New physics searches in angular shapes of photon+jet events in 2012 ATLAS data

Sheir Yarkoni
Department of Physics
McGill University, Montréal

May 2014

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of M.Sc.

Abstract

The Large Hadron Collider at CERN, Switzerland, was built to both refine current Standard Model measurements, as well as discover new physics beyond the Standard Model. Using data from the general-purpose ATLAS detector, we take a step towards answering one of these questions: are quarks point-like, or do they themselves have sub-structure? By investigating the angular correlations in photon+jet events we address this question. The final goal of this thesis was to produce a feasible method to test deviations from the Standard Model using 2012 ATLAS data. A new data-driven background estimation analysis technique was developed to test for significant deviations from Standard Model predictions. We parametrize the angular separation between the leading photon and jet in each event with the variable $\psi = \log(\exp(|\eta_\gamma - \eta_j|)+1)$ (where η is pseudorapidity) using standard ATLAS Monte Carlo generators Pythia, Sherpa, and JetPhox. Kinematic comparisons are performed between 2012 ATLAS data and Monte Carlo samples to show how well data is described by the generators. We take the ratio of the number of narrowly-separated ($\psi < 1.5$) to selected events with $\psi < 5$ as a function of invariant mass of the final state photon and the final state jet, resulting in our observable distribution F_ψ. We fit the numerator and denominator invariant mass distributions separately and obtain the background estimate of the F_ψ distribution in each bin by dividing the results of the fit in each bin. We use a bootstrap method for statistical error estimation. Our analysis techniques were optimized with the Excited Quark model as a benchmark model. Using the developed methodology for background estimation, we proceed to estimate the expected signal sensitivity of this analysis technique to excited quark production in the full 2012 ATLAS dataset of 20.3 fb^{-1}.

Résumé

Le Grand collisionneur de hadrons au CERN, en Suisse, a été construit afin d'améliorer les mesures actuelles du Modèle Standard et pour tenter de découvrir des phénomènes physiques qui ne sont pas décrits par le Modèle Standard. En utilisant les données amassées par le détecteur ATLAS, nous nous penchons sur un de ces phénomènes: les quarks sont-ils des particules ponctuelles ou ont-ils plutôt une structure interne? Nous nous servons des corrélations angulaires entre les photons et les gerbes de particules dans l'état final de collisions à hautes énergies pour étudier la question. Le but de ce mémoire est de démontrer l'efficacité de cette technique d'analyse et d'en étudier la performance avec les données prises par ATLAS en 2012. Nous présentons une nouvelle technique d'estimation du bruit de fond causé par les processus physiques du Modèle Standard et un test qui permet de déterminer si les données d'ATLAS dévient significativement par rapport à ce bruit de fond. Nous paramétrons la séparation angulaire entre le photon le plus énergétique et la gerbe la plus énergétique dans chaque événement par $\psi = \log(\exp(|\eta_\gamma - \eta_j|) + 1)$ (où η est la pseudorapidité) en utilisant des données simulées par les générateurs Monte Carlo Pythia, Sherpa et JetPhox. Nous comparons les distributions des variables cinétiques entre les données ATLAS et les événement simulés pour démontrer que les données simulées reproduisent bien les caractéristiques des données réelles. Nous définissons un nouvel observable F_ψ qui représente la fraction des événements qui ont une petite séparation angulaire (ψ est plus petit que 1.5) parmi les événements acceptés (ψ est plus petit que 5) en fonction de la masse au repos du système photon+gerbe. Un ajustement de courbe est fait séparément pour les distributions de masse au repos qui constituent le numérateur et le dénominateur de l'observable F_ψ. La prédiction du bruit de fond en F_ψ dans chaque intervalle de masse au repos est obtenue par le résultat de la division des valeurs des deux courbes ajustées dans cet intervalle. L'incertitude statistique est calculée à l'aide d'une méthode de type "bootstrap". La méthode analytique est optimisée en utilisant comme modèle de référence le modèle du quark excité. Avec notre nouvelle méthode d'estimation du bruit de fond, notre

technique d'analyse et notre signal de référence, nous mesurons la valeur attendue pour la détection d'un signal de quark excité dans les 20.3 fb^{-1} de données ATLAS de 2012.

Acknowledgements

An enormous thank you to Steve and Camille for their help and continual support throughout this project, I know it couldn't have been easy. Camille, without our nearly daily talks and your repeated explanations, I don't know if I'd have ever finished this thesis. Thank you to the rest of the McGill ATLAS group for their comradery, whether it was sunny or snowy outside (mostly snowy), it was always a pleasure to come to the office. And lastly, thank you to all my friends and family for supporting me throughout this project.

Statement of Originality

The author, Sheir Yarkoni, claims the following aspects of the work contained herein constitute original scholarship and an advancement of knowledge:

- Optimizing the selection criteria for the angular analysis.
- The background estimation used for F_ψ.
- The method in which statistical errors are estimated for the background estimation for F_ψ.
- Expected signal sensitivity testing procedure used to calculate excited quark production in 2012 ATLAS data.

Contents

List of Figures

List of Tables

1

Introduction

The Standard Model of Particle Physics is the theory that describes all known elementary particles of matter and all interactions between them via the strong nuclear force, the weak nuclear force, and the electromagnetic force [1]. Despite the experimental success of the Standard Model, there are still many fundamental questions left to answer: Why is the top quark so massive? What is responsible for dark matter? Are there extra dimensions? Are quarks truly indivisible, or are they themselves composite? The subject of this thesis relates to this last question.

As of now, quarks, which are the building blocks of protons and neutrons, are assumed to be point-like in the Standard Model, which is consistent with all experimental observations to date. However, this is an assumption built into the theory, and there is no fundamental reason why nature should behave this way. It is possible that quarks are comprised themselves of smaller constituents. Should this be the case, it could be detectable at the Large Hadron Collider, at energies up to the TeV scale.

In this thesis, by analyzing the angular separation between quarks and photons, we develop a new method to test for quark substructure physics. We optimize the analysis' sensitivity to deviations from expected Standard Model production using the Excited Quark model (also denoted as q*) [12]. More generally, using angular variables under study, any significant deviation from the expected background would be evidence for new physics beyond the Standard Model.

In Chapter 2 we review Standard Model theory and the experimental conditions at ATLAS. In Chapter 3 we discuss the detection and reconstruction of jets and photons in the ATLAS calorimeters. Chapter 4 describes the Monte Carlo and data selection in this thesis. In Chapters 5-7 we motivate and present the analysis techniques, as well as results. Conclusions are drawn from these results in Chapter 8, as well as remarks concerning future photon+jet analyses at ATLAS.

2

Theory and ATLAS detector description

2.1 The Standard Model

The Standard Model of particle physics contains elementary particles in two general categories: bosons and fermions [1]. Bosons have integer values of intrinsic spin, whereas fermions have half-integer values. The fermions are further separated into two subgroups- quarks and leptons. These particles are differentiated based on whether or not they carry a quantum number called color charge. Every particle in the Standard Model has a corresponding anti-particle, with some being their own anti-particle. These anti-particles have opposite signed quantum numbers (for example, electromagnetic charge), but have the same mass. The groupings of the particles in the Standard Model can be found in Figure 2.1.

Bosons in the Standard Model are, in general, force carriers of one of the three fundamental forces: the strong nuclear force, the weak nuclear force, and the electromagnetic force. The one exception is the recently discovered Higgs boson [7], which is an artifact of the complex scalar doublet Higgs field and the symmetry-breaking mechanism [2]. Through the Higgs mechanism, the Higgs field gives mass to the weak gauge bosons in the Standard Model.

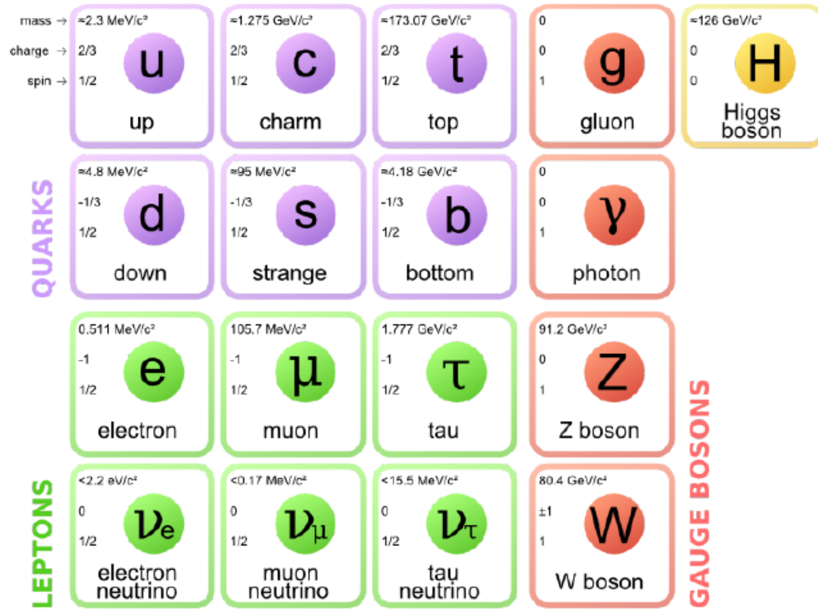


Figure 2.1: Table of Standard Model particles and their properties (mass, spin, and electric charge) [31]

The theory of Quantum Electrodynamics describes the interaction of the photon using the $U(1)_q$ gauge group. At high energies, the electromagnetic and weak forces can be unified into a single force called the electroweak force (the GSW model), by using the gauge group $SU(2)_L \times U(1)_Y$. Through the electroweak interaction, the W, Z, and γ bosons interact with the quarks and leptons [2]. At lower energies, below roughly 100 GeV, the weak nuclear force and the electromagnetic force are described separately. The spontaneous symmetry-breaking mechanism, also known as the Higgs Mechanism [2], gives mass to the gauge bosons W^+, W^-, and Z^0. The Z^0 and γ are neutral in electric charge, whereas the W^+ and W^- carry a +1 and -1 electric charge respectively.

Fermions are divided into three generations of quarks and leptons in the Standard Model (SM). Each generation of leptons is specified by a flavor (type), and contains a charged lepton and a neutrino of that flavor: electron, muon, and tau. There are 6 quarks: up, down, charm, strange, top, and bottom, divided into the three generations. Quarks also have flavor, and by emitting or absorbing a W boson, quarks in hadrons can change flavor. This is the process that is responsible for many radioactive decays. Quarks have fractional electric charge and and carry ”color charges”, labeled red, blue, and green (or the corresponding anti-color). The leptons have a quantum number associated with them, conveniently called the lepton number. All leptons have $+1$ lepton number, and their anti-particles have -1. Lepton number is a conserved quantity in interactions. However, despite this conservation law, the *flavor* is not always conserved. It has been shown experimentally that in decays that produce neutrinos, some oscillate to a different flavor of neutrino [13].

Quantum Chromodynamics (QCD) describes the strong nuclear interactions between gluons and quarks, and is described mathematically using the gauge group $SU(3)$. The gauge boson that mediates the strong nuclear force is the gluon. The quantum numbers associated with QCD interactions are called ”color charges”. Each gluon carries one colors and one anti-color, a quantity that is conserved in strong nuclear interactions. These combine to give a total of 8 unique gluons [3]. Due to a concept called ”color confinement”, quarks can not exist as free particles. Instead, they form bound states with each other, either as colorless mesons (combination of quark/anti-quark as a color/anti-color pair), or as color singlet baryons (3 quarks in either a red-green-blue state, or the corresponding anti-color state) [3].

As the distance between two quarks increases, the potential energy from the strong nuclear force will produce more particles from the vacuum between them to form new bound states. This process is known as ”hadronization”, and produces collimated streams of mesons, baryons, and other particles that deposit energy in detectors. These streams are called ”QCD jets”. Although in ATLAS any energy deposit can result in a jet candidate, we are interested only in energy deposits from hadronization. This is explained in detail in Chapter 3.

2.2 Photon+jet production in the Standard Model

In high energy particle physics, the mathematical description of a physical process is represented graphically using Feynman diagrams. We use these diagrams to calculate the production ”cross sections” of the physical process in the diagram. The more vertices a process has, the more Feynman diagrams there are that contribute to the cross section calculation. Perturbative expansions are used to calculate these additional cross section terms [1–3]. The order of the perturbative expansion is dictated by the number of vertices in the process. Thus increasing the order makes the calculation more difficult. The first order non-zero contributing terms are called ”Leading Order (LO)”, first-order corrections are ”Next-to-Leading Order (NLO)”, and so on.

In the Standard Model there are two ways to produce a final state γ+jet at leading order for proton-proton collisions (for this process also called tree-level): Compton scattering of a quark-gluon pair, and quark-antiquark annihilation, with their Feynman diagrams shown in Figure 2.2. At NLO, the γ+jet final state can also be produced through the gluon annihilation process shown in Figure 2.3.

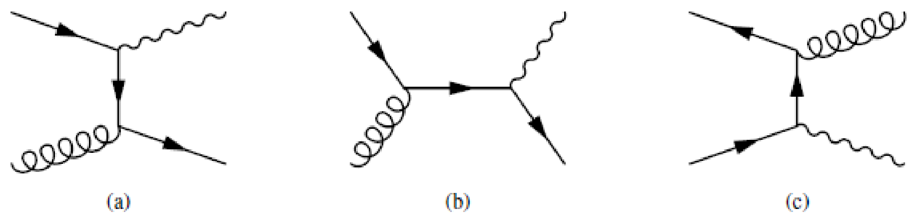


Figure 2.2: Tree-level SM γ+jet production showing the (a) quark-gluon exchange, (b) quark-gluon fusion, and (c) quark-antiquark annihilation process Feynman diagram

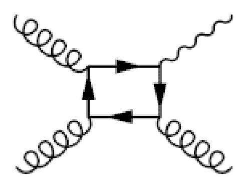


Figure 2.3: Gluon annihilation process for NLO SM γ+jet production (also called "box diagram")

It is sometimes possible for photons to be created in close proximity to a jet. These are called "fragmentation photons", and they are produced when hard-scattering quarks and gluons in a highly energetic jet produce a high energy photon within the high energy jet. In single-jet events, we would see a diphoton final state and simply not select the event. However, in multi-jet events we have multiple candidates for the leading jet, meaning the jet with the highest p_T in the event (the sub-leading jet is defined as having the next highest p_T, and so on). To correctly identify the interesting photon+jet pair, we need to remove these fragmentation photons. As will be shown in Chapter 5, fragmentation photons usually fail the isolation criteria of the analysis, but not always, since they can carry away large fractions of the jet's original energy [8]. However, due to the high cross section of dijet production compared to γ-jet at the LHC [12], these events end up becoming an irreducible background to the analysis.

2.3 The Large Hadron Collider

The Large Hadron Collider (LHC), shown in Figure 2.4, is a particle accelerator located at the European Organization for Nuclear Research (CERN) laboratory, near Geneva, Switzerland. At 27 km in circumference, it is the largest particle accelerator in the world, and is designed to reach energies of up to 14 TeV in center of mass energy for proton-proton collisions. The LHC is capable of delivering bunches of protons in intervals as short as 25 ns, with up to 1.69×10^{11} protons per bunch. The design instantaneous luminosity is 10^{34} cm^{-2} s^{-1} [19]. As of 2014, the LHC has delivered a total integrated luminosity of 28.9 fb^{-1} [20]. [1]

The purpose of the LHC is to conduct searches for physics beyond the Standard Model, as well as perform precision measurements of Standard Model physics at new energies. Many experiments use the LHC, and the accelerator chain that feeds it. There are currently 4 main detectors that operate directly at the LHC: ATLAS, CMS, ALICE, and LHCb. The ATLAS and CMS detectors are general purpose detectors, designed to record events for a broad variety of physics processes. The redundancy in dual general-purpose detectors provides a cross check for all possible new physics discoveries, as was done with the Higgs boson discovery [6,7].

1. "Barn" (1 barn = 10^{-28} m^2) is a measure of particle interaction cross section. Since the (dimensionless) interaction rate is the product of (instantaneous) luminosity and cross section, luminosity can be expressed in units of barns^{-1}/second, and integrated luminosity in barns^{-1}. This measures the size of a dataset.

Figure 2.4: Map of the LHC's geographical position with the four main detectors situated on the LHC ring [35]

The ALICE detector is optimized for heavy ion collisions (either lead-proton or lead-lead), and the LHCb detector is designed primarily for physics resulting from b-quark production. In this thesis, data from the ATLAS detector was used.

2.4 The ATLAS Detector

The ATLAS detector [17,18], an abbreviation for **A** **T**oroidal **L**HC **A**pparatu**s**, is a general purpose physics detector at the LHC. A schematic of the detector and its sub-components is shown in Figure 2.5. The basic hardware components of the ATLAS detector are: the inner detector, the solenoid magnet system, the calorimetry systems, the toroid magnets, and the outer muon detectors [17]. These components are described in more detail in the subsequent sections.

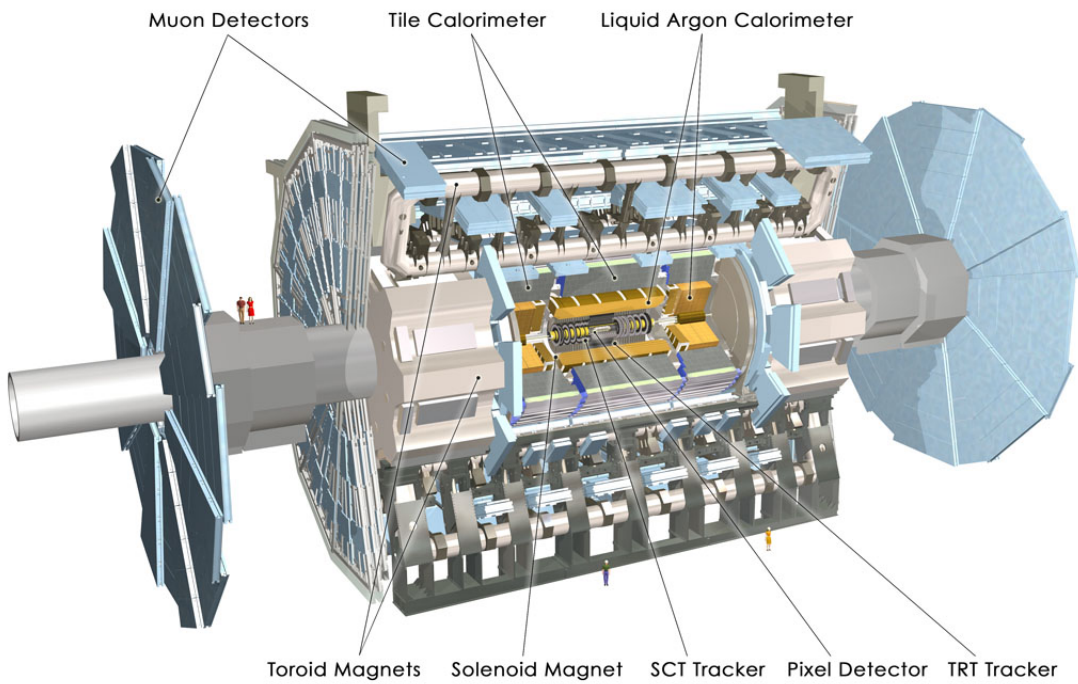


Figure 2.5: An ATLAS detector schematic [32], with human figures for scale. The LHC proton beams enter from the sides of the detector (perpendicular to the disk-shaped muon detectors) and collide at the center of the pixel detector. The ATLAS detector is 46 meters long with a diameter of 25 meters [27].

