CHAPTER 5

Events Selection

The goal of the event selection is to achieve the highest possible fraction of tZq events in data as predicted by Monte-Carlo Simulation. Data events are selected if they pass the preselection criteria discussed in Section 4.1. The preselection treatments are also applied on MC events. In order to verify the modeling of physics processes in the relevant areas of kinematic phase space, a set of signal, control and validation regions are defined. Each region is defined by a set of selection cuts on the reconstructed variables. These reconstructed variables can also be the one from intermediate-state particles which are calculated from the final state observables.

5.1 Full Event Reconstruction

The tZq final state at the leading order for which the analysis is performed consists of three leptons, one neutrino, one b-quark and a light quark which is expected in the forward direction. The two of the three leptons come form the Z decay and the one from the leptonic W decay. In order to reconstruct the Z boson, an opposite-sign, same-flavor (OSSF) lepton pair is needed. In the $ee\mu$ and $e\mu\mu$ channels, this is uniquely identified. For the eee and $\mu\mu\mu$ events, both possible combinations are considered and the pair that has the invariant mass closest to the Z-boson mass is chosen.

The remaining lepton and the missing momentum from neutrino is used to reconstruct the W boson. The missing part of the neutrino four-vector is the longitudinal component along the z-axis (P_z^ν), which can be obtained using the mass constraint of the W boson which is M_W, 80.4 GeV. From the four-momentum conservation

$$(P^W)^2 = (P^l + P^\nu)^2 = M_W^2 \tag{5.1}$$

The solution of the quadratic equation given in Equation 5.1 in terms of the P_z^ν can be expressed as follows

$$P_z^\nu = \frac{\alpha . P_z^l \pm \sqrt{(E^l)^2(\alpha^2 - P_T^l . \not{E}_T)}}{(P_T^l)^2} \tag{5.2}$$

where α is given by

$$\alpha = \frac{M_W^2}{2} + \overrightarrow{P}_T^l . \overrightarrow{\not{E}}_T \tag{5.3}$$

when the quantity under the square root is positive ($\alpha^2 \geq P_T^l.\not{E}_T$), then there are two real solutions, and the smallest one in magnitude is taken. Since the W boson is expected to be produced with small rapidity. For some events, Equation 5.1 has imaginary solution ($\alpha^2 \leq P_T^l.\not{E}_T$), which is interpreted as a mis-measurement of $\not{E}_T$. In this case the transverse mass, $m_T(W)$, is greater than M_W and $m_T(W)$ is explicitly set to equal M_W and the neutrino 4-vector is rescaled. Technically this is resolved by introducing a new scale factor β, which is defined by

$$\beta = \frac{M_W^2}{2P_T^l}.\not{E}_T - \overrightarrow{P}_T^l.\overrightarrow{\not{E}}_T \tag{5.4}$$

β is used to scale $\not{E}_x, \not{E}_y$ and $\not{E}_T$ and then α is recalculated as shown in Equation 5.3. P_z^ν is found by considering only the offset part of the Equation 5.2.

Then, the reconstructed W boson and the b-tagged jet are used for the t-quark reconstruction as follows

$$P^t = P^W + P^{b\text{-}jet} \tag{5.5}$$

A summary of relevant symbols representing the reconstructed objects is presented in Table 5.1

Symbol	Description
l_1^Z	Highest p_T lepton from the reconstructed Z boson
l_2^Z	Lowest p_T lepton from the reconstructed Z boson
Z	Reconstructed Z boson
l^W	Lepton from the reconstructed W boson from the t-quark decay
W	Reconstructed W boson from the t-quark decay
b-jet	b-tagged jet
t	Reconstructed t quark
j_f	Forward jet
j_r	Radiation jet
$l_{1/2/3}$	p_T ordered leptons
$j_{1/2/3}$	p_T ordered jets

Table 5.1: Object reconstruction.

5.2 Signal Region

As shown in the Feynman diagrams in section ...., the tZq signal consists of final states with three leptons, one neutrino, and two jets, one of which originates from a b-quark and the other tends to be more forward. The QCD calculations at NLO suggests that there can exist significant QCD radiation present in the event which can manifest itself in the form of a third reconstructed jet.

In order to increase acceptance as much as reasonably possible two orthogonal signal regions (SRs) named as *SR-2j1b* and *SR-3j1b* are defined. In SR-2j1b, events with three leptons, one b-tagged jet

and one untagged jet as forward jet are selected. In SR-3j1b, events are selected identically to the SR-2j1b except for the inclusion of a second untagged jet. One of the two untagged jets, that one that gives, with the b-jet, the highest invariant mass, m_{bj_f} is selected to be the forward jet. The remaining jet is called radiation jet. The same nomenclature is used for the jets in the control regions (CRs)

5.2.1 Lepton Isolation Working Point

In order to differentiate the prompt production of leptons in the signal processes, for instance decay of heavy resonances like Higgs, W and Z bosons, from the fake leptons background processes, isolation criteria are implemented. Isolation consists in accessing the activity surrounding the trajectory of the leptons in the tracker and the calorimeter. To calculate the isolation variables, a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ in the η and ϕ plane is defined around the particle and the energy of close-by objects falling in the cone is added, having subtracted the contribution of the particle itself. Both calorimter- and track-based isolation variables are used in ATLAS. Based on the cuts applied on these two isolation variables different working points (WPs) are defined which are presented in reference [].

In this analysis, the yield for a set of isolation WPs shown in the Table 5.2 is presented. The yield are obtained with the following selection cuts: The three leptons are sorted by their p_T, irrespective of flavour, and required to have transverse momenta of at least 28, 20 and 20 GeV, respectively. Jets are required to have $P_T > 35$ GeV with one b-tagged jet with 70% WP.

