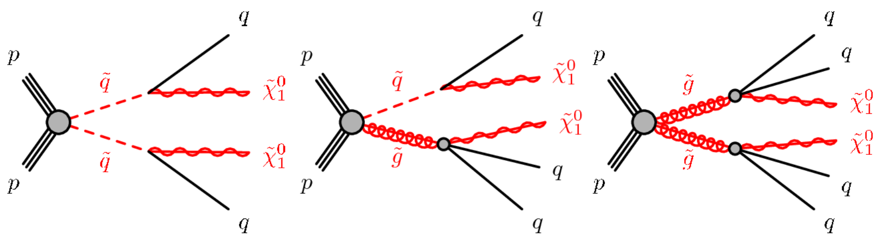
Background estimation in 0-lepton (2-6) jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ analysis of SUSY searches at 13 TeV

Yuanyuan Zhang
Supervisor: Dr. Yu Higuchi Nakahama

August 20, 2015

1 Introduction

0-lepton (2-6) jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ analysis concerns SUSY signals of pair-produced squarks and gluinos ($\tilde{q}\tilde{q}$ $\tilde{q}\tilde{g}$ $\tilde{g}\tilde{g}$) direct decay and longer decay chains. For direct decay, squark decays as $\tilde{q} \to q\tilde{\chi}_1^0$ and gluino $\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$. These pair-produced squarks and gluinos can generates two, three and four jets (Figure 1 below show the $\tilde{q}\tilde{q}$ $\tilde{g}\tilde{g}$ decay modes). With longer decay chains of squarks and gluinos, the jet multiplicity will increase. The lightest SUSY particles (LSP) $\tilde{\chi}_1^0$ contribute to the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$. In pp collision, a variety of processes in the Standard Model (SM) also generate (2-6) jets, 0-lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$. The estimation of the SM background is vital to searching of SUSY signal. The aim of this study is a first look at the new 13 TeV data in the signal and control regions of the 0-lepton analysis in order to check if the Run-1 like semi-data driven method can be applicable to the new Run-2 analysis.



(a) $\tilde{q}\tilde{q}$ (b) $\tilde{q}\tilde{g}$ (c) $\tilde{g}\tilde{g}$

Figure 1: Direct decay mode of pair-produced supersymmetric particles $\tilde{q}\tilde{q}$ $\tilde{q}\tilde{g}$ $\tilde{g}\tilde{g}$

2 Analysis Strategy

Applying cuts on discriminating kinematic variables to increase the signal background ratio is the basic strategy of the signal region optimisation. The inclusive SUSY searches try to achieve maximal reach over the ($m_{\tilde{g}}$, $m_{\tilde{q}}$)-plane, so a variety of signal regions(SRs) is defined. For one signal region, several control regions(CRs) are defined, with each CR enriched in particular SM background (four SM backgounds) and minimized in SUSY signal contamination. By plotting the events distribution in CRs over several discriminating kinematic variables(m_{eff}(incl.),$\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})$ etc.), the cut on this variable can be determined by reducing the enriched background to maximum extent. After the cut, the events observation in the CRs $N(\mathrm{CR}, \mathrm{obs})$ can be converted into background events estimation(normalized) in SRs $N(\mathrm{SR}, \mathrm{nor})$ by

$$N(\mathrm{SR}, \mathrm{nor}) = N(\mathrm{CR}, \mathrm{obs}) \left[\frac{N(\mathrm{SR}, \text{un-nor})}{N(\mathrm{CR}, \text{un-nor})} \right]$$

the ratio in the bracket is defined to be transfer factor, while $N(\mathrm{SR}, \text{un-nor})$ and $N(\mathrm{CR}, \text{un-nor})$ are estimated events number contributed by the same process to SR and CR. This ratio can be determined by data-driven techniques or MC simulations, and also its associated uncertainties.

3 Backgrounds Composition

There are five SM backgrounds in 0-lepton (2-6) jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ analysis.

3.1 Z + jets

Z+jets background consists of four components Z($\rightarrow e\bar{e}$)+jets, Z($\rightarrow \mu\bar{\mu}$)+jets, Z($\rightarrow \tau\bar{\tau}$)+jets and Z ($\rightarrow \nu\bar{\nu}$)+jets. The first three can be suppressed by requirements of 0-lepton. There are still some small contribution from the first three components when electron or muon is not identified .The largest background contribution of Z + jets results from the last one, and neutrinos generated by it will bring large $E_{\mathrm{T}}^{\mathrm{miss}}$.

