First observation of electroweak Z boson plus two jet production

Christian Gütschow
UCL

October 10, 2014

Declaration

I, Christian Gütschow, confirm that the work presented in this thesis is my own.
Where information has been derived from other sources,
I confirm that this has been indicated in the thesis.

Christian Gütschow
October 2014

Abstract

Events with jets produced in association with a Z boson in proton-proton collisions can be used to study distributions sensitive to the vector boson fusion (VBF) process at CERN's Large Hadron Collider. This process is interesting to study because of its similarity to the VBF production of a Higgs boson as well as its sensitivity to new physics via the WWZ triple gauge coupling. Evidence for electroweak Zjj production beyond the 5σ level is presented using data collected at a centre-of-mass energy of $\sqrt{s} = 8$ TeV by the ATLAS experiment in 2012. This constitutes the first observation of a VBF-like process. The detector-corrected cross sections measured in two fiducial regions are in excellent agreement with the Standard Model expectations and have also been used to constrain anomalous triple gauge couplings. Furthermore, detector-corrected cross sections and differential distributions for inclusive Zjj production are presented in different regions of phase space with varying sensitivity to the electroweak Zjj component. In addition, a performance study of the tracking algorithms used in the ATLAS high-level trigger system is presented.

We have a habit in writing articles
published in scientific journals
to make the work as finished as possible,
to cover up all the tracks,
to not worry about the blind alleys
or describe how you had the wrong idea at first,
and so on. So there isn't any place to publish,
in a dignified manner,
what you actually did
in order to get to do the work.

— Richard Phillips Feynman

Preface

Modern particle physics experiments have become incredibly complex and continue to inspire decades of collaborative effort. The ATLAS experiment alone has over $3\,000$ members from all over the world who seemingly defy all cultural and geographical boundaries in an attempt to gain a better fundamental understanding of how the universe works. None of the measurements presented in this thesis would have been possible without everyone's commitment to the ATLAS detector, its maintenance, performance monitoring, the calibration and so on. My analysis work relies heavily on software tools that have been developed by other members of the collaboration and so I should like to thank my ATLAS colleagues for their dedicated engagement and hard work. The ATLAS collaboration as a whole is indebted to CERN and its supporting staff for their successful operation of the Large Hadron Collider without which there would be no collision data to analyse, no new physics to explore.

The analysis team responsible for the Zjj measurements consisted of Andy Pilkington, Emily Nurse, Graham Jones, Kiran Joshi and myself. Although I have contributed to every part of the analysis, my level of involvement varies for different analysis aspects and so I would like to give credit where credit is due. All `Powheg` predictions have been produced by Graham, whose adept `Powheg` knowledge has been invaluable. The ATLAS unfolding framework has been heavily extended by Kiran, who has also developed, lost, redeveloped and run the unfolding machinery for the differential distributions and their associated uncertainties. My plotting script has merely combined his unfolded data curves with the `Sherpa` and `Powheg` predictions provided by Andy and Graham, respectively. Kiran has also been responsible for running `Sherpa` on numerous occasions and provided me with various `Sherpa` curves (at 8 TeV, at 14 TeV, with and without interference – you name it) as well as scripts and a Rivet routine to go with it. Emily has provided me with the original fitting framework that I have built upon and extended in order to extract the electroweak Zjj component and its associated uncertainties. However, the experimental uncertainties associated with the signal extraction are based on systematic variations that Andy has produced. In return, I have provided him with all sorts of numerical input, so that he could run the limit-setting machinery for anomalous triple gauge couplings. I have also been responsible for the inclusive cross-section measurements, for which Andy provided an estimate of the multijet background. The constructive critique of our editorial board, led by Monica Dunford, has helped improve the analysis significantly and their support and patience have been much appreciated. It has been a great pleasure working with everyone on this analysis.

Moreover, I am particularly grateful to Jiri Masik and Mark Sutton for all their patience and insight regarding the inner detector trigger software. I could not have produced the here presented efficiency measurements without their help. The original

Contents

List of figures

List of tables

1

A theory of almost everything

The Standard Model of particle physics is a theory of all the fundamental interactions occurring in nature, apart from gravity. The Standard Model explains a wide range of experimental results and many predictions by the Standard Model have been subsequently verified experimentally with very high precision, too, making it the most successful quantum field theory to date. Formally, the Standard Model is based on the $SU(3) \times SU(2) \times U(1)$ gauge symmetry group, giving rise to three distinct fundamental types of interactions:

- The electromagnetic and weak interactions observed in nature are remnants of an underlying electroweak symmetry, as shown by Sheldon Glashow in 1961 [1]. The electroweak symmetry is spontaneously broken through the Brout-Englert-Higgs mechanism [2–7], giving mass to some of the electroweak gauge bosons in the process. The discovery of these heavy gauge bosons in 1983 provided direct experimental evidence in support of the theory.

- The electroweak gauge bosons acquire mass through interactions with the so-called Higgs field, a scalar quantum field which manifests itself in a scalar boson that couples to all massive particles including itself. The recent discovery of a scalar boson consistent with the Standard Model expectations [8, 9] is generally considered to be another major success of the Standard Model of particle physics. The Brout-Englert-Higgs mechanism was incorporated into the theoretical framework in 1967 by both Abdus Salam [10] and Steven Weinberg [11].

- Strong interactions are described by Quantum Chromodynamics (QCD) which postulates the existence of colour charge in loose analogy to the electric charge of quantum electrodynamics (which describes interactions between light and matter). Unlike its electric counterpart, however, colour charge comes in three types that mix in a conceptually similar fashion as the primary colours of the RGB colour model, such that a combination of all three types of colour charge renders a colour-neutral ('colourless') state. The transition between different colour

states is administered by the force carriers of the strong interaction, the so-called gluons, which are themselves colour-charged particles and can therefore participate in strong interactions. This has some important consequences concerning the coupling strength of the strong interaction. Consider the vacuum surrounding an electrically charged particle. By virtue of Heisenberg's uncertainty principle, it is possible to create an electrically charged particle-antiparticle pair (a virtual quantum loop) for a very short amount of time. Such quantum fluctuations become polarised in the vicinity of a (real) electrically charged particle and, as a result, the electrical field of the real charge is partially cancelled out. The observed electrical charge therefore appears to be screened and it follows that the effective strength of the electromagnetic force decreases with distance. In the vicinity of colour charge, however, there are additional quantum fluctuations involving gluon quantum loops. The effect of polarised virtual gluons is to enhance the colour charge, such that the strong force between colour-charged particles does not decrease with distance and, as a result, colour-charged particles are presumed to be confined to colour-neutral systems. Albeit a generally accepted concept, the confinement has never formally been proven and still remains one of the unresolved Millenium Prize Problems [12]. At length scales of just a few fermi and larger, the strong coupling is about a factor of 10^2 stronger than the electromagnetic coupling, a factor of 10^6 stronger than the weak coupling and a factor of 10^{39} stronger than the gravitational coupling. Consequently, the strong force is responsible for binding the nucleons of an atom together, making it the strongest force observed in nature. At the other end of the distance spectrum, the strong coupling becomes smaller as the colour-enhancing effect of virtual gluons diminishes. For high enough energies (or, conversely, at small enough distances), the strong coupling is sufficiently small, such that the mathematical expression for the scattering interaction between particles can be expanded in powers of the strong coupling constant, α_S, such that approximate calculations up to a certain order in α_S become feasible in QCD. The leading order (zeroth power in α_S) terms in the series represent the simplest possible scattering interaction, whilst the higher-order terms correspond to more complicated processes. For example, the additional radiation of a gluon is represented by a small correction (a perturbation) to the leading order calculation requiring next-to-leading order terms (one power in α_S). Many predictions of perturbative calculations in QCD have been verified experimentally in precision measurements using data collected at the Large Electron-Positron Collider between 1989 and 2000.

The elementary particles postulated by the Standard Model are illustrated in figure 1.1. The outer (black) ring shows the twelve spin-$\frac{1}{2}$ particles predicted by the Standard Model. These obey Fermi-Dirac statistics and are thus subject to the Pauli exclusion principle. The corresponding antiparticles are omitted for clarity. The fermions in the top half are called quarks, whilst the fermions in the bottom half are called leptons. The electron (e), muon (μ) and tau (τ) lepton shown in the lower left quarter carry each one negative unit of electric charge $|e|$. Each of these comes with an associated
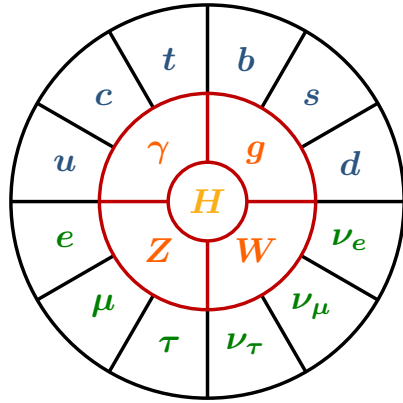
Figure 1.1
Elementary particles predicted by the Standard Model. The outer (black) circle shows the spin-$\frac{1}{2}$ fermions, where quarks (in blue) are distinguished from leptons (in green). The remaining particles in the (red) centre of the diagram are the bosons predicted by the Standard Model, where the spin-1 force carriers (in orange) are distinguished from the spin-0 boson (in yellow) at the very centre of the diagram. Refer to main text for details.

neutrino (ν) – shown in the lower right quarter – that does not carry any electric charge, which makes the neutrino particularly hard to detect experimentally as it only interacts weakly. The up (u), charm (c) and top (t) quark shown in the top left quarter carry $+\frac{2}{3}$ units of electric charge, while the bottom (b), strange (s) and down (d) quark in the top right quarter carry $-\frac{1}{3}$ units of electric charge. In addition, the quarks are also colour-charged which means that they are both strongly and electroweakly interacting particles. By contrast, the leptons are colour-neutral particles and therefore do not interact strongly. The colour-charged quarks, however, cannot be observed in isolation due to colour confinement: As two quarks become separated from one another, it is energetically more favourable for a new quark-antiquark pair to appear between them, such that they can recombine to form colour-neutral bound states, also known as hadrons. It is the strong force that keeps quarks to the confines of the hadronic system, such as a proton which is the colour-neutral state of two up quarks and a down quark for instance[1]. The consistent failure to detect free quarks experimentally is considered further evidence in favour of QCD. The existence of individual quarks can instead be inferred in a detector from the collimated stream of hadrons they turn into as a result of confinement. This QCD decay cascade is commonly referred to as a 'jet'.

Opposite pairs of fermions in the left and right halves of the diagram in figure 1.1 share similar properties based on the various flavour quantum numbers of the theory, giving rise to three generations of quarks and leptons. For instance, charged leptons share the same lepton flavour with their associated neutrinos. Another flavour quantum number is the weak isospin which is of opposite sign for the charged leptons and the neutrinos and similarly of opposite sign for the up and down quark. The complexity of the quark-flavour structure is increased by additional quantum numbers related to the strange-, charm-, bottom- and top-quark flavours. Moreover, it is worth noting that the mass distribution among the quarks is blatantly skewed over several orders of magnitude, ranging from a mere few MeV for the up quark to a few GeV for the

---

[1]Note that the colour-charged quarks within the hadronic system constantly undergo colour-state transitions through gluon exchange, which may proceed either directly or involve additional gluon radiation as well as gluon splitting into a virtual quark-antiquark pair. The hadronic system is therefore subject to a constant flux of virtual quarks and gluons that may participate in scattering interactions at hadron colliders just like the primary hadron constituents.

bottom quark and shooting all the way up to about 173 GeV for the top quark. This disproportionality is similar for the leptons and remains an open question, coined the 'mass hierarchy problem', as it is not explained by the Standard Model.

The remaining particles predicted by the Standard Model are shown in the inner (red) circle of the diagram, all of which obey Bose-Einstein statistics and thus are not subject to the Pauli exclusion principle. The outer ring hosts the spin-1 gauge bosons which act as the force carriers. The massless and electrically neutral photon (γ) mediates the electromagnetic force between electrically charged particles. The heavy Z and W bosons are the intermediaries of the weak interaction. While the Z boson carries no electric charge, there are two W bosons carrying one positive and one negative unit of electric charge, respectively. The strong force is carried by the gluon (g) which can be in one of eight colour states. The gluon couples to all colour-charged particles including itself, since it carries an effective colour charge. The only spin-0 boson predicted by the Standard Model is the Higgs boson (H) at the centre of the diagram. The Higgs boson couples to all massive particles with the coupling strength being roughly proportional to the mass of the particle to which it couples.

It is worth noting that about a third of the particle content had been predicted by the Standard Model before the particles were discovered experimentally. However, despite its enormous success, the Standard Model only describes the ordinary matter content in the universe. That is to say, it fails to account for both dark energy and dark matter, which make up a non-negligible 95% of the physics in the known universe. As already mentioned, gravity has yet to be incorporated into the Standard Model and the theory also fails to explain the mass hierarchy observed across the particle content of the Standard Model. Furthermore, in the original formulation of the Standard Model, neutrinos are predicted to be massless. However, it has been experimentally established that neutrinos can oscillate between the lepton flavours, implying that they must have a non-vanishing mass, which could still be accommodated for in the Standard Model though. Moreover, some of the rare Standard Model processes remain unobserved due to lack of experimental sensitivity. The production of a Z boson through a WWZ vertex is one such rare interaction. For the first time this channel has now become experimentally accessible in a measurement [13] of two jets produced in association with a Z boson (Zjj) using proton-proton collision data. This measurement shall be discussed in some detail in the remainder of this document, which is outlined as follows.

A description of the experimental setup can be found in chapter 2. Chapter 3 introduces the process of interest and its various production mechanisms. It will be shown that the two jets can be produced either strongly or electroweakly in association with the Z boson and details of how the different production mechanisms are simulated in order to obtain theoretical predictions are provided. Kinematic differences between strong and electroweak Zjj production are then discussed and regions of phase space are defined in which Zjj production is to be studied. A measurement of inclusive Zjj production in these regions is then presented in chapter 4, while a measurement of differential distributions is discussed in chapter 5. The latter is used to test the theoretical modelling of strong Zjj production in extreme regions of phase space, whilst

probing the impact of the electroweak component. The electroweak component of Zjj production is then extracted and the first significant measurement of electroweak Zjj production is presented in chapter 6. Similarities between the production of a Z boson and a Higgs boson in association with two jets are exploited in chapter 7, in an attempt to probe the theoretical modelling even further. An outlook regarding potential future Zjj measurements is presented in chapter 8 and, finally, a summary is given in chapter 9.

2

Design and performance of the ATLAS detector

The Large Hadron Collider (LHC) is the name given to the largest of a number of particle accelerators currently operated by the *European Organization for Nuclear Research* (CERN) near Geneva, Switzerland. The LHC is placed in a 27 km underground tunnel below the Swiss Canton of Geneva as well as the French Pays de Gex. It is designed to accelerate and collide protons at an energy of 14 TeV, which makes the LHC to date the world's largest and most powerful machine ever to be constructed. Protons are injected into the main accelerator ring after passing through a chain of pre-accelerators that successively increase the proton energy to 450 GeV. In the 2012 run, the LHC has been used to accelerate the injected protons within about 20 minutes to a peak energy of 4 TeV and to subsequently collide them in four interaction points, each hosting a particle detector. The ATLAS (**A T**oroidal **L**HC **A**pparatu**S**) experiment is located close to the main CERN site and is one of the two general purpose detectors, designed to learn as much as possible about rare or new interactions that may occur at the unprecedented collision energies provided by the LHC.

2.1 The ATLAS coordinate system

The coordinate system adopted by ATLAS is illustrated in figure 2.1 in relation to the LHC ring. ATLAS uses a right-handed coordinate system, in which the origin coincides with the nominal interaction point. The x-axis points from the nominal interaction point towards the centre of the LHC ring, while the y-axis points upwards, towards the surface of the earth. The z-axis runs along the centre of the beam pipe, pointing counterclockwise, with the azimuthal angle ϕ being measured right-handedly around it, starting from the x-axis. A Cartesian momentum vector $\vec{p} = (p_x, p_y, p_z)$ can then be expressed in terms of its longitudinal component $p_z = |\vec{p}| \cos\theta$ (projected onto the beam pipe) and its transverse component $p_{\mathrm{T}} = |\vec{p}| \sin\theta$ (projected onto the x-y plane)
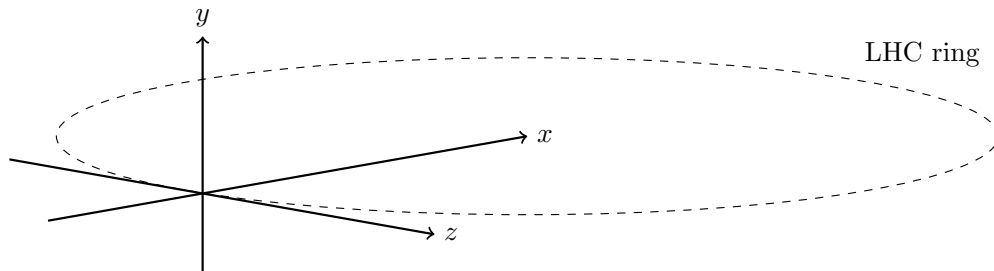
Figure 2.1
The ATLAS coordinate system in relation to the LHC ring. ATLAS uses a right-handed coordinate system with the nominal interaction point defined as the origin. The x-axis points from the nominal interaction point towards the centre of the LHC ring, while the y-axis points upwards, towards the surface of the earth; the z-axis runs along the centre of the beam pipe, pointing counterclockwise.

as

$$\vec{p} = (p_\mathrm{T}\cos\phi, p_\mathrm{T}\sin\phi, p_z) \tag{2.1}$$

where the polar angle θ is measured with respect to the z-axis. The Lorentz transformation between the lab frame and the particle's rest frame can be specified in terms of rapidity

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) = \frac{1}{2}\ln\left(\frac{E+|\vec{p}|\cos\theta}{E-|\vec{p}|\cos\theta}\right) \tag{2.2}$$

where E is the energy of the particle. Rapidity is a convenient choice of variable since differences in rapidity are invariant under Lorentz boosts along the beam axis. Moreover, in the massless limit

$$\lim_{E\to|\vec{p}|} y = \frac{1}{2}\ln\left(\frac{1+\cos\theta}{1-\cos\theta}\right) = \frac{1}{2}\ln\left(\frac{2\cos^2(\theta/2)}{2\sin^2(\theta/2)}\right) = -\ln\left(\tan\frac{\theta}{2}\right) = \eta \tag{2.3}$$

which is known as the pseudorapidity and only depends on the trajectory of the particle. Because it is independent of the particle energy, the pseudorapidity is the preferred choice of variable to parameterise the position of a particle in the detector. Figure 2.2 shows a selection of polar angles expressed as pseudorapidities for easier visualisation.
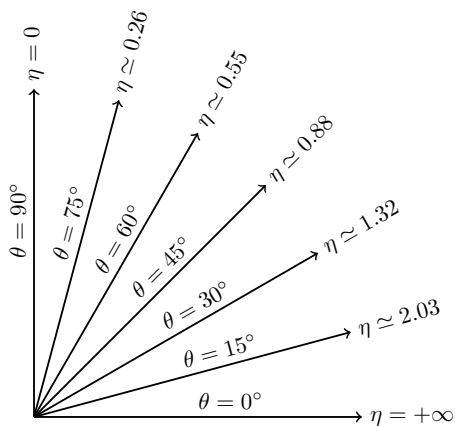
Figure 2.2
A selection of polar angles expressed as pseudorapidities. The peusorapidity η is related to the polar angle θ via $\eta = -\ln(\tan(\theta/2))$.

2.2 The ATLAS detector

The ATLAS detector is described extensively in refs. [14, 15] and so at this point only a comparatively short summary is given. The detector comprises a number of subdetectors assembled in concentric cylindrical layers as illustrated in figures 2.3 and 2.4.

Particle trajectories and interaction vertices are reconstructed with the inner detector (ID) tracking system, which consists of a silicon pixel detector, a semiconductor tracker (SCT) and a transition radiation tracker (TRT). The pixel detector is the innermost component, surrounding the beam pipe just a few centimetres away from the interaction point. It provides the highest precision measurements due to the high granularity of its three pixel layers, with an intrinsic resolution of about 10 μm in the transverse direction and about 115 μm in the longitudinal direction. The main purpose of the pixel detector is to detect short-lived particles that already decay close to the interaction point. The SCT is a silicon microstrip detector which surrounds the pixel detector and typically provides additional eight precision measurements per reconstructed track, thus contributing to both momentum measurement and vertex positioning. The intrinsic resolution of the SCT is about 17 μm in the transverse direction and about 580 μm in the longitudinal direction. The outermost component of the ID tracking system is a straw detector filled with a xenon gas mixture that ionises as charged particles pass through it, allowing for a drift-time measurement with a spatial resolution of about 130 μm per straw. Furthermore, the space between the straws is filled with materials of very different refractive indices, causing transversing particles to radiate photons. The TRT owes its name due to the fact that heavier particles are less likely to produce this transition radiation than the lighter ones, such that the tracking of transition radiation can be used to distinguish charged pions from electrons for instance. The ID tracking system is immersed in a 2 tesla axial magnetic field, produced by a solenoid magnet surrounding the TRT, which allows for charge and momentum identification of charged particles. The ID tracking system provides full hermetic coverage in ϕ and within $|\eta| < 2.5$. The transverse momentum resolution achieved by the combined ID tracking system is about 75 MeV for a track of around 5 GeV and between 3.5 GeV and 11 GeV for a track of around 100 GeV [16, 17], depending on the pseudorapidity of the track. Furthermore, for a track of around 5 GeV, the ID tracking system achieves an angular resolution of about 11 mrad in the polar direction and about 7 mrad in the azimuthal direction. The corresponding angular resolutions for a track of around 100 GeV are about 9 mrad and 2 mrad in the polar direction and the azimuthal direction, respectively.

The ATLAS calorimeter system surrounds the solenoid magnet that encloses the ID tracking system, providing fine-grained measurements of particle energy depositions over a wide range of pseudorapidity. The calorimeter system consists of an inner electromagnetic and an outer hadronic sampling calorimeter; their overall arrangement with respect to the ID tracking system is illustrated in figure 2.3. The electromagnetic calorimeter (ECAL) uses lead and steel as the absorbing material that interacts electromagnetically with incident electrons and photons for instance, resulting in an electro-
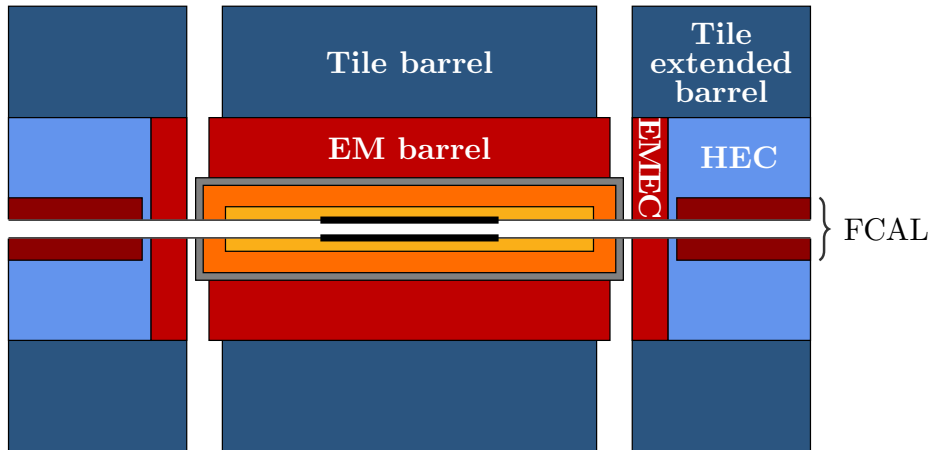
Figure 2.3
Schematic view of the ATLAS calorimeter and ID tracking system (not to scale) along the beam pipe. The muon spectrometer has been omitted for clarity. The subdetector labels equivalently apply to detector parts reflected across both the horizontal and the vertical axis. Refer to main text for details.

magnetic decay cascade in the calorimeter. The cascade particles ionise the liquid argon that is used as the active sampling material in between alternating layers of absorbing material and electrodes. The energy deposition in the calorimeter can then be inferred from a measurement of the voltage due to the ionised active material, with an energy resolution to within 2 % for electrons of around 50 GeV depending on their pseudorapidity. The electromagnetic calorimeter is divided into a central barrel covering $|\eta| < 1.475$ and a complementary end cap (EMEC) on either side, covering $1.375 < |\eta| < 3.2$. The region beyond the EMEC and close to the beam pipe (the so-called 'forward' region of the detector) is supplemented with additional liquid-argon electromagnetic and hadronic calorimeters. This foward calorimeter (FCAL) system is used to measure particles from small-angle scattering interactions in the range $3.1 < |\eta| < 4.9$. The electromagnetic modules use copper as the absorbing material, while tungsten is used for the hadronic FCAL modules. Analyses sensitive to the hadronic activity in the forward region of the detector rely on the FCAL which achieves energy resolutions of 20 % and 10 % for pions of around 20 GeV and 200 GeV [18], respectively. Each FCAL is surrounded by a liquid-argon hadronic end cap (HEC) calorimeter, covering the range $1.5 < |\eta| < 3.2$ and joining up with the EMEC calorimeter discs. This entire calorimetric system is enclosed by an additional hadronic sampling calorimeter which is divided into a central barrel covering $|\eta| < 1.0$ and complementary extended barrel parts on each end covering the range $0.8 < |\eta| < 1.7$. The outer hadronic calorimeter (HCAL) uses iron plates as the absorbing material that interacts strongly with incident hadrons from a jet, producing charged particles in the process. The space between the iron plates is filled with scintillating plastic tiles, serving as the active sampling material. The energy deposition in a given tile cell can then be inferred from the light signal of the scintillator as recorded by two photomultiplier tubes on each end of the tile cell. The intrinsic energy resolution of the tile calorimeter (excluding the ID tracking information) is about
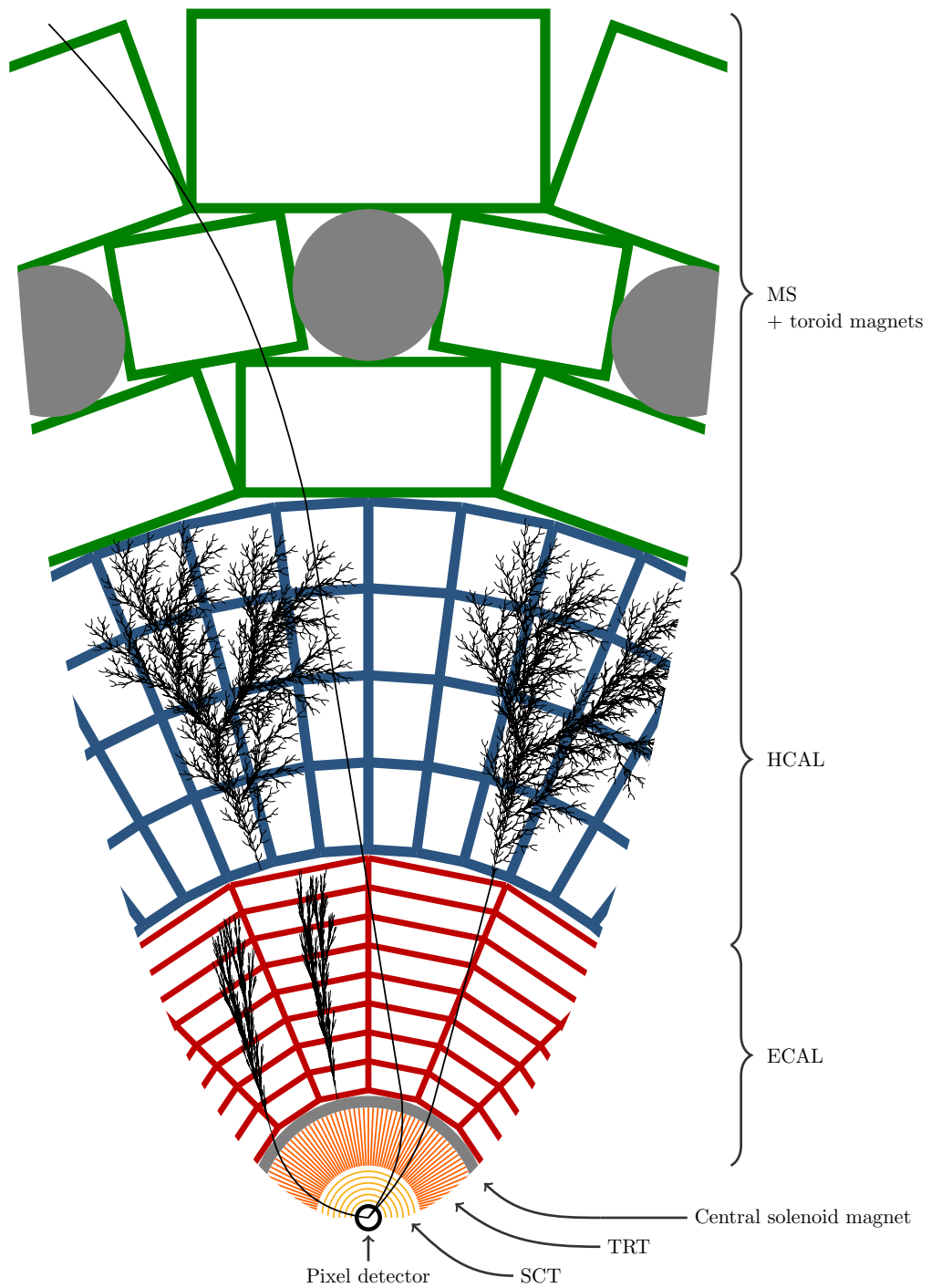
Figure 2.4
Schematic view of the ATLAS detector and its various subdetectors (not to scale). The view is limited to a conic section from the pixel detector surrounding the beam pipe out to the muon spectrometer. Typical (simplified) particle decay chains are indicated as well. Refer to main text for details.

15 % for a jet of around 100 GeV and about 3 % for a jet of around 1 TeV, depending on the pseudorapidity of the jet.

The muon spectrometer (MS) is by far the largest component of the ATLAS detector, surrounding the calorimeter system with additional end caps covering each end of the barrel. Muons and neutrinos are the only Standard Model particles expected to pass through the muon spectrometer. Whilst neutrinos escape the detector undetected, such that their presence can only be inferred via the apparent momentum imbalance in the transverse plane, muon trajectories can be directly identified with the MS tracking system. Design variations of multiwire proportional chambers are used for either fast triggering or precision tracking. Muons are detected in the barrel region ($|\eta| < 1.05$) using resistive plate chambers, whereas thin-gap chambers are used for the end-cap regions ($1.05 < |\eta| < 2.4$) due to their higher rate capability. Both technologies achieve timing resolutions within 25 ns, which is the spacing between proton bunches in the beams delivered by the LHC. The MS tracking system is therefore capable of identifying the associated proton bunch crossing.[1]. The MS incorporates a toroid magnet system in the barrel region as well as the end-cap regions to allow for standalone momentum measurements, as illustrated in figure 2.4. Monitored drift tubes are used to precisely measure the muon momentum in both barrel and end-cap regions ($|\eta| < 2.0$), except close to the beam pipe ($2.0 < |\eta| < 2.7$) where cathode strip chambers are employed due to their superior timing resolution and rate capability. In the former case, an incoming muon ionises a gas mixture within the drift tube, where an electrical field subsequently causes the liberated charge to drift towards the coaxial anode wire running along the centre of the tube. The precise position of the muon with respect to the anode wire can then be inferred from a measurement of the drift time which is proportional to the drift distance. The conceptual variation in the latter case is that here the position of the muon is estimated by interpolating charge depositions on adjacent cathode strips. The standalone momentum resolution of the MS tracking system is about 4 % for a muon of around 100 GeV and between 8 % and 11 % for a muon of around 1 TeV, depending on the pseudorapidity of the muon.

2.3   The ATLAS trigger system

In 2012 the LHC delivered beams of proton bunches with a spacing of 50 ns (twice as long as the design value of 25 ns), corresponding to an event rate of 20 MHz. With on average 20 proton-proton interactions per bunch crossing, the average interaction rate was around 400 MHz in 2012, a number far too high to be fully reconstructed on the fly, let alone recorded. The ATLAS trigger system is a chain of algorithms used to identify potentially interesting signatures in the detector and ensure that only these interesting events are recorded and stored on disk.

⚛   The hardware-based Level 1 trigger system reduces the event rate down to 75 kHz by looking for interesting events using coarse data input from the calorimeter

---

[1]Bunch-crossing identification becomes relevant if particles of more than one proton bunch crossing are present in the detector.

system as well as the muon spectrometer. The electronics of the Level 1 trigger system are capable of identifying cones in the η-ϕ direction, so-called 'regions of interest' (RoI), containing muon-candidate tracks or clusters of energy depositions in the calorimeter. Regions of interest are flagged up by the Level 1 trigger if the energy estimate for these objects exceeds predefined thresholds. The system has a target latency of around 2 μs, about a quarter of which is used to scan the detector and make a decision. Most of the time is spent on transmitting the data of an accepted event via cables to a computer farm near the detector for further analysis. If no RoI is found, the event is rejected and the next event can be processed. Pipelined front-end electronics allow for incoming data to be held in memory, while the Level 1 algorithms are busy processing events in parallel.

⚛ An RoI found by the Level 1 trigger will subsequently seed the software-based Level 2 (L2) trigger system, which is run on a dedicated computer farm close to the detector. The L2 trigger has access to the ID tracking information in addition to the input from the calorimeter system and the muon spectrometer, however, only the RoI identified by the Level 1 trigger is read out in the interest of speed. On average the L2 algorithm has around 40 ms to scan the RoI and make a decision by reconstructing tracks and employing pattern-recognition techniques. The L2 trigger reduces the event rate to a maximum of 5 kHz.

⚛ An event that has been deemed worthy by the L2 trigger will seed the event filter (EF) trigger which is run on the same computer farm close to the detector. This software-based trigger has around a second to read out the full event information and make a final decision as to whether the event is to be stored on disk for offline analysis. The EF software is very similar to that used in the offline analysis and is also able to compensate for some detector effects such as calibration corrections. The EF algorithm reduces the event rate to below 400 Hz.

The two software-based triggers are commonly referred to as the high-level trigger (HLT), which employs object reconstruction algorithms in order to achieve very high event selection efficiencies. The HLT algorithms reduce the event rate down to just a few hundred hertz which can be recorded to disk, requiring a few hundred megabytes of disk space per second. In 2012 ATLAS alone recorded around six petabytes worth of data, which is the equivalent of over a million DVD-type optical disks (each single-layered with 4.7 GB of storage space). In more scientific terms, the integrated luminosity L_{int}, which has dimensions of inverse area, is a measure for the amount of data collected in a given period of time. Figure 2.5 shows the total time-integrated luminosity delivered by the LHC (in green), recorded by ATLAS (in yellow) and deemed of good quality for physics analyses (in blue) for proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 8$ TeV in 2012 [19]. The recorded luminosity represents the inefficiency of the full ATLAS data acquisition system, including the front-end electronics, buffering, triggering and data handling.
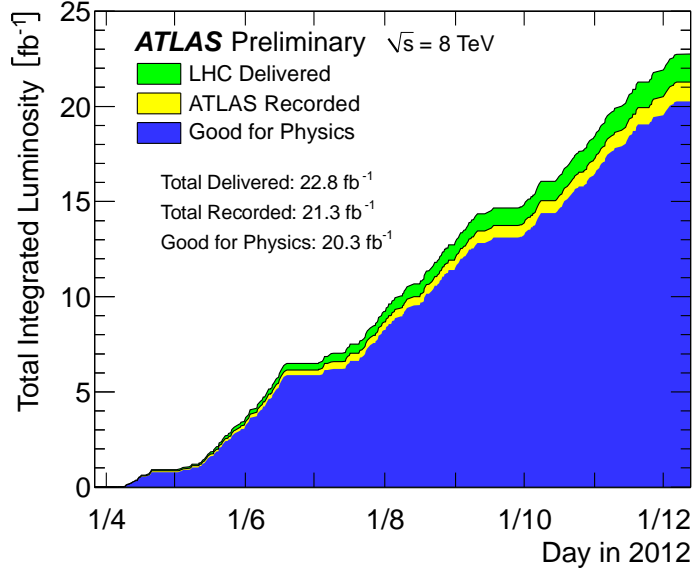
Figure 2.5
Total integrated luminosity delivered by the LHC (green), recorded by ATLAS (yellow) and of good quality during stable beam conditions (blue) for proton-proton collisions at $\sqrt{s} = 8$ TeV in 2012. The luminosity shown represents the preliminary luminosity calibration for the 2012 run. The delivered luminosity accounts for the luminosity delivered from the start of stable beam conditions until the LHC requests ATLAS to put the detector in a safe standby mode to allow a beam dump or beam studies. The recorded luminosity represents the inefficiency of the ATLAS data acquisition. The data quality has been assessed after reprocessing of the recorded data [19].

