Top-quark background estimation for physics BSM in the di-lepton and jets final state with the ATLAS detector

Jeremiah Kgomotso Monnakgotla
School of Physics and Institute for Collider Particle Physics
University of the Witwatersrand

Supervisor
Professor Bruce Mellado

In partial fulfillment of the requirements for the degree of
Master of Science in Physics

30 July 2020

Acknowledgements

National Research Foundation (NRF) financial support to this research is hereby acknowledged. I would like to express my heartfelt thanks to Prof Bruce Mellado, my thesis supervisor. Thank you for advising, making recommendations and encouraging me through this research. Thanks are also due to Dr Yesenia Hernández Jiménez for her continuous monitoring, encouragement and advice she has provided throughout my time.

I would like to thank my colleagues in the HEP group in particular Lehumo Mashishi, Nthabiseng Lekalakala and Thabo Masuku for their assistance in writing this thesis.

I am truly thankful to my parents for their incredible love and care. They also inspired me to explore my possibilities and fulfill my dreams. They have helped me a lot in my life to get to this point.

DECLARATION

I, the undersigned, hereby declare that the work contained in this dissertation is my original work and that any research done by others or by myself has been acknowledged and cited in compliance with the requirements of the Faculty of Science of the University of the Witwatersrand.

30 July 2020

Jeremiah Kgomotso Monnakgotla — Date

Abstract

This thesis presents the top-quark background estimation for the search of a heavy scalar, H, that decays into the Standard Model Higgs boson and a Higgs-like scalar, S. This model is an extension of the Standard Model, where the S and H interact with the Standard Model Higgs boson. This work uses the dataset corresponding to an integrated luminosity of 36 fb^{-1} of pp collisions at a center-of-mass energy of $\sqrt{s}$=13 TeV collected by the ATLAS detector. The final state considered for this analysis is composed of two oppositely charged leptons with different flavour ($e^{\pm}\mu^{\mp}$) and jets. The dominant Standard Model background in this search is the top-quark processes ($t\bar{t}$ and Wt). Top quark control and validation regions are defined and studied. The validation region is used to test the MC modeling of the top quark processes in regards to leptonic kinemates. In this region, data-driven corrections have been extracted to improve Monte Carlo modeling. Higher order electroweak corrections are implemented in the Monte Carlo simulation to improve $t\bar{t}$ modeling. The derived normalization factor in the top-quark control region is applied to the top-quark processes in both the control and the signal regions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

ATLAS and CMS [1, 2] experiments announced the discovery of the Standard Model (SM) Higgs boson (h) in 2012 [3, 4]. This opened up a broad range of research areas to explore whether the observed Higgs boson is unique or is one of the SM Higgs boson spectra. A wide range of SM results published by the ATLAS and CMS collaborations pertaining to multi-lepton production at the Large Hadron Collider (LHC) observes discrepancies between data and Monte Carlo (MC) simulation. It is clear that the available LHC data in the area of multi-lepton production is not well understood. Excesses may well be explained by the Madala hypothesis [5, 6] (see Section 1.3). In view of this, the additional scalars H and S were introduced such that H decays predominantly to Sh, which subsequently decay to a wide range of possible final states, dominated by leptons and jets.

In this work, the top quark background estimation is performed for the $H \rightarrow Sh$ search using a dataset corresponding to an integrated luminosity of 36.1 fb^{-1}

recorded by the ATLAS detector at the centre-of-mass energy ($\sqrt{s}$) of 13 TeV. In Chapter 1 an introduction to the SM of particle physics and the Madala hypothesis is given. Chapter 2 gives an outline of the LHC and a description of the ATLAS detector. Chapter 3 describes the mass range of the beyond Standard Model (BSM) scalars H and S as well as the details about SM background processes considered in the study. Chapter 4 describes the physics objects and the trigger selection used for this analysis. In Chapter 5 the top quark background study is presented and the conclusions are given in Chapter 6.

1.2 The Standard Model of Particle Physics

The SM [7, 8] is a successful theory which provides the accurate description of the interaction of elementary particles. The elementary particles are classified as fermions or bosons. The SM is based on the quantum field theory (QFT) where there exist one field for each type of particle. The SM consists of 12 spin-half fermions that are defined by their interactions. The SM predicts three fundamental forces which concerns four types of gauge vector bosons: the gluon (g), photon (γ), $W^{\pm}$ and Z bosons. These particles have been observed experimentally.

1.2.1 The Elementary Particles

The fermions exist in two main groups: quarks and leptons. There are six particles in each group, each of them related in pairs or "generations". The three generations have identical characteristics and properties with the exception of their masses. The six quarks are divided in three groups, the first generation being "up quark" and the "down" quark, the second generation being "charm quark" and "strange quark," and finally the "bottom quark" and the "top quark". Quarks are distinguishable by their colour charge under the strong force. The six leptons are

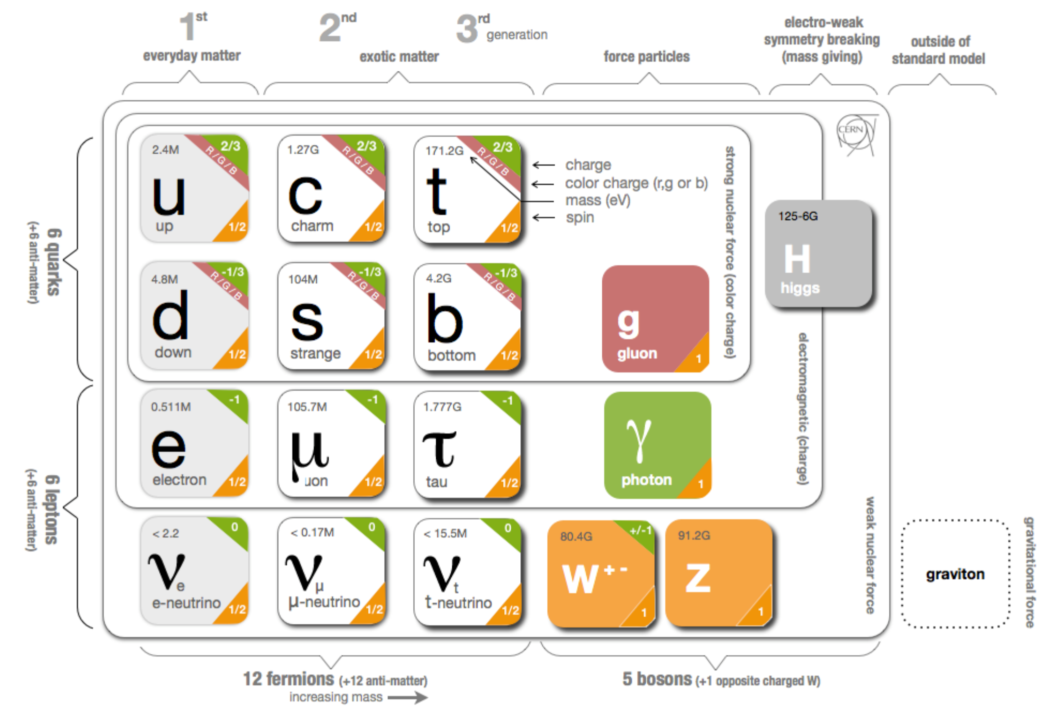


Figure 1.1: The SM particle content.

likewise divided into three generations: electron, electron neutrino, muon, muon neutrino, tau and tau neutrino. The muon has the same electric charge (Q = -1) but it's about 200 times heavier than the electron. The greater mass of the muon means that in electric fields, it accelerates less than the electron and thus, when it is transmitted through the material produce less electromagnetic irradiation. Thus, muons penetrate significantly. The tau is about 3500 times heavier than the electron with the same electrical charge (Q = -1). The tau lepton can decay rapidly, in 2.9×10^{-13} s, to either an electron or a muon including the associated neutrinos.

The forces are carried by spin-1 gauge bosons. The massless gluons mediate the strong force between quarks. The electromagnetic force is carried by a massless γ with no electrical charge, which is transmitted through the interactions of charged particles. This force is governed by electromagnetism's quantum theory of quantum electrodynamics (QED). The shorter range of the weak force is mediated by three gauge bosons, the charged W and neutral Z bosons discovered at CERN in 1983 [9]. The SM particles are summarized in Figure 1.1

The W and the Z bosons couple to the γ implying the unification of the electromagnetic and weak interactions of leptons under gauge group $SU(2)_L \times U(1)_Y$. Subscripts L and Y pertain to the left-handed and weak hypercharge: $Y = Q - T^3$, where Q is the operator of electric charge and T^3 is the third component of the weak operator respectively. The quantum theory of strong force is called quantum chromodynamics (QCD). Spin-1, massless particles classified as g, mediate the strong force. Gluons couple to colour-charged particles like photons couple to electromagnetic charge. The theory of the strong interactions upon the quark is well described by the gauge group with a local $SU(3)_C$ symmetry, where the subscript pertains to the colour charge. The gluon fields, $G^i (i = 1, ..., 8)$ are the $SU(3)_C$ gauge fields. The strong force binding quarks together prohibits the detection of individual quarks through a process called colour confinement. Hence, the quarks and gluons cannot be found as free particle. Finally, the Higgs boson (a complex scalar field) was postulated in 1964 [1, 3, 4] and discovered in 2012 at CERN. This boson generates mass to the fundamental particles through the Brout-Englert-Higgs (BEH) mechanism which will be discussed in Section 1.2.2. The SM Lagrangian with scalar and fermion fields is presented in Equation 1.1:

$$\mathcal{L}_{SM} = \mathcal{L}_G + \mathcal{L}_F + \mathcal{L}_Y + \mathcal{L}_H, \tag{1.1}$$

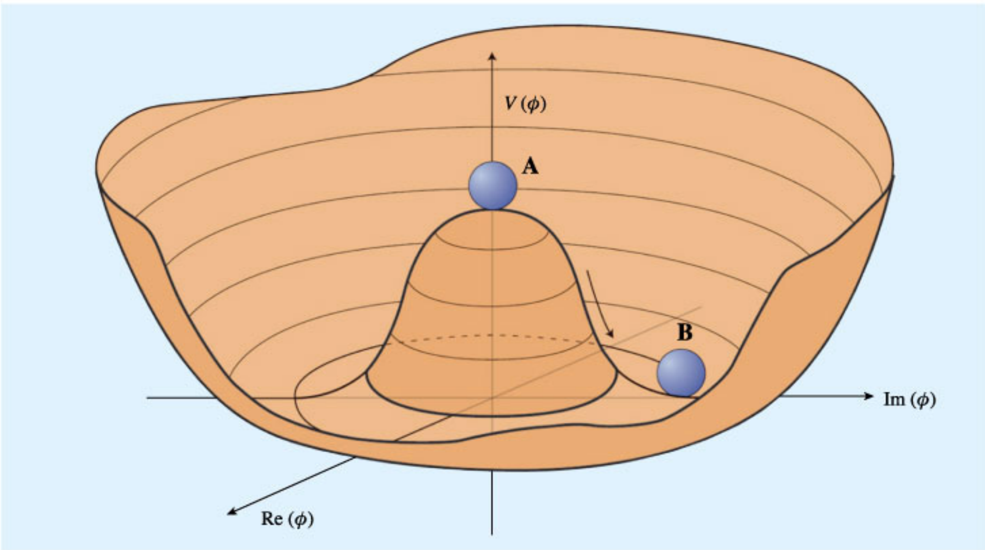


Figure 1.2: The potential corresponding to Equation 1.7 for μ^2 <0.

where the components in Equation 1.1 are given by:

$$\mathcal{L}_G = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{1.2}$$

$$\mathcal{L}_F = \quad i\bar{\psi}D\psi + h.c., \tag{1.3}$$

$$\mathcal{L}_Y = \quad \bar{\psi}_i y_{ij}\psi_j\phi + h.c., \tag{1.4}$$

$$\mathcal{L}_H = \quad |D_\mu\phi|^2 - V(\phi). \tag{1.5}$$

$\mathcal{L}_G$ describes the kinetic energies of the gauge fields, $\mathcal{L}_F$ describes the couplings to fermions, $\mathcal{L}_Y$ describes the Yukawa couplings and $\mathcal{L}_H$ describes couplings to the Higgs boson. The SM Lagrangian presented in Equation 1.1 is invariant under gauge symmetry $U(1)_Y \times SU(2)_L \times SU(3)_C$.

1.2.2 The BEH Mechanism

The application of the BEH mechanism [1, 3, 4] applied to the SM resolves the mass problem for the fermions and gauge bosons through the Spontaneous Symmetry Breaking (SSB) of $SU(2)_L \times U(1)_Y$. In the SM Lagrangian, the gauge symmetry does not allow a mass term for gauge bosons. In order for these gauge bosons to acquire mass, the SSB is then needed. Similarly, the SM fermions are forced to have no mass to satisfy the $SU(2)_L \times U(1)_Y$ symmetry. For illustration purposes one can consider one real scalar field ϕ with Lagrangian given by:

$$\mathcal{L} = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - V(\phi), \tag{1.6}$$

and the potential V is defined by:

$$V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4, \tag{1.7}$$

where the coefficient of the quadratic term in ϕ is interpreted as the mass of the particle that the field describes.

The Lagrangian given in Equation 1.6 is invariant under the transformation: $\phi \rightarrow -\phi$. In this system if $\mu^2 < 0$ the potential will have a minimum in a circle with radius $v = \sqrt{-\mu^2/\lambda}$ which is not equivalent to the value of zero. The vacuum expectation value (*vev*) of ϕ is the constant v. Ultimately, it could be said that the symmetry is "hidden" or spontaneously broken, as shown in Figure 1.2. The field ϕ is the Higgs field and the quantum excitation of this field is the massive spin-0 particle called the Higgs scalar: $m_h^2 = 2\mu^2 = 2\lambda v^2$. The value m_h is 125.09 ± 0.24 GeV [10]. The fermions and the gauge bosons obtain mass from interacting with the Higgs field.

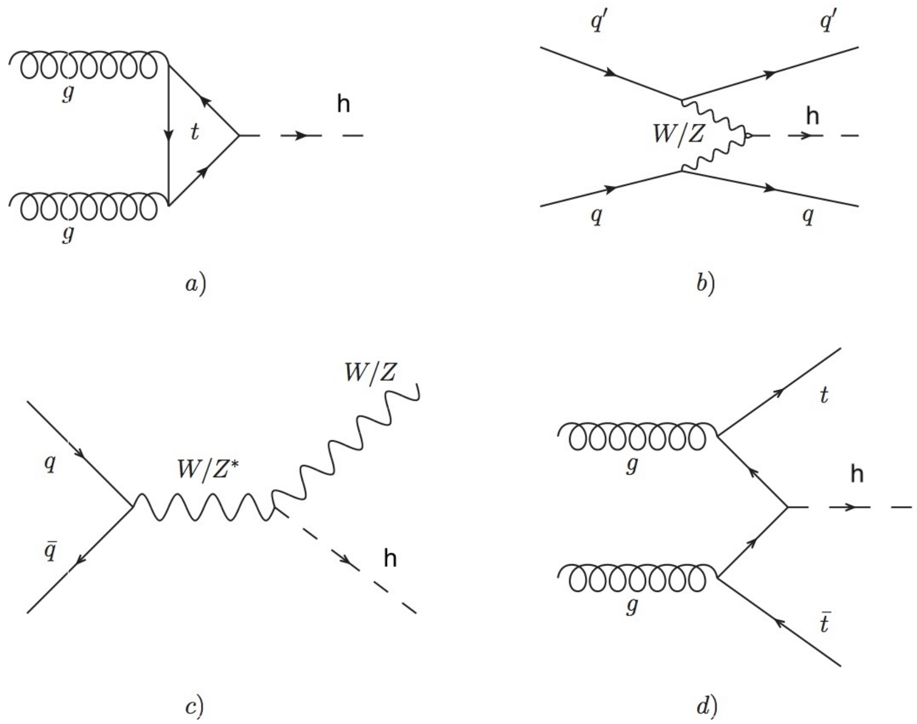


Figure 1.3: Leading order Feynman diagrams for the Higgs production mechanism: (a) gluon-gluon fusion, (b) vector boson fusion, (c) W/Z associated production and (d) top associated production.

1.2.3 Higgs Boson Production

The SM Higgs boson is produced by various mechanisms at the LHC at CERN. The main Higgs boson production mechanism is the gluon-gluon fusion (ggF) which is mediated by virtual heavy top quark, as shown in Figure 1.3a. The second largest cross section of the SM Higgs production mode is vector boson fusion (VBF), as shown in Figure 1.3b. The next most important mechanisms is the Higgs boson associated production with W boson $q\bar{q} \to Wh$ (Wh) and Z boson $q\bar{q}/qq \to Zh$ (Zh) processes, as shown in Figure 1.3c. Contributions are also expected from top fusion ($t\bar{t}h$ and th), as shown in Figure 1.3d. Figure 1.4 illustrates the cross sections for the various Higgs production mechanism as a function of its mass. Higgs boson couples to fermions and gauge boson as dis-

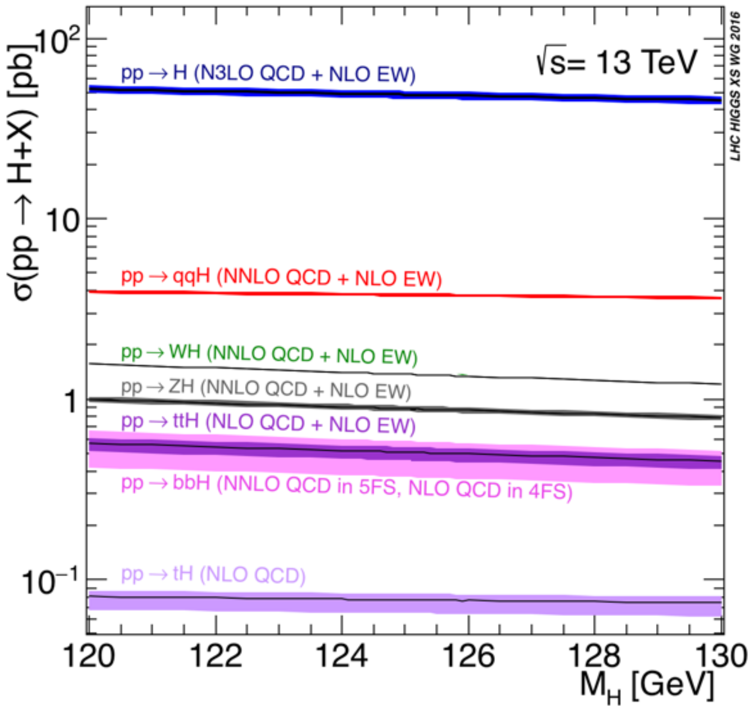


Figure 1.4: The production modes of the Higgs boson [10].

cussed in Section 1.2.4.

1.2.4 SM Higgs Boson Decay Modes

The fermionic decay $h \to b\bar{b}$ is the dominant Higgs decay mode which occurs 57.1% of the time. The fermionic decay $h \to \tau^+\tau^-$ and $h \to c\bar{c}$ occur 6% and 3% of the time, respectively. The off-shell Higgs decays $h \to WW^*$ and $h \to ZZ^{(*)}$ occur 23.1% and 2.9% of the time, respectively. The W and Z off-shell effects are important and lead to $h \to ZZ^{(*)} \to 4\ell$ final state as one of the SM Higgs boson discovery modes. The Higgs boson can also decay to a pair of massless gauge bosons, the gg and $\gamma\gamma$ through a loop of virtual heavy quarks. The gg decay is mainly mediated by top and bottom quark loops, the latter contributing

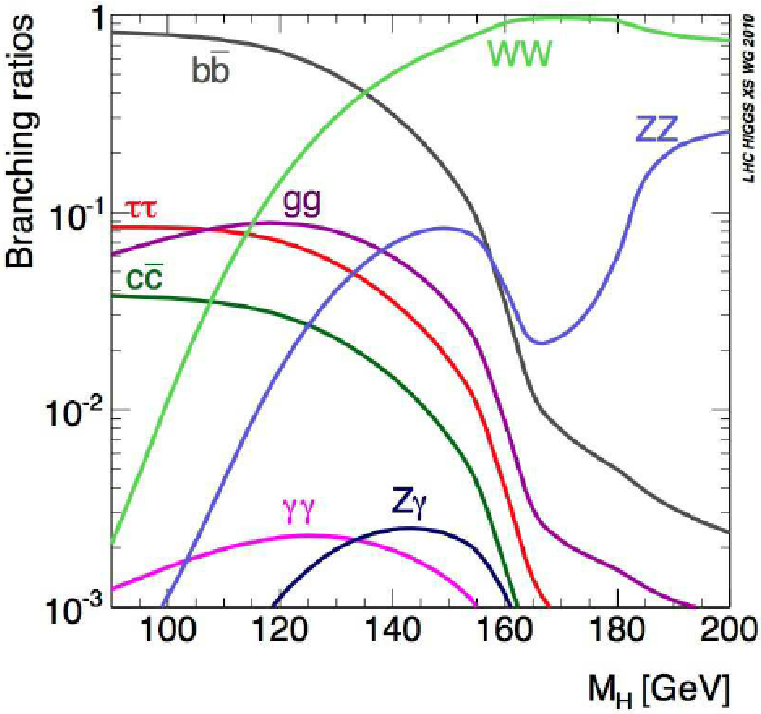


Figure 1.5: Higgs boson decay modes [10].

8.5%. Finally, the $\gamma\gamma$ decay is mainly mediated by the top-quark and the W loops which occur 0.22% of the time. The branching ratios for the various Higgs decay modes are presented in Figure 1.5 and summarised in Table 1.1.

Table 1.1: Branching fractions (BF) of the SM Higgs boson with a mass of 125.09 ± 0.24GeV [10].

Channel	Branching Fraction (%)
$b\bar{b}$	57.1
$\tau^+\tau^-$	6
$c\bar{c}$	3
WW^*	23.1
ZZ^*	2.9
gg	8.5
$\gamma\gamma$	0.22

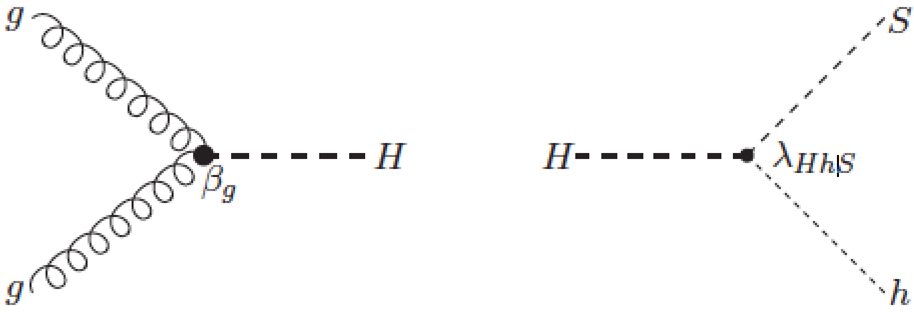


Figure 1.6: Feynman diagrams contributing to the dominant production mode of H (ggF) on the left, and the dominant decay mode ($H \rightarrow Sh$) on the right. The coupling parameters β_g and λ_{HhS} are defined in Equation 1.11 and Equation 1.17.

1.3 The Madala Hypothesis

The discovery of the SM h at the LHC has opened a window for physics BSM. A number of features in the data collected during Run 1 of the LHC led to the formulation of the Madala hypothesis [5, 6]. The Madala hypothesis predict the existence of two new scalar bosons, H and S. In subsequent to the features observed in Run 1 data, both ATLAS and CMS experiments have observed significant excesses in multi lepton data which can be accomodated by the Madala hypothesis as discussed in Ref. [11]. In this work, the Madala hypothesis considers a simplified 2HDM+S model, where the S and H interact with SM h. The S scalar is assumed to be a Higgs-like scalar, such that the number of free parameters are reduced. This allows the S scalar to inherit the Higgs boson branching ratios as illustrated in Figure 1.5 depending on its mass. The $H \rightarrow Sh$ decay is assumed to have a 100% BR. The production and possible decay modes for H are presented in Figure 1.6. The total effective Lagrangian for a simplified model of H and S is:

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_{BSM}, \tag{1.8}$$

where all of the new interactions and states hold in the $\mathcal{L}_{BSM}$ term which is given by:

$$\mathcal{L}_{BSM} = \mathcal{L}_H + \mathcal{L}_S. \tag{1.9}$$

The terms in this Lagrangian are defined as follows:

$$\mathcal{L}_H = \frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2} m_H^2 HH + \mathcal{L}_G + \mathcal{L}_Y, \tag{1.10}$$

$$\mathcal{L}_G = -\frac{1}{4}\beta_g \kappa_{Hgg}^{SM} G_{\mu v} G^{\mu v} H + \beta_V \kappa_{hVV}^{SM} V_\mu V^\mu H, \tag{1.11}$$

$$\mathcal{L}_Y = -\frac{1}{\sqrt{2}}[y_{ttH}\bar{t}tH + y_{bbH}\bar{b}bH]. \tag{1.12}$$

The Lagrangian for the S scalar can be written as:

$$\mathcal{L}_S = \mathcal{L}_K + \mathcal{L}_{SVV} + \mathcal{L}_{Sf\bar{f}} + \mathcal{L}_{hHS}, \tag{1.13}$$

where:

$$\mathcal{L}_K = -\frac{1}{2}\partial_\mu S \partial^\mu S - \frac{1}{2} m_S^2 SS, \tag{1.14}$$

$$\begin{aligned}\mathcal{L}_{SVV} &= \tfrac{1}{4}\kappa_{Sgg}\tfrac{\alpha_S}{12\pi v} S G^{a\mu v} G^a_{\mu v} + \tfrac{1}{4}\kappa_{S\gamma\gamma}\tfrac{\alpha}{\pi v} S F^{\mu v} F_{\mu v} \\ &+ \tfrac{1}{4}\kappa_{SZZ}\tfrac{\alpha}{\pi v} S Z^{\mu v} Z_{\mu v} + \tfrac{1}{4}\kappa_{SZ_\gamma}\tfrac{\alpha}{\pi v} S Z^{\mu v} F_{\mu v} \\ &+ \tfrac{1}{4}\kappa_{SWW}\tfrac{2\alpha}{\pi S_w^2 v} S W^{+\mu v} W^-_{\mu v},\end{aligned} \tag{1.15}$$

$$\mathcal{L}_{Sf\bar{f}} = -\sum_f K_{Sf}\frac{m_f}{v} S\bar{f}f, \tag{1.16}$$

$$\mathcal{L}_{HhS} = -\frac{1}{2}v[\lambda_{hhS}hhS+\lambda_{hSS}hSS+\lambda_{HHS}HHS+\lambda_{HSS}HSS+\lambda_{HhS}HhS]. \quad (1.17)$$

The V symbol represents Z, W and $W^{+}_{\mu v} = D_\mu W^\pm_v - D_v W^\pm_\mu, D_v W^\pm_\mu = [\partial_\mu \pm ieA_\mu]W^\pm_v$.

The first term in Equation 1.11 presents the effective coupling of gluons to the H boson, where κ^{SM}_{Hgg} is the SM coupling given by $\alpha_s/(3\pi v)$. β_g is a parameter that regulates the production rate of $gg \to H$. The term $\beta_g \kappa^{SM}_{Hgg}$ is taken as the effective coupling strength of gluons to the H boson. The second term from Equation 1.11 denotes the effective coupling of VVH. $\kappa^{SM}_{hVV} = m^2_V/v$ regulates the branching ratio of $H \to VV$. Equation 1.12 presents the SM top Yukawa coupling. The sector containing the bottom Yukawa coupling y_{bbH} allows for the $H \to b\bar{b}$ decay which is counted as negligible since the production of bottom quarks via ggF is insignificant.

The $H \to Sh$ decay produces a wide range of possible final states that includes photons, jets, leptons and missing transverse momentum ($E^{\text{miss,track}}_{\text{T}}$). In this work, three mass points of H and S are considered and the production of the signal MC samples are discussed in Section 3.2.

1.3.1 Testing The Hypothesis In The Dilepton Channel

Figure 1.7 is the dilepton invariant mass distribution which shows the comparison of ATLAS data, SM background and BSM signal after selecting events having two oppositely charged leptons with different flavour and at least 1 b-tagged jet [11]. The invariant mass distribution observes discrepancies in the region below 100 GeV which can be accommodated by the $H \to Sh$ signal. The H and S masses are fixed to $m_H = 270$ GeV and $m_S = 150$ GeV, respectively. Simultaneous fits are made for all measurements in Table 1.2 and the profile likelihood ratio is shown in Figure 1.8 for each individual result. The best fit is $\beta_g^2 = 2.92 \pm 0.35$ which corresponds to a significance of 8.04 σ. This is an indication of how the simplified BSM model described here can accommodate the discrepancies in multilepton final states at the LHC.

Table 1.2: A summary of the SM+BSM fit results for each measurement considered Ref. [12].

Selection	**Best-fit** β_g^2	**Significance** (σ)
ATLAS Run 1 $SS\ ll$ and $lll + b$-jets	6.51±2.99	2.37
ATLAS Run 1 $OS\ e\mu + b$-jets	4.09±1.37	2.99
CMS Run 2 $SS\ e\mu\ \mu\mu\ lll + b$-jets	1.41±0.80	1.75
CMS Run 2 $OS\ e\mu$	2.79±0.52	5.45
CMS Run 2 $lll + E_{\mathrm{T}}^{miss}$ (WZ)	9.70±3.88	2.36
ATLAS Run 2 $SS\ ll\ and\ lll + b$-jets	2.22±1.19	2.01
ATLAS Run 2 $OS\ e\mu + b$-jets	5.42±1.28	4.06
ATLAS Run 2 $lll\ + E_{\mathrm{T}}^{miss}$ (WZ)	9.05±3.35	2.52
Combination	2.92±0.35	8.04

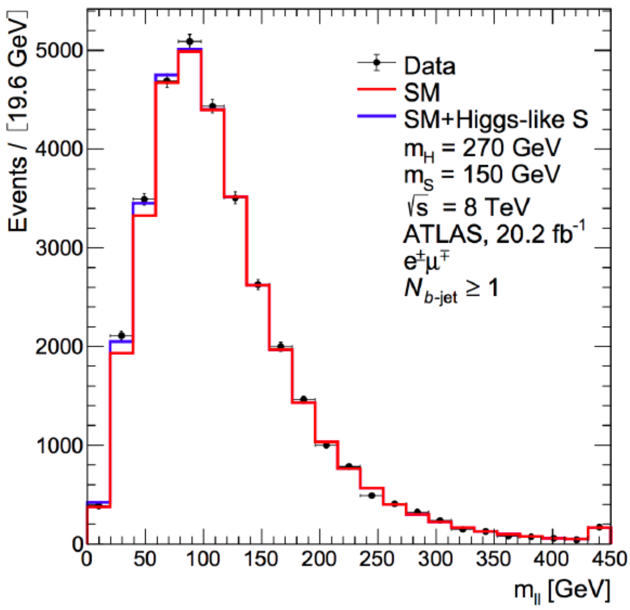


Figure 1.7: Dilepton invariant mass distribution of the ATLAS data (dots) and SM background (red line) in comparison with the BSM signal (blue line) in events where a pair of oppositely charged leptons with different flavour ($e^{\pm}\mu^{\mp}$) with at least 1 b-tagged jet are required [11].

