

Measurement of the electroweak production of a photon and a W boson in association with two jets in proton-proton collisions using the ATLAS Detector

John McGowan

Department of Physics McGill University, Montréal, Québec

2023-02-10

A thesis submitted to McGill University in partial fulfillment of the
requirements of the degree of Doctor of Philosophy

©John McGowan 2023



Table of Contents

Abstract	2
Abrégé	3
Author's Contribution	4
Acknowledgements	7
List of Figures	11
List of Tables	12
A Note on Units	13
1 Introduction	14
2 Theoretical Background and Review of Relevant Literature	16
2.1 The Standard Model	16
2.1.1 Elementary Particles and Interactions	17
2.1.2 The Electroweak Production of a W Boson and a Photon in Associ- ation with Jets	19
2.1.3 The Strong Production of a W Boson and a Photon in Association with Jets	20
2.2 Physics at Hadron Colliders	21
2.2.1 Hadronization	21
2.2.2 Parton Distribution Functions	22
2.2.3 Cross Sections in Proton-Proton Collisions	23
2.3 Beyond the Standard Model	24

2.3.1	Standard Model Shortcomings	24
2.3.2	Standard Model Effective Field Theory	25
2.4	Status and Prospects of Standard Model Multiboson Measurements	26
3	Experimental Setup	28
3.1	The Large Hadron Collider	28
3.2	The ATLAS Detector	30
3.2.1	The Inner Detector	32
3.2.2	Calorimeters	36
3.2.3	The Muon Spectrometer	40
3.2.4	Luminosity Detectors	42
3.2.5	The ATLAS Trigger	44
3.3	Event and Detector Simulation	45
3.3.1	Event Simulation	45
3.3.2	Pileup	46
3.3.3	Detector Response	46
4	Event Reconstruction and Selection	48
4.1	Data Quality and Trigger Requirements	49
4.2	Photons	50
4.3	Electrons	55
4.4	Muons	57
4.5	Jets	58
4.6	Missing Transverse Energy	58
4.7	Object Overlap Removal	59
4.8	Event Selection	60
5	Background Estimation	65
5.1	Jets Misreconstructed as Photons	66

5.1.1	The Template Fit Method	67
5.1.2	Uncertainties on Jets Misreconstructed as Photon Estimate	73
5.1.3	Non-prompt Photon Differential Distributions	76
5.2	Electrons Misreconstructed as Photons	78
5.2.1	The Tag and Probe Method	78
5.2.2	Uncertainties on the Estimate of the Background from Electrons Misreconstructed as Photons	82
5.2.3	Electrons Misreconstructed as Photon Differential Distributions . . .	83
5.3	Jets Misreconstructed as Leptons	83
5.3.1	Uncertainties on the Background from Jets Misreconstructed as Lep- tons	85
5.3.2	Jets Misreconstructed as Leptons Differential Distributions	85
5.4	Photons from Other Collisions	86
5.5	Summary	89
6	Neural Networks for Signal Classification	93
6.1	Introduction to Neural Networks	93
6.1.1	Training	95
6.1.2	Testing and Validation	97
6.1.3	Hyperparameter Tuning	97
6.2	Neural Network Design	98
6.2.1	Training Variables	98
6.2.2	Neural Network Training	99
6.2.3	Neural Network Combination Validation	106
7	Uncertainties	110
7.1	Theoretical Uncertainties	110
7.1.1	Scale Uncertainties	112
7.1.2	PDF Uncertainties	112

7.1.3	Strong Coupling Uncertainties	113
7.1.4	Other Theoretical Uncertainties	113
7.2	Experimental Uncertainties	114
8	Statistical Significance and Cross Section Measurement	119
8.1	Statistical Model and Statistical Significance Calculation	120
8.2	Expected Statistical Significance	124
8.3	Observed Statistical Significance	126
8.4	Fiducial Cross Section Measurement	127
8.4.1	Fiducial Region Definition	129
8.4.2	Fiducial Cross Section	133
9	Differential Cross Section Measurement	137
9.1	Differential Extraction Methodology	137
9.2	Treatment of Uncertainties	140
9.2.1	Experimental Uncertainties	140
9.2.2	Theory Uncertainties	141
9.2.3	Statistical uncertainties	142
9.2.4	Likelihood Fit Results	143
9.3	Unfolding	143
9.3.1	Iterative Bayesian Unfolding	143
9.3.2	Unfolded Results and Treatment of Uncertainties	146
10	Discussion and Conclusion	152
A	Simulated Event Samples	168
B	Jets Misreconstructed as Photons: The 2D Sideband Method	170
C	Jets Misreconstructed as Photons: Prompt Photon Template Closure Test	174

D	Jets Misreconstructed as Photons: Cross-check of real photon purity modelling	179
E	Electrons Misreconstructed as Photons: Signal fraction fits	181
F	Jets Misreconstructed as Leptons: Fraction $F_{j \rightarrow \ell}$	184
G	Comparison of Neural Network Training Variables in the Electron and Muon channels	185
H	Treatment of Scale Uncertainties	187
I	Pulls of Nuisance Parameters in Fiducial Cross Section Fit	190

Abstract

This thesis presents the observation and cross section measurements for the electroweak production of a W boson and a photon associated with two high energy jets where the W boson decays to an electron or muon and its corresponding neutrino. The data analyzed consists of 140 fb^{-1} of proton-proton collision data taken by the ATLAS detector between 2015 and 2018. A neural network is trained to separate the signal from the dominant sources of irreducible background. Backgrounds arising from misreconstructed objects are estimated using data driven approaches. A fit of the neural network output is then performed to extract the observed number of signal events. The physics process of interest is observed with a statistical significance of 9.0σ which constitutes the first observation of this process using data collected by the ATLAS detector. The observed number of signal events is corrected for detector effects to obtain a total production fiducial cross section measurement of $15.8 \pm 3.0 \text{ fb}$, consistent with leading order Standard Model predictions. Differential cross section measurements as a function of 6 kinematic variables chosen for their potential sensitivity to new physics are also obtained. The differential cross section measurements are consistent with leading order Standard Model predictions.

Abrégé

Cette thèse présente l'observation et les mesures de section efficace pour la production d'un boson W et d'un photon avec deux jets de haute énergie où le boson W se désintègre en un électron ou un muon, et un neutrino. Les données analysées consistent en 140 fb^{-1} de données de collision proton-proton prises par le détecteur ATLAS entre 2015 et 2018. Un réseau de neurones est formé pour séparer le signal des sources dominantes de bruit de fond prompt. Les bruits de fond provenant d'objets mal reconstruits sont estimés en utilisant des approches basées sur les données. Le processus physique d'intérêt est observé avec une signification statistique de 9.0σ , ce qui constitue la première observation de ce processus utilisant les données recueillies par le détecteur ATLAS. La performance du détecteur est prise en compte dans le calcul du nombre d'événements de signal mesuré afin d'obtenir une mesure de la section efficace à $15.8 \pm 3.0 \text{ fb}$, compatible avec les prédictions du Modèle Standard calculés au premier ordre de la théorie. Des mesures différentielles de section efficace en fonction de 6 variables choisies pour leur sensibilité potentielle à la nouvelle physique sont également obtenues. Les mesures différentielles de section efficace sont compatibles avec les prédictions du Modèle Standard calculés au premier ordre de la théorie.

Author's Contribution

ATLAS is a collaboration of over 5000 members spread across 182 institutions in 42 different countries. Each paper published by the ATLAS collaboration has an author list of over 3000 scientific authors. The construction and operation of the ATLAS detector is a Herculean effort spanning decades. Hundreds of scientists work on the tools used for simulated event generation, detector simulation and analysis software which are used across ATLAS, including in the analysis presented here. The collective work of all these scientists has made this measurement possible.

While many scientists, engineers and technicians build and operate the ATLAS detector, and contribute to the analysis software used, the results and methods presented in this thesis are the work of a small team. Four PhD candidates (John McGowan (McGill University), Benjamin Honan (University of Manchester), Zhen Wang (Tsung-Dao Lee Institute), and Xi Wang (Shanghai Jiao Tong University)) are responsible for nearly all of the analysis chain, including the background estimates and signal fitting. Prof. Heather Russell (University of Victoria) also contributed to simulated event sample generation and to studies of the pileup background. A small handful of professors and postdoctoral researchers advised the team (Prof. Heather Russell, Prof. Andrew Pilkington (University of Manchester), Prof. Brigitte Vachon (McGill University), Dr. Xingguo Li (McGill University) and Dr. Jing Chen (Shanghai Jiao Tong University)).

The results in this thesis constitute the first observation of this physics process using data collected by the ATLAS detector. The thesis presents the first use of machine learning techniques for the study of this process. This thesis also presents differential cross section

measurements of the process of interest that have never before been measured, including the rapidity ordered azimuthal separation of the two leading jets, the rapidity ordered azimuthal separation of the lepton and the photon, and the lepton transverse momentum.

A summary of the contributions of the author and other members of the analysis team are outlined here by chapter.

1. The author summarizes the context and motivation for the measurement presented in this thesis.
2. The author gives an overview of the physical theory that this thesis tests and gives a brief review of relevant literature.
3. The author describes the experimental apparatus used in the analysis, which has been decades in the making and is the work of many scientists, engineers and technicians. The simulation of physics events and the detector response is also described here, which is the work of many scientists inside and outside the ATLAS collaboration. The author directly contributed to the construction and commissioning of the New Small Wheel (NSW) upgrade of the ATLAS muon spectrometer. The New Small Wheel replaced the Cathode Strip Chambers (CSCs) described in Chapter 3 after the recording of the data used in this thesis.
4. The author describes the reconstruction, calibration and selection of physics objects used in this analysis. The reconstruction and calibration is performed by ATLAS collaborators. The reconstructed and fiducial event selection is the work of the author and the other PhD students in the analysis team.
5. The author describes the training and validation of a multivariate classifier used to distinguish signal from prompt backgrounds. This work is entirely done by the author.
6. The author describes the estimation of the backgrounds in the analysis. The author is responsible for the estimation of the background from jets misidentified as pho-

tons, which is the dominant non-prompt background in the analysis. The author is also responsible for the estimate of the pileup background, and the simulated backgrounds. Xi Wang is responsible for the estimate of the background from electrons misidentified as photons, and the background from jets misidentified as muons. Zhen Wang is responsible for the background of jets misidentified as electrons.

7. The author describes the methods used to estimate the theoretical and experimental uncertainties in the analysis. The author is responsible for preparing data samples with theoretical and experimental uncertainty information for the analysis team. The author is responsible for deriving the effect of the uncertainties on the fit of the multivariate classifier.
8. The author describes a fit of the multivariate classifier and the measurement of a fiducial cross section which is the work of the author.
9. The author describes a fit of various kinematic variables and their unfolding for detector effects. This chapter is largely the work of Benjamin Honan and Zhen Wang. The author provided the estimate of the background arising from non-prompt photons from jets for this measurement and was responsible for implementing systematic variations in the analysis framework used. In addition the author defined the fiducial volume and implemented the analysis RIVET routine to validate the fiducial volume and to preserve the analysis.
10. The author discusses the results and highlights potential avenues of further investigation.

Acknowledgements

This thesis is the culmination of a long journey and I would certainly not have made it without a great deal of help along the way. First to my supervisor, Professor Brigitte Vachon: I could not have asked for a better supervisor. I deeply appreciate your steady guidance over the past five years and the freedom you gave me to pursue different projects and to spend multiple years working at CERN. Throughout it all, I considered our weekly meetings a privilege.

I owe a small handful of professors and post docs with whom I've worked very closely a special thanks for all their technical expertise and for showing me the ropes. To Professor Heather Russell, Dr. Tony Kwan, and Dr. Gerardo Vasquez: Thank you very much for your patience in answering my questions and guiding me through the technical aspects of my work on ATLAS. In addition to teaching me how to be an effective researcher, you've taught me what it means to be a great colleague and team member. I'll do my best to take this with me wherever I go.

I've been incredibly fortunate to work with such a wonderful group as the ATLAS McGill group. To Professors Andreas, Francois, and Steve, postdocs and engineers Jona, Raphael, Steffan, Kathrin and Xingguo, and current and former students Benoit, Charlie, John, Luis, Lia, Auriane, Alessandro, Shreya, Christina, Zhelun and Claire: Thank you for the countless discussions over the years. I could not have asked for a more welcoming environment to learn and grow as a researcher.

I've also learned a great deal from the analysis team who share in the work presented in this thesis, especially the other PhD students who worked alongside me in this analysis.

To Ben, Zhen, and Xi: Thank you for sharing this exciting work with me and for the many exchanges we've had over the past few years. Professor Andy Pilkington was especially generous with his time and expertise: Thank you for your immeasurable guidance in bringing this work to its current form.

Finally I owe everything else to my friends and family. A few words cannot begin to express my gratitude, but here they are nonetheless. To Dad and Andrew: Thank you for your unwavering support and patience from the very beginning. To Mom: I hope this would make you proud. To Gwen: Thank you for your support and perspective, you've reminded me in the toughest parts of this journey to look around and enjoy the view.

List of Figures

2.1	Example Feynman diagrams of the signal process	19
2.2	Example Feynman diagram of the largest prompt background	21
2.3	The running of the strong coupling constant	22
2.4	Example parton distribution function	23
3.1	The CERN accelerator complex	29
3.2	The ATLAS Detector	31
3.3	The ATLAS coordinate system	31
3.4	The ATLAS inner detector	33
3.5	The ATLAS Calorimeter	37
3.6	Section of the Electromagnetic Calorimeter	38
3.7	The ATLAS Muon Spectrometer	41
4.1	Event display of a signal like event	49
4.2	Signals of various particles in ATLAS	51
4.3	Identification efficiency for unconverted and converted photons	53
4.4	Illustration of photon isolation	55
5.1	Prompt photon template fit	69
5.2	Non-prompt photon template fit	70
5.3	Photon template normalization fit	71
5.4	Pulls of template parameters	72
5.5	Non-prompt photon template with real photon leakage	74

5.6	Non-tight photon fit with real photon leakage	75
5.7	Non-prompt photon yield in p_T^γ	77
5.8	Electron misreconstructed as a photon illustration	79
5.9	Electron misreconstructed as photon fraction in bins of <i>Probe</i> transverse momentum and pseudorapidity	81
5.10	Electron misreconstructed as photon yield in p_T^γ	83
5.11	Jet misreconstructed as lepton yield in p_T^γ	86
5.12	The $\Delta z(\gamma, PV)$ distribution in data	87
5.13	Comparison between the distribution of $\Delta z(\gamma, PV)$ in data and simulated QCD $W\gamma jj$ events.	88
5.14	Pileup photon yield in p_T^γ	90
5.15	Photon transverse momentum in baseline events	92
6.1	Schematic diagram of a fully connected neural network with one hidden layer.	94
6.2	Neural network training variable distributions	100
6.3	Neural network loss curves	102
6.4	Receiver operator characteristic curve	103
6.5	Neural network output distribution for training and testing data	104
6.6	Average value of training variables as function of neural network output . .	105
6.7	Neural network combination validation in simulation	108
6.8	Neural network combination validation in data	109
7.1	Theory Uncertainties on the Sherpa QCD $W\gamma jj$ prediction	114
8.1	Background only fit in control region only	125
8.2	Likelihood scan performed on fit to data in control region and Asimov dataset in signal region	126
8.3	Unblinded fit to data	128
8.4	Likelihood scan performed on unblinded fit	129

8.5	Schematic of model dependence of a photon isolation cut at truth level . . .	131
8.6	Truth photon isolation cut determination	132
8.7	Validation of truth photon isolation cut	133
8.8	Fiducial cross section measurement	135
9.1	Electroweak event yields in extracted variables	148
9.2	Response matrices for each unfolded variable	149
9.3	Fractional uncertainties on unfolded M_{jj} distribution	150
9.4	Measured differential cross sections	151
B.1	The 2D sideband method	171
B.2	Correlation and leakage coefficients for 2D sideband method	172
B.3	Template Fit and 2D Sideband Method Comparison	173
C.1	Bias in Template Fit Method in Simulation	175
C.2	<i>Non-tight</i> photon fit in Template Fit Closure Test	176
C.3	Normalization fit in Template Fit Closure Test	177
C.4	Results of prompt photon template closure test	178
D.1	Prompt photon purity modelling in p_T^γ	180
E.1	Signal fraction fits in $M_{e\gamma}$	182
E.2	Signal fraction fits in M_{ee}	183
F.1	$F_{j \rightarrow \ell}$ in electron and muon channel	184
G.1	Comparison of neural network training variables in electron and muon channels	186
H.1	Shape of scale variations in Sherpa electroweak $W_{\gamma jj}$ generator	188
H.2	Shape of scale variations in Sherpa QCD $W_{\gamma jj}$ generator	189
I.1	Nuisance parameter pulls in fit of neural network distribution	191

List of Tables

2.1	Elementary fermions and their properties	18
2.2	Elementary bosons and their properties	18
4.1	Summary of triggers used in analysis	50
4.2	Object and event selection summary	63
4.3	Cutflow table	64
5.1	Selection criteria of <i>Tag</i> electron and <i>Probe</i> objects used for calculating the probability that an electron is misreconstructed as a photon.	80
5.2	Yields of prompt and non-prompt events in baseline events	91
6.1	Gridsearch summary	101
6.2	Neural network variable ranking	106
7.1	Impact of experimental uncertainties on QCD $W_{\gamma jj}$ and electroweak $W_{\gamma jj}$ event yields	118
8.1	Normalization factors after fit to data	127
8.2	Fractional uncertainty contributions to the measured Fiducial cross section .	136
9.1	Summary of signal and control regions used in differential cross section fit .	138
A.1	Summary of simulated samples used in analysis	169

A Note on Units

Throughout this thesis "natural" units are used where $c = \hbar = 1$. As such energy, momentum and mass are all expressed in units of electron-volts (eV) instead of eV, eV/ c , and eV/ c^2 .

Chapter 1

Introduction

Science progresses as an interplay between theory and experiment. A mathematical model makes some predictions about how nature behaves and an experiment tests these predictions and establishes the range of their validity. In turn, new theories are developed to revise the old, or at least expand the old theory range of validity.

The current understanding of nature at the subatomic scale is codified in the Standard Model of particle physics. This mathematical model describes the fundamental constituents of matter and how they interact. To date, it has proven remarkably successful under rigorous test; yet, the Standard Model is known to be incomplete. For example, it does not explain the preponderance of matter in our universe when an equal amount of matter and antimatter is understood to have existed at the beginning of the universe, nor does it explain the nature of dark matter, or the observed non-zero neutrino masses.

These shortcomings of the Standard Model demand that experimentalists continue to test the Standard Model in order to understand its range of validity and limits. To do this, experimental measurements must be performed and compared with the predictions of the Standard Model.

The work presented in this thesis constitutes a test of the validity of the Standard Model. The rate of occurrence of an extremely rare subatomic process is studied. Rare subatomic processes take place in proton-proton collisions produced by the highest en-

ergy collider in the world, the Large Hadron Collider (LHC) at the CERN laboratory. The work presented in this thesis is an analysis of data from the ATLAS detector, which records the energy and direction of particles produced in these proton-proton collisions.

The subatomic process of interest is the production through the weak force of two elementary particles, called a W boson and a photon, along with two sets of fragments originating from each of the initial protons that collided. To identify these rare events from the large background a machine learning approach is developed. The success of the measurement relies on the ability to precisely predict the yield of different sources of background, even when the background is difficult to mathematically model. A data driven technique was developed to estimate the large background from misidentified photons.

A statistical significance compatible with observation of the process of interest is obtained. Fiducial and differential production cross sections are also measured in this work. The measurements are statistically compatible with the Standard Model predictions. The differential cross section measurements can be used in the future to search for new physics phenomena.

Chapter 2

Theoretical Background and Review of Relevant Literature

In this Chapter, a brief overview of the theory relevant to the analysis presented in this thesis is given. Shortcomings of the Standard Model are discussed as well as a potential use of the measurement presented in this thesis to search for beyond Standard Model physics. A review of relevant literature also presented.

2.1 The Standard Model

The Standard Model [1] is a quantum field theory that describes all of the experimentally observed particles and their interactions under three forces: the strong force, which binds protons and neutrons in the nucleus; the electromagnetic force which binds electrons to atoms; and the weak force that is responsible for weak decays. For a complete description of the Standard Model, the reader is referred to [1]. A brief, qualitative overview is given here.

The Standard Model is described by a Lagrangian density function and has 19 free parameters that must be determined by experiments. In addition, the Standard Model obeys local symmetries $U(1)_Y$, $SU(2)_L$, and $SU(3)_C$ which come with conserved charges

hypercharge Y , weak isospin T , and color C , respectively. The better known electric charge Q is related to hypercharge and weak isospin by the Gellmann-Nishijima formula $Q = T_3 + \frac{Y}{2}$ where T_3 denotes the third component of isospin. Elementary particles of the Standard Model can be classified according to the values of their charges.

2.1.1 Elementary Particles and Interactions

The particles of the Standard Model consist of spin-1/2 “fermions”, integer spin “gauge bosons” and the Higgs Boson, which has spin 0. The fermions can be further classified into two categories: quarks and leptons. Quarks and leptons come in three generations, which differ by their masses. Note also that for each fermion, there exists a charge conjugate antiparticle. In this thesis, when a particle is referred to, its charge conjugate state is implied.

The three generations of charged leptons are the electron (e^-), the muon (μ^-) and the tau lepton (τ^-) which interact via the electromagnetic force and the weak force. With each charged lepton comes a corresponding neutrino, ν_e , ν_μ and ν_τ . The neutrino carries no electric charge and only interacts via the weak interaction.

Quarks carry color charge, weak isospin and a fractional electric charge (+2/3 or -1/3). The six quarks are the up (u), down (d), charm (c), strange (s), top (t) and bottom (b). Like the leptons, each flavor of quark comes with an antiparticle of opposite charge ($\bar{u}$, $\bar{d}$, $\bar{c}$, $\bar{s}$, $\bar{t}$, $\bar{b}$). As carriers of weak isospin and electric charge, quarks interact with leptons via the electromagnetic and weak force. As carriers of color charge, quarks also interact with one another by way of the strong force, and bound states of quarks are possible. However, only colorless states are observed in nature. This “color confinement” results in quarks forming colorless hadrons, which are either mesons (a quark-antiquark pair) or baryons (three quarks with a total color charge of 0). For example, protons are a bound state of two up quarks and a down quark.

The spin 1 “gauge bosons” are the force carriers that mediate the electromagnetic, weak, and strong interactions. The weak interaction is mediated by the massive $W^\pm$

and Z bosons. The electromagnetic interaction between charged particles is mediated by the massless photon (γ). At energies far above the mass scale of the weak interaction ($\mathcal{O}(100 \text{ GeV})$) the W and Z masses are negligible and the electromagnetic and weak forces are indistinguishable. These energies are far beyond the energy probed in the analysis presented in this thesis. At lower energies the electromagnetic and weak forces behave differently due to a mechanism called electroweak symmetry breaking [1]. This mechanism gives the $W^\pm$ and Z bosons non-zero mass. The strong interaction is mediated by eight massless gluons (g). As the $W^\pm$ bosons are both electrically and weakly charged, they can interact with themselves, with the photon, and with the Z boson as well as with the leptons and quarks. Similarly, as the gluons carry color charge, they can interact with themselves as well as with the quarks.

Finally there is the spin 0 Higgs boson. The Higgs boson is an excitation of the Higgs field, which gives mass to the fundamental particles.

The elementary fermions and bosons described by the Standard Model are summarized along with their basic properties in Tables 2.1 and 2.2, respectively.

Leptons (spin = 1/2)								
Particle	Mass	Q [e]	Particle	Mass	Q [e]	Particle	Mass	Q[e]
e^-	0.511 MeV	- 1	μ^-	105.7 MeV	- 1	τ^-	1.77 GeV	- 1
ν_e	< 2 eV	0	ν_μ	< 0.19 MeV	0	ν_τ	< 18.2 MeV	0
Quarks (spin = 1/2)								
Particle	Mass	Q [e]	Particle	Mass	Q [e]	Particle	Mass	Q[e]
u	2.2 MeV	+ 2/3	c	1.28 GeV	+ 2/3	t	173.1 GeV	+ 2/3
d	4.7 MeV	- 1/3	s	96 MeV	- 1/3	b	4.18 GeV	- 1/3

Table 2.1: Elementary fermions and their properties. Values are taken from [2].

Particle	Mass	Q [e]	Spin
$W^\pm$	80.38 GeV	± 1	1
Z	91.19 GeV	0	1
γ	0	0	1
g	0	0	1
H	125.09 GeV	0	0

Table 2.2: Elementary bosons and their properties. Values are taken from [2].

Given the Standard Model Lagrangian, a prescription known as Feynman rules [3] can be used to calculate a quantity called the matrix element, $\mathcal{M}$, beginning with so-called Feynman diagrams drawn using the allowed Standard Model interaction couplings. The matrix element encodes the probability of a transition from an initial state to a final state and can be used to calculate observable quantities such as interaction cross sections and particle decay widths. The matrix element calculation involves a perturbative expansion in a coupling constant which is truncated at a given order. Feynman diagrams representing possible interactions can be drawn for each order, with higher order correction diagrams typically including internal loops which come with a suppression factor.

2.1.2 The Electroweak Production of a W Boson and a Photon in Association with Jets

The analysis described in this thesis seeks to measure the electroweak production cross section of a W boson and a photon in association with two jets. At leading order, the Feynman diagrams used to calculate the matrix element for this process include electroweak couplings. Example diagrams of the electroweak production of a W boson and a photon in association with jets are shown in Figure 2.1.

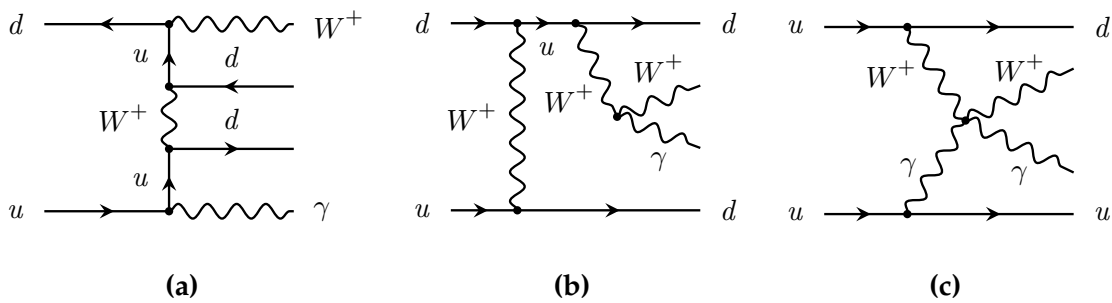


Figure 2.1: Example Feynman diagrams for the electroweak production of a W boson and a photon in association with two jets at leading order. Time flows from left to right.

Following the Feynman rules, each three particle vertex in the diagrams in Figure 2.1 comes with a factor of the relevant coupling while the quartic vertex comes with a factor

of the relevant coupling squared in the rate of production. The signal process in this thesis concerns diagrams with order $\alpha_W + \alpha_{EM} = 4$ at leading order where α_W represents a coupling characterizing an interaction that proceeds via the weak force and α_{EM} represents a coupling characterizing an interaction that proceeds via electromagnetic force. Note at leading order there are no contributions from the strong coupling, α_S . At next-to-leading order, one can add diagrams with loops involving electroweak or strong vertices, and additional factors of α_S , α_{EM} or α_W can enter the calculation.

Note also that the process of interest is sensitive to both the triple gauge boson coupling between two W bosons and a photon (Figure 2.1(b)) and the quartic gauge boson coupling $WW\gamma\gamma$ (Figure 2.1(c)). The diagram including the $WW\gamma\gamma$ coupling is referred to as the “vector boson scattering” diagram. The measurement in this thesis is designed to be most sensitive to the vector boson scattering diagram by performing the analysis in kinematic regions where this diagram is the dominant contribution to the process of interest. The “vector boson scattering” diagram is highly sensitive to new physics.

2.1.3 The Strong Production of a W Boson and a Photon in Association with Jets

The largest prompt background in this analysis arises from the production of a W boson and a photon in association with two jets in which the strong interaction is involved. Since the theory of the strong interaction is called Quantum Chromodynamics (QCD), this background is referred to as the QCD production of a W boson and a photon in association with two jets throughout the thesis. An example Feynman diagram of this background, at leading order, is shown in Figure 2.2.

The cross section of this diagram is proportional to a factor of α_S^2 and a factor of α_W^2 . At the energy scales probed at the LHC, α_S is of order 10 times larger than α_{EM} , and even larger than α_W . The rate of production of this background is therefore of order 100 times larger than that of the reaction of interest.

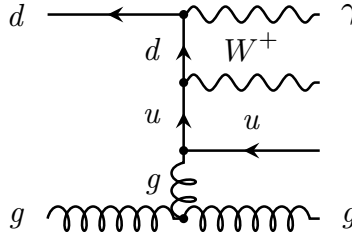


Figure 2.2: Example Feynman diagram for the QCD production of a W boson and a photon in association with two jets at leading order.

2.2 Physics at Hadron Colliders

Matrix elements are used to calculate interaction cross sections from initial state particles, in the case quarks, to final state objects. At hadron colliders, initial state quarks come from the colliding protons and the final state quarks undergo a process called hadronization. The relevant aspects of physics at hadron colliders are briefly discussed here.

2.2.1 Hadronization

The value of the strong coupling constant decreases with the energy scale of the process probed, Q , as shown by multiple experiments in Figure 2.3. At large Q , or small distance, quarks and gluons behave as nearly free particles. This is referred to as “asymptotic freedom” as this behavior occurs as the value of the strong coupling constant asymptotically vanishes. At low Q , or large distance, where the value of the strong coupling constant is large, quarks bind together. If two quarks are pulled apart, the energy density between them becomes so large that quark-antiquark pairs are created from vacuum to form lower energy bound states. As a result of this hadronization process quarks are only observed in bound states. This phenomena is known as “confinement” Due to hadronization and confinement, in collider experiments, the presence of a parton (quark or gluon) in the final state of a reaction is inferred from the production of collimated jets of hadrons whose properties are correlated to the energy and direction of the original parton.

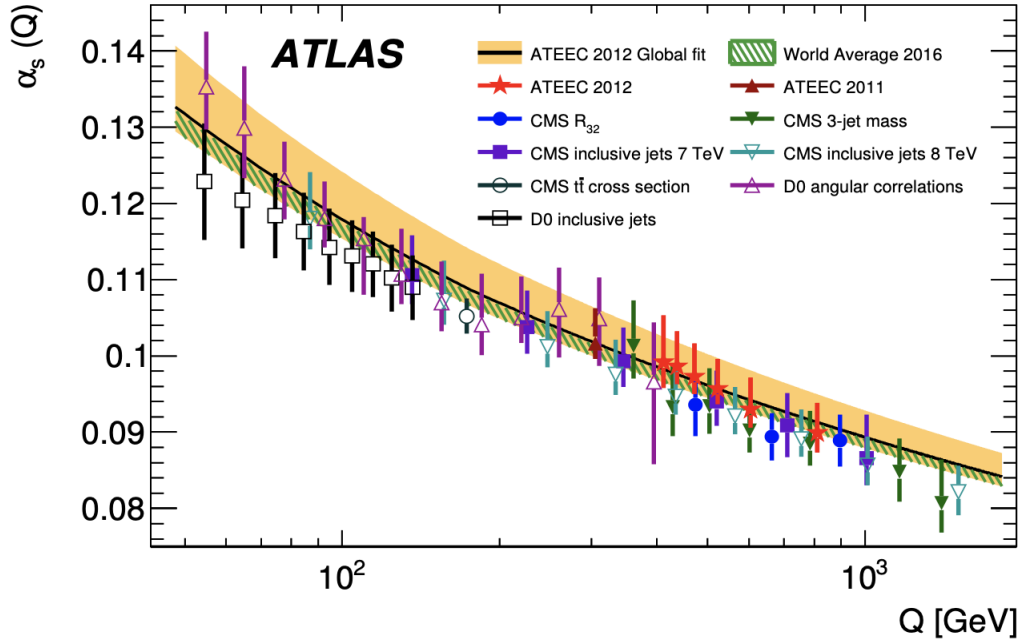


Figure 2.3: The strong coupling constant as a function of the energy probed as measured by various experiments [4]

2.2.2 Parton Distribution Functions

Protons consist of three valence quarks (uud) but also the gluons that bind the quarks together. These gluons can create quark-antiquark pairs called sea quarks. Thus in proton-proton collisions, valence quarks, gluons, or sea quarks can interact. The probability that a quark or gluon is contained in a proton at a given energy is given by a *Parton Distribution Function* (PDF). An example Parton Distribution Function is shown in Figure 2.4. One can see that at an energy scale of $Q^2 = 10 \text{ GeV}^2$ and large values of the fraction of total proton energy carried by a parton, x , the probability for a proton to contain an up or down quark is the greatest, while at small x values the proton content is largely dominated by gluons.

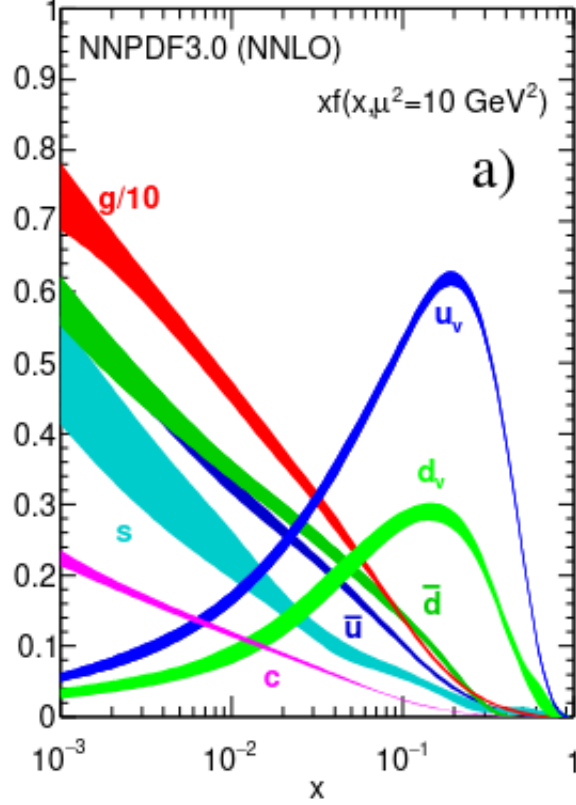


Figure 2.4: The NNPDF3.0 [5] proton parton distribution function at scale $Q^2 = 10 \text{ GeV}^2$ [2]. The symbol x represents the fraction of the total proton energy carried by a parton. The y-axis is the value of the parton distribution function multiplied by x .

2.2.3 Cross Sections in Proton-Proton Collisions

Given a matrix element describing the parton interactions of interest calculated using a perturbative expansion in the strong coupling constant and the proton parton distribution functions, the factorization theorem [3] may be used to calculate a cross section of interest in proton-proton collisions.

The factorization theorem states that the total cross section for two hadrons with longitudinal momentum P_1 and P_2 , can be computed using a weighted combination of the relevant partonic cross section using:

$$\sigma_{pp \rightarrow X} = \sum_{i,j} \int_0^1 dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}(p_1, p_2, \alpha_s(\mu_R^2), Q^2/\mu_R^{2v}, Q^2/\mu_F^2) \quad (2.1)$$

where i, j represent the different possible combinations of initial state partons, $f_{i,j}(x, Q^2)$ are the parton distribution functions, $p_1 = x_1 P_1$ and $p_2 = x_2 P_2$ are the momentum of the two interacting partons and $\hat{\sigma}$ is the parton cross section of the process of interest as calculated from the matrix element. Infrared and ultraviolet divergences arise in the calculation of the matrix element. These are dealt with by introducing a renormalization scale (μ_R) and a factorization scale (μ_F) as described in [3]. The parton distribution functions and parton cross section is dependent on these scales unless the perturbative expansion is carried out to all orders, which is not practical. For fixed order calculations, the calculations are typically done by fixing $\mu = \mu_F = \mu_R = Q$ where Q is the energy scale of the process of interest.

2.3 Beyond the Standard Model

The Standard Model describes interactions at hadron colliders remarkably well; however it does not explain other observed phenomena. Below the shortcomings of the Standard Model are briefly discussed. In addition a tool to use Standard Model measurements to search for signs of beyond the Standard Model physics is outlined.

2.3.1 Standard Model Shortcomings

In the first section, the reader likely noticed that gravity is not mentioned among the forces described by the Standard Model. As a quantum mechanical picture of gravity does not exist, gravity is not included in the Standard Model. Several measurements from astrophysics and cosmology suggest that the universe is made of mostly dark matter and dark energy which does not exist in the Standard Model. These measurements suggest that the Standard Model only describes 5% of all the content in the universe. The Standard Model also fails to describe the matter-antimatter asymmetry observed in the universe. By Standard Model predictions, matter and antimatter should have been created equally in the early universe and the annihilation of matter and antimatter should

have left nothing but radiation, rather than the matter dominated universe observed today.

The nature of neutrinos is not fully described by the Standard Model. In the Standard Model, neutrinos are assumed to be massless, in direct conflict with recent observations [6].

Another hint that the Standard Model may be incomplete is the so-called “Fine-Tuning” problem [7]. The Higgs boson mass observed at the scale of 100 GeV is consistent with the Standard Model only if the parameters of the Standard Model are such that higher order contributions to the Higgs mass perfectly cancel. Many argue that this fine tuning is unnatural and should not be required in a complete theory.

2.3.2 Standard Model Effective Field Theory

The many shortcomings of the Standard model and yet the lack of evidence for new physics at hadron colliders motivate using Standard Model measurements, like that presented in this thesis, to indirectly search for new physics. This may be done using an *effective field theory*.

In the Standard Model Lagrangian, all operators consistent with Standard Model symmetries are included up to dimension four. However one could continue to write down higher order operators consistent with Standard Model symmetries, arriving at the Standard Model Effective Field Theory (SMEFT) [8] Lagrangian

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_{n=5}^{\infty} \frac{f_n}{\Lambda^{n-4}} \mathcal{O}_n \quad (2.2)$$

where f_n are coupling constants, $\mathcal{O}_n$ are n -dimension operators with $n > 4$, and Λ is an energy scale at which new physics appears, which is assumed to be greater than the scale experimentally probed. In such a parameterization, new particles do not appear explicitly, but their effect is indirectly incorporated in these higher order operators and in anomalous values of Standard Model couplings.

Both dimension 6 and dimension 8 SMEFT operators modify the couplings probed in this thesis. Some dimension 6 operators violate charge parity conservation and may modify distributions of azimuthal separation of the final state particles measured. Dimension 8 operators may modify the gauge boson quartic couplings and may result in larger than expected cross sections in the tails of the dijet mass or diboson mass distributions of multiboson production processes.

2.4 Status and Prospects of Standard Model Multiboson Measurements

Measurements of vector boson scattering and the electroweak production of gauge bosons in association with jets have become possible thanks to the large data samples recorded at the LHC. Vector boson scattering measurements join other multiboson measurements as important tests of the Standard Model. Before the LHC, measurements of diboson production were performed at the Large Electron Positron collider (LEP) [9] and at the Tevatron [10]. The LHC has made it possible to achieve more precise diboson measurements such as those that have been performed by the ATLAS and CMS collaborations [11–24]. The large LHC data samples have made the study of triboson production possible for the first time. Triboson production in many channels has been studied, including the production of $W\gamma\gamma$ and $WW\gamma$, which probe the same quartic gauge coupling as the production of a W boson and a photon in association with jets [15, 25–28].

Many measurements of the electroweak production of gauge bosons in association with jets have been performed since the beginning of LHC running [29–43]. The electroweak production of a W boson and a photon in association with jets was first observed and measured by CMS [43–45].

To date, no significant deviation from Standard Model predictions have been found but collaboration between the experimental and theory communities have allowed for a better understanding of the modelling of vector boson scattering processes [46].

In absence of deviation from Standard Model predictions, multiboson measurements have been used to set limits on the size of potential anomalous couplings using an effective field theory framework. Some channels have been used to set combined limits and global combinations are foreseen in the near future.

The High Luminosity upgrade of the LHC (HL-LHC) will collide protons at $\sqrt{s} = 14$ TeV and with a projected peak luminosity of $7.5 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Increased statistics will improve the precision of vector boson scattering measurements, which are often dominated by statistical uncertainties. Upgraded detectors will allow for reduced experimental uncertainties. The prospects for vector boson scattering measurements at the HL-LHC are outlined in [47] along with prospects for vector boson scattering measurements at future colliders.

Chapter 3

Experimental Setup

The analysis presented in this thesis is performed using 140 fb^{-1} of proton-proton collision data at a center of mass energy $\sqrt{s} = 13 \text{ TeV}$ collected by the ATLAS detector. The proton-proton collisions are provided by the Large Hadron Collider (LHC). In this chapter, the LHC and the ATLAS detector are briefly described. Simulated proton-proton collisions are used to interpret the data. The simulation of proton-proton collisions and the simulation of the ATLAS detector are also discussed in this chapter.

3.1 The Large Hadron Collider

The CERN accelerator complex is shown in Figure 3.1. The Large Hadron Collider is located inside an underground tunnel of 27 km circumference. The LHC uses radio frequency cavities for proton acceleration and quadrupole and dipole magnets for focusing and bending proton beams. Protons originate from the nuclei of hydrogen atoms in a bottle. The hydrogen atoms are ionized in an electric field and the protons are sent to LINAC2, a linear accelerator. After being accelerated to 50 MeV in LINAC2, the protons are injected into the proton synchrotron booster which accelerates the protons to 1.4 GeV. The protons are then injected into the proton synchrotron where they are accelerated to

and γ_r is the relativistic Lorentz factor to account for protons travelling at a speed near the speed of light. The beam emittance, ϵ_n , is a measure of the spread of the beams in the plane transverse to the beam direction, β^* is the beta function which parameterizes the length of the beam in the longitudinal direction, and F is a factor accounting for the angle at which two beams cross at the time of a collision. Instantaneous luminosity is one of the parameters used to measure the performance of a particle collider.

Using the instantaneous luminosity, the number of events expected from a given physics process in a given period of time can be calculated as

$$N = \mathcal{L} \times \sigma \quad (3.2)$$

where σ is the cross section of the process expressed in units of barn (b), and $\mathcal{L}$ is the integrated luminosity, defined as

$$\mathcal{L} = \int L dt \quad (3.3)$$

where the time integral runs over the amount of time during which particle collisions occur. The integrated luminosity is expressed in units of b^{-1} and provides a measure of the total number of collisions produced over a period of time. The integrated luminosity of the dataset used in this thesis is measured by detectors described in Section 3.2.4.

3.2 The ATLAS Detector

The ATLAS detector is a cylindrical multi-purpose detector located along the LHC. A schematic drawing of the ATLAS detector is shown in Figure 3.2. The ATLAS detector was designed to enable a broad physics program, that includes the observation of the Higgs boson, precision electroweak measurements and searches for new physics. To achieve these goals, high detector granularity and hermiticity is paramount. The multiple subsystems of the ATLAS detector are shown in Figure 3.2 and described in this section.

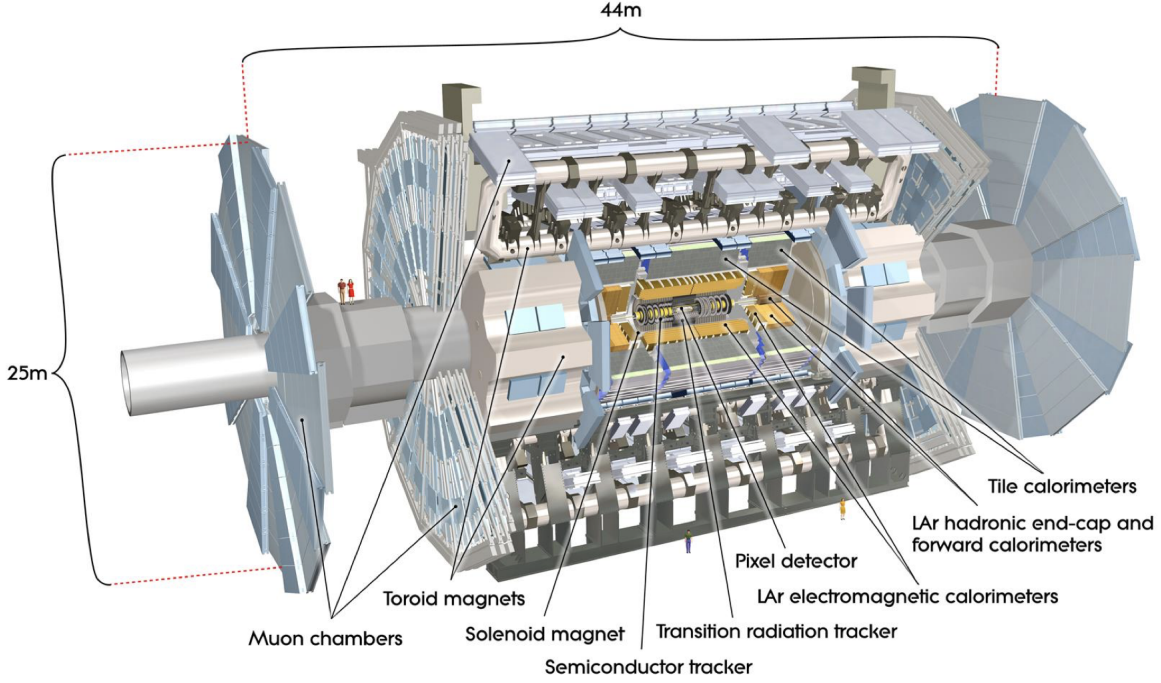


Figure 3.2: Schematic drawing of the ATLAS detector with each detector subsystem indicated [50].

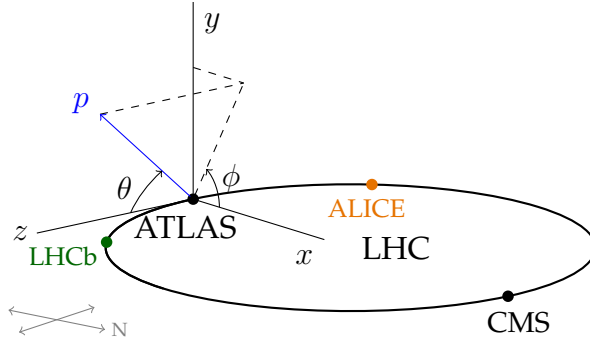


Figure 3.3: The ATLAS coordinate system [51].

The ATLAS coordinate system is shown in Figure 3.3. The interaction point where proton beams collide is taken to be the origin of the coordinate system used to describe the ATLAS detector. The detector is oriented such that the LHC beamline lies long the z -axis while the x -axis points toward the center of the LHC ring and the y -axis points upwards. The azimuthal angle ϕ , is defined from $-\pi$ to π starting from the x -axis. The polar angle θ , is defined from 0 to π starting from the positive z -axis. The detector is

approximately symmetric in ϕ and θ . Throughout this thesis, the variables rapidity and pseudorapidity are used. Rapidity is defined as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad (3.4)$$

where E is the total energy of a particle and p_z is the component of the particle momentum in the z direction. The difference between rapidities of two particles is invariant with respect to Lorentz boosts along the z -axis. This makes this quantity particularly useful in a hadron collider where proton collisions rarely have their center of mass frame coincide with the detector rest frame. Since rapidity can be hard to measure for highly relativistic particles, in practice, pseudorapidity is used. Pseudorapidity is defined as

$$\eta = \ln \tan \frac{\theta}{2}. \quad (3.5)$$

For highly relativistic particles such as those produced in proton-proton collisions at the LHC, pseudorapidity is a good approximation of rapidity.