2.4 Performance of the HLT tracking algorithms

The input to the ID tracking algorithm employed by the L2 trigger system is a set of coordinate triplets corresponding to the signal hits in the pixel and SCT detectors. First the algorithm determines the position of potential proton-proton interaction vertices along the z-direction by extrapolating from the input coordinates in the innermost detector layers back to the beam line. The pseudorapidity of all coordinate points is then calculated with respect to the extrapolated z-position. Tracks originating from the hard scattering interaction tend to have higher transverse momenta, such that their associated signal hits in the pixel and SCT detectors will tend to have similar coordinates in η-ϕ space. Hence clusters of signal hits in η-ϕ space are identified as track candidates if at least four detector layers have contributed to the cluster. This removes a lot of the fake track candidates due to random alignments of coordinate points. In order to remove noise from the remaining track candidates and to improve the track quality, a similar clustering procedure is then performed in ϕ-p_{T}^{-1}-space for the signal hits belonging to the remaining track candidates. The inverse transverse momentum is proportional to the track curvature. Finally, the track parameters are extracted from the surviving clusters using a Bayesian fitting approach suited for noisy measurement environments and a track extension out to the TRT can be used to improve the resolution of the track transverse momenta. Tracks reconstructed by the L2 trigger system are then passed on to the EF tracking software which has a similar level of sophistication as that used in the offline reconstruction.
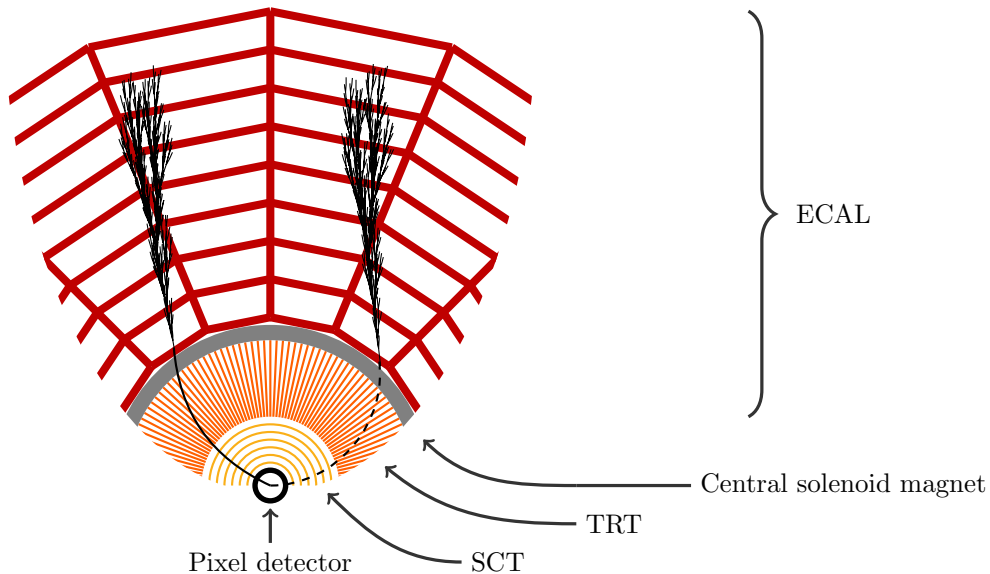
Figure 2.6
Schematic view of the ATLAS detector limited to a conic section from the pixel detector surrounding the beam pipe out to the electromagnetic calorimeter (not to scale). Two (simplified) electron decay chains are indicated to illustrate the principle of a tag-and-probe analysis in the $Z \to e^+ e^-$ channel. Refer to main text for details.

The performance of the algorithms used to reconstruct ID tracks in the high-level trigger systems is evaluated using a tag-and-probe-based technique. The analysis principle is illustrated for a simplified $Z \to e^+ e^-$ candidate event in figure 2.6, which shows a schematic view of the ATLAS detector from the pixel detector surrounding the beam pipe out to the electromagnetic calorimeter. The two simplified electron decay chains are indicated by curved tracks (solid or dashed) that pass through the ID tracking system and turn into electromagnetic cascade showers as they enter the calorimeter system. In a tag-and-probe analysis, a 'tag' lepton is required to be reconstructed by the trigger in both the ID tracking system and an independent subdetector. In the shown example, the RoI associated with the tag electron is required to contain an energy deposition in the ECAL as well as a reconstructed ID track. Furthermore, a 'probe' lepton is required to be reconstructed by the trigger in the independent subdetector only. Note that the ID tracking algorithms are still executed for the probe RoI, even though this tracking information does not affect the trigger decision, which is to say that an ID track is not specifically required to be present in the RoI that contains the energy deposition of the probe lepton in the calorimeter. This is indicated by a dashed track in figure 2.6. The performance of the ID tracking algorithm can be evaluated using $Z \to \ell^+ \ell^-$ candidate events, for which the ID track is expected if the invariant mass of the tag-and-probe pair is close to the Z-boson resonance. Efficiency measurements are performed using both $Z \to e^+ e^-$ and $Z \to \mu^+ \mu^-$ candidate events using proton-proton collision data recorded at a centre-of-mass energy of $\sqrt{s} = 8$ TeV between April and December 2012 [20].

2.4.1 Electron tracking efficiencies

The HLT algorithms reconstruct tag electron candidates by identifying, matching and combining inner detector tracks with clusters of energy in the electromagnetic calorimeter. They are required to have a minimum transverse momentum of 24 GeV and are reconstructed within $|\eta| < 2.5$. Probe electron candidates are reconstructed by only identifying clusters of energy in the electromagnetic calorimeter. These are required to have a minimum transverse momentum of 15 GeV and are reconstructed within $|\eta| < 2.5$. The invariant mass of the tag-and-probe pair is restricted to 70 GeV $\leq m_{\rm tp} \leq$ 110 GeV. In order to suppress the contribution from fake electron candidates further, tag-and-probe pairs are only considered if their calorimeter clusters can be matched to those of well-defined electron candidates that have been fully reconstructed offline, where the maximum matching distance between the HLT clusters and the offline reference clusters is at most $\Delta R = 0.15$ in η-ϕ space[2]. Note that in this case the performance of the HLT algorithms is therefore evaluated with respect to the ID tracking performed offline. Offline electron candidates are reconstructed by identifying, matching and combining inner detector tracks with clusters of energy in the electromagnetic calorimeter. They are reconstructed within $|\eta| < 2.5$ and are required to satisfy a set of 'tight' selection criteria [21] that have been reoptimised for the higher rate of proton-proton collisions per bunch crossing (pileup) observed in the 8 TeV data. Furthermore, the minimum transverse momentum of the offline calorimeter clusters are required to be greater than 24 GeV or 15 GeV, depending on whether the offline electron is associated with the tag electron candidate or the probe electron candidate.

The track expected for the probe electron is considered to have been successfully reconstructed by a given HLT algorithm if an ID track reconstructed in the probe RoI can be matched ($\Delta R < 0.03$) to the offline-reconstructed ID track of the associated reference electron. The tracking efficiency for an HLT algorithm with respect to the offline tracking is therefore defined as the number of events for which an ID track reconstructed by the HLT algorithm can be matched to the offline-reconstructed ID track divided by the number of tag-and-probe pairs being probed. Tracking efficiencies for both the L2 and EF algorithms are presented as a function of the transverse momentum and the pseudorapidity of the offline track belonging to the reference electron that has been matched to the probe electron in figures 2.7(a) and 2.7(b), respectively. The efficiencies are generally very high with a small inefficiency from offline tracks with transverse momenta of around 15 GeV. The efficiencies degrade slightly with pseudorapidity as the edges of the ID tracking system are approached. This is because for larger pseudorapidity values the electrons have to traverse more material in order to reach the calorimeter system and are therefore more likely to lose energy due to bremsstrahlung, which makes it more difficult to reconstruct the ID track. The ratio of the offline track transverse momentum to the transverse energy of the corresponding offline calorimeter cluster is a measure of the energy loss due to bremsstrahlung. Tracking efficiencies for both the L2 and the EF algorithms as a function of this ob-

---

[2]The dimensionless radius parameter is defined via the relation $(\Delta R)^2 = (\Delta \eta)^2 + (\Delta \phi)^2$.
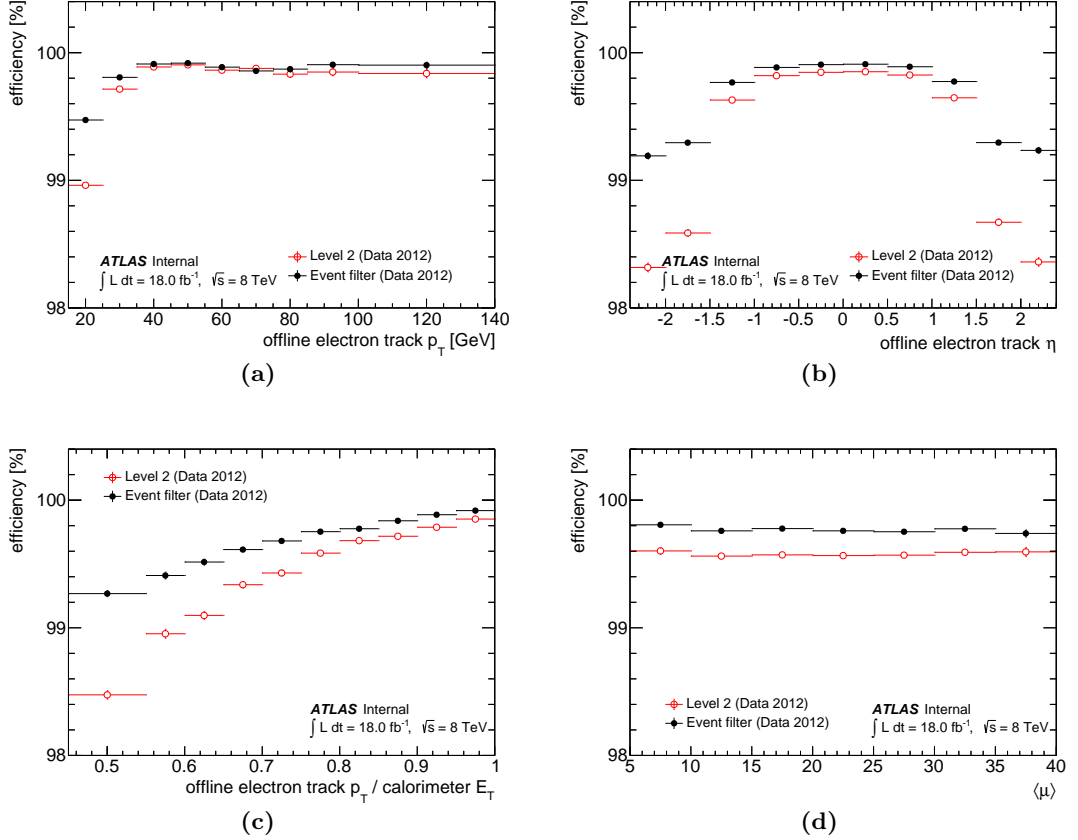
Figure 2.7
L2 and EF electron tracking efficiencies as a function of the (a) transverse momentum and (b) pseudorapidity of the offline track belonging to the reference electron that has been matched to the probe in η-ϕ space. Similar tracking efficiencies are also shown as a function of the (c) ratio of the offline track transverse momentum to the transverse energy of the corresponding offline calorimeter cluster and, furthermore, the (d) average number of proton-proton interactions per bunch crossing, $\langle\mu\rangle$. The error bars indicate the statistical uncertainties of the data.

servable are shown in figure 2.7(c). The efficiencies are highest where the ratio is close to unity, corresponding to little amounts of bremsstrahlung emitted by the electrons. Consequently, a decrease in efficiency can be observed towards smaller ratio values, i.e. the regime of increased bremsstrahlung where it becomes harder to reconstruct the ID track. Finally, the tracking efficiencies for both the L2 and the EF algorithms as a function of the average number of proton-proton interactions per bunch crossings, $\langle\mu\rangle$, are presented in figure 2.7(d). Note that the efficiencies are both very high and essentially flat as function of $\langle\mu\rangle$ for the 2012 dataset which is encouraging.

2.4.2 Muon tracking efficiencies

The HLT algorithms reconstruct tag muon candidates by identifying, matching and combining inner detector tracks with track segments in the muon spectrometer [22]. They are required to have a minimum transverse momentum of 24 GeV and are reconstructed within $|\eta| < 2.5$. Probe muon candidates are reconstructed by identifying track segments in the muon spectrometer only. These are required to have a minimum transverse momentum of 15 GeV and are reconstructed within $|\eta| < 2.5$. The invariant mass of the tag-and-probe pair is restricted to 75 GeV $\leq m_{\mathrm{tp}} \leq$ 105 GeV. No addi-

(a)

(b)

(c)
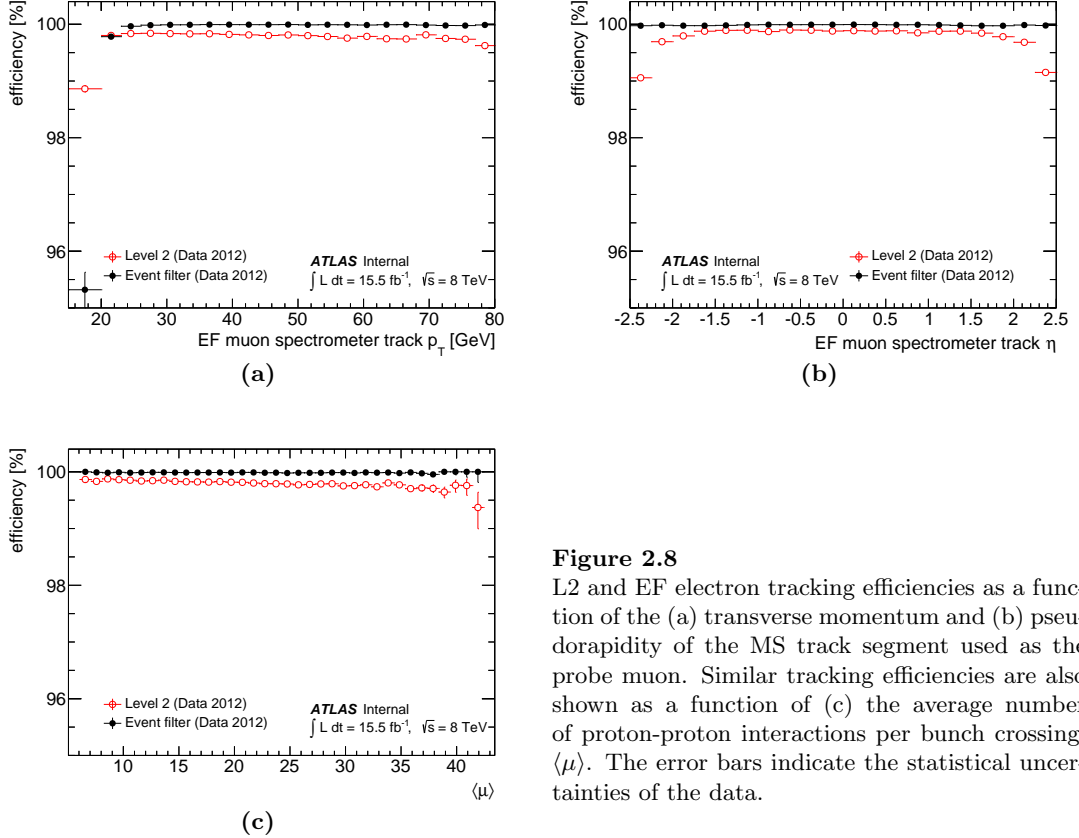
Figure 2.8
L2 and EF electron tracking efficiencies as a function of the (a) transverse momentum and (b) pseudorapidity of the MS track segment used as the probe muon. Similar tracking efficiencies are also shown as a function of (c) the average number of proton-proton interactions per bunch crossing, $\langle\mu\rangle$. The error bars indicate the statistical uncertainties of the data.

tional matching to offline reference muons has been performed as contributions from fake lepton candidates are less of an issue in the muon channel. Note that in this case the performance is evaluated irrespective of any offline processing and therefore reflects the absolute efficiency of the HLT tracking algorithms.

The track expected for the probe muon is considered to have been successfully reconstructed by a given HLT algorithm if an ID track reconstructed in the probe RoI can be paired with the tag muon such that the invariant mass of the combined tag-and-track system lies within 40 to 140 GeV. This ensures that the ID track belongs to the muon coming from the Z boson and avoids potential inefficiencies due to the resolution of the MS track, which may prevent a successful match to the ID track in η-ϕ space. The mass window used for the mass-hypothesis criterion has been extended with respect to the mass window used to select the tag-and-probe pairs in order to account for the different mass resolution. Tracking efficiencies for both the L2 and EF algorithms are presented as a function of the transverse momentum and the pseudorapidity of the MS track segment used as the probe muon in figures 2.8(a) and 2.8(b), respectively. The efficiencies are generally very high for the L2 algorithm with a small inefficiency towards the edges of the ID tracking system as well as a minor drop in efficiency for pseudorapidities around ± 0.9 which may be due to a poor muon spectrometer resolution in that area. Note that the material-dependent drop in efficiency as function of the pseudorapidity is less pronounced than in the electron case since energy losses due to bremsstrahlung are less of an issue for the much heavier muons. A small inefficiency is observed for MS tracks of around 15 GeV. The efficiencies for the EF algorithm are

generally even higher as they have been evaluated with respect to the L2 algorithm, however, a drop in efficiency is also seen for MS tracks of around 15 GeV. The size of the error bar here indicates statistical limitations of the reference sample (given by the tag-and-probe pairs) in this region of phase space. This suggests that the drop in efficiency may not reflect an inefficiency of the HLT algorithms, but that it could instead be due to some phase-space cut that implicitly affects the kinematics of the probe muon. Finally, the tracking efficiencies for both the L2 and the EF algorithms as a function of the average number of proton-proton interactions per bunch crossing, $\langle \mu \rangle$, are presented in figure 2.8(c). These are generally very high and only drop by about $0.1\,\%$ across the shown multiplicity range which is acceptable.

3

Strong and electroweak Zjj production

The production of two jets in association with a leptonically decaying Z boson (Zjj) at the LHC predominantly involves the strong interaction. In addition, Zjj events can be produced as a result of quark-quark scattering proceeding through the intermediary of a vector gauge boson. This production mechanism, however, is a purely electroweak process and therefore much rarer. Electroweak Zjj production is defined as all the contributions to $\ell^+\ell^- jj$ production involving a t-channel exchange of an electroweak gauge boson [23, 24]. At leading order, this gives rise to a gauge invariant set of seven diagrams, as depicted in figure 3.1, including contributions from Z-boson bremsstrahlung, non-resonant production as well as the vector boson fusion (VBF), where the Z boson is radiated off the propagator. The latter production mechanism, as shown in figure 3.1(a), is particularly interesting due to its similarity to the VBF production of a Higgs boson as well as its sensitivity to new physics via the WWZ triple gauge coupling.

Inclusive Zjj production also includes a contribution coming from diboson events (ZV) where one of the vector bosons decays hadronically, as illustrated in figure 3.2 for ZW production in both the t-channel and the s-channel. Albeit a purely electroweak process, diboson-initiated Zjj production does not count towards this definition of electroweak Zjj production since it is lacking the necessary t-channel exchange of the gauge boson. For electroweak Zjj production, the two jets arise from quarks recoiling against the t-channel propagator, whereas for diboson-initiated Zjj production the two jets are the decay products of one of the gauge bosons, resulting in both kinematic and topological differences. The contribution of diboson events to inclusive Zjj production is at the percent level and can be reduced even further by requiring the invariant mass of the dijet system to be well beyond the gauge boson masses.

A selection of leading-order diagrams for strong Zjj production at the LHC is given in figure 3.3. The dominant production mechanism is via the Drell-Yan process with additional parton[1] radiation due to the strong interaction, as shown in figure 3.3(a). Additional contributions to strong Zjj production are due to t-channel and s-channel

---

[1]'Parton' is a collective term for the proton constituents, i.e. quarks and gluons.
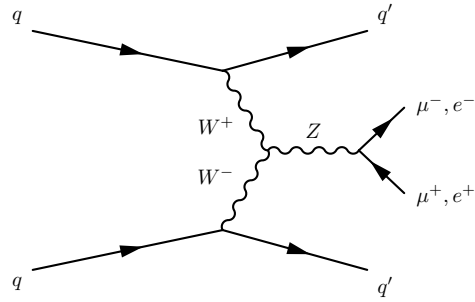
(a) vector boson fusion (with triple gauge boson coupling)
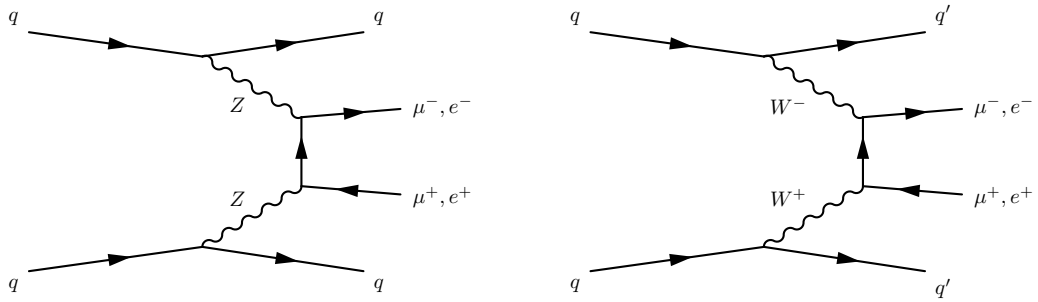
(b) non-resonant $\ell^+\ell^- jj$
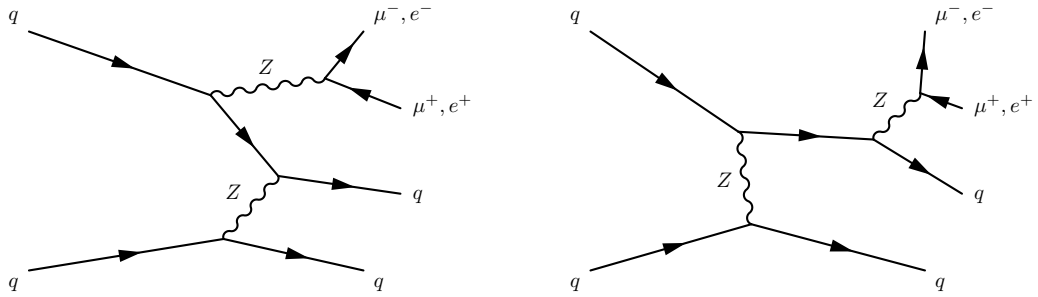
(c) Z-boson bremsstrahlung (and t-channel exchange of a Z boson)
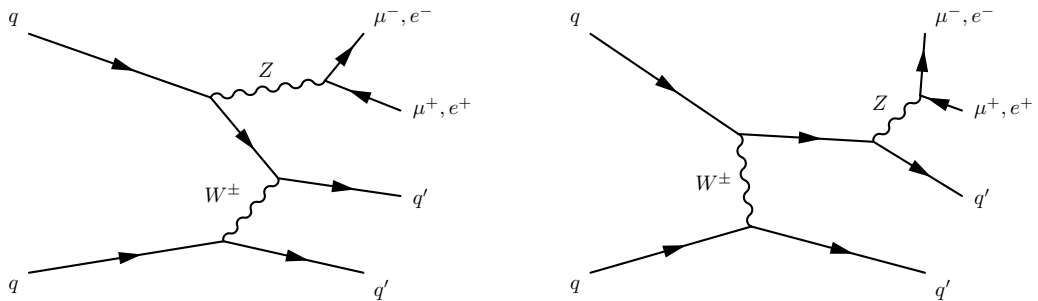
(d) Z-boson bremsstrahlung (and t-channel exchange of a W boson)

Figure 3.1 Leading order Feynman diagrams for electroweak Zjj production at the LHC.

(a) *t*-channel $\qquad\qquad\qquad$ **(b)** *s*-channel
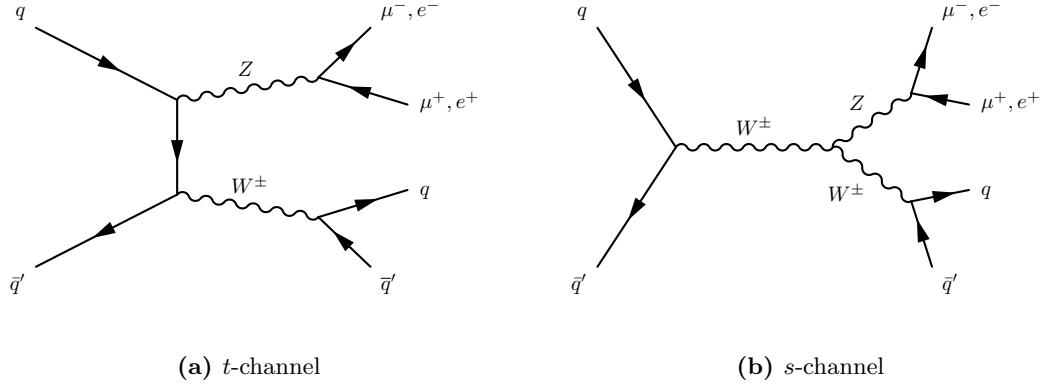
Figure 3.2   Representative leading order Feynman diagrams for diboson-initiated Zjj production.

parton exchanges with additional Z-boson bremsstrahlung off a quark as illustrated in figure 3.3(b) and 3.3(c), respectively.

Inclusive Zjj production consists of both strong and electroweak Zjj production as well as the small diboson-initiated contribution. Typical background processes to inclusive Zjj production stem from $t\bar{t}$ events, single t-quark interactions as well as $W + \text{jets}$ and WW events. As these processes do not contain a real Z boson, their relative contribution can be substantially suppressed by requiring a Z-boson candidate with an invariant mass close to the Z-boson resonance.

(a) Drell-Yan-like interactions with additional parton radiation

(b) *t*-channel exchange with additional Z-boson bremsstrahlung

(c) *s*-channel exchange with additional Z-boson bremsstrahlung
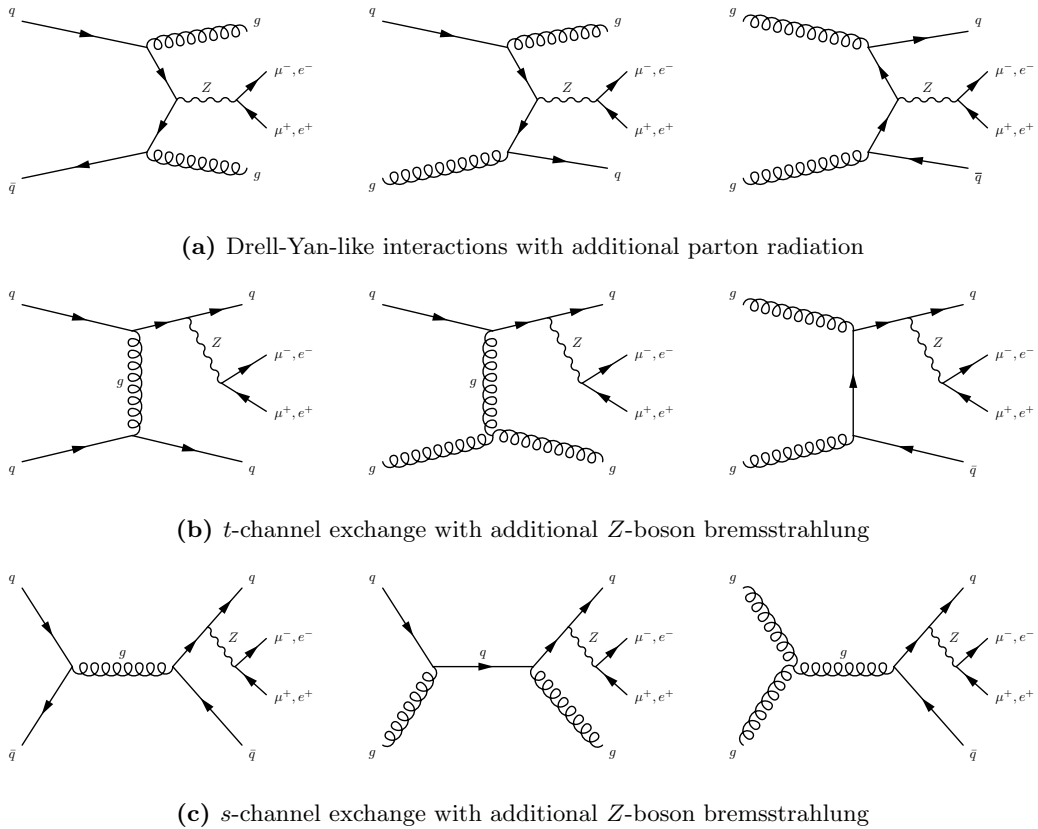
Figure 3.3   Selection of leading order Feynman diagrams for strong Zjj production at the LHC.
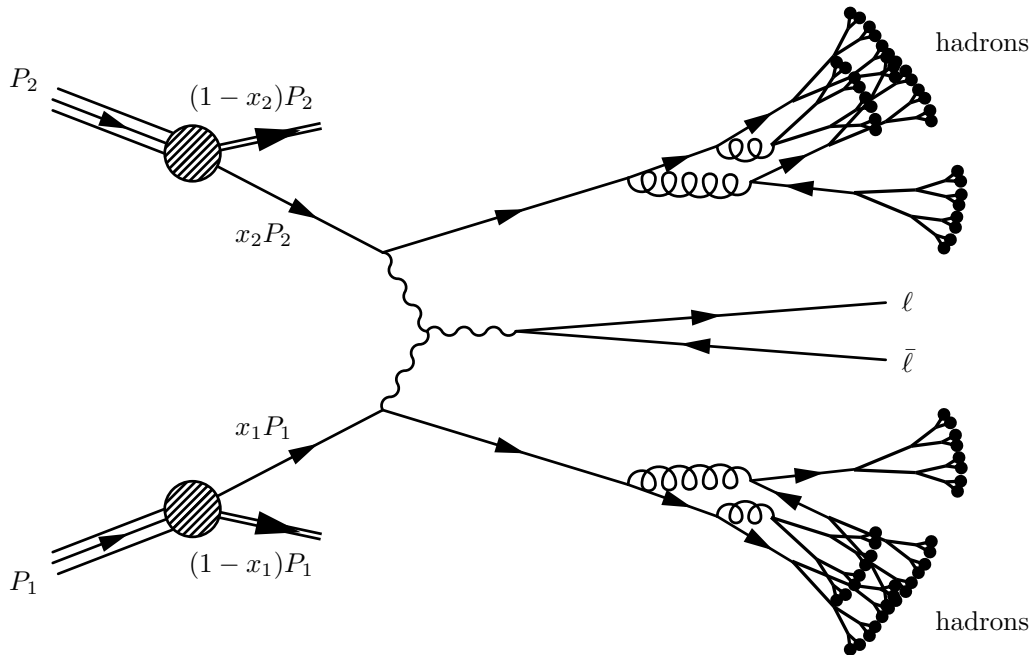
Figure 3.4
Feynman diagram for electroweak Zjj production via the vector boson fusion process. The schematic is extended to include the original protons and shows a simplified representation of the parton evolution into the fully hadronic final state. The evolution of the proton remnants is omitted for clarity.

3.1 Modelling of high energy collisions using Monte Carlo

The hard scattering interaction for processes such as VBF Zjj production, as shown in figure 3.1(a), is only a small part of the proton-proton collision process, most of which is not calculable in perturbative QCD and is usually modelled using different phenomenological approaches instead. Most of the techniques used to model high energy collisions heavily rely on the Monte Carlo method which is based on repeated random sampling and therefore ideally suited for estimating unknown probability distributions, simulating stochastic outcomes as well as fast numerical integration, especially when an analytical solution is challenging to obtain. Figure 3.4 gives a slightly more complete picture of a single collision between two incoming protons carrying momentum P_1 and P_2, respectively. The probability of finding a parton with a momentum fraction x of the overall proton momentum is described by the corresponding parton distribution function (PDF). The functional form of the PDF, however, is not predicted by the theory and instead has to be extracted via a global fit to the available proton collision data collected in deep inelastic scattering experiments.

The momentum exchange between two partons in each of the protons is referred to as the hard scattering interaction and is factorised in the cross-section calculation. This factorisation introduces an artificial energy scale that determines what part of the physics is described by the PDF and the matrix element[2], respectively. This factori-

---

[2]This is the same S-matrix used to calculate the scattering amplitude in quantum field theory.

sation scale is unphysical and any dependence of the final calculation on it provides an estimate of the theory uncertainty associated with the prediction. Although the factorisation of the hard scattering process in the cross-section calculation has never formally been proven, there are strong arguments that the factorisation theorem holds for many processes, including Drell-Yan-like interactions [25, 26].

In figure 3.4, the two partons participating in the hard scatter carry momentum $x_1 P_1$ and $x_2 P_2$, respectively. At small distances (i.e. at high energies) the strong coupling α_S is small and so the matrix element for the hard scattering interaction can be approximately calculated using a perturbative expansion in α_S. Real (parton emission) and virtual (parton loop) corrections are then represented by higher order terms in the series. Individual terms tend to diverge when integrating over the full energy spectrum. For example, corrections due to real emissions become divergent in the collinear limit (i.e. at small angles or at small momenta). It so happens that these divergences are usually exactly cancelled by divergences due to virtual corrections in the infrared limit (i.e. when the loop parton has small momentum), such that their sum yields a finite result. However, an incomplete cancellation between virtual and unresolvable (real) contributions can yield finite remainders of the divergences, giving rise to logarithmically enhanced terms that spoil the convergence of the perturbative series. The convergence can be restored through a resummation of the enhanced collinear and soft terms to all orders in α_S into a so-called Sudakov form factor. Moreover, virtual corrections can also diverge at high energies. This is sometimes referred to as the 'ultraviolet catastrophe' since there are no cancelling terms. These divergences can be circumvented using dimensional regularisation, which in practice requires the introduction of an additional arbitrary energy scale known as the renormalisation scale. Much like the factorisation scale, this energy scale is unphysical and any dependence of the final calculation on it provides an estimate of the theory uncertainty associated with the prediction. In general, going beyond the leading order terms results in a more accurate prediction at the expense of quickly growing computational effort.

The matrix-element calculation yields a parton-level prediction involving colour-charged objects which may or may not radiate additonal partons as they evolve down to smaller energies. A parton shower models the branching of these partons in analogy to the radioactive decay of a nucleus where the probability that no resolvable emission occurs down to a certain energy scale is then just given by a Sudakov form factor. The Sudakov form factor effectively resums the colliner and soft QCD corrections to all orders, thereby avoiding the hard wide-angle emissions that are taken care of by the matrix-element calculation. Various matching and merging schemes exist to improve the accuracy of the parton-shower modelling in relation to the matrix-element calculation [27–35]. The parton evolution is cut off at a small energy scale beyond which the parton-shower model breaks down since the strong coupling can no longer be considered small. This is where a hadronisation model takes over; popular choices include the cluster model [36] as well as the Lund string model [37]. The partons from the parton-shower cascade will be combined according to their colour charge which typically yields resonances (such as excited kaons) that will then decay into the fully

hadronic final state. This collimated QCD particle cascade is commonly referred to as a jet and its precise definiton depends on the choice of jet reconstruction algorithm used to estimate the number of outgoing partons in the hard scattering interaction. Various iterative procedures exist [38–42] that aim to cluster hadrons sequentially based on their proximity for instance.

The proton remnants with momenta $(1 - x_1)P_1$ and $(1 - x_2)P_2$ are evolved using similar parton-shower and hadronisation models and tend to contribute (soft) energy to the final-state objects. This contribution is referred to as the underlying event and is accounted for in the simulation. In principle, the evolution of the proton remnants can also interfere (e.g. via gluon exchange) with the parton-shower cascades stemming from the hard scattering interaction, which adds to the complexity. Furthermore, the simulation of the underlying event also accounts for multiple parton interactions (MPI), i.e. additional hard interactions in the same proton-proton collision.

3.2 Theoretical predictions and Monte Carlo simulation

Particle-level predictions for strong [43] and electroweak [44–46] Zjj production are obtained at next-to-leading order accuracy using the `Powheg Box` [47–49]. The fully hadronic final state is constructed by interfacing the `Powheg Box` with `Pythia 6` [50] for the parton shower model as well as modelling algorithms for the subsequent hadronisation and contribution due to multiple parton interactions in the same proton-proton collision. These particle-level predictions, hereafter referred to as the `Powheg` predictions, are produced using the `CT10` [51] parton-distribution functions and the Perugia 2011 tune [52] for the simulation of underlying event activity. The strong Zjj calculation is augmented with the `MiNLO` feature [53], resulting in an improved resummation of Sudakov logarithms. In summary, the `Powheg` predictions have next-to-leading order accuracy for $Z + 2$ partons, leading order accuracy for the zero-, one- and three-parton configurations and leading logarithmic accuracy for Z production in association with four partons.