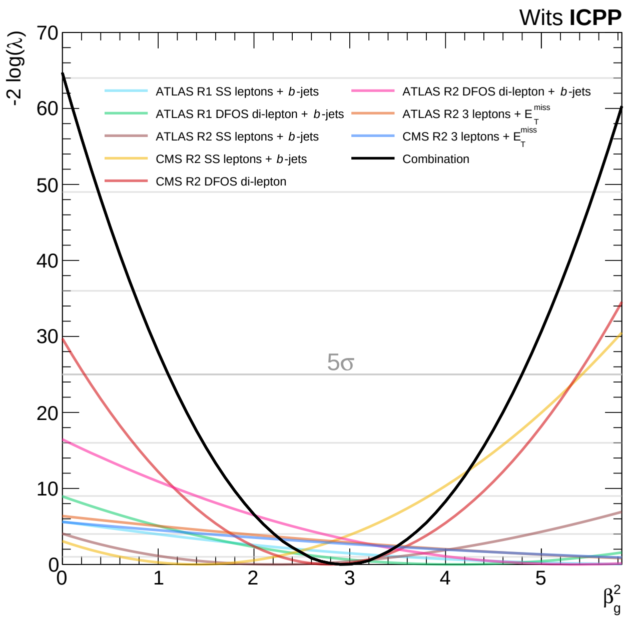


Figure 1.8: The profile likelihood ratios for each result taken into consideration [12].

Chapter 2

The Large Hadron Collider and the ATLAS Detector

2.1 The Large Hadron Collider

The LHC [13, 14] is the most powerful and largest particle collider. It is situated in a tunnel at CERN 100 meters underground and consists of a 27 km ring of superconducting magnets. The CERN accelerator complex consists of machines which inject the proton beam into the next one, before the beam goes to the main LHC ring, as shown in Figure 2.1. To begin with, the protons are extracted from hydrogen atoms taken from a bottle containing hydrogen. This process is performed by stripping the electrons from hydrogen atoms by using an appropriate electric field. The protons accelerating at 750 keV are directed to linear accelerator (LINAC2) and accelerated to 50 MeV. Protons are injected into the PS Booster (PSB) to be accelerated to an energy up to 1.4 GeV. Afterwards, the beam is supplied to the Proton Synchrontron (PS) where they are accelerated to 26 GeV. Subsequently, the beam is fed to a pre-final stage which transfers 450 GeV protons to the main LHC ring (both clockwise and anti-clockwise). The beams

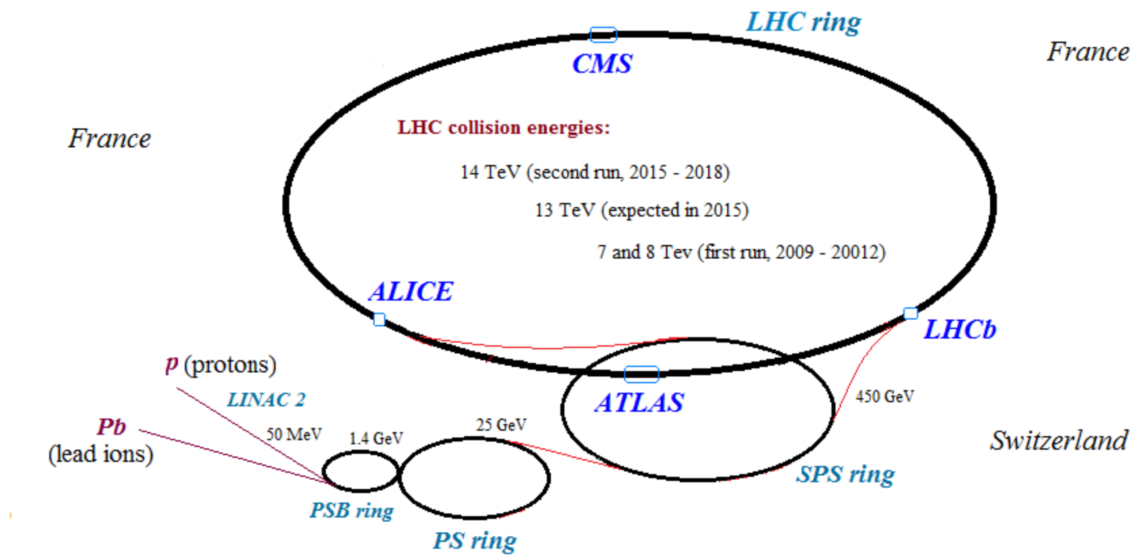


Figure 2.1: A diagram of the LHC particle accelerator.

within the LHC collide around the accelerator ring at four locations that represent the positions of four particle detectors: A Toroidal LHC ApparatuS (ATLAS), the Compact Muon Solenoid (CMS), the Large Hadron Collider beauty (LHCb) and a Large Ion Collider Experiment (ALICE). Figure 2.1 shows the LHC ring and the detectors. ATLAS and CMS are the two largest and general purpose detector which played an important role in the discovery of the SM Higgs boson.

2.2 The ATLAS Detector

The ATLAS detector [15] is one of the two general purpose detector designed to perform precision measurements of the SM Higgs boson and searches for new physics through head on high energetic protons collisions. The weight of the ATLAS detector is approximately 7000 tonnes and it is 46 meters long and 26 meters in diameter. Furthermore, the ATLAS detector is made up of four main

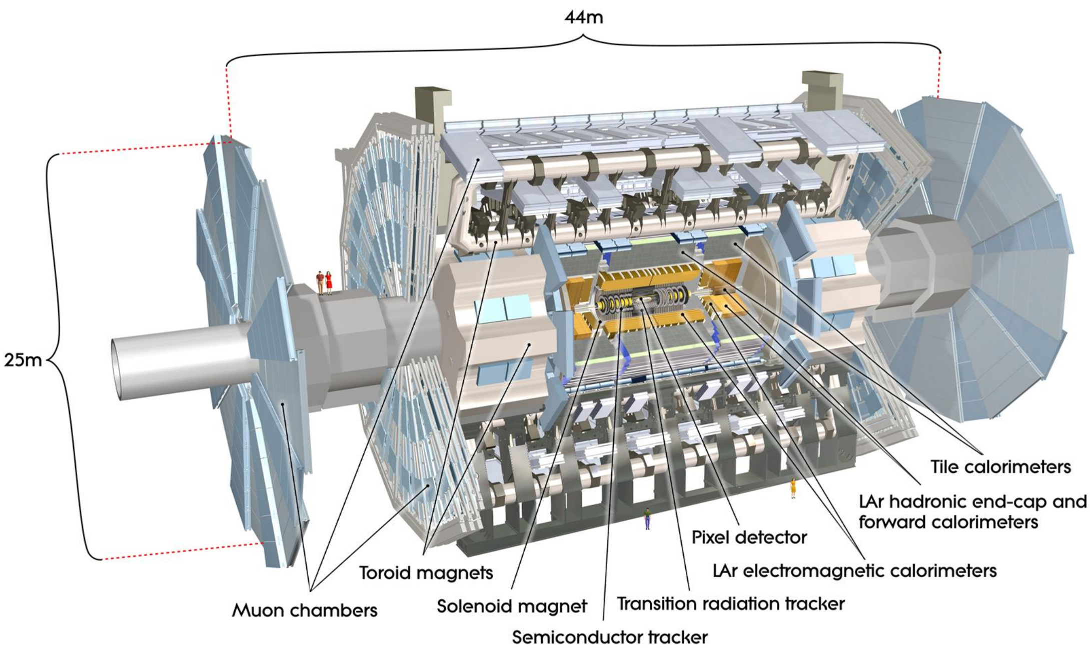


Figure 2.2: The ATLAS detector layout.

components, the Inner Detector (ID), the calorimeters, Muon Spectrometer (MS) and the magnetic system.

ATLAS uses a right-handed Cartesian coordinate system coming from the origin of the detector at the proton-proton (pp) Interaction Point (IP). The positive x-axis points from the IP towards the center of the LHC ring, positive y-axis points upwards towards the earth's surface and the positive z-axis point lies along the proton beam. The x-y plane is transverse to the beam direction, often described in terms of the $r-\phi$ coordinate. The cylindrical coordinate system is defined by:

$$r = \sqrt{x^2 + y^2}, \tag{2.1}$$

$$\phi = \arctan\left(\frac{y}{x}\right), \tag{2.2}$$

$$\theta = \arctan\left(\frac{r}{z}\right), \tag{2.3}$$

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{2.4}$$

The x-axis and r is the radial dimension measuring the distance from the beam line. The azimuthal angle ϕ is measured as usual around the beam axis, and the polar angle θ is the angle from the beam axis to the transverse plane. Finally, η denotes the pseudorapidity which is defined in terms of θ (angle of the positive z-axis).

2.2.1 The Inner Detector

The ID [16, 17] is essentially designed for tracking and identifying charged particles. It is the innermost layer of the ATLAS detector that surrounds the IP at the center where proton beam collisions occur. In order to manage the high density particles formed by the LHC, the ID has been designed to perform high-precision measurements with fine detector granularity. It is surrounded by a 2 T magnetic field generated by the solenoidal superconducting system. It allows the ID to serve as a spectrometer for the reconstruction of the curved trajectories of charged particles. This gives the ability for charged particles with transverse momentum above 500 MeV within $|\eta| < 2.5$ to be reconstructed, since the ones below 500 MeV do not fully cross the ID.

The ID is constructed of three sub-detectors built with two technologies: silicon sensors and straw drift tubes. When charged particles traverse the silicon sensors,

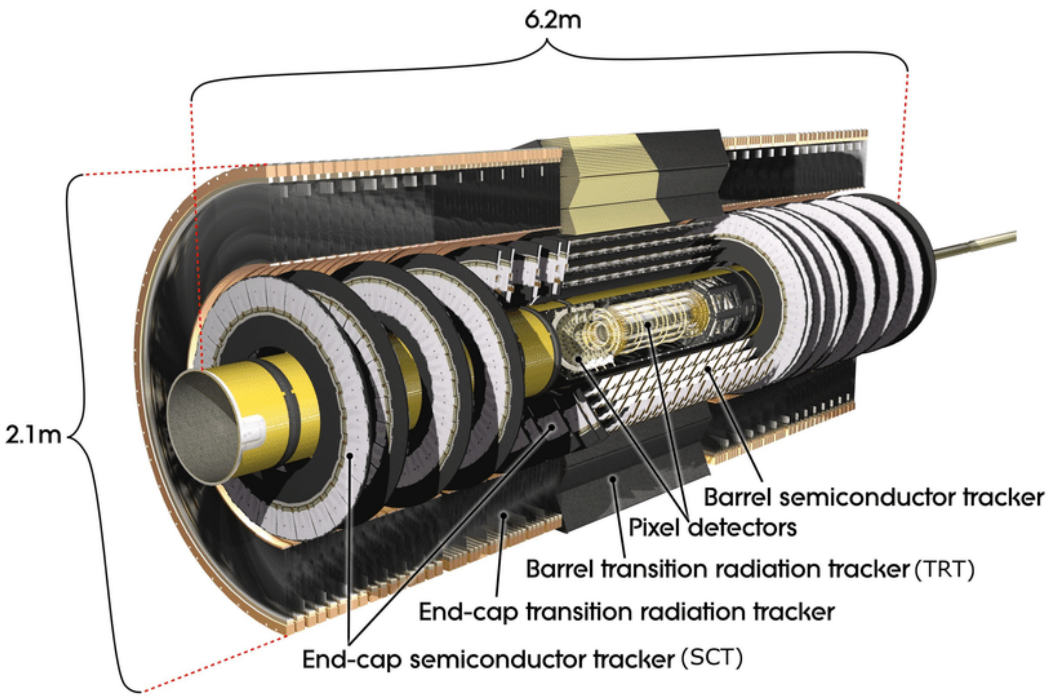


Figure 2.3: A 3D cut-away view of the ID detector.

electron-hole pairs are produced which can be collected using an applied electric field. This charge is locally registered in the sensor, which identifies the particle location. In the straw drift tubes a similar process takes place. The charged particles passing through the drift tubes ionize the gas contained within the straw. Liberated electrons are drifted, with the electron field applied, to the wire in the center of the straw where they are recorded unlike silicon sensors, the primary ionization is multiplied in drift tubes before detection. Silicon pixels are used in the Pixel detector [18, 19], and silicon strips are used in the semiconductor tracker (SCT) [20, 21, 22] . Straw drift tubes are used in the transition radiation tracker (TRT) [23, 24, 25] system. The Pixel detector is the nearest sub-detector to the IP and has the finest granularity. The Pixel detector, consisting of over 80 million channels, produces three measurements per charged particle on the average with a $r - \phi$ plane resolution of 10 m and z-plane resolution of 115 m.

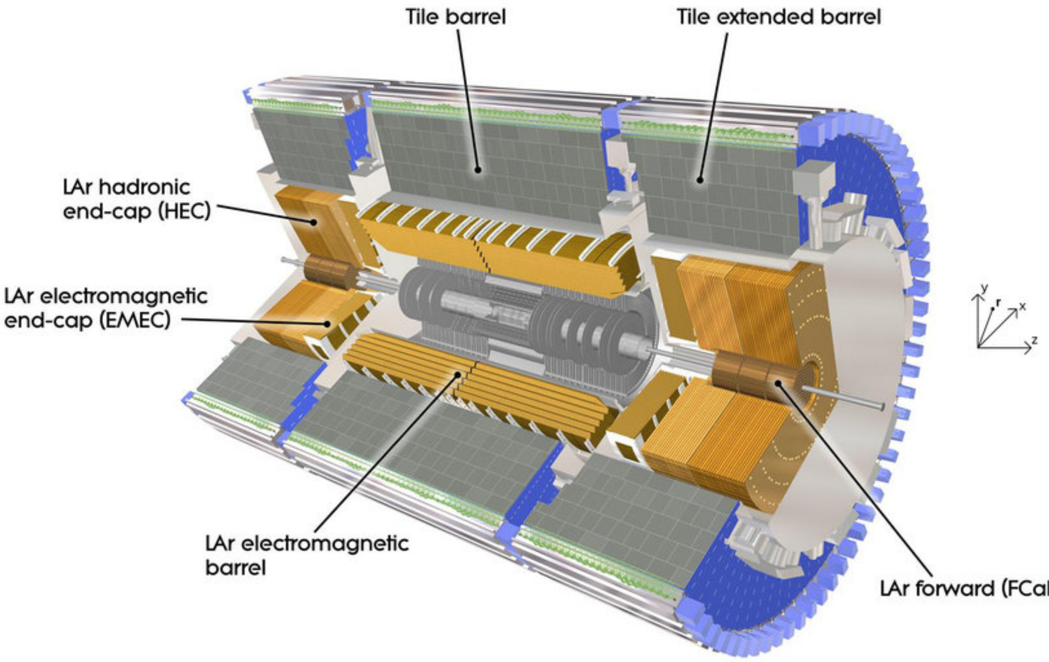


Figure 2.4: A 3D cut-away view of the ATLAS Calorimeter system.

The Pixel detector has a uniform coverage in ϕ, to $|\eta|= 2.5$. The SCT covers the Pixel detectors. Each layer of SCT consists of a double layer silicon strips, the axes of which are leaned by 40 mrad. The TRT is the largest subdetector in the ID. As well as being a tracking detector, the TRT also provides particle identification through transition radiation detection. Charged particles emit photons of transition radiation while traversing the TRT.

2.2.2 The ATLAS Calorimeter System

The ATLAS calorimeter system [26, 27, 28] is designed to measure the energies of the hadrons, electrons and the photons in the range $|\eta| < 4.9$. Figure 2.4 shows a diagram of the calorimeter system. The calorimeter system provides both electromagnetic and hadronic containment, which can stop particles before reaching the muon system. Incident particles generate calorimeter energy shower. The active detection sensors are used to measure only a fraction of the energy emitted by the particle. The energy of the full shower can be determined from the measured energy. This system is composed of four components: the Liquid-Argon (LAr) Electromagnetics (EM) Calorimeter, End-cap calorimeter, the Tile Calorimeter (TileCal) and the LAr hadronic end-cap calorimeter. LAr EM barrel and end-cap calorimeters measure electron and photon energies. The EM calorimeter is a lead LAr detector with a special geometry which provides a full and uniform ϕ coverage and fast readout. The EM calorimeter is separated in three radial sections of various η-ϕ granularities. The energies of the electrons and photons are measured by the LAr EM calorimeter and end-cap calorimeter in the range $|\eta| < 3.2$ and $3.1< |\eta| < 4.9$, respectively. The energies of the hadrons measured by the TileCal and LAr hadronic end-cap in the range $|\eta| < 1.7$ and $1.5< |\eta| < 3.2$, respectively.

2.2.3 The Muon Spectrometer

The MS system [29, 30, 31] measures the momenta and position of charged particles exiting the barrel and end-cap calorimeters in the pseudorapidity range $|\eta| < 2.7$. This is the outermost part of the ATLAS detector, as shown in Figure 2.5. The MS operates in a toroidal magnetic field. The large barrel toroid produces magnetic bending over the range $|\eta| <1.4$. For $1.6 < |\eta| < 2.7$, two smaller end-cap magnets mounted on the two ends of the toroid are used to bend the muon tracks. A combination of the barrel and the end-cap fields produces a bending in the region $1.4 < |\eta| < 1.6$. In the barrel region the precision chambers are placed between the eight coils of the superconducting barrel toroid magnet, while the chambers are located in front and behind the two toroid magnets of the end cap. The muon positions in the barrel region are determined in chambers arranged around the beam axis at radius of approximately 5 m, 7.5 m, and 10

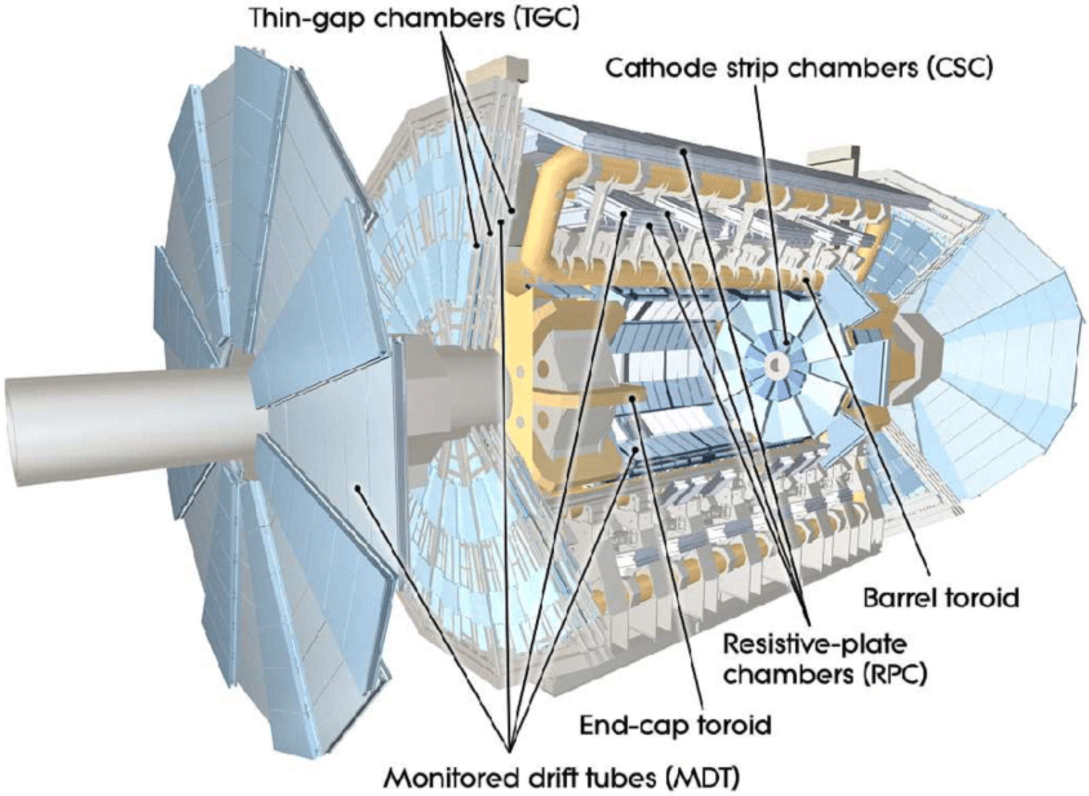


Figure 2.5: Cut-away view of the ATLAS muon system.

m in three cylindrical layers. The chambers are arranged in three planes perpendicular to the beam, in the transition and end-cap regions. The monitored drift tubes (MDT) covers range $|\eta| < 2.7$ which perform the precision momentum measurement. These chambers are made of 3 to 8 layers of drift tube operated under an absolute pressure of 3 bar. Cathode-strip chambers (CSCs) are used in the forward region ($2 < |\eta| < 2.7$) in the innermost tracking layer.

2.2.4 Trigger and Data Acquisition System

The Trigger and Data Acquisition (TDAQ) system is implemented to perform a run-time check on numerous interactions per second to select only interesting events for physics analysis and discard unusable information. The trigger com-

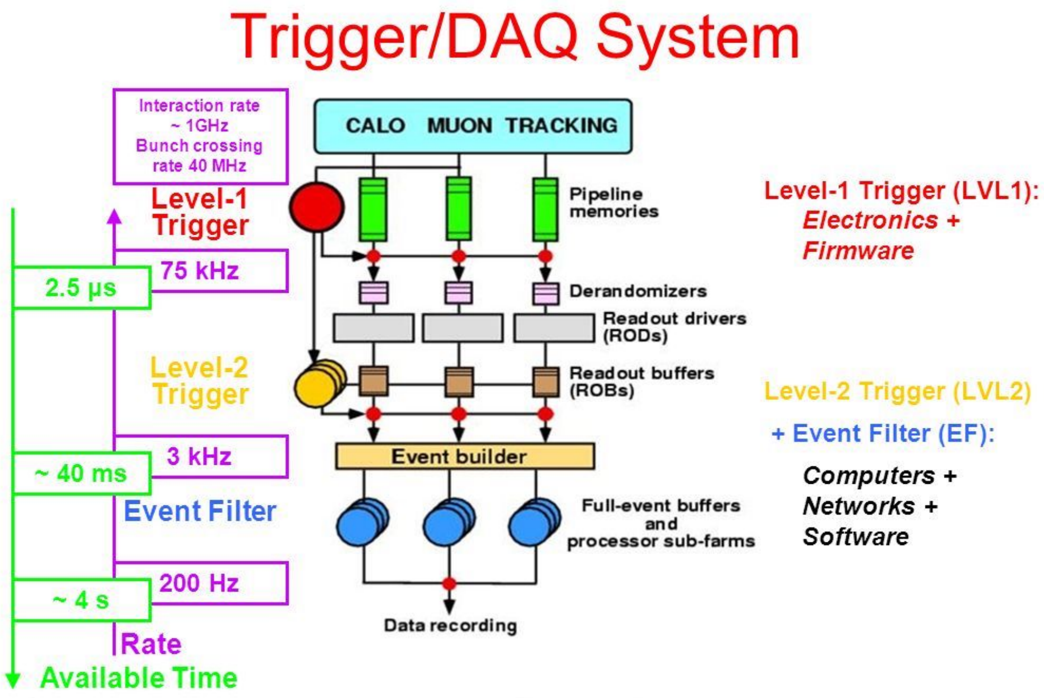


Figure 2.6: ATLAS trigger and data acquisition systems block diagram.

prises three levels of event selection: hardware and software that is Level-1 (L1), Level-2 (L2) and event filter. The combination of trigger L2 and event filter form the High-Level Trigger (HLT), which is predominantly focused on computer and networking devices. The L1 trigger is used to identify signatures from high-p_T muons, electrons, photons and jets decaying into hadrons. This is achieved through the use of information from a subset of detectors: the Resistive Plate Chambers (RPC) and Thin-Gap Chambers (TGC). The maximum acceptable rate of L1 trigger to handle the read-out detector systems is 75 kHz (upgradeable to 100 kHz). Subsequently, the L2 trigger is fed by detector regions where potential trigger events within the event have been identified by the L1 Trigger, to reduce the data to be transmitted from the read-out detector. In that process, this L2 trigger, with an average event processing time of about 40 ms, reduces the event rate to below 3.5 kHz. The event filter uses offline analysis methods on fully constructed events to classify events at a rate to be captured for offline analysis. This decreases the event rate to about 200 Hz, with an interval of four seconds for the processing of events. Figure 2.6 displays a block diagram with triggers and data acquisition systems.

Chapter 3

Data and Monte Carlo samples

3.1 Data Set

This analysis is based on data collected with the ATLAS detector during the 2015 and 2016 data-taking periods of pp collisions at $\sqrt{s}$ =13 TeV with a bunch spacing of 25 ns. The data sample corresponds to an integrated luminosity of 36.1 fb^{-1}. The Good Run Lists (GRL) are applied to guarantee the quality of the analysed collisions as the GRL only contain data collected under good detector conditions. The GRL used for this analysis are:

- data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02_PHYS_StandardGRL_All_Good_25ns.xml
- data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml.

Because of the high LHC peak luminosity and small bunch separation, the number of pp interactions occurring in the same bunch crossing is large. This is referred to as pile-up event. Residual differences in the pile-up between data and MC simulation have been corrected by reweighting the MC events to reproduce the

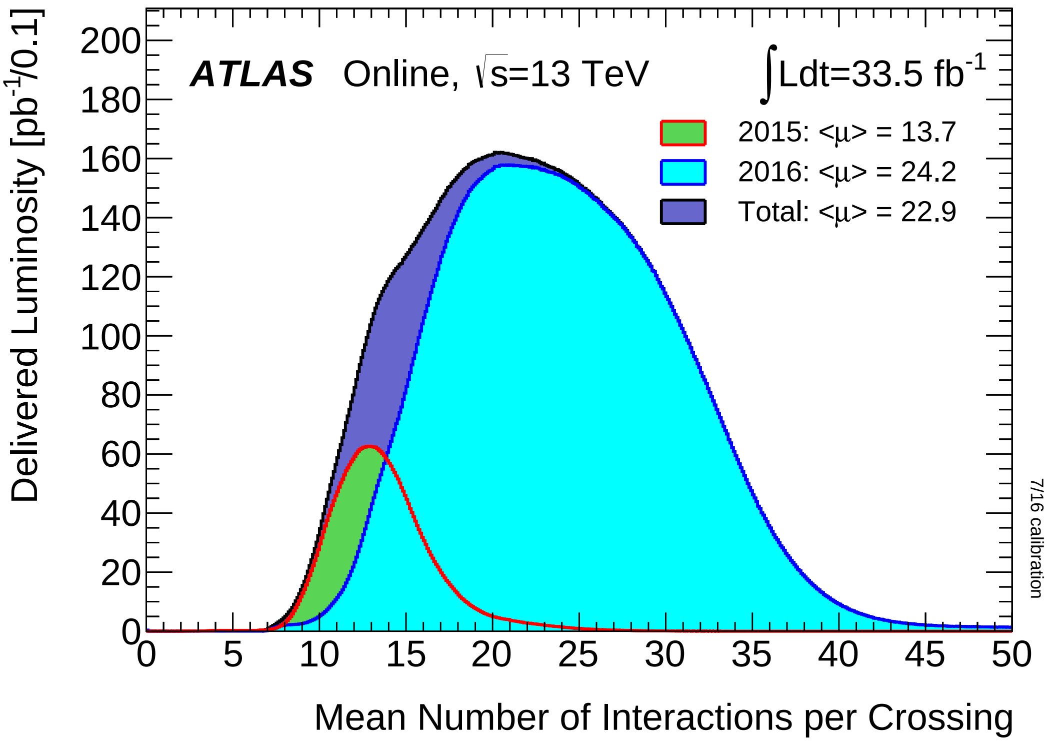


Figure 3.1: Mean number of interactions per bunch crossing, recorded by the ATLAS detector in 2015 and 2016.

average number of interactions per bunch-crossing ($< \mu >$) observed in data as shown in Figure 3.1.

3.2 MC Samples

In order to model signals and background processes, MC event generators are used. The $H \rightarrow Sh$ signal discussed in Section 1.3 is modeled using PYTHIA v8.210 event generator with two lepton filter. All possible decay modes for Sh are allowed. A two lepton filter only saves events that satisfy the dileptonic final states. These leptons are required to have transverse momentum (p_{T}) above 8 GeV with $|\eta| < 2.5$. $H \rightarrow Sh$ signal samples are produced for three mass points

Table 3.1: Summary of the MC $H \to Sh$ signal samples produced.

Signal	m_H	m_S	cross section (σ) [pb]	Number of events	Filter efficiency
	240 GeV	170 GeV	1	30k	0.097021
$H \to Sh$	350 GeV	240 GeV	1	30k	0.177186
	400 GeV	240 GeV	1	30k	0.195975

as presented in Table 3.2. The cross section for $H \to Sh$ is assumed to be 1 pb. The filter efficiency is the ratio of accepted events divided by total number of event generation attempts. The details for the signal production is given in Appendix A.1.

Background processes are top-quark pair production ($t\bar{t}$ without or in association with vector bosons), single-top in association with W and Z boson (Wt and Zt), single vector boson production (Z and W in conjunction with jets) and the Higgs boson production via ggF, VBF and in association with W and Z bosons (Wh, Zh and WW), other diboson production (WW, WZ, ZZ, $V\gamma$ where $V = W$, Z).

For $t\bar{t}$ and Wt samples, the Powheg-box v2 [32] is used with the CT10 [33] next-to-leading order (NLO) parton distributions function (PDF) set. Simulated top events are interfaced with Pythia v8.210 [34] using the A14 tune for $t\bar{t}$ and Pythia v6.428 using Perugia 2012 tune for Wt production for parton shower, hadronization and underlying event. The top-quark mass is set to 172.5 GeV. A parton shower MC event generator completely simulates the final state of high-energy collisions down to the level of stable particles. The underlying event arises from collisions between those partons which are not directly involved in the hard subprocess in the incoming hadrons.

The Sherpa v2.2.1 is used to model the production of $Z(\to ee/\mu\mu/\tau\tau)$ + jets along with the NNPDF 3.0 PDF set at next-to-next-to leading order (NNLO).

Sherpa v2.2.1 provides its own parton shower, hadronization and underlying event modeling. In these samples the minimum invariant mass of the dilepton system is 10 GeV. The samples include up to two additional parton emissions at NLO and up to four at LO.

W ($\rightarrow e\nu/\mu\nu/\tau\nu$)+jets process is simulated using Powheg-box v2 generator using the CT10 NLO PDF, interfaced to Pythia v8.186 for parton showering, hadronization and underlying event. For the WW samples, the AZNLO [35] tune is used for the underlying event in combination with the CTEQ6L1 PDF set. The A14 tuning is used with the NNPDF23LO PDF set for the underlying event. Hadronizing transforms partons into observed hadrons that are eventually detected in the detector.

Background diboson production (WZ, $ZZ, W\gamma$ and $Z\gamma$) is simulated using the Sherpa v2.2.2 generator with the NNPDF 3.0 PDF set at NNLO. The minimum p_{T} in the $W\gamma$ and $Z\gamma$ samples is set to 7 GeV. The WW process is modelled at NLO using the Powheg-box v2 event generator for $q\bar{q}$ initial states at NLO, including NNLO contributions from $gg \rightarrow WW \rightarrow \ell\nu\ell\nu$ production using the Sherpa+OpenLoops framework for non-resonant and resonant Higgs boson production.

The Higgs boson associated production with W and Z bosons is modeled using Powheg-box v2 in conjuction with the CT10 at NLO interfaced to Pythia v8.210 for parton showering, hadronization and underlying event. The Higgs boson production via ggF is simulated at NNLO using the Powheg-box v2 NNLOPS program. VBF Higgs boson production is simulated using the Powheg-box v2 with the CTEQ6L1 at NLO interfaced to Pythia v8.210 for parton showering, hadronization and underlying event. A full list of the background samples is shown in Table 3.2.

Table 3.2: MC generators used to model the background processes, and corresponding cross sections.

