

© 2024 by JianCong Zeng. All rights reserved.

STUDY OF ELECTROWEAK WW/WZ PRODUCTION IN THE SEMILEPTONIC
FINAL STATES IN ASSOCIATION WITH A HIGH-MASS DIJET SYSTEM USING
THE ATLAS DETECTOR

BY

JIANCONG ZENG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2024

Urbana, Illinois

Doctoral Committee:

Associate Professor Benjamin Hooberman, Chair
Professor Mark Neubauer, Director of Research
Professor Brian D. Fields
Assistant Professor Yonatan Frederick Kahn

Abstract

The analysis searches for electroweak diboson ($WW/WZ/ZZ$) production in association with a high-mass dijet system, using data from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\text{TeV}$; the total dataset corresponds to an integrated luminosity of 139 fb^{-1} . The analysis is performed in the semileptonic final states which include a hadronic decay boson and a leptonic decay one. The analysis is split into three channels according to the number of electrically charged leptons from the leptonic boson decay: 0-lepton channel, 1-lepton channel, and 2-lepton channel. The focus of this thesis is an approach to enhance the separation between the VBS signal and the backgrounds in the 1-lepton channel using artificial neural network models. This approach is developed using Monte Carlo simulated events in the semi-leptonic VBS analysis and tested using ATLAS data for improving the sensitivity of the analysis to the target signal. The semileptonic vector boson scattering production of WW/WZ in the 1-lepton channel is measured with an observed (expected) significance of 5.78σ (5.98σ) standard deviations.

To my late grandparents, for every cherished memory.

To my parents, Meizhen Zhu and Yuesong Zeng, for your unwavering support and patience in helping me fulfill my dream. I am determined to make you proud with every step I take.

And finally, to Fiona, whom I never knew.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Science, High Energy Physics, under contract number DE-SC0023365. I am deeply grateful for their financial support which made this research possible.

I extend my sincere gratitude to my advisor, Mark Neubauer, for his invaluable guidance, encouragement, and insights throughout my study.

I would also like to thank Viviana Cavaliere and Matthew Feickert, whose support and mentorship were crucial during my research. Their expertise and advice were invaluable to the success of this work.

I am also thankful to my colleagues in the ATLAS collaboration for their assistance and partnership, which have greatly enriched my research experience.

A special thanks to my family and friends for their unwavering understanding and support during the years of my study. Their encouragement kept me motivated and focused.

Table of Contents

Chapter 1 Introduction	1
Chapter 2 Theoretical Overview	3
2.1 Standard Model	3
2.2 The Higgs Mechanism	6
2.3 Vector Boson Scattering	8
2.4 Beyond the Standard Model	13
2.4.1 Anomalous Quartic Gauge Couplings	13
Chapter 3 Experimental Apparatus	16
3.1 The LHC	16
3.2 The ATLAS Detector	19
3.2.1 The Inner Detector	20
3.2.2 The Calorimeters	21
3.2.3 The Muon Spectrometer	24
3.3 The Trigger and Data Acquisition	26
3.4 Fast Tracker	27
3.4.1 Second Stage Board	29
3.4.2 Hit Warrior FPGA	30
Chapter 4 Experimental Methods	35
4.1 Data Samples	35
4.2 Monte Carlo Simulated Samples	35
4.3 Derivation and CxAOD framework	39
Chapter 5 Object Definition	41
5.1 Electron Definition	41
5.2 Muon Definition	42
5.3 Jet Definitions	42
5.3.1 Small-R jet selection	42
5.3.2 Large-R jet selection	44
5.4 Missing Energy Definition	46
5.5 Jet Flavor (b) Tagging Definition	46
5.6 Overlap Removal	46
Chapter 6 Event Selections	48
6.1 Trigger Requirements	50
6.2 Event Pre-selection	51
6.3 Leptonically Decaying Boson Selection	52
6.4 VBS Selection	54
6.5 Signal Regions (SRs) Selection	55
6.5.1 Selection of $W/Z \rightarrow J$ candidates (merged category)	55

6.5.2	Selection of $W/Z \rightarrow jj$ candidates (resolved category)	56
6.5.3	VV system invariant mass	56
6.6	Control Regions (CRs) Definitions	58
6.6.1	$W+jets$ Control Region (WCR)	58
6.6.2	Top Control Region (TCR)	58
6.7	Signal and Background Cut-flows	61
Chapter 7	Background Modeling and Estimations	63
7.1	Reweighting for the Tag Jets	63
7.2	$W + Jets$ Control Region	68
7.3	Top Background Control Region	74
Chapter 8	Machine Learning Analysis	82
8.1	Motivations and Architecture	82
8.2	Input Variables and Feature Engineering	82
8.3	Trainings and Validations	90
Chapter 9	Systematic Uncertainties	94
9.1	Experimental Uncertainties	94
9.1.1	Baseline uncertainties	94
9.1.2	W/Z -tagging efficiency SF Uncertainty	99
9.1.3	Quark/Gluon jets uncertainty	99
9.1.4	Tracks uncertainties	100
9.2	Background Uncertainties	103
9.2.1	PDF and Scale uncertainties	107
9.2.2	ISR/FSR uncertainties	110
9.2.3	Tag Jets re-weighting	110
9.3	Signal Uncertainties	112
9.3.1	PDF and Scale uncertainties	112
9.3.2	EWK-QCD interference and Shower uncertainty	113
Chapter 10	Statistical Interpretation	114
10.1	Introduction	115
10.2	Fit inputs and options	116
10.3	Fit Model Inspection	118
10.3.1	Prefit Plots and Tables	118
10.3.2	Asimov Fit Results	121
10.4	Fit Results with Unblinded Data	125
10.5	Conclusions	131
Appendix A	List of MC Samples	132
References		148

Chapter 1

Introduction

The investigation of electroweak symmetry breaking (EWSB) has been one of the pivotal objectives of experiments conducted at the Large Hadron Collider (LHC). Within the Standard Model (SM) of particle physics, EWSB is explained by the Brout-Englert-Higgs mechanism, a cornerstone of our understanding of particle interactions [1–3]. While the Higgs boson’s discovery and subsequent measurements have significantly advanced our knowledge, a comprehensive understanding of EWSB requires further investigation. Vector boson scattering (VBS) plays a crucial role in this context, owing to its sensitivity to the interactions involving the longitudinal components of gauge bosons, which are directly tied to the mechanism of EWSB.

The ATLAS and CMS collaborations have reported results from their fully leptonic VBS analyses conducted during the LHC’s Run-II period. These findings, detailed in a series of publications by ATLAS [4–6] and by CMS [7, 8], have led to the official claim of observing the SM VBS process with fully leptonic final states.

In the study of VBS, Beyond Standard Model (BSM) physics is often explored through the effective field theory (EFT) framework [9]. Within this framework, VBS interactions can be influenced by anomalous quartic gauge couplings (aQGCs). Searches for evidence of aQGCs have been conducted across various experiments. These efforts are part of a continuing exploration to understand the potential for new physics beyond what the SM currently describes. aQGCs typically cause an increase in the VBS cross-section. This enhancement is noticeable at high transverse momentum (p_T) of the vector bosons and at high invariant mass of the two-boson system.

Experimentally, VBS is recognized by two key features: the presence of a pair of vector bosons (W , Z , or γ) and two forward jets. These jets are notably separated by a large distance in rapidity and have a high dijet invariant mass. Previous studies searching for aQGCs in VBS events have mainly focused on cases where the bosons decay into leptons ($W(l\nu)$ and $Z(l\bar{l})$)¹ and photons.

The semi-leptonic channels, such as $V(qq')Z(\nu\nu)$, $V(qq')W(l\nu)$ and $V(qq')Z(l\bar{l})$ ($V = W, Z$), present unique advantages. The branching fractions—meaning the probabilities of these decay modes—are much

¹Unless otherwise noted, ℓ stands for either electron (e) or muon (μ) in this thesis.

larger for decays involving $V(qq')$ compared to those that are purely leptonic. This efficiency and sensitivity make the semi-leptonic channels particularly appealing for aQGC searches.

This thesis focuses on analyzing the production of $V(qq')W(l\nu)$ alongside a high-mass dijet system, in a phase space optimized for sensitivity to VBS and aQGCs. This specific setup is referred to as the 1-lepton channel, part of a broader analysis that includes 0-lepton, 1-lepton, and 2-lepton channels [10]. The primary emphasis is on the 1-lepton channel, showcasing a machine learning approach designed to improve the detection sensitivity to SM VBS process.

This thesis is structured as follows: Chapter 2 offers a brief overview of the Standard Model, delves into the physics of the Higgs field, and explains why searching for VBS and aQGCs is important. Chapter 3 describes the experimental setup used to gather the data for this study, focus on the LHC and the ATLAS detector. Chapters 4 to 10 detail the entire workflow of the semi-leptonic VBS analysis and highlight the promising outcomes achieved using my machine learning approach.

Chapter 2

Theoretical Overview

The Standard Model of particle physics is our best description of elementary particles and how they interact. The SM has been tested and proven right by many experiments worldwide at various energy levels. Even though the SM has been very successful, it doesn't answer every question about the universe, leaving space for new theories beyond the SM that try to fill these gaps and explain the mysteries. This chapter gives a simple overview of the SM, discusses the importance of the Higgs field, and explains why studying VBS and aQGCs is essential.

2.1 Standard Model

The Standard Model of particle physics offers a comprehensive framework that describe the fundamental particles and the interactions governing them, excluding gravity [11]. Table 2.1 shows the relative strengths of these four fundamental forces, highlighting the marked disparity between gravity and the other forces within the realm of subatomic particles.

Force	Relative Strength
Strong Nuclear Force	1
Electromagnetic Force	$\sim 1/137$
Weak Nuclear Force	$\sim 10^{-6}$
Gravitational Force	$\sim 10^{-38}$

Table 2.1: Relative Strengths of the Fundamental Forces

The particles of the Standard Model are classified into four distinct categories. Leptons and quarks constitute all known matter, each possessing half-integer spin and adhering to Fermi-Dirac statistics. This principle prohibits any two identical fermions from occupying identical quantum states simultaneously. The model includes six leptons and six quarks, each accompanied by a corresponding antiparticle counterpart. A fundamental distinction between these particle types is that quarks are subject to the strong nuclear force, whereas leptons are not. Furthermore, quarks are perpetually bound within composite particles known as hadrons—protons and neutrons being the most recognized examples.

The remaining particles within the Standard Model facilitate the interactions among leptons and quarks and are characterized by integer spin, thereby following Bose-Einstein statistics. This group encompasses the gauge bosons, which are spin-1 particles responsible for mediating the three fundamental forces, and the Higgs boson [12], a unique scalar particle with a spin-0, crucial for imparting mass to other particles. Figure 2.1 and Table 2.2 show all Standard Model particles, alongside their properties.

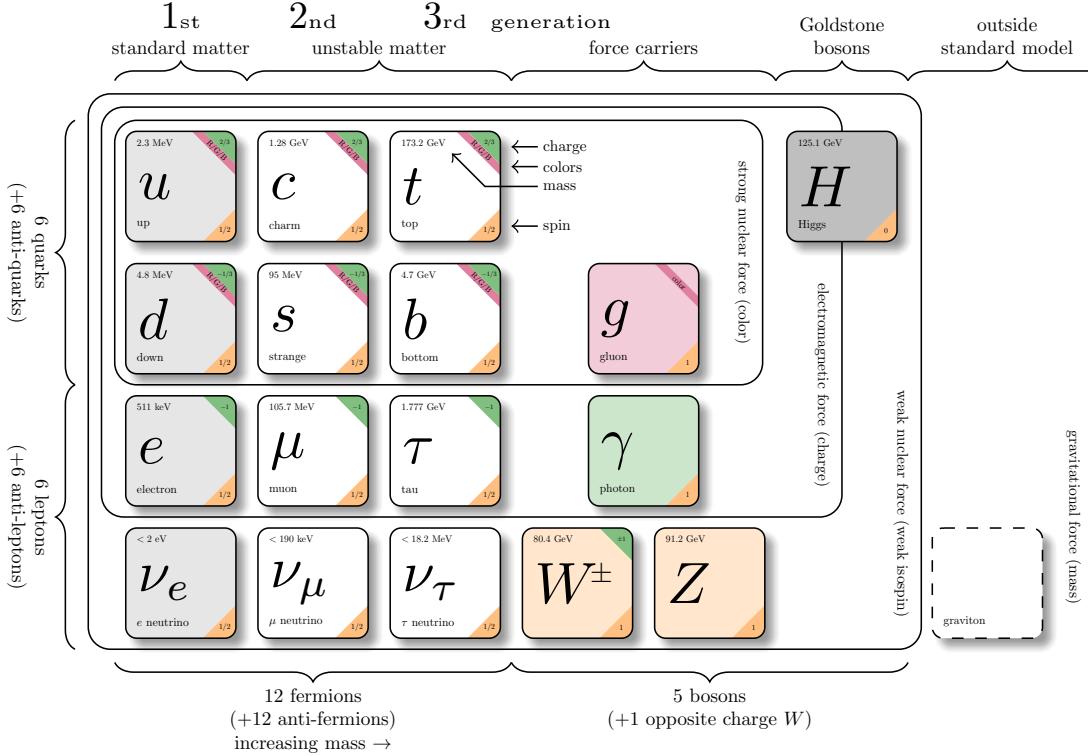


Figure 2.1: Illustration of Standard Model particles [13].

The SM is a quantum field theory anchored in the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where each component of the symmetry group governs a different aspect of particle interactions. In this context, the $SU(3)_C$ symmetry relates to the strong interaction, with the subscript C denoting color charge—a conserved property. Color charge exists in three forms—red, blue, and green—alongside their anticolors (anti-red, anti-blue, and anti-green).

One of the foundational principles of quantum chromodynamics (QCD) [14], the theory of the strong interaction, is that color charge is never observed in isolation. This phenomenon, known as color confinement, dictates that quarks are always found in composite particles called hadrons, which exhibit a neutral color charge. Hadrons are categorized into mesons and baryons. Mesons consist of a quark and an antiquark

Particle Type	Particle	Spin	Charge (e)	Mass
Lepton	Electron (e^-)	1/2	-1	0.511 MeV/c ²
	Muon (μ^-)	1/2	-1	105.7 MeV/c ²
	Tau (τ^-)	1/2	-1	1.777 GeV/c ²
	Electron Neutrino (ν_e)	1/2	0	< 2 eV/c ²
	Muon Neutrino (ν_μ)	1/2	0	< 0.19 MeV/c ²
	Tau Neutrino (ν_τ)	1/2	0	< 18.2 MeV/c ²
Quark	Up (u)	1/2	+2/3	2.3 MeV/c ²
	Down (d)	1/2	-1/3	4.8 MeV/c ²
	Charm (c)	1/2	+2/3	1.275 GeV/c ²
	Strange (s)	1/2	-1/3	95 MeV/c ²
	Top (t)	1/2	+2/3	173.1 GeV/c ²
	Bottom (b)	1/2	-1/3	4.18 GeV/c ²
Gauge Boson	Photon (γ)	1	0	0
	W Boson (W^\pm)	1	± 1	80.379 GeV/c ²
	Z Boson (Z^0)	1	0	91.1876 GeV/c ²
	Gluon (g)	1	0	0
Scalar Boson	Higgs (H)	0	0	125.10 GeV/c ²

Table 2.2: Standard Model Particles and Their Known Properties.

pair, combining a color with its anticolor. Baryons, such as the proton and neutron, are formed from three quarks, each of a distinct color (or anticolor), which collectively neutralize their color charge. This configuration explains why hadrons are observable in nature, contrary to isolated quarks. As quarks attempt to separate, they undergo a process known as *hadronization*. During hadronization, new quark-antiquark pairs spontaneously emerge from the vacuum. These new particles arrange themselves to neutralize the color charge of the initially separating quarks. As a result, instead of isolated quarks, we observe the formation of new hadrons, ensuring that the color charge remains confined. Hadronization and its resulting particle decays produce a shower(cascade) of hadrons and leptons known as a “jet.” These jets, which will be discussed on in Section 5.3.1, are crucial for probing QCD physics.

The electroweak interaction, described by the symmetry group $SU(2)_L \times U(1)_Y$, unifies the electromagnetic and weak forces [15–17]. The conservation of electric charge (Q) is central to this theory, governed by the formula: $Q = \frac{Y}{2} + T_3$. Here, Y denotes the hypercharge, and T_3 is the third component of the weak isospin (T), akin to the third component of spin in quantum mechanics, differentiating “up” and “down” states within isospin doublets. This framework introduces a distinction in symmetry for particles based on their chirality: left-handed particles exhibit doublet configurations under $SU(2)_L$, while right-handed particles form singlets, indicating they do not participate in weak isospin doublets.

Spin-1 particles, known as gauge bosons or vector bosons, are the force carriers in the SM. The photon mediates the electromagnetic force. The weak force is mediated by three vector bosons, the W^\pm and Z bosons. For the strong force, it is mediated by eight different gluons. A key feature of the weak force is

lepton universality, meaning all types of leptons couple to the W^\pm and Z bosons in the same way, leading to identical branching ratios for each lepton type.

2.2 The Higgs Mechanism

In the SM, the requirement of gauge invariance under the group $SU(2)_L \otimes U(1)_Y$ implies that weak bosons and fermions should be massless. However, this stands in contradiction to the observed masses of these particles. The resolution within the SM framework is the Higgs mechanism [12, 18–20], which introduces the Higgs field. Through spontaneous symmetry breaking, this scalar field gives mass to the weak bosons and fermions, reconciling the SM with experimental observations.

Spontaneous symmetry breaking refers to a phenomenon where the ground state of a symmetric system does not reflect the system's symmetry after perturbation. The term “breaking” does not imply the destruction of the Lagrangian symmetry. Instead, the symmetry is not apparent in the perturbed ground state of the system. A classical example of spontaneous symmetry breaking is the magnetization of iron. Above its Curie temperature, iron’s magnetic moments are oriented randomly, respecting rotational symmetry. Upon cooling below this temperature, the moments align uniformly, creating a magnet with distinct poles. Although the magnet displays a specific orientation, the physical principles governing magnetism retain their inherent symmetry.

The scalar field (Higgs field), ϕ , represented as a complex doublet, takes on a non-zero vacuum expectation value (VEV) of $v \approx 246$ GeV, leading to three massive gauge bosons (W^\pm and Z) and the Higgs boson (H).

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (2.1)$$

The Lagrangian density of the Higgs field, $\mathcal{L}_{\text{Higgs}}$, respects local $SU(2)_L \otimes U(1)_Y$ symmetry and includes the Higgs potential $V(\phi)$, as shown in Figure 2.2. This potential is defined with parameters $\mu^2 < 0$ and $\lambda > 0$ to ensure stable minima.

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi), \quad (2.2)$$

where $V(\phi)$ is the Higgs potential, given by

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (2.3)$$

Choosing a ground state breaks the $SU(2)_L \otimes U(1)_Y$ symmetry to the subgroup $U(1)_{\text{QED}}$, reparameterizing

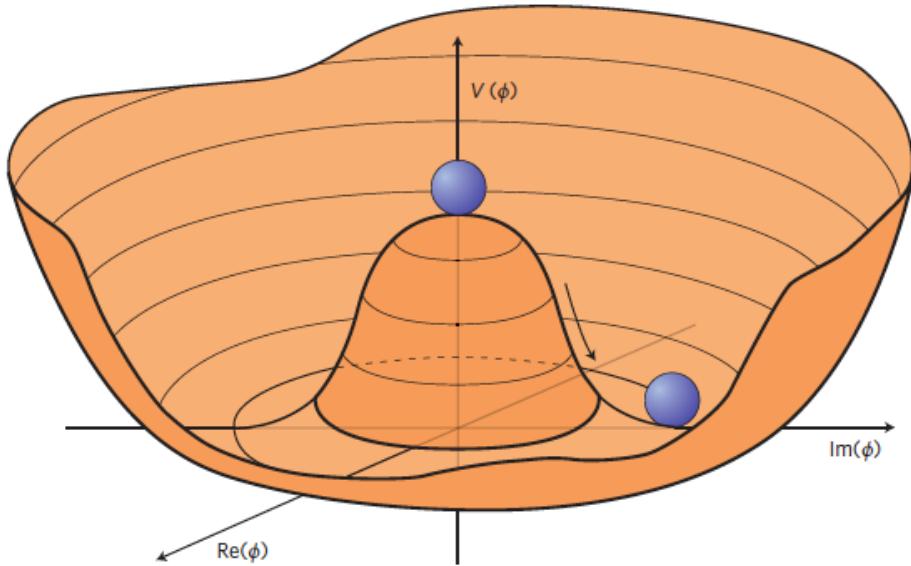


Figure 2.2: The Higgs potential, adapted from Ellis (2013) [21].

the complex scalar doublet's components into observable fields. After this symmetry breaking, the scalar doublet fields are reparameterized as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \exp \left(\frac{i\tau^i \phi^i(x)}{2} \right), \quad (2.4)$$

where $h(x)$ is the Higgs boson field, and the $\phi^i(x)$ are the Goldstone bosons that become gauge degrees of freedom in the unitary gauge $\phi^i(x) = 0$. This simplifies the kinetic term of the Lagrangian, resulting in mass terms for the W^\pm and Z bosons as:

$$m_W = \frac{1}{2} v g, \quad m_Z = \frac{v g}{2 \cos \theta_W}. \quad (2.5)$$

The Higgs boson mass is determined as:

$$m_h = \sqrt{-2\mu^2} = v\sqrt{2\lambda}. \quad (2.6)$$

In the SM, the VEV of the Higgs field, denoted by v , along with the gauge coupling constants g_1 and g_2 , are free parameters. These are not predicted by the theory but are determined through experimental measurements. The VEV has been established to be about 246 GeV [22]. This VEV is crucial as it gives mass to the W^\pm and Z bosons through the mechanism previously discussed.

Quarks and charged leptons also gain their mass by interacting with the Higgs field. These interactions are described by Yukawa couplings [23]. Each type of fermion has its own Yukawa coupling constant c_i , which determines the strength of its interaction with the Higgs field, and therefore, its mass.

$$\mathcal{L}_{\text{Yukawa}} = -\frac{1}{\sqrt{2}}(v+h)(c_1\bar{d}d + c_2\bar{u}u + c_3\bar{e}e) \quad (2.7)$$

$$= -\left(1 + \frac{h}{v}\right)(m_d\bar{d}d + m_u\bar{u}u + m_e\bar{e}e). \quad (2.8)$$

Neutrinos, however, do not acquire mass through the same mechanism. They remain massless in the SM's framework of electroweak symmetry breaking. The fact that neutrinos have been observed to have mass in experiments is an unsolved issue within the SM, hinting at new physics beyond the current theory.

2.3 Vector Boson Scattering

When the scalar field ϕ is expanded around its ground state value, the kinetic term in $\mathcal{L}_{\text{Higgs}}$ becomes

$$|D_\mu\phi|^2 = \frac{1}{2}(\partial_\mu\phi_0)^2 + \frac{1}{2}(\partial_\mu\phi_2)^2 + g^2\phi_0^2A_\mu A^\mu + \sqrt{2}g\phi_0A_\mu\partial^\mu\phi_2 + \dots, \quad (2.9)$$

where the mass term for A_μ is $g^2\phi_0^2A_\mu A^\mu$. The term $\sqrt{2}g\phi_0A_\mu\partial^\mu\phi_2$ provides an extra degree of freedom for A_μ . The directly coupling A_μ to ϕ_2 allows ϕ_2 to serve as an additional degree of freedom for the gauge boson A_μ , enabling it to acquire the necessary longitudinal polarization to be a massive particle.

A vector boson is a boson with spin-1. By definition, the gauge bosons mentioned above are vector bosons, which introduces Vector Boson Scattering (VBS) into the discussion. The SM VBS process involves a quartic gauge coupling (QGC) vertex, as shown in Figure 2.3, which depicts the self-coupling among four electroweak gauge bosons. The discovery of a Higgs boson in LHC [24, 25] motivates further study of the mechanism of EWSB by probing the VBS processes. Given that new physics in the electroweak sector is likely to involve QGCs, the measurements of VBS at high energy are crucial tests of the SM and will determine whether the Higgs is entirely responsible for EWSB.

In Figure 2.4, we present examples of VBS diagrams that contribute to the SM signal process in the analysis. In the Feynman diagrams presented, a dashed line represents the Higgs boson, a solid line represents quark, and a rippled line represents vector boson. It's noteworthy that not all VBS diagrams involve quartic gauge couplings. The decays of the bosons are not explicitly shown.

In Figure 2.5, we present several non-VBS electroweak diagrams. These include purely-electroweak tree-

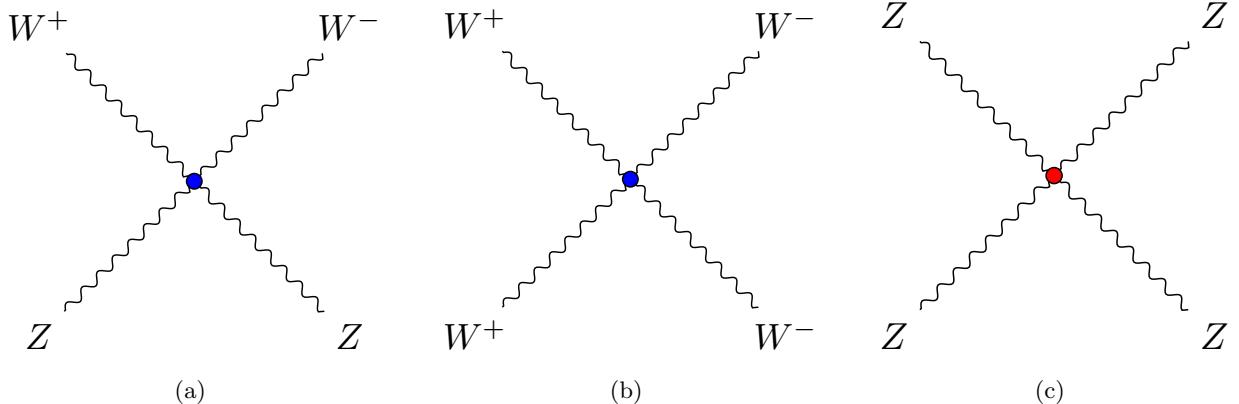


Figure 2.3: Examples of QGC vertices: (a) and (b) are within the SM; (c) is anomalous, discussed further in Section 2.4.1. These and other Feynman diagrams in this thesis are made using the JaxoDraw [26] program.

level diagrams, specifically of order $\mathcal{O}(\alpha_{EW}^6)$, contributing to the final state. Such non-VBS diagrams are expected to be significantly suppressed by the event selection criteria employed in this analysis.

Depending on the decay products of the vector bosons, VBS processes can be categorized into three types: fully leptonic, fully hadronic, and semileptonic. The fully leptonic VBS analysis studies final states where both vector bosons decay into leptons. Similarly, the fully hadronic analysis targets cases where the decay products are exclusively hadronic. While studies of fully leptonic final states have been done, the branching ratio from double leptonic decay modes is relatively small. On the other hand, fully hadronic final states have a significantly higher branching ratio but introduce considerable background noise, making them challenging to analyze. The semileptonic VBS analysis is looking at final states with one hadronic and one leptonic boson, offering a balanced compromise between a higher branching ratio and manageable background levels. As illustrated in Figure 2.6, the two “forward” jets, j_f and j_b , are high p_T jets that scatter through the coupling of the gauge bosons. These jets, pointing roughly along the beamline direction, serve as signatures of the VBS process. Further details on physics object definitions and VBS selection criteria will be discussed in Chapter 5 and Section 6.4.

Figure 2.6 also shows two jets, j_c , originating from the hadronically decaying vector boson. The boson decaying leptonically can have a few combinations of decay products. By counting the number of observable lepton(s), we have three channels in the semileptonic VBS analysis. Although neutrinos are technically leptons, within the context of this and similar analyses, “lepton” typically refers to those detectable directly by the ATLAS and CMS detectors. Therefore, the 0-lepton channel corresponds to final states where the vector boson decays into a pair of neutrinos. The 1-lepton channel involves one observable lepton (e.g., e , μ) and a neutrino, while the 2-lepton channel includes final states without neutrinos. Most of the work presented in this thesis applies to all three channels but is primarily discussed from the 1-lepton channel

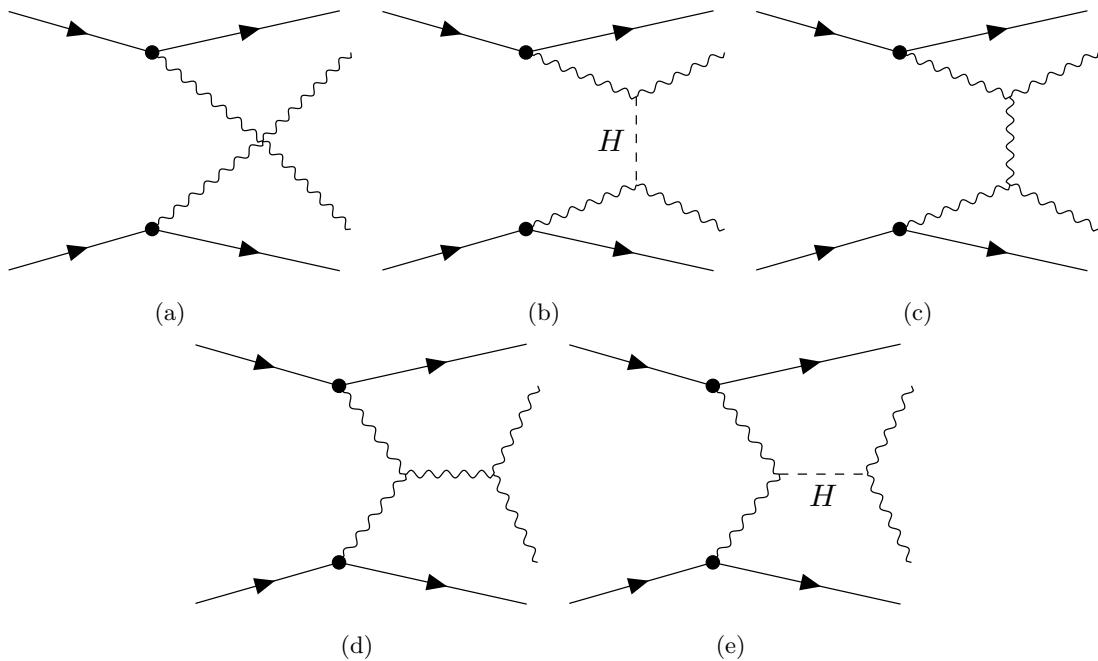


Figure 2.4: Examples of VBS diagrams contributing to the signal. Note that not all VBS diagrams contain quartic gauge couplings.

perspective.

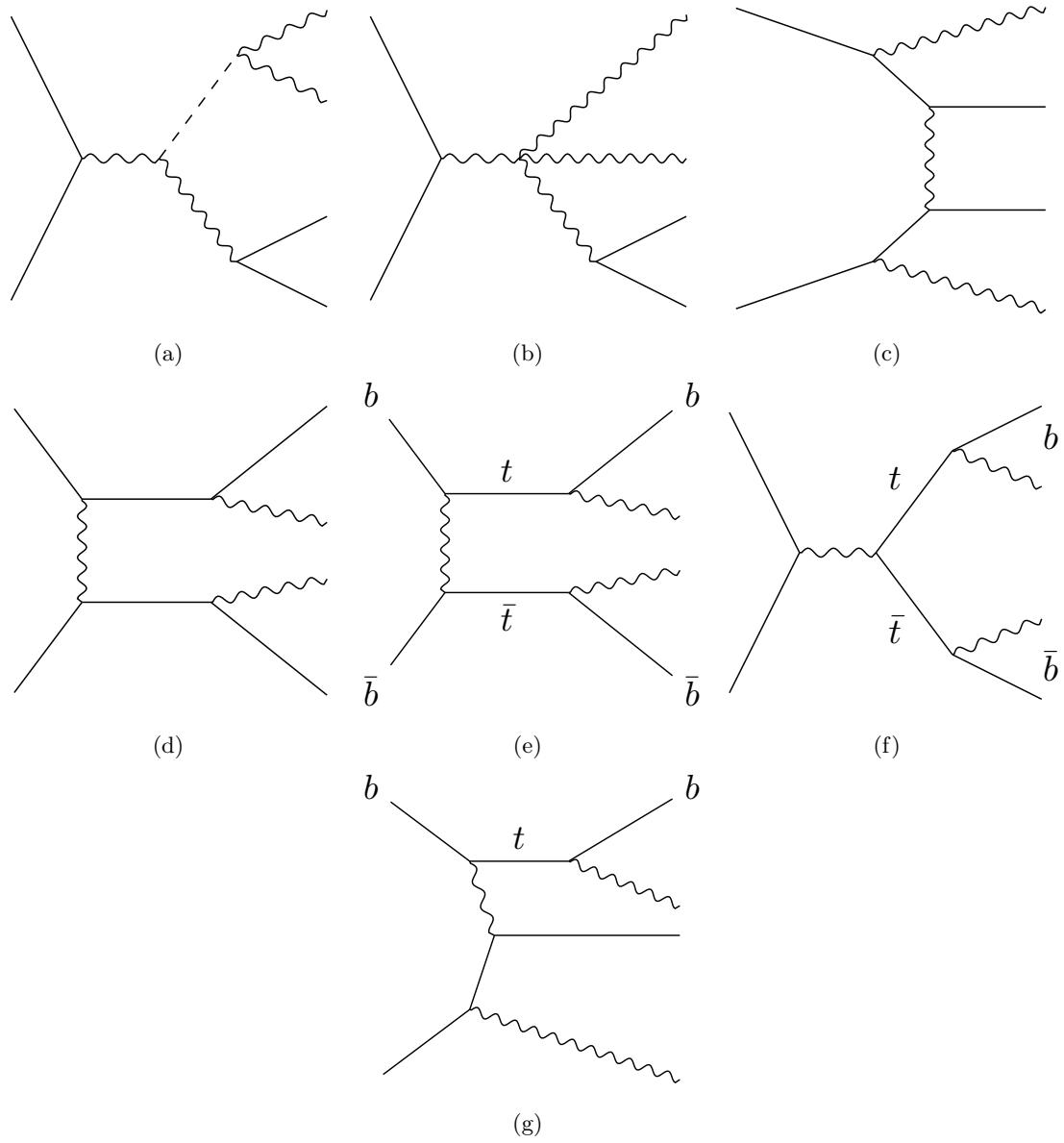


Figure 2.5: Examples of non-VBS $\mathcal{O}(\alpha_{EW}^6)$ diagrams contributin to the signal.

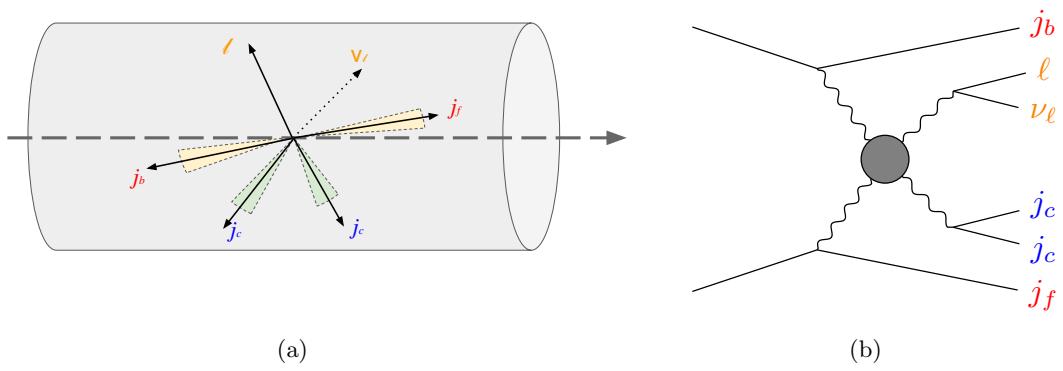


Figure 2.6: A schematic representation of a semileptonic VBS process within the detector, accompanied by the corresponding Feynman diagram.

2.4 Beyond the Standard Model

Despite intensive searches, experiments at the LHC have not yet observed new physics phenomena beyond those predicted by the Standard Model of particle physics. After the discovery of the Higgs boson, a significant number of analyses have been dedicated to searching for traces of new particles. In the meantime, we also explore the possibility of Beyond the Standard Model (BSM) interactions involving known Standard Model particles, such as the Higgs, W , and Z bosons. Any potential new physics related to EWSB could thus modify the interactions of these particles. In our case where no new physics has yet been found, the Standard Model can be considered as a low-energy approximation of a more comprehensive theory. This perspective allows us to describe potential BSM phenomena through an Effective Field Theory (EFT) approach [27] [28].

2.4.1 Anomalous Quartic Gauge Couplings

The concept of anomalous couplings among electroweak vector bosons existed before the Higgs boson was discovered [29]. Even then, it was widely believed that EWSB governed the electroweak interactions, ensuring they remained consistent without violating any unitarity constraints. Given the discovery of the Higgs boson and the solid foundation of the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge symmetry, employing an EFT approach has emerged as a compelling method for predicting and analyzing precise measurements of electroweak processes, including any deviations from expected results.

The EFT framework offers a streamlined and versatile way to explore anomalous couplings without being tied to any specific model [30]. This includes both anomalous triple gauge couplings (aTGCs) and anomalous quartic gauge couplings (aQGCs). Although aTGCs remain a significant area of interest, our focus here will be aQGCs. aQGCs are particularly intriguing because they can be investigated through VBS, providing a clear pathway for probing these potential deviations.

Assuming that new physics emerges only at energies above the scale Λ , we can describe any phenomena beyond the Standard Model (including anomalous gauge couplings) using an effective Lagrangian:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_j \frac{c_j^{(8)}}{\Lambda^4} \mathcal{O}_j^{(8)} + \dots \quad (2.10)$$

where \mathcal{L}_{SM} is the Standard Model Lagrangian, \mathcal{O}_i are the higher-dimensional operators of dimension d_i , c_i are the Wilson coefficients. In the effective Lagrangian, we include only even-dimensional operators because odd-dimensional operators would violate lepton and/or baryon number conservation [30]. The dimension-6 (D-6) terms, $\frac{c_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(6)}$, are often associated with aTGCs. The dimension-8 (D-8) terms, $\frac{c_j^{(8)}}{\Lambda^4} \mathcal{O}_j^{(8)}$, are typically associated with aQGCs, introducing or modifying interactions involving four gauge bosons. Although both

D-6 and D-8 operators can contribute to the VBS processes, the D-6 operators have been tightly constrained to values around zero by previous diboson measurements. Therefore, we will focus our discussion on the D-8 operators here.

Building on the EFT approach to model the effects of possible aQGCs, we adopt the Eboli model, which introduces 21 new D-8 operators adhering to the SM $SU(2) \times U(1)_Y$ gauge symmetry [31]. These operators can be classified into three categories: scalar, tensor, and mixed types. A list of these operators is provided at the end of this section. As shown in Table 2.3, the semileptonic VBS stands out as an exceptional process capable of testing nineteen D-8 operators concurrently, with final states that involved WW , WZ , and ZZ boson pairs.

	WWWW	WWZZ	ZZZZ	WW γ Z	WW $\gamma\gamma$	ZZZ γ	ZZ $\gamma\gamma$	Z $\gamma\gamma\gamma$	$\gamma\gamma\gamma\gamma$
$\mathcal{L}_{S0,2}, \mathcal{L}_{S1}$	X	X	X	—	—	—	—	—	—
$\mathcal{L}_{M0}, \mathcal{L}_{M1}, \mathcal{L}_{M6}, \mathcal{L}_{M7}$	X	X	X	X	X	X	X	—	—
$\mathcal{L}_{M2}, \mathcal{L}_{M3}, \mathcal{L}_{M4}, \mathcal{L}_{M5}$	—	X	X	X	X	X	X	—	—
$\mathcal{L}_{T0}, \mathcal{L}_{T1}, \mathcal{L}_{T2}$	X	X	X	X	X	X	X	X	X
$\mathcal{L}_{T5}, \mathcal{L}_{T6}, \mathcal{L}_{T7}$	—	X	X	X	X	X	X	X	X
$\mathcal{L}_{T8}, \mathcal{L}_{T9}$	—	—	X	—	—	—	X	X	X

Table 2.3: Correspondences between the vertices and operators. The “X” marks indicate the quartic gauge vertices that can be modified by the specified D-8 operators.

The following are the three classes of D-8 operators in the Eboli model [31].

The scalar operators containing just covariant derivatives of the Higgs field, $D_\mu \Phi$:

$$\mathcal{L}_{S0,2} = \left[(D_\mu \Phi)^\dagger D_\nu \Phi \right] \times \left[(D^\mu \Phi)^\dagger D^\nu \Phi \right] \quad (2.11)$$

$$\mathcal{L}_{S1} = \left[(D_\mu \Phi)^\dagger D^\mu \Phi \right] \times \left[(D_\nu \Phi)^\dagger D^\nu \Phi \right] \quad (2.12)$$

Here, the operators \mathcal{L}_{S0} and \mathcal{L}_{S2} are Hermitian conjugates, and can be treated as the same operator in practice.

The tensor operators containing just the field strength tensor:

$$\mathcal{L}_{T,0} = \text{Tr} [\hat{W}_{\mu\nu} \hat{W}^{\mu\nu}] \times \text{Tr} [\hat{W}_{\alpha\beta} \hat{W}^{\alpha\beta}] \quad (2.13)$$

$$\mathcal{L}_{T,1} = \text{Tr} [\hat{W}_{\alpha\nu} \hat{W}^{\mu\beta}] \times \text{Tr} [\hat{W}_{\mu\beta} \hat{W}^{\alpha\nu}] \quad (2.14)$$

$$\mathcal{L}_{T,2} = \text{Tr} [\hat{W}_{\alpha\mu} \hat{W}^{\mu\beta}] \times \text{Tr} [\hat{W}_{\beta\nu} \hat{W}^{\nu\alpha}] \quad (2.15)$$

$$\mathcal{L}_{T,3} = \text{Tr} [\hat{W}_{\alpha\mu} \hat{W}^{\mu\beta} \hat{W}^{\nu\alpha}] \times B_{\beta\nu} \quad (2.16)$$

$$\mathcal{L}_{T,4} = \text{Tr} [\hat{W}_{\alpha\mu} \hat{W}^{\alpha\mu} \hat{W}^{\beta\nu}] \times B_{\beta\nu} \quad (2.17)$$

$$\mathcal{L}_{T,5} = \text{Tr} [\hat{W}_{\mu\nu} \hat{W}^{\mu\nu}] \times B_{\alpha\beta} B^{\alpha\beta} \quad (2.18)$$

$$\mathcal{L}_{T,6} = \text{Tr} [\hat{W}_{\alpha\nu} \hat{W}^{\mu\beta}] \times B_{\mu\beta} B^{\alpha\nu} \quad (2.19)$$

$$\mathcal{L}_{T,7} = \text{Tr} [\hat{W}_{\alpha\mu} \hat{W}^{\mu\beta}] \times B_{\beta\nu} B^{\nu\alpha} \quad (2.20)$$

$$\mathcal{L}_{T,8} = B_{\mu\nu} B^{\mu\nu} B_{\alpha\beta} B^{\alpha\beta} \quad (2.21)$$

$$\mathcal{L}_{T,9} = B_{\alpha\mu} B^{\mu\beta} B_{\beta\nu} B^{\nu\alpha} \quad (2.22)$$

The mixed operators containing $D_\mu \Phi$ and field strength:

$$\mathcal{L}_{M,0} = \text{Tr} [\hat{W}_{\mu\nu} \hat{W}^{\mu\nu}] \times [(D_\beta \Phi)^\dagger D^\beta \Phi] \quad (2.23)$$

$$\mathcal{L}_{M,1} = \text{Tr} [\hat{W}_{\mu\nu} \hat{W}^{\nu\beta}] \times [(D_\beta \Phi)^\dagger D^\mu \Phi] \quad (2.24)$$

$$\mathcal{L}_{M,2} = [B_{\mu\nu} B^{\mu\nu}] \times [(D_\beta \Phi)^\dagger D^\beta \Phi] \quad (2.25)$$

$$\mathcal{L}_{M,3} = [B_{\mu\nu} B^{\nu\beta}] \times [(D_\beta \Phi)^\dagger D^\mu \Phi] \quad (2.26)$$

$$\mathcal{L}_{M,4} = [(D_\mu \Phi)^\dagger \hat{W}_{\beta\nu} D^\mu \Phi] \times B^{\beta\nu} \quad (2.27)$$

$$\mathcal{L}_{M,5} = [(D_\mu \Phi)^\dagger \hat{W}_{\beta\nu} D^\nu \Phi] \times B^{\beta\mu} \quad (2.28)$$

$$\mathcal{L}_{M,6} = [(D_\mu \Phi)^\dagger \hat{W}_{\beta\nu} \hat{W}^{\beta\nu} D^\mu \Phi] \quad (2.29)$$

$$\mathcal{L}_{M,7} = [(D_\mu \Phi)^\dagger \hat{W}_{\beta\nu} \hat{W}^{\beta\mu} D^\nu \Phi] \quad (2.30)$$

Chapter 3

Experimental Apparatus

3.1 The LHC

The Large Hadron Collider (LHC) [32] at CERN is currently the world’s largest and most powerful hadron accelerator and collider. It is located in a circular tunnel roughly 100 m beneath the borders of Switzerland and France. The tunnel is about 27 kilometers long and was initially built for the Large Electron-Positron Collider (LEP), which operated from 1989 to 2000. Construction of the LHC began in 1998 and was completed in 2008. There are eight Interaction Points (IP) where particle beams can collide, four located in caverns, each housing one of the four experiments of the LHC: ATLAS [33], CMS [34], LHCb [35], and ALICE [36]. The ATLAS and CMS experiments search for the same types of collision events and are used to cross-check one another’s results.

The LHC was designed to collide proton beams at a center-of-mass energy of 14 TeV and a maximum instantaneous luminosity of $10 \text{ nb}^{-1} \text{ s}^{-1}$ ($10^{34} \text{ cm}^{-2} \text{s}^{-1}$) [37]. The proton beams are guided by a sophisticated system of superconducting magnets. This system includes 1,232 dipole magnets with an 8.3 T strength for bending the beams and 392 main quadrupole magnets with a 7.5 T strength for focusing the beams [38] [39]. To sustain their superconducting state and counteract external sources of heat, the magnets are immersed in a liquid helium bath at 1.9 K within a vacuum-sealed inner vessel, as seen in Figure 3.1.

Before entering the LHC, the proton beams went through a sequence of accelerators, as shown in Figure 3.2, which gradually increased the beams’ energy. A single bottle of Hydrogen gas serves as the sole proton source for the whole LHC. At the LINAC2 linear accelerator site, the hydrogen gas is introduced into the duoplasmatron, which ionizes the gas to create a plasma of electrons and hydrogen ions [41]. The plasma is then manipulated by strong electrical and magnetic fields, which separate and accelerate the protons out of the duoplasmatron and into the accelerator. The protons enter the LINAC2 at about 90 keV and are accelerated to 50 MeV before arriving in the Booster. The Booster is a (relatively) small synchrotron, followed by the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). The proton beam is accelerated gradually at each stage and split before entering the LHC as two countercirculating beams. The details of

LHC DIPOLE : STANDARD CROSS-SECTION

CERN AC/DI/MM - HE107 - 30 04 1999

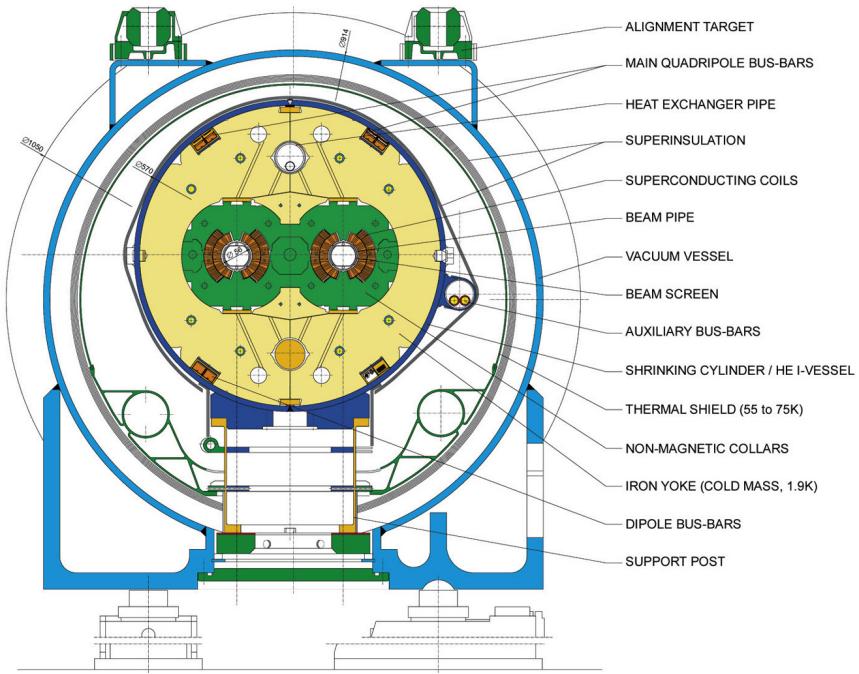


Figure 3.1: The cross-section of an LHC dipole magnet with vacuum chamber [40].

this process are summarized in Table 3.1.

Within the synchrotrons, the proton beam is repeatedly passed through the radiofrequency (RF) cavities, metallic chambers containing an electromagnetic field. This process accelerates the protons and organizes them into bunches with 25 ns spacing. Once the proton bunches circulate in the LHC, they are further accelerated while maintaining a 25 ns spacing. The beams can circulate stably in the LHC for many hours and only need to be refilled if the beam is dumped [37].

Accelerator Name	Type	Dimensions/Circumference	Exit/Final Energy
LINAC2	Linear Accelerator	-	50 MeV
Booster	Synchrotron	157 m	1.4 GeV
Proton Synchrotron (PS)	Synchrotron	628 m	26 GeV
Super Proton Synchrotron (SPS)	Synchrotron	7 km	450 GeV
Large Hadron Collider (LHC)	Synchrotron	27 km	6.5 TeV (per beam)

Table 3.1: Summary of CERN's Accelerator Chain Leading to the LHC

CERN's Accelerator Complex

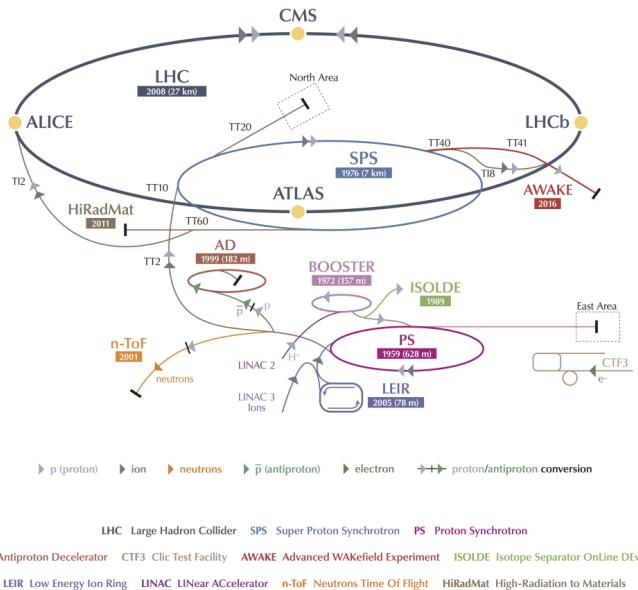


Figure 3.2: The CERN accelerator complex [42]. The LHC proton injector chain is denoted by the light grey arrows.

3.2 The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) detector [33] is a general purpose detector designed to search for all types of interesting physics events at a high luminosity. The ATLAS experiment is one of the four main LHC experiments. The ATLAS detector is located in the experiment cavern at Point 1 of the LHC at the same depth as the tunnel, roughly 100 m underground.

The detector is a large cylindrical shape, 44 m long and 25 m in diameter, and weighs approximately 7000 tonnes. Figure 3.3 shows the sub-detector systems of ATLAS and the relative human scale. The beam pipe is encased by the Inner Detector (ID), which in turn is surrounded by the calorimeters, and these are all nested inside the muon spectrometer (see Figure 3.4). The LHC coordinate system is illustrated as a rectangular coordinate system in Figure 3.5

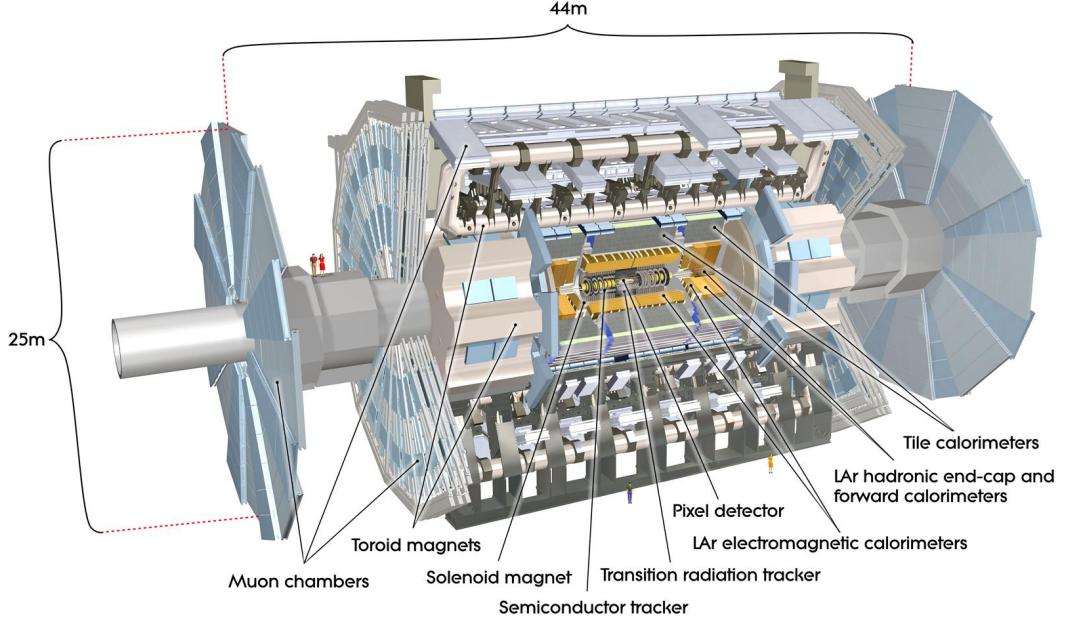


Figure 3.3: ATLAS detector with human models included for scale [43].

