SIGNAL REGION OPTIMIZATION FOR THE SEARCH OF RIGHT HANDED NEUTRINOS

Srishti Patil

1 Introduction

The Standard Model of particle physics (SM) along with the theory of General Relativity (GR) can describe almost everything observed in nature, barring a few experimental and observational facts like baryon assymetry in the universe and the composition and origin of Dark Matter. *Right handed neutrinos* (RHN), if they exist, could be responsible for a lot of these unexplained phenomena. They are promising Dark Matter candidates as well.

This project is a part of the search for RHN using data samples collected by the CMS experiment at the LHC in 2016 [Table 1], corresponding to an integrated luminosity of 36 fb^{-1}. We employ a technique called multilepton analysis [Section 2] for the search and the project focuses on a part of the process called Signal Region (SR) optimization. The results obtained are used to produce limits on the production of right handed neutrinos.

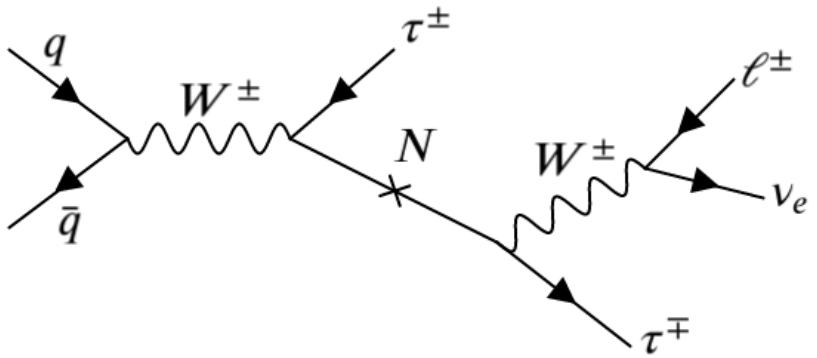


Figure 1: Illustrative leading order Feynman diagram for the production of RHN and its subsequent decay ($N \rightarrow W\tau \rightarrow \nu_\ell \ell \tau$) that may result in a multilepton final state. The taus can decay either hadronically or leptonically, giving us different channels to analyze that have different background profiles

2 Overview

The final state of an event (collision) is the set of particles seen by the detector. A lot of processes Beyond the Standard Model (BSM) produce final states with multiple leptons. Multilepton

Sample (RHN Mass Point)	Total Number of Events (n)	Cross-Section (xsec) (pb)	Luminosity (fb^{-1})
100 GeV	299,200	0.830	360.482
150 GeV	286,200	0.119	2405.042
400 GeV	200,000	0.00272	73529.411

Table 1: Data samples used for this analysis. Three mass points are considered and their respective sample luminosities are calculated using the formula L_{sample} = n_{sample} / $xsec_{sample}$

Analysis is a useful technique for studying such processes. Since a lot of SM processes also produce multilepton final states (called background), we use this strategy to distinguish signal from background.

The model considered here [Figure 1] includes a right handed neutrino (N) that decays into a W boson and a τ, which decay further, giving rise to multilepton final states. For this analysis, we refer to electrons and muons collectively as light leptons (ℓ). Events are then categorized by the multiplicity of light leptons and taus (τ) in these final states, into the following channels: 3ℓ, $2\ell1\tau$, $1\ell2\tau$ and 3τ.

To begin with, we generate the RHN acceptance cutflow table [Section 6, Table 9] to understand the process and how various triggers and selections affect signal acceptance. Next, we attempt to understand the process of RHN production and decay at the gen-level [Section 5.1] (*i.e.* from simulations). Guided by the conclusions drawn from the study of these gen-level plots, we try to carve out a signal region [Section 5] that gives us good signal sensitivity from the phase space. Some selections used to reduce the background are explained in Section 5.2. We use signal significance as a metric to compare signal regions. The analysis is done for two channels: $2\ell1\tau$ and 3ℓ.

3 Object and Event Selections

To better identify particles significant to our analysis, we apply certain basic preselections and triggers to the data. Object selections are listed in Table 2 and Table 3. Triggers are listed below. A few corrections like MET filters (on data and MC both) and tau energy scale corrections are also applied.

Triggers:

- Leading muon $p_{\mathrm{T}} > 26$ GeV
- Leading electron $p_{\mathrm{T}} > 30$ GeV
- For Single Muon PD, HLT_IsoMu24 & HLT_IsoTkMu24 (in data only)
- For Single Electron PD, HLT_Ele27_WPTight_Gsf (in data only)

Muons	Electrons
• $p_{\mathrm{T}} > 10, \lvert\eta\rvert < 2.4$ • Prompt ($d_{xy} < 0.05, d_z < 0.1$ cuts) • POG Medium ID • Relative dB-corrected isolation tight WP (< 0.15 in $\Delta R = 0.4$)	• $p_{\mathrm{T}} > 10, \lvert\eta\rvert < 2.5$ • Prompt ($d_{xy} < 0.05(0.1), d_z < 0.1(0.2)$ in barrel(endcap) region) • Cut-based Medium ID • Relative rho-corrected isolation Medium WP (included in ID)

Table 2: Object selections for light leptons

Taus	
MVA ID (old)	Deep ID (new)
• $p_{\mathrm{T}} > 20, \lvert\eta\rvert < 2.3$ • Prompt ($d_z < 0.2$ cut) • Old decayModeFinding, 1-prong & 3-prong • '2017v2' ID • againstElectron & againstMuon discriminators, loose WP • Cleaning against tight ID light leptons ($\Delta R > 0.4$)	• $p_{\mathrm{T}} > 20, \lvert\eta\rvert < 2.3$ • Prompt ($d_z < 0.2$ cut) • New decayModeFinding, 1-prong & 3-prong • Tau_idDeepTau2017v2p1VSe ≥ 15 & Tau_idDeepTau2017v2p1VSmu ≥ 3 (Loose WP) • Tau_idDeepTau2017v2p1VSjet > 31 (Tight WP)

Table 3: Object selections for taus. A comparision between the two object IDs is given in Section 6.

The events used for analysis are selected only if they pass certain conditions (selections) listed in Table 4. An event is selected for a channel only if it fails the selections for all the previous channels. The order of priority (event selection flow) is illustrated in Figure 2.

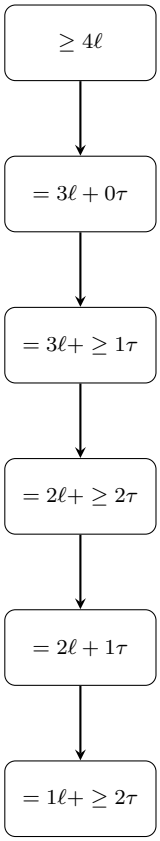
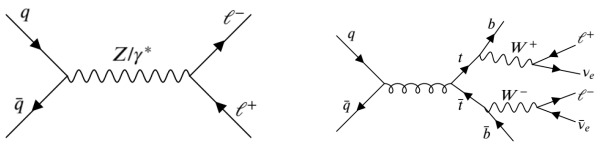


Figure 2: Event selection flow

4 Background

Many SM processes give rise to multilepton final states, and our objective is to distinguish these processes (background) from the signal (RHN). The method used for background estimation in this analysis is Monte-Carlo based. All the background processes considered are listed in Table 5. WZ and ZZ are prompt backgrounds while DY and $t\bar{t}$ are fake (*i.e.* one of the leptons is faked by another particle/jet).



(a) DY *(b)* $t\bar{t}$

Figure 3: Leading order Feynman diagrams for fake background processes. *(a)* The Drell-Yan process. The third lepton can be faked by another particle/jet. *(b)* One of the possible decay products of the $t\bar{t}$ process

Event Selections
3ℓ
• Leading light lepton (e/μ) passing triggers • Altleast two more light leptons (e/μ), $p_{\mathrm{T}} > 10, 10$ GeV • Minimum invariant mass of same flavor (SF) light lepton pair > 12 GeV • Minimum ΔR between all leptons in the event > 0.4 • Triggers (in data only)
$2\ell 1\tau$
• Event should fail 3ℓ selection • Leading light lepton (e/μ) passing triggers • Another light lepton (e/μ), $p_{\mathrm{T}} > 10GeV$ • Minimum invariant mass of same flavor (SF) light lepton pair > 12 GeV • Atleast one hadronic tau, $p_{\mathrm{T}} > 20$ GeV • $\Delta R > 0.4$ among the three leptons • Triggers (in data only)
$1\ell 2\tau$
• Exactly one light lepton (e/μ), for triggering • Atleast two hadronic taus, $p_{\mathrm{T}} > 20, 20$ GeV - Prompt taus pass tight ID, fake taus pass loose ID • $\Delta R > 0.4$ among the three leptons • Triggers (in data only) • Event should fail all other event ($3\ell, 2\ell+ \geq 1\tau$..) selections

Table 4: Event selections for 3ℓ, $2\ell 1\tau$ and $1\ell 2\tau$ channels

Background		Total Number of Events (n)	Cross-Section (xsec) (pb)	Luminosity (fb^{-1})
$2\ell 1\tau$	DY	89,832,690	5765	15.582
	WZ	6,610,401	5.052	1308.472
	ZZ	7,547,891	1.325	5696.521
	$t\bar{t}$	24,265,024	88.29	274.833
3ℓ	DY	89,832,690	5765	15.582
	WZ	1,295,229	5.052	256.378
	ZZ	1,041,601	1.325	786.114
	$t\bar{t}$	24,265,024	88.29	274.833

Table 5: A few background processes and their luminosities. Luminosity is calculated by the formula L = n / xsec

5 Signal Region Optimization

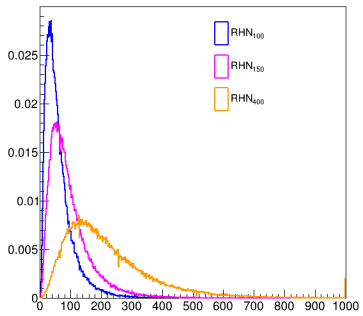
The objective of Signal Region optimization is to find a region in the phase space in which the signal is easily distinguished from background. A metric for the same is signal significance, defined as-

$$\text{Significance} = \frac{\text{Signal}}{\sqrt{\text{Background}}}$$

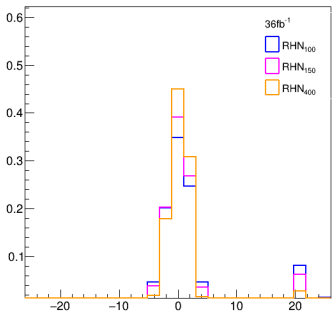
The higher the significance, the better is the signal region. We perform signal region optimization for the $2\ell 1\tau$ and 3ℓ channels separately because the background profiles for these channels are different.

5.1 Understanding the Process with Simulations

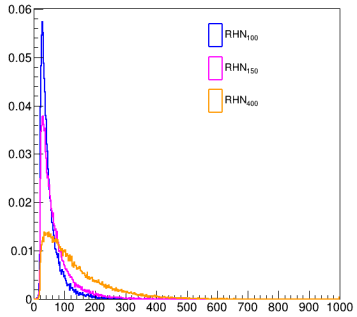
To undrestand the general topology of the events, like the $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ between leptons and the origin (mother particle) of the particles involved, we plot these quantities using simulated data.



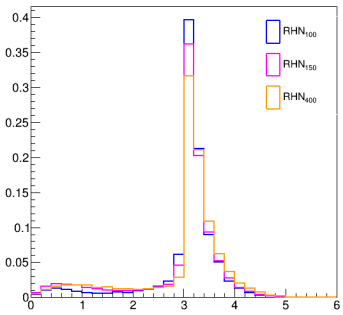
(a) RHN p_{T} [GeV]



(b) PDG-ID of RHN Mothers



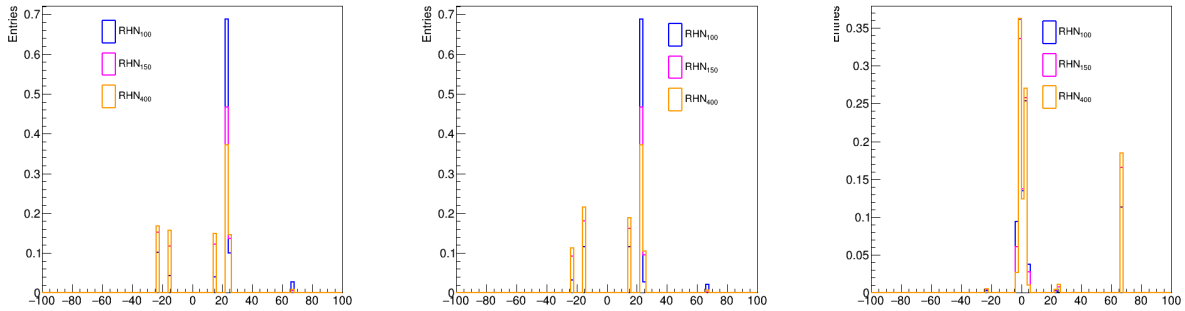
(c) p_{T} of τ coming from PV [GeV]



(d) ΔR (RHN,τ)

Figure 4: Gen plots for the production of RHN (all the plots have been normalized to unity). *(a)* Transverse momenta of RHN of different mass points. We can see that the 400GeV RHN (yellow) is more boosted than the 100GeV RHN (blue). *(b)* PDG-ID of mother particles of RHN. We observe that they are mostly quarks. *(c)* Transverse momentum of the τ coming from the production vertex (PV). *(d)* ΔR between RHN and τ coming from PV. We observe that they are produced back-to-back.

