Differential measurements of Z and γ bosons produced with jets at the CMS experiment

Angelo Giacomo Zecchinelli
Imperial College London
Department of Physics

A dissertation submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

This thesis presents measurements of the differential cross sections of Z + jets and γ + jets, and their ratio, as a function of the boson transverse momentum. In addition, a description of measurements of the angular distribution between the Z boson direction and a jet in events where the Z boson is emitted collinear with a jet is reported. The analysis is based on a data sample of proton-proton collisions at a centre-of-mass energy of 13 TeV corresponding to an integrated luminosity of 35.9 fb^{-1} recorded by the CMS experiment at the LHC. The data are compared to different theoretical predictions after correcting for detector effects. In general, the predictions at higher orders in perturbation theory show better agreement with the data. These results represent the first measurement of the differential cross section ratio of Z + jets and γ + jets at 13 TeV and the first explicit measurement of collinear Z emission.

Declaration

I, the author of this thesis, hereby declare that the work contained within this document is my own. It is produced on top of the existing work and ideas of several individuals and the Compact Muon Solenoid collaboration as a whole. All figures labelled "CMS" are sourced from CMS publications, while the ones labelled "CMS Preliminary" are sourced from unpublished or preliminary public results. Figures that have not been produced by the author are referenced accordingly. The work performed by others is reviewed by the author, as key aspects are relevant in the later stages, in chapters 1 – 4. The qualification task performed by the author, maintaining and developing the Online-to-Offline software for the CMS L1 Trigger, has been omitted in this thesis, providing only a brief description of the current software implementation in 3.2.6.3. Chapter 5 details the event selection and categorisation designed by the author, although inputs are based on work performed by others. All remaining chapters detail the work done solely by the author.

Angelo Giacomo Zecchinelli.

Contents

List of Figures

List of Tables

Chapter 1.

Introduction

The Standard Model of particle physics (SM) has proven to be remarkably successful in describing experimental results and has shown considerable predictive power. With the observation of the Higgs boson by ATLAS [1] and CMS [2] in 2012, the latest particle predicted by the model has been found, and the theory is now complete. Despite this, the SM lacks a candidate for dark matter, the presence of which in the universe was inferred in 1983 [3, 4, 5], a neutrino mass mechanism, and a quantum description of gravitational interaction.

For this reason, numerous direct searches targeting physics beyond the SM (BSM) are performed at the LHC; so far, no clear signal of BSM physics has been observed. While direct searches continue probing new regions of the phase space and new couplings between SM and BSM physics, SM precision measurements provide a complementary approach, looking for eventual deviations from predictions. Among the different standard model processes investigated at the LHC, the production of vector bosons in association with jets is of particular interest as a valuable test of the Standard Model and the search for a wide variety of phenomena beyond the Standard Model.

This thesis presents a measurement of the differential production cross-sections of $\mathrm{Z}+$ jets [1], γ + jets, and their ratio for highly Lorentz-boosted bosons, and the first study at hadron colliders of the collinear emission of a Z boson close to a jet.

Such measurements elucidate our understanding of the SM in challenging regions of the phase space. In particular, the high boson p_{T} regime and collinear Z boson emission provide important landmarks for perturbative quantum chromodynamics (QCD) and electroweak (EW) calculations, as implemented in the state-of-the-art Monte Carlo (MC) event generators. The differential cross-section measurement of the $\mathrm{Z}+$ jets and γ + jets processes can constrain higher-order perturbative QCD and EW calculations that lead to a non-negligible dependence of the cross-section on boson p_{T}. The weak radiative corrections become large and negative at high energies due to large Sudakov logarithms arising from the virtual exchange of soft or collinear massive gauge bosons [6, 7, 8, 9, 10, 11]. The corrections have the form $\alpha \log^2(\hat{s}/M^2_{\mathrm{W,Z}})$,

[1] Throughout the text, the $\mathrm{Z}+$ jets is understood to include Drell–Yan + jets production.

where $\hat{s}$ is the squared parton centre-of-mass energy, α is the gauge coupling of the unified EW theory, and $M_{W,Z}$ is the mass of the W or Z boson.

Moreover, these measurements play an important role in BSM searches, such as supersymmetry, dark matter, and invisible decays of the Higgs boson. The production of $Z + \text{jets}$ and γ + jets at high boson p_T are the key control processes used to estimate the backgrounds to these searches from a Z boson decaying to neutrinos ($Z \to \nu\bar{\nu}$), and the Z/γ ratio is a theoretical input in using γ + jets to predict $Z \to \nu\bar{\nu}$ [12]. Accurate modelling of these processes thus maximises the discovery potential of many new physics searches.

The emission of a real W or Z boson can contribute significantly to inclusive $W + \text{jets}$ and $Z + \text{jets}$ measurements at high energies [13, 14, 15]. These processes can be accessed by selecting a boosted event topology and studying the region of small angular separation between a W or Z boson and a jet. The fully reconstructable decay products from the Z boson (in this case, to muons) allow a direct measurement of the angular separation between the Z boson and the closest jet. This thesis presents the first measurement of the collinear emission of a Z boson, as studies of the real emission of the W boson have previously been performed by the ATLAS and CMS collaborations at 8 TeV [16] and 13 TeV [17] respectively.

The presented thesis is organised as follows: Chapter 2 provides an introduction to the Standard Model of the elementary particles, while Chapter 3 gives an overview of the experimental setup, with a description of the accelerator system and a detailed description of the different components making up the CMS detector. In Chapter 4, the techniques used for the reconstruction of physics objects starting from the signals registered in the subdetectors are discussed, whereas Chapters 5 to 7 are dedicated to the description of the analysis. Chapter 5 describes the event selection and the corrections applied to simulated samples, Chapter 6 discusses the estimation of backgrounds and Chapter 7 shows the differential cross-section measurements. In Chapter 8, the results are discussed, providing a summary of the analysis and possible future outlook.

Chapter 2.

The Standard Model

The SM is a quantum field theory that describes the elementary particles and their interactions and was developed during the 20^{th} century. This theory is extremely well supported by experimental evidence, and it has been tested at a wide range of energies scales up to the order of TeV. In this Chapter, the fundamental particles and their interactions are described, and the spontaneous symmetry breaking through the Higgs mechanism is elucidated. Lastly, Section 2.3 is dedicated to discussing motivations behind precision measurements of the SM at colliders and provides an overview of the latest results performed with the CMS experiment.

2.1. The fundamental interactions in the Standard Model

The SM is a gauge theory based on the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where $SU(3)_C$ is the symmetry group which describes the strong interactions between coloured charged particles (see Section 2.1.2), $SU(2)_L$ describes the weak interactions between isospin left-handed doublets and $U(1)$ represents the interactions between hypercharged particles (see Section 2.1.3). The weak isospin and the hypercharge symmetries ($SU(2)_L \otimes U(1)_Y$) are broken at low energies, while the colour symmetry $SU_C(3)$ and the electromagnetic gauge symmetry $U_{el}(1)$ are preserved.

In the SM, the constituents of ordinary matter are 12 fundamental fermions with spin $\frac{1}{2}$ and their antiparticles, which have opposite quantum numbers and the same mass. Instead, the carriers of the forces are gauge bosons, which are associated with the generators of the symmetry groups. The fermions are divided into two subgroups: quarks, particles that carry a colour charge, and therefore can also interact via the strong force and, leptons, which interact only through the electroweak force. Due to the confinement of the colour charge, i.e. only colourless states can be isolated, isolated quarks can not be observed experimentally. The observable colour singlet bound states are called hadrons; they present an integer charge, and depending on the number of quarks involved in the bound state, they are divided into mesons, spin integer particles formed by a quark and an anti-quark, baryons, fermions formed

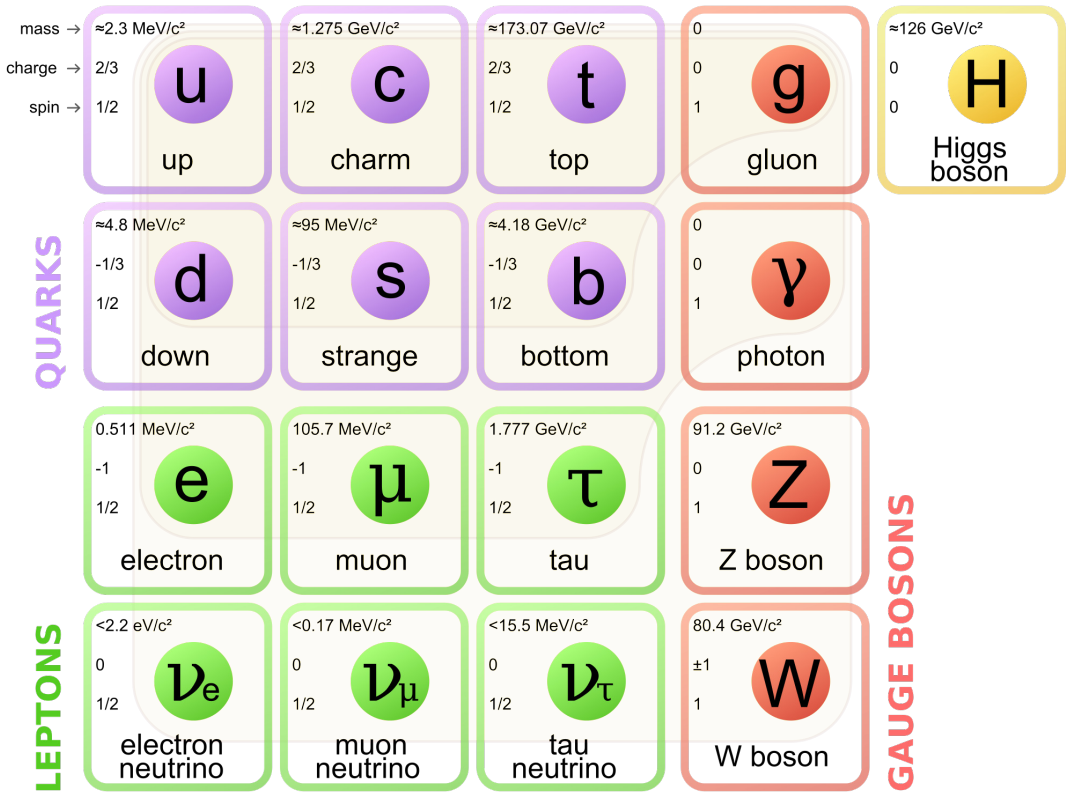


Figure 2.1.: Standard Model fundamental particles [18].

by three quarks, and exotics states, composed of a higher number of quarks. The SM gauge bosons, which play the role of mediators of the forces between particles, are the photon, the Z^0 and $W^{\pm}$ bosons, and the gluons. The photons are massless and mediate the electromagnetic interaction, the Z and W are massive particles that mediate the weak force, while the strong force carriers are the eight gluons, which are massless as well, but present a colour charge themselves and can therefore self interact. Finally, the Higgs boson is the only (neutral) scalar particle introduced in the Standard Model to explain the masses of all the other particles considered in the theory. Figure 2.1 summarises the composition of the Standard Model.

The interactions between the fermions in the SM are built using the local gauge invariance principle, requiring the Lagrangian to be invariant under the group of local symmetry transformations. To preserve the invariance under local gauge transformations, particular vector fields, called gauge fields, must be added to the Lagrangian. Each interaction is associated with a propagator, and the number of propagators depends on the number of symmetry generators.

2.1.1. Quantum Electrodynamics

Quantum electrodynamics (QED) is a gauge theory describing an electrically charged particle and the photon. Starting from the Dirac Lagrangian , which describes a spinor field associated with fermions of spin $\frac{1}{2}$ and masses m:

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi \ , \tag{2.1}$$

one can enforce the invariance of the theory under local $U(1)$ gauge transformations. Considering a generic local variation:

$$\psi \to \psi' = \psi e^{i\theta(x)} \ , \tag{2.2}$$

where θ is a function of the space-time coordinates, it is necessary to introduce additional terms to preserve the Lagrangian invariance. This can be obtained by adding to the Lagrangian a vector field A_{μ} that transforms as

$$A_{\mu} \to A_{\mu} + \frac{1}{q}\partial_{\mu}\theta(x) \ , \tag{2.3}$$

where q is a free parameter corresponding to the charge of the particle associated with the spinor field, and redefining the covariant derivative as

$$D_{\mu} = \partial_{\mu} - iqA_{\mu} \ . \tag{2.4}$$

The Lagrangian for a vector field A_{μ}, describing a spin 1 particle, is derived from Maxwell's equations:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \ , \tag{2.5}$$

where $F^{\mu\nu}$ is the Faraday tensor, defined as:

$$F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu} \ . \tag{2.6}$$

Adding together the two fields and including the proper coupling term, the QED Lagrangian becomes:

$$\mathcal{L}_{QED} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + [i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi] + (q\bar{\psi}\gamma^{\mu}\psi)A_{\mu} \ . \tag{2.7}$$

This Lagrangian is invariant under local $U(1)$ gauge transformations and represents a massive spin $1/2$ particle interacting with a massless spin 1 particle, i.e. a fermion and a photon.

The coupling constant of the electromagnetic interactions, the fine structure constant, is defined as:

$$\alpha = \frac{1}{4\pi\epsilon_0}\frac{e^2}{\hbar c} \approx \frac{1}{137} \ , \tag{2.8}$$

where:

- e is the elementary charge,
- $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant,
- c is the speed of light in vacuum,
- ϵ_0 is the electric constant or permittivity of free space.

2.1.2. Quantum Chromodynamics

The strong interaction is mediated by eight massless gluons, which correspond to the eight generators of $SU(3)$ and the field theory describing these interactions is called Quantum Chromodynamics (QCD). Gluons interact with coloured particles, i.e. quarks, and among themselves due to the non-abelian structure of the $SU(3)_C$ group. This allows for three and four gluons vertices. The strong interaction Lagrangian can be written as:

$$L_{QCD} = -\frac{1}{4}F^{(a)}_{\mu\nu}F^{(a)\mu\nu} + i\sum_q \bar{\psi}^i_q\gamma^{\mu}(D_{\mu})_{ij}\psi^j_q - \sum_q m_q\bar{\psi}^i_q\psi_{qi} \ , \tag{2.9}$$

with

$$F^{(a)}_{\mu\nu} = \partial_{\mu}A^a_{\nu} - \partial_{\nu}A^a_{\mu} - g_s f_{abc}A^b_{\mu}A^c_{\nu} \ , \tag{2.10}$$

and

$$(D_\mu)_{ij} = \delta_{ij}\partial_\mu + ig_s \sum_a \frac{\lambda^a_{ij}}{2} A^a_\mu \,, \tag{2.11}$$

where g_s is the coupling constant of the strong interaction, f_{ijk} (i, j, k =1,...,8) are the $SU(3)$ structure constants, ψ^i_q is the Dirac spinor associated to a quark field with colour i and flavour q, while A^a_μ are the gluon fields. Eventually, the λ^a_{ij} are the matrix representation of the $SU(3)$ generators. It should be noticed that this definition of the covariant derivative makes $\mathcal{L}_{QCD}$ invariant under local gauge transformations.

At first order, the coupling constant of strong interactions is expressed as:

$$\alpha_s(\mu^2) = \frac{g_s^2(\mu^2)}{4\pi} = \frac{12\pi}{(33 - n_f)\ln(\mu^2/\Lambda^2_{QCD})} \,, \tag{2.12}$$

where μ is the energy transferred in the interaction, n_f represents the number of quark flavours considered, and Λ_{QCD} is the energy scale above which perturbative QCD is valid, $\Lambda_{QCD} \sim 200$ MeV.

Depending on the energy scale, i.e. the magnitude of α_S, QCD has two peculiar properties:

- *Colour Confinement*, which means that colour-charged particles cannot be isolated. This particular phenomenon is caused by the force between quarks not diminishing as they are separated. Thus, when two quarks are separated, the gluon field's energy increases to the point it is favourable to create another quark pair. As a result, free quarks are bound into hadrons before they can be detected, and their properties can only be inferred by measuring the resulting colour-free particles (hadronic jets).
- *Asymptotic freedom*, meaning that quarks and gluons are weakly coupled in high-energy interactions. In the limit of $\mu^2 \to \infty$, quarks behave as free particles. In this particular regime, it is possible to use perturbative theory to calculate the strong interaction matrix element. In fact, the perturbative expansion of the S-matrix is valid only if the coupling constant is small, i.e. $\alpha_S << 1$. This is verified when the energy scale of the interaction is larger than Λ_{QCD}, $\mu^2 >> \Lambda_{QCD}$. The dependency of α_S on μ^2 is related to the distance between quarks in the interaction. For scatterings happening at short distances, which means at large value of μ^2, the coupling between quarks weakens, and they behave as free particles. In the opposite case, when μ^2 is small or at large distances, the coupling between quarks is so strong that they cannot be extracted from hadrons.

2.1.3. Electroweak unification

The modern theory of weak interaction has been developed from the theory of β decay, established by Enrico Fermi in 1933 [19]. This interaction is characterised by a Vector-Axial structure that violates parity conservation. In fact, the charged weak current couples only to left-handed fermions and right-handed antifermions. The weak interaction is described by the $SU(2)_L$ symmetry group, where the subscript L refers to the left-handed chirality of the particles participating in the interaction; left-handed fermions are grouped into weak isospin doublets with $T = \frac{1}{2}$ and $T_3 = \pm\frac{1}{2}$, right-handed fermions are instead singlets in weak isospin ($T = 0$, $T_3 = 0$). The weak interaction is mediated by the $W^{\pm}$ and Z bosons.

It is now known that the weak and electromagnetic interactions are different aspects of a unique interaction, called the electroweak interaction, described by a larger gauge symmetry group. The concept of electroweak unification was introduced by Glashow [20] in 1961 and refined by including symmetry preserving mass terms for gauge bosons, independently by Weinberg [21] and Salam [22] in 1967, historically marking the advent of the Standard Model. This group is $SU(2)_L \times U(1)_Y$ where Y represents the weak hypercharge, which depends on the electric charge Q and the third component of isospin T_3, according to the equation:

$$Q = \frac{Y}{2} + T_3 \, . \tag{2.13}$$

The $SU(2)_L \times U(1)_Y$ gauge symmetry group has four generators, which can be identified with the bosons W^1, W^2, W^3 for $SU(2)_L$ and B for the $U(1)_Y$. As discussed with the electromagnetic interaction, local gauge invariance can be restored by defining an appropriate covariant derivative D^μ, defined as:

$$D^\mu = (\partial_\mu + igT_i W^i_\mu + \frac{ig'}{2} Y B_\mu) \, , \tag{2.14}$$

where T_i are the three isospin components and Y corresponds to the hypercharge operator. Thus, the interaction term of the electroweak Lagrangian can be expressed as:

$$\mathcal{L}_{EW} = -\frac{g'}{2}(\bar{\psi}\gamma^\mu\psi)B_\mu - g(\bar{\psi_L}\gamma^\mu T_i \psi_L)W^i_\mu \, . \tag{2.15}$$

The $SU(2)_L \times U(1)_Y$ generators cannot be directly associated with the photon and the weak interaction carriers. In fact, the B boson interacts with both components of the weak isospin doublet (e_L and ν_L), while photons do not couple to neutral particles. Moreover, the W^i coming from $SU(2)_L$ are not the physical weak interaction fields.

The bosons for the electromagnetic interaction and the neutral weak current interaction arise from the mixing of the W^3 and B fields:

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} , \tag{2.16}$$

where θ_W is the weak-mixing angle, known as Weinberg angle, with $\sin 2\theta_W = 0.23122 \pm 0.00003$ [23].

This also implies that the $W^\pm$ bosons originate from a linear combination of W^1 and W^2 :

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) . \tag{2.17}$$

Similarly, the electric charge e is dependent on the weak mixing angle and the coupling constants of the W^3 and B fields:

$$g \sin\theta_W = g' \cos\theta_W = e . \tag{2.18}$$

Finally, the coupling constant for the weak neutral current, mediated by the Z boson, can be expressed as:

$$g_Z = \frac{e}{\sin\theta_W \cos\theta_W}(T_3 - Q\sin^2\theta_W) . \tag{2.19}$$

2.2. The physics of the Higgs boson

2.2.1. The Higgs mechanism

The Standard Model Lagrangian, as discussed so far, describes vector bosons and fermions that must be necessarily massless to preserve local gauge invariance. However, it is known that the weak interaction mediators are massive, as well as most of the fermions. The problem of introducing masses into the Standard Model while preserving local gauge invariance was solved by introducing the *Higgs mechanism*. This solution was formulated by theoretical

physicists Higgs, Brout, Englert, Guralnik, Hagen and Kibble in 1964 [24] [25] [26]. Here, an example [27] is given to show how this mechanism works. Consider a complex scalar field:

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \, , \tag{2.20}$$

coupled to an electromagnetic field and itself, described by the Lagrangian:

$$\mathcal{L} = -\frac{1}{4}(F_{\mu\nu})^2 + |D_\mu \phi|^2 - V(\phi) \, , \tag{2.21}$$

with $D_\mu = \partial_\mu + ieA_\mu$. This Lagrangian is invariant under $U(1)$ local transformation

$$\phi(x) \to e^{i\alpha(x)}\phi(x) \qquad A_\mu(x) \to A_\mu(x) - \frac{1}{e}\partial_\mu \alpha(x) \, , \tag{2.22}$$

if the potential $V(\phi)$ is chosen to be of the form:

$$V(\phi) = -\mu^2 \phi^* \phi + \frac{\lambda}{2}(\phi^* \phi)^2 \, . \tag{2.23}$$

The λ parameter drives the behaviour of the potential at $\pm\infty$ and it is bounded to be real to ensure the vacuum stability. When $\mu^2 > 0$, the field ϕ acquires a non zero vacuum expectation value (VEV). The minimum of the potential occurs at

$$\langle \phi \rangle = \phi_0 = e^{i\theta} \left(\frac{\mu^2}{\lambda} \right)^{1/2} \, , \tag{2.24}$$

and this results in the spontaneously breaking of the $U(1)$ symmetry. There are now infinite states in the circle of minimum potential, and the field ϕ acquires a non zero VEV, moving along an arbitrary direction. The particular minimum of the chosen potential is not symmetric anymore under $U(1)$ transformations.

For the sake of simplicity, θ can be taken equal to zero, hence $\phi(x)$ can be rewritten as:

$$\phi(x) = \phi_0 \frac{1}{\sqrt{2}}(\phi_1(x) + i\phi_2(x)) \ . \tag{2.25}$$

Now the potential 2.2.1 becomes:

$$V(\phi) = -\frac{1}{2\lambda}\mu^4 + \frac{1}{2}\dot{2}\mu^2\phi_1^2 + \mathcal{O}(\phi_i^3) \ , \tag{2.26}$$

so that the field ϕ_1 acquires a mass $m = \sqrt{2}\mu$ while the field ϕ_2 is the massless Goldstone boson. The Goldstone theorem, in fact, states that given a Lagrangian that is invariant under a group of continuous transformations with N generators, if M out of N symmetries are spontaneously broken, then there will be M massless bosons in the particle spectrum of the theory. Given the lack of empirical observations, the Goldstone bosons arising with the spontaneous symmetry breaking cannot appear as independent physical particles in the theory. This can be obtained using the local $U(1)$ gauge symmetry (2.2.1). In fact, it is possible to choose the $\alpha(x)$ parameter in such a way that the field $\phi(x)$ becomes real (*unitary gauge*), resulting in the cancellation of the ϕ_2 terms. With this choice, the Lagrangian becomes:

$$\mathcal{L} = -\frac{1}{4}(F_{\mu\nu})^2 + (\partial_\mu\phi)^2 + e^2\phi^2 A_\mu A^\mu - V(\phi) \ . \tag{2.27}$$

The Lagrangian now includes a mass term for the gauge boson that is different from zero if the potential $V(\phi)$ has a non zero VEV. The degree of freedom of the disappeared Goldstone boson has been absorbed by the massive field, acquiring a longitudinal polarisation.

This simplified example starts with a $U(1)$ gauge-invariant Lagrangian. A particular potential with a non-zero VEV is added, and the Lagrangian is rewritten, expanding the potential around its minimum. This leads to a Lagrangian that is not $U(1)$ gauge invariant anymore, in which the gauge bosons have acquired a mass and a longitudinal polarisation. This mechanism, generalized to non-Abelian models, is known as *the Higgs mechanism*.

2.2.2. The Standard Model Higgs Boson

The Higgs mechanism can be used to introduce mass terms to the gauge bosons in the Lagrangian of the Standard Model. The local gauge symmetry group to start from is $SU(2) \otimes U(1)$, and the theory, after spontaneous symmetry breaking, needs to include the QED unbroken $U(1)$ symmetry as well as a massless photon field. The vacuum expectation value,

therefore, has to be invariant under $U(1)$. This can be achieved via the Higgs mechanism by introducing a complex doublet of scalar fields:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} , \tag{2.28}$$

with the Lagrangian:

$$\mathcal{L} = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 . \tag{2.29}$$

Typically the potential ground state (see Section 2.2.1) is chosen to be:

$$\phi_1 = \phi_2 = \phi_4 = 0 \qquad \phi_3 = v \qquad v = \left(\frac{-\mu^2}{\lambda} \right)^{1/2} \qquad \phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{2.30}$$

It is then possible to write the field ϕ as an expansion from the vacuum expectation value:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} . \tag{2.31}$$

Notably, adopting the unitarity gauge, $H(x)$ is a real scalar field.
The gauge fields $W_\mu^{1,2,3}$ and B_μ can be written as mass eigenstates in terms of $W^\pm$, Z^0 and A fields. Combining the gauge Lagrangian $SU(2)_L \otimes U(1)_Y$ together with the scalar Lagrangian L_ϕ, the theory becomes:

$$\begin{aligned} \mathcal{L}_\phi \mathcal{L}_{gauge} =& \frac{1}{2} \partial_\mu H \partial^\mu H + \mu^2 H^2 - \lambda v \sigma^3 - \frac{\lambda}{4} H^4 + \\ &+ \left(\frac{gv}{2} \right)^2 W_\mu^+ W^{-\mu} + \frac{1}{2} \left(\frac{g^2 + g'^2}{4} \right) v^2 Z_\mu Z^\mu + \\ &+ \frac{1}{2} g^2 v W_\mu^+ W^{-\mu} H + \frac{g^2 v}{2} \frac{1}{2\cos^2 \theta_W} Z_\mu Z^\mu H + \\ &+ \frac{g^2}{4} W_\mu^+ W^{-\mu} H^2 + \frac{g^2 v}{2} \frac{1}{2\cos^2 \theta_W} Z_\mu Z^\mu \sigma^2 . \end{aligned} \tag{2.32}$$

The first line provides the kinetic and the self-interaction terms for the scalar field H. The second line presents the mass terms of the W and Z fields. The third and fourth lines describe the coupling between the Higgs field and the gauge bosons.
From the Lagrangian, the masses of the W and Z vector bosons are given by:

$$m_W = \frac{gv}{2} = 80.379 \pm 0.012 \text{ GeV} \qquad m_Z = \frac{v\sqrt{g^2+g'^2}}{2} = 91.1876 \pm 0.0021 \text{ GeV} \ \ [23], \tag{2.33}$$

and the field $H(x)$ corresponds to scalar particle, with mass:

$$m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2}\,. \tag{2.34}$$

It emerges that the scale of the Higgs mass is controlled by the VEV of the H field, which is calculated to be $\sim$ 246 GeV [23], but its precise value is a free parameter of the theory, depending on the new coupling λ. This result has important phenomenological consequences, implying that the SM does not predict the mass range of the Higgs boson.
From the interaction terms in Eq. 2.2.2, it can be observed that the coupling between the Higgs and gauge vector bosons depends on the W and Z masses.

Fermion mass terms do not derive directly from the symmetry breaking mechanism. However, with the addition of a complex Higgs doublet in the Langrangian, it is possible to introduce a new term of the form:

$$\mathcal{L}_f = -\lambda_f[\bar{\psi_L}\phi\psi_R + \bar{\psi_R}\phi^\dagger\psi_L]\,, \tag{2.35}$$

which is manifestly invariant under $SU(2)_L \times U(1)_Y$. The λ_f term, known as the Yukawa coupling constant, is a free parameter, independent for each fermion. Considering for simplicity only the electronic flavour and the scalar field ϕ, expressed in unitarity gauge as in Eq. 2.2.2, then Eq. 2.2.2 becomes

$$\begin{aligned}\mathcal{L}_e &= -\frac{\lambda_e}{\sqrt{2}}\left[(\bar{\nu},\bar{e})_L\begin{pmatrix}0\\ v+H\end{pmatrix}e_R + \bar{e}_R(0,v+H)\begin{pmatrix}\nu\\ e\end{pmatrix}_L\right] \\ &= -\frac{\lambda_e v}{\sqrt{2}}\bar{e}e - \frac{\lambda_e}{\sqrt{2}}H\bar{e}e\,,\end{aligned} \tag{2.36}$$

where the first term gives the mass of the electron, while the second describes the interaction between the electron and the Higgs field. The Higgs coupling to the fermions is, therefore, proportional to the mass of the fermion itself.

Since the ϕ field is an $SU(2)$ doublet, a gauge-invariant coupling with the left-handed lepton doublet E_L and the right-handed singlet e_R is included in the Lagrangian as

$$-\lambda_e(\bar{E}_L\phi)e_R + \text{h.c.}\,, \tag{2.37}$$

where h.c. represents the Hermitian conjugate of the terms already present and λ_e is a new dimensionless coupling constant [27]. The new term in the Lagrangian leads to a lepton mass of the form $m_e = \lambda_e v/\sqrt{2}$. Similarly, the equivalent $SU(2)$ invariant coupling for the left-handed quark doublet Q_L and right-handed singlets d_R and u_R is included in the Lagrangian as

$$-\lambda_d(\bar{Q}_L\phi)d_R - \lambda_u\epsilon^{ab}(\bar{Q}_L\phi_b^\dagger)u_R + \text{h.c.}\,, \tag{2.38}$$

where ϵ^{ab} is a two-dimensional Levi-Civita tensor [27]. Following the lepton discussion, the quark fields acquire mass terms with $m_d = \lambda_d v/\sqrt{2}$ and $m_u = \lambda_u v/\sqrt{2}$. Specific to the quark sector is the mixing between generations of up and down-type quarks, with couplings determined by the 3×3 unitary Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix [28, 29]. In general in the unitarity gauge, for any quark or lepton flavor, the Higgs boson couples according to:

$$\mathcal{L}_f = -m_f\bar{f}f\left(1 + \frac{H}{v}\right)\,. \tag{2.39}$$

As neutrinos do not have a right-handed chiral state, no couplings to the h-field can be included without breaking a gauge symmetry. Therefore, the neutrinos are massless in the SM. However, the neutrino flavours oscillate and, therefore, the presence of neutrino mass eigenstates has been measured experimentally and included ad-hoc into the SM through the Pontecorvo-Maki-Nakagawa-Sakata matrix [30, 31], analogous to the CKM matrix.

2.3. Standard Model precision measurements

Despite its success, the Standard Model does not provide a full description of all the phenomenological observations. It does not include a description for the gravitational interaction and does not provide a candidate for dark matter. These problems suggest the existence of BSM physics, and the search for particles outside the SM is one of the main physics goals that drive the experimental programme of colliders. Besides exploring the direct production

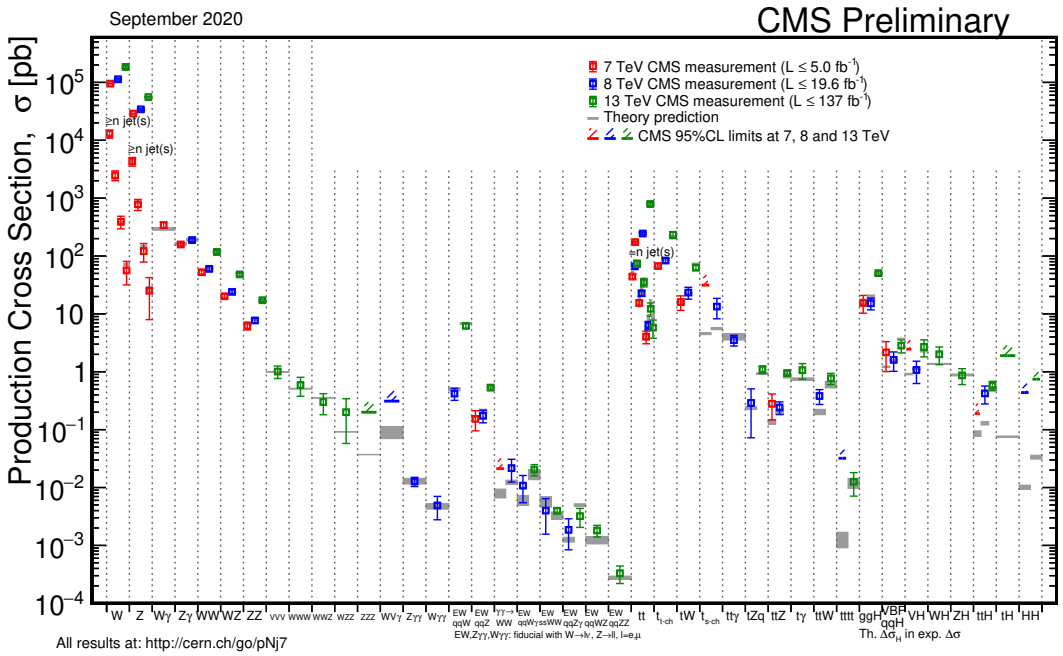


Figure 2.2.: Overview of Standard Model cross-section measurements performed by the CMS collaboration. The red (blue) points are measurements from Run 1 at 7 TeV (8 TeV). The green points refer to Run 2 measurements, and the grey boxes are the theoretical predictions for the different processes cross-section. The list of measurements is updated to September 2020 [32].

of BSM, a complementary approach is provided by precision testing of SM predictions. Precision measurements' impact on new physics is two-fold, as improving predictions on SM processes provides a stronger constraint on BSM backgrounds, but also the presence of new physics can be indirectly tested by observing deviations (or lack of them) from expected SM behaviour. Moreover, theoretical predictions are based on perturbation theory and to limit the computational cost, higher-order interactions are neglected. However, with the increase in the centre-of-mass energy and statistical significance of the collected data, the effects of higher-order processes can be explored, especially targeting regions of the parameter space where such corrections dominate and are nowadays accessible. Therefore, a wide range of SM measurements is carried out at colliders, exploiting the large datasets collected at 13 TeV during LHC Run 2. Fig.2.2 offers a general overview of SM measurements performed by the CMS collaboration,

2.3.1. V + jets measurements

Among the different SM processes explored at the LHC, events where vector bosons are produced in association with one or more jets (V + jets) are of particular interest for their versatility and relatively large cross-section. They provide benchmarks for precision tests of

the SM, constraining Parton Distribution Functions (PDFs), probing higher-order perturbative QCD and also contributing as dominant backgrounds to many BSM searches.

