UNIVERSITY OF CALIFORNIA
SANTA CRUZ

MEASUREMENT OF COLLINEAR W-BOSON PRODUCTION WITH HIGH TRANSVERSE MOMENTUM JETS AT $\sqrt{s}$= 13 TEV USING THE ATLAS DETECTOR

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

Yuzhan Zhao

June 2024

The Dissertation of Yuzhan Zhao
is approved:

Bruce Schumm, Chair

Abraham Seiden

Michael Hance

Matthew Gignac

Peter Biehl
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

List of Tables

Abstract

Measurement of collinear W-boson production with high transverse momentum jets at $\sqrt{s}$= 13 TeV using the ATLAS detector

by

Yuzhan Zhao

This thesis presents a measurement of collinear W-boson radiation from high-momentum jets using the full Run-2 dataset collected by the ATLAS collaboration at $\sqrt{s}$ = 13 TeV, with an integrated luminosity of 140 fb^{-1}. The W+ jets signal, identified through leptonic decays in the electron or muon channels, focuses on events with small angular separation between the jet and lepton, often involving numerous QCD emissions. Accurate modeling of this phase-space requires sophisticated multi-jet merged setups. Over 86,000 W-boson candidates were measured across various differential distributions with a precision of 10-20%, and the data was compared against newly developed state-of-the-art generator predictions using Sherpa 2.2.11 and MadGraph5_aMC@NLO with FxFx multi-jet merged setups, which are accurate to next-to-leading order in the strong and weak coupling constants.

Acknowledgments

First, I would like to thank my advisor, Bruce Schumm, who has provided invaluable advice and guidance throughout my graduate work. Bruce has offered strong support for all my research projects, giving me advice on pursuing an academic path and preparing me to be a well-trained and well-prepared scientist. He also assisted significantly by providing funding and support for me to participate in conferences and workshops that I was interested in.

I would also like to thank Hartmut Sadrozinski and Abraham Seiden for giving me a wonderful opportunity to pursue research in instrumentation and detector technology. The skills and experience I gained while working in the Ultra-Fast Silicon Detector (UFSD) group were invaluable. Additionally, Hartmut provided not only research support but also valuable advice and discussions that helped shape my academic path.

I would like to thank Matthew Gignac for his invaluable training and assistance in starting and learning how to perform ATLAS analysis. Matthew was very patient in training me for the analysis, providing invaluable guidance and assistance in tackling problems. Without Matthew's help, I would not have been able to complete this analysis work. I would also like to thank Giordon Stark for his critical support and mentorship throughout my analysis. It was a pleasure working with Giordon, and I learned many computational skills and experimental techniques that will be very useful in my future career.

Finally, I want to thank my family for supporting me in pursuing what I wanted to do

and always being there as a strong backing in my life. None of the accomplishments I have achieved would have been possible without their financial and emotional support.

Chapter 1

Introduction

The Standard Model (SM) of particle physics has been successfully describing the electromagnetic, weak, and strong interaction between elementary particles since it was created in the late 1960s. The unification of the electromagnetic and weak interaction, the so called electroweak (EW) interaction, provides a prediction and theoretical description of the W boson and its interaction properties. The production of W boson in the high energy proton-proton collision experiment at the Large Hadron Collider (LHC) is one of the dominant SM process. Given the large amount of energy accessible to the LHC collisions, the W boson is often produced in association with quarks and gluons in the final state. These quarks and gluons will undergo a hadronization process to form composite subatomic particles and deposit energy clusters within the detector, and the energy clusters are reassembled as jets. Hence, W+jets production is one of the dominant SM process in high energy proton-proton collisions, and the precise measurement of W+jets production is an important way to probe the fundamental structure of both the electroweak and strong interaction. Moreover, W+jets production is a significant irreducible SM background to diverse range of analyses within the LHC physics program: For example, Higgs production in association with a W boson [1, 2], top quarks measurements [3, 4], and searches for new physics phenomena beyond the SM [5–7].

This thesis reports on measurement of W boson production in association with a high transverse momentum jet, with a specific focus on the region in phase-space where the angular separation between this jet and the associated W-boson is small. Such regions of phase-space are dominated by multi-jet events with electroweak (EW) emissions of a

relatively soft W-boson that are typically produced with many additional jets in the final state. The W+jets production in this phase-space had been measured previously, but with a lower center-of-mass energy of 8 TeV, and only the muon channel was measured [8]. This thesis extends this previous measurement, with much higher center-of-mass energy at 13 TeV, making use of both the electron and muon channels and presents differential cross section for a number of final-state observables. The final results of both channels are combined.

In addition to the previous 8 TeV W+jets measurement, a 13 TeV measurement of collinear Z+jets production has also been performed with the full Run-2 dataset [9]. The W+jets measurement reported in this thesis will be largely complementary to the collinear Z+jets measurement, but benefits from a 10 times higher production cross-section compared to Z+jets. This higher cross-section enables probing higher energy tails of final-state observables.

To accurately describe the data for W+jets events, where the jet multiplicity is high, multi-leg Monte Carlo (MC) generators that combine multiple exclusive jet emission with highest possible perturbative accuracy in the strong and electroweak couplings are necessary. The legacy modeling of the W+jets production in such a challenging phase-space exhibited large discrepancies with the 8 TeV data. In preparation for the 13 TeV analysis, a collaborative effort was made with theorists and MC generator experts to fix and improve the physics and modeling of the W+jets production in the collinear regime. The measurements reported in this thesis are compared against the

resulting state-of-the-art MC predictions for the simulation of W+jets events. These prediction include SHERPA 2.2.11 MC generator with 0-2 jets@NLO + 3,4,5 jets@LO QCD accuracy and approximate NLO EW virtual corrections for up to 2 jets, and a new MADGRAPH5_AMC@NLO+PYTHIA8 generator with 0,1,2,3 jets@NLO QCD acurracy using FxFx merging [10] is used as an alternative prediction to compare against data. These new predictions are shown to significantly improve the description of data. The details of the generators will be discussed in a later chapter.

This thesis is structured as follows. Chapter 2 provides theoretical background for W+jets production. Chapter 3 offers an overview of the LHC experiment and the ATLAS detector. Chapter 4 details the dataset and triggers employed in the analysis. Chapter 5 outlines the Monte Carlo generators and simulated samples for signal prediction and background estimation. Chapter 6 covers object definition and reconstruction. Chapter 7 describes the measurement region defined at the particle and reconstruction levels, event selection procedure and observable of interested in this measurement. The estimation of backgrounds is expounded upon in Chapter 8, while Chapter 9 compares the signal modeling in the measurement regions at the reconstruction level. The sources of systematic uncertainties are discussed in Chapter 11. Chapter 10 covers the unfolding of reconstructed events to the particle level. Finally, Chapter 12 presents the measured cross-section and final results.

Chapter 2

Overview of W+jets Production in the Standard Model

2.1 Overview of the Standard Model

The Standard Model (SM) of particle physics is a mathematical framework that divides the set of fundamental particles into two classes: fermions and bosons. Fermions, characterized by half-odd-integer spin, constitute ordinary matter at the atomic level. In contrast, bosons with integer spin act as force mediators, and are known to give rise to three of the four known fundamental interactions: electromagnetic, weak, and strong interactions. These interactions are described mathematically through the local gauge symmetries SU(3)⊗SU(2)⊗U(1). The SU(3) symmetry gives rise to eight gluons that are responsible for the strong interaction. At energies surpassing the unification scale, the electromagnetic and weak interactions merge into the electroweak interaction, governed by the SU(2)⊗U(1) symmetry. The electroweak symmetry undergoes spontaneous breaking with the introduction of a scalar field through the Higgs mechanism. The resulting hidden SU(2)⊗U(1) symmetry gives rise to the massless photon and the massive $W\pm$ and Z bosons. Both the photon and Z boson are electrically neutral, while the $W\pm$ bosons carry electric charge due to a non-zero combination of weak hyper-charge and weak isospin. Lastly, the rest mass of all massive particles in the Standard Model is acquired through the Higgs mechanism [11], and the SM also incorporates antiparticles for each fermion, with antiparticles having exactly the same mass but opposite charge to their counterparts.

2.2 The Fermions and Bosons of the Standard Model

Ordinary matter in nature consists of fermions, including leptons and quarks, organized into three generations, with particles arranged in increasing order of mass within each generation.

The lepton sector consists of six elementary particles, further categorized into three lepton flavors. The electron, the most commonly known fundamental particle with unit electrical charge, is the lightest charged lepton, corresponding to the first generation. More massive and unstable charged leptons include the muon and τ (tau), corresponding to the second and third generations, respectively. The remaining three elementary leptons are the charge-neutral neutrinos, each associated with the same flavor as their charged lepton partner. Unlike their charged partners, neutrinos are nearly massless and interact very weakly with ordinary matter.

The quark sector also consists of six particles, each distinguished by a unique flavor. Within each generation of the quark sector, quarks are paired as up and down types, distinguished by their relative masses, with the up-type generally being more massive than the down-type. Up-type quarks carry an electrical charge of +2/3, while down-type quarks carry a negative charge of -1/3. Quarks also exhibit an additional property known as color charge symmetry, where each quark can be assigned the colors red, green, or blue. The combination of a quark and antiquark of the same color, or of three quarks of each of the three colors is colorless, and free particles must be colorless due to a property of the strong interaction called color confinement.

The interactions among fermions are described by the exchange of vector boson particles with a spin of 1. The electromagnetic interaction between charged particles occurs through the exchange of massless photons, whereas the Z and W are massive bosons governing the weak interactions among all fermions. The Z boson is electrically neutral and slightly more massive than the W boson, whereas the W boson carries a unit charge of either +1 or -1. Quarks are bound together by the strong interaction via the exchange of gluons, forming protons and neutrons, the constituents of atoms.

On the other hand, the Higgs boson is a massive scalar boson with zero spin, produced by the excitation of the scalar Higgs field. This scalar field imparts rest mass to all massive elementary particles in the SM via the Higgs mechanism [11].

2.3 W Boson Production from Proton-Proton Collision

The interaction leading to the production of a W boson in proton-proton collisions, constrained by the requirement that the W boson can only carry either +1 or -1 electrical charge, requires an interaction between a quark and an anti-quark pair to conserve charge. The proton consists of three valence quarks, two up (u) and one down (d) quark, while the anti-quark must originate from quark/antiquark sea generated by gluon splitting within the proton. Furthermore, the massive W boson subsequently decays into either lepton or quark pairs. The primary process for the leading-order production of the W boson is depicted in the Feynman diagram Figure 2.1(left). The analysis documented in this thesis makes use of the leptonic final-state, which is produced by the leptonic decay of

the W boson as depicted in Figure 2.1(right).

The interaction between the W boson and fermions is governed by the electroweak charged current component of the SM Lagrangian ($\mathcal{L}_{CC}$), as specified by Equation (2.1). Here, the g is the universal SU(2) coupling constant, and the $W_\mu^\dagger$ represents the W boson field. d_i ($\bar{u}_i$) is the down(up)-type (anti-)quark field, and ℓ and $\bar{\nu}$ correspond to the charged lepton and neutrino fields, respectively. The term regulating the interaction between the W boson and quarks is distinguished by an additional matrix element factor V_{ij}, describing quark flavor mixing and provided by the 3x3 unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix. The $h.c.$ term is the hermitian conjugate of the displayed Lagrangian.

$$\mathcal{L}_{CC} = \frac{g}{2\sqrt{(2)}} \left\{ W_\mu^\dagger \left[\bar{u}_i \gamma^\mu \left(1 - \gamma_5\right) V_{ij} d_j + \bar{\nu} \gamma^\mu \left(1 - \gamma_5\right) \ell \right] + h.c. \right\} \tag{2.1}$$

2.4 W+jets Production in the Collinear Phase-space

As mentioned above, particles in the final state of an interaction must be colorless due to color confinement. Consequently, free quarks or gluons cannot exist in isolation; they must undergo a hadronization process to form hadrons. The collection of particles created by the hadronization of a single quark or gluon is called a jet. These jets are recognized as energy clusters observed in particle detectors.

A leading-order model for W+1-jet production can be obtained by a simple modification of the diagram in Figure 2.1 with an additional next-to-leading order (NLO)

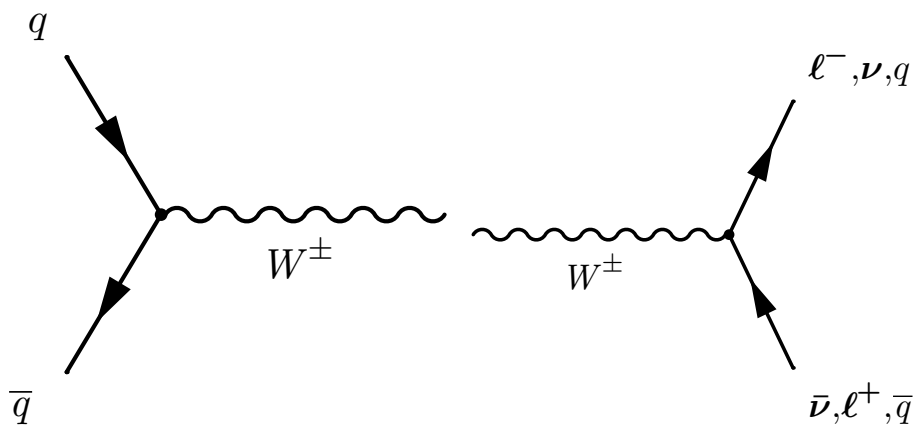
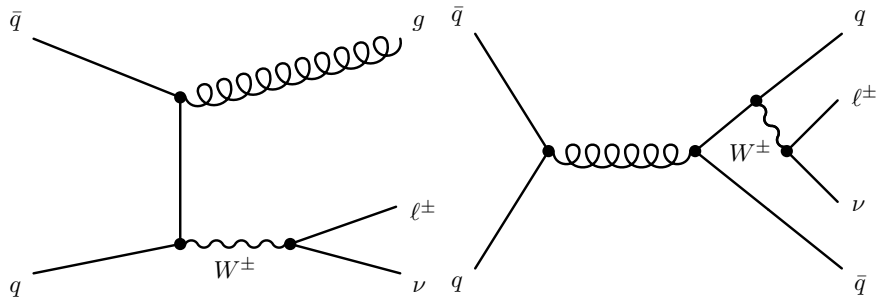


Figure 2.1: Tree level Feynman diagrams of W boson production. (a) shows on-shell W boson production from quark and anti-quark annihilation in proton-proton collisions. The W boson then decays leptonically or hadronically as shown in figure (b). The leptonic decay channel will be used for the measurements described in this thesis.

QCD vertex, representing initial gluon emission from the initial quark line, as shown in Figure 2.2(a). This configuration results in the W boson recoiling back-to-back against a single gluon jet, with a large angular separation between the W boson and the gluon jet. In contrast, this thesis focuses on collinear production where the angular separation between the W boson and the closest jets is small. This phase space is dominated by a higher-order W+ >1-jet production.

At higher order, the W+jets production process is more complicated and rich. The W boson production at higher order is not mostly from quark and anti-quark annihilation, but instead the W boson is emitted from initial or final state radiation from quark lines. In order to produce W+ >1 jets, for example, W+2 jets production can be treated as an electroweak (EW) correction to dijets production, as shown in Figure 2.2(b), where the W boson is emitted from either one of the initial or final state quark lines. The emitted W boson then tends to be in collinear with the jet from which it originated. Thus, to achieve collinear W boson production with multiple jets ($\geq$ 2 jets) in the final state, higher-order QCD and EW corrections, or the use of parton shower techniques, will be necessary. However, there is an overlap between diagrams involving higher-order corrections and leading order, requiring careful handling to avoid double counting. Therefore, the measurements of collinear W+jets production described in this thesis act to probe the current state-of-the-art theoretical calculations of higher order QCD and EW corrections in multi-jet final states, as well as test the precision of parton shower techniques.



(a) (b)

Figure 2.2: Representative feynman diagrams for the production of a W boson in association with jets. The t-channel diagram shown in (a) typically leads to large angular separation between the recoiling jet and the lepton, while the s-channel diagram shown in (b) can lead to small angular separations between the lepton and the outgoing quark line.

Chapter 3

The LHC Experiment and the ATLAS Detector

This chapter introduces the Large Hadron Collider (LHC) and the ATLAS detector. The LHC accelerates and collides protons at high center-of-mass energies, enabling the study of SM particles and their behaviors using data collected at $\sqrt{s}$=13 TeV. The ATLAS detector, a general-purpose detector, consists of multiple sub-detectors designed to accurately reconstruct particles produced in collision events. It provides precise measurements of their momentum, energy, and other kinematic properties, which are essential for probing and precisely measuring parameters within the SM, as well as for searching for phenomena beyond the SM.

3.1 The Large Hadron Collider

The LHC is currently the most powerful hadron collider, located on the France-Swiss border near Geneva, Switzerland. The LHC accelerator is built 100 m underground at the European Organization for Nuclear Research (CERN) and has a circumference of approximately 27 km. Protons are initially accelerated by a series of smaller accelerators before being injected into the LHC, where they are further accelerated to an energy of 6.5 TeV per proton beam, achieving a center-of-mass energy of up to 13 TeV at the collision points. Protons in the LHC travel along the accelerator ring in bunches of 1.1×10^{11} protons per bunch, with approximately 2233 bunches per beam. Two proton beams circulate in opposite directions with a bunch spacing of 25 ns. Interactions produced during bunch crossings are measured and recorded at the four major detection sites, as shown in Figure 3.1: CMS, LHCb, ALICE, and ATLAS.

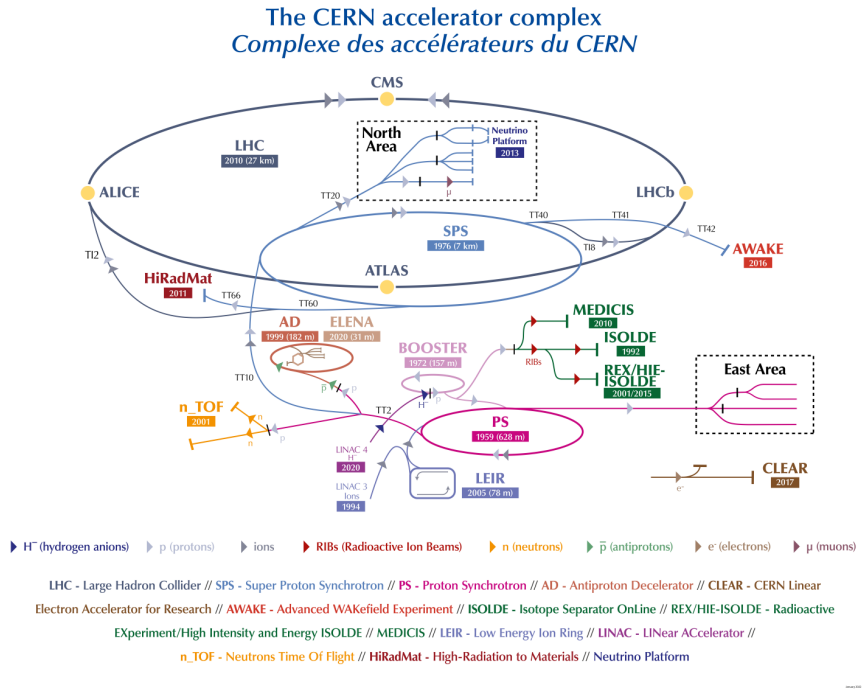


Figure 3.1: A schematic layout of the Large Hadron Collider. The four major detection sites, labeled with yellow circles, are CMS, LHCb, ALICE, and ATLAS.

The instantaneous per bunch luminosity (L) is a crucial parameter in collider physics that quantifies the number of collisions per unit area per unit time, as shown in Equation (3.1),

$$L = \frac{f(N_{p_1}N_{p_2})^2}{4\pi\sigma_x\sigma_y}, \tag{3.1}$$

where f is the bunch crossing frequency, N_{p_1} and N_{p_2} are the number of protons within a bunch, and σ_x and σ_y are the x and y components of the colliding beam widths. The expected instantaneous luminosity per bunch crossing is estimated, assuming that the number of protons is the same for the two colliding bunches and a uniform cross section of the bunch crossing, as shown in Equation (3.2)

$$L = \frac{(1.1\times10^{11})^2}{4\pi(25\times10^{-9}s)(16\times10^{-4}cm)^2} = 10^{34}cm^{-2}s^{-1}. \tag{3.2}$$

It is also convenient to define the integrated luminosity $L_{int} = \int Ldt$, which corresponds to the total data collected by the detector over the LHC operation period. The total integrated luminosity during the full Run-2 period is shown in Figure 3.2. At the end of Run-2, 139 fb^{-1} of data were collected.

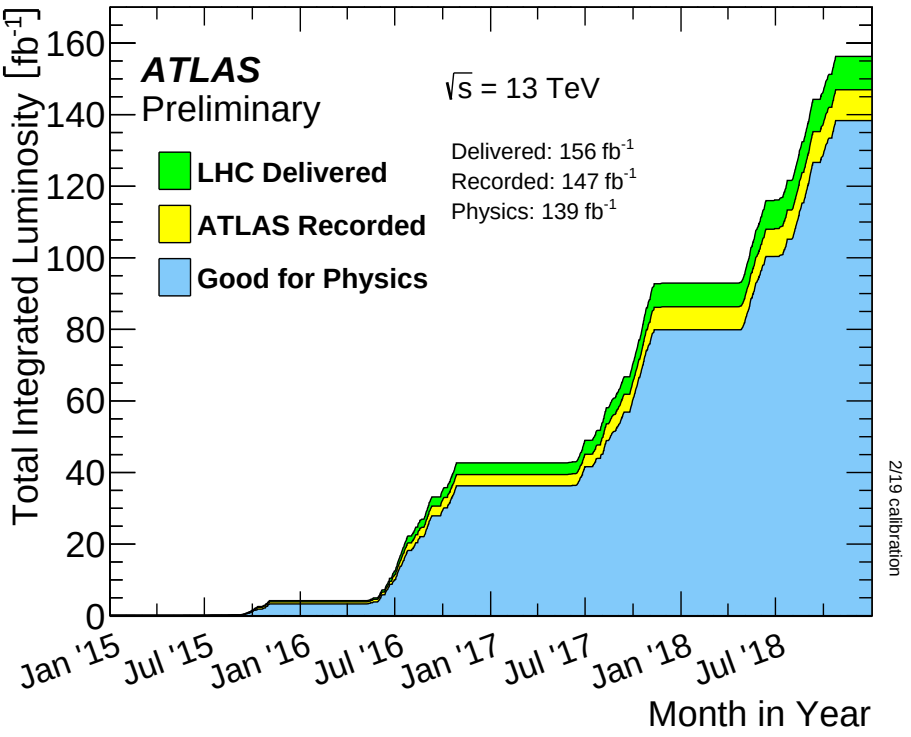


Figure 3.2: Total integrated luminosity delivered over the months during the full Run-2 period. This result is published on the ATLAS Run-2 luminosity public results page [12].

Another useful parameter for assessing the LHC operation conditions is the number of interactions per bunch crossing, μ, as shown in Equation (3.3)

$$\mu = L \times \sigma_{inel} \times f, \tag{3.3}$$

where L is the instantaneous per bunch luminosity, σ_{inel} is the inelastic proton-proton cross section, and f is the beam circulation frequency [12]. The average number of interactions $< \mu >$ over the course of the Run-2 data-taking period is shown in Figure 3.3,

with an average of about 30 interactions per bunch crossing throughout the entire Run-2 period.

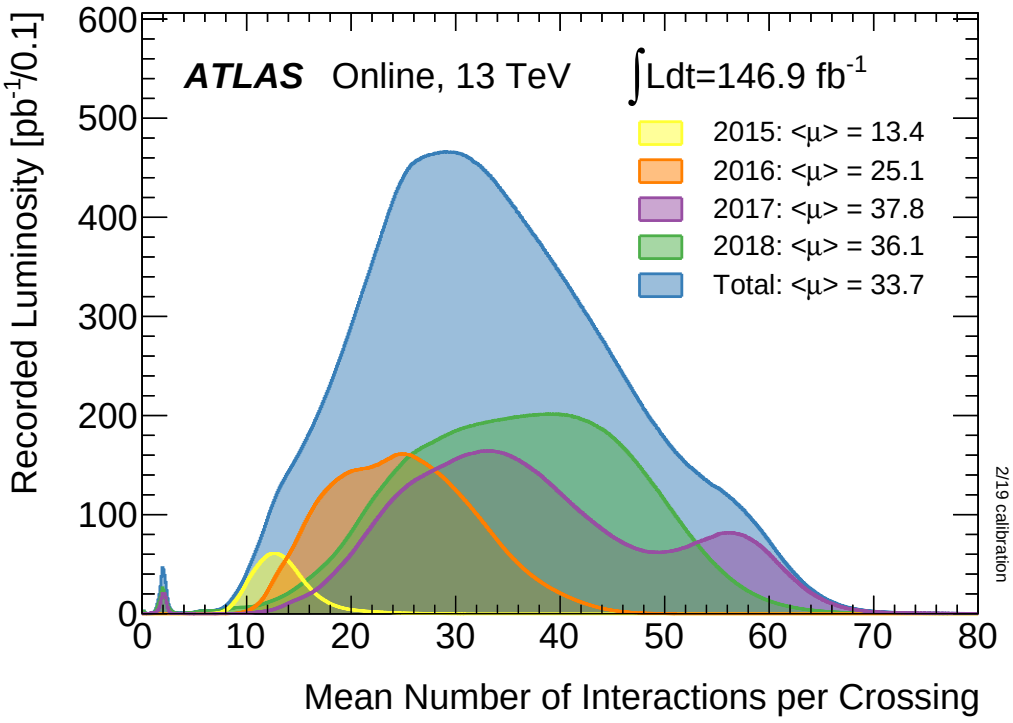


Figure 3.3: The average number of interactions per bunch crossing for different years in the Run-2 data-taking period. This result is published on the ATLAS Run-2 luminosity public results page [12].

3.2 ATLAS detector

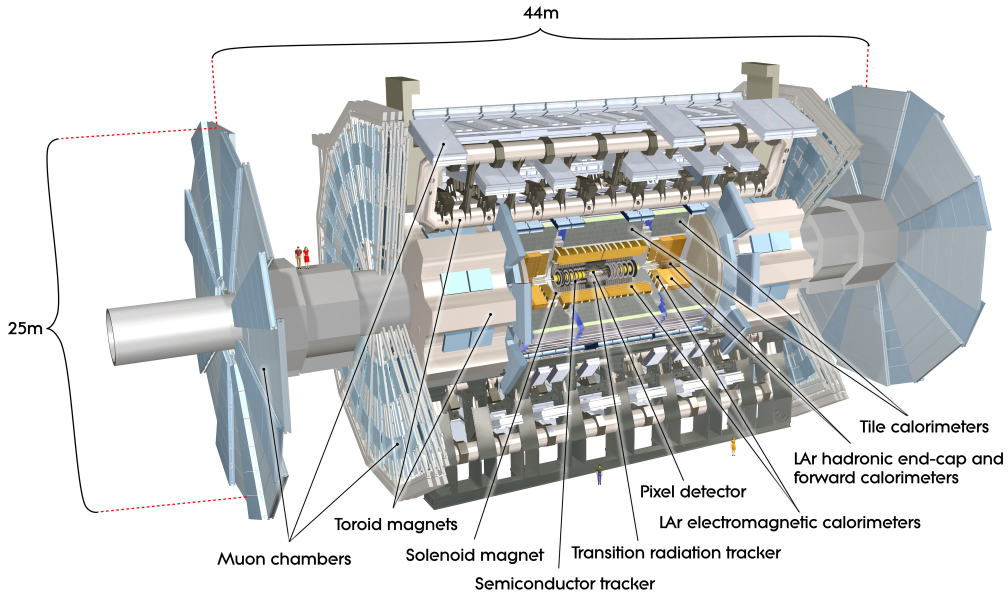


Figure 3.4: A schematic layout of the ATLAS detector and its major sub-detector components.

The ATLAS (A Toroidal LHC Apparatus) [13] is a cylindrical detector measuring 46 m in length and 25 m in diameter, as shown in Figure 3.4. It is a general-purpose detector located at one of the four detection sites at the LHC, designed for a wide range of physics studies, including precision measurements of SM processes and searches for elementary particles not described by the SM.

The ATLAS detector consists of multiple sub-detector systems and components, each designed to efficiently reconstruct specific types of physical objects from collision events. The major sub-detector systems, from innermost to outermost, are the Inner Detector

(ID), the electromagnetic (EM) and hadronic calorimeter, and the Muon Spectrometer (MS). Other essential components include the superconducting magnets and the trigger and data acquisition (DAQ) system. Each of sub-detectors systems will be introduced in the following sections.

3.2.1 Coordinate System

Given the geometry of the ATLAS detector, it is convenient to use a right-handed coordinate system centered at the nominal interaction point. The z-axis runs along the direction of the beam pipe, while the x-y plane is perpendicular to the beam direction. In this case, cylindrical coordinates (r, ϕ, z) are used, where (r, ϕ) parametrize the transverse plane and ϕ is the azimuthal angle around the z-axis. The momentum of objects is usually described using $\bar{p} = (p_{\mathrm{T}}, p_z)$, with p_{T} being the momentum of a particle in the transverse plane and p_z being the momentum along the beam (z-axis).

Rapidity (y) and pseudorapidity (η) are also often used to define the kinematics of outgoing particles, and they are defined as follows:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{3.4}$$

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right), \tag{3.5}$$

where E is the energy of the particle, p_z is the particle momentum in the z direction, and θ is the polar angle measured from the beam axis. When $\eta = 0$, it is perpendicular to

the beam axis. Rapidity (y) is useful for describing outgoing particles from a collision, as the difference in rapidity is Lorentz invariant under boosts in the beam direction. However, for highly relativistic particles, rapidity is difficult to measure precisely, whereas pseudorapidity (η) is much easier to measure experimentally. Pseudorapidity corresponds to rapidity in the case of massless particles or high-energy approximations. Moreover, η is often used to describe the coverage for each of the ATLAS detector components, which covers an $|\eta|$ range from 0 to 4.9.

3.2.2 Inner Detector

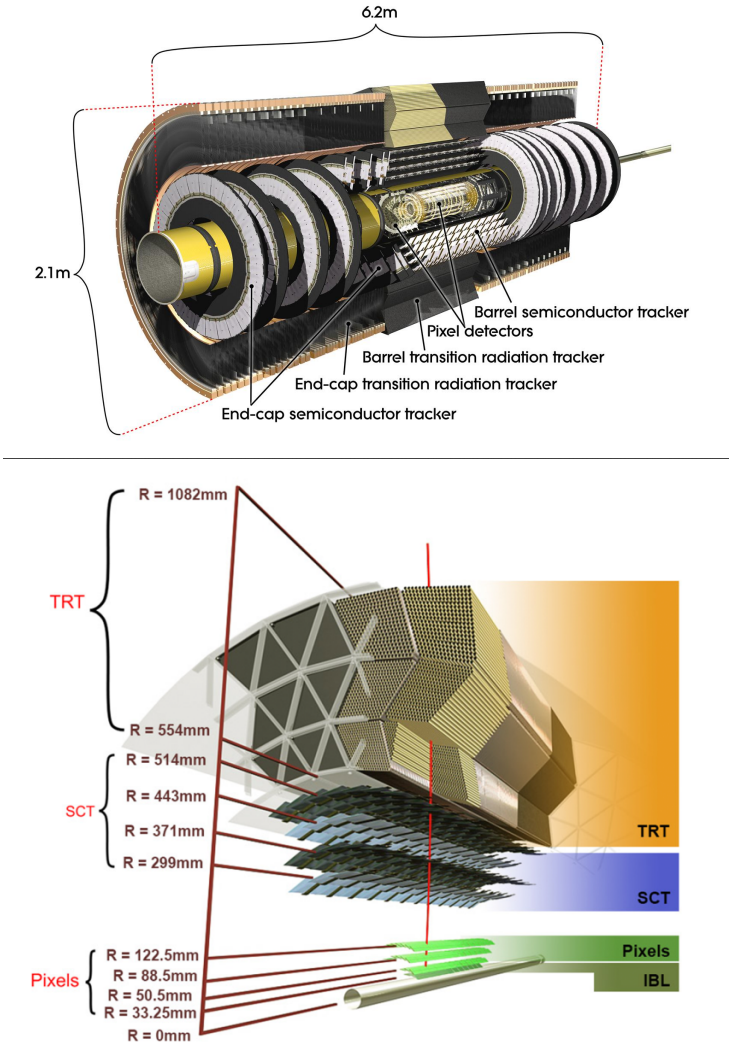


Figure 3.5: A schematic layout of the ATLAS Inner Detector (ID), showing the Pixel Detector (PD), Semiconductor Tracker (SCT) in both the end-cap and barrel regions, and the Transition Radiation Tracker (TRT) in both the end-cap and barrel regions. A radial cut view illustrates the radial coverage for the PD from 0 to 122.5 mm, SCT from 299 to 514 mm, and TRT from 554 to 1082 mm. This figure is sourced from the CERN website.

The Inner Detector (ID) system, as shown in Figure 3.5, is designed for tracking, determining the location of interactions, identifying charged particles, and measuring their trajectories and momenta. The ID system is encased in a cylindrical envelope around the collision points and is the detector system closest to the interaction point, making it the first to encounter outgoing particles. The ID system consists of three sub-detectors: the Pixel Detector (PD), the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). The entire ID system is immersed in a 2T magnetic field oriented in the z-direction, provided by the superconducting solenoid magnet.

The PD system is designed to have three-dimensional vertexing capabilities with excellent impact parameter resolution. It is the closest to the beam pipe, providing a total area coverage of 1.7 m^2 and pseudorapidity coverage of $|\eta| < 2.5$ with over 80 million silicon pixel sensors. These sensors are grouped into base units called modules, each containing a sensor tile with 47,232 pixels equipped with front-end readout electronics for signal amplification and discrimination. Each pixel sensor has dimensions of 50 µm x 400 µm and a thickness of 250 µm. The pixel sensors have a position resolution of 14 µm in ϕ and 115 µm in the z-direction. The PD system provides one measurement per barrel layer for each charged particle track and full pattern recognition capability to reconstruct tracks and interaction vertices. The PD system is cooled to approximately -5 to -10 $^{\circ}$C to maintain a good signal-to-noise ratio and minimize the effects of radiation damage.

The SCT system surrounds the PD system and consists of four layers with 6 million

silicon microstrips, with a total surface area of 60 m^2 and covers $|\eta| < 2.5$. In the barrel, the length of the strips is oriented along the z-axis and segmented in ϕ, while in the end-cap, they extend radially from the beam axis. The sensors are arranged into modules, each with two layers of strips rotated by 40 mrad with respect to each other to allow charged particle tracking along the beam direction. A track is therefore expected to have 8 hits in the SCT. The SCT shares the cooling system with the PD and is cooled between -5 to -10° C.

The last layer is the TRT system, which uses a drift tube system with over 350,000 drift tubes covering 12 m^3 of volume for $|\eta| < 2.0$. It provides spatial resolution in ϕ up to 130 µm per straw. In the barrel (endcap), there are 52,544 (245,760) straws over 73 layers (160 straw planes) that provide transition radiation tracking for charged particle identification. All charged particle tracks will traverse through at least 36 straws. The charge collection time in the straw is an important parameter for tracking performance. At a fixed transverse momentum for a charged particle, a light-mass charged particle will emit more transition radiation photons than a heavier-mass charged particle. Therefore, the TRT provides robust tracking information along with standalone track separation for electron and pion, or other charged hadron.

3.2.3 Calorimeters

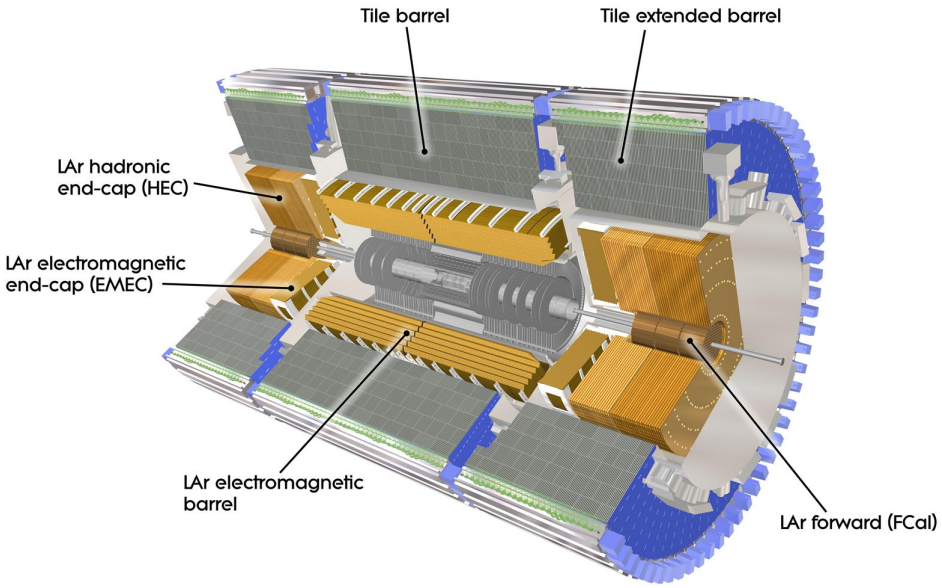


Figure 3.6: A longitudinal, cut-away view of the ATLAS calorimeter system, showing the tile calorimeter barrel, the tile calorimeter extended barrel, the liquid argon electromagnetic barrel, the liquid argon electromagnetic end-cap, the liquid argon hadronic end-cap, and the forward calorimeter. This figure is sourced from the CERN website.

The calorimeter system, as shown in Figure 3.6, consists of electromagnetic (EM) and hadronic calorimeters designed to measure the energy and position of particles produced in collisions. This system employs a sampling design with alternating layers of absorber and active materials. The absorbers cause the incident particles to interact and lose energy, which is subsequently detected by the active materials. The amount of charge generated by energy deposition and collected by the active material is proportional

to the energy of the incident particle.

The calorimeters are located outside the Inner Detector (ID) system and the magnetic solenoid field, providing extensive pseudorapidity coverage of $|\eta| < 4.9$. They offer sufficient granularity for precision measurements and ensure containment for both electromagnetic and hadronic showers from electrons, photons, and hadrons.

Different technologies and designs are used in different $|\eta|$ regions of the detector to meet the varying requirements of various physics processes. In the $|\eta|$ region covered by the ID, the calorimeters are finely segmented and well-suited for measurements of electrons and photons. Beyond this region, the calorimeters have a coarser granularity, designed to reconstruct jets, and contribute to the reconstruction of missing transverse momentum.

3.2.3.1 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (EMC) is designed to absorb and measure the energy deposited by showers generated from electromagnetic interactions of particles. The EMC uses a combination of lead plates as absorbers and liquid argon as the active material. The EMC consists of a barrel calorimeter within $|\eta| < 1.475$, divided into two halves with a 4 mm gap at $z = 0$, and two end-cap calorimeters covering $1.375 < |\eta| < 3.2$, with an inner wheel for $2.5 < |\eta| < 3.2$ and an outer wheel for $1.375 < |\eta| < 2.5$.

To provide complete and uniform coverage over ϕ without azimuthal cracks and allow

low latency readout of the data, the EMC uses accordion-shaped electrodes and absorber plates. The barrel and outer end-cap wheels are segmented into three radial sampling layers with varying thicknesses and radiation lengths. The first sampling layer is finely segmented in $\Delta\eta = 0.0031$ with 8 strips in front of each calorimeter cell. This fine segmentation in the first layer is essential for the capability of separating electromagnetic showers from electrons and photons from those of energetic pions. The second sampling layer, which collects the largest fraction of the energy of the electromagnetic shower, has a fine segmentation of $\Delta\eta = 0.025$ and $\Delta\phi = 0.0245$. The last layer collects the tail end of the electromagnetic shower with a coarser segmentation of $\Delta\eta = 0.05$. A schematic of a section of the EMC in the barrel region is shown in Figure 3.7.

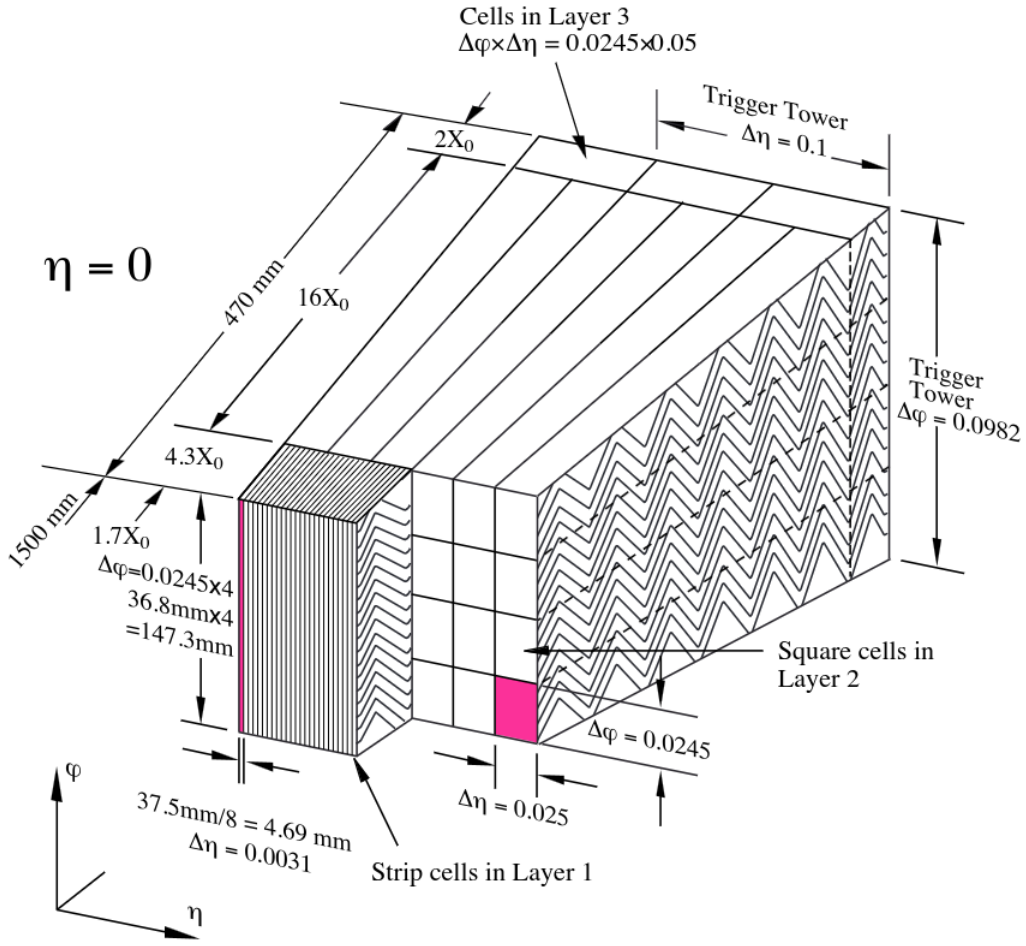


Figure 3.7: Schematic of the ATLAS Electromagnetic Calorimeter. This figure is sourced from the CERN website.

3.2.3.2 Hadronic Calorimeter

The Hadronic Calorimeters (HC) are utilized for measuring and containing the energy of particles interacting hadronically. The HC comprises a tile calorimeter, two hadronic end-cap calorimeters, and a liquid argon forward calorimeter. The tile calorimeter, functioning as a sampling calorimeter, employs steel as the absorber and scintillator

tiles as the active medium, covering the region $|\eta| < 1.7$. Positioned behind the liquid argon EM calorimeter, it is divided into a long central barrel, measuring 5.8 m in length and covering $|\eta| < 1.0$, and two extended barrels on each side of the detector, each 2.6 m in length and covering $0.8 < |\eta| < 1.7$, with radii ranging from 2.28 to 4.25 m. The Hadronic End-cap Calorimeters (HEC), located behind the end-cap EMC wheels, utilize copper as the absorber and liquid argon as the active material, covering the range $1.5 < |\eta| < 3.2$. Lastly, the liquid argon forward calorimeter (FCal), positioned approximately ±4.7 m from the interaction point, covers the range $3.1 < |\eta| < 4.9$ with much coarser granularity. It consists of three modules: an electromagnetic module with Copper and Liquid-Argon, and two hadronic modules with Tungsten and Liquid-Argon.

3.2.4 Muon Spectrometer

The Muon Spectrometer (MS) is the outermost detection system of the ATLAS detector, comprising four types of detectors: Monitored Drift Tubes (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC), and Thin Gap Chambers (TGC). These detectors specialize in detecting muons and facilitating the reconstruction of their momentum. Due to the higher mass of muons compared to electrons, muons tend to penetrate the Inner Detector (ID) without initiating showers in the calorimeter, making the MS critical for their detection.

The MS relies on the magnetic deflection of muon tracks within large superconducting toroidal magnets. For $|\eta| < 1.4$, the large barrel toroid provides the magnetic bending,

while for $1.6 < |\eta| < 2.7$, smaller end-cap magnets inserted into both ends of the barrel toroid contribute to the bending. In the transition region $1.4 < |\eta| < 1.6$, both the end-cap and barrel fields contribute to the bending.

Muon tracks within the barrel, transition, and end-cap regions are captured by chambers arranged in three cylindrical layers parallel to the beam axis. The MDT and CSC systems offer precision tracking measurements for the MS. MDTs comprise two sets of multi-layers, with each set comprising three to four layers of drift tubes filled with a gas mixture of Argon and CO2, providing tracking resolutions of 80 µm per tube and 50 µm per multi-layer, covering $|\eta| < 2.7$. In the innermost end-cap regions ($2.0 < |\eta| < 2.7$), CSCs, consisting of two disks with eight chambers per disk, provide tracking resolutions of 40 µm in the plane containing the beam axis and 5 mm in the transverse plane.

RPCs and TGCs contribute to muon triggering by offering fast response times for bunch-crossing identification and well-defined p_T trigger thresholds. RPCs select muon tracks within $|\eta| < 1.07$, forming three concentric layers (RPC1, RPC2, and RPC3) around the beam axis. In contrast, TGCs, mounted in the end-cap region ($1.07 < |\eta| < 2.7$), trigger on muon tracks and provide azimuthal measurements complementary to those from the MDTs. TGCs utilize a quenching gas mixture of CO2 and Pentane, offering a timing resolution of 4 ns and chamber tracking resolutions of 2 to 6 mm in the radial direction and 3 to 7 mm in the azimuthal direction.

3.2.5 Trigger System

The ATLAS Trigger System employs a multi-level approach to select events of interest for offline analysis. Given the extremely high instantaneous luminosity and the cross section for inelastic proton-proton collisions at $\sqrt{s}$ = 13 TeV, managing resource availability to store the vast amount of data poses a significant challenge.

To reduce the data rate and select a small fraction of events with the most interesting physics signatures, a set of highly selective triggers is used. This ensures that relevant events are stored at a frequency of approximately 1 kHz, within the bandwidth and storage capacities of ATLAS. Physics processes with final state signatures exceeding these rate limits are pre-scaled.

The ATLAS trigger system comprises three levels: Level-1 (L1), Level-2 (L2), and the Event Filter (EF). At L1, a coarse Region of Interest (RoI) is defined using a subset of detectors that are fast enough to identify the bunch crossing. The bunch crossing ID uniquely identifies two colliding bunches from all other bunches circulating in the LHC. Additionally, L1 is hardware-based and includes the calorimeter (L1Calo) and the muon spectrometer (L1Muon) triggers. L1Calo identifies energy clusters of interest using specific thresholds on energy deposits and η and ϕ values. These triggered regions are known as Region of Interests (RoIs), and they are defined in terms η and ϕ. Similarly, L1Muon uses information from the RPCs and TGCs in the MS system to select events of interest. At L2, additional detectors refine the measurement of the trigger candidate using the RoIs formed at L1. The EF level constructs variables using full event information to

further select events of interest.

After the hardware-based trigger, ATLAS uses a software-based high-level trigger (HLT). The HLT runs on a computing farm of around 40,000 processor cores, achieving an event recording rate of approximately 1 kHz. It employs both online and offline algorithms to analyze the momentum and angular distribution of final state particles and reconstruct the event. Various trigger "menus" determine whether an event is interesting enough to be kept for offline analysis. The specific trigger menus used for this measurement are discussed in Section 4.3.

3.2.6 Magnet System

The ATLAS magnet system generates strong magnetic fields that are crucial for detecting charged particles, reconstructing their trajectories, and measuring their momenta. This magnetic field is an essential component of both the ID and MS systems. The magnet system comprises one solenoid, one barrel toroid, and two end-cap toroids (one in each end-cap). The entire system spans a length of 26 m and a diameter of 22 m, storing approximately 1.6 GJ of energy. The volume where the total magnetic field exceeds 50 mT is around 12.000 m^3.

The solenoid, located in the barrel region between the ID and EMC systems, provides an axial field of 2 T for track measurements in the ID system. The toroids, installed within the MS, facilitate momentum and charge identification measurements of muons, generating a magnetic field of approximately 0.5 T in the barrel and around 1.0 T in the

end-cap regions.

Chapter 4

Data, Software, and Triggers

4.1 Software

An extensive software suite [14] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

This subsection describes the major software packages used in this analysis, as summarized in Table 4.1. This analysis uses the `STDM4` derivation samples with software packages released from `AnalysisBase, 21.2.226`, along with `SusySkimModelling`, a dedicated C++ analysis routine that relies on the `SusySkimAna` package. These tools are used to produce physics objects and measurement observables for further statistical analysis using downstream tools, which include `collinearw`, `RooUnfold`, and the scientific Python ecosystem. The `collinearw` is a Python-based analysis package that handles all of the histogramming, bookkeeping, and systematic evaluation. It integrates unfolding through `RooUnfold` via custom-written Python bindings. The combination of measurement results for the electron and muon channels is performed using `combiner`.

Software package	Usage
ATLAS Athena / Release 21.2.226	Tools from combined performance groups for object reconstruction, object containers, etc.
SusySkimAna / SusySkim-Modelling @ b13e420	Analysis routine to generate physical objects and obervables needed for this analysis.
collinearw @ 6d86869	Statistical analysis: histograming, systematic processing and bookkeeping, etc.
RooUnfold @ c9438d4	Unfolding: remove detector effects and retrieve measurement at particle level.
combiner @ 3832ca6	Tool for combining results from electron and muon channels.

Table 4.1: Summary of major software packages used in this analysis.

4.2 Dataset

The data used in this analysis were recorded with the ATLAS detector from 2015 to 2018 in pp collisions at $\sqrt{s}$ = 13 TeV (full Run-2 data set) and correspond to a total integrated luminosity of 140 fb^{-1} [15]. The uncertainty in the combined 2015–2018 integrated luminosity is 0.83% [16], obtained using the LUCID-2 detector [17] for the primary luminosity measurements, complemented by measurements using the inner

detector and calorimeters. The mean number of pp interactions per bunch crossing, including hard scattering and pile-up events, was $\langle\mu\rangle = 34$. The data are required to satisfy criteria that ensure that the detector was in good operating condition as described in Section 6.2.6.

4.3 Single Lepton Triggers

This analysis relies on the lowest p_{T} un-prescaled single lepton triggers available at the time of data collection. The trigger is emulated in MC simulation, and corresponding data/MC scale factors from the relevant combined performance groups are used to correct for differences between data and MC samples [18, 19]. Table 4.2 lists the triggers used to select the analysis channels across various data-taking periods optimized for different luminosity ranges.

Different sets of requirements and stages of the trigger systems are used for the muon and electron channels, respectively. Each set of requirements is connected with a **logical OR (||)** condition. The Level-1 (L1) trigger requirements are implemented first, followed by a list of High-Level Trigger (HLT) requirements. The nomenclature of lepton HLT triggers usually consists of an energy or momentum threshold, identification criteria, and isolation requirements. For example, the muon trigger menu for 2015, "HLT_mu20_iloose_L1MU15", indicates preselection at the L1 trigger with one muon with $p_{\mathrm{T}} > 15$ GeV, followed by the selection of at least one muon with $p_{\mathrm{T}} > 20$ GeV that satisfies the "iloose" identification criteria at the HLT trigger. A summary of the

trigger menu names and their corresponding requirements is shown in Table 4.3.

	Run Number		
Trigger Menu	Start	End	Year (Periods)
Muon channel			
HLT_mu20_iloose_L1MU15 || HLT_mu40	276262	284484	2015
HLT_mu26_ivarmedium || HLT_mu50	297730	-	2016, 2017, 2018
Electron channel			
e24_lhmedium_L1EM20VH || e60_lhmedium || e120_lhloose	276262	284484	2015
e26_lhtight_nod0_ivarloos || e60_lhmedium_nod0 || e140_lhloose_nod0	297730	-	2016, 2017, 2018

Table 4.2: Single lepton triggers used for the the electron and muon channels for various taking periods. The set of trigger requirements are connected by logical OR(||). The description of the physical quantities for each item in the trigger menu are summarized in Table 4.3.

Trigger Menu	Level-1 (L1)	Hight-Level Trigger (HTL)
HLT_mu20_iloose_L1MU15	At least one muon with $p_{\mathrm{T}} > 15$ GeV.	at least one muon with $p_{\mathrm{T}} > 20$ GeV, satisfy loose isolation.
HLT_mu40	-	At least one muon with $p_{\mathrm{T}} > 40$ GeV.
HLT_mu26_ivarmedium	-	At least one muon with $p_{\mathrm{T}} > 26$ GeV, satisfy medium, variable threshold isolation.
HLT_mu50	-	At least one muon with $p_{\mathrm{T}} > 50$ GeV.
e24_lhmedium_L1EM20VH	$E_T > 20$ GeV, E_T threshold dependent hadronic calorimeter veto.	$E_T > 24$ GeV, medium likelikhood ID.
e60_lhmedium	-	$E_T > 60$ GeV, medium likelikhood ID.
e120_lhloose	-	$E_T > 120$ GeV, loose likelikhood ID.
e26_lhtight_nod0_ivarloos	-	$E_T > 26$ GeV, tight likelikhood ID, no impact parameter(d0) requirement.
e60_lhmedium_nod0	-	$E_T > 60$ GeV, medium likelikhood ID, no impact parameter(d0) requirement.
e140_lhloose_nod0	-	$E_T > 140$ GeV, loose likelikhood ID, no impact parameter(d0) requirement.

Table 4.3: Summary of the requirements for trigger menu shonw in Table 4.2.

Chapter 5

Monte Carlo Simulated Samples

Simulated Monte Carlo (MC) samples are used to estimate various Standard Model background processes. The main characteristics of all physics objeccts reconstructed in the simulated events are corrected to the data using the scale factors provided by the combined performance groups. Corrections applied to the simulation include scale factors for differences in the b-tagging, lepton identification/isolation/trigger efficiencies and "jet vertex tagger"(JVT) pile-up suppression algorithm. Additionally, the average number $< \mu >$ of simulated interactions is reweighted to match the data taking conditions.

State-of-the-art merged MC signal predictions are compared against the measured cross-sections. This section describes the MC setups used in this analysis. A condensed summary of the MC samples is given in Table 5.1, and Sections 5.1 to 5.4 provide details of their configurations.

Process	Matrix element	Parton shower	PDF set	Tune
W + jets	Sherpa 2.2.11 [20–22]	Sherpa 2.2.11 [23, 24]	NNPDF30NNLO [25]	Sherpa default
W + jets	MadGraph v2.6.5 [26]	Pythia 8.240 [27]	NNPDF3.1luxQED [28]	A14+FxFx
Diboson	Powheg v2 [29–32]	Pythia 8.210 [27]	NLO CT10 [33]	AZNLO [34]
$t\bar{t}$	Sherpa 2.2.12 [20–22]	Sherpa 2.2.21 [23, 24]	NNPDF30NNLO [25]	Sherpa default
t (s-channel)	Powheg v2 [29–32]	Pythia 8.230 [27]	NLO CT10 [33]	Perugia2012 [35]
t (t-channel)	Powheg v2 [29–32]	Pythia 8.230 [27]	NLO CT10f4 [33]	Perugia2012 [35]
$t + W$	Powheg v2 [29–32]	Pythia 8.230 [27]	NLO CT10 [33]	Perugia2012 [35]
Dijets	Pythia 8.230 [27]	Pythia 8.230 [27]	NNPDF2.3lo [36]	A14
EWWjj	Sherpa 2.2.11 [20–22]	Sherpa 2.2.11 [23, 24]	NNPDF30NNLO [25]	Sherpa default

Table 5.1: Overview of simulated Monte Carlo samples used in this analysis.

5.1 Single V + jets

The signal process measured in this analysis is $pp \rightarrow \mathrm{W} \rightarrow \ell\nu$ plus jets, illustrated in Figures 2.2(a) and 2.2(b). Two state-of-the-art predictions are compared against the measured data. The nominal prediction uses Sherpa 2.2.11 [20–22], and an alternative signal model uses MadGraph5_aMC@NLO event generators [26]. All results shown in this analysis are at NLO QCD precision (unless specified otherwise). Additionally, a newly available NLO EW_{virt} correction is introduced to the Sherpa 2.2.11 prediction to study the effects and contributions of the EW virtual correction on the signal. All single W+jets predictions are normalized to the cross-section predicted by the MC event generator.

New modeling and statistical enhancement techniques were introduced in the Sherpa 2.2.11 package to improve modeling relative to the legacy Sherpa 2.2.1 setup [37]. In the Sherpa 2.2.11 samples, NLO-accurate matrix elements for up to 2 jets are computed, and LO-accurate matrix elements for up to 5 jets are calculated with the Comix [21] and OpenLoops [22, 38] libraries. The default Sherpa parton shower [23] based on Catani-Seymour dipoles and the cluster hadronization model [39] are used. The splitting function for initial state radiation was modified to include extra terms that improve the prediction of the parton shower and matrix element near the merging scale (a small effect for this analysis due to the high-energy jet selection).

The sample was statistically enhanced using the $\max(S_{\mathrm{T}}, p_{\mathrm{T}}^{V})$ variable to capture the different production cross-sections in different phase-space configurations, where S_{T}

refers to the scalar sum of the transverse momenta of all final-state partons excluding those from the boson decay, and p_{T}^{V} is the transverse momentum of the boson. The $\max(S_{\mathrm{T}}, p_{\mathrm{T}}^{V})$ variable extracts the maximum between the S_{T} and p_{T}^{V} in a given event, adjusting the sampling rate efficiently based on differences in cross-section of different parton multiplicities [10]. QCD scale uncertainties have been evaluated on-the-fly [40] using 7-point variations of the factorization and renormalization scales in the matrix elements. The scales are varied independently by factors of 0.5 and 2, avoiding opposite factors. Unlike Sherpa 2.2.1, the scale variations in Sherpa 2.2.11 include coherent variations in the matrix elements and parton shower, which are the default for the Sherpa 2.2.11 prediction.

An alternative NLO-accurate multi-jet merged setup is provided using MadGraph5_aMC@NLO +Pythia with the FxFx merging procedure. Samples are generated using the MadGraph5_aMC@NLO v2.6.5 program to generate matrix elements for V+ 0, 1, 2, and 3 additional partons in the final state to NLO accuracy. The showering and subsequent hadronization are performed using Pythia8.240 [41] with the `A14` tune [42], employing the `NNPDF2.3lo` PDF set [36] with α_{s} =0.130. The different jet multiplicities are merged using the FxFx prescription [43]. MadGraph5_aMC@NLO performs a 5-flavor scheme (5FS) calculation with massless b- and c-quarks at the matrix element level and massive quarks in the Pythia shower. The PDF set used for event generation is `NNPDF3.1luxQED` [28] with α_{s} = 0.118, and the samples have been generated with additional weights for the PDF replicas and scale variations of the renormalization and

factorization scales [44].

Electroweak production of $\ell\ell jj$, $\ell\nu jj$, and $\nu\nu jj$ final states was simulated with SHERPA 2.2.12 [45] using leading-order (LO) matrix elements with up to two additional parton emissions. The matrix elements were merged with the SHERPA parton shower [23] following the MEPS@LO prescription [46] and using the set of tuned parameters developed by the SHERPA authors. The NNPDF3.0NNLO set of PDFs [25] was employed. The samples were produced using the VBF approximation, which avoids overlap with semi-leptonic diboson topologies by requiring a t-channel color-singlet exchange. The electroweak production of $\ell\nu jj$ is treated as a background in this analysis and is subtracted from the data prior to unfolding.

5.2 Top quark

The simulation of top quark backgrounds is divided into pair-produced top quarks ($t\bar{t}$) and single top quark production. The production of $t\bar{t}$ events is modeled using SHERPA 2.2.12. In these samples, NLO-accurate matrix elements for up to 1 jet are computed, and LO-accurate matrix elements for up to 4 jets are calculated with the Comix [21] and OpenLoops [22, 38] libraries. The same general configuration as the signal W+jets described above is used for these samples.

The $t\bar{t}$ prediction uses the POWHEGBOX v2 [29–32] generator. This generator provides matrix elements at next-to-leading order (NLO) in the strong coupling constant α_s with the NNPDF3.0NLO [25] parton distribution function (PDF) and the h_{damp} parameter

set to 1.5 m_{top} [47], which controls the matching in POWHEG and effectively regulates the high-p_{T} radiation against which the $t\bar{t}$ system recoils. The functional form of the renormalization and factorization scale is set to the default scale $\sqrt{m_{\text{top}}^2 + p_{\text{T}}^2}$.

The $t\bar{t}$ sample is normalized to the cross-section prediction at next-to-next-to-leading order (NNLO) in QCD, including the resummation of next-to-next-to-leading logarithmic (NNLL) soft-gluon terms calculated using TOP++2.0 [48–54]. For proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, this cross-section corresponds to $\sigma(t\bar{t})_{\text{NNLO+NNLL}} = 832 \pm 51\,\text{fb}$, using a top quark mass of $m_{\text{top}} = 172.5\,\text{GeV}$. The uncertainties on the cross-section due to PDF and α_{s} are calculated using the PDF4LHC prescription [55] with the MSTW2008 68% confident-level (CL) NNLO [56, 57], CT10 NNLO [33, 58], and NNPDF2.3 5f FFN [36] PDF sets, are added in quadrature to the scale uncertainty.

Single-top tW associated production is modeled using the POWHEGBOX v2 [30–32, 59] generator, which provides matrix elements at next-to-leading order (NLO) in the strong coupling constant α_{s} in the five-flavor scheme with the NNPDF3.0NLO [25] PDF set. The functional form of the renormalization and factorization scale is set to the default scale, which is equal to the top quark mass ($m_{\text{top}} = 172.5\,\text{GeV}$). The diagram removal scheme [60] is employed to handle the interference with $t\bar{t}$ production [47]. The events are interfaced with PYTHIA8.230 [27] using the A14 tune [42] and the NNPDF23LO PDF set. The decays of bottom and charm hadrons are simulated using the EVTGEN v1.6.0 program [61]. The inclusive cross-section is corrected to the theory prediction

calculated at NLO in QCD with NNLL soft-gluon corrections [62, 63]. For proton-proton collisions at a center-of-mass energy of $\sqrt{s}$ = 13 TeV, this cross-section corresponds to $\sigma(tW)_{\text{NLO+NNLL}} = 71.7 \pm 3.8\,\text{pb}$, using a top quark mass of $m_{\text{top}} = 172.5\,\text{GeV}$. The uncertainty on the cross-section due to PDF is calculated using the MSTWNNLO 90% CL PDF set [56, 57] and is added in quadrature to the scale uncertainty.

5.3 Diboson

The Powheg Box v2 [30–32] generator was used to simulate the WW, WZ, and ZZ production processes at NLO accuracy in QCD [64]. The simulations included the effects of singly resonant amplitudes and interference effects due to Z/γ^* and same-flavour lepton combinations in the final state, where appropriate. However, interference effects between WW and ZZ for same-flavour charged leptons and neutrinos were ignored.

Diboson events were interfaced with Pythia 8.210 [27] for parton shower modeling, hadronisation, and underlying event simulation, using parameters set according to the AZNLO tune [34]. The CT10 PDF set [33] was used for the hard-scattering processes, while the CTEQ6L1 PDF set [65] was used for the parton shower. The EvtGen 1.2.0 program [61] was employed to decay bottom and charm hadrons.

The factorisation and renormalisation scales were set to the invariant mass of the boson pair. An invariant mass of $m_{\ell\ell} > 4\,\text{GeV}$ was required at the matrix-element level for any pair of same-flavour charged leptons.

5.4 Di-jet

Di-jet production was generated using Pythia 8.230 [27] with leading-order matrix elements matched to the parton shower. The renormalisation and factorisation scales were set to the geometric mean of the squared transverse masses of the two outgoing particles in the matrix element, $p_{\mathrm{T}}^{\mathrm{hat}} = \sqrt{(p_{\mathrm{T},1}^2 + m_1^2)(p_{\mathrm{T},2}^2 + m_2^2)}$. The NNPDF2.3lo PDF set [36] was used for the matrix element generation, parton shower, and simulation of multi-parton interactions. The A14 [42] set of tuned parameters was used.

Perturbative uncertainties were estimated using event weights [66] that account for variations in the scales at which the strong coupling constant is evaluated in the initial- and final-state showers, as well as the PDF uncertainty in the shower and the non-singular part of the splitting functions. The modeling of fragmentation and hadronisation was based on the Lund string model [67, 68].

To efficiently populate the inclusive jet p_{T} spectrum, the sample used a biased phase-space sampling, which was compensated for by a continuously decreasing event weight. Specifically, events at a scale $p_{\mathrm{T}}^{\mathrm{hat}}$ were oversampled by a factor of $(p_{\mathrm{T}}^{\mathrm{hat}}/10\,\mathrm{GeV})^4$.

Chapter 6

Object Definition

6.1 Definition of objects and observables

In this section we define the various objects and observables that are used to select and categorize our events. The objects we consider are electrons (Section 6.2.1), muons (Section 6.2.2) and jets (Section 6.2.3). In addition, the missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) is defined in Section 6.2.4,

For objects and observables available within the standard `SUSYTools` package, we rely on `AnalysisBase` software for their reconstruction as described in Section 4.1. All physics objects are labeled as either “baseline" or “signal". All criteria applied to baseline leptons are also applied to signal leptons. The baseline objects are used for overlap removal, reconstruction of missing transverse momentum, and estimation of the multi-jet backgrounds. These baseline leptons are also used for the lepton veto described in Section 7.4.

6.2 Objects definition and reconstruction

For a quick summary of the electron, muon and jet selections, see Table 6.1.

6.2.1 Electrons

Electrons are reconstructed by algorithms that provide a good separation between isolated electrons and fake signatures from hadronic jets. We require the electron shower to be consistent with that expected of an electron. Cuts are applied on the fractional

deposit in the hadronic calorimeter and on the shower width in η and ϕ in the first and second layers of the electromagnetic calorimeters, as well as track quality requirements and track-cluster match requirements. Making use of these criteria, the "*MediumLLH*"[1] quality requirement is used to define baseline electrons. In addition, the p_{T} of electrons is required to be larger than 10 GeV, and $|\eta| < 2.47$, removing the crack region of $1.37 < |\eta| < 1.52$, where η of the cluster in the 2nd calorimeter sampling layer is used. When calculating the p_{T} of an electron, the energy is always taken to be from the calorimeter cluster, suitably calibrated, and the value of η is taken from the matched track. The electron energy scale is corrected for data and smeared for Monte Carlo, as specified in Ref [69]. Electrons satisfying the above criteria are referred to as "baseline" electrons.

Additionally, for electrons we apply the "Tight_VarRad" isolation[2] requirement which exploits a selection on the calorimeter isolation within a cone of $R(\eta, \phi) < 0.2$ and a selection on the momentum contained in tracks within a cone of $R(\eta, \phi) < 0.2$ around the electron candidate. We also require for the associated track that $|z_0 \sin\theta|$ be less than 0.5 mm, and that the significance $|d_0/\sigma_{d_0}|$ be less then 5.0 for electrons. Here, the standard helical track parameters z_0 and d_0 are defined relative to the primary vertex, which in turn is derived from fits to groups of high-quality reconstructed tracks consistent with coming from a common origin. In order to pass the p_{T} requirements in the single electron trigger, electrons are required to have $p_{\mathrm{T}} > 30$ GeV. Signal electrons

[1] electron identification from E/gamma CP group

[2] https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/IsolationSelectionTool

are also required to have $|\eta| < 2.4$ to harmonize the selection with the muon channel (described below). Baseline electrons passing the additional criteria above are referred to as “signal" electrons.

6.2.2 Muons

Muon candidates are reconstructed by algorithms that perform a statistical combination of a track reconstructed in the muon spectrometer with a corresponding track in the inner detector up to $|\eta| < 2.4$. All muon identification cuts are taken from recommendations proposed by the muon combined performance group [70]. The p_{T} of the muons is required to be larger than 10 GeV. The pseudo-rapidity of muons must satisfy $|\eta| < 2.4$. Muons are required to be of at least *Medium* quality, as defined by the muon combined performance group[3]. Muons passing the above criteria are referred to as “baseline" muons.

Additionally, muons are then required to pass the “Loose_FixedRad”[4] isolation working point. We also require for the associated track that $|z_0 \sin\theta|$ be less than 0.5 mm, and that $|d_0/\sigma_{d_0}| < 3.0$. In order to pass the p_{T} requirements in the single muon trigger, muons are required to have p_{T}> 30 GeV. Baseline muons passing the additional criteria above are referred to as “signal" muons.

[3]Muon CP group: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonPublicResults
[4]https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/RecommendedIsolationWPs

6.2.3 Jets

This analysis uses $R = 0.4$ anti-k_t jets clustered from particle flow (or "PFlow") objects as summarized in this paper [71], which are constructed using an algorithm which removes calorimeter energy deposits due to charged hadrons and replaces them with tracking information. Particle-flow objects are charged-particle tracks matched to the hard-scatter vertex and calorimeter energy clusters following an energy subtraction algorithm [71] that removes the calorimeter deposits associated with good-quality tracks from any vertex. The tracking information is used to improve the clusters' energy resolutions. This also improves the angular resolution of jets compared to those clustered from topological clusters from the calorimeters only. The momenta of these jets are calibrated as described in this paper [72]. All jets are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|y| < 2.8$ and are referred to as "baseline jets". Any jets failing these requirements are discarded and not used further, except where stated explicitly.

Jets classified as "signal" must have $p_{\mathrm{T}} \geq 30\,\mathrm{GeV}$ and $|y| < 2.5$. Additionally, they must satisfy the "Tight" requirement, which is the default working point corresponding to an average signal efficiency of 96%. For "signal" jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|y| < 2.4$, they must also satisfy an additional requirement based on the output of the multivariate "jet vertex tagger" (JVT) algorithm. This algorithm is used to identify and discard jets primarily originating from pileup interactions, as detailed in [73, 74]. Jets failing these requirements are not considered in this analysis.