In the transverse plane of this right-handed coordinate system, positive x points toward the center of the LHC ring, and positive y points toward the sky. The beam pipe defines the orientation of the z-axis, with the positive direction going counterclockwise around the LHC ring. While the LHC coordinate system is Cartesian, the coordinate system preferred for describing LHC events, particularly in the ATLAS experiment, utilizes a particle's transverse momentum (p_T), pseudorapidity (η), and azimuthal angle (ϕ). The polar angle (θ) is defined relative to the beam axis, whereas the azimuthal angle (ϕ) is measured in the plane perpendicular to the beam. In practice, pseudorapidity is preferred over the polar angle. It is defined

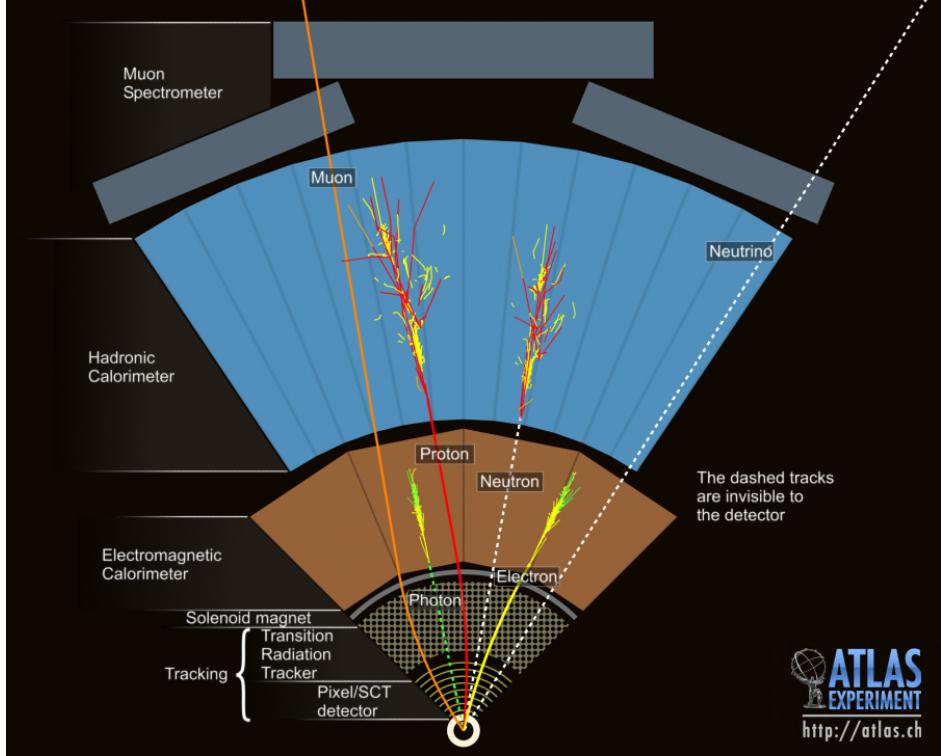


Figure 3.4: The inner layers of ATLAS, depicted in a cross-sectional view along the x-y plane [43].

as

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (3.1)$$

and is a good approximation of the rapidity of a particle in the high-energy regime, which is a measurement of the particle's velocity along the beam axis, given by

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \quad (3.2)$$

While neither the rapidity nor the pseudorapidity are Lorentz invariants, the rapidity difference between the two particles is invariant under Lorentz boosts along the direction of motion. Similarly, the difference in pseudorapidities is approximately invariant under such boosts, especially for highly relativistic particles in LHC/ATLAS events.

3.2.1 The Inner Detector

Located inside the solenoid magnet, the Inner Detector (ID) [45] consists of three subsystems designed to detect charged particles, covering a pseudorapidity range of $|\eta| < 2.5$. As seen in Figure 3.6, these subsystems

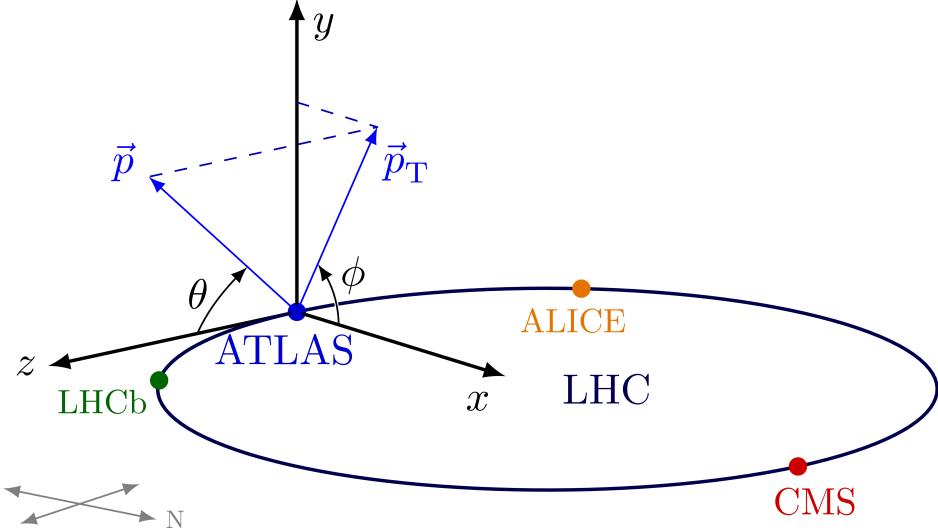


Figure 3.5: The LHC coordinate system as seen from the ATLAS detector [44].

are described from the innermost to the outermost as follows.

The first, the Pixel detector, consists of four layers of pixel detectors. These pixel layers comprise a two-dimensional array of silicon pixels, providing high-precision measurements close to the interaction point. The innermost layer is the Insertable B-Layer (IBL). The “B” in its name stands for “Barrel,” though the IBL is crucial in detecting short-lived particles, like hadrons with b quarks. Due to its closeness to the interaction point, the IBL enhances the performance of the Pixel detector by improving the precisions of vertexing and track reconstruction.

The second subsystem, the SemiConductor Tracker (SCT), comprises four double layers of silicon strips. The SCT behaves similarly to the Pixel detector, except the measurement is one-dimensional. When a charged particle moves through a silicon layer, it bumps electrons to a higher energy level, creating “holes” in their place. This generates a current and signal the detector can read as a “hit.”

The last subsystem is the Transition Radiation Tracker (TRT) [47], a straw tracker comprised of approximately 300,000 4 mm-diameter straw tubes. A straw tube is a long tube with a wire in the center filled with gas (Xenon). When a charged particle passes through, the gas becomes ionized, creating an electronic signal (a “hit”). The superconducting solenoid magnet bends the paths of charged particles based on their momentum, with the numerous hits in the TRT greatly enhancing momentum measurement accuracy.

3.2.2 The Calorimeters

The calorimeter system, enclosing the solenoid magnet and the inner detector, provides coverage up to $|\eta| < 4.9$, although the coverage varies for each subdetector. The calorimeters employ a fundamentally

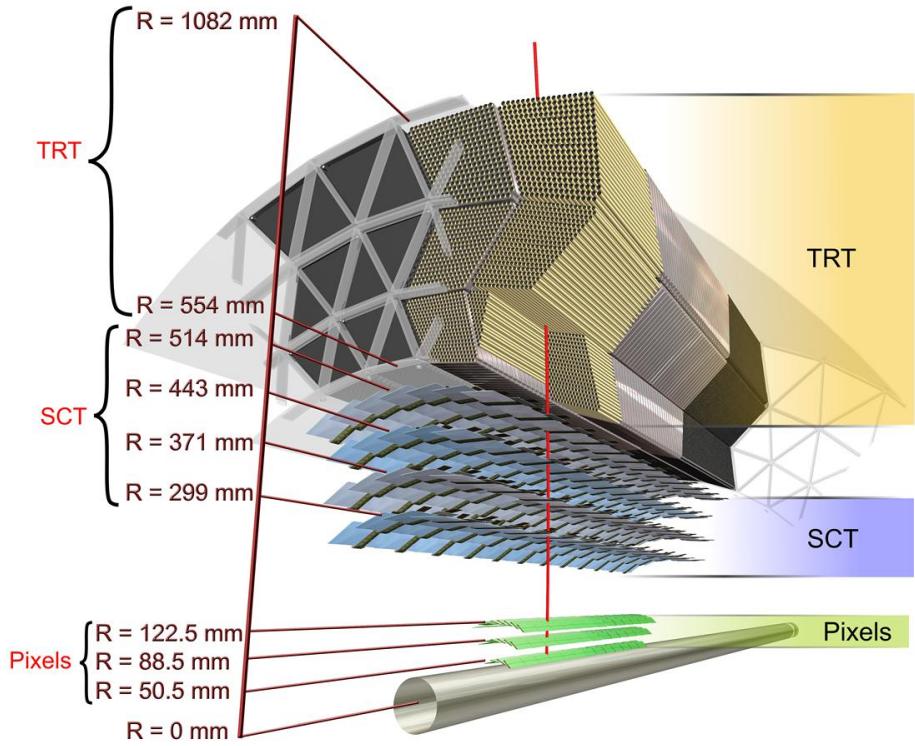


Figure 3.6: ATLAS Inner Detector cut-away: Pixel Detector, Semiconductor Tracker, and Transition Radiation Tracker [46].

different detection method compared to the ID. While the ID is designed to track particles with minimal interaction, the calorimeters are designed to completely absorb the particles. This absorption prevents most particles from reaching further layers of detection. A significant advantage of calorimeters is their ability to detect neutral particles, with the exception of neutrinos. The calorimeter system comprises various subsystems, each responsible for covering a specific range of particles. Figure 3.7 illustrates these subsystems.

Calorimeters operate on a fundamental principle where particles enter a region filled with dense matter, leading to the creation of particle showers as they release their energy within the calorimeter. These devices are typically constructed from alternating layers of materials: a dense material that absorbs the particles, and a sampling material positioned between the absorber layers. The purpose of the sampling material is to detect the particle showers generated by the absorbers. This detection is achieved through sensors connected to the sampling material. There are two prevalent types of sampling materials used in calorimeters: scintillating plastics and ionizable liquids. Scintillating plastics emit light when struck by particles and are read by

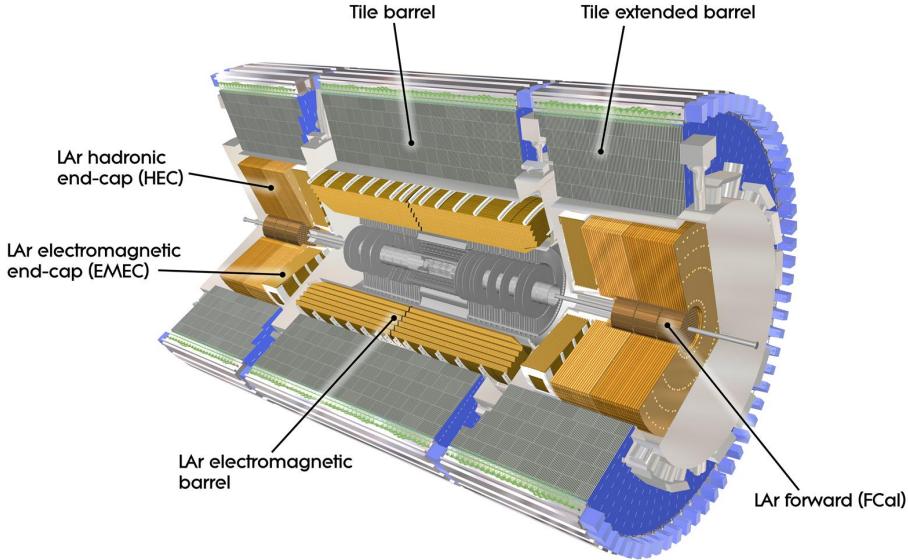


Figure 3.7: Cut-away view of the ATLAS calorimeter system [48].

photodiodes, while the ionizable liquids create ions that are detected by electrodes.

The innermost section of the calorimeter is occupied by the Electro-Magnetic Calorimeter (ECAL) [49]. Noted for its unique accordion design, the ECAL ensures complete azimuthal coverage, eliminating gaps along the azimuthal direction. Despite this, a gap exists between the barrel and endcap sections of the ECAL along the z -direction, leading to a common practice of excluding particles within the pseudorapidity range of $1.37 < |\eta| < 1.52$. The ECAL employs lead as the primary absorbing material, complemented by steel, and utilizes liquid argon as the sampling medium. Upon interaction with incoming particles, the liquid argon becomes ionized. These ionization events are then detected by electrodes shaped to match the accordion design of the calorimeter. The coverage of the ECAL extends to a pseudorapidity range of $|\eta| < 3.2$.

The principal hadronic calorimeter, known as TileCAL [50], is designed to accommodate the penetrating nature of hadrons, which travel deeper into materials before stopping. TileCAL is considerably larger, utilizing steel as its absorbing material and scintillating plastic as the sampling material, ensuring effective energy measurement of hadrons. TileCAL is specifically engineered to cover a pseudorapidity range of $|\eta| < 1.7$. In contrast, the endcaps of the calorimeter, which are constructed differently, employ copper as the absorbing material and liquid argon as the sampling medium. These endcaps extend the calorimeter's coverage to a range of $1.5 < |\eta| < 3.2$, complementing TileCAL. Due to the endcaps being nested within TileCAL's structure, there is a significant degree of overlap between these components, ensuring that there

are no gaps in the hadronic calorimeter's coverage.

The Forward Calorimeter (FCAL) [51] plays a crucial role in capturing particles that travel almost parallel to the beam line, essential for accurately measuring missing energy in experiments. Its design enables it to cover an extended pseudorapidity range of $3.1 < |\eta| < 4.9$, significantly beyond the coverage of the central detectors. Although most physics analyses focus on events within a much narrower range, the ability of the FCAL to detect particles in this extended range is vital for ensuring that no energy goes unaccounted for, which is crucial for precision measurements. The FCAL comprises two distinct sections: an inner section with copper as the absorbing material and an outer section that utilizes tungsten. Liquid argon is employed as the sampling material across the entire FCAL, facilitating the detection and measurement of particle energies.

3.2.3 The Muon Spectrometer

The muon spectrometer (MS) [52], designed to detect muons within $|\eta| < 2.7$, encircles the calorimeters and forms the detector's outermost layer. Its barrel region is defined in the range of $|\eta| < 1.05$, with the endcap regions covering the remaining area. The MS is supplemented by three large toroid magnets surrounding the calorimeters, bending the charged particles in the z direction with a strength of about $2.5 \text{ T} \cdot \text{m}$ in the barrel region and up to $6 \text{ T} \cdot \text{m}$ in the endcaps. Only muons and neutrinos, due to their unique properties, can penetrate the calorimeters to reach the MS, which is specifically designed for precise measurements of muon momentum. As seen in Figure 3.8, the trigger system of the muon system features the Resistive Plate Chamber (RPC) and Thin Gap Chamber (TGC). These trigger chambers offer bunch-crossing identification, well-defined transverse momentum (p_T) thresholds, and measurements of the muon's track coordinate [33]. Combined with the corresponding measurements from the ID, one can determine all three components of a muon's momentum.

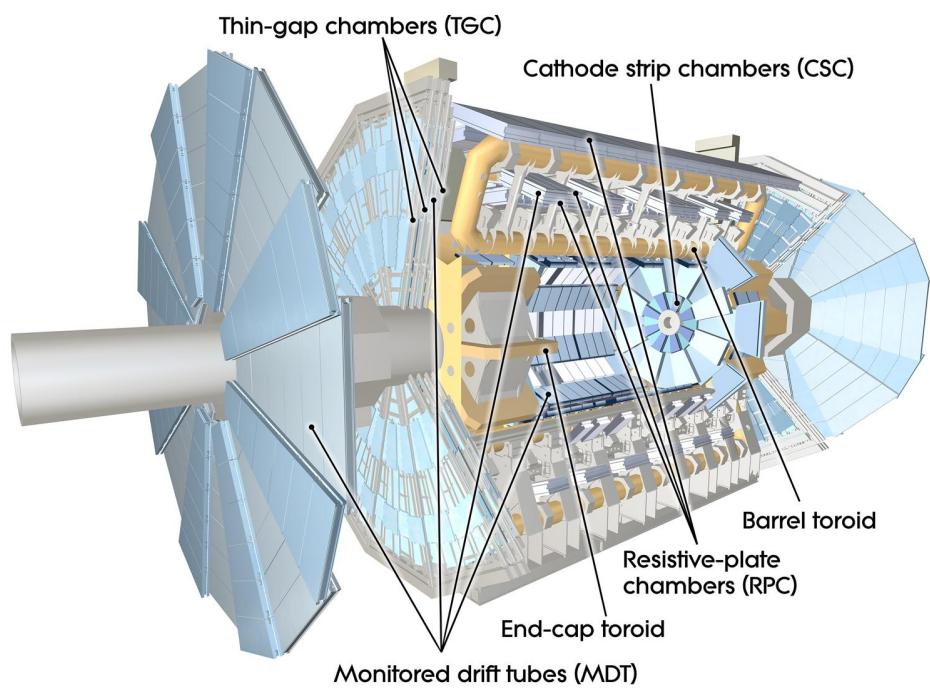


Figure 3.8: Cut-away view of the Muon Spectrometer system [53].

3.3 The Trigger and Data Acquisition

The Trigger and Data Acquisition (TDAQ) system of ATLAS decides which events to record and store. As mentioned in Section 3.1, the 25 ns spacing between proton bunches results in an event rate of 1 GHz, making real-time storage unfeasible. Since the vast majority of the events aren't interesting for physics analysis here at ATLAS, TDAQ's goal is to reduce the collection rate to 1.5 kHz. The trigger system is comprised of two triggers: the hardware-based Level 1 Trigger (L1) [54] and the software-based High Level Trigger (HLT) [55]. Figure 3.9 illustrates how the system works.

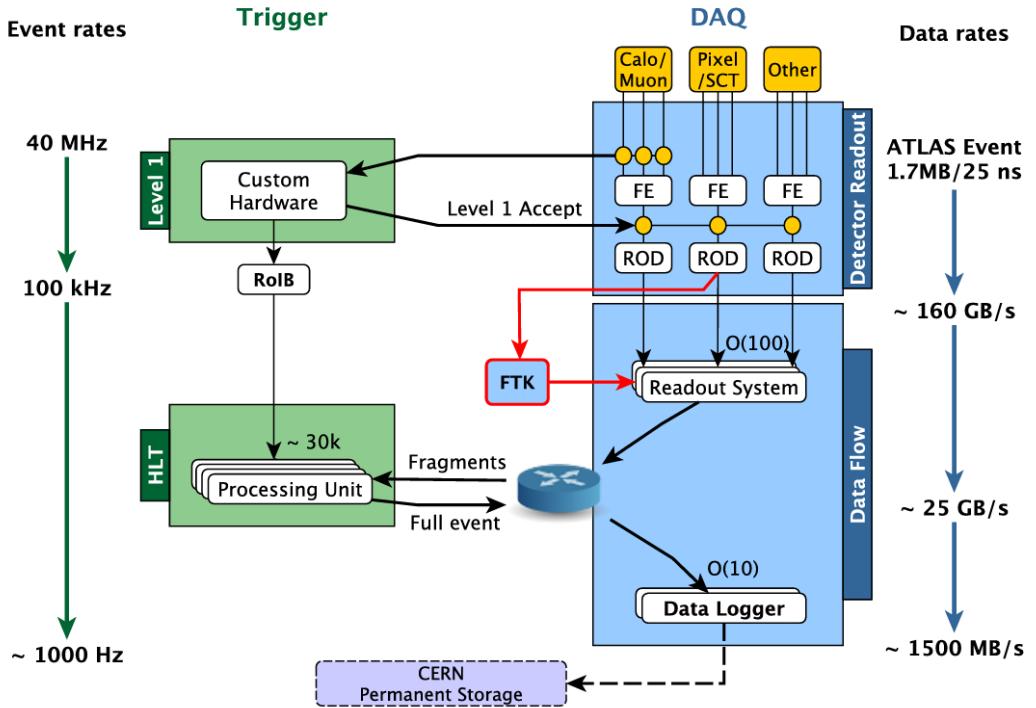


Figure 3.9: ATLAS TDAQ Architecture [56].

The L1 trigger, directly connected to the muon spectrometer and the calorimeters, identifies physics events with interesting objects such as high-momentum muons, significant missing energy, high-energy electrons/photons, or high-energy hadrons. The dedicated hardware for the L1 trigger is situated within the same cavern as the ATLAS detector. Events passing the L1 trigger are transferred from the detectors' Front End (FE) to the Read Out Drivers (ROD) at a reduced the event rate of 100 kHz.

The HLT includes the Level 2 Trigger (L2) and the Event Filter (EF). L1 identifies a region of interest (ROI) of the detector, and the complete event data within this ROI is forwarded to L2. “Tracking” in the ROI is done by L2 to reconstruct the trajectory (or track) of electrically charged particles in the ID. The tracks and calorimeter information are combined to refine the object selection with higher resolution than

what was used in L1, such as tighter requirements on electrons, muons, and missing energy. Events that pass L2 requirements are sent to the EF, which uses data from the entire detector to decide whether to keep the event.

3.4 Fast Tracker

The Fast TrackEr (FTK) system was foreseen as a significant upgrade for the TDAQ system of ATLAS by providing hardware-based tracking information directly to the HLT. As discussed in Section 3.3, although the L1 trigger and HLT perform exceptionally in selecting physics events and reducing the event rate, this implementation does have its limitations. Currently, the HLT receives the "hits" information recorded by the ID, essentially points in space, and conducts tracking calculations within the ROIs determined by the L1 trigger. Tracking calculations in the HLT software are demanding and slow, leading to restrictions on the size and number of ROIs. The ROI approach also resulted in notable inefficiencies in jet finding, due to discrepancies between the online and offline strategies [57] [58]. A solution is to use specialized hardware for this task, which is the primary goal of the FTK system. FTK will be positioned between the L1 trigger and HLT, offering tracking information for all events that pass the L1 trigger across the entire detector and supplying this data to the HLT.

The FTK system consists of various types of boards and components, with details provided in Figure 3.10 and Table 3.2. The Input Mezzanines (IM) receive data directly from the Level 1 trigger (the RODs). There are a total of 12 layers from the Inner Detector: four double SCT layers, three Pixel layers, and one IBL. Clustering is applied to the hits from these layers to reduce data flow. The clustered data is then sent to the Data Formatter (DF) system. The DF system divides the data, sending eight layers (five from the SCT and three from the Pixel) to the main Processing Units (PU) and the remaining four layers (three SCT and one IBL) to the Second Stage Boards (SSB).

Module	Function	Type	Number	
IM	Input Mezzanine	Cluster Pixel and SCT hits, format module data	Mezzanine	128
DF	Data Formatter	Transport and duplicate module hit data to $\eta\phi$ towers	ATCA	32
AUX	Auxiliary Card	Transport coarse-resolution 8-layer hit data to AMB	VME	128
AMB	AM Board	Transport hit data to AM	VME	128
AM	Associative Memory	Match hits to patterns	ASIC	8192
AUX	Auxiliary Card	Evaluate track candidates in matched patterns	VME	128
SSB	Second Stage Board	Add remaining hits to 8-layer tracks, fit, remove overlaps	VME	32
FLIC	HLT Interface Board	Interface to ATLAS readout	ATCA	2

Table 3.2: Overview of the FTK system in the order in which the data are processed. The AUX appears in the table twice because of its dual functions [59].

The main Processing Units are made up of two boards: the Associated Memory Board (AMB) and the

FTK Overview

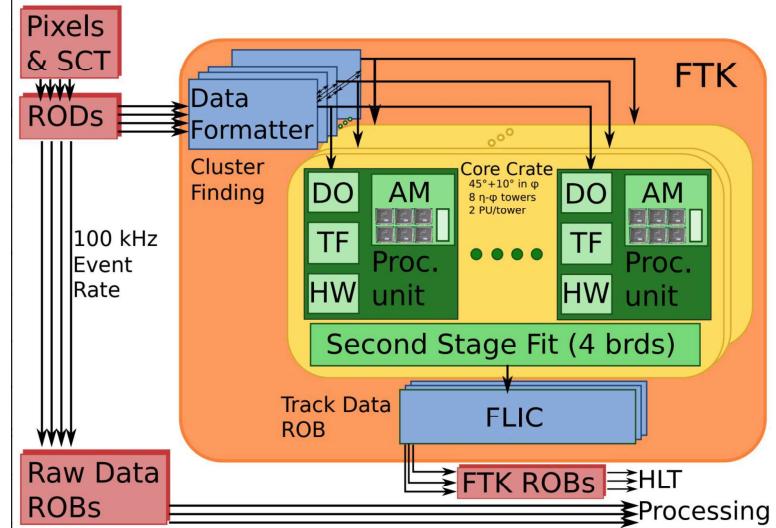


Figure 3.10: Overview of the FTK system.

Auxiliary Board (AUX), with the AUX serving as a rear transmission module. The AMB is essential to the FTK system. To construct tracks rapidly, billions of precomputed tracks are stored in a large memory bank. These track patterns are then matched to the incoming hit coordinates using massive parallel processing. Each AMB can match eight million patterns in parallel to the incoming hit data.

The Auxiliary Board (AUX) communicates with its corresponding Associated Memory Board (AMB) via the VME backplane, performing a range of functions. Its principal duty involves eight-layer track fitting. The procedure is as follows:

- Input data from the DF is forwarded to the AMB.
- The AMB, after matching patterns, returns them to the AUX.
- Utilizing these patterns, the AUX conducts a fit to assess their quality, allowing tracks to be missing a hit on one layer – these are termed *majority tracks* – to boost efficiency.
- Moreover, the AUX undertakes the removal of duplicates among the eight-layer track candidates.
- Valid eight-layer tracks are then forwarded to the Second Stage Boards (SSB).

The Second Stage Board (SSB) system includes 32 boards, each responsible for managing tracks from four PUs. The SSB integrates these eight-layer tracks with hit data for the four additional layers from the corresponding DF, to assemble 12-layer tracks. This process involves:

- Extrapolating the eight-layer track into the missing layers.
- Using hits near the extrapolated coordinates to fit full 12-layer tracks.

Like in the first stage, it is possible for tracks to be missing a hit in one of the new layers. After fitting, tracks undergo a global duplicate check before being forwarded to the FTK Level2 Interface Crate (FLIC). SSBs are paired to cover the entire detector range, with pairs linked in a ring to manage overlap.

The FTK Level2 Interface Crate (FLIC) is the last element in the FTK system, serving two primary purposes:

- It merges the 12-layer track data from the 32 SSBs.
- It formats the consolidated data so that it can be interpreted by the Read Out System (ROS).

3.4.1 Second Stage Board

The Second Stage Boards (SSBs), designed and manufactured at the University of Illinois, perform three critical functions (see Figure 3.11):

- Extrapolating incoming eight-layer tracks to include the missing layers,
- Fitting these extrapolations to form 12-layer tracks,
- Removing any duplicates found across the detector.

The hardware of the SSBs includes five Field-Programmable Gate Arrays (FPGAs) that require firmware. Four of these FPGAs, designated as the EXTF FPGAs, are dedicated to the tasks of extrapolation and track fitting. The fifth FPGA, known as the Hit Warrior (HW) FPGA, is tasked with duplicate removal. Furthermore, the SSBs are equipped with a substantial amount of Reduced Latency Dynamic Random-Access Memory (RLDRAM) for storing the constants necessary for extrapolation and track fitting calculations. The layout and the dataflow through the SSB are shown in Figure 3.12.

The core functionality of the EXTF FPGA is allocated to the Extrapolator (EXP), which oversees the extrapolation process, and the Track Fitter (TF), focused on the fitting of the resulting 12-layer tracks. For a thorough exploration of these firmware aspects, see Markus Atkinson's PhD thesis [60].

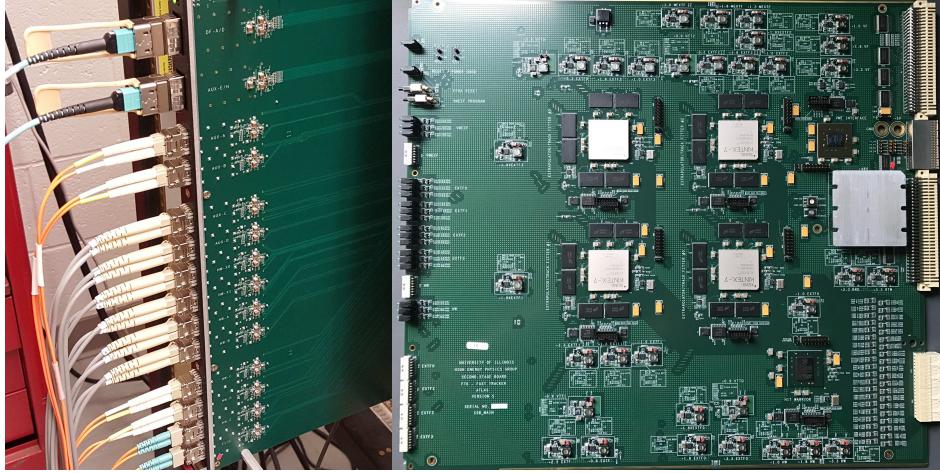


Figure 3.11: The physical Second Stage Boards [59].

3.4.2 Hit Warrior FPGA

The Hit Warrior FPGA on the SSB serves two key roles. Its main job is to eliminate duplicate tracks. Its second role involves merging data. Each SSB houses four EXTF FPGAs, and the Hit Warrior FPGA collects all their output. This output is then scanned for any duplicates, combined into a single large output packet, and forwarded to the FLIC. The top-level diagram of the Hit Warrior firmware design is depicted in Figure 3.13. In the diagram, the Sync Engine module synchronizes the four incoming EXTF streams. The Hit Warrior module is the logic component that performs the comparison and duplicate removal tasks. Additionally, Spy Buffers are utilized for monitoring purposes. More details on the Hit Warrior module follow.

The Track Parser functions as the first component of the Hit Warrior module, where it examines the data to identify the start of track packets. These packets always start with the keyword “BDA” and consist of 28 words. Each track is distributed throughout the FPGA to enable parallel processing of track comparisons. Tracks are processed sequentially, one after another, which makes it practical to use a track counter. The specific processing path for each track depends on this counter. However, the overarching strategy ensures that each new track is compared with all previously received tracks simultaneously. By the time the final track is processed, all necessary track comparisons have been completed.

The first and simplest component to understand in this context is the RAW FIFO (First In, First Out). This acts as an internal storage or buffer for all the tracks corresponding to the current event, storing them in the exact order they were received. The layout is illustrated in Figure 3.14.

Tracks are also sent to units called Comparators. The total number of Comparators is limited by available resources, which in turn sets a maximum on the number of tracks that can be effectively compared for an

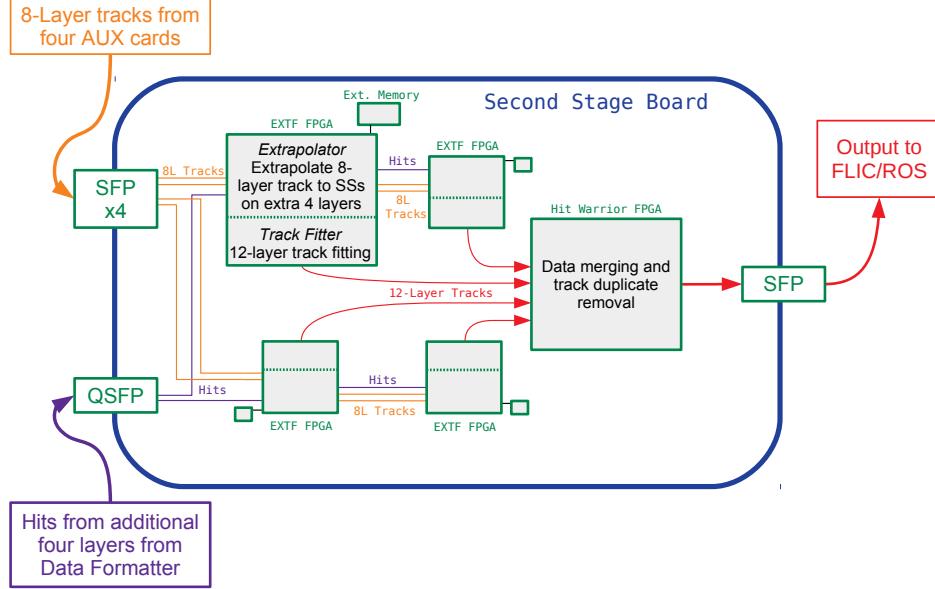


Figure 3.12: Diagram of the dataflow through the Second Stage Board [59].

event. Each Comparator is equipped with a small piece of RAM (Random Access Memory) capable of holding a single track packet. Furthermore, each Comparator is directly connected to the stream of incoming tracks, allowing it to receive and compare new track data as it arrives.

The comparator conducts comparisons between two tracks on a word-to-word basis to identify matches. Each track packet contains 28 words, comprising 12 for the Track Header and 16 for the hit list. These 16 words are distributed across 12 layers. A match requires at least eight layers to be identical, a criterion determined through simulation to ensure maximum efficiency. For 2D pixel layers, both coordinates must match.

When matches are detected, a selection criterion decides the optimal track. Tracks with the same number of layers are evaluated on their χ^2 values to select the best one. If the layer count varies, the track with the greater number of layers is chosen. The selection scheme is outlined in the Table 3.3.

Track 1	Track 2	Decision
Nominal	Nominal	Select Track with lowest χ^2
Nominal	Majority	Select Nominal Track
Majority	Majority	Select Track with lowest χ^2

Table 3.3: Selection scheme for matching tracks.

After each track is compared, the outcome is forwarded to the Decoder, which logs the results of all track comparisons. Each comparator can issue one of four possible outcomes, encoded as two bits: *idle*, *no match*, *delete RAM*, and *delete Stream*. The Decoder synthesizes these outcomes with the current track count and

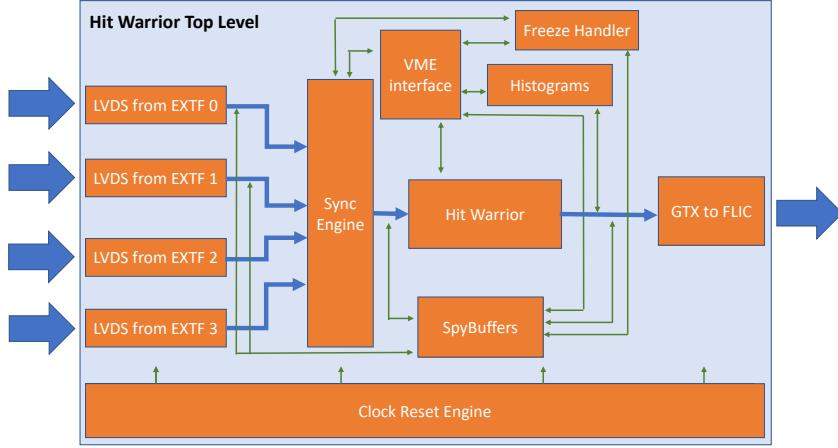


Figure 3.13: Diagram of the Hit Warrior firmware [59].

the comparator's index to generate a “copy vector.” This vector is essential for managing the write process to the CUT FIFO.

Once every track for the event has been examined and the copy vector is complete, the RAW FIFO’s content is moved. This FIFO contains a copy of every track. This data is then transferred to a second FIFO, named the CUT FIFO, guided by the copy vector to ensure only the valid tracks are preserved. This arrangement allows for the removal of gaps in the data by excluding the tracks as dictated by the copy vector.

Comparator Output	Meaning
Idle	No operation
No Match	No duplicate found
Delete RAM	Remove track from RAM
Delete Stream	Remove track from incoming stream

Table 3.4: Comparator Results and Their Meanings

The CUT FIFO then serves as the final repository for the processed tracks, ensuring that only the relevant tracks are stored for further processing. The decoder, using comparator flags, the index number of each comparator, and the track counter, identifies tracks for deletion. For instance, comparator M can only eliminate track M stored in its RAM or the track currently indexed by the track counter. Each processed track is compared in parallel to all prior ones. In each comparison round, any previous track might be marked for deletion by the incoming track. After each round, the decoder updates the copy vector to record any tracks flagged for removal. When the last track of the event has been compared, the final version of the copy vector is used to regulate the transfer of data from the RAW FIFO to the CUT FIFO, ensuring only

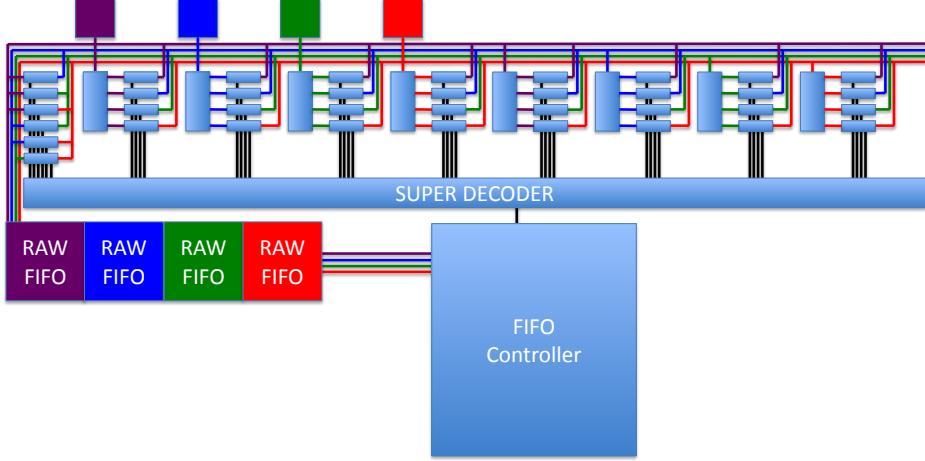


Figure 3.14: Hit Warrior FIFO Diagram

the required tracks are retained.

It became apparent later in development that the prototype Hit Warrior firmware couldn't keep up with the required processing speed. To address this, the system was adapted to process multiple streams in parallel. The Parallel Hit Warrior operates much like the original but at quadruple the speed by handling four streams simultaneously. This choice is aligned with the four EXTF FPGAs on each SSB. Therefore four tracks are inputted into four separate comparators. These comparators have been adapted to match a track in RAM against four incoming stream tracks at the same time. Additionally, a new component, the Cross Comparator, was introduced to compare all six pairings among the four stream tracks. The layout of the Parallel Hit Warrior is shown in Figure 3.15a, and the details of the comparator are revealed in Figure 3.15b.

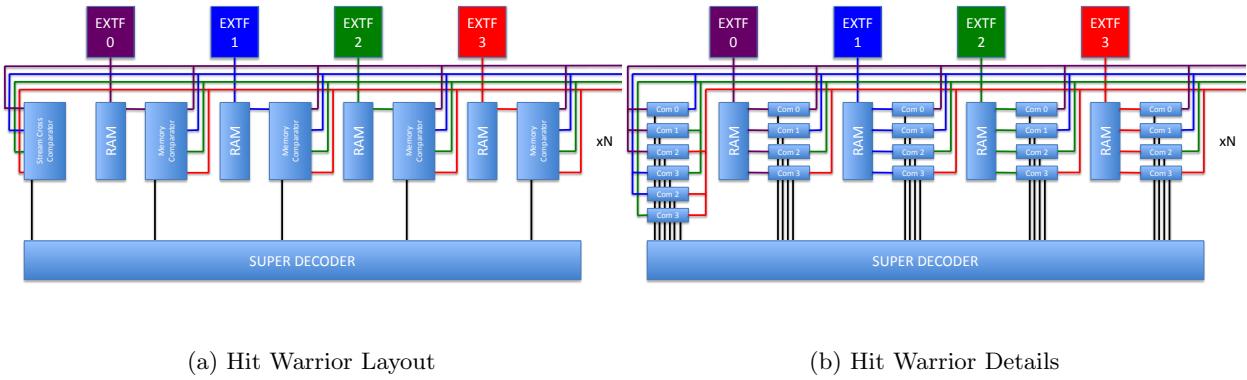


Figure 3.15: Hit Warrior System Components

The firmware design for the HitWarrior FPGA utilizes circular buffers, known as Spy Buffers, to capture

data as it moves through the system, alongside the capability to read out monitoring registers. It also allows for the freezing of monitoring data in the event of errors to aid in debugging. Furthermore, these buffers store synchronized monitoring information and compile histograms of monitoring data. While this firmware design is broadly applied across various components of the FTK system, including the design of synchronization blocks, the SSB version is tailored for the HitWarrior FPGA to meet specific design needs and to accommodate the architecture of the FPGA chip. As shown in Figure 3.13, the Spy Buffers monitor the four parallel streams from the EXTF FPGAs, and the output of the Hit Warrior module before sending it to the FLIC.

My role in the development of the SSB involved updating the firmware for the Hit Warrior FPGA, focusing on parallelizing the prototype Hit Warrior firmware and implementing the Spy Buffers. I was deeply involved, often as the sole representative for the SSB, in its commissioning and integration into the ATLAS trigger system. This required extensive hours in the ATLAS control rooms and the underground facilities where the FTK system is installed.

Chapter 4

Experimental Methods

4.1 Data Samples

The semileptonic VBS analysis is based on proton-proton (pp) collision data collected from 2015 to 2018, amounting to an integrated luminosity of 139 fb^{-1} . The overall uncertainty for the integrated luminosity, measured at 1.7% [61], was determined using the LUCID-2 (Luminosity Cherenkov Integrating Detector) [62] for primary luminosity measurements. Table 4.1 summarizes the integrated luminosity used in this analysis.

Table 4.1: Integrated luminosity data used in the semileptonic VBS analysis.

Year	$\mathcal{L} [\text{fb}^{-1}]$
2015	3.21
2016	32.88
2017	44.31
2018	58.45
total	139.0 ± 2.4

4.2 Monte Carlo Simulated Samples

Monte Carlo (MC) simulations are integral to nearly all stages of the analysis workflow. These simulations are essential for background modeling, signal acceptance evaluation, event selection optimization, systematic uncertainty estimation, and statistical analysis. The MC samples are generated with ATLAS-approved settings. The EvtGen v1.2.0 program [63] simulates bottom and charm hadron decays. To account for pileup, additional pp collisions are simulated using PYTHIA 8.186 [64] and integrated into all MC events. The samples undergo a comprehensive simulation of the ATLAS detector [65] using GEANT4 [66]. All simulated events are processed using the same trigger and reconstruction algorithms as the collected data. Appendix A summarizes the MC samples for SM background and EW signal processes.

Background processes

The V (W/Z) + jets events are simulated with Sherpa 2.2.1 [67] and normalized to the NNLO cross sections. Up to 2 partons at NLO and 4 partons at LO are considered using Comix [68] and OpenLoops [69] for matrix element calculations, integrated with the Sherpa parton shower [70] using the ME+PS@NLO prescription [71]. These simulations use the NNPDF3.0NNLO PDF set alongside specific tuning [72].

As alternative V + jets MC samples for the analysis, QCD V +jets production was simulated using MADGRAPH5_aMC@NLO 2.2.2 [73] with LO-accurate matrix elements (ME) featuring up to four final-state partons. These ME calculations utilized the NNPDF3.0NLO PDF set [74] for H_T -sliced and the NNPDF2.3LO set [75] for N_{parton} -sliced events. The events were then interfaced with PYTHIA 8.186 [76] to simulate the parton shower, hadronization, and the underlying event dynamics. The CKKW-L merging procedure [77, 78] was applied to eliminate overlap between matrix element calculations and parton shower emissions. The A14 tune of PYTHIA 8 [79], in conjunction with the NNPDF2.3LO PDF set [75], was used. Decays of bottom and charm hadrons were handled by EVTGEN 1.2.0 [63]. Although not the nominal samples, these V +jets samples were normalized to NNLO predictions [80] and served to derive modeling uncertainties for the V + jets background.

The $t\bar{t}$ and single-top events are generated using Powheg-Box [81] and NNPDF3.0NLO PDF sets [74] for matrix element calculations, ensuring top quark spin correlations are preserved. Specifically, t-channel top quarks are decayed with MadSpin [82]. PYTHIA8.230 simulates the parton shower, fragmentation, and underlying event dynamics using the A14 tune set [79], with the top quark mass set at 172.5GeV. The cross sections for $t\bar{t}$ and single-top processes are calculated with NNLO precision in QCD, which includes the re-summation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms [83–86]. The HDAMP parameter, which regulates high- p_T radiation in POWHEG, is set to $1.5m_t$ to ensure good data/MC agreements in the high- p_T region [87].

The $q\bar{q}$ -induced diboson processes (WW , WZ , and ZZ) are simulated using SHERPA 2.2.1 or 2.2.2 [88], as per the specific process. These simulations incorporate off-shell effects and, where relevant, Higgs boson contributions. The matrix elements are calculated with NLO QCD accuracy for up to one additional parton and LO accuracy for up to three additional partons.

Loop-induced $gg \rightarrow VV$ processes were simulated using LO-accurate matrix elements with up to one additional parton emission, covering both fully leptonic and semileptonic final states. These calculations were harmonized and combined with the SHERPA parton shower via Catani–Seymour dipole factorisation [68, 70], following the MEPS@NLO prescription [71, 89–91].

QCD diboson production from gg initial states is excluded from the final results due to its significantly

smaller expected cross section compared to the baseline QCD VV production; their impact has been evaluated and found to be negligible.

Signal SM EW $VV+jj$ processes

The EW $VV + jj$ production is simulated with MadGraph5_aMC@NLO v2.3.3 [73] and PYTHIA 8 [76] for fragmentation, using the NNPDF30NLO PDF set [75]. The samples feature two on-shell V bosons: one decaying leptonically ($Z \rightarrow \ell\ell$ where $\ell = e, \mu$; $Z \rightarrow \nu\nu$; $W \rightarrow \ell\nu$ with $\ell = e, \mu, \tau$), and the other hadronically. Both W^+ and W^- decays are included, and for $WWjj$, all charge combinations (W^+W^+ , W^+W^- , and W^-W^-) are considered. Table 4.2 details the EW $VV + jj$ samples used in this analysis. In these samples, we account for all purely-electroweak tree-level diagrams of order $\mathcal{O}(\alpha_{EW}^6)$ that contribute to the final states. This includes both VBS diagrams and non-VBS electroweak diagrams. Detailed examples of these diagrams have been previously discussed and are showcased in Figures 2.4 and 2.5, as found in Section 2.3.

For EW $WW + jj$ production, the non-VBS electroweak $t\bar{t}$ diagrams, as depicted in Fig. 2.5(e) and (f), could significantly contribute. This potential issue is effectively mitigated through the use of a b -veto. Despite this mitigation, the contribution from the non-VBS tZb process, illustrated in Fig. 4.1, remains considerable. To address this, a cut was implemented near the reconstructed top mass, effectively reducing its impact, as detailed in Table 6.2.

Diagrams that feature both electroweak and QCD vertices, with an order of $\mathcal{O}(\alpha_{EW}^4 \alpha_s^2)$, are not included in these signal samples and are not part of the signal definition. Examples of such diagrams are depicted in Fig. 4.2. These processes, unaffected by aQGCs, are considered part of the background, including $t\bar{t}$, single-top, and diboson events.

Table 4.2: List of VBS samples used in the analysis.

Process	DSID	Events - mc16a	Events - mc16d	Events - mc16e	Filter efficiency	cross-section (pb)
$W(\ell\nu)W(qq')jj, b - veto$	364848	1958000	2296000	3320000	0.17465	1.9994
$W(\ell\nu)W(qq')jj, b - filter$	364849	1996000	2400000	3388000	0.83126	1.9777
$W(\ell\nu)Z(qq')jj$	364850	1994000	2394000	3392000	1.0	0.2571
$Z(\nu\nu)W(qq')jj$	364851	1986000	2394000	3356000	1.0	0.15532
$Z(\ell\ell)W(qq')jj$	364852	1996000	2374000	3390000	1.0	0.045609
$Z(\nu\nu)Z(qq')jj$	364853	1998000	2396000	3390000	1.0	0.032238
$Z(\ell\ell)Z(qq')jj$	364854	1990000	2388000	3396000	1.0	0.0096553

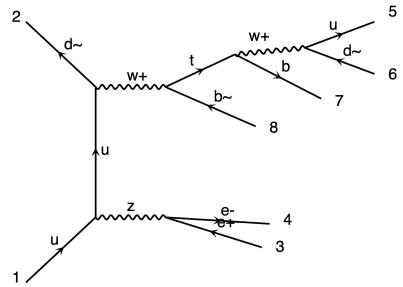


Figure 4.1: The example of the tZb diagram, which included in non-VBS $\mathcal{O}(\alpha_{EW}^6)$ diagrams.

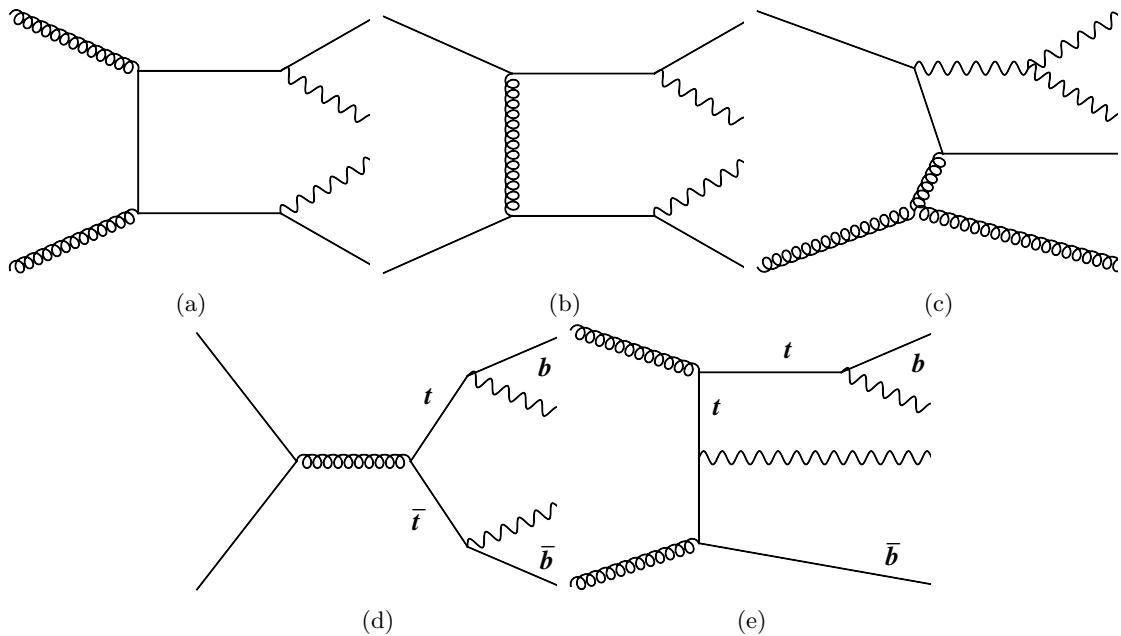


Figure 4.2: Examples of $\mathcal{O}(\alpha_{EW}^4 \alpha_S^2)$ diagrams that lead to the $VV+2$ parton final state.

Signal aQGC processes

As discussed in Section 2.4.1, the Eboli model is used to model potential aQGC effects in the VBS process. This model includes 21 D-8 operators, 19 of which can alter the semileptonic VBS final state. The matrix element for a SM process, incorporating new contributions from EFT, can be expressed as follows:

$$|A_{\text{SM}} + c_i A_i|^2 = |A_{\text{SM}}|^2 + \sum_i c_i^2 |A_i|^2 + \sum_i 2c_i \text{Re}(A_{\text{SM}}^* A_i) + \sum_{i \neq j} c_i c_j \text{Re}(A_i^* A_j) \quad (4.1)$$

Here, $|A_{\text{SM}}|^2$ denotes the squared matrix element for the SM process, $|A_i|^2$ represents the squared matrix element for pure EFT contributions, $2\text{Re}(A_{\text{SM}}^* A_i)$ is the interference between the SM and EFT contributions, and $\text{Re}(A_i^* A_j)$ is the interference among different EFT contributions. This formulation allows for a detailed breakdown of the process, which can be useful with the matrix-element decomposition method available in recent versions of MadGraph5_aMC@NLO, enabling the specific generation of each term individually.

To model the Eboli D-8 EFT processes, we generate a series of samples for each operator A_i at varying coefficient values (c_i). We utilize MadGraph5_aMC@NLO version 2.7.2 for these simulations at leading order (LO), incorporating the NNPDF30LO parton distribution function (PDF) set [75], and Pythia 8.244 [76] for fragmentation processes. The matrix element calculation results in two on-shell W/Z bosons, which are subsequently decayed using MADSPIN [82] to represent the leptonic and hadronic decay channels. Due to the matrix-element weighting requirements, simulations for each weak isospin state of the boson (e.g., W^+W^- , W^+W^+ , W^-W^- , W^+Z , W^-Z , ZZ) are conducted separately.

4.3 Derivation and CxAOD framework

For the 1-lepton analysis, we employ the HIGG5D2 derivation, which is produced by Athena's derivation framework. This framework consists of tools, algorithms, and Python scripts designed to produce analysis data formats known as derived AODs (DAODs), essential for initiating most analyses. DAODs streamline the analysis process by providing a refined dataset, having undergone preliminary selection cuts to minimize data size. These initial cuts are listed in Table 4.3. Typically, DAODs are generated by a central production team, meaning individual researchers usually engage with the framework directly only for testing purposes on a smaller scale.

We use the common analysis CxAOD framework (version r33-12) for this analysis. The CxAOD framework is used to define physics objects, adhere to all CP recommendations for them, and apply pre-selection cuts. These procedures and definitions are detailed in Sections 5 and 6.2, respectively. The CP recommendations

Table 4.3: Pre-selections in the derivation framework for data size reduction. N_j , N_J^{TCC} , and N_J^{LCTopo} represent the counts of different types of jets.

HIGG5D2 ($\ell\nu qq$)	
Trigger	<i>OR</i> of all single-lepton and E_T^{miss} triggers
Number of leptons (lepton p_T)	≥ 1
(electron quality)	$> 20 \text{ GeV}$
(muon quality)	DFCommonLHLoose DFCommonGoodMuon
Number of jets	$N_j \geq 1$ or $N_J^{\text{TCC}} + N_J^{\text{LCTopo}} \geq 1$

are providede by the combined performance (CP) groups for use in analysis.

Chapter 5

Object Definition

This section describes the definitions of the physics objects used in the analysis.

5.1 Electron Definition

Electron candidates for the analysis were chosen based on specific selection criteria, focusing on their momentum and isolation from surrounding detector activity to identify signal electrons effectively. Electrons traversing the calorimeter’s crack were excluded. In this 1-lepton channel analysis, exactly one signal electron is required. Electrons failing to meet basic quality standards during data collection are excluded. The criteria for electron isolation are optimized to reduce the presence of non-prompt electrons. The non-prompt particles, like electrons, are not directly produced at the primary interaction vertex, but rather stem from secondary processes or decays. This analysis uses two electron definitions, Tight and Loose, summarized in Table 5.1.

- “Tight” electron: used to select $W \rightarrow e\nu$ candidate.
- “Loose” electron: used to veto events with additional leptons.

Table 5.1: Summary of Electron Selections

Criteria	Loose	Tight
Pseudorapidity range	$ \eta < 2.47$ (veto in $[1.37, 1.52]$ region)	
Energy calibration	“es2017_R21_v0” (ESModel)	
Transverse momentum	$p_T > 7 \text{ GeV}$	$p_T > 28 \text{ GeV}$
Object quality [92]	Not from a bad calorimeter cluster (BADCLUSELECTRON) Remove clusters from regions with EMEC bad HV (2016 data only)	
Track to vertex association	$ d_0^{BL}(\sigma) < 5$ $ \Delta z_0^{BL} \sin \theta < 0.5 \text{ mm}$	
Identification	Loose	Tight
Isolation	FCLoose at $p_T < 100 \text{ GeV}$ and no isolation requirement at $> 100 \text{ GeV}$	FixedCutHighPtCaloOnly

5.2 Muon Definition

Muons are chosen similarly to electrons, with the key distinction being the use of the muon spectrometer instead of the calorimeter. They must meet comparable momentum and isolation criteria, plus a consistency check aligning tracks from the inner detector to the muon spectrometer [93]. Muons failing basic quality standards during data collection are excluded. The criteria for muon identification and isolation are optimized to reduce the presence of non-prompt muons. This analysis uses two muon definitions, Tight and Loose, summarized in Table 5.2.

- “Tight” muon: used to select $W \rightarrow \mu\nu$ candidate.
- “Loose” muon: used to veto events with additional leptons.

Table 5.2: Summary of Muon Selections

<i>Criteria</i>	<i>Loose</i>	<i>Tight</i>
Pseudorapidity range	$ \eta < 2.5$	
Momentum Calibration	Sagitta Correction [93] Used	
Transverse momentum	$p_T > 7 \text{ GeV}$	$p_T > 28 \text{ GeV}$
d_0 Significance Cut	$ d_0^{BL}(\sigma) < 3$	
z_0 Cut	$ z_0^{BL} \sin \theta < 0.5 \text{ mm}$	
Selection Working Point	Loose	Medium
Isolation Working Point	FixedCutLoose at $p_T < 100 \text{ GeV}$ and no isolation requirement at $> 100 \text{ GeV}$	FixedCutTightTrackOnly

5.3 Jet Definitions

As previously discussed in Section 2.1, the detection of quarks is indirect due to quark confinement. The hadronization processes manifest in the detector as a shower of hadrons. Using clustering algorithms, these hadronic showers can be reconstructed as a cone-shaped energy pattern, known as a jet.