PDF fits variously incorporate V + jets measurements. The gluon PDF can be constrained by γ + jets and low mass Drell-Yan +jets measurements, with the first process being directly sensitive to the gluonic content of the proton via QCD Compton scattering. Measurements involving vector boson production with heavy flavour jets are used to probe the sea content of the proton. In particular, the Z+b and Z+c measurements and especially their ratio are sensitive to the charm quark contribution, and the measurement of W+c production is directly sensitive to the strange content of the proton.

V + jets production can also proceed via vector-boson-fusion (VBF). This topology results in two energetic jets in the forward region which can be used as tags but, of course, is characterized by a lower cross-section. The large amount of data collected during Run 2 of the LHC provides an opportunity to perform measurements of this kind. These rare processes, involving a triple boson vertex, are key measurements to spot possible deviations from SM predictions induced by BSM in the triple gauge coupling. Moreover, studying additional hadronic activity in the rapidity gap between the VBF tagging jets provides an interesting test for perturbative QCD calculations.

Finally, typical BSM processes that can benefit from precision V + jets measurements involve missing transverse momentum in the final state. This is the case for searches for supersymmetry and dark matter production. The dominant irreducible background in such searches comes from the SM $Z \to \nu\bar{\nu}$ process, and V + jets processes constitute fundamental cross-measurements to constrain it.

All these aspects make V + jets processes compelling measurements to pursue at the LHC, exploiting the large amount of data collected at 13 TeV. A summary of V + jets published measurements by the CMS collaboration is provided in Fig. 2.3, where the large number of green points correspond to measurements performed with Run 2 datasets, testifying to the active interest in studying these processes.

2.3.2. Higher-order electroweak corrections

The increased centre-of-mass energy and large dataset collected in the LHC Run 2 presents a great opportunity to perform ever more precise measurements of V + jets processes and do so in regions of phase space previously limited by statistics.

Performing V + jets measurements in particularly extreme kinematic regions poses a challenge to theoretical prediction as higher-order diagrams, which at low energy are suppressed and consequently neglected in theoretical calculations, start playing a role, modifying the

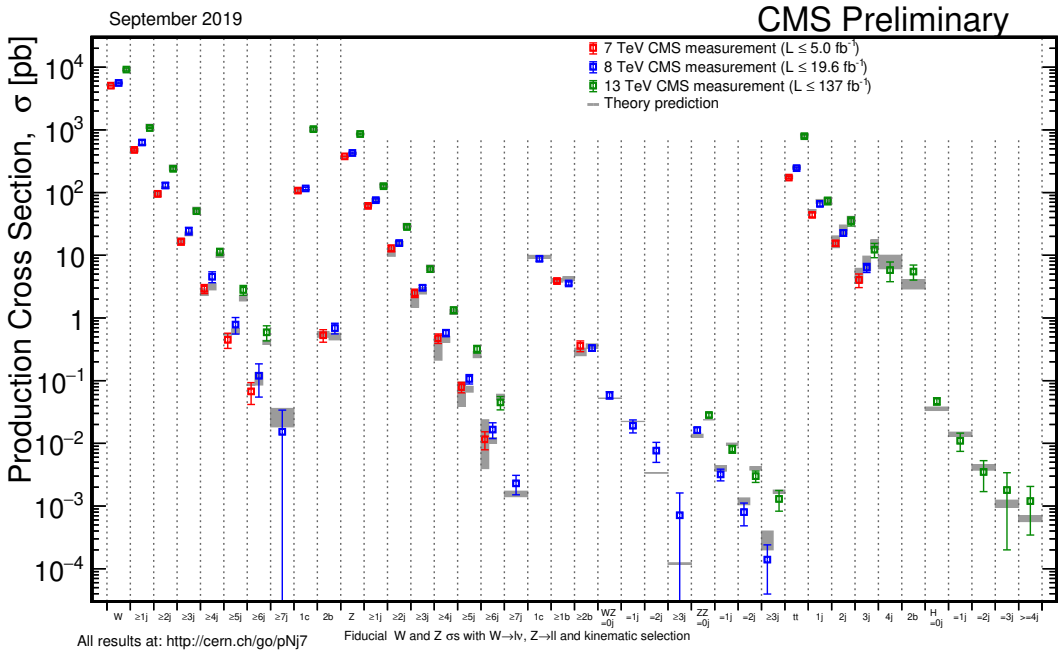


Figure 2.3.: Overview of Standard Model cross-section measurements of V + jets processes performed by the CMS collaboration. The red (blue) points are measurements from Run 1 at 7 TeV (8 TeV). The green points refer to Run 2 measurements, and the grey boxes are the theoretical predictions for the different processes cross-section. The list of measurements is updated to September 2020 [32].

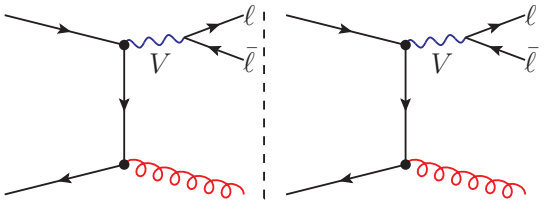
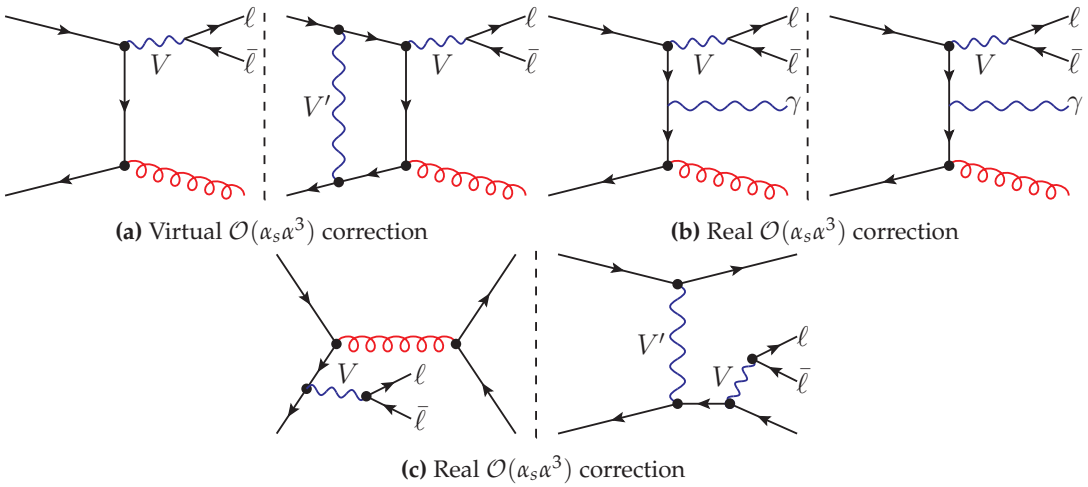


Figure 2.4.: Representative LO contribution to V + 1 jet production.

differential cross-sections. This is the case for higher-order electroweak corrections, which are typically suppressed by the large vector boson masses.

In particular, the differential cross-section measurement of the Z + jets and γ + jets processes can constrain higher-order perturbative QCD and EW calculations that lead to a non-negligible dependence of the cross-section on boson p_{T}. The weak radiative corrections become large and negative at high energies due to the presence of Sudakov logarithms that result from the virtual exchange of soft or collinear massive gauge bosons [6, 7, 8, 9, 10, 11]. The corrections have the form $\alpha \log^2(\hat{s}/M^2_{\mathrm{W,Z}})$, where $\hat{s}$ is the squared parton centre-of-mass energy, α is the gauge coupling of the unified EW theory, and $M_{\mathrm{W,Z}}$ is the mass of the W or Z boson. Figure 2.4 and 2.5 show the Feynman diagrams for LO production and NLO EW corrections to Z + jets and γ + jets proceses.



(a) Virtual $\mathcal{O}(\alpha_s\alpha^3)$ correction

(b) Real $\mathcal{O}(\alpha_s\alpha^3)$ correction

(c) Real $\mathcal{O}(\alpha_s\alpha^3)$ correction

Figure 2.5.: Representative virtual and real NLO EW contributions to V + 1 jet production [33].

The impact of the perturbative corrections has been extensively discussed in the context of measurements and searches targeting high p_{T} regimes (for example searches for dark matter [34]), as the dependence of the EW corrections on the boson momentum can lead to large effects on the order of tens of per cent for large boson p_{T}.

Furthermore, developments in theoretical calculations have led to improved predictions with automated next-to-leading order (NLO) QCD and EW corrections, for instance, SHERPA v2.2.8 + OPENLOOPS v2.0 [35, 36] and MADGRAPH5_aMC@NLO v2.2.2 [37]. Measuring the production cross-section of $\mathrm{Z}+$ jets, γ + jets, and their ratio at high transverse momentum is valuable for probing the magnitude and the dependence of these higher-order corrections on boson p_{T}, playing an important role in validating the latest theoretical calculations. A prediction of the Z/γ ratio p_{T} distribution obtained with the SHERPA v2.2.8 + OPENLOOPS v2.0 software is shown in Fig. 2.6. It can be observed how higher-order EW corrections grow with p_{T}^{V}, and their impact becomes dominant over QCD ones at very high p_{T}. It is also worth noticing that the default approach used in these predictions is to combine EW and QCD corrections additively. The validity of such approach can be tested by sensitive measurements, which can gauge potentially large effects due to the interplay of EW and QCD corrections beyond NLO.

The CMS Collaboration has previously performed a differential measurement of the Z/γ cross-section ratio at $\sqrt{s} = 8\,\mathrm{TeV}$ using a data set corresponding to an integrated luminosity of $19.7\,\mathrm{fb}^{-1}$ [38].

In contrast to quantum electrodynamics and QCD corrections, where the massless gauge bosons lead to logarithms that are cancelled by the corresponding real-emission corrections, the masses of the W and Z gauge bosons act as infrared regulators, providing a physical cutoff. The real emission of a W or Z boson can contribute significantly to inclusive $\mathrm{W}+$ jets

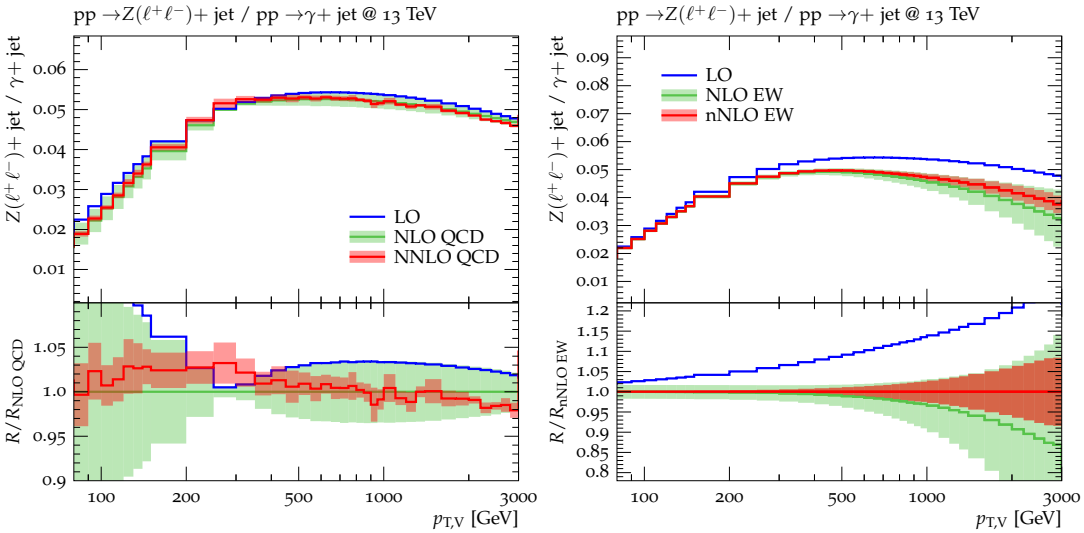


Figure 2.6.: Ratios of p_{T} distributions for $pp \to Z/\gamma$+jet processes at LO, NLO and NNLO QCD (a) and LO, NLO EW and nNLO EW accuracy (b). While higher-order QCD corrections remain flat in the high boson p_{T} region and their overall impact is less than 5%, EW corrections grow with p_{T}^{V} reducing LO cross section by 20% at 2 TeV [34].

and Z + jets measurements at high energies [13, 14, 15]. These processes can be accessed by selecting a boosted event topology and studying the region of small angular separation between a W or Z boson and a jet. The fully reconstructable decay products from the Z boson allow a direct measurement of the angular separation between the Z boson and the closest jet. The ATLAS Collaboration has previously performed a study of the real emission of the W boson at 8 TeV [16] and CMS at 13 TeV [17].

Chapter 3.

LHC and the CMS detector

The data used to perform the analysis presented in this thesis were collected by the CMS detector during the 2016 data taking period of the LHC accelerator, operating at a centre-of-mass energy of 13 TeV.
In this chapter, after a brief historical introduction, an overview of the physics of the events produced at the LHC is given, together with a description of the different systems comprising the LHC complex, focusing in particular on the CMS detector and its sub-detectors.

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [40] is the largest and most powerful particle accelerator in the world. It has been built at *Conseil Européen pour la Recherche Nucléar* (CERN), approximately 100m underground across the Swiss-French border. Fig. 3.1 shows an overview of the LHC site. The project was approved in 1994 by the CERN council and became operative in September 2008. However, nine days after the first beam circulation, an accident occurred due to an electrical fault and 53 magnets were damaged as a consequence [41]. This caused a one year delay in physics operations to repair the damages and improve the safety conditions. During the autumn of 2009, the operations started again, and all four CERN experiments recorded the first physics events on November 30th.

The first period of data taking (Run1) ended in 2012, reaching a 7 TeV centre-of-mass energy for the 2010-2011 runs and 8 TeV for 2012. Thanks to the data collected in this period, the Higgs boson has been discovered, and strong constraints on a range of BSM models have been placed. After this running period, the LHC faced two years of shutdown, and in this time, all the accelerator apparatuses have been upgraded to reach a centre-of-mass energy of 13 TeV.

The first physics beams of the second run period (Run2) collided on April 4th of 2015. Operations continued till the end of 2018. An unprecedented luminosity from both protons

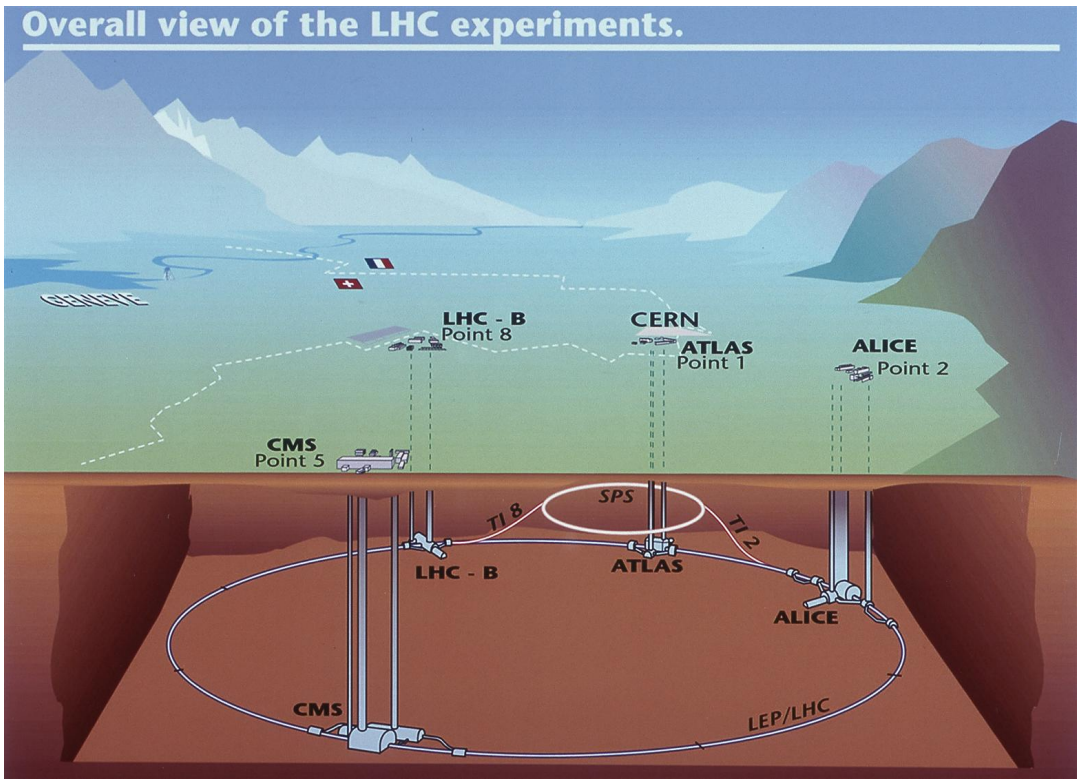


Figure 3.1.: An overview of the LHC complex at CERN [39].

and hadron collisions has been delivered to the different experiments. The evolution of the luminosity delivered by the LHC during Run2 is shown in Fig 3.3a.

3.1.1. Hadronic interactions

The LHC apparatus is situated in the tunnel built in the late 1980s to accommodate the LEP (Large Electron-Positron collider) experiment [42]. Contrary to its predecessor, the LHC has been designed to be a discovery machine. Its physics reach aiming to investigate the nature of particle interactions up to the TeV scale, studying the electroweak symmetry breaking mechanism and testing physics beyond the Standard Model; the choice to accelerate and collide hadrons is strictly related to these goals. Any charged particle, in fact, radiates photons when accelerated, and in particular, the rate of energy loss for a particle that covers a circular orbit with a radius R depends on its mass m in the following way:

$$\frac{dE}{dt} \propto \frac{E^4}{m^4 R} \,. \tag{3.1}$$

For this reason, to produce high energy beams, protons that are two hundred times heavier are used instead of electrons. Furthermore, protons are not elementary objects, as they are

made of quarks and gluons. Therefore, during a collision, the scattering occurs between two of the elementary components, each coming with a random fraction of the total momentum. The energy in the centre-of-mass of each collision is not fixed by the beam energy, giving hadron colliders the capability of exploring a broad region of the phase space.

Two classes of collision events are defined:

- *Soft Collisions.* The most frequent events are generated by protons that do not come close enough to each other to let their own substructure interact. Typically, such collisions are characterised by low momentum transfer, and the particles in the final state are quite collinear to the beam axis.
- *Hard Collisions.* The hard collisions rarely happen compared to the rate of soft interactions, and they are the most interesting from the experimental point of view. These types of interactions happen at a short distance, and the collision involves large momentum transfer. The final state can be characterised by particles at a large angle, high transverse momentum and high mass.

In the hard scattering, the protons' structure can become evident as the collision can be considered as happening between two partons. Therefore, the centre-of-mass energy depends on the fraction of momentum carried by the partons and it is lower than that reached nominally by the accelerator. The parton centre-of-mass energy $\sqrt{\hat{s}}$ is described by:

$$\sqrt{\hat{s}} = \sqrt{x_a x_b s}\,, \tag{3.2}$$

where $\sqrt{s}$ is the centre-of-mass energy reached by the accelerator and x_a , x_b are the fraction of momentum carried by the scattering partons. The general cross section for a hard-scattering interaction is:

$$\sigma = \sum_{a,b} \int dx_a dx_b f_a(x, Q^2) f_b(x, Q^2) \hat{\sigma}_{ab}(x_a, x_b)\,, \tag{3.3}$$

where $\hat{\sigma}_{ab}$ represents the cross-section of elementary interaction between two partons and $f_a(x, Q^2)$, $f_b(x, Q^2)$ are the Parton Distribution Functions (PDFs), which are the probabilities at a given transferred momentum (Q^2) to find a specific parton inside the proton, carrying a particular fraction x of the total proton momentum. Figure 3.2 shows the evolution of the PDFs for a proton with different Q^2 values.

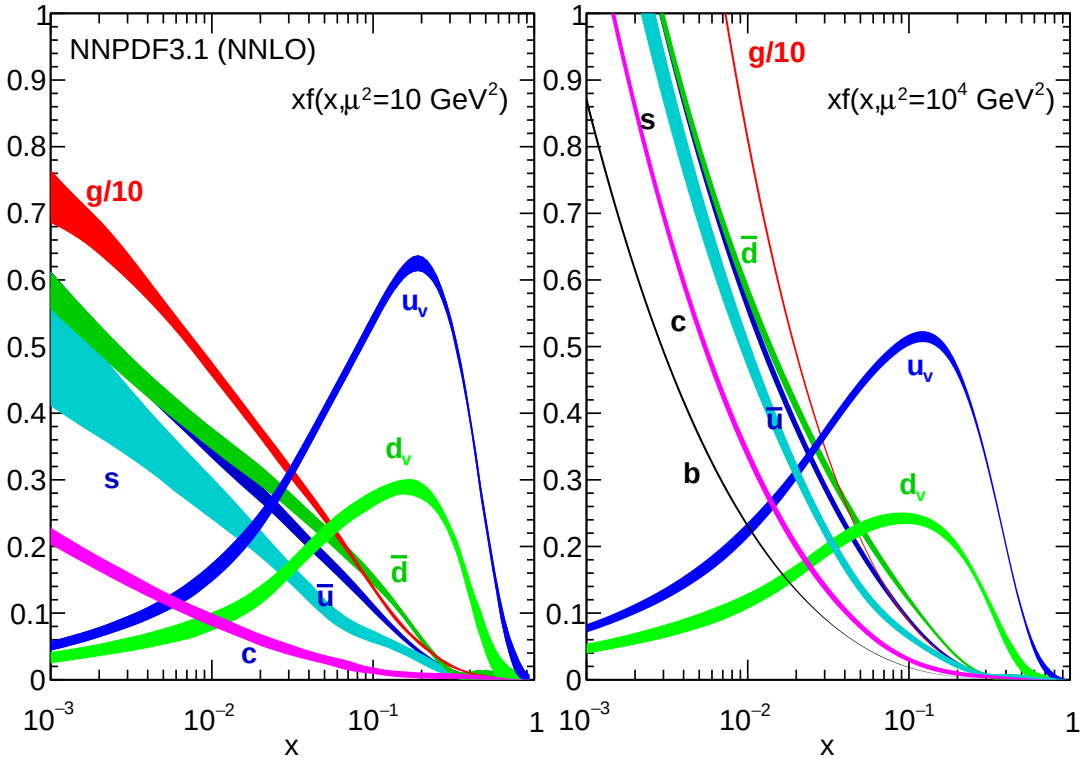


Figure 3.2.: Parton Distribution Functions (PDFs) computed for $Q^2 = 10\,\text{GeV}^2$ (on the left) and $Q^2 = 10$ TeV^2 (on the right), the probability of interacting with sea quarks carrying a good fraction of the proton momentum grows with the transferred momentum [43].

3.1.2. Luminosity and pile-up

In a particle accelerator, the rate of events of a process with a cross-section σ is given by:

$$\frac{dN_{event}}{dt} = \mathcal{L}\sigma_{event} \ , \tag{3.4}$$

where $\mathcal{L}$ the instantaneous luminosity which depends only on the beam parameters and can be written for beams with a Gaussian profile:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n \beta*} F[\mathrm{cm}^{-2}\mathrm{s}^{-1}] \ , \tag{3.5}$$

where (nominal parameters for the 2016 campaign are given in parenthesis)[44]:

- N_b is the number of particles per bunch ($\sim 1.1 \times 10^{11}$)
- n_b is the number of bunches per beam (2200)
- f_{rev} is the revolution frequency (11.236 KHz)
- ϵ_n is the normalized transverse beam emittance (2.2 μm)
- β^* is the beta focusing function at the collision point in CMS and ATLAS (40 cm)
- F is the geometric luminosity reduction factor due to the crossing angle at interaction point (0.836)

The LHC's luminosity is not constant over a physics run but reduces due to the degradation of intensity and emittance of the circulating beams. The main cause for the luminosity decay at the LHC is the collisions themselves. The behaviour of the luminosity as a function of time is given by :

$$\mathcal{L}(t) = \frac{\mathcal{L}}{(1 + t/\tau_{nuclear})^2} \ , \tag{3.6}$$

where

$$\tau_{nuclear} = \frac{N_{tot,0}}{\mathcal{L}\sigma_{tot}k} \ , \tag{3.7}$$

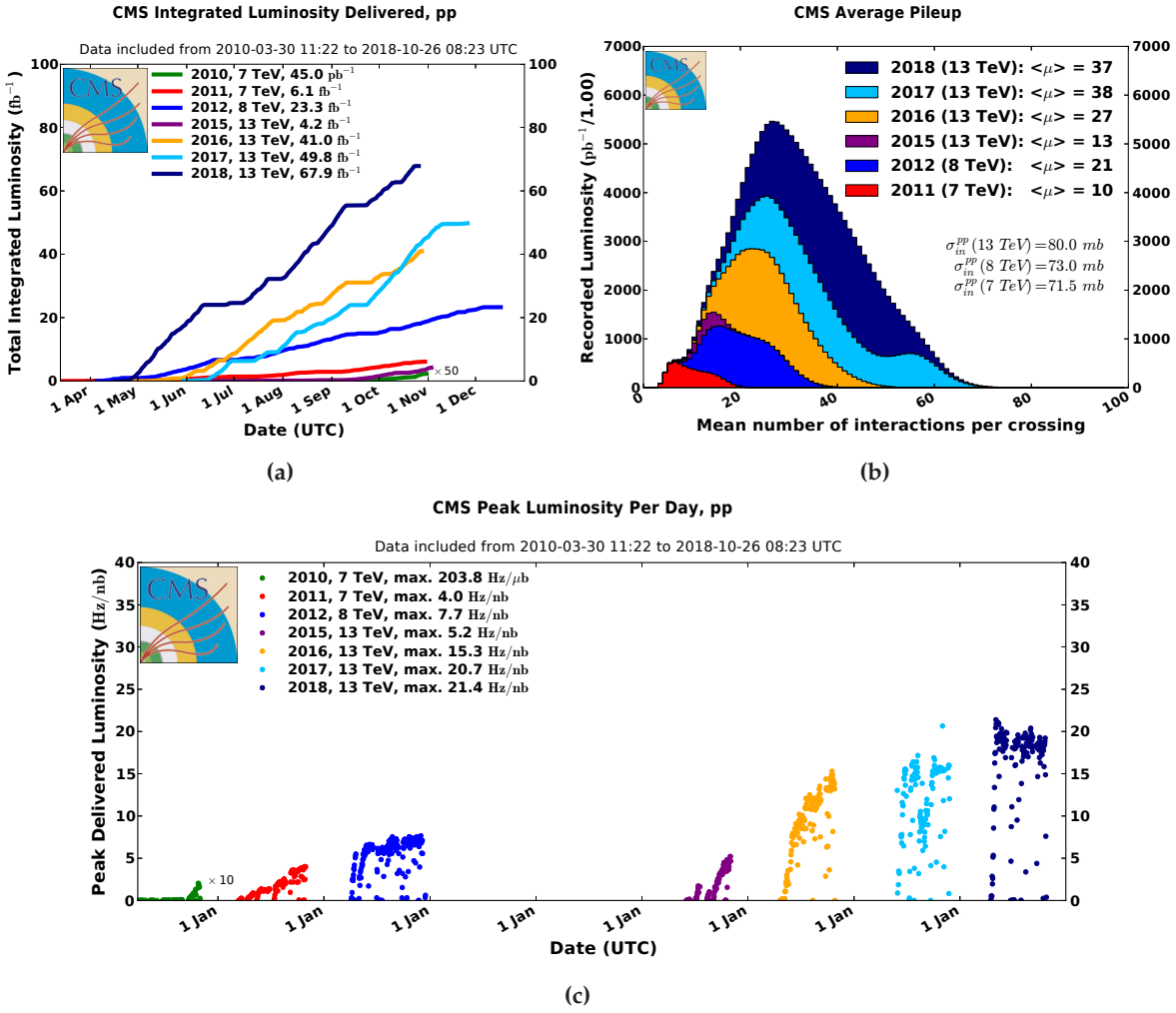


Figure 3.3.: On the top left (a) is displayed the integrated luminosity versus time recorded by CMS in the different data-taking periods. On the top right (b) the distribution of the number of interactions per bunch crossing for each data taking period. The bottom plot (c) depicts the daily peak luminosity for the different years [45].

with $N_{tot,0}$ the initial beam intensity, $\mathcal{L}$ the initial luminosity, σ_{tot} the total pp cross section and k the number of interaction points (4).
Total number of collisions in a given period of time is given by the integrated luminosity $\int \mathcal{L}dt$.

In Figure 3.3, the integrated luminosity, the average number of interactions per bunch crossing and peak luminosity for the different data-taking periods are summarized, while in Figure 3.4, the integrated luminosity collected and the pileup conditions for 2016 are shown.

Since the bunches of protons contain $\sim 10^{11}$ particles, at each bunch crossing multiple $p-p$ interactions can occur. This effect is known as *pile-up* and deteriorates the energy resolution of the object (leptons, photons, jets, E_T^{miss} , etc.) reconstruction, thus making it more difficult

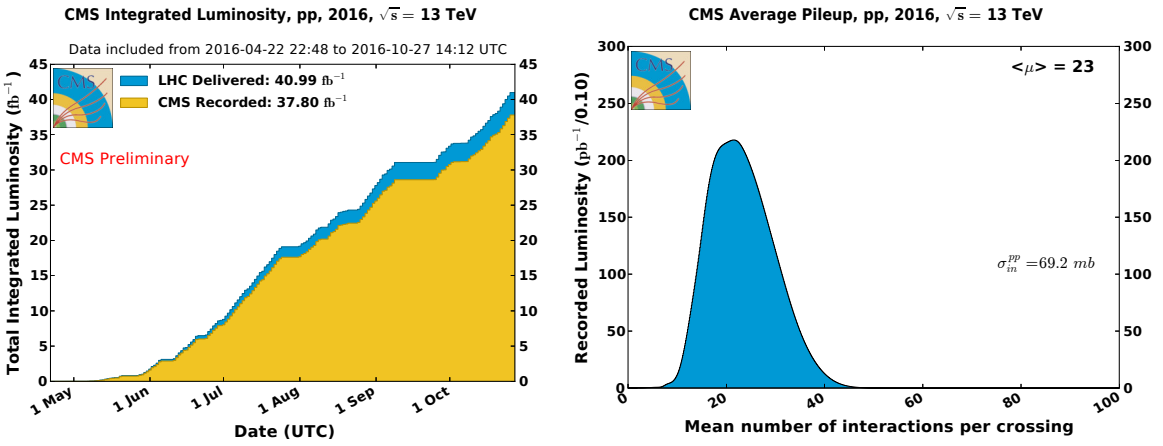


Figure 3.4.: On the left (a) the integrated luminosity versus time delivered to CMS (blue) and recorded by CMS (yellow) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2016. On the right, the distribution of the number of interactions per bunch crossing for the same data taking period [45].

to reconstruct vertices and identify tracks due to the large number of hits in the tracker. The pile-up and the instantaneous luminosity are related by:

$$\mathcal{L} = \frac{rate_{inel}}{\sigma_{inel}} = \frac{\mu \cdot n_b \cdot f_{rev}}{\sigma_{inel}} \,, \tag{3.8}$$

where μ is the number of inelastic interactions per bunch crossing and n_b is the number of bunches. The number of pile-up interactions per bunch crossing, μ, is therefore proportional to $\mathcal{L}/f_{rev}$ and increases with the peak luminosity.

3.1.3. Coordinates system

A right-handed system of coordinate is defined, with the centre of the detector as the origin. The x axis is oriented horizontally towards the LHC ring centre, the y axis points upwards along the vertical direction, and the z axis is directed anticlockwise along the beamline. The $x-y$ plane, therefore, is transverse to the beamline, and the projection of kinematic quantities on this plane, such as the transverse energy E_{T} or transverse momentum p_{T}, are used in hadron colliders as they are invariant under longitudinal Lorentz boosts.

Other useful variables are the azimuthal angle ϕ measured in the $x-y$ plane from the x axis and the polar angle θ, measured on the $y-z$ plane. To express the relativistic velocity of a

particle with respect to the lab frame, the rapidity y is defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_{\mathrm{L}}}{E - p_{\mathrm{L}}} \right) , \tag{3.9}$$

where E is the energy of the particle and p_{L} is the projection of the particle 3-momentum along the beam line. Despite not being Lorentz invariant, the rapidity variable has the property of preserving differences Δy under longitudinal boosts. However, measuring the total momentum of particles is challenging at colliders, and a more convenient variable, the pseudorapidity η, is defined as follows:

$$\eta = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_{\mathrm{L}}}{|\mathbf{p}| - p_{\mathrm{L}}} \right) , \tag{3.10}$$

where $|\mathbf{p}|$ is the 3-momentum magnitude. From this definition it is clear how $\eta \approx y$ for highly relativistic particles. Moreover, by substituting $p_{\mathrm{L}} = p \cos\theta$, 3.10 can be rewritten as:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) , \tag{3.11}$$

showing how η depends only on the angle between the particle trajectory and the beam axis, and therefore it is easier to estimate then y.

Finally, to measure the angular difference between objects the variable $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ is used.

