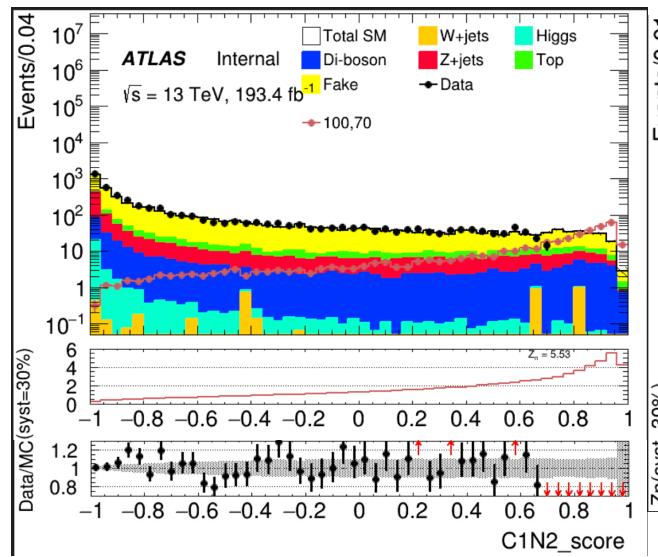
Binary class(HH)

Overfit Check

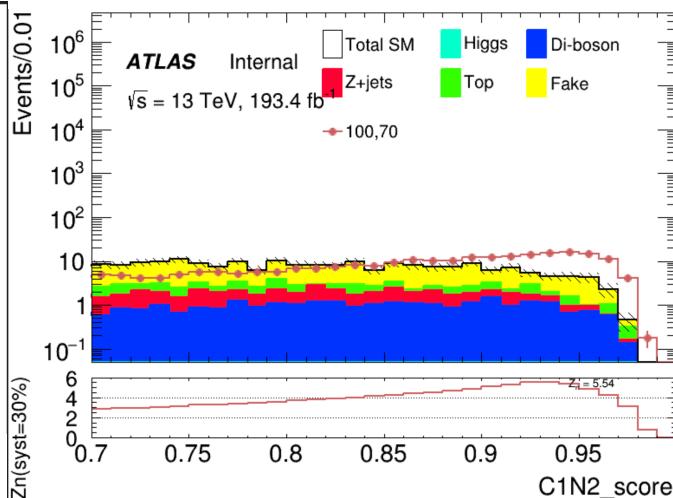


SR(HH) Binary

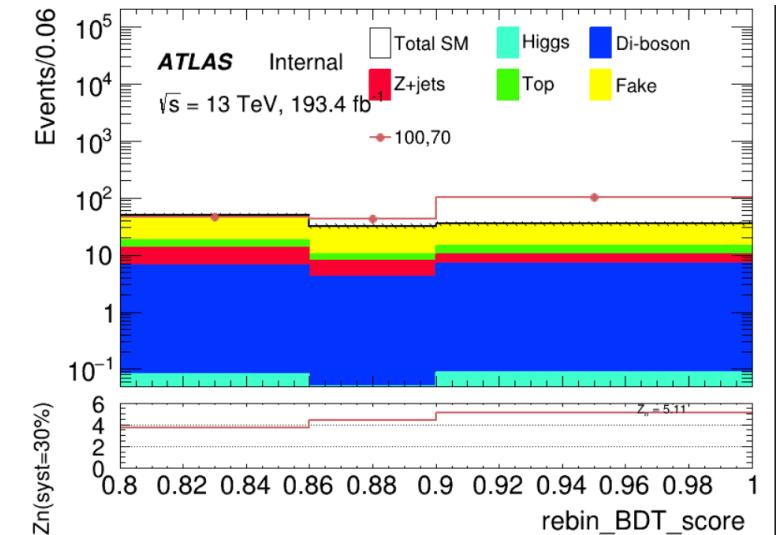
50 bins



Cut at 0.8



rebin



Sum Zn = 7.65

Bin Range	Zn	C1N2 (100_70) Yield ± Error	VV Yield ± Error	Top Yield ± Error	Fake Yield ± Error	Higgs Yield ± Error	Zjets Yield ± Error	Wjets Yield ± Error	Total Bkg Yield ± Error
[0.80,0.85]	3.65	46.569 ± 1.327	6.508 ± 0.368	4.719 ± 0.686	31.458 ± 3.898	0.080 ± 0.023	6.659 ± 0.416	0.978 ± 0.978	50.402 ± 4.042
[0.85,0.90]	4.37	43.128 ± 1.283	4.124 ± 0.341	2.663 ± 0.546	21.525 ± 3.292	0.045 ± 0.019	3.564 ± 0.286	0.000 ± 0.000	31.921 ± 3.354
[0.90,1.00]	5.11	101.059 ± 1.961	6.855 ± 0.398	4.331 ± 0.731	20.438 ± 2.976	0.086 ± 0.031	3.207 ± 0.250	0.000 ± 0.000	34.917 ± 3.118

Zjets(HH) Binary

== 2 medium tau

== 0 lepton

METtrig && MET ≥ 200

OS

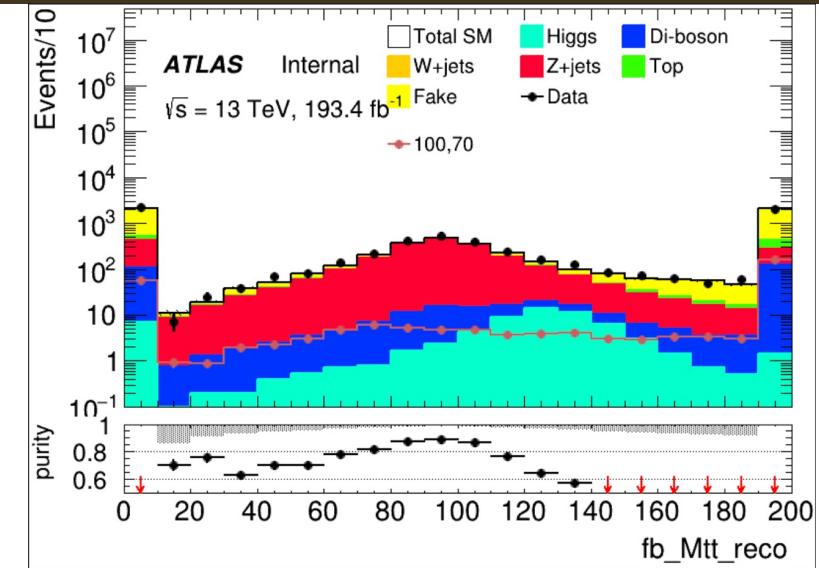
bVeto

nBaseJet ≥ 1

Jet pt $> 100\text{GeV}$

C1N2 score ≤ 0.7 (Othogonal with SR)