The analysis makes use of a b-jet veto, primarily to suppress background contributions

from top quark production. A b-tagging procedure is applied to label jets likely to have originated from a b-quark. This procedure requires the availability of tracking information, and for this reason each jet has assigned to any track within a p_{T}-dependent ΔR of the jet axis ("close to"), provided the track is of sufficient quality and has $p_{\mathrm{T}} > 500\,\mathrm{MeV}$. Where a track is close to more than one jet, it is assigned to the closest jet. The DL1r algorithm [75], a different version of deep neutral network algorithm based on DL1 for flavor-tagging, is then applied to all jets, and they are considered b-tagged (a b-jet) if they pass the fixed cut working point (`FixedCutBEff_60`), chosen to have a 60.64% efficiency on average for jets associated with true b-hadrons in simulated $t\bar{t}$ events. This is a multivariate algorithm which uses a selection of inputs including information about the impact parameters of ID tracks, the presence of displaced secondary vertices, and the reconstructed flight paths of b- and c-hadrons inside the jet are reported in [76–79]. At the chosen working point, the light-jet (charm-jet) rejection measured in $t\bar{t}$ MC simulation is about a factor of 1155 (29) on average for small-R jets. The training and calibration of this algorithm is performed separately for each jet type as described in [80, 81]. Correction factors are applied to the simulated samples to compensate for differences between the b-tagging efficiencies in data and simulation. The configuration used this analysis is available in `2020-21-13TeV-MC16-CDI-2021-04-16_v1.root`.

Property	Signal	Baseline
	Electrons	
Kinematic	$p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$(crack − veto)	$p_{\mathrm{T}} > 10\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.47$(crack − veto)
Identification	TightLLH	MediumLLH
Isolation	Tight_VarRad	–
Impact parameter	$\lvert d_0/\sigma(d_0)\rvert < 5$, $\lvert z_0 \sin\theta\rvert < 0.5$ mm	
	Muons	
Kinematic	$p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$	$p_{\mathrm{T}} > 10\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$
Identification	Medium	Medium
Isolation	Loose_FixedRad	–
Impact parameter	$\lvert d_0/\sigma(d_0)\rvert < 3$ & $\lvert z_0 \sin\theta\rvert < 0.5$ mm	
	Jets	
Kinematic	$p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert y\rvert < 2.5$	$p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $\lvert y\rvert < 2.8$
Clustering	Anti-k_t $R = 0.4$ PFlow	
Pileup mitigation	JVT Medium for $p_{\mathrm{T}} < 60$ GeV, $\lvert y\rvert < 2.4$	–
b-tagging	$\lvert y\rvert < 2.5$, DL1r FixedCutBeff 60%	–
	Missing Transverse Momentum	
Object selection	Baseline	
Working point	Tight	

Table 6.1: Summary of baseline and signal object definitions. The electron crack-veto removes the region $1.37 < |\eta| < 1.52$, where η of the cluster in the 2nd calorimeter sampling layer is used.

6.2.4 Missing transverse momentum

Due to conservation of transverse momentum, if all particles produced in the primary collision are detectable then there should be no $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event unless it arises from detector effects e.g. resolution, material effects, or non-instrumented regions of the detector.

Missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) is computed using an object-based algorithm as described in [82]. All *baseline objects* described above serve as input to the METMaker tool. Calorimeter energy deposits and reconstructed muons up to $|\eta| < 4.9$ are attributed to high-p_{T} objects in the following order: electrons, photons, jets, and muons. Tracks associated with the primary vertex but not linked to any such object are included in the Soft Term[5] contribution to $E_{\mathrm{T}}^{\mathrm{miss}}$, which has demonstrated robustness against pile-up [82, 84]. The $E_{\mathrm{T}}^{\mathrm{miss}}$ vector (bolded) is computed as the sum of the following terms:

$$(\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})_{x(y)} = (\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{e} + (\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{\gamma} + (\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{\mathrm{jet}} + (\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{\mu} + (\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}})_{x(y)}^{\text{Soft Term}}, \quad (6.1)$$

where each term is computed from the negative sum of calibrated reconstructed objects in the respective category. The *Tight* Working Point [85] to reduce the pileup dependence of missing transverse momentum. The Tight $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated without jets beyond $|\eta| > 2.4$ and $20 < p_{\mathrm{T}} < 30$ GeV.

[5] The Run 1 analysis used Cluster Soft Terms [83].

6.2.5 Overlap removal

To ensure accurate event reconstruction, it is important to address instances of energy double-counting within the detector. For instance, consider a scenario where a high-p_{T} muon radiates energy in the hadronic calorimeter before being detected by the muon spectrometer. In this case, a jet containing the muon may be reconstructed with minimal inner detector (ID) tracks. However, such a jet originates from final state radiation rather than being part of the parton shower hadronization process. Therefore, it is necessary to exclude this "jet" if it coincides spatially with the muon. Steps are taken within the ATLAS detector reconstruction framework to eliminate "fake" objects and mitigate energy overlap, ensuring the fidelity of event reconstruction.

Baseline reconstructed objects undergo the default "`SUSYTools`" overlap removal procedure to prevent such double-counting signal objects. The baseline object selections above are applied to the event sample. The following overlap rules are applied to the remaining events to correct for double counting between the different object selections. The overlap rules can change the number of electron, muon, and jet candidates. The signal objects are defined from the remaining non-overlapped objects with an isolation requirement imposed in Table 6.1.

Objects are considered to overlap if they lie less than a given distance ΔR Equation (6.2) from each other

$$\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2} \tag{6.2}$$

where ΔR defines the distance in rapidity y and azimuthal angle ϕ. Overlaps between candidate objects are removed sequentially as described in Table 6.2.

Overlap removal between muons and electrons that share an Inner Detector (ID) track aims to resolve two sources of ambiguity:

1. remove "calo-muons", which are identified using inner detector tracks matched to calorimeter energy deposits compatible with a minimum-ionizing particle[6], while retaining the overlapping electron; and then

2. remove electrons to suppress contributions from muon bremsstrahlung while retaining the overlapping muon.

Note that the baseline muon definition vetos calo-muons, so the central “Overlap Removal“ tool is a redundant check on the overlap removal of calo-muons.

Overlap removal between electron and jet candidates aims to resolve two sources of ambiguity:

1. remove jets that are formed primarily from the showering of a prompt electron ($\Delta R < 0.2$) while retaining the overlapping electron; and

2. remove electrons that are produced in the decay chains of hadrons ($\Delta R < 0.4$) while retaining the overlapping jet.

Overlap removal between muon and jet candidates aims to resolve two sources of ambiguity:

[6]They are only used in the narrow region $|\eta| < 0.1$ to recover acceptance where services cause a gap in muon coverage

1. remove jets[7] that appear in close proximity to muons due to high-p_{T} muon bremsstrahlung while retaining the overlapping muon; and

2. remove muons that are likely to have originated from the decay of hadrons while retaining the overlapping jet.

Reject	Keep	Criteria
electron	muon	shared ID track
jet	electron	$\Delta R < 0.2$
electron	jet	$\Delta R < 0.4$
jet	muon	NumTrack < 3 and $\Delta R < 0.2$
muon	jet	$\Delta R < 0.4$

Table 6.2: Summary of overlap removal procedure which is performed in the order listed from top to bottom. If an electron and a muon share an ID track, the muon is chosen over the electron. If a jet and an electron are found within $\Delta R < 0.2$, the object is interpreted as an electron and the overlapping jet is removed from the event. If a jet and an electron are found within $0.2 < \Delta R < 0.4$, the object is interpreted as jet and the electron is removed. If a jet and a muon are found within $\Delta R < 0.4$, the object is interpreted as jet and the muon is removed.

[7] These jets usually have very few matching ID tracks.

6.2.6 Event cleaning cuts

All events are required to pass the recommended event cleaning cuts.

- Using the recommendations of the DataPrep group[8], Events are vetoed if they have the following, or they are flagged as having an incomplete data event flag.
  - xAOD::EventInfo::LAr: LAr noise burst and data corruption.
  - xAOD::EventInfo::Tile: Tile corrupted events.
  - xAOD::EventInfo::SCT: events affected by the SCT recovery procedure for single event upsets.
- Events are required to have a primary vertex with at least two tracks with p_{T}>500 MeV.
- Events are vetoed if they contain a bad jet that doesn't satisfy the "baseline" requirement.
- Events from a given data taking period are required to pass the appropriate Good-Run-List (GRL) designated by the detector subsystem exports. The GRL used in this measurement is summarized below:
  - 20170619/data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good_25ns.xml
  - 20180129/data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml
  - 20180619/data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml
  - 20190318/data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml

[8] https://twiki.cern.ch/twiki/bin/viewauth/Atlas/DataPreparationCheckListForPhysicsAnalysis

Chapter 7

Selection and Measurement

Observables

Using the objects defined in Section 6.1, kinematic observables that are being unfolded in this analysis will be defined in Section 7.1. The particle-level selection that defines the fiducial region of this measurement will be discussed in Section 7.3, and the reconstruction-level selection that aims to emulate the particle-level selection will be discussed in Section 7.4. Finally, the measurement observables and their tailored selections are discussed in Section 7.5.

7.1 Definition of observables

Making use of the objects defined in Section 6.2, several observables of interest that are based on final state objects will be discussed in this section. In the collinear phase-space, the W+ jets production cross-section will also be unfolded differentially in these observables.

7.1.1 Leading jet transverse momentum

The leading jet transverse momentum $p_{\mathrm{T}}^{\mathrm{leading}}$ is given by the most energetic jet in the event. As shown in Section 7.3, $p_{\mathrm{T}}^{\mathrm{leading}} > 500\,\mathrm{GeV}$ is required to enhance the production of the collinear, soft W-boson emission.

7.1.2 Invariant mass of the leading and sub-leading jets

The di-jet invariant mass m_{jj} is computed using the leading and sub-leading jets. In order for m_{jj} to be physically meaningful, the analysis will require $N_{\text{jet}} \geq 2$ for this observable.

7.1.3 Scalar sum of transverse momentum

In events with high jet multiplicity N_{jet}, it is impractical to look at kinematic observables for each individual jet in the event. Instead, the scalar sum of p_{T} of all signal jets in the final state is used, defined in Equation (7.1).

$$S_{\text{T}} = \sum_{j} p_{\text{T}}^{\text{jet},j} \qquad p_{\text{T}}^{\text{jet},j} > 30\,\text{GeV} \tag{7.1}$$

where $p_{\text{T}}^{\text{jet},j}$ is the scalar magnitude of the jet transverse momentum vector.

7.1.4 W-boson candidate

In each event, a W-boson candidate is reconstructed from the vector sum of the $E_{\text{T}}^{\text{miss}}$ and a lepton with the highest p_{T}. We denote this as p_{T}^{W}.

7.1.5 Minimum angular separation quantity $\Delta_{\min_i}\Phi(W^{\text{cand.}}, \text{jet}_i^{100})$ and $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$

Angular quantities between a jet and the W-boson decay products are constructed to define the collinear phase-space, All of these angular quantities are calculated using jets with $p_{\text{T}}^{\text{jet}} \geq 100\,\text{GeV}$[1]. This is denoted by jet_i^{100}. A two-dimensional minimum azimuthal angular separation quantity $\Delta_{\min_i}\Phi(W^{\text{cand.}}, \text{jet}_i^{100})$ between the W candidate and the list of energetic jets with $p_{\text{T}} \geq 100\,\text{GeV}$ can be computed directly using the azimuthal angle difference, ϕ, between the W candidate and the jet object:

$$\Delta\Phi = \tan^{-1}\left(\frac{P_y^{W^{cand}}}{P_x^{W^{cand}}}\right) - \tan^{-1}\left(\frac{P_y^{jet_i^{100}}}{P_x^{jet_i^{100}}}\right), \tag{7.2}$$

where $P_{x(y)}$ represents the x(y)-components of momumtum vector of the W candidate and jet object.

However, we observe a poor correlation in this observable between the reconstructed and particle-level quantities making its measurement difficult to make use of. Instead, a three-dimensional minimum angular separation quantity $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$ between the jet and the lepton is used, where ΔR is computed as:

$$\Delta R = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}, \tag{7.3}$$

where Δy and $\Delta\phi$ are the rapidity[2] and azimuthal difference between the lepton and closest

[1] Note that other jet counting observables require p_{T}> 30 GeV. It is only in the calculation of the $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$ observable that we apply $p_{\text{T}}^{\text{jet}} \geq 100\,\text{GeV}$

[2] Typically, pseudo-rapidity (η) is used for angular variables. Rapidity (y) is used as it is potentially

jet. As will be shown later in the unfolding section Section 10.4.1, this $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ observable has an excellent correlation between the reconstructed and particle-level selection.

7.1.6 Ratio of the W-boson p_{T} to the closest jet p_{T} (p_{T}^{W} / $p_{T}^{\text{closest jet}}$)

The closest jet to the W-boson candidate is identified using the two-dimensional minimum angular separation $\Delta_{\mathrm{min}_i}\Phi(W^{\mathrm{cand.}}, \mathrm{jet}_i^{100})$, as described in Section 7.1.5. The distribution of the ratio of transverse momentum (p_{T}) of the W-boson candidate to its closest jet provides spatial topology and kinematics of a collision event. In the case of a single W boson produced with a single jet, the W boson has to recoil against the jet by conservation of momentum. Hence, the p_{T} ratio of the W boson to the closest jet will center around a value of 1. In contrast, for events produced with a single W boson and two jets, the W boson is most likely produced by emission from one of the two jets, and the W boson will pick up part of the momentum from the jet where it is emitted. Therefore, this p_{T} ratio will generally be less than 1. For events with more than two jets in the final state, this situation is more complex, with the closest jet not necessarily originating from where the W boson is emitted. Hence, this ratio can be significantly larger than 1. All these situations are equally important to be modeled accurately to describe the W+jets process and its production topology.

more reliable for higher p_{T} jets.

7.2 Collinear & back-to-back phase-space

Using observable $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ defined in Section 7.1.5, this analysis defines the collinear phase-space as $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ <2.6 requirement, and the complementary phase-space, $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ >2.6, is referred to as the back-to-back phase-space.

7.3 Fiducial region definition

This measurement is conducted within a defined fiducial region established by a set of particle-level selection criteria, which specifically targets the enhancement of the collinear phase-space. The truth collection of jets utilized in this measurement is from the `AntiKt4TruthWZJets` object container. Events used in this measurement must satisfy the following conditions: there must be exactly one lepton with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.5$, along with at least one anti-k_t jet with $p_{\mathrm{T}}^{\mathrm{jet}} > 500\,\mathrm{GeV}$. As discussed in Section 6.2.5 and Table 6.2, an overlap removal procedure is implemented for both leptons and jets with $\Delta R < 0.4$ to ensure unambiguous reconstruction of leptons and jets. Consequently, muons and electrons used in this measurement must be separated by $\Delta R > 0.4$ from any jet in the event with $p_{\mathrm{T}}^{\mathrm{jet}} > 30\,\mathrm{GeV}$. The full selection for the fiducial region is summarized in Table 7.1.

The presence of a high-momentum jet enhances events with multiple QCD emissions. To further increase the purity of W+jets in the measurement, we construct an angular variable to separate the phase space into two topologies: collinear and back-to-back.

Cut	Event selection
Angular separation of leptons and jets	$\Delta R(\ell, jet) > 0.4$
Number of leptons	= 1
Lepton momentum	$p_T^{\ell} > 30$ GeV
Lepton pseudo-rapidity	$\lvert\eta\rvert < 2.5$
Number of jets	≥ 1
Jet transverse momentum	$p_T^{jet} > 30$ GeV
Jet rapidity	$\lvert y\rvert < 2.5$
Leading jet transverse momentum	$p_T^{jet} > 500$ GeV
Region	selection
Inclusive	All cuts above
Inclusive 2j	Inclusive, plus $n_{jets} \geq 2$
Collinear	Inclusive plus $\Delta R_{\min_i}(\ell, \text{jet}_i^{100}) < 2.6$
Back-to-back	Inclusive plus $\Delta R_{\min_i}(\ell, \text{jet}_i^{100}) > 2.6$

Table 7.1: Summary of the fiducial selections used to defined the measurement region phase-space.

As discussed in the previous section, the $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ variable is derived from the angular separation of the lepton and the closest jet with transverse momentum greater than 100 GeV. The back-to-back phase space ($\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100}) > 2.6$) is dominated by single-jet emission, corresponding to s-channel W-boson production recoiling against a single jet. The collinear phase space ($\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100}) < 2.6$) is mostly populated by final states with $N_{\mathrm{jet}} \geq 2$, where the W-boson is radiated from a jet emission that balances the hard jet in the event.

To understand the different event topologies, the distribution of $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ for different number of particle-level jet emissions is shown in Figure 7.1. The predictions are obtained using the Sherpa 2.2.11 MC at particle level. This shows that the collinear phase-space is sensitive to vector bosons produced with multiple additional QCD emissions. Distributions of other kinematic observables for different number of final state jets in the inclusive and collinear topologies are shown in Figures 7.2 to 7.4.

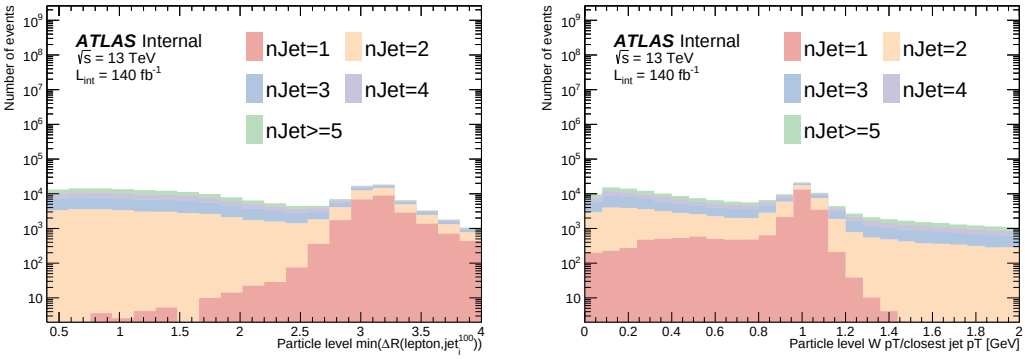
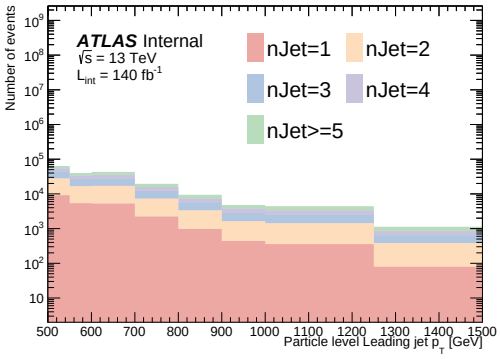
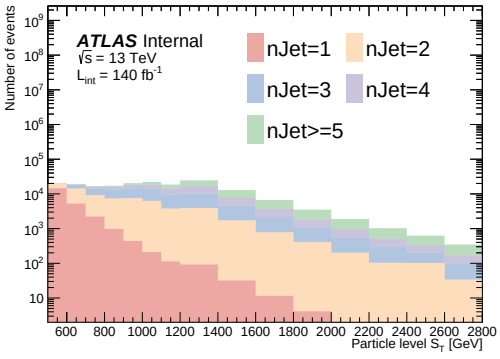


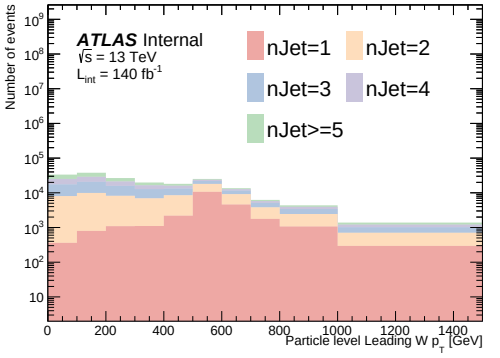
Figure 7.1: Distribution of the number of events as a function of the ΔR between the lepton and closest jet with $p_{\mathrm{T}}^{\mathrm{jet}} > 100\,\mathrm{GeV}$ (a) and the ratio of the W-boson p_{T} and closest jet p_{T}, in the **inclusive** phase-space. The predictions are obtained using the SHERPA 2.2.11 MC at particle level.



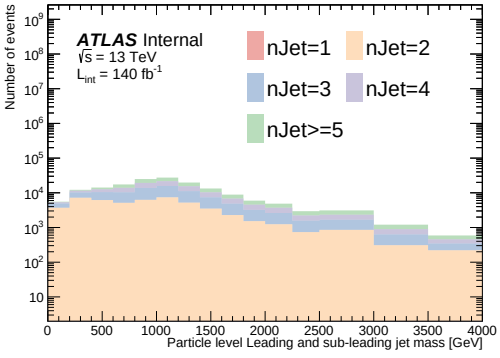
(a) $p_{\mathrm{T}}^{\mathrm{leading}}$



(b) S_{T}

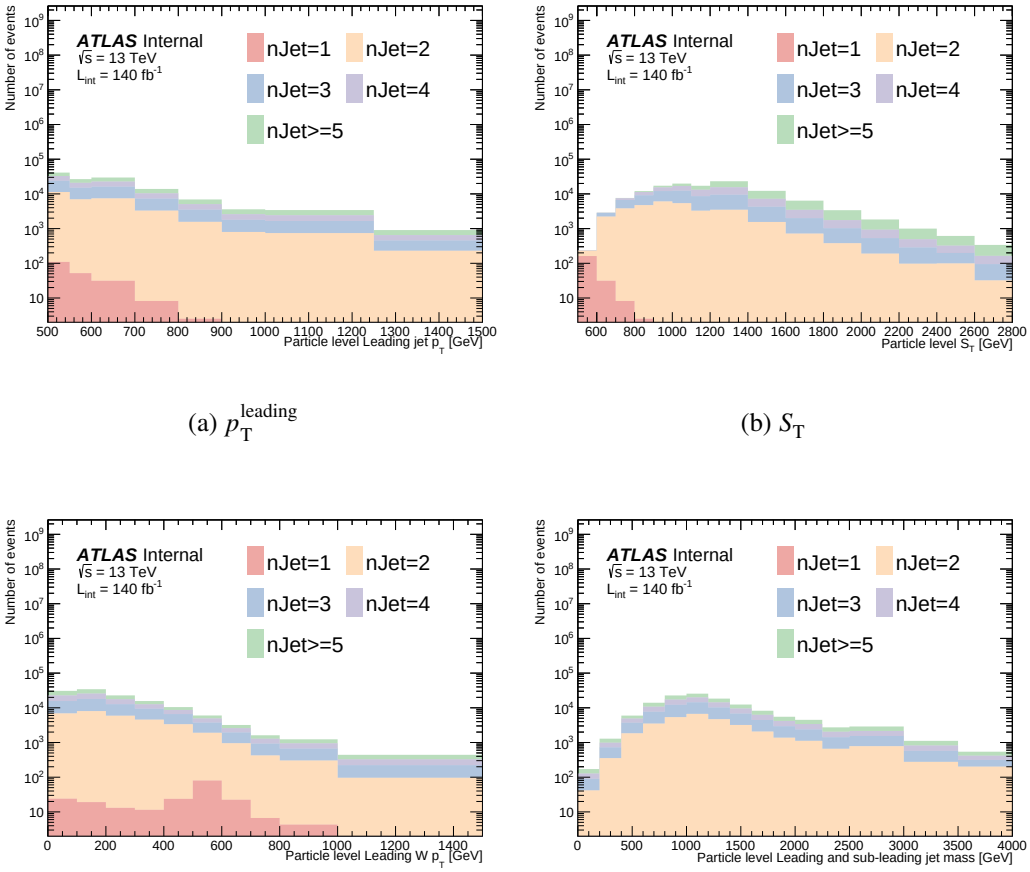


(c) W-boson p_{T}



(d) di-jet mass with at least 2 jets selection

Figure 7.2: Distribution of the number of events as a function of the various observables in the **inclusive** phase-space. The dijet mass distribution has additional requirement of at least 2 jets in the event. The predictions are obtained using the SHERPA 2.2.11 MC at particle level.



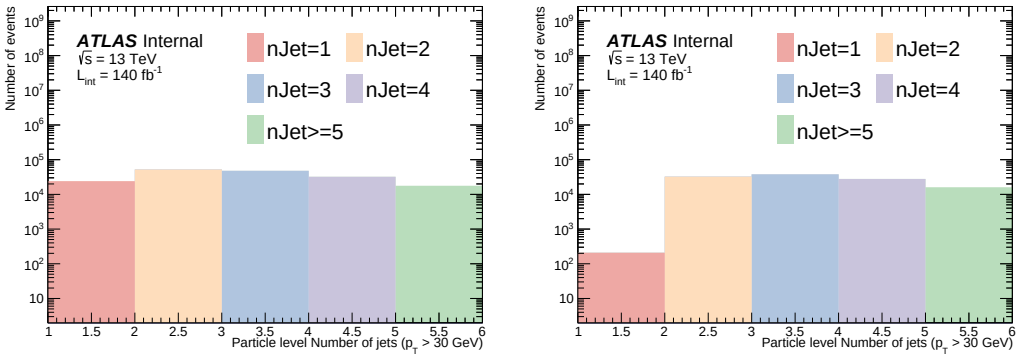
(a) $p_{\mathrm{T}}^{\mathrm{leading}}$

(b) S_{T}

(c) W-boson p_{T}

(d) di-jet mass

Figure 7.3: Distribution of the number of events as a function of the various observables in the **collinear** phase-space. The dijet mass distribution has additional requirement of at least 2 jets in the event. The predictions are obtained using the SHERPA 2.2.11 MC at particle level.



(a) Number of jets (inclusive)

(b) Number of jets (collinear)

Figure 7.4: Distribution of the number of events as a function of the number of jets with p_{T}> 30 GeV in the **inclusive** phase-space (a) and collinear phase-space (b). The predictions are obtained using the SHERPA 2.2.11 MC at particle level.

7.4 Reconstruction-level selection

Events are selected with the same criteria as the particle-level selection described in the previous section, but in addition are required to pass a series of quality criteria to ensure they were collected during periods of nominal detector operations. Specifically the selection requires a lepton with $p_{\mathrm{T}}>$ 30 GeV and at least one jet with $p_{\mathrm{T}}>$ 500 GeV. The lepton must pass all kinematic, geometric and object-level signal criteria defined in Sections 6.2.1 and 6.2.2, and must be matched to a single lepton trigger with offline $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. The event must only contain one signal lepton; any events with an additional signal or baseline leptons in the event are vetoed. Furthermore, a b-jet veto with a 60% efficiency working-point is applied to reduce the top-quark backgrounds. A summary of the reconstruction-level selections can be found in Table 7.2.

The reconstruction-level modeling with full systematic uncertainty is presensted in Chapter 9, following an overview of the background estimation strategy (Chapter 8) and the discussion on systematic uncertainties (Chapter 11).

Reconstructed level-cuts	Event selection
Angular separation of leptons and jets	$\Delta R(\ell, jet) > 0.4$
Number of leptons	= 1
Lepton momentum	$p_T^{\ell} > 30$ GeV
Pseudo-rapidity	electron $\|\eta\| < 2.47$ (electrons have crack veto), muon $\|\eta\| < 2.4$
Lepton trigger	Un-prescaled lepton triggers and matched.
Number of jets	>= 1
Jet transverse momentum	$p_T^{jet} > 30$ GeV
Jet rapidity	$\|y\| < 2.5$
Leading jet transverse momentum	$p_T^{jet} > 500$ GeV
B-jet veto	$\text{no}_{\text{b-jet}}$, $p_T^{\text{b-jet}} > 30$ GeV

Table 7.2: Detector-level selections used to define the measurement region at reconstruction level.

7.5 Measurement observables

This analysis will measure the observables defined in Section 7.1 in various selections beyond the inclusive selection described in the previous sections. In the fully inclusive selection defined in Table 7.1, the $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ and the ratio of the W-boson candidate p_{T} over the closest jet p_{T} observables will be measured. In addition, we will measure the m_{jj} distribution in the inclusive selection with at least two jets that satisfy the signal object definition. While our selection is strongly dominated by events with at least two jets, there remains a small rate for events with a single jet. In order to reconstruct the m_{jj} observable, we therefore request that the event has at least two jets. We refer to this region as the inclusive 2-jet region.

Finally, beyond the inclusive selection, we measure several observables in the collinear phase space defined by $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ <2.6. Within this selection, we measure the differential distributions of jet multiplicity, $p_{\mathrm{T}}^{\mathrm{leading}}$, S_{T}, and p_{T}^{W}. Additionally, the integrated cross-section for the back-to-back phase space is also measured.

A summary of the selection regions and relevant observables measured in each region is provided in Table 7.3. Measurements are performed separately for the electron and muon channels, and the final results are reported as the average of the two measurements. More details about the averaged results can be found in Section 12.1.

Table 7.3: Summary of the observables measured each region in this analysis. The additional selection column specifies what selection is applied above the fully inclusive selection defined in Table 7.1.

Region	Additional selection	Observables
Inclusive		$\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$, p_{T}^{W} / $p_T^{\text{closest jet}}$
Inclusive 2 jet	≥ 2 jet	m_{jj}
Collinear	$\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100}) < 2.6$	n-jet, $p_{\mathrm{T}}^{\text{leading}}$, p_{T}^{W}, S_{T}
Back-to-back	$\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100}) >= 2.6$	integrated cross-section

Chapter 8

Background Estimation

8.1 Overview of backgrounds and estimation methods

Several processes can give rise to a signature similar to collinear W+jets. We choose to subtract the contributions from these *background processes* from data before performing the unfolding. It is important to understand their modeling well as a function of each of the unfolding variables so that a bias is not introduced.

Backgrounds arise from leptons produced in the electroweak (EW) decay of weak bosons or top quarks, and so are associated with leptons from $t\bar{t}$, Z+jets, diboson, and single top-quark production. Backgrounds also arise in more generic multi-jets events, for which the leptons are either non-prompt (arising from heavy-quark decay), part of the hadronic shower in the calorimeter system (significant for muon backgrounds only), or as fakes due to mis-reconstruction. The dominant background are $t\bar{t}$, Z+jets and multi-jets production.

The $t\bar{t}$ and Z+jets processes contribute to both the inclusive and collinear phase-spaces. These contributions are estimated using a data-driven approach, where the MC samples are scaled with normalization factors derived from data to match the observed data yield in separate Z+jets and $t\bar{t}$ control regions. The estimated backgrounds from Z+jets and $t\bar{t}$ processes account for approximately 5% (4%) and 8% (7%) of the collinear phase-space for electrons (muons), respectively.

The production of multi-jets is also a significant background primarily dominating in the back-to-back electron channel. The estimation of multi-jets event background follows a similar approach to that of $t\bar{t}$ and Z+jets processes, utilizing Monte Carlo

samples. However, other sources contributing to the multi-jet background originate partly from the mis-reconstruction of leptons from partial jet energy clusters, which may not be accurately modeled by multi-jets MC samples compared to other more fundamental processes. Therefore, the estimation of the multi-jets background for the electron signal sample is cross-checked with a data-driven "ABCD" approach (described in the Section 8.5), and both approaches are found to be consistent within systematic uncertainties. The multi-jets backgrounds estimated using Monte Carlo are expected to account for 9% (3%) of the collinear phase-space for electrons (muons), respectively.

Finally, the single top-quark and diboson contributions are estimated directly from MC, and are expected to account for 2% (2%) and 3% (3%) of the collinear phase-space for electrons (muons), respectively. The total background composition in each region is summarized in Table 8.1. These fractions are computed with respect to the total background + signal MC yields *after* the application of the correction factors described in the next sections.

	Inclusive		Inclusive 2j		Collinear	
Background (%)	Electron	Muon	Electron	Muon	Electron	Muon
Z + jets	4.1	3.8	4.4	4.0	4.9	4.3
$t\bar{t}$	6.1	6.1	6.7	6.6	7.2	7.3
Single top-quark	1.7	1.8	1.9	1.9	1.9	2.0
Diboson	2.8	3.0	2.8	3.0	2.7	2.8
$\tau\nu$+jets	5.1	4.9	5.0	4.8	4.7	4.4
V+gamma+jets	3.4	4.0	3.4	4.0	3.4	3.8
Dijets	14.7	3.9	13.7	3.9	10.9	3.7
EW W+2j	3.2	3.4	3.2	3.4	3.4	3.5

Table 8.1: Summary of the proportional dominant background contributions to the measurement in the collinear phase-space for electron and muon channels. The contribution is defined where the total is with respect to the sum of all MC (backgrounds plus signal W+ jets). The fractions are extracted *after* the data-driven normalization factors are applied.

8.2 Data-driven Background normalization factors

The MC normalization factors for $Z+$jets, $t\bar{t}$, and multi-jets backgrounds are determined using dedicated control regions, as defined in Section 8.3 for these backgrounds. The goal of this method is to extract normalization factors for each of the backgrounds without introducing bias from contributions of other backgrounds that are known to be mis-modeled in the high momentum region of the phase-space we explore.

In this method, two free-floating fitting parameter functions, $\mu_r(x)$, are used for the $t\bar{t}$ and $Z+$jets contributions, while a separate fitting parameter function, $\nu_{jj}(p_{\mathrm{T}}, \eta)$, is used for the multi-jet contribution. As shown later in Tables 8.3 and 8.4, the $Z+$jets and $t\bar{t}$ control regions (CRs) are highly enriched, and the multi-jet contribution in these regions is minimal. Therefore, the $\mu_r(x)$ parameters are allowed to vary across the control regions in a fully correlated manner to adjust the overall amount of the corresponding background, without bias from multi-jet modeling. The $\nu_{jj}(p_{\mathrm{T}}, \eta)$ parameter is derived independely after applying the μ_r to the corresponding backgrounds.

All other contributions, including $W+$jets, are estimated from MC simulations and subtracted from the data. These contributions are denoted as $N_{other}^{\mathbf{r}^{CR}}(x)$, where $\mathbf{r}$ represents contributions to the $Z+$jets, $t\bar{t}$, or multi-jet CRs. The independent variable x is chosen to be the exclusive jet multiplicity, with a transverse momentum threshold of $p_{\mathrm{T}} > 30$ GeV, and the normalization factors for the Z + jets and tt backgrounds are calculated as a function of this independent variable. This ensures that the method captures the changing perturbative accuracy of the multi-jet merged MC generators

that provide the initial estimates for these two backgrounds in this analysis. For the multi-jet background, a 2D binned parameterization using lepton p_{T} and $|\eta|$ is used to account for the varying fake contributions in different kinematic and detector regions. All normalization factors are extracted in the inclusive phase space and applied across all regions in this analysis.

To derive the normalization factors, a series of two equations with two unknown μ_r parameters can be constructed in each bin of the exclusive jet multiplicity (x), as detailed in Equation (8.1). The normalization factors are then obtained by solving this series of equations. For events involving $t\bar{t}$ and $Z+$jets with 5 or more jets, the bins are merged to derive an inclusive normalization factor. Additionally, to reduce statistical fluctuations, the 2nd and 3rd exclusive jet multiplicity bins of the $t\bar{t}$ normalization factor are averaged and merged. Subsequently, the ν_{jj} parameter for the multi-jet contribution is determined using Equation (8.2), applying the normalization factors obtained for $t\bar{t}$ and $Z+$jets accordingly. Separate binning of p_{T} and $|\eta|$ is used for electrons and muons to ensure sufficient counts per bin and minimize statistical fluctuations.

$$\begin{bmatrix} N_{tot}^{Z^{CR}}(x) \\ N_{tot}^{t\bar{t}^{CR}}(x) \end{bmatrix} = \begin{bmatrix} N_{t\bar{t}}^{Z^{CR}}(x) & N_{Z}^{Z^{CR}}(x) \\ N_{t\bar{t}}^{t\bar{t}^{CR}}(x) & N_{Z}^{t\bar{t}^{CR}}(x) \end{bmatrix} \times \begin{bmatrix} \mu_{t\bar{t}}(x) \\ \mu_{Z}(x) \end{bmatrix} + \begin{bmatrix} N_{other}^{Z^{CR}}(x) \\ N_{other}^{t\bar{t}^{CR}}(x) \end{bmatrix} \tag{8.1}$$

$$N_{tot}^{jj^{CR}} = \nu_{jj}(x_{p_{\mathrm{T}}}, x_{\eta}) \times N_{jj}^{jj^{CR}} + N_{other}^{jj^{CR}} \tag{8.2}$$

Once the normalization factors are extracted, they are applied as an event-by-event weights for the $t\bar{t}$, Z + jets and multi-jet processes, and as a function of the number of reconstructed jets or lepton p_{T} and η in the event using the initial MC estimate in each of the measurement regions. The inclusive normalization factors are applied across all regions.

8.3 Control regions

A control region is enriched in a specific process that is being "controlled" and then normalized to data. In addition to its primary purpose, a control region serves as an in-situ validation for the modeling of backgrounds across various kinematic observables to be measured. Control regions in this analysis are constructed only for backgrounds with a significant contribution in the signal regions, as outlined in Table 8.1. Backgrounds such as semi-leptonic diboson and single top-quark production, which contribute less than 5%, are estimated purely through MC simulations. Normalization factors are extracted from the control regions as a function of jet multiplicity for $t\bar{t}$ and Z + jets, and lepton (p_{T}, η) for multi-jets, as detailed in Section 8.2. These normalization factors are derived in the inclusive selection and then applied to all other regions, including the collinear and inclusive 2-jets regions.

The control regions are constructed so they are kinematically similar to the signal regions. For this purpose, nearly all the selection criteria in the control regions are identical to that of the signal region, with the exception of a minimal number of cuts that

allow the control region to be enriched in the targeted background process. The baseline selection for the control regions is the inclusive phase-space described in Table 7.5. For all control regions, the leading jet p_{T} requirement is relaxed to $p_{\mathrm{T}}>$ 400 GeV to account for the additional phase-space that is added during the unfolding procedure to account for migrations from outside the nominal phase-space into the fiducial region[1]. The additional selection or relaxation of cuts above the inclusive phase-space for each specific control region is described below.

In addition, the control regions are also used to constrain the size of systematic uncertainties for the controlled background processes. We employ the so-called *"offset method"*, wherein the fit is repeated for each source of systematic uncertainty, and new normalization factor is derived. Systematic uncertainties considered in this analysis can be found in Chapter 11. This method effectively removes any normalization effects from sources of uncertainty on the floating backgrounds, since the control regions are fixing the normalization to the data. Instead, the uncertainties become a shape uncertainty on the extrapolation from the control to signal regions. Since our control regions are designed to be kinematically similar to the signal regions, this method will lead to a reduction in the systematic uncertainty for the $t\bar{t}$, $Z+$jets and multi-jet backgrounds. This method is performed for all sources of uncertainties for these backgrounds. For the plots shown below, the offset method has been applied; therefore the uncertainty band is quite small since the systematics have been constrained within the control regions

[1] A comparison of the extracted normalization factors for different leading jet requirements can be found in Appendix A.5.5.

themselves.

8.3.1 Z + jets and $t\bar{t}$ control regions

The $t\bar{t}$ control region is defined within the same phase-space as outlined in Table 7.2, with the exception of replacing the b-jet veto with the requirement of exactly two b-tagged jets. Similarly, the Z + jets control region mirrors the inclusive phase-space, but it replaces the single lepton requirement with exactly two same-flavor opposite-sign (SFOS) signal leptons with an invariant mass between $60 < m_{\ell\ell} < 120\,\text{GeV}$. The threshold for the sub-leading lepton transverse momentum is relaxed to p_{T}>10GeV to enhance statistics in the Z + jets control region. In the $t\bar{t}$ control region, the purity of the $t\bar{t}$ process exceeds 80%, with minor contributions from single top and W+ jets. The purity of the Z + jets process in its control region is approximately 90%, with a minor contribution from the Diboson process. A summary of the selection requirements for these two control regions can be found in Table 8.2.

The pre- and post-fit distributions of jet multiplicity are shown in Figures 8.1 and 8.2 for the electron and muon channels, respectively, in the $t\bar{t}$ control region. Similarly, the pre- and post-fit distributions for the Z + jets control region are shown in Figures 8.3 and 8.4. The normalization factors for $t\bar{t}$ and Z + jets are consistent across lepton flavors. Moreover, in the $t\bar{t}$ control region, since at least two b-tagged jets are required, no normalization factor can be derived for the 1-jet bin[2]. For the 1-jet bin in $t\bar{t}$, we

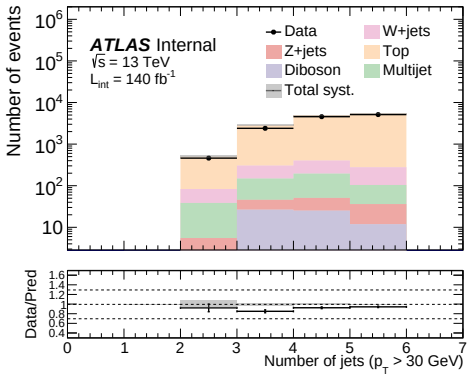
[2] Although a 1 b-tag region was explored, it was found to have significant contamination from the signal W+jets process.

Selection	$t\bar{t}$ CR	Z + jets CR
Number of leptons	1	2
Trigger	Single lepton triggers	
Leading lepton p_{T}	> 30 GeV	
Sub-leading lepton p_{T}	–	> 10 GeV
Leading jet p_{T}	> 400 GeV	
Number of b-jets	2	0
$m_{\ell\ell}$	–	[60,120]

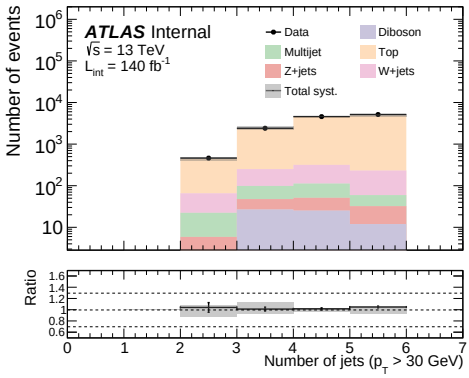
Table 8.2: Selection criteria for the $t\bar{t}$ and Z + jets control region selections. The leptons are required to pass all signal lepton criteria, with the exception of the p_{T} requirement, which is relaxed to 10 GeV for the sub-leading lepton.

assume a normalization factor of one and assign an uncertainty derived from the average uncertainties across the ≥ 2 jet bins. Nevertheless, the contribution of $t\bar{t}$ to the 1-jet multiplicity bin in our measurement region is minimal, accounting for approximately 1% of the total yield in the measurement region, as indicated in Section 9.1.

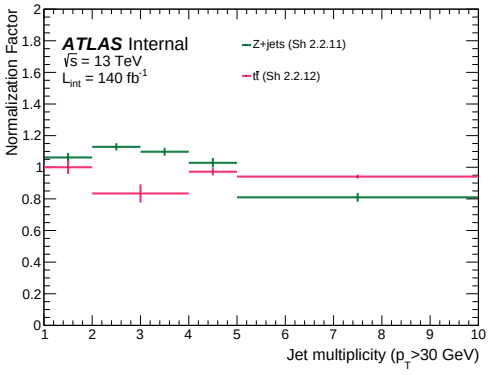
The Z + jets normalization factors exhibit a significant dependence on the jet multiplicity, which is related to the perturbative accuracy of the multi-jet merged predictions. In the Z + jets control region, Z-bosons are produced with numerous additional QCD emissions. The Z + jets production inevitably relies on the leading-order (LO) predictions within the SHERPA multi-jet merged setup. The Z + jets samples used in this analysis are with perturbative accuracy of 0-2 jets at NLO and 3-5 jets at LO. For the bins described by LO contributions (≥ 3) in Figures 8.3 and 8.4, the normalization factor decreases for every additional jet emission.



(a) Pre-fit

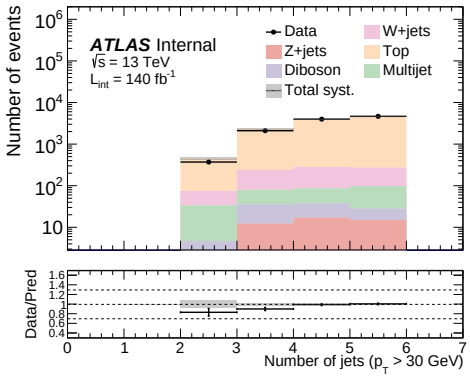


(b) Post-fit

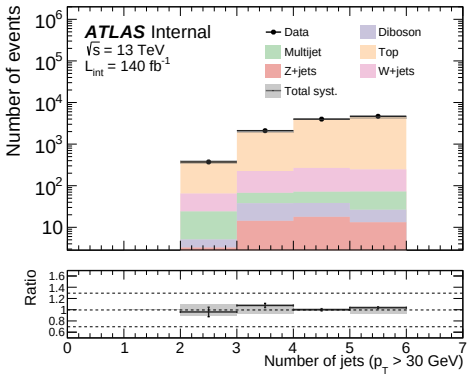


(c) Normalization factors

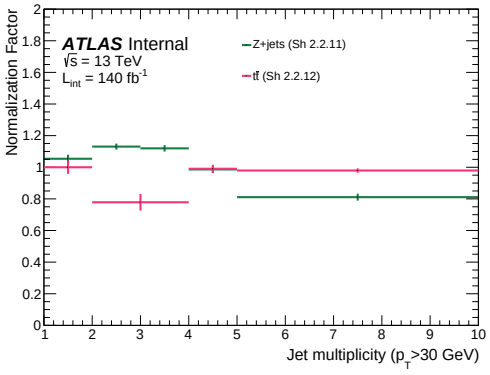
Figure 8.1: Contribution from various background and signal processes in the $t\bar{t}$ **CR electron channel**, as a function of jet multiplicity before (a) and after (b) the fit. The extracted normalization factors from the fit are shown in (c). The grey error band include statistical and systematic uncertainties added in quadrature.



(a) Pre-fit

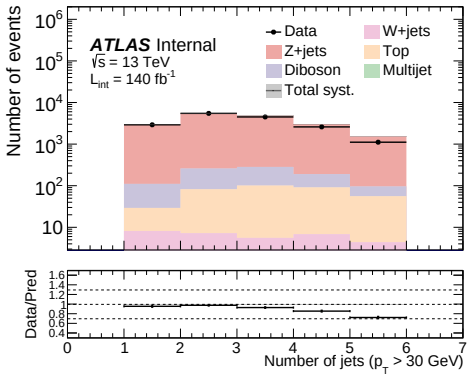


(b) Post-fit

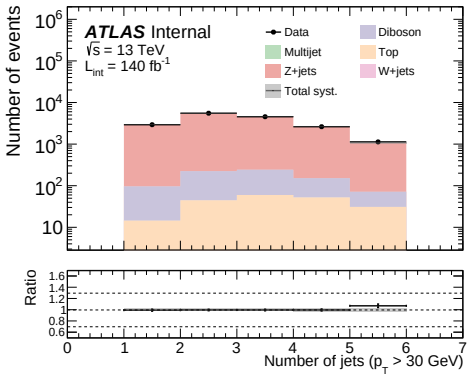


(c) Normalization factors

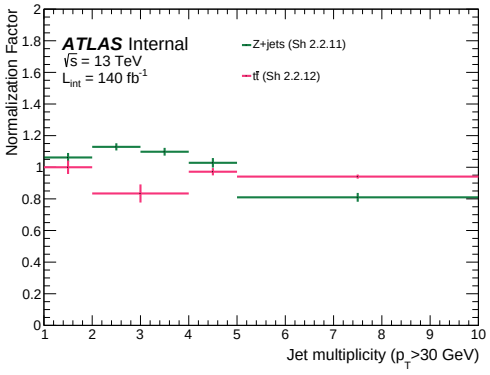
Figure 8.2: Contribution from various background and signal processes in the $t\bar{t}$ **CR muon channel**, as a function of jet multiplicity before (a) and after (b) the fit. The extracted normalization factors from the fit are shown in (c). The grey error band include statistical and systematic uncertainties added in quadrature.



(a) Pre-fit

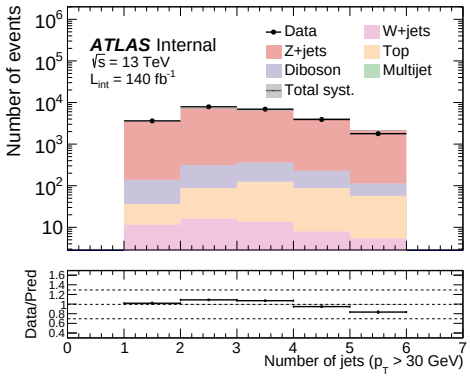


(b) Post-fit

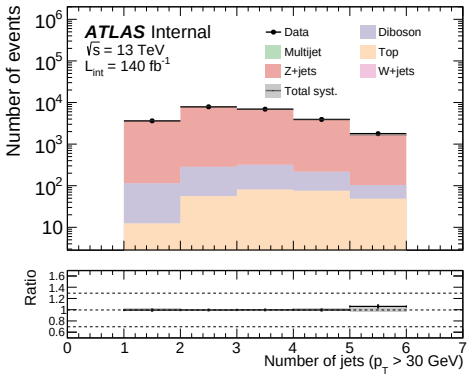


(c) Normalization factors

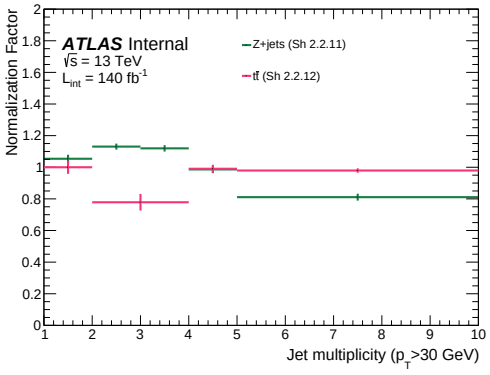
Figure 8.3: Contribution from various background and signal processes in the ***Z* + jets CR electron channel**, as a function of jet multiplicity **after** the fit. The grey error band include statistical and systematic uncertainties added in quadrature.



(a) Pre-fit



(b) Post-fit

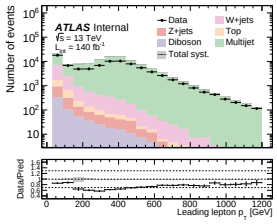


(c) Normalization factors

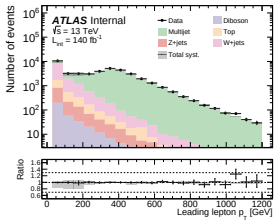
Figure 8.4: Contribution from various background and signal processes in the ***Z* + jets CR muon channel**, as a function of jet multiplicity **after** the fit. The grey error band include statistical and systematic uncertainties added in quadrature.

8.3.2 Multi-jets control region

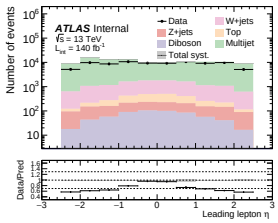
The multi-jets control region is defined within the same phase-space as described in Table 7.2, but it is required to fail either the impact parameter or lepton isolation criteria outlined in Table 6.1. Figure 8.5 and Figure 8.6 show the expected contributions from various background and signal processes as a function of lepton p_{T} and η before and after applying the normalization factors for electrons and muons, respectively. The electron and muon multi-jets control regions are 83% and 88% pure in the multi-jets background for electrons and muons, as shown in Tables 8.3 and 8.4. The normalization factors in the lepton p_{T} projection show consistent values across bins for both electrons and muons. In contrast, the normalization factors in the lepton η projection exhibit noticeable changes for electrons, particularly around $|\eta| = 1.5$, whereas the muon normalization factors remain approximately consistent across all η ranges



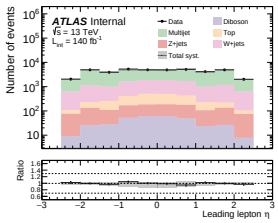
(a) pre-fit lepton p_{T}



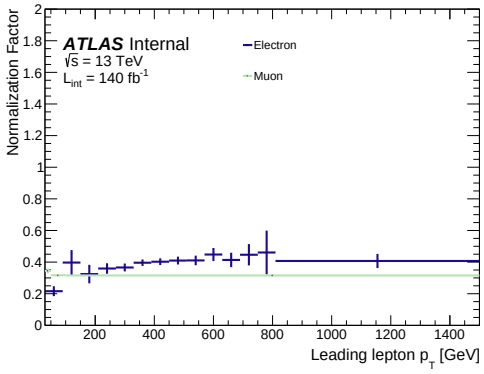
(b) post-fit lepton p_{T}



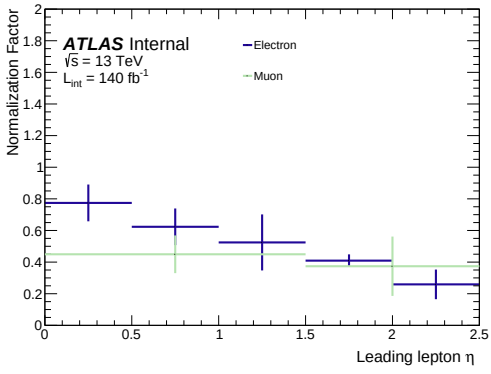
(c) pre-fit lepton η



(d) post-fit lepton η

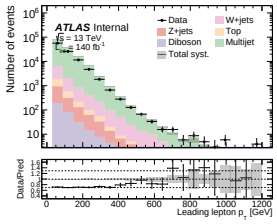


(e) projection to lepton p_{T}

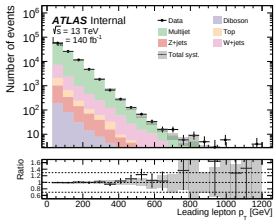


(f) projection to lepton η

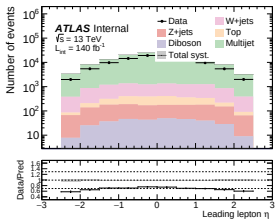
Figure 8.5: Contribution from various background and signal processes in the **di-jets CR** electron channel, as a function of lepton p_{T} and η before (a) and after (b) the fit. The projection of the extracted normalization factors with 2D parameterization from the fit are shown in (c) and (d). The grey error band include statistical and systematic uncertainties added in quadrature.



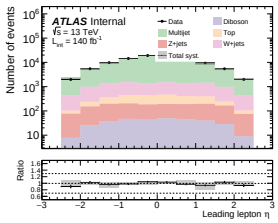
(a) pre-fit lepton p_{T}



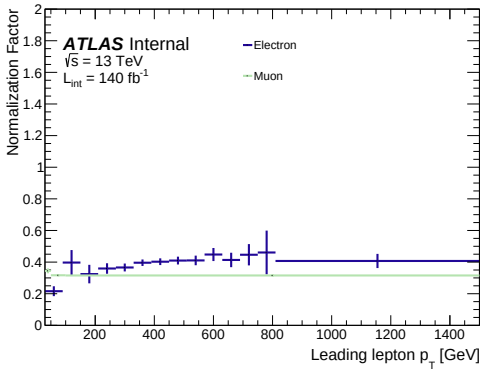
(b) post-fit lepton p_{T}



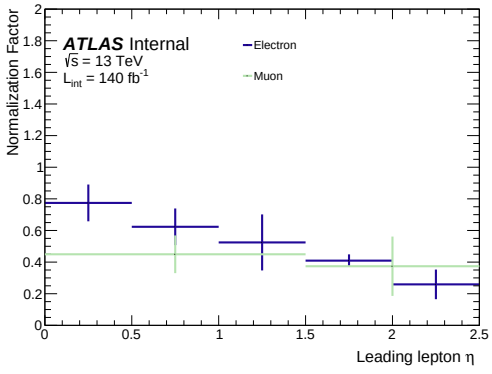
(c) pre-fit lepton η



(d) post-fit lepton η



(e) projection to lepton p_{T}



(f) projection to lepton η

Figure 8.6: Contribution from various background and signal processes in the **di-jets CR** muon channel, as a function of lepton p_{T} and η before (a) and after (b) the fit. The projection of the extracted normalization factors with 2D parameterization from the fit are shown in (c) and (d). The grey error band include statistical and systematic uncertainties added in quadrature.

8.4 Data/MC distributions in control regions

In this section, the comparison between the background models with applied normalization and the data is evaluated. After applying the normalization factors for $t\bar{t}$, Z+jets, and multi-jets backgrounds, as presented in the previous section, the yields for background and signal processes in CRs are showned in Tables 8.3 and 8.4 for electrons and muons, respectively. The statistical and systematic uncertainties, as will be discussed in Chapter 11, have been combined in quadrature, with the systematic uncertainties being symmetrized. The yields before the application of normalization factors can be found in appendix Appendix A.4.

8.4.1 $t\bar{t}$ background modeling

Data/MC distributions for the $t\bar{t}$ control region, after applying normalization factors, are presented for all unfolding observables in this section. Differential distributions with normalization factors applied are shown for electrons in Figures 8.7 and 8.8, and for muons in Figures 8.9 and 8.10. Both statistical and systematic uncertainties are included within the grey error bands[3]. The distributions prior to applying normalization factors can be found in Appendix A.4.3.

As shown in the comparison, the Sherpa 2.2.12 generator for the $t\bar{t}$ production provides an excellent description of the data, with nearly all distributions agreeing within uncertainties. The only exception is a slight 1σ over-prediction in the highest p_{T}^{W} bins that

[3] Further details on these uncertainties are discussed in Chapter 11

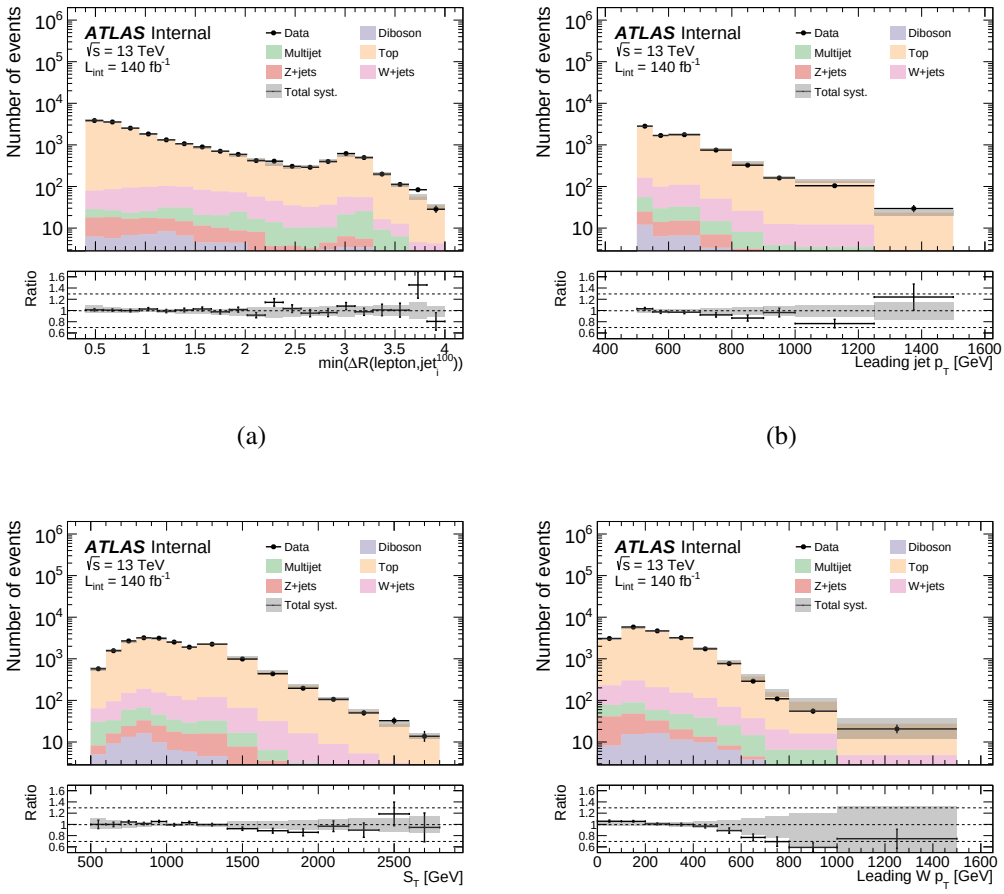
Process	$t\bar{t}$ CR	Z+jets CR	Multi-jets CR
$t\bar{t}$	16152 ± 264	195 ± 19	1378 ± 126
Diboson	76 ± 4	611 ± 15	349 ± 12
$\tau\nu$+jets	57 ± 5	1 ± 1	2613 ± 76
Single top	2395 ± 129	29 ± 2	285 ± 13
EW V+jets	8 ± 1	0 ± 0	330 ± 9
Z+jets	100 ± 6	16757 ± 37	1026 ± 29
Multi-jets	197 ± 26	0 ± 0	29975 ± 1352
W+jets	740 ± 13	6 ± 1	6860 ± 47
Total SM	19724 ± 317	17600 ± 1728	42816 ± 3801
Observed data	19973 ± 141	17591 ± 133	42761 ± 207

Table 8.3: **Electron** channel yields in the $t\bar{t}$, Z + jets and Multi-jets CRs **after** applying normalization factors. Uncertainties include statistical and systematic errors added in quadrature.

Process	$t\bar{t}$ CR	Z+jets CR	Multi-jets CR
$t\bar{t}$	14451 ± 218	273 ± 28	1257 ± 119
Diboson	70 ± 4	803 ± 16	342 ± 11
$\tau\nu$+jets	47 ± 4	1 ± 1	6544 ± 136
Single top	2121 ± 112	38 ± 3	341 ± 13
EW V+jets	8 ± 1	0 ± 0	313 ± 7
Z+jets	67 ± 4	24327 ± 55	1325 ± 19
Multi-jets	174 ± 23	1 ± 1	91396 ± 739
W+jets	768 ± 12	5 ± 1	1561 ± 26
Total SM	17707 ± 232	25448 ± 45	103079 ± 649
Observed data	17910 ± 134	25444 ± 160	103069 ± 321

Table 8.4: **Muon** channel yields in the $t\bar{t}$, Z + jets and Multi-jets CRs **after** applying normalization factors. Uncertainties include statistical and systematic errors added in quadrature.

are explored. However, the contribution from single top becomes non-negligible in this region, and significant uncertainties arises from the treatment of quantum interference between single top tW and $t\bar{t}$ processes. Nevertheless, the contribution of single top is less than 2% of the total yield, and it does not exceed 4% of the total expected yield in any bin of our observables. Thus, we believe that the conservative uncertainties applied to the complex final state are sufficient, and this analysis will not be substantially impacted by this mis-modeling. The modeling when using POWHEG +PYTHIA can be found in Appendix A.1.



(a) (b)

(c) (d)

Figure 8.7: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **electron**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

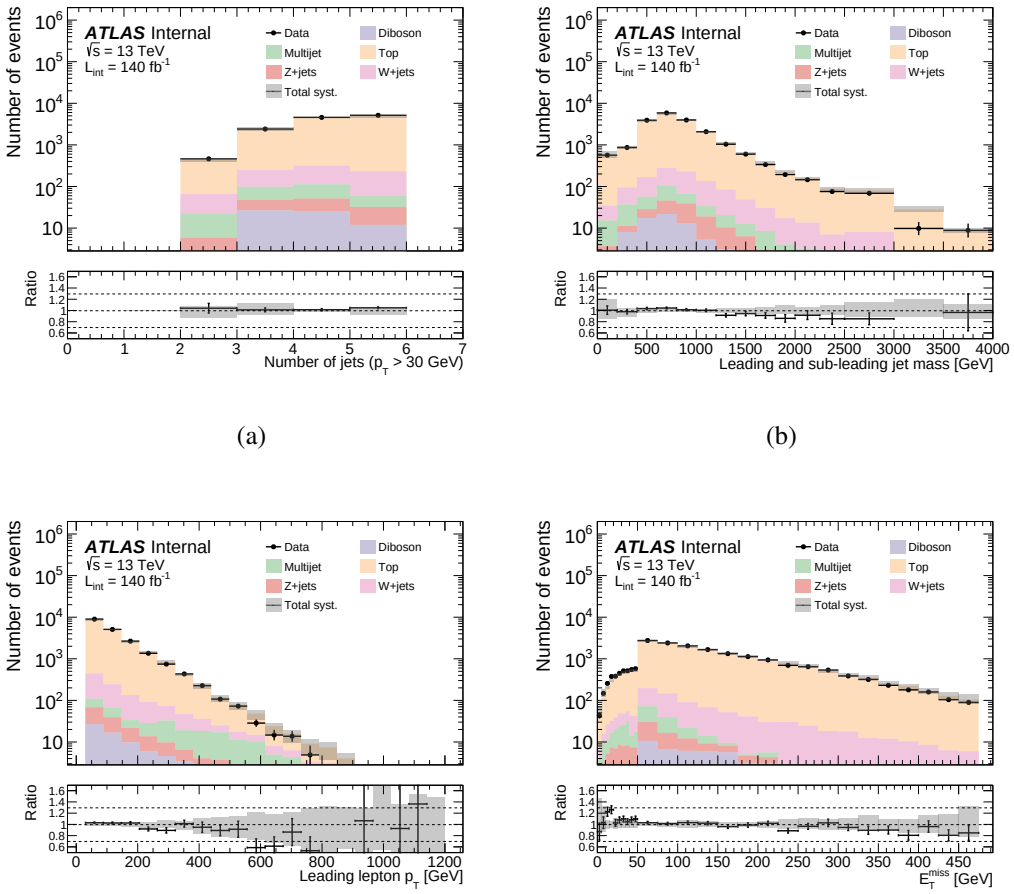
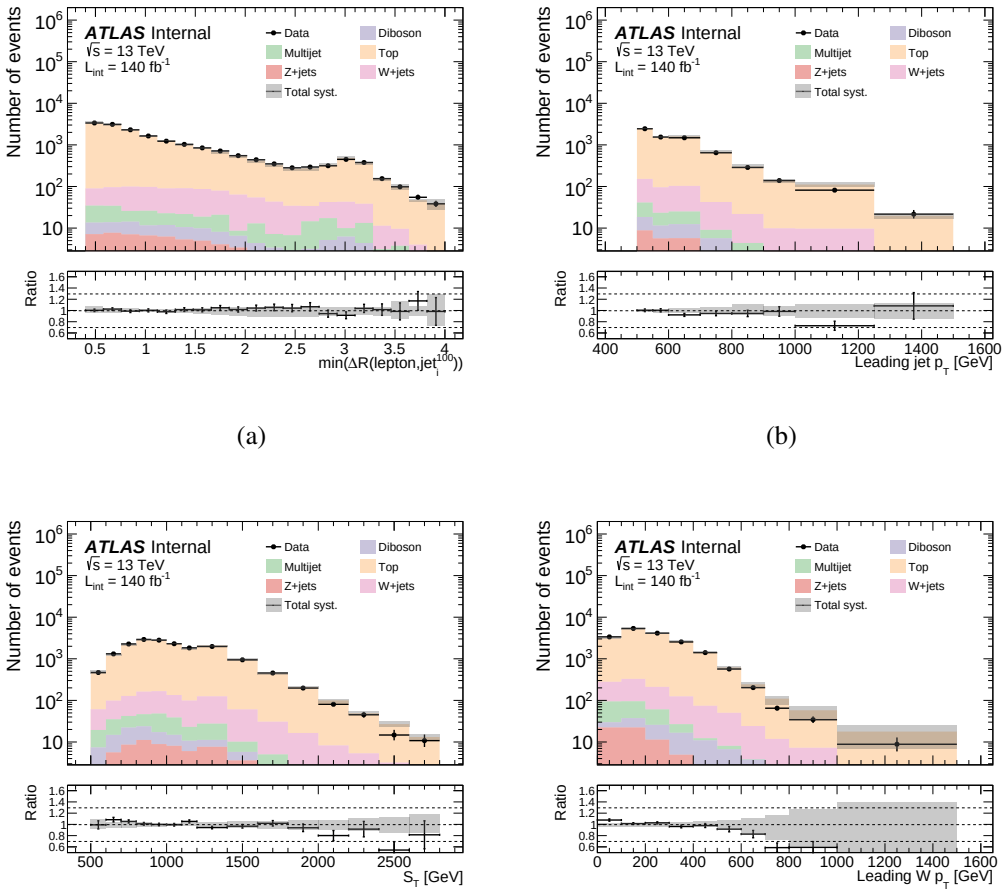


Figure 8.8: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **electron**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.