3.1.4. The LHC complex

The LHC consists of a 27 km ring of superconducting magnets located in the underground tunnel at the border between Switzerland and France that was previously used by LEP. A schematic representation is shown in Figure 3.1. The LHC ring is divided into eight arch sections, each of which is 3 km long, and eight linear sections of 523m. The magnet system is composed of 1232 dipoles which can produce a bending magnetic field of up to 8.36 Tesla, 392 quadrupoles used to focus the beam and thousands of hexapoles, octopoles and decapoles which can correct the beam orbit. The proton beams inside the LHC travel in opposite directions in separate beam pipes. After having been accelerated to the energy limit of 6.5 TeV, they are made to collide at four different points. Each collision point has a different detector: these are ALICE, ATLAS, CMS and LHCb.

The LHC ring is only the final step of a complex acceleration chain (shown in Figure 3.5), starting from an ionized hydrogen source. Protons are separated from the electrons with strong

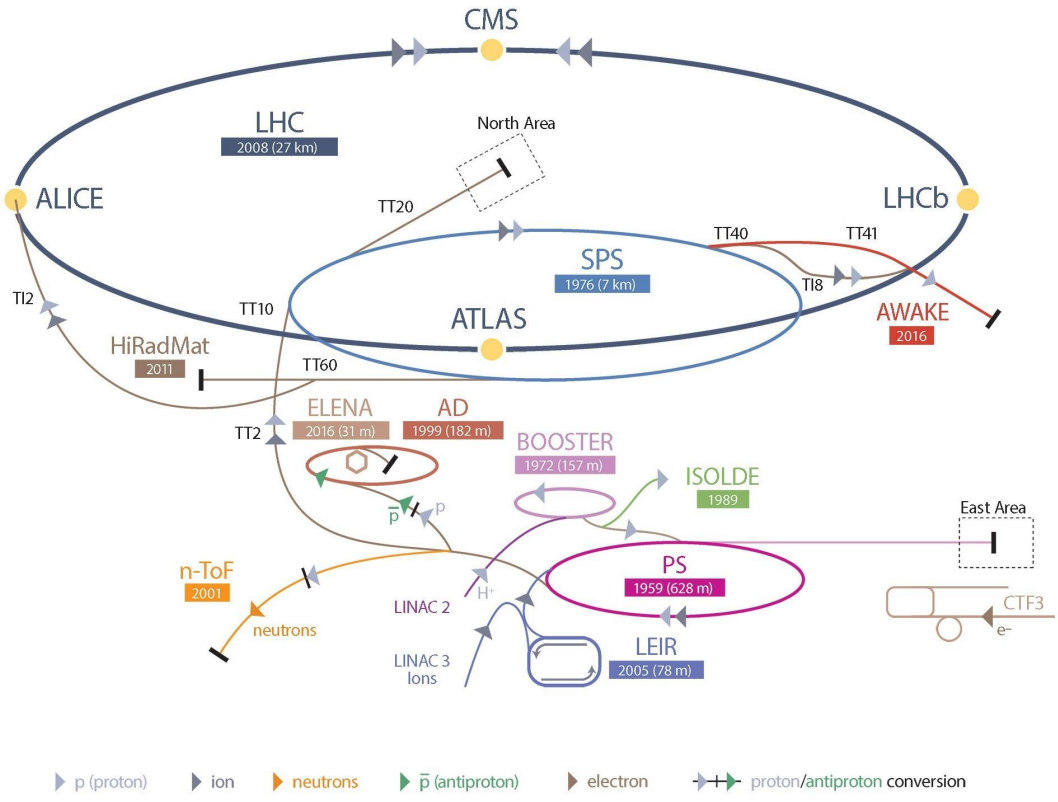


Figure 3.5.: The accelerator complex at CERN [39].

electric fields, and then they are accelerated up to 50 MeV by the linear accelerator called Linac2. The proton beam is then ready to be injected into the Proton Synchrotron Booster (PSB) and reaches an energy of 1.4 GeV. The next step is the Proton Synchrotron (PS), which delivers proton beams with an energy of 45 GeV to the Super Proton Synchrotron (SPS), the last step before the LHC. The SPS was a self-standing accelerator during the eighties, and thanks to the data collected by the UA1 and UA2 experiments, the W boson was discovered in 1983. Nowadays, it is part of the LHC accelerator scheme, and it further increases the energy of the beams up to 450 GeV before the injection into the main accelerator. Moreover, a significant fraction of the beams accelerated by the SPS is dedicated to other experiments, as depicted in Figure 3.5. Heavy ions follow a slightly different path, being accelerated by the Linac3 and by the Low Energy Ion Ring (LEIR) before entering the PS.

3.1.5. LHC experiments

The main LHC experiments (ALICE, CMS, ATLAS and LHCb) are located in caverns corresponding to the four collision points of the LHC, shown in Figure 3.1. Three additional experiments (TOTEM, LHCf and MoEDAL) focus on forward emitted particles and are located along the tunnel, near the main experiments. Before describing in detail the CMS detector, here is a brief description of the different LHC experiments:

- CMS [46] and ATLAS [47] are general-purpose detectors, designed to cover the widest possible range of physics in proton-proton and heavy-ion collisions.
- LHCb [48] specializes in investigating the differences between matter and antimatter by studying light states containing b quarks.
- ALICE [49] is dedicated to heavy-ion collisions to study the strong interactions of quarks and gluon plasma.
- TOTEM [50] makes precise measurements of protons in the forward region. It is located at the CMS interaction point, and covers pseudo-rapidity regions that are inaccessible by the main experiment. Similarly, the ATLAS experiment has a dedicated subdetector, ALFA [51]. Together they are used to monitor the LHC luminosity.
- LHCf [52] is dedicated to the measurement of neutral particles emitted in the very forward region. Particles from collisions at the ATLAS interaction point are used to simulate cosmic rays to provide calibration and help to interpret the results of cosmic-ray experiments.
- MoEDAL [53] This experiment is designed for searches of magnetic monopoles and Stable Massive Particles. It is located near the LHCb interaction point.

3.2. Compact Muon Solenoid (CMS) Experiment

As described in section 3.1.2, conditions provided by the LHC during 2016 data taking resulted in an average number of collisions per bunch crossing of approximately 25, leading to roughly 1000 charged particles every 25 ns bunch crossing. The CMS detector [54] was designed to operate in such a busy environment by using high-granularity subdetectors with good time resolutions allowing for low occupancy. This results in millions of detector electronic channels that require precise synchronization, all with a high radiation tolerance.

The main physics goals that have driven the design choices of the CMS detector, namely the search for the Higgs Boson, supersymmetric particles, new massive vector bosons, extra dimensions, precision studies of the SM, and heavy-ion physics, imply the following performance requirements:

- **Muon identification:** Good muon identification up to $|\eta| = 2.4$ and dimuon mass resolutions of about 1% for transverse momenta of 100 GeV, with an unambiguous charge assignation up to $p < 1$ TeV.
- **Charged particle reconstruction:** good momentum and reconstruction efficiency, particularly for the inner tracker, allowing for the identification of secondary vertices for the triggering and offline tagging of *tau* and *b*-jets. This requires a tracker component close to the interaction region.
- **Electromagnetic energy resolution:** measure the dielectron and diphoton mass with 1% resolution for transverse momenta of 100 GeV up to $|\eta| = 2.4$. Good rejection power of backgrounds from π^0 decays into photons and lepton isolation at high luminosities.
- **Transverse missing momentum and dijet mass resolution:** near-complete reconstruction of the proton interaction to measure the transverse missing momentum. This requires a hermetic detector with hadron calorimeters extending up to $|\eta| = 5.0$ and fine component segmentation in the η-ϕ plane of less than 0.1×0.1.

A complete layout of the entire CMS detector is shown in Fig. 3.6. The main feature that distinguishes the CMS detector is a 3.8 T superconducting solenoid magnet enclosing the tracking and calorimetry systems, making the largest magnet of this type ever made. All subdetectors are divided into two main regions in η, called the barrel (covering central η range, up to $|\eta| = 1.44$) and the endcap (covering up to $|\eta| = 3.0$, with a slight overlap with the barrel). The hadronic calorimeter extends its coverage to $|\eta| < 5$ with an additional segment in the forward region. This is required for improving the missing momentum reconstruction. The different subdetectors are further segmented within the barrel, endcap and forward regions. A discussion with further details is reported in the following subsections. This gives the CMS detector a total length of 21.6 m, a diameter of 14.6 m and a weight of 12 500 tonnes.

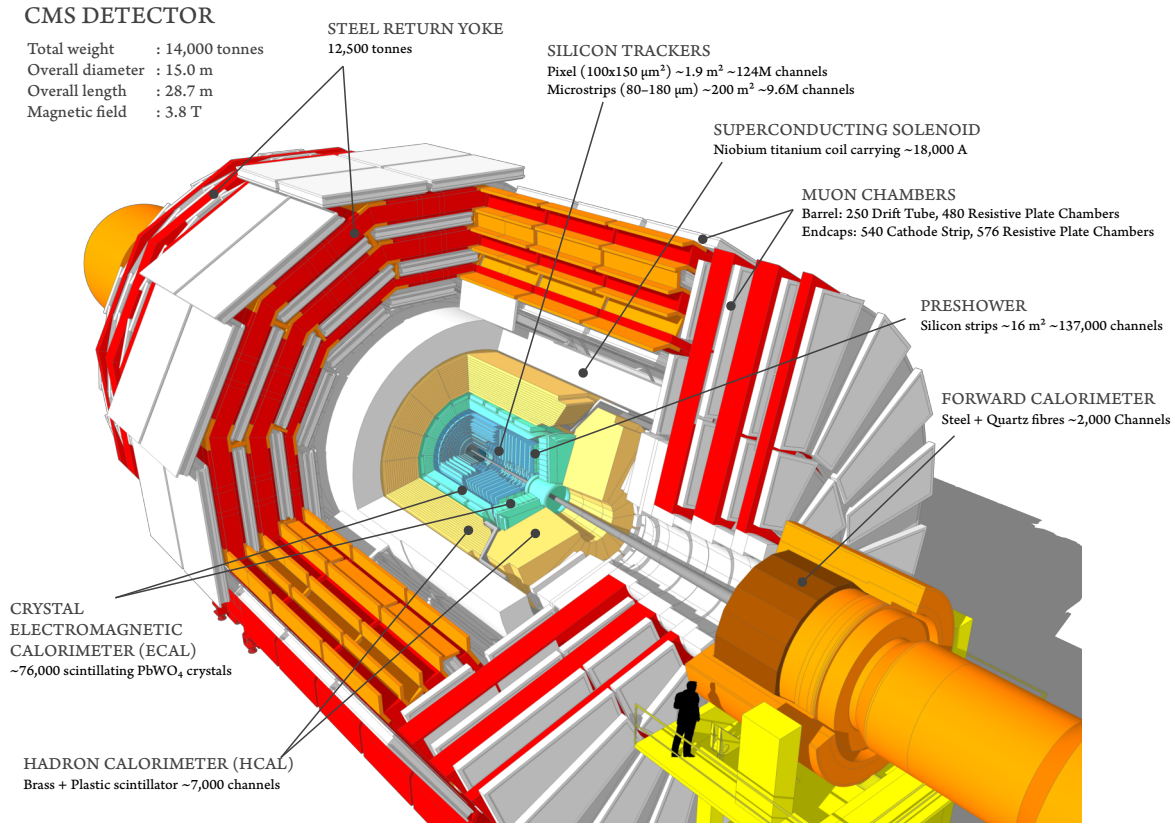


Figure 3.6.: A schematic view of the CMS detector and subdetectors with a section removed to view the internal components [55].

3.2.1. Silicon Tracker

The innermost subdetector is the silicon tracker [54]. It is designed to reconstruct the tracks of charged particles from ionisation deposits (known as hits) left on several silicon sensors layers. Measuring the curvature of the track makes it possible to determine the transverse momentum of a charged particle, defined by:

$$p_{\mathrm{T}} = rqB \; . \tag{3.12}$$

r is the curvature radius, q is the particle charge, and B is the magnetic field module.

Vertices from hard scattering, pile-up and decays are identified from intersections of the reconstructed tracks. All these points are typically located within a few centimetres, and to be able to resolve them, the tracker components surround the interaction point, with the first layer of sensors being at 4.4 cm from the beam line. A representation of the CMS tracker is shown in Fig. 3.7. The inner layer (20 cm), which withstands the highest flux of particles, is made of pixels, whereas the outer tracker is composed of silicon strips. To maintain occupancy at 0.1%

in the inner tracker and precisely measure 3-dimensional position hits, a total of 66 million 100 µm × 150 µm pixels are used. They are arranged into three layers for the barrel region (BPIX) at radii of 4.4 cm, 7.3 cm and 10.2 cm, with each layer 53 cm long, and two additional pixel layers placed in endcap regions (FPIX). Both in the barrel and endcaps, Pixel modules are arranged in an overlapping configuration to ensure that outgoing particles cross at least one module per layer.

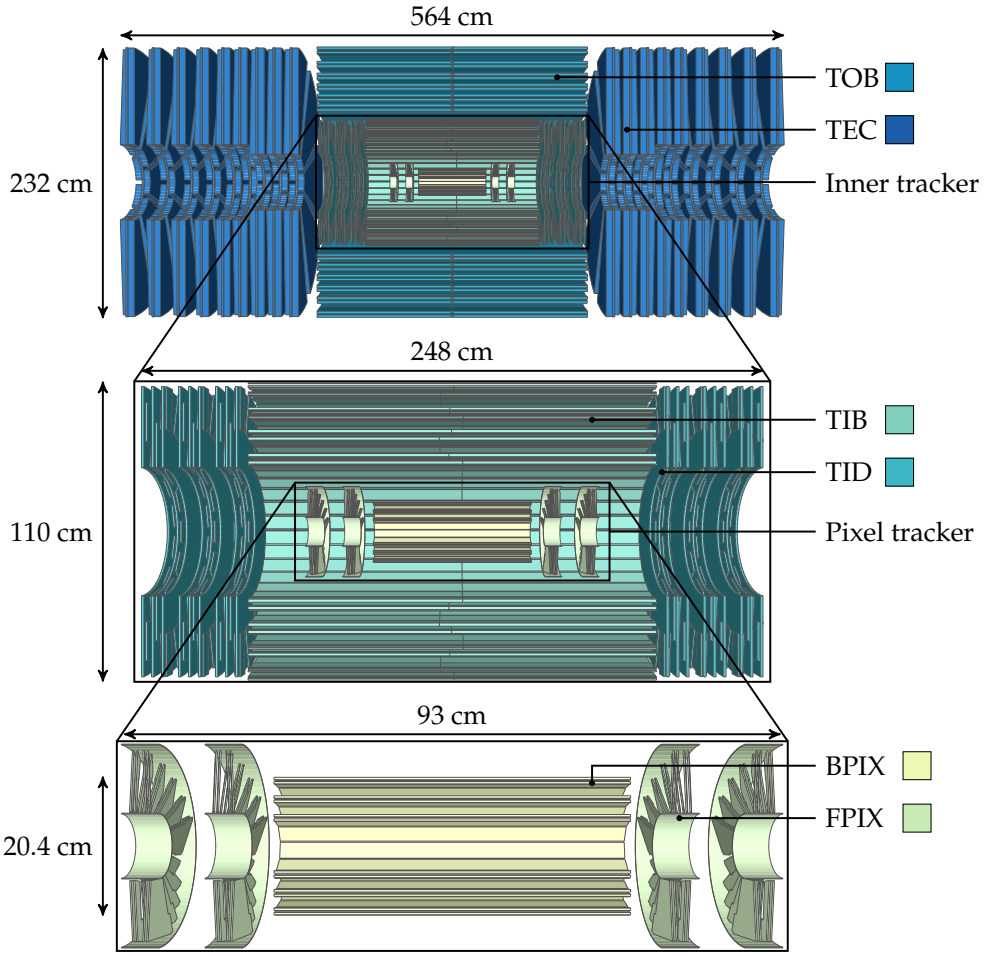


Figure 3.7.: A cross-sectional view, in the y-z plane, of the silicon tracker. The full tracker is shown in the upper part of the figure with a zoomed inset of the inner tracker in the central section, and a further zoomed inset of the pixel tracker in the lowest part. [56].

As the particle flux reduces moving away from the interaction point, silicon strips are used. Such sensors, whose dimensions range from 10 cm × 80 µm to 25 cm × 180 µm, are arranged in a stereo-configuration to allow 3-d measurement. The silicon strips are further separated into an inner and outer component, presenting similarly to the pixel layout, barrel and endcap regions. The tracker inner barrel (TIB) consists of four layers from $20 < r < 55$ cm and covering $|z| < 65$ cm whereas the tracker outer barrel (TOB) has six layers extending the coverage to $|z| < 124$ cm and $55\,\text{cm} < r < 116$ cm. The inner endcap (TID) component provides coverage

for the gap between the TIB and TOB with three small disks oriented perpendicular to the z-axis. The tracker outer endcap (TEC) consists of nine disks extending to 124 cm $< |z| <$ 282 cm [57]. The entirety of the strips tracker consists of 9.6 million silicon sensors and, together with the pixels system, provides a near hermetic coverage up to $|\eta| = 2.4$.

To process the data from the pixels and strips, an on-detector chip is needed. The read-out is attached to the silicon sensor via a bump bonding technique, and the electronics components are designed to operate in a high radiation environment. The analogue signal from particle hits is collected, processed and buffered in the processor, waiting on a decision to accept or reject the data (more details in Sec. 3.2.6.1). Once a decision is made, the signal is either flushed or transmitted over optical links to off-detector chips for digitisation and data formatting.

3.2.2. ECAL

The CMS ECAL [54] is an homogeneous calorimeter made of scintillating crystals realized in lead tungstate ($PbWO_4$) and measures electromagnetic deposits with a coverage of up to $|\eta| = 3.0$. The choice of an homogeneous design over a sampling one has been made to enhance energy resolution of electromagnetic objects, drawbacks of this being a deterioration over time from irradiation and the necessity of using photodetectors in a high magnetic field environment. The material used is characterized by short radiation lengths, X_0 (0.89 cm), and Moliere radius (2.2 cm) with fast responses (80% of the light emitted is within 25 ns) together with a good level of radiation resistance. Such properties are needed to meet design requirements (the calorimeter system is enclosed within the solenoid), providing sufficient containment (both lateral and throughout the crystals' length) for the cascade of electromagnetic particles produced when incoming photons or electrons interact with the ECAL material (electromagnetic showers).

The ECAL is subdivided into the barrel (EB) and the two endcap sections consisting of the endcap crystals (EE) and the preshower (ES) as shown in Fig. 3.8. The EB has an internal radius of $r = 129$ cm with the crystals tilted at a 3° angle with respect to the interaction point to avoid particles falling completely in the interface between two crystals. Each crystal covers 0.0174 radians with a cross-sectional base of 22×22 mm^2 and a length of 230 mm. The crystal dimensions are optimized to fully contain a shower, with the amount of material corresponding to 25.8 radiation lengths. Calorimeter units are organized in submodules, consisting of 5 pairs of crystals held together by a glass-fibre structure. 40–50 submodules, together with their readout electronics, are grouped into modules. A supermodule is made by 4 modules side-by-side, covering half the length of the barrel and 20° in ϕ, hence two halves of 18 supermodules complete the entire EB, covering up to $|\eta| = 1.479$. The EE starts at $|z| = 314$ cm from the nominal vertex and extends the ECAL coverage up to $|\eta| < 3.0$. Each endcap consists of two semi-circular aluminium plates ("Dees") supporting units of 5×5 crystal groups known as

supercrystals. Each crystal in the EE has a frontal cross-section of $28.6 \times 28.6\,\mathrm{mm}^2$ with a length of 220 mm resulting in 24.7 X_0. Between the TEC and the EE is the ES, which covers the range $1.653 < |\eta| < 2.6$. The ES consists of two layers of silicon strips and disks of lead absorbers at depths of $2X_0$ and $3X_0$ respectively to initiate the electromagnetic showers and measure their tracks. The higher granularity of strip modules allows for discrimination of individual electrons or photons from the decay of $\pi^0 \to \gamma\gamma$ and provides spatial information, which is particularly important for reconstructing the direction of photons.

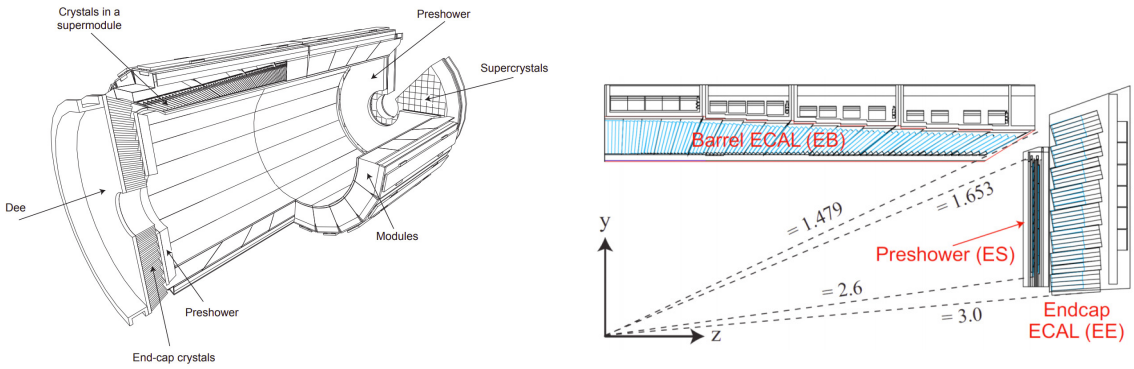


Figure 3.8.: Schematic layout of the CMS electromagnetic calorimeter, presenting the arrangement of crystal modules, supermodules, endcaps and the preshower in front (left). Geometric view of one-quarter of the ECAL (right) [58].

Incoming electromagnetic particles interacting with the crystals produce scintillation light. The photons produced are measured with avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcap. Read-out electronics process and buffer the signals until a decision is made, and if an accept signal is received, data is transferred to off-detector electronics for further processing. Simultaneously, faster algorithms using input gathered from 5×5 clusters are used as an input variable for triggering purposes (more details in Sec. 3.2.6.1).

The reconstructed energy of an electromagnetic shower determined from photodiode signals is given by:

$$E = G\mathcal{F}\sum_i c_i A_i \,, \tag{3.13}$$

where G is the absolute scale and converts the signal in GeV; it is determined from in situ measurements of photons in $Z \to \mu\mu\gamma$. The function $\mathcal{F}$ is a correction factor that depends on the type of particle, its momentum and location in the detector, and the clustering algorithm used. This correction is determined from simulation and validated by test beam and in situ measurements of $Z \to ee$ and $Z \to \mu\mu\gamma$. The coefficients c_i, known as intercalibration coefficients, are obtained from the combination of laboratory measurements, test beam precalibrations, cosmic ray measurements and in situ $W \to e\nu$ measurements and $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$ mass

reconstruction. They are crystal dependent and multiplied by the signal amplitude A_i, and are summed over each crystal in the cluster, labelled by i.

The calorimeter modules energy resolution is measured with test beams by a Gaussian fit to the reconstructed energy distribution. It is parameterised as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \,, \tag{3.14}$$

where S, N and C are respectively the stochastic, noise and constant contributions. They are measured to be $S = 2.8\,\%$, $N = 12\,\%$ and $C = 0.3\,\%$. Overall the CMS electromagnetic calorimeter measures the energy of photons and electrons of 10–250 GeV with a resolution that ranges between 1.50–0.35 %. For more energetic objects the resolution is dominated by the constant term which depends mainly on the gain as well as intercalibration errors and crystal non-uniformity [59].

3.2.3. HCAL

The HCAL [54] design allows for full coverage of the hadronic activity from collisions. Compared to ECAL, it extends the pseudorapidity coverage up to $|\eta| = 5$, thanks to its forward region (HF) subdetector. The other components of the hadronic calorimeter are the barrel (HB), the endcaps (HE), in a similar layout as the ECAL, and an extra section located beyond the solenoid (HO) that acts as a tail catcher for hadronic showers [60]. A diagram showing the layout of the four HCAL components is given in Fig. 3.9. The HE and HB are sampling calorimeters with a brass absorber and plastic scintillator tiles, and they are fully immersed within the 3.8 T magnetic field.

The HB covers the region $0 < |\eta| < 1.3$ with an inner radius of 1777 mm and outer radius of 2876.5 mm. It consists of two longitudinal half-barrels of 18 wedges of 20° in ϕ, with each wedge further segmented into 5° sectors. The wedges are made of 17 layers of plastic scintillators interspersed with brass plates. Each sector's layers are enclosed by a steel structure that provides structural strength, forming a tower. Each tower covers the same area of an ECAL supercluster ($\Delta\eta \times \Delta\phi$ =0.087×0.087). The design choice of having both tracking and calorimeter systems within the magnet results in space-constraints which restrict the HB to 5.7 hadronic interaction lengths (λ_{int}) at $\eta = 0$ and up to 9 at $\eta = 1.3$, where typical interaction length needed for completely stopping hadrons with several hundreds GeV of momentum is $\sim 10\lambda_{int}$. The cascade of hadronic particles produced when scattering with the absorption plates material, produce scintillation photons when traversing the active scintillation tile. Such photons are captured by wavelength shifting (WLS) fibres embedded within the plastic tiles and fused to optical clear fibres running along with the tower. The optical signal is summed across the depth of a tower in three groups: the first layer, layers 2 to 5 and 6 to 17. The

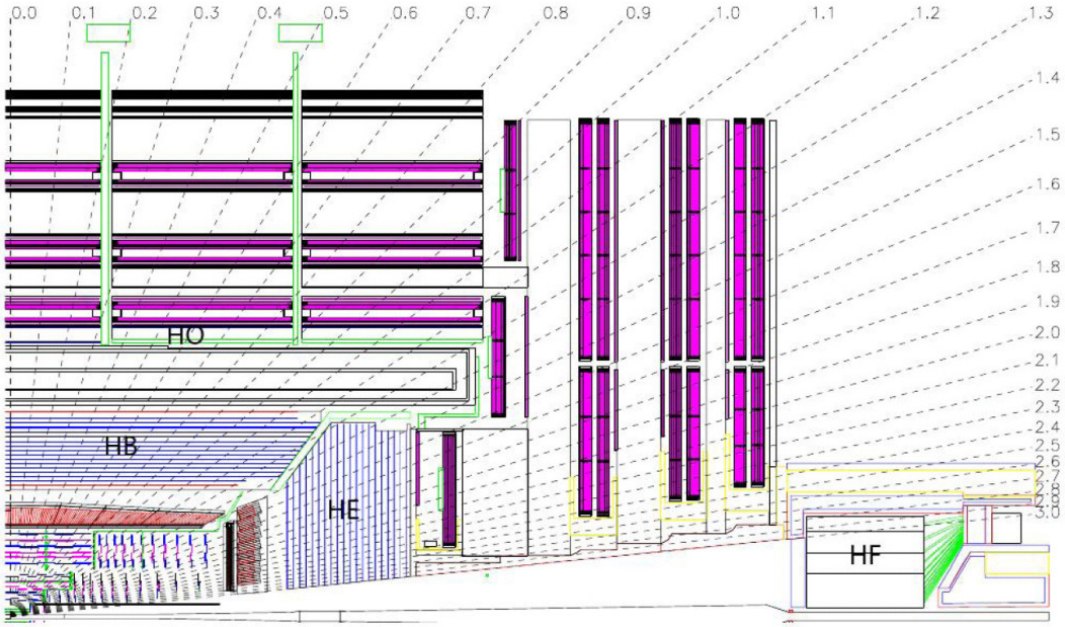


Figure 3.9.: Front cross-sectional view of a quadrant of the CMS HCAL. The four subdetectors' location (HB, HC, HO and HF) constituting the HCAL is displayed [59].

clear fibres are coupled to Silicon photomultipliers (SiPM). The level of depth-segmentation is limited by signal-to-noise in the photomultipliers. For this reason, the layers need to be grouped, and the adopted schema of layer combination is the result of an optimization to improve tracking of hadronic showers [60].

The HE subdetector covers the region $1.3 < |\eta| < 3.0$, overlapping the HB to avoid particles falling completely in the uninstrumented gap. The HE is composed of towers identical to the ones used in the HB for $|\eta| < 1.74$, with wedges covering 20° in ϕ and further segmented to 5°. The wedges in $|\eta| > 1.74$ are made by towers covering 10° in $\Delta\phi$ and 0.09–0.35 in $\Delta\eta$. The HE is segmented longitudinally in four or five groups of layers, following a $\{1+2+3+5+7\}$ scheme. The overall CMS design allows sufficient clearance to accommodate material interaction lengths to mostly contain hadronic showers, however in the barrel the shower tails can propagate through the magnet [60].

The HO consists of one or two layers of plastic scintillator tiles placed beyond the CMS solenoid, measuring the hadronic showers leaking from the HB. The absorbing material, in this case, is made by the solenoid and iron return yoke, extending the hadronic interaction lengths to a minimum of 11 λ_{int} across the HO extension ($\eta < 1.26$). The whole assembly consists of 5 rings with 12 wedges, subdivided into 5° sectors, matching the HB tower size. In the central ring, where the interaction length of HB is minimal, two layers of plastic scintillator at $r = 3850$ mm and $r = 4097$ mm separated by a 19.5 cm thick piece of iron are used. In contrast, the remaining four rings are made of a single layer located at a distance $r = 4097$ mm. Partial

plates are placed in rings ± 1 around the chimney allowing power cables and cryogenics inlet for the magnet.

The HF is placed at $|z| = 11.2\,\text{m}$ from the interaction point and operates in particularly elevated radiation density. For this reason, the design choice for this apparatus differs from the rest of the HCAL, using steel absorbers embedded with radiation-hard quartz fibres that produce Cherenkov light from the interaction with hadronic showers. The radiation is detected by dual-anode photomultipliers [60]. The steel absorbers are segmented in 18 wedges covering 20° each, and subdivided into two towers of 10°, apart from the last two towers up to $|\eta| = 5.0$, which are 20°. All the towers are oriented longitudinally with an η segmentation of approximately 0.175, varying with z. Immersed in the absorbing material are the fibres also parallel to the beam axis. Quartz fibres present two different lengths (1.65 m and 1.43 m); thus, the HF is also segmented in-depth [61].

The HCAL resolution can be expressed by the formula:

$$\frac{\sigma_{p_\text{T}}}{p_\text{T}} = \sqrt{\text{sgn(N)} \cdot \left(\frac{N}{p_\text{T}}\right)^2 + \frac{S^2}{p_\text{T}} + C^2}\,, \tag{3.15}$$

which resembles the classic calorimeter-based resolution, with the noise (N) stochastic (S) and constant (C) terms, but introduces the possibility of the noise term to be negative. This happens at low PU when tracking information is added. Calibration of the calorimeter is done by combining a variety of techniques. Initial calibration is estimated with radioactive sources, test beams with e, π and μ and lasers. The calibration is then improved with in situ measurements using collisions data and cosmic rays. The total energy scale uncertainty is overall smaller than 3%. The hadronic energy resolution is then measured with dijet and γ+jets samples when combining calorimeters and tracking information, an overall resolution of less than 10% for central jets ($\eta < 0.5$) of 100 GeV is measured, growing up to 20% in the forward region.

3.2.4. Solenoid Magnet

CMS design requirements on muon identification aim to identify the charge of 1 TeV muons unambiguously. To achieve this, a niobium-titanium superconducting magnet is used, providing field strength of 3.8 T parallel to the beam direction. The magnet, shown in Fig. 3.10, measures 12.5 m in length and 6 m in diameter, and has a mass of 220 tonnes, which is kept at 4.5 K by a superfluid helium cooling system. To maintain an elevated bending power outside the magnet cylinder, the returning field is focused by a 14 m iron return yoke. Thanks to this structure, the magnet field strength is kept at 2 T in the muon system.

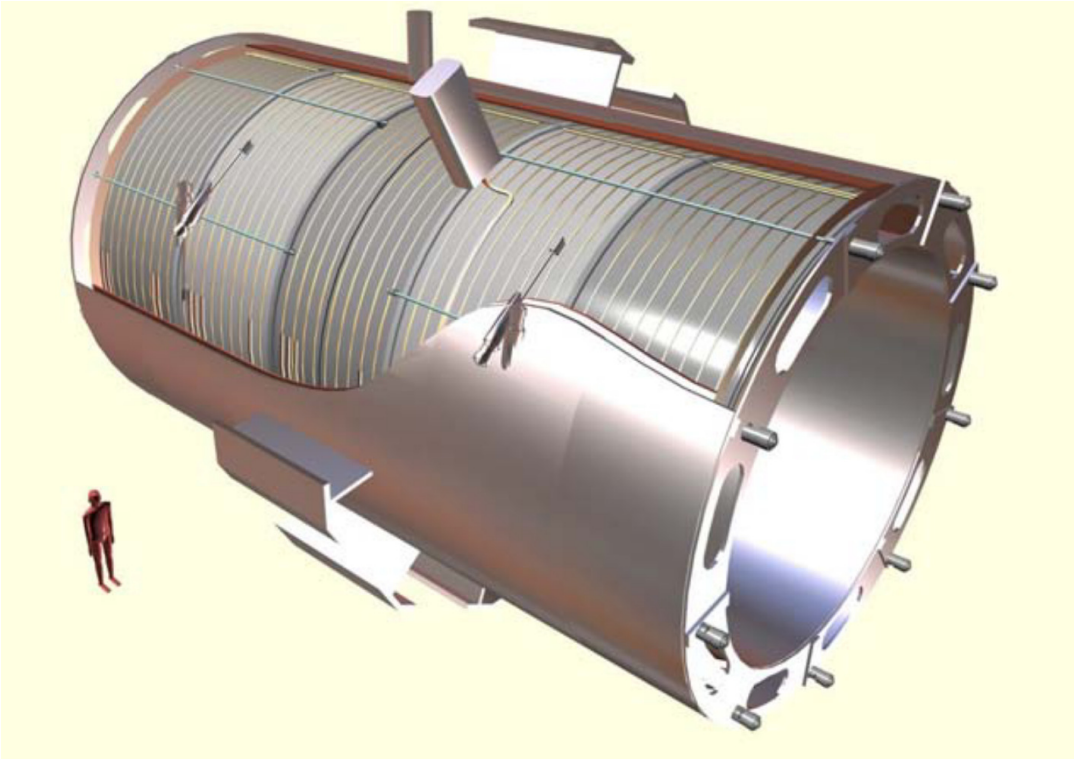


Figure 3.10.: Graphical representation of the CMS solenoid. The five modules composing the cold mass of the cryostat are visible, together with the supporting system [59].

