Weekly Report

Xin Shuiting

Backgrounds estimation on 2LSS analysis

- Primary unblind results with fake factor method

- Download missing part of ttbar , Zjets, problems remained with mc samples.

$t\bar{t}$	Two tight leptons with opposite sign leptons $(M(\ell\ell) - M(Z)) > 20$ GeV At least two b-jets
Zjets	Two tight leptons with opposite sign and same flavor $(M(\ell\ell) - M(Z)) < 45$ GeV b veto, at least one jet

bad validation of MC .

Backup

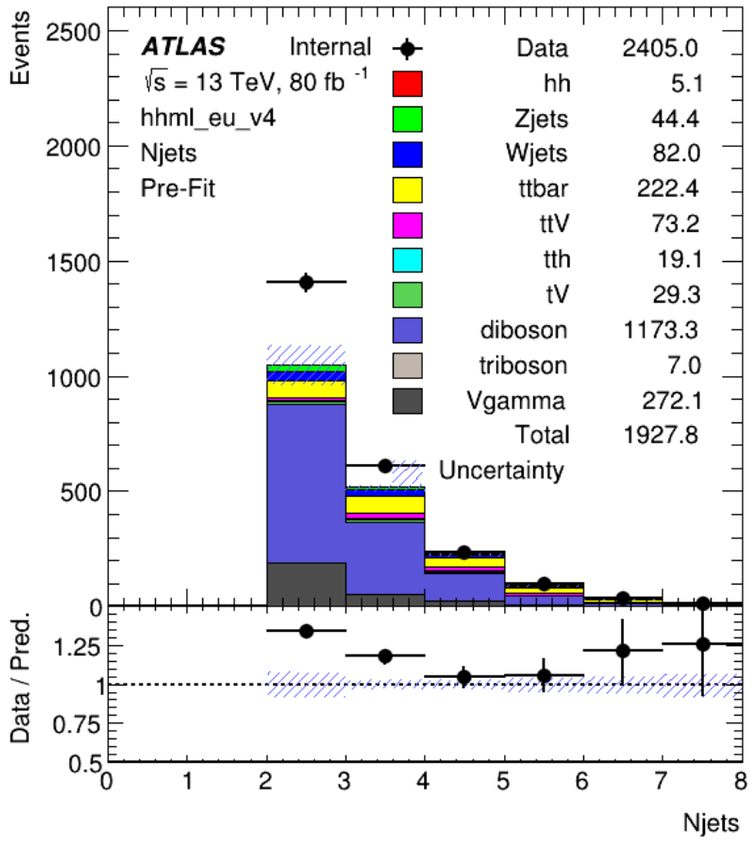
MC sample lists Group v2 mc16a*mc16d samples

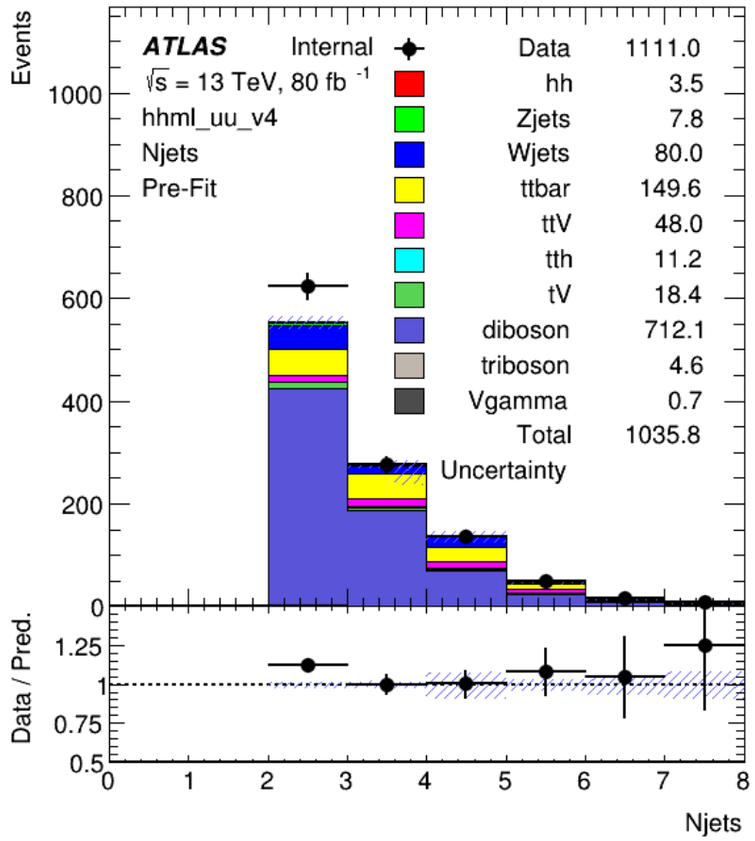
Process	DSID
ZJets	364100-41 364198-215
Wjets	364156-97
ttbar	410470,410471,410472
ttV	410218,410219,410220,410155,410156,410157
tth	345873,345874,345875
tV	410560,410646,410647
diboson	364250,364253-5,364283-7;363355-60, 363489
triboson	364242-364249
Vgamma	364500-364535
hh	450578,450661,450662,450663