3.2 W+jets

W+jets contains W($\rightarrow e\nu$)+jets, W($\rightarrow \mu\nu$)+jets and W($\rightarrow \tau\nu$)+jets. The first two can also be largely reduced by requiring 0-lepton, with some processes contribute to background due to no electron or muon reconstructed. In the last one, τ can decays to hadrons, which will be recognized as jets rather than leptons. So the majority of W+jets background is W($\rightarrow \tau\nu$)+jets.

3.3 Top

Top means top quark pairs ($t\bar{t}$) and single top quark decay modes, i.e. $t\bar{t} \rightarrow 2b + l\nu + 2q$, $t\bar{t} \rightarrow 2b + 2(l\nu)$, $t \rightarrow b + l\nu$ (l can be e, μ or τ). The processes with τ

semileptonic decaying or un-reconstructed leptons contributes to backgrounds.

3.4 Multi-jets

The multi-jets consists of semi-leptonic decays of heavy-flavour quarks, and mis-reconstruction of jets energy which give rise to $E_{\mathrm{T}}^{\mathrm{miss}}$.

3.5 Dibosons

Dibosons (WW, WZ, ZZ), with at least one boson decays to charged leptons or netrinos (according to decay mode of W and Z) also contribute to standard model background.

4 Discriminant Plots for SRs and CRs

We plot several discriminants for SR and CRWT of v44 data at 13TeV. The region selections are listed in Table 1.

Region	Process	Selections
CRW	W($\to l\nu$)+jets	p_{T}(jet1) > 100 GeV, p_{T}(jet$[2,3,4,5,6]$) > 40 GeV $0 < m_{\mathrm{eff}}$(incl.) < 1200 GeV 30 GeV $< m_{\mathrm{T}}(l, E_{\mathrm{T}}^{\mathrm{miss}}) <$ 100 GeV p_{T}(lep1) > 25 GeV, b-veto
CRT	$t\bar{t} \to 2b + l\nu + 2q$ $t\bar{t} \to 2b + 2(l\nu)$ $t \to b + l\nu$	p_{T}(jet1) > 100 GeV, p_{T}(jet$[2,3,4,5,6]$) > 40 GeV $0 < m_{\mathrm{eff}}$(incl.) < 1200 GeV 30 GeV $< m_{\mathrm{T}}(l, E_{\mathrm{T}}^{\mathrm{miss}}) <$ 100 GeV p_{T}(lep1) > 25 GeV, b-tag
SR	signal + background	p_{T}(jet1) > 100 GeV, p_{T}(jet$[2,3,4,5,6]$) > 40 GeV $\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}} > 0.4$, $\Delta R > 0.2$ $0 < m_{\mathrm{eff}}$(incl.) < 1200 GeV $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}(2\mathrm{j}) > 0.2$ $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV

Table 1: Selection criteria for control region W+jets and Top, signal region.

4.1 SR

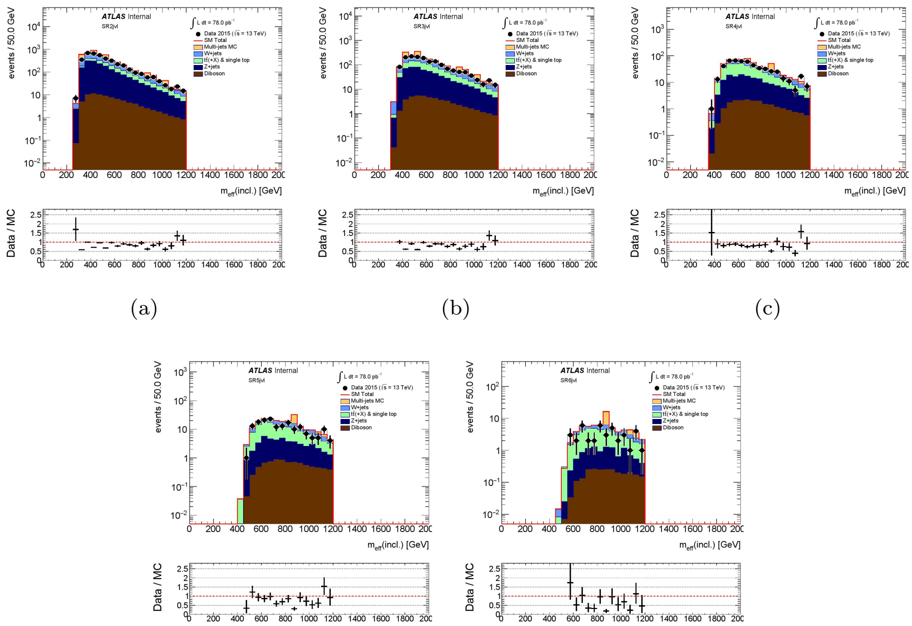
For signal region in Table 1, the p_{T}(jet1) > 100 GeV, p_{T}(jet$[2,3,4,5,6]$) > 40 GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV are the minimum requests for SUSY searches. The cut on angle $\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}}$ between jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ with the cut on $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}(2\mathrm{j})$ are used to suppress the multijet background. The geometric variable cut $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} > 0.2$ helps to remove overlapping between jet and lepton.

