Simulation Validation and Background Estimation for Singly and Doubly Charged Higgs Boson Searches in ATLAS

University of Victoria
Dept. of Physics & Astronomy
CERN Summer Student Programme
IPP Summer Student Fellowship

By

Denaisha Kraft
University of Victoria, Dept. of Physics & Astronomy
denaisha.k@gmail.com
August 23, 2024

Performed at:
University of Victoria, Victoria, Canada
and
CERN, Geneva, Switzerland

Supervisor(s): Heather Russell, John McGowan

Abstract

Measurements of rare physics processes using the ATLAS detector require a large number of Monte Carlo (MC) simulated events to ensure a low statistical uncertainty on the MC prediction. GEANT4 is able to accurately simulate the detector response, but it requires significant CPU resources. AtlFast3 (AF3) is an updated fast simulation tool that is less CPU-intensive. In this report, comparisons are performed for the modelling of various kinematic variables in Run 2 MC simulations of vector boson fusion (VBF) charged Higgs boson production. For these comparisons, the detector is simulated with GEANT4 and AF3. Good agreement is seen between AF3 and GEANT4 in the signal and control regions, with only a small discrepancy observed in the high jet p_T region for the $W^{\pm}Z$ channel. This discrepancy showed no correlation with any other observables. In a separate study, same-sign $W^{\pm}W^{\pm}$ backgrounds are investigated. Non-prompt and fake lepton backgrounds are presented for $t\bar{t}$ and W+jets background samples. These studies provide estimates on the type of non-prompt and fake leptons that are present, and are useful for data-driven estimates of non-prompt backgrounds.

Contents

1 Introduction

1.1 Charged Higgs Boson Searches

Investigating whether the Higgs boson is part of an extended Higgs sector is a current area of research for physics beyond the Standard Model (SM). Extended Higgs sectors, which include additional complex doublets or higher-isospin scalar fields, predict the presence of charged Higgs bosons. In the general two-Higgs doublet model, CP-invariance prevents the charged Higgs boson from decaying into W or Z bosons by preventing tree-level coupling [1]. The Georgi-Machacek (GM) model includes additional isotriplet scalar fields, which permits tree-level coupling to massive vector bosons, such as W and Z bosons. The GM model extends the SM by including one real and one complex triplet, and in the GM model, two isospin triplet scalar fields are added to the SM doublet. This gives rise to 5-plet states corresponding to physical Higgs bosons. The production of $H_5^{\pm}$ and $H_5^{\pm\pm}$ scalars is via vector boson fusion (VBF) in the GM model.

This search is motivated by excesses seen in Run 2 of the LHC (2015-2018), which corresponded to 140 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector [1]. For the $H^{\pm} \to W^{\pm}Z$ channel, the largest excess was seen for $m_H = 375$ GeV. This was a local significance of 2.8 standard deviations, and a global significance of 1.6 standard deviations. For the $H^{\pm\pm} \to W^{\pm}W^{\pm}$ channel, the largest excess was seen for $m_H = 450$ GeV, with a local and global significance of 3.2 and 2.5 standard deviations, respectively.

1.1.1 Vector Boson Fusion and Vector Boson Scattering

Both vector boson fusion (VBF) and vector boson scattering (VBS) are direct probes of SM weak boson interactions. VBF refers to the production of a single boson [2], which in this search involves a charged Higgs boson that further decays into two massive vector bosons (W or Z). VBS refers to the production of a pair of bosons in the final state. Studying VBS ($VV \to VV$) where V is either a W or Z boson, probes the mechanism of electroweak (EW) symmetry breaking in the SM. VBS involves the self-interactions of the electroweak vector bosons containing a three- or a four-boson vertex, and at leading order it involves two initial quarks that each radiate a vector boson [3]. These bosons then interact and decay, and the two outgoing quarks fragment close to the beam direction. This leads to a final state consisting of two vector bosons and two jets, denoted as $VVjj$. With the trajectory of the incoming quarks only being altered slightly when the gauge bosons are radiated, the two associated jets are generally in forward regions of the detector, close to the beam direction. The process is initiated by electroweakly-interacting quarks, and the outgoing quarks hadronize into a spray of particles known as jets.

VBS and VBF only involve EW-interaction vertices (EW $VVjj$), but QCD processes (QCD $VVjj$) can give rise to the same signature in the detector. This analysis searches for the electroweak production of the charged Higgs boson by VBF, where the Higgs boson then decays into massive vector bosons. For the singly charged Higgs boson search, the $W^{\pm}Z$ channel is used, and for the doubly charged Higgs boson search, the same-sign $W^{\pm}W^{\pm}$ channel is used. Figure 1 shows the representative diagrams for the production and decay of the charged Higgs bosons through VBF.

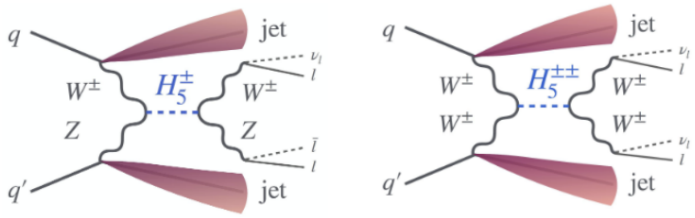


Figure 1: Diagrams for the production and decay of $H^{\pm}$ (left) and $H^{\pm\pm}$ (right) states produced through vector boson fusion.

1.1.2 Same-Sign $W^{\pm}W^{\pm}$ and $W^{\pm}Z$ Channels

The final states for these analyses consist of the leptonic decays of two massive vector bosons produced in association with two jets, denoted as $VVjj$. A more detailed explanation of the signal regions (SR) and control regions (CR) is found in Sections 2.2.1 and 2.2.2 for the $W^{\pm}W^{\pm}$ and $W^{\pm}Z$ channels, respectively. For the $W^{\pm}Z$ SR, three charged leptons are required. The $W^{\pm}W^{\pm}$ SR requires a same-charge lepton pair. Due to the presence of neutrinos in the final state of both channels, requirements are set on the magnitude of the missing transverse momentum, E_T^{miss}. With the VBF topology, the final state is characterized with at least two jets with a large invariant mass, m_{jj}, and a large absolute difference in rapidity, $|\Delta y_{jj}|$ [1].

The SM production of the $VVjj$ final state is the dominant background for these searches [1]. A further discussion of backgrounds is found in Section 3. The production of this final state at leading-order has contributions from modes that involve only electroweak interaction vertices (EW $VVjj$) and modes involving strong interaction vertices (QCD $VVjj$). Jets misidentified as leptons and leptons from hadron decays are known as fake and non-prompt leptons. These lepton backgrounds contribute to the signal region, as the final state is mimicked by these background processes. In the $W^{\pm}W^{\pm}$ SR, these non-prompt backgrounds arise mainly from W+jets and semileptonic $t\bar{t}$ processes [1]. Analysis of these backgrounds is discussed in further detail in Section 3.2. Other non-prompt backgrounds can arise from electron charge misidentification and photon conversions.

2 AtlFast3 and GEANT4 Comparison

To perform precision measurements and search for new particles and new interactions, large and accurate datasets of simulated Monte Carlo (MC) events are required. Events are prepared by generating the desired physics process, simulating interactions with detectors, digitizing the detector response, then reconstructing the events with the use of custom algorithms.

The accordion geometry of the ATLAS electromagnetic calorimeter makes shower development simulation particularly CPU-intensive when using the GEANT4 toolkit (FullSim). GEANT4 is able to accurately simulate events and is currently used for analysis in ATLAS, but less CPU-intensive and faster approaches are being developed. These faster approaches for calorimeter simulation are essential for reducing CPU usage. AtlFast3 (AF3) combines parameterized approaches with machine-learning techniques to meet the simulation needs of the ATLAS experiment [4]. AF3 is being used for a large resimulation campaign of Run 2 MC events, and is planned to be used for Run 3 and beyond.

2.1 Overview of AtlFast3 and GEANT4 Simulations

AF3 is an updated fast simulation tool that follows a previously deployed tool, called AtlFastII (AF2). AF2 has limitations in modelling jets and their detailed substructure, while AF3 has significant improvements in the modelling of jet substructure. AF3 has the same CPU performance as AF2, but improved accuracy in reproducing GEANT4. AF3 consists of two parametric calorimeter simulations, known as FastCaloSim V2 and FastCaloGAN [4]. These simulate the energy of a particle shower as a single step instead of simulating every particle inside the calorimeter, which makes the CPU performance independent of the particle energy. AF3 combines these different simulation tools; GEANT4 is used to simulate muons in all detectors and particles in the inner detector, FastCaloSim V2 is used to simulate electrons and photons of all energies, and FastCaloGAN is used to simulate hadrons with specific energies in the calorimeter. AF3 has a highly accurate performance, which allows it to be used to simulate a large number of events for various physics processes.