In increasing radius, the ATLAS detector consists of an inner detector, electromagnetic and hadronic calorimeters, and a muon spectrometer. The detector is segmented into a cylindrical barrel section with detectors designed to detect particles with a large momentum component transverse to the z -axis, and two wheel-shape endcaps designed to detect particles with a large longitudinal momentum component.

3.2.1 The Inner Detector

The inner detector (ID) [52] comprises the detector subsystems closest to the interaction point. The inner detector serves to precisely measure the trajectory of electrically charged particles and, in doing so, provides a measurement of the particles momentum. The reconstructed particle trajectories are used to identify whether a particle originated from

a high momentum transfer collision, the hard scatter. Reconstructed trajectories, or *tracks*, are also used for particle identification. Momentum and spatial resolution is paramount.

The inner detector consists of, in increasing radius, the insertable B-layer (IBL) [53], the pixel detector [54], the silicon strip tracker (SCT) [52], and the transition radiation tracker (TRT) [52]. A schematic of each inner detector subsystem and its radial position is shown in Figure 3.4. The ID operates in a 2 T solenoidal magnetic field, enabling the identification and momentum measurement of charged particles from the bending of their tracks in the $x-y$ plane. The ID is designed to accurately reconstruct the momentum of charged particles down to 100 MeV. The momentum resolution is momentum dependent and ranges from 1.6% for particles with transverse momentum below 5 GeV to 50% in the TeV range [55].

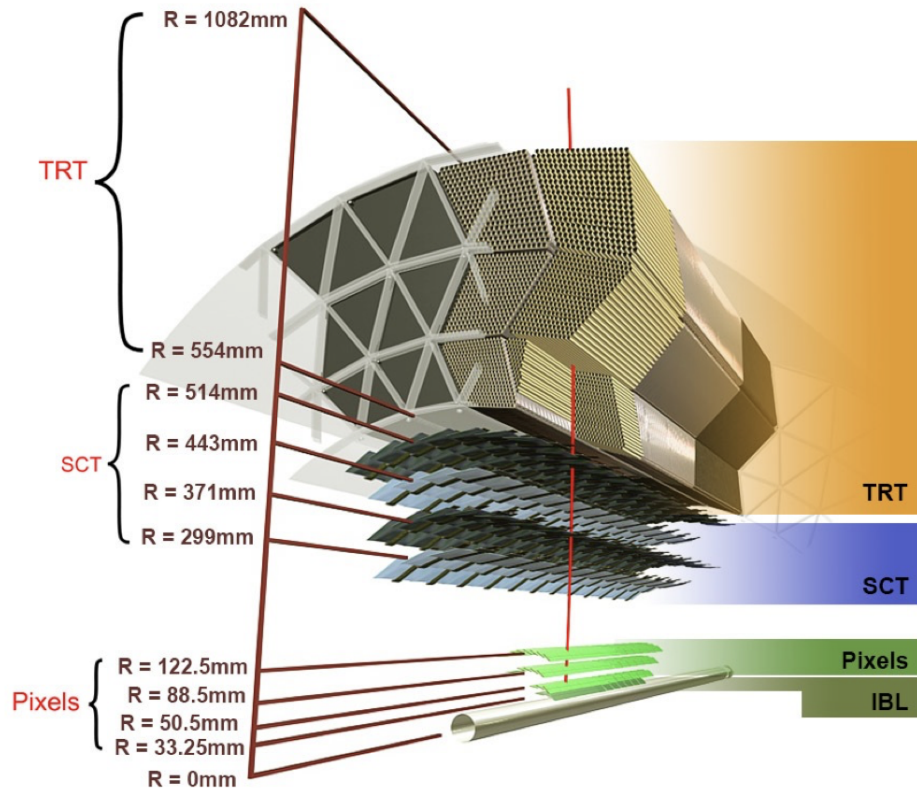


Figure 3.4: Schematic diagram representing a cutaway view of the inner detector and its subsystems: the IBL, pixel detector, silicon tracker and transition radiation tracker. Image taken from [55].

The Pixel Detector and Insertable B-Layer

The pixel detector provides up to three layers of charged particle hits. The pixel detector is segmented into a barrel covering $0 < |z| < 400.5$ mm and $50.5 < R < 122.5$ mm and two endcaps covering $495 < |z| < 650$ mm and $88.8 < R < 149.6$ mm. The barrel layers are concentric around the beam pipe while each endcap layer is a disk in the $x - y$ plane. In both barrel and endcap, each pixel is a silicon p-n junction of 50×400 μm . The spatial resolution of charged particle hits is 10 μm in the $R - \phi$ plane and 115 μm along the z -axis [52].

The insertable B-layer is an additional layer of silicon p-n junctions that was inserted between the pixel detector and the beam pipe in 2014 before the beginning of Run 2. The additional layer covers $|z| < 332$ mm and $31.0 < R < 40.0$ mm and serves to increase the ability of the inner detector to identify whether a particle originated in the hard scatter and to protect the first layer of the pixel detector from radiation damage at Run 2 luminosity.

The Silicon Strip Tracker

A charged particle leaving the pixel detector will next traverse 4 or 9 layers of silicon strip tracker (SCT) depending on its trajectory through the barrel or endcap region of this subsystem. The charged particle crosses the first layer of the SCT at $R = 299$ mm and leaves at $R = 514$ mm in the barrel, which covers $0 < |z| < 749$ mm. In the endcap, the charged particle crosses the first layer at $|z| = 839$ mm and the last at $|z| = 2735$ mm. The endcaps cover $275 < R < 560$ mm. Each SCT layer consists of one layer of silicon strips aligned with the beamline and another offset by 40 μrad . The strip pitch is 80 μm .

In each layer, a position is recorded by fitting the pulse on the strips that record hits above threshold. With the additional constraint that the line between two layers points to the origin of the interaction, the spatial resolution of the SCT is 17 μm in the $R - \phi$ plane in both the barrel and the endcap. The resolution in z in the barrel is 580 μm while the resolution in R in the endcap is 580 μm .

The Transition Radiation Tracker

The final and largest subsystem in the inner detector is the transition radiation tracker (TRT) which covers $0 < |z| < 712$ mm and $563 < R < 1066$ mm in the barrel and $848 < |z| < 2710$ mm and $644 < R < 1004$ mm in the endcaps. The TRT is composed of 4 mm diameter drift tubes, or straws, filled with a xenon gas mixture and each constructed with an anode wire in its center held at a voltage of 1.5 kV. A charged particle passing through a straw ionizes the xenon gas mixture, and the resulting current produced by the drift of ionization electrons and ions is read out on the anode wire as a hit.

The barrel segment of the TRT consists of 73 layers of 1.44 m long straws that run parallel to the beam pipe. In the endcap, there are 160 layers of 37 cm long straws pointing radially outwards from the beam pipe. A charged particle travelling through the TRT in the barrel traverses on average 36 straws. The spatial resolution of the TRT is $130 \mu\text{m}$ in $R \times \phi$ in the barrel and $130 \mu\text{m}$ along the z -axis in the endcap.

The TRT provides additional information in the measurement of transition radiation [56]. When a charged particle passes through the polypropylene fiber between the drift tubes, it emits soft x-ray photons. The intensity of this so called transition radiation is proportional to the particle's Lorentz factor,

$$\gamma = \frac{E}{m} \tag{3.6}$$

where E and m are the energy and mass of the charged particle traversing the TRT. The transition radiation photons are absorbed by the xenon gas mixture inside the drift tubes resulting in gas ionization of a greater magnitude than that generated by the passage of a charged particle. This further aids in particle identification, principally to distinguish electrons, which come with the most transition radiation due to their small mass, from pions.

3.2.2 Calorimeters

Moving radially outward from the interaction region, after the inner detector and the solenoid magnet is the ATLAS calorimeter, schematically represented in Figure 3.5. The ATLAS calorimeter is composed of the electromagnetic calorimeter, whose role is to measure the energy and position of electrons and photons, and the hadronic calorimeter, whose role is to measure the energy and position of hadrons such as neutrons and pions.

Sampling calorimetry is employed for both the electromagnetic calorimeter and the hadronic calorimeter. A sampling calorimeter consists of a dense absorber material, which serves to break up a particle into a particle *shower*, interspersed with an active material which serves to measure the energy of the shower. An electromagnetic shower is a cascade of photons and electrons. In the dense absorber material, photons are radiated from electrons as *bremsstrahlung* and photons convert into electron-positron pairs. The multiplication of electron-positron pairs over multiple layers gives the shower its namesake. The ionization by the electrons of the active material in the electromagnetic calorimeter provides the energy measurement. Hadronic showers consist of hadron production and subsequent decays, including pion and muon decays. Neutral pions decay into two photons which create an electromagnetic component to the hadronic shower.

Both the electromagnetic and hadronic calorimeters have modules in the barrel and endcaps, to measure the energy and position of particles with large transverse or longitudinal momentum, respectively. For particles with a very large longitudinal component, there is a special forward calorimeter located near the beam pipe which also uses sampling calorimetry to detect electromagnetic and hadronic showers.

The Electromagnetic Calorimeter

The electromagnetic calorimeter [50] is composed of liquid argon (LAr) active layers and lead absorber layers. A schematic diagram of a section of the barrel electromagnetic calorimeter is shown in Figure 3.6. Charged particles from an electromagnetic shower

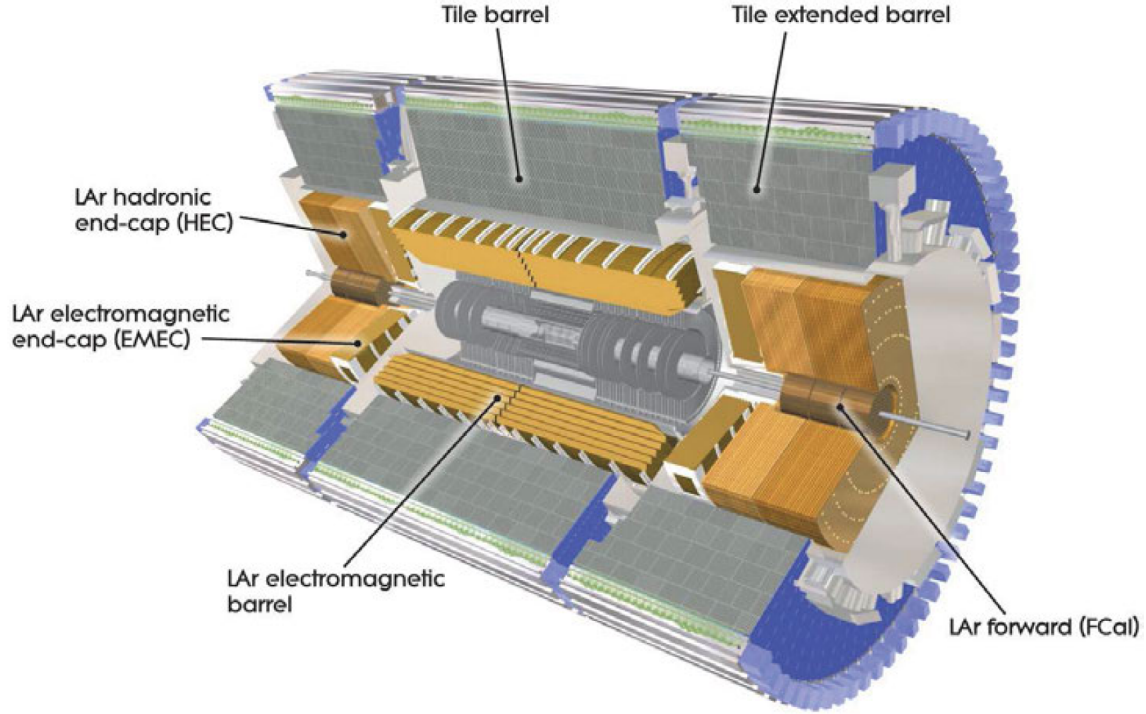


Figure 3.5: Schematic diagram of the ATLAS calorimeter [50].

traversing the liquid argon gap produce ionization electrons. The ionization electrons and positive ions drift under the influence of an electric field across the liquid argon gap. This motion of charge induces a current on electrodes in the gap whose amplitude is proportional to the amount of energy deposited. In the barrel, lead absorber layers and copper electrodes are arranged in an accordion shape to provide full coverage in ϕ . The degree of folding in the accordion shape is varied in η in order to keep the thickness of the calorimeter, in terms of radiation length, constant across the barrel.

The electromagnetic calorimeter covers $|\eta| < 1.475$ in the barrel and $1.375 < |\eta| < 3.2$ in the endcap. A gap exists between the barrel and the endcap which is used for the cabling of the inner detector. This gap, at $1.37 < |\eta| < 1.52$ is referred to as the crack region and is excluded from the analysis. In the endcaps, the electromagnetic calorimeter is segmented into an inner and outer wheel covering $2.5 < |\eta| < 3.2$ and $1.37 < |\eta| < 2.5$, respectively. In the barrel, the electromagnetic calorimeter extends from $R = 2.8 \text{ m}$ to $R = 4\text{m}$ and is segmented into three layers of decreasing granularity. From the first to the third

layer, the granularity is $\eta \times \phi = 0.0031 \times 0.1$, $\eta \times \phi = 0.025 \times 0.025$ and $\eta \times \phi = 0.05 \times 0.025$. The electromagnetic calorimeter thickness is $22X_0$ in the barrel and $24X_0$ in the endcap, where X_0 represents the radiation length². This thickness ensures that electromagnetic showers originating from an electron or a photon are well-contained within the volume of the electromagnetic calorimeter. The energy resolution of the electromagnetic calorimeter is $\frac{\sigma}{E} = \frac{10\%}{\sqrt{E}} \oplus c$ where c is 1-2% depending on η [57].

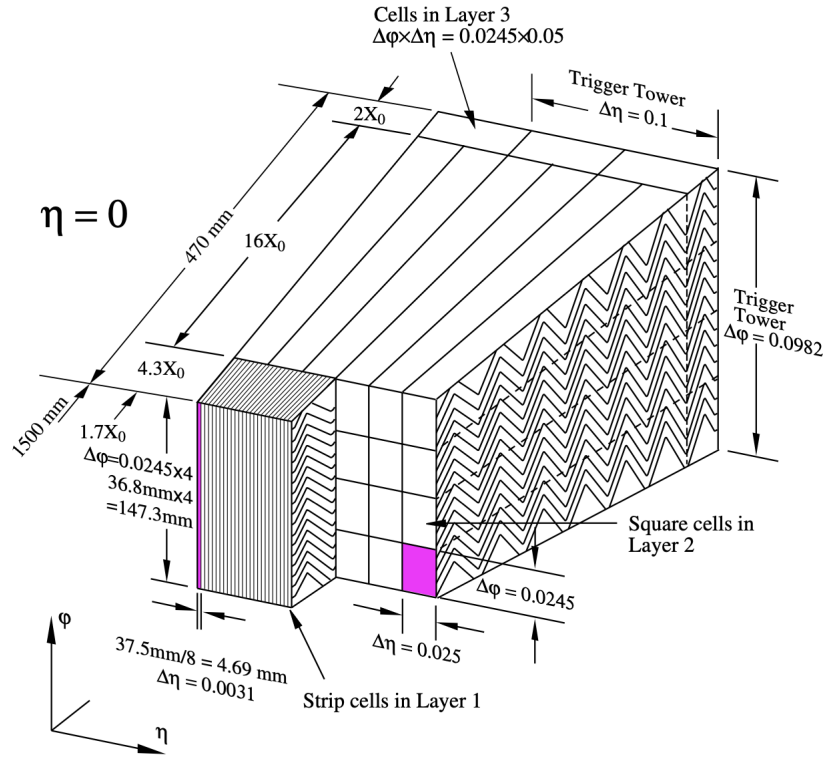


Figure 3.6: Schematic diagram of a section of the electromagnetic calorimeter showing the geometry and granularity of each of the 3 layers in the barrel [58].

The Hadronic Calorimeter

The hadronic calorimeter [50] is composed of active layers of scintillating plastic tiles and absorber layers of steel in the barrel. Charged particles in a hadronic shower produce UV light in the scintillating tiles which is then carried by wavelength shifting fibers to

²The radiation length is the mean distance an electron travels in a material before losing all but $1/e$ of its energy

photomultiplier tubes. In the endcap, active layers of liquid argon are interspersed with absorber layers of copper and the operating principal is the same as that of the electromagnetic calorimeter.

The hadronic calorimeter covers $|\eta| < 1.7$ and $2.28 < R < 4.25$ m in the barrel and $1.7 < |\eta| < 3.2$ and $0.475 < R < 2.03$ m in the endcap. At $\eta = 0$ the hadronic calorimeter thickness is $7.4\lambda_0$, where λ_0 is the nuclear interaction length³ ensuring that the hadronic shower development is contained within the hadronic volume. The hadronic calorimeter energy resolution is $\frac{\sigma}{E} = \frac{52\%}{\sqrt{E}} \oplus 5.7\%$ [59]. Not all of the energy of the hadronic shower is detected within the hadronic calorimeter. Some charged particles from hadronic showers end up being detected in the muon spectrometer located behind the hadronic calorimeter. This "punch through" is one of the many experimental systematic uncertainties in this analysis, which are discussed in detail in Chapter 7.

The Forward Calorimeter

Particles very close to the beam pipe ($3.1 < |\eta| < 4.9$) interact with the forward calorimeter [50]. The forward calorimeter is composed of three layers, each 45 cm in length. The first consists of liquid argon active layers, alternating with copper absorption plates. The other two layers also use liquid argon active layers, but alternate with tungsten absorber layers for their short interaction length. Each active layer consists of electrodes in the shape of rods inserted into a matrix of round tubular holes running parallel to the beam axis and filled with liquid argon. This unique geometry insures the full containment of even the highest energy hadronic showers. The forward calorimeter provides good hermiticity which is important, for example, for the precise calculation of missing momentum in a proton-proton collision.

³The nuclear interaction length λ_0 is the mean free path a particle travels in a given material before undergoing an inelastic nuclear interaction.

3.2.3 The Muon Spectrometer

As muons are an important final state particle in many processes of interest that take place in proton-proton collisions at energies probed at the LHC, the ATLAS detector is equipped with a muon spectrometer [60]. Unlike electrons, photons that are largely contained within the calorimeter system, muons lose a small amount of their energy to bremsstrahlung which initiates electromagnetic showers. In addition, due to their large mass, the bending radius of a muon in a magnetic field is much larger than that of an electron. At large muon transverse momentum of the order 100 GeV, the lack of curvature of muon trajectories in the inner detector degrades the muon transverse momentum resolution of the inner detector [50]. To measure the momentum of high energy muons, the muon spectrometer covers $5 < R < 10$ m in the barrel and $7.4 < |z| < 21.5$ m in the endcap. The rapidity coverage of the muon spectrometer is $|\eta| < 1.4$ in the barrel and $1.6 < |\eta| < 2.7$ in the endcap.

A schematic diagram of the muon spectrometer is shown in Figure 3.7. Four detector technologies are used in the muon spectrometer. Monitored drift tubes (MDTs) and cathode strip chambers (CSCs) are used for precision position measurements while resistive plate chambers (RPCs) and thin gap chambers (TGCs) are used for triggering and for additional position measurements for track reconstruction. Muons bend in the barrel in a 1 T magnetic field provided by a large barrel toroidal magnet and in the endcaps in a 0.5 T magnetic field provided by two smaller endcap toroid magnets. This allows measurement of muon momentum up to 1 TeV. The muon spectrometer is capable of measuring muon momentum down to a transverse momentum of 3 GeV. In combination with the inner detector, muon momentum resolution is better than 3% for a large transverse momentum range and 10% at 1 TeV [61].

Monitored Drift Tubes

Monitored drift tubes are aluminum cathode tubes of 30 mm diameter filled with a 93%-7% argon-CO₂ mixture. A tungsten-rhenium anode wire runs down the center of each

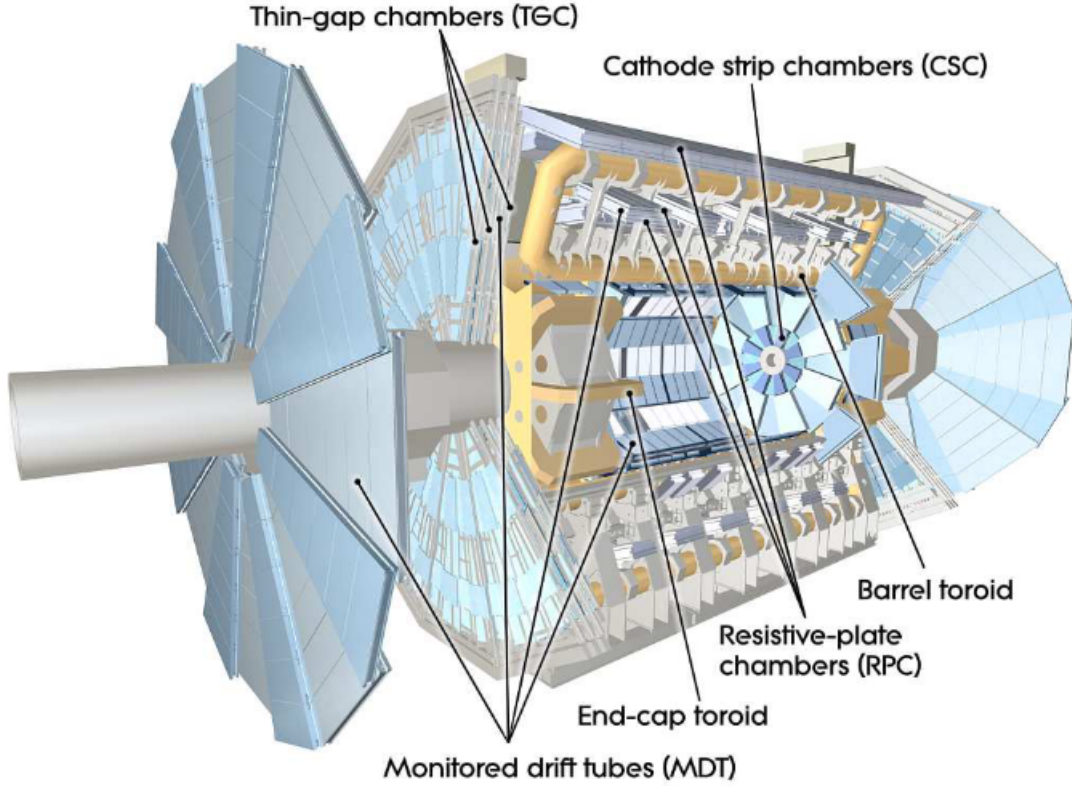


Figure 3.7: Schematic diagram of the ATLAS muon spectrometer [50]. Not shown on the diagram is the first station of Thin-Gap Chambers located on small endcap wheels.

tube and is held at a potential of 3080 V. A muon passing through an MDT ionizes the gas and the subsequent movement of free electrons in the large electric field induces a current on the anode wire which is read out as a signal. The maximum ionization drift time of monitored drift tubes is about 700ns.

MDTs are arranged into multilayers of three or four tubes. Two multilayers make up a chamber. Chambers are arranged into three cylindrical layers in the barrel and four layers in each endcap. The single-hit spatial resolution, averaged over all drift tube chambers is about $80 \mu\text{m}$ [62].

Cathode Strip Chambers

The cathode strip chambers are located in the endcap of the ATLAS detector from $\eta = 2.0$ to $\eta = 2.7$ and must cope with a much higher particle flux than the MDTs. The CSCs are

arranged into two wheels of 8 chambers each. Each chamber is a multi-wire proportional chamber with wires oriented radially out from the beam axis and cathodes segmented into readout strips. The chambers are filled with an argon-CO₂ mixture and the anode wires are held at a potential of 1.9 kV. The spatial resolution of the CSCs is 60 μm per CSC plane.

Resistive Plate Chambers

RPCs are parallel electrode plates operated at a voltage of 9.8 kV. The chambers are filled with a 94.7%-5%-0.3% mixture of C₂H₂F₄-Iso-C₄H₁₀-SF₆ gas. The RPCs are segmented into stations of two chambers. There are three concentric layers of stations in the barrel. A hit coincidence requirement is used as a trigger. While the RPCs have a spatial resolution limited to 1 cm by the electrode strip pitch, they have an excellent timing resolution of 1.5 ns and are therefore excellent systems to use in triggering.

Thin Gap Chambers

Thin Gap Chambers are multiwire proportional chambers. Copper strips and pad electrodes sit 2.8 mm apart. Anode wires separated by 1.8 mm and oriented radially lie between the cathodes and are held at an operating voltage of 3.1 kV. The chambers are filled with a 55%-45% mixture of CO₂ and n-pentane. The TGCs are arranged in four stations in each endcap of the ATLAS detector. The second station has three layers of TGCs while the remaining three have two layers of TGCs. The TGCs provide an azimuthal angular resolution of 2-3 mrad. A timing resolution of better than 25 ns allows them to serve as a trigger in the muon spectrometer endcap.

3.2.4 Luminosity Detectors

The luminosity recorded by the ATLAS detector is measured using the LUCID (Luminosity measurement with a Cherenkov Integrating Detector) detector [63] and the BCM

(Beam Conditions Monitor) [64]. Each detector provides an independent measurement of luminosity, and the combination of the two measurements is used to reduce uncertainty. The LUCID detector is made of 16 photomultiplier tubes (PMT) located at $|z| = 17$ m from the interaction point. Cherenkov photons are produced in quartz windows on the PMTs. The BCM consists of four diamond sensors located at $|z| = 184$ mm and $R = 55$ mm with two sensors oriented along the y -axis and two sensors oriented along the z -axis.

The luminosity produced by a single pair of colliding proton bunches is

$$\mathcal{L} = \frac{\mu f_r}{\sigma_{inel}} \quad (3.7)$$

where μ is the average number of inelastic interactions per bunch crossing, f_r is the LHC bunch revolution frequency, and σ_{inel} is the reference cross section for inelastic proton-proton collisions. Several algorithms can be used for each luminosity detector to go from a number of hits to a luminosity measurement. For example, assuming the number of inelastic proton-proton interactions in a bunch crossing follows a Poisson distribution, the probability of a signal above threshold in one PMT of the LUCID detector is

$$P_{hit} = \frac{N_{hit}}{N_{BC}} = 1 - e^{-\mu_{vis}} \quad (3.8)$$

where N_{hit} is the number of hits above threshold, N_{BC} is the number of proton-proton bunch crossings sampled, and $\mu_{vis} = \epsilon\mu$ where ϵ is the efficiency of the PMT. The quantity σ_{inel} is determined by "van der Meer" scans [65] performed at low luminosity. During Run 2, the LHC provided 156 fb^{-1} of integrated luminosity. Of this total, 140 fb^{-1} of proton-proton collisions were recorded by ATLAS. The uncertainty on the luminosity measurement in LHC Run 2 is 0.8% [66]. The dominant source of uncertainty arises from the extrapolation from the low luminosity regime where the "van der Meer" scans are performed to the high luminosity data taking regime.

3.2.5 The ATLAS Trigger

Proton bunches in the LHC cross at most every 25 ns, corresponding to an event rate of 40 MHz. Available readout bandwidth and data storage prevents recording data from each collision. Thus a *trigger* system is employed to select data in only potentially interesting events. The trigger is segmented into a Level-1 (L1) system which selects events at a rate of up to 100 kHz followed by a high level trigger (HLT) which selects events at a rate of up to 1.5 kHz. The L1 trigger is implemented on custom hardware while the HLT is implemented on commercial off-the-shelf components. If the rate of events passing requirements for the L1 or HLT is higher, a prescale factor is applied to reduce the rate down to these specifications.

The L1 trigger relies on limited granularity information from the calorimeters and muon detectors. In the calorimeters, *trigger towers* of granularity $\Delta\eta \times \Delta\phi = (0.1 \times 0.1)$ along constant η lines are the basic elements used for event selection. Different clustering algorithms for jets, electrons and photons, and taus are used to define regions of interest (ROIs) in the detector volume. The coordinates of the identified regions of interest satisfying a minimum energy threshold are sent to the high level trigger. The information used from the muon spectrometer are hits in the RPCs and TGCs. Event information from events with coinciding hits satisfying a minimum transverse momentum threshold are passed to the HLT.

The HLT runs more complex object reconstruction based on the full detector granularity. The HLT reconstructs only fractions of the detector based on the regions of interest identified by the L1 trigger. In addition, for computationally intensive reconstruction tasks, dedicated algorithms that are less computationally intensive than those used for offline event reconstruction are used at trigger level. This allows the HLT to keep up with the rate of events delivered by the L1 trigger. The HLT processes events within 200 μs of the bunch crossing. Collision events can then be selected based on predefined sets of criteria based on different combinations of detector patterns reconstructed by the L1 trigger and the HLT.

3.3 Event and Detector Simulation

The analysis and interpretation of LHC data requires the ability to simulate proton-proton collisions and the ATLAS detector response. The basic ingredients of event and detector simulation are discussed here.

3.3.1 Event Simulation

The first step in generating simulated events of proton-proton collisions is the computation of the matrix element, which relates the initial state and the final state of a scattering process. This is done to some order in perturbative quantum chromodynamics (QCD), i.e. leading order (LO) or next-to-leading order (NLO). The matrix element calculation involves a complex integration, so the use of Monte Carlo methods are essential. Along with the matrix element, the distribution of parton momentum inside the colliding protons (PDFs) are needed in order to calculate scattering cross sections as discussed in Chapter 2.

The output of event generators is a list of final state particles and their four-vector. This information is interfaced to simulation packages designed to simulate the parton showering and hadronization processes. These packages use parameterized models tuned to real data collected at different particle colliders. The partons that take part in the hard scatter emit gluons which can further emit gluons. This process is called the parton shower and is also simulated. Due to confinement, partons hadronize to form jets. The hadronization itself must be simulated. The partons that do not take part in the hard scatter, i.e. the proton remnant, may also interact in elastic or inelastic collisions. These processes forming the so-called underlying event must also be simulated.

Parton distribution functions as well as inputs to parton shower and hadronization models include many free parameters. These free parameters are determined from experimental measurements. Collaborations, like NNPDF [5], derive parton distribution functions (PDF) from comparisons of theoretical predictions to experimental results.

A table of the event generators, PDFs, parton shower and hadronization packages, and parameter sets used to simulate different processes used in this thesis are shown in Appendix A. In some cases multiple event generators which are interfaced to different hadronization or showering algorithms are available.

3.3.2 Pileup

In addition to events coming from the proton-proton collision that includes the hard scatter, different proton-proton interactions can simultaneously occur in the same proton-proton bunch crossing. This *pileup* is quantified by the number of inelastic collisions in a bunch crossing μ and is measured using the ATLAS minimum bias trigger. The average of that quantity over all proton bunches is quantified by $\langle\mu\rangle$. This quantifies the average number of expected proton-proton collisions per bunch crossing. Pileup is simulated by overlaying the simulated hard scatter with enough simulated minimum bias events to reproduce the observed $\langle\mu\rangle$. Minimum bias interactions refer to any possible inelastic interactions between protons that satisfy the most inclusive and least biased trigger criteria defined during data taking. The pileup distribution $\langle\mu\rangle$ is reweighted to match the $\langle\mu\rangle$ of the recorded dataset.

3.3.3 Detector Response

The output of the simulation of individual proton bunch crossings consists of the momentum 4-vector of all particles generated through the different phases of the simulation chain (proton-proton interaction, underlying event and pileup). These momentum 4-vectors and particle types are then passed to the GEANT4 [67] package to simulate the ATLAS detector response. The detector geometry, materials and electric and magnetic fields are included in the GEANT4 model of the ATLAS detector. Physics processes such as ionization, transition radiation, scintillation, electromagnetic and hadronic showering are simulated. The result is detector output that is in the same format as real data. The

same reconstruction algorithms as well as particle and event selection can then be run on simulated physics samples as is run on real data. This way simulation and data can be compared. In the next chapter, event reconstruction algorithms and event selection are discussed.

Chapter 4

Event Reconstruction and Selection

In this Chapter the reconstruction of photons, leptons, and jets from raw detector signals are described. The selection of events with specific kinematic properties defining a signal enriched data sample is also described.

In the process of interest studied in this thesis, a W boson and a photon are each radiated from the constituent quarks of two protons and scatter off one another as shown in Figure 2.1c. The W boson then promptly decays into a lepton and a neutrino. The expected reconstructed event signature in the detector consists of at least two jets, or collimated sprays of hadronic particles, a photon, either an electron or a muon, and an imbalance of momentum in the transverse plane associated with the production of an undetected neutrino. Event signatures with an electron or positron are hereforward referred to as belonging to the electron channel, while those with a muon or antimuon are referred to as belonging to the muon channel. Signal events are further characterized as having two very high energy forward jets and a centrally produced W and photon system. An event display of a candidate signal like event is shown in Figure 4.1.

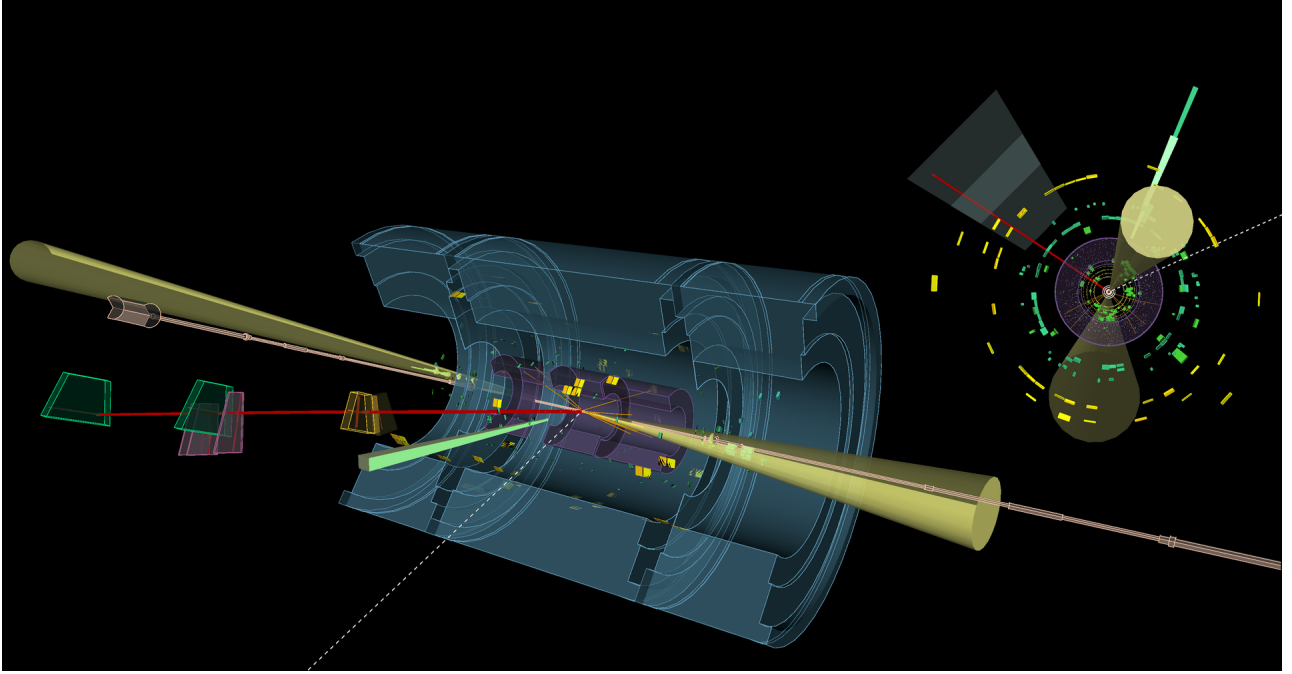


Figure 4.1: Computer reconstruction of a candidate signal event. The event features two forward jets, a muon (indicated in red), a photon (indicated in green), and an imbalance of momentum in the transverse plane (indicated by the dotted line).

4.1 Data Quality and Trigger Requirements

Only data satisfying basic data quality requirements are used in this analysis. The data is required to have been recorded when the ATLAS detector was fully operational, all detector subsystems were calibrated and the trigger and data acquisition systems were functional. The full list of data quality requirements is described in [68]. For an event to be considered, it must pass a single lepton trigger. Single lepton triggers require that an event contained an electron or a muon satisfying a transverse momentum requirement and an identification and isolation requirement. Multiple triggers are used, each with different transverse momentum, identification and identification requirements, depending on the run period. The trigger requirements are summarized in Table 4.1. The raw data from all events satisfying at least one of the single lepton triggers, about 2.9 billion events, are then processed to to identify different types of final state objects.

Channel	Run Period	p_T	Identification	Isolation
Electron	2015	24 GeV	<i>Medium</i>	None
		60 GeV	<i>Medium</i>	None
		120 GeV	<i>Loose</i>	None
	2016-2018	26 GeV	<i>Tight</i>	None
		60 GeV	<i>Medium</i>	None
		120 GeV	<i>Loose</i>	None
Muon	2015	20 GeV	None	<i>Loose</i>
		50 GeV	None	None
	2016-2018	26 GeV	None	<i>Medium</i>
		50 GeV	None	None

Table 4.1: Summary of transverse momentum, identification and isolation requirements for each trigger used in the analysis. For an event to be considered in the analysis, it must satisfy the requirements of at least one trigger.

4.2 Photons

A schematic diagram depicting the expected detector signature from a photon is shown in Figure 4.2. Photons are reconstructed from localized clusters of energy deposited in the electromagnetic calorimeter. Since they do not carry an electric charge, photons do not generate hits in the inner detector. Their trajectory is inferred from the position of the reconstructed energy cluster in the calorimeter. A photon travelling through the inner detector can interact with material in the inner detector and convert into an electron-positron pair. In this case inner detector tracks from the electron and positron can be used in photon reconstruction.

The localized clusters of energy deposited are identified using an algorithm with a sliding window in the $\eta - \phi$ plane of size 3×5 in units of calorimeter cells [69] in the 2^{nd} layer of the electromagnetic calorimeter where the photon deposits most of its energy. Neighboring cells with a combined total energy greater than 2.5 GeV become photon candidate energy clusters.

A photon candidate energy cluster only has a matching track in the inner detector if it has converted into an electron-positron pair. In this case a conversion vertex reconstruction algorithm is used with tracks that are loosely matched to the calorimeter cluster as in-

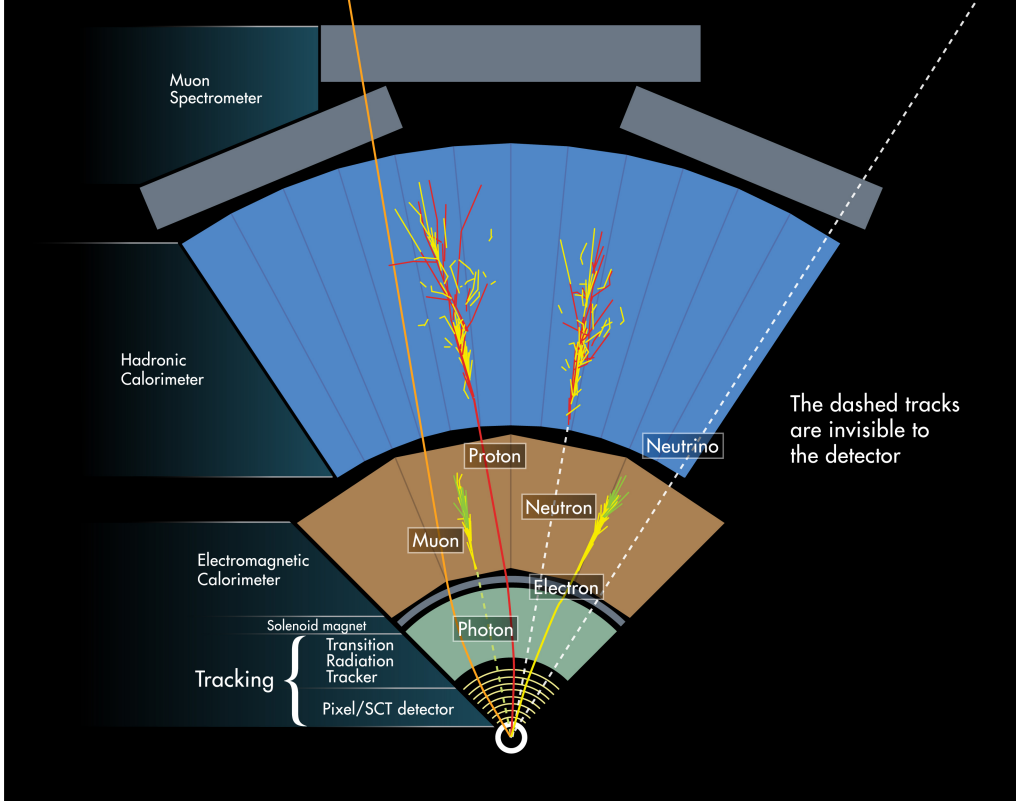


Figure 4.2: Schematic drawing of signals left by various particles in different layers of the ATLAS detector [70].

puts. In some cases where a calorimeter cluster belongs to a converted photon candidate track and an electron candidate track, an ambiguity removal algorithm is applied [69].

Once energy clusters in the electromagnetic calorimeter are identified, additional criteria are used to identify electromagnetic showers most likely originating from the passage of photons through the electromagnetic calorimeter. Photon identification quality is based on nine different variables that parameterize the shape of the reconstructed shower in the electromagnetic (EM) calorimeter. The cuts on these values are dependent on the cluster transverse energy, η , and whether the photon is converted or not. The variable definitions are given below, as well as the cut used to determine if a candidate passes or fails the cut for unconverted photon candidates with $25 \text{ GeV} < E_T < 30 \text{ GeV}$ and $\eta < 0.6$:

- R_{had} : The ratio of transverse energy deposited in the first layer of the hadronic calorimeter to the cluster transverse energy. $R_{had} < 0.042$.

- R_η : Ratio of energy measured in a 3×7 cell area in $\eta \times \phi$ around the photon candidate to the energy measured in a 7×7 cell area in $\eta \times \phi$ around the photon candidate in the middle layer of the EM calorimeter. $R_\eta > 0.920$.
- R_ϕ : Ratio of energy measured in a 3×3 cell area in $\eta \times \phi$ around the photon candidate to the energy measured in a 7×3 cell area in $\eta \times \phi$ around the photon candidate in the middle layer of the EM calorimeter. $R_\phi > 0.935$.
- w_{η_2} : Energy weighted lateral shower width in the middle layer of the EM calorimeter in units of pseudorapidity. $w_{\eta_2} < 0.0115$.
- w_{s3} : Energy weighted lateral shower width in the first layer of the EM calorimeter, in units of cells, calculated from energy deposited in three strips around the strip with the highest energy. $w_{s3} < 0.692$.
- w_{tot} : Energy weighted lateral shower width in the first layer of the EM calorimeter, in units of cells, calculated from energy deposited in 20 strips around the strip with the highest energy. $w_{tot} < 3.71$.
- F_{side} : The fraction of energy outside the three central strips, but within 7 strips. $F_{side} < 0.305$.
- ΔE : The difference in energy between the strip with the second largest energy deposited and the strip with the smallest energy deposit between the two leading strips. $\Delta E < 239 \text{ MeV}$.
- E_{ratio} : The ratio of the energy difference between the largest and second largest energy deposits over the sum of these energies $E_{ratio} > 0.82$.

A signal photon in this analysis is a photon that passes transverse momentum and pseudorapidity dependent cuts on all 9 of the shower shape variables above. A candidate signal photon satisfying all these conditions is referred to as a *tight* photon. The *tight*

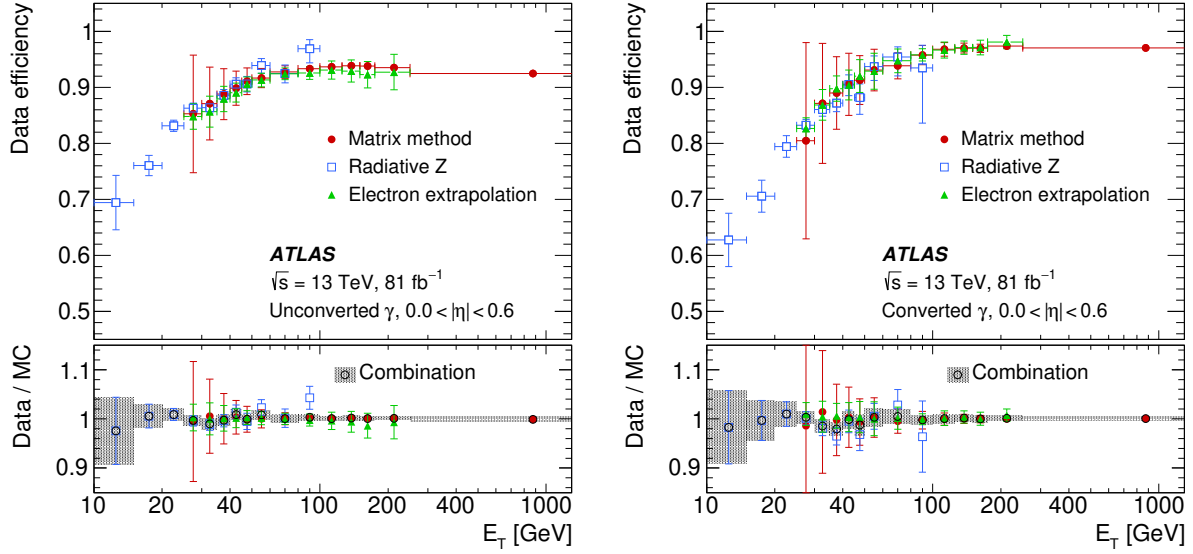


Figure 4.3: Efficiencies of the *tight* photon identification definition for unconverted (left) and converted (right) photons. The ratio of data to simulation efficiencies used as a correction scale factor is also shown. The “Matrix method”, “Radiative Z” and “Electron extrapolation” refer to three different independent methods of assessing the identification efficiency. These plots are taken from [69].

identification requirement selects narrow clusters of energy that are characteristic of photons. The *tight* photon identification efficiency for unconverted and converted photon is shown in Figure 4.3.

In addition to *tight* photons, *non-tight* photons are also used in the estimation of the background due to non-prompt photons from jets described in Section 5.1. The *non-tight* working points are defined as electromagnetic clusters that are allowed to fail at least one of a subset of the *tight* selection criteria cuts. The subsets of the different *non-tight* working points are listed below:

- *Loose'2*: w_{s3}, F_{side}
- *Loose'3*: $w_{s3}, F_{side}, \Delta E$
- *Loose'4*: $w_{s4}, F_{side}, \Delta E, E_{ratio}$
- *Loose'5*: $w_{s4}, F_{side}, \Delta E, E_{ratio}, w_{tot}$

The Loose'4 working point is used to define the default non-prompt photon enriched control region in Section 5.1. The *non-tight* working points allow for no overlap with the *tight* photon sample. Even with the *tight* requirement, hadronic jets may be occasionally misidentified as photons. In addition photons not originating from the hard scatter, such as photons from neutral pion decays may be classified as *tight* photons. To further distinguish signal photons from photons from jets or hadrons, the photon candidate is additionally required to be isolated.

