



HAL
open science

Search for the electroweak production of $Z\gamma$ pairs and measurement of the differential cross section of the $Z\gamma$ production in association with two jets with the ATLAS experiment at LHC

Olympia Dartsis

► **To cite this version:**

Olympia Dartsis. Search for the electroweak production of $Z\gamma$ pairs and measurement of the differential cross section of the $Z\gamma$ production in association with two jets with the ATLAS experiment at LHC. High Energy Physics - Experiment [hep-ex]. Université Grenoble Alpes, 2019. English. NNT : 2019GREAY085 . tel-03028197

HAL Id: tel-03028197

<https://theses.hal.science/tel-03028197>

Submitted on 27 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Physique Subatomique et Astroparticules**

Arrêté ministériel : 25 mai 2016

Présentée par

Olympia DARTSI

Thèse dirigée par **Lucia DI CIACCIO**, Professeur, USMB

préparée au sein du **Laboratoire d'Annecy de Physique des
Particules**
dans l'**École Doctorale de Physique**

**Recherche de la production électrofaible de
paires $Z\gamma$ et mesure de la section efficace
différentielle de production de $Z\gamma$ en
association avec deux jets dans l'ATLAS
expérience du LHC.**

**Search for the electroweak production of
 $Z\gamma$ pairs and measurement of the
differential cross section of the $Z\gamma$
production in association with two jets
with the ATLAS experiment at LHC.**

Thèse soutenue publiquement le **3 Octobre 2019**,
devant le jury composé de :

Monsieur Philippe SCHWEMLING,

Professeur des Universités, Université Paris-Diderot, Rapporteur

Madame Chariclia PETRIDOU,

Professeur des Universités, Aristotle University of Thessaloniki,
Rapporteur

Madame Evelin MEONI,

Professeur associé, Università di Cosenza, Examineur

Monsieur Francesco POLCI,

Chargé de Recherche, Laboratoire de Physique Nucléaire et de Hautes
Energies, Examineur

Monsieur Giovanni LAMANNA,

Directeur de Recherche, L.A.P.P. Laboratoire d'Annecy Le Vieux de
Physique des Particules, Président de jury



Contents

1	Introduction	3
2	Theoretical Introduction	7
2.1	The Standard Model	7
2.1.1	The theory of electromagnetic interactions - QED	8
2.1.2	The theory of strong interactions - QCD	10
2.1.3	The theory of electroweak interaction	12
2.1.4	Spontaneous electroweak symmetry breaking	13
2.1.5	Electroweak gauge boson self-couplings	15
2.1.6	Electroweak gauge boson scattering	16
2.2	Effective field theory	19
2.3	Anomalous Quartic Gauge Couplings - aQGCs	20
2.4	Proton-proton interactions at the LHC	22
2.4.1	Factorisation theorem	23
2.4.2	Parton distribution functions	23
2.4.3	Hadronization and fragmentation	25
3	Experimental Apparatus	27
3.1	The Large Hadron Collider (LHC)	27
3.1.1	Accelerator complex	28
3.1.2	Accelerator parameters	29
3.2	The ATLAS Detector	30
3.2.1	ATLAS Coordinate System	32
3.2.2	The Magnet System	33
3.2.3	Inner Detector (ID)	34
3.2.4	Calorimeters	36
3.2.5	Muon Spectrometer (MS)	40
3.2.6	Trigger System	41
3.3	ATLAS Monte Carlo simulations	42
4	Electron performance studies	43
4.1	Electron trigger	43
4.2	Electromagnetic shower development	44
4.3	Electron and photon reconstruction	44
4.4	Electron identification	45
4.5	Electron isolation	49

4.6	Efficiency measurement methodology	52
4.7	Identification efficiency measurement	53
4.7.1	Tag-and-probe method with $Z \rightarrow ee$ events	54
	Z_{mass} method	56
	Z_{iso} method	57
4.7.2	Tag-and-probe method with $J/\psi \rightarrow ee$ events.	61
4.7.3	Comparison between the Z_{mass} and Z_{iso} results	63
4.7.4	Tag and probe method with $Z \rightarrow ee\gamma$ events.	66
	Motivation	66
	Optimization and event selection	66
	Background estimation and systematic variations	69
4.8	Combined scale factors	72
4.9	Usage of electron selections in physics measurements	77
5	Analysis of $Z\gamma$ electroweak production	79
5.1	Object selection	80
5.1.1	Jet selection	80
5.1.2	Lepton selection	80
5.1.3	Photon selection	82
5.2	Event selection	82
5.3	Background events	84
5.4	Selection of $Z\gamma jj$ signal and control regions	84
5.5	Background estimation	86
5.5.1	Z +jets background	86
	Extrapolation to the signal phase space	89
5.5.2	Initial normalization of the $Z\gamma jj$ -QCD background	90
5.5.3	Evaluation of the $t\bar{t}\gamma$ background	90
5.5.4	Control distributions	91
5.6	Multivariate analysis	97
5.7	Fit procedure	99
5.8	Interference of strong and electroweak production	101
5.8.1	Cross section calculation	101
5.8.2	Effect of the interference	103
5.8.3	Optimization of the kinematic selection	103
5.8.4	Treatment of the interference	106
5.9	Systematic uncertainties	109
5.9.1	Systematic uncertainties from reconstruction and background contribution	109
5.9.2	Theory uncertainties on the predicted shapes	110
5.10	Fit results	111

6	Cross section measurements	113
6.1	Integrated cross section measurement	113
6.1.1	Fiducial cross section extraction	114
6.2	Differential cross section measurements	116
6.2.1	Principle of the unfolding method	116
6.2.2	Overview of unfolding methods	118
6.2.3	Bayesian Iterative	119
	D'Agostini's scheme	120
6.2.4	Differential distributions	121
	Treatment of the background	122
	Systematic uncertainties	123
	Unfolding closure test	124
	$M_{Z\gamma}$ differential distribution	124
	P_T^γ differential distribution	126
	M_{jj} differential distribution	128
	N_{jets} differential distribution	130
7	Conclusion	135
8	Résumé	137
8.1	Introduction	137
8.2	Éléments du Modèle Standard de la physique des particules	138
8.3	Diffusion électrofaible des bosons de jauge dans le modèle standard	139
8.4	Le détecteur ATLAS	140
8.5	Mesure de l'efficacité d'identification des électrons par la méthode Tag-and-Probe.	142
8.6	Analyse de la production électrofaible de $Z\gamma$	143
8.7	Mesures de la section efficace de production de $Z\gamma$	144
8.8	Conclusions	148

Acknowledgements

By reaching the end of my PhD work, a period filled with many ups and downs, I could not but thank all those who have helped, encouraged and supported me all these years.

I would like to start by expressing my gratitude to my supervisor Lucia di Ciaccio, for her guidance all these years. Her deep understanding of the particle physics has been a source of knowledge for me. Also, I appreciate deeply her contribution with the thesis proofreading, which gave the thesis the necessary structure.

I am grateful to Emmanuel Sauvan who was always available to answer my questions and for his generous help in all the technical issues I was facing during these three years. The acknowledgment should be extended to the entire ATLAS team of the LAPP, my host laboratory in Annecy, for their warm welcoming. I would like to thank Louis Helary, Narei Lorenzo and Iro Koletsou with whom I worked in the $Z\gamma$ analysis, for the constructive discussions and the nice working environment. During my work on the electron identification efficiency measurement many people have helped me, Philip Sommer, Eleonora Benhar, Christos Anastopoulos, Kristin Lohwasser, Cyril Pascal Becot, Andreas Petridis, Sarah Heim, Angela Burger and Elias Ruttinger, thank you for your suggestions and help.

I would like to thank the rapporteurs of my thesis, Chariclia Petridou and Philippe Schwemling for their time and careful reading of my thesis and for all their useful suggestion. I am very thankful to Evelin Meoni, Francesco Polci and Giovanni Lamanna for accepting to be in my jury members and for their interest to my thesis work.

I could not forget the ATLAS team members of my previous institute in Greece, AUTH, with whom I started in this field. Thank you (Chariclia Petridou, Kostas Kordas, Dinos Bachas, Dimitris Iliadis, Ioannis Nomidis and Despoina Sampsonidou) for the wonderful experience and time we spend together.

During my PhD, I had the luck to meet and collaborate with other PhD students and post-docs, who made my working time at Lapp a pleasant experience. Thanks Saskia and Peter Falke, Angela Burger, Francesco Constanza, Ana Cueto, Sergii Raspopov, Kiril Grevtsov, Oleh Kivernyk, Louis Portales, Luca Franco and Mohamed Belfkir.

My parents and brothers -Athanasios, Chrisanthi, Christos and Dimos- have contributed discreetly but also profoundly. Being there always present, even from far, by supporting and encouraging me whenever I needed it.

Finally, Daniel, my love, who was undoubtedly my biggest support during these years -thank you for helping me extending my limits and overcome all the difficulties I faced.

Chapter 1

Introduction

The discovery of the Higgs boson at the CERN Large Hadron Collider (LHC) [1, 2] was a major success of the Standard Model (SM). The LHC's high energy and luminosity has offered and offers the opportunity to measure many processes predicted by the SM, some of them not yet observed with previous accelerators. These include vector boson scattering [3, 4, 5, 6], vector boson fusion [7, 8, 9] and multiple gauge bosons production [10, 11].

The main subject of my thesis is the measurement of the cross section of the diboson $Z\gamma$ production in association with a high mass dijet system. Proton-proton collision data from the 2015-2016 ATLAS experiment at the LHC with a center-of-mass energy of $\sqrt{s} = 13$ TeV and an integrated luminosity of 36.1 fb^{-1} , are analyzed. The electroweak production of $Z\gamma jj$ events, provides a direct access to the nature of the electroweak symmetry breaking mechanism, by probing the quantum gauge boson couplings $WWZ\gamma$, $ZZZ\gamma$, $ZZ\gamma\gamma$ and $Z\gamma\gamma\gamma$. The last three couplings are forbidden at the lowest order in the SM. Any deviation from the SM predictions would be a hint at new physics beyond the SM, a realm that could manifest itself in the form of anomalous quartic gauge boson couplings (aQGCs).

Until now, the electroweak production of $Z\gamma jj$ ($Z\gamma jj$ -EW), which involves processes with fourth-order electroweak coupling including vector-boson scattering, has never been observed. ATLAS and CMS have found evidence of the process analyzing Run I data at 8 TeV [12, 13]. The CMS Collaboration has recently confirmed this evidence with Run II data at 13 TeV [14]. The same $Z\gamma jj$ final state can be produced by a strongly-mediated process, with second-order electroweak coupling and second-order strong coupling, resulting in a cross section two orders of magnitude larger than the electroweak cross section. The $Z\gamma jj$ production via the electroweak and strong mechanism interfere, since the initial and final states are the same. The understanding of this effect is an important aspect of this analysis because the ATLAS analysis at 8 TeV had found a hint of a potential significant impact of the interference.

In this thesis, I studied the interference effect and its impact on the observation and on the fiducial cross section measurements of the $Z\gamma jj$ -EW production. Having computed the three contributions (electroweak, strong and interference) to the total cross section by means of Monte Carlo simulations, I performed an optimization of the selection cuts to reduce the impact of the interference even though I found that the effect was smaller than thought in Run I analysis. At the same time, I also

defined a procedure on how to account for the interference in the data analysis. The experimental focus of my work is the measurement of the differential cross section of the $Z\gamma jj$ final state as a function of the invariant mass of the $Z\gamma$ system, the transverse momentum of the photon, the jet multiplicity and the invariant mass of the dijet system. Distributions sensitive to new physics are unfolded and compared with MC expectations.

All physical results rely on the reconstruction and selection of final states consisting of jets, electrons, muons and photons. Electrons and their identification therefore play an important role in the $Z\gamma jj$ analysis. Hence, this thesis also documents my work on data-driven methods for electron identification efficiency. Most importantly, these include the implementation of an additional data-driven method which provides electron identification efficiencies in the range of $E_T = 10-25$ GeV, and the maintenance and update of one of the two main data-driven methods for $E_T = 25-150$ GeV. The results I obtained are used in all ATLAS analyses whose signature contains electrons.

Personal Contribution

In the current chapter, my contribution to the performance and analysis studies presented in this thesis, will be explained. During my PhD, I successfully become an ATLAS member for my work on my qualification task in the Electron Identification group, in charge of measuring the electron efficiency for all ATLAS analyses.

The measurement uses a "tag-and-probe" method which employs events containing well-known resonances decaying to electrons, namely $Z \rightarrow ee$ and $J/\psi \rightarrow ee$. A strict selection on one of the electron candidates (tag) and on the invariant mass of the tag-and-probe electrons allows for the selection of the other electron candidate (probe). The invariant mass of the pair will form a peak near the Z boson or the J/ψ mass, the probe is considered unbiased and used to measure electron identification efficiencies.

For electrons with transverse momentum above 15 GeV, the electron efficiency is measured in $Z \rightarrow ee$ events using two variables to discriminate signal against background: the mass of the tag-probe pair (Z_{mass} method) and the isolation of the probe (Z_{iso} method). For transverse momentum probes between 10-15 GeV a sample of events in which a Z boson is accompanied by a photon issued by a QED final state radiation is used. In this case the invariant mass is formed of three objects, the two electrons and the photon ($Z\gamma$ method). This method allows the extension of the efficiency calculation to electrons with transverse momentum between 10 and 15 GeV by selecting probe electrons that lose their energy due to final state radiation (FSR) reducing the background contamination at the same time. I contributed in the Electron Identification group, producing the electron identification efficiencies for the whole collaboration using the Z_{iso} method, and I ported the analysis code into the new ATLAS release. The $Z\gamma$ method was first implemented by me in the official ATLAS framework. The results of the electron efficiencies were used in the analyses presented at the ATLAS workshop 2017 and at the Moriond 2018.

As concern my work in the $Z\gamma jj$ electroweak analysis, I was responsible for the interference effect measurements. The analysis deals with the study of $Z\gamma$ production in association with a high-mass dijet system. In this analysis final states with a photon and a Z boson decaying into a pair of either electrons or muons are analysed. So far the electroweak production of $Z\gamma jj$ has never been observed. The $Z\gamma jj$ electroweak production, contains processes with fourth-order electroweak coupling and include vector-boson scattering (VBS) processes. The same $Z\gamma jj$ final state can be produced by QCD-mediated processes, with second order electroweak coupling and second order strong coupling. This $Z\gamma jj$ -QCD is the main background of the EW analysis, and two

different approaches are used in ATLAS in order to distinguish the QCD background from the EW signal. The two approaches are:

- BDT approach: using a discriminating variable -which is output of a multivariable analysis- of the signal-likeness of the event, which is fitted to extract the electroweak contribution.
- Cut-based approach: the centrality of the $Z\gamma$ system is used as the sensitive variable which is fitted to extract the electroweak contribution.

In the Run II analysis, I found a small constructive interference ($\sim 4\%$) between the $Z\gamma jj$ -QCD and $Z\gamma jj$ -EW production using SM predictions. One of my tasks was to simulate using Monte Carlo simulations, events of the signal (EW), background (QCD) and the interference between the two, using the MADGRAPH event generator. The goal of my work was to evaluate the size of the interference and to investigate a way to decrease its impact on the results. During my investigation a cut in the $\Delta\eta_{jj}$ was found as the best candidate. To achieve this result I used both the BDT and the Cut-based approach. The last part of my work on the interference effect was to investigate how to treat the interference in the analysis. I showed that the interference can be treated as an uncertainty, which is evaluated from the difference in the shape of the pure EW contribution and a template containing both EW and interference contributions. Moreover I was responsible for producing the control plots for the analysis.

Another aspect of my work, was to measure differential cross sections of the $Z\gamma jj$ final state using a Bayesian iterative unfolding method. To this extend I worked on the unfolding of several distributions as specified below. Unfolding methods are widely used in many analyses, in order to correct experimental measurements for detector effects. The data distributions as well as Monte Carlo simulation for the description of the detector effects, are used as input to the unfolding. The statistical and systematic uncertainties affecting the measurements and/or the simulation are propagated through the unfolding procedure. The systematic uncertainty associated to the unfolding method is also evaluated. The distributions that are unfolded are: the invariant mass of the $Z\gamma$ system ($M_{Z\gamma}$), the transverse momentum of the photon (p_T^γ), the number of jets (N_{jets}) and the invariant mass of the dijet system (M_{jj}). The interest of this work is that the resulting corrected measurements with their uncertainties can be directly compared with the corresponding theoretical predictions as well as with the results from other experiments.

Chapter 2

Theoretical Introduction

Ἡ φύσις μὴδεν μῆτε ἀτελὲς ποιεῖ μῆτε μάτην.

Ἀριστοτέλης 384-322 π.Χ.

The desire to understand the world around us is an innate human drive. Particle physics is at the heart of our understanding of the laws of nature. It studies the fundamental constituents of the universe, the elementary particles, and the interactions between them, the forces. In this chapter, the theory behind the Standard Model of particle physics as well as the concept of Effective Field Theories for searches for new physics beyond the Standard Model are introduced.

2.1 The Standard Model

The Standard Model (SM) [15] of particle physics provides a unified picture of the fundamental particles and of some of their interactions. The SM describes successfully all current experimental data and represents one of the triumphs of modern physics. All the elementary particles and force mediators are mathematically described by fields: scalar for spin 0 particles, bi-spinors for spin 1/2 particles and vectors for spin 1 particles. Additional quantum numbers that characterize the particles are the electric charge, the weak isospin and the color, as it will be explained later. Almost all physical phenomena encountered in everyday life can be described in terms of the electron, electron neutrino, proton and neutron, interacting via the electromagnetic, strong and weak forces. However, a fourth interaction is missing from the SM: gravity, an extremely weak and always attractive force that is responsible for the formation of the large-scale structure of the universe. Possible extensions of the SM, so-called Beyond the Standard Model (BSM) theories, propose solutions that would allow to include gravity into the model.

According to the SM, two types of elementary particles exist in nature, fermions and bosons [16]. Matter consists of point-like fermions with half-integer spin, which interact by exchanging point-like gauge bosons of spin one. There are two subgroups of fermions: the quarks and the leptons. Leptons can be separated into electrically charged particles (e, μ, τ) and neutral charged particles, the neutrinos (ν_e, ν_μ, ν_τ). Quarks (u, d, c, s, t, b) on the other hand, carry electrical and also color charge.

The fermions are categorized into generations or families. The list of the generations of fermions is presented in Figure 2.1. The dynamics of each of the twelve fundamental fermions are described by the Dirac equation of relativistic quantum mechanics. One important consequence of the Dirac equation is that for each of the twelve fermions, an antiparticle state exists with exactly the same mass but opposite charge.

In the SM, particles interact with each other through three fundamental forces: electromagnetic, strong and weak force. Each of them is described by a quantum field theory (QFT). In addition to quarks and leptons, there are twelve bosons of spin one which are the carriers of the electromagnetic, weak and strong interactions. The photon is the carrier of the electromagnetic force, it is massless and interacts with any particle that has an electric charge. The W^\pm and Z bosons are the carriers of the weak interactions and have a mass of approximately 80 GeV and 91 GeV, respectively. The carriers of the strong interactions are the gluons. There are eight massless gluons that interact with particles carrying the strong charge called color. Finally, a spin zero boson, the Higgs particle is present with the role of allowing the introduction of mass terms for the weak gauge bosons, the charged leptons and the quarks without breaking the local gauge invariant.

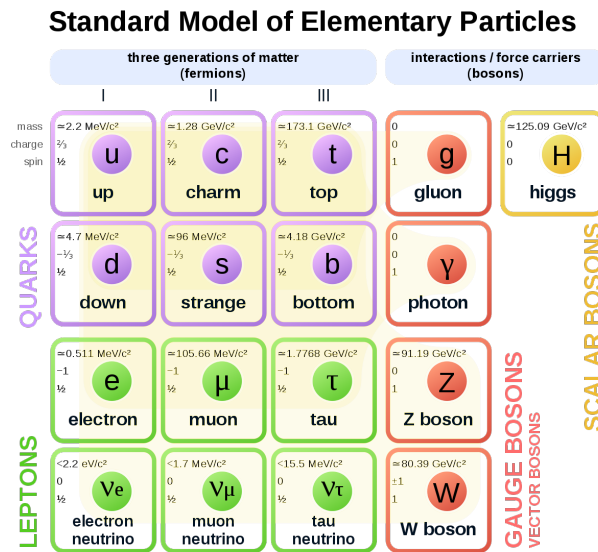


FIGURE 2.1: The fundamental particles of the Standard Model. Some characteristics of fermions and bosons (mass, charge and spin) are illustrated.

2.1.1 The theory of electromagnetic interactions - QED

Quantum electrodynamics (QED) is a relativistic QFT that describes the electromagnetic interactions of particles. It is the first theory that achieves a full integration of quantum mechanics and special relativity. QED describes mathematically all phenomena involving electrically charged particles interacting by exchanging photons and

represents the quantum counterpart of classical electromagnetism. QED provides extremely accurate predictions of quantities like the anomalous magnetic moment of the electron and the Lamb shift of the energy levels of hydrogen.

In QED, particle interactions, and thus the mediator gauge boson, the photon, arise because the Lagrangian describing a free fermion field has to be invariant under local gauge transformations. The Lagrangian \mathcal{L}_0 describing a free Dirac fermion of mass m represented by a field $\psi(x)$, has the form:

$$\mathcal{L}_0 = i\bar{\psi}(x)\gamma^\mu\partial^\mu\psi(x) - m\bar{\psi}(x)\psi(x), \quad (2.1)$$

where γ^μ are the Dirac matrices.

This Lagrangian is invariant under a global $U(1)$ phase transformations:

$$\psi(x) \rightarrow \psi'(x) \equiv e^{iq\theta}\psi(x), \quad (2.2)$$

where q is the electric charge, Q is the fermion charge in unit of e and θ is an arbitrary real constant. However, this free Lagrangian is no longer invariant if the phase θ depends on the space-time coordinates. Thus, once a phase convention has been adopted at one reference point x_0 , the same convention must be used for all space-time points, which is unnatural. The "gauge principle" is the requirement that the $U(1)$ phase invariance should hold locally. In order to maintain the Lagrangian invariant under this local transformation, a modified derivative D_μ must replace the ordinary derivative. The definition of this covariant derivative is:

$$\partial_\mu\psi(x) \rightarrow D_\mu\psi(x) \equiv [\partial_\mu + ieQA_\mu(x)]\psi(x), \quad (2.3)$$

where e is the electron charge, Q and $A_\mu(x)$ is a vector field. The modified Lagrangian, invariant under local $U(1)$ transformations, is:

$$\mathcal{L} = i\bar{\psi}(x)\gamma^\mu D_\mu\psi(x) - m\bar{\psi}(x)\psi(x) = \mathcal{L}_0 - eQA_\mu(x)\bar{\psi}(x)\gamma^\mu\psi(x) \quad (2.4)$$

Here $j^\mu = \bar{\psi}(x)\gamma^\mu\psi(x)$ is the fermion current density and $A_\mu(x)$ corresponds to the photon field, which is the carrier of the electromagnetic interactions. The gauge principle has therefore generated an interaction between a Dirac fermion and a gauge field $A_\mu(x)$. An additional gauge invariant term that accounts for the photon kinetic energy should be added in order that $A_\mu(x)$ can be interpreted as a propagating field, so that the complete QED Lagrangian becomes:

$$\mathcal{L}_{QED} = \mathcal{L}_0 - eQA_\mu(x)\bar{\psi}(x)\gamma^\mu\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (2.5)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. A mass term for the gauge field, $\mathcal{L}_m = \frac{1}{2}m^2 A^\mu A_\mu$, is forbidden because it would violate the local $U(1)$ gauge invariance - hence, the photon is predicted to be massless. Experimentally, we know that $m_\gamma < 10^{-18}\text{eV}$. As it will be discussed in the following section, the absence of a mass term in the Lagrangian

describing the weak interaction was an important issue for particle physics. The problem was solved by the Brout-Englert-Higgs mechanism [17, 18, 19].

2.1.2 The theory of strong interactions - QCD

Already prior to the 1960s, many particles decaying with lifetimes typical of strong interactions (hadrons) were discovered. The results from deep inelastic scattering experiments indicated that hadrons are composed of point-like constituents. In particular, it was proposed that they are formed by partons that determine the hadron properties called valence quarks and by virtual partons composed of quarks, anti-quarks and gluons collectively called the sea. Evidence of gluons, the gauge bosons of the strong interaction, was found in e^+e^- collision events with the discovery of final states with three jets [20]. Hadrons are classified in two categories:

- baryons = qqq (3 valence quarks), like $p(uus)$, $n(udd)$, $\Lambda(uds)$ and
- mesons = $q\bar{q}$ (pair of valence quark - antiquark), like $\pi^+(u\bar{d})$, $K^0(\bar{s}d)$, ψ -meson($c\bar{c}$)

Using this categorization, one can nicely classify the entire hadron spectrum. After the discovery of the $\Delta^{++}(uuu)$ baryon, a new quantum number, the color, was introduced in order to preserve the Pauli principle. This new quantum number can take on three values called blue, red and green (antiblue, antired and antigreen for the antiquarks). The particles we observe, baryons and mesons, are colorless states. The introduction of the color led to the construction of quantum chromodynamics (QCD), the quantum field theory that describes the strong interactions of colored quarks and gluons, which possesses a $SU_C(3)$ local gauge symmetry.

The free QCD Lagrangian reads:

$$\mathcal{L}_0 = \sum_f \bar{q}_f^\alpha (i\gamma^\mu \partial_\mu - m_f) q_f^\alpha, \quad (2.6)$$

where q_f^α is a quark field of color α and flavor f . The free Lagrangian is invariant under global $SU(3)_C$ transformations

$$q_f^\alpha \rightarrow (q_f^\alpha)' = U_\beta^\alpha q_f^\beta, \quad (2.7)$$

where U is a 3×3 unitary matrix, that can be written as

$$U = \exp\left\{i \frac{\lambda^\alpha}{2} \theta_\alpha\right\}, \quad (2.8)$$

where λ^α ($\alpha = 1, 2, \dots, 8$) are the Gell-Mann matrices, the generators of the fundamental representation of the $SU(3)_C$ algebra and θ_α are arbitrary real parameters.

As in the case of QED, the QCD Lagrangian has to be invariant under the local gauge $SU(3)_C$ transformations of Equation 2.7 with $\theta_\alpha = \theta_\alpha(x)$. In order to satisfy

this principle, the derivatives have to be replaced by the covariant derivative¹.

$$\partial_\mu q_f \rightarrow D^\mu q_f \equiv [\partial^\mu + ig_s \frac{\lambda^\alpha}{2} G_\alpha^\mu(x)] q_f, \quad (2.9)$$

where g_s is a coupling constant and G_α^μ ($\alpha = 1, 2, \dots, 8$) are eight vector fields corresponding to the gluon fields. To express the kinetic energy of the gluon field, the corresponding field strengths are introduced:

$$G^{\mu\nu}(x) = \partial^\mu G^\nu - \partial^\nu G^\mu + ig_s [G^\mu, G^\nu]. \quad (2.10)$$

Finally the QCD Lagrangian [21] can be written as:

$$\mathcal{L}_{QCD} \equiv -\frac{1}{4} G_\alpha^{\mu\nu} G_{\mu\nu}^\alpha + \sum_f \bar{q}_f (i\gamma^\mu D_\mu - m_f) q_f. \quad (2.11)$$

A physical consequence of this QCD Lagrangian is that quarks interact by exchanging gluons. The strength of the interactions is characterized by a single universal coupling constant $\alpha_s = (\frac{g_s}{4\pi})^2$ known as the *strong coupling constant*. Moreover, the first term in Equation 2.11 gives rise to self-interactions² among the gauge fields, which is not the case for QED.

Figure 2.2 shows the dependence of α_s on the energy of the interaction. The long distance regime, relevant for low momenta ($Q \leq 1$ GeV), is characterized by a strong coupling $\alpha_s \sim \mathcal{O}(1)$, which explains the fact that quarks are confined in hadrons and cannot appear as free particles. By contrast, in the short distance regime, $Q \gg 1$ GeV, the coupling constant is $\alpha_s \ll 1$. The decrease of α_s with energy adds asymptotic freedom to the properties of the QCD, thus allowing to describe high energy interactions between protons using perturbation theory in terms of scattering among quasi-free partons (quarks and gluons).

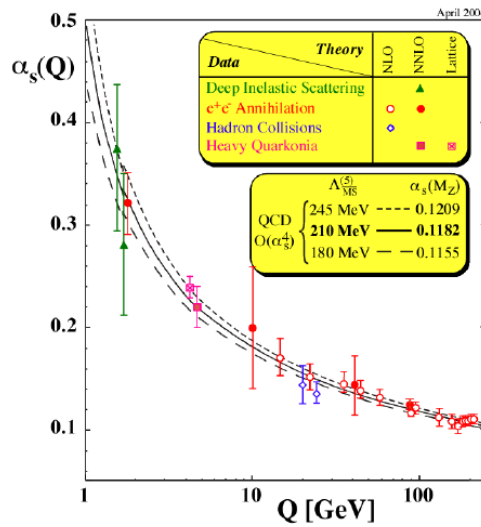


FIGURE 2.2: The dependence of α_s on the energy of the interaction Q [22, 23].

¹To simplify the equations, we consider a vector notation in color space: $q_f \equiv (q_f^1, q_f^2, q_f^3)$.

²The QCD theory is therefore called a non-abelian theory.

2.1.3 The theory of electroweak interaction

The theory was put in its present form in 1968-1978 by Glashow, Salam and Weinberg [24, 25, 26]. The Glashow-Salam-Weinberg model treats the weak and electromagnetic interactions as different manifestations of a single electroweak force, combining them into the same theoretical framework. All particles carrying a weak charge, for instance all quarks and leptons, can participate in a weak interaction. It was shown by C. Wu [27] that the weak interaction is parity-violating and connects only the left-handed states of the fermion fields. The left-handed components of the lepton and quark fields are therefore arranged in "isospin" doublets of an $SU(2)$ symmetry as follows:

$$\psi_L = \begin{pmatrix} e \\ \nu_e \end{pmatrix}_L, \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L, \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L$$

$$\psi_L = \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L$$

and the right-handed component in singlets of $SU(2)$, $\psi_R = q_R$ where $q = (e, \mu, \tau, u, d', c, s', t, b')$. The states d', s' and b' are linear combination of d, s and b quarks; the flavor mixing is given by the Kobayashi-Maskawa Matrix (CKM-Matrix) [28].

To preserve the local gauge invariance, the electroweak (EW) Lagrangian introduces covariant derivatives with three gauge fields for the $SU(2)_L$ symmetry (a triplet of spin 1 fields $W_\mu = W_\mu^1, W_\mu^2, W_\mu^3$ with a coupling constant g) plus one field for the $U(1)_Y$ symmetry (a single spin 1 field B_μ with a coupling constant g'). The underlying symmetry of the EW interaction is therefore $SU(2)_L \times U(1)_Y$, where Y is the hypercharge, and the subscript L refers to the fact that the charged weak gauge bosons only couple to left-handed particles. The relation among the electric charge Q , which is the conserved quantity due to gauge invariance in QED, the third component of the weak isospin T_3 , which is the conserved quantity of the weak interaction, and the hypercharge Y , the conserved quantity in the $U(1)_Y$, is:

$$Q = T_3 + \frac{Y}{2}. \quad (2.12)$$

The EW quantum numbers of the fermions are shown in Table 2.1.

Field	T_3	Q	Y
u_L	1/2	2/3	1/3
d_L	-1/2	-1/3	1/3
ν_L	1/2	0	-1
e_L	-1/2	-1	-1
u_R	0	2/3	4/3
d_R	0	-1/3	-2/3
e_R	0	-1	-2

TABLE 2.1: Quantum numbers of the fermions with respect to the $SU(2)_L \times U(1)_Y$ symmetry.

The physical fields [29] of the photon (A_μ) and of the neutral boson responsible for the weak interaction (Z^0) are related to B_μ and W_μ^3 by

$$\begin{pmatrix} A_\mu \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \quad (2.13)$$

where θ_W is the weak mixing angle, which was determined experimentally to be $\theta_W = 28.75^\circ$ [30]. The weak mixing angle can be expressed in terms of the $SU(2)_L \times U(1)_Y$ gauge field coupling strengths g and g' :

$$\sin \theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}. \quad (2.14)$$

The physical charged EW bosons are a combination of the spin 1 fields introduced in the covariant derivative:

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \quad (2.15)$$

The Lagrangian of the EW theory reads:

$$\begin{aligned} \mathcal{L}_{EW} = & -\frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \\ & + \bar{\psi}_L(i\gamma^\mu\partial_\mu - g\frac{1}{2}\tau_a\gamma^\mu W_\mu^a - g'\frac{Y}{2}\gamma^\mu B)\psi_L \\ & + \bar{\psi}_R(i\gamma^\mu\partial_\mu - g'\frac{Y}{2}\gamma^\mu B)\psi_R, \end{aligned} \quad (2.16)$$

where τ_a are the Pauli matrices and $W_{\mu\nu}$ and $B_{\mu\nu}$ are the field strength tensors of $SU(2)_L$ and $U(1)_Y$, respectively. The first and the second term in the first line of Equation 2.16 represent the kinetic energy of the gauge fields. Until now, no masses have been introduced into the electroweak theory. A naïve introduction of mass terms is not possible since it would break the $SU(2)_L \times U(1)_Y$ local gauge symmetry.

2.1.4 Spontaneous electroweak symmetry breaking

As mentioned earlier, the local gauge invariance $SU(2)_L \times U(1)_Y$ implies that the gauge bosons W and Z of the electroweak model, as well as the fermions, are massless. However, particles with characteristics of charged and neutral EW bosons, the W^\pm and Z bosons, were found and their mass was measured to be $M_{W^\pm} \simeq 80$ GeV and $M_Z \simeq 91$ GeV. The solution to this conundrum, along with the elucidation of the problem of the origin of the mass of quarks and leptons, is the major achievement of the Brout-Englert-Higgs mechanism. Brout, Englert and Higgs independently proposed the existence of a new scalar field ϕ . This field ϕ is expressed in terms of four real scalar fields ϕ_i ($i = 1, 2, 3, 4$) arranged in a complex doublet [31]:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.17)$$

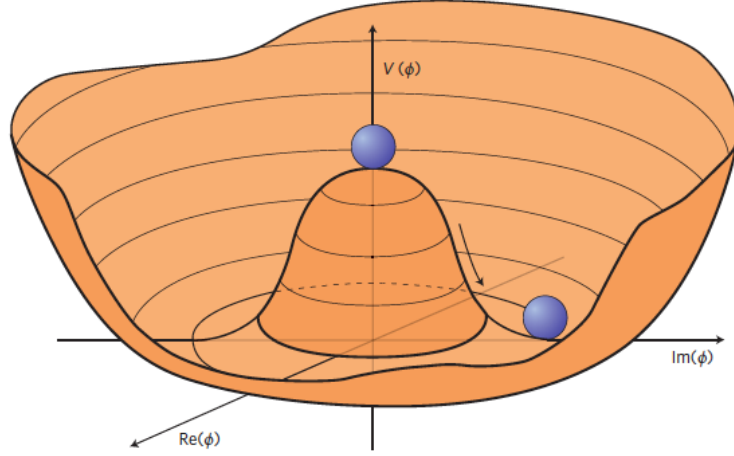


FIGURE 2.3: Representation of the shape of the potential $V(\phi)$ around the minimum of the Higgs field.

The characteristics of this new scalar field ϕ are such that a mass can be attributed to the gauge bosons via the "spontaneous" symmetry breaking mechanism. Such symmetry breaking occurs when the Lagrangian describing a system is invariant under a gauge transformation while its minimum energy state is not. The Lagrangian describing the new scalar field is written as:

$$\mathcal{L}_\phi = D^\mu \phi D_\mu \phi + \mu^2 |\phi|^2 - \lambda |\phi|^4, \quad (2.18)$$

where D^μ is the covariant derivative $D^\mu = i\partial_\mu - g\frac{1}{2}\tau W_\mu - g'\frac{Y}{2}B_\mu$ and $-\mu^2|\phi|^2 + \lambda|\phi|^4$ is the potential $V(\phi)$, with μ and λ real constants.

For $\mu^2 > 0$ and $\lambda > 0$, the potential has a non-zero minimum, the vacuum (see Figure 2.3). At this minimum, the field satisfies the condition $|\phi|^2 = \frac{\mu^2}{2\lambda}$. This condition is met by different choices for the values of ϕ_i^0 ($i = 1, 2, 3, 4$). Selecting a specific vacuum leads to spontaneous symmetry breaking and to the appearance of mass terms for the W^\pm and Z bosons in the Lagrangian. Choosing the vacuum state with $|\phi_1^0|^2 = |\phi_2^0|^2 = |\phi_4^0|^2 = 0$ and $|\phi_3^0|^2 = +\frac{\mu^2}{\lambda} = u^2$, a general expression of the scalar field ϕ is:

$$\phi = \frac{e^{i\tau\theta(x)}}{\sqrt{2}} \begin{pmatrix} 0 \\ u + H(x) \end{pmatrix} \quad (2.19)$$

where θ are three real massless scalar fields called Goldstone bosons and $H(x)$ is the Higgs field. The Higgs field $H(x)$ arises from quantum fluctuations around the minimum. When the field expression of Equation 2.19 is plugged into the Lagrangian of equation 2.18, these fluctuations generate the appropriate mass terms: for the Higgs field $H(x)$, a mass term appears directly, while the gauge fields W^i become massive after "absorbing" the Goldstone bosons. The relevant equations expressing the masses of the W^\pm and Z bosons as a function of the coupling constants g , g' and the u are

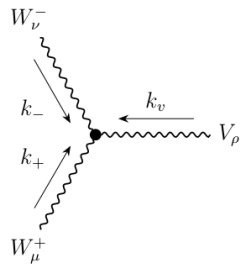
the following:

$$\begin{aligned} M_Z &= \frac{1}{2}u\sqrt{g^2 + g'^2}, & M_W &= \frac{1}{2}ug, \\ M_Z &= \frac{M_W}{\cos\theta_W} \end{aligned} \quad (2.20)$$

The discovery of the W^\pm and Z bosons was considered a major success for the theory and for CERN. The mass of the W boson has been precisely measured at LEP, Tevatron and LHC, yielding a combined result of 80.379 ± 0.012 GeV/ c^2 . The mass of the Z boson was measured using LEP data and was found to be 91.1876 ± 0.0021 GeV/ c^2 [32].

2.1.5 Electroweak gauge boson self-couplings

The electroweak Lagrangian written in Equation 2.16, predicts triple and quartic interactions between the electroweak gauge bosons. Only charged triple gauge boson vertices are allowed: W^+W^-, V with $V \in Z, A$ [33]. They are represented by the diagram in Figure 2.4, where k_+, k_-, k_ν are the four-momentum of the gauge bosons and g_{ab} , ($a, b = \mu, \nu, \rho$) is the metric tensor.



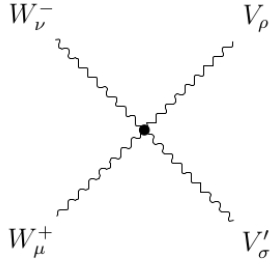
$$= ie C_{WWV} [g_{\mu\nu}(k_+ - k_-)_\rho + g_{\nu\rho}(k_- - k_\nu)_\mu + g_{\rho\mu}(k_\nu - k_+)_\nu]$$

$$\begin{aligned} \text{For } Z \text{ boson: } & C_{WVZ} = \frac{\cos(\theta_w)}{\sin(\theta_w)} \\ \text{For photon: } & C_{WVA} = -1 \end{aligned}$$

FIGURE 2.4: Feynman diagram and rules representing the coupling of three gauge bosons. The term C_{WWV} ($V = W, Z, A$) is a coupling constant.

As for the triple gauge boson interactions, the SM includes only charged quartic electroweak gauge boson interactions: W^+W^-VV' with $VV' = W^+W^-, ZZ, ZA, ZA$. Compared to the triple gauge boson interactions, an additional coupling factor e appears (Figure 2.5).

Another important diagram in the study of the EW boson self couplings is depicted in Figure 2.6. It represents the coupling of the SM Higgs boson H to the W and Z bosons. It is proportional to the electric coupling constant e and the corresponding gauge boson mass m_V .



$$= i e^2 C_{WWVV'} (2g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\sigma\nu} - g_{\mu\sigma}g_{\nu\rho})$$

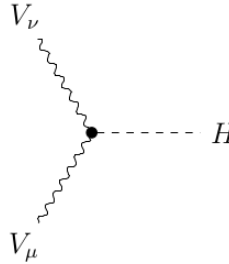
$$\text{For } WW: \quad C_{WWWW} = \frac{1}{\sin^2(\theta_\omega)}$$

$$\text{For } ZZ: \quad C_{WWZZ} = \frac{\cos^2(\theta_\omega)}{\sin^2(\theta_\omega)}$$

$$\text{For } Z\gamma: \quad C_{WWZA} = \frac{\cos(\theta_\omega)}{\sin(\theta_\omega)}$$

$$\text{For } \gamma\gamma: \quad C_{WWAA} = -1$$

FIGURE 2.5: Feynman diagram and rules representing the coupling of four gauge bosons. The term $C_{WWVV'}$ ($V, V' = W, Z, A$) is a coupling constant.



$$= -i e C_{VVH} m_V g_{\mu\nu} = -2i m_V^2 g_{\mu\nu} / u$$

$$\text{For } WWH: \quad C_{WWH} = \frac{1}{\sin(\theta_\omega)}$$

$$\text{For } ZZH: \quad C_{ZZH} = \frac{1}{\sin(\theta_\omega) \cos(\theta_\omega)}$$

FIGURE 2.6: Feynman diagram and rules representing the coupling of the Higgs boson with the massive electroweak bosons. The term C_{VVH} ($V = W, Z$) is a coupling constant.

2.1.6 Electroweak gauge boson scattering

Gauge boson self interactions are responsible of the vector boson scattering (VBS), namely of the process $VV \rightarrow VV$ with $V = W^\pm, Z, \gamma$ [33]. This includes triple and quartic gauge interactions (TGCs and QGCs) as well as interactions between the gauge and the Higgs bosons, as shown in Figure 2.7. For the process $VV \rightarrow VV$, when only subsets of the diagrams in Figure 2.7 are considered, the theoretical predictions show an unbounded rise of the cross section with increasing center-of-mass energy. Therefore the study of the VBS processes may enlighten the gauge sector and the Higgs role in the electroweak theory. In practice, the quartic gauge interactions involved

in the VBS represent one of the main reasons why it is an interesting process to study³. New phenomena could generate additional contributions to quartic gauge interactions with respect to the SM predictions, which manifest in anomalous quartic gauge couplings as described in Section 2.3.

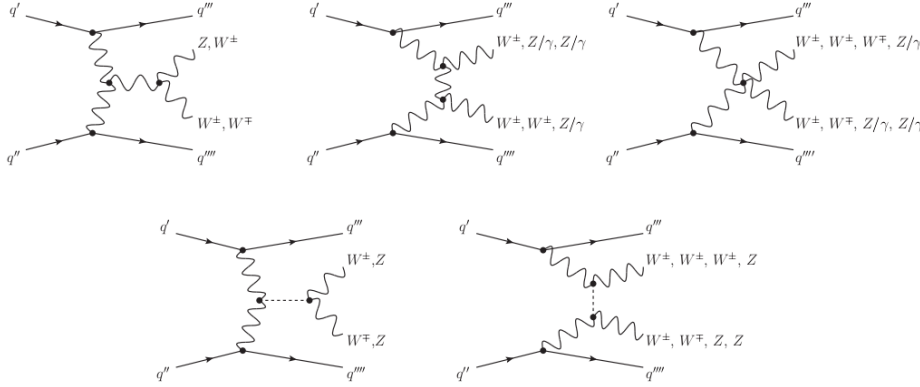


FIGURE 2.7: Diagrams representing the VBS process.

A closer look at Figure 2.7 reveals the characteristics of the scattering process. The VBS processes are studied at the LHC via a process in which two gauge bosons, radiating off the incoming quarks, interact with each other. This results in a final state with two bosons and two quarks (channel $VVjj$).

It is worth mentioning that this figure includes only a fraction of the complete list of possible $VVjj$ processes. In the upper part of Figure 2.8, a set of $VVjj$ -EW diagrams is presented. Cross sections are proportional to the fourth (sixth) power of the electroweak coupling constant α_{em} (including the decay of the vector bosons). No diagrams with strong vertices are included. Feynman diagrams that share the same initial and final state and contain strong interaction vertices are named $VVjj$ -QCD. The lowest order in QCD are represented by the diagrams in the lower part of Figure 2.8, where the order of the coupling constant in the cross section is $\alpha_{em}^2\alpha_s^2$ ($\alpha_{em}^4\alpha_s^2$). These processes are the dominant background in most EW gauge boson scattering analyses. Moreover, the $VVjj$ -EW and $VVjj$ -QCD processes have identical initial and final states and therefore interfere. A dedicated study of this effect in the case of $Z\gamma jj$ is presented in Chapter 5.

High transverse momentum jets in the forward pseudorapidity region, called tagging jets in the analysis, are a very important experimental signature of the $VVjj$ -EW final state. The tagging jets exhibit a large spatial separation in rapidity and a high invariant mass. These properties are caused by the fact that the two jets originate from the incoming quarks, which emit the scattering vector bosons, thus carrying a high fraction of the momentum of the incoming protons. In addition, none or insignificant jet activity (occurrence of jets) between the two tagging jets is expected due to a lack of color flow between the incoming quarks. The vector bosons resulting

³At the LHC QGCs can be also studied using processes with three bosons in the final state.

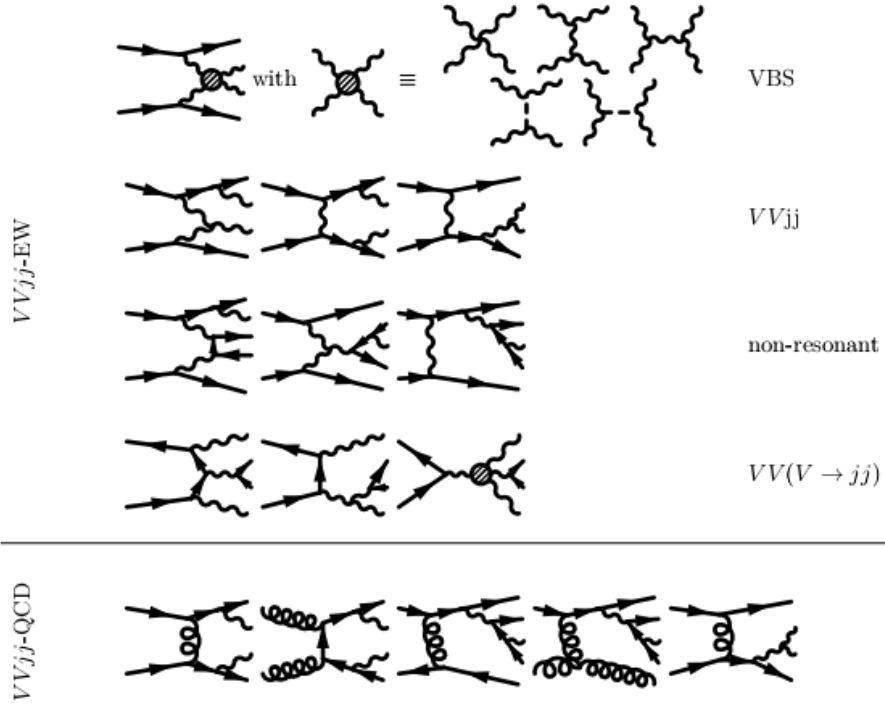


FIGURE 2.8: Diagrams representing the $qq \rightarrow VVjj$ processes [33].

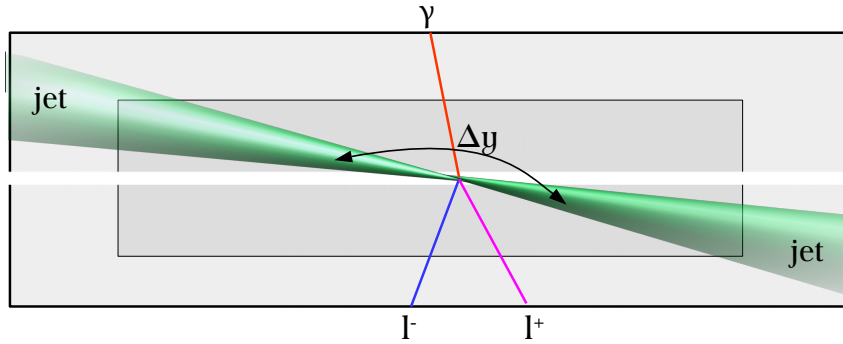


FIGURE 2.9: Topology of a VBS event. In this graph the VBS $Z\gamma jj$ process is illustrated where the Z boson decays into a pair of leptons.

from the scattering are usually located, as Figure 2.9 shows, centrally with respect to the tagging jets [34]. For this reason, a new variable, called centrality of the diboson system is introduced and used in the analysis carried on in this thesis as:

$$\zeta(Z\gamma) = \left| \frac{y_{Z\gamma} - (y_{j1} + y_{j2})/2}{\Delta Y_{jj}} \right|, \quad (2.21)$$

where y_{j1} and y_{j2} are the rapidity of the tagging jets, $\Delta Y_{jj} = |y_{j1} - y_{j2}|$ and $y_{Z\gamma}$ is the rapidity of the $Z\gamma$ system. Decreasingly positive values of the centrality indicate

a large separation between the tagging jets and a large separation of the bosons and the two jets. The properties of the diboson system yield additional variables suitable for analysis, as will be shown in Chapter 5.

New resonances may be also produced as intermediate resonances when two electroweak gauge bosons interact. The properties of the observed diboson system are then influenced by the properties of the intermediate resonance. In case of diboson resonances with large masses, one would expect a high invariant mass of the diboson system on a large separation between the angle ϕ of the outgoing boson momenta, and high bosons transverse momenta. The resonant diboson production is not addressed in this thesis.

2.2 Effective field theory

Despite being an extremely successful theory, the SM is not believed to be the ultimate theory to describe nature. Some of the most far-reaching issues with the SM are, for example, [35]:

- the SM does not include a description of the gravitational force,
- the SM describes only 5% of the existing matter in the universe, without addressing the issue of dark matter,
- the SM provides no explanation regarding the relations of the individual strengths of the forces or the mass hierarchy of the elementary particles.

These and additional problems has led physicists to search for evidence for new physics beyond the SM, which can be carried out in two ways. The first way is to develop a specific model and then analyze data to test the predictions of the model. In this case, a restricted approach is followed, which might however be highly sensitive (an example is the SUSY model). Still, this method may fail if new physics are beyond the kinematic reach of the LHC.

Another way of searching for new physics is to parameterize the effects of new physics on SM processes via effective field theories (EFT). The underlying assumption of effective field theories is that new physics cannot be observed directly but would be found above the current energy threshold of the experiment. According to EFT [36], the low energy effects of new physics may be detectable through deviations from the SM predictions which might appear just above the EW scale, around the TeV scale or even beyond (bottom-up approach).

These effects can be parameterized using an effective Lagrangian (\mathcal{L}_{EFT}) with a given energy scale Λ for new physics. The SM Lagrangian (\mathcal{L}_{SM}) is extended through the introduction of a set of new operators $\mathcal{O}^{(d)}$ of dimension d , accompanied by dimensionless coupling constant $c_i^{(d)}$:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_{d>4} \sum_i \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathcal{O}_i^{(d)}. \quad (2.22)$$

For dimensional reasons, the operators $\mathcal{O}^{(d)}$ have coefficients with dimension of inverse powers in energy, so the operators with the lowest dimensions are dominant. Most of the SM operators are of dimension four, and since only operators with even dimension satisfy the conservation of lepton and baryon number, the new operators have at least dimension six.

2.3 Anomalous Quartic Gauge Couplings - aQGCs

For the description of the BSM contributions in terms of anomalous quartic gauge couplings, two EFT representations are mainly used. These representations, the linear and the non-linear, can be distinguished due to the different ordering of the EFT expansion. In the non-linear representation, the Higgs couplings are treated as additional free parameters and deviations in the Higgs sector can be introduced already at lowest order. The non-linear representation was motivated before the experimental Higgs discovery in case of a heavy or strongly interacting Higgs. However, since no significant deviation from the light SM Higgs predictions has been observed so far, the linear Higgs-doublet representation is used and will be presented in this section. Deviations from the SM predictions of the Higgs couplings with gauge bosons and of triple or quartic gauge boson couplings start to appear at energy dimension-6.

The aQGCs [37] are not independent of changes in the HVV couplings or of anomalous triple gauge boson couplings (aTGCs) which are usually measured respectively in Higgs production (or decay) and in vector boson pair production. The contribution of dimension-6 operators to the VBS cross section due to the aQGC is smaller than the contribution due to TGCs and HVV couplings. Consequently, in order to extract clean information for the aQGCs the expansion of the EFT Lagrangian in higher than dimension-6 terms is required. Out of the higher-order operators, the dimension-8 are the lowest dimension operators which influence the quartic gauge boson coupling without contributing to aTGCs or to HVV couplings. After this extension, the EFT Lagrangian takes the following form:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_i \frac{c_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_j \frac{f_j^{(8)}}{\Lambda^4} \mathcal{O}_j^{(8)}. \quad (2.23)$$

The dimension-8 operators are separated into longitudinal ($\mathcal{O}_{S,j}$), transverse ($\mathcal{O}_{T,l}$) and mixed ($\mathcal{O}_{M,k}$) contributions containing both the Higgs $SU(2)_L$ doublet derivatives and the field strength.