Same with pre-selection



CR: Mtt reco $\geq 80 \text{ && } \text{Mtt reco} \leq 110$

VR: (Mtt reco $\geq 40 \text{ && } \text{Mtt reco} < 80$) || (Mtt reco $> 110 \text{ && } \text{Mtt reco} < 130$)

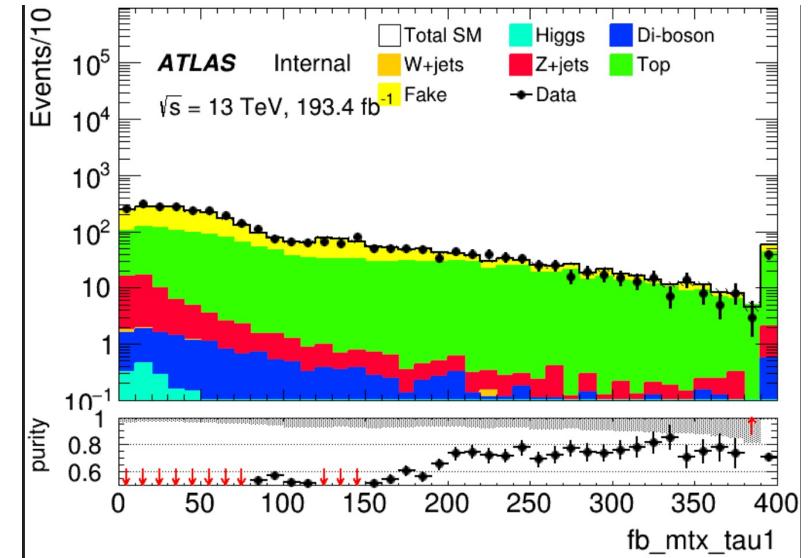
Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	1420.2+-10.9304	1221.51+-5.434	0.860	1559	1.09
VR1	821.998+-9.435	673.277+-4.073	0.819	904	1.10
VR2	465.375+-7.800	320.194+-2.908	0.688	523	1.12

Top(HH) Binary

== 2 medium tau
 == 0 lepton
 METtrig && MET \geq 200
 OS
 nBaseJet \geq 1
 Jet pt $>$ 100GeV
 Mtt_reco $<$ 40GeV && Mtt_reco $>$ 130GeV

 ≥ 1 bTag(improve top events)
C1N2 score \leq 0.7(Othogonal with SR)

Same with pre-selection



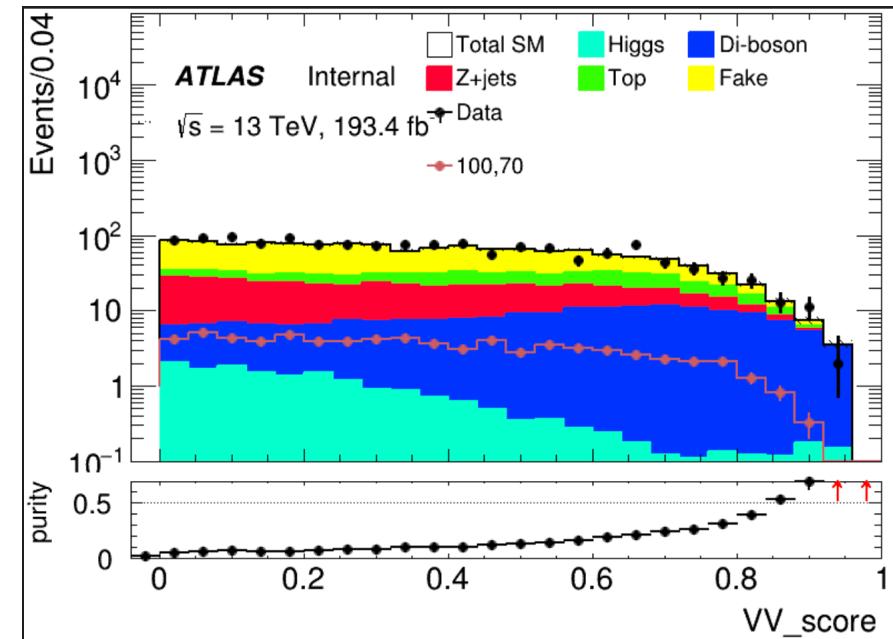
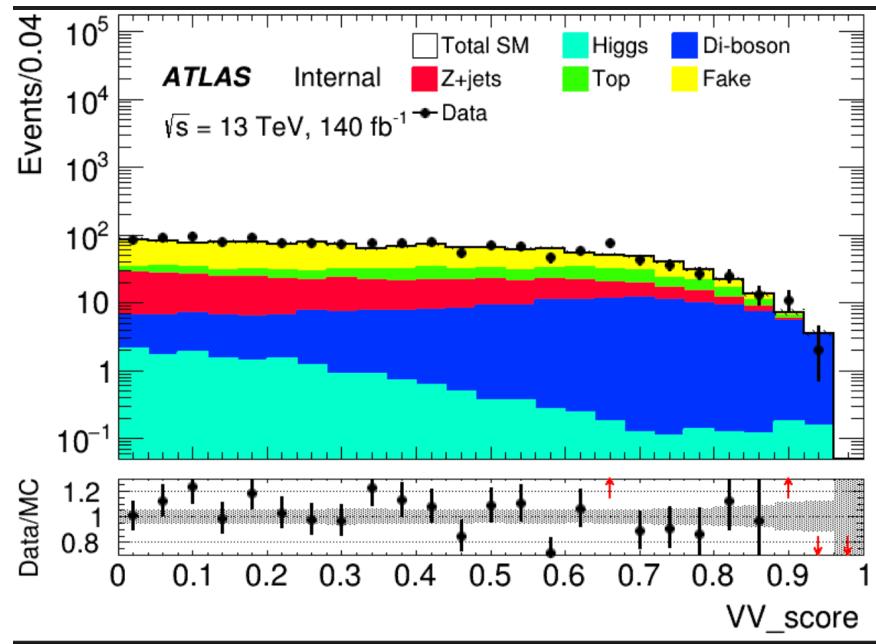
CR: $270 < M_T(\tau, MET) < 400$
 VR: $230 < M_T(\tau, MET) < 270$