Name	Electron	Muon
Gradient	Gradient	FCTightTrackOnly_FixedRad
PLV	PLVTight	PLVTight
PLVLoose	PLVLoose	PLVLoose
PLIV	PLImprovedTight	PLImprovedTight
PLIVV	PLImprovedVeryTight	PLImprovedVeryTight
Pflow	PflowTight	PflowTight_VarRad
TrackOnly	TightTrackOnly	TightTrackOnly_VarRad
Tight	Tight	Tight_VarRad

Table 5.2: Combination of the Isolation Working Point analysed

Figure 5.1 and 5.2 shows the relative yield for the different lepton WPs. It is found that PLV, PLIV, and PLIVV have large signal and small fake-background compared to others. While PLVLoose has comparatively large fake background. The significance calculated in Table 5.3 shows that PLVLoose has the highest significance for both SR-2j1b and SR-3j1b. The significance is also high for PLV and PLIV compared to others. So, the lepton WPs PLV, and PLVLoose are chosen for the optimization of selection cuts.

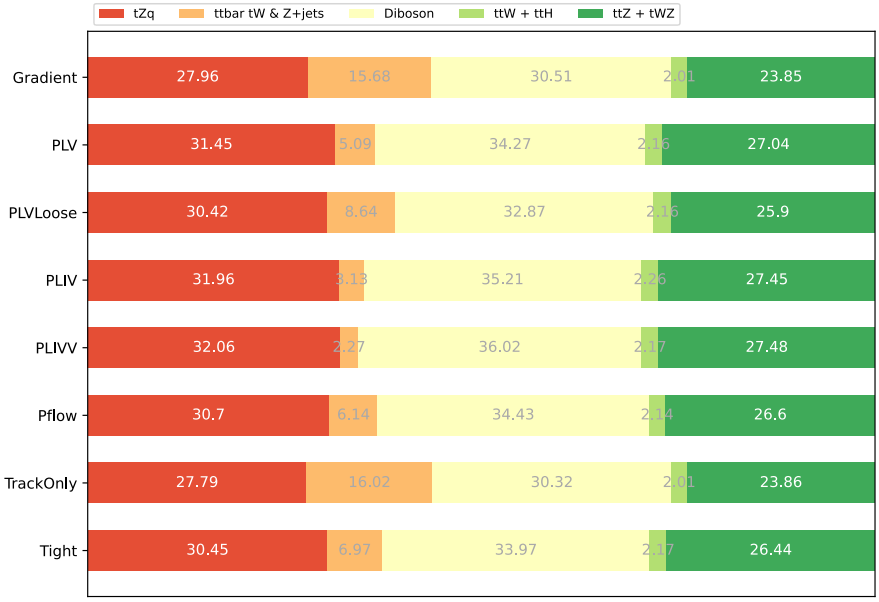


Figure 5.1: Relative yield for SR 2j1b with jet p_T > 35 GeV, lepton p_T > 28, 20, 20 , GeV & btag-eff: 70%

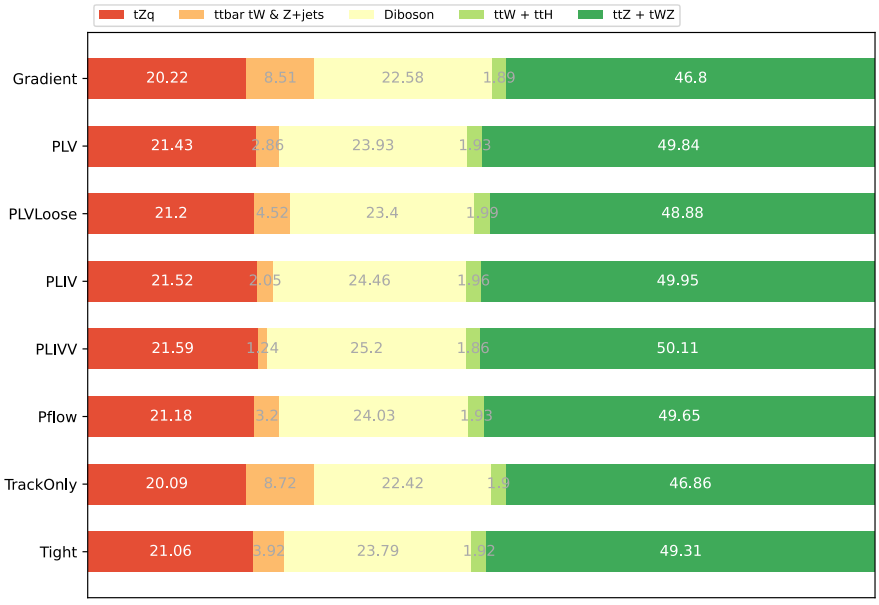


Figure 5.2: Relative yield for SR 3j1b with jet p_T > 35 GeV, lepton p_T > 28, 20, 20 , GeV & btag-eff: 70%

	SR 2j1b			SR 3j1b		
Lepton WPs	S	B	$\frac{S}{\sqrt{S+B}}$	S	B	$\frac{S}{\sqrt{S+B}}$
Gradient	89.40	230.38	5.00	50.89	200.73	3.21
PLV	83.73	182.51	5.13	47.67	174.82	3.20
PLVLoose	91.50	209.26	5.28	52.10	193.61	3.32
PLIV	80.70	171.8	5.08	46.10	168.10	3.15
PLIVV	66.50	140.90	4.62	38.30	139.10	2.88
Pflow	79.86	180.27	4.95	45.79	170.38	3.11
TrackOnly	92.59	240.56	5.07	52.61	209.22	3.25
Tight	81.82	186.88	4.99	46.96	176.06	3.14

Table 5.3: Values of the Significance for SR 2j1b and SR 3j1b for various Isolation WPs

5.2.2 Optimization of Cuts

The analysis aims to set new requirements on the phase space so that the contribution of the signal grows relative to the number of background events. The significance is calculated for different selection cuts and are compared to the previous selections. The cuts are optimized by varying leptons p_T, jets p_T and the b-tagging efficiency of the b-tagged jet.