2.4.1 Coordinate System

The coordinate system defined at ATLAS is as follows: the positive x direction is defined as towards the center of the ring, the positive y is pointing directly up, and the z-axis is defined along the proton beam line (also known as the beam axis), with direction following the right hand rule. The xy-plane is parameterized using ϕ, the azimuthal angle in the xy-plane with respect to the x-axis. In particle physics, it is often convenient to boost the frame of reference from the lab frame to the frame in which the particle of interest travels perpendicular to the beam axis.

The variable used for this is rapidity, and is defined as:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right). \tag{2.1}$$

Here, E is the energy of the particle, and p_z is the momentum of the particle along the beam axis. However, at high energies, where the mass of the particle is negligible ($m << E$), we can use the variable pseudorapidity, η, which is defined as:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{2.2}$$

where θ is the polar angle relative to the z-axis. At high energies pseudorapidity is preferred over rapidity, since it only depends on the position of the particle, and particle production is constant as a function of η. Using this, a distance between two points can be defined in $\eta - \phi$, as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{2.3}$$

Two variables of particular importance in particle physics are the transverse (in the xy-plane) momentum and energy, p_T and E_T[2], defined as:

$$p_T = \frac{|p|}{\cosh\eta},\ E_T = \sqrt{m^2 + p_T^2}, \tag{2.4}$$

where $|p|$ is magnitude of momentum, m is invariant mass, and η is the particle's pseudorapidity. These variables are useful because they measure the momentum and energy in the transverse ($\eta - \phi$) plane. Since incoming particles

2. The unit of measurement for both momentum and energy in this thesis is eV (electron volts), using units where $\hbar = c = 1$.

have no transverse energy or momentum by construction, p_T and E_T in the outgoing particles are a measure of the underlying physical process, as high p_T particles can only result from "hard" scattering.

2.4.2 Inner Detector

When a charged particle passes through a material, it ionizes it. If the resulting ionization can be measured, then the trajectory of the passing particle can be reconstructed. The trajectories of these charged particles are called "tracks", which are reconstructed by collecting the hits in the inner detector subcomponents. The tracks are then used in calculating the momentum of particles, identifying charge, and also to reconstruct the position of the primary vertex. The inner detector is surrounded by a central solenoid magnet with a magnetic field strength of 2 T to curve the charged particles in order to measure their momentum [17]. The bent tracks are matched to calorimeter deposits for particle identification purposes.

To provide adequate precision for the high intensity environment at the LHC, the inner detector is comprised of 3 subcomponents which together cover a total pseudorapidity range of $|\eta| < 2.5$. These systems are: the pixel detector, the silicon microstrip detector (also known as the Semiconductor Tracker, or SCT for short), and the transition radiation tracker (TRT). The pixel detector is used to provide high accuracy and high granularity hits near the interaction point. The SCT provides additional high precision hits in intermediate range from the interaction point. The TRT is used not only for hits, but also for particle identification by recording the transition radiation produced by incoming particles. The high granularity provided by the inner detector's components allows for high-precision position measurements. Together with the magnetic field, this allows for high-precision momentum measurements [17].

The inner detector has a momentum reslution of $\sigma\left(\frac{1}{p_T}\right) = 0.4\,\mathrm{TeV}^{-1}$ at high-p_T, and impact parameter resolutions of $\sigma(d_0) = 11 \bigoplus \frac{60}{p_T\sqrt{\sin\theta}}\ \mu$m, and $\sigma(z_0) = 70 \bigoplus \frac{100}{p_T\sqrt{\sin\theta}}\ \mu$m. The angular resolution is $\sigma(\phi) = 0.08$ mrad, and $\sigma(\theta) \leq 1$ mrad for high-p_T tracks [28].

2.4.3 Calorimeters

The ATLAS detector has two sets of calorimeters: the electromagnetic (EM) calorimeter, and the hadronic calorimeter. Both are sampling calorimeters, filled with dense passive material to produce particle showers, and active detection layers used to measure the energy deposits in the calorimeters.

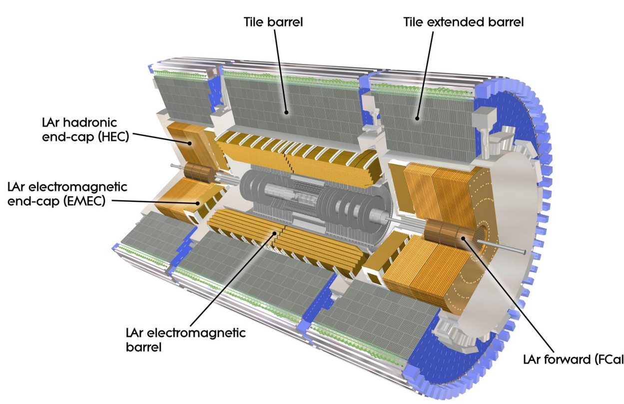


Figure 2.6: A schematic diagram of the ATLAS calorimeter systems with labeled electromagnetic and hadronic sub-systems [33]

The EM calorimeter is divided into two parts: the barrel ($|\eta| < 1.475$), and the endcaps ($1.375 < |\eta| < 3.2$). Thus the EM calorimeter has full ϕ and $|\eta| < 3.2$ coverage, except for a region at $1.37 < |\eta| < 1.52$ (referred to as the "crack") which contains material not instrumented with active detector elements (such as cables and cryogenic services for the calorimeter systems), and is excluded from most analyses. The EM calorimeter uses liquid argon (LAr) as the active material, and lead plates as absorbers. The detector employs an accordion-shaped geometry, as seen in Figure 2.7. This shape provides continuous coverage in ϕ without introducing azimuthal cracks [17]. Below $\eta < 1.8$ the EM calorimeter is preceded by a presampler detector [17]. The high granularity of the presampler makes it useful in compensating for lost energy in the inner detector. The EM calorimeter has an energy resolution of $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \bigoplus 0.7\%$ (order of 1% for 100 GeV) [17].

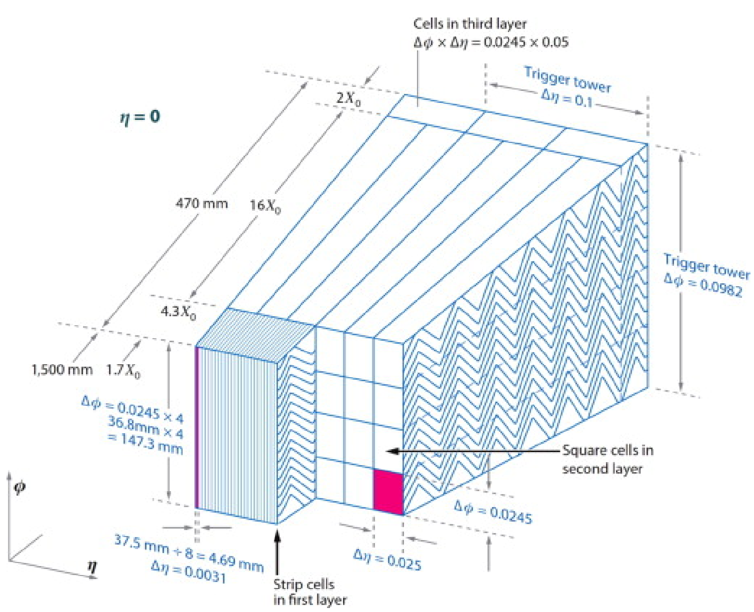


Figure 2.7: An illustration of the accordion-shaped geometry used by the ATLAS EM calorimeter, with the high-granularity presampler in front [37]

The hadronic calorimeter covers the range $|\eta| < 4.9$. It also has a barrel/endcap geometry: the barrel covers the range $|\eta| < 1.7$ and uses plastic scintillator tiles as the active material and iron absorbers. Over $1.5 < |\eta| < 3.2$, the hadronic end-cap range, LAr calorimeters with copper absorbers are used. The forward calorimeters extend to $|\eta| = 4.9$, and also employ LAr calorimeters with copper and tungsten absorbers. The hadronic calorimeter has an energy resolution of $\frac{\sigma_E}{E} = \frac{60\%}{\sqrt{E}}$ [17].

2.4.4 Muon Detectors

The muon spectrometer is a system of detectors designed to track and measure the energy of muons. This is the outermost layer of the ATLAS detector. Muons do not interact strongly and do not radiate, and typically pass through the EM and hadronic calorimeters without depositing a significant amount of energy in them (although these deposits can later be used to reconstruct the track of the muon). The muon momentum is calculated by measuring the curvature of the muon track as it passes through the detector systems. The muon system is also made of a barrel and an endcap, with a total coverage of $|\eta| < 2.7$. The barrel components include monitored drift tubes and resistive plate chambers for precision tracking and triggering. The endcaps use thin gap chambers for triggering and cathode strip chambers for precision measurements.

The muon system surrounds eight toroidal barrel coils and two toroidal endcap magnets, with a peak field of 4 T. The toroidal magnetics are in charge of bending the muons for identification and momentum measurements. The muon system was designed to have a 3% momentum resolution on 3 GeV muons, and increasing to 10% on muons with energy of 1 TeV [17].

2.4.5 Trigger and Data Acquisition

The ATLAS trigger system is responsible for reducing the physics event rate from the 40 MHz bunch crossing frequency to a final recording rate of 200-600 Hz. The ATLAS trigger is the system that decides if an event is of sufficient physics interest to be recorded for long-term storage and analysis. To combat the high rates, a three-level trigger was introduced, with each consecutive level only accepting a subset of the previous level's accepted events. The level-1 (L1) trigger uses information directly from the hardware to make decisions. This information is transfered to the level-2 (L2) trigger to make decisions based on specific regions of the detector. The event-filter (EF) trigger builds entire events using software reconstruction. The L2 and EF are grouped together as the High-Level Trigger (HLT) [19]. The advantage that comes from having a multi-level trigger is that every subsequent step has more time to analyze a candidate event, and can thus employ more sophisticated decision making algorithms.

The L1 trigger is comprised of programmable hardware. A unit called the Central Trigger Processor (CTP) is in charge of analyzing information from the different sub-components (along with the internal LHC clock), and sending decisions to the L2 trigger. The CTP receives information from each sub-component's Local Trigger Processor, which sends Trigger, Timing, and Control signals to the CTP. Using these signals, the L1 trigger collects data in what is called "Regions of Interest" (RoI) for the L2 to consider. These RoI are information from slices of various parts of the detector that will later be used by the HLT to decide whether or not an event should be recorded. The L1 decision rate occurs at a maximum 75 kHz.

At L2, the trigger requests the data fragments in the RoI in order to build ″event fragments″. These event fragments include information such as track and muon identification, and thus more sophisticated algorithms can be used to analyze them. Based on this information, events can be passed to the event filter at a rate of roughly 5 kHz.

Since the latency between events is rather high at this point (50 ms), the event filter builds a fully reconstructed event to analyze, using information from all components of the detector. This is done to match the online event reconstruction to the offline reconstruction as much as possible, but is still within the alloted time between events. The energy deposits from various components of the detector in events that pass trigger criteria are reconstructed as physics objects and categorized as such (muons, jets, photons, etc). The algorithms used to find photons and jets will be discussed in the next chapter. The events are then recorded at a rate on the order of several 100 Hz.

2.5 Previous photon+jet analyses at ATLAS

The work done in this thesis is based on a previous analysis performed at ATLAS [8], where a direct search for resonant states in the photon+jet invariant mass spectrum was performed. The shape of the photon+jet invariant mass distribution in data, as well as the background fit result and potential signal from the excited quark model can be seen in Figure 2.8. The smoothly varying background shape in the photon+jet invariant mass allows for straightforward discrimination against a resonance-like signal. The analysis in this thesis builds upon the analysis in [8], using similar tools and the same Monte Carlo samples, but instead searches for resonant states in an angular observable.

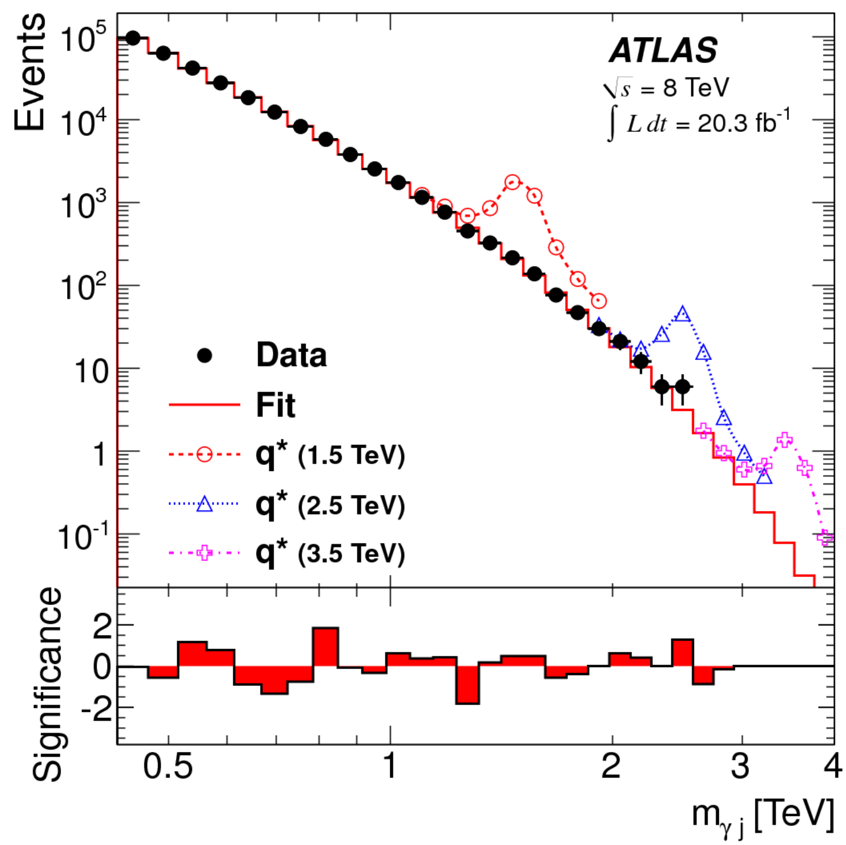


Figure 2.8: Invariant mass of photon+jet pair, with background fit and potential signal samples, and significances per mass bin [8]

One of the challenges that was faced in the photon+jet mass analysis was the characterization of the background. There are limitations in the MC that prevent a direct MC-to-data comparison in the observables. Thus to be able to search for signal in data, a parametrization of background needed to be done while simultaneously including potential signal. The following fit function is chosen to describe the background:

$$f(x = m_{\gamma j}/\sqrt{s}) = p_1\,(1-x)^{p_2}\,x^{-(p_2+p_4 \ln x)} \tag{2.5}$$

The validity of this fit function was tested by the analysis team in the resonant mass search with several Monte Carlo samples, as well as multiple data control regions [8]. It was shown to be appropriate for the invariant mass distributions that were used in this thesis.

3

Photons and Jets at ATLAS

3.1 Photons

In order to identify the photons correctly, characteristics of the energy they deposit in the calorimeters must be used. For example, shower width, depth, and amount of energy deposited in the EM calorimeter compared to the hadronic calorimeter are all looked at to identify photons. Only standard ATLAS selection criteria will be reviewed here, as they were used in our analysis. There are several different sets of criteria which could be used to determine whether or not a calorimeter shower is a photon. The most and least stringent are the *tight* and *loose* photon selections respectively [14]. While the *tight* selection provides a lower fake rate, it also has a lower efficiency, as less energy deposits will pass the photon selection. In this analysis, we exclusively use *tight* photons. The *tight* photon selection requirements are presented in Table 3.1. The cuts on these variables are provided by expert ATLAS groups to the analysis teams. Detailed description of optimization and performance of these selections can be found in [14].

One issue that arises when dealing with photon/jet identification is discriminating between jets and photons. A prime example of this is the $\pi^0 \rightarrow \gamma\gamma$ decay process. The resulting shower in the calorimeter from this decay can be seen in the diagram in Figure 3.1. The fact that the shower is made of two photons makes this process difficult to identify when compared to a single photon. For this reason the high-granularity presampler calorimeter is used to identify the two separate photons produced from the decay process. Primary photons are produced in electroweak processes, while photons from a π^0 are part of hadronic jets. The ability to distinguish between the two processes is important, since we select one final state photon and jet per event in this analysis.

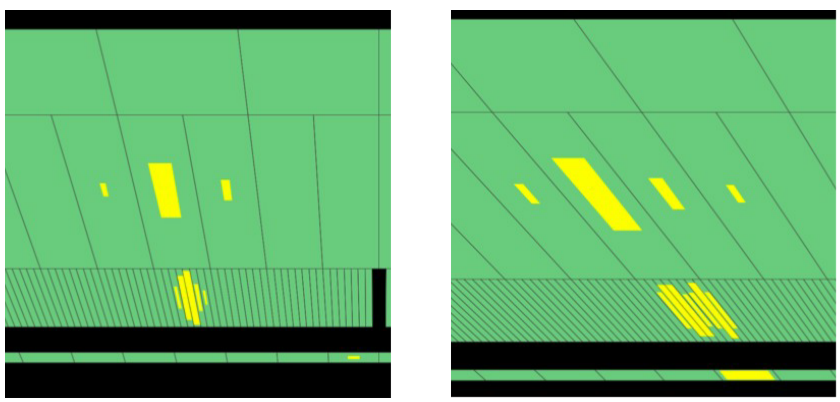


Figure 3.1: A photon signature (left) compared side-by-side to a π^0 jet candidate (right) in the calorimeters. [36]

In this analysis the isolation of the photon is important. We want the photons that ultimately pass our selection criteria to be directly produced in the hard parton interaction. Other photons, usually produced by fragmentation or bremsstrahlung, diminish the sensitivity of the analysis, and are actively removed. These photons are typically either in close proximity to a jet, or have other significant energy deposits close by in the calorimeter. Thus we impose a series of cuts to remove these non-isolated photons to ensure the purity of our sample. This isolation cut is further discussed in Chapter 4.