Huge overestimation

Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	182.692+-5.738	140.981+-4.009	0.771	150	0.82
VR	140.228+-5.236	103.866+-3.445	0.740	134	0.94

VV(HH) Binary

Pre-selection && C1N2_score <= 0.7 && VV_score >= 0.80



Region	TotalBkg	VV	purity	Data	Data/Bkg
VR	46.6004+-2.77	24.193+-0.816	0.519	51	1.108

Binary class(LH)

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_frac_MET_tau2	: 8.270e-02
:	2 : fb_dRtt	: 6.684e-02
:	3 : fb_dPhitt	: 6.226e-02
:	4 : fb_frac_MET_tt	: 5.197e-02
:	5 : fb_frac_jet_tau2	: 5.179e-02
:	6 : fb_MT2_50	: 5.077e-02
:	7 : fb_dPhiMax_tj	: 4.779e-02
:	8 : fb_dPhiMin_xj	: 4.343e-02
:	9 : fb_mt_tau_min	: 3.547e-02
:	10 : fb_Mll	: 3.511e-02
:	11 : fb_mtx_tau1	: 3.408e-02
:	12 : fb_nBaseJet	: 3.146e-02
:	13 : fb_frac_jet_tt	: 3.110e-02
:	14 : fb_mtx_tau2	: 2.941e-02
:	15 : fb_frac_MET_tau1	: 2.898e-02
:	16 : fb_METsig	: 2.824e-02
:	17 : fb_pt_Vframe	: 2.726e-02
:	18 : fb_Mwh	: 2.684e-02
:	19 : fb_Proj_j	: 2.678e-02
:	20 : fb_frac_MET_sqrtHT_40	: 2.560e-02
:	21 : fb_frac_jet_tau1	: 2.518e-02
:	22 : fb_MCT	: 2.254e-02
:	23 : fb_Mwl	: 2.185e-02
:	24 : fb_mt_quad_sum	: 2.165e-02
:	25 : fb_Proj_tt	: 2.038e-02
:	26 : fb_ht_tau	: 1.992e-02
:	27 : fb_e_tau2	: 1.819e-02
:	28 : fb_mt_sum_ttj	: 1.624e-02
:	29 : fb_mt_tau2	: 1.618e-02

Weight choose: no weight, abs(weight)

No weight have better performance
but abs(weight) fit our analysis requirement

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

Binary class(LH)

Hyperparameter tune:
use optuna to auto-optmize

constraint:

average of AUC need to ≥ 0.6

penalty function: $\text{score} = \text{test_auc} - 0.3 * \text{auc_gap}$ ($\text{auc_gap} = \text{abs}(\text{train_auc} - \text{test_auc})$)
 $\text{maximum}(\text{score})$

Class: C1N2, bkg

$\text{Test_auc} = \sum \{\text{Test_auc_class}\}$
 $\text{Train_auc} = \sum \{\text{Train_auc_class}\}$

Grid Search

Ntrees: [200, 300, 400]

MaxDepth: [4, 6, 8, 10]

MinNode: [1, 3, 5, 7]

Learning rate: [0.001, 0.005, 0.01, 0.05, 0.1]



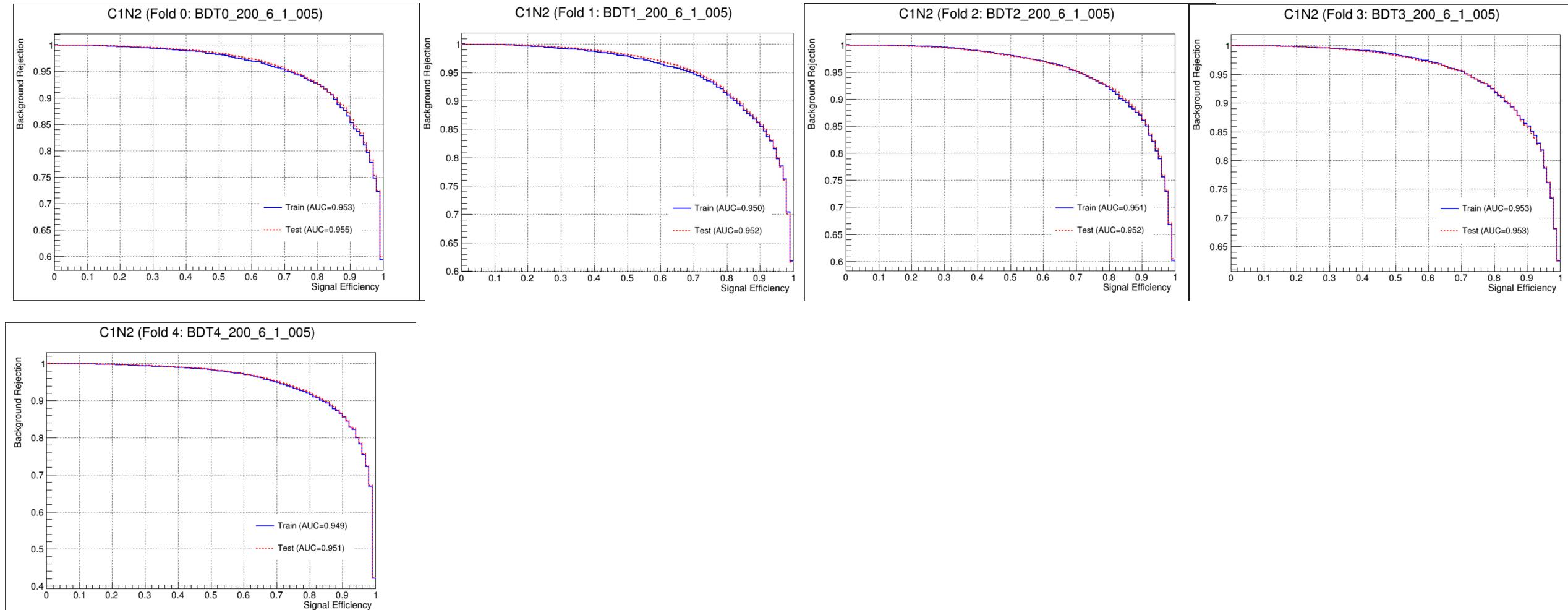
Best one: Ntree=200, MaxDepth=8, MinNode=1%, Learning Rate=0.05