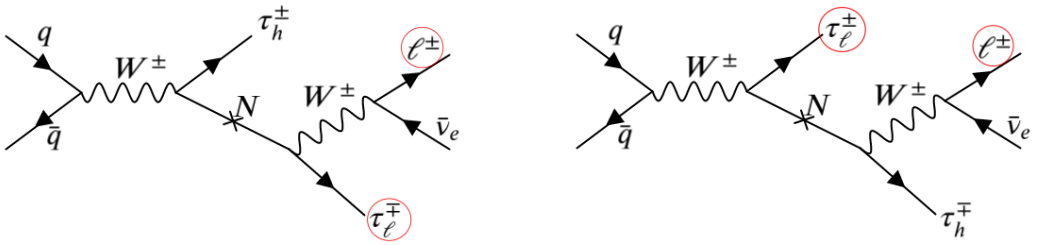
Next, we look at the final state leptons and their plots. Here, we divide the $2\ell 1\tau$ channel further into two categories: (1) OS (opposite-sign) in which the two light leptons have opposite signs and (2) SS (same-sign) in which they have the same sign. The motivation behind this is that these two categories will have different dominating backgrounds. For instance, the $2\ell 1\tau$-OS category is likely to have DY in large numbers because DY cannot produce same-sign dileptons.



(a) ℓ_1 Mother PDG-ID *(b)* ℓ_2 Mother PDG-ID *(c)* τ Mother PDG-ID

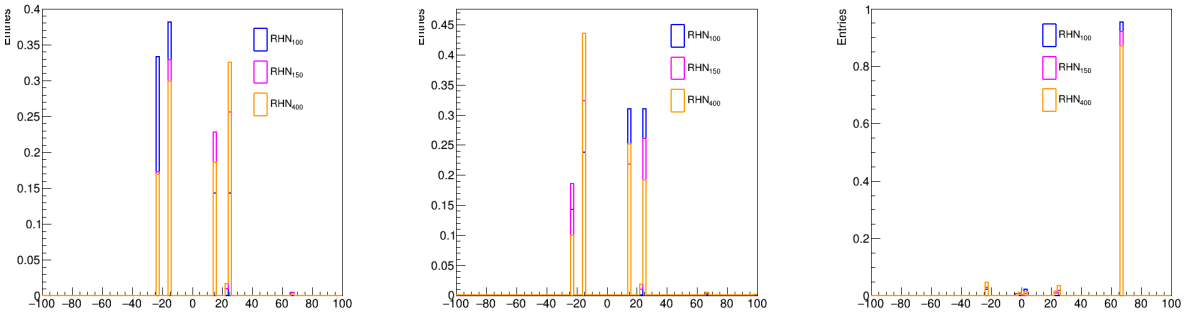
Figure 5: Gen plots for the $2\ell 1\tau$-OS channel (all the plots have been normalized to unity). All the entries for RHN have been made at ID = 66 (instead of 9900012). *(a)* PDG-ID of the mother of the leading light lepton ℓ_1. We see peaks at ± 15 and ± 24 which correspond to taus and W bosons respectively. *(b)* PDG-ID of the mother of the sub-leading light lepton ℓ_2. *(c)* PDG-ID of the mother of the τ. We see peaks at 66 and within the range (-6,6), which correspond to RHN and quarks respectively.

When we compare Figure 5 to the feynman diagram in Figure 6 (a), we see that the observations match. From the diagram, we can further hypothesize that the light lepton pair should be close together (when boosted). Similarly, the observations from Figure 7 (SS gen plots) match with the ones from the Feynman diagram [Figure 6 (b)].



(a) OS *(b)* SS

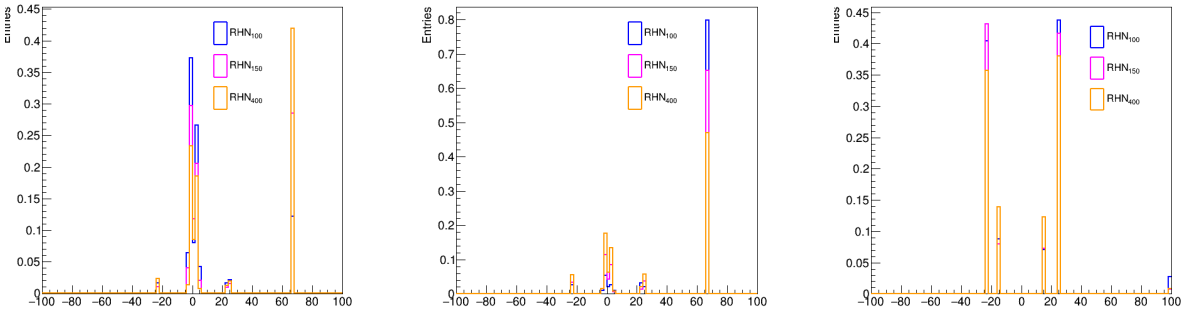
Figure 6: Feynman diagrams for $2\ell 1\tau$ OS and SS channels. *(a)* The PV τ decays hadronically and the τ coming from the RHN decays leptonically. Since RHN is electrically neutral, the light lepton pair in the final state will be OS in this case. *(b)* The only way an SS pair can be produced is for one of the light leptons to be a product of the PV τ decay, since the daughters of RHN will have to be OS (note that this can result in an OS pair as well).



(a) ℓ_1 Mother PDG-ID *(b)* ℓ_2 Mother PDG-ID *(c)* τ Mother PDG-ID

Figure 7: Gen plots for the $2\ell 1\tau$-SS channel (all the plots have been normalized to unity). All the entries for RHN have been made at ID = 66 (instead of 9900012). *(a)* PDG-ID of the mother of the leading light lepton ℓ_1. Again, we see peaks at taus and W bosons, but this time the peak at τ is comparable to the one at W boson. *(b)* PDG-ID of the mother of the sub-leading light lepton ℓ_2. *(c)* PDG-ID of the mother of the τ. We see a peak only at RHN.

In conclusion, we decide to divide the $2\ell 1\tau$ channel into OS and SS categories, owing to the differences in their production and backgrounds. For a better understanding, we study the $1\ell 2\tau$ channel in a similar manner [Figure 8]. We observe that the PV τ is almost always the leading τ.



(a) τ_1 Mother PDG-ID *(b)* τ_2 Mother PDG-ID *(c)* ℓ Mother PDG-ID

Figure 8: Gen plots for the $1\ell 2\tau$ channel (all the plots have been normalized to unity). All the entries for RHN have been made at ID = 66 (instead of 9900012). *(a)* PDG-ID of the mother of the leading tau τ_1. We see peaks at RHN and quarks. *(b)* PDG-ID of the mother of the sub-leading tau τ_2. *(c)* PDG-ID of the mother of the light lepton ℓ.

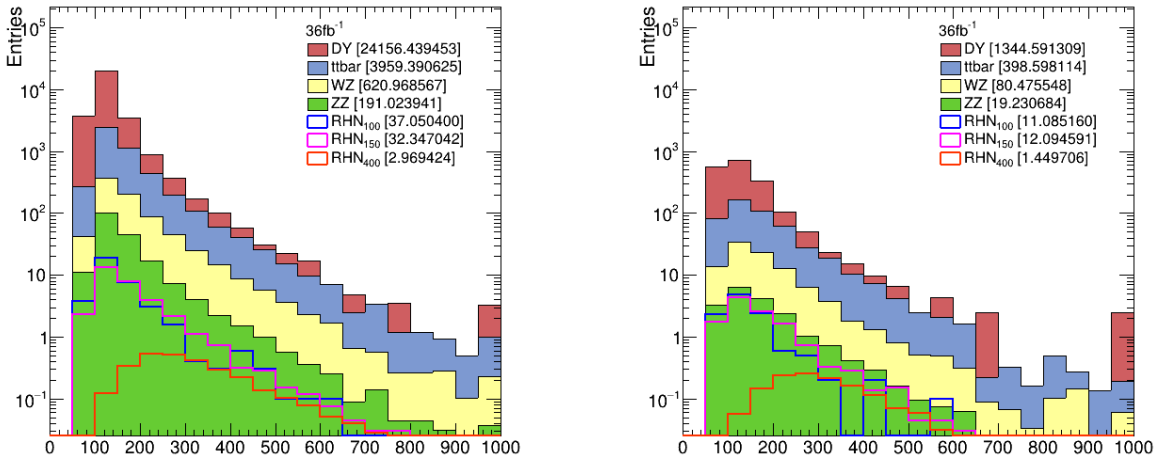
5.2 Reducing Background

Based on our understanding of the SM background processes and the RHN decay, we can apply specific selections to kill the background to a large extent. The following conditions or vetoes were applied:

1. **On-Z veto**: In the process of RHN decay, the light leptons are daughters of W bosons or τ. In the background processes (like ZZ, WZ and DY), many of the light leptons are daughters of the Z boson. Thus, we exclude the on-Z region, *i.e.* we exculde events for which the invariant mass of the light lepton pair ($M_{\ell_1\ell_2}$) is in the on-Z range. We define this range to be 76 GeV $-$ 106 GeV (Z-mass is 91 GeV). Table 6 and Figure 9 show how this veto affects the background and signal.

	Region	Events Before Veto	Events After Veto
DY	$2\ell 1\tau$	$10,512$	745
	$2\ell 1\tau$-OS	$10,456$	738
	$2\ell 1\tau$-SS	56	7
$t\bar{t}$	$2\ell 1\tau$	$3,719$	$1,046$
	$2\ell 1\tau$-OS	$3,377$	942
	$2\ell 1\tau$-SS	342	104
M100	$2\ell 1\tau$	413	154
	$2\ell 1\tau$-OS	371	125
	$2\ell 1\tau$-SS	42	27
M150	$2\ell 1\tau$	$2,560$	$1,199$
	$2\ell 1\tau$-OS	$2,161$	880
	$2\ell 1\tau$-SS	390	305

Table 6: The number of events before and after applying the On-Z veto. The signal also reduces with the background but the overall significance improves in every region.



(a) L_{T} before On-Z veto [GeV]

(b) L_{T} after On-Z veto [GeV]

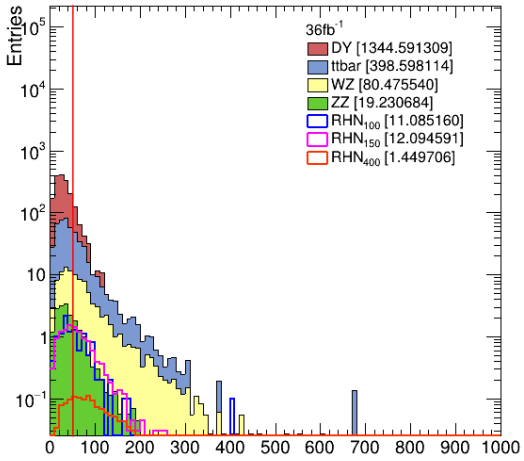
Figure 9: L_{T} plots before and after the application of the On-Z veto (L_{T} is defined as the scalar sum of the p_{T} for the final state leptons). All the sample luminosities have been scaled to the integrated luminosity of 36 fb^{-1} so that the comparision is accurate.

2. $\boldsymbol{N_{Bjets}}$ **veto**: This veto is majorly for cutting down $t\bar{t}$ background. The $t\bar{t}$ decay produces b-jets [Figure 3] while RHN does not. So, we include events which have no b-jets ($N_{Bjets} = 0$).