Category	Description	Name
Acceptance	$\|\eta\| < 2.37$, $1.37 < \|\eta\| < 1.52$ excluded	
Hadronic leakage	Ratio of E_T in the first sampling of the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| < 0.8$ and $\|\eta\| > 1.37$	R_{had_1}
	Ratio of E_T in all the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \|\eta\| < 1.37$)	R_{had}
EM Middle layer	Ratio in η of cell energies in 3×7 versus 7×7 cells	R_η
	Lateral width of shower	$w_{\eta 2}$
	Ratio in ϕ of cell energies in 3×3 and 3×7 cells	R_ϕ
EM Strip layer	Shower width for three strips around strip with maximum energy deposit	w_{s3}
	Total lateral shower width	$w_{s\text{tot}}$
	Energy outside core of three central strips but within seven strips divided by energy within the three central strips	F_{side}
	Difference between the energy associated with the second maximum in the strip layer, and the energy reconstructed in the strip with the minimal value found between the first and second maxima	ΔE
	Ratio of the energy difference associated with the largest and second largest energy deposits over the sum of these energies	E_{ratio}

Table 3.1: Variables used for tight photon identification cuts [14]

3.2 Jets

In ATLAS, jets are identified by the energy deposits in the calorimeters resulting from physical particles produced in hard interaction processes. In this analysis, we are interested in QCD jets, meaning only deposits in the calorimeter from the hadronization of quarks and gluons (partons). Since jets are defined as any energy deposit, not all jets are QCD jets. The hadronization of partons results in localized, eliptical-shaped (in $\eta - \phi$) energy deposits in the calorimeters. In addition to the hadronization process, it is also possible for partons to emit radiation after they are produced (called ”final-state radiation”) before they reach the calorimeters. We use the information from the calorimeters and the inner detector in order to reconstruct the energy and trajectory of the original parton. This is shown schematically in Figure 3.2. While the inner detector's TRT is used for particle ID, tracking is not used in jet reconstruction algorithms.

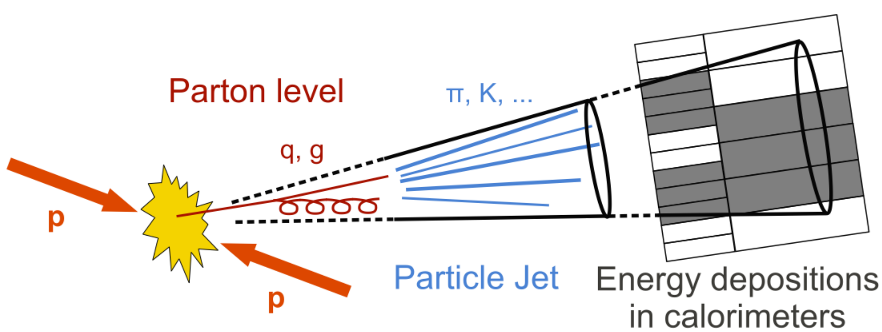


Figure 3.2: Schematic diagram of the evolution of a jet from the partonic level into what is observable in the calorimeter [34].

Jet reconstruction systems in ATLAS use a characteristic radius (defined using the angular ΔR variable) in order to reconstruct jets in the calorimeter. There are both online and offline systems that do this jet building. The offline reconstruction algorithms are more robust, and allow for more delicate analyses to be performed. The main algorithm used in ATLAS is called the anti-k_t algorithm [26], which uses the inverse of transverse momentum with respect to the beam axis as the weight for the calorimeter cells. The anti-k_t algorithm assumes a cone-like shape for the jet, with an origin at the primary vertex. There are several ATLAS-specific standard radii (again in ΔR) for the anti-k_t algorithm: 0.4, 0.6, and 0.8. To remain consistent with previous analyses of similar final states [8,12], this analysis uses the anti-k_t 0.6 algorithm.

4

Analysis of photon+jet final state

4.1 Monte Carlo Samples

We use Monte Carlo excited quark signal and SM background simulations to study the sensitivity of a search which exploits the angular separation between the photon and jet in photon+jet final states. In this chapter we perform a comparison study between LO MC, NLO MC, and ATLAS data to establish the reliability of MC background predictions. Since we rely directly on data for background estimation, we use these MC samples to determine the robustness of our signal extraction method. The MC samples are presented and characterized in Sections 4.1.1-4.2, and MC comparisons to ATLAS data are shown in Section 4.2.2.

The MC samples are generated using the physical cross section of the process being modeled, and are then scaled to data luminosity for comparison. By comparing kinematic distributions of MC to data we show the reliability of the generators in modeling ATLAS data without looking at our signal region in data, thus keeping the analysis blind. The generation is divided into two broad steps, a ”physics” part, and is then followed by a detector simulation, using the Geant4 simulator [23–25]. The physics part is called the ”truth” level, and represents the actual physical process, whereas the detector simulation shows the detector's response to such events, and is called "reconstructed" level.

4.1.1 Standard Model Monte Carlo samples

The direct photon Standard Model processes in this analysis are modeled by three different Monte Carlo generators: Pythia [15], Sherpa [16], and JetPhox [8]. Pythia and Sherpa are fully-simulated, meaning all events are passed through a complete simulation of the ATLAS detector [23]. Pythia simulates events in a 2→2 manner, meaning two particles are used as input, and two particles come out as output. Initial- and final-state radiation are added independently of the matrix element calculation. Pythia is a LO generator [15].

In Sherpa the total number of output particles is variable, typically denoted as 2→ n generation. This allows a higher accuracy in the inclusive jet cross section calculation. This also allows for the parton showering to be done as part of the matrix element calculation, which results in a more physical representation of the output particles [16]. As is done to Pythia, Sherpa is also run through a full detector simulation. Sherpa is considered an NLO generator due to its simulation of multijet events, but can only simulate photon+jet final states to LO.

JetPhox has no interface with detector simulation, producing particle level events only. JetPhox generates strictly $2 \rightarrow 2$ or $2 \rightarrow 3$ events. The advantage of JetPhox is that on top of the fragmentation simulation, JetPhox also simulates the gg box process [8], and is an NLO generator in the photon+jet final state. However, its lack of full event simulation means it is only comparable to the truth-level production of Pythia and Sherpa. This limits the ability to compare JetPhox to data or fully-simulated MC directly.

Pythia and Sherpa are generators that produce γ-jet processes as well as fragmentation photons and initial-/final-state radiation. Due to rapidly decreasing cross section as a function of p_T, the MC production mechanism of Pythia and Sherpa for ATLAS involves separating phase-space into distinct photon p_T ranges. To ensure adequate statistics over the entire phase space, separate samples are generated in these p_T ranges with different equivalent luminosities. To assemble the subsamples produced in the various photon p_T ranges into a cohesive SM direct photon sample, simulated events are assigned weights that get contributions from the cross section and the number of events in each sample. In the case of Pythia, a generator efficiency is included as well. This is due to the fact that Pythia automatically vetoes at the generation step events that are not within the acceptance of the detector [15]. The samples are then recombined (or ”stitched” together) using leading truth photon p_T cuts to produce non-overlapping regions of phase space, with ”leading” defined as the highest p_T photon in the event. The samples and cuts used for Pythia and Sherpa can be found in Tables 4.1 and 4.2 respectively.

Sample ID	p_T^γ range $[GeV]$	Cross section $[pb]$	Events	Luminosity $[pb^{-1}]$
129172	100.0-220.0	3.4250×10^6	2999985	1.535×10^3
129173	220.0-400.0	1.2217×10^5	999994	8.449×10^3
129174	400.0-650.0	3.3487×10^3	999879	2.065×10^5
129175	650.0-1000.0	1.1563×10^2	999875	4.789×10^6
129176	1000.0-1150.0	4.9226×10^0	99895	1.066×10^7
129177	1150.0-1.0E10	8.7493×10^1	99999	6.211×10^7

Table 4.1: Pythia direct photon samples with sample IDs, p_T cuts, generator cross section, number of events, and equivalent integrated luminosity

Sample ID	p_T^γ range $[GeV]$	Cross section $[pb]$	Events	Luminosity $[pb^{-1}]$
113715	95.0 -170.0	2.1530×10^3	2496794	1.160×10^3
113716	170.0- 300.0	1.3785×10^2	1499992	1.088×10^4
113717	300.0 -525.0	5.9627×10^0	999690	1.677×10^5
126371	525.0- 820.0	2.7645×10^1	988873	3.613×10^6
126955	820.0- 1050.0	1.3346×10^2	99999	7.493×10^6
126956	1050.0- 1.0E10	2.3821×10^3	99997	4.198×10^7

Table 4.2: Sherpa direct photon samples with sample IDs, p_T cuts, generator cross section, number of events, and equivalent integrated luminosity

4.1.2 Signal Monte Carlo samples

The signal simulation samples used in this analysis are Pythia excited quark (q*) samples [5]. The samples are passed through a full detector simulation, and only the decay to the photon+jet final state is simulated. The q* model has a single parameter, the assumed mass of the excited quark resonant state, denoted m_{q*}, and are produced via quark-gluon fusion [8]. Each sample covers the entire photon p_T range, and the samples differ in the assumed resonant mass. The datasets used and their corresponding production masses and cross sections can be found in Table 4.3.

Sample ID	Signal Mass [TeV]	$\sigma\times$ BR [nb]	Number of Events
158011	2500	1.0290×10^5	10000
158012	3000	1.9045×10^6	10000
158013	3500	3.5460×10^7	10000
158014	4000	6.5735×10^8	10000
158015	4500	1.2745×10^8	10000

Table 4.3: Excited quark (q*) signal samples with their respective resonant masses and cross sections

4.2 Kinematic Comparisons

Due to the different limitations of each generator in this analysis, and the differences in cross sections calculations (LO vs NLO), it is necessary to compare the performances of the generators to each other. In order to show that the MC samples used are a reasonable approximation of the ATLAS data, comparisons are also performed between MC and data. All MC plots in this thesis are normalized to 20.3 fb^{-1}, which is the total integrated luminosity of ATLAS data in 2012 that satisfy beam and detector quality requirements [21]. This allows for direct comparison of simulation to data.

4.2.1 Monte Carlo validation

At truth-level, we deal with 4-vector representations of particles that result from a matrix element calculation. These vectors are then used as input for reconstruction algorithms. The only criteria we require an event to fulfill are kinematic cuts on the position and energy of the particle, isolation requirements of photons and jets, and removal of fake jets. The values used for the leading jet/photon candidate selection in Table 4.4 were used previously by the photon+jet mass resonance search and have been adopted here unless explicitly mentioned otherwise [8].

The biggest difference is the leading photon η acceptance. Because this angular analysis requires having widely separated photon-jet pairs we extend the photon η acceptance region to $|\eta| < 2.37$. However, the detector performance in detecting photons past the barrel region ($|\eta| > 1.37$) is not as well behaved as central photons due to the crack in the detector mentioned in Section 2.4.3. The kinematic characteristics of only Pythia and Sherpa are compared in this work, as the characteristics of JetPhox have been studied in this context before [9].

Since particles and processes other than jets can produce energy deposits in the calorimeters, it is possible for the anti-k_t algorithm to reconstruct and store a jet candidate that did not arise from the presence of a real QCD jet. These fake jets need to be identified and removed during the analysis stage. The most relevent type of fake jet in this analysis is when the anti-k_t algorithm reconstructs a photon as a jet candidate. This results in a photon candidate list and a jet candidate list which are not mutually exclusive. In order to ensure there is no double-counting, any jet candidate which overlaps in $\Delta R < 0.1$ with a high p_T photon is removed from the list of jet candidates. Since fragmentation photons are typically in close proximity to a high-p_T jet, we remove these photons from the analysis by cutting on the angular separation between the leading photon and the leading jet.

Distributions of various kinematic and angular variables obtained from Pythia and Sherpa are compared in Figure 4.1. The distributions of photon p_T and η agree fairly well between the two generators. However there was a large discrepancy in the leading jet p_T distribution. Based on studies performed, we have concluded that this departure is not a selection-based issue, and is likely a gener-

Jets:	$p_T > 125$ GeV $\lvert\eta\rvert < 2.8$
Photons:	$p_T > 125$ GeV $\lvert\eta\rvert < 2.37$
Isolation and overlap:	$\Delta R_{\gamma j} > 1$ Exactly 1 jet within .1 of leading photon No other jets with $p_T > 30$ GeV within 1 in ΔR of leading jet

Table 4.4: Truth-level cuts used for Pythia and Sherpa direct photon samples

ator effect. Initial- and final-state radiation are entered at different points in the truth-level generation for Pythia and Sherpa. Due to a different effective definition of what constitutes a "leading jet", this would cause a shape difference in the leading jet p_T distribution between Pythia and Sherpa. This effect disappears at the reconstructed level and does not impact the analysis or signal sensitivity, as seen in Figure 4.2. Another issue to note in the jet p_T plots are the single- or few-bin large fluctuations seen throughout the distribution (more visible in the high p_T tail). This is due to the recombination of the samples after generation. Since we do not restrict the leading jet p_T within each sample (like we do with the photons), statistical fluctuations allow for extremely high-p_T jets to be generated. They are, however, physically appropriate and necessary. Each point also has a high error associated with it, making the overall distribution statistically consistent.

Once an event passes the truth-level requirements, a selection is applied with a similar set of cuts at the reconstructed level. Each particle generated at the truth level is passed through a simulation of the ATLAS detector, and reconstructed. These reconstructed particles can then be compared to the truth-level particles, and the effect the detector had on the distributions of the kinematic vari-

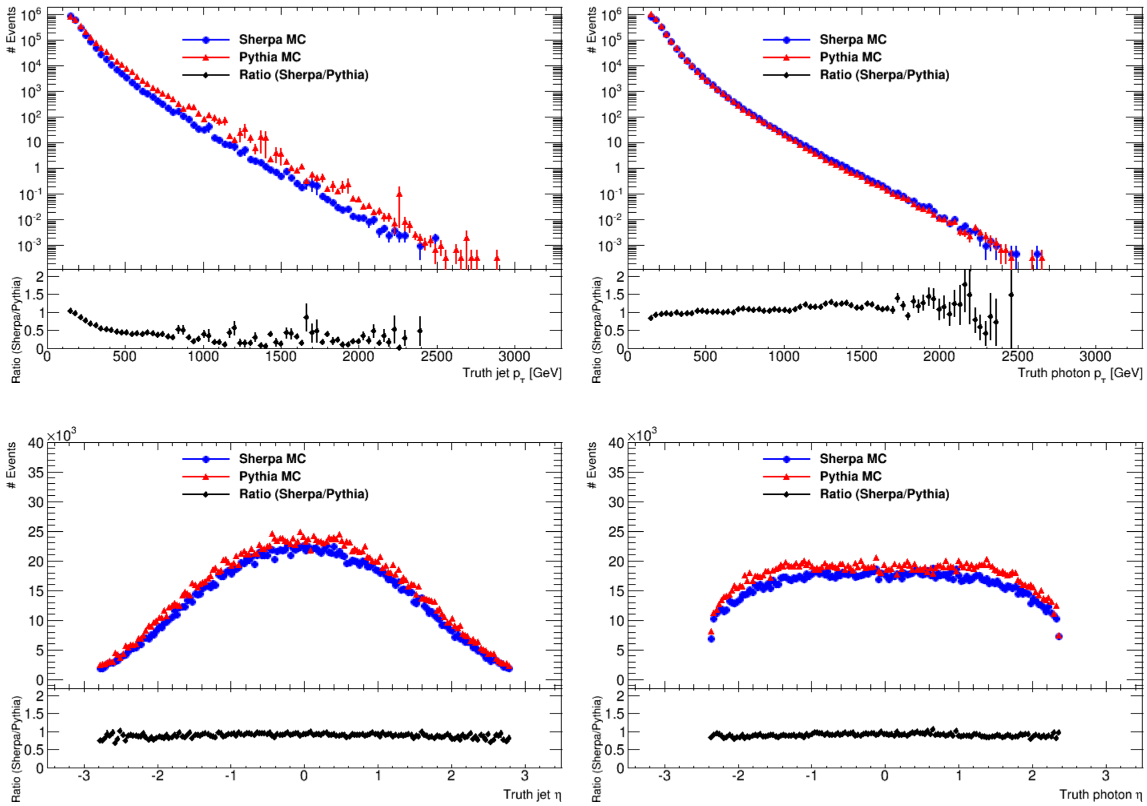
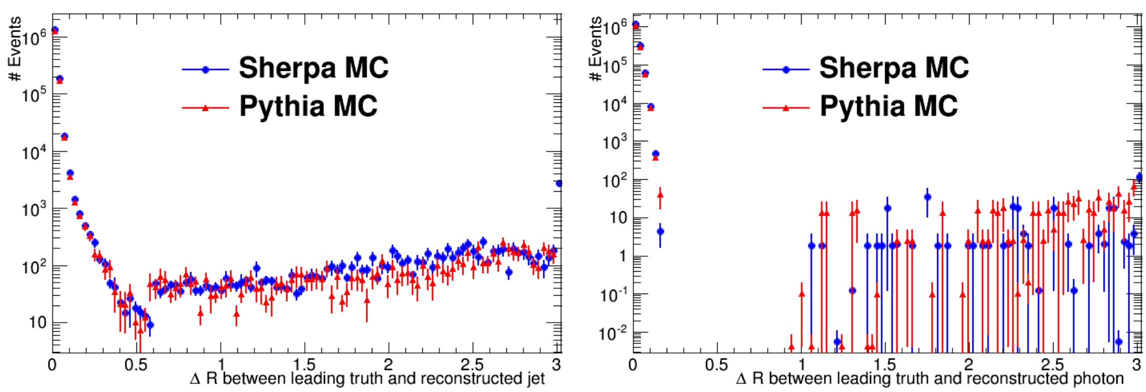


Figure 4.1: Truth-level jet, photon, p_T and η comparisons between Pythia and Sherpa (p_T on log scale,η is linear)

ables can be determined. It is also useful to estimate the efficiency and acceptance of the detector. The cuts used to select events for reconstructed-level comparisons can be found in Table 4.5. When a leading reconstructed photon/jet has been identified, it is then matched to the closest truth-level counterpart by measuring distance in ΔR. For the photons, the isolation requirements to remove fragmentation photons form an implicit cut on the matching criteria. As can be seen in Figure 4.2 (b), any cut of $0.1 < \Delta R < 1$ is effectively the same as a cut of 1.

