Evidence of Higgs Boson Production through Vector Boson Fusion

by

Benjamin C. Cerio

Department of Physics
Duke University

Date: _____

Approved:

_____
Ashutosh V. Kotwal, Supervisor

_____
Ayana T. Arce

_____
Steffen A. Bass

_____
Christopher W. Walter

_____
Ying K. Wu

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Physics
in the Graduate School of Duke University
2015

Abstract

Evidence of Higgs Boson Production through Vector Boson Fusion

by

Benjamin C. Cerio

Department of Physics
Duke University

Date: _____

Approved:

_____
Ashutosh V. Kotwal, Supervisor

_____
Ayana T. Arce

_____
Steffen A. Bass

_____
Christopher W. Walter

_____
Ying K. Wu

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics in the Graduate School of Duke University
2015

Abstract

The discovery of the Higgs boson in 2012 provided confirmation of the proposed mechanism for preserving the electroweak $SU(2) \times U(1)$ gauge symmetry of the Standard Model of particle physics. It also heralded in a new era of precision Higgs physics. This thesis presents a measurement of the rate at which the Higgs boson is produced by vector boson fusion in the $WW^{(*)} \to \ell\nu\ell\nu$ decay channel. With gauge boson couplings in both the production and decay vertices, a VBF measurement in this channel is a powerful probe of the VVH vertex strength. Using 4.5 fb^{-1} and 20.3 fb^{-1} of pp collision data collected at respective center-of-mass energies of 7 and 8 TeV in the ATLAS detector, measurements of the statistical significance and the signal strength are carried out in the Higgs mass range $100 \leq m_H \leq 200$ GeV. These measurements are enhanced with a boosted decision tree that exploits the correlations between eight kinematic inputs in order to separate signal and background processes. At the benchmark Higgs mass of 125.36 GeV, the significance of the data assuming the background-only hypothesis to be true has been observed to be 3.2σ (2.7σ expected), constituting evidence of VBF Higgs boson production. The measured signal strength (ratio of observed cross section times branching ratio to that predicted by the SM) is $1.27^{+0.53}_{-0.45}$. The inclusive cross section times branching ratio is found to be $0.51^{+0.22}_{-0.17}$ pb at $\sqrt{s} = 8$ TeV, consistent with the SM prediction of 0.34 pb. No significant deviations from the SM predictions for VBF Higgs boson production are observed.

Contents

List of Tables

List of Figures

List of Abbreviations

Symbols

u	Up quark
d	Down quark
c	Charm quark
s	Strange quark
t	Top quark
b	Bottom quark
q	Quark
j	Jet
p	Proton
ℓ	Lepton
e	Electron
μ	Muon
ν	Neutrino
γ	Photon
g	Gluon
Z	Z boson
W	W boson

V	Weak vector boson (W or Z)
H	Higgs boson
m_H	Higgs boson mass
$\mathcal{L}$	Lagrangian density
σ	Cross section
$\mathcal{B}$	Branching fraction
η	Pseudo-rapidity
p_{T}	Transverse momentum
E_{T}	Transverse energy
ϕ	Azimuthal angle
d_0	Distance of closest approach in transverse plane
z_0	Distance of closest approach in longitudingal direction
ϵ	Efficiency
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse energy
$m_{\ell\ell}$	Dilepton invariant mass
m_{jj}	Dijet invariant mass
m_{T}	Transverse mass of the WW system
$\Delta\phi_{\ell\ell}$	Azimuthal opening angle between leptons
ΔY_{jj}	Rapidity difference between two leading jets
$p_{\mathrm{T}}^{\mathrm{tot}}$	Vector sum of hard object p_{T}
ΔR	Radial η-ϕ distance
Z/DY	Drell-Yan process
$\mathcal{L}$	Likelihood
μ	Signal strength
θ	Nuisance parameter
q_μ	Test statistic

Acronyms

SM	Standard Model
QFT	Quantum Field Theory
QCD	Quantum Chromodynamics
EM	Electromagnetic
vev	Vacuum Expectation Value
MC	Monte Carlo
HS	Hard scatter
PS	Parton Shower
pdf	Parton Distribution Function
ggF	Gluon-Gluon Fusion
VBF	Vector Boson Fusion
BF	Branching Fraction
LHC	Large Hadron Collider
ATLAS	A Toroidal LHC ApparatuS
ID	Inner Detector
MS	Muon Spectrometer
SCT	Silicon Tracker
TRT	Transition Radiation Tracker
HEC	Hadronic Endcap Calorimeter
FCal	Forward Calorimeter
MDT	Monitored Drift Tube
CSC	Cathode Strip Chamber
ROI	Region of Interest
LCW	Local Cell Weighting
JES	Jet Energy Scale

JER	Jet Energy Resolution
JVF	Jet Vertex Fraction
BDT	Boosted Decision Tree
ST	Single Top
SF	Scale Factor
BJV	b-jet veto
CJV	Central Jet Veto
OLV	Outside Lepton Veto
CR	Control Region
SR	Signal Region
NF	Normalization Factor
ISR	Initial State Radiation
POI	Parameter of Interest
NP	Nuisance Parameter
PE	Pseudo-Experiment
PDF	Probability Density Function
CDF	Cumulative Distribution Function
LO	Leading Order
NLO	Next-to-Leading Order

Acknowledgements

Earning a doctoral degree is a challenging and time-consuming endeavor. Earning a doctoral degree in particle physics at a lab on the other side of the Earth poses its own challenges. Fortunately for me, I had an amazing network of supportive and understanding family, friends, and colleagues during—and leading up to—my graduate studies.

First, I would like to thank my father, Jay Cerio, for implicitly expressing to his children the value of education, and for introducing me to the wonder of numbers and the natural world early in my childhood. Thanks to him, I think I was the only 2nd grader who knew about imaginary numbers and atoms. I am also grateful to my mom, Wendy Webb, who has always been a phone call away during difficult times. Thanks to my sisters, Cait and Anna, for being supportive throughout these years.

I also extend loving thanks to my soon-to-be wife, Julie Colvin, who has lost countless hours of sleep due to my late nights and early morning presentations. She has helped me get through the stress-filled years of graduate school with her limitless optimism and her resonating laugh. I am extremely grateful for the sacrifices she has made for me and her unfailing support.

No one has influenced my trajectory towards graduate school more than my undergraduate mentor, Robert Holtzapple. As my first research mentor, he taught me how to translate from the classroom to the laboratory, how to communicate results effectively, how to write and debug code, and he also introduced me to the

excitement of high energy physics research. I extend my sincere gratitude to Robert.

Thanks to my ATLAS colleague, Tae Min Hong, for initially convincing me of the importance of the VBF measurement, and more generally, for championing me in HSG3, helping with my visibility in a large working group. Thanks to both Tae and Philip Chang for the memorable late nights in Cobra working on the first incarnation of the BDT-based analysis.

A myriad of thanks to the VBF BDT team—Doug Schouten, Koos van Nieuwkoop, Jennifer Hsu, Bernd Stelzer, Heberth Torres, Simon Viel, Ashutosh Kotwal, and Oliver Stelzer-Chilton. It was a long road to approval, but we pushed out the first primary BDT-based analysis in HSG3. I thank every one of you for making the work in this thesis possible. Thanks also to the other VBF colleagues with whom I worked closely, including Johanna Bronner, Bonnie Chow, and Tomo Lazovich.

Thanks to Jonathan Long and Christian Meineck for your tireless efforts in producing samples for HSG3, to Aaron Armbruster for developing the statistics code that I built on for the VBF result, to Doug Schaefer for your help in implementing the fake estimate in the BDT analysis, to Magda Chelstowska for your help with the 2011 re-analysis samples, and to Chris Hays for your guidance in computing WW uncertainties. I also want to thank the HSG3 conveners whose tenure coincided with my time in the group in Run-1: Jianming, Pierre, Biagio, Tatsuya, Corrinne, and Olivier.

Thanks to the Duke HEP group. In particular, I would like to thank my advisor, Ashutosh Kotwal, who provided guidance in my first years of graduate school, and left me free to pursue my own interests in later years. I am grateful to him for always making time for spontaneous physics discussions, for facilitating collaboration, and for helping to teach me how to think analytically in the context of particle physics. Also, thanks to Professor Mark Kruse and Professor Al Goshaw, and in particular to Professor Ayana Arce for her guidance in my first year of research in ATLAS.

The talented graduate students in the Duke HEP group were a valuable resource through the years, whether offering help in solving a computing issue or advice on navigating ATLAS politics. I would like to thank the first Duke graduate students in ATLAS, my former officemates, and friends: Chris Pollard, Mia "Meow" Liu, and Kevin Finelli. I also thank Dave Bjergaard for many fun physics and non-physics discussions, and Lei Li and Chen Zhou for tolerating my office behavior.

I would like to acknowledge the Duke physics grads class of 2008. Thanks to my friends Chris Coleman-Smith and Hannah Guilbert, without whose support I wouldnt've made it through the first few years of graduate school. Thanks to Taritree Wongjirad for our many late nights doing advanced quantum and particle physics problem sets together. Thanks to George Laskaris for many interesting discussions about physics, Greek politics, and life. I also spent time doing problem sets, navigating grad school life, and exploring Durham with the remaining members of my class, including Huaixiu Zheng, Abe Clark, and Yuan Lin. Thanks to all for the pleasant memories.

Lastly, I would like to extend my thanks to the rest of my Durham network of friends, including my band-mates Nancy Scott and Marco Bertolini. Playing music provided much-needed respite from my research obligations. Thanks to the "Gregson girls"—Jessalee, Nancy, and Nikki. Thanks to Christy and Matty. Thanks to my math friends, Dave, Josh (not math, but close enough), Phil, Sarah, Brian, and Albert. You have all helped to make graduate school a chapter of my life that I will reflect on with great fondness.

Preface

The field of high energy particle physics entered an exciting new chapter when the LHC first collided stable proton beams. Shortly after these first collisions, in the spring of 2010, I entered the ATLAS collaboration. I started by completing service work in the inner detector alignment group, where I developed and maintained code for computing corrections to the intrinsic errors for the sub-detectors of the inner detector. In my first few years, I was also involved in searches for resonances producing two same-sign leptons. In particular, I was interested in Higgs triplet models which predict the existence of a doubly-charged Higgs that decays leptonically. For these searches, I studied backgrounds due to photon conversions with a sample of pions, validating the material model in simulation. Starting in the fall of 2011, I switched gears from searches for exotic Higgs to a search for the SM Higgs in the $WW^{(*)} \to \ell\nu\ell\nu$ decay channel. I helped to develop a matrix-element-based analysis, and investigated the possibility of using a linear discriminant to combine the two competing multivariate analyses at the time. The Higgs discovery was announced in the summer of 2012, at which time a 3.2σ excess of events was observed in the $WW^{(*)} \to \ell\nu\ell\nu$ channel. Later that summer, my focus shifted to the two jet bin of the same decay channel. Having many correlated particles in the final state, it was well-suited for multivariate techniques. I was on the team to first motivate the use of a boosted decision tree, and it was subsequently adopted as the primary vector boson fusion analysis. This thesis presents the BDT-based vector boson fusion anal-

ysis from Run-1. Most of the results presented in this thesis are summarized in a paper that has been submitted for publication.

I am extremely grateful to have been involved in a scientific endeavor as ambitious and important as ATLAS. Working with such talented and driven individuals has proven to be both humbling and motivating. Given the highly collaborative nature of the field, the work presented in this thesis has, by no means, all been completed by me. And the work that I have done myself would not have been possible without the support of my immediate subgroup, without the years of analysis development prior to my joining the $H \to WW^{(*)}$ subgroup, and most importantly, without the considerable resources spent on the construction of ATLAS and the development of its robust reconstruction and identification algorithms.

In the following thesis, I attempt to write at a level that should be understandable to a beginning graduate student. In many cases, it was not feasible to adhere to this convention without diluting the content or providing an unreasonable amount of background explanation. In such cases, the content is explained at a level that is appropriate for the high energy physics experimentalist. The chosen convention is relatively constant for a given chapter. On a scale of 1-4, with 1 being for the beginning grad student and 4 being the HEP specialist, the following serves as a guide for the reader: Theoretical Background - 1(2), The ATLAS Detector - 1, Particle and Event Reconstruction - 3, Analysis Overview - 2, Collision Data and Simulation - 3, Object and Event Selection - 4, Background Estimates - 4, Systematic Uncertainties - 4, Results - 4, Statistical Treatment - 1(2), Boosted Decision Trees - 1. The parenthesis indicates that certain sections of the chapter may assume a different convention.

1

Introduction

Since the time of the ancient Greeks, humans have been pondering the basic question "what is matter?" The early Greek philosophers Leucippus and Democritus hypothesized the existence of fundamental building blocks of matter, termed "atomos", or "that which can not be split." Thus was born the idea of an elementary particle as a unit of matter that is not divisible.

With the advent of the scientific method, the Greek philosophy of the atom was developed into a testable theory. John Dalton's atomic theory introduced the idea of a chemical element—a species of matter that is composed of immutable atoms of the same mass and size—that can be combined with other elements to form chemical compounds. His theory, however, still considered atoms to be indivisible. This idea changed in 1897 when J.J. Thomson discovered that atoms were composed of negatively charged electrons with a mass on the order of 1000 times smaller than that of the hydrogen atom. Since it was known at the time that atoms were electrically neutral, Thomson proposed that in addition to the electron, there is an additional fundamental particle with a positive charge that is uniformly distributed through the atom. Earnest Rutherford subsequently showed that the positive charge in the atom

is in fact concentrated in an atomic nucleus, and hypothesized the existence of an additional nucleon—the neutron—that helps to bind the positively charged particles of the nucleus together.

Over the course of the twentieth century, scientists learned that nucleons are bound states of particles known as quarks, which, along with electrons, are currently considered indivisible. As the technology in experiments improved, additional quarks and electron-like particles, as well as an electrically neutral particle known as the neutrino, were discovered. Moreover, a theory that describes the interactions of elementary particles as the exchange of force-carrying particles was developed and found to agree with experimental evidence. A major deficiency in this theory, related to the origin of the masses of the elementary particles, was fixed in the 1960s, and with the fix came the prediction of yet another elementary particle, the Higgs boson. This theory of elementary particles is known as the Standard Model of particle physics. The particles in this model are summarized in Figure 1.1.

Prior to 2012, all of the particles in the Standard Model had been observed, with the exception of the Higgs boson. The fundamental interactions among the particles predicted in the Standard Model had also been widely confirmed by experiment. In the summer of 2012, the ATLAS and CMS experiments at the Large Hadron Collider announced the discovery of a particle whose properties closely resembled that of the Higgs boson. Subsequent measurements by these experiments have confirmed that the Higgs boson exists, a great triumph for the Standard Model.

This thesis builds on the centuries of work attempting to understand the building blocks of matter with a measurement of the production of the Higgs boson through a mechanism known as vector boson fusion. The measurement has been performed with proton-proton collision data collected with the ATLAS detector between 2011 and 2013 in the Higgs decay channel $H \to WW^{(*)} \to \ell\nu\ell\nu$. It is organized as follows. Chapter 2 motivates the Higgs boson measurement with a discussion of the gauge

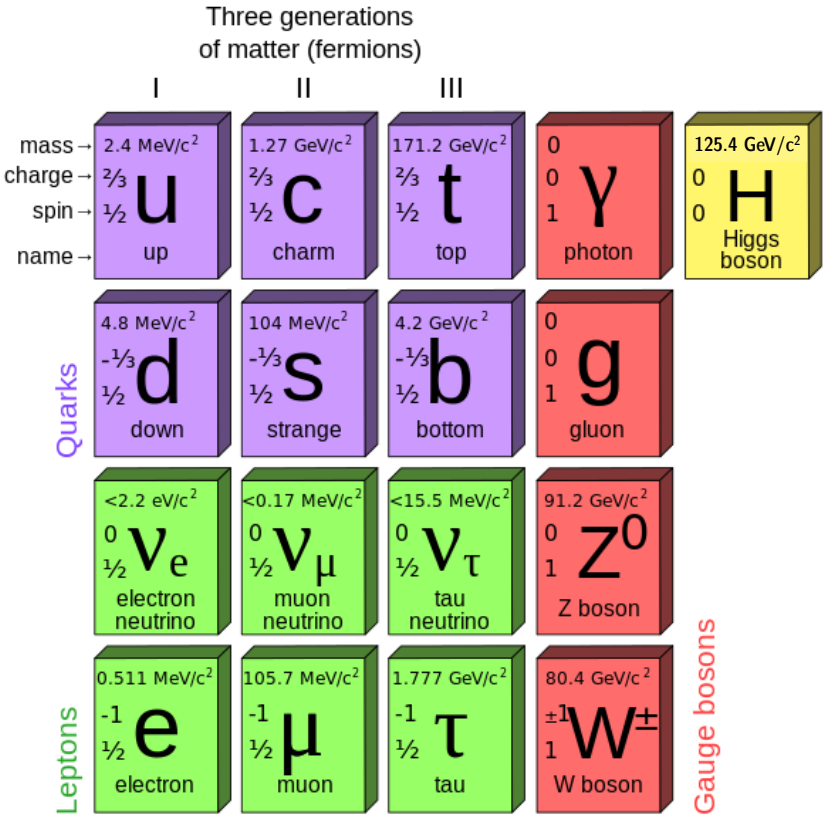


FIGURE 1.1: Summary of the particles in the Standard Model of particle physics.

structure of the Standard Model, its applications in collider physics, and the physics of the Higgs boson at the LHC. In Chapter 3 the ATLAS detector is introduced, with a discussion of its various subsystems. The reconstruction of proton-proton collision events in the ATLAS detector is explained in Chapter 4. Starting with Chapter 5, the measurement of vector boson fusion is presented. Chapter 5 presents an overview of the analysis. In Chapter 6, there is detailed information regarding the collision data on which the measurement is performed and the Monte Carlo simulation samples that are used to make predictions. This chapter also outlines the expected background processes for this measurement. Chapter 7 examines the

selection of particles and events, and introduces the multivariate classifier that is used in the analysis. Chapter 8 explores data-driven background estimation techniques, and Chapter 9 outlines the experimental uncertainties that are assigned. Chapters 10 and 11 present the results and their statistical interpretation, respectively. The statistical formalism is discussed in detail in the beginning of Chapter 11. Finally, the thesis concludes with a brief statement in Chapter 12.

2

Theoretical Background

Nature uses only the longest threads to weave her patterns, so that each small piece of her fabric reveals the organization of the entire tapestry.

– Richard Feynman

Evolving over the course of decades of theoretical insights and experimental discoveries, the Standard Model (SM) of particle physics can be viewed as a success of the scientific method. In its current incarnation, the SM describes three of the four elementary particle interactions, predicting with surprising accuracy the spectrum of elementary particles, the strengths with which they interact, and the form of their interactions. Governing the interactions within stars, nuclear reactors, and in the universe $\sim 10^{-32}$ s after the Big Bang, these predictions are derived from fundamental symmetries known to exist in nature. In this chapter, the theoretical framework for the Higgs boson measurement presented in this thesis is laid out, starting with a description of the SM. From there, an application of the SM to physics in hadron colliders is discussed, and the chapter concludes with a look at the nature of the Higgs boson at the Large Hadron Collider.

2.1 The Standard Model

The SM seeks to predict the spectrum of, and interactions among, the particles that constitute matter. Since these elementary particles are infinitesimally small, their behavior is governed by the postulates of quantum mechanics (QM). Specifically, the SM is built on a theoretical framework known as quantum field theory (QFT), a relativistic and field-theoretic extension of QM. In addition to being Lorentz invariant, QFT improves on QM by allowing particle number to be violated in a closed system, a phenomenon that is observed in, for example, an atomic energy level transition whereby a photon is created or absorbed. Particles in QFT are described as excited states of space-time fields. Each particle has a field that is realized as a mathematical operator serving to create or annihilate particles. The dynamics of a field and its associated particles are obtained from a Lagrangian density, $\mathcal{L}(\phi, \partial_\mu \phi)$, the field-theoretic analogue of the Lagrangian in classical mechanics. In the context of QFT, if two different fields appear in the same term of $\mathcal{L}$, the two particles associated with these fields are said to "couple." It is these couplings that lead to the fundamental interactions.

Elementary particles broadly fall into two categories: half-integer spin fermions and integer spin bosons. Quarks and leptons, for example, are fermions with spin-1/2, while the particle that mediates the electromagnetic (EM) interaction, the photon, is a spin-1 boson. Dirac formulated a relativistic analogue of the Schrodinger equation for spin-1/2 fermions, and the corresponding Lagrangian is [17]

$$\mathcal{L}_{\text{fermion}} = i\bar{\psi}\gamma_\mu \partial^\mu \psi - m\bar{\psi}\psi. \tag{2.1}$$

To incorporate the two spin states of the fermion and the fact that each particle has an antiparticle, the field ψ has four degrees of freedom. In the above equation, $\bar{\psi} = \psi^\dagger \gamma^0$, where $\psi^\dagger$ denotes the adjoint. The γ^μ are the Dirac γ-matrices, whose anti-

commutators form the Dirac algebra: $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$. The first term of Equation 2.1 is the kinetic energy term, and the second is the mass term that gives a mass m to the particle associated with the field.

Equation 2.1 describes a quark or a lepton that does not interact with other particles through the four fundamental interactions. In the SM, the EM, weak, and strong interactions are incorporated as additional terms in $\mathcal{L}_{\text{fermion}}$, while the fourth interaction, gravity, has not been successfully synthesized into the model. Inter-field interactions arise when a local gauge symmetry is imposed. These so-called symmetries are in fact internal mathematical degrees of freedom that are required to preserve the equations of motion. For example, the Lagrangian should be invariant under the transformation $\psi \rightarrow e^{-ig\alpha(x)}\psi$, where α is an arbitrary local phase. The non-interacting Lagrangian in Equation 2.1 does not fulfill this requirement. However, if a new field is introduced through a term of the form $g\bar{\psi}\gamma^\mu\psi A_\mu$, $\mathcal{L}_{\text{fermion}}$ becomes invariant as long as the new field transforms as $A_\mu \rightarrow A_\mu + \partial_\mu\alpha(x)$. The vector field A_μ is called a gauge field, and it represents the photon. Because this new field represents a physical particle, a kinetic energy term associated with A_μ needs to be introduced. A choice that preserves gauge and Lorentz invariance is a kinetic energy term of the form $-(1/4)F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu}$ is the EM field strength tensor, defined as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. A Klein-Gordon photon mass term of the form $(1/2)m^2 A_\mu A^\mu$ is not included in the Lagrangian because it transforms as $A_\mu A^\mu \rightarrow A_\mu A^\mu + 2\partial_\mu\alpha A^\mu + \partial_\mu\partial^\mu\alpha$, thereby breaking gauge invariance. The photon, therefore, is required to be massless. It can be shown that a Lagrangian of this form for A_μ recovers Maxwell's equations [18].

An important property of the gauge transformation $e^{-ig\alpha(x)}$ is that it forms a Lie group under multiplication [19]. In general, any element of a Lie group can be written as $e^{-i\alpha_j(x)X_j}$, where the X_j are the group generators and the parameters $\alpha_j(x)$ identify each element of the group. For a gauge theory, the number of generators is equal to

the number of gauge fields that arise when invariance is imposed in the Lagrangian. Furthermore, the nature of the resulting gauge interactions is determined by the commutation relations among the generators: $[X_i, X_j] = if_{ijk}X_k$, where the f_{ijk} are constants called the structure constants.

For the EM interaction, the symmetry group is $U(1)$, an abelian group that has a single generator, resulting in a single gauge field associated with the photon. For the weak interaction, the internal symmetry is weak isospin, which is described by the Lie group $SU(2)$. With three generators, $SU(2)$ gauge invariance produces three fields associated with the weak vector bosons W^+, W^-, and Z. In the theory of the strong interaction, called quantum chromodynamics (QCD), an $SU(3)$ color symmetry is required, resulting in eight gauge fields associated with gluons. Apart from producing more gauge fields, the $SU(2)$ and $SU(3)$ symmetries differ from the $U(1)$ symmetry of the EM interaction because they have non-zero structure constants. An important consequence of this differing group structure is that self-interaction terms arise in the Lagrangian. Hence, both gluons and weak vector bosons can couple to themselves.

In EM, the inclusion of a mass term for A_μ breaks the $U(1)$ gauge symmetry, which is easily resolved by setting the mass of the photon to zero. The same problem arises in both strong and weak Lagrangians. In the case of QCD, the eight gluon masses are set to zero, and color symmetry is restored. The masses of the three weak gauge bosons, on the other hand, have been measured to be non-zero. In fact, these particles are quite massive: $m_W = 80.4$ GeV and $m_Z = 90.1$ GeV. If the gauge symmetry is broken by introducing mass terms, infinities are induced in the perturbative expansion of the path integral which can not be renormalized, rendering the theory non-predictive.

In addition to the weak boson mass terms, the fermion mass terms in the weak Lagrangian violate the $SU(2)$ symmetry. Since the weak interaction has been measured to maximally violate the discrete symmetry known as parity, gauge invariance

is only required for left-handed fields, though both left and right-handed fermion fields exist in nature and hence, in the Lagrangian. With a mixture of helicity states in the mass term, each transforming differently under $SU(2)$, the symmetry is broken. The prediction implied by setting all fermion masses to zero is at odds with a myriad of experimental evidence. These theoretical problems with the weak boson and fermion masses led to the development of the idea of spontaneous symmetry breaking.

An approach for simultaneously introducing both gauge boson and fermion mass terms into the Lagrangian, while preserving gauge symmetry and renormalizability, was developed in the 1960s [20, 21, 22, 23, 24, 25, 26, 27], and was subsequently adapted for the unified electroweak theory [28, 29]. A complex scalar field that transforms as a weak isospin doublet, and is described by the Lagrangian

$$\mathcal{L}_{\text{Higgs}} = \left(D_\mu \phi^\dagger\right)\left(D^\mu \phi\right) - \left[\mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2\right] \tag{2.2}$$

is introduced into the Lagrangian describing both EM and weak interactions. Here the derivative D_μ is the covariant derivative. This form is chosen to be invariant under the gauge symmetry $SU(2) \times U(1)$, and due to the restrictions $\mu^2 < 0$ and $\lambda > 0$, to yield non-zero ϕ at the potential energy minimum. The ϕ state that minimizes the potential energy, denoted ϕ_0, corresponds to the vacuum state, and the magnitude of the field is known as the vacuum expectation value (vev). From the invariance of $\mathcal{L}_{\text{Higgs}}$, the manifold defined by ϕ_0 is manifestly invariant under $SU(2) \times U(1)$. The $SU(2)$ symmetry is then broken "spontaneously" when a point on the manifold is chosen. This process is spontaneous in the sense that the symmetry is broken in the ground state wave function and not the Lagrangian itself.

By performing a gauge transformation on ϕ_0, the non-zero expectation value can be projected into the real part of the neutral component of the doublet, leaving the

three remaining real scalar degrees of freedom with a magnitude of zero [30]. To recover the dynamics, the non-zero scalar component is perturbed about the vev—$v+H(x)$—where the field $H(x)$ is a real scalar field known as the Higgs field. Putting this expression for ϕ into Equation 2.2, mass terms for $W^{\pm}$ and Z are generated in the coupling of the v part of the field and bilinear gauge boson terms in the covariant derivative D_{μ}. Additional gauge invariant terms that couple ϕ and bilinear fermion terms are added to the Lagrangian, resulting in fermion mass terms when ϕ acquires a vev. This mechanism, therefore, is successful in mitigating both the gauge boson and fermion mass deficiencies outlined above.

An important bi-product of this mechanism for generating masses is the prediction of the existence of a spin-0 particle, hereafter referred to as the Higgs boson or the Higgs. The Higgs boson couples to fermions and gauge bosons with a strength proportional to their masses, implying that direct production of such a particle is possible at colliders. Moreover, such couplings induce higher order corrections in measurable quantities such as the top quark and W boson masses. Consequently, the existence of the Higgs boson can be indirectly established by measuring deviations in these observables. Though the Lagrangian 2.2 appears to introduce two new parameters to the SM, one can be related to existing SM input parameters, leaving only one parameter that can be expressed in terms of the mass of the Higgs boson (m_H).

The above procedure, known popularly as the Higgs mechanism, restores the gauge symmetry of the SM Lagrangian, thereby ensuring renormalizability. The underlying $SU(2)$ symmetry of the weak interaction is hidden, or spontaneously broken, resulting in a tangible prediction, the existence of a massive, spin-0 boson whose coupling to other particles is related to their masses. Other scalar degrees of freedom in the field ϕ become the longitudinal polarization of the weak gauge bosons, a necessary property of massive spin-1 particles. This form of the Higgs mechanism

is not unique; the chosen representation of $SU(2)$, namely that ϕ is a complex scalar doublet, is merely the most parsimonious. Additional degrees of freedom arise if another representation is chosen or if an additional $SU(2)$ invariant Higgs fields are introduced. In some cases, these extended Higgs scenarios, which predict additional Higgs particles, fix other theoretical limitations of the SM. However, because every incarnation of the Higgs mechanism has at least one neutral scalar Higgs boson, the minimalistic form is chosen in the SM.

In its current form, the SM is a gauge theory obeying $SU(3) \times SU(2) \times U(1)$ symmetry, resulting in a total of 12 gauge bosons: 8 massless, bi-colored gluons, 3 massive weak vector bosons, and a single massless photon. Three fermion generations, each with a charged lepton, a neutrino, two quarks, as well as their corresponding anti-matter particles, have been observed, totaling 48 fermions. Adding to the list the all-important Higgs boson, 61 particles are predicted to exist by the SM, all of which have been experimentally observed.

2.2 The SM at Hadron Colliders

Starting with Rutherford's gold foil experiment, scattering experiments that measure the deflection of particles in the presence of some form of matter have been effective in probing the structure of matter and the forces with which it interacts. Such experiments have proved to be invaluable testing grounds for the SM. The gauge structure of the SM has historically been tested by looking for evidence of particles introduced in the process of enforcing gauge symmetry or by measuring couplings in the gauge sector. The coupling strength between fields is related to a quantity called the cross section which is a measure of the probability with which a scattering process occurs. For the scattering of two particles into n particles, the cross section

can be written

$$\hat{\sigma}_{qq\to n} = \int_{\mathcal{V}_n} |M(q_1, q_2; y_1, ..., y_n)|^2 d\Phi_n(q_1 + q_2; y_1, ..., y_n) \tag{2.3}$$

where the q_i (y_i) are the momentum 4-vectors of the incoming (outgoing) particles. The differential $d\Phi_n(q_1 + q_2; y_1, ..., y_n)$ is the Lorentz invariant phase space term that enforces conservation of energy and momentum, and $M(q_1, q_2; y_1, ..., y_n)$ is the matrix element that captures the dynamics of the Lagrangian. The integration is performed in the phase space region of interest $\mathcal{V}_n$. $M(q_1, q_2; y_1, ..., y_n)$ is calculated from the Lagrangian with the perturbative techniques of QFT (see, for example [31, 32]). Therefore, measuring the cross section $\hat{\sigma}$ for a given scattering process is an effective test of the relevant terms in the Lagrangian.

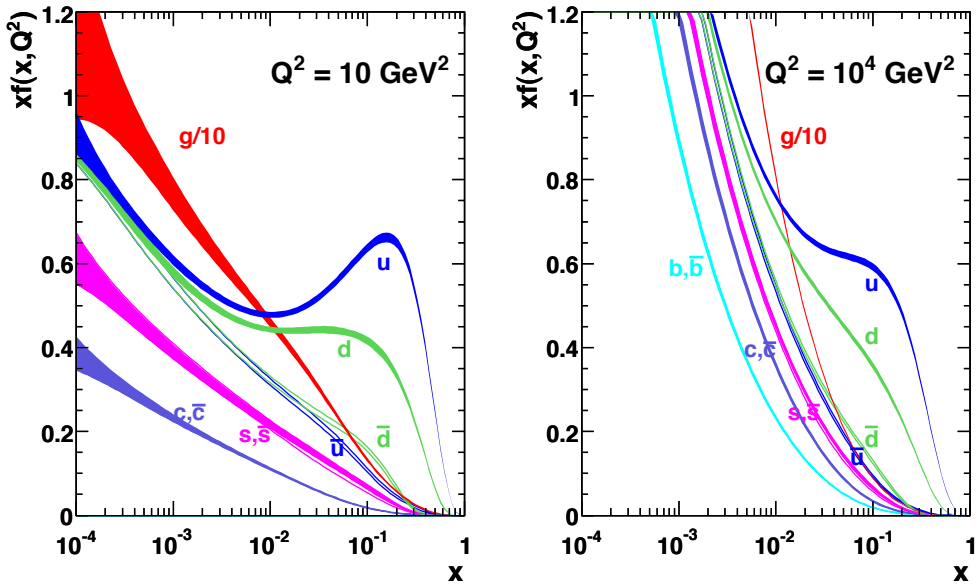


FIGURE 2.1: Parton distribution functions for the partons within the proton, shown at two different energy scales, 10 GeV (left) and 10^4 GeV (right), from the MSTW 2008 NLO pdf set [2]. pdfs are shown as a function of the parton energy fraction x.

The expression for σ defined in Equation 2.3 applies to the scattering of two elementary particles into n particles. At the LHC, the two incident particles are

protons, which are composite particles composed of quarks interacting through gluon exchange. Due to the non-abelian structure of the gauge group governing these interactions, as well as the fact that gluons are massless, quarks and gluons can be described as free particles at sufficiently high energy scales, or equivalently at short distances. This phenomenon, a consequence of the fact that the strong coupling constant decreases with increasing energy, is called "asymptotic freedom" [33, 34]. It allows the perturbation techniques of QFT to be used for QCD predictions at energy scales above the threshold Λ_{QCD}. Below this scale, in the long-distance regime, the strong interaction becomes extremely strong. As a consequence, it is not possible to isolate quarks or gluons. Instead, the strong field pulls quarks from the vacuum to form a configuration of two or three quarks that transforms as an $SU(3)$ color singlet. This property of QCD is known as confinement.

Protons are composed of three valence quarks—two up quarks and a down quark—as well as a "sea" of gluons and quarks of other flavors due to excitations between the valence quarks. Parton distribution functions (pdfs) quantify the probability that a given parton within the proton carries a momentum fraction x of the total proton momentum. These pdfs have been determined as a function of the energy scale at which the proton is probed through a global fit of data from deep inelastic scattering and other high energy collider experiments [2]. In Figure 2.1, the results of these fits are shown as the product of the parton momentum fraction and the pdf, $xf(x, Q^2)$, which represents the momentum density for a given parton. At high x, the valence quarks carry the majority of the momentum, and at low x the sea partons begin to contribute momentum, with the gluon pdf dramatically larger than the others (note that the gluon distribution is scaled down by a factor of 10). This behavior turns out to be important in Higgs physics at the LHC (Section 2.3).

For proton-proton scattering, the cross section is computed by weighting the quark-quark scattering cross sections (e.g. Equation 2.3) by the pdfs and integrating

over the momentum fractions [35]:

$$\sigma_{pp\to n} = \int dx_1 dx_2 f_{q_1}(x_1, \mu_F) f_{q_2}(x_2, \mu_F) \hat{\sigma}_{qq\to n}. \tag{2.4}$$

Here, f_{q_1} (f_{q_2}) is the pdf for a quark of flavor q_1 (q_2) and momentum fraction x_1 (x_2). The cross section is now expressed in terms of the perturbative cross section associated with the free partons, with the non-perturbative part factorized into the pdfs. The scale that separates the two regimes is called the factorization scale, μ_F. As more terms are included in the perturbative expansion of $\hat{\sigma}_{qq\to n}$, divergences arise due to unrestricted momenta in loop diagrams. To make the theory predictive again, $\hat{\sigma}_{qq\to n}$ is renormalized at some scale μ_R. The resulting cross section calculated at all orders in perturbation theory is invariant with respect to changes in μ_F and μ_R. Due to calculational difficulties, the cross section is computed at fixed order in perturbation theory, making it necessary to vary μ_F and μ_R, and quantify the change in the predicted cross section. This change is then assigned as a theoretical uncertainty.

In most scattering experiments, a prediction is obtained from Monte Carlo (MC) simulation, whereby an event is generated probabilistically by drawing from the differential cross section distribution defined by Equation 2.4. MC generators are able to generate a hard scattering up to some fixed order, and to augment fixed order calculations, parton shower (PS) programs are typically interfaced to the MC generator, allowing diagrams with more vertices to be modeled. For a quark or gluon in the final state, the PS program uses the DGLAP equation [36] to model the emission of additional quarks and gluons down to some cut-off energy scale in a process known as fragmentation. The resulting partons, which are not confined to color singlet configurations, are then hadronized using a non-perturbative model, forming a collection of hadrons that are observable to the detector. The shower of

hadrons associated with a final state quark or gluon is known as a jet (discussed in more detail in Chapter 4).

In the electroweak sector of the SM, the gauge structure is tested by measuring evidence for diagrams that only arise when gauge invariance is imposed, and by comparing the cross sections of such processes to the SM predictions. The LHC, with its high center-of-mass energy, is sensitive to many of these processes, as shown in Figure 2.2, which summarizes the cross sections measured by the ATLAS detector for some important SM processes. The production of a single W or Z boson is precisely measured to be consistent with the SM prediction. Processes involving the production of two weak gauge bosons—WW, WZ, and ZZ—are also in agreement with the SM. Such self-consistent predictions provide strong evidence for the gauge structure of the SM.

2.3 Higgs Physics at the LHC

Prior to the summer of 2012, the gauge structure of the SM had held up to repeated experimental tests, with the exception of one crucial part—evidence of the Higgs boson associated with $SU(2) \times U(1)$ symmetry breaking. This long sought after particle had been one of the primary motivations for the construction of the LHC, a pp collider with a center-of-mass energy ($\sqrt{s}$) expected to be large enough to observe the particle. Following the discovery of the Higgs boson in July 2012, LHC experiments entered a new phase of precision measurements of the Higgs couplings. In the following section, Higgs physics at the LHC is discussed.

The introduction of the Higgs field into the SM Lagrangian gives rise to a consistent set of testable predictions. The Higgs boson should behave like a chargeless spin-0 particle. It should couple to weak gauge bosons through terms of the form HVV ($V = W/Z$), with a strength proportional to the square of the mass of the gauge boson. Moreover, as a consequence of fermion mass generation, the Higgs

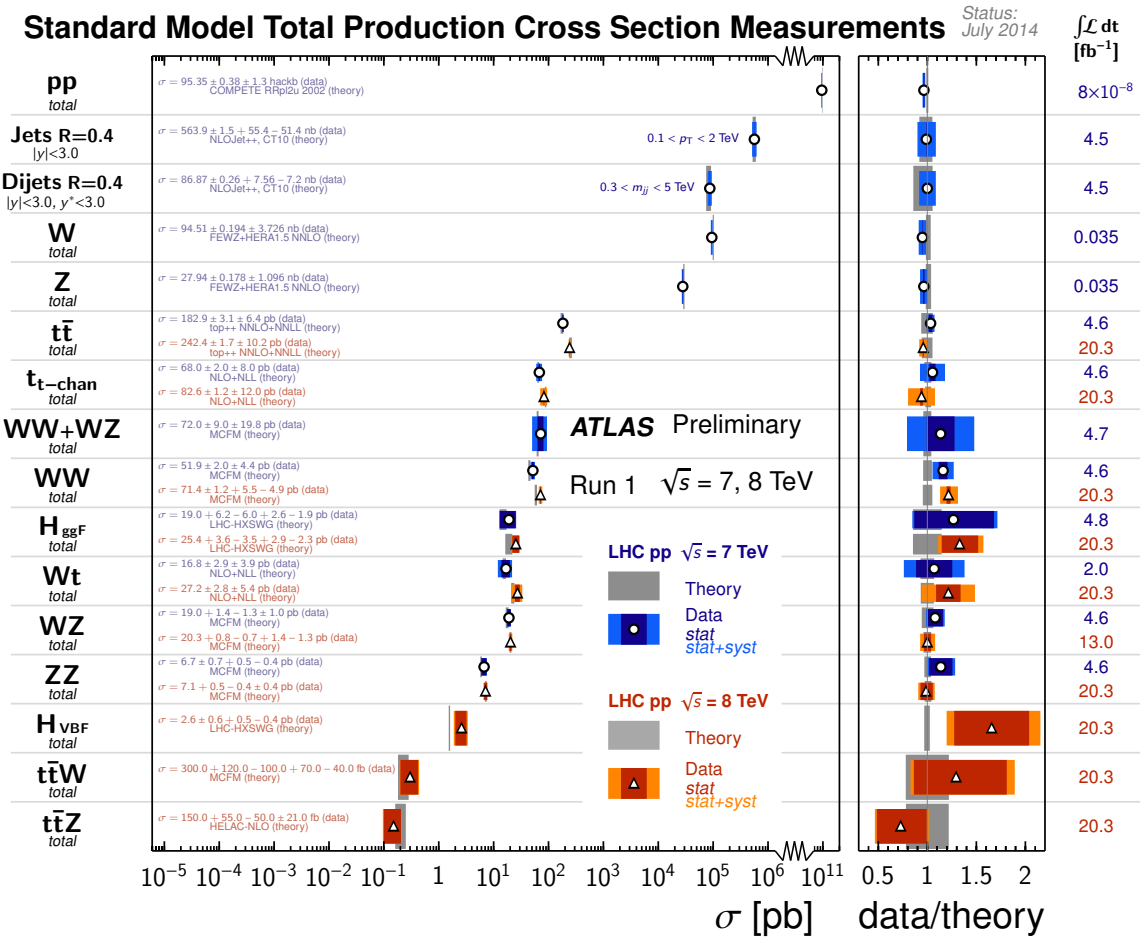
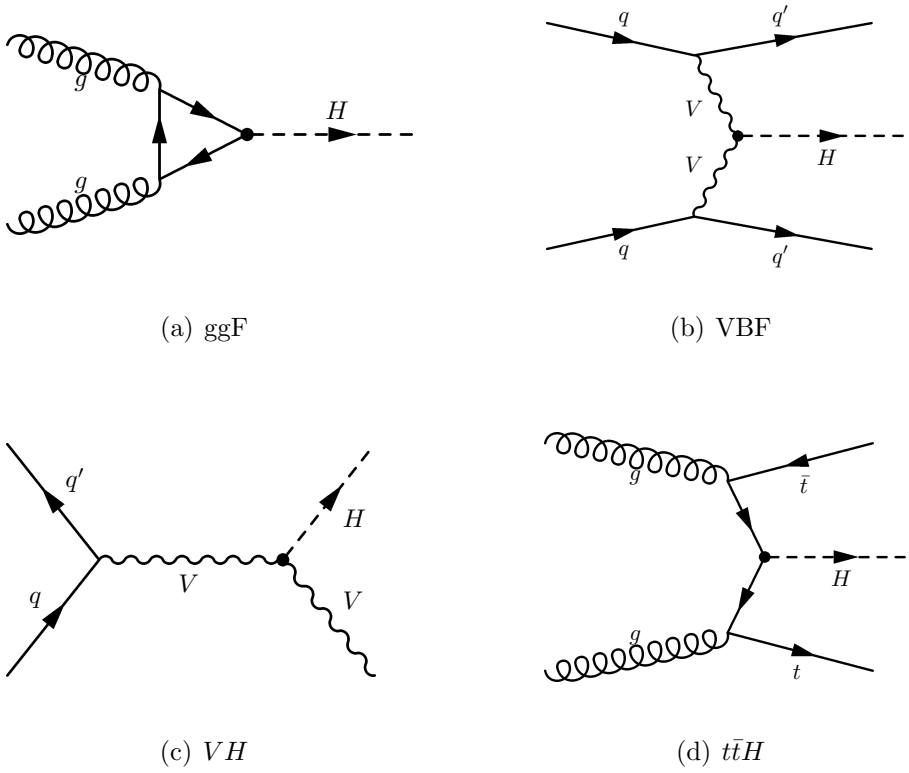


FIGURE 2.2: Total production cross section measurements for Standard Model scattering processes as measured in ATLAS. The sub-plot on the right is the ratio of the measured cross section to the cross section as predicted by the Standard Model.

boson should couple to fermions with a strength that grows linearly their masses. Because the masses of the gauge bosons and fermions are well-measured, the Higgs couplings are determined. In fact, the only remaining parameter to which Higgs predictions are sensitive is the mass of the Higgs boson, which is not constrained by the theory itself. It is, however, argued that, if the theory is to remain unitary and have a stable vacuum, the Higgs mass should lie in the range 50 GeV $\lesssim m_H \lesssim$ 800 GeV [37]. Since the Higgs boson is a massive particle that couples to both fermions and gauge bosons, it is found in loop diagrams at higher orders in perturbation theory. These

diagrams contribute non-negligible corrections to SM input parameters that are observable in precision electroweak measurements. Using such measurements from the LEP, SLC, and Tevatron experiments, m_H had been constrained to be 91^{+30}_{-23} GeV at the one standard deviation level prior to the discovery [38]. A direct search done at LEP2 placed a lower bound of $m_H > 114.4$ GeV at the 95% confidence level. These experimental constraints helped to guide the LHC experiments prior to the Higgs discovery. Following the discovery, the Higgs mass has been measured to be 125.7 ± 0.4 by the CMS collaboration [39] and 125.36 ± 0.41 by the ATLAS collaboration [40].



(a) ggF (b) VBF

(c) VH (d) $t\bar{t}H$

FIGURE 2.3: Feynman diagrams of the dominant Higgs production mechanisms at the LHC. Gluon-gluon fusion (a) has the largest cross section, followed by vector boson fusion (b). Associated production of the Higgs with weak vector bosons (c) and top quarks (d) amount to smaller contributions.

The dominant mechanisms for the production of a Higgs boson at the LHC, shown

in terms of Feynman diagrams in Figure 2.3, are dictated by the fact that the Higgs boson preferentially couples to more massive particles. The process with the largest cross section is gluon-gluon fusion (ggF), $gg \to H$, characterized by two incoming gluons that effectively couple to the Higgs through a quark loop. Only the heavy top and bottom quarks contribute in this loop. If the center-of-mass energy is large with respect to m_H, then the Higgs can be produced with a small fraction of the incoming proton momentum, or in the x region where the gluon pdf is large (Figure 2.1). Higgs production by direct coupling between the Higgs and incoming quarks is suppressed by the pdfs for the heavy quarks for which there is a non-negligible coupling to the Higgs.

The second largest Higgs production mechanism is the vector boson fusion (VBF) process, whereby two incoming quarks radiate virtual weak gauge bosons that then fuse to form the Higgs, $qq \to q'q'V^*V^* \to q'q'H$. Though this process can proceed via either W or Z fusion, the contribution of the W diagram is around three times that of the Z, due to the fact that W bosons couple more strongly to fermions [37]. In spite of the smaller cross section, this process is a powerful probe of the Higgs sector due to its characteristic final state. Since the energy of the radiated weak bosons is significantly less than that of the incoming quarks, the deflection of these quarks is small, and therefore the outgoing quarks manifest as high energy forward jets. Another feature of these events is that there is little QCD radiation between the outgoing quarks, due to the absence of the flow of color between them. These two characteristics allow such Higgs events to be efficiently isolated from the background events, which, in hadron colliders, are likely to include central jets.

Associated production of the Higgs with either weak gauge bosons or top quarks is also visible at the LHC. The former occurs when the initial state quarks form an off-shell gauge boson that then splits into a Higgs and a gauge boson: $qq \to V^* \to VH$. The latter process is similar to ggF in that an effective coupling between gluons and

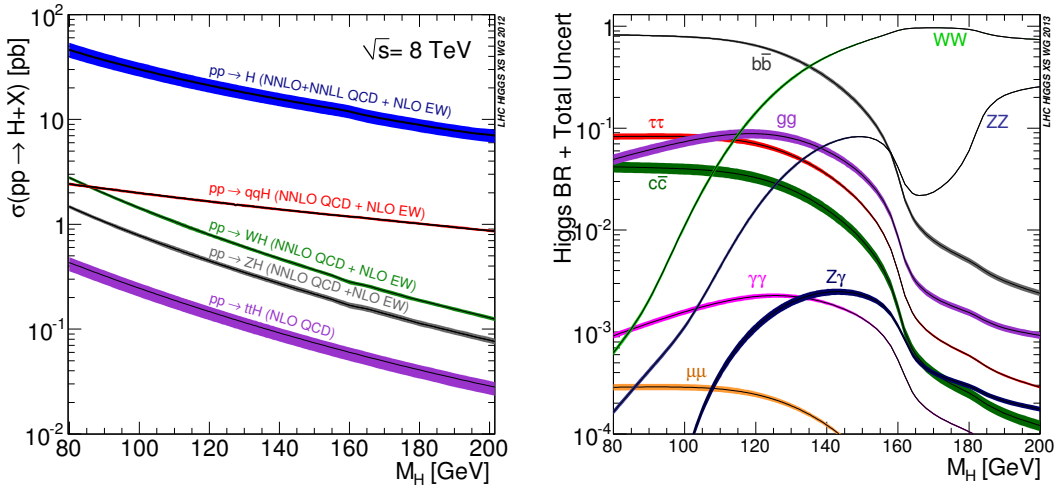
Table 2.1: Higgs boson branching fractions (BFs) at $m_H = 125.4$ GeV for all of the possible decay channels, which are ordered by BF magnitude [1].

Higgs Decay	Branching Fraction
$b\bar{b}$	0.571
WW	0.221
gg	8.53×10^{-2}
$\tau\tau$	6.25×10^{-2}
$c\bar{c}$	2.88×10^{-2}
ZZ	2.74×10^{-2}
$\gamma\gamma$	2.28×10^{-3}
$Z\gamma$	1.57×10^{-3}
$\mu\mu$	2.17×10^{-4}

the Higgs is mediated through top quarks, but in this case the quarks appear as outgoing particles: $gg \to t\bar{t}H$.

The cross sections of these four production processes at $\sqrt{s} = 8$ TeV are shown in Figure 2.4(a)[3]. Across the m_H range shown, the ggF process cross section is approximately an order of magnitude larger than that of VBF. At the measured value of $m_H = 125.4$ GeV, $\sigma_{\text{ggF}} = 19.15$ pb and $\sigma_{\text{VBF}} = 1.573$ pb [41]. The WH (ZH) cross section is 0.6970 pb (0.4112 pb), and $\sigma_{t\bar{t}H} = 0.1280$, about two orders of magnitude smaller than σ_{ggF}.

Once produced, the unstable Higgs boson decays instantaneously into the particles to which it couples. The probability for decaying into a given set of particles is quantified by the branching fraction (BF), shown for the allowed Higgs decays in Figure 2.4. In the $m_H < 130$ GeV region, the dominant decays are $H \to b\bar{b}$, $H \to \tau^+\tau^-$, and $H \to gg$, where the gg decay arises through a heavy quark loop. Above this mass, the branching ratios for decays into WW and ZZ become dominant as m_H nears the threshold for the decay into on-shell Ws and Zs. At the measured Higgs mass, the dominant decays, ordered by $\mathcal{B}$, are $b\bar{b}$, W^+W^-, gg, $\tau^+\tau^-$, $c\bar{c}$, and



(a) Higgs production cross section.

(b) Higgs decay branching fractions.

FIGURE 2.4: Higgs boson production cross sections and branching fractions as a function of the Higgs mass parameter [3].

ZZ. Though there is no direct coupling to $\gamma\gamma$ and $Z\gamma$, the Higgs can decay to these particles through a heavy quark or a W boson loop. These two decays, along with $H \to \mu\mu$, are the remaining three decays at $m_H = 125.4$ GeV. The exact values are shown in Table 2.1.

Scattering experiments that aim to measure the Higgs boson are guided by the above SM predictions for Higgs cross sections and branching fractions. In principle, the best experimental sensitivity is achieved by isolating the dominant production mechanism and decay; however, in practice, this is not always feasible due to experimental considerations. Hadron colliders produce enormous backgrounds from QCD processes with final states consisting of quarks and gluons. The Higgs process that is produced at the highest rate at $m_H = 125.4$ GeV, $gg \to H \to b\bar{b}$, lies in a region of phase space that is saturated by irreducible QCD background, making it impossible to observe this Higgs decay. To suppress QCD backgrounds, final states are required to have photons or at least one charged lepton. The most sensitive decay channels at the LHC are $\gamma\gamma$, $ZZ^{(*)} \to \ell\ell\ell\ell$, and $WW^{(*)} \to \ell\nu\ell\nu$. The former two channels

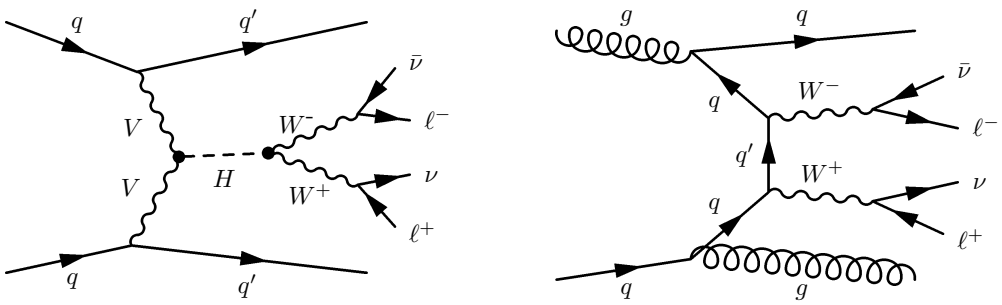
compensate for a smaller $\mathcal{B}$ because the final state particles allow the resonant peak of the Higgs to be resolved against the background. Furthermore, there are few SM backgrounds that produce the 4ℓ final state associated with the $ZZ^{(*)}$ decay channel.

Despite its relatively large $\mathcal{B}$, the $WW^{(*)}$ channel is of a comparable sensitivity to $\gamma\gamma$ and $ZZ^{(*)}$ channels due to the nature of the W boson decay. The W boson decays predominantly (67.6%) to hadrons, and to a lesser extent a charged lepton and a neutrino (32.4%). The large background from QCD processes prohibits finding $WW^{(*)}\rightarrow\ qqqq$, and large backgrounds remain even if one of the W decays is leptonic. The most sensitive $WW^{(*)}$ decay channel is therefore $WW^{(*)}\rightarrow\ell\nu\ell\nu$. The neutrinos in the final state only interact through the weak interaction, and are therefore not detected. Without the ability to reconstruct the neutrino momenta, it is impossible to reconstruct the invariant mass of the parent Higgs boson. This makes it challenging to resolve the Higgs mass peak against the background from non-resonant WW production, thereby degrading the sensitivity of this decay channel. In spite of this challenge, it is currently the most sensitive channel in ATLAS.

2.4 The $H\rightarrow WW^{(*)}\rightarrow\ell\nu\ell\nu$ Channel

In this thesis, a measurement of vector boson fusion production in the $WW^{(*)}\rightarrow\ell\nu\ell\nu$ decay channel is presented. The Feynman diagram for this signal process is shown in Figure 2.5(a). The products of the Higgs decay manifest as two oppositely charged leptons and missing transverse energy due to the undetectable neutrinos. Additionally, the VBF production mechanism results in two forward jets between which QCD radiation is suppressed. The selection of such a signature is detailed in Chapter 7. In this section, the discussion focuses on the kinematic properties of the $H\rightarrow WW^{(*)}\rightarrow\ell\nu\ell\nu$ decay chain.

Global conservation laws and the group structure of the SM dictate the kinematic behavior of the final state leptons in the decay $H\rightarrow WW^{(*)}\rightarrow\ell\nu\ell\nu$. The Higgs boson



(a) VBF $H \to WW^{(*)} \to \ell\nu\ell\nu$ (b) Continuum $WW^{(*)} \to \ell\nu\ell\nu$

FIGURE 2.5: Feynman diagram for (a) Higgs production via VBF in the $WW^{(*)} \to \ell\nu\ell\nu$ decay channel, and (b) an example of non-resonant production of a pair of W bosons in association with two jets.

is spin-0 and the two daughter W bosons are spin-1. In the rest frame of the Higgs boson, in order to conserve angular momentum, the W bosons are required to have oppositely aligned spins. Similarly, the leptons from the W boson decays are spin-correlated to their parents.

Since the $W \to \ell\nu$ decay proceeds through a weak current, it violates parity, a discrete spatial symmetry that reverses the sign of the spatial coordinates, $(x, y, z) \to (-x, -y, -z)$. As discussed in Section 2.1, the requirement of gauge invariance in EM gives rise to an interaction term $g\bar{\psi}\gamma^\mu\psi A_\mu$. In other words, the current of the EM interaction has a bilinear form $\bar{\psi}\gamma^\mu\psi$, and transforms as a Lorentz 4-vector, preserving parity. The weak interaction current has a parity-violating bilinear form $\bar{\psi}\gamma^\mu\frac{1}{2}(1-\gamma^5)\psi$, where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. Under parity, the $\frac{1}{2}\bar{\psi}\gamma^\mu\psi$ term transforms as a vector like the EM current, while the space-like components of the other term, $-\bar{\psi}\gamma^\mu\gamma^5\psi$, do no flip sign. A term with the latter property is considered an "axial vector", and the above form of the weak current is referred to as the "V-A" structure of the weak force. The $\frac{1}{2}(1-\gamma^5)$ operator projects the left-handed (negative helicity) eigenstate, where helicity is defined as the component of the spin in the direction of motion, $\frac{1}{2}\boldsymbol{\sigma}\cdot\hat{\boldsymbol{p}}$. Massless neutrinos are eigenstates of the helicity operator, and due

to the form of the weak current, and to the fact that neutrinos only interact weakly, only left-handed neutrinos—and right-handed anti-neutrinos—exist in nature.

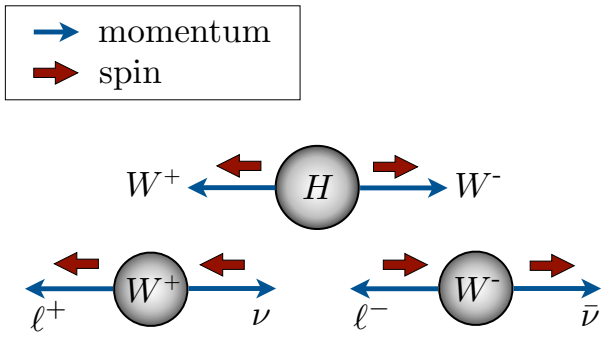


FIGURE 2.6: Schematic diagram of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay chain illustrating how conservation of angular momentum enhances the kinematic configuration in which the opening angle between the charged leptons is small.