In addition to the `Powheg` predictions, particle-level predictions for all contributions of inclusive Zjj production have been obtained using the `Sherpa` v1.4.3 [54] event generator. `Sherpa` combines a leading order matrix-element generator for the calculation of the $Z + n$-parton scattering process ($n = 0, 1, 2, \ldots$) with a parton-shower model for the QCD parton cascades as well as additional parton radiation in the inital and final states. The approach by Catani, Krauss, Kuhn and Webber (CKKW) is employed [55] in order to merge the various final-state topologies and match to the parton shower. Multi-parton matrix elements are calculated for two and three partons in the final state of the electroweak scattering process, while the strong scattering process also includes the configuration with four partons in the final state. The zero- and one-parton configurations have also been calculated for strong Zjj production in order to include contributions from double-parton scattering. To this end, the `MENloPS` feature [56] has been employed for the strong Zjj sample which merges an inclusive $Z/\gamma^* \to \ell^+\ell^-$ calculation performed at next-to-leading order accuracy with the leading order calculations

for $Z + n$ jets, thereby reducing theoretical uncertainties associated with the inclusive Z/γ^*-boson production[3]. Diboson-initiated Zjj production has been calculated with up to three partons in addition to the partonically decaying boson. The fully hadronic final states are then again constructed from the parton-shower model and modelling algorithms for subsequent hadronisation and MPI contributions. All `Sherpa` samples are produced using the `CT10` PDF predictions as well as the default generator tune for the simulation of underlying event activity. The leading order `Sherpa` predictions for strong and electroweak Zjj production are normalised to the next-to-leading order predictions taken from `Powheg`, using factors of 1.23 and 1.02, respectively.

Background events arising from $t\bar{t}$ and single t-quark interactions have been simulated with `MC@NLO` v4.03 [57] interfaced with `Herwig` [58, 59] and `Jimmy` [60] (using the `AUET2` [61] tune) for the parton-shower model and contributions due to the underlying event activity. The generator modelling of the $t\bar{t}$ background events has been examined by comparing it to a prediction taken from the `Powheg Box` interfaced with `Pythia 6` using the Perugia 2011 tune. The $t\bar{t}$ predictions are normalised to a next-to-next-to-leading order calculation in perturbative QCD with resummation of next-to-next-to-leading logarithmic soft gluon terms [62]. The background contribution due to $W + $jets and WW events are simulated using `Sherpa`.

The particle-level predictions for all the aforementioned processes, with the exception of the `Powheg` Zjj predictions, have been passed through `GEANT4` [63, 64] for a full simulation [65] of the ATLAS detector. Additional proton-proton interactions in the same bunch crossing of the two beams (pileup) are simulated in the detector-level predictions with `Pythia 8` [66] using tune `A2` [67] in conjunction with the `MSTW2008lo` PDF set [68]. All generated samples have been reconstructed with the same analysis chain that has been used for the data.

3.2.1 Theoretical uncertainties

Theoretical uncertainties on the strong and electroweak Zjj predictions are obtained by generating a number of theory variation samples using the `Sherpa` and `Powheg` generators, respectively. Scale variations are obtained for strong and electroweak Zjj production individually by doubling and halving the factorisation and renormalisation scales used in the calculation. While the renormalisation scale sets a cut-off for loop corrections that ensures convergence of the perturbative calculation, the factorisation scale determines a cut-off that separates the hard scattering interaction from long-distance effects, e.g. due to the parton-distribution functions. Both scales are varied separately with respect to the nominal scale choices in order to examine their impact on the final prediction. Additional theory variation samples are generated for the `Sherpa` predictions by doubling the CKKW matching scale, changing the recoil strategy for dipoles with initial-state emitter and final-state spectator in the parton-shower model from the default [69] to the one proposed in ref. [70], changing the MPI activity uniformly by 10 % [71] and by changing the MPI spectrum altogether, such that more

---

[3]However, note that the analysis is not expected to benefit from this improvement as at least two jets will be required in all regions of phase space (cf. section 3.3.1).

jets from double-parton interactions are produced[4]. Similar modelling uncertainties are derived for the `Powheg` predictions from the set of Perugia 2011 tunes, with the largest effect coming from those tunes with increased or decreased parton-shower activity or increased MPI activity.

The calculation of separate predictions for strong and electroweak Zjj production relies on the assumption that the interference contribution between the strong and electroweak processes is both colour and kinematically suppressed, which so far has only been proven for dijet production in association with a Higgs boson (Hjj) [73–76]. Although no such studies are available for the Zjj system, the interference effects arise from the same sources as in the case of Hjj production and should therefore be small. The assumption of negligible interference is examined in the simulation using an inclusive leading order prediction for Zjj production from `Sherpa`. This inclusive sample combines matrix elements for strong and electroweak Zjj production at the amplitude level and thus formally calculates the interference between them. The interference contribution is then determined by subtracting the individual predictions for strong and electroweak Zjj production. The measurements of inclusive Zjj production are found to be unaffected by the interference contribution, while its impact on the extraction of the electroweak component is found to be at the level of a few percent and discussed in more detail in section 6.3.2.

3.3 Event selection

Theoretical predictions at the particle-level are obtained from final-state objects with a mean lifetime ($c\tau$) longer than $10\,\mathrm{mm}$. Leptons are defined at the so-called 'dressed' level, which is to say they are constructed from the four-momentum combination of the respective lepton (a muon or an electron) and all nearby photons within a cone of radius $\Delta R = 0.1$ centred on the lepton. The dimensionless radius parameter is defined in η-ϕ space via the relation $(\Delta R)^2 = (\Delta\eta)^2 + (\Delta\phi)^2$. Leptons are selected within a pseudorapidity range of $|\eta| < 2.47$ and are required to have a minimum transverse momentum of 25 GeV. Jets are reconstructed using the anti-k_t jet clustering algorithm [40] with a jet-radius parameter of 0.4 and within a rapidity range of $|y| < 4.4$. Jets are required to have a minimum transverse momentum of $p_\mathrm{T} > 25$ GeV and to be well separated from selected leptons, meaning that the distance between a jet and any of the selected leptons ought to be $\Delta R_{j,\ell} \geq 0.3$ in η-ϕ space.

The cross section for inclusive dijet production in association with a Z boson is measured in five fiducial regions, each with varying sensitivity to the electroweak component of Zjj production. A summary of the selection criteria for the different regions is provided in table 3.1 and is discussed in more detail in the following section.

---

[4]This latter variation is motivated by data from an analysis of the underlying event [72]. Specifically, the parameters for this `Sherpa` variation are `SIGMA_ND_FACTOR=0.14` and `SCALE_MIN=4.0`.

Table 3.1
Summary of the selection criteria that define the fiducial regions. 'Interval jets' refer to the selection criteria applied to the jets that lie in the rapidity interval bounded by the two leading jets in the event.

Object	*baseline*	*high-mass*	*search*	*control*	*high-p_{T}*
Leptons			$\|\eta^\ell\| < 2.47,\ p_{\mathrm{T}}^\ell > 25$ GeV		
Dilepton pair			81 GeV $\leq m_{\ell\ell} \leq 101$ GeV		
	—		$p_{\mathrm{T}}^{\ell\ell} > 20$ GeV		—
Jets			$\|y^j\| < 4.4,\ \Delta R_{j,\ell} \geq 0.3$		
		$p_{\mathrm{T}}^{j_1} > 55$ GeV			$p_{\mathrm{T}}^{j_1} > 85$ GeV
		$p_{\mathrm{T}}^{j_2} > 45$ GeV			$p_{\mathrm{T}}^{j_2} > 75$ GeV
Dijet system	—	$m_{jj} > 1$ TeV	$m_{jj} > 250$ GeV		—
Interval jets		—	$N_{\mathrm{jet}}^{\mathrm{gap}} = 0$	$N_{\mathrm{jet}}^{\mathrm{gap}} \geq 1$	—
Zjj system		—	$p_{\mathrm{T}}^{\mathrm{balance}} < 0.15$	$p_{\mathrm{T}}^{\mathrm{balance},3} < 0.15$	—

3.3.1 Fiducial regions

The *baseline* region contains events with a Z-boson candidate, defined as two oppositely charged, same-flavour leptons (muons or electrons) with a dilepton invariant mass of 81 GeV $\leq m_{\ell\ell} \leq 101$ GeV and at least two jets that satisfy $p_{\mathrm{T}}^{j_1} > 55$ GeV and $p_{\mathrm{T}}^{j_2} > 45$ GeV, where j_1 and j_2 refer to the leading and subleading transverse momentum jets in the event, respectively. This is the most inclusive region in the analysis and contains the events in all other regions. The tight cut on the dilepton invariant mass suppresses backgrounds from events that do not contain a Z boson. Note that the contribution from diagrams where the Z boson is replaced by a virtual photon is substantially reduced due to this tight mass requirement. Furthermore, the contribution to the electroweak cross section from non-resonant production (cf. diagrams in figure 3.1(b)) is less than 1 % after applying the selection criteria of the *baseline* region.

The moderately high p_{T} requirements reduce jet-based experimental uncertainties associated with low-p_{T} jets, particularly so in the forward region of the detector. In addition, the p_{T} cuts on the leading two jets suppress strong Zjj events in favour of the electroweak signal. This is illustrated in figure 3.5 which shows shape comparisons of the leading and subleading jet-p_{T} spectra for the signal and background predictions after requiring a Z-boson candidate and at least two jets with $p_{\mathrm{T}} > 25$ GeV. In the electroweak process, the jets are recoiling against a heavy vector boson which results in notably harder signal p_{T} spectra when compared to the corresponding p_{T} spectra due to the background processes, which typically involve the exchange of a quark or gluon.

The *high-mass* region constitutes a subset of the events in the *baseline* region for which the invariant mass of the leading two jets satisfies $m_{jj} > 1$ TeV, whereas the events in the *high-p_{T}* region are a subset of the events in the *baseline* region for which the leading two jets also satisfy the higher p_{T} requirements of $p_{\mathrm{T}}^{j_1} > 85$ GeV and $p_{\mathrm{T}}^{j_2} > 75$ GeV, respectively. The *high-mass* and the *high-p_{T}* regions are useful to probe

(a) leading jet
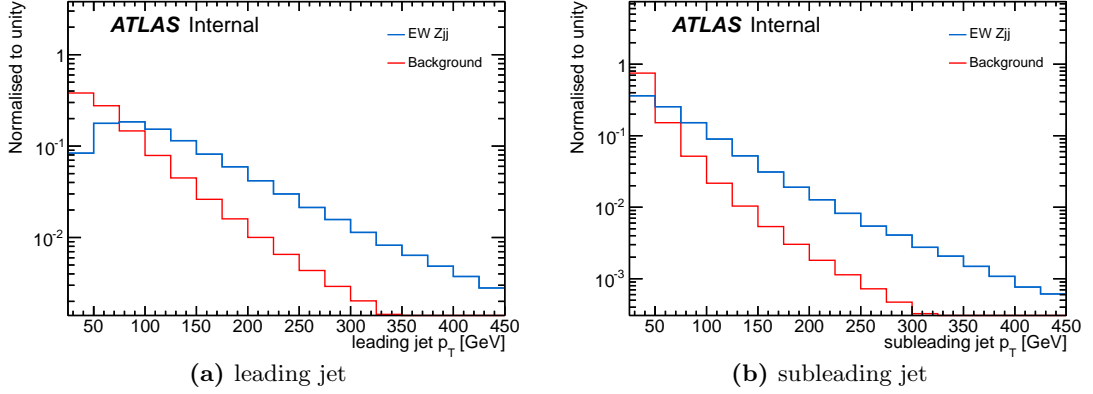
(b) subleading jet

Figure 3.5
Particle-level shape comparisons of the (a) leading and (b) subleading jet p_{T} spectra for the electroweak Zjj signal and the background, which includes strong Zjj production as well as small contributions from diboson processes and $t\bar{t}$ interactions. The distributions are normalised to unity after selecting events requiring a Z-boson candidate and at least two jets with $p_{\mathrm{T}} > 25$ GeV.

(a) $|\Delta y|$

(b) m_{jj}
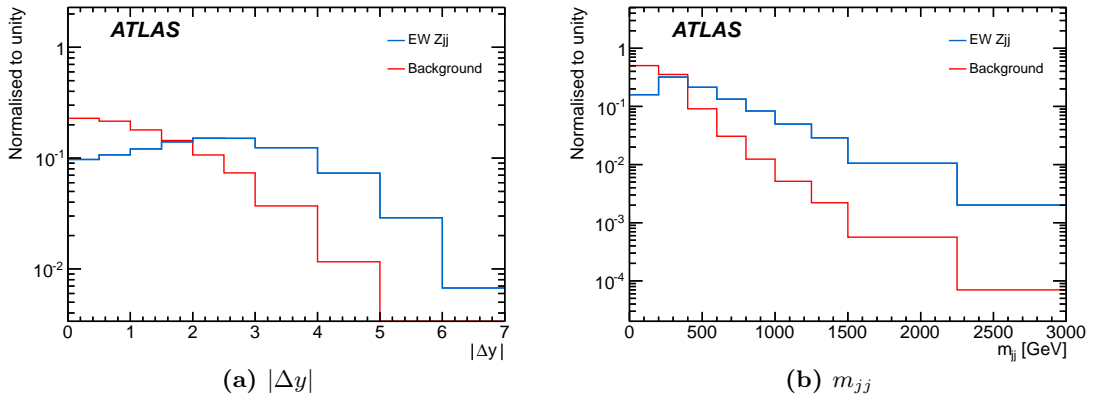
Figure 3.6
Particle-level shape comparisons of the (a) rapidity difference between the leading two jets and the (b) invariant mass spectrum of the dijet system for the electroweak Zjj signal and the background, which includes strong Zjj production as well as small contributions from diboson and $t\bar{t}$ interactions. The distributions are normalised to unity and are shown in the *baseline* region.

the impact of the electroweak process, which produces harder jet-p_{T} spectra, as shown in figure 3.5, and similarly harder distributions sensitive to the dijet kinematics. This can be seen in figure 3.6, which shows shape comparisons of the rapidity difference between the leading two jets, $|\Delta y|$, and the invariant mass spectrum of the dijet system, m_{jj}, for signal and background predictions using the selection criteria of the *baseline* region. It emerges that the electroweak signal is characterised by the presence of two high-p_{T} jets in the forward region of the detector, i.e. widely separated in rapidity, which leads to a large dijet invariant mass.

The t-channel exchange of the colourless gauge boson in the electroweak process implies that additional jet activity due to QCD radiation between the leading two jets is suppressed for the signal. Figure 3.7(a) shows the shape comparison of the number of jets in the rapidity interval between these leading two jets, $N_{\mathrm{jet}}^{\mathrm{gap}}$, for the signal and background predictions. It can be seen that the number of interval jets provides excellent discrimination between the signal and background processes and therefore a veto on additional interval-jet activity is used to split the *baseline* region into a

(a) $N_{\text{jet}}^{\text{gap}}$ **(b)** $p_{\text{T}}^{\text{balance}}$
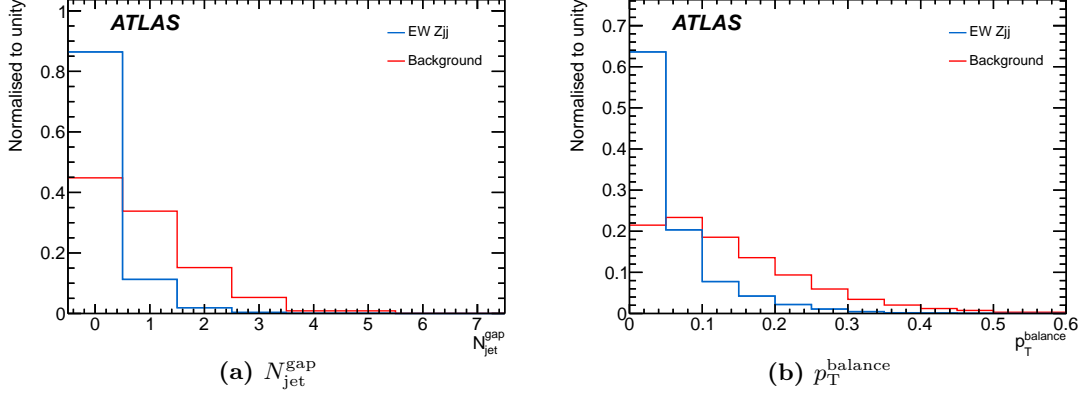
Figure 3.7
Particle-level shape comparisons of the (a) number of jets in the rapidity interval between the two leading jets as well as the (b) $p_{\text{T}}^{\text{balance}}$ distribution for the electroweak Zjj signal and the background, which includes strong Zjj production as well as small contributions from diboson and $t\bar{t}$ interactions. The distributions are normalised to unity and shown in the *high-mass* region.

signal-enhanced and a signal-suppressed subregion. While the former serves to extract the electroweak component, the latter can be used to directly examine the modelling of strong Zjj production. The signal-enhanced *search* region consists of all *baseline*-region events that also satisfy the following selection criteria:

⚛ No additional jets with $p_{\text{T}} > 25$ GeV in the rapidity interval between the leading two jets.

⚛ The invariant mass of the dijet system is required to satisfy $m_{jj} > 250$ GeV, which removes a large fraction of diboson events.

⚛ The transverse momentum of the dilepton pair is required to satisfy $p_{\text{T}}^{\ell\ell} > 20$ GeV. This reduces the impact of events with two jets that do not originate from the hard scattering interaction.

⚛ The normalised transverse-momentum balance between the two leptons and the two leading jets, $p_{\text{T}}^{\text{balance}}$, is required to be less than 0.15. Specifically, the variable $p_{\text{T}}^{\text{balance}}$ is defined as

$$
p_{\text{T}}^{\text{balance}} = \frac{\left| \vec{p}_{\text{T}}^{\,\ell_1} + \vec{p}_{\text{T}}^{\,\ell_2} + \vec{p}_{\text{T}}^{\,j_1} + \vec{p}_{\text{T}}^{\,j_2} \right|}{\left| \vec{p}_{\text{T}}^{\,\ell_1} \right| + \left| \vec{p}_{\text{T}}^{\,\ell_2} \right| + \left| \vec{p}_{\text{T}}^{\,j_1} \right| + \left| \vec{p}_{\text{T}}^{\,j_2} \right|}, \tag{3.1}
$$

where $\vec{p}_{\text{T}}^{\,i}$ is the transverse momentum of object i, and the two selected leptons defining the Z-boson candidate are labelled by ℓ_1 and ℓ_2, respectively. A shape comparison of the $p_{\text{T}}^{\text{balance}}$ distribution for signal and background is shown in figure 3.7(b). Apart from its discriminative value, the $p_{\text{T}}^{\text{balance}}$ requirement also removes events with poorly measured jets and reduces the impact of those events which contain jets that do not originate from the hard scattering interaction.

The signal-suppressed *control* region is defined using similar selection criteria as the *search* region, with two modifications. The veto on interval jets is reversed such that at

least one jet with $p_{\mathrm{T}} > 25$ GeV is present in the rapidity interval between the two leading jets. Consequently, the definition of the transverse-momentum balancing variable is modified to incorporate the highest transverse momentum interval jet. Analogously, the variable $p_{\mathrm{T}}^{\mathrm{balance},3}$ is defined as

$$
p_{\mathrm{T}}^{\mathrm{balance},3} = \frac{\left| \vec{p}_{\mathrm{T}}^{\,\ell_1} + \vec{p}_{\mathrm{T}}^{\,\ell_2} + \vec{p}_{\mathrm{T}}^{\,j_1} + \vec{p}_{\mathrm{T}}^{\,j_2} + \vec{p}_{\mathrm{T}}^{\,j_3} \right|}{\left| \vec{p}_{\mathrm{T}}^{\,\ell_1} \right| + \left| \vec{p}_{\mathrm{T}}^{\,\ell_2} \right| + \left| \vec{p}_{\mathrm{T}}^{\,j_1} \right| + \left| \vec{p}_{\mathrm{T}}^{\,j_2} \right| + \left| \vec{p}_{\mathrm{T}}^{\,j_3} \right|},
\tag{3.2}
$$

where the label j_3 refers to the leading interval jet. The $p_{\mathrm{T}}^{\mathrm{balance},3}$ requirement similarly removes events with poorly measured jets and reduces the impact of those events which contain jets that do not originate from the hard scattering interaction.

It is worth noting that the *search* region, which shall be used to extract the electroweak Zjj contribution, cannot be optimised with respect to e.g. VBF Z production alone (cf. figure 3.1(a)). This has been confirmed – with the input and advice from the Sherpa authors – using a private version of Sherpa in which the source code has been altered to set various Z-boson couplings to zero, thus effectively removing specific Feynman diagrams from the calculation. The VBF contribution to the electroweak cross section then comes out to about $400\,\%$ (!) in the *baseline* region – the contribution from Z-bremsstrahlung diagrams (cf. figures 3.1(c) and 3.1(d)) is of similar size. This demonstrates the large destructive interference between the two types of diagrams. It is somewhat interesting to note though that the VBF contribution to the electroweak cross section rises to $600\,\%$ in the *high-mass* region, implying that a 'VBF-only cross section' would be a lot harder if there was such a thing. However, at the end of the day, the only viable conclusion to be drawn from this exercise is that any attempt to optimise a region of phase space based on individual diagrams is not only pointless, but also extremely foolish, since diagram-specific calculations are not gauge invariant. The phase-space requirements presented in this section have instead been chosen, so as to optimise the expected significance using the full gauge invariant set of electroweak Zjj diagrams (cf. figure 3.1).

3.3.2   Event reconstruction at the detector level

The measurement is performed using proton-proton collision data corresponding to an integrated luminosity of 20.3 fb^{-1} recorded at a centre-of-mass energy of $\sqrt{s} = 8$ TeV between April and December 2012. Events containing a Z-boson candidate in the muon channel are retained for further analysis based on single-muon triggers with minimum transverse momentum requirements of 24 GeV or 36 GeV (isolation criteria are applied for the lower p_{T} threshold). Events containing a Z-boson candidate in the electron channel are retained for further analysis using a dielectron trigger, requiring both electrons to have $p_{\mathrm{T}} > 12$ GeV.

Events are analysed if they have been recorded in a data-taking period in which the detector has been fully operational and, furthermore, if they contain at least one reconstructed collision vertex, defined by three associated inner detector tracks with

$p_T > 400$ MeV. For a given event, the primary vertex is then defined as the collision vertex yielding the highest sum of squared transverse momenta of its associated inner detector tracks.

Muon candidates are reconstructed by identifying, matching and combining inner detector tracks with track segments in the muon spectrometer [22]. They are reconstructed within $|\eta| < 2.4$ and are required to have a minimum transverse momentum of 25 GeV. Track quality requirements are imposed in order to suppress backgrounds and impact parameter requirements ensure that the muon candidates originate from the primary vertex. Furthermore, the muon candidates are required to be isolated, meaning the scalar sum of the transverse momenta from the tracks within a cone of $\Delta R = 0.2$ centred around (but excluding) the muon track is required to be less than 10% of the p_T of the muon.

Electron candidates are reconstructed by identifying, matching and combining inner detector tracks with clusters of energy in the electromagnetic calorimeter. They are required to have a minimum transverse momentum of 25 GeV and are reconstructed within $|\eta| < 2.47$, but excluding $1.37 < |\eta| < 1.52$ which is the transition region between the barrel and endcap electromagnetic calorimeters. The electron candidates are required to satisfy a set of 'medium' selection criteria [21] that have been reoptimised for the higher rate of proton-proton collisions per bunch crossing (pileup) observed in the 8 TeV data. Furthermore, impact parameter requirements ensure that the electron candidates originate from the primary vertex.

Jets are reconstructed from three-dimensional topological clusters of energy in the calorimeter [77] using the anti-k_t jet clustering algorithm [40] with a jet-radius parameter of 0.4. The reconstructed jet energies are initially corrected to account for soft energy arising from pileup [78]. Furthermore, the energy and direction of each jet are then corrected for calorimeter non-compensation, detector material and the transition between calorimeter regions, using a combination of simulation-derived calibration constants and in situ data-driven calibration constants [79, 80]. Jets are required to have a minimum transverse momentum of 25 GeV and are reconstructed within the rapidity range of $|y| < 4.4$. Additional data quality requirements are imposed in order to minimise the effect of noisy calorimeter cells. Furthermore, the jet vertex fraction (JVF) is used to identify jets from the primary interaction, so as to suppress jets from overlapping proton-proton collisions (pileup). Inner detector tracks are associated with the reconstructed jets using ghost association [81], whereby the inner detector tracks are assigned negligible momentum and clustered to the jet using the anti-k_t algorithm. The JVF is then defined as the ratio of the scalar summed transverse momentum of the ghost-associated tracks originating from the primary vertex to the scalar summed transverse momentum of all the ghost-associated tracks. Each jet with $p_T < 50$ GeV within the inner detector tracking acceptance ($|y| < 2.4$) is required to have a JVF greater than 50%. Finally, jets are required to be well separated from selected leptons, meaning that the distance between a jet and any of the selected leptons ought to be $\Delta R_{j,\ell} \geq 0.3$ in η-ϕ space. Jets that fall within such a cone are removed from the analysis.

4

Inclusive Zjj cross sections

From a frequentist's perspective, the probability that a specific process occurs in a given data sample is proportional to the number of candidate events counted in the data divided by the amount of data looked at. A measure for the amount of data collected in a given period of time is the integrated luminosity L_{int} which has dimensions of inverse area. Thus the ratio of number of events to integrated luminosity has dimensions of area; it is referred to as the cross section in analogy to scattering theory. This cross section is said to be fiducial if it is only valid in a well-defined phase space, e.g. one where the detector operates at a high efficiency or where the process tends to dominate over other (background) processes. The number of background events, N_{bkg}, expected in this fiducial phase space can be estimated using Monte Carlo or the data and is usually subtracted from the number of events observed in the data, N_{obs}. Furthermore, one can apply a correction factor $\mathcal{C}$ to correct for detector inefficiencies and resolutions in order to be able to compare the measured cross section to a particle-level prediction. It follows that the fiducial cross section σ_{fid} is given by

$$\sigma_{\mathrm{fid}} = \frac{N_{\mathrm{obs}} - N_{\mathrm{bkg}}}{\mathcal{C} \times L_{\mathrm{int}}} \tag{4.1}$$

Terms appearing in the numerator of equation (4.1) are discussed in section 4.1, while the determination of the correction factor $\mathcal{C}$ as well as the systematic uncertainties arising from the terms in the denominator of equation (4.1) are discussed in section 4.2. The typical size of the experimental uncertainties is illustrated for a few detector-level distributions in section 4.3. Finally, the results of the cross-section measurement are presented in section 4.4.

4.1 Backgrounds

Table 4.1 shows the expected process composition for N_{obs} by percentage as predicted by the Monte Carlo. Processes containing a Z boson in the final state are separated from the contribution of non-Z-background events (N_{bkg}) by a horizontal line. It can be seen that the event sample is dominated by real Zjj events which is largely due to the tight cut around the Z-mass peak. The dominant background contribution is

Table 4.1

Process composition for each phase space (%) for the combined muon and electron channels. The strong Zjj, electroweak Zjj, diboson, $t\bar{t}$, $W+$ jets and single top rates are estimated by running the analysis chain over MC samples fully simulated in the ATLAS detector. The multijet background is estimated using a data-driven technique.

	Composition (%)				
process	*baseline*	*high-p_{T}*	*search*	*control*	*high-mass*
Strong Zjj	95.8	94.0	94.7	96.0	85
Electroweak Zjj	1.1	2.1	4.0	1.4	12
WZ and ZZ	1.0	1.3	0.7	1.4	1
$t\bar{t}$	1.8	2.2	0.6	1.0	2
Single top	0.1	0.1	< 0.1	< 0.1	< 0.1
Multijet	0.1	0.2	< 0.1	0.2	< 0.1
WW, $W+$ jets	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1

coming from $t\bar{t}$ events and is found to be at the level of $2\,\%$ or less, depending on the fiducial region. Contributions from single-t events and processes involving $W+$ jets are found to be negligible.

A contribution due to multijet events arises when two jets are misidentified as isolated leptons. While this background contribution is estimated using a data-driven approach analogous to the one presented in ref. [82], all the other process contributions are estimated using dedicated Monte Carlo samples which have been passed through GEANT4 for a full simulation of the ATLAS detector. In order to estimate the multijet contribution a two-component template fit to the dilepton invariant mass distribution is performed where the shape of the multijet template is obtained by reversing some of the electron identification criteria or by reversing some of the muon isolation criteria, respectively. In addition, a Monte Carlo template is formed from all the other processes for which dedicated Monte Carlo samples are available. A maximum-likelihood fit to the data is then performed as illustrated in figure 4.1 for the $Z \to e^+e^-$ channel. For each decay channel, the multijet contribution is extracted and found to be $\lesssim 0.2\,\%$ depending on the phase-space region.
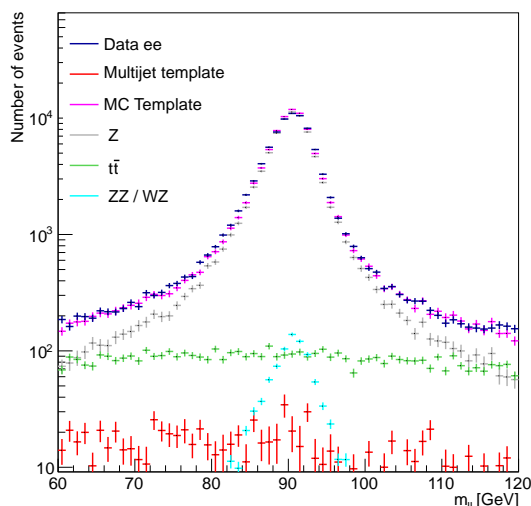
Figure 4.1

The dielectron mass spectrum in the *baseline* region after performing a maximum-likelihood fit to the data using a multijet template (found by reversing some of the electron identification criteria) and an MC template (formed from all other processes). The individual $t\bar{t}$ and diboson background contributions are also shown in order to indicate the relative size of the multijet contribution.

44

4.2   Systematic and statistical uncertainties

The number of events that pass a given phase-space selection at the particle level is expected to be different at the detector level due to various detector effects, such as reconstruction inefficiencies and resolutions. Since these effects are fully simulated in the Monte Carlo and the event record prior to the detector simulation is retained as well, one can use the simulation to correct the normalisation back to the particle level by comparing the event yields before and after the detector simulation. A correction factor, $\mathcal{C}$, is then obtained by constructing the ratio of the number of events that pass the selection at the detector level to the number of events that pass the same selection at the particle level. Many events are expected to pass the selection at both levels and so in order to take into account the statistical correlation between the numerator and the denominator, it is useful to first rewrite them in the following fashion

$$\mathcal{C} = \frac{d + b}{p + b} \tag{4.2}$$

where the event yields at detector and the particle level have been separated into those events that pass the selection at both levels (b), those that pass the selection only at the particle level (p) and those that pass the selection only at the detector level (d). The statistical uncertainty on the correction factor due to limited Monte Carlo events is then given by

$$\delta\mathcal{C}_{\text{stat}} = \frac{1}{(p+b)^2} \times \sqrt{(\delta d \times (p+b))^2 + (\delta p \times (d+b))^2 + (\delta b \times (p-d))^2}. \tag{4.3}$$

With this result in mind, standard error propagation can be used to show that the total uncertainty on the measured fiducial cross section defined in equation (4.1) is given by

$$\delta\sigma_{\text{fid}} = \sigma_{\text{fid}} \times \sqrt{\frac{(\delta N_{\text{obs}})^2 + (\delta N_{\text{bkg}})^2}{(N_{\text{obs}} - N_{\text{bkg}})^2} + \left(\frac{\delta\mathcal{C}_{\text{stat}}}{\mathcal{C}}\right)^2 + \left(\frac{\delta\mathcal{C}_{\text{syst}}}{\mathcal{C}}\right)^2 + \left(\frac{\delta L_{\text{int}}}{L_{\text{int}}}\right)^2}. \tag{4.4}$$

Systematic uncertainties on the correction factor are divided into those that are correlated between the lepton channels (such as jet-based systematics and theory modelling uncertainties) and those that are not. The uncertainties that are uncorrelated between the lepton channels arise from the lepton reconstruction, identification, isolation and trigger efficiencies as well as the lepton momentum scale and resolution and have been defined in refs. [21, 83]. The effect on the correction factor is determined by repeating the analysis for every source of systematic uncertainty and applying a systematic variation to the Monte Carlo event weights or lepton four-momenta as applicable. The variations correspond to $\pm 1\sigma$ shifts in the determined momentum scale, the resolution smearing or efficiency corrections respectively. Note that since these variations arise from the unknown 'true' energy scale of the reconstructed objects as well as resolution and efficiency mismodelling in the simulation, they are only applied at the detector level. The resulting change in $\mathcal{C}$ with respect to the nominal value is then taken as a

systematic uncertainty on the correction factor for a given source. The various sources are assumed to be uncorrelated and hence the individual contributions are added in quadrature to obtain the final uncorrelated systematic uncertainty which is found to be at the level of 3% and 2% in the electron and muon channel, respectively.

Jet-based systematic uncertainties on the correction factor are correlated between the lepton channels and thus they are estimated by first combining the two channels, then applying the systematic variations and calculating the resulting change in $\mathcal{C}$ with respect to the nominal value. Differences between the calorimeter response in simulation and data are accounted for by the jet energy scale (JES) and jet energy resolution (JER) uncertainties [79, 80]. The JES uncertainty for the 2012 data set includes components for the simulation-based and data-driven calibration constants, soft energy pileup corrections (which are applied as part of the jet calibration procedure), the calibration of forward jets as well as the unknown jet flavour composition. The latter refers to the different calorimeter response for quark- and gluon-initiated jets. Hard jets originating from pileup interactions are also reconstructed in the event at the detector level. Any mismodelling of pileup jets in the simulation is a source of systematic uncertainty. In the central calorimeter region (i.e. for $|\eta| < 2.4$), the JVF cut removes a large fraction of these, but also jets associated with the primary vertex. Hence any mismodelling of the JVF distribution in the Monte Carlo with respect to data introduces a possible bias and so the analysis is repeated with modified JVF cuts in order to cover possible differences in efficiency between the data and the simulation. In the forward region, i.e. outside the inner detector acceptance, no track-based cut can be applied to remove jets originating from pileup interactions. To estimate the impact of jets originating from pileup, the analysis is repeated using only the Monte Carlo jets that can be matched ($\Delta R < 0.3$) to a particle-level jet with $p_{\mathrm{T}} > 10$ GeV. Note that the particle level does not contain contributions from pileup interactions. The effect of pileup on the cross-section measurement is then determined by comparing the simulated detector-level event yield after applying this jet-matching procedure to the one with no matching applied. Studies of the central jet transverse momentum in a pileup-enhanced sample (JVF $< 10\%$) and the transverse energy density in the forward region of the detector [84], indicate that the simulation could be mismodelling the number of pileup jets by up to 35%. The difference between the event yields obtained with and without jet matching is therefore scaled by 0.35 and taken as a two-sided systematic uncertainty on the measured cross section.

In addition to these experimental uncertainties, systematic uncertainties on the correction factor due to possible Monte Carlo generator mismodelling are evaluated. As discussed earlier, events that pass a phase-space selection at the particle level may not necessarily also pass the selection at the detector level due to detector effects. The amount of phase-space migrations can be different if there is a discrepancy between the data and the simulation in the distributions of those variables that are used to define the region of phase space. For a given distribution the effect is estimated by reweighting the simulation to match the data and the difference of the resulting correction factor with respect to the nominal correction factor is taken as a systematic uncertainty.