3.2.5. Muon Chambers

The outer part of the CMS detector comprises the muon system, which consists of several tracking detectors that provide 3-dimensional position measurements for muons[54], which interact weakly with the detector materials and therefore are not stopped in the calorimeters or magnet. Given the large area to be covered, characterized by different radiation levels, different types of detectors are needed, and the design choice was to use three different gaseous detectors.

The layout of the muon detectors is shown in Fig. 3.11. In the barrel region, drift tubes (DT) are used. These provide 2-dimensional position measurements, which are used to collect the ionisation charge deposited from charged particles flying through the gas. Since the radiation and muon density is higher in the endcaps, another type of detector, the cathode strips chamber (CSC), has been utilised. Such detectors allow for 3-dimensional position measurement and therefore perform better than DTs. Resistive plate chambers (RPC) are used in both barrel and endcaps, matching DT and CSC hits. These detectors use two plates presenting large voltage differentials to collect the ionization charge and are typically faster than DT and CSC, providing good timing resolution, which is needed for bunch identification and triggering purposes.

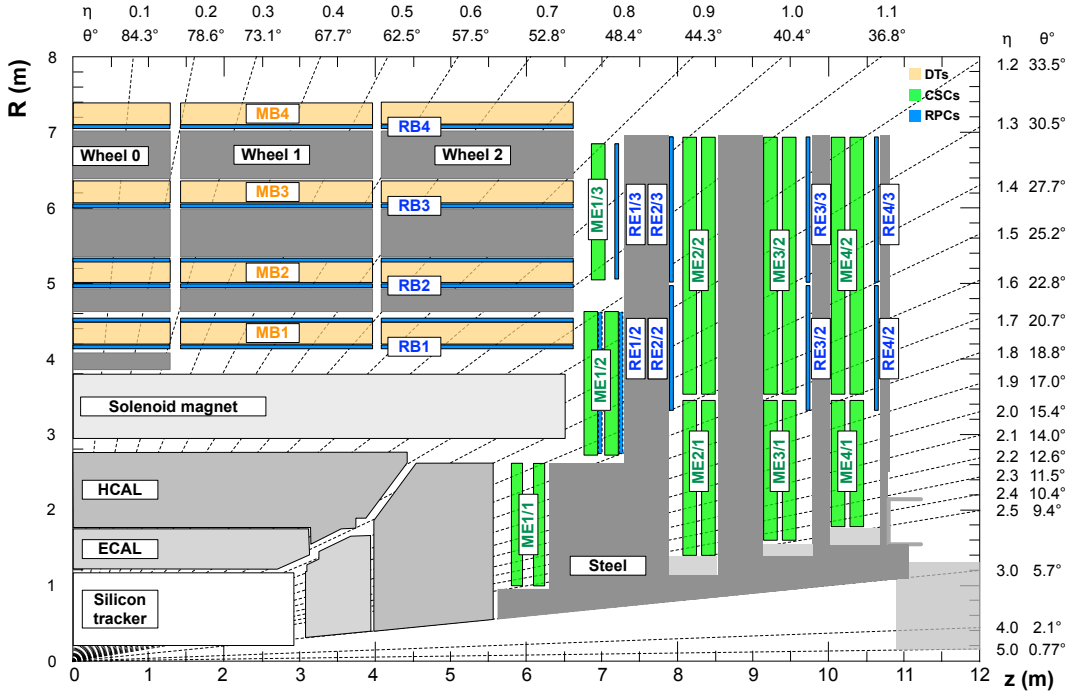


Figure 3.11.: The figure depicts one quadrant of the CMS detector, with the Muon detectors in colour: DTs are in yellow, CSCs in green and RPCs are coloured in blue. The interaction point (IP) on the origin. The division of the muon system into stations (MB), wheels (YB) and disks (ME) are discussed in the text. [62].

The muon barrel (MB) covers up to $|\eta| < 1.2$ and consists of four layers labelled MB1–MB4 positioned at $r = 4.0$, 4.9, 5.9 and 7.0 m respectively. Each layer, also known as "station", is subdivided into sectors, each covering $\Delta\phi = 30°$, with adjacent sectors overlapping to avoid uninstrumented gaps. Similarly to the HO layout, the MB is divided into five wheels along the beam axis, labelled YB $-$ 2 to YB $+$ 2. MB1 and MB2 stations are made of a DT chamber between two RPCs, whereas MB3 and MB4 have one RPC placed underneath the DT chamber. Each DT chamber consists of several layers of drift tubes grouped in superlayers (SL). Each SL is made of four drift tubes layers with their anodes oriented in the same direction. A drift chamber contains two SLs oriented along the beam axis, and therefore measuring coordinates on the r-ϕ plane (SL_ϕ) and a third SL perpendicular to the SL_ϕ, measuring the z coordinates (SL_θ). Drift chambers contain also a honeycomb support structure (HC) and overall they are assembled as follows: SL_ϕ, HC, SL_θ, SL_ϕ for MB1 to MB3 and SL_ϕ, HC, SL_θ for the MB4 station.

The muon endcaps (ME) cover the range $0.9 < |\eta| < 2.4$ overlapping with the barrel. Both endcaps are composed of four disks, labelled ME1–ME4, placed perpendicular to the beam axis and consisting of two concentric rings divided into 18 (inner ring of MB2–4) or 36 sectors. The ME is composed of CSCs and RPCs; the RPCs are similar to those used in the barrel and cover up to $|\eta| = 2.1$. The CSCs are multi-wire proportional chambers with cathode strips perpendicular to the anode wires. Each chamber presents 6 gas gaps that provide the

longitudinal position. Cathode strips are placed radially measuring ϕ position, and anode wires are placed roughly into rings around the beam axis to provide the radial position.

The muon chamber energy resolution is about 10% for muons in the barrel region and with $p_T < 200$ GeV, growing to 15% – 40% for 1TeV muons, depending on η. However, if combined with tracker information, the p_T resolution improves to 5% and 10%.

3.2.6. Trigger system

As described in Sec. 3.1.2, the LHC collides protons at 40 MHz. With an average of 25 interactions per bunch crossing, this results in roughly 1-1.5 MB of data per event, and if all the collisions were recorded, it would require a throughput of up to 60 Tb/s. This large amount of data cannot be recorded, and readout electronics cannot work at such a frequency; therefore, the data flow has to be reduced. Moreover, as mentioned in Sec. 3.1.1, the majority of interactions result in low transferred momentum collisions which are not useful to the CMS physics programme. The trigger system has the crucial role of reducing the event rate to ~ 1 kHz and at the same time recognizing and writing to tape all the interesting events. The trigger system is implemented in two stages: the Level-1 trigger (L1) and the high-level trigger (HLT) The L1 is a hardware trigger based on custom-designed electronics [63, 64], whereas the HLT is software-based and operates on several thousand CPUs [65].

3.2.6.1. Level-1 trigger

The L1T implements the highest rate reduction, bringing it from 40 MHz to 100 kHz. To achieve that, the maximum latency allowed is 3.8 µs, with a standard buffer capacity of 128 bunches and the possibility of storing 12 extra bunches for contingency. The signals are extracted from the detector and fed to a processing farm situated in an adjacent cavern via optical fibres. Data from each bunch crossing is stored in a buffer waiting for the trigger decision. Therefore the allowed processing time is less than 3 µs. For this reason, most L1T operations consist of simple arithmetic and queries of lookup tables (objects storing expensive computations results). Moreover, these restrictions prevent the L1T from accessing the full event reconstruction, and only the muon system and a coarser representation of the calorimeters provide input for decision making. Fig. 3.12 gives a schematic representation of the L1T dataflow, for both muons and calorimeter triggers consist of 2 layers of processing for the signal's treatment. The L1T performs a reconstruction of individual objects such as photons, electrons, muons, jets, transverse energy sum of missing transverse energy (E_T^{miss}). The physics requirements that drives the L1T final decision are related to the efficient reconstruction of W,Z and Higgs bosons in their lepton and photon channel, as well as ensuring sensitivity for BSM processes which typically involve sizeable E_T^{miss} in the final state. In particular, the goal is triggering with

an efficiency of above 95% for events containing a single muon or a pair of muons or a single electron/photon with $|\eta| < 2.4$ and $p_T > 40\,\text{GeV}$, as well as events presenting E_T^{miss}.

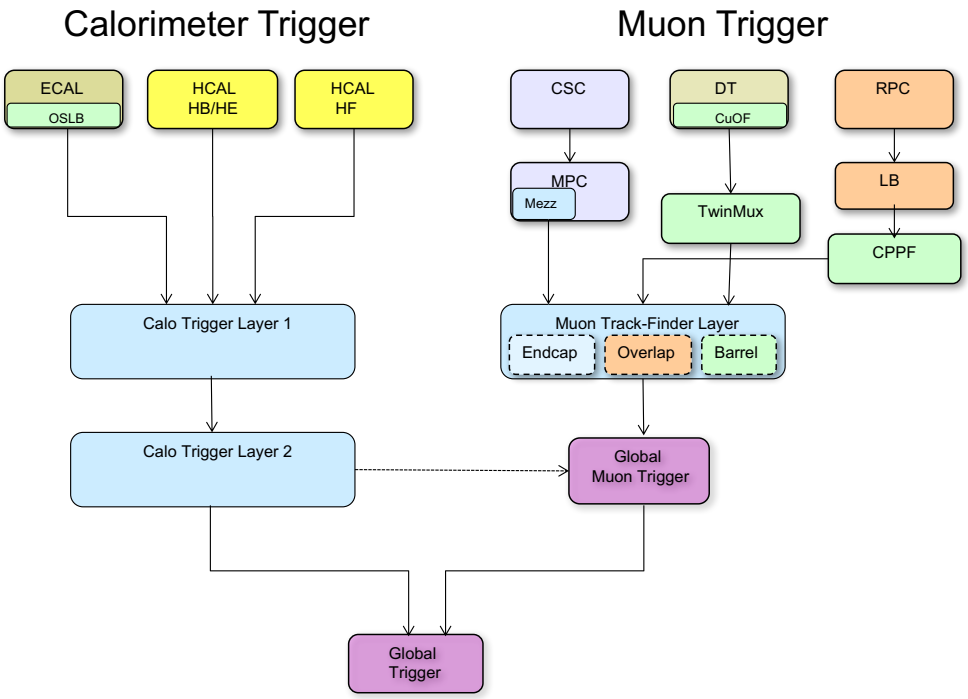


Figure 3.12.: Graphical representation of the L1 trigger dataflow. The calorimeter trigger collects data from ECAL, HCAL and HF separately. Their outputs are sent to the calorimeter trigger layer 1, which pre-process the inputs and subsequently forwarded to layer 2, which implements the triggers algorithms. The inputs from RPC, DT and CSC are separately used in the track finder layer track candidates are then fed in the sorting and merging layer. The global muon trigger merges muons candidates from different systems, including calorimeter isolation information. Finally, the global trigger combines information from both systems to form the Level-1 decision [60].

Calorimeter trigger

The calorimeter trigger collects signals from both ECAL and HCAL, grouped into the so-called trigger towers: in the barrel, a trigger tower is made of a 5×5 group of crystals, of total dimension $0.087\,\eta \times 0.087\,\phi$, corresponding to the size of one tower in the HCAL. In the endcap, the towers are more irregular but are matched to the size of the corresponding HCAL towers as close as possible. The data from each trigger tower are sent to the calorimeter trigger, which implements a time-multiplexed architecture realised over two layers of FPGA cards to process the data. In the first layer of the time-multiplexed trigger, each card is mapped to a calorimeter system slice, with the entire detector described by 18 cards. The layer-1

cards preprocess calorimeter inputs and re-transmit them to a single layer-2 node. Time multiplexing is implemented in this step as the transmission time exceeds bunch crossing frequency, and therefore, layer-1 cards have to redirect the output from subsequent events over different layer 2 nodes. The calorimeter trigger implements a multiplexing factor of 9, meaning that a minimum of 9 layers and 2 nodes are needed to keep up with the collision rates. The layer 2 nodes implement triggering algorithms having the full calorimeter description available, constructing particle candidates that are forwarded to the global trigger for the final L1 decision.

Muon trigger

The muon trigger system implements track finding from successive chamber muon hits. The trigger is subdivided into three regions: the endcap track finder (EMTF) using hits from CSCs and RPCs and covering from $|\eta| = 1.25$, the barrel track finder (BMTF) using DTs and RPCs signals up to $|\eta| = 0.8$ and the overlap track finder (OMTF) which uses inputs from the region where barrel and endcap muon systems overlap and therefore receive hits from RPCs, CSCs and DTs. The muon chambers signal is collected, refined, and assigned to a specific bunch crossing by dedicated hardware located in the detectors' proximity. The regional track finding algorithms then identify muon track candidates, which are sent to the global muon trigger. The global muon trigger can receive up to 16 muons per bunch crossing and performs operations to merge and remove duplicate muon candidates, as well as combining the isolation information from the calorimeter trigger. Finally, the four best muon candidates are forwarded to the global trigger.

Global trigger

The global trigger receives objects and variables from the muon and calorimeter systems, correlating and synchronizing trigger candidates.

The L1T decision depends on events fulfilling a set of requirements for each object topology, aiming to guarantee the physics performances described in Sec. 3.2.6.1. An object satisfying these is called a L1T seed.
The collection of all the active seeds for a particular run is called a L1T Menu. Limitations in the L1T bandwidth require the global trigger to discard events even if the current L1T menu would have selected them. This process, known as prescaling, reduces the trigger efficiency for seeds that target objects characterised by a large production cross-section. For example, a prescale of 3 means that only one in three events firing the given seed is accepted. After the prescaling is applied, the L1T decision is sent to the buffers, and data are either discarded or passed to the high-level trigger.

3.2.6.2. High-level trigger

The high-level trigger, HLT, builds from L1T seeds and performs the final decision for events to be saved on disk. Given the reduced event frequency, it can exploit the full detector granularity and all subdetector information, including tracker inputs. The HLT consists of a processor farm of about 1000 cores located on the surface at the CMS site. Each core runs a copy of the CMS software code and can make a decision based on a set of requirements, known as paths, to L1T selected events. A Dynamical allocation of the data ensures optimal resources usage. To meet the timing constraint, reconstruction of the event is based on the received L1T seeds, gradually performing the full reconstruction verifying each step contained in the targeted HLT path is satisfied. HLT reconstruction is close to offline reconstruction (discussed in Chap. 4), lacking only calibrations and specific run conditions information. After accepting an event, the data is archived and made readily available for further operations such as data quality monitoring and trigger studies (for which run conditions are actually needed, further details in Sec. 3.2.6.3), as well as first offline reconstruction, which happens within 48 hours from the production.

3.2.6.3. Conditions database system

As explained in Sec. 3.2.6.2, the HLT workflow builds on L1T outputs. Together with L1T seeds, other non-event data are needed. The prescales information and the L1T Menu containing the active L1T seeds are needed online by the HLT to re-emulate L1 decisions. Moreover detector calibration, setup and alignment configurations are used for online data-quality monitoring, emulating trigger rates, and offline event reconstruction. Thus, it is crucial to keep a log of the detector conditions for each run configured in CMS. This large amount of data is stored in ORACLE databases located in the CMS experimental site. All possible detector setup data are loaded on the Online Master Database System (OMDS). When a particular setup is loaded for data taking, an automated logging workflow, known as online-to-offline (O2O), fetches the configuration data and writes it to the Offline Reconstruction Condition Database Online System (ORCON). The ORCON database is then replicated on another database connected to the CERN IT network, and it is accessed by offline workflows [66]. A depiction of the CMS condition database architecture is shown in Fig. 3.13.

The OMDS has a purely relational database structure, and each subsystem has designed its own schema. L1T configurations are stored in the form of XML files organized by keys; to load a particular configuration at run time, a set of keys is selected, one for each subdetector and each one pointing to one or more XMLs. Given the complexity of the detector, the size of the OMDS database is considerable (several TBs) and contains heterogeneous configuration schemes. For this reason, the logging of the configuration uses a data model which optimises

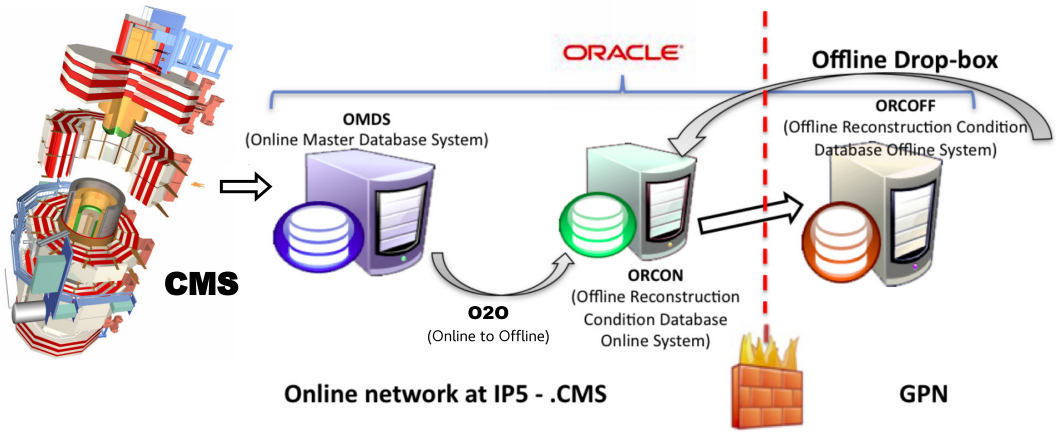


Figure 3.13.: CMS condition Database Architecture [67].

resource querying and encapsulates the underlying configuration data completely.
The bulk data is made by a boosted serialized object, the payload, which represents the set of configuration parameters consumed in the various workflows of the physics data processing. Each payload is exchanged and stored as an unstructured binary array, with no assumption on its internal layout, and to be used, needs a dictionary class to extract the information. To each payload is assigned time information for the validity of the particular configuration. In this context, time is measured in terms of run numbers. The interval of runs for which a specific payload contains valid conditions is known as Interval of Validity (IOV). The status of a particular subdetector, evolving in time, is described by a set of payloads and their associated set of IOVs. Payloads and IOVs are grouped in tags, labels that identify sets of conditions consumed by the same workflow. The IOVs collected in a tag are contiguous; therefore, an interval can be identified by a lower bound only. The last IOV in the sequence has a default infinite upper bound. This design choice has the great advantage of not requiring a continuous extension of payload validity, ensuring that access to the database is limited to the changes involving new payloads. A collective label called Global Tag identifies the set of tags involved in a given workflow. In this way, specific collections of tags can be handled simultaneously, providing conditions for the full detector, allowing a specific configuration to be used by various jobs for large-scale data production.

3.2.7. Worldwide LHC Computing Grid (WLCG)

Despite the important reduction of throughput performed by CMS trigger system, a vast amount of data is still recorded. Events accepted by the trigger require a prompt full reconstruction in the first hours from being registered and several reconstruction steps to incorporate improved object reconstruction and calibrations. Moreover, simulations of proton-proton col-

lisions for both SM processes and BSM signals, which include a simulation of the detector response, the emulation of the trigger decision given the recorded conditions as well as a reproduction of PU regimes with full event reconstruction, demands a significant amount of computing power. To cope with this amount of processes, the WLCG was established [68]. The WLCG is a highly distributed computing cluster organized into Tier levels. A schematic representation of the WLCG is reported in Fig. 3.14.

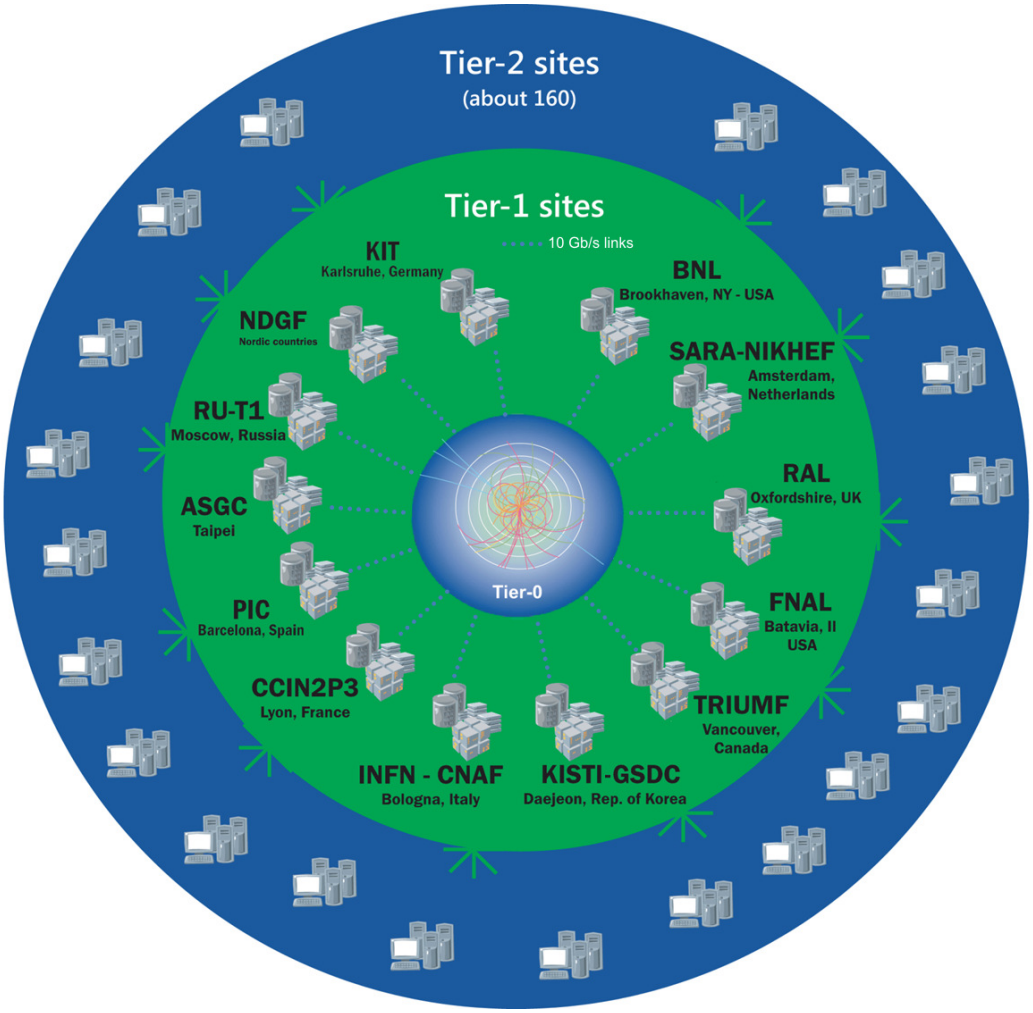


Figure 3.14.: Schematic depiction of the WLCG TIER structure [69].

The Tier-0 is located at the main CERN site and provides resources to all LHC experiments for immediate reconstruction and storing on tape of recorded events. It also provides computing resources for simulation to the experiments on an opportunistic basis. The Tier-1 level is made of 13 large computing centres worldwide, providing additional storage for data from Tier-0 and simulated events, which are produced directly at this Tier-level. At Tiers 1, data reprocessing is also performed, and several collaboration wide services are hosted. The Tier-2 level provides the bulk of storage and resources to CMS members. The computing centre at this level can run CMS user analysis and host on-disk temporary replicas of samples used for

analysis purposes. Smaller computing centres hosted by universities and laboratories are often referred to as Tier-3 sites. They generally do not provide computing resources or storage to the collaboration, mainly fulfilling local users' requests.

Chapter 4.

Event reconstruction

Starting from the raw inputs of the CMS subsystems, which are stored after the trigger decision, the reconstruction of the full event final state is implemented. The precise identification and calibration of physics objects are crucial to executing the CMS research program. The reconstruction of particles is performed with a technique that combines all the available information from the different subsystems into a coherent picture of the event particles, known as "particle flow" (PF). The building blocks used by PF to identify particle candidates are tracks of charged particles from the inner tracker and the muon systems and calorimeter clusters containing particle energy deposits. In the following chapter, a description of the tracking, clustering and PF processes is presented, together with details on the physics object identification selections.

4.1. Tracks

Charged particles, originating from the collision point, travel in helical trajectories across the magnetic field. A fitting technique known as the combinatorial track finder (CTF) reconstructs the silicon tracker's hits into tracks. Inside the pixel tracker, hits are created by clustering charge deposits in adjacent pixels above a certain threshold. The hit position for single-pixel clusters is taken at the middle of the pixel. In contrast, multi-pixel clusters have a charge-weighted position, correcting for the drift of ionisation electrons due to the magnetic field. The final track fit involves a χ^2-fit of the predicted charge deposits using a large number of simulated trajectories, also accounting for pixel deterioration. Strip tracker hits are built from adjacent strips clusters with charge deposits above a noise limit. Similarly to the pixel clusters, the hit's position is determined from a charge-weighted cluster average, correcting for the electron drift in the magnetic field and calibrating the measured charge. All hits have an associated uncertainty which is taken into account when performing the track fitting procedure [70].

The CTF consists of a series of Kalman filters (KF) [71] — an iterative parameter estimator. The Kalman filter algorithm is initialised with a seed, typically from a guess or estimation,

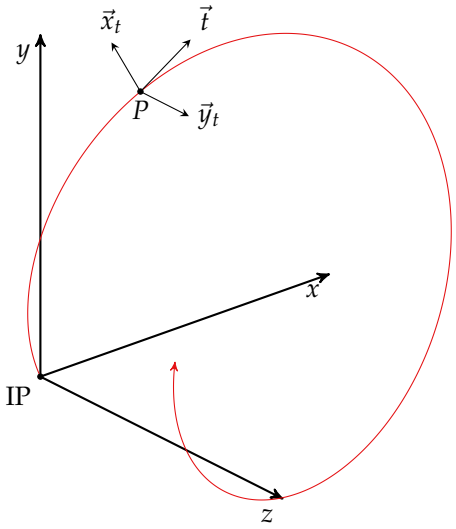


Figure 4.1.: The diagram shows a track (shown in red) starting from the interaction point (IP) and the CMS coordinate system. The local coordinate system along the trajectory can be observed, having the point P as the origin. The local system is defined by $\vec{t}$, the tangential vector, $\vec{x}_t$, perpendicular to both the z-axis and $\vec{t}$ and $\vec{y}_t$ which is the remaining vector forming a right-handed coordinate system. The angle of the projection of $\vec{t}$ onto the x-y plane is ϕ, whereas θ is the angle between of the y-z projection of $\vec{t}$ [56].

which is used to build a first track candidate. Comparing iteratively the prediction with the observed hits, the initial state is updated. The helicoidal trajectory is fitted using the perigee parametrisation, characterised by five independent parameters as follows:

$$x = \left(\frac{q}{p_{\mathrm{T}}}, \theta, \phi, x_t, y_t \right) , \tag{4.1}$$

where:

- q/p_{T} is the ratio between the charge and the momentum associated with the track.
- θ and ϕ are respectively the dip and the azimuthal angle in the global reference frame.
- x_t and y_t are the coordinates of a point along the trajectory in the $\vec{x}_t$ and $\vec{y}_t$ axes, which are the basis of a local right-handed reference system defined by the tangential vector $\vec{t}$ to the particle trajectory.

A schematic representation of a track and the defining parameters is shown in Fig. 4.1.

The Kalman filter's initial state is obtained by combining a set of pixel hits selected for the low hit uncertainty and pixel occupancy. Each KF iteration includes hits in the next tracker layer within a χ^2 window. Tracks arising from detector noise typically present layers where no hits are found within the acceptance window and can be discarded. Additional

hit candidates undergo further checks for consistency with estimated track p_T and vertex constraints. Candidate tracks are cross-cleaned so that each seed is associated with a unique track and vice versa. Once the full set of tracks is reconstructed, a new Kalman filter procedure is implemented, but this time only hits belonging to the track candidates are considered. This allows precision fitting of the track parameters, accounting also for inter-layer material and magnetic field inhomogeneities. To avoid bias towards the initial KF, a third fit is repeated, considering seeds formed by hits from the outermost layer and iterating inwards.

The trajectory building is repeated a total of ten times, starting from different seeds to target various particle classes. After each step, hits associated with fitted candidate tracks are removed:

- The first three iterations target prompt high p_T tracks, and b-hadron decays. Seeds are formed using three of the four available hits in the pixel detector, applying constraints on the distance of the nearest approach to the beam axis and p_T of the track.
- Two iterations are used to reconstruct tracks with one or two missing hits in the pixel detector which may be caused by particle decays or detector inefficiencies. These are seeded by two or three pixel hits combined with hits in the strips layers.
- The sixth and seventh iterations are seeded by strip hits, targeting displaced tracks.
- The eighth iteration aims to reconstruct high p_T jets with overlapping constituent tracks. The initial seed is characterized by a pair of hits in the pixel and strips, compatible with tracks pointing to high energy calorimeter deposits.
- The last two iterations target muons that have not been reconstructed in prior iterations, using muon system hits to form seeds and propagating tracks inwards to the tracker.

The tracks obtained with these iterations are propagated from the outermost hit to the calorimeter system, checking for overlapping clusters and towards the beam pipe to estimate the distance of closest approach. The overall momentum resolution is 0.7 (5)% at 1 (1000) GeV in the barrel region. Track reconstruction efficiency has been measured to be 98% (99%)for tracks with p_T above 500 MeV $(2\,\text{GeV})$ [72].

4.1.1. Electron tracking

Due to the magnetic field's presence in the tracker volume, a large fraction of electrons emit Bremsstrahlung radiation along their trajectories. The CTF algorithm might miss tracks irradiating high energy photons. Therefore, to recover efficiency with these particular tracks, another fitting procedure is implemented based on Gaussian-sum filters (GSF). The GSF method fits track hits with a weighted sum of Gaussian distributions, treating radiative losses

as non-Gaussian noise. The fit output is further processed via a boosted decision tree (BDT) to improve reconstruction efficiency and maximise the rejection of noise generated tracks. The same fitting procedure is also used to reconstruct tracks of electron-positron pairs arising from photon conversions in the tracker.

4.1.2. Muon tracking

Muon tracks are reconstructed with the CTF; however, since muon system hits are also taken into account, various typologies of muon tracks are defined:

- **Standalone muon:** these tracks are reconstructed using hits from the muon system only. Seeds are formed with inputs from the DT or CSC detectors, and the fitting procedure includes all muon chambers. To improve momentum resolution, the fitted muon tracks are constrained to be compatible with the beam spot.
- **Tracker muon:** in this case, all reconstructed tracks are considered potential muon candidates. This hypothesis is tested including calorimetry and muon systems information. Such reconstruction is beneficial to measure low p_{T} muons, which do not leave significant hits in the muon tracker.
- **Global muon:** standalone muon tracks are further matched to inner tracker candidates. The tracks providing the best match are selected, and a new fit considering both the muon system and tracker hits is performed, providing the best momentum resolution.

4.2. Vertex reconstruction

As mentioned in Sec. 3.1.2, 2016 LHC conditions yielded roughly 23 proton-proton interaction per bunch crossing. Given the low cross-section of hard collisions, the final state objects firing the CMS trigger are typically produced in a single scattering. It is therefore important to precisely identify the primary vertex (PV) among the several collision points in each event.

To reconstruct the PV, tracks with good quality parameters are selected and grouped based on the point of the closest approach to the beamline.
Each group of tracks forms a vertex seed. The vertex reconstruction is performed using an adaptive vertex fitter (AVF) [73]. The AVF procedure is similar to the KF algorithm, starting from the vertex seeds and iterating over the selected tracks. Each track entering the fitting procedure is associated with a weight encoding the compatibility between the track and the candidate vertex [74]. The vertex position is updated after each iteration applying a least-squares method to the weighted tracks and the procedure is iterated for all vertex candidates

in the event. The vertices are then ordered based on the $\sum p_T^2$ of the outgoing tracks, with the primary vertex being identified in the leading candidate [75].

4.3. Calorimeter clustering

As particle showers typically develop in several calorimeter units, a clustering algorithm is in place to group the different deposits belonging to the same object. These clusters are used in the PF algorithm to be linked with reconstructed tracks or identified as deposits from neutral particles. Each calorimeter subsystem runs independent clustering algorithms, apart from the HF as forward showers are extremely collimated. Cells registering energies twice as big as the noise threshold (80 MeV in the EB, 300 MeV in the EE, and 800 MeV in the HCAL) and above neighbouring ones are the input seeds initializing the clustering algorithms. Clusters are formed merging iteratively the cells fulfilling the threshold requirement to the most energetic neighbouring one.

The measured energy from the clusters is calibrated to reflect the energy of the incoming particles accurately. The calibrations are estimated independently in the electromagnetic and hadronic calorimeters. They are determined in several steps using test beam data, radioactive sources, cosmic rays, in situ collisions data and simulated events.

