Study of $t\bar{t}Z$ with invisible Z decay in Dark Matter searches

Alberto Ressa, University of Rome-La Sapienza, Italy

Supervisors: Marco Rimoldi, Federico Meloni

September 4, 2019

Abstract

In this report I will present the characterization of the $t\bar{t}$ plus Z background in a search for Dark Matter in association with a pair of top quarks using the ATLAS detector. Thanks to the high statistic of the ATLAS Run 2 dataset, I developed a data-driven method to validate the $t\bar{t}Z \to ll\nu\nu$ normalisation using a four lepton selection.

Contents

1 Introduction

The understanding of the nature of Dark Matter (DM) is one of the priorities in modern physics. What we know about it comes from astrophysical evidences. We know that a kind of matter must exist which interact only by gravity and eventually some kind of weak force. At the Large Hadron Collider (LHC), it is possible to search for the production of Dark Matter particles out of the proton proton collisions. During this project the data collected from 2015 to 2018 at the LHC Run 2 has been used, corresponding to a total luminosity of 139 fb^{-1} at the centre of mass energy of 13 TeV. Figure 1 shows two types of collisions at two very different physical scale, but both involved in the Dark Matter searches. On the left, the pp collision recorded by the ATLAS experiment and, on the right, a collision between two clusters of galaxies. In the latter figure in blue it is shown the halo of Dark Matter estimated from the gravitational effect during the collision, in purple the ordinary matter distribution.

Figure 1: Left, pp collisions in ATLAS. Right, the Bullet Cluster phenomenon; it is a collision between two cluster of galaxies. In blue is shown the DM halo distribution, in purple the ordinary matter distribution.

1.1 A Dark Matter Model

In the Standard Model (SM), there are no particles which could explain DM. Neutrinos are the only ones which interact only by weak force and gravity, but their masses (still not well known) are too small to explain the gravitational phenomena observed in astrophysics.

Many different models (including SuperSymmetry or Extra Dimensions) try to explain the nature of DM. The general class of new particles which could match the DM properties is called WIMPs, Weakly Interacting Massive Particles; this acronym describes the main characteristics these particles must have.

The models taken into account for this project, are characterized by the presence of a neutral and scalar or pseudo-scalar (spin 0) mediator which would interact with both

SM and DM particles. This mediator would have a Yukawa-like coupling with these particles, such that the heaviest particles would have the highest cross section for this process. That's why this search is looking for DM particle produced in association with top quarks, which are the heaviest elementary particles in the SM.
In Figure 2 is shown the Feynmann diagram for the phenomenon of interest in this project. From SM we know that top quark decay mainly in a bottom quark and a W

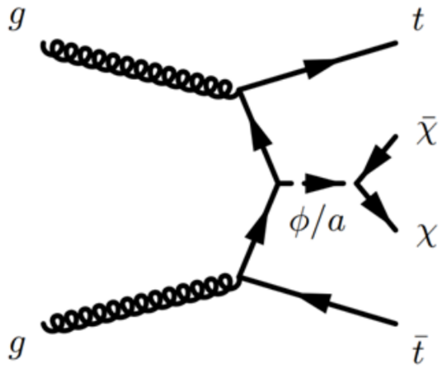


Figure 2: Feynmann diagram for the DM process involved in this search.

boson. The project is focused on the di-leptonic decay of the top quarks. It means that are selected events with 2 leptons in the final state. These are coming from the decay $W \to l\nu$ or from other processes that will be discussed in the next chapters. The number of jets and b-tagged jets (which are jets candidate from b-hadrons decay) are also taken into account. The di-leptonic top decay mode has been used since leptons, rather than hadronic objects, leave a clear experimental signature in the detector.

1.2 Missing transverse momentum

The Missing transverse momentum is main variable involved in DM processes. DM particles, eventually produced at LHC, will cross the detector without leaving any signal since they don't interact by electromagnetic of strong force. Thus, events of this kind are characterised by a large momentum unbalance. Since the interactions happen at parton level, and we are not able to know the fraction of momentum carried by each parton inside the proton, we can use the momentum conservation only in the transverse plane. So the vector sum of all the momenta projected in the transverse plane in a single event must give zero. Thus we can extract the momentum carried by the invisible particles using the following formula:

$$E_T^{miss} = -\left(\sum_{i\in\mu} p_T^i + \sum_{i\in e} p_T^i + \sum_{i\in\gamma} p_T^i + \sum_{i\in had.\ \tau} p_T^i + \sum_{i\in jets} p_T^i + \sum_{i\in Soft\ Term} p_T^i \right) \quad (1)$$

The sum runs over all the energy deposits of identified physical objects (e, μ ...) or non-identified contributions (Soft Term). In Figure 3 is shown the ATLAS detector

projected in the transverse plane where it is possible to observe an event with momentum unbalance. The red arrow shows the vector for the missing transverse momentum.
Of course not only DM or new physics events contribute to missing energy events; a lot of SM processes give a similar signature. In particular neutrinos do not interact with the detector, leaving a missing energy signature. Also leptons and jets could be misidentified generating a momentum unbalance. In this project this kind of events are estimated from a Monte Carlo simulation, while it is expected that an estimation from data would give a better result.

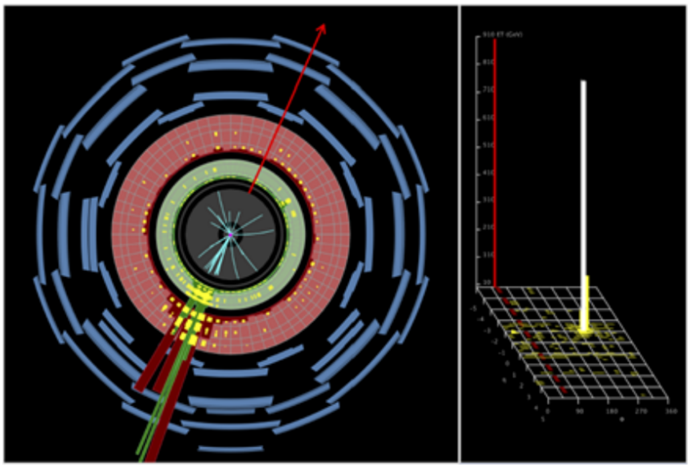


Figure 3: An example of missing transverse energy event in ATLAS detector.