This section focuses on the validation of AF3 compared to GEANT4 (FullSim). Comparisons are made for MC campaigns corresponding to Run 2 datasets from 2015-2018. The validation is performed for singly charged $H^{\pm}$ and doubly charged $H^{\pm\pm}$ Higgs boson searches via $W^{\pm}Z$ and same-sign $W^{\pm}W^{\pm}$ channels, respectively. Mass points ranging from 450 GeV to 3000 GeV are simulated and compared for both the singly and doubly charged Higgs boson samples.

2.2 Selection Regions

The event selections for the signal region (SR) and control region (CR) requirements used for the simulation comparisons are outlined below. The signatures of each channel are also discussed below.

2.2.1 Same-Sign $W^{\pm}W^{\pm}$ Channel Selections

$W^{\pm}W^{\pm}jj$ events are characterized by the presence of two high-energy forward jets located in opposite hemispheres, a same-sign charged-lepton pair of either muons or electrons, and missing transverse momentum, E_T^{miss}. In this search, selected events involve the W bosons decaying into a lepton and a neutrino. The emission of a vector boson from the initial quark line causes the final-state quarks to have large momenta and have high absolute values of rapidity [5]. This leads to the two jets having a large invariant mass, m_{jj}, and a large absolute difference in rapidity, $|\Delta y_{jj}|$.

For each event, two signal leptons with the same charge are required. Each lepton must have a transverse momentum $p_T > 27$ GeV, and the dilepton invariant mass is required to be above 20 GeV. A third lepton veto is applied to suppress backgrounds from processes with more than two leptons in the final state, such as $W^{\pm}Z$ and ZZ. The presence of two neutrinos in the final state leads to a large E_T^{miss}, so events must satisfy $E_T^{miss} > 30$ GeV. At least two jets are required for each event, with the requirement that the transverse momentum, p_T, exceeds 65 GeV and 35 GeV for the leading and subleading jet, respectively. Additional jets are required to have $p_T > 25$ GeV, and all jets are required to have an absolute pseudorapidity $|\eta| < 4.5$. For the SR, the dijet invariant mass must satisfy $m_{jj} > 500$ GeV. This and the $|\Delta y_{jj}|$ requirement aim to limit the background contribution from QCD-induced processes [3]. For the CR, the dijet invariant mass is set to be within the bounds of $200 < m_{jj} < 500$ GeV. Table 1 outlines the SR and CR selections used for the simulation comparisons in this report.

Table 1: Signal region and control region selections used for the same-sign $W^{\pm}W^{\pm}$ channel.

Requirement	SR	Low-m_{jj} CR
Leading and subleading lepton p_T	> 27 GeV	
Electron $\|\eta\|$	< 2.47 (1.37 in ee), excluding $1.37 \le \|\eta\| \le 1.52$	
Muon $\|\eta\|$	< 2.5	
Leading (subleading) jet p_T	> 65 (35) GeV	
Additional jet p_T	> 25 GeV	
Jet $\|\eta\|$	< 4.5	
$m_{\ell\ell}$	> 20 GeV	
E_T^{miss}	> 30 GeV	
Charge misid. $Z \to ee$ veto	$\|m_{ee} - m_Z\| > 15$ GeV	
m_{jj}	> 500 GeV	$200 < m_{jj} < 500$ GeV
$\|\Delta y_{jj}\|$		> 2

2.2.2 $W^{\pm}Z$ Channel Selections

Similarly to the same-sign $W^{\pm}W^{\pm}$ channel, the $W^{\pm}Z$ VBF channel is characterized by the presence of two jets with a large absolute difference in rapidity. The final state consists of three charged leptons and missing transverse momentum. Selected events must have a Z candidate, which is defined by two leptons of the same flavour and opposite charge with an invariant mass that is consistent with the Z mass: $|m_{\ell\ell} - m_Z| < 20$ GeV [6]. If there is more than one pair that is a Z candidate, the one with the invariant mass closest to the Z boson mass is chosen. The remaining third lepton is taken to be the W boson lepton candidate. For selected events, the missing transverse momentum, E_T^{miss}, is required to be greater than 25 GeV. Analysis and simulation validation was done for the different lepton channels $\mu Z\mu W$, μZeW, $eZeW$, and $eZ\mu W$. These four channels correspond to the possible combinations of leptons and whether they are consistent with coming from the Z or W boson based on charge and invariant mass.

Electron and muon candidates have identification and isolation criteria applied to them, which is either *loose*, *medium*, or *tight* [6]. These correspond to increasing orders of background rejection. Each

working point uses the same set of variables for identification, but the requirements become more rigorous for the tighter working points. This identification criteria is based on shower shapes and track parameters for electrons, and on track parameters for muons. *Baseline* leptons satisfy the *loose* identification and isolation criteria. *Loose* and *tight Z* and *tight W* leptons are *baseline* leptons, and *tight Z* leptons satisfy the *medium* identification and *tight* isolation criteria. *Tight W* leptons satisfy the *tight* identification and *tight* isolation criteria [6]. Table 2 highlights the baseline selection used. Lepton p_T was selected to be > 25 GeV, and jet selection was set to be the same as shown in Table 1. For the CR, $|\Delta y_{jj}| < 3.5$ was used, and for the SR, $|\Delta y_{jj}| > 3.5$ was used. Events were also required to satisfy $m_{jj} > 500$ GeV.

Table 2: Baseline selections used for $W^{\pm}Z$ signal and control regions.

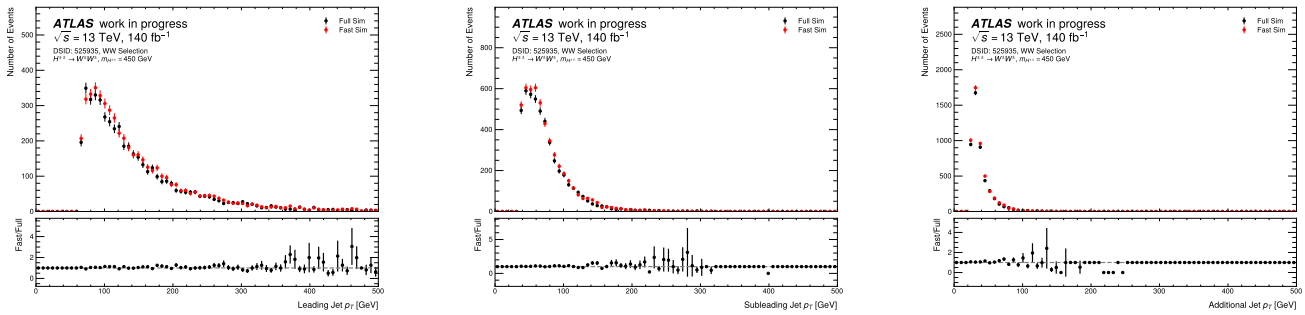
Baseline *WZ* selection
Event cleaning and primary vertex
Single-electron or single-muon trigger
Exactly 3 *Loose* leptons (e or μ) with p_T >25 GeV (p_T >27 GeV for the trigger-matched lepton)
ZZ veto: veto events with additional *Baseline* leptons
Z candidate: A *Tight Z* same-flavour-opposite-sign lepton pair with $\lvert m_{\ell\ell} - m_Z \rvert < 20$ GeV
W candidate: *Tight W* lepton requirements on 'non-*Z* leptons' and $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV

2.3 Simulation Comparisons

The following sections show a sample of selected plots that compare AF3 and FullSim. Additional comparison plots are found in Appendix A. The full set of signal and control region comparisons for all mass points is linked here.

2.3.1 $W^{\pm}W^{\pm}$ Channel Results

Plots were obtained for mass points ranging from $m_{H^{\pm\pm}} = 450$ GeV to 3000 GeV. In total, six $H^{\pm\pm}$ mass points were compared. The comparisons were performed using both signal region and control region selections, and the selection process used for these regions is outlined in Section 2.2.1. Figure 2 shows sample SR plots for $m_{H^{\pm\pm}} = 450$ GeV. The leading and subleading jet p_T are shown, as well as the additional jet p_T if an additional jet is present. No major discrepancies are noticed in these comparisons for this sample. CR plots for the same kinematic variables showed similar distributions and revealed no major discrepancies. Comparisons were also made for the pseudorapidity, η, for all mass points. The η plots for all mass points showed strong agreement between AF3 and FullSim.