In the $WW^{(*)} \to \ell\nu\ell\nu$ decay, since the spins of the final state particles are correlated, and the spin can communicate with the momentum of a particle through the helicity properties of the weak interaction, the momenta of the final state leptons are correlated. This is illustrated schematically in Figure 2.6 for a single WW spin configuration. In this regime, the mass of the charged leptons is negligible, implying that these leptons (anti-leptons) are, like the massless neutrinos, left-handed (right-handed) eigenstates of the helicity operator. Because the spins of the $\ell\nu$ pairs have opposite sign, and matter (anti-matter) predominantly decays anti-parallel (parallel) to the spin direction, the opening angle between the charged leptons in the plane transverse to the incoming particles is typically small. This kinematic arrangement is powerful in rejecting non-resonant WW background processes. An example of such a process is shown in Figure 2.5(b). Without the spin correlations through the Higgs, the distribution of the opening angle is relatively flat (see Section 7.3).

2.5 Conclusion

Proton-proton scattering experiments such as ATLAS provide rich testing grounds for the diversity of phenomena predicted by the $SU(3)\times SU(2)\times U(1)$ gauge structure of the SM. Included in these predictions is the existence of a Higgs boson, the bi-product of the spontaneous breaking of $SU(2)$. The Higgs boson was discovered in 2012, and thus far, measurements of its couplings have been in agreement with the SM predictions. In pp scattering at the LHC energies, the Higgs is primarily produced through gluon-gluon fusion in which the gluons couple to the Higgs through a fermion loop. As a consequence, this coupling has been measured with the highest precision. With increasingly more scattering data recorded, it is becoming possible to precisely measure sub-dominant production processes, like vector boson fusion, in which the Higgs is produced through a coupling to W or Z bosons. This thesis focuses on such a measurement in the $WW^{(*)}\to\ell\nu\ell\nu$ Higgs decay channel, which is of particular interest since it probes the WWH coupling in both the production and decay of the Higgs. Measuring deviations from the SM coupling may hint at a richer underlying group structure in our universe.

3

The ATLAS Detector

The ATLAS experiment is one of the four large experiments at the Large Hadron Collider (LHC) located at the European Organization for Nuclear Research (CERN) outside of Geneva, Switzerland. Buried between 45 m and 170 m under the French-Swiss border, the LHC is designed to accelerate two counter-rotating beams of protons to 7 TeV each[1]. The proton beams are steered with superconducting magnets around an evacuated ring that is 26.7 km in circumference [42]. At several points along the ring, the beams are steered and focused such that the constituent protons collide with high probability at a center-of-mass energy of 8 TeV. One such point is surrounded by the ATLAS detector, a general-purpose detector designed to take a snapshot of the collision remnants [4]. In this chapter, the various subsystems of the ATLAS detector are discussed, and in the following chapter, the reconstruction of a *pp* collision from the ATLAS readout channels is considered.

[1] For technical reasons, the beam energy was lowered to 3.5 and 4 TeV for the dataset used in this thesis.

3.1 Overview

The ATLAS detector is situated along the LHC beam pipe. Its global coordinate system is defined with the z coordinate coinciding with the direction of the proton beam, x pointing towards the center of the LHC ring, and y pointing upwards. The azimuthal angle ϕ is the angle in the $x-y$ plane as measured from the x axis, and the polar angle θ is the angle from the z axis. The origin of this coordinate system lies nominally at the center-of-mass of the detector, though in particle reconstruction, it is shifted to be the pp collision point. In most cases, the θ coordinated is replaced by the pseudo-rapidity, defined as $\eta = -\log(\tan(\theta/2))$, because for massless particles, differences in this quantity are Lorentz invariant with respect to boosts in the z direction. Detector components or particles at small angles with respect to the beam axis, considered to be "forward", lie at large values of $|\eta|$.

ATLAS is designed to be sensitive to the broad range of scattering events expected at the TeV scale. One of the primary design considerations was the search for the Higgs boson. Since many Higgs final states include charged leptons, a high precision tracking system was required. Moreover, the $H \to \gamma\gamma$ channel calls for high performance EM calorimetry to identify and measure electrons and photons. Another equally important consideration in the design was the high expected LHC collision rate and the large QCD backgrounds expected at a pp collider. At its design luminosity, the collision rate of the LHC is 1 GHz. The proton beam is partitioned into ~1500 bunches with up to 10^{11} protons per bunch, making it likely that more than one proton collision occurs every time two bunches cross paths. Also, because the bunches are separated by 50 ns[2], particles from adjacent bunches may appear to be from the colliding bunch. These two phenomena, referred to as pile-up, also shaped the design of ATLAS. To deal with these overlapping events, the ATLAS

[2] The design bunch spacing is 25 ns with proton beams of 3000 bunches.

components have fast recovery times and fine granularity. Also, the tracking system close to the collision region has high resolution, allowing overlapping collisions to be distinguished in reconstruction.

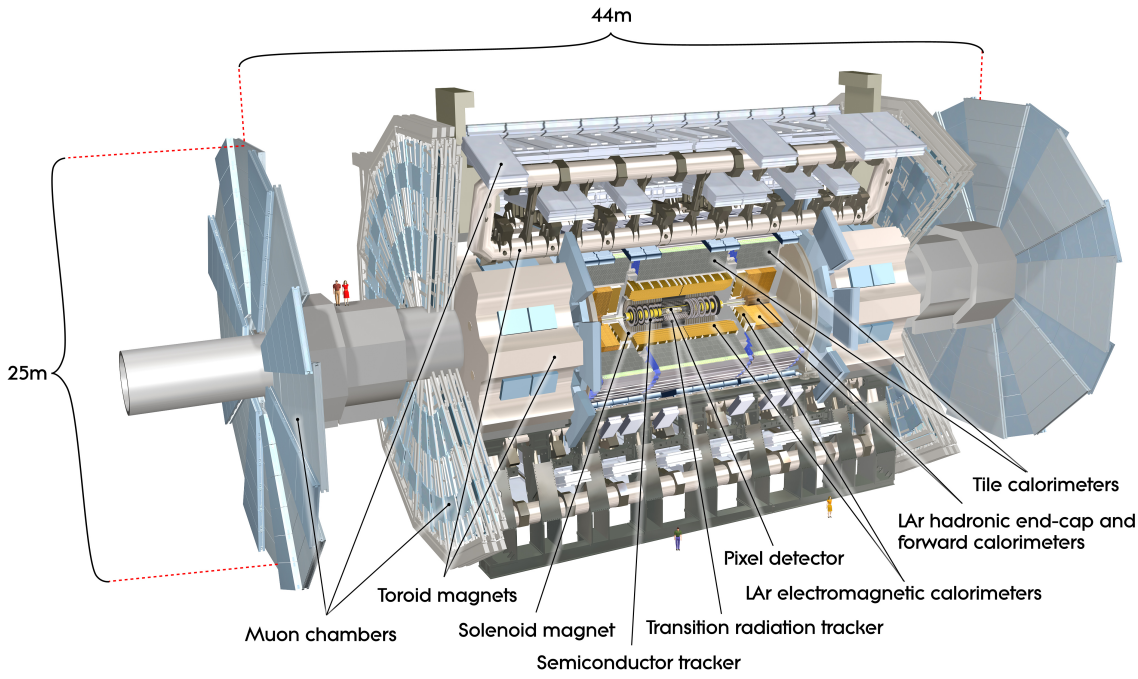


FIGURE 3.1: Diagram of the entire ATLAS detector. The cross sectional view exposes the inner detector, the calorimeter system, and the muon spectrometer system.

The resulting design of the ATLAS detector is displayed in Figure 3.1. Lying closest to the collision point, the inner detector (ID) is the primary tracking system. To resolve the momentum of charged particles, it is immersed in a uniform 2 T magnetic field. Beyond the ID are the EM and hadronic calorimeters which measure the energy of photons, electrons, and hadrons produced in the collision. The outermost detectors comprise the muon spectrometer (MS) system. These detectors are embedded in a high-bend toroidal magnetic field, resulting in precision tracking across a large momentum range. The detector is nominally $\pm z$-symmetric and has eight-fold azimuthal symmetry due to the toroid magnet system. The various subsystems are

segmented in the z direction into a barrel region with a concentric cylinder geometry, and two end-cap regions with components that are "wheels" or "disks" that fit against the barrel ends, thereby increasing acceptance.

3.2 Inner Detector

Charged particles produced in collisions traverse the ID, depositing on sensors signals that are recorded as spatial coordinates. These coordinates are then processed through a pattern recognition and reconstruction algorithm to extract the particle tracks from which the momentum four vectors are derived (see Section 4.1). The ID is designed to measure tracks across a large momentum range, from $O(100$ MeV) to $O(1$ TeV), falling in the pseudorapidity range $|\eta| < 2.5$. It is composed of three subdetectors: the pixel tracker, the silicon microstrip tracker (SCT), and the transition radiation tracker (TRT). A cross-sectional diagram of the ID in the barrel region is shown in Figure 3.2.

3.2.1 Silicon Detectors

With nearly 50 pp collisions per beam crossing producing $O(1000)$ particles, the detectors near the collision point are required to have high resolution, fine granularity, fast response, and radiation hardness. These requirements are satisfied by the silicon pixel detectors. The 1744 identical pixel sensors of the pixel tracker are arranged in three concentric layers in the barrel region (Figure 3.2) and in three disks in each end-cap. Each sensor is composed of ~47k pixels of size 50 μm $\times$ 400 μm, corresponding to 80M readout channels. The intrinsic spatial resolution of each pixel is 10 μm $\times$ 115 μm, and the sensor is placed such that the precision pixel direction is the global azimuthal direction in which charged particles bend. Due to the high efficiency of the pixel sensors, the average charged particle track in the ID volume will result in three precision spatial measurements from the pixel tracker.

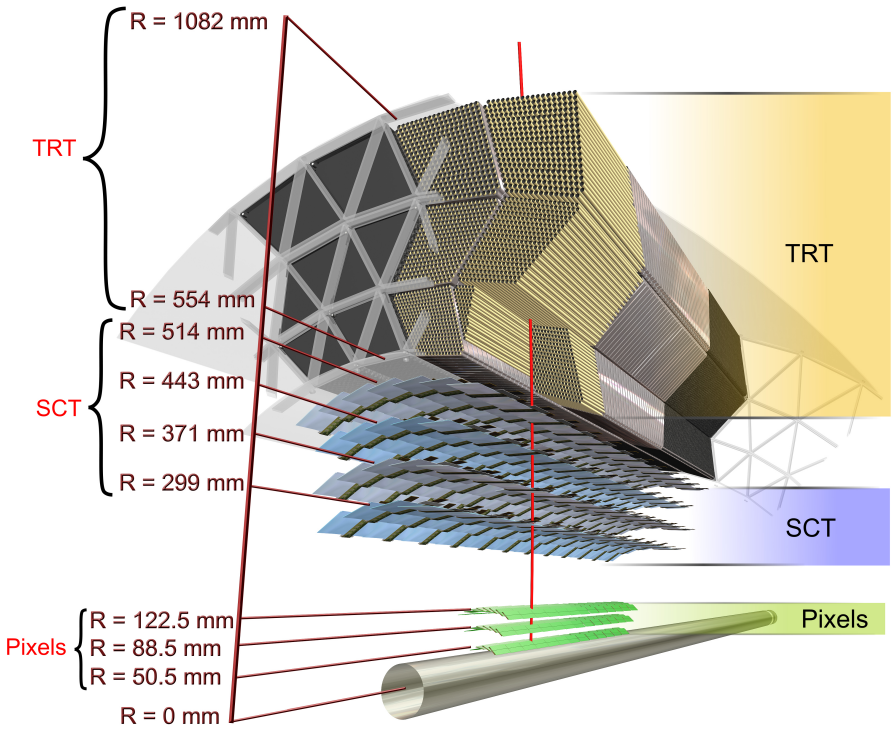


FIGURE 3.2: Cross-sectional diagram of the inner detector in the barrel region illustrating the radial layout of the subsystems.

Beyond the pixel detector lies the second silicon sub-detector in the ID, the SCT. The SCT relies on more traditional technologies at the cost of degraded detector precision. In the barrel region, the SCT modules are tiled to form four concentric layers (Figure 3.2), while in each end-cap, the modules form nine disks, amounting to a total of 4088 modules. Each barrel module is composed of four nearly square (64.0 mm $\times$ 63.6 mm) silicon sensors, each with 768 readout strips that are 22 μm in width [43]. Two sensors are placed side-to-side—with a 2 mm gap for readout electronics—on top of a thermal pyrolitic graphite substrate, which provides mechanical support and the thermal conductivity necessary for cooling the sensors.

Another two sensors are placed on the other side of the substrate, and displaced by an angle of 40 mrad. This stereo angle configuration allows another spatial degree of freedom to be measured (z in barrel, R in endcap). The end-cap sensors and modules are similar except for adjustments to the dimensions. The intrinsic resolution of the SCT is 17 μm in the azimuthal direction, and the effective resolution in the z (R) direction in the barrel (endcap) is 580 μm. An average track in the ID volume will result in four precision spatial measurements in the SCT.

3.2.2 Transition Radiation Tracker

The TRT is the outermost sub-detector in the ID, spanning the region 56 cm $< R <$ 108 cm in the barrel [44]. It is a collection of polyimide-based drift chamber tubes (straws) that are 4 mm in diameter. The tube wall is the high voltage cathode composed of layers of polyimide, graphite-polyimide, polyurethane, and a 0.2 μm layer of aluminium to achieve the requisite electrical and mechanical properties [4]. The anode is a 31 μm gold-plated tungsten wire running directly through the center of the straw with a radial offset of less than 300 μm. Each straw is filled with a gas mixture of 70% Xe, 27% CO_2, and 3% O_2.

In the TRT barrel region, spanning $|\eta| < 1.0$, the TRT straws run parallel to the beam axis. They are placed in carbon-fiber-shelled modules in which the straws form uniform arrays with an average spacing of 6.6 mm. The straws in the module are 144 cm long. To accommodate high particle multiplicity, the wires within each straw are electrically disconnected at the straw center, and the $+z$ and $-z$ halves of the straw record independent hits. At either end of the module, there is a high voltage plate that couples to the straw walls and a front-end electronics board. A gas inlet circulates CO_2 outside of the straws, thereby preventing electrical discharges and the accumulation of any leaked Xe. This gas bath also cools the straws by conducting heat to the highly thermally-conductive shell of the module. With a quadrilateral

prism geometry, the modules are arranged into three concentric rings, each comprised of 32 modules.

The TRT end-caps provide tracking in the region $1.0 < |\eta| < 2.0$. Each end-cap consists of two sets of wheels. The set closer to the collision point is composed of 12 wheels, each with eight straw layers spaced at a wire-to-wire distance of 8 mm [45]. The other wheel set has only eight wheels and a layer spacing of 15 mm. In each wheel, there are 768 straws of length 37 cm oriented radially outwards with uniform azimuthal spacing. To achieve better uniformity in the number of straws that are traversed, each successive wheel is rotated by 3/8 of the azimuthal spacing. Similar to the barrel modules, each wheel is a self-contained unit in which CO_2 circulates around the straws.

In total, there are 52544 straws in the barrel and 122880 straws in each end-cap. On average, a charged particle with a momentum greater than 500 MeV and $|\eta| < 2.0$ will traverse 36 straws, except in the transition region between the barrel and end-cap, $0.8 < |\eta| < 1.0$. Each straw is capable of providing a measurement of the azimuthal distance of approach of the track at an intrinsic resolution of 130 μm. Therefore, in the barrel (end-cap), the TRT measures an $R - \phi$ ($z - \phi$) coordinate but no z (R) information.

An important design feature of the TRT is the presence of radiator material between straws. Electromagnetic transition radiation occurs when a charged particle traverses the interface between two media with different dielectric constants. The average energy of the radiated photon grows linearly with the relativistic γ factor of the charged particle. Since electrons are less massive than charged hadrons, they have larger γ ($E = \gamma mc^2$), and hence more transition radiation. To take advantage of this phenomenon, the TRT barrel straws are embedded in a matrix of polypropylene fibers that serve as the radiation material. In the endcap, polypropylene foils lie between the layers of TRT straws. When an electron traverses the radiator material, it emits

photons that are then absorbed by the straw gas, producing significantly larger signal amplitudes than expected with minimum-ionizing charged particles. The front-end electronics distinguish between the high threshold hits typical of electrons and the low threshold hits observed for heavy charged particles. An electron with p_{T} greater than 2 GeV will typically have seven to ten high threshold TRT hits along its trajectory.

3.3 Calorimetry

The ATLAS calorimeters are designed to efficiently capture, and precisely measure, the energy of photons, electrons, and hadrons. They are symmetric about the beam axis and extend beyond the ID to $|\eta| = 4.9$ [4]. The electromagnetic calorimeter, optimized for electrons and photons, is segmented into a barrel and two end-caps. Its barrel is housed in a liquid argon cryostat. The two end-caps have separate cryostats that also contain the hadronic end-cap calorimeter (HEC) and the forward calorimeter (FCal). Beyond the barrel EM calorimeter is the hadronic tile calorimeter. A drawing of entire calorimeter system is displayed in Figure 3.3.

3.3.1 Liquid argon electromagnetic calorimeter

The barrel of the EM calorimeter spans the region $|\eta| < 1.48$, and is split into two halves at $\eta = 0$. Lead plates with an accordion geometry provide the needed absorbing material. This geometry, shown in Figure 3.4, yields uniform azimuthal coverage and fast detector response [4]. The readout electrodes are positioned parallel to the lead plates in the center of the gap between adjacent plates. Each electrode consists of three layers of copper separated by a thin insulating layer of polyimide. The outer layers are set to the high voltage potential, and the inner layer carries the signal to the electronics by capacitive coupling. These electrodes are etched in order to define the granularity in the R and η directions. The ϕ granularity is set by the spacing of the electrodes and how they are integrated into the front-end

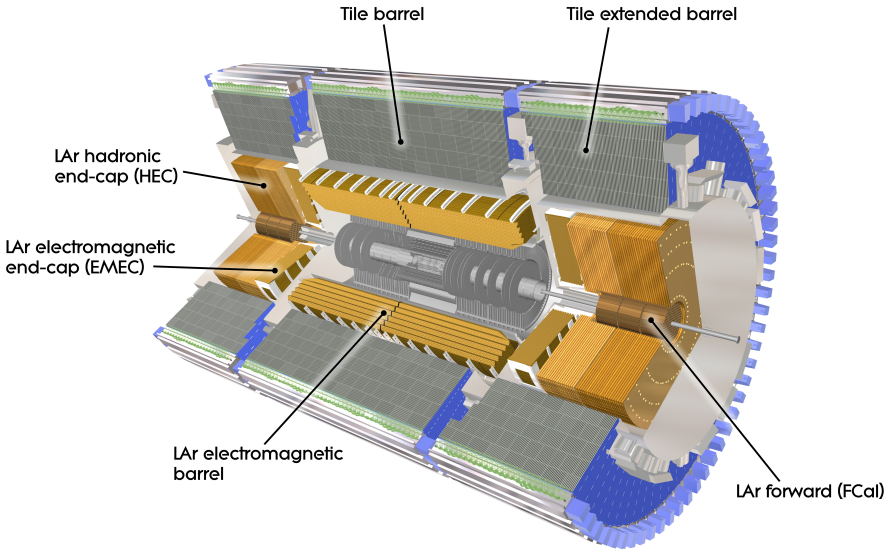


FIGURE 3.3: Diagram of the ATLAS calorimeter systems. In the barrel region, the liquid argon EM calorimeter, shown in gold, lies just beyond the inner detector, and the tile hadronic calorimeter, shown in gray, is concentric with the EM calorimeter. In the forward region, the liquid argon EM calorimeter, hadronic end-cap calorimeter, and forward calorimeter are shown with various shades of gold. Outside of these lies the extended tile calorimeter.

electronics [46]. The calorimeter is subdivided in R into three regions of varying granularity, with the finest granularity closest to the collision point. In both the ϕ and η directions, the granularity in the middle layer is 0.025. On either side of the electrode, the liquid argon gap is 2.1 mm, corresponding to a drift time of 250 ns at a nominal voltage of 2.0 kV. The total thickness of the barrel calorimeter is at least 22 radiation lengths, increasing to 33 as $|\eta|$ increases.

The end-cap segments of the EM calorimeter cover the region $1.38 < |\eta| < 3.2$, with the overlap ensuring that there is no loss of resolution in the barrel-end-cap transition region. Each end-cap is segmented into two concentric wheels, with the inner (outer) wheel composed of 768 (256) lead plates. As for the barrel calorimeter, the plates are accordion-shaped; however, in the end-cap, the perforations are in the

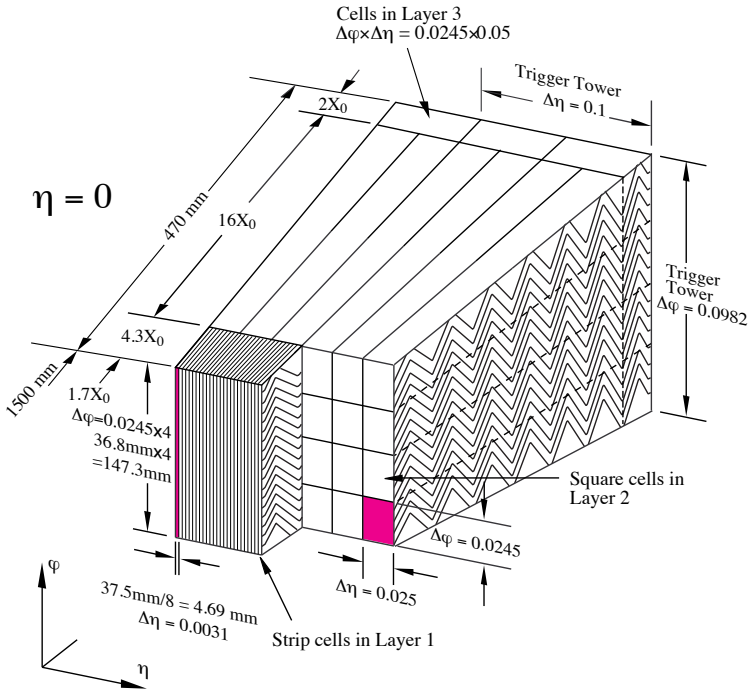


FIGURE 3.4: Diagram of a liquid argon calorimeter module at $\eta = 0$, illustrating the accordion geometry and η-ϕ granularity [4].

R direction instead of the z direction. The R-η segmentation is set by etches in the inter-plate electrodes, forming three layers of varying granularity, with a granularity of $\Delta\phi \times \Delta\eta = 0.025 \times 0.025$ in the middle layer. The thickness of the end-cap calorimeter varies from 24 to 38 radiation lengths, depending on $|\eta|$.

Another important component of the EM calorimeter is the presampler, located in front of the barrel calorimeter and just outside of the solenoid coils [4]. Spanning the region $|\eta| < 1.52$, the function of the presampler is to collect the energy lost by incident particles before reaching the calorimeter. This correction to the calorimeter measurement improves the energy resolution by as much as 40% [47]. The presampler provides full azimuthal coverage with 32 modules of width $\Delta\phi = 0.2$ for each half barrel of the calorimeter. It has a single active liquid argon layer without any additional absorbing material. To collect signal, sheets of electrodes are positioned

parallel to the x-y plane. Two electrode types—the anode and cathode—are interweaved to achieve a potential difference of 2.0 kV across the active medium gap of 2 mm. The anode has the same three layer configuration as the barrel and end-cap calorimeters, allowing the signal to be read out by the central conductor. Each electrode is etched in the center to give a ϕ granularity of 0.1. The desired η granularity is achieved by electrically connecting adjacent electrodes in the longitudinal direction, resulting in a constant granularity of 0.025.

3.3.2 Hadronic calorimeters

Hadrons from the pp collision undergo nuclear interactions in the detector material. Because these interactions occur at a low rate with respect to the EM processes that deposit energy in the EM calorimeter, additional calorimeters with greater thickness are positioned beyond the EM calorimeters. There are three types of hadronic calorimeters: the barrel tile, the end-cap liquid argon, and the forward liquid argon.

The tile calorimeter is segmented into a barrel that covers the region $|\eta| < 1.0$ and two extended barrels on either side, increasing the coverage up to $|\eta| = 1.7$ [48]. Each segment is divided azimuthally into 64 modules. These modules house 4 mm and 5 mm steel plates, oriented parallel to the x-y plane, that act as the absorbing material (Figure 3.5). In the gaps between plates, there are polystyrene-based scintillating tiles of thickness 3 mm [4]. Ionizing particles crossing the tiles produce high frequency visible light that is collected by optical fibers on either side of the tile. The wavelength of the scintillation light is shifted by the fibers that also carry the light to photo-multiplier tubes (PMTs) on the outer edge of the module, where it is converted to an electrical signal. By grouping fibers together for collection into the same PMT, the desired granularity is obtained. These fiber groups form three layers in the R direction, with the first two having a granularity of $\Delta\phi \times \Delta\eta = 0.1 \times 0.1$ and the third with $\Delta\phi \times \Delta\eta = 0.1 \times 0.2$. The radial depth of the tile calorimeter is

approximately 7.4 nuclear interaction lengths.

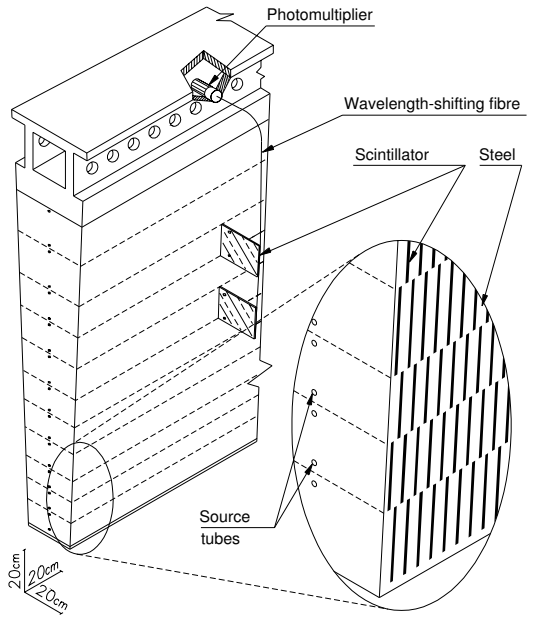


FIGURE 3.5: Diagram of tile calorimeter barrel module, illustrating the layout of the steel plates and scintillating tiles, as well as the mechanism by which scintillation light is converted to an electrical signal.

The acceptance of the tile calorimeter is extended by the hadronic end-cap calorimeter which spans the range $1.5 < |\eta| < 3.2$. Each end-cap consists of two wheels with the same radius but positioned at different z values. Both wheels are partitioned in ϕ into 32 modules. In the inner (outer) wheel, each module contains 24 (16) 25 mm (50 mm) thick copper plates running parallel to each other and to the x-y plane. The plates are separated by an 8.5 mm gap that is further separated into four 1.8 mm liquid-argon-filled drift zones by three electrodes. The central electrode is a 35 μm sheet of copper with 150 μm of insulating material on each side, while the outer two electrodes are 75 μm sheets of carbon-loaded polymide between two insulating layers [49]. With the two outer electrodes connected to high voltage, this electrode configuration forms an electrostatic transformer that is equivalent to

two 3.6 mm drift gaps at two times the high voltage of 1800 V. This configuration has been chosen to minimize ion build up. The detector granularity is defined by etches in the central copper layer, and in the $|\eta| < 2.5$ ($|\eta| > 2.5$) region, it is $\Delta\phi \times \Delta\eta = 0.1 \times 0.1$ (0.2×0.2).

The last hadronic calorimeter component is the forward calorimeter, covering the region $3.1 < |\eta| < 4.9$ [4]. Like the electromagnetic calorimeter and HEC, the ionizing material in the FCal is liquid argon. Each FCal is segmented into three sub-detectors. The first (FCal1), which is closest to the IP, is optimized for electromagnetic showers, and the other two (FCal2 and FCal3) have been designed to contain high energy hadronic showers. Within the 45 mm FCal1 module are plates of copper running parallel to the x-y plane. A total of 12260 electrodes are positioned in a 7.5 mm hexagonal array through holes in the plates. Each electrode is a copper tube coaxial with a copper rod with a gap of 0.269 mm between the copper surfaces. This small drift length has been chosen to avoid ion saturation due to high collision rates [50]. To ensure near uniformity in the FCal1 response, the size of the gap between the copper tube and rod is fixed to within 1% by an insulating fiber that is wound in a helical pattern about the copper rod. A potential of 250 V is applied across the gap (inner rod at high voltage, outer tube at ground), and the signal is read out by coaxial cables. The signals from groups of adjacent electrodes are summed in the front-end electronics, forming a calorimeter cell. Due to the hexagonal geometry of the electrodes, it is not possible to segment into cells of fixed η-ϕ dimensions. Instead the electrodes are grouped into 16 ϕ bins, each with four η bins.

The FCal2 and FCal3 modules are similar in structure to the FCal1. The primary difference is that the electrodes are surrounded by tungsten slugs and the electrode rod is tungsten in order to increase the absorptivity. Additionally, the spacing of the electrodes is increased such that the η distance spanned by a group of electrodes is approximately constant across the three FCal modules. The gap spacing between

the electrode rod and tube also increases, to 0.369 mm in FCal2 and 0.508 mm in FCal3. Finally, just beyond FCal3 is a passive brass plug to shield the muon system from radiation that has punched through the FCal.

3.4 Muon Spectrometers

With a rest mass that is 200 times greater than that of the electron, muons deposit a small fraction of their energy in the calorimeter. The muon spectrometer system, positioned just outside of the calorimeters, precisely measures the muon momentum in the range $|\eta| < 2.7$ to complement the measurement provided by the ID [4]. It also triggers on muons in $|\eta| < 2.4$. This system has been designed to measure muon momenta up to $O(1$ TeV) with a relative resolution of 10%. In order to do so, there are three layers of precision tracking chambers; in the barrel, these layers are concentric cylinders at ($R = 5$, 7.5, 10 m), and in each end-cap the chambers form four parallel disks in the x-y plane at $z = 7.4$, 10.8, 14.0, 21.5 m. Superconducting toroidal magnets provide a B-field for the precision measurement. Fast trigger chambers are capable of providing rough muon track information in 10s of nanoseconds, allowing a trigger decision to be made. The high temporal resolution of these chambers makes it possible to match a muon to the associated beam crossing. Figure 3.6 shows the geometry of the muon system.

3.4.1 High precision tracking chambers

The precision tracking system consists of three (four) layers of chambers in the barrel (end-cap) that are immersed in a toroidal magnetic field. In the barrel, there are eight toroidal magnetic coils positioned symmetrically in ϕ, and for each coil, there is a pair of chambers. The chamber pairs form an alternating set of large and small rectangular chambers, with adjacent chambers overlapping in ϕ to minimize acceptance loss. In the end-caps, each wheel is also composed of overlapping large

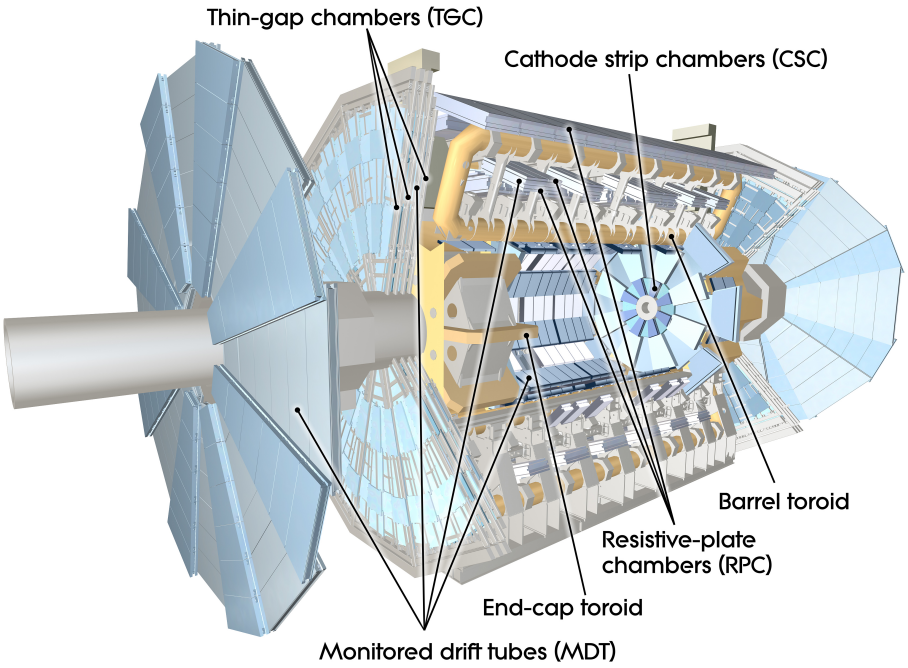


FIGURE 3.6: Cross-section diagram of the ATLAS muon spectrometer system. The locations of the four types of chambers—resistive plate chambers, cathode strip chambers, thin gap chambers, and monitored drift tubes—are indicated.

and small chambers, but the geometry of the chambers is trapezoidal.

Two precision tracking technologies have been adopted. The predominant chamber type is the monitored drift tube chamber (MDT). These chambers contain parallel 3.0 cm copper tubes that run parallel to the ϕ direction in both the barrel and end-caps. Inside of the tube is a 50 μm tungsten-rhenium wire at 3080 V that collects the ionization electrons from interactions between incident muons and the $ArCO_2$ gas between the tube and the anode. Tubes are arranged into layers and segmented in η and ϕ. Parallel tubes are separated by a 60 μm layer of glue for structural support. The spatial resolution of an MDT is limited by the degree to which the position of the tube is known. Each chamber is equipped with an internal optical alignment system capable of measuring deformations at the level of a few μm. Moreover, to avoid resolution degradation due to a sagging anode wire, which is expected to be

~1 mm, the tension of the wire can be adjusted. The resulting spatial resolution of a single tube is 80 μm.

In the inner region of the first end-cap layer, the counting rate exceeds the limit for MDTs, and therefore another chamber technology, the cathode strip chamber (CSC), is introduced. These chambers span the region $2.0 < |\eta| < 2.7$ and have the same alternating large and small plate configuration as the end-cap MDTs. Each chamber contains four layers of side-by-side parallel anode wires with the central wire oriented radially. The cathodes, which are 2.5 mm from the wire plane, are strips on either side of the wires forming two planes. In one plane, the strips run perpendicular to the wires, providing a precision spatial measurement in the bending direction. The other cathode has strips running parallel to the wires and is segmented more coarsely. With this configuration, the CSC provides four two-dimensional measurements with a spatial resolution of 40 μm in one direction and 5 mm in the other. Moreover, the temporal resolution is about 7 ns per plane. This makes bunch-crossing identification possible in a high particle density environment.

3.4.2 Trigger chambers

The muon trigger system is an important component of the muon spectrometer. It is designed to (1) discriminate muons based on transverse momentum, (2) associate a muon with a particular bunch crossing, (3) provide fast and coarse tracking for high level triggers (see Section 3.5), (4) give another spatial measurement to complement the MDT, and (5) be robust against neutron and photon backgrounds [4]. Due to the fact that the environments are quite different in the barrel and end-caps, two different technologies are in place in these regions.

In the barrel region ($|\eta| < 1.05$), three layers of resistive plate chambers (RPCs) form the trigger system. Two of the layers sandwich the second MDT layer while the third is located immediately in front of, or behind, the last MDT layer. Each RPC

has two 2 cm drift layers with parallel plates and a series of strips positioned such that both ϕ and η are measured. In each of the end-caps, where backgrounds are more problematic, thin gap chambers (TGCs) are used instead of RPCs. There is one TGC layer in front of the second MDT wheel, two behind the same wheel, and a fourth layer immediately in front of the first tracking chamber layer. Each chamber in the TGC system has two or three drift layers, each providing an independent measurement of R and ϕ. A drift layer consists of two parallel graphite plates that are parallel to the x-y plane. Between the plates are parallel anode wires that run tangential to ϕ in the center of the chamber. The gaseous ionizing medium is a mixture of CO_2 and n-pentane. On the sides of the plates that do not face the anodes, copper strips that are in contact with the graphite and run tangent to R allow the azimuthal coordinate to be read out.

3.5 Trigger System

Due to limited resources, it is not possible to store the information associated with each collision event at the design collision rate of 1 GHz. The trigger system selects potentially interesting events, thereby reducing the effective collision rate, before events are stored for downstream analysis. It is organized into three levels—Level-1 (L1), Level-2 (L2), and event filter (EF). This three level structure seeks to deal with a fundamental problem of selecting potentially interesting events, namely that such a decision requires partial event reconstruction, which by its nature, requires time and memory. The levels progress from very fast and coarse reconstruction applied to events which are collected at a high rate to a more refined reconstruction applied to a small subset of these events.

The L1 trigger integrates a limited amount of information from all of the calorimeter subsystems and the muon trigger chambers to identify high E_{T} electrons, photons, jets, and muons. The calorimeter trigger system searches for energy deposits in a

coarse granularity cell of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ that fall above a series of configured thresholds. If the multiplicity of these deposits falls above an energy-threshold-dependent cut-off value, then the event passes the trigger. The muon trigger system uses RPC (TGC) measurements in the barrel (end-caps) to identify patterns that are consistent with a muon originating from the collision point. Starting with a muon hit in the second chamber layer, the trigger searches for additional hits along a path formed by extrapolating to the collision point, and if a minimum number of hits is found, the pattern is considered a muon candidate. If the number of muon candidates for the same bunch crossing falls above a threshold, the L1 muon trigger is accepted. The trigger is binned into 6 transverse momentum bins. The L1 trigger reduces the event rate from 1 GHz to 75 kHz with each trigger decision executed in less than 2.5μs. L1 trigger processing is done in the front-end electronics system on the detector, as is the storage of event data in buffers. Once the L1 trigger is passed, the data are sent away from the detector to readout drivers, awaiting the L2 trigger.

The L2 trigger is considered a high level trigger (HLT) in that it makes a trigger decision based on the full granularity of the detector and even some information from the ID. As opposed to the L1 triggers that use front-end hardware to process an event, the L2 trigger system is built around a specialized software-based framework that runs on a computer farm. L2 triggers consider small event data fragments associated with regions of interest (ROIs) that are defined by the L1 trigger. Data for each ROI is typically on the order of 1% of the total data in the event, which speeds up the L2 trigger processing and minimizes the amount of data transferred to the farm. The L2 trigger algorithms iteratively pull data from the readout drivers and determine whether an ROI satisfies the hypothesis for a given particle. If the algorithm determines that the ROI is not consistent with a particle, the next ROI data is pulled and the process repeats. If none of the ROIs are found to be consistent with particles, the event is rejected. In its current form, the L2 system can only

accommodate an event rate of 40 kHz, about half of the design trigger rate for the L1 trigger. L2 further reduces the trigger rate to 3.5 kHz, with a processing time of around 40 ms for each event.

The final trigger level is the event filter (EF) trigger. It is also a HLT, but instead of using full granularity portions of the event (ROIs), it improves on the L2 trigger by incorporating all of the event data into the processing algorithms. Event processing is done offline in computer clusters at an average rate of four seconds per event, and the resulting trigger rate is reduced to 200 Hz. Events that pass the EF trigger are transferred to the CERN computer center for permanent storage. The raw data for each event amounts to around 1.3 MB on average.

4

Particle and Event Reconstruction

A set of algorithms has been developed to efficiently reconstruct the stable particles produced by a proton-proton collision event. In the following chapter, the various reconstruction algorithms are discussed, starting with tracking and calorimeter clustering and then focusing on the physical objects relevant to the VBF analysis: electrons, muons, jets, b-hadron jets, and missing transverse energy.

4.1 Inner Detector Tracks

Interesting scattering processes nearly always include charged particles in the final state. The two tracking systems in ATLAS, the ID and the muon system, are designed to measure the trajectories of these charged particles, from which momentum 4-vectors are derived. A track is fully specified by five parameters $\boldsymbol{\alpha} = (q/p, \theta, \phi, d_0, z_0)$, where q is the charge, p is momentum, θ is the angle with respect to the beam line, ϕ is the azimuthal angle, d_0 is the distance of closest approach (usually to the vertex associated to the track) in the transverse plane, and z_0 is the distance of closest approach in the longitudinal direction. In the following section, the algorithms for fitting these parameters from detector hits will be

described.

4.1.1 Inside-out Tracks

The primary tracking algorithm in ATLAS builds tracks starting with hits in the pixel detectors, and progressively adding hits at larger r values. This is known as the inside-out track reconstruction paradigm [51]. Starting with hits from the pixel and SCT detectors, collectively known as the silicon detectors, clusters of contiguous hits are identified and associated with a coordinate in space. The track-finding algorithm is seeded by a group of three clusters in different layers that is consistent with a track. This seed defines a "road" along which the track candidate is built. The track trajectory is propagated through the detector in a Kalman filter based approach [52]. At the k^{th} detector layer, the hit that is most consistent with the track parameters from the previous layer is assigned to the track. The track parameter vector is then updated with the new hit included, yielding $\boldsymbol{\alpha}_k$. The propagation to the subsequent layer is then given by

$$\boldsymbol{\alpha}_{k+1} = M_k \boldsymbol{\alpha}_k + \boldsymbol{\epsilon}_k \tag{4.1}$$

where M_k is the linear map representing the magnetic field between the two detector layers, and ϵ_k is a stochastic term that accounts for multiple scattering of the charged particle. The procedure is then repeated at layer $k+1$. The above approach for fitting the track parameter vector is equivalent to a global minimum least-squares fit.

After all of the seeds have been evaluated by the Kalman-filter algorithm, ambiguities are removed from the resulting track candidate collection. These ambiguities include track candidates that share one or more hits, are incomplete, or include hits that arise from more than one charged particle. In order to resolve these ambiguities, the tracks are first refitted using a detector geometry with a more realistic description of the detector material. Each track is then assigned a score that quantifies the

quality of the track. Tracks with more hits receive higher scores, and these scores are weighted according to the precision of the detector subsystem in which the hits reside. Given the high intrinsic efficiencies of the silicon detectors, if a hit is absent along a track trajectory, the track score is penalized. After track scoring, hits that are shared by more than one track are assigned to the track with higher score. The remaining track is then refitted and re-scored. Tracks that fail to pass a score threshold are discarded.

The silicon only track candidates that pass the quality threshold are the seeds for the extension of the track into the TRT. In the first pass, the silicon hits that define the track seed for the TRT extension are not changed. Again, the Kalman filter approach is used to map the track from one detector surface to another. Along the trajectory, TRT hits are added depending on the distance between the Kalman filter prediction and the actual hit location. With the TRT hits identified, the track is refitted with the hits from all sub-detectors—silicon and TRT. In this pass, the silicon hits are allowed to vary. The score of the refitted track is compared to that of the silicon only track, and if it is smaller, the silicon only track is retained, with the hits from the TRT as outliers, but still associated to the track. If the score of the refitted track is larger, this new track is retained.

4.1.2 Outside-in Tracks

Because the inside-out algorithm is seeded by silicon clusters associated to a primary vertex, it fails to find tracks arising from secondary decays that occur late in the silicon system or in the TRT. These tracks include those associated with electrons from photon conversions or leptons from in-flight hadron decays. To recover secondary tracks, an outside-in algorithm, starting in the TRT, is used in concert with the inside-out algorithm.

Since TRT hits do not have information about the global z coordinate in the

barrel, it is not possible to use pattern-finding procedure based on three-dimensional space coordinates as is done in inside-out tracking. Instead, the TRT hits are projected to the $r-\phi$ plane in which they form straight lines. The hits are then Hough transformed [53] to the parameter space of a line, where the coordinates are the y-intercept, or the initial ϕ of the line, and the slope, or $1/p_{\text{track}}$, where p_{track} is the momentum of the track. In this space, track candidates appear as maxima at their respective $(\phi_0, 1/p_{\text{track}})$ coordinates. This technique, which defines the hit $r-\phi$ coordinate to be the center of the straw, fails to incorporate information about the drift time. In a second step, the track candidates are refitted with a Kalman-filter-based algorithm that incorporates information about the drift time, thereby improving the accuracy of the fitted track parameters.

The final step of the outside-in track reconstruction algorithm is the extension of the TRT segments into the silicon tracker. Scanning longitudinally in a small $r-\phi$ wedge defined by the TRT segment, the algorithm searches for at least two clusters of SCT hits in the three outermost layers of the SCT. Once found, the curvature defined by these two three-dimensional points and the first TRT hit in the TRT segment is computed, and if the result is unreasonable, the two clusters are disregarded. The SCT clusters that are retained seed the track-fitting algorithm, which propagates the track to smaller r, extracting the fitted track parameters.

4.1.3 Vertex Reconstruction

The vertex of a track refers to the three-dimensional spatial coordinate from which it originates. Primary vertices are associated with tracks that have likely been produced as a result of pp scattering in the interaction region, while secondary vertices are associated with tracks that result from in-flight decays of unstable particles, e.g. K^0_S decays, or from interactions with the detector material, e.g. photon conversions. Robust reconstruction of primary vertices is crucial for suppressing backgrounds due

to in-time pileup. Moreover, the identification of b-quark initiated jets, which is needed to reject top quark processes in the VBF analysis, relies on the ability to detect displaced vertices close to the interaction region.

Reconstruction of primary vertices is split into two steps [54, 55]. The first is the identification of vertex candidates based on track parameters from the reconstructed track collection defined in the previous two sections. In order to isolate tracks that are consistent with primary vertices, requirements are placed on the track p_{T}, d_0, $\sigma(d_0)$, $\sigma(z_0)$, and the number of hits in the silicon detectors. In this case, d_0 and z_0 are measured with respect to the center of the luminous region, also known as the beam spot. The luminous region is determined for each collision run in an unbinned likelihood fit to the distribution of primary vertices recorded in the run. Primary vertices in this fit are reconstructed on-the-fly with a fast reconstruction algorithm, trading off accuracy for speed [56]. With the tracks selected, the vertex-finding algorithm scans the z_0 distribution, and each maximum is retained as a seed for a vertex candidate.

The second step in vertex reconstruction is the fitting of the vertex position and error with the constituent tracks. Starting with an initial guess of the vertex position, the χ^2 is computed for each track. Tracks that are incompatible with the vertex position, as measured by the χ^2, are de-weighted. For the next iteration, the vertex position is updated by minimizing

$$L(\boldsymbol{v}) = \frac{1}{2}\sum_{i=1}^{n} w(\chi_n(\boldsymbol{v}))\chi_n^2(\boldsymbol{v}) \tag{4.2}$$

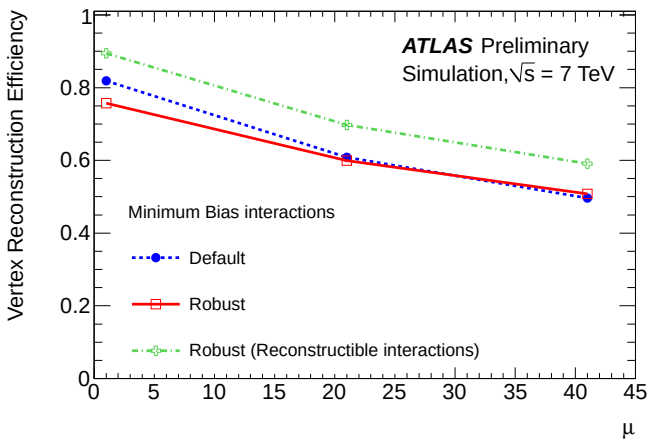
with respect to $\boldsymbol{v}$, the vertex position. Here the sum over the weighted χ^2s is over the tracks associated with the vertex. Tracks that are incompatible with the fitted vertex by more than 7σ are used to seed a new vertex.

As the amount of pileup in the ATLAS detector increases, the vertex reconstruc-

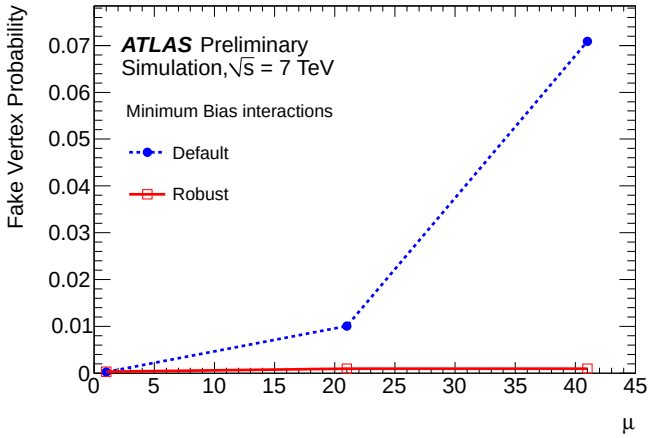
tion efficiency decreases [5]. The primary cause of this is that with high vertex density, two true vertices are more likely to be reconstructed as a single vertex. This effect can be mitigated by placing more strict quality requirements on the tracks that define vertices. In addition to degraded reconstruction efficiency, high pileup results in a greater likelihood to reconstruct a fake vertex due to high hit densities in the ID. The amount of in-time pileup is reflected by the mean number of interaction per bunch crossing, $\langle\mu\rangle$, defined as

$$\langle\mu\rangle = \frac{L \cdot \sigma_{\text{inel}}}{n_{\text{bunch}} f_r} \tag{4.3}$$

where L is the measured instantaneous luminosity, σ_{inel} is the inelastic cross section of pp scattering, n_{bunch} is the number of proton bunches in the beam, and f_r is the frequency of revolution in the LHC ring. The dependence of the vertex reconstruction efficiency and the fake probability on $\langle\mu\rangle$ is shown in Figures 4.1(a) and (b). With $\langle\mu\rangle = 1$, the efficiency of reconstructing a vertex with strict track requirements is 90%, decreasing to about 60% for $\langle\mu\rangle = 41$, a large dependence on pileup in spite of the robust track cuts. The probability of reconstructing a fake vertex, on the other hand, is insensitive to pileup with robust tracks. If the default track selection is used, the fake probability goes from 0% at $\langle\mu\rangle = 1$ to 7% at $\langle\mu\rangle = 41$. In Figure 4.2, the number of reconstructed vertices (N_{PV}) is plotted as a function of $\langle\mu\rangle$ with strict requirements on the vertex tracks for ATLAS data collected at $\sqrt{s} = 7$ TeV. The relationship between N_{PV} and $\langle\mu\rangle$ is linear at low $\langle\mu\rangle$, but as pileup increases, since the vertex reconstruction efficiency decreases, N_{PV} falls below the linear extrapolation.



(a) Vertex reconstruction efficiency



(b) Fake vertex probability

FIGURE 4.1: The vertex reconstruction efficiency (a) and fake vertex probability (b) for minimum bias MC simulation. Vertices built from default track requirements are shown in blue and those built from more strict requirements are shown in green and red.[5]

4.2 Calorimeter Clustering

The ATLAS calorimeter system, described in Section 3.3, is crucial for reconstructing electrons, photons, jets, and missing transverse energy. Just as tracks are reconstructed from hits in the ID, calorimeter cells are grouped into clusters that represent the transverse energy of incident particles. With the energy and direction information from the clusters, it is possible to reconstruct the 4-momentum of a particle.

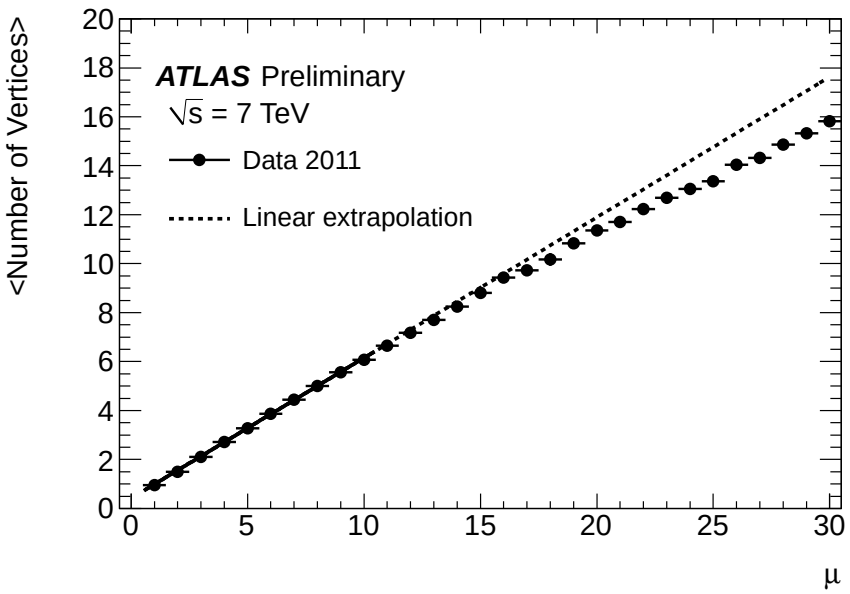


FIGURE 4.2: The number of primary vertices as a function of $\langle\mu\rangle$ for ATLAS data collected at $\sqrt{s} = 7$ TeV.

Moreover, cluster shapes are used to identify the particles in the collision. Two types of clustering algorithms are used in ATLAS: sliding window and topological clustering [57]. The former is effective in the reconstruction of electrons, photons, and hadronic τ decays, while the latter is most performant for jets and missing transverse energy.

4.2.1 Sliding window clustering

In sliding window clustering, calorimeter cells in a fixed rectangular $\eta - \phi$ window are summed, and clusters are built around windows for which E_{T} is maximized. The first step is to partition the calorimeter into an $\eta - \phi$ grid with each grid element having dimension $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. For each grid element, the enclosed cell energies are summed across the calorimeter layers, yielding a "tower", the object from which clusters are constructed. A window of area $\eta_{\mathrm{window}} \times \phi_{\mathrm{window}}$ is scanned across the grid of towers, and at each point, the transverse energy sum of all towers

within the window is computed. If a local maximum above an energy threshold is found, the location associated with this point is computed as the weighted vector sum of the tower energies, and the resulting (η, ϕ) coordinate is used to seed the final step in which the clusters are built. If two of these seeds fall within two η or two ϕ grid units of eachother, the seed with the lower E_{T} is discarded.

Calorimeter cells are assigned to a cluster if they are enclosed by the rectangle of size $\eta_{\mathrm{cluster}} \times \phi_{\mathrm{cluster}}$ centered at the seed defined in the previous step. The dimensions of the cluster rectangle as well as the seed location depend on the calorimeter layer. Furthermore, η_{cluster} and ϕ_{cluster} vary depending on the predicted particle type. In the barrel region, the window size for electrons and converted photons is 3×7 in units of $\Delta\eta \times \Delta\phi$. Because electrons bend in the ϕ direction in the ID B-field and emit *Bremsstrahlung* photons, electron calorimeter showers tend to be elongated in the ϕ direction. Similarly, photons that convert to electron-positron pairs also have showers that are elongated in ϕ, and therefore the window is set to 3×7. Unconverted photons, on the other hand, tend to shower in a smaller ϕ region, and consequently the window dimension is 3×5.

4.2.2 Topological clustering

In contrast to sliding window clustering, the topological clustering algorithm iteratively adds neighboring cells to seed cells, resulting in variable size clusters. Seed cells are calorimeter cells with a signal-to-noise ratio that falls above a large threshold t_{seed}. Signal is defined as the energy deposited in the cell for the EM calorimeter, and as the absolute value of the energy in the hadronic calorimeter. There are two sources of noise: readout electronics and pileup. Noise from the electronics is computed based on the cell gain, while the expected noise due to pileup is estimated from the beam conditions, and the two components are added in quadrature. The noise in hadronic calorimeter cells in the outermost barrel layer as a function of the

mean number of interactions per bunch crossing is shown in Figure 4.3. In these cells, the pileup noise increases fairly linearly with $\langle\mu\rangle$.

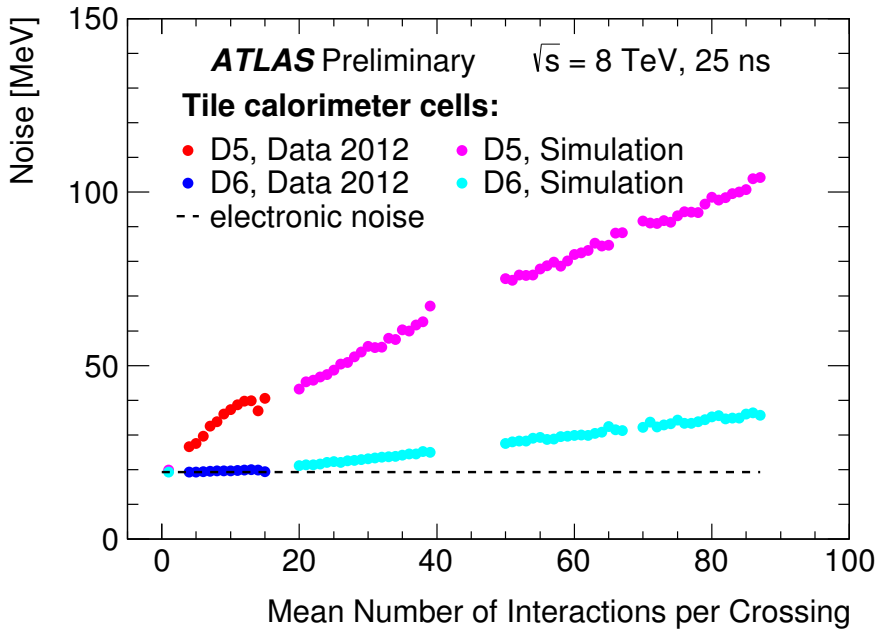


FIGURE 4.3: Noise in hadronic calorimeter cells located in the outermost layer in $0.9 < |\eta| < 1.3$ as a function of $\langle\mu\rangle$. Collision data at $\sqrt{s} = 7$ TeV with 50 ns bunch spacing is shown in blue and red. Data points at $\langle\mu\rangle > 20$ are from simulation. The dashed line is the expected noise from electronics.[6]

Seed cells, called proto-clusters, are ranked according to S/N and for each seed, if a neighboring cell passes another S/N threshold, t_{neighbor}, the neighbor is added to the proto-cluster. In cases where the neighboring cell is adjacent to two nearby seeds, the two proto-clusters are merged. In the next iteration, the neighbors with $S/N > t_{\text{neighbor}}$ become the new seeds, and the procedure is repeated. Neighboring cells which pass a threshold t_{cell}, but fall below t_{neighbor}, are added to the proto-cluster, but do not become seeds in the subsequent iteration. These low S/N cells are retained to better model the tails of the calorimeter shower. Cluster building continues until the list of seeds is exhausted. Because neighboring cells include those in adjacent calorimeter layers in addition to those in the $\eta - \phi$ plane, the resulting

topological clusters are three-dimensional objects which can span across the EM and hadronic calorimeters.