1.3 Signal/Background study and kinematic Variables

It is fundamental for this kind of searches to understand the different Standard Model background processes which have the same final state as the signal one. To study them, it is necessary to apply some cuts based on some kinematics variables in order to define some Regions where a specific background is enhanced. These regions are called Control Regions (CR). Similarly, it is possible to define Signal Regions (SR) by applying different cuts in order to enhance the expected number of signal events with respect to the expected SM number of events. It is important for the SR to have the lowest possible background contamination, but it is also important to have a low signal contamination in the CR in order to have a better background estimation. That's why the CR are defined to have kinematic properties which make them orthogonal to the SR.
The aim of each CR is to compare the yields of data with the MC predictions.
Assuming the following relation:

$$\frac{N^{SR}_{data}}{N^{SR}_{MC}} = \frac{N^{CR_i}_{data}}{N^{CR_i}_{MC}} \tag{2}$$

it is possible to compute a scale factor from the ratio between data and MC in each control region and multiply it for the MC predictions in the SR to have a better estimation

of the background. In particular the scale factor is computed by subtracting from data the non-dominant backgrounds predicted by MC simulation and dividing it by the MC prediction of the dominant background.
Here I will briefly describe some of the kinematic variables used for the definition of the CR and SR selections:

- m_{T2}: as already explained, it is possible to use the momentum conservation only in the transverse plane. The transverse mass of some final state objects can be useful to reconstruct the mass of decaying particles also when invisible decay products are present. But when more than one particle escapes detection it is not possible to define a transverse mass. We know only the total momentum of the invisible particles from the missing transverse momentum. This variable use this information as a constraint to set a lower bound for invisible particles transverse mass ([2]).

- $E_{T,corr}^{miss}$: the main aim of this project is to study the background coming from the production of a Z boson decaying in neutrinos in association with two top quarks. It is possible to use events in which the Z boson decays in 2 leptons for this purpose. These events can emulate the invisible Z decay background identifying the two leptons coming from the Z boson and treating them as invisible particles. This means "correcting" the E_T^{miss} excluding these two leptons from its computation in the events containing 3 or more leptons.

- $E_{T,sig}^{miss}$: this variable comes out from an object-based computation. The value of the E_T^{miss} is compared with the p-value of the distribution of no real missing energy hypothesis for that object. An high value of the missing energy significance means that it is reasonable to associate it with the momentum carried by an invisible particle ([4]).

- $\Delta\Phi_{boost}$: it is the azimuth angular distance between the missing momentum and the vector sum of it and the leptons momentum.

Some preliminary cuts has been applied in each region; for example the transverse momentum of the leading lepton is required to be greater than 25 GeV and greater than 20 GeV for the other leptons. In the next chapters I will describe each regions defined in this project, and I will present the results obtained.

2 Signal Region

As already briefly described, the aim of this region is to enhance a possible signal. Two leptons in the final state are required. In order to reduce the SM background coming from the leptonic decay of the Z boson, same flavor (SF) leptons events are vetoed if their invariant mass is closer than 20 GeV from the Z mass (91.2 GeV). Different flavor (DF) events are selected with invariant mass greater than 20 GeV. Then it is required an $m_{T2} > 110 GeV$ in order to deal with massive invisible particles. In table 1 are summarized all the cuts applied in the SR.

	SR
Number of Leptons	2
$p_T^{lep1}[GeV]$	> 25
$p_T^{lep2}[GeV]$	>20
$m_{ll}[GeV]$	[20, 71.2] >111.2 if SF
	>20 if DF
n_{bjet}	≥ 1
$\Delta\Phi_{boost}$	<1.5
$E_{T,sig}^{miss}$	>12
m_{T2}	> 110 GeV

Table 1: Selection cuts for the Signal Region.

In Figure 5 (Appendix), the distributions of the main variables in the SR are shown. Only the MC simulations are shown.

3 Control Regions

In the introduction the important role played in this kind of analysis by the CR has been explained. Different types of SM backgrounds are identified with the particles in their final state. In particular during my project I studied CR for the two main backgrounds, $t\bar{t}$ (CRttbar) and $t\bar{t}Z$ (CRttZ). As shown in the next sections, the contamination of the $t\bar{t}$ background in the CR built to enhance $t\bar{t}Z$ is equal to 0. Thus first the scale factor for the $CRttZ$ has been computed and then it has been applied on $t\bar{t}Z$ background in CRttbar to optimise the estimation of the scale factor for the $t\bar{t}$.

3.1 Control Region for $t\bar{t}$

The aim of this region is to look at the process which give a pair of top quarks in the final state. It is an important background since the top pair could decay leptonically producing a pair of leptons as required for the SR. The missing energy in this process can arise from neutrinos produced from $W \to l\nu$, or from misidentified jets. To emulate the background process in SR, 2 leptons in the final state are required. Only different flavors leptons are selected to avoid the contamination from the $t\bar{t}Z$ background. Cuts on other variables as m_{T2}, $E_{T,sig}^{miss}$ and $\Delta\Phi_{boost}$ has been applied in order to make CRttbar orthogonal to the SR from the kinematic point of view. Event selection for the $CRttbar$ is summarized in Table 2. In Figure 11 (Appendix) the distributions for data and MC of

	$CRttbar$
Number of Leptons	2
Leptons Flavor	DF
$p_T^{lep1}[GeV]$	> 25
$p_T^{lep2}[GeV]$	>20
$m_{ll}[GeV]$	>20
n_{bjet}	≥ 1
$\Delta\Phi_{boost}$	>1.5
$E_{T,sig}^{miss}$	>8
$m_{T2}[GeV]$	[100, 120]

Table 2: Selections for the CRttbar.

some variables in the CRttbar are shown. In Table 3, the yields in CRttbar are reported. The number of events observed in the data are shown too. The total background MC prediction is shown before and after applying the scale factor. The agreement between data and expected number of SM events from MC is good.

3.2 Control Region for $t\bar{t}Z$

This region has been built to study the background coming from processes with $t\bar{t}Z$ production in the final state. This SM background is the main object of this project.

	CRttbar
Observed	245
ttbar	205.94
Z+jets	0
ttZ	0.70
ttW	0.75
Wt	32.61
VV	0.70
VVV	0.03
other	0.99
MCFakes	3.29
Nominal MC ttbar	218.15
Scale Factor	0.94
Total Background unscaled	257.19
Total Background fitted	245.00

Table 3: Yields and scale factor for the CRttbar

It give rise to the most important irreducible background in this search. Irreducible background means a process which have the same final state signature of the signal process. It is the case since the produced Z boson can decay in a pair of neutrinos (Figure 4), which leave no signal in the detector.

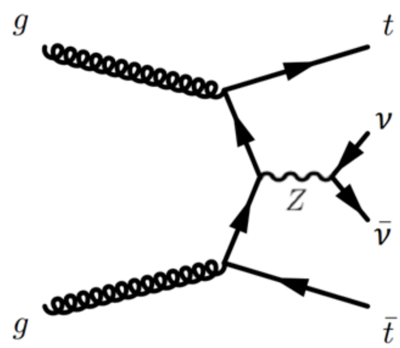


Figure 4: Feynmann diagram for the irreducible background in which the Z boson produced out of the top quarks decays in neutrinos.