However, if we plot the ΔR distances between the leading truth and reconstructed jet, again in Figure 4.2 (a), we see a sharp dip, followed by a smoothly decreasing tail before the discontinuity at $\Delta R = 0.6$. This discontinuity is a consequence of the radius of our jet reconstruction algorithm, which is 0.6. We are only interested in the jets that are matched within the 0.6 radius, as these are more likely to be the same object at both truth- and reconstructed-level. Therefore we choose a cut of 0.2 between the leading truth and reconstructed jets for the matching. We also choose a cut of 0.2 for the photons, for consistency, as we do not lose any hard photons with this cut. These matching criteria are added as part of the reconstructed cuts in Table 4.5. If a leading reconstructed jet or a photon cannot be matched to a truth-level counterpart, the event is rejected.



(a) (b)

Figure 4.2: ΔR between leading truth- and leading reconstructed-level (a) jets and (b) photons

Another addition to the reconstructed-level cuts is the "crack removal". The crack is the region $1.37 < |\eta| < 1.52$ between the barrel and endcap of the calorimeter in which there is non-instrumented material, and has a low efficiency of photon identification. We select the highest-p_T photon that does not lie within this region as the leading photon.

Jets:	$p_T > 125$ GeV
	$\|\eta\| < 2.8$
Photons:	$p_T > 125$ GeV
	$\|\eta\| < 2.37$ (excluding crack)
Isolation and overlap:	$\Delta R_{\gamma j} > 1$
	Exactly 1 jet within .1 of leading photon
	No other jets with $p_T > 30$ GeV within 1 in ΔR of leading jet
	Object and event quality requirements, and photon isolation energy criteria
Matching:	Highest $p_T > 30$ GeV truth particle
	within 0.2 in ΔR of leading reconstructed particle

Table 4.5: Matched reconstructed-level cut table

The resulting kinematic distributions of both Pythia and Sherpa are compared at reconstructed-level in Figure 4.3. At lower p_T the photon p_T distributions are in better agreement with each other than the jet p_T distribution, whereas at high p_T the performance is similar. There is also a small structural difference in the shape of the jet η distributions. But as evidenced by the ratio plots in Figure 4.3, the points are close to unity within errors. The deficits seen in the photon η plots are a result of the calorimeter crack removal mentioned previously.

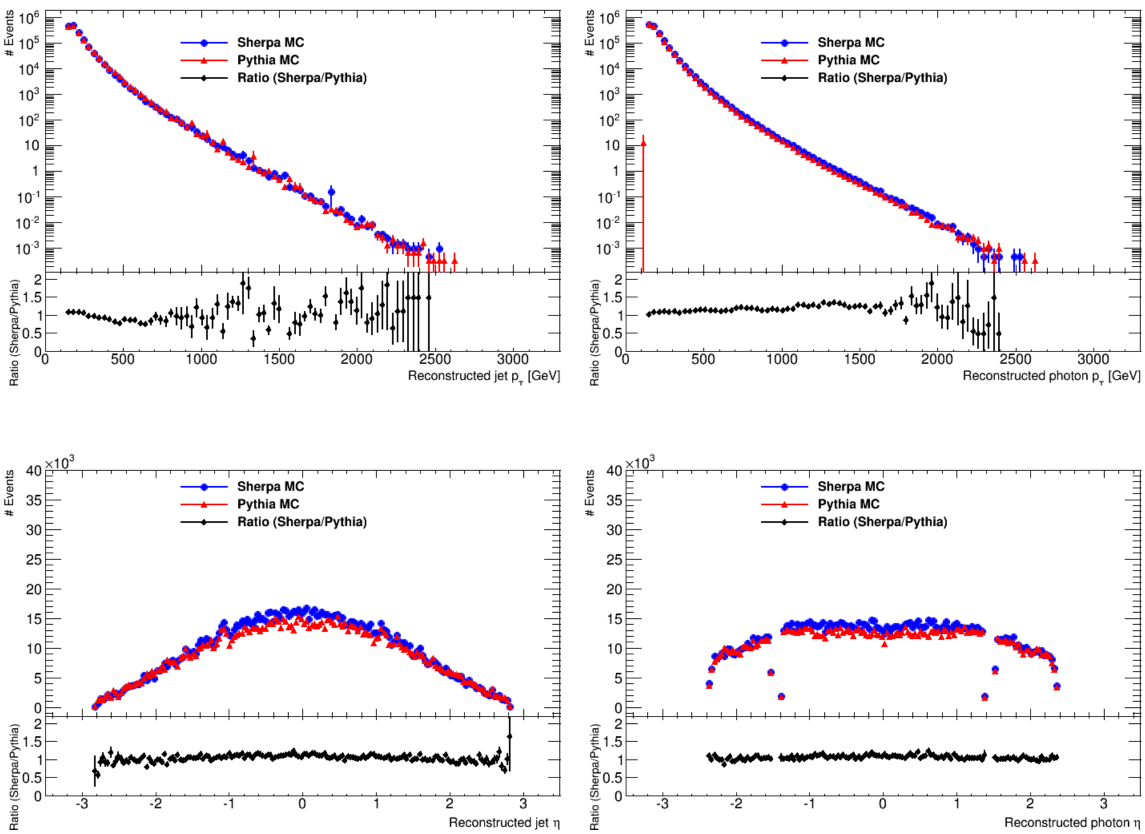


Figure 4.3: Matched reconstructed-level jet, photon, p_T and η comparisons between Pythia and Sherpa (p_T on log scale,η is linear)

4.2.2 Data to Monte Carlo comparisons

The final step in validating the MC samples is to compare their performance to 2012 ATLAS data. The full 2012 dataset (20.3 fb^{-1}) was used for this study. The selection criteria used for this are based on the mass analysis cuts, which can be found in the paper [8]. The only differences are the photon η region being extended to 2.37, and the elimination of the $\Delta\eta$ cut between the leading photon and the leading jet. This last cut is removed since this would result in looking directly at our signal region, introducing potential bias in our signal sensitivity optimization. The kinematic distribution comparisons between data and MC leading photon and jets (p_T and η) shown in Figure 4.4 are sufficient to show the generators' ability to model data. A normalization was done on the full Pythia and Sherpa datasets to match the number of events in MC to the number of events in data. This was done to compensate for the cross section differences between LO generators and data, and doesn't affect the overall shape of the MC distributions. Since our observable is a ratio, shape discrepancies dominate over scale uncertainties, and thus are treated as primary contributions to potential biases.

At lower energies, Sherpa shows better agreement with data than Pythia in jet p_T. In photon p_T, Pythia and Sherpa agree well with data within errors. The structure differences between Pythia and Sherpa, as seen in the ratio plots in Figure 4.4, are due to the fact that Pythia and Sherpa are LO generators in the photon+jet state. Sherpa has slightly more favorable results with less deviation from data in the p_T spectra of both photon and jets. In η, the generators agree with data within errors, and again we see the deficits that result from the calorimeter crack removal. We conclude that both generators are appropriate to use for optimizing the analysis' sensitivity to potential new physics in the 2012 ATLAS dataset. However, it is observed that there are slight shape differences in the η distributions between MC and data. By re-weighting the MC events to match data using the η distributions, we can understand the systematic limitations the MC generators induce on our analysis. This will be shown in Chapter 7.

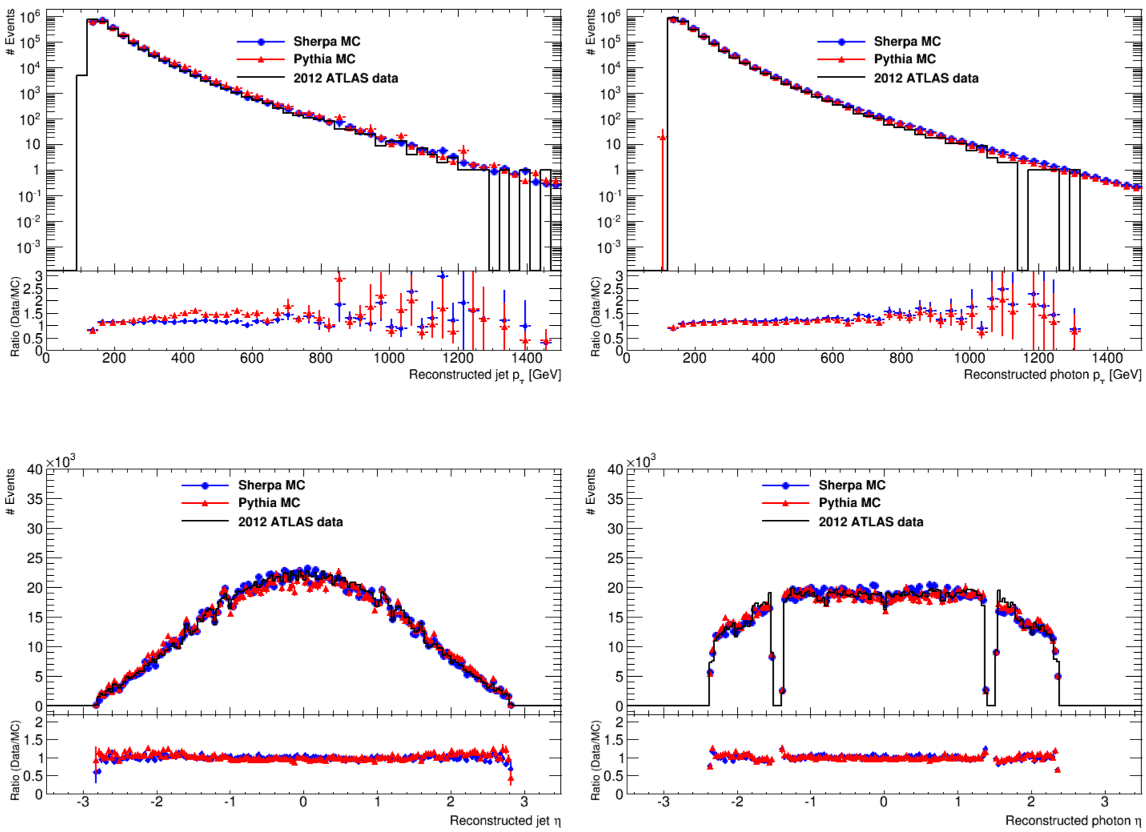


Figure 4.4: The performance of the final Pythia and Sherpa selection is compared to the full 2012 dataset

5

Analysis of angular distributions in photon+jet final state particles

5.1 Introduction

Angular analyses of decay products have been used in the past to probe the sub-structure of known objects. While in modern accelerator-based particle physics experiments these analyses are commonly complementary to direct resonance searches, they generally benefit from being less sensitive to certain types of systematic uncertainties. For ATLAS and the LHC, these would include the jet energy scale (JES) uncertainty, the photon energy scale (γES) uncertainty, luminosity uncertainties, and other similar uncertainties.

In this analysis, our observable is the ratio of narrowly separated events to widely separated events per photon+jet invariant mass range ($m_{\gamma j}$). Instead of using the usual $\eta-\phi$ angular variables, we characterize the angular distance between the leading photon and leading jet using the variable $\psi = \ln\left(\exp(|\eta_\gamma - \eta_j|) + 1\right)$. This is done to linearize the Standard Model photon+jet production cross section. Due to this, the resulting distribution (F_ψ) is flat as a function of $m_{\gamma j}$.

5.2 The ψ variable

Much of the work done in the angular photon+jet analysis is inspired by an analogous dijet angular analysis. In that analysis, a choice of variable, named χ, was made to flatten the cross section of dijet production as a function of this variable with [11,12]:

$$\chi = \exp(|\eta_1 - \eta_2|). \tag{5.1}$$

Here, η_1 and η_2 are the leading and subleading jet η respectively. The dijet cross section as a function of χ can be seen in Figure 5.1. The first step to performing an analogous analysis for the photon+jet channel is to make a choice of variable that presents some of the same advantages as the choice of χ for the dijet analysis. For our analysis, we use the leading photon and leading jet in every event, and define ψ as:

$$\psi = \ln(\chi + 1) = \ln\left(\exp(|\eta_\gamma - \eta_j|) + 1\right). \tag{5.2}$$

The samples used in the photon+jet analysis consist mostly of events with direct photon production, but receive a significant contribution from dijet events where one jet fragments into a photon, carrying away a large fraction of the jet's momentum. This is an irreducible background with a different cross section dependence on ψ than the direct photon production, as shown in Figure 5.2. Thus, any choice of variable that flattens the leading order cross section of SM photon+jet production will not result in the same shape in dijet production. This will invariably result in a non-flat overall SM contribution. However, this does not limit our sensitivity to signal. This is because jets combined with fragmentation photons peak at high values of ψ, whereas signal contributions from the q* model peak at low values of ψ. For different mass ranges, the shape seen in Figure 5.2 does not differ. Whereas the dijet analysis exploited the flatness the overall cross section, in this analysis we use the consistency between different mass ranges.

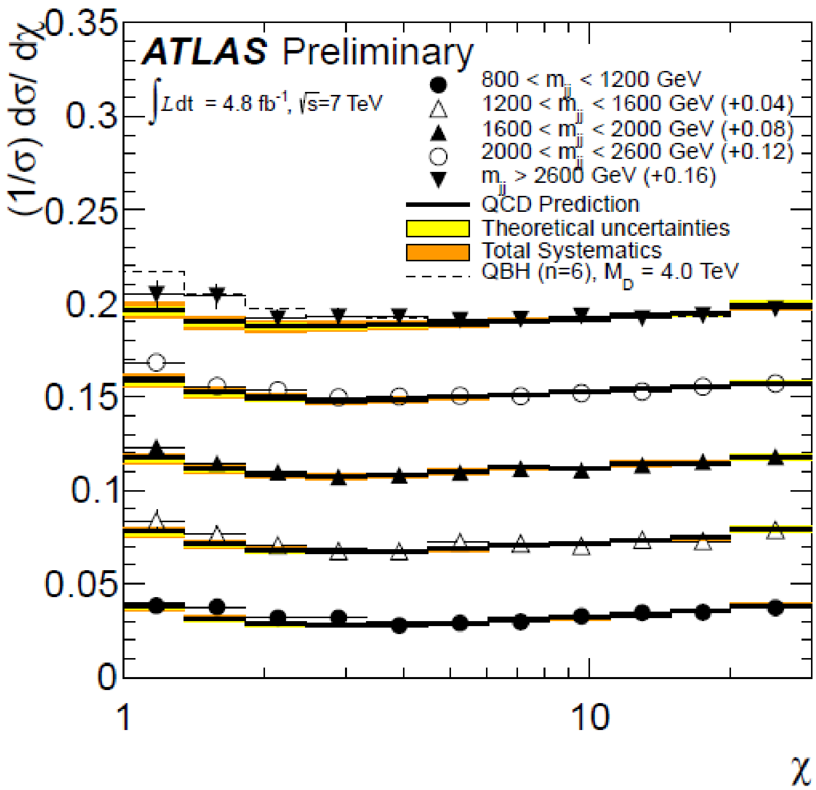
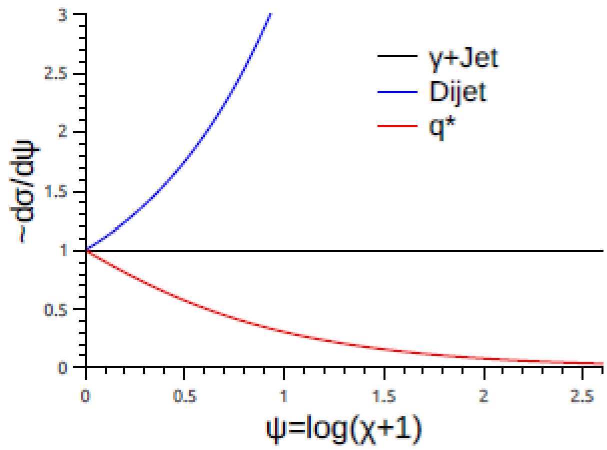
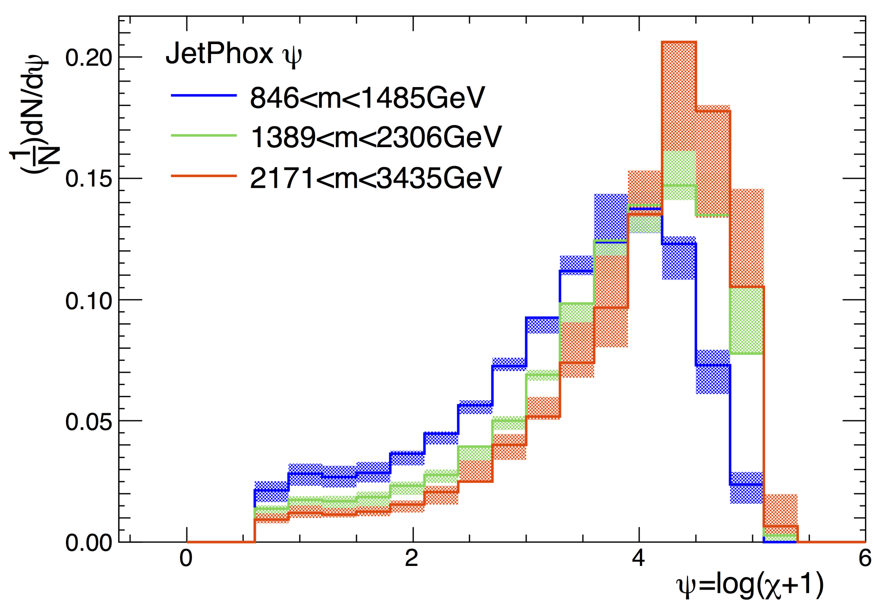


Figure 5.1: The dijet cross section as a function of χ for different dijet invariant mass ranges [12]



(a)



(b)

Figure 5.2: JetPhox simulation showing (a) expected cross section contributions from the leading order processes for gamma+jet (black), dijet (blue) and excited quark (red) production with respect to ψ, and (b) photon+jet events in test mass ranges as a function of ψ [9]

While the ψ variable does not behave in the same way as the dijet's χ, it is still well suited for an angular search. This is because the shape of the cross section production of photon+jet final states does not change much between mass ranges, as shown in Figure 5.2. The proof-of-concept for this technique was studied using JetPhox in [9]. The focus of this thesis was to develop a methodology that allows us to perform a search on ATLAS data, optimized using Pythia and Sherpa.