For events with low particle multiplicity, the clustering algorithm described above is sufficient. However, due to the high energy density of actual ATLAS events, the above algorithm creates clusters that span large regions of the calorimeter. To resolve individual particles, the clusters are split into sub-clusters, using local energy maxima within the clusters. If a cell has $E > 400$ MeV, this energy is greater than that of the neighboring cells, and four or more neighbors have S/N above a threshold, the cell is added to a list of seeds. These seeds are then used to grow clusters with the same algorithm in the previous step, except that only the cells that fall within the parent cluster are added, no threshold is applied, and overlapping clusters are not merged, i.e. cells can be associated with two clusters, and the energy of the cell is shared between the two clusters. The resulting collection of clusters is combined with the parent clusters without local maxima to form the final list of topological clusters in the event.

There are two topological cluster types in ATLAS, defined by the S/N thresholds for cluster building. The "electromagnetic 633" ("combined 420") cluster type has $t_{\text{seed}} = 6$ (4), $t_{\text{neighbor}} = 3$ (2), and $t_{\text{seed}} = 3$ (0). Electromagnetic 633 clusters only use cells from the EM calorimeter, while combined 420 clusters can include cells from either calorimeter. With higher S/N thresholds, the rate at which noise results in a cluster is smaller for EM 633 clusters. Combined 420 clusters, on the other hand, have been optimized to find low energy deposits with sufficient noise rejection power.

4.3 Electrons

4.3.1 Electron reconstruction

Electrons and photons appear in many interesting final states, including Higgs decay chains, making it crucial to efficiently reconstruct these particles. In the ATLAS detector, an electron is loosely defined as energy deposits in the EM calorimeter matched to a reconstructed track. Sliding window clustering, described in Section 4.2.1, is used. The efficiency for finding the EM showers associated with a true electron with sliding window clustering is 95% at $E_{\mathrm{T}} = 7$ GeV, increasing to 99.9% at $E_{\mathrm{T}} = 45$ GeV [8].

Tracks associated with electron candidates are reconstructed with the algorithms discussed in Section 4.1, with some important differences to account for the propensity for electrons to emit radiation as they pass through matter. Electron track candidates are extended through the inner detector with the Kalman filter approach. In the first fit attempt, scattering in the detector material assumes the particle is a pion. Upon reaching the calorimeter, if the track does not overlap with a region of interest (ROI), defined as the $\Delta R < 0.3$ region around the center of an EM cluster which satisfies loose shower shape requirements, then starting from the track seed, the track is propagated again, instead assuming an electron-like interaction with the detector material. The electron hypothesis allows for up to 30% energy loss due to *Bremsstrahlung*. All track candidates found to overlap with ROIs are then re-fitted using the global χ^2 track-fitting algorithm. These tracks are extrapolated to the middle layer of the EM calorimeter. For each of these track candidates, one of two requirements must be satisfied: (1) the track falls within $\phi = 0.2$ in the direction of deflection and $\phi = 0.05$ in the opposite direction, and within $\eta = 0.05$ of the center of the EM cluster, (2) after rescaling the track momentum by the measured cluster energy, the track is within $\phi = 0.1$ in the direction of deflection and $\phi = 0.05$

in the other direction of the center of the EM cluster. These two requirements are adjusted slightly for track candidates with less than 4 silicon hits (TRT-only tracks). Requirement (2) is designed to retain low p_{T} tracks that may have suffered significant energy losses before reaching the calorimeter.

With the exception of TRT-only tracks, tracks that pass either of the requirements are then re-fitted again with the Gaussian Sum Filter (GSF) algorithm [58], which is a generalization of the Kalman filter algorithm that accounts for nonlinear effects from *Bremsstrahlung*. The re-fitted tracks are matched to EM clusters again, this time with tighter requirements on the $\Delta\phi$ and $\Delta\eta$ distances. For cases in which more than one track is associated to a cluster, tracks with more pixel hits that fall closer to the cluster center are favored.

With tracks matched to clusters, the calorimeter clusters are then rebuilt in each layer using sliding window dimensions that have been optimized for electrons. The size of the window is 3×7 (5×5) in the barrel (endcaps). In order to reduce the uncertainties on the measured energy, a multivariate regression technique is used to correct for instrumental effects. The regressor, which has been trained on single electron events, determines a correction factor for the total cluster energy, using as independent variables the total energy measured in the calorimeter, the ratio of the presampler energy to the EM calorimeter energy, the shower depth, the pseudorapidity of the center of the cluster in the global coordinate system, and the η and ϕ of the cluster in the calorimeter coordinate system [7]. The performance of the MVA technique is summarized in Figure 4.4. The most probable value (MPV) of the ratio of the measured energy to the true energy as a function of η_{electron}, an estimate of the linearity of the technique, is less than 0.5% across the E_{T} range in the barrel region, while in the forward calorimeter the effect is slightly larger due to more detector material in front of the calorimeter. Similarly, the relative energy resolution (σ/E) grows with $|\eta|$ due to material effects.

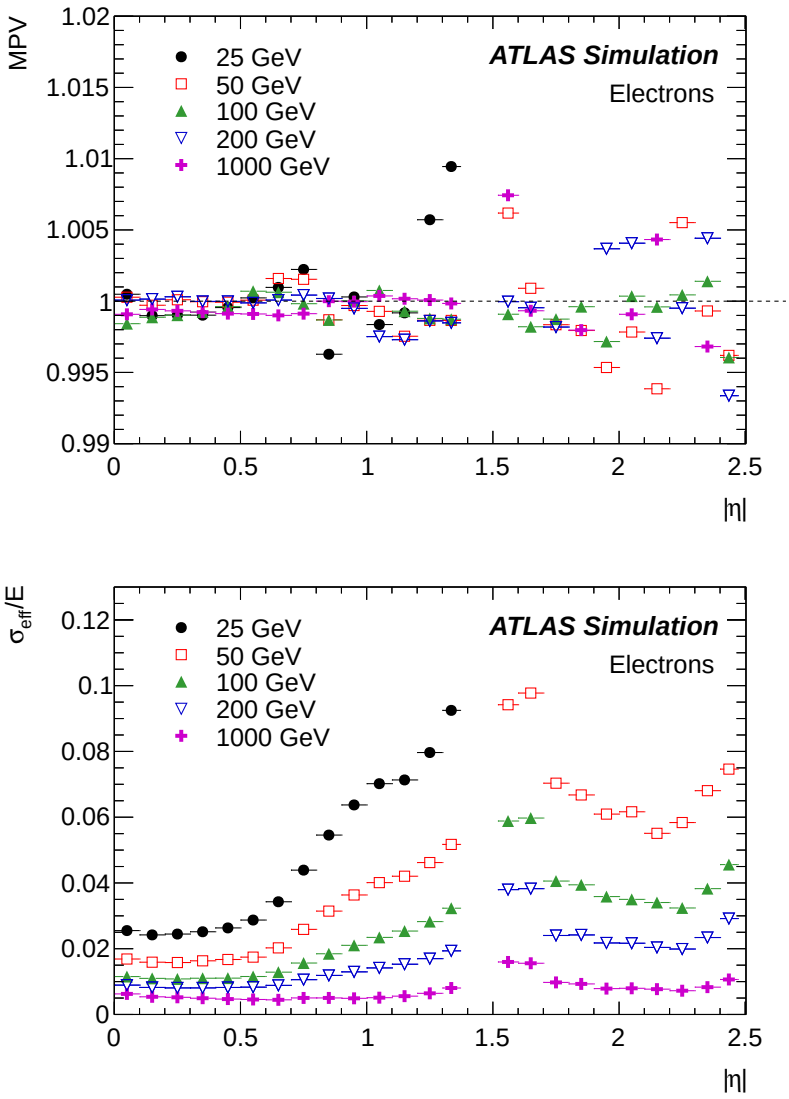


FIGURE 4.4: The most-probable-value (MPV) of the ratio of the measured energy to the true energy (a) and the energy resolution (b) as a function of η_{electron} for electrons of various p_{T}. [7]

The 4-momentum of the reconstructed electron candidate is derived from both calorimeter and track information. The electron energy is taken from the calorimeter cluster, while the η and ϕ are taken from the best-matched track. For TRT-only tracks, η and ϕ from the cluster are used.

4.3.2 Electron identification

Electrons which are reconstructed according to the algorithm described in the previous section can arise from non-prompt backgrounds, including charged hadrons, semi-leptonic heavy flavor hadron decays, the Dalitz decay of the pion ($\pi_0 \to e^+e^-\gamma$), and photon conversions. To suppress these backgrounds while retaining true prompt electrons, candidates are identified using a collection of discriminating variables, either placing sequential cuts on the variables, or using them as inputs to a multivariate algorithm. In the tracking region ($\eta < 2.47$), signal and background electrons are distinguished by variables which describe the EM shower shapes, properties of tracks, as well as the matching between the track and EM cluster [8].

Cut-based electron identification criteria are categorized based on the rejection power. In order of increasing rejection—and decreasing efficiency—the categories are called *loose*, *medium*, and *tight*. Electrons selected with the *tight* criteria are a subset of *loose* and *medium*, and those selected with *medium* are a subset of *loose*. These selection criteria have been optimized in bins of electron E_{T} and η to account for the fact that shower shapes vary with these quantities.

To improve on the cut-based identification approach, the same discriminating variables are used in a likelihood-based multivariate technique. The electron likelihood is defined by the equation

$$L_{s(b)}(\boldsymbol{x}) = \prod_{i=1}^{n} P_{s(b),i}(x_i) \tag{4.4}$$

where $P_{s(b),i}(x_i)$ is the signal (background) probability density function (p.d.f.) for the i^{th} variable evaluated at x_i. The p.d.f.s are obtained from data in control regions. Track hit requirements are not included in the likelihood and are instead left as cuts. As in cut-based electron identification, there are three selection categories

based on background rejection. The LOOSE, MEDIUM, and VERY TIGHT categories correspond to different cuts on the likelihood-based discriminant $L_s/(L_s + L_b)$, and the cut values are chosen such that the efficiencies are approximately the same as the corresponding cut-based categories. In addition to different discriminant cut-values, each likelihood category uses a different set of input variables in the likelihood. The LOOSE category features variables that are efficient in rejecting electron candidates in light-flavor jets, while MEDIUM and VERY TIGHT include additional variables that help in the discrimination against heavy flavor jets and photon conversions. The p.d.f.s that define the likelihood are subdivided into nine $|\eta|$ bins and six E_{T} bins whose boundaries roughly correspond to those in cut-based identification. Due to limited data statistics, however, the binning is coarser for likelihood-based electron identification.

In addition to the identification requirements above, hadronic backgrounds are further rejected by requiring that the electron candidates are isolated. Two types of isolation cuts are applied. Calorimeter-based isolation requires that the sum of the transverse energy in a $\eta - \phi$ radius around the electron is less than some value. Similarly, track-based isolation requires that the sum of the p_{T} of tracks with $p_{\mathrm{T}} > 4$ GeV in a $\eta - \phi$ radius around the electron is small. The specific isolation cuts used in the analysis presented in this thesis are discussed in Section 7.1.1.

Understanding the efficiency—and associated uncertainties—for selecting an electron is crucial to all ATLAS analyses with electrons in the final state. In order to measure the efficiencies, a sample of true electrons is needed. This can be obtained by using the truth record in MC simulation, whereby a true electron is matched to a reconstructed electron, and the fraction of reconstructed electrons in the sample of true electrons is extracted. However, due to the strong dependence on the material model in the simulation, the electron efficiency measurements use ATLAS data in phase space regions where true electrons are likely to reside.

Electron efficiencies are measured with a data-driven technique called "tag and probe" [8]. A tag electron is identified with *tight* requirements, the remaining electron candidates in the events are scanned over, and if the invariant mass of the electron pair falls within 15 GeV of the Z pole mass, the second electron is considered a probe electron and put into a set called the "denominator". Because the invariant mass falls at the Z resonance, the probe electron is likely a true electron from the Z decay. Each electron in the denominator set is then required to pass either a reconstruction or an identification condition, depending on the efficiency being measured, and if a given electron passes, it is placed in the "numerator" set. The ratio of the number of electrons in the numerator to that of the denominator is the efficiency, which is binned in electron E_{T} and η.

The efficiency for selecting an electron can be factorized into four components

$$\epsilon = \epsilon_{\mathrm{reco}} \cdot \epsilon_{\mathrm{identification}} \cdot \epsilon_{\mathrm{trigger}} \cdot \epsilon_{\mathrm{additional}}. \tag{4.5}$$

The efficiency for reconstruction (ϵ_{reco}) is computed for electrons that pass the reconstruction algorithm described in the previous section with respect to a denominator sample of EM calorimeter deposits. The identification efficiency ($\epsilon_{\mathrm{identification}}$) is computed for each identification criteria, where the denominator sample is reconstructed electrons with track quality cuts. The trigger efficiency is obtained from electrons that have already been identified, and the final term is the efficiency associated with isolation cuts or other electron quality cuts. The latter two efficiencies, being dependent on analysis-level cuts, will be discussed in Chapter 7.

Tag-and-probe efficiencies for each identification category are computed in bins of electron E_{T} and ϕ. Tag-probe pairs are required to fall in the Z peak for $E_{\mathrm{T}} >$ 15 GeV, while at lower E_{T}, J/ψ decays are used. The numerator sample consists of probe electrons that have passed a given identification criteria, while the denominator

sample is the superset of those electrons that have passed reconstruction requirements and have one pixel hit and seven silicon hits. Robust estimation of backgrounds to $Z \to ee$ is critical for avoiding biases in the efficiency measurements. To get the M_{ee} shape template for backgrounds, identification cuts are inverted, resulting in an orthogonal background rich sample. These templates are then normalized using the high M_{ee} sideband for the denominator, and for the numerator sample, the region in which the electrons have the same reconstructed charge is used to constrain the normalization.

The tag-and-probe efficiencies for identifying electrons in each category are shown in Figure 4.5, binned in E_{T} (a) and η (b). Electron samples have been obtained with 20.3 fb^{-1} at $\sqrt{s} = 8$ TeV. Efficiencies for the cut-based categories agree with the associated likelihood categories by design, with likelihood identification achieving better background rejection. The efficiency increases with increasing E_{T} for each of the categories with a value of ~95% (89%) for *loose* (*tight*) in the highest E_{T} bin. Features in the η distribution are well-understood. As the identification criteria are tightened, isolated efficiency dips become more pronounced. The dip at $\eta \sim 0$ is due to a small gap between the calorimeter and the TRT, while the drop at $1.37 < |\eta| < 1.52$ is due to the barrel-endcap transition in the calorimeter. Also, the drop at high $|\eta|$ is due to the presence of more detector material in this region.

The efficiency as measured in data is compared to simulated $Z \to ee$ events. Tag and probe electrons are defined as they are in data, but in the case of MC, the numerator set is the set of probe electrons that fall within $\Delta R < 0.2$ of a true electron in the simulation. The resulting efficiencies from MC are compared to those of data, and the ratio of data to MC is computed as a correction which is applied to MC. These "scale factors" are applied to each electron identified in the event as a function of E_{T} and η.

In order to assess the uncertainties on the efficiency measurement—and the asso-

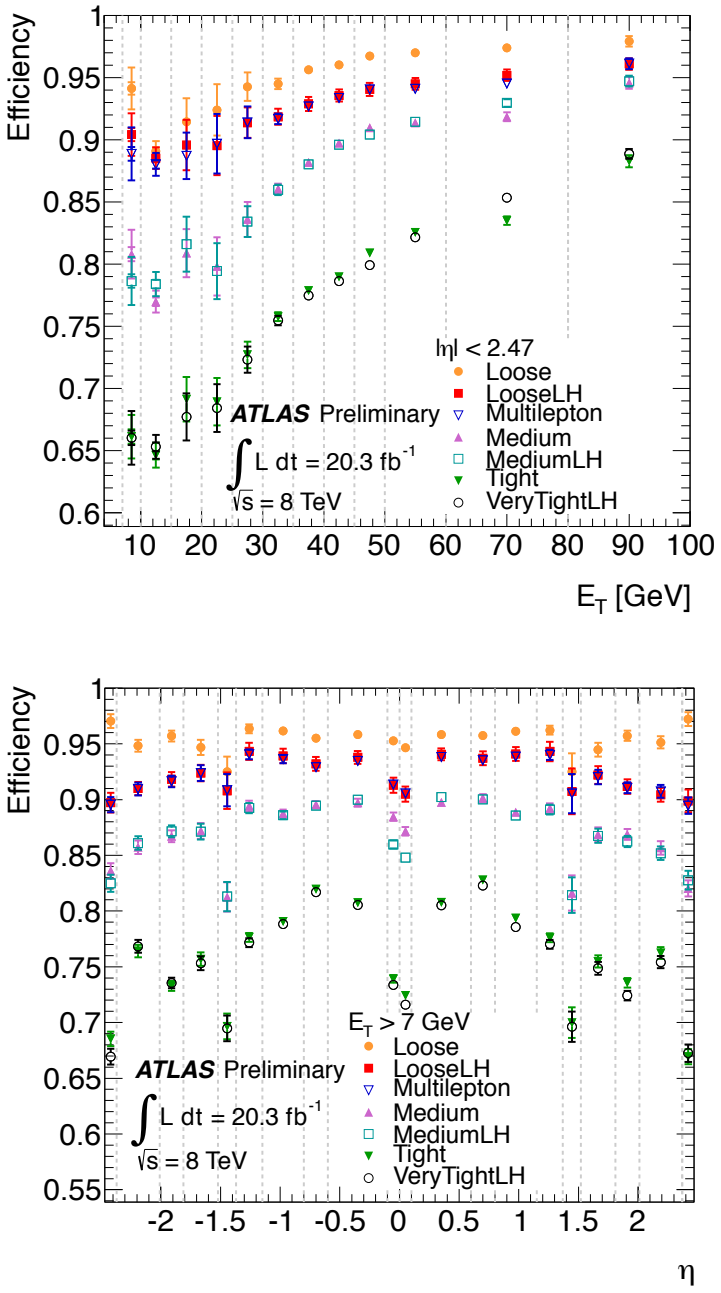


FIGURE 4.5: Electron identification efficiencies for different cut-based and likelihood categories, binned in electron E_{T} (a) and η (b) [8].

ciated scale factors—several alternative measurements are carried out. In addition to using the Z peak to select denominator electrons, a calorimeter isolation variable is

used to select denominator electrons and normalize the background templates. Moreover, the width of the Z mass window is varied, and the definition of the tag electron is varied. In total, there are 90 systematic variations on the efficiency measurement, with the central efficiency set to the average and the systematic uncertainty the RMS.

4.4 Muons

4.4.1 Muon reconstruction

Muons are reconstructed with information from the inner detector (ID), the muon spectrometer (MS), and to a lesser extent, the calorimeters. In the MS, track segments are first identified in each layer, and then combined in a track fit. Muon candidate tracks are reconstructed in the ID according to the procedure in Section 4.1.

4.4.2 Muon identification

There are four muon identification schemes that take advantage of the information from the different detector subsystems [9]. Stand-alone (SA) muons are reconstructed with tracks from the MS only. The track of the muon candidate is extrapolated back to the distance of closest approach with the beam line, accounting for energy lost in the calorimeter. These muons are required to have hits in at least two MS layers. Because the MS extends to $|\eta| < 2.7$, SA muons recover acceptance for muons which fall beyond the ID tracking volume $|\eta| < 2.5$. For combined (CB) muons, muon tracks are independently reconstructed in both the ID and the MS and then combined. Segment-tagged (ST) muons are those for which there exists an ID track which, when extrapolated to the MS, corresponds to a track segment in a MDT or CSC layer. This type recovers acceptance associated with muons that only traverse a single MS layer due to the limited coverage of the MS or because the muon

p_{T} is low. Finally, for calorimeter tagged (CaloTag) muons, an ID track is associated with an energy deposit in the calorimeter that is consistent with a minimum-ionizing particle.

For the analysis presented in this thesis, only CB muons are used. The combination of the ID and MS tracks is done by statistically combining the two tracks with their respective parameters and covariance matrices. Starting with a set of ID tracks and a set of MS tracks, MS-ID track pairs are matched in η and ϕ, and the MS-ID track pair with the smallest combined χ^2 is retained as a CB muon. The constituent tracks are then removed from the set and the next iteration proceeds until there are no more candidate tracks. To suppress background, the ID tracks that are used require at least 1 pixel hit, 5 SCT hits, and at least 9 TRT hits in the region with full TRT coverage. Moreover, tracks can only have a maximum of 2 active pixel or SCT sensors without hits along the track trajectory.

As with electrons, it is crucial to know the reconstruction and identification efficiencies for muons of various transverse momentum and pseudorapidity. The efficiencies are measured with the tag-and-probe method on $Z \to \mu\mu$ events in which two oppositely charged and isolated muons with $p_{\mathrm{T}} > 25(10)$ GeV have dilepton invariant mass that falls within 10 GeV of m_Z. Tag muons are required to satisfy CB requirements, and the efficiency of identifying a probe CB muon is measured as a function of p_{T} and η, as shown in Figure 4.6. The efficiency for CB muons is between 95% and 100% across most of the η range, although it falls to zero at $|\eta| < 0.1$, due to a gap in the MS coverage. In general, the efficiency as measured in data agrees quite well with that of simulation, with the exception of the barrel-endcap transition region ($0.9 < \eta < 1.3$), where there are imperfections in the modeling of the detector. The efficiency measurements are compatible for data and simulation across the p_{T} spectrum as well (Figure 4.6(b)).

As for electrons, the ratio of the efficiency measured in data to simulation is com-

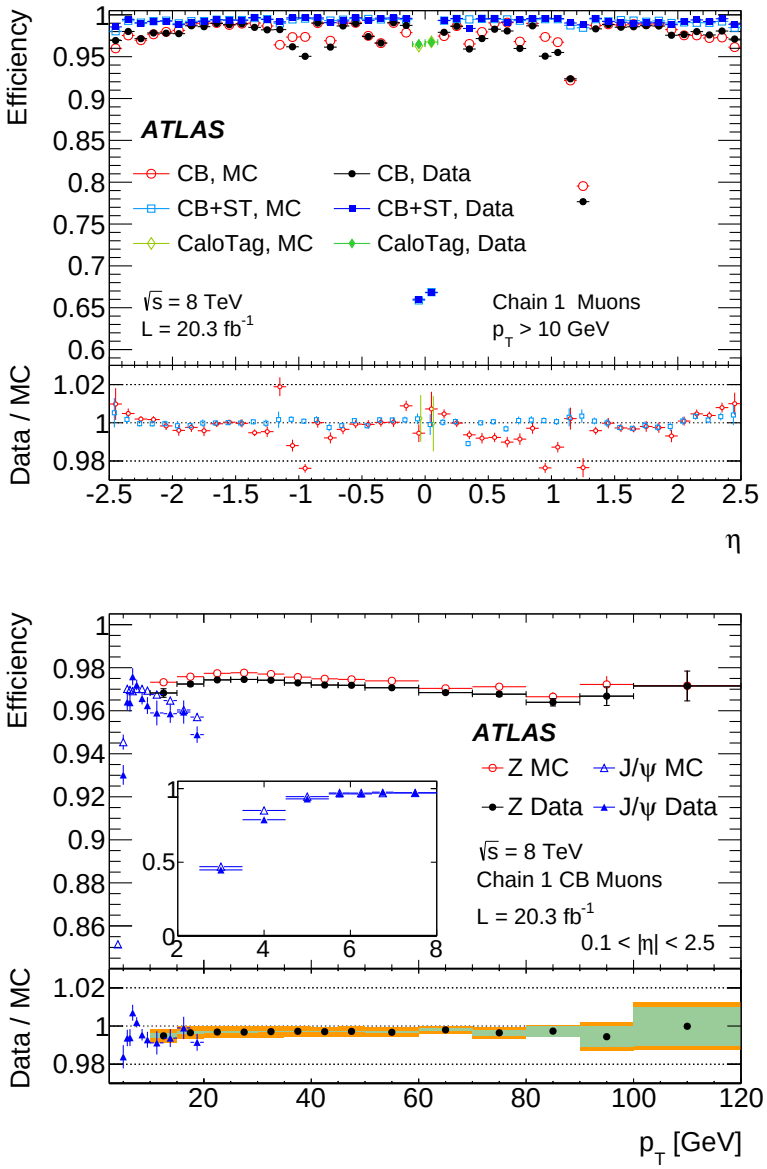


FIGURE 4.6: Muon identification efficiency (a) as a function of η for the different muon types with a lower p_{T} threshold of 10 GeV, and (b) as a function of p_{T} for CB muons with $0.1 < |\eta| < 2.5$ [9].

puted. In order to recover the correct reconstruction and identification efficiencies in simulation, these scale factors are applied to all predictions from simulation in two-dimensional $\eta - \phi$ bins. In addition to efficiency corrections, the muon momentum scale and resolution is corrected in simulation, with corrections that are determined in a likelihood fit to data. Scale corrections are less than 0.1% in most η regions, and as high as 0.4% in the region $1.25 < \eta < 1.5$. Muon resolution is corrected in MC by applying a "smearing" term to the reconstructed p_{T}. This smearing correction is less than 10% for the ID and less than 15% for the MS. Systematic uncertainties on these corrections are discussed in an analysis-specific context in Section 9.2.

4.5 Jets

A jet is loosely defined as a collection of collimated, high p_{T} hadrons arising as a result of the fragmentation and hadronization of a parton or partons. Because partons do not exist as observable particles, the constituent particles in jets—which manifest as tracks and calorimeter clusters—must be combined in a way that best reflects the parton 4-momentum. Jet algorithms are designed to do this, while avoiding complications due to perturbative QCD (pQCD). In pQCD, the probability for a quark to radiate a gluon diverges in the soft ($E_{\mathrm{gluon}} \to 0$) or collinear ($\Delta R_{\mathrm{quark,gluon}} \to 0$) limits. Jet constituents are formed by a series of gluon emissions with subsequent quark splittings and are therefore difficult to model with pQCD. To ensure that the processes which form jets can be predicted with pQCD calculations, jet constituents are iteratively recombined until some condition that protects against infrared divergences is satisfied

4.5.1 Anti-k_t jets

In ATLAS, jets are built with the anti-k_{T} algorithm [59]. The jet constituents are the topological clusters defined in Section 4.2.2. For each event, starting with the

full set of clusters, the quantity

$$d_{ij} = \min\left(p_{ii}^{-2}, p_{ij}^{-2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{4.6}$$

is computed for each cluster pair. For the pair that yields the minimum d_{ij}, if $d_{ij} < p_{ii}^{-2}$, the clusters are combined and the next iteration begins with the new set of clusters. If the above condition is not true, cluster i is considered a "jet" and is removed from the set of clusters. This process is repeated until no more clusters exist. The parameter R is called the distance measure. If cluster i has no clusters within R, then $\frac{\Delta R_{ij}^2}{R^2} > 1$, making p_{ii}^{-2} smaller than d_{ij}, and the cluster is added to the list of jets. Therefore, the distance measure controls the jet area and protects against collinear divergences. To protect against soft divergences, a minimum energy requirement is placed on the clusters. The form of d_{ij} in Equation 4.6 is chosen to favor jets in which soft clusters are grown around a hard "seed" cluster, due to the $\min\left(p_{ii}^{-2}, p_{ij}^{-2}\right)$ term. In ATLAS, both $R = 0.4$ and $R = 0.6$ jets are constructed with the implementation in FastJet [60, 61].

4.5.2 Jet energy calibration

Before combining calorimeter clusters with the anti-k_{T} algorithm, the clusters are calibrated [11]. For the initial calibration, the cluster energy is corrected to produce an accurate energy measurement for a particle that showers electromagnetically in the calorimeter. A secondary calibration, called local cell signal weighting (LCW) [62], aims to accurately measure hadronic particles in the calorimeter. Using energy density and shower depth distributions, clusters are classified as EM or hadronic, and then corrections are applied to clusters to account for partial calorimeter response for hadrons, signal loss due to the noise thresholds, and energy lost in inactive regions. These corrections are obtained from simulations of interactions with the calorimeter

of single charged and neutral pions.

With jets built from calibrated clusters (EM or LCW), additional corrections are applied on the jets themselves. The jet response is defined as

$$\mathcal{R}_{\text{jet}} = \frac{E_{\text{jet}}^{\text{calo}}}{E_{\text{jet}}^{\text{truth}}} \tag{4.7}$$

where $E_{\text{jet}}^{\text{calo}}$ is the jet energy as measured by the calorimeter and $E_{\text{jet}}^{\text{truth}}$ is the jet energy obtained from running the anti-k_{T} algorithm on the truth-level particles associated with the jet. Jet energy scale (JES) corrections are applied to the numerator, with the goal of restoring this ratio to one. Four corrections are applied [11]:

(1) Energy offset due to in-time and out-of-time pile-up.

(2) Correction associated with shifting jet direction to be consistent with primary vertex and not the global center-of-mass of the detector.

(3) Scale correction from MC simulations in bins of jet p_{T} and η.

(4) Data-derived *in situ* corrections to account for differences in data and MC.

A description of the first two corrections can be found in [11]. Correction (3) is determined from an inclusive, isolated jet sample from simulation. The correction itself is just the inverse of $\mathcal{R}_{\text{jet}}$ from Equation 4.7. Because this correction relies on simulation to model prompt jets, the underlying event, and pile-up, data-driven techniques are used for additional corrections. In general, these techniques exploit the p_{T} balance between a well-measured reference object, such as a lepton or a photon, with a jet. Assuming that the jet energy scale is perfectly calibrated, that there are not additional prompt particles in the event, and that there is not additional radiation in the event, the p_{T} of the reference object should match that of the jet. In practice, this is not the case, and the deviation from unity is the scale correction, with the other effects accounted for with systematic uncertainties. For jets which fall in the p_{T} range [20 GeV,200 GeV], the reference object is a Z boson which has been reconstructed

from electrons, while for higher p_{T} jets in the range [100 GeV,800 GeV] the reference object is a prompt photon. Finally, for p_{T} that falls in the range [200 GeV,1200 GeV], multi-jet events in which a single high p_{T} jet is recoiling against several lower p_{T} jets are used. In this measurement, the low p_{T} jets, which have been calibrated with Z or γ momentum balance, are the reference objects. The above three momentum balance techniques require the jets to be central in rapidity ($|\eta_{\mathrm{jet}}| < 1.2$). To evaluate the dependence on η_{jet}, an η intercalibration factor (c) is measured from dijet events, in which there exists, with high probability, a jet that is subsequently calibrated with one of the above techniques, i.e. a central jet. The factor c is a measure of the momentum balance between the well-measured reference jet and the probe jet: $c = p_{\mathrm{T}}^{\mathrm{ref}}/p_{\mathrm{T}}^{\mathrm{probe}}$. This correction, which is applied after (3) but before (4), is binned in jet p_{T} and η, and in practice, the data-to-MC ratio is applied to the predictions from simulation.

The *in situ* JES corrections for central jets are shown in Figure 4.7 for each of the three momentum balance techniques. In the figure, the data-to-MC ratio of the jet response (Equation 4.7) is shown in bins of jet p_{T}. The three techniques are statistically combined and the result is shown as a solid black line, which is consistent with unity across the momentum range. The largest corrections, observed at high p_{T}, are at the level of 2%, and are also consistent with unity within the $\pm 1\sigma$ systematic and statistical uncertainties.

The techniques for evaluating the *in situ* JES corrections rely on assumptions that are only approximately true. For example, the p_{T} balance between the reference object and the jet can be spoiled by radiation that falls outside of the jet or from underlying event contributions within the jet. Uncertainties related to such assumptions are quantified by varying the relevant selection and testing the impact on the resulting correction factor. In addition, there are uncertainties associated with the selection of the reference object, MC generator modeling, and the resolu-

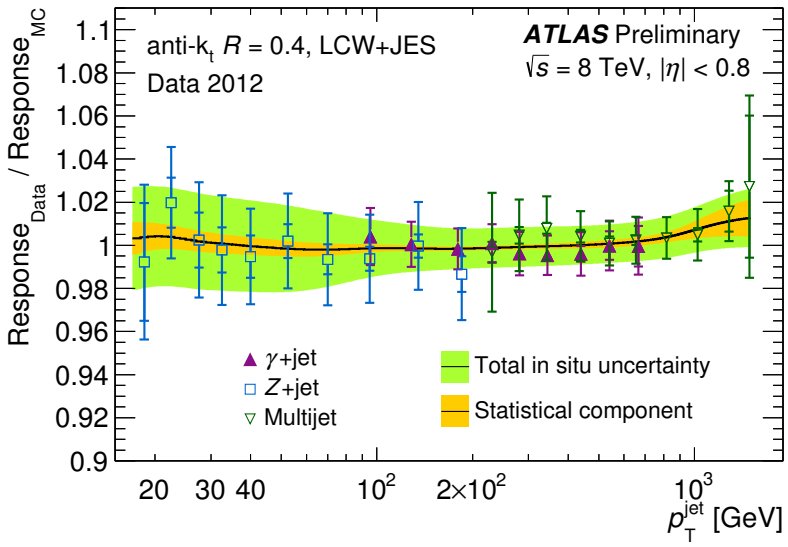
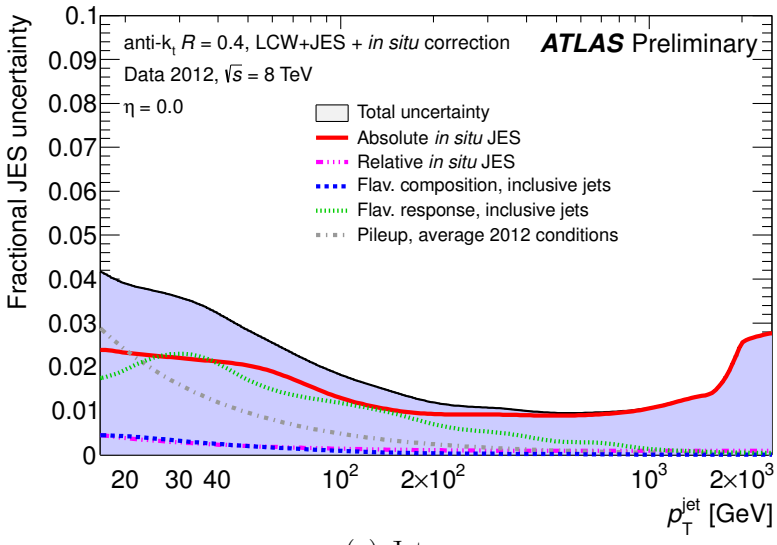


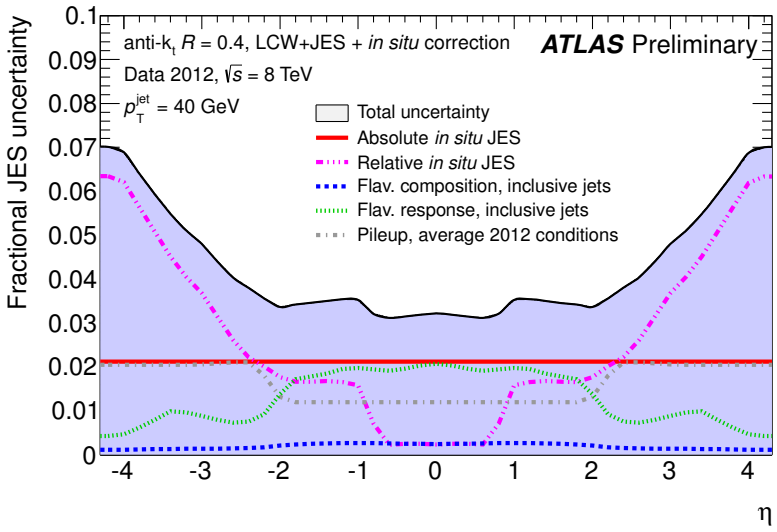
FIGURE 4.7: Ratio of the data-derived (*in situ*) jet response to that derived from simulation as a function of the jet p_{T} for anti-k_T $R = 0.4$ jets calibrated with local cell weighting [10].

tion modeling of jets. The resulting uncertainties on the JES correction are shown in red in Figure 4.8. The η inter-calibration correction uncertainty is dominated by the uncertainty due to a large difference observed between PYTHIA and HERWIG. This is the dominant JES uncertainty at high $|\eta|$, shown in Figure 4.8(b). Additionally, there is an uncertainty to account for the fact that the *in situ* techniques derive the correction with a sample of jets with a quark to gluon ratio that may differ from a more inclusive sample. Similarly, there is an uncertainty to account for differences in the detector response between quark and gluon jets.

In situ techniques are also used to measure the jet energy resolution (JER) [12]. The relative momentum resolution $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$, which is equivalent to the relative energy resolution at fixed rapidity, is measured with two independent methods. In the di-jet balance method, the p_{T} asymmetry between the two leading jets in di-jet events is measured. The inherent assumption is that a deviation from perfect balance is due to resolution effects, and therefore the resolution of the asymmetry



(a) Jet p_T



(b) Jet η

FIGURE 4.8: Fractional jet energy scale uncertainties for anti-k_T $R = 0.4$ jets calibrated with local cell weighting. (a) shows the uncertainty as a function of jet p_T at $\eta = 0$, and (b) shows the uncertainty as a function of jet η for jets with $p_T = 40$ GeV [11].

can be propagated to $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$. To account for the breakdown in this assumption due to the presence of additional QCD radiation in the event, a correction is applied to the measured resolution. In the second method, called the bi-sector method, the di-jet p_{T} vector is projected onto a new coordinate system in the transverse plane (x', y'), which is defined such that the y' axis coincides with the line that bi-sects the angle between the leading jets $\Delta_{\theta 1,2}$. Again, the assumption is that the di-jet p_{T}, and therefore the projections of the p_{T} into the new coordinate system, will deviate from zero due to resolution effects. The resolutions associated with the two projections, $\sigma_{y'}$ and $\sigma_{x'}$, are extracted by fitting Gaussians to the projected p_{T} distributions, then used to compute $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$.

The relative resolutions, as measured by the above two methods, are shown in Figure 4.9 as a function of the average p_{T} of the two leading jets. Though these are $R = 0.6$ anti-k_{T} jets calibrated with the EM scheme, the relative resolutions for $R = 0.4$ LCW jets are approximately the same, with $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} \sim 20\%$ at $p_{\mathrm{T}}^{\mathrm{jet}} = 20$ GeV and decreasing to 5% at high jet p_{T}. The dependence of $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$ on $p_{\mathrm{T}}^{\mathrm{jet}}$ follows the expected functional form of the energy resolution for a calorimeter, given by

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C. \tag{4.8}$$

N denotes the resolution contribution from noise due to electronics and pile-up. Since this term is independent of p_{T}, it becomes dominant for jet p_{T} below 20 GeV. The S term accounts for statistical fluctuations and is a significant component of the resolution for intermediate jet p_{T}. Finally, the C term captures fluctuations which are linear in jet p_{T}, including jet energy lost in passive detector material and non-uniformity of detector response. This term dominates for jet $p_{\mathrm{T}} > 400$ GeV.

JER uncertainties are evaluated by computing the variation in the extracted resolution under the change in a particular source. Sources considered include pile-up,

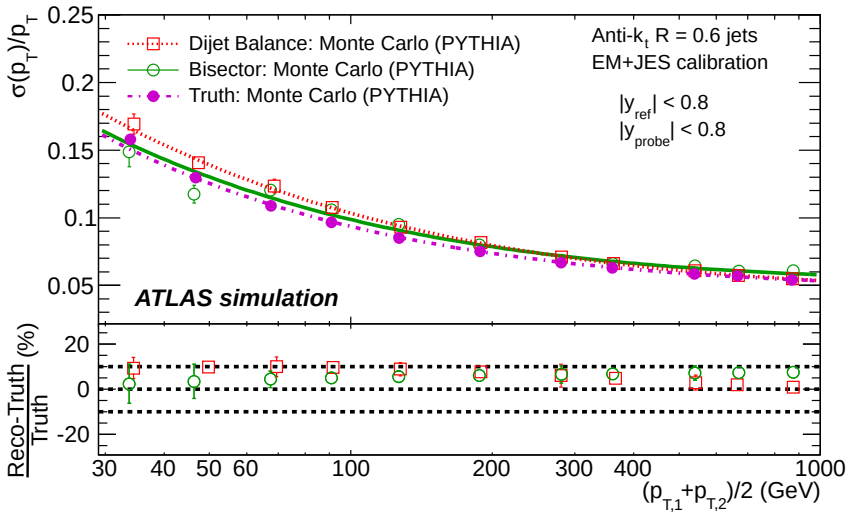


FIGURE 4.9: Relative jet resolution as a function of jet p_{T}as measured with the di-jet balance method (red), the bi-sector method (green), and for truth level jets (purple). Figure is from [12].

JES, and MC modelling. There is also an uncertainty associated with the disagreement in the measured resolution and the truth resolution from MC simulation, which is the dominant uncertainty with respect to the other sources. For $p_{\mathrm{T}} \sim 50$ GeV jets, the total uncertainty is on the order of 20%, while for $p_{\mathrm{T}} \sim 400$ GeV jets, the uncertainty falls in the range 11%-18%, depending on the $|\eta|$ bin.

4.5.3 Jet quality requirements

To distinguish high p_{T} jets coming from the hard scatter from fake jets, quality cuts are applied on the jets. The dominant background jets arise from beam gas events in which the proton beam interacts with residual gas in the beam pipe, beam halo events in which the beam interacts with accelerator material far from the detector, cosmic ray muons which are coincident with a collision event, and calorimeter noise. As in electron identification, several jet quality categories are defined, each achieving different levels of background rejection. In order of increasing rejection—and decreasing efficiency—the categories are LOOSER, LOOSE, MEDIUM, and TIGHT. Jets

are distinguished from background by placing requirements on the fraction of jet cells from certain detector regions or with poor signal shape quality, the timing of the jet with respect to the trigger time-stamp, and the fraction of jet energy closest to the interaction point [63]. The efficiencies of the resulting jets are measured with a tag-and-probe technique, resulting in a measured efficiency of greater than 99% for LOOSER jets and 96%-99% for LOOSE jets. At jet p_{T} values near 25 GeV , MEDIUM (TIGHT) jets have are selected at an efficiency of 96% (85%).

In addition to the cuts in the jet identification categories above, there is often a requirement on a quantity called the jet vertex fraction (JVF) [13]. Using reconstructed tracks, the JVF variable is computed for each jet as

$$\mathrm{JVF} = \frac{\sum_k p_{\mathrm{T}}(\mathrm{track}_k)}{\sum_j p_{\mathrm{T}}(\mathrm{track}_j)}, \tag{4.9}$$

where index k runs over the jet tracks matched the hard-scatter PV, and the index j runs over all jet tracks. The association between tracks and jets is carried out using a "ghost" association technique whereby tracks of infinitesimal p_{T} are included in the jet algorithm in addition to calorimeter clusters, and are assigned to the jet to which they are recombined. Because such ghost tracks are so soft, the resulting jet 4-momentum is unchanged. JVF can be interpreted as the fraction of energy in the jet from the HS: values close to zero indicate that the majority of the tracks come from vertices other than the HS vertex, while values close to one indicate that the jet is likely from the HS. Therefore, placing a cut on JVF efficiently suppresses in-time pileup. For jets without any associated tracks, which is always the case for jets that fall outside of the tracking volume ($|\eta| > 2.47$), JVF is set to -1.

The discriminating power of JVF is illustrated in Figure 4.10, where the JVF distributions for HS and pileup jets are shown for Z + jets simulation.

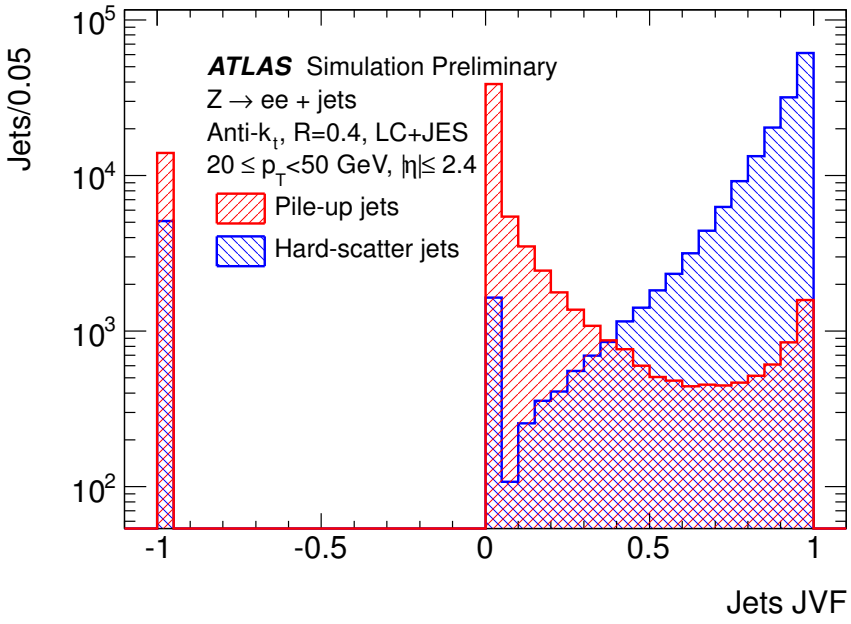


FIGURE 4.10: Distribution of JVF for HS (blue) and pileup (red) jets with 20 GeV $< p_{\mathrm{T}} <$ 50 GeV and $|\eta| < 2.5$ in simulated Z + jets events. [13]

4.6 b-Hadron Jets

Many interesting scattering processes have b-quarks in the final state. The hadronization of these quarks produces b-hadrons, which preferentially decay via the weak force to c-hadrons. These c-hadrons also subsequently decay weakly, and because the two decays occur at weak interaction lifetimes (10^{-12} s), the decay vertices are measurably displaced from the primary vertex. b-tagging algorithms exploit this fact by finding jet-associated tracks with secondary vertices.

The b-tagging algorithms in ATLAS are designed to efficiently find b-jets and minimize the rate at which light-flavor jets from u, d, and s quarks, as well as gluons, are tagged as b-jets. A multi-layered multivariate algorithm has been developed in order to synthesize information about track impact parameters and reconstructed secondary vertices [15]. As a first step, tracks are selected, requiring that they have seven silicon hits, two of which are pixel hits, and one of which is in the pixel b-

layer. Only tracks with $p_{\mathrm{T}} > 1$ GeV are considered. Moreover, the following cuts are applied to the impact parameter distributions: $|d_0| < 1$ mm and $|z_0| \sin\theta < 1.5$ mm. The resulting track collection is associated to jets by ΔR

matching, where the ΔR value varies with jet p_{T}. With the tracks defined for each jet, the tagging algorithm computes a likelihood ratio for two hypothesis: the b-jet hypothesis, and the light-flavor hypothesis. The simulation-derived PDFs for the two hypotheses are the impact parameter significances $d_0/\sigma(d_0)$ and $z_0/\sigma(z_0)$. In order to take advantage of correlations between the two, the PDFs are two-dimensional.

In addition to the impact-parameter-based algorithm above, a second algorithm uses the properties of the reconstructed secondary vertex to identify b-jets [15]. For this approach, the track requirements are looser in order to select background tracks from light-flavor hadrons or material interactions. Considering only tracks associated to a jet, vertices are reconstructed with an approach similar that of primary vertices (Section 4.1.3), resulting in a single reconstructed vertex. A two-hypothesis likelihood is computed with the following PDFs: the invariant mass of the tracks associated with the vertex, the ratio of the sum of energies of the vertex tracks to the sum of energies of all tracks in the jet, the number of two track vertices, and the ΔR between the jet and the line connecting the primary vertex to the secondary vertex. The output of the algorithm is a likelihood ratio. An additional algorithm which uses information about the reconstructed tracks has been shown to improve the tagging efficiency [64]. Under the assumption that the b and c hadron decays are collinear, a Kalman-filter fit is performed to reconstruct the line connecting the primary vertex, b-hadron decay vertex, and c-hadron decay vertex, as well as the three-dimensional positions of these vertices. Then the same discriminating quantities as above are computed for the reconstructed vertices. Additionally, the flight length significance $L/\sigma(L)$ is considered. These distributions are used to train an artificial neural network (NN) for which b-jets are signal and light flavor jets are background. The NN output then

quantifies the probability that a jet is a b-jet.

Outputs from the three b-tagging approaches described above are used to train another NN classifier, thereby boosting the performance with respect to any one approach. The resulting algorithm, called "MV1", is the most widely used b-tagging algorithm in ATLAS. The performance of the MV1 tagger is summarized in Figure 4.11, showing the efficiency for tagging a true b-jet as a function of the light flavor rejection rate, or the "mis-tag" rate, for jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ in a simulated $t\bar{t}$ sample. At a b-jet efficiency of 70%, the mis-tag rate is less than 1%, and as the b-jet efficiency is increased to 85%, there is a corresponding increase in the mis-tag rate to 10%.

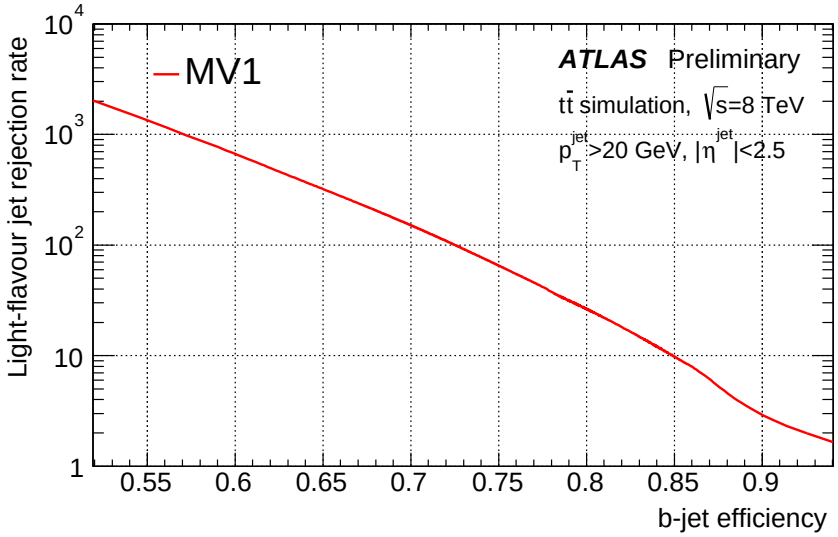


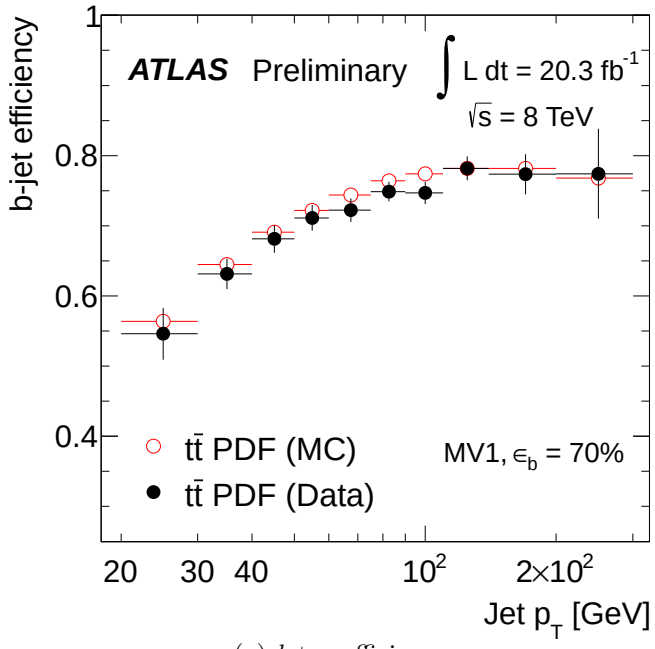
FIGURE 4.11: The efficiency for tagging a true b-jet plotted against the light flavor jet rejection factor, or equivalently the inverse of the mis-tag rate, for jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ in a simulated $t\bar{t}$ sample [14].

Because the distributions with which the NN is trained are from simulation and because tagging is sensitive to the condition of the detector, which is not always accurately captured in simulation, the b-tagging efficiencies are measured in actual collision data [15]. In order to obtain a b-hadron-rich sample, oppositely charged

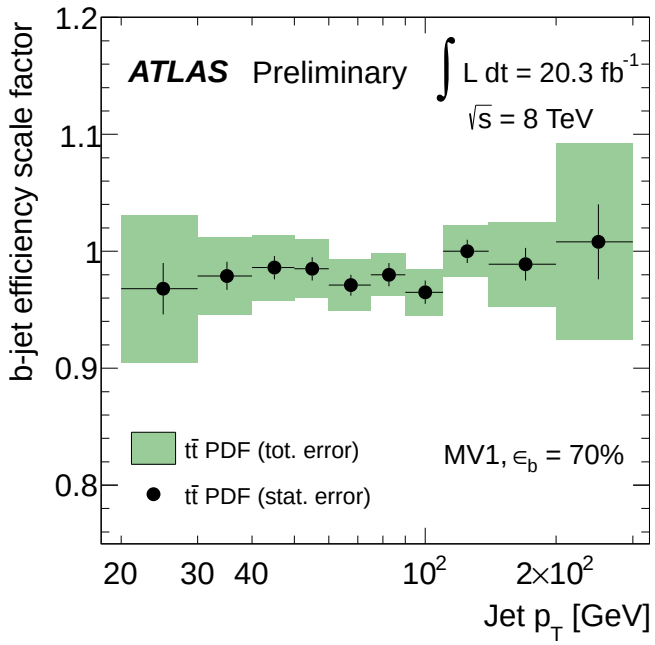
dilepton events, where the leptons satisfy tight selection criteria, are selected. Additionally, these events are required to have two or more jets with standard quality requirements, and the invariant mass of the two leptons is required to fall outside of the Z boson mass window. Such a selection favors $t\bar{t}$ events in which each top quark decays to a W boson and a b quark, with a subsequent leptonic W boson decay. The overall b-jet purity in this region, as detemined from simulation, is 67% for two jet events, and 52% for three jet events. Using a likelihood-based approach that properly treats event flavor correlations, the efficiency for b-tagging a jet is obtained in bins of p_{T} at several different NN output values. In Figure 4.12(a), the efficiency is shown for both data and simulation as a function of jet p_{T} for the NN output corresponding to an integrated efficiency of 70%. Efficiencies obtained from data agree well with those from $t\bar{t}$ simulation, also illustrated in Figure 4.12(b), where the scale factors are consistent with unity within the systematic uncertainties. These systematic uncertainties are assessed by varying the MC simulation model, re-running the likelihood fit, and re-computing the SFs. The deviation from the central SF is computed and the differences are added in quadrature for the total uncertainty. Theoretical systematic variations for $t\bar{t}$ MC are evaluated for the following sources: hadronization model, hard-scatter matrix element, top p_{T} weighting, and the amount of radiation in the parton shower. Similarly, for the background processes, Z/DY and diboson, the hard-scatter matrix element has been varied, as have the normalizations of these backgrounds. Finally, detector uncertainties related to pile-up, lepton selection efficiencies, jet energy scale and resolution, JVF, and the mis-tag rate are also considered.

4.7 Missing Transverse Energy

Many physics analyses seek final states with particles that are weakly interacting and just pass through the detector undetected. Because the initial momenta of



(a) b-tag efficiency



(b) b-tag scale factor

FIGURE 4.12: b-tagging efficiencies and scale factors as a function of jet p_{T} measured with $t\bar{t}$ samples in both data and simulation [15].

the incoming partons is approximately zero in the transverse plane, the presence of these undetected particles can be inferred by measuring the missing momentum in this direction. Mathematically, this is seen by requiring momentum conservation

$$\begin{aligned} \boldsymbol{p}_{\mathrm{T}}^{\mathrm{incoming}} = 0 &= \boldsymbol{p}_{\mathrm{T}}^{\mathrm{outgoing}} \\ &= \boldsymbol{p}_{\mathrm{T}}^{\mathrm{visible}} + \boldsymbol{p}_{\mathrm{T}}^{\mathrm{missing}}, \end{aligned} \tag{4.10}$$

which implies that $\boldsymbol{p}_{\mathrm{T}}^{\mathrm{missing}}$ is obtained by measuring $-\boldsymbol{p}_{\mathrm{T}}^{\mathrm{visible}}$. Since the longitudinal momenta of the incoming partons is not known, this quantity can not be inferred through momentum conservation, making it impossible to reconstruct the 4-momentum of the undetected particles. In this thesis, the missing transverse momentum of the visible particles $\boldsymbol{p}_{\mathrm{T}}^{\mathrm{visible}}$ is reconstructed using either calorimeter deposits or tracks. The former is denoted $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ while the latter is denoted $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$. Generically, missing transverse energy will be referred to as $E_{\mathrm{T}}^{\mathrm{miss}}$.

4.7.1 Calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ Reconstruction

Calorimeter-based missing transverse energy uses information from the calorimeter as well as the tracking systems for muons, and can therefore be decomposed into two terms [65]. The calorimeter term is further decomposed into the high-p_{T} physics objects which are reconstructed from calorimeter information, as well as a soft term for the low energy deposits not associated to a high-p_{T} object:

$$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss,CALO}} = \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss},e} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss},\gamma} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss},\tau} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss,jets}} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss,soft}} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss},\mu} \tag{4.11}$$

To avoid double-counting energy in the calorimeter, selection of the objects going into Equation 4.11 is done sequentially. First, the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss},e}$ term is defined using reconstructed electrons (Section 4.3) with $p_{\mathrm{T}} > 10$ GeV. Then photons with $p_{\mathrm{T}} >$

10 GeV and without any overlap with reconstructed electrons are selected. High-p_{T} jets ($p_{\mathrm{T}} > 20$ GeV) are reconstructed with topological clusters using the anti-k_{T} algorithm with distance measure $R = 0.4$ (Section 4.5). These clusters are required to be isolated from those associated to electrons or photons. To reject jets from pile-up, jets with $|\eta| < 2.4$ and 20 GeV $< p_{\mathrm{T}} <$ 50 GeV are required to have $|\mathrm{JVF}| > 0$ (Section 4.5.3). The final calorimeter term represents energy in the event that is not associated with high-p_{T} objects. This soft term is built from LCW-calibrated topological clusters which do not have an explicit E_{T} requirement but are robust against electronic and pileup noise (Section 4.2). Given that it is not possible to reliably match a cluster to a primary vertex, the soft term is highly sensitive to in-time pileup, as shown in Figure 4.13. In an inclusive $Z \to \mu\mu$ sample, the average soft term at $N_{\mathrm{PV}} = 1$ is around 10 GeV and increases to 40 GeV at $N_{\mathrm{PV}} = 25$. This term significantly degrades both the magnitude and direction resolutions for $E_{\mathrm{T}}^{\mathrm{miss}}$. Various corrections which help to mitigate the dependence on pileup have been used in ATLAS [16], but for the analysis presented in this thesis, the baseline calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ definition without corrections is used.