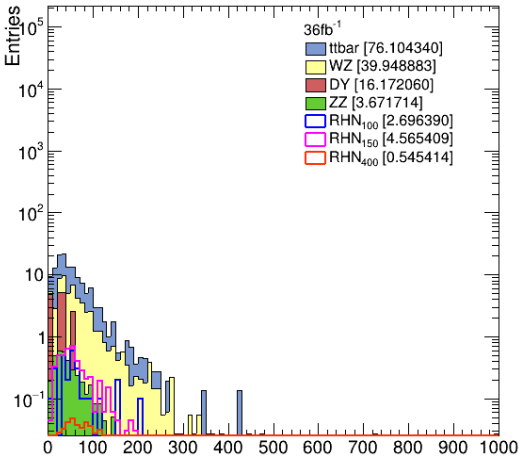
3. $\boldsymbol{H_T^{50} < 100}$ **veto**: RHN decay usually does not result in high energy jets, while other background processes might. So we put a upper limit on the sum of p_{T} of all jets in an event. This variable is named H_T. We use a modified variable H_T^{50} which puts a lower limit of 50 on jet p_{T}. This is to avoid pileup and consider real jets only.

5.3 $2\ell 1\tau$

Guided by the observations and conclusions from the gen plots, and after applying vetoes to kill background, we try to find a region with appropriate selections. To begin, we study some basic plots:



(a) $p_{\mathrm{T}}^{\mathrm{miss}}$ ($2\ell 1\tau$-OS) [GeV]



(b) $p_{\mathrm{T}}^{\mathrm{miss}}$ ($2\ell 1\tau$-SS) [GeV]

$2\ell 1\tau$-OS	$2\ell 1\tau$-OS $p_{\mathrm{T}}^{\mathrm{miss}} < 50$	$2\ell 1\tau$-OS $p_{\mathrm{T}}^{\mathrm{miss}} > 50$
$S/\sqrt{B}$ (100) = 0.2582 $S/\sqrt{B}$ (150) = 0.2817 $S/\sqrt{B}$ (400) = 0.0337	$S/\sqrt{B}$ (100) = 0.1564 $S/\sqrt{B}$ (150) = 0.1399 $S/\sqrt{B}$ (400) = 0.0079	$S/\sqrt{B}$ (100) = 0.2877 $S/\sqrt{B}$ (150) = 0.3789 $S/\sqrt{B}$ (400) = 0.0633

Figure 10: $p_{\mathrm{T}}^{\mathrm{miss}}$ for $2\ell 1\tau$ OS and SS channels. In *(a)*, we observe that most of the DY background lies in the low $p_{\mathrm{T}}^{\mathrm{miss}}$ range. This is confirmed by the significance table. Thus, $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ will be our first selection, *i.e.* we exculde the region below 50 GeV.

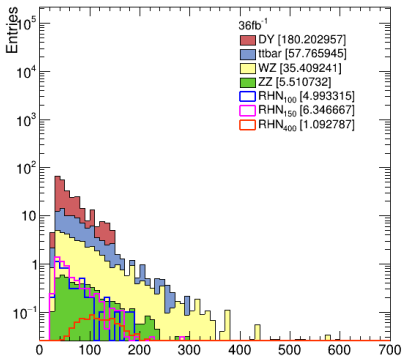
5.3.1 $2\ell 1\tau$ - OS

After studying the $p_{\mathrm{T}}^{\mathrm{miss}}$ plots for the OS channel [Figure 10], we decide to select the region in which $p_{\mathrm{T}}^{\mathrm{miss}}$ is above 50. This eliminates a lot of DY background. All the following plots are of the $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ region. In Figure 11, we have the p_{T} and ΔR plots. We observe that the light lepton pair for RHN events is generally closer together [Fig. 11*(d)*] and that the sub-leading light lepton and the tau are far apart [Fig. 11*(f)*](as expected from simulations).

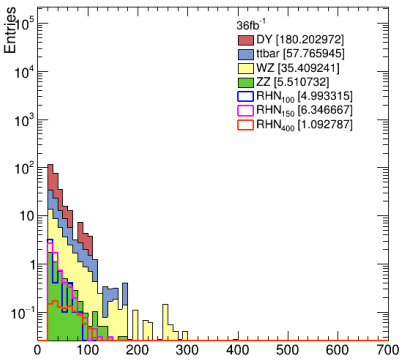
Next, we plot a number of physical quatities like $\Delta\phi$, $m_{\ell_1\ell_2}$, transverse mass (m_T), visible transverse mass, etc. We also define a few new variables like $E_{\mathrm{T}}^{\mathrm{miss}}/m_T$ and m_T (ℓ_{res}, $E_{\mathrm{T}}^{\mathrm{miss}}$) where ℓ_{res} is defined as the vector sum of one of the two OS pairs. Transverse mass is defined as -

$$m_T = \sqrt{2p_{\mathrm{T}}^{miss}p_{\mathrm{T}}^{\ell}[1 - cos(\Delta\phi_{m_T})]}$$

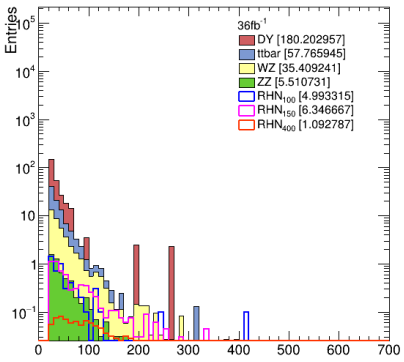
where p_{T}^{ℓ} is the p_{T} of the lepton under consideration.



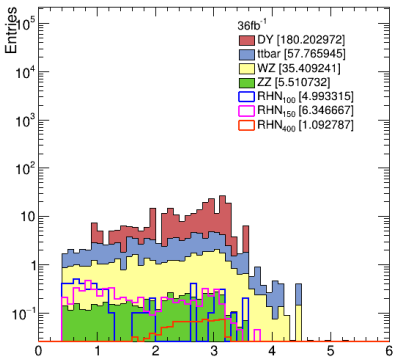
(a) p_{T} of leading light lepton ℓ_1[GeV]



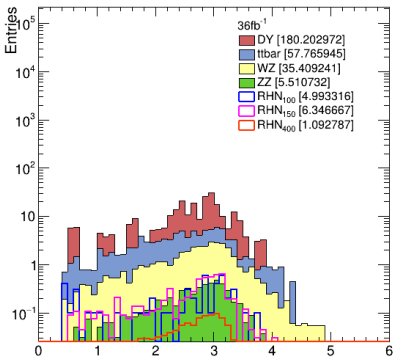
(b) p_{T} of sub-leading light lepton ℓ_2[GeV]



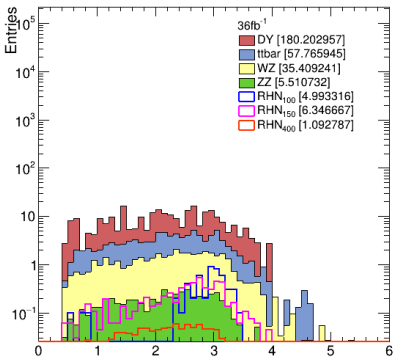
(c) p_{T} of τ [GeV]



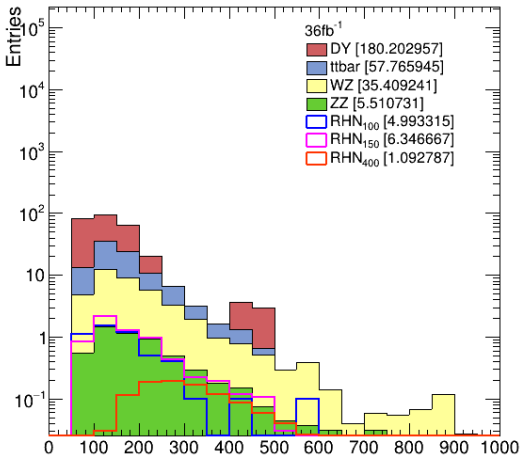
(d) $\Delta R\,(\ell_1,\ell_2)$



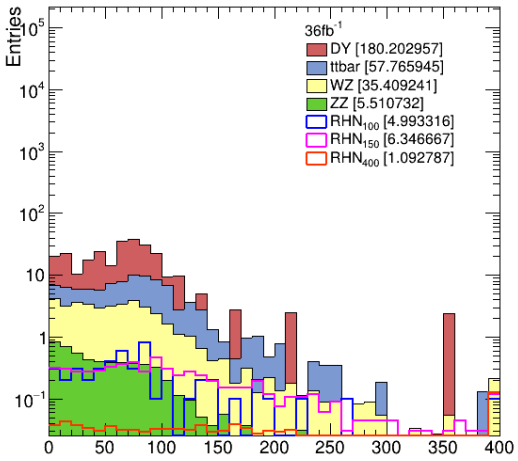
(e) $\Delta R\,(\ell_1,\tau)$



(f) $\Delta R\,(\ell_2,\tau)$



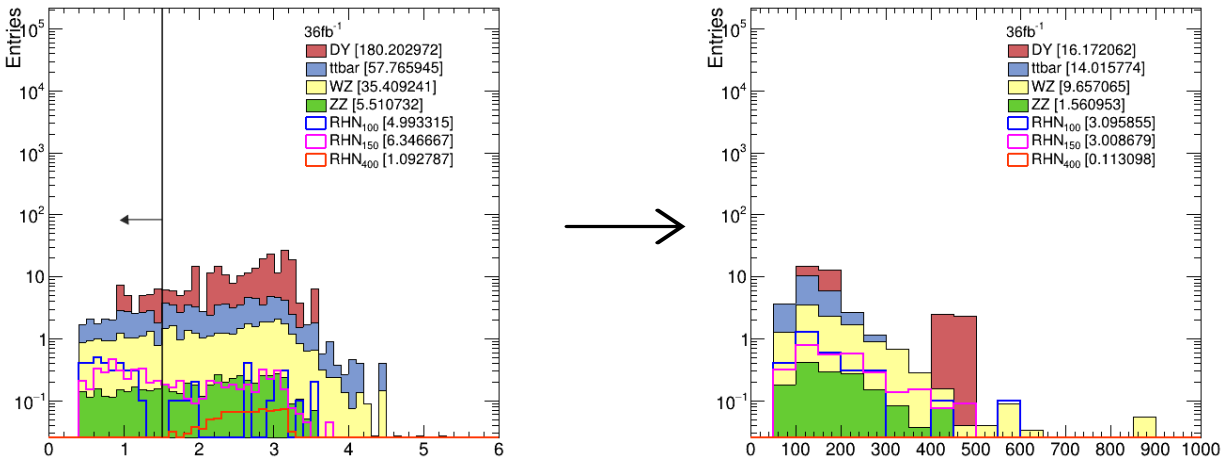
(g) L_{T} [GeV]



(h) $m_T\,(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ [GeV]

Figure 11: Some basic distributions for the $2\ell1\tau$-OS $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ region.

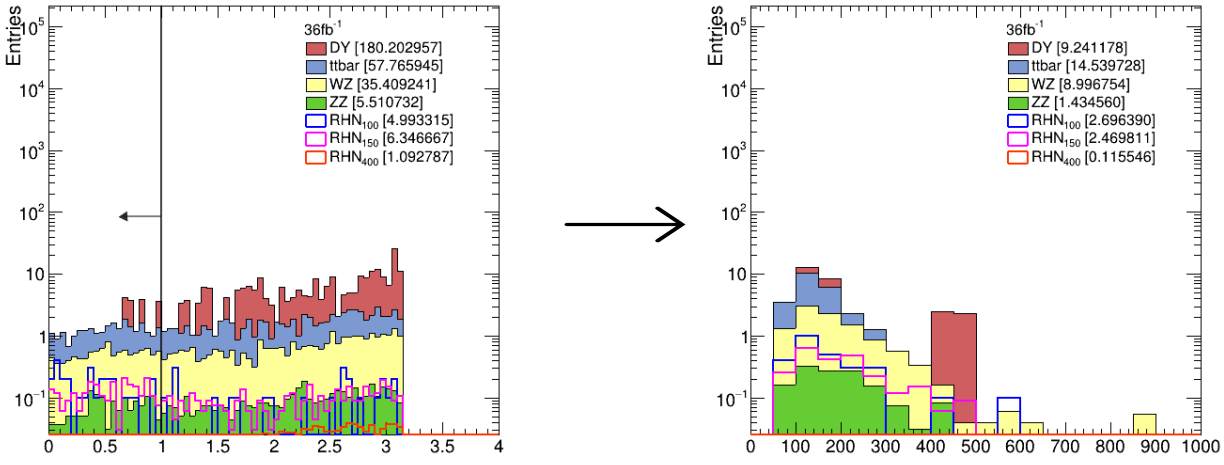
Based on these plots, we attempt to find the optimal signal region so that signal significance is maximized. We try a bunch of selections (or cuts) for the same. The ones that yield the best results are given in Figures 12 to 16 (L_{T} is used as the discriminating variable). A summary of all the selections is given in Table 7.