(a) (b)

(c) (d)

Figure 8.9: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **muon**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

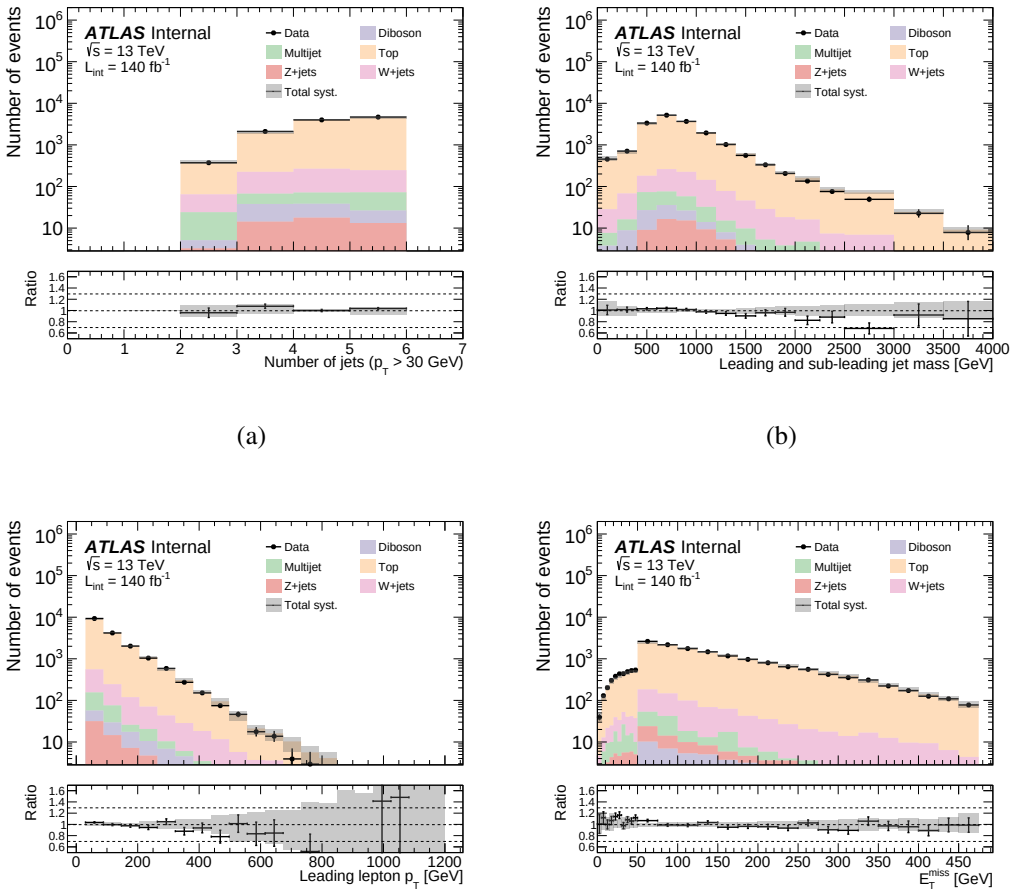


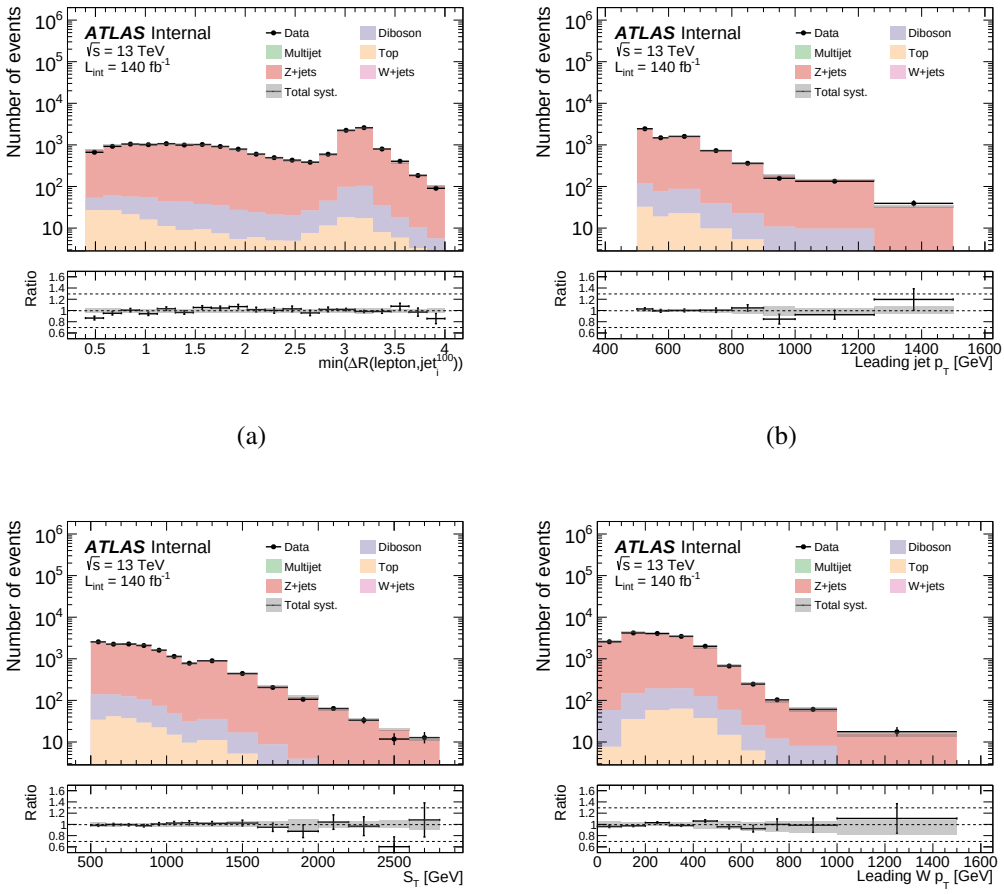
Figure 8.10: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **muon**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

8.4.2 Z + jets background modeling

Data/MC distributions for the Z + jets control region, after applying normalization factors, are presented for all unfolding observables in this section. Differential distributions with normalization factors applied are shown for electrons in Figures 8.11 and 8.12, and for muons in Figures 8.13 and 8.14. Both statistical and systematic uncertainties are included within the grey error bands[4]. The distributions prior to applying normalization factors can be found in Appendix A.4.4.

For all observables, good agreement is observed between the MC prediction and the data. Notably, the modeling of the leading jet and sub-leading jet mass (m_{jj}) in the highest bins is at the 1σ level. Additionally, distribution with modeling using the MadGraph5_aMC@NLO +Pythia8 FxFx merged setup can be found in Appendix A.3.2, and the FxFx merged setup exhibits good agreement with data and shows similar modeling performance to the Sherpa 2.2.11 prediction.

[4]Further details on these uncertainties are discussed in Chapter 11



(a) (b)

(c) (d)

Figure 8.11: Reconstructed level comparisons of data and background predictions in the $Z+$jets control region for events with two **electrons**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

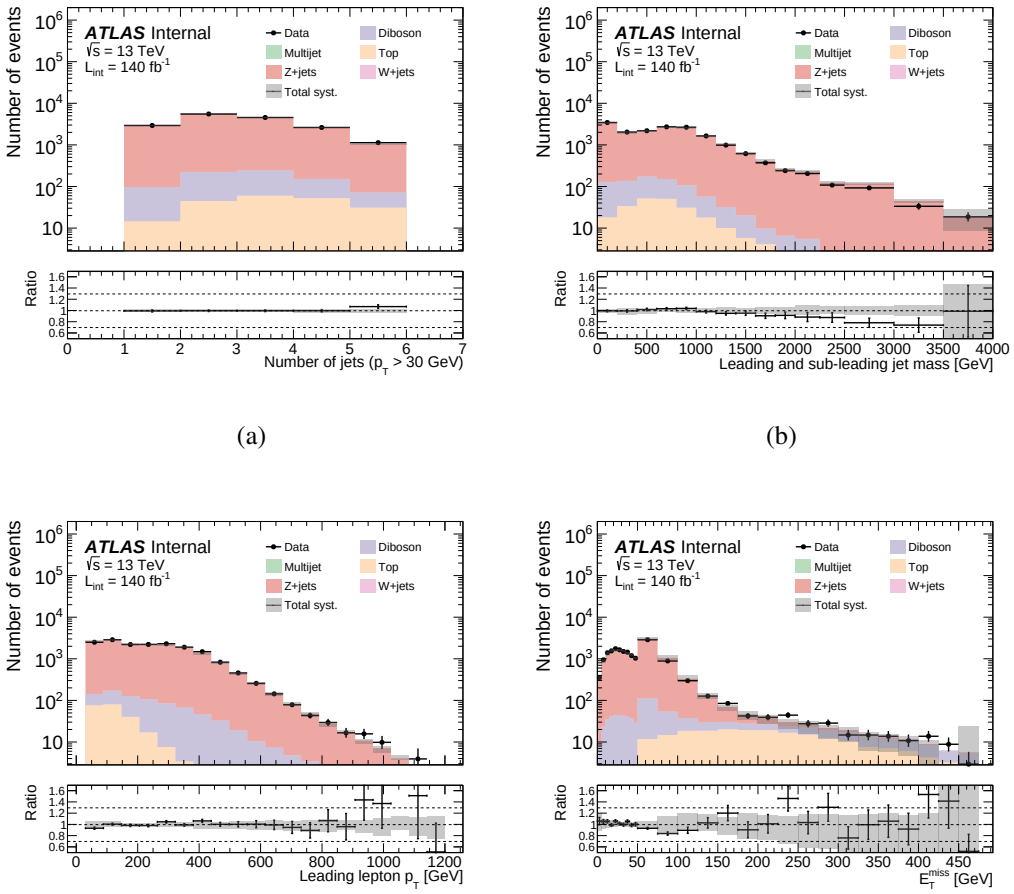
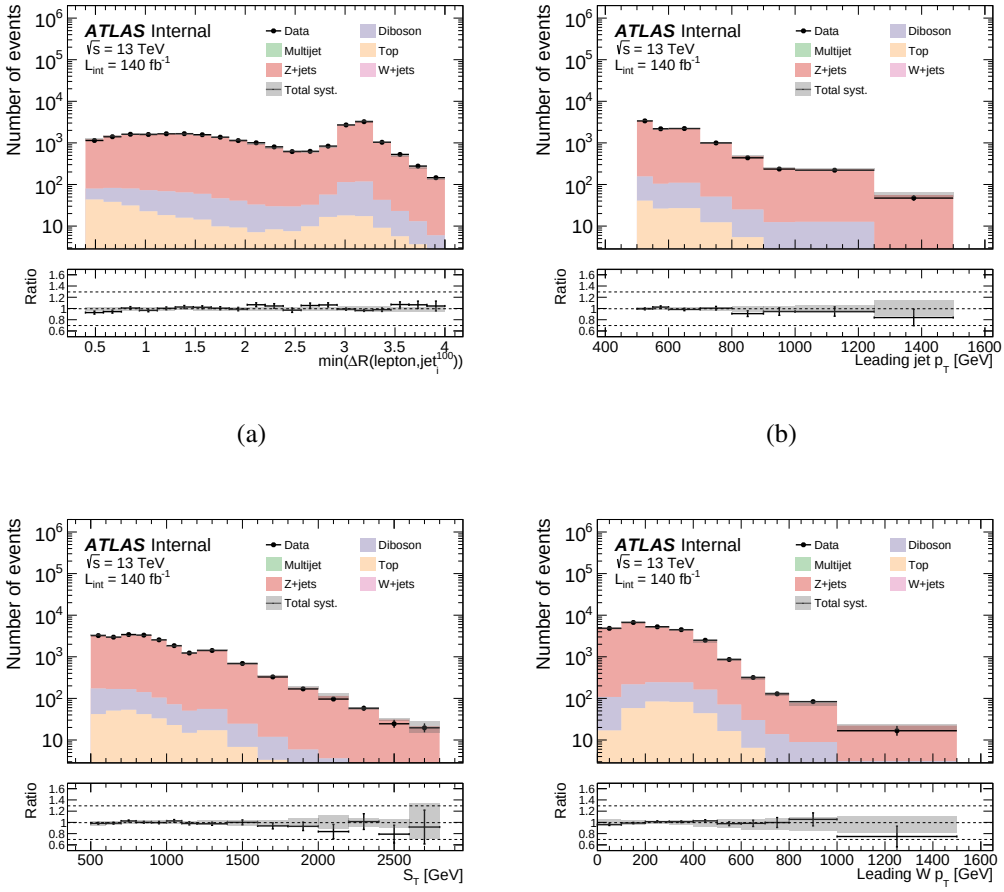


Figure 8.12: Reconstructed level comparisons of data and background predictions in the $Z+$jets control region for events with two **electrons**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.



(a) (b)

(c) (d)

Figure 8.13: Reconstructed level comparisons of data and background predictions in the $Z+$jets control region for events with two **muons**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

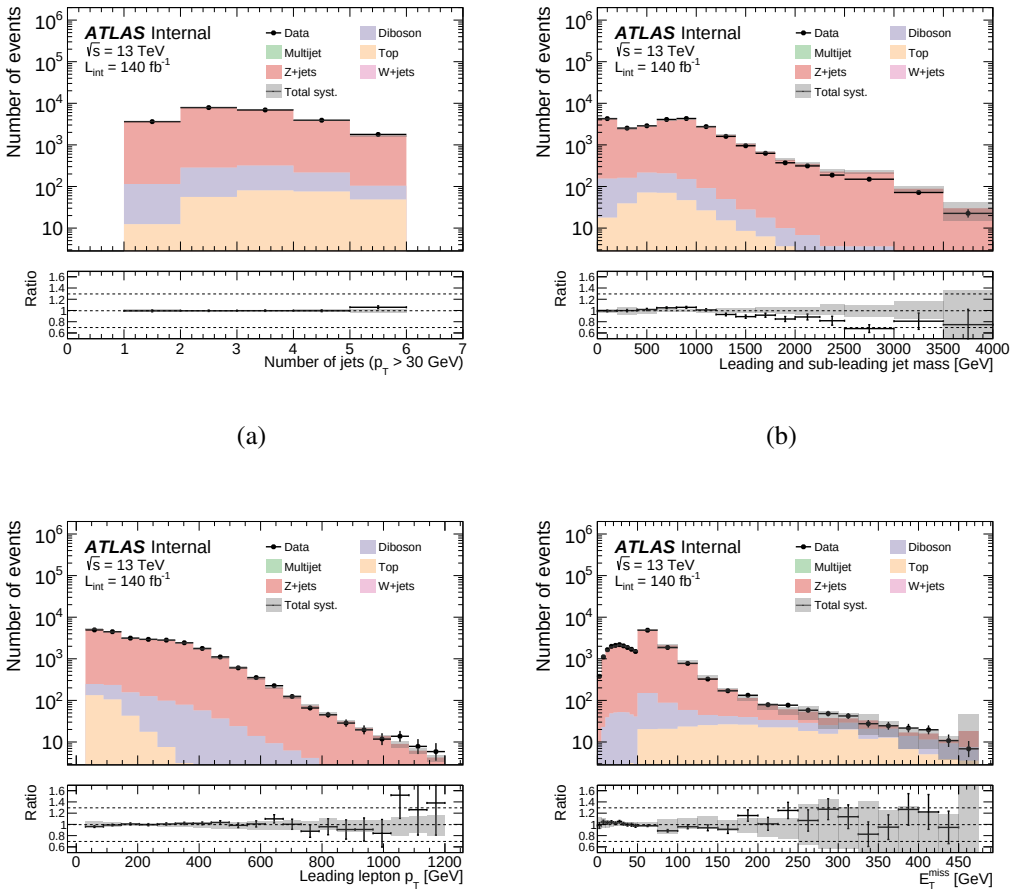
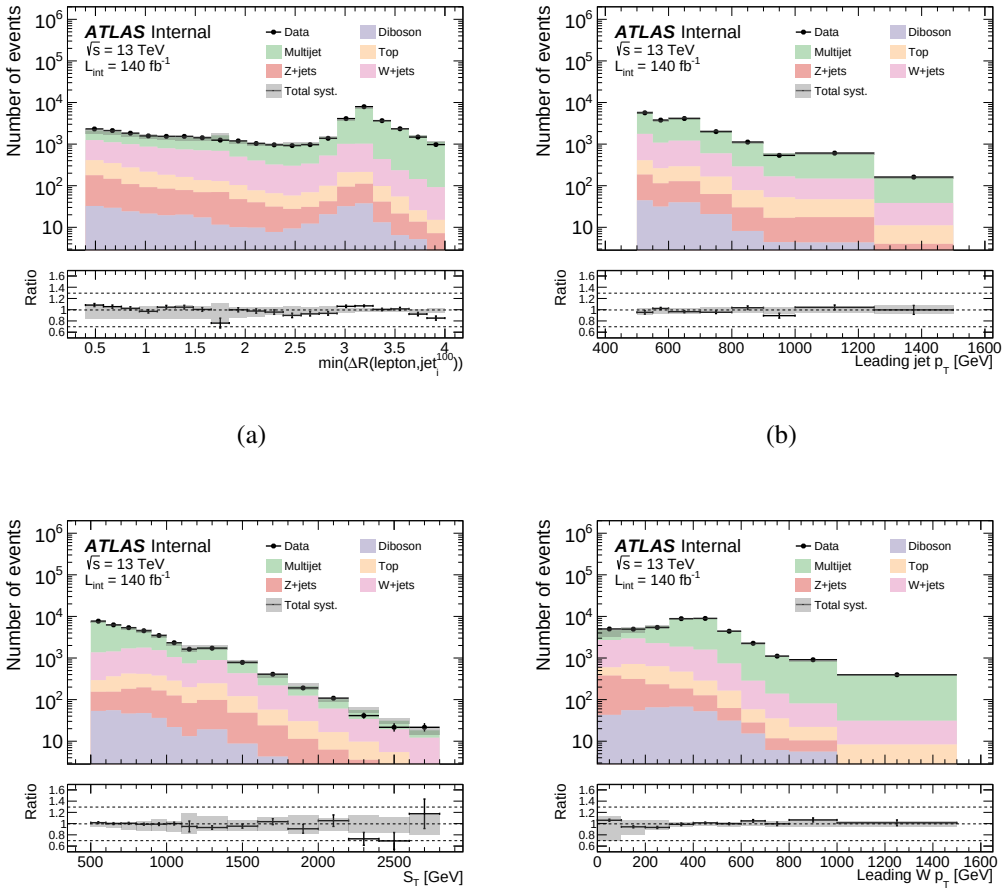


Figure 8.14: Reconstructed level comparisons of data and background predictions in the Z + jets control region for events with two **muons**. Normalization factors for $t\bar{t}$ and Z + jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

8.4.3 Multi-jet background modeling

Data/MC distributions for the multi-jets control region, after applying normalization factors, are shown for electrons in Figures 8.15 and 8.16 and for muons in Figures 8.17 and 8.18. Both statistical and systematic uncertainties are included in the grey error bands. Further details on these uncertainties are discussed in Chapter 11. The pre-fit distributions can be found in Appendix A.4.5.

The modeling of most observables in the multi-jets control regions shows good agreement with data, with similar 1σ level of difference observed in the highest bin of the m_{jj} distribution for the electron channel. Additionally, there is a noticeable modeling difference in the $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ observable for the muon channel, particularly in the region close to the back-to-back ($\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100}) > 2.6$) phase-space. This discrepancy may be explained by the substantial contributions from signal and other backgrounds that cannot be corrected with normalization factors, such as the τ channel of the W+ jets process. Therefore, alternative approaches using data-driven ABCD methods are employed to validate the normalization factor corrected multi-jets MC prediction. Validation regions are also utilized to assess the estimation of multi-jets and fakes.



(a) (b)

(c) (d)

Figure 8.15: Reconstructed level comparisons of data and background predictions in the multi-jets control region for **electrons**. Normalization factors for multi-jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

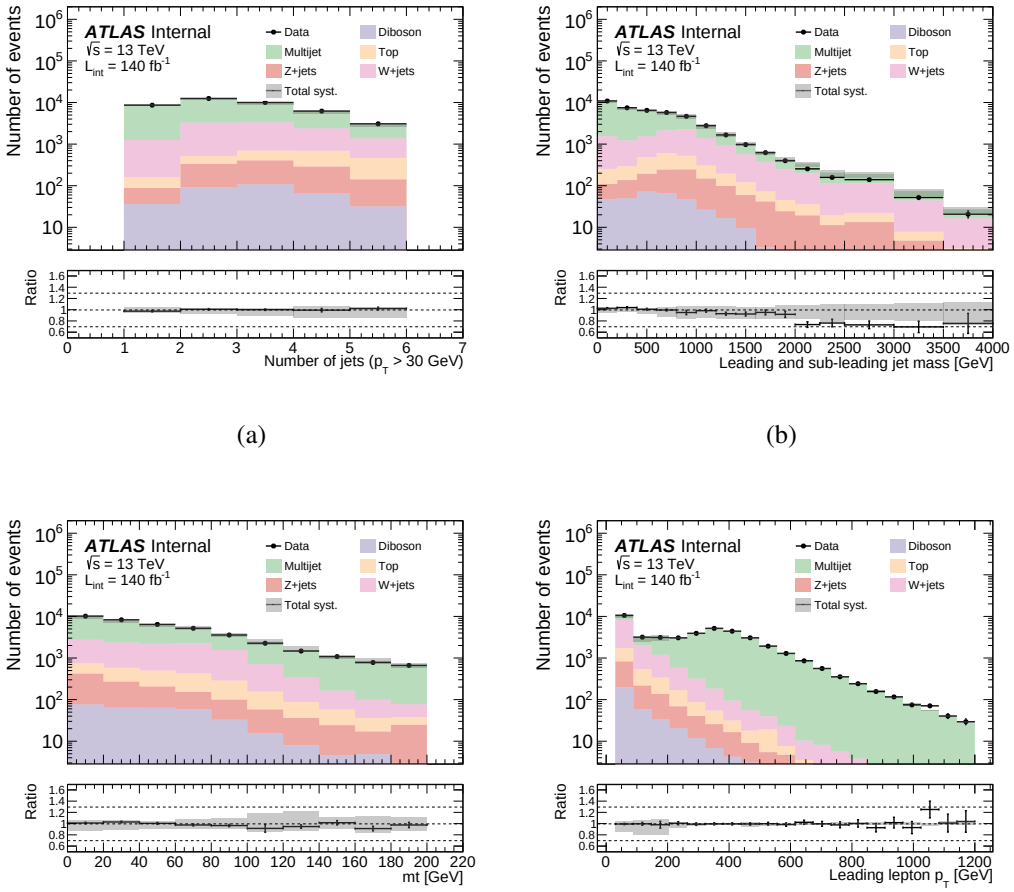
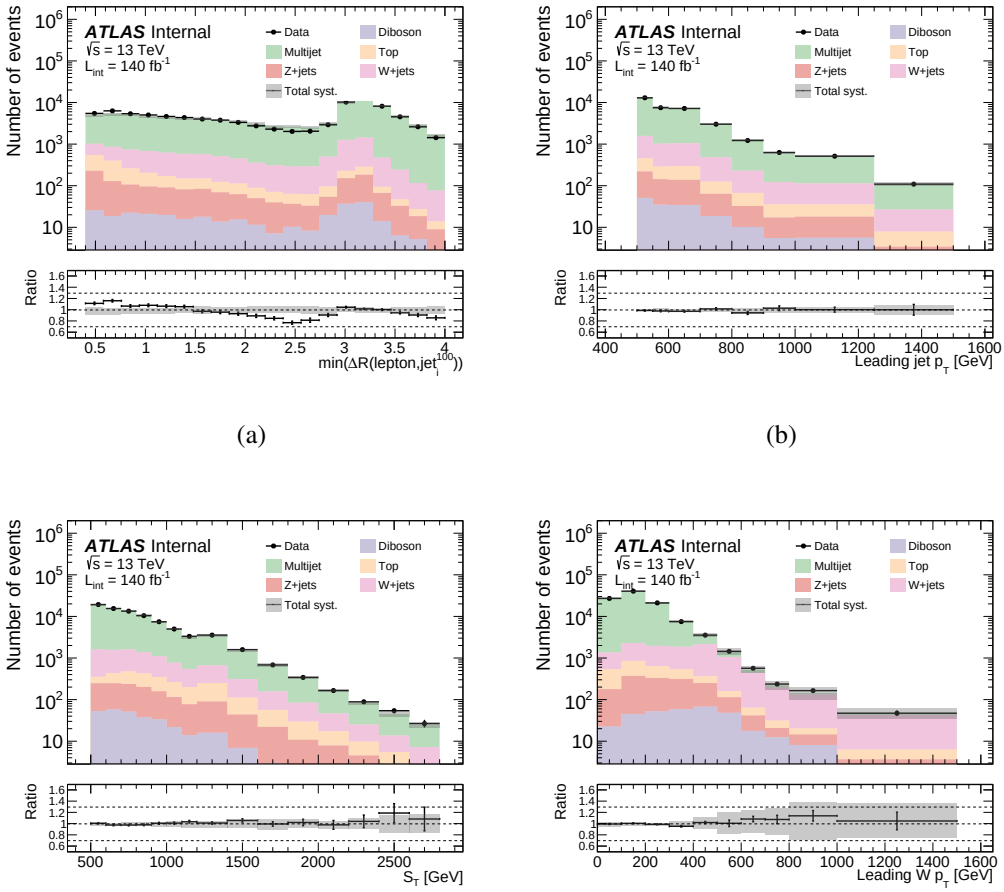


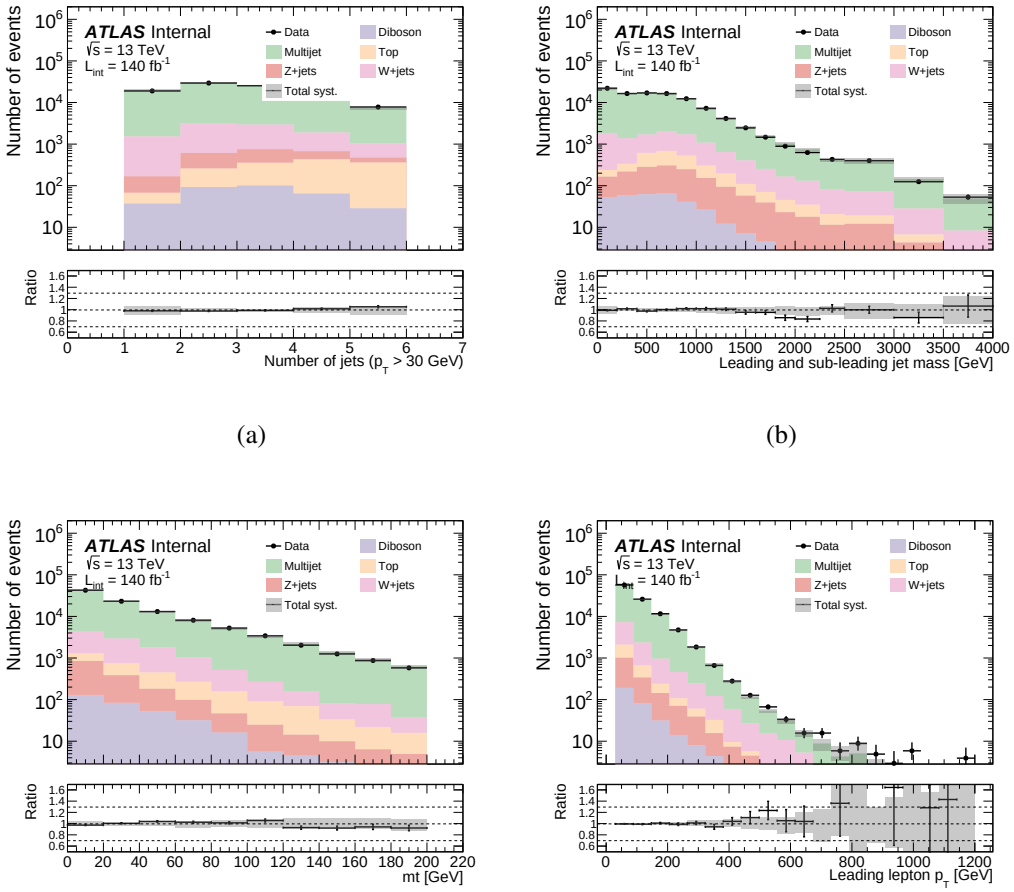
Figure 8.16: Reconstructed level comparisons of data and background predictions in the multi-jets control region for **electrons**. Normalization factors for multi-jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.



(a) (b)

(c) (d)

Figure 8.17: Reconstructed level comparisons of data and background predictions in the multi-jets control region for **muons**. Normalization factors for multi-jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.



(a) (b)

(c) (d)

Figure 8.18: Reconstructed level comparisons of data and background predictions in the multi-jets control region for **muons**. Normalization factors for multi-jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

8.5 Alternative multi-jets and fakes estimation: ABCD method

The multi-jets background and fakes, as described in Section 8.4.3, rely heavily on MC prediction, scaled by normalization factors derived from the multi-jets control region. Consequently, the estimation of such background can be significantly biased by potential MC mis-modeling, considering that multi-jets background is challenging to simulate accurately with parton shower models. Therefore, this section provides a description of an approach that aims to provide an alternative validation method for the multi-jets background, namely the ABCD method. This method requires only minimal input from MC simulations of the multi-jets and estimates this background in a fully data-driven manner.

However, it's important to note that this alternative method only serves as validation and is not used as the primary method to estimate the multi-jets background due to statistical limitations. Moreover, it requires extraordinary and meticulous tuning to mitigate any potential statistical fluctuations. As the comparison will show later, the multi-jets MC prediction corrected using normalization factors is in agreement with the ABCD method. Therefore, the analysis in subsequent chapters will rely on the scaled MC prediction for multi-jets.

8.5.1 Composition of fakes from Monte Carlo multi-jets prediction

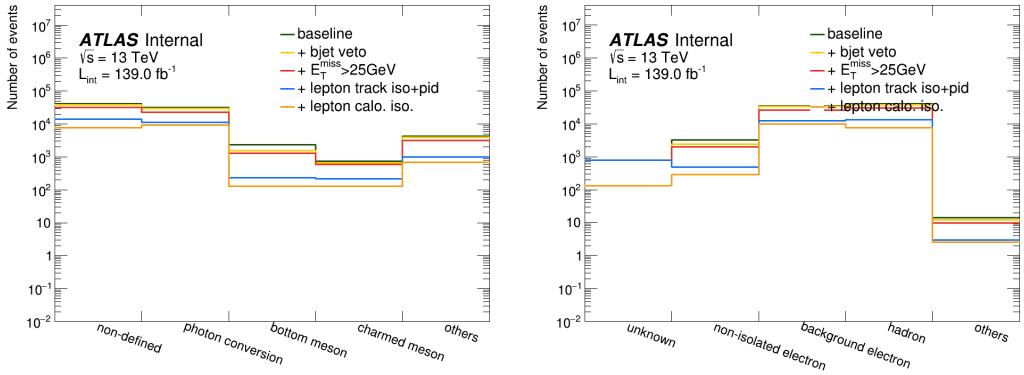
Fake backgrounds comprise processes resulting from fake leptons, like jets misidentified as electrons or non-isolated leptons from jet hadronization, such as muons from b-decays. A data-driven ABCD method can be used to estimate these backgrounds. Comparisons between the data-driven estimate and multi-jets MC event predictions, particularly in validation regions with significant fake background contributions, are crucial.

Before introducing the ABCD method, it's beneficial to illustrate the composition of fake backgrounds in the measurement region using multi-jets MC samples. The origin and types of these backgrounds at the particle-level are shown in Figures 8.19 and 8.20 for the electron and muon channels, respectively. The *MCTruthClassifier* determines origin and type classification codes, categorized into dominant fake categories: photon conversions, bottom mesons, charmed mesons, and others minor contributions. The composition is displayed at various selection stages using different colors, with each criterion applied above the previous one.

For electrons, the primary fake sources arise from the "undefined" and photon conversion categories. The "undefined" category denotes jets from secondary proton interactions[5]. Track isolation significantly reduces contributions from semi-leptonic heavy-flavor decays, while calorimeter isolation above track-based isolation notably reduces "unknown" category contributions. For muons, the combination of track-

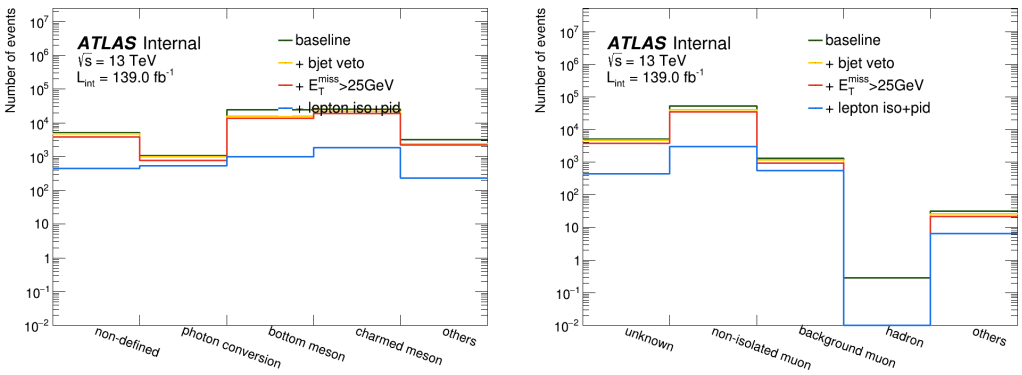
[5] Pile-up hits aren't saved in the truth record, leading to undefined matches.

based isolation and impact parameter cuts most effectively reduces fakes, as the main contribution consists of non-isolated muons from semi-leptonic heavy-flavor decays.



(a) (b)

Figure 8.19: MC truth origin (a) and type (b) for **electron** fakes. The others category includes all other category that are not shown in the plot.



(a) (b)

Figure 8.20: MC truth origin (a) and type (b) for **muon** fakes. The others category includes all other category that are not shown in the plot.

8.5.2 The ABCD method

The ABCD method relies on two linearly independent variables that are used to derive a transfer factor to estimate the multi-jet background. In this method, the signal lepton criteria and an $E_{\mathrm{T}}^{\mathrm{miss}}$ cut were the variables used to construct the ABCD planes.

To construct the ABCD planes, two sets of observables from the selection definition of the signal region, termed *region A*, are loosened to include control regions where fakes are enriched, as illustrated in Figure 8.21. Then, *regions B, C, and D* are defined as phase-spaces, which can be obtained by inverting one or both of the observable requirement from region A according to the pass or fail condition given by the two observable sets. Region B is populated by baseline lepton events that fail the tighter

signal criterion, while region C contains events with $E_{\mathrm{T}}^{\mathrm{miss}} < 30$GeV that otherwise pass all signal criteria. Finally, region D contains events with $E_{\mathrm{T}}^{\mathrm{miss}} < 30$GeV that fail the signal criterion but pass the baseline requirements. For the purpose of this study, we require the signal region A to satisfy $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV, and notes that the component of the signal region $E_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV is dominated by the multijet background.

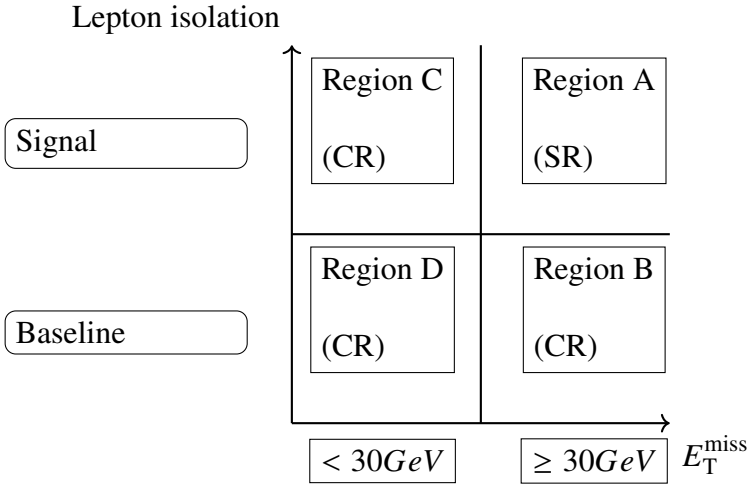


Figure 8.21: Schematic of ABCD phase-spaces for the estimation of the multi-jet background.

It is helpful if the observables used in constructing the ABCD plane are uncorrelated in the background processes. However, finding a perfect set of uncorrelated observables is challenging, and any residual correlation needs to be examined and corrected. One approach to correcting the correlation is to parameterize it along one of the ABCD observables, such as $E_{\mathrm{T}}^{\mathrm{miss}}$ in this case. Using MC multi-jets samples, the correlation correction can be obtained as following:

$$
\begin{aligned}
Corr(E_{\mathrm{T}}^{\mathrm{miss}}) &= \frac{N_{C\cup A}^{bkg}}{N_{D\cup B}^{bkg}}(E_{\mathrm{T}}^{\mathrm{miss}}), \\
[A, B, C, D] &= Corr(E_{\mathrm{T}}^{\mathrm{miss}}) \times \left[A', B', C', D'\right],
\end{aligned}
\tag{8.3}
$$

where $N_{C\cup A}^{bkg}(N_{D\cup B}^{bkg})$ represents the combined regions of C and A (D and B), and $Corr(E_{\mathrm{T}}^{\mathrm{miss}})$ is the correlation correction parameterized in terms of $E_{\mathrm{T}}^{\mathrm{miss}}$. The events from the original ABCD plane (denoted with prime) with residual correlation can then be corrected with this correction factor, ensuring the following condition holds for the corrected ABCD plane:

$$
\frac{N_C^{bkg}(\vec{x})}{N_D^{bkg}(\vec{x})} = \frac{N_A^{bkg}(\vec{x})}{N_B^{bkg}(\vec{x})},
\tag{8.4}
$$

where $\vec{x}$ is set of observables that give the bin-to-bin parameterization. In the case for which $\vec{x}$ has just a single bin, this equation will just be the ratio of event total yield in each of the ABCD regions. The number of background events in the signal region can be obtained with rearrangement of Equation (8.4):

$$
\begin{aligned}
N_A(\vec{x}) &= N_B(\vec{x}) \times TF(\vec{x}), \\
TF(\vec{x}) &= \frac{N_C^{bkg}(\vec{x})}{N_D^{bkg}(\vec{x})},
\end{aligned}
\tag{8.5}
$$

where $TF(\vec{x})$ is called the transfer factor. The transfer factor is parametrized bin-by-bin in a given observable. Since the non-isolated lepton and fake lepton that is reconstructed from a jet are the main sources of non-prompt background in this study,

the parameterization of transfer factor is done in the two dimensions of lepton transverse momentum p_{T} and its pseudo-rapidity η.

8.5.3 Comparison of ABCD method and multi-jets MC prediction

The ABCD method was compared with the default method (MC-based multi-jets with data normalization corrections) in both the signal regions and in an electron validation region that will be introduced below in Section 8.5.4.

As shown in Figure 8.22, the two results are consistent within uncertainties across most distributions in the electron validation region (defined in the next section). However, discrepancies are observed in the m_{jj} distribution, as illustrated in Figure 8.23. Additionally, noticeable differences are observed between the two predictions for some observables in the inclusive and collinear regions, as shown in Figures 8.24 to 8.26. These differences are not a concern, as the good agreement with data for multi-jets will be demonstrated in the next section. Given the good agreement demonstrated with the default method shown in Section 8.5.4, we conclude that the default method provides a more reliable estimate of the multi-jets background compared to the statistically limited ABCD method.

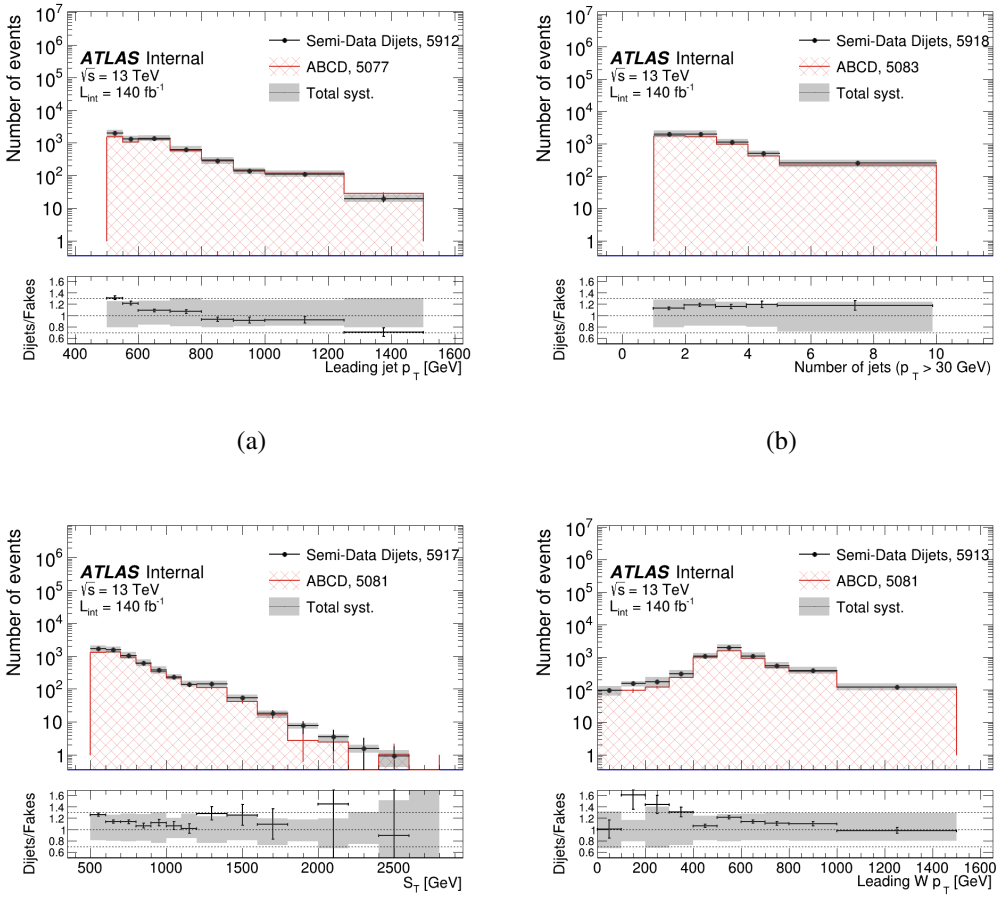


Figure 8.22: Comparison of background estimated by ABCD-based and MC scaling approaches in the validation region of **electrons**. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven di-jet background estimate. The uncertainties on the ABCD method are statistical only.

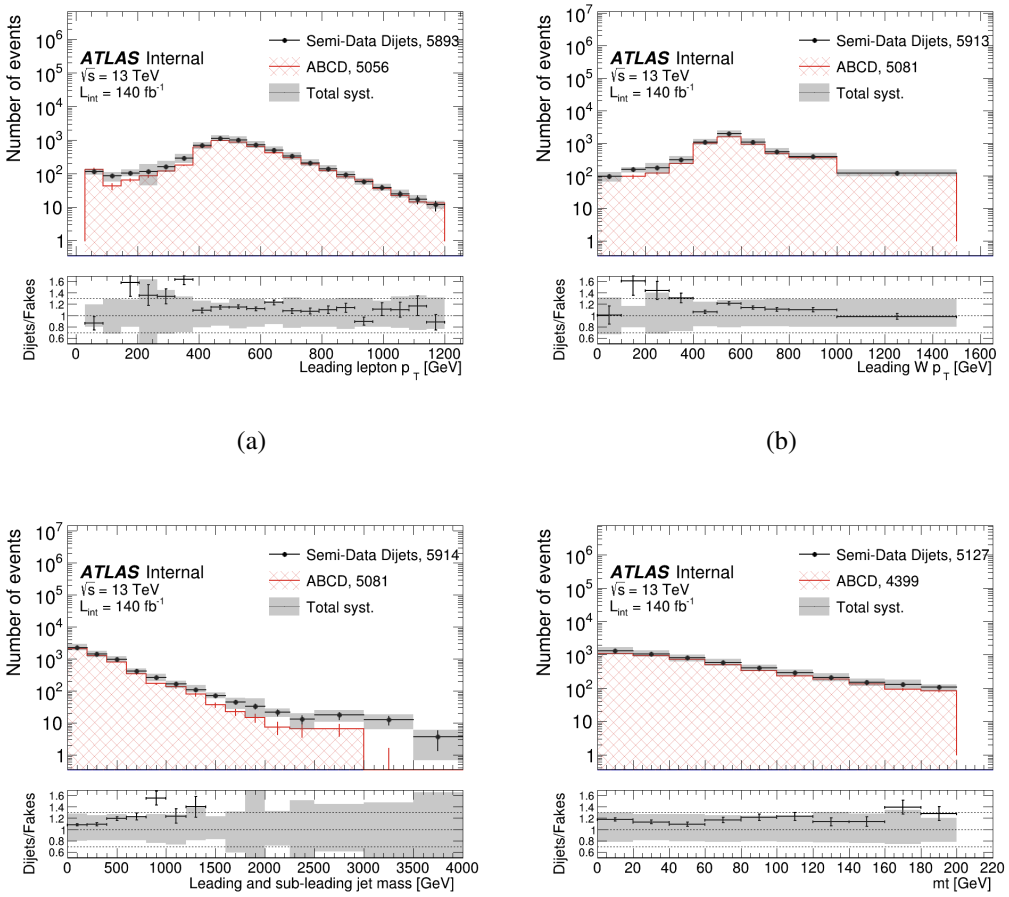
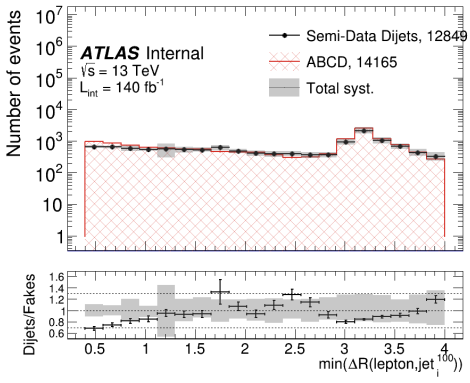
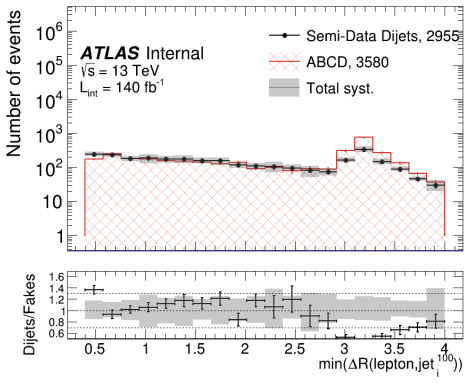


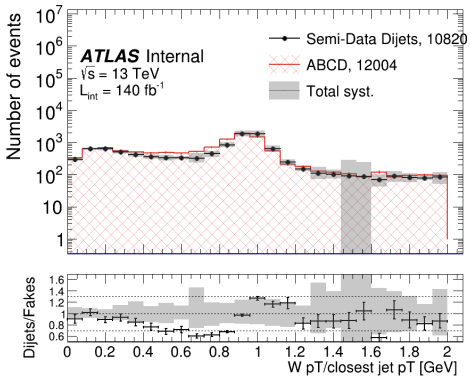
Figure 8.23: Comparison of background estimated by ABCD-based and MC scaling approaches in the validation region of **electrons**. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven di-jet background estimate. The uncertainties on the ABCD method are statistical only.



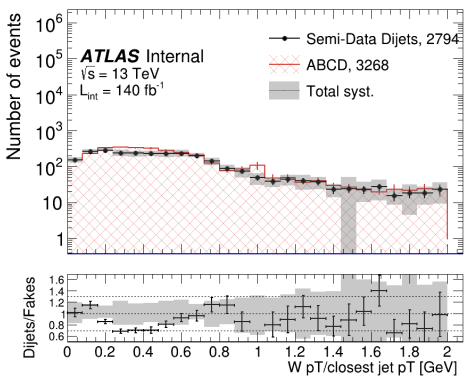
(a) electron



(b) muon



(c) electron



(d) muon

Figure 8.24: Comparison of background estimated by ABCD-based and MC scaling approaches in inclusive region. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven multi-jet background estimate. The uncertainties on the ABCD method are statistical only. Some of the bins have extremely large uncertainties is due to the statistical limitation and fluctuation in some of the MC samples used for calculating the uncertainties.

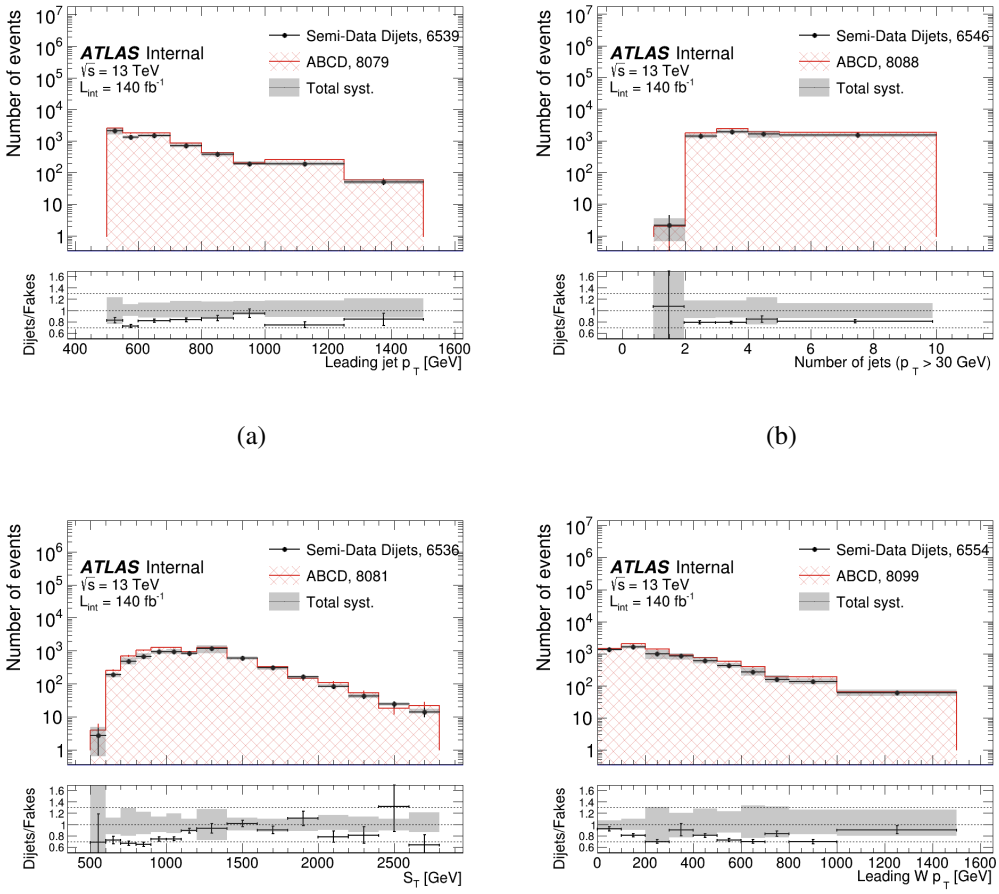


Figure 8.25: Comparison of background estimated by ABCD-based and MC scaling approaches in the collinear region of **electrons**. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven multi-jet background estimate. The uncertainties on the ABCD method are statistical only.

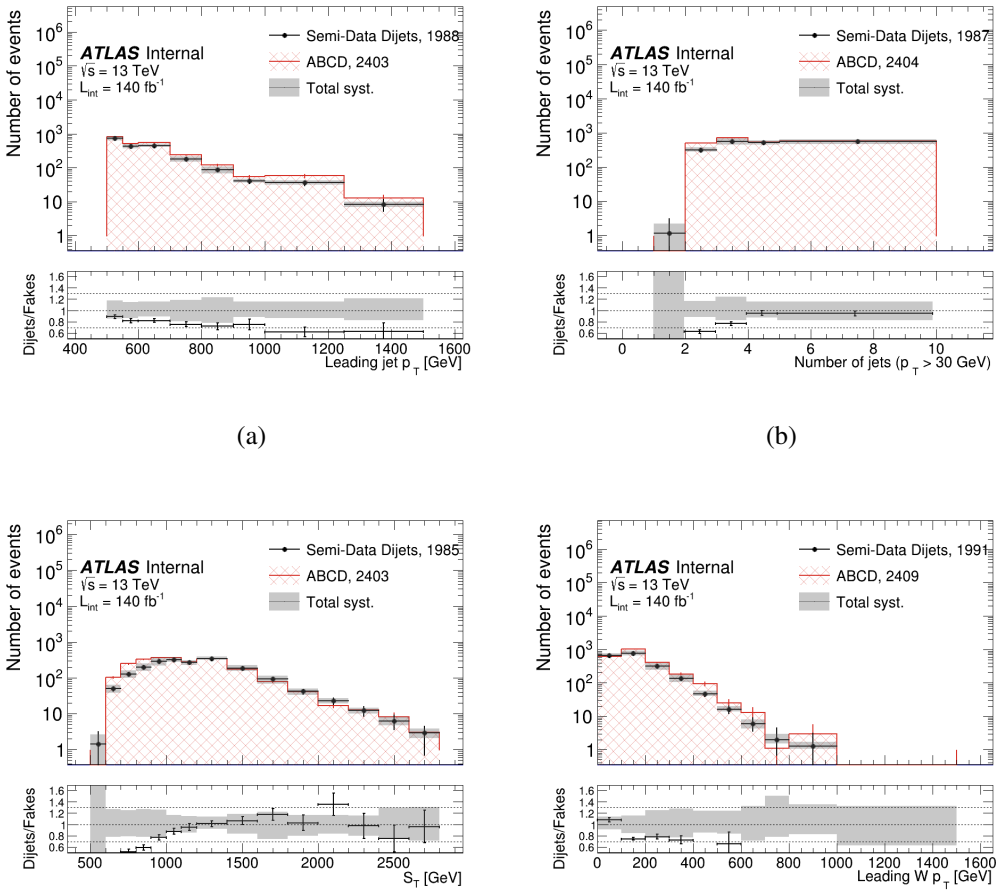


Figure 8.26: Comparison of background estimated by ABCD-based and MC scaling approaches in the collinear region of **muons**. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven multi-jet background estimate. The uncertainties on the ABCD method are statistical only.

8.5.4 Low $E_{\mathrm{T}}^{\mathrm{miss}}$ multi-jet validation region

The multi-jet contribution in the signal region is dominant in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ and back-to-back phase-spaces. By explicitly selecting this region, we can enhance the multi-jet contribution and perform a meaningful validation of the multi-jets yield and shapes in various observables.

The low-$E_{\mathrm{T}}^{\mathrm{miss}}$ validation region (VR) in the electron channel is defined with the same selection criteria as the signal region (see Section 7.4), but in the back-to-back phase-space ($\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100}) > 2.6$) with $E_{\mathrm{T}}^{\mathrm{miss}} < 100\,\mathrm{GeV}$. However, in forming the validation region, we also require that $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV, since the region below 30 GeV has already been explored in the context of the ABCD method. In this region, the multi-jet process is roughly 50% of the predicted yield; the contamination from the signal W+jets process is roughly 38% of the predicted yield.

The inclusive yields for multi-jets, as estimated with the MC based approach and the data-driven ABCD procedure, and other processes in the VR for electron channel are listed in Table 8.5. Both the semi-data driven and ABCD prediction provide a good agreement with the data within the uncertainties, noting that systematic uncertainties were not considered for the ABCD method in these comparisons.

For the muon channel, a similar validation region to that of the electron channel is defined, but with an additional selection criterion on the $p_{\mathrm{T}}^{W}/p_{T}^{\text{closest jet}}$ ratio, set within the range $[0.3, 0.75]$. As illustrated in Figure 8.20 (a), the primary sources of fake signals in the muon channels originate from b- and c-mesons. The decay of b- and

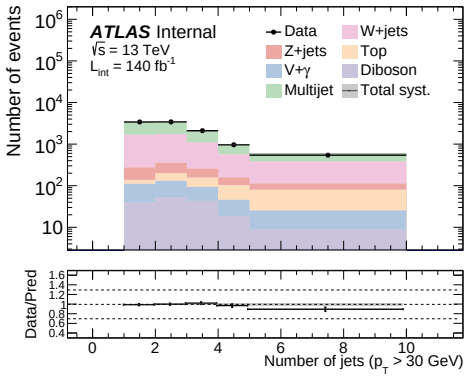
c-quarks typically produces a W boson and a lighter quark, with these decay products each carrying approximately half of the momentum of their parent quark. Therefore, this selection criterion ensures a region enriched with fakes for the muon channel.

The differential distributions in the validation region for electron channel are shown in Figure 8.27 to Figure 8.30. The statistical and systematic uncertainties, which will be discussed in Chapter 11, have been summed in quadrature into the grey error band. For both the MC based method and the ABCD procedure, the background predictions agree with the data within the combined statistical and systematic uncertainties. The exception is the m_{jj} distribution in Figure 8.30. However, the miss-modeled bins at extreme m_{jj} have significant contributions from the signal W+jets process, which contributes at the 60-75% of level of the total background model in the bins that have $m_{jj} > 2000$ TeV. Since the W+jets is observed to be significantly mis-modeled (e.g. see Figure 9.6) in these bins and significantly contributes to this distribution, we attribute this mis-modeling to W+jets and not the multi-jet model.

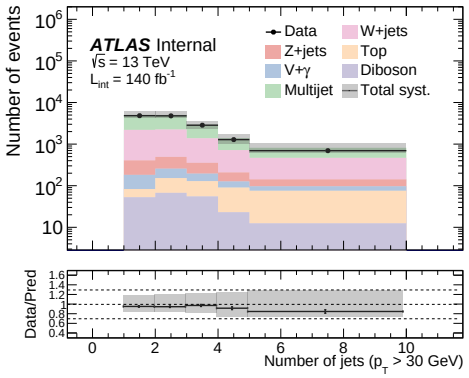
Similarly, the differential distributions in the validation region for the muon channel are presented in Figures 8.31 and 8.32.

Table 8.5: Yields in the Low-$E_{\mathrm{T}}^{\mathrm{miss}}$ VR for the MC based multi-jet and data-driven ABCD predictions. The summed backgrounds include all backgrounds in the first component of the table, but only the multi-jet background grouped in a common row as the summed backgrounds. Uncertainties on the MC backgrounds and the Multi-jet (MC based) include statistical and systematic uncertainties added in quadrature, while the Multi-jet (Data-driven) ABCD estimate contains statistical uncertainties only.

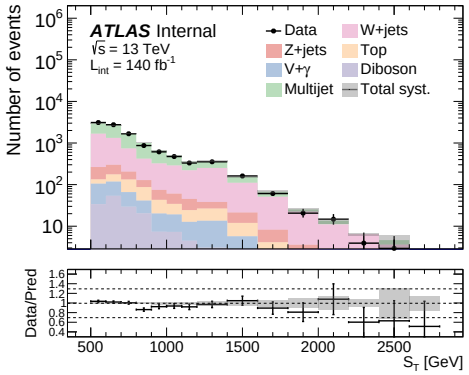
Process	Low-$E_{\mathrm{T}}^{\mathrm{miss}}$ VR
$t\bar{t}$	207.45 ± 62.88[0.30]
Diboson	165.22 ± 39.10[0.24]
Single top	66.73 ± 32.22[0.48]
Z+jets	467.31 ± 120.54[0.26]
W+2j	190.94 ± 44.75[0.23]
W+jets	4109.16 ± 834.52[0.20]
W+jets (tau)	57.21 ± 40.73[0.71]
V+γ	249.45 ± 73.26[0.29]
Multi-jet (MC based)	5788.72 ± 1916.17[0.33] (statistical and systematics)
Total background	11302.19 ± 1923.87[0.17]
Multi-jet (ABCD)	5082.01 ± 48.81 (statistical only)
Total background	10612.85 ± 62.47
Observed data	10569.00 ± 102.81



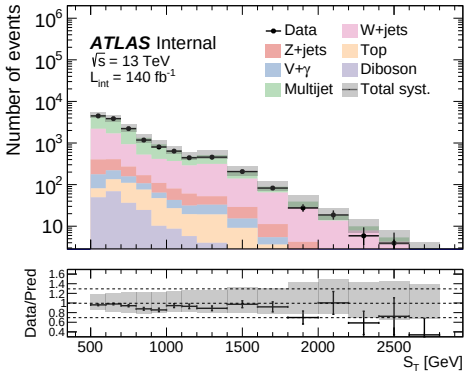
(a) ABCD



(b) Fake MC

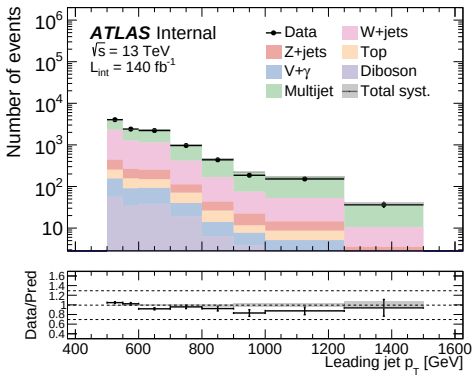


(c) ABCD

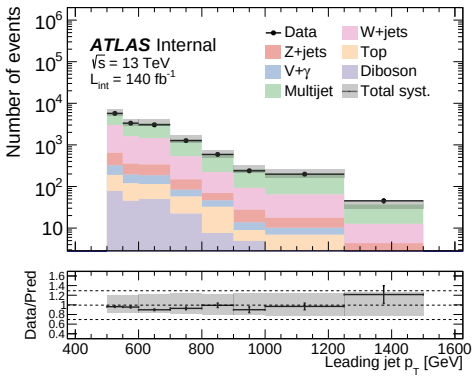


(d) Fake MC

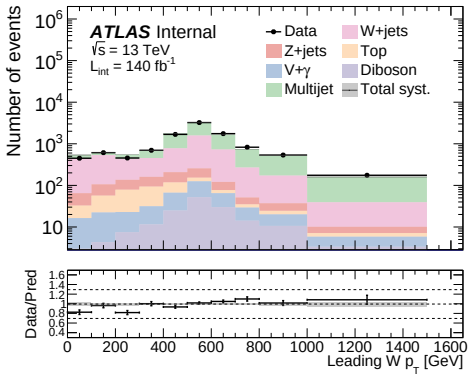
Figure 8.27: Differential distributions in the validation region for electron channel, comparing the MC based estimate and the data-driven ABCD prediction. The grey error band include statistical and systematic uncertainties added in quadrature. The ABCD method only contains statistical uncertainties on its estimate.



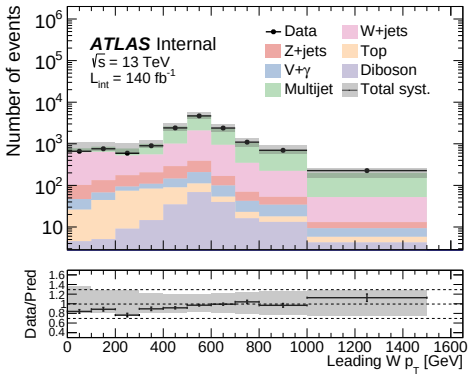
(a) ABCD



(b) Fake MC

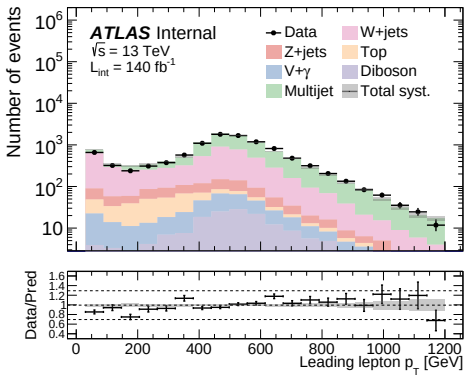


(c) ABCD

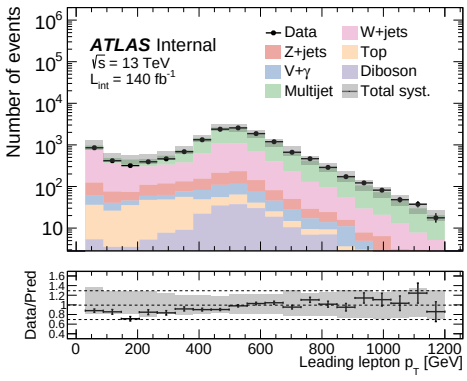


(d) Fake MC

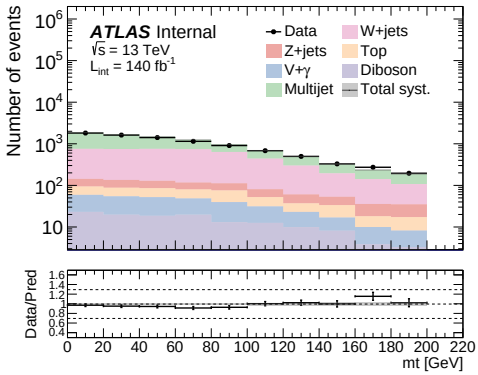
Figure 8.28: Differential distributions in the validation region for electron channel, comparing the MC based estimate and the data-driven ABCD prediction. The grey error band include statistical and systematic uncertainties added in quadrature. The ABCD method only contains statistical uncertainties on its estimate.



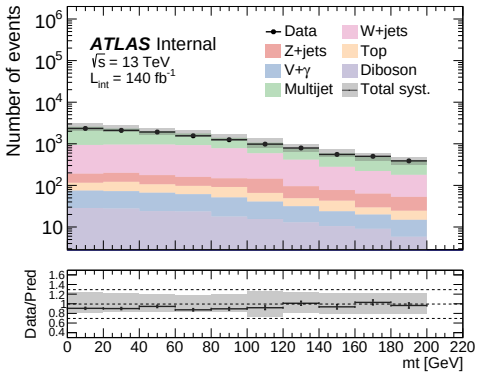
(a) ABCD



(b) Fake MC

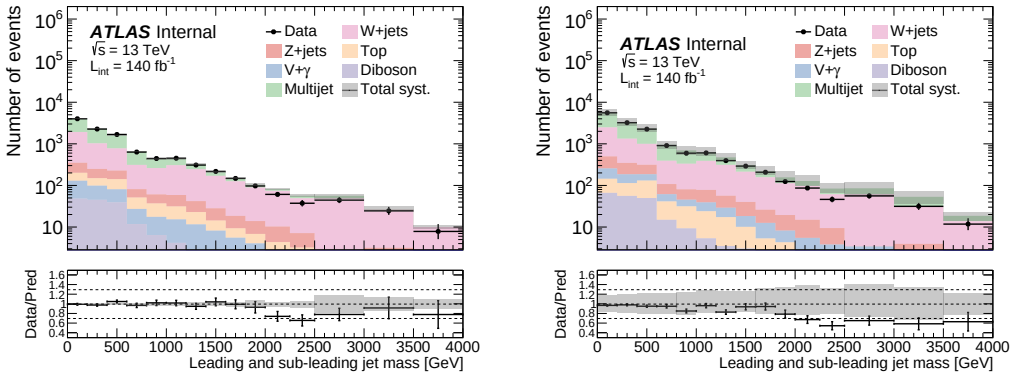


(c) ABCD



(d) Fake MC

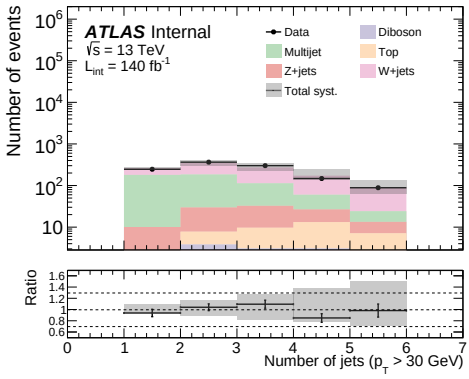
Figure 8.29: Differential distributions in the validation region for electron channel, comparing the MC based estimate and the data-driven ABCD prediction. The grey error band include statistical and systematic uncertainties added in quadrature. The ABCD method only contains statistical uncertainties on its estimate.



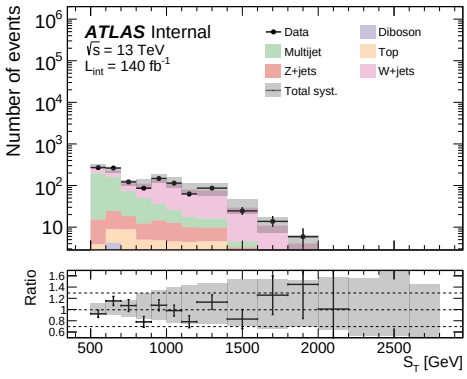
(a) ABCD

(b) Fake MC

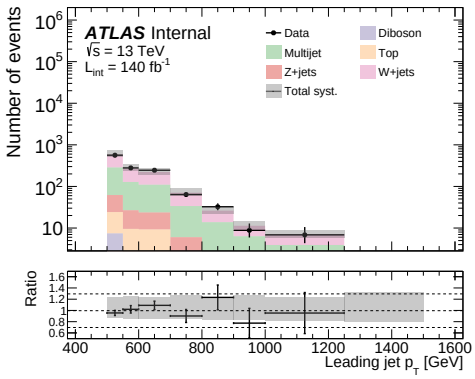
Figure 8.30: Differential distributions in the validation region for electron channel, comparing the MC based estimate and the data-driven ABCD prediction. The grey error band include statistical and systematic uncertainties added in quadrature. The ABCD method only contains statistical uncertainties on its estimate.



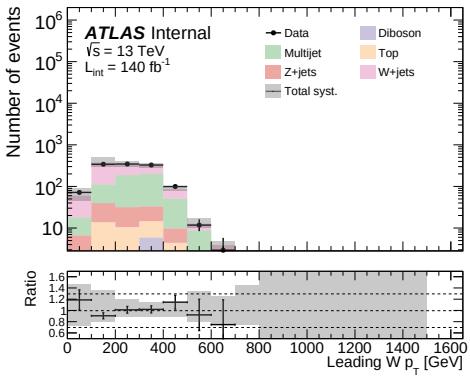
(a) jet multiplicity



(b) S_t

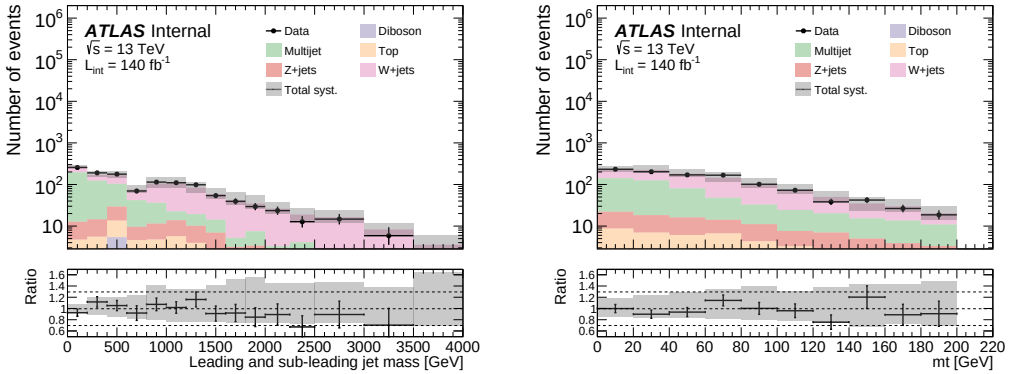


(c) leading jet p_{T}



(d) W p_{T}

Figure 8.31: Differential distributions in the validation region for muon channel. The grey error band include statistical and systematic uncertainties added in quadrature.



(a) sum of leading and sub-leading jet masses (b) M_t

Figure 8.32: Differential distributions in the validation region for muon channel. The grey error band include statistical and systematic uncertainties added in quadrature.

Chapter 9

Signal Region Model and Comparison to Data

This chapter aims to provide a comprehensive model of the content of the signal region at the reconstruction level and to compare it with observed data. Since the data collected in the experiments contains all possible physical processes, accurately estimating the background yield is crucial for extracting the number of signal events. The background compositions in the data are estimated using MC samples with normalization factors applied, as discussed in Chapter 8. In contrast, the $W+$jets signal is taken directly from MC samples without any scaling. Subsequently, these backgrounds will be subtracted from the data prior to the unfolding procedure, which will estimate the signal yield at the particle level. Details of the unfolding procedure will be discussed later inin Chapter 10.

9.1 Expected signal contributions

The estimated signal contribution in the inclusive, inclusive 2-jet, and collinear phase-spaces is shown in Tables 9.1 and 9.2 for the electron and muon channels, respectively. The sum of the signal and background yields agrees well with the observed data and falls well within the 1σ uncertainty range. The uncertainty in the tables represents the quadrature sum of the statistical and symmetrized systematic uncertainties, and the systematic uncertainties are computed after applying the offset method, for which the details will be discussed later in Chapter 11,

The expected signal region yields estimated without applying the scaling factors from the CR are shown in Tables 9.3 and 9.4. These yields include only statistical

uncertainties. The significant disagreement between the data and prediction emphasize the importance of the background normalization factors.

The total backgrounds are estimated to contribute approximately 31% (41%) of the total yields in the inclusive region of the muon (electron) channels. The dominant background in the muon channel is the $t\bar{t}$ process, while the electron channel has significant contribution from the multi-jet background. The uncertainties on the $t\bar{t}$ ($Z+$jets) background are around the 24% (10%) level for both electron and muon channels, and nearly independent of the region. In the electron channel, the multi-jet background uncertainty is around 30% in all regions, and arises primarily from the uncertainty on the normalization factor dependency in the control regions. Finally, the single top uncertainty is nearly 50% and arises from the diagram-reduction (DR) vs diagram-subtraction (DS) scheme comparison for the treatment of the tW quantum interference with $t\bar{t}$ [86]. Despite this being a large uncertainty, the overall contribution of the single top is very small ($< 2\%$).