(a) close-up of low-p_T region
(prior to reweighting)

(b) prior to reweighting

(c) reweighted distribution
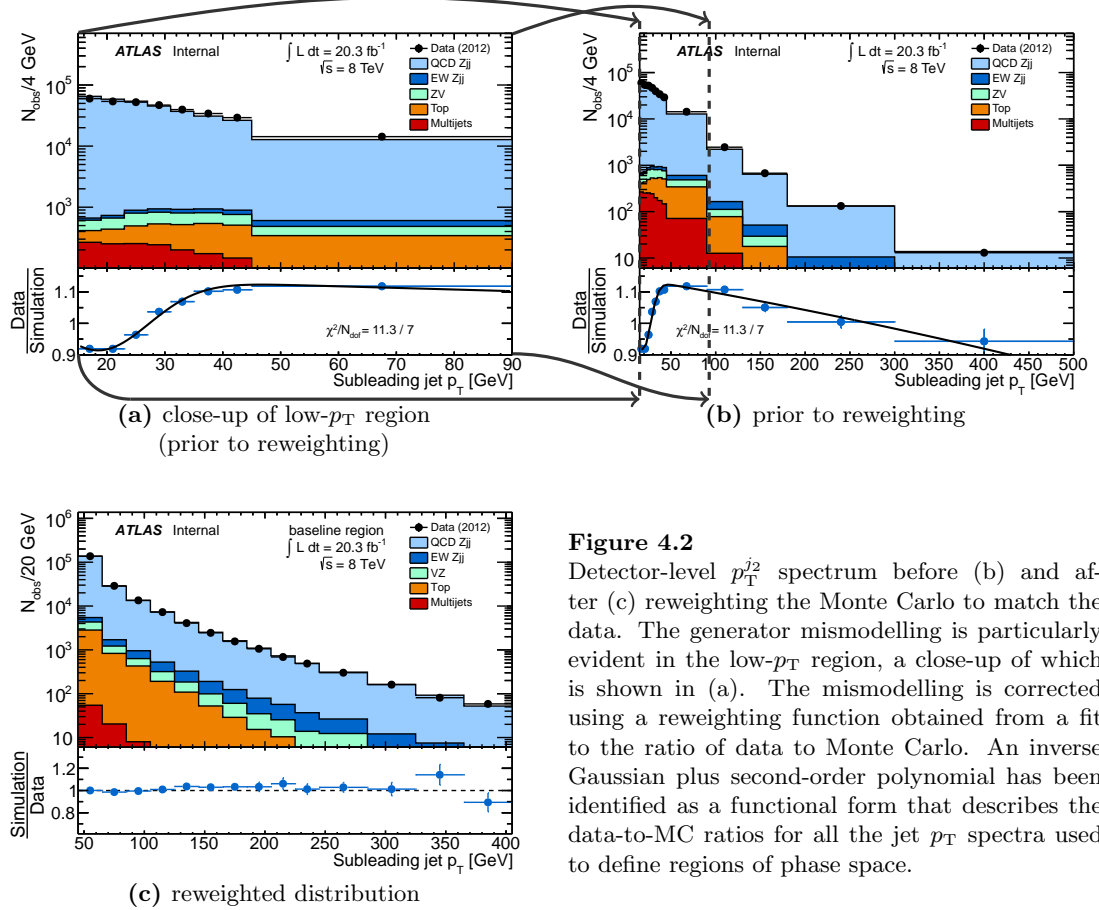
Figure 4.2
Detector-level p_T^{j2} spectrum before (b) and after (c) reweighting the Monte Carlo to match the data. The generator mismodelling is particularly evident in the low-p_T region, a close-up of which is shown in (a). The mismodelling is corrected using a reweighting function obtained from a fit to the ratio of data to Monte Carlo. An inverse Gaussian plus second-order polynomial has been identified as a functional form that describes the data-to-MC ratios for all the jet p_T spectra used to define regions of phase space.

For example, the steps necessary to assess the mismodelling of the subleading jet p_T spectrum are illustrated in figure 4.2. The distribution of p_T^{j2} is shown in figure 4.2(b) all the way down to the jet calibration limit of 15 GeV and after requiring a well-defined leading jet in the event. The ratio of data to simulation is also shown. For convenience, a close-up of the lower end of this spectrum is provided in figure 4.2(a). It can be seen that the simulation overpredicts the data around and below 25 GeV, whereas it generally underpredicts the data above that. This feature is possibly due to a transition between different p_T spectra used to model jets originating from the hard scatter and those which come from e.g. pileup or the underlying event. A reweighting function for this distribution is then obtained by fitting the ratio of data to simulation with an inverse Gaussian (that is to say a Gaussian with negative amplitude) plus a first-order polynomial over the range 15–700 GeV, giving a χ^2 of 11.3 with 7 degrees of freedom.

For a given event, the weight is determined using the particle-level information and applied as an additional event weight, so as to change both particle- and detector-level distributions. The correction factor is then recalculated using the reweighted events and the difference with respect to the nominal correction factor is taken as a systematic uncertainty. Note that reweighting the particle-level p_T^{j2} spectrum according to this fit will decrease the number of events containing jets with low transverse momentum, whilst increasing the proportion of events with higher transverse momentum. As a result, the contribution of those events that pass the detector-level selection but fail to pass the selection at the particle level will be significantly decreased. Conversely, the

	$\delta\mathcal{C}/\mathcal{C}$ (%)				
component	*baseline*	*high-p_{T}*	*search*	*control*	*high-mass*
Jet energy resolution	± 1.7	± 0.1	± 0.5	± 0.6	± 5.1
Jet energy scale	± 9.9	± 9.1	± 7.6	± 12.9	± 16.7
Jet vertex fraction[a]	± 0.2	—	± 1.0	± 2.2	—
pileup jet modelling	± 2.4	± 0.8	$< \pm 0.1$	± 1.3	± 2.3
generator modelling	± 0.8	± 0.3	± 0.5	± 0.7	± 3.6

[a]Jets used to define the *high-p_{T}* and *high-mass* regions tend to have large transverse momenta and so the jet-vertex fraction (which is only used for jets with $p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.4$) has no effect in these regions of phase space.

reweighting leads to an increased contribution of those events that pass the particle-level selection but fail to pass the selection at the detector level due to smearing effects around the 25 GeV cut-off used to reconstruct a jet in the event[1]. Figure 4.2(c) shows the reweighted p_{T}^{j2} spectrum in the *baseline* region. Good agreement is found between data and simulation near the cut value used to define the phase space which is essential in order to be able to capture the effect of phase-space migrations adequately.

The procedure is repeated to extract systematic uncertainties due to generator mismodelling for all kinematic variables used to reconstruct an object or to define a fiducial region. The final systematic uncertainty due to generator mismodelling is obtained by adding the individual uncertainties in quadrature.

Table 4.2 shows a breakdown of the various sources of systematic uncertainties which are correlated between the two decay channels for each region of phase space. The dominant uncertainty component is due to the jet energy scale, ranging from 7.6 % in the *search* region up to 16.7 % in the *high-mass* region. The uncertainty due to the jet energy resolution on the other hand is generally much smaller, ranging from 0.1 % in the *high-p_{T}* region up to 5.1 % in the *high-mass* region. The uncertainty due to the generator mismodelling is typically at the sub-percent level, except in the *high-mass* region where it is 3.6 %. However, even in this extreme region of phase space the generator modelling uncertainty is negligible compared to the JES uncertainty. The uncertainty on the integrated luminosity is 2.8 %, estimated using the methodology detailed in ref. [85] for beam-separation scans performed in November 2012.

4.3 Detector-level comparisons of data and simulation

Detector-level comparisons between data and simulation are shown in figure 4.3 in order to illustrate the size of the experimental uncertainties discussed in section 4.2 as well as to examine the contribution of the non-Z-background events discussed in section 4.1.

---

[1]Since the p_{T} cuts on the two leading jets are 55 and 45 GeV in the *baseline* region respectively, which is much higher than 25 GeV, the effect on the correction factor is expected to be small though.
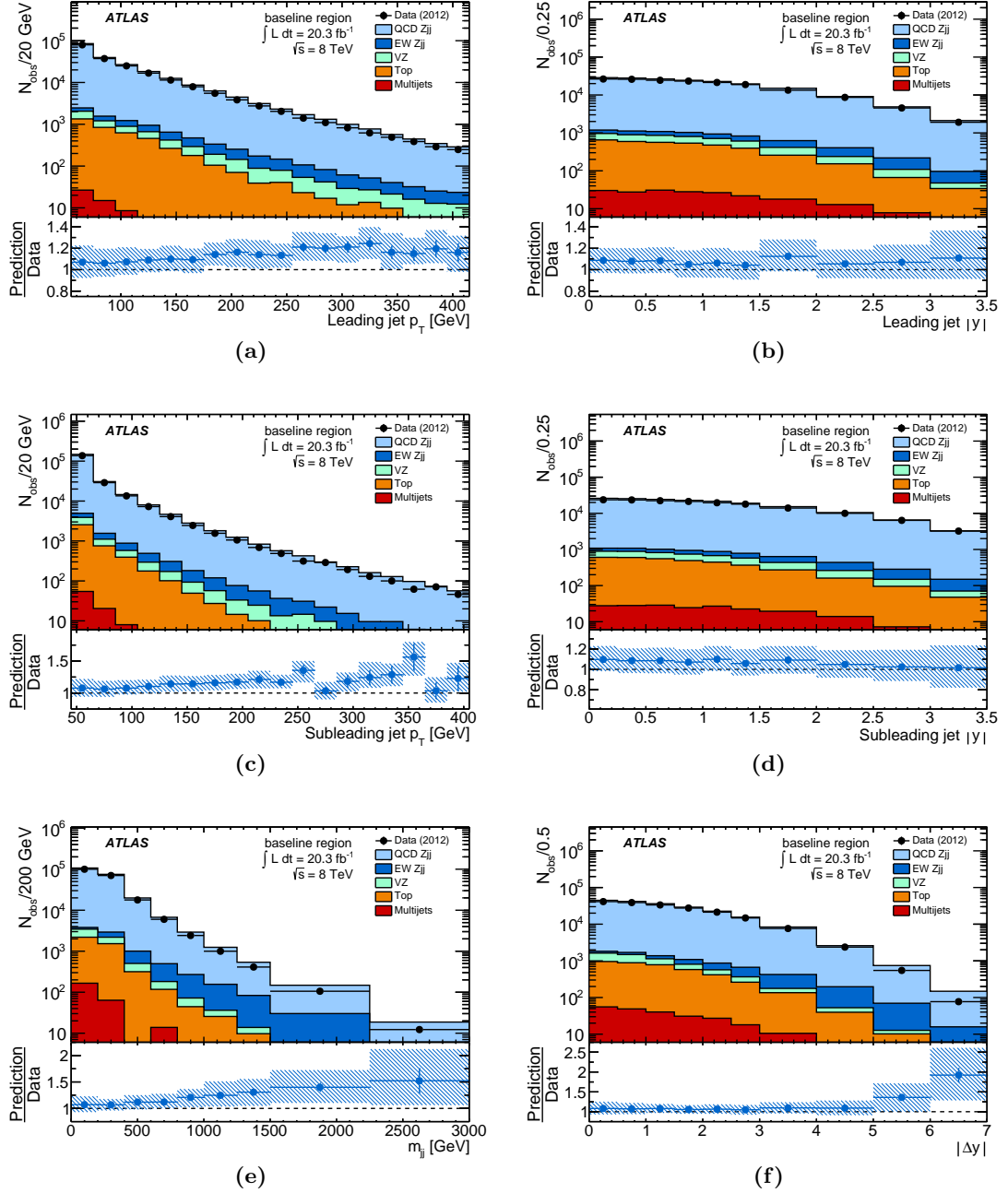
Figure 4.3

Detector-level comparisons of data and simulation in the *baseline* region for the leading jet (a) transverse momentum and (b) rapidity, the subleading jet (c) transverse momentum and (d) rapidity as well as the (e) invariant mass and (f) rapidity span of the dijet system. The simulated samples are normalised to the expected cross-section predictions discussed in section 3.2 and then stacked. The hatched band in the ratio reflects the total experimental systematic uncertainty on the simulation, while the error bars indicate the statistical uncertainties of the data.

The Monte Carlo simulation is compared to the data in the *baseline* region as a function of the leading jet transverse momentum and rapidity, the subleading jet transverse momentum and rapidity as well as the invariant mass and rapidity separation of the two leading jets. The uncertainty on the simulation due to experimental systematic uncertainties is indicated in the ratio as a hatched (blue) band. Specifically, this includes the uncertainties due to the JES, JER, JVF as well as the pileup jet modelling. The simulation generally gives an adequate description of the data, although there are

Table 4.3

Fiducial cross sections for inclusive Zjj production, measured in the $Z \to \ell^+\ell^-$ decay channel.

Fiducial region	σ_{fid} (pb)
baseline	5.88 ± 0.01 (stat) ± 0.62 (syst) ± 0.17 (lumi)
high-p_{T}	1.82 ± 0.01 (stat) ± 0.17 (syst) ± 0.05 (lumi)
search	1.10 ± 0.01 (stat) ± 0.09 (syst) ± 0.03 (lumi)
control	0.447 ± 0.004 (stat) ± 0.059 (syst) ± 0.013 (lumi)
high-mass	0.066 ± 0.001 (stat) ± 0.012 (syst) ± 0.002 (lumi)

indications of generator mismodelling at high jet transverse momentum and large dijet invariant mass. Note that the contribution from $t\bar{t}$ interactions (the dominant background process) and multijet events remains small in each bin of the distributions.

4.4 Results

The cross sections are measured using the formula given in equation (4.1) for the muon and electron channel separately. In each fiducial region the cross sections are found to be compatible between the two decay channels with a maximum difference of 1.1σ considering only sources of uncertainty that are uncorrelated between the channels. The results of the individual decay channels are then combined using a weighted average, with the weight set to the squared inverse of the corresponding uncorrelated uncertainty component.

The measured cross sections for inclusive Zjj production are presented in table 4.3 together with their statistical and systematic uncertainties. The corresponding theoretical predictions are estimated using `Powheg` for both strong and electroweak Zjj production as well as a small contribution of ZV events obtained from `Sherpa`. Table 4.4 shows the Standard Model predictions together with their statistical and theoretical uncertainties. The latter is broken down into scale and generator-modelling uncertainty components, both estimated from the envelope of the corresponding `Powheg` sample variations discussed in section 3.2, as well as a PDF uncertainty component derived from the `CT10` eigenvectors using the procedure described in ref. [86].

Table 4.4

Theoretical predictions for inclusive Zjj production in the $Z \to \ell^+\ell^-$ decay channel. Both strong and electroweak Zjj events are produced using `Powheg`. A small contribution of ZV events from `Sherpa` is also included. The PDF uncertainty is estimated from the `CT10` eigenvectors using the procedure described in ref. [86]. The scale and modelling uncertainties are each estimated from the envelope of the `Powheg` sample variations discussed in section 3.2.

Fiducial region	σ_{theory} (pb)
baseline	6.26 ± 0.06 (stat) $^{+0.50}_{-0.60}$ (scale) $^{+0.29}_{-0.35}$ (PDF) $^{+0.19}_{-0.25}$ (model)
high-p_{T}	1.92 ± 0.02 (stat) $^{+0.17}_{-0.20}$ (scale) $^{+0.09}_{-0.10}$ (PDF) $^{+0.05}_{-0.07}$ (model)
search	1.23 ± 0.01 (stat) $^{+0.11}_{-0.13}$ (scale) $^{+0.06}_{-0.07}$ (PDF) $^{+0.03}_{-0.04}$ (model)
control	0.444 ± 0.005 (stat) $^{+0.051}_{-0.054}$ (scale) $^{+0.021}_{-0.025}$ (PDF) $^{+0.032}_{-0.034}$ (model)
high-mass	0.068 ± 0.001 (stat) $^{+0.009}_{-0.009}$ (scale) $^{+0.004}_{-0.003}$ (PDF) $^{+0.004}_{-0.002}$ (model)
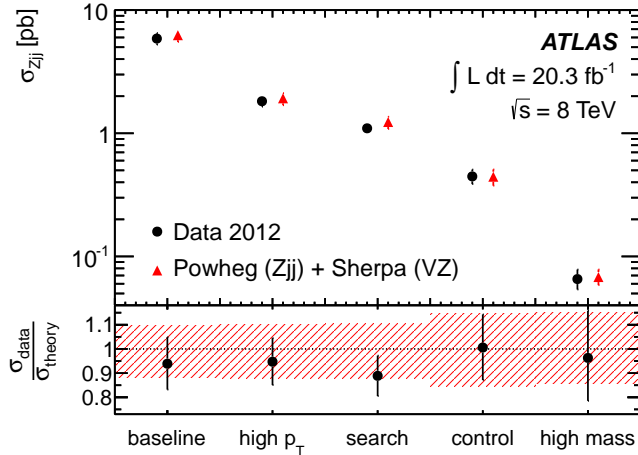
Figure 4.4
Fiducial cross-section measurements for inclusive Zjj production in the $Z \to \ell^+\ell^-$ decay channel, compared to the `Powheg` prediction for strong and electroweak Zjj production and the small contribution from ZV production predicted by `Sherpa`. The (black) circles represent the data and the associated error bar is the total uncertainty in the measurement. The (red) triangles represent the theoretical prediction, the associated error bar (or hatched band in the lower panel) is the total theoretical uncertainty on the prediction.

For easier comparison a summary is provided in figure 4.4. Good agreement between measurement and theory is observed across all fiducial regions.

4.5 Cross check: pileup dependence

The selection efficiency ϵ for a given phase space can be defined as the number of events that pass any additional jet-based selection with respect to the complete sample of events containing a Z-boson candidate. A jet originating from the hard scatter of a different proton-proton collision in the same bunch crossing (a so-called pileup jet) may be incorrectly associated with the primary vertex and hence influence the event selection since all the fiducial regions in this measurement are defined in terms of the jet multiplicity. The selection efficiency therefore depends on the amount of pileup in the event. Note though that the cross-section measurement itself does not depend on the amount of pileup in the event as long as the pileup dependence in the data is correctly described by the simulation. Although in general the selection efficiencies may very well be different for data and simulation, the ratio of the two should be the same as a function of the pileup activity in the event. A measure for the amount of pileup in the event is given by the number of reconstructed primary vertices N_{PV} or, alternatively, by the average number of proton-proton collisions per bunch crossing $\langle\mu\rangle$. In order to assess whether the cross-section measurement exhibits a pileup dependence, the analysis is repeated in three bins of N_{PV} as well as three bins of $\langle\mu\rangle$. In each case, the bin widths have been chosen such that all bins contain approximately the same amount of statistics, giving rise to a low ($N_{\mathrm{PV}} \leq 8$), medium ($9 \leq N_{\mathrm{PV}} \leq 13$) and high ($N_{\mathrm{PV}} \geq 14$) region for N_{PV} as well as a low ($\langle\mu\rangle \leq 15$), medium ($16 \leq \langle\mu\rangle \leq 20$) and high ($\langle\mu\rangle \geq 21$) region for $\langle\mu\rangle$. Figure 4.5 shows the ratio of selection efficiencies in Monte Carlo and data for the various fiducial regions for the three N_{PV} bins. There is a noticeable trend in the *high-p_{T}* region across the N_{PV} range which is not covered by the statistical uncertainty alone, as indicated by the black error bars in figure 4.5(b). The inconsistency between these three bins in N_{PV} is at the 1.8–4.1σ level. This N_{PV} trend

(a) *baseline* region

(b) *high-p_T* region

(c) *search* region

(d) *control* region

(e) *high-mass* region
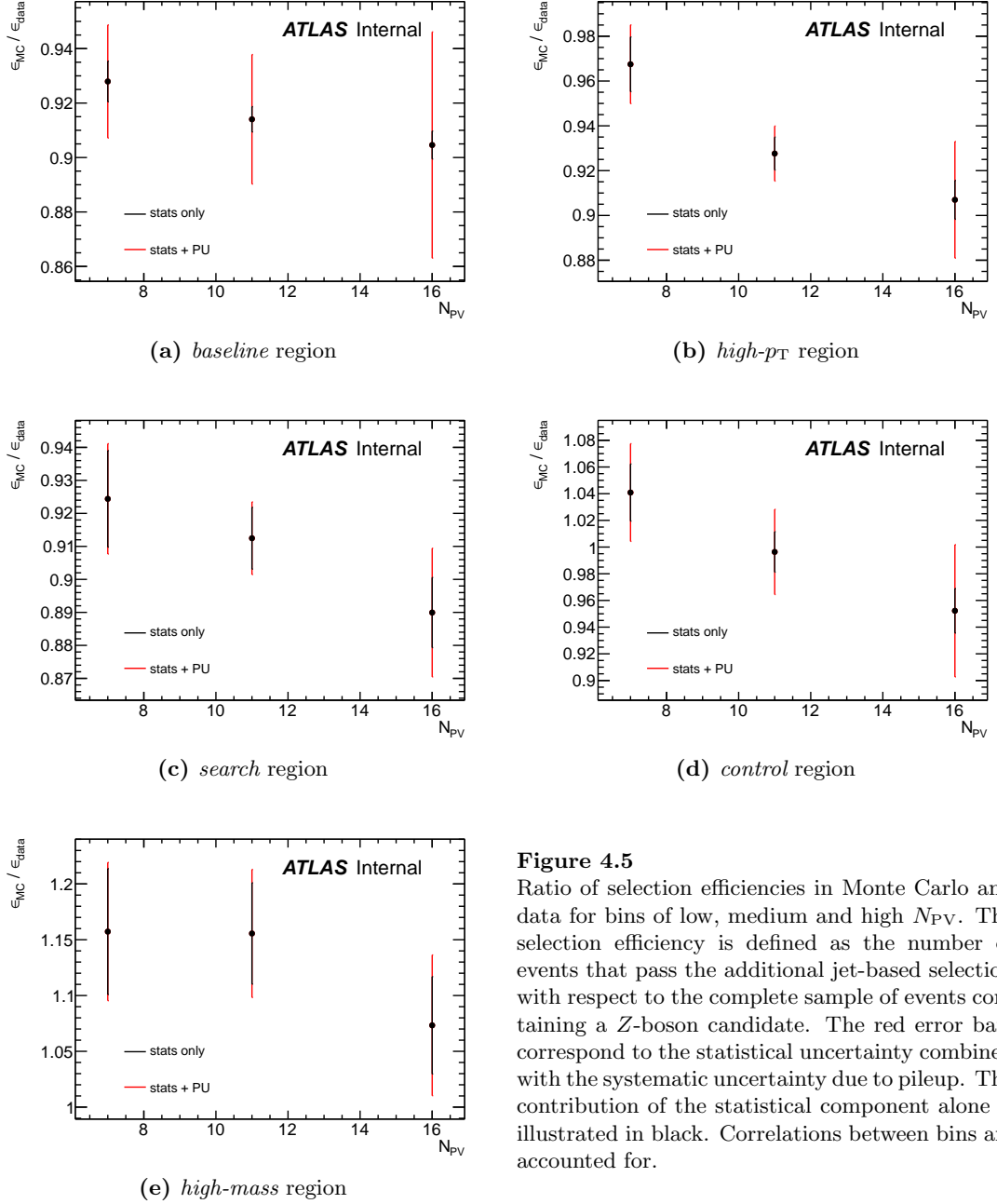
Figure 4.5
Ratio of selection efficiencies in Monte Carlo and data for bins of low, medium and high N_{PV}. The selection efficiency is defined as the number of events that pass the additional jet-based selection with respect to the complete sample of events containing a Z-boson candidate. The red error bars correspond to the statistical uncertainty combined with the systematic uncertainty due to pileup. The contribution of the statistical component alone is illustrated in black. Correlations between bins are accounted for.

is less pronounced in the other fiducial regions[2]. However, adding in the systematic uncertainty due to pileup ('PU') and accounting for correlations between N_{PV} bins, one finds that the ratios are consistent at the 1.1–1.6σ level in the *high-p_T* region (and generally at the level of 1σ or better in all the other fiducial regions). It is therefore concluded that the uncertainties cover any discrepancies due to pileup.

Figure 4.6 shows similar plots for the three $\langle \mu \rangle$ bins. In this case the various ratios have been found to be consistent at the level of 1σ or better across all fiducial regions.

---

[2]Recall that the JVF is not used to remove pileup jets in the *high-p_T* region where the two leading jets are required to have a $p_T > 85$ GeV and 75 GeV respectively (the JVF cut is only applied for jets with $p_T < 50$ GeV).
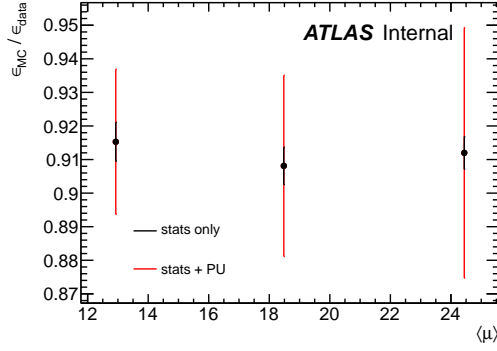
(a) *baseline* region
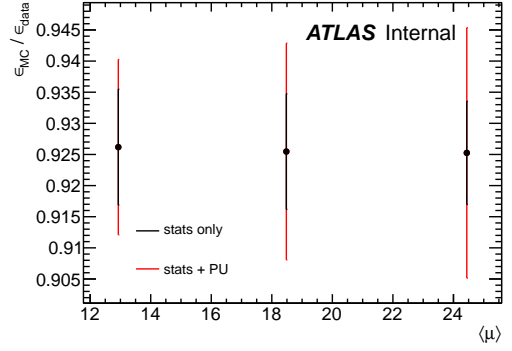
(b) *high-p_T* region
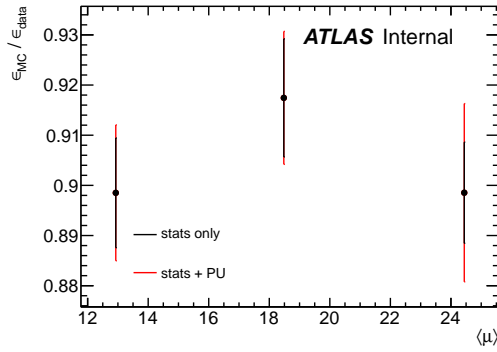
(c) *search* region
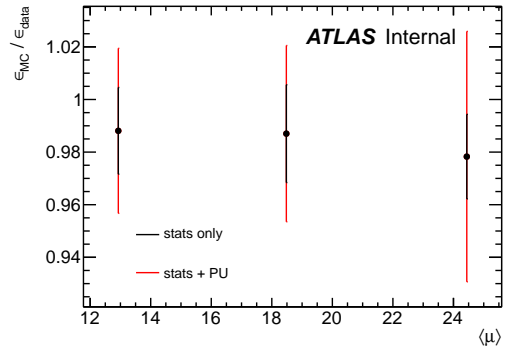
(d) *control* region
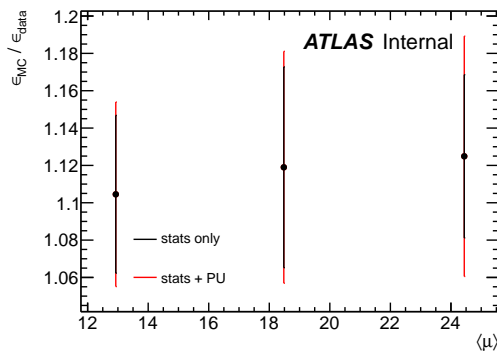
(e) *high-mass* region

Figure 4.6
Ratio of selection efficiencies in Monte Carlo and data for bins of low, medium and high $\langle\mu\rangle$. The selection efficiency is defined as the number of events that pass the additional jet-based selection with respect to the complete sample of events containing a Z-boson candidate. The red error bars correspond to the statistical uncertainty combined with the systematic uncertainty due to pileup. The contribution of the statistical component alone is illustrated in black. Correlations between bins are accounted for.

5

Differential distributions

In addition to the inclusive cross-section measurement presented in section 4, a number of inclusive Zjj differential distributions have been measured in order to confront the theoretical modelling of strong Zjj production in regions with varying sensitivity to the electroweak Zjj component. The data have been corrected for detector effects and are compared to particle-level predictions from both `Sherpa` and `Powheg`.

The complete set of unfolded distributions is provided in appendix A and the unfolded data are also available in HEPDATA [87] along with estimates of the statistical correlation between bins of different unfolded distributions to allow the quantitative comparison of all distributions simultaneously. Here, only a handful of the distributions shall be shown in order to discuss qualitative differences between the particle-level predictions from `Sherpa` and `Powheg`. Both of these are shown for combined electroweak and strong Zjj production as well as for strong Zjj production only. The contribution from diboson events has been neglected for the theoretical prediction as their impact has been found to be negligible.

The distributions sensitive to the kinematics of the two leading jets are

- ⚛ $\frac{1}{\sigma} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}}$, the normalised distribution of the dijet invariant mass, m_{jj},

- ⚛ $\frac{1}{\sigma} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}|\Delta y|}$, the normalised distribution of the rapidity difference between the two leading jets, $|\Delta y|$, and

- ⚛ $\frac{1}{\sigma} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}|\Delta\phi(j,j)|}$, the normalised distribution of the difference in azimuthal angle between the two leading jets, $|\Delta\phi(j,j)|$. Although this distribution is not very sensitive to differences between strong and electroweak Zjj production, it is of interest in Hjj studies, as the azimuthal structure of the vector boson fusion and gluon fusion production channels have been shown to be quite different [88–90].

In the electroweak process, the two leading jets tend to be produced in the forward region with large transverse momentum as they recoil against the heavy spin-1 gauge bosons involved in the t-channel exchange. In the strong process, the t-channel typically involves a spin-$\frac{1}{2}$ quark exchange and thus the invariant mass spectrum of the dijet system tends to be a lot steeper for strong Zjj production than it is for electroweak Zjj production. Figure 5.1 shows the unfolded m_{jj} distribution for the *baseline* and

(a) *baseline* region   **(b)** *search* region
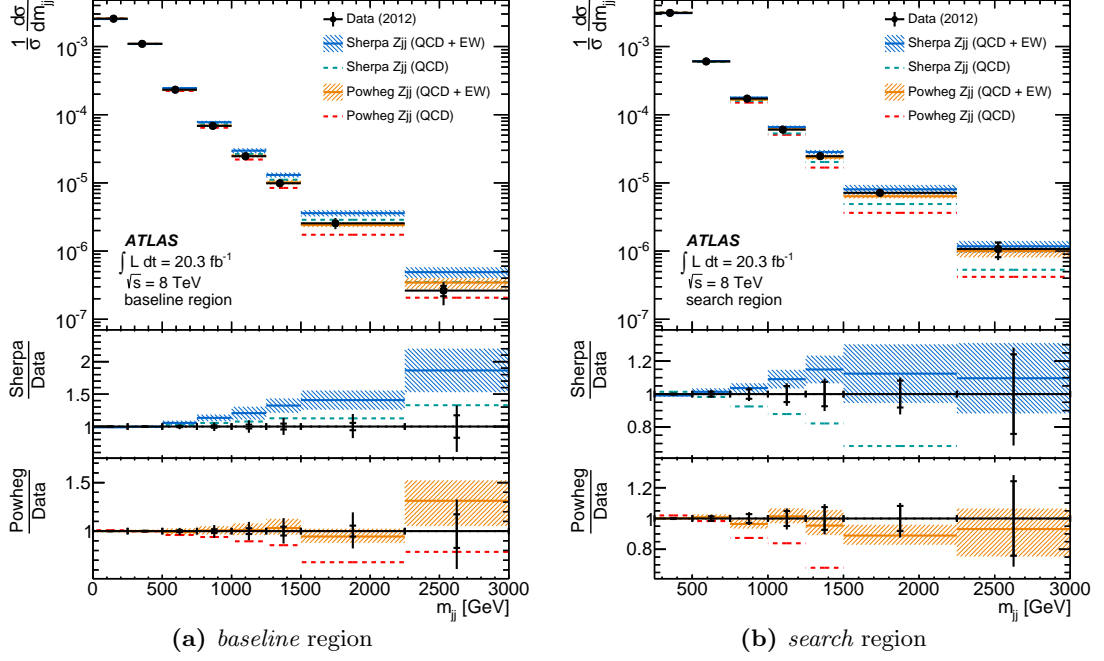
Figure 5.1
Unfolded $\frac{1}{\sigma} \cdot \frac{d\sigma}{dm_{jj}}$ distribution in the (a) *baseline* and (b) *search* region. The data are shown as filled (black) circles. The vertical error bars show the size of the total uncertainty on the measurement, with tick marks used to reflect the size of the statistical uncertainty only. Particle-level predictions from `Sherpa` and `Powheg` are shown for combined strong and electroweak Zjj production (labelled as QCD + EW) by hatched bands, denoting the model uncertainty, around the central prediction, which is shown as a solid line. The predictions from `Sherpa` and `Powheg` for strong Zjj production (labelled QCD) are shown as dashed lines.

the *search* region. Indeed, the strong-only prediction fails to capture the data at large m_{jj} for both predictions, particularly so in the signal-enhanced *search* region where the electroweak component is required by both predictions in order to describe the data. In the *baseline* region, `Sherpa` already overpredicts the data at large m_{jj} even if the electroweak component is not included, whereas `Powheg` needs the electroweak contribution in order to achieve good agreement with the data. However, note that in this region `Powheg` is accurate to next-to-leading order in perturbative QCD, whereas `Sherpa` is only accurate to leading order. `Sherpa` has been known to predict too many events at large m_{jj} and $|\Delta y|$ from previous measurements at the LHC and Tevatron [82, 91]. In the *search* region, both generators give an adequate description of the data (provided the electroweak component is included), even though the veto on additional jet activity in the *search* region means that both generators are formally accurate to leading order only.

The unfolded distributions that are sensitive to the difference in the t-channel colour flow between electroweak and strong Zjj production are

- ⚛ $\frac{1}{\sigma} \cdot \frac{d\sigma}{dN_{jet}^{gap}}$, the normalised distribution of the number of jets with $p_T > 25$ GeV in the rapidity interval bounded by the two leading jets, N_{jet}^{gap},

- ⚛ $\frac{1}{\sigma} \cdot \frac{d\sigma}{dp_T^{balance}}$, the normalised distribution of the p_T-balancing variable $p_T^{balance}$, which is defined in equation (3.1) in section 3.3.1,

- the average number of jets with $p_T > 25$ GeV in the rapidity interval bounded by the two leading jets, $\langle N_{\text{jet}}^{\text{gap}} \rangle$, as a function of m_{jj} and $|\Delta y|$,

- the *jet veto efficiency*, i.e. the fraction of events that contain no additional jets with $p_T > 25$ GeV in the rapidity interval bounded by the two leading jets, as a function of m_{jj} and $|\Delta y|$

- as well as the p_T^{balance} *cut efficiency*, i.e. the fraction of events with $p_T^{\text{balance}} < 0.15$, as a function of m_{jj} and $|\Delta y|$.

In the electroweak process, the t-channel exchange of colourless gauge bosons means there is no colour flow between the two leading jets and thus additional jet activity due to quark and gluon radiation is very unlikely. Figure 5.2(a) shows the $N_{\text{jet}}^{\text{gap}}$ distribution in the *high-mass* region. The biggest differences between the combined and the strong-only prediction can be seen in the low-multiplicity region as expected which also demonstrates the effectiveness of the jet veto as a discriminative variable.

The distributions in the remaining subfigures probe the theoretical modelling of additional quark and gluon radiation as a function of the energy scale of the dijet system. The $\langle N_{\text{jet}}^{\text{gap}} \rangle$ distribution is shown in figure 5.2(b) as a function of m_{jj}. As expected, the average number of interval jets is predicted to be larger in the absence of the electroweak component. The biggest differences between the combined and the strong-only prediction can again be seen at large values of m_{jj}. Finally, figures 5.2(c) and 5.2(d) show the *jet veto efficiency* as function of m_{jj} and $|\Delta y|$ respectively. These correspond to the fraction of events with no additional interval jets. In the absence of the electroweak component, the *jet veto efficiency* is expected to be lower which again can be seen clearly at large values of m_{jj} and $|\Delta y|$. In general, `Sherpa` and `Powheg` give an adequate description of the data for all of these distributions, although `Sherpa` does a slightly better job than `Powheg`. Both `Sherpa` and `Powheg` have previously given a good description of interval-jet activity in pure dijet topologies [92, 93].

Since neither generator is able to fully reproduce the data for all the differential distributions in the various fiducial regions, the unfolded data can be used to constrain the modelling of Zjj production in the extreme regions of phase space probed in this measurement.