(a) $\Delta R\,(\ell_1, \ell_2)$ *(c)* L_{T} [GeV]

Significance before cut	Significance after cut
$S/\sqrt{B}\,(100) = 0.2990$ $S/\sqrt{B}\,(150) = 0.3800$ $S/\sqrt{B}\,(400) = 0.0654$	$S/\sqrt{B}\,(100) = 0.4811$ $S/\sqrt{B}\,(150) = 0.4675$ $S/\sqrt{B}\,(400) = 0.0176$

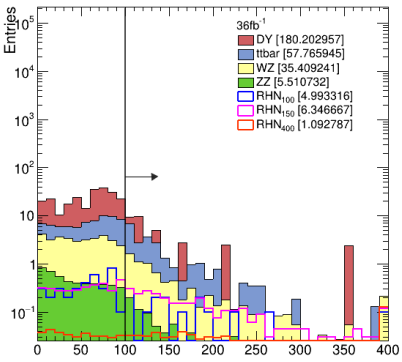
Figure 12: $\Delta R\,(\ell_1, \ell_2) < 1.5$ for $2\ell 1\tau$-OS $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ events.



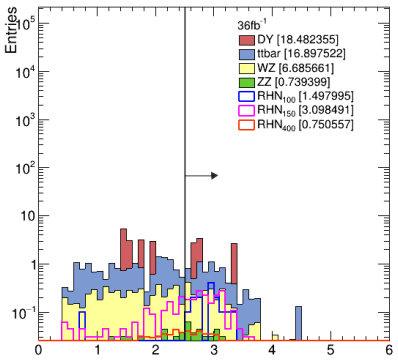
(a) $\Delta\phi\,(\ell_1, \ell_2)$ *(c)* L_{T} [GeV]

Significance before cut	Significance after cut
$S/\sqrt{B}\,(100) = 0.2990$ $S/\sqrt{B}\,(150) = 0.3800$ $S/\sqrt{B}\,(400) = 0.0654$	$S/\sqrt{B}\,(100) = 0.4610$ $S/\sqrt{B}\,(150) = 0.4222$ $S/\sqrt{B}\,(400) = 0.0197$

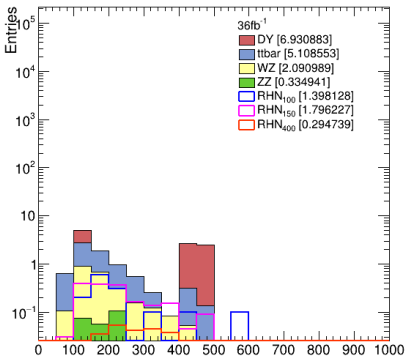
Figure 13: $\Delta\phi\,(\ell_1, \ell_2) < 1$ for $2\ell 1\tau$-OS $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ events.



(a) $m_T\,(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ [GeV]



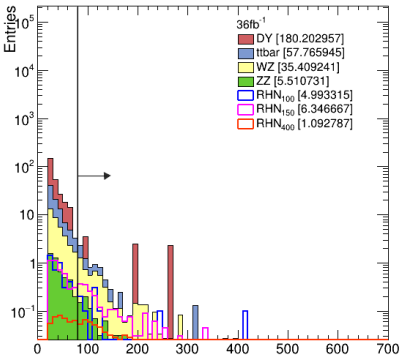
(c) $\Delta R\,(\ell_2, \tau)$



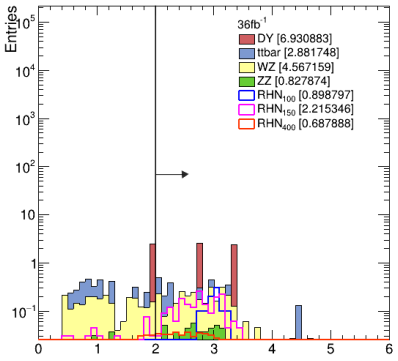
(e) L_{T} [GeV]

Significance before cuts	Significance after m_T cut	Significance after ΔR cut
$S/\sqrt{B}\,(100) = 0.2990$ $S/\sqrt{B}\,(150) = 0.3800$ $S/\sqrt{B}\,(400) = 0.0654$	$S/\sqrt{B}\,(100) = 0.2289$ $S/\sqrt{B}\,(150) = 0.4736$ $S/\sqrt{B}\,(400) = 0.1147$	$S/\sqrt{B}\,(100) = 0.3676$ $S/\sqrt{B}\,(150) = 0.4723$ $S/\sqrt{B}\,(400) = 0.0774$

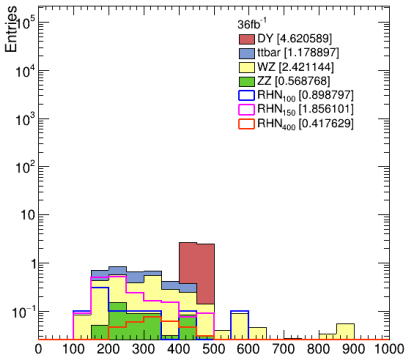
Figure 14: $m_T\,(\tau, E_{\mathrm{T}}^{\mathrm{miss}}) > 100$ and $\Delta R\,(\ell_2, \tau) > 2.5$ for $2\ell 1\tau$-OS $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ events.



(a) p_{T} of τ [GeV]



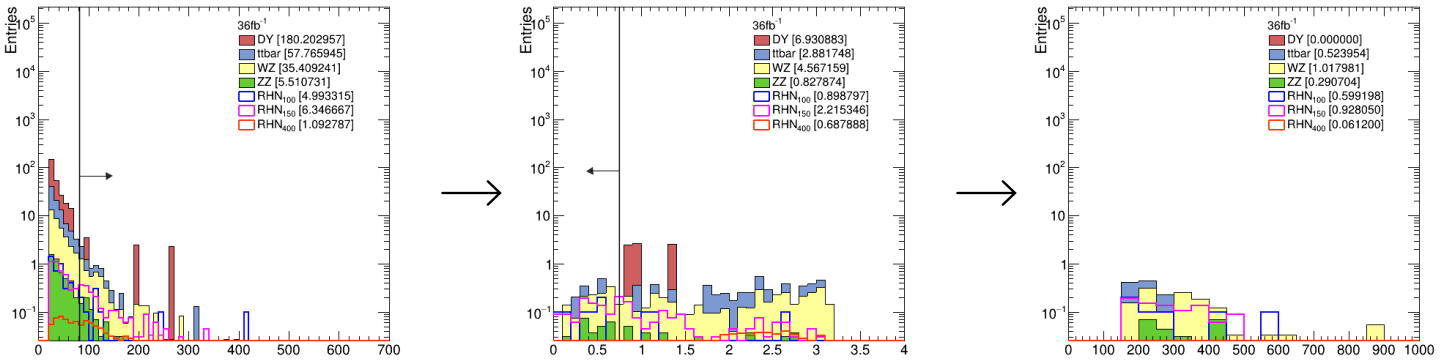
(c) $\Delta R\,(\ell_2, \tau)$



(e) L_{T} [GeV]

Significance before cuts	Significance after p_{T} cut	Significance after ΔR cut
$S/\sqrt{B}\,(100) = 0.2990$ $S/\sqrt{B}\,(150) = 0.3800$ $S/\sqrt{B}\,(400) = 0.0654$	$S/\sqrt{B}\,(100) = 0.2305$ $S/\sqrt{B}\,(150) = 0.5681$ $S/\sqrt{B}\,(400) = 0.1764$	$S/\sqrt{B}\,(100) = 0.3032$ $S/\sqrt{B}\,(150) = 0.6260$ $S/\sqrt{B}\,(400) = 0.1408$

Figure 15: p_{T} of $\tau > 80$ and $\Delta R\,(\ell_2, \tau) > 2$ for $2\ell 1\tau$-OS $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ events.

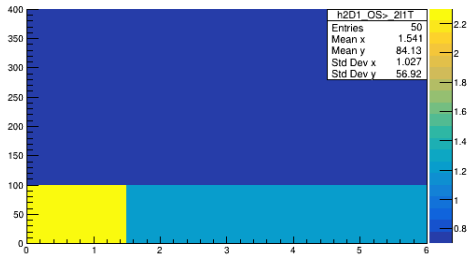


(a) p_{T} of τ [GeV] *(c)* $\Delta\phi\,(\ell_1, \ell_2)$ *(e)* L_{T} [GeV]

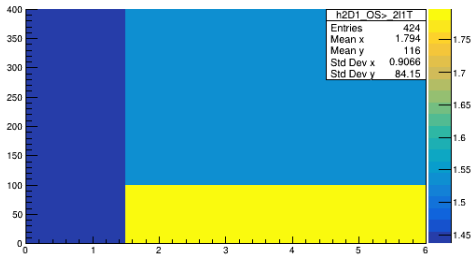
Significance before cuts	Significance after p_{T} cut	Significance after $\Delta\phi$ cut
$S/\sqrt{B}\,(100) = 0.2990$ $S/\sqrt{B}\,(150) = 0.3800$ $S/\sqrt{B}\,(400) = 0.0654$	$S/\sqrt{B}\,(100) = 0.2305$ $S/\sqrt{B}\,(150) = 0.5681$ $S/\sqrt{B}\,(400) = 0.1764$	$S/\sqrt{B}\,(100) = 0.4426$ $S/\sqrt{B}\,(150) = 0.6855$ $S/\sqrt{B}\,(400) = 0.0452$

Figure 16: p_{T} of $\tau > 80$ and $\Delta\phi\,(\ell_1, \ell_2) < 0.75$ for $2\ell 1\tau$-OS $p_{\mathrm{T}}^{\mathrm{miss}} > 50$ events.

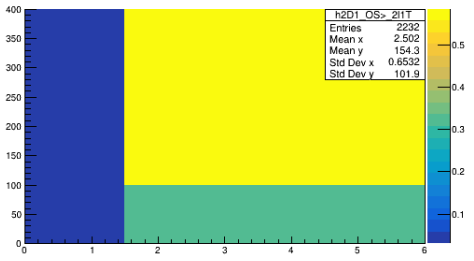
As expected, the ΔR variables are important and selecting regions based on them can give us good signinificance for all the mass points. We therefore plot 2-D histograms with these variables on one axis. Different regions on the 2-D plane can then be selected as signal regions. These plots are shown in Figures 17 to 19.



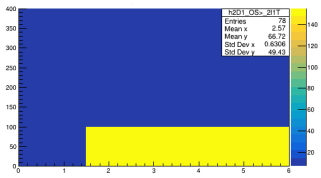
$m_T(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ v/s $\Delta R\,(\ell_1, \ell_2)$ (100 GeV)



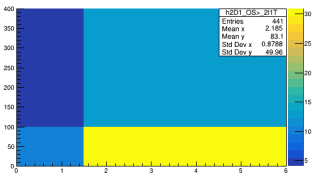
$m_T(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ v/s $\Delta R\,(\ell_1, \ell_2)$ (150 GeV)



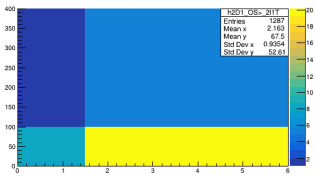
$m_T(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ v/s $\Delta R\,(\ell_1, \ell_2)$ (400 GeV)



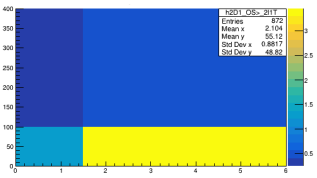
DY



$t\bar{t}$



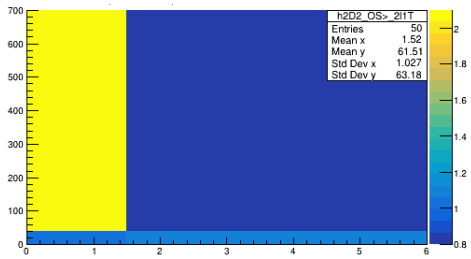
WZ



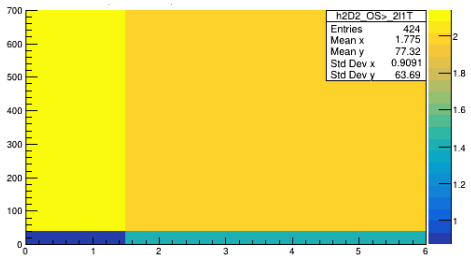
ZZ

Significance before cuts	Significance after cuts
$S/\sqrt{B}\,(100) = 0.2990$ $S/\sqrt{B}\,(150) = 0.3800$ $S/\sqrt{B}\,(400) = 0.0654$	$S/\sqrt{B}\,(100) = 0.443$ [bin $(1, 1)$] $S/\sqrt{B}\,(150) = 0.376$ [bin $(1, 2)$] $S/\sqrt{B}\,(400) = 0.110$ [bin $(2, 2)$]

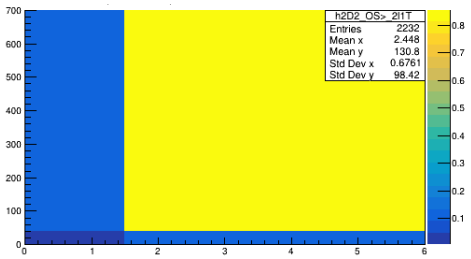
Figure 17: [$2\ell 1\tau$-OS 2-D histogram] $m_T(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ v/s $\Delta R\,(\ell_1, \ell_2)$ (all m_T axes are in GeV units). The bins (x,y) with the best significance for each mass point have been mentioned in parentheses.



