Top-quark background estimation for BSM physics search $H \rightarrow Sh \rightarrow 2\ell+$ 2 jets with the ATLAS detector at LHC

Jeremiah Kgomotso Monnakgotla[1], Yesenia Hernández Jiménez[1] and Bruce Mellado[1,2]

[1] School of Physics and Institute for Collider Particle Physics, University of the Witwatersrand, Johannesburg, Wits 2050, South Africa
[2] iThemba LABS, National Research Foundation, PO Box 722, Somerset West 7129, South Africa

E-mail: jeremiah.kgomotso.monnakgotla@cern.ch

Abstract. This paper presents the top-quark background estimation for a heavy scalar H that decays to Standard Model Higgs boson (h) and a Higgs-like scalar (S). This work uses the dataset corresponding to an integrated luminosity of 36 fb^{-1} of pp collisions at center-of-mass energy $\sqrt{s}$ =13 TeV. For this analysis the final state consists of two oppositely charged leptons with different flavour ($e^{\pm}\mu^{\mp}$) and jets. In this analysis the dominant Standard Model background is made of the top-quark processes ($t\bar{t}$ and Wt). The top quark control and validation regions are defined in the 1 b-tagged jet and 2 b-tagged jets phase space, respectively. The construction of the top-quark validation region is defined in order to correct the Monte Carlo mismodeling observed in the lepton kinematics.

1. Introduction

A wide range of results validating Standard Model (SM) processes are published by ATLAS [1, 2] and CMS [3] collaborations pertaining the multi-lepton production at the Large Hadron Collider (LHC) which observe discrepancies between data and Monte Carlo (MC). Excesses can be explained by the Madala hypothesis [4, 5, 6, 7]. In that hypothesis, H decays to Sh which subsequently decays into a wide range of possible final states, dominated by leptons and jets.

This paper presents the top-quark background estimation for the $H \rightarrow Sh$ search using a dataset corresponding to an integrated luminosity of 36 fb^{-1} recorded by the ATLAS detector at center-of-mass energy $\sqrt{s}$ = 13 TeV. The Madala hypothesis uses a simplified model that considers gluon-gluon fusion production mode in association with the heavy scalar (H) that decays into a Higgs-like scalar (S) and the Standard Model (SM) Higgs boson (h). The S boson is assumed to have the either of or Higgs boson's branching ratio depending on its mass (m_S).

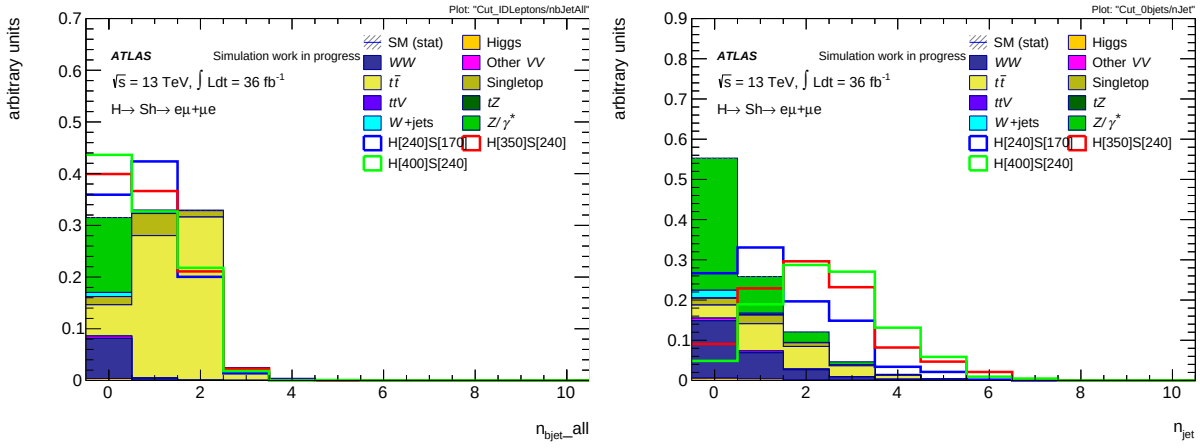
2. Object Selection

The final state consists of two oppositely charged leptons with different flavor and at least two jets. In this analysis, the main physics objects are the electrons, muons, jets and missing transverse momentum ($E^{\text{miss}}_{\text{T, track}}$). Electron candidates must have pseudorapidity range $|\eta| < 2.47$, excluding the transition region between the barrel and endcaps in the LAr calorimeter ($1.37 > |\eta| > 1.52$). The muon candidates must have pseudorapidity range $|\eta| < 2.5$. The leading lepton to satisfy $p_{\text{T}} > 27$ GeV, sub-leading lepton $p_{\text{T}} > 15$ GeV and veto events having additional leptons with $p_{\text{T}} > 10$ GeV. The leading lepton and sub-leading lepton are the leptons with the highest p_{T} and second highest p_{T}, respectively.