5.1 Unfolding method

Data distributions for combined strong and electroweak Zjj production are obtained by subtracting the small background contribution from $t\bar{t}$ and multijet events from the data in each bin, before normalising the differential distributions to unity. These detector-level distributions are then corrected back to the particle level by applying an iterative Bayesian unfolding procedure [94, 95]. In probability theory, the Bayesian interpretation of probability is a measure for the degree of belief. Bayes' theorem then describes how the initial degree of belief, prior to making a measurement, is changing in light of the collected data. Here, the normalised particle-level predictions from `Sherpa` for combined strong and electroweak Zjj production serve as the initial degree of belief
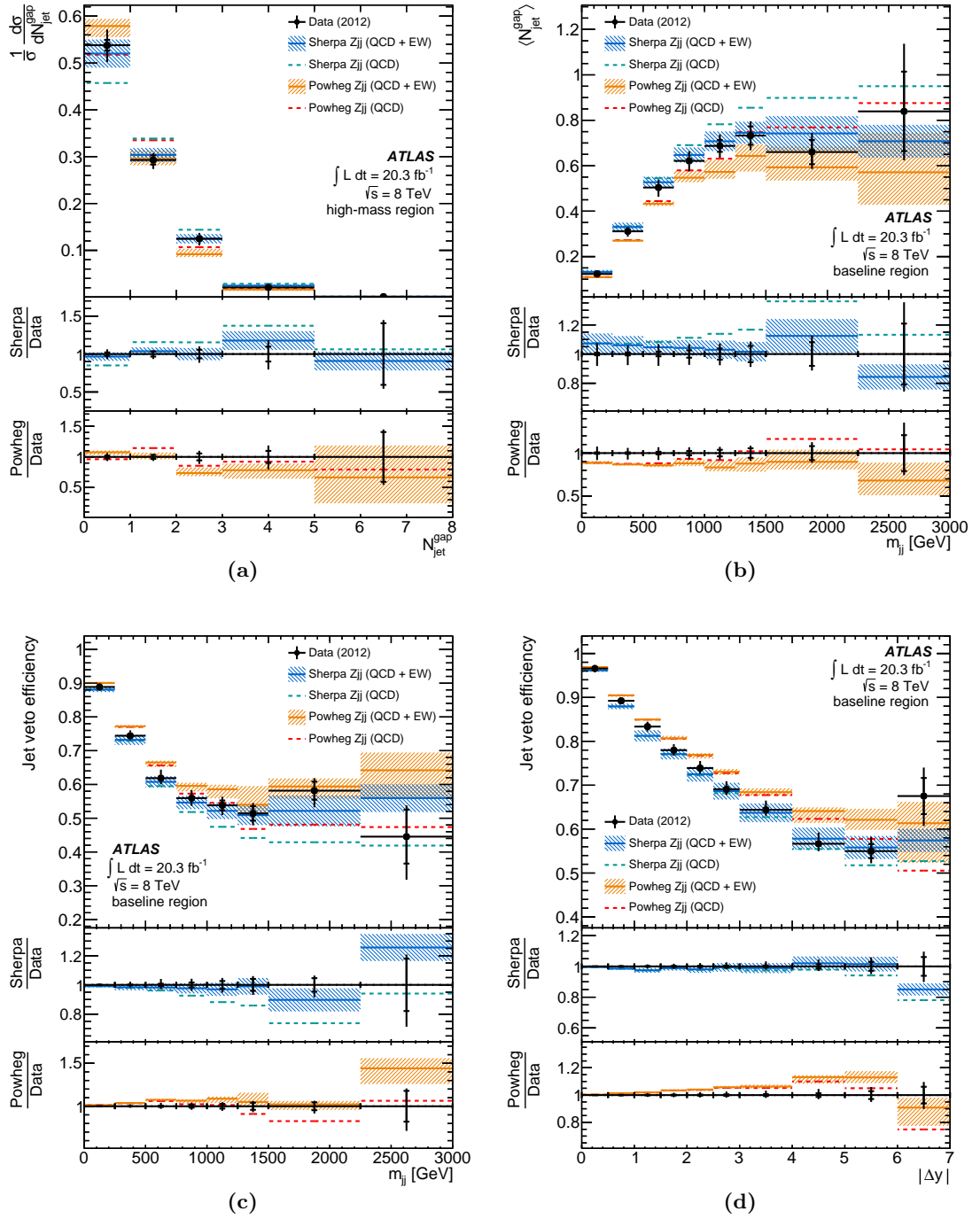
Figure 5.2
Unfolded (a) $\frac{1}{\sigma} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}N_{\mathrm{jet}}^{\mathrm{gap}}}$ distribution in the *high-mass* region and unfolded (b) $\langle N_{\mathrm{jet}}^{\mathrm{gap}} \rangle$ distribution as a function of m_{jj} in the *baseline* region, as well as the unfolded *jet veto efficiency* as a function of (c) m_{jj} and (d) $|\Delta y|$ in the *baseline* region. The data are shown as filled (black) circles. The vertical error bars show the size of the total uncertainty on the measurement, with tick marks used to reflect the size of the statistical uncertainty only. Particle-level predictions from `Sherpa` and `Powheg` are shown for combined strong and electroweak Zjj production (labelled as QCD + EW) by hatched bands, denoting the model uncertainty, around the central prediction, which is shown as a solid line. The predictions from `Sherpa` and `Powheg` for strong Zjj production (labelled QCD) are shown as dashed lines.

of what the underlying particle-level distributions might be. After the first unfolding iteration, the input prior is replaced with the output distribution from the previous iteration and the procedure is repeated. In practice, it is found that two iterations are already sufficient to achieve convergence.

In the unfolding procedure, a detector response matrix is constructed from the combined simulated samples for strong and electroweak Zjj production for each differential distribution. The response matrix is then used to reverse the bin migrations caused by the finite detector resolution and to correct for phase-space migrations, i.e. events that pass the selection requirements of a given fiducial region only at the particle level but not at the detector level (or vice versa). The statistical uncertainty on the unfolded data is estimated using pseudoexperiments.

In order to correctly unfold the phase-space migrations in the efficiency distributions, the unfolding procedure is applied to a two-dimensional distribution that is constructed from the dijet observable in one dimension and the binary outcome of the efficiency criterion in the other dimension. The efficiencies are then constructed from the unfolded two-dimensional distributions. Similarly, the correlations in the $\langle N_{\mathrm{jet}}^{\mathrm{gap}} \rangle$ distributions are retained by unfolding two-dimensional distributions constructed from the corresponding $N_{\mathrm{jet}}^{\mathrm{gap}}$ distribution in one dimension as a function of the respective dijet observable along the other dimension. Moreover, statistical correlations between bins of different unfolded distributions are estimated using a bootstrap method [96].

5.2 Systematic uncertainties

The sources of experimental and theoretical uncertainties include the ones discussed for the inclusive Zjj fiducial cross-section measurement in section 4.2. However, both the uncertainty on the integrated luminosity as well as the lepton-based systematic uncertainties have negligible impact on the unfolding as they do not result in significant shape changes of the differential distributions. The remaining experimental systematic uncertainties are therefore due to JES, JER, JVF and the modelling of pileup jets.

In order to estimate to what extent the unfolding procedure depends on the theoretical modelling of the various kinematic variables used to define the fiducial regions, the corresponding kinematic distributions are again reweighted such that the simulated spectrum for a given variable matches the one observed in the data. For each unfolded distribution an additional uncertainty is assigned that estimates to what extent the unfolding procedure depends on the modelling of the distribution that is being unfolded. These additional uncertainties are obtained by reweighting the respective particle-level predictions based on the ratio of the corresponding detector-level distributions in data and simulation.

For each source of systematic uncertainty, a new response matrix is constructed using the systematically varied distribution in place of the nominal prediction. The unfolding procedure is then repeated with the new response matrix and the resulting shift in the unfolded spectrum is taken as a systematic uncertainty. The total systematic uncertainty in each bin is then obtained by adding the individual uncertainties in

quadrature. The dominant sources of systematic uncertainty are again found to be due to the JES and the JER.

Measure what is measurable,
and make measurable what is not so.

— Galileo Galilei

6

Electroweak Zjj cross sections

The electroweak component of Zjj production is extracted by a fit to the detector-level m_{jj} distribution in the *search* region. The fitting procedure is described in section 6.1 and various cross checks related to the fitting procedure are presented in section 6.2. Systematic uncertainties are discussed in section 6.3 and a significance estimate is given in section 6.4. The extracted number of events are then converted into a fiducial cross section and corrected back to the particle level, details of which along with the measurement results can be found in section 6.5. In addition, the extracted signal has been used to set limits on anomalous triple gauge couplings as outlined in section 6.6.

6.1 Signal extraction

The data in the signal-suppressed *control* region is used to directly evaluate and constrain the modelling of the background. The signal model is taken from the `Sherpa` electroweak Zjj sample and the background model is constructed from the `Sherpa` strong Zjj sample as well as the small contributions from diboson and $t\bar{t}$ samples. Contributions from $W +$ jets and multijet events are found to have negligible impact on the results as they do not cause significant shape changes in the m_{jj} spectrum in neither *search* nor *control* region. The detector-level m_{jj} spectrum in the *control* region is shown in figure 6.1 for the individual lepton channels as well as their combination and in each case, the lower panel shows the ratio of data to simulation. It can be seen that `Sherpa` predicts slightly too many events at large m_{jj}, a feature observed in previous measurements at the LHC and Tevatron [82, 91]. In order to improve the background modelling, a reweighting function is obtained by fitting the ratio of simulation to data with a second-order polynomial. This data-driven correction obtained in the *control* region is then directly applied to the background Monte Carlo model in the signal-enhanced *search* region. Hence the data in the *control* region is used to constrain the generator modelling of the background m_{jj} shape and the simulation is only used to extrapolate this constraint between the *control* and the *search* region. This data-driven technique has the advantage of minimising both the experimental and theoretical uncertainties associated with the background model.

(a) $Z \to e^+ e^-$ channel

(b) $Z \to \mu^+ \mu^-$ channel
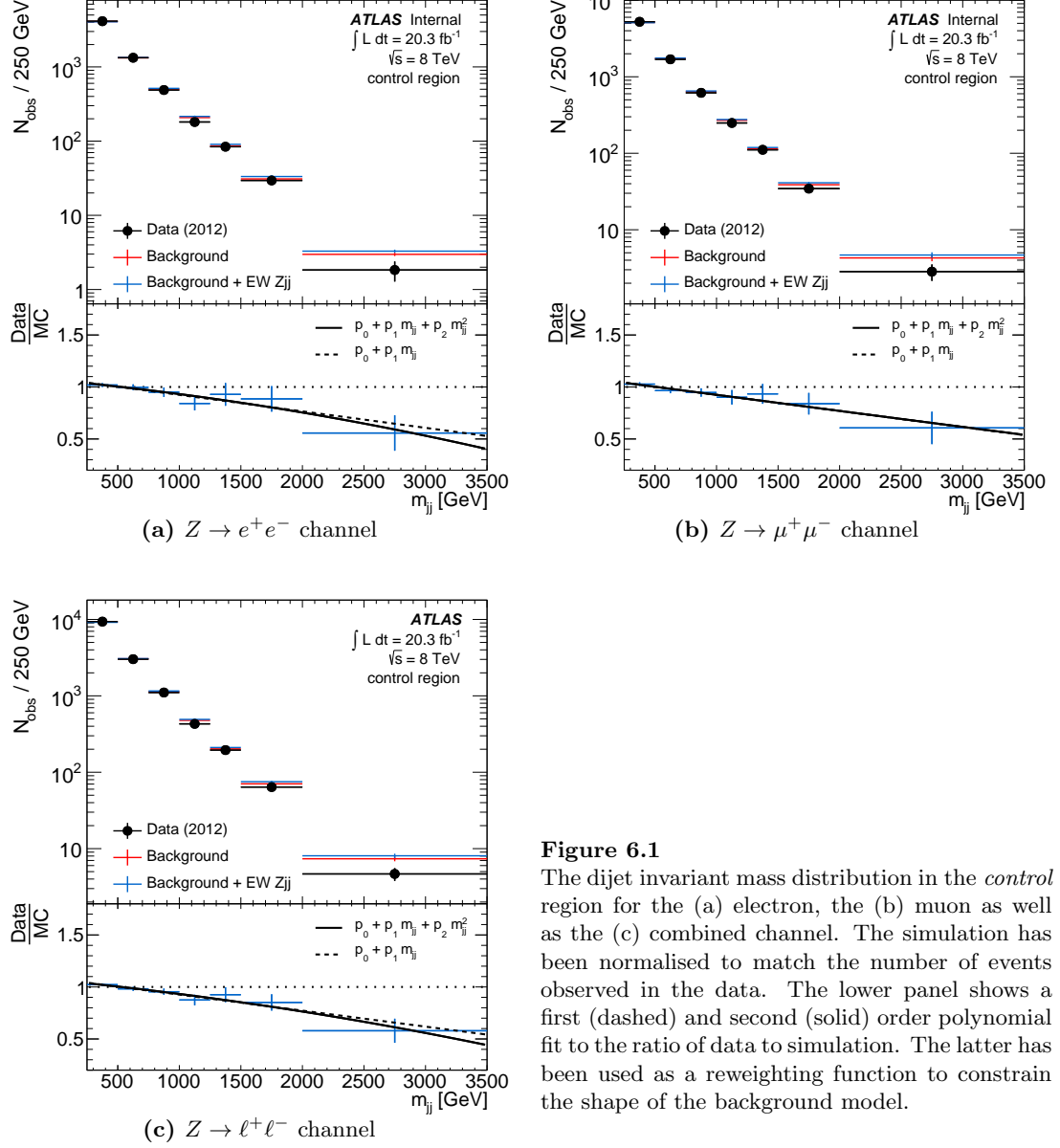
(c) $Z \to \ell^+ \ell^-$ channel

Figure 6.1

The dijet invariant mass distribution in the *control* region for the (a) electron, the (b) muon as well as the (c) combined channel. The simulation has been normalised to match the number of events observed in the data. The lower panel shows a first (dashed) and second (solid) order polynomial fit to the ratio of data to simulation. The latter has been used as a reweighting function to constrain the shape of the background model.

Table 6.1

The number of background events (N_{bkg}), including contributions from strong Zjj production, diboson and $t\bar{t}$ events, and the number of electroweak Zjj events (N_{EW}) as predicted by the MC simulation and as obtained from a fit to the data. The first and second uncertainties on the MC prediction are the experimental and theoretical systematic uncertainties, respectively. The first and second uncertainties on the fitted yields are due to statistical uncertainties in data and simulation, respectively. The sum of the fitted signal and background events is also given for convenience.

	$Z \to e^+ e^-$ channel	$Z \to \mu^+ \mu^-$ channel	$Z \to \ell^+ \ell^-$ channel
Predicted N_{bkg}	$13\,700 \pm 1\,200 \, {}^{+1\,400}_{-1\,700}$	$18\,600 \pm 1\,500 \, {}^{+1\,900}_{-2\,300}$	$32\,600 \pm 2\,600 \, {}^{+3\,400}_{-4\,000}$
Predicted N_{EW}	$602 \pm 27 \pm 18$	$731 \pm 29 \pm 22$	$1\,333 \pm 50 \pm 40$
Fitted N_{bkg}	$13\,351 \pm 144 \pm 29$	$17\,201 \pm 161 \pm 31$	$30\,530 \pm 216 \pm 40$
Fitted N_{EW}	$897 \pm 92 \pm 27$	$737 \pm 98 \pm 28$	$1\,657 \pm 134 \pm 40$
$N_{bkg} + N_{EW}$	$14\,248$	$17\,938$	$32\,187$

(a) $Z \to e^+ e^-$ channel

(b) $Z \to \mu^+ \mu^-$ channel
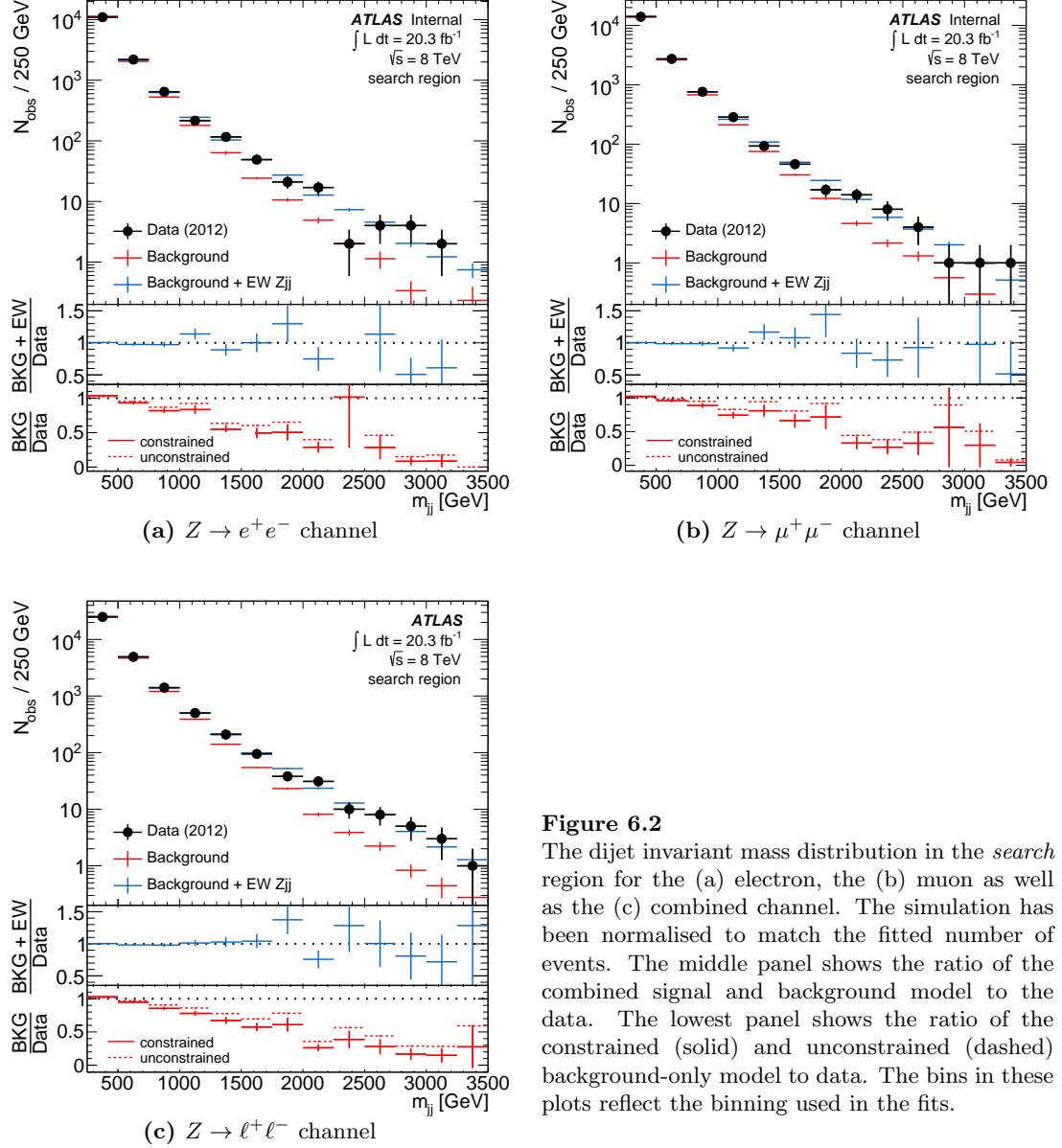
(c) $Z \to \ell^+ \ell^-$ channel

Figure 6.2
The dijet invariant mass distribution in the *search* region for the (a) electron, the (b) muon as well as the (c) combined channel. The simulation has been normalised to match the fitted number of events. The middle panel shows the ratio of the combined signal and background model to the data. The lowest panel shows the ratio of the constrained (solid) and unconstrained (dashed) background-only model to data. The bins in these plots reflect the binning used in the fits.

A fit to the data is then performed in the *search* region using the signal model and the constrained background model. The normalisation of each model is allowed to float and the signal and background yields are extracted after performing a log-likelihood maximisation [97]. The number of signal (N_{EW}) and background (N_{bkg}) events obtained from the fit are summarised in table 6.1, along with the statistical uncertainties due to the data (first component) and the MC simulation (second component). In addition, the corresponding number of signal and background events expected by the MC simulation are given with their experimental (first component) and theoretical (second component) systematic uncertainties. The sum of the fitted signal and background events is also given for convenience. Note that the expected number of events taken from the `Sherpa` simulation have been normalised to reproduce the next-to-leading order calculations for Zjj production obtained from `Powheg` in the *baseline* region. The results are shown for both the individual lepton channels and their combination. The latter is obtained by combining the two channels for the data and the respective MC

predictions before performing the fit. The resulting yield for the combination is used to determine systematic uncertainties that are correlated between the lepton channels (cf. section 6.3), whereas the yields from the individual decay channels are used to calculate the fiducial cross section (cf. section 6.5).

Figure 6.2 shows the detector-level m_{jj} distribution in the *search* region for the individual lepton channels as well as their combination. The simulation is normalised to match the number of events obtained from the fit. The ratio of the constrained background-only model to the data is shown in the lowest panel. The dashed line shows the ratio using the unconstrained background-only model, thus demonstrating that the effect of the *control*-region constraint is small compared to the disagreement in the tail of the m_{jj} spectrum between the background-only model and the data. However, good agreement is observed between the data and the simulation if the electroweak component is included. The ratio of the combined signal and (constrained) background model to the data is shown in the middle panel.

6.2   Validation of the data-driven background constraint

A key feature of the analysis is the reweighting function derived in the *control* region. This data-driven correction is applied to the background model in the *search* region, improving the modelling of the m_{jj} spectrum, whilst reducing both experimental and theoretical systematic uncertainties associated with the background model. A number of cross checks have been performed in order to validate the constraint procedure.

6.2.1   Choice of polynomial

The reweighting function is obtained by fitting the ratio of data to simulation with a second-order polynomial. The analysis chain has been repeated using a first-order polynomial instead. This alternative reweighting function is indicated with a dashed line in the lower panels in figure 6.1 and it can be seen that both choices of polynomial lead to very similar reweighting functions. The second-order polynomial describes the ratio of data to simulation slightly better at large m_{jj} than the first-order polynomial. The change in the signal yield is less than 2 % if a first-order polynomial is used instead of the second-order polynomial.

6.2.2   Choice of event generator

The detector-level m_{jj} spectrum simulated with `Sherpa` has been reweighted using the ratio of the `Powheg` and `Sherpa` particle-level predictions. This reweighting is done for the *search* and the *control* region separately as `Powheg` has been shown to give a better description of the m_{jj} spectrum in all of the fiducial regions (cf. appendix A). Figure 6.3 shows the ratio of the reweighted m_{jj} spectrum to the nominal `Sherpa` prediction in the *control* region along with the best fit of a second-order polynomial as indicated by a black curve. In addition, the nominal reweighting function is shown as a red curve. The striking similarity of the two curves demonstrates that the reweighting to `Powheg`
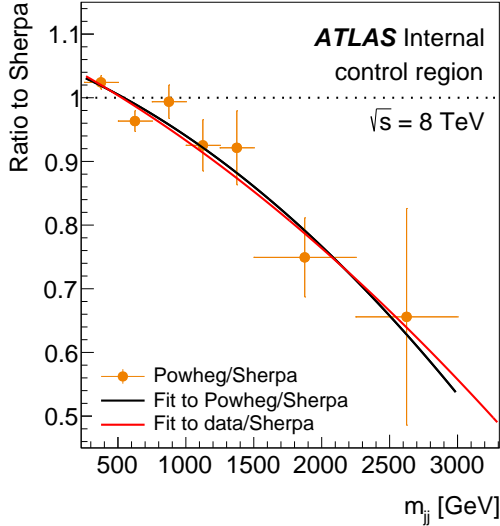
Figure 6.3
The detector-level m_{jj} distribution taken from `Sherpa` is reweighted using the ratio of the `Powheg` and `Sherpa` particle-level predictions and compared to the nominal `Sherpa` detector-level prediction. The black curve shows the best fit of a second-order polynomial to this ratio. In addition, the nominal *control*-region constraint is shown with a red curve, demonstrating that the reweighting to `Powheg` improves the modelling of the m_{jj} spectrum such that the reweighting function becomes much flatter.

improves the modelling of the m_{jj} spectrum such that the data-driven correction obtained using the reweighted `Sherpa` spectrum will become much flatter. Repeating the analysis chain including this reweighting to `Powheg` results in a signal yield consistent at $0.8\,\%$ with the signal yield based on the nominal `Sherpa` predictions.

6.2.3 Choice of control region

The *control* region has been chosen to maximise the available statistics, whilst keeping the expected electroweak contribution in the simulation to a minimum[1]. A jet veto is used to distinguish the *control* region from the *search* region, while the cut on $p_{\mathrm{T}}^{\mathrm{balance},3}$ removes events with poorly measured jets or events which contain jets that do not originate from the hard scattering interaction. This leaves little phase space to construct an orthogonal region serving the same purpose as the *control* region. However, a potential mismodelling of the additional jet activity in the rapidity interval between the two leading jets could still affect the background-model constraint. The impact of any such mismodelling is studied by probing this interval-jet activity. Three pairs of orthogonal subregions are constructed by splitting the nominal *control* region based on

- the transverse momentum of the leading interval jet p_{T}^{j3} (one subregion with $25 < p_{\mathrm{T}}^{j3} \leq 38$ GeV and another subregion with $p_{\mathrm{T}}^{j3} > 38$ GeV),

- the rapidity of the leading interval jet $|y^{j3}|$ (a subregion for $|y^{j3}| \leq 0.8$ and another subregion for $|y^{j3}| > 0.8$)

- as well as the interval-jet multiplicity $N_{\mathrm{jet}}^{\mathrm{gap}}$ (one subregion requiring $N_{\mathrm{jet}}^{\mathrm{gap}} = 1$, another subregion requiring $N_{\mathrm{jet}}^{\mathrm{gap}} \geq 2$).

The phase-space boundaries have been chosen so as to roughly balance the available statistics in the data between orthogonal subregions. Furthermore, the impact of MPI on the *control*-region constraint is examined by defining an MPI-suppressed subregion

---

[1] Note that the signal contribution in the *control* region has been halved and doubled in the simulation and no significant effect on the extracted signal yield has been observed.
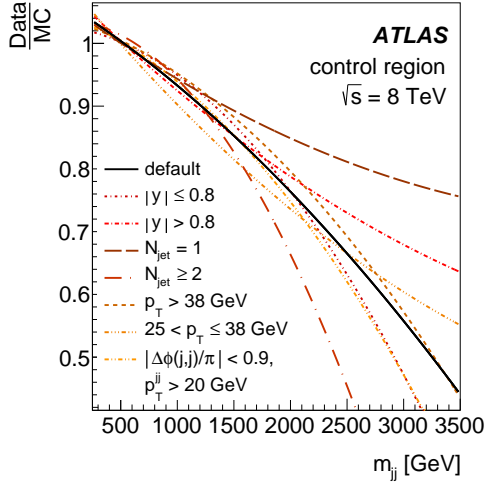
Figure 6.4
Background reweighting functions obtained for different choices of control region.

with the additional requirements of $|\Delta\phi(j,j)/\pi| < 0.9$ and $p_T^{jj} > 20$ GeV, where p_T^{jj} is the transverse momentum of the dijet system.

Each of these subregions is used to derive a data-driven correction and the corresponding reweighting functions are illustrated in figure 6.4. The default reweighting function obtained using the nominal *control* region is shown as a solid black line. The analysis chain is then repeated for each of these background-model constraints and the spread in the extracted signal yields is found to be at most 5 % between subregions. This spread is likely to be statistical in origin as pairs of orthogonal subregions are found to agree to better than 1σ considering only the statistical uncertainty associated with the reweighting functions. This result may still seem surprising given that the spread of the various curves in figure 6.4 appears rather large at high m_{jj}, however, the background modelling in the tail of the m_{jj} spectrum has only a small impact on the extracted number of signal events. The biggest impact comes from the statistics-dominated region of the m_{jj} spectrum around and below 1 TeV, where the spread between the curves is just a few percent.

6.2.4 Proof of principle

While the data in the *control* region is used to constrain the modelling of the background, the simulation is used to extrapolate this constraint from the *control* region to the *search* region. This extrapolation procedure has been examined using the different orthogonal subregions defined in section 6.2.3. Figure 6.5 shows the ratio of data to simulation both before and after applying the background-model constraint derived in a complementary subregion. For instance, the background model in the subregion with $25 < p_T^{j3} \le 38$ GeV has been corrected with a data-driven reweighting function obtained in the orthogonal subregion with $p_T^{j3} > 38$ GeV, as illustrated in figure 6.5(a). Similarly, the data in the subregion with $|y^{j3}| \le 0.8$ is used to constrain the simulation in the subregion with $|y^{j3}| > 0.8$, as shown in figure 6.5(b), and the data-driven correction obtained in the subregion with $N_{jet}^{gap} \ge 2$ has been applied to the background model in the subregion requiring $N_{jet}^{gap} = 1$, as presented in figure 6.5(c). In all cases, the corrected simulation describes the data better than the uncorrected simulation.

(a) split based on p_T^{j3}

(b) split based on y^{j3}

(c) split based on $N_{\text{jet}}^{\text{gap}}$
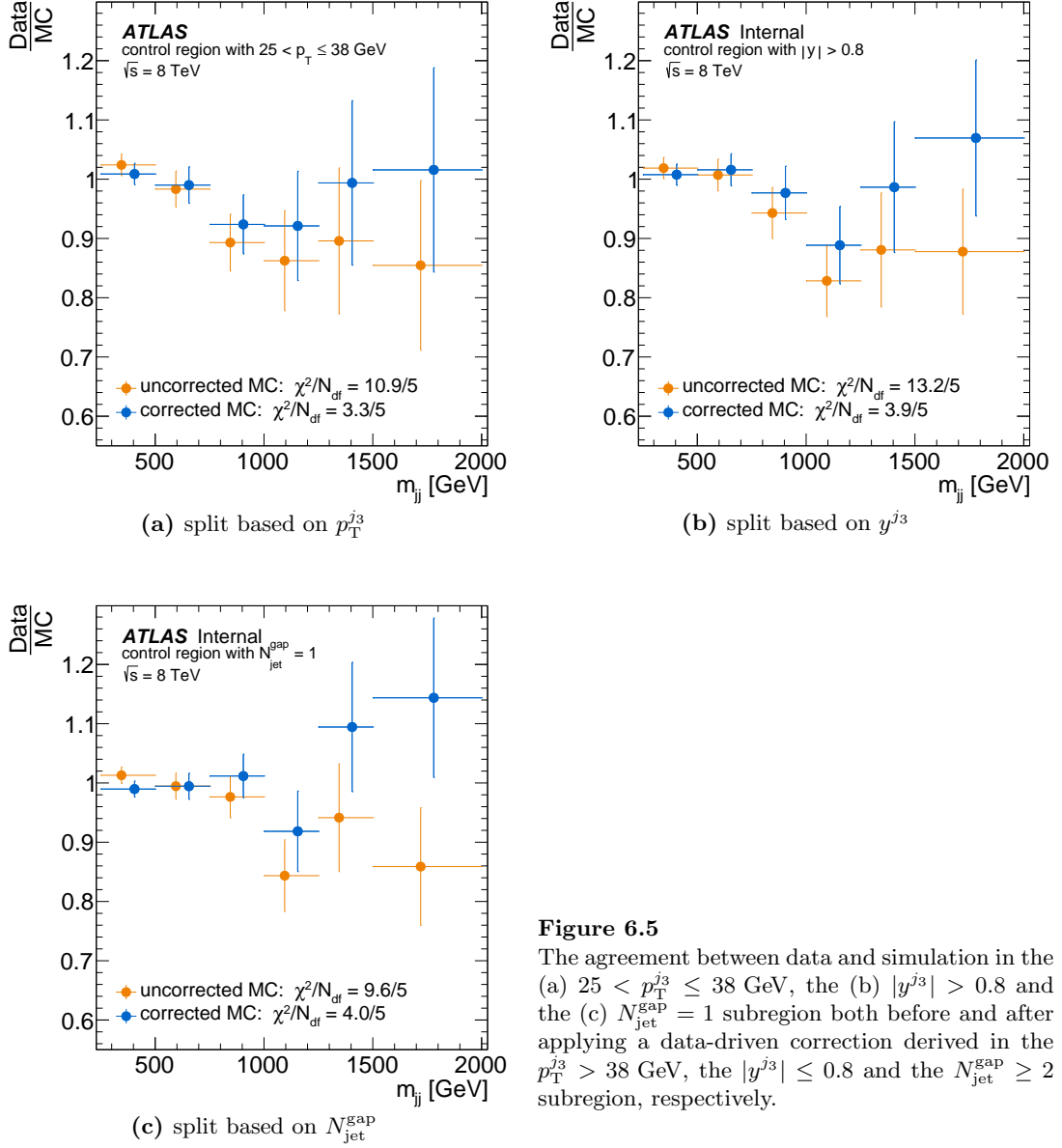
Figure 6.5
The agreement between data and simulation in the (a) $25 < p_T^{j3} \leq 38$ GeV, the (b) $|y^{j3}| > 0.8$ and the (c) $N_{\text{jet}}^{\text{gap}} = 1$ subregion both before and after applying a data-driven correction derived in the $p_T^{j3} > 38$ GeV, the $|y^{j3}| \leq 0.8$ and the $N_{\text{jet}}^{\text{gap}} \geq 2$ subregion, respectively.

6.3 Systematic uncertainties on the signal yield

Systematic uncertainties on the extracted number of electroweak Zjj events, N_{EW}, arise from the jet-based experimental uncertainties, the theoretical modelling of the Zjj simulations and the *control*-region reweighting function. The lepton-based experimental uncertainties do not cause a significant shape change in the m_{jj} spectrum and so the corresponding systematic uncertainty on N_{EW} is negligible. A summary of the systematic uncertainties associated with the signal extraction is presented in table 6.2.

6.3.1 Experimental uncertainties

The systematic uncertainty due to the limited number of events in the *control* region is estimated using pseudoexperiments. This is achieved by fluctuating each bin of the m_{jj} spectrum in the *control* region independently according to a Poisson distribution and a Gaussian distribution for data and MC simulation, respectively. The analysis chain

	$\delta N_{\mathrm{EW}}/N_{\mathrm{EW}}$	
Source	Electrons	Muons
Control region statistics	$\pm\,8.9\,\%$	$\pm\,11.2\,\%$
Jet energy scale	$\pm\,5.6\,\%$	
Jet vertex fraction	$\pm\,1.1\,\%$	
Jet energy resolution	$\pm\,0.4\,\%$	
Pileup jet modelling	$\pm\,0.3\,\%$	
PDF	$^{+1.5}_{-3.9}\,\%$	
Signal modelling	$\pm\,8.9\,\%$	
Background modelling	$\pm\,7.5\,\%$	
Signal/background interference	$\pm\,6.2\,\%$	

Table 6.2
Systematic uncertainties – expressed in percentages – on the number of fitted signal events, N_{EW}, in the *search* region.

is then repeated with the fluctuated spectra and the electroweak Zjj contribution is extracted in the usual fashion. This procedure is repeated 1000 times and the spread in the extracted signal yields is found to be 8.9 % and 11.2 % in the $Z \to e^+e^-$ and the $Z \to \mu^+\mu^-$ channel, respectively.

Jet-based experimental uncertainties cause shape changes in the m_{jj} spectrum and can thus affect the extracted signal yield. The corresponding experimental uncertainties are the same as those discussed in section 4.2, namely uncertainties due to the JES, JER, JVF as well as the pileup jet modelling. Their effect on the extracted signal yield is examined simultaneously for the signal model and the background model in order to account for correlations.

The systematic variations in the signal model due to the experimental uncertainties are evaluated by constructing the ratio of the systematically varied m_{jj} spectrum and the nominal m_{jj} spectrum for each source of systematic uncertainty. The fit of a second-order polynomial to such a ratio then provides a reweighting function for the corresponding systematic variation. The use of the polynomial reduces statistical fluctuations in the tail of the m_{jj} spectrum. The analysis chain is repeated with the additional reweighting function applied to the signal model and the shift in the resulting number of signal events with respect to the nominal signal yield is taken as a systematic uncertainty.

As the shape of the m_{jj} spectrum in the background model is constrained by the data in the *control* region, only the effect of the systematic variations on the extrapolation between *control* and *search* region needs to be evaluated. This is achieved by constructing a double ratio from the ratios of the systematically varied to nominal m_{jj} spectra in the *search* and the *control* region. The fit of a first-order polynomial to such a double ratio then provides a reweighting function that can be used to assess the effect of the corresponding systematic variation on the extrapolation. However, this double ratio is particularly sensitive to statistical fluctuations in the simulation because the systematic shifts in both regions are often very similar. In order to avoid adding up many statistical fluctuations in quadrature and double counting statistical uncertainties in the simulated samples, only statistically significant shifts are included. The gradient of a fitted first-order polynomial is considered a statistically significant shift if the gra-

dient parameter is greater than 1.64 times its uncertainty. The parameter uncertainty is scaled up so as to remove 90 % of statistical fluctuations instead of the usual 68 %. Note that the choice of significance criterion has been investigated by changing the scale factor to 1.0 and 2.0, corresponding to a removal of 68 % and 95 % of statistical fluctuations, respectively. The resultant systematic uncertainties are unchanged with respect to the nominal choice of removing 90 % of statistical fluctuations. For each significant source of systematic uncertainty, the reweighting function obtained from the first-order polynomial fit is applied to the background model in the *search* region in addition to the corresponding reweighting function obtained for the signal model.

The dominant source of experimental uncertainty is the JES (5.6 %). The uncertainty is mostly due to the uncertainty on the signal model shape because the background model is constrained using the *control* region, which significantly reduces the uncertainty on the background model. The uncertainty due to the JVF is small (1.1 %), whereas the contributions from the JER and the pileup jet modelling are both at the subpercent level and thus effectively negligible.

6.3.2 Theory uncertainties

Additional systematic uncertainties on the extraction of the electroweak component arise from the theoretical modelling in the simulation. The theory uncertainties affect the signal model as well as the extrapolation between the *control* and the *search* region for the background model. Uncertainties due to the PDF modelling and the generator modelling are considered separately.