The discriminants for SR are : m_{eff}(incl.) , N_{jets}, $E_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}}$, p_{T}(jet1), $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}$(incl.). In this study, the MC event are normalised to

78pb^{-1} which corresponds to the compared dataset size. We only show m_{eff}(incl.) plots, with other plots summarized in appendix.

$$m_{\text{eff}}(\text{incl.}) = \sum_{i=1}^{N_{\text{jets}}} |\mathbf{p}_T^{\text{jet},i}| + \sum_{j=1}^{N_{\text{leptons}}} |\mathbf{p}_T^{\text{lep},j}| + E_{\text{T}}^{\text{miss}}$$

m_{eff}(incl.) (connected with m_{SUSY}) is a sensible variable to discriminate SUSY signal with SM background. In the distribution plot Figure 2, one can see : 1. The MC and data agrees well. 2. Diboson background is negligible. 3. With the increasing of N_{jets}, Top background(tend to generate multi-jet) increase its persentage in total background. 4. With one more cut $E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$(incl.) to suppress multi-jet background, 2-jets got less multi-jet contribution. 5. There is a significant multijet contribution around m_{eff}(incl.) $\approx$ 900 GeV in 4-6 jets, due to a multi-jet MC event with huge weight.



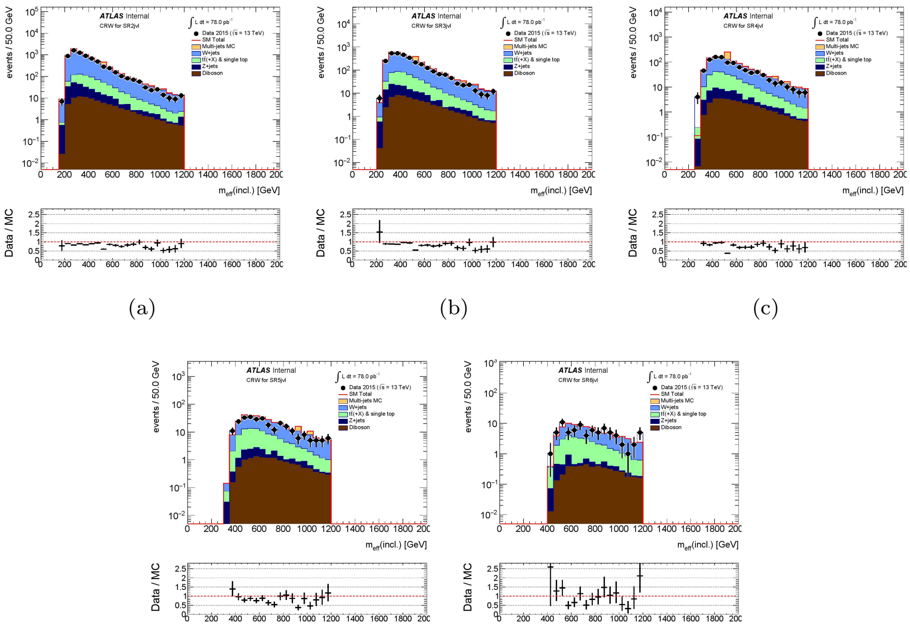
(a) (b) (c)

(d) (e)

Figure 2: m_{eff}(incl.) of signal region for 78pb^{-1} at 13TeV, comparison between v44 data and normalized Monte Carlo samples.