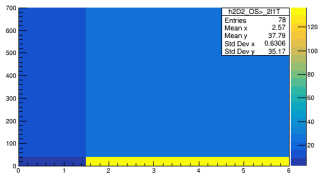
$\tau\ p_{\mathrm{T}}$ v/s $\Delta R\,(\ell_1, \ell_2)$ (100 GeV)



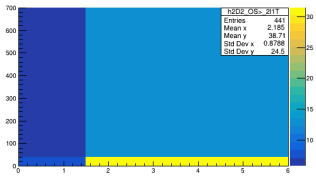
$\tau\ p_{\mathrm{T}}$ v/s $\Delta R\,(\ell_1, \ell_2)$ (150 GeV)



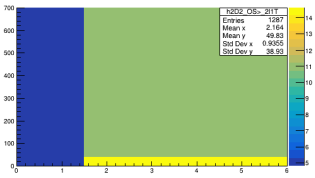
$\tau\ p_{\mathrm{T}}$ v/s $\Delta R\,(\ell_1, \ell_2)$ (400 GeV)



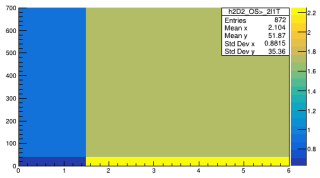
DY



$t\bar{t}$



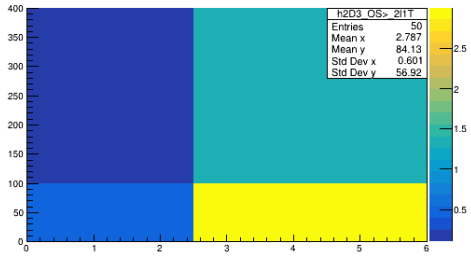
WZ



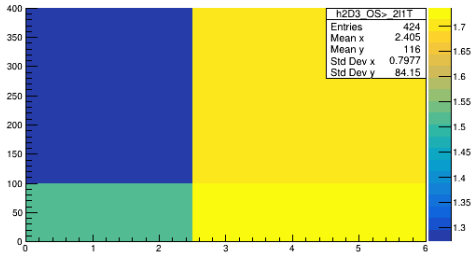
ZZ

Significance before cuts	Significance after cuts
$S/\sqrt{B}\,(100) = 0.2990$ $S/\sqrt{B}\,(150) = 0.3800$ $S/\sqrt{B}\,(400) = 0.0654$	$S/\sqrt{B}\,(100) = 0.416$ [bin $(1, 2)$] $S/\sqrt{B}\,(150) = 0.425$ [bin $(1, 2)$] $S/\sqrt{B}\,(400) = 0.117$ [bin $(2, 2)$]

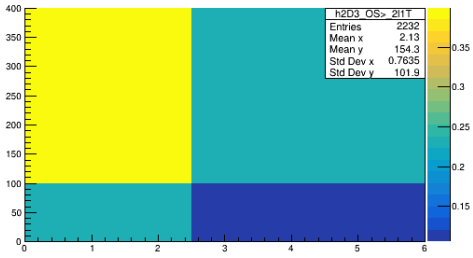
Figure 18: [$2\ell 1\tau$-OS 2-D histogram] $\tau\ p_{\mathrm{T}}$ v/s $\Delta R\,(\ell_1, \ell_2)$ (all p_{T} axes are in GeV units). The bins (x,y) with the best significance for each mass point have been mentioned in parentheses.



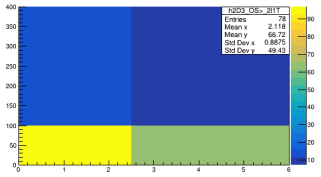
$m_T(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ v/s $\Delta R\,(\ell_2, \tau)$ (100 GeV)



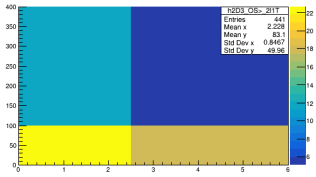
$m_T(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ v/s $\Delta R\,(\ell_2, \tau)$ (150 GeV)



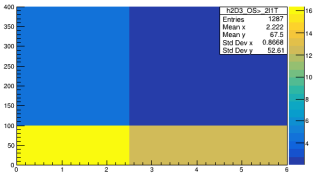
$m_T(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ v/s $\Delta R\,(\ell_2, \tau)$ (400 GeV)



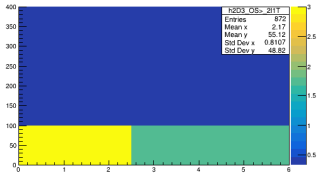
DY



$t\bar{t}$



WZ



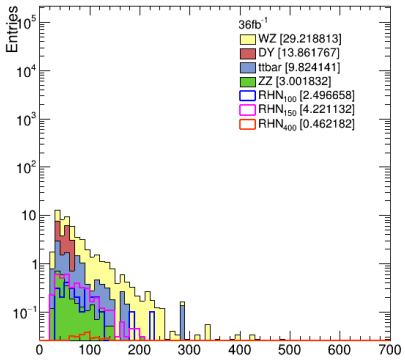
ZZ

Significance before cuts	Significance after cuts
$S/\sqrt{B}\,(100) = 0.2990$ $S/\sqrt{B}\,(150) = 0.3800$ $S/\sqrt{B}\,(400) = 0.0654$	$S/\sqrt{B}\,(100) = 0.342$ [bin $(2, 2)$] $S/\sqrt{B}\,(150) = 0.450$ [bin $(2, 2)$] $S/\sqrt{B}\,(400) = 0.075$ [bin $(1, 2)$]

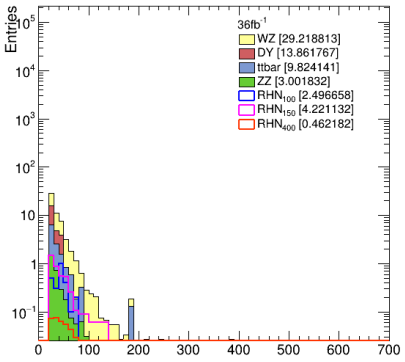
Figure 19: [$2\ell 1\tau$-OS 2-D histogram] $m_T(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ v/s $\Delta R\,(\ell_2, \tau)$ (all m_T axes are in GeV units). The bins (x,y) with the best significance for each mass point have been mentioned in parentheses.

5.3.2 $2\ell 1\tau$ - SS

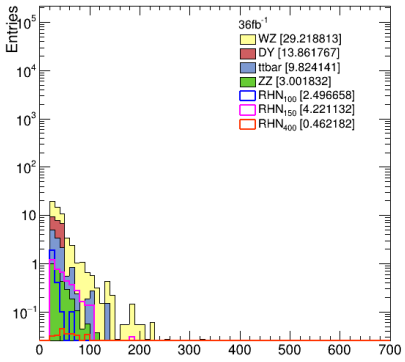
Following a strategy similar to that of the $2\ell 1\tau$-OS channel, we look at the basic p_{T}, ΔR, m_T and L_{T} plots [Figure 20] and then apply the appropriate selections. We try a number of cuts on newly introduced variables specific to this channel (like ΔR between the same-sign lepton pair). We also make cuts based on the presence of an OS lepton pair. Unfortunately, none of the selections give a significance better than the initial one. The optimal signal region seems to be the region with events that have an OS pair. A summary of all the selections is given in Table 7.



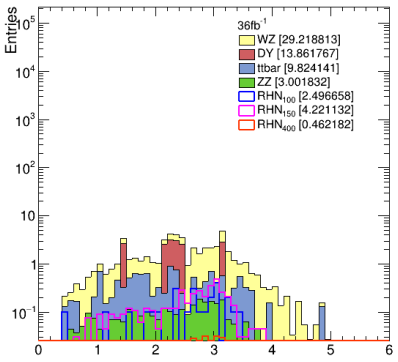
p_{T} of leading light lepton ℓ_1[GeV]



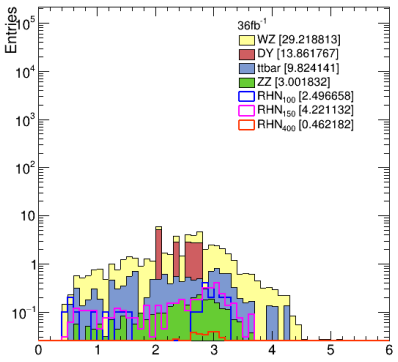
p_{T} of sub-leading light lepton ℓ_2[GeV]



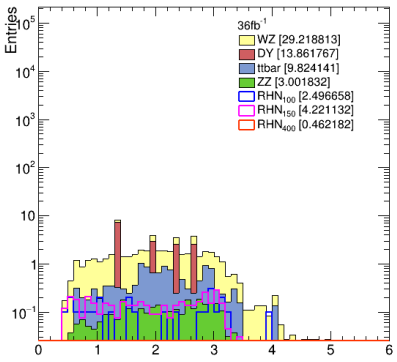
p_{T} of τ [GeV]



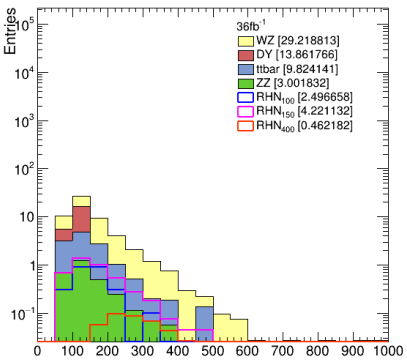
$\Delta R\,(\ell_1,\ell_2)$



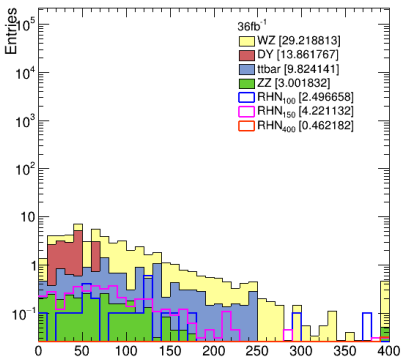
$\Delta R\,(\ell_1,\tau)$



$\Delta R\,(\ell_2,\tau)$

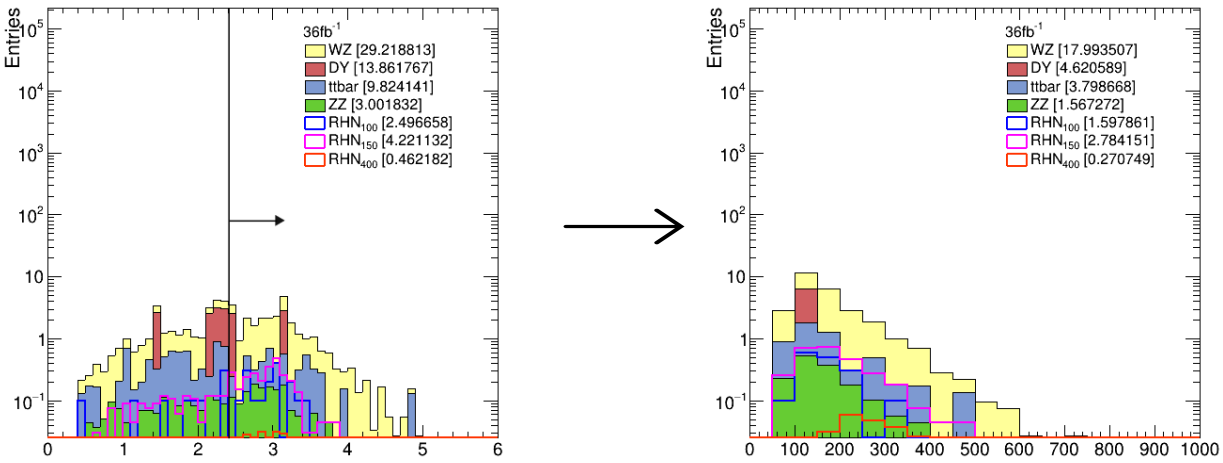


L_{T} [GeV]



m_T (τ, $E_{\mathrm{T}}^{\mathrm{miss}}$) [GeV]

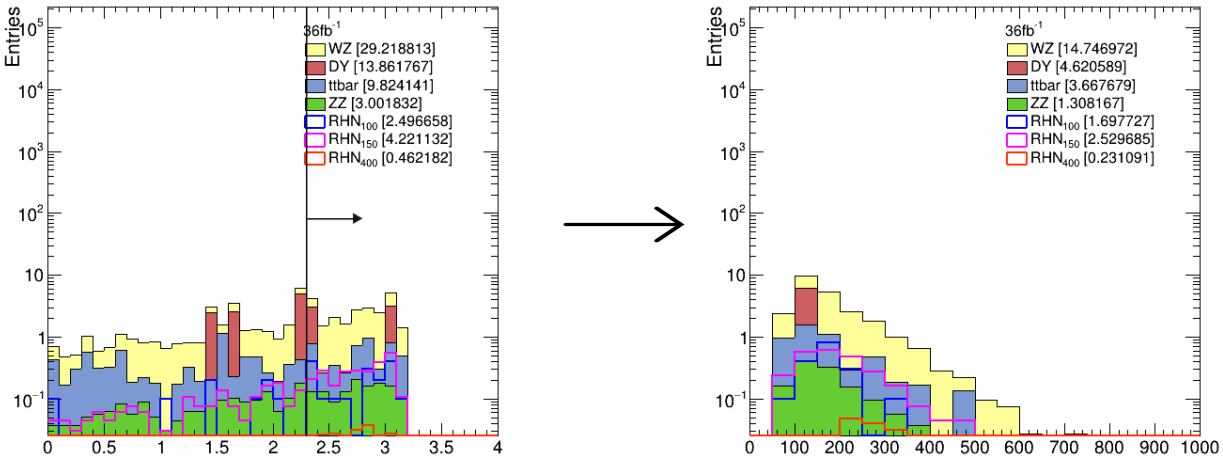
Figure 20: Some basic distributions for the $2\ell 1\tau$-SS region.