4.4. Particle flow

Hits of ionising particles in the tracking and muon systems together with energy deposits of particle showers are used to reconstruct muons, electrons and photons, and charged and neutral hadrons. The first step in the reconstruction, needed to provide object candidates to the PF method, consists of forming tracks of charged particles from hits in the tracker and the muon chambers and calorimeter clusters, obtained by combining energy deposits in the calorimeters. Starting from these ingredients, the PF algorithm combines pieces of information into a coherent description of particle candidates propagating through the detector. The compact design of the detector and the relatively small amount of inactive material between different subsystems allows for precision matching between tracks and energy clusters, with the calorimeter deposits linked to tracks by extrapolating trajectories to the ECAL and HCAL. A distance parameter quantifies the $\eta - \phi$ difference, and only the track minimizing the cluster distance is kept. Radiation photons are linked to electrons if their deposits are found compatible with tangent trajectories to GSF tracks. Photon conversions are identified with a dedicated algorithm, creating a link between the electron pairs and the photon. As an example, electrons are reconstructed by combining tracks and corresponding ECAL deposits, also taking into account secondary deposits consistent with Bremsstrahlung photons from the primary electron

track and combined with tracks coming from secondary vertices compatible with radiation photons converting to electron pairs. Calorimeter information is also linked across preshower ECAL and HCAL to form a unique object across a full calorimeter tower. In parallel, groups of at least three tracks compatible with secondary vertices are linked together and attributed to nuclear interactions. Once links are established, the PF algorithm implements corrections for reconstruction efficiencies, cleaning noise-originated object candidates and classification of particles. Apart from improving the reconstruction of particle properties, the PF algorithm ensures consistent estimates of event-related variables, such as the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation and lepton isolation.

4.4.1. Muons

Global muon tracks are propagated into the ECAL and HCAL volume, and clusters compatible with their trajectory are selected. The energy collected in these clusters is used together with tracker information for implementing isolation requirements on muon candidates. This is done to mitigate the rate of not confined charged hadrons misidentified as muons. In particular, the total energy in a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ centred around the muon trajectory must be less than 10% of the muon candidate p_{T}:

$$I_{\mathrm{track,calo}} = \frac{1}{p_{\mathrm{T}}} \left(\left| \sum_{\substack{i \in \mathrm{tracks} \\ \Delta R < 0.3}} \vec{p}_{\mathrm{T},i} \right| + \sum_{\substack{j \in \mathrm{clusters} \\ \Delta R < 0.3}} E_{\mathrm{T},j} \right) < 0.1 \,, \tag{4.2}$$

where $I_{\mathrm{track,calo}}$ is the isolation energy estimated by combining tracker and calorimeter information, the sum of the transverse momenta $\vec{p}_{\mathrm{T}}$ is done over all tracks within $\Delta R = 0.3$ and the sum over transverse energies E_{T} includes all clusters within $\Delta R = 0.3$.

In CMS reconstruction, non-isolated muons whose tracks and calorimeter footprints meet high quality criteria are also identified as muons. The isolation variable is included in the event content, and further decisions on such objects is dependent on analysis selections.

PF muons p_{T} is reconstructed from tracker hits up to $p_{\mathrm{T}} < 200\,\mathrm{GeV}$. For more energetic muons, to improve track curvature estimation, the p_{T} is taken from the best track fitting both tracker and muon system hits. After identification and reconstruction of all PF muon candidates, the corresponding PF blocks are excluded from further object identification.

4.4.2. Electrons and isolated photons

As described in Sec. 4.1.1, electron and photon reconstruction must consider Bremsstrahlung photons and electron-positron pairs from photon conversions to properly measure the energy

of the particle candidate. These secondary particles are linked to a seed candidate, i.e. a photon candidate, consisting of an ECAL supercluster with energy above 10 GeV and no GSF track link, or an electron candidate, which is made by a GSF track associated to an ECAL cluster. In both cases, it is required that HCAL clusters within $\Delta R = 0.15$ of the candidate should register an energy deposit less than 10% of the ECAL supercluster energy. The candidate's calorimeter energy is corrected adding all the energies of the linked deposits and calibrating the detector response. For electron candidates, it is also combined with GSF p_{T} estimation.

To further reduce misreconstruction, additional identification criteria are in place. For photon candidates, calorimeter clusters need to be isolated from tracks, other than conversion pair ones, and the HCAL over ECAL energy ratio must be compatible with the average photon shower signature. For electrons, instead, a BDT is used. The features considered in training are radiation energy from the GSF track, the compatibility between track and calorimeter cluster, HCAL to ECAL energy ratio, number of hits in the tracker and track χ^2. The BDT is trained independently for the EB and EE electrons and isolated and non-isolated candidates.

As for muons, once the electromagnetic objects are fully reconstructed, the PF blocks linked with the candidates are removed.

4.4.3. Hadrons and non-isolated photons

After fragmentation and hadronisation, the bulk of hadronic objects interacting with the detector are made of charged and neutral mesons, namely $\pi^{\pm}$, $K^{\pm}$ and K^0_L together with protons and neutrons. The π^0 lifetime is extremely short and, therefore, the lightest neutral mesons are reconstructed as non-isolated and highly collimated photon pairs. More rarely, also non-isolated muons can be produced in prompt hadron decays. The CMS design does not include systems dedicated to hadronic discrimination. Therefore the PF identification is limited to charged and neutral hadrons and electromagnetic showers.

PF hadrons are reconstructed after muons, electrons and isolated photons. Therefore all remaining ECAL clusters within the tracker acceptance and without matched tracks are reconstructed as non-isolated photons, and all HCAL clusters not linked to tracks are labelled as neutral hadrons. This identification is easily implemented as both tracker and calorimetry information is available ($|\eta| < 2.4$). From $|\eta| = 2.4$ up to the ECAL acceptance ($|\eta| < 3.0$), electromagnetic clusters linked to HCAL ones are assigned to unknown charged hadrons, whereas non-linked ECAL deposits are reconstructed as non-isolated photons. Showers developing in the HF ($|\eta| < 5.0$) are identified as electromagnetic or hadronic activities depending on their depth and lateral development.

After calorimetric deposits are calibrated, charged hadrons candidates are cross-checked comparing the track p_{T} and the sum of HCAL and ECAL energy. Larger calorimeter energy

over track-based p_{T} estimation might be due to neutral objects superposed with charged hadrons. The difference between the calorimeter and tracker energy, δ_{T}, is defined as follows:

$$\delta_{\mathrm{T}} \equiv E_{\mathrm{T}} - |\sum \vec{p}_{\mathrm{T}}| \,, \tag{4.3}$$

where $E_{\mathrm{T}} = E_{\mathrm{T}}^{\mathrm{ECAL}} + E_{\mathrm{T}}^{\mathrm{HCAL}}$ and $|\sum \vec{p}_{\mathrm{T}}|$ is the sum of track momenta linked to the calorimeter clusters. Different corrections to the reconstruction are implemented, depending on δ_{T} size with respect to the expected calorimeter resolution and relative contribution of ECAL and HCAL to the total energy. If δ_{T} is smaller then E_{T}^{ECAL}, but not compatible with expected resolution effects, the energy difference is reconstructed as a photon deposit. In contrast, for larger δ_{T}, the extra ECAL energy is still reconstructed as a photon, but the HCAL contribution is assigned to a neutral hadron. Non isolated muons are characterized by an opposite sign difference and tracker based p_{T} larger than calorimeter deposits. Such leptons are, however, rarely produced in hadron decays. Any remaining difference is associated with misreconstructed tracks.

4.4.4. Jets

Outgoing quarks and gluons from the hard scattering form collimated jets of hadrons. The properties of the parton originating the jet can be reconstructed if all the particles from its fragmentation and hadronization are grouped. To reconstruct such objects, the anti-k_{T} algorithm [76, 77] is used to cluster PF candidates into cones of radius $R = 0.4$. The usage of PF objects instead of purely calorimeter deposits improves jet reconstruction thanks to the complementary energy and position information provided by the tracker. To reject jets that are misreconstructed, additional selections are implemented on the jet energy contribution and the number of PF constituents. Moreover, the minimum clustering threshold is set to 15 GeV, as softer objects do not provide a reliable estimate of the originating parton momentum.

Due to detector inefficiencies and inhomogeneities, reconstructed jet energy can differ from the energy of the original partons. In order to mitigate this effect, a complex procedure is implemented and it is parametrized as:

$$p_{\mu}^{\mathrm{corr}} = C_{\mathrm{offset}}(p_{\mathrm{T}}^{\mathrm{raw}}) \cdot C_{\mathrm{rel}}(\eta) \cdot C_{\mathrm{abs}}(p_{\mathrm{T}}^{'}) \cdot C_{\mathrm{res}}(p_{\mathrm{T}}^{''}, \eta) \cdot p_{\mu}^{\mathrm{raw}} \,, \tag{4.4}$$

where C_{offset}, C_{rel}, C_{abs} and C_{res} are respectively the offset, relative, absolute and residual corrections, whereas p_{T}^{raw} , $p_{\mathrm{T}}^{'}$, $p_{\mathrm{T}}^{''}$ and p_{T}^{corr} is the p_{T} before and after each correction step. The different correction steps are [78]:

- C_{offset}: corrects for the energy contributions coming from pileup. Charged hadrons whose tracks are not compatible with the PV are subtracted (charged hadrons subtraction, CHS) before jets are formed. To remove neutral components of pileup, a method called

Effective Area (EA) is implemented. The average energy density ρ is multiplied by the A_{eff} parameter, which depends on the object footprint.

- C_{rel}: used to make uniform the calorimeter response in each region of the detector. The correction is derived from simulation.
- C_{abs}: used for correcting calorimeter energy as a function of p_{T}. This correction is derived using Z+jets and γ+jets events.
- C_{res}: implemented to remove any residual difference between simulation and data, and it depends on both p_{T} and η. This correction is applied only to data.

4.4.5. Missing transverse momentum

As the PF algorithm fully reconstructs event objects, determining $p_{\text{T}}^{\text{miss}}$ is fairly straightforward and is defined as:

$$\vec{p}_{\text{T,raw}}^{\,\text{miss}} = -\sum_i \vec{p}_{\text{T},i} \,, \tag{4.5}$$

summing over the PF candidates. However, different detector effects, errors in measuring PF candidates p_{T} and reconstruction inefficiencies can lead to artificially large $p_{\text{T}}^{\text{miss}}$ estimation. This constitutes a challenging background to standard model and new physics processes which produce genuine $p_{\text{T}}^{\text{miss}}$ in the final state, and it is thus crucial to keep this effect under control. For this reason, the following set of filters dedicated to identify and suppress anomalous high-$p_{\text{T}}^{\text{miss}}$ events are used [79]:

- **HCAL filters** In the HCAL noise in the readout electronics or the photo-diodes can result in spurious energy. In addition, particle interactions with the HF photomultipliers might lead to the same result. To mitigate such effects, the sensors' geometrical patterns and the signal spike shape and timing are used by noise filter algorithms. An isolation-based filter implements a topological algorithm that combines HCAL deposits to ECAL and tracking information, aiming to identify isolated anomalous activity in HCAL.
- **ECAL filters** In the ECAL, crystals reporting anomalous high energy deposit and channels for which readout electronics fail are identified with dedicated noise filters and their energies are not considered for reconstruction.
- **Beam Halo filter** Beam halo components can cause large $p_{\text{T}}^{\text{miss}}$ as such particles travelling close to the collision point can interact with the detector, leaving energy deposits across the calorimeter, at a constant ϕ. If such particles interact with the CSC, they can leave a muon segment inline with calorimeter deposits, being identified as muons. The beam halo

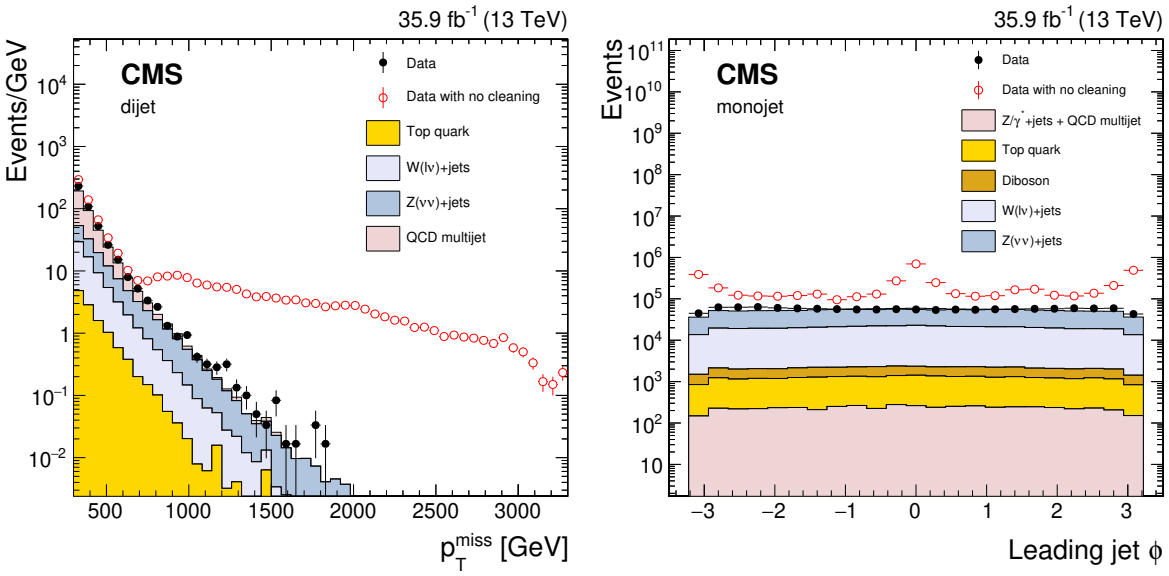


Figure 4.2.: The $p_{\mathrm{T}}^{\mathrm{miss}}$ (left) and jet ϕ (right) distributions for dijet (left) and monojet (right) selections with the event filtering algorithms applied (filled markers), without the event filtering algorithms applied(open markers), and from simulation (solid histograms) [79].

filter, combining calorimeter and CSC information, identifies and rejects such spurious objects.

- **Reconstruction filters** To improve $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, particles that have the largest effect on the magnitude and direction of the $p_{\mathrm{T}}^{\mathrm{miss}}$ vector are reconstructed for a second time. Such particles are mainly high p_{T} muons. This reconstruction revision consists of cross-checking for cosmic ray contamination, refitting the tracks selecting the resulting p_{T} that minimises the event $p_{\mathrm{T}}^{\mathrm{miss}}$ and cross-checking for overlaps with hadronic activity or hadrons misidentified as muons. After these steps, the relevant PF objects are updated, and the event $p_{\mathrm{T}}^{\mathrm{miss}}$ is recomputed.

Fig. 4.2 shows a comparison of $p_{\mathrm{T}}^{\mathrm{miss}}$ and jet ϕ distributions before and after the event filters. The filters efficiency on spurious high $p_{\mathrm{T}}^{\mathrm{miss}}$ events is around 85–90% with a mistag rate of less than 0.1%.

Chapter 5.

Event selection and simulation

The measurements presented in this thesis rely on the reconstruction of events containing the Z and γ bosons in association with jets. The analysis has been carried out using proton-proton collision data at $\sqrt{s} = 13$ TeV taken by the CMS experiment in 2016 only, and corresponding to an integrated luminosity of 35.9fb^{-1} [80]. The Z boson candidates are reconstructed using the $Z \to \mu\mu$ decay channel, therefore final states with two muons and jets are selected. The limited person-power, the lower reconstruction efficiency and the higher background contamination resulted in the pragmatic choice of not exploiting the electron channel for Z boson reconstruction For the γ + jets channel, the presence of one isolated photon in association with jets is required.

This chapter describes the data selections targeting the final states of interest. Details are provided on particle identification criteria starting from the PF objects and the definition of the fiducial regions for the measurement.

5.1. Samples

5.1.1. Data collection

As described in Sec. 3.2.6.2, the data satisfying at least one HLT path are reconstructed and stored, adding the fired triggers to the events information. The quality of the reconstructed data is monitored online and offline analysing several event and object level variables, using simulations as reference. Events presenting pathological distributions cannot be used for physics analysis, and they are vetoed. This analysis uses certified events, belonging to the so-called "golden JSON" 5.1, which is a list containing all the events satisfying the data quality requirements, evaluated using the latest detector calibrations and alignment for 2016.

Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt

Table 5.1.: List of certified quality data for the 2016 run, used in the analysis.

The reconstruction of Z boson candidates is performed using a combination of HLT paths targeting single muon events: events are selected applying a logical "OR" between `HLT_IsoMu24` and `HLT_isoTkMu24`. These paths require the presence of a global muon with $p_{\mathrm{T}} > 24$ GeV or a tracker muon satisfying the same requirement. Besides the p_{T} selection, the muon candidates need to fulfil an isolation requirement that combines calorimeter tracker information.

To reconstruct events containing a prompt isolated photon, events firing the `HLT_Photon175` path are selected. This HLT path requires at least one isolated HLT photon, with $\mathrm{E_T} > 175$ GeV. The isolation requirement is implemented at HLT level comparing the energy deposits in the ECAL and HCAL, with accepted candidates fulfilling $(E^{HCAL}/E^{ECAL}) < 0.15$. The chosen HLT path guarantees high reconstruction efficiency for prompt photons with energies above 190 GeV. Further details on the `HLT_Photon175` efficiency measurement are described in Sec. 3.2.6.

5.1.2. Simulated event samples

Several simulated samples, exploiting different event generators, are used in the analysis to model signal and background events as well as to provide theoretical predictions that are compared with the measurement presented in this thesis. The full list of simulated event samples used in this analysis is summarised in Tables 5.3 and 5.2.

For the Z boson analysis the signal process of Drell-Yan ($q\bar{q} \rightarrow \mathrm{Z}/\gamma^* \rightarrow \ell^+\ell^-$) + jets is generated using the matrix element (ME) generator MADGRAPH5_aMC@NLO v2.2.2 [81, 82]. The matrix element includes Z+0,1,2 jets at next-to-leading order (NLO) in the strong coupling constant. The event production is split into different bins of the vector boson p_{T}. This is done to guarantee an adequate event population of the high p_{T} regions, which are the focus of the analysis. The simulation of the parton showering (PS) and hadronization is performed by interfacing the ME final state partons with the PYTHIA 8.2 [83] program, using the CUETP8M1 underlying event description. The FxFx merging and matching technique [84] is used to combine ME partons and the PS jets and avoid possible double counting of jets in the final state. Several background processes leading to final states similar to the signal one are considered. W + jets, tt + jets, WZ, ZZ $\rightarrow 2\ell 2Q$, ZZ $\rightarrow 2\nu 2Q$ and ZZ $\rightarrow 4Q$, s-channel production of single top, ttW and ttZ events are also generated at NLO using MADGRAPH5_aMC@NLO v2.2.2. WW and ZZ$\rightarrow 4\ell$ events as well as t-channel and tW-channel single top and ttH samples are generated at NLO using the POWHEG [85, 86, 87] generator. Lastly, the MADGRAPH5_aMC@NLO v2.2.2 code is also used at LO to generate W + jets and Z + jets processes where the vector

boson is produced via vector boson fusion (also referred to as EWK V+jets events). As for the signal simulation, all the background samples parton showering simulation is performed with PYTHIA 8.2. NLO samples adopted the FxFx merging scheme, whereas LO ones use MLM [88].

The γ + jets signal process is also generated using the MADGRAPH5_aMC@NLO v2.2.2 generator where the matrix element includes γ+0,1 jets at NLO in α_s. The same PYTHIA 8.2 tune is used for the modelling of the underlying event, and the matching and merging is performed with the FxFx approach. The background for this final state is composed of: QCD multi-jet events, which are modelled with MADGRAPH5_aMC@NLO v2.2.2 at LO, ttγ processes which have been generated with POWHEG interfaced with PYTHIA 8.2, and V+γ+jets events generated at NLO with MADGRAPH5_aMC@NLO v2.2.2.

Additional signal samples, generated at LO for both Z + jets and γ + jets, are used to perform cross-checks on simulated event variables as well as providing independent theoretical predictions to compare with the measurements. These samples are also generated at LO using MADGRAPH5_aMC@NLO v2.2.2 with up to four hard partons in the matrix element. Comparisons of the data are also made to γ + jets samples generated using SHERPA v2.2.2, which includes 0,1,2 jets in the matrix element.

All the samples mentioned so far are fully reconstructed to detector level, with the detector response simulated using the GEANT4 [89] package.

The Z + jets and γ + jets processes are also generated at particle level using SHERPA v2.2.8 + OPENLOOPS v2.0 (v2.1.0) [90] with a matrix element calculation for up to 2 additional partons at NLO in QCD and up to 4 partons at LO in QCD and the approximate NLO EW calculation using the Comix [91] and OPENLOOPS v2.0 [36] matrix element generators. This is merged with CSSHOWER [92], the default parton shower in SHERPA v2.2.8, using the ME-PS matching implemented according to the aMC@NLO method [93, 94]. Moreover, the γ + jets process is simulated at the particle level using the JETPHOX generator as well [95]. This program includes the full NLO QCD calculation of both direct and fragmentation contributions to the photon+jet cross-section. The NLO parton-to-photon fragmentation functions BFG set II [96] are used, and five massless quark flavours are considered.

The cross-section of particular samples are rescaled to NNLO and NLO calculations performed with FEWZ (v3.1) [97] Tab. 5.3 and Tab. 5.2 report both the generator and the final cross-section precision for each sample.

Both the NNPDF3.0 LO and NLO [98] PDFs are used across all the samples used in this analysis, with the LO and NLO generators as described above.

Table 5.2.: Simulated event samples for γ + jets signal and background processes. Where not specified, the precision refers to the calculation order in α_s. Samples processed simulating the CMS reconstruction are marked accordingly. The native cross-section of some samples has been scaled to match higher order theoretical predictions obtained with FEWZ, and the final cross-section normalisation precision is also reported.

process	generator	precision	reco	normalisation
γ + jets	MADGRAPH5_aMC@NLO v2.2.2 γ+0,1 jets + PYTHIA 8.2	NLO QCD	✓	NNLO
	SHERPA v2.2.8 γ+0,1,2 jets NLO 3,4 LO	NLO QCDxEW		NLO
	MADGRAPH5_aMC@NLO v2.2.2 γ+0,1,2,3,4 jets	LO	✓	NNLO
	JETPHOX γ+0,1,2	NLO		NLO
	SHERPA v2.2.2	LO	✓	NLO
Vγ+jets	MADGRAPH5_aMC@NLO v2.2.2 + PYTHIA 8.2	NLO	✓	NLO
$t\bar{t}\gamma$	MADGRAPH5_aMC@NLO v2.2.2 + MADSPIN + PYTHIA 8.2	NLO	✓	NLO
W + jets	MADGRAPH5_aMC@NLO v2.2.2 + PYTHIA 8.2	NLO	✓	NNLO
QCD	MADGRAPH5_aMC@NLO v2.2.2 + PYTHIA 8.2	LO	✓	LO

Table 5.3.: Summary of the simulated samples used in the Z+jets analysis. Where not specified, the precision refers to the calculation order in α_s. Samples processed simulating the CMS reconstruction are marked accordingly, and cross section normalisation precision is highlighted.

process	generator	precision	reco	normalisation
DY+jets	MADGRAPH5_aMC@NLO v2.2.2 Z+0,1,2 jets +PYTHIA 8.2	NLO	✓	NNLO
	SHERPA v2.2.8 Z+0,1,2 jets (NLO) 3,4 (LO)	NLO QCDxEW		NLO
	MADGRAPH5_aMC@NLO v2.2.2 Z+0,1,2,3,4 jets + PYTHIA 8.2	LO	✓	NNLO
Diboson	MADGRAPH5_aMC@NLO v2.2.2 + MADSPIN + PYTHIA 8.2	NLO	✓	NLO
	POWHEG	NLO	✓	NNLO
VBF Z+2 jets	MADGRAPH5_aMC@NLO v2.2.2 + PYTHIA 8.2	LO	✓	LO
W + jets	MADGRAPH5_aMC@NLO v2.2.2 + PYTHIA 8.2	NLO	✓	NNLO
$t\bar{t}$	MADGRAPH5_aMC@NLO v2.2.2 + PYTHIA 8.2	NLO	✓	NNLO+NNLL
ttV	MADGRAPH5_aMC@NLO v2.2.2 + MADSPIN + PYTHIA 8.2	NLO	✓	NLO
	MADGRAPH5_aMC@NLO v2.2.2 + PYTHIA 8.2	LO	✓	LO
Single top	MADGRAPH5_aMC@NLO v2.2.2 + PYTHIA 8.2	NLO	✓	NLO
	POWHEG + PYTHIA 8.2	LO	✓	NLO
	POWHEG + MADSPIN + PYTHIA 8.2	LO	✓	NLO
ttH	POWHEG + PYTHIA 8.2	LO	✓	LO

5.2. Physics objects

The physics objects, which undergo the reconstruction procedure described in Sec. 4 are further selected, applying isolation and identification requirements. This procedure is needed to target the processes of interest for the analysis while providing excellent background rejection.

5.2.1. Jets

Jets are defined as particle-flow (PF) candidates clustered by the anti-k_T jet clustering algorithm [76] with a distance parameter of 0.4. The four-momenta of jets, which are defined as the four-vector sum of the four-momenta of the constituent particle-flow candidates, are corrected to match the particle-level objects' energy. Further details on the correction procedure are discussed in Sec. 5.3.4. To mitigate the effect of pileup, the Charged Hadron Subtraction (CHS) algorithm is also applied, excluding from the PF jet constituents charged hadrons that can be traced back to pileup vertices.

Several sets of requirements on PF jets are defined, known as working points, to provide different selection efficiency and background rejection rates. The "loose" working point Jet-Id selection criterion is chosen, and the cuts characterizing it are listed in Tab. 5.4. In addition, working points are also defined for the identification of jets arising from pileup interactions. To identify effectively PF jets not compatible with the interaction vertex, we require the "loose" pileup ID working point on all jets with p_{T} <50 GeV. This working point is defined on the output score of a boosted decision tree (BDT) built from tracking and hadronic shower variables.

Table 5.4.: The "loose" jet ID requirements.

Selection variable	value	notes
$-3.0 < \eta_{\mathrm{jet}} < 3.0$		
Neutral Hadron Fraction	< 0.99	-
Neutral EM Fraction	< 0.99	-
Number of constituents	> 1	-
Charged Hadron Fraction	> 0	only for $\lvert\eta_{\mathrm{jet}}\rvert < 2.4$
Charged Multiplicity	> 0	only for $\lvert\eta_{\mathrm{jet}}\rvert < 2.4$
Charged EM Fraction	< 0.99	only for $\lvert\eta_{\mathrm{jet}}\rvert < 2.4$
$\lvert\eta_{\mathrm{jet}}\rvert > 3.0$		
Neutral EM Fraction	< 0.90	-
Number of Neutral Particles	> 10	-

5.2.2. Muons

Muons are selected using the "tight" working point, which aims to suppress muons from decay in flight and from high p_{T} hadrons punching-through the HCAL and the solenoid. Tab. 5.5 summarizes the cuts implemented by the selected working point. Muons are also required to be well isolated, i.e. with low activity in their track's close vicinity. The transverse momenta of PF neutral and charged candidates, as well as photons, lying within a cone around the lepton, are summed. The PF relative isolation is combined with $\Delta\beta$ corrections to remove the effects of

the pileup. The relative combined isolation I^{rel}_{comb} is then defined as follows:

$$I^{rel}_{comb} = \frac{1}{p_{\mathrm{T}}(\mu)} \left(\sum p_{\mathrm{T}}(\text{ch Had from PV}) + \right. \\ \left. + max(0, \sum \mathrm{E_T}(\text{neu Had}) + \sum \mathrm{E_T}(\text{photon}) - 0.5 \cdot \sum p_{\mathrm{T}}(\text{ch Had from PU}))\right) , \quad (5.1)$$

where the factor of 0.5 results from the average of neutral to charged particles. Muons are defined to be isolated if they fulfill the criterion $I^{rel}_{comb} < 0.15$ [99].

Table 5.5.: Object definition for muons used in the analysis. The cuts shown here correspond to the "Tight" working point definition.

Selection variable	value
Reconstructed as Global Muon	True
Particle-Flow muon	True
χ^2/ndof of global muon track fit	< 10
No. of Matched Stations	> 1
No. of muon chamber hits in global-muon track fit	> 0
Transverse Impact parameter w.r.t primary vertex d_{xy} (cm)	< 0.2
Longitudinal distance of tracker track w.r.t primary vertex d_Z (cm)	< 0.5
Number of pixel hits	> 0
Number of tracker layers with hits	> 5
Combined relative isolation I^{rel}_{comb}	< 0.15

5.2.3. Photons

Photons are identified according to the "tight" working point definition (which gives ~ 71 % efficiency) of the cut-based photon identification algorithm [100] and required to be well isolated. PF-based isolation is used with a cone size $\Delta R < 0.3$ after corrections are applied to remove the effects of pileup. PF corrected isolation is defined as:

$$\mathrm{PF}^{corr}_{Iso} = max(\mathrm{PF}_{Iso} - \rho \times A_{eff}\ , 0)\ , \quad (5.2)$$

where ρ is the event energy density and A_{eff} is the "effective area", a pre-calculated parameter that depends on the isolation cone pseudorapidity and the type of PF object considered.

Table 5.6 summarises the identification and isolation selection used.

Table 5.6.: Photon identification selection.

Selection variable	Barrel	Endcap
Conversion safe electron veto	Yes	Yes
H/E	0.0269	0.0213
$\sigma_{i\eta i\eta}$	0.00994	0.03
PF charged hadron isolation	0.202	0.034
PF neutral hadron isolation	$0.264 + 0.0148 \times p_{\mathrm{T},\gamma} + 0.000017 \times p_{\mathrm{T},\gamma}^2$	$0.586 + 0.0163 \times p_{\mathrm{T},\gamma} + 0.000014 \times p_{\mathrm{T},\gamma}^2$
PF photon isolation	$2.362 + 0.0047 \times p_{\mathrm{T},\gamma}$	$2.617 + 0.0034 \times p_{\mathrm{T},\gamma}$

5.3. Corrections to simulated samples

To improve the data description provided by simulated samples, corrections to the MC events are applied. Such corrections are typically expressed as scale-factors that are multiplied to either the event weight or a particular object, with the event weight being determined by the calculated process cross-section and the integrated luminosity of the data. The corrections described in the following paragraphs introduce systematic uncertainties which are propagated to the analysis results. The impacts of the different systematic sources due to simulation corrections are discussed in Sec. 7.2.

5.3.1. Pileup

To model the effects of multiple pp collisions within the same or neighbouring bunch crossings (pileup), all simulated events are generated with a nominal distribution of pp interactions per bunch crossing, which differs from data conditions. The simulated event samples are then reweighted to match the pileup distribution as measured in the data. This procedure is called *pileup reweighting*.

The reweighting factors are a function of the variable called `nTrueInt`, which is the parameter of the Poisson distribution representing the expected distribution of interactions per bunch crossing. In each simulated event, the number of pileup interactions is modelled by a Poisson distribution stitched to a polynomial distribution for higher vertex multiplicities. In the data, the `nTrueInt` is measured independently in different sets of runs characterized by stable luminosity conditions, named "lumi sections". The `nTrueInt` in the data is derived from the measured instantaneous luminosity for each colliding bunch pair in each lumi section and the cross-section of the total inelastic pp interaction. The latter is measured by CMS to be 68.6 mb [101] for 2016 conditions, with an uncertainty of $\pm 5\%$.

The pileup reweighting factors are the ratios of the distributions of `nTrueInt` in the data and the simulated events, with the overall sum of weights being normalized to preserve the

number of the simulated events. The same procedure is repeated for the $\pm 1\sigma$ variations of the minimum bias cross-section, i.e., 72.03 mb and 65.17 mb, to estimate the uncertainty associated with the reweighting procedure.

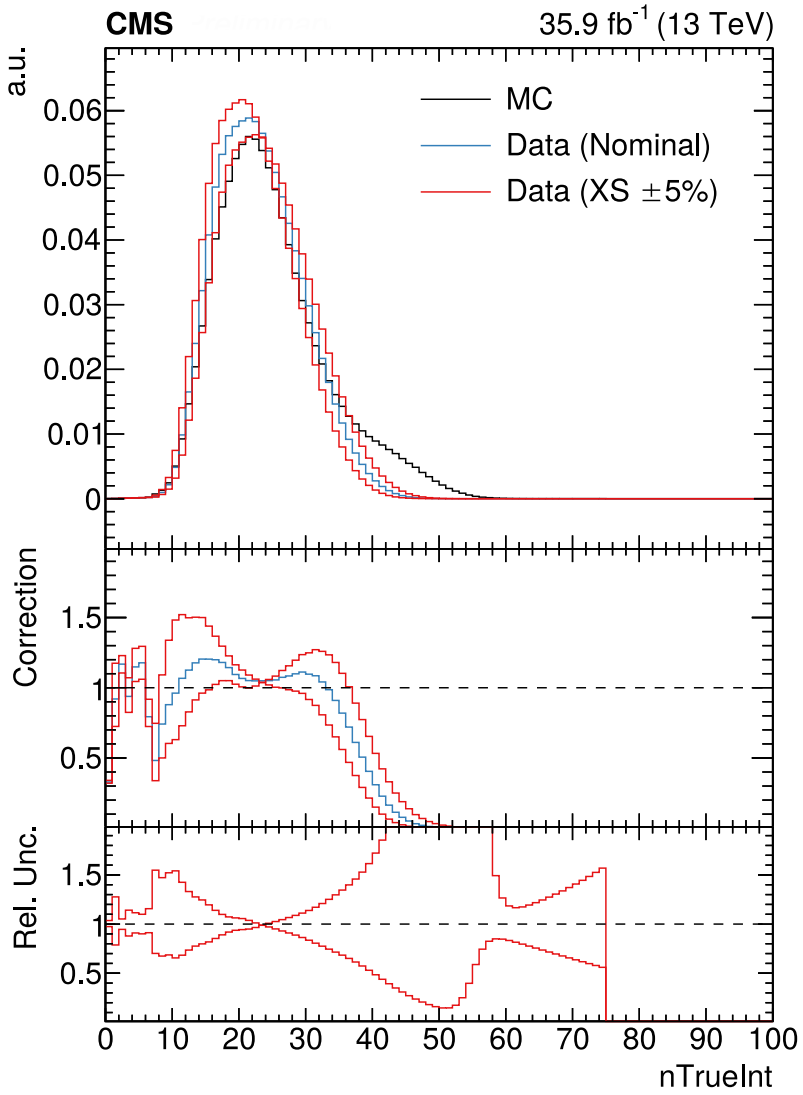


Figure 5.1.: The distributions of the average number of inelastic interactions per colliding bunch pair in simulated events and three data scenarios with the nominal minimum bias cross-section and the up/down variations by 5% (top panel). The data distributions ratio to the MC distribution gives the correction required for MC (middle panel). The nominal correction is set to unity to show the relative uncertainty from varying the minimum bias cross section (bottom panel).