Process	Inclusive	Collinear	Inclusive 2-jets
$t\bar{t}$	5530 ± 2101	4845 ± 1891	5492 ± 2086
Diboson	2881 ± 159	1929 ± 107	2593 ± 143
Single top	1742 ± 511	1369 ± 409	1693 ± 495
Z+jets	4328 ± 1445	3585 ± 1347	4149 ± 1465
EW V+jets	3572 ± 620	2636 ± 457	3233 ± 565
$\tau\nu$+jets	5096 ± 629	3321 ± 382	4484 ± 565
Multi-jets	3637 ± 368	2461 ± 220	3251 ± 302
W+jets (Sh 2.2.11)	72016 ± 21722	52024 ± 19654	64652 ± 22107
Total Signal+Bkg	98802 ± 21904	72170 ± 19808	89548 ± 22280
Observed data	94229 ± 307	67969 ± 261	85128 ± 292

Table 9.1: Yield of each process in various **muon** measurement regions **after** applying the background normalization factors from the CR. The values in brackets indicate the relative uncertainties on each of the respective backgrounds. Uncertainties include statistical and systematic errors added in quadrature, and systematic errors will be discussed later in Chapter 11.

Process	Inclusive	Collinear	Inclusive 2-jets
$t\bar{t}$	7229 ± 2653	5808 ± 2254	7073 ± 2629
Diboson	3443 ± 189	2224 ± 123	3073 ± 169
Single top	2149 ± 614	1608 ± 471	2056 ± 586
Z+jets	6352 ± 1921	5110 ± 1744	5990 ± 1911
EW V+jets	4182 ± 726	3029 ± 527	3746 ± 656
$\tau\nu$+jets	6494 ± 856	4034 ± 548	5619 ± 746
Multi-jets	14628 ± 2414	6344 ± 909	11598 ± 1767
W+jets (Sh 2.2.11)	77339 ± 22342	53989 ± 19899	68367 ± 22568
Total Signal+Bkg	121816 ± 22756	82146 ± 20148	107522 ± 22906
Observed data	119880 ± 346	79887 ± 283	105740 ± 325

Table 9.2: Yield of each process in various **electron** measurement regions **after** applying the background normalization factors from the CR. The values in brackets indicate the relative uncertainties on each of the respective backgrounds. Uncertainties include statistical and systematic errors added in quadrature, and the systematic errors will be discussed later in Chapter 11.

Process	Inclusive	Collinear	Inclusive 2-jets
$t\bar{t}$	6112.59 ± 66.12	5320.54 ± 54.40	6073.58 ± 65.32
Diboson	2877.95 ± 26.62	1926.75 ± 21.86	2590.95 ± 25.33
$\tau\nu$+jets	5159.52 ± 37.52	3355.58 ± 29.61	4536.50 ± 34.91
Single top	1736.87 ± 15.09	1364.46 ± 13.27	1688.23 ± 14.86
Z+jets	4092.25 ± 33.61	3420.40 ± 18.13	3928.05 ± 33.52
EW V+jets	3583.05 ± 12.44	2644.12 ± 10.67	3242.59 ± 11.83
V+γ+jets	3788.84 ± 17.63	3420.48 ± 17.57	2609.85 ± 16.92
Multi-jets	5348.47 ± 64.96	3607.61 ± 55.76	4783.69 ± 62.85
W+jets (Sh 2.2.11)	72309.79 ± 219.41	52232.95 ± 187.98	64906.64 ± 217.35
Total Signal+Bgk	101220.48 ± 245.68	73872.41 ± 208.28	91750.23 ± 242.49
Observed data	93876.00 ± 306.39	67704.00 ± 260.20	84795.00 ± 291.20

Table 9.3: Yield of each process in **muon** measurement regions **without** applying background normalization factors. The uncertainties are statistical only.

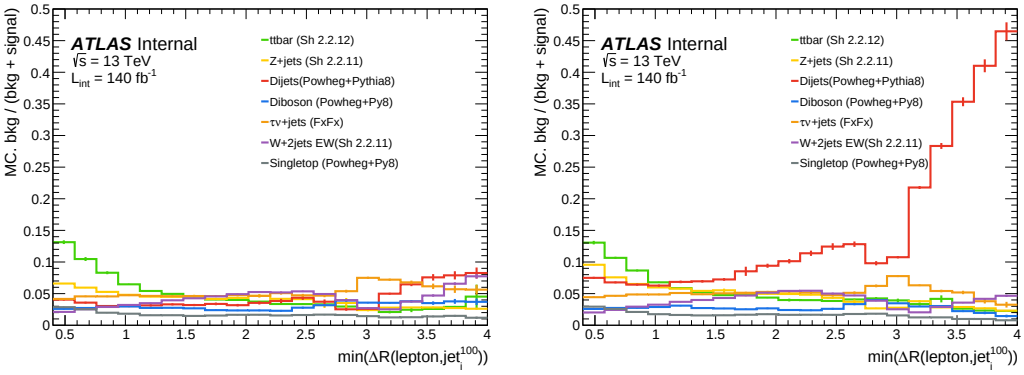
Process	Inclusive	Collinear	Inclusive 2-jets
$t\bar{t}$	8480.92 ± 77.11	6726.70 ± 46.48	8289.48 ± 70.17
Diboson	3429.93 ± 29.83	2214.49 ± 23.99	3061.01 ± 28.18
$\tau\nu$+jets	6862.21 ± 63.04	4290.36 ± 56.27	5949.24 ± 60.59
Single top	2139.20 ± 17.36	1598.84 ± 14.73	2045.23 ± 16.80
Z+jets	6452.00 ± 32.92	5255.23 ± 22.86	6099.52 ± 25.91
EW V+jets	4279.47 ± 14.02	3102.37 ± 11.92	3832.42 ± 13.25
V+γ+jets	4158.65 ± 15.96	2793.43 ± 14.38	3722.15 ± 15.09
Multi-jets	26832.15 ± 192.48	13259.80 ± 172.99	22059.05 ± 187.04
W+jets (Sh 2.2.11)	79421.47 ± 187.72	55531.21 ± 169.40	70251.88 ± 183.45
Total Signal+Bkg	137897.34 ± 291.00	91979.02 ± 255.74	121587.84 ± 281.35
Observed data	118829.00 ± 344.72	79151.00 ± 281.34	104739.00 ± 323.63

Table 9.4: Yield of each process in **electron** measurement regions **without** applying background normalization factors. The uncertainties are statistical only.

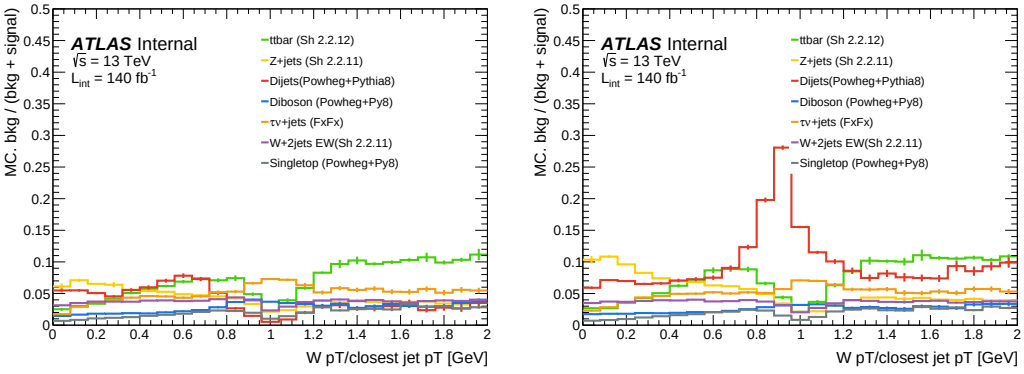
The composition of the background processes contributing to inclusive signal regions are shown in Figures 9.1 and 9.2 for the inclusive and inclusive 2-jet observables. Similar to the inclusive numbers above, these fractions are computed with respect to the total background + signal MC yields *after* applying the normalization factors described previously in Chapter 8. The electron channel is shown on the left and the muon channel on the right. In the electron channel, the multi-jet background dominates the back-to-back

phase-space and reaches nearly 45% for large angular separation. For both the muon and electron channel, the $t\bar{t}$ process contributes at the 5% level in the back-to-back region and increases up to 15% for the most collinear regions. For the inclusive 2-jet m_{jj} distribution, in the higher bins the electroweak V+jj process contributes at the 10-12% level in the highest m_{jj} bins, while for the electron channel the multi-jet background contributes at a constant 10% level above 1 TeV.

The background composition in the collinear signal region is shown in Figures 9.3 and 9.4 for electron and muons, respectively. For both the electron and muon channels, all backgrounds contribute less than the 20% level in any given bin. Notably, $t\bar{t}$ (diboson) contributes in the highest (lowest) jet multiplicity bin at the 15% level. In the electron channel, the multi-jet background increases as the W-boson momentum increases and reaches the 15% level in the last bin.

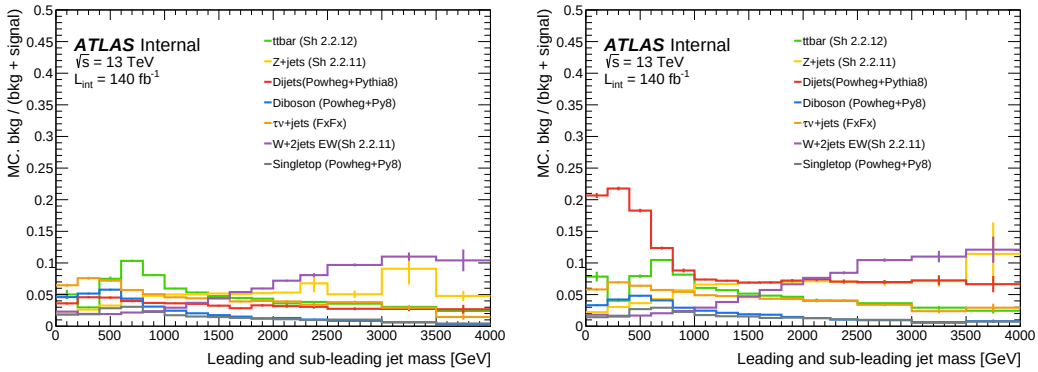


(a) Muon (b) Electron



(c) Muon (d) Electron

Figure 9.1: Differential distributions showing the fraction of background events in the **inclusive** region for the muon (a,c) and electron (b,d) channels. The fractions are defined with respect to the sum of all MC (backgrounds plus signal W+jets). The fractions are extracted *after* the data-driven normalization factors are applied.



(a) Muon (b) Electron

Figure 9.2: Differential distributions showing the fraction of background events in the **inclusive 2-jet** region for the muon (a) and electron (b) channels. The fractions are defined with respect to the sum of all MC (backgrounds plus signal W+jets). The fractions are extracted *after* the data-driven normalization factors are applied.

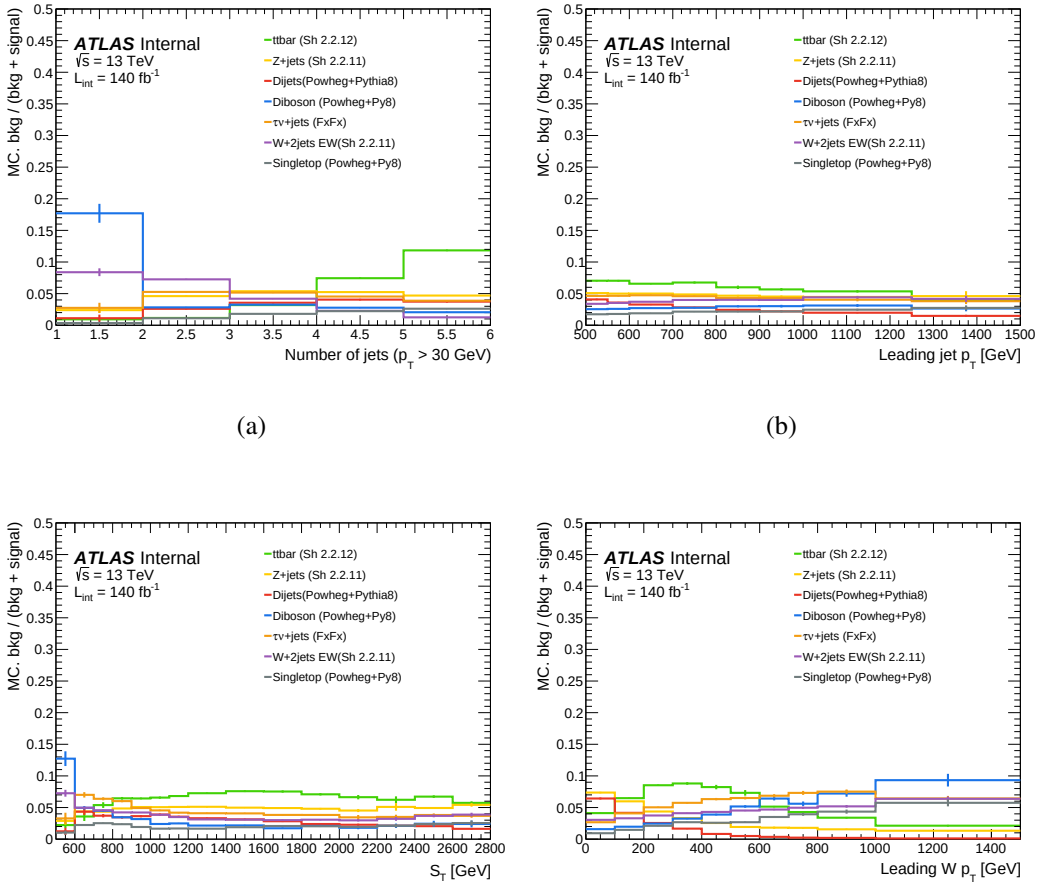
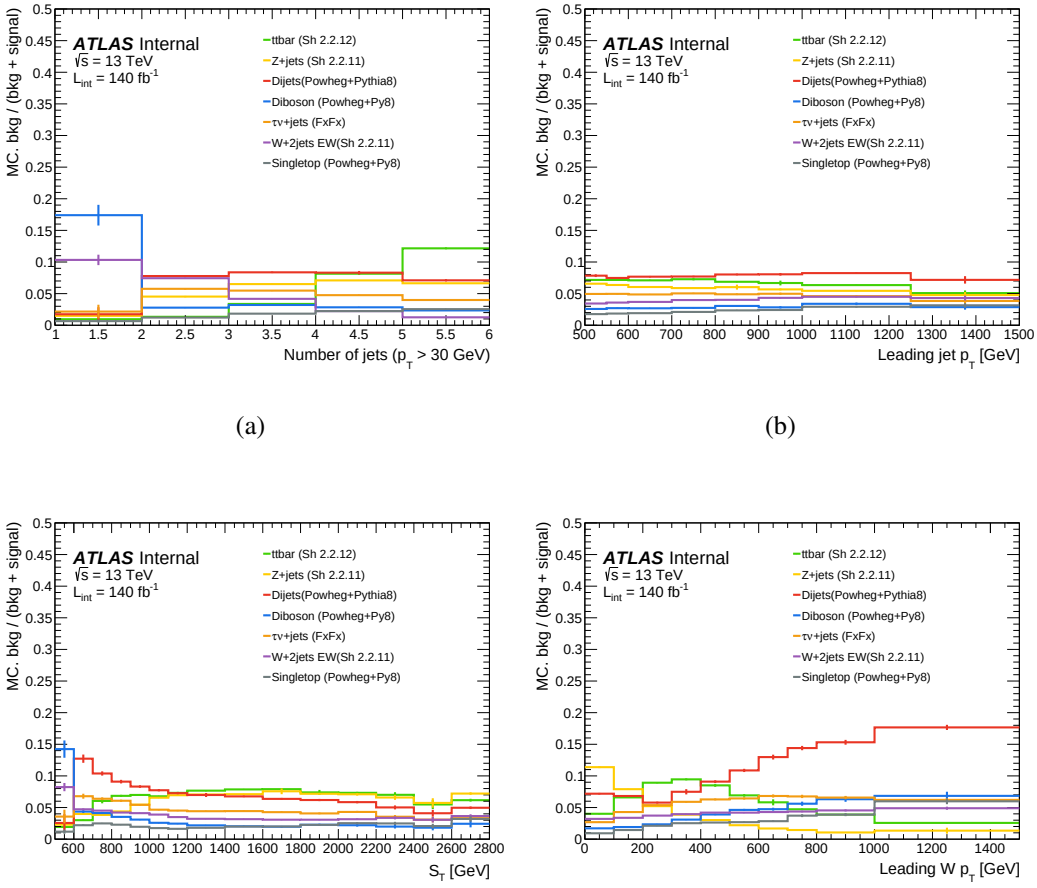


Figure 9.3: Differential distributions showing the fraction of background events in the **collinear** region for the **muon** channel. The fractions are defined with respect to the sum of all MC (backgrounds plus signal W+jets). The fractions are extracted *after* the data-driven normalization factors are applied.



(a) (b)

(c) (d)

Figure 9.4: Differential distributions showing the fraction of background events in the **collinear** region a for the **electron** channel. The fractions are defined with respect to the sum of all MC (backgrounds plus signal W+jets). The fractions are extracted *after* the data-driven normalization factors are applied.

9.2 Comparison with data

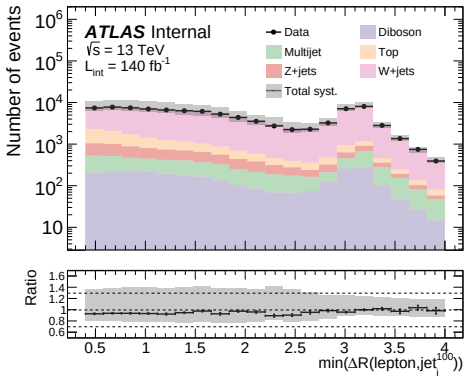
In this section, we compare the signal plus background predictions in the signal inclusive and collinear phase-spaces at the reconstruction-level. The $t\bar{t}$, Z+jets and dijet backgrounds include the n-jet dependent normalization scale factors discussed throughout Chapter 8. All distributions include statistical and systematic uncertainties. The experimental and theory systematics uncertainties will be discussed in further detail in Chapter 11.

In the inclusive phase-space, the only distributions that will be formed for measurement are the angular separation between the lepton and closest jet ($\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$) and the p_{T} ratio between W boson to its closest jet. The differential distributions for electrons and muons are shown in Figure 9.5. The invariant mass distribution of the leading and sub-leading (m_{jj}) will only be measured in the inclusive region with at least 2 jets, and is shown in Fig. 9.6. The sum of backgrounds and W+jets signal prediction provided by SHERPA 2.2.11 models the data well across the full phase-space, as already shown in the total yield in Tables 9.1 and 9.2.

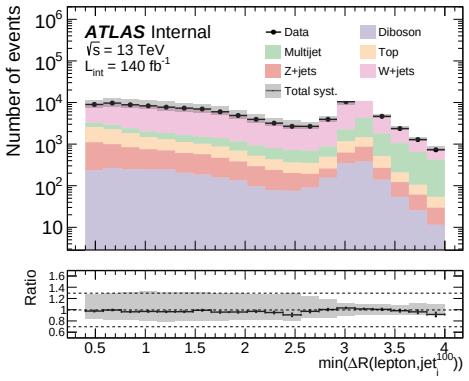
Additional kinematic distributions in the inclusive phase-space are shown in Figures 9.9 and 9.10, although these variables will not be unfolded. In the collinear phase-space, several other important kinematic observables are measured and will be unfolded to particle-level. These are shown in Figures 9.7 and 9.8.

The SHERPA 2.2.11 prediction provides a good description of the data for all kinematic variables, except the most extreme m_{jj} regions, for which above approximately 1.5

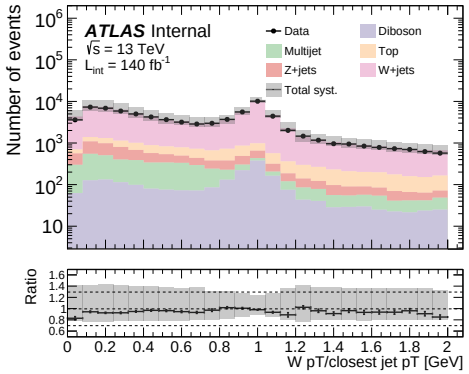
TeV signal is over-predicted consistently by roughly 1σ variation in scale uncertainty as shown in Figure 9.6. While the modeling has been improved in the SHERPA 2.2.11 compared to previous legacy versions, this variable continues to be mis-modeled by state-of-the-art merged MC setups. For the benefit of searches that rely on this variable, this analysis will measure the m_{jj} variable to provide input for MC generator experts to work towards improving the modeling of this challenging variable.



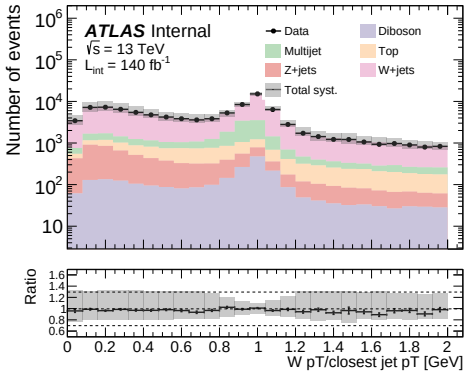
(a) Muon channel



(b) Electron channel

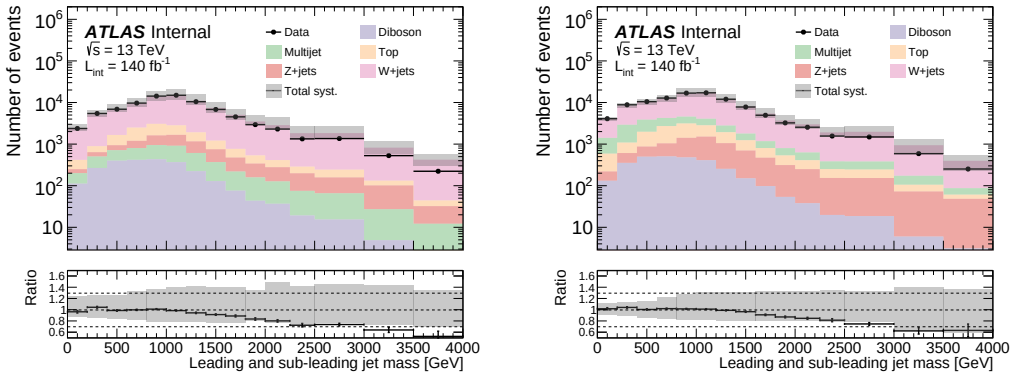


(c) Muon channel



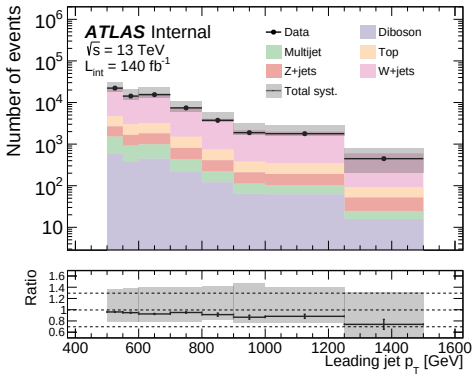
(d) Electron channel

Figure 9.5: $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$ distribution at reconstruction-level in the **inclusive** selection for the muon channel (a,c) and electron channel (b,d). Leading jet threshold of 500 GeV is shown. The grey band includes include statistical and systematic uncertainties added in quadrature.

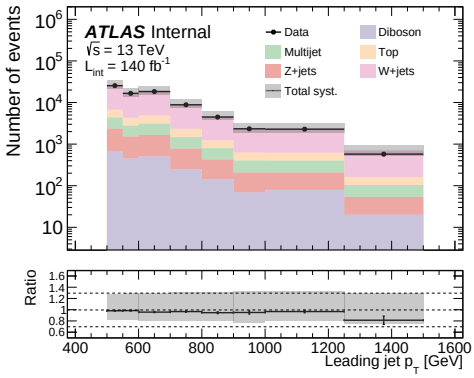


(a) Muon (b) Electron

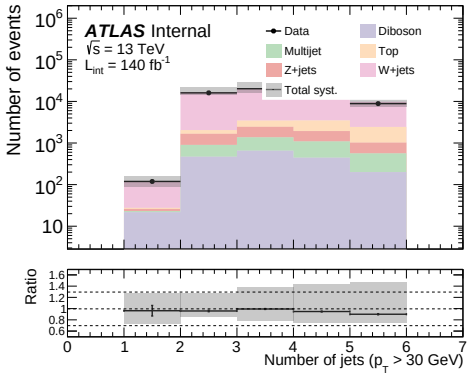
Figure 9.6: m_{jj} distribution at reconstruction-level in the **inclusive 2-jet** selection for the muon channel (a) and electron channel (b). Leading jet threshold of 500 GeV is shown. The grey band includes include statistical and systematic uncertainties added in quadrature.



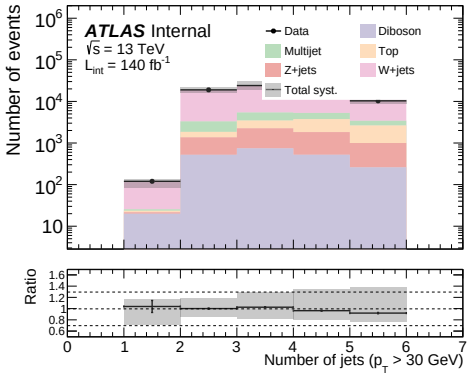
(a) Muon



(b) Electron

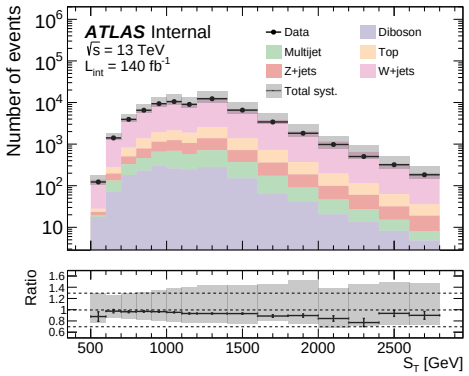


(c) Muon

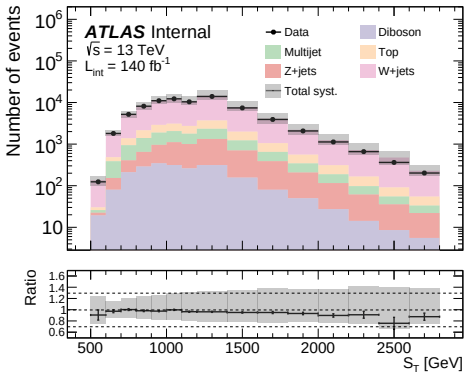


(d) Electron

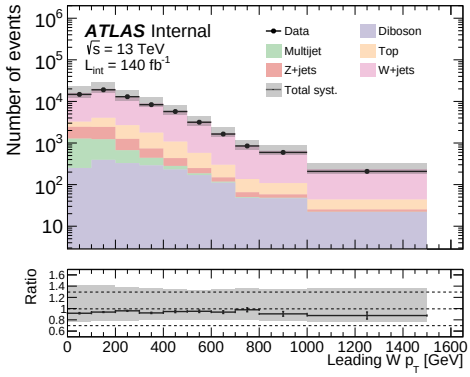
Figure 9.7: Differential distributions at reconstruction-level in the **collinear** phase-space. (left:electron channels, right:muon channel) The grey band includes statistical and systematic uncertainties added in quadrature.



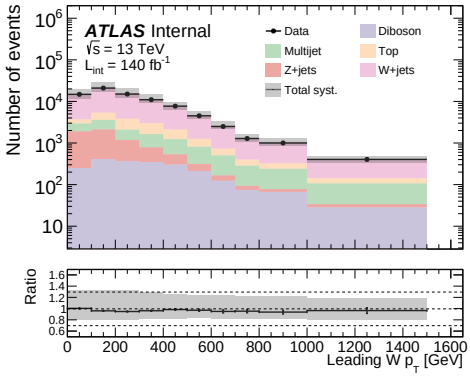
(a) Muon



(b) Electron

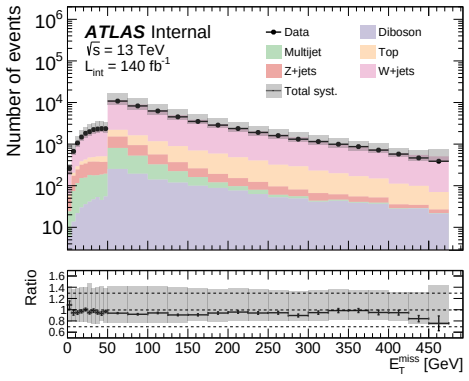


(c) Muon

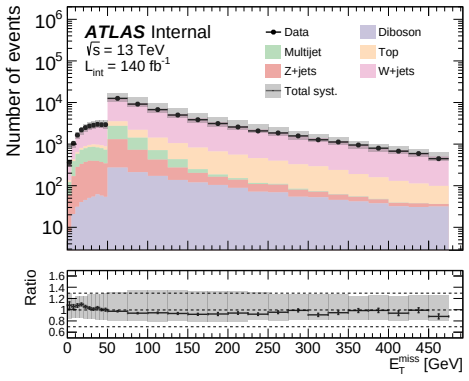


(d) Electron

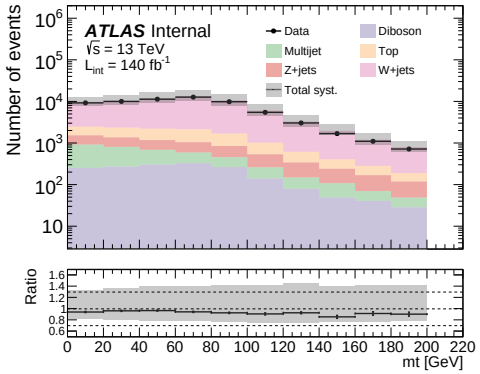
Figure 9.8: Differential distributions at reconstruction-level in the **collinear** phase-space. (left:electron channels, right:muon channel) The grey band includes statistical and systematic uncertainties added in quadrature.



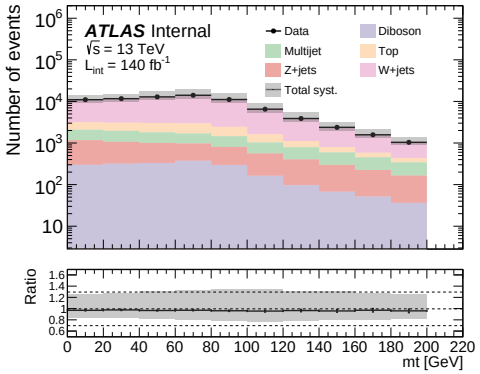
(a) Muon



(b) Electron

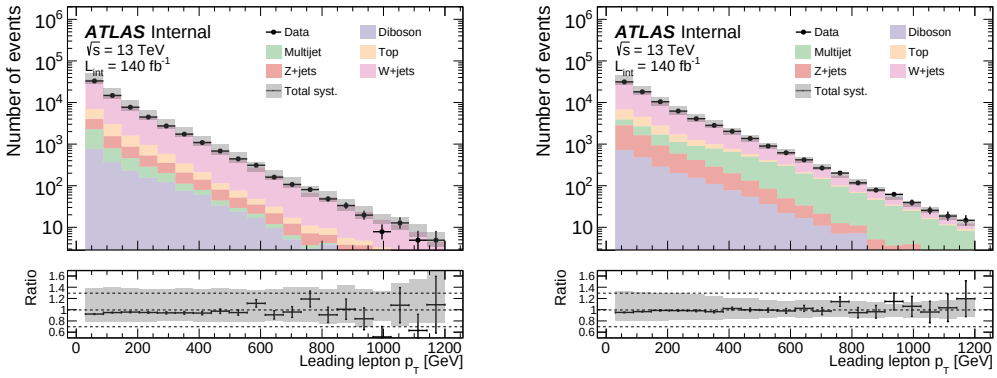


(c) Muon



(d) Electron

Figure 9.9: Additional differential distributions at reconstruction-level in the **collinear** phase-space. (left:electron channels, right:muon channel) The grey band includes statistical and systematic uncertainties added in quadrature. The variables shown here are not planned to be unfolded, and only added for cross checks.



(a) Muon

(b) Electron

Figure 9.10: Additional differential distributions at reconstruction-level in the **collinear** phase-space. (left:electron channels, right:muon channel) The grey band includes statistical and systematic uncertainties added in quadrature. The variables shown here are not planned to be unfolded, and only added for cross checks.

Chapter 10

Unfolding

The measurement of physical processes at the particle level is essential for obtaining results corrected for the detector smearing due to limitations in detector inefficiency and resolution. Particle-level results are required for comparison with other model predictions to validate and refine the modeling of kinematic observables in collisions. To achieve this, a procedure known as "unfolding" is employed to eliminate detector effects and obtain accurate measurements at the particle level.

In this analysis, an iterative Bayesian unfolding procedure, based on the method proposed by D'Agostini [87] and implemented using the RooUnfold package [88], is utilized to derive cross-section results at the particle level. The details of this unfolding procedure are explained in Section 10.1.

The unfolding process exclusively employs reconstruction-level objects that are truth-matched to particle-level objects. The matching criteria compare objects at both the truth and reconstruction levels to ensure that the reconstructed object originates from the truth object. The details of the matching procedure will be described in Section 10.2.

The number of matched events is obtained by summing events across all particle-level (truth) bins using the corresponding conditional probability described in Equation (10.9). Contributions from non-truth-matchable reconstructed events are estimated from MC simulations and subtracted from the total data yield, as represented in Equation (10.8).

Additionally, several aspects of the unfolding metrics for the W+jets signal are addressed in the following sections. These include the response matrix, which describes the correlation between matched object kinematic quantities at truth and reconstruction

levels, selection efficiency due to inefficiencies in object reconstruction, the unmatched fraction due to pile-up jets faking signal events, and purity, which describes the diagonality of the response matrix.

10.1 Formulation

In any experiment, the distribution of measured observables can deviate from their expected values due to a combination of physics modeling and reconstruction smearing effects. The unfolding procedure is expressed in Equation (10.5), where a conditional probability term $P\left(N_j|\eta_i\right)$ encodes the correlation information of events from the particle-level bin (η_i) to the reconstruction-level bin (N_j), and a β_i^{reco} term describes the backgrounds. Here, the index i represents the chosen particle-level bins, and j runs over all reconstruction-level bins. The conditional probability can be split into two parts based on truth-matching criteria, as shown in Equation (10.1), each of which is independently estimated by the MC simulation:

$$P\left(N_i|\eta_j\right) = P\left(N_i^{\text{matched}}|\eta_j\right) + P\left(N_i^{\text{non-matched}}|\eta_j\right), \tag{10.1}$$

where N_i^{matched} represents the number of reconstructed objects that are associated and truth-matched to particle-level objects. Reconstructed objects that cannot find the corresponding particle-level objects are enumerated by $N_i^{\text{non-matched}}$. The rate at which the objects are truth-matched is quantified by the truth-matching efficiency m_j, or equivalently the unmatched fraction $\mathcal{P}_j$, as shown in Equation (10.2)

$$1 - \mathcal{P}_j = m_j = \frac{\sum_k P\left(N_k^{\text{matched}} | \eta_j\right)}{\sum_i P\left(N_i | \eta_j\right)} \tag{10.2}$$

The overall reconstruction efficiency r_j in Equation (10.3) represents the probability that a particle-level event in bin j will pass the selection criteria, regardless of which reconstruction-level bin i it falls into and whether it was truth-matched or not.

$$r_j = \sum_i \frac{P\left(N_i | \eta_j\right)}{\eta_j} \tag{10.3}$$

The selection efficiency ϵ_j, which will be discussed in Section 10.4.2, defined by Equation (10.3) as a simple product of two previous terms defined above in Equation (10.4):

$$\epsilon_j = m_j \times r_j = \frac{\sum_k P\left(N_k^{\text{matched}} | \eta_j\right)}{\eta_j} \tag{10.4}$$

Using the MC W+jets samples, the number of reconstructed (N_i) and reconstructed and truth matched (N_i^{matched}) events can be wrriten as follows:

$$N_i = \sum_j P\left(N_i|\eta_j\right)\eta_j + \beta_i^{\text{reco}} \tag{10.5}$$

$$= \sum_j P\left(N_i^{\text{matched}}|\eta_j\right)\eta_j + \sum_j P\left(N_i^{\text{non-matched}}|\eta_j\right)\eta_j + \beta_i^{\text{reco}} \tag{10.6}$$

$$= \sum_j P\left(N_i^{\text{matched}}|\eta_j\right)\eta_j + N_i^{\text{non-matched}} + \beta_i^{\text{reco}} \tag{10.7}$$

$$N_i^{\text{matched}} = N_i - N_i^{\text{non-matched}} - \beta_i^{\text{reco}} \tag{10.8}$$

$$= \sum_j P\left(N_i^{\text{matched}}|\eta_j\right)\eta_j \tag{10.9}$$

The goal of the unfolding procedure, described in the next section, is to accurately estimate the number of particle-level events η_j in each bin of the given observable, based on the observed number of reconstructed events N_i in each bin of the observable.

10.2 Matching of truth and reconstructed objects

Since all regions include a selection on the leading jet and the lepton, these leading jet and lepton in the events are required to be within $\Delta R < 0.4$ of the corresponding selected particle-level objects for all observables. For measurements in the inclusive, collinear, and back-to-back regions, the jet found to be closest to the lepton must be truth-matched to the closest particle-level jet. In the case of measuring the dijet mass m_{jj}, the sub-leading jet must also be truth-matched in addition to the closest jet matching criteria. For observables involving many jets, such as S_{T} and jet multiplicity, no additional matching criteria are applied at the object level.

10.3 Unfolding procedure

As a reminder, the measured cross sections presented in this note are evaluated within the fiducial phase space summarized in Table 7.1. The cross sections are defined at the particle level using the input container `AntiKt4TruthWZJets`. Electrons and muons originating from the $W \to \ell\nu$ decay are *dressed* by summing the four-vector of the lepton with all prompt photons within a cone of $\Delta R < 0.1$ around the lepton.

Reconstructed events are unfolded to the particle level using an iterative Bayesian technique based on Bayes' theorem [87]. The conditional probability is obtained according to the Bayes' theorem:

$$P\left(\eta_j | N_i^{\text{matched}}\right) = \frac{P\left(N_i^{\text{matched}} | \eta_j\right) \times P\left(\eta_j\right)}{P\left(N_i^{\text{matched}}\right)} \tag{10.10}$$

$$= \frac{P\left(N_i^{\text{matched}} | \eta_j\right) \times P\left(\eta_j\right)}{\sum_k P\left(N_i^{\text{matched}} | \eta_k\right) \times P\left(\eta_k\right)}, \tag{10.11}$$

where $P\left(\eta_j\right)$ is the prior probability of particle-level bin j estimated with MC samples. The unfolded number of events at the particle level is then given by Equation (10.12):

$$\hat{\eta}_j = \sum_i P\left(\eta_j | N_i^{\text{matched}}\right) \times N_i^{\text{matched}}, \tag{10.12}$$

where $\hat{\eta}_j$ is the unfolded particle-level result at an intermediate iteration. This value is used to update the prior probability for the next iteration. The $\hat{\eta}_j$ is expected to converge to η_j after sufficient number of iterations. The iterative procedure is implemented in the

RooUnfold package [88], and two Bayesian iterations are used in the unfolding process (see Appendix A.5.3).

As described in Section 10.2, for MC events, leptons and jets at reconstruction and particle levels are matched using the angular separation ΔR(particle, reco) criteria. If an object fails to satisfy the matching criteria, it does not enter the response matrix, a correlation matrix between kinematic observables described in Section 10.4.1, and is considered an unmatched fake, which is subtracted from the data along with other backgrounds. The SHERPA 2.2.11 prediction is used to construct the response matrices and estimate the unmatched events when unfolding the background-subtracted data sample.

Events with generator-level leading jet $p_{\mathrm{T}}>$ 500 GeV can migrate outside of the fiducial phase-space when the reconstructed level leading jet $p_{\mathrm{T}}<$ 500 GeV (due to the limited jet energy resolution). In order to account for these migration effects, we include an additional underflow bin for all observables where the selection on the reconstructed level leading jet p_{T} satisfies $400 < p_{\mathrm{T}} < 500$ GeV. Similar to every other bin, all background events and unmatched events are subtracted from the data sample. Note that the fraction of unmatched events is large in this underflow bin since most events do not match to generator level events with $p_{\mathrm{T}}>$ 500 GeV.

A table showing the truth-matching efficiency in various bins with $p_{\mathrm{T}} < 500$ GeV is shown below in Section 10.3. The majority of the migrations are accounted for by extending the selection down to 400 GeV and including this in the underflow bin. From

MC, approximately 9% of matched events exist in this underflow bin. The response matrices below include this underflow bin.

Reconstruction-level selection	Fraction of total MC
$400 < p_{\mathrm{T}}^{\mathrm{jet}} \leq 500$	9.12%
$300 < p_{\mathrm{T}}^{\mathrm{jet}} \leq 400$	0.15%
$100 < p_{\mathrm{T}}^{\mathrm{jet}} \leq 300$	0.03%

Table 10.1: Fraction of events not accounted for due to the phase-space migrations out of the reconstruction-level selections used by the analysis. The nominal fiducial selection is applied, and thus includes the generator level leading jet $p_{\mathrm{T}} > 500$ GeV requirement.

10.4 Unfolding metrics

In this section, the response matrices and various unfolding metrics are presented.

10.4.1 Response matrix

The response matrices for the unfolding procedure for each observable are developed using MC simulated samples. These matrices encode the event-wise mapping between reconstruction-level and particle-level events, corresponding to the term $P\left(N_i^{\mathrm{matched}}|\eta_j\right)$ in Equation (10.9). Events are required to pass all selection requirements at both the

reconstruction and particle levels, in addition to the object-matching criteria described above.

The response matrices are shown below in Figures 10.1 to 10.4. For a given reconstruction-level bin column, all the truth bin contents are normalized to unity, such that y-values value in a given reconstructed (x-axis) bin correspond to the rate at which a given particle-level event migrates around. The underflow bin contains events outside the nominal reconstructed-level selection that migrate into the fiducial measurement phase-space (e.g. they have generator level leading jet p_{T}> 500 GeV, but their reconstructed level jet p_{T} is between [400,500] GeV).

The migrations in the response matrix off-diagonal elements for the $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ observbale, as shown in Figure 10.1, are very small. This is because it uses the well-measured angular directions of the lepton and the closest jet. In contrast, larger migrations are observed for quantities such as jet multiplicity and S_T, where only the leading jet is matched, while other jets are not truth-matched, leading to more off-diagonal contributions. For $p_{\mathrm{T}}^{\mathrm{leading}}$, as illustrated in Figure 10.3 (c) and (d), some migrations correspond to situations where two particle-level jets are clustered as separate jets, but appear as a single jet at the reconstruction level. In these cases, the total energy and geometric center of the individual particle-level jets are roughly consistent with the single reconstruction-level jet. For the case of W p_{T}, as shown in Figure 10.4 (c) and (d), matching only the lepton is insufficient to achieve a one-to-one correspondence between truth and reconstruction levels. The energy uncertainty introduced in the $E_{\mathrm{T}}^{\mathrm{miss}}$ object

inherently affects the W boson, leading to off-diagonal contributions in the response matrix.

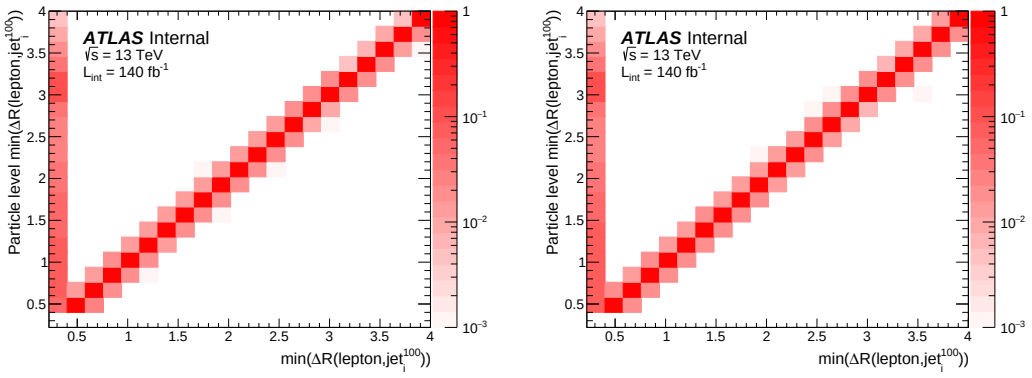


Figure 10.1: Response matrices for unfolding of the $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ observable in the inclusive phase-space for the (a) electron and (b) muon channel. Note that the first bin is a special underflow bin containing all events reconstructed with 400 GeV $< p_{\mathrm{T}} <$ 500 GeV, wheree vents migrating from outside the reco-level selection.

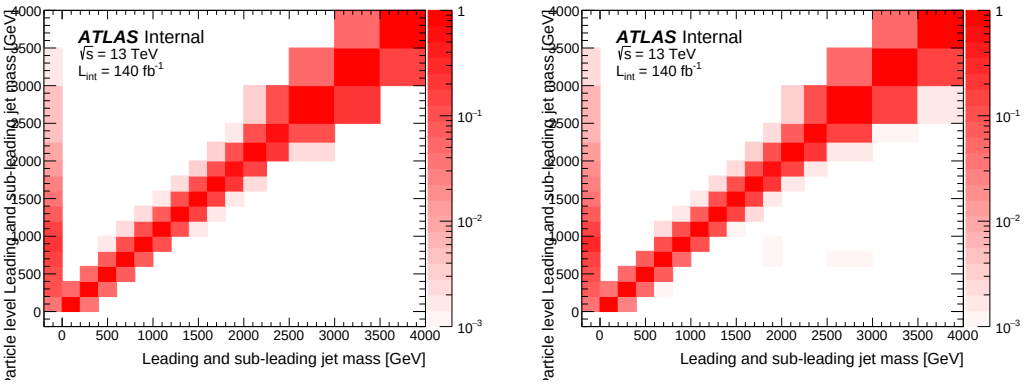
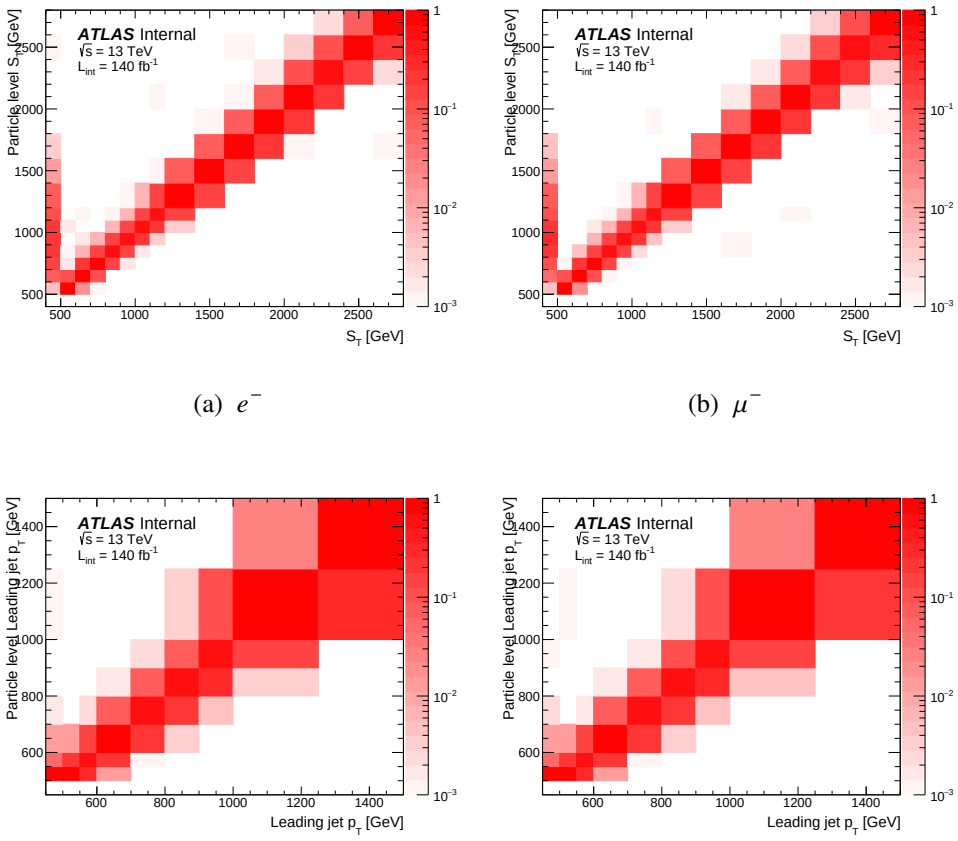


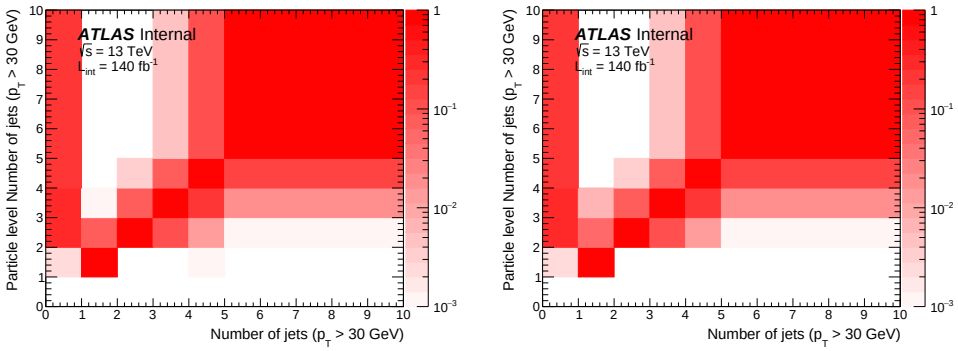
Figure 10.2: Response matrices for unfolding of the m_{jj} observable in the inclusive phase-space for the (a) electron and (b) muon channel with the truth-matching additionally applied to the sub-leading jet. Note that the first bin is a special underflow bin containing all events reconstructed with 400 GeV $< p_{\mathrm{T}} <$ 500 GeV, wheree vents migrating from outside the reco-level selection.



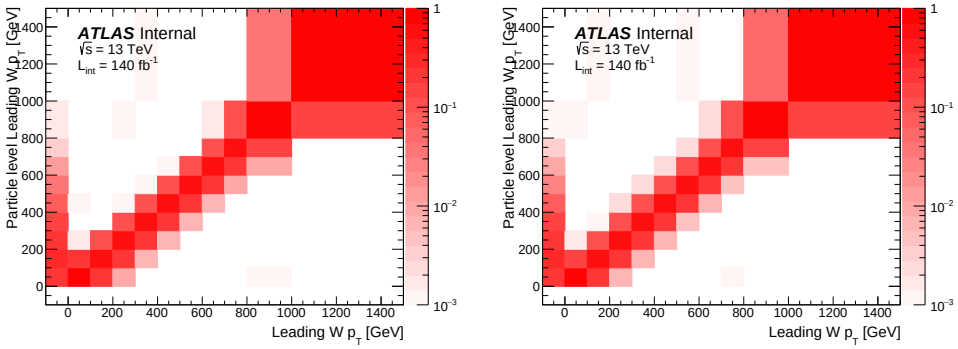
(a) e^- (b) μ^-

(c) e^- (d) μ^-

Figure 10.3: Response matrices for unfolding of the S_T and leading jet p_{T} observables in the collinear phase-space. Note that the first bin is a special underflow bin containing all events reconstructed with 400 GeV $< p_{\mathrm{T}} <$ 500 GeV, wheree vents migrating from outside the reco-level selection.



(a) e^- (b) μ^-



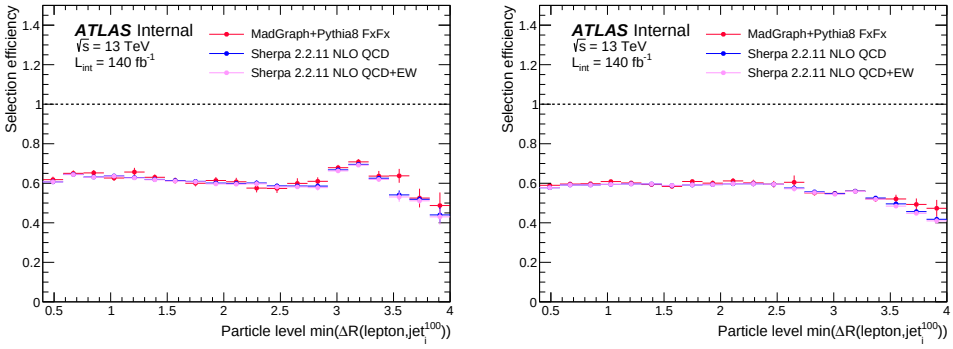
(c) e^- (d) μ^-

Figure 10.4: Response matrices for unfolding of jet multiplicity and W p_{T} observables in the collinear phase-space. Note that the first bin is a special underflow bin containing all events reconstructed with 400 GeV $< p_{\mathrm{T}} <$ 500 GeV, wheree vents migrating from outside the reco-level selection.

10.4.2 Selection efficiency

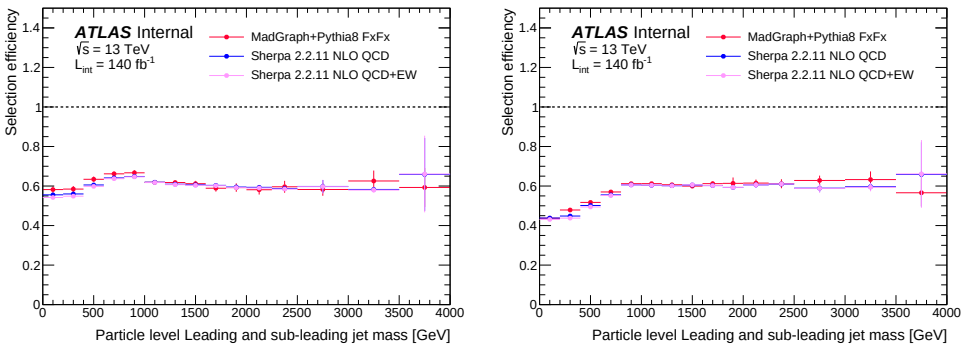
Due to detector inefficiencies related to the trigger, object reconstruction, or offline selection criteria, the number of events matched between the reconstruction level and particle level will always be greater than the number of events selected at the particle level. The *selection efficiency*, determined from MC samples as shown in Equation (10.4), quantifies this rate. Here, N_k^{matched} represents the number of reconstructed events matched to a particle-level event, and η_j is the number of particle-level events. The numerator includes events that pass both particle and reconstruction-level selections, as well as the object-level matching described in Section 10.3, while the denominator includes events that pass only the corresponding particle-level selections.

The selection efficiencies for electrons and muons in the inclusive and collinear selections are shown in Figures 10.5 to 10.8. The selection efficiencies are approximately 50% to 60% and good agreement is obtained between the various generator predictions (SHERPA 2.2.1 vs SHERPA 2.2.11). The turn-on of efficiency at low W p_{T} arises from the lepton PID requirements and is slower for electrons than muons due to the inefficiency present in the TightLLH electron working point used in this analysis.



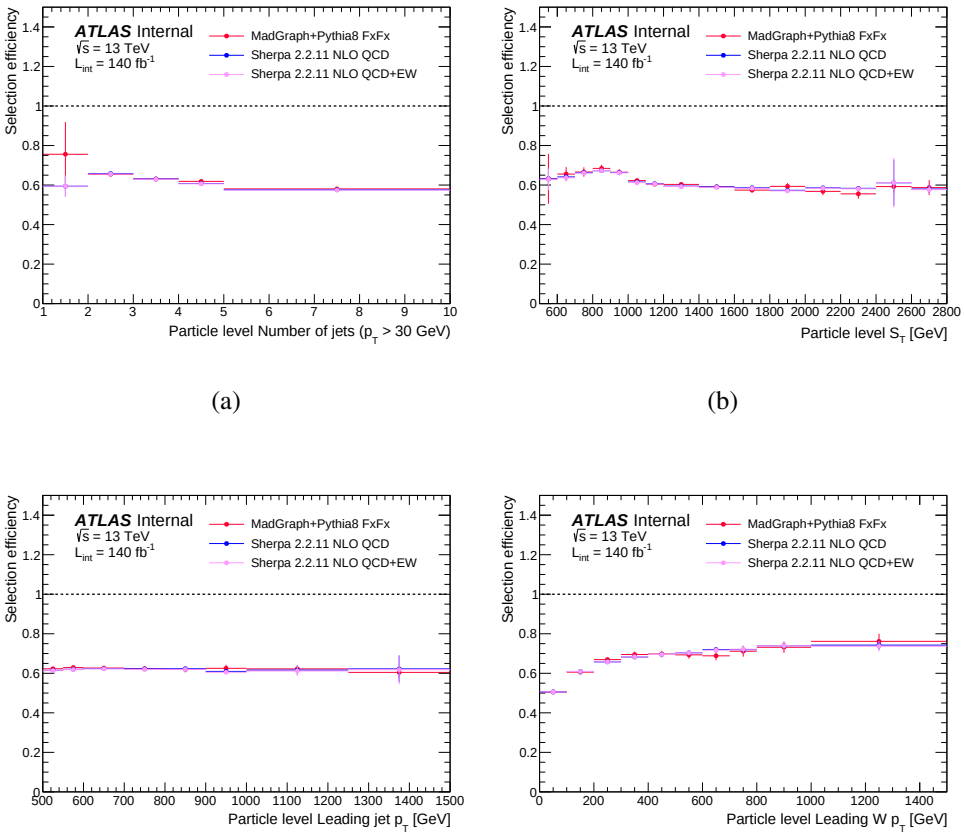
(a) e^- (b) μ^-

Figure 10.5: Selection efficiency for $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ observable in the inclusive phase-space for electrons (a) and muons (b).



(a) e^- (b) μ^-

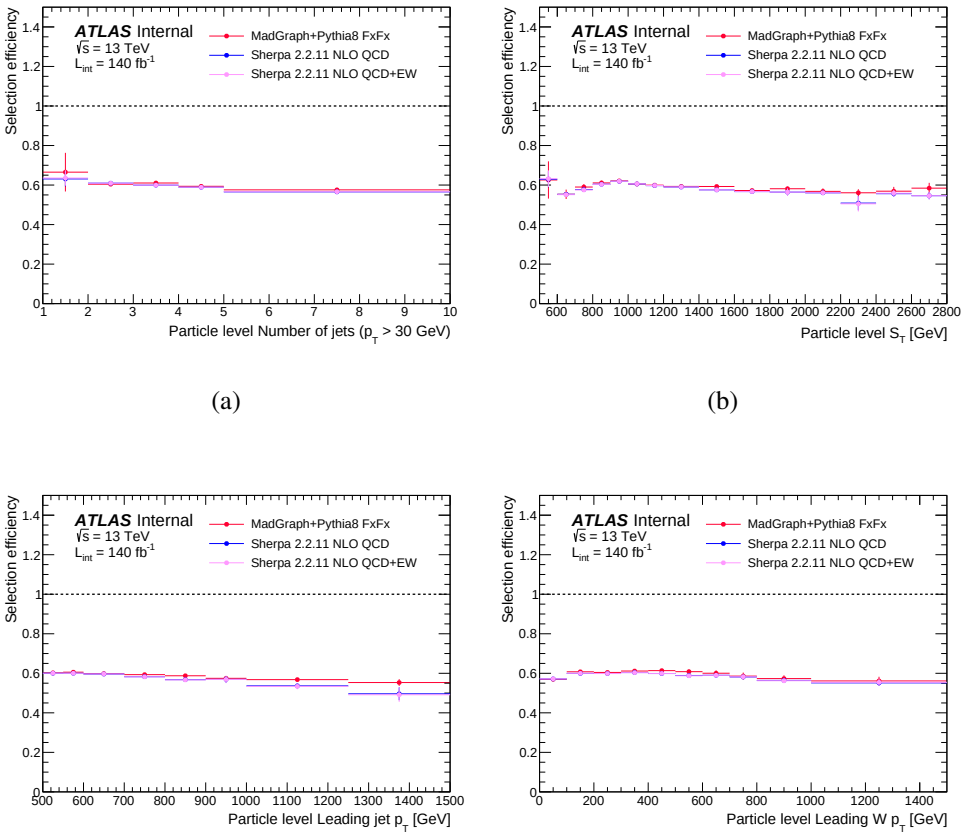
Figure 10.6: Selection efficiency for sum of leading and sub-leading jets mass (m_{jj}) in the inclusive phase-space for electrons and muons with the truth-matching additionally applied to the sub-leading jet.



(a) (b)

(c) (d)

Figure 10.7: Selection efficiency for jet multiplicity, S_T, leading jet p_{T}, and W boson p_{T} in the collinear phase-space for **electron** channel.



(a) (b)

(c) (d)

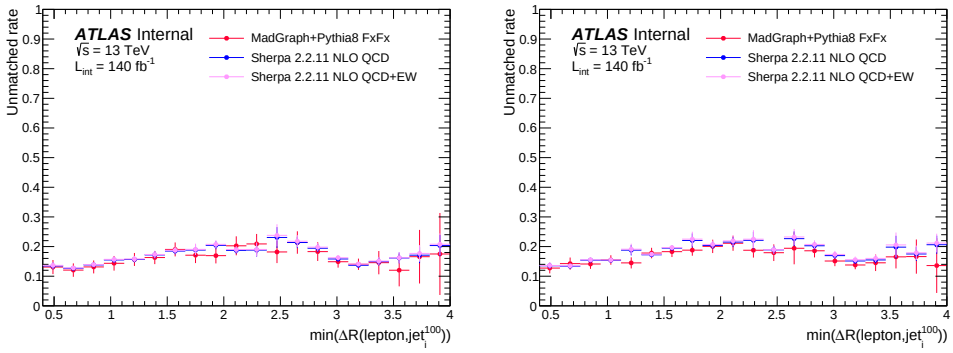
Figure 10.8: Selection efficiency for jet multiplicity, S_T, leading jet p_{T}, and W boson p_{T} in the collinear phase-space for **muon** channel.

10.4.3 Unmatched fraction

Reconstructed events might fail to match to a particle-level event or objects due to factors such as pile-up jets faking those from the hard-scatter event of interest. The unmatched fraction quantifies this mismatch and is defined in Equation (10.2) with algebraic rearrangement, resulting in Equation (10.13). Here, $P\left(N_k^{\text{not-matched}}|\eta_j\right)$ represents the rate of reconstructed events *failing* the event or object-level matching requirements, while $P\left(N_i|\eta_j\right)$ represents the total rate of reconstructed events passing the offline selection criteria.

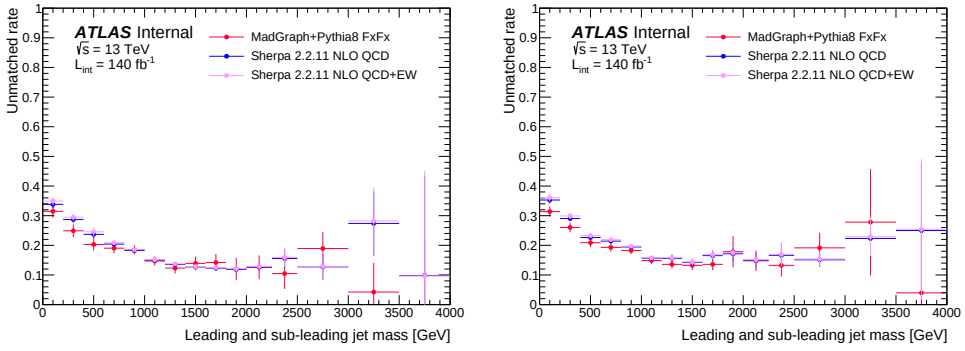
$$\mathcal{P}_i = 1 - m_i = \frac{\sum_k P\left(N_k^{\text{not-matched}}|\eta_j\right)}{\sum_i P\left(N_i|\eta_j\right)} \tag{10.13}$$

The unmatched fraction for electrons and muons in the inclusive and collinear regions are shown below in Figures 10.9 to 10.12. Most distributions have a flat unmatched fraction, with a few exceptions; the leading jet p_{T} has large unmatched fraction close to the phase-space selection of $p_{\text{T}}>$ 500 GeV. This arises from events where the leading jet in the truth selection is below the $p_{\text{T}}>$ 500 GeV threshold, but is reconstructed above this threshold. This effect is strongest closest to the cut boundary, then decreases below the 10% level. Since the leading jet is also considered in other observables, such as S_{T} and m_{jj}, similar effects are observed but considerably more smeared out since these observables include other sub-leading jets in their definitions.



(a) e^- (b) μ^-

Figure 10.9: Non-truth-matching rate as function of $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$ observabe for the inclusive phase-space of electron and muon channels.



(a) e^- (b) μ^-

Figure 10.10: Non-truth-matching rate as function of the sum of leading and sub-leading jets mass (m_{jj}) in the inclusive phase-space for electron and muon channels. The truth-matching additionally applied to the sub-leading jet.

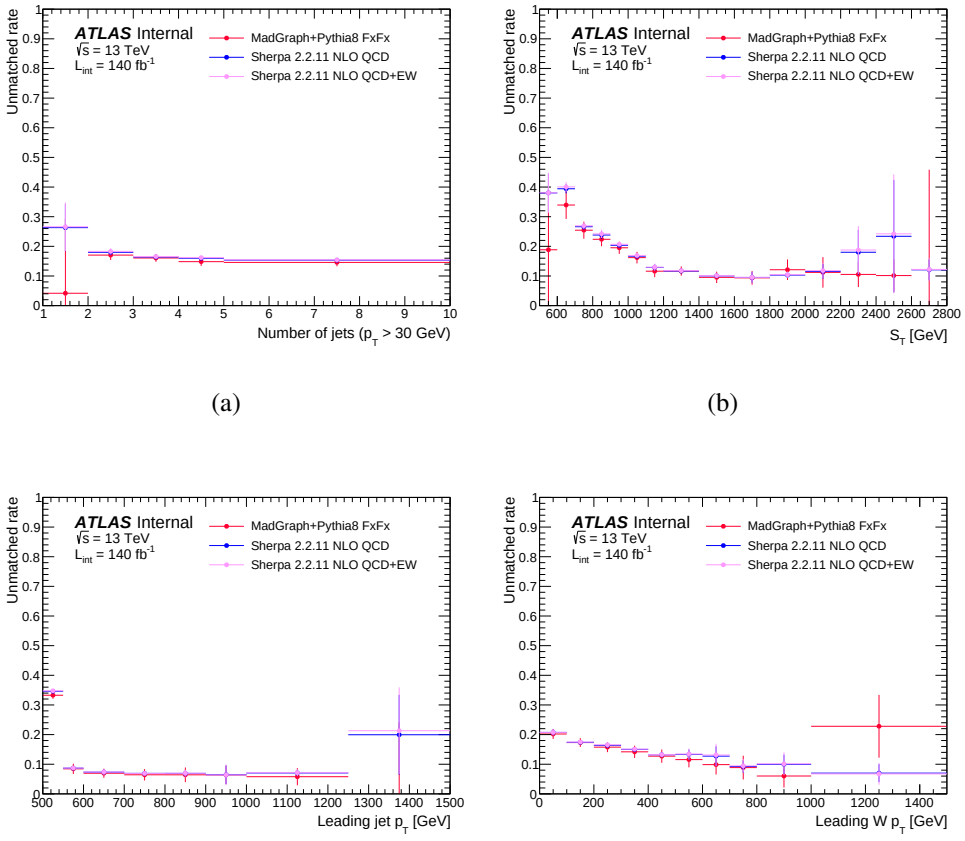


Figure 10.11: Non-truth-matching rate as function of jet multiplicity, S_T, leading jet p_{T}, and W boson p_{T} in the collinear phase-space for **electron** channel.

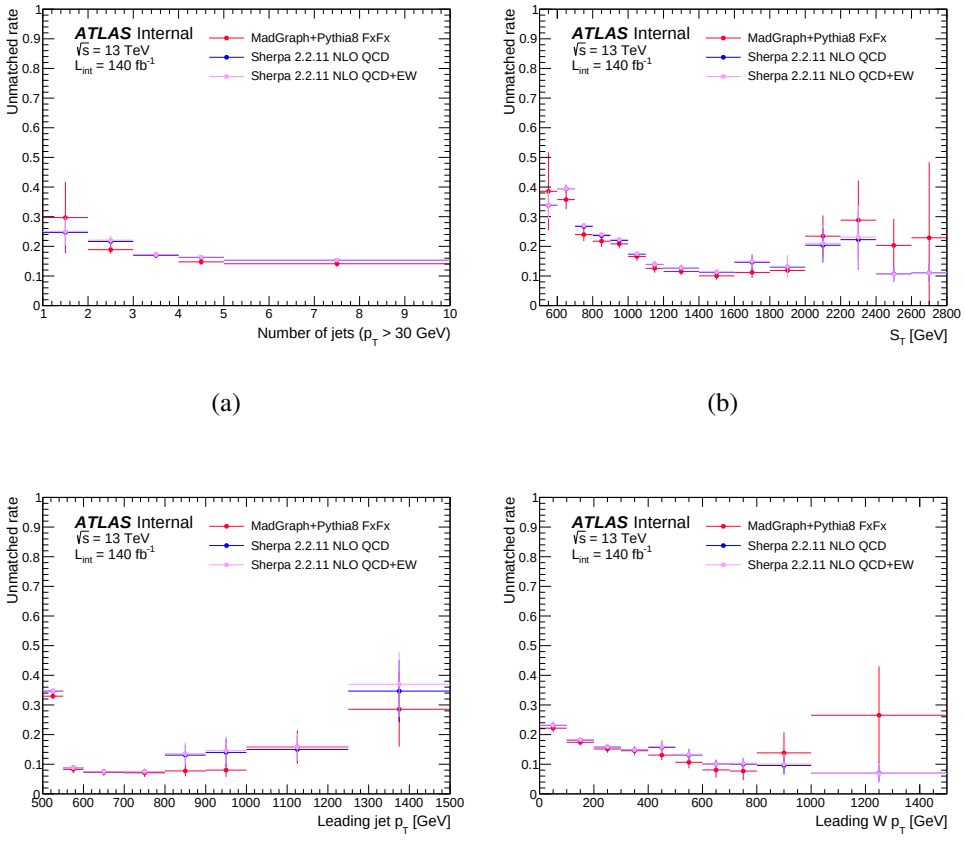


Figure 10.12: Non-truth-matching rate as function of jet multiplicity, S_T, leading jet p_{T}, and W boson p_{T} in the collinear phase-space for **muon** channel.