The jets are reconstructed by the anti-k_t algorithm [8] with a distance parameter R=0.4 using a cluster of energy deposit in the calorimeters. The jets are required to have $p_{\text{T}} > 25$ GeV within the range $|\eta| < 2.5$. However, jets with $p_{\text{T}} < 60$ GeV are required to have Jet-Vertex-Tagger (JVT) [9] less than 0.59, where JVT is a technique used to separate jets from pile-up events. In Addition, jets originating from b-hadrons are identified using the MV2c10 b-tagging algorithm [10, 11] with an efficiency of 85%. Finally, b-tagged jets are required to have $p_{\text{T}} > 25$ GeV.

3. Signal Region Definition

Phase space with high signal sensitivity and low background contamination determines the signal region (SR). The top-quark background is reduced by selecting events without b-tagged jets as shown in Figure 1 (a). After applying a b-tagged jet veto, the signal is more dominant at ≥ 2 jets region as shown in Figure 1 (b).



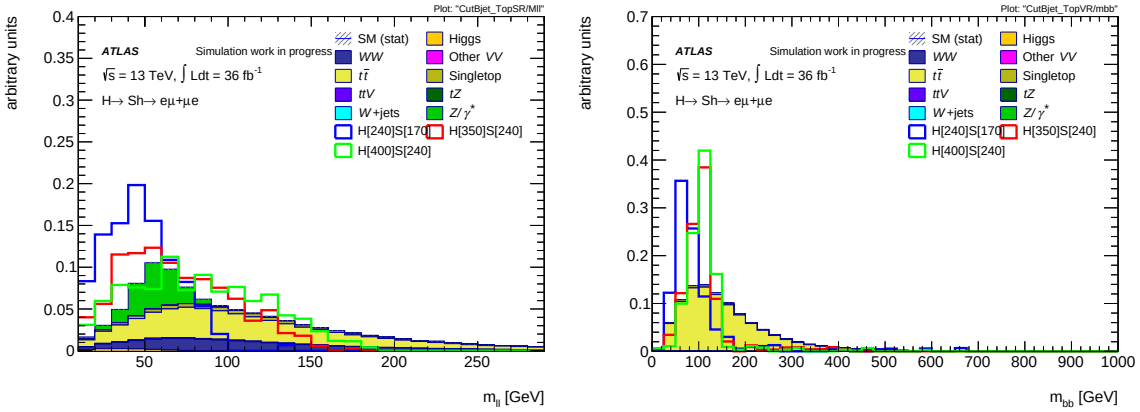
(a) (b)

Figure 1: Shape Comparison of signal and background (both normalized to unity) after selecting events with two high p_{T} leptons: (a) Number of b-tagged jets and (b) Number of jets distributions

4. Top-quark Background Estimation

The top-quark background estimation is based on MC simulation and the experimenta data. In this analysis the top-quark background 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. After applying the requirements discussed in the section 3, the SR is dominated by the top-quark production with 60% contribution. Other relevant background contributions arise from the production of WW, W+jets and Z+jets. The top-quark is estimated by defining a control region (top CR) and a validation region (top VR) to be orthogonal in the SR and required to be signal depleted. The top CR is constructed in order to derive the normalization factors and extrapolate the top-quark contribution to the SR. In order to test the modeling of the top-quark processes a top VR is constructed.

The top CR is selected by applying the SR selection with an additional requirement on the invariant mass of the di-lepton system ($m_{\ell\ell}$) to be greater than 150 GeV in order to achieve a signal depleted region, as shown in Figure 2 (a). The top VR is selected by same SR selection with the exception of the b-tagged jets requirement. 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 2 (b). The event selection summary is presented in Table 1 and the contributions from different SM processes are shown in Table 2



(a) (b)

Figure 2: Comparison shape of signal (normalized to total background) and background after selecting events with two high p_{T} leptons and at least two jets: (a) $m_{\ell\ell}$ and (b) m_{bb} distributions.