Figure 5.1 shows the distributions of `nTrueInt` in data and simulation. In simulated samples, events with `nTrueInt` larger than 74 are not included, whereas the data distribution goes beyond that. However, the fraction of events in data above 74 interactions per bunch crossing is negligible.

5.3.2. Lepton and photon scale factors

Simulation mismodelling of efficiencies related to object identification and isolation requirements is mitigated by using scale factors (SF). The technique used to measure efficiencies both in data and simulations is the so-called "tag-and-probe" method [102]. Using Z decays to leptons, candidate pairs compatible with the Z boson's invariant mass are selected. In each pair, one of the leptons is required to fulfil a set of tight restrictions, hence "tagging" a genuine $Z \to ee$ or $Z \to \mu\mu$ event. In contrast, the other (the probe) constitutes an unbiased object which can be used as a reference for measuring efficiencies of particular selections. Corrections for photons are also derived using tag and probe $Z \to ee$ samples, exploiting the similar signature left in the ECAL by electromagnetic objects.

5.3.2.1. Muon scale factors

For muons, SF are derived on a $Z \to \mu\mu$ sample, selecting pairs fulfilling $70\,\text{GeV} < m_{\mu\mu} < 130\,\text{GeV}$, and applying to the tag muon a p_T threshold of 26 GeV and tight isolation and identification requirements, whereas the probe has only a lower p_T requirement of 20 GeV.

As muons are reconstructed from tracks with identification and isolation requirements applied in succession, the efficiency in selecting muons can be factorised as follows:

$$\varepsilon_\mu = \varepsilon_\mu(\text{track})\varepsilon_\mu(\text{ID}|\text{track})\varepsilon_\mu(\text{Iso}|\text{ID}) \,, \tag{5.3}$$

where parameter $\varepsilon_\mu(\text{track})$ is the track reconstruction efficiency, $\varepsilon(\text{ID}|\text{track})$ is the identification efficiency for a given set of tracks and $\varepsilon(\text{Iso}|\text{ID})$ is the isolation efficiency given a collection of muons passing identification requirements.

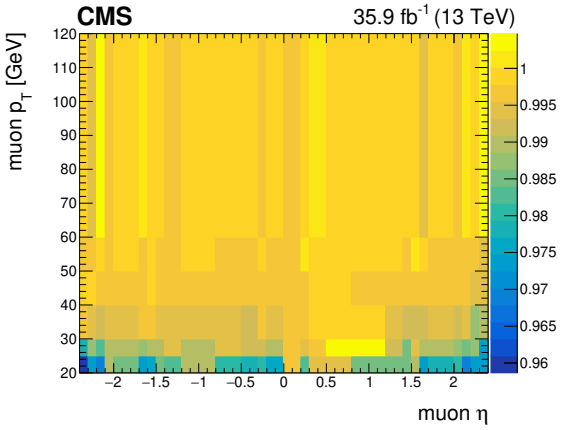
Efficiencies are measured in bins of p_T and η, and a fit to the invariant mass distribution is performed in each bin to extract the signal component in data, whereas for MC the signal component is directly extracted using the generator level identification information [99].

It has to be noticed that due to different performances in the pixel detector between the initial months and the final period of the 2016 data-taking campaign, muon ID performance is significantly different as a function of time. To account for this behaviour independent SFs have been derived, and the overall weight applied to simulation events is an average of the two SFs. The event weight derived from the scale factors of the selected muons is obtained by averaging over the different eras using the integrated luminosities corresponding to the initial and final data-taking conditions as in the expression below:

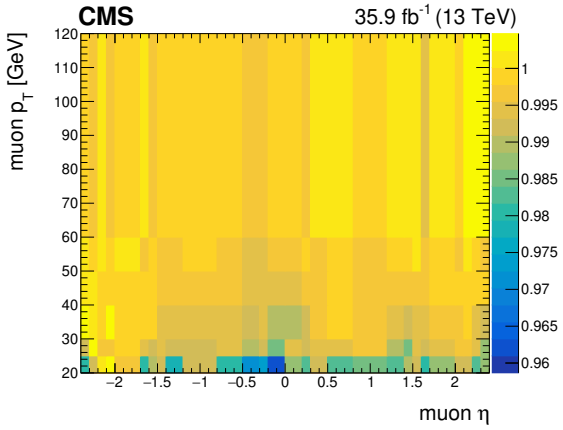
$$w_{\mu\mu} = (\int \mathcal{L}^{initial} \cdot w_{\mu\mu}^{initial} + \int \mathcal{L}^{final} \cdot w_{\mu\mu}^{final}) / (\int \mathcal{L}^{initial} + \int \mathcal{L}^{final}) \,, \tag{5.4}$$

where $w_{\mu\mu}$ is the weight obtained multiplying the SF for each muon selected in the event.

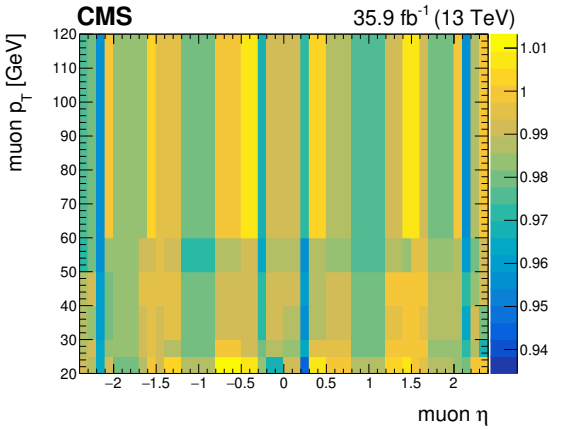
Figure 5.2 shows the SF for the tight identification working point and the tight isolation requirement, measured in bins of p_T and η for the two data-taking periods. The sources of uncertainties on the SF measurements are given by the statistical significance of the bins considered and the systematic uncertainties related to the invariant mass fit procedure. The overall uncertainty is propagated through the analysis results and its impact on the final results is discussed in Section 7.2.



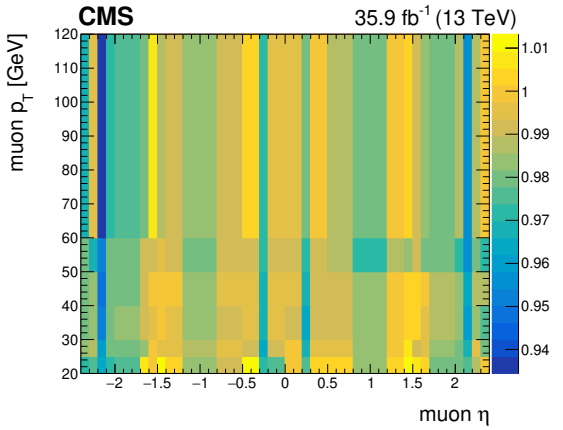
(a) Initial data-taking period



(b) Final data-taking period



(c) Initial data-taking period



(d) Final data-taking period

Figure 5.2.: Muon identification (id) scale corrections applied to MC, parameterised with the muon p_T and $|\eta|$. The top row shows the scale-factors for veto muons with corrections less than 1% as a result of the basic selection. The bottom row shows the scale-factors for selection muons with corrections from 1–4% with the largest discrepancies on the edges of the endcaps ($0.9 < |\eta| < 1.2$ and $2.1 < |\eta| < 2.4$) where the muon identification modelling does not perform as well in MC.

5.3.3. Photon scale factors

Photon corrections are derived using a sample of electrons from Z decays [103]. In this case, tag and probe objects are selected in a 60-120 GeV mass window, with a p_{T} threshold of 35 GeV on the tag electron together with tight identification and isolation requirements and loose identification requirements and a p_{T} threshold of 10 GeV for probes.

Photons are reconstructed from ECAL deposits to which isolation requirements are applied, therefore the photon efficiencies can be parametrised into:

$$\varepsilon_{\gamma} = \varepsilon_{\gamma}(\mathrm{reco})\varepsilon_{\gamma}(\mathrm{ID}-\mathrm{Iso}|\mathrm{reco}) \ , \tag{5.5}$$

where the $\varepsilon_{\gamma}(\mathrm{reco})$ term is the efficiency of reconstruction, which is considered 100% for superclusters presenting H/E<0.5, whereas the term $\varepsilon_{\gamma}(\mathrm{ID}-\mathrm{Iso}|\mathrm{reco})$ represents the selection efficiency. Similarly to the muon case, the efficiency is measured in bins of p_{T} and η, as reported in Figure 5.3, and the SF is derived from the ratio of the efficiency measured in data and MC. The uncertainty on the photon SF is propagated as systematic error to the measured cross-section; details on the procedure followed and the impact on the results are discussed in Section 7.2.

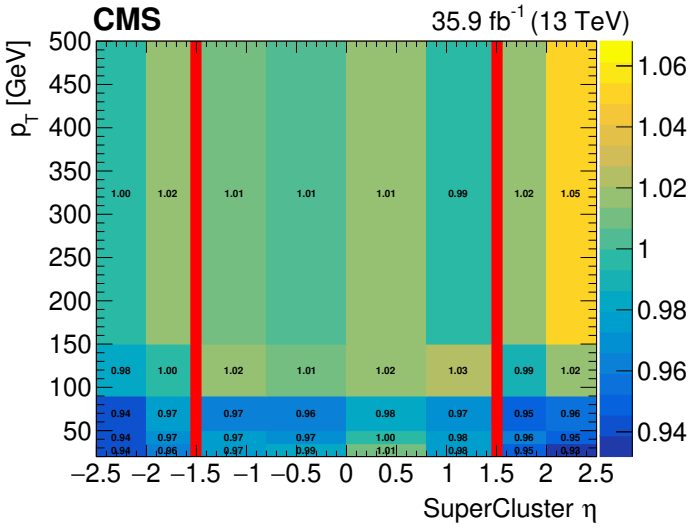


Figure 5.3.: Photon identification scale-factors for the working point that is used in the analysis. The bands at $1.444 < |\eta| < 1.566$ correspond to the partially uninstrumented gap in the ECAL. The corrections typically arise from the mismodelling of electromagnetic particles' interactions with the inner detector material resulting in photon conversions and Bremsstrahlung.

5.3.4. Jet corrections

Jets are particularly challenging objects for simulation due to their composite nature and the non-linear detector response to clustered particles. A complex set of corrections is applied to map the reconstructed jet energies to their particle level equivalents, originating from hard scattering. A three-level correction process is implemented, with each step addressing a different effect of reconstruction [104].

The first step aims to estimate the contribution of particles coming from pileup interactions that fall within the jet cone, causing an offset in energy estimation. To estimate this contribution, a method known as "random cone" is implemented using zero bias datasets (i.e. randomly selected events before the trigger decision), and MC events generated by switching off hard interactions. The idea is to measure the average contribution from pileup events, clustering particles falling inside cones constructed around a random direction. This energy contribution is estimated for each type of PF particle, parametrised to the number of interactions per bunch crossing μ and jet kinematics. For each analysis event, such contributions are subtracted from the reconstructed jets energies, and an SF is defined to map simulations to data. Figure 5.4 shows the energy offset due to pileup for different PF objects. A systematic uncertainty is determined by comparing the correction from the random cone method to the generator-level in MC.

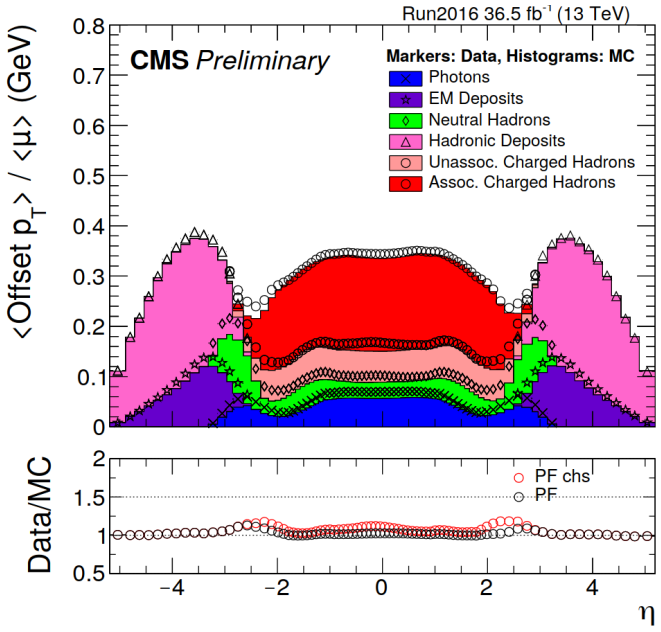


Figure 5.4.: Data/MC comparison for the average offset per additional pileup interaction (μ), which is calculated for each type of PF particle [105].

The second step involves the correction for the detector response to jet p_T to match to particle-level momentum. This is determined from QCD dijet simulations by comparing the

reconstructed jets to particle-level ones, deriving corrections parameterised by the jet p_T and η. Figure 5.5 reports the response correction for different jet kinematics, and it can be observed how detector response is stable in the barrel region, $|\eta| < 1.3$. A systematic uncertainty on this procedure is derived from its dependence on the underlying detector calibrations using test beam studies results.

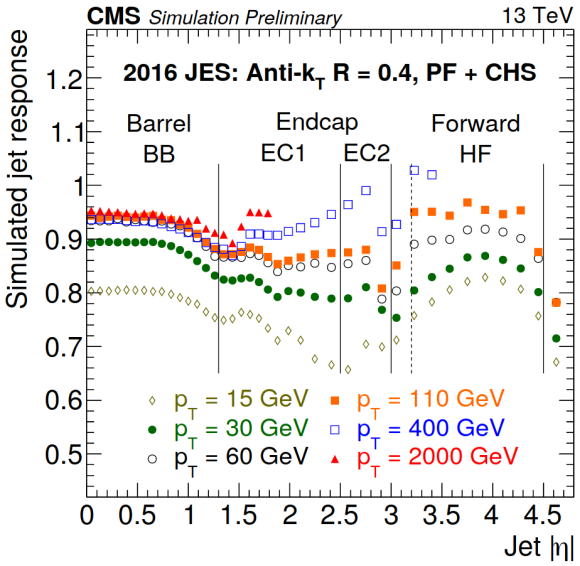


Figure 5.5.: The detector response to different jet kinematics, obtained in simulation. A stable response is found in the barrel, with a drop in energy reconstruction for $p_T < 30$ due to HCAL acceptance. Strong p_T dependence is observed in HE (in both EC1 and EC2) and HF regions [105].

The third and final step resolves residual differences between simulation and data. These last set of corrections are applied to data, split into two steps, called "relative" and "absolute" residual corrections. The relative residual correction is estimated in the function of p_T and η in dijet events, balancing probe jets using as reference jets reconstructed in the well-calibrated $|\eta|$ <1.3 region. The residual correction is obtained with the missing energy projection fraction method (MPF). The idea is that dijet-like topology presents no intrinsic $\vec{\not{E}}_T$, therefore measured imbalance can be attributed to mismeasurement of the hadronic recoil. Figure 5.6 shows relative residual corrections averaged over p_T in different η bins and determined for different data-taking periods.

Absolute scale corrections are instead derived from $Z/\gamma^*(\to \ell\bar{\ell}) +$ jets and $\gamma +$ jets events. In these processes, the response is calculated in events that present limited additional hadronic activity. Jets are required to be reconstructed in the $|\eta|$ <1.3 region, back-to-back to the vector boson. The MPF methodology is used to measure the imbalance in jet p_T, which is then used to probe scale effects due to initial (ISR) or final state radiated (FSR) jets. Figure 5.7 shows absolute scale corrections estimated from fitting a functional form to $Z/\gamma^*(\to \ell\bar{\ell}) +$ jets, $\gamma +$ jets

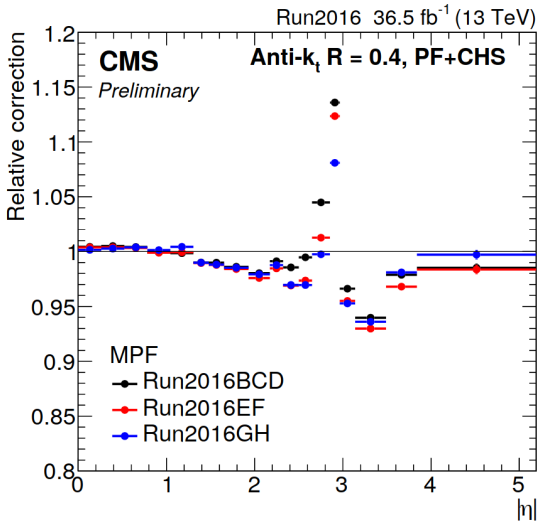


Figure 5.6.: Relative scale correction value at average p_T in each η bin is shown for beginning (Run2016BCD), middle (Run2016EF) and end (Run2016GH) of data taking [105].

and dijets measurements. The correction is reported in p_T bins and measured independently in different data-taking periods. Along with these corrections are associated systematic uncertainties to cover the modelling in MC from an alternative event generator, statistical uncertainties in the data due to trigger prescales, the dependence of the corrections over the data taking period, constituent particle energy uncertainties, and the residual difference between the dijet, γ + jets and $Z/\gamma^*(\to \ell\bar{\ell})$ + jets samples after all corrections have been applied.

Overall jet energy correction systematic uncertainties, which are propagated to the analysis results, are summarized for jet p_T and η in Figure 5.8.

5.3.5. Trigger efficiencies

5.3.5.1. Muon trigger efficiency

As described in Sec 5.1, the $\mu\mu$ + jets sample from which $Z \to \mu\mu$ events are reconstructed is selected with the logical OR of the `HLT_IsoMu24` and `HLT_IsoTkMu24` triggers. The trigger efficiencies are measured for both data and simulated samples (containing a trigger emulation) using a tag and probe method, [106], similarly to muon reconstruction and identification efficiencies. Comparing data and MC efficiencies, which are estimated in bins of p_T and η, a SF is derived and applied to correct event yields from simulations. The fraction of true muons in each bin is determined with a maximum likelihood fit to the invariant mass distribution. The signal component is modelled by a Breit-Wigner convoluted with a Gaussian distribution and background contribution with a falling exponential. It is straightforward to define the trigger

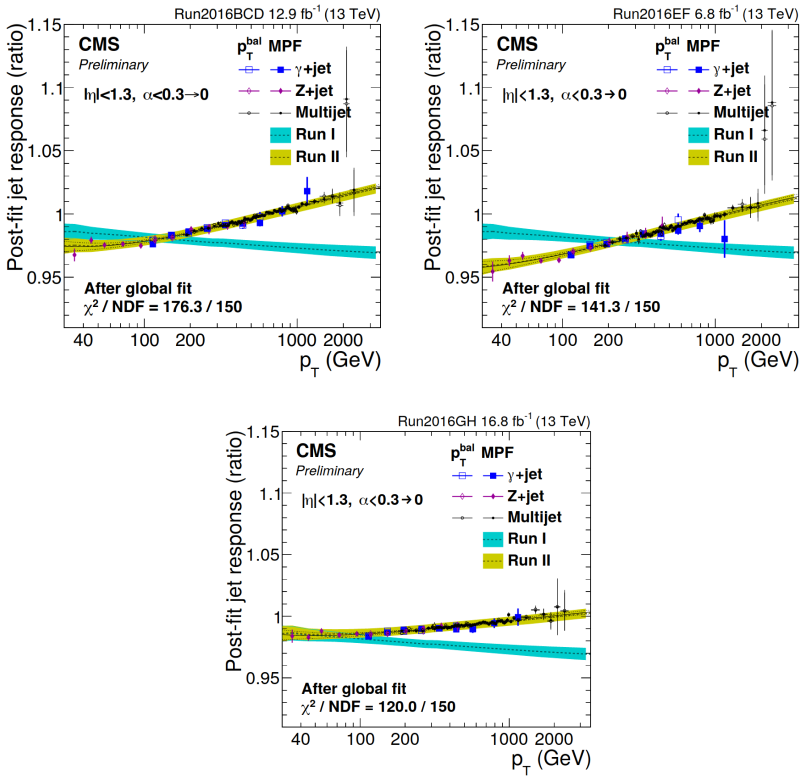


Figure 5.7.: Absolute scale corrections to detector response in bins of jet p_{T} for different moments of data-taking. The corrections are obtained from fitting dijet, γ + jets and $\mathrm{Z}/\gamma^{*}(\rightarrow \ell\bar{\ell})$ + jets events. Improvement in detector response, and therefore smaller corrections, can be observed for data collected at the end of data taking, driven by improved tracking performances [105].

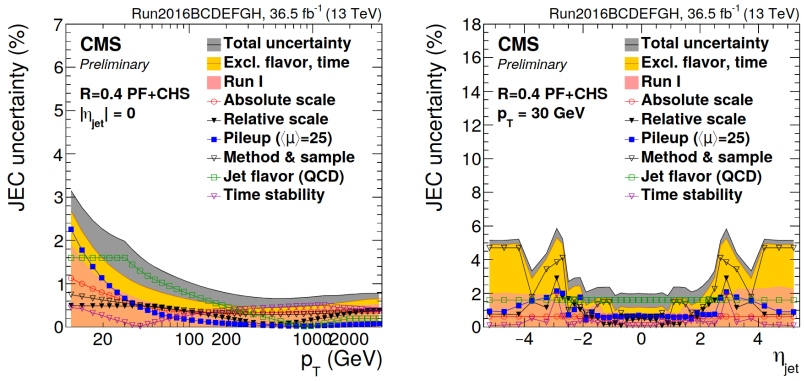
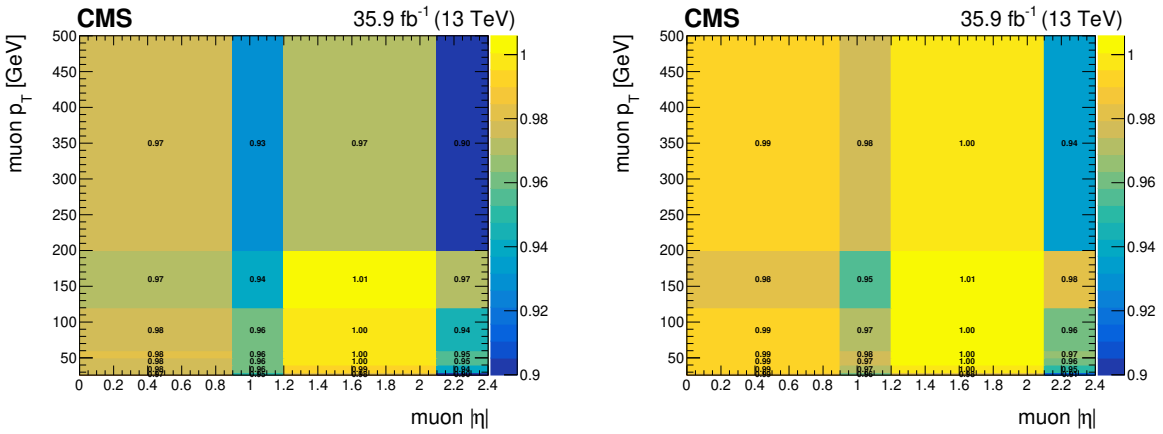


Figure 5.8.: Jet energy correction uncertainties for 2016 conditions. The plots show the uncertainty of jet energy scale corrections, reporting the total uncertainty (obtained summing all the uncertainties in quadrature) and the contribution of each source [105].

efficiency as the fraction of signal events where the probe passes the muon trigger selection over the overall number of true muons. The statistical uncertainty on the signal events is propagated through to the efficiencies. Additional systematic uncertainties are determined

by repeating the fitting procedure varying the functional forms for signal and background, the number of invariant mass bins, the invariant mass range considered and tag and probe selection requirements. As already discussed in Sec. 5.2.2, during the 2016 data-taking period, tracking performances were not constant throughout the year. The loss of tracking efficiency was mitigated in the later part of 2016. Despite better performances being achieved in the latest data reprocessing, muon trigger efficiencies are split into the initial part and final part of the data-taking period, as shown in Figure 5.9.



(a) Initial data-taking period

(b) Final data-taking period

Figure 5.9.: Muon trigger efficiencies parameterised by p_{T} and $|\eta|$, split by the data-taking period. To maintain a consistent background rejection, the trigger efficiency is lower in the endcaps where the muon background rate is larger and for low p_{T} muons where the track reconstruction is poorer. Inefficiencies for high p_{T} muons result from misidentified charged hadrons or issues in reconstructing the p_{T} for straight tracks. Most of the trigger efficiency arises from the L1T decision, which is heavily constrained by the timing budget. The differences between the data-taking periods are at most 2%, with significant overlap within the uncertainties.

The procedure described so far provides an efficiency, ε_i, associated with a single muon; however, as this analysis targets Z boson reconstruction, events are required to have two muons in the final state, with the single muon trigger being fired by any of the two leptons. Applying a binomial distribution, the trigger efficiency for at least one of the two muons passing the trigger selection ε, can be expressed as:

$$\varepsilon = 1 - (1 - \varepsilon_{\mu 1})(1 - \varepsilon_{\mu 2}) \,, \tag{5.6}$$

where $\varepsilon_{\mu 1}$ and $\varepsilon_{\mu 2}$ are the trigger efficiencies for the two muons in the selected events. This efficiency correction is applied to the MC event weight scaling the number of selected events in the analysis, with a luminosity-weighted average of the initial and final data-taking periods. The individual muon trigger efficiency uncertainties are combined to derive the event level one, which is then propagated as source of systematic error through the analysis.

5.3.6. Photon trigger efficiency

The photon trigger efficiency is calculated using two orthogonal methods: the "absolute" ratio technique and the so-called "bootstrap" method. The latter technique uses two identical HLT paths with different thresholds, in particular to measure the efficiency of the analysis reference trigger `HLT_Photon175`, a lower p_{T} threshold `HLT_Photon120` trigger is selected. The relative trigger efficiency is calculated as the ratio of the events that have a photon passing both triggers `HLT_Photon175` and `HLT_Photon120` to events that have a photon passing only `HLT_Photon120`. To apply this efficiency measurement to the analysis phase space, identical offline photon requirements are applied on the events firing these triggers. The efficiency is then evaluated as follows:

$$\varepsilon_{\texttt{HLT\_Photon175}} = \frac{N_{\texttt{HLT\_Photon175} \cap \texttt{HLT\_Photon120} \cap \texttt{Offline}}}{N_{\texttt{HLT\_Photon120} \cap \texttt{Offline}}} \,. \tag{5.7}$$

Only events passing photon tight identification criteria are selected, using 2016 Single Photon datasets.

Since both HLT_Photon175 and HLT_Photon120 may hide the same kind of inefficiency, the results may be biased. Hence the photon trigger efficiency is also measured using the orthogonal technique called the absolute trigger efficiency, where the Single Jet trigger paths are used.

Different single jet trigger paths are considered, and the efficiency is obtained by averaging the results obtained. The considered paths are HLT_PFJet40, 60, 80, 140, 200, 260, 320, 400, 450, 500, which requires at least one jet in the event with a p_{T} above the threshold. The efficiency is estimated using the same expression of Eq. 5.7, using the HLT_PFJetXX reference trigger and adding jet requirements to the offline selection. Events need to have at least one offline reconstructed particle flow jet, matched to the HLT jet object, with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ and passing the loose jet ID. The tight photon candidate in the event must be away from the reconstructed jet by $\Delta R = 0.7$ to avoid overlapping in the isolation cone.

The results of the two techniques *relative efficiency* and *absolute efficiency* are compared to MC results using GJets samples as shown in Figure 5.10. As can be seen from the top plot, both techniques give consistent results. The bottom plot shows the trigger efficiency used in the analysis, including statistical and systematic uncertainties. The systematic uncertainty is calculated as the average difference between the two trigger efficiency results, while the statistical error is taken from the HLT PFJetHT path. Since the photon trigger is not fully efficient at high photon p_{T} (Figure 5.10), this loss in efficiency is corrected for in the simulation.

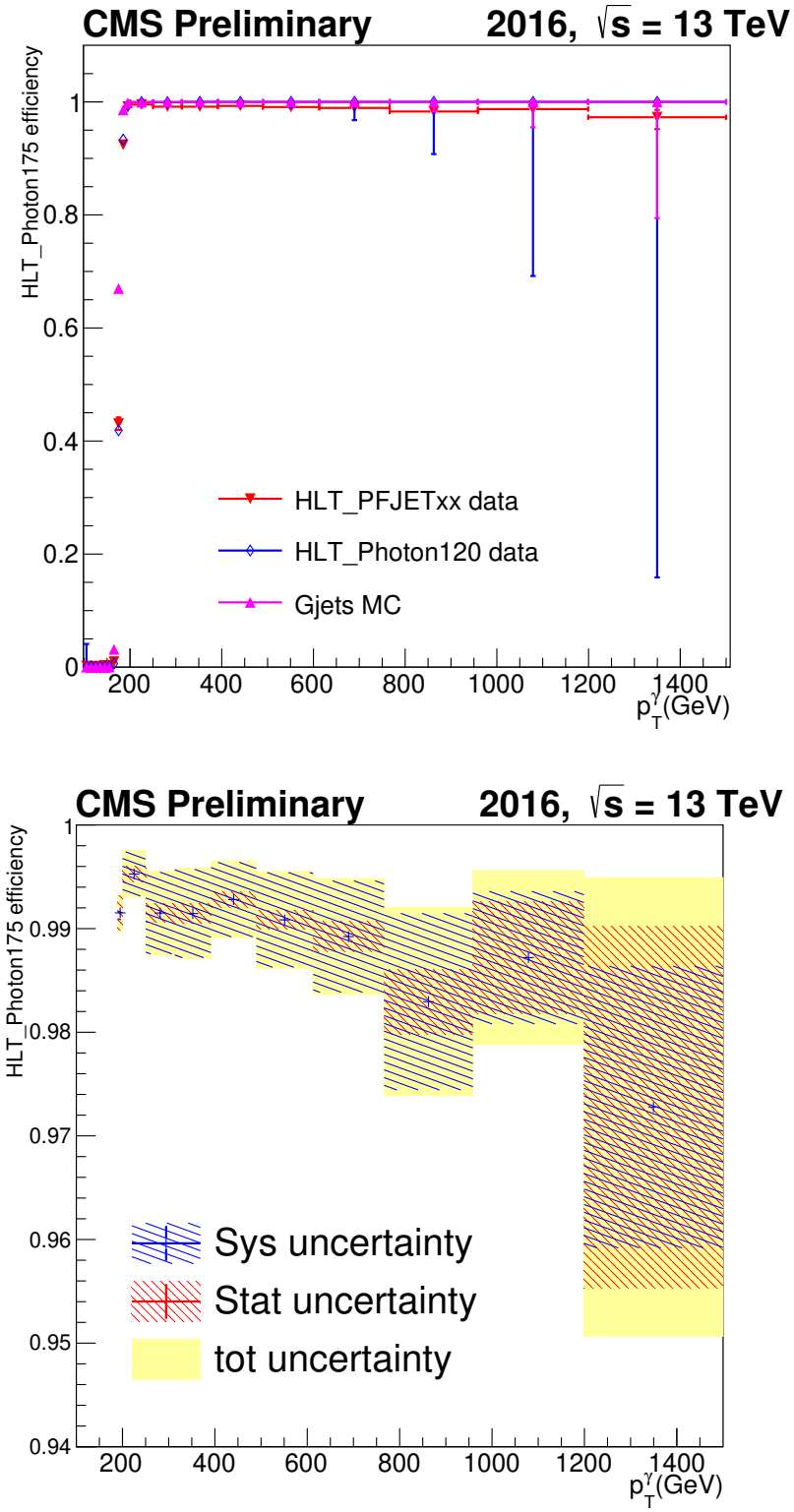


Figure 5.10.: (a) The trigger efficiency for HLT_Photon175 as measured in data using the ratio method (blue) and the orthogonal method (red) and compared to the simulated efficiency in the γ + jets MC sample (pink), as a function of the leading photon transverse momentum. (b) The distribution of the trigger efficiency and its associated systematic and statistical uncertainties [107].

A cross-check studying the dependence of the HLT efficiency on the photon η, $\sigma_{i\eta i\eta}$, and the number of primary vertices $nvtx$ has also been performed and results are shown in Figure 5.11. The trigger efficiency is relatively flat as a function of these variables and does not show any significant dependence.

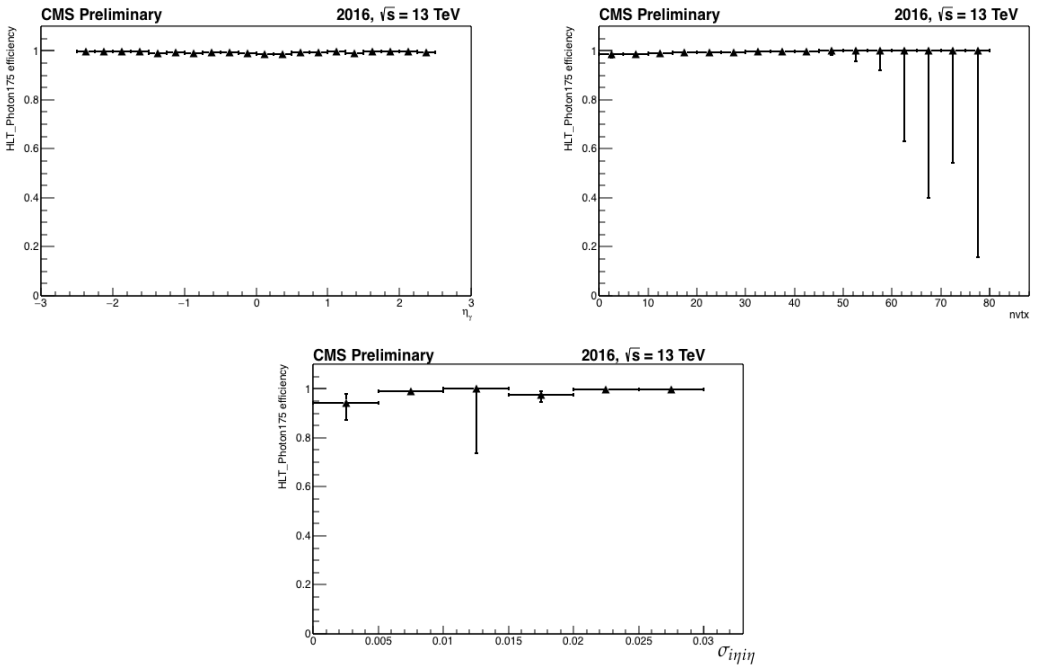


Figure 5.11.: Trigger efficiency as a function of photon η, nvtx and $\sigma_{i\eta i\eta}$ [107].