For the muon term, combined muons (Section 4.4) with $p_{\mathrm{T}} > 6$ GeV are used where there is both ID and MS coverage ($0.1 < |\eta| < 2.5$). Because muons usually leave energy clusters in the calorimeter, and the combined muon fit accounts for this loss, the combined muon energy is double-counted. This is corrected by subtracting the parameterized energy loss in the calorimeter from the fitted muon momentum. Outside of the ID, standalone muons are used, while in regions with limited MS coverage ($|\eta| < 0.1$), ID track muons are used.

4.7.2 Track $E_{\mathrm{T}}^{\mathrm{miss}}$ Reconstruction

As mentioned in the previous section, $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ is sensitive to in-time pileup, degrading its resolution for large N_{PV} events. Track $E_{\mathrm{T}}^{\mathrm{miss}}$ seeks to suppress pileup

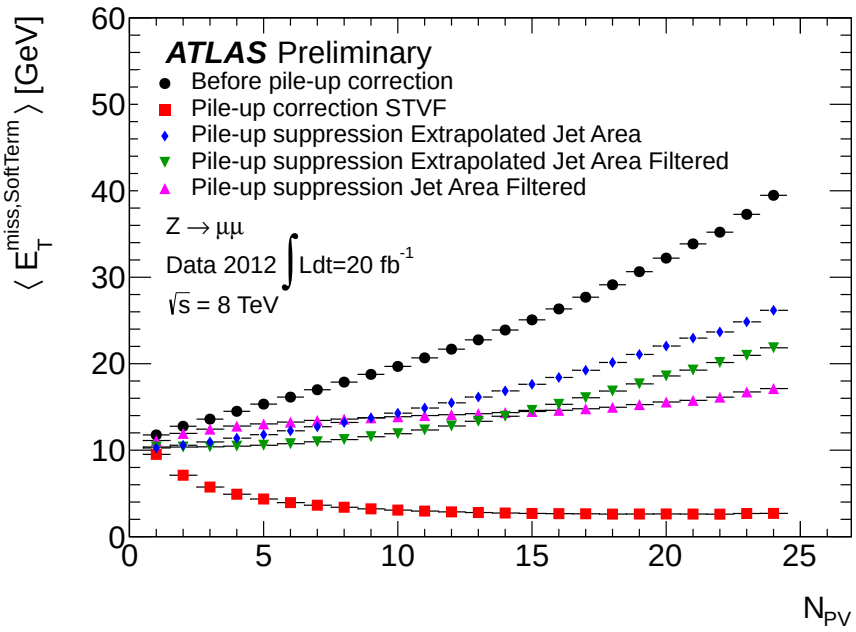


FIGURE 4.13: The average calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term $\left\langle E_{\mathrm{T}}^{\mathrm{miss,soft}} \right\rangle$ (shown in black) as a function of the number of primary vertices for an inclusive $Z \to \mu\mu$ sample in $\sqrt{s} = 8$ TeV data [16].

dependence by replacing the clusters defining the soft term with tracks matched to the primary vertex of the event (Section 4.1). These tracks are required to have $p_{\mathrm{T}} > 500$ MeV and $|\eta| < 2.5$, with additional quality requirements: $|d_0| < 1.5$ mm, $|z_0 \sin(\theta)| < 1.5$ mm, $N_{\mathrm{hits}}^{\mathrm{pixel}} \geq 1$, and $N_{\mathrm{hits}}^{\mathrm{SCT}} \geq 6$. If a track fails these quality criteria, but passes requirements associated with electrons and muons, then the associated lepton p_{T} replaces the track. The lepton requirements are similar to those in the $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ definition, with some important exceptions which are discussed in Chapter 7. Moreover, tracks that fall within a cone of $R = 0.4$ of a jet are not included in $E_{\mathrm{T}}^{\mathrm{miss}}$; instead the jet itself is used. The rationale is that if tracks are used to define $E_{\mathrm{T}}^{\mathrm{miss}}$ instead of reconstructed jets, neutral tracks associated with the jets are missed, thereby degrading the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. The specific definition of a jet is analysis-dependent, and will therefore be described in Chapter 7.

5

Analysis Overview

The primary aim of the analysis presented in this thesis is to measure Higgs boson production via vector boson fusion and to compare such a signal to the SM predictions. Previous Higgs measurements, including the ones in which the Higgs was discovered, have combined all Higgs production mechanisms (see Figure 2.3) into a single signal controlled by the the same strength. This analysis, by contrast, considers only Higgs production through a VVH coupling, VBF and VH, to be signal. Gluon-gluon fusion production is assumed to have been "discovered" and is therefore included in the SM background prediction.

This VBF analysis is performed in the W^+W^- Higgs decay channel, which, at the current best-fit Higgs mass 125.4 GeV has the second largest branching fraction of 0.22 (see Table 2.1). Though the W boson predominantly decays to final state hadrons, this signature is not visible over the enormous QCD backgrounds in pp scattering. Instead, both W bosons are required to decay leptonically: $W \to \ell\nu$. This final state is more sensitive than the semi-leptonic channel, $WW \to \ell\nu qq$, due to increased background rejection. In addition to the Higgs decay products, there are two final state quarks associated with VBF Higgs production. Therefore, VBF

Higgs production in the $\ell\nu\ell\nu$ decay channel manifests in the ATLAS detector as two, oppositely-charged high-p_{T} leptons, missing transverse energy from the neutrinos, and two forward jets initiated by the VBF quarks.

A preliminary Run-1 VBF measurement in the $WW^{(*)} \to \ell\nu\ell\nu$ channel has been published by ATLAS [66]. The observed significance of the data assuming that the background-only hypothesis is true has been measured to be 2.5σ, corresponding to a signal strength that is consistent with the SM expectation, 1.66 ± 0.79. The analysis presented in this thesis improves on the previous analysis in many respects, most notably the use of a multivariate algorithm called a boosted decision tree (BDT). A detailed description of this algorithm can be found in Appendix A.

The following chapters examine the BDT-based VBF analysis. Chapter 6 briefly outlines the ATLAS collision datasets and simulation samples used in the analysis. In Chapter 7, the selection of the final state physics objects and VBF-like events is described. This chapter also details input distributions for the BDT algorithm (Section 7.3) and the validation of the BDT (Section 7.4). The background estimation methods are outlined in Chapter 8, and the associated systematic uncertainties are discussed in Chapter 9. Finally, the results are reported in Chapters 10 and 11, with Chapter 11 focusing on the statistical formalism and results. Chapter 10 includes a summary of the analysis of a smaller dataset collected at $\sqrt{s} = 7$ TeV in 2011 that is analyzed along with the 8 TeV dataset in the final statistical results (Section 10.2), and also concludes with a discussion of a cut-based analysis that is performed as a cross-check of the BDT analysis (Section 10.3).

6

Collision Data and Simulation

The VBF Higgs boson measurements presented in this thesis are performed on pp collision data recorded in the ATLAS detector in Run-1 of the LHC. This short chapter describes the actual collision data and the samples used to simulate the collision data for signal and background processes.

6.1 Data collected in ATLAS

The analysis is performed on pp collision data recorded by the ATLAS detector (Chapter 3) over the course of Run-1 of the LHC in 2011 and 2012. In this run, the center-of-mass collision energy ($\sqrt{s}$) was set below the design value of 14 TeV for technical reasons. In total, 20.3 fb^{-1} of data were collected at $\sqrt{s} = 8$ TeV, and a smaller dataset of 4.5 fb^{-1} was collected at $\sqrt{s} = 7$ TeV. Figure 6.1 displays as a function of time the integrated luminosity delivered to ATLAS (green) and recorded by ATLAS (yellow). If the detector is fully operational in a given block of time (approximately 5 min), the events recorded in that block are qualified for physics analysis. The integrated luminosity of such events is shown in blue in Figure 6.1.

During Run-1, each beam consisted of approximately 1400 proton bunches, cor-

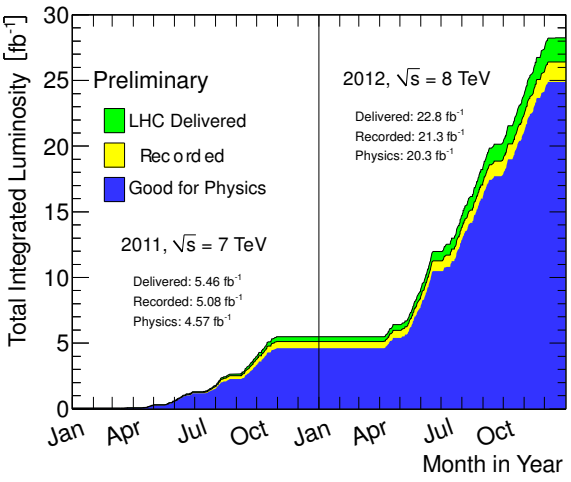


FIGURE 6.1: The total integrated luminosity delivered to ATLAS (green), recorded by ATLAS (yellow), and considered of sufficient quality for physics analysis (blue) in Run-1 of ATLAS. The horizontal axis is split by year, with 2011 on the left and 2012 on the right.

responding to a bunch spacing of 50 ns. The instantaneous luminosity increased over the course of the run, reaching an average value of around 7×10^{33} cm^{-2}s^{-1}, nearly the LHC design luminosity of 1×10^{34} cm^{-2}s^{-1} [42]. As a result, there were overlapping pp collision events, known as pile-up. Pile-up is categorized into two types: in-time and out-of-time. In-time pile-up refers to proton collisions that occur in the same bunch crossing as the collision of interest. It degrades the reconstruction of physics objects that rely on calorimeter signals, such as jets and missing transverse energy, since it is impossible to associate a calorimeter deposit to a primary collision vertex. In-time pile-up is quantified by the mean number of collisions per bunch crossing ($\langle\mu\rangle$). The distributions of $\langle\mu\rangle$ for both $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV datasets are shown in Figure 6.2. For the 8 TeV dataset, there were, on average, approximately 20 collisions per bunch crossing. Out-of-time pile-up refers to interactions in adjacent bunches that result in detector signals coincident with the colliding bunch. It also primarily impacts calorimeter-derived observables, since the response

time of the calorimeter is longer than the bunch spacing.

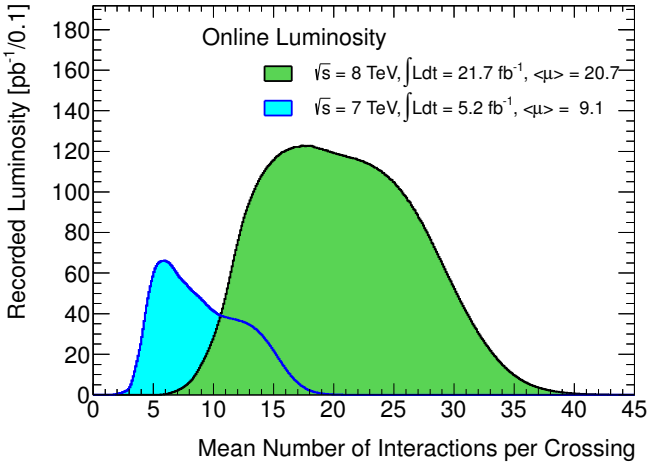


FIGURE 6.2: Mean number of collisions per bunch crossing for both 7 TeV and 8 TeV datasets collected in Run-1 of the LHC. The area under the distributions is the total integrated luminosity for each dataset.

To select interesting collision events, the Run-1 datasets are filtered with a three level trigger system (see Section 3.5). For the 7 TeV data, only single lepton triggers are required. The lower p_{T} threshold for electron (muon) candidates is 20 (18) GeV. The trigger requirements are tightened for data integrated later in this dataset, since the instantaneous luminosity, and hence the pile-up, increases.

For the 8 TeV dataset, trigger requirements were tightened due to increased pile-up. At Level-1, the electron (muon) p_{T} threshold is 18 (15) GeV, and at the event-filter level, the threshold is 24 GeV for both electrons and muons, and they are required to be isolated. Alternatively, the electron (muon) can pass an event-filter level trigger with a 60 (36) GeV threshold and no isolation requirement. For electrons, this second higher threshold trigger is only considered if the electron passes a 30 GeV threshold at Level-1. The above single lepton triggers are logically OR-ed with dilepton triggers at both Level-1 and the event-filter level in order to increase acceptance.

6.2 MC Simulation Data

Monte Carlo simulations are relied on for the prediction of signal and background observables. The hard scattering process, subsequent showering and hadronization of final state quarks and gluons, and the response of the detector are all simulated with MC programs. The absolute predictions are then obtained by scaling the MC distributions by the factor

$$\text{norm} = \mathcal{L} \cdot \sigma / N_{\text{total}}, \tag{6.1}$$

where $\mathcal{L}$ is the integrated luminosity, σ is the theoretical cross section for the process, and N_{total} is the total number of events generated, or in the case of weighted MC, the sum of weights. Table 6.1 summarizes the MC programs used for each signal and background process, as well as the cross section to which the prediction is scaled. Each cross section includes the relevant branching fraction for the $\ell\nu\ell\nu$ final state integrated over all lepton combinations.

6.2.1 Higgs Boson Samples

The signal process, VBF Higgs production in the $WW^{(*)} \to \ell\nu\ell\nu$ decay channel, is modeled with POWHEG [67] interfaced to PYTHIA8 [68] for the parton shower. The cross section is computed at NLO in QCD and EW [69, 70, 71], with approximate NNLO QCD corrections [72]. The branching fraction for the WW decay for the VBF process and other Higgs processes in this analysis is derived from the HDECAY program [73]. The final $\sigma \cdot \text{Br}$ is 0.036 pb at $m_H = 125$ GeV.

Because WH and ZH Higgs production probes the same Higgs couplings as VBF, these processes are considered signal. The cross sections for these processes are calculated at NNLO in QCD [74, 75] with NLO EW radiative corrections [76]. At $m_H = 125$ GeV the resulting cross sections are $\sigma_{ZH} = 9.4$ fb and $\sigma_{ZH} = 15.9$ fb

Table 6.1: Signal and background process summary with MC generators used and process $\sigma \cdot$Br, where the branching ratio is integrated over all flavor channels resulting in two charged leptons.

Process	Generator	$\sigma \cdot$ Br(8 TeV) (pb)
VBF $H \to WW$	POWHEG + PYTHIA8	$36 \cdot 10^{-3}$
ggF $H \to WW$	POWHEG + PYTHIA8	0.435
WH/ZH $H \to WW$	PYTHIA8 (PYTHIA6)	$25 \cdot 10^{-3}$
$t\bar{t}$ dileptonic	POWHEG +PYTHIA6	26.6
tW/tb leptonic	POWHEG + PYTHIA6	4.17
tqb leptonic	ACERMC + PYTHIA6	28.4
QCD $WW + 2$ jets	SHERPA	0.568
EW $WW + 2$ jets	SHERPA	0.039
$gg \to WW$	GG2WW + HERWIG	0.20
inclusive W	ALPGEN + HERWIG	$37 \cdot 10^3$
inclusive $Z/\gamma^\star(m_{ll} \geq 10GeV)$	ALPGEN + HERWIG	$16.5 \cdot 10^3$
EW $Z/\gamma^\star$	SHERPA	5.36 (inc. t-ch)
$W(Z/\gamma^*)$	POWHEG + PYTHIA8	12.7
$W(Z/\gamma^*)(m_{(Z/\gamma^*)} < 7$ GeV)	SHERPA	12.2
$Z^{(*)}Z^{(*)} \to 4l(2l2\nu)$	POWHEG + PYTHIA8	0.73(0.50)
EW $WZ + 2$ jets	SHERPA	$13 \cdot 10^{-3}$
EW $ZZ + 2$ jets $(4l, ll\nu\nu)$	SHERPA	$73 \cdot 10^{-5}(12 \cdot 10^{-4})$
$W\gamma$	ALPGEN + HERWIG	369
$Z\gamma(p_T^\gamma > 7GeV)$	SHERPA	163

for the processes in which the W bosons from the Higgs decay leptonically.

The ggF+2j process is also simulated with POWHEG [77] interfaced with PYTHIA8. The cross section is computed at NNLO in QCD [78, 79, 80, 81, 82, 83], with NLO electroweak corrections [84, 85] and soft gluon resummations up to next-to-next-to-leading-log (NNLL) [86]. Due to known deficiencies in the Higgs p_{T} spectrum in POWHEG, the Higgs p_{T} is re-weighted to the NLO + NNLL prediction from HQT [87]. $\sigma \cdot$ Br is 0.435 pb, a factor of 12 larger than VBF.

6.2.2 Background Samples

Top background is comprised mainly of $t\bar{t}$, with smaller contributions from s and t channel single top and Wt (labeled collectively as ST). The hard scatter for $t\bar{t}$ is simulated in POWHEG [88] at NLO in QCD, while PYTHIA6 [89] is used for parton showering and hadronization. The LO pdf set CTEQ6L1 is used with PERUGIA 2011

as the underlying event (UE) tune. For ST, s channel and Wt use POWHEG [90, 91] + PYTHIA6 as well, while ACERMC [92] + PYTHIA6 is used to model t-channel. The same pdf set and UE tune is used for all top processes. The $t\bar{t}$ normalization is scaled to the cross section for pp collisions at $\sqrt{s} = 8$ TeV for a top quark mass of 172.5 GeV/c^2: $\sigma_{t\bar{t}} = 252.9^{+15.3}_{-16.3}$ pb. This value has been calculated at NNLO in QCD including a resummation of NNLL soft gluon terms [93, 94, 95, 96, 97, 98, 99]. The total cross section for $t\bar{t}$ has been measured in ATLAS to be 237.7 ± 11.3 pb in the leptonic decay channel with one electron and one muon in the final state [100], agreeing with the theory calculation within the QCD scale uncertainties. The s channel ST, t channel ST, and Wt normalizations are also scaled to the NNLO and NNLL cross sections of 5.61 ± 0.22 pb [101], $87.76^{+3.44}_{-1.91}$ pb [102], and 22.37 ± 1.52 pb [103], respectively. These computations are compatible with the respective measurements in ATLAS [104, 105].

Standard model production of a pair of W bosons is split into two classes, depending on how the final state quarks or gluons are produced. "QCD WW" jets come from QCD vertices, while "EW WW" jets are produced via electroweak couplings. The QCD WW prediction is from SHERPA 1.4.1 [106], which is used to simulate the hard scatter, parton shower, and hadronization for all $q\bar{q}/qg/\bar{q}g \to WW$ diagrams (see Figure 2.5(b) for an example diagram). Events are generated at LO in QCD with up to three jets in the hard scatter, and the W bosons are forced to decay leptonically. The MC prediction is normalized to $\sigma(WW^{(*)} \to \ell\nu\ell\nu) = 5.679$ pb, as computed in MCFM [107]. Given that the branching fraction of $W \to \ell\nu$ is 0.11 and there are nine lepton combinations, the corresponding total cross section is $\sigma_{tot}(WW) = 53.90$ pb. This has been measured in ATLAS in the WW leptonic decay channels and the exclusive zero jet bin [108]. The result, which includes resonant WW production from the Higgs boson, is $\sigma_{tot}(WW) = 71.4^{+1.2}_{-1.2}(\text{stat})^{+5.0}_{-4.4}(\text{syst})^{+2.2}_{-2.1}(\text{lumi})$ pb, significantly larger than the MCFM computation which includes H, $\sigma_{MCFM}(WW) = 58.7^{+3.0}_{-2.7}$ pb. How-

ever, the ATLAS measurement is done in a phase space region that is orthogonal to this analysis, and in fact, the MCFM calculation does not include diagrams with two partons in the final state. The normalization in the 2j bin relies on the LO jet multiplicity prediction from SHERPA. For the production of $gg \to WW$ through a quark loop, the NLO generator GG2WW [109] is used with the parton showering program HERWIG [110], though this contribution is insignificant compared to the other WW processes.

EW WW processes are also simulated with SHERPA, but with exactly two jets in the hard scatter. Additional jets are obtained at the parton shower level. The LO cross section for EW WW with the W bosons decaying to leptons, computed in SHERPA, is $\sigma(WW^{(*)} \to \ell\nu\ell\nu) = 39.68$ fb. Events with Higgs couplings are not simulated; instead, the interference between Higgs and WW processes is assessed as a uncertainty on the cross section.

Like WW, SM Z + jets processes, also referred to as Drell Yan or Z/γ^*, are split into QCD and EW categories depending on the nature of the jets. QCD Z + jets is simulated with the LO event generator ALPGEN [111] interfaced to HERWIG. To enhance statistics, events are generated with a dilepton filter, requiring $m_{\ell\ell} > 10$ GeV, and a filter to select events with jets that are VBF-like (at least two jets with $p_{\mathrm{T}} > 15$ GeV, $|\eta| < 5.0$, $m_{jj} > 200$ GeV, and $\Delta Y_{jj} > 2.0$). VBF-filtered samples are merged with unfiltered samples for full phase space coverage. To enhance statistics for events with a high p_{T} photon ($p_{\mathrm{T}} > 7$ GeV) in addition to the Z boson, a dedicated SHERPA sample is used, and the phase space overlap is removed at truth level. EW Z + jets processes are modeled with SHERPA, with a generator filter to select dilepton events with $m_{\ell\ell} > 7$ GeV.

Non-WW diboson backgrounds are sub-dominant contributions in the VBF analysis. For $WZ/W\gamma^*$, events are generated in POWHEG [112], which treats interference between Z and γ^* diagrams properly. Because POWHEG can not produce events at

low dilepton mass, the phase space in which $m_Z/\gamma^* < 7$ GeV is modeled with SHERPA with up to one final state parton computed at the matrix element level. This prediction is then scaled to the NLO cross section computed in MCFM. ZZ processes are also modeled with POWHEG. For the EW processes without QCD vertices at LO, SHERPA is used for WZ, $W\gamma^*$, and ZZ.

The stable particles produced after the generation of the hard scatter and hadronization are then propagated through a full simulation of the ATLAS detector implemented in GEANT [113]. The resulting simulated signals are then converted into hits in a digitization step. In this step, the effects of pile-up are simulated by overlaying events on top of the primary hard scatter event. Detector noise is also simulated at this stage.

7

Object and Event Selection

The final state physics objects of the signal process, VBF $H \to WW^{(*)} \to \ell\nu\ell\nu$, are two leptons from the decays of the W bosons, missing transverse energy associated with the neutrinos from W bosons, and two jets arising from the quarks involved in the production of the Higgs boson via VBF. A general discussion of the reconstruction of these objects can be found in Chapter 4. In this chapter, analysis-specific object selections and the selection of the VBF phase space are discussed.

7.1 Object Definitions

7.1.1 Electrons

With two leptons in the final state, it is crucial to efficiently select leptons, while minimizing contamination from backgrounds with jets faking leptons. Such backgrounds are W + jets and multi-jet production via QCD. Because these backgrounds fall off quickly with increasing lepton p_{T}, lepton selections have been optimized in bins of p_{T}. For electron identification, the likelihood-based identification category VERY TIGHT (Section 4.3.2), which has the highest background-rejecting power of

all of the identification categories, is required at low electron E_{T} ($E_{\mathrm{T}} < 25$ GeV). To recover efficiency at high E_{T}, where the fake background contribution is smaller, the identification category is relaxed to cut-based *medium*. The standard *medium* definition is modified slightly to improve rejection against electrons from photon conversions. Since the pixel b-layer is positioned just outside of the beam pipe, most photon conversions occur beyond this detector layer. Therefore, to suppress photon conversions, electrons are required to have a hit in the pixel b-layer if one is expected given the track parameters. Also, if an electron track coincides with a track that is associated to a reconstructed conversion vertex, the electron is flagged as a photon conversion and removed from the electron collection.

Further rejection of fake background is accomplished by requiring electrons to be isolated. Cuts are placed on the ratio of the sum of the E_{T} of topological clusters in a cone of radius R around the electron to the electron E_{T}, $E_{R=r}/E_{\mathrm{T}}$, where r and the cut value have been optimized in bins of E_{T}. To avoid including the energy from the electron itself, the energy in a window of dimension $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ centered on the electron candidate is removed from the isolation energy sum. An additional correction to the isolation energy is applied on an event-by-event basis to account for calorimeter energy arising from pileup and the underlying event [114].

In addition to this calorimeter-based isolation cut, electrons are required to satisfy track-based isolation requirements in which the p_{T} sum of the tracks in a cone of radius R divided by the electron E_{T} is the discriminating variable. Track-based isolation is more robust against pileup, making it a more powerful discriminant against fakes. Finally, cuts are applied to the transverse electron impact parameter significance ($d_0/\sigma(d_0) < 3$), as well as the longitudinal impact parameter ($z_0 \sin(\theta) < 0.4$ mm), for rejection against fake electrons arising from pileup vertices. Table 7.1 summarizes the electron selection.

To account for mis-modeling of electron variables in simulation, scale factors (SFs)

Table 7.1: Electron selection summary in E_{T} bins, including the identification category, the calorimeter and track isolation cuts, and the impact parameter cuts.

E_{T} (GeV)	electron ID	calo. isolation	track isolation	impact parameters
10-15	VERY TIGHT likelihood	$E_{\mathrm{T}}^{R=0.3}/E_{\mathrm{T}} < 0.20$	$p_T^{R=0.4}/E_{\mathrm{T}} < 0.06$	$d_0/\sigma(d_0) < 3.0$, $z_0 \sin\theta < 0.4$ mm
15-20		$E_{\mathrm{T}}^{R=0.3}/E_{\mathrm{T}} < 0.24$	$p_T^{R=0.3}/E_{\mathrm{T}} < 0.08$	
20-25		$E_{\mathrm{T}}^{R=0.3}/E_{\mathrm{T}} < 0.28$	$p_T^{R=0.3}/E_{\mathrm{T}} < 0.10$	
> 25	*medium* + conversion			

are applied to each electron selected in the MC prediction (Section 4.3.2). These SFs correct the efficiency differences between data and simulation for the trigger, reconstruction, identification, isolation, and impact parameter requirements. All of these efficiencies are measured with an electron-rich $Z \to ee$ sample in data with the tag-and-probe technique described in Chapter 4. The total electron efficiency is shown in Table 7.2 in bins of E_{T} for a $H \to WW^{(*)}$ sample at $m_H = 125$ GeV. Due to the use of VERY TIGHT identification for $E_{\mathrm{T}} < 25$ GeV, the efficiency in the low E_{T} bins is relatively low, ranging from 40% to 70%.

Table 7.2: Total lepton selection efficiencies and associated uncertainties for a $m_H = 125$ GeV Higgs signal sample. The uncertainties are split into isolation and reconstruction/identification, respectively. For ϵ_{muon}, if the uncertainty is less than 0.005, then 0.005 is shown. The total uncertainty can be obtained by adding the two components in quadrature.

E_T	$\epsilon_{\mathrm{electron}}$	ϵ_{muon}
10-15	$0.412 \pm 0.016 \pm 0.016$	$0.574 \pm 0.027 \pm 0.005$
15-20	$0.619 \pm 0.009 \pm 0.024$	$0.808 \pm 0.012 \pm 0.005$
20-25	$0.668 \pm 0.008 \pm 0.027$	$0.904 \pm 0.007 \pm 0.005$
25-30	$0.755 \pm 0.007 \pm 0.014$	$0.924 \pm 0.006 \pm 0.005$
30-35	$0.770 \pm 0.007 \pm 0.005$	$0.932 \pm 0.006 \pm 0.005$
35-40	$0.796 \pm 0.006 \pm 0.003$	$0.942 \pm 0.005 \pm 0.005$
40-45	$0.798 \pm 0.006 \pm 0.002$	$0.943 \pm 0.005 \pm 0.005$
45-50	$0.813 \pm 0.006 \pm 0.002$	$0.944 \pm 0.005 \pm 0.005$

7.1.2 Muons

This analysis requires muons to be of the "combined" type in which the combined muon track is a statistical combination of independent ID and MS track fits (Section 4.4). The ID track associated with the muon must satisfy the requirements: (1) the sum of pixel hits and dead pixel sensors crossed by the track must be greater than zero, (2) the sum of SCT hits and dead SCT sensors crossed by the track must be greater than four, (3) the number of missing hits in a crossed SCT sensor that is not dead must be less than three, and (4) a TRT extension is found if the track falls within the TRT acceptance. These cuts have been studied rigorously by a dedicated performance group that also computes the SFs associated with such a selection using tag-and-probe.

To suppress backgrounds from processes with "fake" muons, requirements are placed on isolation variables as well as the transverse and longitudinal impact parameters. Isolation cut values have been optimized in bins of muon p_{T}. To account for pileup dependence, the calorimeter isolation is corrected event-by-event according to N_{PV}. A sensitivity scan has been performed over the two-dimensional parameter space for the $d_0/\sigma(d_0)$ and $z_0 \sin(\theta)$ cuts to find the optimal cut values. The resulting selection is shown in Table 7.3.

Table 7.3: Muon selection summary in p_{T} bins, including the identification category, the calorimeter and track isolation cuts, and the impact parameter cuts.

p_{T} (GeV)	calo. isolation	track isolation	impact parameters
10-15	$E_{\mathrm{T}}^{R=0.3}/p_{\mathrm{T}} < 0.06$	$p_{\mathrm{T}}^{R=0.4}/p_{\mathrm{T}} < 0.06$	$d_0/\sigma(d_0) < 3.0$, $z_0 \sin\theta < 1.0$ mm
15-20	$E_{\mathrm{T}}^{R=0.3}/p_{\mathrm{T}} < 0.12$	$p_{\mathrm{T}}^{R=0.3}/p_{\mathrm{T}} < 0.08$	
20-25	$E_{\mathrm{T}}^{R=0.3}/p_{\mathrm{T}} < 0.18$	$p_{T}^{R=0.3}/p_{\mathrm{T}} < 0.12$	
> 25	$E_{\mathrm{T}}^{R=0.3}/p_{\mathrm{T}} < 0.30$		

Data-derived efficiencies associated with the impact parameter and isolation cuts

are also measured with tag-and-probe and those from simulation are corrected to reflect data. The resulting total efficiency for selecting a muon is shown in Table 7.2. Compared to electrons, the muon efficiencies associated with the optimal cuts are 20%-40% higher, due to the fact that electrons are required to have tighter quality cuts to suppress problematic backgrounds such as jets and photon conversions.

7.1.3 Jets

As discussed in Section 4.5, jets are reconstructed with the anti-k_{T} $R = 0.4$ algorithm and calibrated with LCW. These jets are required to satisfy the quality criteria for the LOOSER category, which has the highest efficiency of all of the jet cleaning categories. The remaining jet cuts are optimized for maximum sensitivity, while minimizing fake jets from pileup which induce migrations into and out of the signal regions. In the optimization, cut values for the jet p_{T} and the JVF (defined in Section 4.5.3) are simultaneously scanned, extracting both an estimate of the expected significance, as well as the uncertainty due to bin migrations. The optimal configuration has been found to be jet $p_{\mathrm{T}} > 25$ GeV and JVF > 0.5 in the $|\eta| < 2.4$ region. In the region outside of the tracker, $2.4 \leq |\eta| < 4.5$, where JVF is unavailable, the p_{T} threshold is increased to 30 GeV. Finally, jets in $|\eta| < 2.4$ with $p_{\mathrm{T}} > 50$ GeV do not have any JVF requirement, since the rate of pileup jets in this p_{T} range is relatively small, and therefore a JVF requirement only serves to degrade the efficiency. The resulting uncertainty due to jet bin migrations for these cuts is 6.3% in a VBF-rich region of phase space, significantly smaller than uncertainties due to JES and JER (Section 4.5).

The jet definition above defines the high-p_{T} tag jets characteristic of VBF. In addition to these jets, another jet collection is considered. Jets which fall in between the pseudorapidity range defined by the two tag jets with $p_{\mathrm{T}} > 15$ GeV and which satisfy the same JVF requirement as the tag jets are placed into a separate jet

collection. This central jet collection is used to veto events with high p_{T} central jets.

7.1.4 b-hadron Jets

The tagging of b-hadron (heavy flavor) jets, described in Section 4.6, is crucial for (1) rejecting background processes with top quarks, and (2) establishing a control region from which the normalization of such backgrounds is extrapolated. In this analysis, b-tagging is performed with the MV1 algorithm, a multi-layered neural network approach that incorporates impact parameter and reconstructed vertex information from tracks associated with jets. Each jet is assigned a score quantifying the probability that it is a b jet, and if the score falls over a threshold, then the jet is tagged. The threshold has been chosen such that 85% of true b-jets are tagged.

The associated b-tagging efficiencies are measured in a $t\bar{t}$-rich control region in data, using the likelihood procedure described in Section 4.6. Scale factors, defined as the ratio of the tagging efficiency in data to that in simulation, are then applied in simulation. These SFs, which are evaluated in six p_{T} bins, are shown with their associated uncertainties, in Table 7.4. The SFs deviate from unity by less than 5% and are consistent with unity within statistical and systematic uncertainties.

For the uncertainties on scale factors, the eigenvector method is used to reduce the number of uncorrelated variations. A covariance matrix is constructed for each source of uncertainty, and the total covariance matrix is then obtained by summing each source matrix. The total covariance matrix is transformed to its eigenbasis, and the square root of the matrix eigenvalues are then taken as the fully uncorrelated systematic variations on the b-tag SFs.

In addition to the b-tag SFs, there are SFs to correct the mis-tag efficiency for c-jets and light flavor jets. The uncertainties on these two factors consider the same sources as those for b-tag SFs, but in this case, the uncertainties for each source are summed in quadrature, resulting in a single uncertainty.

Table 7.4: b-tagging scale factors and uncertainties in p_{T} bins. Uncertainties are split into statistical and systematic components, and the total uncertainty is the sum in quadrature of the two.

p_{T} bin	Scale Factor	Statistical Uncertainty	Systematic Uncertainty	Total Uncertainty
$20-30$	0.999	1.5%	4.7%	4.9%
$30-60$	1.006	0.5%	1.8%	1.9%
$60-90$	0.989	0.5%	1.6%	1.6%
$90-140$	0.996	0.6%	1.4%	1.5%
$140-200$	0.965	1.2%	2.9%	3.1%
$200-300$	1.046	2.9%	6.6%	7.2%

7.1.5 $E_{\mathrm{T}}^{\mathrm{miss}}$

Missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) is discussed at length in Section 4.7. In this analysis, $E_{\mathrm{T}}^{\mathrm{miss}}$ arises due to the presence of undetected neutrinos from the W boson decays. Two different $E_{\mathrm{T}}^{\mathrm{miss}}$ definitions are used. Calorimeter-based $E_{\mathrm{T}}^{\mathrm{miss}}$ ($E_{\mathrm{T}}^{\mathrm{miss,CALO}}$) reconstructs the transverse energy in the event with calibrated physics objects and the soft energy in the event is reconstructed with calorimeter topological clusters. Track-based $E_{\mathrm{T}}^{\mathrm{miss}}$ ($E_{\mathrm{T}}^{\mathrm{miss,TRK}}$), on the other hand, uses tracks matched to the primary vertex to reconstruct the soft energy, and is therefore less sensitive to in-time pileup. Moreover, $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ is a standardized definition which is used across ATLAS analyses, while $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ uses physics object definitions from analysis-specific optimizations, yielding better scale and angular resolution than $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$.

Quality cuts on the tracks entering $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ are enumerated in Section 4.7. If a track fails this selection, but is associated with either an electron or a muon, it is added to the track collection. Electrons are required to satisfy either (1) *medium* identification requirements, $E_{\mathrm{T}}^{\mathrm{cluster}} > 10$ GeV, and $|\eta| < 2.47$ or (2) the analysis-level electron requirements described in Section 7.1.1. Muons are required to satisfy cuts which are looser than those in the analysis, namely combined muons with $p_{\mathrm{T}} > 6$ GeV, $|\eta| < 2.5$ and $|z_0 \sin(\theta)| < 1$ mm. Tracks that overlap with reconstructed

electrons and muons are removed from the track collection, as are tracks that fall within a cone of 0.4 of an analysis-level jet. This improves the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution because neutral hadrons associated with jets which do not produce tracks but do deposit energy in the calorimeter are not excluded.

The systematic uncertainties from all of the reconstructed physics objects which define $E_{\mathrm{T}}^{\mathrm{miss}}$ must be combined coherently to obtain the total uncertainty on $E_{\mathrm{T}}^{\mathrm{miss}}$. For the hard objects, such as leptons and jets, the uncertainties are evaluated independently and propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$. For the soft term, uncertainties associated with the scale and resolution calibrations are evaluated with data-driven techniques for both $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ [65] and $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$. For the calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties, a $Z \to \mu\mu$ sample is isolated, and to simplify $E_{\mathrm{T}}^{\mathrm{miss}}$ such that only muons and the soft term contribute, exactly zero jets with p_{T} greater than 20 GeV are required. To evaluate the scale uncertainty, the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector is projected onto $\mathrm{p_T}$ of the Z boson, and the mean value of this quantity is evaluated in bins of $\sum E_{\mathrm{T}}$. The difference between data and MC simulation is assigned as the scale uncertainty on the soft term, which, for $\sqrt{s} = 8$ TeV data, is 3.6% averaged over all $\sum E_{\mathrm{T}}$. The soft term resolution uncertainty is determined by comparing data and simulation, but the quantities of interest are the widths of the x and y components of $E_{\mathrm{T}}^{\mathrm{miss}}$. Integrating over $\sum E_{\mathrm{T}}$, the mean uncertainty is 2.3%. Another approach for evaluating the uncertainty, which utilizes the balance between hard and soft objects, yields similar results.

The $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ uncertainties are evaluated with a data-driven approach that balances the p_{T} of hard objects and the soft term. Defining the energy components of the hard objects as

$$E_{\mathrm{x(y)}}^{\mathrm{hard}} = \Sigma_{\mu}\, E_{\mathrm{x(y)}}^{\mu} + \Sigma_{ele}\, E_{\mathrm{x(y)}}^{ele} + \Sigma_{\nu}\, E_{\mathrm{x(y)}}^{\nu} + \Sigma_{\mathrm{jets}}\, E_{\mathrm{x(y)}}^{\mathrm{jets}}, \tag{7.1}$$

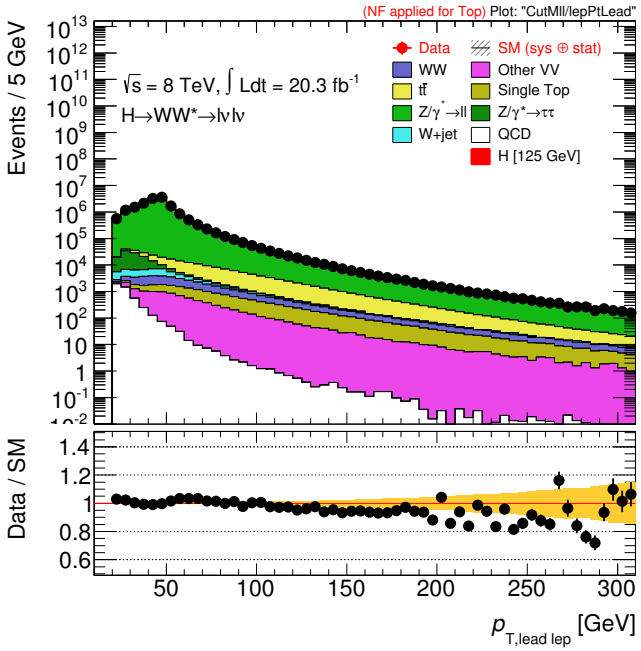
the neutrino term is set to zero in data and extracted from the truth information in

simulation. The soft term is then just the reconstructed $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ with the hard objects subtracted. To evaluate the uncertainty, the scale and resolutions of $E_{\mathrm{T,soft}}^{\mathrm{miss,trk}}$ are measured in data and simulation as a function of $E_{\mathrm{T}}^{\mathrm{hard}}$ and the average number of interactions per bunch crossing $\langle\mu\rangle$. Binning in $E_{\mathrm{T}}^{\mathrm{hard}}$ allows the uncertainty to be extrapolated to events which are kinematically different from $Z \to \mu\mu$ events, while separating the uncertainty into $\langle\mu\rangle$ bins accounts for potential effects from in-time pileup.

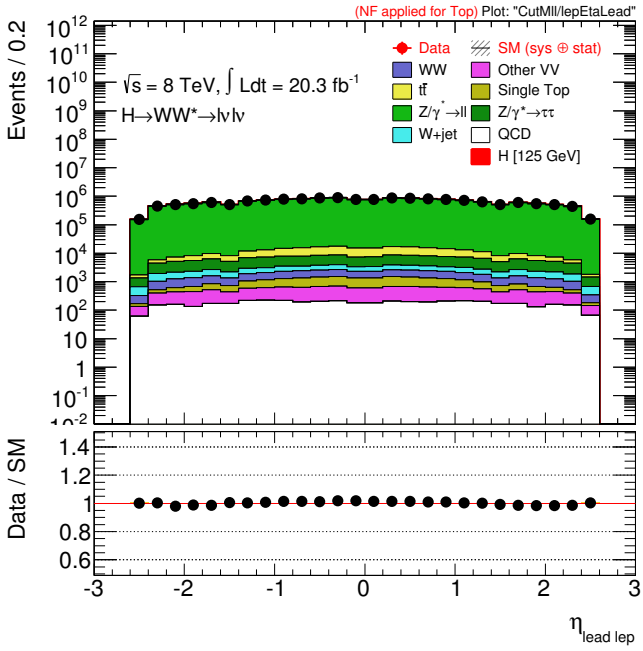
The $E_{\mathrm{T,soft}}^{\mathrm{miss,trk}}$ term is decomposed into the component which is parallel to the $E_{\mathrm{T}}^{\mathrm{hard}}$ and one which is perpendicular to $E_{\mathrm{T}}^{\mathrm{hard}}$. The former is sensitive to the modeling of the soft hadronic activity against which the hard system is recoiling, while the latter is a measure the intrinsic $E_{\mathrm{T,soft}}^{\mathrm{miss,trk}}$resolution. To extract the uncertainty, for the parallel component, the simulation is smeared with a Gaussian to reproduce the resolution in data and then shifted to the scale observed in data. These two corrections are applied as systematic uncertainties. For the perpendicular component, only smearing is performed. The resulting uncertainties are summarized in Table 7.5. Very little pile-up dependence has been observed, and therefore the uncertainties are not ultimately assigned as a function of $\langle\mu\rangle$. Moreover, it has been checked that these uncertainties are robust against different MC generators as well as the jet scale and resolution uncertainties, with the largest effect coming from the parton shower model in simulation. Uncertainties are adjusted accordingly to account for this.

7.2 Event Selection

With the physics objects defined, event-level selection cuts are applied sequentially in order to enhance signal. These cuts can be categorized into those which are shared with the gluon fusion $H \to WW^{(*)} \to \ell\nu\ell\nu$ analyses, and those which are specific to the VBF analysis.



(a) Transverse momentum



(b) Pseudo-rapidity

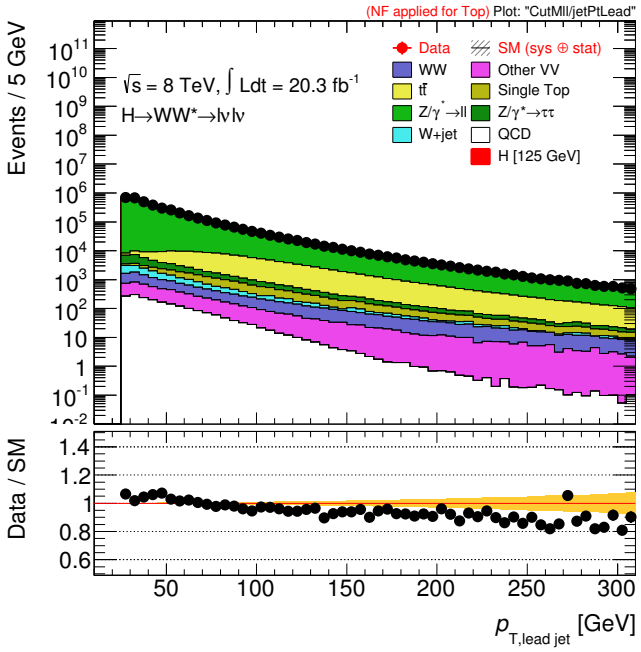
FIGURE 7.1: The leading lepton (a) p_{T} and (b) η at pre-selection for all lepton flavors combined. Data-driven corrections to backgrounds are not applied at this stage. The error band includes statistical uncertainties only.

Table 7.5: The absolute systematic variations (GeV) on the $E_{\mathrm{T,soft}}^{\mathrm{miss,trk}}$ term of $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$. Scale and resolution uncertainties are shown in bins of $E_{\mathrm{T}}^{\mathrm{hard}}$. The scale uncertainty corresponds to the shift factor which is applied to the parallel projection of $E_{\mathrm{T,soft}}^{\mathrm{miss,trk}}$ in simulation. The resolution uncertainties, split into the parallel and perpendicular components, correspond to additional smearing applied to the soft terms to match the width of the distribution observed in data.

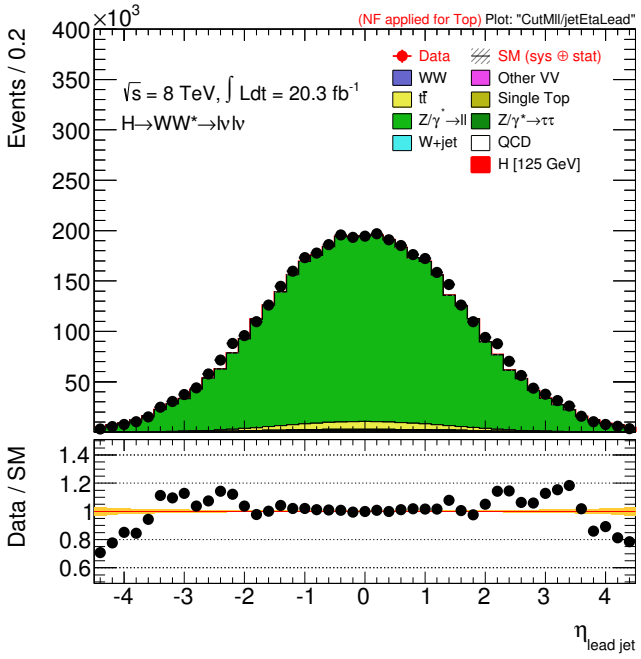
$E_{\mathrm{T}}^{\mathrm{hard}}$ bin (GeV)	scale (GeV)	reso. para. (GeV)	reso. perp. (GeV)
0-5	0.28	1.57	1.68
5-10	0.38	1.60	1.58
10-15	0.58	1.60	1.61
15-20	0.73	1.75	1.72
20-25	0.82	1.88	1.72
25-30	0.99	2.13	1.84
35-40	1.12	2.35	2.10
40-45	1.20	2.58	2.15
50+	1.44	3.33	2.68

7.2.1 Common Preselection

For all $H \to WW^{(*)} \to \ell\nu\ell\nu$ analyses, exactly two leptons with opposite charge are required, with the leading lepton $p_{\mathrm{T}} > 22$ GeV and the subleading lepton $p_{\mathrm{T}} > 10$ GeV. These cuts have been optimized to maximize signal acceptance, while minimizing contamination from background due to jets faking leptons. A dilepton mass cut of $m_{\ell\ell} > 10$ (12) GeV is applied in the $e\mu/\mu e$ ($ee/\mu\mu$) lepton channel, and to reject background from Z/DY in the $ee/\mu\mu$ channel, events with dilepton mass falling within 15 GeV of the Z pole mass are discarded. The p_{T} and η distributions of the the leading lepton (jet) are shown in Figure 7.1 (7.2) after the common pre-selection cuts have been applied. MC simulation models the data adequately in this phase space region. Discrepancies in the jet distributions at high p_{T} and high $|\eta|$ which lie outside of the statistical uncertainty band are covered by the systematic uncertainties associated with the JES and JER calibrations (Section 4.5).

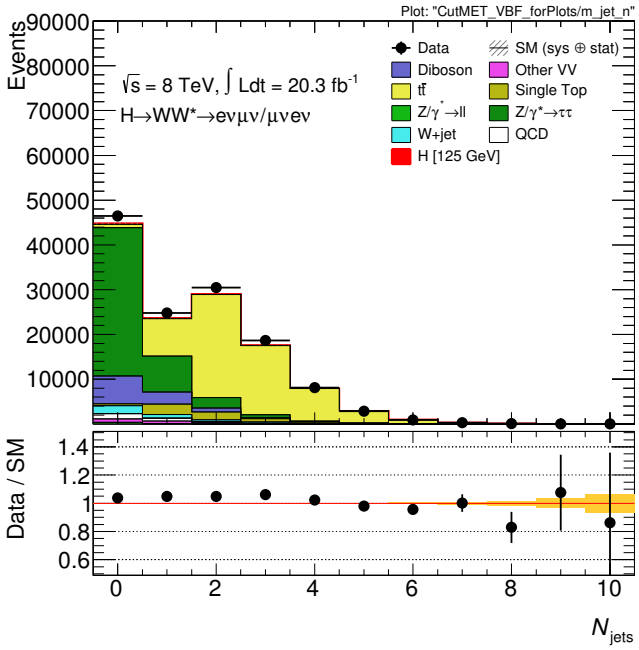


(a) Transverse momentum

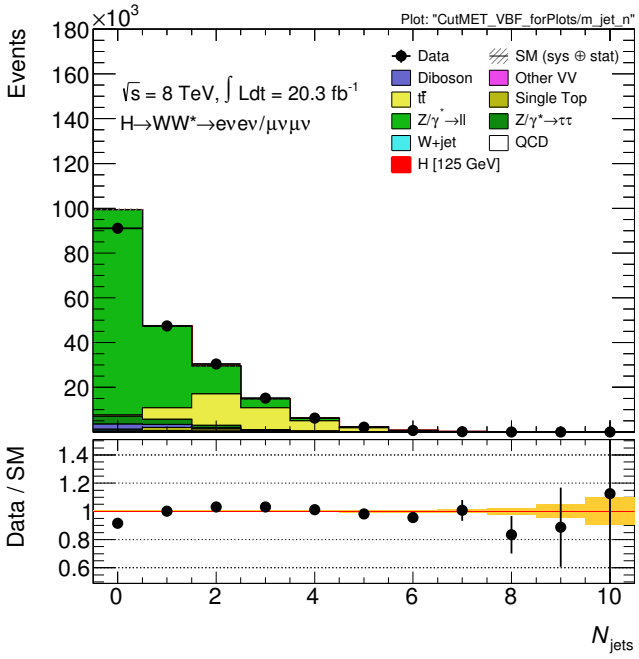


(b) Pseudo-rapidity

FIGURE 7.2: The leading jet (a) p_{T} and (b) η at pre-selection for all lepton flavors combined. Data-driven corrections to backgrounds are not applied at this stage. The error band includes statistical uncertainties only.



(a) $e\mu/\mu e$ channel



(b) $ee/\mu\mu$ channel

FIGURE 7.3: Jet multiplicity distributions for (a) $e\mu/\mu e$ channel and (b) $ee/\mu\mu$ channel after common pre-selection. In (b) $E_{\mathrm{T}}^{\mathrm{miss}}$ cuts and Z veto are also applied. Error band represents statistical uncertainties.

7.2.2 VBF-specific pre-selection

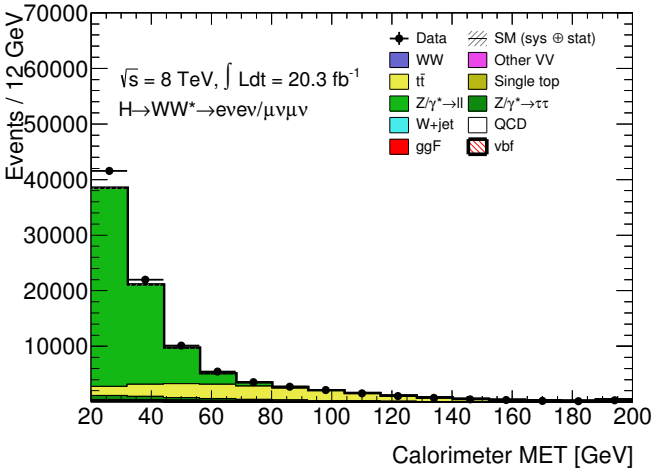
The VBF analysis selection departs from the common selection starting with a cut on the number of jets in the event, which is required to be greater than or equal to two, with the two leading jets considered to be the tag jets that are characteristic of VBF Higgs production. Looking at the jet multiplicity distributions after pre-selection (Figure 7.3), the dominant backgrounds with $N_{\text{jet}} \geq 2$ are $t\bar{t}$ in both lepton channels and also Z/DY in the $ee/\mu\mu$ channel. Additional pre-selection cuts are designed to suppress these large backgrounds.

In the $e\mu/\mu e$ channel, although signal has $E_{\text{T}}^{\text{miss}}$ from final state neutrinos, there is not an $E_{\text{T}}^{\text{miss}}$ cut applied, since the dominant backgrounds also have neutrinos in the final state. Without a cut on $E_{\text{T}}^{\text{miss}}$, there is a gain in signal acceptance corresponding to a sensitivity gain of 6% compared to the cut value of 20 GeV in the previous analysis [66].

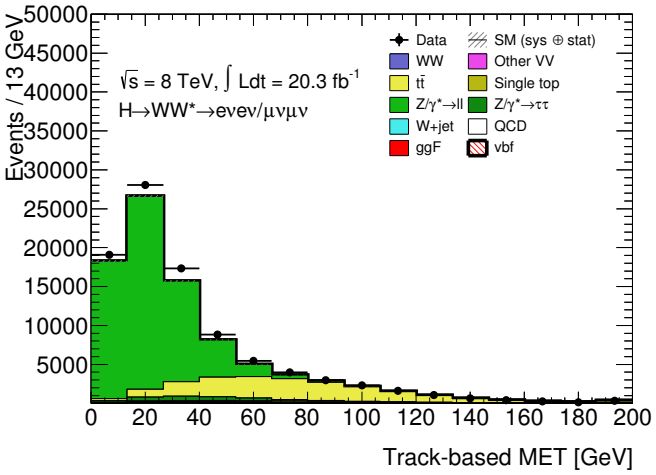
Calorimeter and track $E_{\text{T}}^{\text{miss}}$ distributions in the $ee/\mu\mu$ channel after the $N_{\text{jet}} \geq 2$ cut are shown in Figure 7.4. $Z/\text{DY} \rightarrow ee/\mu\mu$ events have small $E_{\text{T}}^{\text{miss}}$, as this process does not have neutrinos in the final state. To reject this background, $E_{\text{T}}^{\text{miss}}$ cuts are applied, with the optimal cuts being $E_{\text{T}}^{\text{miss,TRK}} > 40$ GeV and $E_{\text{T}}^{\text{miss,CALO}} > 45$ GeV.

To suppress top quark background, events with one or more b-tagged jets are vetoed (b jet veto or BJV), a cut that removes 94% of $t\bar{t}$ and 87% of ST, while retaining ~70% of signal. In the majority of the remaining top background, one of the tag jets is a true heavy flavor jet that is not tagged because it falls outside of the tracking volume for b-tagging ($|\eta| > 2.5$), while the other tag jet is a light flavor jet from ISR, and is therefore not tagged. In such events, the other heavy flavor jet is not b-tagged due to inefficiency in the tagging algorithm, or because the jet p_{T} is below the 20 GeV threshold for b-tagging.

Compared to VBF, processes with a pair of high p_{T} jets arising from QCD vertices



(a) Calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$



(b) Track $E_{\mathrm{T}}^{\mathrm{miss}}$

FIGURE 7.4: Distributions of (a) $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ and (b) $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ in the $ee/\mu\mu$ channel with the common pre-selection cuts and $N_{\mathrm{jet}} \geq 2$. Data-driven background corrections are not applied at this stage, and the error band represents statistical uncertainties only.

have more soft QCD activity in the rapidity region between the jets due to color exchange. The central jet veto (CJV) exploits this by requiring that the highest p_{T} jet that falls between the two leading jets has $p_{\mathrm{T}} < 20$ GeV. This cut retains events without any central jets because, in this case, the lead p_{T} of the central jets is set to a value of -1. In addition to the CJV, an outside lepton veto (OLV)

is applied, exploiting the tendency for leptons resulting from the top quark decay, $t \to Wb \to \ell\nu b$, to fall outside of the pseudorapidity of the jets. The OLV is defined such that if either of the leptons falls outside of the pseudorapidity gap, the event is vetoed.

The final preselection cut suppresses $Z \to \tau\tau$ processes, which contribute in the signal region if both tau leptons decay leptonically: $\tau \to \ell\nu_\ell\bar{\nu}_\tau$. The presence of neutrinos in the final state complicates suppression of this process, because the mass of the Z can not be fully reconstructed from its daughter leptons. Instead, the invariant mass $m_{\tau\tau}$ is estimated using the collinear approximation [115], which assumes that (1) for each tau lepton decay, the neutrinos are collinear with the visible decay products, and (2) the only source of $E_{\mathrm{T}}^{\mathrm{miss}}$ in each event is the neutrinos from the τ decays. Assumption (1) is valid in the limit that the masses of the particles involved in the decay are negligible compared to their momenta, and (2) is valid up to detector resolution effects. Under these assumptions, the momentum of the invisible decay products for each tau lepton is obtained by solving the following system of equations for p_1^{miss} and p_2^{miss}:

$$\begin{aligned} E_{\mathrm{T,x}}^{\mathrm{miss}} &= p_1^{\mathrm{miss}} \sin\theta_{\mathrm{lep1}} \cos\phi_{\mathrm{lep1}} + p_2^{\mathrm{miss}} \sin\theta_{\mathrm{lep2}} \cos\phi_{\mathrm{lep2}} \\ E_{\mathrm{T,y}}^{\mathrm{miss}} &= p_1^{\mathrm{miss}} \sin\theta_{\mathrm{lep1}} \sin\phi_{\mathrm{lep1}} + p_2^{\mathrm{miss}} \sin\theta_{\mathrm{lep2}} \sin\phi_{\mathrm{lep2}}, \end{aligned} \tag{7.2}$$

where θ (ϕ) denotes the polar (azimuthal) angle of the visible decay product. Using $p_{1(2)}^{\mathrm{miss}}$, the momentum fraction carried by the visible leptons, $x_{1(2)} = p_{\mathrm{lep1(2)}}/(p_{\mathrm{lep1(2)}} + p_{1(2)}^{\mathrm{miss}})$, is used to define $m_{\tau\tau}$:

$$m_{\tau\tau} = \frac{m_{\ell\ell}}{\sqrt{x_1 x_2}}. \tag{7.3}$$

If the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector does not fall between the momentum vectors of the charged leptons, at least one of x_1 and x_2 is negative. This is typically true of signal events

given the decay topology of $H \to WW^{(*)} \to \ell\nu\ell\nu$. It is also common in top and WW backgrounds where the assumption of collinearity breaks down since the mass of the parent W is large. For events in which x_1 or x_2 is less than zero, $m_{\tau\tau}$ is set to -9999. The distribution of $m_{\tau\tau}$ after the OLV in the $e\mu/\mu e$ channel is displayed in Figure 7.5. The underflow bin illustrates that the majority of signal, top, and WW events have negative x_1 or x_2. $Z \to \tau\tau$ peaks at the Z pole mass, and is suppressed by requiring that $m_{\tau\tau} < m_Z - 25$ GeV.