The nominal signal yield has been extracted using the `CT10` PDF set. An uncertainty due to the PDF modelling is estimated by repeating the full analysis chain using simulated samples that have been reweighted according to the `CT10` uncertainties as well as the central values and uncertainties of two other PDF sets, namely `MSTW2008nlo` [68] and `NNPDF2.3` [98]. Each PDF variation is applied to the signal and background model simultaneously in order to account for correlations. For each PDF set, the uncertainty on the signal yield is calculated using the procedure recommended by the respective collaboration [99, 100]. Since the `CT10` uncertainties are given at 90 % confidence level, the default uncertainties have been scaled to reflect the conventional 68 % tolerance interval instead. The uncertainty due to the strong coupling α_S is found to be negligible. The overall systematic uncertainty due to the PDF modelling is found to be $^{+1.5}_{-3.9}$ %, calculated from the envelope of uncertainties obtained from the individual PDF sets.

The generator modelling uncertainty is estimated by running pseudoexperiments using the suite of `Sherpa` particle-level samples introduced in section 3.2.1, which include variations of the factorisation and renormalisation scale, the MPI activity as well as changes to the parton-shower scheme or the CKKW matching parameters. Pseudoexperiments are performed for each sample variation independently in order to obtain a shift in N_EW along with the corresponding standard deviation for each of the variations.

For a given signal model variation, the bins of the particle-level m_{jj} spectrum are fluctuated independently according to a Gaussian distribution. The fit of a second-order polynomial to the ratio of the fluctuated simulation and the nominal particle-level

simulation provides a reweighting function for the corresponding systematic variation. The nominal particle-level prediction is not fluctuated, so as to avoid double counting the statistical uncertainty in the simulation. The analysis chain is then repeated with this additional reweighting function applied to the signal model and the number of signal events is extracted in the usual fashion. This procedure is repeated 1000 times resulting in a distribution of signal yields that is fitted with a Gaussian in order to extract both mean and standard deviation of the resulting shift in N_{EW}. The effect of the signal modelling uncertainty on the nominal signal yield is found to be 8.9 %, calculated from the envelope of shifts (mean plus standard deviation) produced from the individual signal model variations.

A separate uncertainty on N_{EW} is obtained from the envelope of the shifts (mean plus standard deviation) produced from the individual background model variations. For a given variation, the bins of the particle-level m_{jj} spectra in *search* and *control* region are fluctuated independently according to a Gaussian distribution. A double ratio is constructed from the ratios of the fluctuated to nominal m_{jj} spectra in the *search* and the *control* region. The nominal simulation is again not fluctuated, so as to avoid double counting the statistical uncertainty in the simulation. The fit of a first-order polynomial to this double ratio then provides a reweighting function that can be used to assess the effect of the corresponding modelling variation on the extrapolation between *control* and *search* region. The analysis chain is then repeated with this additional reweighting function applied to the background model in the *search* region and the number of signal events is extracted in the usual fashion. This procedure is repeated 1000 times resulting in a distribution of signal yields that is again fitted with a Gaussian in order to extract both mean and standard deviation. The effect of the background modelling uncertainty on N_{EW} is found to be 7.5 %. As the systematic variations may affect the signal and background model in different ways, the uncertainties due to the signal and background modelling are assumed to be uncorrelated.

In order to cross check the assigned modelling uncertainties, the `Sherpa` signal and background m_{jj} spectra have been reweighted to the `Powheg` prediction using the ratio of the `Powheg` and `Sherpa` particle-level predictions. As discussed in section 6.2.2, the reweighting of the background model to `Powheg` changes the signal yield by 0.8 %, which is covered by the background modelling uncertainty obtained from the `Sherpa` variation samples. Reweighting the signal model to `Powheg` changes the signal yield by 4.6 %, which too is covered by the signal modelling uncertainty obtained from the corresponding `Sherpa` variation samples.

A systematic uncertainty associated with possible interference between electroweak and strong Zjj production is estimated by reweighting the background template to account for the interference contribution. The interference is examined using a set of dedicated `Sherpa` samples as discussed in section 3.2. Since these samples are generated using only leading-order matrix elements for Zjj production, the change to the background model due to interference is estimated in the *baseline* region, i.e. prior to applying the jet veto. The impact of interference is then evaluated by repeating the full analysis chain after reweighting the background model in either the *search* or the

control region alone. This approach assumes that the potential interference only affects one of the two regions and therefore has a maximal impact on the signal extraction. The signal yield is reduced by 6.2 % if the background model is reweighted only in the *search* region to account for interference. Conversely, if the background model is reweighted only in the *control* region, the signal yield increases by 6.2 %. A conservative systematic uncertainty of ±6.2 % has been assigned to the final measurement.

6.4   Significance estimate

Regardless of whether the simulation has been taken directly from `Sherpa` or the modelling has been constrained with the data in the *control* region, the background model alone fails to describe the tail of the m_{jj} spectrum for both lepton channels, as shown in figure 6.2. However, the simulation agrees with the data if the electroweak component is included. Assuming there was no electroweak contribution, an unfortunate statistical fluctuation in the data could be mistakenly extracted as signal. The probability for such a fluctuation to result in a signal yield that is at least as large as the one observed in this measurement is commonly referred to as the *p*-value. The *p*-value can be expressed in terms of the standard deviation σ using the conversion rule

$$Z = \text{probit}\,(1 - p) = \sqrt{2}\,\text{erf}^{-1}\,(1 - 2p) \tag{6.1}$$

where Z is the number of standard deviations, probit is the quantile function of the normal distribution and erf^{-1} is the inverse Gauss error function. The *p*-value of this measurement has been estimated using pseudoexperiments in order to evaluate systematic effects in much the same way as in the principal analysis chain.

The background simulation in both *search* and *control* region is first corrected using the nominal data-driven constraint derived in the *control* region and then scaled such that the integral of the m_{jj} spectrum in the *control* region matches the corresponding number of events observed in the data. Pseudodata are generated for both regions by fluctuating each bin of the respective background-only m_{jj} spectrum independently according to a Poisson distribution. Signal and background pseudomodels are constructed from the nominal models in a similar fashion. Each bin of the respective m_{jj} spectrum is fluctuated independently according to a Gaussian distribution in order to account for the statistical uncertainty in the simulation. In addition, the shape of the m_{jj} spectra in both pseudomodels are smeared according to their experimental and theoretical uncertainties. For each source of systematic uncertainty, the bins of the corresponding m_{jj} spectrum are shifted by the respective uncertainty interval times a normally distributed random number. The full analysis chain is then repeated, including the use of a pseudodata-driven reweighting function obtained in the *control* region. The number of signal events are extracted and the entire procedure is repeated.

Figure 6.6 shows the resulting signal yield distribution after running one million pseudoexperiments. The different curves show the impact of systematic uncertainties if the corresponding smearing is applied to the signal only, the background only, none or
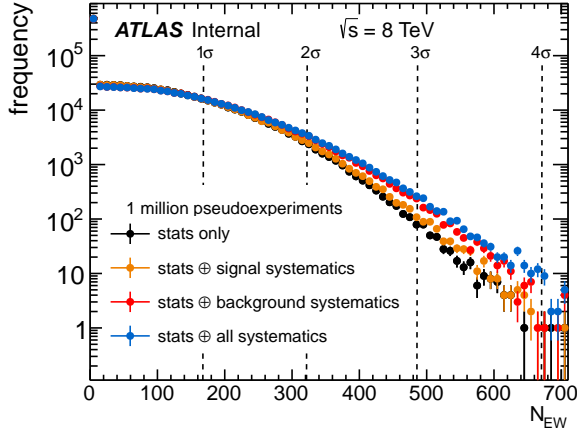
Figure 6.6
Signal yield distribution after running one million pseudoexperiments. The analysis chain is repeated using pseudodata generated by Poisson fluctuating the background model. The different curves show the impact of systematic uncertainties if none, only the signal model, only the background model or both models are subjected to Gaussian smearing according to the experimental and theoretical uncertainties.

both of the pseudomodels, respectively. The large spike for $N_{\mathrm{EW}} = 0$ is expected since the pseudodata are generated from the background-only model and so the signal model is not needed to describe the data in most of the pseudoexperiments. By construction, values of $N_{\mathrm{EW}} > 0$ correspond to statistical fluctuations in the pseudodata that have been misinterpreted as signal. Systematic variations in the pseudomodels change the shape of the corresponding m_{jj} spectrum, thus increasing the chance for the pseudodata to be misinterpreted as signal. This is indicated by broader signal yield distributions when the pseudomodels are subjected to systematic smearing. Systematic variations in the signal model have the smallest effect because the pseudodata is generated from the background-only model.

Figure 6.7 shows the signal yield distribution after running one billion pseudoexperiments. None of the trials have led to a value of N_{EW} greater than or equal to the 1657 events observed in the data. One in a billion is equivalent to a p-value of 10^{-9} which corresponds to 6.0σ significance. The background-only hypothesis is therefore rejected at greater than 6σ significance.

Running a billion pseudoexperiments is a computationally expensive endeavour. Based on the curvature of the signal yield distribution in figure 6.7 though, it appears that around half a trillion pseudoexperiments would be required in order to reach the observed value of 1657 signal events. This is a prohibitively large number and hence no attempt has been made to exactly quantify the significance of this measurement.

Figure 6.7
Signal yield distribution after running one billion pseudoexperiments. The analysis chain is repeated using pseudodata generated by Poisson fluctuating the background model. The signal and background model have been subjected to Gaussian smearing according to the experimental and theoretical uncertainties. The background-only hypothesis is rejected at greater than 5σ significance.
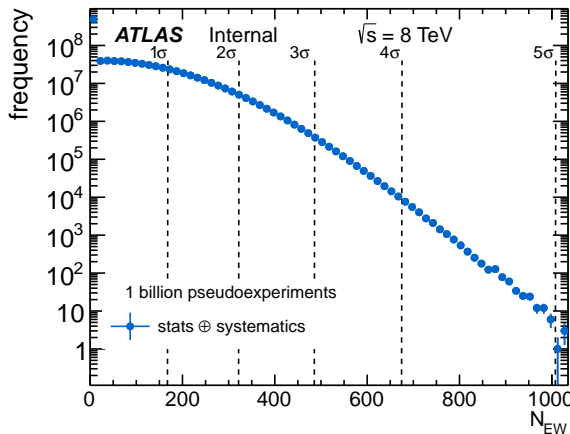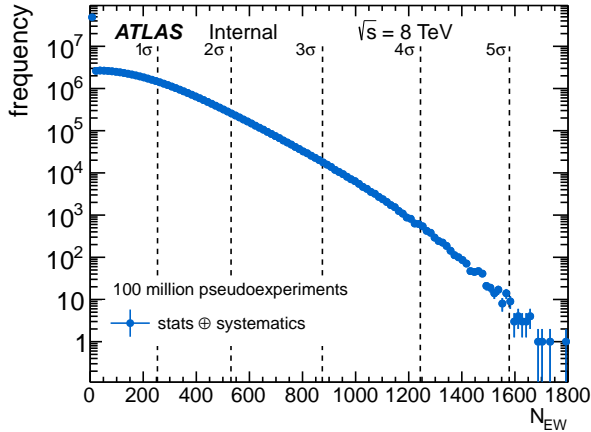
Figure 6.8
Signal yield distribution after running 100 million pseudoexperiments. The analysis chain is repeated using pseudodata generated by Poisson fluctuating the background model. The signal and background model have been smeared for each source of systematic uncertainty according to a uniform distribution in the range -5 to $+5$ times the respective uncertainty. Even in this extremely conservative approach, the background-only hypothesis is rejected at greater than 5σ significance.

6.4.1 Possible impact of non-Gaussian tails

In order to evaluate systematic effects, the signal and background pseudomodels are smeared according to their experimental and theoretical uncertainties. For each source of systematic uncertainty, the bins of the m_{jj} spectra are shifted by the corresponding uncertainty interval times a normally distributed random number. This approach assumes that the systematic uncertainties are Gaussian distributed. In order to investigate the possible impact of non-Gaussian tails in the systematic uncertainties, the pseudoexperiments have been repeated using a random number that is uniformly distributed in the range -5 to $+5$. This implies that for each source of systematic uncertainty, 5σ effects in either direction are just as likely as no effect at all. The signal yield distribution shown in figure 6.8 is the result of running 100 million of these modified pseudoexperiments. It can be seen that even in this extremely conservative approach, the background-only hypothesis is still rejected at greater than 5σ significance, and hence the significance is formally only claimed to be beyond the 5σ level.

6.5 Fiducial cross sections

Cross sections for electroweak Zjj production are measured in the *search* region defined in section 3.3.1. The signal yields extracted in the electron and the muon channel are each converted into a fiducial cross section using the formula

$$\sigma_{\mathrm{EW}} = \frac{N_{\mathrm{EW}}}{\mathcal{C}_{\mathrm{EW}} \times L_{\mathrm{int}}} \tag{6.2}$$

where N_{EW} is the extracted number of electroweak events, $\mathcal{C}_{\mathrm{EW}}$ is a correction factor based on the ratio of the detector- and particle-level predictions for electroweak Zjj production taken from `Sherpa`, and L_{int} is the integrated luminosity.

The correction factors for the electron and muon channel are 0.66 and 0.80, respectively. The difference between the two decay channels is primarily due to different reconstruction and identification efficiencies for muons and electrons. Systematic uncertainties on the correction factor are divided into those that are correlated between the lepton channels (such as jet-based systematics, generator modelling and PDF un-

Source	$\delta N_{\mathrm{EW}}/N_{\mathrm{EW}}$		$\delta \mathcal{C}_{\mathrm{EW}}/\mathcal{C}_{\mathrm{EW}}$	
	Electrons	Muons	Electrons	Muons
Lepton systematics	—	—	$\pm 3.2\,\%$	$\pm 2.5\,\%$
Control region statistics	$\pm 8.9\,\%$	$\pm 11.2\,\%$	—	—
Jet energy scale	$\pm 5.6\,\%$		$^{+2.7}_{-3.4}\,\%$	
Jet vertex fraction	$\pm 1.1\,\%$		$^{+0.4}_{-1.0}\,\%$	
Jet energy resolution	$\pm 0.4\,\%$		$\pm 0.8\,\%$	
Pileup jet modelling	$\pm 0.3\,\%$		$\pm 0.3\,\%$	
PDF	$^{+1.5}_{-3.9}\,\%$		$\pm 0.1\,\%$	
Signal modelling	$\pm 8.9\,\%$		$^{+0.6}_{-1.0}\,\%$	
Background modelling	$\pm 7.5\,\%$		—	
Signal/background interference	$\pm 6.2\,\%$		—	

certainties) and those that are not. Correlated systematic uncertainties are evaluated by first combining the two channels. The systematic uncertainties on $\mathcal{C}_{\mathrm{EW}}$ are summarised in table 6.3, along with the corresponding systematic uncertainties on N_{EW} for completeness. The latter uncertainties are the same as those provided in table 6.2.

For the correction factor, the uncorrelated uncertainties and the jet-based uncertainties have been evaluated using the same procedure described in section 4.2 for the inclusive cross-section measurement. The lepton-based uncertainties are found to be at the level of 3 % for both electrons and muons. The dominant source of experimental uncertainty is due to the JES which is also at the level of 3 %. The uncertainty due to the JVF is about 1 %, whilst the uncertainties due to the JER and the pileup jet modelling are both at the subpercent level.

The generator modelling can affect the kinematics of final-state particles and hence result in phase-space migrations which in turn affect the correction factor. However, the systematic uncertainty on the correction factor due to possible generator mismodelling cannot be evaluated in the same way as for the inclusive cross sections (discussed in section 4.2) because the simulation would then have to be reweighted to a data distribution of the signal process alone, which of course is not available. A systematic uncertainty is instead obtained from the PDF variations and the suite of `Sherpa` particle-level samples introduced in section 3.2.1, which includes variations of the factorisation and renormalisation scale, the MPI activity as well as changes to the parton-shower scheme or the CKKW matching parameters.

The PDF uncertainty is estimated in a similar fashion as described in section 6.3.2. The signal model is reweighted according to the `CT10` uncertainties as well as the central values and uncertainties taken from `MSTW2008nlo` and `NNPDF2.3`. The overall systematic uncertainty due to the PDF modelling is found to be $\pm 0.1\,\%$, calculated from the envelope of uncertainties obtained from the individual PDF sets. For the systematic uncertainty obtained from the theory variation samples, the amount of phase-space

migrations depends on the differences between the systematically varied and the nominal distributions of those variables that are used to define the *search* region. The corresponding uncertainty for a given variation sample is estimated by reweighting the nominal simulation such that the particle-level distribution matches the one predicted by the systematically varied simulation. The procedure is repeated to extract systematic uncertainties due to generator modelling for all `Sherpa` variation samples and for all kinematic variables used to define the *search* region, including the reconstructed objects. The final systematic uncertainty due to the generator modelling is found to be at the level of 1 %, obtained by adding the individual uncertainties in quadrature.

The impact of N_{EW} and $\mathcal{C}_{\mathrm{EW}}$ is found to be anticorrelated for each source of systematic uncertainty in table 6.3. For instance, an increase in the JES results in a larger correction factor, simply because more events will pass the detector-level selection in the simulation, whereas the number of events that pass the particle-level selection in the simulation remains the same. However, an increase in the JES also results in a harder m_{jj} spectrum for the electroweak component. The normalisation of the m_{jj} spectrum in the signal model is fixed by a fit to the data. The optimal normalisation will be the one that best fills the discrepancy between the background model and the data in the tail of the distribution and so the extracted signal yield will have to decrease in order to compensate for the harder spectrum. Therefore, the impact of N_{EW} and $\mathcal{C}_{\mathrm{EW}}$ is anticorrelated and the uncertainty on the measured cross section is thus given by the linear combination of the individual uncertainties on N_{EW} and $\mathcal{C}_{\mathrm{EW}}$. The total systematic uncertainty on the measured cross section is then obtained by summing the individual sources of systematic uncertainty in quadrature.

The fiducial cross sections in the electron and muon channel are

$$\sigma_{\mathrm{EW}}^{ee} = 67.2 \pm 6.9\,(\text{stat})\,^{+12.7}_{-13.4}\,(\text{syst}) \pm 1.9\,(\text{lumi})\,\text{fb} \tag{6.3}$$

$$\text{and} \quad \sigma_{\mathrm{EW}}^{\mu\mu} = 45.6 \pm 6.1\,(\text{stat})\,^{+9.1}_{-9.6}\,(\text{syst}) \pm 1.3\,(\text{lumi})\,\text{fb}. \tag{6.4}$$

These measurements are consistent at the 1.7σ level, accounting for only those uncertainties that are uncorrelated between the two channels, namely the statistical uncertainty in the data and the simulation, the uncertainty due to limited statistics in the *control* region as well as the lepton-based experimental uncertainties. The individual lepton channels are then combined using a weighted average, with the weight set to the squared inverse of the corresponding uncorrelated uncertainty component.

The combined fiducial cross section is then given by

$$\sigma_{\mathrm{EW}} = 54.7 \pm 4.6\,(\text{stat})\,^{+9.8}_{-10.4}\,(\text{syst}) \pm 1.5\,(\text{lumi})\,\text{fb}. \tag{6.5}$$

The theoretical prediction from `Powheg` for the electroweak Zjj cross section is $46.1 \pm 0.2\,(\text{stat})\,^{+0.3}_{-0.2}\,(\text{scale}) \pm 0.8\,(\text{PDF}) \pm 0.5\,(\text{model})\,\text{fb}$, which is in good agreement with the measurement. Figure 6.9 shows the measurement in relation to several other total and fiducial Standard Model production cross sections measured by ATLAS [101], the fiducial electroweak Zjj cross section being by far the smallest.

A particle-level fiducial cross section for electroweak Zjj production has also been
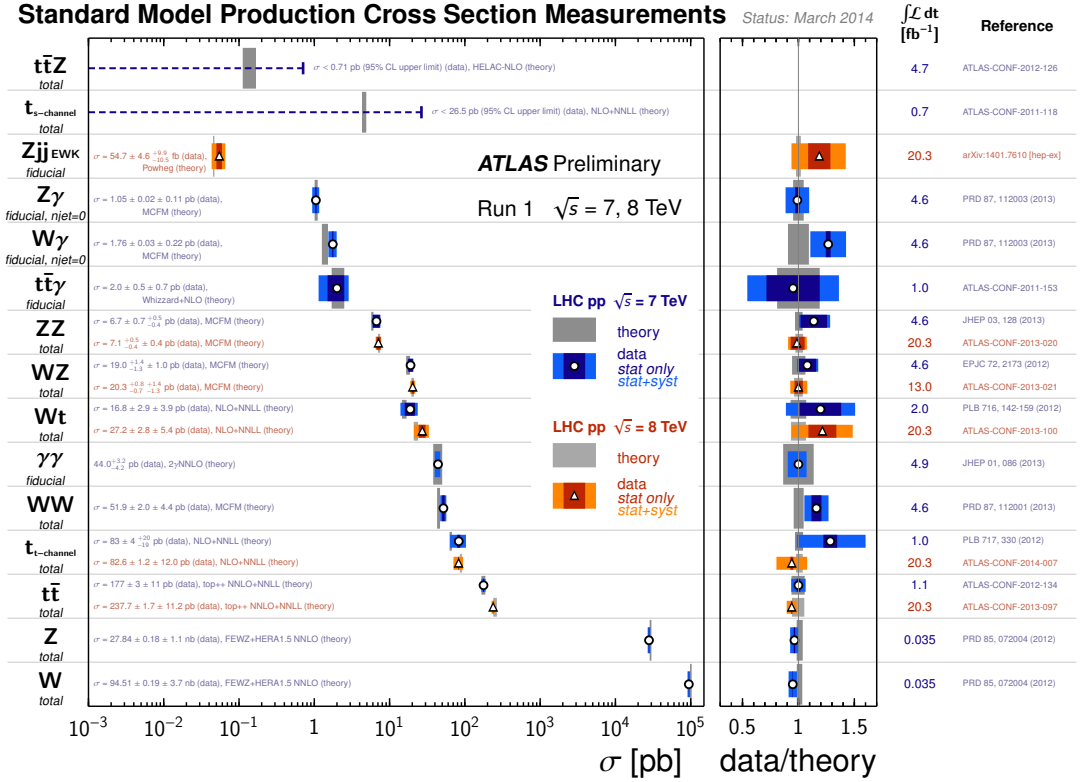
Figure 6.9
Overview of several total and fiducial Standard Model production cross sections measured by ATLAS. The measurements are corrected for leptonic branching fractions and compared to the corresponding theoretical expectations that were calculated at next-to-leading order or higher. The W and Z inclusive cross sections have been measured with 35 pb^{-1} of integrated luminosity from the 2010 dataset; all other measurements use the 2011 and 2012 datasets. The error bar corresponds to the full uncertainty, including systematic and luminosity uncertainties. The statistical component is indicated by a darker colour. The uncertainties on the theoretical predictions are taken from the original ATLAS papers and may not always have been evaluated using the same prescriptions for PDF and scale uncertainties [101].

determined for the *search* region with $m_{jj} > 1$ TeV by getting the fraction of events that pass this cut from the `Sherpa` signal model. In this region, the electroweak Zjj production accounts for about a third of the events. The tail of the m_{jj} distribution is hence the part of the spectrum that is most sensitive to the electroweak component and least sensitive to the normalisation of the background model. The measured cross section for electroweak Zjj production in the *search* region with $m_{jj} > 1$ TeV is given by

$$\sigma_{\text{EW}} \left(m_{jj} > 1 \text{ TeV} \right) = 10.7 \pm 0.9 \, (\text{stat}) \pm 1.9 \, (\text{syst}) \pm 0.3 \, (\text{lumi}) \text{ fb}, \qquad (6.6)$$

which is again in good agreement with the theoretical prediction from `Powheg`, given by $9.38 \pm 0.05 \, (\text{stat}) {}^{+0.15}_{-0.24} \, (\text{scale}) \pm 0.24 \, (\text{PDF}) \pm 0.09 \, (\text{model})$ fb. A breakdown of the experimental and theoretical systematic uncertainties associated with the extracted signal yield and the correction factor is provided in table 6.4 for completeness. Uncertainties associated with the extracted signal yield have been obtained using the same methodology described in section 6.3, with an additional uncertainty for the fraction of events that pass the additional m_{jj} cut in the signal model taken from `Sherpa`. The impact of systematic variations in the m_{jj} spectrum on the extracted yield and the

Table 6.4

Systematic uncertainties, expressed in percentages, on the number of fitted signal events in the *search* region with $m_{jj} > 1$ TeV, N_{EW}, as well as the correction factor to the particle level, $\mathcal{C}_{\mathrm{EW}}$. The uncertainties are anticorrelated between N_{EW} and $\mathcal{C}_{\mathrm{EW}}$.

Source	$\delta N_{\mathrm{EW}}/N_{\mathrm{EW}}$		$\delta\mathcal{C}_{\mathrm{EW}}/\mathcal{C}_{\mathrm{EW}}$	
	Electrons	Muons	Electrons	Muons
Lepton systematics	—	—	$\pm 3.2\,\%$	$\pm 2.7\,\%$
Control region statistics	$\pm 8.8\,\%$	$\pm 11.7\,\%$	—	—
Jet energy scale	$\pm 1.6\,\%$		$^{+7.8}_{-8.2}\,\%$	
Jet vertex fraction	$\pm 1.3\,\%$		$^{+0.5}_{-1.4}\,\%$	
Jet energy resolution	$\pm 0.2\,\%$		$\pm 0.7\,\%$	
Pileup jet modelling	$\pm 0.1\,\%$		$\pm 0.5\,\%$	
PDF	$^{+1.2}_{-3.5}\,\%$		$\pm 0.1\,\%$	
Signal modelling	$\pm 0.9\,\%$		$^{+1.8}_{-1.4}\,\%$	
Background modelling	$\pm 7.5\,\%$		—	
Signal/background interference	$\pm 6.2\,\%$		—	

fraction of events passing the additional m_{jj} cut are anticorrelated, such that those uncertainties that are mostly due to the signal model tend to be a lot smaller. Possible event migrations across the additional phase-space boundary are accounted for by a new correction factor that has been evaluated along with its associated systematic uncertainties using the same methodology used for the nominal *search* region.

6.6 Constraints on new physics

The observation of the electroweak component of Zjj production provides access to the VBF Zjj diagram, shown in figure 3.1(a), which is characterised by the WWZ triple gauge coupling[2]. The physics involved in the coupling of two charged gauge bosons with a neutral gauge boson are predicted by the Standard Model and any anomalies with respect to the Standard Model prediction are indicative of new physics. The theoretical description of the triple gauge coupling is somewhat involved and omitted at this stage as it requires some basic familiarity with concepts of quantum field theory. The interested reader is kindly referred to appendix B for a discussion of the effective Lagrangian density describing the physics behind the triple gauge boson coupling. It suffices to say that after symmetry considerations one ends up with three coupling parameters and a unitarisation scale Λ, representing the energy scale associated with new physics that could be somehow modifying the WWZ coupling.

The most common approach to probing the WWZ triple gauge coupling at a hadron collider is by measuring vector boson pair production. Note that in this case all three gauge bosons entering or leaving the WWZ vertex have a time-like four-momentum. The VBF diagram, however, involves a space-like four-momentum transfer for two of the gauge bosons and thus offers a complementary tool to probe the triple gauge boson vertex since the effects of the boson propagator in vector boson fusion are different

---

[2]This is sometimes referred to as a trilinear gauge boson coupling in the literature.

The 95 % confidence intervals for parameters of the WWZ coupling, obtained from counting the number of events with $m_{jj} > 1$ TeV in the *search* region. Observed and expected intervals are presented for unitarisation scales of $\Lambda = 6$ TeV and $\Lambda = \infty$. The parameter Δg_1^Z refers to the deviation of g_1^Z with respect to its Standard Model value.

coupling parameter	$\Lambda = 6$ TeV		$\Lambda = \infty$	
	observed	expected	observed	expected
Δg_1^Z	$[-0.65, 0.33]$	$[-0.58, 0.27]$	$[-0.50, 0.26]$	$[-0.45, 0.22]$
λ_Z	$[-0.22, 0.19]$	$[-0.19, 0.16]$	$[-0.15, 0.13]$	$[-0.14, 0.11]$

from those in vector boson pair production. The potential benefits of using the VBF diagram as a probe of anomalous triple gauge couplings have been discussed previously in the literature, e.g. in ref. [102] which emphasises that full information on triple gauge boson couplings can be obtained only if electroweak vector boson production is measured in addition to vector boson pair production. Since the fiducial cross-section measurements for electroweak Zjj production are in good agreement with the Standard Model predictions, the number of events in the *search* region with $m_{jj} > 1$ TeV is used to set limits on anomalous triple gauge couplings. Note that this region of the m_{jj} spectrum is most sensitive to the electroweak component and least sensitive to the normalisation of the background model. The extracted background yield is 592 events, while 900 events are observed in the data. The expected signal yield for this region is 261 events, estimated using the `Sherpa` simulation. The change in the expected signal yield due to anomalous triple gauge couplings is then estimated by varying the coupling parameters within the `Sherpa` simulation and a profile likelihood test [103] is performed in order to estimate the confidence intervals for these anomalous couplings using a frequentist approach. A Poisson likelihood function is constructed from the number of events observed in the data, the extracted number of background events, the expected signal yield as a function of the anomalous couplings and, furthermore, the systematic uncertainties which are included as nuisance parameters with correlated Gaussian constraints. Confidence intervals are estimated using pseudoexperiments, whereby a given anomalous coupling is rejected at 95 % confidence level if at least 95 % of pseudoexperiments yield a profile likelihood ratio larger than the observed profile likelihood ratio. As discussed in appendix B, the electroweak Zjj channel is most sensitive to two of the three coupling parameters, namely g_1^Z and λ_Z. In the Standard Model these parameters have the values $g_1^Z = 1$ and $\lambda_Z = 0$. It is common practice to quote limits on Δg_1^Z, the deviation of g_1^Z with respect to its Standard Model value, such that all confidence intervals are centred around zero. The 95 % confidence intervals restricting anomalous values of the parameters Δg_1^Z and λ_Z are presented in table 6.5 for two unitarisation scales. Although not as stringent as the ones derived in a recent measurement of WZ production [104] for instance, the confidence intervals shown in table 6.5 represent the first limits on anomalous WWZ couplings at a hadron collider that make use of the vector boson fusion process.

7

Implications for VBF Higgs measurements at the LHC

Vector boson fusion is the second largest production mechanism of the Higgs boson at the LHC. Efforts are under way to seek for evidence of VBF Higgs production in a number of final states as the VBF topology provides a clean environment that can be used to make precision measurements of the Higgs boson couplings [105]. Many ATLAS analyses compare their VBF Higgs cross section measurement with the Standard Model prediction obtained from `Powheg` [106]. In the electroweak Zjj channel, measurement and `Powheg` calculation are found to agree well with each other in two regions of phase space (cf. section 6.5), which begs the question of whether the `Powheg` simulation will do a similarly good job in the case of VBF Higgs production. Note that the production of a Z boson through vector boson fusion is kinematically very similar to the VBF production of a Higgs boson as illustrated in figures 7.1(a) and 7.1(b), respectively. Although the two processes cannot be compared directly as the VBF Zjj diagram interferes strongly with Z-bremsstrahlung diagrams to form an electroweak Zjj component (cf. figure 3.1), both types of processes involve rapidity gaps formed by two electroweakly produced jets. The study of this type of topology is important for our understanding of jet distributions involving rapidity gaps and can be used to improve the modelling in the simulation for the benefit of either channel. In a confidence-building exercise, the electroweak Zjj analysis framework has been extended to test the `Powheg` method in regions of phase space that have been optimised for the VBF production of a Higgs boson in association with two jets. The jet selection of the *search* region used to extract the electroweak Zjj component is modified, so as to employ a jet selection optimised for VBF Hjj production. This cross check has been performed for both the dominant dilepton and diboson decay channel of the Higgs boson, namely $H \to \tau^+\tau^-$ and $H \to WW^*$, respectively.

(a) VBF Zjj production   **(b)** VBF Hjj production
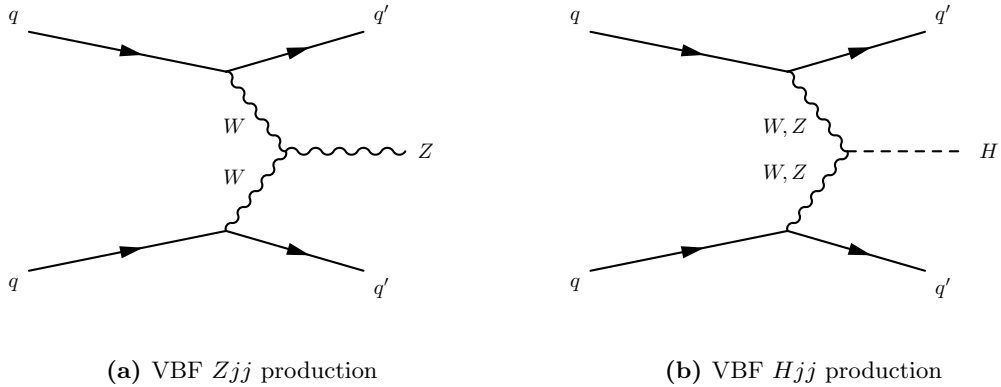
Figure 7.1
Leading order Feynman diagrams for the production of a Z boson and a Higgs boson proceeding through vector boson fusion at the LHC.

7.1   Validation in the dilepton decay channel

A VBF $H \to \tau^+\tau^-$ search region is constructed from a Z-boson candidate, defined exactly as in the nominal analysis (cf. section 3.3.1), and a jet selection optimised for the VBF $H \to \tau^+\tau^-$ channel where the two tau leptons decay leptonically. This cut-based selection is used as a cross check in a dedicated VBF $H \to \tau^+\tau^-$ analysis which uses multivariate techniques to extract the VBF component. Jets are reconstructed using the anti-k_t jet clustering algorithm [40] with a jet-radius parameter of 0.4 and within a rapidity range of $|y| < 4.4$. Jets are also required to be well separated from selected leptons, meaning that the distance between a jet and any of the selected leptons ought to be $\Delta R_{j,\ell} \geq 0.3$ in η-ϕ space. At least two jets are required that satisfy $p_\mathrm{T}^{j_1} > 40$ GeV and $p_\mathrm{T}^{j_2} > 30$ GeV, where j_1 and j_2 refer to the leading and subleading transverse momentum jets in the event, respectively. The invariant mass of the dijet system is required to satisfy $m_{jj} > 250$ GeV, which removes a large fraction of diboson events. Note that in the dedicated VBF $H \to \tau^+\tau^-$ analysis, a tighter cut of $m_{jj} > 400$ GeV is used in order to enhance the signal further. This cut has been relaxed with respect to the dedicated analysis, since the electroweak Zjj component is extracted via a fit to the m_{jj} spectrum. The rapidity separation between the leading two jets must satisfy $|\Delta y| > 3.0$ and ought to be free of additional jets with $p_\mathrm{T} > 25$ GeV, where interval jets are only considered within the inner detector tracking acceptance ($|y| < 2.4$). Finally, the rapidity interval between the two leading jets must contain the lepton pair representing the Z-boson candidate.