5.3 Defining our observable distribution

5.3.1 Observable definition using ψ

In order to improve the sensitivity to new physics in the dijet analysis, instead of looking directly at a χ distribution for a particular mass range, the dijet analysis team took the ratio of two different χ cuts, and tested this resulting distribution for new physics [12]. The resulting distribution is called F_χ, and is the ratio between the number of events with narrowly separated decay products to the number of events with widely separated decay products, for a particular range in m_{jj}. As is seen in Figure 5.1, since the individual χ distributions are approximately flat, the resulting ratio between the two is also flat. Taking this ratio for all mass bins, the final F_χ distribution is shown in Figure 5.3 with signal MC. The figure also shows how potential new physics predict contributions in the shape of large departures from the flat QCD background.

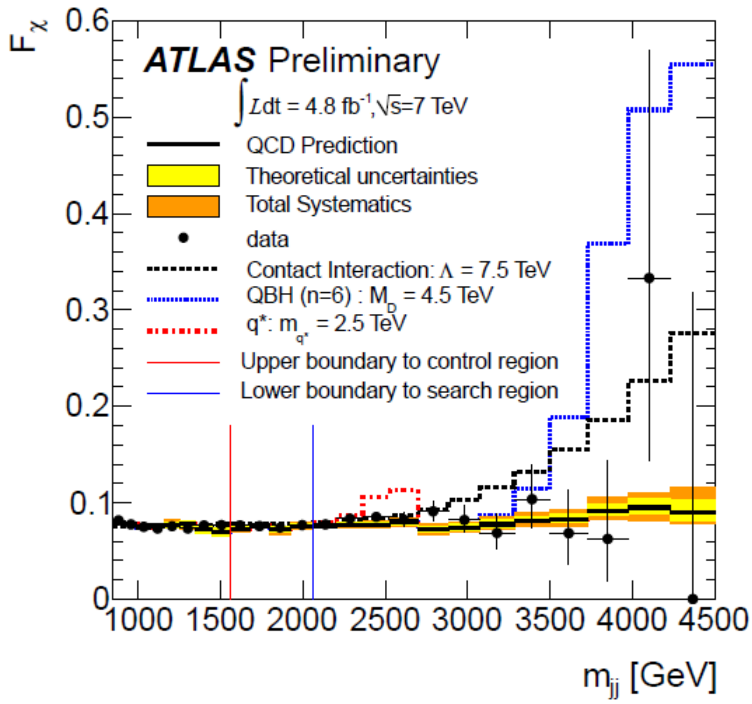


Figure 5.3: The F_χ distribution of ATLAS data plotted with QCD background and potential signals from a variety of new physics contributions [12], as a function of m_{jj}

While the ψ variable flattens the cross section of direct photon+jet production, the influence of the SM dijet background that produces fragmentation photons results in a peaked distribution with respect to ψ. We define $F_\psi(m_{\gamma j})$ to be the ratio of the number of events with wide and narrow separations between decay products, as a function of the invariant mass of the photon+ jet combination, $m_{\gamma j}$:

$$F_\psi(m_{\gamma j}) = \frac{N_{\psi<\psi_n}}{N_{\psi<\psi_d}}, \tag{5.3}$$

where ψ_n is the ψ cut on the numerator, and ψ_d is the ψ cut on the denominator. For clarity, we refer to the photon+jet invariant mass distribution as N_ψ, for a cut value in ψ. Thus the numerator distribution of $F_\psi(m_{\gamma j})$ is referred to as

N_{ψ_n}, the denominator distribution is referred to as N_{ψ_d}, and the ratio between the two is $F_\psi(m_{\gamma j})$. For the remainder of this thesis, $F_\psi(m_{\gamma j})$ will refer to a generic division of two invariant mass distributions, whereas F_ψ will refer specifically to the distribution optimized in this analysis. Thus,

$$F_\psi \equiv F_\psi(m_{\gamma j})|_{\psi_n=1.5,\ \psi_d=5}. \tag{5.4}$$

The values of the ψ cuts were optimized for maximum sensitivity to signal (with the only restriction being that $\psi_d > \psi_n$), and is described in detail in Section 5.3.3. The resulting distribution F_ψ can be seen in Figure 5.4. While F_ψ is not as flat as F_χ, it is relatively smooth after a kinematically biased region at low mass. The high mass region above this bias is the interesting region with respect to searches for new physics, as it has never been searched before. The bias comes from the kinematics of the photon and jet minimum p_T requirements. Since low mass events with high p_T tend to be narrowly separated, imposing a minimum p_T cut will bias the events to be narrowly separated in that region. Since the ATLAS trigger system automatically selects jets and photons that pass a minimum p_T threshold, it is not possible to eliminate this effect. Thus in order to search for new physics, we need to be define a search region that is unaffected by this bias. This study was performed, and is explained in Section 5.3.2.

The immediate benefit of the ratio approach instead of the ψ distribution, or the invariant mass distribution, is that the analysis is potentially less sensitive to scale uncertainties, such as luminosity uncertainties, jet energy scaling, and other similar scaling factors, due to cancellations in the ratio. Since our observable is flat as a function of $m_{\gamma j}$, new physics that generate deviations are straightforward to quantify. The shape of new physics that this approach is most sensitive to is a direct resonance, meaning a narrow mass resonance with a decay width, such as the excited quark model [5]. As shown in Figure 5.4, this would appear as a mass peak on the flat F_ψ distribution, and significance testing software such as BumpHunter [29] could be used for direct detection.

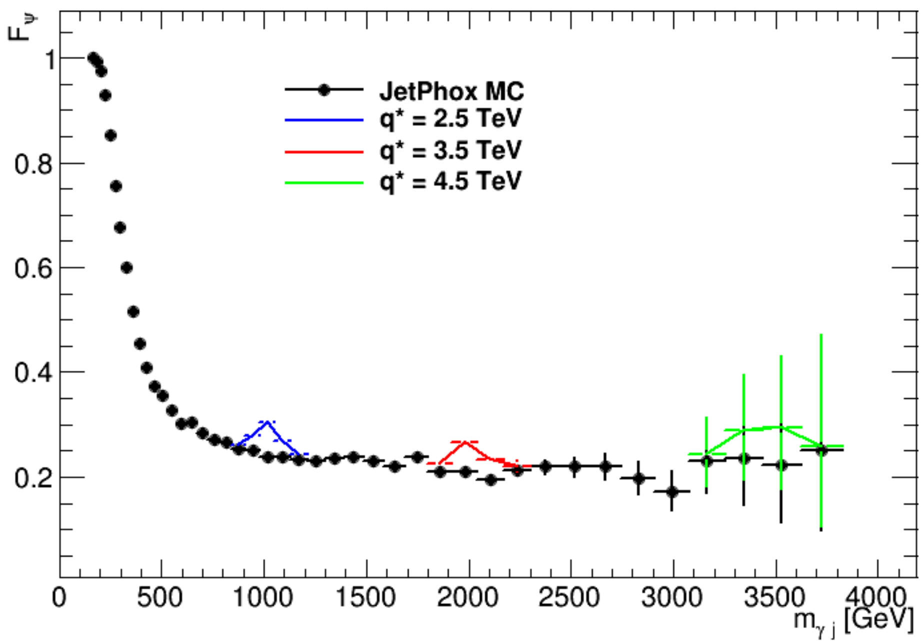


Figure 5.4: JetPhox simulated F_ψ distribution with Pythia q* signal at three test masses: 1 (red), 2 (blue), and 3.5 TeV (green). The scaling of these signal samples is for exemplification only, and is smaller than predicted by the model.

5.3.2 Trigger studies

A trigger cut is a p_T threshold applied to the photon, but in the real-time detector level as opposed to offline in the reconstruction or analysis level. However, the trigger is not 100% efficient with respect to the analysis-level photon p_T

at the threshold value. Thus, the analysis-level photon p_T cut value is chosen to be higher than the trigger threshold. For photons, ATLAS studies have established that the trigger becomes at least 99% efficient when an offline cut value 5 GeV above the trigger threshold value is chosen [22]. Using a near fully-efficient trigger minimizes trigger systematic uncertainties to the level they can be neglected in this analysis. Since at the analysis level we use a 120 GeV photon p$_T$ trigger, we impose a 125 GeV cut on the leading photon p_T.

Due to the kinematic bias at low mass in the F_ψ distribution which results from the photon p_T cut, a study was performed to check whether or not the analysis would benefit from using a lower threshold than the previously stated 125 GeV photon p_T cut. This choice can extend the flat region of F_ψ to lower mass.

In this study we compared different photon p_T thresholds, at: 45, 65, 85, 105, and 125 GeV. The generator used for this particular study was JetPhox. Since there is no detector simulation in JetPhox, a "trigger" is in essence a cut on the truth photon p_T. However, this is a fairly good indicator of the effects the photon p$_T$ cut would induce on the F_ψ distribution, based on our trigger selection. The different F_ψ distributions for each photon p_T threshold can be seen in Figure 5.5.

Since we are interested in searching for new physics on the plateau region of F_ψ, we first need to identify where exactly we are unaffected by the bias of the low mass kinematic region. In order to characterize the kinematic bias of each of these triggers, a set of criteria were selected. A derivative distribution of F_ψ as a

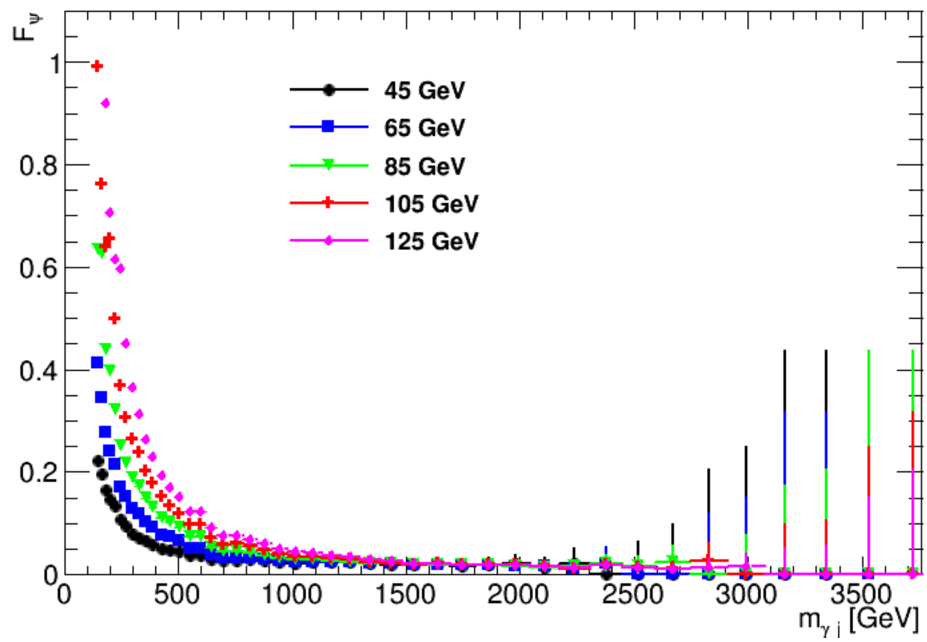


Figure 5.5: A comparison of the photon p_T cut associated to each photon trigger in JetPhox and their effect on the kinematic bias in F_ψ

function of $m_{\gamma j}$ was generated to perform simple optimization, defined as:

$$\frac{dF_\psi}{dm_{\gamma j}} = \frac{F_\psi|_{n+1} - F_\psi|_n}{m_{\gamma j}|_{n+1} - m_{\gamma j}|_n}. \tag{5.5}$$

The derivative distribution of F_ψ for the 125 GeV trigger is shown in Figure 5.6. We then define a set of criteria to reliably identify the first bin of the plateau region. We define it to be the first bin in the derivative distribution of F_ψ to have 10 consecutive bins that satisfy the following criteria:

– 7/10 following bins have dF_ψ value ≥ 0

– 7/10 following bins have dF_ψ value $>$ current bin

The first item comes from simple calculus-based optimization: find the first point where the derivative crosses zero. However, this alone does not take into account points at which F_ψ is still too unstable to be considered "smooth". Thus we introduce the second requirement, making sure that, overall, the F_ψ distribution is not still decreasing dramatically at this point. The resulting plateau thresholds for each trigger was found, with the results presented in Table 5.1.

p_T cut:	45 GeV	65 GeV	85 GeV	105 GeV	125 GeV
Threshold:	421 GeV	892 GeV	1104 GeV	1204 GeV	1191 GeV

Table 5.1: JetPhox photon p_T cut for each trigger with its respective plateau threshold in $m_{\gamma j}$

The kinematic bias at low mass does not allow us to search for new physics in this region. Thus, how far away from it we are when we *do* search for new physics is extremely important, and influences the ability to do data driven background estimation in this analysis. The presence of extremely high statistics at lower masses means that any fitting technique will be heavily influenced by the first few low-mass bins. Therefore, we do not want to include any of these influences in our search region, since it will affect the significance of any possible signal. Based on the trigger selection, we established a quantitative method to determine the viable search region for new physics in F_ψ. We use this method to demonstrate that the low end of the search region can be extended to 421 GeV from a high 1191 GeV by choosing a different trigger. However, for simplicity in comparison with the published resonance search [8], the nominal 120 GeV trigger is used throughout the rest of the thesis.

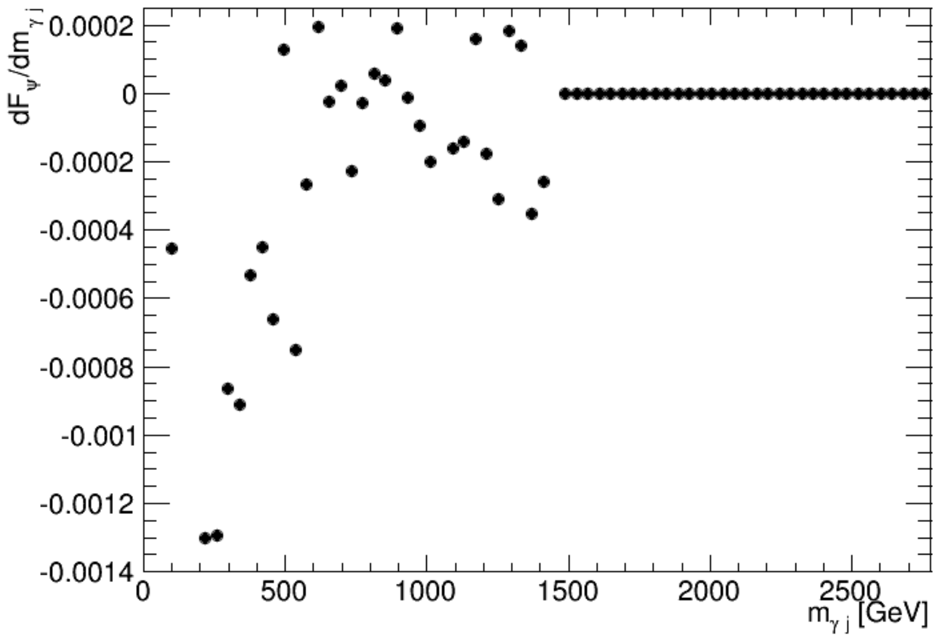


Figure 5.6: The value of the derivative as a function of mass for the 125 GeV photon p_T cut in JetPhox

5.3.3 ψ cut optimization study

After $F_\psi(m_{\gamma j})$ was formally defined, we optimized the selection of ψ_n and ψ_d. The signal samples used were the excited quark (q*) Pythia samples (Table 4.3). The photon+jet Pythia (Table 4.1) and Sherpa (Table 4.2) samples were used as background. In order to test the dependence of the optimal ψ cuts on the value of the resonant mass of the q* model, several q* masses (m_{q*}) were used: 1 TeV, 2 TeV, and 3.5 TeV. The optimization was done using a 2-dimensional sensitivity study in ψ, varying both the numerator and denominator ψ cuts. The background samples were scaled to 10^6 events (approximately the amount of events expected to be present in 2012 ATLAS data sample), and the signal fraction, defined as the amount of signal events divided by the amount of background events, to 10^{-15}.

The sensitivity statistic ξ is defined as:

$$\xi = \frac{\Sigma_{sig}}{\sigma_{bkg}} \tag{5.6}$$

To calculate sensitivity to signal we quantified the excess present above the background in F_ψ, Σ_{sig}, and the uncertainty of the background, σ_{bkg}. The amount of signal is defined as the difference in F_ψ with and without the injection of signal:

$$\Sigma_{sig} = \frac{\frac{N_s}{D_s} - \frac{N_b}{D_b}}{1 + \frac{D_b}{D_s}}. \tag{5.7}$$

Where N and D represent the number of events in the numerator and denominator distributions respectively, and subscript s and b signify the number of signal or background events respectively. For these samples, fully simulated events that pass our selection criteria from Table 4.5 were used. In order to obtain the amount of either signal or background (N_s, N_b, etc.), a 2-dimensional integral over ψ and $m_{\gamma j}$ is performed. This is done iteratively for values in both ψ_n and ψ_d, ranging from 0 to 6 in ψ (in steps of 0.1), again adhering to the restriction that $\psi_d > \psi_n$. Using these sums we then calculate the background uncertainty, σ_{bkg}, using Poisson errors:

$$\sigma_{bkg} = \frac{N_b}{D_b}\sqrt{\frac{1}{N_b} + \frac{1}{D_b}} \tag{5.8}$$

After each time the sensitivity was calculated, the amount of signal events in the sample was scaled by $10^{1/4}$. This was done iteratively until the signal fraction was 1%. While this is many more events than any likely signal, it allows for more robustness in the method.