An illustration of photon isolation energy can be seen in Figure 4.4. The raw isolation energy, $E_{T,raw}^{cone40}$, is the sum of transverse energy clusters in the calorimeter located within a distance of $\Delta R = 0.4^1$ from the photon candidate as found by the sliding window algorithm introduced above. The photon energy ($E_{T,core}$) is subtracted from the raw isolation energy by subtracting the energy contained in a $\Delta\eta \times \Delta\phi = 5 \times 7$ rectangle around the center of the photon cluster. As the photon energy may not be contained in this 5×7 window, a transverse momentum and η dependent leakage correction is applied. An η dependent average pileup contribution is also subtracted from the raw isolation energy [69]. The photon calorimeter isolation is then expressed

$$E_T^{cone40} = E_{T,raw}^{cone40} - E_{T,core} - E_{T,leakage}(E_T, \eta) - E_{T,pile-up}(\eta) \quad (4.1)$$

where $E_{T,raw}^{cone40}$ is the sum of transverse energy in a distance $\Delta R = 0.4$ from the photon center, $E_{T,core}$ is the sum of transverse energy in the $\eta \times \phi = 5 \times 7$ window, and $E_{T,leakage}$ and $E_{T,pile-up}$ are the leakage and pileup corrections, respectively.

An isolated photon is defined as a photon candidate satisfying the condition

$$E_T^{cone40} < 0.022 E_T^\gamma + 2.45. \quad (4.2)$$

The isolation requirement is of the form $X E_T + Y$ as at lower transverse momentum, the rate of background processes is higher. The values of X and Y were chosen to optimize

¹ $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$

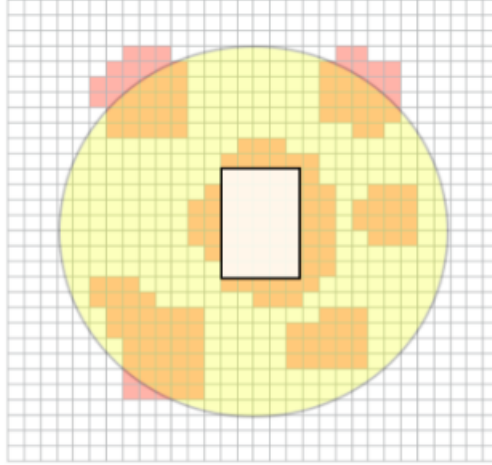


Figure 4.4: An illustration of photon isolation energy. The isolation energy is defined as the sum of energy deposited in a cone of $\Delta R < 0.4$ around a photon candidate minus the sum of energy in a 5×7 cell window. Additional corrections for photon energy leakage and pileup are applied.

signal efficiency for identifying photons originating from Higgs boson decays and rejection of jet misreconstructed as photons.

In summary, photons in this analysis are defined by clusters of energy in the electromagnetic calorimeter. Photons can be either “unconverted”, if they are not matched to tracks in the ID, or “converted” if one or two tracks point back to the proton-proton collision. A signal photon is defined to be a *tight* and isolated photon. For simulated events, scale factors are used to match the efficiency of photon identification and isolation cuts between data and simulation.

4.3 Electrons

Electrons are reconstructed as a cluster of energy deposited in the electromagnetic calorimeter, similar to that from a photon, but with matching tracks that point back to the origin of the proton-proton collision as can be seen in Figure 4.2. Thus electron reconstruction relies on two intermediate objects in addition to clusters in the calorimeter: tracks in the inner detector and the object point of origin, or the vertex.

The first step in track reconstruction is clustering pixels or strips that record energy deposited above a threshold in the pixel tracker and SCT, respectively. Three resulting space points are used to seed a track which is extrapolated using a Kalman filter [71]. Track candidates are then removed against one another using a score based on the number of hits and the χ^2 fit of the track. Tracks with $p_T > 400$ MeV and at least 7 space points within $|\eta| < 2.5$ are kept.

Tracks are then matched to a reconstructed origin of the proton-proton collision, called a vertex. Vertices are reconstructed using only high quality tracks to ensure good spatial resolution. The definition of high quality tracks as well as a description of the vertex finding and fitting procedure is given in [72]. The vertex made up of tracks with the highest total transverse momentum is taken as the *primary vertex* in the event.

A candidate electron consists of a cluster of energy in the electromagnetic calorimeter that is matched to a track that points back to the primary vertex. This association of tracks with a primary vertex is quantified with cuts on the transverse impact parameter and the longitudinal impact parameter. The transverse impact parameter is defined as $|d_0/\sigma_{d_0}|$, where d_0 is the shortest distance between the track and the primary vertex in the transverse plane and σ_{d_0} is the uncertainty on d_0 . The longitudinal impact parameter is defined as $|z_0 \sin \theta|$ where z_0 is the shortest distance between the track and the primary vertex projected on the z-axis. A signal electron is required to have a transverse impact parameter less than 5 and a longitudinal impact parameter less than 0.5 mm. Similarly to photons, additional requirements are applied to select good signal electrons [69]. The identification of electrons is defined using a likelihood discriminant based on track quality, track to cluster matching, detection of transition radiation in the TRT, and the reconstructed electromagnetic shower shape. The isolation requirement is $E_T^{cone20} < 0.06 p_T$ where E_T^{cone20} is defined similarly to E_T^{cone40} as described above, but with a cone size of $\Delta R = 0.2$. There is an additional isolation requirement on the electron track: $p_T^{cone30} < 0.06 p_T$ where p_T^{cone30} is the sum of the transverse momentum of tracks in a cone of $\Delta R = 0.3$ around the electron candidate track. *Tight* and isolated electrons are used as signal electrons in this analysis,

but these cuts are inverted for background estimates discussed in Sections 5.2 and 5.3. Again, scale factors are used to match the efficiency of isolation, identification and trigger requirements for electrons between data and simulation.

4.4 Muons

Muons are identified as a track reconstructed in the inner detector matched, in $\eta - \phi$, with a track reconstructed in the muon spectrometer. In the muon spectrometer, each detector subsystem (i.e. MDTs, CSCs, TGCs, RPCs) has a dedicated algorithm to produce *track segments*. Track segments that are matched to the interaction point are then combined, forming muon spectrometer track candidates. A muon track candidate must have at least two track segments in the barrel or one track segment in the endcap.

A combined muon track can be formed by fitting the combined hits from the inner detector and the muon spectrometer. Since the muon spectrometer and the inner detector do not have full coverage in η , multiple muon reconstruction algorithms are considered to recover efficiency [61].

The requirements on the muon transverse and longitudinal impact parameters are $|d_0/\sigma_{d_0}| < 3$ and $|z_0\sin\theta| < 0.5$ mm, respectively. Track information is used to define isolation and identification working points [61]. The particle flow algorithm is used for muon isolation, and combines track and calorimeter information. An isolated muon candidate satisfies $p_T^{cone30} + 0.4E_N < 0.045p_T$ where E_N is a measure of energy from neutral energy flow objects in the calorimeter. Muon isolation is described in detail in [73]. Signal muons are required to be *tight* and isolated. For background estimates discussed in Section 5.3 the identification, isolation, and impact parameter cuts are inverted. Scale factors are used to match the efficiency of muon isolation, identification and trigger cuts in data and simulation.

4.5 Jets

Both inner detector track information and energy deposited in the calorimeter are used in jet reconstruction. Energy deposited in neighboring cells above a specified signal-to-noise ratio in the hadronic calorimeter are combined to form *clusters* [74]. These energy clusters are combined using the anti- k_t algorithm [75]. The anti- k_t algorithm compares the distance, d_{ij} , between two energy clusters

$$d_{ij} = \min(k_{T,i}^{-2}, k_{T,j}^{-2}) \frac{\Delta_{ij}}{R} \quad (4.3)$$

with the distance between the energy cluster and the jet axis, $d_{i,B}$. The quantity $k_{T,i}$ is the transverse energy of cluster i , and $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the geometrical distance between clusters in $\eta \times \phi$. If $d_{i,B} > d_{ij}$, cluster i is merged into a jet, otherwise, cluster i forms a new jet.

Jet energy clusters are then combined with tracks using the particle-flow algorithm [76]. The leading transverse momentum track is matched to a jet energy cluster in the calorimeter. The expected energy in the calorimeter deposited by the particle that created the track is calculated from the track momentum and the position of the cluster. Based on this calculation, energy clusters are combined until the energy cluster matches the expected energy. The particle flow algorithm is then applied to the remaining tracks and calorimeter clusters in order of decreasing track transverse momentum.

4.6 Missing Transverse Energy

The transverse energy of neutrinos can be inferred using conservation of momentum in the transverse plane. The missing transverse energy, E_T^{miss} , is defined as

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}. \quad (4.4)$$

The quantities E_x^{miss} and E_y^{miss} are defined as the magnitude of the negative vector sum of the reconstructed objects in the event projected on to the x or y axis. Included in the reconstructed objects are tracks that are matched to the primary vertex but not matched to reconstructed particles such as photons, electrons, muons or jets.

4.7 Object Overlap Removal

It is possible for detector signals to be used in the reconstruction of more than one object (e.g. photons, electrons, muons and jets). To avoid this, an object overlap removal procedure is applied. The following requirements are applied in the order they are described.

- If two electron candidates reconstructed in the electromagnetic calorimeter share the same track, the electron candidate with the lower transverse momentum is discarded.
- If a muon candidate reconstructed in the calorimeter shares a track with an electron, the muon is discarded.
- If a muon reconstructed in the muon spectrometer shares a track with an electron candidate reconstructed in the electromagnetic calorimeter, the electron candidate is discarded.
- Photon candidates reconstructed within $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.4$ of electron and muon candidates are discarded.
- Jet candidates reconstructed within $\Delta R < 0.2$ of an electron candidate are discarded.
- Electron candidates reconstructed within $\Delta R < 0.4$ of remaining jet candidates are discarded.
- Jets reconstructed within $\Delta R < 0.2$ of a muon candidate are discarded.

- Muon candidates reconstructed within $\Delta R < 0.4$ of remaining jet candidates are discarded.
- Photons candidates reconstructed within $\Delta R < 0.4$ of remaining jet candidates are discarded.

4.8 Event Selection

With the basic analysis objects defined, various criteria are applied to the events in the dataset in order to select events with kinematic characteristics expected from the physics reaction of interest.

First, events are required to contain at least one reconstructed vertex. In events with greater than one reconstructed vertex, the primary vertex is taken as the vertex with the highest sum of track transverse momentum. The event is required to contain at least 2 jets with transverse momentum greater than 50 GeV and $|\eta| < 4.4$. The large jet transverse momentum requirement limits the background from jets originating from another proton-proton collision. At least one photon with $|\eta| < 2.37$ and transverse momentum greater than 22 GeV is required. Both of these cuts suppress the background originating from hadrons misreconstructed as photons. The finely segmented first layer of the electromagnetic calorimeter does not cover beyond $|\eta| = 2.37$ and the fraction of photons originating from misreconstructed hadrons is greater at low photon transverse momentum. The event is required to have exactly one lepton (electron or muon) that is well identified and isolated. An electron is required to have $|\eta| < 2.47$ consistent with the central electron acceptance of the tracker and the electromagnetic calorimeter. Similarly muons are required to have $|\eta| < 2.5$. Photons and electrons have the additional requirement that they fall outside the region $1.37 < |\eta| < 2.47$ to avoid the crack in the electromagnetic calorimeter between the barrel and the endcap where photon and electron energy is poorly measured. The lepton transverse momentum cut is required to be greater than 30 GeV, to be consistent with the region where the trigger efficiency is well modelled in

simulation [77,78]. To select events where the lepton likely comes from the leptonic decay of a W boson, events are required to have a missing transverse energy above 30 GeV.

The above kinematic requirements select a phase space with good acceptance of final state objects as well as a phase space where the final state object momentum and energy are well measured. These requirements do not eliminate background processes with similar kinematics as the signal process, nor do they completely eliminate backgrounds arising from mis-reconstruction and mis-identification of final state objects. Additional cuts are applied to reduce these sources of background.

A Z boson decays to two leptons approximately 10% of the time. To avoid background from leptonic decays of Z bosons, it is required that there is no second lepton in the event, with the same flavor as the first lepton, that passes loose identification. To suppress the background arising from events where one of the leptons from the Z boson decay is mis-reconstructed as a photon, it is required that the invariant mass of the combined lepton and photon system be at least 10 GeV away from the Z mass, i.e. $|M_Z - M_{l\gamma}| > 10 \text{ GeV}$.

To reduce the background arising from events where jets originate from a different vertex than the primary vertex, an algorithm called a jet vertex tagger is used to select jets that likely come from the vertex of interest based on the information of the tracks in the jet [79]. A similar tagger is used to reduce the background from events with a top quark. Top quarks predominantly decay to a W boson and a bottom quark. A tagger is used to identify events with jets that likely arise from bottom quarks [80]. Requiring that no events contain a so-called *b-jet* reduces the background originating from the production of top quarks.

An additional requirement that the two leading jets be well separated from the missing transverse energy ($\Delta\phi > 0.4$) is required to reduce the background coming from events where mismeasurement of jet transverse momentum generates an artificial large missing transverse energy.

As described in Chapter 2 events with the same final state arising from the QCD production of a W boson and a photon in association with jets constitute a large irreducible

background and there is interference between the QCD and electroweak production. This interference is largest when the two leading jets are close together in rapidity. To limit the effect of interference, the two leading transverse momentum jets are required to be separated by 2 units of rapidity.

Similarly, as described in Chapter 2, multiple processes contribute to the electroweak production of a W boson and a photon in association with jets. The quartic coupling diagram is of interest because of its sensitivity to beyond Standard Model physics. This diagram dominates at large jet energies. Thus an invariant mass cut of $M_{jj} = 500$ GeV is required to be maximally sensitive to the diagram including the quartic coupling.

The object and event selection used to define a data sample for the analysis are summarized in Table 4.2. In addition, Table 4.3 shows the number of events in data at selected points in the event selection. In the following chapters this is referred to as the *Baseline Selection*. With the signal sample defined, the next chapter describes the background estimation for the analysis.

Object	Selection Criteria
Photon	$p_T > 22 \text{ GeV}$ $ \eta < 2.37$ $E_T^{cone40} < 0.022 p_T + 2.45 \text{ GeV}$ <i>Tight Identification</i>
Electrons	$p_T > 30 \text{ GeV}$ $ \eta < 2.5$ $ d_0/\sigma_{d_0} < 5$ $ z_0 \sin\theta < 0.5 \text{ mm}$ $E_T^{cone20} < 0.06 p_T$ $p_T^{cone30} < 0.06 p_T$ <i>Tight Identification</i>
Muons	$p_T > 30 \text{ GeV}$ $ \eta < 2.5$ $ d_0/\sigma_{d_0} < 3$ $ z_0 \sin\theta < 0.5 \text{ mm}$ $p_T^{cone30} + 0.4 E_N < 0.045 p_T$ <i>Tight Identification</i>
Jets	$p_T > 50 \text{ GeV}$ $ \eta < 4.4$ Jet Vertex Tag No b-jets
Event	$ M_Z - M_{l\gamma} > 10 \text{ GeV}$ $E_T^{miss} > 30 \text{ GeV}$ $M_T^W > 30 \text{ GeV}$ Object Overlap Removal $\Delta\Phi(j_{0(1)}, E_T^{miss}) > 0.4$ $M_{jj} > 500 \text{ GeV}$ $\Delta Y(j_0, j_1) > 2$

Table 4.2: Summary of object and event selection criteria used in this analysis.

Selection	Number of Events in Data
Has a Primary Vertex	5.4×10^9
Passes Triggers	2.9×10^9
Passes Photon Criteria	6.8×10^6
Passes Lepton Criteria	4.8×10^6
Passes Jet Criteria	1.0×10^6
$ M_Z - M_{l\gamma} > 10 \text{ GeV}$	6.3×10^5
$E_T^{miss} > 30 \text{ GeV}$	3.4×10^5
$M_T^W > 30 \text{ GeV}$	2.9×10^5
No b-jets	1.9×10^5
$p_{j_{0(1)}}^T > 50 \text{ GeV}$	5.5×10^4
$M_{jj} > 500 \text{ GeV}$	8.1×10^3
$\Delta Y(j_0, j_1) > 2$	6.5×10^3

Table 4.3: Number of events in data at various points in the event selection.

Chapter 5

Background Estimation

In this section the background estimate of the analysis is described. The event selection described in Chapter 4 is designed to select events with two jets, a prompt photon, and a lepton and missing transverse momentum in a phase space sensitive to the quartic coupling between W bosons and photons; however, other physics events invariably enter the selected data sample. First there is a prompt background, that is to say events which contain the same final state objects as the signal process. Principal among these is the QCD production of a W boson, a prompt photon and two jets. This is the largest real background in this analysis. Events containing a prompt photon and a top quark also constitute a large real background. Since top quarks primarily decay into W bosons and b -jets, an event where the W boson decays leptonically will share the final state with the signal. While the b -jet veto is used to reject these events, it is less than perfectly efficient, so a large amount of this so-called top background enters the data sample selected. Events where a leptonically decaying Z boson, a photon, and jets are produced but where one of the leptons from the Z boson is not reconstructed will also resemble signal events. Contributions from the different sources of real background are estimated using simulated event samples. The numbers of events associated with the QCD and electroweak production of $W\gamma jj$ are simultaneously determined from a fit to data as described in Chapters 8 and 9.

In addition to the prompt background, there is a non-prompt background arising from events where either the lepton or the photon do not come from the hard scatter. For example, a π^0 might be produced in a jet and decay to two photons. One of these photons might pass the criteria for a prompt photon. In this way events where a W boson is produced in association with jets or events with a top quark pair might enter the signal region. The non-prompt photon background is the largest non-prompt background of this analysis. Similarly a lepton from a jet could be misidentified as a prompt lepton. Since photons and electrons have similar signals in the electromagnetic calorimeter, tracking inefficiencies and less than perfect calorimeter cluster to inner detector track matching can result in electrons being misidentified as photons. In this way events with Z bosons produced in association with jets can enter the signal region. Finally since the photon point of origin is not well measured for unconverted photons, events where the photon comes from a different collision can also enter into the data sample selected. This is termed the pileup photon background.

Each of the non-prompt sources of background discussed above is estimated using a data driven method. The data driven estimates are described and validated in the sections that follow.

5.1 Jets Misreconstructed as Photons

The modelling of hadronic showers in the electromagnetic calorimeter is imperfect. A data driven method called the template fit method is used to estimate the background arising from jets that are misreconstructed as photons. The results of this method are checked with a 2D sideband method and with a closure test performed on simulated data, discussed in Appendix B and C, respectively.

5.1.1 The Template Fit Method

The template fit method exploits the difference in shape of the distribution of photon isolation energy between *tight* and *non-tight* reconstructed photons. Photon isolation energy is described in Chapter 4. Recall that an isolated photon satisfies the condition

$$E_T^{cone40} < 0.022 E_T^\gamma + 2.45\text{GeV} \quad (5.1)$$

To estimate the number of jets faking *tight* and isolated photons the method proceeds as follows in each bin of a kinematic distribution

1. A template shape of $E_T^{cone40} - 0.022 E_T^\gamma$ expected from prompt photons is obtained by fitting the $E_T^{cone40} - 0.022 E_T^\gamma$ distribution of reconstructed *tight* photons in simulated data that originate from simulated truth photons.
2. A template shape of $E_T^{cone40} - 0.022 E_T^\gamma$ expected from jets misreconstructed as photons (non-prompt photons) is extracted by fitting the $E_T^{cone40} - 0.022 E_T^\gamma$ distribution of events containing a *non-tight* photon in data.
3. A probability distribution function is made by summing the prompt photon and the non-prompt photon templates. The normalization of the prompt and non-prompt contributions are then fit to data that contains a reconstructed *tight* photon. In the fit to data, the parameters of the two template shapes are floated, but are constrained by the statistical uncertainties around their best fit values obtained in steps (1) and (2). In this way statistical uncertainties on the template fits are propagated to the final result.
4. The estimated number of jets misreconstructed as photons is finally obtained by integrating the fitted $E_T^{cone40} - 0.022 E_T^\gamma$ template for non-prompt photons from -20 up to 2.45 GeV. This corresponds to the $E_T^{cone40} - 0.022 E_T^\gamma$ selection criteria used to identify isolated photons in the event selection.

In the following sections, $f_P(x|\theta_P)$ and $f_{NP}(x|\theta_{NP})$ denote the *prompt* photon probability density function of true photons, and the *non-prompt* photon probability density function, respectively. The variable x represents the kinematic variable $E_T^{cone40} - 0.022 E_T^\gamma$ and θ , the free parameters of the fit.

1. Prompt photon template extraction

The prompt photon template is obtained from a fit to the isolation energy distribution of simulated electroweak signal and QCD $W\gamma jj$ events where the basic event selection is applied without the isolation selection.

A double-sided Crystal Ball function [81] is chosen for the prompt photon shape. A Crystal Ball function, $f_{CB}(x)$ is a Gaussian convoluted with a power law tail and is defined as

$$f_{CB}(x) = \begin{cases} e^{-\frac{(x-x_0)^2}{2\sigma^2}} & \frac{x-x_0}{\sigma} > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n e^{-\frac{|\alpha|^2}{2}} \left(\frac{n}{|\alpha|} - |\alpha| - \frac{x-x_0}{\sigma}\right)^{-n} & \frac{x-x_0}{\sigma} \leq -\alpha \end{cases} \quad (5.2)$$

where x_0 and σ are the Gaussian mean and width, α is a cutoff parameter and n is the power of the exponentially decaying tail. A double-sided Crystal Ball is the addition of two Crystal Ball functions sharing the same mean and width. The double-sided Crystal Ball has six free parameters: one parameter defining the Gaussian mean x_0 , one parameter defining the Gaussian width σ , two cutoff parameters $\alpha_{1(2)}$, and two parameters describing the exponentially decaying tail $n_{1(2)}$.

The prompt photon template, $f_P(x|\theta_P)$ is extracted by performing an unbinned fit of $E_T^{cone40} - 0.022 E_T^\gamma$ of *tight*, truth-matched photons from simulated $W\gamma jj$ events with the double-sided Crystal Ball function. The fit minimizes the negative log-likelihood given by

$$-\ln \mathcal{L} = -\ln \left[\prod_i f_P(x|\theta_P) \right] \quad (5.3)$$

where i runs over each event. An example of this fit is shown Figure 5.1.

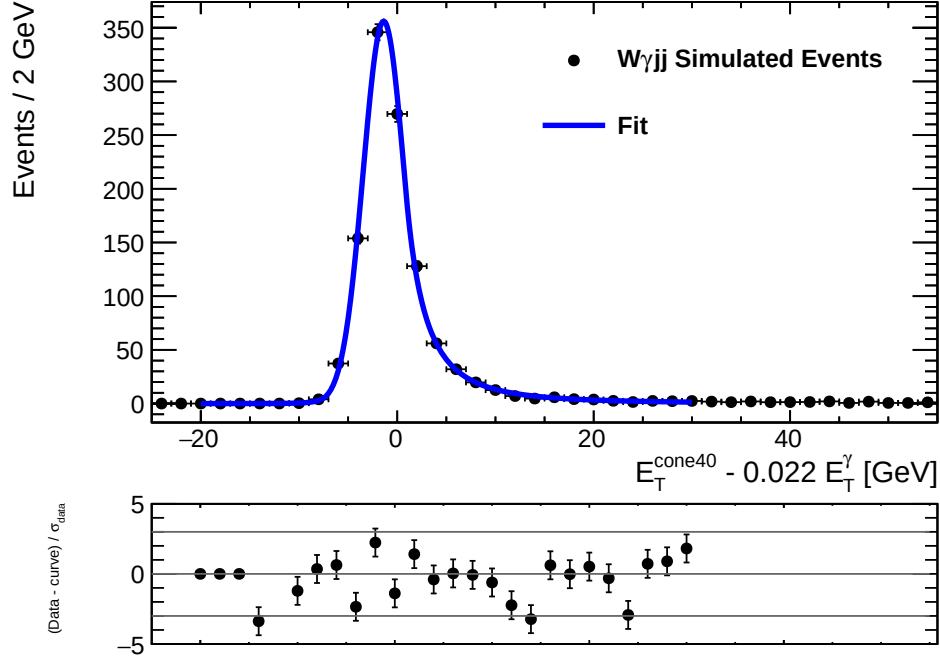


Figure 5.1: Example fit of the photon isolation energy distribution of *tight* photons obtained using simulated electroweak and QCD $W\gamma jj$ events. This fit is used to extract the prompt photon template shape.

2. Non-prompt photon template extraction

The non-prompt photon template is obtained with a fit to the $E_T^{\text{cone40}} - 0.022 E_T^\gamma$ distribution of a data sample defined by *non-tight* identification criteria in place of *tight* identification criteria. This change in the photon identification creates a data sample enriched in jets misreconstructed as a photons.

A Novosibirsk function [82] is chosen to describe the non-prompt photon shape. The Novosibirsk function is defined as

$$f_N(x) = e^{-0.5(\ln q_y)^2 / \Lambda^2 + \Lambda^2} \quad (5.4)$$

where

$$q_y = 1 + \frac{\Lambda(x - x_0)}{\sigma} \times \frac{\sinh(\Lambda\sqrt{\ln 4})}{\Lambda\sqrt{\ln 4}} \quad (5.5)$$

and x_0 is the peak position, σ is the width of the peak and Λ parameterizes the tail of the distribution.

The negative log-likelihood minimized to extract the non-prompt photon template, $f_{NP}(x|\theta_{NP})$ is

$$-\ln \mathcal{L} = -\ln \left[\prod_i f_{NP}(x|\theta_{NP}) \right] \quad (5.6)$$

where i runs over the number of events. An example of the fit used to extract the shape of the non-prompt photon template is shown in Figure 5.2.

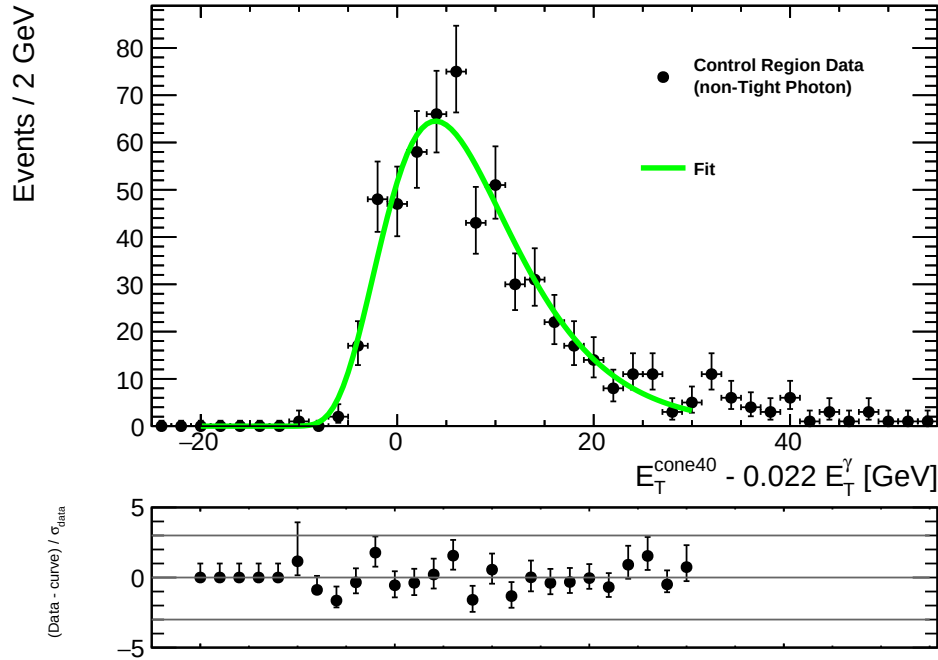


Figure 5.2: Example fit of the photon isolation energy distribution of data events from a control region enriched in jets misreconstructed as photons. The result of the fit is used to extract the shape of the non-prompt photon template.

3. Normalization fit

Finally the resulting prompt and non-prompt photon templates are summed and their relative normalizations fit to data satisfying the signal selection. In this step the template parameters are again floated but constrained by multivariate Gaussian functions.

The multivariate Gaussian functions are defined to have a mean and width given by the best-fit values and statistical uncertainties of the template parameters. The negative log likelihood used to fit the data is given by

$$-\ln \mathcal{L} = -\ln \left[\prod_i f_P(x|\theta_P) f_{NP}(x|\theta_{NP}) G(\theta_P|\sigma_{\theta_P}) G(\theta_{NP}|\sigma_{\theta_{NP}}) \right] \quad (5.7)$$

where G are multivariate Gaussian constraints and i runs over the number of events. The symbols θ_P and θ_{NP} represent the parameters of the *prompt* and *non-prompt* photon templates, respectively, and the quantities σ_{θ_P} and $\sigma_{\theta_{NP}}$ are the statistical errors on those parameters obtained in the fits to extract the *prompt* and *non-prompt* photon shapes. An example of the fit performed to extract the relative normalizations of the two templates is shown in Figure 5.3.

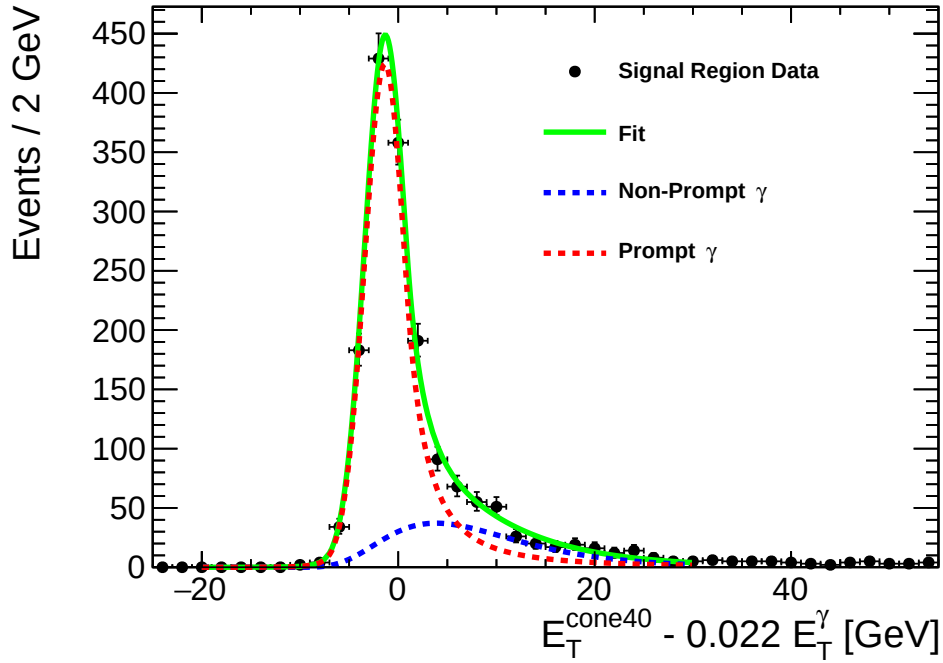


Figure 5.3: Example of the fit to data containing reconstructed prompt photon data performed to extract the relative normalizations of the prompt photon and non-prompt photon templates.

It is valuable to understand how the values of the free parameters of the templates change in the fit to the normalization compared to their values obtained in the fit to extract the template shapes. For this the *pull* is a useful quantity. The pull is defined as

$$pull(\theta) = \frac{\hat{\theta} - \theta_0}{\Delta\theta} \quad (5.8)$$

where $\hat{\theta}$ is the value of the parameter resulting from the fit used to extract the template, $\Delta\theta$ is the uncertainty on the parameter after the fit used to extract the template and θ_0 is the value of the parameter obtained from the normalization fit. A large pull of one of the prompt photon template parameters would indicate that the modelling of photon isolation energy is imperfect, while a large pull of the non-prompt photon template parameters could indicate that the shape of photon isolation energy of non-prompt photons that satisfy the *tight* photon working point is different from the shape of photon isolation energy of non-prompt photons that satisfy the *non-tight* working point. Figure 5.4 shows an example of template fit parameter pulls. Neither large pulls of the prompt photon template parameters nor non-prompt photon template parameters are seen.

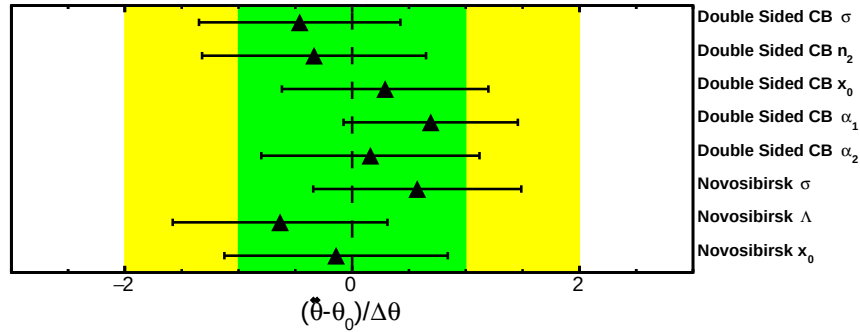


Figure 5.4: Example template parameter pulls resulting from the fit of the prompt and non-prompt photon normalizations to signal region data. The green and yellow bands represent 1σ and 2σ where σ is the statistical uncertainty on the template parameter values obtained in the fits used to extract the prompt and non-prompt photon shapes.

4. Estimation of Event Yields

To obtain a number of events containing a jet misreconstructed as a photon the *non-prompt* photon template after normalization is integrated from -20 to 2.45 GeV in order to apply the criteria $E_T^{cone40} - 0.022 E_T^\gamma < 2.45$ used to define an isolated photon.

5.1.2 Uncertainties on Jets Misreconstructed as Photon Estimate

Four sources of uncertainty on the estimate of the background from jets misreconstructed as photons are considered:

1. Statistical error on the estimated number of events containing a jet misreconstructed as a photon.
2. Uncertainty in the non-prompt photon template shape associated with the leakage of real photons into the non-prompt photon enriched control region.
3. Uncertainty in the non-prompt photon template shape associated with the definition of the *non-tight* photon identification.
4. Uncertainty in the modelling of the prompt photon template shape.

Statistical error on yield: Statistical uncertainty on the yield accounts for the limited statistics in the regions used to defined the template parameters and to estimate the number of jets misreconstructed as photons. Statistical uncertainty is propagated to the final result by allowing the template fit parameters to float during the fit of normalization but constraining parameters by multivariate Gaussian constraints as described above. The statistical uncertainty on the estimated number of events containing a jet misreconstructed as a photon is obtained by integrating the function describing the non-prompt photon contribution with its best fit parameters varied by their statistical uncertainties..

Uncertainty due to real photon leakage: The nominal method does not account for prompt photon leakage into the *non-tight* control region in data used to derive the parameters of the non-prompt photon template. The effect of prompt photon leakage is studied by including an additional template in the non-prompt photon fit to account for prompt photons in the non-prompt region. This template is called the leakage template and is fit to prompt photons from simulated electroweak and QCD $W\gamma jj$ events that fail the *tight* photon identification criteria. As is the case for *tight* photons, a double-sided Crystal Ball function is used to describe the shape of this template. An example leakage template shape is shown in Figure 5.5. The negative log likelihood which is minimized in the fit to *non-tight* photon data then becomes

$$-\ln \mathcal{L} = -\ln \prod_i [f_{NP}(x|\theta_{NP})f_L(x|\theta_L)G(\theta_L|\sigma_{\theta_L})] \quad (5.9)$$

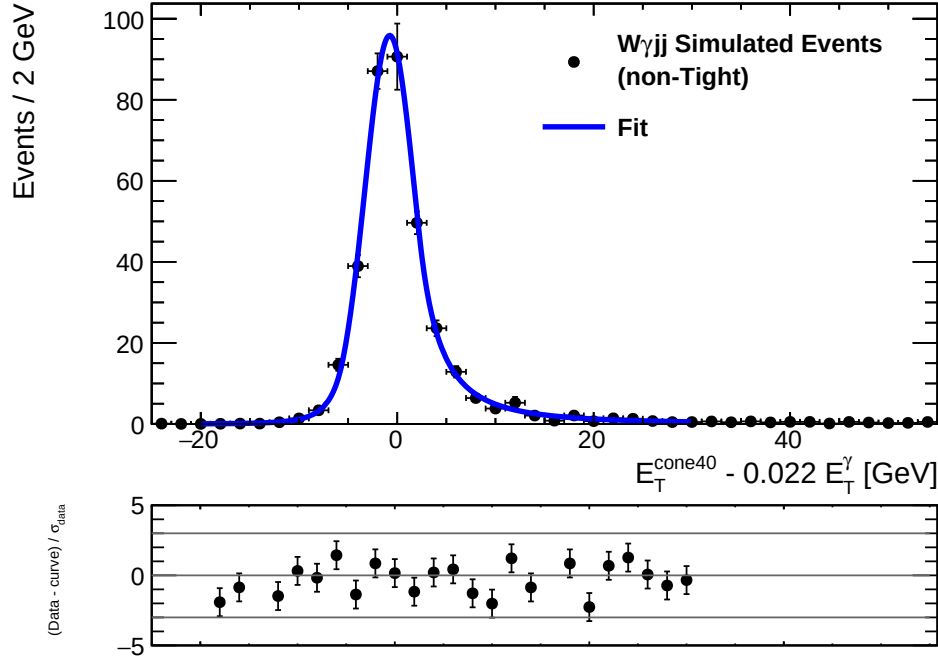


Figure 5.5: Example fit of the photon isolation energy distribution of non-tight photons obtained using simulated electroweak and QCD $W\gamma jj$ events. This fit is used to extract the prompt photon leakage template shape used to study the effect of real photon leakage into the control region used to extract the non-prompt photon shape.

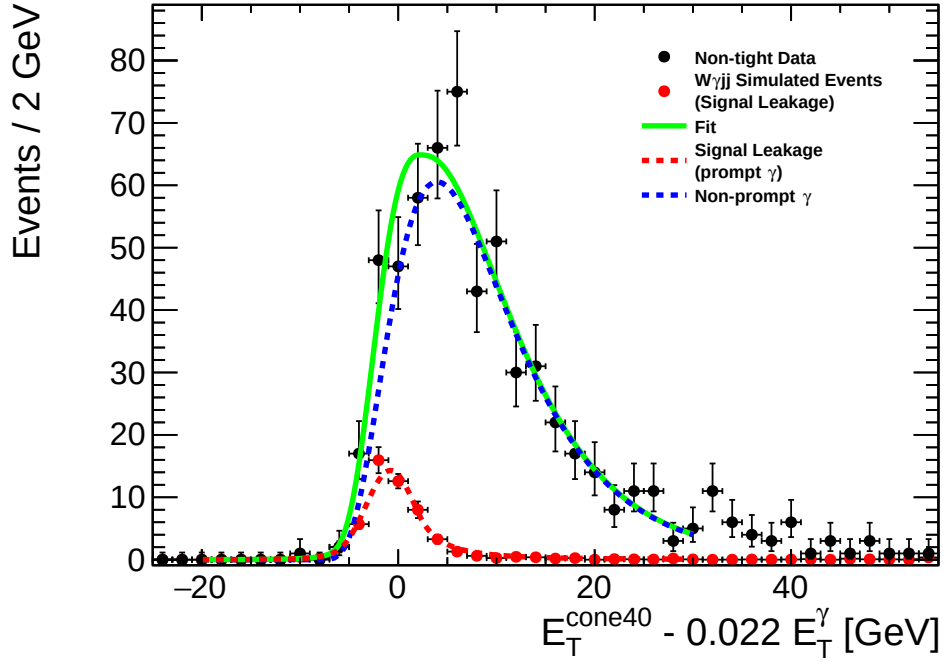


Figure 5.6: Example of a fit of the photon isolation energy distribution of events with a photon satisfying the *non-tight* photon identification criteria where an additional template for prompt photon leakage is included.

where G is a Gaussian constraint defined by the best fit parameters and their uncertainties in the fit to prompt photons that fail *tight* identification. An example of a non-prompt photon fit with the additional leakage template is shown in Figure 5.6.

The inclusion of the additional leakage template has a small effect on the estimated jet misreconstructed as photon yield as the *non-tight* template shape is not changed dramatically when the leakage template is included. The difference in jet misreconstructed as photon yield is taken as a systematic uncertainty on the estimated yield of jets reconstructed as photons.

Uncertainty due to photon identification definition: As discussed in Section 4, four different photon identification definitions can be used to define a non-prompt photon enriched control region. The so-called *Loose'2* working point allows two photon identification criteria to fail and is thus the most similar to criteria used to define prompt photon

data, i.e. the most biased. The *Loose'5* working point allows up to five selection criteria to fail and therefore has the least bias. To account for bias in the choice of *non-tight* photon definition, the template fit method is performed with each photon identification definition in addition to the nominal (the *Loose'4* working point). The maximum deviation from the nominal non-prompt photon yield is taken as the uncertainty on the method due to the choice of the *non-tight* photon identification criteria used to define the non-prompt photon enriched control region.

Uncertainty in modelling of the prompt photon template: The prompt photon template is obtained using simulated electroweak and QCD $W\gamma jj$ events. Since the photon isolation energy modelling of real photons in simulated events may be imperfect, a theory uncertainty is estimated. The factorization (μ_F) and renormalization (μ_R) scales used in the event simulation are varied to create a systematic variation in the photon isolation energy distribution. The $(\mu_F, \mu_R) = (0.5, 0.5)$ and $(\mu_F, \mu_R) = (2.0, 2.0)$ variations are used to define an up and down variation. As the fit allows the prompt photon template shape parameters to be pulled within their statistical uncertainty to fit the data, the method itself accounts for mismodelling of the prompt photon isolation energy distribution. For this reason this systematic is small.

5.1.3 Non-prompt Photon Differential Distributions

To obtain differential distributions of the non-prompt photon from jet yield, the method outlined above is performed in each bin of the observable of interest. As an example, the differential distribution of non-prompt photons from jets and the size of the statistical and systematic uncertainties on the estimate as a function of the photon transverse momentum is shown in Figure 5.7. The dominant source of uncertainty on the estimate of the background from jets misreconstructed as photons comes from the variation of the definition of the non-prompt photon enriched control region.

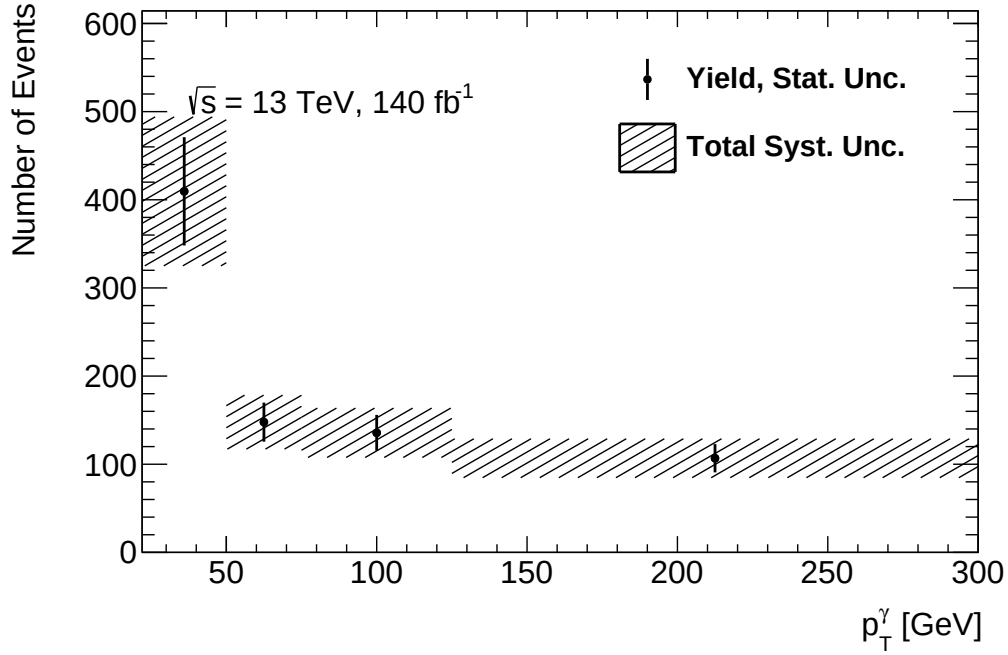


Figure 5.7: Estimated event yield and uncertainties for the non-prompt background originating from jets misreconstructed as photons as a function of photon transverse momentum.

For the differential cross section measurement described in Chapter 9 there are not sufficient statistics to perform the template fit in each bin of each phase space region used. Either there are not enough events in the non-prompt photon enriched control region to get realistic non-prompt photon template shapes, or there are not enough prompt photon events to constrain the normalization. Instead, the template fit is performed inclusively in each phase space region and the prompt photon purity in simulation is used to derive the shape of the differential distribution of jets misreconstructed as photons. The assumption of this method is that the purity dependence on the variable of interest in the differential measurement is well modelled in simulation. To check this assumption, the purity as a function of the variables being extracted is fit with a linear function in both simulation and in data, and the slopes are compared. The assumption is found to be a valid one. The details associated with this cross-check may be found in Appendix D.

5.2 Electrons Misreconstructed as Photons

Photons and electrons are both characterized by a narrow electromagnetic shower in the electromagnetic calorimeter. Electrons are identified by a track in the inner detector that points from the primary vertex to the energy cluster in the calorimeter; however, a photon that converts to an electron-positron pair at low radius in the inner detector can leave a similar signal. While the electromagnetic calorimeter has a high reconstruction efficiency, tracking inefficiencies and less than perfect calorimeter cluster to track matching can result in an electron being wrongly reconstructed as a photon. This source of background is estimated in data by measuring the probability that an electron is misreconstructed as a photon and applying this probability to a selection of events resembling that of the signal selection, but with the reconstructed photon replaced by an electron.

5.2.1 The Tag and Probe Method

To aid the reader, a schematic diagram illustrating the calculation of the probability that an electron is misreconstructed as a photon as a function of the reconstructed object's transverse momentum and pseudorapidity is shown in Figure 5.8.

Assuming a pure sample of true electrons in data, the fraction of electrons misreconstructed as photons can be defined as the ratio of the probability to misreconstruct a true electron as a photon, $P(e \rightarrow \gamma)$, to the probability to reconstruct a true electron as an electron, $P(e \rightarrow e)$,

$$F_{e \rightarrow \gamma} = \frac{P(e \rightarrow \gamma)\epsilon_\gamma}{P(e \rightarrow e)\epsilon_e} = \frac{\left(\frac{N_\gamma^{reco}}{N_e^{truth}}\right)\epsilon_\gamma}{\left(\frac{N_e^{reco}}{N_e^{truth}}\right)\epsilon_e} = \frac{N_\gamma^{reco}\epsilon_\gamma}{N_e^{reco}\epsilon_e} \quad (5.10)$$

where N_γ^{reco} and N_e^{reco} are the numbers of reconstructed photons and electrons, respectively, N_e^{truth} is the number of true electrons, and ϵ_γ and ϵ_e are the selection efficiencies for photons and electrons respectively. Written in this form, this fraction can be estimated purely in data.