4.2 CRWT

For control region W+jets (CRW) and control region Top (CRT) in Table 1, they only vary by b-tag and b-veto. CRW and CRT do not use cuts of $\Delta\phi$



(a) (b) (c)

(d) (e)

Figure 3: m_{eff}(incl.) of control region W+jets(CRW) for 78pb^{-1} at 13TeV

and ΔR in order to increase the statistics of W+jets and Top. $p_{\text{T}}(\text{lep1}) > 25$ GeV was applied to select the events with almost 100% efficiency after trigger selection. The transverse $m_{\text{T}}(l, E_{\text{T}}^{\text{miss}})$ is defined as the invariant mass of lepton and $E_{\text{T}}^{\text{miss}}$.

$$m_{\text{T}}(l, E_{\text{T}}^{\text{miss}}) = \sqrt{2 * |\mathbf{p}_{\text{T}}^{\text{lep}}||\mathbf{p}_{\text{T}}^{\text{miss}}| - 2 * \mathbf{p}_{\text{T}}^{\text{lep}}\mathbf{p}_{\text{T}}^{\text{miss}}}$$

SUSY events are supposed to populate the high m_{T} region, we choose 30 GeV $< m_{\text{T}} <$ 100 GeV to minimize the signal contamination.

We produce N_{jets}, m_{eff}(incl.), $E_{\text{T}}^{\text{miss}}$, $\Delta\phi(\text{jet}, E_{\text{T}}^{\text{miss}})_{\text{min}}$, p_{T}(jet1), p_{T}(lep1) distributions. Here we only demonstrate m_{eff}(incl.) distribution for CRW(Figure 3) CRT(Figure 4), with others in Appendix. In Figure 3 and Figure 4, one can see the W+ jets and Top dominate respectively, which reflect the good performance of the CR selections. In conclusion, data and MC distributions in CRs, in particular shapes, agree well within the statistical uncertainties. The Run-1 like semi data-driven background estimation can be applicable to the Run-2 analysis without any correction to MC shapes.

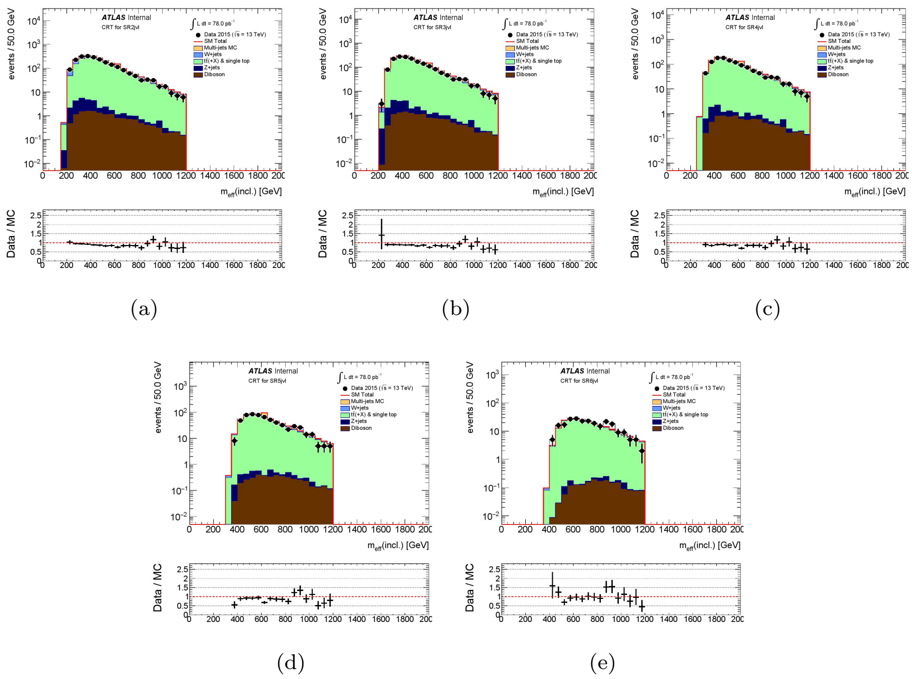


Figure 4: m_{eff}(incl.) of control region Top(CRT) for 78pb^{-1} at 13TeV

5 Cutflowtables

In some situation, the information about exact events number of samples after cuts is vital. The cutflowtables can provide exact event numbers for MC samples and Data after cuts. We produced cutflowtables for several different set of CRW CRT cutflowtables and SR/Data comparision cutflowtables. Here are CRW-2jets (Table 2), CRT-2jets (Table 3) and SR-data-2jets(Table 4), others can be found in Appendix.