From these operators only sixteen linearly independent operators can be constructed for aQGCs, usually grouped into $\mathcal{O}_{S,0,1}^{(8)}$ (with coefficient $f_{S,j}$) contains only covariant derivatives of the Higgs field, $\mathcal{O}_{M,0,\dots,7}^{(8)}$ (with coefficient $f_{M,k}$) containing derivatives of the Higgs field and gauge boson field strength tensors and $\mathcal{O}_{T,0,1,2,5,\dots,9}^{(8)}$ (with coefficient $f_{T,l}$) with just the bosonic field strength tensors included. Figure 2.10 summarizes dimension-8 operators using the Eboli basis [38]. By writing explicitly

$$\begin{aligned}
\mathcal{O}_{S,0} &= [(D_\mu \Phi)^\dagger D_\nu \Phi] \times [(D^\mu \Phi)^\dagger D^\nu \Phi] & \mathcal{O}_{T,0} &= \text{Tr} [\widehat{W}_{\mu\nu} \widehat{W}^{\mu\nu}] \times \text{Tr} [\widehat{W}_{\alpha\beta} \widehat{W}^{\alpha\beta}] \\
\mathcal{O}_{S,1} &= [(D_\mu \Phi)^\dagger D^\mu \Phi] \times [(D_\nu \Phi)^\dagger D^\nu \Phi] & \mathcal{O}_{T,1} &= \text{Tr} [\widehat{W}_{\alpha\nu} \widehat{W}^{\mu\beta}] \times \text{Tr} [\widehat{W}_{\mu\beta} \widehat{W}^{\alpha\nu}] \\
\mathcal{O}_{S,2} &= [(D_\mu \Phi)^\dagger D_\nu \Phi] \times [(D^\nu \Phi)^\dagger D^\mu \Phi] & \mathcal{O}_{T,2} &= \text{Tr} [\widehat{W}_{\alpha\mu} \widehat{W}^{\mu\beta}] \times \text{Tr} [\widehat{W}_{\beta\nu} \widehat{W}^{\nu\alpha}] \\
\mathcal{O}_{M,0} &= \text{Tr} [\widehat{W}_{\mu\nu} \widehat{W}^{\mu\nu}] \times [(D_\beta \Phi)^\dagger D^\beta \Phi] & \mathcal{O}_{T,5} &= \text{Tr} [\widehat{W}_{\mu\nu} \widehat{W}^{\mu\nu}] \times \widehat{B}_{\alpha\beta} \widehat{B}^{\alpha\beta} \\
\mathcal{O}_{M,1} &= \text{Tr} [\widehat{W}_{\mu\nu} \widehat{W}^{\nu\beta}] \times [(D_\beta \Phi)^\dagger D^\mu \Phi] & \mathcal{O}_{T,6} &= \text{Tr} [\widehat{W}_{\alpha\nu} \widehat{W}^{\mu\beta}] \times \widehat{B}_{\mu\beta} \widehat{B}^{\alpha\nu} \\
\mathcal{O}_{M,2} &= [\widehat{B}_{\mu\nu} \widehat{B}^{\mu\nu}] \times [(D_\beta \Phi)^\dagger D^\beta \Phi] & \mathcal{O}_{T,7} &= \text{Tr} [\widehat{W}_{\alpha\mu} \widehat{W}^{\mu\beta}] \times \widehat{B}_{\beta\nu} \widehat{B}^{\nu\alpha} \\
\mathcal{O}_{M,3} &= [\widehat{B}_{\mu\nu} \widehat{B}^{\nu\beta}] \times [(D_\beta \Phi)^\dagger D^\mu \Phi] & \mathcal{O}_{T,8} &= \widehat{B}_{\mu\nu} \widehat{B}^{\mu\nu} \widehat{B}_{\alpha\beta} \widehat{B}^{\alpha\beta} \\
\mathcal{O}_{M,4} &= [(D_\mu \Phi)^\dagger \widehat{W}_{\beta\nu} D^\mu \Phi] \times \widehat{B}^{\beta\nu} & \mathcal{O}_{T,9} &= \widehat{B}_{\alpha\mu} \widehat{B}^{\mu\beta} \widehat{B}_{\beta\nu} \widehat{B}^{\nu\alpha}
\end{aligned}$$

FIGURE 2.10: Field content of dimension-8 operator [38].

these operators, the effective Lagrangian becomes:

$$\begin{aligned}
\mathcal{L}_{EFT} &= \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_{j=0,1} \frac{f_{S,j}}{\Lambda^4} \mathcal{O}_{S,j}^{(8)} \\
&+ \sum_{k=0,\dots,7} \frac{f_{M,k}}{\Lambda^4} \mathcal{O}_{M,k}^{(8)} + \sum_{l=0,1,2,5,\dots,9} \frac{f_{T,l}}{\Lambda^4} \mathcal{O}_{T,l}^{(8)}. \tag{2.24}
\end{aligned}$$

As said before, at the LHC, aQGCs can be studied in electroweak gauge boson scattering processes $VVjj$ and in triple gauge boson production VVV . The coefficient of the operators of dimension-8 affecting the $VVjj$ -EW and VVV final states and the corresponding processes are shown in Table 2.2.

	ZZ	$Z\gamma$	W^+W^-	$W^\pm W^\pm$	$W\gamma$	
$VVjj$ final state		$\gamma\gamma$	WZ			
VVV final state	ZZZ	$ZZ\gamma$	WWZ	WWW	$WV\gamma$	$\gamma\gamma\gamma$
$f_{S,0}, f_{S,1}$	✓	—	✓	✓	—	—
$f_{M,0}, f_{M,1}, f_{M,6}, f_{M,7}$	✓	✓	✓	✓	✓	—
$f_{M,2}, f_{M,3}, f_{M,4}, f_{M,5}$	✓	✓	✓	—	✓	—
$f_{T,0}, f_{T,1}, f_{T,2}$	✓	✓	✓	✓	✓	✓
$f_{T,5}, f_{T,6}, f_{T,7}$	✓	✓	✓	—	✓	✓
$f_{T,8}, f_{T,9}$	✓	✓	—	—	—	✓

TABLE 2.2: Anomalous quartic electroweak gauge coupling parameters of the dimension-8 operators. Checkmark (✓) indicate a dependence of the final state on the corresponding parameter.

2.4 Proton-proton interactions at the LHC

The LHC is the largest hadron accelerator ever built. One type of collision that is investigated with it is the proton-proton collision, where the constituents of the protons (quarks, antiquarks or gluons) interact with each other.

Since perturbative QCD cannot fully describe the proton structure, proton structure functions are introduced. The proton structure functions can be expressed in the quark-proton model in terms of parton density functions (PDFs). PDFs are functions of the four-momentum transfer Q in the collision and the longitudinal momentum fraction x of the proton carried by a parton of flavor i , $f_i = f_i(x_i, Q^2)$. In a proton-proton collision, the interesting interactions happening between two partons are known as *hard scatterings*, leading to hard outgoing particles of high transverse momentum. The term "hard" indicates the large Q of the collision. *Hard scattering* interactions are described by perturbative QCD.

Most of the ATLAS trigger selection criteria are defined to select events produced by *hard scattering* processes. A schematic representation of a *hard scattering* in a proton-proton collision is given in Figure 2.11. Incoming partons may emit QCD radiation, known as initial state radiation (ISR) and outgoing partons, following the same mechanism, may emit the so-called final state radiation (FSR). Incoming and outgoing charged particles may also emit QED radiation, i.e photons. The spectator partons, namely the partons not involved in the *hard scattering*, may also produce additional hard interactions. These additional interactions between the residual partons of the original protons are known as multiple interactions. Particles from ISR, FSR, the proton remnants along with multiple interactions are known as the *underlying event* (UE).

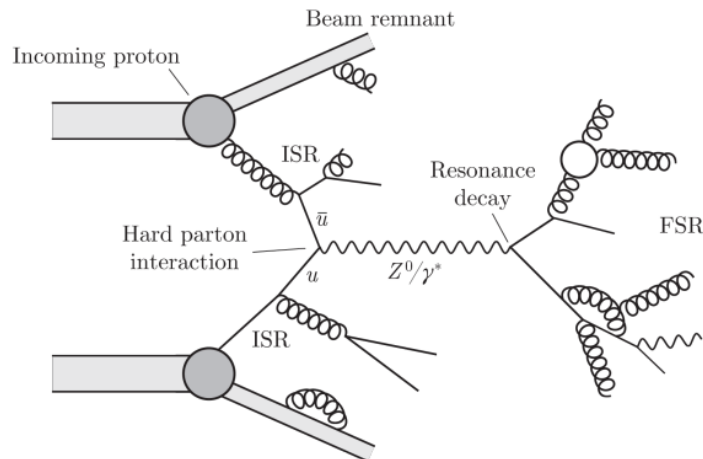


FIGURE 2.11: Schematic representation of a proton-proton *hard scattering* process. The production of a Z^0/γ^* process is illustrated with Z^0/γ^* decaying to quark pairs.

Another effect present at the LHC is the pile-up. Due to the very high number of colliding protons in a bunch crossing and the high proton-proton cross section, there

are more than two protons that interact when the two LHC proton bunches cross each other at the center of the experiment. These pile-up events are essentially produced by soft processes described by non-perturbative QCD.

2.4.1 Factorisation theorem

An important tool in QCD is the factorisation theorem [39] that factors out the hard, short-distance component of the interaction process (calculated using perturbative QCD) from the soft, long-distance one. According to this theorem, in proton-proton collisions the cross section of a *hard scattering* process $\sigma_{pp \rightarrow x}$ can be factorized into two contributions: the first is a term that represents the cross section of the parton interaction and the second is a term that corresponds to the momentum distribution of the partons inside the colliding protons (PDF). Therefore, the cross section $\sigma_{pp \rightarrow x}$ can be written as:

$$\sigma_{pp \rightarrow x} = \sigma_{hard\ scatter} \otimes \text{PDF} = \sum_q \int dx_1 dx_2 f_q(x_1, Q^2) f_{\bar{q}}(x_2, Q^2) \cdot \hat{\sigma}_{q\bar{q} \rightarrow x} \quad (2.25)$$

where x_1 and x_2 are the momentum fractions of the protons carried by the partons q and \bar{q} , respectively, and $f_q(f_{\bar{q}})$ represents the momentum fraction distribution of a parton $q(\bar{q})$ which depends also on the four momentum transfer Q of the process. Given that at short distances (high energy regime) $\alpha_s \ll 1$, perturbation theory can be applied and the partonic cross section can be expressed as a power series expansion of the coupling constant α_s :

$$\hat{\sigma}_{q\bar{q} \rightarrow x} = \underbrace{\hat{\sigma}_0}_{\text{LO}} + \underbrace{\alpha_s \hat{\sigma}_1}_{\text{NLO}} + \underbrace{\alpha_s^2 \hat{\sigma}_2}_{\text{NNLO}} + o(\alpha_s^3) \quad (2.26)$$

Here LO refers to the leading order, NLO to the next-to-leading order and NNLO to the next-to-next-to-leading order of the theoretical calculation. Figure 2.12 shows the predictions of some important SM cross sections at $p\bar{p}$ and pp colliders as a function of the center-of-mass energy.

2.4.2 Parton distribution functions

As introduced earlier, the PDFs define the probability of finding a parton in a proton with a given momentum fraction x , at a momentum transfer Q . The set of distributions $f_i(x, Q^2)$ describes how the momentum of the proton is shared between the individual parton, where f_i refers to valence quarks, sea quarks and gluons. PDFs sets are obtained by a fit to a large number of cross section data points in a large grid of Q^2 and x values from many experiments. The most commonly used procedure consists of parametrizing the dependence of the parton distributions on the variable x at some low value of $Q^2 = Q_0^2$ and evolving these input distributions up in Q^2 through the DGLAP equation [40]. There are several groups which provide the parametrization of PDF. An example of $x f_i(x, Q^2)$ distributions for the valence quarks u and d , the

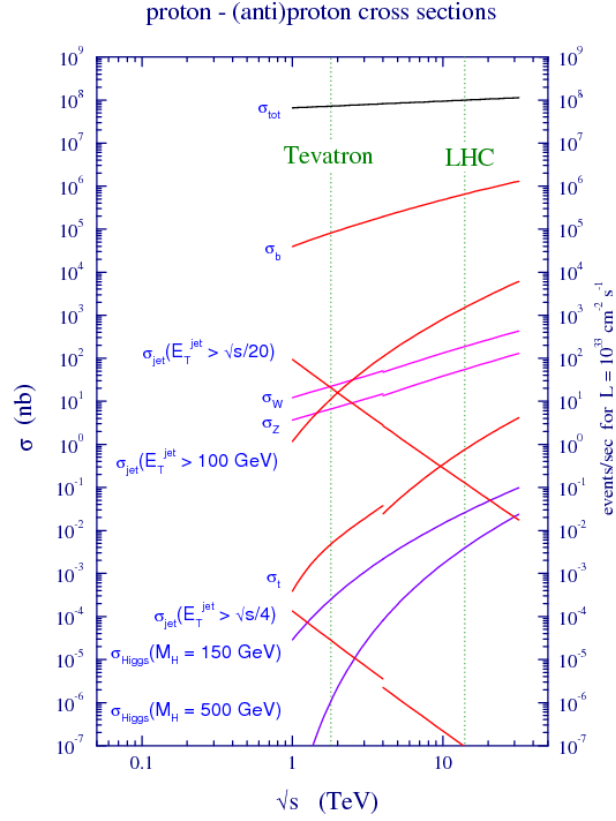


FIGURE 2.12: Cross sections of SM processes in proton-proton collisions as a function of the center of mass energy.

sea quarks and the gluons g for two different values of $Q^2 = 10 \text{ GeV}^2$ and $Q^2 = 10^4 \text{ GeV}^2$ is presented in Figure 2.13. At low x , it is the gluon PDF that dominates while at higher x the contribution from the valence quarks become more important.

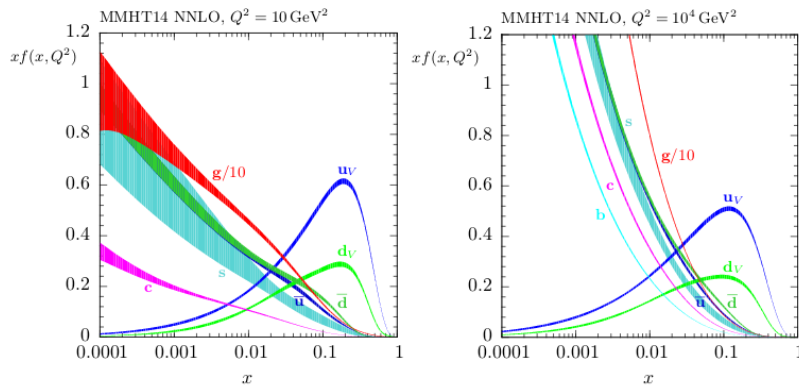


FIGURE 2.13: Parton distribution functions of the proton for two different Q^2 scales as predicted by the MMHT collaboration [41].

2.4.3 Hadronization and fragmentation

The *hard scattering* of the proton constituents leads to outgoing partons with large transverse momenta. Since the partons are colored objects that cannot propagate freely, additional $q\bar{q}$ pairs will be created to build colorless hadrons. The flow of the hadrons will constitute a "jet" structure via a fragmentation process. This process involves the production of hadrons and takes place at an energy scale where the QCD coupling constant is large and perturbation theory cannot be used. Consequently, fragmentation is described using a QCD-motivated model with parameters that must be determined experimentally.

Chapter 3

Experimental Apparatus

Από ομονοιής τα μεγάλα έργα.

Δημόκριτος 470-370 π.Χ.

The importance of theoretical models is beyond doubt. Nevertheless, no such model would bear much weight if there were no experimental results to confirm or reject it. Consequently, the precise design of an experimental apparatus is of central importance. At a particle accelerator, the colliding beams produce individual interactions referred to as events. The detector systems, using a wide range of technologies, detect and measure the properties of the particles produced in these high-energy collisions with the aim of reconstructing the primary particles. Fundamentally, one tries to go from the signals collected in the different detector systems back to the underlying theory.

3.1 The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) [42] - the world's largest particle collider - was built at the European Organization for Nuclear Research (CERN - Conseil Européen pour la Recherche Nucléaire) between 1998 and 2008. The LHC is a hadron-hadron accelerator, designed to accelerate and collide proton beams at a center-of-mass energy of up to 14 TeV with a luminosity of up to $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, as well as heavy ions at an energy of 2.8 TeV per nucleon and luminosities of up to $10^{27} \text{ cm}^{-2}\text{s}^{-1}$. The LHC (previously hosting the Large Electron Positron Collider -LEP) lies in a tunnel of 27 kilometers in circumference and 175 meters below ground level at the France-Switzerland border near Geneva.

Because they accelerate relatively light particles, lepton colliders like the LEP suffer from higher synchrotron radiation¹ losses, compared to hadron colliders of the same size. Consequently, hadron colliders are more suited to search for higher energy phenomena, which is the key to directly observe new particles or interactions. However, hadron collisions are more "messy" compared to lepton collisions due to the composite nature of hadrons.

¹Synchrotron radiation is the electromagnetic radiation emitted when charged relativistic particles are accelerated radially.

3.1.1 Accelerator complex

The journey of the proton beams in an accelerator complex takes place in a sequence of linear and circular particle accelerators. Each accelerator in the chain boosts the energy of the beam before injecting it into the next machine, thus reaching increasingly higher energies. The LHC is the very last part of this chain. Figure 3.1 illustrates the LHC proton beam journey. The proton source is a bottle of hydrogen gas. The hydrogen is ionized, passes through an electric field and only protons enter a linear accelerator, the LINAC 2. At the end of the LINAC 2, the protons have reached an energy of 50 MeV. In order to maximize the intensity of the beam, the protons are split into four beams and are injected into the Proton Synchrotron Booster (PSB), a four-layer circular accelerator which further accelerates them to 1.4 GeV. The Proton Synchrotron (PS) comes next. Here protons are being concentrated in *bunches* and reach an energy of 25 GeV. These proton bunches are then injected into the Super Proton Synchrotron (SPS), where they are accelerated to 450 GeV. At this point, they enter the LHC ring. Two proton beams travel in opposite directions in an ultrahigh vacuum, each one inside separate beam pipes. Finally, the two beams are accelerated to the final energy (6.5 TeV at present), squeezed and put in collision at the interaction points.

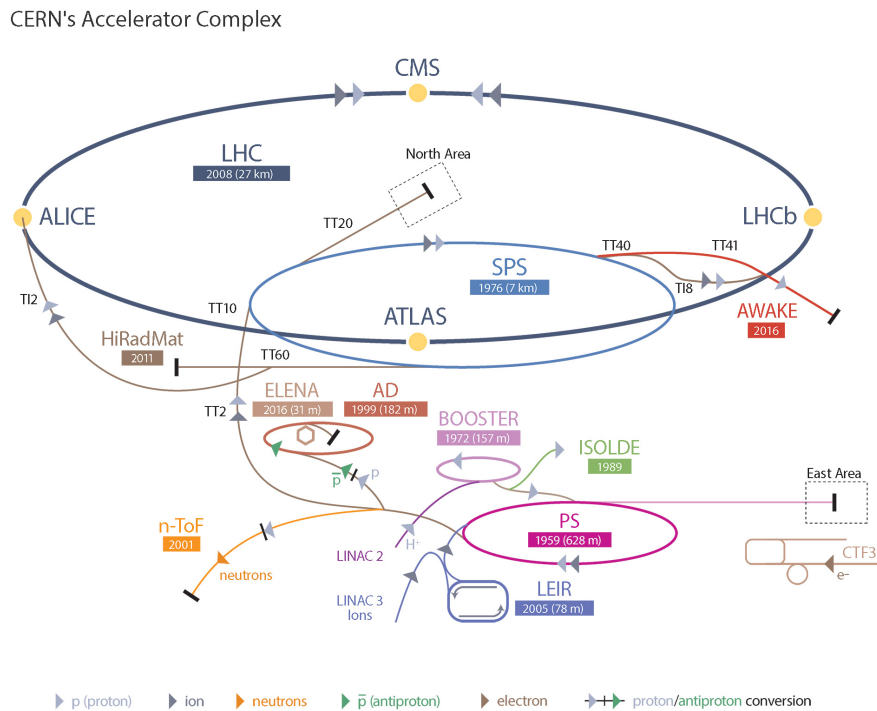


FIGURE 3.1: Graphical illustration of the CERN accelerator complex [43].

The two beams collide at four different points where four detectors, ALICE, ATLAS, CMS and LHCb are installed. The experiments record the events when an

on-line trigger signals the occurrence of an interesting event. Two of the four experiments, ATLAS and CMS, located at opposite sides of the LHC ring, are general purpose experiments. After the Higgs discovery, ATLAS and CMS focus on searching for new physics, on the measurement of the Higgs boson properties and of Standard Model processes. The LHCb focus on precise b-physics studies. The ALICE detector is designed to study heavy ion collisions, the physics of strongly interacting matter at extreme energy densities, where quark-gluon plasma is formed. Both ALICE and LHCb contribute also to the search for new phenomena in the b-physics sector and in the heavy ion collisions, respectively.

3.1.2 Accelerator parameters

The most obviously relevant parameter for particle physics experiments is the available energy at the center-of-mass of the initial particles that might allow to produce new, rare effects. Another very important parameter is the luminosity. The integrated luminosity (\mathcal{L}) relates the produced number of events (N) of a given type to their production cross section (σ) in the following way:

$$N = \mathcal{L} \times \sigma. \quad (3.1)$$

The cross section of a given process is a quantity related to the probability that this process occurs. Therefore a high luminosity is necessary in order to produce rare events (namely events with a small cross section). The luminosity is a process-independent quantity, but depends on the accelerator parameters. The instantaneous luminosity, in turn, is expressed as:

$$\frac{d\mathcal{L}}{dt} = \frac{N_p n_b^2 f_{\text{rev}} \gamma_r}{4\epsilon_n \beta^*} \times F. \quad (3.2)$$

In this relation, N_p is the number of protons per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency of the beams, γ_r is the relativistic gamma factor of the circulating particles, ϵ_n is the normalized transverse beam emittance related to the transverse phase space volume occupied by the beam, β^* is the value of the β amplitude² function at the collision point, and F is a geometric luminosity reduction factor, which depends on the crossing angle at the interaction point.

The LHC data taking happened during two phases, collectively referred to as Run I (2009-2012) and Run II (2015-2018). As an example, Figure 3.2 shows the evolution of the total integrated luminosity for the 2017 and 2018 ATLAS data taking.

One of the challenges of the experiment at the LHC is physics-related. Due to the high LHC collision rate and high proton-proton cross section, multiple proton-proton

²The amplitude β function is determined by the accelerator magnet configuration and powering. In terms of the cross-sectional size of the bunch σ and the transverse emittance ϵ , the amplitude β function becomes $\beta = \pi \times \sigma^2 / \epsilon$. If β is high, simply speaking the beam size is wide and the beam particle trajectories are more parallel.

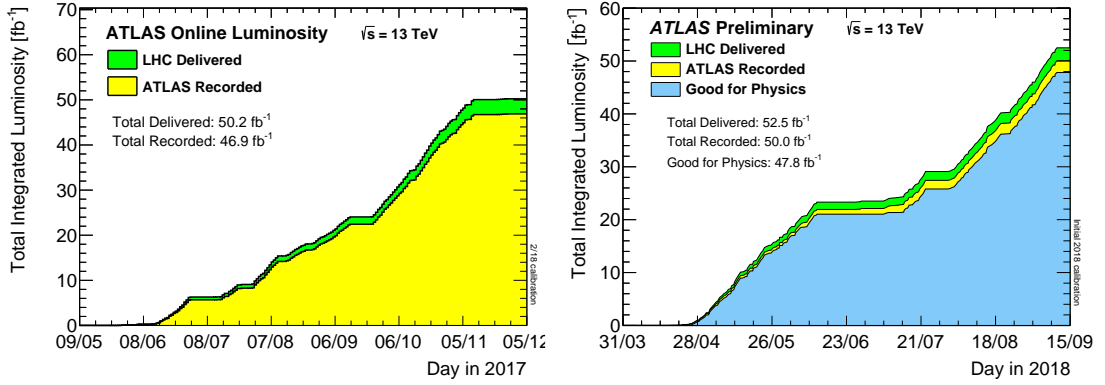


FIGURE 3.2: Evolution of the total integrated luminosity during the 2017 and 2018 ATLAS data taking [44].

interactions are recorded at the same time. This phenomenon, already introduced in Section 2.4, is called pile-up. Particles in a pile-up event have much lower transverse momentum than in the events coming from the *hard scattering*. Two categories of pile-up events exist: in-time and out-of-time. The term in-time refers to multiple events that are produced by the protons interacting in the same bunch crossing. The out-of-time pile-up events result from the superposition of events occurring in close-by bunches due to the detector latency. Figure 3.3 shows the average number of collisions per bunch crossing, a variable called $\langle\mu\rangle$, during the full Run II data taking. The distribution peaks around a value of $\langle\mu\rangle \simeq 30$. The pile-up adds complications, especially in the detection of particles using calorimeters. A good understanding and modeling of pile-up events is essential for the analysis of the ATLAS data.

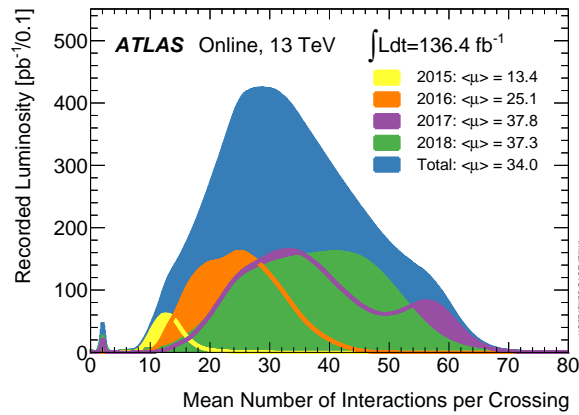


FIGURE 3.3: Distribution of the average number of interactions per bunch crossing, for the whole Run II data taking [44].

3.2 The ATLAS Detector

The ATLAS (**A Toroidal LHC ApparatuS**) detector [45] (Figure 3.4) is one of the two general-purpose detectors at the LHC, with a forward-backward symmetric

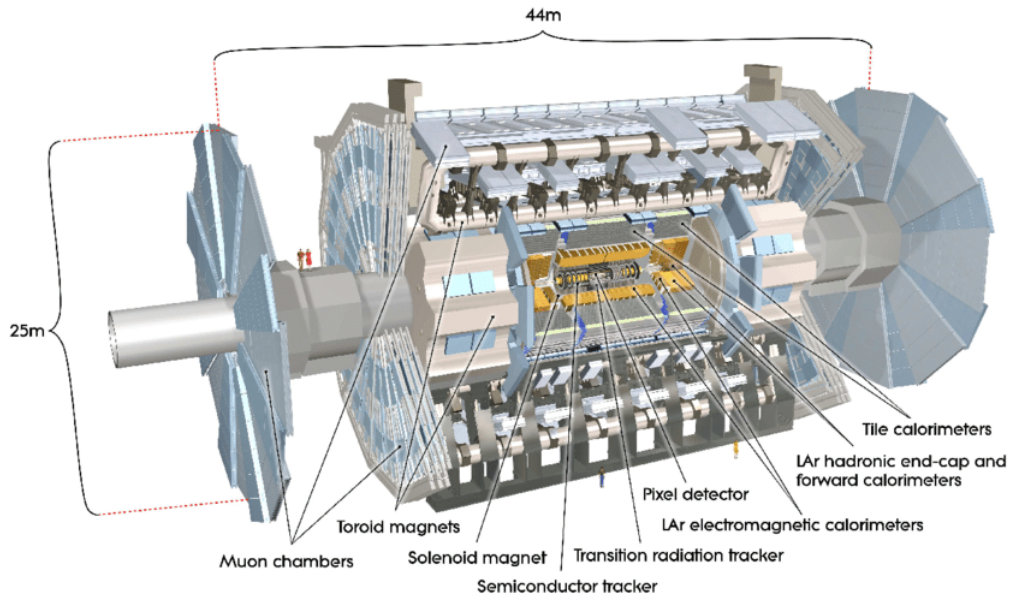


FIGURE 3.4: Schematic view of the ATLAS detector [45].

cylindrical geometry. It was designed to investigate a wide range of physical processes, from the search for the Higgs boson, extra dimensions and dark matter to precise measurement of SM processes. ATLAS is the largest detector ever constructed, being 44 m long and 25 m high, with a total weight of 7000 tons. The ATLAS detector consists of four major components, arranged in layers around the collision point and recording the path, the momentum, and energy of the particles.

Due to the high proton-proton cross section and machine luminosity, the interactions in the ATLAS detector create an enormous amount of events. In order to digest this information, the ATLAS trigger system preselects events which may be of interest and should hence be recorded. To this extent, complex data acquisition and computing systems are employed for on-line analysis and event selection.

The purpose of the ATLAS detector is to record as much information as possible about the final state particles which carry information about the interaction that has generated them. This is accomplished with the help of different detecting subsystems. The four major components are: Inner Detector, Calorimeter, Muon Spectrometer and Magnet System. The reason that detectors are divided into many components is that each component measures a specific set of particle properties.

The **Inner Detector**, immersed in a solenoidal magnetic field, is mainly used to reconstruct charged particle tracks, to measure their momenta and to compute primary and secondary vertex coordinates.

The **Calorimeters** are mainly used to identify and measure the energy of electrons, photons and hadrons. They are designed to stop most types of particles except muons and neutrinos.

The **Muon Spectrometer** is used to detect and measure the properties of muons.

Since muons interact minimally in the detector and have long lifetimes, they are identified and measured in the outermost detector system.

The **Magnet System** curves the trajectory of the charged particles for momentum measurements and charge identification.

Before proceeding to describe in more detail the ATLAS subsystems, the following subsection illustrates the coordinate system used by the ATLAS experiment.

3.2.1 ATLAS Coordinate System

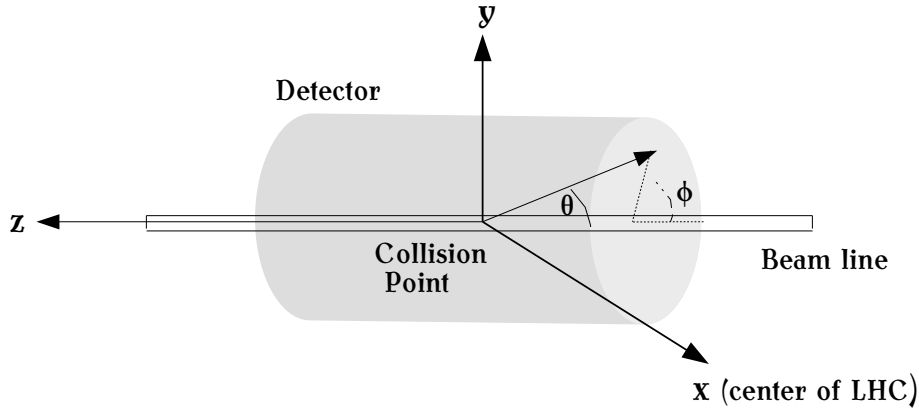


FIGURE 3.5: The coordinate system used by the ATLAS experiment.

The nominal interaction point of ATLAS corresponds to the origin of the coordinate system. The latter is a right-handed system with the z -axis running along the beam line, while the x - y plane is perpendicular to the beam line and referred to as the transverse plane (see Figure 3.5). The positive x -axis points from the interaction point to the center of the LHC ring, and the positive y -axis points upwards. The positive z -axis is referred to as "side-A", while the negative z -axis as "side-C". The transverse plane is described in terms of r - ϕ coordinates, where the azimuthal angle ϕ is measured from the x -axis around the beam and the radial coordinate r measures the distance from the beam line. The transverse momentum of a particle p_T , is defined as the momentum projected onto the r - ϕ plane. The angle θ is the polar angle measured from the beam line. The polar angle of the track of a particle is related to its pseudorapidity η by the following relation:

$$\eta = -\ln[\tan(\theta/2)]. \quad (3.3)$$

If the mass of a particle is negligible with respect to its energy, the pseudorapidity is a very good approximation of the rapidity y of a particle of energy E and longitudinal

momentum p_z :

$$y = \frac{1}{2} \log\left(\frac{E + p_z}{E - p_z}\right). \quad (3.4)$$

The angular distance between particles in the η - ϕ space is defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (3.5)$$

Once the transverse momentum and the pseudorapidity of a particle are measured, the Cartesian components of the momentum p_x, p_y, p_z and the modulo of the momentum p can be calculated from:

$$\begin{aligned} p_x &= p_T \cos(\phi), & p_y &= p_T \sin(\phi), \\ p_z &= p_T \sinh(\eta), & p &= p_T \cosh(\eta). \end{aligned} \quad (3.6)$$

3.2.2 The Magnet System

The momentum of charged particles is measured from the curvature of their trajectory in a magnetic field. For the same value of the magnetic field, low momentum particles bend more than high momentum particles. In addition, the direction in which the deflection occurs determines the charge of the particles. The ATLAS Magnet System [45] consists of four superconducting magnets; one solenoid and three toroids, as shown in Figure 3.6. The cooling of the magnets is carried out with liquid Helium.

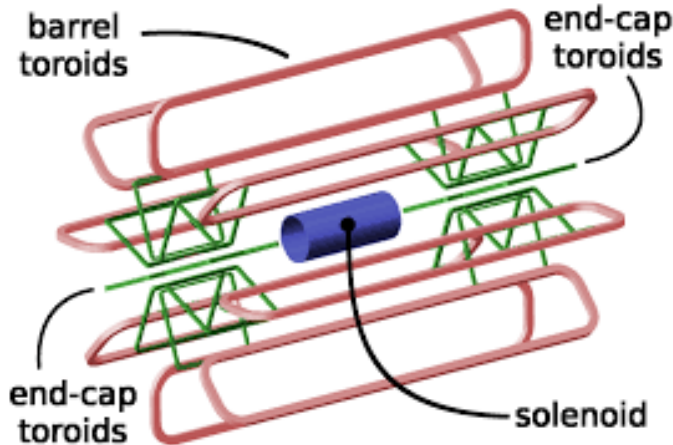


FIGURE 3.6: Schematic view of the ATLAS Magnet System.

The central solenoid provides a 2 T magnetic field which bends the charged particles in the r - ϕ plane. It is located in front of the electromagnetic calorimeter and it is 5.3 m long with a diameter of 2.4 m. It shares the cryostat with the Liquid Argon electromagnetic barrel calorimeter. As the electromagnetic calorimeter is situated outside the solenoid, the coil of the solenoid must have few radiation length, X_0 .

In the Muon Spectrometer the magnetic field is produced by toroid magnets. A total of three toroid magnets, one located in the barrel and two in the end-cap regions, is used. The barrel toroid consists of eight superconducting coils, each 25.3 m long with an outer diameter of 20.1 m, grouped in the shape of a torus. The eight coils in

the torus are kept in position by sixteen support rings. The barrel toroid generates a magnetic field up to 2.5 T with a mean value of approximately 0.5 T. The end-cap toroid magnets, located inside the barrel toroid at both ends of the central solenoid, produce a magnetic field of 0.2 to 3.5 T and provide the required high magnetic field across a radial span of 1.5 to 5 m. The choice of the open-air toroid configuration was made to improve the muon reconstruction and identification performance without relying on the Inner Detector. Such configuration allows to efficiently generate the magnetic field over a large volume with a reduced amount of material. This minimizes the amount of multiple scattering, one of the factors limiting the muon momentum resolution.

3.2.3 Inner Detector (ID)

The ATLAS inner detector is the subsystem closest to the interaction point (IP) immersed in an axial field provided by the solenoid magnet [46]. It is designed to provide excellent momentum resolution and both primary and secondary vertex measurements by reconstructing the charged particles tracks arising from the collisions. The ID consist of three subsystems, based on different detection techniques, providing a precise and robust determination of tracks coordinates. Starting at the innermost layer, we have: the Silicon Pixel Detector (PIX) including the Insertable B-Layer (IBL), the SemiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT), as illustrated in Figure 3.7.

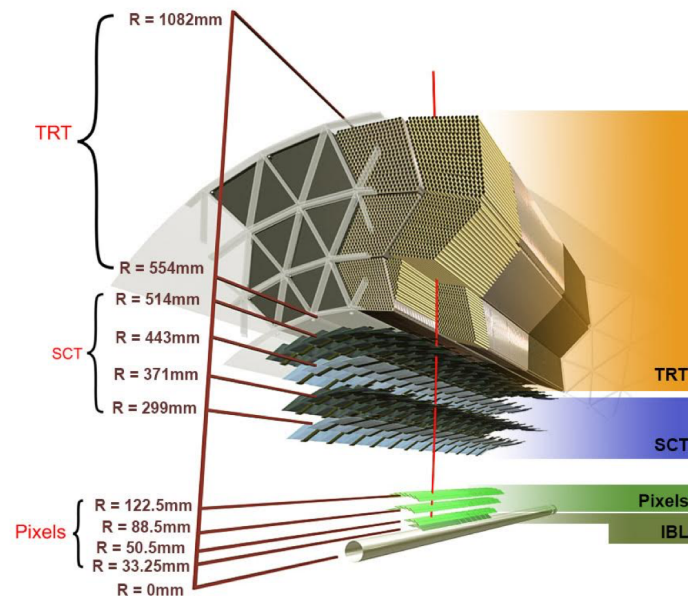


FIGURE 3.7: Scheme of a transverse section of the ID with its three sub-detectors. The IBL is a pixel detector. The radius of the subdetector components is also shown.

The ID acceptance covers the pseudorapidity range $|\eta| < 2.5$. The outer radius of the tracker region is 1.15 m, limited by the inner dimension of the Liquid Argon electromagnetic calorimeter. The total length of the ID is 7 m, limited by the position

of the end-cap calorimeters. In the barrel region, the high-precision detectors are arranged in concentric cylinders around the beam axis, while the end-cap detectors are mounted on disks perpendicular to the beam axis.

Pixel detector

The pixel detector is a silicon detector placed in the innermost part with respect to the beam line of the ID. As a consequence of its proximity to the beam pipe, it experiences the highest amount of particle flux. Each pixel of the detector has a size of $50 \times 400 \mu\text{m}^2$, resulting in approximately 80 million readout channels in total. Due to its fine granularity, it provides a high precision set of measurements close to the IP. The intrinsic spatial resolution in the barrel region is $14 \mu\text{m}$ and $115 \mu\text{m}$ in r - ϕ and z , respectively. The main function of the pixel detector is to reconstruct tracks and measure precisely the primary and displaced vertices arising from a decay. The pixel detector is subdivided into four barrel layers with radii of 33.25 mm, 50.5 mm, 88.5 mm and 122.5 mm.

The very first layer called IBL (Insertable B-layer) was installed during the first long shutdown of the LHC. Thanks to its position close to the IP (33.25 mm), the IBL improves in particular the reconstruction of secondary vertices associated with the production of long-lived particles such as b-hadrons. This information is very useful to identify jets originating from the fragmentation of b-quarks. The addition of the IBL has improved the b-tagging performance by providing high efficiency and excellent separation of b-jets from light quark jets in the presence of high pile-up.

With the IBL, the resolution of the transverse and longitudinal impact parameters (d_0 and z_0) has improved by a factor of two for low energy tracks. Furthermore, the presence of the IBL also has some implications for the reconstruction of converted photons. In particular, one-track conversions (where one electron leg is not reconstructed) are identified by requiring the absence of hits in the IBL. A precise reconstruction of primary vertices plays also a crucial role in the analysis: the correct identification of the tracks originating from a proton-proton collision is necessary to avoid pile-up contributions to the track isolation.

SemiConductor Tracker (SCT)

The SCT, a silicon micro-strip detector, is located in the intermediate ID area and consists of barrel and end-cap detectors. There are 4088 two-sided modules in total and over six million implanted readout strips of silicon distributed over the four cylindrical barrel layers and eighteen planar end-cap discs. In order to achieve high precision tracking, the modules are made of two pairs of daisy-chained silicon strip sensors, glued back to back at a stereo angle of 40 mrad. The above configuration permits the measurement of the three spatial coordinates. The intrinsic spatial resolution in the barrel region is $17 \mu\text{m}$ in r - ϕ and $580 \mu\text{m}$ in z .

Transition Radiation Tracker (TRT)

The TRT detector is the outermost part of the ID. It consists of layers of straw tubes filled with a gas mixture of Xenon. The TRT provides position information in r - ϕ plane in the barrel region and along the z -axis in the end-cap region. The intrinsic accuracy is $170 \mu\text{m}$ per straw in r - ϕ . To increase the electron identification efficiency, the space between the straws is filled with material, so that particles passing through emit transition radiation photons. Since electrons are more likely than hadrons to emit transition radiation photons at low energy, one can distinguish an electron from a pion using the TRT response. The TRT spatial resolution is lower than that of the previously described ID subdetectors. However, because of the larger volume it occupies, a greater number of hits per particle track can be recorded (around 36).

3.2.4 Calorimeters

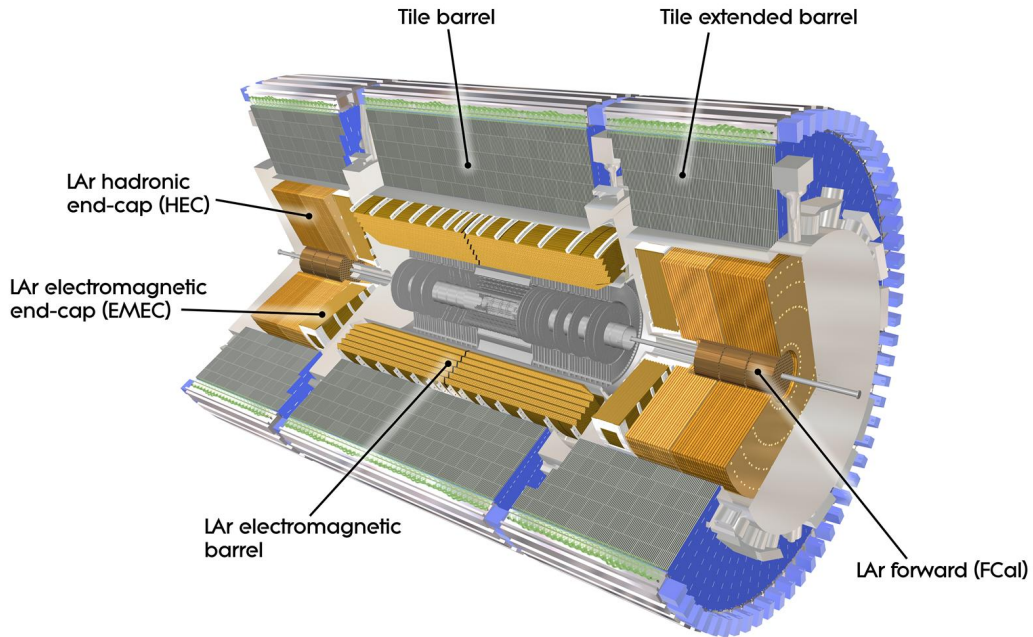


FIGURE 3.8: Schematic view of the ATLAS calorimetry system.

The ATLAS calorimeter system (Figure 3.8) consists of sampling detectors with full ϕ -symmetry. It is composed of electromagnetic (EM) and hadronic components. The calorimeters are designed to measure the energy of incoming particles by fully absorbing them. The electromagnetic calorimeter provides the reconstruction, identification and energy measurement of EM objects (electrons, positrons and photons) with high granularity, while the hadronic calorimeters are used for the measurement of hadrons (protons, neutrons, pions, etc.). The ATLAS calorimeters with their large coverage, ensure the measurement of the missing transverse energy E_T^{miss} , defined

as the modulo of the negative vector sum of the transverse momenta (p_T) of the registered particles.

The EM barrel and end-cap calorimeters housed in three cryostats, one barrel and two end-caps, cover a rapidity range of $|\eta| < 3.2$. The barrel hadronic calorimeter covers $|\eta| < 1.7$, the hadronic end-cap calorimeter covers $1.5 < |\eta| < 3.2$, and the forward calorimeter the range between $3.1 < |\eta| < 4.9$.

Electromagnetic barrel and end-cap calorimeter

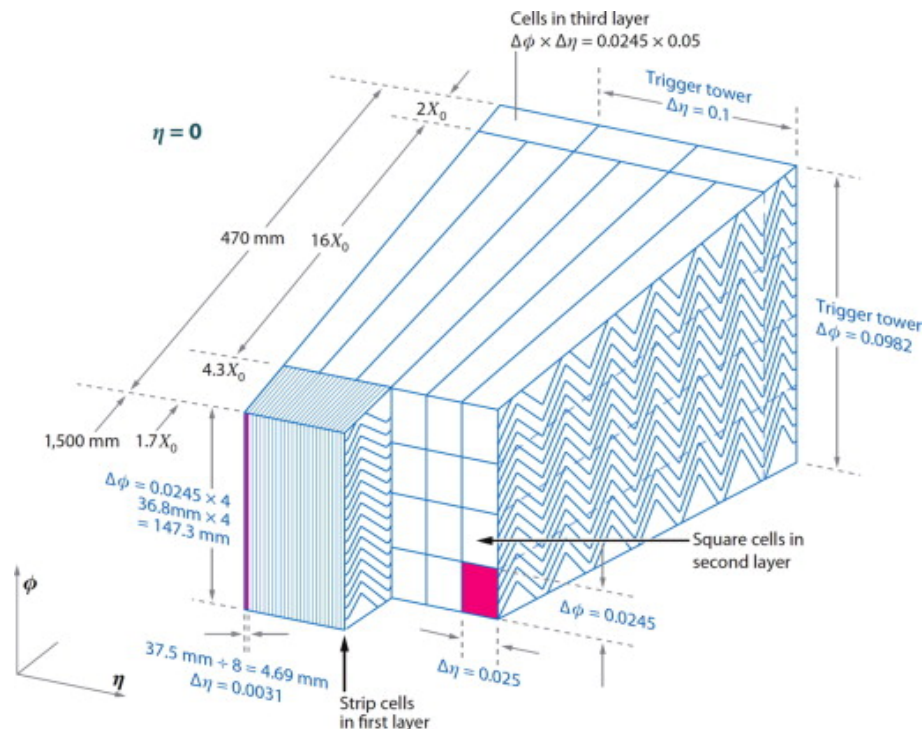


FIGURE 3.9: Sketch of a part of the barrel module where the different layers are clearly visible with the ganging of electrodes in ϕ . The granularity in η and ϕ of the cells of each of the three layers and of the trigger towers is also shown.

The electromagnetic calorimeter (ECAL), provides information about the energy, position and identity of the particle. It is a sampling calorimeter that uses lead as absorber and Liquid Argon (LAr) as active material. Lead has a short radiation length, allowing for a compact construction of the ECAL. Liquid Argon is well suited as active medium due to its ionization yield and its radiation hardness. The electromagnetic interaction of photons, electrons and positrons generates in the lead plates secondary particles mainly through pair production and bremsstrahlung. A cascade of particles called shower is produced up to the point where the electrons, positrons or photons have too little energy to produce secondary particles, losing their remaining energy through inelastic collisions with atoms (electrons and positrons), Compton scattering or photoelectric effect (photons). The form of the shower itself (width, length, location of maximum) provides information about the identity of the particle. In the active material, the Argon atoms are ionized by the charged particles in the shower, and

the ions and electrons drift to high-voltage readout electrodes situated between the absorbers. The electric signal has a triangular shape whose peak is proportional to the energy loss of the particle. The ECAL achieves an energy resolution of

$$\frac{\sigma}{E} \sim 10\%/\sqrt{E[GeV]} \oplus 0.7\%.$$

The LAr calorimeters covers pseudorapidities of $|\eta| < 1.475$ for the electromagnetic barrel (EMB) and $1.375 < |\eta| < 3.2$ for the electromagnetic end-cap (EMEC). An accordion geometry has been chosen for the absorbers and the electrodes of the electromagnetic calorimeter (Figure 3.9). The advantages of the accordion shape is that it provides a high-granularity, full azimuthal symmetry without cracks and fast signal read-out.

The calorimeter depth is about 22 radiation lengths in the barrel and 24 radiation lengths in the end-cap. In the region $|\eta| < 2.5$ the ECAL is segmented into three longitudinal sections as shown in Figure 3.9. The first layer (strips) has a very fine granularity in η mainly for π^0 - γ separation, the second layer (middle) is where most of the energy of the electrons and photons is deposited and the third layer (back) measures the energy of the tail of the shower. In the range $2.5 < |\eta| < 3.2$ the ECAL (inner wheel) is segmented into two layers and has a coarser granularity. In the region $|\eta| < 1.8$, a shower detector (presampler) is used to correct for the energy lost upstream of the calorimeter. The presampler consists of an active Liquid Argon layer of thickness 1.1 cm (0.5 cm) in the barrel (end-cap) region.

Hadronic calorimeters: TileCal and HEC

The hadronic calorimeter consists of three parts: the tile calorimeter in the barrel region (TileCal), the hadronic end-cap calorimeter (HEC) and the forward calorimeter (FCAL), extending all the way to the region close to the beam pipe. Each of these calorimeters is described in detail below. The granularity of the different layers is displayed in Table 3.1.

Scintillator Tile calorimeter		
	Barrel	Extended Barrel
$ \eta $ coverage	$ \eta < 1.0$	$0.8 < \eta < 1.7$
Number of layers	3	3
Granularity $\Delta\eta \times \Delta\phi$	0.1×0.1	0.1×0.1
Last layer	0.2×0.1	0.2×0.1
Readout Channels	5760	4092 (both sides)
LAr hadronic end-cap (HEC)		
$ \eta $ coverage		$1.5 < \eta < 3.2$
Number of layers		4
Granularity $\Delta\eta \times \Delta\phi$		0.1×0.1 $1.5 < \eta < 2.5$ 0.2×0.2 $2.5 < \eta < 3.2$
Readout Channels		5632 (both sides)

TABLE 3.1: Main parameters of Hadronic and Forward Calorimeters.

The TileCal, the central section of the ATLAS hadronic calorimeter, is a key component to detect hadrons, jets and taus, and to measure the missing transverse energy [45]. Due to the excellent muon signal-to-noise ratio, it assists the spectrometer in the identification and reconstruction of muons. The TileCal is built of steel and scintillating tiles coupled to optical fibers and read out by photomultipliers. The calorimeter is equipped with systems that allow to monitor and calibrate each stage of the readout system by exploiting different signal sources: laser light, charge injection and a radioactive source. Its hadron energy resolution is given by:

$$\frac{\sigma}{E} \sim 50\%/\sqrt{E}[\text{GeV}] \oplus 3\%. \quad (3.7)$$

The response linearity lies within approximately 1% up to a few TeV energies. TileCal has a fixed central barrel (LB) and two movable extended barrel (EB). The scintillating tiles are placed in the plane perpendicular to the colliding beams and are radially staggered in depth.

The calorimeter modules are segmented into four longitudinal layers (A, B, C, D) of thickness in unit of interaction length λ : 1.5, 4.1 and 1.8 in the barrel and 1.5, 2.6, 3.3 in the extended barrels, with a total depth of 7.4 λ at $\eta = 0$. Each cell has a transversal segmentation of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ (0.2×0.1 in the outer D layer). In addition to the regular cells, gap/crack scintillators have been installed in front of the TileCal extended barrel and LAr electromagnetic end-cap calorimeters to improve the performance in the region occupied by many services and electronics from the inner tracker of the LAr calorimeter.

The HEC [47] ensures the hadronic particle detection in the region $1.5 < |\eta| < 3.2$ and therefore overlaps with the TileCal at its edges. Two independent wheels join the end-cap region directly behind the ECAL and share the same LAr cryostat with the latter. It is a sampling calorimeter with a copper absorber and Liquid Argon as active material. Each hadronic end-cap calorimeter consists of two independent wheels of outer radius 2.03 m. The first wheel is built out of 25 mm copper plates, while the second one uses 50 mm plates. In both wheels the 8.5 mm gap between consecutive copper plates is equipped with three parallel electrodes, splitting the gap into four drift spaces of about 1.8 mm. The hadronic end-cap calorimeter achieves roughly the same energy resolution as the TileCal (Equation 3.7).

LAr forward calorimeter (FCAL)

The FCAL, comprised of an electromagnetic and a hadronic calorimeter, extends the ATLAS coverage to the most forward regions ($3.1 < |\eta| < 4.9$) and allows the measurement of particles at very low angles with respect to the beam axis. Its first layer, a copper/LAr sampling calorimeter, is optimized for electromagnetic measurements, while the second and the third layers are optimized for hadronic measurements. Since the FCAL front face is set back by about 4.7 m with respect to the interaction point,

the calorimeter has to be kept more compact and contain the hadronic showers despite the smaller available space. Therefore, tungsten is used as an absorber material. Liquid Argon is the active material and it shares the same cryogenic vessel with the HEC. The FCAL's importance lies in the fact that it guarantees the hermeticity of the calorimetry necessary for the determination of E_T^{miss} .

3.2.5 Muon Spectrometer (MS)

The MS is devoted to the identification of muons and to the measurement of their charge and momentum. The measurement is based on the magnetic deflection of tracks in the large superconducting air-core toroid magnets. The MS [45] is the outermost ATLAS subdetector located right after the calorimeters. It covers the pseudorapidity range $|\eta| < 2.7$, and is composed of precision chambers and trigger chambers that cover up to $|\eta| < 2.4$. Muons loose about 3 GeV from the interaction in the material in the calorimeters. Hence, the MS measured momenta values are approximately 3 GeV lower than their original momenta.

Resistive plate chambers (RPC, three doublet layers in the barrel, $|\eta| < 1.05$) and Thin Gap Chambers (TGC, one triplet layer followed by two doublets in the end-caps, $1.0 < |\eta| < 2.4$), provide triggering capability, as well as (η, ϕ) position measurements with typical spatial resolution of 5-10 mm [48]. A precise momentum measurement for muons with pseudorapidities up to $|\eta| = 2.7$ is provided by three layers of Monitored Drift Tube chambers (MDTs), with each chamber providing six to eight η measurements along the muon trajectory. For $|\eta| > 2$, the inner layer is equipped with a quadruplet of cathode strip chambers (CSC) instead of MDTs.

Monitored drift tubes (MDT)

The MDTs provide a measurement of space points with high accuracy. They cover a pseudorapidity range of $|\eta| < 2$. The drift tubes are filled with pressurized Ar/CO_2 gas, which is ionized by charged particles when they transverse the tube. The chambers are composed of several layers from three to eight depending on the distance to the interaction point. The ionization charge is collected with a high potential anode. The single MDT hit resolution is about 80 μm .

Cathode strip chambers (CSC)

In the innermost layer of the forward region ($2 < |\eta| < 2.7$) the MDTs are replaced with CSCs. The rationale behind this is that when the particle flow exceeds the rate of 150 Hz/cm², the MDTs will not longer be in a safe functioning mode. CSCs guarantee a safe operation for rates up to 1000 Hz/cm². A CSC is a multiwire proportional chamber with wires oriented in the radial direction. It consist in two large disks with eight chambers each. Each chamber is built of four planes, allowing for four independent (η, ϕ) measurements for each muon track. Similarly to MDTs, the CSCs

also have anode wires and the strips act as the cathode. The single hit resolution of a CSC is about $60 \mu\text{m}$.

Resistive plate chambers (RPC) and the thin gap chambers (TGC)

The RPCs in the barrel module of the MS and the TGCs in the end-cap region of the MS are part of the triggering system of the MS. The RPCs are arranged in three concentric cylinders around the beam line - the trigger stations - that, in addition to triggering, also provide measurements of a track in the barrel region. Each station is composed of two detector layers that permit a separate (η, ϕ) measurement. The TGCs also provide trigger capacity and measurement of the coordinate in the ϕ direction, which completes the measurement of the MDT in the η direction. The TGCs are composed of multiwire proportional chambers with an excellent time resolution.