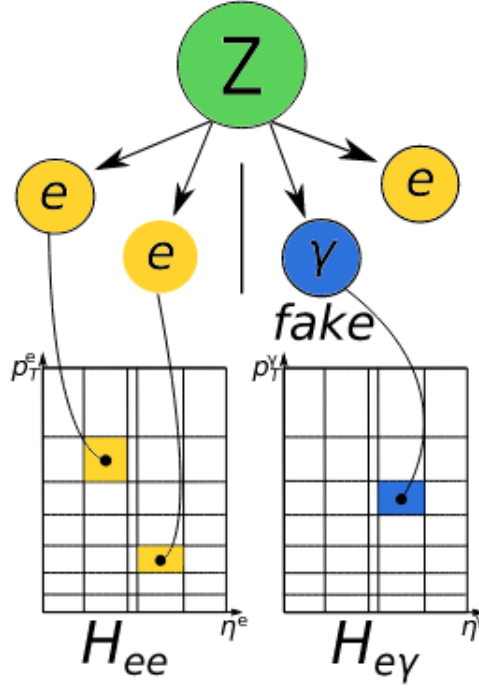


Figure 5.8: A schematic diagram depicting the calculation of the probability that an electron is misreconstructed as a photon. A pure sample of dielectron events is chosen by selecting events in data with a dielectron or electron and photon mass consistent with the objects coming from a Z boson. The number of events with one reconstructed photon and one reconstructed electron are divided by the number of events with two reconstructed electrons to calculate the probability that an electron is misreconstructed as a photon. This is done in bins of transverse momentum and pseudorapidity.

To do this estimate in data, two control regions are needed: one region that is rich in true electrons reconstructed as electrons, and another region enriched in true electrons misreconstructed as photons. These two control regions are identified by selecting events with a reconstructed invariant mass, M_{ee} or $M_{e\gamma}$, within 10 GeV of the Z boson mass, i.e. $|M_Z - M_{ee}| < 10 \text{ GeV}$ or $|M_Z - M_{e\gamma}| < 10 \text{ GeV}$. This results in a relatively pure sample of true dielectron events where a reconstructed photon is very likely to originate from a true electron. A “Tag and Probe” method is used to select Z boson events with two reconstructed electrons or one reconstructed electron with one reconstructed photon. The *Tag* is a reconstructed electron that passes the analysis requirements of electrons.

The *Probe* electron is an electron with opposite charge from the *Tag*, and since it is used to calculate the probability that an electron is misreconstructed as a photon, the same kinematic cuts as those applied to the photon are applied to the *Probe* electron. The object selection for the *Tag* and *Probe* objects are summarized in Table 5.1.

Tag Electron	Probe electron	Probe photon
$p_T > 30 \text{ GeV}$ $ \eta < 2.5$ Isolated <i>Tight</i> $ z_0 \times \sin \theta < 5 \text{ mm}$ $ d_0/\sigma_{d_0} < 3$ $m_T^{W(e\nu)} > 30 \text{ GeV}$ Pass Trigger	$p_T > 22 \text{ GeV}$ $ \eta < 2.37$ Isolated <i>Tight Electron</i> $ z_0 \times \sin \theta < 5 \text{ mm}$ $ d_0/\sigma_{d_0} < 3$ Opposite charge to <i>Tag</i> electron	$p_T > 22 \text{ GeV}$ $ \eta < 2.37$ Isolated <i>Tight Photon</i>

Table 5.1: Selection criteria of *Tag* electron and *Probe* objects used for calculating the probability that an electron is misreconstructed as a photon.

To determine the number of true electrons reconstructed as either an electron or a photon, two additional points must be considered. The first point to consider is that background contamination in the two regions may lead to different signal to background ratios. The background contamination must then be corrected for. This is done by assigning a weight to each type of *Tag* and *Probe* pair. The weight is the signal fraction $\frac{S}{S+B}$, determined by an $S + B$ fit to the invariant mass of the appropriate type of *Tag* and *Probe* pair. For the Z boson signal shape, a double-sided Crystal Ball function, as described in Section 5.1, is used while an exponentially decaying function is used to model the background. Both shape parameters and normalizations are floated in the fits. The fits for the *Tag* electron-*Probe* electron pair and the *Tag* electron-*Probe* photon pair are shown in Appendix E.

The second point to consider is that in the case where the *Probe* electron also passes the *Tag* requirement, either electron can contribute to the denominator in the calculation of the fraction $F_{e \rightarrow \gamma}$. To account for this, the number of electrons in the denominator should be filled twice in the case where there is a *Probe* electron that passes the *Tag* requirements

and once in the case that the *Probe* electron does not pass the *Tag* requirements. In other words, the fraction of true electrons misreconstructed as photons can be calculated as

$$F_{e \rightarrow \gamma} = \frac{N_{e_{tag}\gamma_{probe}}}{2N_{e_{tag}e_{tag}} + N_{e_{tag}e_{probe}}} \quad (5.11)$$

where $N_{e_{tag}\gamma_{tag}}$ is the number of events with one electron satisfying the *Tag* criteria and one photon satisfying the photon *Probe* criteria, $N_{e_{tag}e_{tag}}$ is the number of events in which both electrons satisfy the *Tag* criteria, and $N_{e_{tag}e_{probe}}$ is the number of events where one electron satisfies the *Tag* criteria and one electron satisfies the *Probe* criteria.

Taking these two considerations into account, the resulting probability that an electron is misreconstructed as a photon in bins of the probe object pseudorapidity and transverse momentum is shown in Figure 5.9.

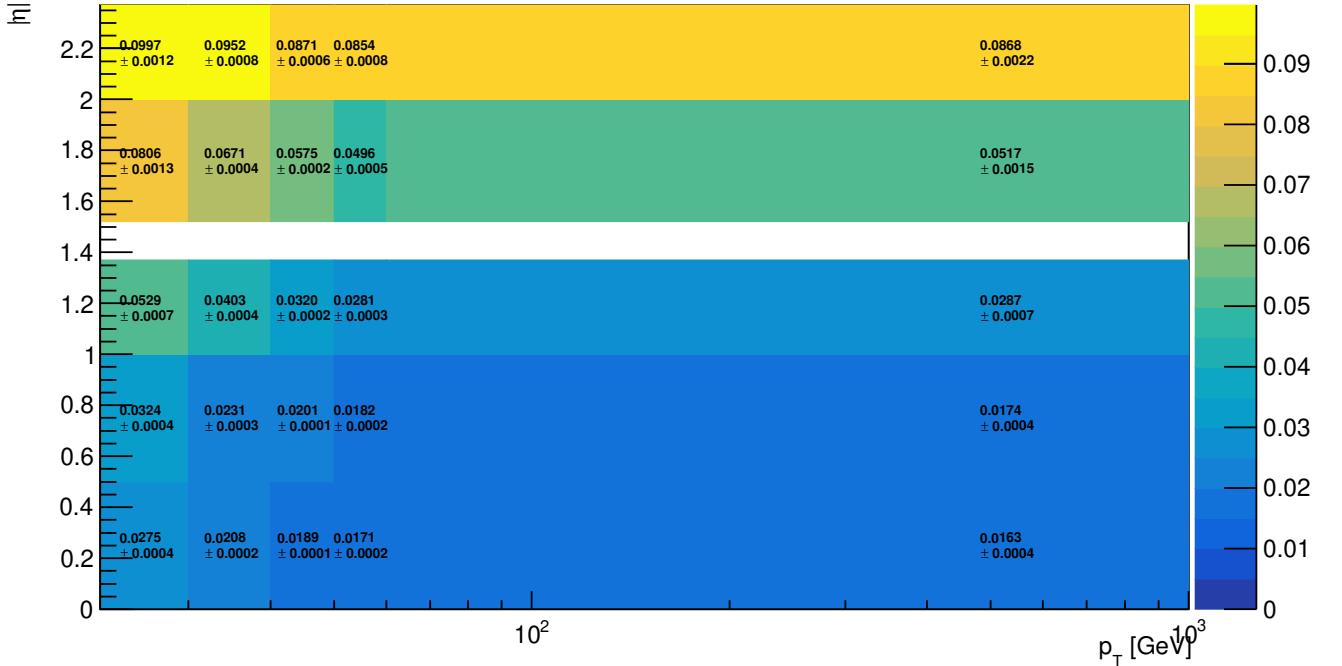


Figure 5.9: Fraction of true electrons misreconstructed as photons, $F_{e \rightarrow \gamma}$, in bins of *Probe* object pseudorapidity and transverse momentum with statistical uncertainties.

The background coming from true electrons misreconstructed as photons is estimated by applying the fraction $F_{e \rightarrow \gamma}$ to data events in a signal-like region. The signal-like region is defined by the analysis selection described in Chapter 4 but with the photon selection criteria replaced with selections for an electron satisfying the probe electron criteria described above. The background coming from true electrons misreconstructed as photons is estimated separately in the electron and muon channels in the analysis. The result of the application of the fraction $F_{e \rightarrow \gamma}$ to these selected events in data is the distribution of events where an electron is misreconstructed as a photon in the analysis region.

5.2.2 Uncertainties on the Estimate of the Background from Electrons Misreconstructed as Photons

Systematic uncertainties are included on the data driven estimate of electrons misreconstructed as photons. The systematic effects considered are the following:

- The fitting range in the fit to correct for background contamination is varied by ± 5 GeV.
- The definition of the region enriched in dielectron or electron and photon misreconstructed as electron events is varied by ± 5 GeV around the Z boson mass.
- The size of the binning of the fraction $F_{e \rightarrow \gamma}$ in pseudorapidity is varied by adding or subtracting two bins.
- The size of the binning of the fraction $F_{e \rightarrow \gamma}$ in transverse momentum is varied by adding or subtracting two bins.
- The choice of fitting functions for the signal fraction weights is changed to a Gaussian with exponential tails for the signal modelling and a 4th order Bernstein polynomial [83] for the background modelling.

5.2.3 Electrons Misreconstructed as Photon Differential Distributions

The differential distribution of electrons misreconstructed as photons and the size of the statistical and systematic uncertainties on the estimate as a function of the photon transverse momentum is shown in Figure 5.10. The systematic uncertainties are small everywhere except for the first bin where changing the fitting range has a large effect due to high background contamination at low lepton and photon transverse momentum.

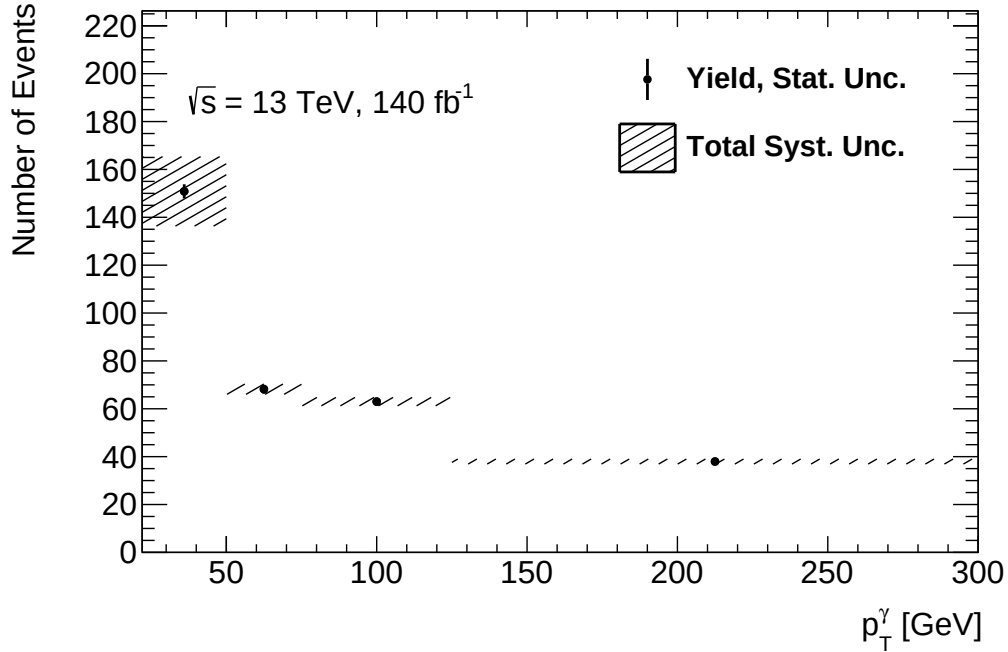


Figure 5.10: Electron misreconstructed as photon yield and uncertainties as a function of photon transverse momentum.

5.3 Jets Misreconstructed as Leptons

An additional source of non-prompt background arises from jets misreconstructed as an electron or muon (lepton). This can occur, for example, when a true electron or muon is produced within a jet as part of the decay of particles created from the hadronic cascade and its interaction with the detector material. This background contribution is estimated using a data driven approach. The number of events from this source of background,

$N_{j \rightarrow \ell}$, is estimated as

$$N_{j \rightarrow \ell} = (N_{Data}^{non-tight} - N_{prompt}^{non-tight}) \times \frac{N^{tight}}{N^{non-tight}} = N_{Data, non-prompt}^{non-tight} \times F_{j \rightarrow \ell} \quad (5.12)$$

where $N_{Data}^{non-tight}$ represents the number of events in data satisfying the signal selection, except for the lepton failing the *tight* identification criteria and $N_{prompt}^{non-tight}$ represents simulated events with prompt leptons that satisfy the signal selection criteria but the lepton failing the *tight* identification criteria. Subtracting this prompt contribution from data with a *non-tight* lepton gives an estimate of the number of events in data with a non-prompt lepton that fails the *tight* identification criteria, $N_{Data, non-prompt}^{non-tight}$. The quantity $F_{j \rightarrow \ell} = \frac{N^{tight}}{N^{non-tight}}$ represents the number of jets containing a *non-prompt* lepton that satisfies the *tight* lepton identification selection divided by the number of jets containing a *non-prompt* lepton that fails the *tight* lepton identification criteria.

The factor $F_{j \rightarrow \ell}$ is calculated using data in a region that is enriched in jets. This region is defined by selecting data events that contain one reconstructed lepton and one jet. These requirements primarily select events originating from dijet process or processes where a W or Z boson is produced in association with jets. In order to suppress the contribution from the W+jet events in this data sample, events are further required to have missing transverse energy $E_T^{miss} < 20$ GeV and a transverse invariant mass, calculated using the lepton transverse momentum and missing transverse momentum vector, less than 20 GeV. To increase the purity of the data samples containing jets, the simulated contributions from physics processes other than dijet processes is subtracted from these data samples. The resulting data yields are used to calculate the factor $F_{j \rightarrow \ell}$, introduced in Equation 5.12, as a function of lepton transverse momentum and pseudorapidity. The calculated values of $F_{j \rightarrow \ell}$ are presented in Appendix F.

In order to calculate differential yields of jets misreconstructed as leptons, $F_{j \rightarrow \ell}$ is applied in a signal-like region with the lepton identification cut inverted. Prompt backgrounds are subtracted from the signal-like region where $F_{j \rightarrow \ell}$ is applied.

5.3.1 Uncertainties on the Background from Jets Misreconstructed as Leptons

The following sources of systematic uncertainty are considered in the estimate of the event yield from jets misreconstructed as leptons:

- The factor $F_{j \rightarrow \ell}$ is calculated using dijet simulated events and using $\gamma + \text{jet}$ simulated events. Since these event samples have a different composition of the origin of non-prompt leptons, the difference in values of $F_{j \rightarrow \ell}$ is applied to the data driven estimate as a systematic.
- The background subtraction in the dijet enriched region and in the non-prompt lepton enriched region where $F_{j \rightarrow \ell}$ applied is done where the $W + \text{jet}$ background scale variations are applied and where an alternate QCD $W\gamma$ simulation is used.
- The choice of the binning in transverse momentum and pseudorapidity used to calculate $F_{j \rightarrow \ell}$ is varied.
- The value of the selection cuts on E_T^{miss} and M_T^W used to define the data samples enriched in dijets are varied by ± 5 GeV.
- An alternate QCD $W\gamma jj$ sample is subtracted from the region where $F_{j \rightarrow \ell}$ is applied.

5.3.2 Jets Misreconstructed as Leptons Differential Distributions

The differential distribution of jets misreconstructed as leptons and the size of the statistical and systematic uncertainties on the estimate as a function of the photon transverse momentum is shown in Figure 5.11. The estimate is done independently for jets misreconstructed as electrons and for jets misreconstructed as muons. In Figure 5.11 the resulting yields and their uncertainties from both the electron and muon channel are combined.

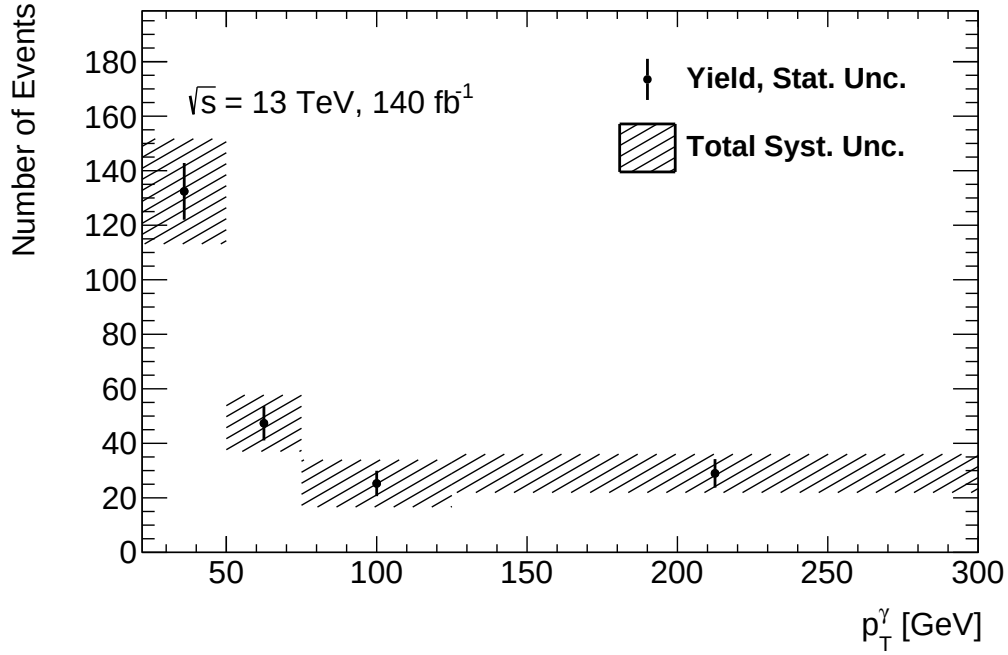


Figure 5.11: Jets misreconstructed as lepton yields and uncertainties as a function of photon transverse momentum.

5.4 Photons from Other Collisions

The final non-prompt background considered is the background arising from so-called “pileup” photons. Unlike the lepton selection, as part of the photon reconstruction there is no explicit requirement on the point of origin of the photon as this is not well measured in most events. It is then a possibility that a photon from a different proton-proton interaction in the same LHC bunch crossing combines with a W +jet event from the primary vertex resulting in a signal-like event in the final state.

The assumption of the data driven method used to estimate the pileup photon background is that the distribution of the difference between the photon and primary vertex z position at $x = y = 0$ ($\Delta z(\gamma, PV)$) will have a different shape for events where the photon is from pileup and events where the photon is from the hard scatter. To minimize the impact of the photon pointing resolution, only converted photons with hits in the silicon pixel detector are used as these have a better z position resolution.

The $\Delta z(\gamma, PV)$ distribution in data is shown in Figure 5.12 where good resolution photons have been selected. The $\Delta z(\gamma, PV)$ distribution is peaked at 0 and has a width of ap-

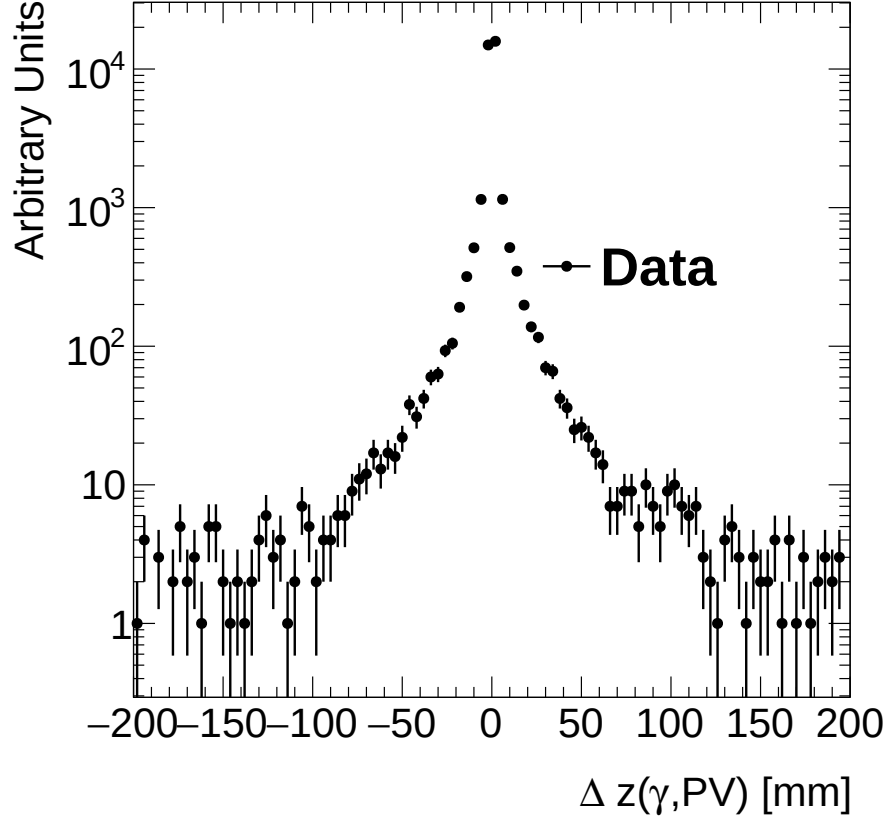


Figure 5.12: The distribution of $\Delta z(\gamma, PV)$ in data with the baseline selection applied using only photons with a good vertex spatial resolution.

proximately 35 mm, which is the longitudinal spread of proton bunches in the LHC. This reflects the fact that good resolution photons have been selected and most events have a reconstructed photon that originates from the hard scatter. However for pileup photons, the $\Delta z(\gamma, PV)$ distribution is the difference of two uncorrelated Gaussian distributions of width 35 mm and should thus be a Gaussian distribution of width $35 \text{ mm} \times \sqrt{2} = 50 \text{ mm}$. The tails of the $\Delta z(\gamma, PV)$ can be used to calculate the fraction of events in data where the photon is from pileup. The expected number of events from pileup in the tails of the $\Delta z(\gamma, PV)$ distribution is $N_{data}^{|\Delta z| > 50 \text{ mm}} - N_{simulation}^{|\Delta z| > 50 \text{ mm}}$ where $N_{data}^{|\Delta z| > 50 \text{ mm}}$ is the number of events in data with $z(\gamma, PV) > 50 \text{ mm}$ and $N_{simulation}^{|\Delta z| > 50 \text{ mm}}$ is the number of simulated QCD

$W\gamma jj$ events with $z(\gamma, PV) > 50$ mm. Since $\Delta z(\gamma, PV)$ is Gaussian distributed for pileup photons, 32% of the pileup photon distribution should have $\Delta z(\gamma, PV) > 50$ mm. Thus the fraction of pileup photons in data can be calculated as

$$f_{PU} = \frac{N_{data}^{|\Delta z|>50mm} - C \times N_{simulation}^{|\Delta z|>50mm}}{N_{data} * 0.32} \quad (5.13)$$

where C is a normalization factor derived by comparing the QCD $W\gamma jj$ simulation to data. The $W\gamma jj$ simulation is chosen for its relatively large statistics. To validate that a normalization factor is all that is needed, the shape of the simulation is compared with data in Figure 5.13. No mismodelling of the shape of the $\Delta z(\gamma, PV)$ distribution is observed. With the baseline selection, the normalization factor is found to be $C = 1.84 \pm 0.06$.

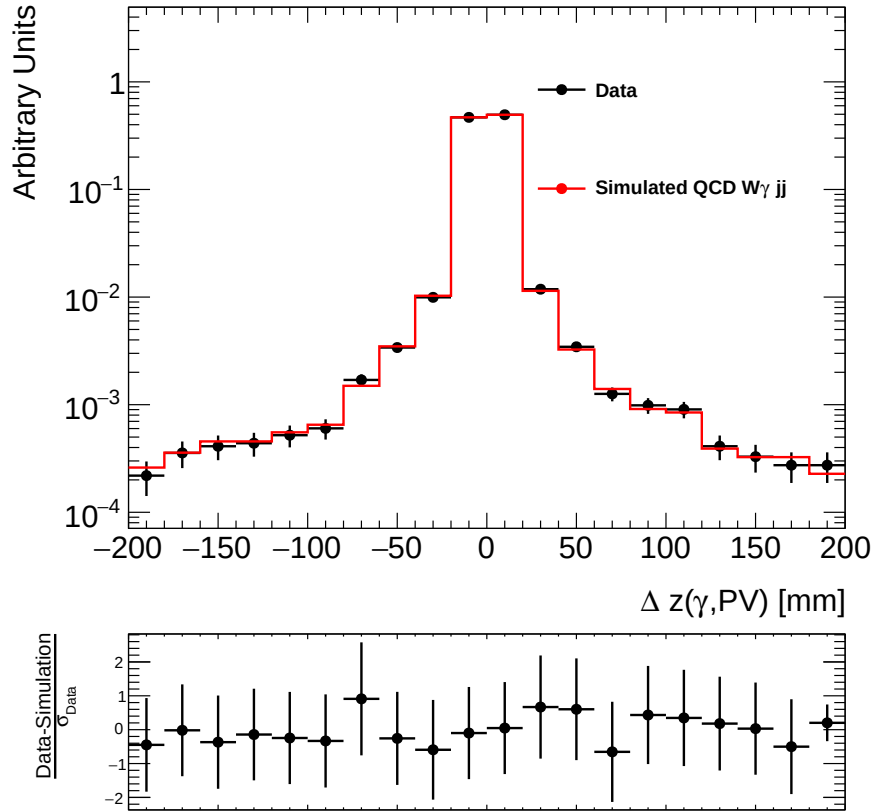


Figure 5.13: Comparison between the distribution of $\Delta z(\gamma, PV)$ in data and simulated QCD $W\gamma jj$ events.

Using the $\Delta z(\gamma, PV)$ distribution shown in Figure 5.12 and the value of the correction factor C , the fraction of pileup photons is calculated using Equation 5.13 to be

$$f_{PU} = 0.019 \pm 0.005 \quad (5.14)$$

The pileup background contains a component where a pileup jet is misreconstructed as a photon. This component is already accounted for in the data driven background estimate of jets misreconstructed as photons (Section 5.1). In order to avoid double counting, the data driven estimate of jets misreconstructed as photons is performed on a sample of data containing a single reconstructed photon. The photon purity of this sample is assumed to be the same as in pileup events. The photons in this sample are required to pass the same baseline event selection and are required to convert in the silicon pixel tracker. The photon purity in this sample is found to be $56.7 \pm 0.1\%$ using the data driven method described in Appendix B.

Applying the photon purity to the pileup fraction found above, the fraction of pileup photons in data is found to be 0.011 ± 0.003 . This fraction is applied to data with non-prompt backgrounds subtracted to derive the differential distribution of the background arising from pileup photons. The background arising from pileup photons as a function of photon transverse momentum is shown in Figure 5.14.

5.5 Summary

This section summarizes the estimated contribution of various types of background processes to the data sample selected using the kinematic criteria described in Chapter 4. Yields from prompt background sources are estimated using simulated events. As described in Sections 5.1 to 5.4 yields from non-prompt background sources are estimated using data driven methods.

The estimated yield of signal events and events from each source of background satisfying the baseline selection criteria defined in Section 4 as well as the number of selected

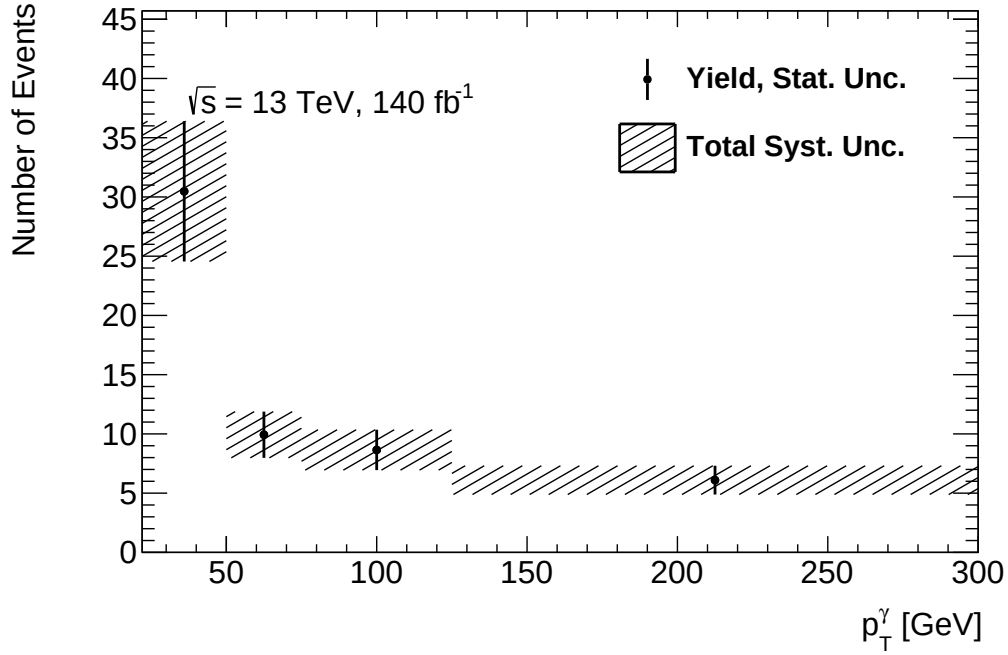


Figure 5.14: Estimated yield and uncertainty of the background where the photon comes from another collision as a function of photon transverse momentum.

data events are shown in Table 5.2. The yields as a function of the photon transverse momentum are shown in Figure 5.15. The uncertainty shown includes the systematic uncertainties on the data driven estimates described in this chapter, the statistical uncertainty on the simulated events, and the uncertainties described in Chapter 7.

From Table 5.2 it is observed that QCD $W\gamma jj$ is the largest background and jets misreconstructed as photons constitute the largest non-prompt background. The non-prompt photon from jets estimate has the largest contribution to the total uncertainty of the non-prompt backgrounds. Signal events are predicted to make up approximately 10% of the events satisfying the baseline selection. As shown in Figure 5.15 while statistically compatible with the observed data distribution, the total number of events expected nevertheless displays a systematic trend. As described in Chapters 7, 8 and 9, this small difference in the shape of the distribution between the observed and expected yields is taken into account by allowing the expected yield from the dominant QCD $W\gamma jj$ background to vary within uncertainties in the fit used to extract the fiducial and differential cross sections.

	Process	Yield
Signal	EW $W\gamma jj$	618 ± 183
Prompt Background	QCD $W\gamma jj$	3357 ± 1569
	Top Bkg.	283 ± 52
	$Z\gamma jj$	283 ± 119
Non-prompt Background	$j \rightarrow \gamma$	778 ± 137
	$e \rightarrow \gamma$	320 ± 11
	$j \rightarrow l$	234 ± 25
	pileup	57 ± 1
Total		5930 ± 1591
Data		6367

Table 5.2: Estimated event yields of signal, prompt backgrounds and non-prompt backgrounds satisfying the baseline selection compared with the observed yield in data.

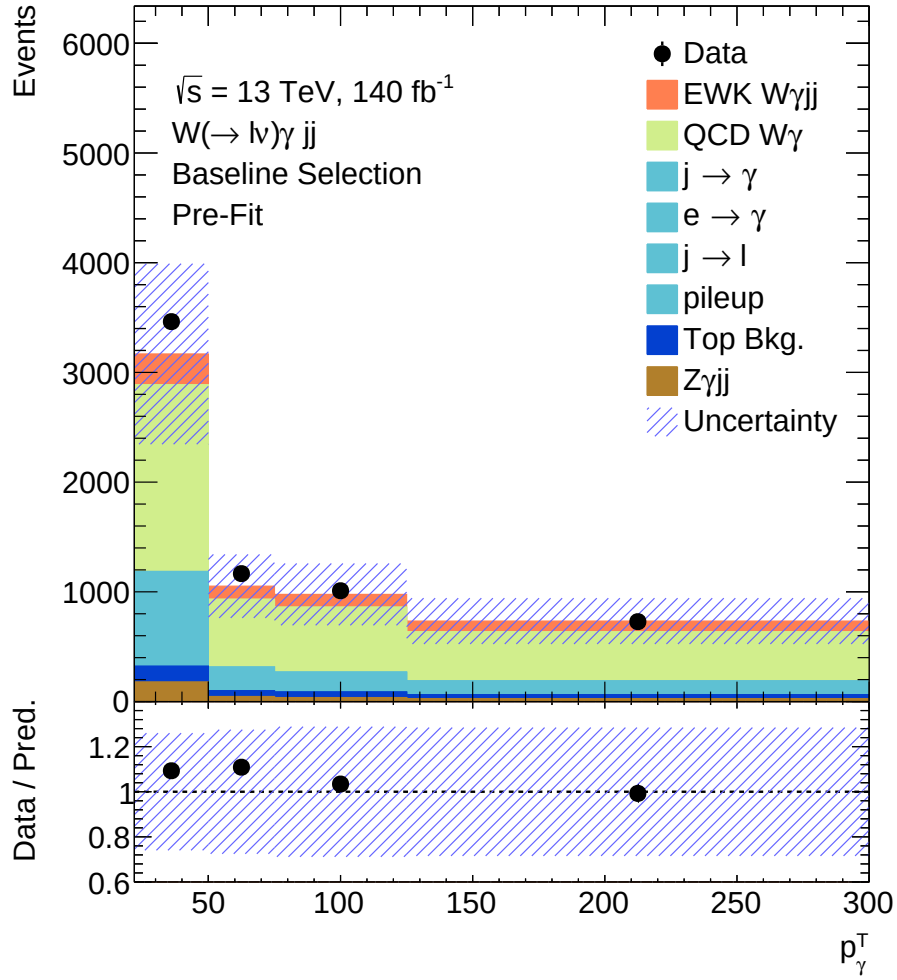


Figure 5.15: Estimated yields of signal and background events satisfying the baseline selection. The uncertainties include statistical uncertainty on the simulated events, uncertainties on the data driven backgrounds and uncertainties described in Chapter 7. The statistical uncertainty on the number of data events measured is not visible because it is smaller than the size of the data markers.

Chapter 6

Neural Networks for Signal Classification

The challenge of identifying a small signal in the presence of a large irreducible background makes this analysis an ideal ground to explore the use of machine learning techniques. In this chapter, the development of a neural network for signal identification is presented. Neural network training, performance and validation are discussed and the neural network used to discriminate the signal from the dominant sources of prompt background is presented.

6.1 Introduction to Neural Networks

Neural networks are universal approximators, i.e. they are capable of modelling any continuous function to arbitrary accuracy given sufficient complexity. In general a neural network is a function f that maps a vector of input features $\mathbf{x}$ in $\mathcal{R}^n$ to a vector of outputs $\mathbf{y}$ in $\mathcal{R}^m$ through the use of layered linear transformations along with a non-linear activation function $\mathcal{R} \rightarrow \mathcal{R}$. To help guide the reader, a schematic diagram of a simple neural network with one so-called hidden layer is shown in Figure 6.1

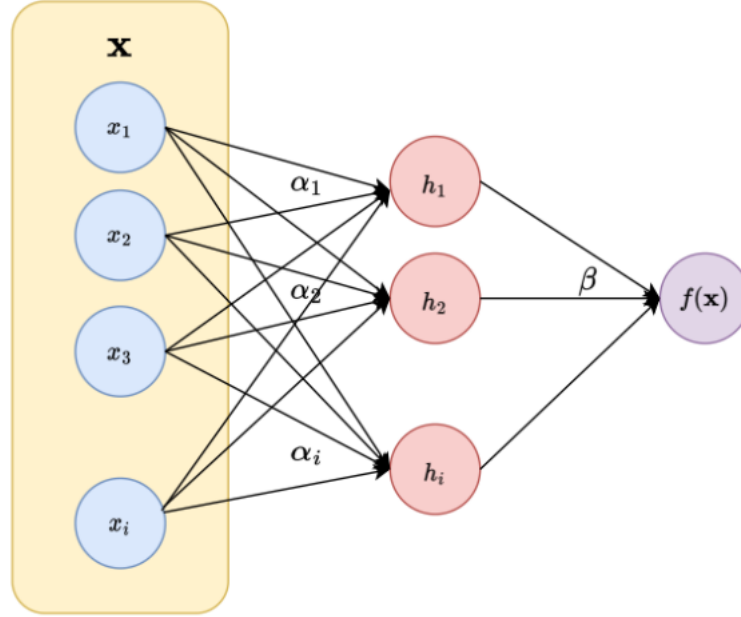


Figure 6.1: Schematic diagram of a fully connected neural network with one hidden layer.

The operation of the so-called hidden layer can be described by

$$\mathbf{h}_i = \sigma(\alpha_i^T \mathbf{x} + \alpha_0) \quad (6.1)$$

where α_i is a vector of weights and α_0 is a bias term. The weights and bias terms are the free parameters of the neural network that are fit during network *training*. The hidden layer contains the transformation σ , which is a non-linear activation function. This hidden layer is then mapped to an output $f(\mathbf{x})$ as

$$f(\mathbf{x}) = \phi_{out}(\beta^T \mathbf{h} + \beta_0) \quad (6.2)$$

where ϕ_{out} is another non-linear activation function, and β and β_0 are another set of weights and a bias term, respectively, that are fit during training. This can be generalized to an arbitrary number of hidden layers and an arbitrary number of h_i in each layer, which are henceforth referred to as nodes. It is not necessary for all input variables to be

fed to each hidden node, or each hidden layer node to be fed to each of the next hidden layer nodes. In other words, weights may be fixed to zero. A neural network with a floating weight for every connection between each node or input layer to the next is called *densely connected* or *fully connected*.

6.1.1 Training

Like any procedure to fit a function to data, neural networks are trained by minimizing some metric or loss function. For classification tasks this commonly takes the form of the binary cross entropy (BCE) loss function [84]

$$\mathcal{L}_{BCE} = \mathbb{E}_{\mathbf{x} \in \mathbf{X}} [-y \log f(\mathbf{x}|\theta) - (1 - y) \log(1 - f(\mathbf{x}|\theta))] \quad (6.3)$$

where $f(\mathbf{x}|\theta)$ is the neural network output given parameters θ and input $\mathbf{x}$ and y is the true value of the data. The error is averaged over a set of inputs $\mathbf{x} \in \mathbf{X}$ which is denoted by $\mathbb{E}_{\mathbf{x} \in \mathbf{X}}$. In the case of classification, the neural network output is restricted to have a value $0 < f(\mathbf{x}) < 1$ by the choice of non-linear activation function.

Typically, to minimize the value of the loss function, the gradient of the loss function is computed with respect to each parameter, and then a step is taken in the parameter space towards the minimum in a *gradient descent* procedure. In neural network training, the gradient is derived using the chain rule and this is used for weight updates. This procedure is called *backpropagation* [85].

To simply illustrate backpropagation, the cross-entropy loss is replaced with the sum of squared loss

$$L = \mathbb{E}_{\mathbf{x} \in \mathbf{X}} [(y - f(\mathbf{x}))^2] \quad (6.4)$$

though the same procedure can be applied to any differentiable loss function. Using the chain rule, the derivatives of the loss with respect to a unit b_m of the vector β and a unit

α_{ml} of one of the vectors α_l are

$$\begin{aligned}\frac{\partial L}{\partial \beta_m} &= -2(y - f(x))\phi'(\beta^T \mathbf{h})h_m \\ \frac{\partial L}{\partial \alpha_{ml}} &= -2(y - f(\mathbf{x}))\phi'(\beta^T \mathbf{h})\beta_m \sigma'(\alpha_m^T \mathbf{x})x_l.\end{aligned}\tag{6.5}$$

Using these derivatives, the weights of the neural network may be updated as

$$\begin{aligned}\beta_m^{(r+1)} &= \beta_m^{(r)} - \eta \frac{\partial L}{\partial \beta_m^{(r)}} \\ \alpha_{ml}^{(r+1)} &= \alpha_{ml}^{(r)} - \eta \frac{\partial L}{\partial \alpha_{ml}^{(r)}}\end{aligned}\tag{6.6}$$

where η is a parameter called the *learning rate* that controls the size of each weight update.

The derivatives above may be written as

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \beta_m} &= \delta z_m \\ \frac{\partial \mathcal{L}}{\partial \alpha_{ml}} &= s_{ml}x_{il}\end{aligned}\tag{6.7}$$

where δ and s_{ml} are called the errors of the neural network at the output and hidden layers, respectively. The output at the hidden layer may also be expressed as

$$s_{ml} = \sigma'(\alpha_m^T \mathbf{x})\beta_m \delta\tag{6.8}$$

The backpropagation algorithm is a two step procedure. First $f(\mathbf{x})$ is computed using Equations 6.1 and 6.2. Next the error at the output node, δ , is computed and used to calculate the errors s_{ml} using Equation 6.8. The errors are then used for the weight updates described by Equations 6.6 and 6.7. This can be generalized to an arbitrary number of layers.

6.1.2 Testing and Validation

Neural networks for classification and regression which have been described here are trained on labelled samples of data $(\mathbf{x}, \mathbf{y})$ termed *training data* and then applied to data $\mathbf{x}'$ to generate labels $\mathbf{y}'$. The ability to generate large labelled simulation samples to compare with large samples of data makes high energy physics analyses like the one presented in this thesis an ideal ground to use neural networks for classification and regression. The use of neural networks however relies on the assumption that randomly sampled training data is representative of the data that the neural network is applied on. Given the neural network power as a universal approximator, it is prone to learning attributes of training data that are not representative of the data it is to be applied on. This is termed *overfitting*. To judge if a network has been overfit, labelled data must be split into training and testing sets, and only the training set used to calculate gradients and update weights. With each training iteration the loss function can be calculated on both the training set and the testing set. A neural network overfits when the value of the loss function evaluated on the testing set begins to increase as the value of the loss function evaluated on the training set continues to decrease.

6.1.3 Hyperparameter Tuning

While weights and bias vectors are trained in neural network training, the choice of the number of layers, l , in the neural network, number of nodes per layer, i , activation function of intermediate layers, σ , and learning rate, η , are all chosen by the user. These parameters defining the structure of the neural network that are chosen by the user are called hyperparameters.

Adaptive learning rate algorithms [86] are often used such that the learning rate is updated with each learning step based on the Hessian of the loss function. Typically, the other remaining hyperparameters are chosen by a grid search, i.e. all combinations are tried for a limited number of iterations, and the best performing combination of hyper-

parameters is used in the final neural network training. More powerful hyperparameter training algorithms are a subject of much research and the interested reader is referred to [87] and [88] for example. A grid search is used in this thesis and other hyperparameter tuning procedures are considered to be outside the scope of this work.

6.2 Neural Network Design

In this section the variables used in neural network training, the neural network architecture, and the training itself is described. Throughout this section, signal refers to electroweak $W\gamma jj$ production while background refers to other processes that produce the same final state in the detector listed in Table 5.2 as “prompt background”.

6.2.1 Training Variables

A neural network is trained to discriminate signal from background using the 13 kinematic variables listed and explained below:

- p_T^γ : The photon transverse momentum.
- p_T^l : The lepton transverse momentum.
- $p_T^{l\gamma}$: The transverse momentum of the lepton and photon system.
- M_T^W : The transverse mass of the W boson.
- $|\Delta\phi(l\gamma, jj)|$: The separation in azimuthal angle between the lepton-photon and dijet systems.
- $\Delta R(l\gamma, jj)$: The angular separation between the lepton-photon and dijet systems.
- ξ_γ : The centrality of the photon defined by $\xi_\gamma = \left| \frac{y_\gamma - (y_{j_0} + y_{j_1})}{\Delta y(j_0, j_1)} \right|$ where y represents the rapidity of the reconstructed object.
- $\xi_{l\gamma}$: The centrality of the lepton-photon system defined by $\xi_{l\gamma} = \left| \frac{y_{l\gamma} - (y_{j_0} + y_{j_1})}{\Delta y(j_0, j_1)} \right|$.

- $\Delta R(l, \gamma)$: The angular separation between the lepton and the photon.
- $p_T^{j_0}$: The transverse momentum of the leading jet.
- η_{j_0} : The pseudorapidity of the leading jet.
- $\Delta R(\gamma, j_0)$: The angular separation between the photon and the leading jet.
- $|\Delta\eta(j_0, j_1)|$: The separation in pseudorapidity of the two leading jets.

These variables are chosen based on their potential to distinguish signal from background alone, or for the potential that correlations between the variables are different between signal and background. The comparison of the shape of these variables between signal and background is shown in Figure 6.2. Several variables have a clear discriminatory power. Some variables exhibit the same shape in signal and in background, however the neural network may exploit subtle differences in correlations between variables in discriminating the signal from the background.

6.2.2 Neural Network Training

In this section the neural network architecture used in this analysis is described.

The neural network consists of a layer that normalizes the input variables, followed by three fully connected hidden layers of 512 nodes each. The model is trained with the Adam learning rate optimizer [86], which adjusts the learning rate based on the Hessian of the loss function. The initial learning rate chosen is 3×10^{-5} . The number of hidden layers, number of nodes and learning rate were determined using a grid search, i.e. the neural network training is repeated with different combinations of number of hidden layers, nodes, and learning rate and the model architecture with the lowest value of the loss function on test data is chosen. The different sets of parameters used for the grid search are summarized in Table 6.1.