10.4.4 Purity

The purity, defined in Equation (10.14), measures the fraction of events in any given bin of the reconstructed distribution that are correctly reconstructed in the corresponding particle-level bin, remaining on the diagonal of the response matrix described in Section 10.4.1. This metric is useful for visualizing the rate at which events migrate off the diagonal during unfolding. The purity is defined as:

$$\mathcal{F}_j = \sum_i \frac{N_{i==j}}{N_{ij}}, \tag{10.14}$$

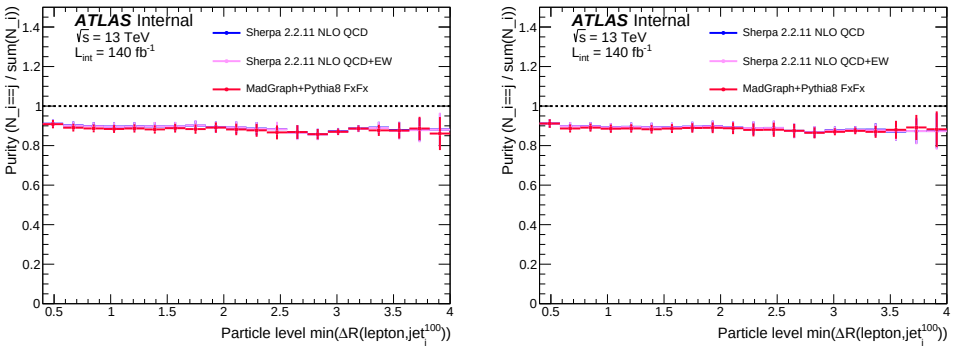
where N_{ij} represents the number of events in a given bin of the response matrix, j indicates the particle-level bins, and i runs over all reconstruction-level bins. To enter the response matrix, events must pass both the event and object-level matching criteria described earlier.

Figures 10.13 and 10.14 shows the purity for electrons and muons in the inclusive phase-space, while Figure 10.15 shows the purity for electrons in the collinear phase-space, and Figure 10.16 for muons.

Due to the non-zero energy and pointing resolution for both reconstructed leptons and jets, the response matrices are not perfectly diagonal as shown in the previous sections. As such, the purity generally tends to increase as a given kinematic observable increases to larger values when the bin width is fixed[1]. Therefore, in order to keep the purity high the bin width is enlarged for increasing kinematic values. As a result, the purity plots

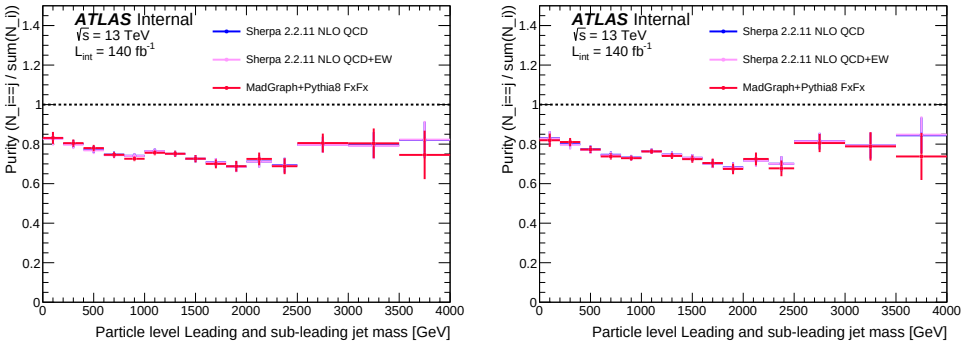
[1] This doesn't necessarily imply the resolution degrades towards large values, but in many cases the *fractional resolution* is fixed, which leads to larger migrations for a fixed bin width

exhibit jumps when the bin width is increased; when the bin width grows, the migration off the diagonal are minimized since it's more likely that the reconstructed and truth event are in the same bin as the bin width increases.



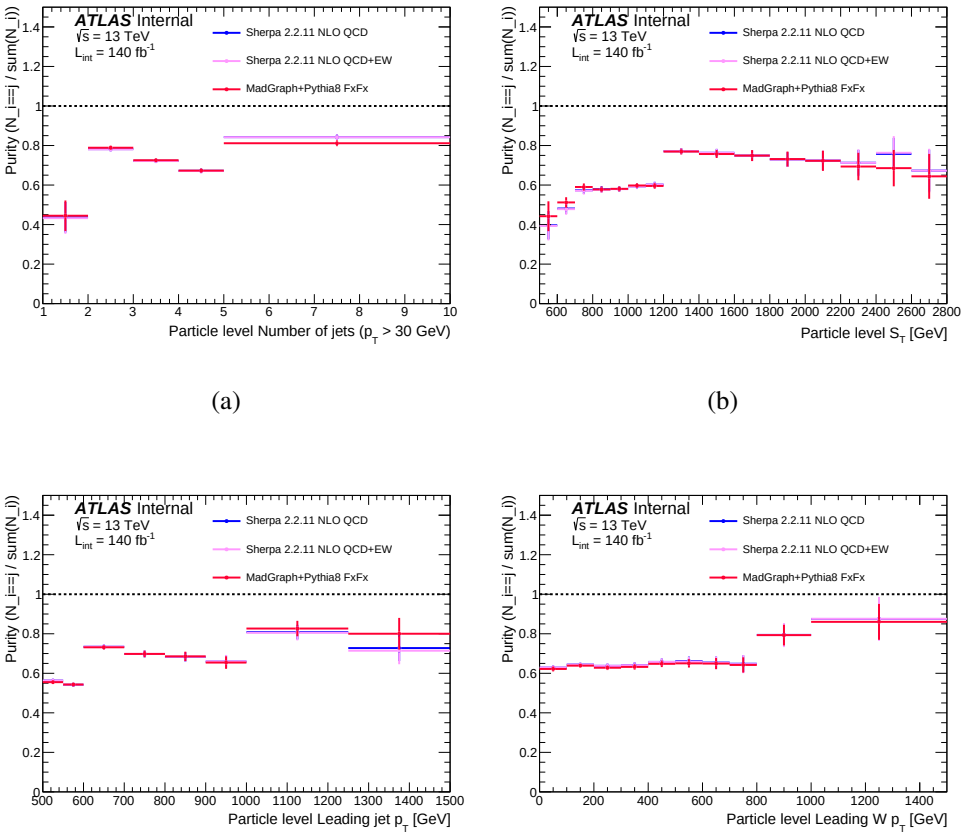
(a) e^- (b) μ^-

Figure 10.13: Distribution of the purity for the $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ observable in inclusive phase-space for electrons (a) and muons (b). Uncertainties are statistical only.



(a) e^- (b) μ^-

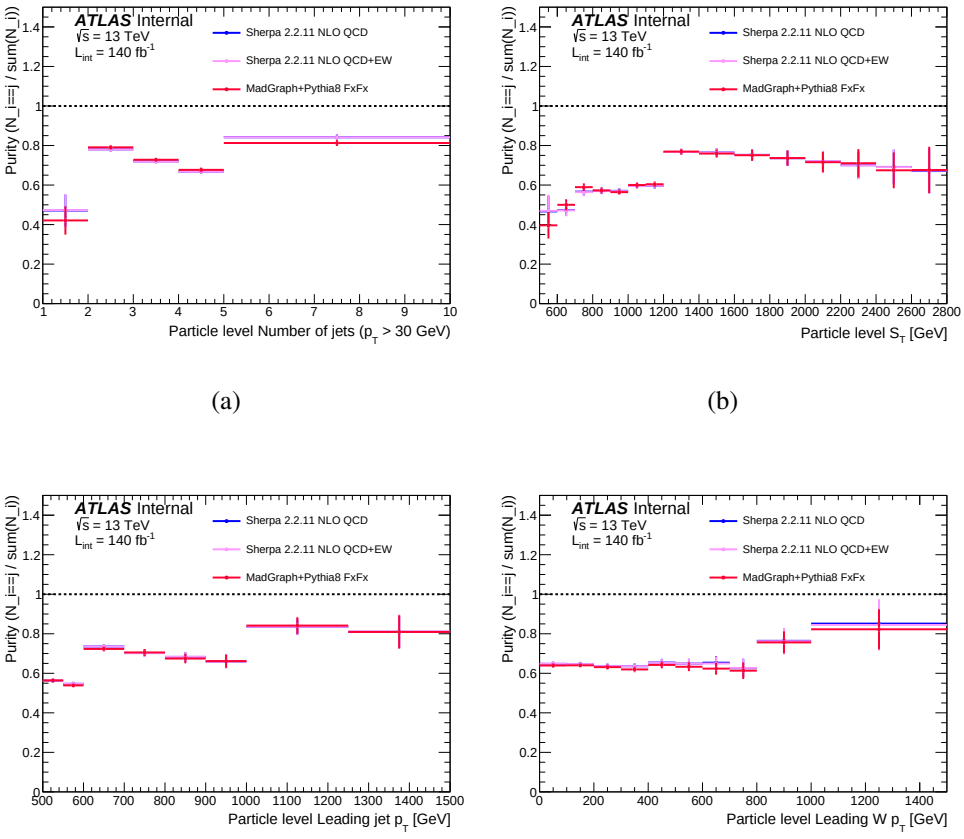
Figure 10.14: Distribution of the purity variable for the m_{jj} distribution in the inclusive phase-space with jet multiplicity >= 2 for electrons (a) and muons (b). Uncertainties are statistical only.



(a) (b)

(c) (d)

Figure 10.15: Distribution of the purity for jet multiplicity, S_T, leading jet p_{T}, and W boson p_{T} in the electron collinear phase-space. Uncertainties are statistical only.



(a) (b)

(c) (d)

Figure 10.16: Distribution of the purity for jet multiplicity, S_T, leading jet p_{T}, and W boson p_{T} in the muon collinear phase-space. Uncertainties are statistical only.

Chapter 11

Systematic Uncertainties

This analysis addresses several sources of uncertainties: *statistical* (Section 11.1), *experimental systematics* (Section 11.2), and *theoretical systematics* (Section 11.3). Systematic uncertainties are propagated to the final results using the *offset method*. In this method, the normalization factor described in Chapter 8 is recalculated for each systematic uncertainty. The signal MC is varied in a correlated manner with the systematic uncertainty, and the background is scaled by the newly derived normalization factor for each systematic uncertainty. The unfolding process is then re-performed, and this procedure for measuring the cross-section is repeated for all theoretical and experimental uncertainties.

11.1 Statistical Uncertainties

The *bootstrapping technique* is used to evaluate the statistical uncertainties and correlations [89]. An ensemble of $N_{\text{rep}} = 10,000$ replicas of the original dataset (in data) is generated, by weighting each event with an integer weight sampled from a Poisson distribution with an expectation value of 1. To generate the replica datasets, a unique and deterministic Poisson weight is assigned to each event based on the event number and index of the replica ($i \in [0 \ldots 9999]$) which contains them. In this way, the technique allows for the measurement of complex statistical correlations between the observables of interest $\hat{O}$ that we measure, exploiting the fact that the replica datasets will contain the same statistical fluctuation for the events in common among the various measurements. From this, we can evaluate the statistical covariance via Equation (11.1a) (correlation

via Equation (11.1b)). In this method, the covariance and correlation coefficient ρ_{ab} between observables a and b are given by

$$\text{covariance}(\hat{a}, \hat{b}) = \frac{1}{N_{\text{rep}}} \sum_{i=0}^{N_{\text{rep}}} (a_i - \bar{a})(b_i - \bar{b}) \tag{11.1a}$$

$$\rho_{ab} = \text{correlation}(\hat{a}, \hat{b}) = \frac{\text{covariance}(\hat{a}, \hat{b})}{\sigma_a \sigma_b} \tag{11.1b}$$

where a_i, b_i are the measured values of the observables $\hat{a}, \hat{b}$ in bootstrap replica i, the variance of $\hat{a}$ (or $\hat{b}$) is the square of the statistical uncertainty σ_a (σ_b) as in Equation (11.2).

$$\sigma_a = \sqrt{\text{covariance}(\hat{a}, \hat{a})} \qquad \sigma_b = \sqrt{\text{covariance}(\hat{b}, \hat{b})} \tag{11.2}$$

More details on bootstrapping uncertainties of this technique, and the determination of covariance can be found in [89].

11.2 Experimental Uncertainties

The experimental uncertainties considered in this measurement arise from detector limitations and the algorithms used in physics object reconstruction and identification. These include uncertainties associated with the reconstructed jet energy due to calorimeter energy resolution and sensitivity, as well as uncertainties in lepton reconstruction and identification. These uncertainties are propagated through the combined performance tools and recorded in the output ROOT trees. They are classified into weight-based and tree-based systematics, depending on whether their impact changes the selection

acceptance of events. Weight-based systematics do not change acceptance, while tree-based systematics do; hence, tree-based systematics require an assessment of their impact on the acceptance of events passing our phase-space cuts. Consequently, tree-based systematics are stored in separate trees to enable the analysis to be re-run. The following sub-sections will discuss the sources of experimental uncertainties in this measurement.

11.2.1 Pile-up

Simulated events are reweighted to match the average number of reconstructed vertices in the data. The number of reconstructed vertices and the actual number of interactions are sensitive to the visible interaction rate, as determined by independent detector measurements. To account for the pile-up uncertainty, a scale factor of $\sigma = \pm 1.09$ is applied to each MC simulated event.

11.2.2 Jet energy scale and resolution

In this measurement, we utilize a strongly reduced parameter set, including uncertainties in both the jet energy scale (JES) and the jet energy resolution (JER). The jet energy scale and resolution are calibrated using in-situ measurements to align the data with the MC simulations. Additionally, η-intercalibration uncertainties are included to account for variations in detector modeling across different η regions and transitions between detector components. Pile-up corrections, which depend on the number of interactions, reconstructed vertices, and the jet p_{T}, are also considered. Each systematic

applies a $\pm 1\sigma$ (1-up/down) variation with respect to the nominal value to MC events. The full list of jet energy scale and resolution uncertainties is provided by the JetETmiss group, resulting in the following systematic names [72]:

1. Uncertainties on (b-)jet falvour response and composition:

   - JET_BJES_Response
   - JET_Flavor_Composition
   - JET_Flavor_Response

2. Uncertainties on Jet η-tercalibration that contains statistical, methodology, and detector modeling:

   - JET_EtaIntercalibration_NonClosure_2018data
   - JET_EtaIntercalibration_NonClosure_highE
   - JET_EtaIntercalibration_NonClosure_negEta
   - JET_EtaIntercalibration_NonClosure_posEta
   - JET_EtaIntercalibration_TotalStat

3. Uncertainties on jet energy resolution:

   - JET_JER_DataVsMC_MC16
   - JET_JER_EffectiveNP_1
   - JET_JER_EffectiveNP_2
   - JET_JER_EffectiveNP_3
   - JET_JER_EffectiveNP_4
   - JET_JER_EffectiveNP_5
   - JET_JER_EffectiveNP_6
   - JET_JER_EffectiveNP_7
   - JET_JER_EffectiveNP_8
   - JET_JER_EffectiveNP_9
   - JET_JER_EffectiveNP_10
   - JET_JER_EffectiveNP_11
   - JET_JER_EffectiveNP_12restTerm

4. Uncertainties on jet energy scale:

- JET_EffectiveNP_Detector1
- JET_EffectiveNP_Detector2
- JET_EffectiveNP_Mixed1
- JET_EffectiveNP_Mixed2
- JET_EffectiveNP_Mixed3
- JET_EffectiveNP_Modelling1
- JET_EffectiveNP_Modelling2
- JET_EffectiveNP_Modelling3
- JET_EffectiveNP_Modelling4
- JET_EffectiveNP_R10_Mixed1
- JET_EffectiveNP_R10_Modelling1
- JET_EffectiveNP_Statistical1
- JET_EffectiveNP_Statistical2
- JET_EffectiveNP_Statistical3
- JET_EffectiveNP_Statistical4
- JET_EffectiveNP_Statistical5
- JET_EffectiveNP_Statistical6
- JET_Rtrk_Baseline_frozen_mass
- JET_Rtrk_ExtraComp_Baseline_frozen_mass
- JET_Rtrk_ExtraComp_Modelling_frozen_mass
- JET_Rtrk_Modelling_frozen_mass
- JET_Rtrk_TotalStat_frozen_mass
- JET_Rtrk_Tracking1_frozen_mass
- JET_Rtrk_Tracking2_frozen_mass
- JET_Rtrk_Tracking3_frozen_mass

5. Uncertainties on single-hadron response component in the high p_{T} range used as extension of the in-situ measurements:

   - JET_PunchThrough_MC16

6. Uncertainties on punch-through component resulting from the punch-through correction in the jet calibration:

   - JET_SingleParticle_HighPt

11.2.3 Lepton energy scale and resolution

The lepton energy scale and resolution corrections are applied to each lepton event to improve agreement with the data. Similar to other experimental uncertainties, the uncertainties on each of the lepton energy scale and resolution corrections are evaluated

by applying a $\pm 1\sigma$ (1-up/down) variation to the nominal setup and reweighting each MC event. The uncertainties on lepton energy scale and resolution, provided by the ATLAS EGAM[1][90] and MUON[2][91] performance groups, considered in this measurement include the following:

1. Uncertainties on electron energy smearing in simulation, and uncertainties on electron energy scale calibration in simulation with specific setup and parameterization of ATLAS detectors simulation in GEANT4:

   - EG_RESOLUTION_ALL
   - EG_SCALE_ALL

2. Uncertainties on charge-agnostic smearing of simulated muon p_{T} in the Inner Detector (ID) and Muon Spectrometer (MS), and muon momentum scale and sagitta biases in simulation:

   - MUON_ID
   - MUON_MS
   - MUON_SAGITTA_RESBIAS
   - MUON_SAGITTA_DATASTAT
   - MUON_SCALE

11.2.4 $E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term

All sources of uncertainties related to the inputs for $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction are consistently propagated, ensuring that changes in parameters such as the jet energy scale are

[1] ATLAS Electron and Photon group
[2] ATLAS Muon group

reflected in the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ value. Moreover, additional uncertainties associated with the track soft-term are taken into account [92]:

- MET_SoftTrk_ResoPara
- MET_SoftTrk_ResoPerp
- MET_SoftTrk_ScaleDown
- MET_SoftTrk_ScaleUp

11.2.5 JVT

The calibrations for the Jet Vertex Tagger (JVT), which is an algorithm used for pile-up rejection in the jet reconstruction process as described in Section 6.2.3, are derived from a selection involving $Z \rightarrow \mu^{+}\mu^{-}$ + jets events and are provided by the JetEtMiss group. These calibrations contains uncertainties associated with both the statistical component and the estimation of pile-up jets within the measurement region used for extraction. Details regarding the uncertainties on JVT provided by the JetEtMiss group tool can be found in [92], and the following are the two JVT uncertainties included in this measurement:

- JET_JvtEfficiency__1down
- JET_JvtEfficiency__1up

11.2.6 Flavor tagging

The flavor of jet objects is determined by flavor tagging algorithms provided by the ATLAS Flavor Tagging Performance Group[93]. The uncertainties related to flavor tagging efficiency are estimated by applying $\pm 1\sigma$ (1-up/down) variations to the jet p_{T}, η, flavor, and scale factors in the tagging algorithm. The set of flavor tagging uncertainties

considered in this measurement includes the following:

- FT_EFF_B_systematics
- FT_EFF_C_systematics
- FT_EFF_Light_systematics
- FT_EFF_extrapolation
- FT_EFF_extrapolation_from_charm
- bTag_B_0
- bTag_Light_0
- bTag_Light_1

11.2.7 Trigger

Trigger systematics associated with the single lepton trigger are considered separately for each channel and are provided by the ATLAS EGAM and Muon Combined Performance groups, as detailed in [18] and [19]. These systematics are estimated by applying $\pm 1\sigma$ (1-up/down) variations of the trigger parameters to the nominal setup. The lepton trigger systematics considered in this measurement are listed below:

- MUON_EFF_TrigStatUncertainty
- MUON_EFF_TrigSystUncertainty
- EL_EFF_Trigger_TOTAL_1NPCOR_PLUS_UNCOR
- EL_EFF_TriggerEff_TOTAL_1NPCOR_PLUS_UNCOR

11.2.8 Lepton identification, reconstruction, and isolation

The lepton identification, reconstruction, and isolation scale factors are used to correct the efficiencies in MC to match those in the data. The associated systematics for these measurements are provided by the EGAM and Muon Combined Performance groups, with details for electron objects in [94] and for muon objects in [95]. This measurement includes the following systematic uncertainties, with $\pm 1\sigma$ (1-up/down) variations applied relative to the nominal setup:

1. Uncertainties on muon identification, reconstruction, and isolation:

   - MUON_EFF_ISO_STAT
   - MUON_EFF_ISO_SYS
   - MUON_EFF_RECO_STAT
   - MUON_EFF_RECO_STAT_LOWPT
   - MUON_EFF_BADMUON_SYS
   - MUON_EFF_RECO_SYS
   - MUON_EFF_RECO_SYS_LOWPT
   - MUON_EFF_TTVA_STAT
   - MUON_EFF_TTVA_SYS

2. Uncertainties on electron identification, reconstruction, and isolation:

   - EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR
   - EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR
   - EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR
   - EL_EFF_ChargeIDSel_TOTAL_1NPCOR_PLUS_UNCOR
   - EL_CHARGEID_SYStotal
   - EL_CHARGEID_STAT

11.3 Theoretical uncertainties

Sources of theoretical uncertainty relating to the choice of PDF set, scale uncertainties, and other parameters unique to each MC sample, are considered for the signal and backgrounds in this analysis. This section presents the theoretical uncertainties considered for each of the dominant backgrounds as well as for W+ jets.

11.3.1 Top-quark Production

The uncertainties for the SHERPA 2.2.12 $t\bar{t}$ prediction are summarized in Table 11.1. Variations in the renormalization (μ_R) and factorization (μ_F) scales, as well as the value of the strong coupling constant(α_s), are described in [96]. The renormalization and

factorization scale variations include a 5-point variation with respect to the nominal value $(\mu_R, \mu_F) = (1.0, 1.0)$, varying the scales by a factor of 2 as follows: $(\mu_R, \mu_F) = (0.5, 0.5)$, $(0.5, 1.0)$, $(1.0, 0.5)$, $(1.0, 2.0)$, $(2.0, 1.0)$, and $(2.0, 2.0)$. The PDF set used for generating top quark events employs the Hessian eigenvector variations of the NNPDF3.0NNLO PDF set [97]. The variation in the α_s is introduced by using an alternative PDF set provided by the NNPDFCT14NLO set.

The single top uncertainty includes similar 5-point variations in the renormalization and factorization scales, as well as the eigenvector variations of the PDF set. However, a different generator was used, as detailed in Section 5.2. Additionally, since $t\bar{t}$ and W-t production have similar final states, there exists a quantum interference effect [86]. To handle this interference, an additional uncertainty arises from the use of two different techniques, namely the Diagram-Reduction and Diagram-Subtraction schemes. The difference between these techniques is added as an uncertainty to the W-t component of the single top.

11.3.2 Diboson Production

For the WW, WZ, and ZZ backgrounds, variations in the renormalization and factorization scales are computed by POWHEG and provided as internal weights. The CT10nlo Hessian PDF eigenvector variations are used to estimate the PDF uncertainty. Details on the scale setting and relevant PDF sets are provided in Section 5.3.

Source of uncertainty	Nominal	Systematic
Renormalization scale.	SHERPA 2.2.12	internal weights
Factorization scale	SHERPA 2.2.12	internal weights
α_s variations	SHERPA 2.2.12	internal weights
PDF	SHERPA 2.2.12	NNPDF3.0NNLO weights

Table 11.1: Theory uncertainties considered for the pair-production of top-quarks ($t\bar{t}$).

11.3.3 Multi-jet Background

The multi-jet background includes variations in both initial-state radiation (ISR) and final-state radiation (FSR), with ISR and FSR varied independently using the factors $0.5, 0.625, 0.75, 0.875, 1.0, 1.25, 1.5, 1.75, 2.0$. Uncertainties on the A14 set of tuned parameters [42] for parton shower and hadronization are also included. All multi-jet modeling variations were computed using internal weights, and the corresponding variations are applied to each of the nominal events using these weights.

11.3.4 V + jets Production

The theoretical uncertainties considered for both W+jets and Z+jets processes are generated with the same setup, resulting in identical uncertainty prescriptions. Five-point variations for renormalization and factorization scales, similar to those used in $t\bar{t}$ production, are performed. Additionally, Hessian eigenvector variations of the

NNPDF3.0NNLO PDF set and variations in the value of the α_s are included. A summary of the theoretical uncertainties considered for V+jets processes is shown in Table 11.2.

Source of uncertainty	Nominal	Systematic
Renormalization scale.	SHERPA 2.2.11	internal weights
Factorization scale	SHERPA 2.2.11	internal weights
α_s variations	SHERPA 2.2.11	internal weights
PDF	SHERPA 2.2.11	NNPDF3.0NNLO weights

Table 11.2: Theory uncertainties considered for V+jets processes

11.4 Unfolding uncertainties

This analysis follows the guidelines mentioned in the Standard Model unfolding prescription[3] which details three different types of uncertainties associated with the unfolding procedure:

- Detector simulation: we determine this by systematically varying our pseudo-data, unfolding, and treating the difference as an uncertainty. In this procedure, the nominal response matrix is used.

- Basic unfolding uncertainty: we determine this by treating any difference between

[3] SM unfolding tiwki: https://twiki.cern.ch/twiki/bin/viewauth/UnfoldingNewAtlasProtected/StandardModel

W+ jets unfolded-reweighted-reconstructed MC and W+ jets reweighted-truth MC as an uncertainty (called "Basic unfolding uncertainty"). The procedure for determining this error is outlined as follows:

1. Perform a data-driven W+ jets reconstructed MC fit to background-subtracted data to get a reco-level transfer factor.
2. Apply the reco-level transfer factor to W+ jets reconstructed MC to get W+ jets reweighted-reconstructored MC.
3. Transform the reco-level transfer factor to a truth-level transfer factor (using a normalized response matrix).
4. Apply the truth-level transfer factor to W+ jets truth MC to get W+ jets reweighted-truth MC.
5. Unfold the W+ jets reweighted-reconstructed MC using (original) nominal response matrix.
6. Compare the W+ jets unfolded-reweighted-reconstructed MC to W+ jets reweighted-truth MC, and the resulting difference is the basic unfolding uncertainty.

- Hidden Variables: uncertainty due to variation in the response matrix. We use an alternative response matrix from MadGraph5_aMC@NLO +Pythia8 with FxFx merging to determine this uncertainty. The prescription is described below.
- Signal modeling uncertainty: In addition to the basic and hidden variable uncer-

tainties, an uncertainty on the measured cross section is assessed from theoretical uncertainties in the underlying signal processes. The scale and PDF variations in the nominal SHERPA prediction are varied in the response matrix, and the unfolded data is compared to the nominal unfolded cross section in data to derive an uncertainty.

The unfolding uncertainty is shown together with the other sources of uncertainty on the measured cross section in the next section. The basic, hidden and signal modeling uncertainties are added in quadrature into a single unfolding uncertainty, and can be found in Figures 11.1 to 11.4. Further studies on the basic and hidden variable and signal modeling unfolding uncertainties, along with a breakdown of each of their contribution to the total unfolding uncertainty can be found in Appendix A.5.2 for all of the observables considered in this analysis.

11.5 Uncertainty on measured cross-section

The yield of the measured $W+$jets at the reconstruction level is calculated by subtracting all backgrounds from the data. To estimate the uncertainties on the unfolded cross-section, the unfolding process is repeated for each source of uncertainty. The relative uncertainty for a given type of systematic uncertainty on the measured $W+$jets cross-section in a given bin i is quantified by the absolute ratio described in Equation (11.3).

$$\delta_i = \left| \frac{N_i^{\text{nominal}} - N_i^{\text{variation}}}{N_i^{\text{nominal}}} \right| \tag{11.3}$$

where N_i^{nominal} is the nominal unfolded cross-section and $N_i^{\text{variation}}$ is the unfolded cross-section for systematics. To remove statistical fluctuations introduced in MC samples, a smoothing algorithm is applied to the relative uncertainty δ_i for each source of systematic uncertainty. In this analysis, a weighted rolling average using a Gaussian kernel is applied for each bin i, starting from the left and then repeating from the right. The mean (μ) and width (σ) of the Gaussian kernel for each bin are chosen to be $\mu = x_i$ and $\sigma = \Delta x_i * 0.8$, where x_i is the relative uncertainty value and Δx_i is the width of bin i. This parameterization is tuned to preserve the original relative shape after the smoothing procedure. The smoothing process is applied to each unfolding observable independently for the electron and muon channels, but it is not applied to the final combined results.

The relative uncertainties on the total cross-section for inclusive, inclusive 2-jets, and collinear phase-space are shown in Tables 11.3 and 11.4. These uncertainties are symmetrized by averaging both positive and negative deviations from the central values. In the electron channel, the dominant sources of uncertainty are from dijet modeling and JES/JER. Additionally, significant uncertainties arise from signal modeling introduced during the unfolding process, as well as theory uncertainties on $W \rightarrow \tau\nu$ and electroweak Wjj backgrounds. The combined "Others" category includes theory uncertainties on sub-dominant background processes such as diboson and single top, along with experimental uncertainties on b-tagging and missing transverse momentum,

all added in quadrature. This contributes at the 2-3% level for both electron and muon channels. In the inclusive and collinear regions, the total uncertainty on the integrated cross-section in the electron (muon) channel is approximately 6% (5%)

Region	Inclusive	Inc. 2 jets	Collinear	Back-to-back
Cross section [fb]	794 ± 51	699 ± 41	542 ± 39	253 ± 18
Relative systematic uncertainties %				
Statistical	0.52	0.55	0.61	0.95
lumi	0.83	0.83	0.83	0.83
JES/JER	3.29	2.88	3.69	4.02
Lepton	1.33	1.01	0.98	2.06
Top Modelling	1.00	1.05	1.21	0.63
Z+jets Modeling	0.91	0.91	1.05	0.60
Dijets Modeling	2.25	1.74	1.57	3.42
Unfolding	0.76	0.88	1.04	0.42
Signal Modeling	1.82	2.06	2.52	0.39
Pile-up	0.53	0.37	0.55	0.60
W(tau), W+2 jets EW	1.89	1.63	2.69	2.19
Others	1.77	1.80	1.94	1.68
total	6.43	5.91	7.22	7.11

Table 11.3: Systematic uncertainties on the total cross-section for **electron** channel. The relative uncertainties are shown in terms of percentage.

Region	Inclusive	Inc. 2 jets	Collinear	Back-to-back
Cross section [fb]	766 ± 40	673 ± 33	523 ± 31	243 ± 13
Relative systematic uncertainties %				
Statistical	0.51	0.54	0.60	0.92
lumi	0.83	0.83	0.83	0.83
JES/JER	2.99	2.66	3.29	3.45
Lepton	0.97	0.78	0.86	1.31
Top Modelling	0.50	0.53	0.61	0.29
Z+jets Modeling	0.21	0.22	0.24	0.26
Dijets Modeling	0.81	0.78	0.95	0.84
Unfolding	0.71	0.84	1.00	0.41
Signal Modeling	1.75	1.89	2.15	0.41
Pile-up	0.33	0.27	0.41	0.34
W(tau), W+2 jets EW	1.54	1.28	1.87	1.95
Others	1.64	1.69	1.83	1.35
Total	5.20	4.91	5.96	5.19

Table 11.4: Systematic uncertainties on the total cross-section for **muon** channel. The relative uncertainties are shown in terms of percentage.

The relative uncertainty on the measured $W+$jets differential cross-section is shown

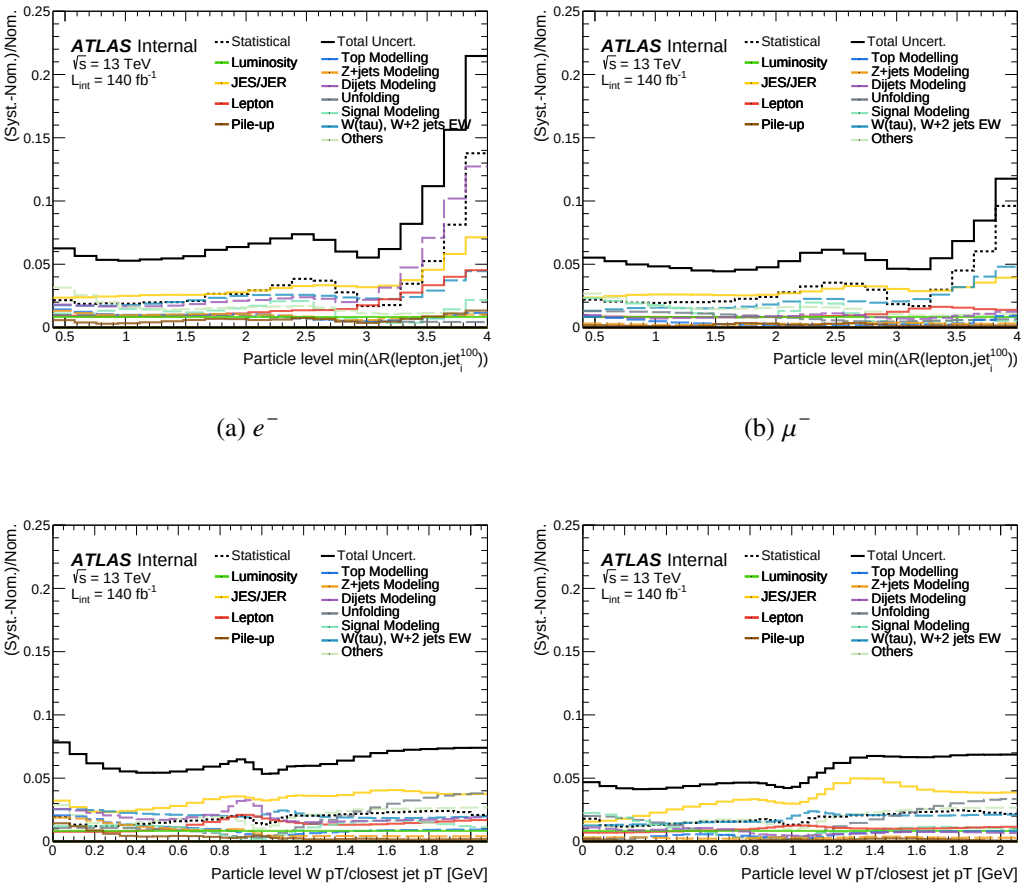
in Figures 11.1 and 11.2 for the inclusive phase-space. In the electron channel, the uncertainty is approximately 6% in the collinear phase-space, largely driven by JES/JER uncertainties. Notably, for the ratio of W-boson p_{T} to the closest jet p_{T}, the multi-jets modeling uncertainties are notably significant, particularly in the region just below 1.0. This region corresponds to the back-to-back phase-space, where multi-jet contributions are most pronounced, as observed in Figure 9.1.

For the sum of leading and sub-leading jet masses (m_{jj}) observable in the inclusive 2-jets phase-space, the dominant uncertainty arises from JES/JER uncertainties, given that this observable depends on the energies of both leading and sub-leading jets. Additionally, particularly in large m_{jj} values, uncertainties related to signal modeling in unfolding start to become significant due to increased off-diagonal contributions in the response matrix.

In the back-to-back phase-space, multi-jets modeling uncertainties emerge as the primary source of uncertainty, given the notable contribution of multi-jets background in this region. For muons, uncertainties are comparable to those in the electron channel in the collinear phase-space and primarily dominated by JES/JER uncertainties. Given the generally lower multi-jets contribution in the muon channel, uncertainties in the back-to-back regions are comparatively smaller than in the electron channel, largely driven by statistical uncertainties.

The uncertainties for collinear unfolding variables are shown in Figures 11.3 and 11.4. These uncertainties typically range around 10-15%, with JES/JER being the dominant

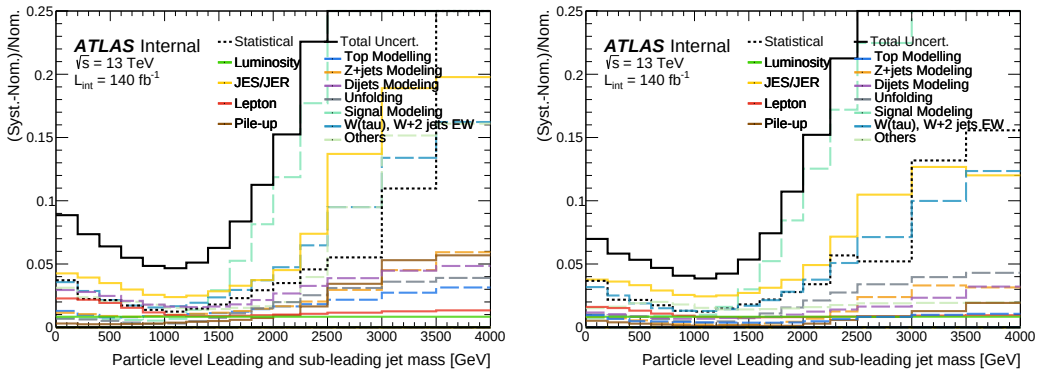
source. In bins where the response matrix with significant off-diagonal contributions, uncertainties related to signal modeling start to increase and contribute noticeably.



(a) e^- (b) μ^-

(c) e^- (d) μ^-

Figure 11.1: Relative systematic uncertainties on unfolded $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ and the ratio of the W-boson candidate over the closest jet p_{T} distributions in electron and muon channel in the **inclusive** phase-space.



(a) e^- (b) μ^-

Figure 11.2: Relative systematic uncertainties on unfolded m_{jj} distribution of electron and muon channel in the **inclusive 2-jets** phase-space.

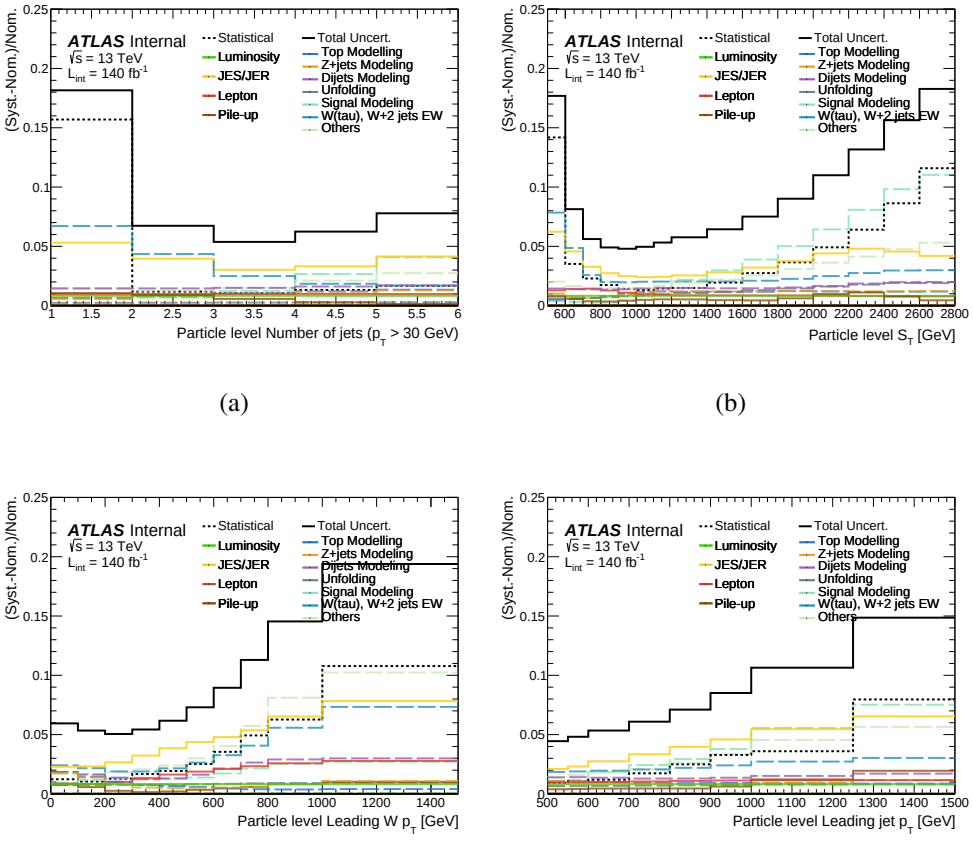
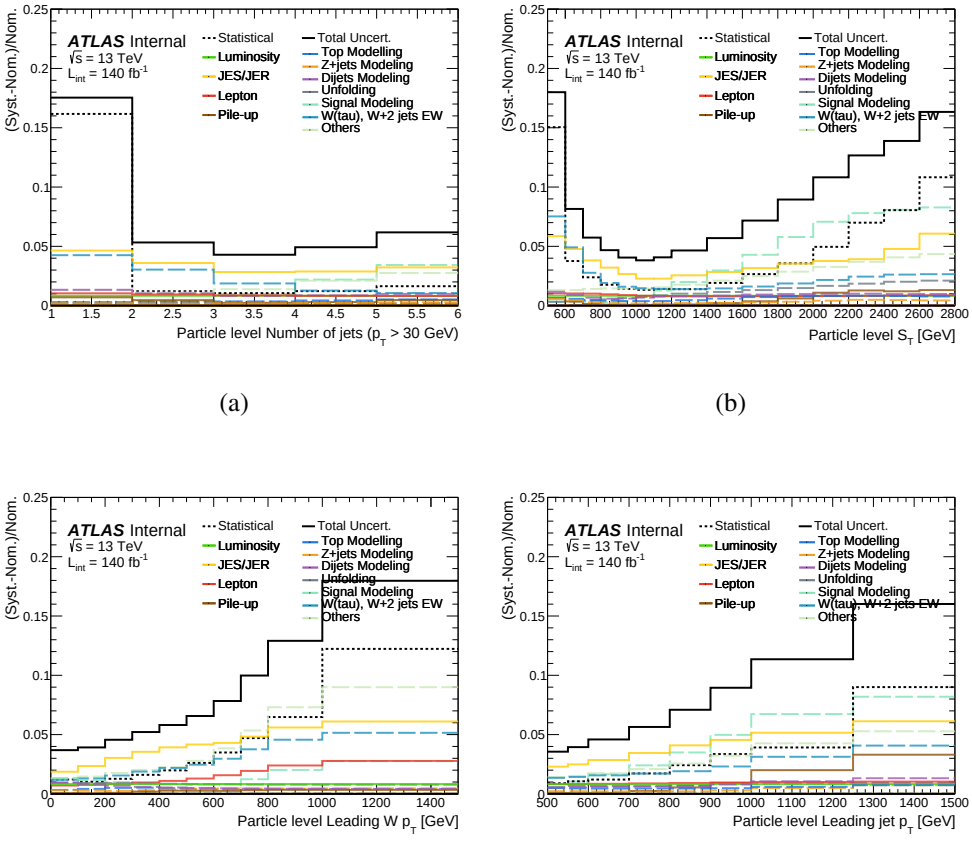


Figure 11.3: Relative systematic uncertainties on unfolded observables of **electron** channel in the **collinear** phase-space.



(a) (b)

(c) (d)

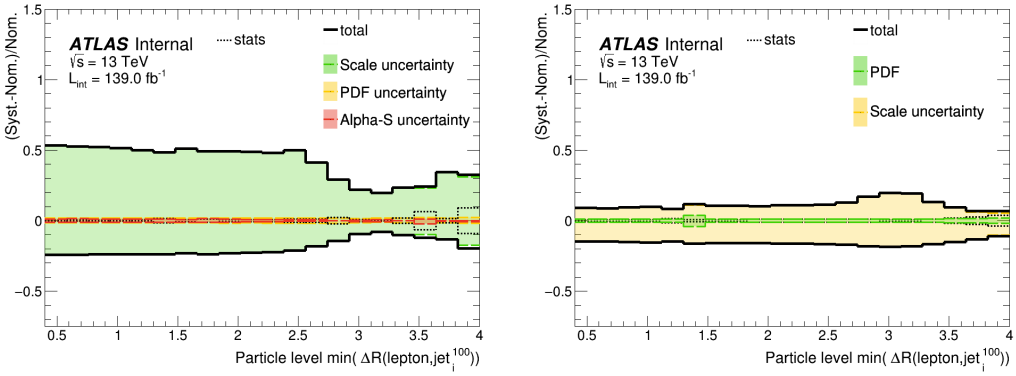
Figure 11.4: Relative systematic uncertainties on unfolded observables of **muon** channel in the **collinear** phase-space.

11.6 Uncertainties on W+ jets theoretical prediction

This section quantifies the systematic uncertainties associated with the theoretical prediction of W+ jets. These uncertainties are intrinsic to each theory prediction and will be compared against the measured data in subsequent sections. A comparison of theoretical systematic uncertainties arising from different generators is provided.

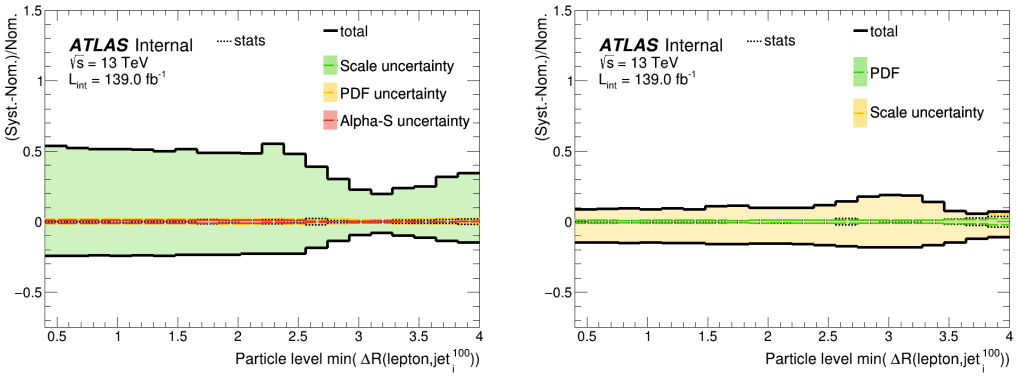
The prediction derived from MadGraph5_aMC@NLO +Pythia8, using FxFx merging, shows the smallest overall uncertainties. Conversely, the Sherpa 2.2.11 prediction demonstrates larger overall uncertainties compared to MadGraph5_aMC@NLO +FxFx. This disparity is largely attributed to the leading-order dependency extending beyond 2-jet configurations. Given that our explored phase-space is predominantly dominated by these leading-order contributions, it results in considerable uncertainties.

The relative uncertainties pertaining to observables in both the inclusive and inclusive 2-jets regions are illustrated in Figures 11.5 and 11.6 and Figures 11.7 and 11.8. Furthermore, relative systematic uncertainties for observables in the collinear regions are shown in Figures 11.9 and 11.10.



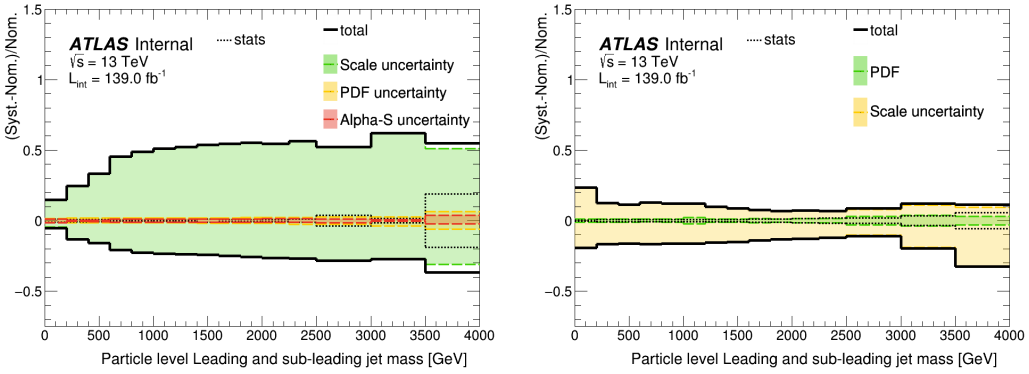
(a) Sherpa 2.2.11 (b) MadGraph5_aMC@NLO +Pythia8 with FxFx merging

Figure 11.5: Relative systematic uncertainties on theory prediction of $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ for **electron** channel in the **inclusive** phase-space.



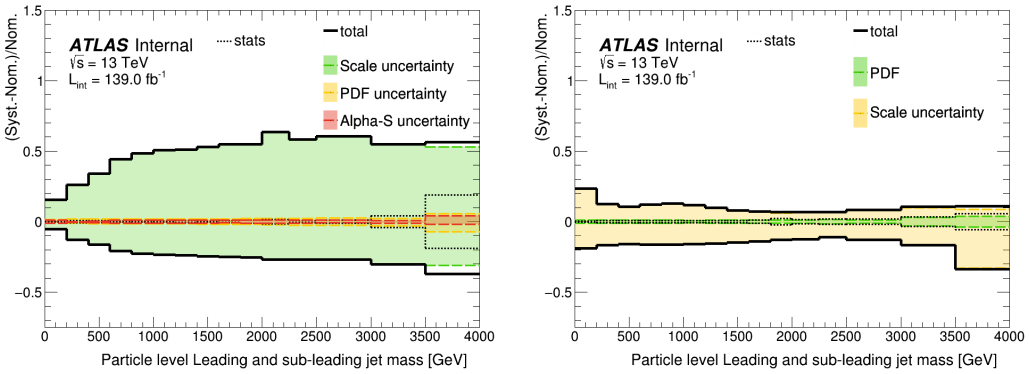
(a) Sherpa 2.2.11 (b) MadGraph5_aMC@NLO +Pythia8 with FxFx merging

Figure 11.6: Relative systematic uncertainties on theory prediction of $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ for **muon** channel in the **inclusive** phase-space.



(a) SHERPA 2.2.11 (b) MADGRAPH5_AMC@NLO +PYTHIA8 with FxFx merging

Figure 11.7: Relative systematic uncertainties on theory prediction of m_{jj} for **electron** channel in the **inclusive 2-jets** phase-space.



(a) SHERPA 2.2.11 (b) MADGRAPH5_AMC@NLO +PYTHIA8 with FxFx merging

Figure 11.8: Relative systematic uncertainties on theory prediction of m_{jj} for **muon** channel in the **inclusive 2-jets** phase-space.

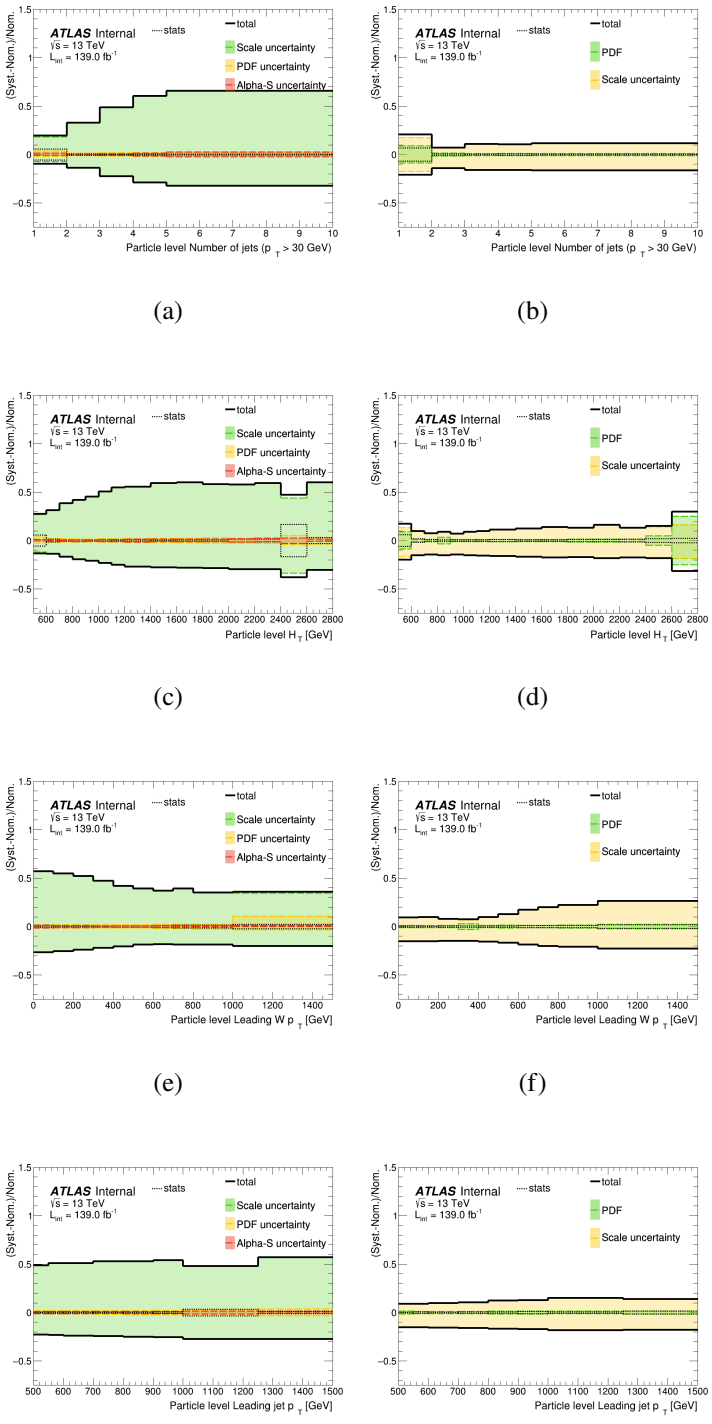
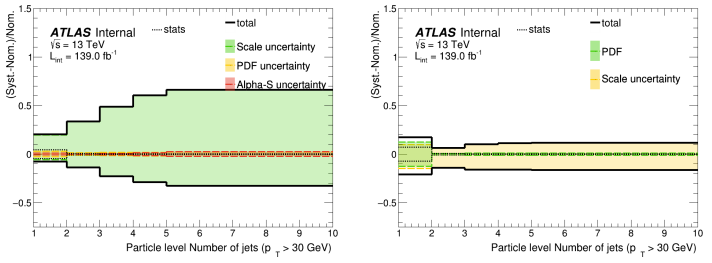


Figure 11.9: Relative systematic uncertainties on theory prediction of observables in the **electron** channel in the **collinear** phase-space.



(a)

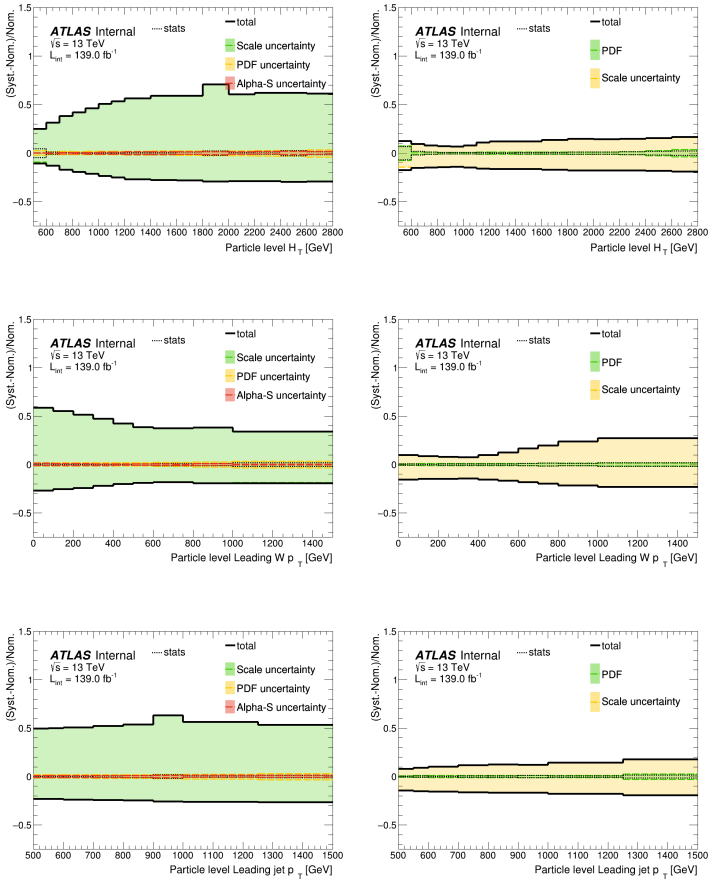


Figure 11.10: Relative systematic uncertainties on theory prediction of observables in the **muon** channel in the **collinear** phase-space.

Chapter 12

Cross-section Results

This chapter presents the final measured particle-level cross-section values both separately and averaged over both the electron and muon selections, for both differential and inclusive measurements.

12.1 Comparison and combining the electron and muon channel results

For the final results, the cross sections are combined over the electron and muon channels. The measurements are combined using a χ^2 likelihood fit with the `Combiner` tool [98]. A key caveat is that `Combiner` does not support asymmetric uncertainties, so all uncertainties are symmetrized, and the tool reports only symmetrized averaged systematic uncertainties. Symmetrized systematics are correlated between the electron and muon channels unless they are specific to lepton reconstruction. For instance, the jet energy scale is correlated, but the muon ID and electron ID remain uncorrelated. The statistical uncertainties described in Section 11.1 are included in `Combiner`.

To compare the electron and muon channels, each channel is compared against the combined cross section. The uncertainties displayed for the individual electron and muon measurements only contain uncorrelated systematic uncertainties between the two channels. In contrast, the uncertainty band on the combined cross-section includes the total measurement uncertainty after the combination, summed in quadrature with the statistical uncertainty. The data points show the combined measurement in data along

with its statistical uncertainty. A comparison of the inclusive cross section measurement in the electron and muon channels, plotted with the combined result, can be found in Figure 12.1. The combined result is in good agreement with both the electron and muon channels across all regions.

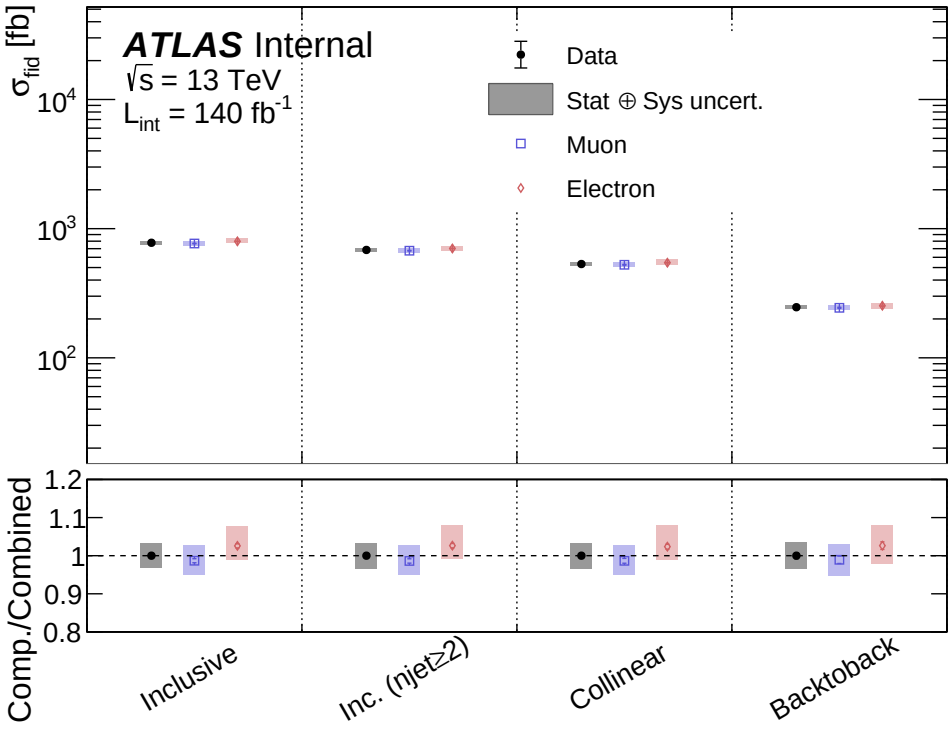


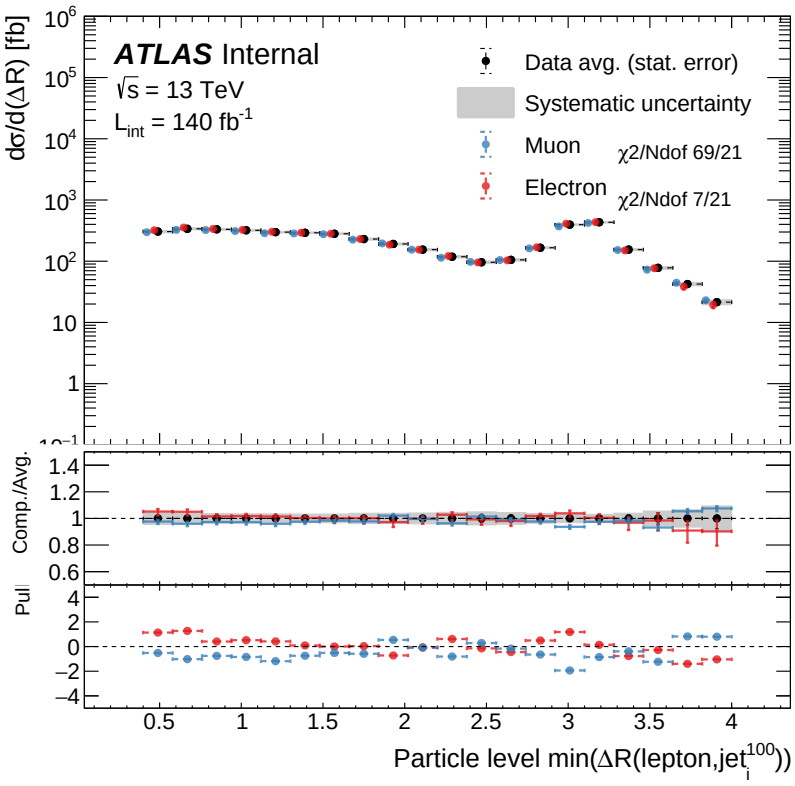
Figure 12.1: Integrated cross-section in each of the phase-space under study in this analysis: inclusive, inclusive with a 2-jet selection, back-to-back, and collinear. The black dot corresponds to the combined unfolded cross-section of the electron and muon channels. The electron and muon channels are super-imposed onto the plot, and only contain uncorrelated systematic uncertainties between both channels in their uncertainty bands. The statistical band is shown using error bars while the total systematic plus statistical uncertainty band is shown using a black filled area.

For the differential distributions, we also compute the pull in each bin to assess the bin-by-bin compatibility of the two measurements. The pull is defined as:

$$\text{pull}_{\ell,i} = \frac{\mu_{\ell,i} - \bar{\mu}_i}{\bar{\sigma}_i} \qquad i \equiv \text{bin}, \ell \in [\text{electron}, \text{muon}]. \tag{12.1}$$

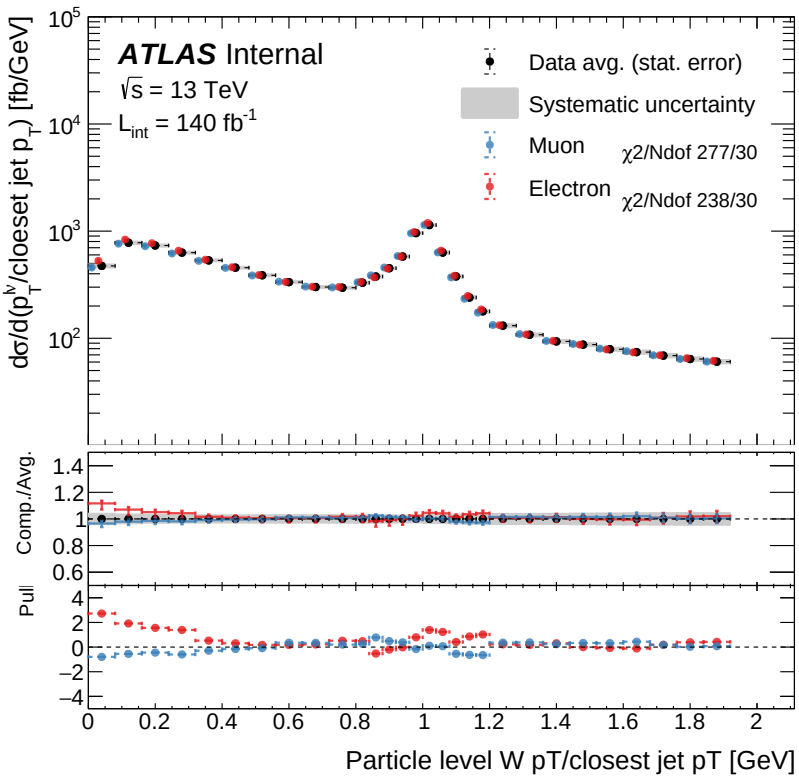
where $\mu_{\ell,i}$ represents the cross section results for the electron and muon channels, respectively. The term $\bar{\mu}_i$ denotes the combined results, and $\bar{\sigma}_i$ is the total uncertainty of the combined results.

For most distributions shown in Figures 12.2 to 12.5, the combined cross-section agrees very well with the individual electron and muon cross-sections, which indicating a high level of compatibility between the measurements. However, there are a few exceptions, particularly in the low value range for the ratio of p_{T}^{W} to the closest jet p_{T}, where the pull on the electron channel is around 2.5σ. A similar level of pull on the electron channel in the collinear phase-space is observed in the p_{T}^{W} around 200 GeV. This is correlated with the small values of the ratio of p_{T}^{W} to the closest jet p_{T}, where this phase-space is mostly dominated by soft W boson emission, carrying a small fraction of the momentum from its parent jet. This large pull on the electron can be explained by the muon measurement that exhibits significantly improved precision over the electron measurement, as shown in Figure 11.1 (c) and (d).



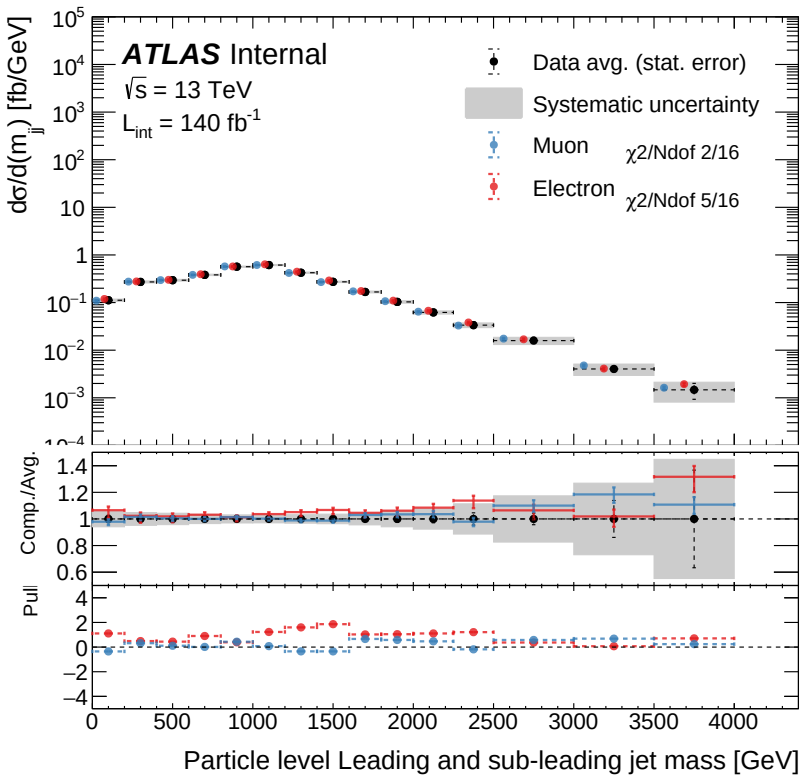
(a)

Figure 12.2: Differential cross-section as a function of the minimum angular separation between the lepton and any jet with transverse momentum greater than 100 GeV in the inclusive selection. Statistical uncertainties on the measured cross section are shown on the black data points; the total systematic uncertainty on the measured cross section summed in quadrature with the statistical uncertainty is shown by the gray band; the individual electron and muon measurements only shown uncorrelated systematic uncertainties between the two channels. Errors on the theory prediction include only sources of theoretical uncertainties.



(a)

Figure 12.3: Differential cross-section as a function of the W-boson p_{T} over the p_{T} of the closest jet to the lepton. Statistical uncertainties on the measured cross section are shown on the black data points; the total systematic uncertainty on the measured cross section summed in quadrature with the statistical uncertainty is shown by the gray band; the individual electron and muon measurements only shown uncorrelated systematic uncertainties between the two channels. Errors on the theory prediction include only sources of theoretical uncertainties.



(a)

Figure 12.4: Differential cross-section as a function of the invariant mass of the two leading jets in the inclusive 2-jet selection. Statistical uncertainties on the measured cross section are shown on the black data points; the total systematic uncertainty on the measured cross section summed in quadrature with the statistical uncertainty is shown by the gray band; the individual electron and muon measurements only shown uncorrelated systematic uncertainties between the two channels. Errors on the theory prediction include only sources of theoretical uncertainties.

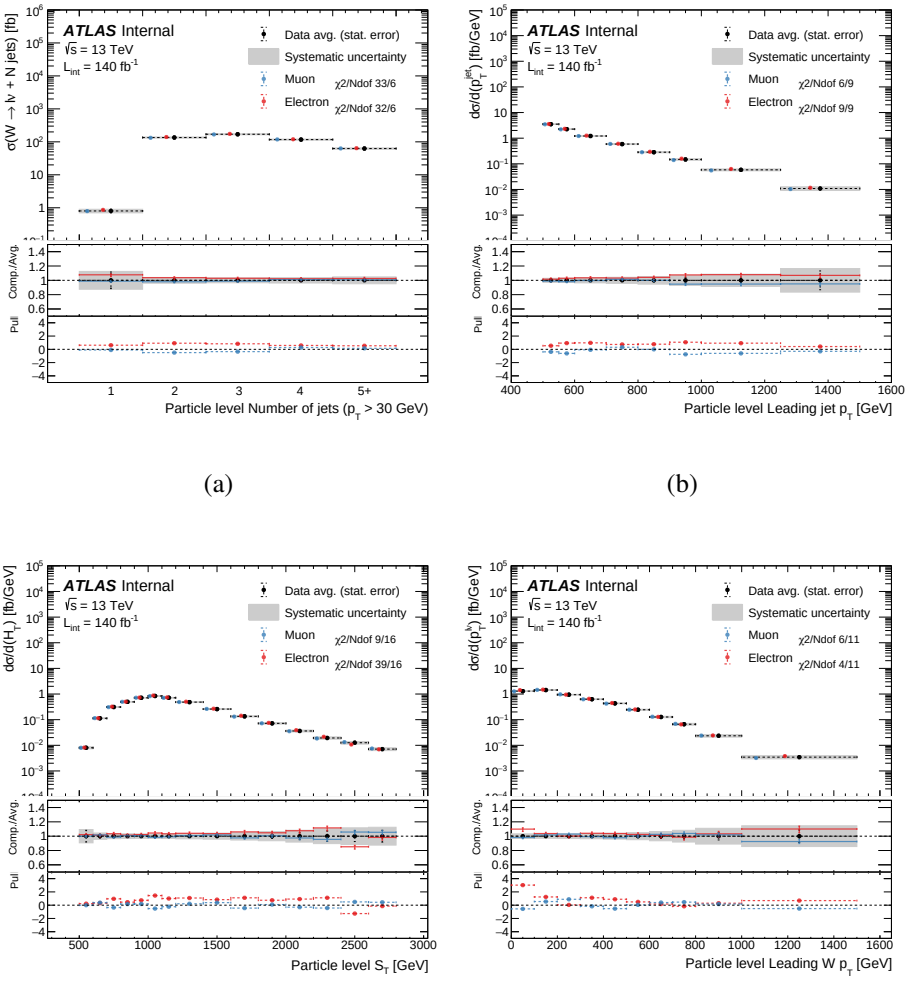


Figure 12.5: Differential cross-section as a function of various observables in the **collinear** phase space. Statistical uncertainties on the measured cross section are shown on the black data points; the total systematic uncertainty on the measured cross section summed in quadrature with the statistical uncertainty is shown by the gray band; the individual electron and muon measurements only shown uncorrelated systematic uncertainties between the two channels. Errors on the theory prediction include only sources of theoretical uncertainties.