- Refer to ttH mc https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HtopMC13TeVMC16
- Scale to 80fb-1 by weight=scale_nom* efficiency* lumi
- But some DSID corrupted, Z->ll, diboson->ZqqZll etc. Times a factor to rescale currently. Hope we can use all samples in the future.

data/MC plots at pre-selection level

Add two BDT variables to SR, PLV and QmisID



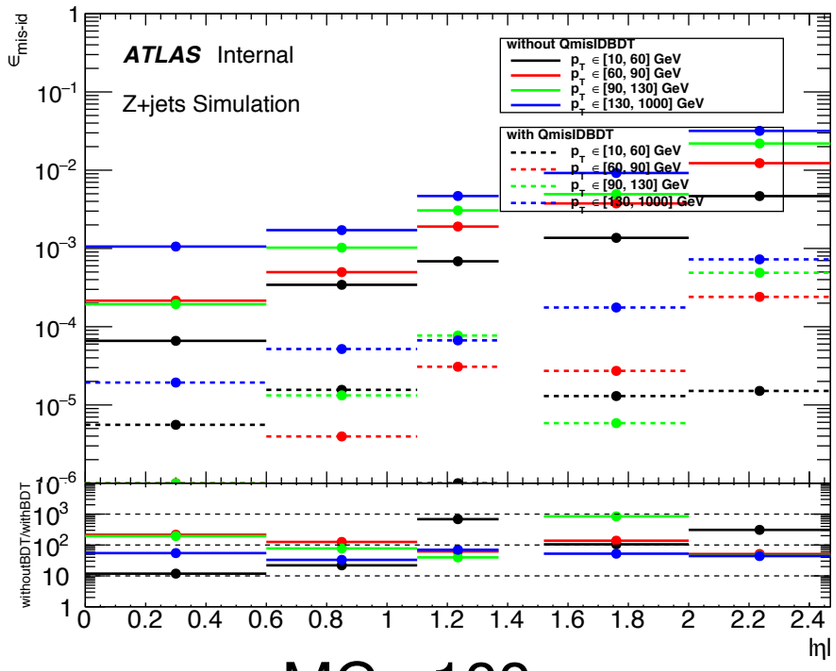


Discrepancy due to : not well modeled with MC samples
Need to apply data-driven methods to estimate QmisID and jet fakes