$\Delta R\,(\ell_1,\ell_2)$ L_{T} [GeV]

Significance before cut	Significance after cut
$S/\sqrt{B}\,(100) = 0.3339$ $S/\sqrt{B}\,(150) = 0.5645$ $S/\sqrt{B}\,(400) = 0.0618$	$S/\sqrt{B}\,(100) = 0.3021$ $S/\sqrt{B}\,(150) = 0.5263$ $S/\sqrt{B}\,(400) = 0.0512$

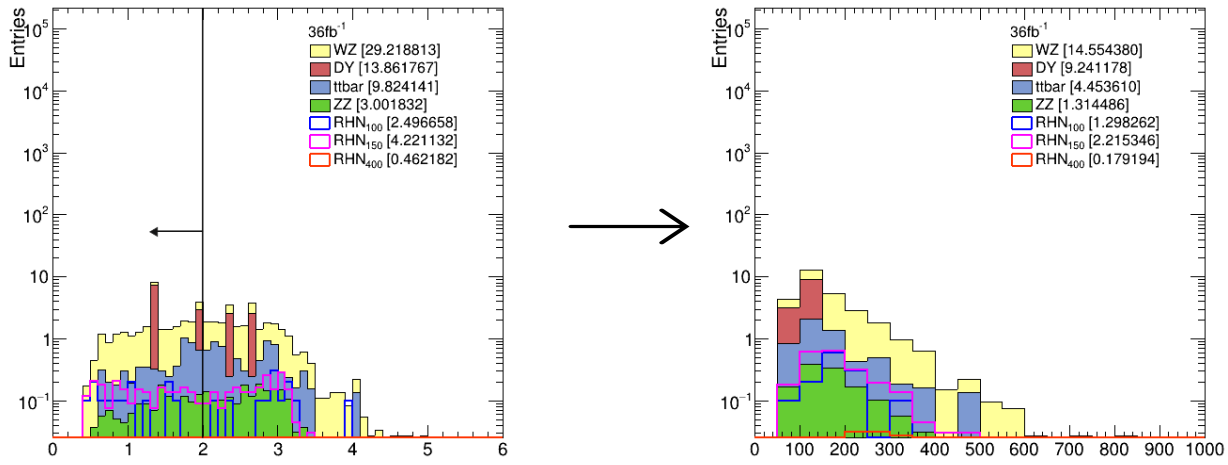
Figure 21: $\Delta R\,(\ell_1,\ell_2) > 2.4$ for $2\ell 1\tau$-SS channel.



$\Delta\phi\,(\ell_1,\ell_2)$ L_{T} [GeV]

Significance before cut	Significance after cut
$S/\sqrt{B}\,(100) = 0.3339$ $S/\sqrt{B}\,(150) = 0.5645$ $S/\sqrt{B}\,(400) = 0.0618$	$S/\sqrt{B}\,(100) = 0.3441$ $S/\sqrt{B}\,(150) = 0.5127$ $S/\sqrt{B}\,(400) = 0.0468$

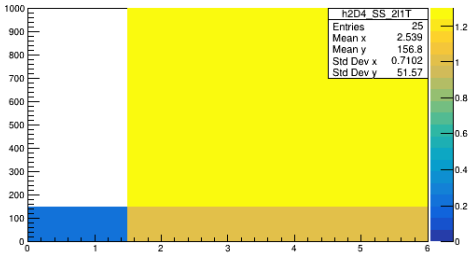
Figure 22: $\Delta\phi\,(\ell_1,\ell_2) > 2.3$ for $2\ell 1\tau$-SS channel.



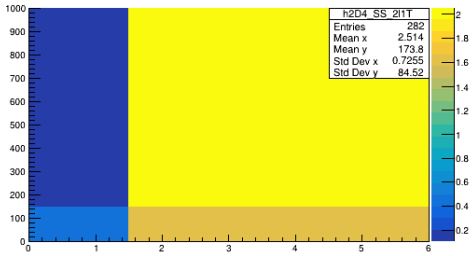
$\Delta R\,(\ell_2, \tau)$ L_{T} [GeV]

Significance before cut	Significance after cut
$S/\sqrt{B}\,(100) = 0.3339$ $S/\sqrt{B}\,(150) = 0.5645$ $S/\sqrt{B}\,(400) = 0.0618$	$S/\sqrt{B}\,(100) = 0.2388$ $S/\sqrt{B}\,(150) = 0.4074$ $S/\sqrt{B}\,(400) = 0.0329$

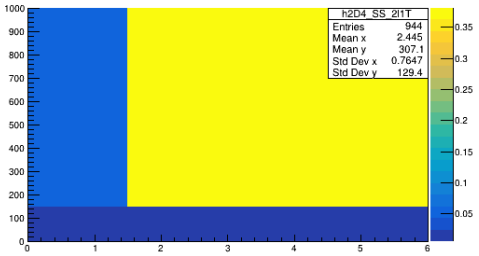
Figure 23: $\Delta R\,(\ell_2,\tau) < 2$ for $2\ell 1\tau$-SS channel.



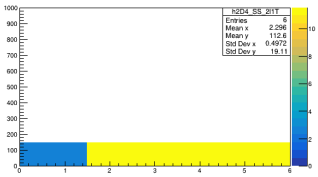
L_{T} v/s $\Delta R\,(\ell_1, \ell_2)$ (100 GeV)



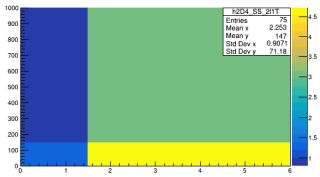
L_{T} v/s $\Delta R\,(\ell_1, \ell_2)$ (150 GeV)



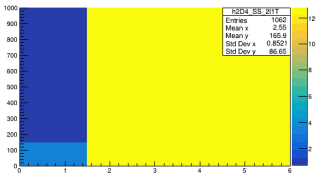
L_{T} v/s $\Delta R\,(\ell_1, \ell_2)$ (400 GeV)



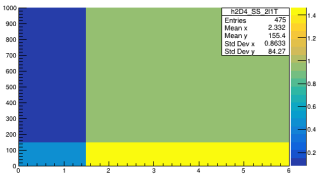
DY



$t\bar{t}$



WZ



ZZ

Significance before cuts	Significance after cuts
$S/\sqrt{B}\,(100) = 0.3339$ $S/\sqrt{B}\,(150) = 0.5645$ $S/\sqrt{B}\,(400) = 0.0618$	$S/\sqrt{B}\,(100) = 0.317$ [bin (2, 2)] $S/\sqrt{B}\,(150) = 0.501$ [bin (2, 2)] $S/\sqrt{B}\,(400) = 0.093$ [bin (2, 2)]

Figure 24: [$2\ell 1\tau$-SS 2-D histogram] L_{T} v/s $\Delta R\,(\ell_1, \ell_2)$ (all L_{T} axes are in GeV units). The bins (x,y) with the best significance for each mass point have been mentioned in parentheses.

	Selection	$S/\sqrt{B}$ (100)	$S/\sqrt{B}$ (150)	$S/\sqrt{B}$ (400)
$2\ell1\tau$-OS	$p_{\mathrm{T}}^{miss} > 50$ + $\Delta R\ (\ell_1, \ell_2) < 1.5$	0.4811	0.4675	0.0176
	$p_{\mathrm{T}}^{miss} > 50$ + $p_{\mathrm{T}}(\tau) > 80$ + $\Delta\phi\ (\ell_1, \ell_2) < 0.75$	**0.4426**	**0.6855**	**0.0452**
	$p_{\mathrm{T}}^{miss} > 50$ [2-D histogram] $p_{\mathrm{T}}(\tau)$ v/s $\Delta R\ (\ell_1, \ell_2)$	0.416	0.425	0.117
$2\ell1\tau$-SS	Events with an OS pair	**0.3657**	**0.6417**	**0.0693**
	$\Delta\phi\ (\ell_1, \ell_2) > 2.3$	0.3441	0.5127	0.0468
	[2-D histogram] L_{T} v/s $\Delta R\ (\ell_1, \ell_2)$	0.317	0.501	0.093

Table 7: Summary table for the $2\ell1\tau$ channel.

5.4 3ℓ

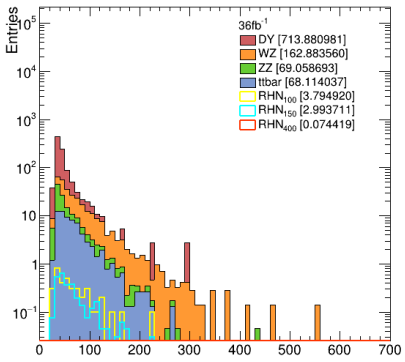
The 3ℓ channel is similar to the $2\ell1\tau$ channel in terms of event topology (we have a light lepton fake instead of a τ fake). 3ℓ has fewer events than the $2\ell1\tau$ channel (after applying vetoes). In addition to that, there is a lot of background (majorly DY and WZ). So, for this part of the analysis, we start by dividing the channel into 7 categories based on the sign (S) and flavour (F) of each lepton and where it lies with respect to the on-Z range:

1. OSSF Below-Z
2. OSSF Above-Z
3. SSSF Below-Z
4. SSSF Above-Z
5. OSOF Below-Z
6. OSOF Above-Z
7. All same-sign leptons

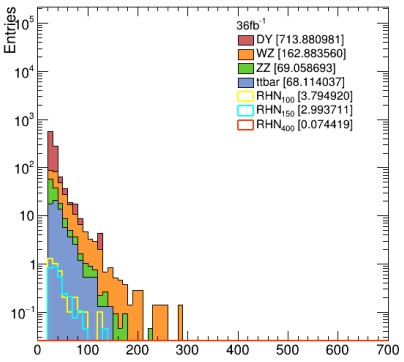
Here, the below-Z and above-Z is defined as < 76 GeV and > 106 GeV respectively. After analyzing a couple of basic plots from each of these categories, we discard the ones which have very less signal (like SSSF and all same-sign categories). Finally, we are left with two categories: *(1)* OSSF Below-Z and *(2)* OSOF Below-Z.

5.4.1 OSSF Below-Z

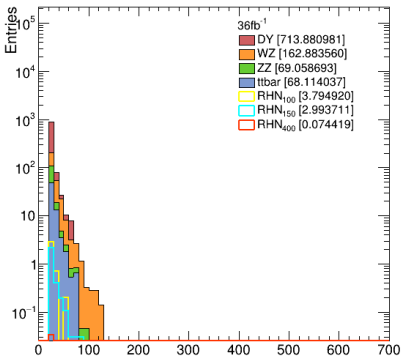
All the events that have at least one opposite-sign same-flavour (OSSF) pair belong to this category. In a lot of cases, we have an ambiguity since two such pairs can be formed (this is also true when we apply vetoes). To resolve this ambiguity, we pick the pair that is closer to the Z mass, *i.e.* we check $|M_{\ell^+\ell^-} - 91|$ and choose the OSSF pair for which this quantity is minimum. The third lepton is named the odd lepton. Some basic distributions are shown in Figure 25 and a summary of the best set of cuts is given in Table 8.