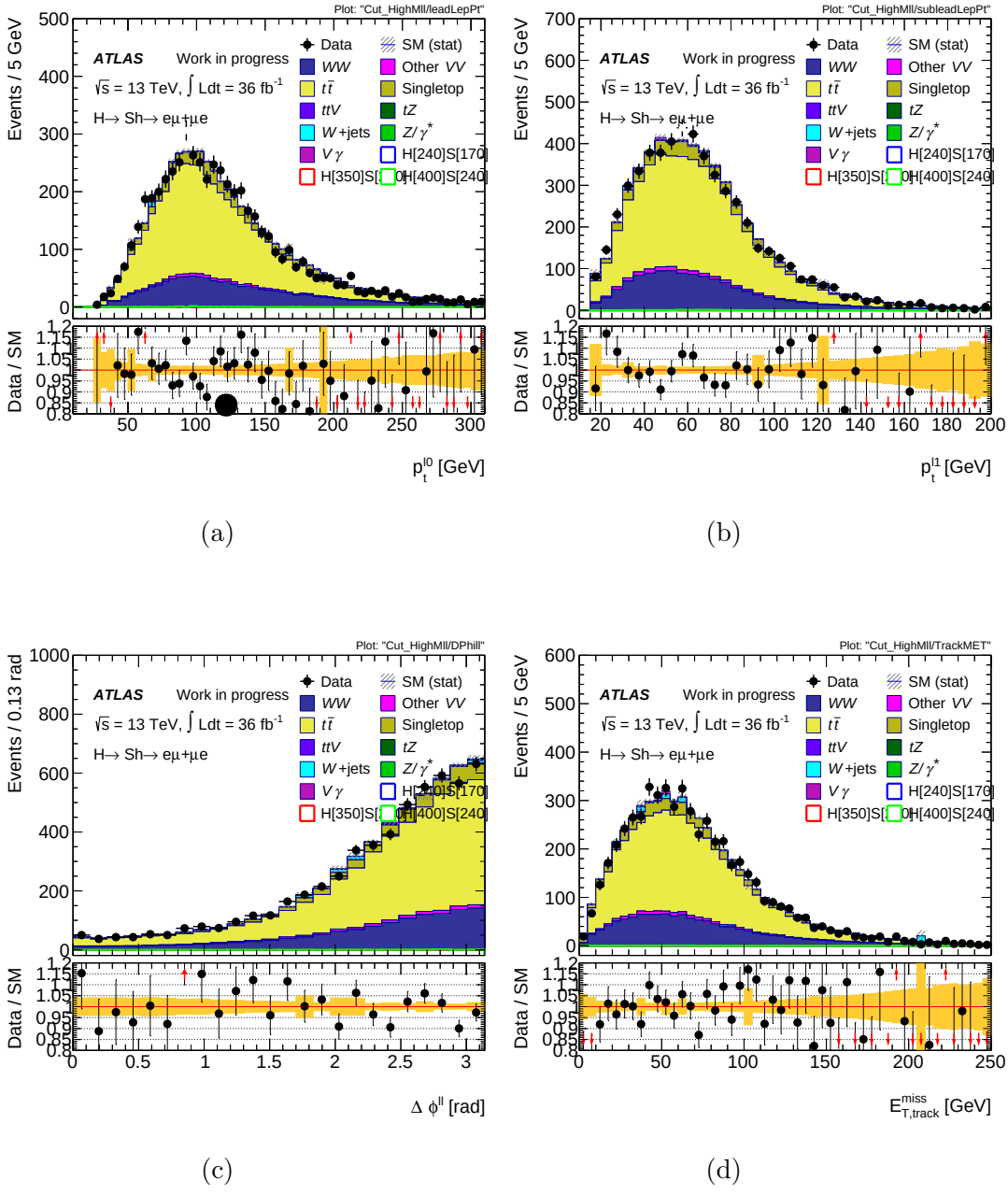
Table 1: Analysis selection for SR, top CR and top VR.

Observable	SR	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	$=0$	$=0$	$=2$
$m_{\ell\ell}$	-	>150 GeV	-
m_{bb}	-	-	>150 GeV

 ISBN: 978-0-620-88875-2

4.1. Kinematic Distributions in the Top Control Region

In this region the agreement between data and MC is good. The ratio of data over MC is 0.97 ± 0.01 with top-quark purity of 72%. The implementation of NLO Electro Weak (EW) corrections (purely based on MC simulation)[12, 13, 14] have a very small impact on the ratio of data over MC, as shown in Figure 3. Other background contributions are from the non-resonant WW which is about 25%, W+jets, other VV, $Z/\gamma*$ processes add up to 3% to the total background.