There still have rooms to optimize for lr

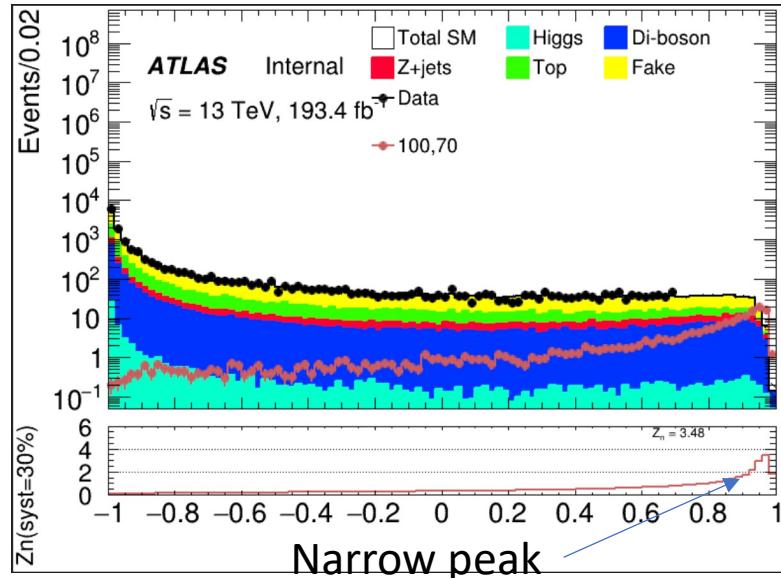
Binary class(LH)

Overfit Check

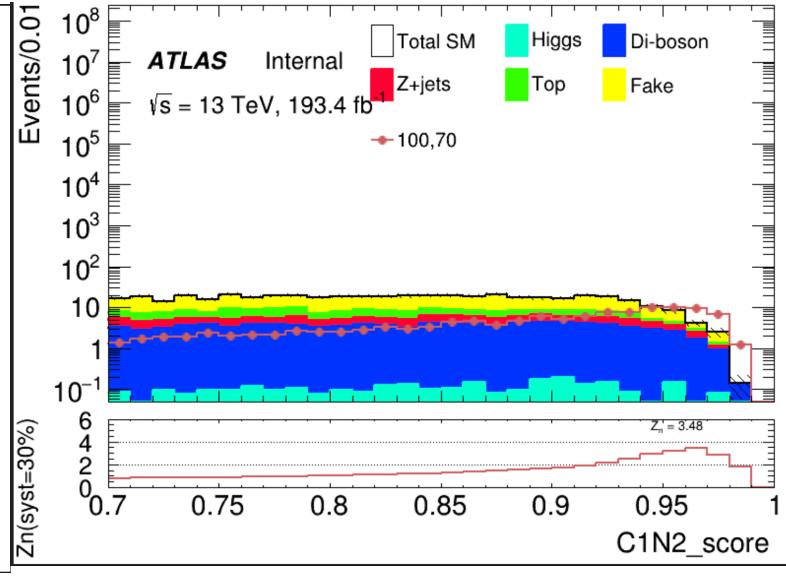


SR(LH) Binary

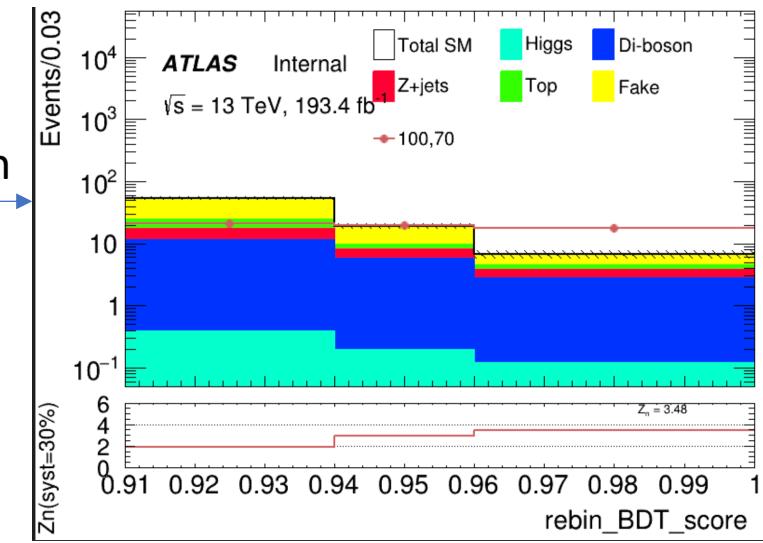
100 bins



Cut at 0.91



rebin



Sum Zn = 4.934

Bin Range	Zn	C1N2 (100_70) Yield ± Error	VV Yield ± Error	Top Yield ± Error	Fake Yield ± Error	Higgs Yield ± Error	Zjets Yield ± Error	Wjets Yield ± Error	Total Bkg Yield ± Error
[0.91,0.94]	1.90	21.447 ± 0.901	11.105 ± 0.565	7.432 ± 0.956	28.625 ± 3.295	0.376 ± 0.057	5.515 ± 0.357	0.173 ± 0.142	53.226 ± 3.497
[0.94,0.96]	2.94	19.678 ± 0.874	5.550 ± 0.367	1.524 ± 0.441	9.404 ± 1.929	0.189 ± 0.052	2.251 ± 0.222	0.000 ± 0.000	18.918 ± 2.018
[0.96,1.00]	3.47	17.543 ± 0.820	2.666 ± 0.263	0.578 ± 0.237	2.400 ± 1.009	0.118 ± 0.040	1.026 ± 0.118	0.000 ± 0.000	6.788 ± 1.067