The number of leptons required is 3, such that 2 of these must come from the Z boson decay. Also the invariant mass of the pair of same flavor and opposite sign (OS) leptons is required to fall in a range around $\sim 91\ GeV$ in order to enhance events containing the Z boson. The top quark decays ($t \rightarrow Wb$) provide a bottom quark which eventually develop a b-tagged jet. Thus, if only 1 b-tagged jet appear in the final state we must take it into account for the counts of the total number of jets. Then only the events having missing energy corrected higher than 140 GeV are selected. This requirement is important to emulate better the invisible Z decay with high missing energy. In this way the background source studied in this region will be more similar to the one in the SR. Selections are summarized in Table 4. In Figure 12 (Appendix) the distributions for the

	$CRttZ$
Number of Leptons	3
$p_T^{lep1}[GeV]$	> 25
$p_T^{lep2}[GeV]$	>20
$p_T^{lep3}[GeV]$	>20
$m_{ll}[GeV]$	SFOS pairs [71.2, 111.2]
n_{bjet}	>1, $n_{jet} > 2$
	=1, $n_{jet} > 3$
$E_{T,corr}^{miss}[GeV]$	>140

Table 4: Selections for the CRttZ

main variables in CRttZ are shown. In Table 5 the yields for data and MC predictions in $CRttZ$ are presented.

	CRttZ
Observed	398
ttbar	0
Z+jets	0.11
ttZ	259.75
ttW	4.94
Wt	0
VV	62.03
VVV	0.39
other	56.68
MCFakes	14.10
Nominal MC ttZ	250.58
Scale Factor	1.04
Total Background unscaled	388.83
Total Background fitted	398

Table 5: Yields and scale factor for the CRttZ

4 Four Leptons signature Validation Region

The main aim of this project is the definition of a validation region for the $t\bar{t}Z$ background. A validation region (VR) is a selection of data and MC having kinematic properties similar to the SR, but in which a specific background process is enhanced. In this way it is possible to check whether the estimation of that background is good

enough to be applied on the SR.
It is possible to validate the $t\bar{t}Z$ background using a 4 leptons signature in the final state. As shown in Figure 5, the process of interest requires 2 leptons coming from the Z

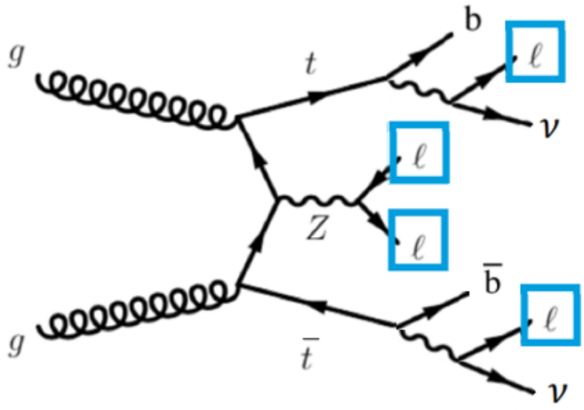


Figure 5: Feynmann diagram for the process of interest in the Validation Region. Highlighted in blue the 4 leptons required in the final state.

boson decay, and other 2 coming from top quark leptonic decays. The 2 leptons coming from the top quarks, in fact, match the selection of the SR. The idea is to treat the 2 leptons from the Z boson as invisible particles in order to emulate the Z decaying in neutrinos. This is the main background in the SR and so it is important to validate it checking the agreement between data and MC in a validation region.
A new variable called m_{T2}^{4lep} has been used in this analysis . It has the same definition of m_{T2} but it is computed using $\vec{E}_{T,corr}^{miss}$ instead of $\vec{E}_T^{miss}$. In this way, the 2 leptons from the Z decay are excluded from the missing momentum computation, resulting as invisible particles. Thus it makes sense to do a comparison between m_{T2}^{4lep} or $E_{T,corr}^{miss}$ in the Validation Region and m_{T2} or E_T^{miss} from the background predictions in the Signal Region, because we expect to find similar distributions if the irreducible background has been well estimated.

4.1 Pre-selection

A preselection study of the 4 leptons selection has been carried out. Events are selected with 4 leptons and 1 or more b-jet in the final state. In Figure 6 the distribution of the missing momentum corrected is shown. It is possible to observe which are the main processes involved in this study. As expected, the $t\bar{t}Z$ background process is the most important one. The production of 2 or 3 vector bosons (W or Z) can also give rise to 4 leptons in the final state as well as the process containing tWZ, which is included in "other". To achieve the same kinematic properties of the SR the idea is to apply similar cuts on the VR. In particular events in SR are selected to have $m_{T2} > 110\ GeV$. The equivalent variable m_{T2}^{4lep} can be used to apply a cut on this region.

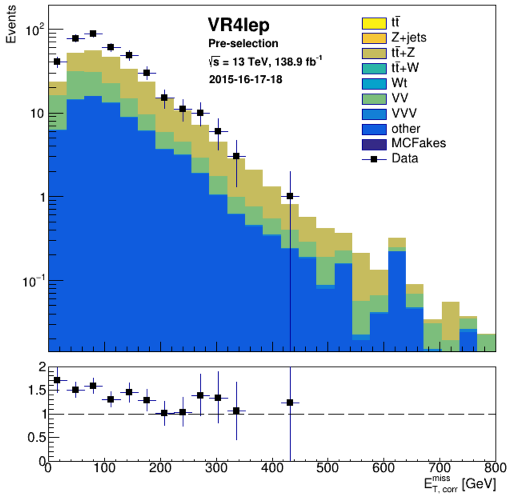


Figure 6: Distribution of the missing transverse momentum corrected in the Pre-selection Validation region with 4 leptons in the final state.

4.2 Post-selection

The VR has been built requiring $m_{T2}^{4lep} > 110\ GeV$. The cut is shown in Figure 7. The events selected are in the tail of the distribution; this is a reason why it is needed high luminosity in order to have enough statistics to study this process. It has been possible to observe 28 events which satisfy the selection.

After the selection on m_{T2}^{4lep}, a comparison between the distributions of some variables in SR and VR has been carried out. In Figure 8 the distributions of $E_{T,corr}^{miss}$ and E_T^{miss} respectively for the VR and the SR, before (on the left) and after (on the right) the selection are shown. Distributions have been normalised in the overlap area. The two distributions have a good overlap after the selection; it means that the range of interest of the kinematic variables is similar in SR and VR. It is possible to notice how the cut on m_{T2}^{4lep} has influenced also the $E_{T,corr}^{miss}$ distribution.