12.2 Combined results and comparison with models

The integrated cross-section from the combined electron and muon channels are presented in this section and compared to model expectations. The results are shown for the inclusive, inclusive-2j, collinear, and back-to-back regions in Figure 12.6. In general, all theoretical predictions agree with the measured data within uncertainties across all regions. In these inclusive regions, the electroweak corrections tend to have a small impact on the agreement with the data. The central value of the MadGraph5_aMC@NLO +Pythia8 prediction with FxFx merging slightly over-predicts the measured data cross sections, but it remains in agreement with the measured data within its corresponding theoretical uncertainties.

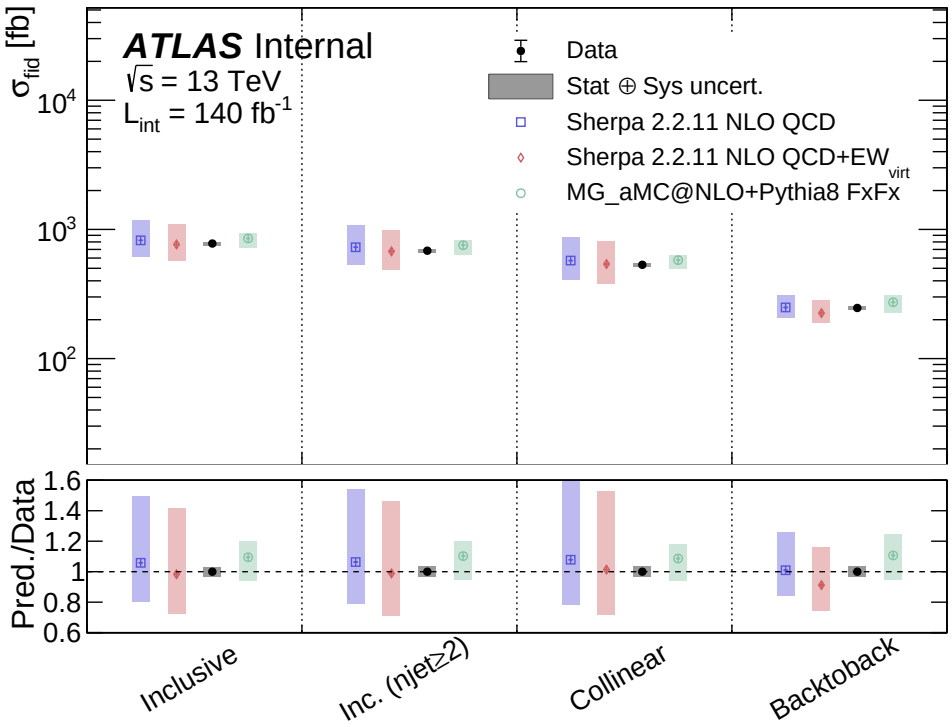


Figure 12.6: Integrated cross-section from combined electron and muon channels in each of the phase-space under study in this analysis: inclusive, inclusive with a 2-jet selection, and collinear. The black dot corresponds to the combined unfolded cross-section of the electron and muon channels. All the non-black colors represent the truth prediction from different MC generators (for the muon channel only here). The statistical band is shown using error bars while the total systematic plus statistical uncertainty band is shown using a gray filled area.

12.3 Differential measurement

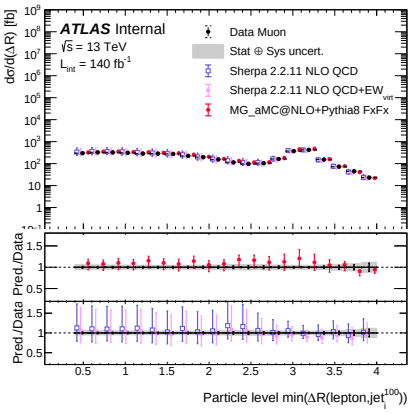
Using the approach discussed in the preceding chapters, the measured W+jets events are first unfolded independently for the electron and muon channels, and then the unfolded results are combined across channels. The unfolded results are compared to the particle-level predictions in the inclusive (Figures 12.7 to 12.9) and collinear (Figures 12.10 to 12.13) phase-spaces for the two lepton channels, as well as their combined results.

The cross sections are unfolded differentially in terms of binned distributions for each observable as shown below. The black data points represent the measurements in data along with their statistical uncertainties, which account for both data statistics and the MC subtracted from the data before unfolding. The grey uncertainty band indicates the total cross-section measurement uncertainty.

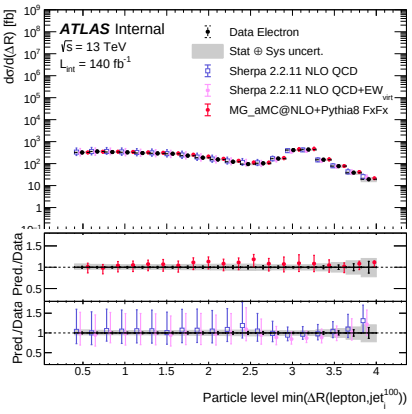
Theoretical predictions are overlaid on the measured cross sections. These predictions include the newly developed Sherpa 2.2.11 and MadGraph5_aMC@NLO +Pythia8 with FxFx merging configurations. Both of these predictions model the data well, except for the extreme m_{jj} tails. They represent a significant improvement over the legacy Sherpa 2.2.1 prediction, which over-predicts the cross section in this region of phase space by over 30%. Furthermore, the agreement with the data is improved for important variables such as $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$ and the leading jet p_{T}, where the previously rising disagreement between the Sherpa 2.2.1 prediction and the data is reduced.

In conclusion, both Sherpa 2.2.11 and MadGraph5_aMC@NLO +Pythia8 with

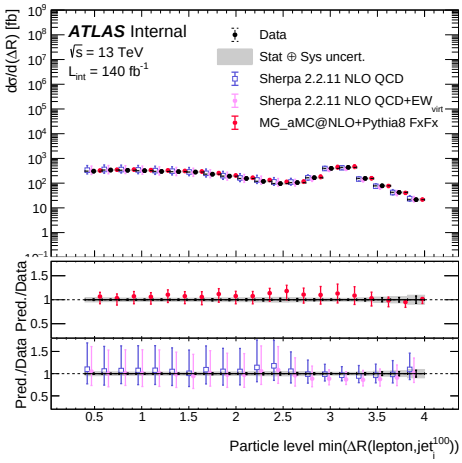
FxFx merging configurations provide a good description of the differential distributions for kinematic observables when compared to data. The SHERPA 2.2.11 prediction performs slightly better for jet-related observables, with the exception of the extremely high tail of the m_{jj} distribution. The electroweak correction in SHERPA 2.2.11 shows slightly better performance for the W boson p_{T} distribution. However, all theoretical predictions shown in this analysis model the data very well, and the differences between these theoretical models are within systematic uncertainties.



(a) Muon

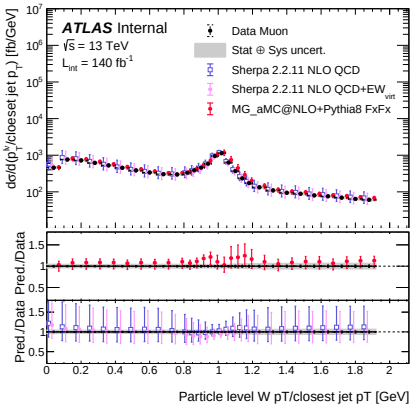


(b) Electron

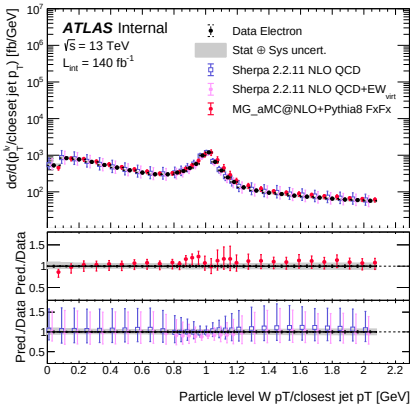


(c) Combined

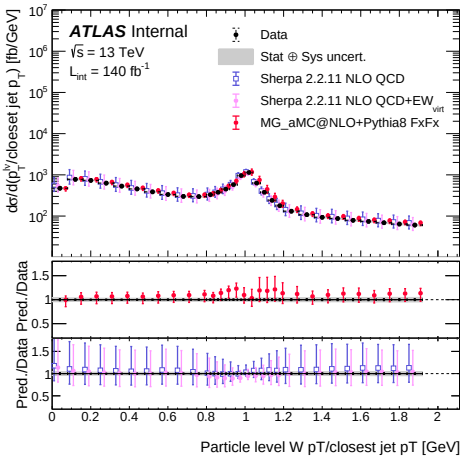
Figure 12.7: Differential cross-section as a function of the minimum angular separation between the lepton and any jet with transverse momentum greater than 100 GeV in the **inclusive** phase-space. Statistical uncertainties on the measured cross section are shown on the black data points, while the grey band shows the systematic and statistical uncertainties added in quadrature. Errors on the theory prediction include theoretical uncertainties. Overflow bins are included and summed together with the last visible bin.



(a) Muon

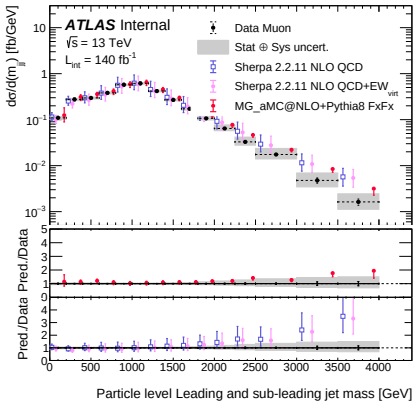


(b) Electron

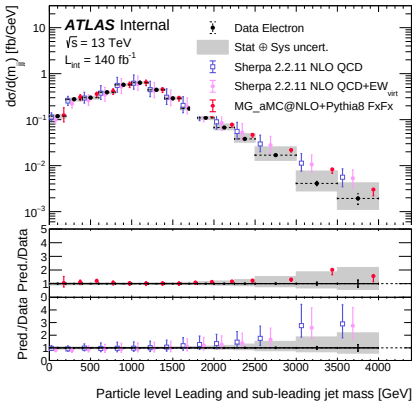


(c) Combined

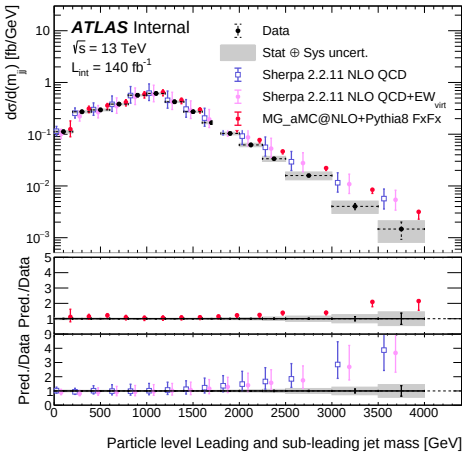
Figure 12.8: Differential cross-section as a function of the ratio of W-p_{T} and $p_{\mathrm{T}}^{\mathrm{jet}}$ in the **inclusive** phase-space. Statistical uncertainties on the measured cross section are shown on the black data points, while the grey band shows the systematic and statistical uncertainties added in quadrature. Errors on the theory prediction include theoretical uncertainties.



(a) Muon

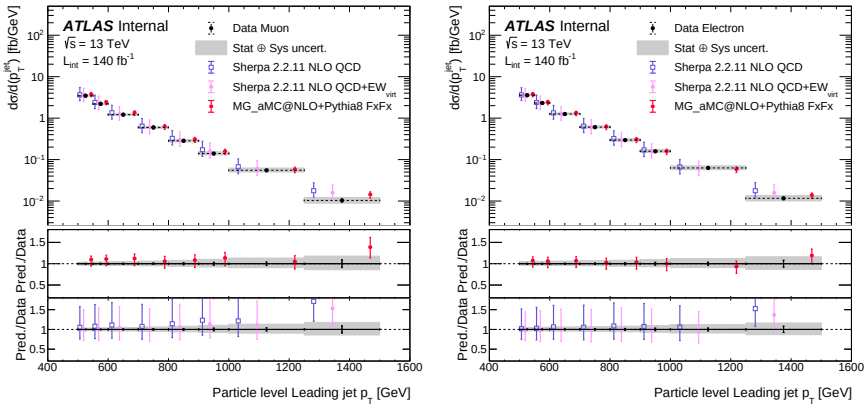


(b) Electron

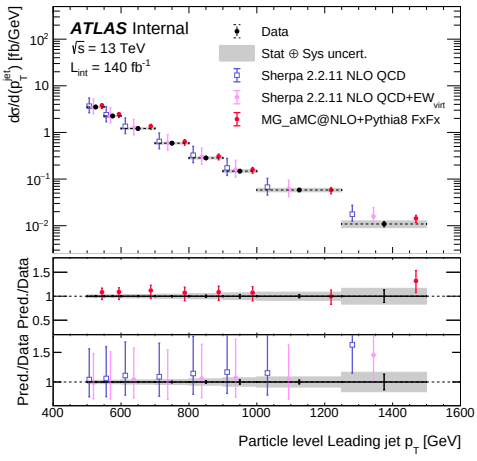


(c) Combined

Figure 12.9: Differential cross-section as a function of the invariant jet mass m_{jj} in the **inclusive, 2-jet** phase-space. Statistical uncertainties on the measured cross section are shown on the black data points, while the grey band shows the systematic and statistical uncertainties added in quadrature. Errors on the theory prediction include theoretical uncertainties.

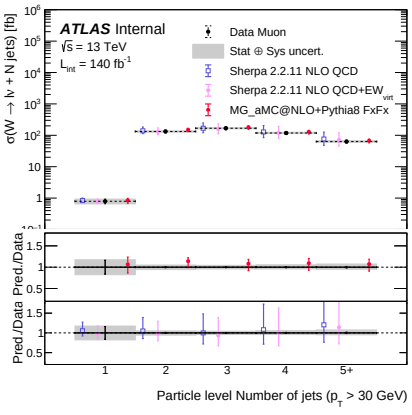


(a) Muon (b) Electron

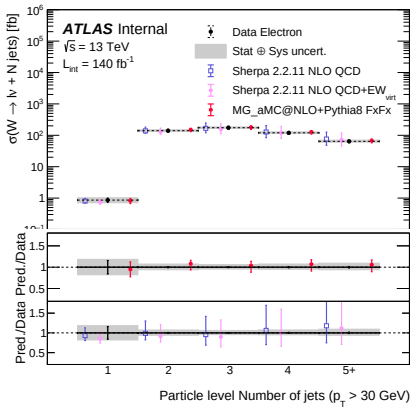


(c) Combined

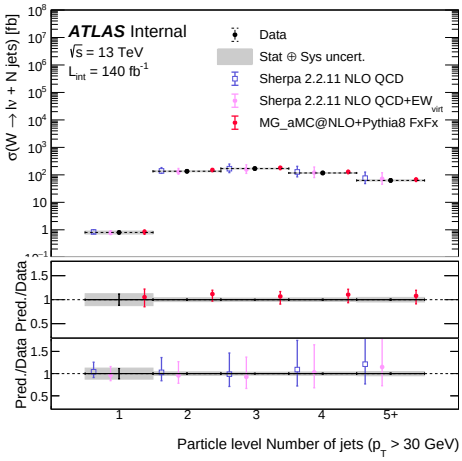
Figure 12.10: Differential cross-section as a function of the leading $p_{\mathrm{T}}^{\mathrm{jet}}$ in the **collinear** phase-space. Statistical uncertainties on the measured cross section are shown on the black data points, while the grey band shows the systematic and statistical uncertainties added in quadrature. Errors on the theory prediction include theoretical uncertainties.



(a) Muon

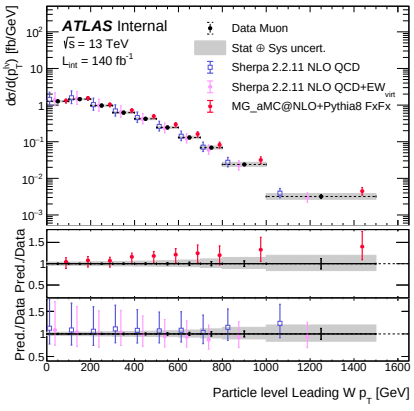


(b) Electron

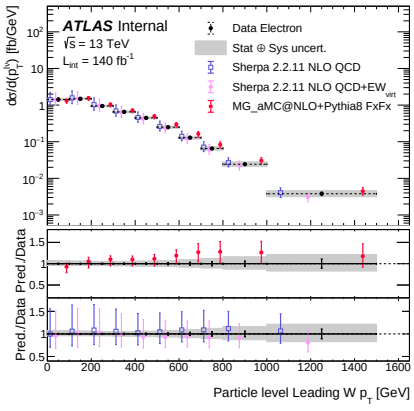


(c) Combined

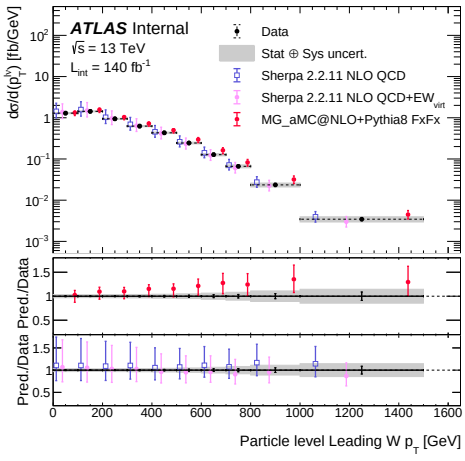
Figure 12.11: Differential cross-section as a function of the jet multiplicity in the **collinear** phase-space. Statistical uncertainties on the measured cross section are shown on the black data points, while the grey band shows the systematic and statistical uncertainties added in quadrature. Errors on the theory prediction include theoretical uncertainties.



(a) Muon

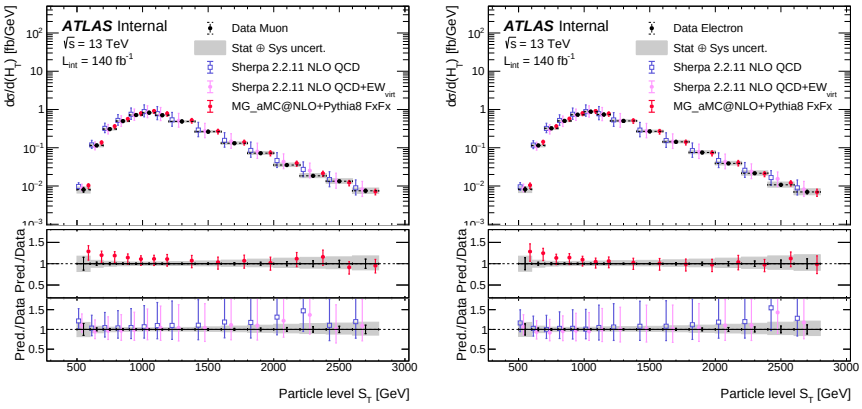


(b) Electron

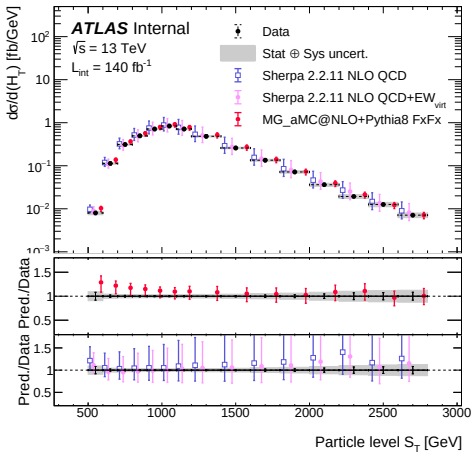


(c) Combined

Figure 12.12: Differential cross-section as a function of the leading W-p_{T} in the **collinear** phase-space. Statistical uncertainties on the measured cross section are shown on the black data points, while the grey band shows the systematic and statistical uncertainties added in quadrature. Errors on the theory prediction include theoretical uncertainties.



(a) Muon (b) Electron



(c) Combined

Figure 12.13: Differential cross-section as a function of the S_{T} in the **collinear** phase-space. Statistical uncertainties on the measured cross section are shown on the black data points, while the grey band shows the systematic and statistical uncertainties added in quadrature. Errors on the theory prediction include theoretical uncertainties.

Chapter 13

Conclusion

This thesis presents a measurement of collinear W-boson radiation from high-momentum jets using the full Run-2 dataset collected by the ATLAS collaboration at a center-of-mass energy of $\sqrt{s}$ = 13 TeV, with a total integrated luminosity of 140 fb^{-1}.

The W+jets signal is defined by a W-boson radiated from a high-momentum jet, identified through its leptonic decays $W \to \ell\nu$, where ℓ is either an electron or a muon. The collinear phase-space contains events with a small angular separation between the jet and lepton, often involving numerous additional QCD emissions. Accurately modeling this phase-space requires sophisticated multi-jet merged setups, which makes the modeling of the collinear W+jets process challenging and highlights the importance of experimental constraints on production rates and kinematic distributions.

In this study, over 86,000 W-boson candidates were measured across various differential distributions using both electron and muon channels. The precision of the measured cross-sections was within 10-20% of the measured values, with the jet energy scale and unfolding bias as dominant sources of uncertainty. However, these experimental uncertainties were smaller compared to the dominant theoretical uncertainties on signal predictions. The measured cross-sections were used to highlight significant scale changes in state-of-the-art theoretical predictions, demonstrating an improved description of the data compared to legacy setups.

Looking forward, this measurement has shown that modeling this phase-space is extremely challenging. However, its modeling remains critical for the SM and BSM ATLAS physics programs. Advances in the Sherpa event generator, particularly in

handling unordered parton-shower histories, have enhanced the data description, though theoretical uncertainties remain large. These results clearly indicate a strong sensitivity to the renormalization scale treatment in theoretical predictions. Further theoretical efforts are essential to improve the theoretical precision of estimates in these phase-spaces.

Appendix A

Appendix

A.1 $t\bar{t}$ MC studies

A.1.1 $t\bar{t}$ CR b-jets multiplicity comparisons: ≥ 2 vs ≥ 1 b-jets

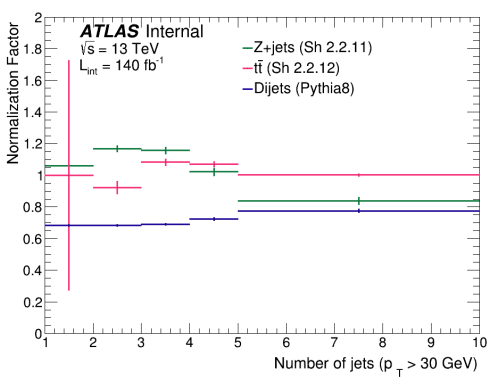
In this section, we explore the possibility of relaxing the requirement on the number of b-jets from ≥ 2 vs ≥ 1 b-jets in the $t\bar{t}$ control region. As shown in Table A.1 the signal W+jets contribution to total MC processes increases from from 3.8%(3.1%) to 11.9%(10.2%) in the muon (electron) channel when relaxing the b-jet multiplicity requirement. Furthermore, overall purity of the $t\bar{t}$ events *decreases* from 82.3%(82.9%) to 68.1%(70.5%) in the muon (electron) channel.

Process	≥ 1 b-jet (Muon)	≥ 1 b-jet (Electron)	≥ 2 b-jet (Muon)	≥ 2 b-jet (Electron)
$t\bar{t}$	42468.83 ± 161.48	52299.49 ± 191.38	14313.22 ± 88.64	17146.09 ± 105.29
Dijets	2803.85 ± 88.21	3101.57 ± 85.24	211.88 ± 27.35	370.86 ± 36.55
Single top	7783.79 ± 29.24	8925.42 ± 33.93	2007.58 ± 14.52	2271.85 ± 16.09
Z+jets (Sherpa 2.2.11)	490.93 ± 4.95	699.25 ± 4.24	51.66 ± 1.06	76.12 ± 1.33
Diboson	428.11 ± 7.81	482.53 ± 8.45	64.92 ± 2.25	70.47 ± 2.19
V+γ+jets	399.22 ± 4.82	427.50 ± 5.70	36.50 ± 1.74	41.45 ± 1.81
$\tau\nu$+jets	451.71 ± 11.71	551.03 ± 13.69	41.65 ± 3.21	54.45 ± 3.48
W+2j	142.21 ± 2.42	163.42 ± 2.67	5.86 ± 0.48	6.86 ± 0.54
W+jets (Sherpa 2.2.11)	7414.02 ± 45.78	7543.99 ± 74.82	659.26 ± 10.96	646.28 ± 12.02
Total SM	62382.66 ± 192.50	74194.20 ± 225.74	17392.51 ± 94.64	20684.43 ± 113.35
Observed data	62803.00 ± 250.61	70606.00 ± 265.72	16879.00 ± 129.92	18730.00 ± 136.86
$t\bar{t}$ purity	68.1 ± 0.3%	70.5 ± 0.3%	82.3 ± 0.7%	82.9 ± 0.7%

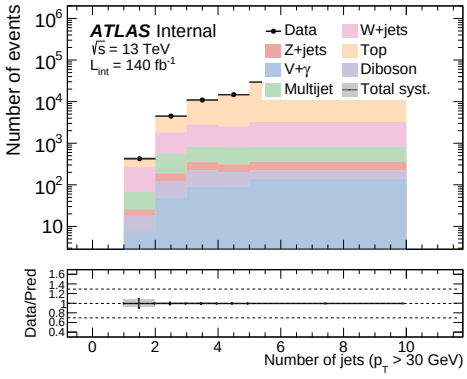
Table A.1: Yield comparison of MC processes in the ***$t\bar{t}$* control region** with ≥ 2 and ≥ 1 b-jets selections.

Aside from the effect on the purity and background considerations mentioned above, relaxing the b-jet multiplicity requirement does not statistically populate the single jet bin enough to derive a meaningful normalization factor. This is demonstrated in Figure A.1, where the simultaneous fit to all of the control regions (see Section 8) was performed with the b-jet multiplicity relaxed to ≥ 1 in the $t\bar{t}$ control region. As can be seen, the statistical uncertainty on extracted normalization factor for the $t\bar{t}$ background in the 1-jet bin is nearly 50% in both the electron and muon channels. The large statistical uncertainty is expected, given the small contribution of $t\bar{t}$ and large contamination of the

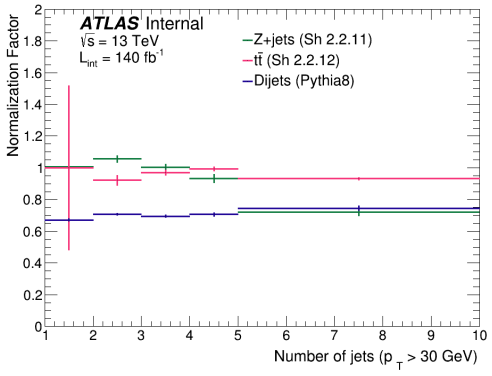
W+ jets process in this bin. All these points considered, we see that the central value of the normalization factor is consistent with one, which is consistent with the value assumed for the 1-jet bin in this analysis. No additional uncertainties are considered on the 1-jet bin from this alternative $t\bar{t}$ control region.



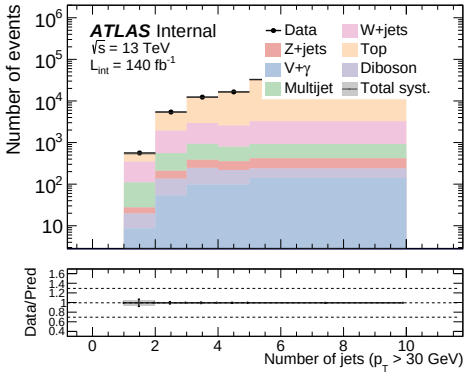
(a) Muon



(b) Muon



(c) Electron



(d) Electron

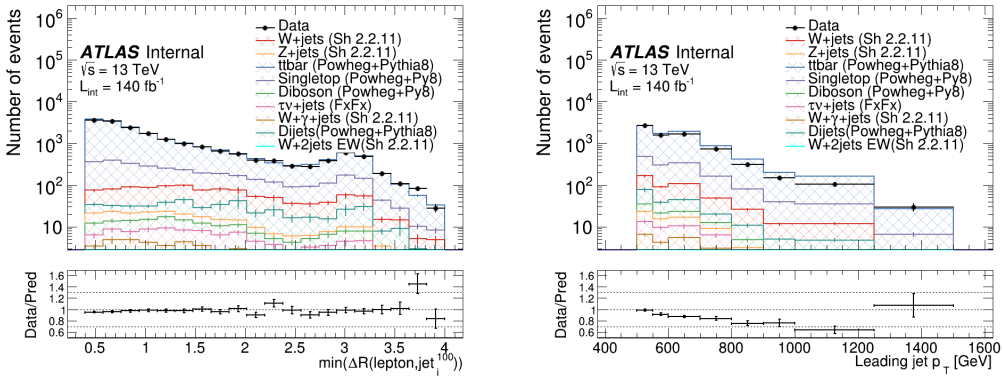
Figure A.1: Extracted normalization factors and distributions of the post-fit jet multiplicity distributions in the $t\bar{t}$ control region when relaxing the b-jet requirement to ≥ 1. Muons are shown on the top, and electrons on the bottom. Uncertainties are statistical only.

A.1.2 POWHEG +PYTHIA $t\bar{t}$ model

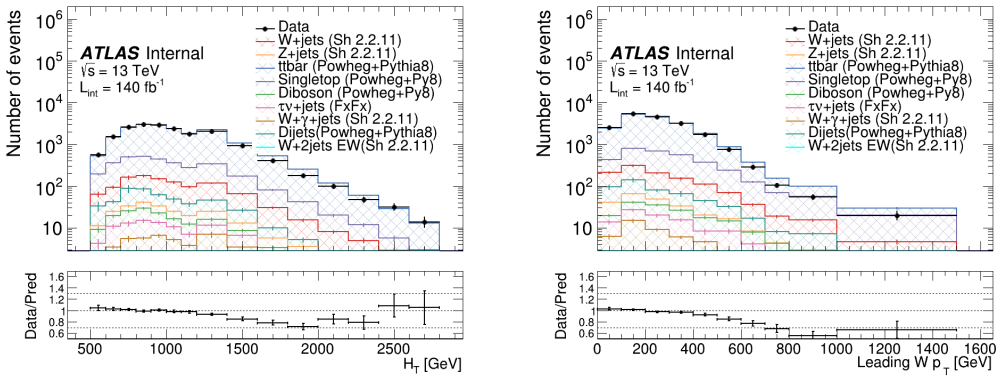
In this section, the modeling of $t\bar{t}$ with POWHEG +PYTHIA is assessed and compared against the default SHERPA 2.2.12 prediction.

A.1.2.1 Comparison in control region

The control region fit is re-performed using the POWHEG +PYTHIA $t\bar{t}$ model, and compared against the data in the $t\bar{t}$ control region as shown in Figures A.2 and A.5. The same semi-data driven method described in Section 8.2 is used to obtain the normalization factor for this alternative $t\bar{t}$ model. In general, the modeling of POWHEG +PYTHIA is worse than the nominal SHERPA 2.2.12 $t\bar{t}$ samples, likely due to the fact that the latter has leading-order accuracy in our final state with at least one hard jet. Mis-modeling in the leading jet and S_{T} distributions can be seen in the POWHEG +PYTHIA sample in the plots below. As such, we choose the SHERPA 2.2.12 $t\bar{t}$ as our nominal prediction in our analysis, and do not consider the POWHEG +PYTHIA model as part of our systematic uncertainties on the $t\bar{t}$ background.

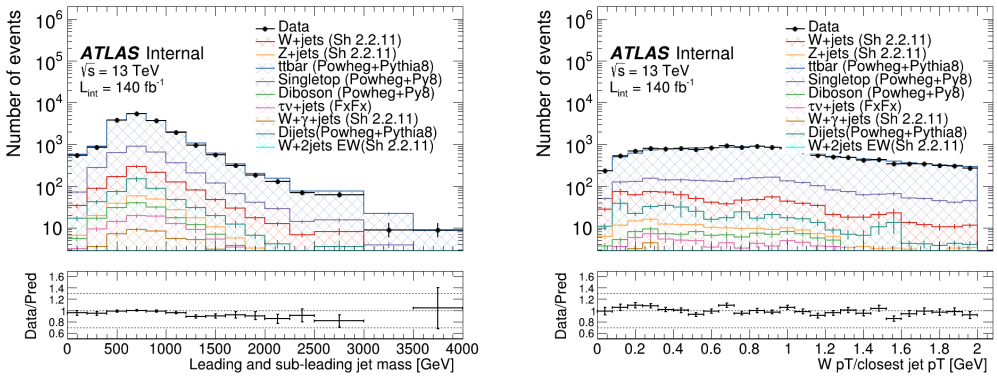


(a) (b)



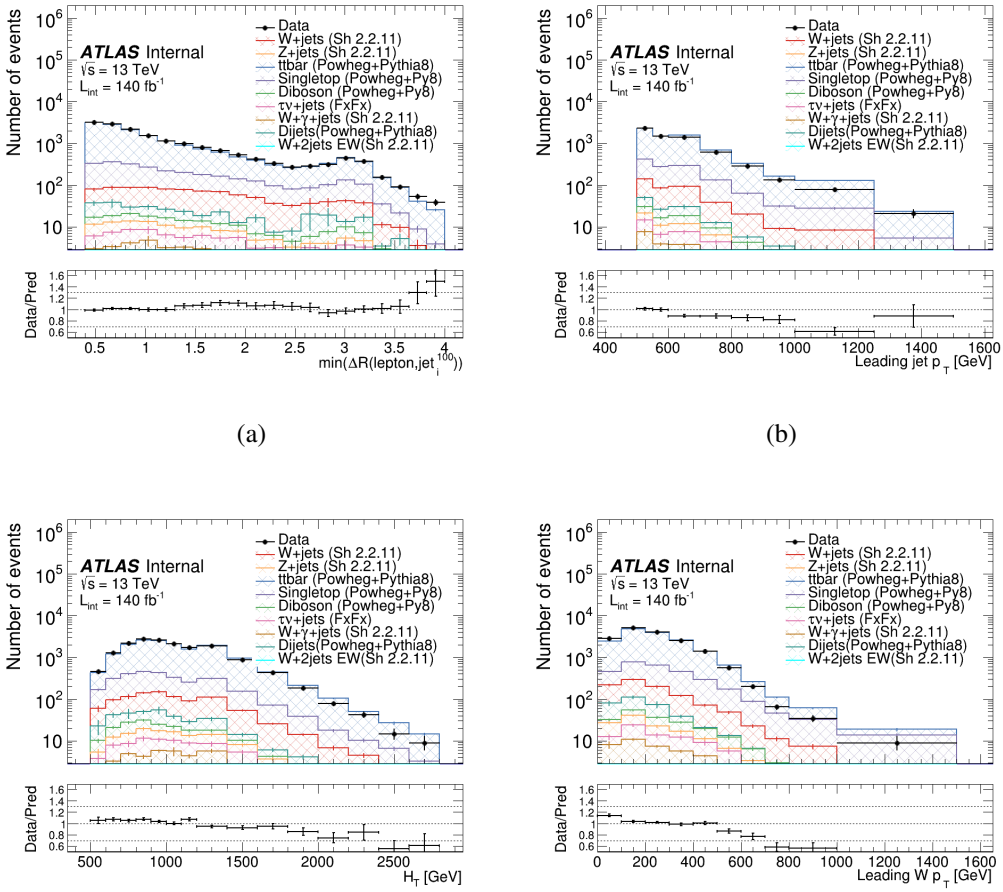
(c) (d)

Figure A.2: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **electron**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. Uncertainties are statistical only.



(a) (b)

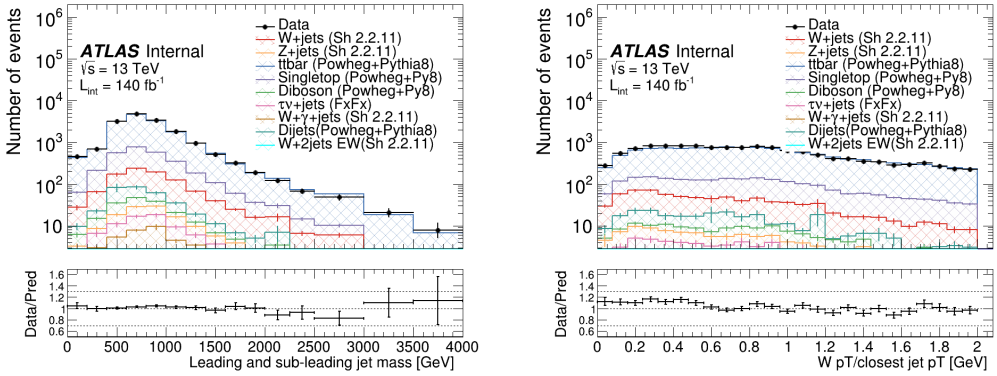
Figure A.3: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **electron**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. Uncertainties are statistical only.



(a) (b)

(c) (d)

Figure A.4: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **muon**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. Uncertainties are statistical only.

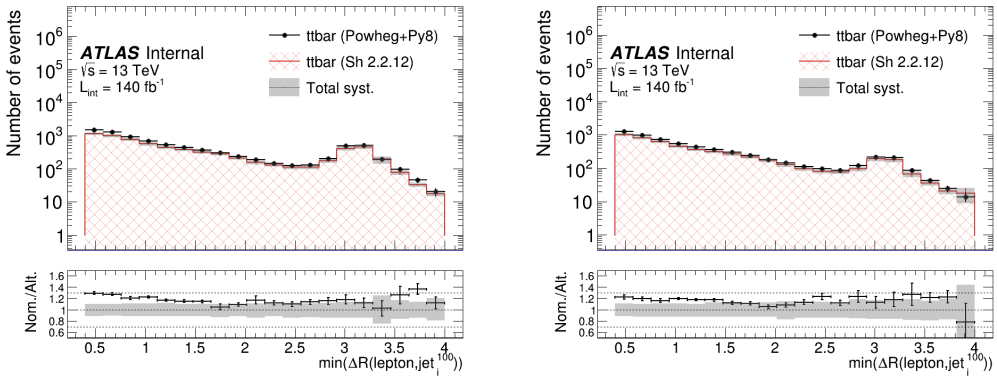


(a) (b)

Figure A.5: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **muon**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. Uncertainties are statistical only.

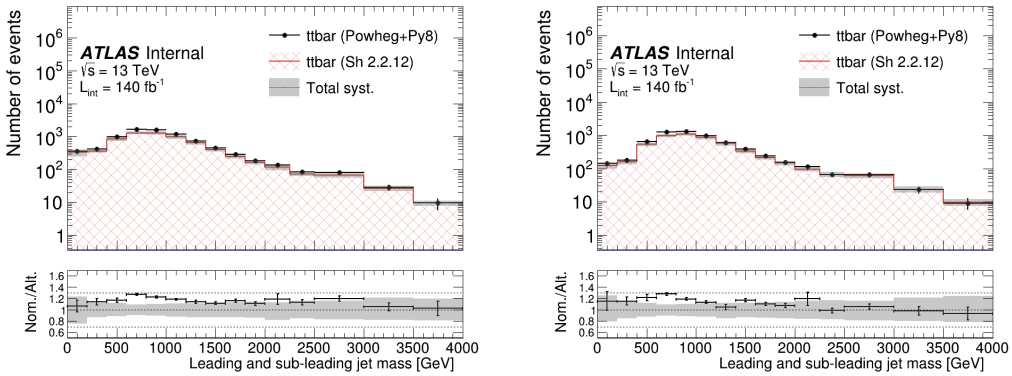
A.1.2.2 Comparison with nominal model in signal region

An additional comparison is performed in the signal regions by comparing the yields and shapes. As can be seen in Figure A.6, the modeling of the $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ distribution has a similar shape but the two models disagree at the $1-2\sigma$ level. Similar conclusions are drawn in the inclusive 2 jet region (Figure A.7), and the collinear regions (Figures A.8 and A.9). As mentioned in the previous section, the POWHEG +PYTHIA $t\bar{t}$ model only provides a leading-order (+PS) model description of our phase-space and leads to differences with the next-to-leading order model. Despite the differences observed in this section, we regard the SHERPA model as providing a better description of the $t\bar{t}$ process in our region of phase-space, as supported by the good modeling presented in Section 8.4.1 by SHERPA and the relatively poor modeling by POWHEG +PYTHIA in the section above. We do not compare to POWHEG +PYTHIA to evaluate sources of uncertainty on the SHERPA model.



(a) Electron (b) Muon

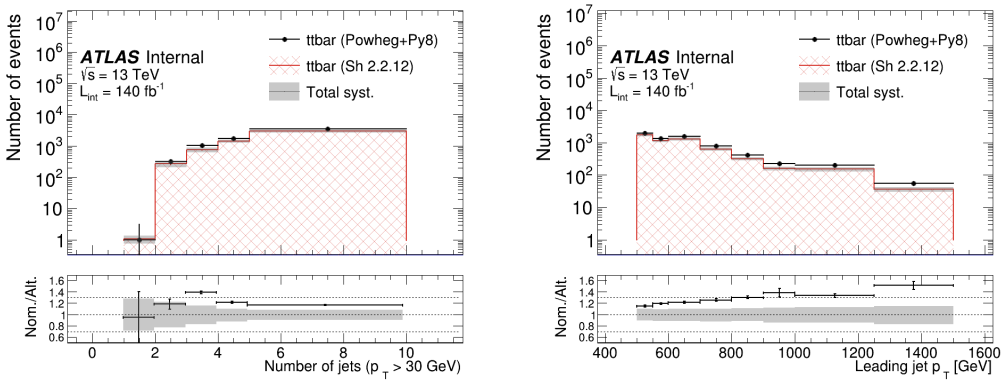
Figure A.6: Reconstructed level comparisons of the SHERPA and POWHEG +PYTHIA models in the inclusive selection Uncertainties on the SHERPA model contain both statistical and systematic uncertainties, while the POWHEG +PYTHIA model is statistical only.



(a) Electron

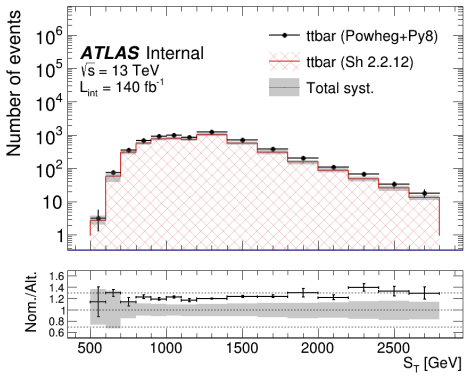
(b) Muon

Figure A.7: Reconstructed level comparisons of the SHERPA and POWHEG +PYTHIA models in the inclusive 2-jet selection Uncertainties on the SHERPA model contain both statistical and systematic uncertainties, while the POWHEG +PYTHIA model is statistical only.

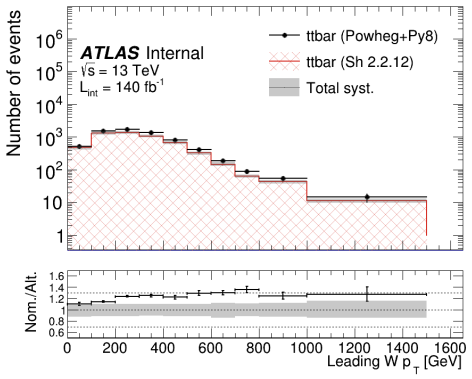


(a)

(b)

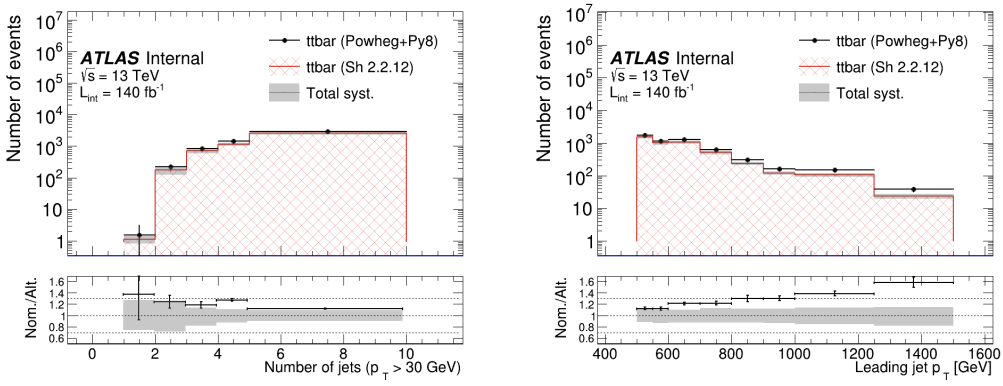


(c)



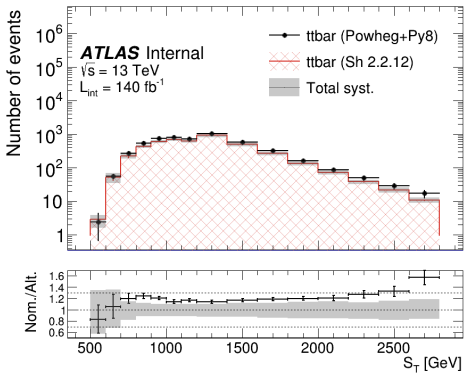
(d)

Figure A.8: Reconstructed level comparisons of the SHERPA and POWHEG +PYTHIA models in the collinear electron selection Uncertainties on the SHERPA model contain both statistical and systematic uncertainties, while the POWHEG +PYTHIA model is statistical only.

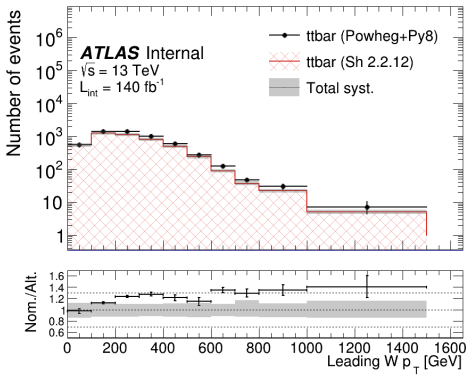


(a)

(b)



(c)



(d)

Figure A.9: Reconstructed level comparisons of the SHERPA and POWHEG +PYTHIA models in the collinear muon selection Uncertainties on the SHERPA model contain both statistical and systematic uncertainties, while the POWHEG +PYTHIA model is statistical only.

A.2 $t\bar{t}$ control region study

This section explore the high W p_{T} mis-modelling in the ttbar control region. The possible explanation for this is electroweak correction is not accounted in the powheg+pythia ttbar samples. Alternative samples with Sherpa 2.2.11 with EW correction is used for verifyting. Another possibility is singletop also contributes to the high W p_{T}. To address this, singletop and ttbar samples are treated as single process and SF is derived. All of this possibility are compared in the appendix.

A.3 Z + jets MC studies

This section contains studies performed on the Z + jets control region.

A.3.1 Di-Lepton mass selection in the Z+jets CR

In this section we estimate the purity the of the Z+jets process in its CR with tighter di-lepton mass m_{ll} selection from $[60, 120]$ GeV to $[81, 101]$ GeV. The yield for each MC process in the Z+jets CR is shown in the Table A.2. The purity of the Z+jets process is defined as the Z+jets yield to the total MC yield. No significant improvement on the Z+jets purity with narrower di-lepton mass selection in its CR.

Process	m_{ll} = [81, 101]GeV (Electron)	m_{ll} = [81, 101]GeV (Muon)	m_{ll} = [60, 120]GeV (Electron)	m_{ll} = [60, 120]GeV (Muon)
$t\bar{t}$	43.59 ± 1.07	41.90 ± 1.01	311.84 ± 3.45	353.45 ± 11.95
Diboson	235.23 ± 3.97	261.50 ± 3.95	593.30 ± 6.17	712.23 ± 6.37
Dijets	0.00 ± 0.00	0.00 ± 0.00	0.89 ± 0.58	1.73 ± 0.82
$\tau\nu$+jets	0.20 ± 0.27	0.73 ± 0.29	1.62 ± 0.92	3.07 ± 0.99
Single top	4.01 ± 0.73	4.10 ± 0.68	28.03 ± 1.92	33.73 ± 2.00
W+2j	0.09 ± 0.06	0.11 ± 0.06	0.98 ± 0.21	1.43 ± 0.24
W+jets (Sherpa 2.2.11)	3.80 ± 0.68	7.15 ± 0.62	27.69 ± 2.73	46.15 ± 1.95
V+γ+jets	239.96 ± 2.84	325.23 ± 4.00	745.39 ± 7.66	1125.43 ± 11.44
Z+jets (Sherpa 2.2.11)	5851.93 ± 18.29	6591.39 ± 21.63	16035.00 ± 25.62	19589.20 ± 32.83
Total SM	6378.81 ± 18.99	7232.12 ± 22.39	17744.75 ± 27.88	21866.41 ± 37.44
Observed data	6409.00 ± 80.06	8602.00 ± 92.75	17579.00 ± 132.59	25995.00 ± 161.23
Z+jets purity	91.7 ± 1.2%	91.1 ± 0.4%	90.4 ± 0.2%	89.6 ± 0.2%

Table A.2: Comparson of yield for nominal ([60,120]) and alternative ([81,101]) di-lepton mass m_{ll} selection.

A.3.2 MadGraph5_aMC@NLO +Pythia8 FxFx Z + jets model

In this section, the modeling of Z + jets with the MadGraph5_aMC@NLO +Pythia8 FxFx Z + jets model is assessed and compared against the default Sherpa 2.2.11 prediction.

A.3.2.1 Comparison in control region

The MadGraph5_aMC@NLO +Pythia8 FxFx merged sample can be used as an alternative prediction for the Z + jets process. In this section, we re-derive the normalization factors with this alternative generator and assess the change of the Z + jets background in the signal region.

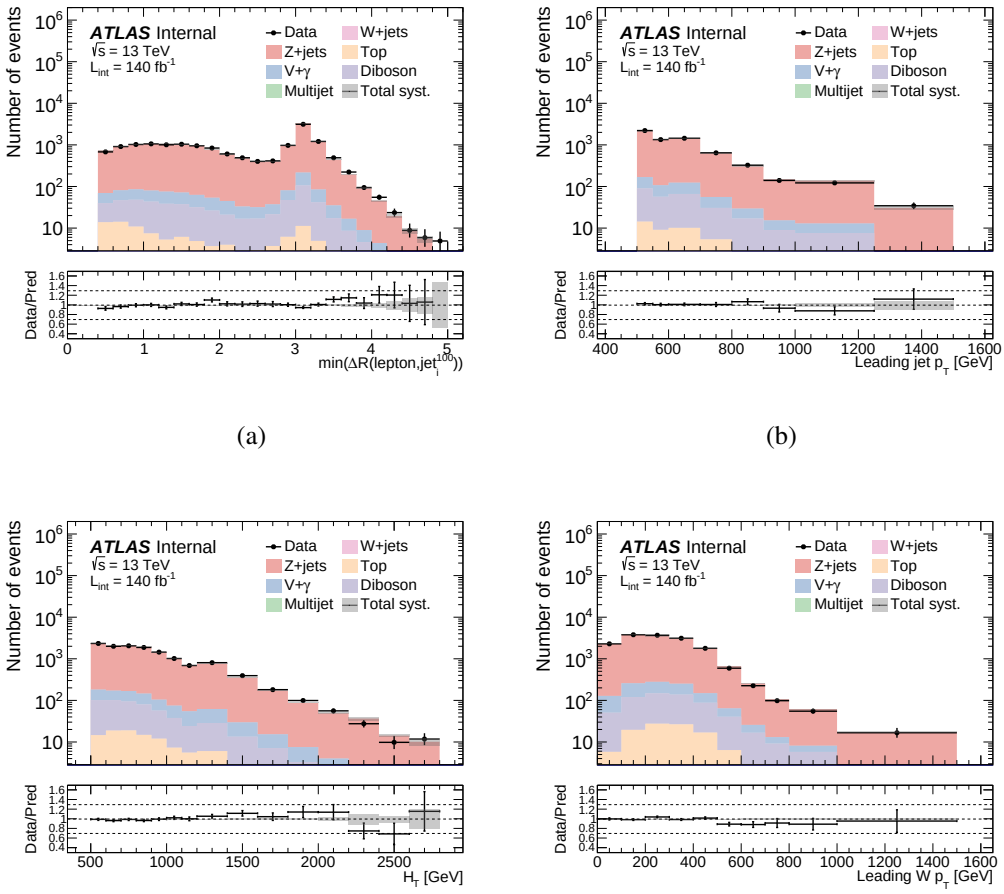


Figure A.10: Reconstructed level comparisons of data and background predictions in the $Z+$jets control region for events with two **electrons**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

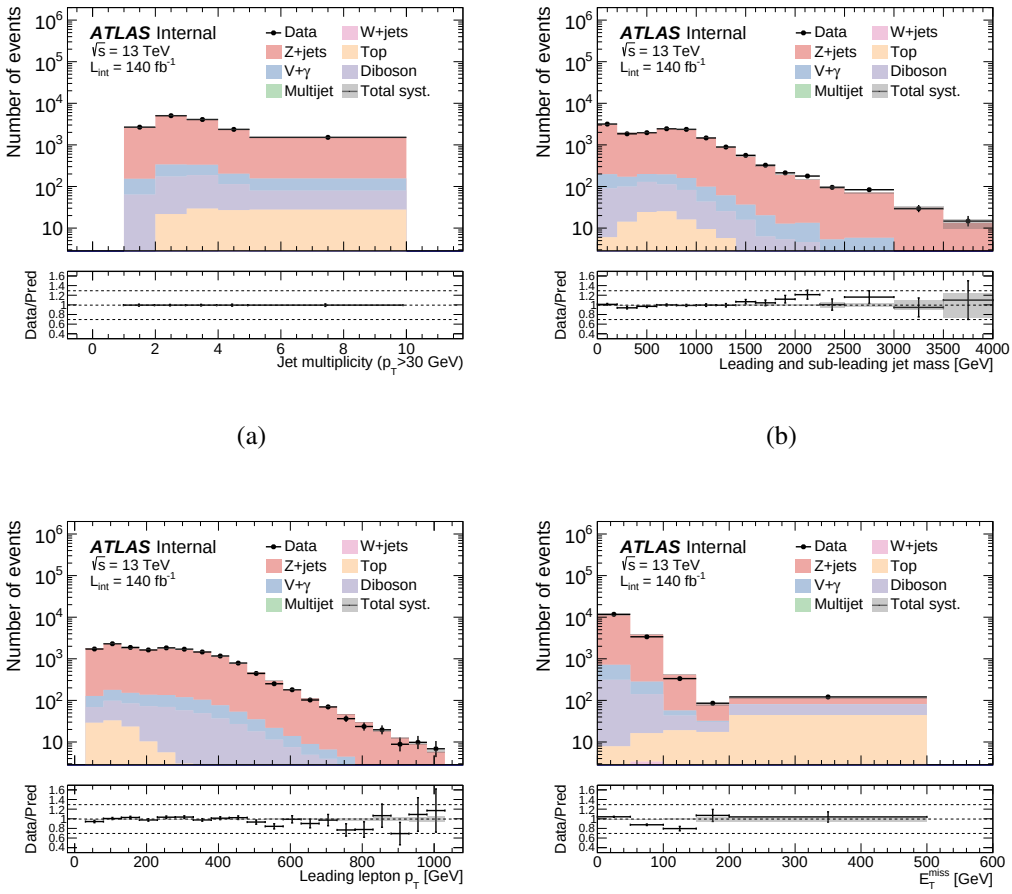


Figure A.11: Reconstructed level comparisons of data and background predictions in the $Z+$jets control region for events with two **electrons**. Normalization factors for $t\bar{t}$ and $Z+$jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

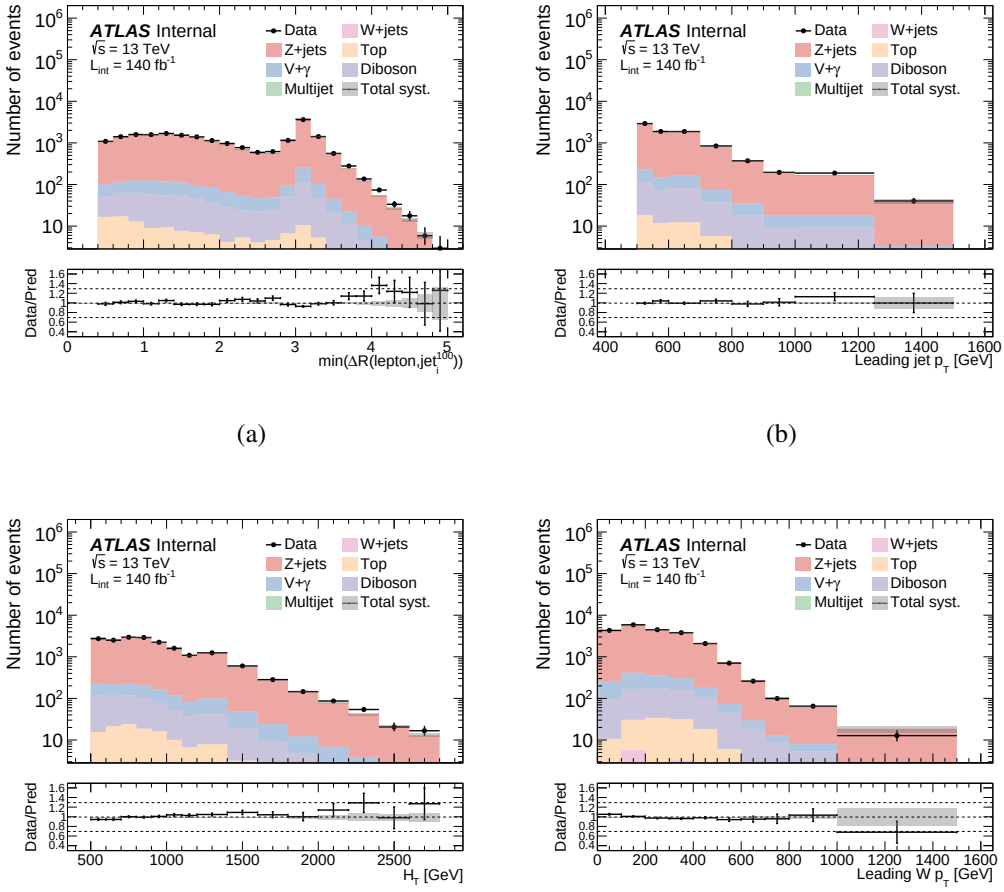


Figure A.12: Reconstructed level comparisons of data and background predictions in the Z + jets control region for events with two **muons**. Normalization factors for $t\bar{t}$ and Z + jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

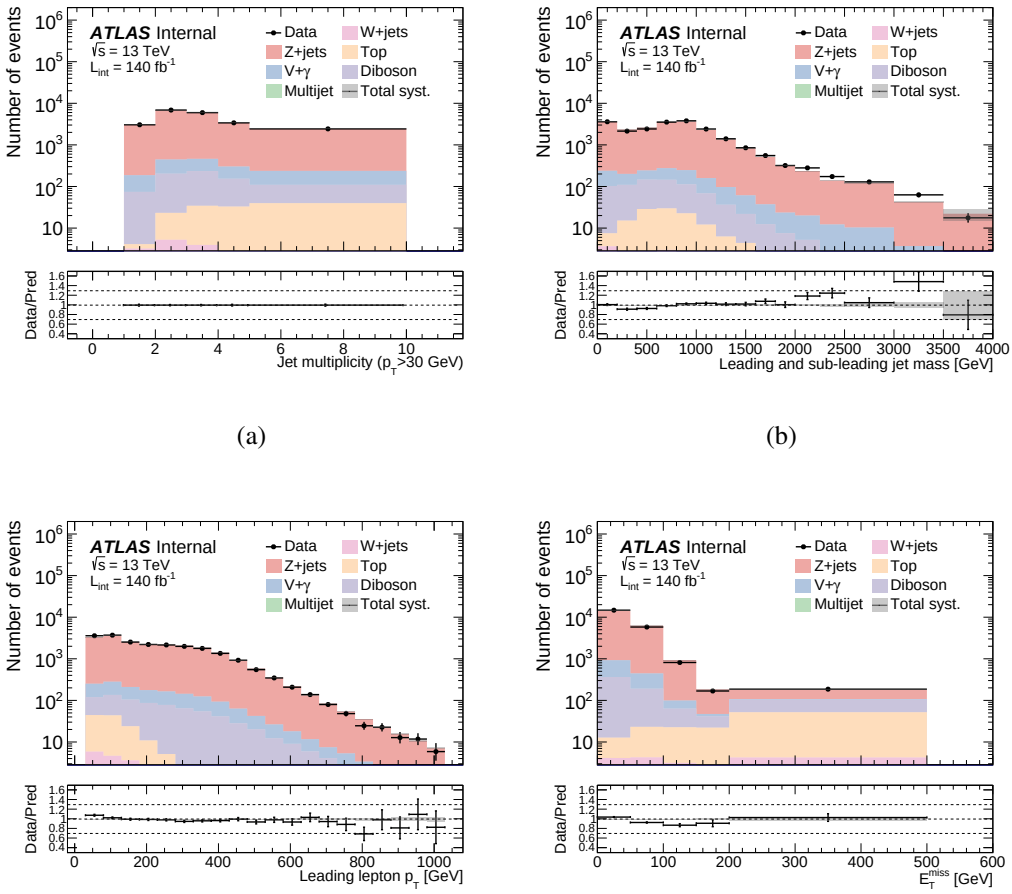
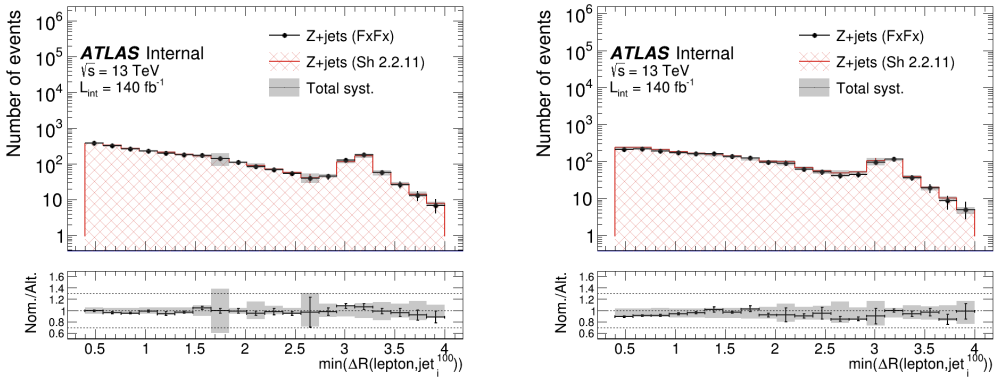


Figure A.13: Reconstructed level comparisons of data and background predictions in the Z + jets control region for events with two **muons**. Normalization factors for $t\bar{t}$ and Z + jets are applied as described in the text. The grey error band include statistical and systematic uncertainties added in quadrature.

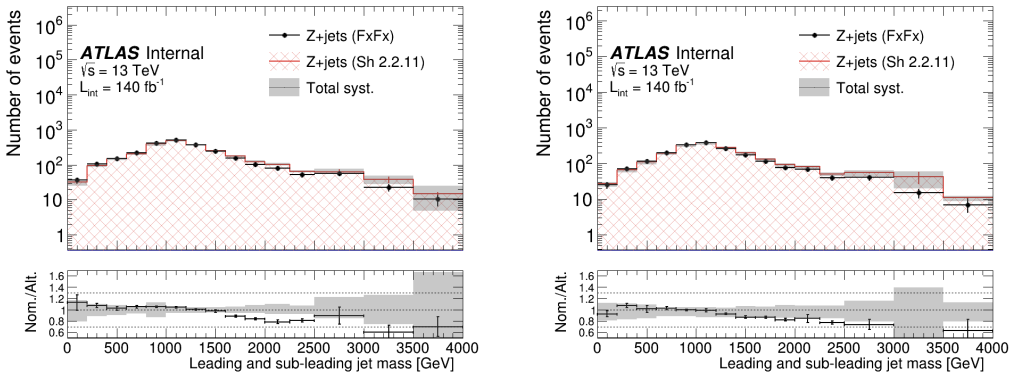
A.3.2.2 Comparison with nominal model in signal region

An additional comparison is performed in the signal regions by comparing the yields and shapes. As can be seen in Figure A.14, the modeling of the $\Delta R_{\text{min}_i}(\ell, \text{jet}_i^{100})$ distribution agrees well between the two models in both the electron and muon channels. In Figure A.15, the MadGraph5_aMC@NLO +Pythia8 FxFx Z + jets model predicts a slightly *softer* m_{jj} spectrum compared to the Sherpa model, but is within $1-2\sigma$ of the combined statistical and systematic uncertainty band. Additional distributions in the collinear region are shown in Figures A.16 and A.17 for the electron and muon selections, respectively. Good agreement is observed for all distributions.



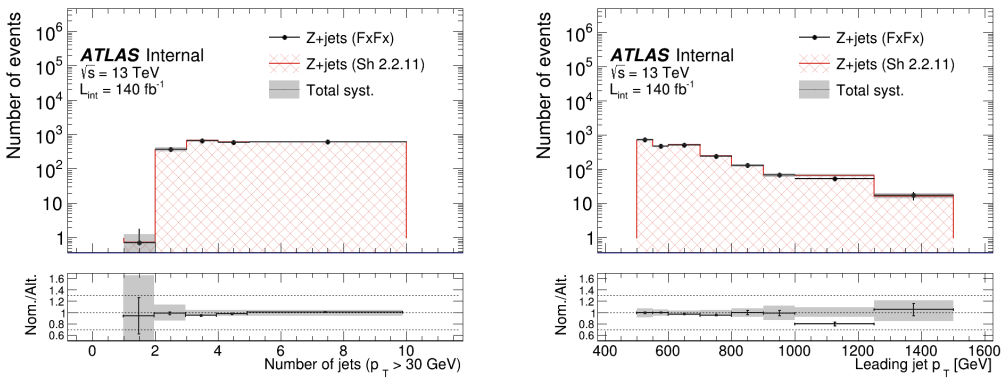
(a) Electron (b) Muon

Figure A.14: Reconstructed level comparisons of the Sherpa and MadGraph5_aMC@NLO +Pythia8 FxFx models in the inclusive selection Uncertainties on the Sherpa model contain both statistical and systematic uncertainties, while the MadGraph5_aMC@NLO +Pythia8 FxFx model is statistical only.



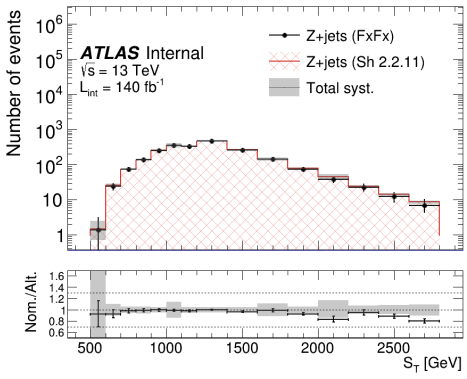
(a) Electron (b) Muon

Figure A.15: Reconstructed level comparisons of the SHERPA and MADGRAPH5_aMC@NLO +PYTHIA8 FxFx models in the inclusive selection Uncertainties on the SHERPA model contain both statistical and systematic uncertainties, while the MADGRAPH5_aMC@NLO +PYTHIA8 FxFx model is statistical only.

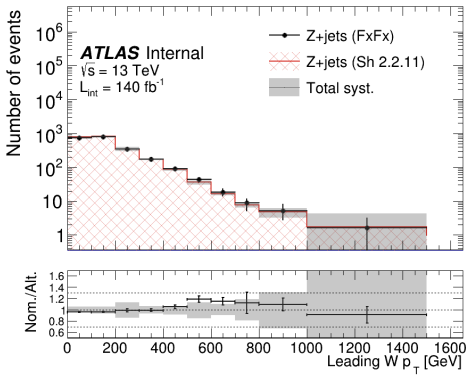


(a)

(b)

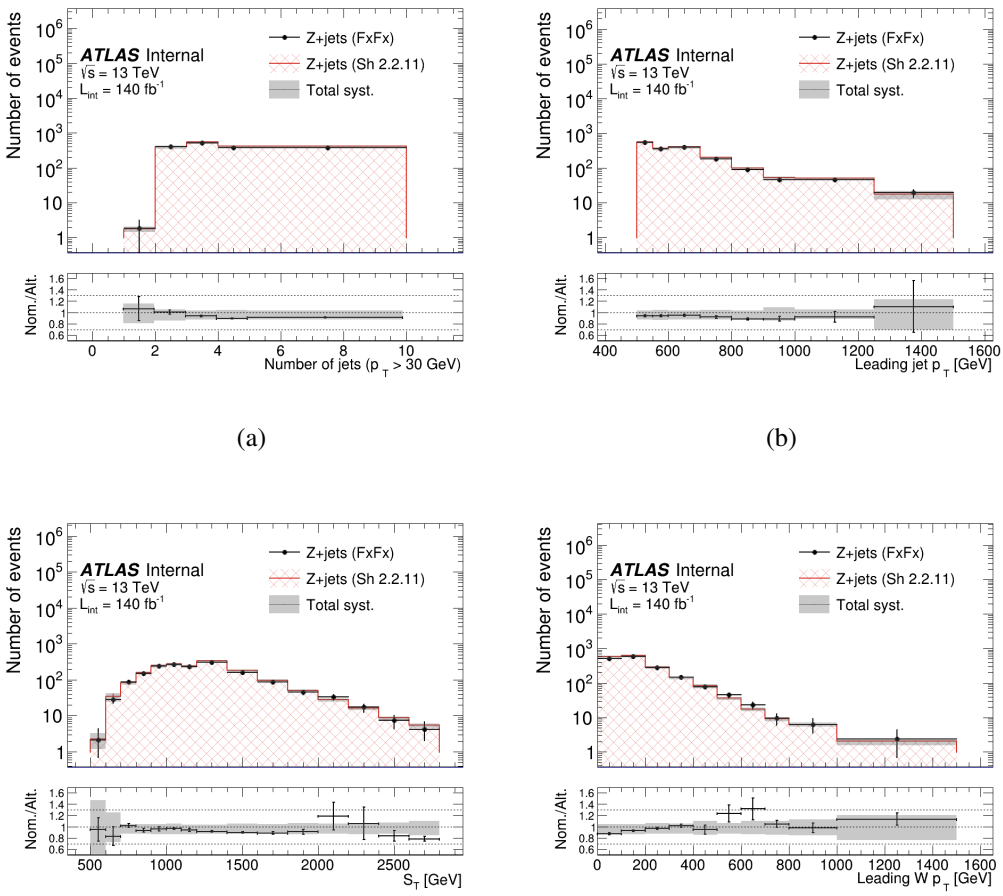


(c)



(d)

Figure A.16: Reconstructed level comparisons of the SHERPA and MADGRAPH5_aMC@NLO +PYTHIA8 FxFx models in the inclusive selection Uncertainties on the SHERPA model contain both statistical and systematic uncertainties, while the MADGRAPH5_aMC@NLO +PYTHIA8 FxFx model is statistical only.