3.2.6 Trigger System

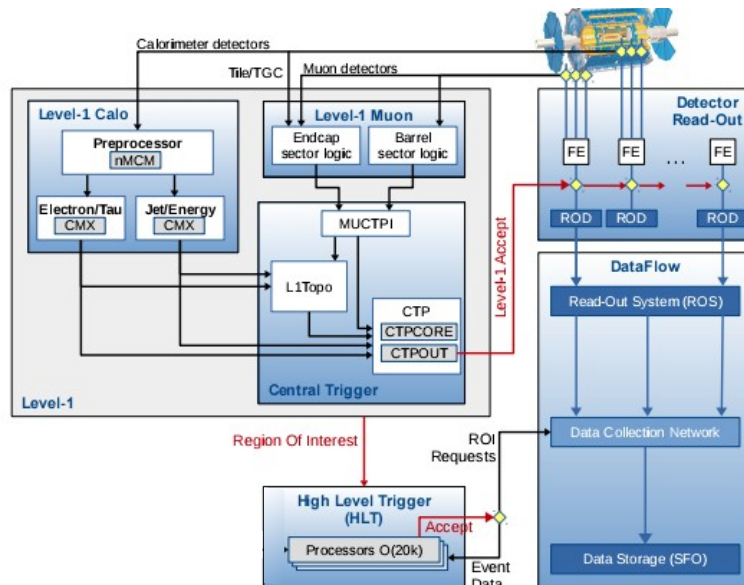


FIGURE 3.10: Schematic layout of the ATLAS trigger and data acquisition system in Run II [49].

An essential component for the data acquisition (DAQ) system of ATLAS is the trigger system. Its role is to take the decision of whether or not to record a given collision event for later study. During the LHC Run I, the ATLAS trigger system operated efficiently at instantaneous luminosities of up to $8 \times 10^{33} \text{cm}^{-2} \text{s}^{-1}$, at center-of-mass energies between 900 GeV - 8 TeV, and collected more than three billion events. In the LHC Run II, the collision energy was increased to 13 TeV. Higher luminosity and higher energy lead to a higher interaction rate; therefore, the trigger selections rates had to be increased by up to a factor of five compared to Run I. The current (Run II) trigger system [49] consists of two levels for event selection: a hardware-based first level trigger (L1) and a software-based high-level trigger (HLT), as shown in Figure 3.10. The L1 trigger is the initial step to identify the interesting

events before proceeding to the following selection stage. The L1 trigger uses custom electronics to determine regions of interest (RoIs) in the detector, taking as input the signals with coarse granularity from the calorimeter and muon detector. With decision times of less than $2.5 \mu\text{s}$, the L1 trigger reduces the event rate from the LHC bunch crossing rate from approximately 30 MHz to 100 kHz. The L1 trigger identifies event features such as high p_T leptons, photons, jets and missing transverse energy. The RoIs formed at L1 are sent to the HLT, in which sophisticated selection algorithms, using full granularity detector information, are run. At this stage, within 200 ms the event rate is reduced from 100 kHz to approximately 1 kHz on average. The events which are accepted by the HLT are transferred to the local storage space located at the experimental site, and then exported to the Tier-0 facility at CERN's computing center for offline reconstruction.

3.3 ATLAS Monte Carlo simulations

The Monte Carlo detector simulation is an important tool used in high energy physics to mimic the response of a detector to individual particles emerging from collision events. ATLAS developed a very detailed description of the detector geometry. The physics analysis Monte Carlo (MC) [50] simulations are used to predict the outcome of an interaction process and allow to compare observations to expectations.

The first step in the simulations is the generation of events, followed by the decays of short lived particles, using various sets of generators such as PYTHIA [51], SHERPA [52], MADGRAPH [53] or POWHEG [54]. Starting from a proton-proton initial state, each generated event contains the particles resulting from the *hard scattering*. PDFs are used in order to describe the structure of the proton. Generally, the matrix element of the *hard scattering* is calculated at LO or NLO in QCD. *Parton showers* are included by adding QCD and QED radiations from the partons. In many cases, the generators also simulates the *underlying event* using non perturbative models. As a last step the *hadronization* takes place, which conclude the process of the complete event generation. The combination of the processes listed above represents the *generated event*. Afterwards, the simulation of the detector response to the particle from the event is performed, using the GEANT4 simulation toolkit [55]. The ATLAS software converts the hits produced by the simulation into detector responses. As a final step, both data events and simulated events are brought into the same format, with the simulated events containing extra information from the parton and particle level event.

Both the simulated events and the data are run through the same ATLAS reconstruction. This makes a detailed comparison of the two possible. The ATHENA framework [56] is the main software infrastructure of the ATLAS collaboration and is used for all the steps involved in MC simulation, data reconstruction, and to some extent also the physical analysis.

Chapter 4

Electron performance studies

Μηδέν είναι μήτε τέχνην άνευ μελέτης μήτε
μελέτην άνευ τέχνης.

Πρωταγόρας, 487-412 π.Χ.

Electrons¹ play a crucial role in many analyses of the ATLAS experiment and in particular in the study of the $Z\gamma$ vector boson scattering with a Z boson decaying to an electron and positron pair. Electrons give rise to tracks in the tracking detector and energy deposits in the electromagnetic calorimeter. The electron online selection and the electron reconstruction aim to combine the track with the energy deposit providing high efficiency. In order to differentiate electrons from other particles several sets of identification criteria are developed. These criteria rely on the quality of the tracks, the shapes of the electromagnetic showers and track-to-cluster matching. Electron identification aims to provide high efficiency and rejection against other particles (background). In this chapter the electron trigger, reconstruction and identification algorithms used by the ATLAS Collaboration are briefly described as well as identification efficiency measurements. The electron identification measurements are a central part of my thesis work and are thus described in more detail here.

4.1 Electron trigger

The ATLAS online data processing reconstructs and identifies electron candidates both at L1 trigger and at the HLT. The electron triggers at L1 [57] use the signals recorded in both the ECAL and HCAL within regions of interest (RoI) of 4×4 trigger towers corresponding to $\Delta\eta \times \Delta\phi \approx 0.4 \times 0.4$. Different E_T thresholds are used in the trigger decision. If the energy measured in the hadronic calorimeter behind the core of the EM cluster is less than a given threshold, L1 "vetoes" against the hadronic leakage. Moreover, an isolation cut using the transverse energy in an annulus of the calorimeter towers around the EM candidate relative to the EM cluster transverse energy may be also used.

The electron candidates at the HLT are reconstructed and selected in several steps in order to reject potential background candidates early, in a way that reduces

¹In this chapter electrons and positrons are often collectively referred to as electrons.

the event rate to a level where more precise algorithms can be applied in the allowed latency range. Clusters are build from the calorimeter cells within the RoI identified at the L1 step. Then tracks are reconstructed using a simplified track algorithm, with a minimum p_T of 1 GeV and are associated to clusters within $|\Delta\eta| < 0.2$. Subsequently, among these objects, electron candidates are reconstructed as clusters matched within $|\Delta\eta| < 0.05$ and $|\Delta\phi| < 0.05$ to tracks extrapolated to the second layer of the EM calorimeter. The identification of electron candidates at the HLT is performed using the same discriminating variables as for the offline identification, combined in a Likelihood (LH) method. The cut on the LH identification discriminant is adjusted as a function of the average number of interactions per crossing to account for pile-up effects.

4.2 Electromagnetic shower development

When an electron (or a photon) crosses an electromagnetic calorimeter, will lose its energy through interactions with the material of the calorimeter. These interactions, mainly within the lead absorbers, create an electromagnetic shower. The charged component of the shower ionises the LAr. The summed ionization current measured by the calorimeter is proportional to the incident particle energy.

Depending on the energy of the electromagnetic particle, different processes take place. For high energies ($E > 10$ GeV) the two dominant processes by which electrons and photons respectively interact with matter are the bremsstrahlung emission, and electron-positron pair production. The electromagnetic showers produced by secondary electrons, positrons and photons, are stopped when the cascade electrons reach an energy where the energy loss by ionization starts to be dominant, or when the cascade photons are not able to create pairs any more ($E \lesssim 1.2$ MeV). An important parameter which characterizes the shower development is the radiation length, X_0 , defined as the mean distance over which an electron is left with 1/e of its incoming energy. It is necessary to keep the amount of radiation length in front of the calorimeter as small as possible, to avoid an early shower development, which can lead to unaccounted energy loss before the particle will enter the calorimeter.

The response of the ECAL to the electromagnetic shower forms clusters which are then used in the reconstruction of electromagnetic particles.

4.3 Electron and photon reconstruction

The electron offline reconstruction begins with seed-cluster reconstruction in the ECAL, using a sliding window algorithm and track reconstruction with an electron specific track fit, including an optimized Gaussian-Sum Filter (GSF) [59],[60] to better account for energy loss of charged particles. Afterwards, tracks are matched in η and ϕ to the cluster and the determination of the final cluster size is carried out. Finally, the four-momentum of the candidate electron is computed using information

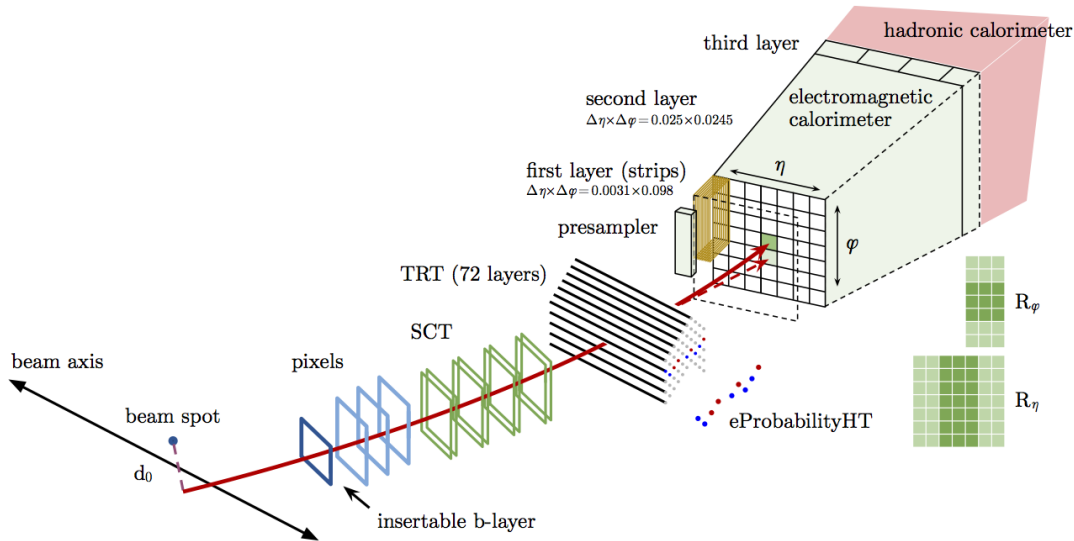


FIGURE 4.1: Schematic illustration of the path of an electron through the ATLAS detector. The red trajectory shows the hypothetical path of an electron, which first traverses the tracking system (pixel detectors, then silicon-strip detectors and lastly the TRT) and then enters the electromagnetic calorimeter. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron with the material of the tracking system (bremsstrahlung) [58].

from both the final calibrated energy cluster and the best track matched to the originated seed cluster. The energy is given by the final calibrated cluster, while the η and ϕ directions are obtained from the corresponding track parameters. Figure 4.1 provides a schematic illustration of the elements that enter into the reconstruction of an electron.

The building of the cluster for the photon reconstruction [61] is performed in a similar way as for the electrons. For photons, no tracks are expected in the tracker unless the photon has converted. In this case one track or two tracks forming a conversion vertex are associated to the cluster.

4.4 Electron identification

To classify the reconstructed electron candidate as signal or background object, algorithms for electron identification are applied. The baseline identification algorithm used in Run II data analyses relies on a Likelihood method. It is an analysis technique that combines several properties of the electron candidate into an estimator which is used to take a decision about the nature of the candidate.

The properties of the electron candidate are related to the tracker and cluster measurements including calorimeter shower shapes, track-cluster matching related quantities, and variables measuring bremsstrahlung effects. These quantities are summarized in Table 4.1 and some of them are illustrated schematically in Figure 4.1.

Likelihood functions of signal (\mathcal{L}_S) and background (\mathcal{L}_B), are constructed as the product of n probability density functions (pdfs), for signal (P_S) and background

Type	Description	Name	Rejects			Usage
			LF	γ	HF	
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $ \eta < 0.8$ or $ \eta > 1.37$)	R_{had1}	x	x		LH
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \eta < 1.37$)	R_{had}	x	x		LH
Third layer of EM calorimeter	Ratio of the energy in the third layer to the total energy in the EM calorimeter. This variable is only used for $E_T < 80$ GeV, due to inefficiencies at high E_T , and is also removed from the LH for $ \eta > 2.37$, where it is poorly modelled by the simulation.	f_3	x			LH
Second layer of EM calorimeter	Lateral shower width, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	$w_{\eta 2}$	x	x		LH
	Ratio of the energy in 3×3 cells over the energy in 3×7 cells centred at the electron cluster position	R_ϕ	x	x		LH
	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centred at the electron cluster position	R_η	x	x	x	LH
First layer of EM calorimeter	Shower width, $\sqrt{(\sum E_i (i - i_{\text{max}})^2)/(\sum E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η , and i_{max} is the index of the highest-energy strip, used for $E_T > 150$ GeV only	w_{stot}	x	x	x	C
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies	E_{ratio}	x	x		LH
	Ratio of the energy in the first layer to the total energy in the EM calorimeter	f_1	x			LH
Track conditions	Number of hits in the innermost pixel layer	n_{Blayer}		x		C
	Number of hits in the pixel detector	n_{Pixel}		x		C
	Total number of hits in the pixel and SCT detectors	n_{Si}		x		C
	Transverse impact parameter relative to the beam-line	d_0		x	x	LH
	Significance of transverse impact parameter defined as the ratio of d_0 to its uncertainty	$-d_0/\sigma(d_0)-$		x	x	LH
	Momentum lost by the track between the perigee and the last measurement point divided by the momentum at perigee	$\Delta p/p$	x			LH
TRT	Likelihood probability based on transition radiation in the TRT	eProbabilityHT	x			LH
Track-cluster matching	$\Delta\eta$ between the cluster position in the first layer and the extrapolated track	$\Delta\eta_1$	x	x		LH
	$\Delta\phi$ between the cluster position in the second layer of the EM calorimeter and the momentum-rescaled track, extrapolated from the perigee, times the charge q	$\Delta\phi_{\text{res}}$	x	x		LH
	Ratio of the cluster energy to the track momentum, used for $E_T > 150$ GeV only	E/p	x	x		C

TABLE 4.1: Quantities used in the electron identification. The columns labeled "Rejects" indicate whether a quantity has significant discrimination power against light-flavor (LF) jets, photon conversions (γ), or non-prompt electrons from the semileptonic decay of hadrons containing heavy-flavor (HF) quarks. In the column labeled "Usage" an "LH" indicates that the probability density function of this quantity is used in the likelihood method and a "C" indicates that this quantity is used directly as a selection criterion [58].

(P_B):

$$\mathcal{L}_{S(B)}(\vec{x}) = \prod_{i=1}^n P_{S(B),i}(x_i). \quad (4.1)$$

In the above expression, x_i indicates the values of each discriminating property as specified in Table 4.1, while $P_{S,i}(x_i)$ and $P_{B,i}(x_i)$ are the values of the signal and background pdfs respectively, for quantity i at value x_i . Whereas the signal consist of

prompt² electrons, the background includes jets that mimic the signature of prompt electrons, electrons from photon conversions in the detector material, and non-prompt electrons from the decay of hadrons containing heavy flavors. The chosen electron properties are mostly uncorrelated with each other in such a way that the LH can be expressed as the product of pdfs of the corresponding properties. Any residual correlation is neglected. The signal and background likelihood for a given electron are then combined into a discriminant $d_{\mathcal{L}}$ on which selected criteria are applied.

This discriminant $d_{\mathcal{L}}$ defined as:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}, \quad (4.2)$$

presents a sharp peak at one (zero) for signal (background), which makes it inconvenient for the working point definition that are used in the analyses as it would require an extremely fine binning. For that reason an inverse sigmoid function is used to transform the distribution of the discriminant of Equation 4.2 as follows:

$$d'_{\mathcal{L}} = -\tau^{-1} \ln(d_{\mathcal{L}}^{-1} - 1), \quad (4.3)$$

where the parameter τ is fixed to 15 [62]. As a consequence, the range of values of the transformed discriminant no longer varies between zero and unity. For each working point, a threshold value of the transformed discriminant is chosen, and electron candidates for which $d'_{\mathcal{L}}$ is above the threshold are considered as signal.

There are two main advantages of using a LH-based electron identification over a cut-based identification. First, a prompt electron that may fail the cut-based identification because it does not satisfy the selection criterion for a single quantity can still pass the identification criteria in a LH-based selection, because the LH selects the candidate using a combination of many discriminating quantities with different optimal weights. Second, the use of discriminating quantities with very similar distributions between signal and background results in large losses in efficiency when they are used in a cut-based identification, while they may be used in a LH-based identification without penalty. Two examples of quantities that are used in the LH-based identification, but not in the cut-based identifications, are R_{ϕ} and f_1 , which are defined in Table 4.1. In Figure 4.2 the distributions for these two quantities for signal and background are compared.

The pdfs are derived for each E_T and η bins (Tables 4.2 and 4.3) from finely binned simulation histograms of the individual identification quantities. These, in turn, are smoothed using an adaptive kernel density estimation (KDE) [62] in order to avoid fluctuations due to the limited statistics of the simulations samples.

As a consequence of the imperfect detector modeling, differences between the simulated quantities and the corresponding quantities in data cannot be fully eliminated. To overcome this problem, corrections to the simulated quantities are performed for

²Prompt leptons are defined as leptons that do not originated from hadron decays or photon conversion.

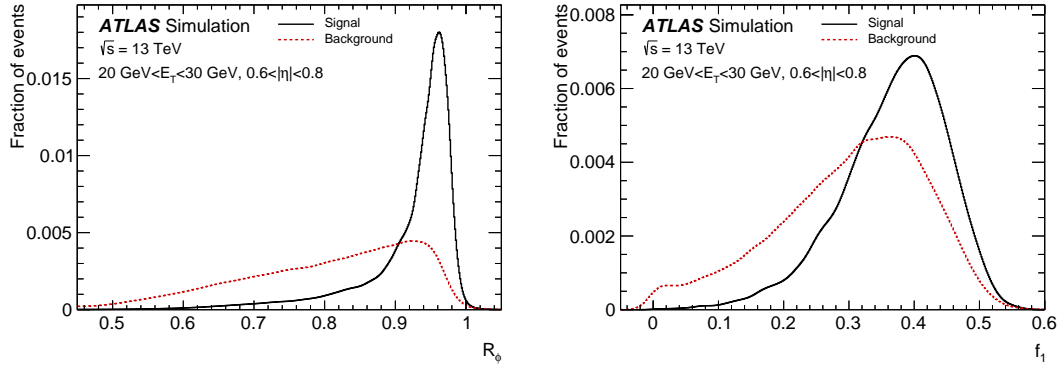


FIGURE 4.2: Distributions of $R\phi$ and f_1 for $20 \text{ GeV} < E_T < 30 \text{ GeV}$ and $0.6 < |\eta| < 0.8$. In each figure, the red-dashed distribution is determined from a background simulation sample and the black-line distribution is determined from a $Z \rightarrow ee$ simulation sample. These distributions are for reconstructed electron candidates before applying any identification [58].

some of the variables so that the simulation reproduces the data as closely as possible. These corrections are determined by comparing samples of simulated and data electrons, using the electronic decay of the Z boson. Two main effects typically characterize the difference between data and simulation. One is a shift of the distribution (offsets) and the other is a difference in the width. There are also some cases where both shift and width corrections are applied.

The optimal values of the shifts and of the width-scaling factors are determined by minimizing a χ^2 that compares the distributions in data and simulation. When a shift needs to be applied, the simulated distribution is shifted by a fixed amount to improve its agreement with data. When a difference in the width is observed, each entry in the simulation is scaled by a multiplicative factor.

In order to satisfy the needs of the different physics analyses carried out within the ATLAS Collaboration, three fixed values of the $d'_{\mathcal{L}}$ discriminant are used to define three main working points. These are referred to, in order of increasing background rejection, as *Loose*, *Medium* and *Tight*. These working points are defined such that the samples selected by a given criteria are subsets of the sample selected by the less restrictive criteria. Each working point uses the same variables to define the LH discriminant, but the value of the selection cut on this discriminant is different for each working point. This value is optimized in bins of cluster η (Table 4.2), which are defined according to the calorimeter geometry, detector acceptances and the variation of the amount of material in the inner detector, and in bins of the electrons E_T (Table 4.3) defined according to the variations of the pdfs of the various electron-identification quantities. In order to have a relatively smooth variation of electron-identification efficiency with E_T , the value of the cut on the discriminant is varied in finer bins than for the pdfs. The bin edges are specified in Table 4.3.

Bin boundaries in $ \eta $										
0.0	0.1	0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37	2.47

TABLE 4.2: Low edges and the upper edge (the last) of the bin boundaries in $|\eta|$.

	Bin boundaries in E_T [GeV]														
PDFs	4.5	7	10	15	20	30	40							∞	
Discriminant	4.5	7	10	15	20	25	30	35	40	45	50	60	80	150	∞

TABLE 4.3: Low edges and the upper edge (the last) of the bin boundaries in E_T .

All working points apply the same requirements on tracking variables, independently of η and E_T : the *Loose*, *Medium*, and *Tight* working points require at least two hits in the pixel detector and seven hits in total in the pixel and silicon-strip detectors combined. For the *Medium* and *Tight* working points, one of these pixel hits must be in the innermost pixel layer (or in the next-to-innermost layer if the innermost layer is non-operational). This requirement helps to reduce the background from photon conversions. A variation of the *Loose* working point -*LooseAndBLayer*- uses the same LH discriminant cut value as the *Loose* one, and adds this requirement on the innermost pixel layers. A *VeryLoose* working point is also defined which does not include an explicit requirement on the innermost pixel layer and requires only one hit in the pixel detector, in order to provide relaxed identification requirements for background studies.

The efficiencies of the LH-based electron identification for the *Loose*, *Medium*, and *Tight* working points for data as well as the corresponding data-to-MC ratios are presented in Figure 4.3.

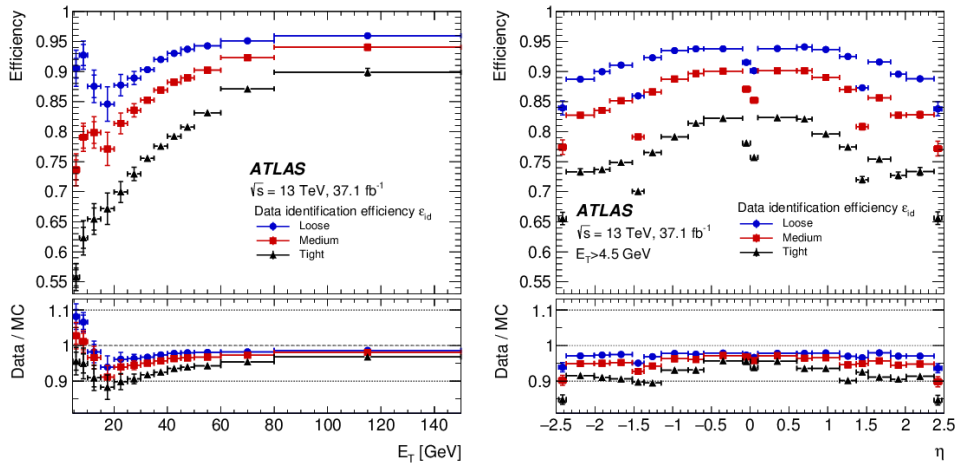


FIGURE 4.3: The measured LH electron-identification efficiencies in $Z \rightarrow ee$ events for the *Loose*, *Medium*, and *Tight* working points as a function of E_T (left plot) and as a function of η (right plot). The vertical bars represent the statistical (inner bars) and the total (outer bars) uncertainties [58].

4.5 Electron isolation

One of the challenges of the LHC experiments is to disentangle the prompt production of electrons arising from the decay of heavy resonances (such as W and Z bosons) from background processes (like the semileptonic decays of heavy quarks) and background particles (e.g. hadrons misidentified as leptons, or electrons from photon conversion

in the detector material upstream of the electromagnetic calorimeter). The characteristic signature of a prompt electron signal shows relatively little activity (both in the calorimeter and in the ID) in the area surrounding the candidate object.

To this extent variables are constructed that quantify the amount of activity in the vicinity of the candidate object. These variables, called *isolation variables*, are computed by summing the transverse energy of topological clusters in the calorimeter ($E_{T,\text{topocluster}}$), or else by adding the transverse momentum of the tracks in a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the direction of the electron candidate, excluding the candidate itself. These isolation variables quantify therefore the energy of the particles produced around the electron candidate and allow to distinguish prompt electrons from non-isolated electron candidates. Discriminating isolation variable based on tracking and calorimeter measurements, are described more in detail below.

For the **calorimeter-based isolation** ($E_T^{\text{cone},\Delta R}$) a cone of size ΔR is build around the direction of candidate electron. The energies of all topological clusters, whose barycenters fall within a cone of radius ΔR (as shown in Figure 4.4) are summed. The core energy E_T , is subtracted by removing the cells included in a $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ rectangle around the direction of the candidate.

$$E_T^{\text{cone},\Delta R} = \left(\sum_{i \in \Delta R} E_{T,\text{topocluster}} \right) - E_T. \quad (4.4)$$

Typically, $\Delta R = 0.2$. An (E_T, η) dependent correction is then applied to account for the electron energy leakage in the isolation cone. The contribution from pile-up and the underlying event activity is also corrected for on an event-by-event basis, using the technique described in [63]. These corrections may cause a slightly negative value of $E_T^{\text{cone},\Delta R}$. The distribution of this variable peaks around zero for isolated electrons, and around positive values for non-isolated ones. This characteristic helps to define a signal and a background dominated region.

The **track-based isolation** ($p_T^{\text{cone},R_{max}}$): is defined as the sum of transverse momenta of all tracks with $p_T > 1$ GeV that satisfy certain quality requirements inside a cone ΔR around the candidate electron track excluding the electron associated track. These tracks are required to originate from the reconstructed primary vertex of the hard collision. The cone size is defined as:

$$\Delta R = \min \left(\frac{10 \text{ GeV}}{p_T [\text{GeV}]}, R_{max} \right), \quad (4.5)$$

where R_{max} is an upper limit, typically ranging from 0.2 to 0.4. The value of 10 GeV in the numerator is optimized to reject the background from simulated $t\bar{t}$ events.

In order to minimize the impact of pile-up, a requirement is placed on the longitudinal impact parameter z_0 , corrected for the reconstructed position of the primary

vertex and multiplied by the sine of the track polar angle: $|z_0 \sin \theta| < 3$ mm. This requirement aims to select tracks that originate from the vertex that is chosen to be the relevant vertex of the process.

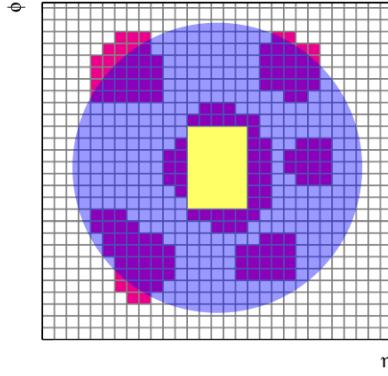


FIGURE 4.4: Schematic representation of the isolation procedure. The grid represents the second-layer calorimeter cells in the η and ϕ directions. The candidate electron is located at the center of the blue circle, which represents the isolation cone. All topological clusters (represented in red) for which the barycentres lie inside the isolation cone are included in the computation of the isolation variable. The 5×7 cells (which cover an area of $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$), represented by the yellow rectangle correspond to the subtracted cells [58].

To match the needs of the different ATLAS analyses, several isolation operating points have been established that use either calorimeter-based isolation with a cone of radius $\Delta R = 0.2$ typically, or track-based isolation using a variable-cone-size with $R_{max} = 0.2$ or 0.4 , or both types of isolation simultaneously. The requirements for each working point are established in bins of electron p_T and η .

The working points are defined in three categories in order to achieve:

- a fixed value of the isolation efficiency ϵ_{iso} , namely constant in the p_T and η of the electron candidate (*Loose* isolation),
- a fixed value of the isolation efficiency ϵ_{iso} , in the η but dependent on the p_T of the electron candidate (*Gradient* isolation),
- fixed requirements on the value of the isolation variables (*Fix* isolation).

Table 4.4 summarizes some of the electron isolation working points. The values of the requirements given in Table 4.4 are determined with simulation, using a $J/\psi \rightarrow ee$ sample for $E_T < 15$ GeV and a $Z \rightarrow ee$ sample for $E_T > 15$ GeV. Figure 4.5, shows the isolation efficiencies measured in data and the corresponding data-to-MC ratios as a function of the electron E_T and η for some of the working points given in Table 4.4 and for candidate electrons satisfying *Tight* identification requirements. The overall difference between data and simulation is less than approximately $\pm 2\%$.

Operating point	$E_T^{cone} (\Delta R = 0.2)$	$p_T^{cone} (R_{max} = 0.2)$	Total ϵ_{iso}
<i>Loose</i> (track only)	—	$\epsilon_{iso} = 99\%$	99%
<i>Loose</i>	$\epsilon_{iso} = 99\%$	$\epsilon_{iso} = 99\%$	98%
<i>Gradient</i>	$\epsilon_{iso} = 0.1143 \times p_T + 92.14\%$	$\epsilon_{iso} = 0.1143 \times p_T + 92.14\%$	90(99)%at25(60) GeV
<i>Gradient (Loose)</i>	$\epsilon_{iso} = 0.057 \times p_T + 95.57\%$	$\epsilon_{iso} = 0.057 \times p_T + 95.57\%$	95(99)%at25(60) GeV
<i>Fix (Loose)</i>	$\frac{E_T^{cone}}{E_T} < 0.20$	$\frac{p_T^{cone}}{p_T} < 0.15$	—
<i>Fix (Tight)</i>	$\frac{E_T^{cone}}{E_T} < 0.06$	$\frac{p_T^{cone}}{p_T} < 0.06$	—
<i>Fix (Tight, track only)</i>	—	$\frac{p_T^{cone}}{p_T} < 0.06$	—
<i>Fix (calo only)</i>	$E_T^{cone} < 3.5 \text{ GeV}$	—	—
<i>Fix (track $R_{max} = 0.4$)</i>	$\frac{E_T^{cone}}{E_T} < 0.11$	$\frac{p_T^{cone}}{p_T} < 0.06$	—

TABLE 4.4: Definition of the electron-isolation operating points and isolation efficiency. The units of E_T and p_T are GeV. All working points shown in the table use a cone size of $\Delta R = 0.2$ for calorimeter isolation and $R_{max} = 0.2$ for track isolation, except for the final entry *Fix* (track) which uses $R_{max} = 0.4$.

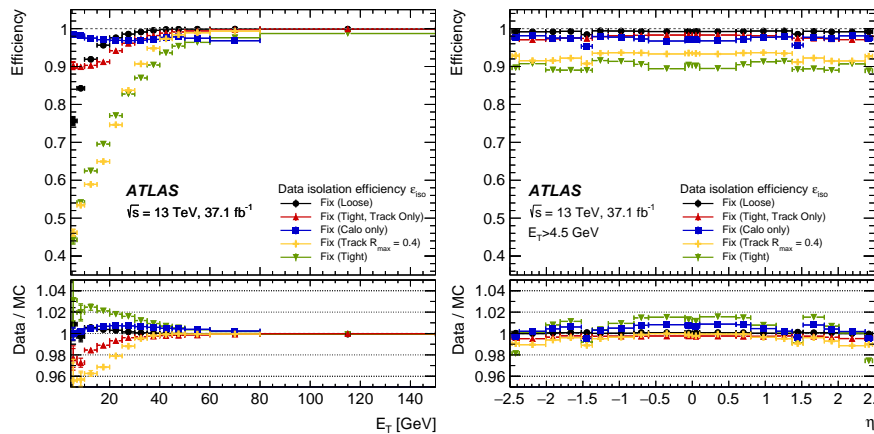


FIGURE 4.5: Isolation efficiencies derived from a $Z \rightarrow ee$ data sample as a function of electron E_T and η for $E_T > 4.5 \text{ GeV}$, for the different Fixed working points. The lower panels show data-to-MC ratios [58].

4.6 Efficiency measurement methodology

In SM measurements and searches for physics beyond the SM, the experimentally determined electron spectra must be corrected for selection efficiencies related to the trigger, as well as to particle reconstruction, identification and isolation. For electrons, these efficiencies are estimated directly from data, using the tag-and-probe method. The method employs events with well-known resonances decaying to electron-positron pairs, namely $Z \rightarrow ee(\gamma)$ and $J/\psi \rightarrow ee$. The events are selected on the basis of the electron-positron invariant mass or the isolation distribution of the probe electron in the signal mass window around the Z -boson peak. Then the method selects unbiased samples of electrons (probes) by placing strict selection requirements on the companion object (tags) produced in the decay. The efficiency of a given identification or isolation requirement is determined by applying this requirement to the probe sample after subtracting the residual background contamination. To avoid biased results, each combination of electron tag-probe pairs in an event is considered, such that an electron can be the tag in one pair and the probe in another. Similar procedures are used to measure reconstruction, identification, isolation and trigger efficiencies. The tag-and-probe method described here covers a range of electron transverse momentum, E_T ,

from 7 to 150 GeV and $|\eta| < 2.47$. The low E_T range is studied by using $J/\psi \rightarrow ee$ (from 7 to 20 GeV) and $Z \rightarrow ee\gamma$ (from 10 to 30 GeV, focusing in the 10-15 GeV bin), while $Z \rightarrow ee$ events are used for transverse energies above 15 GeV. The total detection efficiency for a single electron can be expressed as:

$$\begin{aligned} \epsilon_{\text{total}} &= \epsilon_{\text{EMcluster}} \times \epsilon_{\text{reconstruction}} \times \epsilon_{\text{identification}} \times \epsilon_{\text{isolation}} \times \epsilon_{\text{trigger}} \\ &= \left(\frac{N_{\text{cluster}}}{N_{\text{all}}} \right) \times \left(\frac{N_{\text{reco}}}{N_{\text{cluster}}} \right) \times \left(\frac{N_{\text{ID}}}{N_{\text{reco}}} \right) \times \left(\frac{N_{\text{iso}}}{N_{\text{ID}}} \right) \times \left(\frac{N_{\text{trigger}}}{N_{\text{iso}}} \right). \end{aligned} \quad (4.6)$$

In the majority of the ATLAS analyses, the electron spectra are not corrected directly for the efficiency determined from the data. Instead, a different procedure is used. The different efficiencies are measured in both data and MC by applying the same selection criteria for the tag and the probe selection and the ratio of data-to-MC efficiencies is used as a correction factor called "scale factor". Usually the scale factors are close to unity. Deviations stem from the imperfect modeling of tracking properties or shower shapes in the calorimeters. Using the ratio of efficiencies, residual effects coming from differences of the measured physics processes used ($Z \rightarrow ee$, $Z \rightarrow ee\gamma$ or $J/\psi \rightarrow ee$) are expected to cancel out in good approximation. In order to estimate the impact of the values of cut choices and potential imperfections in the background modeling on the scale factors, different variations of the efficiency measurement are carried out by modifying for example the selection of the tag electron or the background estimation method. The same variations of the selection are applied consistently in data and MC. The central value of a given efficiency measurement using one of the $Z \rightarrow ee$, $Z \rightarrow ee\gamma$ or $J/\psi \rightarrow ee$ processes is taken to be the average value of the results from all the variations. The systematic uncertainty is estimated to be equal to the root mean square (RMS) of the results of the variations with the intention of modeling a 68% confidence interval, as it will be explained in Section 4.8. The statistical uncertainty on a single variation of the measurement is calculated following the approach of [64]. The statistical uncertainty is taken to be the average of the statistical uncertainties over all combined investigated variations of the measurement. The following sections focus on the measurement of the electron efficiency which is the subject on which I have contributed directly.

4.7 Identification efficiency measurement

In this section, the identification efficiency measurements using the three different decays introduced above are described. The electron identification efficiencies are calculated as the ratio of the number of probe electron candidates passing a certain identification selection (numerator) to the number of all reconstructed probe electrons (denominator). In data both the numerator and the denominator suffer from background contamination, in particular for $E_T < 25$ GeV. In order to remove these background electrons, template distributions are built using relevant variables for

signal and background electrons, as it will be explained later. The measured data distribution is then separately adjusted in the numerator and denominator as the sum of two template distributions (signal and background). This procedure returns the number of signal electrons to be used for the efficiency measurement in the numerator and denominator. In practice, the identification efficiency gives the probability that a prompt electron is identified given it is reconstructed, and is estimated in data as:

$$\epsilon_{\text{ID}}^{\text{Data}} = \frac{N_{\text{pass ID}} - N_{\text{pass ID}}^{\text{BKG}}}{N_{\text{pass ID}} + N_{\text{fail ID}} - (N_{\text{pass ID}}^{\text{BKG}} + N_{\text{fail ID}}^{\text{BKG}})}, \quad (4.7)$$

where the numerator refers to the number of prompt electrons that pass the identification criteria, corrected for the background, and the denominator includes also the number of prompt electrons which fail the identification criteria, corrected for the background. A correction factor (scale factor, SF) is calculated as the ratio of the identification efficiency in data and in MC:

$$SF = \frac{\epsilon_{\text{ID}}^{\text{Data}}}{\epsilon_{\text{ID}}^{\text{MC}}} \quad (4.8)$$

and is applied as a weight in the analysis to the MC event. The value of $\epsilon_{\text{ID}}^{\text{MC}}$ is determined from the corresponding MC sample of prompt electrons ($Z \rightarrow ee$, $Z \rightarrow ee\gamma$ or $J/\psi \rightarrow ee$).

4.7.1 Tag-and-probe method with $Z \rightarrow ee$ events

To determine the identification efficiency in data two different methods are used to build data-driven background template distributions: the Z_{mass} method which uses the invariant mass of the tag-probe pair as the discriminating variable for the background subtraction, and the Z_{iso} method which relies on the calorimetric isolation distribution of the probe electron. The two methods are combined with a procedure which treats them as systematic variations of a single measurement as motivated in Section 4.8. Figure 4.6, illustrates the tag-and-probe $Z \rightarrow ee$ methodology.

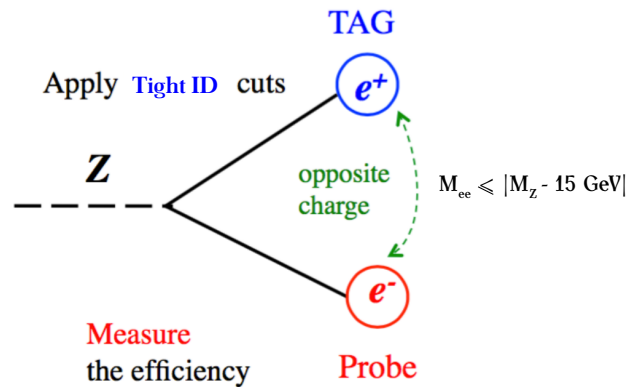


FIGURE 4.6: Graphical illustration of the $Z \rightarrow ee$ tag-and-probe method.

Bin boundaries in E_T [GeV]										
15	20	25	30	35	40	45	50	60	80	150

TABLE 4.5: Low edges and the upper edge (last) of the bin boundaries in E_T used in $Z \rightarrow ee$ methods.

Since the identification efficiency depends on the electron transverse energy and the detector η region, the measurement is performed in different bins of η given in Table 4.2 and in different E_T bins as summarized in Table 4.5.

Events are selected using the "OR" of two unprescaled single-electron triggers, requiring the candidates to pass a minimum E_T threshold. One trigger has a minimum E_T threshold of 24 (26) GeV for 2015 (2016 and 2017) data samples and requires *Tight* trigger identification and track isolation, while the other trigger has a minimum E_T threshold of 60 GeV and *Medium* trigger identification. Moreover the events must be included in the GoodRunList, which ensures that all the relevant subdetectors, in this case, the ID, LAr and TileCal subdetectors, were fully functional when the event information was collected.

It is further required to have at least two reconstructed electron candidates. Tag electrons are required to have $E_T > 27$ GeV and lie outside of the calorimeter transition region, $1.37 < |\eta| < 1.52$. Furthermore they must be associated with the object that fired the trigger within $\Delta R < 0.07$, and pass *Tight* identification requirements. The probe electrons are required to have $E_T > 15$ GeV and pass track quality criteria. The event selection cuts common to both methods (Z_{mass} and Z_{iso}) are shown in Table 4.6.

Events included in the GRL		
Events pass single electron triggers		
Number of vertices ≥ 1		
≥ 2 tracks assigned to the vertex		
Object quality criteria on ECAL cluster of the tag and probe		
≥ 2 electrons in the event		
Reject probe electrons within $\Delta R < 0.4$ to jet with $E_T^{\text{jet}} > 20$ GeV		
For MC events: successful truth matching for tag and probe electrons		
Tag electron	Probe electron	Tag-Probe pair
$E_T > 27$ GeV	$E_T \geq 15$ GeV	$75 \leq m_{ee} \leq 105$ GeV opposite charge pairs
$-2.47 \geq \eta \leq 2.47$ excluding $1.37 \geq \eta \leq 1.52$	$-2.47 \geq \eta \leq 2.47$	
match to trigger	–	
pass <i>Tight</i> ID	–	

TABLE 4.6: Summary of the selection cuts used in the $Z \rightarrow ee$ tag-and-probe method.

Z_{mass} method

To obtain a data-driven background template distribution, representing the shape of the tag-probe invariant mass distribution in which the probe candidate is a background electron, probes are chosen as reconstructed electrons satisfying "inverted" identification and isolation criteria. The signal template is obtained from $Z \rightarrow ee$ MC events.

The background template is normalized using the side-band of the tag-probe invariant mass distribution obtained in data. The number of electron candidates in the denominator of Equation 4.7 before background subtraction is affected by more background than the numerator since the probe electrons in the denominator are selected before identification. On the other hand, since the probe electrons in the numerator satisfy the identification criteria, in the background template of the numerator there is significant signal contamination. For this reason, the normalization procedure of the background template in the numerator and denominator is performed in different ways, even though the same kind of distribution is used. The normalization in the denominator is performed by counting the number of events in the high invariant mass region $120 \text{ GeV} < m_{ee} < 250 \text{ GeV}$ using the opposite-sign tag-probe pairs. For the numerator, the template is normalized by counting the number of events in the high invariant mass region $120 \text{ GeV} < m_{ee} < 250 \text{ GeV}$ using the same-sign invariant mass distribution. This is done because the same-sign distribution has less signal contamination than the opposite-sign distribution.

The efficiency is obtained as the ratio of events in the numerator and denominator (after background subtraction) with an invariant mass m_{ee} within 15 GeV around the Z -boson mass [32] in which the tag and probe electrons have opposite electron charge.

To evaluate the systematic uncertainties, the efficiency measurements from the variations of the following requirements are considered:

- **Mass window requirement:** within 10 and 20 GeV around the Z -boson mass.
- **Tag isolation requirement:** applying or not a cut on the calorimetric isolation variable $E_T^{cone,0.4} < 5$ GeV of the tag electron.
- **Tag identification requirement:** loosening the identification requirement on the tag electron from *Tight* to *Medium*.

Moreover, since for $E_T < 30$ GeV the background contamination is larger, also a low invariant mass normalization region is considered to estimate the systematic uncertainty: $60 \text{ GeV} < m_{ee} < 70 \text{ GeV}$. Due to kinematic reasons, there are far fewer events in the low mass region for $E_T > 30$ GeV. Therefore, for the $E_T > 30$ GeV only the normalization region $120 \text{ GeV} < m_{ee} < 250 \text{ GeV}$ is used.

Z_{iso} method

In the Z_{iso} method, the calorimetric isolation $E_T^{cone,0.3}$ with a cone of $\Delta R = 0.3$ for the probe electron is used as a discriminating variable between the signal and the background. In the $E_T^{cone,0.3}$ distribution, signal electrons are accumulated at low values, while background objects are found at high values of $E_T^{cone,0.3}$. This allows the definition of a signal and background-dominated region.

The data-driven background template is obtained by constructing a distribution which represents the shape of the isolation distribution of background electrons. The signal template is obtained from $Z \rightarrow ee$ MC. Due to a small correlation between the isolation and identification variables and a small bias introduced by badly reconstructed prompt electrons which pass the background selection, the scale factors are sensitive to the background template definition. Three different background templates are defined either by inverting cuts on the shower shape variables and/or on track-cluster matching variables or by requiring the electron to fail the loose cut-based identification selection. In order to avoid statistical overlap between the data sample of electrons used for the efficiency measurement and the electron sample used for the template, only same sign tag-and-probe pairs are selected. Below, the three templates used in the analysis are described.

Template 1. This template is built by requiring that the probe electrons fail the cuts on the variables w_{tot} and eProbabilityHT, defined in Table 4.1. The template distributions depend on E_T and η of the probe to minimize shape biases because some variables are not defined for some (E_T, η) bins. For instance, no TRT information is available for $|\eta| > 2.0$, while signals from the TRT are necessary to compute the eProbabilityHT. In the case of missing information, the template is constructed either ignoring the selection or the template shape is assumed to be the same as in the adjacent η -bins. In the plots this template is called TemplateVar1.

Template 2. This template is built by requiring that the cuts on at least two of the following four variables (defined in Table 4.1): E_{ratio} , $\Delta\eta_1$, f_1 and w_{stot} , are not satisfied by the probe electron. The value of the cuts on these variables depends on E_T and η . In the plots this template is called Template6.

Template 3. This template is built by requiring that the probe fails the cut-based *Loose* identification criteria. In the plots this template is called TemplateFailLoose.

In all cases the background templates are normalized to the isolation distribution of the probes using the upper end of the isolation distribution which is background dominated, namely using events above a given threshold ($E_T^{cone,0.3} = 12.5$ GeV) of the isolation variable. To evaluate the effects of the systematic uncertainties, the efficiency measurements obtained by applying the following variations are considered:

- two additional thresholds for the definition of the background normalization region are used: $E_T^{cone,0.3} > 10$ GeV and $E_T^{cone,0.3} > 15$ GeV,
- the **mass window** is varied between 10 and 20 GeV around the Z -boson mass,
- the **tag isolation requirement** is varied by applying or not a cut on the calorimetric isolation variable $E_T^{cone,0.3}/E_T < 0.3$ in addition to the *Tight* selection,
- different identification cuts are inverted to form two alternative templates, and an alternative isolation variable with a larger isolation cone size namely $E_T^{cone,0.4}$ is used as a discriminant.

Figure 4.7 shows the probe isolation distribution corresponding to the denominator and the numerator of Equation 4.7 for the *Loose* identification criteria. The final scale

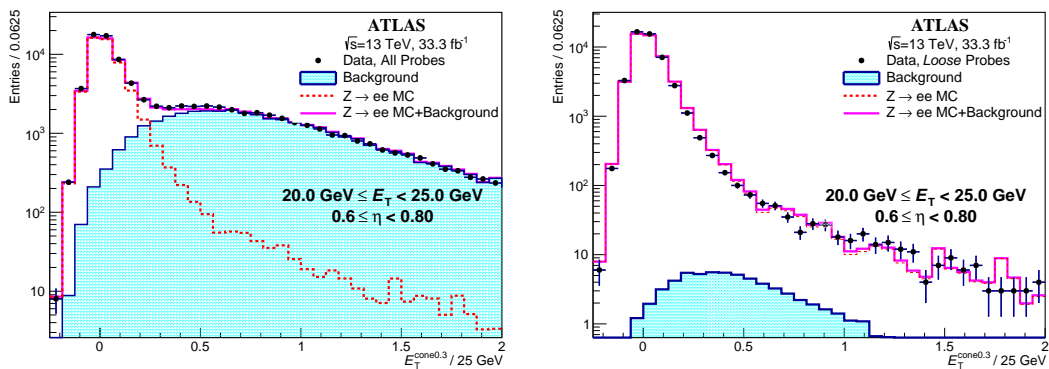


FIGURE 4.7: Distributions of the probe isolation for the bin $20 \text{ GeV} < E_T < 25 \text{ GeV}$ and $0.6 < \eta < 0.8$. The left plot shows the data, the simulated signal distribution and the background template of all probe electrons and the right plot shows the same distributions when the probe electrons satisfy the *Loose* identification criteria. The purple line represents the sum of the $Z \rightarrow ee$ MC and the background distribution.

factors for the three different working points (*Loose*, *Medium* and *Tight*) are presented in Figures 4.8, 4.9 and 4.10, using data collected in 2015, 2016 and 2017. The scale

factors are presented as function of the pseudorapidity η for three representative E_T bins. These results are obtained using the ATLAS software Release 20.7³. Figure 4.11, shows the scale factors for the three different working points for the 2017 data set, using the current software release Release 21, developed for the analysis of full Run II data. The scale factors are consistent with one except in the transition region between the electromagnetic barrel and end-cap calorimeter and at the edges of η region. In 2016 and 2017 data the scale factor are closer to one due to improvements in the MC description of the electron variable.

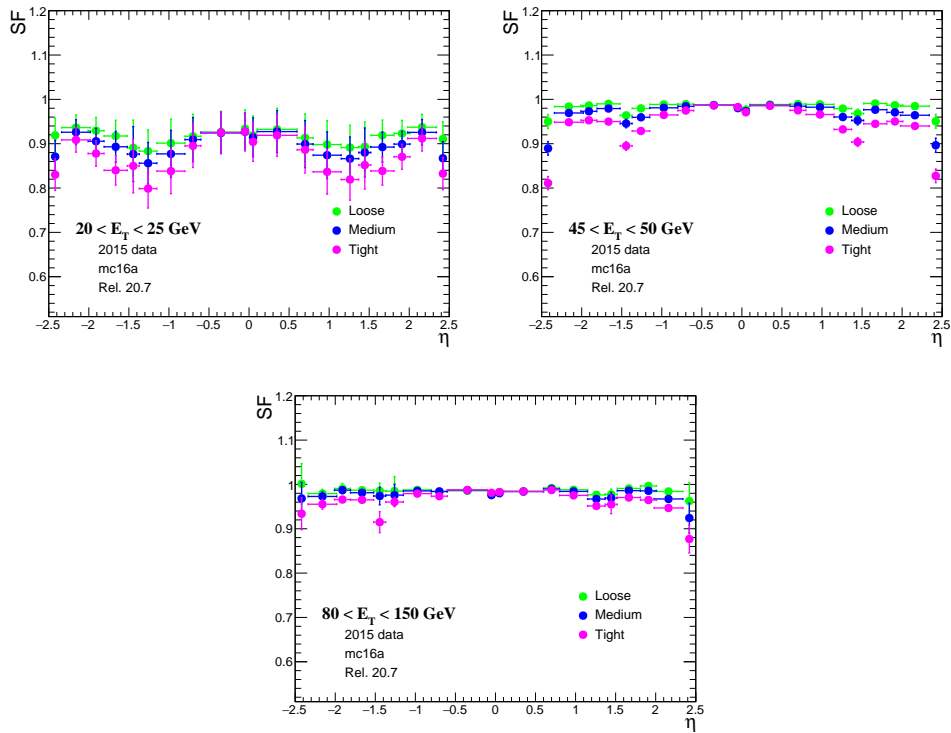


FIGURE 4.8: Scale factors for three representative E_T bins as a function of η using the full 2015 data set, for the three different working point: *Loose*, *Medium* and *Tight*.

³Different versions of the ATLAS reconstruction and identification software are distinguished by a "release number".

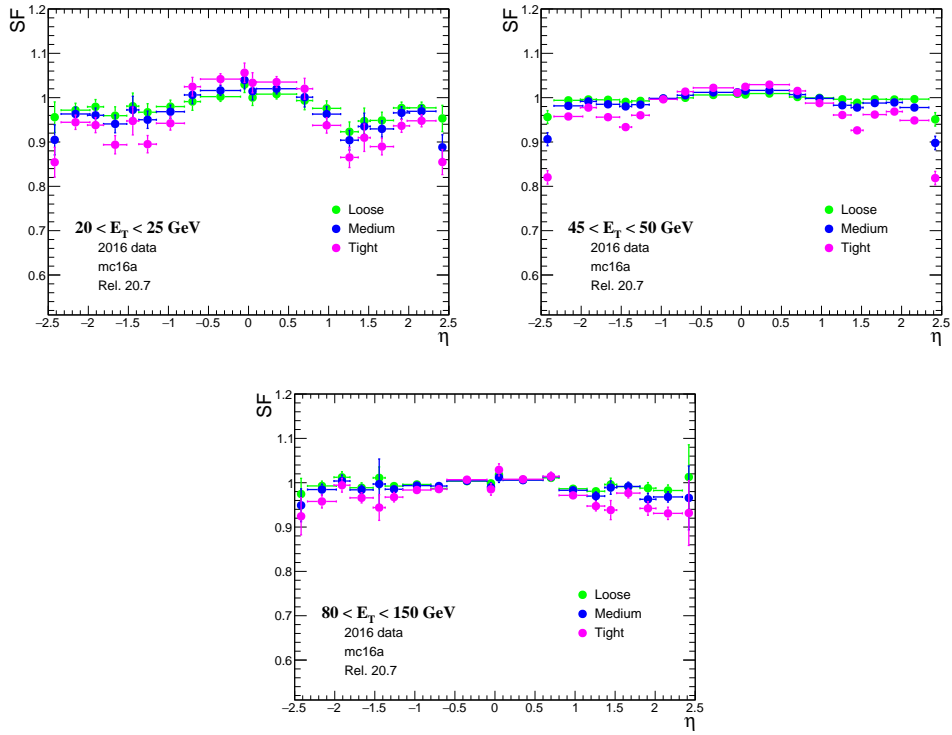


FIGURE 4.9: Scale factors for three representative E_T bins as a function of η using the full 2016 data set, for the three different working point: *Loose*, *Medium* and *Tight*.