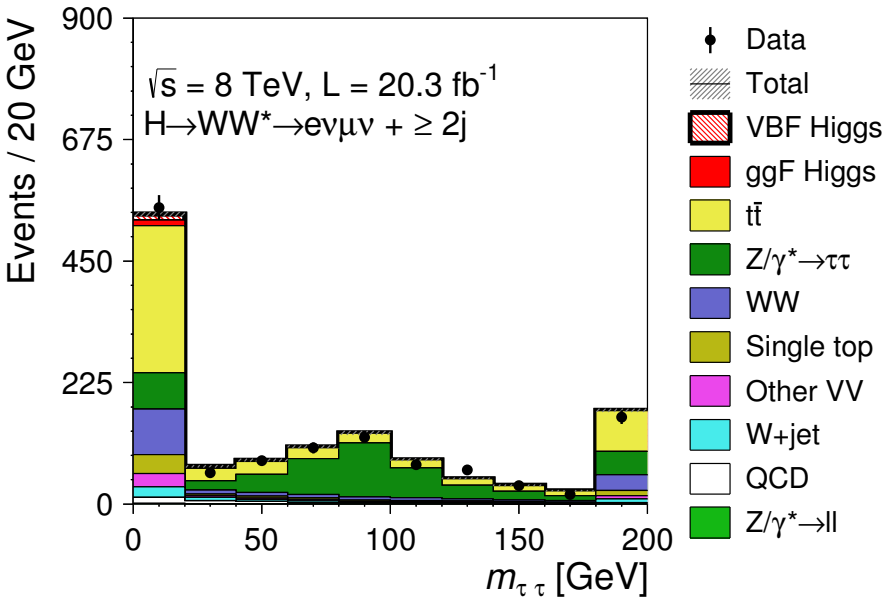


FIGURE 7.5: Distribution of $m_{\tau\tau}$ after the outside lepton veto in the $e\mu/\mu e$ channel. Underflow and overflow bins are shown. The hatched uncertainty band includes statistical uncertainties only. Normalization factors are not applied.

In the case where the leptons are back-to-back, i.e. $\phi_{\text{lep1}} = \phi_{\text{lep1}} + \pi$, the system in Equation 7.2 becomes singular. This causes degradation in the mass resolution of $m_{\tau\tau}$.

7.3 BDT Inputs

The BDT algorithm, described in Appendix A, is well-suited for the classification of VBF signal and its associated backgrounds, as the final state includes many correlated physics objects. The BDT training algorithm builds a model $F(\boldsymbol{x})$ that maps the input variables $\boldsymbol{x}$ to a number y in the range [-1,1], with -1 corresponding to events that are background-like and 1 corresponding to those that are signal-like. In the following, the BDT response y will be referred to as BDT or BDT score.

7.3.1 Higgs Decay

There are eight BDT inputs: $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, m_{T}, m_{jj}, ΔY_{jj}, $p_{\mathrm{T}}^{\mathrm{tot}}$, η_{lep} centrality, and $\sum_{\ell,j} M_{\ell j}$. These can be categorized into those that are sensitive to the topology of the $H\to WW^{(*)}\to\ell\nu\ell\nu$ decay and those that pick out the products of Higgs production via VBF.

The absolute value of the difference in azimuthal angle of the two leptons, $\Delta\phi_{\ell\ell} = |\phi_{\mathrm{lep1}} - \phi_{\mathrm{lep2}}|$, is an example of an input falling into the first category. As discussed in Section 2.4, the final state leptons in the decay chain $H\to WW^{(*)}\to\ell\nu\ell\nu$ fall close in azimuthal angle due to the spin of the SM Higgs boson and the V-A structure of the weak force. Shown at the pre-selection stage in Figure 7.6, $\Delta\phi_{\ell\ell}$ is peaked near zero for signal. This distribution is relatively flat for the dominant backgrounds from top quark and non-resonant WW processes.

Another variable that picks out the Higgs decay topology is $m_{\ell\ell}$. Because $m_{\ell\ell} \simeq |p_1||p_2|\,[1-\cos(\Delta\phi_{\ell\ell})]$ and signal peaks at $\Delta\phi_{\ell\ell}\sim 0$, the signal lies at low values of $m_{\ell\ell}$. As shown in Figure 7.6, top quark, Z/DY, and non-resonant WW processes typically have higher values of $m_{\ell\ell}$.

The transverse mass of the $WW^{(*)}\to\ell\nu\ell\nu$ system (m_{T}) [116, 117] is designed to capture the mass of the Higgs and is therefore a powerful discriminant against

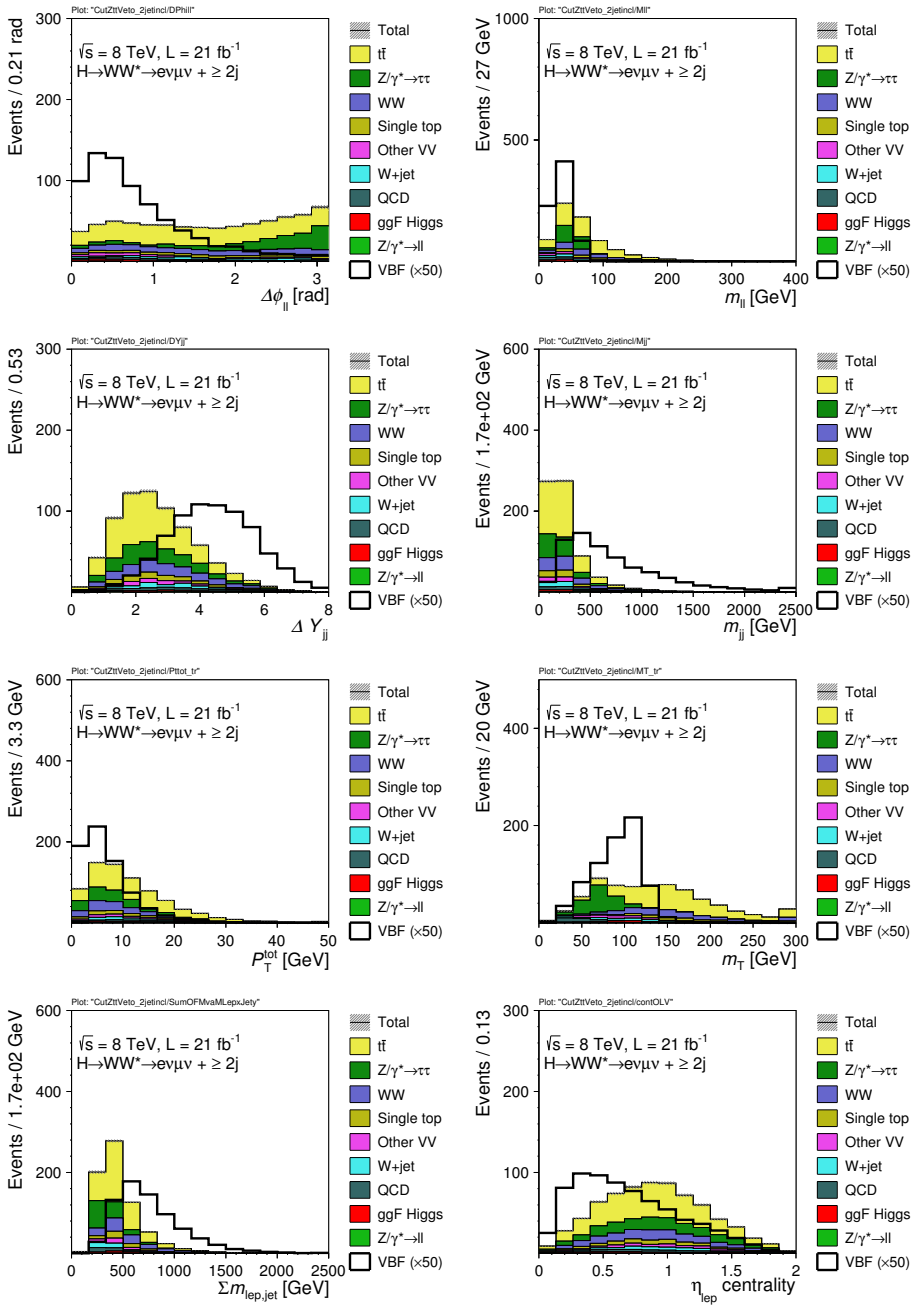


FIGURE 7.6: Distributions of the eight BDT inputs $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, ΔY_{jj}, m_{jj}, $p_{\mathrm{T}}^{\mathrm{tot}}$, m_{T}, $\sum_{\ell,j} M_{\ell j}$, and η_{lep} centrality in the $e\mu/\mu e$ channel after pre-selection. Signal is enhanced by a factor of 50 to illustrate the separation between signal and background.

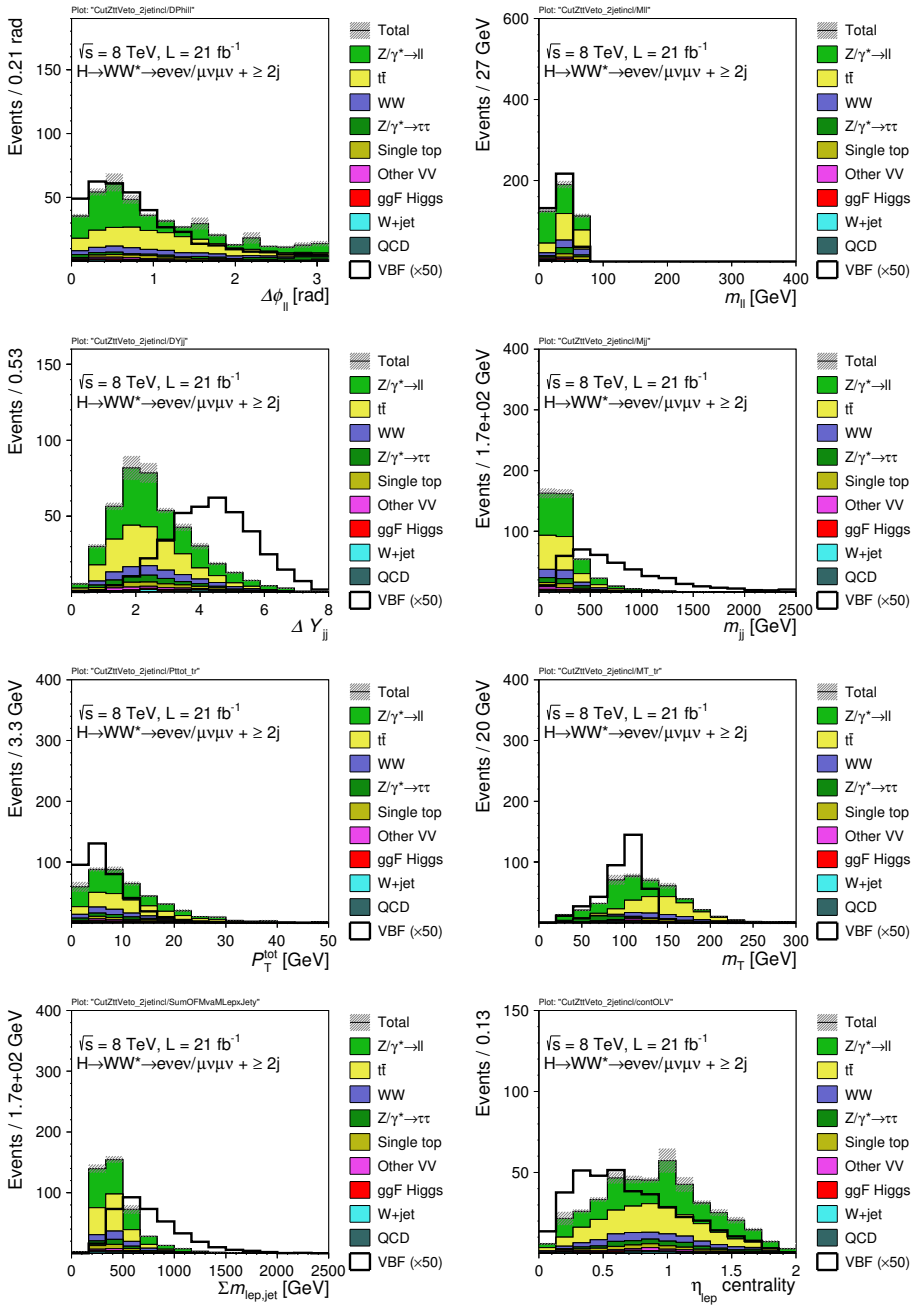


FIGURE 7.7: Distributions of the eight BDT inputs $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, ΔY_{jj}, m_{jj}, $p_{\mathrm{T}}^{\mathrm{tot}}$, m_{T}, $\sum_{\ell,j} M_{\ell j}$, and η_{lep} centrality in the $ee/\mu\mu$ channel after pre-selection. Signal is enhanced by a factor of 50 to illustrate the separation between signal and background.

non-resonant WW background. With neutrinos in the final state, it is not possible to fully reconstruct the mass of the Higgs, $m_H = \sqrt{(p^{\ell\ell} + p^{\nu\nu})^2}$, where $p^{\ell\ell} = p_{\ell 1} + p_{\ell 2}$ and $p^{\nu\nu} = p_{\nu 1} + p_{\nu 2}$, because the neutrino 4-vectors are not reconstructed. A partial reconstruction of m_H is achieved by setting the longitudinal components of each 4-vector to zero in the expression above, resulting in a transverse mass: $m_{\mathrm{T}} = \sqrt{(p_{\mathrm{T}}^{\ell\ell} + p_T^{\nu\nu})^2}$. Expanding gives the expression for the true transverse mass

$$\begin{aligned}(m_{\mathrm{T}}^{\mathrm{true}})^2 &= E_{\mathrm{T}}^{\ell\ell\,2} - |\boldsymbol{p}_{\mathrm{T}}^{\ell\ell}|^2 + E_{\mathrm{T}}^{\nu\nu\,2} - |\boldsymbol{p}_{\mathrm{T}}^{\nu\nu}|^2 + 2E_{\mathrm{T}}^{\ell\ell}E_{\mathrm{T}}^{\nu\nu} - 2\boldsymbol{p}_{\mathrm{T}}^{\ell\ell}\cdot\boldsymbol{p}_{\mathrm{T}}^{\nu\nu} \\ &= (E_{\mathrm{T}}^{\ell\ell} + E_{\mathrm{T}}^{\nu\nu})^2 - |\boldsymbol{p}_{\mathrm{T}}^{\ell\ell} + \boldsymbol{p}_{\mathrm{T}}^{\nu\nu}|^2,\end{aligned} \tag{7.4}$$

where $E_{\mathrm{T}}^{\ell\ell} = \sqrt{|\boldsymbol{p}_{\mathrm{T}}^{\ell\ell}|^2 + m_{\ell\ell}^2}$, and $\boldsymbol{p}_{\mathrm{T}}^{\ell\ell}$ ($\boldsymbol{p}_{\mathrm{T}}^{\nu\nu}$) is the vector sum of the lepton (neutrino) momentum. In this expression, all dilepton terms are measureable, since the lepton 4-momenta are fully reconstructed. The experimental proxy for $\boldsymbol{p}_{\mathrm{T}}^{\nu\nu}$ is $E_{\mathrm{T}}^{\mathrm{miss}}$ (Section 7.1.5). To measure $E_{\mathrm{T}}^{\nu\nu}$, it is necessary to choose a convention for the invariant $m_{\nu\nu}$. If $m_{\nu\nu} = 0$, the m_{T} definition has the desirable property that it is bounded above by m_H if the Higgs is produced at rest. In practice, this edge in the distribution is smeared out due to non-zero Higgs p_{T} and detector resolution effects. Choosing this convention, the experimental m_{T} is given by the equation

$$m_{\mathrm{T}} = \sqrt{(E_{\mathrm{T}}^{\ell\ell} + E_{\mathrm{T}}^{\mathrm{miss}})^2 - |\boldsymbol{p}_{\mathrm{T}}^{\ell\ell} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}|^2}. \tag{7.5}$$

The shape of the m_{T} distribution is illustrated in Figure 7.6 for a $m_H = 125$ GeV Higgs sample. At this Higgs mass, there is little signal above $m_{\mathrm{T}} = 150$ GeV. Non-resonant WW and $t\bar{t}$, on the other hand, have broad m_{T} distributions that extend well beyond the edge of the signal distribution.

7.3.2 VBF Topology

The remaining five BDT inputs isolate the region of phase space associated with the VBF topology. The defining characteristic of VBF is two "tag" jets with a large rapidity gap. The rapidity distance between the two leading jets, $\Delta Y_{jj} = |Y_{\text{jet1}} - Y_{\text{jet2}}|$, is therefore used as a BDT input. As shown in Figure 7.6, signal peaks at $\Delta Y_{jj} \sim 4.5$, while background processes peak at $\Delta Y_{jj} \sim 2$. The dijet invariant mass, m_{jj}, is also a BDT input. This quantity is a function of ΔY_{jj}:

$$\begin{aligned} m_{jj}^2 &= m_{j1}^2 + m_{j2}^2 + 2\left[E_{\mathrm{T}}^{j1} E_{\mathrm{T}}^{j2} \cosh\left(\Delta Y_{jj}\right) - \boldsymbol{p}_{\mathrm{T}}^{j1} \cdot \boldsymbol{p}_{\mathrm{T}}^{j1}\right] \\ &\simeq 2 p_{\mathrm{T}}^{j1} p_{\mathrm{T}}^{j2} \left[\cosh\left(\Delta Y_{jj}\right) - \cos\left(\Delta\phi_{jj}\right)\right]. \end{aligned} \tag{7.6}$$

The second equality assumes that the mass of each jet is small with respect to the jet p_{T}. Given that m_{jj} grows as $\cosh\left(\Delta Y_{jj}\right)$, signal has large m_{jj} with respect to background. In spite of the high degree of correlation between ΔY_{jj} and m_{jj}, m_{jj} contributes discriminating power through the other terms in Equation 7.6, the coefficient $p_{\mathrm{T}}^{j1} p_{\mathrm{T}}^{j2}$ and the $\cos\left(\Delta\phi_{jj}\right)$ term. Jets from $t\bar{t}$ tend to be back-to-back, while VBF jets are fairly uniformly distributed in ϕ. Therefore, on average, at a fixed m_{jj}, ΔY_{jj} is smaller for $t\bar{t}$. Moreover, the BJV favors $t\bar{t}$ events with an ISR jet and a b-jet outside of the tracking volume, resulting in a softer jet p_{T} spectrum for $t\bar{t}$. The coefficient is therefore smaller on average, shifting the m_{jj} distribution to lower values.

Another BDT input that enhances the VBF signal is $p_{\mathrm{T}}^{\mathrm{tot}}$, the modulus of the vector sum of all of the physics objects in the event:

$$p_{\mathrm{T}}^{\mathrm{tot}} = |\boldsymbol{p}_{\mathrm{T}}^{l1} + \boldsymbol{p}_{\mathrm{T}}^{\mathrm{l2}} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} + \sum \boldsymbol{p}_{\mathrm{T}}^{\mathrm{jets}}|. \tag{7.7}$$

The sum over jets runs over all jets that pass the jet selection criteria described in Section 7.1. Track $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is used in $p_{\mathrm{T}}^{\mathrm{tot}}$, and because this version of $E_{\mathrm{T}}^{\mathrm{miss}}$ is merely

the vector sum of all of the physics objects in the event with an additional soft track term, there is cancellation in the above expression, resulting in

$$p_{\mathrm{T}}^{\mathrm{tot}} \simeq |\boldsymbol{E}_{\mathrm{T,soft}}^{\mathrm{miss,trk}}|. \tag{7.8}$$

This quantity is the modulus of the vector sum of soft tracks falling outside of selected jets, which probes the amount of soft QCD radiation in a given event. Backgrounds with QCD jets have more soft gluon radiation and lie at larger values of $p_{\mathrm{T}}^{\mathrm{tot}}$ with respect to VBF events which have relatively little QCD activity.

Lepton eta centrality (η_{lep} centrality) is an extension of the OLV discussed in Section 7.2.2. It is a measure of the centrality of the leptons with respect to the two tag jets, and is defined by the following equations:

$$\eta_{l0}\,\mathrm{cent.} = 2 \cdot |\frac{\eta_{l_0} - \bar{\eta}}{\eta_{j_0} - \eta_{j_1}}|$$

$$\eta_{l1}\,\mathrm{cent.} = 2 \cdot |\frac{\eta_{l_1} - \bar{\eta}}{\eta_{j_0} - \eta_{j_1}}|$$

$$\eta_{\mathrm{lep}}\,\mathrm{centrality} = \eta_{l0}\,\mathrm{cent.} + \eta_{l1}\,\mathrm{cent.}. \tag{7.9}$$

where $\bar{\eta} = (\eta_{j_0} + \eta_{j_1})/2$. As defined above, for a given lepton, the quantity $\eta_l\,\mathrm{cent.}$ is zero if the the lepton falls directly in between the tag jets, is less than one if the lepton falls at some η value between the tag jets, and is greater than one if it falls outside of the tag jets. The OLV, which requires that one lepton falls outside of the pseudorapidity gap, is equivalent to applying the cuts $\eta_{l0}\,\mathrm{cent.} < 1$ and $\eta_{l1}\,\mathrm{cent.} < 1$. After pre-selection, η_{lep} centrality is therefore constrained to the range [0,2]. Because leptons in top processes fall closer to the b-jets, the top background events that do survive the OLV have leptons with $\eta_l\,\mathrm{cent.}$ closer to one. The more central VBF leptons fall closer to $\bar{\eta}$ and η_{lep} centrality is smaller, as illustrated in Figure 7.6.

The final BDT input is the sum of the lepton-jet invariant masses for all lepton-jet pairs, denoted $\sum_{\ell,j} M_{\ell j}$. The sum runs over the two tag (leading) jets, and not the other selected jets. Like η_{lep} centrality, this observable is designed to exploit the fact that VBF jets fall at high pseudorapidity, while the Higgs decay leptons are central. This large opening angle between the leptons and jets results in large $M_{\ell j}$ terms compared to background.

The BDT input distributions are shown for signal and background for the $e\mu/\mu e$ ($ee/\mu\mu$) channel in Figure 7.6 (7.7).

7.3.3 Input Performance

The eight BDT inputs described above have been found to be the most discriminating kinematic distributions from an initial set of 27 (see Section A.4 for a description of the BDT optimization). The relative importance of each BDT input is quantified with a metric that counts the number of times a BDT input is used to split a decision tree node, weighting each count by the square of the gain in separation. The separation gain is given by the Gini index, defined in Equation A.1. The input ranking is shown in Table 7.6.

Table 7.6: BDT input ranking. The "Score" is the weighted sum of the number of times a given input is used to split a decision tree node, with the weight factor quantifying the separation gain for a given split.

Rank	BDT input	Score
1	m_{T}	0.26
2	m_{jj}	0.20
3	$\Delta\phi_{\ell\ell}$	0.16
4	$m_{\ell\ell}$	0.14
5	ΔY_{jj}	0.13
6	$\sum_{\ell,j} M_{\ell j}$	0.07
7	η_{lep} centrality	0.02
8	$p_{\text{T}}^{\text{tot}}$	0.01

7.4 BDT Validation

7.4.1 Training

The BDT training procedure, described in Appendix A, is performed after the common preselection and the BJV. For $ee/\mu\mu$ events, the two $E_{\mathrm{T}}^{\mathrm{miss}}$ cuts are also applied (track $E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV and calo $E_{\mathrm{T}}^{\mathrm{miss}} > 45$ GeV). The CJV, OLV, and $Z \to \tau\tau$ veto are not applied for training as these cuts considerably decrease the background sample size, making it difficult for the training algorithm to learn the kinematics of the backgrounds. Moreover, a smaller training sample is more susceptible to overtraining, a phenomenon in which the BDT has learned the statistical or systematic fluctuations of the training sample and therefore does not generalize to statistically independent samples.

The MC samples outlined in Table 6.1 are used in the training of the BDT. In order to use all of the available MC statistics in the analysis, a cross-validation approach is used. The training set is split into two statistically independent samples—even events and odd events—and a separate BDT is trained for each sample, yielding two BDT functions: $F_{\mathrm{even}}(\boldsymbol{x})$ and $F_{\mathrm{odd}}(\boldsymbol{x})$. For a given event, the BDT score is obtained by applying the BDT that was trained on the orthogonal training set, i.e.

$$\mathrm{BDT}(\boldsymbol{x}_i) = \begin{cases} F_{\mathrm{even}}(\boldsymbol{x}_i) & : \text{event i is odd} \\ F_{\mathrm{odd}}(\boldsymbol{x}_i) & : \text{event i is even} \end{cases} \tag{7.10}$$

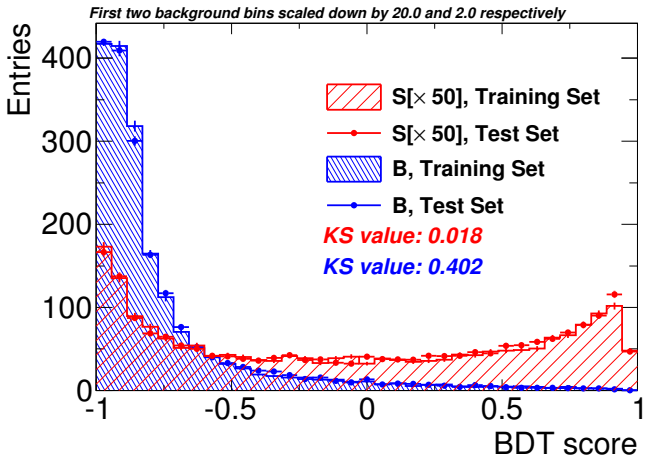
The background training sample is composed of all of the processes listed in Table 6.1, with the exception of W+jets and QCD. These two data-driven backgrounds are neglected, as they include events with negative event weights, which are ignored in the BDT implementation. The only Higgs process considered to be signal in training is VBF. Given that the aim of the analysis is to probe the signal strength of VBF, gluon fusion Higgs production is treated as a background. Moreover, despite the fact

that the VH processes are considered signal in the likelihood fit, in the BDT training they are ignored. Due to the negligible contribution of these processes in the $\ell\nu\ell\nu$ channel, the impact of removing them in the training is small.

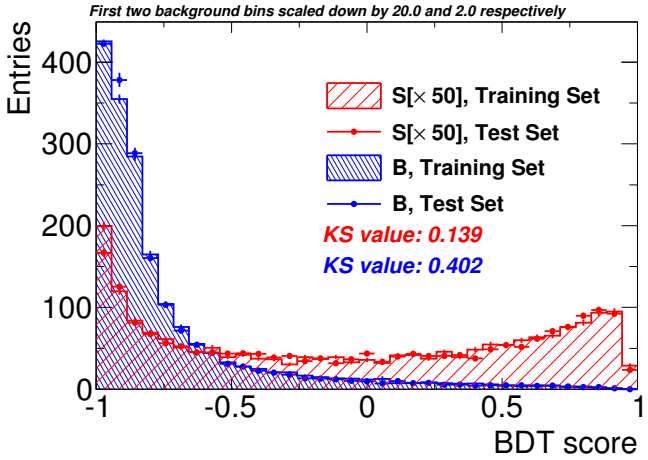
The level of overtraining for the BDT is evaluated by plotting the BDT response distribution $F(\boldsymbol{x})$ for the two statistically independent samples. Disagreement in the shape is evidence of overtraining. A comparison of the $F_{\text{even}}(\boldsymbol{x})$ distributions for the training set (even events) and the test set (odd events) is shown in Figure 7.8(a), and $F_{\text{odd}}(\boldsymbol{x})$ is shown in Figure 7.8(b). Signal, shown in red, peaks close to a BDT score of 1, and background in blue peaks at -1. The event yields are scaled to luminosity, with signal enhanced by a factor of 50. Because background peaks so sharply at -1, the first two background bins are scaled down by 20 and 2, respectively. A Kolmogorov-Smirnov goodness-of-fit test is used to quantify the level of agreement of the BDT response between the two subsamples. In this test, the null hypothesis is that the two distributions have the same underlying shape. For $F_{\text{even}}(\boldsymbol{x})$, the signal (background) p-value is 0.02 (0.40) and for $F_{\text{odd}}(\boldsymbol{x})$, the p-values are 0.14 (0.40). These p-values indicate that there is some degree of overtraining for the signal. The discrepancy between the two subsamples lies at low BDT values, a phase space region that is ultimately cut away in the analysis, as it does not contribute any sensitivity. Moreover, because the same BDT function is applied to the signal prediction, background prediction, and the actual collision data, overtraining will not invalidate the statistical results of an analysis; it can only result in a suboptimal BDT.

7.4.2 Data-MC Comparisons

The BDT training relies on MC simulation to model the shapes of, and the correlations among, the BDT inputs. In order to validate the modeling, the level of agreement between data and MC in the BDT input distributions is quantified in



(a) Overtraining check for BDT 1



(b) Overtraining check for BDT 2

FIGURE 7.8: Overtraining check for the optimized boosted decision tree. Training sets are shown with hatched distributions, while the statistically independent test sets are shown with points. The first two background bins are scaled down by factors of 20 and 2, respectively, in order to illustrate agreement at high BDT values. Moreover, the signal normalization is multiplied by a factor of 50.

signal-depleted validation regions. There are three such regions in this analysis: a low BDT region, a top-rich region, and a Z/DY-rich region. The latter two are discussed in Sections 8.1 and 8.2, respectively.

The low BDT validation region has the same preselection cuts as the signal region, with an additional cut on the BDT score—BDT < -0.48. The cut value

for this region has been determined through a binning optimization algorithm for the BDT distribution. Binning optimization is performed on the BDT distribution because the binned likelihood fit is performed on this distribution. Hence, the statistical sensitivity of the analysis is highly sensitive to the choice of binning. A crucial consideration in the optimization of binning is over-fitting on statistical fluctuations. If the metric for statistical sensitivity is not prudently chosen, sensitivity can be arbitrarily inflated with ever-finer binning, as backgrounds fluctuate out of signal-rich bins. Moreover, if there is over-fitting in the binning optimization, and bins become entirely depleted of background, then systematic uncertainties are not properly accounted for in the fit algorithm, again resulting in an optimistic estimate of the statistical sensitivity.

The binning optimization algorithm for the BDT distribution has been chosen to account for the above considerations. First, the preselection cuts are applied and the BDT distribution is computed for signal at $m_{\mathrm{H}} = 125$ GeV, as well as all of the backgrounds. Because BDT peaks sharply at one for signal, the algorithm starts at the right side of the BDT distribution, and in BDT steps of 0.02, integrates to the left. At each step, the Poisson significance estimate, given by

$$Z_{\mathrm{Pois}}(S, B) = \sqrt{2((S + B)\log(1 + S/B) - S)}, \tag{7.11}$$

where S (B) is the signal (background) event count normalized to luminosity, is computed. When a maximum in $Z_{\mathrm{Pois}}(S, B)$ is reached, the event yield for each background is checked, and if each background is represented, a bin boundary is set at that BDT value. The next iteration then begins at the new boundary until another maximum is found. The procedure continues until the maximum $Z_{\mathrm{Pois}}(S, B)$ falls below some threshold as the integration nears the background-dominated low BDT region. The significance curves obtained in this procedure are shown in Figure 7.9.

The resulting bin boundaries are -0.48, 0.3, and 0.78. The region BDT < -0.48 is dominated by background (in $e\mu/\mu e$, $B \sim 660$), with little signal (in $e\mu/\mu e$, $S \sim 5$), and is consequently not included in the signal region. Instead, this region is used to validate the modeling of BDT inputs.

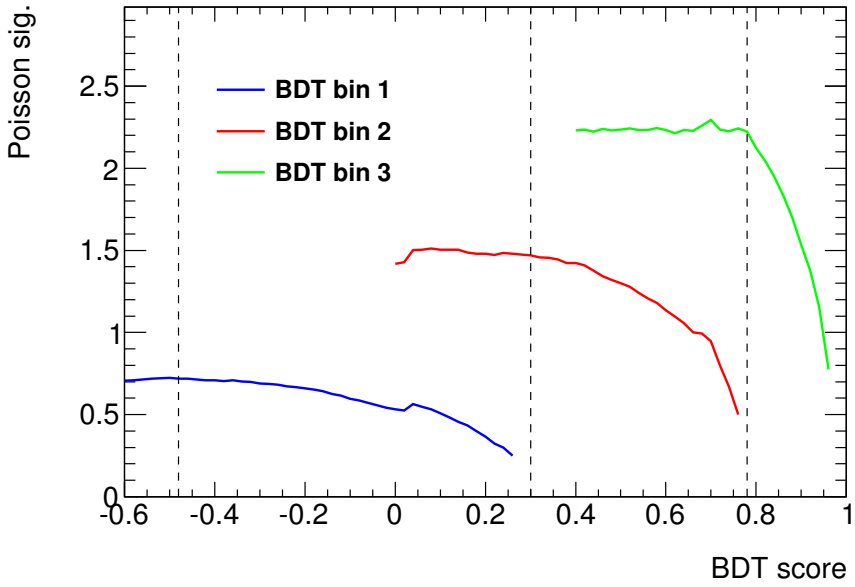


FIGURE 7.9: Poisson significance scans used in the BDT bin optimization procedure. The resulting bin boundaries are shown as dashed lines: [-0.48,0.3,0.78]. Events falling in the region BDT < -0.48 are not included in the SR. Instead, this region is used to validate the modeling of the BDT inputs.

Data-MC comparisons for the eight BDT inputs in the BDT < -0.48 validation region (VR) are shown in Figure 7.10. Despite the fact that there are not any data-driven corrections applied to the MC predictions in these plots, the MC models the data well in this region, as indicated by the p-value from the KS test for each BDT input. The lowest p-value, for η_{lep} centrality, is 0.38—a discrepancy of less than one sigma.

In addition to the one-dimensional BDT input distributions, the modeling of the correlations among the BDT inputs has been investigated. This is accomplished by

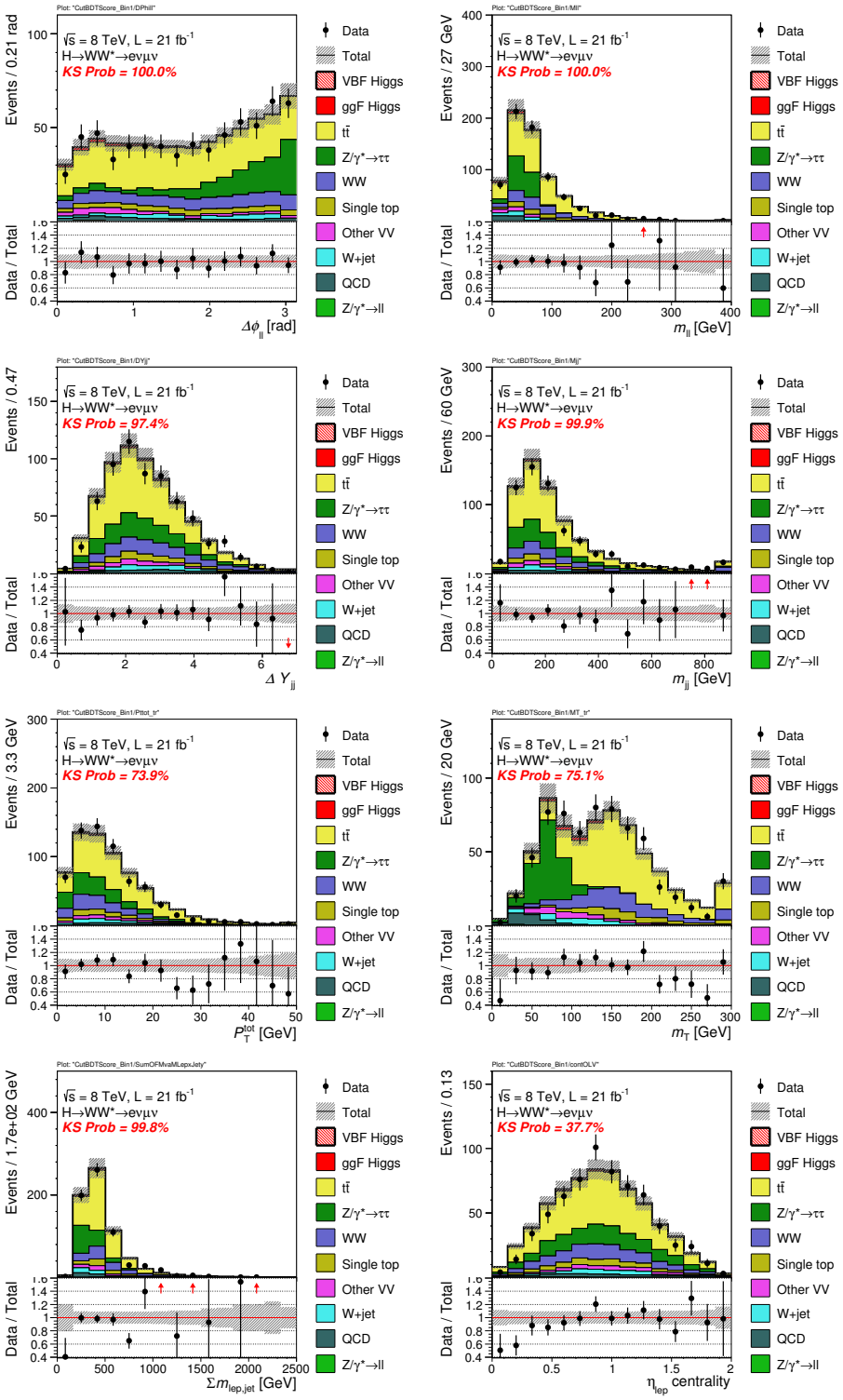


FIGURE 7.10: Distributions of the eight BDT inputs $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, ΔY_{jj}, m_{jj}, $p_{\mathrm{T}}^{\mathrm{tot}}$, m_{T}, $\sum_{\ell,j} M_{\ell j}$, and η_{lep} centrality in the $e\mu/\mu e$ validation region (BDT score < -0.48). Error band represents statistical uncertainties. Data-driven corrections to $t\bar{t}$ and Z/DY are not applied.

plotting the average value of the i^{th} BDT input against the j^{th} input for all possible pairs of BDT inputs. The resulting matrix of plots is shown in Figure 7.11 for the low BDT VR. In this region, the MC models the correlations observed in data well. Mis-modeling in the correlations will also manifest as data-MC discrepancies in the BDT response distribution.

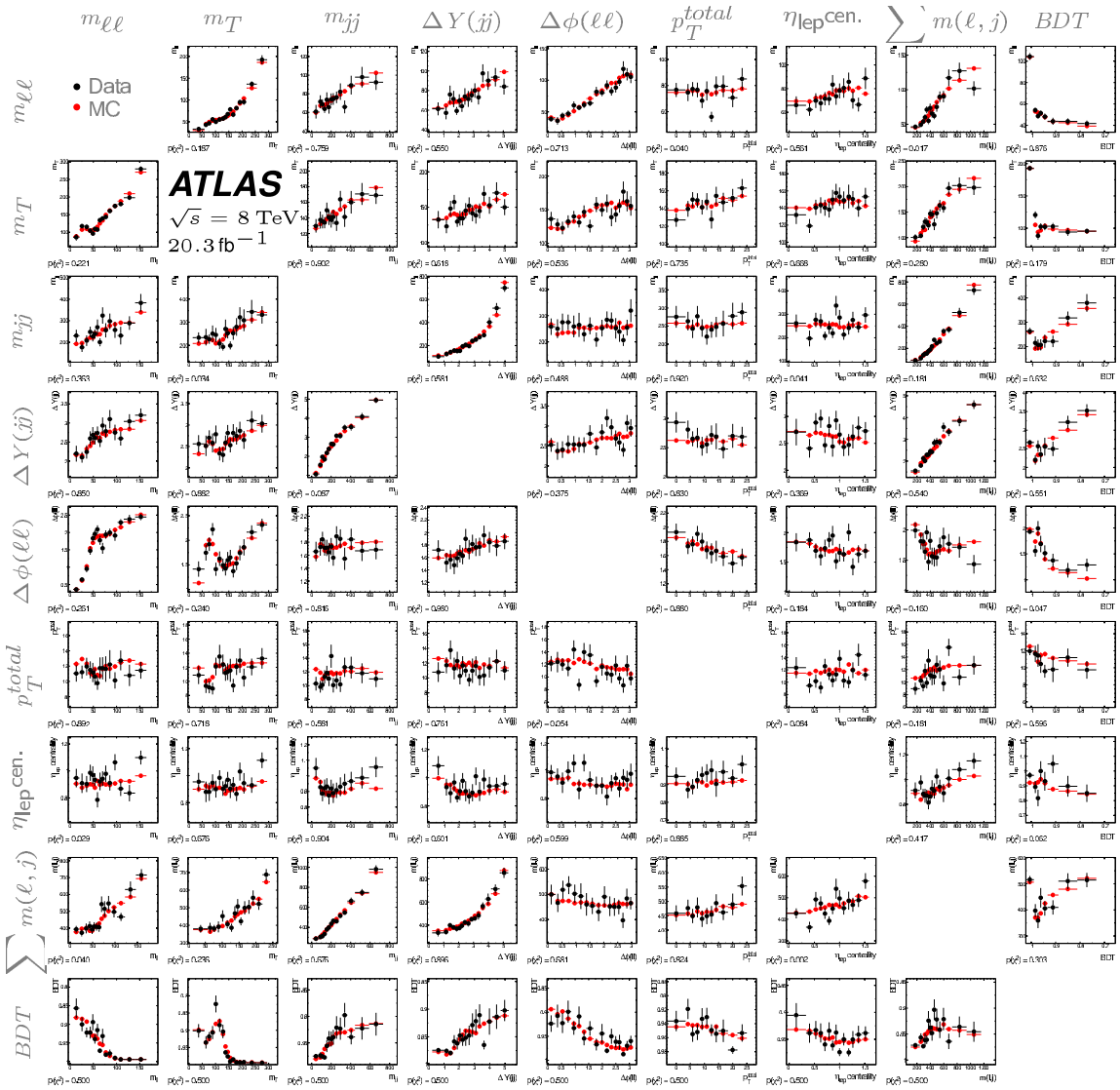


FIGURE 7.11: Correlation plots of BDT inputs in the low BDT VR. Distributions of $< X_i >$ vs X_j are shown for each BDT input pair. Data is shown in black, while the MC prediction for the background is in red.

8

Background Estimates

As discussed in Section 6.2, background predictions are derived from MC simulation. In some cases, the accuracy of the prediction from simulation can be improved upon by incorporating information from the data collected in the ATLAS detector. These "data-driven" approaches are summarized in the following section.

8.1 Top Quark Processes

In spite of the high level of agreement observed between theory calculations and ATLAS top quark measurements, the top quark background normalization is constrained using a top-rich control region (CR). The cross section measurements in ATLAS require the selected jets to be central ($|\eta| < 2.5$) [100, 105], whereas in the VBF analysis, due to the signal topology, the η requirement is looser, $|\eta| < 4.5$. Dedicated top quark measurements have yet to probe this region to test existing theoretical models, motivating the use of a data-driven top background estimate. Because the kinematic shapes are similar for $t\bar{t}$ and ST, these processes share a common normalization. All of the selection cuts applied in the SR are also applied in the top CR, with the exception of the BJV. Instead of requiring zero b-tags, exactly

one b-tag is required, enriching this region with top background and minimizing contamination from sources without heavy flavor quarks in the final state. The top CR is subdivided into BDT bins, and the top normalization is computed separately in each bin. This approach is used because the BDT spans a large and relatively unknown region of phase space, and there is no *a priori* reason to assume that the normalization is constant across the BDT spectrum.

The normalization factor (NF) in each BDT bin is the ratio of the number of top background events in the CR in data to number of top background events predicted by MC, expressed as

$$\tau_i = \frac{N^{CR}_{data,i} - N^{CR}_{non-top,MC,i}}{N^{CR}_{top,MC,i}} \tag{8.1}$$

where the index i represents the i^{th} BDT bin. The top estimate in the SR is obtained by scaling the MC prediction by the NF from the CR:

$$\begin{aligned} N^{SR,est}_{top,i} = \tau_i N^{SR}_{top,MC,i} &= \frac{N^{SR}_{top,MC,i}}{N^{CR}_{top,MC,i}}(N^{CR}_{data,i} - N^{CR}_{non-top,MC,i}) \\ &\equiv \alpha_i (N^{CR}_{data,i} - N^{CR}_{non-top,MC,i}) \end{aligned} \tag{8.2}$$

In this expression, the only quantities derived from MC simulation are the extrapolation factor α and the number of non-top events in the CR. Systematic uncertainties on top quark backgrounds are therefore assessed by computing the change in α under the variation of a given systematic source. The uncertainty associated with $N^{CR}_{non-top,MC,i}$ is included as an auxiliary term in the fit, but given the low level of non-top quark backgrounds in the top CR, such uncertainties have little impact.

The statistical uncertainty on the top NF is approximately $\sqrt{N^{CR}_{data}}/N^{CR}_{top,MC}$ which is just $1/\sqrt{N^{CR}_{data}}$ for a sufficiently pure top CR. More data events in the CR corresponds to a lower statistical uncertainty on the top NF. Motivated by this, the top

Table 8.1: Event yields for $t\bar{t}$, single top, non-top backgrounds, and data in the top control region, shown at each cut stage starting with $N_{b\text{-jet}} = 1$. The corresponding top normalization factor is shown in blue.

	$t\bar{t}$	ST	non-top	data	top NF
$N_{\text{jet}} \geq 2$	64944	3571	9895	80986	1.04 ± 0.004
$N_{b\text{-jet}} = 1$	21952	1865	2586	27791	1.06 ± 0.007
CJV	16455	1521	1912	21071	1.07 ± 0.008
OLV	3284 ± 7	305 ± 2	344 ± 10	4183	1.07 ± 0.02
$Z \to \tau\tau$ veto	2387 ± 6	217 ± 1	184 ± 7	2964	1.07 ± 0.02
BDT bin 0	2305 ± 6	204 ± 1.259	169 ± 7	2807	1.05 ± 0.02
BDT bin 1	71.9 ± 1.0	10.4 ± 0.4	13.1 ± 1.1	143	1.58 ± 0.15
BDT bins 2+3	10.1 ± 0.4	2.1 ± 0.2	2.4 ± 0.4	14	0.95 ± 0.31

CR is merged for the two flavor channels $ee/\mu\mu$ and $e\mu/\mu e$. Also, since the BDT is trained to efficiently reject top quark background, the two high BDT bins are depleted of this process in the b tag region. To improve the statistical precision, these two bins are merged in the top CR and a common NF is applied to the two high BDT SR bins. The event yields and the corresponding top NF are shown at each cut stage in the top CR in Table 8.1. Distributions of the BDT inputs in the top CR after the $Z \to \tau\tau$ veto and without any cut on BDT score are shown in Figures 8.1 and 8.2. The error bars shown are MC statistical uncertainties only. These plots illustrate that Powheg+Pythia models $t\bar{t}$ remarkably well in this phase space region. There is some degree of mis-modeling in the $\Delta\phi_{\ell\ell}$ distribution at low values in both $e\mu/\mu e$ and $ee/\mu\mu$ channels. This feature is covered by the uncertainties on the top quark background.

As mentioned above, since the extrapolation factor $\alpha = \frac{N^{SR}_{top,MC}}{N^{CR}_{top,MC}}$ is determined by MC, the theoretical uncertainties on top quark backgrounds are evaluated by measuring the variation of this ratio. These uncertainties are binned in BDT score, accounting for the fact the extrapolation is done separately in each BDT bin. Four sources of uncertainty are considered: the parton shower (PS) model, pdf, the QCD

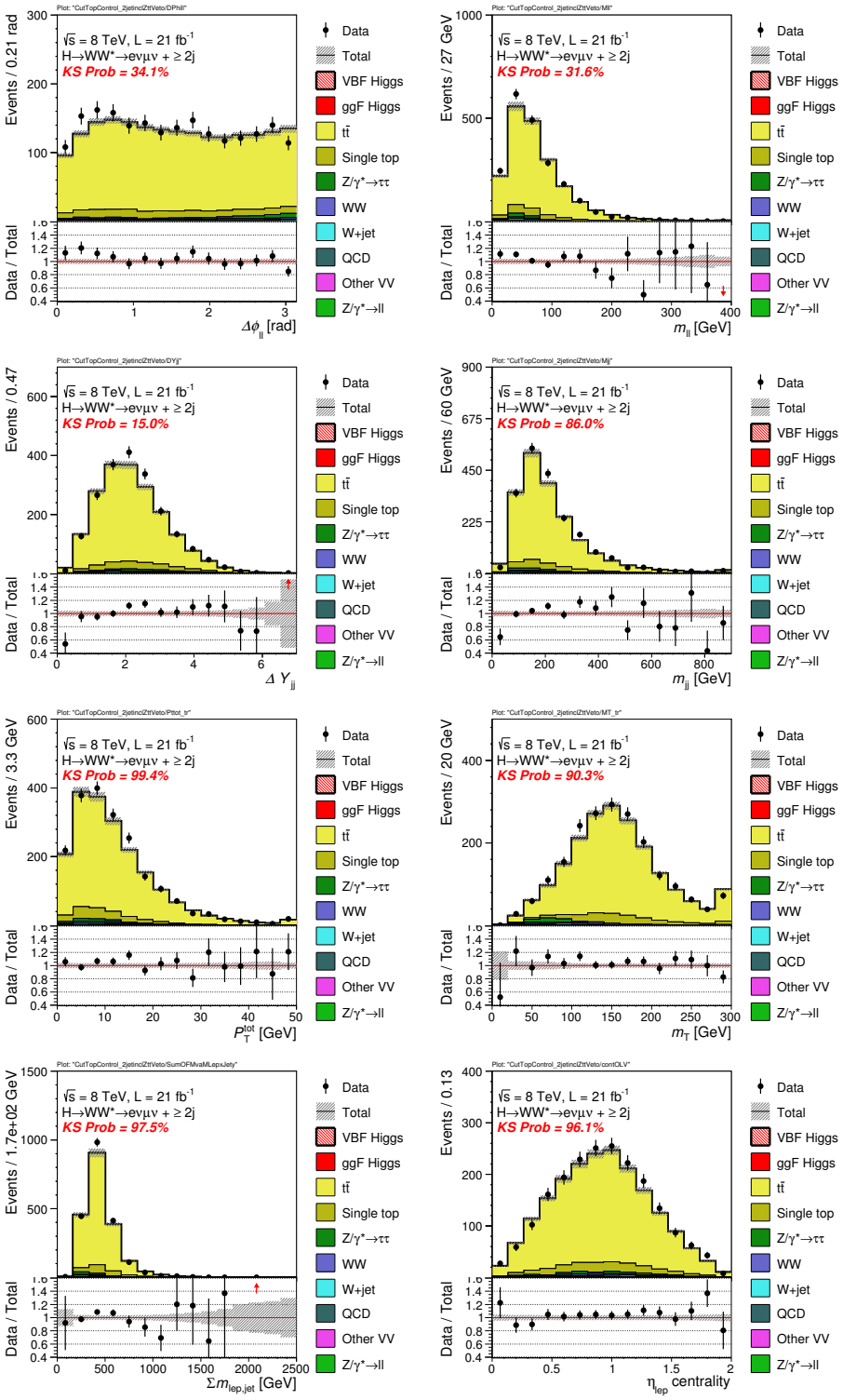


FIGURE 8.1: Distributions of the eight BDT inputs $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, ΔY_{jj}, m_{jj}, $p_{\mathrm{T}}^{\mathrm{tot}}$, m_{T}, $\sum_{\ell,j} M_{\ell j}$, and η_{lep} centrality in the $e\mu/\mu e$ top CR after the $Z \to \tau\tau$ veto. Error band represents statistical uncertainty. Data-driven corrections to $t\bar{t}$ and Z/DY are not applied.

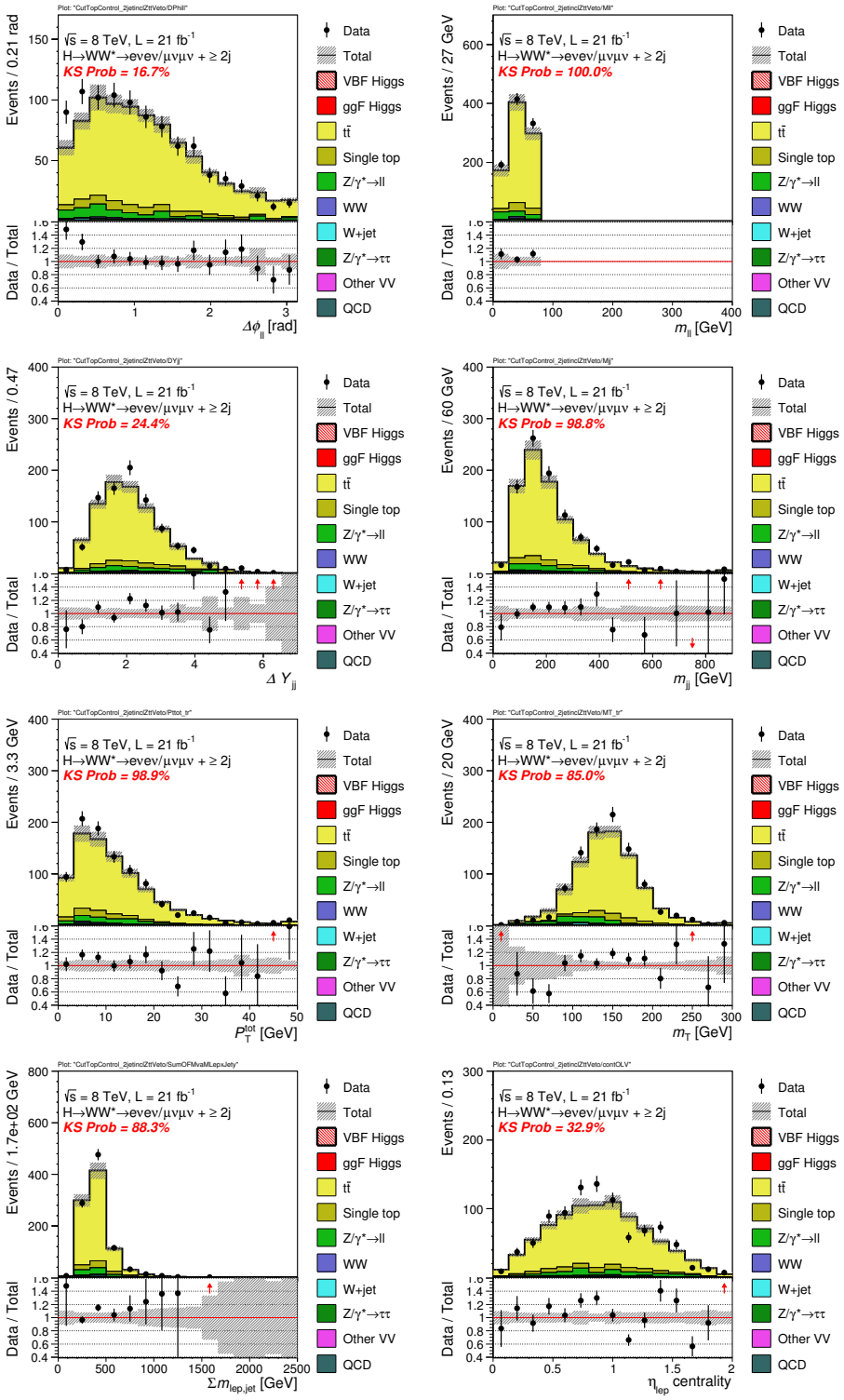


FIGURE 8.2: Distributions of the eight BDT inputs $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, ΔY_{jj}, m_{jj}, $p_{\mathrm{T}}^{\mathrm{tot}}$, m_{T}, $\sum_{\ell,j} M_{\ell j}$, and η_{lep} centrality in the $ee/\mu\mu$ top CR after the $Z \to \tau\tau$ veto. Error band represents statistical uncertainty. Data-driven corrections to $t\bar{t}$ and Z/DY are not applied.

scale, and the matrix element model. For the PS modeling uncertainty, the default sample with PYTHIA showering is compared to the same hard scatter generator but with HERWIG showering. The differences in the BDT spectrum is $< 5\%$ in the low BDT bin and less than the statistical uncertainties in the high BDT bins. For the uncertainty from the choice of pdf, the CT10 error sets are used and the uncertainty is ~1% across the BDT spectrum. Uncertainties are of the same order if they are derived by instead comparing CT10 and NNPDF.

The remaining two sources of uncertainty are evaluated at truth level with the true momentum 4-vectors of the final state particles after showering and hadronization, but before interacting with the detector. Truth level quantities are defined to best reflect the corresponding observable, and are then used as BDT inputs, defining a "truth-level BDT". Truth-level charged leptons are required to be decay products of the W bosons from the top quark decays. Truth-level jets are built using the anti-k_t algorithm on truth hadrons instead of topo-clusters, with the same p_T and η cuts used in the selection of reconstruction level jets. Track $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined using tracks from charged, interacting particles that fall within the tracking volume, $|\eta| < 2.47$, where the track selection efficiency is modeled by a logistic turn-on in track p_T. To capture the reconstruction level track $E_{\mathrm{T}}^{\mathrm{miss}}$ definition (see Section 7.1.5), tracks that fall within a ΔR of 0.4 of a selected jet are excluded to avoid double-counting the transverse energy associated with jets. Truth-level $E_{\mathrm{T}}^{\mathrm{miss}}$ is then the vector sum of the lepton, jet, and soft track transverse momenta.