5.3.1 Small-R jet selection

When a heavy particle decays, it can produce two quarks that rapidly move away from each other, leading to the formation of two distinct jets of particles in the detector. If the decaying particle is moving slowly, the resulting jets tend to be more compact and easily distinguishable. These compact jets, characterized by a smaller radius parameter (R) in their clustering algorithm, are referred to as Small-R jets.

Small-R jets are used for reconstructing the less boosted $W/Z \rightarrow qq$ candidates, identifying qq pairs from V boson decays as “Signal” jets and forward jets in vector-boson scattering as “VBS” jets. Sections 6.4 and 6.5.1-6.5.2 provide in-depth discussions on Signal and VBS jets. We use the ATLAS baseline jet reconstruction algorithm, the `AntiKt4EMPFlowJets`, which is anti- k_t clustering algorithm [94]. Standard jet calibrations are implemented, including the use of Jet Vertex Tagging (JVT) with a Medium Working Point (WP) to reduce pile-up interactions [95]. Pile-up interactions refer to the multiple, simultaneous collisions of other proton pairs occurring alongside the primary collision event. To further suppress pile-up jets, especially in the forward-like topology relevant to our analysis, the forward-JetVertexTagger (fJVT) [96] is applied, enhancing the focus on jets from the targeted VBS processes. Small-R jet selection criteria is summarized in Table 5.3.

Table 5.3: Summary of Selection Criteria and Calibration Methods for Small-R Jets

Jet Reconstruction Parameters	
Parameter	Value
algorithm	anti- k_T
R-parameter	0.4
input constituent	EMPFlow
Analysis Release Number	21.2.164
Calibration sequence (Data)	JetArea_Residual_EtaJES_GSC_Insitu
Calibration sequence (MC)	JetArea_Residual_EtaJES_GSC_Smear
Selection Requirements	
Observable	Requirement
Jet cleaning	LooseBad
BatMan cleaning	Yes
p_T	>20 GeV ($ \eta < 2.5$) and >30 GeV ($2.5 < \eta < 4.5$)
$ \eta $	< 4.5
JVT	> 0.5 for $60 \text{ GeV} < p_T < 120 \text{ GeV}$ and $ \eta < 2.4$
WP	Medium
fJVT	> 0.5 (and $ timing < 10 \text{ ns}$) for $p_T < 120 \text{ GeV}$ and $2.5 < \eta < 4.5$
WP	Loose
b -tagging (See Sec. 5.5)	Tagged, or not tagged

5.3.2 Large-R jet selection

For high- p_T $W/Z \rightarrow qq$ candidates, the angle between the two jets narrows, leading to the merging of jets, as shown in Figure 5.1. Clustering algorithms then reconstruct them into a single, larger-radius jet, known as a large-R jet, which contains the two merged sub-jets. Following the trimming procedure [97] to reduce pile-up and soft radiation effects, the “jet mass”, m_J , is reconstructed by summing the four-vectors of jet constituents. The large-R jets then undergo baseline kinematic cuts:

- $p_T^J > 200 \text{ GeV}$
- $|\eta|^J < 2$
- $m^J > 50 \text{ GeV}$

The (large-R) jet substructure variable, D_2 , derived from the energy correlation functions based on energies and pair-wise angles of the sub-constituents [98,99], is sensitive to the expected 2-prong sub-structure from the boosted W/Z bosons decay. The variable D_2 is defined as

$$D_2^{(\beta=1)} = E_{CF3} \left(\frac{E_{CF1}}{E_{CF2}} \right)^3 \quad (5.1)$$

where the energy correlation functions (E_{CF}) are defined as:

$$\begin{aligned} E_{CF1} &= \sum_i p_{T,i} \\ E_{CF2} &= \sum_{ij} p_{T,i} p_{T,j} \Delta R_{ij} \\ E_{CF3} &= \sum_{ijk} p_{T,i} p_{T,j} p_{T,k} \Delta R_{ij} \Delta R_{jk} \Delta R_{ki} \end{aligned} \quad (5.2)$$

The track multiplicity, n_{Tracks} , of the ungroomed large-R jet is also used to enhance background rejection. It is particularly sensitive to QCD jets from single-quark and gluon decays. The W/Z boson tagger, based on three variables— m_J , D_2 , n_{Tracks} —is used to identify boson jets from large-R jet candidates. The application of this tagger and the three variables in defining the merged category is elaborated in Section 6.5.1.

Summary of selections and calibrations of the large-R jet is shown in Table 5.4.



Figure 5.1: Illustration of jets merging

Table 5.4: Summary of Selection Criteria and Calibration Methods for Large-R Jet

Jet Reconstruction Parameters	
Parameter	Value
algorithm	anti- k_T
R-parameter	1.0
input constituent	LCTopoCluster
grooming algorithm	Trimming
f_{cut}	0.05
R_{trim}	0.2
Analysis Release Number	21.2.164
Calibration sequence (Data)	EtaJES_JMS_Insitu_InsituCombinedMass
Calibration sequence (MC)	EtaJES_JMS
Selection Requirements	
Observable	Requirement
p_T	>200 GeV
$ \eta $	<2.0
mass	> 50 GeV
SmoothedWZTagger	
Object	Working point
W/Z	3-var tagger working point with $V = W, Z$ and $XX = 50, 80$

5.4 Missing Energy Definition

Missing transverse energy, E_T^{miss} , is the total energy of all the undetected particles in an event. In this analysis, with just one neutrino in the final state, it's assumed that the E_T^{miss} directly corresponds to this single neutrino's energy. The E_T^{miss} reconstruction is done based on the signals of detected particles in the final state.

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss}, e} + E_{x(y)}^{\text{miss}, \gamma} + E_{x(y)}^{\text{miss}, \tau} + E_{x(y)}^{\text{miss, jets}} + E_{x(y)}^{\text{miss}, \mu} + E_{x(y)}^{\text{miss, soft}} \quad (5.3)$$

After the calibrations for pileup, two types of contributions are taken into account in the reconstruction.

- hard-event signals: fully reconstructed and calibrated particles (e, γ, τ, μ) and small-R jets
- soft-event signals: soft tracks in the inner detector not parts of any reconstructed physical object

When taking the negative vectorial sum of all contributing physics objects, only the calorimeter signals are used to avoid double counting [100].

5.5 Jet Flavor (b) Tagging Definition

Due to the short lifetime of the t quark, it decays before hadronization, with a 99.8% chance of decaying into a b quark. As a result, a b -tagging method is used to reject events originating from $t\bar{t}$ and single top events. In this analysis, small-R jets are tagged using the new DL1r b -tagging score available in ATLAS [101]. The DL1r b -tagging is based on a deep-learning neural network using distinctive features of b -hadrons [102]. As specified in Section 6.6, the b -tagged jets are used to define the TopCR within the 1-lepton phase space. A summary of the b -tagging requirements in the resolved category is shown in Table 5.5.

Table 5.5: Summary of b -tag requirements in the resolved category

Jet collection	AntiKt4PFlowJets
Jet selection	“Signal jet” and “VBS jet” selection in Table 5.3
Algorithm	DL1r
Operating Point	Fixed, Efficiency = 70%

5.6 Overlap Removal

We utilize the standard overlap removal tools, `AssociationUtils` [103], to manage the overlaps between analysis-level objects (electrons, muons, jets, etc.). This approach ensures that the same energy deposits are not used in the reconstruction of multiple analysis-level objects. The specific overlap removal tools employed

in this analysis are summarized in Table 5.6. The overlap between small-R and large-R jets will be addressed in the later stages of the selection process by adopting a “Merged over Resolved” regime categorization, as detailed in Sections 6.5 and 6.5.1.

Table 5.6: The sequential application of overlap removal conditions. ΔR is calculated using rapidity by default.

Reject	Against	Criteria
electron	electron	shared track, $p_{T,1} < p_{T,2}$
muon	electron	is calo-muon and shared ID track
electron	muon	shared ID track
jet	electron	$\Delta R < 0.2$
electron	jet	$\Delta R < 0.4$
jet	muon	NumTrack < 3 and (ghost-associated or $\Delta R < 0.2$)
muon	jet	$\Delta R < \min(0.4, 0.04 + 10 \text{ GeV}/p_{T,\mu})$
large-R-jet	electron	$\Delta R < 1.0$

Chapter 6

Event Selections

This chapter outlines the event selection process, optimized to enhance the expected significance of the signal relative to the background. A series of preselection criteria are employed, including basic quality cuts (event cleaning) and trigger requirements. The preselection cuts applied in the `CxAODMaker` are included.

In the first phase of event selection, the focus is on identifying the V boson that decays leptonically. For the 1-lepton channel, the candidate boson is selected with the target decay mode $W \rightarrow \ell\nu$. Next, the focus shifts to the hadronic components of the final state. This step aims to enhance the identification of objects (jets) associated with VBS and to select the other V boson that decays hadronically. For the selection of VBS-jet candidates, a standard jet-tagging procedure is implemented. The hadronically decaying V candidate is identified either as a large- R jet in the merged regime or as a pair of small- R jets in the resolved regime.

After the two bosons are identified, topological cuts are applied to further reduce the background. In the merged regime, the boson tagger defines the signal regions (SRs), and a reversed cut specifies the control regions (CRs) for the $V +$ jets background. In the resolved regime, a signal dijet system is constructed, and its invariant mass is utilized to define a mass window for SRs or sidebands for $V +$ jets CRs. Additionally, dedicated top control regions (TopCRs) are established by including events with additional b -jets.

The selections applied are summarized in Tables 6.1-6.2. Subsequent sections of this chapter will provide detailed discussions of each selection criterion, following the sequence in which they were applied.

Table 6.1: A summary of regions event selection for 1-lepton channel in the merged regime.

Selection	SR		W CR (WR)		$t\bar{t}$ CR (TR)	
	HP	LP	incl	HP	LP	
$W \rightarrow \ell\nu$	Num of Tight leptons			1		
	Num of Loose (!Tight) leptons			0		
	E_T^{miss}			> 80 GeV		
	$p_T(\ell)$			> 28 GeV		
VBS jets candidates	Leading Tag jet p_T			> 30 GeV		
	Subleading Tag jet p_T			> 30 GeV		
	m_{jj}			> 400 GeV		
$W/Z \rightarrow J$	Num of large- R jets			≥ 1		
	3-Var Tagger	pass50WP	pass80WP && !pass50WP	fail80WP	pass50WP	pass80WP && !pass50WP
Top veto	Num of b -tagged jets outside of large- R jet			0	≥ 1	

Table 6.2: A summary of regions event selection for 1-lepton channel in the resolved regime.

Selection	SR	W CR (WR)	$t\bar{t}$ CR (TR)
$W \rightarrow \ell\nu$	Number of Tight leptons	1	
	Number of Loose (!Tight) leptons	0	
	E_T^{miss}	> 80 GeV	
	$p_T(\ell)$	> 28 GeV	
VBS jets candidates	Leading Tag jet p_T	> 30 GeV	
	Subleading Tag jet p_T	> 30 GeV	
	m_{jj}	> 400 GeV	
$W/Z \rightarrow jj$	Number of small- R jets	≥ 4	
	Leading jet p_T	> 40 GeV	
	Subleading jet p_T	> 20 GeV	
	$Z \rightarrow q\bar{q}$ and $W \rightarrow qq'$	64 < m_{jj} < 106 GeV 50 < m_{jj} < 64 GeV or $m_{jj} > 106$ GeV	64 < m_{jj} < 106 GeV
Top veto	Number of additional b -tagged jets	0	≥ 1
VBS enhancing	m_{jj}	> 220 GeV	

6.1 Trigger Requirements

The trigger requirements are summarized in Table 6.3 and include the lowest unprescaled single-lepton or E_T^{miss} triggers. This trigger setup is employed in studies conducted for the VV semi-leptonic search [104] and in the SM $VH \rightarrow bb$ analysis [105]. Analyses from these studies feature semi-leptonic final states that closely resemble those in this analysis.

Differences between data and MC simulations in lepton trigger efficiency are addressed by applying scale factors. These factors are implemented using the following packages, which ensure accurate adjustments to the simulated events:

- `AsgElectronEfficiencyCorrectionTool`
- `MuonEfficiencyScaleFactors`

Table 6.3: The list of triggers used in the analysis.

Data-taking period	$e\nu qq$ and $eeqq$ channels	$\mu\nu qq$ ($p_T \mu\nu < 150 \text{ GeV}$) and $\mu\mu qq$ channels	$\mu\nu qq$ ($p_T \mu\nu > 150 \text{ GeV}$) and $\nu\nu qq$ channels
2015	HLT.e24_lhmedium_L1EM20 OR HLT.e60_lhmedium OR HLT.e120_lhloose	HLT.mu20_iloose_L1MU15 OR HLT.mu50	HLT.xe70
2016a (run < 302919) $(L < 1.0 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1})$	HLT.e26_llhtight_nod0_ivarloose OR HLT.e60_lhmedium_nod0 OR HLT.e140_lhloose_nod0 HLT.e300_etcut	HLT.mu26_ivarmedium OR HLT.mu50	HLT.xe90_mht_L1XE50
2016b (run ≥ 302919) $(L < 1.7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1})$	same as above	same as above	HLT.xe110_mht_L1XE50
2017	same as above	same as above	HLT.xe110_pufit_L1XE55
2018	same as above	same as above	HLT.xe110_pufit_xe70_L1XE50

6.2 Event Pre-selection

Table 6.4 details the selection cuts implemented within the CxAOD framework, which is specifically designed to optimize data management by significantly reducing the data size. All CP recommendations, provided by the Combined Performance (CP) groups for analysis use, are already applied at this stage. Detailed definitions and criteria for the following are provided in their respective sections:

- Definitions of Tight and Loose leptons — Sections 5.1 to 5.2.
- N_j , the number of small-R jets — Section 5.3.1.
- N_J^{LCTopo} , the number of large-R jets — Section 5.3.2, with $p_T > 200 \text{ GeV}$ and $|\eta| < 2.5$.

Table 6.4: List of the pre-selections applied in the CxAOD framework to reduce the data size.

Common	
Trigger	listed in Table 6.3
Event Cleaning	official GRL (GoodRunsLists)
Number of jets	$N_j \geq 2$ or $N_J^{\text{LCTopo}} \geq 1$
1-lepton channel	
Number of Tight leptons	= 1
Number of additional Loose leptons	= 0
$p_{T,\ell\nu}$	$> 75 \text{ GeV}$

6.3 Leptonically Decaying Boson Selection

The event selections for this analysis are largely based on the methodology used in a previous study with the 36 fb^{-1} dataset [106]. In the 1-lepton channel, events are required to contain a $\ell\nu$ pair indicative of a W boson decay, along with a $W/Z \rightarrow q\bar{q}$ candidate. The candidate for the $W \rightarrow \ell\nu$ decay is identified by the presence of exactly one Tight lepton and the absence of any Loose leptons in the final state.

The following anti-QCD cuts are applied to reject the QCD multijet and non-collision background:

- $E_{\text{T}}^{\text{miss}} > 80\text{ GeV}$
- $p_{\text{T},\ell} > 28\text{ GeV}$

Figure 6.1 shows the distributions of missing transverse energy ($E_{\text{T}}^{\text{miss}}$), lepton transverse momentum ($p_{\text{T},\ell}$), and large- R jet mass (m_J) in the 1-lepton channel at different stages of the analysis. The “preselection merged” and “preselection resolved” stages refer to the application of initial selection criteria in the merged and resolved regimes, respectively. These stages exclude the use of boson tagger cuts for the large- R jet and mass window cuts for signal jets, which are detailed further in Section 6.5.

In the distributions before any event selection, the observed data that are not adequately represented by the simulated events primarily originate from multijet and non-collision backgrounds. These discrepancies mainly arise due to the mis-measurement of jet energy. Although these backgrounds are relatively minor, it is crucial to acknowledge that MC simulations do not accurately estimate their shape or rate. The multijet background, often referred to as the QCD background, typically occurs when jets are misidentified as leptons or when leptons are produced later in the decay process of jets. Electron-related backgrounds are predominantly caused by jets generating extensive showers in the Electromagnetic Calorimeter (ECAL), which subsequently meet the criteria for electron trigger. For muons, the background typically arises from hadron decays within jets. Therefore, events from either QCD multijet or non-collision backgrounds typically contain “soft” (low p_T) leptons, which are effectively suppressed by the reasonable requirements on $E_{\text{T}}^{\text{miss}}$ and $p_{\text{T},\ell}$. After the anti-QCD cuts, we observe good agreement between the data and MC simulation in both merged and resolved regimes. In the m_J distribution, the W -peak in the $t\bar{t}$ process is also clearly visible.

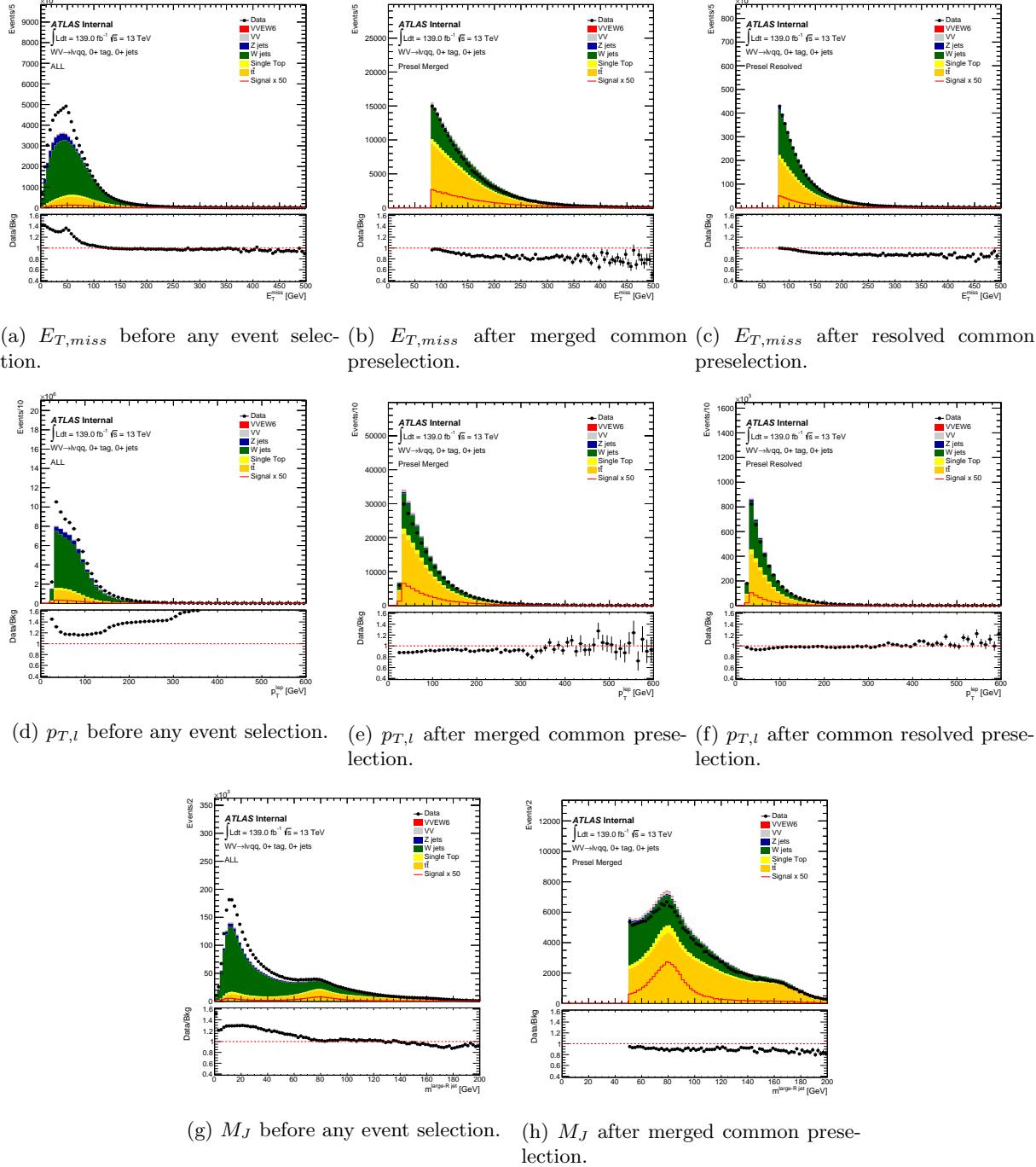


Figure 6.1: Distributions of $E_{\text{T}}^{\text{miss}}$, $p_{T,\ell}$, and m_J in the 1-lepton channel at various analysis stages.

6.4 VBS Selection

In the VBS selection, all events are required to have a pair of small- R jets, designated as Tag Jets, which serve as candidates for VBS jets. The hadronically-decaying W/Z candidate is reconstructed using either one large- R jet (J) or two small- R jets (j), as detailed in Sections 6.5.1 and 6.5.2. In the merged case, each event must include one large- R jet and two small- R jets, while in the resolved case, the requirement is for at least four small- R jets. Therefore, the resolved case involves a more complex selection strategy. Tag Jets are selected first, which are essential for defining the VBS jet candidates. Next, among the remaining jets, those emanating from the W/Z decay are identified as signal jets. This approach prioritizes establishing the forward topology (VBS jets) before identifying the decay products of the target process.

The Tag Jets algorithm employed in this analysis adheres to standard ATLAS practices. As discussed in Section 6.5.2, key requirements for Tag Jets include passing the fJVT selection, specifically using the Loose fJVT working point (WP). This Loose WP is selected for its simplicity and because it is the recommended default WP. Additionally, the corresponding scale factors (SFs) are readily available in our analysis framework.

Tag Jets candidates are required to be in opposite hemispheres, satisfying $\eta_{\text{tag } j_1} \cdot \eta_{\text{tag } j_2} < 0$. They should also have the highest dijet invariant mass among all pairs of small- R jets in the event that meet the fJVT requirements, as previously described. This criterion ensures the selection of the most physically significant jet pairs for analysis.

Once the Tag Jet pair is selected, the following criteria are applied to both jets to enhance the signal-to-background ratio:

- Each Tag Jet must have a transverse momentum (p_T) greater than 30 GeV.
- The invariant mass of the two Tag Jet system should exceed 400 GeV.

This approach is based on optimization studies conducted in the previous round of the analysis, which utilized the 36 fb^{-1} dataset [106], to inform our choice of both p_T and invariant mass cuts.

6.5 Signal Regions (SRs) Selection

Multiple Signal Regions (SRs) are defined to optimize the signal sensitivity and to accommodate the different reconstruction regimes of the hadronically decaying boson ($V \rightarrow qq$). The Merged regime is prioritized over the Resolved regime. Detailed descriptions of these two regimes are provided in this section.

6.5.1 Selection of $W/Z \rightarrow J$ candidates (merged category)

For boosted (high-energy) boson production, where the p_T of hadronically decaying W/Z bosons is at least 200 GeV, each boson is frequently reconstructed as a single large- R jet. In this “merged” category, the selection requires at least one large- R jet, with the leading large- R jet utilized for W/Z candidate reconstruction. To avoid double-counting jet energy, the large- R jet must be separated by a distance greater than $|\Delta R| = 1.4$ from both VBS Tag Jets. The final step in selecting boosted $W/Z \rightarrow qq$ candidates involves boson tagging based on three variables: jet mass (m_J), substructure variable D_2 , and ungroomed track multiplicity (n_{Tracks}). We adopt tagger working points (WPs) recommended for 50% and 80% signal efficiencies, applied at the jet level to optimize the selection.

To establish orthogonal regions, we define:

- High-Purity (HP) Region: Includes events (and their corresponding single large-R jet) that meet the 50% WP criteria.
- Low-Purity (LP) Region: Includes events that satisfy the 80% WP but do not fulfill the 50% WP criteria.

Thus, events in the merged regime conforming to the 50% WP criteria are categorized into the HP Signal Region (SR), while those meeting the 80% WP standards but not the 50% are allocated to the LP SR. Table 6.1 contains the complete definitions of the HP and LP SRs.

For the HP SR, the 50% WP W/Z -tagging scale factor is applied. For the LP SR, a custom scale factor is defined:

$$SF_{LP} = \frac{\epsilon_{loose} SF_{eff,loose} - \epsilon_{tight} SF_{eff,tight}}{\epsilon_{loose} - \epsilon_{tight}} \quad (6.1)$$

Here, ϵ represents the efficiency estimated in $t\bar{t}$ events for signal and $\gamma+jets$ and multijet events for background. “Loose” corresponds to the 80% WP, while “tight” refers to the 50% WP. Detailed discussions on W/Z -tagging scale factors are available in Section 9.1.2.

In the baseline boson tagger, exclusive selections for Z and W candidates are used. However, due to significant overlap in these selections, it is not feasible to define two orthogonal regions for the hadronic decays of W and Z . Therefore, an inclusive $V \rightarrow qq$ selection is utilized, which is a logical OR of the W and Z boson tagger selections. This inclusive selection adopts the lower mass cut from the W selection and the upper cut from the Z selection.

6.5.2 Selection of $W/Z \rightarrow jj$ candidates (resolved category)

In the lower p_T range for hadronically decaying W/Z bosons, two distinct jets are typically resolvable. This resolved regime offers the highest efficiency for EW $VV + jj$ signals, although its sensitivity is marginally lower compared to the merged regime due to increased background acceptance. Within this regime, events are selected based on the presence of at least two “signal” jets. These signal jets are identified from a pool of candidates, excluding the two VBS Tag Jets.

In the search of W/Z candidates, we select the two signal jets with the highest p_T . This approach slightly reduces signal efficiency within the mass window compared to previous analyses. However, it allows for a more relaxed application of the Close- V selection algorithm, which pairs jets with invariant masses closest to the nominal $V(W/Z)$ boson masses.

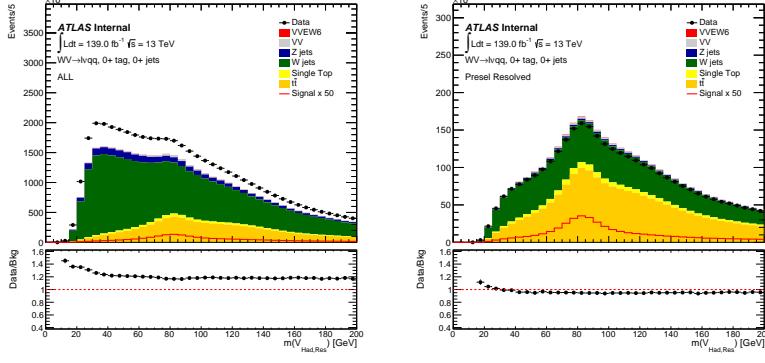
After selecting the two jets of interest, the leading jet is required to have $p_T > 40\text{ GeV}$, a criterion set to further reduce background and enhance sensitivity. To identify events indicative of a hadronically decaying W/Z boson, the dijet mass (m_{jj}) is constrained within a mass window of $64 < m_{jj} < 106\text{ GeV}$.

Figure 6.2 shows the full range distributions for m_{jj} without the mass window constraints, with the W peak coming from top-associated processes clearly visible.

As mentioned in Section 4.2, a VBS-enhancing cut, $m_{jjj} > 220\text{ GeV}$, is introduced later in the analysis to reduce the contribution from the non-VBS tZb process, as illustrated in Fig. 4.1. This cut, implemented near the reconstructed top mass, is shown in Figure 6.4 in the next section. Events passing this cut are classified into “tight” regions, while those that do not are classified into “loose” regions. The resolved “tight” regions are used in calculations and statistical studies, whereas the resolved “loose” regions serve as a sanity check.

6.5.3 VV system invariant mass

The mass of the WV system, m_{WV} , is calculated using the lepton, neutrino, and the hadronically-decaying boson (represented by either a large- R jet or two small- R jets). To determine the neutrino’s momentum in the z -direction (p_z), we apply the Particle Data Group (PDG) value for the W boson mass to the lepton-



(a) Before any event selection. (b) After resolved preselection.

Figure 6.2: Reconstructed mass distribution of the leading two jets.

neutrino system. This results in a quadratic equation. The solution for p_z is chosen based on the following criteria: if the solutions are complex, the real component is taken; if they are real, the solution with the smaller absolute value is selected.

6.6 Control Regions (CRs) Definitions

In this analysis, we adopt the conventional approach for background estimation, which combines MC simulations with data-driven methods. MC simulations are used to model SM background processes, while dedicated Control Regions (CRs) are established to adjust the background normalization within the SRs. This section will detail the definition of CRs tailored to $W+\text{jets}$ and $t\bar{t}$ events. Furthermore, we distinguish between CRs for the resolved and merged categories to address potential discrepancies in background modeling across different p_T ranges.

6.6.1 $W+\text{jets}$ Control Region (WCR)

The $W+\text{jets}$ Control Region (WCR) is defined using mass sidebands relative to the hadronically decaying vector boson. Notably, in the resolved category, distinct sidebands outside the mass windows of both the W and Z bosons are identifiable. In contrast, in the merged category, the definition becomes more conceptual due to the event selection methodology. The criteria for event selection in the WCR are aligned with those in the SR to ensure analytical consistency:

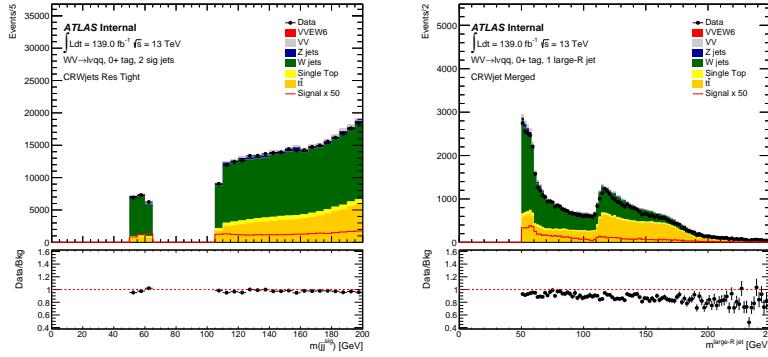
- In the resolved category, events must satisfy $50 < m_{jj} < 64 \text{ GeV}$ or $m_{jj} > 106 \text{ GeV}$, effectively inverting the m_{jj} requirement for the SR.
- In the merged category, the criterion is failing the high efficiency WP ($\epsilon = 80\%$) of the vector boson tagger.

Figure 6.3a shows the sideband distributions for m_{jj} in the resolved WCR. Figure 6.3b shows the large-R jet mass distributions in the merged control region. Given that the merged CR is defined by the vector boson tagger's 80% working point, the mass distribution encompasses events within both the sidebands and the mass window.

In the resolved category, the “Loose” and “Tight” regions are defined similarly to those in the resolved SR. The m_{jjj} distributions are shown in Figure 6.4, where the peak at about 180 GeV, corresponding to the top quark mass, is observed in the “Loose” region. This indicates that the cut at 220 GeV was effective in excluding this peak from the “Tight” region.

6.6.2 Top Control Region (TCR)

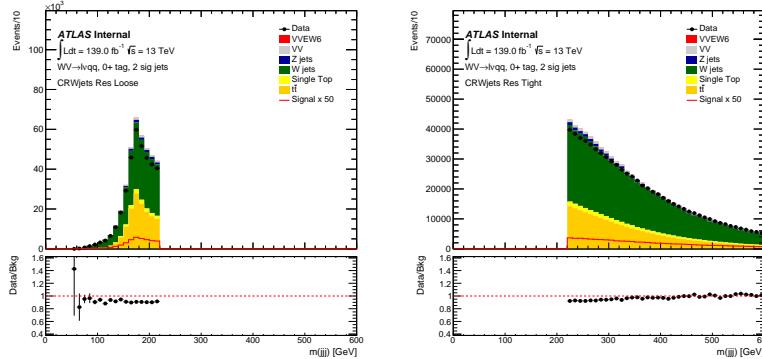
The TCR is established by requiring the presence of at least one b -tagged jet, contrary to a b -veto in other regions. Figure 6.5 displays the b -jets multiplicity distributions for both the resolved and merged regimes.



(a) m_{jj} plot in the sidebands.
Resolved WCR.

(b) m_J plot in the “sidebands.”
Merged WCR.

Figure 6.3: $m_{VH_{ad}}$ (m_{jj} or m_J) plots in the resolved and merged WCRs.

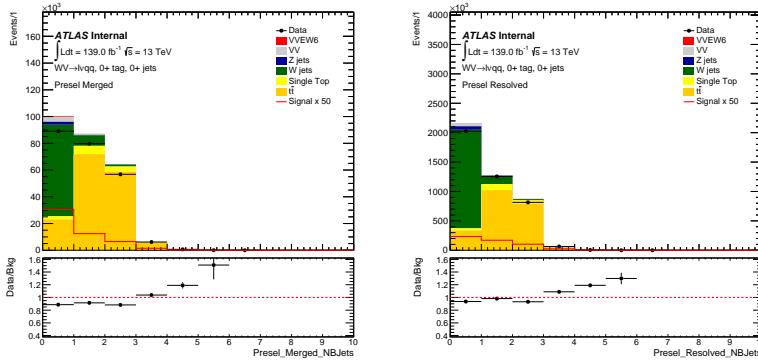


(a) Resolved “Loose” WCR.

(b) Resolved “Tight” WCR.

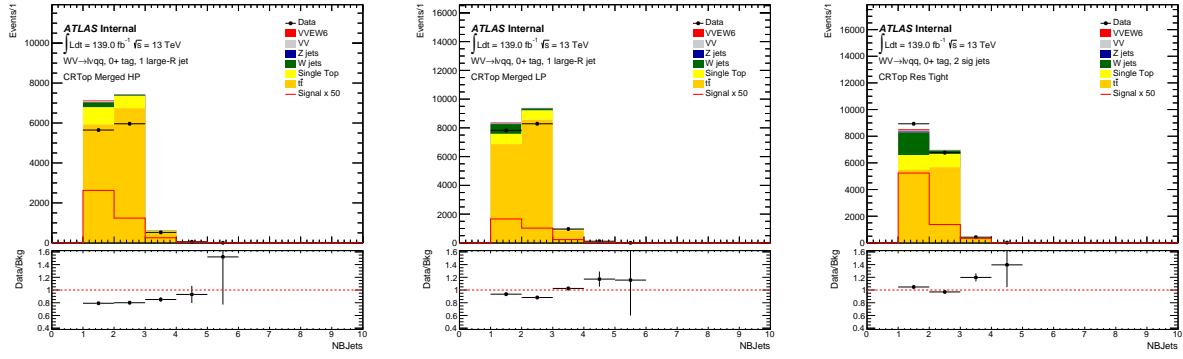
Figure 6.4: m_{jjj} plots in the “Loose” and “Tight” WCRs.

The defining feature of the TCR, the inclusion of additional b -jets, creates orthogonal cuts that align with the mass window phase spaces. In the merged regime, the TCR adopts the High Purity (HP) and Low Purity (LP) categorization analogous to the SRs.



(a) After merged preselection

(b) After resolved preselection



(c) HP TCR

(d) LP TCR

(e) Resolved TCR

Figure 6.5: *B*-jets multiplicity across various stages of event selection.

6.7 Signal and Background Cut-flows

Cutflow tables for various signal and control regions are presented in Figure 6.6.

$\sqrt{s} = 13\text{TeV}$, $\mathcal{L} = 139.0\text{fb}^{-1}$	Wjets	Zjets	Diboson	tbar	singletop	EW6Signal	Data	Data/MC	Total BG MC
All	55443762.55 ± 26272.85	4886817.85 ± 6264.10	803232.01 ± 438.14	13260543.35 ± 1383.63	2078599.51 ± 456.40	65029.87 ± 37.86	96003695.00 ± 9802.74	1.26	76537985.13 ± 27052.13
pass Preselection	46048672.73 ± 23530.55	4150236.33 ± 5749.44	677542.96 ± 392.57	10808712.01 ± 1245.69	1698957.94 ± 413.25	52525.00 ± 33.73	77082259.00 ± 8779.65	1.22	63436646.97 ± 24261.51
TagJet30Merged	16187888.27 ± 1194.73	1428334.96 ± 2810.34	27962.81 ± 238.29	6122801.37 ± 937.99	849321.98 ± 289.34	33248.92 ± 26.11	28311498.00 ± 5320.86	1.14	24894558.30 ± 11586.24
OneLargeJet	418590.80 ± 373.92	32411.03 ± 67.22	17354.54 ± 51.96	379099.80 ± 233.44	39686.42 ± 66.82	2490.74 ± 5.92	88854.00 ± 942.79	1.00	889633.34 ± 453.90
MET 80	204364.62 ± 260.85	6936.63 ± 31.67	8784.94 ± 37.69	221498.83 ± 179.60	24013.19 ± 52.79	1443.64 ± 4.57	442322.00 ± 665.07	0.95	467041.86 ± 324.85
Mjj400Merged	172408.73 ± 233.09	5916.31 ± 25.17	7466.54 ± 34.67	182188.18 ± 163.10	20394.61 ± 48.55	1304.23 ± 4.24	366558.00 ± 605.28	0.94	389678.59 ± 291.80
LargeRJWP50	4703.71 ± 38.61	161.72 ± 4.04	7027.78 ± 10.83	21394.66 ± 55.97	2287.58 ± 17.46	349.32 ± 2.11	24373.00 ± 156.12	0.82	29599.77 ± 71.18
MerTagBJetVeto	4394.02 ± 38.05	147.08 ± 3.97	659.06 ± 10.54	8102.81 ± 34.49	768.06 ± 10.03	266.13 ± 1.59	12178.00 ± 110.35	0.85	14337.15 ± 53.55

(a) SR High Purity Merged

$\sqrt{s} = 13\text{TeV}$, $\mathcal{L} = 139.0\text{fb}^{-1}$	Wjets	Zjets	Diboson	tbar	singletop	EW6Signal	Data	Data/MC	Total BG MC
All	55443762.55 ± 26272.85	4886817.85 ± 6264.10	803232.01 ± 438.14	13260543.35 ± 1383.63	2078599.51 ± 456.40	65029.87 ± 37.86	96003695.00 ± 9802.74	1.26	76537985.13 ± 27052.13
pass Preselection	46048672.73 ± 23530.55	4150236.33 ± 5749.44	677542.96 ± 392.57	10808712.01 ± 1245.69	1698957.94 ± 413.25	52525.00 ± 33.73	77082259.00 ± 8779.65	1.22	63436646.97 ± 24261.51
TagJet30Merged	16187888.27 ± 1194.73	1428334.96 ± 2810.34	27962.81 ± 238.29	6122801.37 ± 937.99	849321.98 ± 289.34	33248.92 ± 26.11	28311498.00 ± 5320.86	1.14	24894558.30 ± 11586.24
OneLargeJet	418590.80 ± 373.92	32411.03 ± 67.22	17354.54 ± 51.96	379099.80 ± 233.44	39686.42 ± 66.82	2490.74 ± 5.92	88854.00 ± 942.79	1.00	889633.34 ± 453.90
MET 80	204364.62 ± 260.85	6936.63 ± 31.67	8784.94 ± 37.69	221498.83 ± 179.60	24013.19 ± 52.79	1443.64 ± 4.57	442322.00 ± 665.07	0.95	467041.86 ± 324.85
Mjj400Merged	172408.73 ± 233.09	5916.31 ± 25.17	7466.54 ± 34.67	182188.18 ± 163.10	20394.61 ± 48.55	1304.23 ± 4.24	366558.00 ± 605.28	0.94	389678.59 ± 291.80
LargeRJWP50	18366.96 ± 81.02	612.47 ± 7.73	1536.17 ± 15.85	48318.18 ± 84.04	4636.14 ± 24.41	586.82 ± 2.77	63726.00 ± 252.44	0.86	74056.74 ± 120.59
MerTagBJetVeto	17217.12 ± 80.02	558.73 ± 7.61	1431.72 ± 15.40	18626.71 ± 52.25	1710.09 ± 14.67	444.47 ± 2.08	34336.00 ± 185.30	0.86	39988.84 ± 98.22

(b) SR Low Purity Merged

$\sqrt{s} = 13\text{TeV}$, $\mathcal{L} = 139.0\text{fb}^{-1}$	Wjets	Zjets	Diboson	tbar	singletop	EW6Signal	Data	Data/MC	Total BG MC
All	55443762.55 ± 26272.85	4886817.85 ± 6264.10	803232.01 ± 438.14	13260543.35 ± 1383.63	2078599.51 ± 456.40	65029.87 ± 37.86	96003695.00 ± 9802.74	1.26	76537985.13 ± 27052.13
pass Preselection	46048672.73 ± 23530.55	4150236.33 ± 5749.44	677542.96 ± 392.57	10808712.01 ± 1245.69	1698957.94 ± 413.25	52525.00 ± 33.73	77082259.00 ± 8779.65	1.22	63436646.97 ± 24261.51
TagJet30Merged	16187888.27 ± 1194.73	1428334.96 ± 2810.34	27962.81 ± 238.29	6122801.37 ± 937.99	849321.98 ± 289.34	33248.92 ± 26.11	28311498.00 ± 5320.86	1.14	24894558.30 ± 11586.24
OneLargeJet	418590.80 ± 373.92	32411.03 ± 67.22	17354.54 ± 51.96	379099.80 ± 233.44	39686.42 ± 66.82	2490.74 ± 5.92	88854.00 ± 942.79	1.00	889633.34 ± 453.90
MET 80	204364.62 ± 260.85	6936.63 ± 31.67	8784.94 ± 37.69	221498.83 ± 179.60	24013.19 ± 52.79	1443.64 ± 4.57	442322.00 ± 665.07	0.95	467041.86 ± 324.85
Mjj400Merged	172408.73 ± 233.09	5916.31 ± 25.17	7466.54 ± 34.67	182188.18 ± 163.10	20394.61 ± 48.55	1304.23 ± 4.24	366558.00 ± 605.28	0.94	389678.59 ± 291.80
LargeRJWP50	4703.71 ± 38.61	161.72 ± 4.04	7027.78 ± 10.83	21394.66 ± 55.97	2287.58 ± 17.46	349.32 ± 2.11	24373.00 ± 156.12	0.82	29599.77 ± 71.18
MerTagBJetVeto	4394.02 ± 38.05	147.08 ± 3.97	659.06 ± 10.54	8102.81 ± 34.49	768.06 ± 10.03	266.13 ± 1.59	12178.00 ± 110.35	0.85	14337.15 ± 53.55

(c) SR Resolved

$\sqrt{s} = 13\text{TeV}$, $\mathcal{L} = 139.0\text{fb}^{-1}$	Wjets	Zjets	Diboson	tbar	singletop	EW6Signal	Data	Data/MC	Total BG MC
All	55443762.55 ± 26272.85	4886817.85 ± 6264.10	803232.01 ± 438.14	13260543.35 ± 1383.63	2078599.51 ± 456.40	65029.87 ± 37.86	96003695.00 ± 9802.74	1.26	76537985.13 ± 27052.13
pass Preselection	46048672.73 ± 23530.55	4150236.33 ± 5749.44	677542.96 ± 392.57	10808712.01 ± 1245.69	1698957.94 ± 413.25	52525.00 ± 33.73	77082259.00 ± 8779.65	1.22	63436646.97 ± 24261.51
TagJet30Merged	16187888.27 ± 1194.73	1428334.96 ± 2810.34	27962.81 ± 238.29	6122801.37 ± 937.99	849321.98 ± 289.34	33248.92 ± 26.11	28311498.00 ± 5320.86	1.14	24894558.30 ± 11586.24
TwoSigJets	11847347.41 ± 10326.72	345534.86 ± 1323.02	203602.87 ± 215.01	4607952.73 ± 813.99	619119.48 ± 253.52	20683.88 ± 20.65	18171059.00 ± 4262.75	1.03	17642421.24 ± 10448.21
OneLargeJet	418590.80 ± 373.92	32411.03 ± 67.22	17354.54 ± 51.96	379099.80 ± 233.44	39686.42 ± 66.82	2490.74 ± 5.92	88854.00 ± 942.79	1.00	889633.34 ± 453.90
MET 80	204364.62 ± 260.85	6936.63 ± 31.67	8784.94 ± 37.69	221498.83 ± 179.60	24013.19 ± 52.79	1443.64 ± 4.57	442322.00 ± 665.07	0.95	467041.86 ± 324.85
Mjj400Merged	172408.73 ± 233.09	5916.31 ± 25.17	7466.54 ± 34.67	182188.18 ± 163.10	20394.61 ± 48.55	1304.23 ± 4.24	366558.00 ± 605.28	0.94	389678.59 ± 291.80
LargeRJWP50	4703.71 ± 38.61	161.72 ± 4.04	7027.78 ± 10.83	21394.66 ± 55.97	2287.58 ± 17.46	349.32 ± 2.11	24373.00 ± 156.12	0.82	29599.77 ± 71.18
OneBJet	309.69 ± 6.56	14.64 ± 0.71	43.72 ± 2.49	13291.85 ± 44.08	1519.53 ± 14.29	83.19 ± 1.39	12195.00 ± 110.43	0.80	15262.62 ± 46.90

(d) TopCR High Purity Merged

$\sqrt{s} = 13\text{TeV}$, $\mathcal{L} = 139.0\text{fb}^{-1}$	Wjets	Zjets	Diboson	tbar	singletop	EW6Signal	Data	Data/MC	Total BG MC
All	55443762.55 ± 26272.85	4886817.85 ± 6264.10	803232.01 ± 438.14	13260543.35 ± 1383.63	2078599.51 ± 456.40	65029.87 ± 37.86	96003695.00 ± 9802.74	1.26	76537985.13 ± 27052.13
pass Preselection	46048672.73 ± 23530.55	4150236.33 ± 5749.44	677542.96 ± 392.57	10808712.01 ± 1245.69	1698957.94 ± 413.25	52525.00 ± 33.73	77082259.00 ± 8779.65	1.22	63436646.97 ± 24261.51
TagJet30Merged	16187888.27 ± 1194.73	1428334.96 ± 2810.34	27962.81 ± 238.29	6122801.37 ± 937.99	849321.98 ± 289.34	33248.92 ± 26.11	28311498.00 ± 5320.86	1.14	24894558.30 ± 11586.24
OneLargeJet	418590.80 ± 373.92	32411.03 ± 67.22	17354.54 ± 51.96	379099.80 ± 233.44	39686.42 ± 66.82	2490.74 ± 5.92	88854.00 ± 942.79	1.00	889633.34 ± 453.90
MET 80	204364.62 ± 260.85	6936.63 ± 31.67	8784.94 ± 37.69	221498.83 ± 179.60	24013.19 ± 52.79	1443.64 ± 4.57	442322.00 ± 665.07	0.95	467041.86 ± 324.85
Mjj400Merged	172408.73 ± 233.09	5916.31 ± 25.17	7466.54 ± 34.67	182188.18 ± 163.10	20394.61 ± 48.55	1304.23 ± 4.24	366558.00 ± 605.28	0.94	389678.59 ± 291.80
LargeRJWP50	4703.71 ± 38.61	161.72 ± 4.04	7027.78 ± 10.83	21394.66 ± 55.97	2287.58 ± 17.46	349.32 ± 2.11	24373.00 ± 156.12	0.82	29599.77 ± 71.18
OneBJet	2112.66 ± 22.71	98.53 ± 3.62	94.77 ± 3.74	14153.94 ± 45.85	2725.19 ± 17.54	183.47 ± 1.96	19177.00 ± 138.48	0.99	19368.55 ± 54.38

(e) TopCR Low Purity Merged

$\sqrt{s} = 13\text{TeV}$, $\mathcal{L} = 139.0\text{fb}^{-1}$	Wjets	Zjets	Diboson	tbar	singletop	EW6Signal	Data	Data/MC	Total BG MC
All	55443762.55 ± 26272.85	4886817.85 ± 6264.10	803232.01 ± 438.14	13260543.35 ± 1383.63	2078599.51 ± 456.40	65029.87 ± 37.86	96003695.00 ± 9802.74	1.26	76537985.13 ± 27052.13
pass Preselection	46048672.73 ± 23530.55	4150236.33 ± 5749.44	677542.96 ± 392.57	10808712.01 ± 1245.69	1698957.94 ± 413.25	52525.00 ± 33.73	77082259.00 ± 8779.65	1.22	63436646.97 ± 24261.51
TagJet30Merged	16187888.27 ± 1194.73	1428334.96 ± 2810.34	27962.81 ± 238.29	6122801.37 ± 937.99	849321.98 ± 28				

Chapter 7

Background Modeling and Estimations

The modeling of major background processes, such as $W + \text{jets}$ and top quark pair ($t\bar{t}$) productions, is examined in this section. Since $Z + \text{jets}$ production has no meaningful contribution to the $V + \text{jets}$ backgrounds in the 1-lepton channel, we focus solely on $W + \text{jets}$ for simplicity. Initially, attention is directed towards the study of mismodeling in the invariant mass of the VBS Tag Jets, denoted as m_{jj}^{tag} , specifically within $W + \text{jets}$ events in the 1-lepton channel. Subsequently, various control regions for $W + \text{jets}$ and $t\bar{t}$ backgrounds are established. These regions are instrumental in validating the modeling and constraining the normalization based on data.

7.1 Reweighting for the Tag Jets

This section examines the modeling of the $W + \text{jets}$ background for the 1-lepton channel, focusing on the mismodeled invariant mass of the two selected Tag Jets (m_{jj}^{tag}) in the Sherpa 2.2.1 $V + \text{jets}$ samples. The forward topology of our final state makes m_{jj}^{tag} a critical variable. To address the observed mismodeling, we applied a linear reweighting procedure to the $W + \text{jets}$ MC samples, ensuring alignment with data in the CRs. Figure 7.1 illustrates the m_{jj}^{tag} distributions for both merged and resolved regions, highlighting the noticeable discrepancies between the primary background source and the observed data.

To correct the mismodeling, which we assume varies linearly with m_{jj}^{tag} , we apply the following reweighting formula to the $W + \text{jets}$ events:

$$w(m_{jj}^{\text{tag}}) = p_0 + p_1 \cdot m_{jj}^{\text{tag}}, \quad (7.1)$$

where p_0 and p_1 are coefficients calculated independently for both merged and resolved regions. The function $w(m_{jj}^{\text{tag}})$ represents the weight adjustment applied to each event, modifying the original event weight based on the m_{jj}^{tag} value. After determining these coefficients from CRs (see Table 7.1), we apply the corrections to the respective SRs. For both merged and resolved scenarios, p_0 and p_1 are obtained by fitting the Sherpa $W + \text{jets}$ model-to-data ratio, after subtracting non- $W + \text{jets}$ backgrounds.

In the merged case, a single fit is applied to the entire m_{jj}^{tag} distribution without subdividing the sam-

ple based on the m_{jj} range. This approach not only simplifies the analysis but also maintains statistical robustness, as illustrated in Figure 7.3.

In the resolved case, the fit to m_{jj}^{tag} is performed within specific m_{jj}^{sig} bins ([50, 60, 70, 100, 120, 150, 200, 300] GeV), excluding the [70, 100] GeV bin as it corresponds to the resolved signal region and we aim to fit using only control region data. This approach segments both MC and data samples according to m_{jj}^{sig} , as illustrated in Figure 7.4. Subsequently, the parameters p_0 and p_1 are determined across these bins. These parameters are then fitted to define the definitive reweighting factors for the signal region's m_{jj}^{sig} , as shown in Figure 7.5. An interpolation from the control to the signal region follows this step.

The study encompasses all data periods and corresponding MC campaigns (mc16a, mc16d, and mc16e). To verify consistency across various data periods and pileup conditions, these analyses were independently repeated for each MC campaign. The outcomes are presented in Figure 7.6. Although a slight dependence on the MC period was observed, the benefits of adjusting the m_{jj}^{tag} reweighting for each period proved to be minimal. Thus, the combined period results were utilized for the final analysis.

Table 7.2 shows the final parameters derived for the 1-lepton channel. The m_{jj}^{tag} distributions for the three SRs before and after reweighting are depicted in Figures 7.1 and 7.2, respectively.

Table 7.1: Definition of control regions used to derive $W + \text{jets}$ reweighting factors.

Resolved	pass Resolved Selection (NoMassWindowCut) BjetVeto
Merged	pass Merged Selection (NoMassWindowCut) BjetVeto

Table 7.2: 1-lepton $m(jj)^{\text{tag}}$ reweighting factors.

Parameter	Merged CRVjet	Resolved CRVjet
p_0 (slope) [GeV $^{-1}$]	$(-51.4 \pm 2.9)10^{-5}$	$(-14.4 \pm 6.1)10^{-5}$
p_1 (constant)	1.47 ± 0.03	1.13 ± 0.02

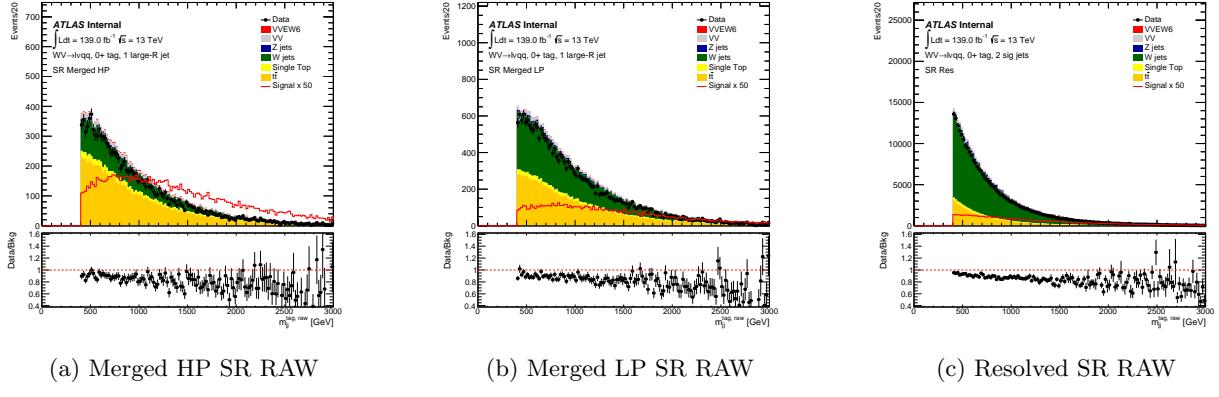


Figure 7.1: m_{jj}^{tag} distributions without any corrections in the signal regions.

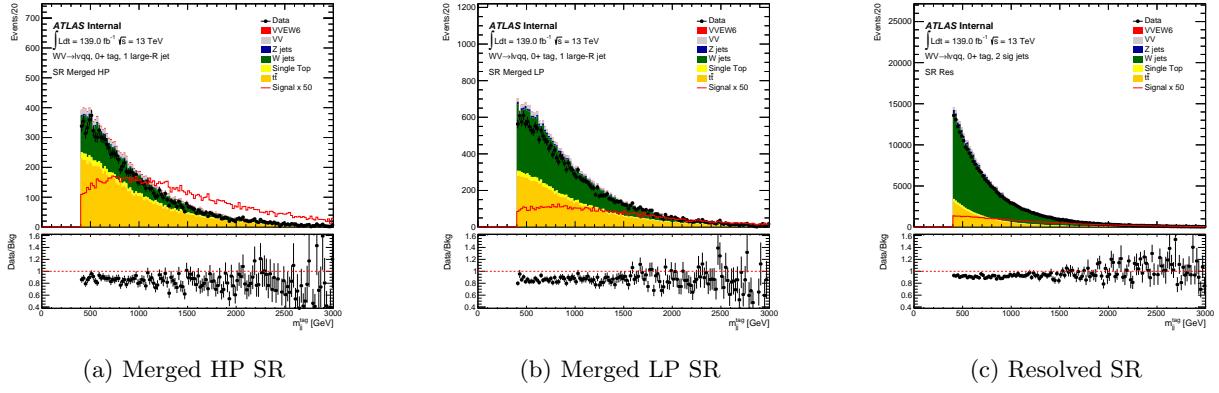


Figure 7.2: m_{jj}^{tag} distributions after applying reweighting in the signal regions.