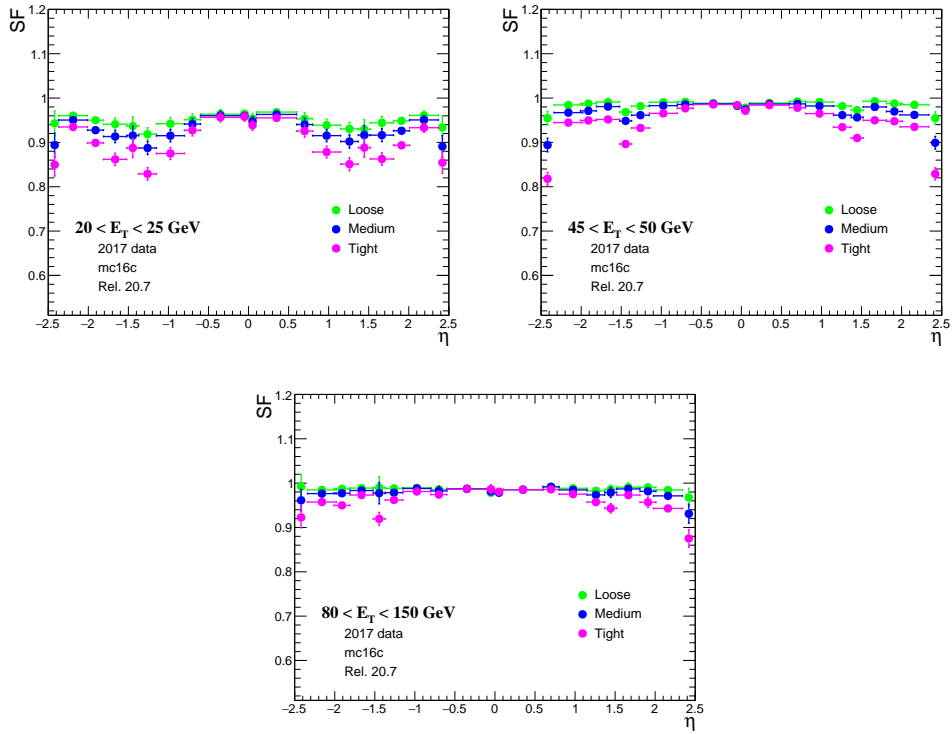


FIGURE 4.10: Scale factors for three representative E_T bins as a function of η using the full 2017 data set, for the three different working point: *Loose*, *Medium* and *Tight*.

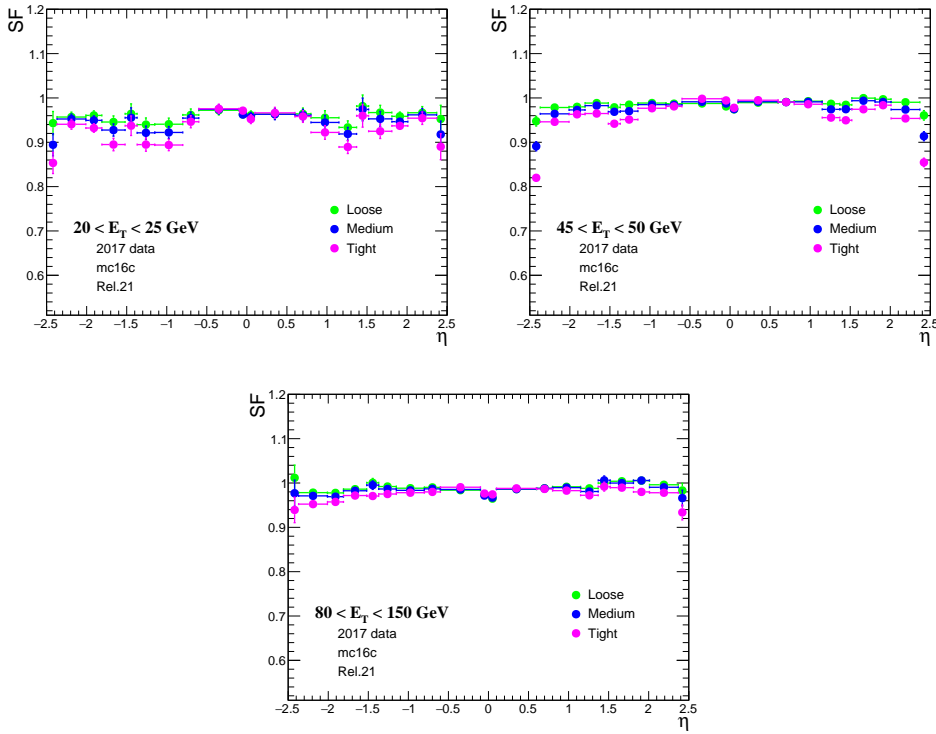


FIGURE 4.11: Scale factors for three representative E_T bins as a function of η using the full 2017 data set using the current ATLAS software release, for the three different working point: *Loose*, *Medium* and *Tight*.

4.7.2 Tag-and-probe method with $J/\psi \rightarrow ee$ events.

This method targets electrons with transverse energy $4.5 \text{ GeV} < E_T < 20 \text{ GeV}$. At such low energies the contamination of the background is significant. Since the J/ψ sample is composed of two contributions (prompt and non-prompt) with significantly different characteristics, these two components need to be distinguished when measuring the efficiency. The prompt J/ψ mesons, are produced directly in proton-proton collisions or in radiative decays of directly produced heavier charmonium states. The non-prompt J/ψ mesons come from b-hadron decays. Since the electrons from the decay of prompt J/ψ particles are expected to be produced at the primary vertex and to be isolated, their efficiencies are expected to be more similar to those of isolated electrons from physics processes of interest such as $H \rightarrow ZZ^* \rightarrow 4l$ in the same transverse momentum range, compared to those from non-prompt production.

The two kinds of productions are distinguished by measuring the position of the J/ψ decay vertex with respect to the primary vertex. Due to the long lifetime of b-hadrons, the electron-positron pair from the non-prompt J/ψ will form a displaced

vertex, while prompt decays occur at the primary vertex. This displacement is estimated using a variable called pseudo-proper time and defined as:

$$\tau = \frac{L_{xy} \times m_{PDG}^{J/\psi}}{p_T^{J/\psi}}, \quad (4.9)$$

where L_{xy} , is the distance between the J/ψ decay vertex and the primary vertex in the transverse plane, while $m_{PDG}^{J/\psi}$ and $p_T^{J/\psi}$ are the mass and the reconstructed transverse momentum of the J/ψ meson, respectively.

A cut (on τ) method and a fit (of τ) method were developed to extract the prompt component. The baseline event selection is the same for both methods. Events with at least two electron candidates with $E_T > 4.5$ GeV and $|\eta| < 2.47$ are collected with dedicated prescaled di-electron triggers with electron E_T thresholds ranging from 4 to 14 GeV. Each of these triggers requires *Tight* online identification and E_T above a certain threshold only for one trigger object, while it demands the transverse energy E_T of the second object to be higher than a lower threshold. Tag electron candidates must match a trigger electron object within $\Delta R < 0.07$ and satisfy the offline *Tight* identification selection. Moreover, the tag and probe candidates must be separated by $\Delta R_{\text{tag-probe}} > 0.15$ to prevent one object to affect the identification of the other. Finally, isolation criteria are applied both on the tag and the probe candidates for most of the variations. All the possible tag-probe pairs are considered.

The invariant mass of the tag-probe pair is used to discriminate signal electrons against background. The most important contribution to the background, comes from random combination of two particles which do not originate from a resonance decay. This is evaluated assuming charge symmetry, namely using the mass spectrum of same-sign (SS) charge pairs. The remaining background is small and can be described with an analytic model. For this, the invariant mass of the tag and probe candidate with opposite charge is fitted in the range of $1.8 \text{ GeV} < m_{ee} < 4.6 \text{ GeV}$, with the sum of three contributions: J/ψ , $\psi(2S)$ and background from hadronic jets, heavy flavor and electrons from conversions. In order to model the signal J/ψ component, a Crystal-Ball function [65, 66] is used. The $\psi(2S)$ is modeled with the same kind of function except for an offset corresponding to the mass difference between the J/ψ and $\psi(2S)$ states. The ratio of J/ψ over $\psi(2S)$ is constrained around the measured value [67]. Finally, J/ψ candidates are considered as those within a mass window of $2.8 \text{ GeV} < m_{ee} < 3.3 \text{ GeV}$.

In the τ -cut method, to select prompt J/ψ the pseudo-proper time is required to be $-1 < \tau < 0.2$ ps (or 0.4 ps, for systematic evaluation). The non-prompt contamination is estimated to be below $\sim 20\%$ (decreasing with decreasing probe E_T) and is subtracted using simulated samples.

In the τ -fit method, the prompt component is extracted by fitting the pseudo proper time distribution in the range $-1 < \tau < 3$ ps, after subtracting the contribution from the estimated OS background using the τ distribution in the mass sidebands

$2.3 \text{ GeV} < m_{ee} < 2.5 \text{ GeV}$ and $4.0 \text{ GeV} < m_{ee} < 4.2 \text{ GeV}$. The non-prompt component is modeled by an exponential decay function convoluted with the sum of two Gaussians, while the shape of the prompt component is described by the sum of the same Gaussians describing the detector resolution, as shown in Figure 4.12.

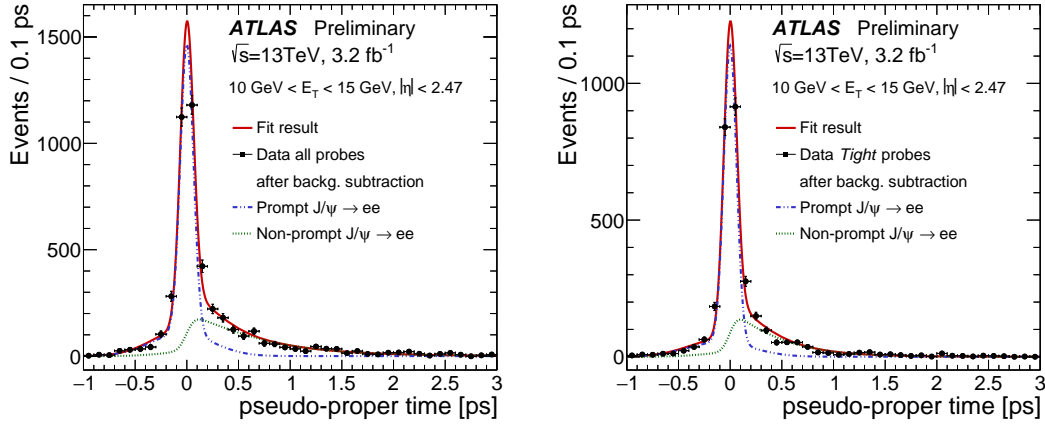


FIGURE 4.12: Pseudo-proper time fit for all probe electrons passing reconstruction and track quality criteria (left) and for probe electrons passing the Tight identification criteria (right) for $10 \text{ GeV} < E_T < 15 \text{ GeV}$, integrated over $|\eta| < 2.47$. Dots with error bars represent the OS minus SS data with the residual background subtracted using the reconstructed dielectron mass distribution sidebands. The signal prompt component is shown by the dashed blue line (sum of two Gaussians) and the signal non-prompt component is shown by the dashed green line (exponential decay function convoluted with the sum of two Gaussians) [57].

Systematic uncertainties are estimated by varying the tag and probe selection (such as the isolation), the fitting procedure (background and signal shapes, fit window, and side-band definitions) and the mass window for signal counting after the mass fit. The main systematic uncertainty in this method arise from the background estimation and the probe definition. The main effects in this estimate are found to be related to the isolation cut on the probe, the choice of the background function, as well as the choice of the Crystal-Ball tail parameters (either fixed to values obtained from simulation or allowed to vary).

4.7.3 Comparison between the Z_{mass} and Z_{iso} results

Figure 4.13, shows the comparison between the efficiencies measured with the Z_{mass} and Z_{iso} methods as function of η for two representative E_T bins: $20 < E_T < 25$ and $40 < E_T < 45 \text{ GeV}$. It is observed that in the low E_T bin, the scale factors obtained with Z_{iso} method are systematically lower than the scale factors obtained with the Z_{mass} method, while a fair agreement is found in the higher E_T bins. Quantitatively the effect is estimated of $\sim 5\%$ in the lowest E_T bin.

In order to understand the origin of the observed difference, the effect of the variation of relevant parameters of the Z_{mass} and Z_{iso} algorithms on the scale factors was examined. A significant change of the scale factors may point out to a possible

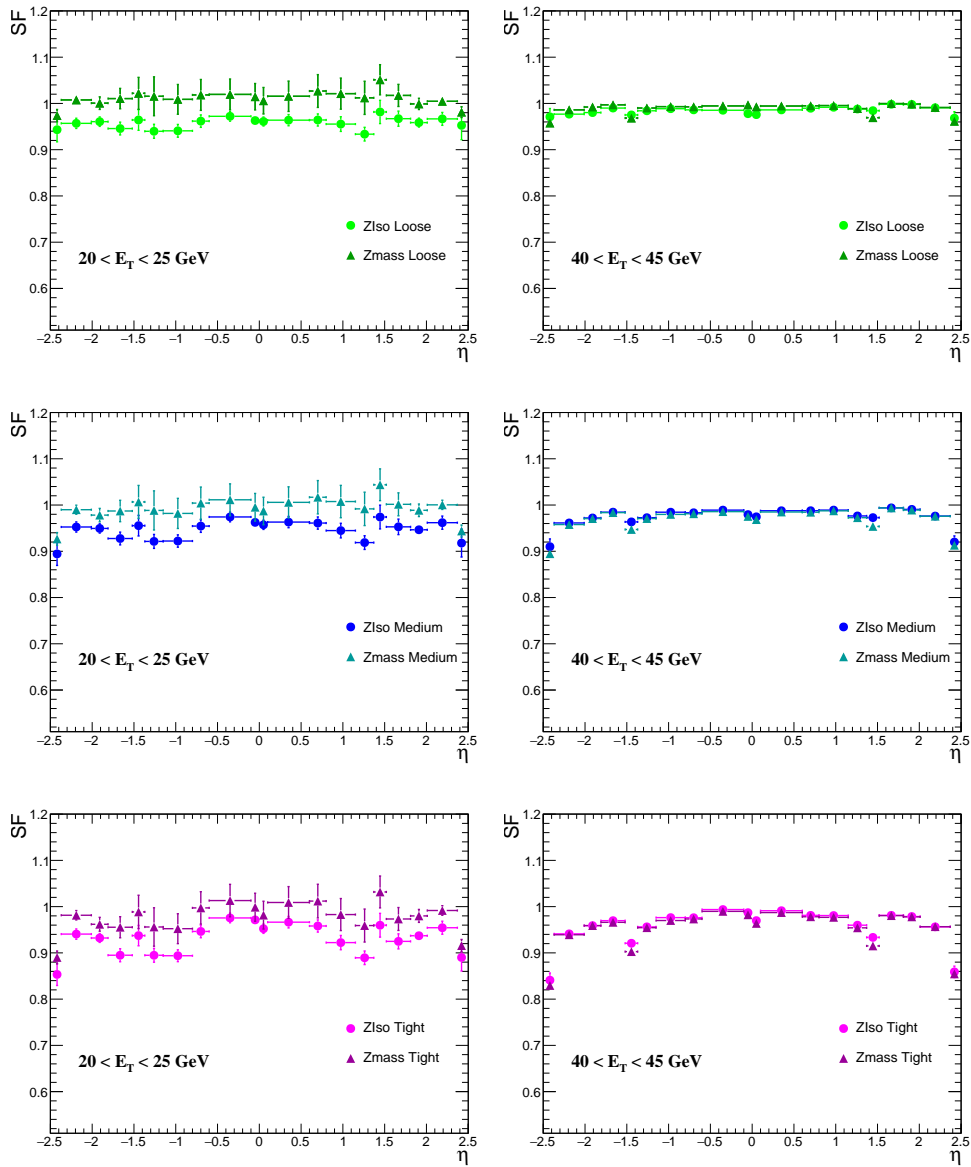


FIGURE 4.13: Comparison of the scale factors obtained with the two $Z \rightarrow ee$ methods in the 20-25 GeV E_T bin (left) and 40-45 GeV E_T bin (right) for different working points (*Loose*, *Medium* and *Tight*). Error bars represent the total uncertainty.

source of discrepancy. Only plots referring to the Z_{iso} method are presented here since they are the results of my work.

Several variations were investigated. The main impact (up to 3%) on the average scale factors per E_T bin was found related to the modeling of the shape of the data driven background as shown in Figure 4.14. Following that the plots that show the two biggest effects on the scale factors are presented in Figures 4.15 and 4.16 using the full 2017 data set. The result of the investigation concluded that the difference between the Z_{mass} and Z_{iso} is related mainly to the background modeling in both methods. As will be discussed in Section 4.8 the difference is finally treated as systematic uncertainty.

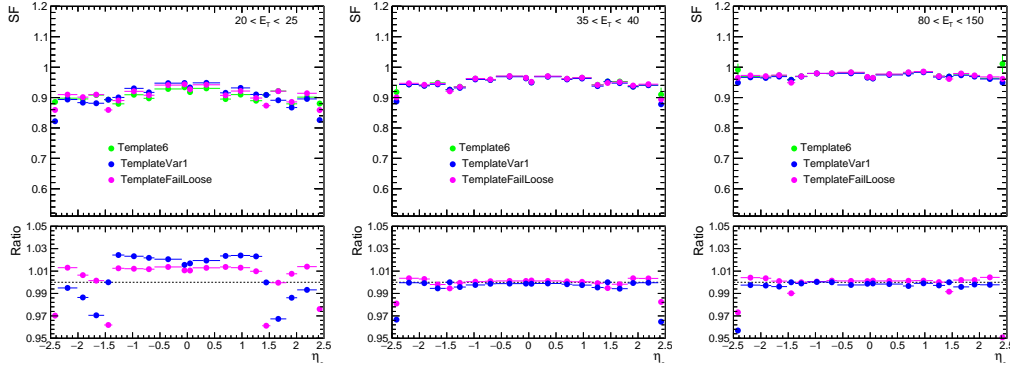


FIGURE 4.14: Impact of the shape of the data-driven background template on the scale factors as a function of η for three different E_T bins for the *Medium* selection criteria. The ratio plots are taken with respect to the Template6. The template definition is given in the text.

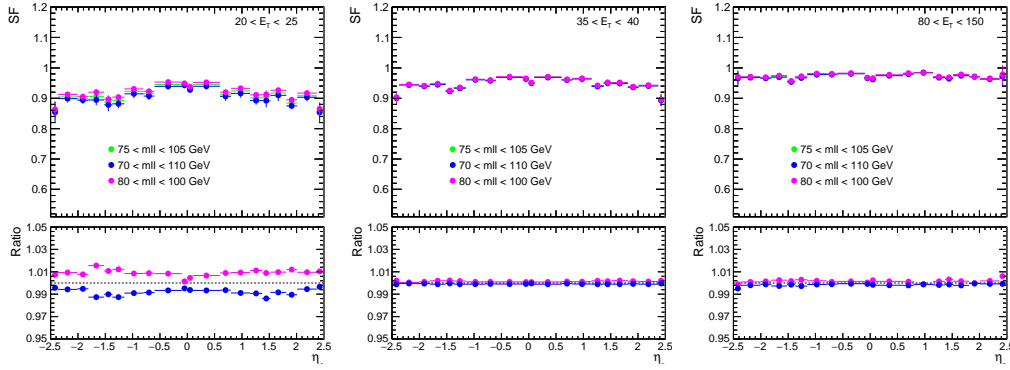


FIGURE 4.15: Impact of a tag-and-probe pair invariant mass cut on the scale factors as a function of η for three different E_T bins for the *Medium* selection criteria. The ratio plots present the ratio between the up and down variations with respect to the central variation.

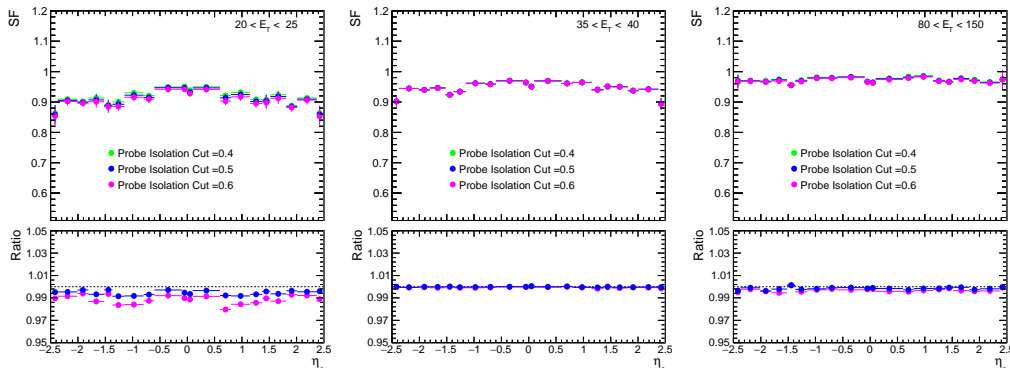


FIGURE 4.16: Impact of the probe isolation selection on the scale factors as a function of η for three different E_T bins for the *Medium* selection criteria. The ratio plots present the ratio between the variation applying a probe isolation cut of 0.4 and 0.6 with respect to applying a cut at 0.5.

4.7.4 Tag and probe method with $Z \rightarrow ee\gamma$ events.

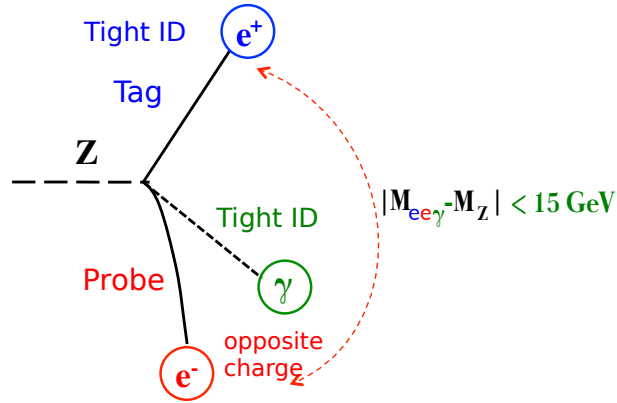


FIGURE 4.17: Graphical illustration of the $Z \rightarrow ee\gamma$ tag-and-probe methodology.

The $Z \rightarrow ee\gamma$ method selects $Z(ee)$ decays in which one of the electrons (probe) has emitted a photon with high transverse momentum (typically above 10 GeV) due to the electromagnetic process called final state radiation (FSR). The identification efficiency is measured using the invariant mass computed from three objects: the tag electron, the probe electron and the photon. The invariant mass of these three objects will form a peak at the Z -boson mass. Probe electrons coming from $Z \rightarrow ee\gamma$ decays, have a lower E_T spectrum compared to those from the $Z \rightarrow ee$ methods, since they have lost much of their energy due to the final state radiation process. This method targets mainly electron efficiency measurements in the transverse energy range $10 < E_T < 15$ GeV. Figure 4.17, illustrates the tag-and-probe $Z \rightarrow ee\gamma$ methodology.

Motivation

For the reason explained above, the $Z \rightarrow ee\gamma$ tag-and-probe method provides a measurement of the electron efficiency in which the distribution of the selected probes is affected by less background with respect to the Z_{mass} and Z_{iso} methods in the low transverse energy region. This feature is also of interest since it may help in understanding the observed different scale factors at low transverse energy between the Z_{mass} and Z_{iso} methods. The $Z \rightarrow ee\gamma$ method addresses the E_T interval between the region where the J/ψ and $Z \rightarrow ee$ methods have the highest statistical power. This is of great interest for analysis like $H \rightarrow ZZ^* \rightarrow 4l$.

Optimization and event selection

The $Z \rightarrow ee\gamma$ method was initially proposed during Run I [68], but the limited statistics did not allow for an effective use of it. Before implementing the method in the official electron tag-and-probe framework an optimization of several selection criteria was carried out using a sample of simulated $Z \rightarrow ee\gamma$ events and concentrating

on the number of events with a probe electron with transverse energy $10 < E_T < 15$ GeV.

Table 4.7 summarizes the starting point of the particle and event pre-selection used in the implementation of the method. The pre-selection criteria applied to define the tag and probe electrons are inspired by the Z_{mass} method. Additional selections are shown in Table 4.8. In order to avoid possible ambiguities, a minimum angular distance between the photon and the electron candidates are required: $\Delta R_{\text{tag-}\gamma} > 0.4$ and $\Delta R_{\text{probe-}\gamma} > 0.2$. The reason for the difference in this cut value lies in the need of selecting a good isolated tag avoiding to bias the probe electron sample. Furthermore, since it is the probe electron that emits a FSR photon, the photon is expected to be closer to the probe electron than to the tag electron. Further requirements are placed on the tag-probe and tag-photon invariant masses, indicated with M_{ee} and $M_{\text{tag-}\gamma}$ respectively, in order to select $Z \rightarrow ee$ events with a photon emitted by FSR: $40 < M_{ee} < 90$ GeV, $M_{\text{tag-}\gamma} < 80$ GeV and $E_{T\text{probe}} + E_{T\gamma} > 30$ GeV .

Events included in the GRL		
Events pass the single electron trigger		
number of vertices ≥ 1		
≥ 2 tracks assigned to the vertex		
Object quality criteria on ECAL cluster of tag and probe		
≥ 2 electrons in the event		
≥ 1 photons in the event		
Reject probe electrons within $\Delta R < 0.4$ to ANTIK4 jet with $E_T^{\text{jet}} > 20$ GeV		
Tag electron	Probe electron	Photon
$E_T \geq 27$ GeV	$E_T \geq 10$ GeV	$E_T \geq 10$ GeV
$-2.47 \geq \eta \leq 2.47$ excluding $1.37 \geq \eta \leq 1.52$	$-2.47 \geq \eta \leq 2.47$	$-2.37 \geq \eta \leq 2.37$
match to trigger	–	converted & non-converted with two tracks
pass <i>Tight</i> ID	–	pass <i>Tight</i> ID & isolated (FixedCutTight)

TABLE 4.7: Summary of the event pre-selection cuts used in the $Z \rightarrow ee\gamma$ tag-and-probe method.

Tag and probe have opposite charge
$40 < M_{ee} < 90$ GeV
$M_{\text{tag-}\gamma} < 80$ GeV
$\Delta R_{\text{tag-}\gamma} > 0.4$
$\Delta R_{\text{probe-}\gamma} > 0.2$
$E_{T\text{probe}} + E_{T\gamma} > 30$ GeV

TABLE 4.8: Event selection used for the $Z \rightarrow ee\gamma$ method, applied after the pre-selection cuts in order to select $Z \rightarrow ee$ events with a photon emitted by FSR.

Table 4.9 presents the effects in MC of the cuts that select radiative Z events. The table shows that these selections are very efficient. Nevertheless comparing the number of selected events with the number of events selected in the ATLAS Z radiative

analysis carried at 8 TeV [68] it was found that the expected 2.5 increase in statistics was not obtained.

Selection Cut	MC Number of events	MC %
$40 < M_{ee} < 90$ GeV	76241	100%
$M_{\text{tag-}\gamma} < 80$ GeV	70595	93%
$\Delta R_{\text{tag-}\gamma} > 0.4$	70298	99.6%
$\Delta R_{\text{probe-}\gamma} > 0.2$	70298	100%
$E_{T\text{probe}} + E_{T\gamma} > 30$ GeV	67085	95%

TABLE 4.9: Cut flow of the percentage of events that passing each selection cut for MC.

To increase the size of the tag-and-probe electron sample, an optimization procedure was performed. Among the different analyzed criteria, in the following part of this section the effect of the **trigger** selection, the **photon isolation** and the **tag electron identification** criteria were reported.

Trigger

So far in the Z tag-and-probe methods, only single electron triggers were used with an E_T trigger threshold that from Run I to Run II was increased from 24 to 26 GeV⁴. To recuperate potential lost events due to this change two electron-photon combined triggers were added in OR to the single electron trigger. The first named HLT_e20_lhmedium_nod0_g35loose requires at least one *Medium* electron with transverse energy above 20 GeV and a *Loose* identified photon with transverse energy above 35 GeV. The second named HLT_e24_lhmedium_nod0_g25medium requires at least one *Medium* electron with transverse energy above 24 GeV and a *Medium* identification photon with transverse energy above 25 GeV. The addition of these triggers allowed the possibility to relax the cut on the transverse energy of the tag electron from 27 to 25 GeV. The two changes above described brought an increase of 1.5 in the number of simulated $Z \rightarrow ee\gamma$ events with a probe electron in the range $10 < E_T < 15$ GeV.

Photon isolation

Isolation consists in assessing the activity surrounding the trajectory of a particle in the tracker and the calorimeters. As in the case of electrons the isolation is one of the most powerful tools to select prompt photons. To calculate the photon isolation variables, a cone is defined around the direction of the photon candidate in the same way as already described in the case of electrons in Section 4.5.

The **track-based isolation** estimates the activity surrounding the photon by summing over the transverse momenta of close-by tracks. The **calorimeter-based isolation** calculates the sum of the clusters energy in a cone around the photon candidate.

Working points are then defined by applying cuts to these two variables. The *FixedCutTight* photon isolation working point was identified as another possible source of loss of events. The use of *FixedCutLoose* [69] resulted in an increase of a factor of

⁴Consequently, in Run I the cut on the p_T of the tag electron was 25 GeV.

1.2 in the number of simulated $Z \rightarrow ee\gamma$ events with a probe electron in the range $10 < E_T < 15$ GeV. It was verified with MC that the number of non-prompt photon in the selected sample did not increase.

Tag-electron identification working point

Switching from the *Tight* working point to a *Medium* working point for the tag identification increased the number of selected events by $\sim 3\%$ in a $Z \rightarrow ee\gamma$ MC sample and $\sim 10\%$ in data. This may indicate that this change increased the fraction of fake tag electrons. This change is not applied since the method relies on the quality of the tag electrons and the increase is modest.

Background estimation and systematic variations

A template method is used to disentangle in data signal from background electrons. In order to form the data-driven background template, probes are chosen in a similar way as in the Z_{mass} method (Section 4.7.1). The amount of signal electron in the background templates is small and is estimated using MC simulations. The signal templates in obtained from $Z \rightarrow ee\gamma$ MC. The background template is then normalized to the data luminosity using the side-band method, as described for the Z_{mass} method. The normalization region for the radiative Z method is the high invariant mass region $100 \text{ GeV} < m_{ee\gamma} < 250 \text{ GeV}$. The number of background probes in data when the *Tight* selection is applied is found negligible. Globally the amount of background probes in the $Z \rightarrow ee\gamma$ distribution is much less than in the Z_{mass} method. Figure 4.18 shows the distribution of the invariant mass of the tag-probe-photon system corresponding to the denominator and numerator of the formula in Equation 4.7 for the *Tight* criteria. Figures 4.19 and 4.20 shows the efficiency in data and in MC and the scale factors obtained with the $Z \rightarrow ee\gamma$ method.

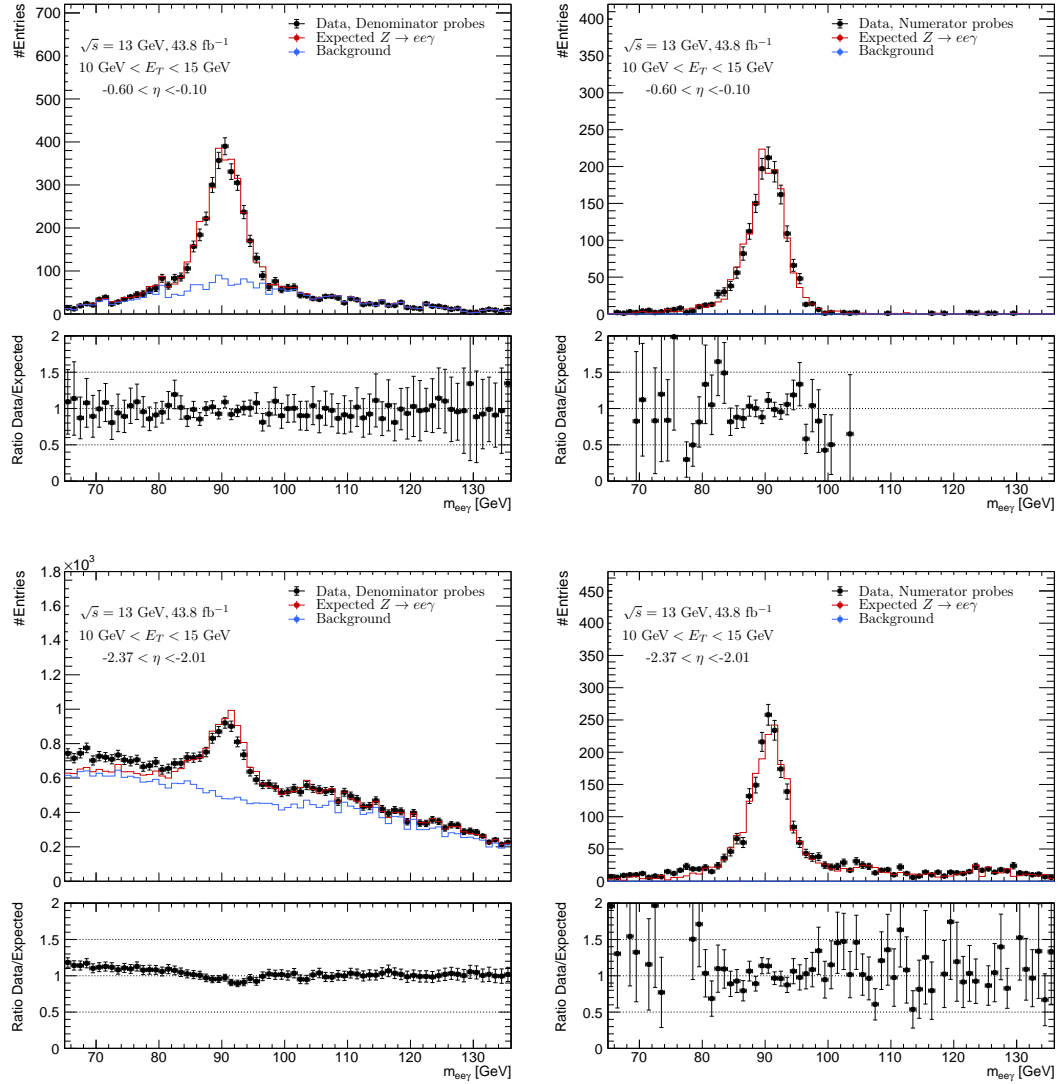


FIGURE 4.18: Distribution of the invariant mass of the tag-probe electrons and the photon. The dots represent the data, the red histogram the signal MC and the blue histogram the background data-driven template. The distributions are shown for the transverse energy interval $10 \text{ GeV} < E_T < 15 \text{ GeV}$, for two representative pseudorapidity bins $-0.60 < \eta < -0.10$ bin (top) and $-2.37 < \eta < -2.01$ bin (bottom). The distributions on the left are obtained at reconstruction level while those on the right are after applying the *Tight* selection.

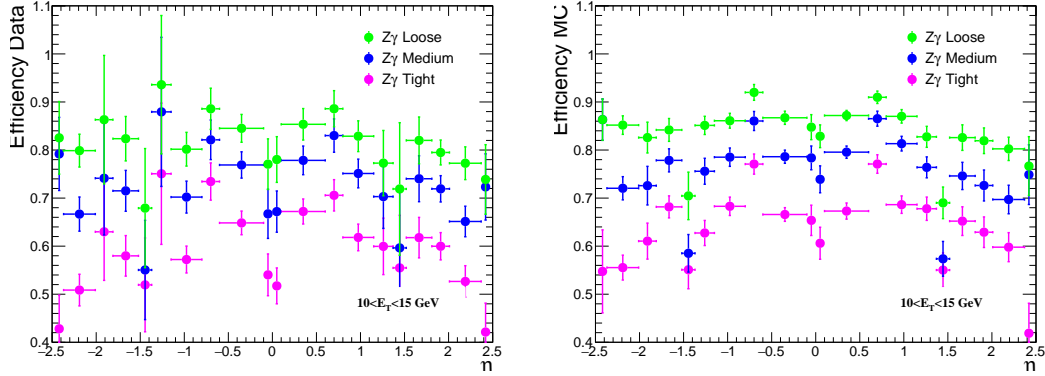


FIGURE 4.19: Data (left) and MC efficiency (right) obtained using the $Z \rightarrow ee\gamma$ method, for the 10-15 GeV energy range for the *Loose*, *Medium* and *Tight* identification criteria. The error bars represent the combined statistical and systematic uncertainty.

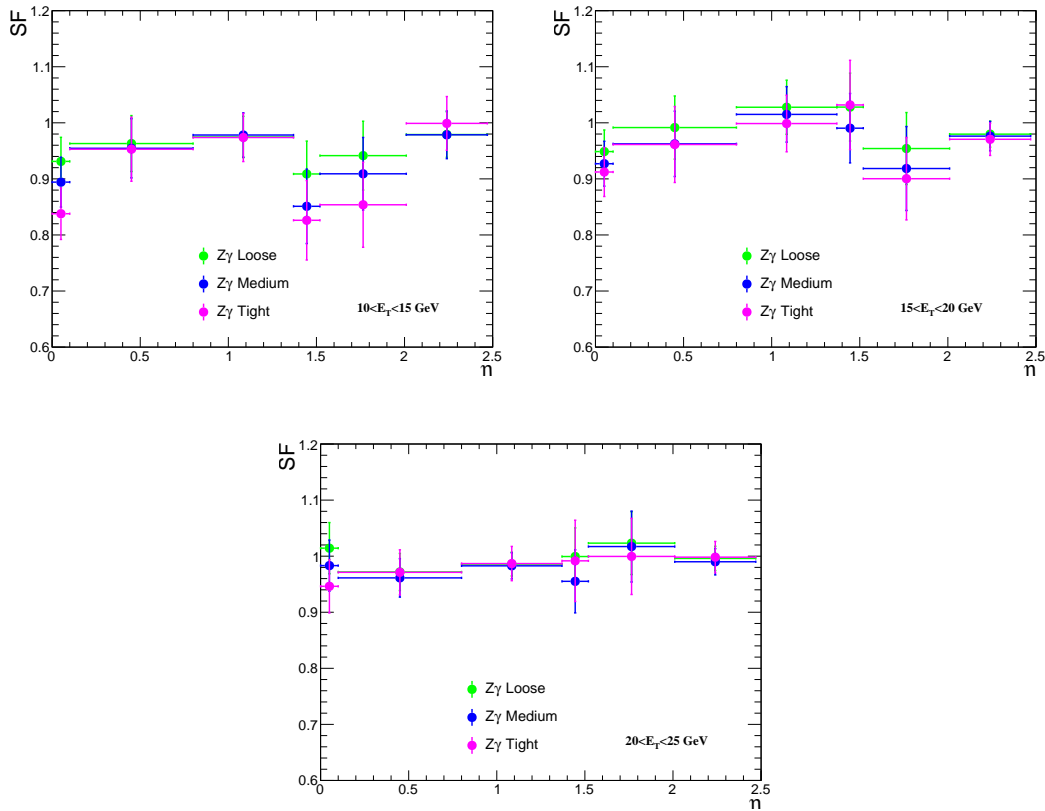


FIGURE 4.20: Scale factors as function of $|\eta|$ obtained using the $Z_{ee\gamma}$ method, for the 10-15 GeV (top left), 15-20 GeV (top right) and 20-25 GeV (bottom) energy range for the *Loose*, *Medium* and *Tight* identification criteria. The error bars represent the combined statistical and systematic uncertainty.

4.8 Combined scale factors

Figures 4.21 and 4.22 compare the efficiencies in data and in MC obtained with Z_{mass} and $Z_{ee\gamma}$ and with Z_{iso} and $Z_{ee\gamma}$, respectively. A fair agreement is obtained even though the values obtained with $Z_{ee\gamma}$ are systematically higher than those obtained with the Z_{iso} method. In the low E_T bins the results of the $Z_{ee\gamma}$ method are closer to those of the Z_{mass} method.

To obtain the final results, the scale factors calculated with different methods are combined. Since the calculation of the scale factors with the Z_{mass} and Z_{iso} methods are performed using the same data sample the measurements are statistically correlated. The same observation applies to the J/ψ τ -cut and τ -fit methods. Moreover, the results are also systematically correlated, since part of the variables chosen for the variations are the same among methods. Therefore, in the combination, the individual results from the three Z methods on one side and the two J/ψ methods on the other side, are considered as systematic variations of one another.

In Figure 4.24 the scale factors as a function of the pseudorapidity for the combination of the τ -cut and τ -fit method are displayed as well as for the Z_{iso} and Z_{mass} in representative bins of E_T . The methods based on the $J/\psi \rightarrow ee$ decays are found to be in agreement with each other. In the methods based on the $Z \rightarrow ee$ decays, the uncertainty of the combination in the low E_T bins does not cover the envelope of the results from the Z_{iso} and Z_{mass} methods. In order to ensure that the systematic uncertainty on the combined results includes the difference between results from the two methods, the uncertainty is inflated such that the RMS covers at least 68% of all variations [70]. It was found that increasing the uncertainties of the measurements by 10% covers the observed differences between the methods. For the low E_T bins, where there is discrepancy between the two $Z \rightarrow ee$ methods, the use of the $Z \rightarrow ee\gamma$ method adds additional information. Figure 4.26 shows the comparison between the $Z_{ee\gamma}$ method and the Z inclusive combined method. The $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ results are combined with a χ^2 -fit, where the systematic uncertainties are included as nuisance parameters (Figure 4.23).

The following list presents for the different E_T bins the methods that are combined to get the final result:

- 7-10 GeV: $J/\psi \rightarrow ee$ method (top plot of Figure 4.24),
- 10-15 GeV: $J/\psi \rightarrow ee$ and $Z \rightarrow ee\gamma$ methods combined (Figure 4.25)
- 15-20 GeV: all methods combined (Figure 4.3),
- above 20 GeV: $Z \rightarrow ee$ method (bottom plot of Figure 4.24).

Finally, the efficiencies measured in data and in MC simulation are presented in Figure 4.27 as a function of the number of measured primary vertices. The plot shows that the algorithm to measure the scale factor is robust against the increase of pile-up.

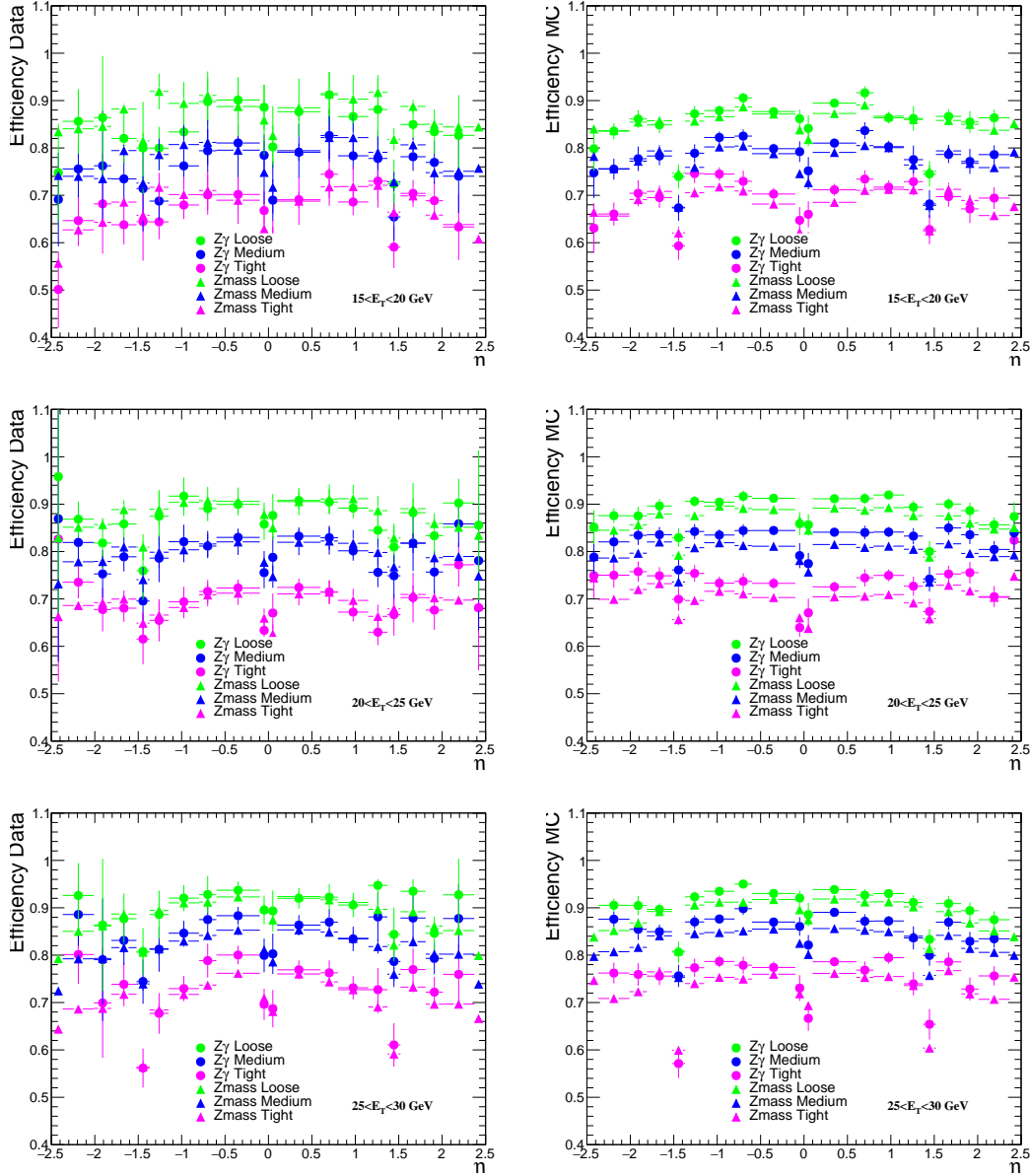


FIGURE 4.21: Data (left) and MC efficiency (right) comparison in the $15 < E_T < 20$ GeV, $20 < E_T < 25$ GeV and $25 < E_T < 30$ GeV bins between Z_{mass} and $Z_{ee\gamma}$ methods. Shown are numbers for all identification criteria. Only systematic uncertainties are reported.

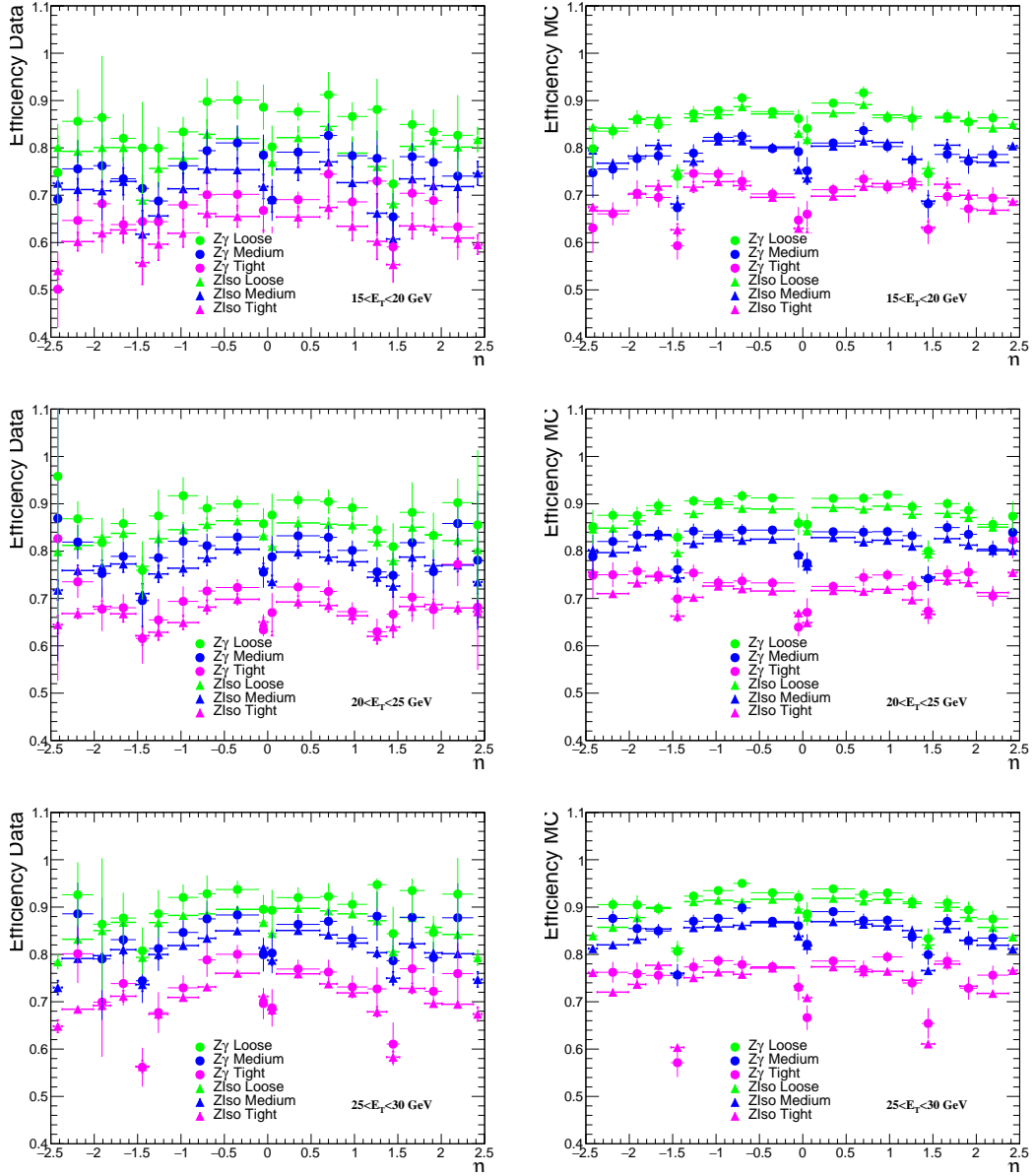


FIGURE 4.22: Data (left) and MC efficiency (right) comparison in the $15 < E_T < 20$ GeV bin (top plots) and in the $20 < E_T < 25$ GeV bin (middle plots) and between the $25 < E_T < 30$ GeV bin (top plots) Z_{iso} and $Z_{ee\gamma}$ methods. Shown are numbers for all identification criteria. Only systematic uncertainties are reported.

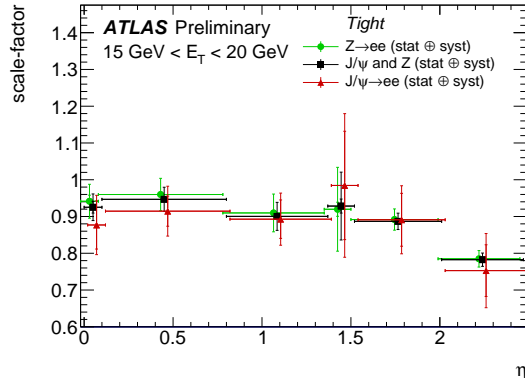


FIGURE 4.23: Scale factor comparison of $Z \rightarrow ee$ with $J/\psi \rightarrow ee$ methods. The scale factors are shown for the *Tight* identification criteria as a function of the pseudorapidity $|\eta|$ for probe transverse energies in the range $15 < E_T < 20$ GeV. The error bars represent the combined statistical and systematic uncertainty [57].

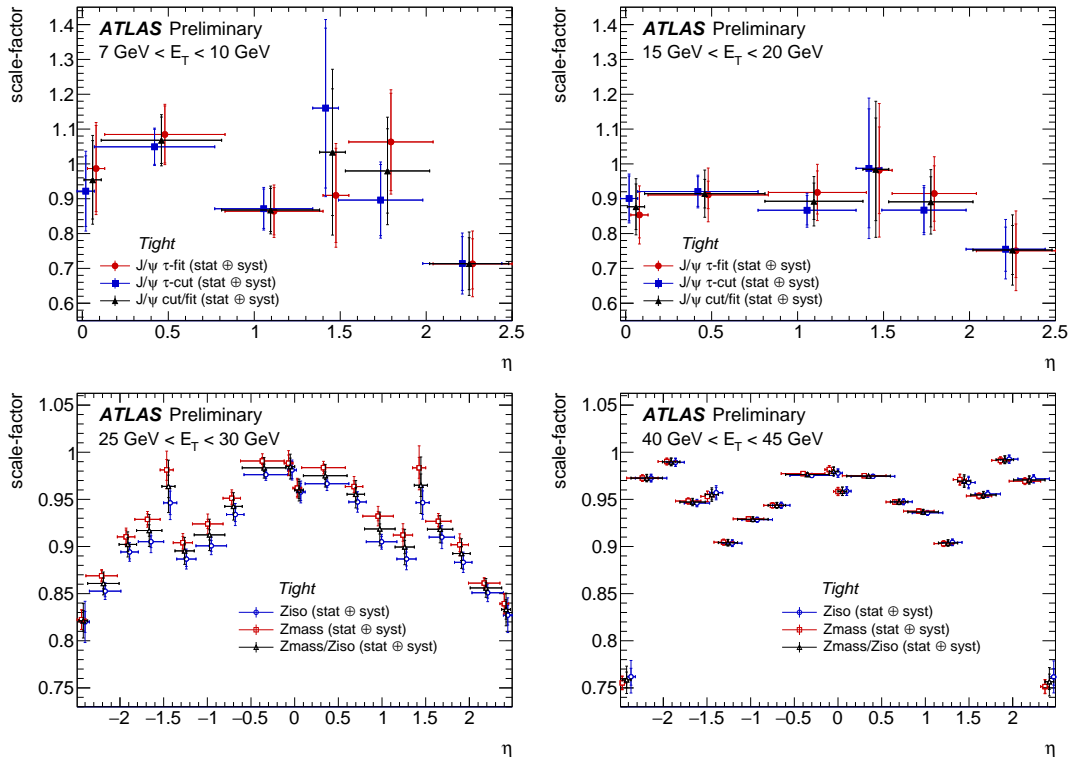


FIGURE 4.24: Scale factor measured in $J/\psi \rightarrow ee$ events with the τ -fit and τ -cut methods and the combined result (top). The scale factors are shown as a function of the pseudorapidity $|\eta|$ for probe transverse energy in the ranges $7 \text{ GeV} < E_T < 10 \text{ GeV}$ (left) and $15 \text{ GeV} < E_T < 20 \text{ GeV}$ (right), for the *Tight* identification criteria. Scale factor measured for $Z \rightarrow ee$ events with the Z_{mass} and Z_{iso} methods, and the combined results are shown (bottom) as a function of pseudorapidity η for probe transverse energies in the ranges $25 \text{ GeV} < E_T < 30 \text{ GeV}$ (left) and $40 \text{ GeV} < E_T < 45 \text{ GeV}$ (right). The error bars represent the combined statistical and systematic uncertainty [57].