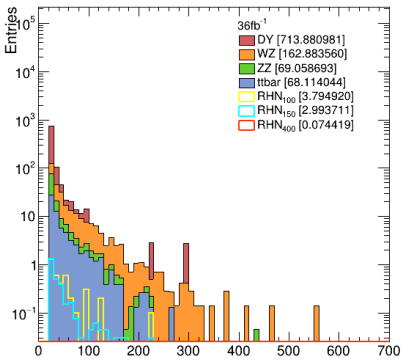
p_{T} of leading light lepton ℓ_1[GeV]



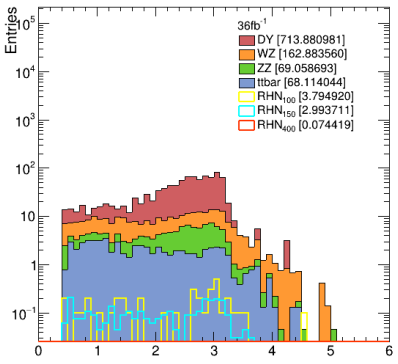
p_{T} of sub-leading light lepton ℓ_2[GeV]



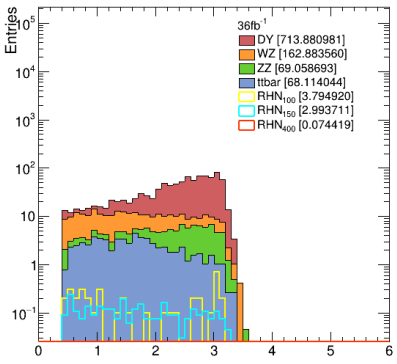
p_{T} of sub-sub-leading lepton ℓ_3[GeV]



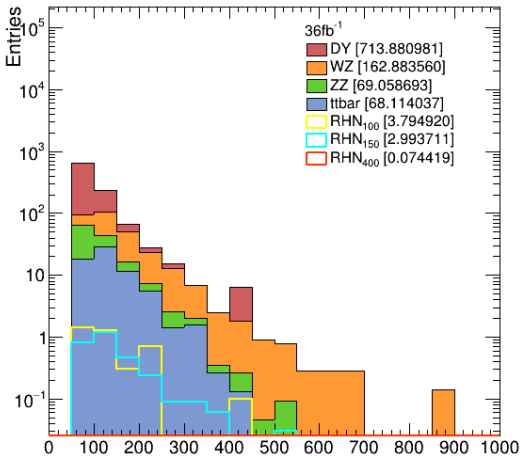
p_{T} of odd lepton ℓ_{odd}[GeV]



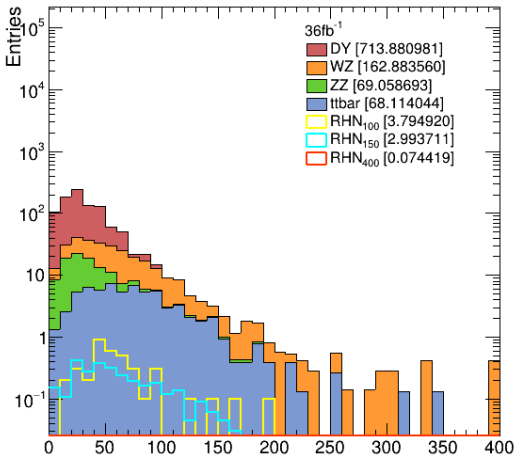
ΔR (ℓ_1,ℓ_2)



ΔR (OSSF pair)

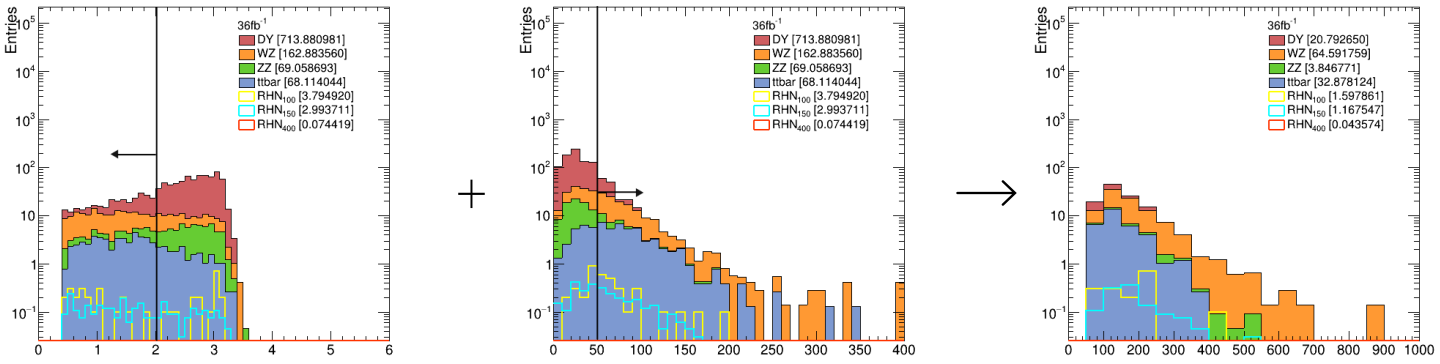


L_{T} [GeV]



p_{T}^{miss} [GeV]

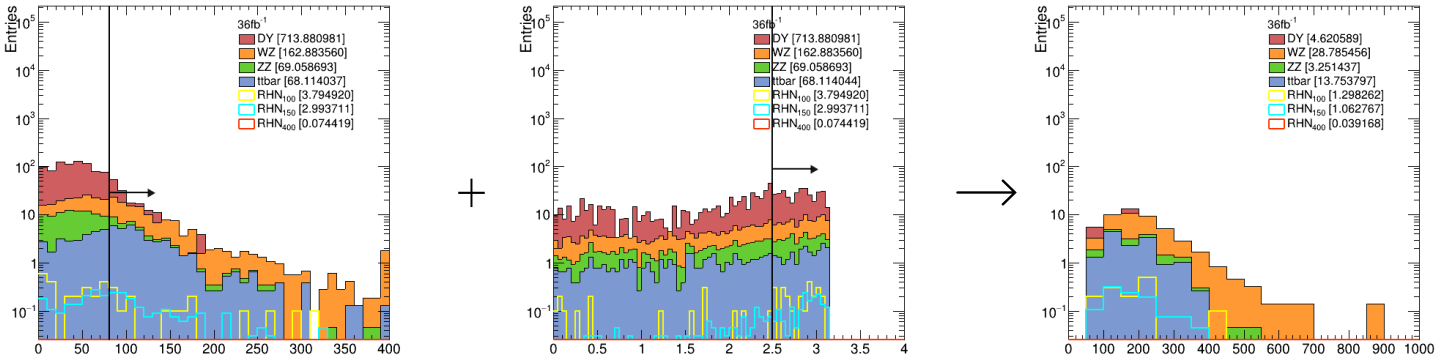
Figure 25: Some basic distributions for the 3ℓ-OSSF below-Z channel. From the p_{T} plots, we observe that the odd lepton is often the leading lepton, *i.e.* the OSSF pair is often of the lower p_{T} leptons.



ΔR (OSSF pair) p_{T}^{miss} [GeV] L_{T} [GeV]

Significance before cuts	Significance after cuts
$S/\sqrt{B}\,(100) = 0.1192$ $S/\sqrt{B}\,(150) = 0.0940$ $S/\sqrt{B}\,(400) = 0.0023$	$S/\sqrt{B}\,(100) = 0.1446$ $S/\sqrt{B}\,(150) = 0.1056$ $S/\sqrt{B}\,(400) = 0.0039$

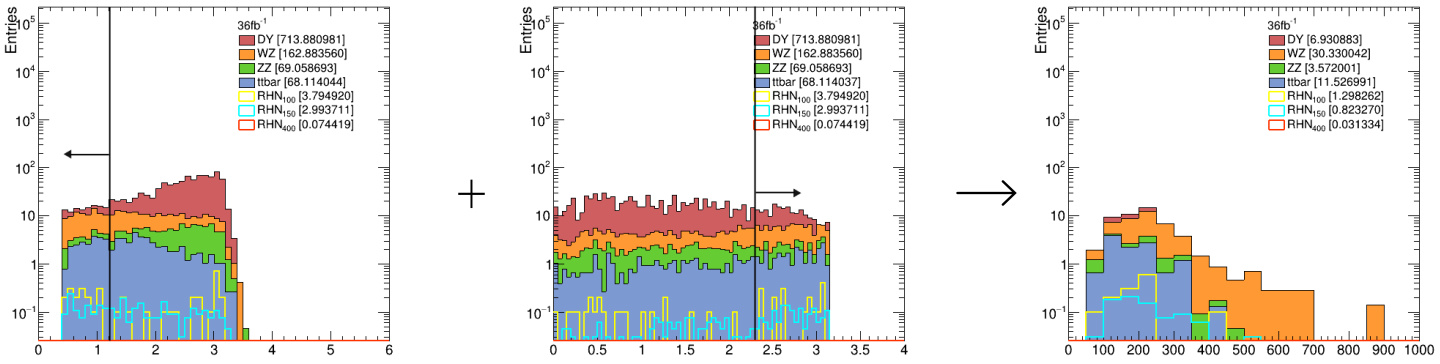
Figure 26: ΔR (OSSF pair) < 2 and $p_{\mathrm{T}}^{miss} > 50$ for 3ℓ-OSSF below-Z events.



$m_T\,(\ell_1, E_{\mathrm{T}}^{\mathrm{miss}})$ [GeV] $\Delta\phi\,(\ell_{odd}, \ell_{OSSF_1})$ L_{T} [GeV]

Significance before cuts	Significance after cuts
$S/\sqrt{B}\,(100) = 0.1192$ $S/\sqrt{B}\,(150) = 0.0940$ $S/\sqrt{B}\,(400) = 0.0023$	$S/\sqrt{B}\,(100) = 0.1828$ $S/\sqrt{B}\,(150) = 0.1497$ $S/\sqrt{B}\,(400) = 0.0055$

Figure 27: $m_T\,(\ell_1, E_{\mathrm{T}}^{\mathrm{miss}}) > 80$ and $\Delta\phi\,(\ell_{odd}, \ell_{OSSF_1}) > 2.5$ for 3ℓ-OSSF below-Z events.



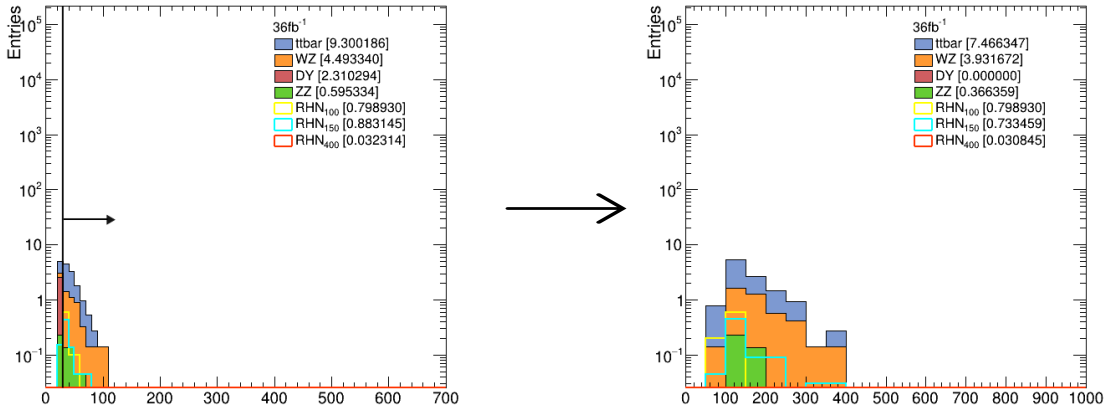
ΔR (OSSF pair) $\Delta\phi\ (\ell_{odd}, \ell_{OSSF_2})$ L_{T} [GeV]

Significance before cuts	Significance after cuts
$S/\sqrt{B}\ (100) = 0.1192$ $S/\sqrt{B}\ (150) = 0.0940$ $S/\sqrt{B}\ (400) = 0.0023$	$S/\sqrt{B}\ (100) = 0.1794$ $S/\sqrt{B}\ (150) = 0.1138$ $S/\sqrt{B}\ (400) = 0.0043$

Figure 28: ΔR (OSSF pair) < 1.2 and $\Delta\phi\ (\ell_{odd}, \ell_{OSSF_2}) > 2.3$ for 3ℓ-OSSF below-Z events.