An example of the resulting sensitivity plot as a function of both numerator and denominator ψ cut is shown in Figure 5.7. The most visible effect is the "forbidden region", which is across the main diagonal of the graph, which comes from the requirement that $\psi_n < \psi_d$. The 3-dimensional graph shows a clear narrow range in the value of ψ_n where the sensitivity is maximal. For ψ_d we observe that as the value of ψ_d is increased, the sensitivity gradually increases up to a threshold beyond which no further gain in the sensitivity is observed. This is a purely kinematic effect and is a result of the fact that past this threshold (roughly at 5 in ψ_d), there are very few events added to the denominator since such widely separated events are rare, and thus no change is seen in sensitivity.

As an example, Figure 5.8 shows the optimal ψ_d (blue) and ψ_n (black) for each iteration of the study for one particular sample ($m_{q*} = 2$ TeV). At each step, the optimal cuts remained virtually constant. The values of all optimal ψ_n and ψ_d cuts (per q* mass) are given in Table 5.2. Pythia and Sherpa agreed well on the optimal values of ψ per m_{q*} using this optimization method.

m_{q*} [GeV]	1000		2000		3500	
	ψ_d	ψ_n	ψ_d	ψ_n	ψ_d	ψ_n
Pythia	4	1.3	5	1.6	4.9	1.3
Sherpa	4.1	1.6	4.9	1.6	5.1	1.4

Table 5.2: Table of optimal ψ_n and ψ_d cuts for Pythia and Sherpa per m_{q*}

In order to maximize sensitivity to signal independently of the assumed resonant q* mass, and at the same time remaining unbiased in generator choice, cut values of $\psi_n = 1.5$ and $\psi_d = 5$ are chosen.

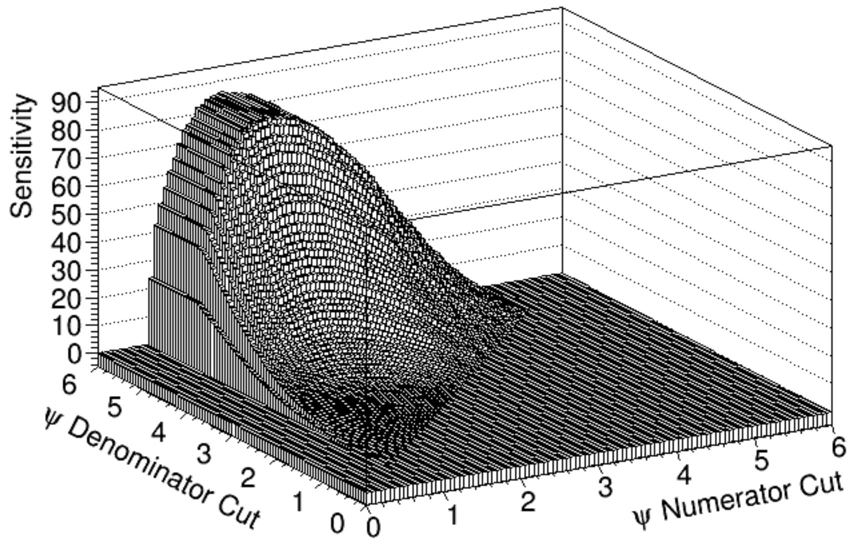


Figure 5.7: A three dimensional plot of the sensitivity to signal/background (here = 1%) for m_q* = 2000 GeV in Sherpa, plotted as a function of both ψ_n and ψ_d

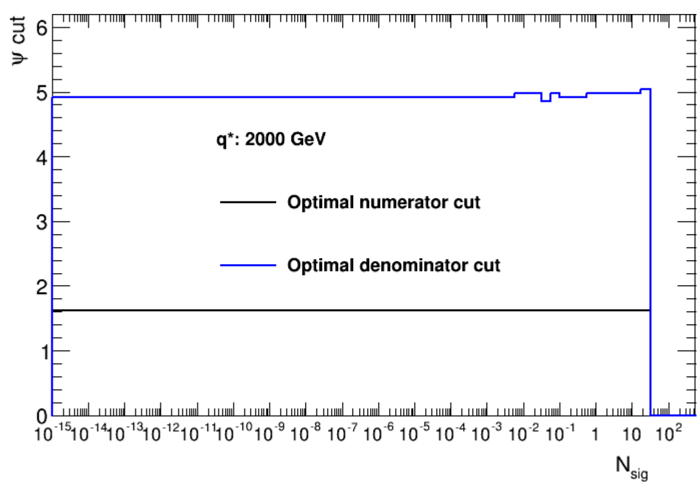
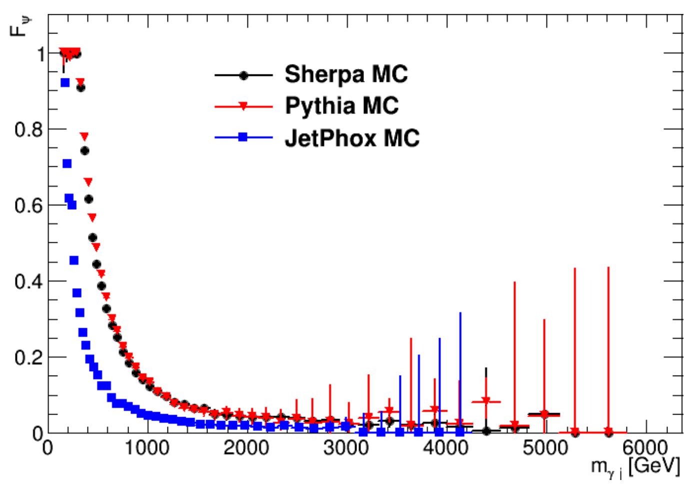


Figure 5.8: The optimal numerator and denominator cuts as a function of signal for the m_{q*} = 2000 GeV in Sherpa

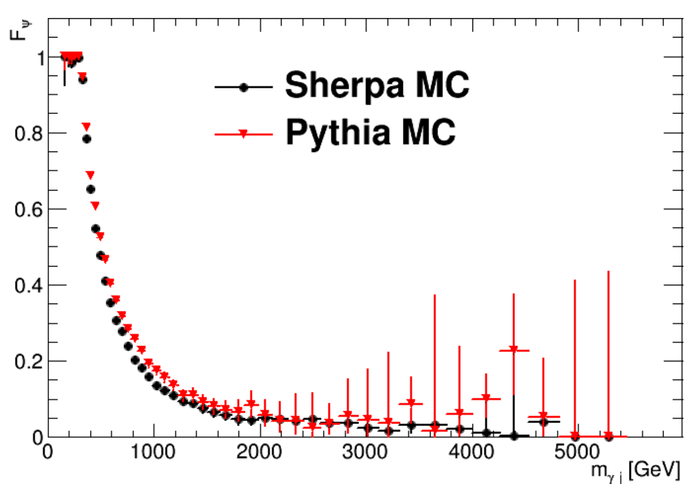
5.4 F_ψ comparisons

Once the selection criteria and optimization of the F_ψ distributions have been finalized, the performance of all three generators are compared. The F_ψ distributions from each generator are plotted together in Figure 5.9. As JetPhox cannot be interfaced with a detector simulation, JetPhox events are compared to truth-level Pythia and Sherpa events. It is seen that there is a large discrepancy in the F_ψ prediction between Pythia, Sherpa and JetPhox at truth-level. As JetPhox is the only NLO generator in photon+jet, we can conclude that the NLO contributions result in a different shape compared to the LO generators Pythia and Sherpa. However, Pythia and Sherpa agree well with each other at truth-level, with slight differences at the reconstructed-level. Specifically, the value in F_ψ at which the plateau stabilizes is slightly different at the reconstructed-level. Sherpa shows little change, while Pythia shifts upwards. The likeliest explanation for this is found by looking at the leading truth jet p_T distributions for Pythia. In comparison to Sherpa, the spectrum has more energetic leading jets. This would likely cause more events to be accepted in the denominator distribution, thus shifting the plateau value down in F_ψ. These differences in jet p_T and $|\eta|$ between the generators were shown and discussed previously in Chapter 4. However, since the shapes between Pythia, Sherpa, and data agreed well after scaling the MC to match the number of events in data, it is reasonable to conclude that the shape of F_ψ would also be similar between MC and data. ATLAS data is not compared in the observable F_ψ since that would unblind the analysis and is beyond the scope of this thesis (and would also require ATLAS approval).

The conclusion that can be drawn from the F_ψ comparison is that all the generators that were tested produced flat F_ψ distributions. Therefore all the generators used in this thesis are sensitive to signal, and can be used for the optimization of the analysis technique. It is noteworthy that since there is a noticeable difference between the LO and NLO F_ψ distributions, ATLAS would benefit from writing a detector simulation interface for JetPhox.



(a)



(b)

Figure 5.9: F_ψ plotted as function of $m_{\gamma j}$ for (a) truth- and (b) reconstructed-level events

6

Background estimation

6.1 Fitting the F_ψ distribution

Due to the fact that we lack an interface between NLO MC and detector simulation, the expectation is that MC will not correctly model data in a direct comparison in our observable, F_ψ. Thus we develop a method to estimate the background from data while simultaneously determining the signal yield. There are several ways to do such an estimation, the simplest of which is fitting a function representing both the signal and background components directly to the F_ψ distribution. A particular characteristic of the F_ψ distribution that makes this challenging is that, as we go higher in $m_{\gamma j}$, there is an exponential loss in entries in histogram bins, and consequently the statistical uncertainty on each bin is exponentially higher. This results in the first several bins of the plateau region dominating the determination of the the fit parameters (each progressive bin contributes less and less to the χ^2 minimization). This becomes a problem once signal is introduced in the sample, because at higher masses the fit function will either contort itself to fit signal points, or will not accurately describe the data at all, depending on the choice of fit function. Since the distribution is expected to be smooth, but not completely flat, the initial fit function chosen was:

$$f(x = m_{\gamma j}/\sqrt{s}) = a + b(1-x)^c \tag{6.1}$$

Where $a, b,$ and c are fit parameters, and $\sqrt{s}$ is the LHC center-of-mass energy (in these samples $\sqrt{s} = 8$ TeV). The result of the direct fitting procedure with Equation 6.1 of a JetPhox SM sample using an 125 GeV leading photon p_T cut is shown in Figure 6.1.

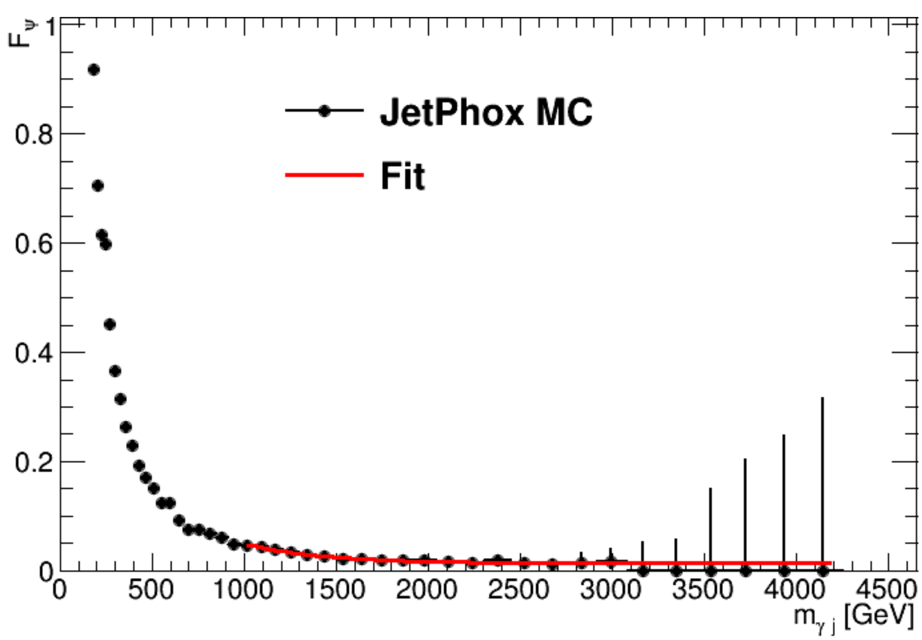


Figure 6.1: JetPhox F_ψ background simulation with statistical errors only plotted as a function of $m_{\gamma j}$ using the 125 photon p_T cut with the fit obtained from Equation 6.1

It is important to note that the χ^2 goodness-of-fit test is not useful in evaluating this fit, since the errors on the F_ψ bins are binomial, not Gaussian. At first glance the fit looks qualitatively adequate. However, a complication arises once signal is introduced into the sample. Due to exponential growth of errors in F_ψ, the fit becomes an inappropriate estimation of background in masses past m_{q*}. An example of the contortion of the fit function can be seen in Figure 6.2. Pythia excited quark signal was used, with $m_{q*} = 2$ TeV.

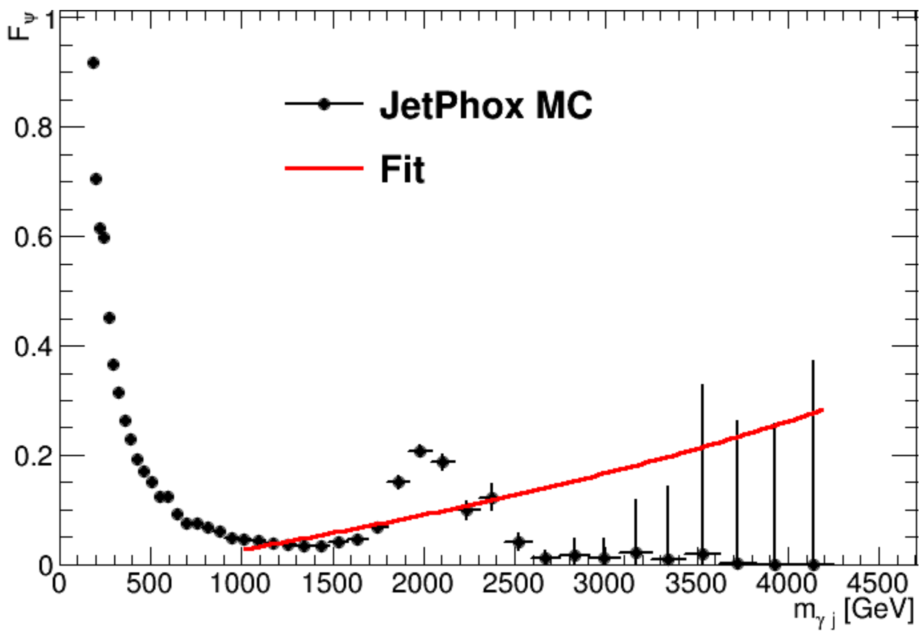


Figure 6.2: JetPhox F_ψ with 125 GeV photon p_T cut, plotted as a function of $m_{\gamma j}$ with m_{q*} = 2 TeV signal, fitted using Equation 6.1

By eye it can be seen in Figure 6.2 that most of the bins past the m_{q*} are very inadequately described by the fit function once signal is introduced. This is a large departure from what was seen in Figure 6.1. This stems from the fact that in data the F_ψ distribution could be either monotonically decreasing or increasing, but is roughly flat. Since we do not know *a priori* what the case will be, the fit function needs to be able to describe both, which is why this method is influenced by the presence of signal. This needs to be done without sacrificing signal sensitivity, or the function's ability to describe the background with the addition of signal. Since Figure 6.2 shows that Equation 6.1 is inadequate in describing background past

m_{q*}, another approach was considered. In order to better describe the F_ψ distribution, a slight modification of the function in Equation 6.1 is done to describe the edge of the kinematic region at low mass. Since bins in this region have roughly similar statistics as the beginning of the plateau region, the ankle of the kinematic bias region can be used to anchor the fit. Since we assume that the number of events in each bin decreases exponentially, we can add an exponential component to the original fit function. The following modified fit function was introduced and tested in Figure 6.3:

$$f(x = m_{\gamma j}/\sqrt{s}) = \exp(-a \cdot x) \cdot \left((1-x)^b + c\right) + d \tag{6.2}$$

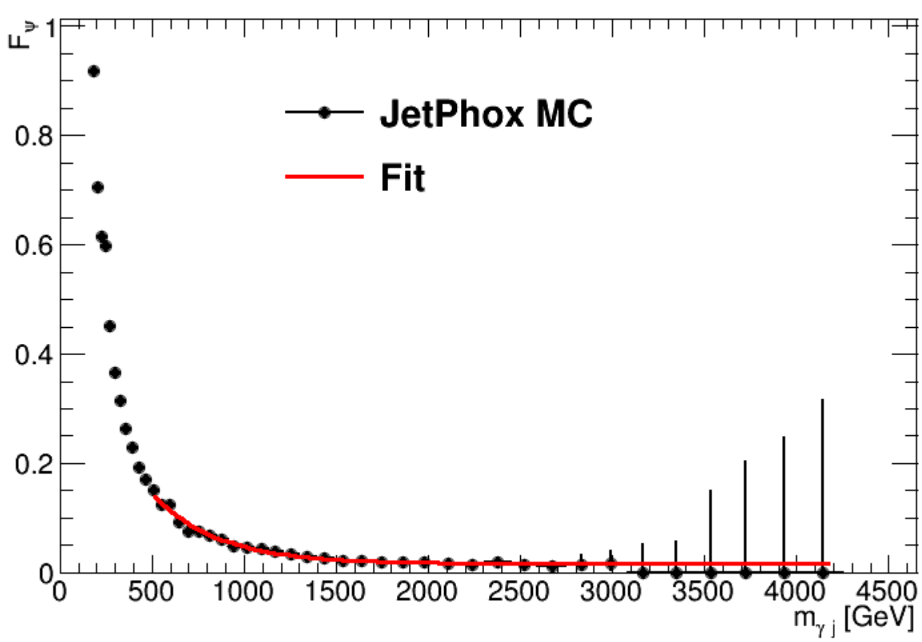


Figure 6.3: JetPhox F_ψ simulation using the 125 GeV trigger and modified background fit function from Equation 6.2

While by eye, again, this modification does seem to anchor the fit, the fact that a non-physical region was used in order to describe a physically meaningful region made the choice of fit function difficult to justify.