Process	cross section [nb]	Generator	PDF
$gg \to h \to WW \to \ell\nu\ell\nu$	$1.102{\times}10^{-3}$	POWEHEG+PYTHIA8	CTEQ6L1
$qq \to h \to WW \to \ell\nu\ell\nu$	$8.579{\times}10^{-5}$	POWEHEG+PYTHIA8	CTEQ6L1
$Wh \to e\nu eW\ell qq$	$5.601{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to qqW\ell$	$5.601{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to qqW\ell$	$8.080{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to qqWW\ell\nu qq$	$3.37210{\times}10^{-5}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to \tau\nu\tau WW\ell\nu qq$	$5.601{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to e\nu eWW\ell\nu qq$	$8.824{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to \mu\nu\mu\nu WW\ell\nu qq$	$8.824{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to qqWW\ell\nu\ell\nu$	$1.274{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to qqWW\ell\nu qq$	$5.316{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to \tau\nu\tau WW\ell\nu qq$	$8.824{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Wh \to \ell\nu W(\ell\nu)W(\ell\nu)$	$4.027{\times}10^{-6}$	POWEHEG+PYTHIA8	NNPDF3
$Zh \to (ee/\mu\mu/\tau\tau)WW\ell\nu(\ell\nu/qq)$	$2.404{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Zh \to (e\nu e\nu/\mu\nu\mu\nu/\tau\nu\tau\nu)WW\ell\nu qq$	$4.775{\times}10^{-6}$	POWEHEG+PYTHIA8	CT10 NLO
$Zh \to W(\tau\nu)W(\tau\nu)\ell\nu qq$	$1.192{\times}10^{-5}$	POWEHEG+PYTHIA8	NNPDF3
$gg \to Zh \to \ell\ell WW$	$2.786{\times}10^{-7}$	POWEHEG+PYTHIA8	NNPDF3
$qq \to WZ\gamma^* \to Vqq \to qq\ell\ell$	$3.2881{\times}10^{-3}$	POWEHEG+PYTHIA8	CTEQ6L1
$qq \to WZ\gamma^* \to WZ \to \ell\nu\ell\ell$	$4.5765{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$qq \to WZ\gamma^* \to WZ \to \ell\nu\ell\ell$	$2.9664{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$qq \to ZZ \to \ell\ell q\bar{q}$	$1.4555{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$qq \to ZZ \to \ell\ell\nu\nu$	$1.252{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$gg \to WW \to \ell\nu\ell\nu$	$2.0524{\times}10^{-4}$	SHERPA	NNPDF30NNLO
$qq \to WWqq \to \ell\nu\ell\nu jj$ (same sign, 4EW coupling)	$2.521{\times}10^{-5}$	SHERPA	NNPDF30NNLO
$qq \to WWqq \to \ell\nu\ell\nu jj$ (6EW coupling)	$4.0864{\times}10^{-5}$	SHERPA	NNPDF30NNLO
Wt dilepton filter	$3.997{\times}10^{-3}$	POWEHEG+PYTHIA8	CT10 NLO
$t\bar{t}$ hdamp258p75	$7.2977{\times}10^{-1}$	POWEHEG+PYTHIA8	CT10 NLO
tZ 4fl t-channel trilepton	$2.0736{\times}10^{-5}$	MADGRAPH+PYTHIA	CT10 NLO
$t\bar{t}W$	$5.4830{\times}10^{-4}$	POWEHEG+PYTHIA8	CT10 NLO
$Z \to ee/\mu\mu/\tau\tau$ ($m_{\ell\ell} > 40$ GeV)	1.98114	SHERPA	NNPDF30NNLO
$Z \to ee/\mu\mu/\tau\tau$ (10 GeV $< m_{\ell\ell} <$ 40 GeV)	2.4137	SHERPA	NNPDF30NNLO
$W^+(e/\mu/\tau)\nu$	$1.1306{\times}10^{1}$	POWEHEG+PYTHIA8	CTEQ6L1
$W^-(e/\mu/\tau)\nu$	8.2826	POWEHEG+PYTHIA8	CTEQ6L1
$Z\gamma \to (ee/\mu\mu)\gamma$ (7 GeV $> p_T^\gamma \leq$ 15 GeV)	$57.654{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$Z\gamma \to (ee/\mu\mu)\gamma$ (15 GeV $> p_T^\gamma \leq$ 35 GeV)	$34.568{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$Z\gamma \to (ee/\mu\mu)\gamma$ (35 GeV $> p_T^\gamma \leq$ 70 GeV)	$6.2934{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$Z\gamma \to (ee/\mu\mu)\gamma$ (70 GeV $> p_T^\gamma \leq$ 140 GeV)	$0.49159{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$Z\gamma \to (ee/\mu\mu)\gamma$ ($p_T^\gamma >$ 140 GeV)	$0.063052{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$W\gamma \to (e/\mu/\tau)\nu\gamma$ (7 GeV $> p_T^\gamma \leq$ 15 GeV)	$199.28{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$W\gamma \to (e/\mu/\tau)\nu\gamma$ (15 GeV $> p_T^\gamma \leq$ 35 GeV)	$134.41{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$W\gamma \to (e/\mu/\tau)\nu\gamma$ (35 GeV $> p_T^\gamma \leq$ 70 GeV)	$19.081{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$W\gamma \to (e/\mu/\tau)\nu\gamma$ (70 GeV $> p_T^\gamma \leq$ 140 GeV)	$1.9228{\times}10^{-3}$	SHERPA	NNPDF30NNLO
$W\gamma \to (e/\mu/\tau)\nu\gamma$ ($p_T^\gamma >$ 140 GeV)	$0.29968{\times}10^{-3}$	SHERPA	NNPDF30NNLO

Chapter 4

Event Reconstruction and Selection

4.1 Leptons

The event candidates must have at least one vertex with at least two associated tracks with p_{T} above 400 MeV [36]. The vertex with the highest summation of the p_{T}^2 of the associated tracks is considered to be the primary vertex (PV).

4.1.1 Electrons

Electrons are required to originate from PV. Electron candidates are reconstructed from a localised cluster of energy deposits in the EM calorimeter matched to their track in the ID. The signal electron candidates are required to have E_{T} above 15 GeV and pseudorapidity is required to be $|\eta| < 2.47$, excluding the range of $1.37 < |\eta| < 1.52$, which is the transition region between the barrel and endcaps in the LAr calorimeter. This is due to large amount of material in the transition region, where the electromagnetic objects are measured with worse resolution and efficiency and they are not used in this analysis.

In addition, electron candidates are required to satisfy the TightLH quality for

Table 4.1: Electron selections.

p_{T} range	Electron ID	Impact parameter
< 25GeV	TightLH	$\lvert z_0 \times sin(\theta)\rvert < 0.5$ mm, $\lvert d_0/\sigma_{d_0}\rvert < 5$
> 25GeV	MediumLH	

Table 4.2: Definition of the lepton isolation working points.

p_{T} range	Lepton	Calo isolation	Track isolation	Operating point
< 25GeV	electron	$E_{\mathrm{T}}^{cone20}/p_{\mathrm{T}} < 0.11$	$p_{\mathrm{T}}^{varcone40}/p_{\mathrm{T}} < 0.6$	-
> 25GeV	-	-	-	passIsoGradient
> 15GeV	muon	$E_{\mathrm{T}}^{cone20}/p_{\mathrm{T}} < 0.09$	$p_{\mathrm{T}}^{varcone30}/p_{\mathrm{T}} < 0.06$	-

$p_{\mathrm{T}} < 25$ GeV and MediumLH definition for $p_{\mathrm{T}} > 25$ GeV as discussed in Ref. [37]. Furthermore the longitudinal impact of each electron track ($\lvert z_0 \times sin(\theta)\rvert$) is required to be less than 0.5 mm, where z_0 is the distance between the tracks and the primary vertex along the beam line, as shown in Figure 4.1. Moreover, the significance of the transverse impact parameter computed with respect to the beam line ($\lvert d_0/\sigma_{d_0}\rvert$) is required to be less than 5.0 where d_0 and σ_{d_0} denote transverse impact parameter and its estimated uncertainty, respectively. The signal electron candidates with $p_{\mathrm{T}} < 25$ GeV are isolated using calorimeter and track isolation variables: $E_{\mathrm{T}}^{cone20}/p_{\mathrm{T}} < 0.11$ and $p_{\mathrm{T}}^{varcone40}/p_{\mathrm{T}} < 0.6$. Subsequently, signal electron candidates with $p_{\mathrm{T}} > 25$ GeV are isolated using a so-called gradient working point [37, 38]. The electron selections and the electron isolation working points are shown in Table 4.1 and Table 4.2, respectively.

4.1.2 Muons

Muon candidates are reconstructed by matching a tracking information in the ID with a track in the MS. The combined tracks are required to satisfy the Medium identification criteria as defined in Ref. [38]. Signal muons are required to have $p_{\mathrm{T}} > 15$ GeV and $\lvert\eta\rvert < 2.5$. Furthermore, the parameter $\lvert z_0 \times sin(\theta)\rvert$ is required

Table 4.3: Muon selections.

p_{T} range	Electron ID	Impact parameter
> 15GeV	Medium	$\|z_0 \times sin(\theta)\| < 0.5$ mm, $\|d_0/\sigma_{d_0}\| < 3$

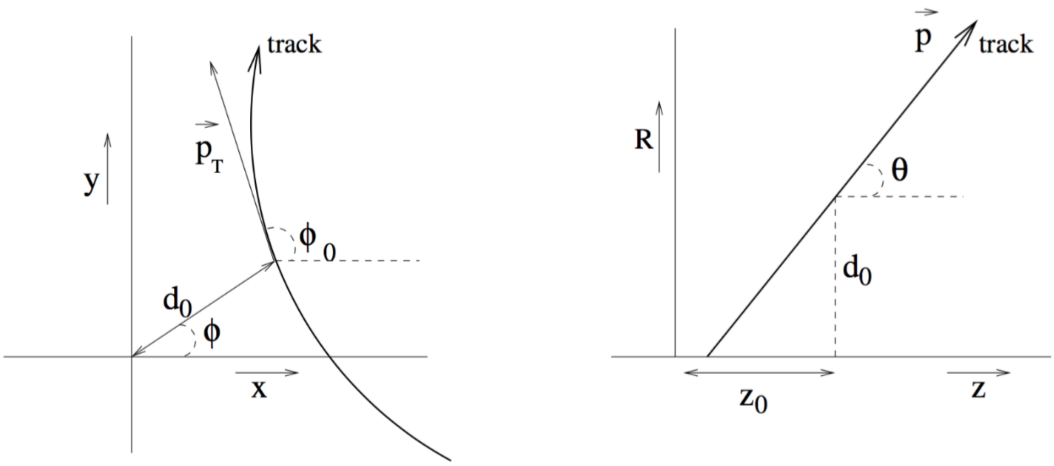


Figure 4.1: Illustration of the perigee parameters of the track in the transverse plane (left) and RZ-plane (right) as defined in the global ATLAS tracking frame.

to be less than 0.5 mm and $|d_0/\sigma_{d_0}|$ to be less than 3. The signal muon candidates are isolated using calorimeter and track isolation variables: $E_{\mathrm{T}}^{cone20}/p_{\mathrm{T}} < 0.09$ and $p_{\mathrm{T}}^{varcone30}/p_{\mathrm{T}} < 0.06$. The muon isolation working points and the muon selections are shown in Table 4.2 and Table 4.3, respectively.

4.2 Jets

The jets are reconstructed by the anti-k_T algorithm [39] with a distance parameter of R=0.4 using a cluster of energy deposit in the calorimeters. The jets are required to have $p_{\mathrm{T}} > 25$ GeV within the range $|\eta| < 2.5$. Jets with $p_{\mathrm{T}} < 60$ GeV are required to have Jet-Vertex-Tagger (JVT) [40] less than 0.59, where JVT is a technique used to separate jets from pile-up events. Furthermore, jets originating from b-hadrons are identified using the MV2c10 b-tagging algorithm [41, 42] with an efficiency of 85%. In addition, b-tagged jets are required to have p_{T} >20 GeV. The b-tagging algorithm rejects the light-quarks and gluon jet by a factor of 34 as estimated by simulated $t\bar{t}$ sample and validated with data [40]. The rejection for selecting jets originating from the c-hadrons is about 3 [40].

4.3 Overlap Removal

If candidate jets are inside a cone of size ΔR =0.2 around an electron or around a muon candidate for any surviving jets with less than three associated tracks, they are discarded. If there is any surviving jet with three or more associated tracks inside ΔR <0.4 of a muon, or 0.2 < ΔR <0.4 of an electron, the corresponding lepton candidate is discarded. If a combined muon shares an ID track with an electrons, the electron is removed. If a calorimeter-tagged muon shares an ID track with an electron, then the muon is removed.

4.4 Missing Transverse Momentum

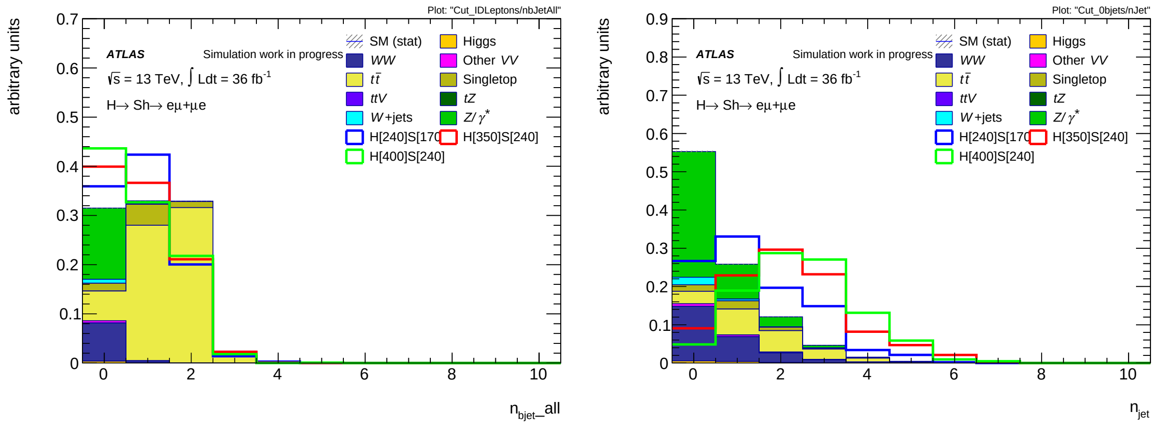
The missing transverse momentum is calculated as the negative vector sum of the transverse momenta of the tracks associated with electrons, jets and muons, as well as the tracks in the ID that are not associated with any other component. This is denoted as $\vec{E}_{\mathrm{T}}^{\mathrm{miss,track}}$ and the module of the vector as $E_{\mathrm{T}}^{\mathrm{miss,track}}$. Via strict criteria on longitudinal and transverse impact parameters, the tracks considered must have a transverse momentum above 400 MeV and be connected to the hard scatter vertex. Pseudorapidity coverage of $E_{\mathrm{T}}^{\mathrm{miss,track}}$ is limited to the tracking volume of $|\eta|$ <2.5, which is lower than calorimeter coverage of $|\eta|$ <4.9.

Table 4.4: Summary of the minimum lepton p_{T} requirements used at different levels of trigger selection. The symbols T, M and L stand for the Tight, Medium and Loose electron identification requirement, respectively. The letter i indicates extra isolation requirement.

Electron trigger	**Muon trigger**	**dilepton trigger**
	2015 dataset	
HLT-e24-lhmedium-L1EM20VH HLT-e60-lhmedium	HLT-mu20-iloose-L1MU15 HLT-mu50	HLT-e17-lhloose-mu14 HLT-e7-lhmedium-mu24
HLT-e120-lhloose		
	2016 dataset	
HLT-e26-lhtight-nod0-ivarloose HLT-e60-lhmedium-nod0	HLT-mu26-ivarmedium HLT-mu50	HLT-e17-lhloose-nod0-mu14 HLT-e7-lhmedium-nod0-mu24
HLT-e140-lhloose-nod0		

4.5 Pre-selection

A set of pre-selection requirements are applied to guarantee the quality of the data used for the analysis. The GRL contains data collected under good operation conditions. The GRL used in this analysis are described in Section 3.1. The event is required to have two oppositely charged leptons with different flavour and no additional lepton with p_{T} >10 GeV. Both leptons must satisfy the quality criteria discussed in Section 4.1. Low-mass resonances decaying to leptons, in which $\tau\tau$ decays may contribute, constitute a significant background and are rejected by requiring $m_{\ell\ell}$ >10 GeV. It is required that as least one of the lepton candidates must be matched to the triggering-lepton. The single-lepton triggers and the dilepton triggers are used in this analysis for electrons and muons in each data period as listed in Table 4.4. The single-lepton trigger efficiencies on the plateau is about 70% for muons within $|\eta| < 1.05$, 90% for muons in the range $1.05 < |\eta| < 2.4$, and above 90% for electrons within $|\eta| < 2.47$ [37, 38].



(a) (b)

Figure 4.2: Shape comparison of signal and background both normalized to unity, after applying the preselection mentioned in Section 4.5: (a) number of b-tagged jets and (b) number of jets distributions (with b-tagged jet veto applied).

4.6 Signal Region Definition

A phase space with high signal sensitivity and low background contamination determines the signal region (SR), which is where measurements are performed. This study requires two oppositely charged leptons with different flavour ($e^{\pm}\mu^{\mp}$). This study requires leading lepton (ℓ_0) to satisfy $p_{\mathrm{T}} > 27$ GeV, subleading lepton (ℓ_1) $p_{\mathrm{T}} > 15$ GeV and veto events with additional $p_{\mathrm{T}} > 10$ GeV leptons. The leading lepton p_{T} and subleading lepton p_{T} are the leptons with the highest p_{T} and second highest p_{T}, respectively. To minimize the contribution from the top quark processes, events are required to have no b-tagged jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. The jet p_{T} requirements are optimized to maximize the signal sensitivity. Figure 4.2 (a) shows the b-tagged jet multiplicity after applying the preselection mentioned in Section 4.5. The b-tagged jet multiplicity shows shape comparison between signal and the background, both normalized to unity. The first bin is selected as the SR since the ratio of signal over background is the highest. The second and third bins are dominated by the top quark background.

After applying the preselection and b-tagged jet veto, the signal over background ratio is higher in the region with at least two jets as shown in Figure 4.2 (b).

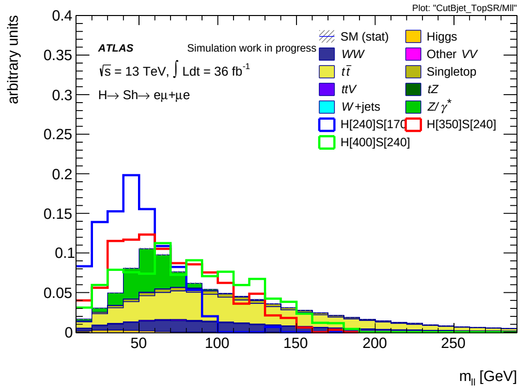
With that in mind, events containing exactly one jet are worth studying as the ratio of the lowest signal mass point (m_H= 240 GeV and m_S= 170 GeV) over the background is non-negligible. In addition, the 1-jet bin contains a significant amount of background from $t\bar{t}$, WW and Z/γ^* processes. However, the study focuses on the region with at least two jets. This region simplifies the analysis, as it is dominated mainly by the $t\bar{t}$ process.

The summary of the event selection in the SR is shown in Table 4.5. Clearly the dominant background is the top quark ($t\bar{t}$ and Wt), which contributes about 60% to the total background. Another important contribution in the total background are the WW and Z/γ^* processes, with a purity of 19% and 18% respectively. Table 4.6 shows the event for each process estimated by the MC simulation in the SR.

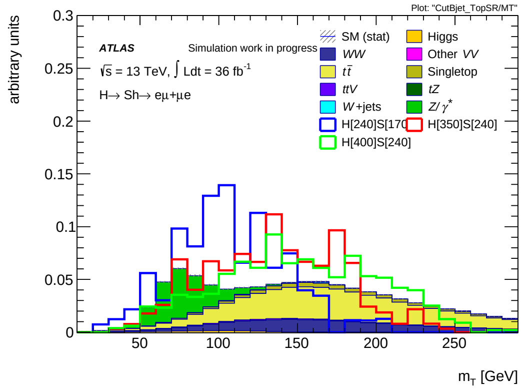
The distributions shown in Figure 4.3 illustrate the MC simulation in the SR for shape comparison of SM background processes and signal. These distributions can be used to optimize the signal sensitivity relative to background contamination. Figure 4.3 (a) shows that the majority of the signal is located in the $m_{\ell\ell}$ region below 150 GeV. Selecting events with $m_{\ell\ell}$ less than 150 GeV can therefore reduce both the top quark and the WW background processes. Figure 4.3 (b) shows the transverse mass between the dilepton system and the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ (m_{T}) distribution. The m_{T} is defined as:

$$\sqrt{2p_{\mathrm{T}}^2 E_{\mathrm{T}}^{\mathrm{miss,track}}(1 - cos(\Delta(\ell\ell, E_{\mathrm{T}}^{\mathrm{miss,track}})))}. \tag{4.1}$$

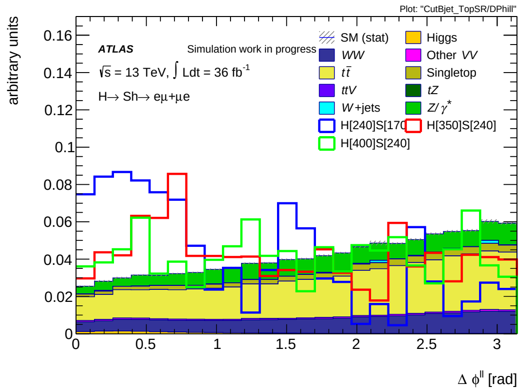
The m_{T} distribution shows that most of the signal is located in the central region



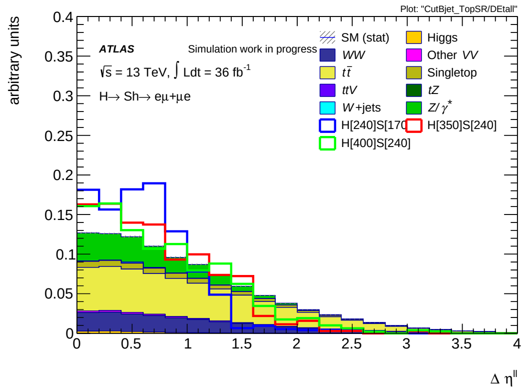
(a)



(b)



(c)



(d)

Figure 4.3: Shape comparison of signal and background, both normalized to unity in the SR: (a) dilepton invariant mass, (b) transverse mass , (c) angular separation between the two leptons and (d) difference in pseudorapidity between the two leptons.

and decreases towards the edges of the spectrum. The distribution of the angular separation between the two leptons ($\Delta\phi^{\ell\ell}$) is shown in Figure 4.3 (c). The $\Delta\phi^{\ell\ell}$ distribution observes a higher ratio of signal to background for lower values of $\Delta\phi^{\ell\ell}$. The distribution of the difference in pseudorapidity between the two leptons ($\Delta\eta^{\ell\ell}$) is shown in Figure 4.3 (d).

Table 4.5: Event selection in the SR.

Observable	SR
Leading lepton p_{T}	$>$ 27 GeV
Subleading lepton p_{T}	$>$ 15 GeV
η^{ℓ}	$\|\eta^{e}\| < 2.47$ (excluding $1.37 < \|\eta^{e}\| < 1.52$), $\|\eta^{\mu}\| < 2.5$
No. of additional leptons ($p_{\mathrm{T}} > 10$ GeV)	0
Number of jets	≥ 2
Number of b-tagged jets	$= 0$

Table 4.6: Number of expected events from MC simulation in the SR. Uncertainties are statistical.

Process	SR
SM Higgs	297.94 ± 1.56
WW	4816.49 ± 13.32
Other VV	561.33 ± 4.34
$t\bar{t}$	$13\,191.14 \pm 23.40$
Wt	1678.65 ± 16.17
ttV	29.02 ± 0.46
tZ	1.72 ± 0.04
W+jets	135.49 ± 35.63
Z/γ^{*}	4723.98 ± 57.17
$V\gamma$	292.37 ± 15.65
Total Background	$25\,728.13 \pm 76.10$

Chapter 5

Top Quark Background Studies

5.1 MC Truth Studies

In this analysis the top quark is the dominant background and it is essential to have a good understanding of this process. This section addresses the MC truth information on the top background in the various b-tagged jet bins at the reconstruction level for $t\bar{t}$ and Wt MC samples. In essence, "particle level" is the information before the events reach the detector simulation, also known as "MC particle level" or "MC truth level". Reconstruction level is the last phase information after the reconstruction algorithms have been applied over the data from the simulated detector.

In this section, b-hadrons at the level of truth and reconstruction are referred to as truth b-jets and b-tagged jets, respectively. The results in Section 4.6 shows that the SR is defined in the 0 b-tagged jet bin which is dominated by the top quark background. The 1 b-tagged jet and 2 b-tagged jets bins are top quark enriched phase space. Therefore these three b-tagged jet bins are used in this study to explore the top quark background. A variable called "HadronConeExclTruthLa-

Table 5.1: Average percentage of truth b-jets for $t\bar{t}$ and Wt samples.

	truth b-jets composition		
	0 b-jet bin (%)	1 b-jet bin (%)	2 b-jet bin (%)
Leading jet	43	62	75
Subleading jet	38	54	74

belID" is used on each b-tagged jet bin to determine if the leading and subleading jets are originated from a truth b-jets. This variable returns three particle data group (PDG) values [43]: 0, 4 and 5 corresponding to jets originating from light quarks (u and d), c-quarks and b-quarks, respectively.

The distributions in Figure 5.1 and Figure 5.2 indicate the truth jet origin for the leading and subleading jets for $t\bar{t}$ and Wt samples, respectively. Table 5.1 summarises the average rate of the leading and subleading jet to be originated from a truth b-jet in each b-tagged jet bin for these samples. In the 0 b-tagged jet bin, 43% and 38% of the leading jets and subleading jets are truth b-jets, respectively. In the 1 b-tagged jet bin, 62% and 54% of the leading jets and subleading jets are truth b-jets. In the 2 b-tagged jet bin, 75% and 74% of the leading jets and subleading jet are truth b-jets. The difference in the 0 b-tagged jet bin and 1 b-tagged jet bin relative to the 2 b-tagged jet bin is about 30% and 20%, respectively. It is therefore clear that the origin of truth b-jets for leading and subleading jets in different b-tagged jet bins are comparable. The results show that one can study the top quark background in the 1 b-tagged jet and 2 b-tagged jets bins.

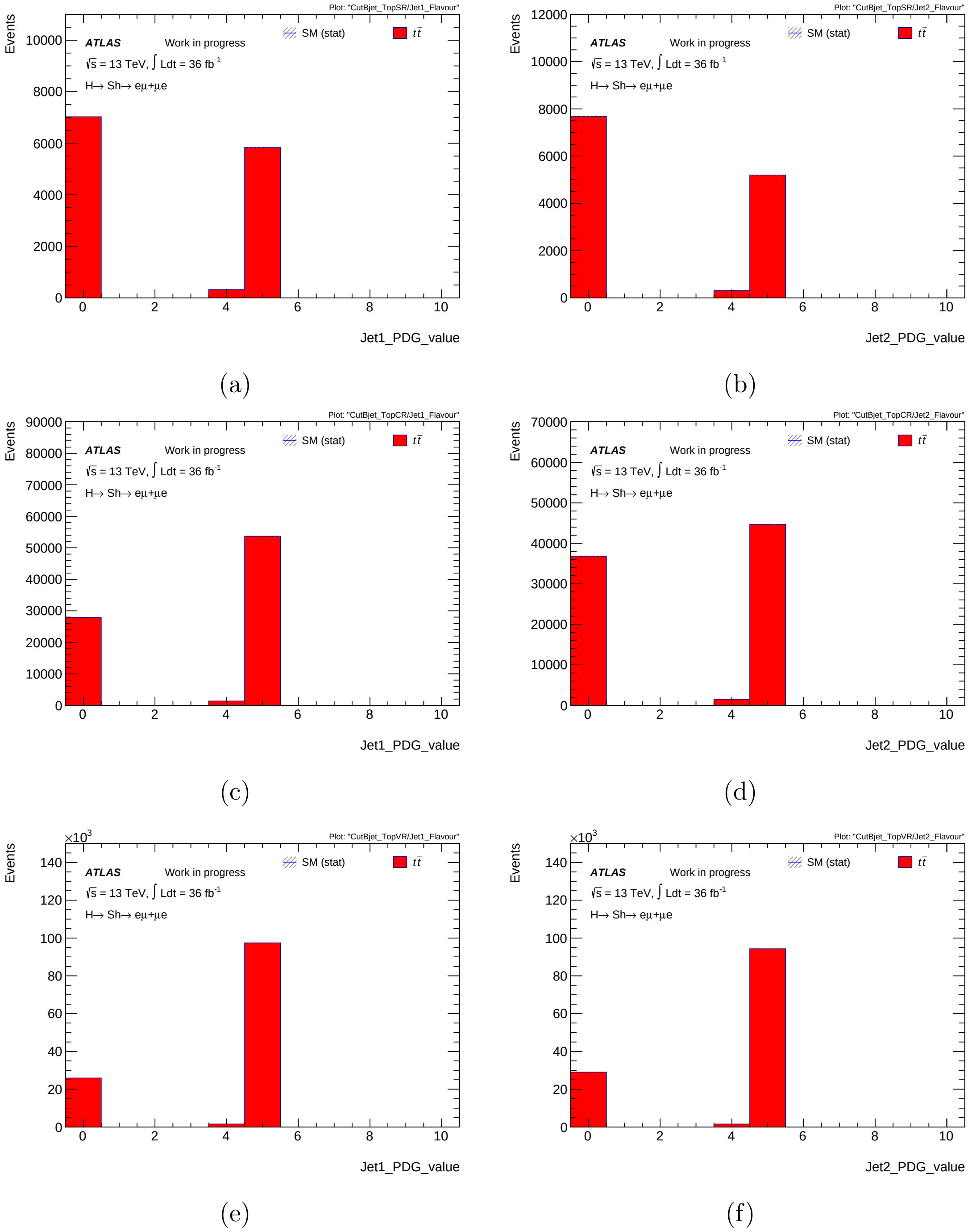


Figure 5.1: b-tagged jets composition for $t\bar{t}$ samples: (a) and (b) 0 b-tagged jets, (c) and (d) 1 b-tagged jets, (e) and (f) 2 b-tagged jets. The distributions show the leading jet on the left and the subleading jet on the right. Uncertainties are statistical.