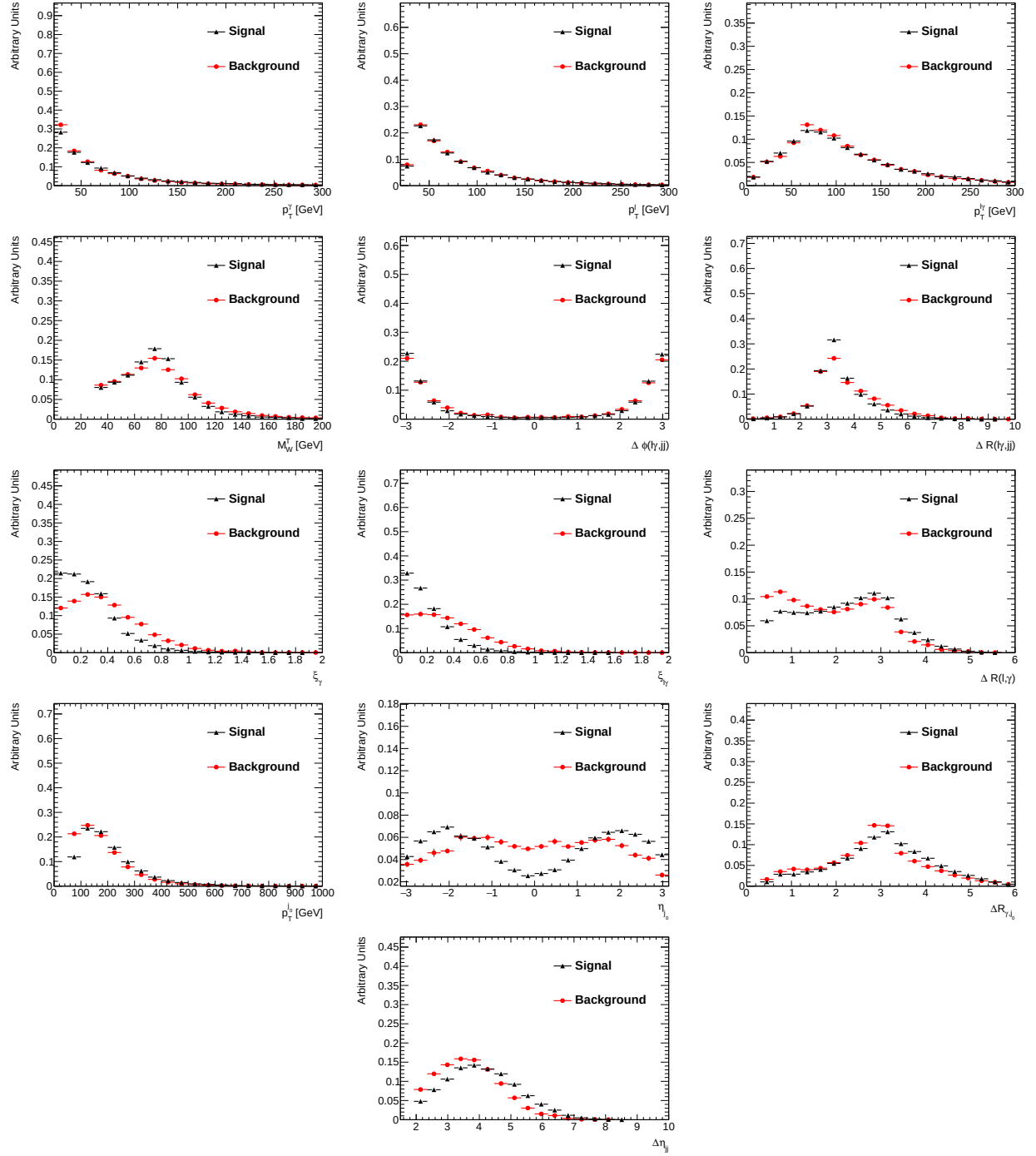


Figure 6.2: Comparison of kinematic variables used in neural network training in signal and background simulation. All distributions are normalized to one.

Hyperparameter	Values
Number of hidden layers	2, 3 ,4,5,7
Nodes per layer	64,128,256, 512 ,1024
Learning rate	1×10^{-5} , 3×10^{-5} , 1×10^{-4} , 3×10^{-4} , 1×10^{-3}

Table 6.1: Summary of the hyperparameter values that were considered in the grid search to find the model with the smallest value of the loss function. The best values are shown in boldface

Each hidden node uses a Leaky Rectified Linear Unit activation function, defined as

$$\sigma(x) = \begin{cases} 0.01x, & \text{if } x < 0 \\ x, & \text{otherwise.} \end{cases} \quad (6.9)$$

Rectified Linear Units are a common choice of activation function in hidden layers as they do not saturate at large or small input values, which leads to the so-called vanishing gradient problems discussed in [89]. The term “Leaky” refers to the function mapping to a small negative value for negative inputs, as opposed to mapping to 0 as for a standard Rectified Linear Unit. This modification is introduced to avoid that a node output becomes 0 and remains at 0 as training proceeds, as discussed in [90]. The output layer uses a sigmoid activation function defined as

$$\phi(x) = \frac{1}{1 + e^{-x}} \quad (6.10)$$

such that the output of the neural network has a range between 0 and 1.

To take advantage of the full simulated dataset, two independent neural networks are trained on simulated events with even and odd numbers, respectively. This conveniently partitions the simulation data into training samples and validation samples. The neural network is fitted to the training sample and is evaluated on validation sample to ensure there is no overfitting. To avoid overfitting, *early stopping* is used where the training is

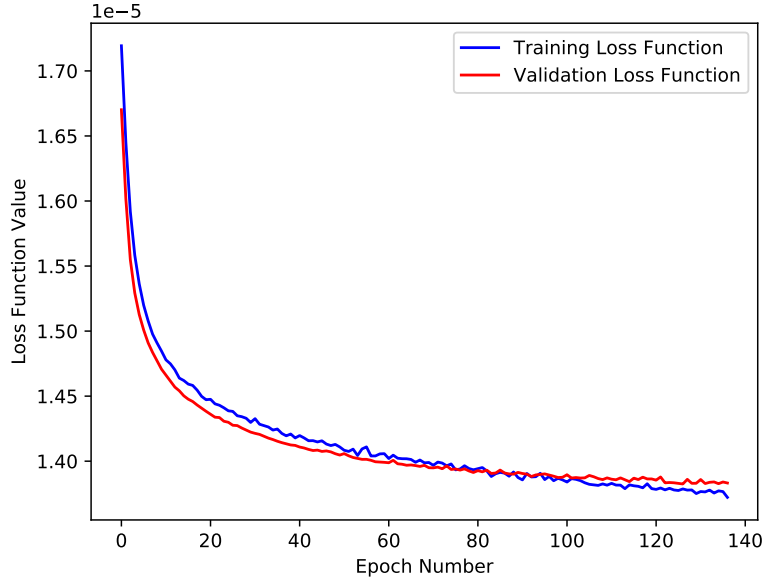


Figure 6.3: Loss function curves for the neural network trained on odd simulated event numbers and tested on even simulated event numbers. Training stops when the validation loss function value does not decrease over 10 training epochs.

stopped when the value of the loss function on validation data does not decrease over 10 iterations through the training data, referred to as training *epochs*. When deploying the networks on real data, which is not used in neural network training, the output values of the two networks are averaged. The combination of the two networks is validated in Section 6.2.3. Separate neural networks are not trained for the electron and muon channel because the shape of the training variables is for the most part the same between channels, leaving little to be gained by training a separate neural network on each channel. For completeness, a channel comparison is shown in Appendix G.

The training and validation loss function value as a function of training epoch of the neural network trained using odd simulated event numbers and tested using even simulated event numbers is shown in Figure 6.3. The smooth loss curve indicates that the network is converging nicely to a minimum. As the validation loss does not have a positive slope at any point, there is no indication of overfitting.

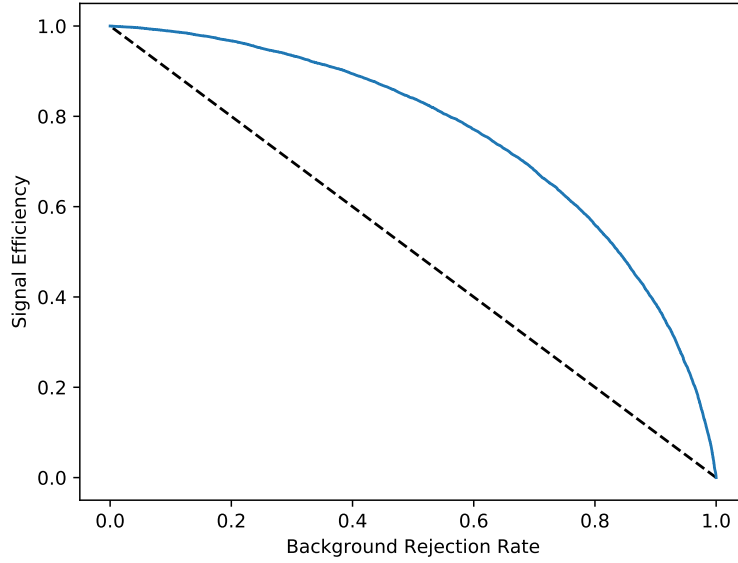


Figure 6.4: Receiver operator characteristic curves for the neural network trained using odd simulated event numbers.

The neural network performance is evaluated using the area under the receiver operating characteristic curve (ROC curve) which is the area under the curve defined by signal efficiency as a function of the background rejection rate. An example ROC curve for the neural network trained on odd event numbers is shown in Figure 6.4. The area under the curve is 0.75 for both the even and odd neural network. To put this in context, an area under the curve of 0.5 would correspond to a classifier that does no better than random guessing. The area under the curve for lepton-photon centrality alone is 0.68 and for dijet mass alone the area under the curve is 0.67. This indicates that the neural network has a significantly better sensitivity to signal than a single variable.

Most importantly is to be able to see the separation in the neural network distribution for signal events and background events to see if this is a quantity that can be fit to extract the signal process. Figure 6.5 shows this distribution on both training sample and testing sample for the neural network trained using odd simulated event numbers. This shows not only that the neural network is a good discriminating variable, but that it generalizes well as the test sample is similar in shape to the training sample. This assures that the network is not overfitting the training sample.

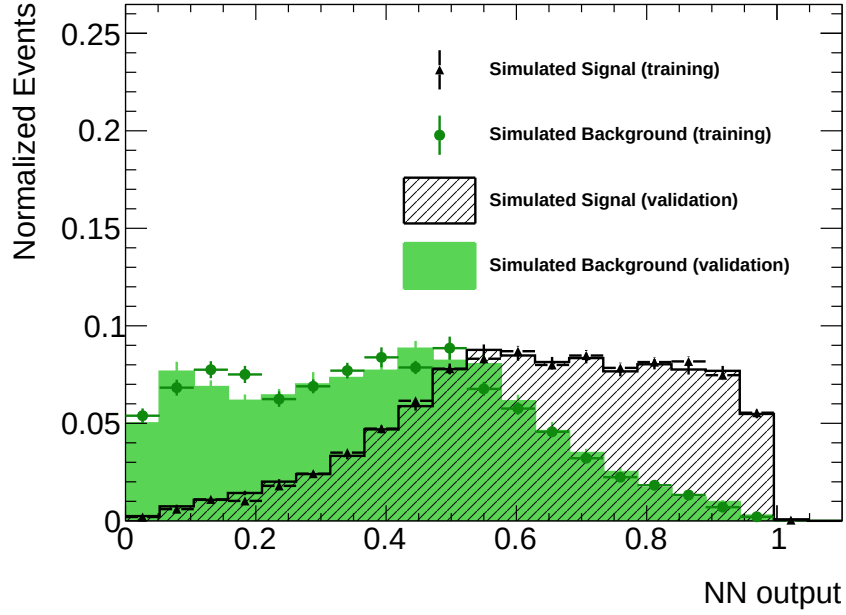


Figure 6.5: Distribution of neural network output of signal and background for training and validation samples for the neural network trained using odd simulated event numbers.

Finally, to get a handle on what the neural network is learning, the profiles of the training variables as a function of neural network output are plotted. The amount of correlation between a training variable and the neural network output value indicates the importance of the training variable. In Figure 6.6 it can be seen that there is a strong correlation between the centrality of the $l\gamma$ system and the neural network output, while $\Delta\phi(W\gamma, jj)$ is largely flat as a function of neural network output. To validate that the network should generalize well to data, in Figure 6.6, the neural network output value in data is also shown.

The importance of each of the training variables is also evaluated in a more quantitative manner by training the neural network without a given training variable and comparing the networks performance with that of the neural network trained on all variables. The most important training variable is the variable that results in the largest drop in area under the receiving operating characteristic curve when removed. The ranking of all training variables is shown in Table 6.2. The most important training variable is $\xi_{l\gamma}$.

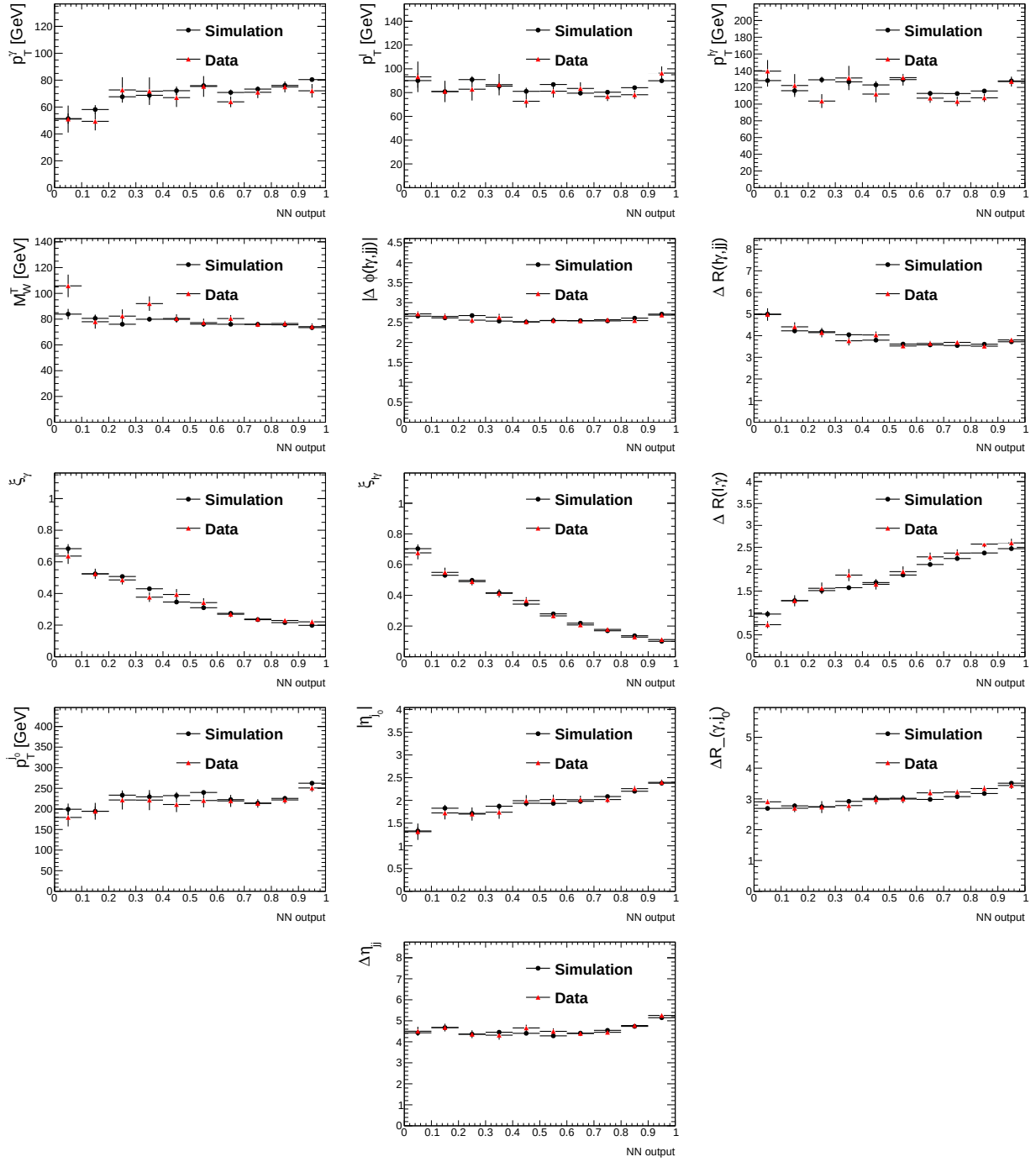


Figure 6.6: Average value of training variables in bins of neural network output value in simulated signal and background events and in data. The neural network output for data is the average of the odd and even event number neural networks. The baseline selection is applied with an additional veto on additional jets in the rapidity gap between the two leading jets.

Training Variable	Even NN $\frac{AUC_{N-1}}{AUC_N}$	Odd NN $\frac{AUC_{N-1}}{AUC_N}$
$\xi_{l\gamma}$	0.96	0.96
$\Delta\eta_{jj}$	0.96	0.96
$\Delta R(l\gamma, jj)$	0.98	0.98
p_T^γ	0.99	0.99
p_T^l	0.99	0.99
ξ_γ	0.99	0.99
$\Delta R(j_0, \gamma)$	0.99	0.99
$\Delta R(l, \gamma)$	0.99	0.99
$p_T^{l\gamma}$	0.99	0.99
$\Delta\phi(l\gamma, jj)$	0.99	0.99
M_T^W	0.99	0.99
$p_T^{j_0}$	0.99	0.99
η_{j_0}	0.99	0.99

Table 6.2: Ranking of training variables by impact on the area under the receiver operating characteristic curve. Each row represents the area under the receiver operating characteristic curve of a neural network trained without a given variable, relative to the area under the receiver operating characteristic curve of a neural network trained on all 13 training variables.

It can be seen that the most important variables have noticeable correlation with neural network output as shown in Figure 6.6. It is interesting to note that some variables with noticeable correlations in Figure 6.6 are of less importance to the neural network as quantified in Table 6.2. This is likely because the neural network can reconstruct the variable from other input variables such that removing the variable from training does not cause a large change in area under the receiving operating characteristic curve.

6.2.3 Neural Network Combination Validation

With an optimal network configuration on a motivated set of training variables, the combination of the neural network trained on simulated events with odd event numbers and simulated events with even event numbers must be validated, i.e. it must be shown that the two neural networks are learning similar physics. For each training variable, the distribution of the training variable that yields a neural network score in a specific range is

plotted. The distribution of the training variable in even event number events in bins of the neural network trained on odd event number events is compared to the distribution of the training variable in odd event number events in bins of the neural network trained on even event number events to validate that the two neural networks are learning the same quantities in simulated events. The comparison is shown for the most important training variable, $\xi_{l\gamma}$ in 5 bins of neural network output in Figure 6.7. The same comparison can then be made where each neural network is evaluated on the full dataset. This comparison is shown in Figure 6.8 for $\xi_{l\gamma}$. In both cases the distributions are statistically compatible suggesting that the two neural networks are learning the same physical quantities and will generalize similarly to data.

In training a neural network to differentiate signal from the prompt backgrounds in the analysis, an observable with a markedly different shape for signal and for background has been created. All the different sources of background described in Chapter 5 are estimated as a function of neural network output values. In Chapter 8 a template fit to the average neural network output distribution in data is performed to extract the number of signal events, from which a statistical significance of observation and a fiducial cross section are calculated.

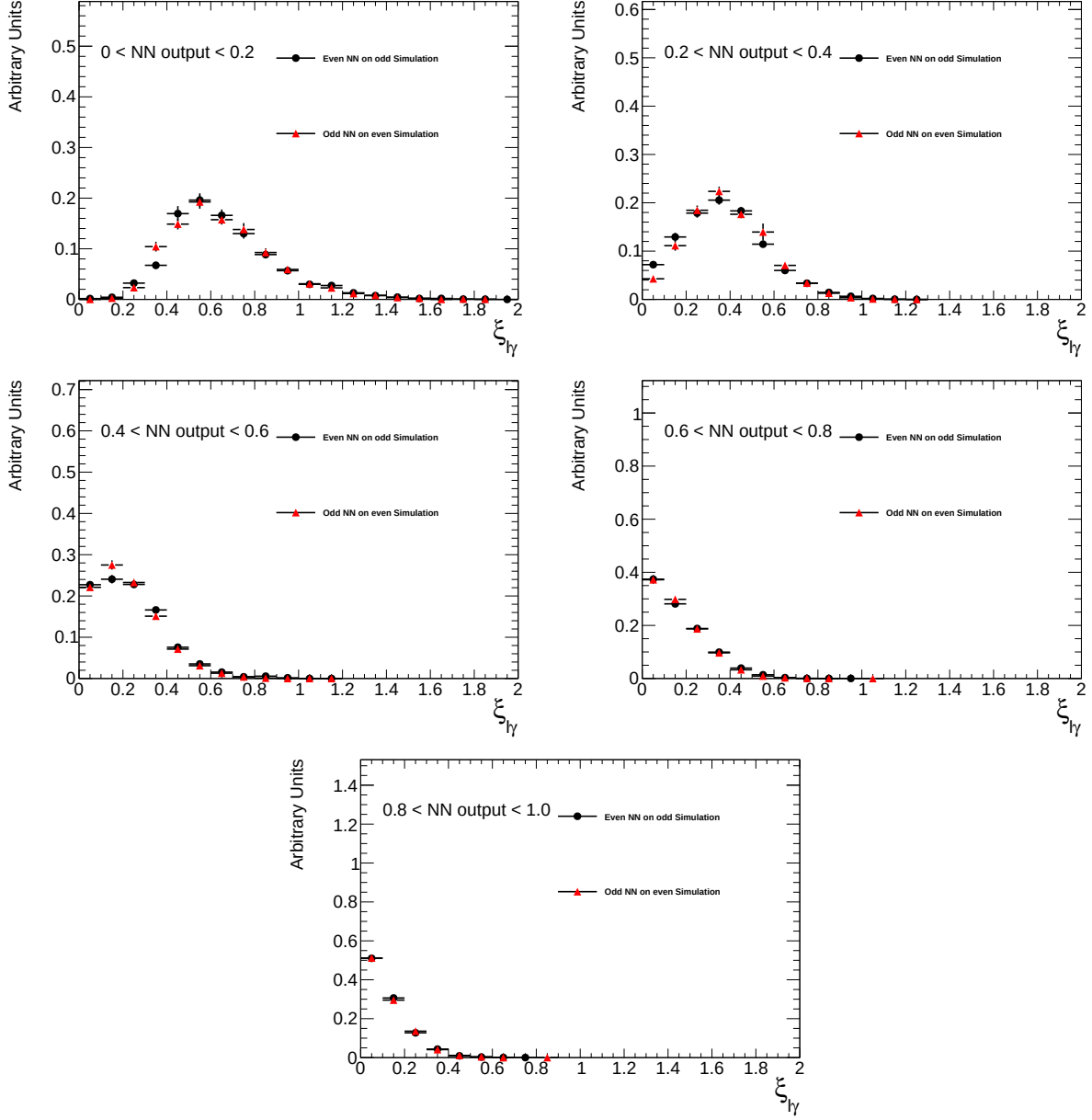


Figure 6.7: Comparison of the $\xi_{l\gamma}$ distributions of simulated events with even(odd) event number in bins of neural network output trained on odd(even) event numbers.

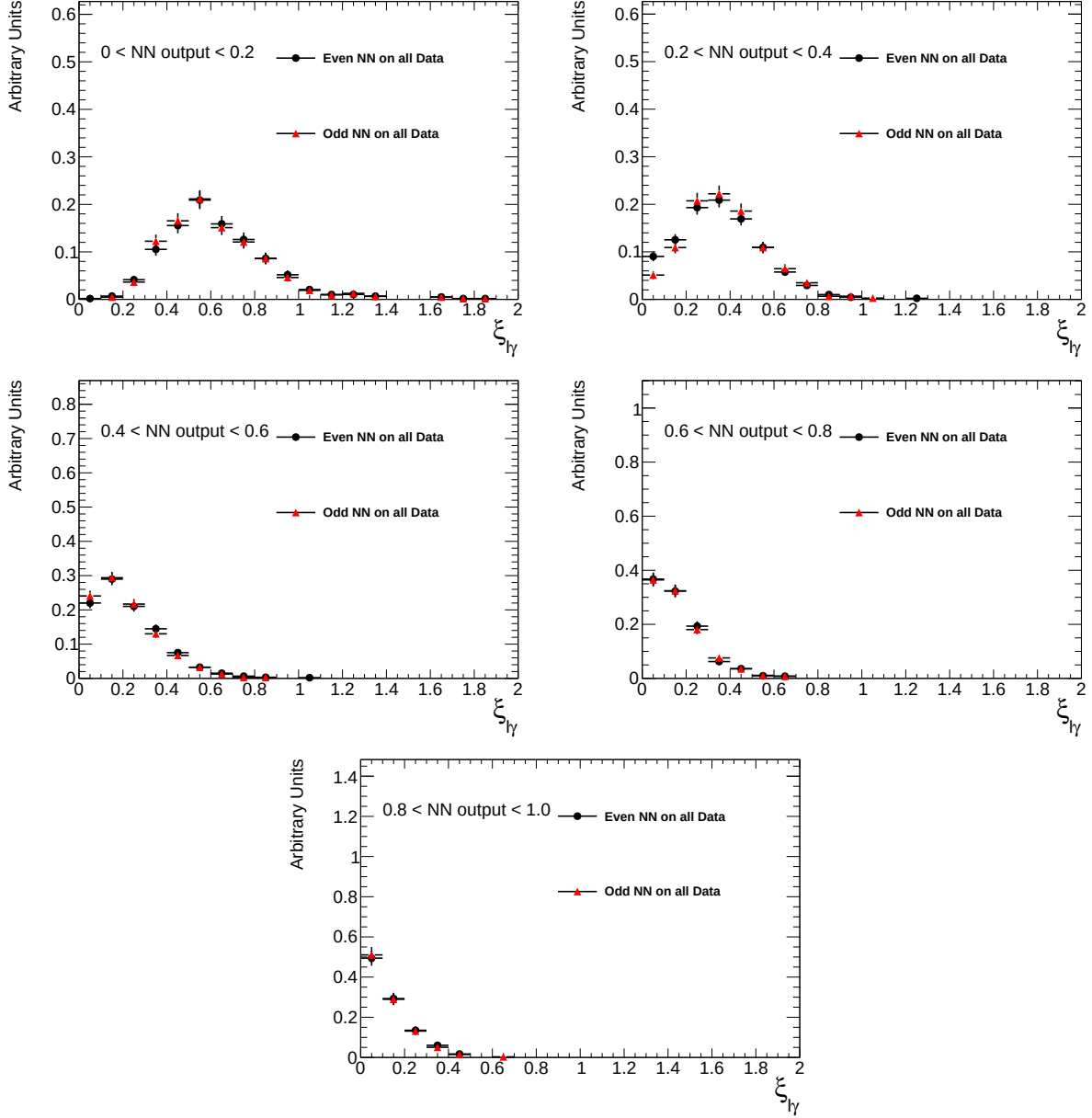


Figure 6.8: Comparison of the $\xi_{l\gamma}$ distributions of data events in bins of neural network output trained on even or odd numbered simulated events

Chapter 7

Uncertainties

In addition to the uncertainties on the data driven background estimates discussed in Chapter 5, other sources of uncertainty must be considered on the background sources estimated using simulated events. These broadly fall into two categories: uncertainties in the modelling of the physics and uncertainties associated with the detector response. In this chapter, these two categories of uncertainties are referred to as theoretical and experimental uncertainties.

7.1 Theoretical Uncertainties

Theoretical uncertainties on the predicted event yields of all simulated event samples used in the analysis are estimated. Simulated event samples are produced following the approach described in Section 3.3. In particular, Equation 2.1 allows for the calculation of an inclusive cross section where $\hat{\sigma}_{ij}$ is calculated to order n in a perturbative expansion in terms of the strong coupling constant α_S . The inclusive cross section calculation is performed assuming a functional form of the parton distribution functions at an energy scale Q and using the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution equations [91–94] to estimate the parton distribution functions at different energy scales. Similarly, the value of α_S is experimentally determined at the energy scale $Q = M_Z$ and

the renormalization group equations [3] are used to estimate the value of α_S at different energy scales. Note that the DGLAP and renormalization group equations are also truncated at some order in a perturbative expansion. It follows that three sources of theoretical uncertainties are readily identified:

1. **Scale Uncertainties:** The quantities $\hat{\sigma}_{ij}$, $d\alpha_S/d\mu_R$ and $df(x, \mu_F)/d\mu_F$ in Equation 2.1 all have expansions in power of α_S with coefficients known up to a certain order. The scale uncertainties refer to uncertainties in the calculated cross section due to the missing higher order terms in these expansions. Here, the two relevant energy scales are the renormalization scale, μ_R , and the factorization scale, μ_F . The renormalization scale is used in the renormalization group equations to eliminate loop divergences arising from the truncation of the α_S expansion [3], while the factorization scale is used in the DGLAP equations to eliminate collinear divergences [3]. The dependence of the inclusive cross section calculation on these two energy scales can be used to estimate the impact of missing higher order terms in the calculation of an inclusive cross section.
2. **PDF Uncertainties:** The parton distribution functions (PDFs) used in Equation 2.1 are obtained from experimental fits to data. Typically, multiple datasets from multiple experiments are included in the fit and these come with corresponding statistical and systematic uncertainties. In addition, theory uncertainties enter into the determination of the PDF calculation from theoretical assumptions made in the PDF fit (e.g. the mass of heavy quarks). Finally, the PDFs derived from a fit to experimental data also depend on the functional form used to describe the PDFs. Collaborations that measure PDFs provide error sets. A comparison of the cross section values obtained using the PDF error sets is used to evaluate this source of uncertainty. In addition, PDFs from different collaborations are used to further quantify this source of uncertainty.

3. **Strong Coupling Uncertainties:** The value of α_S is experimentally determined based on a fit to multiple datasets. Uncertainties on α_S arise from statistical and systematic uncertainties on the measurements used to extract a value of α_S .

All these uncertainties are determined in accordance with PDF4LHC [95] and ATLAS Physics Modelling Group guidelines.

7.1.1 Scale Uncertainties

Both the factorization and normalization scales are varied by a factor of 0.5 and 2.0. Six scale variations corresponding to $(\mu_F, \mu_R) = (1, 0.5), (1, 2), (0.5, 1), (2.0, 1), (0.5, 0.5)$ and $(2.0, 2.0)$ are used to create an envelope, i.e. $\max[\mathcal{O}(\mu_{Fi}, \mu_{Ri}) - \mathcal{O}(\mu_F = 1, \mu_R = 1)]$ in each bin where $\mathcal{O}$ is the observable of interest and i is the scale variation. The envelope then defines the uncertainty arising from the choice of scales. While the envelope approach is the recommendation by the ATLAS Physics Modelling Group, it is worth noting that envelopes can result in underestimating the scale uncertainty if the envelope is flat and individual scale variations are not. The shape of individual scale variations are shown in Appendix H. Since each individual variation is flat, an envelope is a suitable approach in this analysis. The scale uncertainties represent the largest theoretical uncertainty in this analysis and range from 10% to 50% depending on the process simulated.

7.1.2 PDF Uncertainties

The NNPDF collaboration PDF [5] is used as a baseline PDF for all simulated samples. The definition of the PDF variation depends on the set of PDF variations that come with the simulation. For those that come with variations based on Monte Carlo methods (described in [96]) the prescription is to use the standard deviation of the variations as the uncertainty on the PDF while for PDF sets based on Hessian eigenvectors (described in [97]) the prescription is to add the difference with the nominal PDF in quadrature. An additional check on the PDF uncertainties is performed by comparing the results obtained

using PDFs provided by the CTEQ-TEA group [98] with the NNPDF PDF and its uncertainties to confirm that the PDF uncertainties cover the difference between the nominal PDF provided and different PDF predictions. The total PDF uncertainty is below 5% for all simulated samples used in the analysis.

7.1.3 Strong Coupling Uncertainties

The uncertainty on simulated event yields due to uncertainties in the value of α_S is evaluated by comparing the nominal cross section prediction with the predictions when α_S is varied by its uncertainties around its central value. The strong coupling constant central value is $\alpha_S(M_Z) = 0.1180 \pm 0.0015$ [5]. The size of the α_S uncertainty on simulated event yields ranges from 0% for electroweak $W\gamma jj$, where α_S does not enter into the cross section calculation at leading order, to 1-2% for the other simulated event samples.

7.1.4 Other Theoretical Uncertainties

The three sources of theoretical uncertainties described above do not cover all uncertainties in the modelling of physics processes. Specifically, Equation 2.1 does not specify the behaviour of the spectator quarks in the proton that are not involved in the hard scatter. Particles can be produced in soft collisions between spectator partons. These in addition to the beam hadrons make up the underlying event. Since different event generators model the underlying event differently, a theory uncertainty can be assigned by evaluating the difference between two different event generators.

Similarly, uncertainties in the parton shower and hadronization models used to simulate events are estimated by comparing the predicted event yields obtained using different event generators. In this analysis, the difference between two different event generators and parton showering algorithms are used to estimate an uncertainty in the event yield of the signal process, as well as on the event yield of the largest simulated background, the QCD $W\gamma jj$ background.

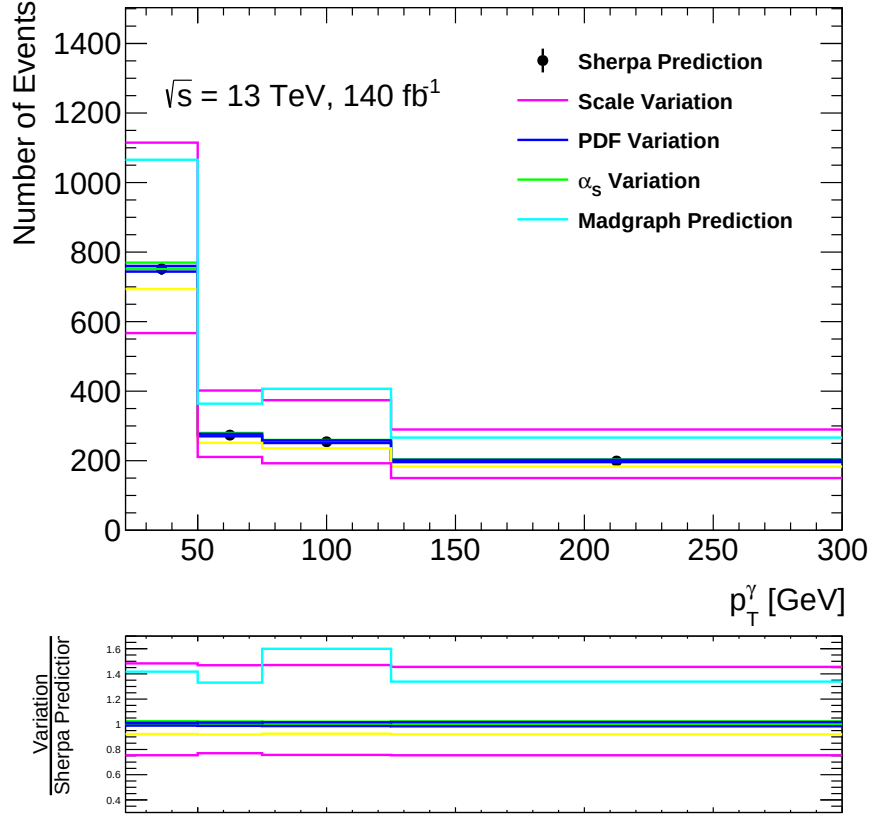


Figure 7.1: The effect of scale, PDF, and α_S variations on the QCD $W\gamma jj$ event yield predicted by the Sherpa event generator as a function of p_T^γ . The yield predicted by the Madgraph event generator is also shown.

The scale, PDF and α_S uncertainties on the simulated QCD $W\gamma jj$ event yield predicted by the Sherpa event generator as a function of p_T^γ are shown in Figure 7.1. In addition, the yield predicted by the Madgraph event generator is shown.

7.2 Experimental Uncertainties

Experimental uncertainties cover uncertainties arising from efficiencies of cuts on object selections, detector resolution and calibration. The methods used to estimate these uncertainties are outlined in this section. The systematic uncertainties are grouped according to the physical object reconstruction or analysis steps in the analysis chain affected. The

effect of each source of uncertainty on the final measurements are discussed in Chapters 8 and 9.

- **Photons:** The selection cuts used to identify *tight* photons that are also isolated have different efficiencies in simulated samples and data. The simulated event yields are scaled to match data distributions using scale factors. The scale factors are calculated with methods that come with their own modelling uncertainties and their own biases [69]. For example, the uncertainty on the identification scale factor ranges from 0.5% (high p_T , unconverted photons) to 12% (low p_T , converted photons) while the uncertainty on the photon isolation energy scale factor is around 1%.

In addition to the photon selection cut efficiencies, the simulation of the calorimeter energy scale and resolution differs from what is observed in data. The scale and resolution of the photon energy is calibrated to match data using $Z \rightarrow e^+e^-$ events. This calibration comes with uncertainties associated with electronics calibration, detector geometry and uncertainties in simulation. Overall 64 different sources of uncertainty describe the uncertainty in photon energy scale and 9 different sources of uncertainty describe that of the photon energy resolution [99].

- **Electrons:** Electrons are similarly affected by the uncertainties in the energy scale and resolution of the electromagnetic calorimeter. Scale factors to correct for differences between data and simulation in the efficiency of the electron isolation and identification cuts are also used and carry their own uncertainties. In addition, other scale factors with their uncertainties are used to correct the electron trigger efficiency and the electron reconstruction efficiency [100].
- **Muons:** Muons come with scale factors accounting for differences observed in data and simulation in their identification, isolation, trigger and reconstruction efficiencies. An additional scale factor is also applied to muons in simulated events to correct for the track to vertex association efficiency. Uncertainties on the momentum scale and the resolution of the muon spectrometer are also taken into account.

- **Jets:** The jet energy scale is calibrated using the data to simulation ratio of the transverse momentum balance between jets and a reference object [101]. Uncertainties in the jet energy scale come from the different reference objects used, i.e. Z bosons, photons, and multijets. The value of each uncertainty depends on the jet transverse momentum and pseudorapidity. Additional uncertainties arise from the jet energy that is not fully contained in the hadronic calorimeter (punch-through uncertainty) and uncertainty in the pileup subtraction used to correct the jet energy. Over 100 different uncertainty sources describe the uncertainty on the jet energy scale [101]. Similarly, there are several uncertainties on the jet energy resolution that come from the difference in jet energy resolution between data and simulation. The jet energy resolution also has several sources of uncertainties associated with the pileup subtraction technique used to correct for the jet energy.

Another source of uncertainty is the difference in the detector response to light and heavy flavor jets in different generators [102]. These uncertainties are derived by comparing different hadronization models in the Sherpa event generator [103], different shower models in the Herwig event generator [104], as well as the differences between Pythia [105] and Sherpa which capture differences in multi-parton interactions, underlying event and hadronic shower effects.

Additional uncertainties are included in the calculation of scale factors to correct for differences between data and simulation in the efficiency of b-jet tagging and jet vertex tagging.

- **Missing Transverse Energy:** As E_T^{miss} is calculated by balancing the transverse momentum of objects in the transverse plane along with tracks not matched to physics objects, the momentum scale and resolution of the physics objects as well as the tracks affect the missing transverse energy calculation [74].
- **Pileup Reweighting:** The simulated event overlay of proton-proton interactions used to simulate pileup conditions is reweighted to match the average number of

interactions per bunch crossing observed in data. The statistical uncertainty associated with this reweighting technique is taken into account.

- **Luminosity:** The luminosity measurement is performed with multiple detectors using data recorded during designated “van der Meer” scans performed with a small number of protons per bunch (low- μ) and fewer proton bunches in the bunch train. The instantaneous luminosity during these special LHC runs designated for luminosity measurements is orders of magnitude below that of the instantaneous luminosity during normal Run 2 physics data taking. The dominant source of uncertainty in the luminosity measurement comes from the uncertainty in extrapolating from the lower instantaneous luminosity of the special runs to the physics data taking conditions. Other large sources of uncertainty come from beam-beam effects, magnetic non-linearity in “van der Meer” scans, and scan-to-scan reproducibility [66].

The uncertainties on individual objects are propagated to event yields by repeating the analysis selection with scale factors or resolution and scale parameters varied by their uncertainties. To provide the reader with an example, the effect of the different categories of experimental uncertainties on the overall yield of electroweak $W\gamma jj$ and QCD $W\gamma jj$ events are summarized in Table 7.1. As can be seen, the largest sources of experimental uncertainty come from the jet energy scale and calibration.

Uncertainty Source	Impact on EW $W\gamma jj$ yield [%]	Impact on QCD $W\gamma jj$ yield [%]
Photons	1.9	2.0
Electrons	0.6	0.6
e/ γ Scale and Resolution	0.4	0.4
Muons	0.8	0.9
Jets	3.2	4.4
E_T^{miss}	0.7	0.8
Pileup	1.9	1.3
Luminosity	0.8	0.8

Table 7.1: Relative uncertainty values on simulated electroweak and QCD $W\gamma jj$ event yields for different categories of experimental uncertainties. The relative values from individual sources of uncertainty in a category are added in quadrature.

Chapter 8

Statistical Significance and Cross Section Measurement

This chapter describes the method used to extract the number of signal events observed in data. This is done using a profile likelihood fit of the neural network output distribution described in Chapter 6 taking into account the number of background events and the uncertainties on the predicted event yields as described in Chapter 7.

As described below, the free parameters of the fit are normalization factors for the electroweak signal process and the QCD $W\gamma jj$ background. The normalization factor for the electroweak signal μ_{EW} is the *parameter of interest* in the fit. The fit uses two regions: a *signal region* defined by requiring that an event has no jets in the rapidity gap between the two leading transverse momentum jets, and a *control region* defined by requiring that an event has at least one jet in the rapidity gap between the two leading jets. The control region is enriched in QCD $W\gamma jj$ events while the signal region is enriched in electroweak $W\gamma jj$ events. Both regions are used to constrain the electroweak $W\gamma jj$ and QCD $W\gamma jj$ normalization factors but because of the relative size of the QCD $W\gamma jj$ and electroweak $W\gamma jj$ yields in the signal and control region, the QCD normalization factor is primarily constrained in the control region while the signal normalization factor is primarily constrained in the signal region. To calculate an expected statistical significance of signal

observation, as described in Section 8.2, data is unblinded in the control region but a so-called *Asimov* dataset is used in the signal region. The observed significance reported in Section 8.3 is obtained from a fit to data in both the signal and control regions. In Section 8.4, the observed number of signal events extracted is used to calculate a fiducial cross section.

8.1 Statistical Model and Statistical Significance Calculation

A profile likelihood fit is performed to maximize the likelihood of observing n events given that ν events are expected [106]. For this analysis, the likelihood function $\mathcal{L}$ can be written in a simplified form as the product of Poisson distributions

$$\mathcal{L} = \prod_i^{N_{bins}} \prod_r^{N_{reg.}} P(n_{i,r} | \mu_{EWK} S_{i,r}(\theta) + B_{i,r}(\mu_{QCD}, \theta)) \quad (8.1)$$

where N_{bins} is the number of bins of the kinematic distribution, $N_{reg.}$ is the number of regions, P represents a Poisson distribution, $n_{i,r}$ is the number of observed events in data in bin i and region r , $S_{i,r}$ is the predicted number of signal events in bin i and region r , $B_{i,r}$ is the predicted number of background events in bin i and region r , and θ generically refers to a nuisance parameter associated with an uncertainty in the signal prediction or the background estimation. The normalization factors μ_{EW} and μ_{QCD} are parameters that are allowed to float in the fit. The parameter μ_{QCD} is defined as

$$\mu_{QCD} = \frac{\sigma_{QCD}^{measured}}{\sigma_{QCD}^{predicted}} \quad (8.2)$$

and μ_{EW} , defined as

$$\mu_{EW} = \frac{\sigma_{EW}^{measured}}{\sigma_{EW}^{predicted}} \quad (8.3)$$

where $\sigma_{EW(QCD)}^{measured}$ and $\sigma_{EW(QCD)}^{predicted}$ denote the measured and predicted cross sections of EW (QCD) $W\gamma jj$ production, respectively.

Nuisance parameters, θ , are categorized into *Gaussian nuisance parameters* and *Gamma nuisance parameters*. Gaussian nuisance parameters, α , correspond to the sources of uncertainty that have been estimated with auxiliary measurements but for which their true parent distribution is not known. These include the uncertainties associated with the expected yields from non-prompt background sources, the theoretical uncertainties on the prompt background yield estimate, and the experimental uncertainties. These uncertainties are included in the likelihood as a product of Gaussian constraint terms

$$\mathcal{L}(\alpha) = \prod_j^{N_\alpha} G(\alpha_j|0, 1) = \prod_j^{N_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha_j^2}{2}} \quad (8.4)$$

where N_α represents the number of uncertainties and α_j are free parameters of the fit. As such, the likelihood favors values of α that are close to their estimated value.

The Gamma nuisance parameters, γ , account for the statistical uncertainty on the sum of backgrounds in a given bin. They are constrained using a product of Poisson distributions

$$\mathcal{L}(\gamma) = \prod_k^{N_\gamma} P(b_k|\gamma_k b_k) = \prod_k^{N_{bins}} \frac{(\gamma_k b_k)^{b_k} e^{-\gamma_k b_k}}{\Gamma(\gamma_k b_k + 1)} \quad (8.5)$$

where b_k is the number of expected events in bin k , γ_k are free parameters of the fit and

$$\Gamma = \int_0^\infty t^{x-1} e^{-t} dt \quad (8.6)$$

is the Gamma function which replaces the factorial $\gamma_k b_k!$ in a typical Poisson distribution to account for the possibility that the expected number of events is not an integer.

Putting all these components together, a more complete description of the likelihood function used in the fit to data is

$$\begin{aligned} \mathcal{L}(\mu_{EW}, \mu_{QCD}, \alpha, \gamma) = & \prod_i^{N_{bin}} \prod_r^{N_{reg.}} P(n_{i,r} | \mu_{EW} S_{i,r}(\alpha, \gamma) + B_i(\mu_{QCD}, \alpha, \gamma)) \\ & \times \prod_j^{N_\alpha} G(\alpha | 0, 1) \prod_k^{N_\gamma} P(b_k | \gamma_k b_k) \end{aligned} \quad (8.7)$$

Each experimental uncertainty is assigned one nuisance parameter since experimental uncertainties are correlated across different physics processes simulated and analysis regions. Theory uncertainties like PDF uncertainties that are constrained by measurements are assigned one nuisance parameter per physics process simulated. There is no *a priori* reason to believe uncertainties arising from missing orders of a perturbative expansion are correlated between regions. Scale uncertainties are thus assigned one nuisance parameter per region and per physics process simulated. The difference between the two QCD $W\gamma jj$ event generators is assigned two nuisance parameters, one derived from the difference in the shape between the two generators and another derived from the difference in normalization between the two generators. These two nuisance parameters along with the scale uncertainty may be pulled in the fit to account for mismodelling of the QCD $W\gamma jj$ background. Similarly, one nuisance parameter is derived from difference of shape between the two electroweak $W\gamma jj$ generators.

To find the maximum likelihood estimators of μ_{EW} , μ_{QCD} and all nuisance parameters the negative log likelihood is minimized. Using the symbol θ to represent the Gaussian and Gamma nuisance parameters, the best values of μ_{EW} , μ_{QCD} and θ parameters are obtained when

$$-\frac{\partial [\ln \mathcal{L}]}{\partial \mu_{EW} \partial \mu_{QCD} \partial \theta} = 0 \quad (8.8)$$

Next a frequentist significance test can be used to establish whether or not the measured value, $\hat{\mu}_{EW}$, constitutes an observation. The significance test makes use of the profile

likelihood ratio

$$\lambda(\mu_{EW}) = \frac{\mathcal{L}(\mu_{EW}, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}_{EW}, \hat{\theta})}. \quad (8.9)$$

where $\hat{\hat{\theta}}$ is the value of nuisance parameters which maximize $\mathcal{L}$ for a given value of μ_{EW} and $\hat{\mu}_{EW}$ and $\hat{\theta}$ are values of the estimators that maximize $\mathcal{L}$. These are referred to as the conditional and unconditional maximum likelihood estimators, respectively. The profile likelihood ratio, $\lambda(\mu_{EW})$, is close to one when data agrees well with the hypothesized value of μ_{EW} .