(a) (b)

(c) (d)

Figure 3: Lepton kinematics of data and MC simulation in the top CR after applying requirements shown in Table 1: (a) Leading lepton p_{T}, (b) Sub-leading lepton p_{T}, (c) Angular separation between two leptons ($\Delta\phi^{\ell\ell}$) and (d) $E^{\mathrm{miss}}_{\mathrm{T,track}}$. Uncertainties are statistical only.

4.2. Kinematic Distributions in the Top Validation Region

In this region the agreement between data and MC is reasonable. The ratio of data over MC is 0.95 with top-quark puritiy of 99%. Other non-top-quark background contributions add up to 1%. The implementation of the NLO EW corrections sligthly improves the agreement between data and MC, as shown in the tail of the distributions in Figure 4.

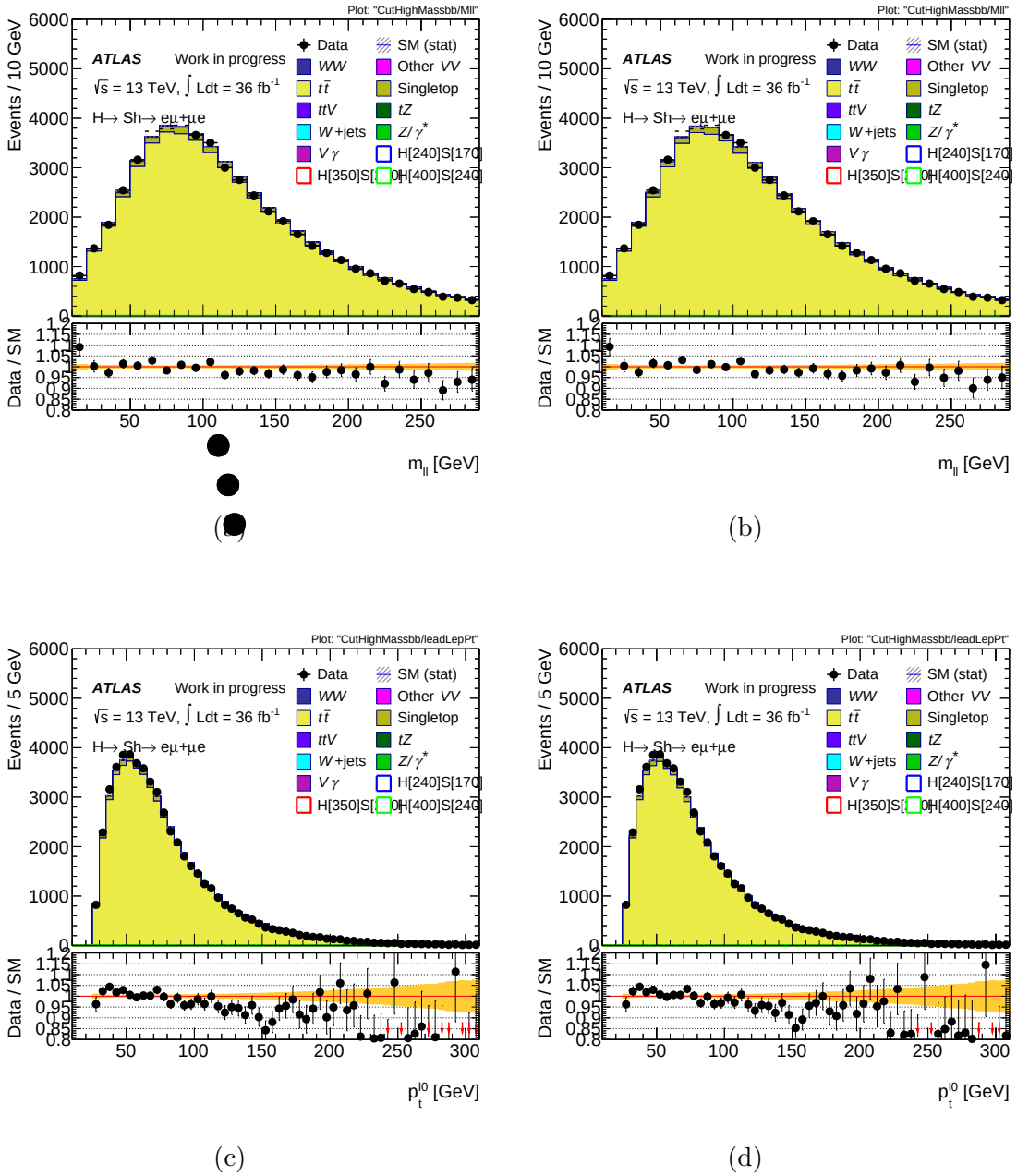


Figure 4: Comparison of data and MC in the top CR after applying requirements shown in Table 1: (a) Di-lepton invariant mass (b) Di-lepton invariant mass after NLO EW corrections (c) Leading lepton p_{T}, (d) Leading lepton p_{T} after NLO EW corrections. Uncertainties are statistical only.