5.4. Event selection and fiducial regions

Events containing physics objects that satisfy identification requirements, after being corrected, are further selected on their kinematic properties. The set of kinematic requirements defines the fiducial regions in which the measurement presented in this thesis is performed. For the γ + jets analysis, events containing one tight isolated photon are selected. The photon is required to have $p_{\mathrm{T}} > 200$ GeV where the trigger reaches a plateau and becomes $> 99\%$ efficient and only photons inside the barrel region of the detector ($|y| < 1.4$) are considered. This is done because photon identification performs better in the barrel, and the analysis targets the high p_{T} region. The selected events are also required to contain at least one hard jet with the leading jet p_{T} of 100 GeV and $|\eta| < 2.4$. Subleading jets are required to have $p_T > 40\,\mathrm{GeV}$. The photon fiducial region in which the presented measurement is carried out is characterized therefore by one photon with $p_{\mathrm{T}} > 200$ GeV and $|y| < 1.4$, together with at least one jet with $p_{\mathrm{T}} > 100$ GeV and $|\eta| < 2.4$. Selected photons need to be isolated from jets in a $\Delta R = 0.5$ cone.

For the Z+jets samples, only the decay of the Z to muons is considered. Events are selected where the two tight and isolated leading muons have $p_{\mathrm{T}} > 30$ GeV and $\eta < 2.4$ with an invariant mass (m_{ll}) satisfying 71 GeV $< m_{ll} <$ 111 GeV. In addition, the Z p_{T} is required to be > 200 GeV and the rapidity $|y| < 1.4$, to be consistent with the photon rapidity range. As in the case of γ + jets selection, the presence of at least one jet with $p_{\mathrm{T}} > 100$ GeV within $|\eta| < 2.4$ is required. All other jets are required to have $p_{\mathrm{T}} > 40$ GeV and $|\eta| < 2.4$. The fiducial region for the Z+jets analysis is defined by requiring one Z boson with $p_{\mathrm{T}} > 200$ GeV and $|y| < 1.4$ and at least one jet with $p_{\mathrm{T}} > 100$ and and $|\eta| < 2.4$. The Z has to be reconstructed from a pair of muons with invariant mass within 20 GeV of the Z boson mass, 71 GeV $< m_{ll} <$ 111 GeV, with $p_{\mathrm{T}} > 30$ GeV and $\eta < 2.4$. Both muons are required to be isolated from jets in a $\Delta R = 0.5$ cone.

The event selection for the cross-section measurement of the collinear Z emission also selects the muon decay mode, requiring 2 tight, isolated muons with $p_{\mathrm{T}} > 30$ GeV and $\eta < 2.4$ and an invariant mass (m_{ll}) satisfying 71 GeV $< m_{ll} <$ 111 GeV. Since the Z boson is typically soft, there is no p_{T} requirement applied to it. Moreover the rapidity requirement on the Z boson is removed. However, to enhance these events, the leading jet must have a minimum p_{T} of 300 GeV, and additional jets are required to have $p_{\mathrm{T}} > 40$ GeV and $|\eta| < 2.4$. This selection targets dijet systems, which are expected to be balanced in jets p_{T} before the Z boson emission, therefore harder p_{T} thresholds on the subleading jets could be adopted. However given the limited statistics available in this particular selection, a lower and more inclusive requirement has been implemented. The fiducial region for this selection is defined by requiring one Z boson reconstructed from a pair of muons with an invariant mass within 20 GeV of the Z boson mass, $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$, and the presence of at least one jet with a p_{T} threshold on the leading jet of 300 (400, 500) GeV.

A summary of the fiducial region and event selection criteria for the different analysis categories is shown in Tables 5.7 and 5.8.

From simulations, the efficiencies of the event selections can be estimated with respect to events generated within the fiducial regions. The overall efficiencies for the Z+jets and γ + jets p_{T} selections are 86% and 80% respectively, whereas the efficiency for the collinear Z emission selection is 80%.

Table 5.7.: Summary of the fiducial region selections for each analysis category.

Fiducial region for Z/γ^*+jets analysis
2 muons with $p_T > 30$ GeV, $\|\eta\| < 2.4$
Vetoing events presenting jets within $\Delta R = 0.5$ of the muons
Dimuon invariant mass 71 GeV $< m_{ll} <$ 111 GeV
Z boson $p_T > 200$ GeV, rapidity $\|y\| < 1.4$
Leading jet $p_T > 100$ GeV, $\|\eta\| < 2.4$
All jets $p_T > 40$ GeV, $\|\eta\| < 2.4$
Fiducial region for γ+jets analysis
1 photon with $p_T > 200$ GeV, rapidity $\|y\| < 1.4$
Vetoing events presenting jets within $\Delta R = 0.5$ of the photon
Leading jet $p_T > 100$ GeV, $\|\eta\| < 2.4$
All jets $p_T > 40$ GeV, $\|\eta\| < 2.4$
Fiducial region for collinear Z emission analysis
2 muons with $p_T > 30$ GeV, $\|\eta\| < 2.4$
Vetoing events presenting jets within $\Delta R = 0.5$ of the muons
Dimuon invariant mass 71 GeV $< m_{ll} <$ 111 GeV
Leading jet $p_T > 300$ GeV (400 GeV, 500 GeV), $\|\eta\| < 2.4$
All jets $p_T > 40$ GeV, $\|\eta\| < 2.4$

Table 5.8.: Summary of the event selection criteria for each analysis category.

Event Selection for Z/γ^*+jets analysis
2 tight, isolated muons with $p_T > 30$ GeV, $\|\eta\| < 2.4$
Dimuon invariant mass 71 GeV $< m_{ll} <$ 111 GeV
Z boson $p_T > 200$ GeV, rapidity $\|y\| < 1.4$
Leading jet $p_T > 100$ GeV, $\|\eta\| < 2.4$
All jets $p_T > 40$ GeV, $\|\eta\| < 2.4$
Event Selection for γ+jets analysis
1 tight isolated photon γ $p_T > 200$ GeV, rapidity $\|y\| < 1.4$
Leading jet $p_T > 100$ GeV, $\|\eta\| < 2.4$
All jets $p_T > 40$ GeV, $\|\eta\| < 2.4$
Event Selection for collinear Z emission analysis
2 tight, isolated muons with $p_T > 30$ GeV, $\|\eta\| < 2.4$
Dimuon invariant mass 71 GeV $< m_{ll} <$ 111 GeV
Leading jet $p_T > 300$ GeV (400 GeV, 500 GeV), $\|\eta\| < 2.4$
All jets $p_T > 40$ GeV, $\|\eta\| < 2.4$

Chapter 6.

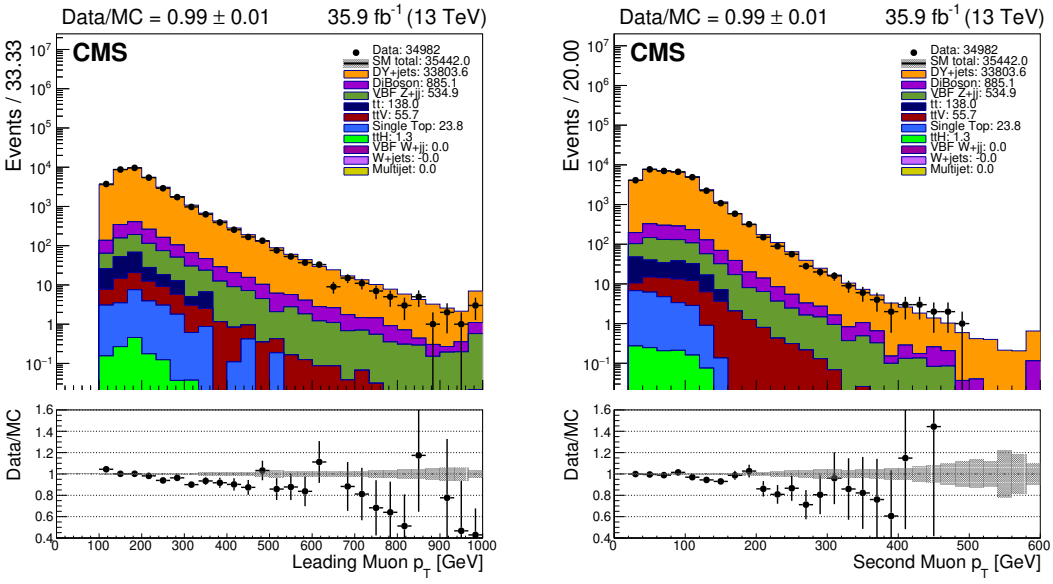
Background evaluation

6.1. Z+jets

6.1.1. Data-MC comparison

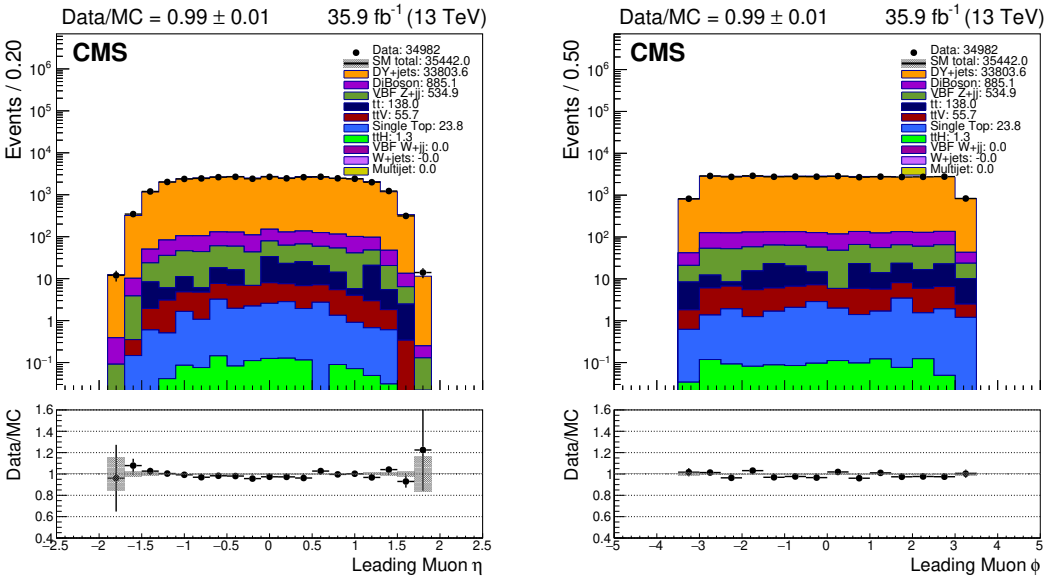
This section shows the data-MC comparison using out-of-the-box MC for some of the key observables in the Z+jets channel, after the selection described in Section 5.4 is applied. Fig. 6.1 shows the p_{T} distribution for the leading and subleading muons, Figures 6.2 , 6.3 show the η and ϕ distribution for the lead and sublead muons, respectively, and Fig. 6.4 shows the invariant mass distribution of the dimuon pair and the p_{T} distribution of the reconstructed Z boson. Figures 6.5, 6.6 and 6.7 show jet related variables, the scalar sum of the p_{T} of all selected jets (HT), the jet multiplicity, the p_{T} distribution of the 3 most energetic jets and the η distribution of the two leading jets.

In general, a good agreement is found between data and MC predictions, with the overall normalisation yielding a data/MC ratio of 0.99 ± 0.01 (statistical) for the analysis selection. The p_{T} distributions for both muons and the reconstructed Z bosons show good agreement in the first bins, while MC predictions overshoot data in the high momentum tail. Jet p_{T} variables (p_{T} of the first three jets and H_{T}) show a similar trend, where the low p_{T} bins in data are slightly above MC predictions, and a downward trend is observed for mid to high p_{T} values. Angular variables for both muons and jets show remarkable agreement across the full η and ϕ range. The N_{jet} distribution is well modelled by simulation up to two jets, while for higher jet multiplicities, a trend in shape can be observed, with simulation overpredicting events with three and four jets and undershooting data for the highest multiplicities. The invariant mass of the muon pair is overall well reproduced by simulation. A small shift can be observed in the Z peak, with MC peaked at a higher mass value, which leads to a 5 % disagreement. It should be noticed that the Drell-Yan sample used for the comparison is the MADGRAPH5_aMC@NLO v2.2.2 NLO in QCD, with up to two jets in the matrix element. Moreover, the error reported in the comparison plots is purely statistical, without including any systematic effect.



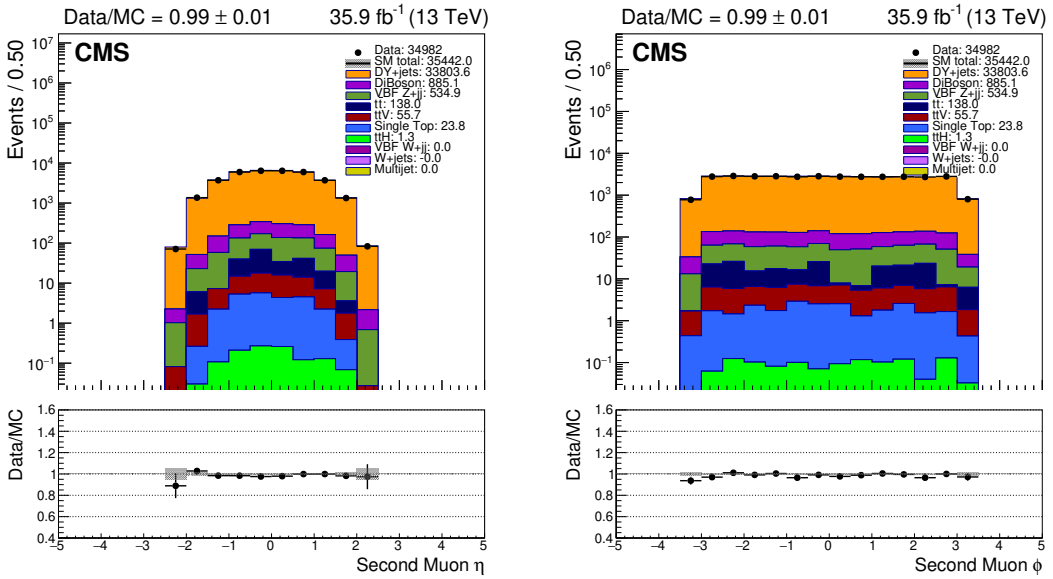
(a) **(b)**

Figure 6.1.: Distributions comparing data with simulation for (a) the leading muon p_T and (b) the subleading muon p_T. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.



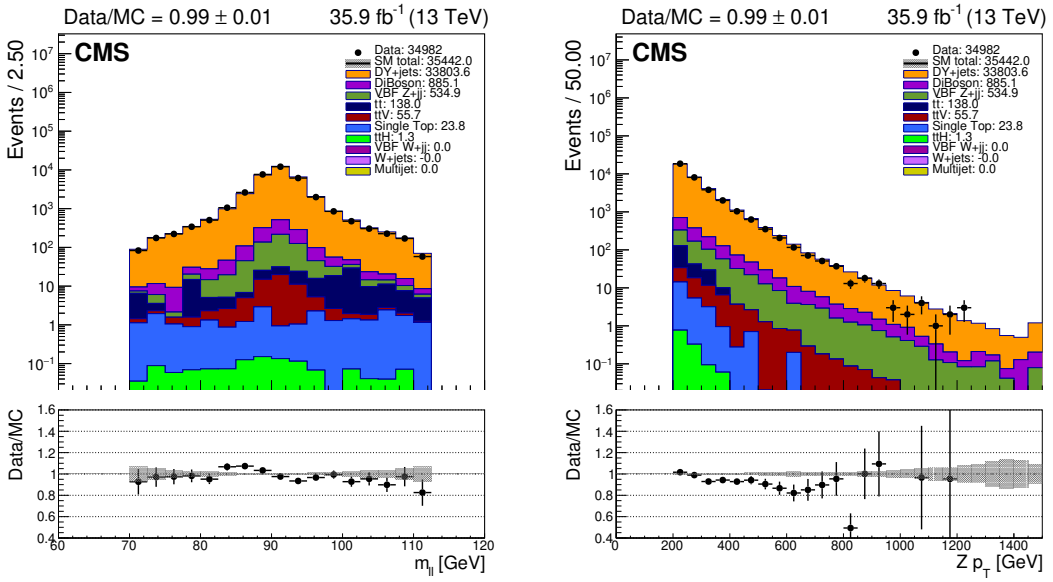
(a) **(b)**

Figure 6.2.: Distributions comparing data with simulation for (a) the leading muon η and (b) the leading muon ϕ. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.



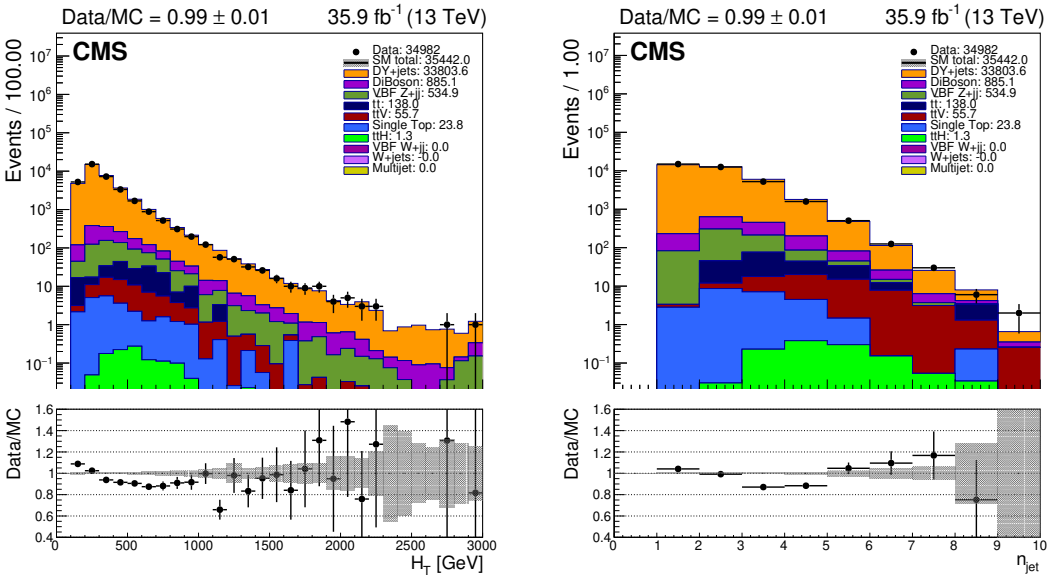
(a) **(b)**

Figure 6.3.: Distributions comparing data with simulation for (a) the subleading muon η and (b) the subleading muon ϕ. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.



(a) **(b)**

Figure 6.4.: Distributions comparing data with simulation for (a) the invariant mass of the two leading muons m_{ll} and (b) the transverse momentum of the Z. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.



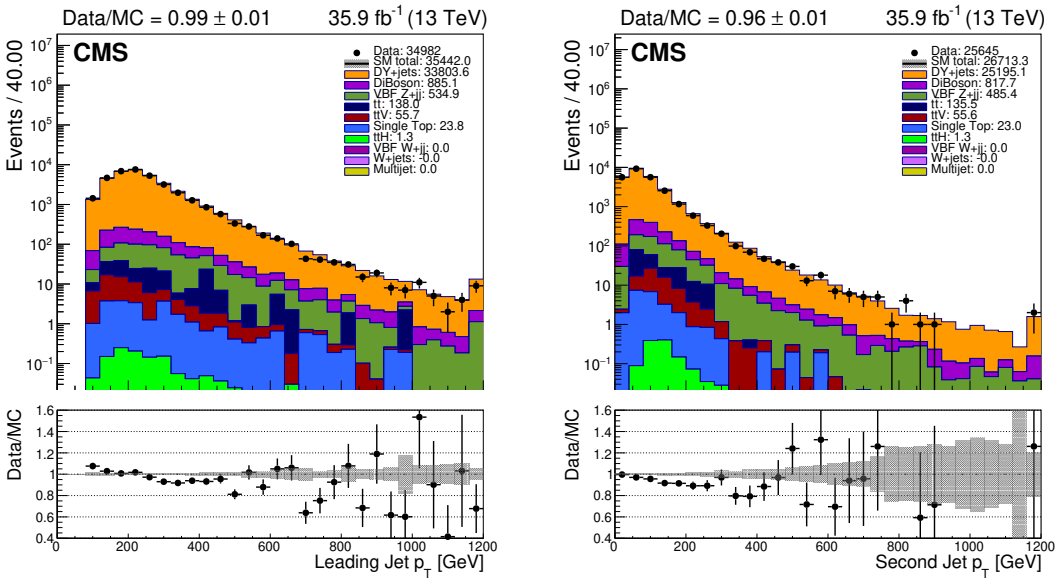
(a) **(b)**

Figure 6.5.: Distributions comparing data with simulation for (a) the scalar sum of all the jets in the event, H_T, and (b) the jet multiplicity. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.

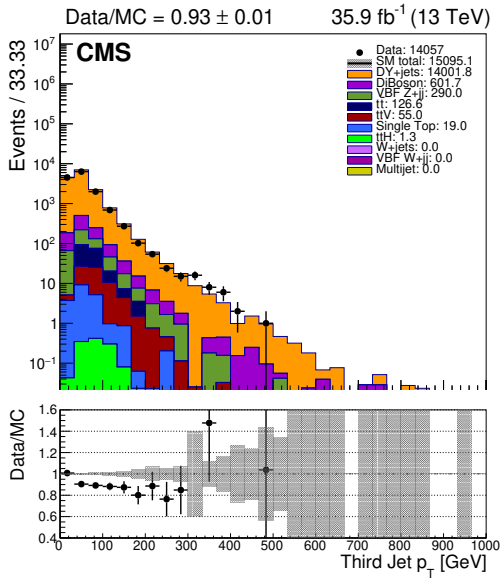
The level of background contamination is kept under control (<5%), proving the effectiveness of the adopted identification and isolation requirements. The spurious events passing the Z + jets selection are mainly from irreducible resonant background from Diboson (2.5%),VBF Z+2jets (1.5%) and $t\bar{t}Z$ (0.1%) processes, and combinatorial background from $t\bar{t}$ (0.4%) and single top (<0.1 %) processes.

6.1.2. Background subtraction

Given the small contribution expected from background events, coming from genuine Z + jets production, the background contribution is taken from MC predictions, and it is subtracted before extracting the Z + jets cross-section. The uncertainty on the backgrounds is taken as the squared sum between the statistical MC uncertainty and the systematic uncertainty, derived from $\pm 1\sigma$ variation on the object efficiencies and simulation corrections. A detailed discussion of the considered systematic effects is provided in Sec. 7.2.

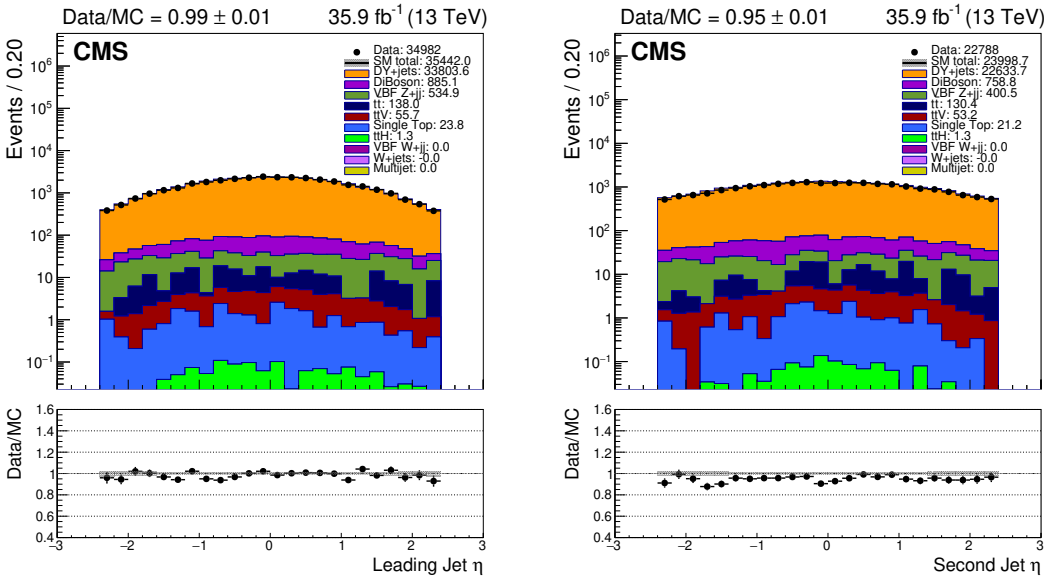


(a) (b)



(c)

Figure 6.6.: Distributions comparing data with simulation for (a) the p_{T} of the leading jet, (b) the p_{T} of the subleading jet and (c) the p_{T} of the third jet in the event. As the analysis selection ensures the presence of one jet, the distributions for higher jet multiplicity present smaller normalisations. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.



(a) **(b)**

Figure 6.7.: Distributions comparing data with simulation for (a) the η of the leading jet (b) the η of the subleading jet. The subleading distribution shows a smaller number of events as the analysis selection requires at least one jet. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.

6.2. γ + jets

6.2.1. Data-MC comparison

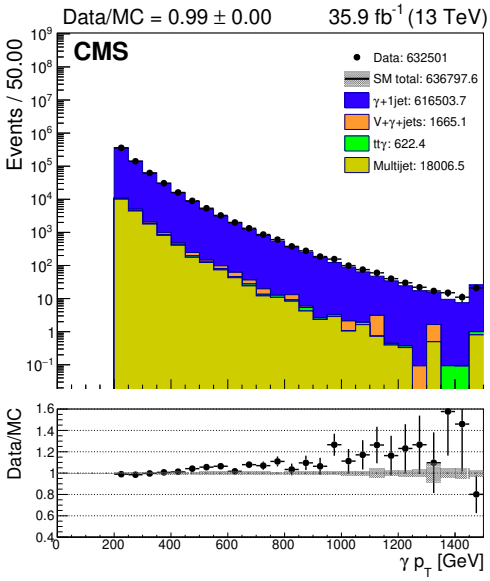
The data-MC comparison for γ + jets events is shown in Figures 6.8 to 6.12. The data are compared with the NLO γ + 1 jet simulation from MADGRAPH5_aMC@NLO v2.2.2. The photon p_T, η and ϕ distributions are shown in Figure 6.8. The p_T distribution for the leading jet and the scalar sum of all jets in the event passing the jet selection requirements are shown in Figure 6.9. The p_T distribution for the second and third leading jet is shown in Figure 6.10 and the η distribution for the first and second jet in Figure 6.12. The jet multiplicity and the vector sum of all jets in the event passing the jet selection requirements, denoted as H_T^{miss}, is shown in Figure 6.11.

The comparison between data and MC shows reasonable agreement for most of the variables investigated, with an overall ratio of 0.99 between data and MC cross-sections. The background contribution is around 3%, and its composition is dominated by multijet events, where hadrons are misidentified as photons. There are subdominant contributions from Vγ+jets and t$\bar{t}\gamma$ events. This minor background is estimated via MC and subtracted from the event yields before extracting the γ + jets cross-section. However, the contribution from QCD multijet events requires a data-driven technique to be measured as simulations are not reliable when reproducing misidentified events.

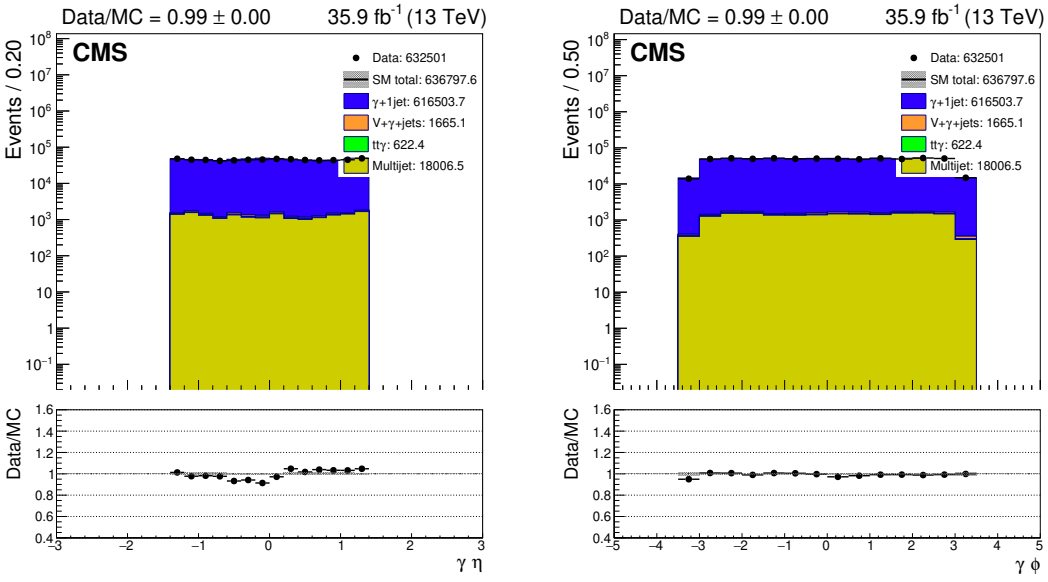
6.2.2. Purity Extraction

The QCD multijet processes provide events in which an electron or π^0 from a jet is misidentified as a photon candidate and passes the analysis offline selection. The contribution from such fake photon events is estimated from the "purity" of γ + jets events, which is defined as the fraction of true, isolated photons from the hard scattering versus the number of all photon candidates after the full selection criteria is applied. A *Template fit* method is used to extract the purity of the photon selection in each p_T bin of the analysis by fitting to the data with a sum of the signal and background template, where the signal denotes the distribution of true photons and the background is the distribution of QCD induced fake photon events.

The number of isolated photons coming from the hard scattering is extracted from a fit to the shower shape variable $\sigma_{i\eta i\eta}$, which measures the extent of the shower across the ECAL crystals in the η direction within a 5×5 cluster [107]. The signal template is obtained from simulated γ + jets events generated at NLO using MADGRAPH5_aMC@NLO v2.2.2, selecting all candidates passing the analysis selection criteria and matched to a particle level isolated photon coming from the hard scattering. The particle level photon is defined as a prompt

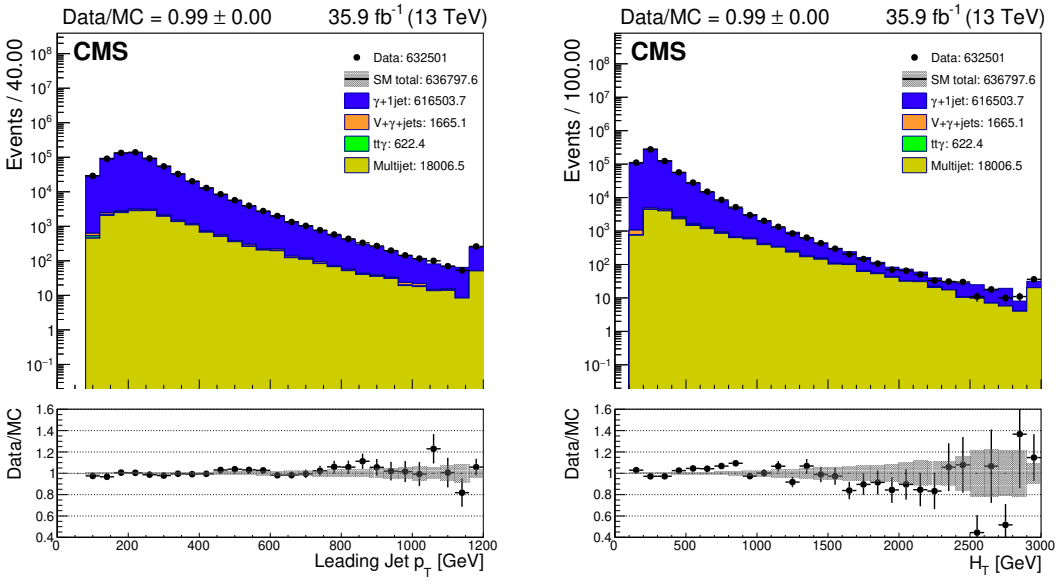


(a)



(b) (c)

Figure 6.8.: Distributions comparing data with simulation for the photon p_T (a) the photon η and (b) the photon ϕ (c). The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.