5.4.2 OSOF Below-Z

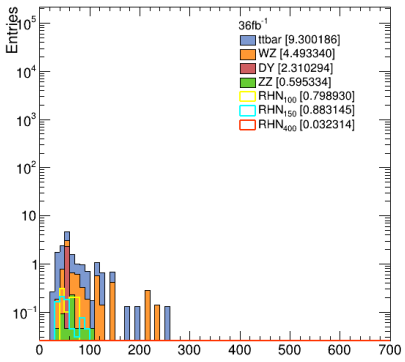
The events that do not have an OSSF pair will have an OSOF pair in the final state (excluding all same-sing events). Again, we pick the pair closer to the Z mass and the third lepton is called the odd lepton. We try a few selections on the distributions [Figure 25] but they cut down the signal even further. A very small number of events (< 1) will fail to give us meaningful results. One of the selections has been shown in Figure 29.



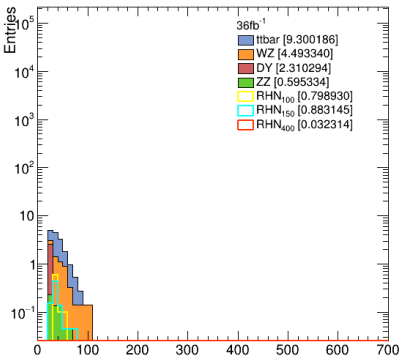
$p_{\mathrm{T}}\ (\ell_2)$ [GeV] L_{T} [GeV]

Significance before cut	Significance after cut
$S/\sqrt{B}\ (100) = 0.1955$ $S/\sqrt{B}\ (150) = 0.2161$ $S/\sqrt{B}\ (400) = 0.0079$	$S/\sqrt{B}\ (100) = 0.2329$ $S/\sqrt{B}\ (150) = 0.2138$ $S/\sqrt{B}\ (400) = 0.0089$

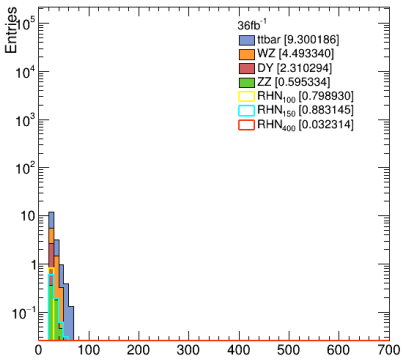
Figure 29: $p_{\mathrm{T}}\ (\ell_2) > 30$ for 3ℓ-OSOF below-Z events.



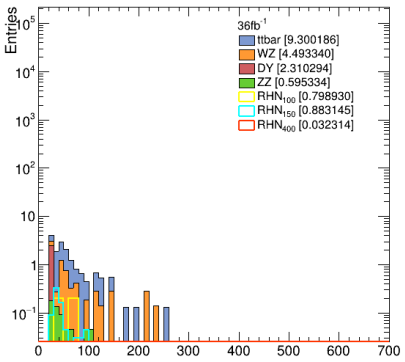
p_{T} of leading light lepton ℓ_1[GeV]



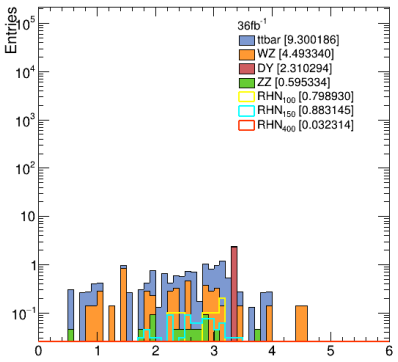
p_{T} of sub-leading light lepton ℓ_2[GeV]



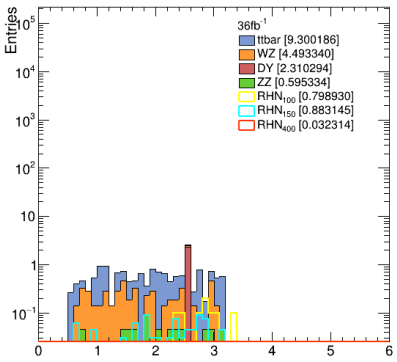
p_{T} of sub-sub-leading lepton ℓ_3[GeV]



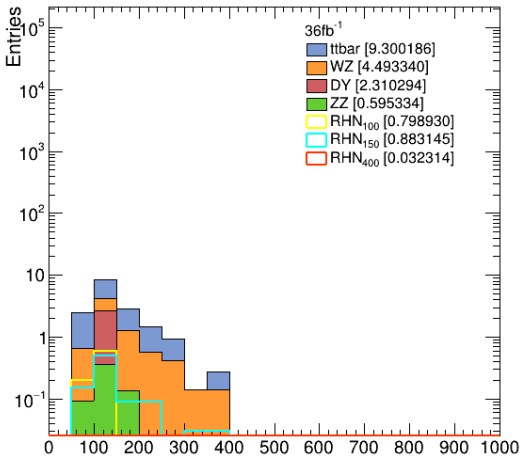
p_{T} of odd lepton ℓ_{odd}[GeV]



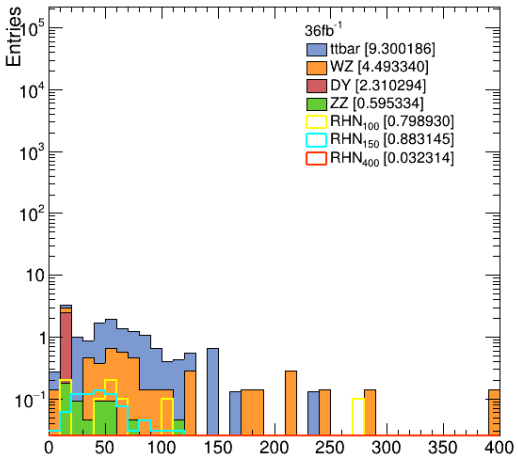
ΔR (ℓ_1,ℓ_2)



ΔR (OSOF pair)



L_{T} [GeV]



p_{T}^{miss} [GeV]

Figure 30: Some basic distributions for the 3ℓ-OSOF below-Z channel.

Channel	Selection	$S/\sqrt{B}$ (100)	$S/\sqrt{B}$ (150)	$S/\sqrt{B}$ (400)
3ℓ-OSSF	$m_{\ell\ell}$ (OSSF pair) < 76 + ΔR (OSSF pair) < 2 + $p_{\mathrm{T}}^{miss} > 50$	0.1446	0.1056	0.0039
	$m_{\ell\ell}$ (OSSF pair) < 76 + m_T (ℓ_1, $E_{\mathrm{T}}^{\mathrm{miss}}$) > 80 + $\Delta\phi$ (ℓ_{odd}, ℓ_{OSSF_1}) > 2.5	**0.1828**	**0.1497**	**0.0055**
	$m_{\ell\ell}$ (OSSF pair) < 76 + ΔR (OSSF pair) < 1.2 + $\Delta\phi$ (ℓ_{odd}, ℓ_{OSSF_2}) > 2.3	0.1794	0.1138	0.0043
3ℓ-OSOF	$m_{\ell\ell}$ (OSSF pair) < 76 + p_{T} (ℓ_2) > 30	0.2329	0.2138	0.0089

Table 8: Summary table for the 3ℓ channel.

6 RHN Acceptance Cut-flow

The RHN acceptance cutflow gives us meaningful insights into how different cuts and selections affect the number of events in each channel. This can help us decide which selection to apply for finding the optimal signal region. We can also observe the difference in various object IDs (MVA IS and Deep ID for τ). Table 9 summarises the cutflow and the numbers at each stage of selection. The main channels used for analysis have been highlighted.

	MVA						Deep					
Selection	150 GeV	100 GeV	DY	WZ	$t\bar{t}$	W + jets	150 GeV	100 GeV	DY	WZ	$t\bar{t}$	W + jets
Total events ran	285816	298880	23237	52560	24235935	57358383	285816	298880	23237	52560	24235935	57358383
Total events	286200	299200	23237	52560	24265024	57402435	286200	299200	23237	52560	24265024	57402435
4ℓ events	2	4	2	9	855	0	2	4	2	9	855	0
$N_{1\ell}$ events	166649	140098	23237	52560	18340314	18917151	166649	140098	23237	52560	18340314	18917151
$N_{1\ell\_trigg}$ events	99912	72436	21406	47589	13065106	10297682	99912	72436	21406	47589	13065106	10297682
$N_{1\ell\_trigg\_1\ell}$ events	22994	10915	21033	38399	2835198	1538	22994	10915	21033	38399	2835198	1538
$N_{1\ell\_trigg\_1\ell p_{\mathrm{T}}10}$ events	21495	9111	20680	36568	2726718	396	21495	9111	20680	36568	2726718	396
$N_{1\ell\_trigg\_1\ell p_{\mathrm{T}}10\_1\ell}$ events	2681	857	54	825	16869	2	2681	857	54	825	16869	2
$N_{1\ell\_trigg\_2\ell p_{\mathrm{T}}10}$ events (3ℓ events)	2251	619	37	746	10002	1	2251	619	37	746	10002	1
$N_{1\ell\_trigg\_1\ell p_{\mathrm{T}}10\_1\tau}$ events	4524	1022	16333	29722	1819	0	3389	663	11742	29813	1294	0
$N_{1\ell\_trigg\_1\ell p_{\mathrm{T}}10\_1\tau p_{\mathrm{T}}20}$ events ($2\ell1\tau$ events)	4280	908	16329	29706	1696	0	3389	663	11742	29813	1294	0
$N_{1\ell\_trigg\_1\tau}$ events	23319	10796	189	7478	876646	7	17542	7315	284	7562	590115	6
$N_{1\ell\_trigg\_1\tau p_{\mathrm{T}}20}$ events	22069	9684	189	7476	805080	6	17452	7315	284	7562	590115	6
$N_{1\ell\_trigg\_1\tau p_{\mathrm{T}}20\_1\tau}$ events	3049	506	1	136	23	0	1681	190	0	25	7	0
$N_{1\ell\_trigg\_2\tau p_{\mathrm{T}}20}$ events ($1\ell2\tau$ events)	2670	363	0	104	21	0	1681	190	0	25	7	0
$N_{1\tau}$ events	37486	22362	0	0	1182328	1301502	27692	14793	0	0	803086	763144
$N_{1\tau p_{\mathrm{T}}20}$ events	35331	20113	0	0	1090836	1124963	27692	14793	0	0	803086	763144
$N_{1\tau p_{\mathrm{T}}20\_1\tau}$ events	6960	2038	0	0	71619	0	3724	916	0	0	33355	0
$N_{2\tau p_{\mathrm{T}}20}$ events	6118	1637	0	0	60823	0	3724	916	0	0	33355	0
$N_{2\tau p_{\mathrm{T}}20\_1\tau}$ events	520	79	0	0	0	0	231	28	0	0	0	0
$N_{3\tau p_{\mathrm{T}}20}$ events (3τ events)	434	54	0	0	0	0	231	28	0	0	0	0

Table 9: RHN acceptance cutflow

After comparing the numbers for the old and new τ IDs, we see that the old (MVA) ID seems to result in higher number of events as compared to the new (Deep) ID. This might be because the new ID has a lower false positive rate, and that it has fewer fakes. With a deeper analysis and truth matching, we can draw more definitive conclusions. We use the Deep ID for our analysis.

7 Results

The search for right-handed neutrinos (RHN) involves setting limits on the signal cross section. To select the signal regions for this analysis, we apply a number of selections to get the highest possible signal significance. Possible signal regions (regions with a the best signal significance so far) have been listed in Table 10. Multiple regions can be combined for better results.

Channel	Region	$S/\sqrt{B}$ (100 GeV)	$S/\sqrt{B}$ (150 GeV)	$S/\sqrt{B}$ (400 GeV)
$2\ell 1\tau$-OS	$p_{\mathrm{T}}^{miss} > 50$ + p_{T} (τ) > 80 + $\Delta\phi$ (ℓ_1,ℓ_2) < 0.75	0.4426	0.6855	0.0452
$2\ell 1\tau$-SS	Events with an OS pair	0.3657	0.6417	0.0693
3ℓ-OSSF	$m_{\ell\ell}$ (OSSF pair) < 76 + m_T (ℓ_1, $E_{\mathrm{T}}^{\mathrm{miss}}$) > 80 + $\Delta\phi$ (ℓ_{odd}, ℓ_{OSSF_1}) > 2.5	0.1828	0.1497	0.0055
3ℓ-OSOF	$m_{\ell\ell}$ (OSSF pair) < 76 + p_{T} (ℓ_2) > 30	0.2329	0.2138	0.0089

Table 10: A summary of all the channels considered for signal region optimization.

8 Acknowledgements

I would like to thank Dr. Sourabh Dube for giving me this opportuninty to delve into the subject and for greatly improving my understanding of the same, and Angira Rastogi for her guidance and help with the project. I thank them both for all the crucial insights and discussions during the course of the project.

List of Tables

List of Figures