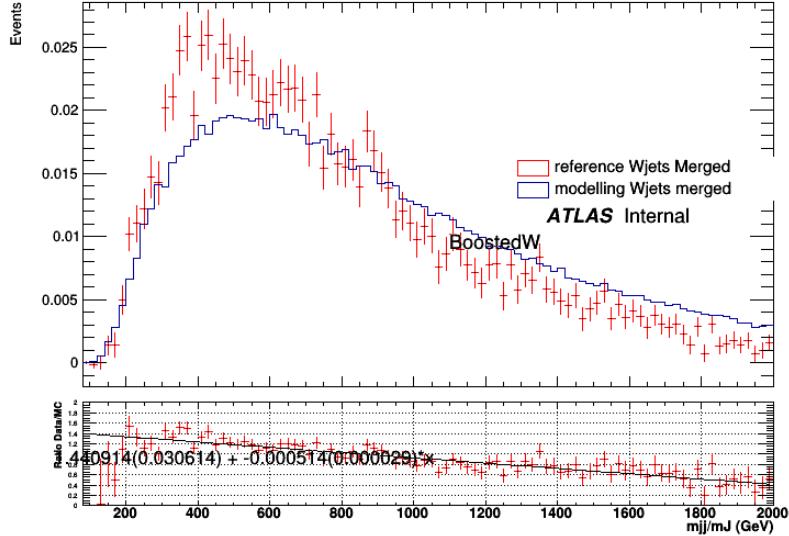


Figure 7.3: Fit of m_{jj}^{tag} slope across the full m_J range in the merged region.

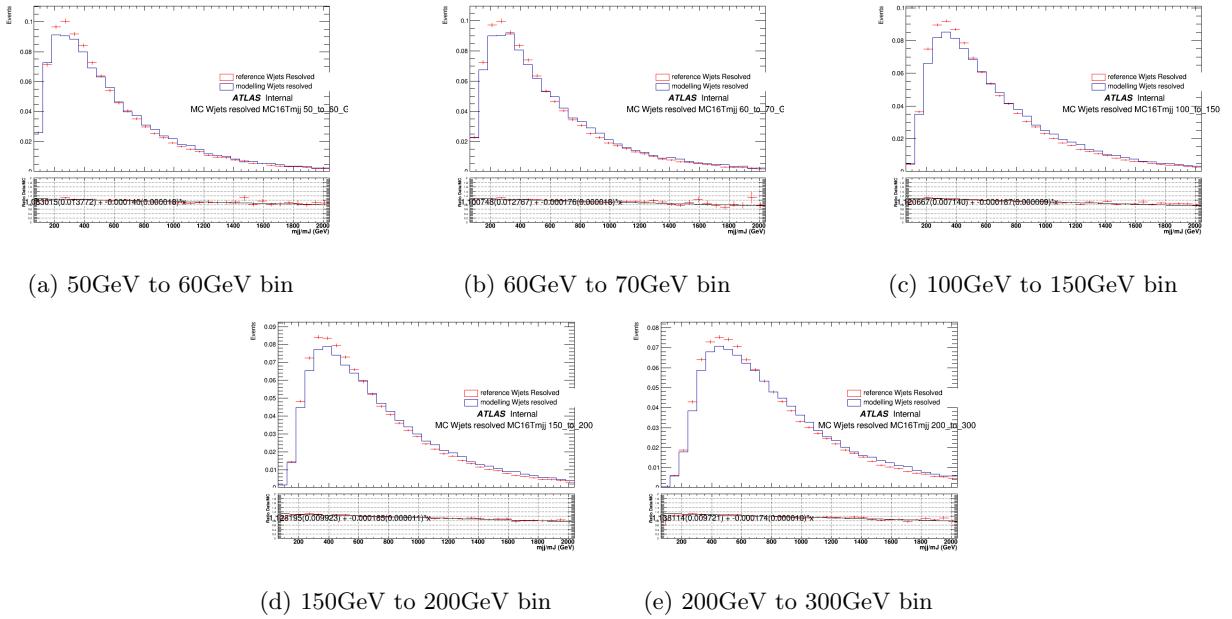
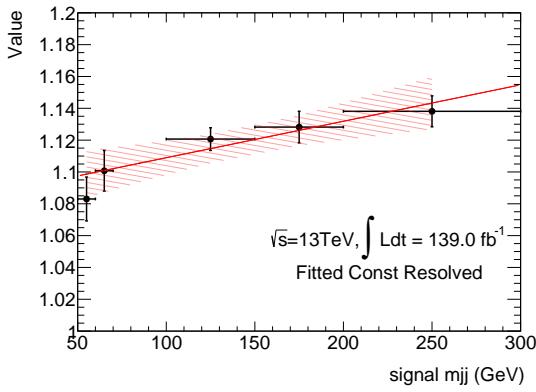
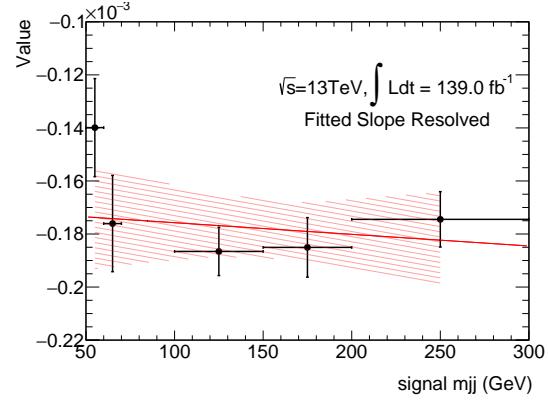


Figure 7.4: Fit of m_{jj}^{tag} slope across various m_{jj}^{sig} slices in $W + \text{jets}$ resolved control region.

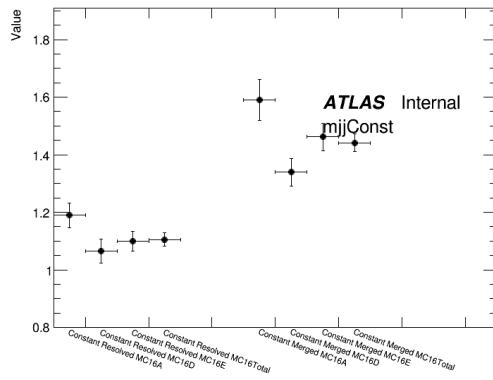


(a) Fit results for “Constant”, p_0 .

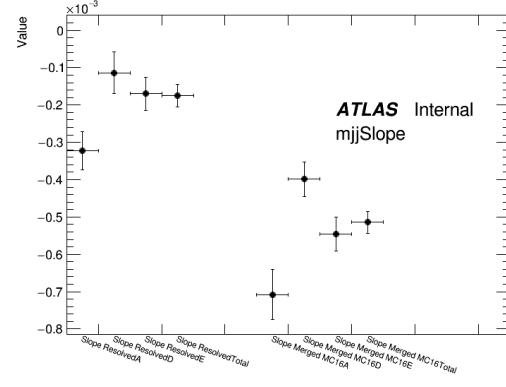


(b) Fit results for “Slope”, p_1 .

Figure 7.5: Slopes and constants as a function of m_{jj}^{sig} , with the red band indicating the 68% confidence level of the fit.



(a) Fitted results for “Constant”, p_0 .



(b) Fitted results for “Slope”, p_1 .

Figure 7.6: Fitted results for slopes and constants for each data period.

7.2 $W + \text{jets}$ Control Region

Figures 7.7, 7.8, 7.9, 7.10, 7.11, and 7.12 show the main relevant kinematic variable distributions in the $W + \text{jets}$ CRs for the 1-lepton channel, after m_{jj}^{tag} reweighting has been applied. The definitions of these CRs are provided in Section 6.6.1. It is worth noting that the resolved loose $W + \text{jets}$ control region is utilized solely to study and validate the agreement between the data and MC samples and is not included in the inputs for the statistical analysis.

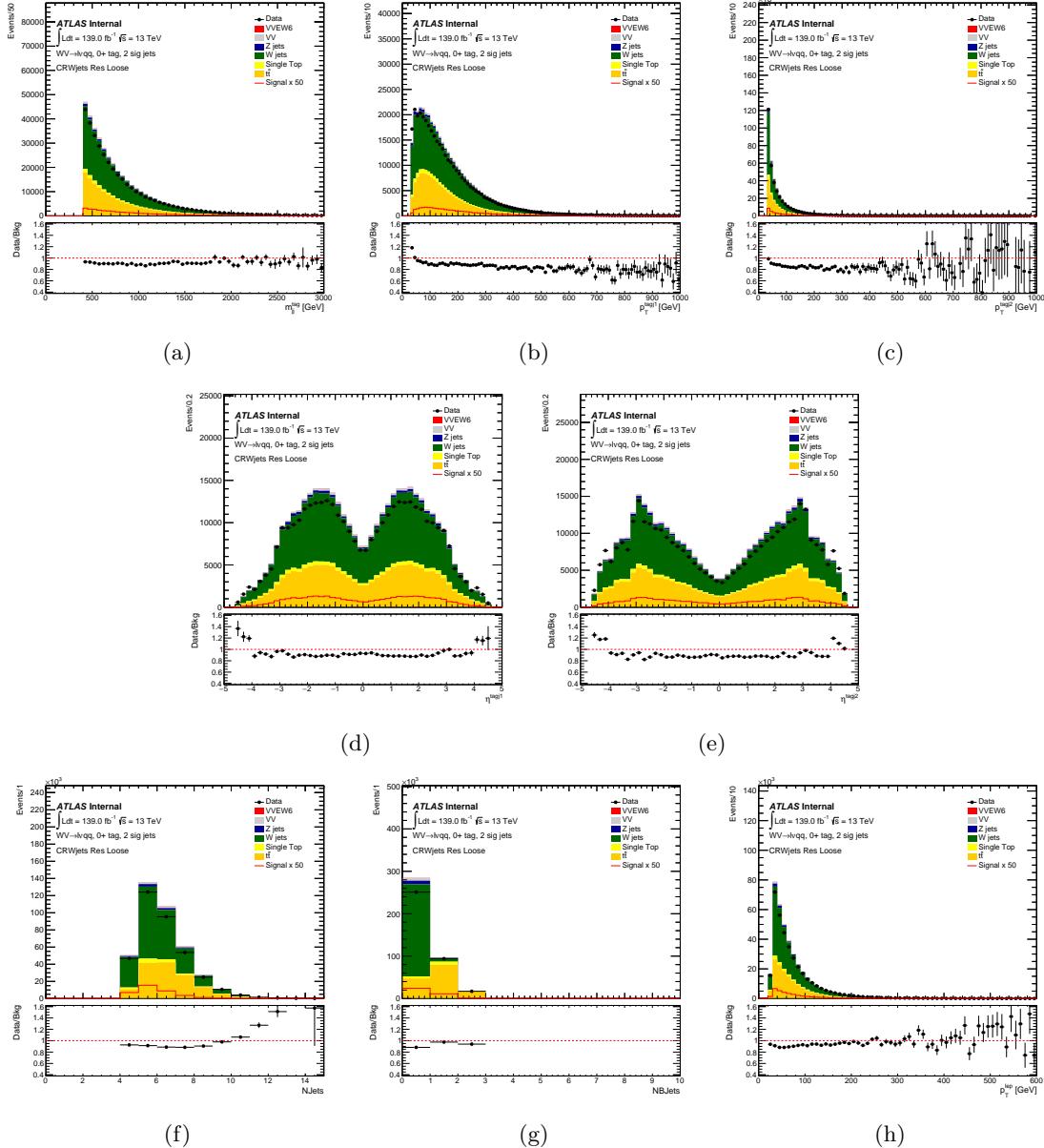


Figure 7.7: Data-MC checks for the resolved loose $W + \text{jets}$ control region in the 1-lepton channel.

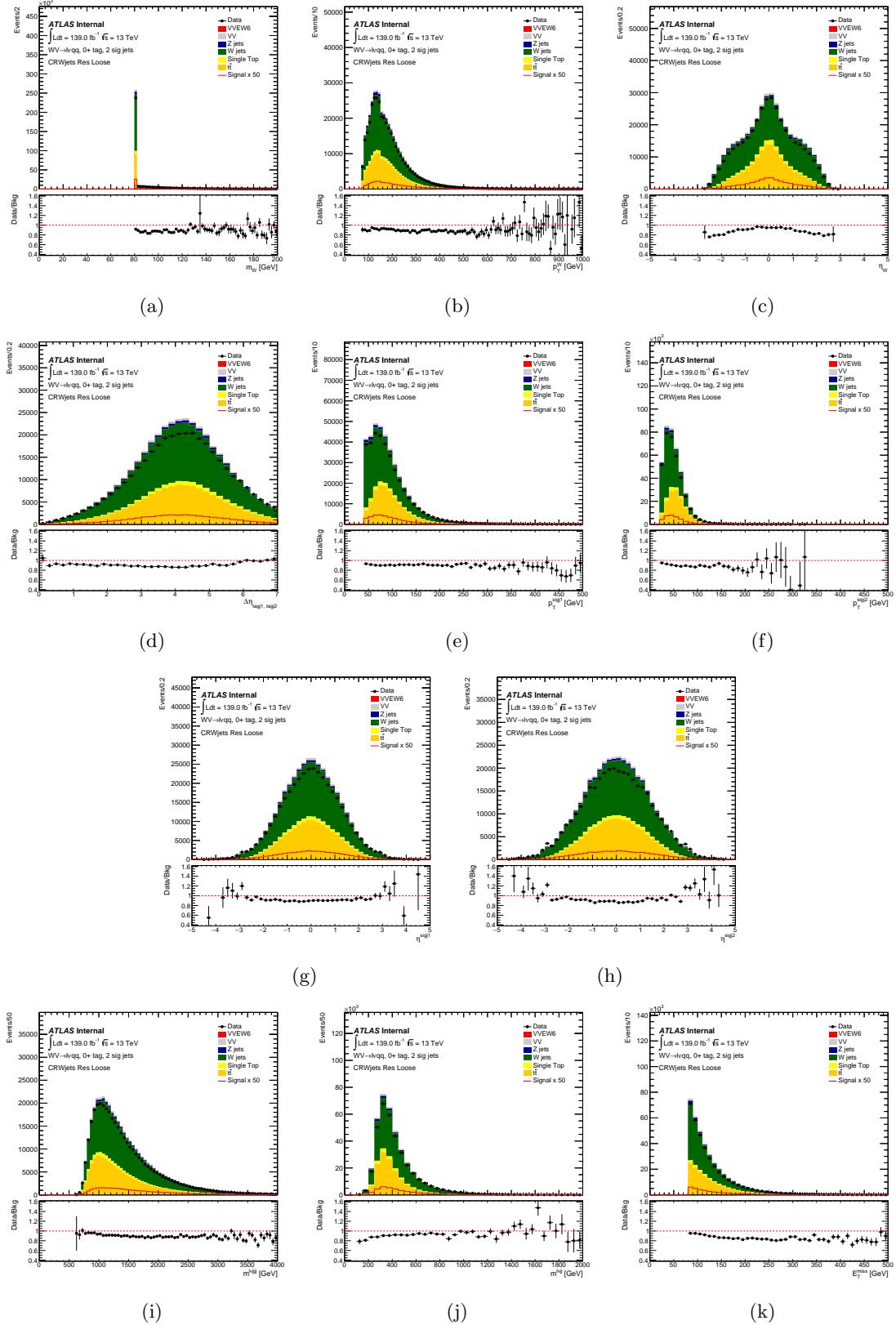


Figure 7.8: Data-MC checks for the resolved loose $W + \text{jets}$ control region in the 1-lepton channel.

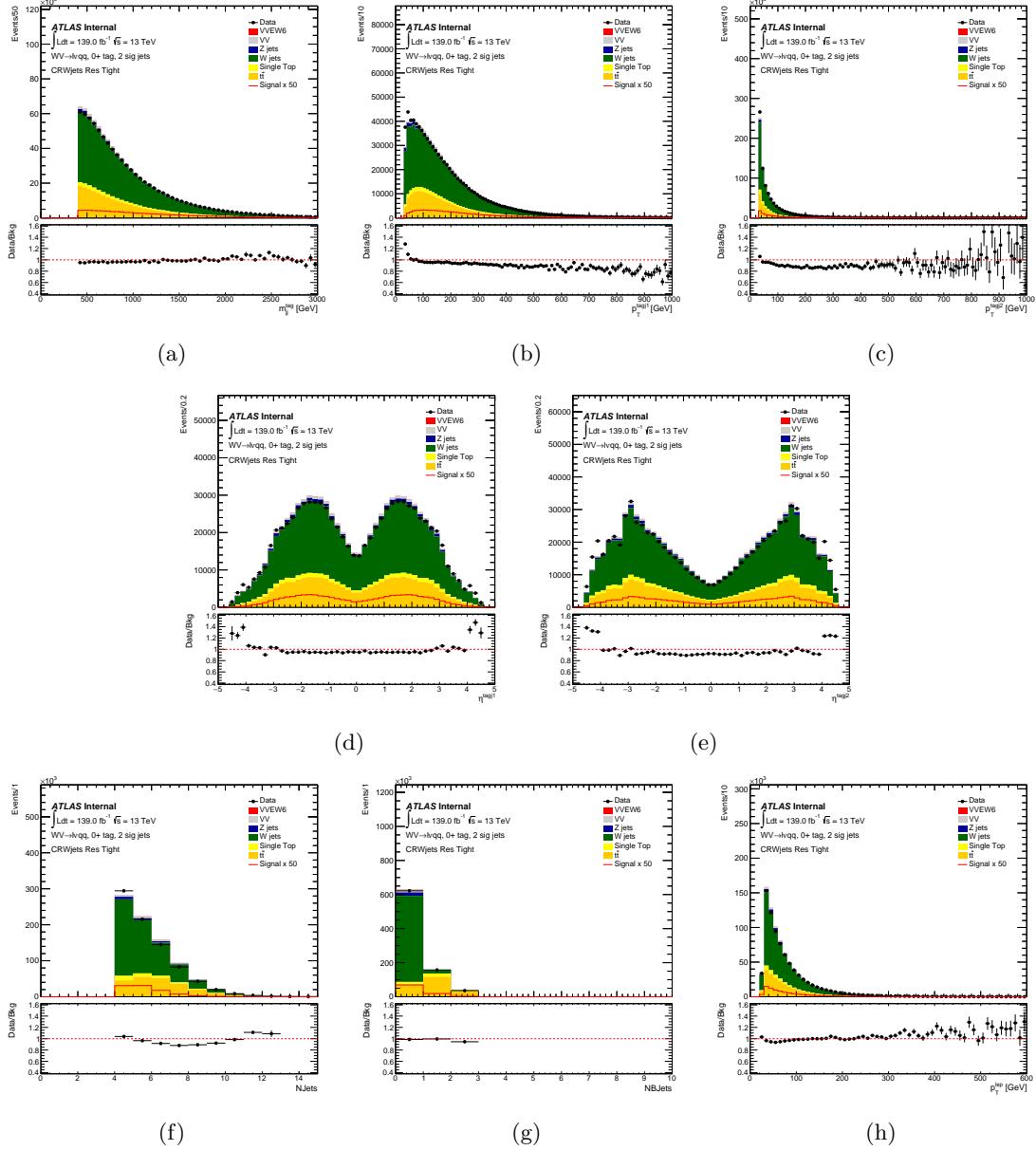


Figure 7.9: Data-MC checks for the resolved tight $W + \text{jets}$ control region in the 1-lepton channel.

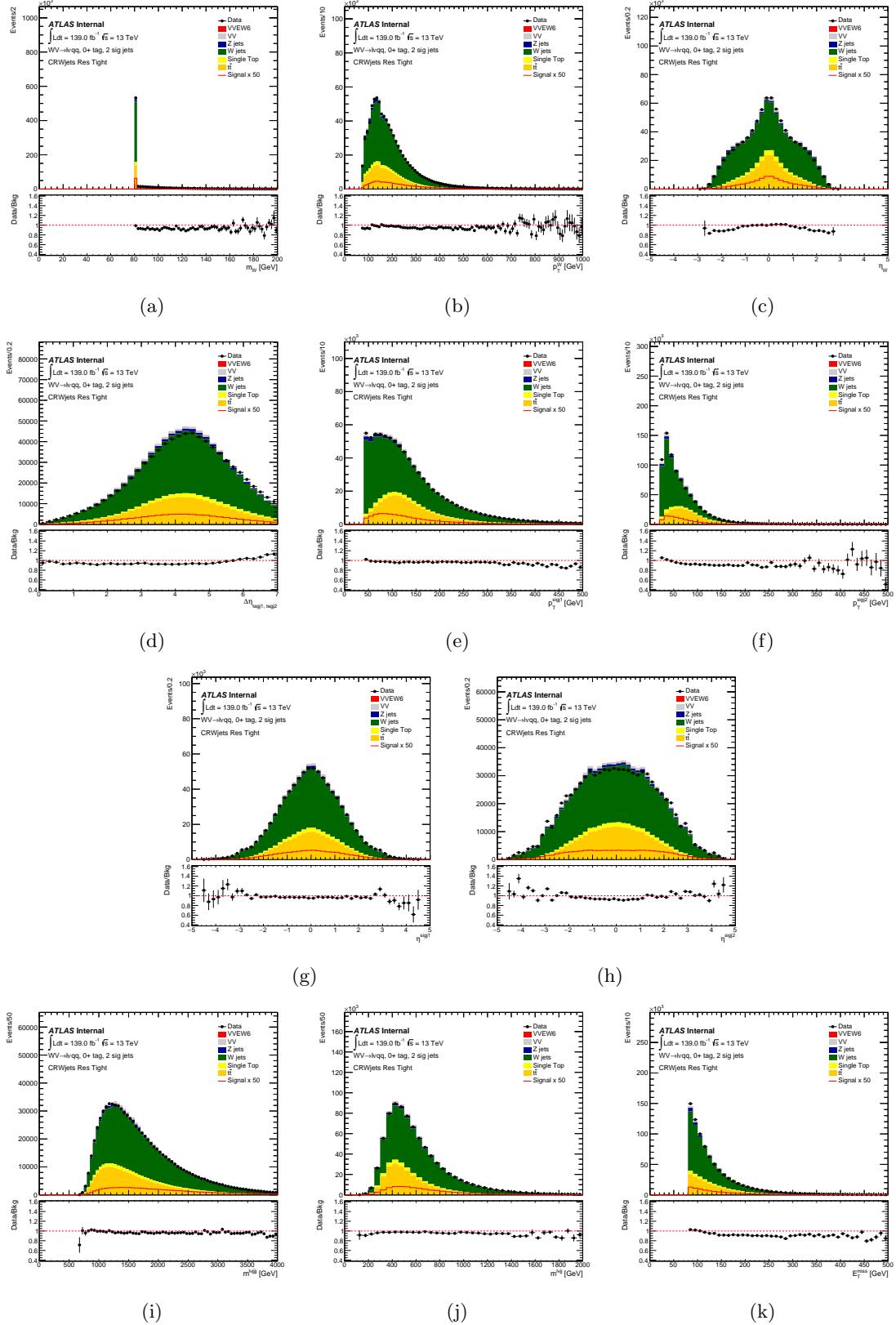


Figure 7.10: Data-MC checks for the resolved tight $W + \text{jets}$ control region in the 1-lepton channel.

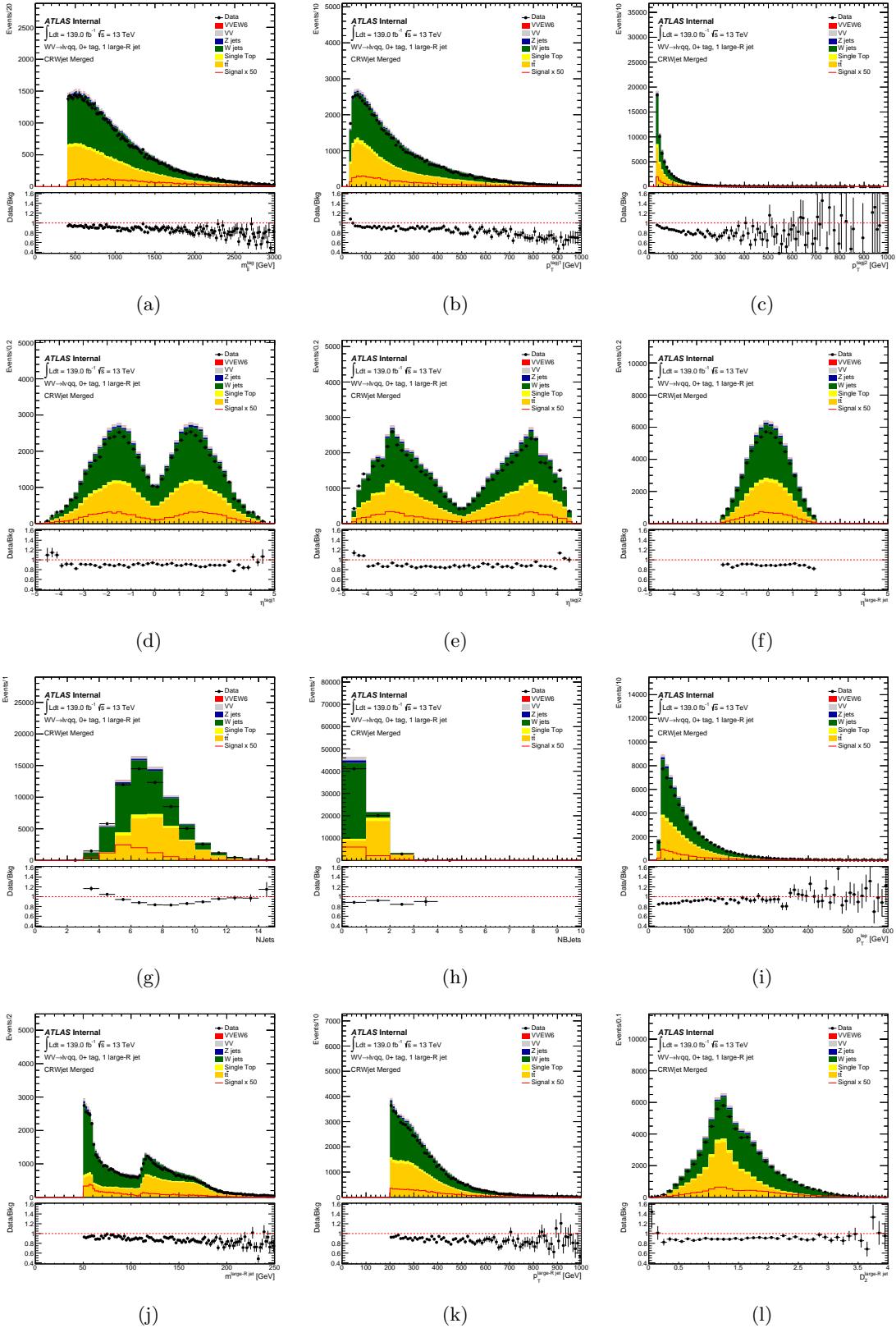


Figure 7.11: Data-MC checks for the merged $W + \text{jets}$ control region in the 1-lepton channel.

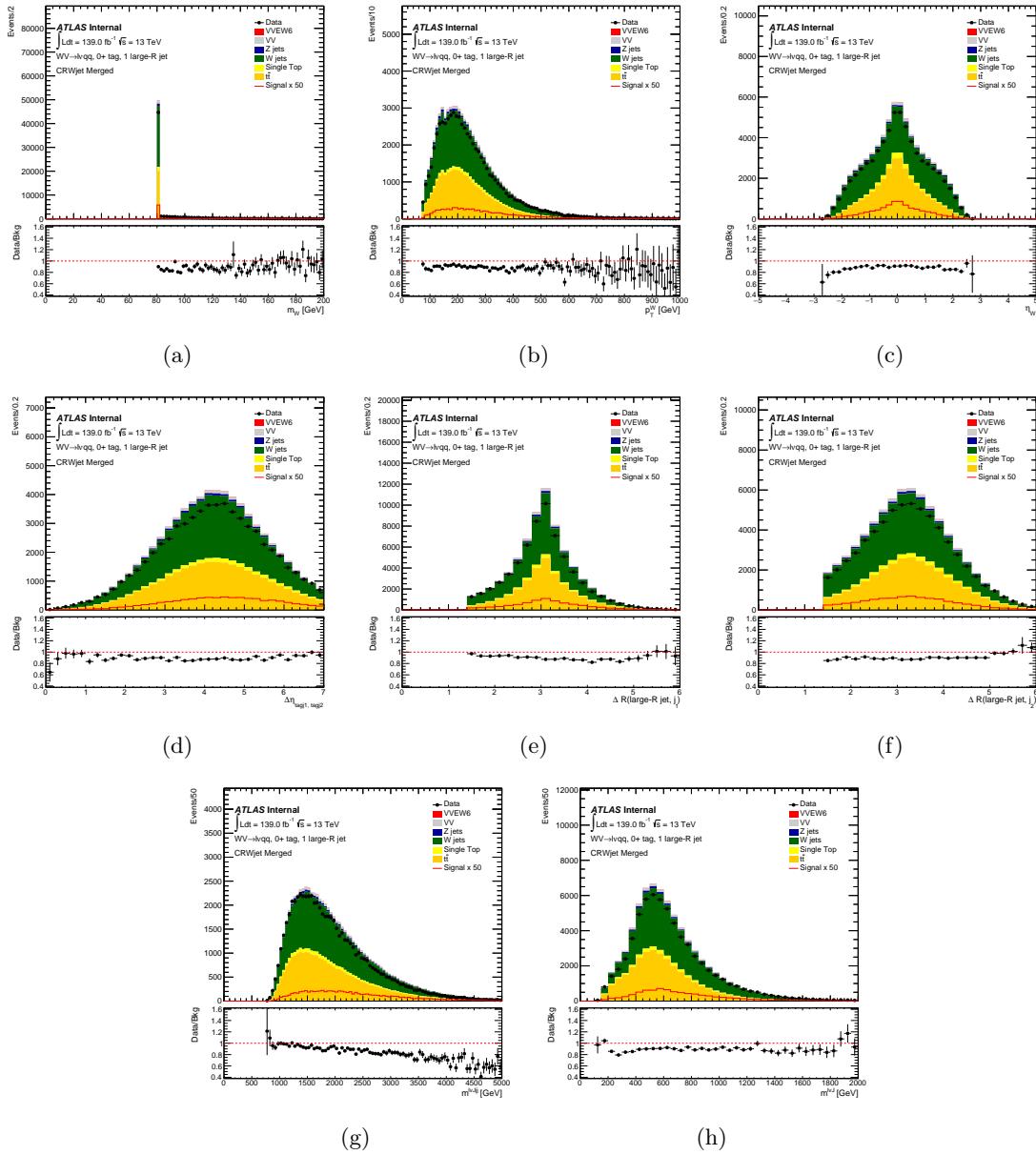


Figure 7.12: Data-MC checks for the merged $W + \text{jets}$ control region in the 1-lepton channel.

7.3 Top Background Control Region

Figures 7.13, 7.15, 7.17, and 7.19 show the main relevant kinematic variable distributions in the Top CRs for the 1-lepton channel, after m_{jj}^{tag} reweighting has been applied. The definitions of these CRs are provided in Section 6.6.2. It is worth noting that the resolved loose top control region is utilized solely to study and validate the agreement between the data and MC samples and is not included in the inputs for the statistical analysis.

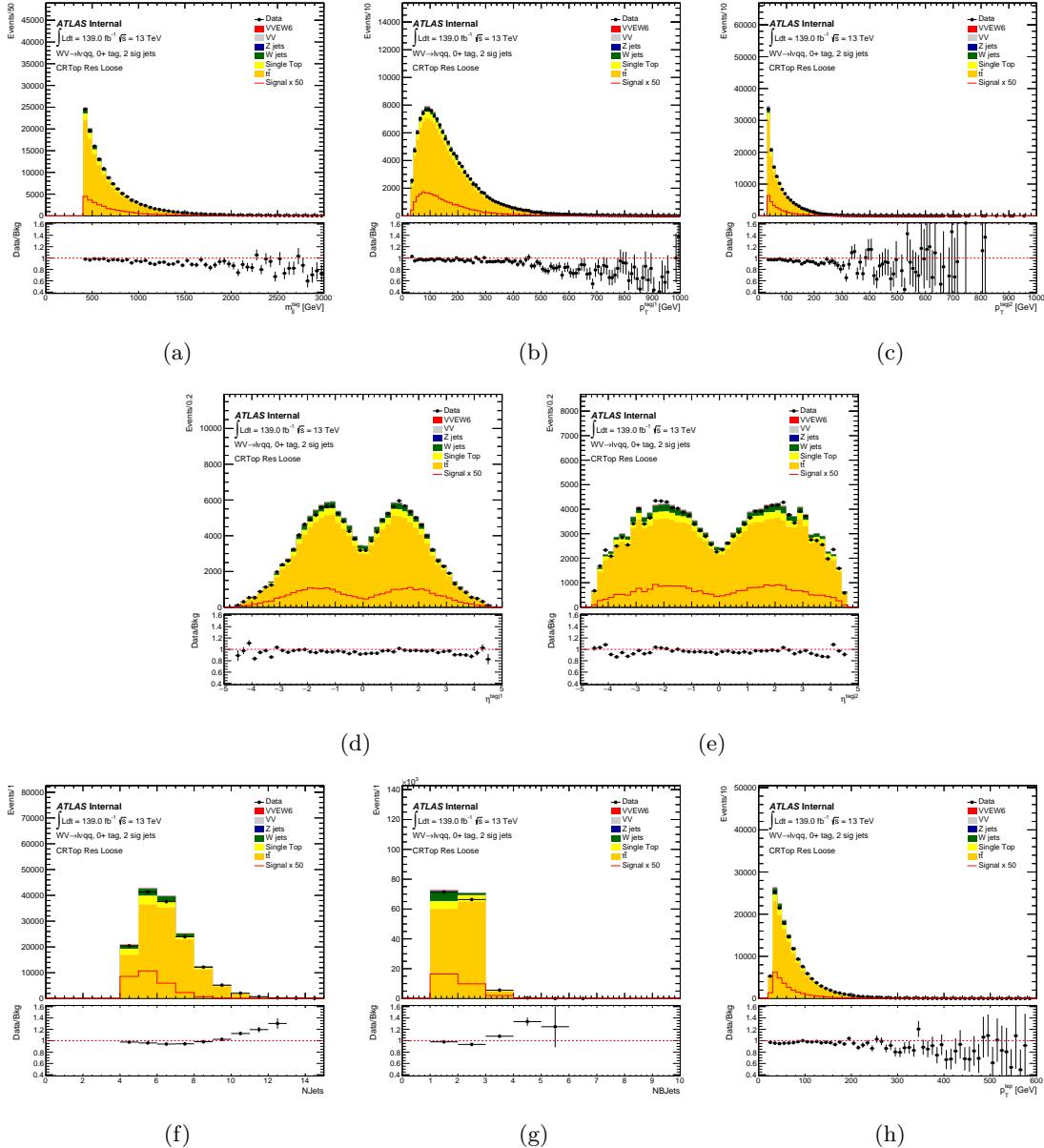


Figure 7.13: Data-MC checks for the resolved loose top control region in the 1-lepton channel.

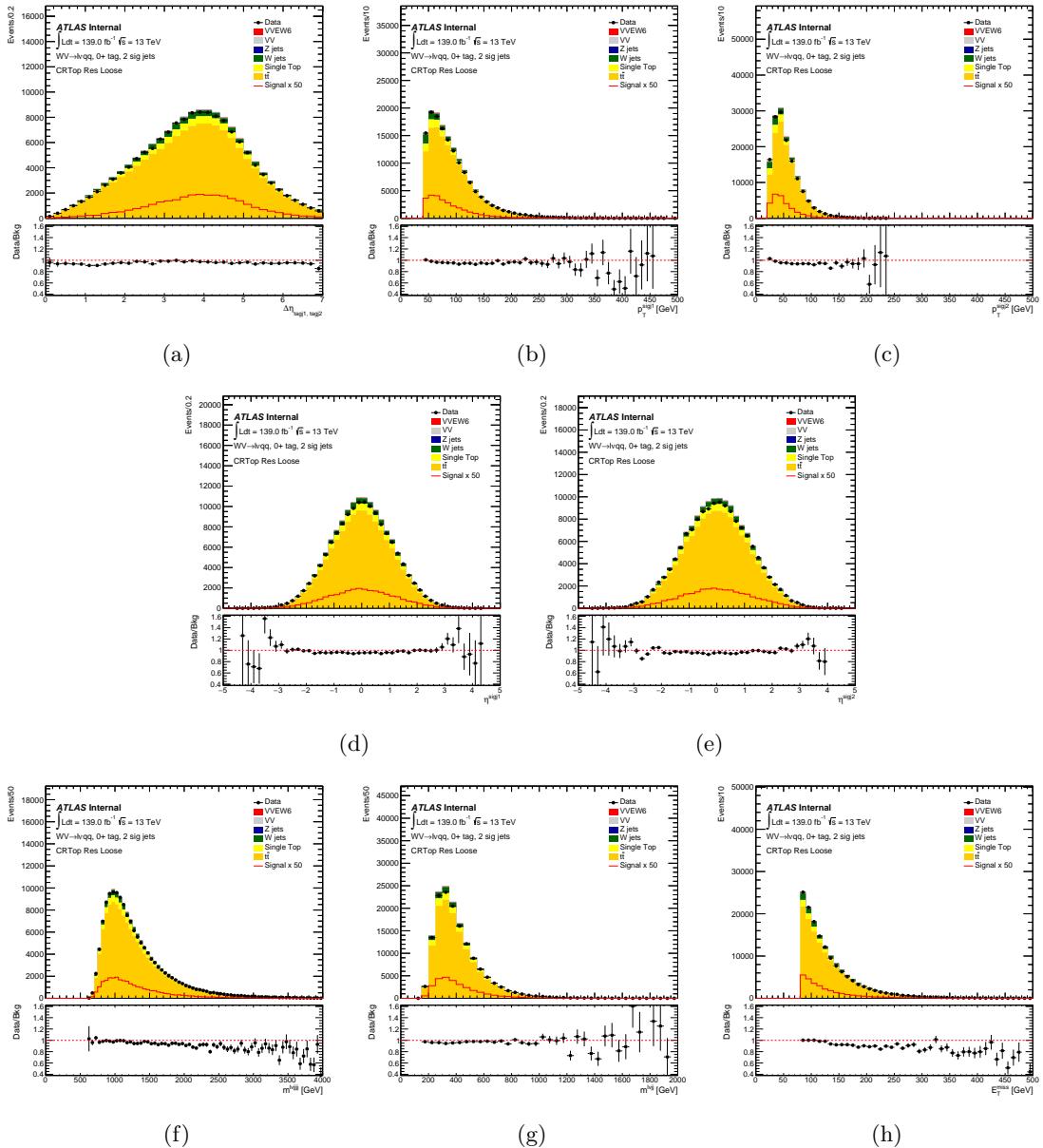


Figure 7.14: Data-MC checks for the resolved loose top control region in the 1-lepton channel.

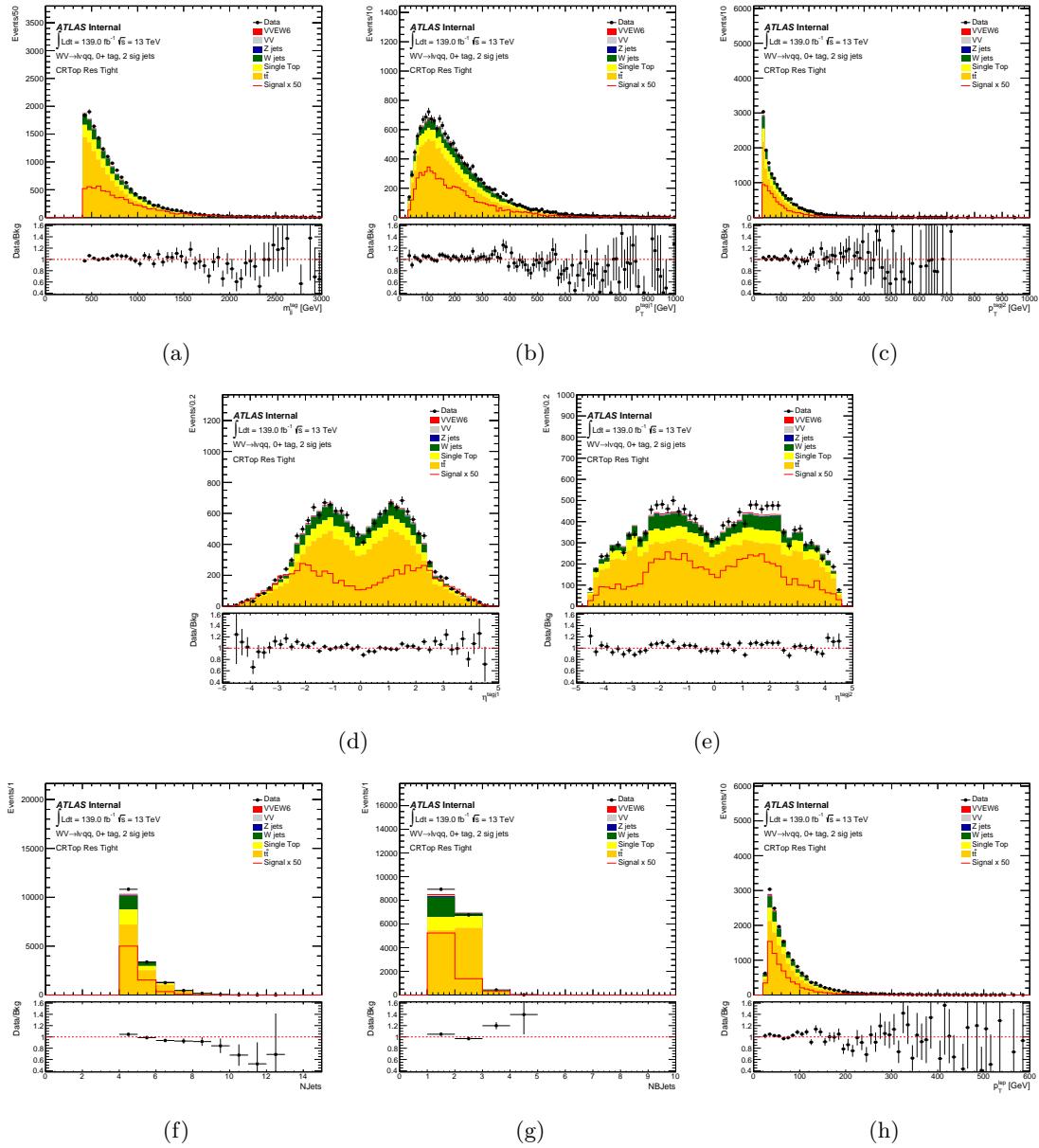


Figure 7.15: Data-MC checks for the resolved tight top control region in the 1-lepton channel.

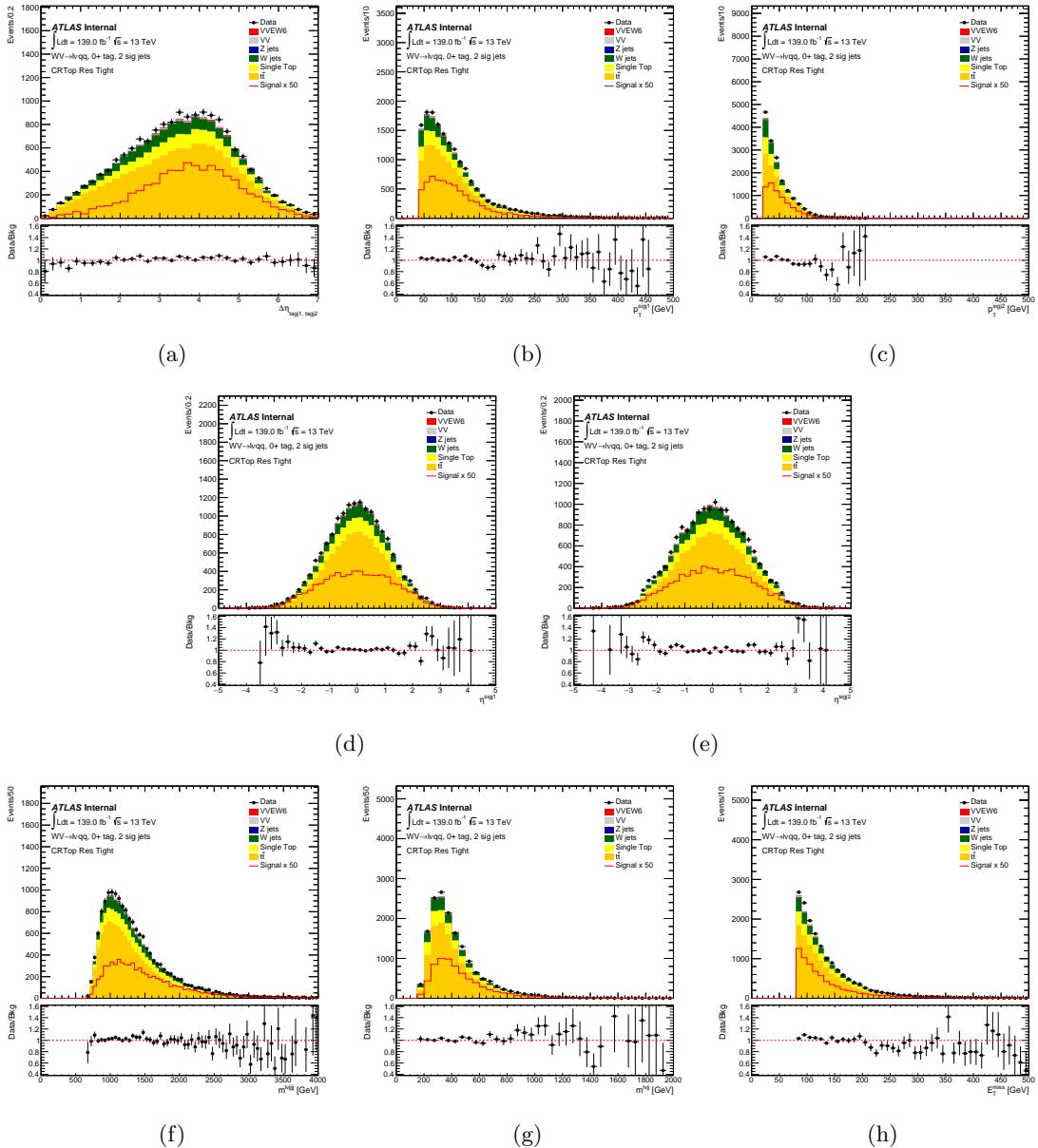


Figure 7.16: Data-MC checks for the resolved tight top control region in the 1-lepton channel.

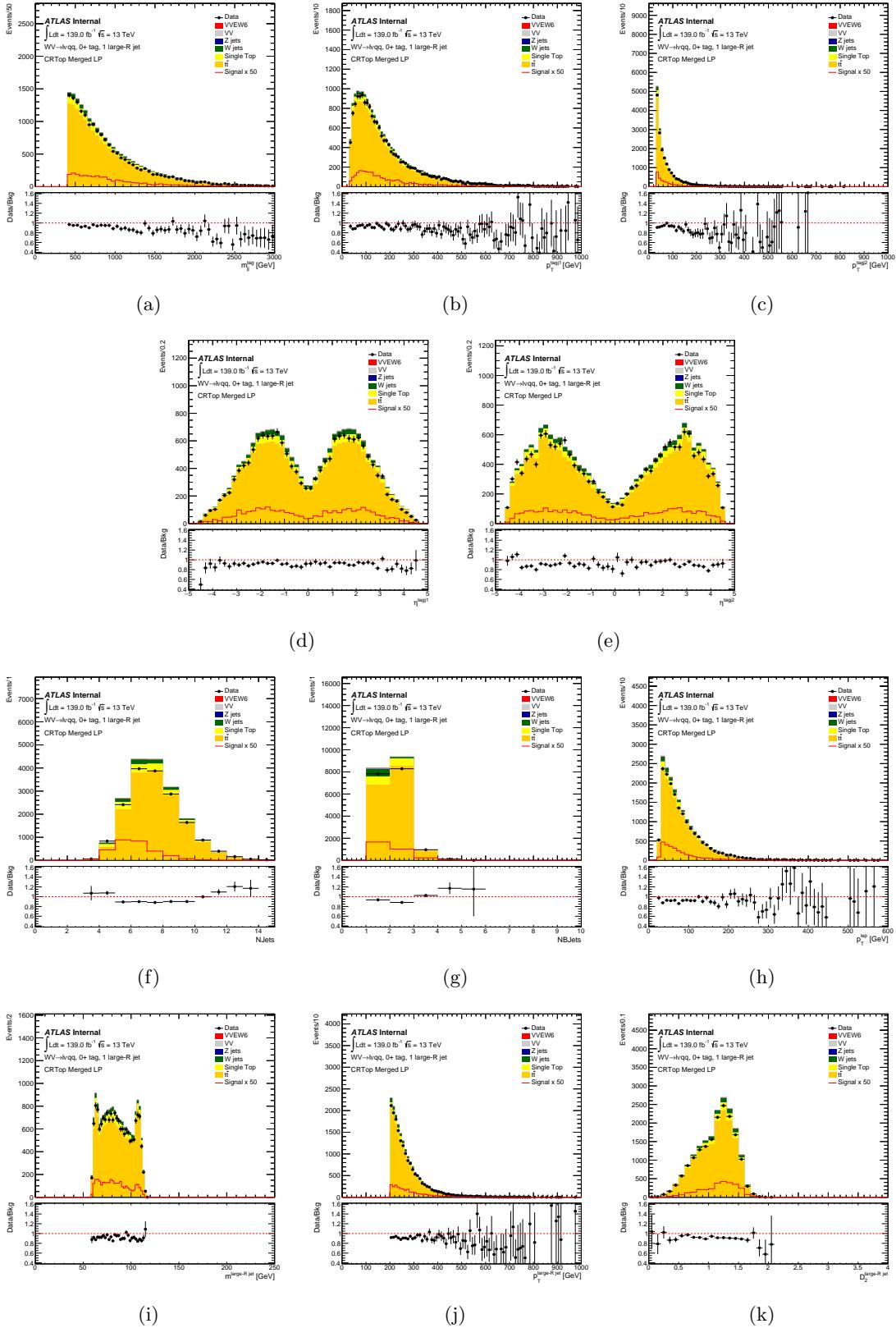


Figure 7.17: Data-MC checks for the merged low-purity top control region in the 1-lepton channel.

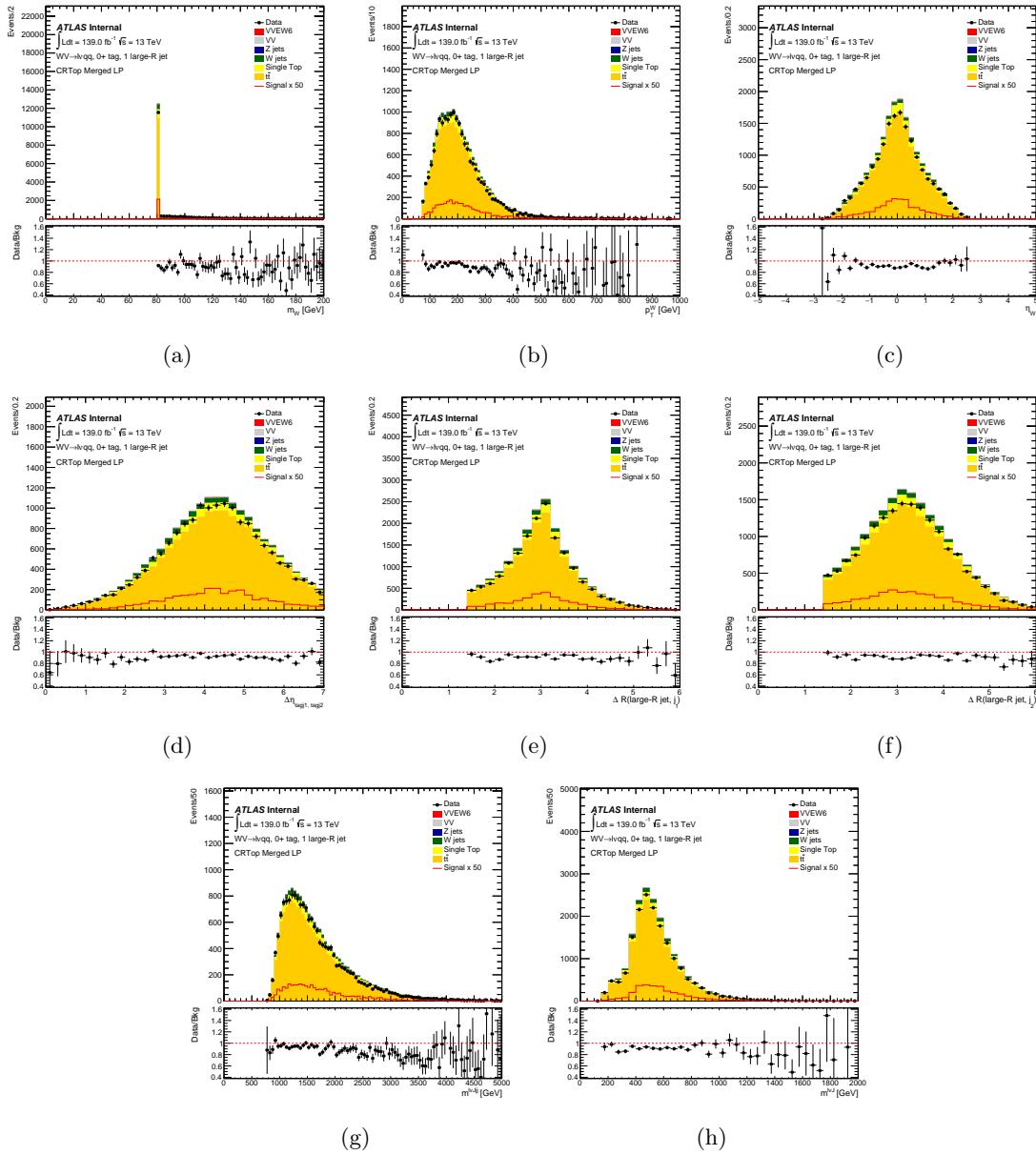


Figure 7.18: Data-MC checks for the merged low-purity top control region in the 1-lepton channel.