A test statistic can then be defined to quantify the compatibility between μ_{EW} and the measured data. The profile likelihood ratio test statistic is denoted

$$t_{\mu_{EW}} = -2 \ln[\lambda(\mu_{EW})] \quad (8.10)$$

such that the test statistic is 0 when data is in perfect agreement with the model under a hypothesized value of μ_{EW} and the test statistic value grows with incompatibility between the hypothesized value of μ_{EW} and data. This incompatibility is quantified using the p -value, defined as

$$p = \int_{t_{\mu_{EW}}^{obs}}^{\infty} f(t_{\mu_{EW}} | \mu_{EW}) dt_{\mu_{EW}} \quad (8.11)$$

where $f(t_{\mu_{EW}} | \mu_{EW})$ is the probability density function of $t_{\mu_{EW}}$ for a given μ_{EW} . According to Wilke's theorem [107], in the large statistics limit, this probability density function is a χ^2 distribution. This can be converted to an *equivalent significance*, Z , using

$$Z = \Phi^{-1}(1 - p) \quad (8.12)$$

where Φ^{-1} is the inverse of the cumulative distribution of the standard Gaussian. This means a Gaussian distributed variable has probability p of being found Z standard deviations away from the mean. The frequentist significance test is described to compare the measured value of μ_{EW} with the null hypothesis, $\mu_{EW} = 0$, otherwise described as the

background only hypothesis. In this case, the equivalent significance is given by

$$Z = \sqrt{-2 \ln \lambda(\mu_{EW} = 0)}. \quad (8.13)$$

It is an accepted standard in the field that rejecting the null hypothesis with an equivalent significance of $Z = 5$ constitutes an observation of the process.

8.2 Expected Statistical Significance

Before unblinding the analysis, a fit is performed to obtain an *expected statistical significance* using a so-called *Asimov dataset* constructed from data in the control region and from the expected yields in the signal region. In this way the fit model can be studied without risk of bias that might come with observing the data in the signal region.

The Asimov dataset is constructed by first performing a background only fit in the control region only. The nuisance parameters and QCD $W\gamma jj$ normalization factor obtained in this fit are then applied to the expected yields in the signal region to create the Asimov dataset in the signal region. Expected results are obtained by simultaneously fitting the normalization factors and nuisance parameters to data in the control region and to the Asimov dataset in the signal region. By construction, the fitted value of the parameter of interest should be 1 and the fitted value of the QCD normalization factor will be the value to which it is constrained in the control region only.

The background only fit used to construct the Asimov dataset is shown in Figure 8.1. As can be seen before the fit there is good agreement between the data and simulation in the control region except for a slight mismodelling of shape. This is corrected for by the nuisance parameters in the fit associated with the QCD $W\gamma jj$ theory uncertainties. The resulting fitted value of the QCD $W\gamma jj$ normalization factor is

$$\mu_{QCD} = 1.03 \pm 0.46 \quad (8.14)$$

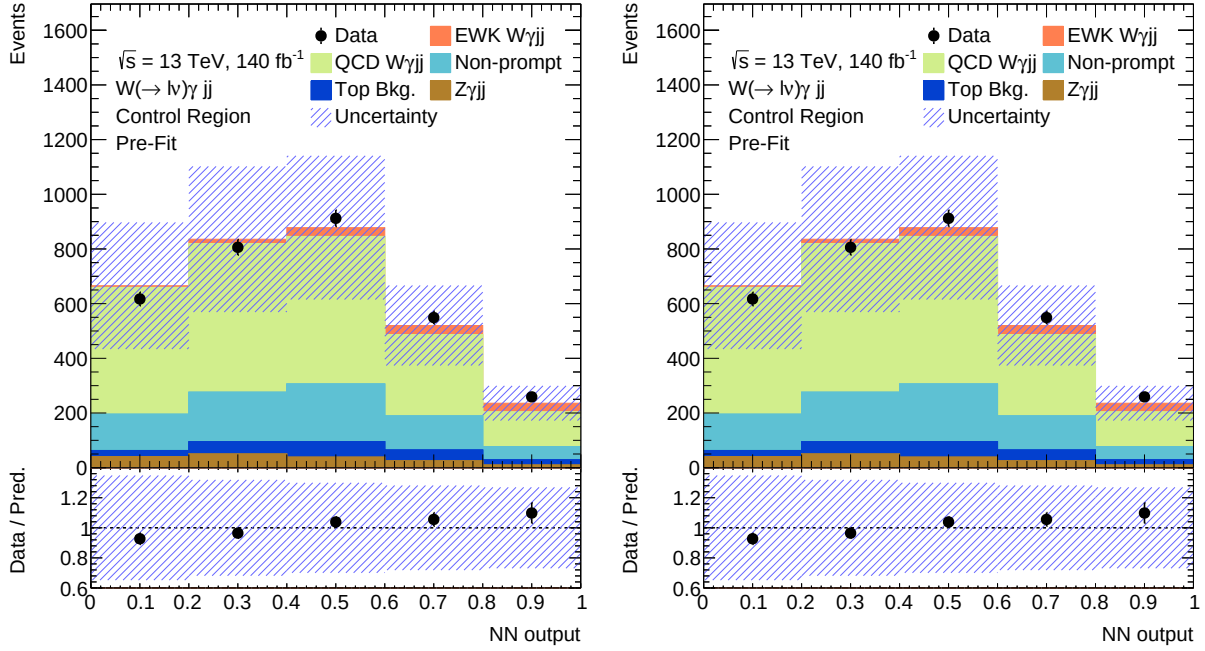


Figure 8.1: Comparison of data and the signal and background model in the control region before (left) and after (right) the background only fit to data. The obtained value of μ_{QCD} and the nuisance parameters in this fit are used to construct the Asimov dataset in the signal region.

where the uncertainty on the parameter includes all uncertainties affecting the background prediction in the analysis. This fitted value of the normalization factor, along with the fitted values of the nuisance parameters, are injected into the templates in the signal region and the full fit as described above is performed. The resulting fitted values of the normalization factors in this fit are

$$\mu_{EW} = 1.00 \pm 0.17 \quad (8.15)$$

and

$$\mu_{QCD} = 1.03 \pm 0.36 \quad (8.16)$$

as expected given that this fit is performed using an Asimov dataset.

The expected statistical significance is calculated using the procedure described above assuming the null hypothesis $\mu_{EW} = 0$. The expected significance obtained from the simultaneous fit of data in the control region and the Asimov dataset in the signal region is 6.3σ . The value of the likelihood ratio test statistic as a function of the parameter of interest, μ_{EW} , is plotted in Figure 8.2. The significance value can be calculated using Equation 8.13 and the value of λ at $\mu_{EW} = 0$, corresponding to the background only hypothesis.

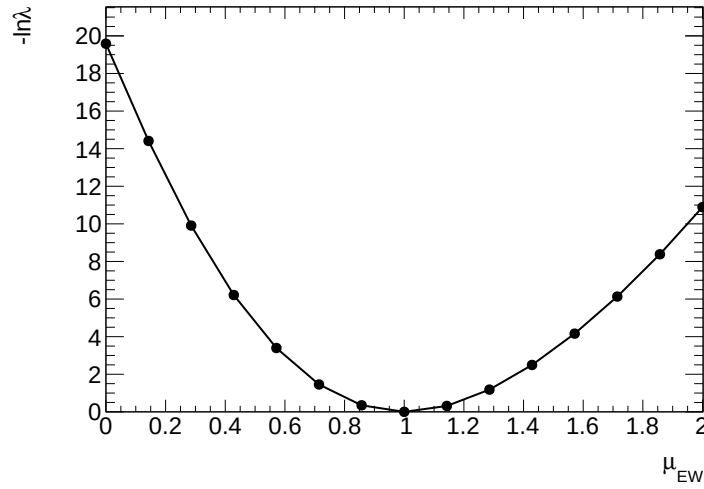


Figure 8.2: Value of the likelihood ratio test statistic as a function of the parameter of interest, μ_{EW} , in the fit to data in the control region and an Asimov dataset in the signal region.

Given there is reasonable agreement between data and the model in the control region, the fit converges and no unreasonable pulls of nuisance parameters are seen, and a significance well above the observation threshold is expected, the fit is next applied to real data in both the signal and control regions.

8.3 Observed Statistical Significance

The fit to data in both the signal and control regions is shown in Figure 8.3. As can be seen, there is reasonable agreement in the shape of data compared to predictions in

the signal and control regions, but a statistically significant difference in normalization in the signal region. The electroweak normalization μ_{EW} corrects for the normalization difference between data and predictions and good agreement is seen post-Fit.

The fitted values of the normalization factors obtained are shown in Table 8.1 and the

Signal Value	Parameter
μ_{EW}	1.51 ± 0.19
μ_{QCD}	1.01 ± 0.36

Table 8.1: Measured electroweak normalization factor (μ_{EW}) and QCD normalization factor (μ_{QCD}) after performing the simultaneous fit to data in both signal and control regions.

pull of all nuisance parameters in the fit can be seen in Appendix I. The background only hypothesis is rejected with a significance of 9.0σ as can be seen in the scan of the test statistic in the parameter of interest in Figure 8.4. This constitutes the first observation of the electroweak production of $W\gamma jj$ using data recorded by the ATLAS detector.

8.4 Fiducial Cross Section Measurement

To provide a result that can be compared to measurements performed by other experiments and to predictions of beyond Standard Model theories, a measured number of events or kinematic distribution should be corrected for the finite resolution and limited acceptance of the detector. This correction is called *unfolding*. Generally in the procedure of unfolding, reconstructed quantities from detector objects must be compared with *truth* quantities in the simulation before simulating detector effects. This can be done with an unfolding correction factor

$$C = \frac{N^{reco}}{N^{fid}} \quad (8.17)$$

where N^{reco} is the number of simulated signal events passing the signal region selection, while N^{fid} is the number of simulated events in the sample passing a set of selection

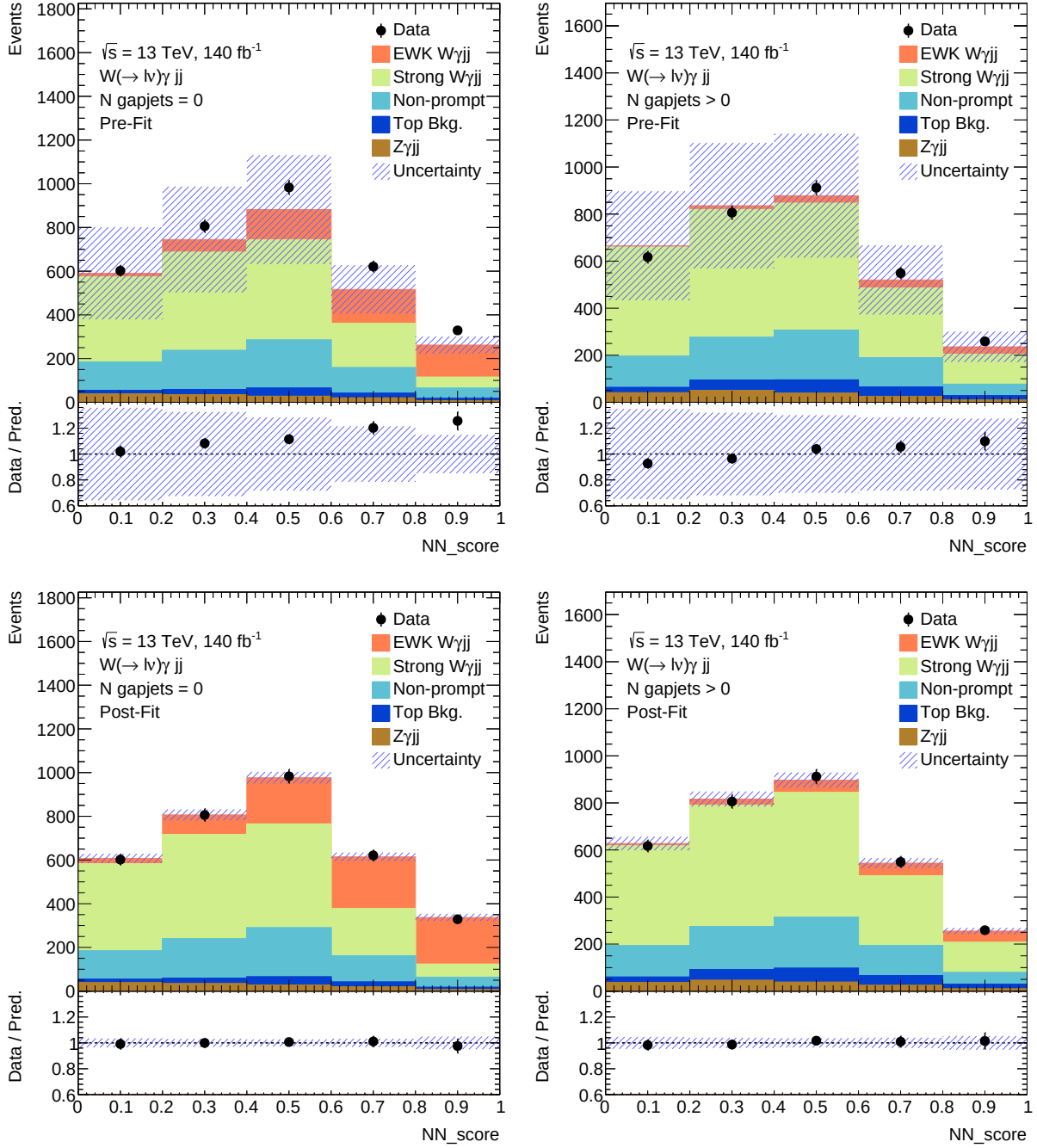


Figure 8.3: Comparison between data and the signal and background model in the signal region (left) and the control region (right) before (top) and after (bottom) the simultaneous fit to data in both regions is performed.

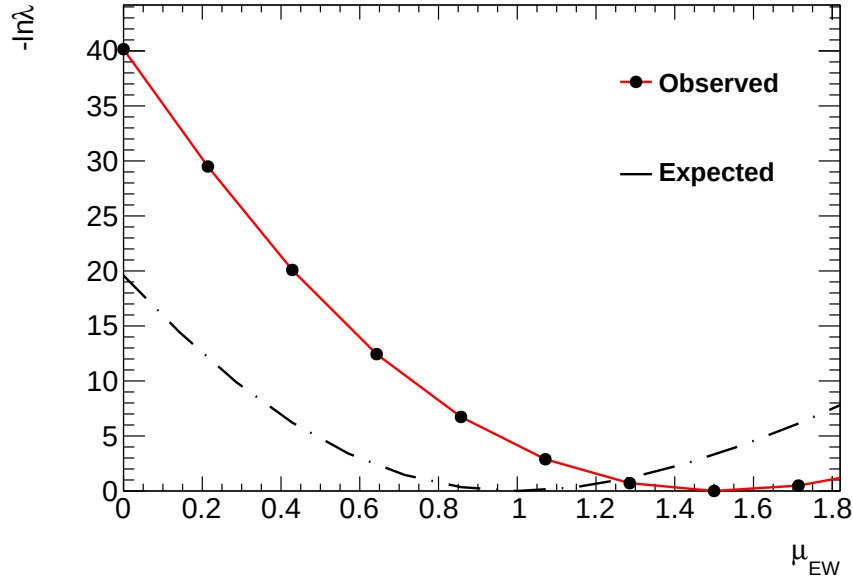


Figure 8.4: Value of the likelihood ratio test statistic as a function of the parameter of interest μ_{EW} in the fit to data in the signal and control regions (observed) and in the fit to data in the control region and an Asimov dataset in the signal region (expected) taken from Figure 8.2.

criteria that is made to mimic the signal region selection using only truth information. The set of criteria applied to truth-level information defines the “fiducial region”.

8.4.1 Fiducial Region Definition

A fiducial region is defined with three goals in mind. The fiducial region should represent an area of phase space that is as close as possible to the reconstructed region to avoid extrapolating the measurement into phase space that is not sampled in data. The fiducial region should be chosen such that theoretical predictions do not change dramatically with different modelling of Standard Model physics processes; in other words, the fiducial volume should be largely model independent. Finally, the fiducial region should be simple and easy to interpret. RIVET [108] is a standard tool for comparing measured results with simulated predictions. A RIVET routine for this analysis is provided for comparison with theory predictions.

With these considerations in mind, final state objects in the fiducial region are defined as follows:

- **Leptons:** Leptons are prompt electrons and muons in the event generator truth record. Leptons generated from hadron decays are excluded. Photons within a cone of $\Delta R < 0.1$ from the prompt leptons are removed from the event and their transverse momentum is added to the lepton to account for radiation of photons from leptons.
- **Jets:** Jets are defined at the event generator truth level by collecting all stable particles in the event, removing the leptons as defined above, and clustering them with the anti- k_t algorithm described in Section 4. As a final step, isolated photons in the truth record are then removed from the resulting list of jets.
- **Missing Transverse Energy:** The missing transverse energy is calculated as the negative of the transverse momentum of all neutrinos in the generator record, excluding neutrinos from hadron decays.
- **Photons:** Photons considered in the definition of the fiducial region are those appearing in the event generator truth record. These photons are then further required to satisfy an isolation requirement.

At truth level, photon isolation energy is defined as the sum of the transverse energy of all stable particles in the event generator truth record that lie within a cone of radius of $R = 0.4$ from the photon, excluding muons, neutrinos and the photon itself. This is different from the photon isolation energy calculated at detector level. At detector level, a subtraction of calorimeter cell energy at the center of the photon candidate accounts for the photon transverse energy.

A lack of photon isolation cut would result in unfolding correction factors that are highly dependent on the final state production mechanism. Since a photon is more likely to fail the reconstructed isolation cut in events with high jet multiplicities, the same be-

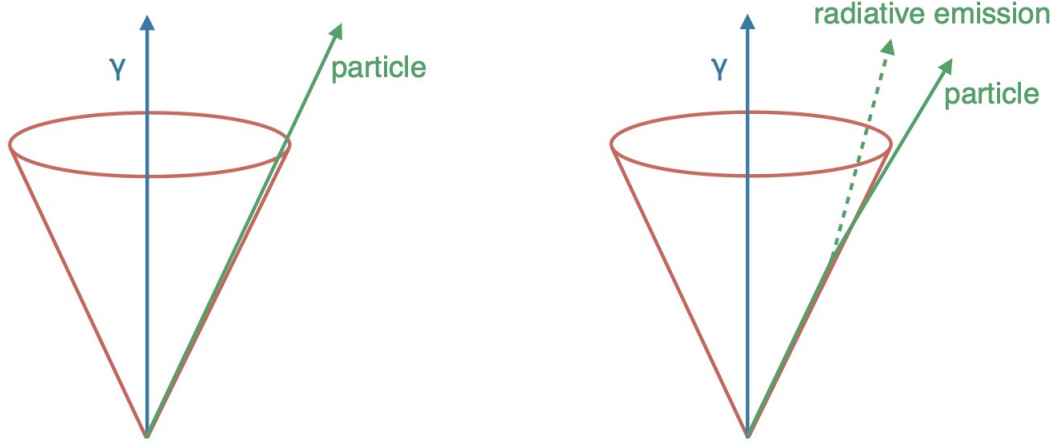


Figure 8.5: Schematic diagram representing model dependence considerations in truth photon isolation definition. On the left, the energy of the particle in the photon isolation cone is fully included in the isolation energy calculation. On the right radiative emission from the particle migrates into the photon isolation cone, introducing a dependence on the fragmentation model

havior should be mimicked at truth level. However a photon isolation cut can also introduce a dependence on the modelling of fragmentation, or the splitting of a parton into the multiple hadrons that make up a jet. One can imagine a parton created at the same vertex as a photon. If the momentum vector of the parton falls within the photon isolation cone, the full parton energy is included in the isolation energy calculation; however if the momentum vector lies just outside the photon isolation cone, radiative emission from the parton could migrate into the isolation cone boundary. This introduces a dependence on the modelling of real emission into the photon isolation calculation. This consideration is illustrated in Figure 8.5.

To minimize model dependence, a truth photon isolation cut such that half the photons with truth isolation at the value of the cut pass the reconstructed level isolation cut and half the photons at the value fail this cut is chosen. The value of the cut satisfying this criteria is found by mapping the truth isolation energy, scaled by the photon transverse momentum, to the reconstructed isolation energy and choosing the value that maps to

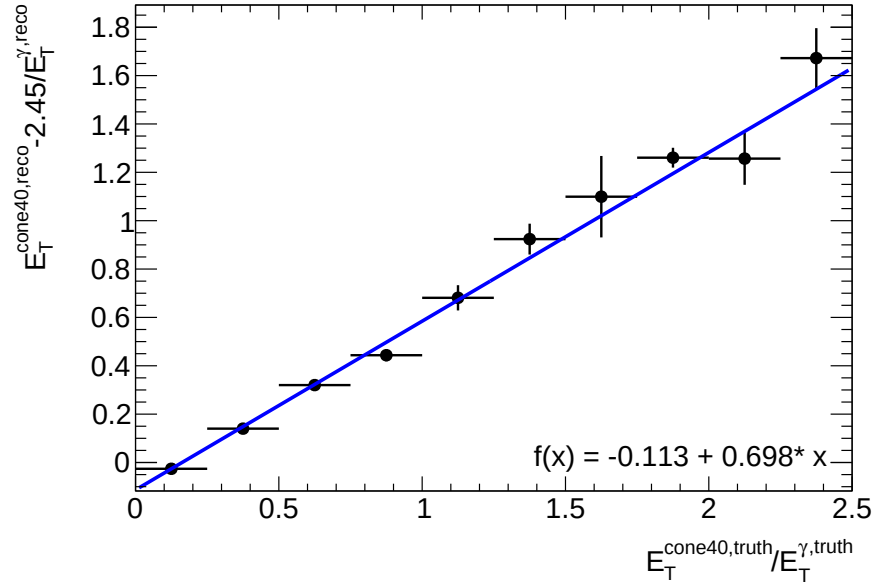


Figure 8.6: The average value of the isolation energy of reconstructed photons as a function of the truth value of isolation energy in simulated signal events. Both values are scaled by the photon transverse momentum.

the value of the cut used at reconstructed level, $E_T^{\text{cone40, reco}} < 0.022 E_T^{\gamma, reco} + 2.45$. This fit is shown in Figure 8.6.

This method suggests an isolation cut at truth level of $E_T^{\text{cone40, truth}} < 0.2 E_T^{\gamma, truth}$. Plotting the ratio of events that pass the reconstructed isolation to those that fail the reconstructed isolation as a function of the truth photon isolation cut (Figure 8.7) shows that the proposed value of the cut has the desired behavior of accepting approximately 50% of events. Note that the same cut is obtained with a different simulated event generator used to model the signal process.

Truth-level objects have now been defined following the goals outlined in defining the fiducial region. The same overlap removal algorithm is applied to truth objects as is done on reconstructed objects and the identical event selection requirements are applied. As a result, the unfolding procedure does not extrapolate through large regions of phase space that are not sampled by the experiment.

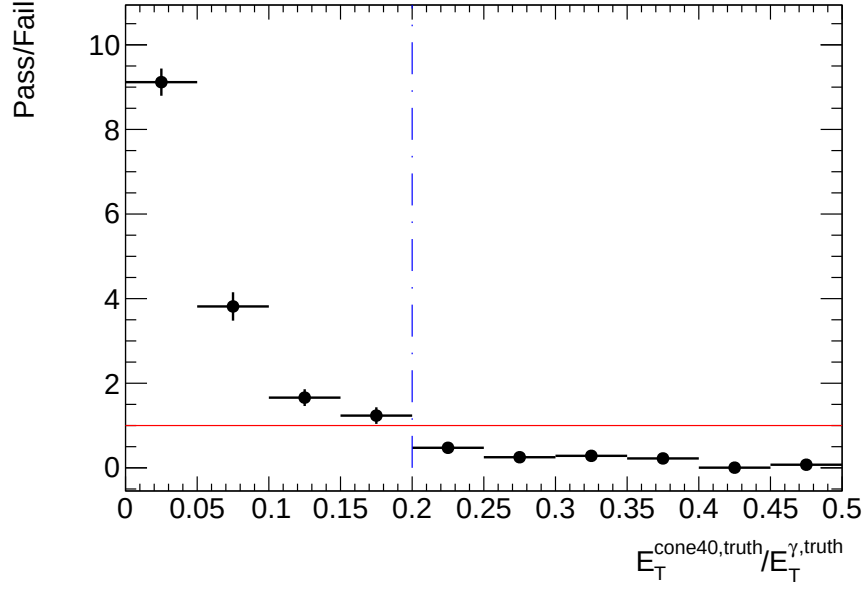


Figure 8.7: The ratio of events in which the photon passes the reconstructed detector level isolation energy cut to those that fail in bins of truth level photon isolation energy scaled by truth photon transverse energy.

8.4.2 Fiducial Cross Section

With the fiducial region defined, a fiducial cross section is calculated as

$$\sigma^{fid} = \frac{\mu_{EW} N_{EW}^{predicted}}{C \mathcal{L}} \quad (8.18)$$

where $N_{EW}^{predicted}$ is the number of predicted signal events in the signal region, μ_{EW} is the measured signal normalization factor obtained in the fit, $\mathcal{L}$ is the integrated luminosity of the dataset, and C is the unfolding correction factor used to account for the imperfect reconstruction efficiency of the detector as described above. The unfolded correction factor is calculated using simulated reconstructed and truth level events from the Sherpa event generator. The unfolding correction factor is $C = 0.43$. With a perfect reconstruction efficiency and detector resolution, the unfolding correction factor would be 1. Various factors contribute to make the correction factor less than 1 including, but not limited to, less than

perfect resolution, limited muon trigger efficiency and limited electron identification efficiency.

The error on the measured fiducial uncertainty due to a specific source of uncertainty is calculated as

$$\Delta\sigma^{fid} = \sigma_{nominal}^{fid} - \frac{(\mu_{EW} + \Delta\mu_{EW})N_{EW}^{predicted}}{(C + \Delta C)\mathcal{L}} \quad (8.19)$$

where C and μ_{EW} are the nominal unfolding correction factor and measured electroweak normalization factor, ΔC is the change in unfolding correction factor resulting in changes in N^{reco} and N^{fid} induced by varying a source of systematic or theory uncertainty, and $\Delta\mu_{EW}$ is the change in μ_{EW} resulting from the fit repeated with a nuisance parameter corresponding to an uncertainty fixed to $\hat{\theta} \pm \Delta\hat{\theta}$. In this way, the correlation between μ_{EW} , N^{reco} , and N^{fid} are accounted for in calculating the contribution of a specific uncertainty to $\Delta\mu_{EW}$.

Note that there is a statistically significant difference in the correction factor predicted by the Madgraph and Sherpa event generators. Using the Madgraph event generator, $C = 0.47$. This is likely due to differences in the modelling of jet kinematics between the two event generators, in particular, the leading jet transverse momentum and the rapidity separation between the two leading jets, which are both used to define the fiducial volume. This difference between the correction factors is included as an uncertainty on the fiducial cross section measurement.

Using Equation 8.18 and the prescription for calculating the uncertainty on the fiducial cross section described in Equation 8.19, the measured fiducial cross section is $\sigma^{fid} = 15.8 \pm 3.0$ fb. The comparison of the measurement to the Sherpa and Madgraph event generator predictions is shown in Figure 8.8. Note that while the measurement is statistically different from the Sherpa event generator prediction, the measurement is consistent with the Madgraph event generator prediction. The contribution of various sources of error to the fiducial cross section measurement are shown in Table 8.2. The measurement is dominated by statistical uncertainty on the simulated samples. Uncertainties on the pileup

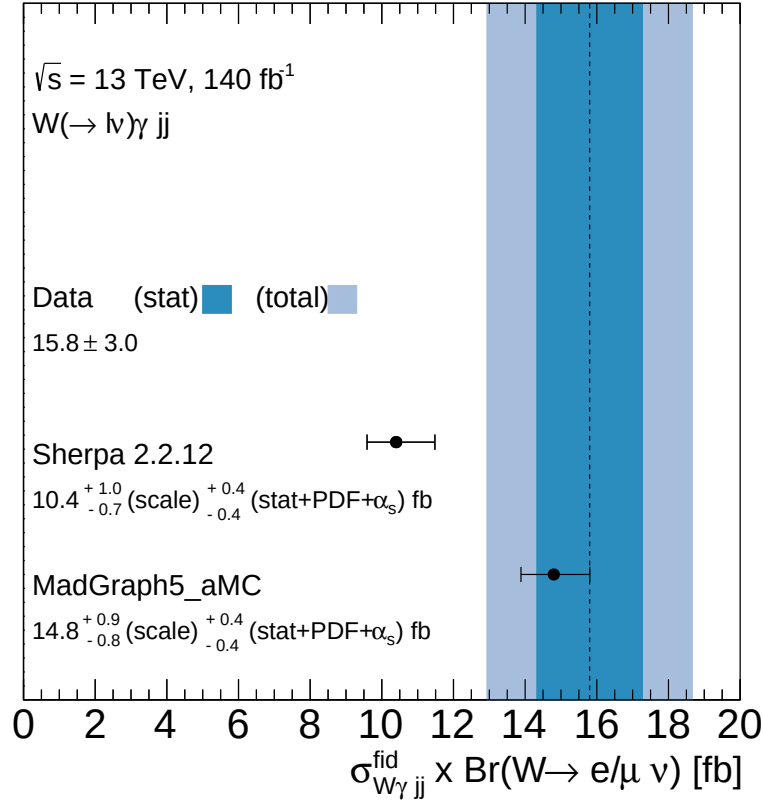


Figure 8.8: The measured fiducial cross section compared with the fiducial cross sections predicted by the Sherpa and Madgraph event generators.

reweighting, the jet energy scale and resolution, and uncertainties on the electroweak $W\gamma jj$ and QCD $W\gamma jj$ predictions also are large contributions to the total uncertainty.

The difference in total predicted cross section between Sherpa and Madgraph has been observed in other analyses of electroweak production of gauge bosons in association with jets. The predictions of different generators for $W^\pm W^\pm$ production in association with two jets was studied in detail in [46]. In the Madgraph sample used in this analysis, matrix elements are calculated for two jet events. The event generation is then interfaced to the Pythia 8 [105] parton shower simulation. The simulated events are matched with the parton shower using a method called FKS subtraction [109]. In the Sherpa signal event sample the matrix element is calculated with up to one additional jet. Events are then merged with the parton shower using the CKKW matching scheme [110]. A non-

Uncertainty Source	Fractional Uncertainty
Simulated Event Statistics	0.11
Jets	0.08
Lepton,Photon,Pileup	0.08
EW $W\gamma jj$ Model	0.07
Strong $W\gamma jj$ Model	0.06
Non-prompt Background Model	0.02
Luminosity	0.02
Other Background Model	0.02
E_T^{miss}	0.01

Table 8.2: Fractional uncertainty of various sources of uncertainty to the measured fiducial cross section.

optimal handling of the flow of color charge between the simulated events and the parton shower likely leads to the lower overall cross section. Sherpa also predicts a narrow Δy_{jj} distribution and a softer leading jet transverse momentum spectrum, both of which are used to define the phase space in which the measurement is performed.

This measurement provides a test of the validity of the Standard Model. The measurement is in agreement with the Madgraph event generator prediction. The measurement is compatible with the Sherpa event generator within two standard deviations, which likely points to incorrect modelling in the Sherpa sample.

Chapter 9

Differential Cross Section Measurement

The observation of the electroweak production of a W boson and a photon in association with two jets was presented in the previous chapter. Given this observation, it is valuable to further characterize the kinematic properties of electroweak $W\gamma jj$ events. This process will become a background to future ATLAS Run 3 analyses and it is important to understand the kinematic properties of these events as well as how well event generators simulate their characteristics. Some of the kinematic distributions of electroweak $W\gamma jj$ events are also sensitive to potential non-Standard Model contributions to the quartic $WW\gamma\gamma$ gauge coupling. In this chapter, the measurement of the differential cross section of the electroweak production of a W boson and a photon in association with two jets is presented in various kinematic variables. The measured differential cross sections are compared with the predictions from the Madgraph and Sherpa event generators to assess the quality of the modelling of this process.

9.1 Differential Extraction Methodology

The extraction of a differential cross section is not motivated by a desire to observe the process, but by the potential sensitivity to new physics. For this reason, the differential extraction is done in a more exclusive phase space than that used for the fiducial cross

section measurement. Here the dijet mass is required to be greater than 1 TeV. The measurement of electroweak $W\gamma jj$ production presented in Chapter 8 had large uncertainties from the modelling of the large QCD $W\gamma jj$ background. In order to constrain the modelling uncertainties, a different methodology is used to extract differential cross sections where the normalization and shape of QCD $W\gamma jj$ background is constrained in three control regions. The data is split into a signal region and three control regions based on the amount of hadronic activity in the rapidity gap between the two leading jets and the centrality of the lepton-photon system. The signal and control regions are summarized in Table 9.1 where M_{jj} is the dijet mass, Δy_{jj} is the rapidity gap between the two leading jets, N gapjets is the number of jets in the rapidity gap between the two leading jets and $\xi_{l\gamma}$ is the centrality of the lepton and photon system, defined as $\xi_{l\gamma} = \left| \frac{y_{l\gamma} - (y_{j_0} + y_{j_1})}{\Delta y(j_0, j_1)} \right|$ where j_0 and j_1 are the highest and second highest transverse momentum jet, respectively.

SR	CRA	CRB	CRC
$M_{jj} > 1 \text{ TeV}, \Delta y_{jj} > 2$			
N gapjets = 0 $\xi_{l\gamma} < 0.35$	N gapjets > 0 $\xi_{l\gamma} < 0.35$	N gapjets > 0 $0.35 < \xi_{l\gamma} < 1$	N gapjets = 0 $0.35 < \xi_{l\gamma} < 1$

Table 9.1: Table summarizing the criteria used to define the signal region (SR) and three control regions (CRA,CRB,CRC) used to extract differential distributions of electroweak $W\gamma jj$ events.

Similar to the one used for the measurement of the fiducial cross section described in Chapter 8, the differential cross section measurements rely on a maximum likelihood fit. In this case, the likelihood is defined as

$$L = \prod_{r,i} \text{P}(N_{ri}^{\text{data}} | \gamma_{ri} \nu_{ri}(\alpha)) \text{P} \left(\left(\frac{\nu_{ri}(\mathbf{1})}{\delta_{ri}} \right)^2 \middle| \gamma_{ri} \left(\frac{\nu_{ri}(\mathbf{1})}{\delta_{ri}} \right)^2 \right) \quad (9.1)$$

where N_{ri}^{data} is the number of data events in bin i and region r , α are the free parameters of the fit described below. The term γ_{ri} is a nuisance parameter which accounts for the

finite statistics in the simulated sample which is constrained by the Poisson constraint term

$$P \left(\left(\frac{\nu_{ri}(\mathbf{1})}{\delta_{ri}} \right)^2 \middle| \gamma_{ri} \left(\frac{\nu_{ri}(\mathbf{1})}{\delta_{ri}} \right)^2 \right) \quad (9.2)$$

where δ_{ri} is the statistical uncertainty of the simulated sample. The symbol ν_{ri} is the predicted event yield in region r and bin i and is given by

$$\nu_{ri} = \mu_i \nu_{ri}^{\text{EW}} + \nu_{ri}^{\text{QCD}} + \nu_{ri}^{\text{other}}. \quad (9.3)$$

where ν_{ri}^{EW} is the simulated prediction for the electroweak $W\gamma jj$ event yield, μ_i is the signal strength that is a free parameter of the fit, ν_{ri}^{QCD} represents the QCD $W\gamma jj$ event yield and ν_{ri}^{other} represents the predicted number of events from the other simulated and data driven estimated backgrounds. The parameter ν_{ri}^{QCD} is constrained by the free parameters in the fit $b_{L,i}$, $b_{H,i}$ and C according to

$$\begin{aligned} \nu_{\text{SR},i}^{\text{QCD}} &= b_{L,i} C \nu_{\text{SR},i}^{\text{QCD,MC}}, \\ \nu_{\text{CRA},i}^{\text{QCD}} &= b_{L,i} \nu_{\text{CRA},i}^{\text{QCD,MC}}, \\ \nu_{\text{CRB},i}^{\text{QCD}} &= b_{H,i} \nu_{\text{CRB},i}^{\text{QCD,MC}}, \\ \nu_{\text{CRC},i}^{\text{QCD}} &= b_{H,i} C \nu_{\text{CRC},i}^{\text{QCD,MC}}. \end{aligned} \quad (9.4)$$

Revisiting Equation 9.1, α can be defined as the vector of free parameters in the fit, $\alpha \equiv (\mu_1, \dots, \mu_n, b_{L,1}, \dots, b_{L,n}, b_{H,1}, \dots, b_{H,n}, C)$.

Along with the signal strength μ_i , the free parameters in the fit are: $b_{L,i}$ which can be thought of as a term to reweight the QCD background bin-by-bin in regions of low centrality (SR and CRA); $b_{H,i}$ which is a similar reweighting for high centrality (CRB and CRC), and a correction factor C . The bin-by-bin reweighting factors are primarily constrained in the QCD background-rich regions with enhanced hadronic activity in the gap between the leading jets (CRA and CRB). The correction factor is included to account

for the extrapolation from high to low hadronic activity. A constant correction factor is chosen.

There is a tradeoff between statistical and systematic uncertainties when adding additional control regions to the likelihood fit. With only one control region, there are not enough degrees of freedom to include the correction factor C , resulting in larger QCD theory uncertainties, but this comes with smaller statistical uncertainties on a smaller number of free parameters constrained in the fit. A choice is also made in the use of CRA to constrain the bin-by-bin reweighting factors used in the signal region and the use of a correction factor to account for uncertainties in the extrapolation across the N-gapjets boundary. The bin-by-bin reweighting factors could just as well be shared by the signal region and CRC and C be used to cross the $\xi_{l\gamma}$ boundary.

Unlike the likelihood fit described in Section 8.1, there are no Gaussian constraint terms parameterizing the systematic uncertainties affecting the measurement here. The same set of experimental and theory uncertainties as those described in Chapter 7 are taken into account, but their treatment in the extraction of differential cross section measurements is different than the approach used to measure the fiducial cross section. This is done because with the smaller statistics available in each control region, it is important to gauge the statistical significance of the systematic uncertainties.

9.2 Treatment of Uncertainties

9.2.1 Experimental Uncertainties

Each experimental uncertainty discussed in Chapter 7 is treated using the so-called *bootstrapping* procedure. The bootstrapping procedure is used to separate systematic effects from effects arising from the limited statistics available to evaluate systematic variations. In this procedure, $N_{bootstrap} = 10000$ pseudo-experiments are performed where each pseudo-experiment is the fit with the simulated predictions varied by the systematic. Concretely, this means creating a nominal prediction and a different systematically varied predic-

tion by multiplying the simulated event weights by a random number sampled from a Poisson distribution to simulate statistical fluctuations around the systematically shifted predicted event yields. The random number seed is the same for both the nominal and the varied template, and thus preserves the correlation between nominal and varied templates. If this correlation were not preserved, the difference between the nominal and systematic bootstrap samples would also reflect statistical fluctuations between the two. The fit described in Section 9.1 is performed on each pseudo-experiment and the relative experimental uncertainty is defined as

$$R_{\text{exp}}^{\text{sys}} = \frac{\langle N_{\text{EW},i}^{\text{Nom}} - N_{\text{EW},i}^{\text{exp}} \rangle}{\langle N_{\text{EW},i}^{\text{Nom}} \rangle}. \quad (9.5)$$

where $\langle \rangle$ denotes the average over 10000 pseudo-experiments and $N_{\text{EW},i}^{\text{Nom}}$ and $N_{\text{EW},i}^{\text{exp}}$ are the fitted number of electroweak signal events from pseudo-experiment i for the nominal and systematically varied prediction, respectively. The statistical uncertainties on the systematic uncertainty are then defined by taking the RMS of the distribution of the difference between the nominal and varied pseudo-experiments. The statistical uncertainty on the systematic uncertainty provides information regarding how likely it is that the difference between the nominal yield and the yield under a systematic variation is only due to statistics. If a systematic uncertainty is deemed to be statistically insignificant, i.e. below some σ_{thresh} , then it is justified to ignore the systematic uncertainty. Multiple σ_{thresh} were considered, but ultimately all systematics are kept. The systematic variations that are statistically insignificant are a small contribution to the final uncertainty.

9.2.2 Theory Uncertainties

The theory uncertainties are treated as described in Chapter 7. The scale uncertainties and the uncertainty derived from the difference between the QCD $W\gamma jj$ prediction of the Sherpa and Madgraph event generator are evaluated using the bootstrap procedure. The

uncertainty for the variation of the factorization and renormalization scale is estimated as

$$R_{(\mu_F, \mu_R)}^{\text{sys}} = \frac{\langle N_{EW,i}^{\text{Nom}} - N_{EW,i}^{(\mu_F, \mu_R)} \rangle}{\langle N_{EW,i}^{\text{Nom}} \rangle}. \quad (9.6)$$

where $N_{EW,i}^{\text{Nom}}$ is the fitted number of electroweak signal events obtained when using the predictions without scale variations for bootstrap i and $N_{EW,i}^{(\mu_F, \mu_R)}$ is the fitted number of electroweak signal events obtained when using a prediction with the scale variation (μ_R, μ_F) . The resulting scale uncertainties are then combined in an envelope as described in Chapter 7. The uncertainty due to the different predictions for the QCD $W\gamma jj$ events is evaluated by performing the fit twice for each bootstrap sample with a different generator (Sherpa or Madgraph) used to create the QCD $W\gamma jj$ prediction. The uncertainty is then defined as

$$R_{\text{Gen Choice}}^{\text{sys}} = \frac{\langle N_{EW,i}^{\text{Nom, Sherpa}} - N_{EW,i}^{\text{Nom, MG5}} \rangle}{\langle N_{EW,i}^{\text{Nom, Sherpa}} \rangle}. \quad (9.7)$$

where $N_{EW}^{\text{Nom, Sherpa}}$ is the fitted number of electroweak signal events where the Sherpa QCD $W\gamma jj$ prediction is used and $N_{EW}^{\text{Nom, MG5}}$ is the fitted number of electroweak signal events obtained when the Madgraph QCD $W\gamma jj$ prediction is used. As there is no correlation between the Madgraph and Sherpa predictions, the bootstraps used to evaluate the systematic uncertainty arising from the choice of event generator are uncorrelated.

9.2.3 Statistical uncertainties

Statistical uncertainties on the data and the simulated event samples are evaluated by performing the fit for 10000 pseudo-experiments where the data or simulated events are fluctuated by a random value sampled from a Poisson distribution with a mean value taken as the number of observed or predicted events. The standard deviation of the resulting distribution of fitted number of electroweak signal events divided by the nominal fitted electroweak number of electroweak signal events is taken as the statistical uncertainty on the data or the simulated samples.

9.2.4 Likelihood Fit Results

The likelihood fit described above is performed for the distribution of six kinematic variables: M_{jj} , p_T^{jj} , p_T^l , $M_{l\gamma}$, $\Delta\phi(jj)$ and $\Delta\phi(l\gamma)$. The variables $\Delta\phi(jj)$ and $\Delta\phi(l\gamma)$ are defined through the rapidity ordering of the jets, i.e. the azimuthal angle of the jet with subleading rapidity is subtracted from the azimuthal angle of the jet with leading rapidity. The first four variables are chosen as they characterize the process of interest and are expected to be sensitive to dimension-8 effective field theory operators that give rise to anomalous values of the $WW\gamma\gamma$ quartic gauge coupling. The signed azimuthal jet and $l\gamma$ separation is expected to be sensitive to operators that violate charge-parity conservation.

The fitted distributions in the signal region with uncertainties calculated as described above resulting from a fit to data are shown in Figure 9.1.

9.3 Unfolding

The measured kinematic distributions differ from the *true* distribution of the electroweak production of $W\gamma$ in association with two jets due to less than perfect detector resolution and reconstruction inefficiencies, as described in Chapter 8. In this section the *unfolding* of the measured distributions to true event yield distributions is described.

9.3.1 Iterative Bayesian Unfolding

The challenge of unfolding a differential measurement is in inferring a true distribution from a measured distribution that is affected by detector efficiency and resolution. Let $\mathbf{x}$ be the true distribution of a kinematic variable and $\mathbf{y}$ be the measured distribution. Each entry corresponds to the number of events observed in a given bin. The true and reconstructed distributions are related through a *response matrix* $\mathbf{A}$,

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} \tag{9.8}$$

Each element of the response matrix is the conditional probability that a reconstructed value y falls into bin i given the true value x in a bin j , i.e.

$$A_{ij} = P(y \in i | x \in j) \quad (9.9)$$

The response matrix can also be called a *folding matrix* as it encodes the detector resolution and efficiency effects that transform a true distribution into a reconstructed distribution.

Similarly one can define an *unfolding matrix* that transforms the measured distribution y into the truth distribution x . The entries of the unfolding matrix represent the conditional probability for a true event x to fall in a bin j given that a reconstructed event y falls in bin i , i.e.

$$\tilde{A}_{ij} = P(x \in j | y \in i) \quad (9.10)$$

The unfolding matrix, $\tilde{A}$, is likely not invertible and even if it were, a simple matrix inversion may not correctly handle statistical fluctuations. A less naive method is needed to approach the unfolding problem.

The method outlined used in this analysis is an *Iterative Bayesian Unfolding* and is described in detail in [111]. The interested reader is referred there for a more complete description of the unfolding method, including the estimation of the uncertainties as well as the application of the unfolding method to multidimensional problems.

As the name suggests, the method relies on Bayes' theorem

$$P(x|y) = \frac{P(y|x)P(x)}{\sum_{l=1}^{n_x} P(y|x)P(x)} \quad (9.11)$$

where x_i denotes a cause and y denotes an effect. The quantity $P(x)$ is the initial probability of cause x , or the *Bayesian prior* of cause x , and $P(y|x)$ is the conditional probability, or probability of y given x . The number of possible causes is given by n_x . Addressing the problem of unfolding measured distributions in the language of Bayes' theorem, an effect is an observed event in a given bin of a reconstructed observable, and the cause is

the true bin in which the event belongs. With a perfect detector and modelling the bin in which an event belongs at the truth and reconstructed level would be the same.

With Bayes' theorem, Equation 9.10 can be rewritten

$$\tilde{A}_{ij} = \frac{P(y \in i | x \in j) P(x \in j)}{P(y \in j)} \quad (9.12)$$

and making use of Equation 9.9

$$\tilde{A}_{ij} = \frac{A_{ij} P(x \in j)}{\sum_j^{N_{bins}} A_{ij} P(x \in j)} \quad (9.13)$$

where $P(x \in j)$ is the truth distribution that is of interest. Written in this form, an initial guess can be taken for the truth distribution, giving an initial value for $\tilde{A}_{ij}$. The initial guess of the unfolding matrix can then be used to estimate the true distribution as

$$\hat{x}_j = \frac{1}{\epsilon_j} \sum_i^{N_{bins}} \tilde{A}_{ij} y_i \quad (9.14)$$

where y_i is the measured data and ϵ_j is the reconstruction efficiency in bin j . Simulated signal event samples are used to construct the unfolding matrix and to calculate ϵ_j . It is logical to take the initial guess of $P(x \in j)$ from the simulated signal event samples.