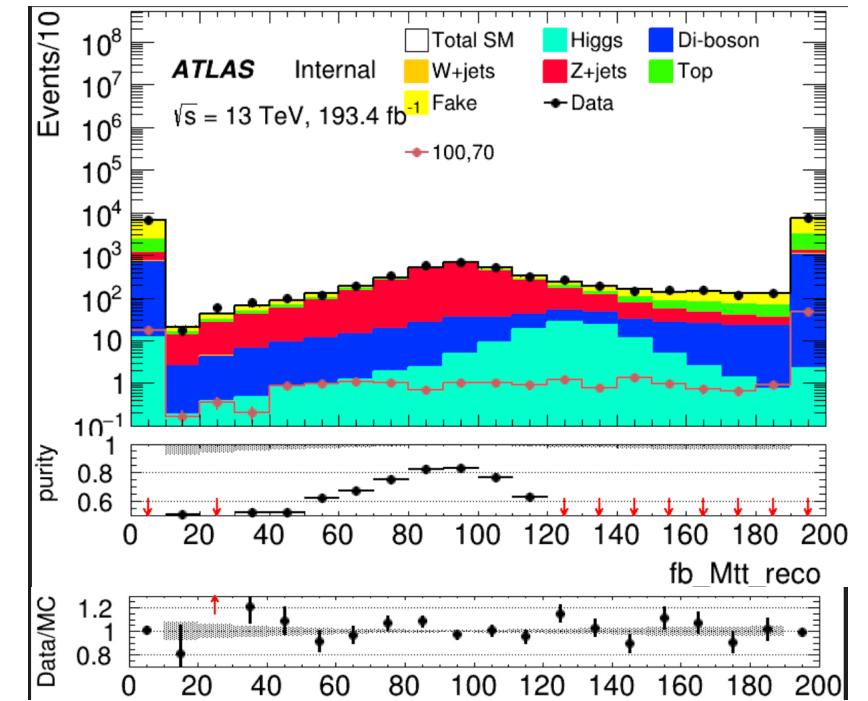
Zjets(LH) Binary

== 1 medium tau
 >= 1 lepton
 METtrig && MET >= 200
 OS
 bVeto
 nBaseJet >= 1
 Jet pt > 100GeV
 C1N2 score <= 0.7(Othogonal with SR)

Same with pre-selection

CR: Mlt reco >= 80 && Mlt reco <= 110
 VR: (Mlt reco >= 40 && Mlt reco < 80) || (Mlt reco > 110 && Mlt reco < 130)

Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	2048.51+-12.873	1600.7+-6.313	0.781	2064	1.0078
VR1	1238.27+-10.537	915.681+-4.726	0.739	1296	1.046
VR2	746.998+-9.369	385.547+-3.280	0.516	774	1.037



Top(LH) Binary

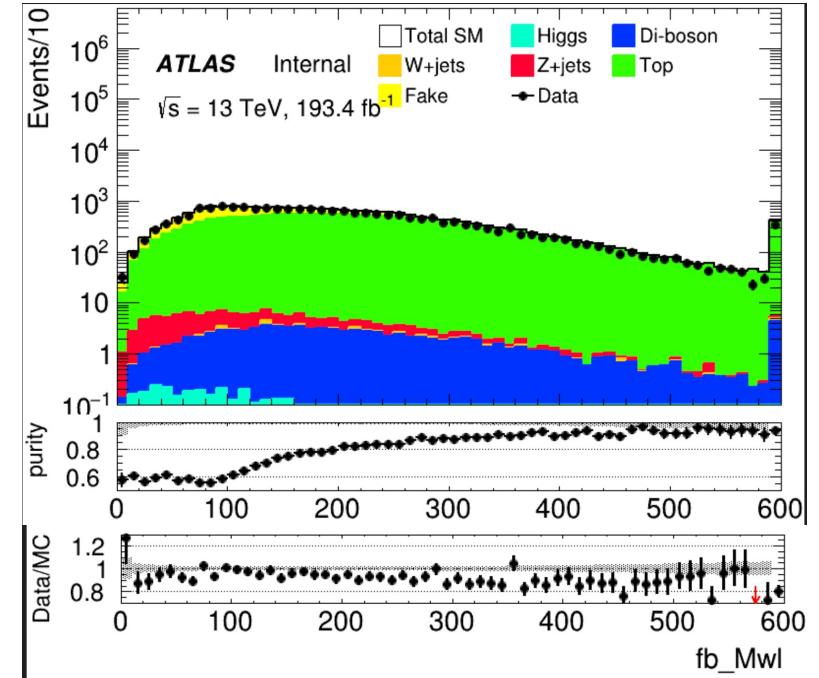
≤ 1 medium tau
 ≥ 0 lepton
 METtrig && MET ≥ 200
 OS
 nBaseJet ≥ 1
 Jet pt $> 100\text{GeV}$
 Mtt_reco $< 40\text{GeV} \&\& \text{Mtt}_\text{reco} > 130\text{GeV}$

≥ 1 bTag(improve top events)
 C1N2 score ≤ 0.7 (Orthogonal with SR)

CR: $300 < M_{inv}(l, MET) < 550$

VR: $250 < M_{inv}(l, MET) < 300$

Same with pre-selection

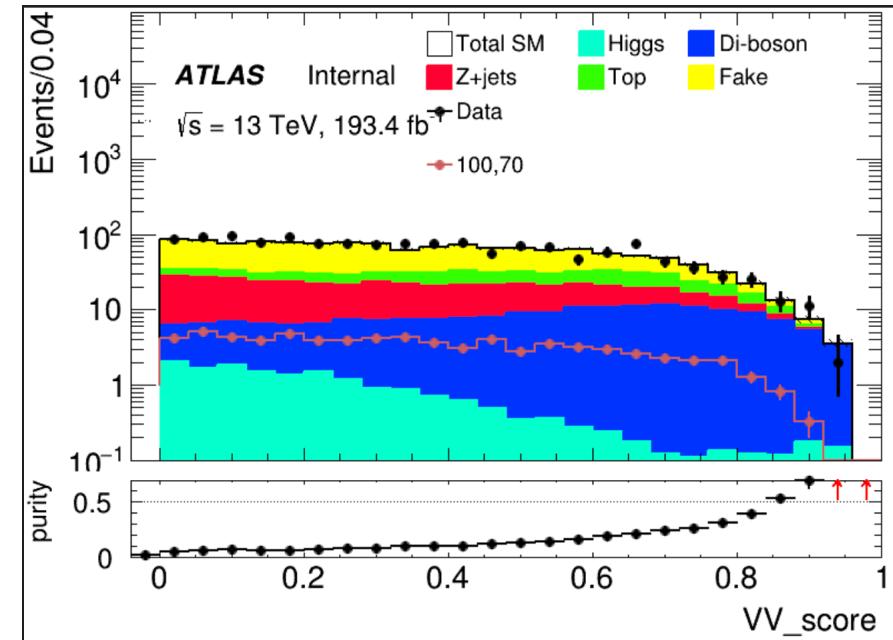
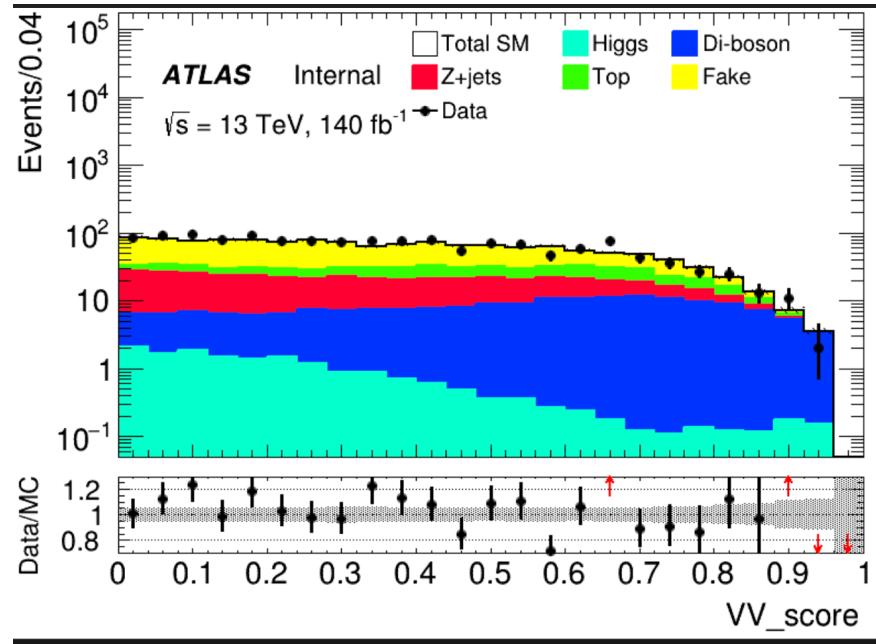