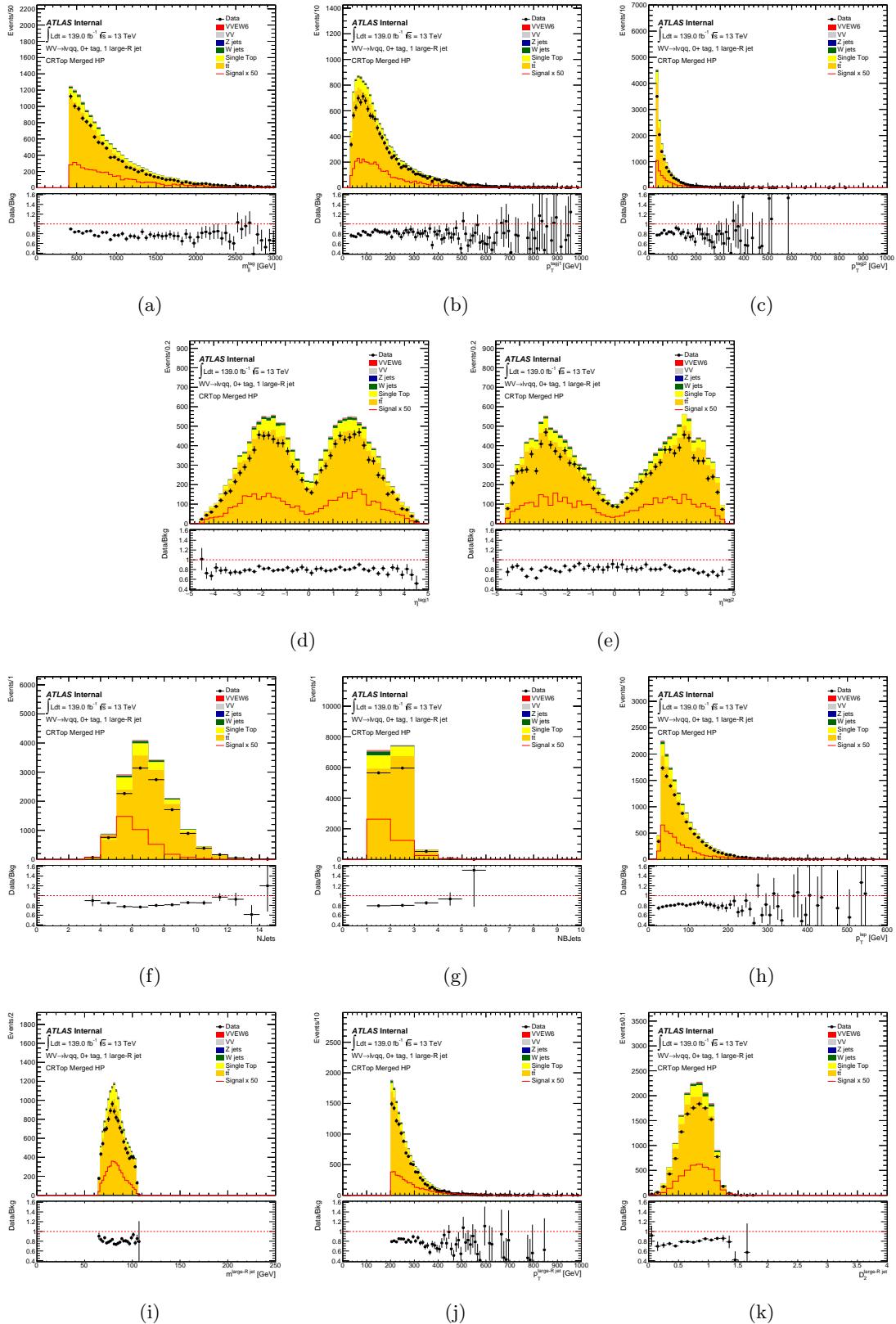


Figure 7.19: Data-MC checks for the merged high-purity top control region in the 1-lepton channel.

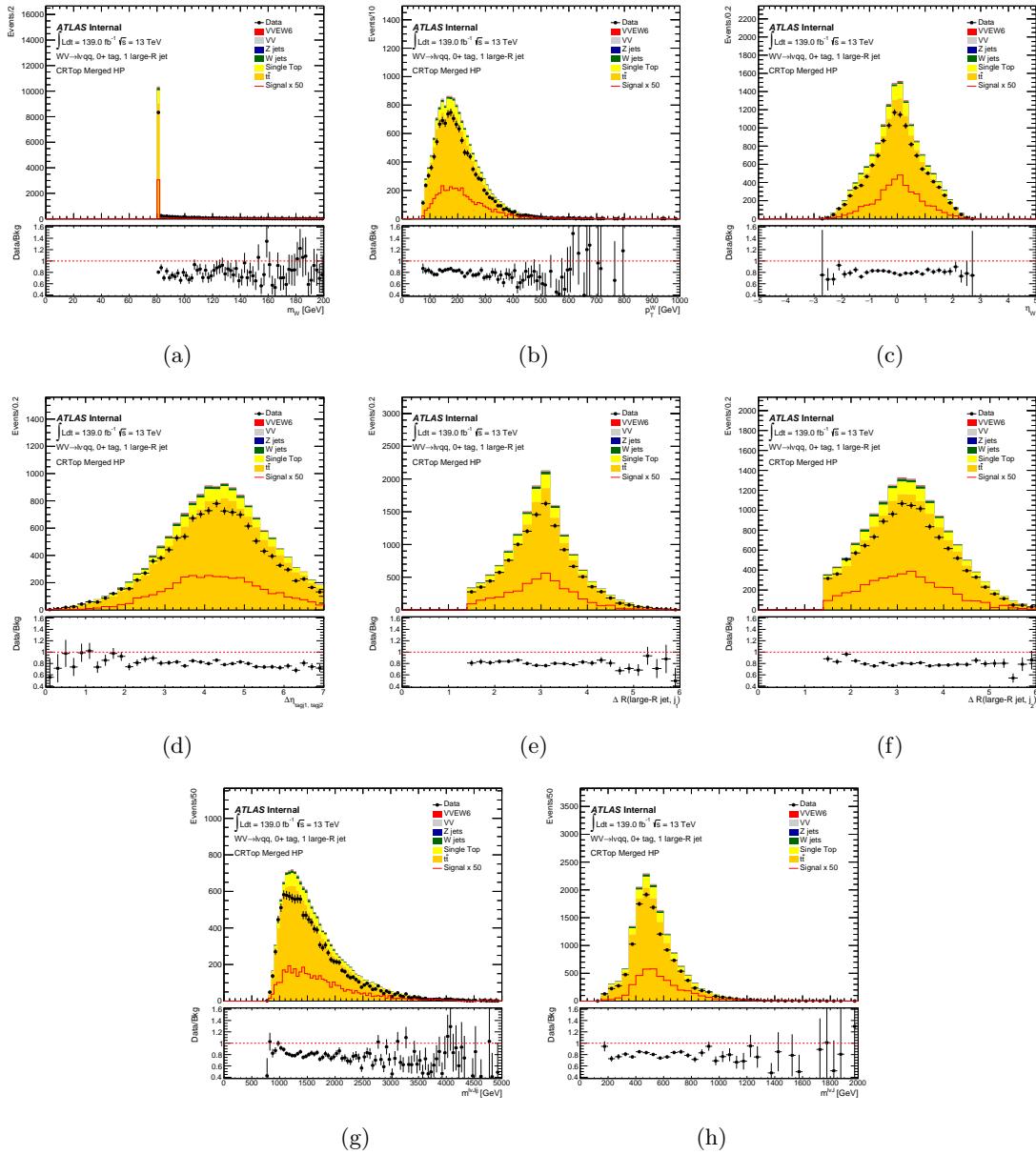


Figure 7.20: Data-MC checks for the merged high-purity top control region in the 1-lepton channel.

Chapter 8

Machine Learning Analysis

8.1 Motivations and Architecture

Within the analysis framework, I have adopted a machine learning strategy, termed the Deep Neural Network (DNN) approach, to improve sensitivity to the VBS signal. This method utilizes neural network models implemented using Keras [107], with TensorFlow serving as the backend [108]. Figure 8.1 illustrates the architecture of the DNN, which features multiple dense layers. The configuration of these layers, including the number of nodes and the specific setup, is detailed in Table 8.1. The hidden layers of the DNN employ the ReLU (rectified linear unit) function for activation. Alongside, L2 regularization is incorporated to minimize the risk of overfitting by penalizing the magnitude of the coefficients, thus maintaining model simplicity and robustness. The output layer utilizes a sigmoid activation function, ideal for binary classification tasks as it maps predictions to a probability scale, generating DNN scores that span a continuous range from zero to one.

Layer(type)	Number of Neurons	Activation Functions	Regularizers
Layer1(dense)	64	ReLU	L2
Layer2(dense)	32	ReLU	L2
Layer3(dense)	32	ReLU	L2
Layer4(dense)	32	ReLU	L2
Layer5(dense)	16	ReLU	L2
Layer Out(dense)	1	Sigmoid	-

Table 8.1: 1-lepton DNN structure in merged and resolved signal regions.

8.2 Input Variables and Feature Engineering

We assess the impact of input features on our DNNs using SHAP (SHapley Additive exPlanations), as introduced by Lundberg and Lee [109]. SHAP quantifies the contribution of each feature to the deviation of a particular prediction from the average outcome across the dataset. This method provides a game theory-based interpretable approximation of complex neural networks, facilitating a deeper understanding of how

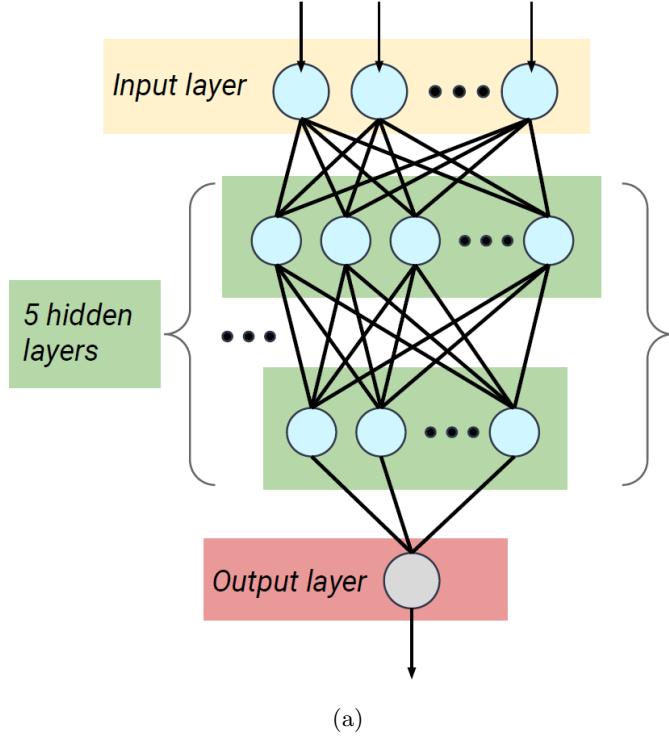


Figure 8.1: Simplified DNN architecture visualisation.

individual features influence the model’s predictions (DNN scores) for specific events. Employing SHAP values helps create a simplified yet faithful model that mirrors the original network’s performance, making the decision-making process transparent and easier to interpret.

To optimize the DNN models, we use a backward feature elimination method, guided by SHAP value rankings, which is a recognized technique in feature selection. This approach involves systematically removing the least important input features, as determined by SHAP values, and then retraining the DNN models after each round to assess the impact of these eliminations on model performance. SHAP values are recalculated after each training session to ensure that the feature rankings are continuously updated. In the end, we retain 15 input features for the merged category and 17 for the resolved category, demonstrating tailored optimizations for each. The final selection of input features for both categories is documented in Table 8.2, and their corresponding SHAP value rankings are illustrated in Figure 8.3.

The term “full system” refers to variables associated with the entire set of signal jet(s), lepton(s), and tagging jets. The boson centrality, $\xi(V)$, is defined as :

$$\xi(V) = \min(\Delta\eta_-, \Delta\eta_+) \quad (8.1)$$

where

$$\Delta\eta_- = \min(\eta_{V_{lep}}, \eta_{V_{had}}) - \min(\eta_{tag,j_1}, \eta_{tag,j_2})$$

and

$$\Delta\eta_+ = \max(\eta_{tag,j_1}, \eta_{tag,j_2}) - \max(\eta_{V_{lep}}, \eta_{V_{had}})$$

Figures 8.4, 8.5, 8.6, and 8.7 display the distributions of input features within the SRs. The noticeable differences in distribution shapes between the signal and background MC samples, as shown in these figures, highlight the effectiveness of our feature selection process.

Figure 8.2 presents a comparison of the performance metrics for the initial and final DNN models in both the merged and resolved categories, demonstrating that both models achieve comparable levels of effectiveness. To better understand this comparison, we define signal efficiency and background rejection as follows:

$$\text{Signal Efficiency} = \frac{\text{Number of Signal events with DNN} > X}{\text{Total Number of Signal events}} \quad (8.2)$$

$$\text{Background Rejection} = \frac{\text{Number of Background events with DNN} < X}{\text{Total Number of Background events}}. \quad (8.3)$$

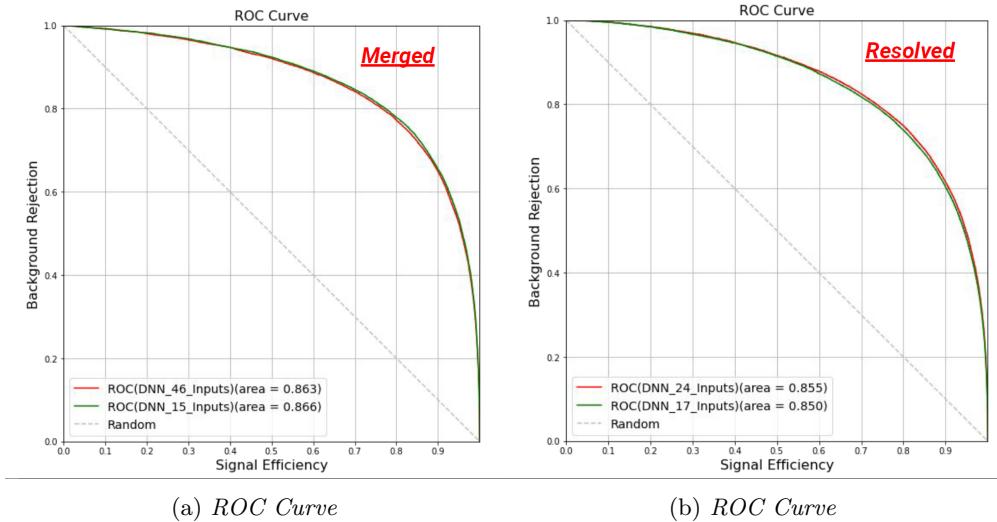


Figure 8.2: ROC curves for the DNN models in the merged (a) and resolved (b) categories. Here, “area” refers to the AUC (Area Under the ROC Curve). The AUC values are computed to offer quantitative measures of performance.

feature	merged	resolved
signal jet(s) mass	$m(J^{\text{sig}})$	$m(jj^{\text{sig}})$
signal jet transverse momentum	—	$p_T(j_{\text{lead}}^{\text{sig}})$ and $p_T(j_{\text{sublead}}^{\text{sig}})$
signal jet width	—	$W(j_{\text{lead}}^{\text{sig}})$ and $W(j_{\text{sublead}}^{\text{sig}})$
dijet(signal jet) transverse momentum	—	$p_T(jj^{\text{sig}})$
number of tracks associated to the signal jet(s)	$N_{\text{trk}}(J^{\text{sig}})$	—
multiplicity of B-tagged jets	$N(j^{\text{B-tagged}})$	—
multiplicity of forward jets	$N(j^{\text{forward}})$	—
multiplicity of track jets	—	$N(j^{\text{track}})$
diboson mass	$m(V^{\text{had}}V^{\text{lep}})$	—
leading tagging jet mass	$m(j_{\text{lead}}^{\text{tag}})$	—
full system mass	—	$m(V^{\text{had}}V^{\text{lep}} + jj^{\text{tag}})$
tagging jet transverse momentum	$p_T(j_{\text{sublead}}^{\text{tag}})$	$p_T(j_{\text{lead}}^{\text{tag}})$ and $p_T(j_{\text{sublead}}^{\text{tag}})$
tagging jet width	$W(j_{\text{lead}}^{\text{tag}})$ and $W(j_{\text{sublead}}^{\text{tag}})$	—
boson centrality	$\xi(V)$	—
pseudo-rapidity of tagging jets	—	$\eta(j_{\text{lead}}^{\text{tag}})$ and $\eta(j_{\text{sublead}}^{\text{tag}})$
pseudo-rapidity of lepton	—	$\eta(l)$
lepton transverse momentum	$p_T(l)$	—
number of tracks associated to the tagging jets	$N_{\text{trk}}(j_{\text{lead}}^{\text{tag}})$ and $N_{\text{trk}}(j_{\text{sublead}}^{\text{tag}})$	—
jets multiplicity	$N(j)$	$m(jj^{\text{tag}})$
tagging jets mass	—	$\Delta\eta(j_{\text{lead}}^{\text{tag}}, j_{\text{sublead}}^{\text{tag}})$
tagging jet separation	—	E_{ℓ}
lepton energy	—	—

Table 8.2: Input variables for the 1-lepton DNN in merged and resolved signal regions.

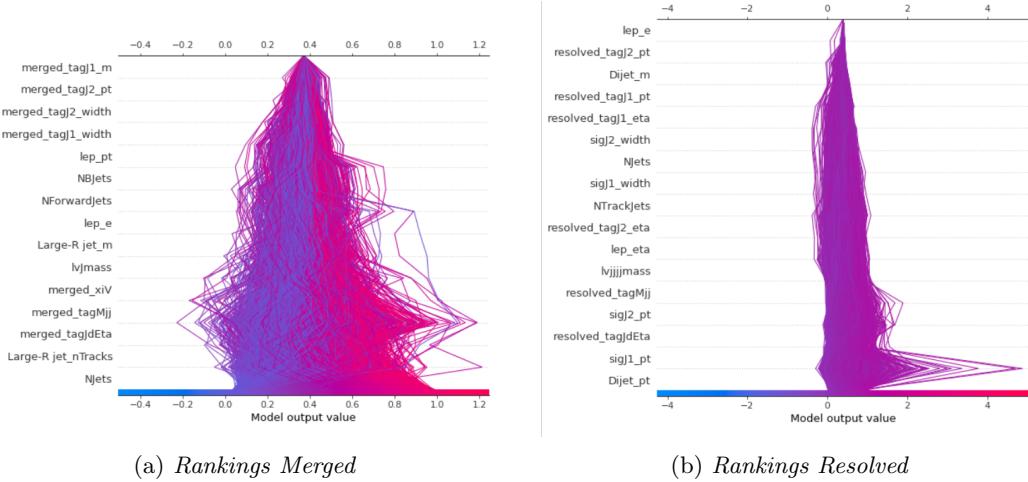


Figure 8.3: SHAP value rankings of input variables for both merged and resolved regimes; variables with lower impact appear at the top.

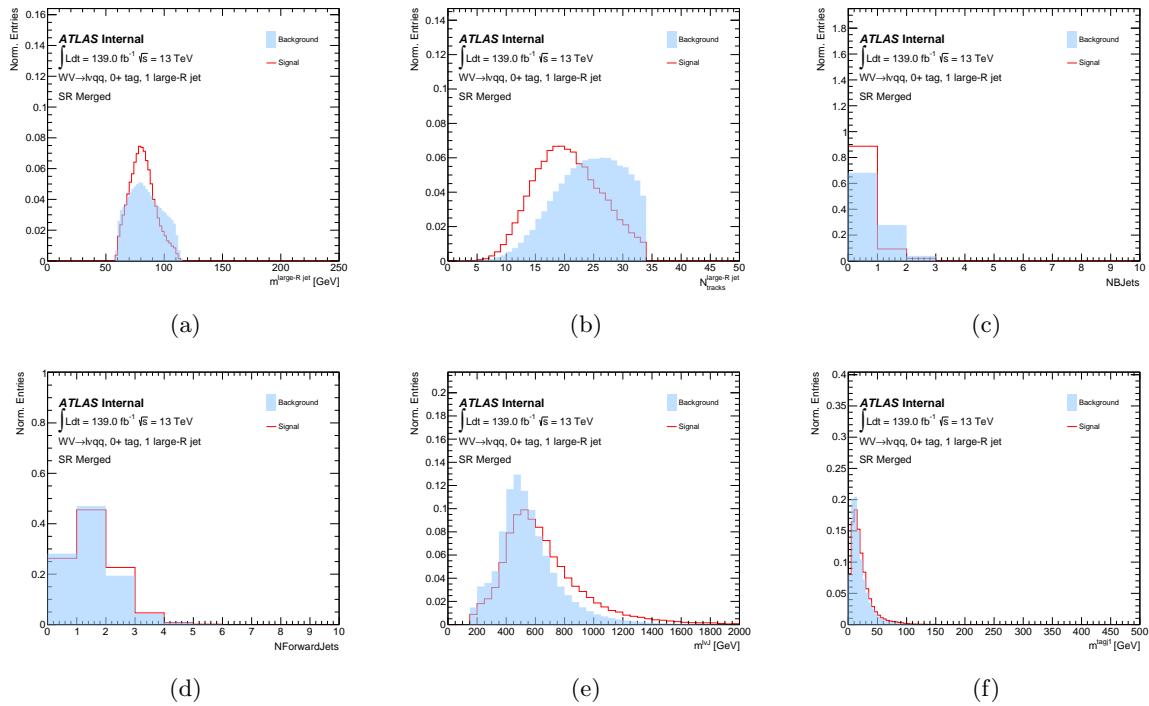


Figure 8.4: Distributions of input variables in the Merged SR (Continued on next page)

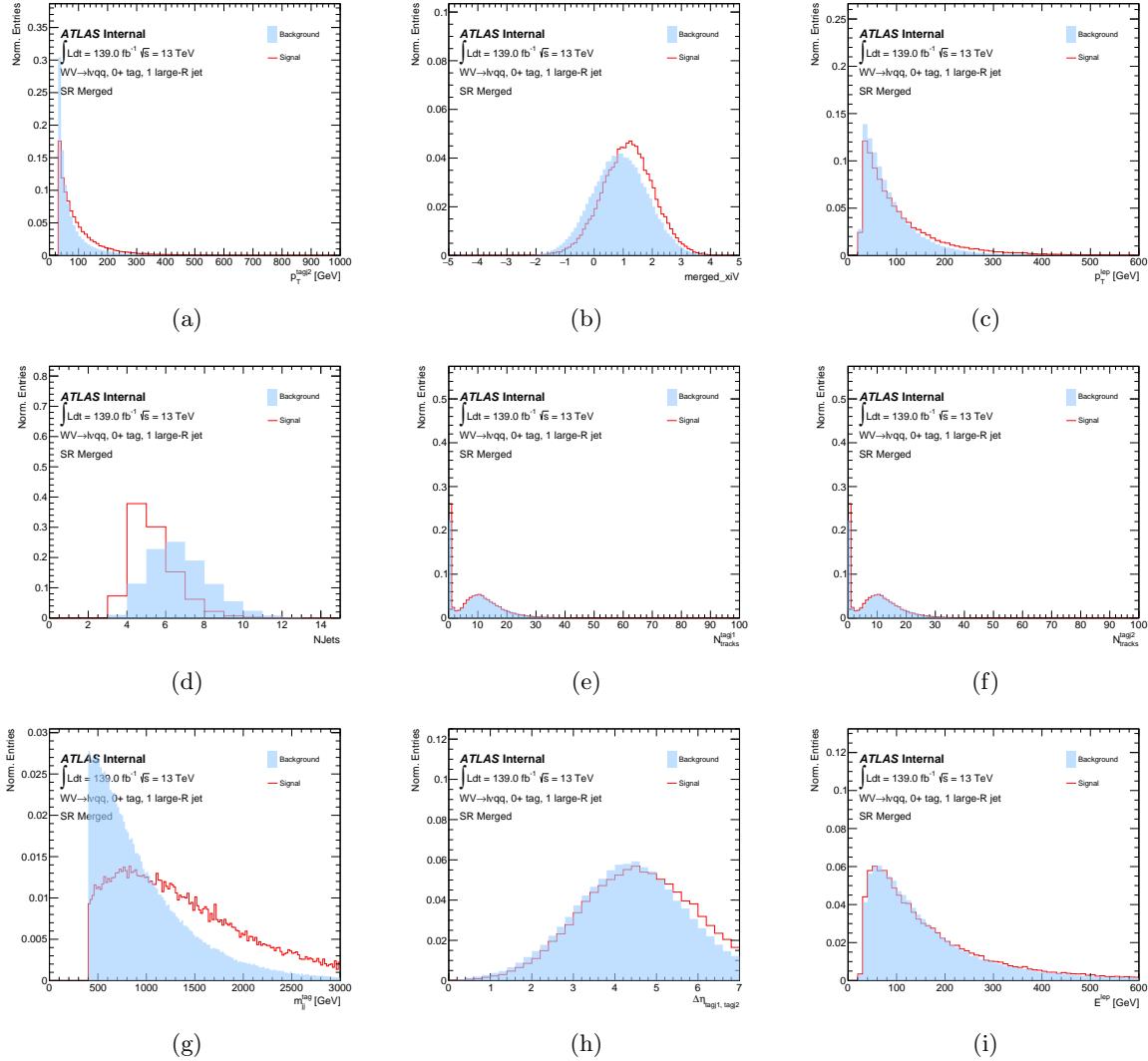


Figure 8.5: Distributions of input variables in the Merged SR (Continued)

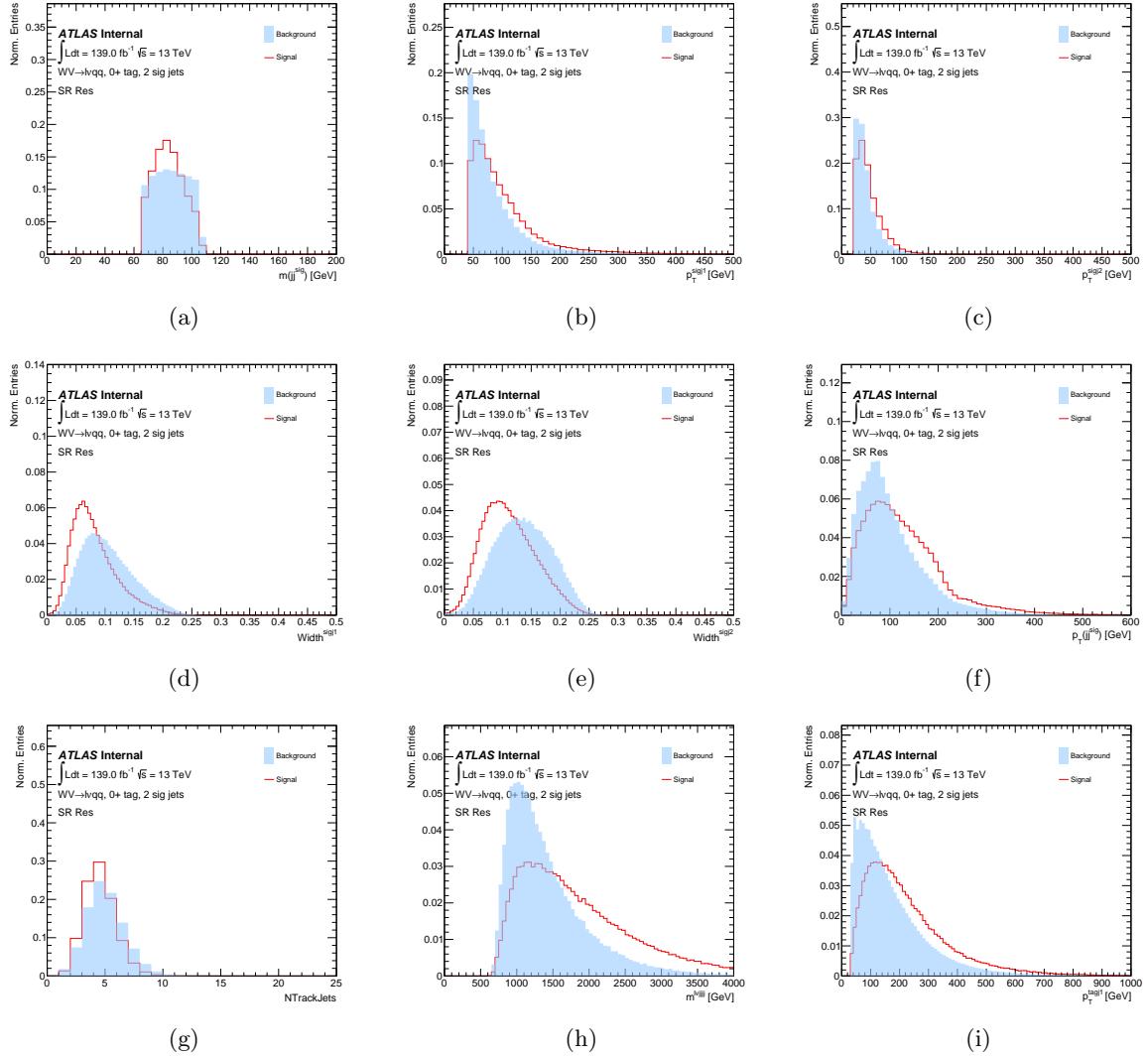


Figure 8.6: Distributions of input variables in the Resolved SR (Continued on next page)

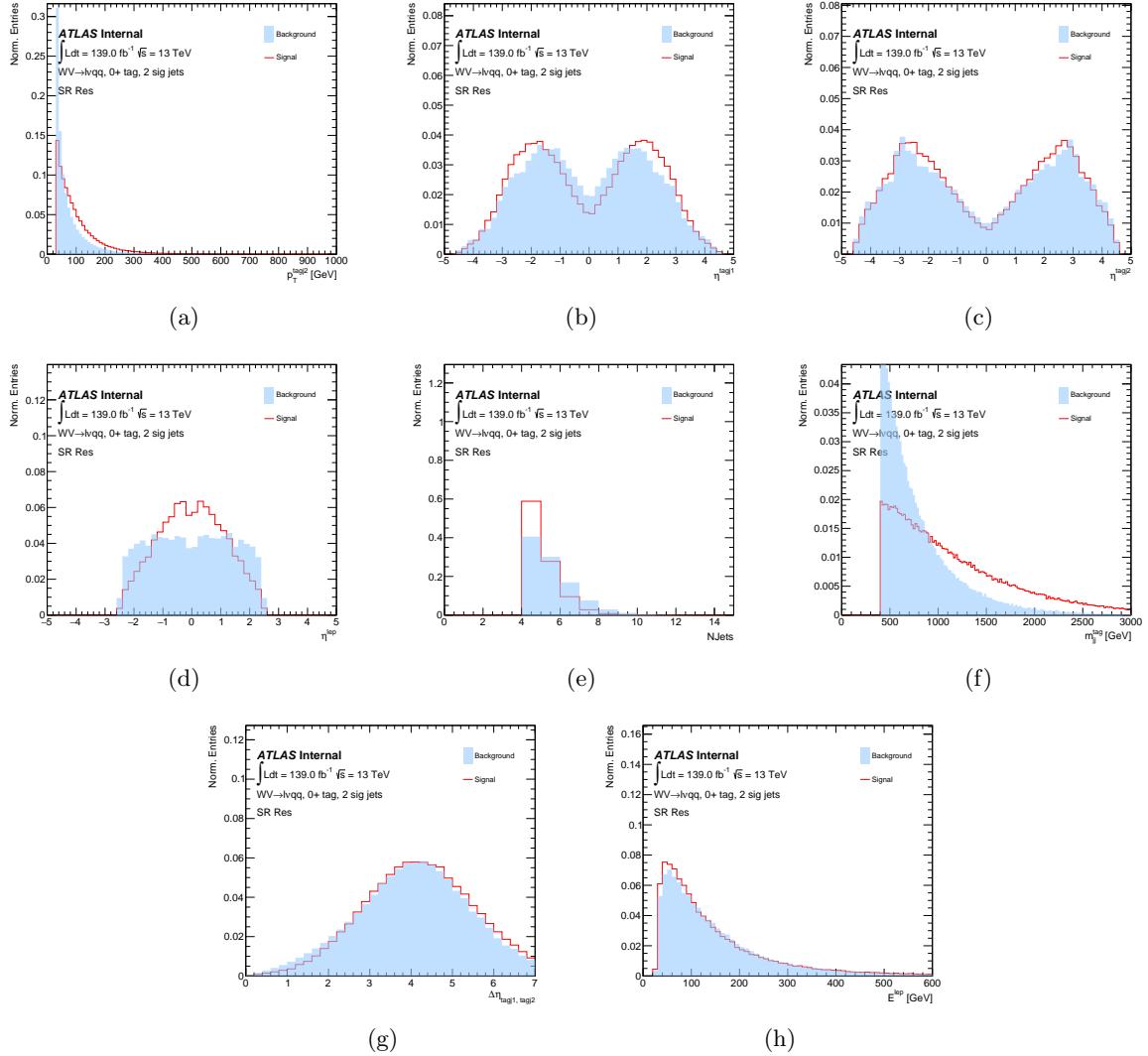


Figure 8.7: Distributions of input variables in the Resolved SR (Continued)

8.3 Trainings and Validations

We train separate DNN models for the merged and resolved categories. In the merged category, a single model is trained using events from both high and low purity SRs, whereas the resolved category uses its own dedicated model. For each category, the MC sample is divided into two parts: one for training and the other for validation. Within this framework, signal EW $VV + jj$ events are labeled as “1” and background events as “0”.

The loss function employed is binary cross-entropy, which is well-suited for models that output binary labels, as demonstrated in Figure 8.8. Experiments with Adaptive Moment Estimation (Adam) and Stochastic Gradient Descent (SGD) optimizers reveal a preference for SGD due to its effectiveness in fine-tuning the distribution of DNN scores, likely attributed to its more consistent convergence behavior in our specific application.

In this analysis, we employ the k-fold cross-validation method to fully leverage our MC samples, ensuring an unbiased evaluation of model performance. We utilize a 5-fold cross-validation strategy, which divides the dataset into five equally sized parts (or folds). The DNN is trained and validated five times, using a cyclical process. During each iteration, four folds are used for training the model, while the remaining fold serves as the validation set. This approach ensures that every MC event in the Signal Regions (SRs) is used for both training and validation across the iterations, thereby enhancing the robustness and generalizability of our model. Figure 8.9 provides a clear indication of the reliability of the DNN approach across different subsets of the dataset.

For practicality and reliability, I implement a 2-fold cross-validation method for the DNN approach. The complete MC sample is evenly divided into two parts: one for training and the other for validation. The consistent performance across these datasets, highlighted in Figure 8.10, indicates an absence of training set bias. This consistency ensures the model’s robustness and unbiased evaluation.

Figure 8.11 demonstrates the effective separation between signal and background in the DNN distributions for both merged and resolved regimes. Figure 8.12 presents the distributions of DNN scores across all three signal regions, with some data bins partially blinded.

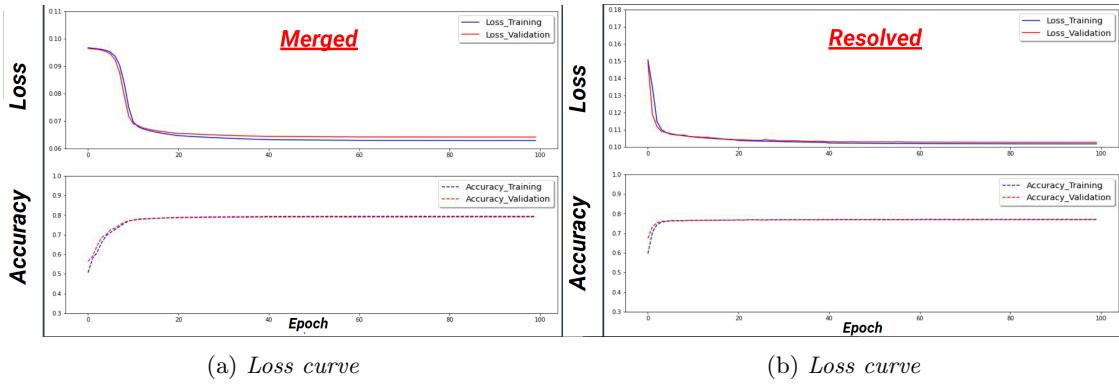


Figure 8.8: Trainings for the DNN models in the merged and resolved regimes.

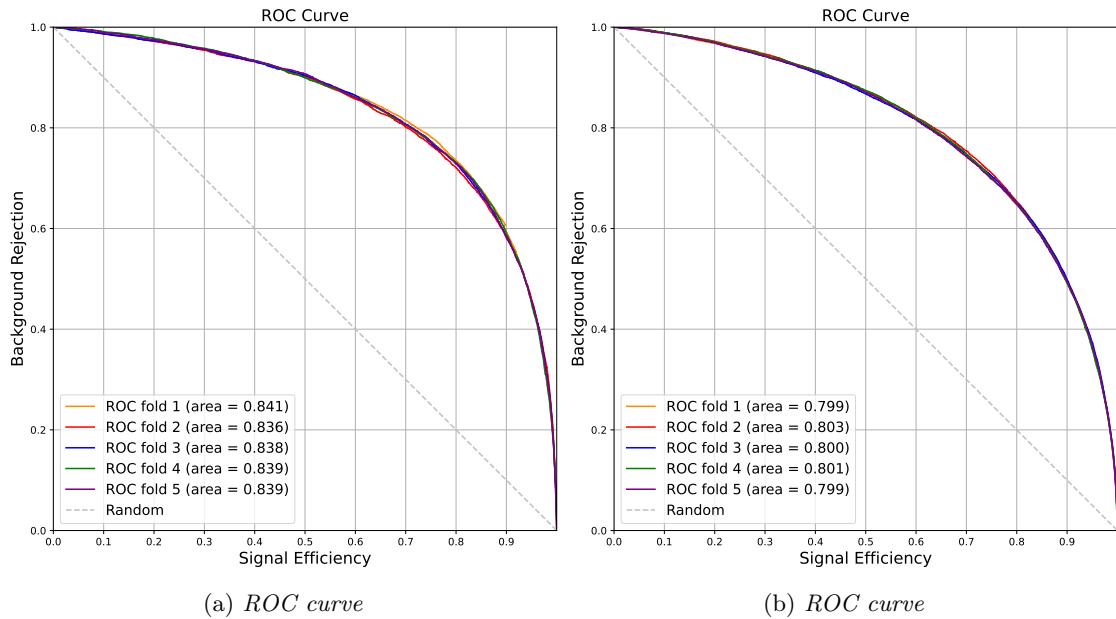


Figure 8.9: K-fold cross-validation for the DNN model trained in the merged and resolved regimes.

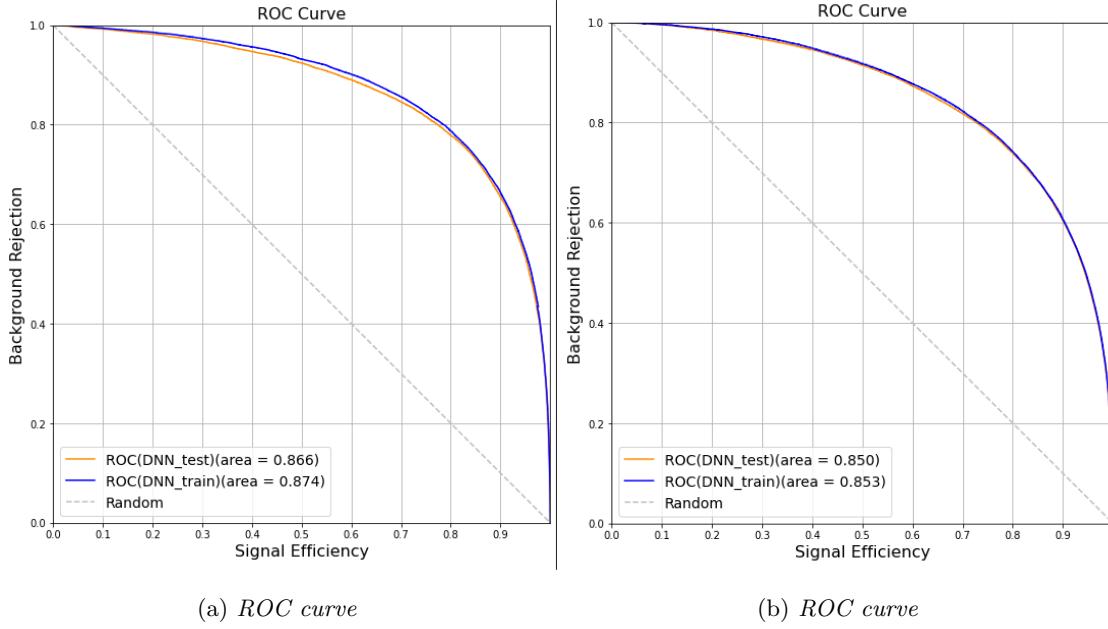


Figure 8.10: Validation for the DNN model trained in the merged and resolved regimes.

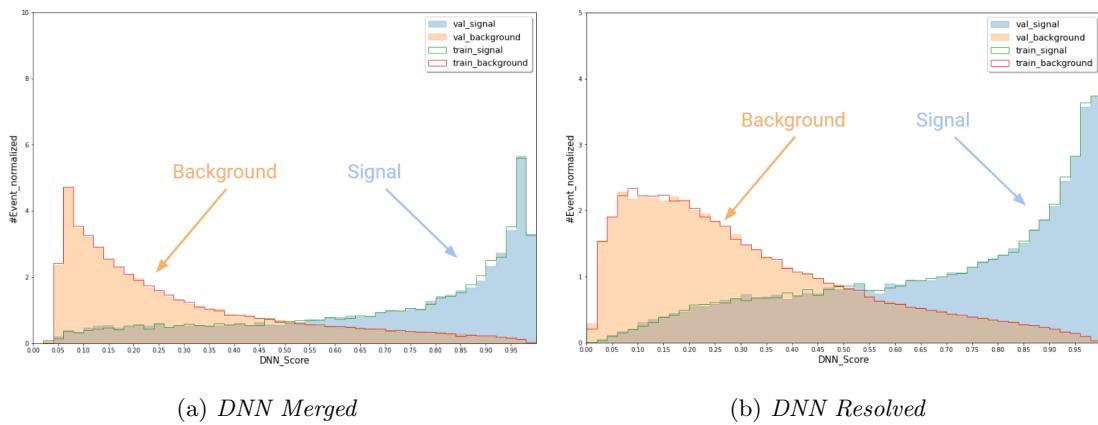


Figure 8.11: Performance for the DNN model trained in the merged and resolved regimes.

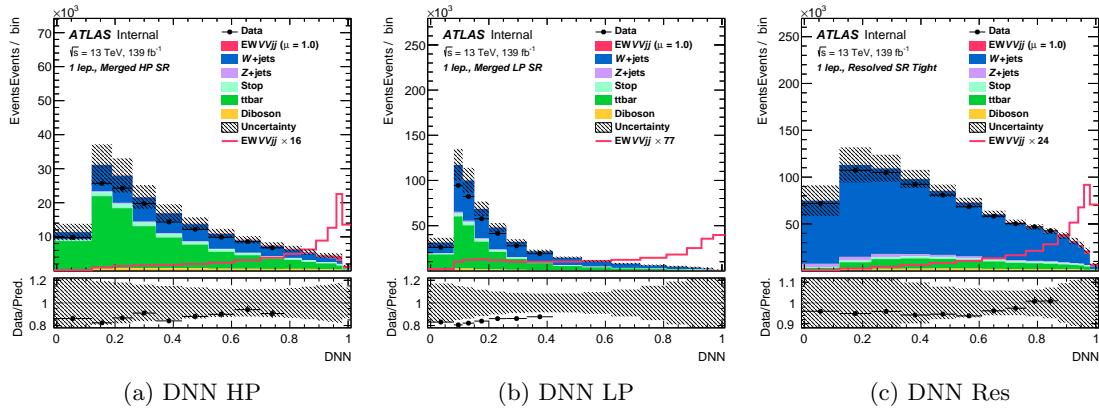


Figure 8.12: DNN score distributions for the high purity merged, low purity merged, and resolved signal regions.

Chapter 9

Systematic Uncertainties

This chapter describes the sources of systematic uncertainties considered in the analysis, divided into three categories: experimental uncertainties, uncertainties on the modelling of background processes, and theoretical uncertainties on the signal processes. In the statistical analysis, each systematic uncertainty is treated as a nuisance parameter (NP); the names of these parameters are defined below. Unless specified, these NPs are included in the final binned maximum likelihood fit, which will be discussed further in Chapter 10.

9.1 Experimental Uncertainties

9.1.1 Baseline uncertainties

The summary of experimental uncertainties is presented in Tab. 9.1 to 9.3.

Table 9.1: Qualitative summary of the systematic uncertainties included in this analysis.

Source	Description	Analysis Name	Notes
Electrons	Energy scale	EG SCALE ALL	
Electrons	Energy resolution	EG RESOLUTION ALL	
Electrons	Trigger	EL EFF Trigger TOTAL 1NPCOR PLUS UNCOR	
Electrons	ID efficiency SF	EL EFF ID TOTAL 1NPCOR PLUS UNCOR	
Electrons	Isolation efficiency SF	EL EFF Iso TOTAL 1NPCOR PLUS UNCOR	
Electrons	Reconstruction efficiency SF	EL EFF Reco TOTAL 1NPCOR PLUS UNCOR	
Muons	p_T scale	MUONS SCALE	
Muons	p_T scale (charge dependent)	MUON SAGITTA RHO	
Muons	p_T scale (charge dependent)	MUON SAGITTA RESBIAS	
Muons	p_T resolution MS	MUONS MS	
Muons	p_T resolution ID	MUONS ID	
Muons	Isolation efficiency SF	MUON ISO SYS	
Muons	Isolation efficiency SF	MUON ISO STAT	
Muons	Muon reco & ID efficiency SF	MUONS EFF STAT	
Muons	Muon reco & ID efficiency SF	MUONS EFF STAT LOWPT	
Muons	Muon reco & ID efficiency SF	MUONS EFF SYST	
Muons	Muon reco & ID efficiency SF	MUONS EFF SYST LOWPT	
Muons	Track-to-vertex association efficiency SF	MUON TTVA SYS	
Muons	Track-to-vertex association efficiency SF	MUON TTVA STAT	
Muons	Trigger	MUON EFF TrigSystUncertainty	
Muons	Trigger	MUON EFF TrigStatUncertainty	
MET	Soft term	MET SoftTrk ResoPerp	
MET	Soft term	MET SoftTrk ResoPara	
MET	Soft term	MET SoftTrk Scale	

Table 9.2: Qualitative summary of the systematic uncertainties included in this analysis.

Source	Description	Analysis Name	Notes
Small-R Jets	JES category reduction	JET_CR_JET_BJES_Response	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Detector1	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Detector2	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Mixed1	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Mixed2	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Mixed3	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Modelling1	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Modelling2	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Modelling3	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Modelling4	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Statistical1	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Statistical2	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Statistical3	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Statistical4	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Statistical5	
Small-R Jets	JES category reduction	JET_CR_JET_EffectiveNP_Statistical6	
Small-R Jets	JES category reduction	JET_CR_JET_Flavor_Composition	
Small-R Jets	JES category reduction	JET_CR_JET_Flavor_Response	
Small-R Jets	JES category reduction	JET_CR_JET_Pileup_OffsetMu	
Small-R Jets	JES category reduction	JET_CR_JET_Pileup_OffsetNPV	
Small-R Jets	JES category reduction	JET_CR_JET_Pileup_PtTerm	
Small-R Jets	JES category reduction	JET_CR_JET_Pileup_RhoTopology	
Small-R Jets	JES category reduction	JET_CR_JET_PunchThrough_MC16	
Small-R Jets	JES category reduction	JET_CR_JET_SingleParticle_HighPt	
Small-R Jets	JES category reduction	JET_CR_JET_EtaIntercalibration_TotalStat	
Small-R Jets	JES category reduction	JET_CR_JET_EtaIntercalibration_Modelling	
Small-R Jets	JES category reduction	JET_CR_JET_EtaIntercalibration_NonClosure_highE	
Small-R Jets	JES category reduction	JET_CR_JET_EtaIntercalibration_NonClosure_negEta	
Small-R Jets	JES category reduction	JET_CR_JET_EtaIntercalibration_NonClosure_posEta	
Small-R Jets	JER	JET_CR_JET_JER_DataVsMC	
Small-R Jets	JER	JET_CR_JET_JER_EffectiveNP_1	
Small-R Jets	JER	JET_CR_JET_JER_EffectiveNP_2	
Small-R Jets	JER	JET_CR_JET_JER_EffectiveNP_3	
Small-R Jets	JER	JET_CR_JET_JER_EffectiveNP_4	
Small-R Jets	JER	JET_CR_JET_JER_EffectiveNP_5	
Small-R Jets	JER	JET_CR_JET_JER_EffectiveNP_6	
Small-R Jets	JER	JET_CR_JET_JER_EffectiveNP_7restTerm	
Small-R Jets	JVT	JET_JvtEfficiency	

Table 9.3: Qualitative summary of the systematic uncertainties included in this analysis.

Source	Description	Analysis Name	Notes
Large-R Jets	p_T scale	FATJET_Medium_JET_Rtrk_Baseline_pT	
Large-R Jets	p_T scale	FATJET_Medium_JET_Rtrk_Modelling_pT	
Large-R Jets	p_T scale	FATJET_Medium_JET_Rtrk_TotalStat_pT	
Large-R Jets	p_T scale	FATJET_Medium_JET_Rtrk_Tracking_pT	
Large-R Jets	p_T scale	FATJET_BJT_JET_EtaIntercalibration_Modelling	
Large-R Jets	p_T scale	FATJET_BJT_JET_Flavor_Composition	
Large-R Jets	p_T scale	FATJET_BJT_JET_Flavor_Response	
Large-R Jets	Mass resolution	FATJET_JMR	
Large-R Jets	JER	FATJET_JER	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_B_0_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_B_1_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_B_2_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_C_0_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_C_1_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_C_2_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_C_3_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_Light_0_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_Light_1_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_Light_2_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_Eigen_Light_3_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_extrapolation_AntiKt4PFlowJets	
B-tagging	Flavor tagging scale factors	FT_EFF_extrapolation_from_charm_AntiKt4PFlowJets	
Pileup reweighting	PRW_DATASF	PRW_DATASF	
Luminosity	LumiNP	ATLAS_LUMI_2015_2018	

Luminosity

Luminosity uncertainty is applied to both the signal and background MC samples. The uncertainties on the integrated luminosity for the datasets are as follows:

- 2015+2016 dataset: 2.1%
- 2017 dataset: 2.4%
- 2018 dataset: 2.0%
- Combined Run-2 dataset (2015-2018): 1.7%

Pileup reweighting

The uncertainty associated with the pileup reweighting is accounted for as `PRW_DATASF`. A variation in the pileup reweighting of MC simulations is included to address the uncertainty in the ratio of the predicted to the measured inelastic cross-section within the fiducial volume defined by $M_X > 13 \text{ GeV}$, where M_X is the mass of the non-diffractive hadronic system [110].

Trigger

Systematic uncertainties on the efficiency of the electron or muon triggers are evaluated using the tag-and-probe method, applied to both the signal and background MC samples. Efficiencies are obtained using the `ElectronEfficiencyCorrection` and `MuonEfficiencyCorrections` tools. The uncertainty in the E_T^{miss} trigger is derived from the scale factor estimation, incorporating statistical contributions and efficiency discrepancies between MC samples, specifically $t\bar{t}$ and $W + \text{jets}$ backgrounds [111]. Upon evaluation, trigger uncertainties were found to be less than 1% and were subsequently excluded from the final fit results due to their minimal impact.

Muons and electrons

The following systematic uncertainties are applied to electrons and muons for simulation-based estimations:

- **Identification and reconstruction efficiencies:** Measured using the tag-and-probe method centered around the Z mass peak.
- **Isolation efficiency:** The scale factor and its uncertainty are derived via the tag-and-probe method, also utilizing the Z mass peak.

- **Energy and Momentum scales:** Determined through the Z mass line shape analysis, with contributions from the CP groups.
- **Track-to-vertex association efficiency:** This applies solely to muons.

Missing transverse energy

The missing transverse energy (E_T^{miss}) is calculated using physics objects, as outlined in Section 5.4. Systematic uncertainties in the reconstructed components, such as the jet energy scale, directly affect the E_T^{miss} and represent its primary sources of uncertainty. Additionally, the uncertainty known as the “Soft Term,” which arises from tracks in the inner detector not associated with any reconstructed objects, is also considered. The resolution and scale of the Soft Term are varied within their respective uncertainties to assess their impact on the overall E_T^{miss} uncertainty, utilizing the `METUtilities` tool.

Small- R Jet Energy Scale and Resolution Uncertainty

The jet energy scale (JES) and resolution (JER) for small- R jets are determined by comparing the response in MC samples to data across various bins in kinematic phase space. The analysis utilizes the configuration `R4_CategoryReduction_SimpleJER`, which incorporates 30 JES and 8 JER uncertainty components. These uncertainties are particularly significant in the boosted analysis due to their impact on the calculation of E_T^{miss} . The efficiency uncertainty of the Jet Vertex Tagger (JVT) is evaluated; however, since its effect was found to be below 1%, it was excluded from the final fit.

Large- R Jet Energy Scale and Resolution Uncertainty

The uncertainties for the large- R jet energy scale are incorporated according to the prescription provided by the jet substructure group. The uncertainty related to the jet p_T scale is assessed through the Rtrk method, which involves comparing the ratio of jet p_T to track-jet p_T in dijet data versus simulation. Beyond the “baseline” uncertainty, additional considerations include uncertainties related to track measurements (“Tracking”), variations between Pythia and Sherpa dijet simulations (“Modelling”), and the statistical uncertainty in dijet data (“TotalStat”).

B-tagging systematics

Systematic uncertainties related to b -tagging arise from scaling factors designed to correct discrepancies in b -tagging efficiency between data and MC simulations. Separate scale factors and their respective systematic

uncertainties are determined for jets originating from b -quarks, c -quarks, and light-flavor quarks, based on various measurements.

9.1.2 W/Z -tagging efficiency SF Uncertainty

For the uncertainties associated with the boson tagger’s background efficiency, we consider both the large- R jet-related uncertainties and the modeling uncertainties from multijet and γ +jets processes. The modeling uncertainties are estimated by comparing the nominal Pythia 8 and Sherpa samples for multijet and γ +jets, respectively, against their alternative samples. For more details on how the tagger is defined, see Section 6.5.1.

Systematic uncertainties associated with the scale factors, which assess the boson tagger’s relative performance in data versus MC, are thoroughly evaluated. These uncertainties include various aspects: for background, considerations include matrix element variations, hadronization, radiation effects, and the impacts from dijets or γ +jets events. For the signal, uncertainties include considerations like extrapolation at high p_T .

Plots illustrating the impact of uncertainties on the Data/MC ratio for the 1-lepton channel are presented in Figure 9.1.

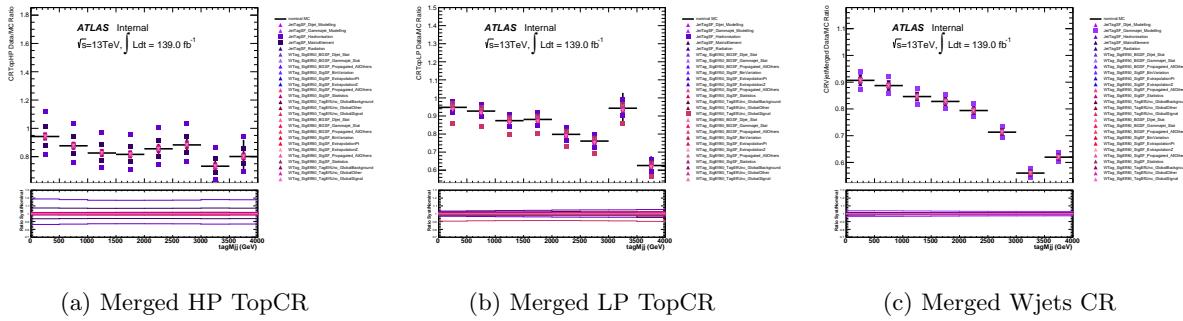


Figure 9.1: Comparison of boson tagging scale factors in the 1-lepton channel.

9.1.3 Quark/Gluon jets uncertainty

Jet flavor response and composition uncertainties consider the different responses of quark- and gluon-initiated jets. The response uncertainties are centrally derived from dijet events, using alternative MC samples, specifically Pythia 8 and Herwig++. For flavor composition, uncertainties are typically assumed based on a presumed 50/50 quark/gluon mix, with a conservative approach that assumes 100% uncertainty. However, in VBS topologies, which tend to be quark-enriched, these uncertainties can limit measurement sensitivity. To mitigate this, we analyze the gluon fraction within our analysis phase-space. This analysis

allows us to rederive jet flavor-related uncertainties using custom gluon fractions, potentially reducing these uncertainties.

The estimation of the gluon fraction across various analysis regions and samples is performed as a function of the small- R jet p_T and η . This is achieved by utilizing truth parton label information in MC samples to discern the proportions of quarks and gluons. For the purpose of quark estimation, all jets except those initiated by b-quarks are considered in the denominator.