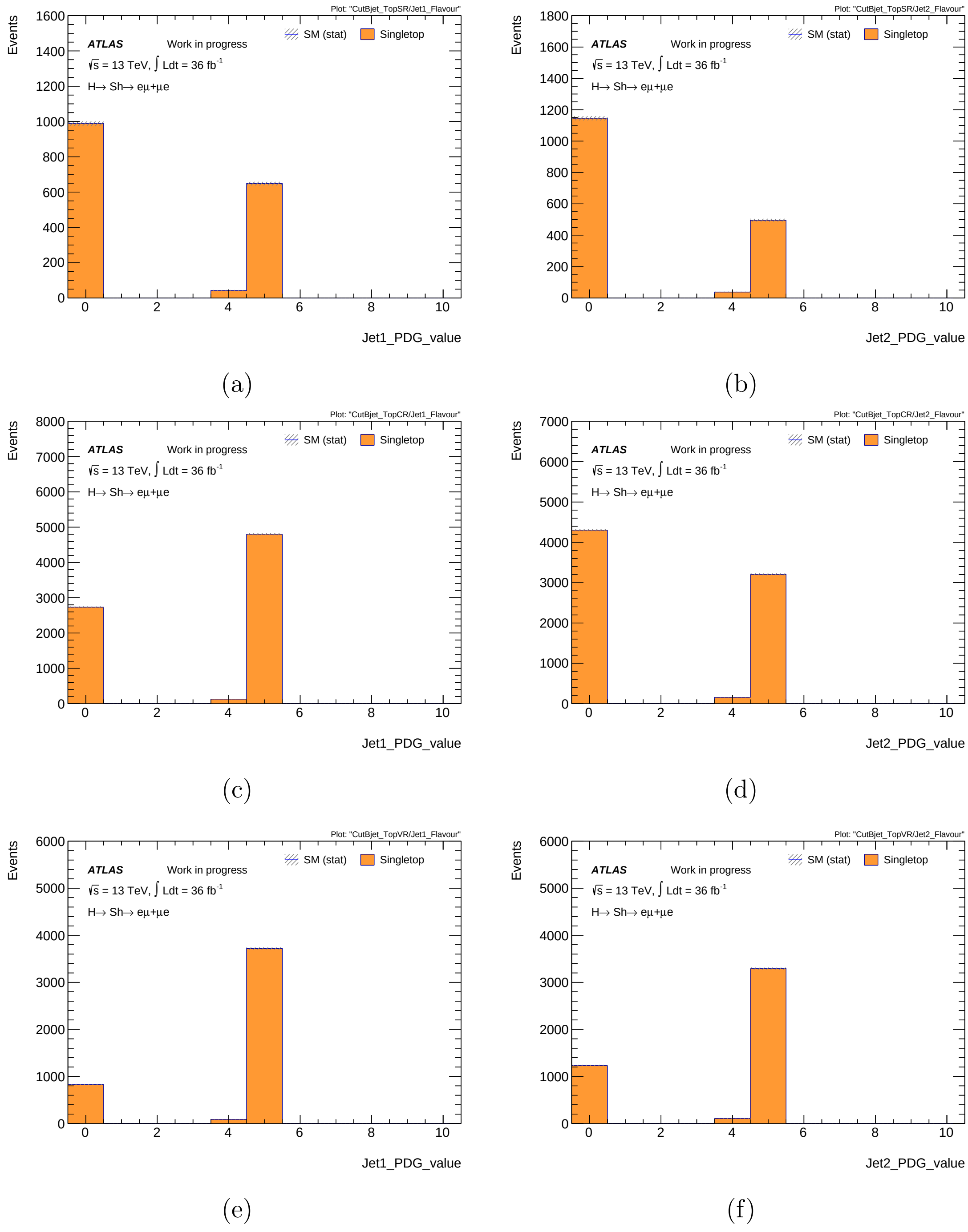


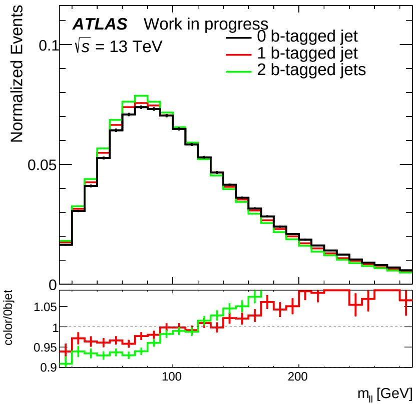
Figure 5.2: Jets composition for Wt samples: (a) and (b) 0 b-tagged jets, (c) and (d) 1 b-tagged jets, (e) and (f) 2 b-tagged jets. The distributions show the leading jet on the left and the subleading jet on the right. Uncertainties are statistical.

5.2 MC Shape Comparison in Different b-tagged Regions

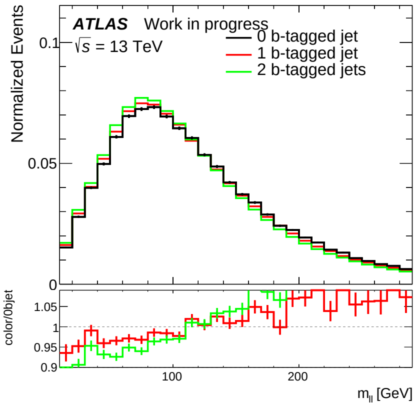
For 0 b-tagged jet bin, 1 b-tagged jet bin and 2 b-tagged jets bin, the shape of the leptonic kinematics for $t\bar{t}$ and Wt events are analysed. It is essential to verify the consistency of the shape of the $m_{\ell\ell}$ distribution in various b-tagged jet bins since the signal is expected at low mass region. The distributions for $e\mu$, μe and $e\mu + \mu e$ channel are provided in this section. The $e\mu$ channel means that the leading lepton is the electron and the subleading lepton is the muon. The μe channel is the opposite. The $e\mu+\mu e$ channel is the combination of $e\mu$ and μe channels.

For $t\bar{t}$ process the $m_{\ell\ell}$ distribution shape is very similar in the $e\mu$, μe and $e\mu + \mu e$ channels. Nevertheless, in the region of less than 100 GeV, the average difference is about 2.3% and 6.7% for 1 b-tagged jet and 2 b-tagged jet phase space relative to the 0 b-tagged jet phase space. Also in the region above 100 GeV, the average difference is about 7% for 1 b-tagged jet and 2 b-tagged jet phase space relative to the 0 b-tagged jet phase space. These features are observed in the $e\mu$, μe and $e\mu + \mu e$ channel, as shown in Figure 5.3.

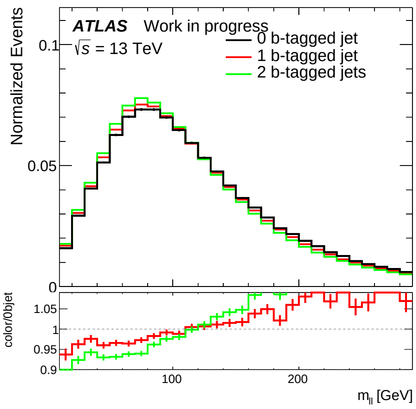
Figure 5.4 shows the shape comparison for the $m_{\ell\ell}$ distributions in the 0 b-tagged jet, 1 b-tagged jet and 2 b-tagged jet phase space for Wt process. The $m_{\ell\ell}$ distributions observes a consistent shape but there is a sligthly difference which is about 2.5% for 1 b-tagged jet and 2 b-tagged jet phase space relative to the 0 b-tagged jet phase space. This is observed throughout the $m_{\ell\ell}$ distribution for $e\mu$, μe and $e\mu + \mu e$ channel as shown in Figure 5.4. The results show that 0 b-tagged jet, 1 b-tagged jet and 2 b-tagged jets phase space are kinematically close since the shape of the leptonic quantities are very similar.



(a)

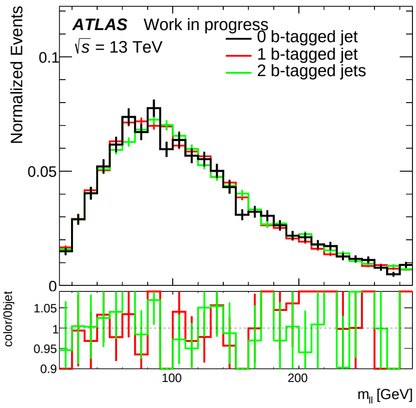


(b)

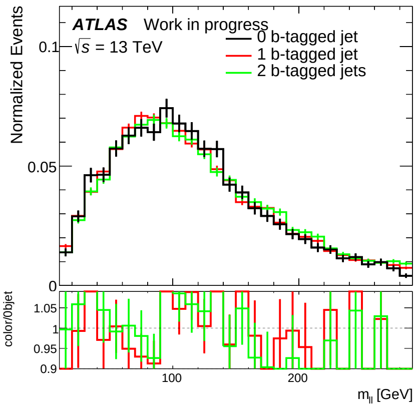


(c)

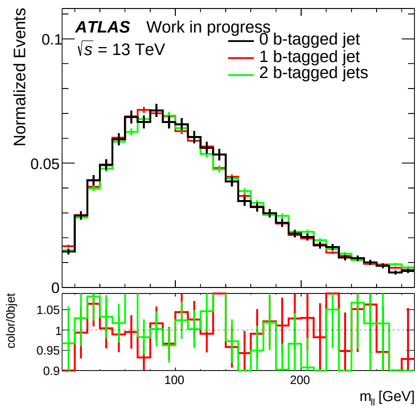
Figure 5.3: $t\bar{t}$ MC simulation comparison shape for $m_{\ell\ell}$ for 0 b-tagged bin (black line), 1 b-tagged bin (red line) and 2 b-tagged bin (green line) in the (a) $e\mu$ channel, (b) μe channel and (c) $e\mu + \mu e$ channel.



(a)



(b)



(c)

Figure 5.4: Wt MC simulation comparison shape for $m_{\ell\ell}$ for 0 b-tagged jet bin (black line), 1 b-tagged jet bin (red line) and 2 b-tagged jet bin (green line) in the (a) $e\mu$ channel, (b) μe channel and (c) $e\mu + \mu e$ channel.

5.3 Top Quark Production

After applying the requirements presented in Section 4.6, the SM processes in the SR are expected to be top-quark, WW production, Z/γ^*. Other relevant background contributions arise from W+jets (fakes), Z+jets production, $V\gamma$, other VV, $t\bar{t}V$ and tZ where V stands for both W and Z bosons.

This analysis relies on MC simulation of SM top quark processes as $t\bar{t}$ pair and Wt. Such processes are integrated as a single process known as the top quark background. The top quark process is dominated by the $t\bar{t}$ events due to its large cross section and the contribution from the single-top (Wt) is relatively small.

The results of the MC truth study in Section 5.1 showed that the composition of the origin of the b-hadrons in the leading and subleading jets is comparable in three b-tagged jet bins: 0 b-tagged jet bin, 1 b-tagged jet bin and 2 b-tagged jet bin. The $m_{\ell\ell}$ study in Section 5.2 shows similar shape in the same b-tagged bins with an average difference of about 5%. In view of this, this provides a way of designing two top quark regions: top control region (top CR) and top validation region (top VR). Top CR is constructed in order to derive the normalization factor and extrapolate the top-quark contribution to the SR. In order to test the modeling of the top-quark processes a top VR is defined. Top CR and top VR are required to be orthogonal to the SR selection and signal depleted. The top CR and top VR are discussed in Section 5.4 and Section 5.5.

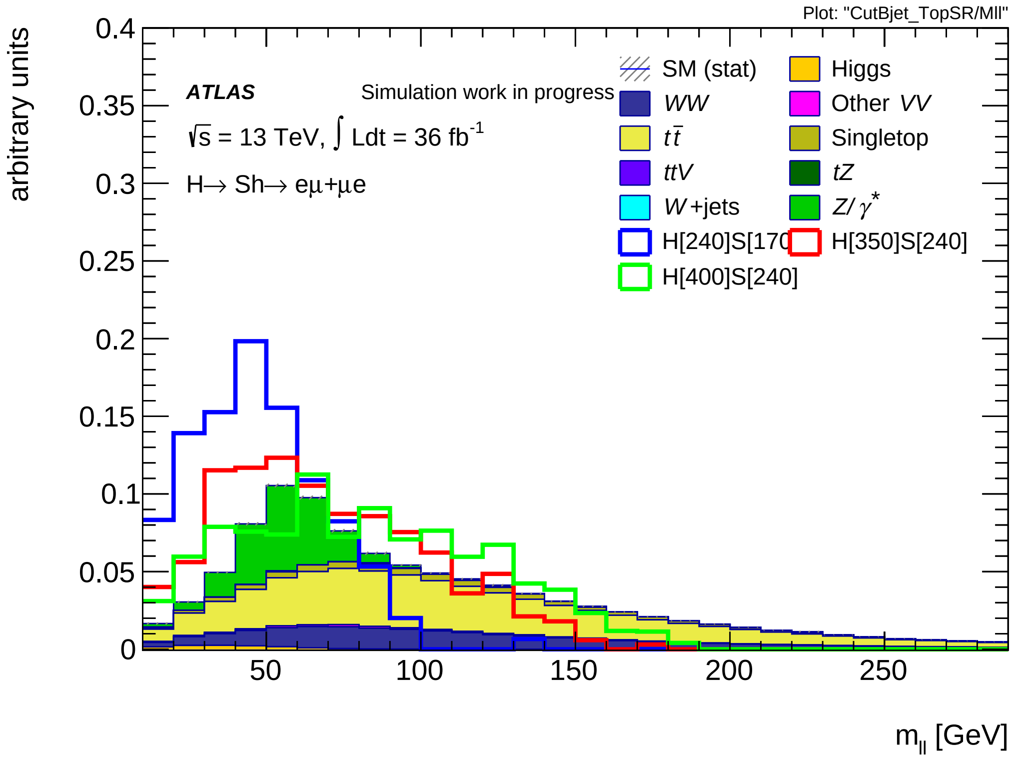


Figure 5.5: The $m_{\ell\ell}$ distribution for shape comparison of signal and background both normalized to unity in the SR.

5.4 Top Control Region

The top CR is selected by applying the same selection as in the SR with an additional requirement on the $m_{\ell\ell}$ to be above 150 GeV in order to achieve a signal depleted region, as shown in Figure 5.5. The definition of the top CR is summarised in Table 5.2. These requirements significantly reduce about 98% of signal contamination in the top CR as shown in Table 5.3.

Table 5.2: Analysis selection for top CR.

Observable	top CR
Leading lepton p_{T}	> 27 GeV
Subleading lepton p_{T}	> 15 GeV
Number of jets	≥ 2
Number of b-tagged jets	$=0$
$m_{\ell\ell}$	> 150 GeV

Table 5.3: Number of data events compared to the number of background events predicted by the MC simulation. Uncertainties are statistical.

	SR	top CR
$H[240]S[170]$	9.67 ± 0.75	0.00 ± 0.00
$H[350]S[240]$	71.99 ± 3.89	0.64 ± 0.37
$H[400]S[240]$	116.57 ± 5.20	2.26 ± 1.22
SM Higgs	297.97 ± 1.56	3.10 ± 0.08
WW	4816.49 ± 13.32	1372.68 ± 7.00
Other VV	561.33 ± 4.34	137.64 ± 2.41
$t\bar{t}$	$13\,191.14 \pm 23.40$	3719.38 ± 12.51
Singletop	1678.65 ± 16.17	490.82 ± 8.83
ttV	29.02 ± 0.46	9.98 ± 0.27
tZ	1.72 ± 0.04	0.57 ± 0.02
W+jets	135.49 ± 35.63	40.43 ± 20.30
Z/γ^*	4723.98 ± 57.17	18.96 ± 10.78
$V\gamma$	292.37 ± 15.65	25.88 ± 6.15
Total Background	$25\,728.13 \pm 76.10$	5819.44 ± 29.25
Data	-	5699
Data/MC	-	0.98 ± 0.01

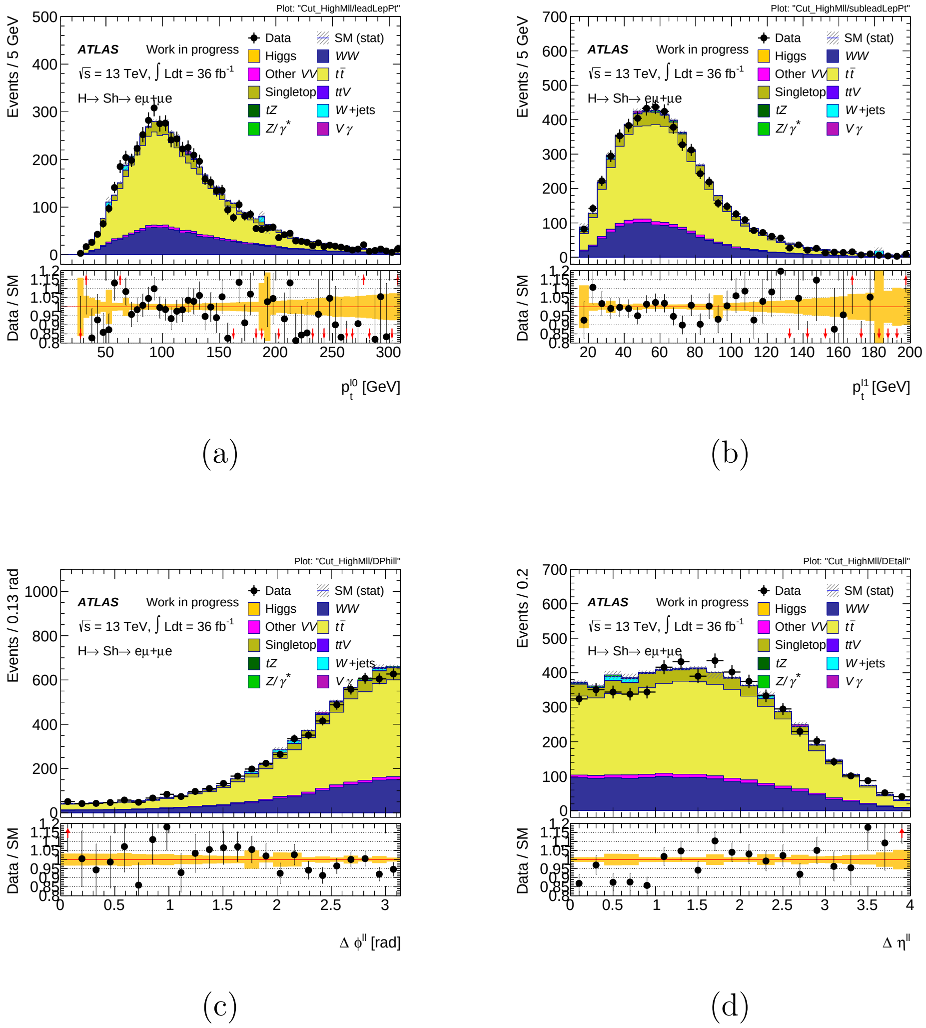


Figure 5.6: Leptonic kinematics of data and MC simulation in the top CR after applying the requirements shown in Table 5.2: (a) leading lepton p_{T}, (b) subleading lepton p_{T}, (c) angular separation between the two leptons and (d) difference in pseudorapidity between the two leptons. Uncertainties are statistical.

In this region, the agreement between data and MC is reasonable. The ratio of data over MC is 0.98 ± 0.01 for $e\mu + \mu e$ channel. The top quark processes contributes 72% to the total background. Other background contributions arise from the non-resonant WW which is about 25%, W+jets, other VV , $Z/\gamma*$ processes add up to 3% to the total background. The top CR distributions for lepton kinematics are shown in Figure 5.6. In the p_{T}^{l0} distribution, the average difference between data and MC is about 5% across the spectrum. The p_{T}^{l1} distribution is well described by the MC and the average difference between data and MC is about 2%.

The $\Delta\phi^{\ell\ell}$ and $\Delta\eta^{\ell\ell}$ distributions as shown in Figure 5.6 (c) and Figure 5.6 (d), respectively. Most of the jet kinematics distributions in Figure 5.7 have the same characteristics observed from the leptonic kinematics.

The ratio of data to MC for both $e\mu$ and μe channels separately is 0.98 ± 0.01, as shown in Appendix B.1 and B.2. Thus one may conclude that these results are consistent with the $e\mu$ and μe channel. Extra distributions are shown in Figure B.1 and Figure B.2 for $e\mu$ and μe channel separately: $\Delta\phi^{\ell\ell}$, $\Delta\eta^{\ell\ell}$, p_{T}^{jj}, $E_{\mathrm{T}}^{\mathrm{miss,track}}$, p_{T}^{j0} and p_{T}^{j1}.

For the top-quark processes ($t\bar{t}$ and Wt) in the top CR, a normalization factor (NF) may be extracted. This normalization factor can be extended to the top quark processes in the SR to assess the MC agreement with the data. This is then used in the SR as well as in the Top CR to scale these processes. The top quark NF is obtained by comparing the top quark background events from MC and the data events observed, after subtracting the expected contributions from processes other than the top quark (non-top quark),

$$NF(\text{top quark}) = \frac{N_{\text{data}} - N_{\text{MC}}^{\text{non-top quark}}}{N_{\text{MC}}^{\text{top quark}}}.$$

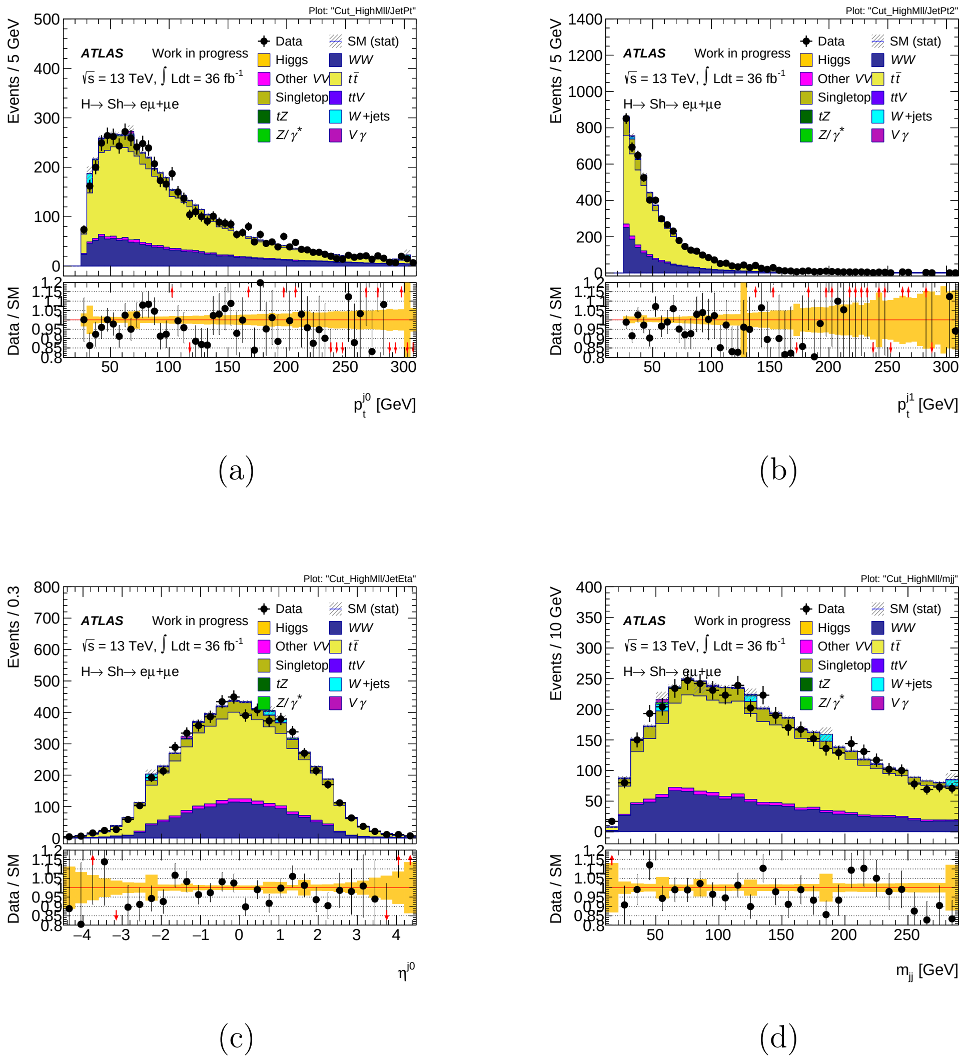


Figure 5.7: Jet kinematics of data and MC simulation in the top CR after applying requirements shown in Table 5.2: (a) leading jet p_{T}, (b) subleading jet p_{T}, (c) leading jet η and (d) dijet invariant mass (m_{jj}). Uncertainties are statistical.

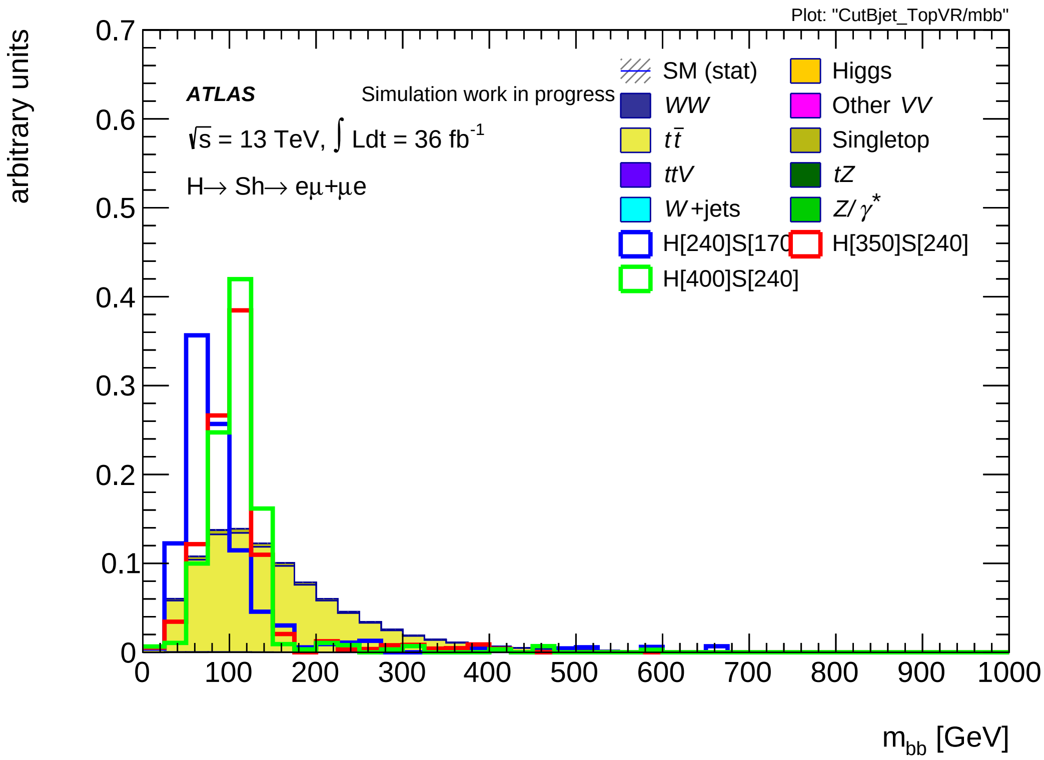


Figure 5.8: Comparison shape of signal and backgound, both normalized to unity for m_{bb} distributions.

5.5 Top Validation Region

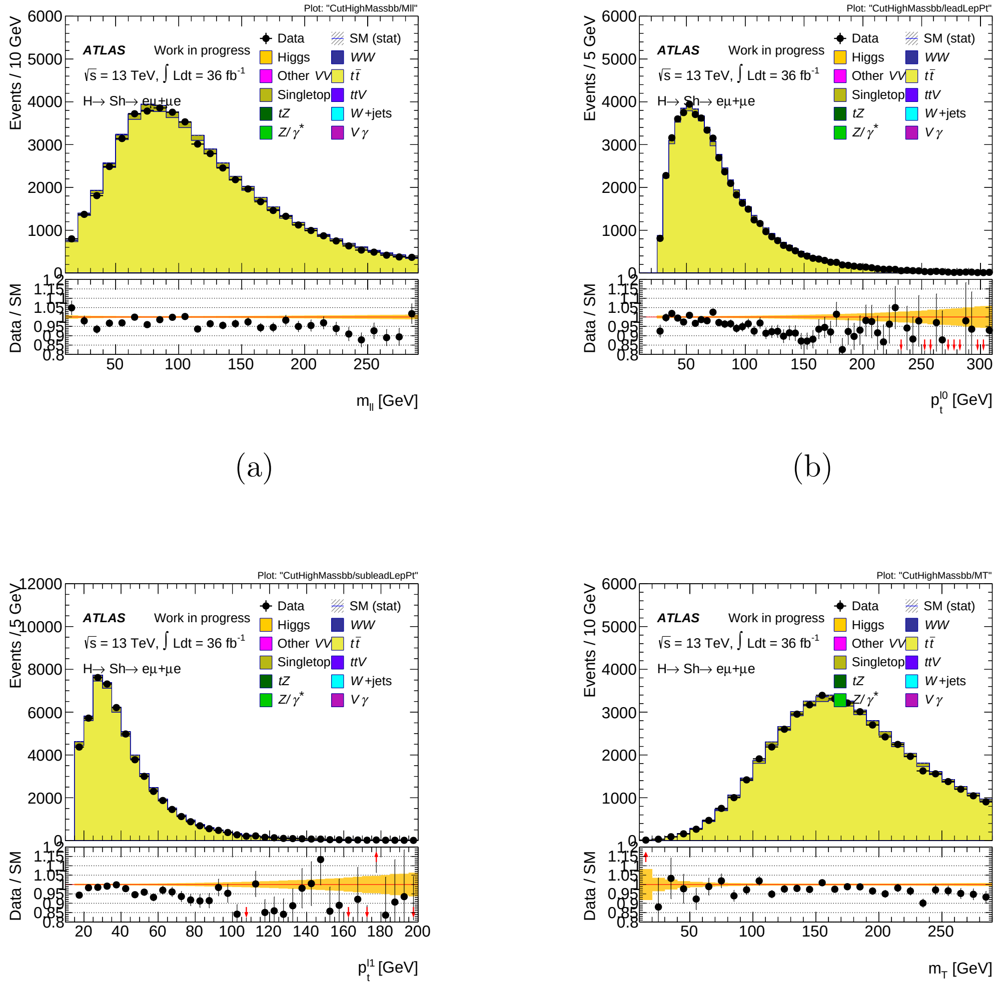
The top VR is selected using the same selection as in the SR but requiring events with exactly two b-tagged jets. Moreover, in order to reduce the signal contamination in the top VR the invariant mass of the two b-tagged jets (m_{bb}) is required to be above 150 GeV as shown in Figure 5.8. The definition of the top VR is summarised in Table 5.4. These requirements significantly reduce about 95% of signal contamination in the top VR as shown in Table 5.5.

Table 5.4: Analysis selection for top VR.

Observable	top VR
Leading lepton p_{T}	> 27 GeV
Sub-leading lepton p_{T}	> 15 GeV
Number of jets	≥2
Number of b-tagged jets	=2
m_{bb}	> 150 GeV

Table 5.5: Number of data events compared to the number of background events predicted by the MC simulation. Uncertainties are statistical.

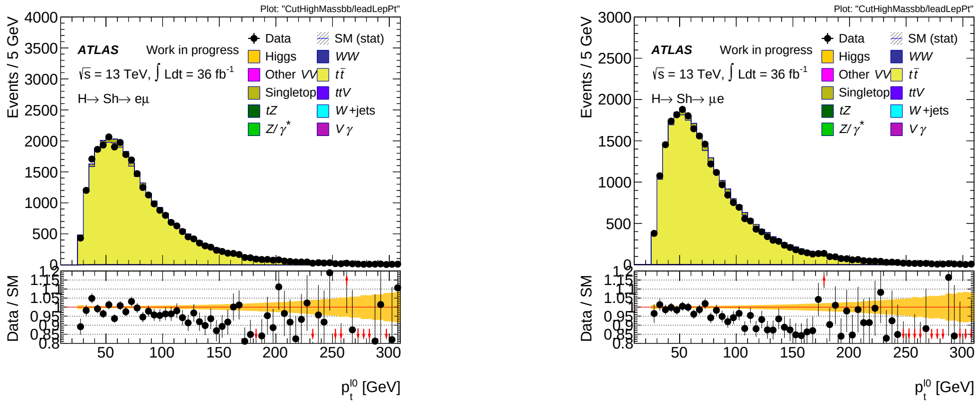
	SR	top VR
$H[240]S[170]$	9.67 ± 0.75	1.11 ± 0.25
$H[350]S[240]$	71.99 ± 3.89	4.19 ± 0.95
$H[400]S[240]$	116.57 ± 5.20	5.51 ± 1.10
SM Higgs	297.97 ± 1.56	26.70 ± 0.25
WW	4816.49 ± 13.32	33.90 ± 0.86
Other VV	561.33 ± 4.34	6.24 ± 0.32
$t\bar{t}$	13 191.14 ± 23.40	53 897.05 ± 46.25
Singletop	1678.65 ± 16.17	2134.64 ± 17.98
ttV	29.02 ± 0.46	86.64 ± 0.76
tZ	1.72 ± 0.04	1.93 ± 0.04
W+jets	35.49 ± 35.63	0.0 ± 0.00
Z/γ^*	723.98 ± 57.17	61.91 ± 3.53
$V\gamma$	292.37 ± 15.65	2.57 ± 1.67
Total Background	25 728.13 ± 76.10	56 285.41 ± 49.84
Data	–	54448
Data/MC	–	0.97 ± 0.01



(a) (b)

(c) (d)

Figure 5.9: Comparison of data and MC in the top CR after applying requirements shown in Table 5.4: (a) dilepton invariant mass, (b) leading lepton p_{T}, (c) subleading lepton p_{T} and (d) transverse mass. Uncertainties are statistical.