To have a validation of the background for the $Z \to \nu\bar{\nu}$ it is useful to plot the distribution of m_{T2}^{4lep} from data in the VR and compare it with the one of m_{T2} from the $t\bar{t}Z$ MC prediction in the SR. As shown in Figure 9, the two distributions overlap well, assuming a reasonable uncertainty.

Using m_{T2}^{4lep} and $E_{T,corr}^{miss}$, the aim is to emulate the $t\bar{t}Z$ background in the SR. So it is expected to find a good agreement once achieved kinematic properties similar to the SR ones.

In table 6, the yields for each of the background predicted from the MC simulations and data are shown. It has been used a Poissonian uncertainty on the yields for each background to compute an uncertainty for the total MC prediction. Thus it is possible to conclude that data are compatible with the MC prediction.

In Figure 13 (section 5) are shown the distributions of some variables in the VR.

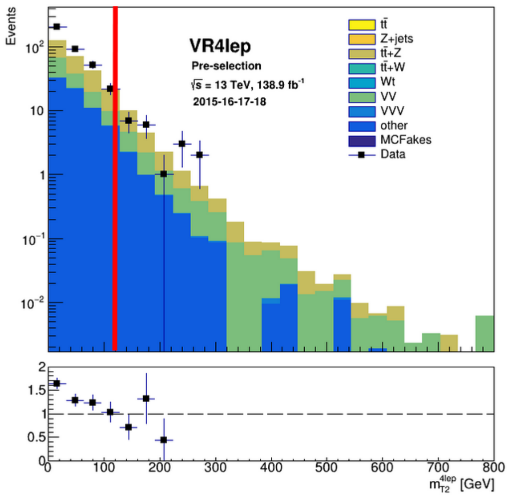


Figure 7: Distribution of m_{T2}^{4lep} in the Pre-selection validation region with 4 leptons in the final state. The red line indicates the lower threshold for the cut on this region.

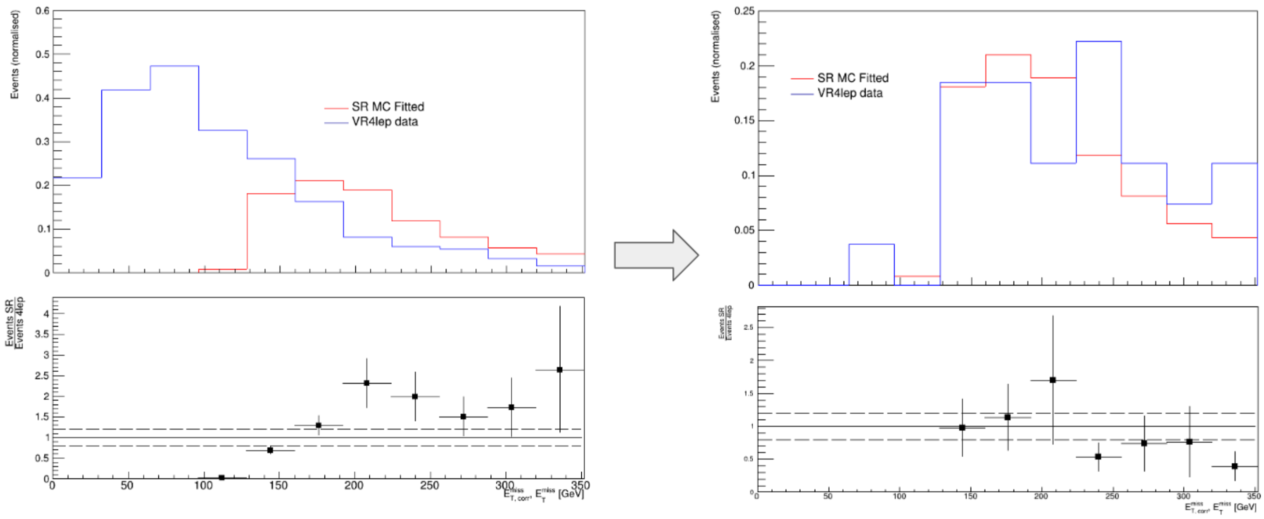


Figure 8: On the left, the distribution of E_T^{miss} from the MC prediction in the SR plotted in red, while in blue the distribution of $E_{T,corr}^{miss}$ from the data in the VR before the selection is plotted. On the right the same distributions but after the selection in the VR.

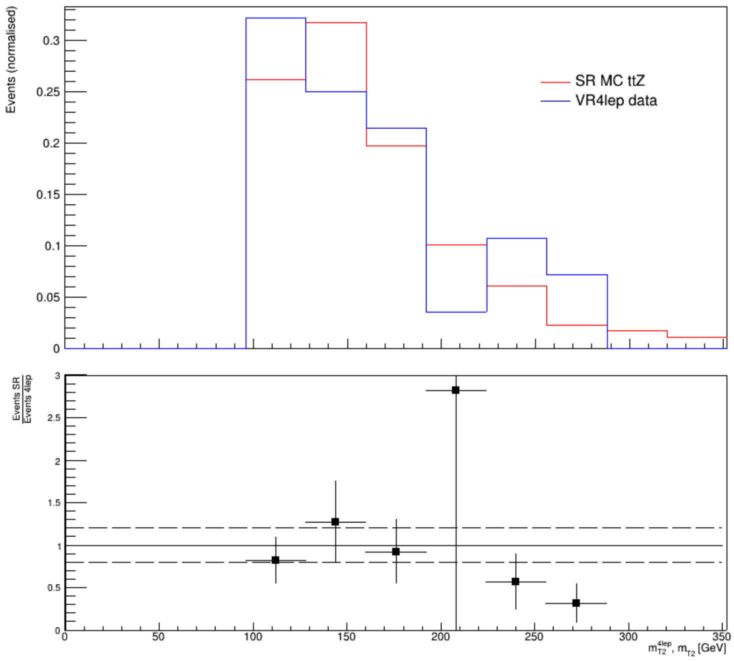


Figure 9: Comparison between the distributions of $t\bar{t}Z$ MC prediction in the SR and data in VR for the variable m_{T2} and m_{T2}^{4lep} respectively.

	VR4lep
Observed	28
ttZ	15.49
VV	7.32
VVV	0.16
other	7.05
Nominal MC ttZ	14.95
Total Background unscaled	29.47
Total Background fitted	30.02 ±6.77

Table 6: Yields and scale factor for the VR.