From the first guess, $\mathbf{x}^{k=0}$, the iterative Bayesian unfolding algorithm proceeds as follows:

1. Estimate the unfolding matrix $\tilde{A}_{ij}^k$ using Equation 9.13 where simulated events are used to construct the folding matrix A_{ij} .
2. Estimate $\hat{x}^{k+1}$ using Equation 9.14 where ϵ_j is taken from simulated signal events.
3. Update $\tilde{A}^{k+1}$ with the new value of $P(x \in j) = \frac{\hat{x}_j}{\sum_j^{N_{bins}} \hat{x}_j}$.
4. Continue until the χ^2 between $\hat{x}^k$ and $\hat{x}^{k-1}$ is small.

The number of iterations for unfolding must be chosen before analysis unblinding. Two iterations are chosen as this was shown to minimize the statistical uncertainty on the unfolded distribution as well as the bias. In general more iterations leads to less bias and a higher statistical uncertainty. As statistical uncertainties are the largest uncertainties on the extracted electroweak yield, fewer iterations is preferred.

9.3.2 Unfolded Results and Treatment of Uncertainties

The unfolding technique described above is performed on the six differential event yield distributions extracted. The response matrices of the unfolded variables are shown in Figure 9.2. The degree to which the response matrix is diagonal reflects the choice of binning and the resolution of the observable which is unfolded.

To evaluate the uncertainty on the unfolded number of events, the pseudo-experiments used in calculating the systematic, theory, and statistical errors are also unfolded and the unfolded distributions are used to construct the uncertainties post-unfolding the same way the pseudo-experiments are used to calculate the errors on the extracted distributions. To extract a systematic due to the choice of simulation for signal events, the same signal prediction is used to construct the response matrix and calculate efficiencies as is used as the signal prediction in the differential signal extraction. The largest uncertainties on the differential measurement vary based on the distribution measured. In Figure 9.3 the fractional uncertainty of different uncertainty sources on the measurement of M_{jj} are shown. Consistent with the fiducial cross section measurement, statistical uncertainties, jet energy scale and resolution uncertainties, and electroweak and QCD $W\gamma jj$ modelling uncertainties are the dominant uncertainties.

The unfolded event yield distributions are divided by the total integrated luminosity of the dataset to obtain unfolded differential cross sections. The unfolded differential cross sections are shown with all uncertainties in Figure 9.4. Each differential cross section is shown along with the predictions from the Sherpa and Madgraph event generators. The measured cross sections fall for the most part between the Sherpa and Madgraph

predictions. Consistent with what is observed in the fiducial cross section measurement, Sherpa falls systematically below the measured differential cross sections but is consistent with measurements within two standard deviations. As well as a difference in normalization, it can be seen that the jet kinematic variables are modelled differently between Sherpa and Madgraph. The differences between Sherpa and Madgraph were discussed in Chapter 8.

The differential measurements presented here provide another test of the validity of the Standard Model and provide valuable inputs to understanding the accuracy of different event generators in describing the process of interest. The measured differential cross sections are consistent with Standard Model predictions and will be used in the future to set limits on the size of potential anomalous quartic gauge couplings.

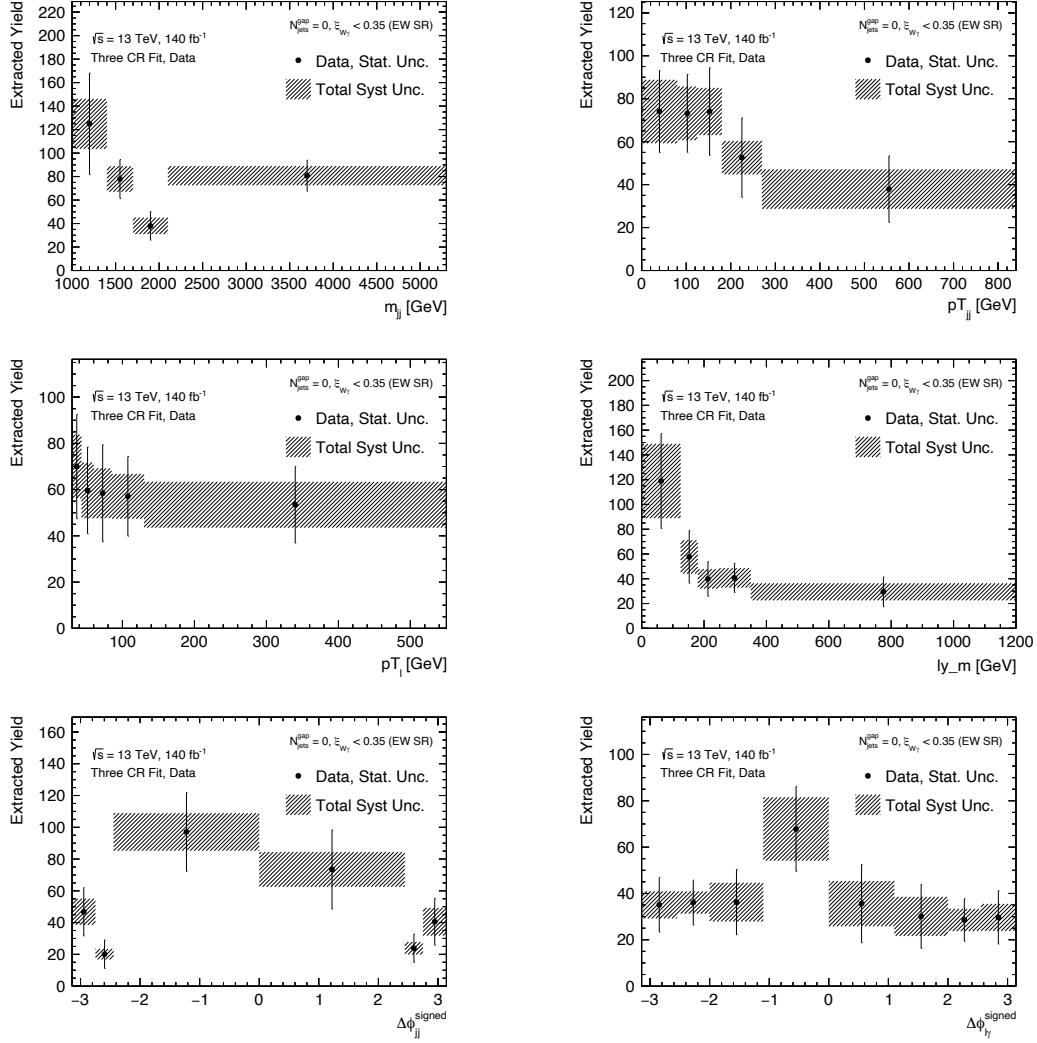


Figure 9.1: The electroweak $W\gamma jj$ event yield in the signal region extracted from a fit to data, along with the total uncertainty on the measurements.

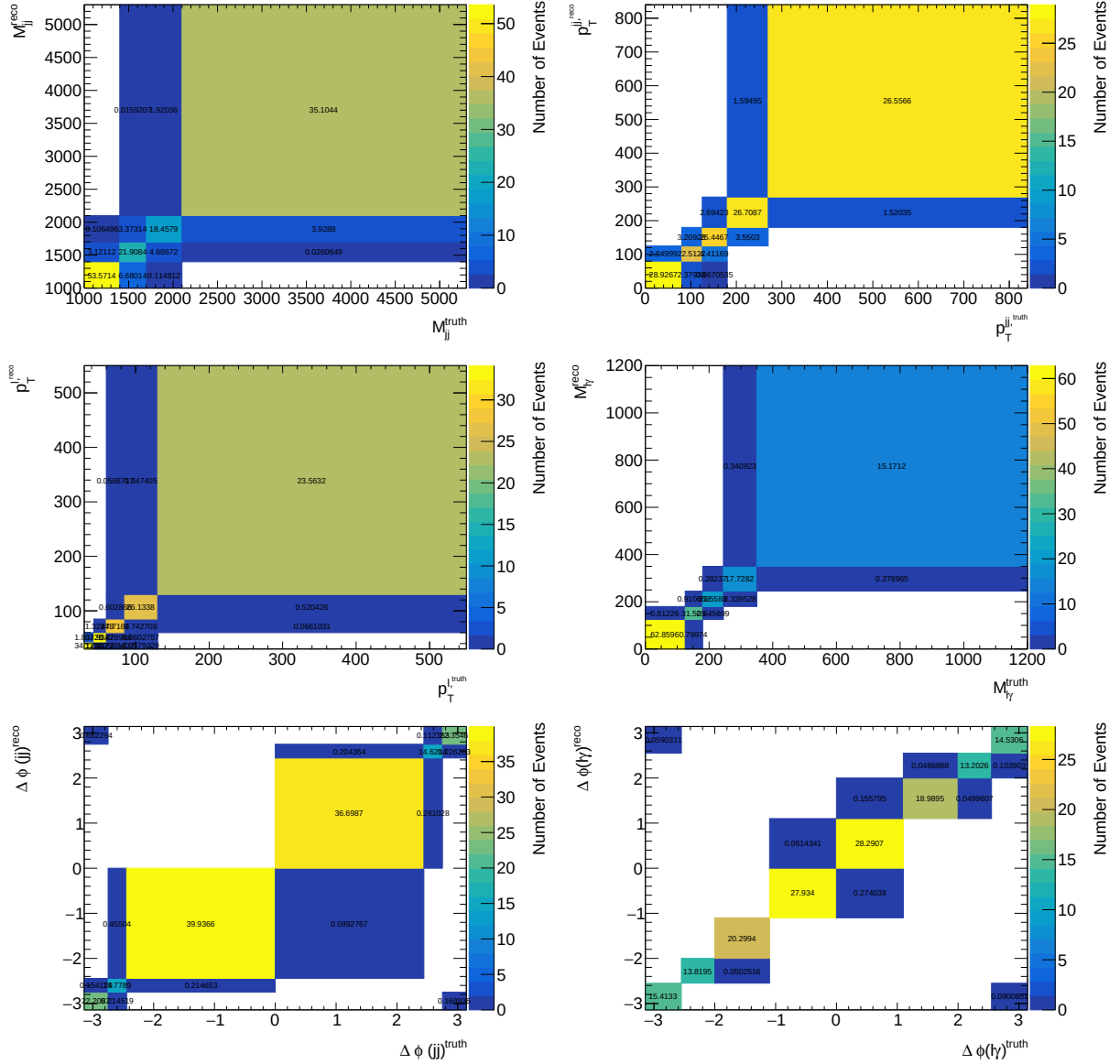


Figure 9.2: The response matrices for each unfolded variable. The response matrix is constructed from signal events simulated with the Sherpa event generator.

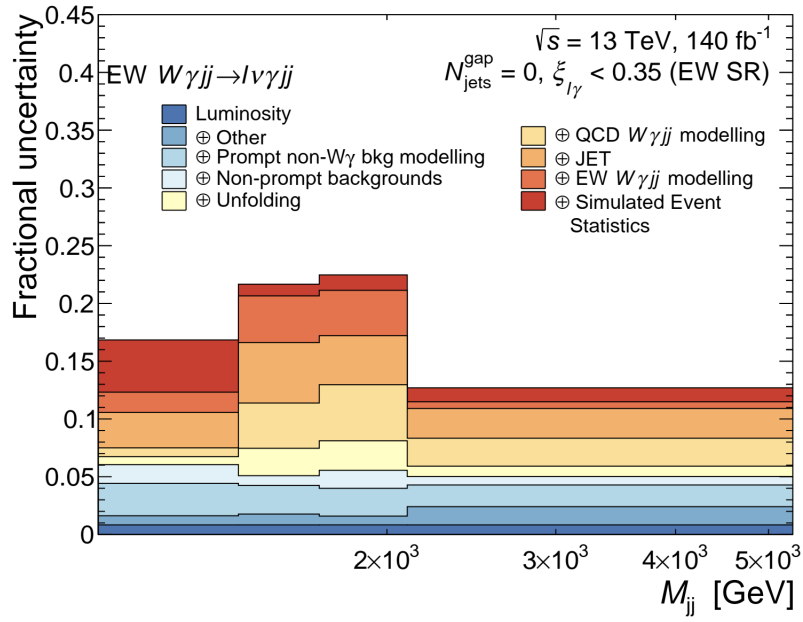


Figure 9.3: The fractional uncertainty from various sources of uncertainty on the unfolded M_{jj} distribution.

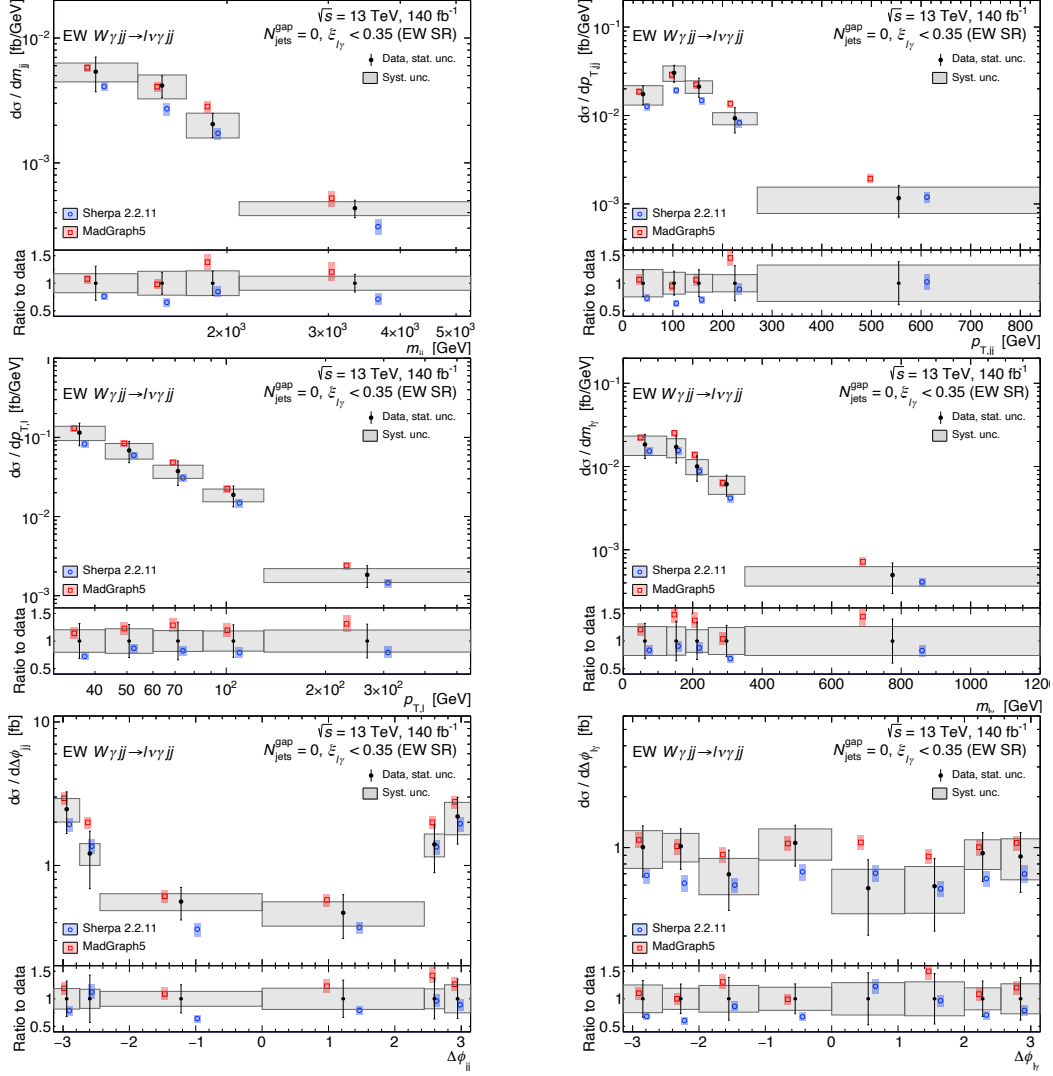


Figure 9.4: Electroweak $W\gamma jj$ unfolded differential cross section distributions as a function of kinematic variables M_{jj} , p_T^{jj} , $\Delta\phi_{jj}$, p_T^l , $\Delta\phi_{l\gamma}$ and $M_{l\gamma}$ along with the predictions from the Madgraph and Sherpa event generators. The measured cross section includes all uncertainties described in Chapter 7 while the error bars on the theoretical predictions are determined from scale, PDF, α_s and statistical uncertainties of the simulated event samples.

Chapter 10

Discussion and Conclusion

This thesis presents the analysis of 140 fb^{-1} of proton-proton collision data taken by the ATLAS detector between 2015 and 2018 in search of the electroweak production of a W boson and a photon in association with two jets. This analysis constitutes the first observation of this process using ATLAS data with an observed(expected) statistical significance of $9.0(6.3)\sigma$. A fiducial cross section of $15.8 \pm 3.0 \text{ fb}$ is measured and differential cross sections are measured in 6 kinematic variables: $M_{jj}, p_T^{jj}, \Delta\phi(jj), M_{l\gamma}, p_T^l$ and $\Delta\phi(l\gamma)$. The measurements are found to be statistically consistent with Standard Model predictions.

Prompt backgrounds in the analysis consist of QCD $W\gamma jj$ production, which is the largest background in this analysis, electroweak and QCD $Z\gamma jj$ production, and events with a top quark and a photon. The prompt backgrounds are estimated using simulated samples. Non-prompt backgrounds arise from jets misreconstructed as photons, electrons misreconstructed as photons, leptons misreconstructed as jets, and events where the photon comes from a different proton-proton collision. Jets misreconstructed as photons constitute the largest non-prompt background in this analysis. A template fit method is used to estimate the number of jets misreconstructed as photons. The electron misreconstructed as photon background is estimated using a tag and probe method, the jet misreconstructed as photon background is estimated using the fake factor method, and

the background arising from photons from other proton-proton collision is estimated by calculating a pileup fraction in the tails of the $\Delta z(\gamma, PV)$ distribution of events with converted photons.

In order to make an observation, a neural network is trained to distinguish simulated signal events from simulated prompt background events. This results in a distribution which is very signal pure at high values of neural network output. This is the first use of machine learning to study this process and results in greater sensitivity than previous studies of this process [43, 45].

A binned likelihood fit of the neural network output is performed to extract a statistical significance of observation. The fit uses a control region defined by additional hadronic activity in the rapidity gap between the two leading jets to constrain the QCD $W\gamma jj$ background and a signal region defined by no additional hadronic activity in the rapidity gap between the two leading jets to constrain the normalization of the signal process. Nuisance parameters are included in the fit to account for experimental uncertainty, uncertainty on the data driven background estimates, statistical uncertainty on the simulated backgrounds and uncertainties arising from the modelling of prompt processes. The profile likelihood ratio statistic is used to reject the background only hypothesis.

The fitted number of signal events is used to calculate a fiducial cross section. The fiducial region is designed to be as close as possible to the measured phase space to avoid extrapolating over regions that have not been probed by the experiment. A photon isolation is included in the fiducial volume to limit model dependence. The measured cross section is 15.8 ± 3.0 fb. This is consistent with the Madgraph prediction of 14.8 ± 1.0 fb and within two standard deviations of the Sherpa prediction of 10.4 ± 1.1 fb. The largest source of uncertainties are the statistical uncertainty, uncertainties on the jet calibration and uncertainties in QCD $W\gamma jj$ and electroweak $W\gamma jj$ modelling.

Differential cross sections are measured in M_{jj} , p_T^{jj} , $\Delta\phi(jj)$, $M_{l\gamma}$, p_T^l and $\Delta\phi(l\gamma)$. The variables M_{jj} , p_T^{jj} , $M_{l\gamma}$ and p_T^{jj} are chosen as they characterize the process of interest and are expected to be sensitive to dimension 8 effective field theory operators that give rise to

anomalous values of the $WW\gamma\gamma$ quartic gauge coupling. The variables $\Delta\phi(jj)$ and $\Delta\phi(l\gamma)$ are defined through the rapidity ordering of the jets and are expected to be sensitive to operators that violate charge-parity conservation.

The differential cross sections are measured using a fit with three control regions used to constrain the shape and normalization of the QCD $W\gamma jj$ background. The extracted yields are unfolded to a truth number of events in the fiducial volume using iterative Bayesian unfolding. The size of the systematic, modelling, and statistical uncertainties are estimated using the *bootstrapping* procedure. The measured differential cross sections agree with the differential cross sections predicted by the Madgraph event generator in most bins and are above the cross section predicted by the Sherpa event generator but within two standard deviations in most bins. The uncertainty on the differential cross sections range from 12% to 35%. The dominant source of uncertainty on the differential cross section are statistical uncertainties, uncertainties on the jet calibration and uncertainties in the modelling of electroweak $W\gamma jj$ and QCD $W\gamma jj$ production.

Both the fiducial cross section measurement and differential cross section measurement are statistically consistent with the Standard Model; however both show disagreement with predictions from the Sherpa event generator at the level of two standard deviations. Sherpa predicts a lower total cross section than Madgraph. Sherpa predicts slightly different shapes of kinematic variables related to jets, specifically the transverse momentum of the leading jet and the rapidity difference between the two leading jets. This is likely due to how Sherpa handles the matching of generated events and the simulated parton shower.

At the time of writing this thesis, ATLAS Run 3 is ongoing and is to be followed by the High Luminosity LHC. The larger datasets collected in Run 3 and at the HL-LHC will open the door to measurements of even rarer processes and more precise measurements of processes that have already been observed. This measurement is statistically limited. There is a large statistical uncertainty on the dominant non-prompt background, and limited statistics in control regions limit the ability to constrain the large QCD $W\gamma jj$

background. The larger datasets to be collected in years to come will provide much more statistical power to constrain these dominant backgrounds.

Modelling uncertainties also constitute a large uncertainty on the fiducial and differential cross sections here. Differences in prediction between event generators for both the electroweak production of $W\gamma jj$ and QCD $W\gamma jj$ are observed and both generators come with large scale uncertainties. This analysis would greatly benefit from the smaller scale uncertainties that would come with a next to leading order in QCD prediction of the signal process.

As the measurement agrees with Standard Model predictions, there is no obvious hint of anomalous quartic gauge couplings. Nevertheless, the measurements included in this thesis will be used in the future to set limits on dimension 8 effective field theory operator couplings. This measurement joins a host of other recent ATLAS and CMS measurements that are sensitive to quartic electroweak couplings. These measurements open the door to combined limit setting on dimension 8 effective field theory couplings which will provide invaluable insights in possible new physics phenomena at energy scales beyond those probed at the LHC.

Bibliography

- [1] C. Burgess and G. Moore. *The Standard Model: A Primer*. Cambridge University Press, 2006.
- [2] R. L. Workman *et. al.* Review of Particle Physics. *PTEP*, 2022:083C01, 2022.
- [3] M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.
- [4] The ATLAS Collaboration. Determination of the strong coupling constant α_s in multijet production with the ATLAS detector at the LHC. Technical Report ATL-PHYS-PROC-2018-004, CERN, Geneva, 2018.
- [5] The NNPDF Collaboration. Parton distributions for the LHC run II. *JHEP*, 2015(4), Apr 2015.
- [6] Q. R. Ahmad *et. al.* Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, Jun 2002.
- [7] C. Wetterich. Fine-tuning problem and the renormalization group. *Physics Letters B*, 140(3):215–222, 1984.
- [8] O. J. P. Éboli and M. C. Gonzalez-Garcia. Classifying the bosonic quartic couplings. *Phys. Rev. D*, 93(YITP-SB-16-09):093013, 2016.

- [9] Electroweak measurements in electron–positron collisions at W-boson-pair energies at LEP. *Physics Reports*, 532(4):119–244, Nov 2013.
- [10] H. Schellman A. V. Kotwal and J. Sekaric. Review of physics results from the Tevatron: Electroweak physics. *International Journal of Modern Physics A*, 30(06):1541004, Feb 2015.
- [11] The ATLAS Collaboration. Measurement of isolated-photon pair production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 2013(1), Jan 2013.
- [12] The CMS Collaboration. Measurement of differential cross sections for the production of a pair of isolated photons in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 74(11), Nov 2014.
- [13] The ATLAS Collaboration. Measurements of $W\gamma$ and $Z\gamma$ production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC. *Phys. Rev. D*, 87(11), Jun 2013.
- [14] The CMS Collaboration. Measurement of the $W\gamma$ and $Z\gamma$ inclusive cross sections in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous triple gauge boson couplings. *Phys. Rev. D*, 89(9), May 2014.
- [15] The ATLAS Collaboration. Measurements of $Z\gamma$ and $Z\gamma\gamma$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. D*, 93(11), Jun 2016.
- [16] The CMS Collaboration. Measurement of the $Z\gamma \rightarrow \nu\bar{\nu}\gamma$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV and limits on anomalous $ZZ\gamma$ and $Z\gamma\gamma$ trilinear gauge boson couplings. *Phys. Let. B*, 760:448–468, Sep 2016.
- [17] The CMS Collaboration. Measurement of the W^+W^- cross section in pp collisions at $\sqrt{s} = 8$ TeV and limits on anomalous gauge couplings. *Eur. Phys. J. C*, 76(7), Jul 2016.
- [18] The ATLAS Collaboration. Measurement of total and differential W^+W^- production cross sections in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS de-

- detector and limits on anomalous triple-gauge-boson couplings. *JHEP*, 2016(9), Sep 2016.
- [19] The ATLAS Collaboration. Measurement of the WW+WZ cross section and limits on anomalous triple gauge couplings using final states with one lepton, missing transverse momentum, and two jets with the ATLAS detector at $\sqrt{s} = 7$ TeV. *JHEP*, 2015(1), Jan 2015.
- [20] The CMS Collaboration. Measurement of the sum of WW and WZ production with W+dijet events in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 73(2), Feb 2013.
- [21] The ATLAS Collaboration. Measurements of $W^\pm Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector and limits on anomalous gauge boson self-couplings. *Phys. Rev. D*, 93(9), May 2016.
- [22] The CMS Collaboration. Measurement of the WZ production cross section in pp collisions at $\sqrt{s} = 7$ and 8 TeV and search for anomalous triple gauge couplings at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, 77(4), Apr 2017.
- [23] The CMS Collaboration. Measurement of WZ and ZZ production in pp collisions at $\sqrt{s} = 8$ TeV in final states with b-tagged jets. *Eur. Phys. J. C*, 74(8), Aug 2014.
- [24] The ATLAS Collaboration. Measurement of the ZZ production cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV using the $ZZ \rightarrow l^- l^+ l'^- l'^+$ and $ZZ \rightarrow l^- l^+ \nu \bar{\nu}$ decay channels with the ATLAS detector. *JHEP*, 2017(1), Jan 2017.
- [25] The ATLAS Collaboration. Evidence of $W\gamma\gamma$ Production in pp Collisions at $\sqrt{s} = 8$ TeV and Limits on Anomalous Quartic Gauge Couplings with the ATLAS Detector. *Phys. Rev. Lett.*, 115(3), Jul 2015.
- [26] The CMS Collaboration. Search for $WW\gamma$ and $WZ\gamma$ production and constraints on anomalous quartic gauge couplings in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev. D*, 90(3), Aug 2014.

- [27] The ATLAS Collaboration. Search for triboson $W^+W^+W^-$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 77(3), Mar 2017.
- [28] The ATLAS Collaboration. Observation of WWW Production in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. *Phys. Rev. Lett.*, 129:061803, Aug 2022.
- [29] The CMS Collaboration. Measurement of electroweak production of a W boson and two forward jets in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 2016(11), Nov 2016.
- [30] The ATLAS Collaboration. Measurement of the electroweak production of dijets in association with a Z boson and distributions sensitive to vector boson fusion in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *JHEP*, 2014(4), Apr 2014.
- [31] The CMS Collaboration. Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, 75(2), Feb 2015.
- [32] The ATLAS Collaboration. Observation of Electroweak Production of a Same-Sign W Boson Pair in Association with Two Jets in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. *Phys. Rev. Lett.*, 123:161801, Oct 2019.
- [33] The CMS Collaboration. Observation of Electroweak Production of a Same-Sign W Boson Pair in Association with Two Jets in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. *Phys. Rev. Lett.*, 123:161801, Oct 2019.
- [34] The CMS Collaboration. Observation of Electroweak Production of Same-Sign W Boson Pairs in the Two Jet and Two Same-Sign Lepton Final State in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 120:081801, Feb 2018.

- [35] The ATLAS Collaboration. Observation of electroweak $W^\pm Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Let. B*, 793:469–492, Jun 2019.
- [36] The ATLAS Collaboration. Observation of electroweak production of two jets and a Z boson pair. *Nat. Phys.*, 19, Feb. 2023.
- [37] The CMS Collaboration. Evidence for electroweak production of four charged leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Let. B*, 812:135992, Jan 2021.
- [38] Measurement of the cross-section of the electroweak production of a $Z\gamma$ pair in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2021-038, CERN, Geneva, 2021.
- [39] The CMS Collaboration. Measurement of the electroweak production of the electroweak production of $Z\gamma$ and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on anomalous quartic gauge couplings. *Phys. Rev. D*, 104(7), Oct 2021.
- [40] The ATLAS Collaboration. Measurement of electroweak $Z(\nu\bar{\nu})\gamma jj$ production and limits on anomalous quartic gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 06:082, 2023.
- [41] The ATLAS Collaboration. Observation of photon-induced W^+W^- production in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Phys. Let. B*, 816:136190, May 2021.
- [42] The CMS Collaboration. Search for high-mass exclusive $\gamma\gamma \rightarrow WW$ and $\gamma\gamma \rightarrow ZZ$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 07:229, 2023.
- [43] The CMS Collaboration. Observation of electroweak production of $W\gamma$ with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Let. B*, 811:135988, Dec 2020.

- [44] The CMS Collaboration. Measurement of electroweak-induced production of a $W\gamma$ with two jets in pp collisions at $\sqrt{s} = 8$ TeV and constraints on anomalous quartic gauge couplings. *JHEP*, 2017(6), Jun 2017.
- [45] CMS Collaboration. Measurement of the electroweak production of $W\gamma$ in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. arXiv 2212.12592, 2022.
- [46] The ATLAS Collaboration. Modelling of the vector boson scattering process $pp \rightarrow W^\pm W^\pm jj$ in Monte Carlo generators in ATLAS. Technical Report ATL-PHYS-PUB-2019-004, 2019.
- [47] D. B. Franzosi *et. al.* Vector boson scattering processes: Status and prospects. *Reviews in Physics*, 8:100071, Jun 2022.
- [48] C. Lefèvre. The CERN accelerator complex. Complexe des accélérateurs du CERN. 2008.
- [49] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008.
- [50] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3(08):S08003–S08003, Aug 2008.
- [51] I. Neutelings. CMS coordinate system, Dec 2022.
- [52] The ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1*. Technical design report. ATLAS. CERN, Geneva, 1997.
- [53] The ATLAS Collaboration. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013, ATLAS-TDR-19, 2010.
- [54] The ATLAS Collaboration. *ATLAS pixel detector: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1998.

- [55] The ATLAS Collaboration. Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-018, CERN, Geneva, 2015.
- [56] V. L. Ginzburg and I. M. Frank. Radiation of a uniformly moving electron due to its transition from one medium into another. *J. Phys. (USSR)*, 9:353–362, 1945.
- [57] The ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data. *JINST*, 14(CERN-EP-2018-296):P03017, 2019.
- [58] The ATLAS Collaboration. Performance of the ATLAS electromagnetic calorimeter under beam tests. *Nucl. Instrum. and Meth. in Phys. Res.*, 518(1):31–35, 2004.
- [59] The ATLAS Collaboration. Testbeam studies of production modules of the ATLAS tile calorimeter. *Nucl. Inst. Meth. A*, 606(ATL-TILECAL-PUB-2009-002, ATL-COM-TILECAL-2009-004):362–394, 2009.
- [60] *ATLAS muon spectrometer: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1997.
- [61] The ATLAS Collaboration. Measurement of muon momentum resolution of the ATLAS detector. *EPJ Web Conf.*, 28:12039, 2012.
- [62] The ATLAS collaboration. Resolution of the ATLAS muon spectrometer monitored drift tubes in LHC Run 2. *JINST*, 14(09):P09011–P09011, Sep 2019.
- [63] The ATLAS Collaboration. The LUCID-2 detector. *Nucl. Instrum. Meth. A*, 936(ATL-FWD-PROC-2018-001):152–153, 2019.
- [64] The ATLAS Collaboration. The ATLAS beam conditions monitor. *JINST*, 3:P02004, 2008.

- [65] P. Grafström and W. Kozanecki. Luminosity determination at proton colliders. *Progress in Particle and Nuclear Physics*, 81:97–148, 2015.
- [66] The ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. (CERN-EP-2022-281), 12 2022.
- [67] S. Agostinelli *et. al.* GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.
- [68] The ATLAS Collaboration. ATLAS data quality operations and performance for 2015–2018 data-taking. *JINST*, 15(04):P04003–P04003, Apr 2020.
- [69] The ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data. *JINST*, 14(CERN-EP-2019-145):P12006, 2019.
- [70] J. Pequeno and P. Schaffner. How ATLAS detects particles: diagram of particle paths in the detector. 2013.
- [71] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth. A*, 262(2):444–450, 1987.
- [72] The ATLAS Collaboration. ATLAS strategy for primary vertex reconstruction during Run-2 of the LHC. Technical Report ATL-SOFT-PROC-2015-016, CERN, Geneva, 2015.
- [73] The ATLAS Collaboration. Muon Isolation Studies on Full Run 2 Dataset. Technical Report ATL-COM-PHYS-2019-177, CERN, Geneva, 2019.
- [74] The ATLAS Collaboration. ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2. Technical Report ATL-PHYS-PROC-2017-045, CERN, Geneva, 2017.

- [75] M. Cacciari *et. al.* The anti- k_t jet clustering algorithm. *JHEP*, 2008(04):063–063, Apr 2008.
- [76] The ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C*, 77(CERN-EP-2017-024):466, 2017.
- [77] The ATLAS Collaboration. Performance of the ATLAS muon triggers in Run 2. Technical Report ATL-COM-DAQ-2019-154, CERN, Geneva, 2019.
- [78] The ATLAS Collaboration. Performance of electron and photon triggers in ATLAS during LHC Run 2. *Eur. Phys. J. C*, 80(1):47, 2020.
- [79] The ATLAS Collaboration. Tagging and suppression of pileup jets with the ATLAS detector. Technical Report ATLAS-CONF-2014-018, CERN, Geneva, 2014.
- [80] The ATLAS Collaboration. Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run. Technical Report ATL-PHYS-PUB-2017-013, CERN, Geneva, 2017.
- [81] M. Oreglia. A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi$. Other thesis, Dec 1980.
- [82] H. Ikeda *et. al.* A detailed test of the CsI(Tl) calorimeter for BELLE with photon beams of energy between 20-MeV and 5.4-GeV. *Nucl. Instrum. Meth. A*, 441:401–426, 2000.
- [83] S. Bernstein. Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités. *Comm. Soc. Math. Kharkov*, 13:1–2, 1912.
- [84] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [85] D. E. Rumelhart *et. al.* Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

- [86] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. arXiv 1412.6980, 2014.
- [87] J. Snoek *et. al.* Practical Bayesian Optimization of Machine Learning Algorithms. arXiv 1206.2944, 2012.
- [88] Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization.
- [89] X. Glorot *et. al.* Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, Apr 2011. PMLR.
- [90] Andrew L. Maas. Rectifier nonlinearities improve neural network acoustic models. 2013.
- [91] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15(IPTI-381-71):438–450, 1972.
- [92] L. N. Lipatov. The parton model and perturbation theory. *Yad. Fiz.*, 20:181–198, 1974.
- [93] G. Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.
- [94] Y. L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.
- [95] J. Butterworth *et. al.* PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.
- [96] The NNPDF Collaboration. A Determination of parton distributions with faithful uncertainty estimation. *Nucl. Phys. B*, 809:1–63, 2009.

- [97] R. S. Thorne A. D. Martin, W. J. Stirling and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63(2):189–285, Jul 2009.
- [98] D. Sayipjamal *et. al.* New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93(3), Feb 2016.
- [99] The ATLAS Collaboraion. Energy scale and resolution systematics: summary and correlations. Technical Report ATL-COM-PHYS-2017-754, Geneva, 2017.
- [100] The ATLAS Collaboration. Performance of electron and photon triggers in ATLAS during LHC Run 2. *Eur. Phys. J. C*, 80(CERN-EP-2019-169):47, 2020.
- [101] The ATLAS Collaboration. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 81(8):689, 2021.
- [102] Dependence of the Jet Energy Scale on the Particle Content of Hadronic Jets in the ATLAS Detector Simulation. Technical Report ATL-PHYS-PUB-2022-021, CERN, Geneva, 2022.
- [103] E. Bothmann *et. al.* Event generation with Sherpa 2.2. *SciPost Phys.*, 7(3), Sep 2019.
- [104] M. Bähr *et. al.* Herwig physics and manual. *Eur. Phys. J. C*, 58(4):639–707, Nov 2008.
- [105] A comprehensive guide to the physics and usage of PYTHIA 8.3.
- [106] K. Schörner-Sadenius T Schörner-Sadenius O. Behnke, K. Kröninger and G. Schott. *Data analysis in high energy physics: A practical guide to statistical methods*. Wiley-VCH, Weinheim, Germany, 2013.
- [107] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1):60 – 62, 1938.
- [108] C. Bierlich *et. al.* Robust Independent Validation of Experiment and Theory: Rivet version 3. *SciPost Phys.*, 8(2), Feb 2020.

- [109] F. Maltoni R. Frederix, S. Frixione and T. Stelzer *et. al.* Automation of next-to-leading order computations in QCD: the FKS subtraction. *Journal of High Energy Physics*, 2009(10):003–003, Oct 2009.
- [110] S. Schumann S. Höche, F. Krauss and F. Siegert. 2009(05):053–053, May 2009.
- [111] G. D’Agostini. A multidimensional unfolding method based on Bayes’ Theorem. Technical Report DESY-94-099, DESY, Hamburg, 1994.
- [112] ATLAS Pythia 8 tunes to 7 TeV data. Technical report, CERN, Geneva, 2014.

Appendix A

Simulated Event Samples

A summary of the different event generators and the accuracy of their matrix element calculations, PDF sets, parton shower and hadronization packages, and parameter sets used to simulate different physics processes is shown in Table A.1. The A14 parameter tune is performed to LHC data at $\sqrt{s} = 7$ TeV and is described in detail in [112]. The default tune of the Sherpa event generation shower model is performed to LEP data and is described in [103]

Process	Prompt/Non-prompt	Generator	ME Accuracy	PDF	Shower & Hadronization	Parameter Tune
EW $W\gamma jj$	Prompt	Sherpa 2.2.12	LO	NNPDF3.0 NLO	Sherpa	Default
		Madgraph5	LO	NNPDF3.0 LO	Pythia8+EvtGen	A14
QCD $W\gamma jj$	Prompt	Sherpa 2.2.11	NLO	NNPDF3.0 NLO	Sherpa	Default
		Madgraph5	NLO	NNPDF3.0 NLO	Pythia8+EvtGen	A14
EW $Z\gamma jj$	Prompt	Madgraph5	LO	NNPDF3.1 LO	Pythia8+EvtGen	A14
QCD $Z\gamma jj$	Prompt	Sherpa 2.2.11	NLO	NNPDF3.0 NLO	Sherpa	Default
$t\bar{t}\gamma$	Prompt	Madgraph5	LO	NNPDF2.3 LO	Pythia8+EvtGen	A14
$tW\gamma$	Prompt	Madgraph5	LO	NNPDF3.0 NLO	Pythia8+EvtGen	A14
$tq\gamma$	Prompt	Madgraph5	LO	NNPDF3.0 NLO	Herwig7+EvtGen	Default
Single Top	Non-Prompt	Powheg	NLO	NNPDF3.0 NLO	Pythia8+EvtGen	A14
W+jets	Non-Prompt	Sherpa 2.2.11	NLO	NNPDF3.0 NNLO	Sherpa	Default
Z+jets	Non-Prompt	Sherpa 2.2.11	NLO	NNPDF3.0 NNLO	Sherpa	Default
Diboson	Non-Prompt	Sherpa 2.2.12	NLO	NNPDF 3.0 NNLO	Sherpa	Default
Dijet	Non-Prompt	Pythia8	LO	NNPDF2.3 LO	Pythia8+EvtGen	A14
EW Wjj	Non-Prompt	Sherpa 2.2.1	LO	NNPDF3.0 NNLO	Sherpa	Default
EW Zjj	Non-Prompt	Sherpa 2.2.1	LO	NNPDF3.0 NNLO	Sherpa	Default
$t\bar{t}$	Non-Prompt	Powheg	NLO	NNPDF3.0 NLO	Pythia8+EvtGen	A14
tW	Non-Prompt	Powheg	NLO	NNPDF3.0 NLO	Pythia8+EvtGen	A14

Table A.1: Summary of simulated samples used in the analysis presented in this thesis. Prompt samples are used to provide template shapes to the fits performed in this analysis, while samples of non-prompt backgrounds are only used in control regions for data driven estimates of non-prompt backgrounds or in closure tests of estimates of non-prompt backgrounds.

Appendix B

Jets Misreconstructed as Photons: The 2D Sideband Method

To cross check the template fit method used to estimate the number of jets misreconstructed as photons, an additional data driven method is employed called the 2D sideband method. In this cut-and-count method, the photon isolation and identification are used to define a two dimensional plane consisting of a signal region and three control regions enriched in jets misidentified as photons. The signal region is defined by the requirement that a reconstructed photon be *tight* and isolated and the control regions are defined by inverting the identification requirement, the isolation requirement, or both. The two dimensional plane is illustrated in Figure B.1.

To account for leakage of prompt photons into the control regions in this method, a leakage coefficient is defined for each control region as follows

$$c_B = \frac{N_B^{sig}}{N_A^{sig}}, c_C = \frac{N_C^{sig}}{N_A^{sig}}, c_D = \frac{N_D^{sig}}{N_A^{sig}} \quad (\text{B.1})$$

where N_x^{sig} are the number of true photon simulated events in region x . The assumption of the method is that the isolation and identification variables are uncorrelated for events where a jet is misreconstructed as a photon, however a correlation coefficient R can be

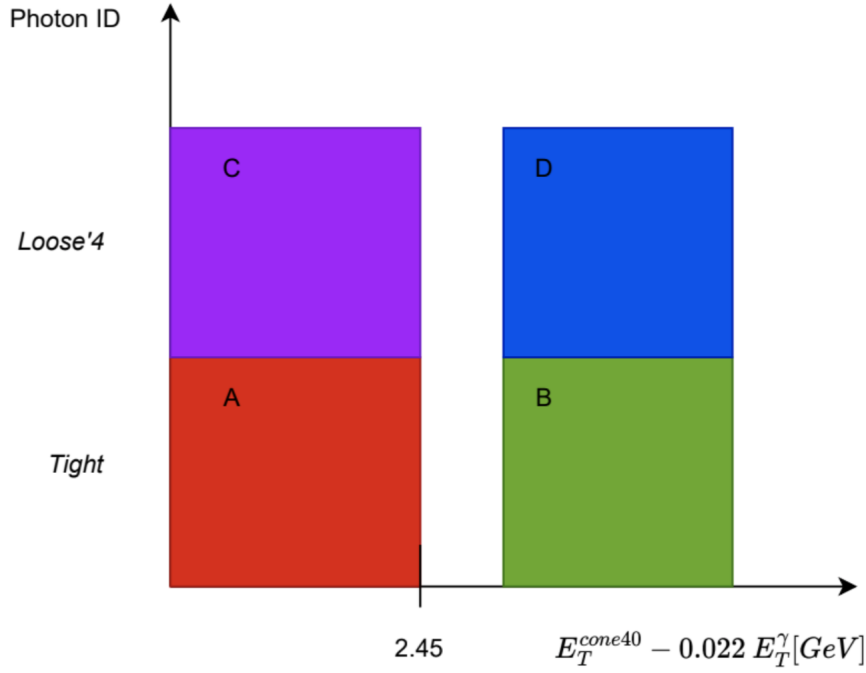


Figure B.1: Regions used for the 2D Sideband method to estimate the yield of jets misreconstructed as photons.

defined to account for any deviation from this assumption

$$R = \frac{N_A^{bkg} \cdot N_D^{bkg}}{N_B^{bkg} \cdot N_C^{bkg}} \quad (\text{B.2})$$

where N_X^{bkg} with $X = A, B, C$ is the number of simulated events where a jet is misreconstructed in region X . Armed with these coefficients, the number of real photon events in data is given by:

$$N_A^\gamma = N_A - N_A^{other} - R \frac{((N_B - N_B^{other}) - c_B N_A)((N_C - N_C^{other}) - c_C N_A)}{((N_D - N_D^{other}) - c_D N_A)} \quad (\text{B.3})$$

where N_X^{other} with $X = A, B, C$ are the other backgrounds (events where the photon is neither a prompt photon or a jet) taken from simulation. Note that the only significant source of this background in this region is simulated electrons misreconstructed as photons. This is also estimated as a data driven estimate. The method was performed both with and

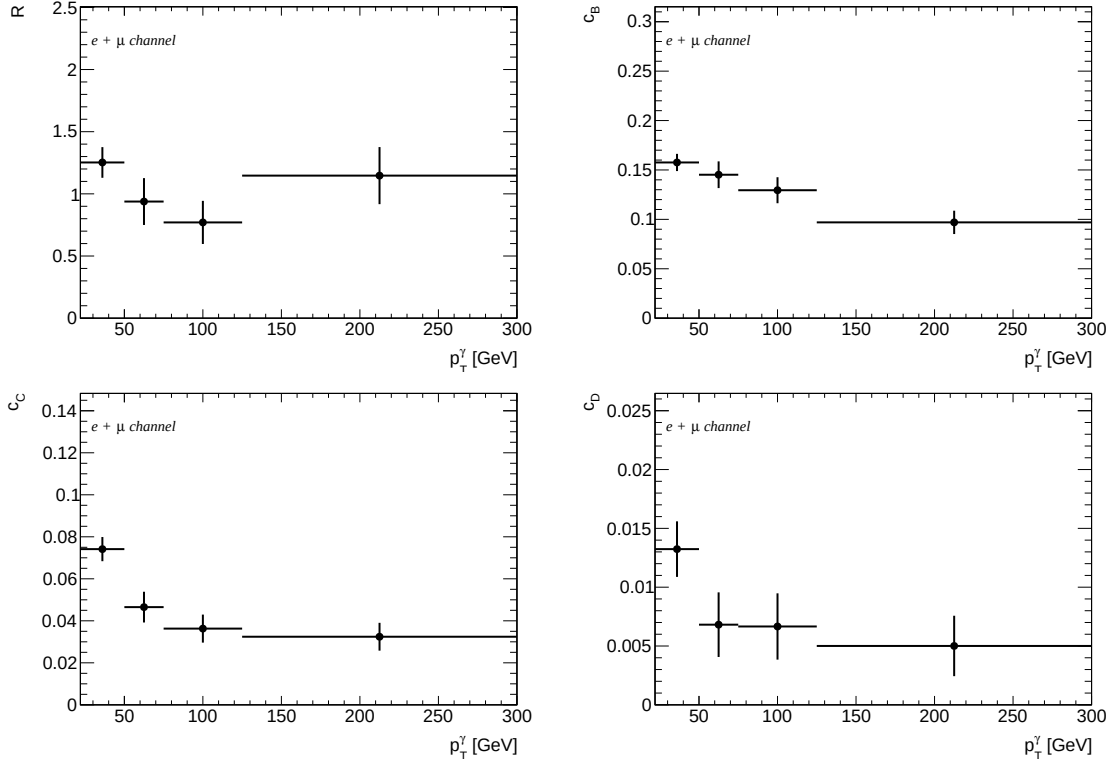


Figure B.2: Correlation coefficient R and signal leakage coefficients (C_B , C_C , C_D) in the three control regions in the 2D sideband method.

without the subtraction of this background and the result was found to be statistically equivalent. As such, there is no background subtraction in the following results. The correlation and leakage coefficients are shown in bins of photon transverse momentum in Figure B.2

The results of the template fit method and the 2D sideband method are compared in Figure B.3. The error bars represent statistical and systematic uncertainties. Agreement is seen between the two methods.