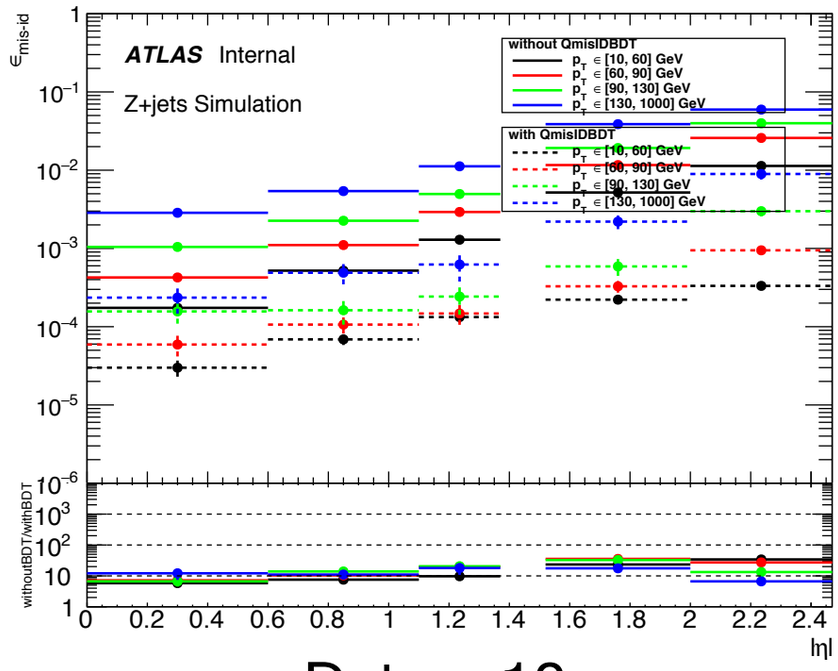
Common selections: tight lepton ,MET,Mll ,b veto…same as previous 36.1fb-1 set
Change: lep_pt>15GeV, PLV <-0.4, QmiID >0.3

Electron QmisID(done a few month)

improvement from QmisID BDT



MC ~100



Data ~ 10

- extract QmisID events numbers from the rate above by pt and nta
- Will be subtracted when analysis fake background

with MC truth, classify background into

- fakes :ttbar, Wjets (heavy b decay)
- QMisID: Zjets,ttbar
- Prompt: diboson ,tV…
- Vgamma

Fake factor method attempt

- fake factor definition $\theta_\ell = \frac{N_{\ell\ell}}{N_{\ell\not\ell}}$

e
- tight:TightLH,isolationFixedCutTight,ChargeIDBDTTight>0.3
- Anti-tight: Fail TightLH,ChargeIDBDTTight>0.3

muon
- tight:Tight,isolationFixedCutTightTrackOnly
- Anti-tight: Fail isolationFixedCutTightTrackOnly

CR(Njet=1,1<=Njet<=2) → **SR(Njet>=2,Njets>=3)**

Fake factor in CR

$$\theta_e(1 \le N_{\text{jet}} \le 2) = \frac{N_{ee}^{\text{data}} - N_{ee}^{\text{promptSS}} - N_{ee}^{V\gamma} - N_{ee}^{\text{QmisID}}}{N_{e\not e}^{\text{data}} - N_{e\not e}^{\text{promptSS}} - N_{e\not e}^{V\gamma} - N_{e\not e}^{\text{QmisIDMC}}} \quad \theta_\mu(1 \le N_{\text{jet}} \le 2) = \frac{N_{\mu\mu}^{\text{data}} - N_{\mu\mu}^{\text{promptSS}} - N_{\mu\mu}^{V\gamma} - N_{\mu\mu}^{\text{QmisID}}}{N_{\mu\not\mu}^{\text{data}} - N_{\mu\not\mu}^{\text{promptSS}} - N_{\mu\not\mu}^{V\gamma} - N_{\mu\not\mu}^{\text{QmisIDMC}}}$$

The value of ff depends on selection much

Fake factor In two different selection

- With PLV cut

Seclction	Fake factor	Value
Njet=1	e	0.4619pm0.026
	muon	2.8604pm0.319
1<=Njet<=2	e	0.4830pm0.0023
	muon	2.9410pm0.2872

CR

Njet=1	VV	ttV	tV	ttH	Vgam	Data
mumu	572.55\pm4.80	4.89\pm0.19	5.17\pm0.94	0.40\pm0.02	0.13\pm0.81	893.00
mu+antiidmu	25.73\pm0.91	0.21\pm0.04	1.00\pm0.45	0.03\pm0.00	-1.29\pm1.29	134.00