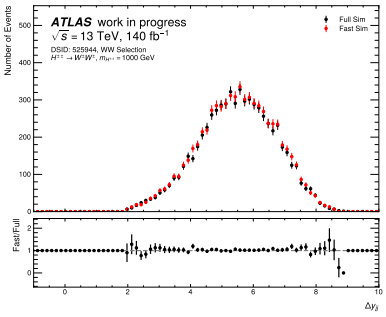


(a) Leading jet p_T. (b) Subleading jet p_T. (c) Additional jet p_T.

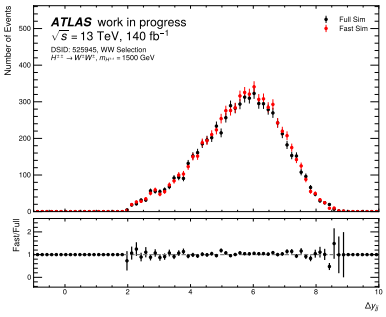
Figure 2: Signal region comparison plots for leading, subleading, and additional jet p_T for $m_{H^{\pm\pm}} = 450$ GeV for the same-sign $W^{\pm}W^{\pm}$ channel.

The difference in jet rapidity between the leading and subleading jet for each sample was also plotted. Figure 3 shows SR and CR plots for three different $m_{H^{\pm\pm}}$ mass points. These mass points are $m_{H^{\pm\pm}} =$ 1000 GeV, 1500 GeV, and 2000 GeV. Figures 3a, 3b, and 3c show the SR comparison, and Figures 3d, 3e, and 3f show the comparison using CR selections for the three different mass points. As shown in the comparisons, for the same-sign $W^{\pm}W^{\pm}$ channel both the CR and SR have good agreement between AF3 and FullSim in jet variables. No significant discrepancies were present in any kinematic variables

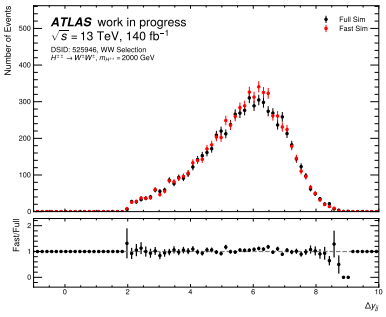
that were compared. For all six $H^{\pm\pm}$ mass points used, plots were made for the leading and subleading W lepton p_T, the dijet invariant mass, m_{jj}, the dilepton invariant mass, $m_{\ell\ell}$, and the missing transverse momentum, E_T^{miss}. All of these distributions validated the accurracy of AF3 compared to FullSim, and good agreement is observed. Sample SR plots of E_T^{miss} are shown in Appendix A.



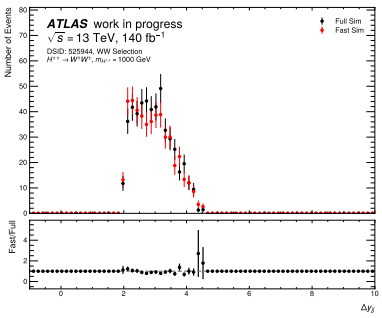
(a) $m_{H^{\pm\pm}} = 1000$ GeV signal region plot of Δy_{jj}.



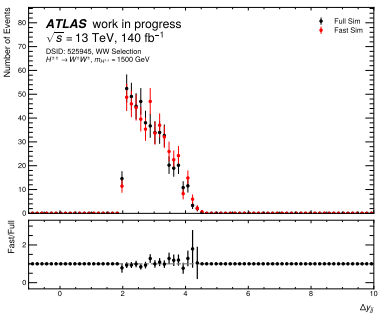
(b) $m_{H^{\pm\pm}} = 1500$ GeV signal region plot of Δy_{jj}.



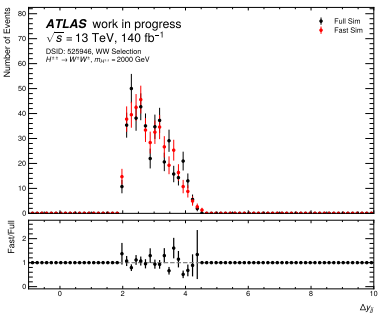
(c) $m_{H^{\pm\pm}} = 2000$ GeV signal region plot of Δy_{jj}.



(d) $m_{H^{\pm\pm}} = 1000$ GeV control region plot of Δy_{jj}.



(e) $m_{H^{\pm\pm}} = 1500$ GeV control region plot of Δy_{jj}.



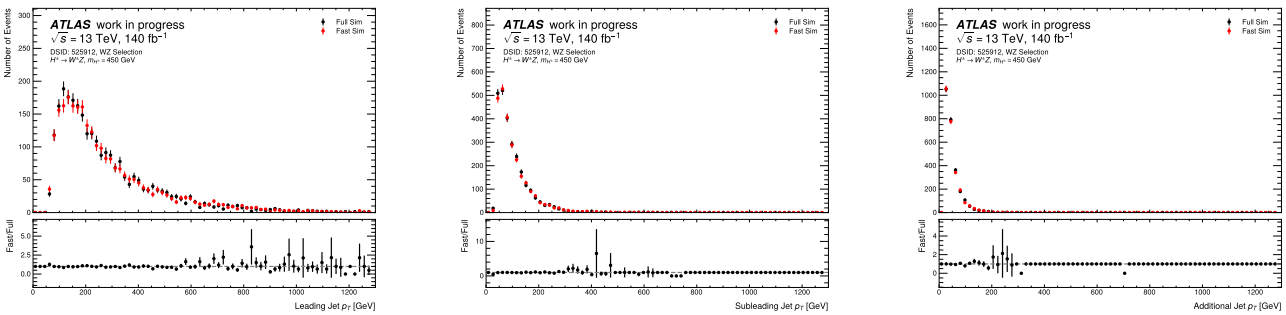
(f) $m_{H^{\pm\pm}} = 2000$ GeV control region plot of Δy_{jj}.

Figure 3: Signal region and control region difference in jet rapidity, Δy_{jj}, comparison plots for three $H^{\pm\pm}$ mass points (1000 GeV, 1500 GeV, 2000 GeV) for the same-sign $W^{\pm}W^{\pm}$ channel.

Figure 2 and Figure 3 only show a sample of some mass points tested during this analysis. Six total $H^{\pm\pm}$ mass points were tested, and similarly to the results shown here, good agreement between AF3 and FullSim was observed for this channel. AF3 was successfully validated in comparison to FullSim MC simulations for the same-sign $W^{\pm}W^{\pm}$ channel.

2.3.2 $W^{\pm}Z$ Channel Results

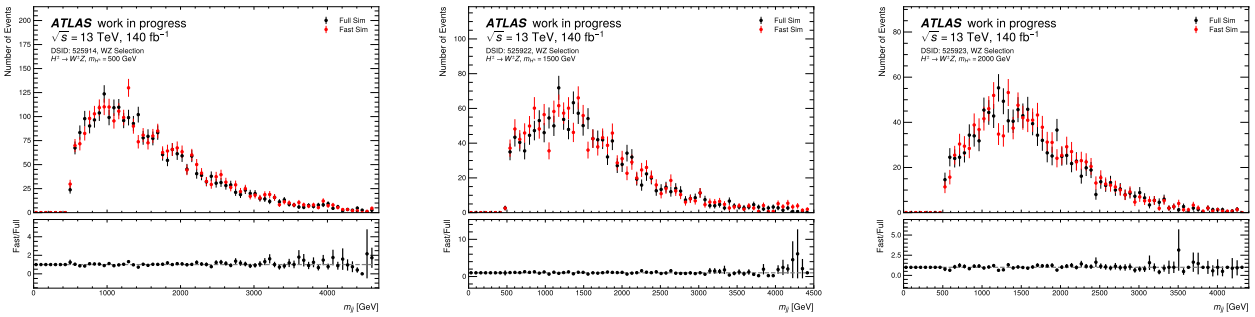
Simulation comparisons were also performed for the $W^{\pm}Z$ channel, which simulates the production of a singly-charged Higgs boson, $H^{\pm}$. The following plots in Figure 4 are for jet p_T for the mass point $m_{H^{\pm}} = 450$ GeV, but five additional mass points were also compared. These are for the leading and subleading jet, as well as the additional jet if an additional jet is present. The signal and control region selections used for this channel are outlined in Section 2.2.2. Figure 4 shows a sample for just the signal region, and good agreement is seen between AF3 and FullSim for the jet variables. Plots comparing η are not shown, but these also showed no major discrepancies between the two simulations.