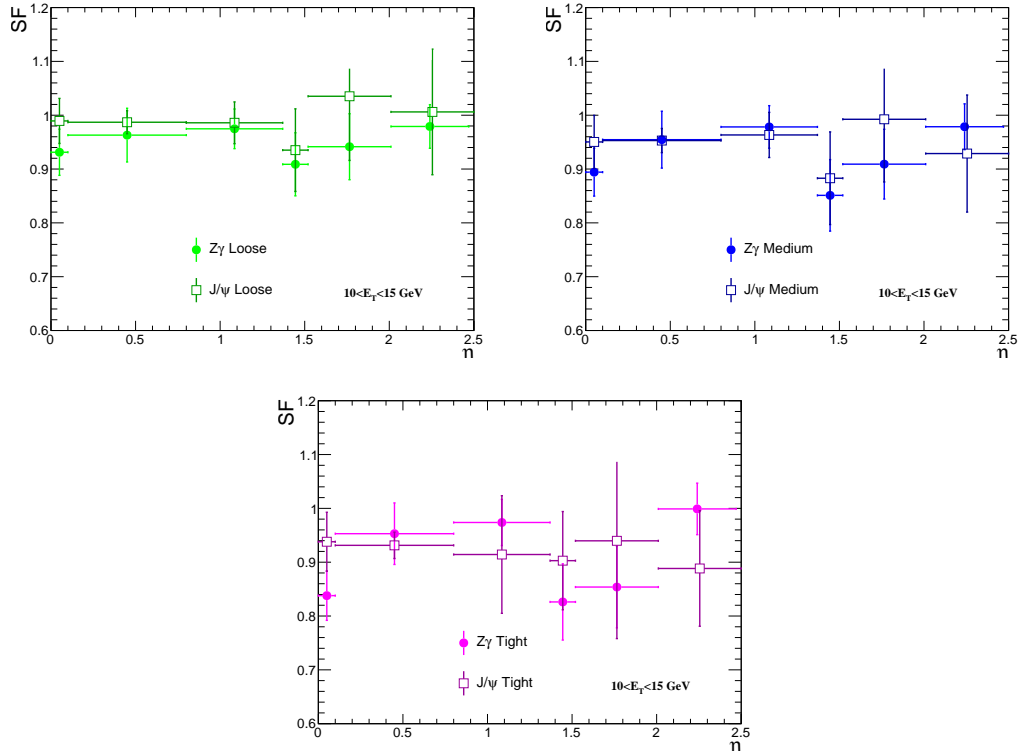


FIGURE 4.25: Scale factor comparison of $Z \rightarrow ee\gamma$ with $J/\psi \rightarrow ee$ methods. The scale factors are shown for the *Loose*, *Medium* and *Tight* identification criteria as a function of the pseudorapidity $|\eta|$ for probe transverse energies in the range $10 < E_T < 15$ GeV. The error bars represent the combined statistical and systematic uncertainty.

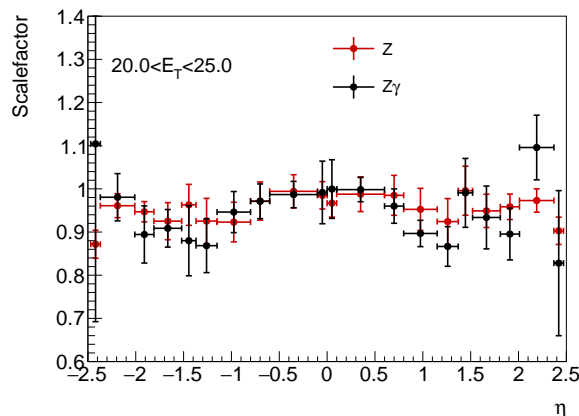


FIGURE 4.26: Scale factor comparison between $Z \rightarrow ee\gamma$ and $Z \rightarrow ee$ methods. The scale factors are shown for the *Tight* identification criteria as a function of pseudorapidity η for probe transverse energies in the range $20 < E_T < 25$ GeV. The error bars represent the combined statistical and systematic uncertainty.

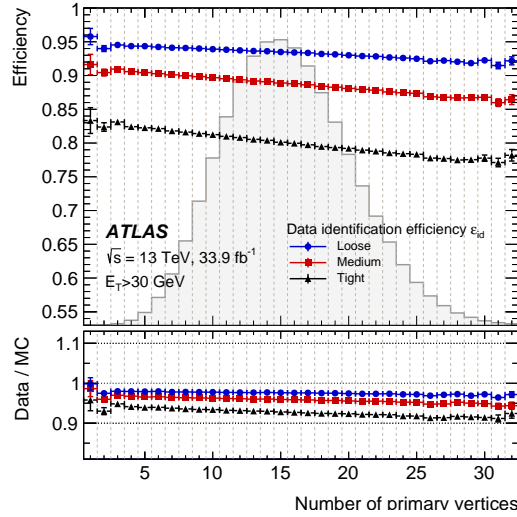


FIGURE 4.27: The LH electron-identification efficiencies for electron candidates with $E_T > 30$ GeV for the *Loose*, *Medium* and *Tight* operating points as a function of the number of primary vertices in the 2016 data using the $Z \rightarrow ee$ process. The shaded histogram shows the distribution of the number of primary vertices for the 2016 data. The inner uncertainties are statistical while the total uncertainties include both the statistical and systematic components.

The bottom panel shows the data-to-MC ratios [58].

4.9 Usage of electron selections in physics measurements

For the electron identification, within the scope of this thesis, scale factors were derived for the full 2015, 2016 and 2017 data sets. These scale factors are applied in all ATLAS analyses using electrons in 2015, 2016 and 2017 data. Examples of physics results that use the scale factors described in this chapter are shown in Figures 4.28, 4.29 and 4.30.

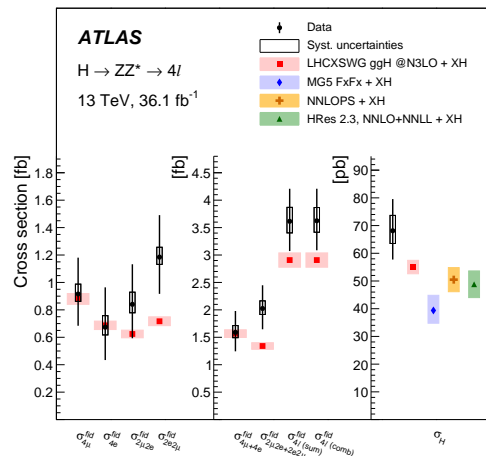


FIGURE 4.28: The fiducial cross sections (left two panels) and total cross section (right panel) of Higgs boson production measured in the 4ℓ final state. The inclusive fiducial cross section is measured as the sum of all channels, as well as by combining the per-channel measurements assuming SM $ZZ^* \rightarrow 4\ell$ branching ratios. The error bars on the data points show the total uncertainties, while the systematic uncertainties are indicated by the boxes. The shaded bands around the theoretical predictions indicate the PDF and scale uncertainties [71].

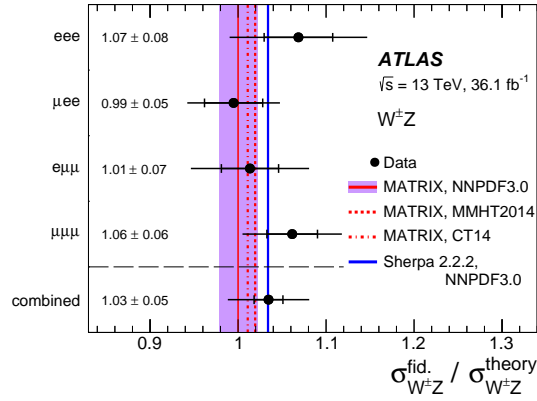


FIGURE 4.29: Ratio of the measured $W^\pm Z$ integrated cross sections in the fiducial phase space to the NNLO SM prediction from MATRIX in each of the four channels and for their combination. The inner and outer error bars on the data points represent the statistical and total uncertainties, respectively. The vertical lines show the predictions for interface with different parton distribution functions. The prediction from MATRIX using the MMHT2014 and CT14 PDF sets and the NLO prediction from SHERPA 2.2.2 are also displayed as dashed-red, dotted-dashed-red and blue lines, respectively [72].

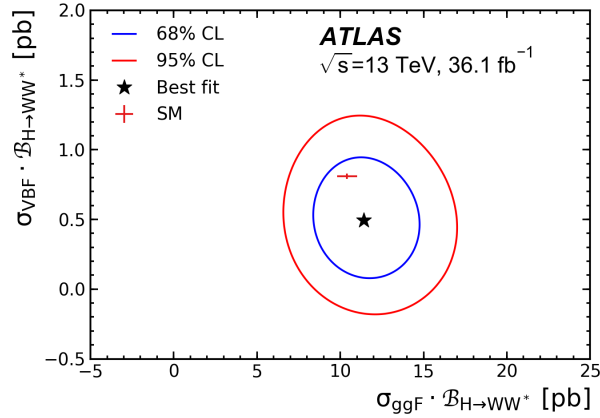


FIGURE 4.30: The 68% and 95% confidence level two-dimensional likelihood contours of $\sigma_{\text{ggF}} \cdot \mathcal{B}_{H \rightarrow WW^*}$ vs. $\sigma_{\text{VBF}} \cdot \mathcal{B}_{H \rightarrow WW^*}$, compared to the SM prediction shown by the red marker. The error bars on the SM prediction represent the ggF and VBF theory uncertainty [73], respectively [74].

Chapter 5

Analysis of $Z\gamma$ electroweak production

Εάν μη έλπηται ανέλπιστον, ουκ εξευρήσει.

Ηράκλειτος 544-484 π.Χ.

In this chapter, the study of the diboson $Z\gamma$ production in association with two high energy jets is presented. Proton-proton collision data from the ATLAS experiment at the CERN Large Hadron Collider at a center-of-mass energy of $\sqrt{s} = 13$ TeV, collected in 2015 and 2016 and corresponding to an integrated luminosity of 36.1 fb^{-1} are analyzed. The electroweak production of $Z\gamma jj$ events ($Z\gamma jj$ -EW), provides a direct access to the nature of the electroweak symmetry breaking mechanism, by probing the gauge boson couplings $WWZ\gamma$, $ZZZ\gamma$, $ZZ\gamma\gamma$ and $Z\gamma\gamma\gamma$. The last three are forbidden at the lowest order in the Standard Model. Thus, any deviation from the Standard Model predictions would hint at new physics which could reveal itself in the form of anomalous quartic gauge boson couplings (aQGCs) [75, 76].

Until now, the production of $Z\gamma jj$ -EW events, which consist in processes with a cross section proportional to the electroweak coupling to the fourth power and include vector-boson scattering processes (Figure 2.8), has never been observed. This production was investigated at the LHC using the Run I data by the CMS and ATLAS collaborations [12, 13] and Run I and Run II data combined by CMS [14]. Up to now, only CMS has reported evidence of the process. One of the reasons that make this process difficult to observe, is that the same $Z\gamma jj$ final state can also be produced by a strongly-mediated process ($Z\gamma jj$ -QCD), with a lowest order cross section proportional to electroweak coupling and the strong coupling to the power two. This cross section is about two order of magnitude larger than the electroweak cross section (Figure 2.8). The $Z\gamma jj$ production via electroweak and strong mechanism interfere since the initial and final states are the same. In this chapter, first the selection of the $Z\gamma jj$ -EW candidate events and the estimation of the non- $Z\gamma jj$ -EW candidate events passing this selection, denoted as background, is presented. A careful object and event selection is important since it ensures a low background contamination. It follows the description of the background estimation methodology and the study on the interference effect between the $Z\gamma jj$ -EW and $Z\gamma jj$ -QCD. The investigation of

the interference effect between the two processes and its impact on the measurement is one of my contributions to the analysis and is described in detail.

5.1 Object selection

In order to disentangle the signal from the background events, several requirements are applied to select the jets, the leptons and the photons used in the analysis. It may happen that the pattern recognition algorithm reconstructs the same object twice. To avoid ambiguities, an overlap removal procedure is applied as will be explained below. In the following, the selection criteria of jets, leptons and photons are described.

5.1.1 Jet selection

Jets are reconstructed from topological clusters [77] using the anti- k_t [78] algorithm with a radius parameter of $R = 0.4$. The energy of jets is calibrated using corrections derived from simulation and data-driven methods [79]. A multivariate combination of track based variables (jet vertex tagger, JVT [80]) is used to suppress jets with $p_T < 60$ GeV originating from pile-up in the ID acceptance ($|\eta| < 2.4$).

Jets are required to:

- have $p_T > 30$ GeV and $|\eta| < 4.5$,
- pass an overlap removal procedure in which selected jets overlapping with an electron within $\Delta R < 0.2$ and selected jets with less than three tracks overlapping with muons within $\Delta R < 0.4$ are discarded.

Jets in the ID region with $p_T > 30$ GeV containing b-hadrons are tagged using a multivariate discriminant (MV2c10) with 70% b-tagging efficiency [81]. The b-tagged jets are used to control the background.

5.1.2 Lepton selection

In the offline analysis electron and muon candidates are required to originate from the primary vertex. One of the variables used for this selection is the impact parameter. The transverse impact parameter (d_0), is defined as the distance of closest approach of a track particle to the interaction point in the plan transverse to the beam-line. Usually, the significance of the impact parameter ($d_0/\sigma(d_0)$) is used where $\sigma(d_0)$ is the resolution of the impact parameter. The significance of the impact parameter plays an important role to reduce pile-up events and background events containing leptons from b-hadron decays. Because of the b-hadron lifetime, these leptons are expected to have larger d_0 significance compared to the leptons which emerge from the primary vertex. The longitudinal impact parameter (z_0), is defined as the z coordinate at the point of the track closest approach to the interaction point. Also this parameter is used to discriminate against pile-up events and particles with long lifetime. In most analyses, instead of applying directly the requirement on the z value, the $|z_0 \sin \theta|$

is used instead. That it is done because the $|z_0 \sin \theta|$ is independent of the track direction, and thus independent of the track resolution.

Muon candidates are reconstructed by combining measurements from the ID and the MS. Their momentum is calibrated before applying any object selection.

Muons are required to:

- have $p_T > 20$ GeV and $|\eta| < 2.5$,
- satisfy the following requirement on the impact parameter [82]: $|d_0/\sigma(d_0)| < 3$ and $|z_0 \sin \theta| < 0.5$ mm,
- satisfy the *Medium* quality selection, determined by the *MuonSelectionTool* [83], which is based on requirements on the number of hits in the ID and MS,
- satisfy the *FixedCutLoose* isolation requirement, which includes calorimeter isolation ($\text{topoetcone20}/p_T < 0.2$) and track isolation ($\text{ptvarcone20}/p_T < 0.15$) criteria,
- pass an overlap removal procedure in order to discard muons originating from the decay of hadrons in a jet: a muon is removed if it is found within $\Delta R < 0.4$ with a jet which survived the jet overlap removal [84].

Electron candidates, as described in Chapter 4, are reconstructed from clusters of energy deposits in the electromagnetic calorimeter, matched to an ID track. The energy-momentum of reconstructed electrons is calibrated using the *ElectronPhoton-FourMomentumCorrection* [85] package.

In the offline analysis **electrons** are required to:

- have $p_T > 20$ GeV and $|\eta| < 2.47$ (excluding the calorimeter transition region $1.37 < |\eta| < 1.52$),
- satisfy a *Medium* identification criterion [86, 87] based on the Likelihood method,
- satisfy the following requirement on the impact parameter: $|d_0/\sigma(d_0)| < 5$ and $|z_0 \sin \theta| < 0.5$ mm,
- satisfy the *Gradient* isolation requirement, which includes a calorimetric isolation ($\text{topoetcone20}/p_T < 0.2$) and track isolation ($\text{ptvarcone20}/p_T < 0.15$) criteria,
- pass an overlap removal procedure: an electron is removed if it is found within $0.2 < \Delta R < 0.4$ with a jet surviving the jet overlap removal.

The above criteria select samples with isolated prompt leptons of high transverse momentum.

5.1.3 Photon selection

Photon candidates are reconstructed from clusters of energy deposits in the ECAL without matched tracks or matched to two tracks which form a vertex in the ID (the latter occurring from converted photons). The four-momentum of reconstructed photons are calibrated using the *ElectronPhotonFourMomentumCorrection* [85] package. The energy measurement of converted photon candidates is further improved with corrections, evaluated as a function of the conversion radius based on dedicated MC simulation studies.

Photons are required to:

- have $E_T > 15$ GeV and $|\eta| < 1.37$ or $1.52 < |\eta| < 2.37$,
- satisfy a *Tight* cut-based identification criteria [87],
- satisfy the *FixedCutLoose* isolation requirement, which includes calorimeter isolation ($\text{topoetcone20}/p_T < 0.065$) and track isolation ($\text{ptvarcone20}/p_T < 0.05$) [88] criteria,
- pass an overlap removal procedure: a photon candidate is removed if it is found within $\Delta R < 0.4$ of an electron candidate.

5.2 Event selection

The event selection starts by applying quality requirements to the recorded data. The events have to be included in the *GoodRunList* (GRL) [89] which defines data taking periods with a good operation and conditions of the ATLAS subdetectors and of the accelerator. During the LHC operation dysfunctions may occur that deteriorate the data quality. Such dysfunctions include unstable beam conditions, magnets ramping (up or down), or detector noise. Information about these occurrences is compiled together and stored in a GRL in order to decide if the data are reliable. Selected events are further required to have a primary vertex with at least two associated tracks and to satisfy trigger selection criteria. Tables 5.1 and 5.2 summarize the trigger selection for the 2015 data taking period and 2016 respectively [90]. Single lepton and dilepton triggers are used. The use of the dilepton triggers allows to lower the lepton transverse momentum threshold to 20 GeV in the offline analysis.

Events are required to have at least two leptons of same flavor and opposite charge (SFOC) and exactly one photon candidate satisfying the selection criteria described above. If there are more than one pair of SFOC leptons, the pair with the invariant mass closest to the mass of the Z boson is selected. The sum of the di-lepton mass and the three-body $\ell\ell\gamma$ invariant mass is required to be larger than 182 GeV. This requirement suppresses the Z boson radiative decays. To select the $Z\gamma jj$ -EW process, events are required to have at least two jets with p_T above 50 GeV (tagging jets), if there are more than two jets with p_T above 50 GeV, the two most energetic jets are chosen as tagging jets. The high value of the p_T cut is justified by the need to reduce

Muon Channel	Electron Channel
<i>Single-muon triggers:</i> <i>HLT_mu20_iloose,</i> <i>HLT_mu50</i>	<i>Single-electron triggers:</i> <i>HLT_e24_lhmedium,</i> <i>HLT_e60_lhmedium,</i> <i>HLT_e120_lhloose</i>
<i>di-muon trigger</i> <i>HLT_2mu14</i>	<i>di-electron trigger</i> <i>HLT_2e12_lhloose</i>

TABLE 5.1: Trigger selections for 2015 data events for the electron and muon channel. Events are required to be selected by at least one of the listed trigger items.

Muon Channel	Electron Channel
<i>Single-muon triggers:</i> <i>HLT_mu26_ivarmedium</i> <i>HLT_mu50</i>	<i>Single-electron triggers:</i> <i>HLT_e26_lhtight_nod0_ivarloose,</i> <i>HLT_e60_lhmedium_nod0,</i> <i>HLT_e140_lhloose_nod0</i>
<i>di-muon trigger</i> <i>HLT_2mu14</i>	<i>di-electron trigger</i> <i>HLT_2e17_lhvloose_nod0</i>

TABLE 5.2: Trigger selections for 2016 data events for the electron and muon channel. Events are required to be selected by at least one of the listed trigger items. Due to the higher event rate in 2016 the single lepton trigger threshold is increased to 26 GeV for both leptons.

the systematic uncertainties due to jet energy scale. The invariant mass of the two jets is defined as:

$$m_{jj} = \sqrt{((E_{j1} + E_{j2})^2 - (\vec{p}_{j1} + \vec{p}_{j2})^2)} \quad (5.1)$$

and is required to be greater than 150 GeV, in order to reduce the contribution of triboson events ($pp \rightarrow VVV + X \rightarrow (\ell\ell)(jj)V + X$). Due to the $Z\gamma jj$ -EW topology, the signal candidate events are characterized by a diboson system centrally located compared to the two tagging jets. For that reason, a variable called centrality of the $Z\gamma$ system ($\zeta(Z\gamma)$) is computed (Equation 2.21). In order to remove events with poor resolution in $\zeta(Z\gamma)$, events with $\zeta(Z\gamma) > 5$ are removed. Finally a cut on the pseudorapidity difference between the tagging jets $\Delta\eta_{jj} > 1.0$ is required since it reduces the interference effect between the $Z\gamma jj$ -EW and $Z\gamma jj$ -QCD production (as it will be explained later). Table 5.3 summarizes the event selections which define the baseline data sample.

Lepton	a pair of SFOC with $p_T^\ell > 20$ GeV $ \eta_\ell < 2.47(2.5)$ for $e(\mu)$ remove e if $\Delta R(e, \mu) < 0.1$
Boson mass	$m_{\ell+\ell^-} > 40$ GeV $m_{\ell+\ell^-} + m_{\ell+\ell-\gamma} > 182$ GeV
Photon	$E_T^\gamma > 15$ GeV $ \eta_\gamma < 2.37$ (excl. $1.37 < \eta_\gamma < 1.52$) remove γ if $\Delta R(\ell, \gamma) < 0.4$
Jet	$N_{jet} \geq 2$, $p_T^{jet} > 30$ GeV, $ \eta_{jet} < 4.5$ remove jets if $\Delta R(\ell, jet) < 0.3$ OR $\Delta R(\gamma, jet) < 0.4$ $\Delta\eta_{jj} > 1.0$
VBS baseline selection	$p_T > 50$ GeV of two tagging jets $ \eta < 4.5$ of two tagging jets $m_{jj} > 150$ GeV $\zeta(Z\gamma) < 5$ $\Delta\eta_{jj} > 1$

TABLE 5.3: Summary of the event selection criteria defining the baseline data sample.

5.3 Background events

Several processes can mimic the experimental signature of the $Z\gamma jj$ -EW production. These processes, collectively called background, are classified into two groups: "irreducible background", which are processes with at least two prompt leptons and a photon, and "reducible background" which are processes passing the signal selection due to a jet which is mis-identified as a photon or a lepton.

The dominant contribution to reducible background processes, comes from the Z +jets production. A smaller contribution comes from WZ events with a lepton being mis-identified as a photon. The dominant contribution to the irreducible background processes comes from the $Z\gamma jj$ -QCD production. A smaller contribution comes from $t\bar{t}\gamma$ production.

5.4 Selection of $Z\gamma jj$ signal and control regions

In this analysis two approaches are used to discriminate the signal from the background. The main one is a multivariate approach, where a Boosted Decision Tree (BDT) is built by using several kinematic variables in order to compute a discriminating variable called BDT score. To cross-check the result, a cut-based approach is developed using the centrality of the $Z\gamma$ system as discriminating variable. The choice of the centrality as the sensitive variable is largely inspired from the analogous $Z\gamma$ VBS analysis performed in Run I [12].

In order to constrain the two main irreducible background (the $Z\gamma jj$ -QCD and the $t\bar{t}\gamma$) two orthogonal regions for the multivariate approach and three orthogonal regions for the cut-based approach are defined (see Figure 5.1). The region common to both approaches, is enriched in $t\bar{t}\gamma$ events and is called "b-CR". This region includes events satisfying the criteria in Table 5.3 and with at least one identified b-jet. It is

used as a validation region in order to check the modeling of the $t\bar{t}\gamma$ background and to correct its normalization.

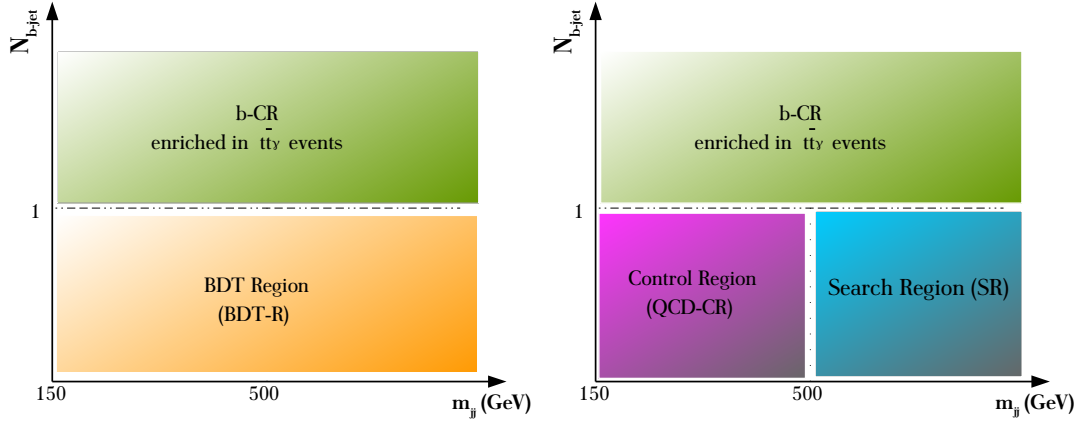


FIGURE 5.1: Illustration of the phase space separation for the multivariate approach (left) and the cut-based approach (right).

In the multivariate approach, a second region ("BDT-R") requires no b-jet. It has the particularity to be enriched in $Z\gamma jj$ -QCD background for the low values of BDT score, and in $Z\gamma jj$ -EW signal for the high values of the BDT score. This region is used both as a control region to constrain the normalization of the $Z\gamma jj$ -QCD background and as search region to extract the $Z\gamma jj$ -EW cross section.

In the cut-based approach, the event sample satisfying the BDT region selection is split into two parts. The first part is enriched in $Z\gamma jj$ -QCD background by requiring that the invariant mass between the two jets is less than 500 GeV and no b-jet. The corresponding region referred as "QCD-CR", is used as a control region to constrain the normalization of the $Z\gamma jj$ -QCD background. The second sample, corresponds to the region called search region "SR", is enriched in $Z\gamma jj$ -EW events by requiring a large dijet invariant mass greater than 500 GeV and no b-jet. In the cut-based approach the SR is the fiducial region of the analysis, where the cross section of the EW signal is extracted with a fit to the centrality distribution.

The selections defining the above described four regions, are summarized in Tables 5.4 and the yield of events in these regions is presented in Table 5.5.

$Z\gamma jj$ Event selection			
b-CR	BDT-R	QCD-CR	SR
$N_{b\text{-jet}} > 0$	$N_{b\text{-jet}} = 0$	$N_{b\text{-jet}} = 0$	$N_{b\text{-jet}} = 0$
-	-	$m_{jj} < 500$ GeV	$m_{jj} > 500$ GeV

TABLE 5.4: Event selections defining the four regions (b-CR, BDT-R, QCD-CR and SR) described in the text. These selections are applied in addition to the baseline criteria.

From Table 5.5 it can be observed that in the SR the fraction of expected signal is at the level of 20%, while the fraction of $Z\gamma jj$ -QCD is about 60%. In the QCD-CR, the fraction of $Z\gamma jj$ -QCD is about 85% and in the b-CR the percentage of $t\bar{t}\gamma$ is about 70%. In the BDT-R, the QCD background dominates with a contribution of 80%,

	Generator	SR	QCD-CR	b-CR	BDT-R
$Z\gamma jj$ -QCD	SHERPA v2.2.2	295 ± 18	818 ± 29	105 ± 11	1113 ± 34
$Z + jets$	SHERPA v2.2.1	58 ± 37	82 ± 19	14 ± 4	140 ± 41
$t\bar{t}\gamma$	MADGRAPH+PYTHIA8	9.0 ± 0.6	25 ± 1	187 ± 3	34 ± 1
$WZjj$	SHERPA v2.2.1	1.7 ± 0.1	2.7 ± 0.2	0.3 ± 0.0	4.4 ± 0.2
Single top	POWHEG+PYTHIA6	1.0 ± 0.5	1.3 ± 0.4	4.0 ± 0.8	2.3 ± 0.7
$Z\gamma jj$ -EW (signal)	MADGRAPH+PYTHIA8	76 ± 1	28 ± 1	4.8 ± 0.2	104 ± 1
Total MC		440 ± 41	957 ± 35	315 ± 12	1398 ± 53
Data		355	866	388	1222

TABLE 5.5: Expected and observed event yields in the four regions described in the text. The backgrounds considered are obtained from MC simulation. Only the statistical uncertainties for the expected number of events are presented.

followed by the Z +jets ($\sim 10\%$). In this region the signal is at the level of 7%. It can be observed that in all regions except the b-CR, the sum of all predictions results in more events compared to data (between 10% and 25%). By contrast, in the b-CR, the predicted event sum is about 19%, lower compared to the data. This indicates that there is a mis-modeling of the MC background predictions, which is addressed in the following sections.

5.5 Background estimation

To discover a possible excess of the $Z\gamma jj$ -EW signal events, the expected backgrounds in data must be well understood. However, while MC simulations may be conceived as a rough approximation of the true background contributions, the associated modeling uncertainties are too large to provide a sufficient and reliable estimate. Therefore methods to extract the main backgrounds from the experimental data themselves are used.

The shape of the $Z\gamma jj$ -QCD distributions is obtained using MC and the normalization is corrected with a fit to data. For the $t\bar{t}\gamma$ background, a dedicated control region "b-CR" is built, as already mentioned, to address its impact. All other backgrounds (WZ and single top), are estimated from MC simulations because their contribution is found to be small. In the following sections, the calculation and extraction methods of the Z +jets, $Z\gamma jj$ -QCD, $t\bar{t}\gamma$ which are the main backgrounds are described.

5.5.1 Z +jets background

For estimating the background from events with a Z -boson and a jet mis-identified as a photon, the data-driven ABCD method (also called two-dimensional side-band method) is used.

This method is one of the most commonly used data-driven background estimation techniques based on the extrapolation of background-enhanced two-dimensional side-band distributions into regions where an excess of signal events is expected. The

ABCD method relies on two nearly uncorrelated variables: the calorimeter-based isolation and the photon identification criteria which define four regions. Region A ("signal region") is enriched in prompt photons, regions B, C, D ("background regions") are enriched in jet faking a photon. Figure 5.2 illustrates the four regions.

The definition of the four regions is given below.

Signal Region

- **A** (Tight and isolated photon sample)

Events with isolated photon candidates ($E_T^{iso,(\Delta R < 0.2)}/E_T < 0.065$) satisfying the *Tight* identification selection criteria.

Control Regions

- **B** (Tight and non-isolated photon sample)

Events with non-isolated photon candidates ($E_T^{iso,(\Delta R < 0.2)}/E_T > 0.065$) and satisfying the *Tight* identification selection criteria.

- **C** (Anti-tight and isolated photon sample)

Events with isolated photon candidates ($E_T^{iso,(\Delta R < 0.2)}/E_T < 0.065$) and satisfying *anti-Tight* identification selection criteria, obtained by reversing some identification cuts.

- **D** (Anti-tight and non-isolated photon sample)

Events with non-isolated photon candidates ($E_T^{iso,(\Delta R < 0.2)}/E_T > 0.065$) and satisfying the *anti-Tight* identification selection criteria.

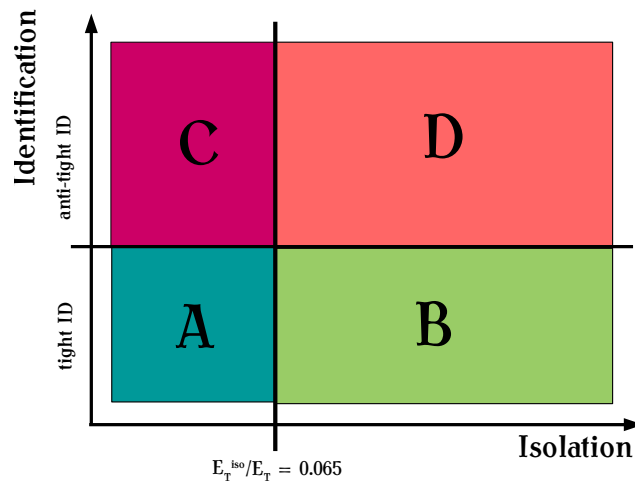


FIGURE 5.2: Graphic illustration of the ABCD regions in the two-dimensional plane. Region A, which requires *Tight* and isolated photons, is the signal region and regions B, C and D are the background control regions.

The ABCD method is applied to a sample of events selected in a region with high statistics and extrapolated to the analysis signal region. The sample used for

the background estimation, is selected by requiring at least two opposite sign leptons, exactly one photon and at least two jets with $p_T > 30$ GeV and invariant mass of the leading jets $m_{jj} < 150$ GeV, and by vetoing the presence of any jet tagged as b-jet.

The number of events N_i in each region $i = A, B, C, D$ can be expressed as:

$$N_A = N_A^{Z\gamma} + N_A^{Zjet} + N_A^{EW} \quad (5.2)$$

$$N_B = c_B \cdot N_A^{Z\gamma} + N_B^{Zjet} + N_B^{EW} \quad (5.3)$$

$$N_C = c_C \cdot N_A^{Z\gamma} + N_C^{Zjet} + N_C^{EW} \quad (5.4)$$

$$N_D = c_D \cdot N_A^{Z\gamma} + N_D^{Zjet} + N_D^{EW} \quad (5.5)$$

where:

- $N_A^{Z\gamma}$ is the number of $Z\gamma jj$ -QCD events in the A region.
- $c_i = \frac{N_i^{Z\gamma}}{N_A^{Z\gamma}}$ are leakage coefficients representing the fraction of $Z\gamma jj$ -QCD events in region i ($i = B, C, D$) with respect to the region A.
- N_i^{Zjet} is the number of events with jets mis-identified as photons in region i ($i = A, B, C, D$).
- N_i^{EW} is the number of events in region i ($i = A, B, C, D$) coming from processes like: WZ , single top, $Z\gamma jj$ -EW and $t\bar{t}\gamma$.

Of these quantities the N_i are measured directly from data, while the c_i and N_i^{EW} are estimated from MC as they do not involve any estimate of the photon mis-identification probability and are usually small corrections. The expected contribution from $Z\gamma jj$ -EW process is negligible.

In this analysis, a likelihood implementation of the ABCD method is used in which the predicted event rates in each region is re-written as:

$$N_A = N_A^{Z\gamma} + N_A^{Zjet} + N_A^{EW} \quad (5.6)$$

$$N_B = c_B \cdot N_A^{Z\gamma} + \eta_B \cdot N_A^{Zjet} + N_B^{EW} \quad (5.7)$$

$$N_C = c_C \cdot N_A^{Z\gamma} + \eta_C \cdot N_A^{Zjet} + N_C^{EW} \quad (5.8)$$

$$N_D = c_D \cdot N_A^{Z\gamma} + \eta_B \cdot \eta_C \cdot R_{MC} \cdot N_A^{Zjet} + N_D^{EW} \quad (5.9)$$

where $\eta_j = N_j^{Zjet}/N_A^{Zjet}$ with $j = B, C$. For the estimation of the N_D^{Zjet} in Equation 5.5, the following formula of the ABCD method is used:

$$N_D^{Zjet} = R_{MC} \cdot \frac{N_C^{Zjet}}{N_A^{Zjet}} \cdot N_B^{Zjet} = R_{MC} \cdot \eta_B \cdot N_C^{Zjet}, \quad (5.10)$$

where R_{MC} is a correction factor, determined using MC, which accounts for possible deviations from the assumption that the isolation and identification variables are not correlated.

The variables η_B and η_C are treated as nuisance parameters that are free to vary in the fit to the data. The other parameters, c_i and N_i^{EW} , are fixed from MC simulation.

The parameters of interest that are fitted in this equation are the $N_A^{Z\gamma}$ and N_A^{Zjet} . These yields, as well as the values of the nuisance parameters, are obtained from a maximum-likelihood fit to the observed number of events N_i . The fractions c_i are of the order of few percent.

The estimation of the Z +jet background in the signal region A (N_A^{Zjet}) should not depend on the definition of the B, C and D control regions. Three variations of the isolation criteria, defining these regions, are studied to estimate the impact on the Z +jet evaluation. There are also multiple ways to invert the set of tight selection criteria and define an *anti-Tight* photon candidate. The result of the ABCD method, should not depend on the choice of the *anti-Tight* definition. The main criteria adopted, in order to choose the isolation criteria and the *anti-Tight* definition, is to minimize the correlation between the isolation and identification variable as predicted by MC simulation such that the R_{MC} factor stays close to one.

Four different versions of *anti-Tight* criteria have been tested. These *anti-Tight* definitions are obtained by requiring that some of the criteria on the shower shape variables computed using the calorimeter first layer that are included in the *Tight* definition are failed. The *anti-Tight* definition and the isolation criteria that gives the smallest statistical error and the value of R_{MC} closest to one is chosen as the baseline choice. The corresponding value of R_{MC} is 1.0 ± 0.2 where the uncertainty derives from the largest deviation observed from the central value when using different isolation definitions of the ABCD regions.

The number of events and the criteria finally adopted for the definition of the regions are given in Table 5.6.

Identification	Isolation	R_{MC}	N_A^{data}	N_B^{data}	N_C^{data}	N_D^{data}	N_A^{EW}	N_B^{EW}	N_C^{EW}	N_D^{EW}	N_A^{Zjet}	$N_A^{Z\gamma}$
LP4	0.065 GeV	1.0 ± 0.2	866	135	224	118	68	5	6	1	431 ± 64	2026 ± 89

TABLE 5.6: Results of the Z +jets background estimate from data. LP4 corresponds to a given *anti-Tight* identification criteria.

The result is therefore found to be:

- $N_A^{Z\gamma} = 2026 \pm 89$ (stat) ± 68 (sys)
- $N_A^{Zjet} = 431 \pm 64$ (stat) ± 15 (sys)

The systematic uncertainties are evaluated from the variation of the result when using different criteria for the ABCD region definition.

Extrapolation to the signal phase space

Since the number of events in the search region is limited, a determination of the photon fake rate in this region is not possible. As said before, to increase the size of the sample used for the estimate of the Z +jets background, it is required that $m_{jj} < 150$ GeV. Thus, in order to estimate the Z +jets contribution in the signal region, a procedure is developed to extrapolate the result to the high m_{jj} region.

This estimate is carried out using a combination of MC and data. The MC simulation is expected to reproduce reasonably well the dependence of the photon fake rate on m_{jj} , since the m_{jj} selection is weakly related to the simulation mis-modeling of the fake rate. The latter is related to the specific hadronization/showering of the jets. Following the Run I approach [12], a data-to-MC correction factor is measured on data at low m_{jj} and then is applied to the MC prediction at high m_{jj} . This factor which represents the rate of the fake photon events is defined as:

$$f_{\text{fake}} = \frac{N^{Z\text{jet}}}{N^{Z\gamma}} \quad (5.11)$$

This procedure is validated with a MC study which shows that f_{fake} does not depend within uncertainties on the m_{jj} selection. The rate of fake photon events in the high dijet invariant mass region is therefore evaluated as:

$$f_{\text{fake}} = 0.21 \pm 0.04(\text{stat}) \pm 0.02(\text{sys}) \quad (5.12)$$

For the estimation of the shape of the distribution of the Z +jets the region D of the data is used after having subtracted the other MC contributions. It is found that the statistical uncertainty of the template distribution is large enough to cover any differences observed with any other templates, therefore no systematic uncertainties on the shape of the background is considered.

5.5.2 Initial normalization of the $Z\gamma jj$ -QCD background

The ABCD method allows the simultaneous measurement of the Z +jets and the $Z\gamma jj$ -QCD contributions in region A. The data-to-MC ratio of $Z\gamma jj$ events in this region for the nominal generator SHERPA v2.2.2 is found to be $0.91 \pm 0.04(\text{stat}) \pm 0.03(\text{sys})$. In the analysis, ultimately, the $Z\gamma jj$ -QCD background normalization in the signal region is estimated with a fit. The ratio cited above is applied to MC as normalization factor to obtain the input parameter of the fit, and is also used to show the pre-fit comparison between data and MC (control plots).

5.5.3 Evaluation of the $t\bar{t}\gamma$ background

As introduced in Section 5.4 in order to compute from data the normalization of the $t\bar{t}\gamma$ background, a control region for the $t\bar{t}\gamma$ is built both for the multivariate and the cut-based approaches (b-CR). This control region is used in the fit to constrain the normalization of the $t\bar{t}\gamma$, by using an additional parameter (see Section 5.7). The variable used to constrain the $t\bar{t}\gamma$ background is the number of tagged b-jets in the event. It is found that the normalization of the $t\bar{t}\gamma$ background must be scaled by a factor of $1.41 \pm 0.05(\text{stat})$. The use of a control region reduces the uncertainties on the modeling of the $t\bar{t}\gamma$ background, by adding a constrain from the data. The MADGRAPH sample is used to model the shape and the POWHEG+PYTHIA6 and

POWHEG+PYTHIA8 are used to estimate the systematic uncertainty deriving from the shape modeling including the parton shower effect.

5.5.4 Control distributions

Table 5.7 summarizes the total number of observed and expected events, after applying corrections to the background estimate previously described. After this first step, the level of agreement between data and the sum of all backgrounds is improved.

	SR	QCD-CR	b-CR	BDT-R
$Z\gamma jj$ -QCD	268 ± 16	745 ± 25	96 ± 10	1013 ± 30
Z +jets	56 ± 3	156 ± 6	20 ± 4	213 ± 7
$t\bar{t}\gamma$	13 ± 1	35 ± 2	264 ± 4	48 ± 2
$WZjj$	1.7 ± 0.1	2.7 ± 0.2	0.3 ± 0.0	4.4 ± 0.2
Single top	1.0 ± 0.5	1.3 ± 0.4	4.0 ± 0.8	2.3 ± 0.7
$Z\gamma jj$ -EW (signal)	76 ± 1	28 ± 1	4.8 ± 0.2	104 ± 1
Total MC	415 ± 16	968 ± 26	389 ± 10	1384 ± 31
Data	355	866	388	1222

TABLE 5.7: Event yields in the four regions considered in this analysis for the signal and all the backgrounds considered. The uncertainties are statistical only, except for the $Z + jets$ contribution, where the data-driven uncertainty is used.

Control distributions of the main kinematic variables are presented in Figures 5.3, 5.4, 5.5 and 5.6 for the following regions:

- Inclusive region (no jet requirement are applied),
- b-CR,
- QCD-CR,
- BDT-R.

Figure 5.7 shows the $Z\gamma$ centrality in the signal region, QCD-CR, b-CR and BDT region. In these distributions, the Z +jets background shape is obtained using the $Z\gamma jj$ -QCD shape as it was done in the Run I analysis [12]. The normalization of the Z +jets and $Z\gamma jj$ -QCD samples is described in Sections 5.5.1 and 5.5.2, respectively. The $t\bar{t}\gamma$ distribution is normalized with the value from the fit. In these figures only the statistical uncertainties on the data and the MC samples are shown. Agreement in the shapes of the distributions between data and expectation is observed.

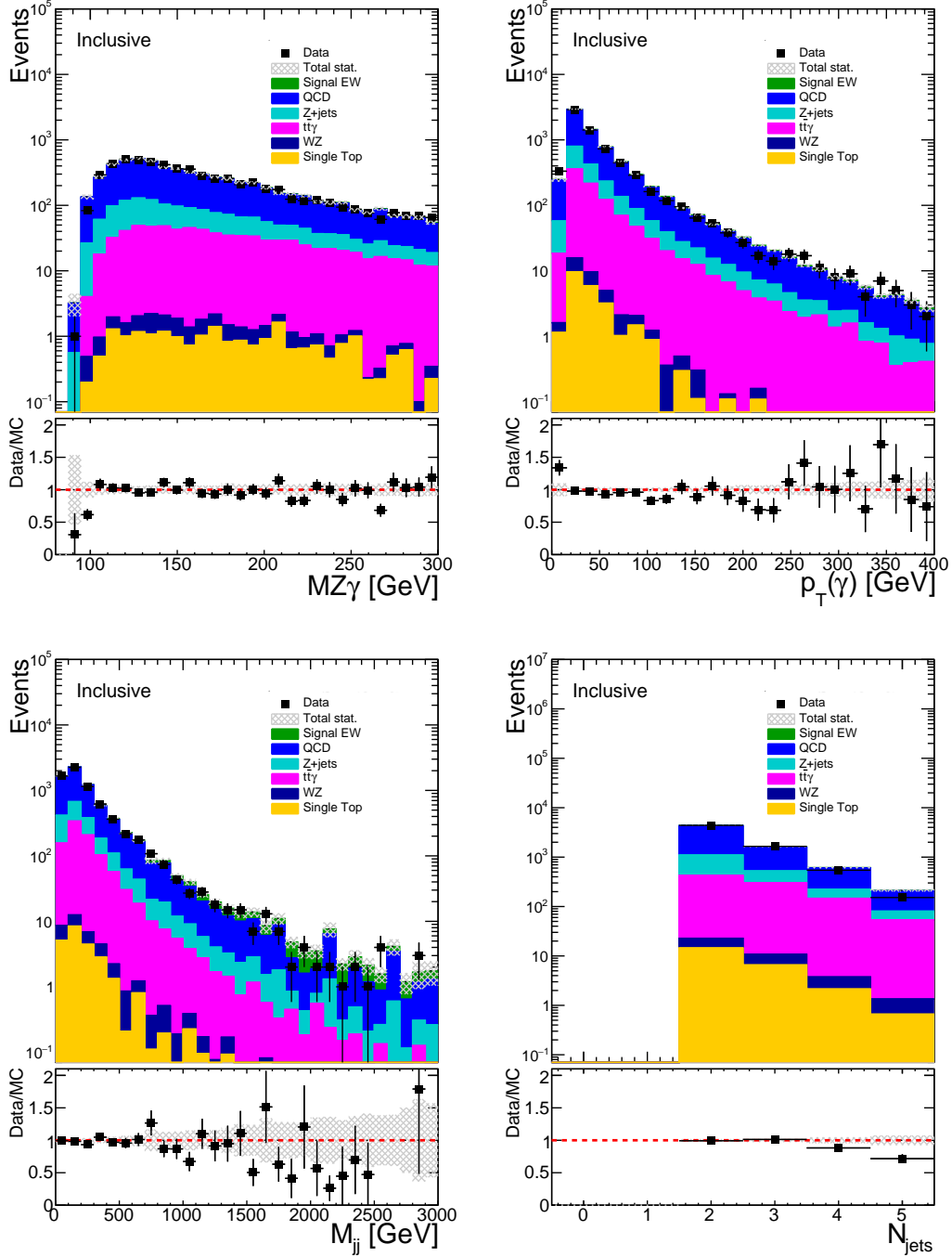


FIGURE 5.3: Distribution of the $M_{Z\gamma}$, $P_T(\gamma)$, M_{jj} and N_{jets} in the Inclusive region. Only the statistical errors on the data and the MC samples is shown. The last bin contains overflows.

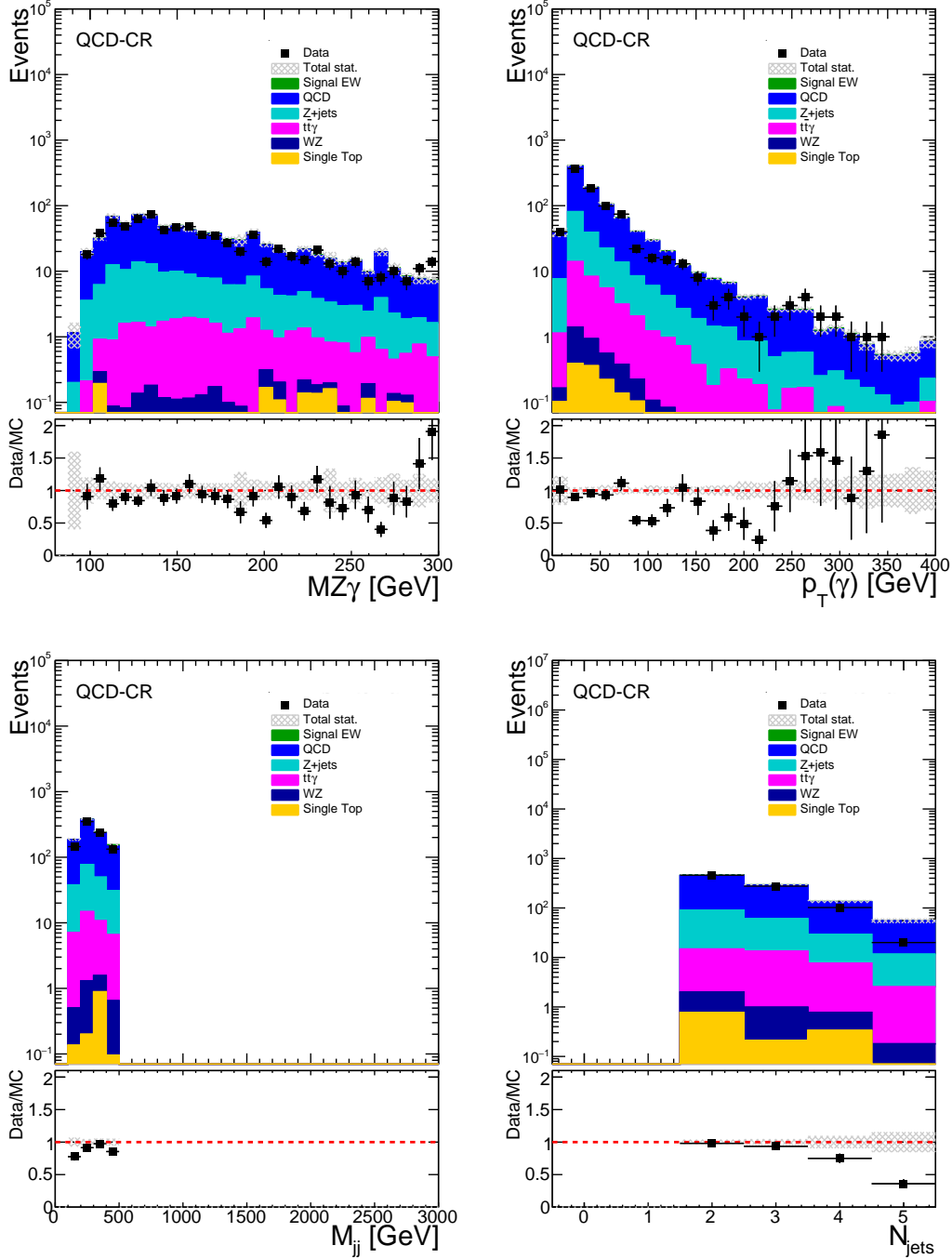


FIGURE 5.4: Distribution of the $M_{Z\gamma}$, $P_T(\gamma)$, M_{jj} and N_{jets} in the QCD-CR. Only the statistical errors on the data and the MC samples is shown. The last bin contains overflows.

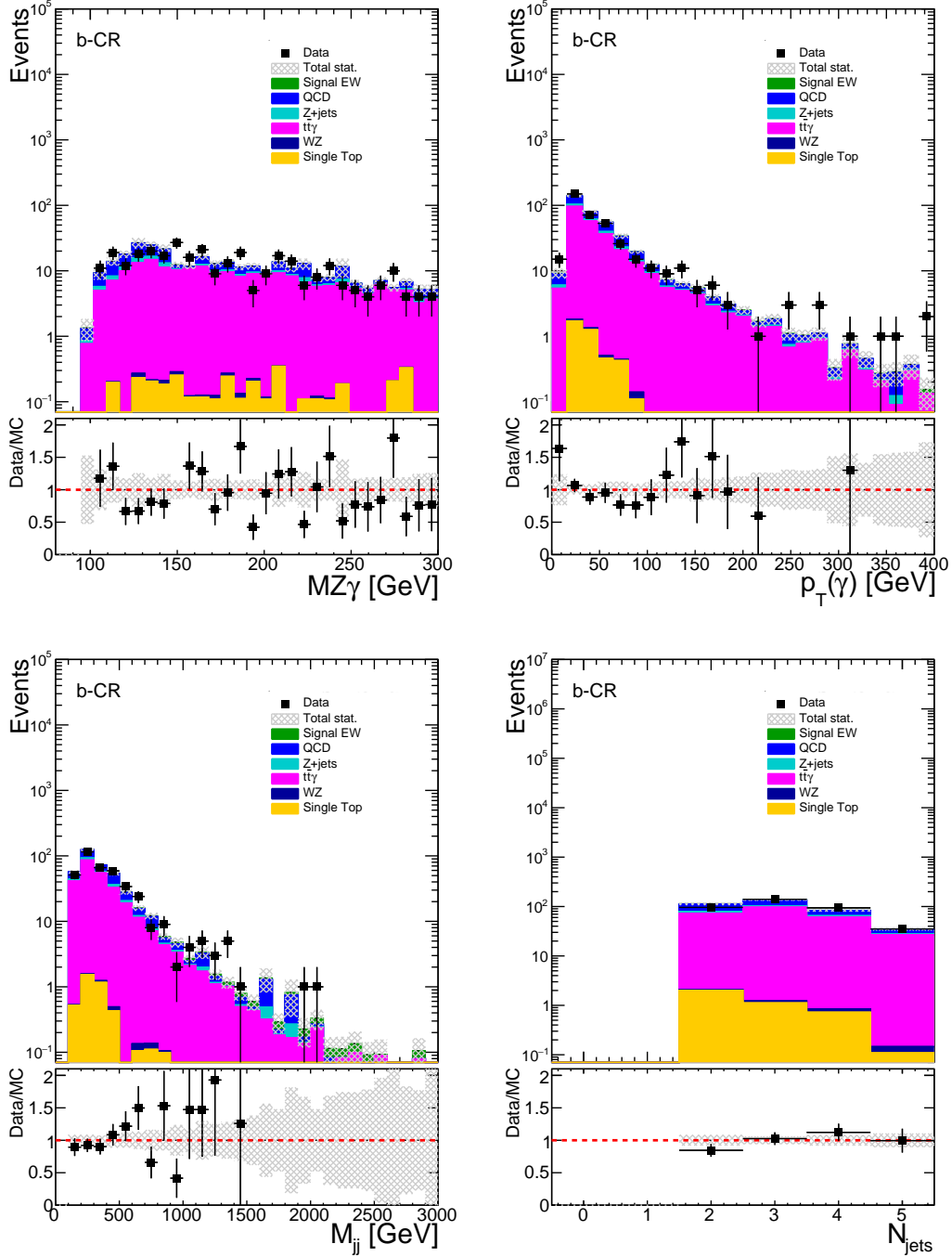


FIGURE 5.5: Distribution of the $M_{Z\gamma}$, $P_T(\gamma)$, M_{jj} and N_{jets} in the b-CR. Only the statistical errors on the data and the MC samples is shown. The last bin contains overflows.