SR

Njet>=2	VV	ttV	tV	ttH	Vgam	Data
mumu +antiid	24/81\pm0.58	1.92\pm0.13	3.64\pm0.94	0.72\pm0.14	1.57\pm1.16	116.00\pm10.77

- Without PLV cut

Seclction	Fake factor	Value
Njet=1	e	0.1734pm0.006
	muon	0.1372pm0.0054
1<=Njet<=2	e	0.1644pm0.0047
	muon	0.1236pm0.0041

CR

	VV	ttV	tV	ttH	Vgam	Data
mumu	995.22\pm5.76	15.54\pm0.33	35.84\pm2.84	1.90\pm0.14	8.00\pm4.61	2085.00
mu+antiidmu	131.93\pm2.92	2.96\pm0.17	254.41\pm7.91	0.55\pm0.10	48.85\pm10.99	8763.00

SR

Njet>=2	VV	ttV	tV	ttH	Vgam	Data
mumu +antiid	89.98\pm2.35	10.21\pm0.39	277.89\pm8.16	3.27\pm0.20	36.64\pm7.76	6317.00\pm79.48

Events yields with PLV, only statistics

Njets>=2	ee	mumu	emu
Fakes	241.37\pm10.56	238.38\pm26.11	299.21\pm18.65
PromptSS	430.889\pm2.95	668.374\pm3.42	1256.15\pm8.42
Vgamma	121.404\pm12.58	0.69\pm0.59	247.435\pm13.9964
QmisID	470.4\pm0.99	-	33.55\pm0.23
Total	1305.65\pm17.00	906.38\pm26.22	1835.21\24.5
Data	1370\pm36.97	856\pm29.25	1879\pm43.3

Njets>=3	ee	mumu	emu
Fakes	84.3\pm6.19	130.29\pm10.82	115.47\pm10.82
PromptSS	192.85\pm1.624	288.53\pm2.00	497.014\pm2.64
Vgamma	51.86\pm6.39	0.067\pm0.067	82.732\pm 6.388
QmisID	172.01\pm0.62	-	10.39\pm0.15
Total	501\pm9.05	419\pm26.22	704.57\pm12.2
Data	543\pm23.28	382\pm19.54	777pm27.87

Don’t list all events number from no PLV region

ee channel

Fake	QmisID	promptSS	Vgamma	Total	data
809.38\pm10.54	-	739.60pm4.31	1.44pm0.90	1549.4\pm10.86	1322pm36.35

in noPLV region, statistics increasing, but worse effect show on Data/MC

Previous research shows

- Optimizing the overall background uncertainty requires good selection of the denominator
- large systematic using ff method. ~60%

Note: prompt and Data number don’t come from page 3 , different selection and tool

Summary and todos

- Check the mc list in using. Modify the Data/MC plot with complete samples. Samples like diboson might need validate.

- Data could be explained with the fake factor method

  Determine the fakes and QmisID background using data-driven method in currently CR. TheCR decisions should be optimized by scan the significance.

- Give the distribution in pre-selection level after estimating QmisID and fake background.

Backup

Lepton origin leading contribution

Zjets QmisID

Evts	Prompt	QmisID	HFb	HFc	LFgamm	LFrest
ee	0	5033	180	28	808	1
uu	0	-1	280	60	0	49
eu	0	42	493	142	923	41

VV Prompt

Evts	Prompt	QmisID	HFb	HFc	LFgamma	LFrest
ee	1199	93	9	1	35	0
uu	2026	0	12	6	0	2
eu	3081	70	21	6	34	1

ttbar Fake from HF

Evts	Prompt	QmisID	HFb	HFc	LFgamma	LFrest
ee	26	1170	2668	91	389	28
uu	32	0.22	3096	209	0	61
eu	57	1405	5758	298	478	85

Wjets Fake from HF

Evts	Prompt	QmisID	HFb	HFc	LFgamma	LFrest
ee	0	0	1229	137	389	-15
uu	0	0	1524	671	0	683
eu	0	0	434	434	2027	179