The signal model is taken from the `Sherpa` electroweak Zjj sample and the background model is constructed from the `Sherpa` strong Zjj sample as well as the small contributions from diboson and $t\bar{t}$ samples. Contributions from $W + $ jets and multijet events are found to have negligible impact on the results as they do not cause significant shape changes in the m_{jj} spectrum. A key feature of the nominal analysis is the reweighting function derived in the *control* region. This data-driven correction is applied to the background model in the *search* region, improving the modelling of the m_{jj} spectrum, whilst reducing both experimental and theoretical systematic un-

(a) VBF $H \to \tau^+\tau^-$ control region   **(b)** VBF $H \to \tau^+\tau^-$ search region
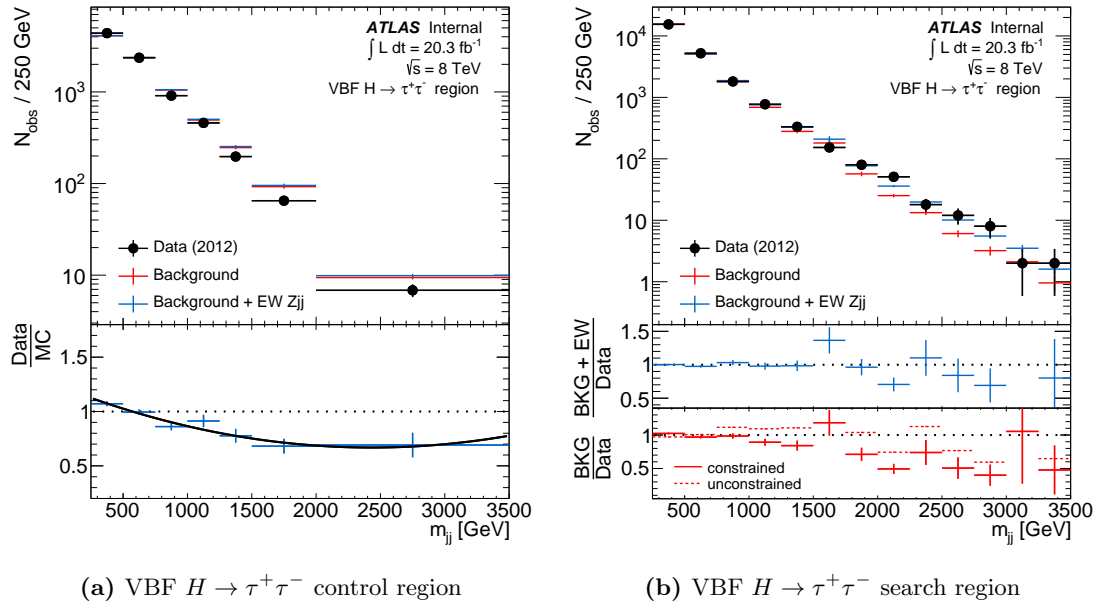
Figure 7.2
The dijet invariant mass distribution in the (a) VBF $H \to \tau^+\tau^-$ control region as well as the (b) VBF $H \to \tau^+\tau^-$ search region for the combined electron and muon channel. The simulation in subfigure (a) has been normalised to match the number of events observed in the data. The lower panel shows a second order polynomial fit to the ratio of data to simulation which has been used as a reweighting function to constrain the shape of the background model in the corresponding search region. The simulation in subfigure (b) has been normalised to match the fitted number of events. The middle panel shows the ratio of the combined signal and background model to the data. The lowest panel shows the ratio of the constrained (solid) and unconstrained (dashed) background-only model to data. The bins in this plot reflect the binning used in the fit.

certainties associated with the background model. In order to maintain the primary analysis chain, a VBF $H \to \tau^+\tau^-$ control region has been constructed by reversing the jet veto requirement imposed in the VBF $H \to \tau^+\tau^-$ search region. The detector-level m_{jj} spectrum in this VBF $H \to \tau^+\tau^-$ control region is shown in figure 7.2(a) for the combined $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ lepton channels. The lower panel shows the ratio of data to simulation as well as a second order polynomial fit to this ratio. This data-driven correction obtained in the VBF $H \to \tau^+\tau^-$ control region is then directly applied to the background Monte Carlo model in the VBF $H \to \tau^+\tau^-$ search region. A fit to the data is then performed using the signal model and the constrained background model. The normalisation of each model is allowed to float and the signal and background yields are extracted after performing a log-likelihood maximisation [97]. The number of signal (N_{EW}) and background (N_{bkg}) events obtained from the fit are summarised in table 7.1, along with the statistical uncertainties due to the data. In addition, the corresponding number of signal and background events expected by the MC simulation are given with an uncertainty due to the limited statistics in the simulation. The sum of the fitted signal and background events is also given for convenience. The expected number of signal events taken from the `Sherpa` simulation have been normalised to reproduce the next-to-leading order calculation for electroweak Zjj production obtained from `Powheg` in the VBF $H \to \tau^+\tau^-$ search region. The statistical uncertainty due to the limited number of events in the control region has been estimated using pseudoexperiments as described in section 6.3.1 and is found to be 22.0 %.

Table 7.1

The number of background events (N_{bkg}) and electroweak Zjj events (N_{EW}) as predicted by the MC simulation and obtained from a fit to the data. The uncertainties on the MC prediction and the fitted yields reflect the statistical uncertainties in the simulation and the data, respectively. The sum of the fitted signal and background events is also given for convenience.

|	
Predicted N_{bkg}	$23\,920 \pm 155$
Predicted N_{EW}	$1\,015 \pm 32$
Fitted N_{bkg}	$23\,158 \pm 113$
Fitted N_{EW}	732 ± 188
$N_{\mathrm{bkg}} + N_{\mathrm{EW}}$	$23\,890$

The extracted and expected signal yield are consistent at the 1.4σ level, accounting only for the uncertainties due to limited statistics in the data, the simulation and the control region.

Figure 7.2(b) shows the detector-level m_{jj} distribution in the VBF $H \rightarrow \tau^+\tau^-$ search region for the combined $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ lepton channels. The simulation is normalised to match the number of events obtained from the fit. The ratio of the constrained background-only model to the data is shown in the lowest panel. The dashed line shows the ratio using the unconstrained background-only model, thus demonstrating that the background-only model struggles to describe the data in the tail of the m_{jj} spectrum regardless of whether the constraint is applied or not. Better agreement is observed, however, between the data and the simulation if the electroweak component is included. The ratio of the combined signal and (constrained) background model to the data is shown in the middle panel.

7.2 Validation in the diboson decay channel

A VBF $H \rightarrow WW^*$ search region is constructed from a Z-boson candidate defined exactly as in the nominal analysis (cf. section 3.3.1) and a jet selection optimised for the VBF $H \rightarrow WW^*$ channel where the two W bosons decay leptonically. This cut-based selection is used as cross check in a dedicated VBF $H \rightarrow WW^*$ analysis which uses multivariate techniques to extract the VBF component. Jets are reconstructed using the anti-k_{t} jet clustering algorithm [40] with a jet-radius parameter of 0.4 and within a pseudorapidity range of $|\eta| < 4.4$. Jets are required to be well separated from selected leptons, meaning that the distance between a jet and any of the selected leptons ought to be $\Delta R_{j,\ell} \geq 0.3$ in η-ϕ space. At least two jets are required that satisfy $p_{\mathrm{T}} > 25$ GeV if they are within the inner detector tracking acceptance ($|\eta| < 2.4$) or to have a minimum transverse momentum of $p_{\mathrm{T}} > 30$ GeV otherwise. The invariant mass of the dijet system is required to satisfy $m_{jj} > 250$ GeV, which removes a large fraction of diboson events. Note that in the dedicated VBF $H \rightarrow WW^*$ analysis, a tighter cut of $m_{jj} > 600$ GeV is used in order to enhance the signal further. This cut has been relaxed with respect to the dedicated analysis, since the electroweak Zjj component is extracted via a fit to the m_{jj} spectrum. The rapidity separation between the leading two jets must satisfy $|\Delta y| > 3.6$ and ought to be free of additional jets with $p_{\mathrm{T}} > 20$ GeV. Finally, the rapidity interval between the two leading jets must contain the lepton pair representing the Z-boson candidate.

Just like before, the signal model is taken from the `Sherpa` electroweak Zjj sample and the background model is constructed from the `Sherpa` strong Zjj sample as well

(a) VBF $H \to WW^*$ control region   **(b)** VBF $H \to WW^*$ search region
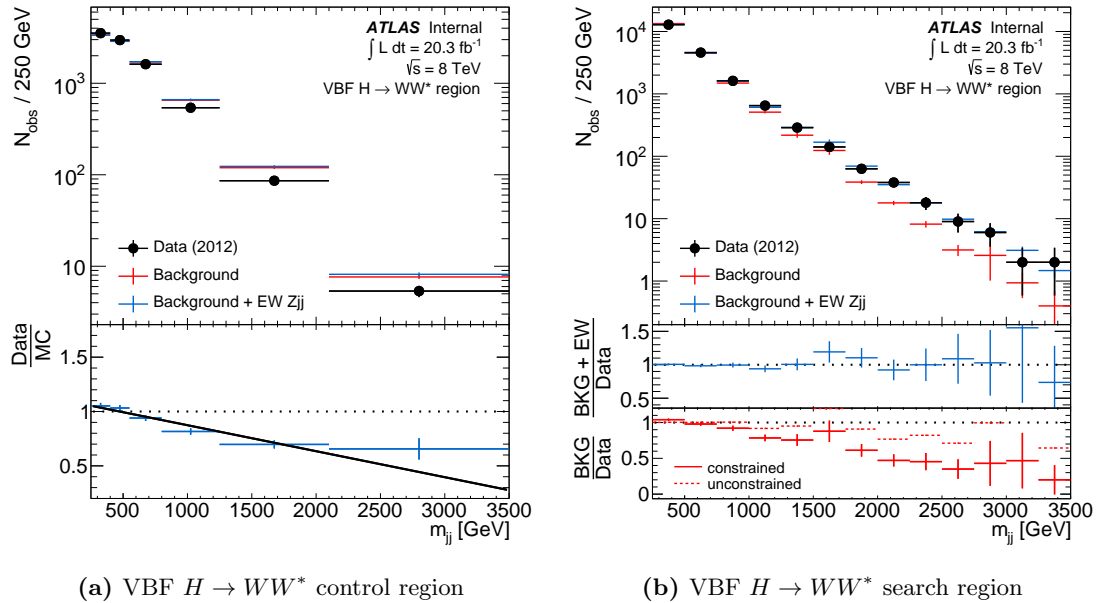
Figure 7.3

The dijet invariant mass distribution in the (a) VBF $H \to WW^*$ control region as well as the (b) VBF $H \to WW^*$ search region for the combined electron and muon channel. The simulation in subfigure (a) has been normalised to match the number of events observed in the data. The lower panel shows a first order polynomial fit to the ratio of data to simulation which has been used as a reweighting function to constrain the shape of the background model in the corresponding search region. The simulation in subfigure (b) has been normalised to match the fitted number of events. The middle panel shows the ratio of the combined signal and background model to the data. The lowest panel shows the ratio of the constrained (solid) and unconstrained (dashed) background-only model to data. The bins in this plot reflect the binning used in the fit.

as the small contributions from diboson and $t\bar{t}$ samples. Contributions from W + jets and multijet events are again found to have negligible impact on the results as they do not cause significant shape changes in the m_{jj} spectrum. The nominal analysis makes use of a data-driven correction that is derived in a signal-suppressed *control* region and applied to the background model in the *search* region, thereby improving the modelling of the m_{jj} spectrum, whilst reducing both experimental and theoretical systematic uncertainties associated with the background model. In order to maintain the primary analysis chain, a VBF $H \to WW^*$ control region has been constructed by reversing the jet veto requirement imposed in the VBF $H \to WW^*$ search region. The detector-level m_{jj} spectrum in this VBF $H \to WW^*$ control region is shown in figure 7.3(a) for the combined $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ lepton channels. The lower panel shows the ratio of data to simulation as well as a first order polynomial fit to this ratio[1]. This data-driven correction obtained in the VBF $H \to WW^*$ control region is then directly applied to the background Monte Carlo model in the VBF $H \to WW^*$ search region. A fit to the data is then performed using the signal model and the constrained background model. The normalisation of each model is allowed to float and the signal and background yields are extracted after performing a log-likelihood maximisation [97].

---

[1] The ratio is particularly prone to statistical fluctuations due to limited statistics in the control region. It has been found that the flexibility of a second order polynomial suffices to capture these fluctuations in the fit, resulting in an extremely conservative uncertainty using the pseudoexperiment approach. A first order polynomial is chosen instead as it manages to capture the trend of the ratio, whilst being less prone to the pull of statistical fluctuations.

Table 7.2

The number of background events (N_{bkg}) and electroweak Zjj events (N_{EW}) as predicted by the MC simulation and obtained from a fit to the data. The uncertainties on the MC prediction and the fitted yields reflect the statistical uncertainties in the simulation and the data, respectively. The sum of the fitted signal and background events is also given for convenience.

Predicted N_{bkg}	$22\,898 \pm 151$
Predicted N_{EW}	697 ± 26
Fitted N_{bkg}	$19\,408 \pm 162$
Fitted N_{EW}	887 ± 88
$N_{\text{bkg}} + N_{\text{EW}}$	$20\,295$

The number of signal (N_{EW}) and background (N_{bkg}) events obtained from the fit are summarised in table 7.2, along with the statistical uncertainties due to the data. In addition, the corresponding number of signal and background events expected by the MC simulation are given with an uncertainty due to the limited statistics in the simulation. The sum of the fitted signal and background events is also given for convenience. The expected number of signal events taken from the `Sherpa` simulation have been normalised to reproduce the next-to-leading order calculation for electroweak Zjj production obtained from `Powheg` in the VBF $H \rightarrow WW^*$ search region. The statistical uncertainty due to the limited number of events in the control region has been estimated using pseudoexperiments as described in section 6.3.1 and is found to be 7.4 %. The extracted and expected signal yield are consistent at the 1.7σ level, accounting only for the uncertainties due to limited statistics in the data, the simulation and the control region.

Figure 7.3(b) shows the detector-level m_{jj} distribution in the VBF $H \rightarrow WW^*$ search region for the combined $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ lepton channels. The simulation is normalised to match the number of events obtained from the fit. The ratio of the constrained background-only model to the data is shown in the lowest panel. The dashed line shows the ratio using the unconstrained background-only model, thus demonstrating that the background-only model struggles to describe the data in the tail of the m_{jj} spectrum regardless of whether the constraint is applied or not. Better agreement is observed, however, between the data and the simulation if the electroweak component is included. The ratio of the combined signal and (constrained) background model to the data is shown in the middle panel.

8

Looking ahead: Electroweak Zjj measurements at the HL-LHC

The Large Hadron Collider (LHC) is the largest machine ever constructed by humans and will remain the most powerful particle accelerator in the world for at least two decades. The European Strategy for particle physics has recently been updated and clearly states that "Europe's top priority should be the exploitation of the full potential of the LHC, including the high-luminosity upgrade of the machine and detectors with a view to collecting ten times more data than in the initial design, by around 2030" [107]. The LHC will have to undergo an extensive upgrade programme from about 2020 in order to achieve this goal and emerge from a decade-long shutdown as the High Luminosity LHC (HL-LHC). Needless to say that such an endeavour ought to be planned carefully. Similar upgrade procedures will be necessary for the LHC detectors, for which the R & D phase has already begun. Possible design choices will mainly be driven by the expected physics potential. One of the many options currently under investigation is an extension of the ATLAS tracking and triggering system out to a range of $|\eta| < 4.0$ in pseudorapidity. The additional tracking information can be used to reduce the impact of jets originating from pileup interactions in the forward region of the detector. The potential damage due to pileup for analyses involving VBF topologies, which are typically characterised by two forward jets, shall be investigated in this chapter using the electroweak Zjj measurement framework.

8.1 Expected event yields for the HL-LHC scenario

Figure 8.1 shows the particle-level cross section as a function of m_{jj} for electroweak and strong Zjj production at centre-of-mass energies of $\sqrt{s} = 8$ TeV and $\sqrt{s} = 14$ TeV. All cross section predictions have been calculated with the `Sherpa` event generator. It can be seen that both cross sections are expected to increase at the higher centre-of-mass energy, however, the cross section for the electroweak contribution is expected to grow much more rapidly as a function of m_{jj} than the cross section for strong Zjj
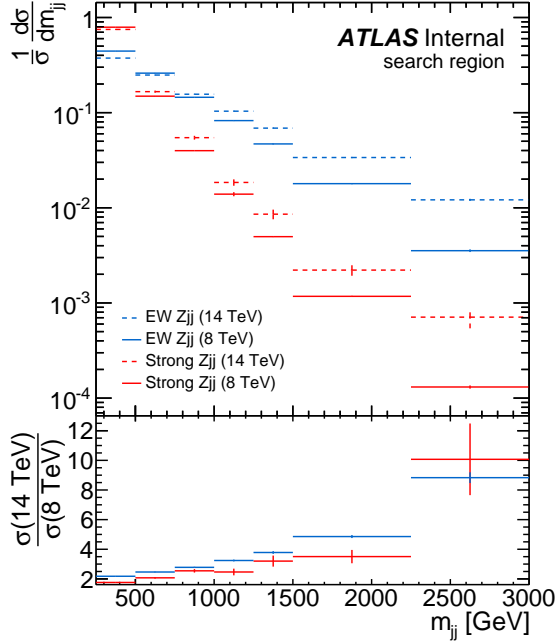
Figure 8.1
Particle-level shape comparisons of the $\frac{1}{\sigma} \cdot \frac{d\sigma}{dm_{jj}}$ distribution in the *search* region for electroweak (blue) and strong Zjj production (red). Curves are shown for two different centre-of-mass energies, namely $\sqrt{s} = 8$ TeV (solid) as well as $\sqrt{s} = 14$ TeV (dashed). The bottom panel shows the ratio of the curves between the two different centre-of-mass energies.

production. The bottom panel in figure 8.1 shows the ratio of the cross-section curves between 14 TeV and 8 TeV. At values of m_{jj} around 2 TeV the electroweak cross section is enhanced by a factor of about 5 at $\sqrt{s} = 14$ TeV with respect to the cross section at 8 TeV. Conversely, the strong Zjj cross section only grows by a factor of about 3.5 at similar values of m_{jj} at $\sqrt{s} = 14$ TeV with respect to the cross section at $\sqrt{s} = 8$ TeV.

An estimate of the expected detector-level event yield N_{d} for a centre-of-mass energy of $\sqrt{s} = 14$ TeV at the HL-LHC is given by

$$N_{\mathrm{d}} = \sigma_{\mathrm{d}}^{8\ \mathrm{TeV}} \times \frac{\sigma_{\mathrm{p}}^{14\ \mathrm{TeV}}}{\sigma_{\mathrm{p}}^{8\ \mathrm{TeV}}} \times L_{\mathrm{int}}^{\mathrm{HL\text{-}LHC}} \tag{8.1}$$

where $\sigma_{\mathrm{d}}^{8\ \mathrm{TeV}}$ is the detector-level cross section at 8 TeV, $\sigma_{\mathrm{p}}^{14\ \mathrm{TeV}}/\sigma_{\mathrm{p}}^{8\ \mathrm{TeV}}$ is the enhancement factor obtained from the particle-level cross sections shown in figure 8.1 and $L_{\mathrm{int}}^{\mathrm{HL\text{-}LHC}} = 3\,000\,\mathrm{fb}^{-1}$ is the integrated luminosity expected for the HL-LHC. The expected event yields for electroweak (N_{EW}) and strong (N_{QCD}) Zjj production are shown in table 8.1 for different regions of the m_{jj} spectrum in the *search* region as well as the *baseline* region. For convenience the expected fraction of signal events is also shown. Note that the `Sherpa` simulation of the background has been reweighted with the *control*-region constraint derived in the nominal analysis (cf. section 6.1). It can be seen that Zjj production is dominated by the electroweak component by about 70 % for values of m_{jj} beyond 2 TeV in the *search* region. Even if the constraint on the m_{jj} spectrum is relaxed, the electroweak process still dominates the tail of the distribution, providing an ideal environment for precision measurements. Note that the uncertainties associated with the background modelling (one of the dominant uncertainties for the 8 TeV analysis using the 2012 dataset of the LHC) would be rendered negligible in this signal-dominated HL-LHC scenario. Although they are generally smaller than in the background-suppressed *search* region, the expected signal fractions in the *baseline*

Table 8.1

Expected signal (N_{EW}) and background (N_{QCD}) yields at the HL-LHC for different regions of the m_{jj} spectrum in the *search* and *baseline* regions. The expected signal fraction is also shown for convenience.

	N_{EW}	N_{QCD}	$\dfrac{N_{\text{EW}}}{N_{\text{EW}} + N_{\text{QCD}}}$
search region with $m_{jj} > 1$ TeV	128 149	231 705	35.6 %
search region with $m_{jj} > 2$ TeV	21 910	9 884	68.9 %
baseline region with $m_{jj} > 1$ TeV	170 296	809 904	17.4 %
baseline region with $m_{jj} > 2$ TeV	27 665	45 020	38.1 %

region still evoke the desire to attempt a signal extraction, thereby avoiding complicated observables such as the *jet veto efficiency* or the p_{T}-balancing variable $p_{\text{T}}^{\text{balance}}$, defined in equation (3.1) in section 3.3.1. The *baseline* region is constructed from a Z-boson candidate and at least two jets in the event, where the leading two jets are required to have transverse momenta greater than 55 GeV and 45 GeV, respectively. The background contribution can potentially be further suppressed by raising the p_{T}-requirements imposed on the leading two jets. It is encouraging to see that for values of m_{jj} beyond 1 TeV, the expected signal yield in either region is such that a statistical precision at the subpercent level could be achieved. In any case, one ought to cater for the possibility of studying inclusive Zjj production in an environment that is much simpler and yet dominated by the electroweak component. This would allow for a precision measurement of the interval-jet distribution, which is sensitive to the difference in the t-channel colour flow between electroweak and strong Zjj production. Note that this would not be possible in the *search* region which is defined via a veto on additional jets in the rapidity interval between the two leading jets. Moreover, a precision measurement of inclusive Zjj production could also be used to study potential interference effects between the strong and the electroweak process and help constrain the associated uncertainties.

8.1.1 Impact of pileup events

In order to estimate the impact of pileup jets on the expected event yields, the analysis has been repeated using only those jets in the event that can be matched ($\Delta R < 0.3$) to particle-level jets in η-ϕ space. The pileup component in the event yield is then given by the difference between the nominal event yield N and the event yield N', which uses only matched jets. The fractional difference $\delta N = (N - N')/N'$ is shown for the *search* region in table 8.2 where the expected signal and background yields are considered both separately and combined. The numbers in the upper half reflect the relative pileup contribution at 8 TeV under LHC conditions, which is found to be generally at the level of 2 % or less. Additional (pileup) jets that are mistakenly associated with the primary collision vertex result in an apparent increase in interval-jet activity, such that the corresponding events are likely to fail the jet veto requirement or the p_{T}-balancing condition used to define the *search* region. Hence, signal and background

Table 8.2
Fractional changes in the expected electroweak (N_EW) and strong (N_QCD) Zjj event yields in the *search* region due to jets originating from pileup events. The resulting change in the combined yields ($\delta N_\mathrm{EW+QCD}$) is also shown for convenience. The numbers in the top half represent the LHC conditions observed in 2012, whereas the numbers in the bottom half have been scaled in order to account for the expected increase in pileup at the HL-LHC. Refer to main text for details.

	search region with	δN_EW	δN_QCD	$\delta N_\mathrm{EW+QCD}$
$\sqrt{s} = 8$ TeV, $\langle\mu\rangle \approx 20$	$m_{jj} > 1$ TeV	$-1.4\,\%$	$-1.0\,\%$	$-1.1\,\%$
	$m_{jj} > 2$ TeV	$-1.9\,\%$	$-1.6\,\%$	$-2.4\,\%$
$\sqrt{s} = 14$ TeV, $\langle\mu\rangle \approx 140$	$m_{jj} > 1$ TeV	$-16.1\,\%$	$-11.9\,\%$	$-13.1\,\%$
	$m_{jj} > 2$ TeV	$-21.7\,\%$	$-18.6\,\%$	$-20.5\,\%$

yield are affected in a similar way. The pileup component is then proportional to the probability for inclusive jet production in proton-proton collisions, given by the ratio of the jet production cross section (σ_jets) to the total inelastic cross section (σ_total). The cross section for inclusive jet production increases by a factor of about 1.8 at 14 TeV [108], whereas the total inelastic cross section rises by a factor of about 1.1 at the higher centre-of-mass energy [109]. The numbers in the lower half of table 8.2, showing the expected pileup contribution at 14 TeV under HL-LHC conditions, have therefore been adjusted by a factor of $\frac{\sigma_\mathrm{jets}^{14\,\mathrm{TeV}}}{\sigma_\mathrm{jets}^{8\,\mathrm{TeV}}} / \frac{\sigma_\mathrm{total}^{14\,\mathrm{TeV}}}{\sigma_\mathrm{total}^{8\,\mathrm{TeV}}} \approx 1.8/1.1$ for the expected increase in pileup jets as well as a factor of 7 in order to account for the expected increase in the mean number of proton-proton interactions per bunch crossing, $\langle\mu\rangle$, from about 20 to around 140 under HL-LHC conditions. It can be seen that the detector-level background yield in the *search* region is expected to decrease by about 12 % to 19 % for values of m_{jj} beyond 1 TeV and 2 TeV, respectively, while the expected losses in the detector-level signal yield are up to 4 % larger than that. An extension of the ATLAS tracking system would allow for part of the event losses to be recovered through pileup jet suppression techniques based on the additional tracking information, such as the jet vertex fraction which is currently only defined within the inner detector tracking acceptance (cf. section 3.3.2).

No jet veto is imposed in the *baseline* region and so the impact of additional pileup jets on the electroweak component is essentially negligible. This can be seen in the upper half of table 8.3, which shows the relative pileup contribution at 8 TeV under LHC conditions, considering the expected signal and background yields both separately and combined. The Zjj final state can also be reconstructed in an event if one or both of the jets originate from a different proton-proton collision in the same bunch crossing than the one which produced the Z-boson candidate. This is less of an issue for the signal yield as electroweak Zjj production already contains two jets. Strong Zjj production, however, is due to inclusive Z production which in turn is dominated at the pileup-free particle level by events containing a Z-boson candidate and no additional jets at all. By virtue of the additional pileup jet activity, these events can migrate into the *baseline* region where no p_T-balancing condition or dilepton p_T cut is applied (as it is in the *search* region) in order to suppress most of this contribution. The cross section

for this pileup scenario is approximately given by

$$\sigma_{\text{pileup}} = \sigma_Z \times \frac{\sigma_{jj}}{\sigma_{\text{total}}} \times \langle\mu\rangle \tag{8.2}$$

where σ_Z is the cross section for Z-boson production, $\langle\mu\rangle$ is the average number of proton-proton collisions per bunch crossing, and the probability for dijet production is just the ratio of the dijet production cross section (σ_{jj}) to the total inelastic cross section (σ_{total}). The relative increase of the cross section for pileup jets between the different centre-of-mass energies can then be written as

$$\frac{\sigma_{\text{pileup}}^{14\,\text{TeV}}}{\sigma_{\text{pileup}}^{8\,\text{TeV}}} = \frac{\sigma_Z^{14\,\text{TeV}}}{\sigma_Z^{8\,\text{TeV}}} \times \frac{\sigma_{jj}^{14\,\text{TeV}}}{\sigma_{jj}^{8\,\text{TeV}}} \times \frac{\sigma_{\text{total}}^{8\,\text{TeV}}}{\sigma_{\text{total}}^{14\,\text{TeV}}} \times \frac{\langle\mu\rangle^{14\,\text{TeV}}}{\langle\mu\rangle^{8\,\text{TeV}}} \approx \frac{1.8}{1.1} \times \frac{\sigma_{Zjj}^{14\,\text{TeV}}}{\sigma_{Zjj}^{8\,\text{TeV}}} \times 7 \tag{8.3}$$

assuming that the cross-section enhancement for dijet production is of similar size as that for strong Zjj production (cf. figure 8.1) and using the fact that the cross section for Z production increases by a factor of about 1.8 at 14 TeV [108], whereas the total inelastic cross section rises by a factor of about 1.1 at the higher centre-of-mass energy [109]. The factor of 7 accounts for the expected increase in $\langle\mu\rangle$ from about 20 to around 140 under HL-LHC conditions. The lower half of table 8.3 shows the fractional changes in the event yields expected due to pileup for the *baseline* region at 14 TeV under HL-LHC conditions. The fractional difference δN represents the ratio of the pileup component to the pileup-free event yield and so the nominator has been adjusted according to equation (8.3) for the cross-section enhancement expected for pileup jets, while the denominator has been scaled by the expected increase in cross section for the signal and background process, respectively. It can be seen that the detector-level background yield is nearly matched in size by a pileup contribution, and so an extension of the ATLAS tracking system would be imperative in order to get a handle on this background contribution through improved pileup jet suppression techniques.

Note that the same scaling has been applied to the pileup component for both the signal and background yields, which is likely to be an overestimate for the signal case. This is considered acceptable though as the impact of pileup on the signal yield is essentially negligible in the *baseline* region anyway. In order to better quantify the impact of pileup on the different processes, fully simulated detector-level predictions would be necessary, which would require the equivalent of one billion simulated events at 14 TeV under HL-LHC pileup conditions, so as to match the available statistics at 8 TeV under LHC pileup conditions.

8.2 Foreseeable trends for systematic uncertainties

A summary of the systematic uncertainties associated with the signal extraction in the 8 TeV analysis is shown in table 6.2 in section 6.3. The total measurement uncertainty is around 20 %, but some improvements are expected for future analyses of this type and are outlined in the following. The uncertainty due to limited statistics in the *control*

Table 8.3

Fractional changes in the expected electroweak (N_{EW}) and strong (N_{QCD}) Zjj event yields in the *baseline* region due to jets originating from pileup events. The resulting change in the combined yields ($\delta N_{\mathrm{EW+QCD}}$) is also shown for convenience. The numbers in the top half represent the LHC conditions observed in 2012, whereas the numbers in the bottom half have been scaled in order to account for the expected increase in pileup at the HL-LHC. Refer to main text for details.

	baseline region with	δN_{EW}	δN_{QCD}	$\delta N_{\mathrm{EW+QCD}}$
$\sqrt{s} = 8$ TeV, $\langle \mu \rangle \approx 20$	$m_{jj} > 1$ TeV	$< 0.1\,\%$	$8.6\,\%$	$7.3\,\%$
	$m_{jj} > 2$ TeV	$< 0.1\,\%$	$2.4\,\%$	$1.6\,\%$
$\sqrt{s} = 14$ TeV, $\langle \mu \rangle \approx 140$	$m_{jj} > 1$ TeV	$0.4\,\%$	$98.7\,\%$	$80.4\,\%$
	$m_{jj} > 2$ TeV	$0.2\,\%$	$27.3\,\%$	$16.8\,\%$

region will be reduced significantly thanks to the high luminosity target of the HL-LHC. The dominant source of experimental uncertainty is the jet energy scale ($5.6\,\%$), but is already known to reduce with an improved η-intercalibration modelling of the jets. The uncertainty due to the theoretical modelling of the signal process can be improved by employing a next-to-leading order calculation, whereas in the 8 TeV analysis the available detector-level simulation only had leading order accuracy. As already mentioned, the background component is suppressed in the tail of the m_{jj} spectrum and hence the corresponding uncertainty associated with the theoretical modelling of the strong Zjj process will be much less of an issue in the HL-LHC scenario. It is not clear, however, whether similar improvements can be expected for the theoretical uncertainty regarding a possible interference between electroweak and strong Zjj production. A conservative systematic uncertainty of $6.2\,\%$ has been assigned in the 8 TeV analysis, but it is conceivable that the provided data can be used to study the interference, such that a better understanding can help constrain this source of systematic uncertainty.

I may not have gone where I intended to go,
but I think I have ended up where I intended to be.

— Douglas Adams

9

Summary

A detailed study of dijets produced in association with a Z boson has been presented using $20.3\,\mathrm{fb}^{-1}$ of proton-proton collision data collected by the ATLAS experiment at the LHC in 2012. Cross sections and differential distributions of inclusive Zjj production have been measured in five fiducial regions. The measured cross sections are in good agreement with the Standard Model prediction for Zjj production obtained from the `Powheg` event generator. The differential distributions are sensitive to the electroweak component of Zjj production and confront the modelling of strong Zjj production in extreme regions of phase space. The data have been fully corrected for detector effects and are compared to predictions from the `Sherpa` and `Powheg` event generators. Neither prediction is able to fully reproduce the data for all distributions and so the data can be used to constrain the theoretical modelling in the extreme regions of phase space probed in this measurement. The electroweak component of Zjj production has then been extracted by fitting the invariant mass distribution of the dijet system and fiducial cross sections for electroweak Zjj production have been measured in two fiducial regions and fully corrected for detector effects. The background-only model has been rejected above the 5σ level and the presented measurements therefore constitute the first observation of the electroweak Zjj process (or any VBF-like process for that matter). The measured cross sections are in good agreement with the Standard Model prediction obtained from `Powheg` and have also been used to constrain new physics by setting limits on anomalous triple gauge couplings. Furthermore, the `Powheg` method has been probed for jet selections that have been optimised with respect to VBF Higgs production and found to be in good agreement with the extracted Zjj yield in these regions as well. Moreover, the potential for future electroweak Zjj measurements at the HL-LHC has been investigated and found to be feasible in much simpler regions of phase space, provided the ATLAS tracking system is extended to allow for an improved suppression of pileup jets. In addition, a performance study of the inner detector tracking algorithms used in the ATLAS high-level trigger system is presented using the tag-and-probe approach for both $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ candidate events which shows excellent tracking efficiencies across the inner detector tracking acceptance.

The covers of this book are too far apart.
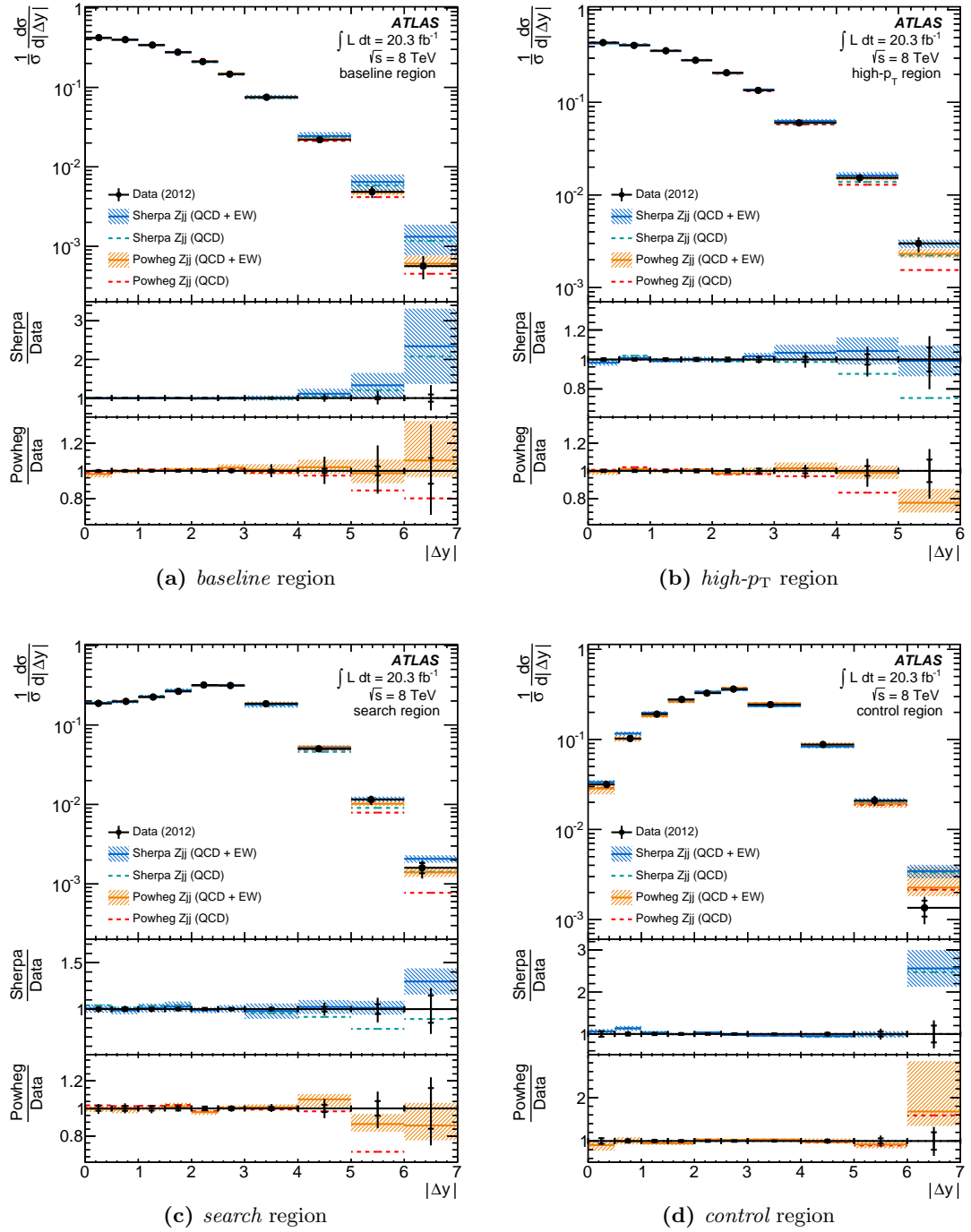
— Ambrose Gwinett Bierce

A

Full set of unfolded differential distributions
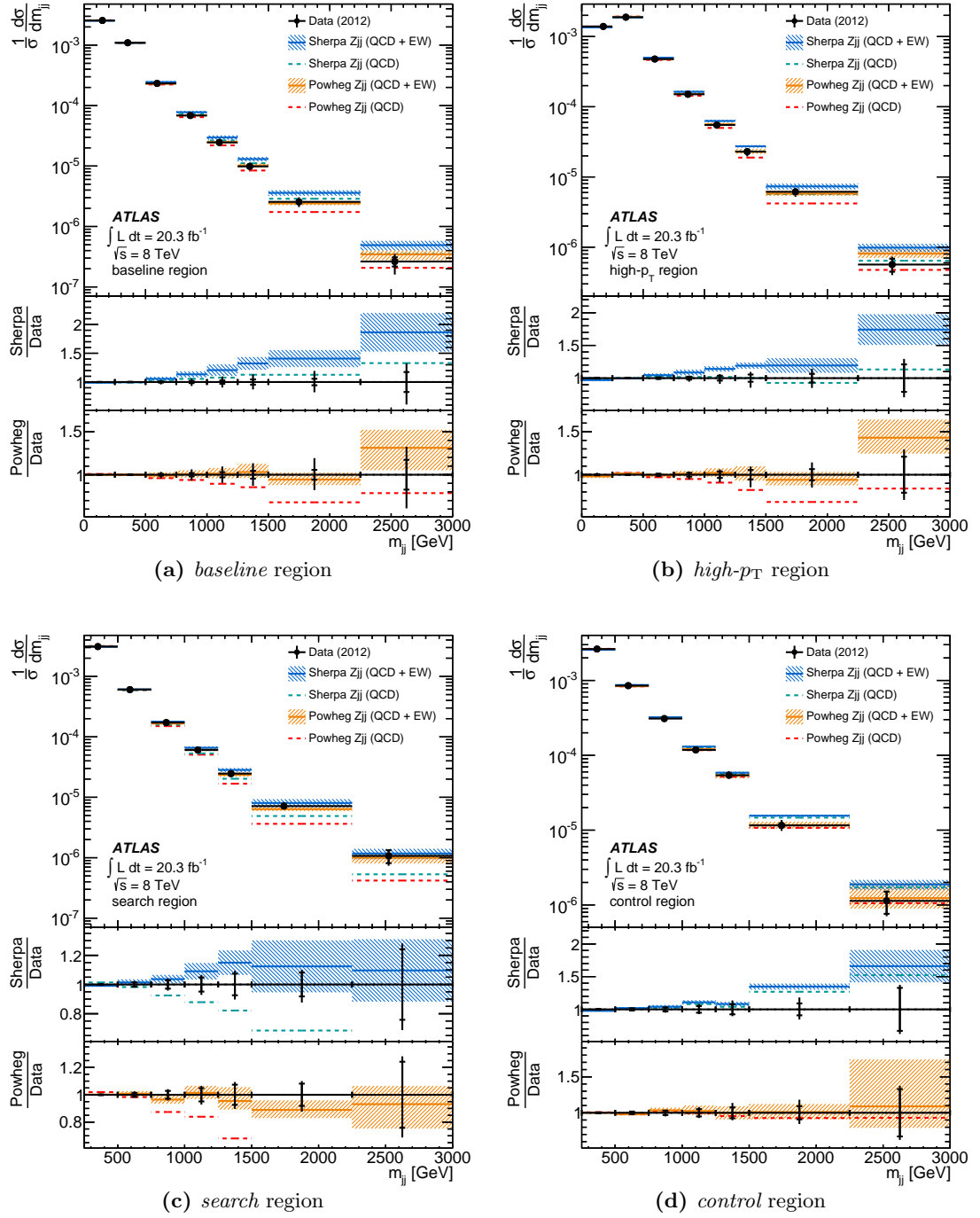
Figure A.1

Unfolded $\frac{1}{\sigma} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}m_{jj}}$ distribution in the (a) *baseline*, (b) *high-p*$_T$ (c) *search* and (d) *control* region. The data are shown as filled (black) circles. The vertical error bars show the size of the total uncertainty on the measurement, with tick marks used to reflect the size of the statistical uncertainty only. Particle-level predictions from `Sherpa` and `Powheg` are shown for combined strong and electroweak Zjj production (labelled as QCD + EW) by hatched bands, denoting the model uncertainty, around the central prediction, which is shown as a solid line. The predictions from `Sherpa` and `Powheg` for strong Zjj production (labelled QCD) are shown as dashed lines.