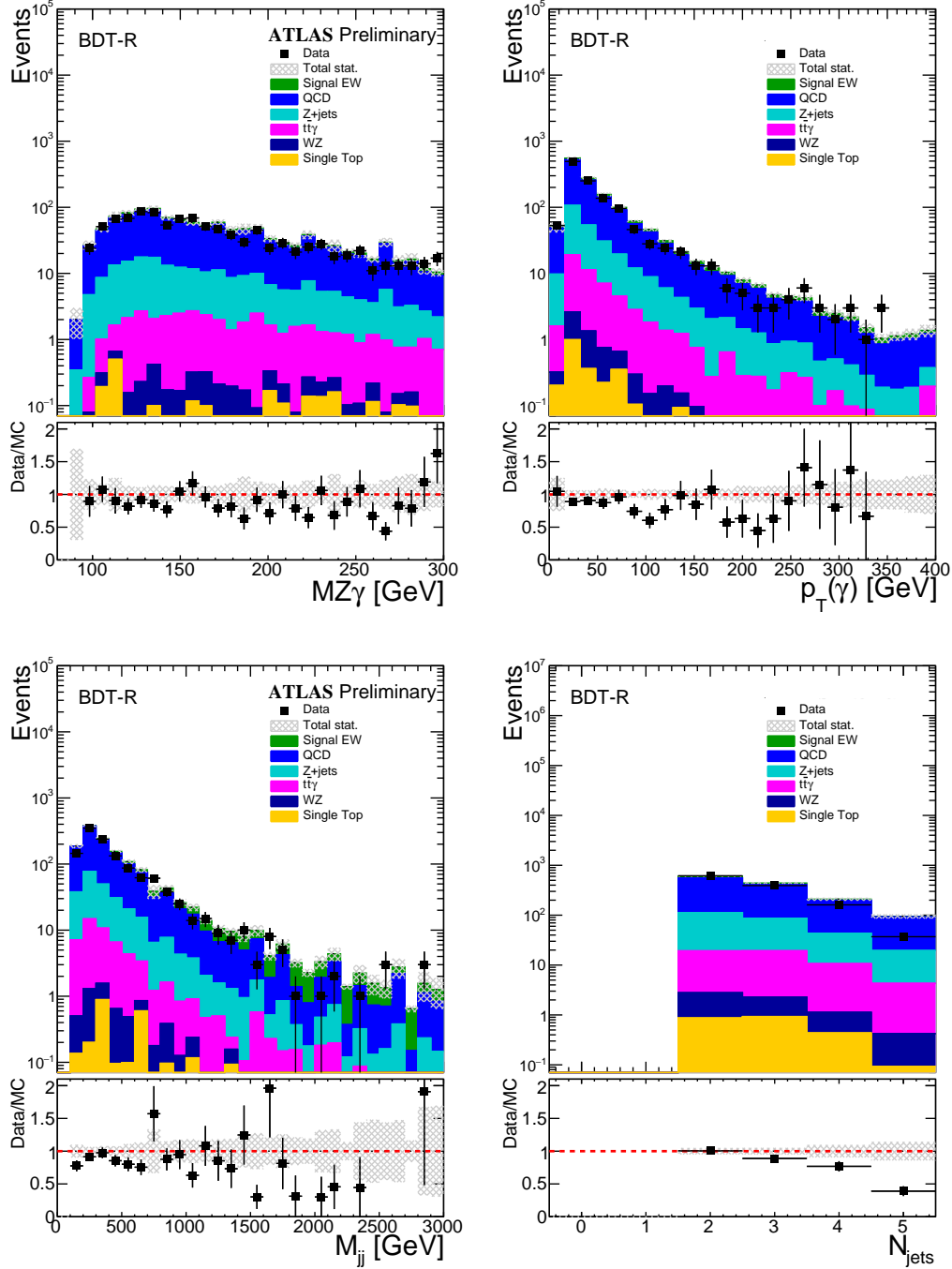


FIGURE 5.6: Distribution of the $M_{Z\gamma}$, $p_T(\gamma)$, M_{jj} and N_{jets} in the BDT-R. Only the statistical errors on the data and the MC samples is shown. The last bin contains overflows.

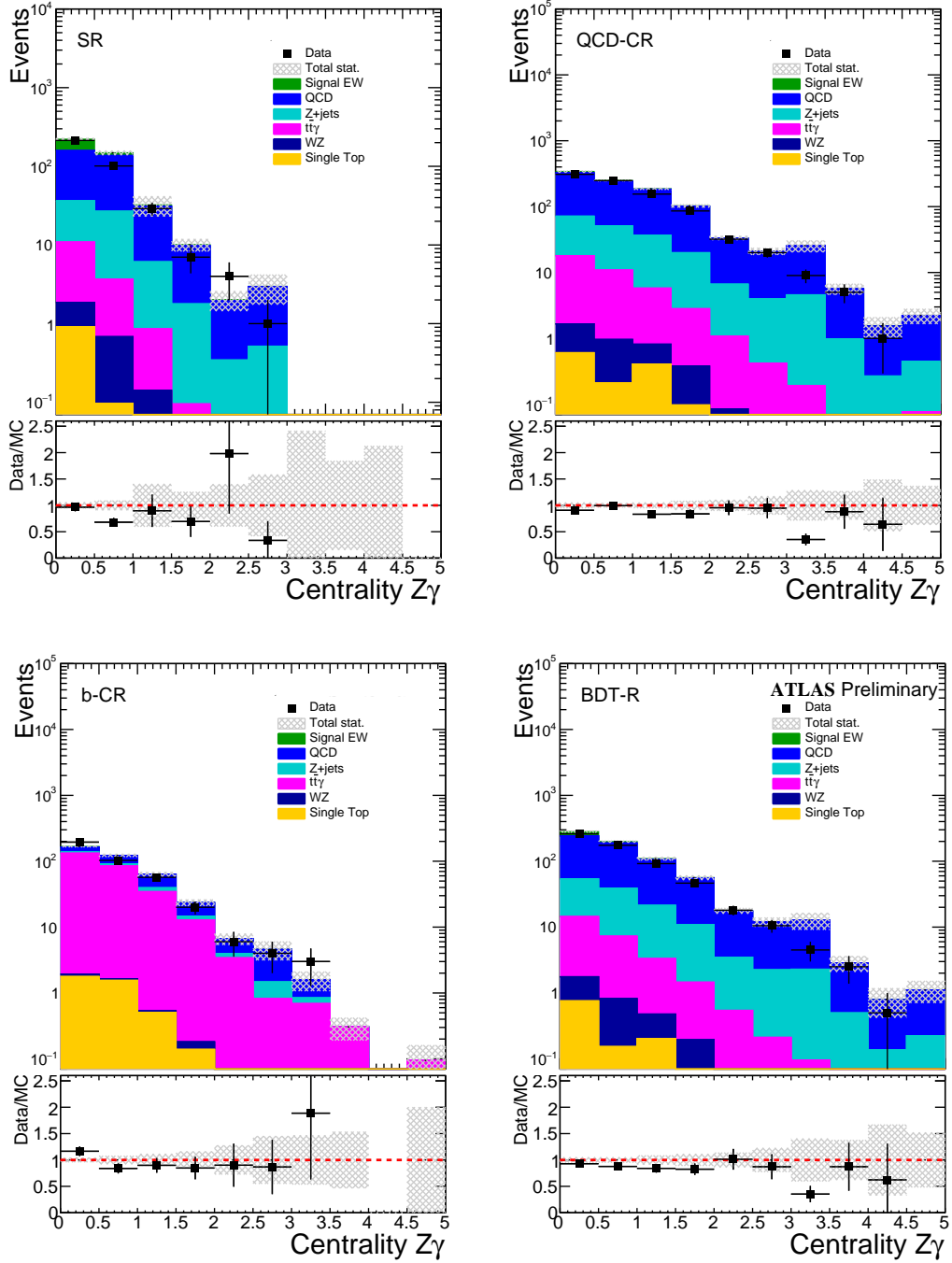


FIGURE 5.7: Distribution of the $Z\gamma$ centrality in the signal region (top left plot), QCD-CR region (top right plot), b-CR (bottom left plot) and BDT region (bottom right plot).

5.6 Multivariate analysis

The main method to extract the signal is based on a Multivariate Analysis (MVA) using a BDT algorithm from the ROOT TMVA package [62]. Compared to other MVA procedures, the BDT is found to be faster to test, train, and optimize, and thus has been chosen. As a result of the procedure, distributions of BDT scores are produced, for each signal and background process in the BDT region. These distributions are then used in a fitting procedure, in order to extract the signal strength described in Section 5.7.

Training of the BDT

The BDT algorithm is a classification tool capable of distinguishing signal and background events. This is achieved by compressing multidimensional discriminatory information down to one single variable called BDT score, which disentangles signal and background. Given that the response of the algorithm is trained with MC, it is important to understand the description by the simulation of the input variables used in this multivariate analysis.

The MC samples, signal and backgrounds, are split into two statistically independent sub-samples. The first set of sub-samples is used for the training of the BDT and is therefore called "training sample". As output, BDT score distributions are produced separately for signal and background. Once trained, the BDT can then be tested on the second set of sub-samples (the "test sample") and the resulting BDT score distributions are compared.

Due to large fluctuations in event weights, the Z +jets MC sample, is not included in training and testing procedures. Instead the shape of the $Z\gamma jj$ -QCD MC sample is used, with a normalization of 0.21 times the cross section of the $Z\gamma jj$ -QCD MC sample, as determined with the ABCD method, since it is found to model the shape of the Z +jets contribution within uncertainties.

BDT Optimization

The BDT construction has been optimized, both in terms of the input variables used and algorithm settings. Many input variables, including kinematic variables of the objects used in the analysis, jet multiplicity and variables of combined objects are investigated. Variables that are found to be mis-modeled in MC samples are excluded. The choice of input variables and the BDT algorithm settings are optimized at the same time. For the optimization, the "N-1" method is used. According to this method, the BDT is initially trained using a "default set" of variables, and its performance is evaluated. Afterwards, one variable or a group of variables is removed each time, and the BDT is retrained and re-evaluated. This process is repeated for all variables and groups of variables using always the same BDT algorithm settings ("default set"). In a similar way the BDT algorithm settings are varied one at the time using the default setting of variables and the training and testing procedure is repeated. The figure of merit for the BDT performance evaluation which drives the choice of variables and settings is the area under the ROC (Receiver Operating Characteristic) curve [91]. A ROC curve is a performance measurement for classification problems. It tells how

much a model is capable of distinguishing between classes (signal from background in this case), this is shown schematically in Figure 5.8. The higher the ROC curve, the better the model is at distinguishing between signal and background. Variables are ranked in terms of importance by testing the improvement of the ROC curve integral after removing a variable or a set of variables from the BDT. The variable-removal or settings-change that gives the greatest improvement is then used.

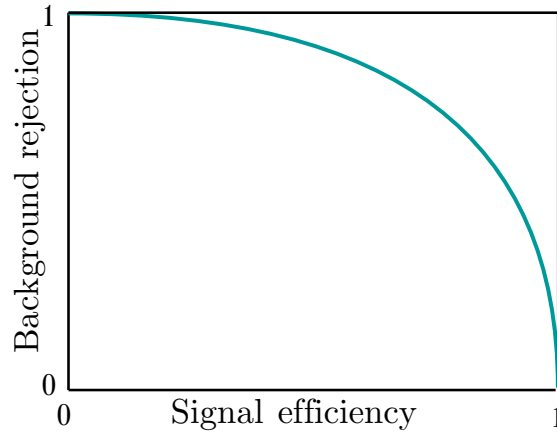


FIGURE 5.8: Shape of a generic ROC curve.

Final chosen BDT configuration

The final set of thirteen input variables used is given in Table 5.8, ranked in terms of importance following the training and testing procedure. This ranking takes into account the signal-background separation power of each variable as well as the effect of possible correlations between variables. A variable with a great separation power can have a low importance on the rank because it is highly correlated with another variable, which in turn has played an important role in the BDT discrimination.

If correlations between variables are found in MC samples but not in the data, that means that the MC does not describe the data well, and thus a BDT trained and tested with simulated events will not perform equally well on data. For this reason, not only individual input variables distributions are compared between data and MC, but also their correlations. The difference between data and MC of the correlation factors between each pair of the final thirteen variables is found to be small.

As can be seen from Table 5.8, the most important discriminant power comes from two correlated variables $p_T(Z\gamma)$ and $m_{Z\gamma}$ and from the centrality $\zeta(Z\gamma)$, followed by the invariant mass of the two tagging jets (m_{jj}).

The final expected BDT distribution in the BDT region, normalized to the 36 fb^{-1} luminosity of data used in this analysis is shown in Figure 5.9.

Rank	Variable	Variable Importance (%)
1	$p_T(Z\gamma)$	22.3
2	Centrality	18.2
3	$m_{Z\gamma}$	18.1
4	m_{jj}	10.7
5	$\min\Delta R(\gamma, j)$	5.6
6	$p_T(Z)$	4.7
7	$p_T(j_1)$	4.0
8	$\Delta\eta(j_1, j_2)$	3.7
9	$\Delta\phi(Z\gamma, jj)$	3.7
10	$\eta(j_1)$	3.1
11	m_Z	3.1
12	$p_T(l_1)$	2.3
13	$\Delta R(Z\gamma, jj)$	0.3

TABLE 5.8: Final BDT input variables, ranked by importance for improvements in the BDT response.

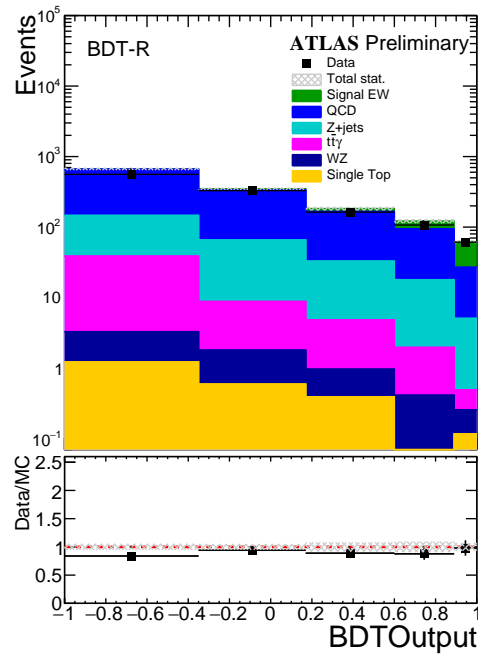


FIGURE 5.9: The observed and expected BDT output, in the BDT region.

5.7 Fit procedure

The signal strength μ_{EW} defined as the number of observed $Z\gamma jj$ -EW events divided by the expected:

$$\mu_{EW} = N_{\text{meas}}^{\text{EW}} / N_{\text{exp}}^{\text{EW}},$$

is extracted with a fit to the data using template distributions. In the multivariate approach, the template distributions are the BDT distributions in the BDT region

and the number of b-jets in the b-CR. In the cut-based approach, the template distributions are the centrality in the QCD-CR and the SR, and the number of b-jets in the b-CR.

The parameter of interest (POI) of the fit is μ_{EW} . The ratio between the observed and expected number of events for the $Z\gamma jj$ -QCD and $t\bar{t}\gamma$, denoted with μ_{QCD} and $\mu_{t\bar{t}\gamma}$, are also extracted from the fit as unconstrained parameters.

The template fit is performed using the HistFactory package from ROOT [92] and is based on a binned likelihood fit. Probability density functions (pdf) are built for the selected variable in the signal and control regions. For the signal region the pdf is:

$$\mathcal{P}(n_b|\mu) = Pois(n_{tot}|\mu S + B) \left[\prod_{b \in bins} \frac{\mu \nu_b^{sig} + \nu_b^{bkg}}{\mu S + B} \right] \quad (5.13)$$

where μ is the signal strength, S and B represent respectively the total expected yield of signal and background events, and ν_b^{sig} and ν_b^{bkg} the expected number of signal and background events in bin b . Similar expressions are used for the control regions. These pdfs also depend on a set of nuisance parameters (NP) which represent the impact of uncertainties on the fit and are constrained by Gaussian distributions. The total likelihood is the product of the likelihood in the signal and control regions. All the parameters implemented are fitted to minimize a test function referred to as "negative log likelihood" $NLL(\mu, \hat{\mu}, \hat{\theta}, \hat{\theta})$, and defined as:

$$NLL(\mu, \hat{\mu}, \hat{\theta}, \hat{\theta}) = -2 \ln \left(\lambda(\mu, \hat{\mu}, \hat{\theta}, \hat{\theta}) \right) = -2 \ln \left(\frac{\mathcal{L}(\mu, \hat{\theta})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right), \quad (5.14)$$

where $\lambda(\mu, \hat{\mu}, \hat{\theta}, \hat{\theta})$ is the likelihood (\mathcal{L}) ratio, $\hat{\theta}$ is the NP combination that maximizes \mathcal{L} for a given value of the POI μ , and $\hat{\mu}$ and $\hat{\theta}$ the "pair" of parameters maximizing \mathcal{L} . The POI μ extracted from the fit is the one minimizing the NLL. The significance $Z_{meas.}$ of the measurement is estimated by looking at the value of the function for the background only hypothesis, corresponding to $\mu = 0$. Its value is computed as:

$$Z_{meas.} = \sqrt{NLL(\mu = 0)}. \quad (5.15)$$

Expected results are obtained using Asimov [93] pseudo-data. An Asimov dataset is a representative dataset and provides a quick method to obtain the median experimental sensitivity of a measurement as well as fluctuations about this expectation. The Asimov dataset is generated with a signal having the same cross section as expected in SM and is built from the templates provided for the fit.

5.8 Interference of strong and electroweak production

Some of the basic processes, as presented in Figure 5.10 leading to the $Z\gamma jj$ -EW and the $Z\gamma jj$ -QCD productions have identical initial and final states, therefore according to quantum mechanics they interfere with each other. In order to find a proper treatment of this effects, the investigation of the interference between the two production modes needs to be studied. First, the computation at LO of all contributions to the $Z\gamma jj$ cross section is presented. Afterwards, the interference effect is estimated in the BDT region and in the search region of the cut-based approach and a study of how to reduce it, is presented. Finally, the treatment of the interference effect in the analysis is described.

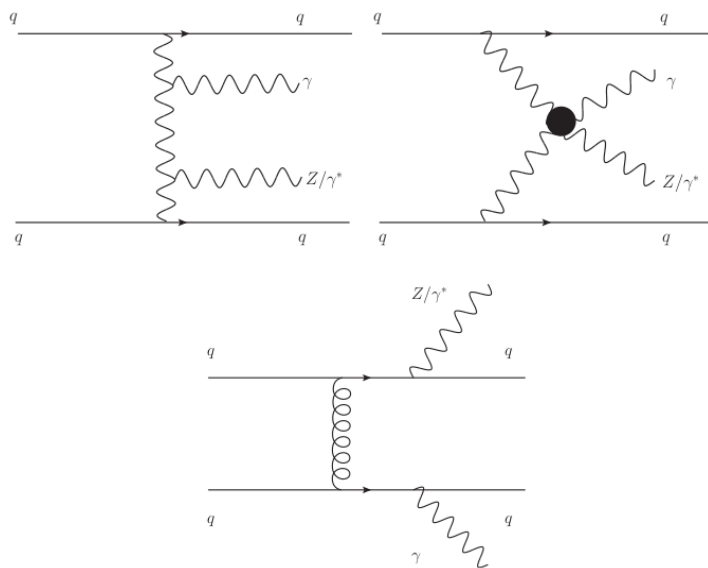


FIGURE 5.10: Tree level diagrams that contribute to the $Z\gamma jj$ process. The two top plots represent two diagrams leading to the $Z\gamma jj$ -EW production while the bottom plot represents one of the diagram leading to the $Z\gamma jj$ -QCD production [12].

5.8.1 Cross section calculation

In practice, there are two ways to evaluate the impact of the EW-QCD interference term on the total $Z\gamma jj$ cross section:

- **Direct way:** calculating directly only the EW-QCD interference contribution.
- **Indirect way:** calculating three quantities:
 - ◇ the cross section of the EW mediated process,
 - ◇ the cross section of the QCD mediated process,
 - ◇ the cross section of the complete (EW+QCD) process. In this case the contribution of the interference is computed by subtracting the EW and QCD cross sections from the complete cross section calculation.

The two methods give the same results when the samples have large statistics. Since the interference contribution to the cross section is much smaller than the QCD and EW contributions alone, to have a statistically significant results many events must be generated. Therefore the indirect method is computationally intensive and inefficient. For this reason, for this study, the direct method is chosen.

Among all the current available generators, only MADGRAPH [53] provides the possibility to compute the interference contribution directly. To save computing time all the calculations are done at LO.

Figure 5.10 reminds few typical diagrams that enter the $Z\gamma jj$ cross section calculation. The QCD contribution consist of diagrams where a gluon connects the two quark lines.

In the amplitude of the scattering process ("Matrix-Element" level), the purely $Z\gamma jj$ -EW contribution has order $\mathcal{O}(g_w^4)$, the $Z\gamma jj$ -QCD contribution has order $\mathcal{O}(g_S^2 g_w^2)$.

The cross section of the process is proportional to:

$$|M|^2 = |M_1 + M_2|^2 = |M_1|^2 + |M_2|^2 + 2 \times \text{Re}(M_2^* \times M_1)$$

where $|M_1|^2$ and $|M_2|^2$ represent the square of the amplitudes of the EW and QCD mediate parts respectively, and the last term gives rise to the interference. The interference component is specified in MADGRAPH using the settings at amplitude square level, while for the QCD and EW components the settings at the amplitude level are used. Therefore, the MADGRAPH settings to generate the three terms are:

- ◇ $Z\gamma jj$ -EW: QCD=0, QED=4.
- ◇ $Z\gamma jj$ -QCD: QCD=2, QED=2.
- ◇ Interference: QCD=2, QED=6.

Three samples of events are generated: a pure EW, a pure QCD and an interference sample. Table 5.9 summarizes the cuts in the generation.

Generation cuts				
Object	p_T cut	η cut	ΔR cut	invariant mass cut
jets	> 15 GeV	< 5.5		
photon	> 10 GeV	< 3.0	$\Delta R(m, k) > 0.1$	$m_{jj} > 0$ GeV
leptons	> 10 GeV	< 3.0	with $m = j, \ell, \gamma$	$m_{\ell\ell} > 40$ GeV
b-quarks	> 15 GeV	< 5.5	and $k = j, \ell$	

TABLE 5.9: Cuts applied in generation of the MADGRAPH samples.

For each of the three contributions (EW, QCD and interference) $N_{\text{tot}} = 10^6$ events are generated. For technical reasons the events are generated in 100 groups of 10^4 events. MADGRAPH provides for each group the corresponding cross section and an uncertainty, which is the mathematical precision of the calculations. For each of the three contributions, the cross section of the 100 samples are averaged and the standard deviation is computed as:

$$\sigma_{\text{gen}} = \sqrt{\frac{\sum_i \sigma_i}{N}}, \quad (5.16)$$

$$\Delta\sigma_{\text{gen}} = \frac{1}{N-1} \sqrt{\sum_i (\sigma - \sigma_i)^2}. \quad (5.17)$$

The standard deviation $\Delta\sigma_{\text{gen}}$ is found in all cases bigger than the MADGRAPH uncertainty on the single generated sample, and is taken as the uncertainty on the combined cross section σ_{gen} .

The resulting cross sections are:

$$\sigma_{\text{gen}}(\text{EW}) = 0.0472 \pm 0.0002 \text{ pb}$$

$$\sigma_{\text{gen}}(\text{QCD}) = 5.0505 \pm 0.0016 \text{ pb} \quad (5.18)$$

$$\sigma_{\text{gen}}(\text{INT}) = 0.0022 \pm 0.0002 \text{ pb} \quad (5.19)$$

5.8.2 Effect of the interference

In order to estimate the impact of the interference in the analysis, the generated events are required to satisfy the phase space selections which define the BDT-R in the multivariate approach and the SR in the cut-based approach. In both cases the cross section, in the corresponding phase space, is computed as:

$$\sigma = \sigma_{\text{gen}} \times \frac{N_{\text{passPS}}}{N_{\text{tot}}} \quad (5.20)$$

with an uncertainty

$$\Delta\sigma = \sigma \times \sqrt{\left(\frac{\Delta\sigma_{\text{gen}}}{\sigma_{\text{gen}}}\right)^2 + \frac{N_{\text{passPS}}}{N_{\text{tot}}^2} \left(1 - \frac{N_{\text{passPS}}}{N_{\text{tot}}}\right)} \quad (5.21)$$

The interference contribution with respect to the EW contribution is found to be:

$$\frac{\sigma(\text{INT})}{\sigma(\text{EW})} = (3.48 \pm 0.26)\%, \text{ in the BDT-R}$$

and

$$\frac{\sigma(\text{INT})}{\sigma(\text{EW})} = (1.87 \pm 0.14)\%, \text{ in the SR,}$$

where the uncertainties are purely statistical.

5.8.3 Optimization of the kinematic selection

In order to minimize further the impact of the interference term while keeping the significance of the pure $Z\gamma jj$ -EW contribution high, further investigations are performed. Figures 5.11 and 5.12, show the distributions of the invariant mass of the dilepton and dijet systems, the transverse energy of the photon, the centrality of the $Z\gamma$ system and the pseudorapidity difference $\Delta\eta$ of the tagging jets, corresponding to the EW and interference parts in the BDT region. These plots show that the $\Delta\eta_{jj}$, has the most discriminant power between the two contributions and can be used to minimize the impact of the interference.

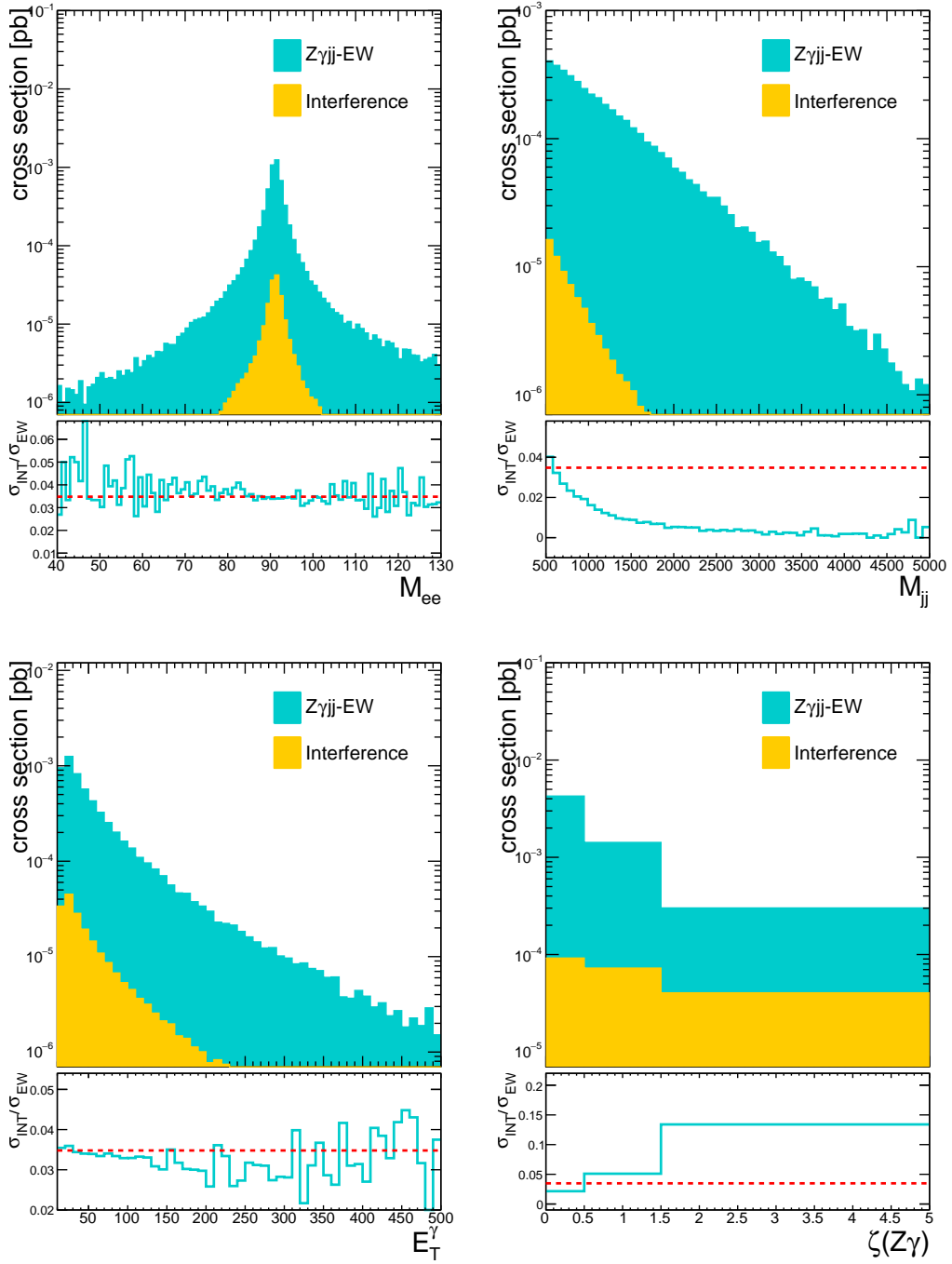


FIGURE 5.11: Distributions of M_{ee} , M_{jj} , E_T^γ , and $\zeta(Z\gamma)$ corresponding to the electroweak production and the interference term in the BDT region. The bottom panel represents the ratio of the distributions on the upper plots.

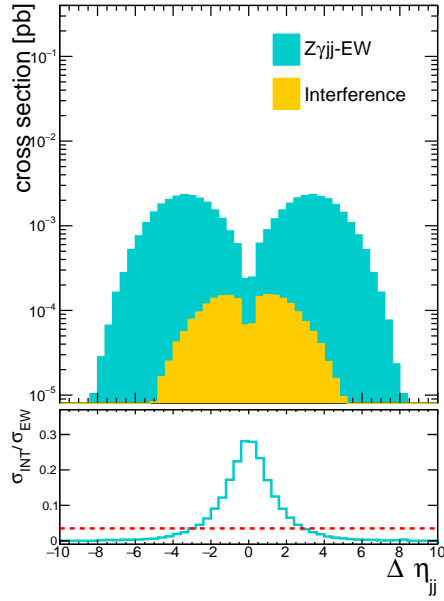


FIGURE 5.12: $\Delta\eta$ distribution of the dijet system, corresponding to the electroweak and interference production in the BDT region. The bottom panel represents the ratio of interference over EW events.

Four different selections on the value of $\Delta\eta_{jj}$ are considered: $|\Delta\eta_{jj}| > 0.5, 1, 1.5$ and 2 . The effect of all these selections is reported in Figure 5.13. This figure shows that the EW cross section gradually decreases with the increase of the $|\Delta\eta_{jj}|$ cut. The selection which corresponds to the middle point ($|\Delta\eta_{jj}| > 1$) is chosen as working point. After applying this selection, the impact of the interference in the electroweak cross section is reduced of $\sim 20\%$ while the EW is reduced of 4% .

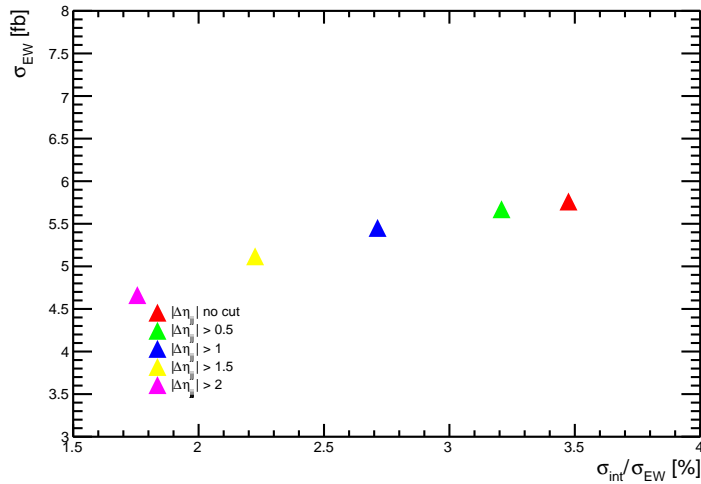


FIGURE 5.13: The electroweak cross section as a function of the interference contribution in the electroweak contribution in the BDT region.

In summary for the "BDT-R":

$$\frac{\sigma(\text{INT})}{\sigma(\text{EW})} = (2.72 \pm 0.20)\% \quad (5.22)$$

and for the cut-based analysis:

$$\frac{\sigma(\text{INT})}{\sigma(\text{EW})} = (1.62 \pm 0.12)\% \quad (5.23)$$

5.8.4 Treatment of the interference

The interference between the electroweak signal and the $Z\gamma jj$ -QCD process is not included in the fit to extract the signal. However since the interference a priori exists it is necessary to estimate the possible distortions that the interference term may have on the shape of the MC $Z\gamma jj$ -EW signal template used in the fit to the experimental data. This effect is included in the systematic uncertainty.

For this estimate, the sum of the contribution of the EW plus the interference term at particle level¹ is used to correct the signal shape at reconstruction level. For this aim centrality and BDT score distributions are created at particle level for the EW, QCD and interference contributions (Figure 5.14), in the SR and BDT-R phase space.

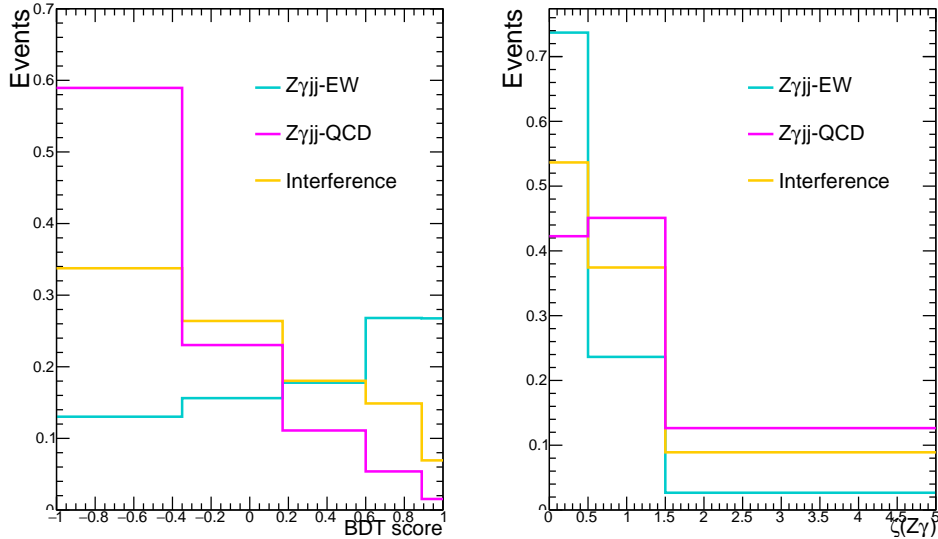


FIGURE 5.14: Distributions of the BDT output and of the centrality of the $Z\gamma$ system for the three $Z\gamma jj$ contributions at particle level. The distributions are normalized to one.

The BDT distribution is built using as input the same variables defined in Table 5.8, but at particle level. Three weights are computed for the centrality and five

¹The particle level is the simulation stage where stable particles, with lifetimes exceeding $c\tau > 10$ mm, are produced from either hard scattering or after hadronization, but before interacting with the detector.

for the BDT defined at particle level in the following way:

$$w_i = \frac{(\text{EW} + \text{INT})_i}{\text{EW}_i}, \quad (5.24)$$

where i indicates the bin number.

The resulting weights are reported in Tables 5.10 and 5.11.

BDT bin	weight
(-1,-0.35)	1.0455 ± 0.0015
(-0.35,0.17)	1.0198 ± 0.0016
(0.17,0.6)	1.0005 ± 0.0018
(0.6,0.89)	0.9873 ± 0.0022
(0.89,1)	0.9788 ± 0.0022

TABLE 5.10: Weights for the BDT distribution, calculated at generated level and applied at reconstructed level to account for the shape of the interference in the $Z\gamma jj$ -EW production.

Centrality bin	weight
(0, 0.5)	0.9779 ± 0.0014
(0.5,1.5)	1.0122 ± 0.0014
(1.5,5)	1.0914 ± 0.0015

TABLE 5.11: Weights for the centrality distribution, calculated at generated level and applied at reconstructed level to account for the shape of the interference in the $Z\gamma jj$ -EW production.

The MC reconstruction level centrality and BDT distributions of the $Z\gamma jj$ -EW nominal sample (MADGRAPH) is reweighted event by event, depending on the value of the particle level centrality and BDT, at particle level. The corresponding reco level distributions after a procedure of symmetrization are used as alternative signal template distributions, to assess the impact of the interference as systematic uncertainty on the shape of the template.

During the procedure described in the previous section, a mismatch is found between the number of reconstructed and truth events: 16% of the events that pass the reco level selection cuts do not satisfy the particle level phase space cuts. For these events it is not possible to attribute to the reconstructed events an appropriate particle level weight. To understand the origin of the mismatch, the cuts on the p_T and η of the leptons, photons and jets at particle level, were relaxed of a quantity corresponding to the p_T and η expected resolution. The percentage of the reco-truth mismatched events did not change significantly, bringing to the conclusion that the effect is related to the object reconstruction efficiency.

To solve this mismatch three methods are considered.

- **Method 1** rejects the events which do not have truth values in the phase space of the measurement.
- **Method 2** uses for the re-weight the reco BDT value instead of the particle level value.

- **Method 3** applies to these events a weight = 1.

The resulting templates are shown in Figure 5.15. From this figure it can be seen that the effect in method 2 and method 1 is identical and is bigger compared to method 3. Although the small systematic that is introduced, the selection of the appropriate method should be the one which introduces the minimum bias in the analysis. The smallest bias is achieved by the method in which the weight is chosen according to the reco BDT value (method 2) for the following reasons:

- Method 3, is the one which may result in the highest bias, as one has a non negligible population of events for which one says that the EW shape is the same as the EW+INT.
- Method 1, might introduce a bias in the case of events that are not distributed in terms of BDT in a similar way with respect to all events way (if for example there is no particle level BDT value mostly for low BDT events).

The method 1 and 2 give very similar results and this indicates that the events without particle level BDT do not have a particular BDT value. Method 2 is chosen to be implemented in the analysis.

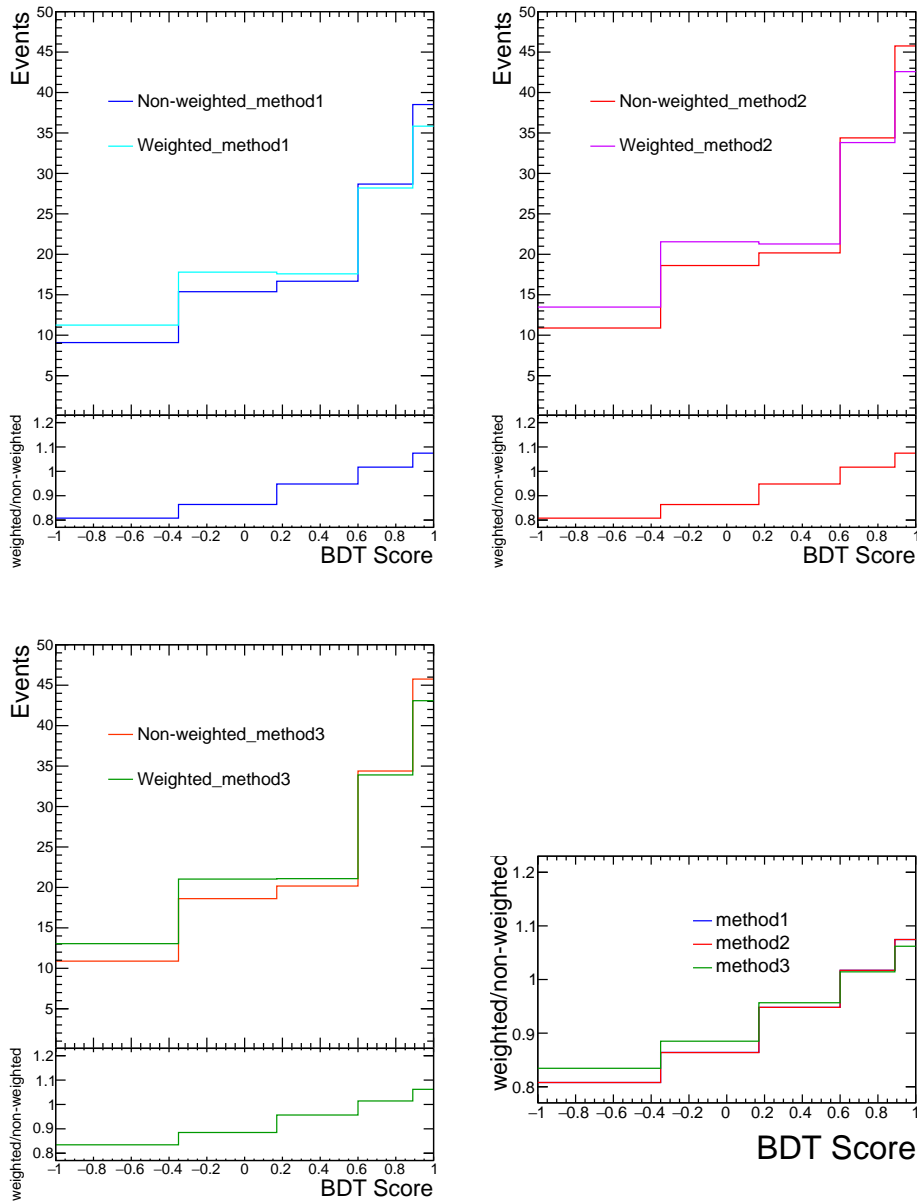


FIGURE 5.15: The BDT score distributions before and after the weighting procedure for method 1, 2 and 3. The right bottom plot shows the comparison of the distributions before and after the weighting.

5.9 Systematic uncertainties

5.9.1 Systematic uncertainties from reconstruction and background contribution

Several experimental sources of the systematic uncertainty are considered. These include uncertainties in the pile-up re-weighting procedure, in the electron and photon energy and muon momentum scale and resolution, in the jet energy scale and resolution, as well as uncertainties in the scale factors applied to simulation in order to reproduce the trigger, reconstruction, identification and isolation efficiencies measured

in data. These uncertainties affect both the signal and the MC-based background predictions. Table 5.13 shows them for the signal and the two main backgrounds.

The uncertainties in the normalization of reducible background events arising from mis-identified photons (Z +jets background) are determined using the data-driven ABCD method as detailed in Section 5.5.2 and is estimated to be $\pm 20\%$. An overall and conservative uncertainty of $\pm 20\%$ is used for the WZ and Single top. The uncertainty of the $Z\gamma jj$ -QCD background is estimated to be $\pm 3.6\%$ from the data-driven background extraction. For the combined 2015 and 2016 data sets the uncertainty on the integrated luminosity is estimated to be 2.1% [94].

5.9.2 Theory uncertainties on the predicted shapes

Theory uncertainties for the $Z\gamma jj$ -EW, $Z\gamma jj$ -QCD and $t\bar{t}\gamma$ include uncertainties due to QCD scale, PDFs and α_S variations. For the $Z\gamma jj$ -EW signal and $t\bar{t}\gamma$ background, parton shower and underlying event uncertainties are obtained with tune variations and reshowering events with HERWIG 7 [95]. For the $Z\gamma jj$ -QCD background, other shape uncertainties are accounted for by comparing two different generators: MADGRAPH and SHERPA v2.2.2. These generators use independent matrix element calculation and shower algorithm. The difference in shape of the BDT score distributions predicted in the BDT-R by SHERPA v2.2.2 and MADGRAPH is used to estimate a model uncertainty on the $Z\gamma jj$ -QCD signal MC template used in the template fit. The difference between the two predictions is symmetrized and used to define an uncertainty band around the SHERPA v2.2.2 prediction. Table 5.12 summarizes the theory uncertainties considered for each sample.

Process	$Z\gamma jj$ -QCD	$Z\gamma jj$ -EW	$t\bar{t}\gamma$
Interference		✓	
Scale	✓	✓	✓
PDF	✓	✓	✓
Comp. gen.	✓		
Parton shower		✓	✓
Underlying event		✓	

TABLE 5.12: Summary of the theory uncertainties considered for the different processes.

Systematic Group	EW	QCD	$t\bar{t}\gamma$
Jets	3.5%	8.4%	5.2%
Leptons and photons	3.3%	3.9%	2.5%
Pileup	2.7%	0.02%	3.7%
Theory	1.8%	1.1%	20.7%
Flavour tagging	0.5%	0.8%	8.8%
MC statistic	3.5%	19.1%	44.9%

TABLE 5.13: Systematic uncertainties in the normalization of the QCD, EW and $t\bar{t}\gamma$ sample in the BDT-R, used as input in the fit procedure.

5.10 Fit results

The measured signal strength with a fit in the BDT region as described in Section 5.7 is:

$$\mu_{EW} = 1.00_{-0.18}^{+0.19} (\text{stat})_{-0.10}^{+0.08} (\text{MCstat})_{-0.08}^{+0.09} (\text{syst})_{-0.10}^{+0.13} (\text{theo}).$$

The measurements performed here formally corresponds to the electroweak production including the interference effects because the interference effect between the $Z\gamma jj$ -QCD and the $Z\gamma jj$ -EW processes is not accounted for in the $Z\gamma jj$ -QCD template. The result is statistically limited. The value of the other unconstrained parameters ($\mu_{QCD}, \mu_{t\bar{t}\gamma}$) are compatible with one within the uncertainties. The observed significance is 4.1σ (with 3.8σ expected). Compatible results are obtained with the fit to the centrality variable.

Chapter 6

Cross section measurements

Αρχή ήμισυ παντός.

Πλάτων, 427-347 π.Χ.

In the present chapter, the measurement of the integrated cross section of the $Z\gamma jj$ -EW process and the differential cross section of the $Z\gamma jj$ process are described. Detector effects (resolution, efficiency and geometric acceptance) distort the distribution of the measured observables. The procedure of correcting these distortions is known as *unfolding*. Unfolding methods are widely used in many analyses in the ATLAS experiment to correct the experimental distributions. This procedure is applied in order to facilitate the comparison between data and predictions and will be also described.

6.1 Integrated cross section measurement

The $Z\gamma jj$ -EW process is measured in a phase space region defined at the particle level, using the objects and event kinematic selection criteria as described in Table 6.1. This phase space region is defined as close as possible to the phase space where the experimental measurement is performed (fiducial phase space), therefore the particle level selection corresponds to the reconstruction selection described in Table 5.3. Some small extrapolations from the experimental phase space to the particle level phase space is applied to correct for the acceptance loss in the calorimeter transition region ($1.37 < |\eta| < 1.52$) for photons, and in the high η region ($2.47 < |\eta| < 2.5$) for electrons. The particle-level phase space choice minimize the impact of the theory uncertainties.

Jets at the particle level are reconstructed in MC events by applying the anti- k_t jet reconstruction algorithm with a radius parameter $R = 0.4$ to all final state stable particles after having excluded prompt muons and electrons. To account for the effect of the final state QED radiation, the energy of the generated leptons at particle level is defined as the energy of the lepton after radiation, plus the energy of all radiated photons within $\Delta R < 0.1$ around the lepton direction ("dressed leptons").

The photon isolation variable E_T^{cone20} at particle level, is estimated by performing the vector sum of all the stable particles (except neutrinos and muons) within a radius

of $\Delta R = 0.2$ around the photon, but excluding the photon itself. The value of the isolation cut used in this analysis is $E_T^{cone20}/E_T^\gamma < 0.05$ [96].

Given that at the particle level, there is no way of knowing which photon is the one from the hard scattering, the selection requirements are applied to all photons that do not originate from hadron decay. Photons from parton fragmentation are therefore included. If more than one photon passes all selection requirements, the one with the highest energy is chosen.

Objects	Particle Level Selection
Leptons	$p_T^\ell > 20$ GeV and $ \eta^\ell < 2.5$ Dressed leptons, OS charge
Photon Kinematic	$p_T^\gamma > 15$ GeV, $ \eta^\gamma < 2.37$ $\Delta R(\ell, \gamma) > 0.4$
Photon Isolation	$E_T^{cone20}/E_T^\gamma < 0.05$
FSR cut	$M_{\ell\ell} + M_{\ell\ell\gamma} > 182$ GeV $M_{\ell\ell} > 40$ GeV
Truth Jets/Outgoing Partons ($p =$ outgoing quarks or gluons)	At least two jets with $E_T^{jet} > 50$ GeV, $ \eta^{jet} < 4.5$ $\Delta R(\ell, jet) > 0.3$ $\Delta R(\gamma, jet) > 0.4$
Search Region	$M_{jj} > 500$ GeV, $N_{bjets} = 0$
BDT Region	$M_{jj} > 150$ GeV, $N_{bjets} = 0$

TABLE 6.1: Phase space selection at particle level. The leptons at particle level are dressed as described in the text.

6.1.1 Fiducial cross section extraction

The measurement of the fiducial cross section is carried out using the measurement of the signal strength $\mu_{EW} = N_{meas.}^{EW}/N_{exp.}^{EW}$ reported in the previous chapter.

The observed fiducial cross section is expressed as:

$$\sigma_{meas.}^{fid. EW} = \frac{N_{meas.}^{EW}}{C \times \mathcal{L}}, \quad (6.1)$$

where $N_{meas.}^{EW}$ is the number of signal events measured in the signal region, C is the event reconstruction efficiency which is $C = \frac{N_{exp.}^{EW, reco.}}{N_{exp.}^{EW, gen.}}$, and \mathcal{L} is the data luminosity.

The expected fiducial cross section is expressed as:

$$\sigma_{exp.}^{fid. EW} = \frac{N_{exp.}^{EW}}{C \times \mathcal{L}}, \quad (6.2)$$

where $N_{exp.}^{EW}$ is the number of simulated signal events expected in the signal region at reconstructed level for the data luminosity \mathcal{L} .

Therefore

$$\sigma_{obs.}^{fid. EW} = \frac{N_{meas.}^{EW}}{N_{exp.}^{EW}} \times \sigma_{exp.}^{fid. EW} = \mu_{EW} \times \sigma_{exp.}^{fid. EW}. \quad (6.3)$$

The SM prediction from MADGRAPH for the $Z\gamma jj$ -EW process without interference¹ in the BDT region is:

$$\sigma_{\text{exp. EW}}^{\text{fid.}} = 7.75 \pm 0.03(\text{stat}) \pm 0.2(\text{PDF}) \pm 0.4(\text{scale}) \text{ fb.} \quad (6.4)$$

The observed $Z\gamma jj$ -EW production cross section, derived from the signal strength μ_{EW} , in the BDT region is:

$$\sigma_{\text{obs. } Z\gamma jj\text{-EW}}^{\text{fid.}} = 7.75_{-1.39}^{+1.47} (\text{stat.})_{-0.99}^{+0.93} (\text{exp.syst})_{-0.78}^{+1.01} (\text{mod.syst}) \pm 0.15(\text{lumi}) \text{ fb.} \quad (6.5)$$

Table 6.2 shows the effect of the uncertainties on the fiducial cross section measurement ($\sigma_{\text{obs. } Z\gamma jj\text{-EW}}^{\text{fid.}}$), where the individual sources are grouped in either theoretical or experimental uncertainties. The main source of theoretical uncertainty comes from the $Z\gamma jj$ -EW theory modeling followed by the $Z\gamma jj$ -QCD theory modeling. For the experimental uncertainty the main impact comes from the background modeling followed by the MC statistics and uncertainties on jets.

Source	Uncertainty [%]
$Z\gamma jj$ -EW theory modeling	+9.6
$Z\gamma jj$ -QCD theory modeling	-6.0
$t\bar{t}\gamma$ theory modeling	+5.5
$Z\gamma jj$ -EW and $Z\gamma jj$ -QCD interference	-6.1
Jets	+2.2
Pile-up	-1.6
Electrons	+3.2
Muons	-2.1
Photons	+7.6
Electrons/photons scale	-8.2
b-tagging	+6.4
MC statistics	-4.3
Backgrounds normalization	+0.7
Luminosity	-0.5
Total Systematics	+2.8
	-2.2
	+1.2
	-0.6
	+0.5
	-0.5
	+1.9
	-1.7
	+7.6
	-8.2
	+8.9
	-8.2
	2.1
	+27.3
	-25.0

TABLE 6.2: Relative uncertainties on the measured fiducial cross section $\sigma_{Z\gamma jj\text{-EW}}^{\text{fid.}}$. The uncertainties are expressed in percentages. The correlations between these uncertainties is taken into account in the computation of the total uncertainty.

The analysis measure also the full $Z\gamma jj$ cross section (QCD + EW) in the same fiducial phase space. The procedure uses the number of events in the BDT region, the efficiency correction C and the luminosity \mathcal{L} and results in:

$$\sigma_{\text{obs. } Z\gamma jj}^{\text{fid.}} = 71.4 \pm 2.4 (\text{stat.})_{-6.5}^{+8.9} (\text{exp.syst})_{-17.0}^{+21.1} (\text{mod.syst}) \pm 0.15(\text{lumi}) \text{ fb.} \quad (6.6)$$

¹The effect of the interference is small, of the order of 3% as can be seen in the Figure 5.13, and its effect is negligible compared to the other uncertainties.

6.2 Differential cross section measurements

In this analysis differential cross sections as function of the photon transverse momentum (P_T^γ), the invariant mass of the diboson system ($M_{Z\gamma}$), the invariant mass of the dijet system (M_{jj}) and the jet multiplicity (N_{jets}) are measured. The choice of the variables is driven by the considerations that they are expected to be sensitive in the high value tails to aQGCs, or have the potential to separate between $Z\gamma jj$ -EW and $Z\gamma jj$ -QCD processes and to constrain the MC description of $Z\gamma jj$ production.

Experimental distributions are corrected for detector effects with an unfolding procedure. They are, together with the MC simulation for the description of detector effects, used as input to the unfolding methods [97]. The statistical and systematic uncertainties affecting the experimental measurements and/or the simulation are propagated through the unfolding procedure. The resulting corrected measurements with their uncertainties can be directly compared to the corresponding theoretical calculations, but also to the results from other experiments. In principle, measured and predicted distributions may be also compared at reconstruction level namely after having corrected the predictions for acceptance, efficiency and resolution. This is maybe the easiest solution when the aim of the analysis is to estimate some parameters in a given theoretical framework.

Nevertheless, without using unfolding methods, the measurement of one experiment cannot be compared directly to the results of other experiments, because the effects of resolution are in general different in different experiments. Moreover, if a new theory is developed many years after that a measurement is carried out, the information needed to correct the theory for the detector effects, may no longer be available. Efforts are presently done in the experiments to preserve all information for future use. In any case, ultimately the choice of the level at which the comparison is performed depends on the particular problem the analyzers deal with.