The tagging of b-hadron jets is captured at truth-level by weighting events according to the probability that the event has zero b-tags or one b-tag. This probability can be expressed in terms of the tagging efficiencies

$$w_0 \;=\; p(0\text{ tag}) \;=\; \prod_i 1-\epsilon_i(p_T,\eta,fl) \tag{8.3}$$

$$w_1 \;=\; p(1\text{ tag}) \;=\; \sum_i \epsilon_i(p_T,\eta,fl)\cdot\prod_{i\neq j}(1-\epsilon_j(p_T,\eta,fl)), \tag{8.4}$$

where ϵ_i, the tagging efficiency for the i^{th} jet, depends on the p_{T}, η, and flavor of the jet. These truth-level definitions have been validated against the corresponding reconstruction level quantities, implying that using the variation induced in the truth BDT is well-motivated for the extraction of uncertainties.

The QCD scale uncertainty was evaluated by independently varying the factorization (μ_F) and renormalization (μ_R) scales up and down by a factor of two in a $t\bar{t}$ sample generated with MC@NLO and showered with Herwig. The largest deviation from the nominal α value is taken as the uncertainty. This uncertainty ranges from 0% in the lowest BDT bin to 5% in the highest. However, in every bin, the uncertainty is statistically compatible with zero due to insufficient MC statistics.

The largest theoretical uncertainty on top is the matrix element modeling uncertainty. The difference in the predicted α value is computed for three generators: MC@NLO, Powheg, and Alpgen. The same PS program, Herwig, is used in each case. The largest difference, observed between MC@NLO and Alpgen, is assigned as the uncertainty. Figure 8.3 displays the ratio $\alpha_{MC@NLO}/\alpha_{\text{Alpgen}}$ binned in BDT. The deviation of this ratio from unity is taken as the uncertainty on the BDT shape—10%, 12%, and 21%, respectively.

8.2 $Z/\text{DY}\to ee/\mu\mu$ and $Z/\text{DY}\to\tau\tau$

In the $ee/\mu\mu$ channel, the dominant background, Z/DY, is significantly suppressed by requiring the events to fall at high values of $E_{\text{T}}^{\text{miss}}$. $Z/\gamma^*\to\ell\ell$ processes do not have "real" $E_{\text{T}}^{\text{miss}}$ from invisible final state particles. Instead, $E_{\text{T}}^{\text{miss}}$ arises due to the

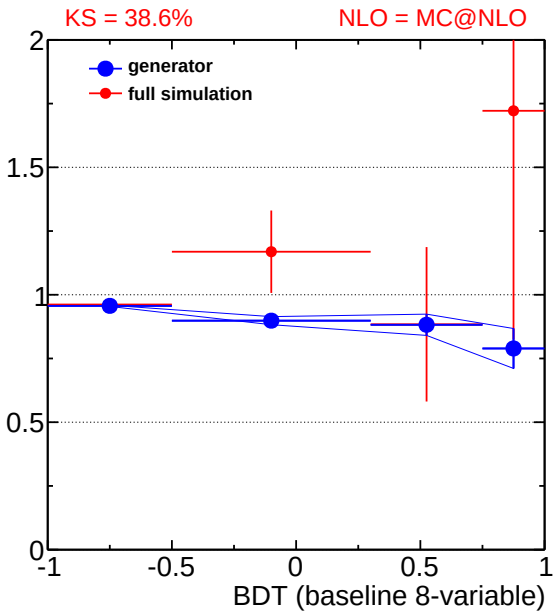


FIGURE 8.3: Ratio of α predicted by ALPGEN to that of MC@NLO. The ratio at truth level is shown in blue and reconstruction level with large statistical uncertainties is shown in red. The deviation from unity for the truth level ratio is taken as an uncertainty.

mis-measurement of leptons or jets, as well as soft QCD activity from the underlying event or pile-up. Such instrumental effects are difficult to model in MC simulation, motivating the use of data-driven approaches.

To correct for mis-modeling of calorimeter-based $E_{\mathrm{T}}^{\mathrm{miss}}$ in Z/DY MC, the efficiency of the $E_{\mathrm{T}}^{\mathrm{miss,CALO}} > 45$ GeV cut is taken from data in the Z peak. After applying the common preselection cuts and the track $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, the $m_{\ell\ell}$-$E_{\mathrm{T}}^{\mathrm{miss}}$ plane is partitioned into four orthogonal regions, labeled A, B, C, and D (Table 8.2). Regions C and D fall in the Z peak ($|m_{\ell\ell} - m_Z| < 15$ GeV), with region C at high $E_{\mathrm{T}}^{\mathrm{miss}}$ (> 45 GeV) and region D at low $E_{\mathrm{T}}^{\mathrm{miss}}$ ($25 <$ GeV $< E_{\mathrm{T}}^{\mathrm{miss}} < 45$ GeV). Regions A and B fall at high and low $E_{\mathrm{T}}^{\mathrm{miss}}$, respectively, but at low $m_{\ell\ell}$ ($m_{\ell\ell} < 75$ GeV). In this way, region A corresponds to the $ee/\mu\mu$ SR. Region B is rich in Z/DY, with ~5% contamination from other backgrounds, and data from this control region is used to

obtain the BDT template for Z/DY processes in the signal region. To extrapolate from region B to region A, the number of events is scaled by the $E_{\mathrm{T}}^{\mathrm{miss}}$ efficiency from regions C and D. The estimate in region A can be written

$$N_{\mathrm{DY},i}^{A,est.} = \frac{N_{\mathrm{DY,data}}^{C}}{N_{\mathrm{DY,data}}^{D}} \cdot N_{\mathrm{DY,data},i}^{B}, \tag{8.5}$$

where $N_{\mathrm{DY,data}}^{X}$ is the number of events measured in data in control region X with the non-DY component subtracted, i.e. $N_{\mathrm{DY,data}}^{X} = N_{\mathrm{data}}^{X} - N_{non-\mathrm{DY},MC}^{X}$. The index i labels the BDT bin. The $E_{\mathrm{T}}^{\mathrm{miss}}$ efficiency is not binned in BDT—i.e. the same ratio is applied for each bin. As for the top background estimate, due to limited statistics in the high BDT bins, the control regions for these bins are merged. Therefore, the relative rates in these two bins are taken from the MC prediction. The resulting NFs, defined as $N_{\mathrm{DY},i}^{A,est.}/N_{\mathrm{DY,i}}^{A,MC}$, are 1.0 ± 0.2 in the lowest BDT bin, and 0.9 ± 0.3 in the two highest bins.

Table 8.2: Summary of the regions used for the data-driven Z/DY estimate (ABCD method). Region A corresponds to the signal region. The BDT template is taken from data in region B, and the efficiency of the $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ cut is determined in the Z peak (regions C and D).

Region A (SR) $E_{\mathrm{T}}^{\mathrm{miss,CALO}} > 45$ GeV $m_{\ell\ell} < 75$ GeV	**Region C** $E_{\mathrm{T}}^{\mathrm{miss,CALO}} > 45$ GeV $\lvert m_{\ell\ell} - m_Z \rvert < 15$ GeV
Region B 25 GeV $< E_{\mathrm{T}}^{\mathrm{miss,CALO}} < 45$ GeV $m_{\ell\ell} < 75$ GeV	**Region D** 25 GeV $< E_{\mathrm{T}}^{\mathrm{miss,CALO}} < 45$ GeV $\lvert m_{\ell\ell} - m_Z \rvert < 15$ GeV

This so-called "ABCD method" is predicated on two assumptions. The first is that the BDT response is not correlated to calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$, allowing the BDT shape

template to be taken from the low $E_{\mathrm{T}}^{\mathrm{miss}}$ CR. The second implicit assumption is that the calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ cut efficiency is not correlated with $m_{\ell\ell}$. This assumption is needed to apply the low to high $E_{\mathrm{T}}^{\mathrm{miss}}$ extrapolation factor from the Z peak in the low $m_{\ell\ell}$ region. The validity of these two assumptions is empirically tested in MC, and any breakdown is accounted for in the assignment of systematics uncertainties.

The absence of a correlation between the BDT response and calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ is due to the fact that jet-corrected $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ is used in the only two BDT inputs that depend on $E_{\mathrm{T}}^{\mathrm{miss}}$, m_{T} and $p_{\mathrm{T}}^{\mathrm{tot}}$. The linear correlation coefficients between the BDT inputs and both calorimeter and track $E_{\mathrm{T}}^{\mathrm{miss}}$ are shown in 8.3. With the exception of m_{T}, the correlation coefficient between calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ and the BDT inputs is less than 0.1. To account for any correlation, the difference in the shape of the BDT template in regions A and B is taken as an uncertainty. This difference is computed in Alpgen+Herwig and Alpgen+Pythia, and the largest uncertainty between the two MC generators is assigned. The template comparisons are shown in Figure 8.4, with uncertainties corresponding to 4%, 10%, and 60% in bins of BDT.

Table 8.3: Linear correlation coefficients between BDT inputs and $E_{\mathrm{T}}^{\mathrm{miss}}$ quantities for Z/DY.

BDT input	$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$	$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$
m_{T}	0.14	0.31
$p_{\mathrm{T}}^{\mathrm{tot}}$	0.00	0.07
m_{jj}	0.08	0.11
$\sum_{\ell,j} M_{\ell j}$	0.08	0.10
$\Delta\phi_{\ell\ell}$	-0.04	0.02
$m_{\ell\ell}$	0.03	0.04
η_{lep} centrality	0.02	0.00
ΔY_{jj}	-0.02	0.05

The second assumption, that the calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ efficiency is not correlated to $m_{\ell\ell}$, is true for the leading order Z+2j processes. For higher order processes with

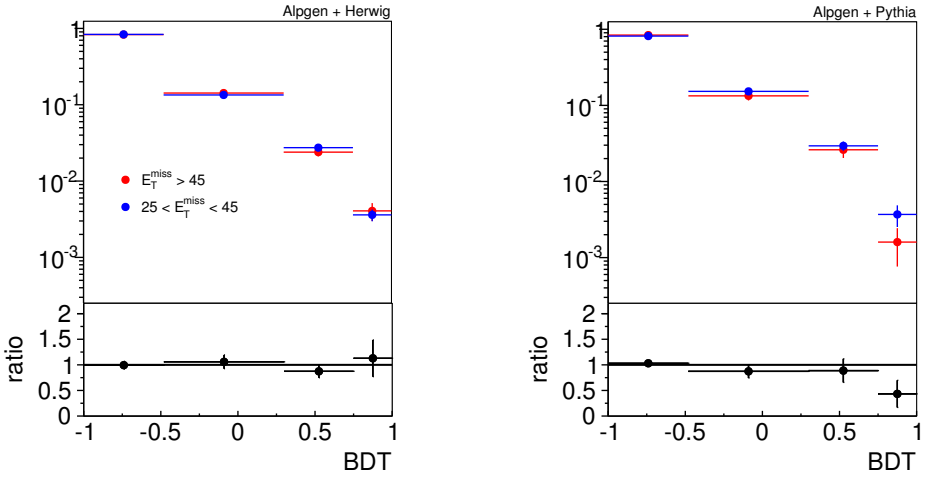


FIGURE 8.4: Comparison of BDT template for Z/DY in 25 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 45 GeV region (blue) and $E_{\mathrm{T}}^{\mathrm{miss}} > 45$ GeV region (red). The baseline Z/DY sample ALPGEN+HERWIG is shown on the left and ALPGEN+PYTHIA is on the right. The difference in template between low $E_{\mathrm{T}}^{\mathrm{miss}}$ region and high $E_{\mathrm{T}}^{\mathrm{miss}}$ region is assigned as an uncertainty on Z/DY.

initial state radiation (ISR) from QCD, the lepton-jet system recoils against the resulting hadronic activity. This soft hadronic is more susceptible to mis-measurement, translating to a larger calorimeter $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. Due to the nature of the pdfs of the incoming partons, increased ISR correlates to increased virtuality of the Z/γ^*, which, in turn, results in increased $m_{\ell\ell}$. Therefore, a correlation is induced between $E_{\mathrm{T}}^{\mathrm{miss}}$ and $m_{\ell\ell}$, and the efficiency of the $E_{\mathrm{T}}^{\mathrm{miss,CALO}} > 45$ GeV cut in the Z peak is expected to be greater than that in the $m_{\ell\ell} < 75$ GeV region. The degree to which the efficiencies disagree is referred to as the non-closure of the method, and is quantified by

$$f_{\text{non-closure}} = \frac{N_A/N_B}{N_C/N_D} \tag{8.6}$$

where the event yields are taken from Z/DY MC. The Z/DY estimate in the SR, given by Equation 8.5, is scaled by this factor $f_{\text{non-closure}}$ to account for the $E_{\mathrm{T}}^{\mathrm{miss}}$

efficiency difference. Moreover, the difference between $f_{\text{non-closure}}$ and unity (17%) is assigned as an uncertainty. The $E_{\text{T}}^{\text{miss}}$ efficiency in the Z peak (N_C/N_D), as measured in data, is 0.43 ± 0.03, which is constituent with the value from Z/DY MC of 0.47 ± 0.04. The statistical uncertainty on this efficiency is taken as an uncertainty.

The Z boson processes in which the Z decays to $\tau\tau$ contribute in both the $e\mu/\mu e$ and $ee/\mu\mu$ channels. The normalization of this background is taken from a control region that includes all of the preselection cuts with the exception of the $Z \to \tau\tau$ veto. Instead, $Z \to \tau\tau$ is enhanced by requiring that $|m_{\tau\tau} - m_Z| < 25$ GeV. Additionally, the cut $m_{\ell\ell} < 80$ GeV is applied to $ee/\mu\mu$ events, and to pick out a more signal-like phase space region, BDT > -0.48 is required. In order to increase the statistics in this region, thereby decreasing the statistical uncertainty on the NF, the $e\mu/\mu e$ and $ee/\mu\mu$ channels are merged for a common NF. The resulting NF is 1.2 ± 0.3. The same normalization is applied across the BDT spectrum due to limited $Z \to \tau\tau$ statistics at high BDT.

8.3 W+jets and QCD

Although W + jets and QCD multi-jet processes do not contain two leptons in the final state, they are expected to contribute in the signal region in cases where a jet is identified as a prompt lepton. Such leptons are considered "fake", whether they are mis-reconstructed from charged hadron tracks or actual non-prompt leptons from heavy hadron decays that are mis-identified. In contrast to the data-driven estimates described above, the procedure described in the following section is purely data-driven in the sense that both the normalization and the kinematic shapes are derived from a control region. In other words, W + jets and multi-jet MC templates are not used at all.

The W + jets CR is the same as the analysis signal region except that the object selection has been adjusted to enrich the region with fake leptons. This is accom-

plished by loosening the quality criteria for one of the leptons and requiring that it does not pass the analysis-level lepton selection, while the other lepton is required to pass. The former lepton type is referred to as "anti-identified". For anti-identified electrons, the calorimeter isolation requirement is loosened to $E_{\mathrm{T}}^{R=0.3}/E_{\mathrm{T}} < 0.30$, the track isolation requirement is $p_T^{R=0.3}/E_{\mathrm{T}} < 0.16$, the conversion flag and b-layer requirements are removed and the electron is required to fail the *medium* identification requirement. For muons, the calorimeter isolation requirements are also loosened, the track isolation requirements are completely removed, and the transverse impact parameter significance cut is removed.

To extrapolate from the $W+$jets CR to the SR, first the non-$W+$jets background is subtracted, and then an extrapolation factor, called the fake factor, is applied. The fake factor is defined as

$$f_{\mathrm{fake}} = \frac{N_{\mathrm{id}}}{N_{\mathrm{anti\text{-}id}}} \tag{8.7}$$

where N_{id} is the number of jets that pass the full lepton selection and $N_{\mathrm{anti\text{-}id}}$ is the number of jets which pass the anti-identified selection. f_{fake} is evaluated with jets in a jet-rich Z CR in bins of lepton p_{T} and η. From the fake factor, the $W+$jets estimate in the SR is given by

$$N_{id+id}^{W+\mathrm{jets}} = f_{\mathrm{fake}} \cdot (N_{\mathrm{id+anti\text{-}id}} - N_{\mathrm{id+anti\text{-}id}}^{\mathrm{QCD}} - N_{\mathrm{id+anti\text{-}id}}^{\mathrm{non\text{-}fake,MC}}) \tag{8.8}$$

where $N_{\mathrm{id+anti\text{-}id}}$ is the total number of events in the $W+$jets CR, $N_{\mathrm{id+anti\text{-}id}}^{\mathrm{QCD}}$ is data-driven QCD multi-jet estimate in the $W+$jets CR, which is described below, and $N_{\mathrm{id+anti\text{-}id}}^{\mathrm{non\text{-}fake,MC}}$ is the MC prediction for the non-fake contribution in the $W+$jets CR.

The dijet control region is similar to the $W+$jets control region, except that instead of the requirement of one anti-identified lepton, both of the leptons satisfy this requirement. The fake factors for this background are obtained using the same

procedure as W + jets, but in a region that is rich in QCD multi-jets. Because there are two fake leptons, two fake factors are derived, where the second fake factor accounts for bias introduced with the requirement of an additional anti-identified lepton. The QCD estimate in the signal region is then expressed as

$$N_{id+id}^{\text{QCD}} = f''_{\text{fake}} \cdot f'_{\text{fake}} \cdot (N_{\text{anti-id+anti-id}} - N_{\text{anti-id+anti-id}}^{W+\text{jets,MC}} - N_{\text{anti-id+anti-id}}^{\text{non-fake,MC}}) \tag{8.9}$$

where f''_{fake} and f'_{fake} are the two dijet fake factors, $N_{\text{anti-id+anti-id}}$ is the total number of data events in the dijet CR, $N_{\text{anti-id+anti-id}}^{W+\text{jets,MC}}$ is the W + jets contamination predicted by simulation in the CR, and $N_{\text{anti-id+anti-id}}^{\text{non-fake,MC}}$ is the non-fake contamination.

To estimate the QCD contamination in the W + jets CR ($N_{\text{id+anti-id}}^{\text{QCD}}$ in Equation 8.8), there is an extrapolation from the dijet CR to the W + jets CR, given by

$$N_{id+anti-id}^{\text{QCD}} = 2 \cdot f''_{\text{fake}} \cdot (N_{\text{anti-id+anti-id}} - N_{\text{anti-id+anti-id}}^{W+\text{jets,MC}} - N_{\text{anti-id+anti-id}}^{\text{non-fake,MC}}) \tag{8.10}$$

where the factor of two arises due to the fact that events with two anti-identified leptons can enter the W + jets CR if either lepton is identified.

The uncertainties associated with this data-driven estimate are relatively large. As mentioned above, the W + jets fake factor is obtained from a Z control region. In order to account for kinematic differences between jets in the Z control region and those associated with W + jets, fake factors from simulated W and Z samples are compared, and the resulting difference is used to correct the Z-derived fake factor from data, and also assigned as a systematic. For electron (muon) fakes, the uncertainty is 20% (22%). In addition, the dependence of the fake factor on the electroweak contamination in the Z control region is evaluated by computing the factor with and without these processes subtracted, yielding uncertainties in

the range 11%-25% (3%-21%) for electron (muon) fakes. Finally, the statistical uncertainty from the control region data is considered. For electron (muon) fakes, this uncertainty falls in the range 18%-52% (10%-34%). Uncertainties associated with the fake factor from the di-jet CR are evaluated similarly, and shown to be dominated by the uncertainty due to the CR sample choice of 60% (40%) for electron (muon) fakes.

Backgrounds due to fakes are relatively small in the BDT signal region. The primary reason for this is that the cross sections of these processes are small in the two jet bin relative to the other jet bins for which the above estimation technique has been developed. Moreover, jets from W + jets and QCD multi-jet processes tend to be more central in rapidity, and because the BDT selects forward jets, these processes are kinematically suppressed. The W + jets estimate in the $e\mu/\mu e$ channel relative to the total background prediction is 2%, 3%, and 0%, in the respective BDT fit bins and the QCD estimate is even smaller—0.4%, 0%, and 0%.

9

Systematic Uncertainties

All Monte-Carlo-based predictions rely on the best estimate for a set of parameters $\hat{\boldsymbol{\theta}}$. These parameters can be theoretical in nature, such as the cross section to which the prediction is scaled, or intrinsic to the ATLAS detector, such as the efficiency for selecting an electron. The uncertainties on these parameters are propagated to uncertainties on event yields in the signal and control regions, and as we will see in the following chapter, these uncertainties are ultimately propagated to an error on the measured VBF signal strength. Uncertainties are always computed at a working point corresponding to $\pm 1\sigma$ and propagated to the statistical fit as such.

9.1 Theoretical sources

Monte Carlo generators are approximate models for any scattering process. To assess theoretical uncertainties, approximations are varied, and the resulting change in the prediction is quantified. The following sources of uncertainty are considered: parton distribution function (pdf), QCD scale, parton showering program, hard-scattering program, and underlying event tune. Depending on the source, these uncertainties are split into two kinds: normalization and acceptance. Normalization uncertainties

are those that vary the overall normalization of a prediction without changing the underlying distributions. Acceptance uncertainties are computed for sources that change kinematic distributions and therefore selection efficiencies. Because the BDT spectrum spans a large and relatively unknown region of phase space, acceptance uncertainties are computed separately in each BDT fit bin.

Varying the factorization (μ_F) and renormalization scales (μ_R) probes sensitivity to the corrections from the next order in the perturbative expansion. For the VBF process, μ_F and μ_R are varied up and down by a factor of two, and the resulting change in event rates in the signal region is on the order of 0.5%. Since the scale change impacts the BDT input distributions associated with tag jets, namely m_{jj} and ΔY_{jj}, there is a distortion of the BDT spectrum of [1%,3%,3%] in BDT bins.

The pdf uncertainty is also evaluated for the VBF signal process. Comparing the baseline pdf CT10 to NNPDF [118], the difference in event yields in the BDT SR is found to be 4%, with a negligible dependence on the BDT score. The baseline hard-scattering generator is POWHEG which simulates the VBF process at NLO in QCD. Another generator, aMC@NLO, utilizes a different prescription for matching the HS and the PS, and in order to ascertain the uncertainty associated with the default prescription, the predictions of the two generators are compared. Although the shape variation of the BDT distribution is negligible, aMC@NLO predicts 4% more events than POWHEG in the BDT SR, which is assigned as a normalization uncertainty. Additionally, uncertainties associated with corrections from higher order EW corrections are assessed by turning these corrections on and off in the generator VBFNLO. These corrections are observed to have little impact on the shape of the BDT spectrum, and increase the overall cross section by 3.5%, a correction which is already accounted for in the NLO cross section calculation to which the baseline prediction is scaled (Chapter 6).

The dominant uncertainty on signal is associated with the parton shower program.

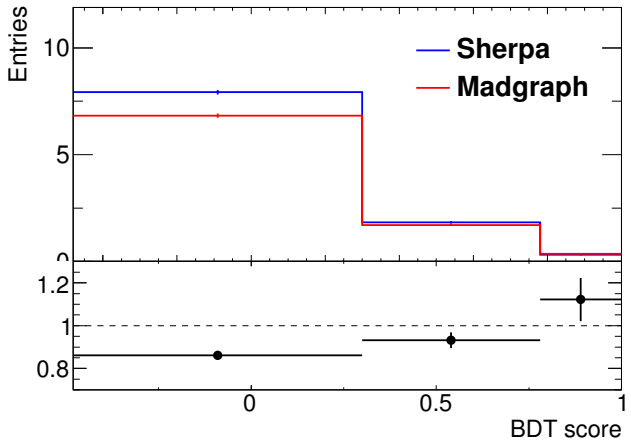
Comparing the baseline program, PYTHIA, to HERWIG, the difference in the event yield in the signal region is 3.4%. The distortion of the BDT spectrum is large due to the fact that the BDT input $p_{\mathrm{T}}^{\mathrm{tot}}$ is highly sensitive to the soft jet activity in an event. Though the effect is as large as 50% in the $p_{\mathrm{T}}^{\mathrm{tot}}$ distribution, in the BDT distribution the largest difference is 10% in the highest BDT bin, because the BDT selects the region of phase space associated with the core of the $p_{\mathrm{T}}^{\mathrm{tot}}$ distribution, where the effect is significantly smaller.

As discussed in Section 8.1, the theoretical uncertainties for $t\bar{t}$ and ST are computed by varying the pdf, QCD scale, parton shower, and the hard-scattering generator. Because this background is normalized with a control region, the change in the extrapolation factor $\alpha = \frac{N_{top,MC}^{SR}}{N_{top,MC}^{CR}}$ (Equation 8.2) is measured in order to assess the uncertainty on the signal region event rate. The dominant uncertainty source has been found to be the hard-scattering generator, derived from a comparison of MC@NLO and ALPGEN. In the respective BDT fit bins, these uncertainties are 10%, 12%, and 21%.

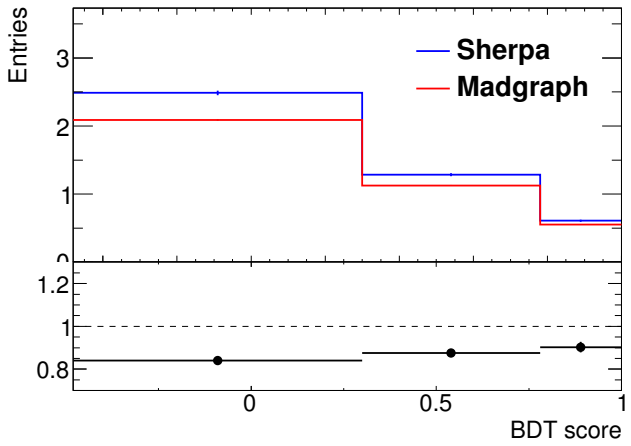
Non-resonant $WW + 2\mathrm{j}$ production is a large background, contributing ~19% of the total background in the $e\mu/\mu e$ SR and ~9% in the $ee/\mu\mu$ SR. QCD $WW+2\mathrm{j}$ rates are predicted with an inclusive SHERPA sample normalized to the NLO MCFM cross section. To assess the generator modelling and QCD scale uncertainties, MadGraph interfaced with PYTHIA for parton showering is used. Like the baseline sample, the MadGraph sample is inclusive, with up to three partons generated in the hard scattering. The final state partons are showered in PYTHIA with the MLM matching scheme [119, 120] at a scale of 20 GeV.

The generator modelling uncertainty is computed by comparing the truth level predictions of MadGraph to those of the SHERPA. In an approach similar to the one used in the evaluation of top theory uncertainties (see Section 8.1), the BDT inputs

are evaluated at truth level, and the resulting distortion of the BDT spectrum is taken as an uncertainty. For the comparison, the two truth level samples are scaled to the integrated luminosity with their respective cross sections, and the event yields for each generator are compared bin-by-bin in BDT score. This comparison is displayed in Figure 9.1(a). In light of the differences, uncertainties of 14%, 8%, and 12% are assigned.



(a) QCD WW



(b) EW WW

FIGURE 9.1: Comparison of MadGraph and SHERPA predictions in the $e\mu/\mu e$ BDT signal region for (a) QCD WW and (b) EW WW processes. Comparison is at truth-level, and the difference between the two generators is assigned as a modeling uncertainty.

Table 9.1: Summary of the theoretical uncertainties assigned to the QCD and EW WW background processes.

Source	QCD WW	EW WW
Generator modeling	(12%,8%,12%)	(16%,12%,10%)
QCD scale	27%	10%
pdf (σ)	4%	3%
pdf (acceptance)	2%	–
QCD-EW interference	–	2%
Higgs interference	–	1.2%

To assess the uncertainty from the choice of QCD scale, the factorization and renormalization scales in MadGraph, which are varied dynamically event-by-event, are coherently scaled up and down by a factor of two. The resulting samples are scaled to their respective post-parton-showering cross sections, and the signal region differences are computed. With the scale set to $\mu_R = \mu_F = \mu_{\text{default}}/2$, the predicted number of events in the SR is 27% higher than the default scale choice, while with $\mu_R = \mu_F = 2\mu_{\text{default}}$, the prediction falls by 2.6%. The larger of the two, 27%, is assigned as the uncertainty, which is consistent with the LO $\sigma(WW)$ uncertainty computed at $\sqrt{s}$= 7 TeV [121], though in a looser phase space region. Because the QCD scale variations induce small shape differences in the BDT input distributions, there is no shape component to this uncertainty. Additional QCD WW uncertainties are summarized in Table 9.1.

For EW WW processes, a hard-scattering modeling uncertainty is computed by comparing the prediction of Sherpa to that of an exclusive two jet MadGraph sample interfaced with Pythia. The cross section computed in MadGraph is 27.37 fb, 30% lower than that of Sherpa, resulting in fewer events predicted. A comparison of the absolute predictions for the two generators is shown in Figure 9.1(b), corresponding to uncertainties of 16%, 12%, and 10% in BDT bins. The QCD scale uncertainty

for EW WW processes is computed to be 10% at LO for a generic VBF phase space region [122]. It is assigned as a normalization uncertainty.

Remaining theory uncertainties for EW WW processes are shown in Table 9.1. The uncertainty from interference between EW WW and Higgs-mediated WW is computed by comparing the cross sections of WW and $H \to WW^{(*)}$ diagrams without interference to the corresponding cross sections with interference. At a Higgs mass hypothesis of $m_H = 125$ GeV, the uncertainty is 1.2%, while at $m_H = 150$ GeV, the uncertainty is 6.4%. Because these are small with respect to the modeling and QCD scale uncertainties, they are neglected in the fit. An uncertainty due to interference between QCD and EW WW processes is evaluated in a similar way and found to be ~2%. This, too, is neglected in the fit.

Gluon-gluon-fusion Higgs production (ggF) is a large background in the most sensitive BDT bins. In order to estimate the uncertainties from higher order terms in the perturbative expansion, the μ_F and μ_R are independently varied in MCFM. Since the final statistical result is derived from a likelihood across the jet bins, it is important to correctly correlate the normalization uncertainty for each jet bin. Moreover, applying a central jet veto effectively limits the phase space to the exclusive two jet bin, calling for the uncertainty to be computed for such a region. An approach that accounts for these considerations has been proposed by Stewart and Tackmann [123]. The exclusive cross section can be expressed as

$$\sigma_{2\mathrm{j}} = \sigma_{\geq 2\mathrm{j}} - \sigma_{\geq 3\mathrm{j}} \tag{9.1}$$

where $\sigma_{\geq 2\mathrm{j}}$ ($\sigma_{\geq 3\mathrm{j}}$) is the inclusive two (three) jet cross section. To a good approximation, the perturbative uncertainties on $\sigma_{\geq 2\mathrm{j}}$ and $\sigma_{\geq 3\mathrm{j}}$ can be considered to be uncorrelated [123]. These uncertainties are evaluated by varying μ_F and μ_R up and down by a factor of two and measuring the change in the cross section in two regions:

for the inclusive 2j bin, the CJV is not applied, and for the inclusive 3j bin, the CJV is inverted. In both regions, all SR cuts are applied apart from the CJV. The resulting relative uncertainties are $\Delta_{\geq 2j} = 22\%$ and $\Delta_{\geq 3j} = 118\%$, corresponding to a normalization uncertainty of $\Delta_{2j} = 34\%$ in the BDT signal region. In addition, the shape variation with QCD scale is computed and found to be [+2.4/-2.5,+2.6/-2.5,+13/-12]% in the BDT fit bins.

Uncertainties on ggF from other sources are small with respect to the QCD scale uncertainty. The pdf uncertainty is 8%. The underlying event and parton shower uncertainty, computed by comparing HERWIG showering to the baseline program POWHEG, is a flat 15% across the BDT bins.

For $Z/\mathrm{DY} \to ee/\mu\mu$ processes, because a data-driven approach is used, the only prediction from MC simulation is the relative rate in the last two BDT bins. Three sources of uncertainty on this quantity have been considered: (1) QCD scale, (2) parton shower, (3) pdf. Of these three sources, evaluated with SHERPA, the largest is QCD scale, corresponding to an uncertainty of 11% in the last BDT bin. The other two sources are neglected. In the $e\mu/\mu e$ channel, the $Z \to \tau\tau$ normalization is constrained from a control region, which, due to limited statistics in this region, results in a large statistical uncertainty of 30%. Due to the negligible contribution of this background in the most sensitive BDT bins and to the fact that the theory uncertainties are *a priori* assumed to be smaller than the statistical uncertainties, theory uncertainties are not computed for this background component.

The theory uncertainties for the remaining backgrounds, which are small in the SRs, will not be discussed here, as they have little impact on the statistical result.

9.2 Instrumental sources

Systematic uncertainty sources due to the reconstruction and identification of electrons, muons, jets, and missing transverse energy are also considered. These sources

are evaluated by varying the relevant parameter in simulation and then measuring the resulting change in the event yields in the BDT signal region. Because the uncertainties are generally binned in the p_{T} and η of the object, the event yield variations are evaluated separately in each BDT bin and correlated. To simplify the presentation of these uncertainties, they are averaged over the three BDT bins in the following discussion.

As discussed in Section 7.1, the efficiencies associated with lepton selection are corrected with data. The uncertainties on these scale factors, summarized in Table 7.2, are split into two uncorrelated components associated with the isolation SF and the identification SF. Propagated to the event yields in the SR, these two uncertainties are both less than 1% for signal and background (Tables 9.2 and 9.3). The remaining lepton uncertainties are due to the momentum scale and resolution calibrations, discussed in Chapter 4. For both signal and background event yields, these uncertainties are $\leq 0.5\%$ for electrons and negligible for muons.

Uncertainties associated with the jet energy scale (JES) and resolution (JER) calibration are significantly larger than the corresponding lepton uncertainties. The JES uncertainties are split into 12 uncorrelated components (Section 4.5), which, added in quadrature, amount to variations of 3.2% (5.3%) on signal (background) event yields integrated over the three SR BDT bins in the $e\mu/\mu e$ channel (Table 9.2). Because the VBF signal region is rich in forward jets, the dominant JES uncertainty is the one associated with the extrapolation of the scale calibration from the central rapidity region to the forward region. The source of this uncertainty is the large difference observed between POWHEG and HERWIG. When propagated to the error on the signal strength, this is the largest instrumental uncertainty in the analysis (see Table 11.2).

The JER uncertainty also has a large impact on the SR predictions, at the level of 5% for signal and background. As discussed in Section 4.5, the JER, or equivalently

Table 9.2: Breakdown of the relative instrumental uncertainties (%) in the $e\mu/\mu e$ channel. Uncertainties are defined as the change in the event yield in the BDT SR, except for top background, for which uncertainties are computed as the variation in $\alpha = N^{SR}_{top,MC}/N^{CR}_{top,MC}$ (Equation 8.2).

Uncertainty Source	Signal	Total Back	top	WW	ggF	Non-WW Diboson	Z/DY	Fakes
Trigger	-	0.1	-	-	-	0.5	0.5	-
electron SF	0.8	0.7	0.8	0.9	0.8	0.9	0.9	-
electron reso.	-	0.6	0.3	0.1	-	1.4	4.7	-
electron scale	-	0.5	0.5	0.6	-	1.3	2.8	-
electron iso.	0.7	0.6	0.7	0.7	0.8	0.8	0.8	-
muon SF	-	-	-	-	-	-	-	-
muon reso.	-	-	-	-	-	-	0.1	-
muon scale	-	0.1	-	-	-	-	0.4	-
muon iso.	0.8	0.7	0.8	0.8	0.8	0.9	0.9	-
JES flavor	0.8	0.9	0.5	2.9	3.0	3.5	0.3	-
JES η model	2.8	4.7	3.9	10.2	5.9	1.3	6.6	-
JES η stat.	-	0.7	0.8	1.3	0.8	1.8	0.2	-
JES flav. response	0.7	0.2	0.1	2.1	1.9	2.9	4.7	-
JES detector	-	0.7	0.9	0.9	0.7	0.9	-	-
JES pileup p_{T}	-	0.6	-	0.6	-	-	4.3	-
JES high p_{T}	-	0.5	-	-	-	-	3.8	-
JES pileup ρ	-	-	0.5	1.5	1.4	0.6	5.6	-
JES modelling	1.1	1.7	1.9	3.8	3.4	5.4	2.6	-
JES $\langle\mu\rangle$	0.1	0.5	0.3	0.2	0.4	2.5	0.5	-
JES N_{PV}	-	0.3	0.2	0.7	0.4	2.8	3.2	-
JES AFII	-	-	0.1	-	-	-	-	-
JER	2.1	3.2	1.8	1.5	0.3	11.8	11.5	-
b-tag SF	-	2.4	6.5	-	-	-	-	-
light tag SF	1.3	1.3	1.2	1.7	1.9	1.8	1.6	-
c-tag SF	-	-	-	0.1	-	-	-	-
$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ scale	-	0.3	-	-	-	-	2.2	-
$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ reso.	-	-	-	-	-	-	-	-
$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ scale	-	2.2	2.6	1.9	1.0	0.3	6.6	-
$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ reso.	0.3	1.8	2.0	0.7	1.1	7.8	4.5	-
di-jet fake rate	-	1.9	-	-	-	-	-	14.5
fake rate μ	-	1.6	-	-	-	-	-	12.0
fake rate e	-	2.2	-	-	-	-	-	17.1
$\langle\mu\rangle$ re-scale	1.2	0.5	0.8	0.3	0.9	3.6	4.4	-

Table 9.3: Breakdown of the relative instrumental uncertainties (%) in the $ee/\mu\mu$ channel. Uncertainties are defined as the change in the event yield in the BDT SR, except for top background, for which uncertainties are computed as the variation in $\alpha = N^{SR}_{top,MC}/N^{CR}_{top,MC}$ (Equation 8.2).

Uncertainty Source	Signal	Total Back	top	WW	ggF	Non-WW Diboson	Z/DY	Fakes
Trigger	1.2	0.5	1.1	1.2	1.2	0.6	0.1	-
electron SF	0.6	0.3	0.6	0.7	0.6	1.9	0.1	-
electron reso.	-	0.4	0.5	0.1	0.3	0.3	0.5	-
electron scale	-	0.1	0.1	1.0	-	3.4	0.2	-
electron iso.	0.6	0.3	0.5	0.6	0.6	1.4	0.1	-
muon SF	-	-	-	-	-	-	-	-
muon reso.	-	0.1	-	0.4	-	-	0.1	-
muon scale	-	-	0.2	-	-	-	0.1	-
muon iso.	1.0	0.4	1.0	1.0	1.0	0.4	0.1	-
JES flavor	1.9	1.1	0.6	4.4	4.0	18.0	0.1	-
JES η model	5.2	3.2	5.4	9.8	8.0	10.2	1.2	-
JES η stat.	0.6	0.6	0.9	1.1	0.6	4.7	0.3	-
JES flav. response	1.4	0.9	0.7	2.3	2.5	14.2	0.1	-
JES detector	0.9	0.7	1.0	0.6	0.8	7.4	0.3	-
JES pileup p_{T}	-	0.4	-	0.1	0.3	0.9	0.6	-
JES high p_{T}	-	0.2	-	-	-	-	0.3	-
JES pileup ρ	0.4	0.5	0.1	2.0	1.0	10.9	-	-
JES modelling	2.1	1.6	1.4	4.6	4.2	15.2	0.6	-
JES $\langle\mu\rangle$	-	0.5	0.1	1.1	0.7	3.3	0.4	-
JES N_{PV}	-	0.4	0.1	1.1	0.3	7.0	0.3	-
JES AFII	-	-	0.1	-	-	-	-	-
JER	2.5	1.5	1.8	4.1	1.0	1.2	1.3	-
b-tag SF	-	1.3	6.8	-	-	0.8	-	-
light tag SF	1.4	0.6	1.2	1.8	2.0	1.3	0.1	-
c-tag SF	-	-	-	0.1	-	0.2	-	-
$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ scale	-	0.2	-	-	-	3.6	0.2	-
$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ reso.	-	0.2	-	-	-	5.9	0.1	-
$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ scale	0.1	1.0	3.0	1.1	1.1	2.7	0.6	-
$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ reso.	0.6	0.9	1.6	1.2	1.1	10.1	0.4	-
di-jet fake rate	-	0.1	-	-	-	-	-	4.9
fake rate μ	-	0.3	-	-	-	-	-	12.6
fake rate e	-	0.3	-	-	-	-	-	13.7
$\langle\mu\rangle$ re-scale	3.4	0.6	0.4	1.2	1.7	13.5	0.2	-

$\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$, is measured *in situ*, and the relative uncertainties are evaluated to be 10%-20%, depending on the p_{T} range. To propagate these uncertainties to this analysis, the p_{T} of all jets with p_{T} greater than some loose threshold is first smeared to match data. The 1σ variation in event yields is obtained by performing additional smearing, in bins of jet p_{T}, to account for the uncertainty on the $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$ measurement. This additional p_{T} smearing migrates jets across the jet multiplicity bins, and any pre-selection cuts correlated to the p_{T} of the tag or central jets. Moreover, because the BDT inputs m_{jj}, ΔY_{jj}, $\sum_{\ell,j} M_{\ell j}$, and η_{lep} centrality are correlated with the jet p_{T}, JER is expected to induce a shape uncertainty.

Uncertainties associated with the efficiencies for tagging b-hadron jets are discussed in Sections 4.6 and 7.1.4. The efficiency SFs for tagging a b-hadron jet, derived from a $t\bar{t}$-rich CR, are shown with their associated uncertainties in Table 7.4. Since there are six p_{T} bins, the uncertainties are split into six uncorrelated sources. In Tables 9.2 and 9.3, the six components are added in quadrature, and the resulting uncertainty is 2.4% (1.3%) on the background event yields in the $e\mu/\mu e$ ($ee/\mu\mu$) channel. This uncertainty primarily impacts top processes because they have true b-hadron jets. Additional uncertainties are assigned for the efficiencies of mis-tagging a light flavor or c-hadron jet. The former is at the level of 1%, while the latter is less than 0.5%.

As described in Section 4.11, $E_{\mathrm{T}}^{\mathrm{miss}}$ is the vector sum of the hard objects in the event with an additional term to capture the soft activity. Because the uncertainties associated with the leptons and jets comprising $E_{\mathrm{T}}^{\mathrm{miss}}$ are already accounted for, the only remaining uncertainties are on the soft term scale and resolution. In the $e\mu/\mu e$ channel, no explicit $E_{\mathrm{T}}^{\mathrm{miss}}$ cuts are applied. However, because the BDT inputs $p_{\mathrm{T}}^{\mathrm{tot}}$ and m_{T} use $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$, and the $Z \to \tau\tau$ veto depends on $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$, there is an event yield variation associated with the $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ soft term. Due to the relatively low QCD activity in VBF events, the uncertainty on this term is small for signal—0.3%

(0.6%) for scale (resolution). For background processes, which generally have more QCD radiation, the event yield variations are larger. In the $ee/\mu\mu$ channel, where cuts are applied on both $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ and $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$, uncertainties on both $E_{\mathrm{T}}^{\mathrm{miss}}$ types are assigned, with the $E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ soft term uncertainties being significantly larger.

The uncertainties on backgrounds from fakes—W+jets and multi-jets—discussed in Section 8.3, are relatively large, but because these backgrounds account for a small fraction of the total background in the SR, the overall uncertainty is at the level of 2% (0.1%) in the $e\mu/\mu e$ ($ee/\mu\mu$) channel.

The uncertainty on the measured integrated luminosity has been evaluated to be 2.8% for the ATLAS data integrated in Run-1 at $\sqrt{s} = 8$ TeV. Predictions which are purely from MC simulation are scaled to this integrated luminosity (Equation 6.1) and are therefore subject to the same uncertainty. Background estimates which use information from data—top, Z/DY, and fakes– are not subject to this uncertainty.

10

Results

The observed number of events recorded in the ATLAS detector in the signal regions are compared to the predictions from MC simulation and associated data-driven approaches. An excess over the background prediction in the absence of VBF signal, if statistically significant, is considered to be associated with VBF Higgs production. In the following chapter, the statistical evaluation of such a scenario is considered, and in this section, the event yields and kinematic plots in the SRs are presented.

10.1 $\sqrt{s} = 8$ TeV Analysis

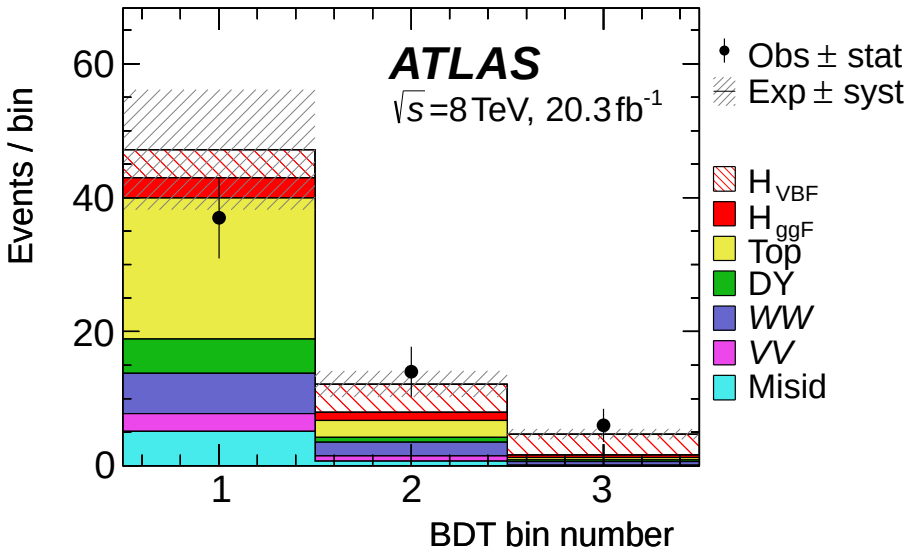
Table 10.1 shows the predicted event yields for signal and backgrounds and the observed event yields for each cut stage in the analysis, starting at the common pre-selection. After the common pre-selection cuts and the requirement of two or more jets, the the expected number of VBF events in the $e\mu/\mu e$ lepton channel is 64, with a background prediction of ~60k events. Applying the additional VBF pre-selection cuts, BJV, CJV, OLV, and the $Z \to \tau\tau$ veto, the background contribution is significantly suppressed to 703 events, with 17 expected signal events. The discriminating power of the BDT is illustrated by the BDT > -0.48 cut which removes 94% of the

remaining background, while preserving 70% of the signal. In this region an excess over the background hypothesis is observed, with 57 events observed and 44 expected. This excess is concentrated in the two highest BDT bins, where the expected S/B is largest (figure 10.1(a)). In the first BDT bin, there is a deficit with respect to the background prediction, due to the large correction of 1.58 ± 0.10 applied to the top background in this bin. The source of this large correction is poor lepton modeling in POWHEG (Section 8.1), which is covered by the modeling systematic. In the $ee/\mu\mu$ channel, there is also an excess in the SR, with 54 background events and 73 events observed. The signal prediction in this region is 6 events. Again, the relative excess in the most sensitive BDT bins is more pronounced (Figure 10.1(b)).

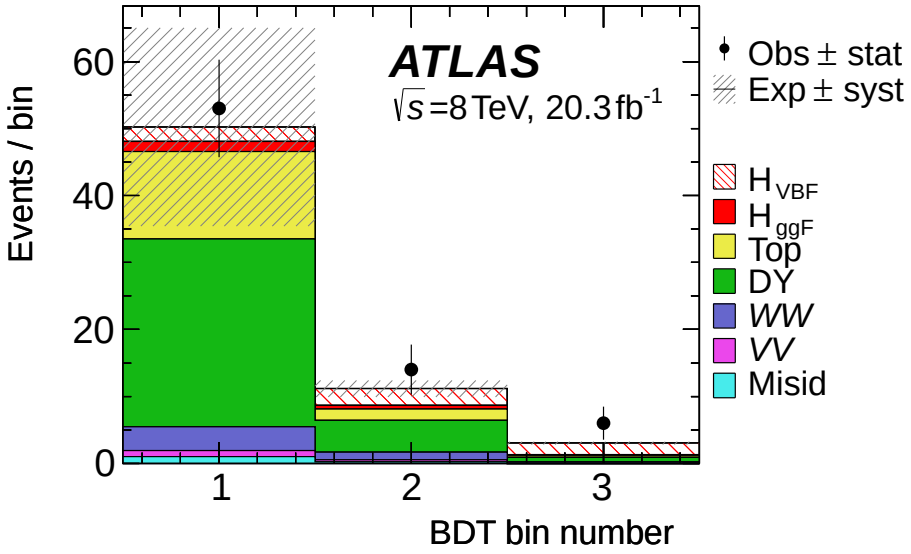
The BDT input distributions in the SR with all flavor channels combined are shown in Figure 10.2. This figure illustrates that the excess in data lies in kinematic regions coincident with VBF, most notably at high m_{jj}, low $\Delta\phi_{\ell\ell}$, low $m_{\ell\ell}$, and low $p_{\mathrm{T}}^{\mathrm{tot}}$. A VBF candidate event falling into BDT bin 2 is displayed in Figure 10.3.

10.2 $\sqrt{s} = 7$ TeV Analysis

The BDT analysis has been optimized for the 20.3 fb^{-1} dataset collected by the ATLAS detector at $\sqrt{s} = 8$ TeV. In addition to this dataset, a smaller dataset of 4.5 fb^{-1} integrated at $\sqrt{s} = 7$ TeV has been analyzed. Since this dataset is substantially smaller, expected gains in sensitivity are expected to be marginal, and therefore an independent optimization has not been performed for this analysis. Instead, the BDT developed for the 8 TeV dataset is applied in the 7 TeV analysis, under the assumption that the difference in $\sqrt{s}$ and pile-up conditions do not have a significant impact on the kinematics of the final state objects. Another compelling consideration motivating the use of the same BDT is that because the BDT defines the signal region of phase space, and the 8 TeV theory uncertainties have been computed for this region, the theory uncertainties can be recycled for the 7 TeV analysis. In the



(a) $e\mu/\mu e$ channel



(b) $ee/\mu\mu$ channel

FIGURE 10.1: BDT response distribution in (a) $e\mu/\mu e$ channel and (b) $ee/\mu\mu$ channel. The error band represents instrumental, theoretical and statistical uncertainties. Top and Z/DY normalization factors are applied. VBF signal is shown in hatched red, not to be confused with ggF, shown in solid red.

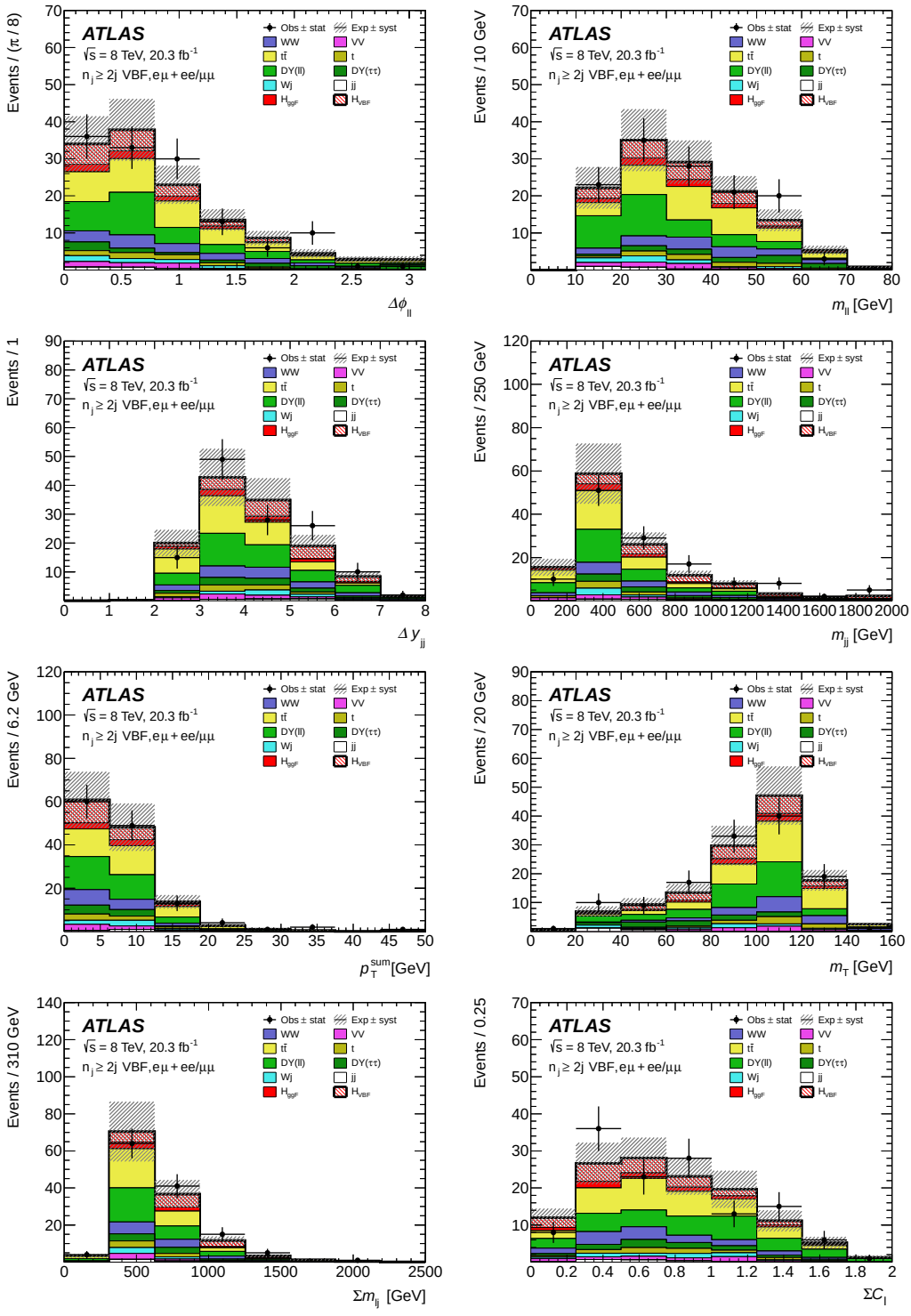


FIGURE 10.2: Distributions of $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, ΔY_{jj}, m_{jj}, $p_{\mathrm{T}}^{\mathrm{tot}}$, m_{T}, $\sum_{\ell,j} M_{\ell j}$, and η_{lep} centrality in both $e\mu/\mu e$ and $ee/\mu\mu$ channel in the BDT signal region (BDT > -0.48).

Table 10.1: Observed and expected event yields at each cut stage, starting with the N_{jet} cut. Expected event yields are split into the background components. The expected signal includes VBF and VH contributions. Highlighted in blue are the yields in the three signal region BDT bins, which go into the likelihood fit. The uncertainties shown are statistical only. Normalization factors are applied to top and Z/DY in the SR, but not at other cut stages.

Cut stage	Observed	Signal	Total Back	top	WW	ggF	Non-WW Diboson	Z/DY	Fakes
$N_{\text{jet}} \geq 2$	676470	134	669892	106081	2779	198	1553	556559	2722
$e\mu/\mu e$ channel	61434	64	59597	53196	1423	103	378	3356	1142
$N_{b\text{-jet}} = 0$	7818	47	7569	3232	1036	76	272	2460	493
CJV	6313	40	6097	2462	859	61	226	2083	408
OLV	1264	23	1273	543	156	18	46	435	76
$Z \to \tau\tau$ veto	718	17	703	370	101	14	31	134	53
BDT > -0.48	57	12	44	16	9	5	3	5	6
BDT bin 1	37	4.4 ± 0.1	42.3 ± 2.2	21.1	6.0	3.0	2.6	4.5	5.1
BDT bin 2	14	4.3 ± 0.1	7.9 ± 0.9	2.6	2.1	1.2	0.8	0.6	0.7
BDT bin 3	6	3.1 ± 0.1	1.5 ± 0.2	0.4	0.5	0.3	0.1	0.2	0.0
$ee/\mu\mu$ channel	615036	70	607896	50497	1356	95	1175	553203	1569
$E_{\text{T}}^{\text{miss,CALO}} > 45$ GeV	119456	45	118395	38710	1025	60	736	77376	488
$E_{\text{T}}^{\text{miss,TRK}} > 40$ GeV	58672	39	56435	35832	944	51	640	18560	408
Z veto	34339	35	33174	28358	756	50	177	3509	323
$m_{\ell\ell} < 75$ GeV	19552	32	18330	14811	365	50	105	2743	256
$N_{b\text{-jet}} = 0$	3367	24	3214	879	265	36	72	1900	62
CJV	2653	20	2515	671	219	29	59	1489	49
OLV	664	12	571	171	47	8	13	325	6
$Z \to \tau\tau$ veto	469	9	403	149	40	6	10	191	6
BDT > -0.48	73	6	54	10	5	2	1	34	1
BDT bin 1	53	2.3 ± 0.1	47.8 ± 5.7	13.1	3.5	1.5	0.9	27.8	1.0
BDT bin 2	14	2.5 ± 0.1	8.7 ± 1.6	1.7	1.1	0.6	0.3	4.8	0.2
BDT bin 3	6	1.7 ± 0.1	1.3 ± 0.3	0.3	0.3	0.2	0.0	0.6	0.0

following section, the $\sqrt{s} = 7$ TeV BDT analysis is briefly documented.