Huge overestimation

Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	4657.04+27.0028	4224.32+21.804	0.907	4152	0.891
VR	2916.32+22.0316	2532.7+-16.8751	0.868	2700	0.925

VV(LH) Binary

Pre-selection && C1N2_score <= 0.7 && VV_score >= 0.80



Region	TotalBkg	VV	purity	Data	Data/Bkg
VR	78.6819+2.26	53.8201+1.124	0.68402	72	0.92

Production issues

Task ID	Task name	N files total	N files done	N files failed	%	Status (JEDI)	Duration, days	Task logged status	Jobs failure, %	Top job errors, count [component:code] "sample message" [log example]
44282337	user.liaoc.542937.MGPy8EG_A14N23LO_C1N2ISR_100p0_0p0_2TFilt_run3.mc20d.r14860_p6284_stauhh_no_done_2/	11	0	0	0	exhausted	4.79	mc20_13TeV:mc20...	0 -	
44282315	user.liaoc.542971.MGPy8EG_A14N23LO_C1N2ISR_180p0_80p0_2TFilt_run3.mc20a.r13167_p6284_stauhh_no_done_2/	2	0	0	0	exhausted	4.79	mc20_13TeV:mc20...	0 -	
44281469	user.liaoc.542937.MGPy8EG_A14N23LO_C1N2ISR_100p0_0p0_2TFilt_run3.mc20d.r14860_p6284_stauhh_no_done_1/	11	0	0	0	exhausted	4.89	mc20_13TeV:mc20...	0 -	
44281451	user.liaoc.542971.MGPy8EG_A14N23LO_C1N2ISR_180p0_80p0_2TFilt_run3.mc20a.r13167_p6284_stauhh_no_done_1/	2	0	0	0	exhausted	4.89	mc20_13TeV:mc20...	0 -	

100_0_mc20d, 180_80_mc20a

Submit them twice but still failed

Bkg decay mode

Wjets: $W \rightarrow e/\mu\text{on} + \nu$

$W \rightarrow \tau + \nu$ (can contribute true τ_{had})
jet misidentified to a fake tau

Zjets: $Z \rightarrow ll/\tau\tau\tau\tau$

jet misidentified to fake tau

Top: $\text{top} \rightarrow W+b$, W can contribute a true τ_{had}
b-quark is a source of fake

VV: W/Z

LH channel: $\geq 1\tau, \geq 1\text{lep}$

Wjets: W contribute lep, jets misidentified to fake

Zjets:

SingleTop: W contribute lep, b-quark misidentified to fake

VV:

HH channel: $\geq 2\tau, == 0\text{lep}$

Wjets: W contribute τ_{had} , plus a fake tau

Zjets: $Z \rightarrow \tau\tau\tau\tau(\text{had})$ or 2 fake tau

SingleTop: W contribute a τ_{had} , plus a fake tau

VV:

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_dRtt	: 8.238e-02
2	: fb_dRMax_xt	: 7.068e-02
3	: fb_METsig	: 6.205e-02
4	: fb_frac_MET_tt	: 6.050e-02
5	: fb_dPhi1x	: 5.751e-02
6	: fb_MIA	: 5.460e-02
7	: fb_mt_taumin	: 5.411e-02
8	: fb_Asy_tt	: 5.363e-02
9	: fb_dEtat2j	: 4.903e-02
10	: fb_MET_Soft	: 4.737e-02
11	: fb_Asy_EH	: 4.625e-02
12	: fb_Mll	: 4.447e-02
13	: fb_frac_MET_MeffInc_40	: 4.317e-02
14	: fb_eta_jet2	: 4.282e-02
15	: fb_eta_jet1	: 4.229e-02
16	: fb_transSphericity	: 4.140e-02
17	: fb_dRMax_jets	: 4.086e-02
18	: fb_frac_MET_Meff	: 4.007e-02
19	: fb_m_jet1	: 3.421e-02
20	: fb_nJets30	: 3.260e-02
:		

Weight choose: no weight, abs(weight)

No weight have better performance
but abs(weight) fit our analysis requirement

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

Multiclass(HH)

Hyperparameter tune:
use optuna to auto-optimize

constraint:

average of AUC need to ≥ 0.6

penalty function: $\text{score} = \text{test_auc} - 0.3 * \text{auc_gap}$ ($\text{auc_gap} = \text{abs}(\text{train_auc} - \text{test_auc})$)
 $\text{maximum}(\text{score})$

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, 0.005, 0.01, 0.05, 0.1]