5 Conclusions

In Table 7 the selections applied for all the regions are summarized while in Table 8 the yields for the 2 CR analysed and the VR are shown. The first step of this project was to analyse the Control Regions for the two dominant background involved in the search for Dark Matter in association with a pair of top quarks. From both the CRttbar and CRttZ, a scale factor has been extracted to be applied on the MC prediction for the $t\bar{t}$ and $t\bar{t}Z$, in both the SR and the VR. This step was necessary to have the best estimation possible for these kind of background. The MC background predictions for the SR has been studied as well to be able to build the VR.
The main aim of the project was to study the irreducible background given by the invisible decay of the Z boson in a pair of neutrinos. To do this the idea was to build a VR to enhance the process characterized by 4 leptons in the final state. Treating the 2 leptons coming from the Z boson as invisible particles, it was possible to compare some distribution from the VR to other of the SR. The variables m_{T2}^{4lep} and $E_{T,corr}^{miss}$ have been used for this purpose. Once verified the overlap between the distributions of kinematic variables from SR and VR, the yields of data and MC in the VR can be compared. They result to be compatible within the uncertaity. Thus it was possible to validate the background of $t\bar{t}Z$ with Z invisible decay.
The next step for this search could be to study the SR including also data, which haven't been considered during this project, to look for some discrepancies with the MC predictions of the background. Eventually other relevant backgrounds could be validated in regions built for this purpose in order to optimise the estimation of SM process.

	SR	CRttbar	CRttZ	VR4lep
Number of Leptons	2	2	3	4
Leptons Flavour	SF/DF	DF		
$m_{ll} [GeV]$	[20, 71.2], >111.2 if SF >20 if DF	>20	SFOS pairs [71.2, 111.2]	
n_{bjet}	≥ 1	≥ 1	>1, $n_{jet} > 2$ =1, $n_{jet} > 3$	≥ 1
$\Delta\Phi_{boost}$	<1.5	>1.5		
$E_{T,sig}^{miss}$	>12	>8		
$m_{T2} [GeV]$	>110	[100, 120]		>110
$E_{T,corr}^{miss} [GeV]$			>140	

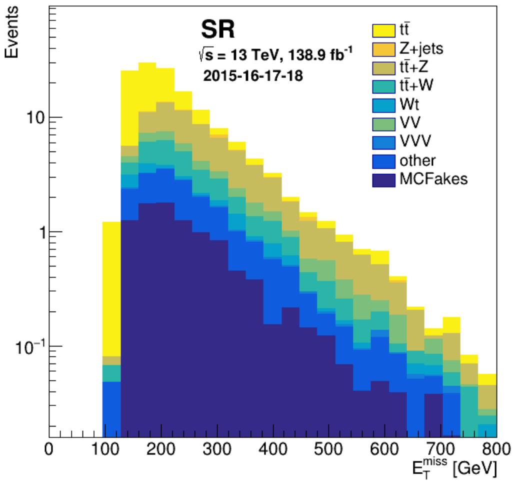
Table 7: Summary of the selections applied in the analysed regions.

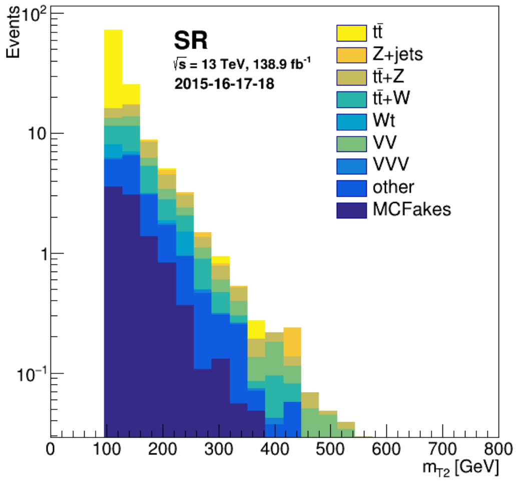
	CRttbar	CRttZ	VR
Observed	245	398	28
ttbar	205.9	0	0
Z+jets	0	0.11	0
ttZ	0.70	259.75	15.49
ttW	0.75	4.94	0
Wt	32.61	0	0
VV	0.70	62.03	7.32
VVV	0.03	0.39	0.16
other	0.99	56.68	7.05
MCFakes	3.29	14.10	0
Total Background unscaled	257.19	388.83	29.47
Total Background fitted	245	398	30.02 ± 6.77

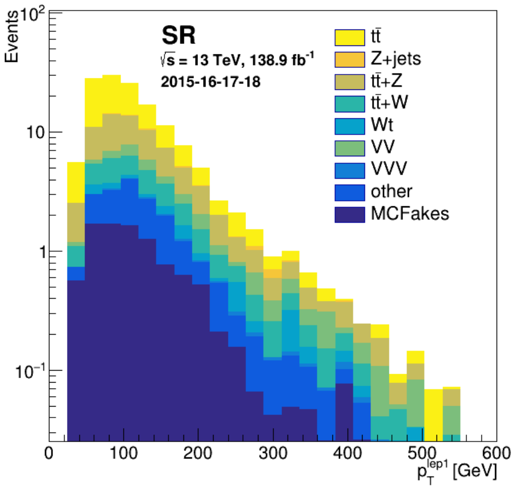
Table 8: Summary of yields.

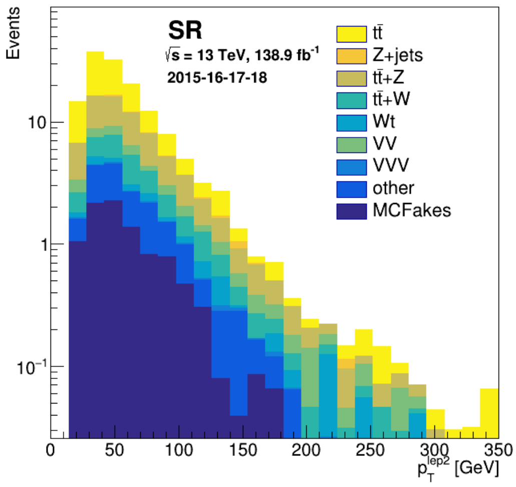
Appendix-Graphics

In this section the plots of different relevant variables will be shown for any region studied for this project.