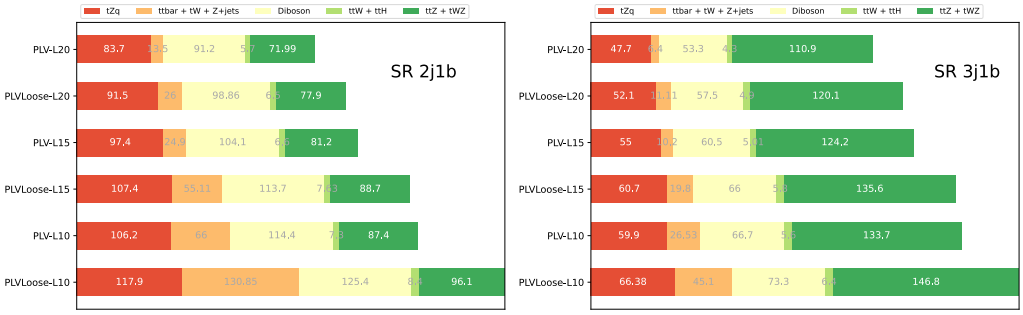


Figure 5.3: Yeild for signal and background at varying third lepton p_T

		SR 2j1b			SR 3j1b		
Lepton WPs	$p_T(\ell_3) >$	S	B	$\frac{S}{\sqrt{S+B}}$	S	B	$\frac{S}{\sqrt{S+B}}$
PLV	20 GeV	83.70	182.39	5.13	47.7	174.9	3.20
	15 GeV	97.40	216.80	5.49	55.00	199.91	3.44
	10 GeV	106.40	275.10	5.45	59.90	232.53	3.50
PLVLoose	20 GeV	91.50	209.26	5.28	52.10	193.61	3.32
	15 GeV	107.40	265.14	5.56	60.7	227.20	3.58
	10 GeV	117.90	360.75	5.35	66.38	271.60	3.61

Table 5.4: Significance calculated at varying third lepton p_T for both SR-2j1b and SR-3j1b

Figure 5.3 shows the yield at varying third lepton p_T, while the first and third lepton lepton p_T remain at 28 GeV and 20 GeV respectively. The btagging WP is 70% and the jet transverse momentum remain above 30 GeV. It shows that both the signal and background increases with loose third lepton p_T cut. The significance in Table 5.4 that PLV at 10 GeV and PLVLoose at 15 GeV have considerable signal to background ratio while significance is above 5σ for SR-2j1b and 3σ for SR-3j1b. Figure 5.5 shows that the change in the first lepton p_T from 28 GeV to 27 GeV doesnot change the yield much for both PLV and PLVLoose. Similarly the yield remains same by changing the second lepton p_T from 20 GeV to 15 GeV. Thus the final cut for analysis for the leptons p_T are set to 27, 20, 10 GeV for PLV and 27, 20, 15 GeV for PLVLoose for first, second and third lepton p_T respectively.

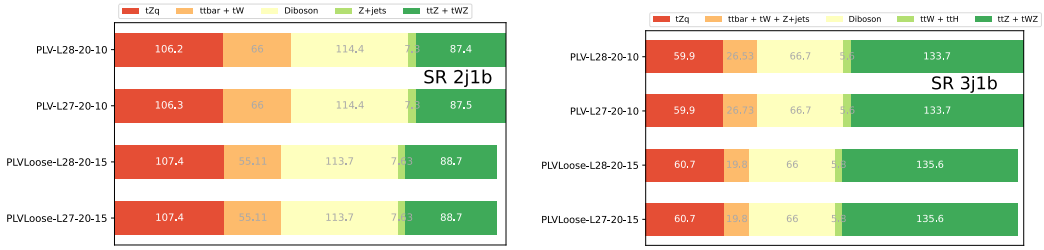


Figure 5.4: Events yeild for signal and background at varying first lepton p_T

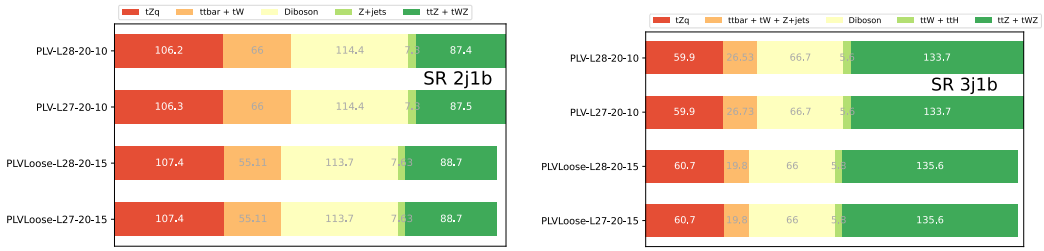


Figure 5.5: Events yeild for signal and background at varying second lepton p_T

Figure 5.6 and 5.7 shows the yield at varying jet p_T, b-tagging efficiency and third lepton p_T. The lower in jet p_T migrate the signal to higher jet multiplicity keeping total number of signal in SR-2j1b and SR-3j1b nearly same. There is large increase in diboson background yield with increase in b-tagging working point. Here, in the plot the diboson contribution is split according to the origin of the associated jets using generator-level information. If one of the jets contains a b- or c-hadron then it is classified as diboson + heavy flavour (VV + HF), otherwise the event is classified as diboson + light flavour (VV + LF). Thus for final cut for the jet transverse momentum and b-jet tagging efficiency for our analysis remain at 35 GeV and 70%.