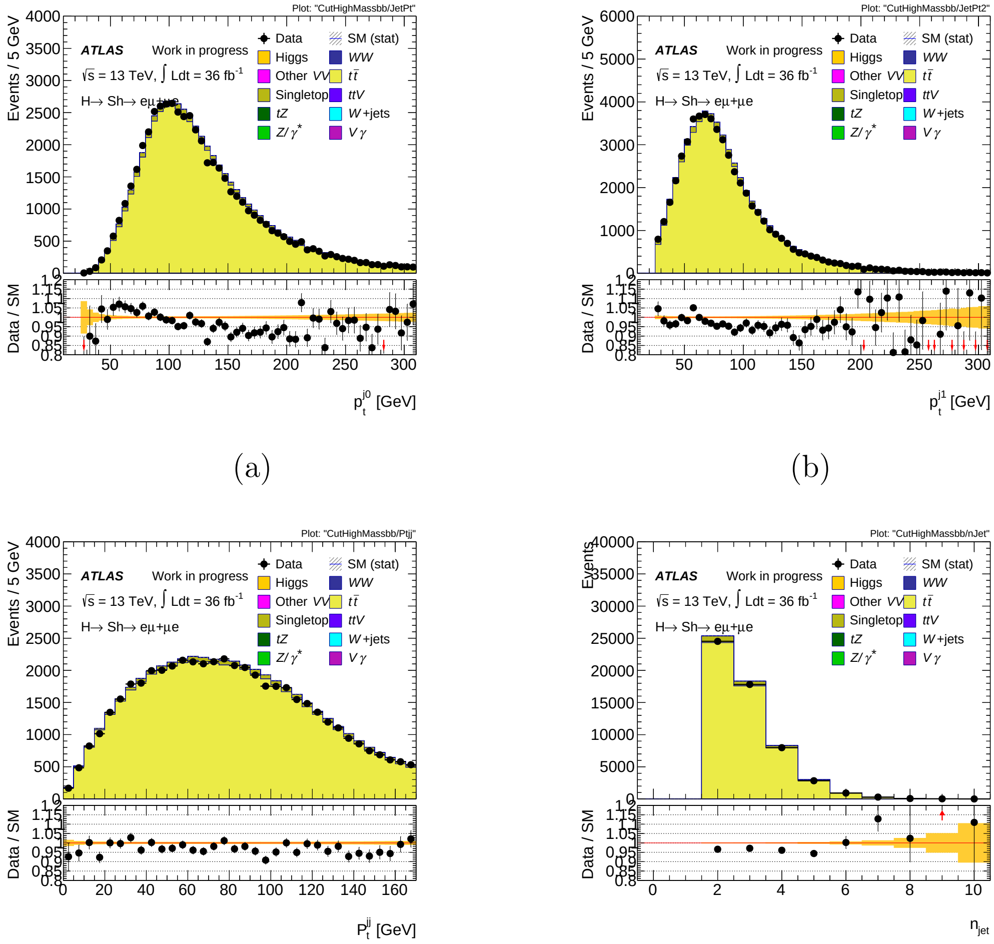


(a) (b)

Figure 5.10: The distribution of the leading lepton p_{T} for: (a) $e\mu$ channel and (b) μe channel. Uncertainties are statistical.

In this region the event yields for MC and data have a reasonable agreement. The ratio of data over MC is 0.97 ± 0.01 with top-quark contribution of 99% to the total background. Other non-top-quark background contributions add up to 1%. The leptonic kinematics for this region are shown in Figure 5.9. Figure 5.9 (a) shows the $m_{\ell\ell}$ distribution. In this distribution the MC overshoots the data by approximately 2%, this can be clearly seen in the ratio plot as most of the points are below 1. Figure 5.9 (b) shows the p_{T}^{l0} which observes a downwards slope in the ratio plot. This observation is also present in the p_{T}^{l0} distribution for $e\mu$ and μe channels as shown in Figure 5.10 (a) and Figure 5.10 (b), respectively. The same features are observed in the p_{T}^{l1} distribution in Figure 5.9 (c).

This slope is referred to as the mismodeling of the MC and related to the MC simulation not to a particular channel. In Section 5.6, the p_{T}^{l0} distribution will be used to derive the reweighting function to improve the MC prediction for $t\bar{t}$ process.



(a) (b) (c) (d)

Figure 5.11: Comparison of data and MC in the top CR after applying requirements shown in Table 5.4: (a) leading lepton p_{T}, (b) subleading lepton p_{T} (c) di-jet p_{T}, (d) jet multiplicity (n_{jet}). Uncertainties are statistical only.

The jet kinematics are shown in Figure 5.11. In the ratio plot the p_{T}^{j0} and p_{T}^{j1} distributions observe the same slope as in the leptonic kinematics. The distributions of p_{T}^{jj} and n_{jet} are not well modelled but the slope in the plot ratio is not observed. The same characteristics are observed for both $e\mu$ and μe channels as shown in Appendix B.3 and Appendix B.4. These appendices show the distributions of the $\Delta\phi^{\ell\ell}$, m_{T}, η^{l0}, p_{T}^{l1}, p_{T}^{j0} and p_{T}^{j1} for $e\mu$ and μe channels separately.

5.6 MC Correction

Correction on the MC simulation are introduced to match the data because MC can not accurately simulate the detector performance. From p_{T}^{l0} distribution in Figure 5.9 (b) one sees clearly a mismodeling between the data and the MC simulation in the top VR. Improving the predictions for the SM allows new physics models or discrepancies in experimental measurements to be identified. In order to improve the MC prediction, two types of MC corrections are implemented separately: the NLO EW corrections in the lepton kinematics and reweighting of top quark events using the p_{T}^{l0} distribution. The NLO electroweak (EW) correction is purely based on MC simulation and reweighting procedure follows a data-driven method. For most observables, the latter gives a very good description.

5.6.1 NLO EW Correction

The EW diagrams at NLO are not included in the $t\bar{t}$ MC simulation. EW corrections are often subject to non-negligible contributions up to 10%. EW corrections are calculated at NLO [44] and the results are obtained from the matrix-element provider Recola [45] in conjunction with the Collier library [46]. The correction factor is obtained by fitting the ratio NLO/LO for $t\bar{t}$ process as a function of the p_{T}^{l0}, as shown in Figure 5.12. The following selection cuts are applied:

- Two b-tagged jets; p_{T} >25 GeV, $|\eta|$ <2.5.
- Two charged leptons; p_{T} >20 GeV, $|\eta|$ <2.5.
- Missing transverse momentum: $E_{\mathrm{T}}^{\mathrm{miss,track}} > 20$ GeV.

The parameters for the fitted line $p0$ and $p1$ are the y-intercept and slope, respectively. These parameters $p0$ and $p1$ are calculated as 1.063 ± 0.009 and

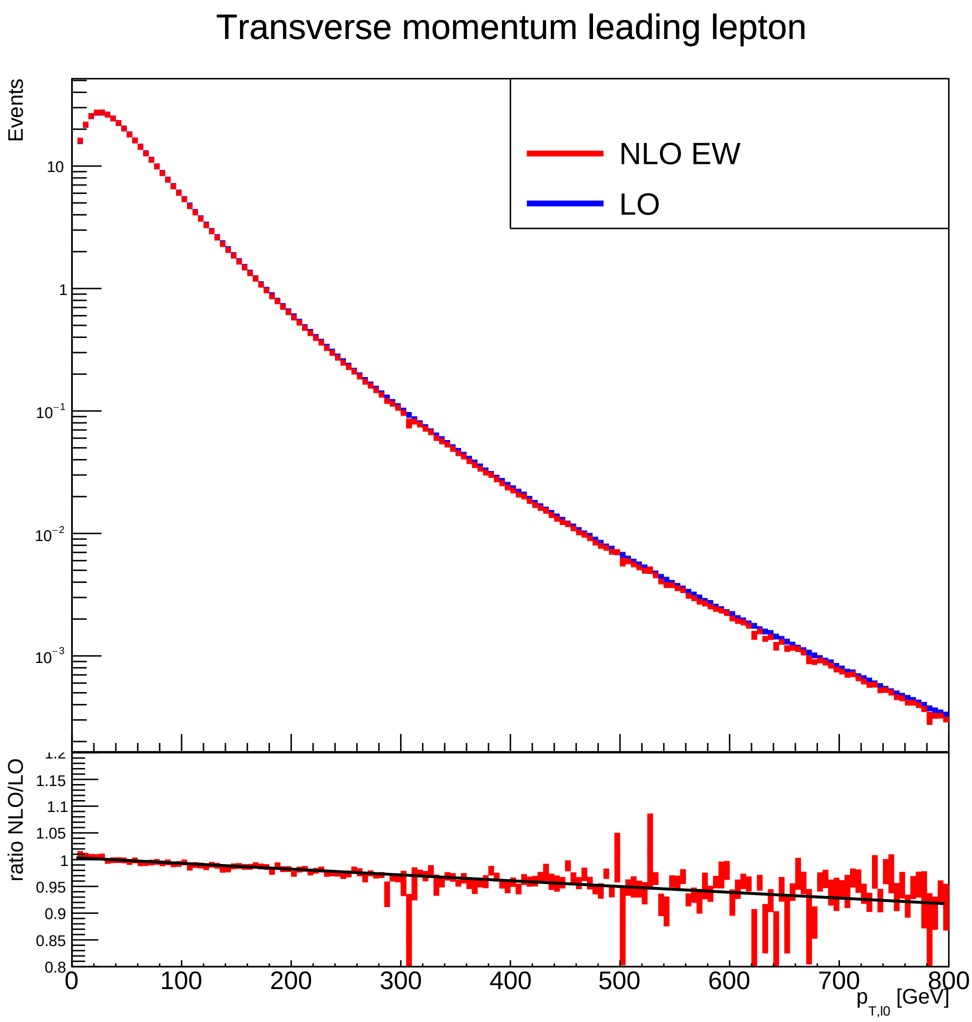


Figure 5.12: The distribution of the leading lepton p_{T} for NLO EW (red) and LO (blue) for $t\bar{t}$ sample.

-0.0007842 ± 0.0001, respectively. The function $p0 + p1 \times p_{\mathrm{T}}^{l0}$ is applied in the top VR. This improves the agreement between the data and MC simulation by about 2%, which is predominantly observed in the tails of the distributions. The top VR distributions after implementing the NLO EW correction are shown in Figure 5.13. Figure 5.13 (a) shows the $m_{\ell\ell}$ distribution after the NLO EW corrections. The agreement between data and MC in the $m_{\ell\ell}$ distribution has improved from 0.87 to 0.89 which is seen on the ratio plot around 250 GeV. The same features are observed in the m_{T} distribution, as shown in Figure 5.13 (c). Figure 5.13 (b) shows the p_{T}^{l0} distribution which shows no improvement and same applies to the p_{T}^{l1} distribution in Figure 5.13 (c). In the jet kinematics the p_{T}^{j0}, p_{T}^{j1}, p_{T}^{jj} and n_{jet} distributions show no improvement after the NLO EW corrections. The agreement between data and MC after NLO EW correction is shown in Table 5.6.

Alternatively, the correction has been extracted using the $m_{\ell\ell}$ distribution and agreement between data and MC improved from 0.97 ± 0.01 to 0.98 ± 0.01 which is seen on the tail of the distributions. The $m_{\ell\ell}$ NLO EW results are shown in Appendix B.1.3. With these results, one can conclude that the NLO EW correction does not cover most of the discrepancies between data and MC. Most of the discrepancies in the high p_{T}^{l0} is probably due to mismodeling, this will be fixed by hand in Section 5.6.2.

Table 5.6: Number of data events compared to the number of background events predicted by the MC simulation after NLO EW corrections. Uncertainties are statistical only.

	top CR	top VR
SM Higgs	3.10 ± 0.08	26.70 ± 0.25
WW	372.68 ± 7.00	33.90 ± 0.86
Other VV	137.64 ± 2.41	6.24 ± 0.32
$t\bar{t}$	3685.90 ± 12.40	$53\,640.18 \pm 46.03$
Singletop	485.68 ± 8.74	2151.93 ± 17.95
ttV	9.87 ± 0.27	86.03 ± 0.75
tZ	0.56 ± 0.02	1.91 ± 0.04
W+jets	40.43 ± 20.30	0.00 ± 0.00
Z/γ^*	8.96 ± 10.78	61.91 ± 3.53
$V\gamma$	25.88 ± 6.15	2.57 ± 1.67
Total Bkg	5780.70 ± 29.18	$56\,011.39 \pm 49.58$
Data	5699	54448
Data/MC	0.99 ± 0.01	0.97 ± 0.01

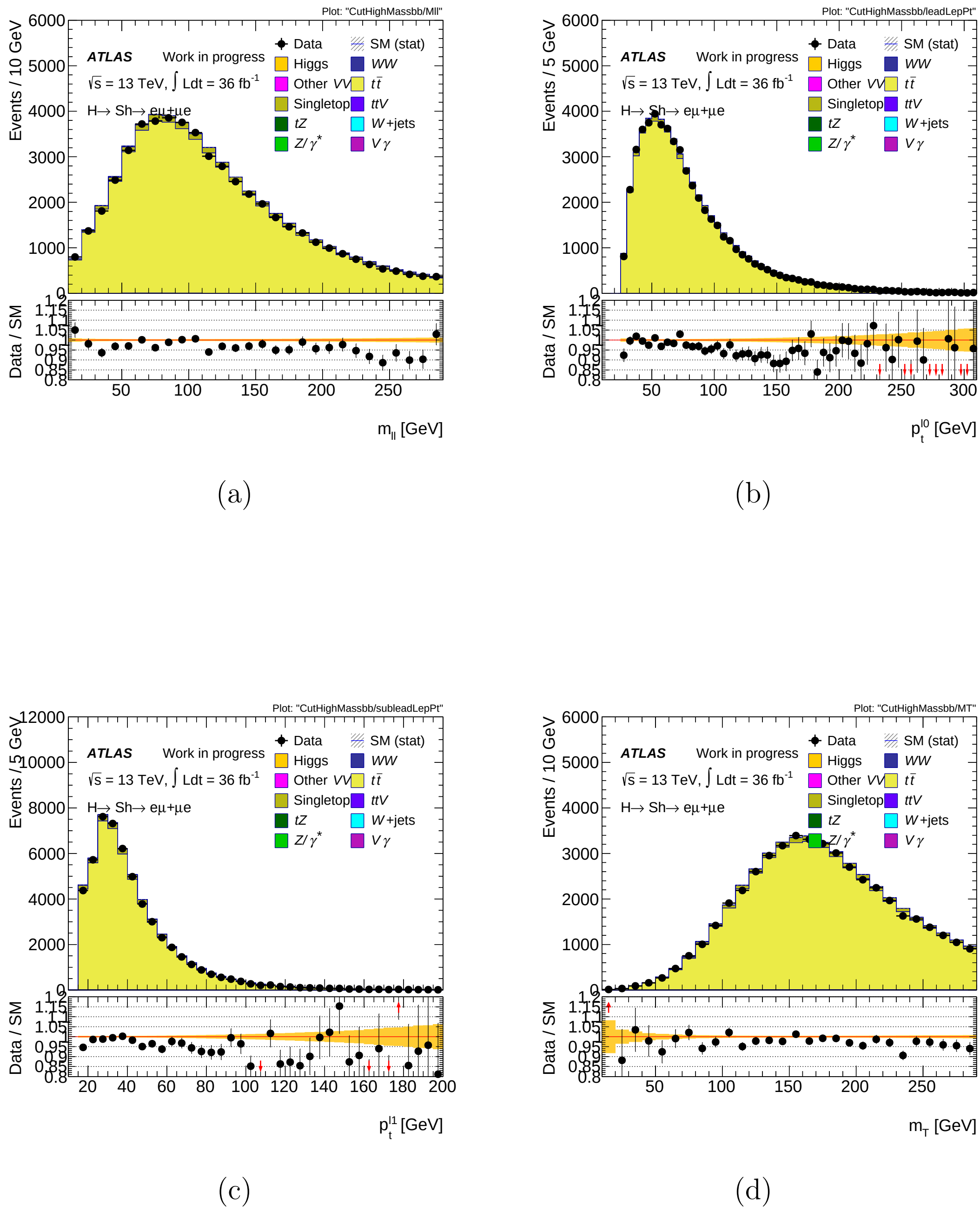


Figure 5.13: Same as Figure 5.9 except that the NLO EW correction has been applied. Uncertainties are statistical.

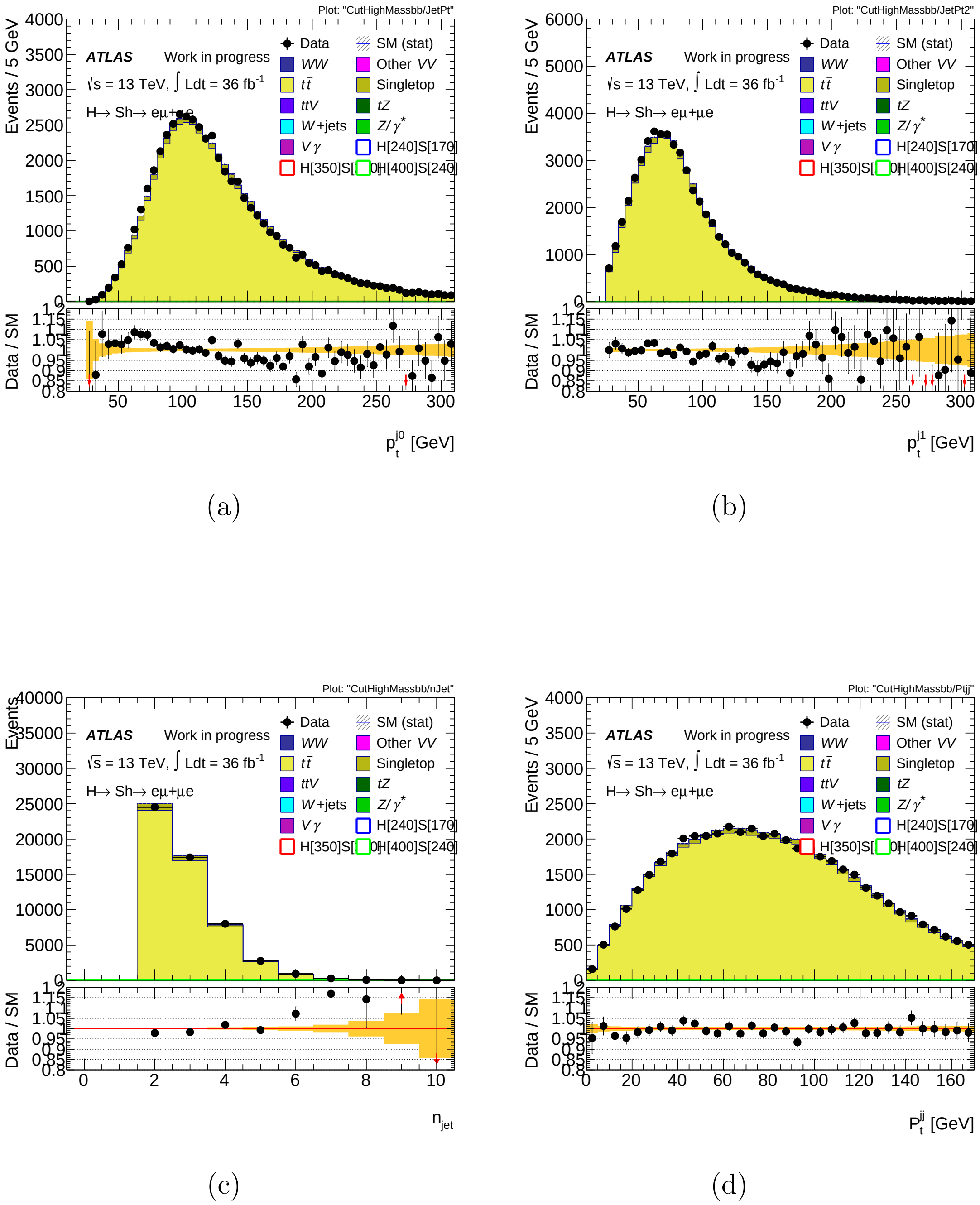


Figure 5.14: Same as Figure 5.11 except that the NLO EW correction has been applied. Uncertainties are statistical.

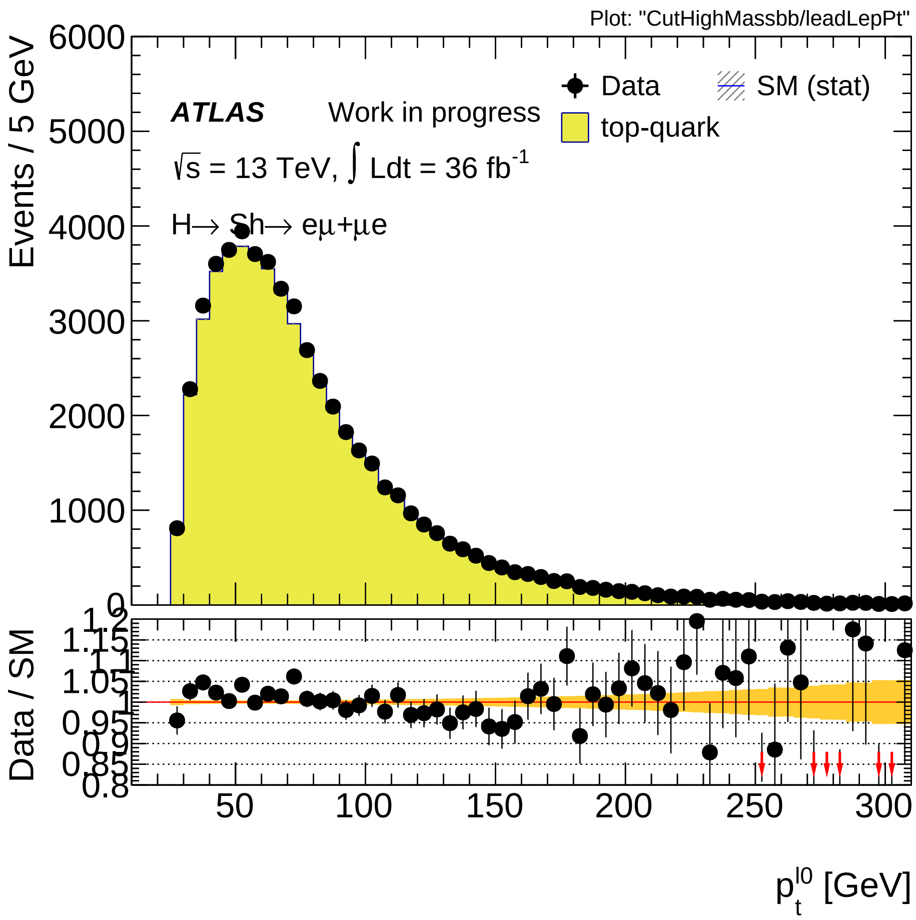


Figure 5.15: The distribution of the leading lepton p_{T} after subtracting non-top quark MC from data compared to top quark background.

5.6.2 MC p_{T}^{l0} Reweighting

In order to improve the modeling of the top MC simulation, the MC reweighting function is derived from the top VR (top quark enriched region), which will be applied in the top VR, top CR and in the SR. The reweigting function is in a form of a straight line, $p0 + p1 \times p_{\mathrm{T}}^{l0}$. The reweighting parameters are computed as follows:

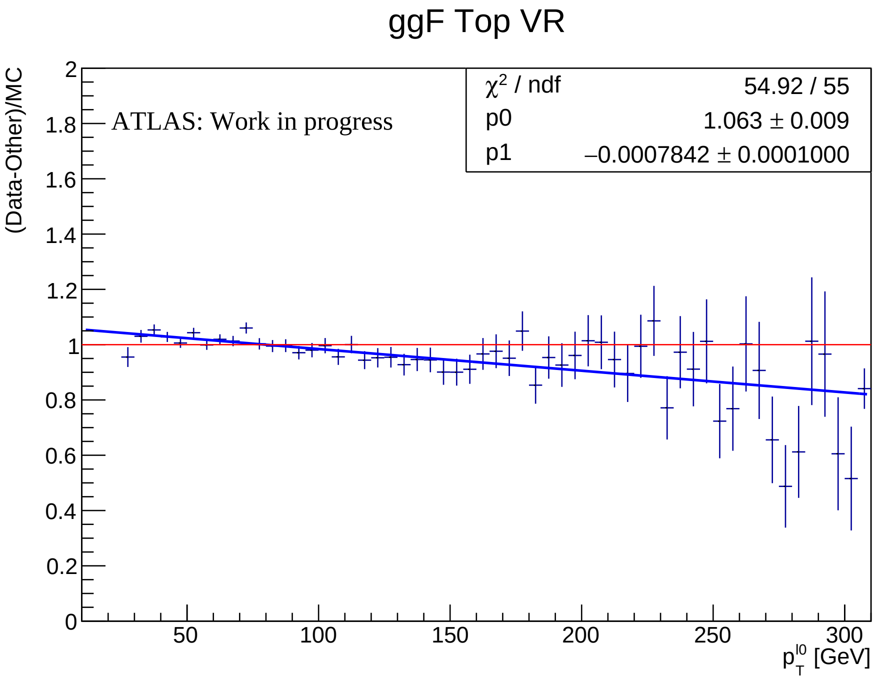


Figure 5.16: Fitted p_{T} correction parameters for the leading lepton p_{T}.

- Remove the non-top quark MC from data and divide by top quark MC: (data-non top MC)/top MC, as shown in Figure 5.15.
- Fit a line to scatter plot of ratio (data-non top MC)/top MC as a function of p_{T}^{l0}, as shown in Figure 5.16.
- The p0 and p1 parameters are found to be 1.063 $\pm$ 0.009 and -7.842$\times 10^{-4} \pm$ 0.001 $\times$ 10^{-1}, respectively.
- The weight is computed as $p0 + p1 \times p_{\mathrm{T}}^{l0}$ is applied to every top quark MC event.

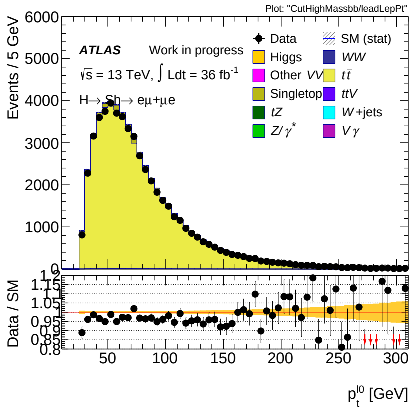
Figure 5.17 shows the lepton kinematics in the top VR after reweighting the top MC events in the $e\mu + \mu e$ channel. From the p_{T}^{l0} and p_{T}^{l1} distributions the slope is more flattened in the ratio plot, this shows the improvement in the MC prediction. The description of the $m_{\ell\ell}$ distribution has also improved, the average between data and MC agreement is around 0.98 ± 0.01. In the m_{T} distribution the same feature is observed. The p_{T}^{j0} distribution shows the agreement of data and MC to be 0.98 ± 0.01 in the region between 50 GeV and 150 GeV, in the region above 200 GeV the points in the ratio fluctuate due to low events statistics. The p_{T}^{j1} distribution shows noticeable improvement in the region between 50 GeV and 200 GeV as the slope in the ratio plot is corrected. Above 200 GeV points fluctuate due to low statistics on the tail of the distribution.

It is important to check that the modeling of the top quark process is validated in the $e\mu$ and μe channels. The p_{T}^{l0} correction works well separately for $e\mu$ and μe channel in the top VR, as shown in Figure 5.18. The slope in the p_{T}^{l0} distribution for $e\mu$ and μe is corrected. In general, there is an improvement between the data and the MC after reweighting for both the $e\mu$ and μe channels.

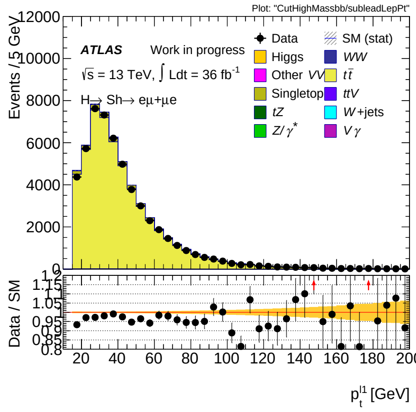
The distributions from top CR after the MC reweighting are shown in Figure 5.19 and Figure 5.20. The ratio of data over MC in the top CR has improved from 0.98 ± 0.01 to 1.00 ± 0.01, as shown in Table 5.7. The same features are observed in the top CR for $e\mu$ and μe channels separately, as illustrated in Figure 5.20. The data to MC agreement for both $e\mu$ and μe channel is 1.00 ± 0.01. Compared to the NLO EW corrections presented in Figures 5.13 and Figure 5.14, the reweighting of the top MC using p_{T}^{l0} distribution covers most of the discrepancies between data and MC. Extra distributions are shown in Appendix B.1.1 for $e\mu$ and μe channel separately: $\Delta\phi^{\ell\ell}$, $\Delta\eta^{\ell\ell}$, m_{T}, n_{jet}, η^{l0} and p_{T}^{j1}. The p_{T}^{l1} corrections are shown in Appendix B.1.2. These results also show a good improvement as the ratio of data to MC is 1.00 ± 0.01. The resulting top quark NF is 1.00 ± 0.01.

Table 5.7: Number of data events compared to the number of background events predicted by the MC simulation after the top re-weighting. Uncertainties are statistical only.