10.2.1 Backgrounds

Backgrounds for the 7 TeV analysis are predicted with MC simulation and, when necessary, data-driven estimates. The MC generators are identical to those of the 8 TeV analysis, with the exception of the QCD WW generator, for which POWHEG interfaced to PYTHIA 6 is used, due to the unavailability of SHERPA. Since POWHEG

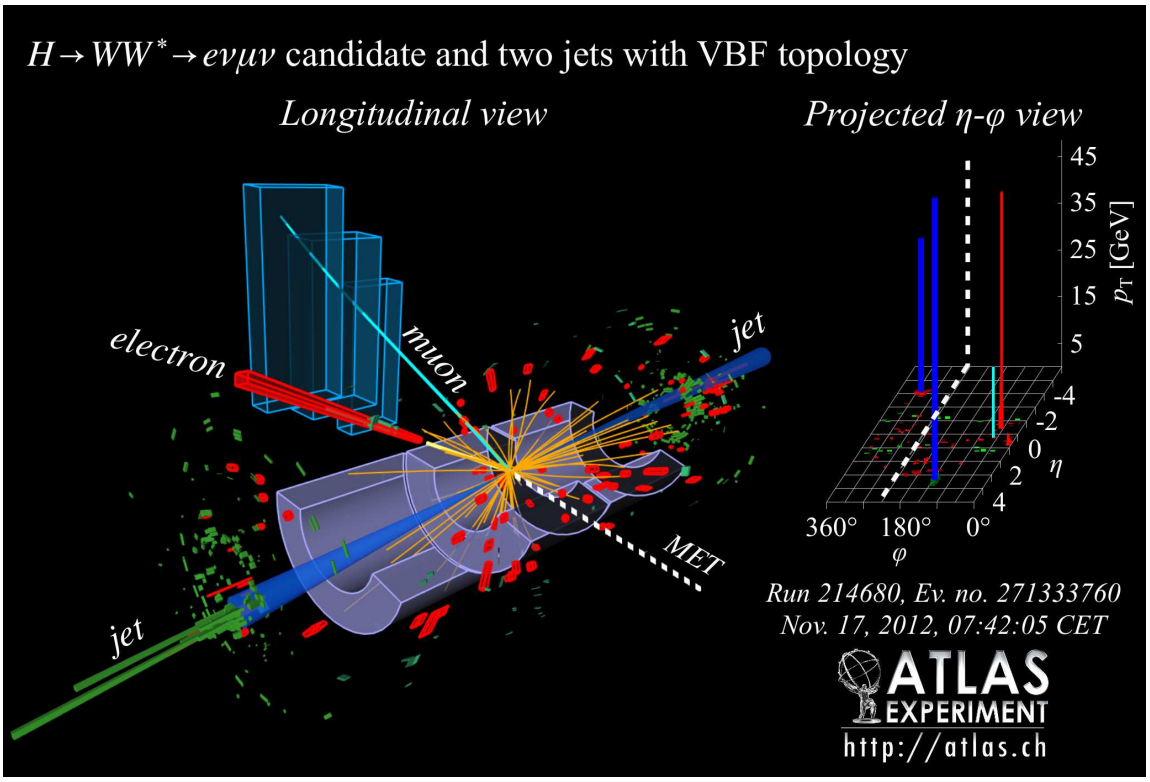


FIGURE 10.3: Display of a collision event passing the $e\mu/\mu e$ event selection and falling into BDT bin 1 (BDT score = 0.26). This VBF candidate illustrates the $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay topology—small $\Delta\phi_{\ell\ell}$ recoiling against $E_{\mathrm{T}}^{\mathrm{miss}}$ (shown with dashed line)—and the VBF topology—forward jets with a large η gap. For this event, the VBF inputs have the following values: $m_{\ell\ell}$ = 21 GeV, $\Delta\phi_{\ell\ell}$ = 0.1, m_{jj} = 1400 GeV, ΔY_{jj} = 6.6, m_{T} = 127 GeV, $p_{\mathrm{T}}^{\mathrm{tot}}$ = 15 GeV, η_{lep} centrality = 0.2, $\sum_{\ell,j} M_{\ell j}$ = 887 GeV.

is only able to generate inclusive WW at NLO, the second high-p_{T} jet is produced in the parton shower, leading to potential mis-modeling of di-jet kinematic variables. Due to the absence of a sufficiently pure WW + 2j control region, the modeling of this background can not be validated. Instead, large theory systematics are assigned.

The MC predictions are normalized to their respective cross sections at $\sqrt{s}$ = 7 TeV, shown in Table 10.2, which are generally on the order of 20% smaller than those at 8 TeV. The VBF cross section decreases by 22%, which is compensated for by smaller background cross sections.

Table 10.2: Cross sections for $\sqrt{s} = 7$ TeV MC predictions and associated change with respect to $\sqrt{s} = 8$ TeV. MC generators are the same as those in the 8 TeV analysis, shown in Table 6.1, except for QCD WW †.

Process	$\sigma \cdot$ Br(7 TeV) (pb)	change w.r.t. 8 TeV
VBF $H \to WW$	$28 \cdot 10^{-3}$	-22%
ggF $H \to WW$	0.341	-22%
WH/ZH $H \to WW$	$21 \cdot 10^{-3}$	-16%
$t\bar{t}$ dileptonic	18.6	-30%
tW/tb leptonic	3.15	-24%
tqb leptonic	20.7	-27%
QCD $WW + 2$ jets †	0.468	-18%
EW $WW + 2$ jets	0.027	-31%
$gg \to WW$	0.14	-30%
inclusive W	$31 \cdot 10^3$	-16%
inclusive $Z/\gamma^\star(m_{ll} \geq 10GeV)$	$14.9 \cdot 10^3$	-10%
EW $Z/\gamma^\star$	2.26	n/a
$W(Z/\gamma^*)$	10.8	-15%
$W(Z/\gamma^*)(m_{(Z/\gamma^*)} < 7$ GeV)	10.6	-13%
$Z^{(*)}Z^{(*)} \to 4l(2l2\nu)$	0.64(0.42)	-12%(-16%)
EW $WZ + 2$ jets	$8.5 \cdot 10^{-3}$	-35%
EW $ZZ + 2$ jets $(4l, ll\nu\nu)$	$53 \cdot 10^{-5}(8.8 \cdot 10^{-4})$	-27%(-27%)
$W\gamma$	313	-15%

Due to limited data statistics in CRs and limited MC simulation statistics in the SRs, the binning of the BDT response is coarser than that of the 8 TeV analysis. However, in order to use the BDT-dependent theory systematics computed at 8 TeV, the same bin boundaries are used and low statistics bins are merged. In the $e\mu/\mu e$ channel, the two high BDT bins in the 8 TeV analysis are merged, resulting in two bins with boundaries [-0.48,0.3,1.0]. In the $ee/\mu\mu$ channel, the three BDT bins are merged into a single bin with boundaries [-0.48,1.0]. The instrumental uncertainties in these SRs are shown in Tables 10.3 and 10.4.

Data-driven background estimation techniques are inherited from the 8 TeV analysis. The top background is normalized from the same CR, with the exception of the BDT binning. Due to limited stats, the NF is computed with a control region with all BDT bins merged and is applied across the BDT spectrum. The resulting purity is 91% and the NF is 0.8 ± 0.3. $Z/\text{DY} \to ee/\mu\mu$ is estimated with the ABCD

method for a single BDT bin, resulting in an estimate of $N_{\mathrm{DY}}^{A,est.} = 2.3 \pm 1.0$, which is compatible with the prediction from MC simulation of 1.4 ± 0.5 events.

10.2.2 Results

The predicted and observed event yields for the 7 TeV analysis are shown in Table 10.5. In the $e\mu/\mu e$ channel, the expected number of background events in the first bin is 3.5, with 6 events observed, and in the second BDT bin 0 events are observed with the expected background prediction of 0.9 events. In the single $ee/\mu\mu$ BDT bin, 4.7 background events are predicted with 0.8 signal events, and the number of events observed is 3.

10.3 Cut-based Analysis

As a cross-check to the BDT analysis, an analysis with sequential selection cuts is performed. Since the signal region of such a cut-based analysis is more easily understood, it is an important test of the predictions of the BDT analysis. Moreover, comparing the SRs of the cut-based analysis to those of the BDT results in a better understanding of BDT analysis.

10.3.1 Selection

The essential difference between the BDT analysis and the cut-based approach is that instead of training a BDT classifier with eight discriminating independent variables, sequential cuts are placed on these variables. In the cut-based analysis, all BDT inputs are used, with the exception of $\sum_{\ell,j} M_{\ell j}$ and η_{lep} centrality, since including these two distributions in the cut-based selection has been shown to yield little sensitivity improvement. Cut values have been determined in a simultaneous optimization procedure, and the resulting cuts are shown in Table 10.6. The same common pre-selection cuts are applied for both the BDT and cut-based analyses, ex-

Table 10.3: Breakdown of instrumental uncertainties in the $e\mu/\mu e$ channel for the 7 TeV analysis. Uncertainties are defined as the change in the event yield in the BDT SR, except for top background, for which uncertainties are computed as the variation in $\alpha = N^{SR}_{top,MC}/N^{CR}_{top,MC}$ (Equation 8.2).

Uncertainty Source	Signal	Total Back	top	WW	ggF	Non-WW Diboson	Z/DY	Fakes
Trigger	-	0.2	0.5	-	-	0.4	-	-
electron SF	1.5	1.1	1.2	1.0	1.4	1.3	1.4	-
electron reso.	0.1	0.4	1.0	0.1	0.4	2.9	1.3	-
electron scale	-	0.2	0.7	0.5	-	3.5	1.3	-
electron iso.	2.0	2.0	2.1	2.1	2.0	2.0	2.1	-
muon SF	-	-	-	-	-	-	-	-
muon reso.	0.1	0.7	1.6	0.1	-	-	0.1	-
muon scale	-	0.5	1.3	-	-	-	0.1	-
muon iso.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	-
JES detector	0.1	0.7	0.1	3.7	0.9	1.2	1.1	-
JES η stat.	-	0.1	-	2.0	-	-	1.3	-
JES modelling	0.7	0.5	3.9	3.0	0.3	0.6	7.4	-
JES Stat.	0.1	1.0	-	1.4	0.3	0.3	6.1	-
JES b jet	-	0.6	1.6	0.1	-	0.2	0.1	-
JES η model	3.5	7.5	13.9	6.2	7.1	0.8	0.6	-
JES flavor	1.6	2.5	2.1	0.4	4.8	11.6	1.3	-
JES flav. response	0.4	2.2	3.2	2.7	0.9	0.4	1.4	-
JES $\langle\mu\rangle$	1.1	2.5	2.3	6.9	1.7	1.0	1.4	-
JES N_{PV}	-	2.3	2.7	0.5	0.7	0.2	6.3	-
JES AFII	0.1	0.5	-	3.4	1.4	0.3	-	-
JER	0.7	1.4	0.3	3.5	3.1	3.1	4.1	-
b-tag SF	-	2.2	5.7	-	-	-	0.1	-
light tag SF	1.3	1.5	1.3	1.4	1.6	1.5	2.4	-
c-tag SF	0.6	0.2	-	1.0	0.5	-	-	-
$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ scale	-	-	-	-	-	-	-	-
$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ reso.	-	-	-	-	-	-	-	-
$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ scale	0.3	2.4	4.7	0.1	1.0	0.6	2.5	-
$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ reso.	0.2	2.9	6.1	1.4	1.0	6.1	2.1	-
di-jet fake rate	-	0.5	-	-	-	-	-	12.8
fake rate μ	-	0.3	-	-	-	-	-	7.9
fake rate e	-	0.9	-	-	-	-	-	24.4
$\langle\mu\rangle$ re-scale	-	-	-	-	-	-	-	-

Table 10.4: Breakdown of instrumental uncertainties in the $ee/\mu\mu$ channel for the 7 TeV analysis. Uncertainties are defined as the change in the event yield in the BDT SR, except for top background, for which uncertainties are computed as the variation in $\alpha = N^{SR}_{top,MC}/N^{CR}_{top,MC}$ (Equation 8.2). Given that the background estimate for $Z/\mathrm{DY} \to ee/\mu\mu$ is data-driven, instrumental uncertainties do not contribute. Non-zero Z/DY uncertainties in the table are from the EW contribution.

Uncertainty Source	Signal	Total Back	top	WW	ggF	Non-WW Diboson	Z/DY	Fakes
Trigger	0.5	0.2	0.1	-	0.5	0.2	0.2	-
electron SF	1.1	0.4	0.8	1.0	1.3	2.0	-	-
electron reso.	-	0.1	-	0.9	0.8	0.6	0.1	-
electron scale	-	1.0	4.4	0.2	0.4	0.6	0.5	-
electron iso.	1.5	0.9	1.3	1.9	1.7	3.3	0.3	-
muon SF	-	0.1	-	-	-	-	0.1	-
muon reso.	-	-	-	0.1	-	0.1	-	-
muon scale	-	-	-	-	-	-	-	-
muon iso.	1.3	0.7	1.4	1.1	1.2	0.4	0.4	-
JES detector	1.3	2.1	9.5	0.7	2.8	-	0.3	-
JES η stat.	0.7	1.2	3.8	1.8	1.8	-	0.3	-
JES modelling	1.1	2.0	5.1	3.3	4.0	1.1	0.7	-
JES Stat.	0.3	1.1	3.8	1.7	0.7	-	0.3	-
JES b jet	-	1.5	8.2	0.1	-	-	-	-
JES η model	5.5	4.8	8.9	13.5	9.2	6.8	0.4	-
JES flavor	3.5	3.1	-	0.2	6.8	0.2	5.3	-
JES flav. response	2.1	1.5	1.3	6.1	5.1	1.1	-	-
JES $\langle\mu\rangle$	1.5	1.8	8.1	4.0	2.8	0.6	1.0	-
JES N_{PV}	0.1	1.8	8.2	0.7	-	-	0.4	-
JES AFII	0.4	1.7	4.8	1.4	0.1	0.2	1.3	-
JER	0.3	4.8	16.3	2.3	2.4	1.2	4.3	-
b-tag SF	-	0.9	5.1	-	-	-	-	-
light tag SF	1.3	1.0	1.3	1.7	1.6	1.8	0.6	-
c-tag SF	0.6	-	-	0.1	-	-	-	-
$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ scale	-	0.8	4.4	-	-	-	-	-
$E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ reso.	-	-	-	0.2	0.1	-	-	-
$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ scale	-	1.3	3.6	0.7	1.4	0.1	1.3	-
$E_{\mathrm{T}}^{\mathrm{miss,TRK}}$ reso.	0.3	1.5	6.3	2.6	1.5	6.2	0.5	-
di-jet fake rate	-	-	-	-	-	-	-	0.2
fake rate μ	-	2.2	-	-	-	-	-	37.9
fake rate e	-	0.3	-	-	-	-	-	4.8
$\langle\mu\rangle$ re-scale	-	-	-	-	-	-	-	-

Table 10.5: Observed and expected event yields for the 7 TeV analysis at each cut stage, starting with the N_{jet} cut. Expected event yields are split into the background components. The expected signal includes VBF and VH contributions. Highlighted in blue are the yields in the three signal region BDT bins, which go into the likelihood fit. The uncertainties shown are statistical only. Normalization factors are applied to top and Z/DY in the SR, but not at other cut stages.

Cut stage	Observed	Signal	Total Back	top	WW	ggF	Non-WW Diboson	Z/DY	Fakes
$N_{\text{jet}} \geq 2$	155205	16	151286	13809	289	27	226	135668	1268
$e\mu/\mu e$ **channel**	8042	8	7668	6831	143	13	54	507	119
$N_{b\text{-jet}} = 0$	949	5	950	372	102	10	39	363	65
CJV	799	5	798	294	91	8	34	310	61
OLV	194	3	172	66	18	2	6	65	14
$Z \to \tau\tau$ veto	100	2	92	43	11	2	4	21	10
BDT > -0.48	6	1	5	2	1	1	0	1	0
BDT bin 1	6	0.6 ± 0.0	3.5 ± 0.5	1.1	0.6	0.4	0.4	0.9	0.1
BDT bin 2	0	0.9 ± 0.0	0.9 ± 0.2	0.4	0.3	0.2	0.0	0.0	0.0
$ee/\mu\mu$ **channel**	147163	8	143618	6978	146	14	172	135160	1149
$E_{\text{T}}^{\text{miss,CALO}} > 45$ GeV	10251	5	9560	5054	103	8	85	4230	79
$E_{\text{T}}^{\text{miss,TRK}} > 40$ GeV	6415	4	6108	4691	96	7	75	1200	39
Z veto	4272	4	4134	3697	76	7	22	310	22
$m_{\ell\ell} < 75$ GeV	2322	4	2221	1888	37	7	13	258	18
$N_{b\text{-jet}} = 0$	328	3	322	106	26	5	9	173	4
CJV	261	3	263	83	23	4	8	142	4
OLV	56	2	56	20	5	1	2	27	1
$Z \to \tau\tau$ veto	39	1	36	17	4	1	1	12	1
BDT bin 1	3	0.8 ± 0.0	4.7 ± 1.5	0.7	0.6	0.2	0.2	2.8	0.2

cept for the $E_{\text{T}}^{\text{miss}}$ cuts, which are tighter for the cut-based analysis. Upon applying the cuts sequentially, a binned likelihood fit is performed on the m_{T} distribution, with optimized bin boundaries of [0,80,130,∞]. In addition to an $m_{jj} > 600$ GeV cut, the m_{T} fit is split into a $600 < m_{jj} < 1000$ GeV bin and a $m_{jj} > 1000$ GeV bin to exploit the larger S/B in the $m_{jj} > 1000$ GeV region.

10.3.2 Results

Backgrounds for the cut-based analysis are simulated with the same generators as the BDT-based analysis, and the analogous data-driven techniques are used for top

Table 10.6: Summary of the selection cuts applied in the cut-based VBF $H \to WW^{(*)}$ analysis which is used to cross check the BDT analysis. After all cuts are applied, a binned likelihood fit is performed on the m_{T} distribution.

Common pre-selection	$p_{\mathrm{T1}(T2)} > 22(10)$ GeV $m_{\ell\ell} > 10(12)$ GeV for $e\mu/\mu e$ ($ee/\mu\mu$) opposite charge leptons $\lvert m_{\ell\ell} - m_Z \rvert > 15$ GeV
Background rejection	$N_{\mathrm{jet}} \geq 2$ $E_{\mathrm{T}}^{\mathrm{miss,CALO}} > 55$ GeV ($ee/\mu\mu$) $E_{\mathrm{T}}^{\mathrm{miss,TRK}} > 50$ GeV ($ee/\mu\mu$) $N_{b\text{-jet}} = 0$ $p_{\mathrm{T}}^{\mathrm{tot}} < 15$ GeV $Z \to \tau\tau$ veto
VBF topology	CJV OLV $\Delta Y_{jj} > 3.6$ $m_{jj} > 600$ GeV, split at 1 TeV
Higgs decay topology	$m_{\ell\ell} < 50$ GeV $\Delta\phi_{\ell\ell} < 1.8$ ($p_{\mathrm{T2}} > 15$ GeV) $\Delta\phi_{\ell\ell} < 2.8$ ($10 < p_{\mathrm{T2}} < 15$ GeV) Fit m_{T} distribution

quark and Z/DY processes, as well as backgrounds due to mis-identified leptons. For the top CR, in order to align the lepton and jet kinematics as closely as possibly with the signal region, the same selection cuts are applied, but with the $N_{b\text{-jet}} = 1$ requirement. The $\Delta\phi_{\ell\ell}$ and m_{jj} cuts are removed from the CR to increase the number of events in these regions, thereby reducing the statistical uncertainty on the top NF. Merging the CR across lepton flavor channels and in the high and low m_{jj} bins, the NF is 1.04 ± 0.19. The ABCD method closely follows that of the BDT analysis. Because the $E_{\mathrm{T}}^{\mathrm{miss,CALO}}$ cut is tighter, the regions are adjusted accordingly, and the pre-selection applied to the regions includes the $p_{\mathrm{T}}^{\mathrm{tot}}$ and m_{jj} cuts. The resulting correction applied to $Z/\mathrm{DY} \to ee/\mu\mu$ events is 0.97 ± 0.42. The systematic uncertainties considered for the cut-based analysis are the same as those of the BDT

analysis. Due to the large phase space overlap between the SRs of the two analyses (Section 10.3.3), the uncertainties for the cut-based analysis are comparable to those in the BDT analysis.

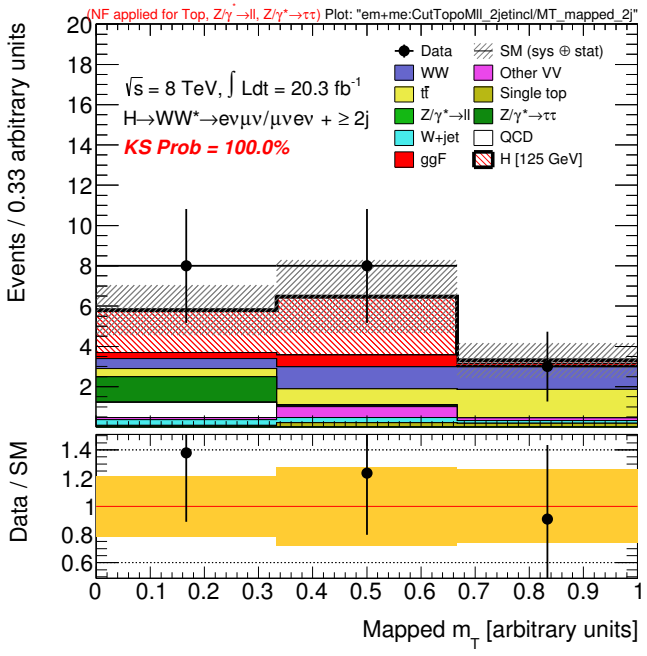
The m_{T} distribution in the cut-based SR is shown in Figure 10.4. Most of the signal lies in the first two m_{T} bins, with the third bin retained to test the background prediction. In the $e\mu/\mu e$ channel, there is an excess over the background prediction in bins 1 and 2. There is also an excess in bin 1 of the $ee/\mu\mu$ channel, but the observed number of events in the bin 2 is more consistent with the background hypothesis. However, in this channel, the statistical uncertainties are large, and the observation is consistent with both signal and background only hypotheses.

CB pass efficiency		BDT bin		
		1	2	3
m_{T} bin	1	8.5%	22.3%	24.8%
	2	5.7%	22.1%	55.1%
	3	1.9%	1.4%	0.3%
	all	16.1%	45.8%	80.2%

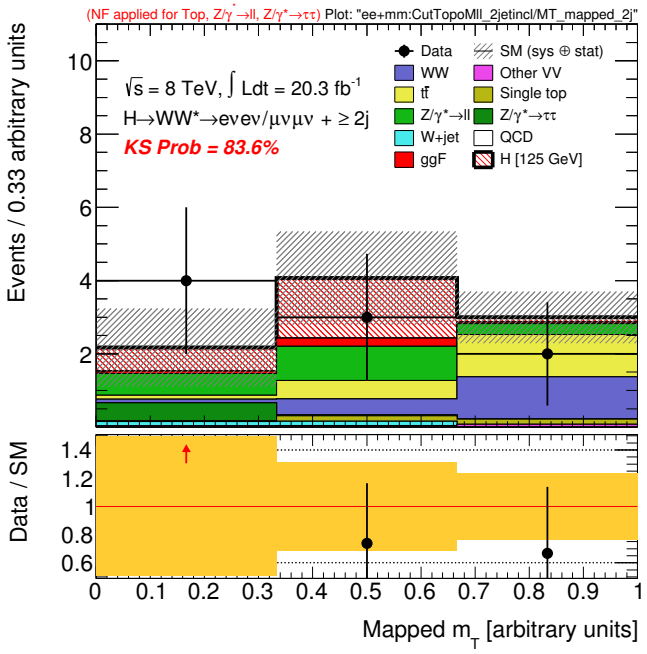
Table 10.7: Fraction of signal events in $e\mu/\mu e$ events in BDT SR that also pass the cut-based selection split into BDT and m_{T} bins (e.g., 5.7% of the signal falling in BDT bin 1 falls into m_{T} bin 2). "All" indicates the fraction when integrated over all m_{T} bins.

10.3.3 Comparison to BDT

By comparing the cut-based and BDT analyses, one gains insight into the phase space being selected by the BDT, and consequently a better understanding of why the BDT analysis out-performs the cut-based analysis. As indicated in Table 10.1, the three BDT SR bins in the $e\mu/\mu e$ channel have increasing S/B—0.1, 0.5, and 2.1, respectively—and the total signal captured in the SR is 11.8 events, or ~20% of all of the $e\mu/\mu e$ events with two jets. The m_{T} bins, on the other hand, have S/B of 0.5, 0.8, and 0.0, and only 8% of the signal is retained. In Table 10.7, the fraction of signal events in the BDT SR that pass the cut-based selection is shown in bins



(a) $e\mu/\mu e$ channel



(b) $ee/\mu\mu$ channel

FIGURE 10.4: Transverse mass distribution in the (a) $e\mu/\mu e$ channel and (b) $ee/\mu\mu$ channel. The error band represents instrumental, theoretical and statistical uncertainties. Top and Z/DY normalization factors are applied. VBF signal is shown in hatched red, not to be confused with ggF, shown in solid red.

of m_{T} and BDT. The overlap between the third BDT bin and the first two m_{T} bins is at the level of 80%. In the first two BDT bins, the respective fractions passing the cut-based selection are 16.1% and 45.8%, suggesting that the gain in acceptance is primarily from these bins. The fraction of signal events in the cut-based signal region which fall into the BDT SR, on the other hand, is significantly higher, ~90% for the first and third m_{T} bins. For the most sensitive m_{T} bin, bin 2, the fraction is 100%.

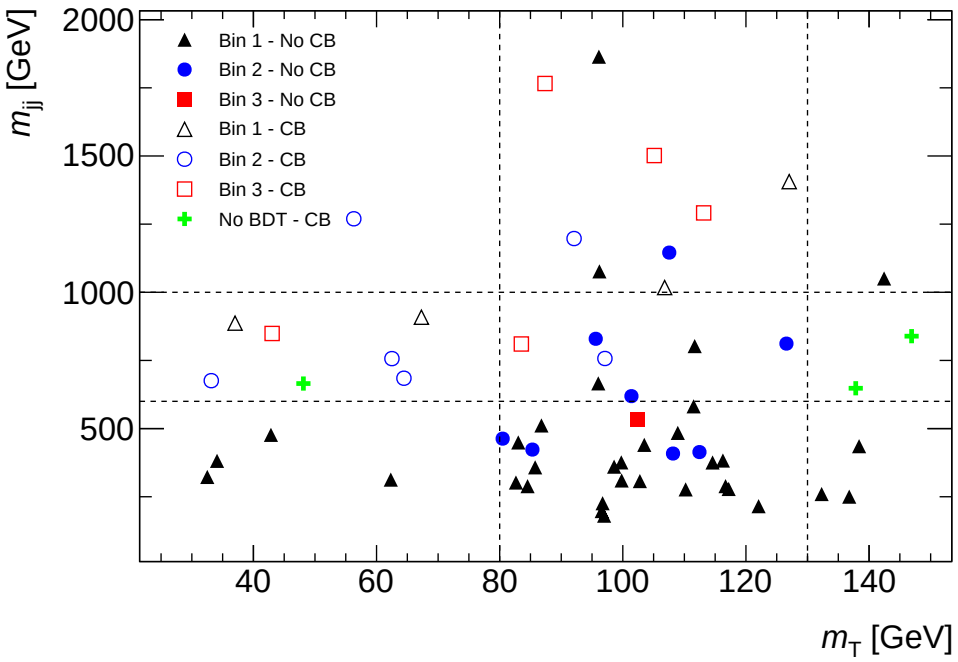


FIGURE 10.5: Scatter plot of the m_{jj} and m_{T} values for data events falling in the BDT or cut-based signal regions ($e\mu/\mu e$ channel only). The dashed lines indicate the binning for the cut-based analysis ($m_{jj} < 600$ GeV events fail the selection). Events which fall into BDT bin 1 are shown in **black**, bin 2 in **blue**, bin 3 in **red**. Solid markers indicate that an event has passed the BDT selection and not the cut-based selection, and hollow markers indicate that both have been passed. Finally, events that pass the cut-based selection and not the BDT are shown in **green**.

From the above numbers, it is clear that the BDT analysis retains more signal events and achieves better background rejection. Given the multivariate nature of

the BDT, understanding why this is the case is not a trivial task. The scatter plot of m_{jj} vs. m_{T} for observed events passing the BDT SR selection is shown in Figure 10.5. In order to compare to the cut-based analysis, the cut-based fit bins are illustrated with dashed lines, and a distinction is drawn between events which pass the cut-based selection and those that do not. As expected, the majority of the events which fall in the BDT SR, but not in the cut-based SR are in the first BDT bin. These events generally fall in the m_{T} window consistent with an $m_H = 125$ GeV Higgs boson decay, but with m_{jj} smaller than expected for a VBF Higgs. Some of these events fluctuate to higher BDT bins, since the values of the other BDT inputs take on values which are more consistent with VBF. For example, the bin 3 event at ($m_{\mathrm{T}} = 102$ GeV,$m_{jj} = 533$ GeV) falls close to the bin 1 event at ($m_{\mathrm{T}} = 109$ GeV,$m_{jj} = 484$ GeV). This event is considered more VBF-like mainly because $p_{\mathrm{T}}^{\mathrm{tot}} = 2.5$ GeV, while the corresponding value for the bin 1 event is 20 GeV. The inputs $\Delta\phi_{\ell\ell}$ and $m_{\ell\ell}$ are also more VBF-like for the bin 3 event, and the remaining inputs have comparable values for the two events. As illustrated in the figure, the bin 3 event fails the $m_{jj} > 600$ GeV cut in the cut-based selection. In fact, of the eight events that fall in bin 2, four fail to pass the m_{jj} cut, but pass all of the cuts related to the Higgs decay topology. Conversely, two events fail the $m_{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ cuts, and pass the VBF cuts. For every event in bin 2, if the event fails a VBF cut, then it passes all Higgs decay cuts, and if an event fails a Higgs decay cut, it passes all VBF cuts, illustrating the source of the increased acceptance. The BDT is doing what it is designed to do—exploiting correlations between the inputs. This finding motivates an alternative approach to the cut-based analysis for the next run. The selections can be split into three categories—loose (tight) Higgs decay (VBF) cuts, tight (loose) Higgs decay (VBF) cuts, tight (tight) Higgs decay (VBF) cuts—with each set of cuts independently optimized. Such an approach is equivalent to a primitive decision tree.

11

Statistical Treatment

Quantum mechanical scattering processes fulfill the requirements of a homogeneous Poisson process [124]. Letting N denote the number of events observed for a given process, the following requirements are satisfied for a Poisson process:

(1) The probability of observing k events at time $t + \Delta t$ is independent of the number of events observed at time t, i.e.

$$\Pr\left[N(t+\Delta t) = k \,|\, N(t) = j\right] = \Pr\left[N(t+\Delta t) = k\right]. \tag{11.1}$$

(2) For a time interval, Δt, the probability of observing a single event depends only on Δt. Mathematically, this can be expressed as

$$\Pr\left[N(t+\Delta t) - N(t) = 1\right] = \lambda \Delta t + O(\Delta t^2), \tag{11.2}$$

where λ is a time-independent constant representing the rate at which the process occurs per unit time.

(3) At $t = 0$, zero events have been observed, i.e. $\Pr\left[N(0) = 0\right] = 1$.

(4) In an infinitesimal time interval, either zero or one event can be observed. With

requirement (2), this implies that

$$\Pr\left[N(t+\Delta t) - N(t) = 0\right] = 1 - \lambda\Delta t - O(\Delta t^2). \tag{11.3}$$

From the above four requirements, it is possible to compute the probability density function (PDF) for N, the number of scattering events observed at fixed time t. The probability of observing 0 events at time $t + \Delta t$ is, according to the above requirements,

$$\begin{aligned}\Pr\left[N(t+\Delta t) = 0\right] &= \Pr\left[N(t) = 0\right] \cdot \Pr\left[N(\Delta t) = 0\right] \\ &= \Pr\left[N(t) = 0\right]\left(1 - \lambda\Delta t - O(\Delta t^2)\right),\end{aligned} \tag{11.4}$$

where the first equality follows from (1) and the second equality follows from (4). Rearranging and taking $\lim \Delta t \to 0$ results in a differential equation for $\Pr\left[N(t) = 0\right]$

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t). \tag{11.5}$$

The notation has been changed from $\Pr\left[N(t) = k\right]$ to $P_k(t)$ to make the time-dependence more clear. Using the initial condition (3), the solution to this equation is $P_0(t) = e^{-\lambda t}$. The differential equation for the more general scenario of observing k events in time t is

$$\frac{dP_k(t)}{dt} = -\lambda\left[P_k(t) - P_{k-1}(t)\right], \tag{11.6}$$

which can be solved recursively starting with the solution for $P_0(t)$, yielding

$$P(N \mid \lambda, t) = \frac{(\lambda t)^N e^{-\lambda t}}{N!}. \tag{11.7}$$

This is the PDF for observing N scattering events for a process that occurs at a rate λ in a duration of time t, also known as the Poisson distribution. The product of λ

and t is the mean number of events expected in time t at a rate of λ and $\sqrt{\lambda t}$ is the standard deviation. Due to the Poisson nature of the scattering process itself, the number of events falling into any phase space region also is also Poisson-distributed.

In a collider experiment, the expected number of events, λt, is generally computed with MC simulation. For a given phase space region, if the predicted number of events for background X is b_X and the prediction for background Y is b_Y, then $P(N|b_X + b_Y) = (b_X + b_Y)^N e^{-(b_X+b_Y)}/N!$ is the probability for observing N events in this region, assuming that the background is properly accounted for by the prediction $b_X + b_Y$. In the situation where N is measured, there is information to be gained about the expected number of events by recasting the PDF as a likelihood function:

$$\mathcal{L}(\mu) = P(N \,|\, \mu s + b) \tag{11.8}$$

This function is no longer a PDF for N. Instead, it is a function of the parameter μ, called the parameter of interest (POI), which scales the signal prediction s. The total background prediction, b, is from MC simulation. If s is the SM prediction, then μ is interpreted as the signal strength $[\sigma \cdot \mathrm{Br}]_{\mathrm{measured}} \,/\, [\sigma \cdot \mathrm{Br}]_{\mathrm{SM}}$. In order to extract the measured μ, the likelihood function is maximized by computing $\partial\mathcal{L}/\partial\mu$, setting it to zero, and solving for $\hat{\mu}$. In the case above, this yields, $\hat{\mu} = (N - b)/s$. If N is measured to be equal to the background-only hypothesis b, then $\hat{\mu} = 0$; if, instead, the observed number of events is consistent with signal produced at the SM rate, $N = s + b$, then $\hat{\mu} = 1$.

11.1 Extended Likelihood Function

The simplified likelihood function defined in Equation 11.8 does not account for the fact that (1) some of the components in b are normalized with control regions, (2) predictions s and b are inherently uncertain (3) the number of events N is generally

measured in several phase space bins for some observable x.

If the normalization of a background is constrained in a control region, the normalization factor (NF) can be incorporated as an additional parameter in the likelihood:

$$\mathcal{L}(\mu, \mu_b) = P(N \,|\, \mu s + \mu_b b^{\mathrm{SR}} + b_{\mathrm{other}}) \cdot P(M \,|\, \mu_b b^{\mathrm{CR}}) \tag{11.9}$$

The second term represents the auxiliary measurement M in the CR, with an expectation which is the product of the MC prediction and a strength parameter: $\mu_b b^{\mathrm{CR}}$. The μ_b parameter then scales the prediction in the SR, b^{SR}. Again, the best-fit values of μ and μ_b are obtained by maximizing the likelihood.

The predictions from MC simulation which enter the likelihood function are inherently uncertain, since they rely on assumptions about theoretical (e.g. renormalization scale) or instrumental (e.g. JES) parameters. Moreover, the number of events falling into the SR in MC simulation is also Poisson-distributed and therefore subject to statistical uncertainties. Because such parameters are not of fundamental interest in the measurement, they are called nuisance parameters (NPs), denoted θ_i. Uncertainties associated with θ_i are generally computed independently and this information is incorporated into the likelihood in two ways. First, the signal and background predictions, denoted generically as E, are written in terms of θ_i:

$$E(\theta_i) = E_0 \nu(\theta_i) \tag{11.10}$$

where $\nu(\theta_i)$ is chosen *a priori* to give $E(\theta_i)$ the expected θ dependence for a given source. Additionally, the uncertainty on θ_i is incorporated with a response term in the likelihood function, denoted $\mathcal{A}(\tilde{\theta}_i \,|\, \theta_i)$, where $\tilde{\theta}_i$ has been determined to be the best guess for θ_i. In the frequentist interpretation, $\mathcal{A}(\tilde{\theta}_i \,|\, \theta_i)$ is the sampling distribution for θ_i, and $\tilde{\theta}_i$ is the mean value.

In the analysis presented in this thesis, three different $\nu(\theta_i)$ and $\mathcal{A}(\tilde{\theta}_i \,|\, \theta_i)$ are

used in the likelihood function, depending on the nature of the uncertainty. For both theoretical and instrumental uncertainties,

$$\begin{aligned} \nu(\theta_i) &= (1+\epsilon)^{\theta_i} \\ \mathcal{A}(\tilde{\theta_i} \,|\, \theta_i) &= \mathcal{G}(\tilde{\theta_i} \,|\, \theta_i, 1) \end{aligned} \tag{11.11}$$

where $\mathcal{G}(\tilde{\theta_i} \,|\, \theta_i, 1)$ is a unit width Gaussian distribution centered at θ_i. The parameter ϵ is obtained by measuring the event yield expectations at their $\pm 1\sigma$ points: $E(1)/E_0 = (1+\epsilon_+)^1$ and $E(-1)/E_0 = (1+\epsilon_-)^{-1}$. Thus the expectation is expressed as

$$E(\theta) = \begin{cases} E_0(1+\epsilon_+)^\theta & \theta \geq 0 \\ E_0(1+\epsilon_-)^\theta & \theta < 0 \end{cases} \tag{11.12}$$

If, for example, the expectation value for an uncertainty source is 3% greater than the nominal expectation value E_0 at $E(1)$, then $(1+\epsilon_+)^1 = E(1)/E_0 = 1.03$, implying that $\epsilon_+ = 3\%$. Therefore, ϵ represents the fractional change in the yields due to a given uncertainty source. With the choices in Equation 11.11, the sampling distribution for the expectation value $E(\theta)$ is a log-normal distribution centered at E_0. An alternative choice is the same $\mathcal{A}(\tilde{\theta_i} \,|\, \theta_i)$ with $\nu(\theta) = (1+\delta\theta)$, which yields an $E(\theta)$ that is normally distributed about E_0 with a width $\sigma = E_0\delta$. If $\theta < -1/\delta$, the expectation value is negative. To avoid such a situation, $E(\theta)$ is set to zero in these cases, resulting in a discontinuity in the expectation value as a function of θ. For this reason, the log-likelihood sampling distribution is chosen. Each NP i is then factorized such that $E(\vec{\theta}) = E_0 \prod_i \nu_i(\theta_i)$.

Similar to the statistical uncertainty on the observed number of data events in the SR, there are statistical uncertainties on the expected number of events due to Poisson fluctuations in the MC sample [125]. The expectation value, E, for a given process can be expressed as the product $\mathcal{L} \cdot \sigma \cdot \epsilon_{\text{SR}}$, where $\mathcal{L}$ is the integrated

luminosity, σ is the process cross section, and ϵ_{SR} is the efficiency for selecting an event in the SR. Re-expressing the efficiency in terms of the raw number of MC events selected in the SR and the total number of events generated, the expectation is

$$E(\vec{\theta}) = \mathcal{L}(\vec{\theta}) \cdot \sigma(\vec{\theta}) \cdot \frac{N_{\mathrm{SR}}(\vec{\theta})}{N_{\mathrm{total}}} \tag{11.13}$$

where the dependence on $\vec{\theta}$ has been explicitly indicated. The random variable N_{SR} is expected to be Poisson-distributed with a mean of $\hat{N}_{\mathrm{SR}}$ and a standard deviation of $\sqrt{\hat{N}_{\mathrm{SR}}}$, where $\hat{N}_{\mathrm{SR}}$ is the measured number of raw events in the SR. Using the fact that $\mathcal{L}$ and σ are constants for this NP, the ratio of $E(\theta)$ to E_0 is equivalent to the ratio of N_{SR} and $\hat{N}_{\mathrm{SR}}$, allowing $E(\theta)$ to be expressed as a fixed constant multiplying E_0: $E(\theta) = E_0\theta$. With this parameterization, the constraint term is set to give the *a priori* sampling distribution:

$$\begin{aligned} \nu(\theta) &= \theta \\ \mathcal{A}(\tilde{\theta}\,|\,\theta) &= P(\hat{N}_{\mathrm{SR}}\,|\,\theta N_{\mathrm{SR}}) \end{aligned} \tag{11.14}$$

Under these definitions, the most likely value of $E(\theta)$ is E_0 and the width Δ_E is $E_0/\sqrt{\hat{N}_{\mathrm{SR}}}$. To keep the number of NPs to a minimum, the parameter for the statistical uncertainty is integrated across all processes in a given bin, yielding a single NP per bin.

The final type of NP—the strength parameter associated with a background that is constrained from a CR—has been discussed (Equation 11.9). To be consistent

with the description of other NPs, the constraint term is written in terms of θ:

$$\begin{aligned} \nu(\theta) &= \theta \\ \mathcal{A}(\tilde{\theta}\,|\,\theta) &= P(\tilde{N}_{\mathrm{CR}}\,|\,N_{\mathrm{CR}}(\theta)) \end{aligned} \tag{11.15}$$

Here, N_{CR} is defined as the expected number of events in the CR: $N_{\mathrm{CR}} = \mu s_{\mathrm{CR}} + \theta b_{\mathrm{CR}}^{\mathrm{target}} + \sum_k b_{\mathrm{CR}}^{\mathrm{other}}$. s_{CR} and $b_{\mathrm{CR}}^{\mathrm{other}}$ are the expected contaminations from signal and backgrounds other than the target background. $\tilde{N}_{\mathrm{CR}}$ is the observed number of events in the CR. In the likelihood expressions that follow, such NPs are explicitly separated from the others to illustrate which backgrounds are constrained with an auxiliary measurement.

Accounting for the systematic and statistical uncertainties, the likelihood function can be written with the additional terms outlined above. Additionally, if the final discriminant in the analysis is binned in more than one bin, a product over the likelihoods for each bin is performed. The likelihood in each bin can be written

$$\mathcal{L}_{\text{bin } i}(\mu, \vec{\mu}_b, \vec{\theta}) = P(N_i\,|\,\mu s(\vec{\theta}) + \sum_j \mu_j b_j(\vec{\theta}) + \sum_k b_k(\vec{\theta})) \cdot \prod_{j=1}^{J} P(M_j\,|\,\mu_j b_j^{\mathrm{CR}}(\vec{\theta})) \cdot \prod_{n=1}^{K} \mathcal{A}(\tilde{\theta_n}\,|\,\theta_n). \tag{11.16}$$

The first term is the likelihood for bin i of the SR having observed N_i events. The second term represents the likelihood for the auxiliary measurement of M_j events in the CR and the final term is the product of the constraint terms for each of the K NPs considered in the analysis. All information in the analysis—the event yields in the SR, NFs from CR measurements, and systematic uncertainties—has been reduced to a single analytic function of the signal strength μ, the NFs $\vec{\mu'}$, and the NPs $\vec{\theta}$. To extract the best-fit μ, the likelihood is maximized with respect to these arguments simultaneously using the MINUIT package [126].

11.2 Hypothesis Testing

In addition to measuring the signal strength, it is crucial to quantify the degree to which the observed data are compatible with some hypothesized μ value. In this analysis, the goal is to quantify the compatibility with the hypothesis that there is not any signal ($\mu = 0$), i.e. the background-only hypothesis.

Because the likelihood function is differentiated to obtain the best-fit value of the POI, its absolute magnitude is arbitrary. Therefore, to extract information from the likelihood, it is necessary to compare two points in parameter space. For example, if there are two competing hypotheses H_a and H_b, the test statistic which is the ratio of the likelihoods for each parameter set, $t = \mathcal{L}(N \mid H_a)/\mathcal{L}(N \mid H_b)$, is greater in magnitude if the data N are more consistent with the parameter set associated with H_a than that of H_b. A test statistic is chosen to maximize the discriminating power between two hypotheses. In other words, if there is some t threshold at which H_a is accepted as the "true" hypothesis, it is important to avoid scenarios in which H_a is accepted as true when in fact H_b is true. Similarly, a test statistic must be chosen to minimize the cases in which H_a is rejected when in reality H_a is true. The Neyman-Pearson lemma [127] states that the ratio of likelihoods maximizes the discriminating power between two hypotheses. Consequently, this analysis utilizes a test statistic of this form. Under the frequentist interpretation, a test is performed by first finding the probability density function for the test statistic under the hypothesis being tested. The p-value for observing the test statistic in the actual realization of the experiment is then computed in light of the distribution of t.

For this analysis, the profile likelihood test statistic (PLTS), given by

$$t_\mu = -2 \log \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \tag{11.17}$$

has been chosen. Given that the log function is monotonic, the Neyman-Pearson lemma still applies for the log of the likelihood ratio. In the numerator of this function, $\hat{\theta}_\mu$ is determined by conditionally maximizing the likelihood with respect to θ with μ fixed to the value for the hypothesis being tested, while in the denominator, the likelihood is maximized with respect to both μ and θ simultaneously.

The PDF for the test statistic, $f(t_\mu \,|\, \mu')$, is determined by either generating pseudo-experiments (PEs) or using an analytic approximation. There is a subtle distinction between μ and μ' in the above notation for the PDF; μ denotes the hypothesis being tested and μ' denotes the μ value assumed in data. In the PE approach, an ensemble of different experimental realizations is generated. Each realization is defined by (1) the number of events observed in the signal region bins N_i, (2) the number of events observed in control regions M_j, and (3) the auxiliary measurement for each NP $\tilde{\theta}_n$. Both (1) and (2) are sampled from Poisson distributions using the expectations shown in Equation 11.16. Similarly, (3) is sampled from the auxiliary constraint term, $\mathcal{A}(\tilde{\theta_n} \,|\, \theta_n)$. To obtain a true sampling distribution, a suitable choice for the parameters (μ, μ_j, θ_n) is needed; otherwise, the mean value of, for example, M_j, is not known, since μ_j is not constrained. The convention is to determine the values of (μ_j, θ_n) from the data in the auxiliary terms, and in the term corresponding to the primary measurement, μ is set to μ'.

With the PDF $f(t_\mu \,|\, \mu')$ from the PEs, the consistency of the observed data with the μ hypothesis is quantified with a p-value, defined as the probability of measuring a t_μ which is greater than or equal to the observed value of the test statistic t_μ^{obs}:

$$p_\mu = \int_{t_\mu^{\text{obs}}}^{\infty} f\left(t_\mu \,|\, \mu'\right) dt_\mu \tag{11.18}$$

If t_μ^{obs} falls in the tail of the $f(t_\mu \,|\, \mu')$ distribution, the integral and therefore p_μ are small, indicating statistical disagreement between the observed data and the μ

hypothesis. In the special case where the aim of the measurement is to reject the background only hypothesis in order to discover a signal, the test statistic is set to t_0, and the p-value is then

$$p_0 = \int_{t_0^{\text{obs}}}^{\infty} f\left(t_0 \,|\, 0\right) dt_0. \tag{11.19}$$

As the observed event yield increases above the expected background, $\hat{\mu}$ increases, and consequently t_0^{obs} increases (Equation 11.17) towards increasingly improbable values of t_0 under the $\mu' = 0$ hypothesis. Taking the integral in Equation 11.19 directly quantifies the probability.

In many cases, an experimenter wishes to know at what level the background only hypothesis is expected to be rejected assuming an alternative hypothesis to be true. Generally, the alternative hypothesis is $\mu' = 1$—i.e. the assumption is that the signal model is true. To compute the expected p_0, two PDFs, $f(t_0 \,|\, 0)$ and $f(t_0 \,|\, 1)$, are determined, and the median t_0 is extracted from $f(t_0 \,|\, 1)$. Denoting this median value as t_0^{exp}, the expected p-value for the alternative hypothesis is computed by replacing t_0^{obs} with t_0^{exp} in Equation 11.19.

The p-value is typically converted to the quantity Z_μ, the number of standard deviations the observation falls above the mean:

$$\begin{aligned} Z_\mu &= \Phi^{-1}\left(1 - p_\mu\right) \\ \Phi(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2} dt. \end{aligned} \tag{11.20}$$

$\Phi(x)$ is the cumulative distribution function (CDF) for the normal distribution, $\mathcal{G}(x \,|\, 0, 1)$. In the particle physics community, in order to claim the discovery of some signal, the null hypothesis is required to be rejected at the 5σ level, or $Z_0 \geq 5$ ($p_0 \leq 2.9 \times 10^{-7}$). "Evidence of" ("observation of") the signal process is claimed if

the null hypothesis is rejected at 3σ (4σ).

Rejecting a hypothesis at the level of 3σ (5σ) requires measuring a p-value to 1 part in 10^3 (10^7). With the PE approach, $f(t_\mu \,|\, \mu')$ is subject to Poisson fluctuations. If the number of events falling above the observed t_μ is denoted $N_{\geq t_\mu^{\text{obs}}}$, then the relative statistical uncertainty on p_μ is $1/\sqrt{N_{\geq t_\mu^{\text{obs}}}}$. For 3σ (5σ) rejection, to achieve a relative uncertainty of 10% on the p-value, approximately 10^5 (10^9) PEs are required, and for 1% uncertainty on p_μ, approximately 10^7 (10^{11}) PEs need to be generated. Each PE generation requires the sampling of K $\mathcal{A}(\tilde{\theta}_n \,|\, \theta_n)$ distributions, where K is the number of NPs, and a Poisson distribution for each primary measurement bin and each auxiliary measurement in order to build the likelihood function. Additionally, each evaluation of t_μ requires two CPU-intensive likelihood minimizations.

In this analysis, there are $O(10^2)$ NPs and $O(10)$ event measurements. Given the sheer number of PEs needed to reject the null hypothesis at 3σ, it is computationally preferable to approximate $f(t_\mu \,|\, \mu')$ with an analytic function. Assuming that the data sample size is N, the best-fit values of the parameters in the likelihood $\hat{\mu}$, $\hat{\mu}_b$, and $\hat{\theta}_i$ are approximately distributed according to a multivariate normal distribution, with corrections at the level of $1/\sqrt{N}$ [128, 129]. From this, it follows that for a single POI, the log-likelihood ratio can be expressed as

$$-2\log\frac{\mathcal{L}(\mu,\hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu},\hat{\theta})} = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + O\left(1/\sqrt{N}\right), \tag{11.21}$$

where $\hat{\mu}$ is a normally distributed random variable with a mean of μ' and width σ, which can be determined from the expectation value of the Fisher information matrix [130], or as described in the following section [131]. Corrections to this analytic expression decrease as $1/\sqrt{N}$, implying that this relationship becomes exact in the limit that the data sample size is infinite. Neglecting the $O(1/\sqrt{N})$ term, $f(t_\mu \,|\, \mu')$

can be derived using the fact that $\hat{\mu}$ is normally distributed.

In this analysis, the test statistic in Equation 11.17 is modified slightly, treating cases in which $\mu \leq \hat{\mu}$ differently from those in which $\mu > \hat{\mu}$:

$$q_\mu = \begin{cases} -2\log \frac{\mathcal{L}\left(\mu,\hat{\theta}_\mu\right)}{\mathcal{L}\left(\hat{\mu},\hat{\theta}\right)} & \hat{\mu} < \mu \\ +2\log \frac{\mathcal{L}\left(\mu,\hat{\theta}_\mu\right)}{\mathcal{L}\left(\hat{\mu},\hat{\theta}\right)} & \hat{\mu} \geq \mu \end{cases} \tag{11.22}$$

This test statistic allows $\hat{\mu}$ to go below the test μ in the likelihood, and is therefore more robust than test statistics which limit the range of $\hat{\mu}$ based on physical grounds. Under the Wald approximation, q_μ can be written

$$q_\mu = \begin{cases} \frac{(\mu-\hat{\mu})^2}{\sigma^2} & \hat{\mu} < \mu \\ \frac{-(\mu-\hat{\mu})^2}{\sigma^2} & \hat{\mu} \geq \mu \end{cases} \tag{11.23}$$

and given that the PDF of $\hat{\mu}$ is $\mathcal{G}(\hat{\mu} \,|\, \mu', \sigma) = 1/(\sqrt{2\pi}\sigma) \exp\left[-(\hat{\mu}-\mu')^2/2\sigma^2\right]$, the PDF of q_μ is

$$f\left(q_\mu \,|\, \mu = \mu'\right) = \begin{cases} \frac{1}{2\sqrt{2\pi}} \frac{1}{\sqrt{q_\mu}} \exp\left[-\frac{1}{2}\left(\sqrt{q_\mu} + \frac{\mu-\mu'}{\sigma}\right)^2\right] & q_\mu > 0 \\ \frac{1}{2\sqrt{2\pi}} \frac{1}{\sqrt{-q_\mu}} \exp\left[-\frac{1}{2}\left(\sqrt{-q_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2\right] & q_\mu \leq 0 \end{cases} \tag{11.24}$$

Instead of computing a p-value from a sampling distribution derived from PEs, it is possible to compute $p(\mu, \mu')$ with the analytic expression for $f(q_\mu \,|\, \mu = \mu')$. A comparison of the sampling distribution for q_0 from PEs to the analytic distribution from the Wald approximation is shown in Figure 11.1 for the simple likelihood function, $\mathcal{L}(s, b) = P(N \,|\, s+b)\mathcal{G}(b_0 \,|\, b, \sigma_b)$, illustrating the validity of the Wald approximation.

Given that the p-value is easily obtained from the CDF of $f(q_\mu \,|\, \mu = \mu')$, it is

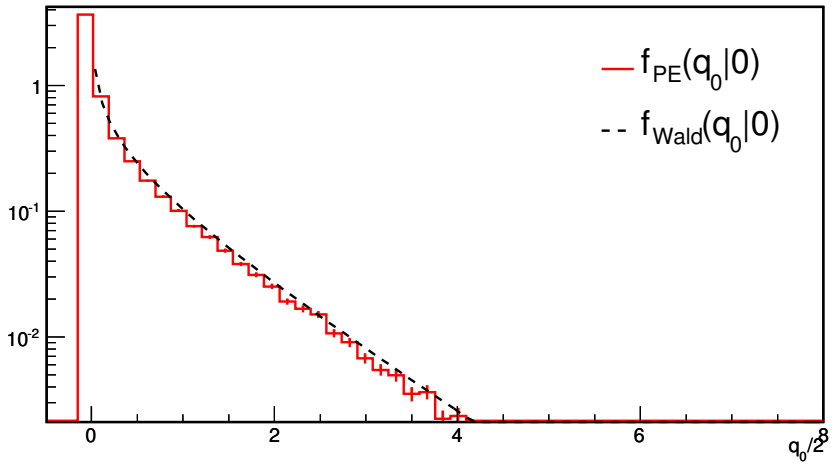


FIGURE 11.1: Comparison of the PDF for the test statistic q_μ derived from pseudo-experiments to that from the Wald approximation for a simple likelihood function, $\mathcal{L}(s,b) = P(N \mid s+b)\mathcal{G}(b_0 \mid b, \sigma_b)$.

useful to compute this function:

$$F\left(q_\mu \mid \mu'\right) = \begin{cases} \Phi\left(\sqrt{q_\mu} + \frac{\mu-\mu'}{\sigma}\right) & q_\mu > 0 \\ \Phi\left(\sqrt{-q_\mu} - \frac{\mu-\mu'}{\sigma}\right) & q_\mu \leq 0 \end{cases} \tag{11.25}$$

where Φ is defined in Equation 11.20. Using this form of the CDF, the p-value for the $\mu = 0$ hypothesis and the data distributed as $\mu' = 0$, the significance of the observed data is simply

$$\begin{aligned} p_0 &= \begin{cases} 1 - \Phi\left(\sqrt{|q_0|}\right) & q_\mu > 0 \\ 1 - \Phi\left(-\sqrt{|q_0|}\right) & q_\mu \leq 0 \end{cases} \\ Z_0 &= \begin{cases} \Phi^{-1}\left(1-p_0\right) = \sqrt{|q_0|} & q_\mu > 0 \\ \Phi^{-1}\left(1-p_0\right) = -\sqrt{|q_0|} & q_\mu \leq 0 \end{cases} \end{aligned} \tag{11.26}$$

Therefore, the only quantity which needs to be computed is the test statistic, q_0,

which can be directly computed from Equation 11.22 without the Wald approximation. In other words, the Wald approximation is only needed to approximate the tail of the sampling distribution of q_μ; it is not used in the evaluation of q_μ.

The observed test statistic, q_0^{obs}, is obtained from the data likelihood. To quantify the expected significance assuming that the signal is produced at the SM rate, a single realization of the experiment, called the Asimov dataset, is defined such that the likelihood is maximized with respect to N_i, M_j, and $\tilde{\theta}_n$ at fixed $\mu = \mu'$, $\mu_j = \hat{\mu}_j$, and $\theta = \tilde{\theta}_n$ [131]. Then the likelihood is defined by the Asimov dataset instead of the observed data, and the PLTS is written in terms of the $\mathcal{L}_A$:

$$\begin{aligned} q_\mu^{\text{Asimov}} &= -2 \log \frac{\mathcal{L}_A(\mu, \hat{\theta}_\mu)}{\mathcal{L}_A(\mu', \tilde{\theta})} \\ &\simeq \frac{(\mu - \mu')^2}{\sigma_A^2} \end{aligned} \tag{11.27}$$

Calculating the expected significance involves computing the median q_0^{Asimov} with the $\mu' = 1$ Asimov dataset, and then using this test point in Equation 11.26, which defines the p-value for the $\mu = 0$, $\mu' = 0$ case.

11.3 The Variance of μ

As discussed, the best-fit μ is determined by maximizing the likelihood with respect to μ, the auxiliary strengths μ_j, and the NPs θ_n simultaneously. To extract the variance on the measured value of μ, the Wald approximation is invoked again [131]. Using the asymptotic test statistic defined in Equation 11.23, the variance of μ can be expressed in terms of μ, $\hat{\mu}$, and q_μ:

$$\sigma(\mu, \mu') = \begin{cases} \frac{|\mu - \hat{\mu}|}{\sqrt{q_\mu}} & \mu > \hat{\mu} \\ \frac{|\mu - \hat{\mu}|}{\sqrt{-q_\mu}} & \mu \leq \hat{\mu} \end{cases} \tag{11.28}$$

In the asymptotic regime, q_μ evaluated at $\mu = \hat{\mu} \pm \sigma$ is ± 1. Therefore, σ can be determined by iteratively computing q_μ and extracting the μ value at which $q_\mu = \pm 1$. This procedure is carried out with the non-asymptotic q_μ, making it a more accurate estimate of σ in cases where the PDF of $\hat{\mu}$ deviates from the normal distribution.

It is often useful to enumerate the degree to which an individual NP contributes to the variance on μ. Fixing all NPs, except the NP of interest (θ_i), to their best-fit values, the likelihood is conditionally maximized with respect to μ and θ_i, and the resulting q_μ is computed. The $\hat{\mu}$ error due to θ_i is extracted as the difference between the μ yielding $q_\mu = 1$ and $\hat{\mu}$.