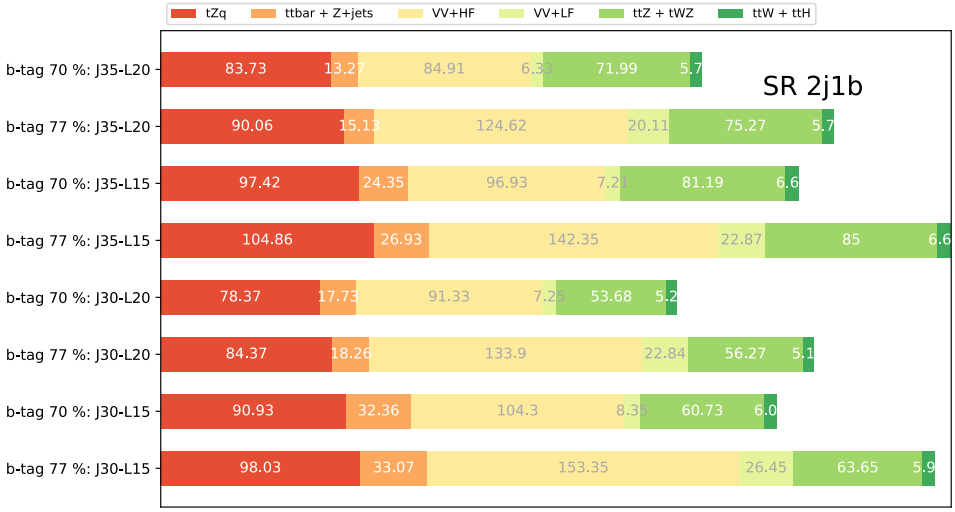


Figure 5.6: Events yeild for signal and background obtained by varing the jet p_T, b-tagging efficiency and the third lepton p_T in the SR-2j1b

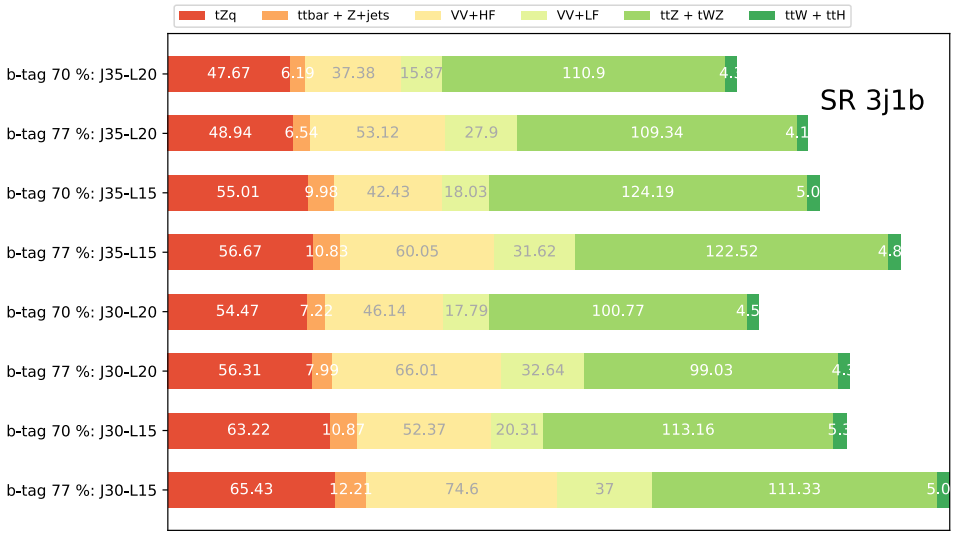


Figure 5.7: Events yeild for signal and background obtained by varing the jet p_T, b-tagging efficiency and the third lepton p_T in the SR-3j1b

5.2.3 Signal Regions (SRs) Plots and Yields

The regions are constructed as described in section 5.2. The full selection cuts applied in the signal regions are listed in Table 5.5. In this table, also the selection cuts for the definition of control regions

Common selections			
Exactly 3 leptons (e or μ) with $\|\eta\| < 2.5$ $p_T(\ell_1) > 27\,\text{GeV}$, $p_T(\ell_2) > 20\,\text{GeV}$, $p_T(\ell_3) > 10\,\text{GeV}$ $p_T(\text{jet}) > 35\,\text{GeV}$			
SR 2j1b	CR Diboson 2j0b	CR $t\bar{t}Z$ 3j2b	CR $t\bar{t}$ 2j1b
≥ 1 OSSF pair $\|m_{\ell\ell} - m_Z\| < 10\,\text{GeV}$ 2 jets, $\|\eta\| < 4.5$ 1 bjet, $\|\eta\| < 2.5$	≥ 1 OSSF pair $\|m_{\ell\ell} - m_Z\| < 10\,\text{GeV}$ 2 jets, $\|\eta\| < 4.5$ 0 bjets	≥ 1 OSSF pair $\|m_{\ell\ell} - m_Z\| < 10\,\text{GeV}$ 3 jets, $\|\eta\| < 4.5$ 2 bjets, $\|\eta\| < 2.5$	≥ 1 OSDF pair No OSSF pair 2 jets, $\|\eta\| < 4.5$ 1 bjet, $\|\eta\| < 2.5$
SR 3j1b	CR Diboson 3j0b	CR $t\bar{t}Z$ 4j2b	CR $t\bar{t}$ 3j1b
≥ 1 OSSF pair $\|m_{\ell\ell} - m_Z\| < 10\,\text{GeV}$ 3 jets, $\|\eta\| < 4.5$ 1 bjet, $\|\eta\| < 2.5$	≥ 1 OSSF pair $\|m_{\ell\ell} - m_Z\| < 10\,\text{GeV}$ 3 jets, $\|\eta\| < 4.5$ 0 bjets	≥ 1 OSSF pair $\|m_{\ell\ell} - m_Z\| < 10\,\text{GeV}$ 4 jets, $\|\eta\| < 4.5$ 2 bjets, $\|\eta\| < 2.5$	≥ 1 OSDF pair No OSSF pair 3 jets, $\|\eta\| < 4.5$ 1 bjet, $\|\eta\| < 2.5$

Table 5.5: Overview of the requirements applied when selecting events in the signal and control regions. OSSF is an opposite-sign same-flavour lepton pair. OSDF is an opposite-sign different-flavour lepton pair.

are reported. The event yields in the SRs after the full selection can be found in Table 5.6 and histogram distributions of reconstructed variables from the top quark and Z boson are given in Fig. 5.8. The predicted number of events to pass selection cuts based on MC simulations for tZq as well as all previously mentioned backgrounds are tabulated. The events in these regions are the primary regions of interest for the statistical analysis described in section 6.4 after having been evaluated by the neural network described in section 6.2.