(a) (b)

(c) (d)

Figure A.17: Reconstructed level comparisons of the SHERPA and MADGRAPH5_aMC@NLO +PYTHIA8 FxFx models in the inclusive selection Uncertainties on the SHERPA model contain both statistical and systematic uncertainties, while the MADGRAPH5_aMC@NLO +PYTHIA8 FxFx model is statistical only.

A.4 Additional background fit studies

A.4.1 Pre-fit control region distributions

This section presents the pre-fit control region distributions.

A.4.2 Yields in the control regions

The before fit yields for the $t\bar{t}$ and Z + jets control regions can be found in Tables A.3 and A.4 for the electron and muon channels, respectively. The multi-jet control region yields can be found in Table A.5. Uncertainties on the pre-fit yields are statistical only.

Table A.3: Yields in the $t\bar{t}$ and $Z+$jets CRs **before** the fit in the **electron** channel. Uncertainties are statistical only.

Process	$t\bar{t}$ CR	Z+jets CR
$t\bar{t}$	18193.84 ± 115.46	341.34 ± 14.45
Diboson	75.77 ± 2.27	607.14 ± 6.34
$\tau\nu$+jets	59.98 ± 3.66	1.95 ± 1.05
Single top	2390.14 ± 16.91	29.51 ± 2.04
EW V+jets	8.16 ± 0.59	1.12 ± 0.22
Z+jets	109.51 ± 2.28	18224.95 ± 26.44
Dijets	441.18 ± 37.46	0.91 ± 0.60
W+jets	754.26 ± 12.78	31.37 ± 3.03
Total SM	22032.84 ± 123.32	19238.29 ± 31.03
Observed data	19887.00 ± 141.02	17398.00 ± 131.90

Table A.4: Yields in the $t\bar{t}$ and $Z+$jets CRs **before** the fit in the **muon** channel. Uncertainties are statistical only.

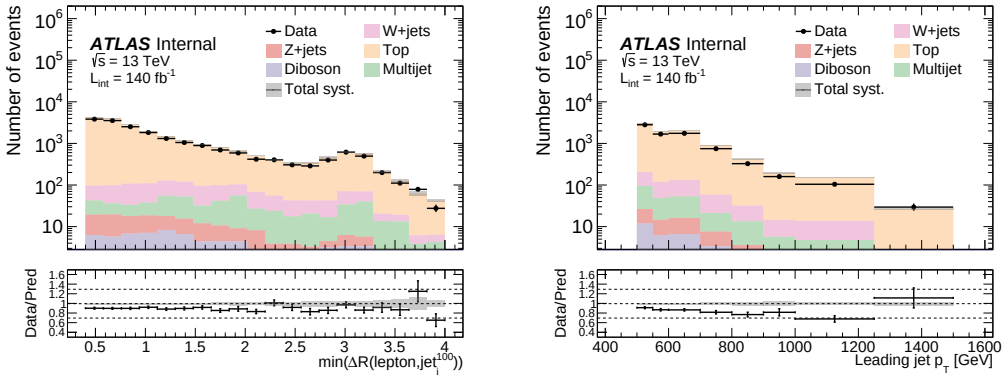
Process	$t\bar{t}$ CR	Z+jets CR
$t\bar{t}$	15254.24 ± 98.55	359.15 ± 15.60
$\tau\nu$+jets	47.93 ± 3.43	3.66 ± 1.08
Diboson	70.15 ± 2.36	802.42 ± 7.17
EW V+jets	7.53 ± 0.55	1.66 ± 0.27
V+γ+jets	39.72 ± 1.84	1214.49 ± 12.83
Single top	2115.84 ± 15.27	38.02 ± 2.29
Z+jets	69.39 ± 1.24	23862.30 ± 39.50
Dijets	259.17 ± 29.30	1.97 ± 0.94
W+jets	771.95 ± 11.63	51.90 ± 2.14
Total SM	18596.21 ± 104.68	25121.08 ± 43.20
Observed data	17878.00 ± 133.71	25400.00 ± 159.37

Table A.5: Yields in the multi-jets control region **before** the fit. Uncertainties are statistical only.

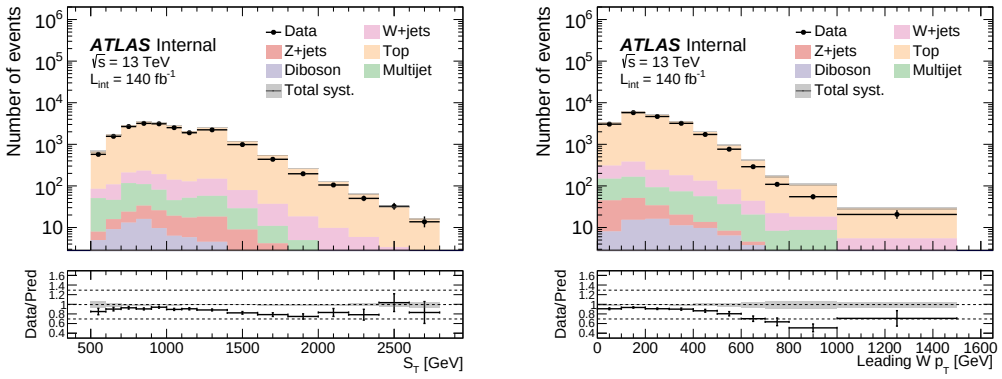
Process	Electron	Muon
$t\bar{t}$	889.48 ± 25.75	1005.51 ± 43.73
Diboson	626.87 ± 12.62	341.11 ± 9.04
$\tau\nu$+jets	3073.08 ± 37.93	6398.52 ± 54.42
Single top	485.84 ± 7.91	342.94 ± 6.63
EW V+jets	394.40 ± 4.28	285.68 ± 3.51
Z+jets	1088.81 ± 11.35	1187.75 ± 10.08
Dijets	111824.17 ± 721.76	133522.27 ± 575.98
W+jets	6241.67 ± 43.58	806.16 ± 23.80
Total SM	124624.31 ± 724.78	143889.94 ± 580.89
Observed data	88548.00 ± 297.57	102885.00 ± 320.76

A.4.3 $t\bar{t}$ control region

The plots shown in this section contain the $t\bar{t}$ CR selection. The electron channel plots are shown in Figures A.18 and A.19, and muons channel plots are shown in Figures A.20 and A.21.

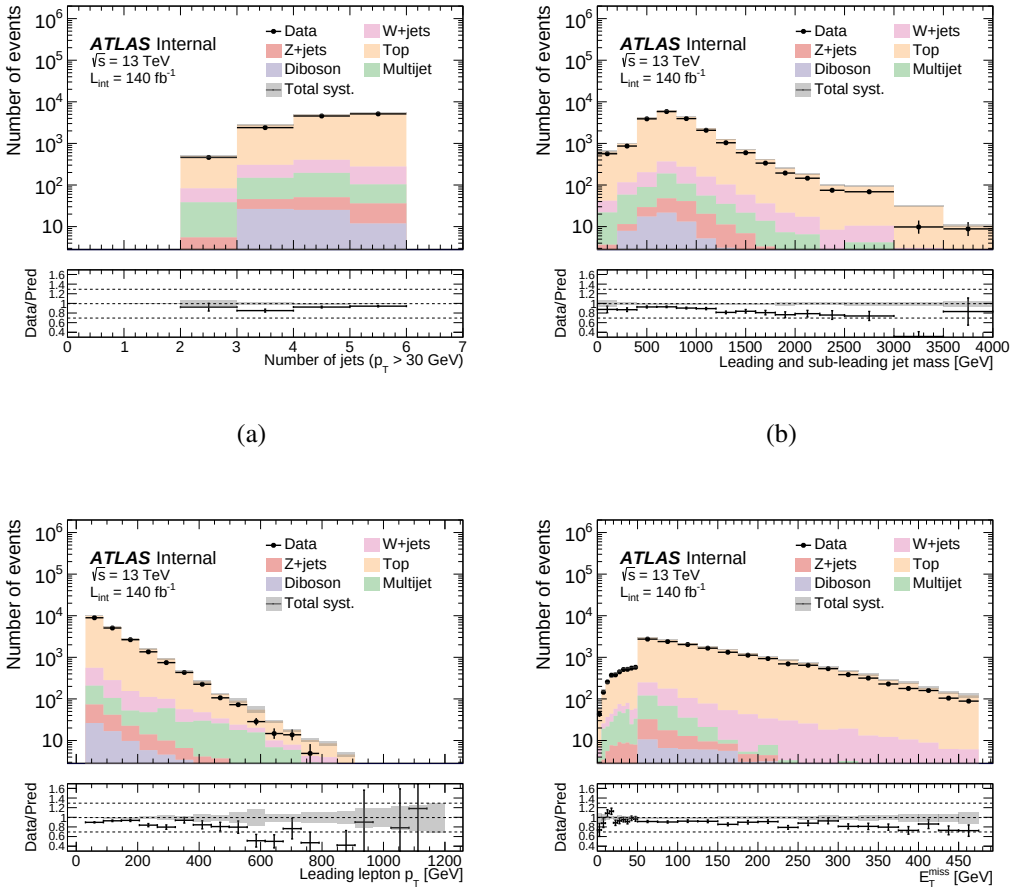


(a) (b)



(c) (d)

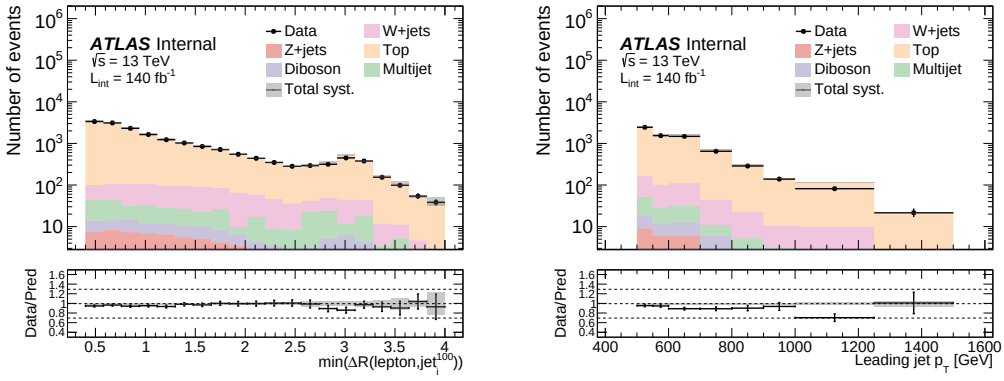
Figure A.18: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **electron**. The grey error band include statistical and systematic uncertainties added in quadrature.



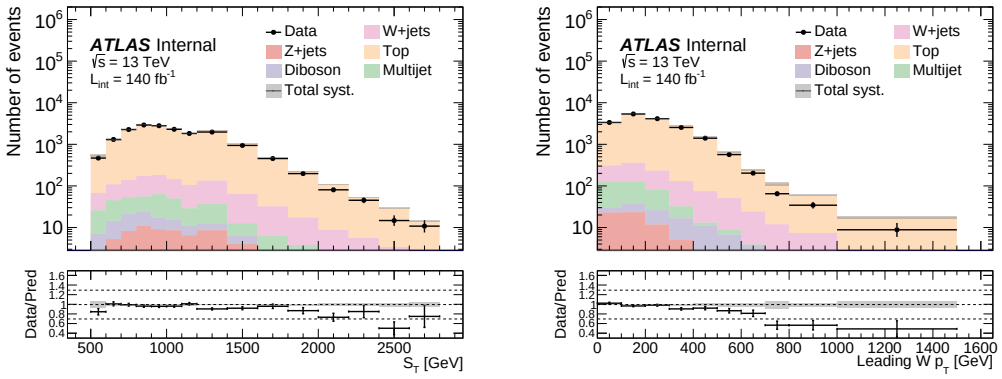
(a) (b)

(c) (d)

Figure A.19: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **electron**. The grey error band include statistical and systematic uncertainties added in quadrature.

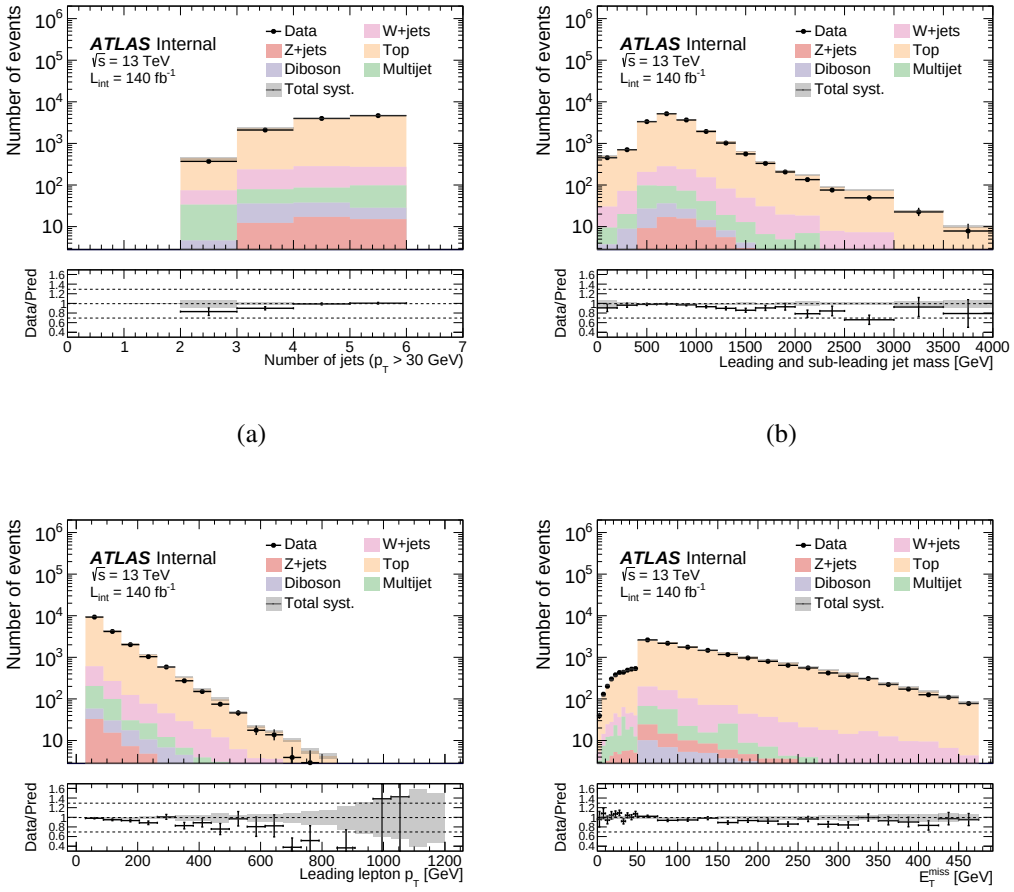


(a) (b)



(c) (d)

Figure A.20: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **muon**. The grey error band include statistical and systematic uncertainties added in quadrature.



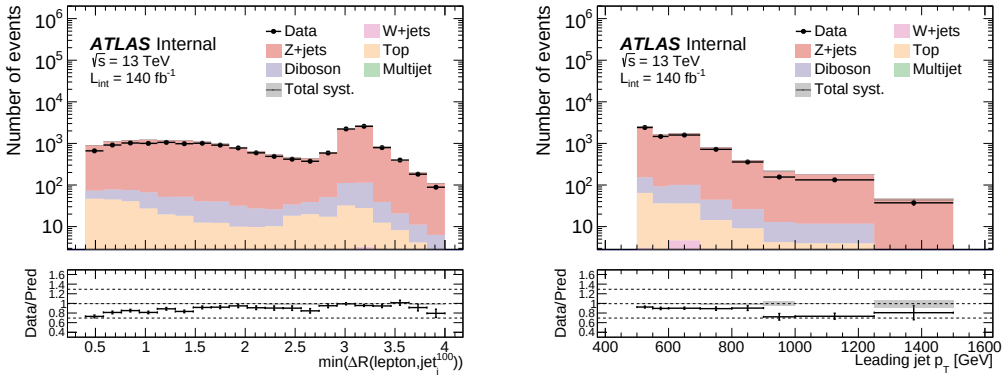
(a) (b)

(c) (d)

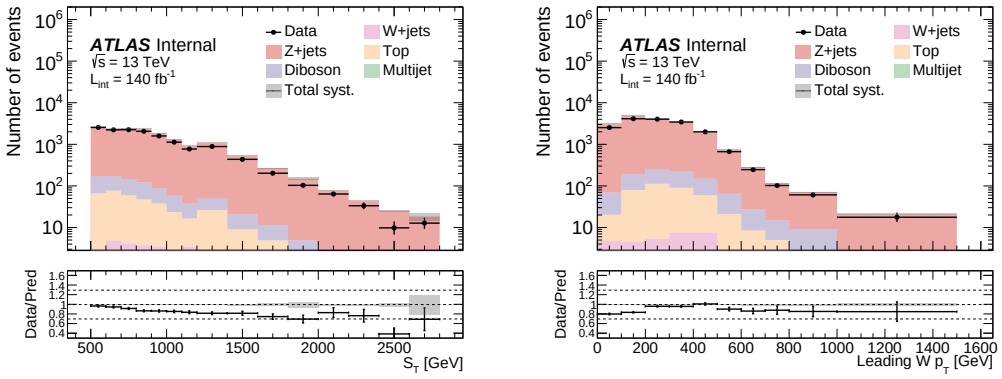
Figure A.21: Reconstructed level comparisons of data and background predictions in the $t\bar{t}$ control region for events with a single **muon**. The grey error band include statistical and systematic uncertainties added in quadrature.

A.4.4 Z + jets control region

The plots shown in this section contain the Z + jets CR selection. The electron channel plots are shown in Figures A.22 and A.23, and muons channel plots are shown in Figures A.24 and A.25.

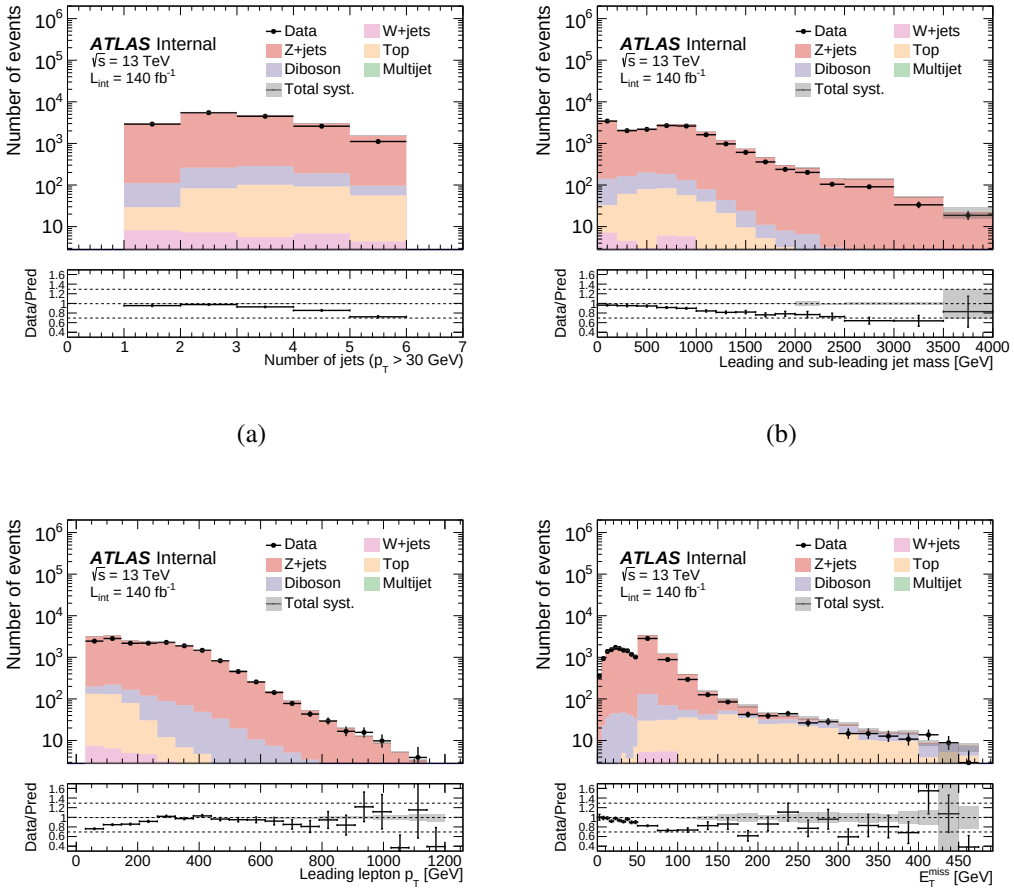


(a) (b)



(c) (d)

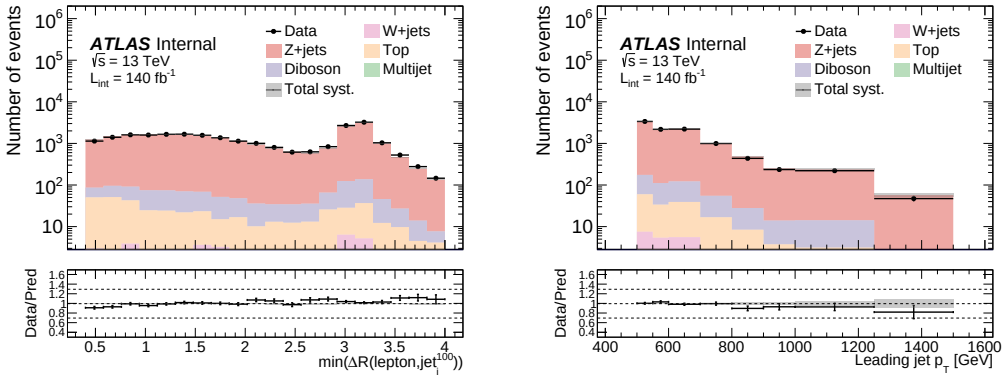
Figure A.22: Reconstructed level comparisons of data and background predictions in the Z + jets control region for events with two **electrons**. The grey error band include statistical and systematic uncertainties added in quadrature.



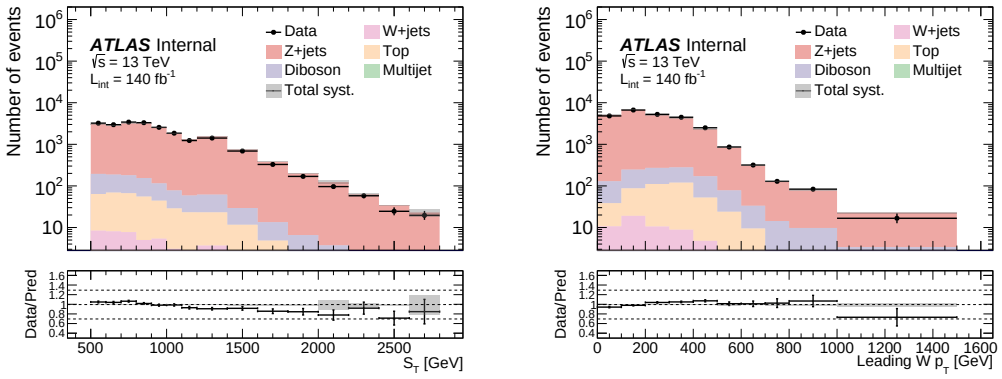
(a) (b)

(c) (d)

Figure A.23: Reconstructed level comparisons of data and background predictions in the Z + jets control region for events with two **electrons**. The grey error band include statistical and systematic uncertainties added in quadrature.

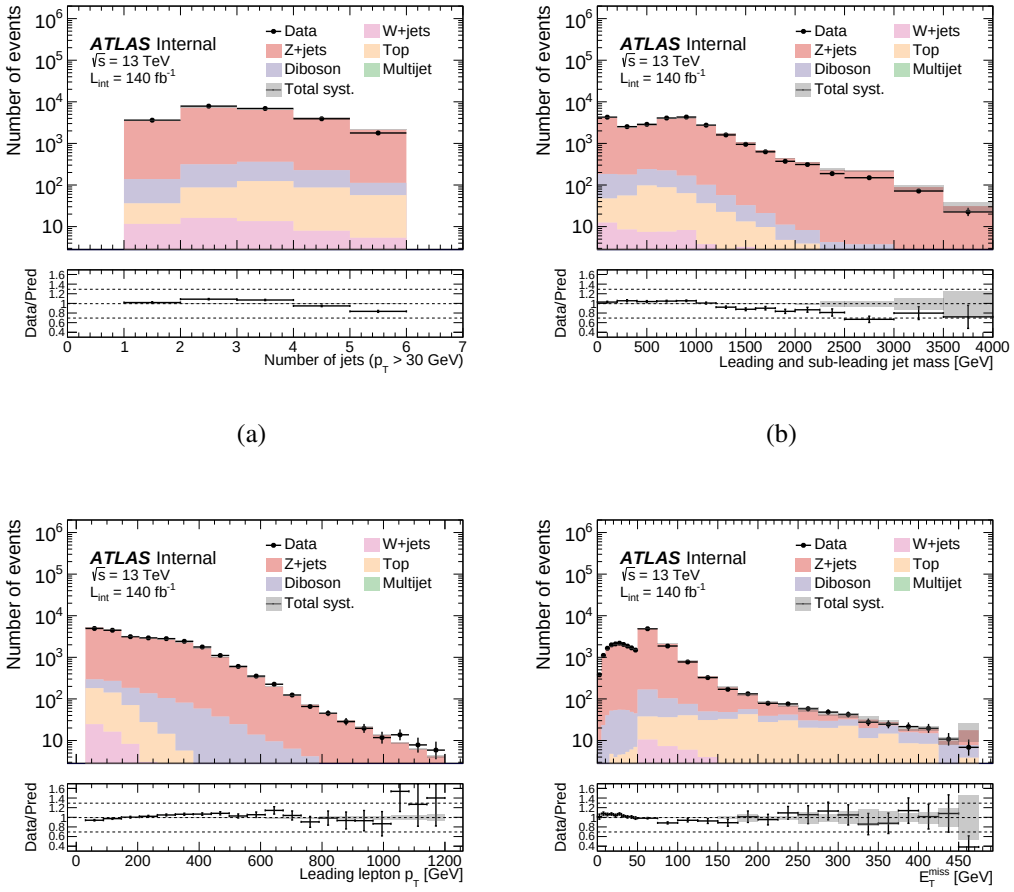


(a) (b)



(c) (d)

Figure A.24: Reconstructed level comparisons of data and background predictions in the Z + jets control region for events with two **muons**. The grey error band include statistical and systematic uncertainties added in quadrature.



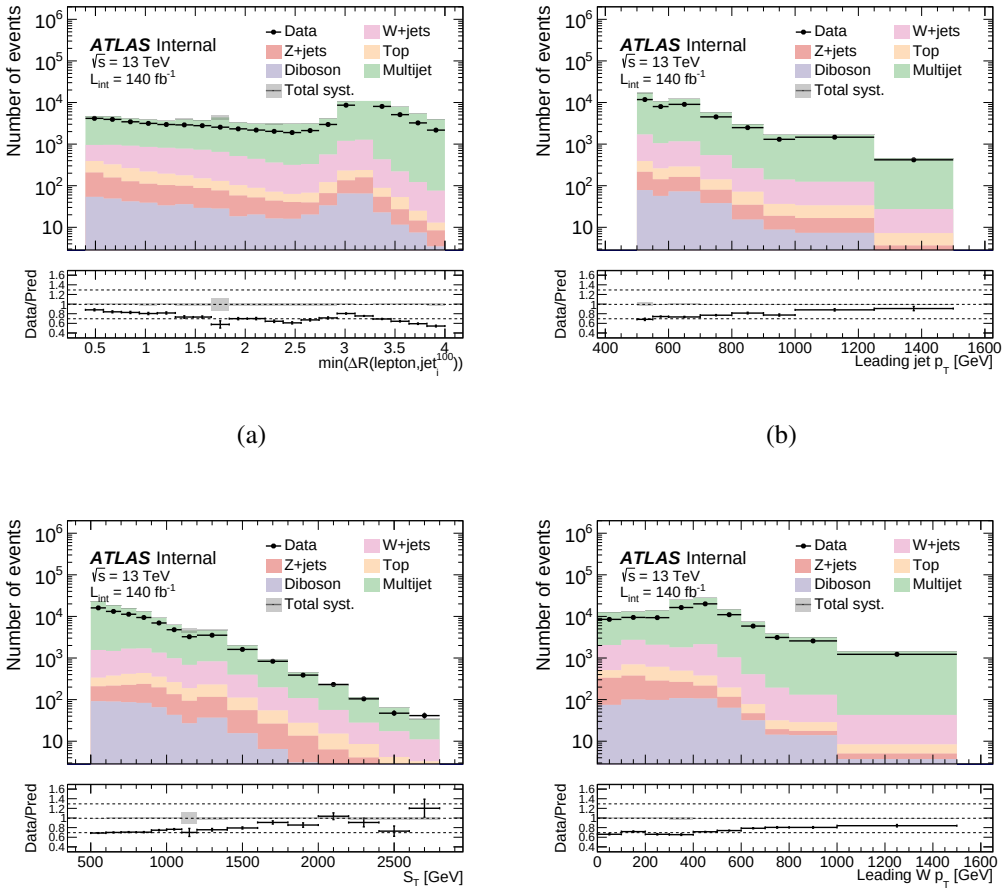
(a) (b)

(c) (d)

Figure A.25: Reconstructed level comparisons of data and background predictions in the $Z+$jets control region for events with two **muons**. The grey error band include statistical and systematic uncertainties added in quadrature.

A.4.5 Multi-jet control region

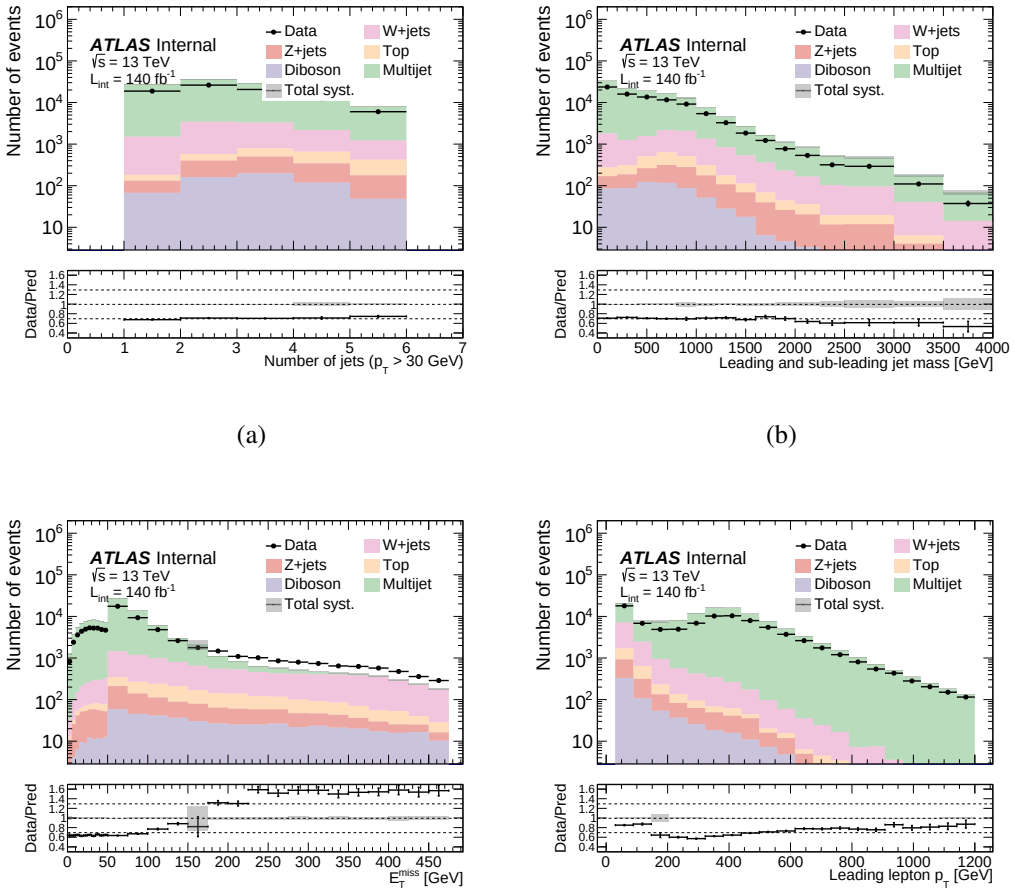
The plots shown in this section contain the multi-jet CR selection. The electron channel plots are shown in Figures A.26 and A.27, and muons channel plots are shown in Figures A.28 and A.29.



(a) (b)

(c) (d)

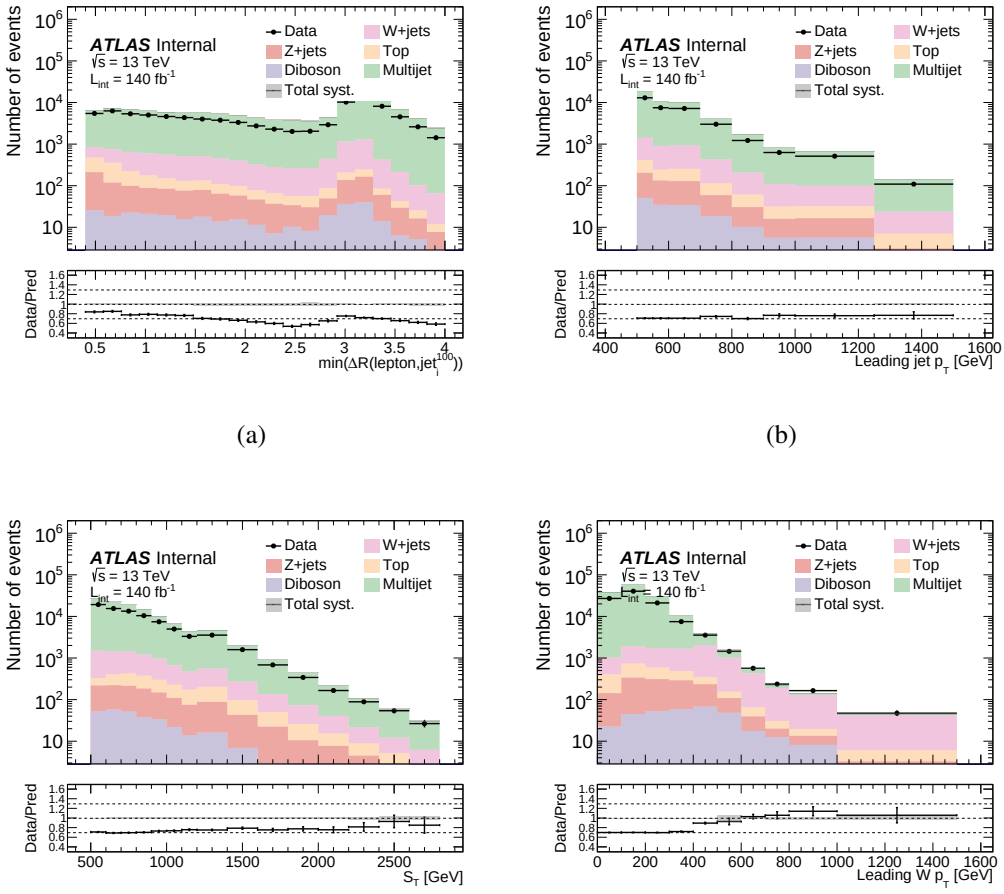
Figure A.26: Reconstructed level comparisons of data and background predictions in the di-jets control region for **electrons** before fit. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven di-jet background estimate.



(a) (b)

(c) (d)

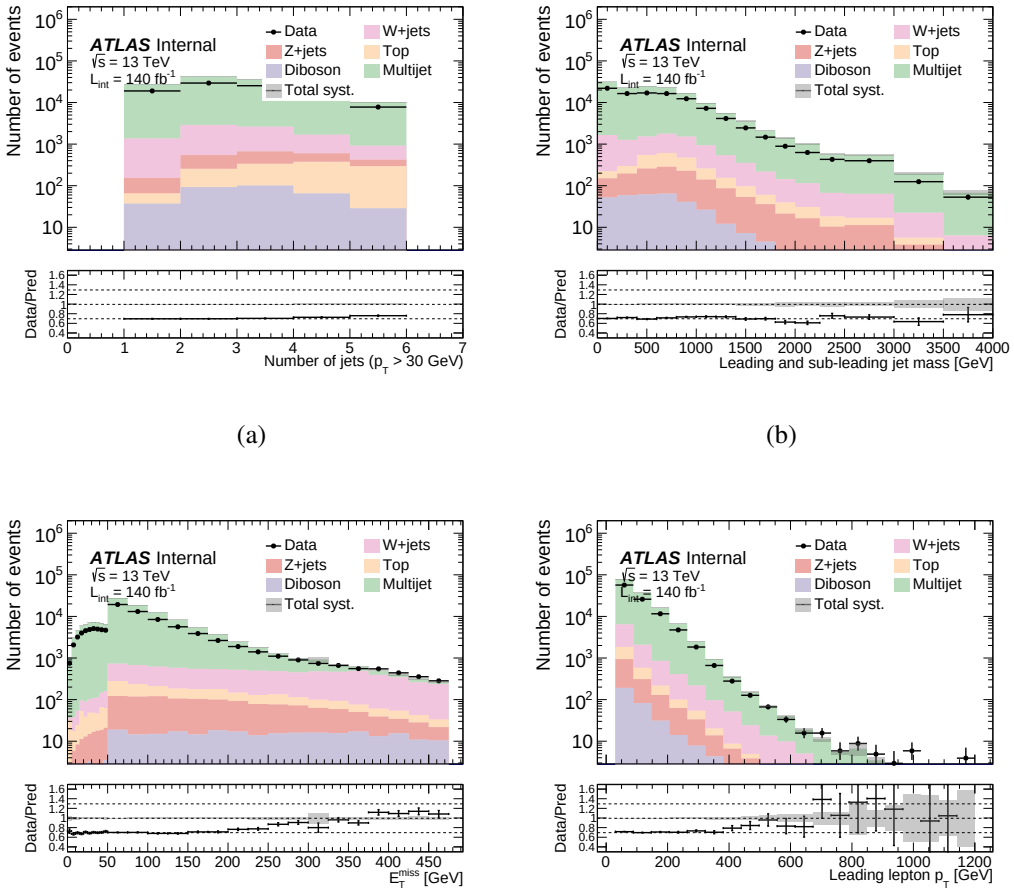
Figure A.27: Reconstructed level comparisons of data and background predictions in the di-jets control region for **electrons** before fit. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven di-jet background estimate.



(a) (b)

(c) (d)

Figure A.28: Reconstructed level comparisons of data and background predictions in the dijets control region for **muons** before fit. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven di-jet background estimate.



(a) (b)

(c) (d)

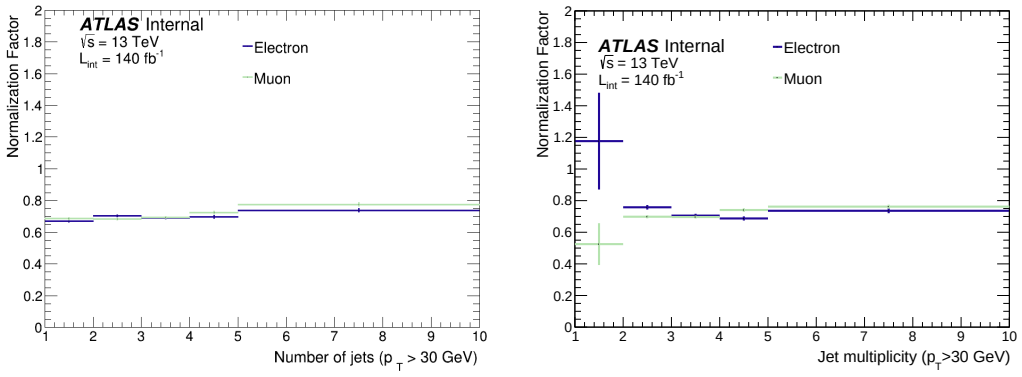
Figure A.29: Reconstructed level comparisons of data and background predictions in the dijets control region for **muons** before fit. The gray band shows the total statistical and systematic uncertainty on the semi-data-driven di-jet background estimate.

A.4.6 Control region dependency on inclusive and collinear selections

As discussed in Section 8.2, we derived background normalization factors (SFs) for $t\bar{t}$, Z+jets, and dijets processes using dedicated control regions. These control regions are defined with inclusive ΔR selection, but applied to all regions in this analysis. In this section, we compare the difference in yields that would be obtain if the normalization factors would have been derived in a control region with the collinear selection applied. As will be shown, the difference between the approaches is negligible, and we will continue using the normalization factors in the inclusive region for all regions.

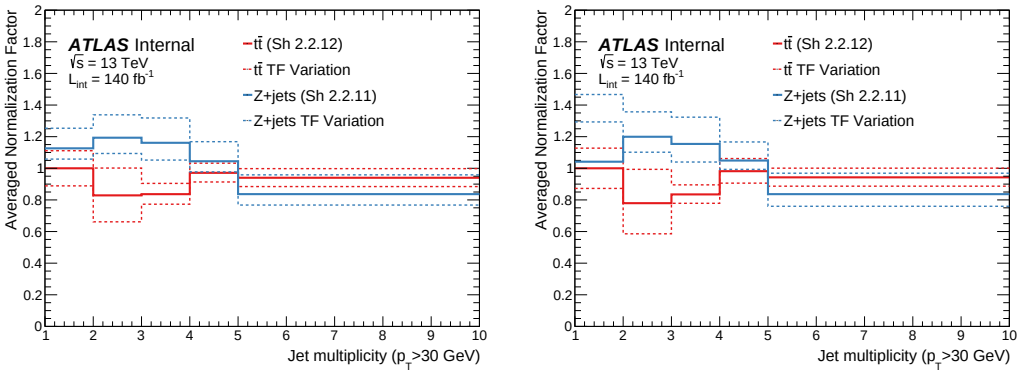
The comparison of the normalization factors in the inclusive and collinear selections is shown in Figures A.30 and A.31. Already at the scale factor level, we can see that the two agree well, indicating a very small dependency on the ΔR phase-space selection. A statistically insignificant difference can be seen in first jet multiplicity bin; since the collinear region is dominated by events with >= 2 jets, such a region contains very few events to derive meaningful normalization factors for the 1-jet bin, and thus contains large statistical uncertainties.

In order to assess the impact on the background estimate in the collinear region, a comparison of the total background yields is performed with the normalization factors derived in the inclusive and collinear regions. This is shown in Figures A.32 and A.33. The overall difference is within 5% and well covered by the systematic uncertainty.



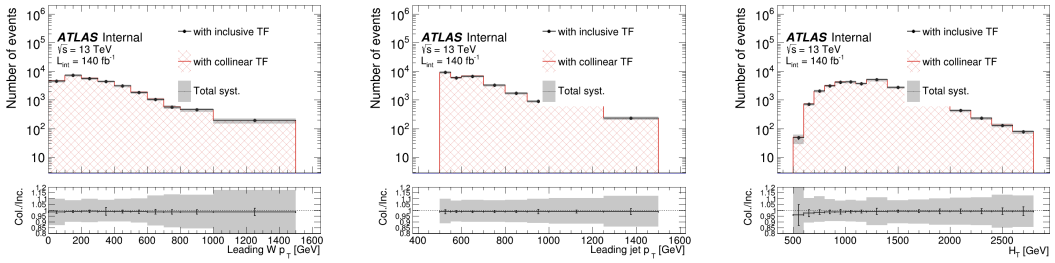
(a) inclusive ΔR selection (b) $\Delta R \leq 2.6$ selection

Figure A.30: Comparison of SFs for dijets background.



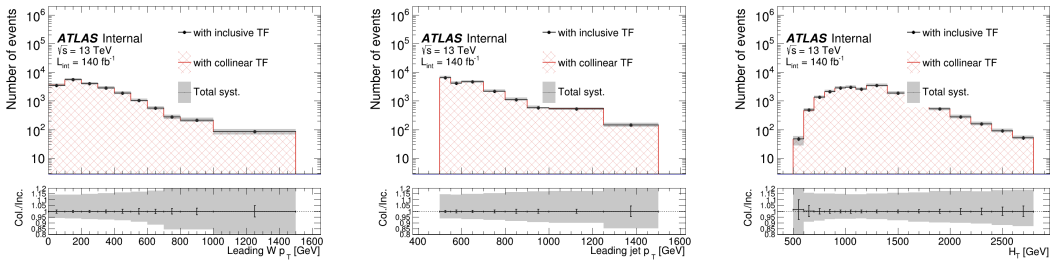
(a) inclusive ΔR selection (b) $\Delta R \leq 2.6$ selection

Figure A.31: Comparison of SFs for $t\bar{t}$ and Z+jets backgrounds.



(a) (b) (c)

Figure A.32: Comparison of the sum of backgrounds in the **electron** collinear region.



(a) (b) (c)

Figure A.33: Comparison of the sum of backgrounds in the **muon** collinear region.

A.4.7 Control region leading jet p_{T} >400 GeVand p_{T} >500 GeV

Process	Inclusive	Collinear	Inclusive 2-jets
$t\bar{t}$	7229 ± 2653	5808 ± 2254	7073 ± 2629
Diboson	3443 ± 189	2224 ± 123	3073 ± 169
Single top	2149 ± 614	1608 ± 471	2056 ± 586
Z+jets	6352 ± 1921	5110 ± 1744	5990 ± 1911
EW V+jets	4182 ± 726	3029 ± 527	3746 ± 656
$\tau\nu$+jets	6494 ± 856	4034 ± 548	5619 ± 746
Multi-jets	14628 ± 2414	6344 ± 909	11598 ± 1767
W+jets (Sh 2.2.11)	77339 ± 22342	53989 ± 19899	68367 ± 22568
Total Signal+Bkg	121816 ± 22756	82146 ± 20148	107522 ± 22906
Observed data	119880 ± 346	79887 ± 283	105740 ± 325

Table A.6: The **post-fit** yield yield of each process in the various **electron** measurement regions using CR leading jet p_{T} >**400** GeV(nominal case).

Process	Inclusive	Collinear	Inclusive 2-jets
$t\bar{t}$	6965.67 ± 65.18	5617.19 ± 36.40	6783.34 ± 57.56
Diboson	3429.93 ± 29.83	2214.49 ± 23.99	3061.01 ± 28.18
$\tau\nu$+jets	6274.84 ± 61.82	3897.82 ± 55.43	5428.32 ± 59.50
Single top	2139.20 ± 17.36	1598.84 ± 14.73	2045.23 ± 16.80
Z+jets	5216.70 ± 32.95	4149.08 ± 22.58	4885.78 ± 25.68
EW V+jets	3942.82 ± 13.46	2840.62 ± 11.41	3522.02 ± 12.70
V+γ+jets	4193.53 ± 16.26	2818.19 ± 14.60	3755.82 ± 15.37
Multi-jets	19092.05 ± 133.37	9480.06 ± 119.18	15781.27 ± 129.59
W+jets (Sh 2.2.11)	72874.26 ± 184.81	50385.77 ± 167.26	64171.16 ± 180.71
Total SM	124129.00 ± 250.47	83002.06 ± 219.59	109433.94 ± 241.73
Observed data	118879.00 ± 344.79	79190.00 ± 281.41	104791.00 ± 323.71

Table A.7: The **post-fit** yield yield of each process in the various **electron** measurement regions using CR leading jet p_{T} >**500** GeV.

Process	Inclusive	Collinear	Inclusive 2-jets
$t\bar{t}$	5530 ± 2101	4845 ± 1891	5492 ± 2086
Diboson	2881 ± 159	1929 ± 107	2593 ± 143
Single top	1742 ± 511	1369 ± 409	1693 ± 495
Z+jets	4328 ± 1445	3585 ± 1347	4149 ± 1465
EW V+jets	3572 ± 620	2636 ± 457	3233 ± 565
$\tau\nu$+jets	5096 ± 629	3321 ± 382	4484 ± 565
Multi-jets	3637 ± 368	2461 ± 220	3251 ± 302
W+jets (Sh 2.2.11)	72016 ± 21722	52024 ± 19654	64652 ± 22107
Total Signal+Bkg	98802 ± 21904	72170 ± 19808	89548 ± 22280
Observed data	94229 ± 307	67969 ± 261	85128 ± 292

Table A.8: The **post-fit** yield yield of each process in the various **muon** measurement regions using CR leading jet p_{T} >**400** GeV(nominal case).

Process	Inclusive	Collinear	Inclusive 2-jets
$t\bar{t}$	5676.85 ± 86.94	4928.81 ± 67.86	5606.42 ± 76.03
Diboson	2877.95 ± 26.62	1926.75 ± 21.86	2590.95 ± 25.33
$\tau\nu$+jets	4698.87 ± 36.01	3048.18 ± 28.39	4120.49 ± 33.48
Single top	1736.87 ± 15.09	1364.46 ± 13.27	1688.23 ± 14.86
Z+jets	3638.05 ± 23.73	3010.96 ± 14.01	3482.95 ± 23.61
EW V+jets	3297.93 ± 11.94	2419.50 ± 10.21	2979.56 ± 11.33
V+γ+jets	3822.60 ± 18.03	2632.90 ± 17.27	3451.35 ± 17.95
Multi-jets	3707.84 ± 45.24	2526.80 ± 39.08	3330.98 ± 43.89
W+jets (Sh 2.2.11)	66296.97 ± 217.57	47382.47 ± 186.36	59265.37 ± 215.59
Total SM	95753.92 ± 245.37	69240.83 ± 207.17	86516.30 ± 239.12
Observed data	93915.00 ± 306.46	67727.00 ± 260.24	84833.00 ± 291.26

Table A.9: The **post-fit** yield yield of each process in the various **muon** measurement regions using CR leading jet p_{T} >**500** GeV.

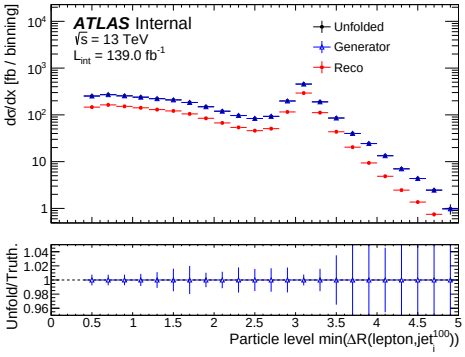
A.5 Studies on unfolding

As an addition to Chapter 10, this appendix contains all the auxiliary studies performed to understand the impact of the various choices made in this analysis on the unfolded results.

A.5.1 Closure tests

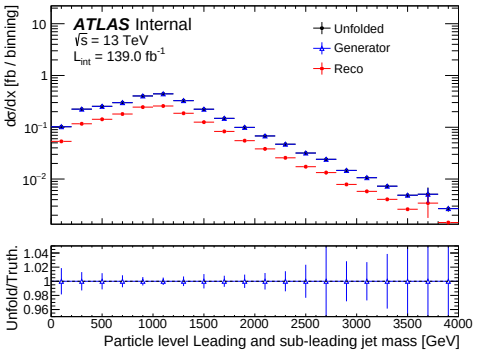
The unfolding procedure is validated by comparing the unfolded reconstruction-level distributions to their true values. As can be seen in Figures A.34 to A.36, perfect agreement is obtained between the particle-level prediction and the reconstructed value unfolded through its own response matrix. Similar conclusions are drawn in the muon channel, shown in Figures A.37 to A.39. The SHERPA 2.2.11 generator is used to perform the closure tests, and two iterations are used during the unfolding procedure.

To probe the sensitivity of closure test on the reconstruction-level distributions, a global scale factor and bin-by-bin poisson variation are introduced separately to the MC W+jets reconstruction-level distributions prior the unfolding closure test. Figure A.40 show the unfolded distribution for different input reconstruction-level distributions scaled by ranges of global scale factors from 1% to 50% in both directions. The unfolded results remain unchange over the number of iteration in the unfolding. Closure can be recovered by scaling the unfolded results by the inverse of the same SF used in the input distribution as shown in Fig. A.41. This results indicate the unfolding procedure preserves linearity for reconstruction-level distributions that are scaled by an overall



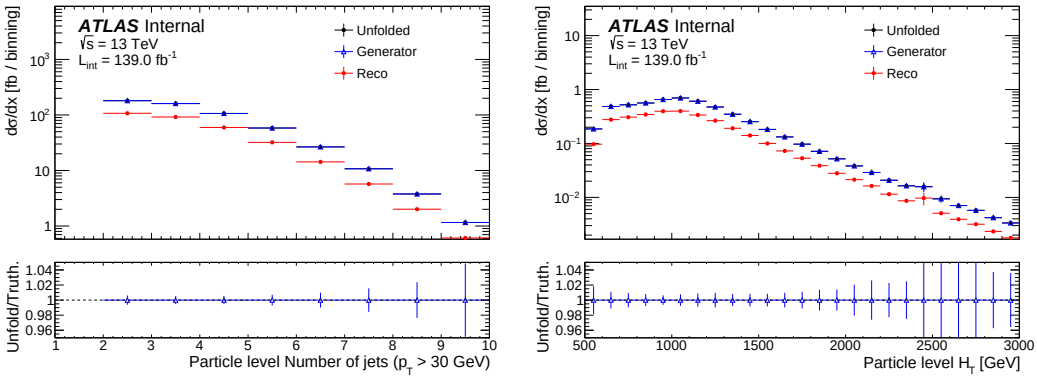
(a)

Figure A.34: Differential distributions shown closure tests of the unfolding procedure. The SHERPA 2.2.11 generator is used to perform the closure tests. Uncertainties are statistical only.

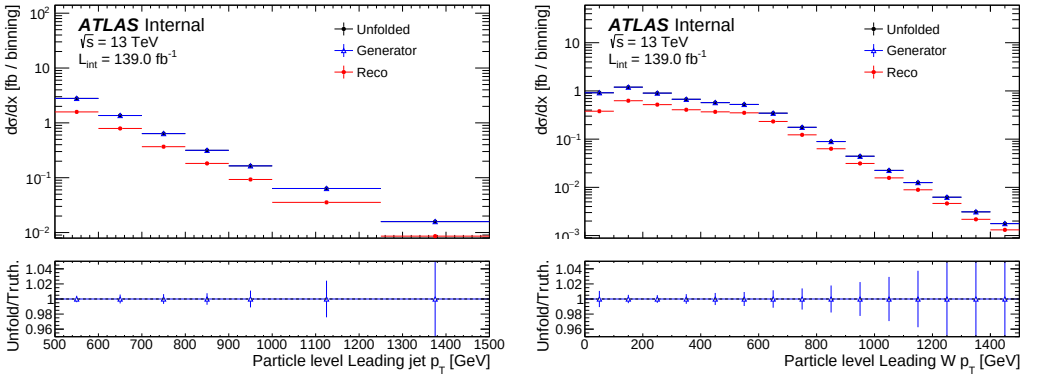


(a)

Figure A.35: Differential distributions shown closure tests of the unfolding procedure. The SHERPA 2.2.11 generator is used to perform the closure tests. Uncertainties are statistical only.

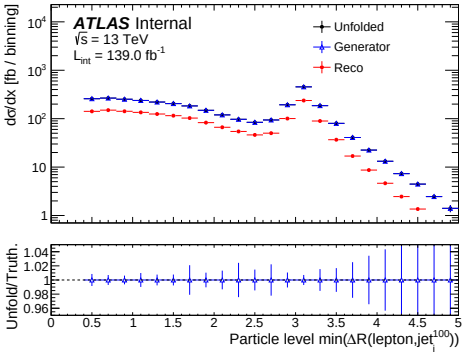


(a) (b)



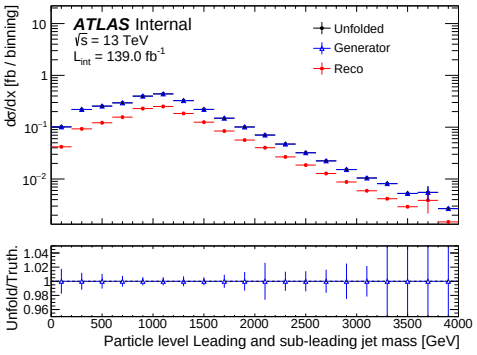
(c) (d)

Figure A.36: Differential distributions shown closure tests of the unfolding procedure. The SHERPA 2.2.11 generator is used to perform the closure tests. Uncertainties are statistical only.



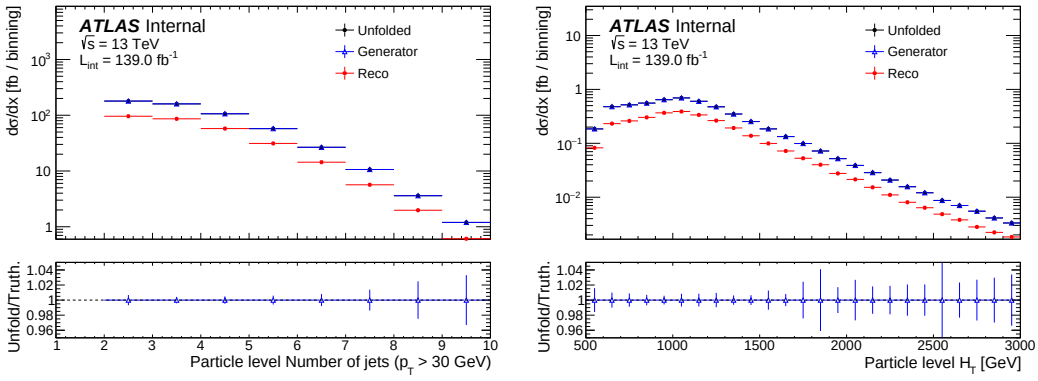
(a)

Figure A.37: Differential distributions shown closure tests of the unfolding procedure. The SHERPA 2.2.11 generator is used to perform the closure tests. Uncertainties are statistical only.

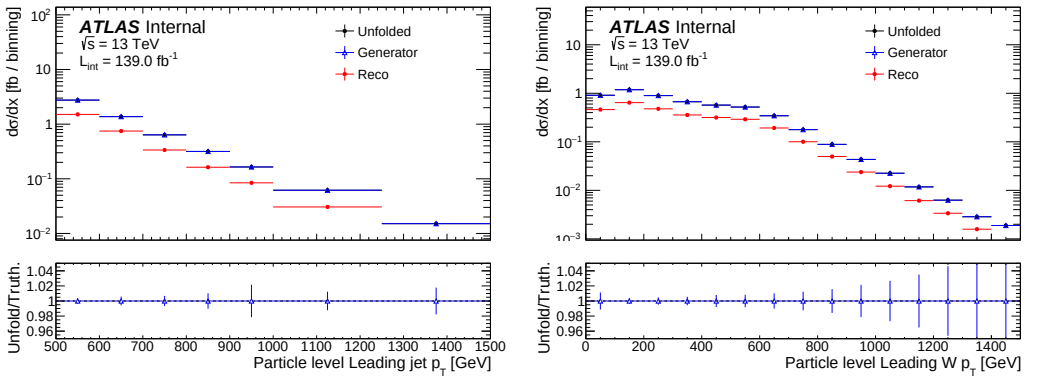


(a)

Figure A.38: Differential distributions shown closure tests of the unfolding procedure. The SHERPA 2.2.11 generator is used to perform the closure tests. Uncertainties are statistical only.



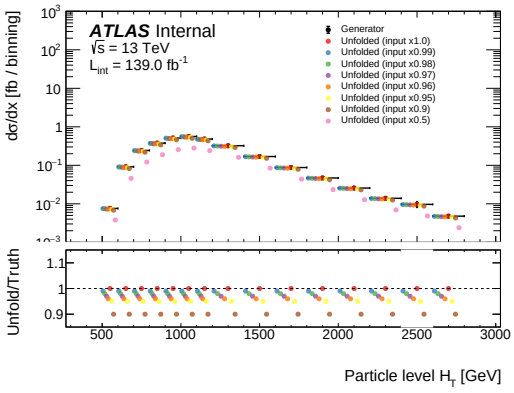
(a) (b)



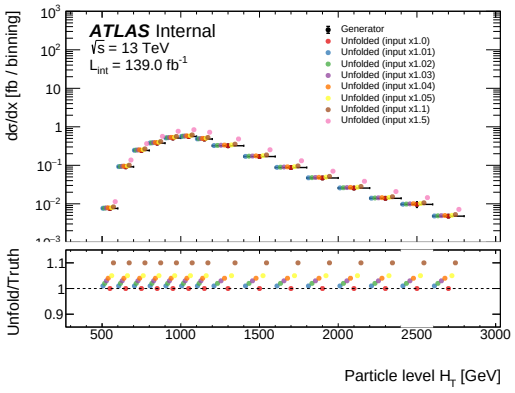
(c) (d)

Figure A.39: Differential distributions shown closure tests of the unfolding procedure. The SHERPA 2.2.11 generator is used to perform the closure tests. Uncertainties are statistical only.

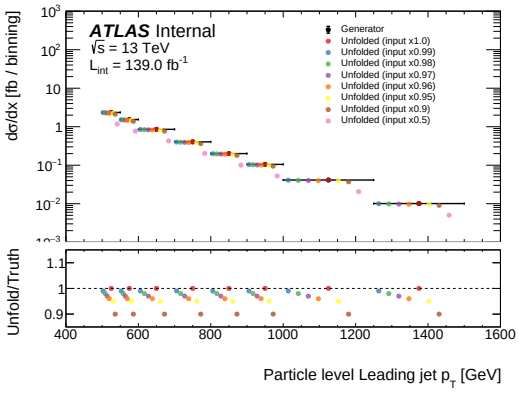
normalization constant.



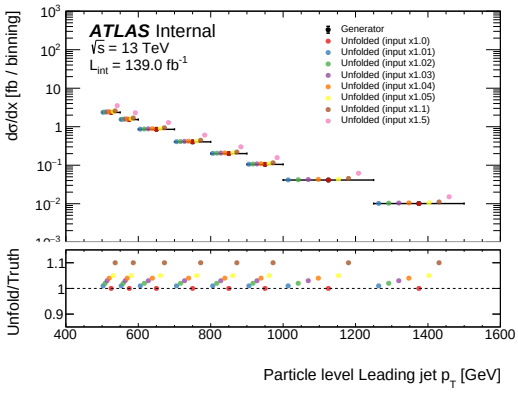
(a) scaled **down** by global SF



(b) scaled **up** by global SF

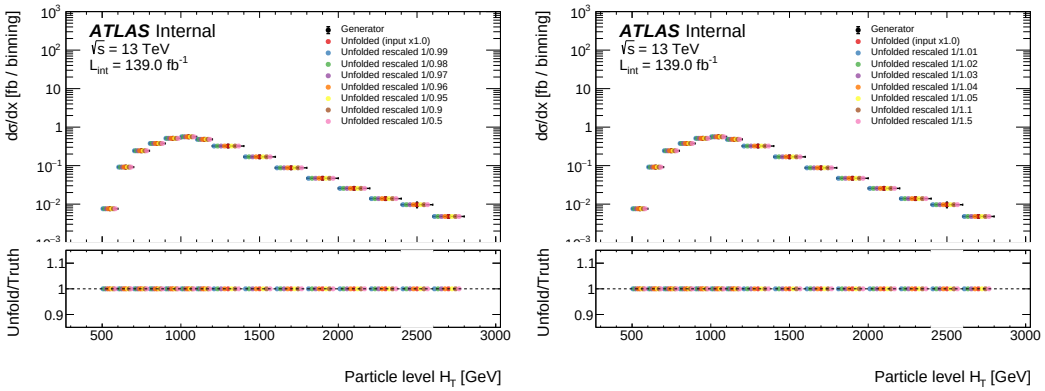


(c) scaled **down** by global SF



(d) scaled **up** by global SF

Figure A.40: Unfolded results for reconstruction-level distributions scaled by ranges of global SF. Only MC W+ jets samples are used, and the global SF is applied only to the reconstruction-level distributions prior the unfolding closure test.



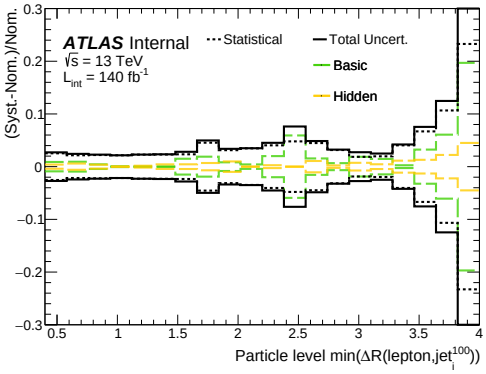
(a) (b)

Figure A.41: Unfolded results rescaled by the inverse of the same SF used in the input reconstruction-level distributions.

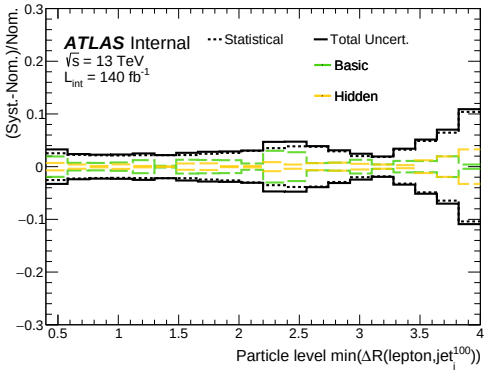
A.5.2 Unfolding uncertainties

A.5.2.1 Basic and hidden variable uncertainties

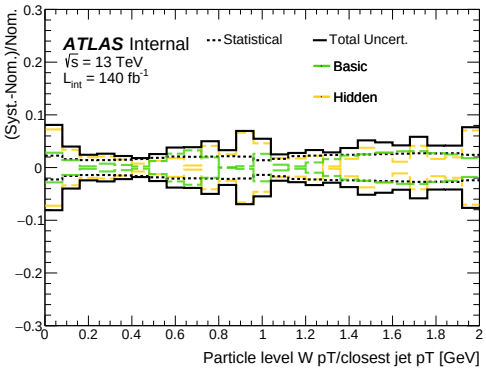
In this section, the basic and hidden unfolding uncertainties described in Section 11.4 for observables of interested are shown in Figures A.42 to A.45. They are displayed alongside the statistical uncertainty.



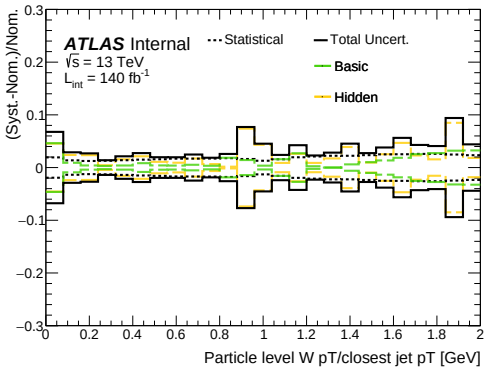
(a) e



(b) μ

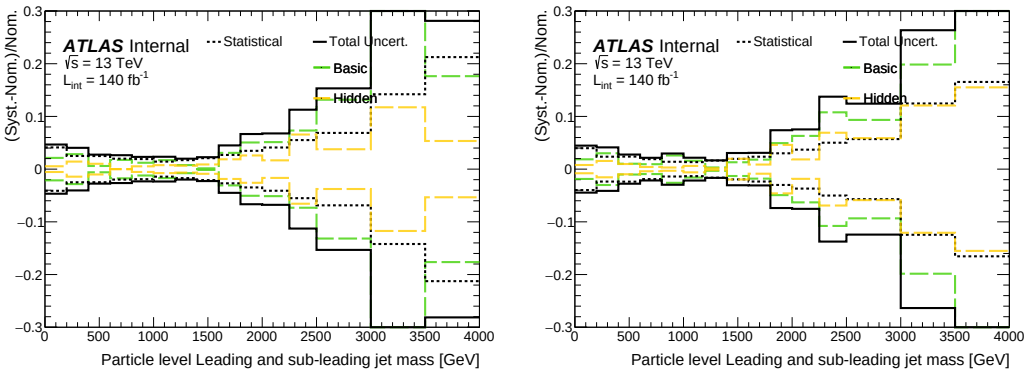


(c) e



(d) μ

Figure A.42: Basic and hidden unfolding uncertainties on unfolded $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$ and the ratio of the W-boson candidate over the closest jet p_{T} distributions in electron and muon channel in the **inclusive** phase-space.



(a) e (b) μ

Figure A.43: Basic and hidden unfolding uncertainties on unfolded m_{jj} distribution of electron and muon channel in the **inclusive 2-jets** phase-space.

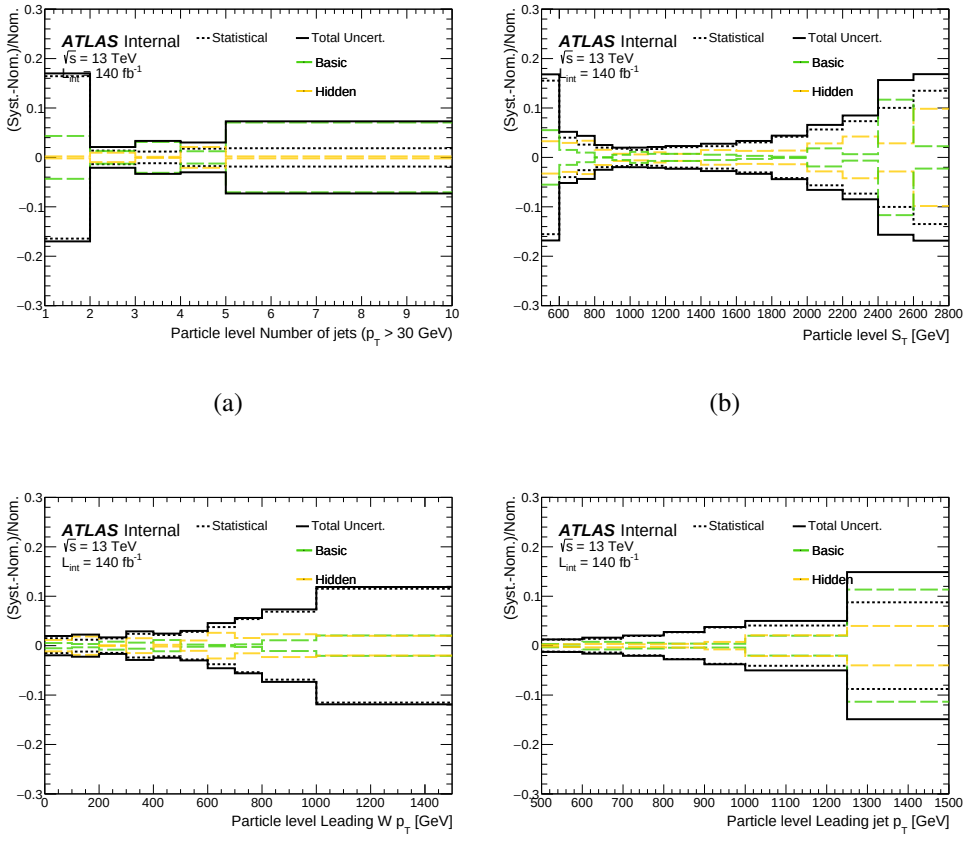


Figure A.44: Basic and hidden unfolding uncertainties on unfolded observables of **electron** channel in the **collinear** phase-space.