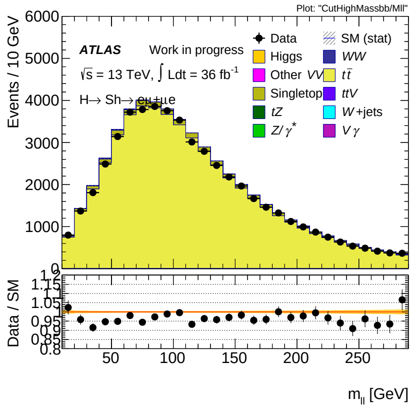
	top CR	top VR
Higgs	3.10 ± 0.08	26.70 ± 0.25
WW	1372.68 ± 7.00	33.90 ± 0.86
Other VV	137.64 ± 2.41	6.24 ± 0.32
$t\bar{t}$	3605.27 ± 12.14	53 896.96 ± 46.26
Singletop	470.48 ± 8.48	2123.58 ± 17.73
ttV	9.50 ± 0.26	85.21 ± 0.75
tZ	0.55 ± 0.02	1.90 ± 0.04
W+jets	40.44 ± 20.30	0.00 ± 0.00
Z/γ^*	18.96 ± 10.78	61.91 ± 3.53
$V\gamma$	25.88 ± 6.15	2.57 ± 1.67
Total Background	5684.48 ± 28.99	56 238.99 ± 9.71
Data	5699	54448
Data/MC	1.00 ± 0.01	0.97 ± 0.00



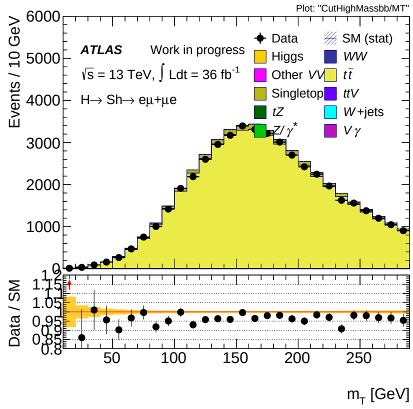
(a)



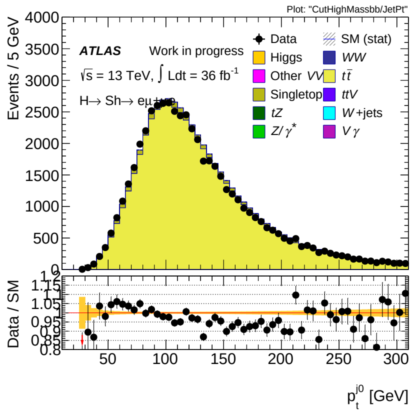
(b)



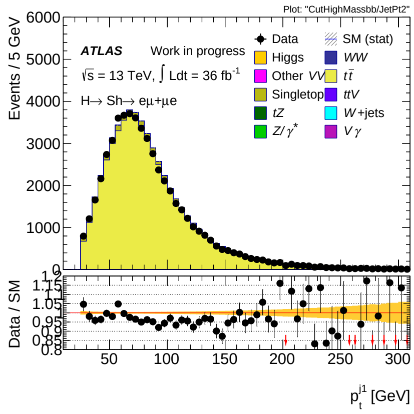
(c)



(d)



(e)



(f)

Figure 5.17: The distributions in the top VR after the p_{T}^{l0} correction has been applied: (a) leading lepton p_{T}, (b) subleading lepton p_{T}, (c) dilepton invariant mass, (d) transverse mass, (e) leading jet p_{T} and (f) subleading jet p_{T}. Uncertainties are statistical.

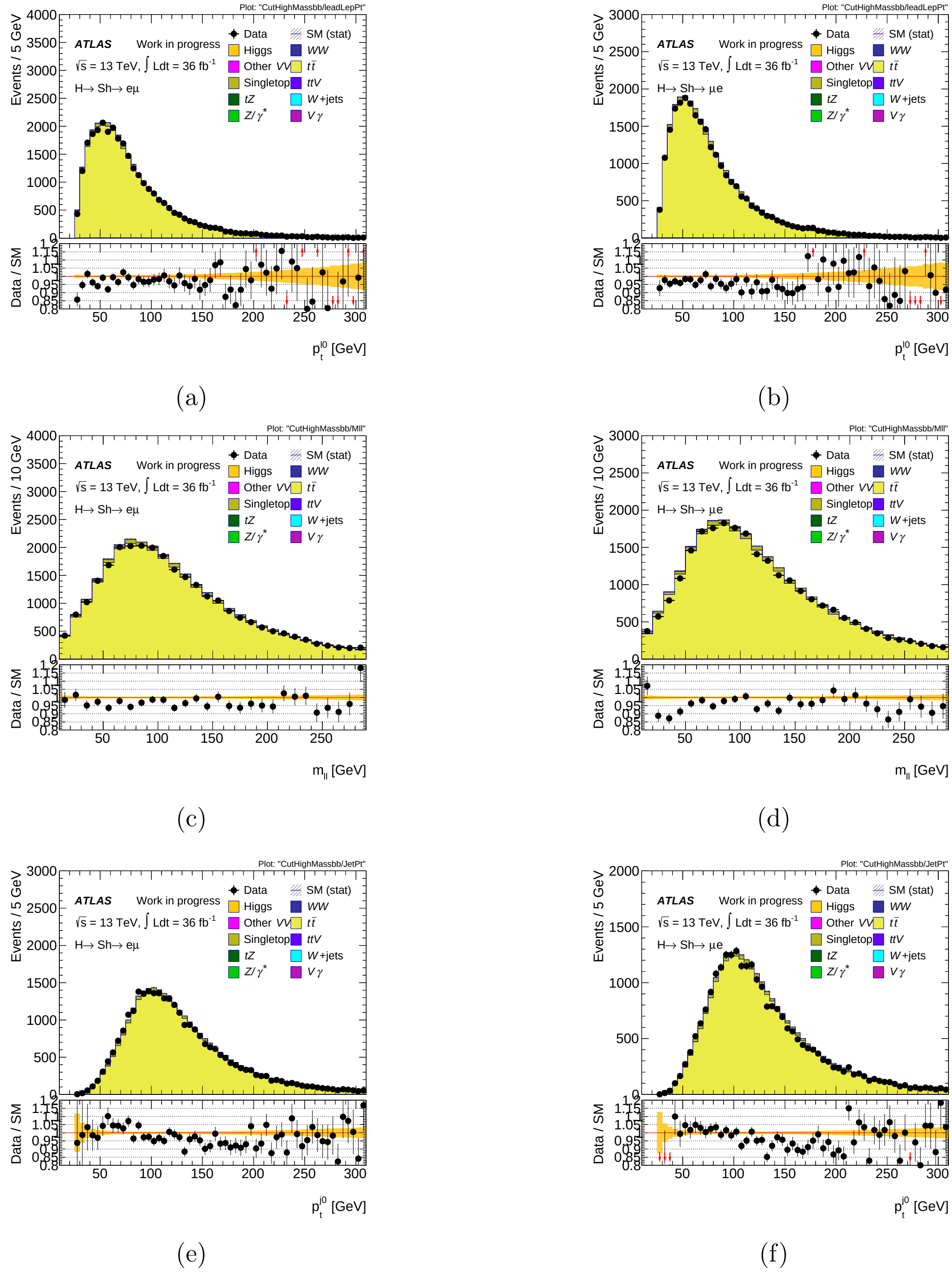
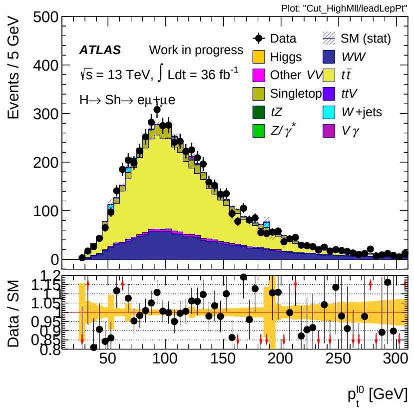
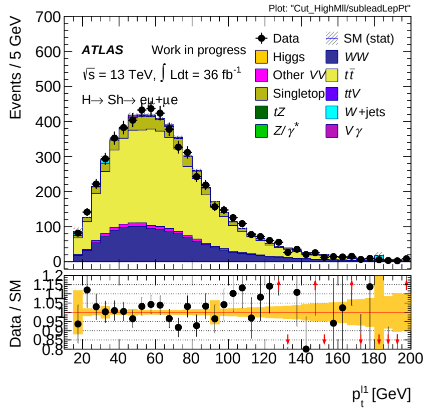


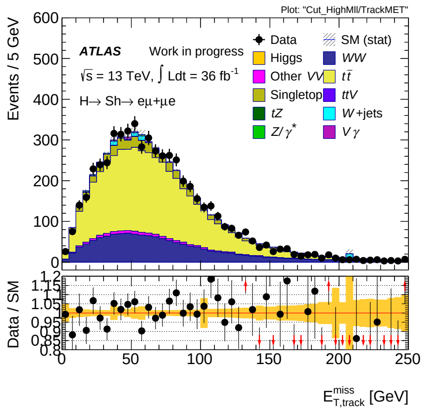
Figure 5.18: The distribution in the top VR after the p_{T}^{l0} correction has been applied for $e\mu$ channel: (a) leading lepton p_{T}, (c) dilepton invariant mass and (e) leading jet p_{T}. For μe channel: (d) leading lepton p_{T} (d) dilepton invariant mass and (f) leading jet p_{T}. Uncertainties are statistical.



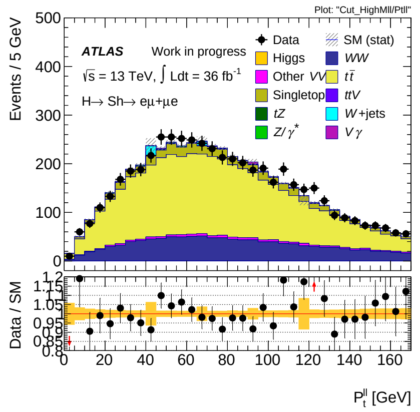
(a)



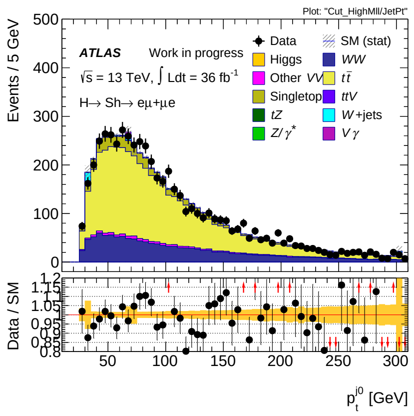
(b)



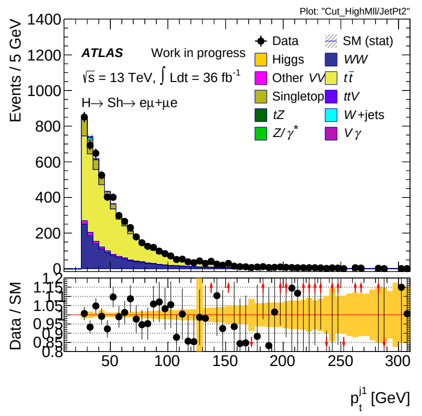
(c)



(d)

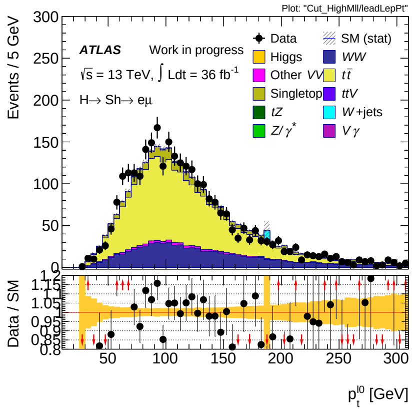


(e)

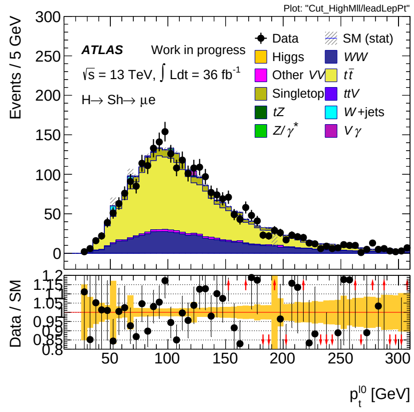


(f)

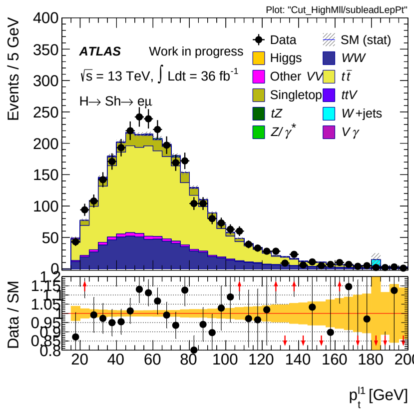
Figure 5.19: The distribution in the top CR after the p_{T}^{l0} correction has been applied: (a) leading lepton p_{T}, (b) subleading lepton p_T, (c) transverse missing momentum, (d) dilepton p_{T}, (e) leading jet p_{T} and (f) subleading jet p_{T}. Uncertainties are statistical.



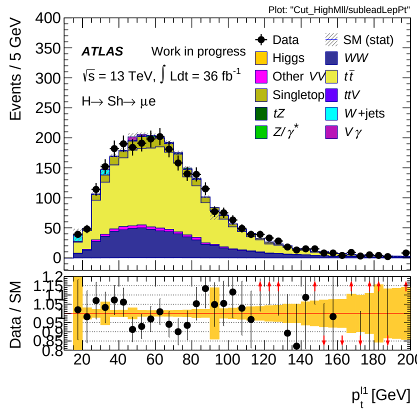
(a)



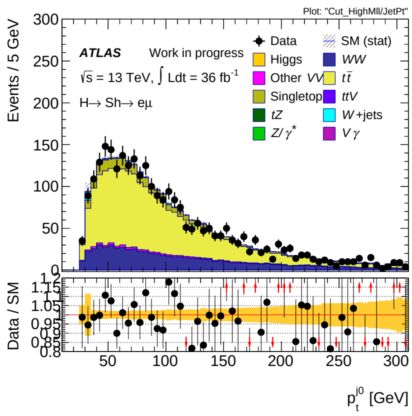
(b)



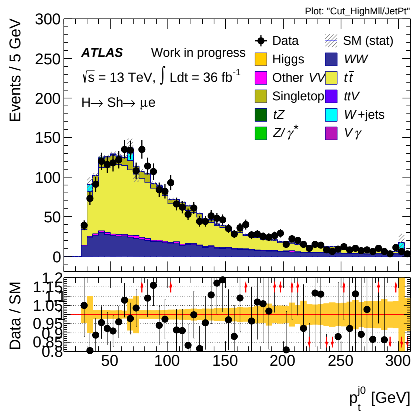
(c)



(d)



(e)



(f)

Figure 5.20: The distribution in the top CR after the p_{T}^{l0} correction has been applied for $e\mu$ channel: (a) leading lepton p_{T}, (c) subleading lepton p_{T} and (e) leading jet p_{T}. For μe channel: (d) leading lepton p_{T}, (d) subleading lepton p_{T} and (f) leading jet p_{T}. Uncertainties are statistical.

Chapter 6

Conclusions

Firstly, in Chapter 1 an introduction of the SM and BEH mechanism is given. Secondly, the discrepancies observed by the ATLAS and CMS experiments in the multilepton final states which can be accommodated by the simplified 2HDM+S model are discussed. This model considers that H decays predominantly to Sh, which then decays to a wide range of final states dominated by leptons and jets. Due to the use of the ATLAS data and the simulation of MC samples during this study, a brief introduction to the LHC and the ATLAS detector is given in Chapter 2.1.

The data taken at the centre-of-mass energy of $\sqrt{s} = 13$ TeV in 2015 and 2016 with an integrated luminosity of 36.1 fb^{-1} was analysed in the $H \rightarrow Sh$ search for the top quark background estimation. For this search, the SR is determined by events having at least 2 jets and 0 b-tagged jets.

The top quark process is the dominant background contributing about 60% to the total background in the SR. The two second important backgrounds are WW and $Z\gamma$, which contribute 19% and 18%, respectively. Two studies were conducted to identify the top CR and top VR: Studies of MC truth and comparison of the $m_{\ell\ell}$ shape. The former observes a comparable amount of truth b-jets. The latter observes a similar $m_{\ell\ell}$ shape in different b-tagged jet bins for the leading and subleading jets at reconstruction level in the $t\bar{t}$ and Wt MC samples.

In this thesis, the top quark background is studied in the top CR/VR. The top quark process ($t\bar{t}$ and Wt) modeling is evaluated in the top VR with such a good agreement between data and MC of 0.97 ± 0.01. Nonetheless, due to the observed mismodeling for top quark processes, the p_{T}^{l0} correction is applied to improve the performance of the MC simulation. This correction is from the top VR and is applied in the top CR as well. The study shows an excellent agreement between data and MC of 1.00 ± 0.01 after the p_{T}^{l0} corrections have been applied.

It should be acknowledged that MC corrections are possible from other leptonic quantities such as the p_{T}^{l1}, $\Delta\phi^{\ell\ell}$ and $m_{\ell\ell}$, but the MC mismodeling is clearly observed in the p_{T}^{l0} distribution. Alternatively, NLO EW corrections are applied in the top VR as well. The effect of this correction is about 2%, only at the tail of the distributions can be observed. Eventually, the full Run 2 data with an integrated luminosity of 139 fb^{-1} will be used for this search.

Appendix A

$H \rightarrow Sh$ MC production

A.1 Job-option

The job-option is a python script that provides all the information about the signal sample MC production. The one shown here is for the m_H= 350 GeV and m_S=240 GeV mass point and the same has been provided similarly for the other m_H= 240 GeV (m_S=170 GeV) and m_H= 400 GeV (m_S=240 GeV). In the job option script, the PDG values for H, S and h is given by 36, 35 and 25, respectively. The "36:onIfMatch" indicates that the scalar H is constraint on the decay of S and h.The "mWidth" tag indicates the decay width of scalar H and S which is set to 0.01.

include("MC15JobOptions/Pythia8-A14-NNPDF23LO-EvtGen-Common.py")
evgenConfig.generators = ["Pythia8", "EvtGen"]
evgenConfig.process = "$gg \rightarrow H \rightarrow Sh \rightarrow \ell\ell + jets$"
evgenConfig.description = "Generation of $gg \rightarrow H \rightarrow Sh$ where h, S goes to everything with a two lepton filter"
evgenConfig.keywords = ["BSMHiggs"]
evgenConfig.minevents = 10000
genSeq.Pythia8.Commands + = ['Higgs:useBSM = on',

```
'ParticleDecays:mSafety = 0.0',
'HiggsBSM:gg2A3 = on',
'Higgs:clipWings = off',
'36:m0 = 350.0',
'36:mWidth = 0.01',
'36:doForceWidth = yes',
'36:addChannel = 1 1 100 25 35',
'36:onMode = off',
'36:onIfMatch = 25 35',
'36:mayDecay = on',
'35:mMin = 50.0',
'25:mMin = 50.0',
'35:m0 = 240.0',
'35:mWidth = 0.01',
'35:doForceWidth = yes'
]
from GeneratorFilters.GeneratorFiltersConf import MultiLeptonFilter
filtSeq += MultiLeptonFilter("MultiLeptonFilter")
MultiLeptonFilter = filtSeq.MultiLeptonFilter
MultiLeptonFilter.Ptcut = 8000.
MultiLeptonFilter.Etacut = 2.5
MultiLeptonFilter.NLeptons = 2
filtSeq.Expression = "(MultiLeptonFilter)"
```

Appendix B

Top Quark Background Study

B.1 Top Control/Validation Region

The top quark background is analysed by defining top CR/VR as shown in Table B.1. The distributions from the top CR show the agreement between data and prediction presented in Figure B.1 and B.2 for $e\mu$ and μe channel, respectively. The validation plots illustrate the modeling of top quark processes ($t\bar{t}$ and Wt) in Figure B.3 and B.4 for $e\mu$ and μe channel, respectively. A full cutflow is shown in Table B.3, B.4 and B.2 for $e\mu$, μe and $e\mu + \mu e$ channel, respectively.

Table B.1: Analysis selection for top CR and VR.

Observable	top CR	top VR
Leading lepton p_T	> 27 GeV	
Sub-leading lepton p_T	> 15 GeV	
Number of jets	≥ 2	
Number of b-tagged jets	$=0$	$=2$
$m_{\ell\ell}$	> 150 GeV	-
m_{bb}	-	> 150 GeV

Table B.2: Number of data events compared to the number of background events predicted by the MC simulation for $e\mu + \mu e$ channel. Uncertainties are statistical.

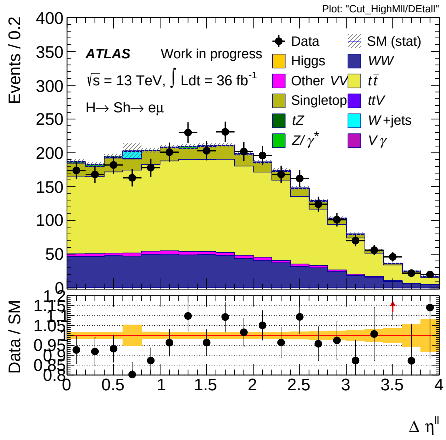
	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	3.10 ± 0.08	1372.68 ± 7.00	137.64 ± 2.41	3719.38 ± 12.51	490.82 ± 8.83	9.98 ± 0.27	0.57 ± 0.02	40.43 ± 20.30	18.96 ± 10.78	25.88 ± 6.15	5819.44 ± 29.25	0.98 ± 0.01	5699
NBjets=2:$M_{bb} > 150$ GeV	26.70 ± 0.25	33.90 ± 0.86	6.24 ± 0.32	53897.05 ± 46.25	2168.46 ± 18.09	86.64 ± 0.76	1.93 ± 0.04	0.00	61.91 ± 3.53	2.57 ± 1.67	56285.41 ± 49.84	0.97 ± 0.00	54448

Table B.3: Number of data events compared to the number of background events predicted by the MC simulation for $e\mu$ channel. Uncertainties are statistical.

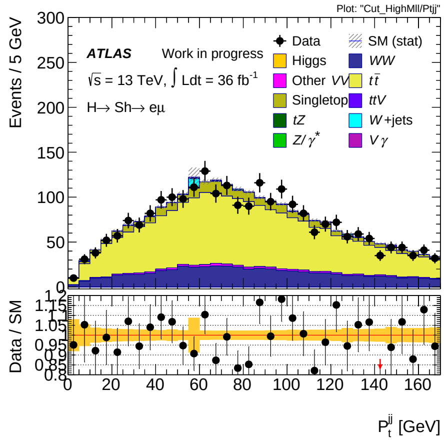
	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	1.57 ± 0.06	698.51 ± 5.00	70.67 ± 1.11	1918.71 ± 9.01	259.36 ± 6.44	5.27 ± 0.19	0.32 ± 0.02	10.32 ± 10.32	15.07 ± 3.09	10.32 ± 2.12	2990.11 ± 16.42	0.98 ± 0.02	2933
NBjets=2:$M_{bb} > 150$ GeV	14.28 ± 0.18	18.25 ± 0.66	3.31 ± 0.24	28525.68 ± 33.77	1146.07 ± 13.20	45.54 ± 0.55	1.06 ± 0.03	0.00	35.57 ± 2.68	0.20 ± 0.11	29789.96 ± 36.37	0.97 ± 0.01	28932

Table B.4: Number of data events compared to the number of background events predicted by the MC simulation for μe channel. Uncertainties are statistical.

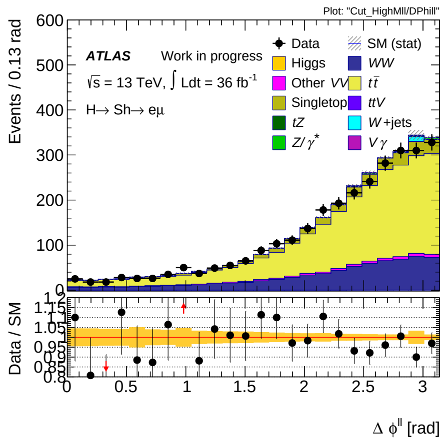
	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	1.53 ± 0.06	674.17 ± 4.90	66.98 ± 2.14	1800.67 ± 8.69	231.46 ± 6.05	4.71 ± 0.19	0.25 ± 0.01	30.10 ± 17.48	3.89 ± 10.33	15.56 ± 5.78	2829.33 ± 24.21	0.98 ± 0.02	2766
NBjets=2 :$M_{bb} > 150$ GeV	12.41 ± 0.17	15.65 ± 0.55	2.94 ± 0.22	25371.37 ± 31.61	1022.39 ± 12.37	41.11 ± 0.52	0.86 ± 0.03	0.00	26.35 ± 2.30	2.37 ± 1.67	26495.45 ± 34.07	0.96 ± 0.01	25516



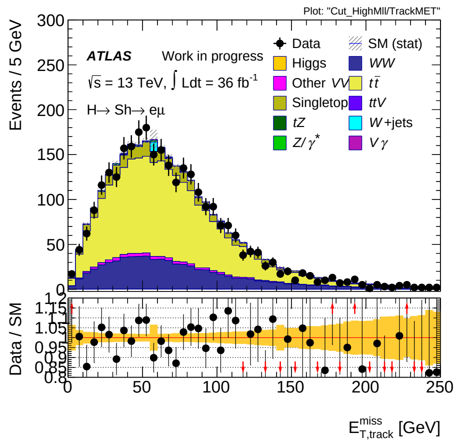
(a)



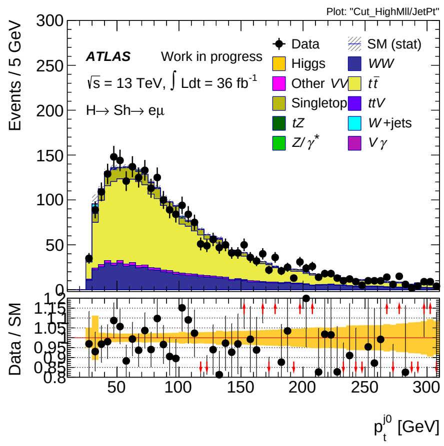
(b)



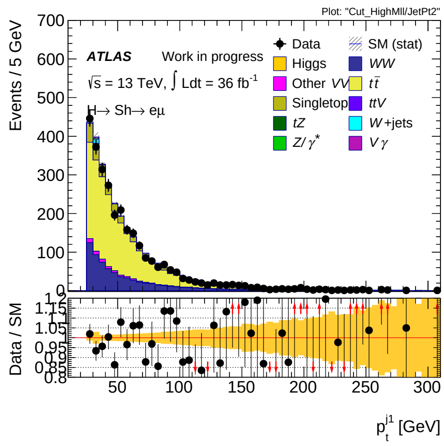
(c)



(d)

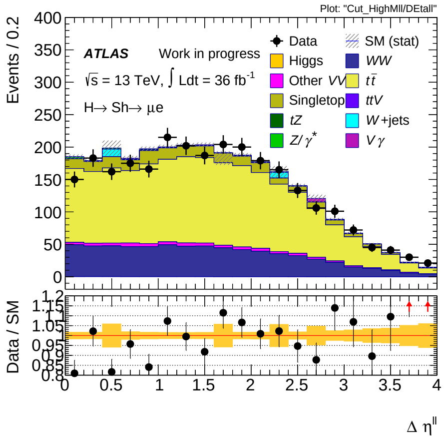


(e)

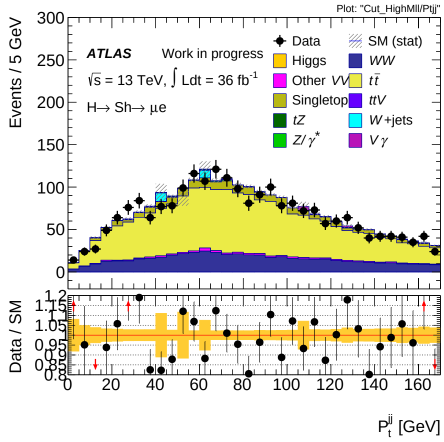


(f)

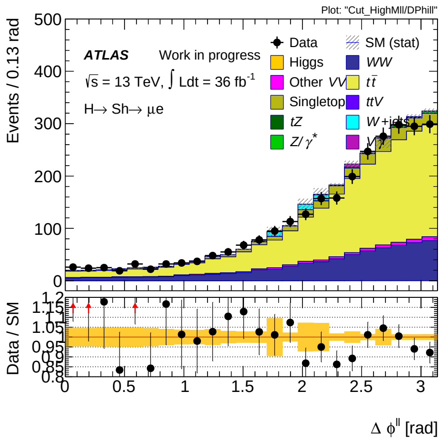
Figure B.1: Comparison of data and background in the top CR in the $e\mu$ channel: (a) $\Delta\eta^{\ell\ell}$, (b) p_{T}^{jj}, (c) $\Delta\phi^{\ell\ell}$, (d) $E_{\mathrm{T}}^{\mathrm{miss,track}}$, (e) p_{T}^{j0} and (f) p_{T}^{j1}.



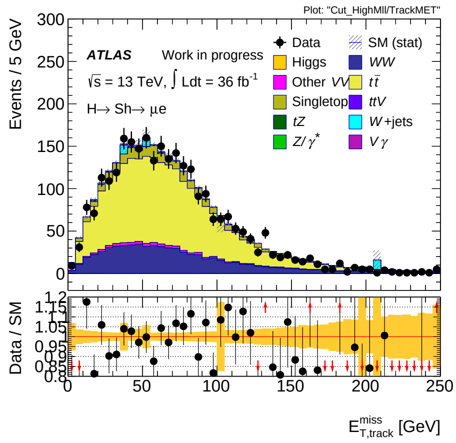
(a)



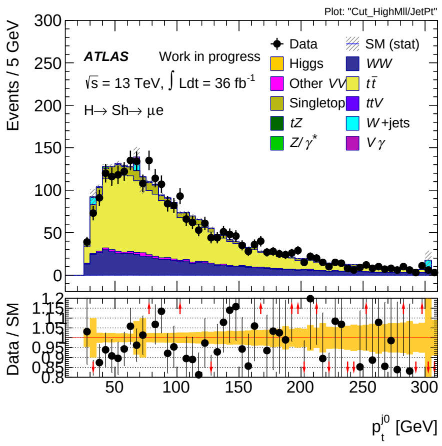
(b)



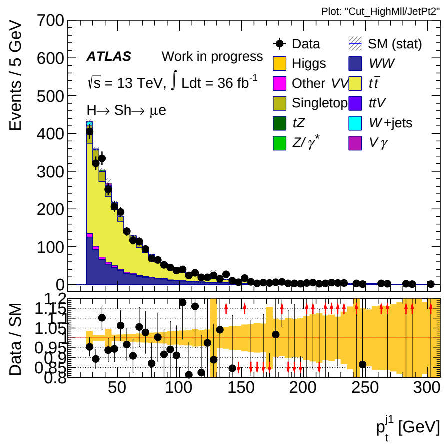
(c)



(d)



(e)



(f)

Figure B.2: Comparison of data and background in the top CR in the μe channel: (a) $\Delta\eta^{\ell\ell}$, (b) p_{T}^{jj}, (c) $\Delta\phi^{\ell\ell}$, (d) $E_{\mathrm{T}}^{\mathrm{miss,track}}$, (e) p_{T}^{j0} and (f) p_{T}^{j1}.