Process	Number of events	Number of raw events
tZq	97.46 ±0.67	56838800
tt	22.72 ±0.88	104389
tW	0.89 ±0.61	1112
Z+jets	36.66 ±4.13	228516
Diboson	115.60 ±1.09	5741400
ttZ	62.88 ±2.30	7001290
ttW	4.56 ±0.18	255621
tWZ	19.34 ±0.58	489419
ttH	2.11 ±0.04	715711
Total expected	362.22 ±4.65	71376200
Data	443	443

Process	Number of events	Number of raw events
tZq	55.31 ±0.56	38569200
tt	10.79 ±0.61	49206
tW	0.33 ±0.58	417
Z+jets	13.32 ±1.14	102582
Diboson	66.05 ±0.70	3615390
ttZ	101.39 ±0.68	12441100
ttW	2.34 ±0.13	134969
tWZ	22.92 ±0.65	596449
ttH	2.80 ±0.05	753658
Total expected	275.24 ±1.84	56262900
Data	307	307

Table 5.6: Numbers of expected events in the SR-2j1b (Left) and SR-3j1b (Right) broken down by process. The uncertainty shown contains only the statistical component.

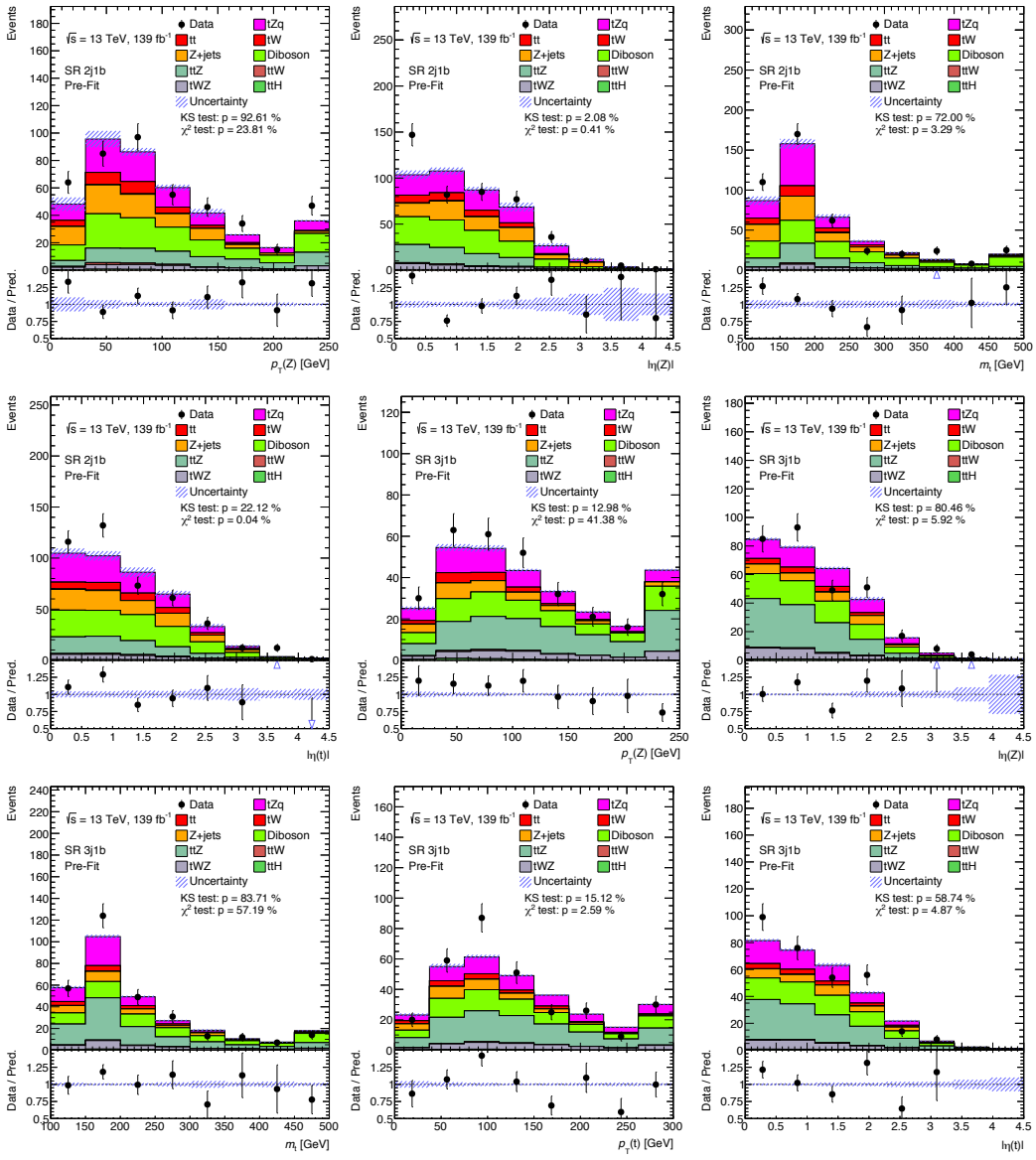


Figure 5.8: Comparison of data and MC predictions for reconstructed event- related quantities for events in the SR-2j1b and SR-3j1b. The uncertainty shown is the statistical uncertainty

5.3 Control Regions (CRs)

In order to ensure proper modeling of each relevant background, a series of control regions as listed in Table 5.5 are defined. These regions are constructed such that they are enriched in three of the main sources of backgrounds: diboson, $t\bar{t}Z$, and $t\bar{t}$ production. There are two CRs for each background,

and one corresponding to each signal region.