Two-dimensional histograms, representing the gluon fraction as a function of jet p_T and η , serve as inputs for recalculating jet flavor and composition uncertainties. Each MC sample is associated with a single input file. The gluon fraction for a specific (p_T, η) bin is determined by aggregating data across all regions.

To account for the uncertainty in the gluon fraction, additional inputs are utilized. The uncertainty for a particular (p_T, η) bin, σ_{gfrac} , is calculated as follows:

$$\sigma_{gfrac} = \sqrt{\sigma_{region}^2 + \sigma_{gen}^2}$$

Here, σ_{region} represents the maximal deviation in gluon fraction between the nominal and any analysis region. Meanwhile, σ_{gen} denotes the generator uncertainty, obtained from comparing alternative Pythia 8 and Herwig++ MC samples.

The inputs for the gluon fraction and their corresponding uncertainties in the 1-lepton channel are illustrated in Figures 9.2 and 9.3, respectively. Comparison between old flavor uncertainties and newly derived flavor uncertainties is shown in Figure 9.4. It is evident from this comparison that flavor composition uncertainties, in general, are reduced.

9.1.4 Tracks uncertainties

We incorporate dedicated uncertainties related to track reconstruction, particularly for tracks associated with small- R jets. The following procedure [112] is used to derive these track multiplicity-related uncertainties, which comprise five main components:

- fake track efficiency: the uncertainty from the track mis-reconstruction rate;
- track efficiency: the uncertainty from the track reconstruction efficiency;
- experimental: the uncertainty in charged particle multiplicity, derived as in [113];
- PDF: the theory uncertainty from the parton distribution function;
- matrix element: the theory uncertainty from the matrix element calculation;

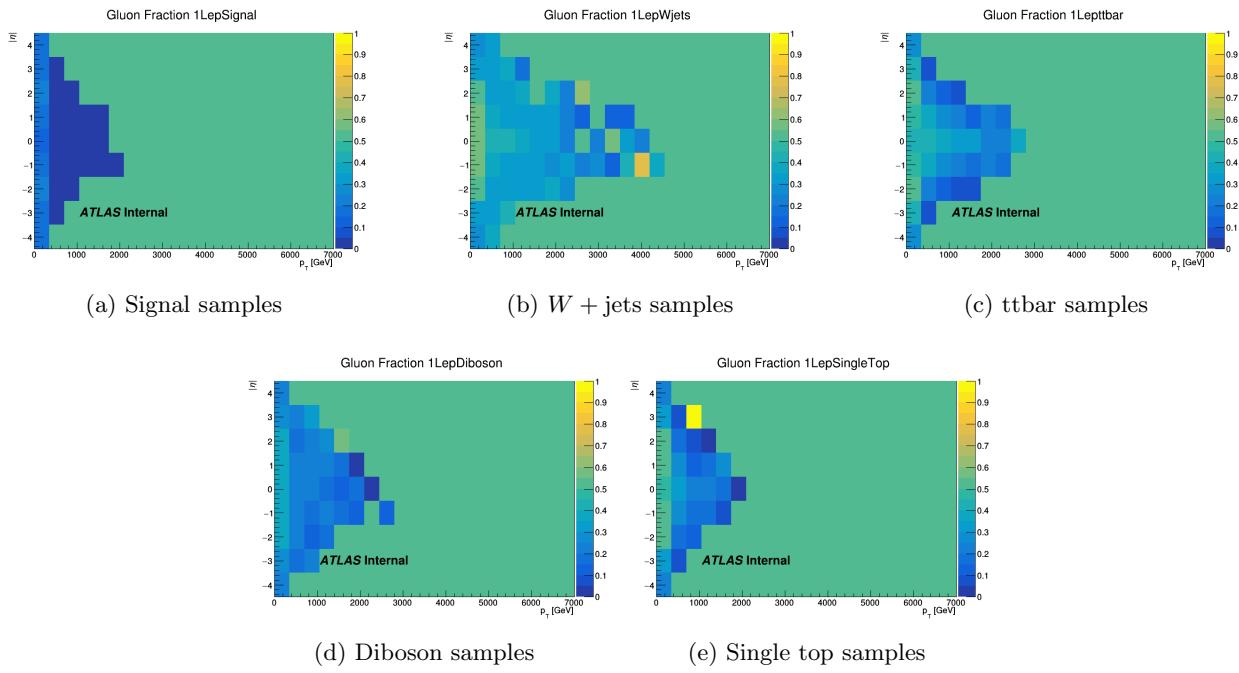


Figure 9.2: Gluon fraction inputs used to calculate jet uncertainties for the 1 lepton channel.

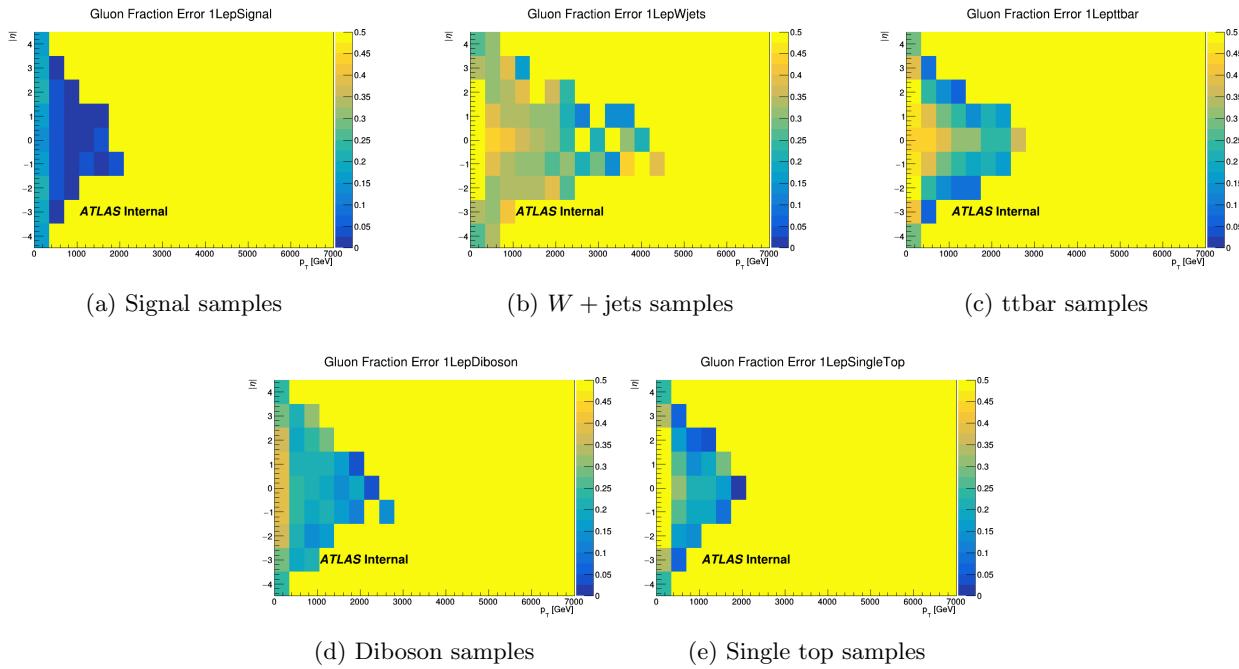


Figure 9.3: Gluon fraction error inputs used to calculate jet uncertainties for the 1 lepton channel.

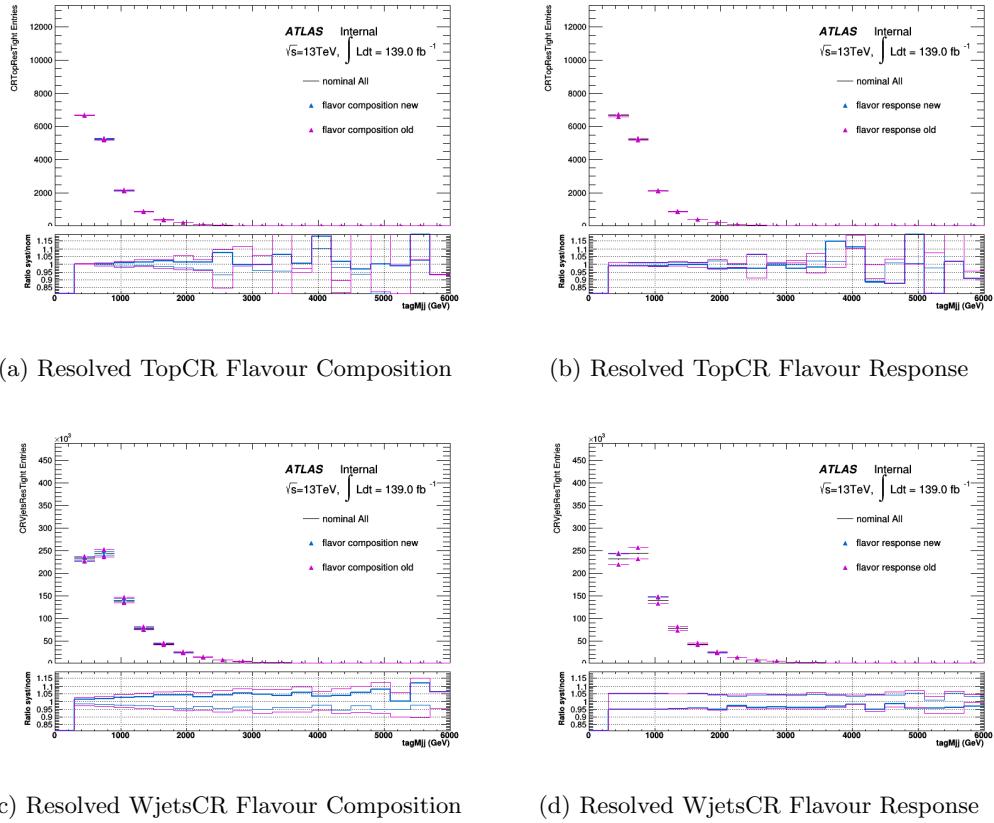


Figure 9.4: Comparison between old flavor uncertainties and newly derived flavor uncertainties.

The impacts from these uncertainties are expected to be minimal.

9.2 Background Uncertainties

Several systematic uncertainties associated with the modeling of backgrounds have been assessed. These uncertainties include a shape systematic related to the discriminant used in the analysis and a normalization systematic for predictions derived solely from simulations. The complete list of the alternative samples utilized to derive these background uncertainties is provided in the Tables 9.4-9.10.

Table 9.4: $W \rightarrow e\nu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR} [\text{pb}]$	k-factor	ϵ_{filter}
363600	MGPy8EG_N30NLO_Wenu_Ht0_70_CVetoBVeto	16719.0	1.12	8.38E+03
363601	MGPy8EG_N30NLO_Wenu_Ht0_70_CFilterBVeto	16720.0	1.12	1.38E+03
363602	MGPy8EG_N30NLO_Wenu_Ht0_70_BFilter	16717.0	1.12	2.42E+02
363603	MGPy8EG_N30NLO_Wenu_Ht70_140_CVetoBVeto	755.1	1.12	7.12E+03
363604	MGPy8EG_N30NLO_Wenu_Ht70_140_CFilterBVeto	755.77	1.12	2.40E+03
363605	MGPy8EG_N30NLO_Wenu_Ht70_140_BFilter	755.73	1.12	4.83E+02
363606	MGPy8EG_N30NLO_Wenu_Ht140_280_CVetoBVeto	318.96	1.12	6.66E+03
363607	MGPy8EG_N30NLO_Wenu_Ht140_280_CFilterBVeto	319.93	1.12	2.64E+03
363608	MGPy8EG_N30NLO_Wenu_Ht140_280_BFilter	319.45	1.12	6.94E+02
363609	MGPy8EG_N30NLO_Wenu_Ht280_500_CVetoBVeto	73.528	1.12	6.19E+03
363610	MGPy8EG_N30NLO_Wenu_Ht280_500_CFilterBVeto	73.562	1.12	2.85E+03
363611	MGPy8EG_N30NLO_Wenu_Ht280_500_BFilter	73.556	1.12	9.52E+02
363612	MGPy8EG_N30NLO_Wenu_Ht500_700_CVetoBVeto	11.529	1.12	5.87E+03
363613	MGPy8EG_N30NLO_Wenu_Ht500_700_CFilterBVeto	11.517	1.12	2.98E+03
363614	MGPy8EG_N30NLO_Wenu_Ht500_700_BFilter	11.51	1.12	1.15E+03
363615	MGPy8EG_N30NLO_Wenu_Ht700_1000_CVetoBVeto	40.158	1.12	5.66E+03
363616	MGPy8EG_N30NLO_Wenu_Ht700_1000_CFilterBVeto	4.014	1.12	3.04E+03
363617	MGPy8EG_N30NLO_Wenu_Ht700_1000_BFilter	40.123	1.12	1.28E+03
363618	MGPy8EG_N30NLO_Wenu_Ht1000_2000_CVetoBVeto	13.243	1.12	5.48E+03
363619	MGPy8EG_N30NLO_Wenu_Ht1000_2000_CFilterBVeto	13.286	1.12	3.08E+03
363620	MGPy8EG_N30NLO_Wenu_Ht1000_2000_BFilter	1.326	1.12	1.43E+03
363621	MGPy8EG_N30NLO_Wenu_Ht2000_E.CMS.CVetoBVeto	0.042294	1.12	5.24E+03
363622	MGPy8EG_N30NLO_Wenu_Ht2000_E.CMS.CFilterBVeto	0.041884	1.12	3.17E+03
363623	MGPy8EG_N30NLO_Wenu_Ht2000_E.CMS.BFilter	0.047382	1.12	1.51E+03

Table 9.5: $W \rightarrow \mu\nu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR} [\text{pb}]$	k-factor	ϵ_{filter}
363624	MGPy8EG_N30NLO_Wmumu_Ht0_70_CVetoBVeto	16720.0	1.12	8.38E+03
363625	MGPy8EG_N30NLO_Wmumu_Ht0_70_CFilterBVeto	16717.0	1.12	1.38E+03
363626	MGPy8EG_N30NLO_Wmumu_Ht0_70_BFilter	16719.0	1.12	2.42E+02
363627	MGPy8EG_N30NLO_Wmumu_Ht70_140_CVetoBVeto	755.19	1.12	7.12E+03
363628	MGPy8EG_N30NLO_Wmumu_Ht70_140_CFilterBVeto	755.62	1.12	2.40E+03
363629	MGPy8EG_N30NLO_Wmumu_Ht70_140_BFilter	755.77	1.12	4.83E+02
363630	MGPy8EG_N30NLO_Wmumu_Ht140_280_CVetoBVeto	318.83	1.12	6.66E+03
363631	MGPy8EG_N30NLO_Wmumu_Ht140_280_CFilterBVeto	319.89	1.12	2.64E+03
363632	MGPy8EG_N30NLO_Wmumu_Ht140_280_BFilter	319.41	1.12	6.94E+02
363633	MGPy8EG_N30NLO_Wmumu_Ht280_500_CVetoBVeto	73.585	1.12	6.19E+03
363634	MGPy8EG_N30NLO_Wmumu_Ht280_500_CFilterBVeto	73.548	1.12	2.85E+03
363635	MGPy8EG_N30NLO_Wmumu_Ht280_500_BFilter	73.569	1.12	9.52E+02
363636	MGPy8EG_N30NLO_Wmumu_Ht500_700_CVetoBVeto	11.522	1.12	5.86E+03
363637	MGPy8EG_N30NLO_Wmumu_Ht500_700_CFilterBVeto	11.524	1.12	2.96E+03
363638	MGPy8EG_N30NLO_Wmumu_Ht500_700_BFilter	11.522	1.12	1.16E+03
363639	MGPy8EG_N30NLO_Wmumu_Ht700_1000_CVetoBVeto	40.194	1.12	5.67E+03
363640	MGPy8EG_N30NLO_Wmumu_Ht700_1000_CFilterBVeto	40.252	1.12	3.05E+03
363641	MGPy8EG_N30NLO_Wmumu_Ht700_1000_BFilter	40.139	1.12	1.28E+03
363642	MGPy8EG_N30NLO_Wmumu_Ht1000_2000_CVetoBVeto	13.262	1.12	5.45E+03
363643	MGPy8EG_N30NLO_Wmumu_Ht1000_2000_CFilterBVeto	13.215	1.12	3.08E+03
363644	MGPy8EG_N30NLO_Wmumu_Ht1000_2000_BFilter	13.287	1.12	1.44E+03
363645	MGPy8EG_N30NLO_Wmumu_Ht2000_E.CMS.CVetoBVeto	0.041697	1.12	5.20E+03
363646	MGPy8EG_N30NLO_Wmumu_Ht2000_E.CMS.CFilterBVeto	0.042199	1.12	3.17E+03
363647	MGPy8EG_N30NLO_Wmumu_Ht2000_E.CMS.BFilter	0.042049	1.12	1.59E+03

Table 9.6: $W \rightarrow \tau\nu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
363648	MGPy8EG_N30NLO_Wtaunu_Ht0_70_CVetoBVeto	16702.0	1.12	8.38E+03
363649	MGPy8EG_N30NLO_Wtaunu_Ht0_70_CFilterBVeto	16702.0	1.12	1.38E+03
363650	MGPy8EG_N30NLO_Wtaunu_Ht0_70_BFilter	16702.0	1.12	2.42E+02
363651	MGPy8EG_N30NLO_Wtaunu_Ht70_140_CVetoBVeto	754.91	1.12	7.12E+03
363652	MGPy8EG_N30NLO_Wtaunu_Ht70_140_CFilterBVeto	755.3	1.12	2.39E+03
363653	MGPy8EG_N30NLO_Wtaunu_Ht70_140_BFilter	755.14	1.12	4.85E+02
363654	MGPy8EG_N30NLO_Wtaunu_Ht140_280_CVetoBVeto	315.31	1.12	6.67E+03
363655	MGPy8EG_N30NLO_Wtaunu_Ht140_280_CFilterBVeto	316.05	1.12	2.64E+03
363656	MGPy8EG_N30NLO_Wtaunu_Ht140_280_BFilter	315.64	1.12	6.93E+02
363657	MGPy8EG_N30NLO_Wtaunu_Ht280_500_CVetoBVeto	72.275	1.12	6.22E+03
363658	MGPy8EG_N30NLO_Wtaunu_Ht280_500_CFilterBVeto	72.19	1.12	2.85E+03
363659	MGPy8EG_N30NLO_Wtaunu_Ht280_500_BFilter	72.226	1.12	9.52E+02
363660	MGPy8EG_N30NLO_Wtaunu_Ht500_700_CVetoBVeto	11.44	1.12	5.88E+03
363661	MGPy8EG_N30NLO_Wtaunu_Ht500_700_CFilterBVeto	11.463	1.12	2.97E+03
363662	MGPy8EG_N30NLO_Wtaunu_Ht500_700_BFilter	11.459	1.12	1.16E+03
363663	MGPy8EG_N30NLO_Wtaunu_Ht700_1000_CVetoBVeto	40.265	1.12	5.67E+03
363664	MGPy8EG_N30NLO_Wtaunu_Ht700_1000_CFilterBVeto	4.027	1.12	3.03E+03
363665	MGPy8EG_N30NLO_Wtaunu_Ht700_1000_BFilter	40.283	1.12	1.30E+03
363666	MGPy8EG_N30NLO_Wtaunu_Ht1000_2000_CVetoBVeto	13.264	1.12	5.46E+03
363667	MGPy8EG_N30NLO_Wtaunu_Ht1000_2000_CFilterBVeto	13.205	1.12	3.12E+03
363668	MGPy8EG_N30NLO_Wtaunu_Ht1000_2000_BFilter	13.222	1.12	1.43E+03
363669	MGPy8EG_N30NLO_Wtaunu_Ht2000_E_CMS_CVetoBVeto	0.043501	1.12	5.21E+03
363670	MGPy8EG_N30NLO_Wtaunu_Ht2000_E_CMS_CFilterBVeto	0.042721	1.12	3.18E+03
363671	MGPy8EG_N30NLO_Wtaunu_Ht2000_E_CMS_BFilter	0.042683	1.12	1.66E+03

Table 9.7: $Z \rightarrow ee + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
363147	MGPy8EG_N30NLO_Zee_Ht0_70_CVetoBVeto	1719.7	1.141	0.83292
363148	MGPy8EG_N30NLO_Zee_Ht0_70_CFilterBVeto	1719.4	1.141	0.10775
363149	MGPy8EG_N30NLO_Zee_Ht0_70_BFilter	1719.4	1.141	0.059156
363150	MGPy8EG_N30NLO_Zee_Ht70_140_CVetoBVeto	85.105	1.141	0.71754
363151	MGPy8EG_N30NLO_Zee_Ht70_140_CFilterBVeto	85.041	1.141	0.17377
363152	MGPy8EG_N30NLO_Zee_Ht70_140_BFilter	85.175	1.141	0.10763
363153	MGPy8EG_N30NLO_Zee_Ht140_280_CVetoBVeto	36.005	1.141	0.67279
363154	MGPy8EG_N30NLO_Zee_Ht140_280_CFilterBVeto	36.028	1.141	0.19996
363155	MGPy8EG_N30NLO_Zee_Ht140_280_BFilter	36.06	1.141	0.12486
363156	MGPy8EG_N30NLO_Zee_Ht280_500_CVetoBVeto	82.054	1.141	0.62846
363157	MGPy8EG_N30NLO_Zee_Ht280_500_CFilterBVeto	82.126	1.141	0.22726
363158	MGPy8EG_N30NLO_Zee_Ht280_500_BFilter	82.474	1.141	0.14193
363159	MGPy8EG_N30NLO_Zee_Ht500_700_CVetoBVeto	12.733	1.141	0.5966
363160	MGPy8EG_N30NLO_Zee_Ht500_700_CFilterBVeto	1.273	1.141	0.24847
363161	MGPy8EG_N30NLO_Zee_Ht500_700_BFilter	12.722	1.141	0.15256
363162	MGPy8EG_N30NLO_Zee_Ht700_1000_CVetoBVeto	0.44546	1.141	0.57676
363163	MGPy8EG_N30NLO_Zee_Ht700_1000_CFilterBVeto	0.44611	1.141	0.26137
363164	MGPy8EG_N30NLO_Zee_Ht700_1000_BFilter	0.44603	1.141	0.16181
363165	MGPy8EG_N30NLO_Zee_Ht1000_2000_CVetoBVeto	0.15208	1.141	0.55543
363166	MGPy8EG_N30NLO_Zee_Ht1000_2000_CFilterBVeto	0.15248	1.141	0.27476
363167	MGPy8EG_N30NLO_Zee_Ht1000_2000_BFilter	0.15327	1.141	0.16618
363168	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_CVetoBVeto	0.0056989	1.141	0.53136
363169	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_CFilterBVeto	0.0057408	1.141	0.2923
363170	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_BFilter	0.0057164	1.141	0.17489

Table 9.8: $Z \rightarrow \mu\mu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
363123	MGPy8EG_N30NLO_Zmumu_Ht0_70_CVetoBVeto	1714.5	1.141	0.83157
363124	MGPy8EG_N30NLO_Zmumu_Ht0_70_CFfilterBVeto	1715.5	1.141	0.10835
363125	MGPy8EG_N30NLO_Zmumu_Ht0_70_BFilter	1715.7	1.141	0.059162
363126	MGPy8EG_N30NLO_Zmumu_Ht70_140_CVetoBVeto	84.57	1.141	0.71809
363127	MGPy8EG_N30NLO_Zmumu_Ht70_140_CFfilterBVeto	84.588	1.141	0.17404
363128	MGPy8EG_N30NLO_Zmumu_Ht70_140_BFilter	84.752	1.141	0.10804
363129	MGPy8EG_N30NLO_Zmumu_Ht140_280_CVetoBVeto	35.883	1.141	0.67432
363130	MGPy8EG_N30NLO_Zmumu_Ht140_280_CFfilterBVeto	35.908	1.141	0.19951
363131	MGPy8EG_N30NLO_Zmumu_Ht140_280_BFilter	35.887	1.141	0.12607
363132	MGPy8EG_N30NLO_Zmumu_Ht280_500_CVetoBVeto	81.871	1.141	0.62802
363133	MGPy8EG_N30NLO_Zmumu_Ht280_500_CFfilterBVeto	81.805	1.141	0.2282
363134	MGPy8EG_N30NLO_Zmumu_Ht280_500_BFilter	81.705	1.141	0.14263
363135	MGPy8EG_N30NLO_Zmumu_Ht500_700_CVetoBVeto	1.271	1.141	0.59722
363136	MGPy8EG_N30NLO_Zmumu_Ht500_700_CFfilterBVeto	12.672	1.141	0.24952
363137	MGPy8EG_N30NLO_Zmumu_Ht500_700_BFilter	12.699	1.141	0.15292
363138	MGPy8EG_N30NLO_Zmumu_Ht700_1000_CVetoBVeto	0.43601	1.141	0.57118
363139	MGPy8EG_N30NLO_Zmumu_Ht700_1000_CFfilterBVeto	0.44623	1.141	0.25948
363140	MGPy8EG_N30NLO_Zmumu_Ht700_1000_BFilter	0.44567	1.141	0.15965
363141	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_CVetoBVeto	0.14899	1.141	0.54908
363142	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_CFfilterBVeto	0.14625	1.141	0.27164
363143	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_BFilter	0.14705	1.141	0.17299
363144	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_CVetoBVeto	0.005538	1.141	0.56337
363145	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_CFfilterBVeto	0.0055466	1.141	0.29294
363146	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_BFilter	0.0056422	1.141	0.16307

Table 9.9: $Z \rightarrow \nu\nu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
361515	MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np0	7518.4	12.283	1.0
361516	MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np1	1200.1	12.283	1.0
361517	MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np2	387.16	12.283	1.0
361518	MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np3	110.08	12.283	1.0
361519	361519.MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np4	43.389	12.283	1.0

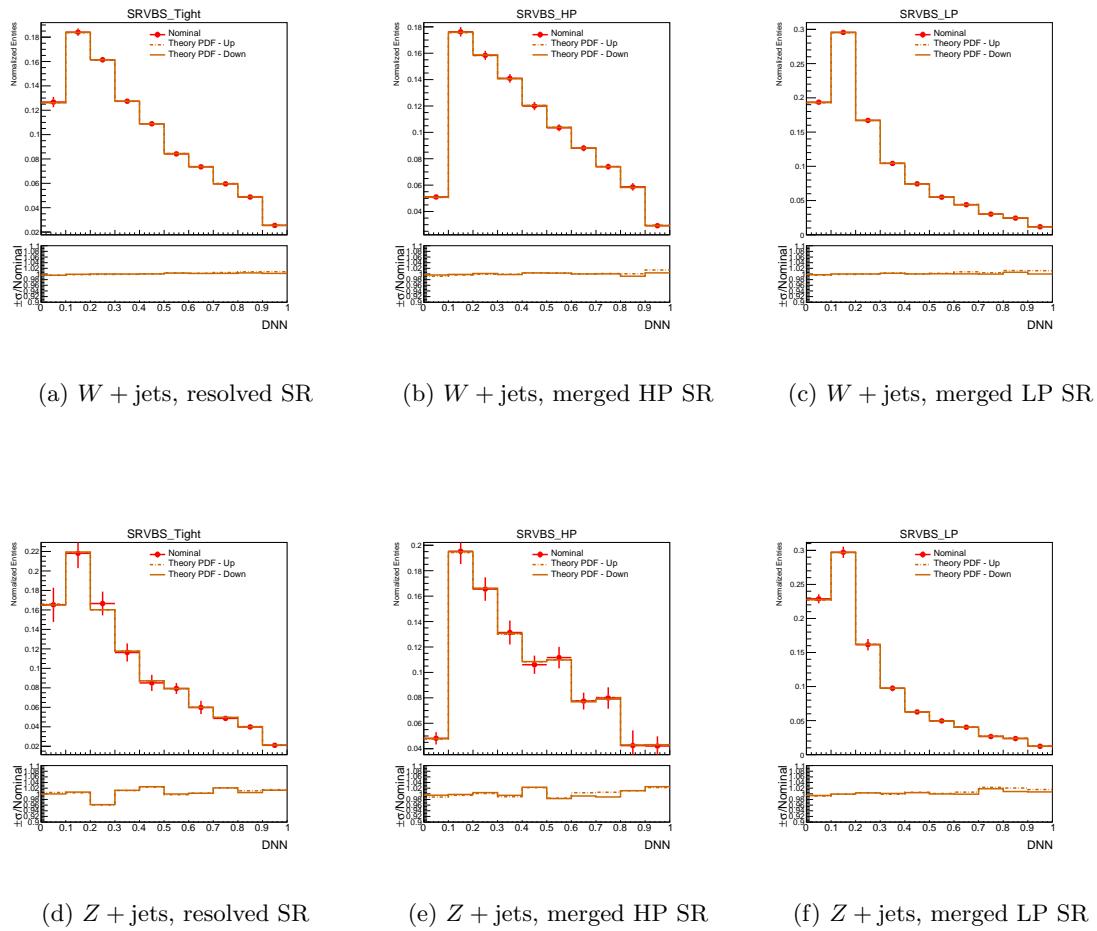
Table 9.10: $t\bar{t}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
410465	aMcAtNloPy8EvtGen_MEN30NLO_A14N23LO_ttbar_noShWe_dil	712.02	11.681	1.07E+03
410464	aMcAtNloPy8EvtGen_MEN30NLO_A14N23LO_ttbar_noShWe_SingleLep	711.43	11.691	4.40E+03
410467	PhPy8EG_A14_ttbar_hdamp517p5_nonallhad_AK10J160	729.74	1.0	1.56E+03
410468	PowhegHerwig7EG_H7UE_tt_hdamp258p75_704_nonallhad_AK10J160	730.15	1.0	1.50E+03
410469	aMcAtNloPy8EG_MEN30NLO_A14N23LO_ttbar_nonallhad_AK10J160	711.41	1.0	1.42E+03
410480	PhPy8EG_A14_ttbar_hdamp517p5_SingleLep	729.74	1.0	4.39E+03
410482	PhPy8EG_A14_ttbar_hdamp517p5_dil	729.74	1.0	1.05E+03
410557	PowhegHerwig7EvtGen_H7UE_tt_hdamp258p75_704_SingleLep	730.14	1.0	4.39E+03
410558	PowhegHerwig7EvtGen_H7UE_tt_hdamp258p75_704_dil	730.15	1.0	1.05E+03

9.2.1 PDF and Scale uncertainties

The uncertainties related to Parton Distribution Functions (PDFs), which include alternative PDF sets and variations in the strong coupling constant α_s , are addressed following the guidelines from the Physics Modelling Group (PMG). An uncertainty envelope, encompassing these variations, is designated as the PDF uncertainty. Examples involving alternative PDFs are presented in Figure 9.5, and the impacts of combined NNPDF and α_s uncertainties are shown in Figure 9.6. The uncertainty for each bin of the considered distributions is determined by calculating the standard deviation of the mean values from the 100 MC replicas of the NNPDF set. For a detailed introduction to NNPDF, see Section 4.2.

The scale uncertainty associated with the QCD renormalization (μ_r) and factorization (μ_f) scales is determined by the envelope of variations. These variations arise from independently scaling μ_r and μ_f by factors of 2, 1, or 0.5, excluding the extreme cases where μ_r is multiplied by 2 and μ_f by 0.5 simultaneously, and vice versa. The combined results of these scale variations are shown in Figure 9.7.



(a) $W + \text{jets}$, resolved SR (b) $W + \text{jets}$, merged HP SR (c) $W + \text{jets}$, merged LP SR

(d) $Z + \text{jets}$, resolved SR (e) $Z + \text{jets}$, merged HP SR (f) $Z + \text{jets}$, merged LP SR

Figure 9.5: Alternative PDF uncertainties in the 1-lepton channel. Histograms are normalized.

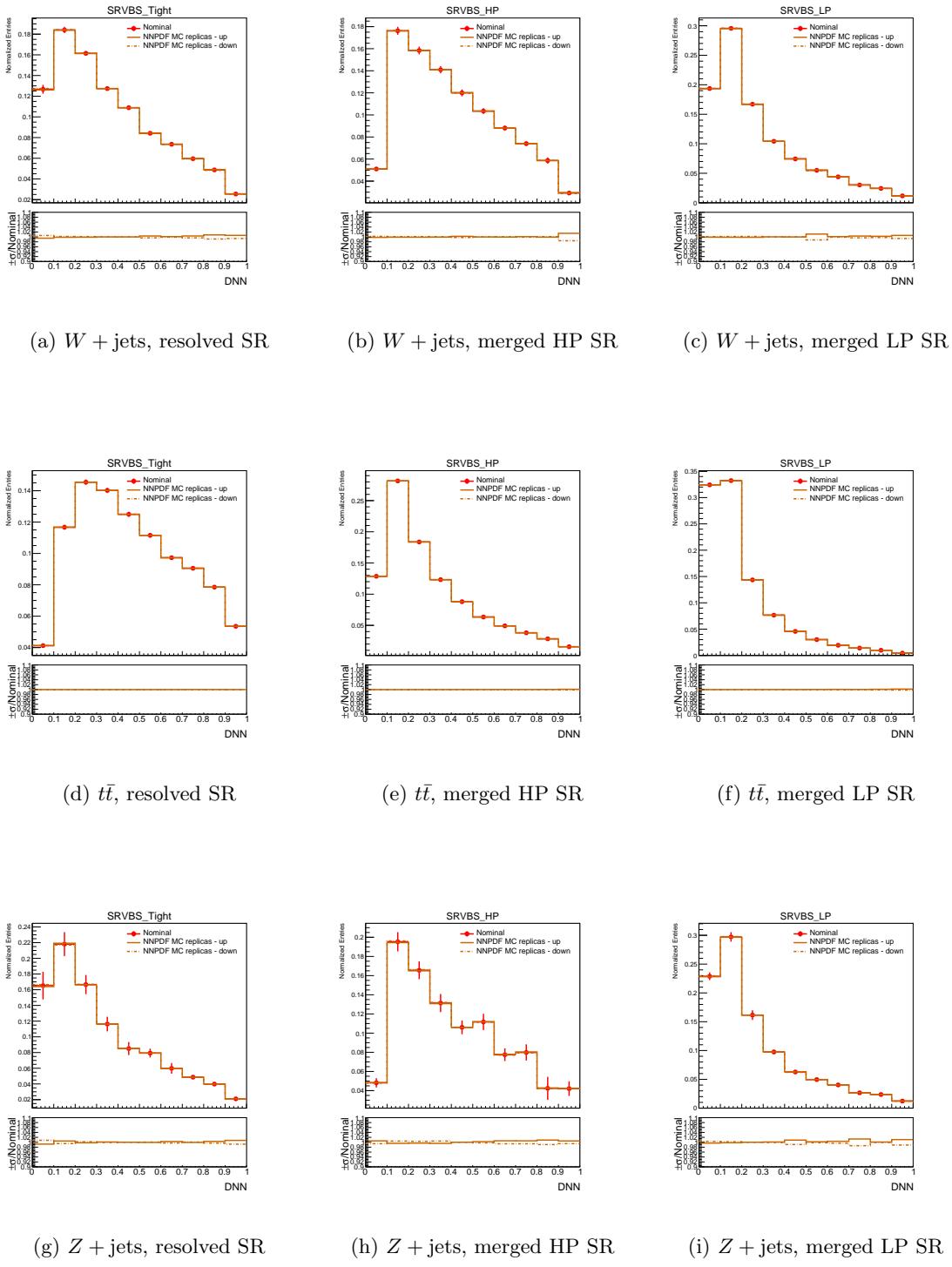


Figure 9.6: PDF uncertainties using the NNPDF set in the 1-lepton channel. Histograms are normalized.

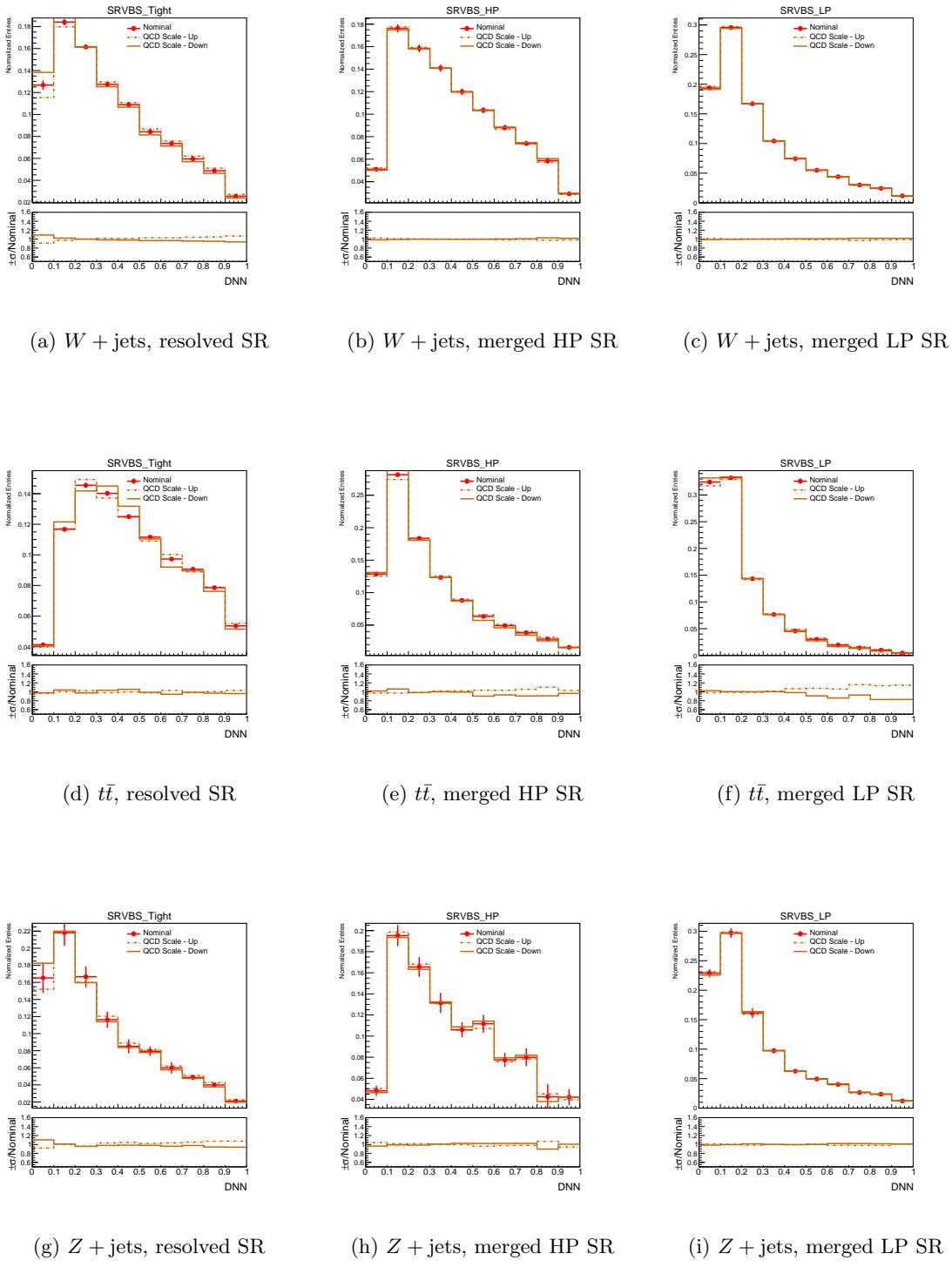


Figure 9.7: QCD scale uncertainties in the 1-lepton channel. Histograms are normalized.

9.2.2 ISR/FSR uncertainties

The impact of initial and final state radiation (ISR/FSR) on top quark backgrounds is investigated using dedicated samples that vary the amount of additional radiation. Figures 9.8 and 9.9 illustrate the effects of these ISR and FSR uncertainties in the 1-lepton channel.

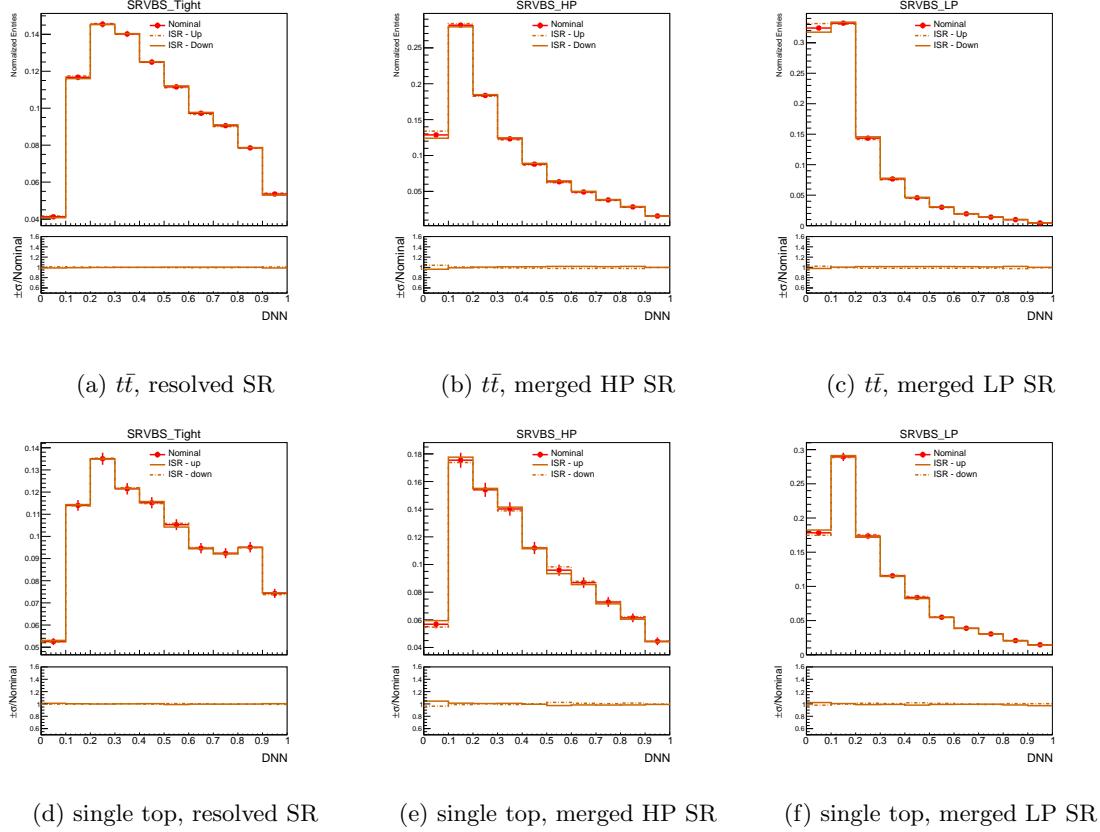


Figure 9.8: ISR uncertainties in the 1-lepton channel.

9.2.3 Tag Jets re-weighting

As outlined in Section 7.1, a reweighting procedure is applied to the $V + \text{jets}$ samples based on the m_{jj}^{tag} variable. An uncertainty is assigned to this reweighting by assuming a 100% uncertainty in the parameters of the linear fit. This is treated as a modeling systematic uncertainty for both $W + \text{jets}$ and $Z + \text{jets}$ backgrounds.

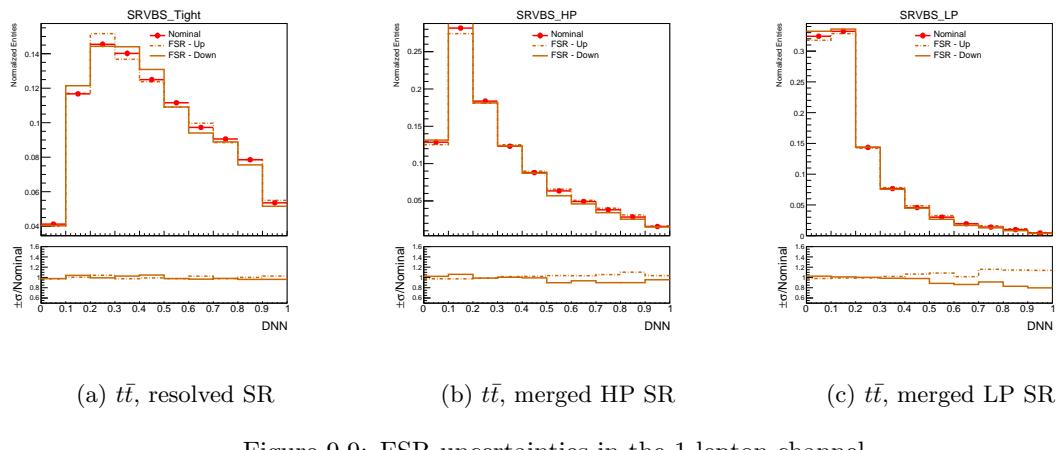


Figure 9.9: FSR uncertainties in the 1-lepton channel.

9.3 Signal Uncertainties

9.3.1 PDF and Scale uncertainties

Similar to the treatment of backgrounds, additional systematic uncertainties arise for signals due to modeling differences among various signal MC generators. These uncertainties include PDF uncertainties and QCD scale uncertainties for signal samples generated using Next-to-Leading Order (NLO) QCD. These uncertainties, which affect signal acceptance, are analyzed using truth weight information stored in the current analysis framework.

For the scale uncertainty, we adopt the envelope method employing the usual variations, akin to the approach utilized for background samples, as detailed in Section 9.2.1. Examples of these uncertainties are illustrated in Figures 9.10 through 9.11.

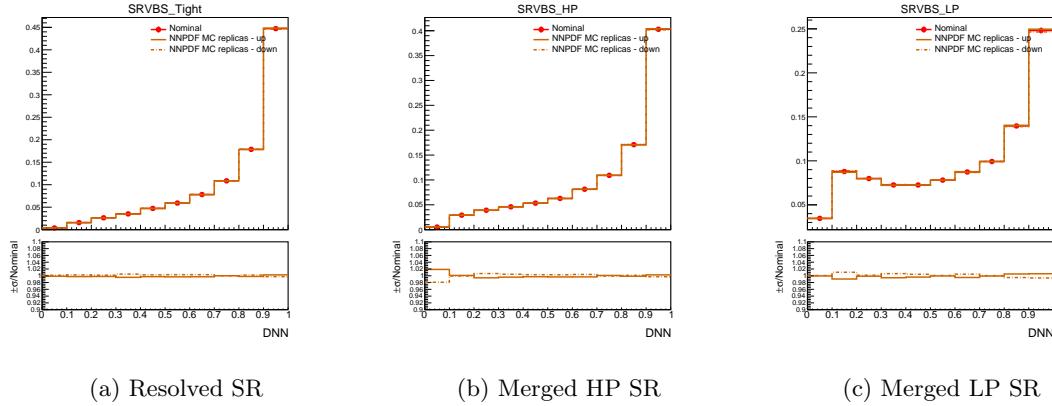


Figure 9.10: PDF uncertainties of the signal MC sample in the 1-lepton channel. Histograms are normalized.

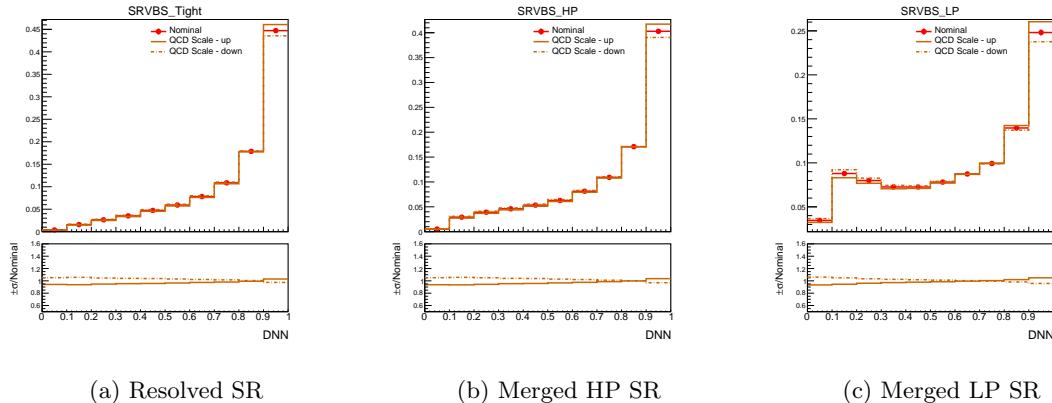


Figure 9.11: QCD scale uncertainties of the signal MC sample in the 1-lepton channel. Histograms are normalized.

9.3.2 EWK-QCD interference and Shower uncertainty

Due to time constraints and the extensive scope of this study, certain aspects could not be comprehensively addressed, including the potential effects of Electroweak-Quantum Chromodynamics (EWK-QCD) interference and the impact of parton showering on the signal MC samples. Notably, the EWK and QCD vector boson plus jets ($VV+jj$) samples were generated separately, implying that no interference contributions were considered in the analysis. While these factors are acknowledged for their potential minor impact on the results, a detailed investigation was beyond the scope of this thesis. Alternative MC samples necessary for assessing these uncertainties were not available when my DNN study concluded. Future studies could explore these aspects to enhance the understanding of the systematics involved.

Chapter 10

Statistical Interpretation

In this chapter, we outline our fitting approach and present the main results of the fit. Our goal is to measure the signal strength (μ_{VBS}) of the EW $VV+jj$ production within the VBS-enhanced phase space for the 1-lepton channel, using a DNN method. Detailed discussions will be provided in Section 10.4.

The fit model is constructed using MC-based templates for both the signal and background processes. In addition to estimating the primary parameters of interest (POIs), the model aims to constrain the normalization of the predominant background processes, specifically $W + \text{jets}$ and $t\bar{t}$ production, using dedicated control CRs established in earlier chapters. Minor backgrounds, such as VV and single-top processes, are adjusted using priors, with specific extrapolation parameters derived for them. The model incorporates the main sources of systematic uncertainties, applying normalization and/or shape effect schemes as appropriate.

10.1 Introduction

To determine the signal strength parameter μ (μ_{VBS}), a binned maximum likelihood fit is used in the statistical analysis. The likelihood function is formulated as follows:

$$\mathcal{L}(N, \tilde{\theta} | \mu, \theta) = \text{Pois}(\mu | \mu s + b) \cdot p(\tilde{\theta} | \theta) \quad (10.1)$$

Here, $\text{Pois}(\mu | \mu s + b)$ represents the product of Poisson probability terms across all histogram bins:

$$\text{Pois}(\mu | \mu s + b) = \prod_{i=1}^{N_{\text{bins}}} \frac{(\mu s_i(\theta) + b_i(\theta))^{N_i} e^{-(\mu s_i(\theta) + b_i(\theta))}}{N_i!} \quad (10.2)$$

In this formulation, μs_i and b_i denote the expected numbers of signal and background events in bin i , respectively, while N_i represents the number of observed events in that bin. The second part of Equation 10.1, $p(\tilde{\theta} | \theta)$, typically referred to as the prior, incorporates our knowledge about the systematic effects considered in the analysis. This binned likelihood approach allows for a robust estimation of μ while accounting for various statistical and systematic uncertainties.

The sensitivity of the signal and background predictions to systematic uncertainties is quantified through nuisance parameters (NPs), denoted as θ . These NPs are typically modeled using either Gaussian or log-normal distributions, with log-normal priors being preferred for normalization uncertainties to ensure that the likelihood remains positive. The expected counts of signal and background events in each bin are modeled as functions of θ . This modeling is structured so that event rates across different categories exhibit a log-normal behavior when θ is governed by normal distributions.

Priors are used to constrain the NPs towards their nominal values within their assigned uncertainties. This is achieved by incorporating penalty terms or auxiliary measurements into the likelihood function. These additions cause the likelihood to increase whenever an NP deviates significantly from its nominal value. Consequently, the likelihood function $\mathcal{L}(\mu, \theta)$ depends on both the signal strength parameter μ and the NPs θ . This setup facilitates the integration of our understanding and uncertainties about the system into the analysis, ensuring that the estimation of μ is consistent with the prior knowledge encapsulated in θ .

The nominal fit result, in terms of the signal strength parameter μ and its uncertainty σ_μ , is determined by maximizing the likelihood function across all parameters. This process yields the maximized log-likelihood value (MLL).

10.2 Fit inputs and options

For the statistical interpretation, we utilize outputs from the CxAODReader framework, which provides histograms of relevant distributions within designated regions. These histograms are then supplied to the WSMaker statistical framework for this analysis.

In the fitting stage of the analysis, the relevant variables include the m_{jj}^{tag} and the final DNN score distributions. Table 10.1 summarizes the regions and variables used in the fit of the analysis. These regions are simultaneously fitted using a binned profile likelihood method to extract the parameter of interest (POI), which, for the Standard Model measurement, is the signal strength of the EW VBS production.

The control regions employ bins of m_{jj}^{tag} with boundaries defined at $\{400, 600, 800, 1000, 1200, 1600, 6000\}$ GeV, where the final bin contains all overflow. For the DNN, binning is based on the transformation D , as detailed in reference [114], with a parameterization specified as $Z(\text{signal}, \text{background}) = (10, 5)$.

Regions	1-lepton channel fit model		
	Merged high-purity	Merged low-purity	Resolved
SR	DNN	DNN	DNN
WCR		m_{jj}^{tag}	m_{jj}^{tag}
TopCR	One bin	One bin	m_{jj}^{tag}

Table 10.1: Summary of the regions from 1-lepton channel entering the likelihood of the fit models. “One bin” implies that a single bin without any shape information is used in the corresponding fit region.

There are two types of NPs: those arising from experimental uncertainties and those associated with MC modeling uncertainties. A detailed discussion of these uncertainties is provided in Chapter 9.

For the experimental uncertainties, various methods to vary particle reconstruction have been developed by the ATLAS collaboration, which introduce approximately a hundred NPs. These parameters are governed by Gaussian priors, which serve to decrease the likelihood whenever the fitted values deviate from their nominal values.

For background MC modeling uncertainties, we consider the main background samples such as $t\bar{t}$, $W+\text{jets}$, $Z+\text{jets}$, and QCD-produced diboson. Additional NPs are obtained by comparing the default generator with an alternative. As shown in Table 10.2, background normalizations are left floating, except for the single-top background normalization, which is constrained with a 30% Gaussian prior.

For the signal, MC modeling uncertainties arise from variations in the scale (α_S), PDF, alternative PDF sets, and radiation parameters within the PYTHIA simulation, as mentioned in Section 9.3.

Treatment	NP name	NP name simplified
Float	ATLAS_norm_WjetsMerged	W norm. Merged
Float	ATLAS_norm_WjetsResolved	W norm. Resolved
Float	ATLAS_norm_ZjetsMerged	Z norm. Merged
Float	ATLAS_norm_ZjetsResolved	Z norm. Resolved
Float	ATLAS_norm_ttbarMerged	$t\bar{t}$ norm. Merged
Float	ATLAS_norm_ttbarResolved	$t\bar{t}$ norm. Resolved
Floating	mu_SemileptonicVBS	Signal strength μ (VBS)

Table 10.2: Floating parameter (includes signal strength parameter). The normalization factors for $Z + \text{jets}$ have minimal impact due to their small contribution in the 1-lepton channel.

10.3 Fit Model Inspection

This section presents the results of the model inspection for the 1-lepton channel.

10.3.1 Prefit Plots and Tables

Figure 10.1 presents the pre-fit plots for the analysis regions included in the fit model. It features the distributions of DNN scores in the SRs, and the m_{jj}^{tag} distributions in the $W + \text{jets}$ and top control regions (CRs). Initially, only some bins on the left side displayed real data in the SRs, which have now been unblinded. Tables 10.3-10.4 show the expected yields for signal and background processes in all the control and signal regions for the 1-lepton channel.