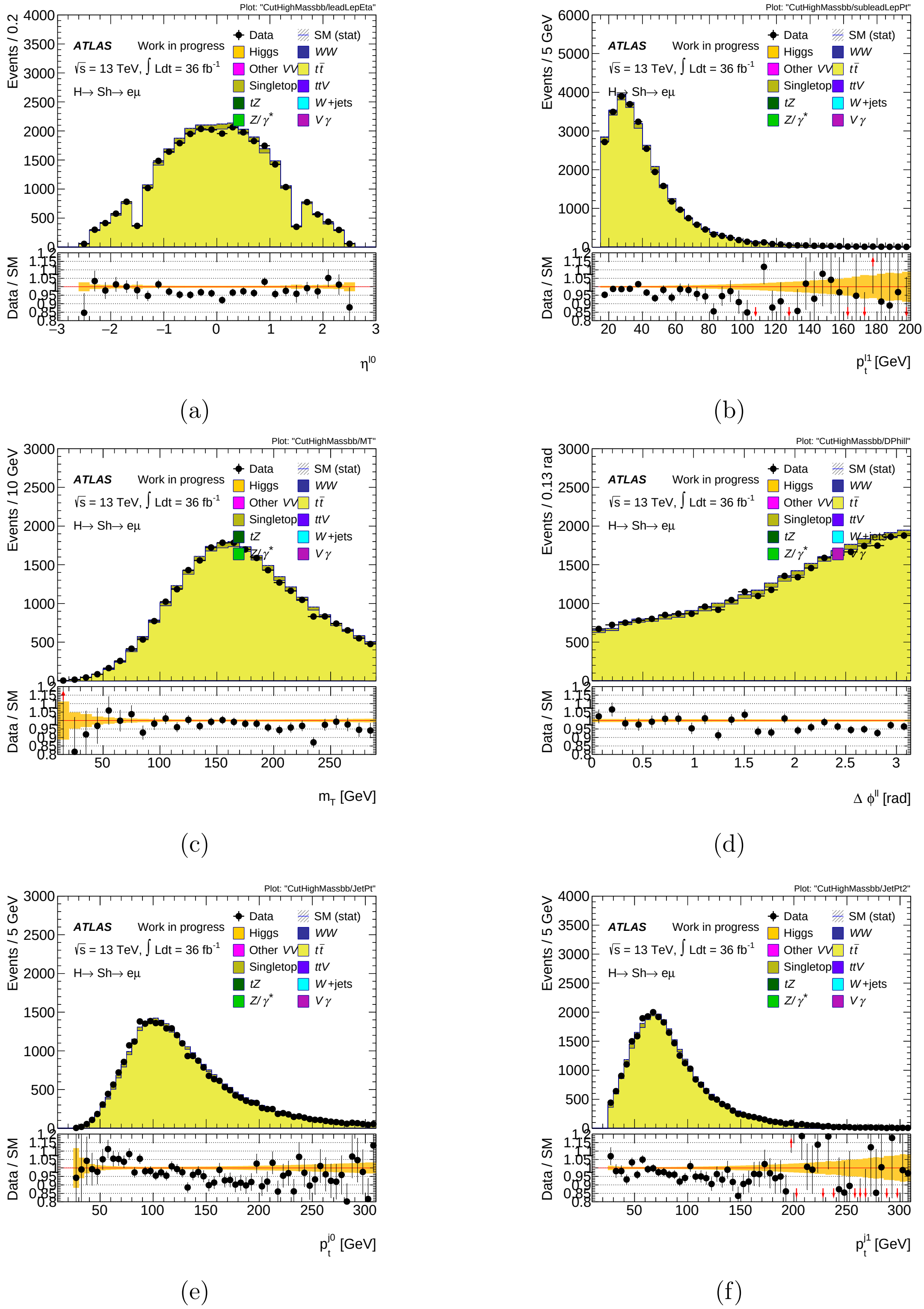


Figure B.3: Comparison of data and MC in the top VR in the $e\mu$ channel: (a) p_{T}^{l0}, (b) p_{T}^{l1}, (c) m_{T}, (d) $\Delta\phi^{\ell\ell}$, (e) p_{T}^{j0} and (f) p_{T}^{j1}.

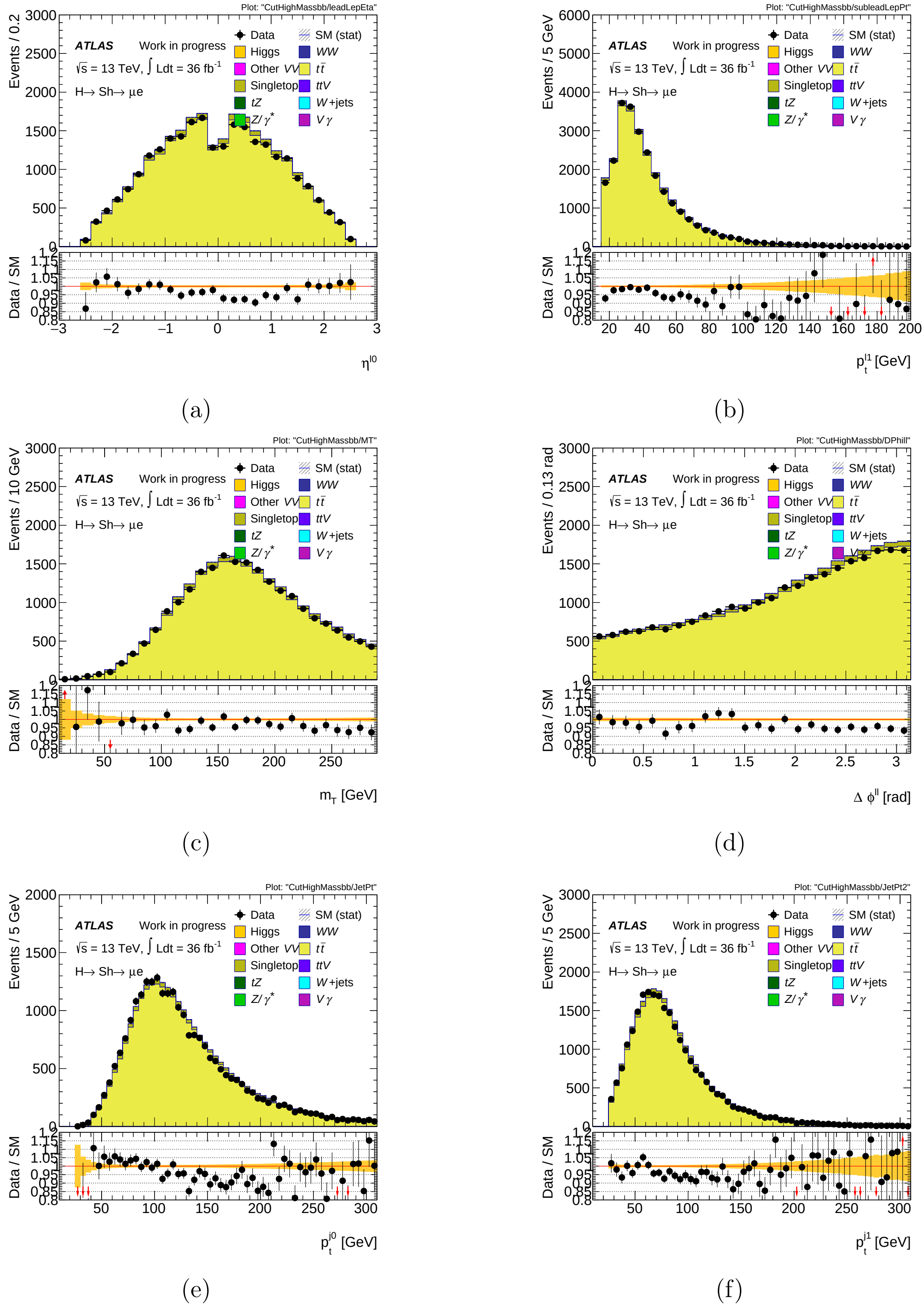


Figure B.4: Comparison of data and MC in the top VR in the μe channel: (a) p_{T}^{l0}, (b) p_{T}^{l0}, (c) m_{T}, (d) $\Delta\phi^{\ell\ell}$, (e) p_{T}^{j0} and (f) p_{T}^{j1}.

B.1.1 MC Corrections using $p_{\mathbf{T}}^{l0}$ distribution

The section presents the leptonic quantities after p_{T}^{l0} corrections has been applied. Top VR distributions are shown in Figure B.5 and Figure B.6 for $e\mu$ and μe channel, respectively. For $e\mu$ channel in top VR the ratio of data over MC is 1 which is the same for μe channel. The leptonic quantities in top VR are shown in Figure B.7 and Figure B.12 for $e\mu$ and μe channel, respectively. In this region, the agreement between data and MC is reasonable and the ratio of data over MC is 0.97. The cutflows show the number of events survived after applying cuts on the leptonic quantities along with the agreement between data and MC. This is shown in Table B.6 and B.7 for $e\mu$ and μe channel. Their combination is shown in Table B.5.

Table B.5: Number of events recorded in the data compared to the number of events predicted in the background for $e\mu + \mu e$. channel. Uncertainties are statistical.

	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	3.10 ± 0.08	1372.68 ± 7.00	137.64 ± 2.41	3605.27 ± 12.14	470.48 ± 8.48	9.50 ± 0.26	0.55 ± 0.02	40.43 ± 20.30	18.96 ± 10.78	25.88 ± 6.15	5684.48 ± 28.99	1.00 ± 0.01	5699
NBjets=2:$M_{bb} > 150$ GeV	26.70 ± 0.25	33.90 ± 0.86	6.24 ± 0.32	53896.96 ± 46.26	2123.58 ± 17.73	85.21 ± 0.75	1.90 ± 0.04	0.00	61.91 ± 3.53	2.57 ± 1.67	56238.99 ± 49.71	0.97 ± 0.00	54448

Table B.6: Number of data events compared to the number of background events predicted by the MC simulation for $e\mu$ channel. Uncertainties are statistical.

	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	1.57 ± 0.06	698.51 ± 5.00	70.67 ± 1.11	1860.09 ± 8.74	248.73 ± 6.18	5.03 ± 0.19	0.31 ± 0.02	10.32 ± 10.32	15.07 ± 3.09	10.32 ± 2.12	2920.62 ± 16.17	1.00 ± 0.02	2933
NBjets=2:$M_{bb} > 150$ GeV	14.28 ± 0.18	18.25 ± 0.66	3.31 ± 0.24	28529.13 ± 33.78	1124.65 ± 12.96	44.80 ± 0.54	1.05 ± 0.03	0.00	35.57 ± 2.68	0.20 ± 0.11	29771.23 ± 36.29	0.97 ± 0.01	28932

Table B.7: Number of data events compared to the number of background events predicted by the MC simulation for μe channel. Uncertainties are statistical.

	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	1.53 ± 0.06	674.17 ± 4.90	66.98 ± 2.14	1745.17 ± 8.43	221.75 ± 5.80	4.47 ± 0.18	0.24 ± 0.01	30.10 ± 17.48	3.89 ± 10.33	15.56 ± 5.78	2763.86 ± 24.06	1.00 ± 0.02	2766
NBjets=2:$M_{bb} > 150$ GeV	12.41 ± 0.17	15.65 ± 0.55	2.94 ± 0.22	25367.83 ± 31.60	998.93 ± 12.10	40.41 ± 0.51	0.85 ± 0.03	0.00	26.35 ± 2.30	2.37 ± 1.67	26467.75 ± 33.97	0.96 ± 0.01	25516

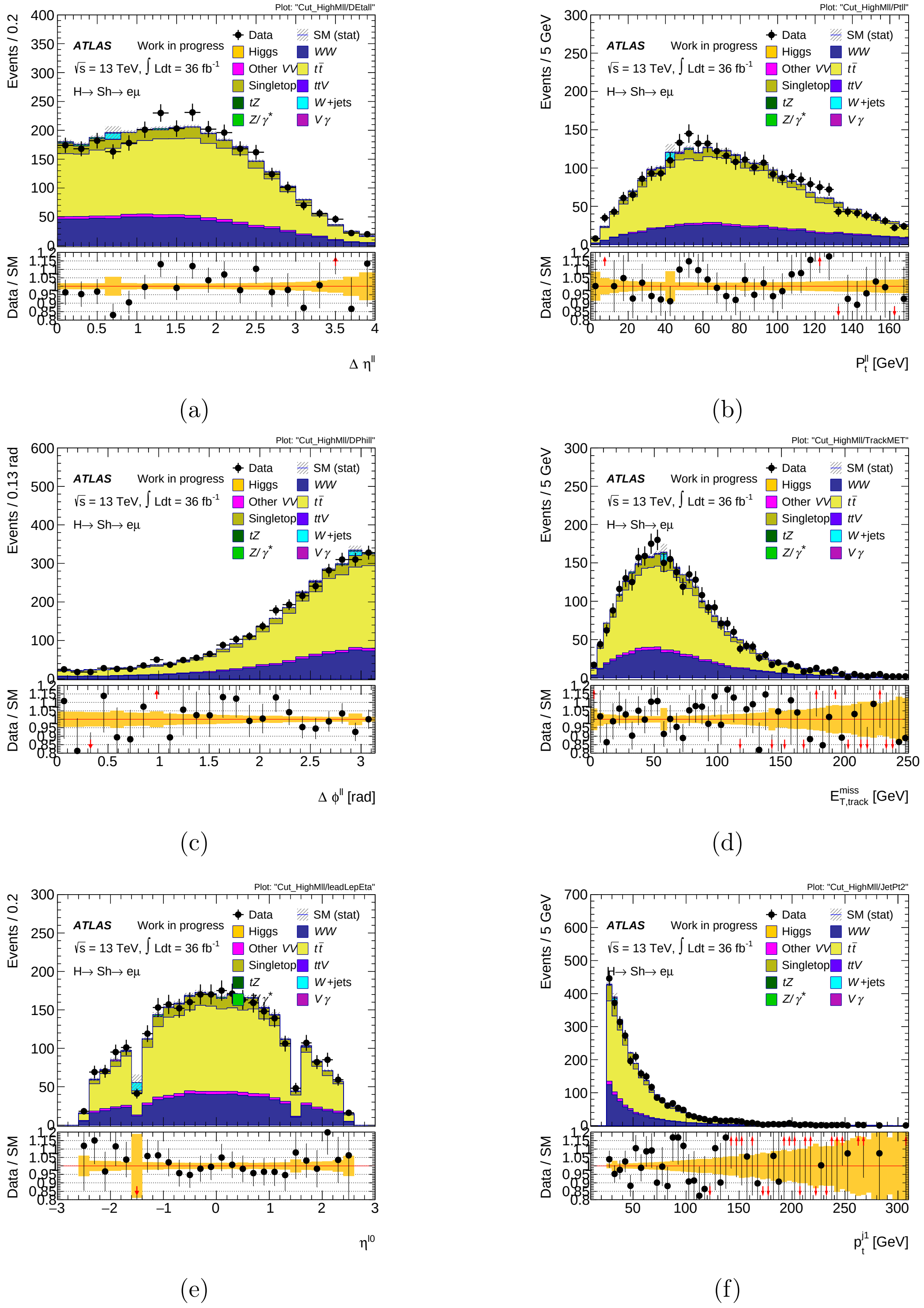


Figure B.5: Comparison of data and MC in the top CR with p_T^{l0} correction applied in the $e\mu$ channel: (a) $\Delta\phi_{\ell\ell}$, (b) $p_{\mathrm{T}}^{\ell\ell}$, (c) $\Delta\phi^{\ell\ell}$, (d) $E_{\mathrm{T}}^{\mathrm{miss,Track}}$, (e) η^{l0} and (f) p_{T}^{j1}.

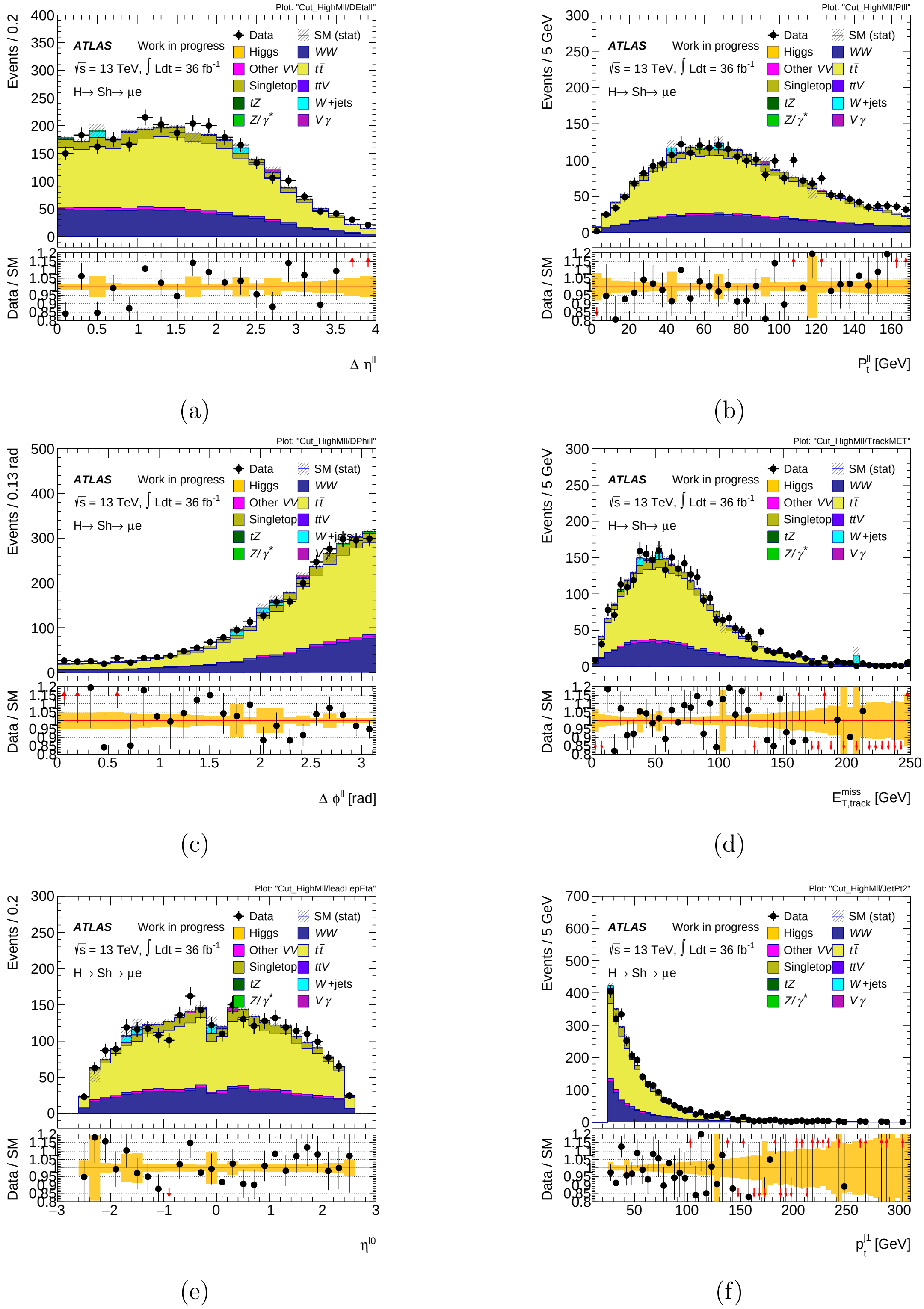
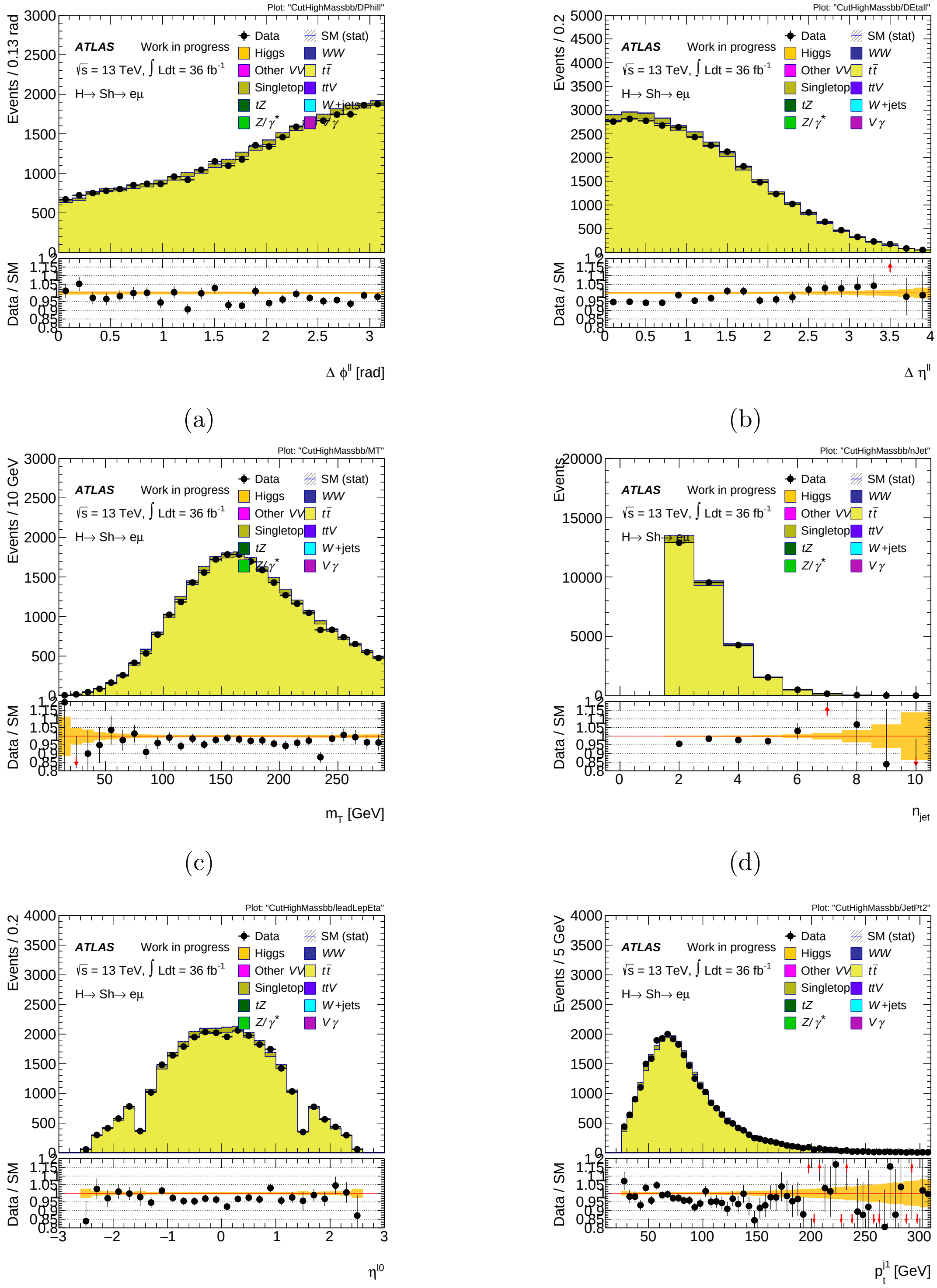


Figure B.6: Comparison of data and MC in the top CR with p_{T}^{l0} correction applied in the μe channel: (a) $\Delta\phi_{\ell\ell}$, (b) $p_{\mathrm{T}}^{\ell\ell}$, (c) $\Delta\phi^{\ell\ell}$, (d) $E_{\mathrm{T}}^{\mathrm{miss,track}}$, (e) η^{l0} and (f) p_{T}^{j1}.



(a) (b)

(c) (d)

(e) (f)

Figure B.7: Comparison of data and MC in the top VR with p_{T}^{l0} correction applied in the $e\mu$ channel: (a) $\Delta\phi^{\ell\ell}$, (b) $\Delta\eta^{\ell\ell}$, (c) m_{T}, (d) N_{jet}, (e) η^{l0} and (f) p_{T}^{j1}.

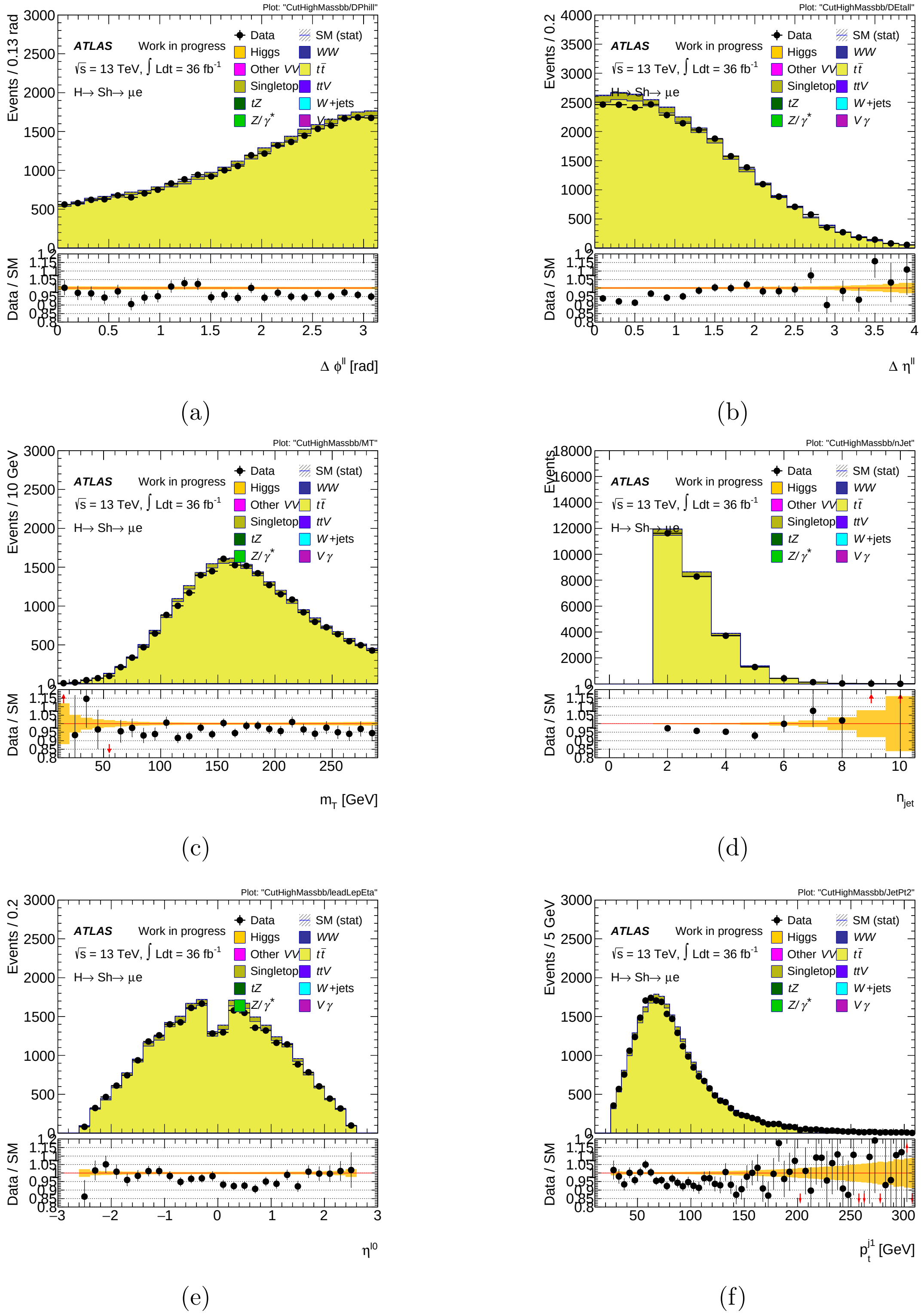


Figure B.8: Comparison of data and MC in the top VR with p_{T}^{l0} correction applied in the μe channel: (a) $\Delta\phi^{\ell\ell}$, (b) $\Delta\eta^{\ell\ell}$, (c) m_{T}, (d) N_{jet}, (e) η^{l0} and (f) p_{T}^{j1}.

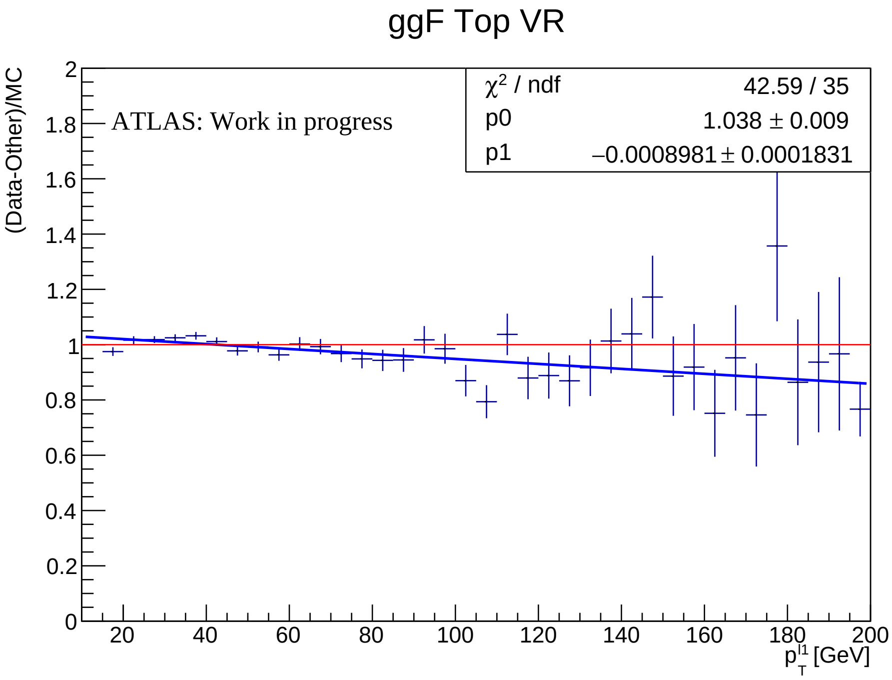


Figure B.9: Fitted p_{T} correction parameters for the subleading lepton p_{T}.

B.1.2 MC Corrections using p_{T}^{l1} distribution

The section presents the distributions after applying p_{T}^{l1} corrections. Top VR distributions are shown in Figure B.10 and Figure B.11 for $e\mu$ and μe channel, respectively. For $e\mu$ channel in top VR the ratio of data over MC is 1.00 ± 0.01 which is the same for μe channel. The leptonic quantities in top VR are shown in Figure B.12 and Figure B.13 for $e\mu$ and μe channel, respectively. In this region, the agreement between data and MC is reasonable and the ratio of data over MC is 0.97. Table B.9 and B.10 for $e\mu$ and μe channel shows the event yield in the top VR and top CR after correction using p_{T}^{l1}. Their combination is shown in Table B.8.

Table B.8: Number of events recorded in the data compared to the number of events predicted in the background for $e\mu + \mu e$. channel. Uncertainties are statistical.

	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	3.10 ± 0.08	1372.68 ± 7.00	137.64 ± 2.41	3637.55 ± 12.24	479.57 ± 8.63	9.72 ± 0.26	0.56 ± 0.02	40.43 ± 20.30	18.96 ± 10.78	25.88 ± 6.15	5726.08 ± 29.08	1.00 ± 0.01	5699
NBjets=2:$M_{bb} > 150$ GeV	26.70 ± 0.25	33.90 ± 0.86	6.24 ± 0.32	53857.90 ± 46.21	2159.50 ± 18.01	86.13 ± 0.75	1.92 ± 0.04	0.00	61.91 ± 3.53	2.57 ± 1.67	56236.78 ± 49.77	0.97 ± 0.00	54448

Table B.9: Number of data events compared to the number of background events predicted by the MC simulation for $e\mu$ channel. Uncertainties are statistical.

	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	1.57 ± 0.06	698.51 ± 5.00	70.67 ± 1.11	1877.36 ± 8.82	253.58 ± 6.30	5.15 ± 0.19	0.31 ± 0.02	10.32 ± 10.32	15.07 ± 3.09	10.32 ± 2.12	2942.86 ± 16.26	1.00 ± 0.02	2933
NBjets=2:$M_{bb} > 150$ GeV	14.28 ± 0.18	18.25 ± 0.66	3.31 ± 0.24	28528.64 ± 33.77	1142.68 ± 13.16	45.30 ± 0.55	1.06 ± 0.03	0.00	35.57 ± 2.68	0.20 ± 0.11	29789.28 ± 36.36	0.97 ± 0.01	28932

Table B.10: Number of data events compared to the number of background events predicted by the MC simulation for μe channel. Uncertainties are statistical.