5.3.1 Diboson CRs Plots and Yields

To define regions of phase space enriched in diboson production, the b-jet is vetoed. This leads to events such as WZ to dominate while effectively removing the events containing the top quark. This region also has significant contamination from Z+jets events.

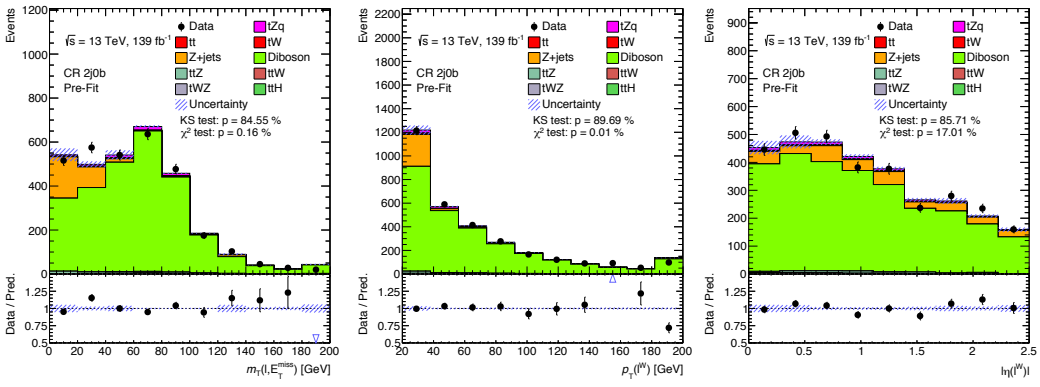
The two diboson control regions are listed in Table 5.5. The events yields in diboson CRs after the full section are shown in Table 5.7. Some of the reconstructed variables in this control regions are shown in Fig. 5.10.

Process	Number of events	Number of raw events
tZq	61.94 ±0.57	40215200
tt	14.58 ±0.73	64357
tW	0.49 ±0.75	556
Z+jets	152.32 ±18.75	624666
Diboson	2625.59 ±5.02	146163000
ttZ	48.59 ±4.79	4766730
ttW	1.85 ±0.11	98134
tWZ	17.10 ±0.54	422977
ttH	1.14 ±0.03	323175
Total expected	2923.59 ±17.11	192679000
Data	3116	3116

Process	Number of events	Number of raw events
tZq	25.58 ±0.42	21530100
tt	5.59 ±0.46	24603
tW	0.60 ±0.42	695
Z+jets	52.53 ±5.89	244640
Diboson	973.45 ±2.35	60037700
ttZ	47.15 ±1.42	6095710
ttW	0.88 ±0.09	51847
tWZ	13.13 ±0.51	375439
ttH	1.07 ±0.03	268270
Total expected	1119.97 ±6.51	88629000
Data	1083	1083

Table 5.7: Numbers of expected events in the CR-2j0b (Left) and CR-3j0b (Right) broken down by process. The uncertainty shown contains only the statistical component.

The large number of observed events in this control region helps to provide a significant constraint on the overall rate of diboson events. Because this region is comprised of almost entirely two diboson backgrounds, if one increases, the other must decrease to maintain agreement with data, and the two are inversely correlated. The two diboson bacgrounds are diboson + heavy flavour (VV + HF) and, diboson + lightflavour (VV + LF).



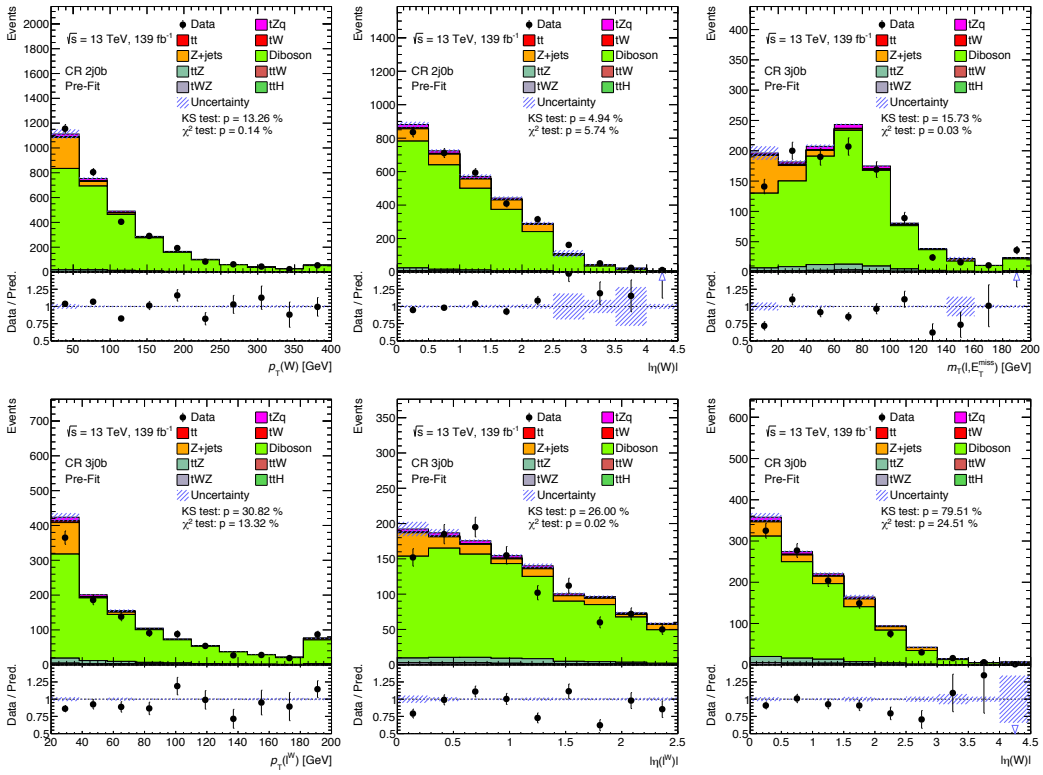


Figure 5.10: Comparison of data and MC predictions for reconstructed event- related quantities for events in the CR-2j0b and CR-3j0b. The uncertainty shown contains only the statistical component.