6.2.1 Principle of the unfolding method

An observed event, from a given signal process, is characterized by two quantities: a *true value* t , which is unknown, and an observed or *reconstructed value* r . Neglecting resolution effects, the probability that an event leads to some measured value is called the *detector efficiency* $\epsilon(t)$.

The distribution of a given variable, κ , at particle level ("true distribution") is represented by a vector $\vec{\kappa}_{\text{true}}$ of components κ_j , where κ_j are the number of events at particle level in bin j and $j = 1, \dots, M$ with M equal to the number of bins of the "true distribution".

Hence, the expected number of events in bin j is:

$$\kappa_j = \kappa \cdot p_j = \kappa \cdot \int_{(\Delta t)_j} f(t) dt, \quad (6.7)$$

where κ is the total number of expected events and $f(t)$ is the probability density function of the true value, and Δt is the bin width.

The distribution of the same variable, ν , at detector level ("reco distribution") is represented by a vector $\vec{\nu}$ of components ν_i , where ν_i are the number of events at detector level after background subtraction in bin i and $i = 1, \dots, N$ with N equal to the number of bins of the "reco distribution". The aim of the folding/unfolding procedure is to relate $\vec{\kappa}$ with $\vec{\nu}$.

The number of events measured in bin i , ν_i , can be expressed as:

$$\nu_i = \kappa \int_{\text{bin } i} dr \int dt p(r|t)\epsilon(t)f(t). \quad (6.8)$$

Here $p(r|t)$ is the conditional probability density function that the true value t is reconstructed as r such that $\int p(r|t)dr = 1$. The Equation 6.8 expresses the fact that the true distribution is folded with the response function ($s(r|t) = p(r|t)\epsilon(t)$). The task of estimating $f(t)$ is called unfolding. Multiplying both numerator and denominator by κ_j , using the Equation 6.7 and considering all discrete bins j of the true value distribution, the expected number of entries in bin i becomes:

$$\begin{aligned} \nu_i &= \sum_j^M \frac{\int_i dr \int_j dt p(r|t)\epsilon(t)f(t)}{(\kappa_j/\kappa)} \kappa_j \\ &= \sum_j^M R_{ij} \kappa_j \end{aligned} \quad (6.9)$$

where the *response matrix* R is given by

$$\begin{aligned} R_{ij} &= \frac{\int_i dr \int_j dt p(r|t)\epsilon(t)f(t)}{\int_j dt f(t)} \\ &= \sum_i \sum_j \frac{p(r_i|t_j)\epsilon(t_j)f(t_j)}{\sum_l f(t_l)}. \end{aligned} \quad (6.10)$$

The response matrix element R_{ij} thus represents the conditional probability that an event with measured value r will be found in bin i , assuming that the true value t was in bin j . As it can be seen from the first line in Equation 6.10, the response matrix depends on the (unknown) pdf $f(t)$, which one attempts to infer. In practice, the response matrix will be determined using the best approximation of $f(t)$ that is available. The response matrix element R_{ij} is often determined using a MC simulation based on an understanding of the physical processes that take place in the detector. In practice, the model dependence may not be negligible, and the understanding of the detector itself is never perfect. Hence, both of them are treated as a possible sources of systematic uncertainty.

Taking into account that background events may also contribute, the number of reconstructed events in bin i is:

$$\nu_i = \sum_j R_{ij} \kappa_j + \beta_i, \quad (6.11)$$

where β_i is the number of expected background events in bin i . This relation can also be written in matrix form:

$$\nu = R\kappa + \beta. \quad (6.12)$$

The aim of the unfolding is to compute $\vec{\kappa} \equiv R^{-1}(\vec{\nu} - \vec{\beta})$ [98].

6.2.2 Overview of unfolding methods

The unfolding procedure starts with the choice of the observable. Next comes the choice of the binning which is performed on the base of the statistical uncertainty in each bin of the distribution and the resolution of the observable. As a third step, the *migration matrix* is built. If resolution effects are properly accounted for in the binning choice, the value of its element along the diagonal is close to one ("diagonal matrix").

It is also crucial to perform validation or closure tests of the results, because the unfolding method can rarely rely on exact solutions, and the approximate solutions are subject to potential instabilities.

A number of methods to correct for detector effects were developed. The following list presents few procedures:

- **Bin-by-bin correction.** This procedure is the simplest but it is not a truly unfolding method. It uses MC in order to extract factors with which data bins are corrected. It is assumed that, correlations between bins are the same in data and MC. The method yields usually smaller statistical and experimental uncertainties than any other unfolding method, but may depend on the MC description in a more significant way with respect to the other methods.
- **Bayesian Iterative by D'Agostini.** Iterative unfolding is a well establish concept, however, D'Agostini [99] interpreted it in the context of the Bayesian statistics.
- **The singular Value Decomposition (SVD) [100].** This method uses the singular value decomposition of the response matrix to overcome wildly oscillating solutions that emerge from the system of linear equations derived from Equation 6.9.
- **Iterative Dynamically Stabilized method (IDS) [101].** It is based on the idea that if the MC simulation provides a relatively good description of the data and of the detector effects, then one can use the transfer matrix to compute a matrix of unfolding probabilities. This method uses a regularization function

which allows to build an improved normalization procedure for MC spectra, unbiased by the presence of possible new structures in data.

Generally speaking, the choice of the method strongly depends on the analysis and the problem the analysis aims to solve. In this analysis, the method of choice is D'Agostini's Iterative Bayesian, for the following reasons:

- the mathematical treatment of the response matrix, being a collection of probabilities of measured data given the true data, avoids the "direct attack" of finding the inverse migration matrix,
- it is fast, allowing uncertainties to be computed in reasonable time,
- it is already implemented in the RooUnfold [102] framework,
- there is only one regularization parameter (the number of iterations), and the regularization process is straightforward,
- the dependence of the posterior (output of an iteration during unfolding) on the prior (input distribution) is negligible.

6.2.3 Bayesian Iterative

The iterative approach is derived from the Bayes'theorem [103], which can be expressed as

$$P(true|obs) = \frac{P(obs|true)f(true)}{\int f(true)P(obs|true)dtrue} = \frac{P(obs|true)f(true)}{g(obs)}, \quad (6.13)$$

where $P(obs|true)$ is the probability of observing the measured value of a variable, given a true value in the interval $[true, true + dtrue]$, $f(true)$ is the probability density function for the true observable and $g(obs)$, is defined as:

$$g(obs) = \int f(true)P(obs|true)dtrue. \quad (6.14)$$

By inverting Equation 6.13 and using the normalization property of $P(obs|true)$, the probability density function for the true observable is:

$$f(true) = \int g(obs)P(true|obs)dobs \quad (6.15)$$

Equation 6.13 uses the approximation that if an initial hypothesis is given by $f(true)$, one can use $P(obs|true)$, derived from simulation, to evaluate $P(true|obs)$ by defining $g(obs)$ as the convolution of $f(true)$ and $P(obs|true)$. This operation is called *folding*. The initial estimate of $f(true)$ can be derived from Equation 6.15 by convoluting $P(true|obs)$ with $g(obs)$ (the probability distribution observed in data). This step is called *unfolding*. Estimates of $f(true)$ can be reused as initial hypotheses for an updated estimate of $P(true|obs)$ in Equation 6.13, setting up an iterative

procedure. The starting point for the $(r + 1)^{th}$ step is obtained as:

$$f^{r+1}(true) = \int f^r(true) \frac{g^{(obs)}_{data}}{g^r(obs)} P(true|obs) dobs, \quad (6.16)$$

where the ratio $\frac{g^{(obs)}_{data}}{g^r(obs)}$ is considered as an *updating function* which includes the information from the data into the new estimate. The number of iterations depends on the step r , in which the modifications to the value of $f^{r+1}(true)$ introduced by additional steps are smaller than a given tolerance value.

D'Agostini's scheme

The method emphasizes the fact that the variables derived by the iteration are the expected contents of bin C_i , the probability that a certain fraction of the total events are found in a given bin and not the overall probability for the distribution. Each bin content of discretized distribution is considered an independent cause of an effect and the probability is written for the bins as:

$$P(C_i|E) = \propto P(E_j|C_i)P(C_i), \quad (6.17)$$

where C_i is the content of the bins of the unknown true distribution to be recovered and E_j are the contents of the bins of the observed distributions. The connection between the two distributions is given by the simulation-derived response matrix $P(E_j|C_i)$ which can be interpreted as the probability that a given cause C_i results in a given effect E_j . The losses due to the limitations in the observation require efficiency correction, again estimated by simulation.

The first estimate of the number of events $n(C_i)$ in the bin i of "causes" can be written as:

$$n(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{N_E} n(E_j)P(C_i|E_j) \quad (6.18)$$

and the additional iterative steps follow the general scheme described above. In general, the iterative scheme is sensitive to "long wavelength" errors in the i^{th} estimate of $g(obs)$; errors that are usually corrected in the initial iterations by incorporating all the useful information into the dataset. Further iterations will progressively include more "shorter wavelength" errors (higher frequency, more detailed information), which are more likely derived from statistical fluctuations in the $g(obs)$ data: the resulting corrections will tend to match $g^r(obs)$ to those fluctuations, rather than to the proper $g(obs)$ value that represents the best estimate of $f(true)$. The iterations are continued until the solution is considered stable. The criterion for stability is analysis-dependent.

6.2.4 Differential distributions

The differential cross sections are measured in the Search Region of the cut-based approach defined in Section 5.4 of the previous chapter. The binning for each observable is chosen depending on the statistics per bin and the resolution of the observable in each bin of the distribution. A main step of the procedures the construction of the *response* or *migration matrix* that contains the description of the detector effects. The *response matrix* is built as a two-dimensional histogram, filled with the *true* vs *reconstructed* values of the observable, and is constructed using MC simulated events.

Starting from the particle level objects in the signal MC (generator level), a selection is imposed to construct the analysis phase space. This selection is summarized in Table 6.1. Then, the analysis proceeds at reconstruction level, where the relevant quantities are extracted. Figure 6.1 illustrates in a schematic way the unfolding procedure, which goes through the following three steps. Initially, the estimated background is subtracted from the data, and acceptance factors are applied to correct for the number of events that fall out of the detector phase space. Then the unfolding matrix is applied with the iterative procedure described above. As a final step, efficiency factors are applied to account for the detector efficiency.

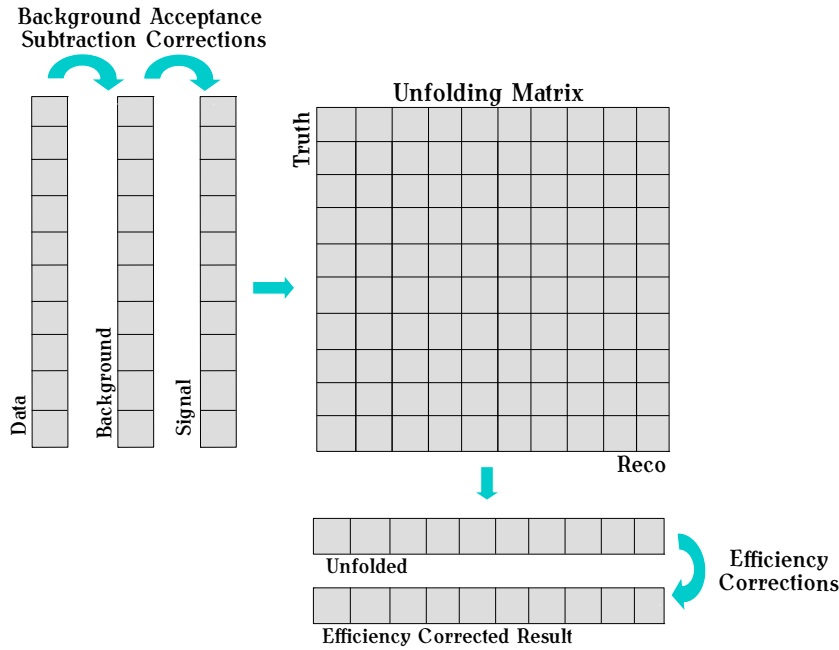


FIGURE 6.1: Schematic illustration of the unfolding procedure. Acceptance corrections are applied to the signal yield before the actual unfolding process. The result is corrected with efficiency factors.

In these measurements the signal events are both kind of $Z\gamma jj$. The *migration matrix* is determined using MC events from the sum of SHERPA v2.2.2 $Z\gamma jj$ -EW and $Z\gamma jj$ -QCD samples. The interference is neglected because it is a small effect.

Each sample is normalized to the luminosity of the data and corrected for the factors obtained in the fit to the centrality distribution.

Treatment of the background

For the background subtraction, the normalization and the shapes of the minor backgrounds ($t\bar{t}\gamma$, single top and WZ) are taken from MC simulations. The normalization and shape of the Z +jets sample is obtained from the ABCD method using the *anti-Tight* definition called LP5. Figure 6.2 shows the distributions of the four observables of the Z +jets background using the reference shape (LP5), the shape obtained from the *anti-Tight* definition called LP4 and the $Z\gamma jj$ -QCD distribution using the SHERPA v2.2.2 sample. The uncertainty on the Z +jets shape is computed as the difference of the shape between the nominal $Z\gamma jj$ -QCD SHERPA v2.2.2 sample and the shape obtained using the LP5 definition. The choice of LP5 instead of LP4 is mainly due to the fact that LP5 has higher statistics input data.

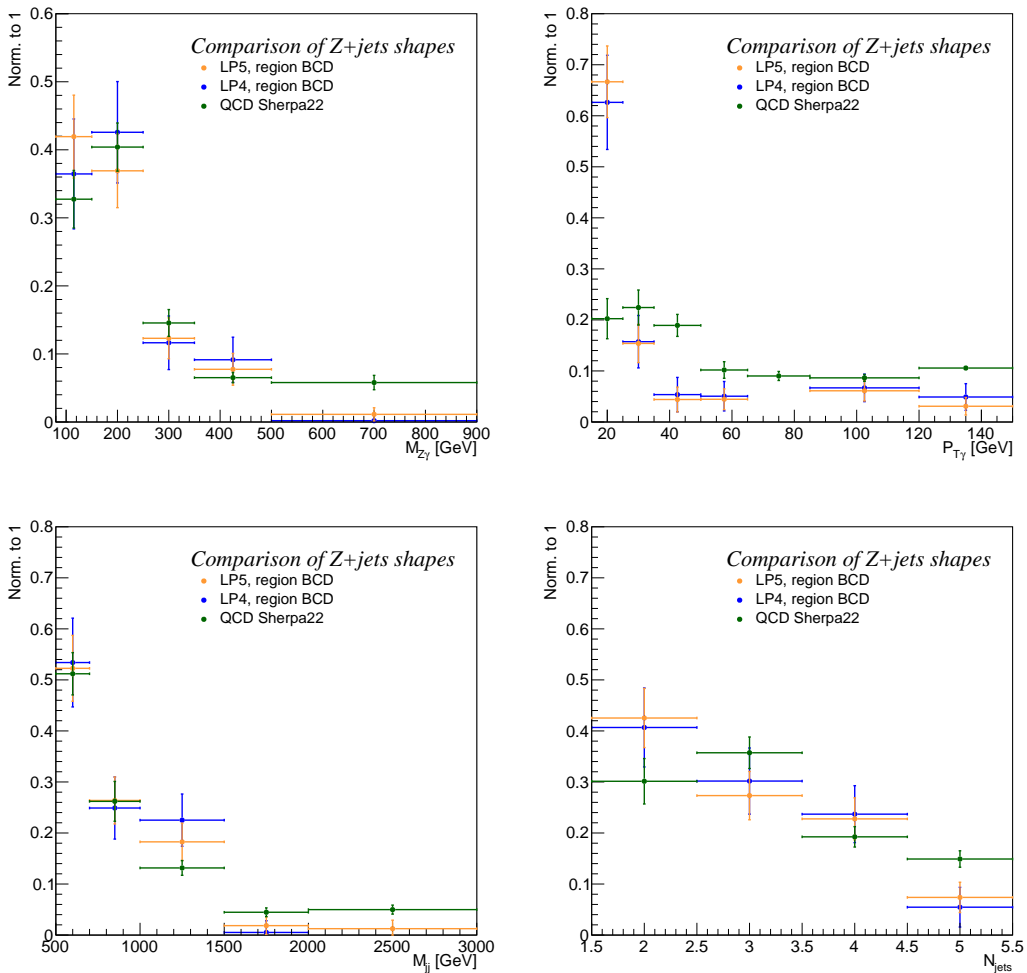


FIGURE 6.2: Normalized distributions of the Z +jets background estimated with the ABCD method using different definitions of the *anti-Tight* selection (LP4, LP5). The $Z\gamma jj$ -QCD distribution, using the nominal SHERPA v2.2.2 sample, is also displayed.

Systematic uncertainties

The systematic uncertainties can be split into three categories. The first category represents uncertainties due to the unfolding procedure itself, while the second category represents uncertainties on the inputs to the unfolding as for example a wrongly modeled energy spectrum or resolution. Finally the third category contains uncertainties related to the subtraction of the background.

Uncertainties due to the unfolding, are calculated with two different methods. In the first method, the unfolding procedure is performed by unfolding the data with a different *migration matrix* (SHERPA v2.1 $Z\gamma jj$ -QCD and MADGRAPH $Z\gamma jj$ -EW). The result is compared to the result of the unfolding procedure of the data using the nominal response matrix (SHERPA v2.2.2 $Z\gamma jj$ -QCD and $Z\gamma jj$ -EW). As cross check in the second method, a pseudo-data sample, using SHERPA v2.1 $Z\gamma jj$ -QCD and MADGRAPH $Z\gamma jj$ -EW, is unfolded using the same *migration matrix* (SHERPA v2.2.2 $Z\gamma jj$ -QCD and $Z\gamma jj$ -EW) as the one used in the analysis. The difference between the differential cross section using data and the one of using pseudo-data is considered as uncertainty in the unfolding method. The results of these two methods are very compatible. The chosen method, which results are presented in this thesis, is the first method since it is use real data. Table 6.3 summarizes the two methods. When combining the EW and QCD sample, the normalization of the uncertainty of μ_{QCD} and μ_{EW} gives a slightly different migration matrix but the effect on the differential cross section is negligible.

	<i>migration matrix</i>	Data to unfold
Method 1	SHERPA v2.2.2 $Z\gamma jj$ -QCD and EW	real data
	SHERPA v2.1 $Z\gamma jj$ -QCD and MADGRAPH $Z\gamma jj$ -EW	real data
Method 2	SHERPA v2.2.2 $Z\gamma jj$ -QCD and EW	real data
	SHERPA v2.2.2 $Z\gamma jj$ -QCD and EW	pseudo-data (SHERPA v2.1 $Z\gamma jj$ -QCD and MADGRAPH $Z\gamma jj$ -EW)

TABLE 6.3: The two methods used to calculate the uncertainty due to the unfolding procedure.

Uncertainties related to the objects of the analysis (electron, muons, photons and jets) derive from different sources and are calculated as the quadratic sum of all the systematic uncertainties in each category separately. For the electrons, the electromagnetic scale and resolution as well as the identification, isolation, reconstruction and trigger efficiencies are taken into account. For the muons, the muons scale, resolution, efficiency and trigger are considered. The jet modeling, calibration, flavor, efficiency and pile-up are taken into account. Finally, for the photons, the photon isolation and identification efficiency are considered. Another source of uncertainty is coming from the pile-up.

The background processes, namely the Z +jets, $t\bar{t}\gamma$, single top and WZ , are all scaled with an uncertainty of $\pm 20\%$, as described in Section 5.9. For the Z +jets

background, the uncertainty on the shape is also taken into account. The uncertainties from background processes are split it into two categories; irreducible and reducible, where the $t\bar{t}\gamma$ belongs to the irreducible background and the Z +jets, single top and WZ to the reducible background. Since the main source of uncertainty of the reducible background is coming from the Z +jets, in the following sections, the reducible background category has been split further in order to present clearly the effect of the Z +jets background.

Unfolding closure test

A technical test, which shows if the unfolding method works as expected is performed. The purpose of this test is to derive possible measures for the bias introduced by the unfolding itself. This technical test ("closure test") is performed using pseudo-data of signal MC samples. By using the same MC events as input to the unfolding procedure than those used to derive the response matrix, if the unfolding method behaves well, the output spectrum is exactly the same as the spectrum at particle level. Applying this test a perfect closure is obtained for all the four observables.

$M_{Z\gamma}$ differential distribution

The chosen binning for the $M_{Z\gamma}$ distribution and the number of events per bin in data are the following:

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
$M_{Z\gamma}$ (GeV)	80 – 150	150 – 250	250 – 350	350 – 500	500 – ∞
N_{data}	108	138	63	29	17

Figure 6.3 shows the migration matrix which relates the distribution at particle and reconstruction level. Each row is normalized to the number of generated events. It can be seen that the *migration matrix* is diagonal, namely the migration effect between bins is very low. In the same figure, the purity and the efficiency corrections for each bin are also presented. The purity is defined as the number of events generated and reconstructed in the same bin divided by the number of generated events in the phase space at particle level. The purity, unlike the migration matrix, includes also migrations outside the fiducial phase space at reconstruction level, but reconstructed inside of it. The purity remains mostly stable at a high value between 75%-80%. The efficiency, is of the order of 35%, largely the result of selection efficiency. For each of the three object categories the official ATLAS identification, efficiency is of the order of 85-95%. As a result, since the process consist of five objects, a low efficiency of the order of 30-35% is expected. This is observed also from the value of the C factor which is of the order of 35-40%, so that it is inevitable that the efficiency (which is computed as the ratio of the reconstructed and generated events over the generated events, $\epsilon = \frac{N_{\text{MC, reco \& gen.}}}{N_{\text{MC, gen.}}}$, while the C factor is the ratio of reconstructed over generated events, $C = \frac{N_{\text{MC, reco}}}{N_{\text{MC, gen.}}}$) is lower than the C factor value.

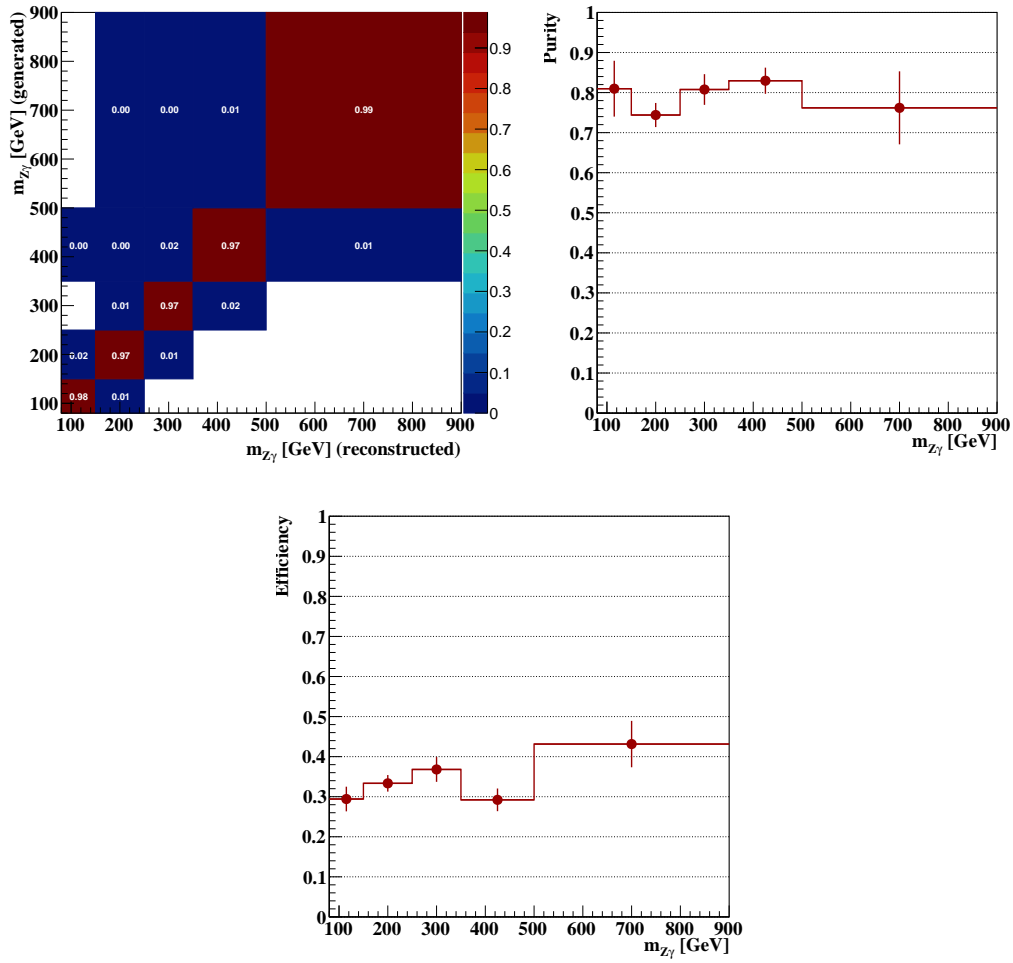


FIGURE 6.3: Distributions of the migration matrix (top left) and the purity (top right) as a function of the invariant mass of the $Z\gamma$ system and the efficiency correction (bottom) for the unfolding of differential cross section.

The number of iterations for the unfolding procedure has been optimized. The final choice is the use of two iterations, since after two iterations the result remain stable. The differential cross section as a function of the invariant mass of the $Z\gamma$ system is presented in Figure 6.4 for $Z\gamma jj$ events of the SR and for all decay channels. A fair agreement with the shape predicted by SHERPA v2.2.2 is observed. The measured cross section is also summarized in Table 6.4, together with the effects of the main systematic sources of uncertainties.

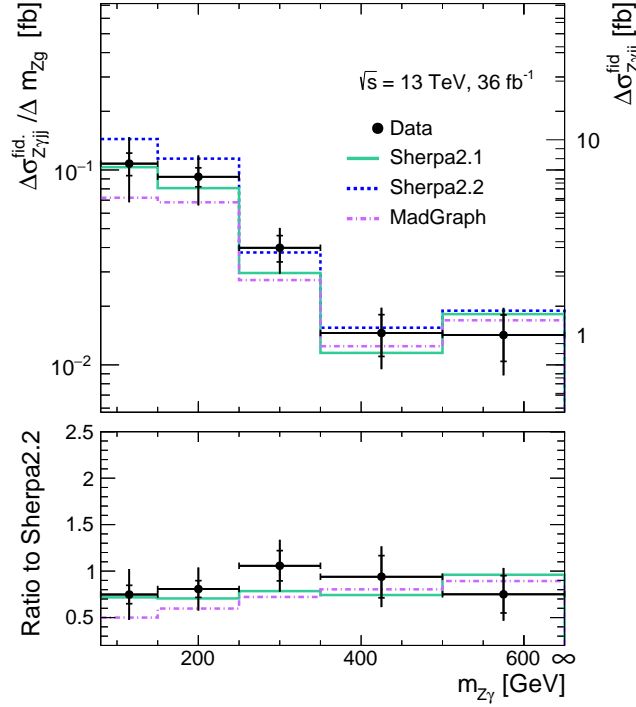


FIGURE 6.4: The measured $Z\gamma\gamma$ differential cross section in the fiducial phase space, defined in Section 6.1, as a function of $m_{Z\gamma}$. The last bin contains overflows. The measurement is compared to the sum of the $Z\gamma\gamma$ -QCD and $Z\gamma\gamma$ -EW predictions for SHERPA v2.2.2.

$m_{Z\gamma}$ [GeV]	80 – 150	150 – 250	250 – 350	350 – 500	≥ 500
$\Delta\sigma_{Z\gamma\gamma}^{\text{fid.}}$ [fb]	7.55	9.23	3.99	2.19	1.00
Relative Uncertainties [%]					
Statistics	13.2	11.1	15.4	24.2	26.7
All systematics	34.2	26.4	21.5	25.0	26.8
Luminosity	2.8	2.6	2.5	2.6	2.3
Total	36.7	28.7	26.4	34.8	37.9
Uncorrelated syst.	1.0	1.2	1.4	1.7	1.3
Unfolding	0.9	0.4	0.7	1.1	0.5
Electrons	1.0	0.9	1.2	1.6	3.6
Muons	1.9	1.9	2.0	2.8	3.2
Photons	1.9	1.2	1.1	1.3	1.4
Jets	11.4	7.3	4.5	8.9	3.8
Z +jets Back.	29.0	24.3	18.3	19.7	21.9
Other Red. Back.	0.2	0.1	0.1	0.1	0.1
Irred. Background	1.0	1.2	1.4	1.7	1.3
Pileup	7.7	3.1	3.9	1.5	0.5

TABLE 6.4: The measured $Z\gamma\gamma$ differential cross section in [fb] as a function of $m_{Z\gamma}$. The relative uncertainties are reported in percentage. The decomposition of the total systematic uncertainty into sources with between-channel correlation and uncorrelated ones is indicated in the bottom rows.

P_T^γ differential distribution

The chosen binning for the P_T^γ distribution and the number of data events per bin are the following:

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7
P_T^γ (GeV)	15 – 25	25 – 35	35 – 50	50 – 65	65 – 85	85 – 120	120 – ∞
N_{data}	86	72	54	34	32	30	47

Figure 6.5 shows the *migration matrix*, which is fairly diagonal and the migration effect is relatively low (at most of 13%). Figure 6.5 presents also the purity for each bin, which is between 70%-85% and the efficiency corrections for each bin, which is on average of 35%. For the unfolding of the P_T^γ distribution a use of three iterations has been chosen since after this number of iterations the result remain stable.

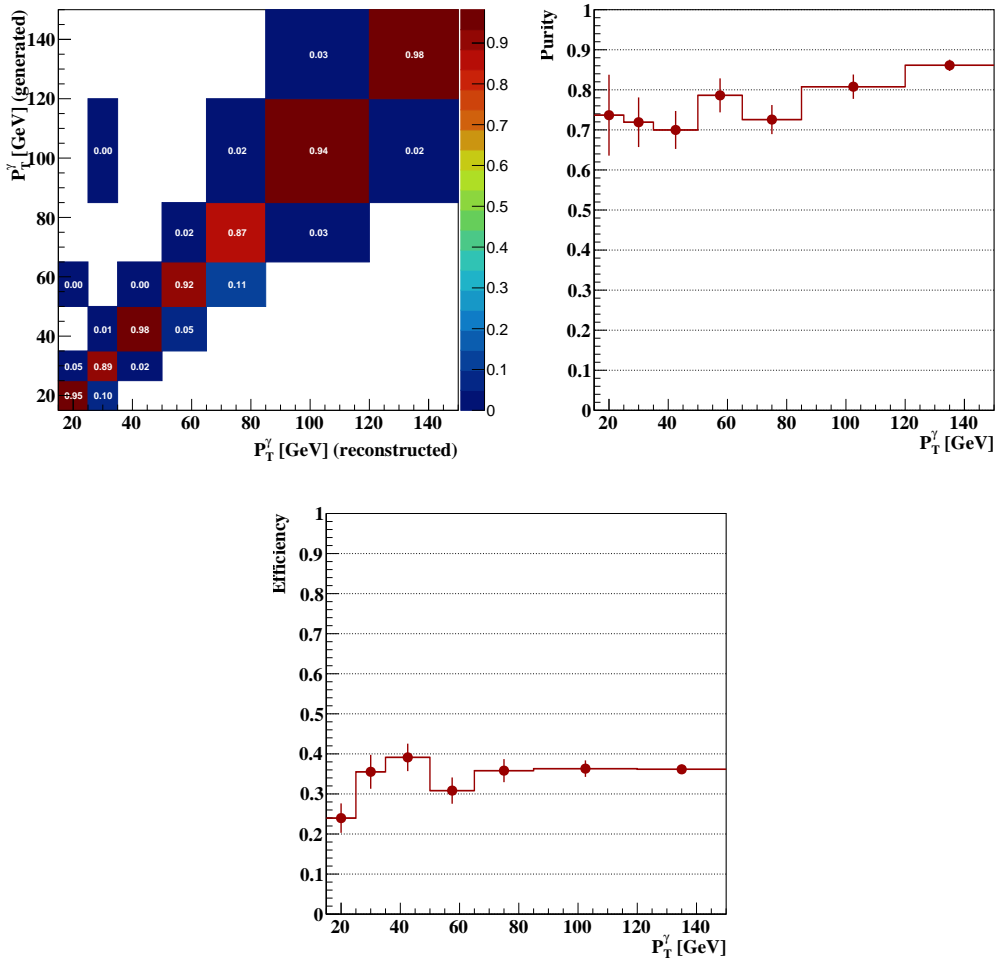


FIGURE 6.5: Distributions of the migration matrix (top left), the purity (top right) as a function of the photon transverse momentum (P_T^γ) and the efficiency correction (bottom) for the unfolding of differential cross section .

The differential cross section as a function of the photon transverse momentum is presented in Figure 6.6 for $Z\gamma jj$ events of the SR for all decay channels. An agreement with the shape predicted by SHERPA v2.2.2 is observed. The measured cross section is also summarized in Table 6.5 together with the effect of the main systematic uncertainty sources.

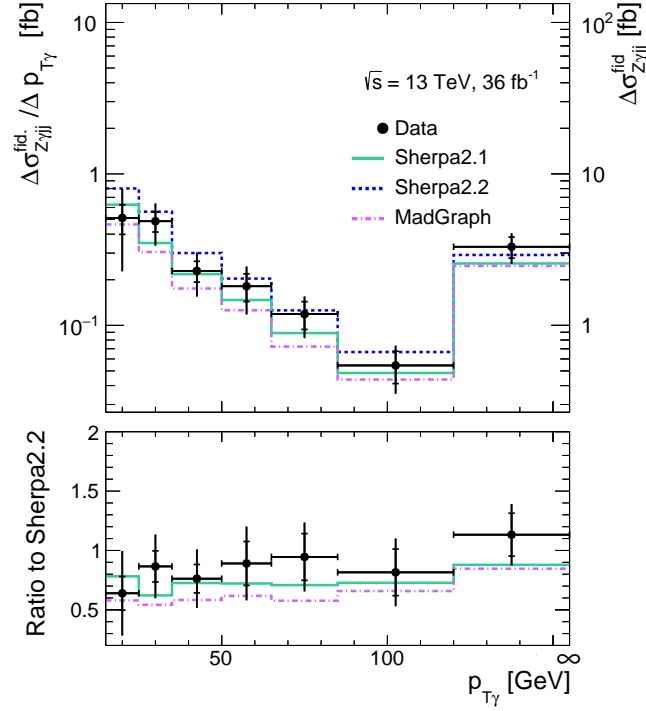


FIGURE 6.6: The measured $Z\gamma jj$ differential cross section in the fiducial phase space, defined in Section 6.1, as a function of P_T^γ . The last bin contains overflows. The measurement is compared to the sum of the $Z\gamma jj$ -QCD and $Z\gamma jj$ -EW predictions for SHERPA v2.2.2.

P_T^γ [GeV]	15 – 25	25 – 35	35 – 50	50 – 65	65 – 85	85 – 120	≥ 120
$\Delta\sigma_{Z\gamma jj}^{\text{fid.}}$ [fb]	5.12	4.87	3.43	2.72	2.37	1.90	3.30
Relative Uncertainties [%]							
Statistics	22.1	15.1	15.7	20.8	20.8	24.1	16.0
All systematics	51.0	27.0	28.2	27.9	22.5	25.3	16.3
Luminosity	4.1	2.4	2.3	2.4	2.2	2.5	2.3
Total	55.6	30.9	32.2	34.8	30.6	34.9	22.8
Uncorrelated syst.	3.8	2.5	3.8	2.8	3.5	3.0	3.6
Unfolding	0.9	0.4	0.7	0.8	0.9	0.1	0.2
Electrons	1.4	1.1	2.0	3.5	6.1	0.9	1.3
Muons	2.4	1.7	2.1	1.7	1.5	1.8	1.9
Photons	3.6	1.0	1.0	0.9	0.7	0.7	1.3
Jets	12.1	7.5	11.3	10.6	6.4	7.7	3.5
Z+jets Back.	43.4	22.2	23.0	20.8	16.7	22.5	14.6
Other Red. Back.	0.4	0.1	0.2	0.1	0.1	0.2	0.1
Irred. Background	3.8	2.5	3.8	2.8	3.5	3.0	3.6
Pileup	11.1	2.4	3.8	8.3	6.7	1.2	1.5

TABLE 6.5: The measured $Z\gamma jj$ differential cross section in [fb] as a function of P_T^γ . The relative uncertainties are reported as percentages. The decomposition of the total systematic uncertainty into sources with between-channel correlation and uncorrelated ones is indicated in the bottom rows.

M_{jj} differential distribution

The chosen binning for the M_{jj} distribution and the number of data events in each bin are the following:

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
M_{jj}	500 – 700	700 – 1000	1000 – 1500	1500 – 2000	2000 – ∞
N_{data}	149	124	55	17	10

From Figure 6.7, it can be seen that the migration effect is relatively low (maximum of 18%). Figure 6.7, also shows the purity which is roughly 75%, and the efficiency corrections which average at around 35%.

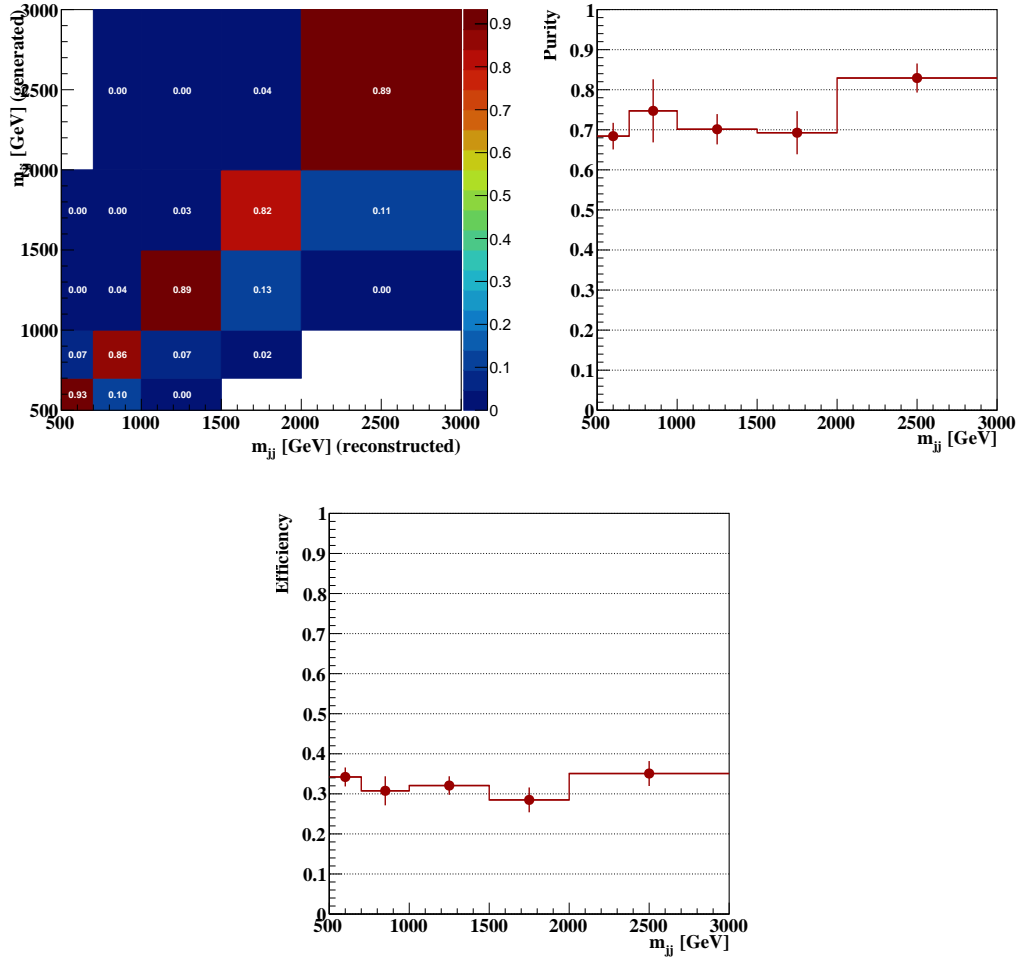


FIGURE 6.7: Distributions of the migration matrix (top left), the purity (top right) as a function of the invariant mass of the dijet system and the efficiency correction (bottom) for the unfolding of differential cross section.

For the unfolding of the M_{jj} distribution three iterations have been chosen, for the same reason as for the other variables. The differential cross section as a function of the invariant mass of the dijet system is presented in Figure 6.8 for $Z\gamma jj$ events of the SR. An agreement with the shape predicted by SHERPA v2.2.2 is observed. The measured cross section is also summarized in Table 6.6, together with the effect of the main systematic uncertainty sources.

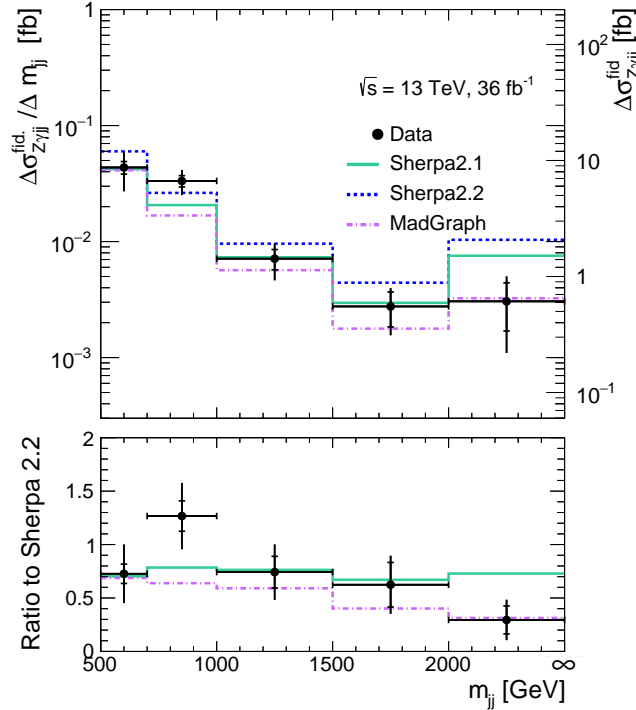


FIGURE 6.8: The measured $Z\gamma jj$ differential cross section in the fiducial phase space, defined in Section 6.1, as a function of m_{jj} . The last bin contains overflow. The measurement is compared to the sum of the $Z\gamma jj$ -QCD and $Z\gamma jj$ -EW predictions for SHERPA v2.2.2.

m_{jj} [GeV]	500 – 700	700 – 1000	1000 – 1500	1500 – 2000	≥ 2000
$\Delta\sigma_{Z\gamma jj}^{\text{fid.}}$ [fb]	8.71	9.99	3.56	1.38	0.61
Relative Uncertainties [%]					
Statistics	12.4	11.2	19.9	33.4	44.5
All systematics	35.6	21.7	28.6	27.9	46.1
Luminosity	2.8	2.4	2.8	2.4	2.4
Total	37.7	24.5	34.9	43.5	64.1
Uncorrelated syst.	4.4	1.9	3.0	5.7	5.4
Unfolding	0.5	0.9	0.9	0.3	1.0
Electrons	0.9	0.7	0.9	0.8	0.7
Muons	1.7	2.1	1.7	3.3	2.4
Photons	1.7	0.9	1.6	1.4	1.9
Jets	9.1	8.4	12.7	15.5	18.4
Z+jets Back.	32.6	16.0	22.8	18.2	37.7
Other Red. Back.	0.2	0.1	0.2	0.1	0.1
Irred. Background	4.4	1.9	3.0	5.7	5.4
Pileup	5.4	5.2	5.9	0.9	9.4

TABLE 6.6: The measured $Z\gamma jj$ differential cross section in [fb] as a function of m_{jj} . The relative uncertainties are reported in percentage. The decomposition of the total systematic uncertainty into sources with between-channel correlation and uncorrelated ones is indicated in the bottom rows.

N_{jets} differential distribution

The chosen binning for the N_{jets} distribution and the number of data events are the following:

	Bin 1	Bin 2	Bin 3	Bin 4
N_{jets} (GeV)	1.5 – 2.5	2.5 – 3.5	3.5 – 4.5	4.5 – ∞
N_{data}	155	116	59	25

Figure 6.9 shows the migration matrix. It can be seen that the migration effect is very low for the $N_{jets} = 2$ (of the order of 7%) and higher for $N_{jets} > 2$ (up to 35%). The migration effect for $N_{jets} > 2$ and for $N_{jets}^{gen.} < N_{jets}^{reco}$ (an effect of 30% at maximum) may come from pile-up jets. At the same time the migration effect for $N_{jets} > 2$ and for $N_{jets}^{gen.} > N_{jets}^{reco}$ (an effect of 10% at maximum) is found to be due to detector resolution, which means that the reconstruction of jet energies will be of a certain precision. Thus, there are generated level jets that pass the cut $p_{T,jets}^{gen} > 30$ GeV, even if, at the same time, not all the reconstructed jets pass the cut $p_{T,jets}^{reco} > 30$ GeV.

Optimization on the number of iterations leads to three iterations, since after this iteration the result remains stable. In the same figure, the purity and the efficiency corrections in each bin are also presented. It can be seen that the purity is of 60%, lower than the value for other observables, but to be expected for the jet multiplicity. The efficiency is of the order of 32%, almost constant in all the bins.

The differential cross section as a function of the jet multiplicity is presented in Figure 6.10 for $Z\gamma jj$ events of the SR. An agreement with the shape predicted by SHERPA v2.2.2 is observed. The measured cross section is also summarized in Table 6.7 together with the effect of the main systematic uncertainty sources.

As can be seen from Tables 6.4, 6.5, 6.6 and 6.7, the main systematic uncertainties come from the Z +jets background and the jets. Another important source of uncertainty is pile-up, while a smaller contribution comes from photons and leptons. The main source for the jets derives from jet calibration, jet flavor response, jet pile-up and jet modeling. For photons, the main source is the photon identification efficiency. For electrons, the uncertainty mostly comes from the electromagnetic scale, while for muons, it can be traced back to the muon efficiency and muon trigger systematics.

From the differential cross section distributions, Figures 6.4, 6.6, 6.8 and 6.10, a fair agreement between the shapes of the measured distributions and the different available MC predictions is found. The results of the differential cross section measurement will be used in the future for extracting limits in the EFT coefficients.

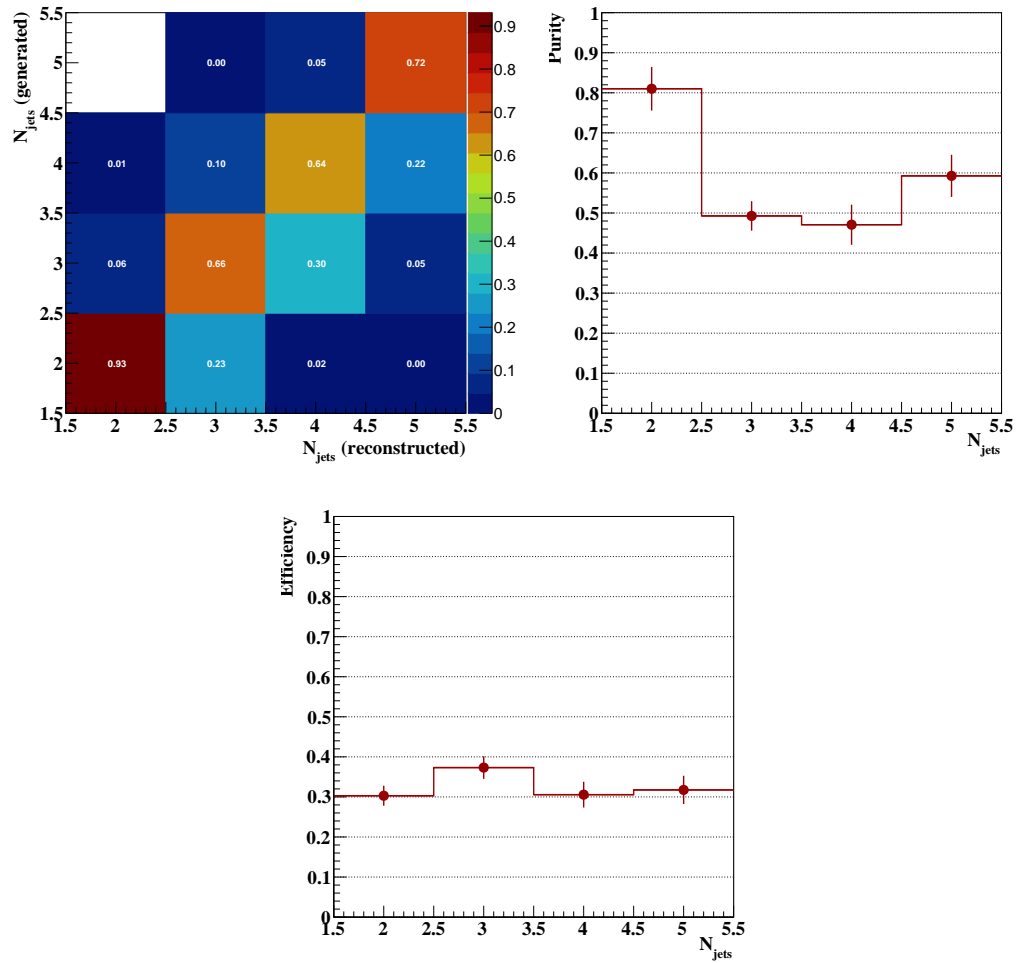


FIGURE 6.9: Distributions of the migration matrix (top left), the purity (top right) as a function of the jet multiplicity and the efficiency correction (bottom) for the unfolding of differential cross section.

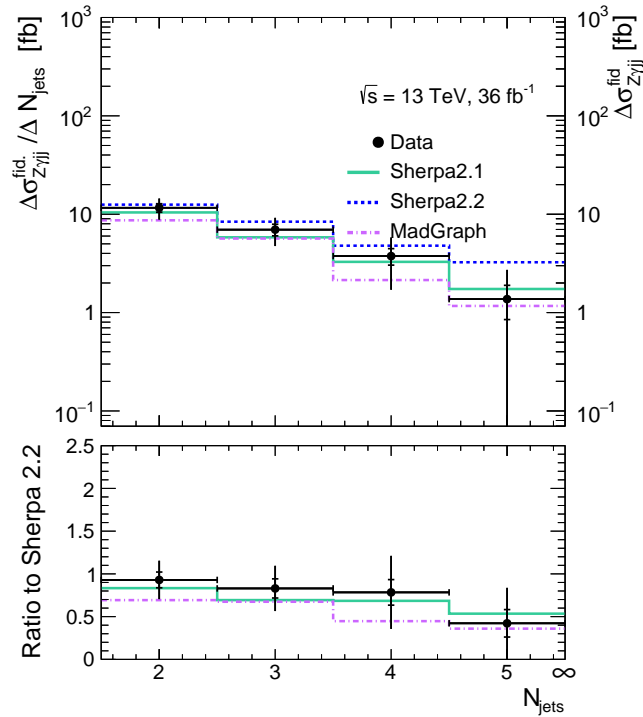


FIGURE 6.10: The measured $Z\gamma jj$ differential cross section in the fiducial phase space, defined in Section 6.1, as a function of N_{jets} . The last bin contains overflow. The measurement is compared to the sum of the $Z\gamma jj$ -QCD and $Z\gamma jj$ -EW predictions for SHERPA v2.2.2.

N_{jets}	1 – 2	2 – 3	3 – 4	≥ 4
$\Delta\sigma_{Z\gamma jj}^{fid.}$ [fb]	11.63	6.97	3.75	1.37
Relative Uncertainties [%]				
Statistics	9.9	13.5	19.1	38.0
All systematics	22.0	28.9	51.1	90.7
Luminosity	2.6	2.6	2.9	2.8
Total	24.1	31.9	54.5	98.3
Uncorrelated syst.	2.0	3.9	4.8	8.6
Unfolding	2.1	1.8	1.7	1.9
Electrons	0.8	0.8	0.9	0.8
Muons	1.9	2.0	1.7	2.9
Photons	1.2	1.6	1.8	1.3
Jets	9.2	7.7	35.0	66.7
Z +jets Back.	16.7	26.1	34.3	57.2
Other Red. Back.	0.1	0.2	0.1	0.3
Irred. Background	2.0	3.9	4.8	8.6
Pileup	5.0	3.3	5.4	9.6

TABLE 6.7: The measured $Z\gamma jj$ differential cross section in [fb] as a function of N_{jets} . The relative uncertainties are reported as percentages. The decomposition of the total systematic uncertainty into sources with between-channel correlation and uncorrelated ones is indicated in the bottom rows.

Chapter 7

Conclusion

Within the scope of this thesis, the study of the $Z\gamma jj$ -EW production, with the Z boson decaying leptonically, and the measurement of its cross section was presented. The differential cross section was measured for the first time for this channel. The data were collected with the ATLAS experiment at the LHC in 2015 and 2016 at a center-of-mass energy of $\sqrt{s} = 13$ TeV and an integrated luminosity of 36.1 fb^{-1} . The main systematic uncertainties in the cross section measurement derive from jets, lepton and photon identification and resolution. In this thesis, detailed studies on the electron identification efficiency were presented. The agreement between data and MC simulation was studied and scale factors were extracted to correct for mis-modeling. Moreover an addition data-driven method for electron identification efficiencies measurement was implemented which addresses electrons with low transverse energy and coming for the decay of the Z boson. These were combined with a complementary electron identification efficiency measurement and provided to the ATLAS collaboration. The main uncertainty contribution results from the background in the measurement. This precision on the electron identification scale factors translates to an uncertainty contribution of less than 1% in the $Z\gamma jj$ -EW cross section measurement. The analysis of $Z\gamma jj$ -EW process result in a signal strength:

$$\mu_{EW} = 1.00^{+0.19}_{-0.18} (\text{stat})^{+0.08}_{-0.10} (\text{MCstat})^{+0.09}_{-0.08} (\text{syst})^{+0.13}_{-0.10} (\text{theo})$$

The integrated cross section measured in the same fiducial region and found to be:

$$\sigma_{\text{obs.}}^{\text{fid.}} Z\gamma jj\text{-EW} = 7.75^{+1.47}_{-1.39} (\text{stat.})^{+0.93}_{-0.99} (\text{exp.syst})^{+1.01}_{-0.78} (\text{mod.syst}) \pm 0.15(\text{lumi}) \text{ fb.} \quad (7.1)$$

The total $Z\gamma jj$ cross section is:

$$\sigma_{\text{obs.}}^{\text{fid.}} Z\gamma jj = 71.4 \pm 2.4 (\text{stat.})^{+8.9}_{-6.5} (\text{exp.syst})^{+21.1}_{-17.0} (\text{mod.syst}) \pm 0.15(\text{lumi}) \text{ fb.} \quad (7.2)$$

Detector effects, such as finite resolution, limited acceptance and imperfect efficiency, distort the distribution of the observable measured elements. In this thesis, a dedicated study of these distortions was presented and the results of the differential measurement of the cross section can now be used to directly compare with the corresponding theoretical predictions as well as with the results of other experiments.