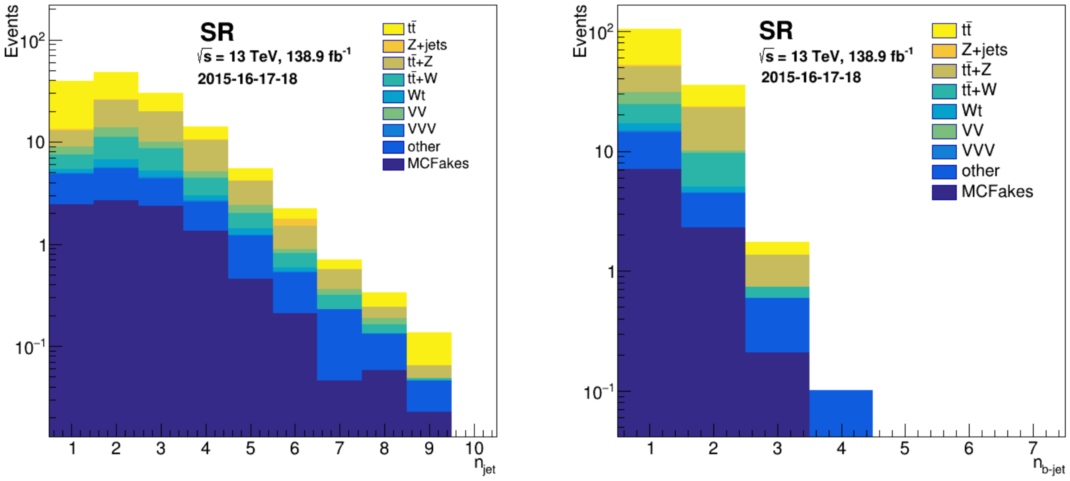
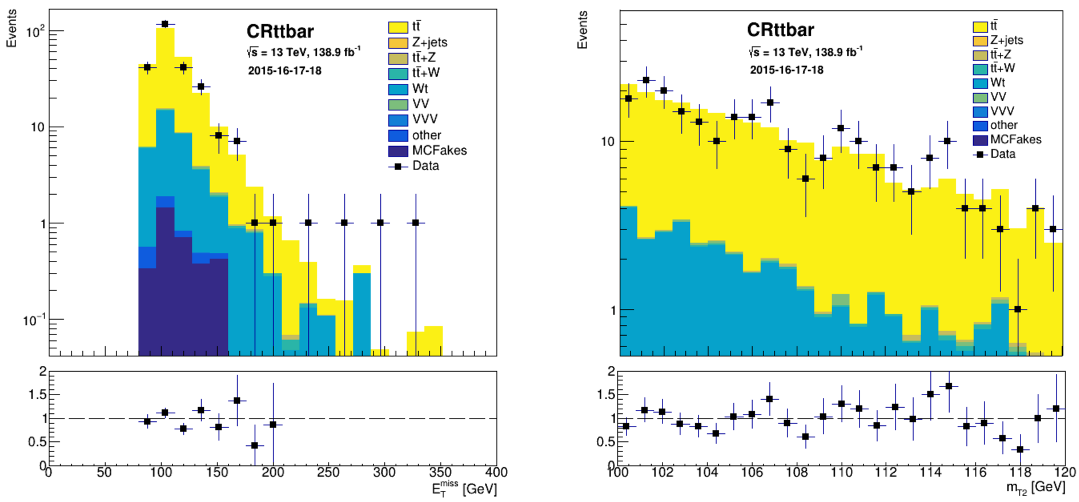


Figure 10: Distributions of the MC predictions of some relevant variables in the SR.



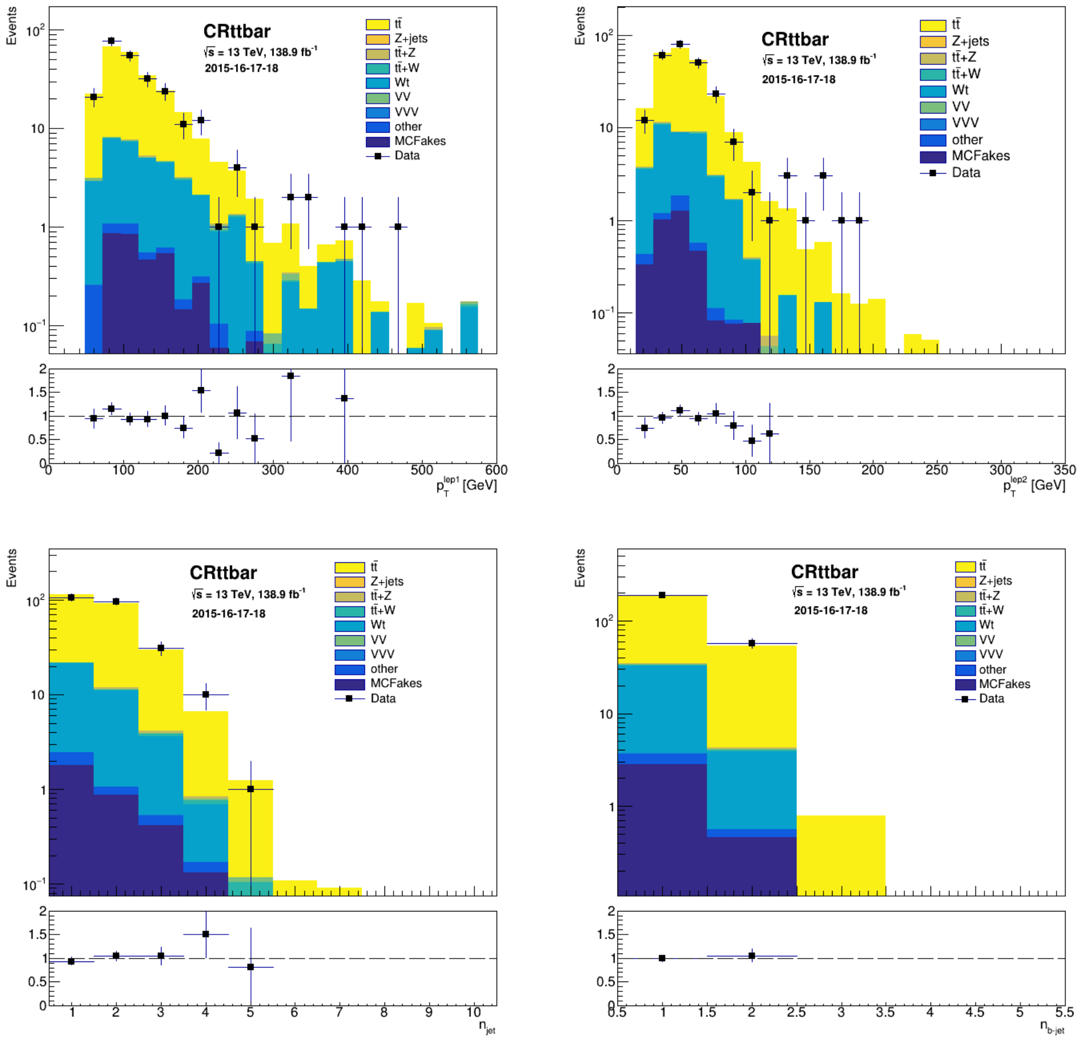


Figure 11: Distributions of MC and data of some relevant variables in the CRttbar.