Figure A.2
Unfolded $\frac{1}{\sigma} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}|\Delta y|}$ distribution in the (a) *baseline*, (b) *high-p*$_\mathrm{T}$ (c) *search* and (d) *control* region. The data are shown as filled (black) circles. The vertical error bars show the size of the total uncertainty on the measurement, with tick marks used to reflect the size of the statistical uncertainty only. Particle-level predictions from `Sherpa` and `Powheg` are shown for combined strong and electroweak Zjj production (labelled as QCD + EW) by hatched bands, denoting the model uncertainty, around the central prediction, which is shown as a solid line. The predictions from `Sherpa` and `Powheg` for strong Zjj production (labelled QCD) are shown as dashed lines.
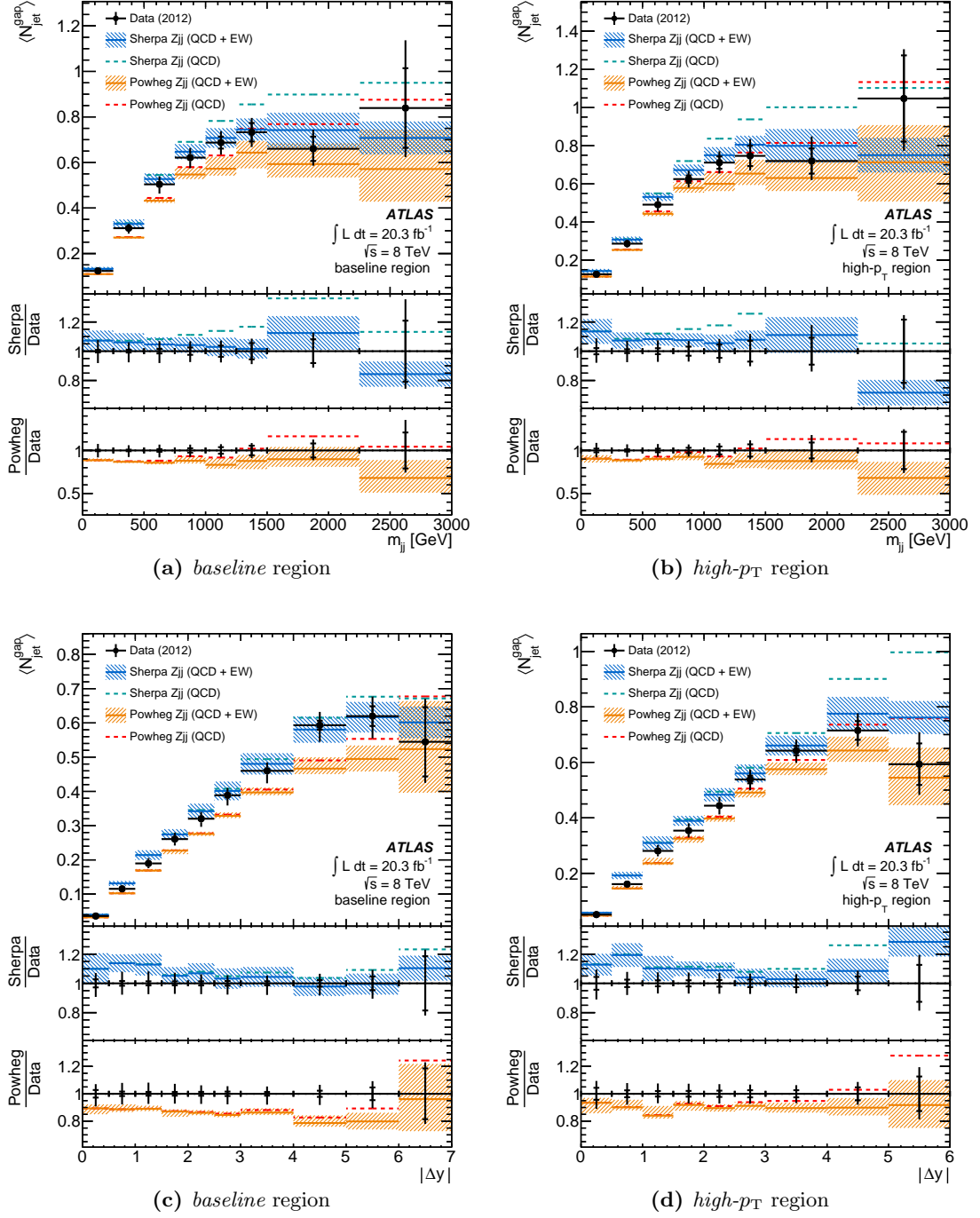
(a) $\frac{1}{\sigma} \cdot \frac{d\sigma}{dN_{jet}^{gap}}$

(b) $\frac{1}{\sigma} \cdot \frac{d\sigma}{dp_T^{balance}}$

(c) $\frac{1}{\sigma} \cdot \frac{d\sigma}{d|\Delta\phi(j,j)|}$

Figure A.3
Unfolded (a) $\frac{1}{\sigma} \cdot \frac{d\sigma}{dN_{jet}^{gap}}$, (b) $\frac{1}{\sigma} \cdot \frac{d\sigma}{dp_T^{balance}}$ and (c) $\frac{1}{\sigma} \cdot \frac{d\sigma}{d|\Delta\phi(j,j)|}$ distributions in the *high-mass* region. The data are shown as filled (black) circles. The vertical error bars show the size of the total uncertainty on the measurement, with tick marks used to reflect the size of the statistical uncertainty only. Particle-level predictions from Sherpa and Powheg are shown for combined strong and electroweak Zjj production (labelled as QCD + EW) by hatched bands, denoting the model uncertainty, around the central prediction, which is shown as a solid line. The predictions from Sherpa and Powheg for strong Zjj production (labelled QCD) are shown as dashed lines.
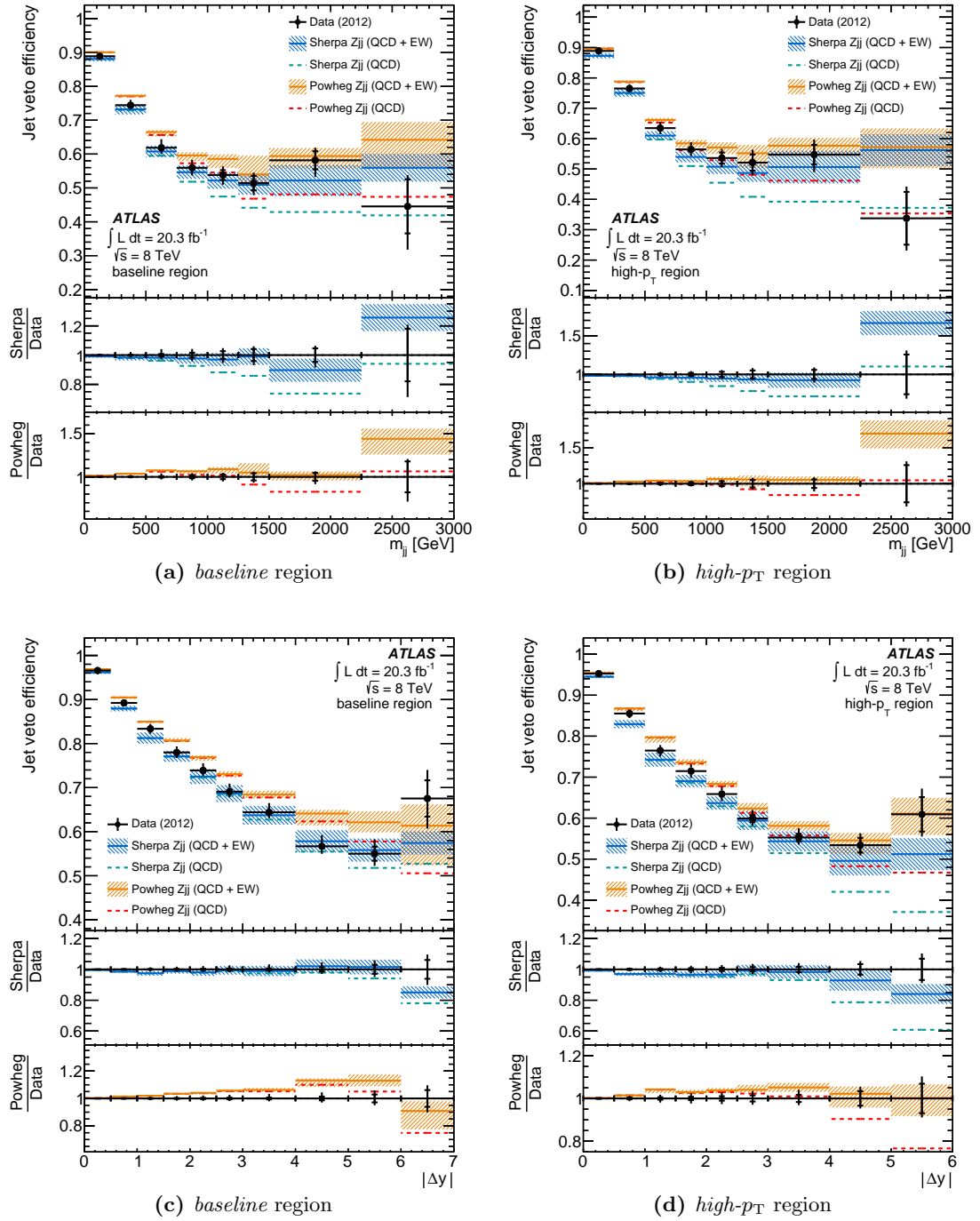
(a) *baseline* region

(b) *high-p*$_T$ region

(c) *baseline* region

(d) *high-p*$_T$ region

Figure A.4

Unfolded $\langle N_{\mathrm{jet}}^{\mathrm{gap}} \rangle$ distributions as a function of m_{jj} in the (a) *baseline* and (b) *high-p*$_T$ region and as a function of $|\Delta y|$ in the (c) *baseline* and (d) *high-p*$_T$ region. The data are shown as filled (black) circles. The vertical error bars show the size of the total uncertainty on the measurement, with tick marks used to reflect the size of the statistical uncertainty only. Particle-level predictions from `Sherpa` and `Powheg` are shown for combined strong and electroweak Zjj production (labelled as QCD + EW) by hatched bands, denoting the model uncertainty, around the central prediction, which is shown as a solid line. The predictions from `Sherpa` and `Powheg` for strong Zjj production (labelled QCD) are shown as dashed lines.
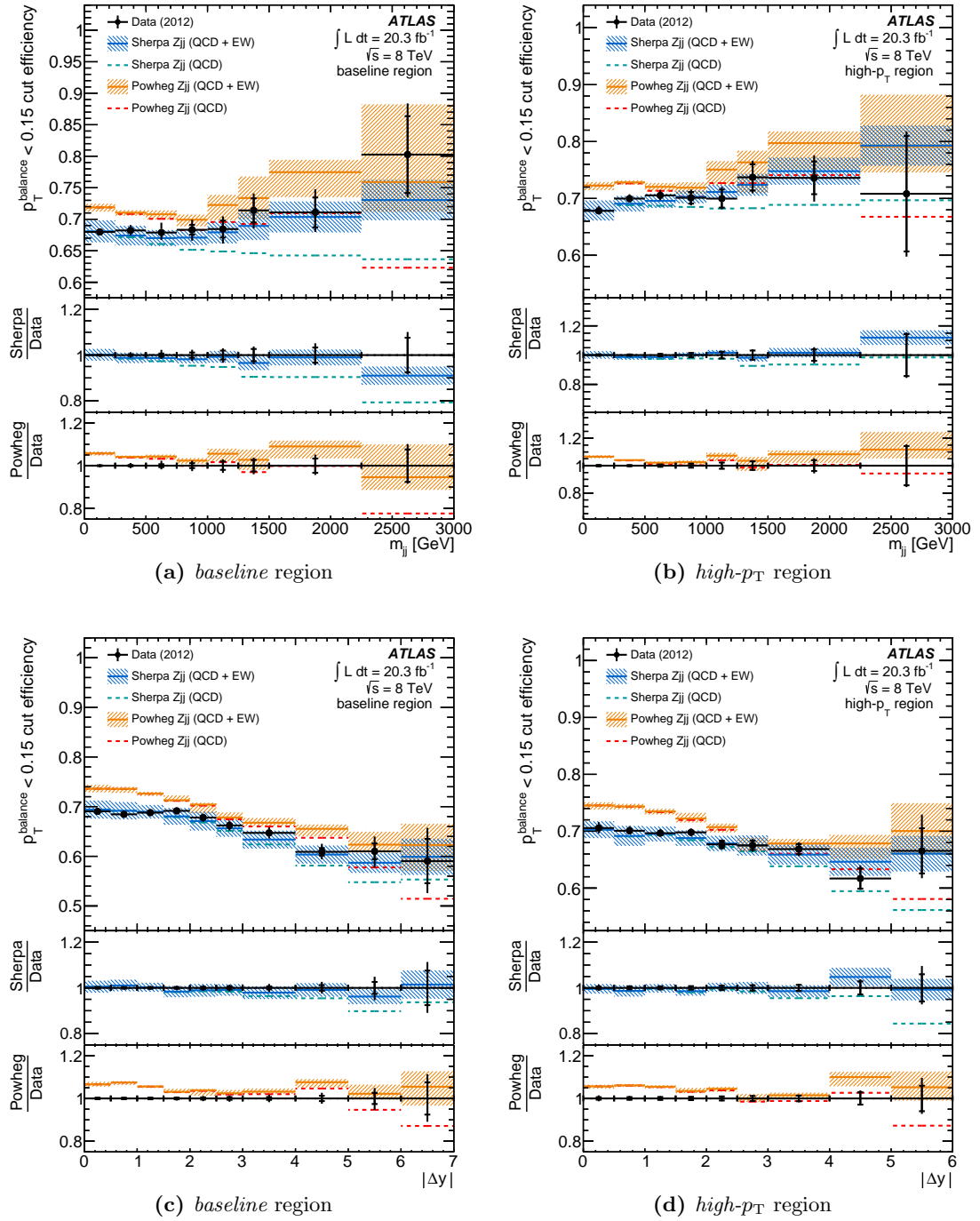
(a) *baseline* region

(b) *high-p_T* region

(c) *baseline* region

(d) *high-p_T* region

Figure A.5

Unfolded *jet veto efficiency* as a function of m_{jj} in the (a) *baseline* and (b) *high-p_T* region and as a function of $|\Delta y|$ in the (c) *baseline* and (d) *high-p_T* region. The data are shown as filled (black) circles. The vertical error bars show the size of the total uncertainty on the measurement, with tick marks used to reflect the size of the statistical uncertainty only. Particle-level predictions from Sherpa and Powheg are shown for combined strong and electroweak Zjj production (labelled as QCD + EW) by hatched bands, denoting the model uncertainty, around the central prediction, which is shown as a solid line. The predictions from Sherpa and Powheg for strong Zjj production (labelled QCD) are shown as dashed lines.

(a) *baseline* region

(b) *high-p*$_\text{T}$ region

(c) *baseline* region

(d) *high-p*$_\text{T}$ region

Figure A.6

Unfolded $p_\text{T}^\text{balance}$ *cut efficiency* as a function of m_{jj} in the (a) *baseline* and (b) *high-p*$_\text{T}$ region and as a function of $|\Delta y|$ in the (c) *baseline* and (d) *high-p*$_\text{T}$ region. The data are shown as filled (black) circles. The vertical error bars show the size of the total uncertainty on the measurement, with tick marks used to reflect the size of the statistical uncertainty only. Particle-level predictions from `Sherpa` and `Powheg` are shown for combined strong and electroweak Zjj production (labelled as QCD + EW) by hatched bands, denoting the model uncertainty, around the central prediction, which is shown as a solid line. The predictions from `Sherpa` and `Powheg` for strong Zjj production (labelled QCD) are shown as dashed lines.

Before I came here I was confused about this subject.
Having listened to your lecture I am still confused.
But on a higher level.

— Enrico Fermi

B

Anomalous triple gauge couplings

The dynamics of any system are fully described by its Lagrangian. The effective Lagrangian density $\mathcal{L}$ describing the most general coupling between two charged gauge bosons and a neutral gauge boson [110] is given by

$$\frac{\mathcal{L}_{WWV}}{ig_{WWV}} = g_1^V \left(W_{\mu\nu}^\dagger W^\mu V^\nu - W_\mu^\dagger V_\nu W^{\mu\nu} \right) \tag{B.1}$$

$$+ \kappa_V W_\mu^\dagger W_\nu V^{\mu\nu} \tag{B.2}$$

$$+ \frac{\lambda_V}{m_W^2} W_{\rho\mu}^\dagger W^\mu{}_\nu V^{\nu\rho} \tag{B.3}$$

$$+ i g_4^V W_\mu^\dagger W_\nu \left(\partial^\mu V^\nu + \partial^\nu V^\mu \right) \tag{B.4}$$

$$- i g_5^V \epsilon^{\mu\nu\rho\sigma} \left(W_\mu^\dagger (\partial_\rho W_\nu) - (\partial_\rho W_\mu^\dagger) W_\nu \right) V_\sigma \tag{B.5}$$

$$+ \frac{\tilde{\kappa}_V}{2} W_\mu^\dagger W_\nu \epsilon^{\mu\nu\rho\sigma} V_{\rho\sigma} \tag{B.6}$$

$$+ \frac{\tilde{\lambda}_V}{2m_W^2} W_{\rho\mu}^\dagger W^\mu{}_\nu \epsilon^{\nu\rho\sigma\mu} V_{\sigma\mu} \tag{B.7}$$

where $V_\mu = V_\mu^\dagger$ can be either the photon or the Z-boson field, W_μ is the W-boson field and $X_{\mu\nu} = \partial_\mu X_\nu - \partial_\nu X_\mu$ (X being either V or W). The constant g_{WWV} takes the values $g_{WWZ} = -|e| \cot \theta_{\mathrm{w}}$ and $g_{WW\gamma} = -|e|$, where $|e|$ is the electric charge of the positron and θ_{w} is the Weinberg angle. The terms (B.4) and (B.5) violate charge conjugation, while parity conservation is violated by the terms (B.5), (B.6) and (B.7). The Standard Model value for the coupling parameters in each of these terms is therefore set to zero, which greatly simplifies the expression to give

$$\frac{\mathcal{L}_{WWZ}}{ig_{WWZ}} = g_1^Z \left(W_{\mu\nu}^\dagger W^\mu Z^\nu - W_\mu^\dagger Z_\nu W^{\mu\nu} \right) + \kappa_Z W_\mu^\dagger W_\nu Z^{\mu\nu} + \frac{\lambda_Z}{m_W^2} W_{\rho\mu}^\dagger W^\mu{}_\nu Z^{\nu\rho} \tag{B.8}$$

where only the terms related to the WWZ coupling have been retained. The remaining coupling parameters have the values $g_1^Z = \kappa_Z = 1$ and $\lambda_Z = 0$ within the Standard Model. When taking into account higher-order terms, the constant coupling parameters lead to unphysically large cross sections at high energies, thereby violating unitarity. Unitarity is restored by form factor effects which ensure that the anomalous couplings will fall off beyond the scale of new physics. Typically the coupling parameters are modified by a dipole form factor of the form

$$g(\hat{s}) = \frac{g_0}{\left(1 + \hat{s}/\Lambda^2\right)^2} \tag{B.9}$$

where g_0 is the bare coupling, $\hat{s}$ is the partonic centre-of-mass energy and Λ is the unitarisation scale, respresenting the energy scale of new physics that could be somehow modifying the WWZ coupling. A unitarisation scale of around 6 TeV is the largest common value for the remaining WWZ coupling parameters that is accessible at the LHC [111, 112]. Finally, note that since the VBF Zjj diagram involves a momentum transfer via the t-channel exchange of a W boson, the vector boson fusion process will be most sensitive to the terms in $\mathcal{L}_{WWZ}$ that contain a derivative of the W-boson field (i.e. the ones associated with the coupling parameters g_1^Z and λ_Z).

Bibliography

[1] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588

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

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

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

[5] G. S. Guralnik et al., *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587

[6] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156–1163

[7] T. W. B. Kibble, *Symmetry breaking in non-Abelian gauge theories*, Phys. Rev. **155** (1967) 1554–1561

[8] 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. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex]

[9] CMS Collaboration, *Observation of a new boson at a mass of* 125 GeV *with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex]

[10] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367–377

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

[12] `claymath.org/millenium-problems/yang%E2%80%93mills-and-mass-gap`

[13] ATLAS Collaboration, *Measurement of the electroweak production of dijets in association with a Z-boson and distributions sensitive to vector boson fusion in proton-proton collisions at* $\sqrt{s} = 8$ TeV *using the ATLAS detector*, JHEP **1404** (2014) 031, arXiv:1401.7610 [hep-ex]

[14] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003

[15] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment – Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex]

[16] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J. **C70** (2010) 787–821, arXiv:1004.5293 [physics.ins-det]

[17] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector and its Performance in 2012*, ATLAS-CONF-2014-047

[18] J. P. Archambault et al., *Performance of the ATLAS liquid argon forward calorimeter in beam tests*, JINST **8** (2013) P05006

[19] `twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`

[20] `twiki.cern.ch/twiki/bin/view/AtlasPublic/HLTTrackingPublicResults`

[21] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. **C72** (2012) 1909, arXiv:1110.3174 [hep-ex]

[22] ATLAS Collaboration, *Measurement of the $W \to \ell\nu$ and $Z/\gamma^* \to \ell\ell$ production cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **1012** (2010) 060, arXiv:1010.2130 [hep-ex]

[23] C. Oleari and D. Zeppenfeld, *QCD corrections to electroweak $\ell\nu_\ell jj$ and $\ell^+\ell^- jj$ production*, Phys. Rev. **D69** (2004) 093004, arXiv:hep-ph/0310156

[24] K. Arnold et al., *VBFNLO: A parton level Monte Carlo for processes with electroweak bosons – Manual for Version 2.6.0*, (2012), arXiv:1107.4038 [hep-ph]

[25] J. C. Collins et al., *Soft Gluons and Factorization*, Nucl. Phys. **B308** (1988) 833

[26] J. C. Collins et al., *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1988), arXiv:hep-ph/0409313

[27] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, JHEP **1212** (2012) 061, arXiv:1209.6215 [hep-ph]

[28] S. Höche et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **1304** (2013) 027, arXiv:1207.5030 [hep-ph]

[29] L. Lönnblad and S. Prestel, *Merging Multi-leg NLO Matrix Elements with Parton Showers*, JHEP **1303** (2013) 166, arXiv:1211.7278 [hep-ph]

[30] S. Plätzer, *Controlling inclusive cross sections in parton shower + matrix element merging*, JHEP **1308** (2013) 114, arXiv:1211.5467 [hep-ph]

[31] S. Höche et al., *Drell-Yan lepton pair production at NNLO QCD with parton showers*, arXiv:1405.3607 [hep-ph]

[32] S. Höche et al., *Higgs-boson production through gluon fusion at NNLO QCD with parton showers*, arXiv:1407.3773 [hep-ph]

[33] K. Hamilton et al., *Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, JHEP **1305** (2013) 082, arXiv:1212.4504 [hep-ph]

[34] K. Hamilton et al., *NNLOPS simulation of Higgs boson production*, JHEP **1310** (2013) 222, arXiv:1309.0017 [hep-ph]

[35] A. Karlberg et al., *NNLOPS accurate Drell-Yan production*, arXiv:1407.2940 [hep-ph]

[36] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. **B238** (1984) 492

[37] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rev. **97** (1983) 31–145

[38] S. Catani et al., *Longitudinally-invariant $K_\perp$-clustering algorithms for hadron-hadron collisions*, Nucl. Phys. **B406** (1993) 187–224

[39] S. D. Ellis and D. E. Soper, *Successive Combination Jet Algorithm For Hadron Collisions* Phys. Rev. **D48** (1993) 3160–3166, arXiv:hep-ph/9305266

[40] M. Cacciari et al., *The anti-k_t jet clustering algorithm*, JHEP **0804** (2008) 063, arXiv:0802.1189 [hep-ph]

[41] Y. L. Dokshitzer et al., *Better Jet Clustering Algorithms*, JHEP **9708** (1997) 001, arXiv:hep-ph/9707323

[42] G. P. Salam and G. Soyez, *A practical Seedless Infrared-Safe Cone jet algorithm*, JHEP **0705** (2007) 086, arXiv:0704.0292 [hep-ph]

[43] J. M. Campbell et al., *W and Z bosons in association with two jets using the* Powheg *method*, JHEP **1308** (2013) 005, arXiv:1303.5447 [hep-ph]

[44] B. Jäger et al., *Next-to-leading order QCD corrections to electroweak Zjj production in the* Powheg Box, JHEP **1209** (2012) 083, arXiv:1207.2626 [hep-ph]

[45] F. Schissler and D. Zeppenfeld, *Parton Shower Effects on W and Z Production via Vector Boson Fusion at NLO QCD*, JHEP **1304** (2013) 057, arXiv:1302.2884 [hep-ph]

[46] E. Re, *NLO corrections merged with parton showers for Z + 2 jets production using the* Powheg *method*, JHEP **1210** (2012) 031, arXiv:1204.5433 [hep-ph]

[47] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv:hep-ph/0409146

[48] S. Frixione et al., *Matching NLO QCD computations with Parton Shower simulations: the* Powheg *method*, JHEP **0711** (2007) 070, arXiv:0709.2092 [hep-ph]

[49] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the* `Powheg Box`, JHEP **1006** (2010) 043, arXiv:1002.2581 [hep-ph]

[50] T. Sjöstrand et al., *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026, arXiv:hep-ph/0603175

[51] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph]

[52] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018, arXiv:1005.3457 [hep-ph]

[53] K. Hamilton et al., *MINLO: Multi-scale improved NLO*, JHEP **1210** (2012) 155, arXiv:1206.3572 [hep-ph]

[54] T. Gleisberg et al., *Event generation with* `Sherpa` *1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph]

[55] S. Catani et al., *QCD Matrix Elements + Parton Showers*, JHEP **0111** (2001) 063, arXiv:hep-ph/0109231

[56] S. Höche et al., *NLO matrix elements and truncated showers*, JHEP **1108** (2011) 123, arXiv:1009.1127 [hep-ph]

[57] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **0206** (2002) 029, arXiv:hep-ph/0204244

[58] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, JHEP **0101** (2001) 010, arXiv:hep-ph/0011363

[59] G. Corcella et al., *HERWIG 6.5 Release Note*, arXiv:hep-ph/0210213

[60] J. M. Butterworth et al., *Multiparton Interactions in Photoproduction at HERA*, Z. Phys. **C72** (1996) 637–646, arXiv:hep-ph/9601371

[61] ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, ATL-PHYS-PUB-2011-008

[62] M. Czakon and A. Mitov, *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*, arXiv:1112.5675 [hep-ph]

[63] S. Agostinelli et al., `GEANT4`: *a simulation toolkit*, Nucl. Inst. Meth. **506** (2003) 250–303, geant4.cern.ch

[64] J. Allison et al., `GEANT4` *Developments and Applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270–278, geant4.cern.ch

[65] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823–874, arXiv:1005.4568 [physics.ins-det]

[66] T. Sjöstrand et al., *A Brief Introduction to PYTHIA 8.1*, Comp. Phys. Comm. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph]

[67] ATLAS Collaboration, *Further ATLAS tunes of PYTHIA6 and Pythia 8*, ATL-PHYS-PUB-2011-014

[68] A. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph]

[69] S. Höche et al. *Hard photon production and matrix-element parton-shower merging*, Phys. Rev. **D81** (2010) 034026, arXiv:0912.3501 [hep-ph]

[70] S. Höche et al., *A parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **0803** (2008) 038, arXiv:0709.1027 [hep-ph]

[71] S. Höche et al., *Uncertainties in NLO + parton shower matched simulations of inclusive jet and dijet production*, Phys. Rev. **D86** (2012) 094042, arXiv:1208.2815 [hep-ph]

[72] CMS Collaboration, *Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV*, JHEP **1109** (2011) 109, arXiv:1107.0330 [hep-ex]

[73] J. R. Andersen and J. M. Smillie, *QCD and Electroweak Interference in Higgs production by Gauge Boson Fusion*, Phys. Rev. **D75** (2007) 037301, arXiv:hep-ph/0611281

[74] M. Ciccolini et al., *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC*, Phys. Rev. **D77** (2008) 013002, arXiv:0710.4749 [hep-ph]

[75] J. R. Andersen et al., *Loop induced interference effects in Higgs Boson plus two jet production at the LHC*, JHEP **0802** (2008) 057, arXiv:0709.3513 [hep-ph]

[76] A. Bredenstein et al., *Mixed QCD-electroweak contributions to Higgs-plus-dijet production at the LHC*, Phys. Rev. **D77** (2008) 073004, arXiv:0801.4231 [hep-ph]

[77] ATLAS Collaboration, *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002

[78] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083

[79] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. **C73** (2013) 2304, arXiv:1112.6426 [hep-ex]

[80] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS 2011 data*, ATLAS-CONF-2013-004

[81] M. Cacciari et al., *The Catchment Area of Jets*, JHEP **0804** (2008) 005, arXiv:0802.1188 [hep-ph]

[82] ATLAS Collaboration, *Measurement of the production cross section of jets in association with a Z boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **07** (2013) 032, arXiv:1304.7098 [hep-ex]

[83] ATLAS Collaboration, *Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data*, (2013), ATLAS-CONF-2013-088

[84] ATLAS Collaboration, *Measurements of the pseudorapidity dependence of the total transverse energy in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS*, JHEP **11** (2012), arXiv:1208.6256 [hep-ex]

[85] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C73** (2013) 2518, arXiv:1302.4393 [hep-ex]

[86] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph]

[87] `hepdata.cedar.ac.uk`

[88] T. Plehn et al., *Determining the structure of Higgs couplings at the LHC*, Phys. Rev. Lett. **88** (2002) 051801, arXiv:hep-ph/0105325

[89] G. Klamke and D. Zeppenfeld, *Higgs plus two jet production via gluon fusion as a signal at the CERN LHC*, JHEP **04** (2007) 052, arXiv:hep-ph/0703202

[90] J. R. Andersen et al., *Azimuthal Angle Correlations for Higgs Boson plus Multi-Jet Events*, JHEP **06** (2010) 091, arXiv:1001.3822 [hep-ph]

[91] D0 Collaboration, *Studies of W boson plus jets production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. **D88** (2013) 092001, arXiv:1302.6508 [hep-ex]

[92] S. Höche and M. Schönherr, *Uncertainties in next-to-leading order + parton shower matched simulations of inclusive jet and dijet production*, Phys. Rev. **D86** (2012) 094042, arXiv:1208.2815 [hep-ph]

[93] ATLAS Collaboration, *Measurement of dijet production with a veto on additional central jet activity in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **09** (2011) 053, arXiv:1107.1641 [hep-ex]

[94] G. D'Agostini, *A Multidimensional unfolding method based on Bayes theorem*, Nucl. Inst. Meth. **A362** (1995) 487

[95] T. Adye, *Unfolding algorithms and tests using RooUnfold*, Proceedings of the PHYSTAT 2011 Workshop, CERN, CERN-2011-006.313, arXiv:1105.1160 [physics.data-an]

[96] K. Hayes et al., *Application of the Bootstrap Statistical Method to the Tau Decay Mode Problem*, Phys. Rev. **D39** (1989) 274

[97] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf C0303241 (2003) MOLT007, arXiv:physics/0306116 [physics.data-an]

[98] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244–289, arXiv:1207.1303 [hep-ph]

[99] J. M. Campbell et al., *Hard Interactions of Quarks and Gluons: a Primer for LHC Physics*, Rept. Prog. Phys. **70** (2007) 89, arXiv:hep-ph/0611148

[100] A. Martin et al., *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. **C64** (2009) 653–680, arXiv:0905.3531 [hep-ph]

[101] `atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/`

[102] U. Baur and D. Zeppenfeld, *Measuring Three Vector Boson Couplings in $qq \rightarrow qqW$ at the SSC*, Conf. Proc. Argonne Accel. Phys. **C93-06-02** (1993) 327–334, arXiv:hep-ph/9309227

[103] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, arXiv:1007.1727 [physics.data-an]

[104] ATLAS Collaboration, *Measurement of WZ production in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. **C72** (2012) 2173, arXiv:1208.1390 [hep-ex]

[105] B. Cox et al., *Extracting Higgs boson couplings using a jet veto*, Phys. Lett. **B696** (2011) 87–91, arXiv:1006.0986 [hep-ph]

[106] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in* `Powheg`, JHEP **1002** (2010) 037, arXiv:0911.5299 [hep-ph]

[107] `council.web.cern.ch/council/en/EuropeanStrategy/esc-e-106.pdf`

[108] M. L. Mangano and J. Rojo, *Cross Section Ratios between different CM energies at the LHC: opportunities for precision measurements and BSM sensitivity*, JHEP **1208** (2012) 010, arXiv:1206.3557 [hep-ph]

[109] A. Donnachie and P. V. Landshoff, *Total cross sections*, Phys. Lett. **B296** (1992) 227–232, arXiv:hep-ph/9209205

[110] K. Hagiwara et al., *Probing the Weak Boson Sector in $e^+e^- \rightarrow W^+W^-$*, Nucl. Phys. **B282** (1987) 252

[111] H. Aihara et al., *Anomalous gauge boson interactions*, arXiv:hep-ph/9503425

[112] U. Baur and D. Zeppenfeld, *Unitarity Constraints on the Electroweak Three Vector Boson Vertices*, Phys. Lett. **B201** (1988) 383