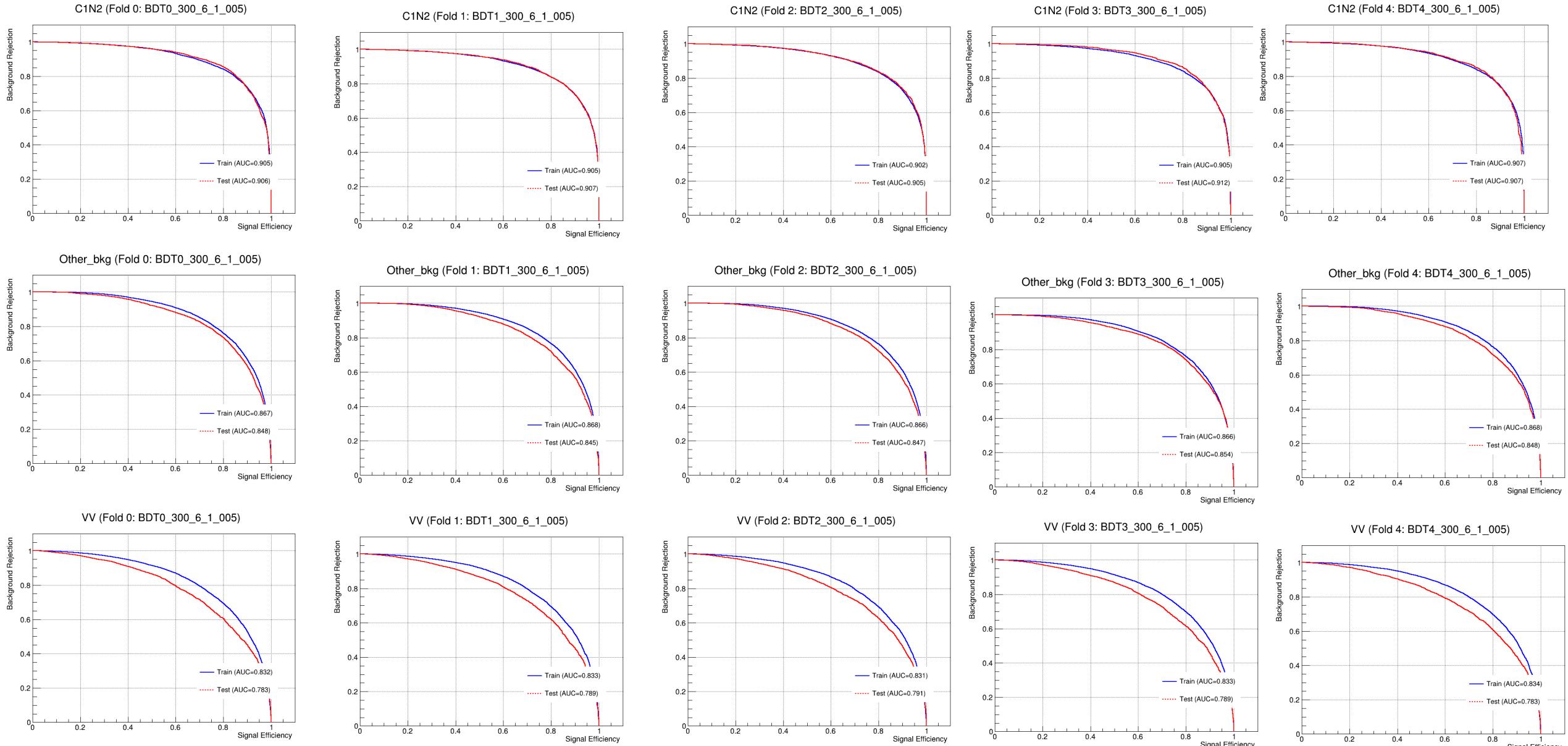
Best one: Ntree=300, MaxDepth=6, MinNode=1%, Learning Rate=0.05



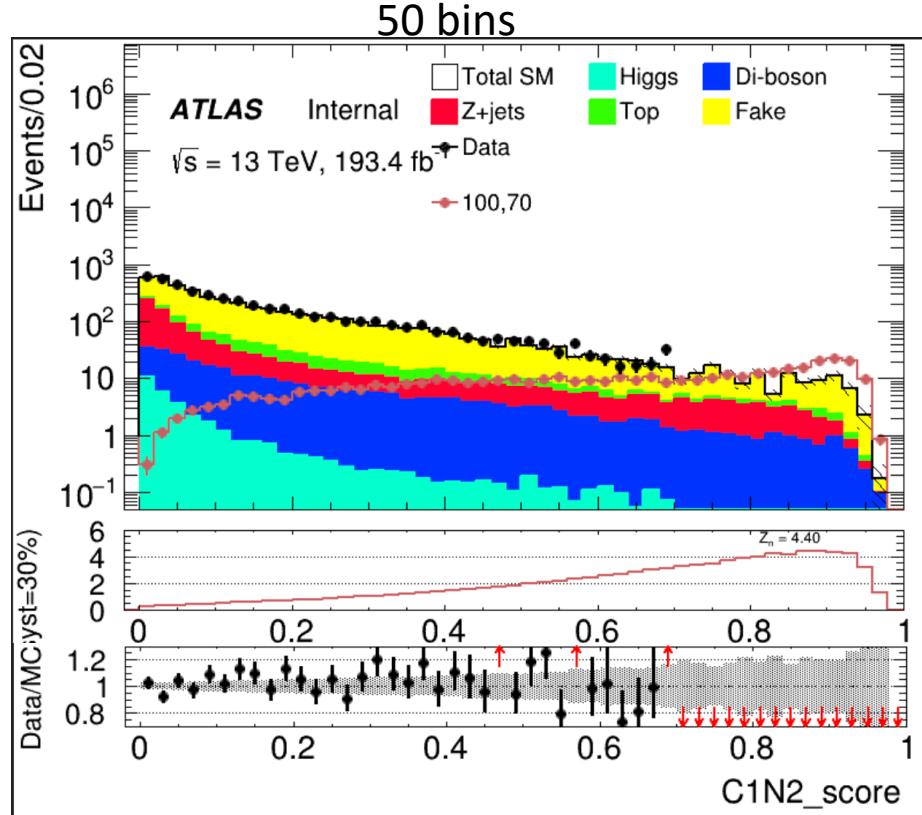
There still have rooms to optimize for lr

Multiclass(HH)

OverFit Check



SR(HH) Multi

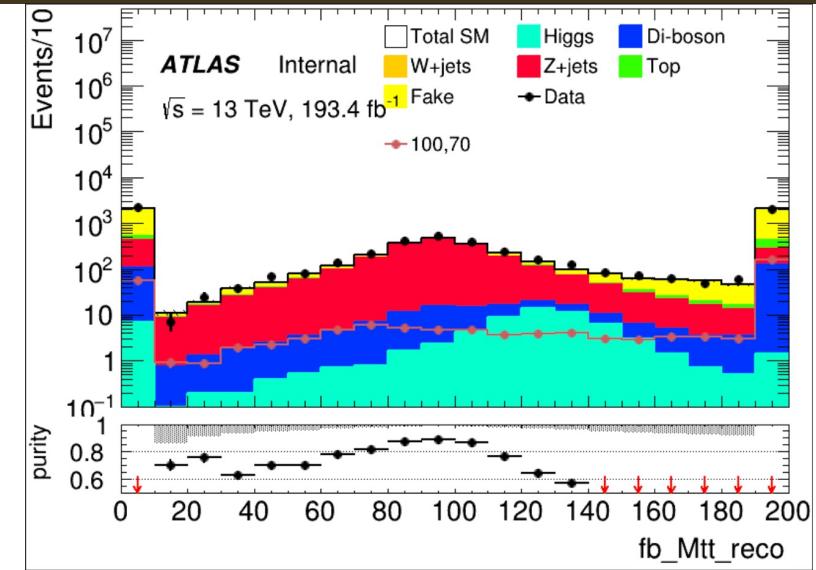


Each index stand for a region, and index 10 stands for sum

Zjets(HH)