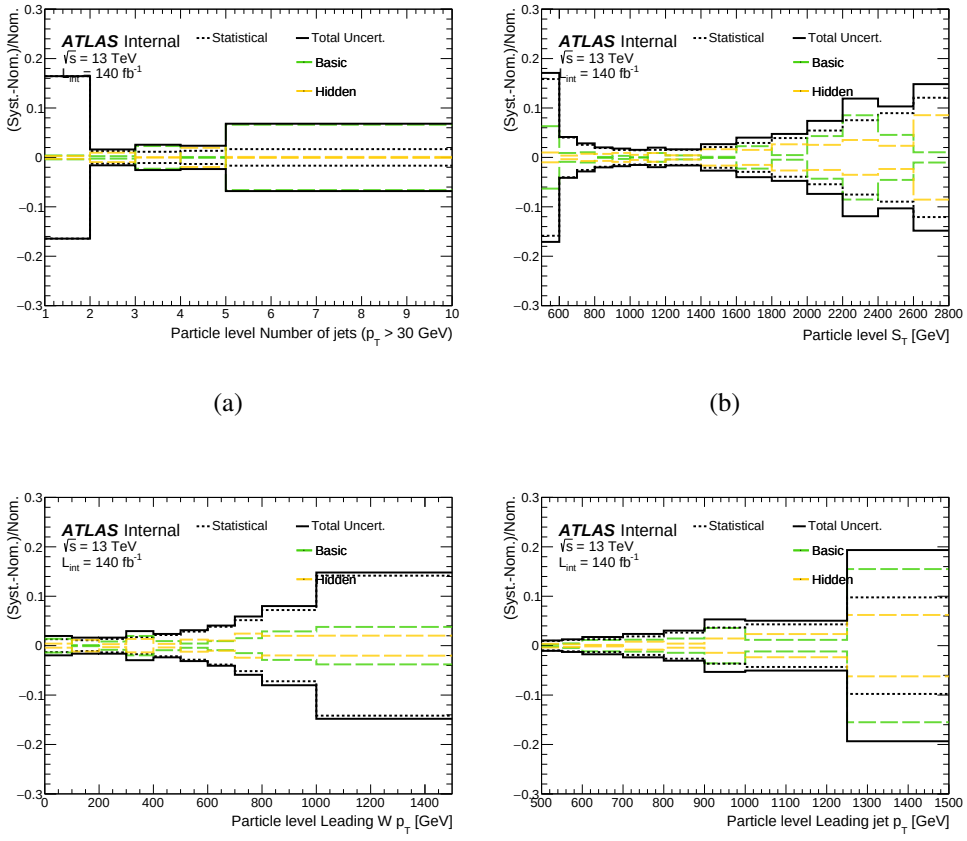
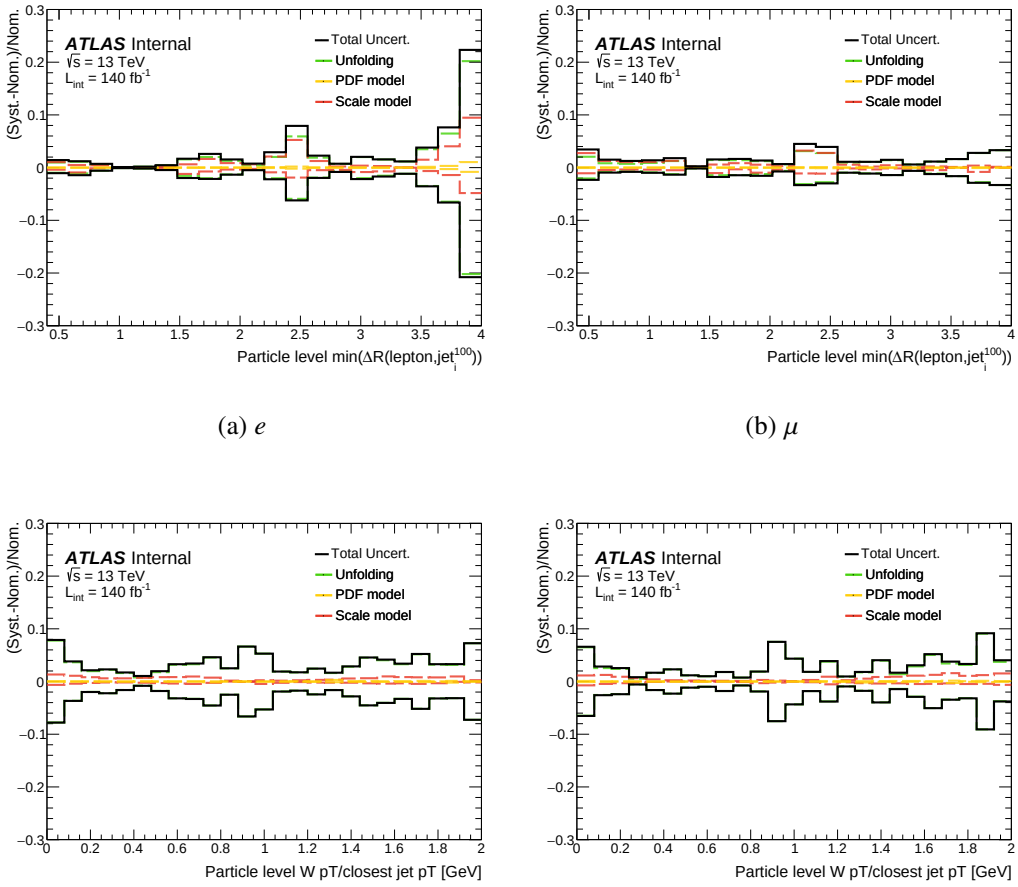


Figure A.45: Basic and hidden unfolding uncertainties on unfolded observables of **muon** channel in the **collinear** phase-space.

A.5.2.2 Signal modeling uncertainties

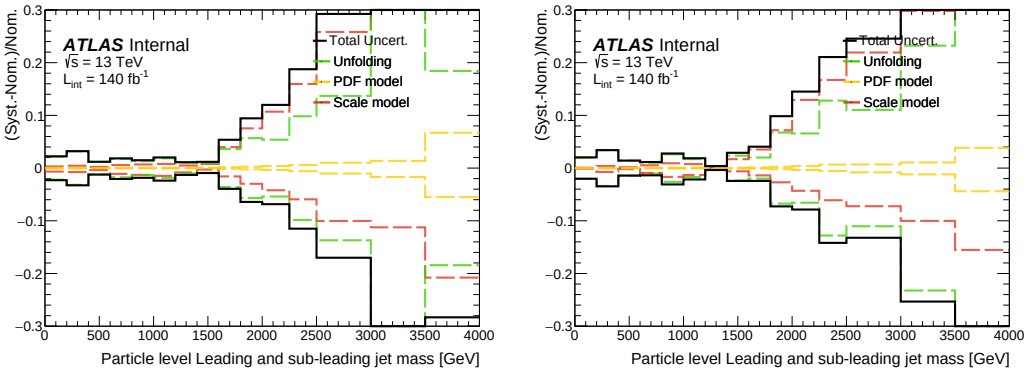
In addition to the basic and hidden variable uncertainties, an uncertainty on the measured cross section is assessed from signal theory uncertainties. The scale and PDF variations in the nominal SHERPA prediction are varied in the response matrix, and compared to the nominal unfolded cross section. As can be seen in Figures A.46 to A.49, the uncertainty is generally smaller than the combined unfolding uncertainty obtained above. The exception is in the inclusive 2j region, where high m_{jj} region above 1.5 TeV; in this region, the scale uncertainties are generally large due to the LO dependency of the SHERPA prediction in the higher jet multiplicity regions. Here, and for the highest jet multiplicity bin in the collinear region, we observe that the modeling uncertainties are simimlar in size to the unfolding uncertainties. In the highest m_{jj} bins, this uncertainty reaches 20%.



(a) e (b) μ

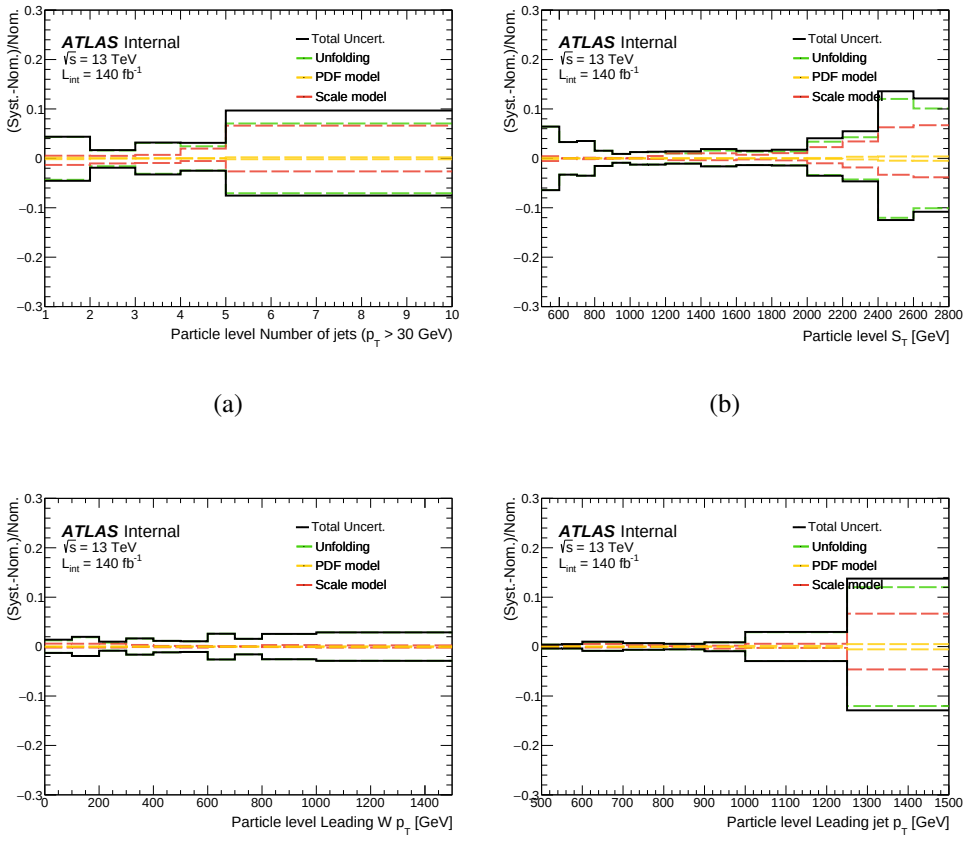
(c) e (d) μ

Figure A.46: Signal modeling uncertainties on unfolded $\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$ and the ratio of the W-boson candidate over the closest jet p_{T} distributions in electron and muon channel in the **inclusive** phase-space. The signal modeling uncertainties are shown together with the combined basic & hidden variable uncertainties, which are all added in quadrature into the total uncertainty band.



(a) e (b) μ

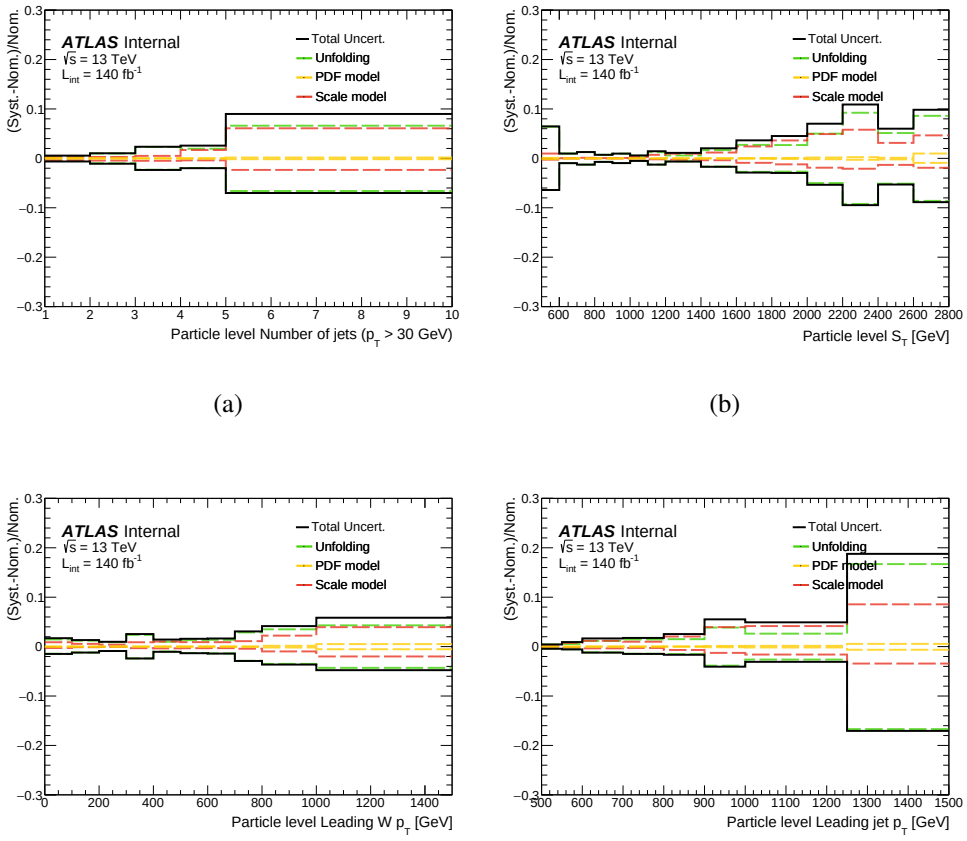
Figure A.47: Signal modeling uncertainties on unfolded m_{jj} distribution of electron and muon channel in the **inclusive 2-jets** phase-space. The signal modeling uncertainties are shown together with the combined basic & hidden variable uncertainties, which are all added in quadrature into the total uncertainty band.



(a) (b)

(c) (d)

Figure A.48: Signal modelling uncertainties on unfolded observables of **electron** channel in the **collinear** phase-space. The signal modeling uncertainties are shown together with the combined basic & hidden variable uncertainties, which are all added in quadrature into the total uncertainty band.



(a) (b)

(c) (d)

Figure A.49: Signal modeling unfolding uncertainties on unfolded observables of **muon** channel in the **collinear** phase-space. The signal modeling uncertainties are shown together with the combined basic & hidden variable uncertainties, which are all added in quadrature into the total uncertainty band.

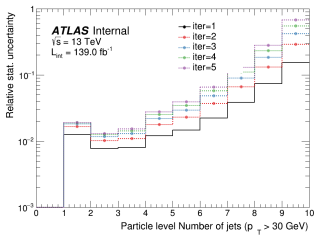
A.5.3 Number of iterations

The method chosen for performing the unfolding is an iterative bayesian approach, which has one free parameter: number of iterations. We need to assess the impact of going from $i \rightarrow i + 1$ iterations with respect to the statistical uncertainty we obtain through the unfolding process.

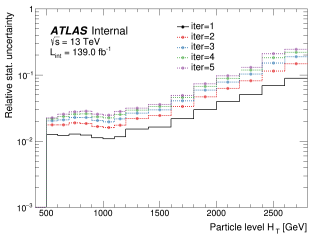
The relative statistical uncertainty as a function of the number of iterations is shown in Figures A.50 and A.51. As can be seen, the statistical uncertainty increases with increasing number of iterations.

Comparisons of unfolded differential cross-sections obtained when varying the number of iterations for various kinematic variables is shown in Figures A.52 and A.53 for the electron and muon channel, respectively. In these plots, the number of bayesian iterations for unfolding is chosen from $n = 2$ to $n = 5$, and displayed by dividing through by $n = 2$, our nominal number of iterations. For $n = 2$, the relative statistical uncertainty is drawn as a gray histogram along the $y = 1$ line. The variations in cross-sections are statistically compatible with the $n = 2$ iteration, with the exception of N_{jet}, where there is around an 8% difference. However, this difference is well covered by the unfolding uncertainties shown above in Figures A.44 and A.45. Given the otherwise stable behavior as a function of the number of iterations, we choose to unfold with $n = 2$ unfolding iterations to keep the statistical uncertainty at a minimum.

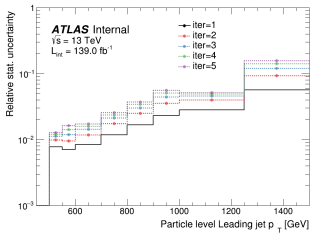
As an additional study on the number of iterations, the unfolding uncertainties and their dependency on the number of iterations was checked. The variation in the



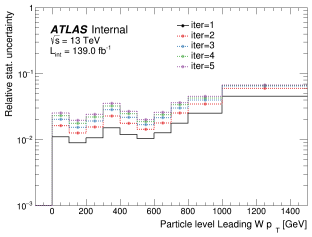
(a) N_{jet}



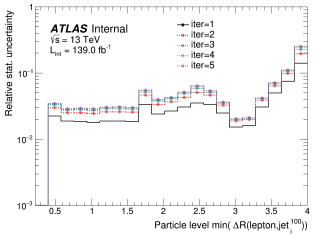
(b) S_{T}



(c) $p_{\text{T}}^{\text{jet}}$

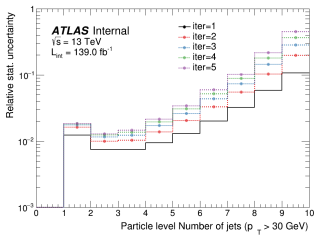


(d) W p_{T}

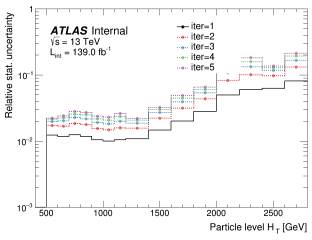


(e) $\Delta R_{\text{min}_i}(\ell, \text{jet}_i^{100})$

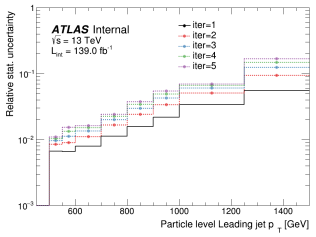
Figure A.50: Comparisons of relative statistical uncertainty in the inclusive phase-space of the electron channel. The number of bayesian iterations is shown for $n = i, i \in 1, 2, 3, 4, 5$ and overlaid as histograms with points. For all except N_{jet}, the size of the uncertainty is close to saturation for $n > 2$.



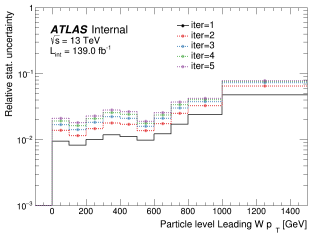
(a) N_{jet}



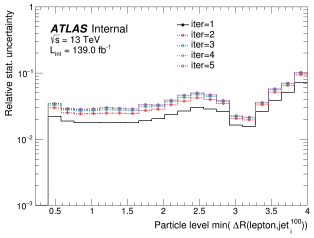
(b) S_{T}



(c) $p_{\text{T}}^{\text{jet}}$

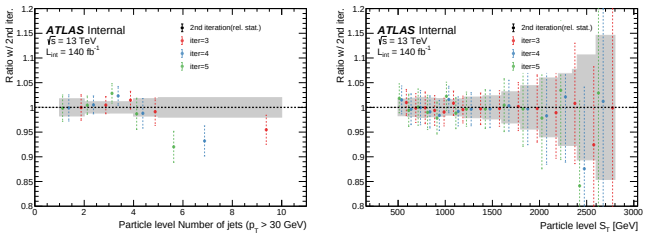


(d) W p_{T}

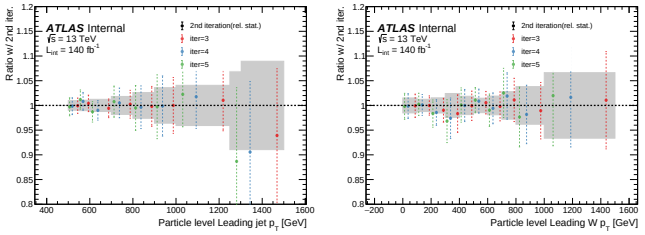


(e) $\Delta R_{\text{min}_i}(\ell, \text{jet}_i^{100})$

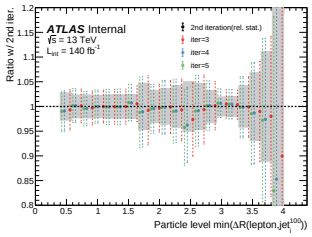
Figure A.51: Comparisons of relative statistical uncertainty in the inclusive phase-space of the muon channel. The number of bayesian iterations is shown for $n = i, i \in 1, 2, 3, 4, 5$ and overlaid as histograms with points. For all except N_{jet}, the size of the uncertainty is close to saturation for $n > 2$.



(a) N_{jet} (b) S_{T}

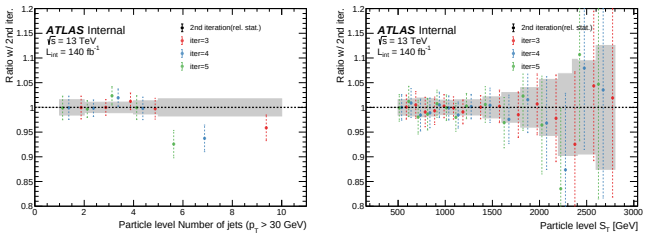


(c) $p_{\text{T}}^{\text{jet}}$ (d) W p_{T}

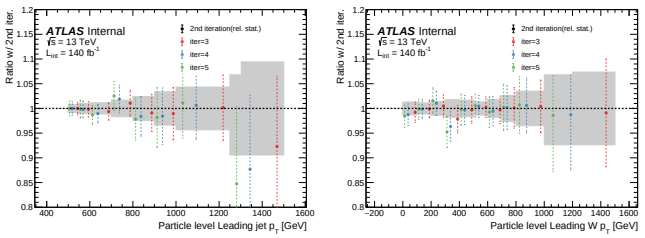


(e) $\Delta R_{\text{min}_i}(\ell, \text{jet}_i^{100})$

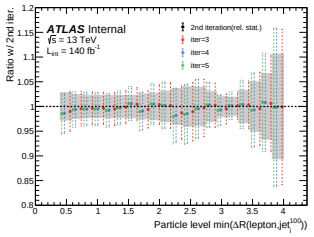
Figure A.52: Comparisons of differential cross-sections for various kinematic variables in the inclusive phase-space of the electron channel. The number of bayesian iterations for unfolding is chosen from $n = 2$ to $n = 5$, divided through by $n = 2$ and overlaid as points. For $n = 2$, the relative statistical uncertainty is drawn as a gray histogram along the $y = 1$ line. For all except N_{jet}, the variations in cross-sections are statistically compatible between iterations.



(a) N_{jet} (b) S_{T}



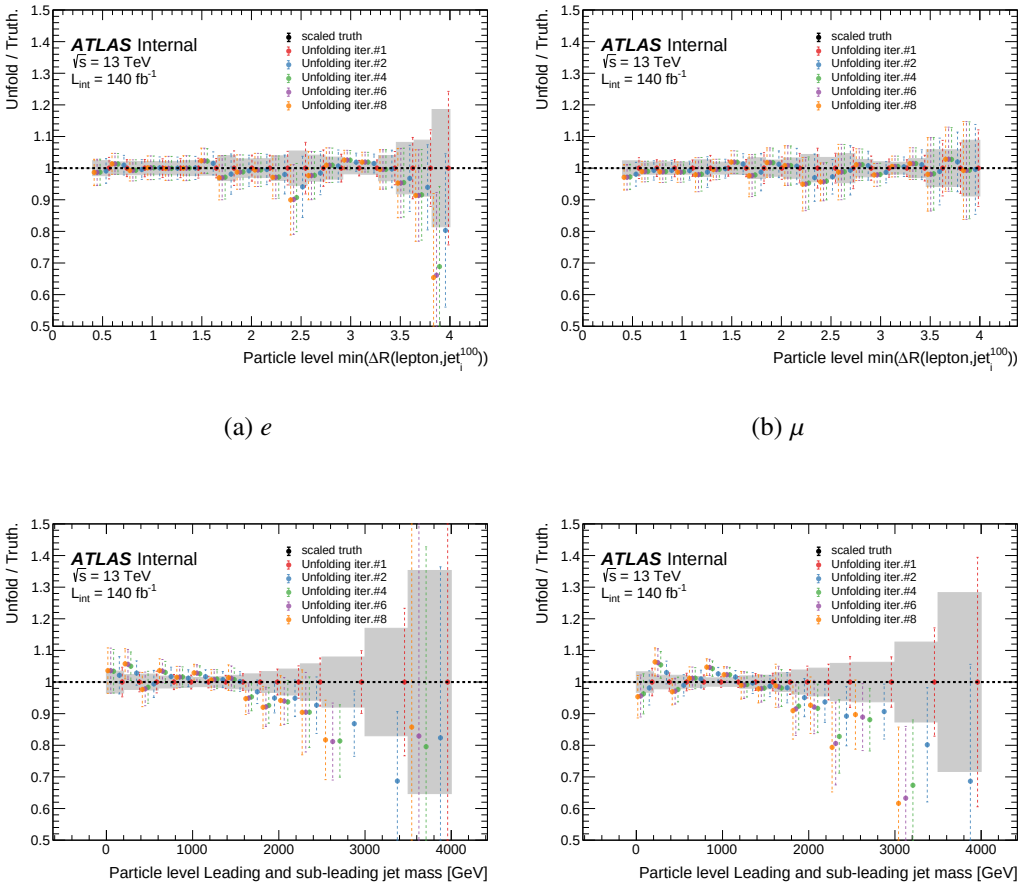
(c) $p_{\text{T}}^{\text{jet}}$ (d) W p_{T}



(e) $\Delta R_{\text{min}_i}(\ell, \text{jet}_i^{100})$

Figure A.53: Comparisons of differential cross-sections for various kinematic variables in the inclusive phase-space of the muon channel. The number of bayesian iterations for unfolding is chosen from $n = 2$ to $n = 5$, divided through by $n = 2$ and overlaid as points. For $n = 2$, the relative statistical uncertainty is drawn as a gray histogram along the $y = 1$ line. For all except N_{jet}, the variations in cross-sections are statistically compatible between iterations.

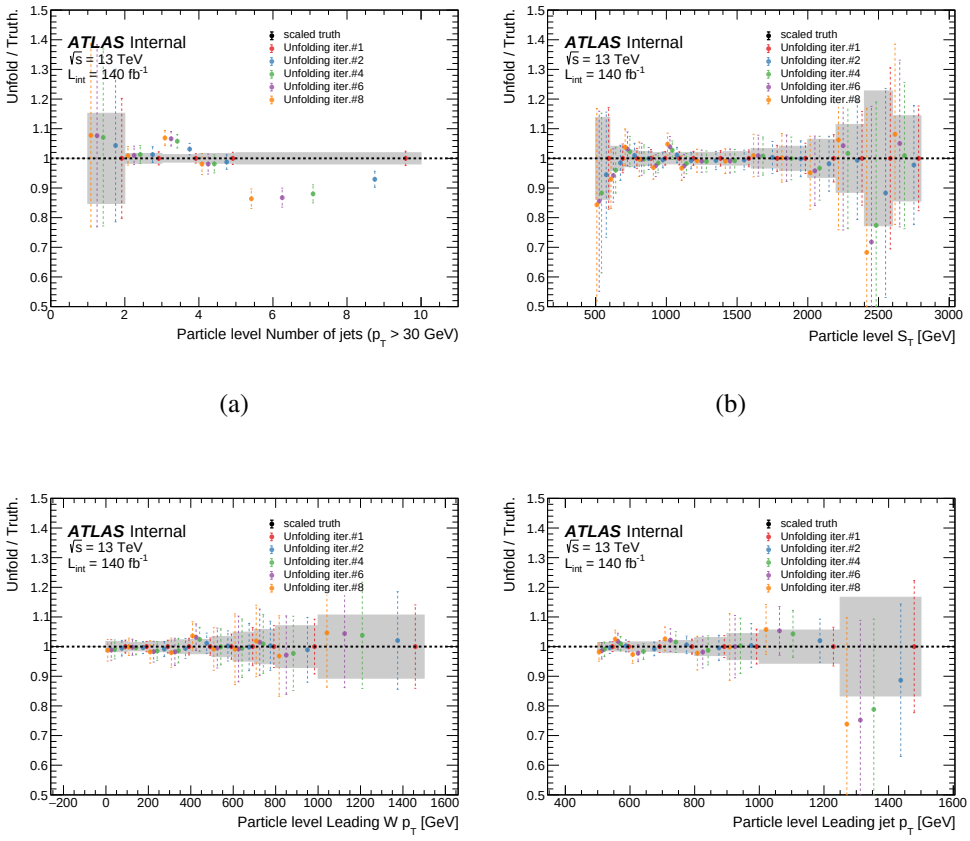
uncertainty for the basic unfolding uncertainty is shown in Figures A.54 to A.56. The dash lines are the relative statistical uncertainty on the unfolded results and the grey band is the relative statistical uncertainty on the known scaled truth distribution used in the re-weighting. We observe that the basic unfolding uncertainty slightly increases with the number of unfolding iterations. However, all of the iterations remain within the statistical fluctuations of the $n = 2$ iteration; moreover, similar to above the higher number of iterations come with larger statistical uncertainties, all of which overlap with the $n = 2$ iteration. Finally, the variation in the uncertainty for the hidden unfolding uncertainty is shown in Figures A.57 to A.59. The hidden unfolding uncertainty has less variation in terms of the number of unfolding iterations and saturated at earlier iteration.



(a) e (b) μ

(c) e (d) μ

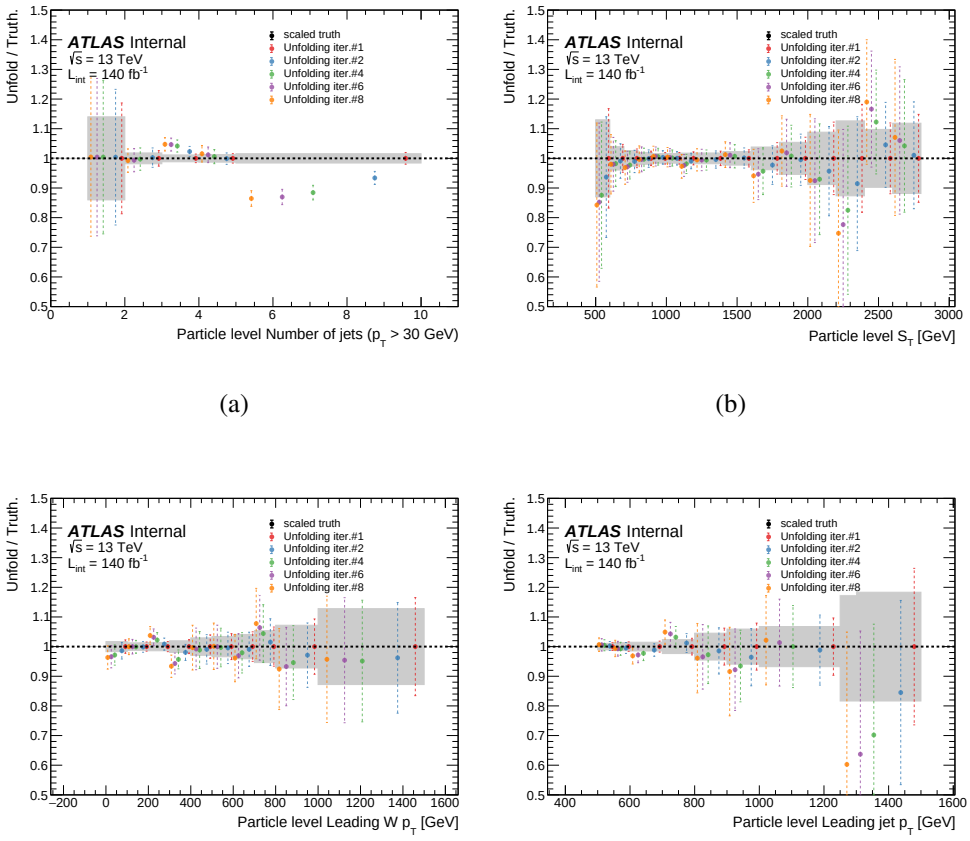
Figure A.54: **Basic** unfolding uncertainty in terms of number of unfolding iterations for $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$ and m_{jj} distributions in the **inclusive(2j)** phase-space of electron and muon channels. The grey band shows the relative statistical uncertainty on the known scaled truth distribution used in the re-weighting.



(a) (b)

(c) (d)

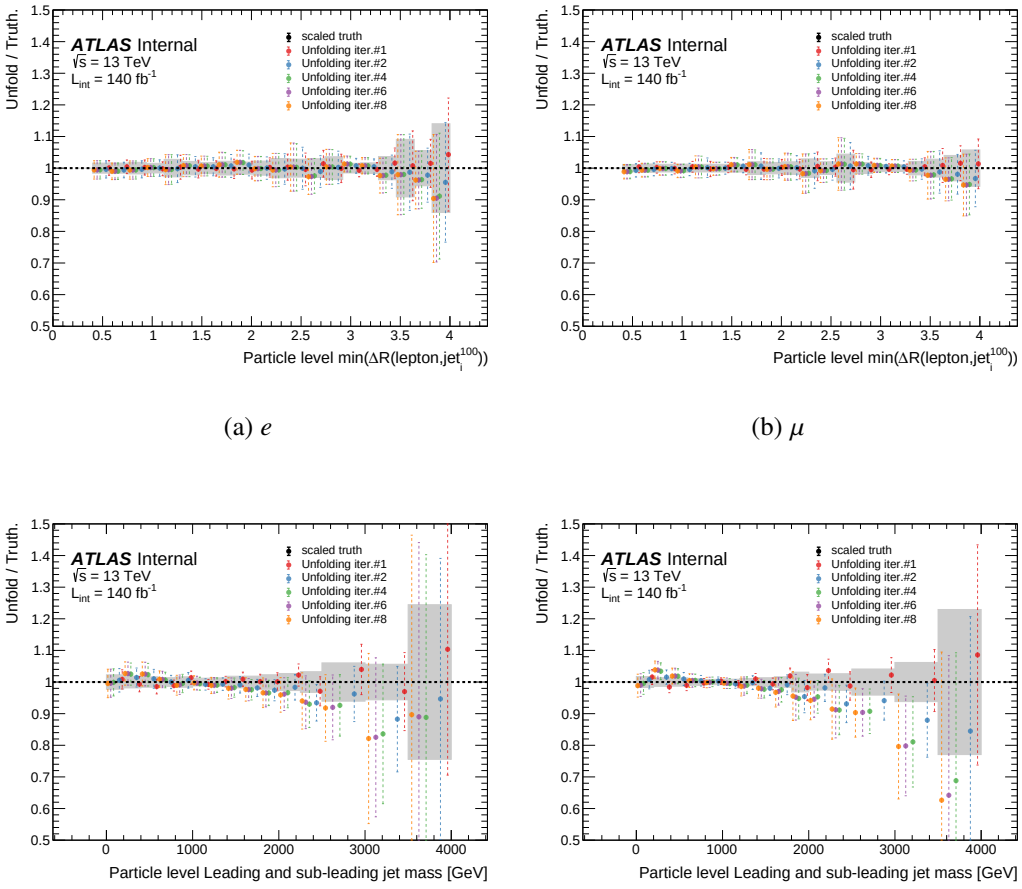
Figure A.55: **Basic** unfolding uncertainty in terms of number of unfolding iterations for observables in the **collinear** phase-space of **electron** channel. The grey band shows the relative statistical uncertainty on the known scaled truth distribution used in the re-weighting.



(a) (b)

(c) (d)

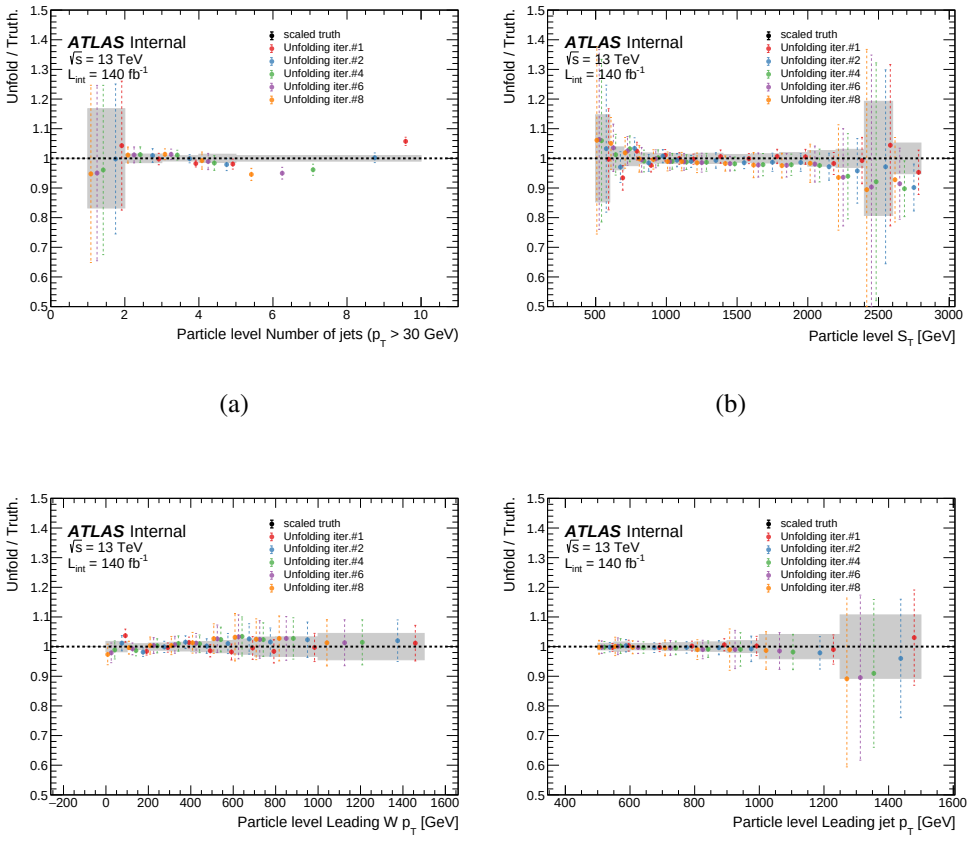
Figure A.56: **Basic** unfolding uncertainty in terms of number of unfolding iterations for observables in the **collinear** phase-space of **muon** channel. The grey band shows the relative statistical uncertainty on the known scaled truth distribution used in the re-weighting.



(a) e (b) μ

(c) e (d) μ

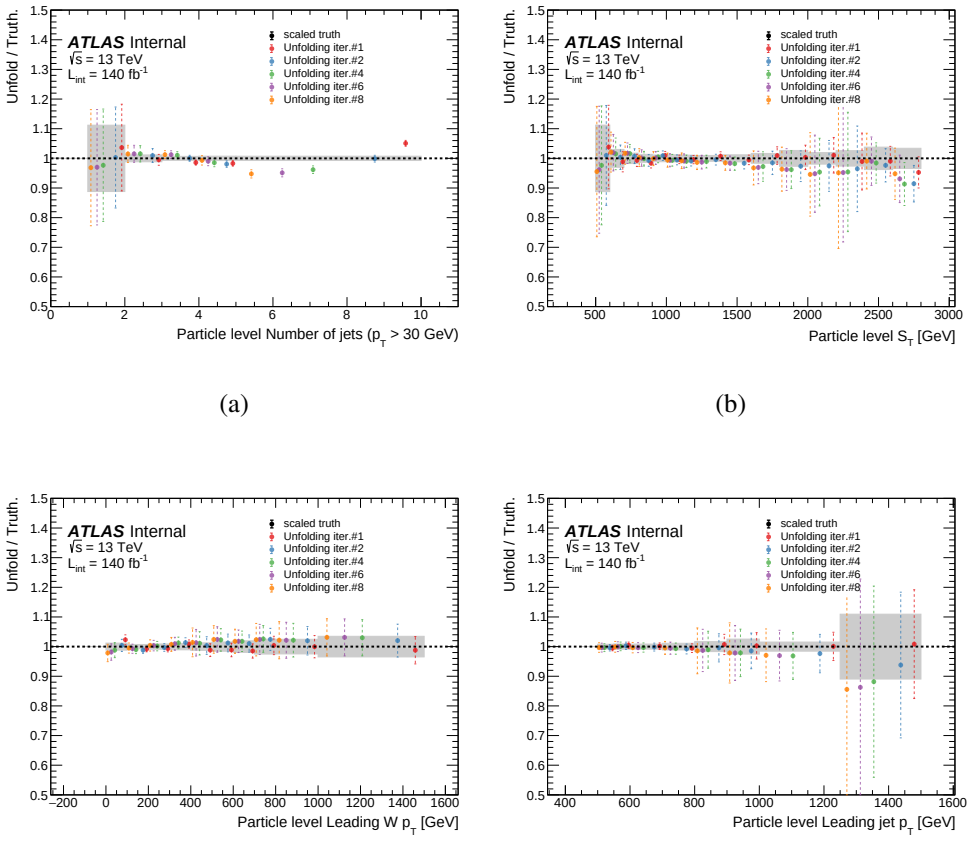
Figure A.57: **Hidden** unfolding uncertainty in terms of number of unfolding iterations for $\Delta R_{\min_i}(\ell, \text{jet}_i^{100})$ and m_{jj} distributions in the **inclusive(2j)** phase-space of electron and muon channels. The grey band shows the relative statistical uncertainty on the known scaled truth distribution used in the re-weighting.



(a) (b)

(c) (d)

Figure A.58: **Hidden** unfolding uncertainty in terms of number of unfolding iterations for observables in the **collinear** phase-space of **electron** channel. The grey band shows the relative statistical uncertainty on the known scaled truth distribution used in the re-weighting.



(a) (b)

(c) (d)

Figure A.59: **Hidden** unfolding uncertainty in terms of number of unfolding iterations for observables in the **collinear** phase-space of **muon** channel. The grey band shows the relative statistical uncertainty on the known scaled truth distribution used in the re-weighting.

A.5.4 Binning of differential distributions

The binning of the differential cross-sections is determined by requiring that the statistical uncertainty in each bin is below 10% and the purity is better then 50%. The chosen bin edges that are used are shown in Table A.10

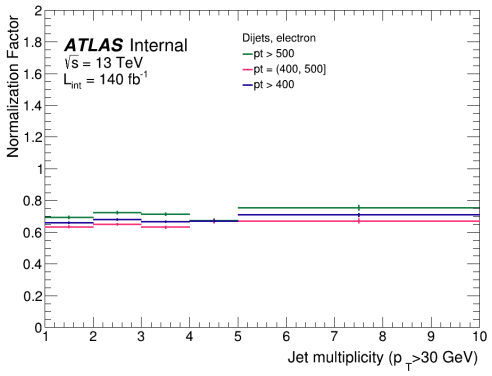
Observable	Bin Edges
Number of jets ($p_{\mathrm{T}} > 30$GeV)	[1.0,2.0,3.0,4.0,5.0,6.0,7.0,8.0,9.0,10.0]
Leading W p_{T} [GeV]	[0.0,100.0,200.0,300.0,400.0,500.0,600.0,700.0,800.0,1000.0,1500.0]
S_{T} [GeV]	[500.0,600.0,700.0,800.0,900.0,1000.0,1100.0,1200.0,1400.0,1600.0,1800.0,2000.0,2200.0,2400.0,2600.0,2800.0]
Leading and sub-leading jet mass (m_{jj}) [GeV]	[0.0,200.0,400.0,600.0,800.0,1000.0,1200.0,1400.0,1600.0,1800.0,2000.0,2250.0,2500.0,3000.0,3500.0,4000.0]
Leading jet p_{T} [GeV]	[500.0,550.0,600.0,700.0,800.0,900.0,1000.0,1250.0,1500.0]
$\Delta R_{\mathrm{min}_i}(\ell, \mathrm{jet}_i^{100})$	[0.0,0.2,0.4,0.6,0.8,1.0,1.2,1.4,1.6,1.8,2.0,2.2,2.4,2.6,2.8,3.0,3.2,3.4,3.6,3.8,4.0,4.2,4.4,4.6,4.8,5.0]

Table A.10: Binning used for observables.

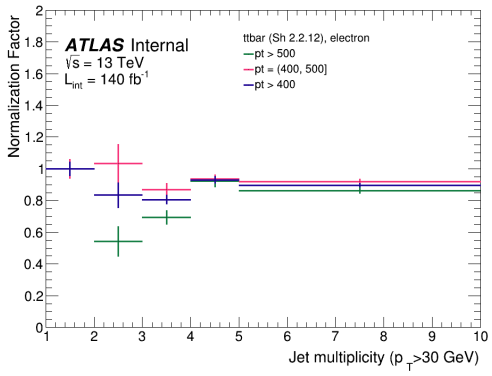
A.5.5 Prompt background correction factors for underflow bins

The nominal leading jet p_{T} requirement in our measurement is p_{T}> 500 GeV. As previously discussed, reconstructed events below this threshold can migrate into our fiducial measurement phase-space and need to be taken into account. When unfolding data, we therefore need to estimate the background in the $400 < p_T < 500$ GeV phase-space. This section compares the data-driven normalization factors obtained in the jet selections that define the nominal jet selection (p_{T}> 500 GeV) and the inclusive and exclusive selections on the underflow bin.

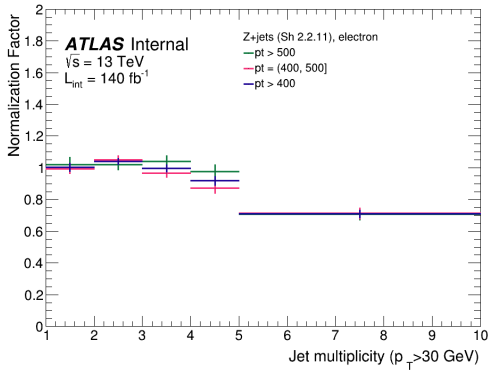
The variation in the normalization factors for different thresholds can be seen in Figure A.60 and A.61 for electrons and muons, respectively. As can be seen, most normalization factors agree well between the different selections. In the final analysis, the p_{T}> 400 GeV set of normalization factors are used.



(a) Dijets

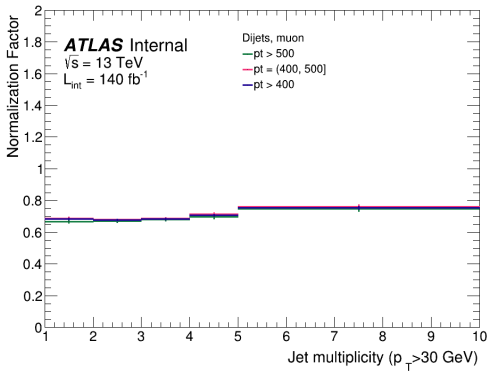


(b) $t\bar{t}$

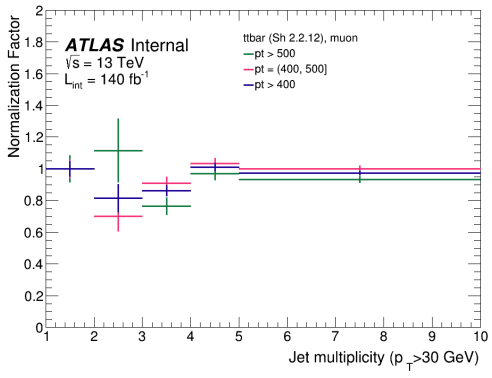


(c) Z+jets

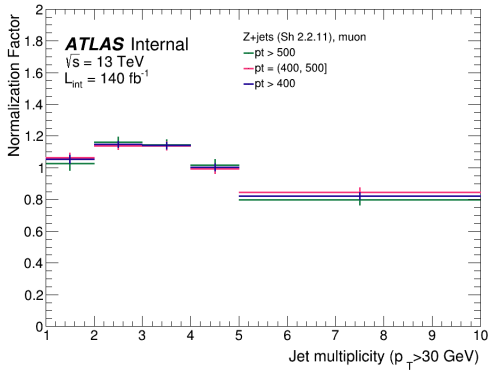
Figure A.60: Semi-data driven correction factors for the prompt backgrounds in the **electron channel** with various leading jet p_{T} requirements.



(a) Dijets



(b) $t\bar{t}$



(c) Z+jets

Figure A.61: Semi-data driven correction factors for the prompt backgrounds in the **muon channel** with various leading jet p_{T} requirements.

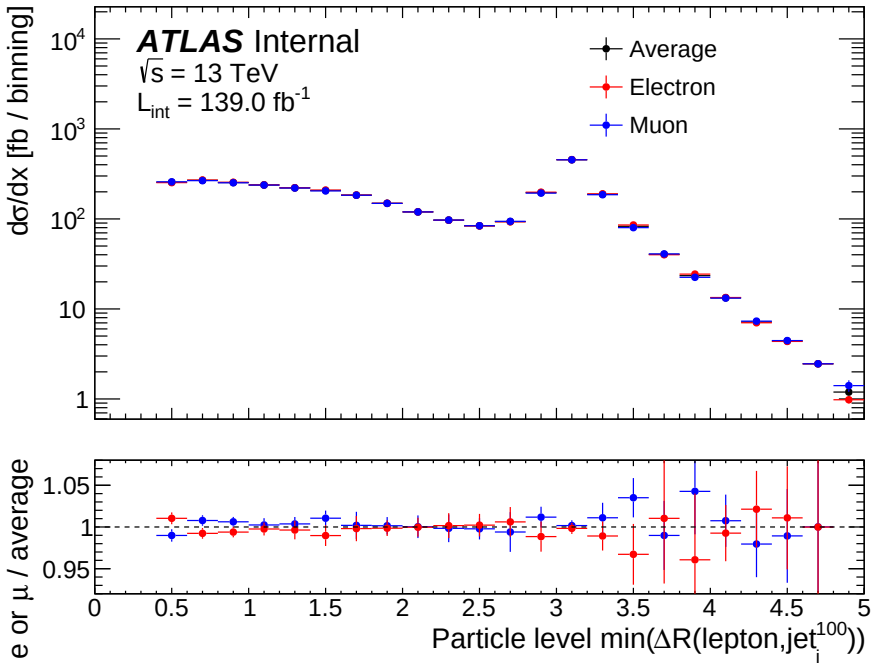
A.5.6 Combination of electron and muon channels cross-sections

The cross-sections reported in this analysis are a combination of the measured values in the electron and muon channels. This is validated for both the theory predictions and the measured cross-sections in data.

:

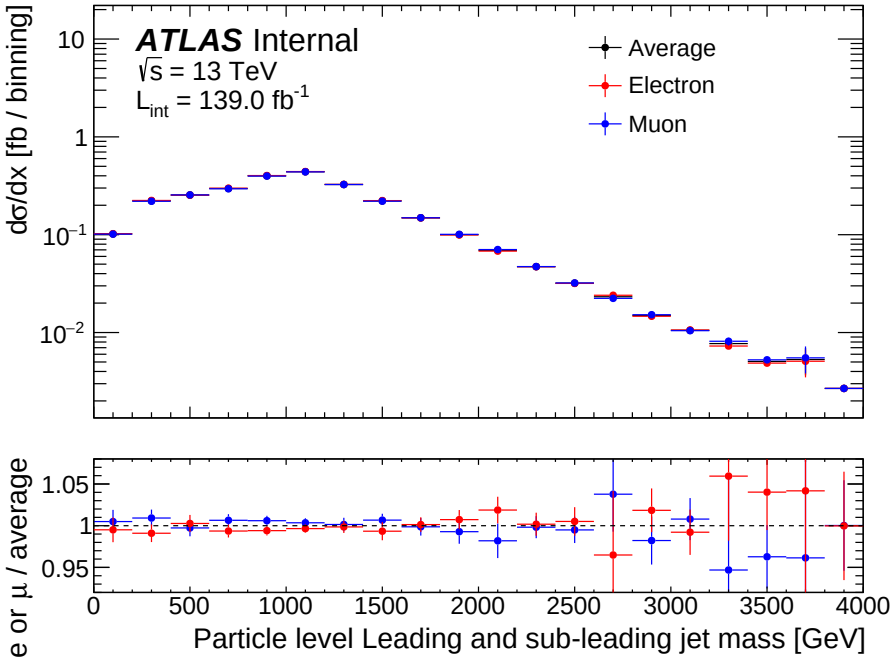
A.5.6.1 Particle level predictions

The cross-sections predicted by the MC generators are similar between the electron and muon channel. This is checked for the default generator predictions that are compared against the data. The comparison of SHERPA 2.2.11 is shown in Figures A.62 to A.64. The largest differences are a few perfect in the tails.



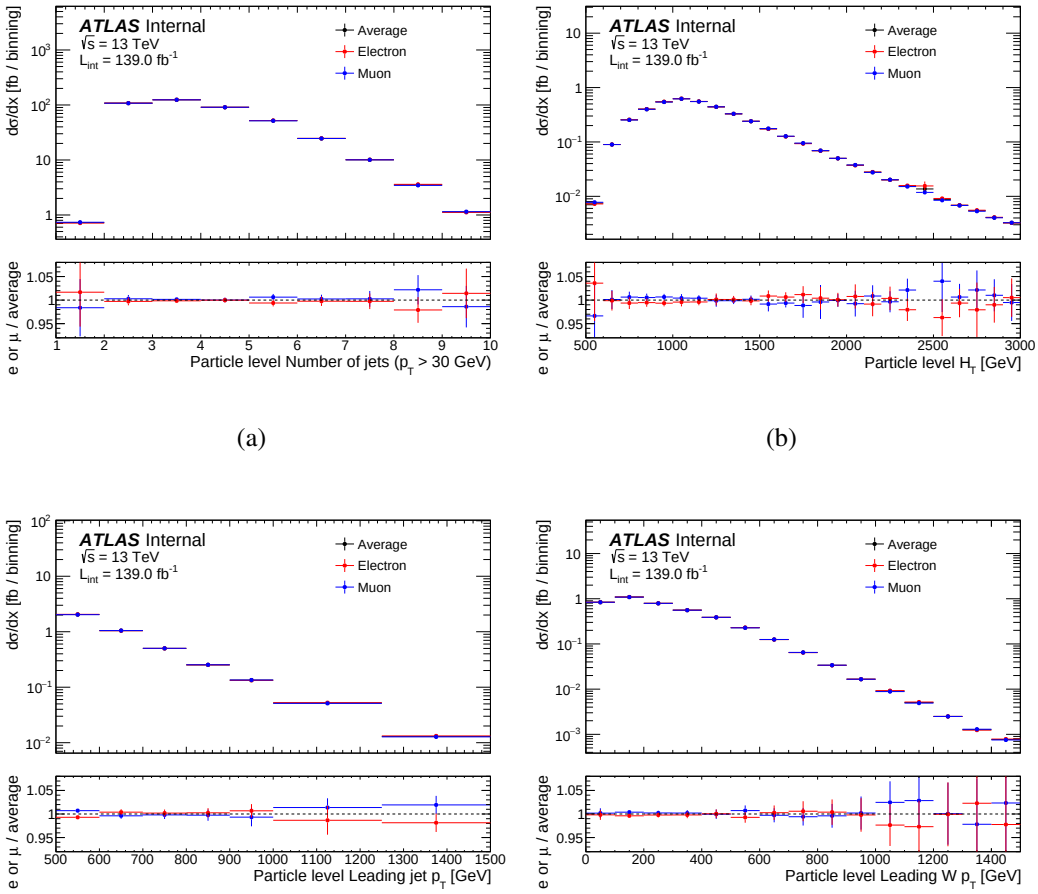
(a)

Figure A.62: Theory predictions for the average cross-section, as well as the individual electron and muon channels the inclusive phase-space. Ratios are taken with respect to the average cross-section.



(a)

Figure A.63: Theory predictions for the average cross-section, as well as the individual electron and muon channels the inclusive 2-jet phase-space. Ratios are taken with respect to the average cross-section.



(a) (b)

(c) (d)

Figure A.64: Theory predictions for the average cross-section, as well as the individual electron and muon channels the collinear phase-space. Ratios are taken with respect to the average cross-section

Bibliography

[1] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Physics Letters B **816** (2021) 136204, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269321001441 (cit. on p. 2).

[2] ATLAS Collaboration, *Measurement of the production cross section for a Higgs boson in association with a vector boson in the $H \to WW^* \to \ell\nu\ell\nu$ channel in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Lett. B **798** (2019) 134949, arXiv: 1903.10052 [hep-ex] (cit. on p. 2).

[3] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s}$ = 13 TeV with the ATLAS experiment*, Phys. Lett. B **810** (2020) 135797, arXiv: 2006.13076 [hep-ex] (cit. on p. 2).

[4] ATLAS Collaboration, *Measurements of top-quark pair spin correlations in the $e\mu$ channel at $\sqrt{s} = 13$ TeV using pp collisions in the ATLAS detector*, Eur. Phys. J. C **80** (2020) 754, arXiv: 1903.07570 [hep-ex] (cit. on p. 2).

[5] ATLAS Collaboration, *Search for squarks and gluinos in final states with one isolated lepton, jets, and missing transverse momentum at $\sqrt{s} = 13$ with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 600, arXiv: 2101.01629 [hep-ex] (cit. on p. 2).

[6] ATLAS Collaboration, *Searches for electroweak production of supersymmetric particles with compressed mass spectra in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, Phys. Rev. D **101** (2020) 052005, arXiv: 1911.12606 [hep-ex] (cit. on p. 2).

[7] ATLAS Collaboration, *Search for new phenomena with top quark pairs in final states with one lepton, jets, and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **04** (2021) 174, arXiv: 2012.03799 [hep-ex] (cit. on p. 2).

[8] M. Aaboud and G, *Measurement of W boson angular distributions in events with high transverse momentum jets at s=8 TeV using the ATLAS detector*, Physics Letters B **765** (2017) 132, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269316307419 (cit. on p. 3).

[9] ATLAS Collaboration, *Cross-section measurements for the production of a Z boson in association with high-transverse-momentum jets in pp collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector*, Journal of High Energy Physics **2023** (2023) 6, arXiv: 2205.02597 [hep-ex] (cit. on p. 3).

[10] ATLAS Collaboration, *Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment*, Journal of High Energy Physics **2022** (2022) 8, arXiv: 2112.09588 [hep-ex] (cit. on pp. 4, 44).

[11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on pp. 6, 8).

[12] *ATLAS Collaboration, Run-2 luminosity public results, (2020)*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (cit. on pp. 17, 18).

[13] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 19).

[14] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187 (cit. on p. 36).

[15] ATLAS Collaboration, *Luminosity monitoring using $Z \to \ell^+\ell^-$ events at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATL-DAPR-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2752951 (cit. on p. 37).

[16] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, (2022), arXiv: 2212.09379 [hep-ex] (cit. on p. 37).

[17] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017 (cit. on p. 37).

[18] *ATLAS Muon Performance*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonPublicResults (cit. on pp. 38, 188).

[19] *ATLAS Electron and Photon Performance*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ElectronGammaPublicCollisionResults (cit. on pp. 38, 188).

[20] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on pp. 42, 43).

[21] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on pp. 42, 43, 45).

[22] F. Cascioli, P. Maierhöfer, and S. Pozzorini,
Scattering Amplitudes with Open Loops, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph] (cit. on pp. 42, 43, 45).

[23] S. Schumann and F. Krauss,
A parton shower algorithm based on Catani–Seymour dipole factorisation, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph] (cit. on pp. 42, 43, 45).

[24] S. Höche, F. Krauss, M. Schönherr, and F. Siegert,
QCD matrix elements + parton showers. The NLO case, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph] (cit. on p. 42).

[25] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on pp. 42, 45, 46).

[26] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on pp. 42, 43).

[27] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on pp. 42, 46–48).

[28] V. Bertone, S. Carrazza, N. P. Hartland, and J. Rojo,
Illuminating the photon content of the proton within a global PDF analysis, SciPost Phys. **5** (2018) 008, arXiv: 1712.07053 [hep-ph] (cit. on pp. 42, 44).

[29] S. Frixione, P. Nason, and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph] (cit. on pp. 42, 45).

[30] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on pp. 42, 45–47).

[31] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on pp. 42, 45–47).

[32] 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 **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on pp. 42, 45–47).

[33] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv: 1007.2241 [hep-ph] (cit. on pp. 42, 46, 47).

[34] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex] (cit. on pp. 42, 47).

[35] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. D **82** (2010) 074018, arXiv: 1005.3457 [hep-ph] (cit. on p. 42).

[36] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on pp. 42, 44, 46, 48).

[37] G. Aad et al., *Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment*, (2021), arXiv: 2112.09588 [hep-ex] (cit. on p. 43).

[38] A. Denner, S. Dittmaier, and L. Hofer, *COLLIER: A fortran-based complex one-loop library in extended regularizations*, Comput. Phys. Commun. **212** (2017) 220, arXiv: 1604.06792 [hep-ph] (cit. on pp. 43, 45).

[39] J.-C. Winter, F. Krauss, and G. Soff, *A modified cluster-hadronization model*, Eur. Phys. J. C **36** (2004) 381, arXiv: hep-ph/0311085 (cit. on p. 43).

[40] E. Bothmann, M. Schönherr, and S. Schumann, *Reweighting QCD matrix-element and parton-shower calculations*, Eur. Phys. J. C **76** (2016) 590, arXiv: 1606.08753 [hep-ph] (cit. on p. 44).

[41] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 44).

[42] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on pp. 44, 46, 48, 191).

[43] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, JHEP **12** (2012) 061, arXiv: 1209.6215 [hep-ph] (cit. on p. 44).

[44] J. Andersen et al., *Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report*, (2014), arXiv: 1405.1067 [hep-ph] (cit. on p. 45).

[45] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph] (cit. on p. 45).

[46] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *QCD Matrix Elements + Parton Showers*, JHEP **11** (2001) 063, arXiv: hep-ph/0109231 (cit. on p. 45).

[47] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: https://cds.cern.ch/record/2216168 (cit. on p. 46).

[48] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, *Hadronic top-quark pair production with NNLL threshold resummation*, Nucl. Phys. B **855** (2012) 695, arXiv: 1109.1536 [hep-ph] (cit. on p. 46).

[49] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic*

soft-gluon resummation, Phys. Lett. B **710** (2012) 612,

arXiv: 1111.5869 [hep-ph] (cit. on p. 46).

[50] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001, arXiv: 1204.5201 [hep-ph] (cit. on p. 46).

[51] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054,

arXiv: 1207.0236 [hep-ph] (cit. on p. 46).

[52] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080,

arXiv: 1210.6832 [hep-ph] (cit. on p. 46).

[53] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004,

arXiv: 1303.6254 [hep-ph] (cit. on p. 46).

[54] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. **185** (2014) 2930,

arXiv: 1112.5675 [hep-ph] (cit. on p. 46).

[55] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*,

J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 46).

[56] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt,
Parton distributions for the LHC, Eur. Phys. J. C **63** (2009) 189,
arXiv: 0901.0002 [hep-ph] (cit. on pp. 46, 47).

[57] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*,
Eur. Phys. J. C **64** (2009) 653, arXiv: 0905.3531 [hep-ph] (cit. on pp. 46, 47).

[58] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*,
Phys. Rev. D **89** (2014) 033009, arXiv: 1302.6246 [hep-ph] (cit. on p. 46).

[59] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: 1009.2450 [hep-ph]
(cit. on p. 46).

[60] S. Frixione, E. Laenen, P. Motylinski, C. White, and B. R. Webber,
Single-top hadroproduction in association with a W boson, JHEP **07** (2008) 029,
arXiv: 0805.3067 [hep-ph] (cit. on p. 46).

[61] D. J. Lange, *The EvtGen particle decay simulation package*,
Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on pp. 46, 47).

[62] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. D **82** (2010) 054018,
arXiv: 1005.4451 [hep-ph] (cit. on p. 47).

[63] N. Kidonakis, "Top Quark Production," *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013)* (JINR, Dubna, Russia, July 15–28, 2013) 139, arXiv: 1311.0283 [hep-ph] (cit. on p. 47).

[64] P. Nason and G. Zanderighi, *W^+W^-, WZ and ZZ production in the POWHEG-BOX-V2*, Eur. Phys. J. C **74** (2014) 2702, arXiv: 1311.1365 [hep-ph] (cit. on p. 47).

[65] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP **07** (2002) 012, arXiv: hep-ph/0201195 (cit. on p. 47).

[66] S. Mrenna and P. Skands, *Automated parton-shower variations in PYTHIA 8*, Phys. Rev. D **94** (2016) 074005, arXiv: 1605.08352 [hep-ph] (cit. on p. 48).

[67] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton fragmentation and string dynamics*, Phys. Rept. **97** (1983) 31 (cit. on p. 48).

[68] T. Sjöstrand, *Jet fragmentation of multiparton configurations in a string framework*, Nucl. Phys. B **248** (1984) 469 (cit. on p. 48).

[69] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. C **74** (2014) 3071, arXiv: 1407.5063 [hep-ex] (cit. on p. 51).

[70] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, (2020), arXiv: 2012.00578 [hep-ex] (cit. on p. 52).

[71] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 53).

[72] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2020), arXiv: 2007.02645 [hep-ex] (cit. on pp. 53, 184).

[73] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 53).

[74] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 580, arXiv: 1705.02211 [hep-ex] (cit. on p. 53), Erratum: Eur. Phys. J. C **77** (2017) 712.

[75] ATLAS Collaboration,

ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset,

The European Physical Journal C **83** (2023) 681, arXiv: 2211.16345 [hep-ex]

(cit. on p. 54).

[76] 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, arXiv: 1907.05120 [hep-ex] (cit. on p. 54).

[77] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013,

2017, URL: https://cds.cern.ch/record/2273281 (cit. on p. 54).

[78] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*,

ATL-PHYS-PUB-2017-003, 2017,

URL: https://cds.cern.ch/record/2255226 (cit. on p. 54).

[79] ATLAS Collaboration,

Expected performance of the ATLAS b-tagging algorithms in Run-2,

ATL-PHYS-PUB-2015-022, 2015,

URL: https://cds.cern.ch/record/2037697 (cit. on p. 54).

[80] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in $t\bar{t}$ events using a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001,

2018, URL: https://cds.cern.ch/record/2306649 (cit. on p. 54).

[81] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using ATLAS proton–proton collision data at* $\sqrt{s}$ = 13 *TeV*, ATLAS-CONF-2018-006, 2018, URL: https://cds.cern.ch/record/2314418 (cit. on p. 54).

[82] 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, arXiv: 1802.08168 [hep-ex] (cit. on p. 56).

[83] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at* $\sqrt{s}$ = 7 *TeV*, Eur. Phys. J. C **73** (2013) 2304, arXiv: 1112.6426 [hep-ex] (cit. on p. 56).

[84] ATLAS Collaboration,
Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton–Proton Collisions recorded in 2012 at $\sqrt{s}$ = 8 *TeV*,
ATLAS-CONF-2013-082, 2013,
URL: https://cds.cern.ch/record/1570993 (cit. on p. 56).

[85] ATLAS Collaboration,
E_T^{miss} *performance in the ATLAS detector using 2015–2016 LHC pp collisions*,
ATLAS-CONF-2018-023, 2018,
URL: https://cds.cern.ch/record/2625233 (cit. on p. 56).

[86] *Quantum interference between doubly and singly resonant top quark production*, URL: https://indico.fnal.gov/event/11999/contributions/11562/attachments/7586/9696/herwig.pdf (cit. on pp. 136, 190).

[87] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **362** (1995) 487, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/016890029500274X (cit. on pp. 155, 159).

[88] T. Adye, "Unfolding algorithms and tests using RooUnfold," *Proceedings, 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (PHYSTAT 2011)* (CERN, Geneva, Switzerland, Jan. 17–20, 2011) 313, arXiv: 1105.1160 [physics.data-an] (cit. on pp. 155, 160).

[89] ATLAS Collaboration, *Evaluating statistical uncertainties and correlations using the bootstrap method*, ATL-PHYS-PUB-2021-011, 2021, URL: https://cds.cern.ch/record/2759945 (cit. on pp. 181, 182).

[90] G. Aad, *Electron and photon energy calibration with the ATLAS detector using LHC Run 2 data*, Journal of Instrumentation **19** (2024) P02009, ISSN: 1748-0221,

URL: http://dx.doi.org/10.1088/1748-0221/19/02/P02009

(cit. on p. 186).

[91] G. Aad, *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), ISSN: 1434-6052,

URL: http://dx.doi.org/10.1140/epjc/s10052-023-11584-x

(cit. on p. 186).

[92] A. Hrynevich, *ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2*, Journal of Instrumentation **12** (2017) C06038,

URL: https://dx.doi.org/10.1088/1748-0221/12/06/C06038

(cit. on p. 187).

[93] G. Aad, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at*

$$\sqrt{s} = 13$$

TeV, The European Physical Journal C **79** (2019), ISSN: 1434-6052,

URL: http://dx.doi.org/10.1140/epjc/s10052-019-7450-8

(cit. on p. 187).

[94] G. Aad,

Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment,

Journal of High Energy Physics **2024** (2024), ISSN: 1029-8479,

URL: http://dx.doi.org/10.1007/JHEP05(2024)162 (cit. on p. 188).

[95] G. Aad, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*,

The European Physical Journal C **81** (2021), ISSN: 1434-6052,

URL: http://dx.doi.org/10.1140/epjc/s10052-021-09233-2

(cit. on p. 188).

[96] E. Bothmann, M. Schönherr, and S. Schumann,

Reweighting QCD matrix-element and parton-shower calculations,

The European Physical Journal C **76** (2016), ISSN: 1434-6052,

URL: http://dx.doi.org/10.1140/epjc/s10052-016-4430-0

(cit. on p. 189).

[97] R. D. Ball et al., *Parton distributions for the LHC run II*,

Journal of High Energy Physics **2015** (2015), ISSN: 1029-8479,

URL: http://dx.doi.org/10.1007/JHEP04(2015)040 (cit. on p. 190).

[98] T. Dado, *Tomas Dado / Combiner - GitLab*, 2023,

URL: https://gitlab.cern.ch/tdado/combiner/ (cit. on p. 210).