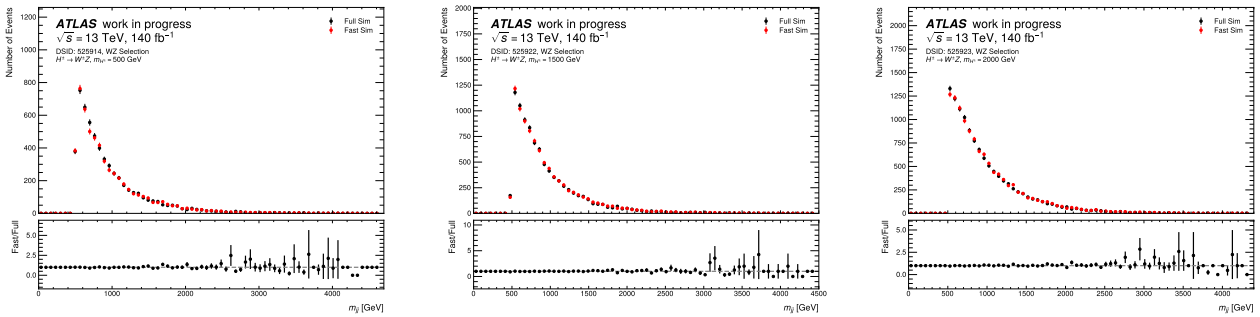
(a) Leading jet p_T. (b) Subleading jet p_T. (c) Additional jet p_T.

Figure 4: Signal region comparison plots for leading, subleading, and additional jet p_T for $m_{H^\pm} = 450$ GeV for the $W^\pm Z$ channel.

Additional comparisons were made for lepton p_T for the $eZeW$, $eZ\mu W$, $\mu Z\mu W$, and μZeW lepton channels, and no discrepancies were observed for the leading Z, subleading Z, or W lepton in any of the control or signal region plots. Δy_{jj} was also compared, and good agreement was seen between AF3 and FullSim. Figure 5 shows a sample of three $H^\pm$ mass points for signal region and control regions plots of the dijet invariant mass, m_{jj}. The three mass points shown are $m_{H^\pm}$ masses of 500 GeV, 1500 GeV, and 2000 GeV. No large discrepancies are evident in any of the signal and control region comparisons between AF3 and FullSim, and the expected kinematic distributions are observed.



(a) $m_{H^\pm} = 500$ GeV signal region plot of m_{jj}. (b) $m_{H^\pm} = 1500$ GeV signal region plot of m_{jj}. (c) $m_{H^\pm} = 2000$ GeV signal region plot of m_{jj}.



(d) $m_{H^\pm} = 500$ GeV control region plot of m_{jj}. (e) $m_{H^\pm} = 1500$ GeV control region plot of m_{jj}. (f) $m_{H^\pm} = 2000$ GeV control region plot of m_{jj}.

Figure 5: Signal region and control region dijet invariant mass, m_{jj}, comparison plots for three $H^\pm$ mass points (500 GeV, 1500 GeV, 2000 GeV) for the $W^\pm Z$ channel.

Similar to the same-sign $W^\pm W^\pm$ channel, six total mass points were compared for the $W^\pm Z$ channel. Plots for the dilepton invariant mass, $m_{\ell\ell}$, trilepton invariant mass, $m_{\ell\ell\ell}$, and missing transverse momentum, E_T^{miss}, were made for the signal region and control region. Sample signal region plots for E_T^{miss} are shown in Appendix A for three different $m_{H^\pm}$ values. The largest discrepancy between AF3 and

FullSim for the $W^{\pm}Z$ channel was identified in leading jet p_T comparisons for high p_T bins. All other kinematic distributions were successfully shown to be consistent between the two MC simulations. The leading jet p_T discrepancy is further discussed in the next section.

2.3.3 High Jet p_T Region: $W^{\pm}Z$ Channel

Overall, simulated $H^{\pm}$ and $H^{\pm\pm}$ events in the VBF channel saw no significant disagreement between AF3 and FullSim. The largest difference seen was observed in jet p_T bins $900 - 1200$ GeV in the $W^{\pm}Z$ channel. Figure 6 shows a comparison for the $W^{\pm}Z$ control region for a mass of $m_{H^{\pm}} = 2000$ GeV.

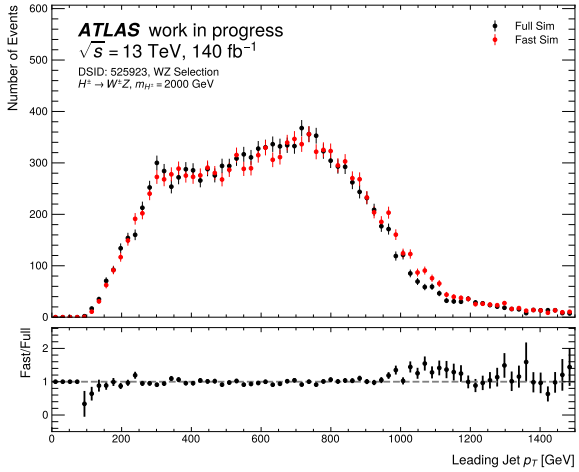


Figure 6: Simulation comparison showing deviation in high jet p_T region for the $W^{\pm}Z$ channel.

The discrepancy was further investigated, and it was found to not show any correlation with other observables, such as η. To identify if the disagreement correlated with a disagreement in η, 2D histograms were plotted for jet p_T and η, but no correlation was found. Despite this discrepancy, all kinematic distributions for both the $W^{\pm}Z$ and same-sign $W^{\pm}W^{\pm}$ channels show good agreement between AF3 and FullSim.

3 Same-Sign $W^{\pm}W^{\pm}$ Background Analysis

Background processes mimic the final state of the channels in the analysis, gaining entry into the signal region of the measurement. The analysis performed on background sources was done for the same-sign $W^{\pm}W^{\pm}$ channel, so the focus of the following sections will be on this channel. The primary focus of the work done was on non-prompt and fake background sources.

3.1 Overview of Backgrounds

Background processes can be classified into various categories. One of these categories is prompt leptons [3]. This background corresponds to SM processes that have final states with two prompt leptons with the same electric charge. One source of this background is QCD $W^{\pm}W^{\pm}jj$, which arises from strong production and has the same final state as the EW $W^{\pm}W^{\pm}jj$ signal. Additionally, the process $W^{\pm}Zjj \rightarrow \ell^{\pm}\nu\ell^{\pm}\ell^{\pm}jj$ is one of the dominant background sources in this analysis. In this process, the Z lepton is out of the detector acceptance region or is not successfully identified. This mimics the $\ell^{\pm}\ell^{\pm} + E_T^{miss} + 2$ jet signature of the signal region [3]. Top backgrounds include $t\bar{t}Z$, $t\bar{t}W$, and tZ processes where some objects in the event are not correctly identified.

Charge misidentification and $V\gamma jj$ processes also contribute to the background. With charge misidentification, the final state consists of two opposite-sign leptons and the electric charge of one of the leptons is incorrectly identified. $V\gamma jj$ processes involve a heavy gauge boson and a prompt photon. The boson decays leptonically and the photon converts into an e^+e^- pair. If the leptons are not properly identified

or reconstructed, this can mimic the final state of the signal.

Non-prompt and fake leptons also contribute to background. This is comprised of leptons from secondary decays and jets that are misidentified as leptons. For electrons, the reconstructed candidates consist of prompt electrons, converted photons, non-prompt electrons, and fake electrons. Prompt electrons are electrons originating from the primary vertex, and are typically from a hard scatter. A hard scatter is a process in which the colliding partons are energetic enough to cause an inelatic collision [3], and the primary vertex is the vertex with the highest scalar sum of the squared transverse momenta of the tracks associated with it [5]. Converted photon electrons result from a photon when it converts into an e^+e^- pair, non-prompt electrons are candidates originating from the decays of hadrons, and fake electrons are misidentified electrons that usually are a jet that has been incorrectly reconstructed. W+jets and $t\bar{t}$ processes are the major sources of the non-prompt and fake background, and are the main focus of the analysis done in this report. While prompt backgrounds are estimated using Monte Carlo simulations, backgrounds from the misidentification of jets as photons, jets as electrons, and electrons as photons are estimated using data-driven techniques [7]. This uses data control regions to estimate the background. The following section discusses these backgrounds in further detail.

3.2 Non-Prompt and Fake Leptons

Even with the identification and isolation selections present in ATLAS, non-prompt and fake lepton background can still occur due to the misidentification or improper reconstruction of leptons. As briefly outlined in Section 3.1, non-prompt leptons originate from a secondary decay instead of the primary vertex, but they are still actual leptons. Fake leptons consist of objects that are incorrectly reconstructed as leptons.