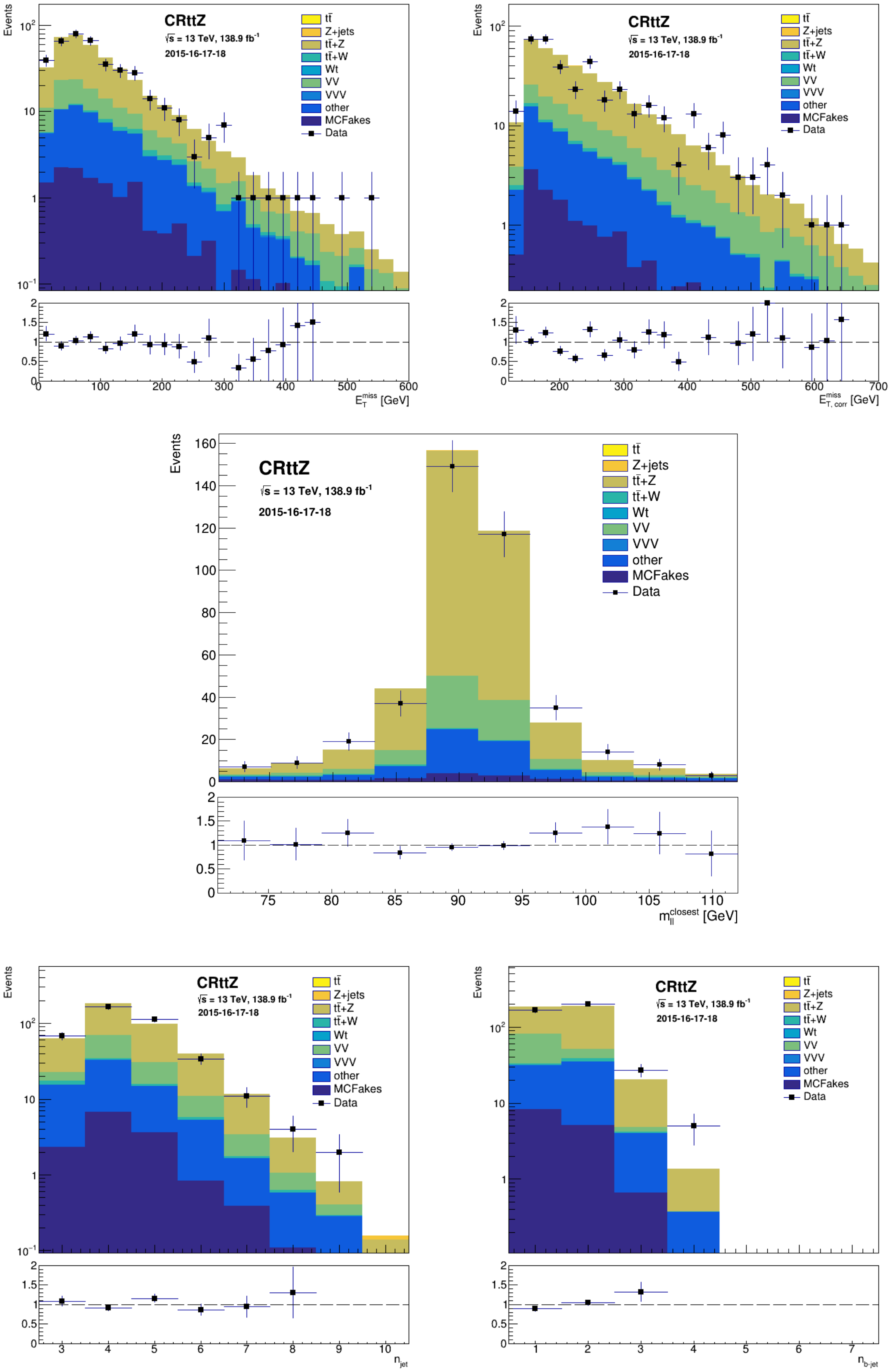
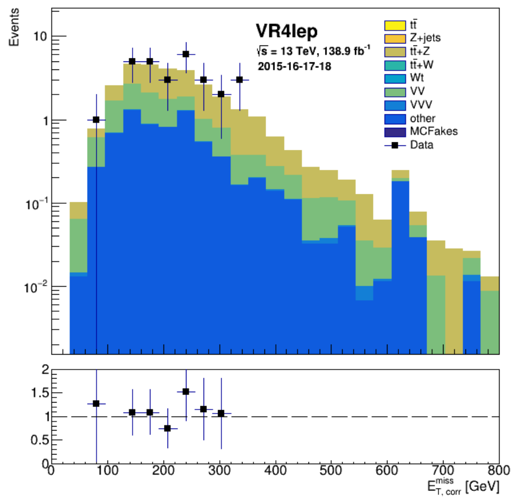
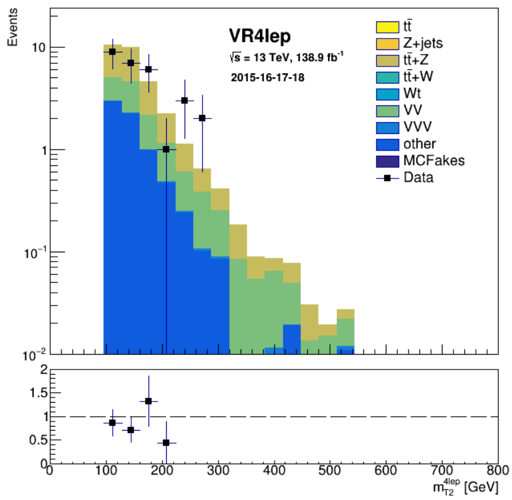
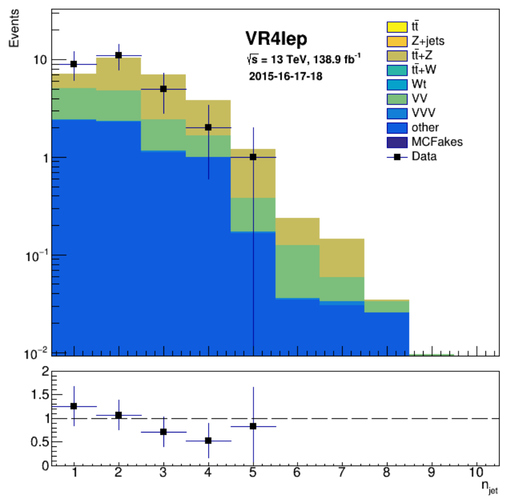
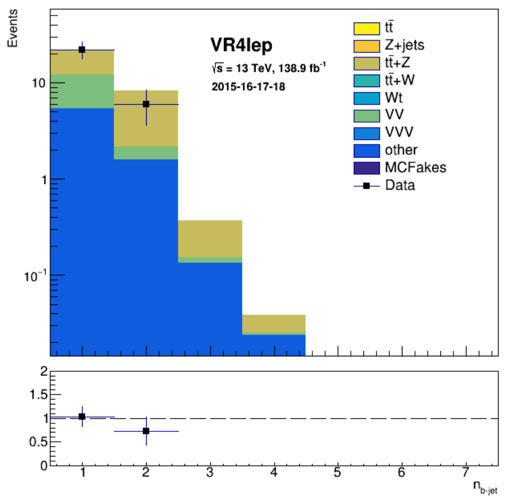