Table 2: Number of events observed in data, compared to the numbers of predicted background events. The analysis is blinded in the SR. Uncertainties are statistical only.

	SR	top CR	top VR
SM Higgs	299.97 ± 1.56	3.19 ± 0.08	25.29 ± 0.24
WW	4789.12 ± 13.38	1355.89 ± 7.01	33.05 ± 0.84
Other VV	421.43 ± 3.63	105.01 ± 1.69	4.56 ± 0.27
$t\bar{t}$	12594.71 ± 32.42	3534.73 ± 17.30	52658.56 ± 64.70
Singletop	1630.74 ± 15.97	469.49 ± 8.66	2134.64 ± 17.98
ttV	22.32 ± 0.41	7.74 ± 0.24	66.83 ± 0.68
tZ	1.96 ± 0.04	0.64 ± 0.02	2.01 ± 0.04
W+jets	115.33 ± 32.44	40.44 ± 20.30	0.00
Z/γ^*	4716.15 ± 57.43	17.77 ± 10.83	62.27 ± 3.76
Total Bkg	24887.97 ± 78.18	5561.66 ± 31.55	54991.04 ± 67.29
Data	-	5376	52217
Data/MC	-	0.97 ± 0.01	0.95 ± 0.00

5. Conclusion

This proceeding presents the top-quark background estimation for the $H \to Sh$ search. This analysis is characterized by two oppositely charged leptons with different flavour and at least two jets. For this search the top-quark is the dominant SM background with 60% contribution in the SR. The top-quark background is normalized in the 0 b-tagged jets bin, with the ratio of data over MC is approximately 0.97 ± 0.01 and top-quark purity of 72%. The modeling of top-quark process is tested in the 2 b-tagged jets bin, the ratio of data over MC is approximately 0.95 and top-quark purity of 99%. Finally, the NLO EW corrections applied on the top-quark MC slightly improve the agreement between data and MC in the top CR and VR.

6. References

[1] ATLAS Collaboration 2016 *JHEP* **09** 029 arXiv:1603.01702
[2] ATLAS Collaboration, Measurement of fiducial and differential W^+W^- production cross-sections at $\sqrt{s}$ =13 TeV with the ATLAS detector 2019 arXiv:1905.04242
[3] CMS collaboration 2018 *JHEP* **10** 117 arXiv:1805.07399
[4] von Buddenbrock S. et al. The compatibility of LHC Run 1 data with a heavy scalar of mass around 270 GeV arXiv:1506.00612 [hep-ph]
[5] von Buddenbrock S. et al. 2016 *Eur. Phys. J.* **C76** 580 arXiv:1606.01674 [hep-ph]
[6] von Buddenbrock S. et al. 2018 *J. Phys.* **G45** 115003 arXiv:1711.07874 [hep-ph]
[7] von Buddenbrock S. et al. The emergence of multi-lepton anomalies at the LHC and their compatibility with new physics at the EW scale 2019 arXiv:1901.05300 [hep-ph]
[8] Cacciari M, Salam G. P., Soyez G 2008 *JHEP* **04** 063 arXiv: 0802.1189 [hep-ph]
[9] ATLAS Collaboration, Tagging and suppression of pileup jets with the ATLAS detector, ATLAS-CONF-2014-018, 2014, https://cds.cern.ch/record/1700870
[10] ATLAS Collaboration, Expected performance of the ATLAS b-tagging algorithms in Run-2, ATL-PHYS-PUB-2015-022, 2015, https://cds.cern.ch/record/2037697
[11] ATLAS Collaboration, Commissioning of the ATLAS b-tagging algorithms using $t\bar{t}$ events in early Run 2 data, ATL-PHYS-PUB-2015-039, 2015, https://cds.cern.ch/record/2047871
[12] Denner A., Mathieu P. 2016 *JHEP* **08** 155 arXiv:1607.05571 [hep-ph]
[13] Actis S. et al. 2017 *Comput. Phys. Commun.* **214** 140-173 arXiv:1605.01090 [hep-ph]
[14] Denner A. et al. 2017 *Comput. Phys. Commun.* **212** 220-238 arXiv:1604.06792 [hep-ph]