Differential cross section measurements were performed for observables selected

for their good sensitivity to potential aQGCs or for their ability to separate processes $Z\gamma jj$ -QCD and $Z\gamma jj$ -EW and to constrain the current MC description of $Z\gamma jj$ production. These distributions are the transverse momentum of the photon (P_T^γ), the invariant mass of the diboson system ($M_{Z\gamma}$), the invariant mass of the dijet system (M_{jj}) and the jet multiplicity (N_{jets}). So far no evidence of aQGCs were observed and the results are in agreement with the SM expectations. Nevertheless the results of the differential cross section measurement can be used to extract limits in the EFT coefficients. The full Run II data are expected to give an observation of the $Z\gamma jj$ -EW. In order to extend the successful physics program at the LHC to high precision studies and searches for new physics beyond the SM, much more data will need to be collected. An upgrade program for LHC, as well as the experiments, will meet this challenge. An upgrade is planned to increase the instantaneous luminosity delivered by the LHC in what is called HL-LHC aiming to deliver a total of about 3000 fb^{-1} of data per experiment. This upgrade will create a very challenging environment with highly increasing pile-up. Thus the upgrade of the ATLAS detector will consist of really precise Inner Tracker which will extend the ATLAS tracking capabilities to a pseudorapidity up to 4.0. Moreover changes to Muon Spectrometer and the timing detector will affect the forward region. These changes it is assumed that will make possible to identify electrons and muons up to $|\eta^{\text{lep}}| = 4$ and jets up to $|\eta^{\text{jet}}| = 3.8$. Since the topology of the electroweak production of di-boson events consist of central bosons associated with two high energetic forward jets, all VBS analyses benefit fully from this upgrade. Already available studies of the prospective of vector boson scattering at the HL-LHC show incredible results. A study for the WZ [104] fully leptonic final state presents a gain of 60% in signal compared to the Run II results. Another study of electroweak production $ZZjj$ [105] predicts a significance of 7σ . More and more studies expected in the VBS sector as it is an important physics goal for the LHC physics program. Definitely the presented work is only a small keystone to the great physics observations that follows.

Chapter 8

Résumé

8.1 Introduction

Le sujet de physique de ma thèse est la mesure de la section efficace de production de diboson $Z\gamma$ production en association avec une paire de dijet de haute masse invariante. Les données de collision protons-protons de l'expérience ATLAS au Grand collisionneur de hadrons du CERN à une énergie dans le centre de masse de $\sqrt{s} = 13$ TeV, recueillies en 2015 et 2016 et correspondant à une luminosité intégrée de 36.1 fb^{-1} sont analysées. La production électrofaible d'événements $Z\gamma jj$ donne un accès direct à la nature du mécanisme de brisure spontanée de la symétrie électrofaible, en sondant les couplages quartiques de boson $WWZ\gamma$, $ZZZ\gamma$, $ZZ\gamma\gamma$ et $Z\gamma\gamma\gamma$. Les trois derniers sont interdits à l'ordre le plus bas dans le Modèle Standard. Tout écart par rapport aux prévisions du Modèle Standard pourrait être un indice d'une nouvelle physique.

Jusqu'à présent, la production électrofaible de $Z\gamma jj$, qui consiste en des processus avec constante couplage électrofaible du quatrième ordre et qui inclut des processus de diffusion de vecteur-boson, n'a jamais été observée. Le même état final $Z\gamma jj$ peut être produit par un processus médié par l'interaction forte, avec une constante couplage électrofaible du second ordre et un couplage fort du second ordre, résultant en une section efficace de deux ordres de grandeur plus grande que la section électrofaible. La production de $Z\gamma jj$ via un mécanisme électrofaible et fort interfère puisque les états initiaux et finaux sont les mêmes. La compréhension de cet effet est un aspect important de l'analyse.

Dans ma thèse, j'ai étudié cet effet et son impact sur l'observation du processus et sur la mesure de la section efficace. Pour cette étude, après avoir effectué le calcul des trois contributions (électrofaible, forte et interférence) à la section efficace totale à l'aide de simulations de Monte Carlo, j'ai effectué une optimisation des coupures de sélection, afin de réduire l'impact de l'interférence, en définissant également une procédure pour la prise en compte de cette interférence dans l'analyse des données. Le point focal expérimental de mon travail est la mesure de la section efficace différentielle de l'état final $Z\gamma jj$ en fonction de la masse invariante du système $Z\gamma$, du momentum transversal du photon, de la multiplicité de jets et de la masse invariante du système dijet. Les distributions sensibles à la nouvelle physique sont corrigées par les effets détecteur et comparées aux attentes de la théorie.

Tous résultat de physique repose sur la reconstruction et la sélection d'états finaux composés de jets, d'électrons, de muons et de photons. Les électrons et leur identification jouent donc un rôle important dans l'analyse $Z\gamma jj$. Ce rapport de thèse documente également le travail que j'ai fait sur l'efficacité de l'identification des électrons, y compris d'une méthode supplémentaire axée sur les données. Les résultats que j'ai obtenus, sont utilisés dans toutes les analyses ATLAS dont la signature contient des électrons.

8.2 Eléments du Modèle Standard de la physique des particules

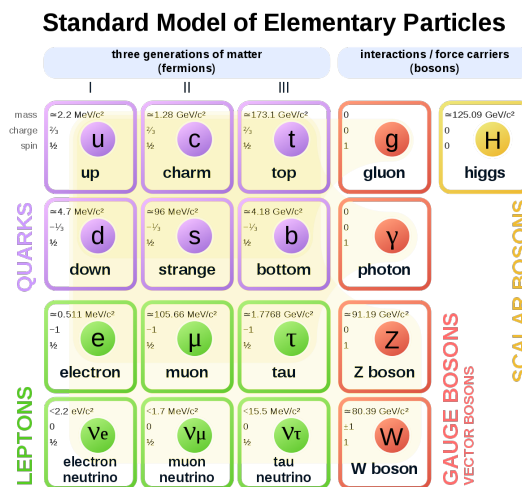


FIGURE 8.1: Le modèle standard des particules élémentaires. Les caractéristiques des fermions et des bosons sont illustrées.

Le Modèle Standard (MS) de la physique des particules, fournit une image unifiée de la nature et de la dynamique des particules et de leurs interactions. Selon le MS, il y a deux types de particules élémentaires : les fermions et les bosons. La matière se compose de fermions ponctuels avec un spin demi-entier, qui interagissent en échangeant des bosons de jauge ponctuels de spin entier. Il y a deux sous-groupes de fermions: les quarks et les leptons. Les leptons peuvent être séparés en particules chargées électriquement (e, μ, τ) et particules chargées neutres, les neutrinos (ν_e, ν_μ, ν_τ). Les quarks, en revanche, sont des particules portant des charges électriques et de couleur (u, d, c, s, t, b). Le MS est basé sur le groupe de symétrie $SU_C(3) \times SU_L(2) \times U_Y(1)$. La théorie de Chromodynamique quantique décrit les interactions entre les quarks qui est médiée par huit gluons. Elle est formulée dans le groupe de symétrie $SU_C(3)$. L'interaction électrofaible est décrite par le groupe de symétrie $SU_L(2) \times U_Y(1)$. Quatre bosons sont associés à cette interaction, trois bosons $W_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$ associé au groupe $SU_L(2)$ et le boson B_μ associé au groupe $U_Y(1)$. A basse énergie, la

symétrie est brisée par un champs scalaire qui couple les champs de gauge. Le potentiel introduit par le champs scalaire a une symétrie continue et les états en énergie les plus bas ne partagent pas la symétrie du Lagrangien. La nature choisit un état fondamentale, donc, la symétrie est brisée. Cette brisure de symétrie permet d'introduire des termes de masse dans le lagrangien sans briser l'invariant locale gauge.

8.3 Diffusion électrofaible des bosons de jauge dans le modèle standard

La structure $SU(2)_L \times U(1)_Y$ du secteur électrofaible du MS prédit des auto-interactions entre les bosons électrofaibles. Le processus VBS est inclus dans le processus $VV \rightarrow VV$ avec $V = W^\pm, Z, \gamma$. Une étude sur les processus VBS peut éclairer le secteur gauge de la théorie électrofaible. Les couplages des jauge quartique impliqués dans le VBS est une grande partie de ce qui en fait du processus $VV \rightarrow VV$ un bon candidat d'étude. Le test de la MS va de pair avec les recherches d'une nouvelle physique. Pour quantifier les effets de la nouvelle physique, on peut choisir une approche indépendante du modèle qui ne fait que paramétrer l'écart par rapport à la MS. Ces écarts sont alors interprétés comme des effets de nouvelle physique hors de la portée cinématique directe de l'expérience, mais dont les effets sont observables à des énergies plus faibles. Principalement, ces écarts sont exprimés en termes de couplages triple et quartique anomalies.

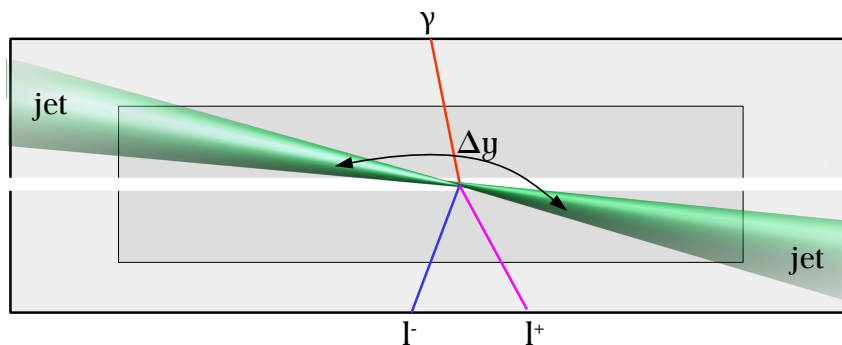


FIGURE 8.2: Topologie générale d'un événement VBS, dans ce graphique le processus VBS $Z\gamma jj$ illustre le boson Z qui se décompose en deux leptons.

La phénoménologie de ce processus de diffusion est relativement unique. Avec la caractéristique de ses deux jets à forte impulsion transversale, appelés "jets de marquage", cette phénoménologie facilite l'analyse. La topologie du procédé électrofaible VBS est ce qui peut le distinguer de l'interaction forte similaire. La topologie électrofaible du VBS est montrée dans la Figure 8.2. Les jets de marquage présentent une grande séparation spatiale en rapidité et une masse invariante élevée. Ces propriétés sont dues au fait que les deux jets proviennent des quarks entrants qui émettent les bosons vecteurs de diffusion portant ainsi une fraction de l'élan des protons entrants. De plus, on s'attend à ce qu'il n'y ait pas ou peu d'activité de jets entre les deux

jets d'étiquetage en raison d'un manque d'écoulement de flux de couleur entre les quarks entrants. Les bosons vecteurs provenant de la diffusion sont généralement situés (comme le montre la Figure 8.2) au centre des jets de marquage. Pour cela, la définition d'une nouvelle variable, appelée centralité, est introduite:

$$\zeta(Z\gamma) = \left| \frac{y_{Z\gamma} - (y_{j1} + y_{j2})/2}{\Delta Y_{jj}} \right| \quad (8.1)$$

La centralité prend des valeurs positives tant que toutes les particules considérées se trouvent entre les jets de marquage et les jets de plus en plus positifs indiquent une grande séparation des jets de marquage entre eux et une grande séparation des fermions/bosons et des deux jets. Les propriétés du système diboson permettent d'obtenir des variables supplémentaires pouvant être analysées. De nouvelles résonances peuvent être produites comme résonance intermédiaire lorsque deux bosons électrofaibles fusionnent. Les propriétés du système diboson observé sont alors influencées par les propriétés de la résonance intermédiaire. En cas de résonances à grande masse, on s'attendrait à une masse invariante élevée du système diboson, une grande séparation en ϕ des bosons sortants, et une momenta transversale élevée.

8.4 Le détecteur ATLAS

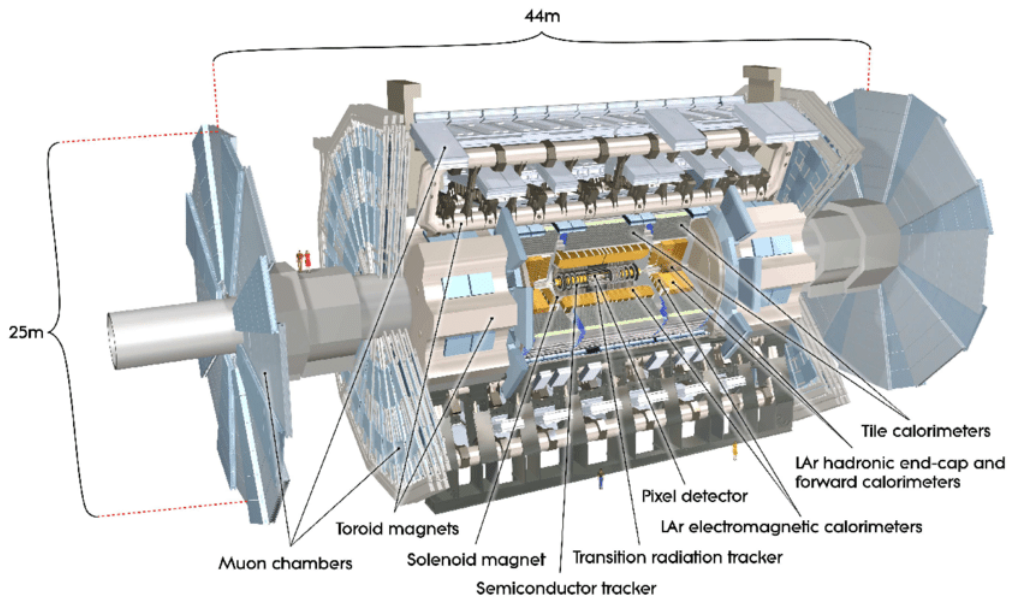


FIGURE 8.3: Illustration graphique du détecteur ATLAS.

Le détecteur ATLAS (Figure 8.3) est l'un des deux détecteurs polyvalents du LHC, avec une géométrie cylindrique symétrique vers l'avant et l'arrière. Il a été conçu pour étudier un large éventail de processus physiques, de la recherche du boson de Higgs aux dimensions supplémentaires et à la matière noire. Les expériences ATLAS et CMS partagent les mêmes objectifs scientifiques, mais utilisent des solutions techniques

différentes et une conception différente des systèmes magnétiques. ATLAS est le plus grand détecteur jamais construit, mesurant 44m de long, 25m de haut et 25m de large, pour un poids total de 7000 tonnes.

Le détecteur ATLAS se compose de quatre composants principaux, disposés en couches autour du point de collision et enregistrant le trajet, l'impulsion et l'énergie des particules, ce qui permet de les identifier individuellement.

En raison de la section efficace élevée et de la luminosité de la machine, les interactions dans le détecteur ATLAS créent une quantité énorme d'événements. Afin de digérer ces informations, le système de déclenchement ATLAS sélectionne les événements susceptibles de présenter un intérêt et qui doivent être enregistrés. Des systèmes complexes d'acquisition de données et d'informatique sont utilisés pour analyser et sélectionner les événements en ligne.

Le but du détecteur ATLAS est d'enregistrer autant d'informations que possible sur les particules de l'état final qui portent des informations sur l'interaction qui les a générées. Ceci est accompli grâce à différents sous-systèmes de détection qui identifient les particules et mesurent leur impulsion et leur énergie. Les quatre principaux composants du sous-système de détection sont : le détecteur interne, le calorimètre, le spectromètre muon et les systèmes magnétiques. La raison pour laquelle les détecteurs sont divisés en plusieurs composants est que chaque composant mesure un ensemble spécifique de propriétés des particules. Le "**Inner Detector**", immergé dans un champ magnétique solénoïdal, est principalement utilisé pour reconstruire les traces des particules chargées, pour mesurer leur impulsions et pour reconstruire les vertex primaires et secondaires.

Les calorimètres (électromagnétiques et hadroniques), sont utilisés pour identifier et mesurer l'énergie des électrons, photons et hadrons. Ils sont conçus pour arrêter la plupart des types de particules sauf les muons et les neutrinos.

Le "**Muon Spectrometer**", est utilisé pour détecter et mesurer les propriétés des muons. Comme les muons interagissent peu dans le détecteur et ont une longue durée de vie, ils sont identifiés et mesurés dans le système de détection le plus à l'extérieur. Les sous-systèmes de détection, mentionnés précédemment, sont immergés dans un champ magnétique généré par un **system d'aimants** qui courbe la trajectoire des particules chargées. Avant de décrire plus en détail les sous-systèmes ATLAS, la sous-section suivante illustre le système de coordonnées utilisé par ATLAS. Le système de coordonnées ATLAS est droitier et son origine se situe au point d'interaction. L'axe z est aligné avec la direction du faisceau, l'angle d'azimut ϕ donne l'angle dans le plan transversal, le plan décrit par les coordonnées x et y . L'angle polaire θ est l'angle avec l'axe z , cet angle est souvent converti en une quantité appelée pseudorapidity qui est calculée avec la formule $\eta = -\ln[\tan(\theta/2)]$.

8.5 Mesure de l'efficacité d'identification des électrons par la méthode Tag-and-Probe.

Dans les mesures du MS et les recherches de physique au-delà du MS, les spectres d'électrons déterminés expérimentalement doivent être corrigés pour tenir compte des efficacités de sélection liées au déclencheur, ainsi que la reconstruction, l'identification et de l'isolement des particules. Les électrons produisent des traces dans le détecteur de trace et des dépôts d'énergie dans le calorimètre électromagnétique. La sélection en ligne des électrons et la reconstruction des électrons visent à combiner la trace avec le dépôt d'énergie pour un rendement élevé. Afin de différencier les électrons des autres particules, plusieurs ensembles de critères d'identification sont développés. Ces critères reposent sur la qualité des traces, les formes des gerbes électromagnétiques et l'appariement trace-à-cluster. L'identification des électrons a pour but d'assurer une haute efficacité et une réjection élevée par rapport aux autres particules (background).

Pour les électrons, ces efficacités sont estimées directement à partir des données, à l'aide de la méthode Tag-and-Probe. La méthode utilise des événements dont les résonances bien connues se désintègrent en paires électron-positon(-photon), à savoir $Z \rightarrow ee(\gamma)$ et $J/\psi \rightarrow ee$. Les événements sont sélectionnés en fonction de la masse invariante électron-positon(-photon) ou de l'isolement de l'électron sonde. Ensuite, la méthode sélectionne des échantillons non biaisés d'électrons (probes) en plaçant des exigences de sélection strictes sur l'objet compagnon (tags) produit dans la désintégration. L'efficacité d'une sélection d'identification ou d'isolement donnée est déterminée en appliquant cette sélection à l'échantillon avec les électrons "probes" après avoir soustrait la contamination de fond résiduelle. Pour éviter des résultats biaisés, chaque combinaison de paires tag-and-probe dans un événement est prise en compte, de sorte qu'un électron peut être le tag dans une paire et le probe dans une autre. Des procédures similaires sont utilisées pour mesurer l'efficacité de reconstruction, l'identification, l'isolement et déclencher des gains d'efficacité.

La méthode tag-and-probe décrite ici couvre une gamme d'impulsion transverse d'électrons, E_T , de 7 à 200 GeV et $|\eta| < 2.47$. La fourchette inférieure E_T est couverte par le $J/\psi \rightarrow ee$ (de 7 à 20 GeV) et le $Z \rightarrow ee\gamma$ (de 10 à 30 GeV, en se concentrant dans l'intervalle de 10-15 GeV), tandis que les événements $Z \rightarrow ee$ sont utilisés aux énergies transversales au-dessus de 15 GeV. Pour identifier des électrons, les analyses appliquent des critères d'identification avec des niveaux de sévérité différents: *Loose*, *Medium* et *Tight*. Il est important pour les analyses de connaître l'efficacité d'identification des électrons dans le détecteur ATLAS et la qualité de la modélisation de cette efficacité par la simulation. Des différences en efficacité d'identification entre la simulation et les données sont corrigées par un facteur de correction.

Des mesures de l'efficacité d'identification des électrons et du facteur de correction sont combinées et la combinaison est mise à la disposition de la collaboration ATLAS. Le facteur de correction est calculé pour un spectre d'énergie de $7 < E_T < 200$ GeV. Les résultats de la combinaison sont montrés en Figure 8.4.

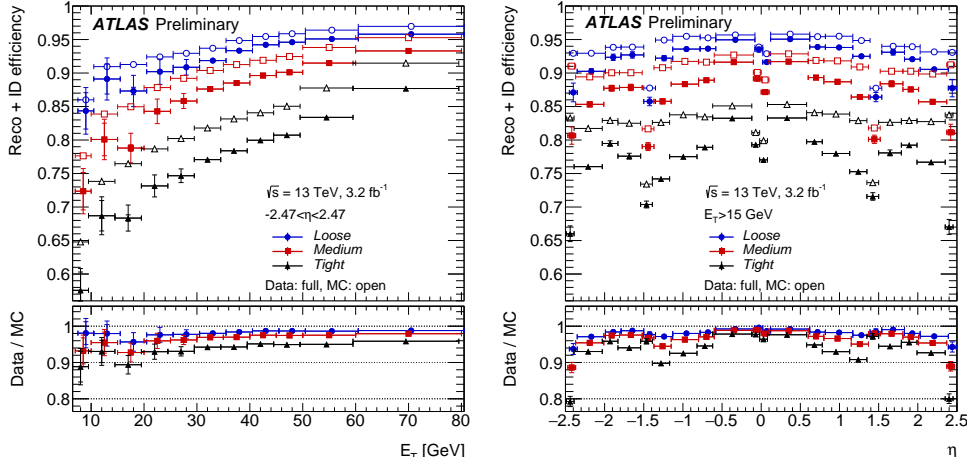


FIGURE 8.4: Efficacité combinée de reconstruction et d'identification des électrons dans les événements $Z \rightarrow ee$ en fonction de l'énergie transversale E_T , intégrée sur toute la plage de pseudo-apidité (gauche), et en fonction de la pseudo-apidité η , intégrée sur toute la plage E_T (droite).

8.6 Analyse de la production électrofaible de $Z\gamma$

Dans ce chapitre, l'étude de la production de diboson $Z\gamma$ en association avec un système de dijet à haute masse est présentée. Les données de collision protons-protons de l'expérience ATLAS au Grand collisionneur de hadrons du CERN à une énergie de centre de masse de $s = \sqrt{13}$ TeV, recueillies en 2015 et 2016 et correspondant à une luminosité intégrée de 36.1 fb^{-1} sont analysées.

Afin de séparer le signal des événements de fond, plusieurs critères de sélection sont appliqués pour sélectionner les jets, les leptons et les photons utilisés dans l'analyse.

Les événements doivent avoir au moins deux leptons de même saveur et de charge opposée (SFOC) et exactement un candidat photon satisfaisant les critères de sélection décrits ci-dessus. S'il y a plus d'une paire de leptons SFOC, la paire avec la masse invariante la plus proche de la masse du boson Z est sélectionnée. Pour sélectionner le processus $Z\gamma jj$ -EW, les événements doivent avoir au moins deux jets avec p_T au-dessus de 50 GeV (marquage des jets).

Plusieurs processus peuvent imiter la signature expérimentale de la production $Z\gamma jj$ -EW. Le fond principal est le $Z\gamma jj$ -QCD dont la forme est obtenue en utilisant MC et la normalisation est corrigée en utilisant des données.

Dans cette étude, une analyse multivariée est utilisée pour extraire le signal. Cette analyse est un outil de classification qui fournit une variable discriminante (score BDT) qui domine à des valeurs faibles (-1) pour le fond et des valeurs élevées (+1) pour le signal. La distribution du score BDT est présentée dans la Figure 8.5.

Puisque les productions $Z\gamma jj$ -EW et $Z\gamma jj$ -QCD ont un état final identique, selon la mécanique quantique, elles interfèrent entre elles. La compréhension de l'effet d'interférence est important pour l'analyse. Des simulations MC sont utilisées pour calculer le terme d'interférence et une étude spécifique a été réalisée pour étudier son impact sur les mesures de la section efficace fiduciale. Cette interférence se situe au

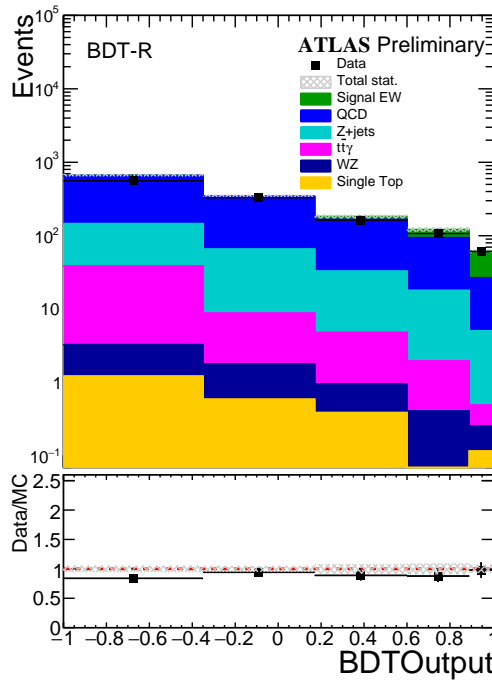


FIGURE 8.5: Résultats du BDT pour les processus de signal et de fond dans la région de la mesure.

niveau de 4% et est traité comme une variation systématique dans la mesure de la section efficace. L'étude de l'interférence permet également de ré-optimiser certaines des variables cinématiques puisqu'elle permet de mieux comprendre les différences cinématiques entre les deux modes de production.

8.7 Mesures de la section efficace de production de $Z\gamma$

La section efficace intégrée est mesurée dans une région définie au niveau des particules à l'aide de critères de sélection cinématique des objets et des événements définis dans le Tableau 6.1. Pour le calcul de la section efficace fiduciale, la formule suivante est utilisée:

$$\sigma_{\text{meas. EW}}^{\text{fid.}} = \frac{N_{\text{meas.}}^{\text{EW}}}{C \times \mathcal{L}}, \quad (8.2)$$

où $\mu_{\text{EW}} = N_{\text{meas.}}^{\text{EW}}/N_{\text{exp}}^{\text{EW}}$ est la force du signal, $N_{\text{meas.}}^{\text{EW}}$ est le nombre de signaux reconstruits (prévus par MC), C est l'efficacité de reconstruction d'événement qui est $C = \frac{N_{\text{exp. reco.}}^{\text{EW}}}{N_{\text{exp. gen.}}^{\text{EW}}}$, \mathcal{L} étant la luminosité intégrée de l'échantillon.

La section efficace de référence attendue est exprimée comme suit:

$$\sigma_{\text{exp. EW}}^{\text{fid.}} = \frac{N_{\text{exp.}}^{\text{EW}}}{C \times \mathcal{L}}, \quad (8.3)$$

où $N_{\text{exp.}}^{\text{EW}}$ est le nombre d'événements de signal simulés attendus dans la région du signal au niveau reconstruit pour la luminosité des données. \mathcal{L} .

par conséquent:

$$\sigma_{\text{obs. EW}}^{\text{fid.}} = \frac{N_{\text{meas.}}^{\text{EW}}}{N_{\text{exp.}}^{\text{EW}}} \times \sigma_{\text{exp. EW}}^{\text{fid.}} = \mu_{EW} \times \sigma_{\text{exp. EW}}^{\text{fid.}}, \quad (8.4)$$

La prédiction MS de MADGRAPH pour le processus $Z\gamma jj$ -EW sans interférence dans la région BDT est:

$$\sigma_{Z\gamma jj\text{-EW}}^{\text{fid. MADGRAPH}} = 7.75 \pm 0.03(\text{stat}) \pm 0.2(\text{PDF}) \pm 0.4(\text{scale}) \text{ fb}, \quad (8.5)$$

et la section efficace de production de $Z\gamma jj$ -EW observée, dérivée de la force du signal μ_{EW} , est:

$$\sigma_{\text{obs. } Z\gamma jj\text{-EW}}^{\text{fid.}} = 7.75_{-1.39}^{+1.47} (\text{stat.})_{-0.99}^{+0.93} (\text{exp.syst})_{-0.78}^{+1.01} (\text{mod.syst}) \pm 0.15(\text{lumi}) \text{ fb} \quad (8.6)$$

Les effets des détecteurs, tels que la résolution finie, l'acceptance limitée et l'efficacité imparfaite, faussent la distribution des observables mesurés. Le résultat de la mesure différentielle de la section efficace des mesures corrigées peut ensuite, avec leurs incertitudes, être directement comparé avec les prédictions théoriques correspondantes ainsi qu'avec les résultats d'autres expériences.

Les simulations MC sont utilisées afin de construire une matrice qui décrit les effets des migrations entre les valeurs générées et reconstruites. En tant que procédure de "unfolding", il s'agit de la correction d'une distribution de mesures. Différentes méthodes sont utilisées en fonction de l'analyse, dans cette analyse. On utilise ici une méthode bayésienne itérative. Des mesures différentielles de section efficace sont effectuées pour des observables sélectionnés pour leur bonne sensibilité à d'éventuels aQGC ou pour leur capacité à séparer les processus $Z\gamma jj$ -QCD et $Z\gamma jj$ -EW et à contraindre la description MC actuelle de la production de $Z\gamma jj$. Ces distributions sont le momentum transversal du photon (P_T^γ), la masse invariante du système diboson ($M_{Z\gamma}$), la masse invariante du système dijet (M_{jj}) et le jet multiplicity (N_{jets}).

Figure 8.6 présente les matrices de migration des quatre variables qui sont étudiées. On peut voir que dans tous les cas la matrice de migration est diagonale et que l'effet de migration est faible, sauf dans le cas de N_{jets} où l'effet de migration est au maximum de 30%. Figure 8.7 montre le résultat des mesures différentielles de section transversale pour tous les observables où l'on peut voir que les données sont en accord avec les prédictions du MS.

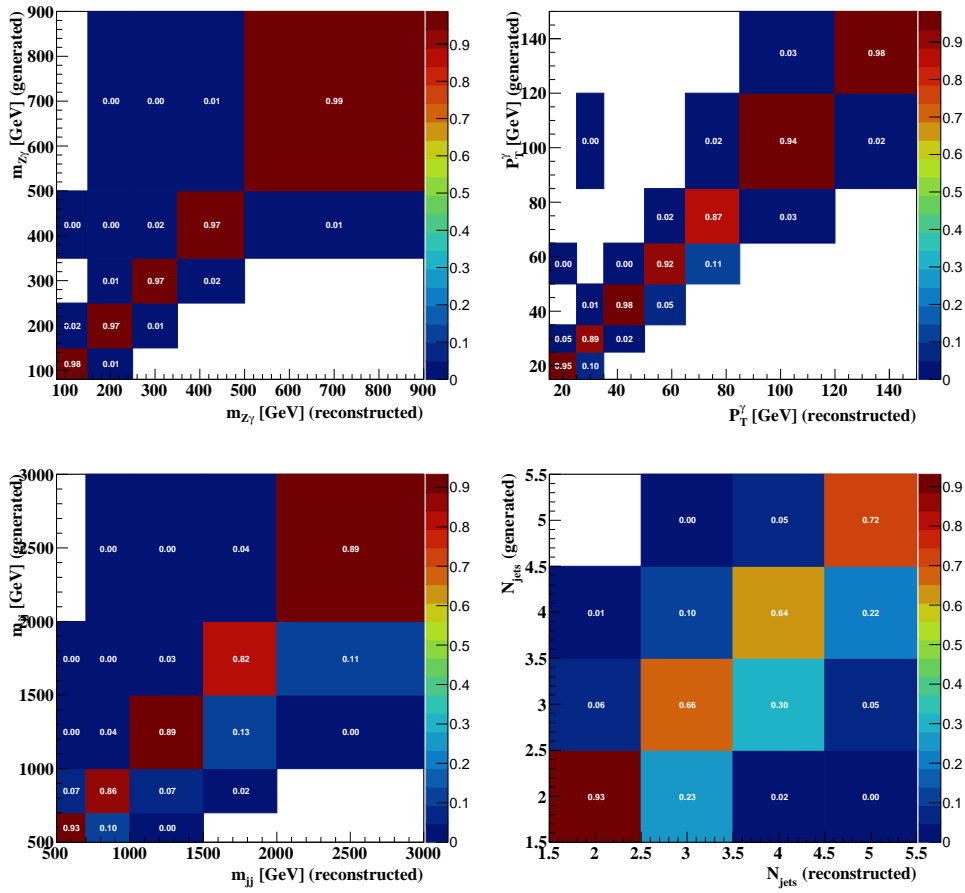


FIGURE 8.6: Présentation des matrices de migration pour les quatre variables qui sont corrigées par les effets détecteur (P_T^γ , $M_{Z\gamma}$, M_{jj} et N_{jets}).

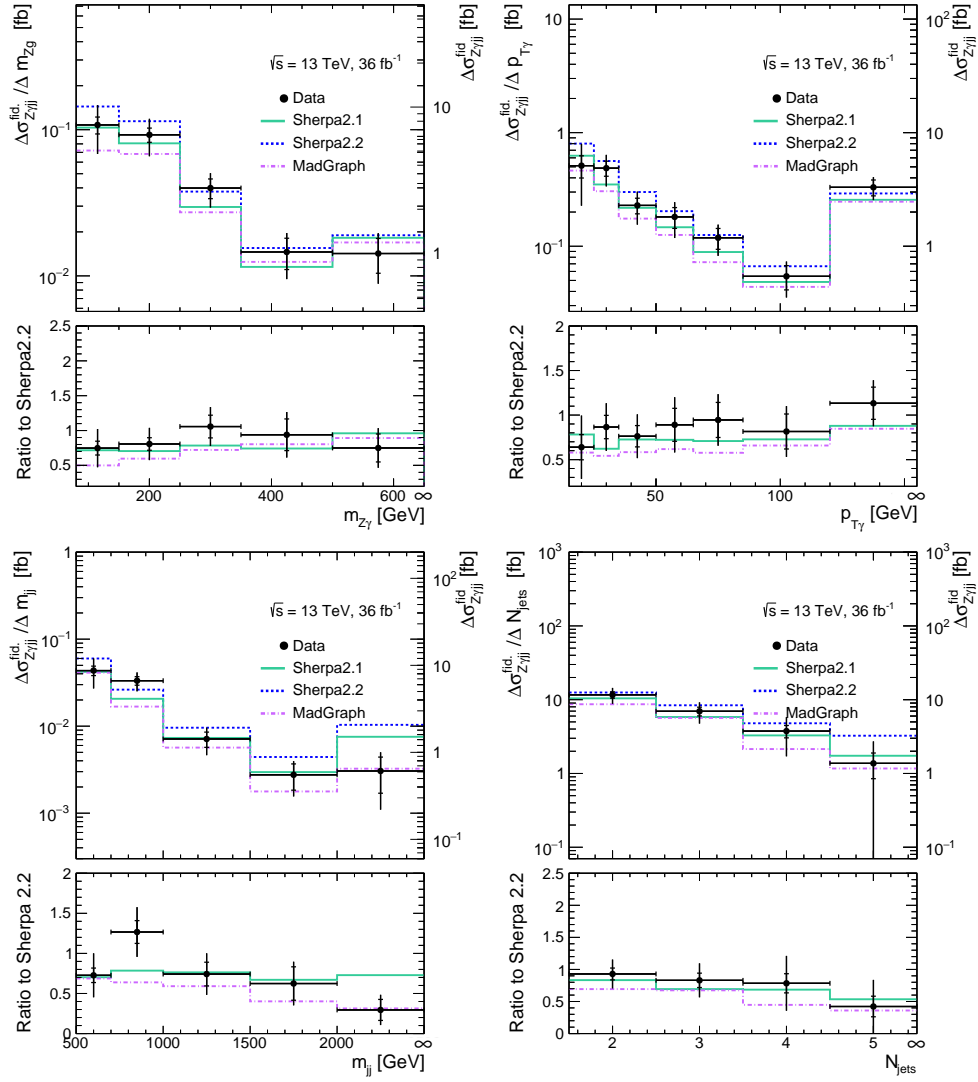


FIGURE 8.7: Présentation des sections efficace différentielles mesurées pour les quatre observables (P_T^γ , $M_{Z\gamma}$, M_{jj} et N_{jets}).

8.8 Conclusions

Dans le cadre de cette thèse, l'étude de la production de $Z\gamma jjj$ -EW avec le boson de Z se désintégrant leptoniequement et de la mesure de sa section efficace a été présentée. La section efficace différentielle a été mesurée pour la première fois pour ce canal. Les données ont été recueillies avec l'expérience ATLAS au LHC en 2015 et 2016 à une énergie centrale de masse de $\sqrt{s} = 13$ TeV et une luminosité intégrée de 36.1 fb^{-1} . Les principales incertitudes systématiques dans la mesure de la section transversale proviennent des incertitudes théoriques (échelle QCD et PDF), de l'incertitude systématique des jets et de l'identification des leptons et photons. Dans cette thèse, des études détaillées sur l'efficacité de l'identification des électrons ont été présentées. L'accord entre les données et la simulation MC a été étudié et des facteurs d'échelle ont été extraits pour corriger les erreurs de modélisation des événements de fond. De plus, une nouvelle méthode de mesure de l'efficacité de l'identification des électrons a été mise en œuvre pour les électrons ayant une faible énergie transversale et venant de la désintégration du boson Z . Celles-ci ont été combinées à une mesure complémentaire de l'efficacité de l'identification des électrons et ont été fournies à la collaboration ATLAS. La principale contribution de l'incertitude provient de l'estimation du fond. Cette précision sur les facteurs de l'échelle d'identification des électrons se traduit par une contribution d'incertitude de 1% dans la section de $Z\gamma jj$ -EW mesure.

Bibliography

- [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012), [arXiv:1207.7214 \[hep-ex\]](#).
- [2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012), [arXiv:1207.7235 \[hep-ex\]](#).
- [3] ATLAS Collaboration, *Evidence for Electroweak Production of $W^\pm W^\pm jj$ in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*, *Phys. Rev. Lett.* **113** (2014), [arXiv:1405.6241 \[hep-ex\]](#).
- [4] ATLAS Collaboration, *Measurement of $W^\pm W^\pm$ vector-boson scattering and limits on anomalous quartic gauge couplings with the ATLAS detector*, *Phys. Rev. D* **96** (2017), [arXiv:1611.02428 \[hep-ex\]](#).
- [5] CMS Collaboration, *Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets*, *Phys. Rev. Lett.* **114** (2015), [arXiv:1410.6315 \[hep-ex\]](#).
- [6] ATLAS Collaboration, *Measurements of $W^\pm Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector and limits on anomalous gauge boson self-couplings*, *Phys. Rev. D* **93** (2016), [arXiv:1603.02151 \[hep-ex\]](#).
- [7] 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* **04** (2014), [arXiv:1401.7610 \[hep-ex\]](#).
- [8] CMS Collaboration, *Measurement of electroweak production of a W boson and two forward jets in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **11** (2016), [arXiv:1607.06975 \[hep-ex\]](#).
- [9] CMS Collaboration, *Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J. C* **75** (2015), [arXiv:1410.3153 \[hep-ex\]](#).
- [10] ATLAS Collaboration, *Measurements of $Z\gamma$ and $Z\gamma\gamma$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Phys. Rev. D* **93** (2016), [arXiv:1604.05232 \[hep-ex\]](#).

- [11] CMS Collaboration, *Search for $WW\gamma$ and $WZ\gamma$ production and constraints on anomalous quartic gauge couplings in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Rev. D* **90** (2014), [arXiv:1404.4619 \[hep-ex\]](#).
- [12] ATLAS Collaboration, *Studies of $Z\gamma$ production in association with a high-mass dijet system in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **07** (2017), [arXiv:1705.01966 \[hep-ex\]](#).
- [13] CMS Collaboration, *Measurement of the cross section for electroweak production of $Z\gamma$ in association with two jets and constraints on anomalous quartic gauge couplings in proton–proton collisions at $\sqrt{s} = 8$ TeV*, *Phys. Lett.* (2017), [arXiv:1702.03025 \[hep-ex\]](#).
- [14] CMS Collaboration, *Measurement of electroweak production of $Z\gamma$ in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Tech. Rep. CMS-PAS-SMP-18-007, CERN, Geneva, 2019. <https://cds.cern.ch/record/2682214>.
- [15] C. P. Burgess and G. D. Moore, *The standard model: A primer*. Cambridge University Press, 2006.
- [16] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.
- [17] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964).
- [18] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, *Phys. Rev.* **145** (1966).
- [19] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012), [arXiv:1207.7214 \[hep-ex\]](#).
- [20] D. P. Barber et al., *Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies*, *Phys. Rev. Lett.* **43** (1979).
- [21] A. Pich, *The Standard Model of Electroweak Interactions*, 2012. [arXiv:1201.0537 \[hep-ph\]](#).
- [22] A. Pich, *Aspects of quantum chromodynamics*, 1999. [arXiv:hep-ph/0001118 \[hep-ph\]](#).
- [23] S. Bethke, *Jet physics at LEP and world summary of $\alpha(s)$* , 1998. [arXiv:hep-ex/9812026 \[hep-ex\]](#).
- [24] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967).
- [25] A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, *Phys. Lett.* **13** (1964).

- [26] S. L. Glashow, *Partial Symmetries of Weak Interactions*, [Nucl. Phys. **22** \(1961\)](#).
- [27] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, [Phys. Rev. **105** \(1957\)](#).
- [28] A. Hocker and Z. Ligeti, *CP violation and the CKM matrix*, *Ann. Rev. Nucl. Part. Sci.* **56** (2006), [arXiv:hep-ph/0605217 \[hep-ph\]](#).
- [29] T. Plehn, *Lectures on LHC Physics*, [Lect. Notes Phys. **844** \(2012\)](#), [arXiv:0910.4182 \[hep-ph\]](#).
- [30] The LHCb collaboration, *Measurement of the forward-backward asymmetry in $Z/\gamma^* \rightarrow \mu^+\mu^-$ decays and determination of the effective weak mixing angle*, [Journal of High Energy Physics **2015** \(2015\)](#).
- [31] S. Dawson, *Introduction to electroweak symmetry breaking*, in *Proceedings, Summer School in High-energy physics and cosmology: Trieste, Italy, June 29-July 17, 1998*. 1998. [arXiv:hep-ph/9901280 \[hep-ph\]](#).
- [32] Particle Data Group Collaboration, M. T. et al, *Review of Particle Physics*, [Phys. Rev. D **98** \(2018\)](#).
- [33] P. Anger, M. Kobel, and S. Lammers, *Probing Electroweak Gauge Boson Scattering with the ATLAS Detector at the Large Hadron Collider*, Jun, 2014. <https://cds.cern.ch/record/1753849>.
- [34] D. L. Rainwater, R. Szalapski, and D. Zeppenfeld, *Probing color singlet exchange in $Z + 2$ jet events at the CERN LHC*, [Phys. Rev. **D54** \(1996\)](#), [arXiv:hep-ph/9605444 \[hep-ph\]](#).
- [35] J. R. Ellis, *Beyond the standard model for hill walkers*, in *1998 European School of high-energy physics, St. Andrews, Scotland, 23 Aug-5 Sep 1998: Proceedings*. 1998. [arXiv:hep-ph/9812235 \[hep-ph\]](#).
- [36] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, [JHEP **10** \(2010\)](#), [arXiv:1008.4884 \[hep-ph\]](#).
- [37] G. Perez, M. Sekulla, and D. Zeppenfeld, *Anomalous quartic gauge couplings and unitarization for the vector boson scattering process $pp \rightarrow W^+W^+jjX \rightarrow \ell^+\nu_\ell\ell^+\nu_\ell jjX$* , [Eur. Phys. J. **C78** \(2018\)](#), [arXiv:1807.02707 \[hep-ph\]](#).
- [38] O. J. P. Eboli, M. C. Gonzalez-Garcia, and J. K. Mizukoshi, *$pp \rightarrow jje^\pm\mu^\pm\nu\nu$ and $jje^\pm\mu^\pm\nu\nu$ at $O(\alpha_{em}^6)$ and $O(\alpha_{em}^4\alpha_s^2)$ for the study of the quartic electroweak gauge boson vertex at CERN LHC*, [Phys. Rev. **D74** \(2006\)](#), [arXiv:hep-ph/0606118 \[hep-ph\]](#).

- [39] J. M. Campbell, J. W. Huston, and W. J. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, *Rept. Prog. Phys.* **70** (2007) **89**, [arXiv:hep-ph/0611148](https://arxiv.org/abs/hep-ph/0611148) [hep-ph].
- [40] A. D. Martin, *Proton structure, Partons, QCD, DGLAP and beyond*, *Acta Phys. Polon.* **B39** (2008), [arXiv:0802.0161](https://arxiv.org/abs/0802.0161) [hep-ph].
- [41] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J.* **C75** (2015), [arXiv:1412.3989](https://arxiv.org/abs/1412.3989) [hep-ph].
- [42] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.
- [43] F. Marcastel, *CERN's Accelerator Complex. La chaîne des accélérateurs du CERN*, CERN Document Server, <https://cds.cern.ch/record/1621583>.
- [44] ATLAS Experiment - public results, *ATLAS Luminosity Public Results Run2*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [45] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008).
- [46] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997. <https://cds.cern.ch/record/331063>.
- [47] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996. <https://cds.cern.ch/record/331061>.
- [48] ATLAS Collaboration, *Muon identification and performance in the ATLAS experiment*, Tech. Rep. ATL-PHYS-PROC-2018-078, CERN, Geneva, Jul, 2018. <https://cds.cern.ch/record/2631716>.
- [49] A. Ruiz-Martinez and ATLAS Collaboration, *The Run-2 ATLAS Trigger System*, tech. rep., CERN, Geneva, 2016. <https://cds.cern.ch/record/2133909>.
- [50] M. G. Pia and G. Weidenspointner, *Monte Carlo Simulation for Particle Detectors*, Tech. Rep. arXiv:1208.0047, Aug, 2012. <https://arxiv.org/pdf/1208.0047.pdf>.
- [51] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008), [arXiv:0710.3820](https://arxiv.org/abs/0710.3820) [hep-ph].
- [52] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009), [arXiv:0811.4622](https://arxiv.org/abs/0811.4622) [hep-ph].

- [53] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014), [arXiv:1405.0301 \[hep-ph\]](#).
- [54] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010), [arXiv:1002.2581 \[hep-ph\]](#).
- [55] GEANT4 Collaboration, GEANT4 Collaboration, *GEANT4: a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003).
- [56] ATLAS Collaboration, *ATLAS Computing: technical design report*. Technical Design Report ATLAS. CERN, Geneva, 2005. <https://cds.cern.ch/record/837738>.
- [57] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, Tech. Rep. ATLAS-CONF-2016-024, CERN, Geneva, Jun, 2016. <http://cds.cern.ch/record/2157687>.
- [58] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV*, [arXiv:1902.04655](#).
- [59] L. Pishdad and F. Labeau, *A New Reduction Scheme for Gaussian Sum Filters*, CoRR (2014), [arXiv:1405.3164](#).
- [60] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, tech. rep., CERN, Geneva, May, 2012. <https://cds.cern.ch/record/1449796>.
- [61] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, *Eur. Phys. J.*, [arXiv:1810.05087 \[hep-ex\]](#).
- [62] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, [arXiv:physics/0703039 \[physics.data-an\]](#).
- [63] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett.* (2008), [arXiv:0707.1378 \[hep-ph\]](#).
- [64] C. Blocker, *Uncertainties on Efficiencies*, CDF/MEMO/STATISTICS/PUBLIC/7168, http://www-cdf.fnal.gov/physics/statistics/notes/cdf7168_eff_uncertainties.ps.

- [65] M. Oreglia, *A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi$* . PhD thesis, SLAC, 1980. <http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=slac-r-236.html>.
- [66] J. E. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'* . PhD thesis, SLAC, 1982. <http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=slac-r-255.html>.
- [67] CMS Collaboration, *J/ψ and $\psi(2S)$ production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* (2012), [arXiv:1111.1557](https://arxiv.org/abs/1111.1557) [[hep-ex](#)].
- [68] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data*, *Eur. Phys. J.* (2017), [arXiv:1612.01456](https://arxiv.org/abs/1612.01456) [[hep-ex](#)].
- [69] ATLAS Isolation Selection Tool, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/IsolationSelectionTool#Photons_AN1.
- [70] K. Olive, *Review of Particle Physics*, *Chinese Physics C* (2014).
- [71] ATLAS Collaboration, *Measurement of inclusive and differential cross sections in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* (2017), [arXiv:1708.02810](https://arxiv.org/abs/1708.02810) [[hep-ex](#)].
- [72] ATLAS Collaboration, *Measurement of $W^\pm Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C79** (2019), [arXiv:1902.05759](https://arxiv.org/abs/1902.05759) [[hep-ex](#)].
- [73] LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [arXiv:1610.07922](https://arxiv.org/abs/1610.07922) [[hep-ph](#)].
- [74] ATLAS Collaboration, *Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \rightarrow WW^* \rightarrow e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett.* **B789** (2019), [arXiv:1808.09054](https://arxiv.org/abs/1808.09054) [[hep-ex](#)].
- [75] F. Campanario, M. Kerner, L. D. Ninh, and D. Zeppenfeld, *$Z\gamma$ production in association with two jets at next-to-leading order QCD*, *Eur. Phys. J.* **C74** (2014) 3085, [arXiv:1407.7857](https://arxiv.org/abs/1407.7857) [[hep-ph](#)].
- [76] F. Campanario, M. Kerner, and D. Zeppenfeld, *$Z\gamma$ production in vector-boson scattering at next-to-leading order QCD*, *JHEP* **01** (2018) 160, [arXiv:1704.01921](https://arxiv.org/abs/1704.01921) [[hep-ph](#)].
- [77] ATLAS Collaboration, G. Aad et al., *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J.* **C77** (2017) 490, [arXiv:1603.02934](https://arxiv.org/abs/1603.02934) [[hep-ex](#)].

- [78] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008), [arXiv:0802.1189 \[hep-ph\]](#).
- [79] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev.* **D96** (2017), [arXiv:1703.09665 \[hep-ex\]](#).
- [80] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, *Eur. Phys. J.* **C77** (2017), [arXiv:1705.02211 \[hep-ex\]](#).
- [81] Connelly, Ian, *Performance and calibration of b-tagging with the ATLAS experiment at LHC Run-2*, *EPJ Web Conf.* **164** (2017).
- [82] Tracking CP recommendations, <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TrackingCPE0YE2015>.
- [83] Latest MCP recommendation for muon selection, <https://twiki.cern.ch/twiki/bin/view/Atlas/MuonSelectionTool>.
- [84] Adams, D et al., *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014*, Tech. Rep. ATL-PHYS-INT-2014-018, CERN, Geneva, Jul, 2014. <https://cds.cern.ch/record/1743654>.
- [85] Electron four momentum correction, <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronPhotonFourMomentumCorrection>.
- [86] ATLAS Collaboration, *Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing*, Tech. Rep. ATL-PHYS-PUB-2015-041, CERN, Geneva, Sep, 2015. <http://cds.cern.ch/record/2048202>.
- [87] Latest egamma tune for electron identification, <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaIdentificationRun2>.
- [88] Official Isolation Working Points, <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/IsolationSelectionTool>.
- [89] T. Golling, H. S. Hayward, P. U. E. Onyisi, H. J. Stelzer, and P. Waller, *The ATLAS Data Quality Defect Database System*, *Eur. Phys. J.* **C72** (2012), [arXiv:1110.6119 \[physics.ins-det\]](#).
- [90] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, *Eur. Phys. J.* **C77** (2017) 317, [arXiv:1611.09661 \[hep-ex\]](#).
- [91] P. V. Tilmann Gneiting, *Receiver Operating Characteristic (ROC) Curves*, arXiv (2018), <https://arxiv.org/pdf/1809.04808.pdf>.

- [92] ROOT Collaboration, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. CERN-OPEN-2012-016, New York U., New York, Jan, 2012. <https://cds.cern.ch/record/1456844>.
- [93] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J.* **C71** (2011), [arXiv:1007.1727](https://arxiv.org/abs/1007.1727) [physics.data-an].
- [94] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, *Eur. Phys. J.* **C76** (2016), [arXiv:1608.03953](https://arxiv.org/abs/1608.03953) [hep-ex].
- [95] *Notes on MC Tuning*, <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MCTuningRecommendations>.
- [96] ATLAS Collaboration, *Measurement of the $Z\gamma$ cross-section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, Tech. Rep. ATL-COM-PHYS-2018-1493, CERN, Geneva, Oct, 2018. <https://cds.cern.ch/record/2644987>.
- [97] G. Cowan, *Statistical Data Analysis*. Oxford science publications. Clarendon Press, 1998. http://www.sherrytowers.com/cowan_statistical_data_analysis.pdf.
- [98] Biondi, Silvia, *Experience with using unfolding procedures in ATLAS*, *EPJ Web Conf.* **137** (2017).
- [99] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem*, *Nucl. Instrum. Meth.* **A362** (1995).
- [100] A. Hocker and V. Kartvelishvili, *SVD approach to data unfolding*, *Nucl. Instrum. Meth.* **A372** (1996), [arXiv:hep-ph/9509307](https://arxiv.org/abs/hep-ph/9509307) [hep-ph].
- [101] B. Malaescu, *An Iterative, Dynamically Stabilized(IDS) Method of Data Unfolding*, in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*. 2011. [arXiv:1106.3107](https://arxiv.org/abs/1106.3107) [physics.data-an].
- [102] T. Adye, *Unfolding algorithms and tests using RooUnfold*, in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, CERN. CERN, Geneva, 2011. [arXiv:1105.1160](https://arxiv.org/abs/1105.1160) [physics.data-an].
- [103] A. Armbruster, K. Kroeninger, B. Malaescu, and F. Spano, *Practical considerations for unfolding*, Tech. Rep. ATL-COM-PHYS-2014-277, CERN, Geneva, Apr, 2014. <https://cds.cern.ch/record/1694351>.

-
- [104] ATLAS Collaboration, *Prospective study of vector boson scattering in WZ fully leptonic final state at HL-LHC*, Tech. Rep. ATL-PHYS-PUB-2018-023, CERN, Geneva, Oct, 2018. <http://cds.cern.ch/record/2645271>.
- [105] ATLAS Collaboration, *Prospect study of electroweak production of a Z boson pair plus two jets at the HL-LHC*, Tech. Rep. ATL-PHYS-PUB-2018-029, CERN, Geneva, Nov, 2018. <http://cds.cern.ch/record/2647219>.