A major source of non-prompt leptons is a jet initiated by a bottom quark, where the jet is known as a b-jet. With this process, the hadronization of a b-quark produces a b-hadron [3]. These b-hadrons have high decay multiplicities of charged particles, so for the signal region, events containing b-jets are removed. The leptonic decays of W+jets and the semileptonic decays of $t\bar{t}$ are the dominant sources of non-prompt and fake leptons in the same-sign $W^{\pm}W^{\pm}$ signal region. These processes are shown in Figure 7. This is due to the presence of one prompt and one non-prompt or fake lepton with jets and E_T^{miss}, which mimics the final state of the signal. The classifications of non-prompt and fake leptons used for the analysis in this report are outlined in the next section.

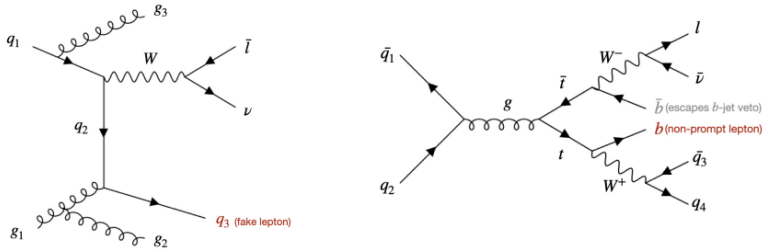


Figure 7: Feynman diagrams for the main processes that contribute to fake and non-prompt leptons for the same-sign $W^{\pm}W^{\pm}$ signal region. W+jets is shown on the left, and $t\bar{t}$ is shown on the right [3].

3.3 Truth-Type and Truth-Origin

For the analysis performed, the truth-origin and truth-type of leptons was mapped to lepton categories, known within ATLAS as "IFF categories" [8]. The IFF types are determined using the truth-origin and truth-type of the lepton, as well as the origin, type, and PDG-ID of the first and last non-Geant mother-particle.

3.3.1 IFF Classification Categories

The different lepton categories used in this analysis for muons and electrons are defined as:

- **Prompt (isolated) electrons**: An electron is in this category if its truth-type or the truth-type of its mother-particle corresponds to an isolated electron. This includes electrons originating from

final state radiation (FSR) photons or bremsstrahlung that has converted, as well as electrons from prompt quarkonium decays ($c\bar{c}/b\bar{b} \to e^+e^-$).

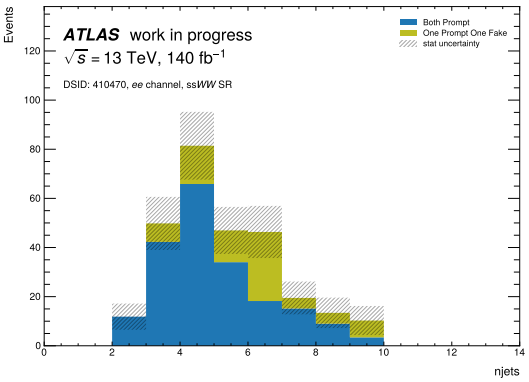
- **Prompt muons**: Isolated muons from prompt sources such as W/Z, top quark, or Higgs decays are in this category. Muons from prompt quarkonium decays also fall into this category.
- **Charge-flip electrons/muons**: This includes electrons or muons with incorrectly identified charges.
- **Prompt photon conversions**: This category includes electrons resulting from the conversion of prompt photons. It covers electrons classified as background electrons with a truth-origin of photon conversions or electromagnetic processes, including isolated and background photons from prompt photon conversions or unhadronized photons.
- **Electrons from muons**: This includes electrons with a non-isolated electron or photon truth-type and a truth-origin of a muon.
- **Tau decays**: Leptons whose truth-type corresponds to a non-isolated muon/electron/photon and originate from tau leptons, including those where the tau lepton radiates a photon that converts into an electron-positron pair, are included in this category.
- **b- and c-hadron decays**: This category contains electrons and muons from heavy-flavour decays, originating from b-hadrons or c-hadrons, with a truth-origin as a bottom-meson/baryon or a charm-meson/baryon.
- **Light-flavour decays**: Leptons from light-flavour decays are identified by a truth-type corresponding to a hadron. This category also includes background electrons, muons, or photons with a truth-origin from light/strange-mesons, or light/strange baryons, as well as those from intermediate photon conversions and Dalitz decays.
- **KnownUnknown and Unknown**: KnownUnknown leptons are those where classification fails due to truth-type and truth-origin information being labelled as unknown or undefined. Unknown leptons are those that cannot be attributed to any of the previously mentioned classes.

3.4 Background Analysis Results

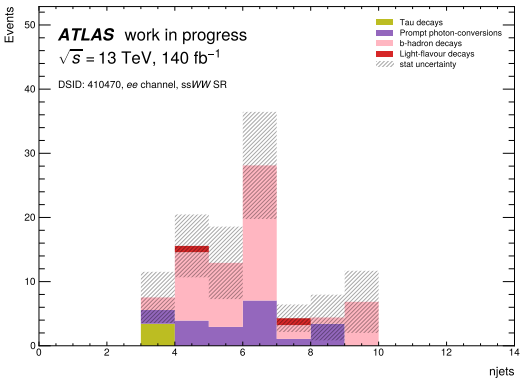
The analysis was done using the same signal region selections as shown in Table 1, with an additional b-jet veto applied and the E_T^{miss} cut set to > 25 GeV due to looser skimming. This analysis was done for the lepton channels ee, $\mu\mu$, and $e\mu$. Figure 8 and Figure 9 show results for the ee channel, while the results from the other two lepton channels can be found in Appendix B in Figure 14.

3.4.1 $t\bar{t}$ Samples

A key contribution to the background of the same-sign sign $W^\pm W^\pm$ channel is $t\bar{t}$ backgrounds. For each lepton in events that passed the signal region requirements, the IFF class of the lepton was determined based on the truth-type and truth-origin information from the lepton. The detailed IFF classification of leptons in each event is found in Appendix B in Figure 16. Figure 8 shows plots that investigate the number of jets in signal region events for the ee channel. Figure 8a shows whether the two leptons in each event were both prompt, both fake/non-prompt, or had one prompt and one fake/non-prompt. Figure 8b shows the IFF classification of the non-prompt/fake lepton in the events that are not entirely prompt leptons. The $t\bar{t}$ sample contains two W bosons, so some events contain all prompt leptons. There are no double-fake events present, which indicates that double-fake backgrounds from $t\bar{t}$ events do not need to be further investigated for this analysis. As expected, contributions from b-hadron decays are present as the main non-prompt background. Other contributions come from tau decays, light-flavour decays, as well as prompt photon conversions. The $e\mu$ and $\mu\mu$ channel results are in Appendix B, and for the $e\mu$ channel a large contribution comes from prompt photon conversions, as well as b-hadron decays.



(a) Fake and prompt classification of the two signal leptons for the ee channel for $t\bar{t}$ background.

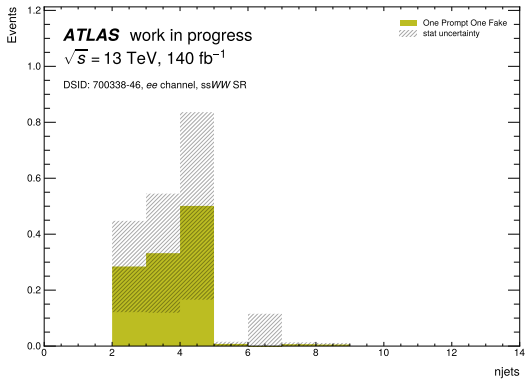


(b) IFF classification for non-prompt/fake leptons in the ee channel for $t\bar{t}$ background.

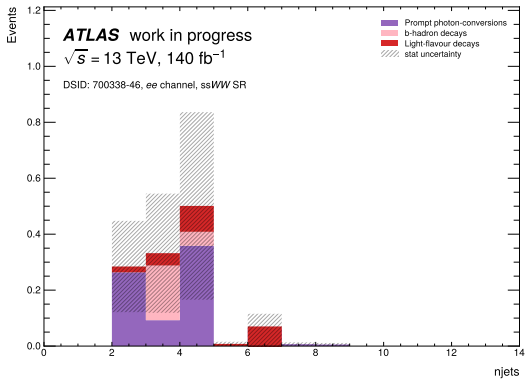
Figure 8: Non-prompt and fake lepton classifications for $t\bar{t}$ with the number of jets in each event for the ee channel.