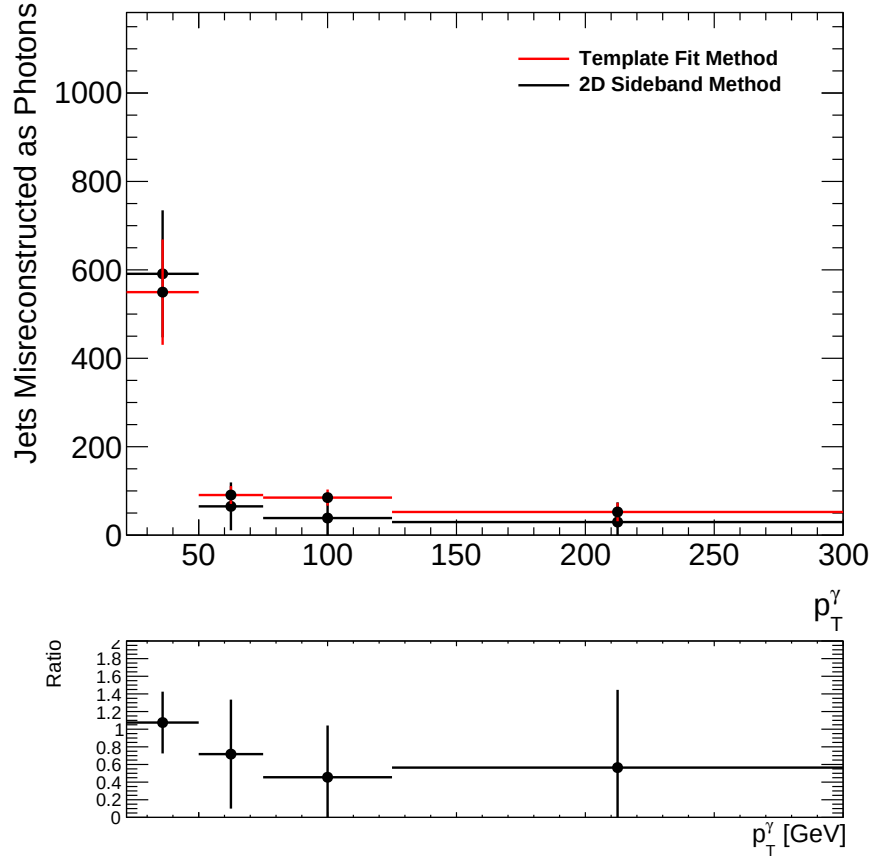


Figure B.3: Comparisons of the estimated event yields coming from jets misreconstructed as photons obtained using the 2D sideband method and template fit method in bins of photon transverse momentum in events satisfying the baseline selection. The error bars represent statistical and systematic uncertainties

Appendix C

Jets Misreconstructed as Photons: Prompt Photon Template Closure Test

As a technical check on the template fit method described in Section 5.1, a closure test is performed using simulated events. In the closure test the template fit method is applied with data replaced with an Asimov dataset made up of a simulated W+jets sample. The method is blind to the truth information of the jets misreconstructed as photons, and the resulting non-prompt photon distribution is compared to non-prompt photons as modelled in simulation. The template closure test is shown here in bins of photon transverse momentum.

First, the $E_T^{cone40} - 0.022E_T^\gamma$ shape of *non-tight*, non-prompt photons in simulated W+jet events is compared to the shape of *tight*, non-prompt photons using a χ^2 shape comparison. This quantifies the amount of bias in taking the non-prompt photon templates from the *non-tight* control region. If the distributions are significantly different in a given bin, full closure is not expected. This comparison is shown in Figure C.1. As can be seen there is a significant difference in shape in the first bin where the $\chi^2 = 0.09$. Non-closure may be expected there. Also note that the closure test is limited by statistics in the distribution it attempts to predict. The fit of the *non-tight* Asimov dataset is shown in Figure C.2 and the fit to Asimov *tight* photons is shown in Figure C.3. Here good agreement can be

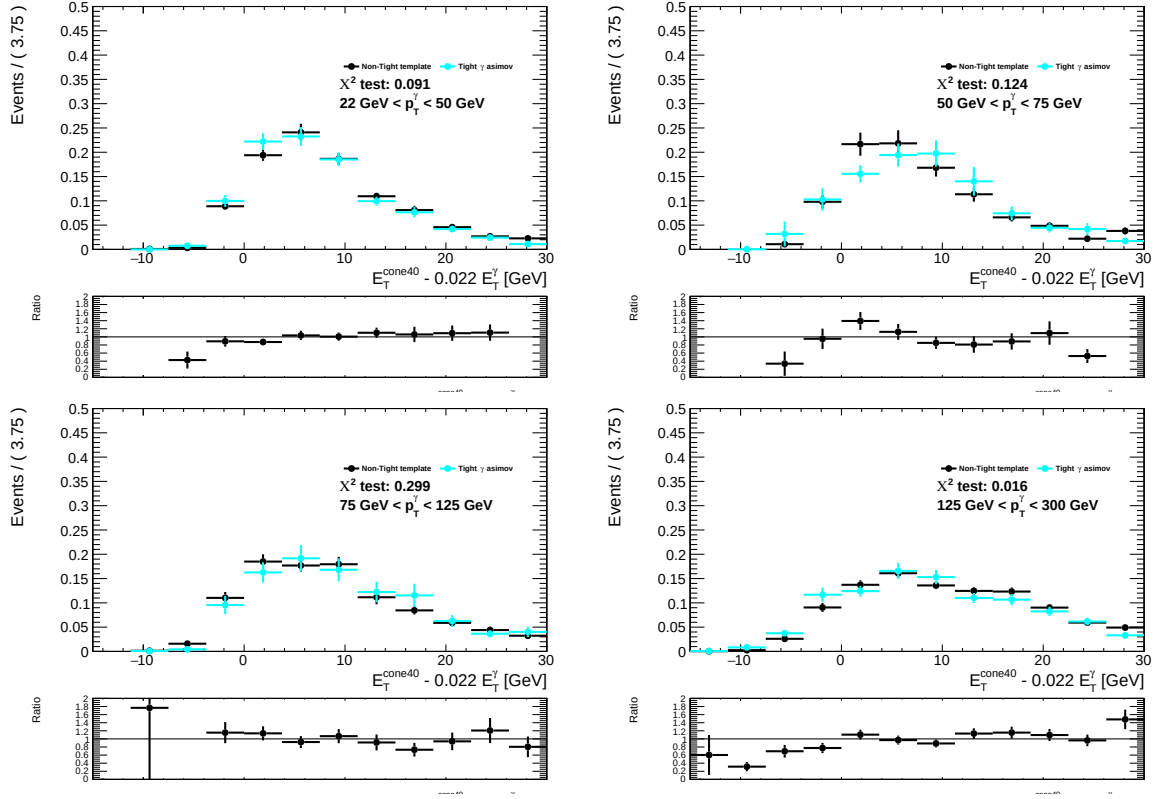


Figure C.1: Shape comparison of *non-tight* photons and tight photons in W+jets simulated events. Each plot represents a bin of photon transverse momentum. This quantifies the bias in taking a jet misreconstructed as photon template from a *non-tight* region.

seen, however the statistical fluctuations in simulated jets misreconstructed as photons prevents the possibility of covering this distribution with a smooth shape. Nonetheless, agreement in the overall yields within statistical errors is found in most bins. The method does not close in bin 1, where it is not expected to. Note also in the first bin the number of jets misreconstructed as photons predicted in simulation is not a smooth distribution. The comparison of the jet misreconstructed as photon yield in simulation and in the method applied to the Asimov dataset as a function of photon transverse momentum is shown in Figure C.4.

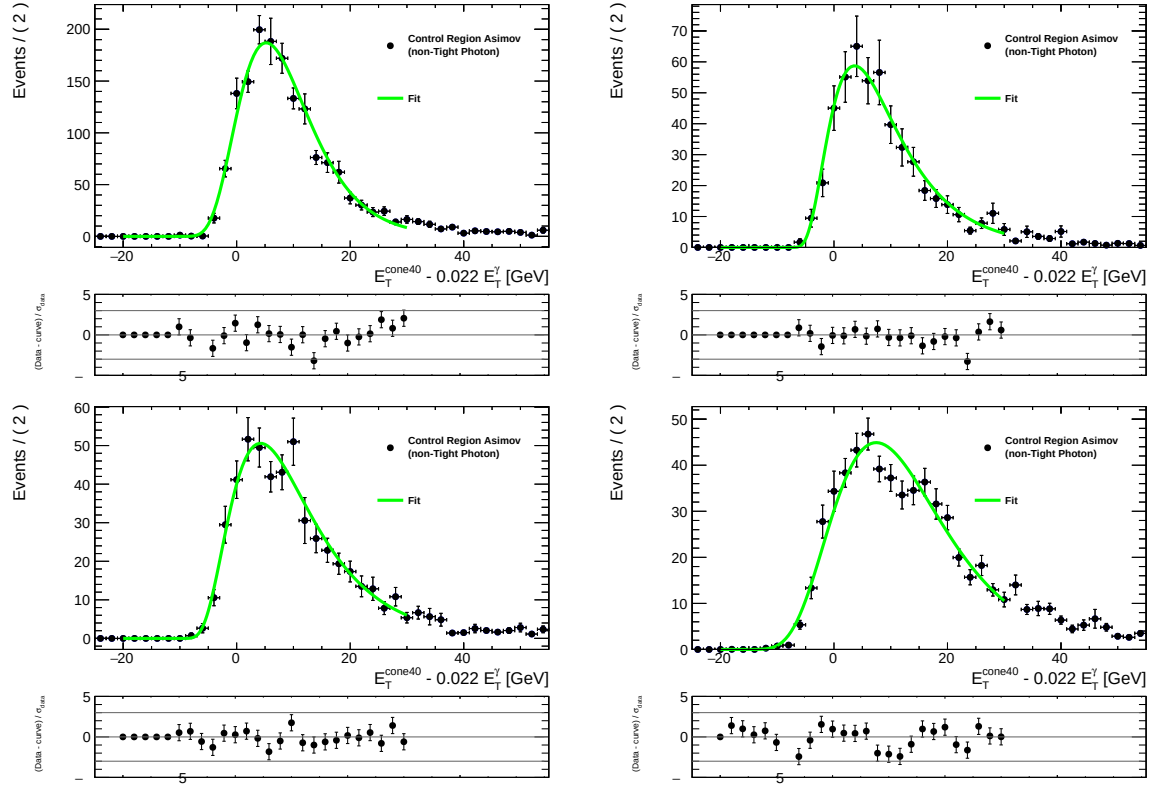


Figure C.2: Fits to *non-tight* Asimov data. Each figure represents the fit done in bins of photon transverse momentum. The *non-tight* Asimov dataset is made of jet *loose'4* photons from the simulated W+jets sample.

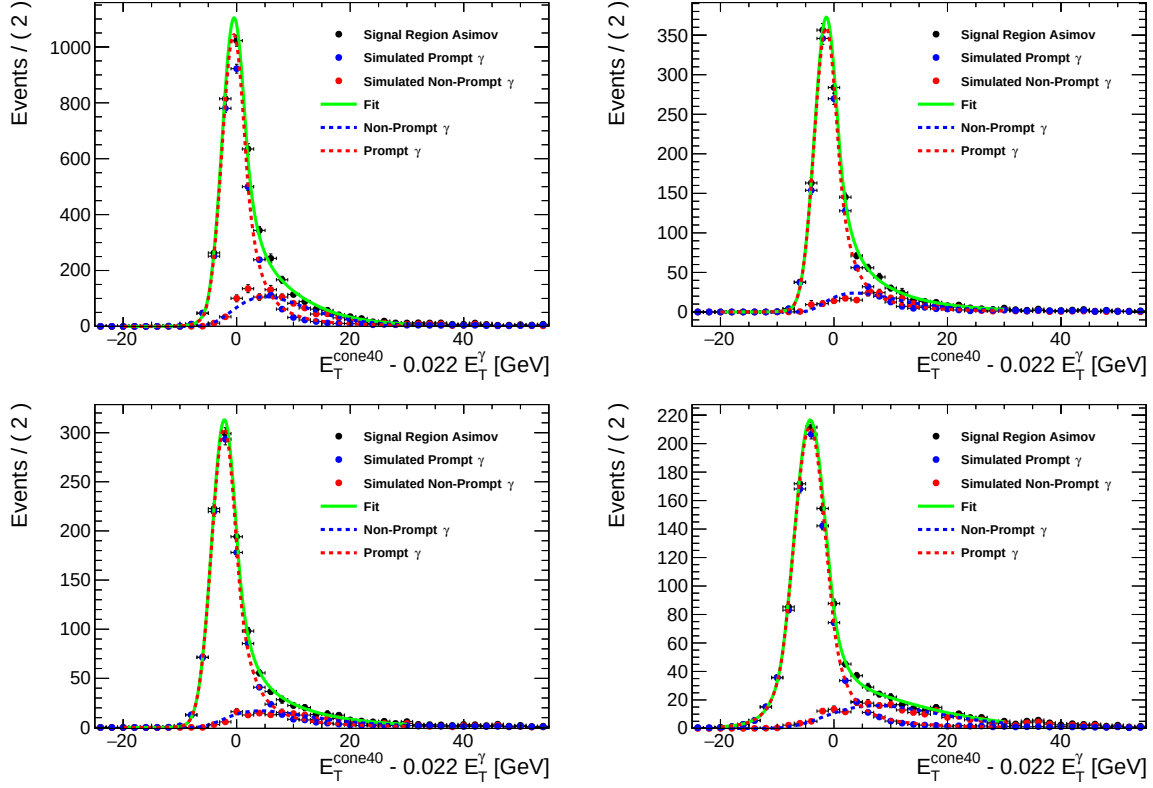


Figure C.3: Fits to *tight* Asimov data. Each figure represents the fit done in a bin of photon transverse momentum. The *tight* Asimov dataset is made of jets misreconstructed as *tight* photons from the simulated W +jets sample and real photons from the simulated QCD $W\gamma jj$ sample. The real and misreconstructed components of the fit are shown.

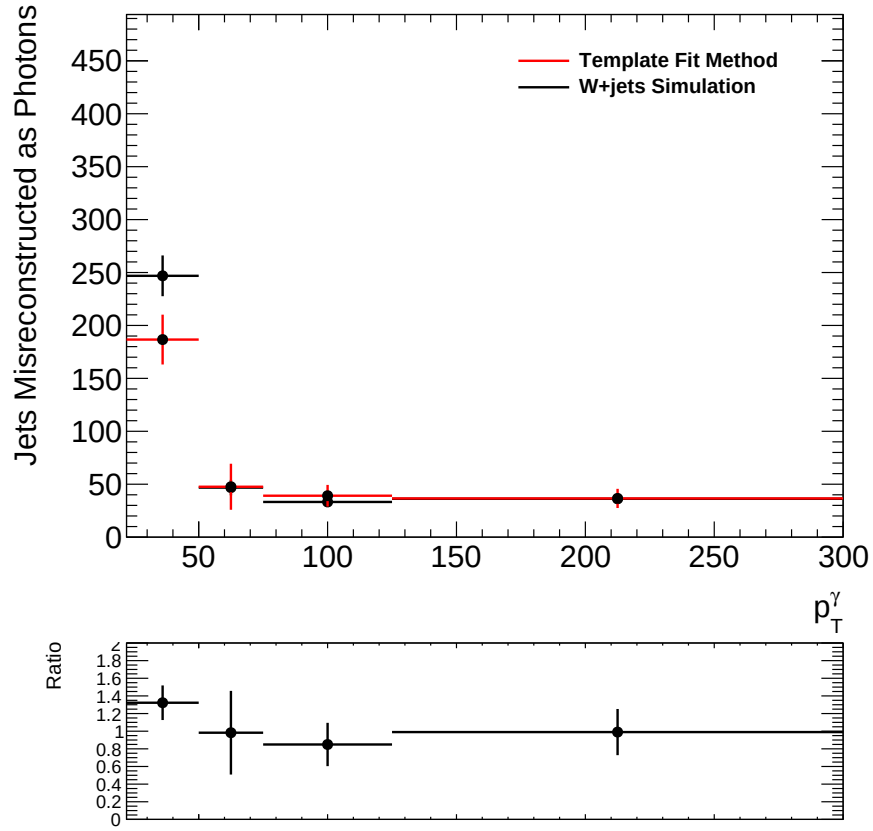


Figure C.4: Comparison of non-prompt photon yield as a function of photon transverse momentum from simulation with that predicted by the template fit method applied to an Asimov dataset.

Appendix D

Jets Misreconstructed as Photons: Cross-check of real photon purity modelling

As described in Chapter 5, there are not sufficient statistics to perform the template fit to measure the background arising from jets misidentified as photons in each bin of the phase space region used in the differential cross section measurement. The template fit is performed inclusively in each region and the prompt photon purity in simulation is used to derive the differential shape of the background from jets misreconstructed as photons. The assumption in this approach is that the non-prompt photon yield relative to the total reconstructed photon yield as a function of the observable being plotted is well modelled in simulation. To check this assumption, the real photon purity as a function of photon transverse momentum is fit with a linear function over two bins of roughly equal number of events in the non-prompt photon control region in both simulation and in data. This is shown in Figures D.1. The slope is found to be consistent between data and simulation, validating the shape extrapolation used for the estimate of the non-prompt photon background used in the differential measurement presented in Chapter 9.

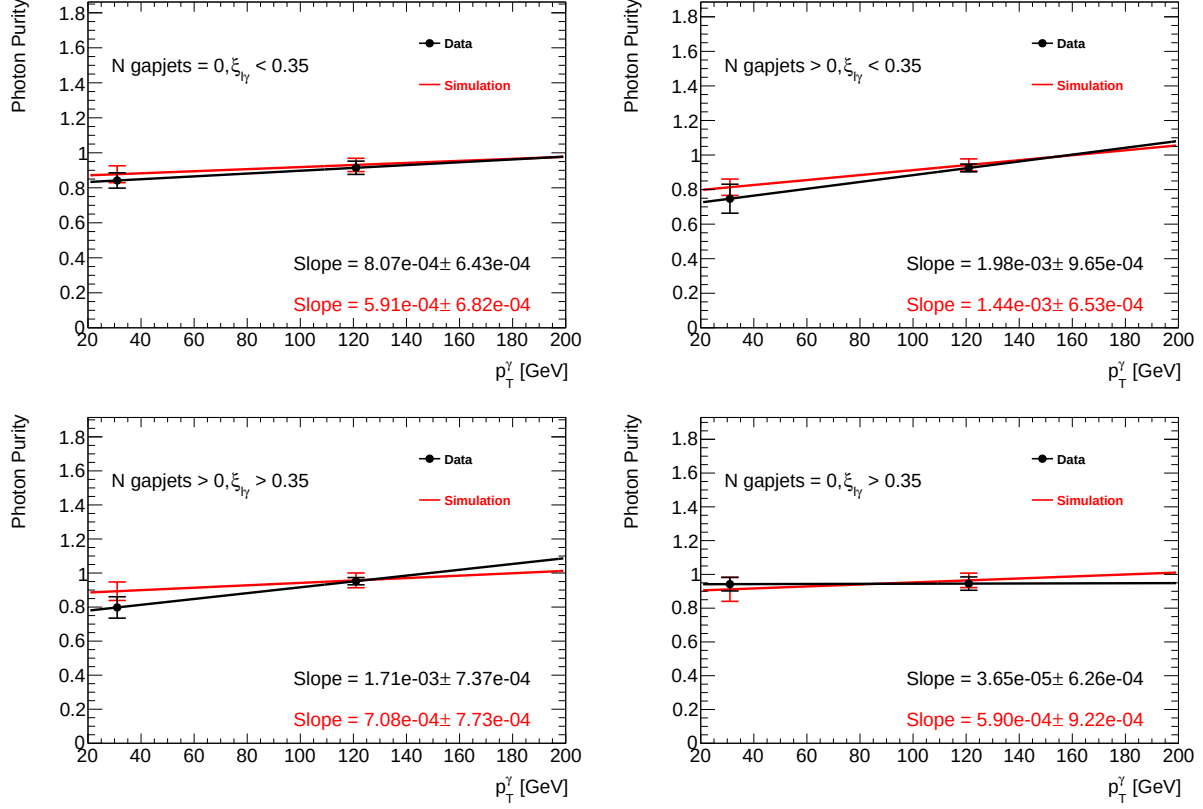


Figure D.1: Prompt photon purity in simulation and data as a function of photon transverse momentum in the signal and control regions used for the differential cross section extraction in Chapter 9.

Appendix E

Electrons Misreconstructed as Photons: Signal fraction fits

The background contamination in the region used to calculate $F_{e \rightarrow \gamma}$ in the data driven estimate of electrons misreconstructed as photons may be different between the region defined by a *Probe* electron and *Tag* electron and the region defined between a *Probe* electron and a *Tag* photon. To account for this, a weight is assigned to each type of *Tag* and *Probe* pair. The weight is the signal fraction $\frac{S}{S+B}$ determined by an $S + B$ fit to the invariant mass of the *Tag* and *Probe* pair, either $M_{e\gamma}$ or M_{ee} . A double-sided Crystal Ball function is used to describe the signal shape, while an exponentially decaying function is used to model the background. Both shape parameters and normalization are floated in the fit. The invariant mass fits in each bin of *Probe* $|\eta|$ and transverse momentum are shown in Figure E.1 and in Figure E.2 for $M_{e\gamma}$ and M_{ee} , respectively.

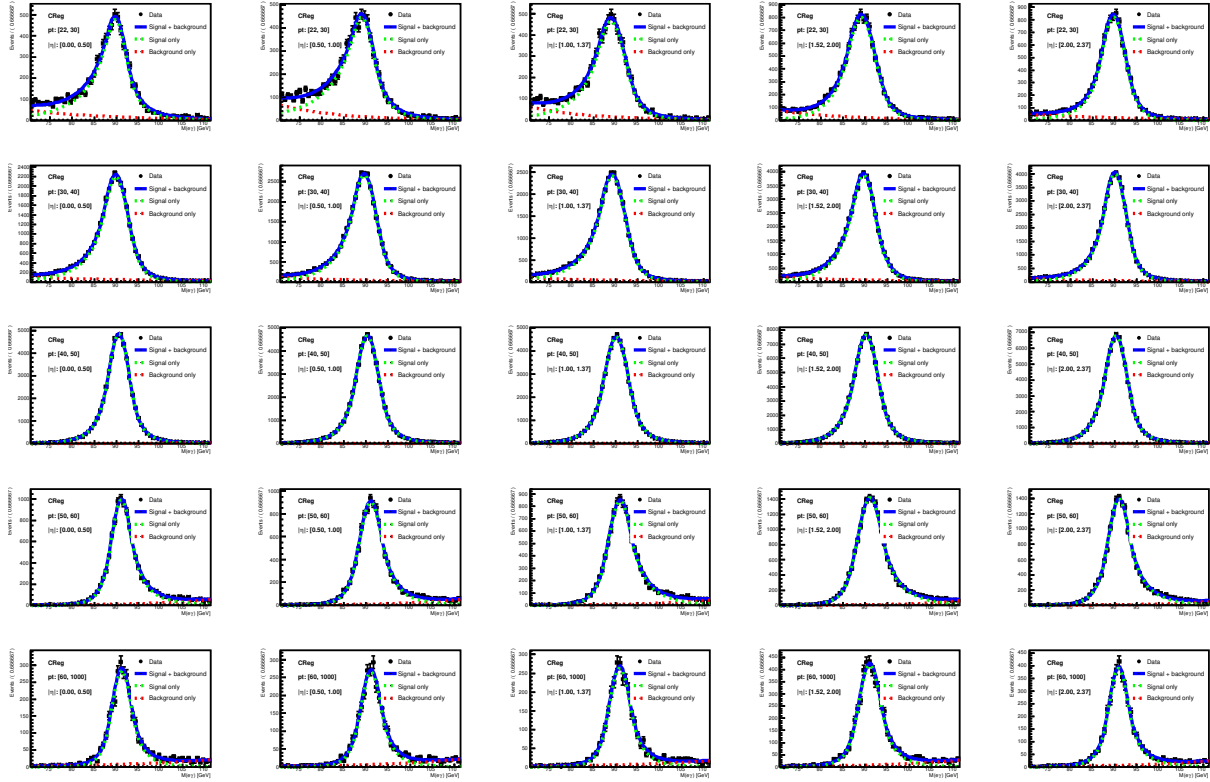


Figure E.1: Signal and background fits to the reconstructed invariant mass M_{ey} for each *Probe* $|\eta|$ and transverse momentum bin.

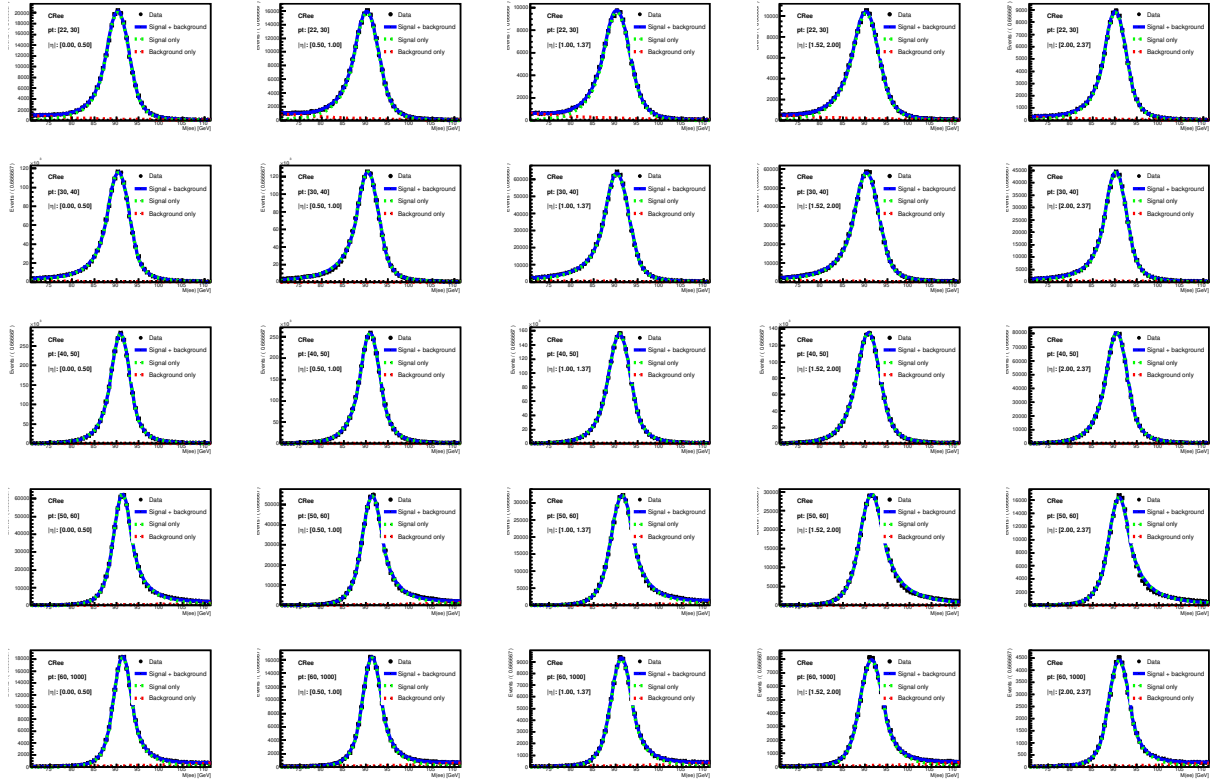


Figure E.2: Signal and background fits to the reconstructed invariant mass M_{ee} for each *Probe* $|\eta|$ and transverse momentum bin.

Appendix F

Jets Misreconstructed as Leptons:

Fraction $F_{j \rightarrow \ell}$

In this appendix the fraction of jets misreconstructed as leptons, $F_{j \rightarrow \ell}$, as a function of lepton transverse momentum and pseudorapidity is shown. This fraction is estimated in a dijet enriched region with one lepton and one jet requirement as described in Section 5.3.

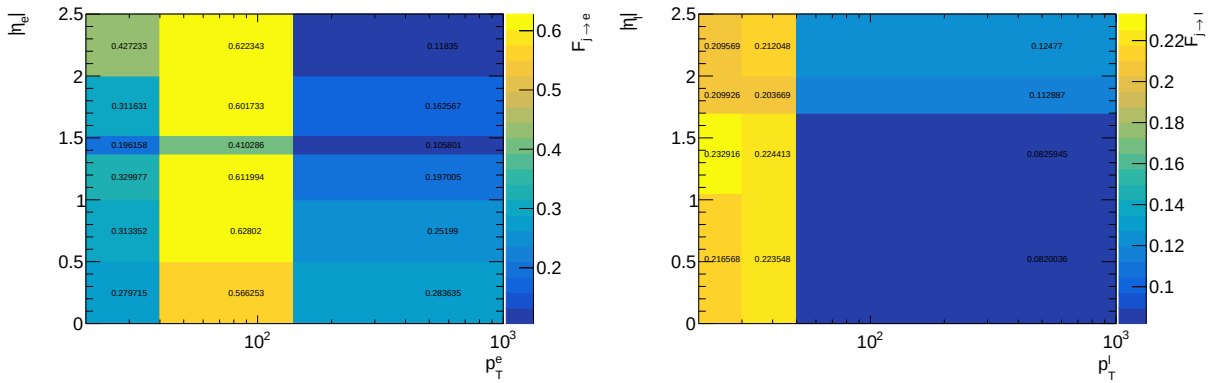


Figure F.1: Fraction $F_{j \rightarrow \ell}$ as a function of lepton transverse momentum and pseudorapidity in the electron (left) and muon (right) channels.

Appendix G

Comparison of Neural Network Training Variables in the Electron and Muon channels

In Figure G.1 the shape of neural network training variables in the simulated samples used to train the neural network presented in Chapter 6 are compared between the analysis muon and electron channels. Since no significant shape differences are seen between the two channels, training is performed on the combined channels.

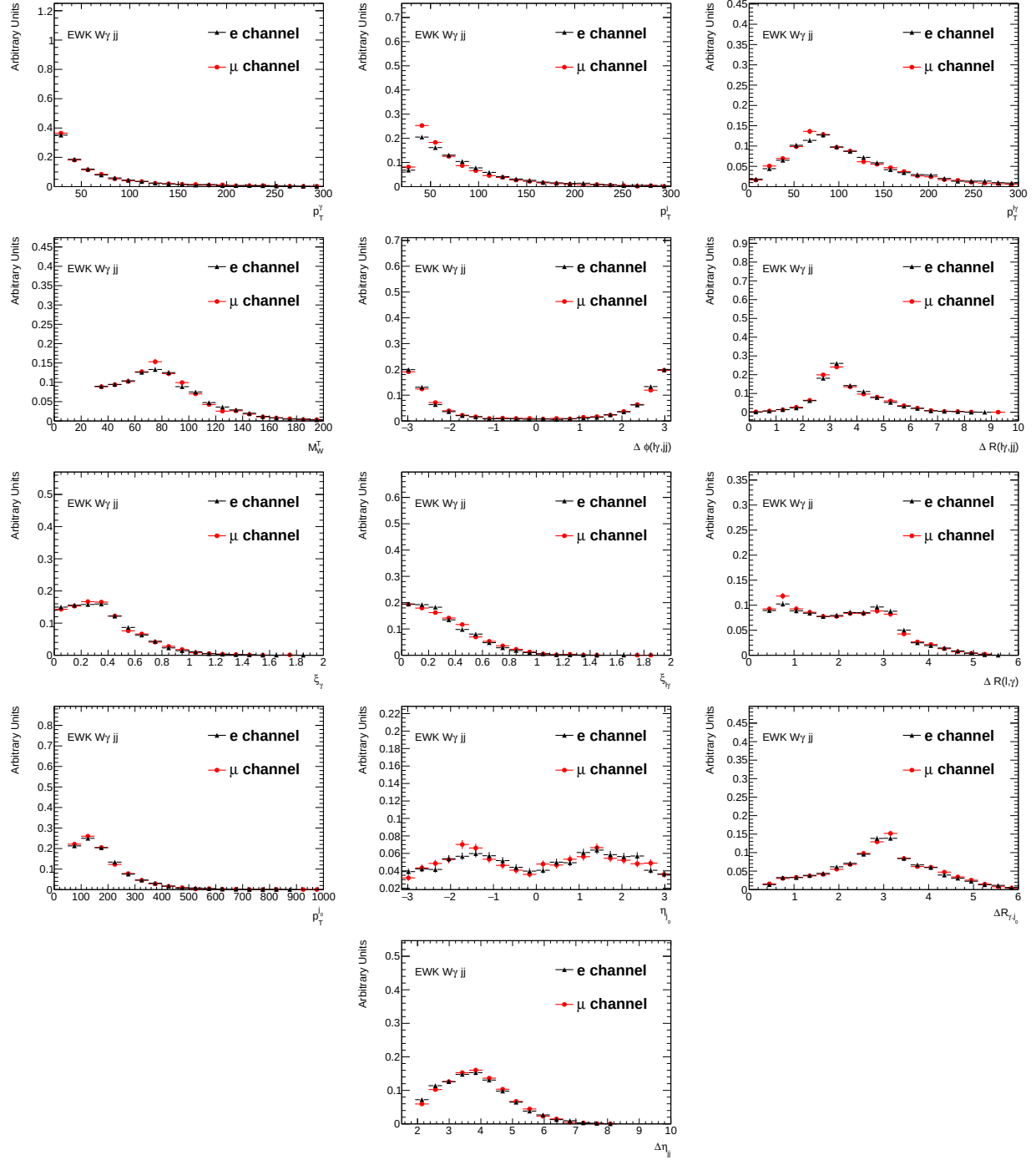


Figure G.1: Shape comparison of kinematic variables used in neural network training between the electron and muon channels in signal and background simulated samples. All distributions are normalized to one.

Appendix H

Treatment of Scale Uncertainties

Typically scale uncertainties are combined in an envelope, however, in the case that a normalization is fit to data and the envelope is flat, this can underestimate the impact of individual scale variations if those scale variations have a shape. To check that this is unlikely the case, the shape of each individual scale uncertainty is compared with nominal in the signal and control region for both the signal sample and the QCD $W\gamma jj$ sample in Figures H.1 and H.2, respectively. All individual scale variations are statistically flat which justifies the envelope approach.

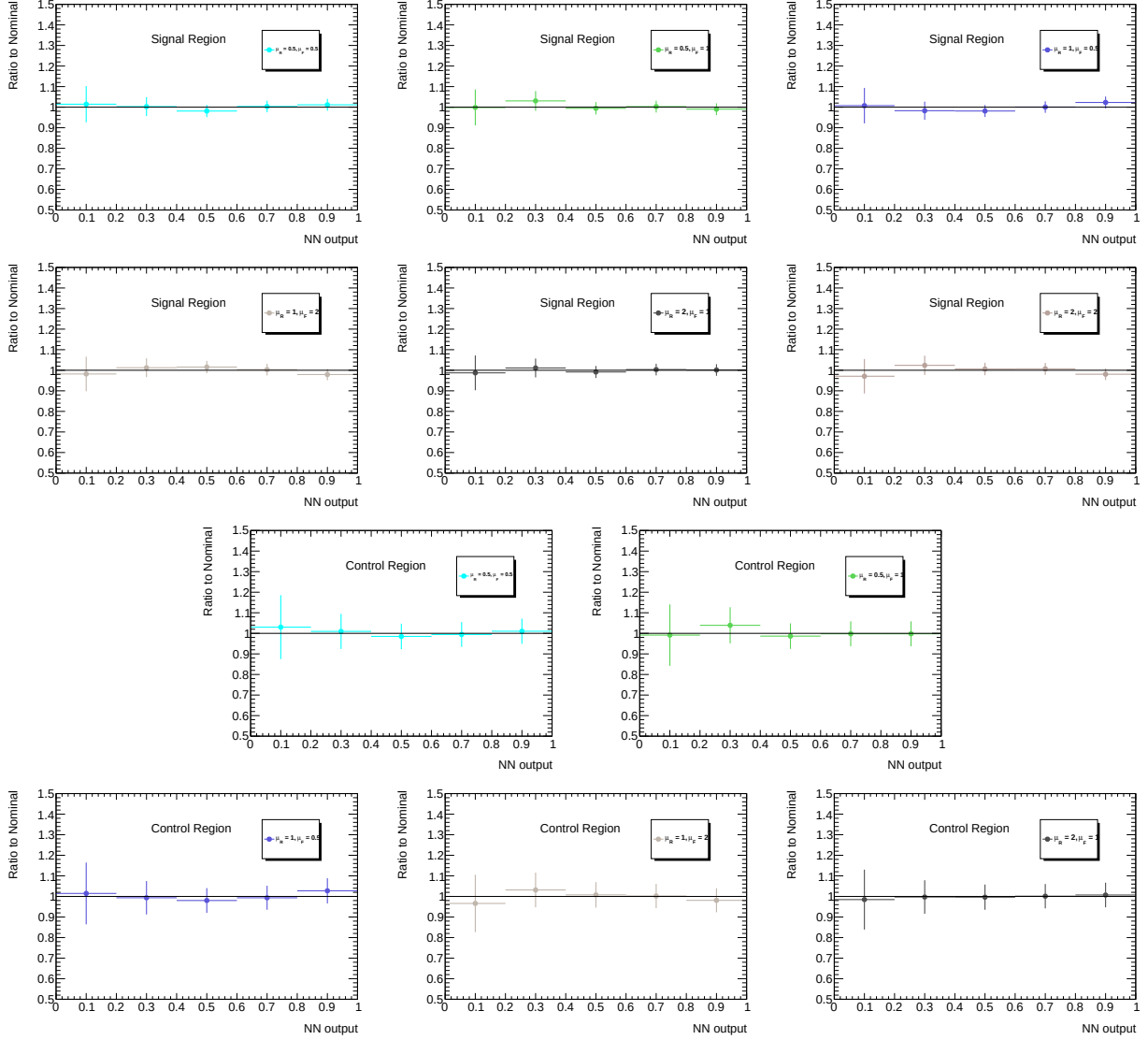


Figure H.1: Shape component of individual scale variations in the neural network output distribution for the Sherpa electroweak $W\gamma jj$ sample in the signal and control region used in the fit to extract a fiducial cross section.

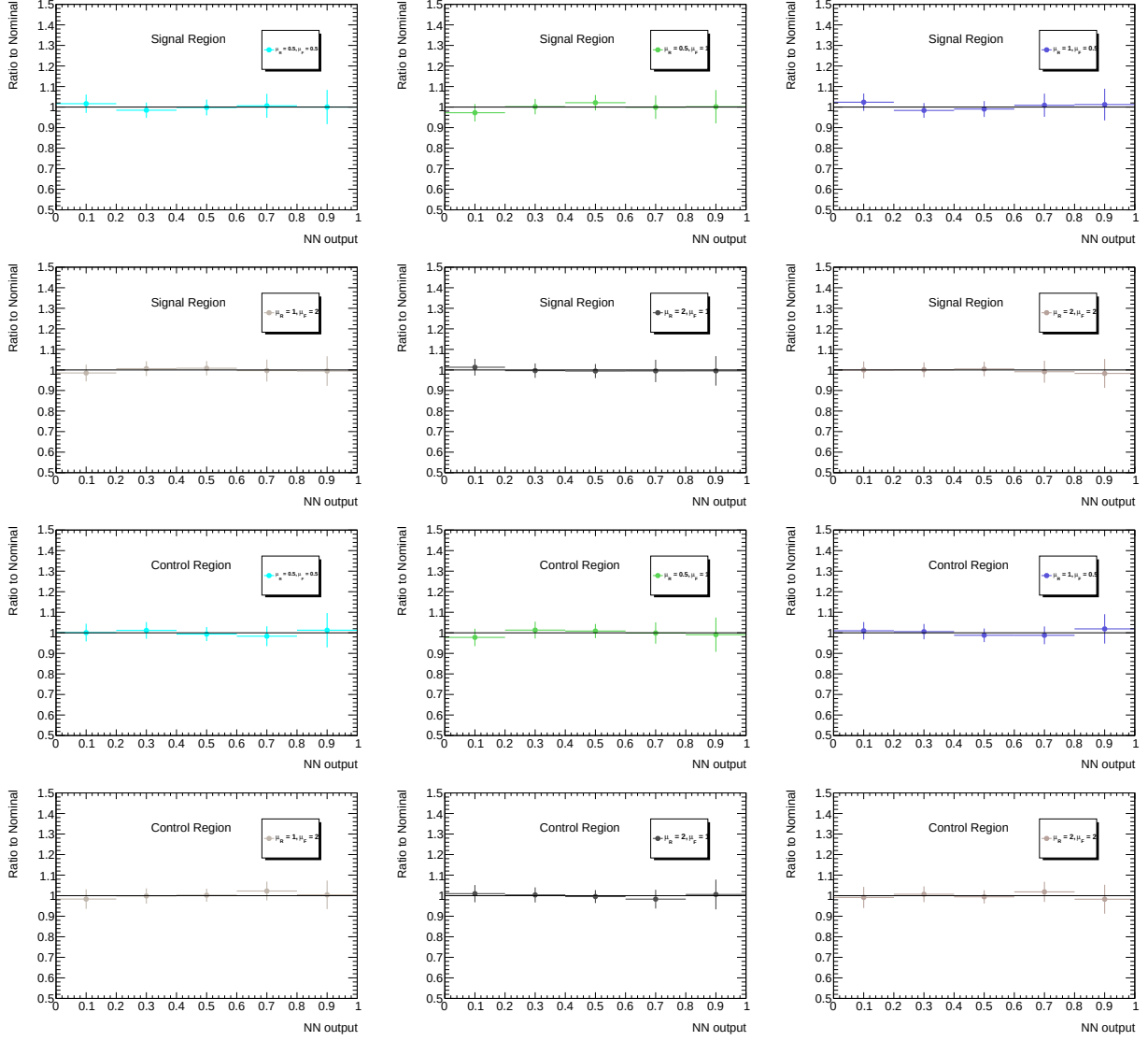


Figure H.2: Shape component of individual scale variations in the neural network output distribution for the Sherpa QCD $W\gamma jj$ sample used in the fit to extract a fiducial cross section.

Appendix I

Pulls of Nuisance Parameters in Fiducial Cross Section Fit

The pulls of nuisance parameters in the fit performed on data in the signal and control region used to extract a signal significance and fiducial cross section are shown here. Some pulls are seen on the nuisance parameters affecting the shape of the electroweak $W\gamma jj$ yield and the shape of the QCD $W\gamma jj$ yield. Overconstraints are seen on very large systematics, like the QCD event generator choice uncertainty. This reflects that the data has sufficient statistical power to constrain these nuisance parameters.

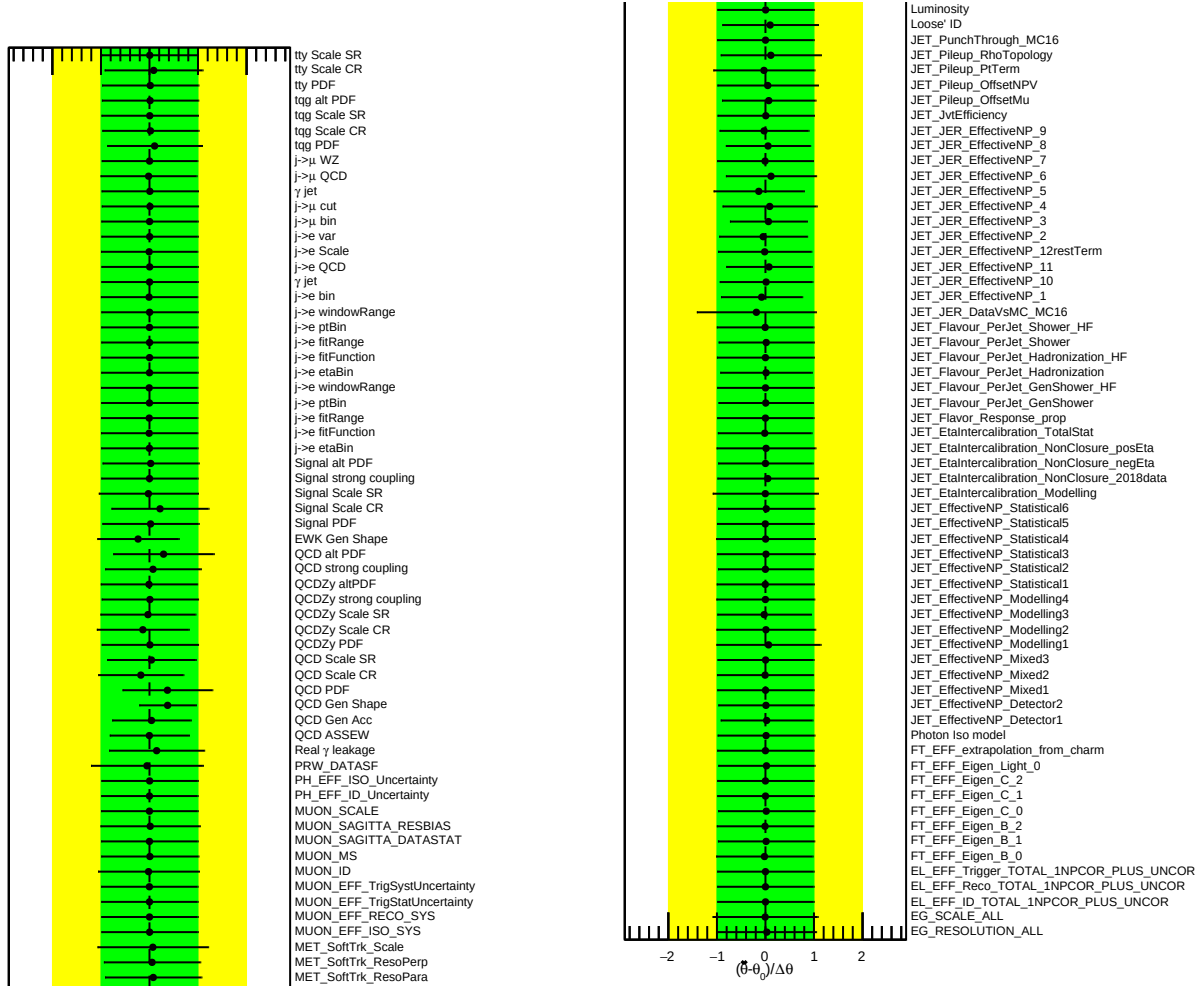


Figure I.1: Pulls of the nuisance parameters in the fit simultaneous to data in the signal and control regions.