	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets0bjets:$M_{ll} > 150$ GeV	1.53 ± 0.06	674.17 ± 4.90	66.98 ± 2.14	1760.18 ± 8.49	225.99 ± 5.90	4.57 ± 0.18	0.24 ± 0.01	30.10 ± 17.48	3.89 ± 10.33	15.56 ± 5.78	2783.22 ± 24.11	0.99 ± 0.02	2766
NBjets=2:$M_{bb} > 150$ GeV	12.41 ± 0.17	15.65 ± 0.55	2.94 ± 0.22	25329.26 ± 31.54	1016.82 ± 12.30	40.83 ± 0.52	0.86 ± 0.03	0.00	26.35 ± 2.30	2.37 ± 1.67	26447.50 ± 33.99	0.96 ± 0.01	25516

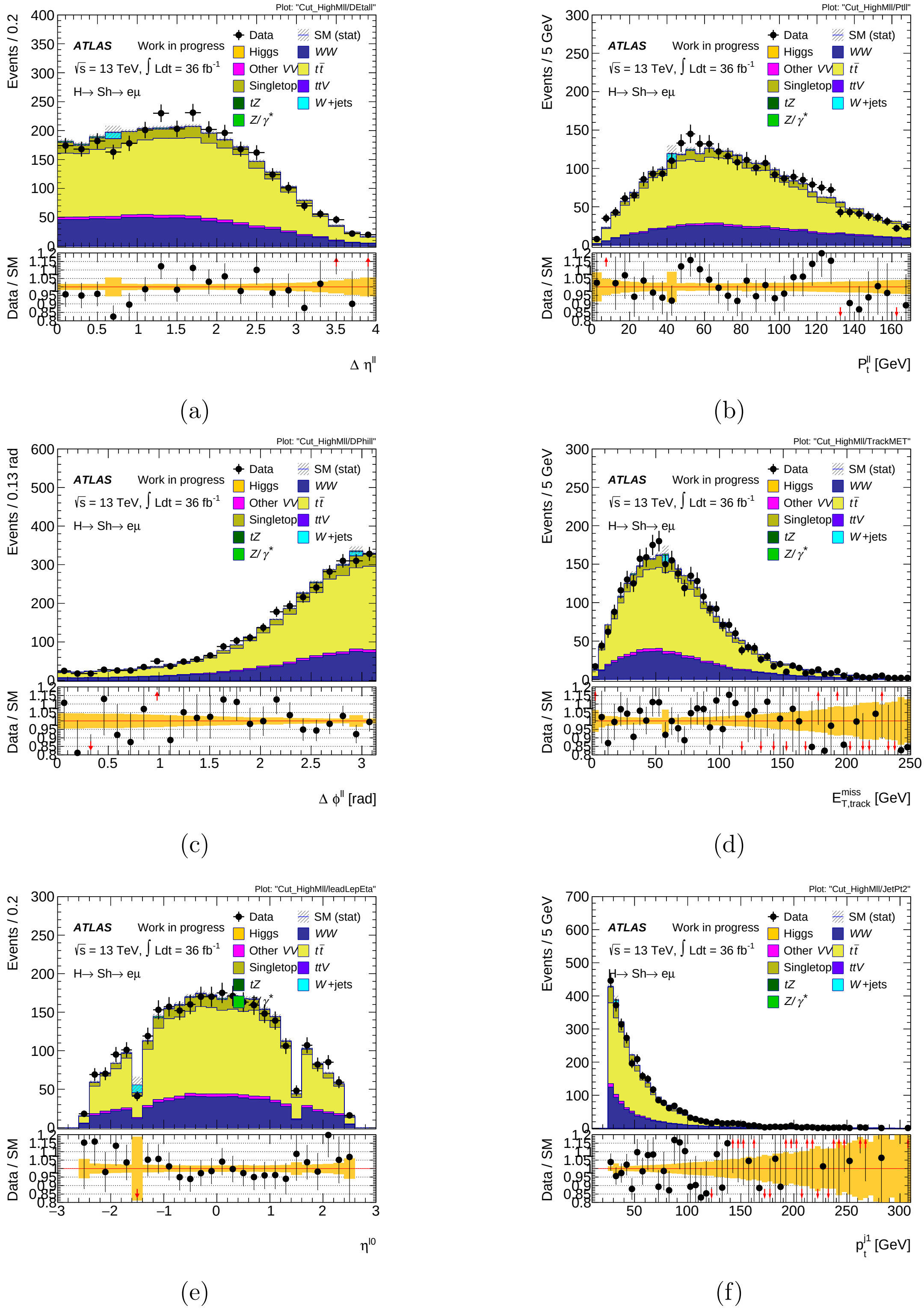


Figure B.10: Comparison of data and MC in the top CR with p_T^{l1} correction applied in the $e\mu$ channel: (a) $\Delta\phi_{\ell\ell}$, (b) $p_{\mathrm{T}}^{\ell\ell}$, (c) $\Delta\phi^{\ell\ell}$, (d) $E_{\mathrm{T}}^{\mathrm{miss,track}}$, (e) η^{l0} and (f) p_{T}^{j1}.

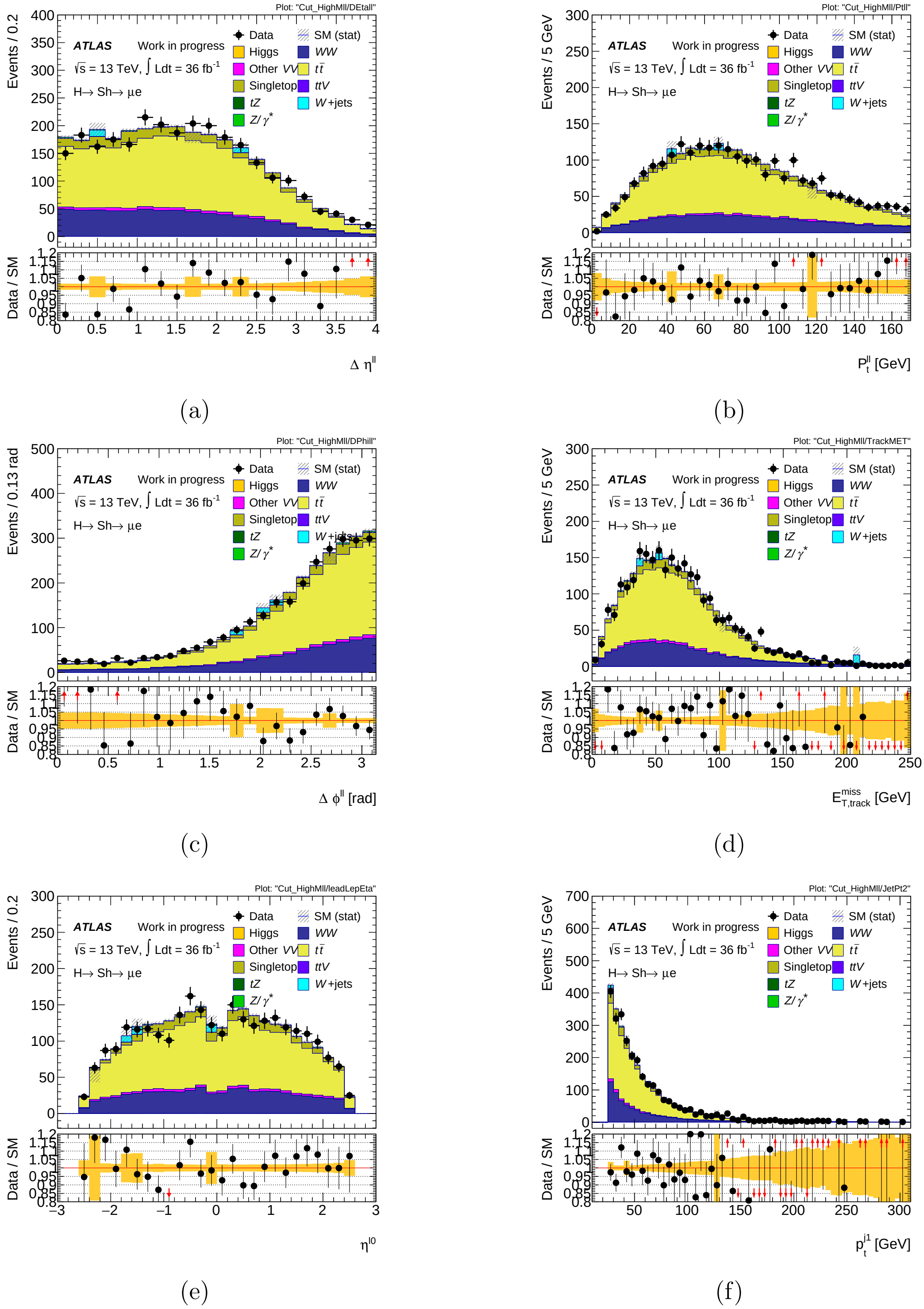


Figure B.11: Comparison of data and MC in the top CR with p_T^{l1} correction applied in the μe channel: (a) $\Delta\phi_{\ell\ell}$, (b) $p_T^{\ell\ell}$, (c) $\Delta\phi^{\ell\ell}$, (d) $E_T^{\text{miss,track}}$, (e) η^{l0} and (f) p_T^{j1}.

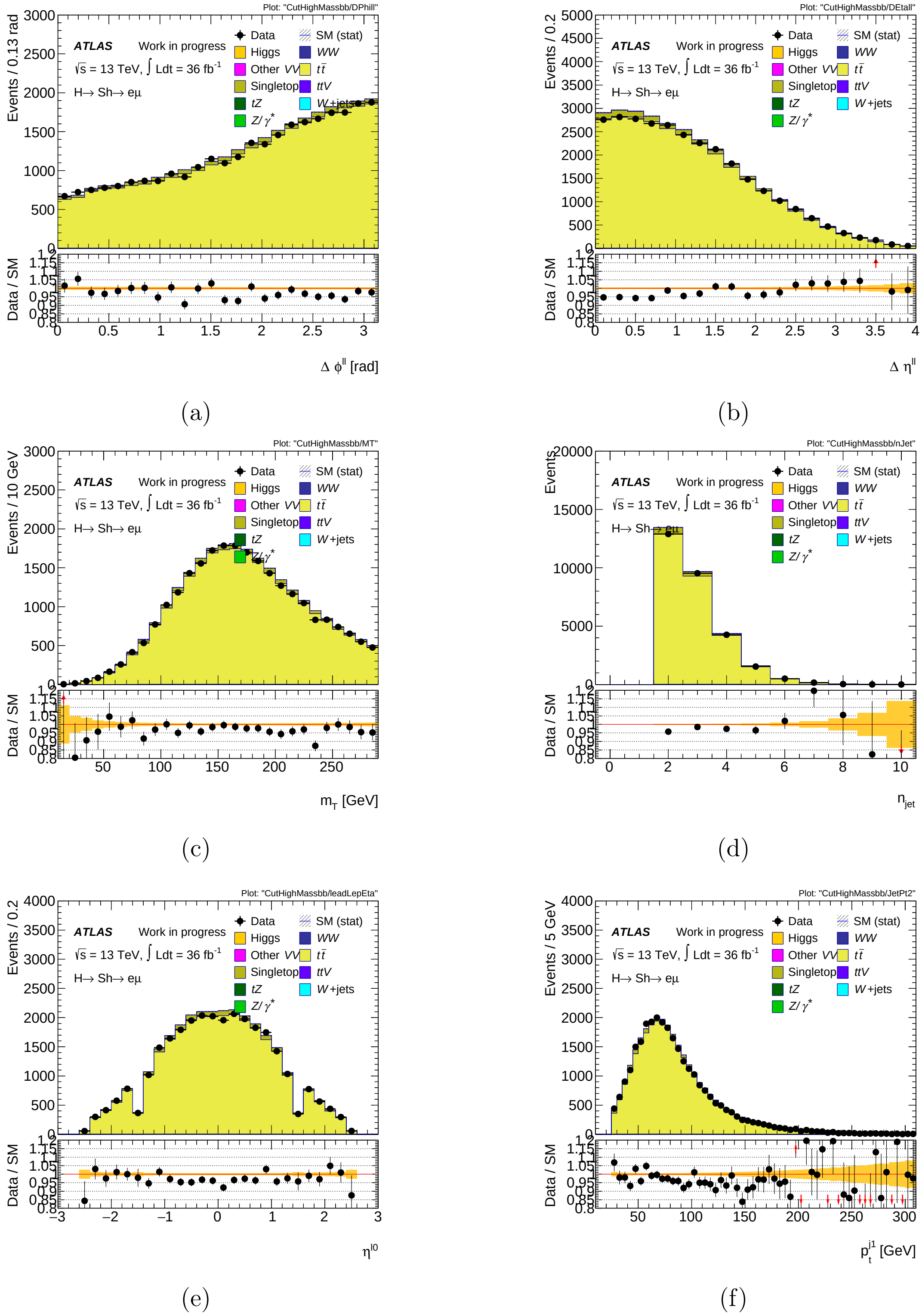
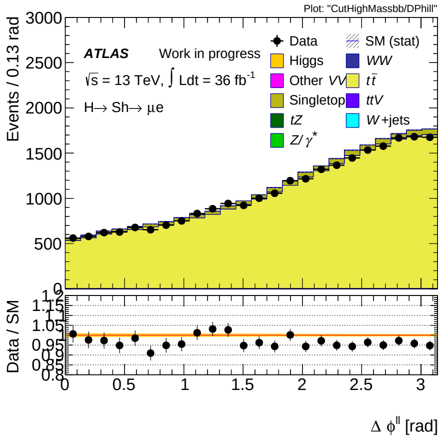
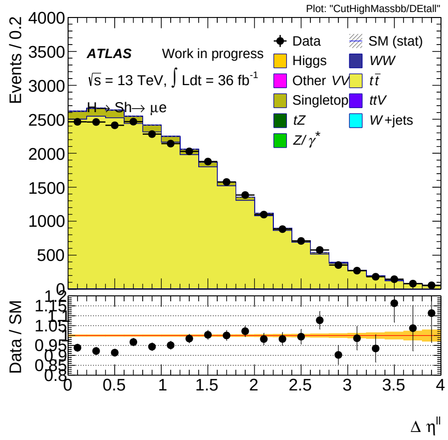


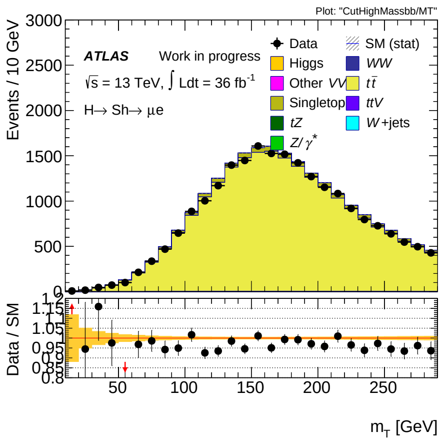
Figure B.12: Comparison of data and MC in the top VR with p_{T}^{l1} correction applied in the $e\mu$ channel: (a) $\Delta\phi^{\ell\ell}$, (b) $\Delta\eta^{\ell\ell}$, (c) m_{T}, (d) N_{jet}, (e) η^{l0} and (f) p_{T}^{j1}.



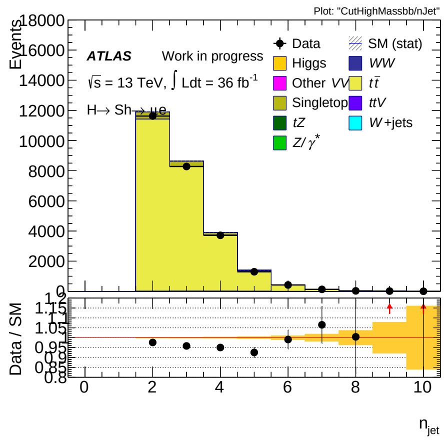
(a)



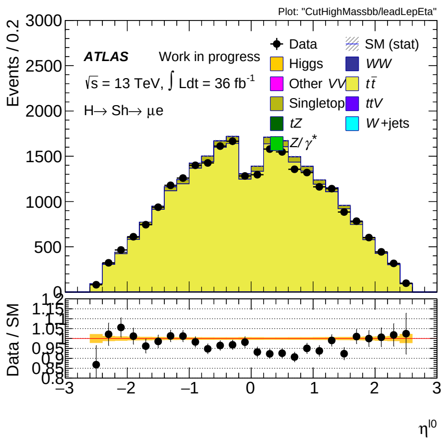
(b)



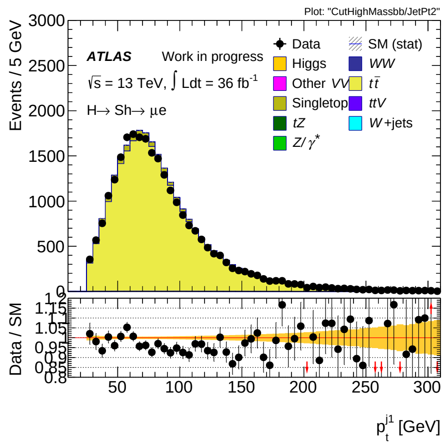
(c)



(d)



(e)



(f)

Figure B.13: Comparison of data and MC in the top VR with p_{T}^{l1} correction applied in the μe channel: (a) $\Delta\phi^{\ell\ell}$, (b) $\Delta\eta^{\ell\ell}$, (c) m_{T}, (d) N_{jet}, (e) η^{l0} and (f) p_{T}^{j1}.

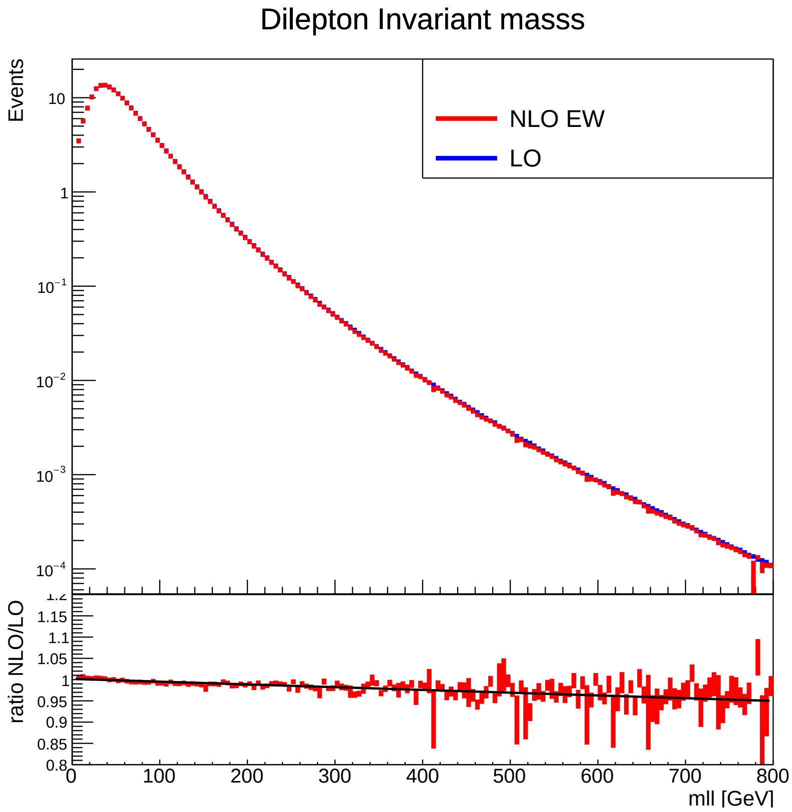


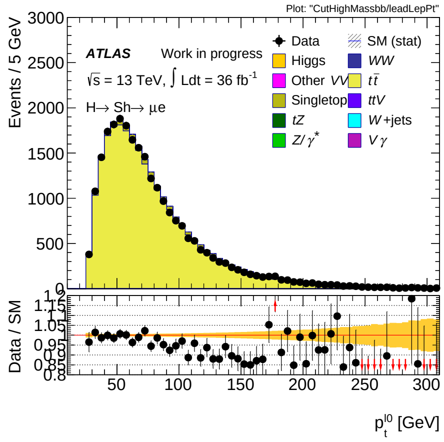
Figure B.14: The distribution of the $m_{\ell\ell}$ shows a shape comparison between NLO EW and LO.

B.1.3 NLO EW Corrections

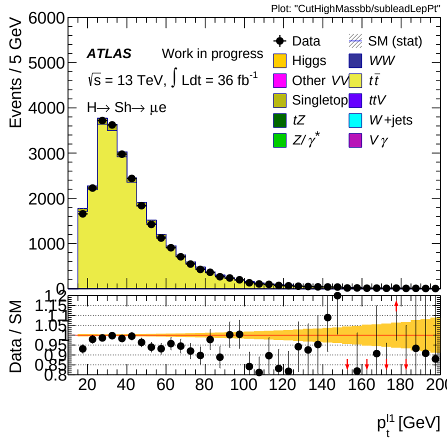
In this section, NLO EW corrections [44, 45, 46] are performed using $m_{\ell\ell}$ distribution, the correction factor are obtained by fitting the ratio NLO/LO as a function of the $m_{\ell\ell}$ as shown in Figure B.14. The obtained parameters for the fitted line shown in the lower panel of $m_{\ell\ell}$ distribution are $p0 = 1.00164 \pm 0.000430458$ and $p1 = -0.000064843 \pm 0.000000307128$. The effect on the leptonic quantities is very small in the top VR as shown in Figure B.15. With this observation it is clear that p_{T}^{l0} corrections applied to the MC works much better than NLO EW corrections.

Table B.11: Number of events recorded in the data compared to the number of events predicted in the background for $e\mu + \mu e$. channel after NLO EW corrections. Uncertainties are statistical.

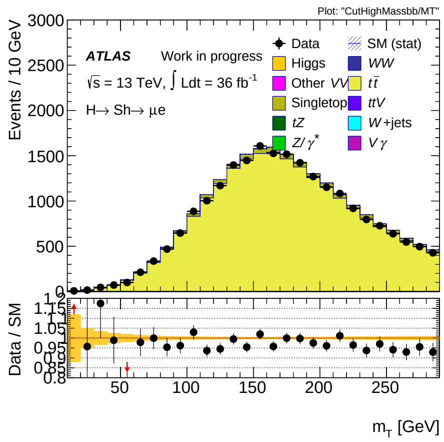
	SM Higgs	WW	Other VV	$t\bar{t}$	Singletop	ttV	tZ	W+jets	Z/γ^*	$V\gamma$	Total Bkg	Data/bkg	data
0bjets:$M_{ll} > 150$ GeV	3.10 ± 0.08	1372.68 ± 7.00	137.64 ± 2.41	3696.67 ± 12.44	487.39 ± 8.77	9.90 ± 0.27	0.56 ± 0.02	40.43 ± 20.30	18.96 ± 10.78	25.88 ± 6.15	5793.22 ± 29.20	0.98 ± 0.01	5699
NBjets=2 :$M_{bb} > 150$ GeV	26.70 ± 0.25	33.90 ± 0.86	6.24 ± 0.32	53704.67 ± 46.09	2157.01 ± 18.00	86.21 ± 0.75	1.92 ± 0.04	0.00	61.91 ± 3.53	2.57 ± 1.67	56081.14 ± 49.65	0.97 ± 0.00	54448



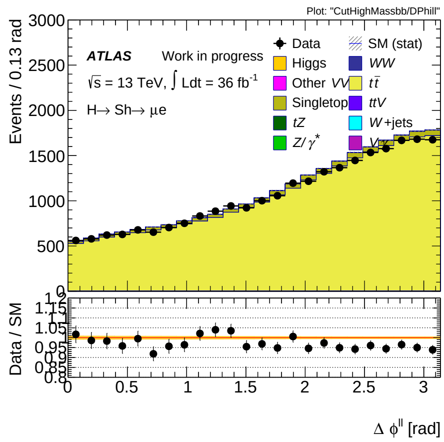
(a)



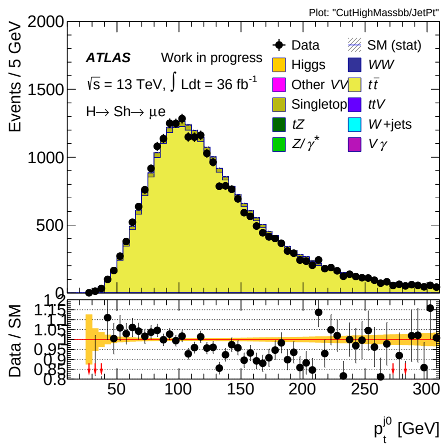
(b)



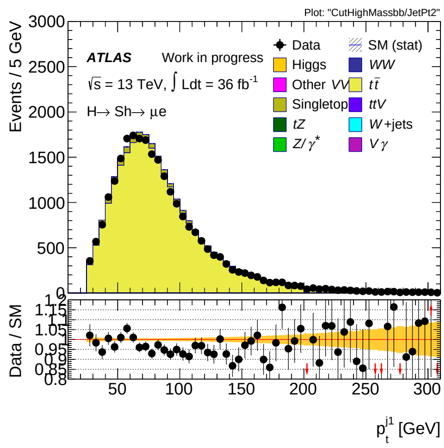
(c)



(d)



(e)



(f)

Figure B.15: Comparison of data and background in the top VR: (a) p_{T}^{l0}, (b) p_{T}^{l0}, (c) m_{T}, (d) $\Delta\phi^{\ell\ell}$, (e) p_{T}^{j0} and (f) p_{T}^{j1}. Uncertainties are statistical.

References

[1] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett.*, vol. B716, pp. 1–29, 2012. 1, 4, 6

[2] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 30 – 61, 2012. 1

[3] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964. 1, 4, 6

[4] P. W. Higgs, "Broken symmetries, massless particles and gauge fields," *Phys. Lett.*, vol. 12, pp. 132–133, 1964. 1, 4, 6

[5] S. von Buddenbrock, N. Chakrabarty, A. S. Cornell, D. Kar, M. Kumar, T. Mandal, B. Mellado, B. Mukhopadhyaya, and R. G. Reed, "The compatibility of LHC Run 1 data with a heavy scalar of mass around 270 GeV," 2015. 1, 11

[6] S. von Buddenbrock, N. Chakrabarty, A. S. Cornell, D. Kar, M. Kumar, T. Mandal, B. Mellado, B. Mukhopadhyaya, R. G. Reed, and X. Ruan, "Phenomenological signatures of additional scalar bosons at the LHC," *The European Physical Journal C*, vol. 76, p. 580, Oct 2016. 1, 11

[7] A. Salam, "*Weak and Electromagnetic Interactions*," *Conf. Proc.*, vol. C680519, pp. 367–377, 1968. 2

[8] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967. 2

[9] A. Astbury, "Found: W and Z," *Physics Bulletin*, vol. 34, pp. 434–436, oct 1983. 4

[10] J. R. Andersen *et al.*, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties," 2013. 6, 8, 9, 10

[11] S. von Buddenbrock, A. S. Cornell, A. Fadol, M. Kumar, B. Mellado, and X. Ruan, "Multi-lepton signatures of additional scalar bosons beyond the Standard Model at the LHC," *J. Phys.*, vol. G45, no. 11, p. 115003, 2018. 11, 13, 15

[12] S. Buddenbrock, A. S. Cornell, Y. Fang, A. Fadol Mohammed, M. Kumar, B. Mellado, and K. G. Tomiwa, "The emergence of multi-lepton anomalies at the LHC and their compatibility with new physics at the EW scale," *JHEP*, vol. 10, p. 157, 2019. 14, 16

[13] L. Evans and P. Bryant, "LHC Machine," *JINST*, vol. 3, p. S08001, 2008. 17

[14] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, "LHC Design Report," *CERN*, 2004. 17

[15] G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003, 2008. 18

[16] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS, Geneva: CERN, 1997. 20

[17] G. Aad *et al.*, "The ATLAS Inner Detector commissioning and calibration," *Eur. Phys. J.*, vol. C70, pp. 787–821, 2010. 20

[18] S. Haywood, L. Rossi, R. Nickerson, and A. Romaniouk, *ATLAS inner detector: Technical Design Report, 2.* Technical Design Report ATLAS, Geneva: CERN, 1997. 21

[19] ATLAS Collaboration, "Technical Design Report for the ATLAS Inner Tracker Pixel Detector," Tech. Rep. CERN-LHCC-2017-021. ATLAS-TDR-030, CERN, Geneva, Sep 2017. 21

[20] A. Abdesselam *et al.*, "The barrel modules of the ATLAS semiconductor tracker," *Nucl. Instrum. Meth.*, vol. A568, pp. 642–671, 2006. 21

[21] A. Abdesselam *et al.*, "The ATLAS semiconductor tracker end-cap module," *Nucl. Instrum. Meth.*, vol. A575, pp. 353–389, 2007. 21

[22] A. Ahmad *et al.*, "The Silicon microstrip sensors of the ATLAS semiconductor tracker," *Nucl. Instrum. Meth.*, vol. A578, pp. 98–118, 2007. 21

[23] E. Abat *et al.*, "The ATLAS TRT barrel detector," *JINST*, vol. 3, p. P02014, 2008. 21

[24] E. Abat *et al.*, "The ATLAS TRT end-cap detectors," *JINST*, vol. 3, p. P10003, 2008. 21

[25] E. Abat *et al.*, "The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: Design and performance," *JINST*, vol. 3, p. P02013, 2008. 21

[26] ATLAS Collaboration, *ATLAS calorimeter performance: Technical Design Report.* Technical Design Report ATLAS, Geneva: CERN, 1996. 22

[27] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1996. 22

[28] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1996. 22

[29] F. Bauer *et al.*, "Construction and Test of MDT Chambers for the ATLAS Muon Spectrometer," *Nucl. Instrum. Meth.*, vol. A461, pp. 17–20, 2001. 23

[30] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1997. 23

[31] T. Argyropoulos *et al.*, "Cathode strip chambers in ATLAS: Installation, commissioning and in situ performance," *IEEE Trans. Nucl. Sci.*, vol. 56, pp. 1568–1574, 2009. 23

[32] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP*, vol. 11, p. 040, 2004. 29

[33] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, "New parton distributions for collider physics," *Phys. Rev.*, vol. D82, p. 074024, 2010. 29

[34] T. Sjostrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015. 29

[35] G. Aad *et al.*, "Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *JHEP*, vol. 09, p. 145, 2014. 30

[36] K. Becker *et al.*, "Optimisation note of the ggF+VBF analysis in $H \to WW^*$ using 36 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV," Tech. Rep. ATL-COM-PHYS-2017-1089, CERN, Geneva, Jul 2017. 33

[37] ATLAS collaboration, "Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data," Tech. Rep. ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016. 34, 37

[38] G. Aad *et al.*, "Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$ =13 TeV," *Eur. Phys. J.*, vol. C76, no. 5, p. 292, 2016. 34, 37

[39] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008. 35

[40] ATLAS Collaboration, "Tagging and suppression of pileup jets with the ATLAS detector," Tech. Rep. ATLAS-CONF-2014-018, CERN, Geneva, May 2014. 35, 36

[41] ATLAS Collaboration, "Expected performance of the ATLAS b-tagging algorithms in Run-2," Tech. Rep. ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015. 35

[42] ATLAS Collaboration, "Commissioning of the ATLAS b-tagging algorithms using $t\bar{t}$ events in early Run-2 data," Tech. Rep. ATL-PHYS-PUB-2015-039, CERN, Geneva, Aug 2015. 35

[43] M. Tanabashi *et al.*, "Particle Data Group," *Phys. Rev.*, vol. D98, no. 3, p. 030001, 2018. 43

[44] A. Denner and M. Pellen, "NLO electroweak corrections to off-shell top-

antitop production with leptonic decays at the LHC," *JHEP*, vol. 08, p. 155, 2016. 60, 102

[45] S. Actis, A. Denner, L. Hofer, J.-N. Lang, A. Scharf, and S. Uccirati, "RECOLA: REcursive Computation of One-Loop Amplitudes," *Comput. Phys. Commun.*, vol. 214, pp. 140–173, 2017. 60, 102

[46] A. Denner, S. Dittmaier, and L. Hofer, "Collier: a fortran-based Complex One-Loop LIbrary in Extended Regularizations," *Comput. Phys. Commun.*, vol. 212, pp. 220–238, 2017. 60, 102

[47] ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012.