3.4.2 W+jets Samples

W+jets events are another key background source for this analysis. Similar plots as shown previously for the $t\bar{t}$ sample were obtained. Figure 9 shows the distribution of jets in the MC simulation for the ee channel. There were no prompt-prompt events, as only one W lepton is present in this background. For the IFF classification of the non-prompt/fake lepton, b-hadron decays are still present, but light-flavour decays and prompt photon conversions are more evident in these events. Figure 9 shows the fake/prompt distribution for the number of jets in each event (Figure 9a), as well as the corresponding IFF category of the fake lepton in each event (Figure 9b). With these plots, there is a large statistical uncertainty in the plots due to the small number of events.



(a) Fake and prompt classification of the two signal leptons for the ee channel for W+jets background.



(b) IFF classifications for non-prompt/fake leptons for the ee channel for W+jets background.

Figure 9: Non-prompt and fake lepton classifications for W+jets with the number of jets in each event for the ee channel.

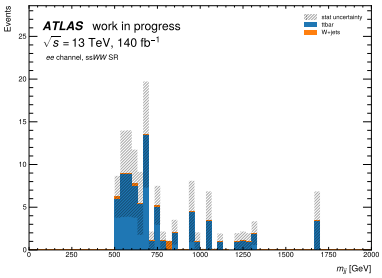
Comparing these with the distributions for the $e\mu$ and $\mu\mu$ channel (Figure 14), there also appears to be more jets per event in $e\mu$ events compared to ee events. To investigate this, the overlap between the non-prompt/fake electron with the closest jet in each event was examined. The distance between the electron and the closest jet, ΔR, as well as the difference between the angular variables $\Delta\phi$ and $\Delta\eta$ was investigated, but no correlation was seen. Further work will involve looking at the distributions of jet p_T and lepton p_T, as well as lepton and jet η distributions.

Monte Carlo is used to study the type and amount of non-prompt/fake leptons from different samples to ensure the selection of data control regions that provide estimates of the correct type of fake

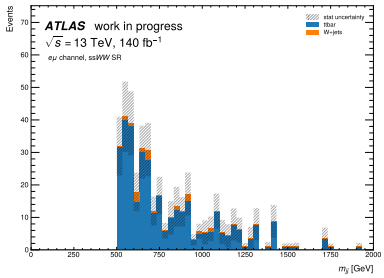
lepton. As shown in Figure 8 and Figure 9, no MC simulated events contained double-fakes. Events containing two fake leptons are very hard to estimate with data, and this result shows that double-fake events likely do not need to be focused on for the analysis.

3.4.3 Combined Background Comparisons

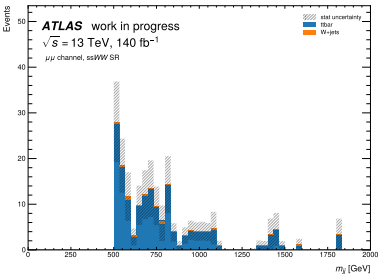
The type of background process was also directly compared for various kinematic variables. Although only $t\bar{t}$ and W+jets samples were used for these comparisons, future work will be to include other background sources in this analysis. Figure 10 shows distributions for the dijet invariant mass, m_{jj}, for each of the three lepton channels, as well as the distribution of the number of jets in each event. For both categories of plots, only events containing at least one non-prompt/fake lepton are shown. Prompt-prompt event N_{jet} and m_{jj} distributions are not shown in Figure 10. Figure 11 shows the comparison for lepton p_T for the ee channel, where only non-prompt/fake leptons are shown. Other lepton channels are shown in Appendix B in Figure 15. For all of these plots, $t\bar{t}$ contributes significantly more than the W+jets background.



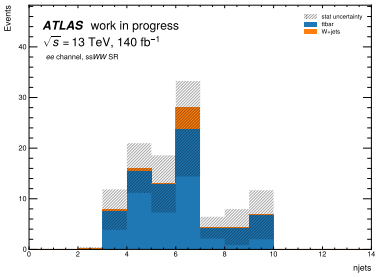
(a) ee channel m_{jj} background comparison.



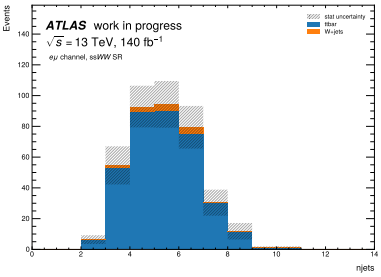
(b) $e\mu$ channel m_{jj} background comparison.



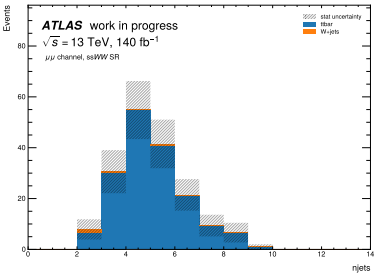
(c) $\mu\mu$ channel m_{jj} background comparison.



(d) ee channel background comparison for number of jets.

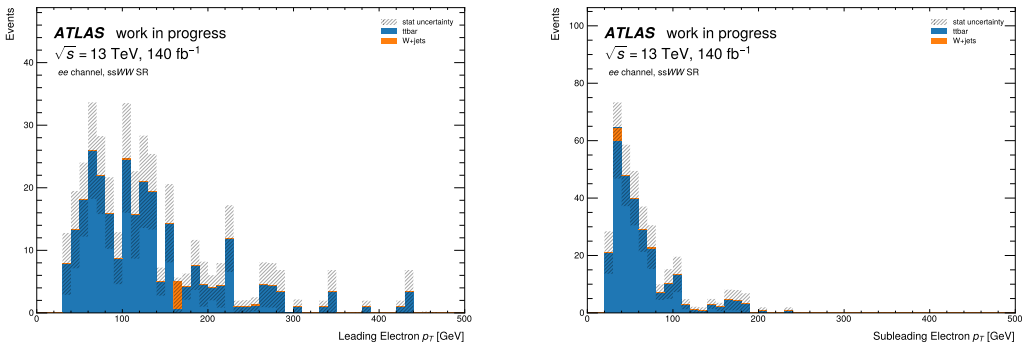


(e) $e\mu$ channel background comparison for number of jets.



(f) $\mu\mu$ channel background comparison for number of jets.

Figure 10: Background comparison plots for m_{jj} and the number of jets in each event. These plots compare the contributions from $t\bar{t}$ and W+jets background samples.



(a) ee channel background comparison for leading electron transverse momentum, p_T.

(b) ee channel background comparison for subleading electron transverse momentum, p_T.

Figure 11: Background comparison plots for lepton p_T in the ee channel. Contributions from $t\bar{t}$ and W+jets background samples are shown.

There is a larger contribution from $t\bar{t}$ samples compared to W+jets, and future work will be to incorporate additional background samples such as Z+jets, $t\bar{t}+\gamma$, single top, $W\gamma$, $Z\gamma$, and opposite-sign EWK and QCD $W^{\pm}W^{\pm}$ backgrounds. While fake and non-prompt leptons are not extremely well simulated in MC, the analysis of background distributions is useful for data-driven estimates of non-prompt background. The number of fake events in simulated events is not an accurate estimate of the signal region. Due to this, data-driven methods are used to estimate background. These methods allow for background estimates to be made using data control regions.

4 Conclusions

Accurate Monte Carlo (MC) simulations are a key aspect of searches for new physics in ATLAS. GEANT4 (FullSim) is currently used for ATLAS analysis and it is able to model physics events accurately, but it is CPU-intensive. AtlFast3 (AF3) uses tools that replace the CPU-intensive calorimeter shower simulations, which uses less CPU resources. These two simulation tools were directly compared using two vector boson fusion channels for Run 2 MC simulations. The same-sign $W^{\pm}W^{\pm}$ and $W^{\pm}Z$ channels were used in this analysis, and both signal region and control region comparisons were made. The results show good agreement between AF3 and FullSim, which validates the accuracy of AF3. All kinematic distributions are consistent between the simulations, but there is a slight discrepancy in high jet p_T regions for the $W^{\pm}Z$ channel. This showed no correlation with any other observables. Future work will be to add Run 3 data and to optimize the analysis as a $H^{\pm}$ and $H^{\pm\pm}$ search.