Selections	Z+jets	W+jets	Top	Diboson	Total Bkg
Pre-sl, $E_{\mathrm{T}}^{\mathrm{miss}} > 200$, $p_{\mathrm{T}}(\mathrm{jet}_1)>200$	86.03 ± 10.33	10868.52 ± 242.16	2986.15 ± 20.23	353.40 ± 6.56	14294.10 ± 243.32
Jet multiplicity	68.45 ± 8.84	7561.46 ± 156.06	2799.29 ± 19.67	317.74 ± 6.17	10746.94 ± 157.66
$\min(\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet}_{1,2,3}))$ cut	14.45 ± 4.88	1061.05 ± 47.81	398.27 ± 7.47	61.64 ± 2.68	1535.41 ± 48.71
$p_{\mathrm{T}}(\mathrm{jet}_2)$ cut	1.26 ± 0.31	213.06 ± 12.14	75.41 ± 3.25	19.51 ± 1.43	309.25 ± 12.65
$p_{\mathrm{T}}(\mathrm{jet}_4)$ cut	1.26 ± 0.31	213.06 ± 12.14	75.41 ± 3.25	19.51 ± 1.43	309.25 ± 12.65
Aplanarity cut	1.26 ± 0.31	213.06 ± 12.14	75.41 ± 3.25	19.51 ± 1.43	309.25 ± 12.65
$E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{\mathrm{H_T}}$ cut	1.26 ± 0.31	213.06 ± 12.14	75.41 ± 3.25	19.51 ± 1.43	309.25 ± 12.65
$E_{\mathrm{T}}^{\mathrm{miss}}/\mathrm{m_{eff}}(\mathrm{Nj})$ cut	1.26 ± 0.31	213.06 ± 12.14	75.41 ± 3.25	19.51 ± 1.43	309.25 ± 12.65
$\mathrm{m_{eff}}$(incl.) cut	0.95 ± 0.27	158.28 ± 9.90	63.83 ± 2.99	15.83 ± 1.29	238.89 ± 10.43

Table 2: v44: Event yields with MC stat uncertainties for the Z+jets, W+jets Top and Diboson backgrounds in CRW for 2jvt at 3.0 fb^{-1}. The cut are : $N_{\mathrm{jets}} \geq 2$, $\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}} \geq 1.0$, $p_{\mathrm{T}}(\mathrm{jet2}) \geq 200$ GeV, $p_{\mathrm{T}}(\mathrm{jet4}) \geq 0$, Aplanarity ≥ 0.0, $E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{\mathrm{H_T}} \geq 0.0$, $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}} \geq 0.0$,$m_{\mathrm{eff}}(\mathrm{incl.}) \geq 1000$ GeV, $N_{\mathrm{bjets}} \geq 0$, $N_{\mathrm{lep}} \geq 1$

Selections	Z+jets	W+jets	Top	Diboson	Total Bkg
Pre-sl, $E_{\mathrm{T}}^{\mathrm{miss}} > 200$, $p_{\mathrm{T}}(\mathrm{jet}_1)>200$	86.03 ± 10.33	10868.52 ± 242.16	2986.15 ± 20.23	353.40 ± 6.56	14294.10 ± 243.32
Jet multiplicity	9.09 ± 0.91	1118.83 ± 30.44	1307.92 ± 13.57	115.35 ± 3.61	2551.18 ± 33.53
$\min(\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet}_{1,2,3}))$ cut	2.60 ± 0.45	514.60 ± 23.03	480.38 ± 8.28	48.25 ± 2.34	1045.84 ± 24.59
$p_{\mathrm{T}}(\mathrm{jet}_2)$ cut	0.83 ± 0.24	175.30 ± 12.79	107.26 ± 3.83	18.38 ± 1.38	301.77 ± 13.43
$p_{\mathrm{T}}(\mathrm{jet}_4)$ cut	0.83 ± 0.24	175.30 ± 12.79	107.26 ± 3.83	18.38 ± 1.38	301.77 ± 13.43
Aplanarity cut	0.57 ± 0.19	101.65 ± 9.96	63.62 ± 2.91	11.70 ± 1.10	177.54 ± 10.44
$E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{\mathrm{H_T}}$ cut	0.57 ± 0.19	101.65 ± 9.96	63.62 ± 2.91	11.70 ± 1.10	177.54 ± 10.44
$E_{\mathrm{T}}^{\mathrm{miss}}/\mathrm{m_{eff}}(\mathrm{Nj})$ cut	0.46 ± 0.18	78.90 ± 9.19	47.18 ± 2.53	8.87 ± 0.96	135.40 ± 9.58
$\mathrm{m_{eff}}$(incl.) cut	0.37 ± 0.16	66.46 ± 8.73	40.57 ± 2.34	7.81 ± 0.90	115.21 ± 9.08