== 2 medium tau
 == 0 lepton
 METtrig && MET ≥ 200
 OS
 bVeto
 nBaseJet ≥ 1
 Jet pt $> 100\text{GeV}$
 C1N2 score ≤ 0.7 (Othogonal with SR)

Same with pre-selection



CR: Mtt reco $\geq 80 \text{ && } \text{Mtt reco} \leq 110$

VR: (Mtt reco $\geq 40 \text{ && } \text{Mtt reco} < 80$) || (Mtt reco $> 110 \text{ && } \text{Mtt reco} < 130$)

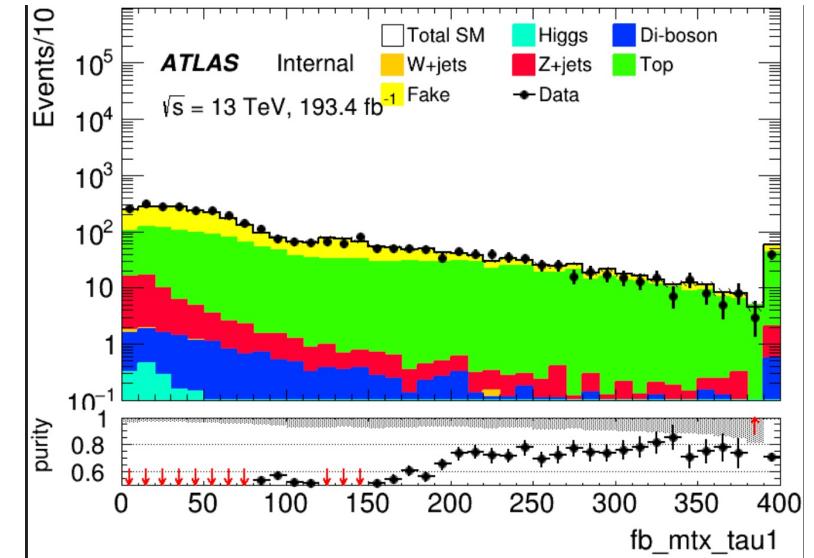
Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	1433.93+-11.003	1232.36+-5.476	0.859	1571	1.096
VR1	824.938+-8.468	674.307+-4.082	0.817	904	1.097
VR2	469.252+-7.834	321.877+-2.913	0.685	524	1.117

Top(HH)

== 2 medium tau
 == 0 lepton
 METtrig && MET ≥ 200
 OS
 nBaseJet ≥ 1
 Jet pt $> 100\text{GeV}$
 Mtt_reco $< 40\text{GeV} \&\& \text{Mtt}_\text{reco} > 130\text{GeV}$

 ≥ 1 bTag(improve top events)
 C1N2 score ≤ 0.7 (Orthogonal with SR)

Same with pre-selection



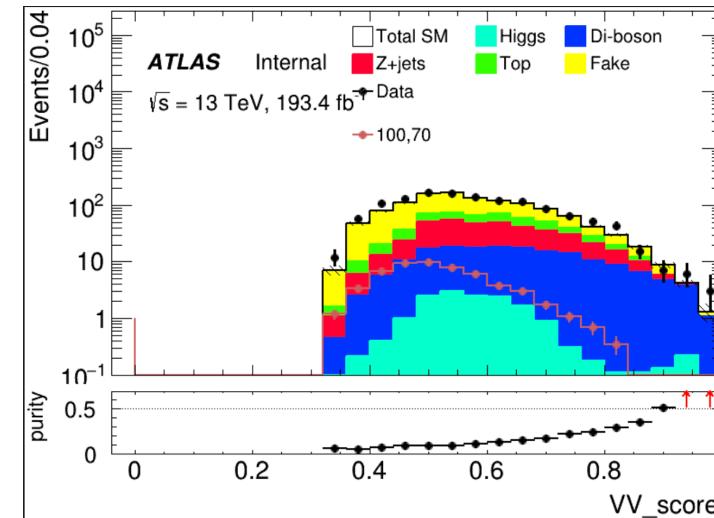
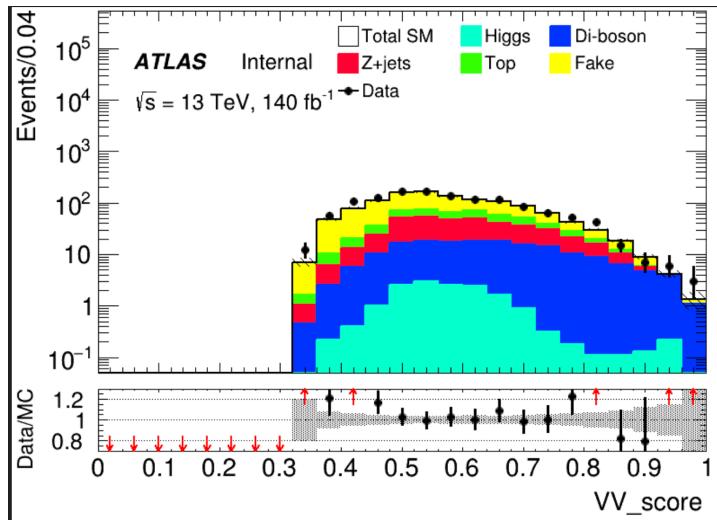
CR: $270 < M_T(\tau, \text{MET}) < 400$
 VR: $230 < M_T(\tau, \text{MET}) < 270$

Huge overestimation

Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	182.692+-5.738	140.981+-4.009	0.771	150	0.82
VR	140.228+-5.236	103.866+-3.445	0.740	134	0.94

VV(HH)

Pre-selection && C1N2_score <= 0.7 && Max(score) == VV_score && VV_score >= 0.85



Region	TotalBkg	VV	purity	Data	Data/Bkg
VR	32.4944+-2.09	15.4902	0.476	31	0.954