MC simulations are also used to study the type and amount of fake leptons in the signal region. This is useful for selecting data control regions that produce estimates of the correct type of non-prompt or fake lepton in events. In this analysis, backgrounds for the same-sign $W^{\pm}W^{\pm}$ channel were investigated. To identify non-prompt and fake lepton classifications, $t\bar{t}$ and W+jets samples were used. The number of jets in each event was investigated, as well as the classifications of the non-prompt leptons in the samples. Significant contributions from b-hadron decays were present in $t\bar{t}$ samples, and the W+jets sample revealed a larger contribution from light-flavour decays. Comparisons between background samples were also made for dijet invariant mass, the number of jets in each event, and lepton p_T for events containing at least one non-prompt or fake lepton. The largest contributions were found to be from $t\bar{t}$ backgrounds. Future work will involve combining more background samples to identify non-prompt lepton classifications, which can be useful for data-driven estimates of background contributions in the signal region.

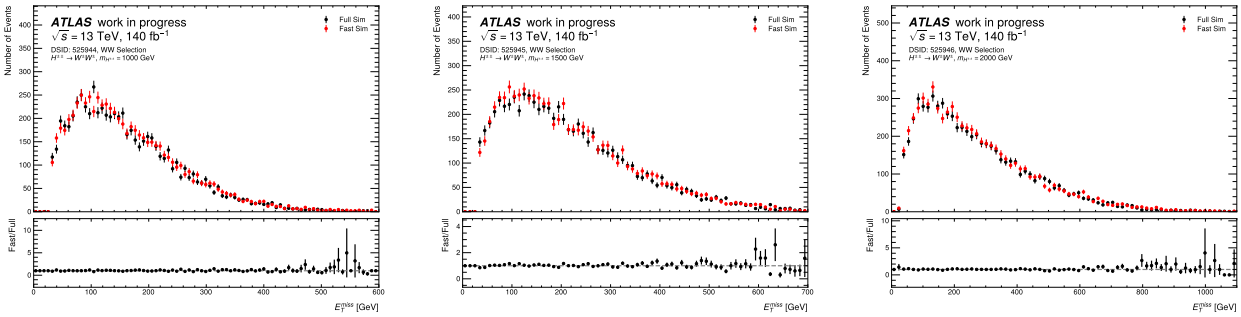
5 Acknowledgements

I would like to thank my supervisor Heather Russell for her guidance and support throughout my time working on this project. I would also like to thank John McGowan for the invaluable assistance he offered during my time at CERN. Additionally, I am sincerely grateful to the Institute of Particle Physics (IPP)

and the University of Victoria for providing funding and giving me the opportunity to participate in the CERN summer student programme.

A Appendix

The following plots are signal region sample comparisons for E_T^{miss}. The complete set of plots for all mass points and all kinematic variables that were studied can be found here.

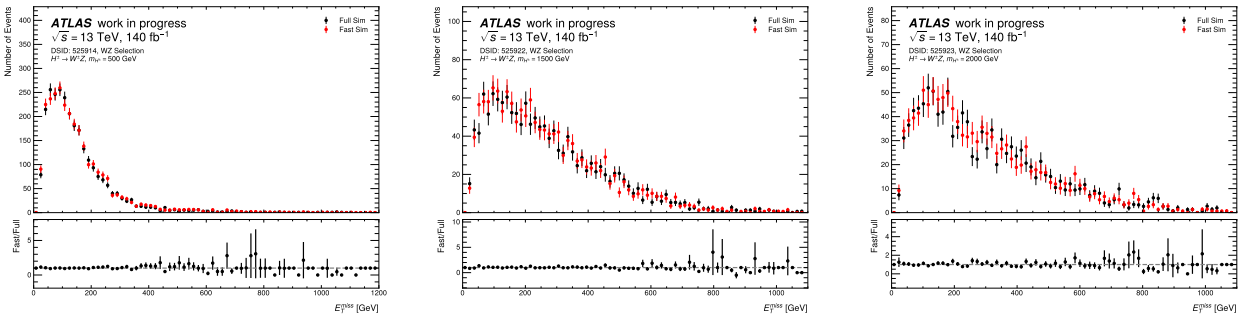


(a) $m_{H^{\pm\pm}} = 1000$ GeV signal region plot of E_T^{miss}.

(b) $m_{H^{\pm\pm}} = 1500$ GeV signal region plot of E_T^{miss}.

(c) $m_{H^{\pm\pm}} = 2000$ GeV signal region plot of E_T^{miss}.

Figure 12: Signal region missing transverse momentum, E_T^{miss}, plots for the $W^{\pm}W^{\pm}$ channel for three mass points (1000 GeV, 1500 GeV, 2000 GeV).



(a) $m_{H^{\pm}} = 500$ GeV signal region plot of E_T^{miss}.

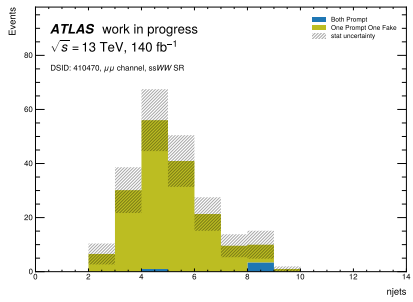
(b) $m_{H^{\pm}} = 1500$ GeV signal region plot of E_T^{miss}.

(c) $m_{H^{\pm}} = 2000$ GeV signal region plot of E_T^{miss}.

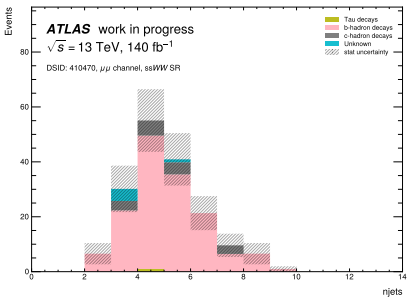
Figure 13: Signal region missing transverse momentum, E_T^{miss}, plots for the $W^{\pm}Z$ channel for three mass points (500 GeV, 1500 GeV, 2000 GeV).

B Appendix

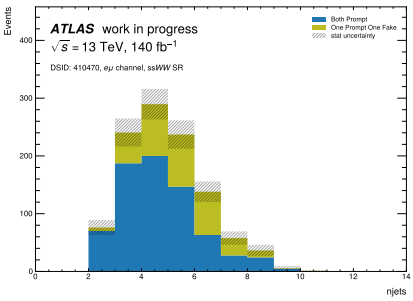
The following plots are for the non-prompt and fake lepton analysis.



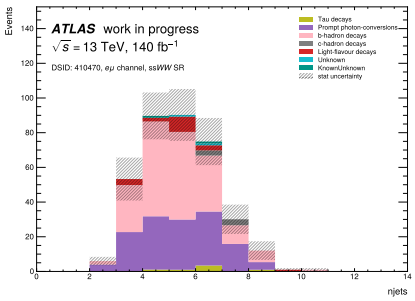
(a) $\mu\mu$ channel fake and prompt classification for $t\bar{t}$.



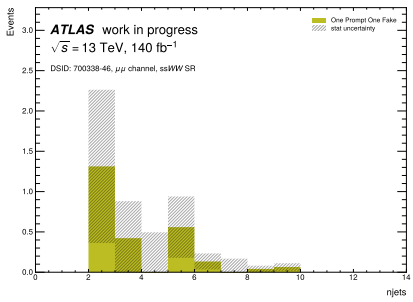
(b) $\mu\mu$ channel fake/non-prompt lepton IFF classification for $t\bar{t}$.



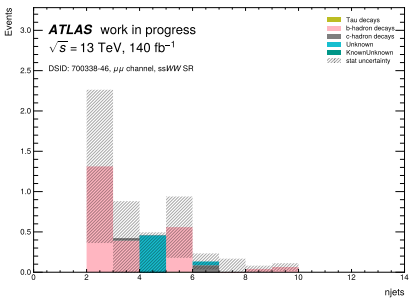
(c) $e\mu$ channel fake and prompt classification for $t\bar{t}$.



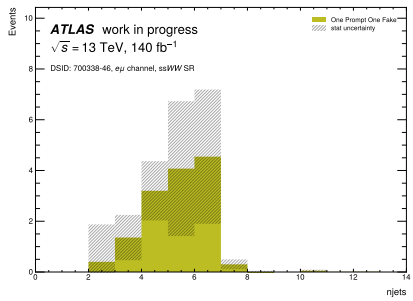
(d) $e\mu$ channel fake/non-prompt lepton IFF classification for $t\bar{t}$.



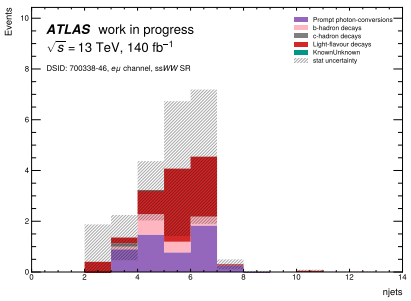
(e) $\mu\mu$ channel fake and prompt classification for W+jets.