11.4 The Fit Model

In the VBF BDT analysis, a likelihood of the form shown in Equation 11.16 is defined. The term that represents the measurement in the signal region is a Poisson product over the lepton flavor channels and the three BDT bins:

$$\mathcal{L}_{\mathrm{SR}}(\mu, \vec{\mu}', \vec{\theta}) = \prod_{\text{bin } i}^{1,2,3\ e\mu/\mu e, ee/\mu\mu} \prod_{j} P[N_{ji} \,|\, \mu \cdot s_{ji}(\vec{\theta}) + \sum_{\ell}^{\text{top}, Z/\mathrm{DY}} \mu_i^\ell \cdot b_{ji}^\ell(\vec{\theta}) + \sum_k b_{ji}^k(\vec{\theta})] \quad (11.29)$$

Background processes indexed by ℓ—top quark background and $Z/\mathrm{DY} \to ee/\mu\mu$ processes—are constrained with an auxiliary strength measurement μ_i^ℓ. In the case of top quark background, this parameter is shared between the two channels, while for Z/DY, the strength parameter is applied only to the $ee/\mu\mu$ SR bins. The background predictions indexed by k, b_{ji}^k, are derived from MC simulation, except for backgrounds from fake leptons, which are estimated with a data-driven technique (Section 8.3) that is external to the likelihood. The next term, representing the

auxiliary measurement of the μ_i^ℓ, can be written explicitly as

$$\mathcal{L}_{\mathrm{CR}}(\mu, \vec{\mu}', \vec{\theta}) = \prod_{k}^{\mathrm{top},Z/\mathrm{DY}} \prod_{\mathrm{bin}\ i}^{1,2} P(M_i^k \,|\, \mu_i^k \cdot \beta_i^k) \tag{11.30}$$

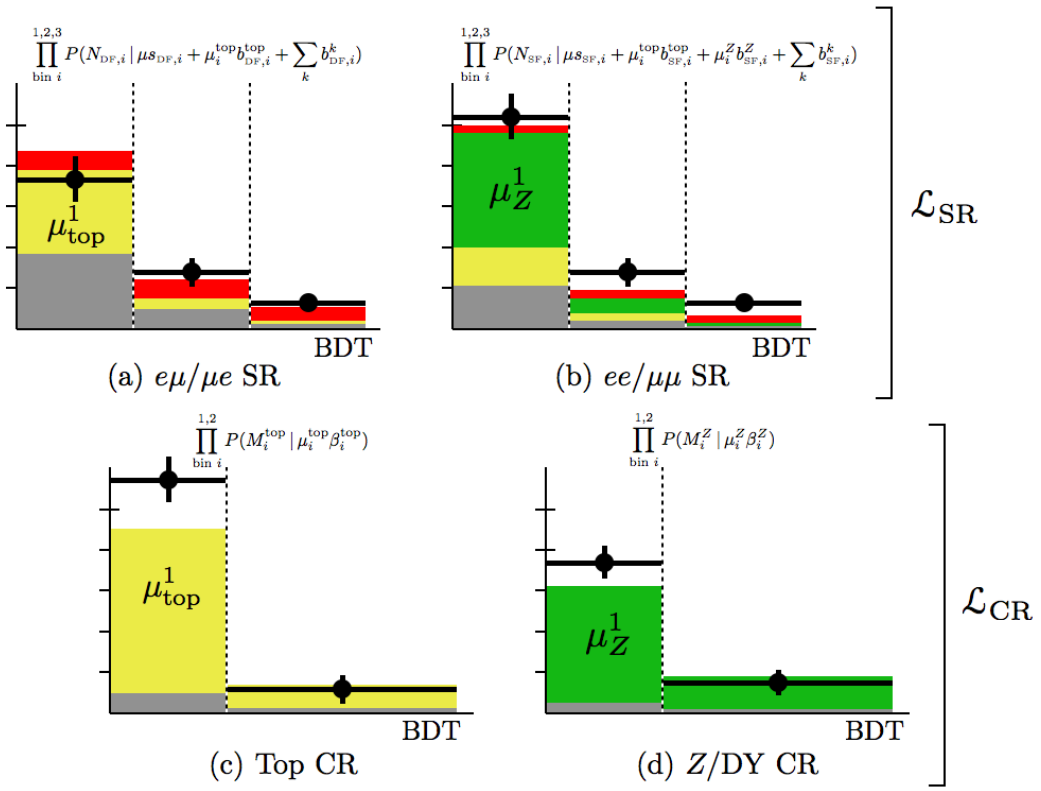
where β is the background k prediction from MC in bin i. The product over bins does not run over the three SR bins, because a common strength parameter is applied in the two high SR bins, i.e. in Equation 11.29, $\mu_2^\ell = \mu_3^\ell$. The likelihood terms defined in Equations 11.29 and 11.30 are shown pictorially in Figure 11.2. The final term constrains the NPs based on auxiliary measurements:

$$\mathcal{L}_{\mathrm{NP}}(\vec{\theta}) = \prod_{i}^{N_{\mathrm{NP}}} \mathcal{G}(\tilde{\theta}_i \,|\, \theta_i, 1) \prod_{j} P(\tilde{\theta}_j \,|\, \theta_j M_j) \tag{11.31}$$

The first term is a product over all NPs which do not represent statistical uncertainties, including both theory and instrumental systematics. Each NP is constrained with a single unit Gaussian term. The additional term, with index j running over all SR and CR bins in Equations 11.29 and 11.30, is the Poisson response to account for statistical uncertainties on the MC prediction. In each bin, all processes are combined for a single term.

The likelihood model for the 7 TeV analysis is adjusted to account for the fact that there are only two (one) BDT bins in the $e\mu/\mu e$ ($ee/\mu\mu$) channel. In addition, the $\mathcal{L}_{\mathrm{NP}}$ term is modified for cases in which the uncertainty sources are not the same as those in the 8 TeV analysis. When the uncertainties are from the same source, they are represented by the same term in $\mathcal{L}_{\mathrm{NP}}$, and are therefore fully correlated.

Higgs production by gluon fusion receives special treatment, as there is information to be gained about the rate of ggF production from other $H \to WW^{(*)} \to \ell\nu\ell\nu$ measurements. In particular, due to the relative dominance of ggF in the zero and



(a) $e\mu/\mu e$ SR (b) $ee/\mu\mu$ SR (c) Top CR (d) Z/DY CR

FIGURE 11.2: Summary of the likelihood terms, $\mathcal{L}_{\mathrm{SR}}$ and $\mathcal{L}_{\mathrm{CR}}$ for the VBF analysis, split into lepton channels. The top CR (c) is used to constrain the $t\bar{t}$and single top normalizations in both flavor channels. The Z/DY CR constrains Z/DY $\rightarrow ee/\mu\mu$ in the $ee/\mu\mu$ channel only.

one jet bins, terms associated with the measurements in these bins are added to the likelihood. Moreover, an additional measurement in the two jet bin, optimized to measure ggF and orthogonal to the VBF regions, is included in the likelihood. In these additional terms, the normalization of ggF is promoted to a nuisance parameter, and correspondingly, the ggF term in $\mathcal{L}_{\mathrm{SR}}$ is changed from b_{ji}^{ggF} to $\mu_{\mathrm{ggF}} \cdot b_{ji}^{\mathrm{ggF}}$. The auxiliary ggF measurements bring with them another set of $\mathcal{L}_{\mathrm{NP}}$ and $\mathcal{L}_{\mathrm{CR}}$ terms, which are collectively denoted as $\mathcal{L}_{\mathrm{other}}(\mu_{VBF}, \mu_{ggF}, \vec{\mu}_b, \vec{\theta})$. The auxiliary μ_{ggF} measurements are subject to the same instrumental systematics as the VBF analysis, motivating the correlation of all such NP terms in the likelihood. In most cases,

uncertainties from theoretical sources are not correlated, as they were evaluated in different phase space regions. The final likelihood can then be written as the product of the likelihoods outlined above:

$$\mathcal{L}_{\text{full}} = \mathcal{L}_{\text{SR}} \cdot \mathcal{L}_{\text{CR}} \cdot \mathcal{L}_{\text{NP}} \cdot \mathcal{L}_{2011} \cdot \mathcal{L}_{\text{other}}. \tag{11.32}$$

11.5 Results

The primary aim of this analysis is to measure the rate at which the Higgs boson is produced by VBF. This is carried out at different Higgs signal mass hypotheses (m_H), and the Higgs predictions from MC simulation are varied depending on the assumed m_H. Because VBF and VH define signal, and ggF is considered a background, both signal and background templates change in the likelihood function for different m_H. The best-fit m_H, as measured in the $H\to ZZ^{(*)}$ and $H\to\gamma\gamma$ channels, is 125.36 GeV. Therefore, this mass point is used as a reference point.

11.5.1 $m_H = 125.36$ GeV

The maximum likelihood estimate of μ_{VBF} is obtained by maximizing $\mathcal{L}_{\text{full}}$, or equivalently, by minimizing the PLTS, q_μ (Equation 11.22). A scan of q_μ as a function of μ is shown in Figure 11.3 for $m_H = 125.36$ GeV. As expected under the Wald approximation, the q_μ curve appears to be parabolic. The minimum is reached at $\hat{\mu} = 1.27$, a value which is larger than the SM expectation due to the excess of observed events in the 8 TeV SRs. The variance of μ is obtained from the scan by interpolating the μ values at which $q_\mu = 1$, resulting in a σ_+ (σ_-) of 0.53 (0.45), or, splitting into statistical and systematic uncertainty components:

$$\boxed{\hat{\mu}_{\text{VBF}} = 1.27^{+0.43}_{-0.39}\ (\text{stat})\ ^{+0.30}_{-0.21}\ (\text{sys}) = 1.27^{+0.53}_{-0.45}\ (\text{total})} \tag{11.33}$$

From the variance, it is evident that $\hat{\mu}$ is consistent with the SM expectation at $m_H =$

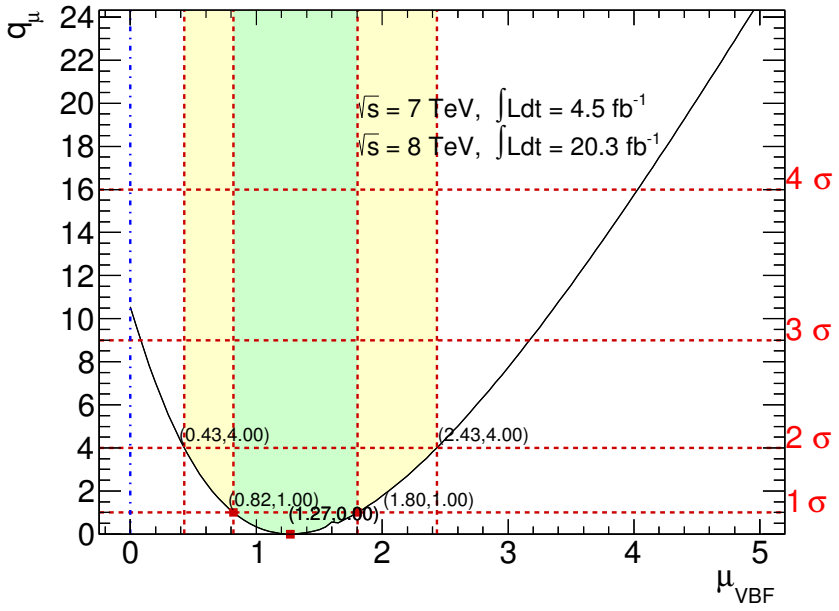


FIGURE 11.3: Scan of the test statistic q_μ at different values of μ at $m_H = 125.36$ GeV. The MLE of μ is obtained from the minimum q_μ, and the total error on $\hat{\mu}$ is obtained from the μ points at which $q_\mu = 1$.

125.36 GeV. The best-fit value of the ggF strength is exactly the SM expectation of 1.0, and the error is 20%, which, as shown in Table 11.2, is one of the largest contributions to the total variance on μ_{VBF}.

In addition to the variance, the test statistic value q_0^{obs} can be determined from the scan. From Equation 11.26, this quantity is just the square of the observed significance with which the null hypothesis is rejected, measured to be

$$\boxed{Z_0^{\text{obs}}\,(Z_0^{\text{exp}}) = 3.2\,(2.7)} \tag{11.34}$$

where the expected significance assuming $\mu' = 1$ is shown parenthetically. This observed significance constitutes evidence of VBF Higgs boson production at $m_H = 125.36$ GeV, as it is over the threshold of 3σ. Consistent with $\hat{\mu}_{\text{VBF}}$, the observed p_0 falls within 1σ of the expected p_0 with $m_H = 125.36$ GeV. To quantify the con-

Table 11.1: The best-fit signal strength and observed and expected significances with respect to the null hypothesis at $m_H = 125.36$ GeV. Results are split by flavor channel and $\sqrt{s}$.

Channel	$\hat{\mu}$	Z_0^{obs}	Z_0^{exp}
7 TeV + 8 TeV	$1.3^{+0.7}_{-0.5}$	3.2	2.7
$e\mu/\mu e$	$1.0^{+0.6}_{-0.5}$	2.4	2.3
$ee/\mu\mu$	$2.0^{+1.0}_{-0.8}$	2.7	1.3
8 TeV only	$1.5^{+0.6}_{-0.5}$	3.6	2.4
$e\mu/\mu e$	$1.2^{+0.6}_{-0.5}$	2.7	2.2
$ee/\mu\mu$	$2.3^{+1.1}_{-0.9}$	2.9	1.3
7 TeV only	$-1.0^{+1.4}_{-1.1}$	-1.0	0.9

sistency with the SM, $Z_1^{\text{Obs.}}$ is computed, with the result of 0.6σ. In other words, assuming SM VBF production, an experimental realization which is equally—or less—probable than the one observed will be obtained in 28% of identical experiments. In contrast, the probability of the observation, assuming that the data are distributed according to $\mu = 0$, is 7×10^{-4}. In Table 11.1, Z_0 is split by flavor channel and $\sqrt{s}$. The excess over the $\mu = 1$ hypothesis is concentrated in the 8 TeV $ee/\mu\mu$ channel, where Z_0^{obs} (Z_0^{exp}) is 2.9 (1.3). This excess is mitigated by the deficit in the 7 TeV analysis, in which $Z_0^{\text{obs}} = -1.0$ with $Z_0^{\text{exp}} = 0.9$.

In Table 11.2, the top 15 NPs are ranked according to their impact on $\hat{\mu}$. The impact of statistical uncertainties is significantly larger than those associated with systematic uncertainties. In fact, statistical uncertainties account for 85% of the total error on $\hat{\mu}$.

11.5.2 Mass scan results

Although m_H has been measured in high resolution Higgs boson decay channels, the rate of VBF Higgs production is independently measured as a function of m_H in the $WW^{(*)} \to \ell\nu\ell\nu$ channel. In Table 11.3, the best-fit VBF and ggF signal strengths are

shown for each test mass. At low m_H, both $\hat{\mu}_{\mathrm{VBF}}$ and $\hat{\mu}_{\mathrm{ggF}}$ are significantly larger than the SM value, a consequence of the fact that the $\mathcal{B}_{H\to WW^{(*)}}$ is an order of magnitude smaller than at the reference point. Without the ability to fully reconstruct the Higgs mass from the decay products, the m_H resolution suffers. Consequently, any excess over the null hypothesis at $m_H = 125.36$ GeV leaks to neighboring mass points. This data excess, in concert with a smaller VBF and ggF predictions at low m_H, pulls the best-fit strength parameters up.

With $\mathcal{B}_{H\to WW^{(*)}}$ reaching a maximum at $m_H \sim 2m_W$, the errors on μ_{VBF} and μ_{ggF} are minimized at m_H points in this region. μ_{ggF} is measured to be consistent with zero. μ_{VBF}, on the other hand, is consistently larger than 0, reaching a minimum at $m_H = 160$ GeV of 0.43 ± 0.13. This is an artifact of the especially poor mass resolution in the VBF analysis. In the optimization of the BDT, mass resolution has not been prioritized. Any mass resolution contribution from the BDT input m_{T} is washed out by the other BDT inputs and the coarse binning of the BDT response. Therefore, the signal template does not vary substantially with m_H, resulting in a large smearing effect.

The observed p_0 is shown in Figure 11.4 as a function of m_H. For comparison, the expected p_0 with the SM signal expectation at $m_H = 125.36$ injected into the null hypothesis is also shown. The observed p_0 closely follows the expectation from an $m_H = 125.36$ GeV signal—though with an offset due to the excess—over the entire mass range. A maximum significance of 3.9σ is reached at $m_H = 190$. Again, this is not surprising given the poor mass resolution and the fact that a dominant background, ggF, is being scaled by its best-fit value of 5×10^{-5}, the lowest measured $\hat{\mu}_{\mathrm{ggF}}$ value across the mass range. Similarly, at $m_H > 160$ GeV, the expected significance for the injected signal is greater than the expectation at the reference mass.

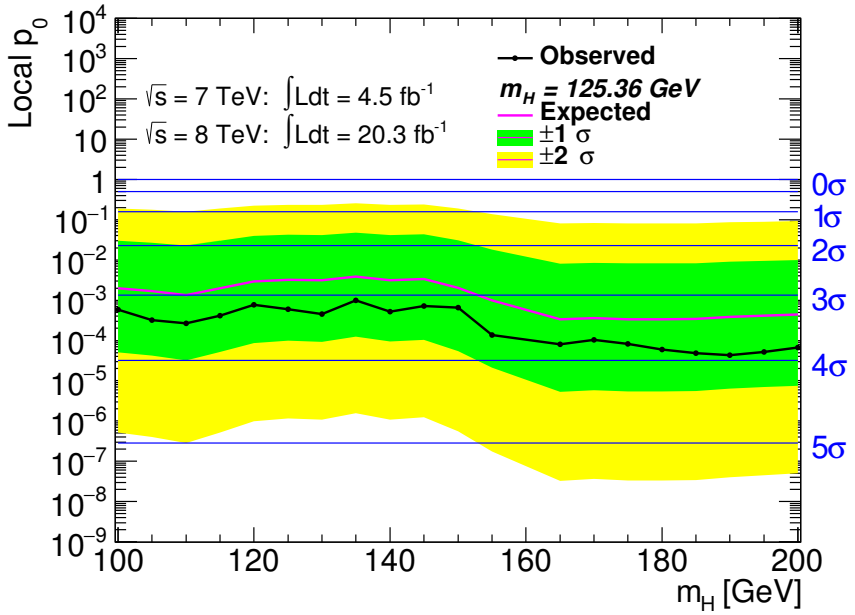


FIGURE 11.4: Significance plotted as a function of the Higgs mass hypothesis. The observed p-value (p_0^{obs}) is shown in black, and p_0^{exp} with signal injected at m_H = 125.36 GeV is shown in magenta. The 1σ (2σ) bands for the p_0^{exp} are shown in green (yellow). The associated Z-score is shown in blue on the right vertical axis.

11.6 Conclusions

At the measured Higgs boson mass of 125.36 GeV, evidence for vector boson fusion Higgs production has been obtained in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ channel at a significance of 3.2σ. The expected significance at this m_H is 2.7σ, indicating that the observed excess is within 1σ of the SM expectation for VBF production. A rigorous analysis optimization—most importantly, the use of a boosted decision tree—has produced an expected sensitivity improvement of 70% with respect to the last published VBF result in this channel on the same dataset [66]. The signal strength for VBF Higgs production has been measured to be $1.27^{+0.53}_{-0.45}$, corresponding to an inclusive $\sigma^{8\ \text{TeV}}_{\text{VBF}} \cdot \mathcal{B}_{H\to WW^{(*)}}$ of $0.51^{+0.22}_{-0.17}$ pb. This measurement is in agreement with the SM prediction of 0.34 pb.

Table 11.2: The top 15 nuisance parameters ranked by impact on $\hat{\mu}$ error (m_H = 125.36 GeV), as well as the total errors split into statistical and systematic components. Absolute errors at $\pm 1\sigma$ are shown.

NP source	$\Delta\mu_{+\sigma}$	$\Delta\mu_{+\sigma}$
Statistical	+0.43	-0.39
JES η model	+0.12	-0.08
Signal parton shower	+0.12	-0.07
μ_{ggF}	+0.09	-0.08
$\mathcal{B}_{H\to WW^{(*)}}$	+0.09	-0.05
ABCD non-closure	+0.06	-0.05
Luminosity (8 TeV)	+0.06	-0.04
WW QCD scale	+0.05	-0.05
Top α	+0.05	-0.05
Signal ME model	+0.07	-0.04
light tag SF	+0.06	-0.04
Signal pdf	+0.07	-0.04
ggF QCD scale	+0.04	-0.04
Top ME model (ggF)	+0.04	-0.03
ggF Higgs p_{T} (0,1j)	+0.04	-0.03
Total Statistical	+0.43	-0.39
Total Systematic	+0.31	-0.22
Total	+0.53	-0.45

Table 11.3: Best-fit μ_{VBF} and μ_{ggF} at each m_H. Errors for the POI μ_{VBF} are computed with the procedure described in Section 11.3, while the error on μ_{ggF} is derived from the Fisher information matrix.

m_H (GeV)	$\hat{\mu}_{\mathrm{VBF}}$	$\hat{\mu}_{\mathrm{ggF}}$
100	$28.5^{+11.5}_{-9.91}$	12.96 ± 5.79
105	$14.0^{+5.59}_{-4.75}$	6.35 ± 1.88
110	$6.66^{+2.68}_{-2.26}$	3.68 ± 0.79
115	$3.45^{+1.42}_{-1.21}$	2.33 ± 0.44
120	$1.94^{+0.83}_{-0.7}$	1.52 ± 0.29
125	$1.31^{+0.54}_{-0.47}$	1.03 ± 0.19
125.36	$1.27^{+0.53}_{-0.45}$	1.0 ± 0.19
130	$0.92^{+0.38}_{-0.33}$	0.78 ± 0.14
135	$0.69^{+0.29}_{-0.25}$	0.57 ± 0.12
140	$0.59^{+0.24}_{-0.21}$	0.47 ± 0.11
145	$0.51^{+0.21}_{-0.18}$	0.4 ± 0.09
150	$0.46^{+0.18}_{-0.16}$	0.29 ± 0.09
155	$0.48^{+0.18}_{-0.15}$	0.13 ± 0.06
160	$0.43^{+0.15}_{-0.14}$	0.05 ± 0.04
165	$0.44^{+0.15}_{-0.13}$	0.03 ± 0.03
170	$0.48^{+0.17}_{-0.15}$	0.02 ± 0.04
175	$0.57^{+0.2}_{-0.17}$	0.01 ± 0.04
180	$0.68^{+0.23}_{-0.2}$	0.0 ± 0.05
185	$0.87^{+0.29}_{-0.26}$	0.0 ± 0.06
190	$0.98^{+0.34}_{-0.3}$	0.0 ± 0.07
195	$1.16^{+0.4}_{-0.35}$	0.0 ± 0.09
200	$1.28^{+0.44}_{-0.39}$	0.0 ± 0.09

12

Conclusion

This thesis has presented a test of the proposed mechanism for preserving the electroweak $SU(2) \times U(1)$ gauge symmetry of the Standard Model of particle physics. Specifically, a measurement of Higgs boson production via vector boson fusion in the $WW^{(*)} \to \ell\nu\ell\nu$ decay channel was described. This measurement was carried out with data collected by the ATLAS detector in Run-1 of the LHC, in which protons were collided at center-of-mass energies of 7 and 8 TeV. A boosted decision tree was used to classify signal and background events using eight kinematic variables. By exploiting the correlations between the inputs, the BDT outperformed traditional cut-based techniques. Evidence for VBF Higgs boson production has been obtained at the 3.2σ level at the best-fit Higgs mass of $m_H = 125.36$ GeV, with an expected significance of 2.7σ. Moreover, the cross section times branching ratio has been measured to be $0.51^{+0.22}_{-0.17}$ pb with the 8 TeV dataset, which is in agreement with the SM prediction of 0.34 pb. This measurement is limited by statistical uncertainties, and will benefit from the additional data collected in Run-2.

Appendices

A

Boosted Decision Trees

This appendix describes the boosted decision tree algorithm and its use in the VBF $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis.

A.1 What is a boosted decision tree?

Boosted decision trees (BDTs) fall into a class of algorithms called machine learning algorithms. In these algorithms, a computer constructs a model of the relationship between independent variables, $\boldsymbol{x}$, and some dependent variable, y, in a process called training. These algorithms can be used for classification of objects that fall into different classes (i.e., signal vs. background) or for regression in the case of a continuously-distributed, real-valued y. Machine learning techniques, and in particular decision trees, are increasingly used in particle physics for particle identification and event classification [132]. At the LHC, BDTs have been used by both ATLAS and CMS [133, 134, 135].

A.2 Decision tree learners

A decision tree is a sequence of selection cuts on the independent variables, which is shown schematically in Fig A.1. In order to grow a decision tree, a training set of events with known $\boldsymbol{x}$ and y is used to construct the model $F(\boldsymbol{x}) = y$. In the case of the VBF analysis, there are two classes, signal, assigned to a numerical value of $y = +1$ and background, $y = -1$, and the training set is fully simulated Monte Carlo samples. The training algorithm starts at the root node, and loops through the $\boldsymbol{x}$, finding the variable that best separates signal and background. The separation is quantified by the Gini index, given by the sum

$$G = \sum_i p_i(1 - p_i), \tag{A.1}$$

where $p_1 = \frac{S}{S+B}$ and $p_2 = \frac{B}{S+B}$. This quantity is designed to be minimized if $p_i = 0$ or $p_i = 1$, and therefore placing a cut on the variable that minimizes it yields orthogonal subsamples, called daughter nodes, that are predominantly either signal or background. The procedure is then repeated at the two daughter nodes, yielding two more cuts and a total of four sub-samples. The process is repeated until a minimum number of events per node or maximum tree depth is reached. The events in the final nodes, are assigned a value $y \in \{-1, 1\}$, depending on the class with more events in the final node.

The discriminating power of a decision tree classifier lies in its ability to take advantage of the correlations among the independent variables. To illustrate this, consider a toy example in which there are two variables, x_1 and x_2, plotted for signal and background in Figure A.2, where the high degree of correlation is visible. Performing a cut optimization on the two input distributions, and choosing a signal acceptance working point of 50% results in the selected box shown in Figure A.2(a).

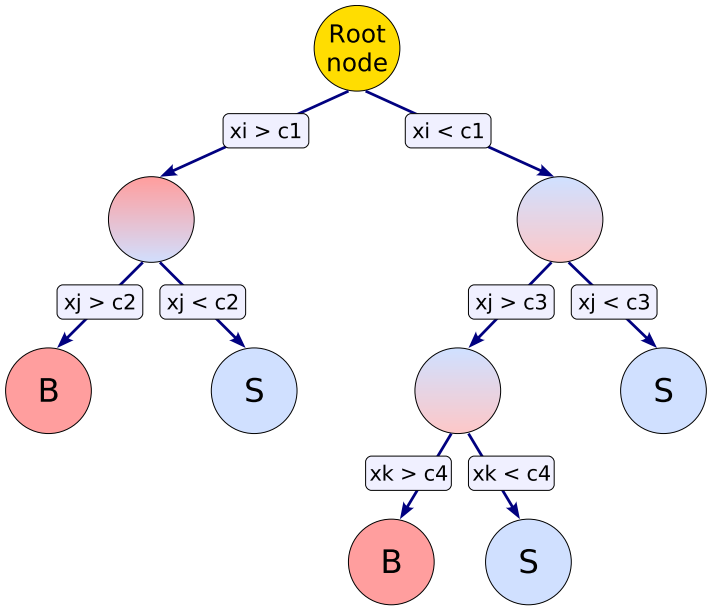


FIGURE A.1: Schematic of a boosted decision tree. [136]

Because a decision tree is a sequence of binary cuts, it can be conceptualized as a collection of boxes in phase space. If the correlations between input variables are different for signal and background, as is the case for this example, then the decision tree partitions the phase space into several signal-rich regions, thereby enhancing the separation power. This is shown in Figure A.2(b), where the decision tree picks out a second signal-rich region. Setting the signal acceptance to the same working point, the decision tree results in a signal purity, $\frac{S}{S+B}$, of 92%, while the signal purity with square cuts is 82%.

In general for classification problems, it is uncommon to use a single decision tree, since this approach is prone to a phenomenon called overtraining. Overtraining refers to the situation in which the model constructed during the training process does not generalize to new data with which a prediction is made. Decision trees are particularly susceptible to overtraining because the split criteria at every node depends on all of the preceding nodes. Therefore, if there is a statistical fluctuation

or some other systematic noise in the data, it will be propagated in the growth of the tree. One procedure for mitigating overtraining is the use of an ensemble of trees, often called a forest.

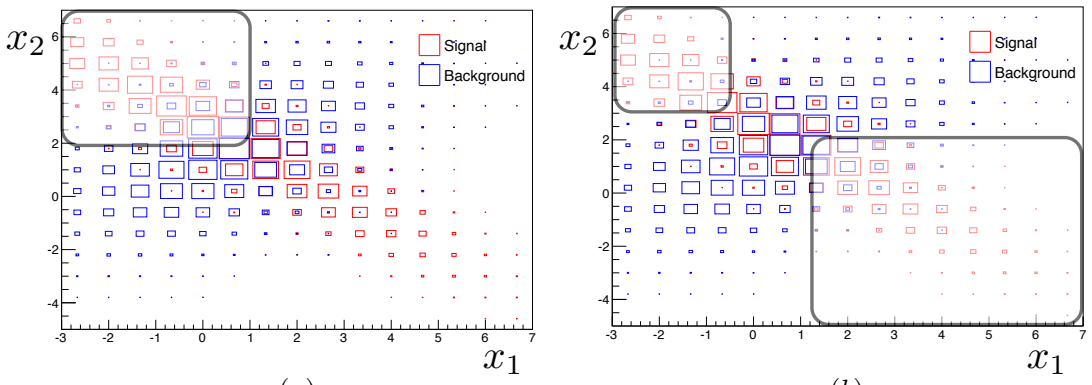


FIGURE A.2: Illustration of improved separation power with a decision tree in the presence of correlated inputs.

A.3 Boosting

In the VBF analysis, a procedure known as boosting is used to grow the decision tree forest. Boosting algorithms grow an ensemble of "base classifiers" sequentially, with the training of each classifier depending on the events that were misclassified by the previous learner. Events that are misclassified are given a higher weight in the next iteration in order for the splitting algorithm to focus on these events. It has been shown that boosting performs well with weak learners [137], or classifiers that do not perform well independently. In the case of decision trees, these base classifiers can be single binary cuts, or a tree with a single node that partitions the space into two subregions.

The most commonly used boosting algorithm, proposed by Freund and Schapire in 1995, is called *AdaBoost* [138]. Suppose there are N events in the training sample,

and the target ensemble size is M trees. In the first iteration of the *AdaBoost* algorithm, each event is weighted democratically with $w_n^1 = 1/N$, and the base classifier is then built with this training set. For the resulting classifier, the misclassification fraction is computed as

$$\epsilon_m = \frac{\sum_{n=1}^{N} w_n^m (1 - \delta(f_m(\boldsymbol{x}_n) - y_n))}{\sum_{n=1}^{N} w_n^m} \tag{A.2}$$

where n indexes the event, m indexes the base classifier, and the $(1-\delta(f_m(\boldsymbol{x}_n)-y_n))$ factor is 0 if the event is corrected classified and 1 if it is misclassified. The weight for event n in the subsequent base classifier is then updated according to the equation

$$w_n^{m+1} = w_n^m \exp\left[\alpha_m (1 - \delta(f_m(\boldsymbol{x}_n) - y_n))\right] \tag{A.3}$$

where $\alpha_m = \log\left(\frac{1-\epsilon_m}{\epsilon_m}\right)$. If the event is correctly classified, the weight is unchanged, while for a misclassified event, the weight in the next training will be larger and therefore, the training algorithm will place an emphasis on this event. The final model is expressed as a weighted linear combination of the base classifiers $F(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m f_m(\boldsymbol{x})$.

The *AdaBoost* algorithm can be reformulated as the minimization of a loss function [139]. The loss function $L(F(\boldsymbol{x}, y))$ measures the deviation between the model and the true value of y in the training set. The specific form of the loss function depends on the classification problem, and for the *AdaBoost* algorithm, it can be shown that the loss function is given by

$$L(F(\boldsymbol{x}), y) = \sum_{n=1}^{N} \exp\left(-F(\boldsymbol{x}_n) y_n\right). \tag{A.4}$$

where $y_n \in \{-1, 1\}$. Again re-expressing $F(\boldsymbol{x})$ as a linear combination of base

classifiers

$$F(\boldsymbol{x}) = \sum_{m=1}^{M} \beta_m f_m(\boldsymbol{x}) = \sum_{m=1}^{M} \beta_m f(\boldsymbol{x}; \boldsymbol{a}_m) \tag{A.5}$$

where $\boldsymbol{a}_m$ defines the parameters of the base classifier (in the case of decision trees, the split variables and values), the task is then to minimize the loss function with respect to the weights β_m and the parameter vectors $\boldsymbol{a}_m$. The minimization is done numerically with the steepest descent algorithm, using the derivative of the loss function, and for this reason, this type of boosting is called "gradient boosting". The *AdaBoost* loss function grows exponentially for large negative values of $F(\boldsymbol{x}_n)y_n$, making it less robust in the face of statistical or systematic outliers. In the VBF analysis, the loss function was chosen to be the negative binomial log-likelihood

$$L(F(\boldsymbol{x}), y) = \log(1 - \exp\left(-F(\boldsymbol{x}_n)y_n\right)). \tag{A.6}$$

Since this function grows linearly with $-F(\boldsymbol{x}_n)y_n$, it performs better with noisy data.

A.4 BDT in VBF $H \to WW^{(*)} \to \ell\nu\ell\nu$

The VBF $H \to WW^{(*)} \to \ell\nu\ell\nu$ final state is composed many of correlated physics objects. In selecting a signal-rich phase space region, cuts are applied on Higgs decay product kinematic distributions, as well as the jet kinematic distributions characteristic of the VBF production mode. The general approach with a BDT is to loosen or remove the rectangular selection cuts and instead train a BDT to find where the signal lies in phase space. In the case of VBF $H \to WW^{(*)} \to \ell\nu\ell\nu$, this is expected to gain in performance over rectangular cuts because the BDT will find an optimal selection in which the VBF topological cuts are loosened with tightened WW decay cuts and vice versa, gaining signal acceptance and improving background

rejection.

A.4.1 Optimization

The software package Toolkit for Multivariate Analysis (TMVA) [136] is used for the implementation of gradient boosting with binary decision trees as the base classifiers. As mentioned above, the negative binomial log-likelihood loss function is used in this implementation. Although the TMVA BDT performs quite well out-of-box, optimization was performed to maximize the expected statistical sensitivity of the analysis. The BDT was optimized with respect to the BDT inputs, the tunable parameters of the BDT, and the preselection cuts applied before and after the training. Due to the computationally intensive nature of training and quantifying the sensitivity, the minimizations along these DOFs were done sequentially as opposed to simultaneously.

The first optimization was done on the BDT inputs, starting with a set of 27 inputs. The inputs in this initial set were identified using a combination of physics intuition– i.e. knowledge of kinematic differences between the dominant backgrounds and signal– past experience from the rectangular cut-based VBF analysis [66], and an exploration of distributions shown to perform well in other MVA-based Higgs analyses [134]. The 27 inputs were then used to train a BDT, and the BDT performance was quantified by computing the Gaussian significance estimate, $\frac{S}{\sqrt{B}}$, obtained after cutting on the BDT response distribution at a value corresponding to a fixed signal acceptance. One of the BDT inputs was then pruned away, and the BDT retrained, yielding a new significance estimate. This procedure was repeated for each of the 50 inputs, yielding 50 unique BDTs, each with 49 input distributions. The input that resulted in the least degradation in the sensitivity was considered extraneous and removed from the set. This process was repeated iteratively until the degradation in the sensitivity with respect to the previous iteration reached some threshold. When

the BDT reached 11 inputs, the figure of merit was changed from the Gaussian significance estimate to the expected sensitivity computed with the full statistical framework (described in detail in Chapter 11) to give a more accurate estimate of the performance of a given BDT. The BDT inputs were pruned from a set of 50 to a set of 8 before significant losses were observed. The 8 BDT inputs are introduced in Section 7.3.

In the TMVA BDT implementation, there are parameters which can be adjusted to achieve maximal performance and robustness. The optimal BDT parameter set is a function of the number of inputs, the number of events in the training set, and the shapes of the input distributions. There is not a general prescription for tuning a BDT for a given training set. For this reason, a brute force optimization was performed on the BDT parameters. The DOFs in the parameter space were the number of trees in the forest, the maximum tree depth, the minimum number of events per node, the shrinkage, and finally the bagging fraction. As discussed above, the forest paradigm is introduced to reduce the possibility of overtraining. The number of trees in the forest, the maximum tree depth and the minimum number of events per node can all be tuned to achieve a model that is not over-fitted, while maintaining the performance of the BDT. The shrinkage parameter, also known as the learning rate, controls the degree to which misclassified events are weighted in the boosting algorithm. Bagging refers to the technique whereby a subset of the total training set is used in the training of a tree. The events in the subset are re-sampled with replacement for each tree in the forest. The bagging fraction parameter in TMVA is the fraction of the total events that is used in the training subset. For the optimization, each parameter was scanned, and at each point, the expected sensitivity was computed, using the same technique as the input optimization. The BDT parameters showing minimal overtraining with optimal performance are shown in Table A.1.

Table A.1: Summary of the optimized BDT settings.

Boost algorithm	gradient
Loss function	$\log(1 - \exp(-F(\boldsymbol{x}_n)y_n))$
Number of trees	1000
Maximum tree depth	5
Minimum number of events per node	1000
Shrinkage (learning rate)	0.125
Bagging fraction	0.25

Bibliography

[1] *CERN Yellow Report Page BR3*, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR3. Accessed 2014-12-11.

[2] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur.Phys.J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[3] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].

[4] ATLAS Collaboration Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[5] ATLAS Collaboration Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, .

[6] J. P. Araque, N. Castro, V. Drugakov, C. Santoni, and P. Starovoitov, *Noise in D5 and D6 cells of the Tile calorimeter at sqrts = 8TeV and 25-ns bunch spacing*, Tech. Rep. ATL-COM-TILECAL-2013-035, CERN, Geneva, Jul, 2013.

[7] ATLAS Collaboration Collaboration, G. Aad et al., *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, arXiv:1407.5063 [hep-ex].

[8] *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, Tech. Rep. ATLAS-CONF-2014-032, CERN, Geneva, Jun, 2014.

[9] ATLAS Collaboration Collaboration, G. Aad et al., *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, arXiv:1407.3935 [hep-ex].

[10] ATLAS Collaboration, *JES Uncertainties from $\sqrt{s} = 8$ TeV data*, . https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2013JESUncertainty.

[11] ATLAS Collaboration Collaboration, G. Aad et al., *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, arXiv:1406.0076 [hep-ex].

[12] ATLAS Collaboration Collaboration, G. Aad et al., *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, Eur.Phys.J. **C73** (2013) 2306, arXiv:1210.6210 [hep-ex].

[13] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, Tech. Rep. ATLAS-CONF-2013-083, CERN, Geneva, Aug, 2013.

[14] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, Tech. Rep. ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014.

[15] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2014-004, CERN, Geneva, Feb, 2014.

[16] ATLAS Collaboration, *Pile-up Suppression in Missing Transverse Momentum Reconstruction in the ATLAS Experiment in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV*, Tech. Rep. ATLAS-CONF-2014-019, CERN, Geneva, May, 2014.

[17] P. A. Dirac, *The Quantum theory of electron*, Proc.Roy.Soc.Lond. **A117** (1928) 610–624.

[18] W. Pauli, *Relativistic Field Theories of Elementary Particles*, Rev.Mod.Phys. **13** (1941) 203–232.

[19] H. Georgi, *Lie Algebras in Particle Physics: From Isospin to Unified Theories*, Front.Phys. **54** (1982) 1–255.

[20] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass*, Phys.Rev. **130** (1963) 439–442.

[21] Y. Nambu, *Axial vector current conservation in weak interactions*, Phys.Rev.Lett. **4** (1960) 380–382.

[22] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cim. **19** (1961) 154–164.

[23] J. Goldstone, A. Salam, and S. Weinberg, *Broken Symmetries*, Phys.Rev. **127** (1962) 965–970.

[24] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys.Rev.Lett. **13** (1964) 321–323.

[25] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys.Lett. **12** (1964) 132–133.

[26] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys.Rev.Lett. **13** (1964) 508–509.

[27] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys.Rev.Lett. **13** (1964) 585–587.

[28] S. Weinberg, *A Model of Leptons*, Phys.Rev.Lett. **19** (1967) 1264–1266.

[29] A. Salam, *Weak and electromagnetic interactions*, . Proc. of the 8th Nobel Symposium on 'Elementary particle theory, relativistic groups and analyticity', Stockholm, Sweden, 1968, edited by N. Svartholm, p.367-377.

[30] Halzen, Francis and Martin, Alan, *Quarks & Leptons: An introductory course in modern particle physics*. John Wiley & Sons, 1984.

[31] Peskin, Michael and Schroeder, Daniel, *An Introduction to Quantum Field Theory*. Perseus Books Publishing, 1995.

[32] M. Srednicki, *Quantum Field Theory*. Cambridge University Press, 2007.

[33] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, Phys.Rev.Lett. **30** (1973) 1343–1346.

[34] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys.Rev.Lett. **30** (1973) 1346–1349.

[35] J. M. Campbell, J. Huston, and W. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, Rept.Prog.Phys. **70** (2007) 89, `arXiv:hep-ph/0611148 [hep-ph]`.

[36] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl.Phys. **B126** (1977) 298.

[37] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, Phys.Rept. **457** (2008) 1–216, arXiv:hep-ph/0503172 [hep-ph].

[38] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Ludwig, et al., *Updated Status of the Global Electroweak Fit and Constraints on New Physics*, Eur.Phys.J. **C72** (2012) 2003, arXiv:1107.0975 [hep-ph].

[39] CMS Collaboration, *Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV*, Tech. Rep. CMS-PAS-HIG-13-005, CERN, Geneva, 2013.

[40] ATLAS Collaboration Collaboration, G. Aad et al., *Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data*, Phys.Rev. **D90** (2014) 052004, arXiv:1406.3827 [hep-ex].

[41] *CERN Yellow Report Page At* 8 TeV, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt8TeV. Accessed 2014-12-11.

[42] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[43] A. Abdesselam, T. Akimoto, P. Allport, J. Alonso, B. Anderson, et al., *The barrel modules of the ATLAS semiconductor tracker*, Nucl.Instrum.Meth. **A568** (2006) 642–671.

[44] ATLAS TRT Collaboration Collaboration, E. Abat et al., *The ATLAS TRT barrel detector*, JINST **3** (2008) P02014.

[45] E. Abat, E. Arik, M. Arik, N. Becerici, O. Dogan, et al., *The ATLAS TRT end-cap detectors*, JINST **3** (2008) P10003.

[46] ATLAS Electromagnetic Barrel Liquid Argon Calorimeter Group Collaboration, B. Aubert et al., *Construction, assembly and tests of the ATLAS electromagnetic barrel calorimeter*, Nucl.Instrum.Meth. **A558** (2006) 388–418.

[47] M. Andrieux, B. Belhorma, A. Belymam, D. Benchekroun, R. Cherkaoui, et al., *Construction and test of the first two sectors of the ATLAS barrel liquid argon presampler*, Nucl.Instrum.Meth. **A479** (2002) 316–333.

[48] ATLAS TileCal Collaboration Collaboration, J. Abdallah et al., *Design, construction and installation of the ATLAS hadronic barrel scintillator-tile calorimeter*, .

[49] D. Gingrich, G. Lachat, J. Pinfold, J. Soukoup, D. Axen, et al., *Construction, assembly and testing of the ATLAS hadronic end-cap calorimeter*, JINST **2** (2007) P05005.

[50] A. Artamonov, D. Bailey, G. Belanger, M. Cadabeschi, T. Chen, et al., *The ATLAS forward calorimeters*, JINST **3** (2008) P02010.

[51] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, et al., *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, .

[52] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl.Instrum.Meth. **A262** (1987) 444–450.

[53] R. Duda and P. Hart, *Use of the Hough Transformation to Detect Lines and Curves in Pictures*, Comm. ACM **15** (1972) .

[54] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2010-069, CERN, Geneva, Jul, 2010.

[55] G. Piacquadio, K. Prokofiev, and A. Wildauer, *Primary vertex reconstruction in the ATLAS experiment at LHC*, J.Phys.Conf.Ser. **119** (2008) 032033.

[56] ATLAS Collaboration, *Characterization of Interaction-Point Beam Parameters Using the pp Event-Vertex Distribution Reconstructed in the ATLAS Detector at the LHC*, Tech. Rep. ATLAS-CONF-2010-027, CERN, Geneva, May, 2010.

[57] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, et al., *Calorimeter clustering algorithms: Description and performance*, .

[58] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, Tech. Rep. ATLAS-CONF-2012-047, CERN, Geneva, May, 2012.

[59] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **0804** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[60] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys.Lett. **B641** (2006) 57–61, arXiv:hep-ph/0512210 [hep-ph].

[61] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur.Phys.J. **C72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[62] ATLAS Liquid Argon EMEC/HEC Collaboration Collaboration, C. Cojocaru et al., *Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region* $1.6 < |\eta| < 1.8$ *in beam tests*, Nucl.Instrum.Meth. **A531** (2004) 481–514, arXiv:physics/0407009 [physics].

[63] ATLAS Collaboration Collaboration, G. Aad et al., *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, Eur.Phys.J. **C73** (2013) 2304, arXiv:1112.6426 [hep-ex].

[64] G. Piacquadio and C. Weiser, *A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS*, J.Phys.Conf.Ser. **119** (2008) 032032.

[65] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at* 8 TeV, Tech. Rep. ATLAS-CONF-2013-082, CERN, Geneva, Aug, 2013.

[66] ATLAS Collaboration, *Measurements of the properties of the Higgs-like boson in the* $WW^{(*)} \to \ell\nu\ell\nu$ *decay channel with the ATLAS detector using 25* fb^{-1} *of proton-proton collision data*, Tech. Rep. ATLAS-CONF-2013-030, CERN, Geneva, March, 2013.

[67] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **1002** (2010) 037, arXiv:0911.5299 [hep-ph].

[68] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput.Phys.Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[69] M. Ciccolini, A. Denner, and S. Dittmaier, *Strong and electroweak corrections to the production of Higgs + 2-jets via weak interactions at the LHC*, Phys. Rev. Lett. **99** (2007) 161803, arXiv:0707.0381 [hep-ph].

[70] M. Ciccolini, A. Denner, and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC*, Phys. Rev. **D77** (2008) 013002, arXiv:0710.4749 [hep-ph].

[71] K. Arnold, M. Bahr, G. Bozzi, F. Campanario, C. Englert, et al., *VBFNLO: A Parton level Monte Carlo for processes with electroweak bosons*, Comput. Phys. Commun. **180** (2009) 1661–1670, arXiv:0811.4559 [hep-ph].

[72] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, *Higgs production via vector-boson fusion at NNLO in QCD*, Phys. Rev. Lett. **105** (2010) 011801, arXiv:1003.4451 [hep-ph].

[73] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A Program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56–74, arXiv:hep-ph/9704448 [hep-ph].

[74] T. Han and S. Willenbrock, *QCD correction to the $pp \to WH$ and ZH total cross-sections*, Phys. Lett. **B273** (1991) 167–172.

[75] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys. Lett. **B579** (2004) 149–156, arXiv:hep-ph/0307206 [hep-ph].

[76] M. Ciccolini, S. Dittmaier, and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, Phys. Rev. **D68** (2003) 073003, arXiv:hep-ph/0306234 [hep-ph].

[77] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO Higgs boson production via gluon fusion matched with shower in POWHEG*, JHEP **0904** (2009) 002, arXiv:0812.0578 [hep-ph].

[78] A. Djouadi, M. Spira, and P. Zerwas, *Production of Higgs bosons in proton colliders: QCD corrections*, Phys. Lett. **B264** (1991) 440–446.

[79] S. Dawson, *Radiative corrections to Higgs boson production*, Nucl. Phys. **B359** (1991) 283–300.

[80] M. Spira, A. Djouadi, D. Graudenz, and P. Zerwas, *Higgs boson production at the LHC*, Nucl. Phys. **B453** (1995) 17–82, arXiv:hep-ph/9504378 [hep-ph].

[81] R. V. Harlander and W. B. Kilgore, *Next-to-next-to-leading order Higgs production at hadron colliders*, Phys. Rev. Lett. **88** (2002) 201801, arXiv:hep-ph/0201206 [hep-ph].

[82] C. Anastasiou and K. Melnikov, *Higgs boson production at hadron colliders in NNLO QCD*, Nucl. Phys. **B646** (2002) 220–256, arXiv:hep-ph/0207004 [hep-ph].

[83] V. Ravindran, J. Smith, and W. L. van Neerven, *NNLO corrections to the total cross-section for Higgs boson production in hadron hadron collisions*, Nucl. Phys. **B665** (2003) 325–366, arXiv:hep-ph/0302135 [hep-ph].

[84] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini, *Two loop light fermion contribution to Higgs production and decays*, Phys. Lett. **B595** (2004) 432–441, arXiv:hep-ph/0404071 [hep-ph].

[85] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, *NLO electroweak corrections to Higgs boson production at hadron colliders*, Phys. Lett. **B670** (2008) 12–17, arXiv:0809.1301 [hep-ph].

[86] S. Catani, D. de Florian, M. Grazzini, and P. Nason, *Soft gluon resummation for Higgs boson production at hadron colliders*, JHEP **0307** (2003) 028, arXiv:hep-ph/0306211 [hep-ph].

[87] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, *Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC*, JHEP **1111** (2011) 064, arXiv:1109.2109 [hep-ph].

[88] S. Frixione, P. Nason, and B. R. Webber, *Matching NLO QCD and parton showers in heavy flavor production*, JHEP **0308** (2003) 007, arXiv:hep-ph/0305252 [hep-ph].

[89] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[90] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **0909** (2009) 111, arXiv:0907.4076 [hep-ph].

[91] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur.Phys.J. **C71** (2011) 1547, arXiv:1009.2450 [hep-ph].

[92] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, Comput.Phys.Commun. **184** (2013) 919–985, arXiv:hep-ph/0405247 [hep-ph].

[93] 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. **B710** (2012) 612–622, arXiv:1111.5869 [hep-ph].

[94] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, *Hadronic top-quark pair production with NNLL threshold resummation*, Nucl.Phys. **B855** (2012) 695–741, arXiv:1109.1536 [hep-ph].

[95] P. Baernreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys.Rev.Lett. **109** (2012) 132001, arXiv:1204.5201 [hep-ph].

[96] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **1212** (2012) 054, arXiv:1207.0236 [hep-ph].

[97] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **1301** (2013) 080, arXiv:1210.6832 [hep-ph].

[98] M. Czakon, P. Fiedler, and A. Mitov, *The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$*, arXiv:1303.6254 [hep-ph].

[99] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, arXiv:1112.5675 [hep-ph].

[100] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s}$= 8 TeVusing eμevents with b-tagged jets*, Tech. Rep. ATLAS-CONF-2013-097, CERN, Geneva, Sept, 2013.

[101] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys.Rev. **D81** (2010) 054028, arXiv:1001.5034 [hep-ph].

[102] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys.Rev. **D83** (2011) 091503, arXiv:1103.2792 [hep-ph].

[103] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, Phys.Rev. **D82** (2010) 054018, arXiv:1005.4451 [hep-ph].

[104] ATLAS Collaboration, *Measurement of the Inclusive and Fiducial Cross-Section of Single Top-Quark t-Channel Events in pp Collisions at* $\sqrt{s}$= 8 *TeV*, Tech. Rep. ATLAS-CONF-2014-007, CERN, Geneva, March, 2014.

[105] ATLAS Collaboration, *Measurement of the cross-section for associated production of a top quark and a W boson at* $\sqrt{s}$= 8 *TeVwith the ATLAS detector*, Tech. Rep. ATLAS-CONF-2013-100, CERN, Geneva, Sept, 2013.

[106] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].

[107] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, JHEP **1107** (2011) 018, arXiv:1105.0020 [hep-ph].

[108] ATLAS Collaboration, *Measurement of the* W^+W^- *production cross section in proton-proton collisions at* $\sqrt{s}$= 8 *TeVwith the ATLAS detector*, Tech. Rep. ATLAS-CONF-2014-033, CERN, Geneva, July, 2014.

[109] T. Binoth, M. Ciccolini, N. Kauer, and M. Kramer, *Gluon-induced W-boson pair production at the LHC*, JHEP **0612** (2006) 046, arXiv:hep-ph/0611170 [hep-ph].

[110] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **0101** (2001) 010, arXiv:hep-ph/0011363 [hep-ph].

[111] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307** (2003) 001, arXiv:hep-ph/0206293 [hep-ph].

[112] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, *W+W-, WZ and ZZ production in the POWHEG BOX*, JHEP **1111** (2011) 078, arXiv:1107.5051 [hep-ph].

[113] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl.Instrum.Meth. **A506** (2003) 250–303.

[114] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys.Lett. **B659** (2008) 119–126, arXiv:0707.1378 [hep-ph].

[115] T. Plehn, D. L. Rainwater, and D. Zeppenfeld, *A Method for identifying $H \to \tau^+\tau^- \to e^\pm\mu^\mp p_T$ at the CERN LHC*, Phys.Rev. **D61** (2000) 093005, arXiv:hep-ph/9911385 [hep-ph].

[116] V. Barger, G. Bhattacharya, T. Han, and B. A. Kniehl, *Intermediate-mass Higgs boson at hadron supercolliders*, Phys. Rev. **D43** (1991) no. 3, 779.

[117] A. J. Barr, B. Gripaios, and C. G. Lester, *Measuring the Higgs boson mass in dileptonic W-boson decays at hadron colliders*, JHEP **0907** (2009) 072, arXiv:0902.4864 [hep-ph].

[118] NNPDF Collaboration Collaboration, R. D. Ball et al., *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, Nucl.Phys. **B855** (2012) 153–221, arXiv:1107.2652 [hep-ph].

[119] Mangano, M., *The so-called MLM prescription for ME/PS matching*, . Presented at the Fermilab ME/MC Tuning Workshop.

[120] J. Alwall, S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad, et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur.Phys.J. **C53** (2008) 473–500, arXiv:0706.2569 [hep-ph].

[121] T. Melia, K. Melnikov, R. Rontsch, and G. Zanderighi, *NLO QCD corrections for W^+W^- pair production in association with two jets at hadron colliders*, Phys.Rev. **D83** (2011) 114043, arXiv:1104.2327 [hep-ph].

[122] B. Jager, C. Oleari, and D. Zeppenfeld, *Next-to-leading order QCD corrections to W+W- production via vector-boson fusion*, JHEP **0607** (2006) 015, arXiv:hep-ph/0603177 [hep-ph].

[123] I. W. Stewart and F. J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, Phys.Rev. **D85** (2012) 034011, arXiv:1107.2117 [hep-ph].

[124] Cassandras, Christos G. and Lafortune, Stephane, *Introduction to Discrete Event Systems*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[125] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput.Phys.Commun. **77** (1993) 219–228.

[126] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput.Phys.Commun. **10** (1975) 343–367.

[127] Neyman, J. and Pearson, E. S., *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society (A) **231** (1933) 289–337.

[128] Wilks, S. S., *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, The Annals of Mathematical Statistics **9** (1938) no. 1, 60–62. `http://www.jstor.org/stable/2957648`.

[129] Wald, Abraham, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Transactions of the American Mathematical Society **54** (1943) no. 3, 426–482.

[130] Reinert, Gesine, *Statistical Theory*. 2009.

[131] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur.Phys.J. **C71** (2011) 1554, `arXiv:1007.1727 [physics.data-an]`.

[132] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, et al., *Boosted decision trees, an alternative to artificial neural networks*, Nucl.Instrum.Meth. **A543** (2005) 577–584, `arXiv:physics/0408124 [physics]`.

[133] ATLAS Collaboration, *Search for the Standard Model Higgs boson in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay mode using Multivariate Techniques with 4.7 $fb^{-}1$ of ATLAS data at $\sqrt{s} = 7$ TeV*, Tech. Rep. ATLAS-CONF-2012-060, CERN, Geneva, June, 2012.

[134] ATLAS Collaboration, *Evidence for Higgs Boson Decays to the $\tau+\tau-$ Final State with the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2013-108, CERN, Geneva, Nov, 2013.

[135] CMS Collaboration, S. Chatrchyan et al., *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, Phys.Rev. **D89** (2014) 012003, `arXiv:1310.3687`.

[136] TMVA Core Developer Team Collaboration, J. Therhaag, *TMVA: Toolkit for multivariate data analysis*, AIP Conf.Proc. **1504** (2009) 1013–1016.

[137] R. E. Schapire, *The Strength of Weak Learnability*, Mach. Learn. **5** (1990) no. 2, 197–227. http://dx.doi.org/10.1023/A:1022648800760.

[138] Y. Freund and R. E. Schapire, *A Decision-theoretic Generalization of On-line Learning and an Application to Boosting*, in *Proceedings of the Second European Conference on Computational Learning Theory*, EuroCOLT '95, pp. 23–37. Springer-Verlag, London, UK, UK, 1995. http://dl.acm.org/citation.cfm?id=646943.712093.

[139] J. H. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*, Annals of Statistics **29** (2000) 1189–1232.

Biography

Ben Cerio was born in upstate New York in 1984. He attended college at Colgate University, graduating *magna cum laude* with a degree in physics and applied mathematics in 2007. Following college, Ben moved to Boston to pursue music with his college rock band, and after a year of writing songs and playing shows, Ben headed south to attend graduate school at Duke University. At Duke, Ben was the recipient of the Charles H. Townes Teaching fellowship. He received the AAPT Teaching Assistant Award for his efforts in teaching introductory physics labs to Duke undergraduates. Ben joined the ATLAS collaboration at the European Organization for Nuclear Research (CERN) in 2010. In ATLAS, he has contributed to the understanding of inner detector error scaling, searches for physics beyond the Standard Model, searches for the Higgs boson, and measurements of the Higgs boson after its discovery. His work has been published in eight papers and conference notes, and he has presented ATLAS collaboration results at two international conferences.