5.3.2 $t\bar{t}Z$ CRs Plots and Yield

To define regions of phase space enriched in $t\bar{t}Z$ production, an additional b-jet is required to enhance events with a second top quark. This region also contains significant amount of signal events.

The two $t\bar{t}Z$ control regions are listed in Table 5.5. The events yields in $t\bar{t}Z$ CRs after the full section are shown in Table 5.8. Some of the reconstructed variables in this control regions are shown in Fig. 5.11 and Fig. 5.12.

The $t\bar{t}Z$ CRs are unique in that they require an extra b-jet which introduces an ambiguity in the event selection and reconstruction criteria. Here, only one of the b-jet is considered and other one is neglected. The forward jet is then selected as in Section 5.2. The $t\bar{t}Z$ CRs also have contamination from tZq signal events which can create a bias in the measurement. This is solved by fitting the neural-network output, O_{NN} distribution (Section 6.2) which shows seperation between the tZq and $t\bar{t}Z$. This allows robust constraint of $t\bar{t}Z$ modeling and also a slight boost to the overall measurement's senstivity.

Process	Number of events	Number of raw events
tZq	15.78 ±0.23	6584710
tt	2.31 ±0.30	11120
tW	0.08 ±0.43	139
Z+jets	1.10 ±0.19	8062
Diboson	5.48 ±0.16	381555
ttZ	48.94 ±0.45	5681490
ttW	1.69 ±0.12	105501
tWZ	3.91 ±0.26	100775
ttH	1.59 ±0.04	488446
Total expected	80.88 ±0.70	13361800
Data	118	118

Process	Number of events	Number of raw events
tZq	10.43 ±0.20	5136470
tt	1.14 ±0.22	5699
tW	0.16 ±0.38	278
Z+jets	0.56 ±0.13	4170
Diboson	3.38 ±0.12	230184
ttZ	53.35 ±0.51	7061060
ttW	0.76 ±0.07	42812
tWZ	4.76 ±0.29	120652
ttH	1.41 ±0.04	365153
Total expected	75.94 ±0.68	12966500
Data	77	77

Table 5.8: Numbers of expected events in the CR-3j2b (Left) and CR-4j2b (Right) broken down by process. The uncertainty shown contains only the statistical component.

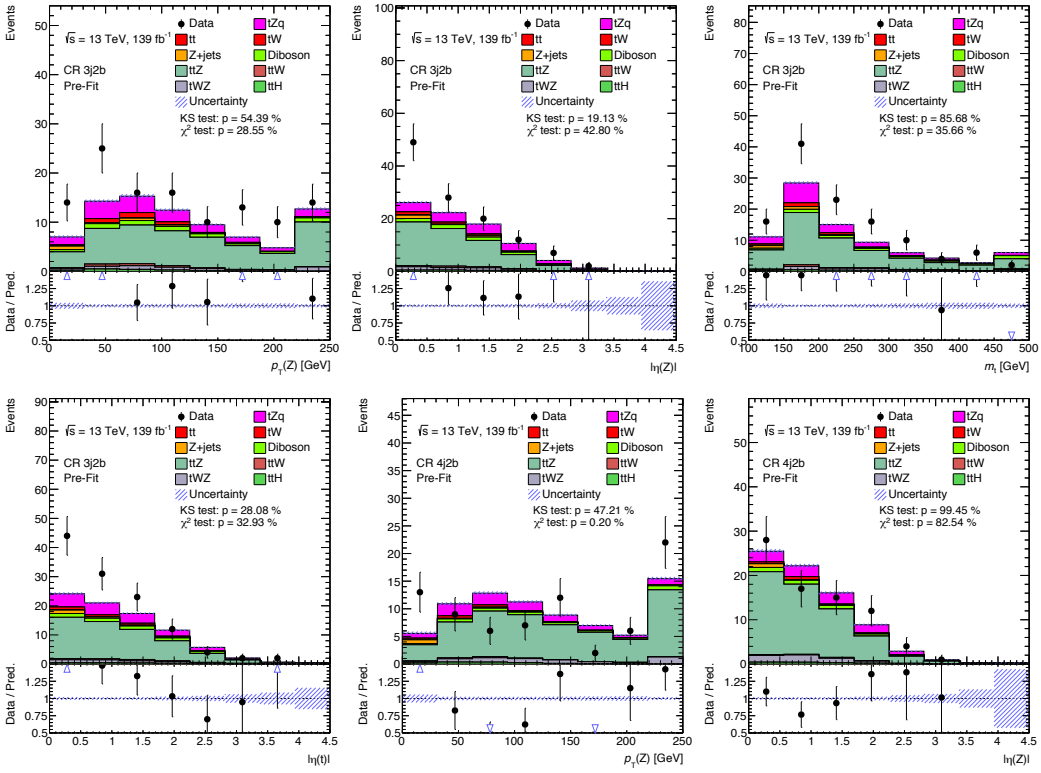


Figure 5.11: Comparison of data and MC predictions for reconstructed event- related quantities for events in the CR-3j2b and CR-4j2b. The uncertainty shown is the statistical uncertainty.