A last variation using direct fitting was tested. This method used the assumption that at high mass the plateau was either flat or smooth, so we can fit a linear function to the F_ψ distribution. In this case, we start with the *end* of the F_ψ spectrum (the first non-zero entry in F_ψ in high mass), and we use the linear function to fit as many bins as possible going back down in $m_{\gamma j}$. This was done on background Monte Carlo only. The result, however, produced a search region that was either too small in range of $m_{\gamma j}$ to be useful, or was still heavily influenced by the lowest $m_{\gamma j}$ bins, and improperly described the higher mass bins. When signal was injected into the background MC samples the problem was exacerbated, as can be seen in Figure 6.4. Since sensitivity to signal is of vital importance, this method was rejected. None of the direct fitting methods investigated produced background estimates that are stable against the choice of fit range and reliable in the presence of signal, so the direct fitting approach was rejected.

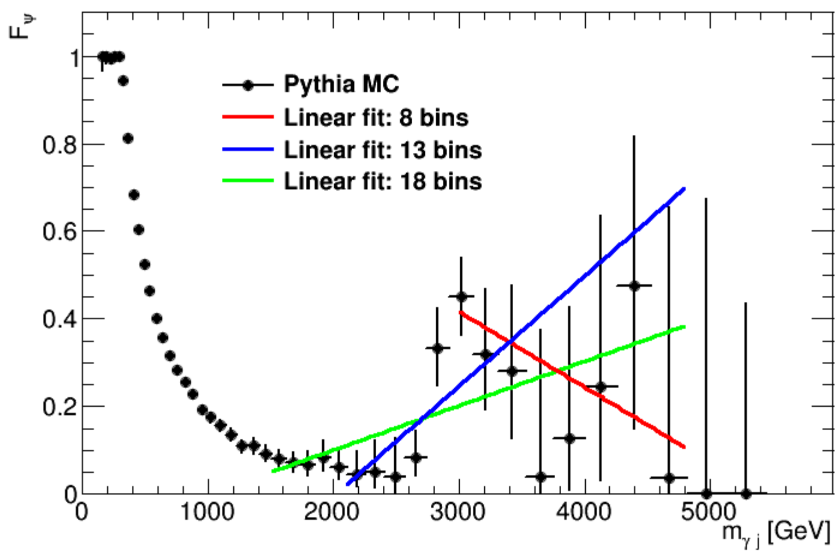


Figure 6.4: F_ψ plotted as a function of $m_{\gamma j}$ using Pythia background injected with $m_{q*} =$ 3000 GeV signal, with multiple linear fits

6.2 Fitting using mass distributions

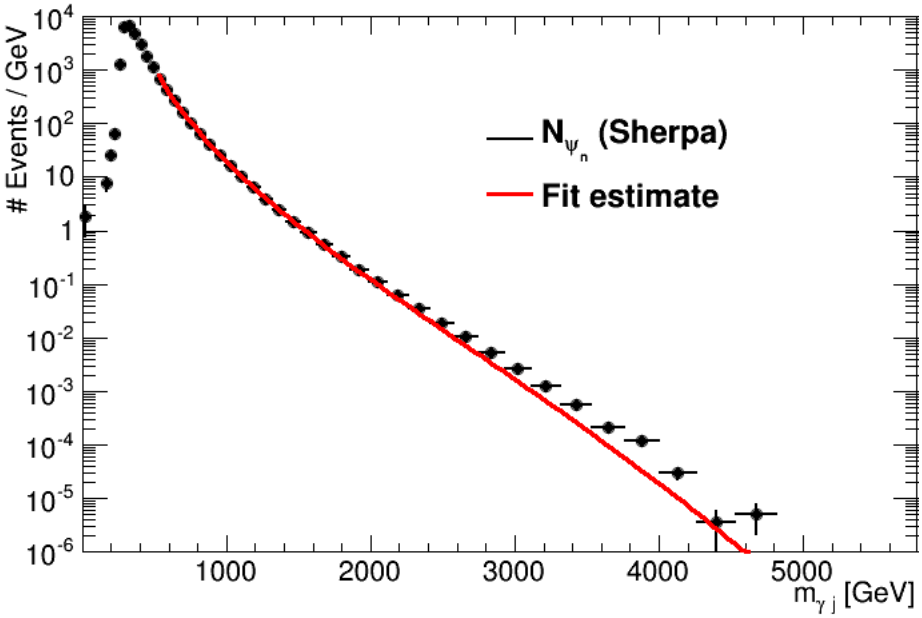
6.2.1 Motivation and calculation of fits

The F_ψ distribution is simply a ratio of two different invariant mass distributions of photon+jet objects, as defined in Section 5.3.1. Since ψ is simply a change of variable from $\Delta\eta$ between the leading photon and leading jet, the numerator distribution is very similar in shape to the photon+jet invariant mass distribution in the published invariant mass search [8]. That analysis also uses a data driven background estimation, using the following fit function [8]:

$$f(x = m_{\gamma j}/\sqrt{s}) = \frac{a(1-x)^b}{x^{\,c+d\ln x}} \tag{6.3}$$

Since the validity of this fit function is well documented (at least for photons with $|\eta| < 1.37$ [8]), a new approach was tested: it could be possible to fit the N_{ψ_n}

and N_{ψ_d} distributions separately, and then divide the fits in order to obtain an estimate of the background in F_ψ. However, due to the variable detector mass resolution as a function of particle energy, the N_ψ distributions have non-uniform bin widths in $m_{\gamma j}$. Therefore each bin was divided by its respective bin width prior to the fitting. An example of how the mass analysis fit function describes the angular distributions can be seen in Figure 6.5. This fit function has been demonstrated to describe the N_ψ distributions well even in the presence of q* signal in [8], and thus suits the needs of this analysis in that regard.



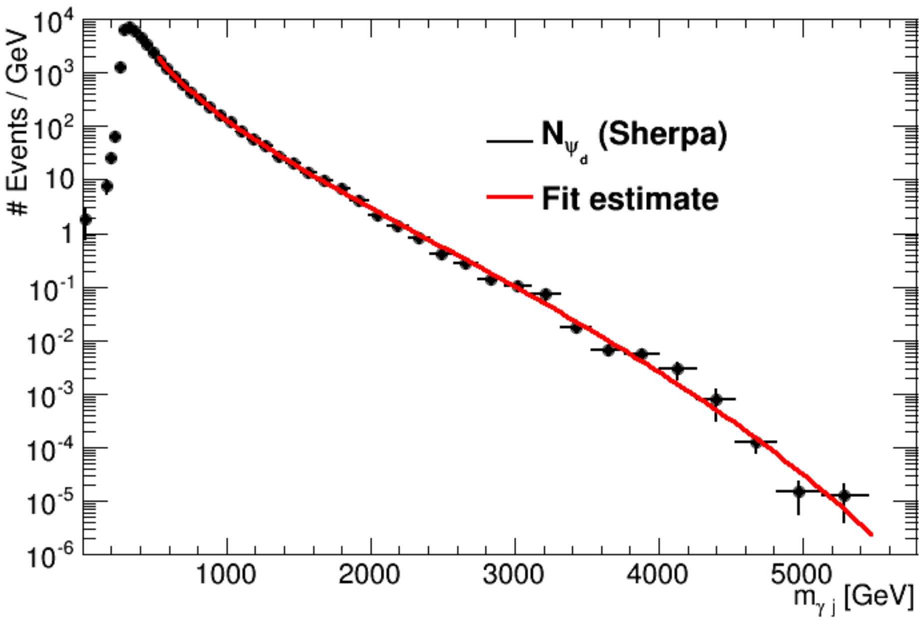


Figure 6.5: Sherpa N_{ψ_n} (top) and N_{ψ_d} (bottom) photon+jet invariant mass distributions fitted with Equation 6.3

Once N_{ψ_d} and N_{ψ_n} were fit, a new histogram was filled by evaluating the fit function at the central value of every bin, thus estimating the bin content at that point. These two new histograms were then divided, producing the estimate of the background contribution to F_{ψ}. Due to exponentially decreasing N_{ψ} distributions, a potential bias could occur in the final estimation of F_{ψ} depending on where the fit was evaluated in each bin. This potential bias was tested by evaluating the fits at three different points: on the bin's low edge, the center of the bin, and at $1/e$ from the bin's low edge. Another method used to test this bias was to integrate the N_{ψ} fits across each mass bin to obtain the fit estimate for that mass bin. However upon comparison, the fit value for the bin using any of these methods were compatible with each other within the fit's statistical power. For simplicity, the chosen method was to evaluate the fits at the center of each mass bin. The final estimation of F_{ψ} is shown in Figure 6.6, without error bars on the points. The correct statistical treatment of correlated errors is introduced and explained in Section 6.2.2. It is noted that in Figure 6.5, in the range past 2.5 TeV, some of the points in N_{ψ_n} are underestimated by the fit. However when calculating F_{ψ} in Figure 6.6, at this high mass we are limited by the statistical uncertainties of the F_{ψ} distribution and the background estimation falls well within these uncertainties. The seemingly higher uncertainties in F_{ψ} compared to N_{ψ} are a consequence of dividing the N_{ψ} distributions.

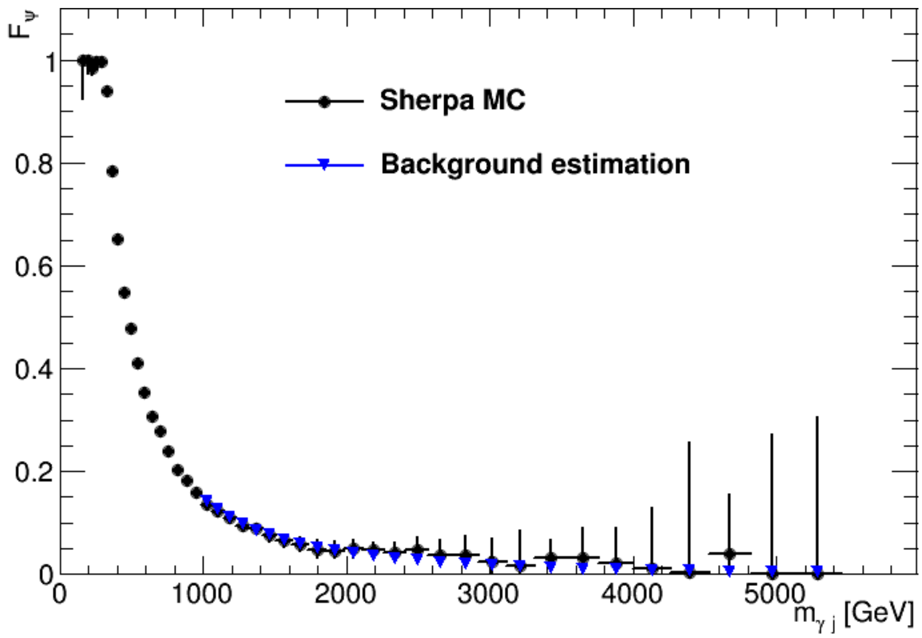


Figure 6.6: Fit on F_ψ using the N_ψ fit division method (simulated using Sherpa)

6.2.2 Bootstrap method error estimation

The statistical errors on F_ψ resulting from the fits on N_ψ were estimated using a bootstrap method. A bootstrap method is a procedure that uses pseudo-random computer generated distributions to simulate statistical fluctuation for a given parent distribution [29]. The parent is used as an input probability density function (PDF) to generate new daughter distributions of equal size. The need to do such an estimation stems from the fact that the bin contents of the N_ψ distributions are highly correlated with each other. Any statistical fluctuation in either N_{ψ_d} or N_{ψ_n} will directly affect the estimation of the final F_ψ distribution.

The daughter distributions were created using the two dimensional distribution N_{ψ_d} and N_{ψ_n} were generated from. This way, when new N_ψ distributions are generated, they will adhere to the fact that N_{ψ_n} is always a subset of N_{ψ_d}. While, by definition, the contents of each of the bins are not expected to match the parent, the new histograms will properly estimate the statistical Poisson fluctuations of the bin contents, while holding the total size of the daughter distributions equal. Then, each daughter distribution is fitted using the fit function from Equation 6.3. From each new N_{ψ_n} and N_{ψ_d} pair a new F_ψ distribution is calculated. The important thing to note is that these new fits are not performed on the original MC, but on the statistically fluctuated daughter histograms.

In order to extract useful uncertainties from this method, a pull distribution was generated for each mass bin $m_{\gamma j}$. Ideally, a pull distribution is a Gaussian distribution centered at 0, defined such that the standard deviation of the distribution is the statistical error being estimated. Every point in the pull distribution, given a mass bin $m_{\gamma j}$, is calculated as:

$$p_i = \bar{x}_{m_{\gamma j}} - x_i. \tag{6.4}$$

Here, $\bar{x}_{m_{\gamma j}}$ is the value of the F_ψ background estimate for the mass bin $m_{\gamma j}$. The x_i is the fit value of the ith bootstrap fit on N_ψ. While we expect the shape of a pull distribution to be Gaussian, the fact that no point in F_ψ can be negative induces a skewness in the pull distributions at higher masses. Thus, instead of the standard deviation of the pull distribution, the root-mean-square (RMS) value of each pull distribution was used to estimate the error for each mass bin. In the case of a Gaussian pull distribution, the RMS would be equal to the standard deviation [30].

It is important to note that the calculated uncertainties are statistical uncertainties to the background estimation only. The usefulness of this method is that the resulting uncertainties correctly treat the correlation effects of this method. The result of the fitting and the bootstrap error estimation can be found in Figure 6.7 for both Pythia and Sherpa.

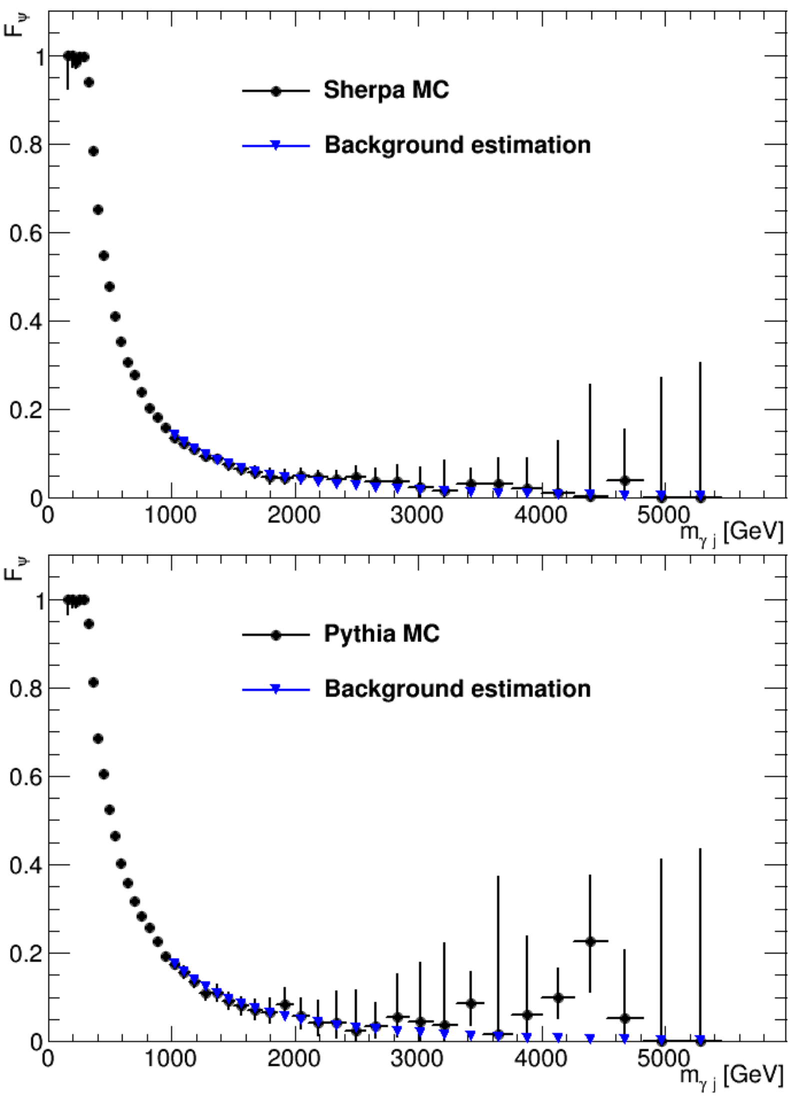


Figure 6.7: F_ψ as a function of $m_{\gamma j}$ in Sherpa (top) and Pythia (bottom) with background estimate from N_ψ fits, using bootstrapping uncertainty calculation

6.2.3 Fit comparisons

The methodology described in Sections 6.2.1-6.2.2 was applied to estimate the Sherpa and Pythia F_ψ background distributions in Figure 6.7. As was shown in Chapter 5, what is expected from the cross section calculation as a function of ψ is a distribution that has a smooth shape [9]. Since the fluctuations in F_ψ can be traced to the similar fluctuations when comparing MC to data in Figure 4.4, as well as the normal statistical fluctuations, we conclude that the resulting background estimation from the bootstrap method is appropriate for this analysis.

7

Signal sensitivity in F_ψ

7.1 Formulation

The ultimate goal of this analysis is to determine whether the angular analysis is sensitive to q*-like signal. To do this, we extract sensitivities from signal injection in SM samples. Specifically, we use the resonance-like signal of the excited quark model. We have demonstrated that a data driven background estimation was necessary, and developed a methodology to do so. Using excited quark signal templates, we can now provide mathematical definitions for signal, significance, and signal sensitivity. In this context, the desired sensitivity is the signal sample cross section that would yield a 5σ significance when injected in SM samples. In this thesis, we assume a known m_{q*}, and calculate local significances only [1].