DNN_SRVBShp		DNN_SRVBslp	
W	4656.33 ± 599.26	W	13507.48 ± 1306.67
Z	155.86 ± 29.63	Z	433.51 ± 75.10
Diboson	564.00 ± 208.79	Diboson	759.72 ± 276.17
stop	720.09 ± 249.77	stop	974.63 ± 316.11
ttbar	7724.82 ± 1531.65	ttbar	10758.23 ± 1251.68
Bkg	13821.09 ± 2012.72	Bkg	26433.56 ± 2560.55
EW6lvqq	228.59 ± 41.37	EW6lvqq	180.21 ± 37.21
Signal	228.59 ± 41.37	Signal	180.21 ± 37.21
SignalExpected	228.59 ± 41.37	SignalExpected	180.21 ± 37.21
S/B	$1.65\text{e-}02$	S/B	$6.82\text{e-}03$
S/sqrt(S+B)	$1.93\text{e+}00$	S/sqrt(S+B)	$1.10\text{e+}00$
data	12178	data	22158

DNN_SRVBsRes	
W	60551.72 ± 5916.55
Z	2114.42 ± 425.26
Diboson	1723.35 ± 582.92
stop	1750.76 ± 541.16
ttbar	7330.60 ± 299.18
Bkg	73470.84 ± 6611.96
EW6lvqq	662.01 ± 56.52
Signal	662.01 ± 56.52
SignalExpected	662.01 ± 56.52
S/B	$9.01\text{e-}03$
S/sqrt(S+B)	$2.43\text{e+}00$
data	71272

Table 10.3: Prefit event yields for the analysis SRs in the 1 lepton channel.

tagMjj_CRTopHP		tagMjj_CRTopLP	
W	352.82 ± 43.34	W	949.62 ± 77.62
Z	16.22 ± 3.13	Z	43.10 ± 7.66
Diboson	41.79 ± 15.53	Diboson	65.63 ± 22.52
stop	1342.33 ± 456.84	stop	1426.00 ± 466.82
ttbar	12267.00 ± 2374.62	ttbar	16508.02 ± 1970.36
Bkg	14020.16 ± 2635.77	Bkg	18992.37 ± 2239.62
EW6lvqq	72.15 ± 13.03	EW6lvqq	59.17 ± 9.99
Signal	72.15 ± 13.03	Signal	59.17 ± 9.99
SignalExpected	72.15 ± 13.03	SignalExpected	59.17 ± 9.99
S/B	$5.15e-03$	S/B	$3.12e-03$
S/sqrt(S+B)	$6.08e-01$	S/sqrt(S+B)	$4.29e-01$
data	12195	data	17195
tagMjj_CRTopRes		tagMjj_CRWjetMerged	
W	2092.27 ± 102.60	W	27122.54 ± 3139.15
Z	96.04 ± 18.42	Z	937.92 ± 172.50
Diboson	84.24 ± 28.43	Diboson	1050.98 ± 363.58
stop	2132.59 ± 646.43	stop	1407.69 ± 442.73
ttbar	11368.00 ± 303.73	ttbar	13290.86 ± 1368.18
Bkg	15773.14 ± 776.66	Bkg	43809.98 ± 4299.96
EW6lvqq	138.07 ± 6.64	EW6lvqq	106.15 ± 11.62
Signal	138.07 ± 6.64	Signal	106.15 ± 11.62
SignalExpected	138.07 ± 6.64	SignalExpected	106.15 ± 11.62
S/B	$8.75e-03$	S/B	$2.42e-03$
S/sqrt(S+B)	$1.09e+00$	S/sqrt(S+B)	$5.07e-01$
data	16137	data	38486
tagMjj_CRWjetRes			
W	527592.75 ± 83268.24	W	527592.75 ± 83268.24
Z	20700.18 ± 6182.92	Z	20700.18 ± 6182.92
Diboson	15641.74 ± 5369.28	Diboson	15641.74 ± 5369.28
stop	33952.01 ± 10877.37	stop	33952.01 ± 10877.37
ttbar	226711.24 ± 18682.68	ttbar	226711.24 ± 18682.68
Bkg	824597.92 ± 112159.27	Bkg	824597.92 ± 112159.27
EW6lvqq	1790.20 ± 151.38	EW6lvqq	1790.20 ± 151.38
Signal	1790.20 ± 151.38	Signal	1790.20 ± 151.38
SignalExpected	1790.20 ± 151.38	SignalExpected	1790.20 ± 151.38
S/B	$2.17e-03$	S/B	$2.17e-03$
S/sqrt(S+B)	$1.97e+00$	S/sqrt(S+B)	$1.97e+00$
data	801406	data	801406

Table 10.4: Prefit event yields for the analysis CRs in the 1 lepton channel.

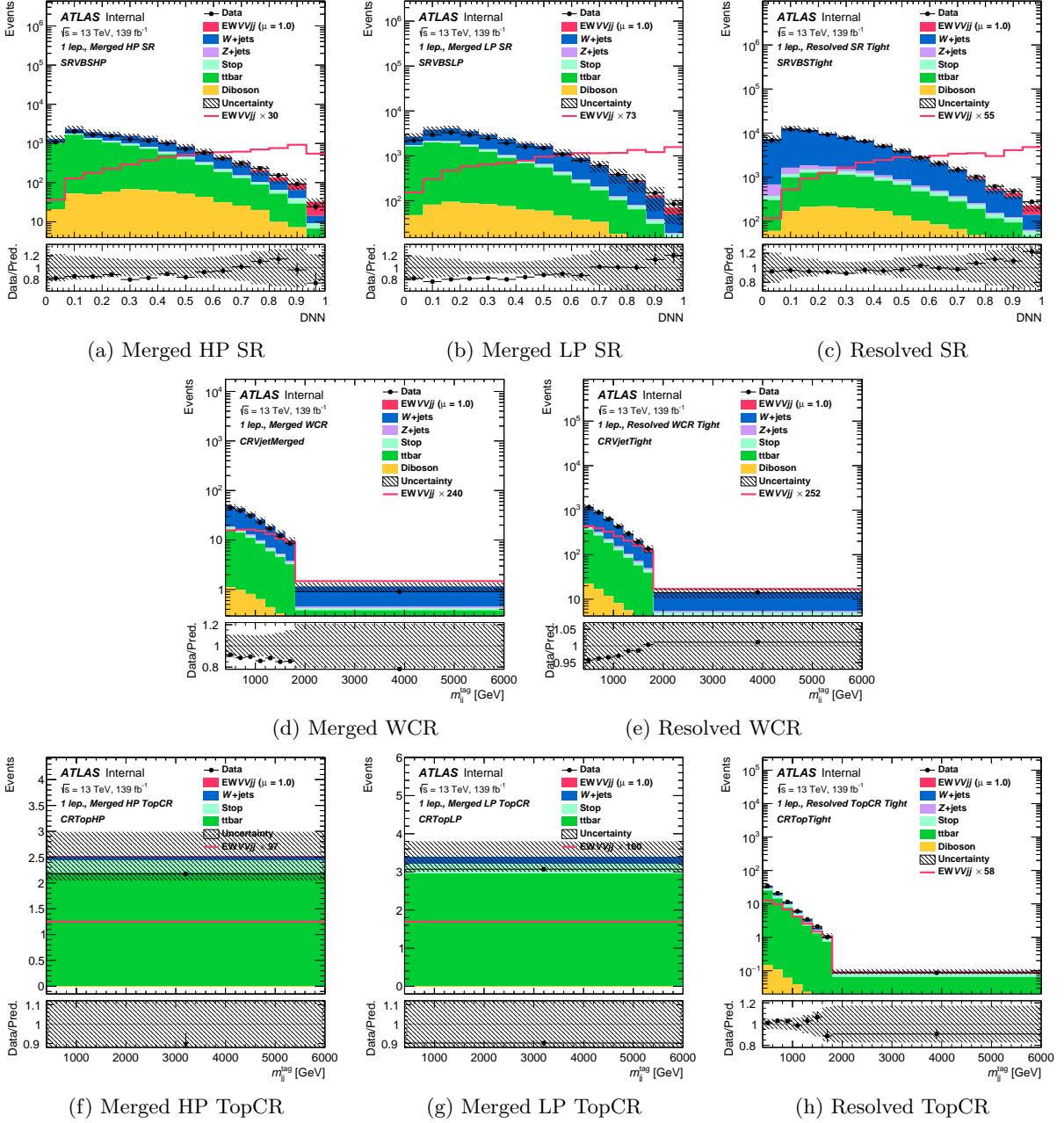


Figure 10.1: Prefit plots for the 1 lepton channel. Data has been unblinded

10.3.2 Asimov Fit Results

A fit to the full SRs using Asimov data is performed to identify constraints within the fit model. Figures 10.2, 10.3, and 10.4 show the pulls, correlations, and rankings, respectively, of the NPs used in the fit.

In the context of the pull plot, the black boxes represent an approximate measure of how well the fitted pull aligns with the postfit constraint for a given NP. A smaller pull that appears less constrained is generally less significant than a larger pull with a tight constraint. This concept is further elaborated in the reference [115], which provides detailed insights into the statistical interpretation of these pulls and constraints in the analysis.

A brief commentary on the identified relevant constraints is provided below:

- `SysMJJREWEIGHT_100per_W+jets` (L1_Fat1 for merged and L1_J2 for resolved) represents uncertainties related to the reweighting of m_{jj}^{tag} in both merged and resolved regions. These constraints are expected due to the application of 100% uncertainties to the m_{jj}^{tag} reweighting, allowing the control regions to effectively constrain these systematic uncertainties (see Section 9.2.3).
- `SysMODEL_Wjets_MadGraph` arises from the shape differences between the Sherpa and MadGraph samples. This uncertainty is significant in our analysis due to the known modeling disparities between these two generators, especially in regions of high m_{jj}^{tag} and jet multiplicity. The influence of this uncertainty on the signal strength μ is notable, approximately 8% in the resolved regime and somewhat lesser in the merged regime, as depicted in the ranking plot (Figure 10.4).
- `SysMODEL_ttbar_PwHwg7` uncertainties refer to the modeling uncertainties associated with the $t\bar{t}$ background, which are derived from alternative sample comparisons. The analysis relies on the constraints obtained from the dedicated TCRs to mitigate these uncertainties.
- `SysPRW_DATASF` and `SysJET_Pileup_Offset` uncertainties are related to pile-up effects. These systematic uncertainties significantly influence the shapes of forward jets and jet track multiplicity (quark-gluon tagging variables). This effect is detailed in Section 9.1.3.
- `SysTheoryQCD_W` represents the QCD theory scale uncertainty impacting the $W + \text{jets}$ prediction. This uncertainty is effectively constrained by the shape of the final discriminant (DNN) in the resolved signal region, as illustrated in Figure 9.7.
- `SysTheoryISR_ttbar` and `SysTheoryFSR_Top` represent theory uncertainties in $t\bar{t}$ production with additional radiation in the initial or final states, as detailed in Section 9.2.2. There is a top modeling

issue in the merged control regions. Given that these nuisance parameters do not directly impact the signal, we simplify our analysis by employing a single bin for the merged TCRs.

- `SysFATJET_BJT_JET_JetTagSF_Hadronisation` represents the uncertainty associated with the hadronisation modeling.

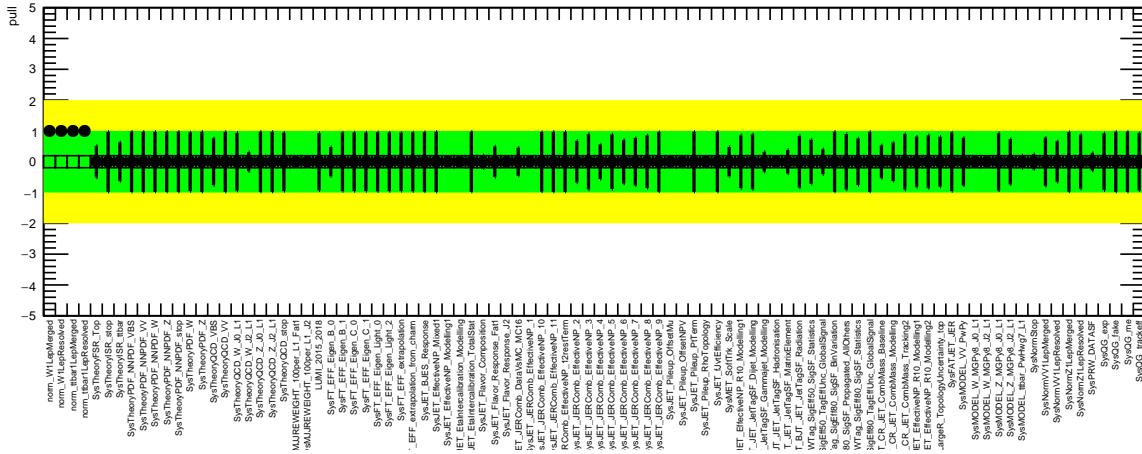


Figure 10.2: Fit cross-check, conditional fit ($\mu = 1$) to asimov data for the 1 lepton channel.

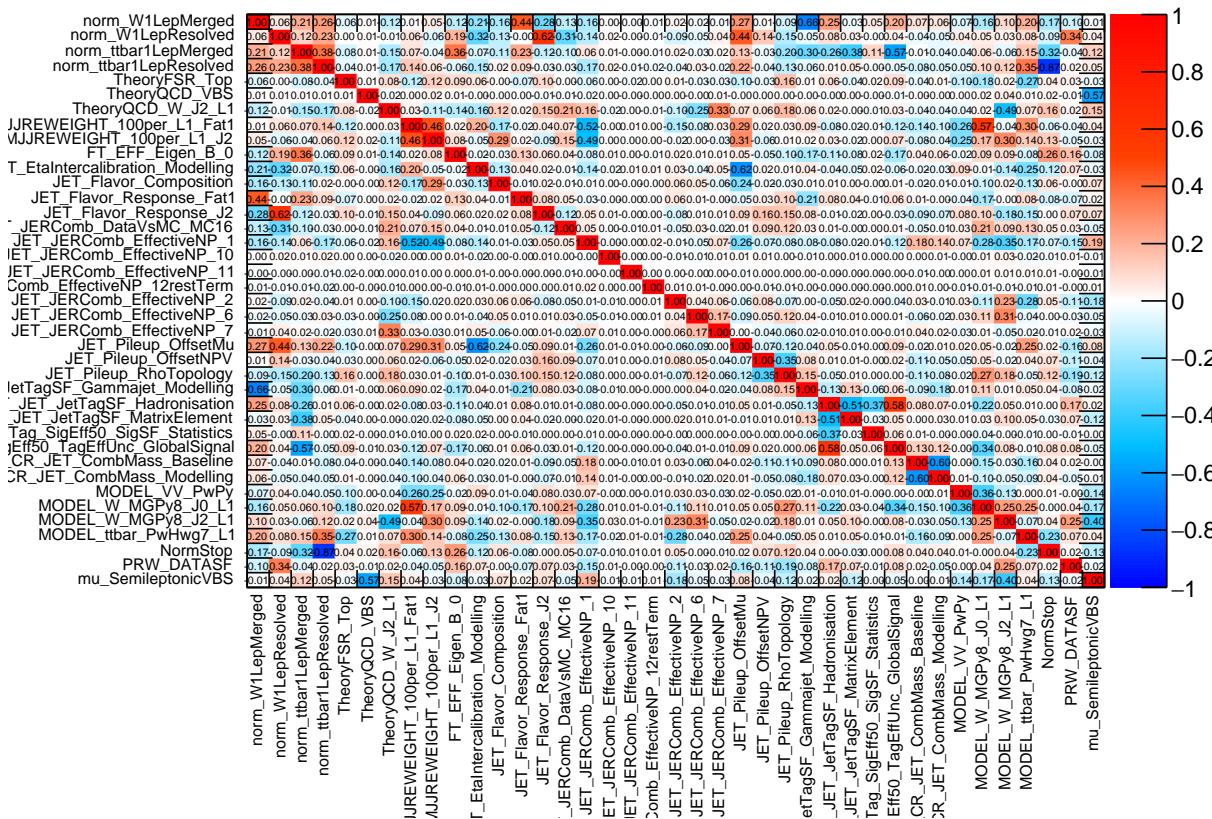


Figure 10.3: Correlations for unconditional fit ($\mu = 1$) to asimov data in the full range, for the 1 lepton channel.

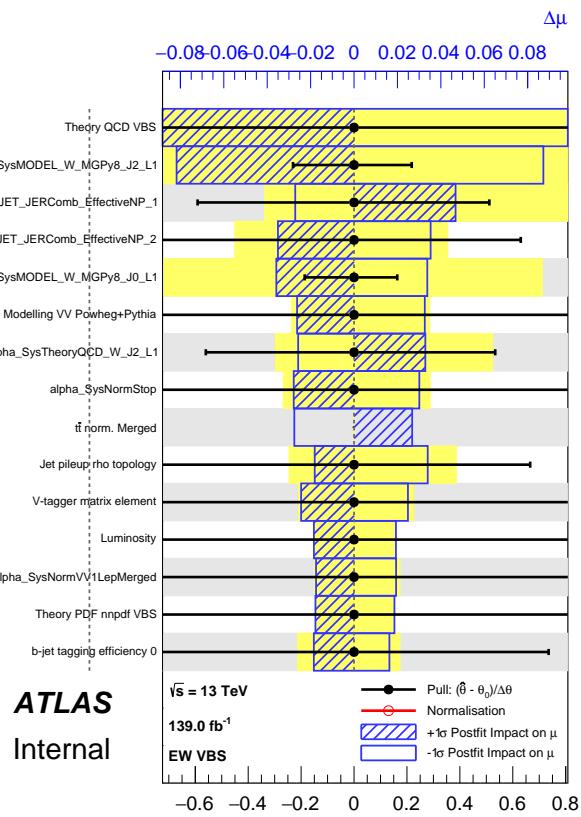


Figure 10.4: Ranking plot for unconditional fit ($\mu = 1$) to asimov data in the full range, for the 1 lepton channel.

10.4 Fit Results with Unblinded Data

Using the methodology detailed in earlier sections of this chapter, we have obtained promising unblinded results using the DNN. Table 10.5 presents both expected and observed significances for the 1-lepton channel. The expected significance is derived from Asimov data, with adjustments for background normalization and nuisance parameters. Remarkably, the observed significance from actual data surpasses the critical 5σ threshold. This achievement marks the first time we reach this level of significance in the semileptonic final states for electroweak $VV + jj$ production within a VBS-enhanced phase space, using only the 1-lepton channel. We observe a deficit compared to the expected significance. Although the DNN fitting approach does not account for certain signal uncertainties, such as EWK-QCD interference and shower systematics, the results are robust enough to serve as a corroborative cross-check against other baseline analyses using different ML approaches.

	1-lep
Expected pre-fit significance	5.39
Expected post-fit significance	5.98
Observed significance	5.78

Table 10.5: Expected and observed significances for the 1-lepton channel.

Preliminary DNN fitting yields a signal strength parameter $\mu_{VBS} = 1.16 \pm 0.25$. Table 10.6 shows the uncertainty breakdown on μ_{VBS} for the 1-lepton channel. Postfit distributions for the SRs with full range data are shown in Figure 10.5, while the corresponding distributions for the CRs are displayed in Figure 10.6. These distributions demonstrate satisfactory agreement between the data and the MC simulations. Tables 10.7-10.8 show the postfit event yields for signal and background processes in all the CRs and SRs.

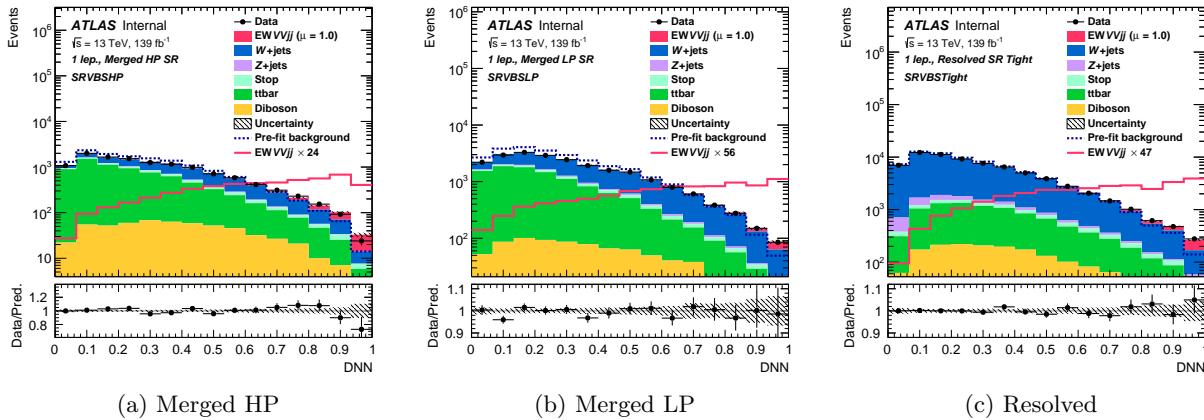


Figure 10.5: Postfit plots for SR distributions

Figure 10.7 shows the correlation matrix for the full-range data fit. Figures 10.8-10.9 display the pulls

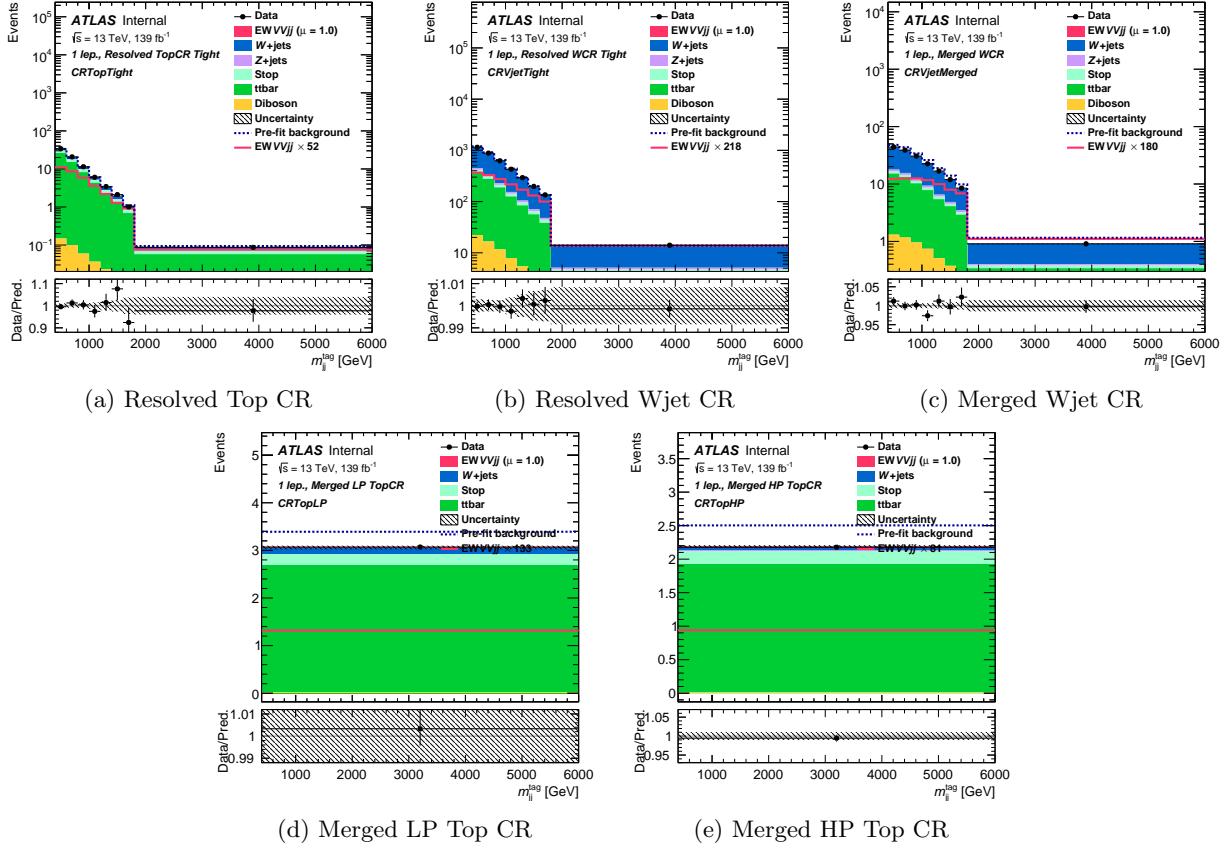


Figure 10.6: Postfit plots for CR distributions

of the NPs and the ranking plots for the unblinded full-range data fit. The leading uncertainties impacting the fit include the theoretical QCD uncertainty on the signal, the modeling uncertainties of the $W + \text{jets}$ background, the reweighting uncertainties of the tagging jets, and `SysJET_JERComb`, the baseline uncertainties from the small- R jets resolution.

Uncertainty Source	Positive Error	Negative Error	Symmetric Uncertainty
Total	+0.263	-0.237	± 0.250
DataStat	+0.107	-0.105	± 0.106
FullSyst	+0.240	-0.212	± 0.226
Floating normalizations	+0.044	-0.035	± 0.039
All normalizations	+0.095	-0.079	± 0.087
All but normalizations	+0.230	-0.202	± 0.216
Jets MET	+0.132	-0.118	± 0.125
BTag	+0.031	-0.031	± 0.031
Leptons	+0.000	-0.000	± 0.000
Luminosity	+0.025	-0.019	± 0.022
Diboson	+0.036	-0.032	± 0.034
Model	+0.148	-0.128	± 0.138
Signal Systematics	+0.166	-0.133	± 0.150
Mjj reweighting	+0.023	-0.037	± 0.030
Quark gluon	+0.015	-0.009	± 0.012
MC stat	+0.082	-0.077	± 0.079

Table 10.6: Uncertainty breakdown on μ_{VBS} in 1-lepton DNN.

DNN_SRVBShp		DNN_SRVBslp	
W	3772.48 ± 207.54	W	10111.99 ± 371.67
Z	175.98 ± 24.93	Z	462.53 ± 58.40
Diboson	574.86 ± 163.38	Diboson	804.42 ± 231.65
stop	598.20 ± 167.59	stop	846.84 ± 244.44
ttbar	6748.42 ± 277.01	ttbar	9856.21 ± 356.81
Bkg	11869.94 ± 110.89	Bkg	22082.00 ± 171.45
EW6lvqq	239.21 ± 42.47	EW6lvqq	195.86 ± 43.72
Signal	239.21 ± 42.47	Signal	195.86 ± 43.72
SignalExpected	206.48 ± 36.66	SignalExpected	169.06 ± 37.74
S/B	$2.02e-02$	S/B	$8.87e-03$
S/sqrt(S+B)	$2.17e+00$	S/sqrt(S+B)	$1.31e+00$
data	12178	data	22158

DNN_SRVBsRes	
W	57446.51 ± 785.36
Z	2168.91 ± 324.79
Diboson	1721.91 ± 552.18
stop	1521.92 ± 416.17
ttbar	7633.82 ± 382.92
Bkg	70493.07 ± 317.65
EW6lvqq	742.72 ± 133.71
Signal	742.72 ± 133.71
SignalExpected	641.09 ± 115.41
S/B	$1.05e-02$
S/sqrt(S+B)	$2.78e+00$
data	71272

Table 10.7: Postfit event yields for the analysis SRs in the 1 lepton channel.

tagMjj_CRTopHP		tagMjj_CRTopLP	
W	279.58 ± 16.61	W	717.29 ± 36.10
Z	18.94 ± 2.83	Z	47.45 ± 6.56
Diboson	46.59 ± 13.86	Diboson	68.76 ± 19.33
stop	1113.16 ± 307.91	stop	1227.37 ± 363.11
ttbar	10728.53 ± 343.94	ttbar	15013.46 ± 390.23
Bkg	12186.80 ± 109.43	Bkg	17074.34 ± 212.66
EW6lvqq	74.71 ± 14.77	EW6lvqq	63.89 ± 14.29
Signal	74.71 ± 14.77	Signal	63.89 ± 14.29
SignalExpected	64.48 ± 12.75	SignalExpected	55.15 ± 12.33
S/B	$6.13e-03$	S/B	$3.74e-03$
S/sqrt(S+B)	$6.75e-01$	S/sqrt(S+B)	$4.88e-01$
data	12195	data	17195
tagMjj_CRTopRes		tagMjj_CRWjetMerged	
W	2044.50 ± 71.00	W	22644.17 ± 842.49
Z	101.67 ± 15.16	Z	1147.58 ± 146.40
Diboson	81.79 ± 26.43	Diboson	1225.46 ± 347.60
stop	1885.19 ± 514.80	stop	1294.60 ± 360.16
ttbar	11853.62 ± 565.44	ttbar	11995.36 ± 787.80
Bkg	15966.76 ± 138.04	Bkg	38307.18 ± 211.01
EW6lvqq	155.57 ± 31.27	EW6lvqq	124.40 ± 23.18
Signal	155.57 ± 31.27	Signal	124.40 ± 23.18
SignalExpected	134.28 ± 26.99	SignalExpected	107.37 ± 20.01
S/B	$9.74e-03$	S/B	$3.25e-03$
S/sqrt(S+B)	$1.23e+00$	S/sqrt(S+B)	$6.35e-01$
data	16137	data	38486
tagMjj_CRWjetRes			
W	501272.25 ± 9637.60		
Z	21630.23 ± 3315.37		
Diboson	15241.39 ± 4878.27		
stop	29343.00 ± 8099.30		
ttbar	231974.93 ± 10369.58		
Bkg	799461.79 ± 2665.92		
EW6lvqq	1989.35 ± 394.05		
Signal	1989.35 ± 394.05		
SignalExpected	1717.15 ± 340.13		
S/B	$2.49e-03$		
S/sqrt(S+B)	$2.22e+00$		
data	801406		

Table 10.8: Postfit event yields for the analysis CRs in the 1 lepton channel.

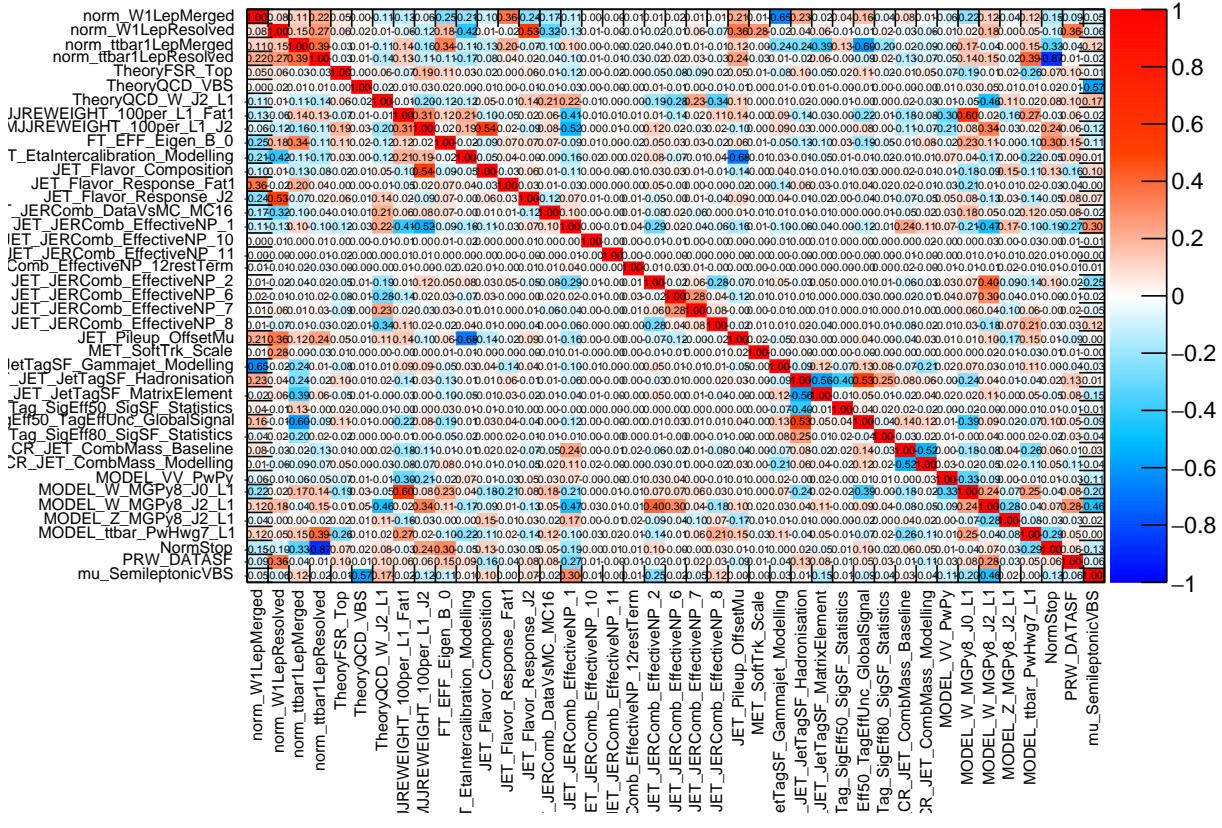


Figure 10.7: Correlations for unconditional fit ($\mu = 1$) to unblinded data in the full range.

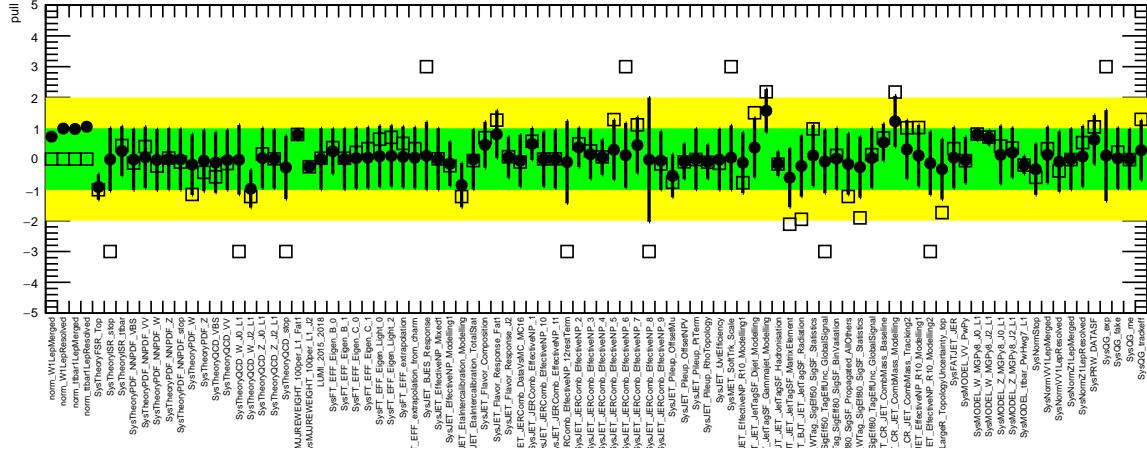


Figure 10.8: Fit cross-check, the pulls of the NPs for the unblinded fit.

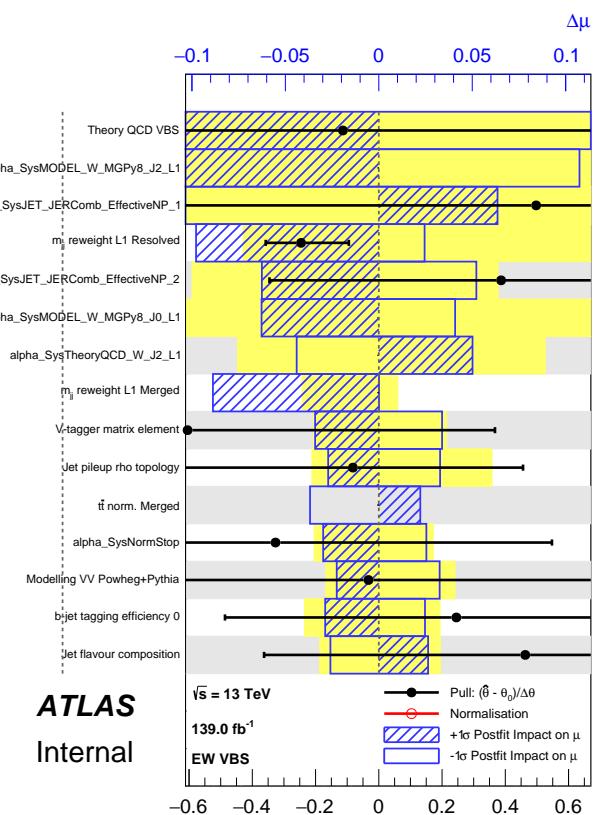


Figure 10.9: Ranking of nuisance parameters postfit-sorted for the unblinded fit.

10.5 Conclusions

This thesis has presented a search for semileptonic Vector Boson Scattering (VBS) events, utilizing the complete Run-II dataset collected by the ATLAS detector at the LHC. Significant emphasis was placed on my contributions, particularly on the deployment of a Deep Neural Network (DNN) approach to analyze and enhance sensitivity to the SM VBS signal in the 1-lepton channel.

The preliminary unblinded results from this study have shown an observed significance of 5.78σ , which exceeds the critical 5σ observation threshold, with a reported signal strength parameter $\mu_{VBS} = 1.16 \pm 0.25$. This achievement is notable as it suggests the feasibility of observing VBS events through the analysis of the 1-lepton channel alone.

The DNN methodology introduced here serves as an alternative machine learning strategy within the broader context of a combined search [10] across three channels (0-lepton, 1-lepton, and 2-lepton), which utilized a Recurrent Neural Network (RNN) approach. While a direct comparison between the DNN and RNN methods does not definitively establish one as superior to the other, it can be argued, based on this work, that the simpler yet effective DNN approach may offer a more suitable solution for future analyses. The significance of this work lies in its demonstration of the DNN's potential to enhance the analysis of semileptonic VBS events. By achieving significant results in the 1-lepton channel, this research contributes to the evolving understanding of VBS processes and their observation at the LHC.

This research encountered several limitations, partly due to the extended duration of the combined search across the three channels. This period coincided with unforeseen challenges posed by the COVID-19 pandemic, which indirectly affected the scope of the investigation. This thesis did not account for signal uncertainties arising from EWK-QCD interference and parton shower systematics. More comprehensive studies of the nuisance parameters and adjustments in the binning strategies for certain NPs could potentially improve the fit results and overall analysis sensitivity.

The study of anomalous quartic gauge couplings (aQGC) is a significant yet ongoing aspect of the combined search, briefly discussed in this thesis. While my involvement in this specific area was limited, the potential for investigating aQGCs exclusively within the 1-lepton channel—especially with additional data from the ATLAS detector—presents an intriguing avenue for future research.

Addressing these limitations presents an opportunity for subsequent research to build upon the foundational work laid out in this thesis, thereby further advancing our understanding and analysis capabilities in the field of semileptonic VBS analysis.

Appendix A

List of MC Samples

The MC samples used in this analysis are listed in this section. Table A.1–A.9 summarizes the MC samples for the SM backgrounds considered in this analysis.

The alternative MC samples used for background systematics are summarised in Table A.10–A.16.

Table A.17 shows the list of the SM EW VV+jj signals samples. Table A.18–A.19–A.20–A.21 A.22–A.23–A.24–A.25 shows the list of the BSM aQGC VV+jj signals samples.

Table A.1: $W \rightarrow e\nu$ +jets samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times BR$ [pb]	k-factor	ϵ_{filter}
364170	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV0_70_CVetoBVeto	1.913E+01	1.0	0.82447
364171	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV0_70_CFilterBVeto	1.913E+01	1.0	0.1303
364172	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV0_70_BFilter	1.914E+01	1.0	0.044141
364173	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV70_140_CVetoBVeto	9.426E-01	1.0	0.66872
364174	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV70_140_CFilterBVeto	9.457E-01	1.0	0.22787
364175	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV70_140_BFilter	9.451E-01	1.0	0.10341
364176	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV140_280_CVetoBVeto	3.398E-01	1.0	0.59691
364177	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV140_280_CFilterBVeto	3.399E-01	1.0	0.28965
364178	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV140_280_BFilter	3.395E-01	1.0	0.10898
364179	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV280_500_CVetoBVeto	7.208E-02	1.0	0.54441
364180	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV280_500_CFilterBVeto	7.213E-02	1.0	0.31675
364181	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV280_500_BFilter	7.211E-02	1.0	0.13391
364182	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV500_1000	1.522E-02	1.0	1
364183	Sherpa_221_NNPDF30NNLO_Wenu_MAXHTPTV1000_E_CMS	1.233E-03	1.0	1

Table A.2: $W \rightarrow \mu\nu$ +jets samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times BR$ [pb]	k-factor	ϵ_{filter}
361334	Sherpa_CT10_Wmumu_Pt280_500_CFilterBVeto	5.747E-03	1.0	0.221
364156	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV0_70_CVetoBVeto	1.914E+01	1.0	0.8238
364157	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV0_70_CFilterBVeto	1.912E+01	1.0	0.1304
364158	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV0_70_BFilter	1.914E+01	1.0	0.044118
364159	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV70_140_CVetoBVeto	9.448E-01	1.0	0.67463
364160	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV70_140_CFilterBVeto	9.378E-01	1.0	0.23456
364161	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV70_140_BFilter	9.446E-01	1.0	0.075648
364162	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV140_280_CVetoBVeto	3.395E-01	1.0	0.62601
364163	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV140_280_CFilterBVeto	3.401E-01	1.0	0.28947
364164	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV140_280_BFilter	3.395E-01	1.0	0.10872
364165	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV280_500_CVetoBVeto	7.207E-02	1.0	0.54647
364166	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV280_500_CFilterBVeto	7.220E-02	1.0	0.31743
364167	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV280_500_BFilter	7.204E-02	1.0	0.13337
364168	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV500_1000	1.501E-02	1.0	1
364169	Sherpa_221_NNPDF30NNLO_Wmumu_MAXHTPTV1000_E_CMS	1.234E-03	1.0	1

Table A.3: $W \rightarrow \tau\nu$ +jets samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times BR$ [pb]	k-factor	ϵ_{filter}
364184	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_CVetoBVeto	1.915E+01	1.0	0.82495
364185	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_CFilterBVeto	1.915E+01	1.0	0.12934
364186	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_BFilter	1.916E+01	1.0	0.044594
364187	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_CVetoBVeto	9.476E-01	1.0	0.67382
364188	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_CFilterBVeto	9.467E-01	1.0	0.22222
364189	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_BFilter	9.433E-01	1.0	0.10391
364190	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV140_280_CVetoBVeto	3.394E-01	1.0	0.59622
364191	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV140_280_CFilterBVeto	3.396E-01	1.0	0.29025
364192	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV140_280_BFilter	3.395E-01	1.0	0.11799
364193	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_CVetoBVeto	7.207E-02	1.0	0.54569
364194	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_CFilterBVeto	7.198E-02	1.0	0.31648
364195	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_BFilter	7.203E-02	1.0	0.13426
364196	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV500_1000	1.505E-02	1.0	1
364197	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV1000_E_CMS	1.234E-03	1.0	1

Table A.4: $Z \rightarrow ee$ +jets samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times BR$ [pb]	k-factor	ϵ_{filter}
364114	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV0_70_CVetoBVeto	1.982E+00	1.0	0.82106
364115	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV0_70_CFilterBVeto	1.981E+00	1.0	0.11295
364116	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV0_70_BFilter	1.982E+00	1.0	0.063809
364117	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV70_140_CVetoBVeto	1.105E-01	1.0	0.69043
364118	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV70_140_CFilterBVeto	1.106E-01	1.0	0.18382
364119	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV70_140_BFilter	1.103E-01	1.0	0.11443
364120	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV140_280_CVetoBVeto	4.073E-02	1.0	0.61452
364121	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV140_280_CFilterBVeto	4.067E-02	1.0	0.23044
364122	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV140_280_BFilter	4.069E-02	1.0	0.14927
364123	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV280_500_CVetoBVeto	8.674E-03	1.0	0.56134
364124	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV280_500_CFilterBVeto	8.671E-03	1.0	0.26294
364125	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV280_500_BFilter	8.677E-03	1.0	0.17223
364126	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV500_1000	1.808E-03	1.0	1
364127	Sherpa_221_NNPDF30NNLO_Zee_MAXHTPTV1000_E_CMS	1.486E-04	1.0	1

Table A.5: $Z \rightarrow \mu\mu$ +jets samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times BR$ [pb]	k-factor	ϵ_{filter}
364100	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV0_70_CVetoBVeto	1.983E+00	1.0	0.8221
364101	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV0_70_CFilterBVeto	1.978E+00	1.0	0.11308
364102	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV0_70_BFilter	1.982E+00	1.0	0.064161
364103	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV70_140_CVetoBVeto	1.089E-01	1.0	0.68873
364104	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV70_140_CFilterBVeto	1.094E-01	1.0	0.18596
364105	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV70_140_BFilter	1.089E-01	1.0	0.11375
364106	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV140_280_CVetoBVeto	3.988E-02	1.0	0.60899
364107	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV140_280_CFilterBVeto	3.980E-02	1.0	0.23308
364108	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV140_280_BFilter	3.991E-02	1.0	0.14618
364109	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV280_500_CVetoBVeto	8.537E-03	1.0	0.55906
364110	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV280_500_CFilterBVeto	8.540E-03	1.0	0.26528
364111	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV280_500_BFilter	8.493E-03	1.0	0.17559
364112	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV500_1000	1.788E-03	1.0	1
364113	Sherpa_221_NNPDF30NNLO_Zmumu_MAXHTPTV1000_E_CMS	1.477E-04	1.0	1

Table A.6: $Z \rightarrow \tau\tau$ +jets samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times BR$ [pb]	k-factor	ϵ_{filter}
364128	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CVetoBVeto	1.982E+00	1.0	0.82142
364129	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CFilterBVeto	1.979E+00	1.0	0.11314
364130	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_BFilter	1.982E+00	1.0	0.064453
364131	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_CVetoBVeto	1.104E-01	1.0	0.68883
364132	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_CFilterBVeto	1.105E-01	1.0	0.1829
364133	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_BFilter	1.109E-01	1.0	0.1283
364134	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CVetoBVeto	4.078E-02	1.0	0.60821
364135	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CFilterBVeto	4.074E-02	1.0	0.22897
364136	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_BFilter	4.076E-02	1.0	0.13442
364137	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_CVetoBVeto	8.550E-03	1.0	0.56036
364138	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_CFilterBVeto	8.671E-03	1.0	0.26245
364139	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_BFilter	8.680E-03	1.0	0.17313
364140	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV500_1000	1.810E-03	1.0	1
364141	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV1000_E_CMS	1.483E-04	1.0	1

Table A.7: $Z \rightarrow \nu\nu + \text{jets}$ samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR} [\text{pb}]$	k-factor	ϵ_{filter}
364142	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV0_70_CVetoBVeto	1.070E+01	1.0	0.8216
364143	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV0_70_CFilterBVeto	1.070E+01	1.0	0.11123
364144	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV0_70_BFilter	1.071E+01	1.0	0.066175
364145	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV70_140_CVetoBVeto	6.032E-01	1.0	0.68924
364146	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV70_140_CFilterBVeto	6.081E-01	1.0	0.18243
364147	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV70_140_BFilter	6.033E-01	1.0	0.1283
364148	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV140_280_CVetoBVeto	2.223E-01	1.0	0.60735
364149	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV140_280_CFilterBVeto	2.219E-01	1.0	0.22527
364150	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV140_280_BFilter	2.225E-01	1.0	0.15185
364151	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV280_500_CVetoBVeto	4.738E-02	1.0	0.55887
364152	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV280_500_CFilterBVeto	4.740E-02	1.0	0.26201
364153	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV280_500_BFilter	4.748E-02	1.0	0.17514
364154	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV500_1000	9.910E-03	1.0	1
364155	Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV1000_E_CMS	8.181E-04	1.0	1

Table A.8: $t\bar{t}$ and single top samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR} [\text{pb}]$	k-factor	ϵ_{filter}
410470	PhPy8EG_A14_ttbar_hdamp258p75_nonallhad	8.318E-01	1.0	0.543
410471	PhPy8EG_A14_ttbar_hdamp258p75_allhad	8.318E-01	1.0	0.45628
410472	PhPy8EG_A14_ttbar_hdamp258p75_dil	8.318E-01	1.0	0.105
410644	PowhegPythiaEvtGen_P2012.SingleTopSchan_noAllHad_top	2.027E-03	1.0	1
410645	PowhegPythia8EvtGen_A14_singletop_schan_lept_antitop	1.267E-03	1.0	1
410658	PhPy8EG_A14_tchan_BW50_lept_top	3.699E-02	1.2	1
410659	PhPy8EG_A14_tchan_BW50_lept_antitop	2.218E-02	1.2	1
410646	PowhegPythia8EvtGen_A14_Wt_DR_inclusive_top	3.794E-02	0.9	1
410647	PowhegPythia8EvtGen_A14_Wt_DR_inclusive_antitop	3.791E-02	0.9	1
410648	PowhegPythia8EvtGen_A14_Wt_DR_dilepton_top	3.997E-03	1.0	1
410649	PowhegPythia8EvtGen_A14_Wt_DR_dilepton_antitop	3.994E-03	1.0	1

Table A.9: Diboson samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR} [\text{pb}]$	k-factor	ϵ_{filter}
363355	Sh_221_NNPDF30NNLO_ZqqZvv	1.557E-02	1.0	0.27978
363356	Sh_221_NNPDF30NNLO_ZqqZll	1.556E-02	1.0	0.14089
363357	Sh_221_NNPDF30NNLO_WqqZvv	6.787E-03	1.0	1
363358	Sh_221_NNPDF30NNLO_WqqZll	3.433E-03	1.0	1
363359	Sh_221_NNPDF30NNLO_WpqqWmlv	2.471E-02	1.0	1
363360	Sh_221_NNPDF30NNLO_WplvWmqq	2.473E-02	1.0	1
363489	Sh_221_NNPDF30NNLO_WlvZqq	1.142E-02	1.0	1

Table A.10: $W \rightarrow e\nu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR} [\text{pb}]$	k-factor	ϵ_{filter}
363600	MGPy8EG_N30NLO_Wenu_Ht0_70_CVetoBVeto	16719.0	1.12	8.38E+03
363601	MGPy8EG_N30NLO_Wenu_Ht0_70_CFilterBVeto	16720.0	1.12	1.38E+03
363602	MGPy8EG_N30NLO_Wenu_Ht0_70_BFilter	16717.0	1.12	2.42E+02
363603	MGPy8EG_N30NLO_Wenu_Ht70_140_CVetoBVeto	755.1	1.12	7.12E+03
363604	MGPy8EG_N30NLO_Wenu_Ht70_140_CFilterBVeto	755.77	1.12	2.40E+03
363605	MGPy8EG_N30NLO_Wenu_Ht70_140_BFilter	755.73	1.12	4.83E+02
363606	MGPy8EG_N30NLO_Wenu_Ht140_280_CVetoBVeto	318.96	1.12	6.66E+03
363607	MGPy8EG_N30NLO_Wenu_Ht140_280_CFilterBVeto	319.93	1.12	2.64E+03
363608	MGPy8EG_N30NLO_Wenu_Ht140_280_BFilter	319.45	1.12	6.94E+02
363609	MGPy8EG_N30NLO_Wenu_Ht280_500_CVetoBVeto	73.528	1.12	6.19E+03
363610	MGPy8EG_N30NLO_Wenu_Ht280_500_CFilterBVeto	73.562	1.12	2.85E+03
363611	MGPy8EG_N30NLO_Wenu_Ht280_500_BFilter	73.556	1.12	9.52E+02
363612	MGPy8EG_N30NLO_Wenu_Ht500_700_CVetoBVeto	11.529	1.12	5.87E+03
363613	MGPy8EG_N30NLO_Wenu_Ht500_700_CFilterBVeto	11.517	1.12	2.98E+03
363614	MGPy8EG_N30NLO_Wenu_Ht500_700_BFilter	11.51	1.12	1.15E+03
363615	MGPy8EG_N30NLO_Wenu_Ht700_1000_CVetoBVeto	40.158	1.12	5.66E+03
363616	MGPy8EG_N30NLO_Wenu_Ht700_1000_CFilterBVeto	4.014	1.12	3.04E+03
363617	MGPy8EG_N30NLO_Wenu_Ht700_1000_BFilter	40.123	1.12	1.28E+03
363618	MGPy8EG_N30NLO_Wenu_Ht1000_2000_CVetoBVeto	13.243	1.12	5.48E+03
363619	MGPy8EG_N30NLO_Wenu_Ht1000_2000_CFilterBVeto	13.286	1.12	3.08E+03
363620	MGPy8EG_N30NLO_Wenu_Ht1000_2000_BFilter	1.326	1.12	1.43E+03
363621	MGPy8EG_N30NLO_Wenu_Ht2000_E.CMS.CVetoBVeto	0.042294	1.12	5.24E+03
363622	MGPy8EG_N30NLO_Wenu_Ht2000_E.CMS.CFilterBVeto	0.041884	1.12	3.17E+03
363623	MGPy8EG_N30NLO_Wenu_Ht2000_E.CMS.BFilter	0.047382	1.12	1.51E+03

Table A.11: $W \rightarrow \mu\nu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR} [\text{pb}]$	k-factor	ϵ_{filter}
363624	MGPy8EG_N30NLO_Wmumu_Ht0_70_CVetoBVeto	16720.0	1.12	8.38E+03
363625	MGPy8EG_N30NLO_Wmumu_Ht0_70_CFilterBVeto	16717.0	1.12	1.38E+03
363626	MGPy8EG_N30NLO_Wmumu_Ht0_70_BFilter	16719.0	1.12	2.42E+02
363627	MGPy8EG_N30NLO_Wmumu_Ht70_140_CVetoBVeto	755.19	1.12	7.12E+03
363628	MGPy8EG_N30NLO_Wmumu_Ht70_140_CFilterBVeto	755.62	1.12	2.40E+03
363629	MGPy8EG_N30NLO_Wmumu_Ht70_140_BFilter	755.77	1.12	4.83E+02
363630	MGPy8EG_N30NLO_Wmumu_Ht140_280_CVetoBVeto	318.83	1.12	6.66E+03
363631	MGPy8EG_N30NLO_Wmumu_Ht140_280_CFilterBVeto	319.89	1.12	2.64E+03
363632	MGPy8EG_N30NLO_Wmumu_Ht140_280_BFilter	319.41	1.12	6.94E+02
363633	MGPy8EG_N30NLO_Wmumu_Ht280_500_CVetoBVeto	73.585	1.12	6.19E+03
363634	MGPy8EG_N30NLO_Wmumu_Ht280_500_CFilterBVeto	73.548	1.12	2.85E+03
363635	MGPy8EG_N30NLO_Wmumu_Ht280_500_BFilter	73.569	1.12	9.52E+02
363636	MGPy8EG_N30NLO_Wmumu_Ht500_700_CVetoBVeto	11.522	1.12	5.86E+03
363637	MGPy8EG_N30NLO_Wmumu_Ht500_700_CFilterBVeto	11.524	1.12	2.96E+03
363638	MGPy8EG_N30NLO_Wmumu_Ht500_700_BFilter	11.522	1.12	1.16E+03
363639	MGPy8EG_N30NLO_Wmumu_Ht700_1000_CVetoBVeto	40.194	1.12	5.67E+03
363640	MGPy8EG_N30NLO_Wmumu_Ht700_1000_CFilterBVeto	40.252	1.12	3.05E+03
363641	MGPy8EG_N30NLO_Wmumu_Ht700_1000_BFilter	40.139	1.12	1.28E+03
363642	MGPy8EG_N30NLO_Wmumu_Ht1000_2000.CVetoBVeto	13.262	1.12	5.45E+03
363643	MGPy8EG_N30NLO_Wmumu_Ht1000_2000.CFilterBVeto	13.215	1.12	3.08E+03
363644	MGPy8EG_N30NLO_Wmumu_Ht1000_2000.BFilter	13.287	1.12	1.44E+03
363645	MGPy8EG_N30NLO_Wmumu_Ht2000_E.CMS.CVetoBVeto	0.041697	1.12	5.20E+03
363646	MGPy8EG_N30NLO_Wmumu_Ht2000_E.CMS.CFilterBVeto	0.042199	1.12	3.17E+03
363647	MGPy8EG_N30NLO_Wmumu_Ht2000_E.CMS.BFilter	0.042049	1.12	1.59E+03