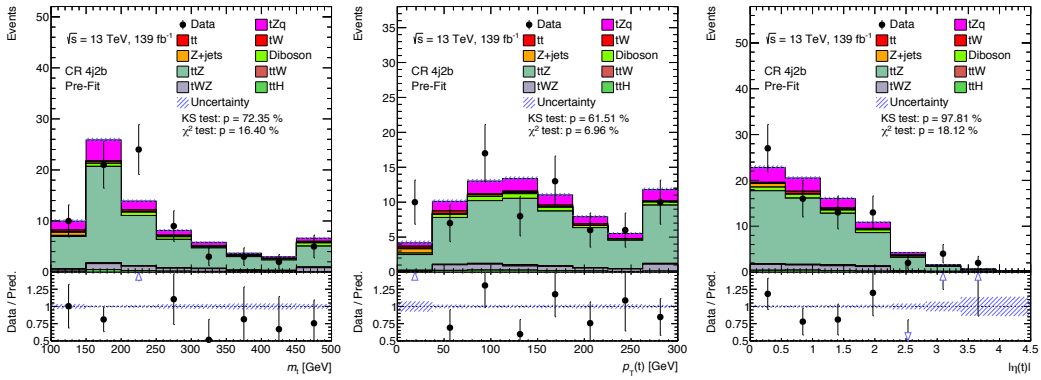


Figure 5.12: Comparison of data and MC predictions for reconstructed event- related quantities for events in the CR-4j2b. The uncertainty shown is the statistical uncertainty.

5.3.3 $t\bar{t}$ CRs Plots and Yields

Finally, the $t\bar{t}$ contribution can be enhanced by requiring the OSSF lepton requirement be removed, effectively removing the requirement on the Z boson and opposite-sign, different-flavor (OSDF) leptons condition is imposed. This phase space is dominated by $t\bar{t}$ events with a fake lepton and a b-jet.

The two $t\bar{t}$ CRs are also defined in Table 5.5. The event yields in the $t\bar{t}$ CRs after the full selection can be found in Table 5.9 and reconstructed variables from the top quark and Z boson are given in Fig. 5.13.

The $t\bar{t}$ CRs suffer from the lowest statistics of all fitted regions. Because of this, two binned histograms per region is used in the statistical analysis.

Process	Number of events	Number of raw events
tZq	0.34 ±0.04	200994
tt	43.40 ±1.23	197797
tW	2.32 ±0.52	3058
Z+jets	0.19 ±0.15	556
Diboson	0.39 ±0.07	21545
ttZ	2.69 ±0.12	262293
ttW	9.76 ±0.26	494284
tWZ	0.44 ±0.10	12649
ttH	3.35 ±0.05	1234180
Total expected	62.87 ±1.37	2427360
Data	71	71

Process	Number of events	Number of raw events
tZq	0.21 ±0.03	139834
tt	20.27 ±0.84	93408
tW	1.06 ±0.61	1251
Z+jets	0.13 ±0.13	417
Diboson	0.30 ±0.04	14178
ttZ	2.47 ±0.12	284950
ttW	5.19 ±0.20	308302
tWZ	0.26 ±0.09	10981
ttH	3.90 ±0.06	1.222370
Total expected	33.80 ±0.96	2075690
Data	49	49

Table 5.9: Numbers of expected events in the CR-2j1b (Left) and CR-3j1b (Right) broken down by process. The uncertainty shown contains only the statistical component.

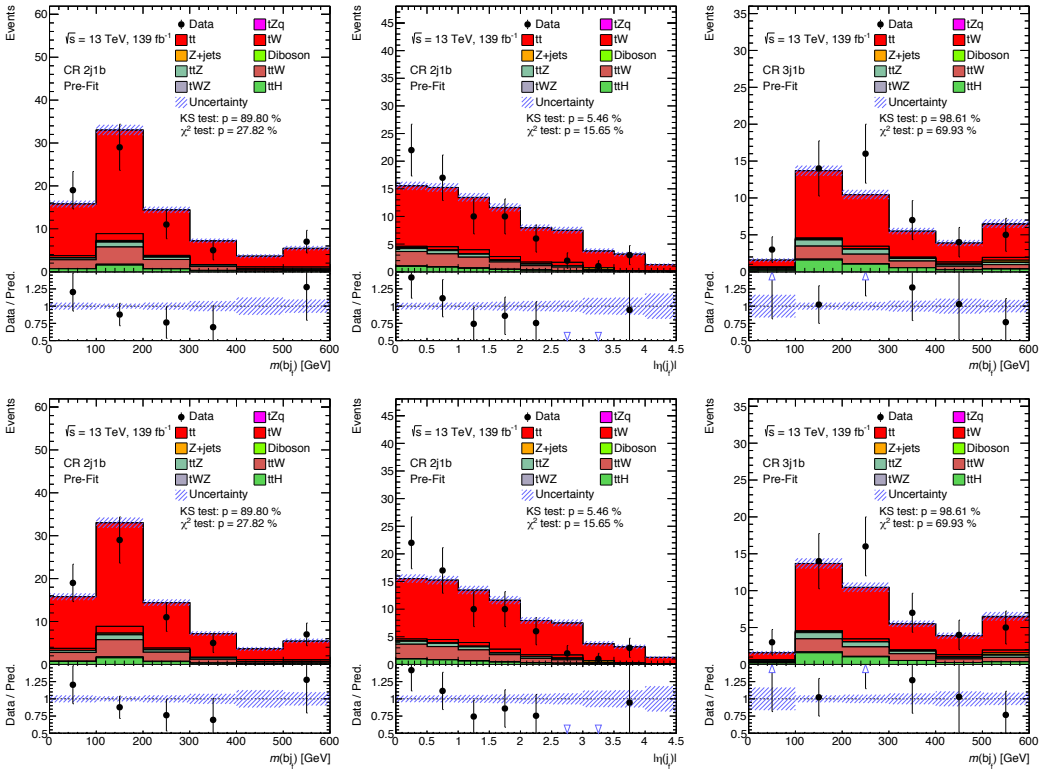


Figure 5.13: Comparison of data and MC predictions for reconstructed event- related quantities for events in the CR-2j1b and CR-3j1b. The uncertainty shown is the statistical uncertainty.