To properly define the presence of signal in F_ψ, we evaluate the significance of the resonance above the nominal background estimation. We define signal in a single mass bin $m_{\gamma j}$ to be:

$$signal(m_{\gamma j}) = F_\psi(m_{\gamma j}) - f(m_{\gamma j}), \tag{7.1}$$

1. In a complete analysis m_{q*} would be treated as an unknown, reducing the local significance of signal. This is known as the ”look elsewhere” effect [30].

where $f(m_{\gamma j})$ is the background estimation of F_ψ for the mass bin $m_{\gamma j}$. The sign in the signal calculation matters as this allows us to search for excesses as well as deficits in F_ψ. Using basic error propagation, we calculate the error on the signal yield per mass bin $m_{\gamma j}$ to be:

$$\sigma^2_{signal}(m_{\gamma j}) = \sigma^2_{F_\psi(m_{\gamma j})} + \sigma^2_{f(m_{\gamma j})}. \tag{7.2}$$

We then sum over the bins of interest in order to compute the total amount of signal in a continuous range of $m_{\gamma j}$ bins:

$$Tot = \sum_{m_{\gamma j}=m_{q*}-2}^{m_{q*}+2} F_\psi(m_{\gamma j}) - f(m_{\gamma j}), \tag{7.3}$$

where m_{q*} is the resonant mass of the chosen signal sample. Since we are currently only interested in computing local significances, we assume a known q* mass, and fix the number of bins in our sum at 5 to cover the excess. Now extending the error propagation from the single bin case, we find that the total error on the signal bump is:

$$\sigma^2_{tot} = \sum_{m_{\gamma j}=m_{q*}-2}^{m_{q*}+2} \sigma^2_{F_\psi(m_{\gamma j})} + \sigma^2_{f(m_{\gamma j})}. \tag{7.4}$$

We can use a straightforward definition to calculate the significance of the observed excess (or deficit):

$$significance = Tot/\sigma_{tot}. \tag{7.5}$$

7.2 Extracting signal sensitivities

In order to compute our cross section sensitivity, we iteratively scale the signal samples and re-evaluate the significance until we observe a 5σ significance. We then compute the cross section that would be associated with this signal yield, and define this as our 5σ cross section sensitivity. As we know both the luminosity of the sample and the number of events in each signal sample, calculating cross sections is straightforward, using:

$$\sigma_{m_{q*}} = N_{events} \;/\; 20.3\; fb^{-1} \tag{7.6}$$

where N_{events} is the number of events in the signal sample. We say that $\sigma_{m_{q*}}$ is the cross section our method is sensitive to at a 5σ level for an excited quark mass m_{q*}. Although a q* sample is used to produce signal templates, this calculation is valid for any model of resonances with a shape similar to that of the q* model. We can also compute a 1σ error band on this cross section, by modifying our definition of signal. For one mass bin:

$$signal \pm \sigma_{f(m_{\gamma j})} = F_\psi(m_{\gamma j}) - \left(f(m_{\gamma j}) \pm \sigma_{f(m_{\gamma j})}\right). \tag{7.7}$$

We then calculate the new 5σ significant cross section for each m_{q*}. This correctly propagates $\pm 1\sigma_{signal}$ into $\pm 1\sigma$ of cross section sensitivity. Based on the Poisson fluctuations in the signal samples, we calculate a one-sided 90% credibility limit for each $\sigma_{m_{q*}}$. We estimate the downward fluctuation in N_{events} by one standard deviation by calculating:

$$N_{events\ 90\%} = N_{events} - Z^*_{90\%} \times \sqrt{N_{events}}, \tag{7.8}$$

where $Z^*_{90\%}$ is the Z-statistic for 90% of the Poisson integral. We utilize the fact that for a high Poisson mean (in our case high N_{events}), the distribution is well estimated by a Gaussian distribution [30]. The number of events in each sample (after scaling to the 5σ level) can be seen in Tables 7.1 and 7.2. This calculation is done for every q* sample in the F_ψ search region for both Pythia and Sherpa. From $N_{events\ 90\%}$ we calculate a new cross section sensitivity, which we interpret as the one-sided 90% frequentist credibility limit on the cross section of each q* sample, quoted in Tables 7.1 and 7.2.

To estimate the expected systematic effect on this sensitivity study of using SM MC samples as the background instead of actual ATLAS data, we re-generate the F_ψ distribution after re-weighing MC events to more closely resemble data. We use the photon η distribution to drive the re-weighting procedure as they displayed the most significant and structured set of discrepancies between MC and data in the comparisons performed in Chapter 4. Since our analysis method relies on the angular distribution between leading photons and jets, η modeling in the MC is expected to be the dominant contribution to systematics. The additional weights applied to MC events is the ratio of the MC and data distribution in the η bin of this event, as shown in Figure 4.4. Computing the RMS value of the differences between the original and re-weighted F_ψ over all bins in $m_{\gamma j}$, there was an 8.6% difference in Sherpa and an 8.8% difference in Pythia. However, there is no overall change to the shape of F_ψ so it seems likely that the analysis methodology will be robust against relatively large variations in calorimeter performance. This robustness is an expected feature of the ratio method, as a result of partial cancellation of these effects in the ratio.

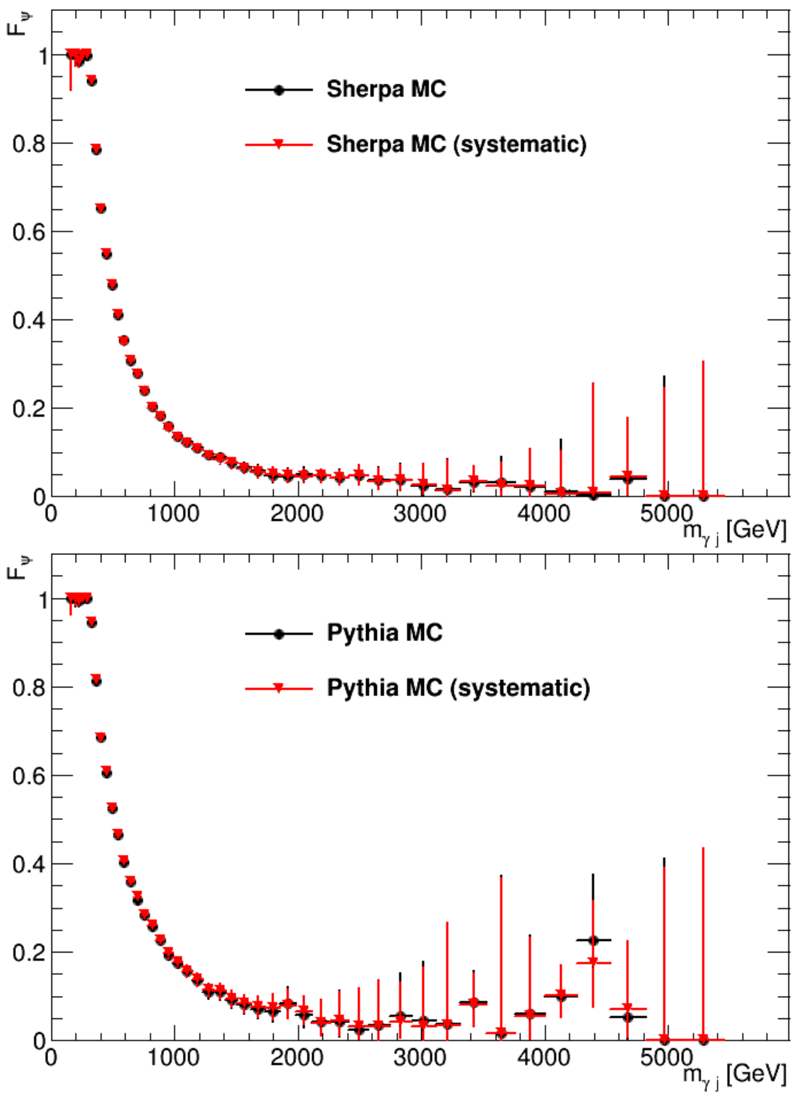


Figure 7.1: Re-weighted F_{ψ} (red) superimposed on original F_{ψ} (black) for both Sherpa (top) and Pythia (bottom), plotted as a function of $m_{\gamma j}$

7.3 Signal sensitivities

Using the developed methodology, we present the expected sensitivity limits of excited quark production with a data sample corresponding to an integrated luminosity of 20.3 fb^{-1} for both Pythia and Sherpa in Tables 7.1 and 7.2. An example showing how such 5σ significances would look in data is shown in Figure 7.2 using Sherpa and the $m_{q*} = 2.5$, 3.5, and 4.5 TeV samples.

m_{q*} [GeV]	N_{events}	$\sigma_{m_{q*}}$ [nb]	90% Limit [nb]
2500	78.06	$3.85^{+0.03}_{-0.07}$	4.50
3000	54.91	$2.70^{+0.06}_{-0.06}$	3.27
3500	34.82	$1.72^{+0.34}_{-0.06}$	2.19
4000	23.60	$1.16^{+0.10}_{-0.01}$	1.58
4500	22.02	$1.08^{+0.21}_{-0.13}$	1.49

Table 7.1: Tabulated results showing the number of signal events and cross section sensitivities for Sherpa per m_{q*} sample

m_{q*} [GeV]	N_{events}	$\sigma_{m_{q*}}$ [nb]	90% Limit [nb]
3000	59.18	$2.92^{+0.12}_{-0.09}$	3.50
3500	39.47	$1.94^{+0.15}_{-0.11}$	2.44
4000	28.32	$1.40^{+0.40}_{-0.02}$	1.84
4500	27.10	$1.33^{+0.23}_{-0.04}$	1.77

Table 7.2: Tabulated results showing the number of signal events and cross section sensitivities for Pythia per m_{q*} sample

Sherpa is found to predict sensitivity to lower cross sections when compared to Pythia per m_{q*} sample. This is due to angular modeling of the final state particles by the detectors, which resulted in a smoother F_ψ distribution in Sherpa.

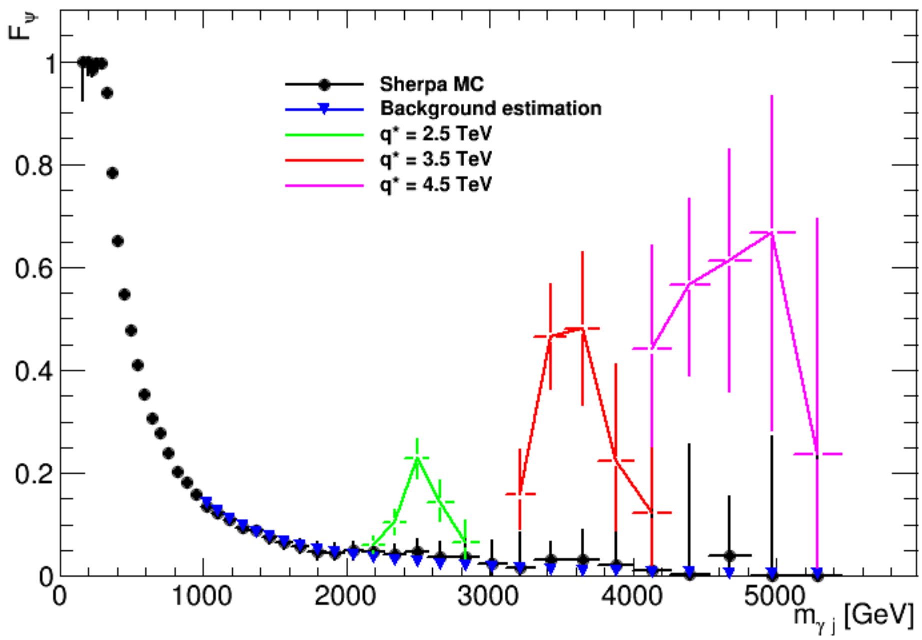


Figure 7.2: F_ψ and background estimate plotted as a function of $m_{\gamma j}$ with 5σ significant signals injected at m_q* = 2.5, 3.5, and 4.5 TeV

8

Conclusions and future work

The analysis method described in this thesis provides a methodology to search for new physics in final state photon+jet events in ATLAS data. Both the Pythia and Sherpa Monte-Carlo samples used were demonstrated to adequately model the shape of the basic photon and jet kinematic variables, despite the structural differences that were discussed in previous chapters. In the observable distribution, F_ψ, some discrepancies were observed between the MC generators, such as the bin-to-bin fluctuations in F_ψ, and in the η re-weighting to match ATLAS data. However, these differences do not hinder the sensitivity of the analysis, which was shown specifically for resonance-like new physics using Pythia excited quark samples. The derived background estimation method using the N_ψ distributions proved to describe the F_ψ distribution well. This method allows for estimation of background even with the presence of signal and for signal significance testing.

The analysis technique demonstrated here provides a solid foundation for background estimation and significance testing, and is a viable method to test for new physics in photon+jet final state events. In a future analysis, systematic uncertainties such as photon ID systematics and mass resolution would need to be calculated, as they would be vital in evaluating the final sensitivity of this analysis to signal in ATLAS data. After including systematic uncertainties, we could then unblind the analysis and use ATLAS data to search for new physics using this method.

This analysis has sensitivity to search for new physics in the current ATLAS dataset. The developed method is a good candidate to be the first angular analysis in the photon+jet final state at ATLAS, an exciting prospect since no angular measurements have been published in this final state since CDF [10].

Bibliography

[1] D. Griffiths, *Introduction to Elementary Particles*. John Wiley & Sons, Sep 26, 2008.

[2] C. Burgess and G. Moore, *The Standard Model: A Primer*. Cambridge University Press, 2007.

[3] B.R. Martin and G. Shaw. *Particle Physics*. Wiley, Chichester, West Sussex,2008.

[4] R. Mann, *An Introduction to Particle Physics and the Standard Model*. Taylor & Francis Group, July 13, 2011.

[5] O. Cakir and R. Mehdiyev, *A search for excited quarks at the CERN LHC*, ATL-COM-PHYS-99-003. `http://cds.cern.ch/record/683799`.

[6] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv:1207.7235 [hep-ex].

[7] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B, **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020.

[8] ATLAS Collaboration, *Search for new phenomena in photon+jet events collected in proton–proton collisions at sqrt(s) = 8 TeV with the ATLAS detector*, Phys. Lett. B **728**(2013) 562, arXiv:1309.3230 [hep-ex].

[9] R. Keyes, A. Warburton, *New angular search for exotic physics in the photon+jet final state using the ATLAS detector on the LHC*, CERN-THESIS-2012-367, `https://cds.cern.ch/record/1648276`.

[10] F. Abe et al., CDF Collaboration, *Center-of-mass angular distribution of prompt photons produced in ppbar collisions*, Phys. Rev. Lett. **71** (1993) 679, `http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.71.679`.

[11] ATLAS Collaboration, *Search for New Phenomena in Dijet Mass and Angular Distributions using 4.8 fb^{-1} of pp Collisions at $\sqrt{s}$ = 7 TeV collected by the ATLAS Detector*, ATLAS-CONF-2012-038, `https://cds.cern.ch/record/1432206`.

[12] ATLAS Collaboration, *ATLAS search for new phenomena in dijet mass and angular distributions using pp collisions at $\sqrt{s}$=7 TeV*, J. High Energy Phys. 01 (2013) 029, arXiv:1210.1718 [hep-ex].

[13] SNO Collaboration, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 1, `http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.89.011301`.

[14] ATLAS Collaboration, *Measurements of the photon identification efficiency with the ATLAS detector using 4.9 fb^{-1} of pp collision data collected in 2011*, ATLAS-CONF-2012-123. `http://cds.cern.ch/record/1473426`.

[15] T. Sjöstrand, S. Mrenna, P. Skands, *A Brief Introduction to Pythia 8.1*, FERMILAB-PUB-07-512-CD-T, arXiv:0710.3820 [hep-ph].

[16] Sherpa 2.0.0 Generator Manual, `https://sherpa.hepforge.org/doc/SHERPA-MC-2.0.0.html`.

[17] ATLAS Collaboration, *ATLAS detector and physics performance : Technical Design Report, 1*, ATLAS-TDR-14/CERN-LHCC-99-014, `https://cds.cern.ch/record/391176`.

[18] ATLAS Collaboration, *Performance of the ATLAS Detector using First Collision Data*, J. High Energy Phys. 1009:056, 2010, arXiv:1005.5254 [hep-ex].

[19] ATLAS Collaboration, *Operational Experience of the ATLAS High Level Trigger with Single-Beam and Cosmic Rays*, ATL-DAQ-PROC-2009-044, `https://cds.cern.ch/record/1228285`.

[20] `http://lhc-commissioning.web.cern.ch/lhc-commissioning/`, Retrieved 14 Feb 2014.

[21] `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults#2012_data`, Retrieved 14 Feb 2014.

[22] ATLAS Collaboration, *Performance of the ATLAS Electron and Photon Trigger in p-p Collisions at* $\sqrt{s} = 7$ *TeV in 2011*, ATLAS-CONF-2012-048, `https://cds.cern.ch/record/1450089`.

[23] ATLAS Collaboration, *ATLAS Detector Simulation with Geant4*, `http://atlas-computing.web.cern.ch/atlas-computing/packages/simulation/geant4/geant4.html`.

[24] J. Allison et al., *Geant4 Developments and Applications*, IEEE Transactions on Nuclear Science, 53 No. 1 (2006) 270-278.

[25] S. Agostinelli et al., *Geant4 - A Simulation Toolkit*, Nuclear Instruments and Methods **A** 506 (2003) 250-303.

[26] M. Cacciari, G. P. Salam, G. Soyez, *The anti-k_t jet clustering algorithm*, J. High Energy Phys. 0804:063, 2008, arXiv:0802.1189 [hep-ph].

[27] ATLAS Home Page, `http://www.atlas.ch/pdf/ATLAS$_$fact$_$sheets.pdf`.

[28] ATLAS Collaboration, *ATLAS inner detector : Technical Design Report, 1*, ATLAS-TDR-4/CERN-LHCC-97-016, `http://cds.cern.ch/record/331063`.

[29] G. Choudalakis, *On hypothesis testing, trials factor, hypertests and the BumpHunter*, arXiv:1101.0390v2 [physics.data-an].

[30] F. James, "Statistical Methods in Experimental Physics (2nd Ed.)", World Scientific Publishing, 2006.

[31] Wikipedia, `http://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`, Retrieved 14 Feb 2014.

[32] `http://www.lppp.lancs.ac.uk/higgs/images/atlas4.gif`

[33] `https://kjende.web.cern.ch/kjende/images/calorimeter_EN.png`, Retrieved 29 Sept 2013.

[34] `http://cms.web.cern.ch/sites/cms.web.cern.ch/files/styles/large/public/field/image/Sketch_PartonParticleCaloJet.png?itok=aWvJLScL`

[35] Vanity Fair, LHC map. `http://www.vanityfair.com/culture/features/2010/01/hadron-collider-201001`

[36] M. Stockton, ATLAS results talk. http://www.physics.mcgill.ca/WNPPC12/talks/WNPPC_2012_Stockton.pdf

[37] Illustration of EM calorimeter accordion geometry, http://ars.els-cdn.com/content/image/1-s2.0-S0010465511003092-gr002.jpg.