Table A.12: $W \rightarrow \tau\nu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
363648	MGPy8EG_N30NLO_Wtaunu_Ht0_70_CVetoBVeto	16702.0	1.12	8.38E+03
363649	MGPy8EG_N30NLO_Wtaunu_Ht0_70_CFilterBVeto	16702.0	1.12	1.38E+03
363650	MGPy8EG_N30NLO_Wtaunu_Ht0_70_BFilter	16702.0	1.12	2.42E+02
363651	MGPy8EG_N30NLO_Wtaunu_Ht70_140_CVetoBVeto	754.91	1.12	7.12E+03
363652	MGPy8EG_N30NLO_Wtaunu_Ht70_140_CFilterBVeto	755.3	1.12	2.39E+03
363653	MGPy8EG_N30NLO_Wtaunu_Ht70_140_BFilter	755.14	1.12	4.85E+02
363654	MGPy8EG_N30NLO_Wtaunu_Ht140_280_CVetoBVeto	315.31	1.12	6.67E+03
363655	MGPy8EG_N30NLO_Wtaunu_Ht140_280_CFilterBVeto	316.05	1.12	2.64E+03
363656	MGPy8EG_N30NLO_Wtaunu_Ht140_280_BFilter	315.64	1.12	6.93E+02
363657	MGPy8EG_N30NLO_Wtaunu_Ht280_500_CVetoBVeto	72.275	1.12	6.22E+03
363658	MGPy8EG_N30NLO_Wtaunu_Ht280_500_CFilterBVeto	72.19	1.12	2.85E+03
363659	MGPy8EG_N30NLO_Wtaunu_Ht280_500_BFilter	72.226	1.12	9.52E+02
363660	MGPy8EG_N30NLO_Wtaunu_Ht500_700_CVetoBVeto	11.44	1.12	5.88E+03
363661	MGPy8EG_N30NLO_Wtaunu_Ht500_700_CFilterBVeto	11.463	1.12	2.97E+03
363662	MGPy8EG_N30NLO_Wtaunu_Ht500_700_BFilter	11.459	1.12	1.16E+03
363663	MGPy8EG_N30NLO_Wtaunu_Ht700_1000_CVetoBVeto	40.265	1.12	5.67E+03
363664	MGPy8EG_N30NLO_Wtaunu_Ht700_1000_CFilterBVeto	4.027	1.12	3.03E+03
363665	MGPy8EG_N30NLO_Wtaunu_Ht700_1000_BFilter	40.283	1.12	1.30E+03
363666	MGPy8EG_N30NLO_Wtaunu_Ht1000_2000_CVetoBVeto	13.264	1.12	5.46E+03
363667	MGPy8EG_N30NLO_Wtaunu_Ht1000_2000_CFilterBVeto	13.205	1.12	3.12E+03
363668	MGPy8EG_N30NLO_Wtaunu_Ht1000_2000_BFilter	13.222	1.12	1.43E+03
363669	MGPy8EG_N30NLO_Wtaunu_Ht2000_E_CMS_CVetoBVeto	0.043501	1.12	5.21E+03
363670	MGPy8EG_N30NLO_Wtaunu_Ht2000_E_CMS_CFilterBVeto	0.042721	1.12	3.18E+03
363671	MGPy8EG_N30NLO_Wtaunu_Ht2000_E_CMS_BFilter	0.042683	1.12	1.66E+03

Table A.13: $Z \rightarrow ee + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
363147	MGPy8EG_N30NLO_Zee_Ht0_70_CVetoBVeto	1719.7	1.141	0.83292
363148	MGPy8EG_N30NLO_Zee_Ht0_70_CFilterBVeto	1719.4	1.141	0.10775
363149	MGPy8EG_N30NLO_Zee_Ht0_70_BFilter	1719.4	1.141	0.059156
363150	MGPy8EG_N30NLO_Zee_Ht70_140_CVetoBVeto	85.105	1.141	0.71754
363151	MGPy8EG_N30NLO_Zee_Ht70_140_CFilterBVeto	85.041	1.141	0.17377
363152	MGPy8EG_N30NLO_Zee_Ht70_140_BFilter	85.175	1.141	0.10763
363153	MGPy8EG_N30NLO_Zee_Ht140_280_CVetoBVeto	36.005	1.141	0.67279
363154	MGPy8EG_N30NLO_Zee_Ht140_280_CFilterBVeto	36.028	1.141	0.19996
363155	MGPy8EG_N30NLO_Zee_Ht140_280_BFilter	36.06	1.141	0.12486
363156	MGPy8EG_N30NLO_Zee_Ht280_500_CVetoBVeto	82.054	1.141	0.62846
363157	MGPy8EG_N30NLO_Zee_Ht280_500_CFilterBVeto	82.126	1.141	0.22726
363158	MGPy8EG_N30NLO_Zee_Ht280_500_BFilter	82.474	1.141	0.14193
363159	MGPy8EG_N30NLO_Zee_Ht500_700_CVetoBVeto	12.733	1.141	0.5966
363160	MGPy8EG_N30NLO_Zee_Ht500_700_CFilterBVeto	1.273	1.141	0.24847
363161	MGPy8EG_N30NLO_Zee_Ht500_700_BFilter	12.722	1.141	0.15256
363162	MGPy8EG_N30NLO_Zee_Ht700_1000_CVetoBVeto	0.44546	1.141	0.57676
363163	MGPy8EG_N30NLO_Zee_Ht700_1000_CFilterBVeto	0.44611	1.141	0.26137
363164	MGPy8EG_N30NLO_Zee_Ht700_1000_BFilter	0.44603	1.141	0.16181
363165	MGPy8EG_N30NLO_Zee_Ht1000_2000_CVetoBVeto	0.15208	1.141	0.55543
363166	MGPy8EG_N30NLO_Zee_Ht1000_2000_CFilterBVeto	0.15248	1.141	0.27476
363167	MGPy8EG_N30NLO_Zee_Ht1000_2000_BFilter	0.15327	1.141	0.16618
363168	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_CVetoBVeto	0.0056989	1.141	0.53136
363169	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_CFilterBVeto	0.0057408	1.141	0.2923
363170	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_BFilter	0.0057164	1.141	0.17489

Table A.14: $Z \rightarrow \mu\mu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
363123	MGPy8EG_N30NLO_Zmumu_Ht0_70_CVetoBVeto	1714.5	1.141	0.83157
363124	MGPy8EG_N30NLO_Zmumu_Ht0_70_CFfilterBVeto	1715.5	1.141	0.10835
363125	MGPy8EG_N30NLO_Zmumu_Ht0_70_BFilter	1715.7	1.141	0.059162
363126	MGPy8EG_N30NLO_Zmumu_Ht70_140_CVetoBVeto	84.57	1.141	0.71809
363127	MGPy8EG_N30NLO_Zmumu_Ht70_140_CFfilterBVeto	84.588	1.141	0.17404
363128	MGPy8EG_N30NLO_Zmumu_Ht70_140_BFilter	84.752	1.141	0.10804
363129	MGPy8EG_N30NLO_Zmumu_Ht140_280_CVetoBVeto	35.883	1.141	0.67432
363130	MGPy8EG_N30NLO_Zmumu_Ht140_280_CFfilterBVeto	35.908	1.141	0.19951
363131	MGPy8EG_N30NLO_Zmumu_Ht140_280_BFilter	35.887	1.141	0.12607
363132	MGPy8EG_N30NLO_Zmumu_Ht280_500_CVetoBVeto	81.871	1.141	0.62802
363133	MGPy8EG_N30NLO_Zmumu_Ht280_500_CFfilterBVeto	81.805	1.141	0.2282
363134	MGPy8EG_N30NLO_Zmumu_Ht280_500_BFilter	81.705	1.141	0.14263
363135	MGPy8EG_N30NLO_Zmumu_Ht500_700_CVetoBVeto	1.271	1.141	0.59722
363136	MGPy8EG_N30NLO_Zmumu_Ht500_700_CFfilterBVeto	12.672	1.141	0.24952
363137	MGPy8EG_N30NLO_Zmumu_Ht500_700_BFilter	12.699	1.141	0.15292
363138	MGPy8EG_N30NLO_Zmumu_Ht700_1000_CVetoBVeto	0.43601	1.141	0.57118
363139	MGPy8EG_N30NLO_Zmumu_Ht700_1000_CFfilterBVeto	0.44623	1.141	0.25948
363140	MGPy8EG_N30NLO_Zmumu_Ht700_1000_BFilter	0.44567	1.141	0.15965
363141	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_CVetoBVeto	0.14899	1.141	0.54908
363142	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_CFfilterBVeto	0.14625	1.141	0.27164
363143	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_BFilter	0.14705	1.141	0.17299
363144	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_CVetoBVeto	0.005538	1.141	0.56337
363145	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_CFfilterBVeto	0.0055466	1.141	0.29294
363146	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_BFilter	0.0056422	1.141	0.16307

Table A.15: $Z \rightarrow \nu\nu + \text{jets}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
361515	MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np0	7518.4	12.283	1.0
361516	MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np1	1200.1	12.283	1.0
361517	MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np2	387.16	12.283	1.0
361518	MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np3	110.08	12.283	1.0
361519	361519.MadGraphPythia8EvtGen_A14NNPDF23LO_Znunu_Np4	43.389	12.283	1.0

Table A.16: $t\bar{t}$ alternative samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times \text{BR}$ [pb]	k-factor	ϵ_{filter}
410465	aMcAtNloPy8EvtGen_MEN30NLO_A14N23LO_ttbar_noShWe_dil	712.02	11.681	1.07E+03
410464	aMcAtNloPy8EvtGen_MEN30NLO_A14N23LO_ttbar_noShWe_SingleLep	711.43	11.691	4.40E+03
410467	PhPy8EG_A14_ttbar_hdamp517p5_nonallhad_AK10J160	729.74	1.0	1.56E+03
410468	PowhegHerwig7EG_H7UE_tt_hdamp258p75_704_nonallhad_AK10J160	730.15	1.0	1.50E+03
410469	aMcAtNloPy8EG_MEN30NLO_A14N23LO_ttbar_nonallhad_AK10J160	711.41	1.0	1.42E+03
410480	PhPy8EG_A14_ttbar_hdamp517p5_SingleLep	729.74	1.0	4.39E+03
410482	PhPy8EG_A14_ttbar_hdamp517p5_dil	729.74	1.0	1.05E+03
410557	PowhegHerwig7EvtGen_H7UE_tt_hdamp258p75_704_SingleLep	730.14	1.0	4.39E+03
410558	PowhegHerwig7EvtGen_H7UE_tt_hdamp258p75_704_dil	730.15	1.0	1.05E+03

Table A.17: EW VV+jj samples used in the analysis. The dataset ID, MC generator, production cross section, filter efficiency and total number of generated events are shown.

DS ID	Name	$\sigma \times BR$ [pb]	k-factor	ϵ_{filter}
364848	MGPy8EvtGen_NNPDF30NLO_A14NNPDF23LO_WWjj_lvqq_EW6_BVeto	1.9994	1.0	1.7465E-01
364849	MGPy8EvtGen_NNPDF30NLO_A14NNPDF23LO_WWjj_lvqq_EW6_BFilter	1.9777	1.0	1.0000E+00
364850	MGPy8EvtGen_NNPDF30NLO_A14NNPDF23LO_WZjj_lvqq_EW6	0.2571	1.0	1.0000E+00
364851	MGPy8EvtGen_NNPDF30NLO_A14NNPDF23LO_WZjj_vvqq_EW6	0.15532	1.0	1.0000E+00
364852	MGPy8EvtGen_NNPDF30NLO_A14NNPDF23LO_WZjj_llqq_EW6	0.045609	1.0	1.0000E+00
364853	MGPy8EvtGen_NNPDF30NLO_A14NNPDF23LO_ZZjj_vvqq_EW6	0.032238	1.0	1.0000E+00
364854	MGPy8EvtGen_NNPDF30NLO_A14NNPDF23LO_ZZjj_llqq_EW6	0.0096553	1.0	1.0000E+00

Table A.18: aQGC samples used in the analysis. The dataset ID, MC generator and total number of generated events are shown.

DSID	Name	Events
503152	MGPy8EG_aQGCFM0_INT_1_WmWm_lvqq	10000
503153	MGPy8EG_aQGCFM0_INT_1_WmZ_llqq	10000
503154	MGPy8EG_aQGCFM0_INT_1_WmZ_lvqq	10000
503155	MGPy8EG_aQGCFM0_INT_1_WmZ_vvqq	10000
503156	MGPy8EG_aQGCFM0_INT_1_WpWm_lvqq	20000
503157	MGPy8EG_aQGCFM0_INT_1_WpWp_lvqq	10000
503158	MGPy8EG_aQGCFM0_INT_1_WpZ_llqq	10000
503159	MGPy8EG_aQGCFM0_INT_1_WpZ_lvqq	10000
503160	MGPy8EG_aQGCFM0_INT_1_WpZ_vvqq	10000
503161	MGPy8EG_aQGCFM0_INT_1_ZZ_llqq	10000
503162	MGPy8EG_aQGCFM0_INT_1_ZZ_lvqq	10000
503163	MGPy8EG_aQGCFM0_QUAD_1_WmWm_lvqq	10000
503164	MGPy8EG_aQGCFM0_QUAD_1_WmZ_llqq	10000
503165	MGPy8EG_aQGCFM0_QUAD_1_WmZ_lvqq	10000
503166	MGPy8EG_aQGCFM0_QUAD_1_WmZ_vvqq	10000
503167	MGPy8EG_aQGCFM0_QUAD_1_WpWm_lvqq	20000
503168	MGPy8EG_aQGCFM0_QUAD_1_WpWp_lvqq	10000
503169	MGPy8EG_aQGCFM0_QUAD_1_WpZ_llqq	10000
503170	MGPy8EG_aQGCFM0_QUAD_1_WpZ_lvqq	10000
503171	MGPy8EG_aQGCFM0_QUAD_1_WpZ_vvqq	10000
503172	MGPy8EG_aQGCFM0_QUAD_1_ZZ_llqq	10000
503173	MGPy8EG_aQGCFM0_QUAD_1_ZZ_lvqq	10000
503174	MGPy8EG_aQGCFM1_INT_1_WmWm_lvqq	10000
503175	MGPy8EG_aQGCFM1_INT_1_WmZ_llqq	10000
503176	MGPy8EG_aQGCFM1_INT_1_WmZ_lvqq	10000
503177	MGPy8EG_aQGCFM1_INT_1_WmZ_vvqq	10000
503178	MGPy8EG_aQGCFM1_INT_1_WpWm_lvqq	20000
503179	MGPy8EG_aQGCFM1_INT_1_WpWp_lvqq	10000
503180	MGPy8EG_aQGCFM1_INT_1_WpZ_llqq	10000
503181	MGPy8EG_aQGCFM1_INT_1_WpZ_lvqq	10000
503182	MGPy8EG_aQGCFM1_INT_1_WpZ_vvqq	10000
503183	MGPy8EG_aQGCFM1_INT_1_ZZ_llqq	10000
503184	MGPy8EG_aQGCFM1_INT_1_ZZ_lvqq	10000
503185	MGPy8EG_aQGCFM1_QUAD_1_WmWm_lvqq	10000
503186	MGPy8EG_aQGCFM1_QUAD_1_WmZ_llqq	10000
503187	MGPy8EG_aQGCFM1_QUAD_1_WmZ_lvqq	10000
503188	MGPy8EG_aQGCFM1_QUAD_1_WmZ_vvqq	10000
503189	MGPy8EG_aQGCFM1_QUAD_1_WpWm_lvqq	20000
503190	MGPy8EG_aQGCFM1_QUAD_1_WpWp_lvqq	10000
503191	MGPy8EG_aQGCFM1_QUAD_1_WpZ_llqq	10000
503192	MGPy8EG_aQGCFM1_QUAD_1_WpZ_lvqq	10000

Table A.19: aQGC samples used in the analysis. The dataset ID, MC generator and total number of generated events are shown.

DSID	Name	Events
503193	MGPy8EG_aQGCFM1_QUAD_1_WpZ_vvqq	10000
503194	MGPy8EG_aQGCFM1_QUAD_1_ZZ_llqq	10000
503195	MGPy8EG_aQGCFM1_QUAD_1_ZZ_vvqq	10000
503196	MGPy8EG_aQGCFM2_INT_1_WmZ_llqq	10000
503197	MGPy8EG_aQGCFM2_INT_1_WmZ_lvqq	10000
503198	MGPy8EG_aQGCFM2_INT_1_WmZ_vvqq	10000
503199	MGPy8EG_aQGCFM2_INT_1_WpWm_lvqq	20000
503200	MGPy8EG_aQGCFM2_INT_1_WpZ_llqq	10000
503201	MGPy8EG_aQGCFM2_INT_1_WpZ_lvqq	10000
503202	MGPy8EG_aQGCFM2_INT_1_WpZ_vvqq	10000
503203	MGPy8EG_aQGCFM2_INT_1_ZZ_llqq	10000
503204	MGPy8EG_aQGCFM2_INT_1_ZZ_vvqq	10000
503205	MGPy8EG_aQGCFM2_QUAD_1_WmZ_llqq	10000
503206	MGPy8EG_aQGCFM2_QUAD_1_WmZ_lvqq	10000
503207	MGPy8EG_aQGCFM2_QUAD_1_WmZ_vvqq	10000
503208	MGPy8EG_aQGCFM2_QUAD_1_WpWm_lvqq	20000
503209	MGPy8EG_aQGCFM2_QUAD_1_WpZ_llqq	10000
503210	MGPy8EG_aQGCFM2_QUAD_1_WpZ_lvqq	10000
503211	MGPy8EG_aQGCFM2_QUAD_1_WpZ_vvqq	10000
503212	MGPy8EG_aQGCFM2_QUAD_1_ZZ_llqq	10000
503213	MGPy8EG_aQGCFM2_QUAD_1_ZZ_vvqq	10000
503214	MGPy8EG_aQGCFM3_INT_1_WmZ_llqq	10000
503215	MGPy8EG_aQGCFM3_INT_1_WmZ_lvqq	10000
503216	MGPy8EG_aQGCFM3_INT_1_WmZ_vvqq	10000
503217	MGPy8EG_aQGCFM3_INT_1_WpWm_lvqq	20000
503218	MGPy8EG_aQGCFM3_INT_1_WpZ_llqq	10000
503219	MGPy8EG_aQGCFM3_INT_1_WpZ_lvqq	10000
503220	MGPy8EG_aQGCFM3_INT_1_WpZ_vvqq	10000
503221	MGPy8EG_aQGCFM3_INT_1_ZZ_llqq	10000
503222	MGPy8EG_aQGCFM3_INT_1_ZZ_vvqq	10000
503223	MGPy8EG_aQGCFM3_QUAD_1_WmZ_llqq	10000
503224	MGPy8EG_aQGCFM3_QUAD_1_WmZ_lvqq	10000
503225	MGPy8EG_aQGCFM3_QUAD_1_WmZ_vvqq	10000
503226	MGPy8EG_aQGCFM3_QUAD_1_WpWm_lvqq	20000
503227	MGPy8EG_aQGCFM3_QUAD_1_WpZ_llqq	10000
503228	MGPy8EG_aQGCFM3_QUAD_1_WpZ_lvqq	10000
503229	MGPy8EG_aQGCFM3_QUAD_1_WpZ_vvqq	10000
503230	MGPy8EG_aQGCFM3_QUAD_1_ZZ_llqq	10000
503231	MGPy8EG_aQGCFM3_QUAD_1_ZZ_vvqq	10000
503232	MGPy8EG_aQGCFM4_INT_1_WmZ_llqq	10000

Table A.20: aQGC samples used in the analysis. The dataset ID, MC generator and total number of generated events are shown.

DSID	Name	Events
503233	MGPy8EG_aQGCFM4_INT_1_WmZ_lvqq	10000
503234	MGPy8EG_aQGCFM4_INT_1_WmZ_vvqq	10000
503235	MGPy8EG_aQGCFM4_INT_1_WpWm_lvqq	20000
503236	MGPy8EG_aQGCFM4_INT_1_WpZ_llqq	10000
503237	MGPy8EG_aQGCFM4_INT_1_WpZ_lvqq	10000
503238	MGPy8EG_aQGCFM4_INT_1_WpZ_vvqq	10000
503239	MGPy8EG_aQGCFM4_INT_1_ZZ_llqq	10000
503240	MGPy8EG_aQGCFM4_INT_1_ZZ_vvqq	10000
503241	MGPy8EG_aQGCFM4_QUAD_1_WmZ_llqq	10000
503242	MGPy8EG_aQGCFM4_QUAD_1_WmZ_lvqq	10000
503243	MGPy8EG_aQGCFM4_QUAD_1_WmZ_vvqq	10000
503244	MGPy8EG_aQGCFM4_QUAD_1_WpWm_lvqq	20000
503245	MGPy8EG_aQGCFM4_QUAD_1_WpZ_llqq	10000
503246	MGPy8EG_aQGCFM4_QUAD_1_WpZ_lvqq	10000
503247	MGPy8EG_aQGCFM4_QUAD_1_WpZ_vvqq	10000
503248	MGPy8EG_aQGCFM4_QUAD_1_ZZ_llqq	10000
503249	MGPy8EG_aQGCFM4_QUAD_1_ZZ_vvqq	10000
503250	MGPy8EG_aQGCFM5_INT_1_WmZ_llqq	10000
503251	MGPy8EG_aQGCFM5_INT_1_WmZ_lvqq	10000
503252	MGPy8EG_aQGCFM5_INT_1_WmZ_vvqq	10000
503253	MGPy8EG_aQGCFM5_INT_1_WpWm_lvqq	20000
503254	MGPy8EG_aQGCFM5_INT_1_WpZ_llqq	10000
503255	MGPy8EG_aQGCFM5_INT_1_WpZ_lvqq	10000
503256	MGPy8EG_aQGCFM5_INT_1_WpZ_vvqq	10000
503257	MGPy8EG_aQGCFM5_INT_1_ZZ_llqq	10000
503258	MGPy8EG_aQGCFM5_INT_1_ZZ_vvqq	10000
503259	MGPy8EG_aQGCFM5_QUAD_1_WmZ_llqq	10000
503260	MGPy8EG_aQGCFM5_QUAD_1_WmZ_lvqq	10000
503261	MGPy8EG_aQGCFM5_QUAD_1_WmZ_vvqq	10000
503262	MGPy8EG_aQGCFM5_QUAD_1_WpWm_lvqq	20000
503263	MGPy8EG_aQGCFM5_QUAD_1_WpZ_llqq	10000
503264	MGPy8EG_aQGCFM5_QUAD_1_WpZ_lvqq	10000
503265	MGPy8EG_aQGCFM5_QUAD_1_WpZ_vvqq	10000
503266	MGPy8EG_aQGCFM5_QUAD_1_ZZ_llqq	10000
503267	MGPy8EG_aQGCFM5_QUAD_1_ZZ_vvqq	10000
503268	MGPy8EG_aQGCFM7_INT_1_WmWm_lvqq	10000
503269	MGPy8EG_aQGCFM7_INT_1_WmZ_llqq	10000
503270	MGPy8EG_aQGCFM7_INT_1_WmZ_lvqq	10000
503271	MGPy8EG_aQGCFM7_INT_1_WmZ_vvqq	10000
503272	MGPy8EG_aQGCFM7_INT_1_WpWm_lvqq	19000

Table A.21: aQGC samples used in the analysis. The dataset ID, MC generator and total number of generated events are shown.

DSID	Name	Events
503273	MGPy8EG_aQGCFM7_INT_1_WpWp_lvqq	10000
503274	MGPy8EG_aQGCFM7_INT_1_WpZ_llqq	10000
503275	MGPy8EG_aQGCFM7_INT_1_WpZ_lvqq	10000
503276	MGPy8EG_aQGCFM7_INT_1_WpZ_vvqq	10000
503277	MGPy8EG_aQGCFM7_INT_1_ZZ_llqq	10000
503278	MGPy8EG_aQGCFM7_INT_1_ZZ_lvqq	10000
503279	MGPy8EG_aQGCFM7_QUAD_1_WmWm_lvqq	10000
503280	MGPy8EG_aQGCFM7_QUAD_1_WmZ_llqq	10000
503281	MGPy8EG_aQGCFM7_QUAD_1_WmZ_lvqq	10000
503282	MGPy8EG_aQGCFM7_QUAD_1_WmZ_vvqq	10000
503283	MGPy8EG_aQGCFM7_QUAD_1_WpWm_lvqq	20000
503284	MGPy8EG_aQGCFM7_QUAD_1_WpWp_lvqq	10000
503285	MGPy8EG_aQGCFM7_QUAD_1_WpZ_llqq	10000
503286	MGPy8EG_aQGCFM7_QUAD_1_WpZ_lvqq	10000
503287	MGPy8EG_aQGCFM7_QUAD_1_WpZ_vvqq	10000
503288	MGPy8EG_aQGCFM7_QUAD_1_ZZ_llqq	10000
503289	MGPy8EG_aQGCFM7_QUAD_1_ZZ_vvqq	10000
503290	MGPy8EG_aQGCFS02_INT_1_WmWm_lvqq	10000
503291	MGPy8EG_aQGCFS02_INT_1_WmZ_llqq	10000
503292	MGPy8EG_aQGCFS02_INT_1_WmZ_lvqq	10000
503293	MGPy8EG_aQGCFS02_INT_1_WmZ_vvqq	10000
503294	MGPy8EG_aQGCFS02_INT_1_WpWm_lvqq	20000
503295	MGPy8EG_aQGCFS02_INT_1_WpWp_lvqq	10000
503296	MGPy8EG_aQGCFS02_INT_1_WpZ_llqq	10000
503297	MGPy8EG_aQGCFS02_INT_1_WpZ_lvqq	10000
503298	MGPy8EG_aQGCFS02_INT_1_WpZ_vvqq	10000
503299	MGPy8EG_aQGCFS02_INT_1_ZZ_llqq	10000
503300	MGPy8EG_aQGCFS02_INT_1_ZZ_vvqq	10000
503301	MGPy8EG_aQGCFS02_QUAD_1_WmWm_lvqq	10000
503302	MGPy8EG_aQGCFS02_QUAD_1_WmZ_llqq	10000
503303	MGPy8EG_aQGCFS02_QUAD_1_WmZ_lvqq	10000
503304	MGPy8EG_aQGCFS02_QUAD_1_WmZ_vvqq	10000
503305	MGPy8EG_aQGCFS02_QUAD_1_WpWm_lvqq	20000
503306	MGPy8EG_aQGCFS02_QUAD_1_WpWp_lvqq	10000
503307	MGPy8EG_aQGCFS02_QUAD_1_WpZ_llqq	10000
503308	MGPy8EG_aQGCFS02_QUAD_1_WpZ_lvqq	10000
503309	MGPy8EG_aQGCFS02_QUAD_1_WpZ_vvqq	10000
503310	MGPy8EG_aQGCFS02_QUAD_1_ZZ_llqq	10000
503311	MGPy8EG_aQGCFS02_QUAD_1_ZZ_vvqq	10000
503312	MGPy8EG_aQGCFS1_INT_1_WmWm_lvqq	10000

Table A.22: aQGC samples used in the analysis. The dataset ID, MC generator and total number of generated events are shown.

DSID	Name	Events
503313	MGPy8EG_aQGCFS1_INT_1_WmZ_llqq	10000
503314	MGPy8EG_aQGCFS1_INT_1_WmZ_lvqq	10000
503315	MGPy8EG_aQGCFS1_INT_1_WmZ_vvqq	10000
503316	MGPy8EG_aQGCFS1_INT_1_WpWm_lvqq	19000
503317	MGPy8EG_aQGCFS1_INT_1_WpWp_lvqq	10000
503318	MGPy8EG_aQGCFS1_INT_1_WpZ_llqq	10000
503319	MGPy8EG_aQGCFS1_INT_1_WpZ_lvqq	10000
503320	MGPy8EG_aQGCFS1_INT_1_WpZ_vvqq	10000
503321	MGPy8EG_aQGCFS1_INT_1_ZZ_llqq	10000
503322	MGPy8EG_aQGCFS1_INT_1_ZZ_vvqq	10000
503323	MGPy8EG_aQGCFS1_QUAD_1_WmWm_lvqq	10000
503324	MGPy8EG_aQGCFS1_QUAD_1_WmZ_llqq	10000
503325	MGPy8EG_aQGCFS1_QUAD_1_WmZ_lvqq	10000
503326	MGPy8EG_aQGCFS1_QUAD_1_WmZ_vvqq	10000
503327	MGPy8EG_aQGCFS1_QUAD_1_WpWm_lvqq	20000
503328	MGPy8EG_aQGCFS1_QUAD_1_WpWp_lvqq	10000
503329	MGPy8EG_aQGCFS1_QUAD_1_WpZ_llqq	10000
503330	MGPy8EG_aQGCFS1_QUAD_1_WpZ_lvqq	10000
503331	MGPy8EG_aQGCFS1_QUAD_1_WpZ_vvqq	10000
503332	MGPy8EG_aQGCFS1_QUAD_1_ZZ_llqq	10000
503333	MGPy8EG_aQGCFS1_QUAD_1_ZZ_vvqq	10000
503334	MGPy8EG_aQGCFT0_INT_1_WmWm_lvqq	10000
503335	MGPy8EG_aQGCFT0_INT_1_WmZ_llqq	10000
503336	MGPy8EG_aQGCFT0_INT_1_WmZ_lvqq	10000
503337	MGPy8EG_aQGCFT0_INT_1_WmZ_vvqq	10000
503338	MGPy8EG_aQGCFT0_INT_1_WpWm_lvqq	20000
503339	MGPy8EG_aQGCFT0_INT_1_WpWp_lvqq	10000
503340	MGPy8EG_aQGCFT0_INT_1_WpZ_llqq	10000
503341	MGPy8EG_aQGCFT0_INT_1_WpZ_lvqq	10000
503342	MGPy8EG_aQGCFT0_INT_1_WpZ_vvqq	10000
503343	MGPy8EG_aQGCFT0_INT_1_ZZ_llqq	9000
503344	MGPy8EG_aQGCFT0_INT_1_ZZ_vvqq	10000
503345	MGPy8EG_aQGCFT0_QUAD_1_WmWm_lvqq	10000
503346	MGPy8EG_aQGCFT0_QUAD_1_WmZ_llqq	10000
503347	MGPy8EG_aQGCFT0_QUAD_1_WmZ_lvqq	10000
503348	MGPy8EG_aQGCFT0_QUAD_1_WmZ_vvqq	10000
503349	MGPy8EG_aQGCFT0_QUAD_1_WpWm_lvqq	20000
503350	MGPy8EG_aQGCFT0_QUAD_1_WpWp_lvqq	10000
503351	MGPy8EG_aQGCFT0_QUAD_1_WpZ_llqq	10000
503352	MGPy8EG_aQGCFT0_QUAD_1_WpZ_lvqq	10000

Table A.23: aQGC samples used in the analysis. The dataset ID, MC generator and total number of generated events are shown.

DSID	Name	Events
503353	MGPy8EG_aQGCFT0_QUAD_1_WpZ_vvqq	10000
503354	MGPy8EG_aQGCFT0_QUAD_1_ZZ_llqq	10000
503355	MGPy8EG_aQGCFT0_QUAD_1_ZZ_vvqq	10000
503356	MGPy8EG_aQGCFT1_INT_1_WmWm_lvqq	10000
503357	MGPy8EG_aQGCFT1_INT_1_WmZ_llqq	10000
503358	MGPy8EG_aQGCFT1_INT_1_WmZ_lvqq	10000
503359	MGPy8EG_aQGCFT1_INT_1_WmZ_vvqq	10000
503360	MGPy8EG_aQGCFT1_INT_1_WpWm_lvqq	20000
503361	MGPy8EG_aQGCFT1_INT_1_WpWp_lvqq	10000
503362	MGPy8EG_aQGCFT1_INT_1_WpZ_llqq	10000
503363	MGPy8EG_aQGCFT1_INT_1_WpZ_lvqq	10000
503364	MGPy8EG_aQGCFT1_INT_1_WpZ_vvqq	10000
503365	MGPy8EG_aQGCFT1_INT_1_ZZ_llqq	10000
503366	MGPy8EG_aQGCFT1_INT_1_ZZ_vvqq	10000
503367	MGPy8EG_aQGCFT1_QUAD_1_WmWm_lvqq	10000
503368	MGPy8EG_aQGCFT1_QUAD_1_WmZ_llqq	10000
503369	MGPy8EG_aQGCFT1_QUAD_1_WmZ_lvqq	10000
503370	MGPy8EG_aQGCFT1_QUAD_1_WmZ_vvqq	10000
503371	MGPy8EG_aQGCFT1_QUAD_1_WpWm_lvqq	20000
503372	MGPy8EG_aQGCFT1_QUAD_1_WpWp_lvqq	10000
503373	MGPy8EG_aQGCFT1_QUAD_1_WpZ_llqq	10000
503374	MGPy8EG_aQGCFT1_QUAD_1_WpZ_lvqq	10000
503375	MGPy8EG_aQGCFT1_QUAD_1_WpZ_vvqq	10000
503376	MGPy8EG_aQGCFT1_QUAD_1_ZZ_llqq	10000
503377	MGPy8EG_aQGCFT1_QUAD_1_ZZ_vvqq	10000
503378	MGPy8EG_aQGCFT2_INT_1_WmWm_lvqq	10000
503379	MGPy8EG_aQGCFT2_INT_1_WmZ_llqq	10000
503380	MGPy8EG_aQGCFT2_INT_1_WmZ_lvqq	10000
503381	MGPy8EG_aQGCFT2_INT_1_WmZ_vvqq	10000
503382	MGPy8EG_aQGCFT2_INT_1_WpWm_lvqq	20000
503383	MGPy8EG_aQGCFT2_INT_1_WpWp_lvqq	10000
503384	MGPy8EG_aQGCFT2_INT_1_WpZ_llqq	10000
503385	MGPy8EG_aQGCFT2_INT_1_WpZ_lvqq	10000
503386	MGPy8EG_aQGCFT2_INT_1_WpZ_vvqq	10000
503387	MGPy8EG_aQGCFT2_INT_1_ZZ_llqq	10000
503388	MGPy8EG_aQGCFT2_INT_1_ZZ_vvqq	10000
503389	MGPy8EG_aQGCFT2_QUAD_1_WmWm_lvqq	10000
503390	MGPy8EG_aQGCFT2_QUAD_1_WmZ_llqq	10000
503391	MGPy8EG_aQGCFT2_QUAD_1_WmZ_lvqq	10000
503392	MGPy8EG_aQGCFT2_QUAD_1_WmZ_vvqq	10000

Table A.24: aQGC samples used in the analysis. The dataset ID, MC generator and total number of generated events are shown.

DSID	Name	Events
503393	MGPy8EG_aQGCFT2_QUAD_1_WpWm_lvqq	20000
503394	MGPy8EG_aQGCFT2_QUAD_1_WpWp_lvqq	10000
503395	MGPy8EG_aQGCFT2_QUAD_1_WpZ_llqq	10000
503396	MGPy8EG_aQGCFT2_QUAD_1_WpZ_lvqq	10000
503397	MGPy8EG_aQGCFT2_QUAD_1_WpZ_vvqq	10000
503398	MGPy8EG_aQGCFT2_QUAD_1_ZZ_llqq	10000
503399	MGPy8EG_aQGCFT2_QUAD_1_ZZ_vvqq	10000
503400	MGPy8EG_aQGCFT5_INT_1_WmZ_llqq	10000
503401	MGPy8EG_aQGCFT5_INT_1_WmZ_lvqq	10000
503402	MGPy8EG_aQGCFT5_INT_1_WmZ_vvqq	10000
503403	MGPy8EG_aQGCFT5_INT_1_WpWm_lvqq	19000
503404	MGPy8EG_aQGCFT5_INT_1_WpZ_llqq	10000
503405	MGPy8EG_aQGCFT5_INT_1_WpZ_lvqq	10000
503406	MGPy8EG_aQGCFT5_INT_1_WpZ_vvqq	10000
503407	MGPy8EG_aQGCFT5_INT_1_ZZ_llqq	10000
503408	MGPy8EG_aQGCFT5_INT_1_ZZ_vvqq	10000
503409	MGPy8EG_aQGCFT5_QUAD_1_WmZ_llqq	10000
503410	MGPy8EG_aQGCFT5_QUAD_1_WmZ_lvqq	10000
503411	MGPy8EG_aQGCFT5_QUAD_1_WmZ_vvqq	10000
503412	MGPy8EG_aQGCFT5_QUAD_1_WpWm_lvqq	20000
503413	MGPy8EG_aQGCFT5_QUAD_1_WpZ_llqq	10000
503414	MGPy8EG_aQGCFT5_QUAD_1_WpZ_lvqq	10000
503415	MGPy8EG_aQGCFT5_QUAD_1_WpZ_vvqq	10000
503416	MGPy8EG_aQGCFT5_QUAD_1_ZZ_llqq	10000
503417	MGPy8EG_aQGCFT5_QUAD_1_ZZ_vvqq	10000
503418	MGPy8EG_aQGCFT6_INT_1_WmZ_llqq	10000
503419	MGPy8EG_aQGCFT6_INT_1_WmZ_lvqq	10000
503420	MGPy8EG_aQGCFT6_INT_1_WmZ_vvqq	10000
503421	MGPy8EG_aQGCFT6_INT_1_WpWm_lvqq	20000
503422	MGPy8EG_aQGCFT6_INT_1_WpZ_llqq	10000
503423	MGPy8EG_aQGCFT6_INT_1_WpZ_lvqq	10000
503424	MGPy8EG_aQGCFT6_INT_1_WpZ_vvqq	10000
503425	MGPy8EG_aQGCFT6_INT_1_ZZ_llqq	10000
503426	MGPy8EG_aQGCFT6_INT_1_ZZ_vvqq	10000
503427	MGPy8EG_aQGCFT6_QUAD_1_WmZ_llqq	10000
503428	MGPy8EG_aQGCFT6_QUAD_1_WmZ_lvqq	10000
503429	MGPy8EG_aQGCFT6_QUAD_1_WmZ_vvqq	10000
503430	MGPy8EG_aQGCFT6_QUAD_1_WpWm_lvqq	20000
503431	MGPy8EG_aQGCFT6_QUAD_1_WpZ_llqq	10000
503432	MGPy8EG_aQGCFT6_QUAD_1_WpZ_lvqq	10000

Table A.25: aQGC samples used in the analysis. The dataset ID, MC generator and total number of generated events are shown.

DSID	Name	Events
503433	MGPy8EG_aQGCFT6_QUAD_1_WpZ_vvqq	10000
503434	MGPy8EG_aQGCFT6_QUAD_1_ZZ_llqq	10000
503435	MGPy8EG_aQGCFT6_QUAD_1_ZZ_vvqq	10000
503436	MGPy8EG_aQGCFT7_INT_1_WmZ_llqq	10000
503437	MGPy8EG_aQGCFT7_INT_1_WmZ_lvqq	10000
503438	MGPy8EG_aQGCFT7_INT_1_WmZ_vvqq	10000
503439	MGPy8EG_aQGCFT7_INT_1_WpWm_lvqq	20000
503440	MGPy8EG_aQGCFT7_INT_1_WpZ_llqq	9000
503441	MGPy8EG_aQGCFT7_INT_1_WpZ_lvqq	10000
503442	MGPy8EG_aQGCFT7_INT_1_WpZ_vvqq	10000
503443	MGPy8EG_aQGCFT7_INT_1_ZZ_llqq	10000
503444	MGPy8EG_aQGCFT7_INT_1_ZZ_vvqq	10000
503445	MGPy8EG_aQGCFT7_QUAD_1_WmZ_llqq	10000
503446	MGPy8EG_aQGCFT7_QUAD_1_WmZ_lvqq	10000
503447	MGPy8EG_aQGCFT7_QUAD_1_WmZ_vvqq	10000
503448	MGPy8EG_aQGCFT7_QUAD_1_WpWm_lvqq	20000
503449	MGPy8EG_aQGCFT7_QUAD_1_WpZ_llqq	10000
503450	MGPy8EG_aQGCFT7_QUAD_1_WpZ_lvqq	10000
503451	MGPy8EG_aQGCFT7_QUAD_1_WpZ_vvqq	10000
503452	MGPy8EG_aQGCFT7_QUAD_1_ZZ_llqq	10000
503453	MGPy8EG_aQGCFT7_QUAD_1_ZZ_vvqq	10000
503454	MGPy8EG_aQGCFT8_INT_1_ZZ_llqq	10000
503455	MGPy8EG_aQGCFT8_INT_1_ZZ_vvqq	10000
503456	MGPy8EG_aQGCFT8_QUAD_1_ZZ_llqq	10000
503457	MGPy8EG_aQGCFT8_QUAD_1_ZZ_vvqq	10000
503458	MGPy8EG_aQGCFT9_INT_1_ZZ_llqq	20000
503459	MGPy8EG_aQGCFT9_INT_1_ZZ_vvqq	20000
503460	MGPy8EG_aQGCFT9_QUAD_1_ZZ_llqq	10000
503461	MGPy8EG_aQGCFT9_QUAD_1_ZZ_vvqq	10000

References

- [1] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321–323. 1
- [2] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508–509. 1
- [3] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, *Phys. Rev. Lett.* **13** (1964) 585–587. 1
- [4] ATLAS Collaboration, *Observation of electroweak production of two jets and a Z-boson pair with the ATLAS detector at the LHC*, 2004.10612. 1
- [5] ATLAS Collaboration, *Observation of Electroweak Production of a Same-Sign W Boson Pair in Association with Two Jets in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, *Phys. Rev. Lett.* **123** (2019) 161801, [1906.03203]. 1
- [6] ATLAS Collaboration, *Observation of electroweak $W^\pm Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **793** (2019) 469, [1812.09740]. 1
- [7] CMS Collaboration, *Evidence for electroweak production of four charged leptons and two jets in proton-proton collisions at $s=13$ tev*, *Physics Letters B* **812** (2021) 135992. 1
- [8] CMS Collaboration, *Evidence for ww/wz vector boson scattering in the decay channel $lvqq$ produced in association with two jets in proton-proton collisions at $s=13$ tev*, *Physics Letters B* **834** (2022) 137438. 1
- [9] A. C. Longhitano, *Low-Energy Impact of a Heavy Higgs Boson Sector*, *Nucl. Phys. B* **188** (1981) 118. 1
- [10] D. Charlton, T. Kono, L. Iconomidou-Fayard, F. Conventi, C. Zhu, K. Nikolopoulos et al., *Study of electroweak WW/WZ/ZZ production in the semileptonic final states in association with a high-mass dijet system and limits on anomalous quartic gauge couplings with 13 TeV full run-2 ATLAS data*, tech. rep., CERN, Geneva, 2022. 2, 131
- [11] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*. Addison-Wesley, Reading, USA, 1995. 3
- [12] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132–133. 4, 6
- [13] C. Burgard, “TikZ diagram of the Standard Model of physics.” <http://www.texample.net/tikz/examples/model-physics/>, 2019. 4
- [14] J. Campbell, J. Huston and F. Krauss, *The Black Book of Quantum Chromodynamics*. Oxford University Press, 2017. 4

- [15] S. L. Glashow, *Partial-symmetries of weak interactions*, *Nuclear Physics* **22** (1961) 579–588. 5
- [16] J. Goldstone, A. Salam and S. Weinberg, *Broken symmetries*, *Phys. Rev.* **127** (Aug, 1962) 965–970. 5
- [17] S. Weinberg, *A model of leptons*, *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266. 5
- [18] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509. 6
- [19] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323. 6
- [20] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global conservation laws and massless particles*, *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587. 6
- [21] J. Ellis, “Higgs physics.” Lecture Notes in 2013 European School of High-Energy Physics, 2013. 10.5170/CERN-2015-004.117. 7
- [22] T. Plehn and M. Rauch, *Quartic higgs coupling at hadron colliders*, *Physical Review D* **72** (Sept., 2005) . 7
- [23] H. Yukawa, *On the Interaction of Elementary Particles. I*, *Progress of Theoretical Physics Supplement* **1** (01, 1955) 1–10, [<https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.1.1/5310694/1-1.pdf>]. 8
- [24] The 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* **716** (2012) 1–29. 8
- [25] CMS Collaboration, *Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc*, *Physics Letters B* **716** (2012) 30–61. 8
- [26] D. Binosi and L. TheuBl, *Jaxodraw: A graphical user interface for drawing feynman diagrams*, *Computer Physics Communications* **161** (2004) 76–86. 9
- [27] S. Weinberg, *Phenomenological Lagrangians*, *Physica A* **96** (1979) 327–340. 13
- [28] S. Weinberg, *Conceptual foundations of the unified theory of weak and electromagnetic interactions*, *Rev. Mod. Phys.* **52** (Jul, 1980) 515–523. 13
- [29] K. Gaemers and G. Gounaris, *Polarization amplitudes for $e^+e^- \rightarrow w^+w^-$ and $e^+e^- \rightarrow zz$* , *Zeitschrift für Physik C Particles and Fields* **1** (Sept., 1979) 259–268. 13
- [30] C. Degrande, N. Greiner, W. Kilian, O. Mattelaer, H. Mebane, T. Stelzer et al., *Effective field theory: A modern approach to anomalous couplings*, *Annals of Physics* **335** (Aug., 2013) 21–32. 13
- [31] O. J. Éboli, M. Gonzalez-Garcia and J. Mizukoshi, *$p p \rightarrow j j e^\pm\mu^\pm\nu\nu$ and $j j e^\pm\mu^\pm\nu\nu$ at o (α_{em} 6) and o (α_{em} 4 α_s 2) for the study of the quartic electroweak gauge boson vertex at cern lhc*, *Physical Review D* **74** (2006) 073005. 14
- [32] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole et al., *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. 16
- [33] The ATLAS Collaboration, *The atlas experiment at the cern large hadron collider*, *Journal of Instrumentation* **3** (aug, 2008) S08003. 16, 19, 24
- [34] The CMS Collaboration, *The cms experiment at the cern lhc*, *Journal of Instrumentation* **3** (aug, 2008) S08004. 16
- [35] The LHCb Collaboration, *The lhcb detector at the lhc*, *Journal of Instrumentation* **3** (aug, 2008) S08005. 16

- [36] T. A. Collaboration, *The alice experiment at the cern lhc*, *Journal of Instrumentation* **3** (aug, 2008) S08002. 16
- [37] L. Evans and P. Bryant, *Lhc machine*, *Journal of Instrumentation* **3** (aug, 2008) S08001. 16, 17
- [38] L. Rossi, *The lhc superconducting magnets*, in *Proceedings of the 2003 Particle Accelerator Conference*, vol. 1, pp. 141–145 Vol.1, 2003. DOI. 16
- [39] L. Rossi, *Superconducting magnets for the lhc main lattice*, *IEEE Transactions on Applied Superconductivity* **14** (2004) 153–158. 16
- [40] A. Team, “Diagram of an LHC dipole magnet. Schéma d'un aimant dipôle du LHC.” 1999. 17
- [41] J.-P. Burnet, C. Carli, M. Chanel, R. Garoby, S. Gilardoni, M. Giovannozzi et al., *Fifty years of the CERN Proton Synchrotron: Volume 1*. CERN Yellow Reports: Monographs. CERN, Geneva, 2011, 10.5170/CERN-2011-004. 16
- [42] J. Haffner, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, . 18
- [43] J. Pequenao, “Computer generated image of the whole ATLAS detector.” 2008. 19, 20
- [44] I. Neutelings, “The lhc coordinate system as seen from the atlas detector.” https://tikz.net/axis3d_cms/, 2023. 21
- [45] ATLAS collaboration, M. S. Alam et al., *ATLAS pixel detector: Technical design report*, . 20
- [46] J. Pequenao, “Computer generated image of the ATLAS inner detector.” 2008. 22
- [47] The ATLAS TRT collaboration, *The atlas transition radiation tracker (trt) proportional drift tube: design and performance*, *Journal of Instrumentation* **3** (feb, 2008) P02013. 21
- [48] J. Pequenao, “Computer Generated image of the ATLAS calorimeter.” 2008. 23
- [49] ATLAS collaboration, *ATLAS liquid argon calorimeter: Technical design report*, . 23
- [50] ATLAS collaboration, *ATLAS tile calorimeter: Technical design report*, . 23
- [51] A. Artamonov, D. Bailey, G. Belanger, M. Cadabeschi, T. Y. Chen, V. Epshteyn et al., *The atlas forward calorimeter*, *Journal of Instrumentation* **3** (feb, 2008) P02010. 24
- [52] ATLAS collaboration, *ATLAS muon spectrometer: Technical design report*, . 24
- [53] J. Pequenao, “Computer generated image of the ATLAS Muons subsystem.” 2008. 25
- [54] ATLAS collaboration, *ATLAS level-1 trigger: Technical Design Report*, . 26
- [55] ATLAS collaboration, *ATLAS high-level trigger, data acquisition and controls: Technical design report*, . 26
- [56] B. Abbott, R. Blair, G. Crone, B. Green, J. Love, J. Proudfoot et al., *The evolution of the region of interest builder for the atlas experiment at cern*, *Journal of Instrumentation* **11** (feb, 2016) C02080. 26
- [57] M. Tamsett, *Performance and improvements of the atlas jet trigger system*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **718** (2013) 253–254. 27
- [58] ATLAS Collaboration, *The performance of the jet trigger for the atlas detector during 2011 data taking*, *The European Physical Journal C* **76** (2016) 526. 27
- [59] The ATLAS collaboration, *The atlas fast tracker system*, *Journal of Instrumentation* **16** (jul, 2021) P07006. 27, 30, 31, 32

- [60] M. Atkinson, *Search for Diboson resonances in 8 TeV and 13 TeV proton-proton collisions at the Large Hadron Collider with the ATLAS detector*. PhD thesis, University of Illinois at Urbana-Champaign, 7, 2019. 29
- [61] ATLAS collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, . 35
- [62] G. Avoni, M. Bruschi, G. Cabras, D. Caforio, N. Dehghanian, A. Floderus et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, *JINST* **13** (2018) P07017. 35
- [63] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152–155. 35, 36
- [64] T. Sjostrand, *PYTHIA 8 Status Report*, 0809.0303. 35
- [65] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, [1005.4568]. 35
- [66] GEANT4 collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250–303. 35
- [67] T. Gleisberg, S. Höche, F. Krauss, M. Schönher, S. Schumann et al., *Event generation with SHERPA 1.1*, *JHEP* **0902** (2009) 007, [0811.4622]. 36
- [68] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **0812** (2008) 039, [0808.3674]. 36
- [69] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, [1111.5206]. 36
- [70] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **0803** (2008) 038, [0709.1027]. 36
- [71] S. Höche, F. Krauss, M. Schönher and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, *JHEP* **04** (2013) 027, [1207.5030]. 36
- [72] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte et al., *Parton distributions for the lhc run ii*, *Journal of High Energy Physics* **2015** (Apr., 2015) . 36
- [73] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, 1405.0301. 36, 37
- [74] NNPDF collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [1410.8849]. 36
- [75] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244–289, [1207.1303]. 36, 37, 39
- [76] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [0710.3820]. 36, 37, 39
- [77] L. Lönnblad, *Correcting the Colour-Dipole Cascade Model with Fixed Order Matrix Elements*, *JHEP* **05** (2002) 046, [hep-ph/0112284]. 36
- [78] L. Lönnblad and S. Prestel, *Matching tree-level matrix elements with interleaved showers*, *JHEP* **03** (2012) 019, [1109.4829]. 36
- [79] ATLAS Pythia 8 tunes to 7 TeV data, tech. rep., CERN, Geneva, 2014. 36

- [80] C. Anastasiou, L. J. Dixon, K. Melnikov and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, *Phys. Rev. D* **69** (2004) 094008, [[hep-ph/0312266](#)]. 36
- [81] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **1006** (2010) 043, [[1002.2581](#)]. 36
- [82] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **1303** (2013) 015, [[1212.3460](#)]. 36, 39
- [83] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Compt/ Phys. Commun.* (2014) 2930, [[1112.5675](#)]. 36
- [84] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, *Phys. Rev.* **D83** (2011) 091503, [[1103.2792](#)]. 36
- [85] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, *Phys. Rev.* **D81** (2010) 054028, [[1001.5034](#)]. 36
- [86] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, *Phys. Rev.* **D82** (2010) 054018, [[1005.4451](#)]. 36
- [87] ATLAS collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, tech. rep., CERN, Geneva, 2016. 36
- [88] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, [[1905.09127](#)]. 36
- [89] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049, [[1111.1220](#)]. 36
- [90] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, *QCD Matrix Elements + Parton Showers*, *JHEP* **11** (2001) 063, [[hep-ph/0109231](#)]. 36
- [91] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053, [[0903.1219](#)]. 36
- [92] ATLAS Collaboration. TWiki. 41
- [93] ATLAS Collaboration, *Studies of the muon momentum calibration and performance of the atlas detector with pp collisions at $\sqrt{s} = 13$ tev*, *The European Physical Journal C* **83** (2023) 686. 42
- [94] M. Cacciari, G. P. Salam and G. Soyez, *The anti-kt jet clustering algorithm*, *Journal of High Energy Physics* **2008** (apr, 2008) 063. 43
- [95] *Tagging and suppression of pileup jets with the ATLAS detector*, tech. rep., CERN, Geneva, 2014. 43
- [96] ATLAS collaboration, *Forward jet vertex tagging using the particle flow algorithm*, tech. rep., CERN, Geneva, 2019. 43
- [97] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **1002** (2010) 084, [[0912.1342](#)]. 44
- [98] A. J. Larkoski, I. Moult and D. Neill, *Power Counting to Better Jet Observables*, *JHEP* **12** (2014) 009, [[1409.6298](#)]. 44
- [99] A. J. Larkoski, I. Moult and D. Neill, *Analytic boosted boson discrimination*, [1507.03018](#). 44
- [100] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **78** (2018) 903, [[1802.08168](#)]. 46

- [101] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019) 970, [1907.05120]. 46
- [102] ATLAS Collaboration, *Atlas flavour-tagging algorithms for the lhc run 2 pp collision dataset*, *The European Physical Journal C* **83** (2023) 681. 46
- [103] S. Farrell, *Associationutils - overlap removal tools*, 2023. 46
- [104] D. Bachas, A. Buzatu, G. Carlino, G. Chiodini, F. Conventi, Y. El Ghazali et al., *Search for heavy WW/WZ/ZZ resonances in semi-leptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., CERN, Geneva, Oct, 2018. 50
- [105] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector*, *Phys. Lett. B* **786** (2018) 59, [1808.08238]. 50
- [106] A. Ryzhov, A. Myagkov, C.-y. Lin, D. Tsybychev, V. Pascuzzi, R. Les et al., *Study of electroweak WW/WZ/ZZ production in the semileptonic final states in association with a high-mass dijet system and limits on anomalous quartic gauge couplings with 13 TeV ATLAS data*, tech. rep., CERN, Geneva, Mar, 2018. 52, 54
- [107] F. Chollet et al., *Keras*, 2015. 82
- [108] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. 82
- [109] S. M. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, in *Advances in Neural Information Processing Systems*, pp. 4–7, 2017. DOI. 82
- [110] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **117** (2016) 182002, [1606.02625]. 97
- [111] ATLAS collaboration, *Object selections for analyses involving a Higgs boson produced in association with a vector boson in which $h \rightarrow b\bar{b}$ and V decays leptonically with Run-2 data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV*, ATL-COM-PHYS-2016-480 (2016) . 97
- [112] ATLAS collaboration, *Quark versus Gluon Jet Tagging Using Charged Particle Multiplicity with the ATLAS Detector*, tech. rep., CERN, Geneva, 2017. 100
- [113] A. Collaboration, *Measurement of the charged-particle multiplicity inside jets from $\sqrt{s} = 8$ tev pp collisions with the atlas detector*, *The European Physical Journal C* **76** (jun, 2016) . 100
- [114] D. Buscher, *Search for Higgs bosons with b-jets in the final state in proton-proton collisions with the ATLAS experiment*, 2016. 116
- [115] N. Morange, *Beautiful Higgs Boson*. Habilitation à diriger des recherches, Université Paris-Saclay, July, 2021. 121