Table 3: v44: Event yields with MC stat uncertainties for the Z+jets, W+jets Top and Diboson backgrounds in CRT for 4jvt at 3.0 fb^{-1}. The cut are : $N_{\mathrm{jets}} \geq 4$, $\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}} \geq 0.4$, $p_{\mathrm{T}}(\mathrm{jet2}) \geq 200$ GeV, $p_{\mathrm{T}}(\mathrm{jet4}) \geq 10$ GeV , Aplanarity ≥ 0.02, $E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{\mathrm{H_T}} \geq 0.0$, $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}} \geq 0.2$,$m_{\mathrm{eff}}(\mathrm{incl.}) \geq 1000$ GeV, $N_{\mathrm{bjets}} \geq 1$, $N_{\mathrm{lep}} \geq 1$

6 Conclusion

In this project, the Run-1 like semi-data driven method was tested to be applicable to Run-2 data. With discriminant distribution plots in control region & signal region and good agreement of MC with data, the proper cuts to suppress the SM background can be determined. In some situation,the cutflowtables can be used to check the exact number of events of different SM backgrounds or data after cuts. In the future, with more statistics and detailed study, more sophisticated data-driven strategies can be used in inclusive SUSY particle searches.

Selections	gluino pair m(gluino,N1)=(1350, 0) GeV	DataMain	Total Bkg
Pre-sl, $E_{\rm T}^{\rm miss}$ > 200, $p_{\rm T}(\rm jet_1)$ > 200	2.50 ± 0.02	9360.00 ± 96.75	1946.00 ± 21.98
Jet multiplicity	2.50 ± 0.02	1610.00 ± 40.12	949.17 ± 7.01
min($\Delta\phi(E_{\rm T}^{\rm miss}, \rm jet_{1,2,3})$) cut	2.01 ± 0.02	701.00 ± 26.48	646.31 ± 6.04
$p_{\rm T}(\rm jet_2)$ cut	2.01 ± 0.02	125.00 ± 11.18	108.24 ± 1.79
$p_{\rm T}(\rm jet_4)$ cut	2.01 ± 0.02	125.00 ± 11.18	108.24 ± 1.79
Aplanarity cut	2.01 ± 0.02	124.00 ± 11.14	106.95 ± 1.77
$E_{\rm T}^{\rm miss}/\sqrt{\rm H_T}$ cut	2.01 ± 0.02	4.00 ± 2.00	4.82 ± 0.15
$E_{\rm T}^{\rm miss}/\rm m_{eff}(Nj)$ cut	2.01 ± 0.02	4.00 ± 2.00	4.82 ± 0.15
$\rm m_{eff}(incl.)$ cut	2.01 ± 0.02	4.00 ± 2.00	4.20 ± 0.13

Table 4: v44: Event yields for signal, data, and total backgrounds with MC stat uncertainties in SR2jvt at 0.078 $\rm fb^{-1}$. The cut are : $N_{\rm jets} \geq 2$, $\Delta\phi(\rm jet, E_{\rm T}^{\rm miss})_{\rm min} \geq 0.4$, $p_{\rm T}(\rm jet2) \geq 200$ GeV, $p_{\rm T}(\rm jet4) \geq 0$, Aplanarity ≥ 0.0, $E_{\rm T}^{\rm miss}/\sqrt{\rm H_T} \geq 20.0$, $E_{\rm T}^{\rm miss}/m_{\rm eff} \geq 0.0$,$m_{\rm eff}(\rm incl.) \geq 1000$ GeV