(f) $\mu\mu$ channel fake/non-prompt lepton IFF classification for W+jets.

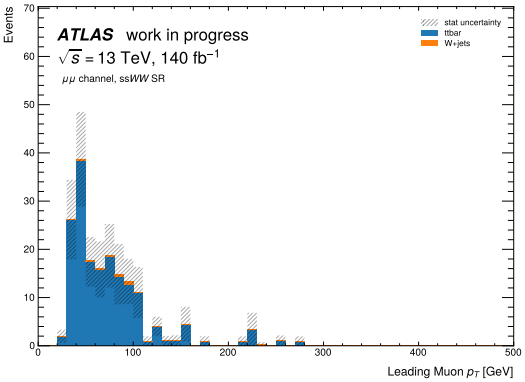


(g) $e\mu$ channel fake and prompt classification for W+jets.

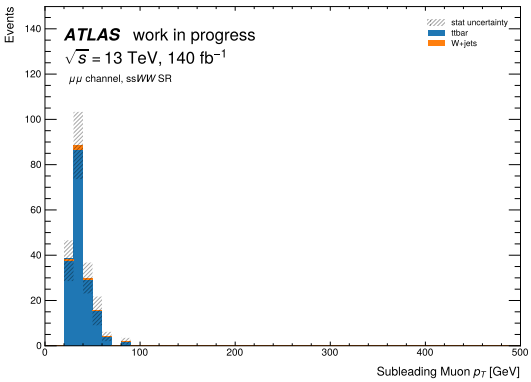


(h) $e\mu$ channel fake/non-prompt lepton IFF classification for W+jets.

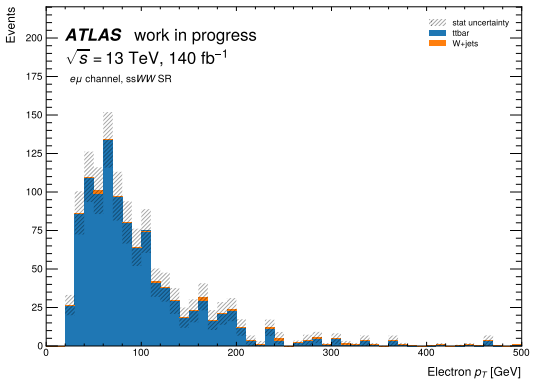
Figure 14: Non-prompt and fake lepton plots for the number of jets in $t\bar{t}$ and W+jets background samples. Figures 14a, 14c, 14e, and 14g show the prompt and fake classification. Figures 14b, 14d, 14f, and 14h show the IFF category for the non-prompt/fake lepton if it exists in each event.



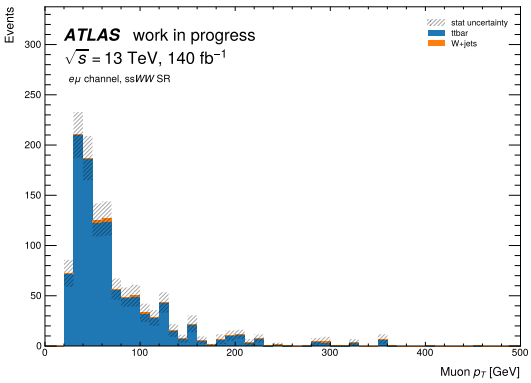
(a) $\mu\mu$ channel leading muon p_T background comparison.



(b) $\mu\mu$ channel subleading muon p_T background comparison.

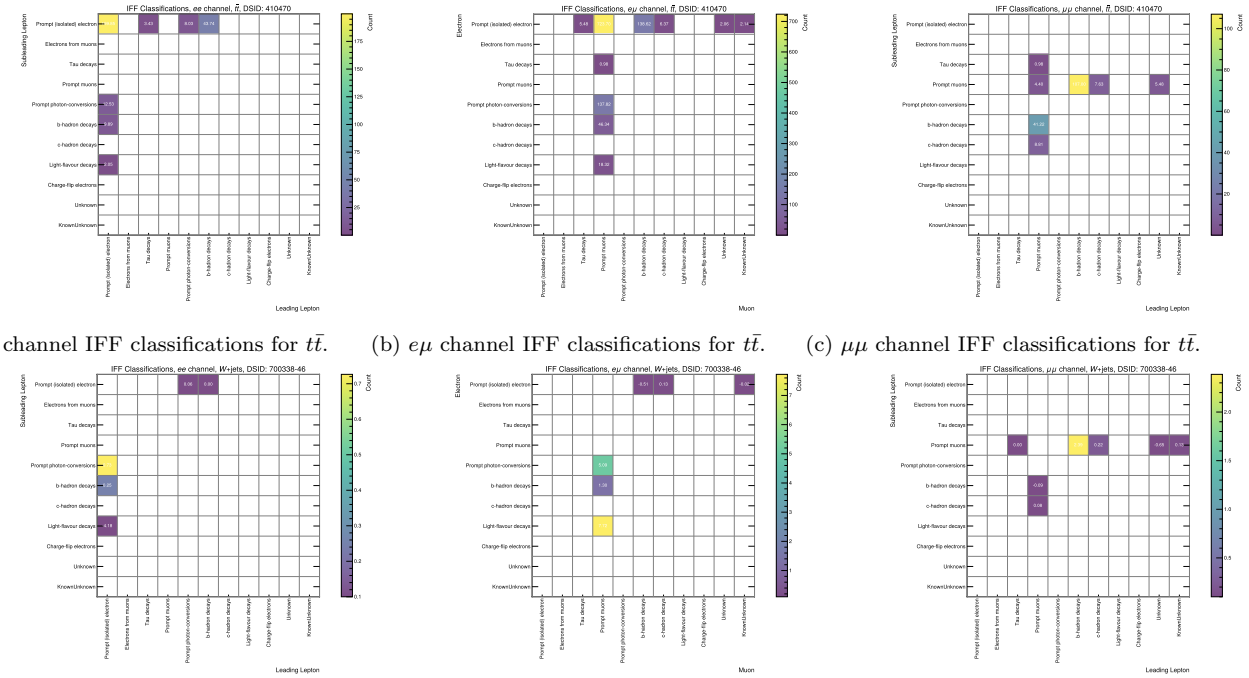


(c) $e\mu$ channel electron p_T background comparison.



(d) $e\mu$ channel muon p_T background comparison.

Figure 15: $t\bar{t}$ and W+jets comparison plots for lepton p_T for the $e\mu$ and $\mu\mu$ channels.



(a) ee channel IFF classifications for $t\bar{t}$.

(b) $e\mu$ channel IFF classifications for $t\bar{t}$.

(c) $\mu\mu$ channel IFF classifications for $t\bar{t}$.

(d) ee channel IFF classifications for W+jets.

(e) $e\mu$ channel IFF classifications for W+jets.

(f) $\mu\mu$ channel IFF classifications for W+jets.

Figure 16: IFF classifications for the ee, $e\mu$, and $\mu\mu$ lepton channels for $t\bar{t}$ and W+jets background samples.

References

[1] ATLAS Collaboration (2024). *Combination of searches for singly and doubly charged Higgs bosons produced via vector-boson fusion in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, https://arxiv.org/abs/2407.10798v1.

[2] J. Nielsen et al. (2022). *Vector Boson Fusion and Vector Boson Scattering Measurements at the Large Hadron Collider*, ATL-PHYS-PROC-2022-091.

[3] S. Solomon (2022). *The First Measurement of the Differential Cross-Section of Electroweak* $W^{\pm}W^{\pm}jj$ *Production at 13 TeV with the ATLAS Detector*, https://cds.cern.ch/record/2865127?ln=en.

[4] ATLAS Collaboration (2022). *AtlFast3: the next generation of fast simulation in ATLAS*, https://arxiv.org/abs/2109.02551v2.

[5] ATLAS Collaboration (2024). *Measurement and interpretation of same-sign W boson pair production in association with two jets in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, https://arxiv.org/abs/2312.00420v2.

[6] ATLAS Collaboration (2018). *Search for resonant WZ production in the fully leptonic final state in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, https://arxiv.org/abs/1806.01532v2.

[7] A. Ambler et al. (2023). *Observation of* $W(\to l\nu)\gamma\gamma$ *in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector at the LHC*, https://cds.cern.ch/record/2798380/files/ATL-COM-PHYS-2021-1089.pdf.

[8] ATLAS Collaboration (2019). *Athena* (Version 22.0.1), https://doi.org/10.5281/zenodo.2641997.