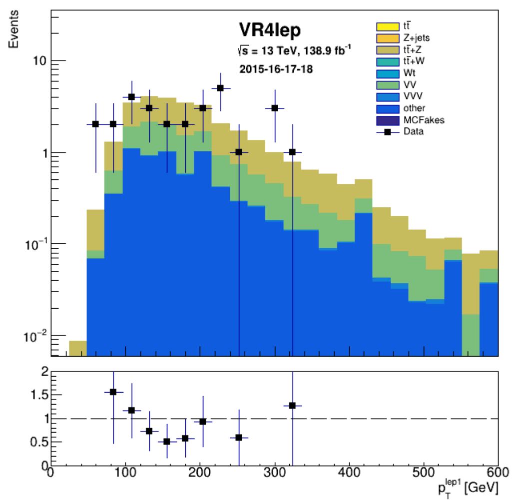
Figure 12: Distributions of MC and data of some relevant variables in the CRttZ.

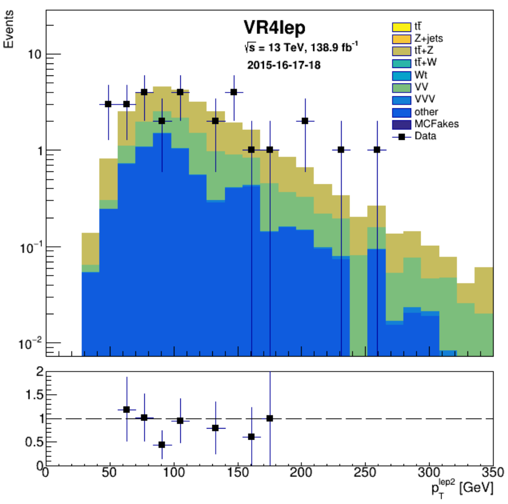

VR4lep
√s = 13 TeV, 138.9 fb-1
2015-16-17-18
tt̄
Z+jets
tt̄+Z
tt̄+W
Wt
VV
VVV
other
MCFakes
Data
Events
E_T, corr^miss [GeV]


VR4lep
√s = 13 TeV, 138.9 fb-1
2015-16-17-18
tt̄
Z+jets
tt̄+Z
tt̄+W
Wt
VV
VVV
other
MCFakes
Data
Events
m_T2^4lep [GeV]


VR4lep
√s = 13 TeV, 138.9 fb-1
2015-16-17-18
tt̄
Z+jets
tt̄+Z
tt̄+W
Wt
VV
VVV
other
MCFakes
Data
Events
n_jet


VR4lep
√s = 13 TeV, 138.9 fb-1
2015-16-17-18
tt̄
Z+jets
tt̄+Z
tt̄+W
Wt
VV
VVV
other
MCFakes
Data
Events
n_b-jet


VR4lep
√s = 13 TeV, 138.9 fb-1
2015-16-17-18
tt̄
Z+jets
tt̄+Z
tt̄+W
Wt
VV
VVV
other
MCFakes
Data
Events
p_T^lep1 [GeV]


VR4lep
√s = 13 TeV, 138.9 fb-1
2015-16-17-18
tt̄
Z+jets
tt̄+Z
tt̄+W
Wt
VV
VVV
other
MCFakes
Data
Events
p_T^lep2 [GeV]

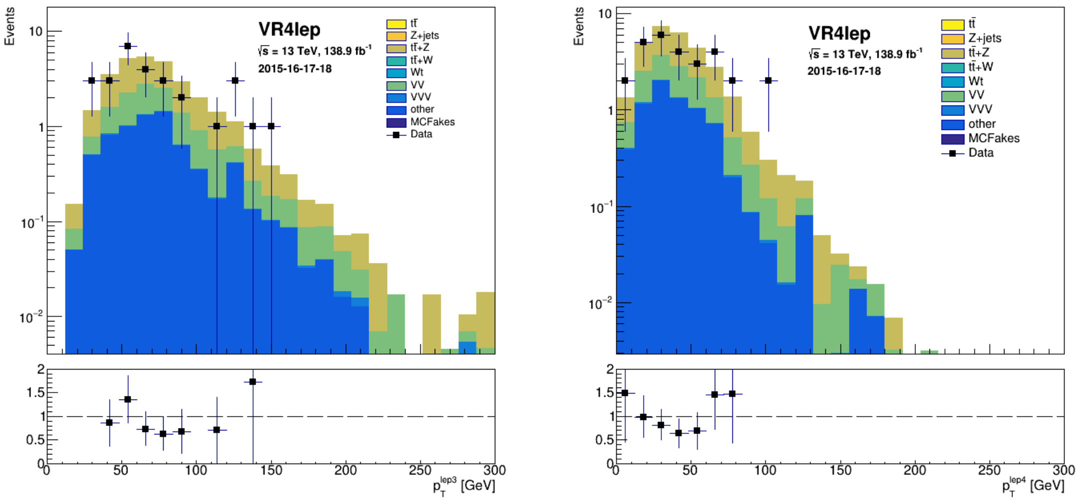


Figure 13: Distributions of the MC and data of some relevant variables in the VR.

References

[1] "Search for dark matter produced in association with bottom or top quarks in $\sqrt{s} = 13TeV$ pp collisions with the ATLAS detector" *The ATLAS Collaboration*

[2] "mT2: the truth behind the glamour" (arXiv:hep-ph/0304226v1) *Alan Barr, Christopher Lester, Phil Stephens*

[3] "Object-based missing transverse momentum significance in the ATLAS detector" *The ATLAS Collaboration*

[4] "Search for direct top squark pair production in final states with two leptons in $\sqrt{s} = 13TeV$ pp collisions with the ATLAS detector" *The ATLAS Collaboration*