(a) **(b)**

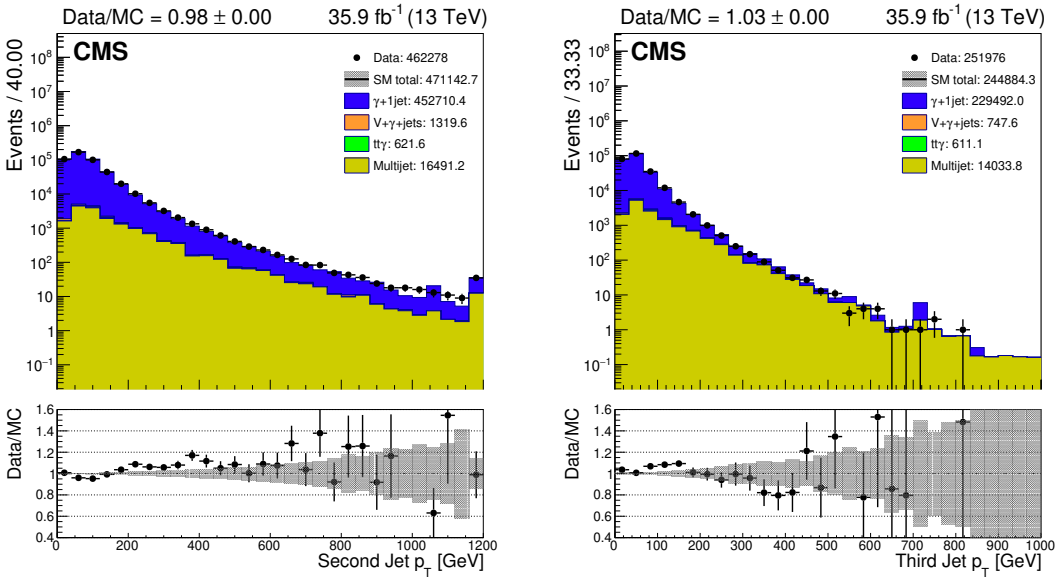
Figure 6.9.: Distributions comparing data with simulation for (a) the leading jet p_{T} and (b) the scalar sum of the jets in the event, HT. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.

photon around which the scalar sum of the p_{T} of all stable particles in a cone of radius $\Delta R = 0.4$ is less than 5 GeV.

Since discrepancies between data and simulation are observed in the $\sigma_{i\eta i\eta}$ variable, the MC distribution of $\sigma_{i\eta i\eta}$ has been reweighted to better match data using a histogram remapping method. Starting from unbiased MC and data samples of true electrons, obtained with a tag and probe selection, the percentile $\sigma_{i\eta i\eta}$ distributions are built. The mapping is then done assigning a new $\sigma_{i\eta i\eta}$ value to each MC event, finding the value in the data distribution with the same percentile. Fig.6.13 shows a comparison between data and MC $\sigma_{i\eta i\eta}$ distributions before and after the remapping. The tag and probe samples used for the remap are:

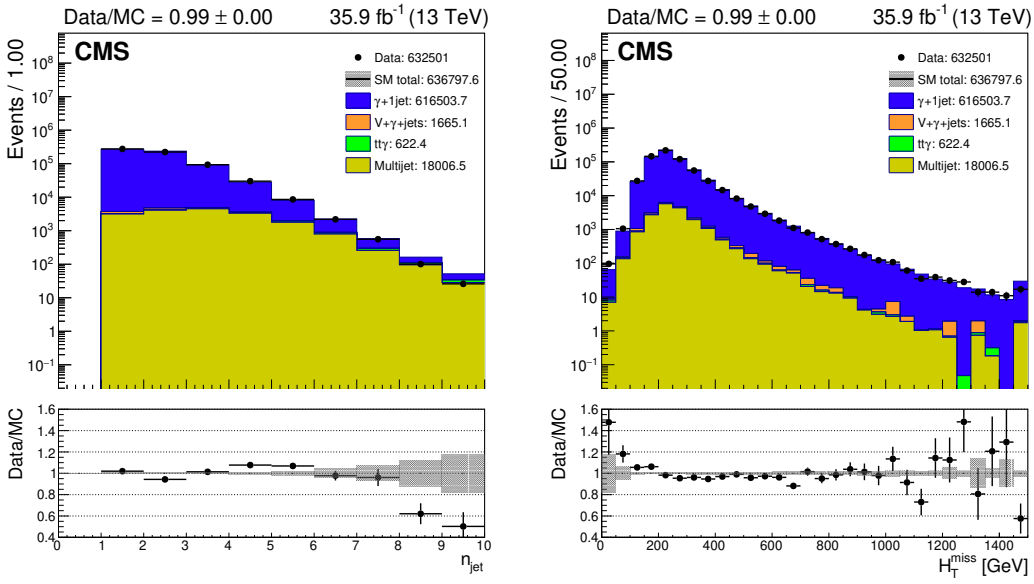
- Data Sample: 2016 Single Electron dataset
- MC Sample: DY+Jets MADGRAPH5_aMC@NLO v2.2.2 MLM + PYTHIA 8.2 @ LO

An uncertainty on the shape of the signal template is derived from this reweighting, considering maximum percentile variations within statistical uncertainties of the $\sigma_{i\eta i\eta}$ bins. Another uncertainty on the signal is assigned to the choice of the MC, using alternative γ + jets MC simulations; the signal $\sigma_{i\eta i\eta}$ distribution as predicted by SHERPA v2.2.8 and LO MADGRAPH5_aMC@NLO v2.2.2 are used, and a comparison is made between this and the nominal distribution from MADGRAPH5_aMC@NLO v2.2.2. The shape of the $\sigma_{i\eta i\eta}$ distribution



(a) **(b)**

Figure 6.10.: Distributions comparing data with simulation for (a) the subleading jet p_{T} and (b) the third jet p_{T}. As the analysis selection ensures the presence of one jet, the distributions for higher jet multiplicity present smaller normalisations with respect to the nominal ones. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.



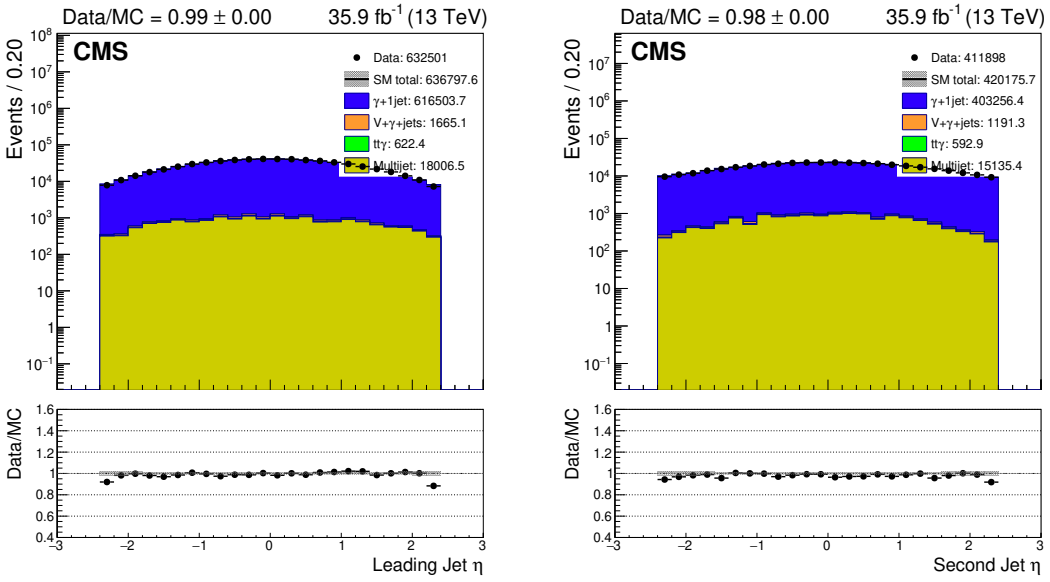
(a) **(b)**

Figure 6.11.: Distributions comparing data with simulation for (a) the jet multiplicity and (b) the $H_{\mathrm{T}}^{\mathrm{miss}}$. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.

from the two simulations is found to be similar. The purity estimation extracted using the signal template from SHERPA v2.2.8 is taken as a systematic uncertainty. More details on this can be found in Appendix A.

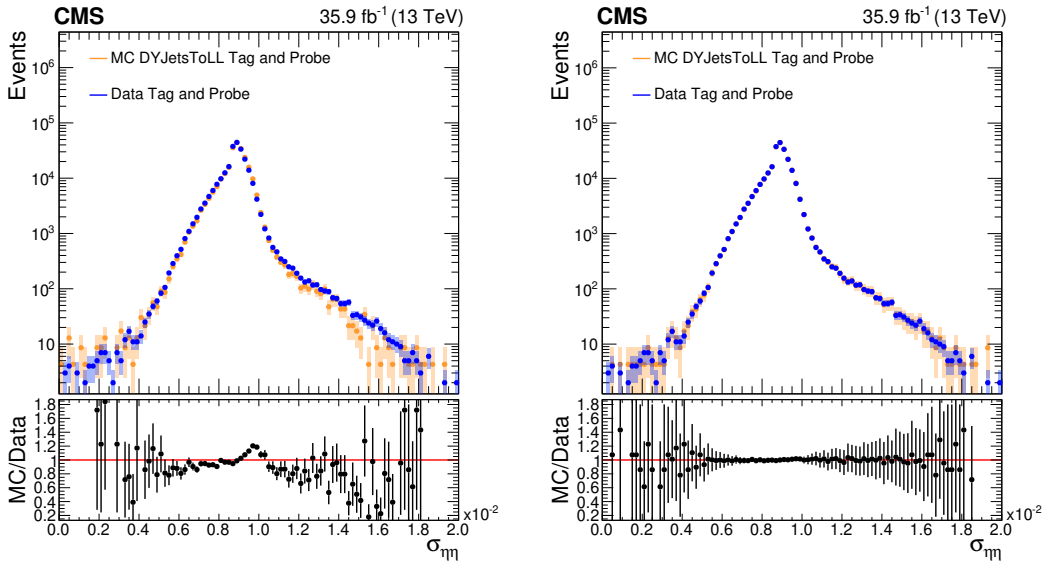
The background template is taken from data in a sideband of the charged hadron isolation variable ($\mathrm{I}_{\mathrm{ph}}^{\mathrm{chHad}}$). The fiducial $\mathrm{I}_{\mathrm{ph}}^{\mathrm{chHad}}$ interval for "tight" photons is $\mathrm{I}_{\mathrm{ph}}^{\mathrm{chHad}} < 0.202$ GeV. To produce the background template, the $\sigma_{i\eta i\eta}$ distribution from photon candidates with $\mathrm{I}_{\mathrm{ph}}^{\mathrm{chHad}} \in [10.0 - 15.0]$ GeV is taken. The choice of the sideband is optimized to reduce signal contamination and, at the same time, provide good statistics of fake photons. More details on the sideband selection can be found in Appendix A.

After the templates are constructed, a binned maximum-likelihood fit to the $\sigma_{i\eta i\eta}$ data distribution is performed. The statistical framework used is based on Roofit [108] and RooStat [109]. To take into account the limited statistics of the templates, the Beeston-Barlow method [110] has been implemented, smearing underpopulated bins with a Poisson distribution. The template fit takes also into account systematic effects by considering alternative template shapes. Practically, templates obtained with a $\pm 1\sigma$ variation on the nominal parameters are interpolated with sixth order splines between the $\pm 1\sigma$ and linearly outside that, to obtain a continuous parametrisation. The uncertainties are added to the likelihood and profiled to determine the



(a) **(b)**

Figure 6.12.: Distributions comparing data with simulation for (a) the η of the leading jet (b) the η of the subleading jet. The subleading distribution shows a smaller number of events as the analysis selection requires at least one jet. The different MC samples are stacked while the data are superimposed. In the ratio box, the gray band around $y = 1$ represents the statistical uncertainty on the total MC prediction and the vertical lines on the data points show the statistical uncertainty of the measurement.



(a) **(b)**

Figure 6.13.: $\sigma_{i\eta i\eta}$ distribution before (a) and after (b) correction. In the ratio box of both plots the Data/MC agreement can be observed. The remapping procedure improves the agreement

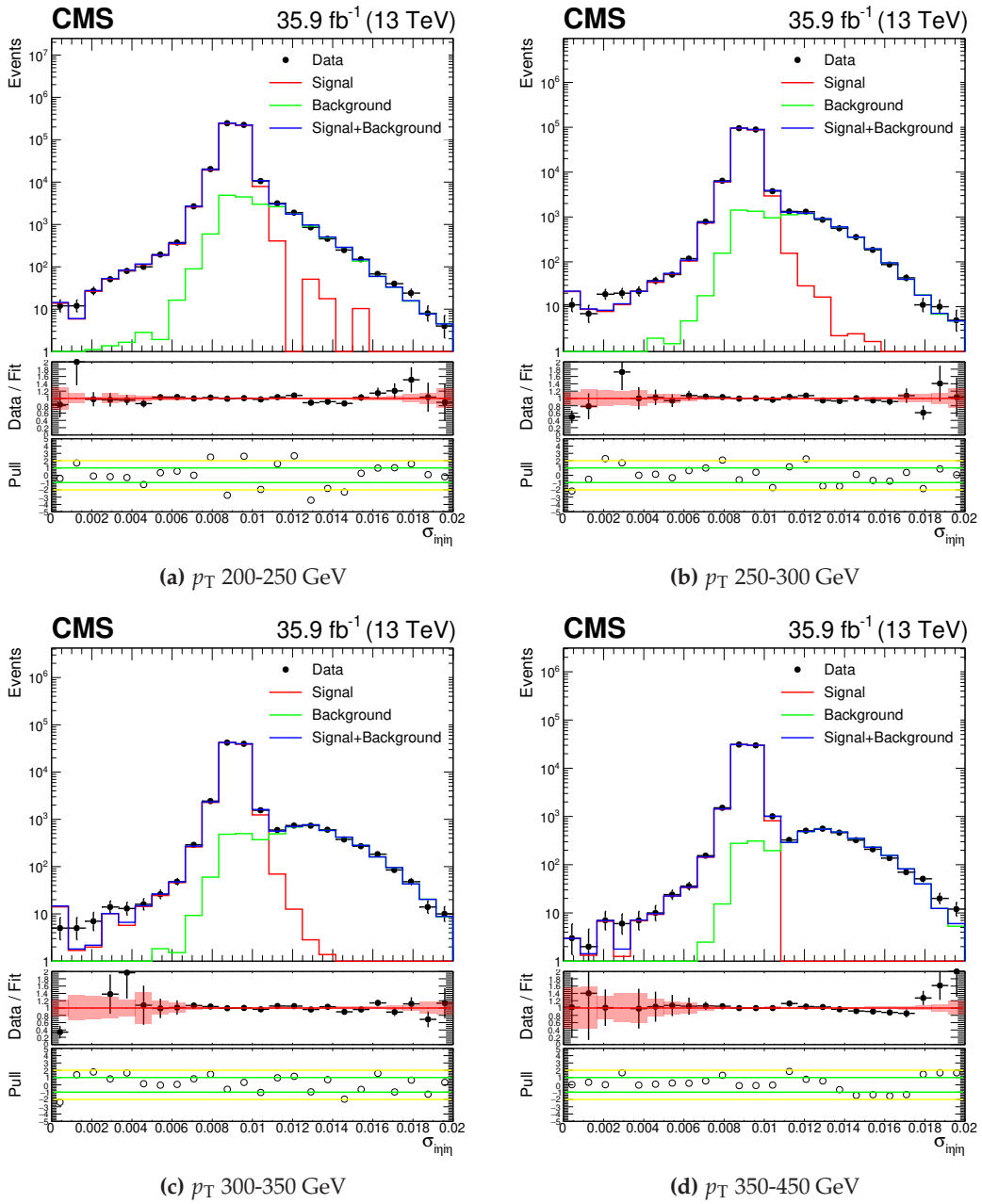
preferred value of the parameters and their impact on the fit uncertainty. A detailed description of the systematic effects considered in the fit is discussed in 7.2.

The purity fraction is then estimated by integrating the fitted template over the "tight" photon $\sigma_{i\eta i\eta}$ fiducial region, i.e. $\sigma_{i\eta i\eta} < 0.00994$. The fit results for the photon $\sigma_{i\eta i\eta}$ shower shape variable, in different p_{T} bins, can be seen in Fig. 6.14 and Fig.6.15, whereas the purity values with their uncertainty from the fit are shown in Table 6.1.

Table 6.1.: Photon purity, binned in p_{T}. The error includes both systematic and statistical uncertainties.

p_{T} bin [GeV]	Purity	error
200-250	0.980	0.003
250-300	0.985	0.007
300-350	0.987	0.008
350-450	0.990	0.013
450-600	0.993	0.010
600-800	0.995	0.018
800-Inf	0.997	0.043

Due to limited statistics, especially in the high p_{T} range, a coarser binning with respect to the analysis is chosen to perform the purity fit. The extrapolation to the analysis bins is carried out by fitting the purity values with an error function. The results of this fit can be seen in Fig. 6.16. The extrapolated purity using the functional form from this fit for the p_{T} bins corresponding to the analysis is shown in Table 6.2.



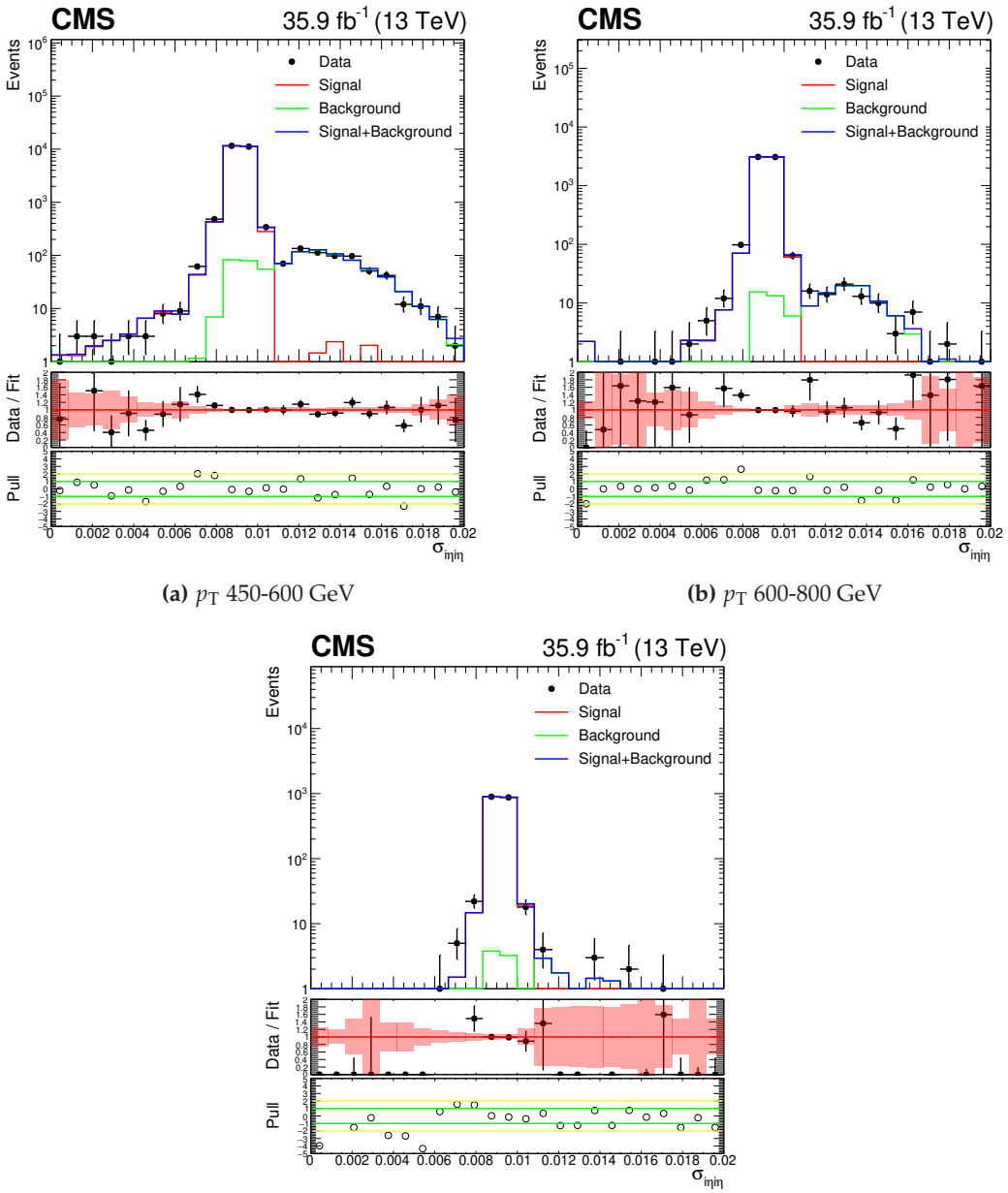
(a) p_T 200-250 GeV

(b) p_T 250-300 GeV

(c) p_T 300-350 GeV

(d) p_T 350-450 GeV

Figure 6.14.: The purity fit on the photon shower shape variable $\sigma_{i\eta i\eta}$ in the photon transverse momentum bins. The black dots are the data points, the red line is the signal template, and the green line represents the background component, and the blue line is the fit model obtained combining the templates. The middle panel shows the ratio between the data and the fit, and the bottom panel shows the pulls.



(a) p_{T} 450-600 GeV

(b) p_{T} 600-800 GeV

(c) p_{T} 800-Inf GeV

Figure 6.15.: The purity fit on the photon shower shape variable $\sigma_{i\eta i\eta}$ in the photon transverse momentum bins. The dots are the data points, the red line is the signal template, and the green line represents the background component. The blue line is the fit of the combined templates. Below the fit plot, the ratio and pulls distribution between the data and the fit are reported.

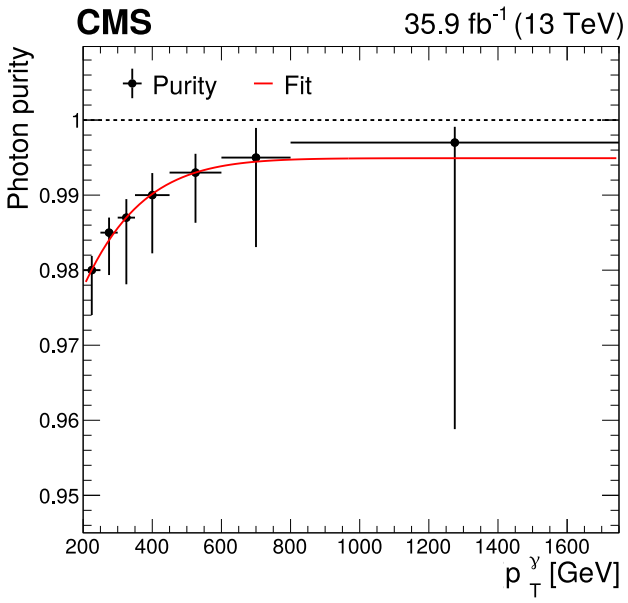


Figure 6.16.: The purity results obtained in bins of photon p_T are fitted with an error function. This fit parameterisation is used to determine the purity for the p_T binning used in the unfolding.

Table 6.2.: Extrapolated photon purity for the p_T bins corresponding to the analysis. The error includes both systematic and statistical uncertainties.

p_T bin [GeV]	Purity	error
200-230	0.98	0.03
230-250	0.98	0.02
250-290	0.98	0.02
290-320	0.99	0.03
320-370	0.99	0.03
370-410	0.99	0.03
410-460	0.99	0.03
460-520	0.99	0.03
520-590	0.99	0.04
590-670	0.99	0.05
670-750	1.00	0.06
750-850	1.00	0.06
850-960	1.00	0.07
960-1080	1.00	0.07
1080-1220	1.00	0.07
1220-1380	1.00	0.07
1380-1550	1.00	0.07
1550-1750	1.00	0.07

Chapter 7.

Differential cross section measurements

In this chapter, the extraction of the differential Z/γ + jets ratio is described. Starting from the event yield obtained after the event selection described in Chapter 5 and the background predictions as derived in Chapter 6, the Z + jets and γ + jets distributions are corrected to particle level for detector reconstruction effects, performing a procedure known as "unfolding". In Sec. 7.1 a detailed description of this procedure is provided; Sec. 7.2 discusses the treatment of systematic uncertainties, while unfolded differential distributions for the Z + jets, γ + jets and their ratio are compared to several theoretical predictions in Sec. 7.3, 7.4 and 7.5 respectively.

7.1. Unfolding

The detector-level distributions from data are unfolded to the particle level after removing background contributions. This step aims to undo the smearing and efficiency loss due to detector finite resolution and acceptance. In particular, when performing a differential measurement, events might be reconstructed in different bins with respect to their particle level counterparts or not selected at all even if the objects producing deposits in the detector satisfy the analysis selection.

The unfolding problem can be written as:

$$\tilde{y}_i = \sum_{j=1}^{m} A_{ij} x_j, \; 1 \leq i \leq n \; , \tag{7.1}$$

where the bins x_j are the particle level distribution, A_{ij} represents the matrix of probabilities for the migrations from the particle-level bin j to any of the n detector-level bins (migration matrix) and $\tilde{y}_i$ is the expected event count at detector level. The observed event counts y_i can differ from the expected $\tilde{y}_i$ due to statistical fluctuations. A schematic view is given in Figure 7.1. This becomes more complicated when contamination from background processes is involved.

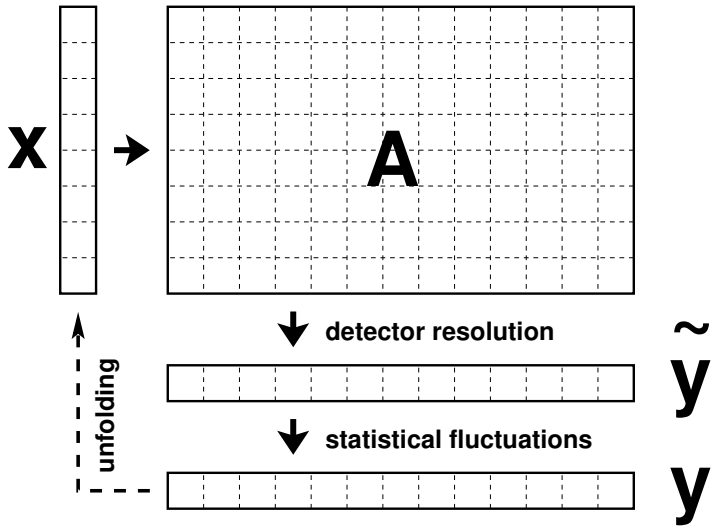


Figure 7.1.: Schematic depiction of the reconstruction process and the unfolding problem [111].

In that case the $\tilde{y}_i$ distribution receives an additional contribution from background, and Eq. 7.1 becomes:

$$\tilde{y}_i = \sum_{j=1}^{m} A_{ij} x_j + b_i,\ 1 \leq i \leq n\ , \tag{7.2}$$

where b_i is the background contributing to bin i.

In general, the presence of statistical fluctuations prevents the problem from being solved by simply inverting the migration matrix, as small variations can be over-amplified. Such fluctuations can be damped by imposing particular smoothness conditions on the x_j. This procedure is known as "regularisation".

The unfolding procedure in this analysis is carried out with the TUnfold software package [111] which implements a least squares method for migration matrix inversion with Tikhonov regularisation [112], providing the option for enforcing area preservation. The usage of least-squares minimization causes the measured data distribution to be finer than the particle level one, making the number of degrees of freedom, $n - m$, larger than zero.

The algorithm finds the stationary point of the problem "lagrangian":

$$\begin{aligned}
\mathcal{L}(x,\lambda) =& \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 && \text{with} \\
\mathcal{L}_1 =& (y - Ax)^{\mathsf{T}} V_{yy}^{-1} (y - Ax), \\
\mathcal{L}_2 =& \tau^2 (x - f_b x_0)^{\mathsf{T}} (L^{\mathsf{T}} L)(x - f_b x_0), \\
\mathcal{L}_3 =& \lambda (Y - e^{\mathsf{T}} x) && \text{and} \\
& Y = \sum_i y_i, \quad e_j = \sum_i A_{ij}
\end{aligned} \quad . \tag{7.3}$$

The first term $\mathcal{L}_1$ is the standard least square minimisation, where V_{yy} is the covariance matrix of the n-dimensional y distribution. The vector x is the unfolded distribution and has m rows. The elements A_{ij} of A describe for each row j of x the probability of migrating to bin i of y.

The term $\mathcal{L}_2$ describes the regularisation step, which damps fluctuations in x otherwise amplified when determining the stationary point of $\mathcal{L}_1$. The parameter τ^2 gives the regularisation strength, whereas the matrix L is the regularisation matrix, and the parameter $f_b x_0$ is the bias vector. The L matrix can have different forms, depending on the regularisation implemented. In the analysis, the regularisation is performed on the second derivative of x and, therefore, the L matrix has $n-2$ rows and the non-zero elements are $L_{i,i} = 1$, $L_{i,i+1} = -2$, $L_{i,i+2} = 1$. The bias vector is composed of a normalisation factor f_b and a vector x_0, and when $f_b \neq 0$, deviations of x from x_0 are suppressed.

The term $\mathcal{L}_3$ enforces the area constraint condition, including a dedicated variable λ in the lagrangian. The sum over all observations is given by Y, whereas the efficiency vector e has m rows and is calculated from A as indicated in equation 7.1. This constraint is implemented to limit biases on the unfolded distribution normalization due to the least square methodology being valid only for normal distributed measurements.

The minimum or stationary point of $\mathcal{L}(x, \lambda)$ is determined by setting the first derivatives to zero:

$$\begin{aligned} \frac{\partial \mathcal{L}(x,\lambda)}{\partial x_j} &= -2\left(A^{\mathsf{T}} {V_{yy}}^{-1}(y - Ax)\right)_j + 2\tau^2\left((L^{\mathsf{T}}L)(x - f_b x_0)\right)_j - \lambda e_j \\ \frac{\partial \mathcal{L}(x,\lambda)}{\partial \lambda} &= Y - e^{\mathsf{T}} x \end{aligned} \quad . \tag{7.4}$$

The migration matrix is determined from MC simulations, accessing both particle and detector level objects. With the TUnfold algorithm, it is foreseen to initialise the unfolding from a matrix M, containing $n+1$ rows and m columns, where the extra row (the underflow bins) is used to count the events which are generated within the acceptance but are not reconstructed. The unfolding algorithm uses such events to adjust the normalization of the x distribution, and therefore the unfolding step corrects the distributions for the overall event selection efficiency. Events which are instead reconstructed within the selected phase space, but are associated with generators objects not from the fiducial regions are treated as background.

When unfolding, the strength of the regularisation, τ^2, is an unknown parameter, and it is crucial to select an appropriate value for it. A large value for τ^2 results in a bias towards the $f_b x_0$ obtained from the MC prediction, whereas if it is too small, the unfolding result often has large fluctuations and correspondingly large negative correlations of adjacent bins. Different

methods for determining the regularisation strength have been developed. In particular, the TUnfold package implements a simple version of the so-called "L-curve" method.

The idea of the L-curve method is to look at the graph of two variables named L_x^{curve} and L_y^{curve} and locate the point where the curvature is maximal. These variables are defined as follows:

$$\begin{aligned} L_x^{\text{curve}} &= \log \mathcal{L}_1 \\ L_y^{\text{curve}} &= \log \frac{\mathcal{L}_2}{\tau^2} \end{aligned} \quad . \tag{7.5}$$

L_x is proportional to the difference between the unfolded vector x and the data, whereas L_y assesses the agreement of x with the regularisation condition. The optimal choice for τ implies a trade-off between the regularisation error and perturbation error; therefore, the maximum curvature point on the L^{curve} graph provides the value of τ for which these two effects have a similar impact in determining the unfolding result.

In TUnfold, the L-curve algorithm is implemented repeating the unfolding for an arbitrary number of points, scanning the curvature of the L-curve and identifying the kink point. In the analysis presented, the number of iterations has been fixed to $n = 30$. Varying the number of iterations shows no impact on the unfolded results.

The background contributions are subtracted from the data prior to unfolding, with the data distribution y and the covariance matrix V_{yy} being modified accordingly to the background prediction distribution and total uncertainty. The systematic errors affecting the background processes are summed in quadrature to the statistical uncertainty and they are considered uncorrelated with respect to the systematics affecting the signal yields. This is a conservative approach, but especially for MC based background subtraction, the overall background uncertainty is driven by the statistical error.

The migration matrix, A, usually receives uncertainties from various sources. First, statistical uncertainties are originating from counting the Monte Carlo events in the matrix M. On top of that, simulation corrections described in Sec. 5.3 introduce systematic uncertainties which result in a variation $M \to M + \delta M$, corresponding to a variation of experimental conditions. While the statistical uncertainties are bin-to-bin independent uncertainties and can be propagated through the unfolding formalism, systematic uncertainties might be correlated across bins. To estimate the impact on the result, the unfolding procedure is repeated, providing alternative migration matrices obtained by applying a $\pm 1\sigma$ shift to each implemented correction. A detailed description of the systematic effects considered in the analysis is reported in Sec. 7.2.

The detector level distributions are unfolded into the fiducial regions described in Sec. 5.4 and summarized in Tab. 5.7. To keep reconstruction level selections and unfolding fiducial

regions as close as possible, and thus minimizing the extrapolation in the cross-section measurement, an overlap removal at generator level is implemented, mimicking the reconstruction isolation requirements. In particular photon events are required to present one photon isolated from jets in a $\Delta R = 0.5$ cone, similarly both muons from the Z boson are required to be isolated from jets in a $\Delta R = 0.5$ cone. In case the generator isolation is not satisfied, the event is treated as background.

The NLO MADGRAPH Z + jets sample has been used to build the migration matrix. To properly fill the migration matrix underflow bin, a set of MC samples is used that include all generator level events satisfying the fiducial region selection and the full reconstruction without any further selection criteria applied. This ensures the efficiency of the event reconstruction and selection to be fully incorporated into the unfolding procedure and the resulting distribution to be comparable with theoretical predictions at the particle level.

Every generator level event is matched to a reconstructed level candidate. If a match is found, the event is used to build the migration matrix, using the event weights derived for implementing reconstruction corrections. If a viable match is not found, the event is put in the underflow bins of the migration matrix, and it is used in the unfolding process to correct for the selection efficiencies. Events presenting reconstructed objects passing the analysis offline selection and the fiducial region requirements, without a match to generator-level ones produced within the acceptance, are treated as background.

The binning for the unfolded distribution has to be coarser than the reconstruction level binning to help the unfolding algorithm to converge. The final binning of the different particle-level distributions has been obtained by merging groups of two reconstruction level adjacent bins to keep the bin migration below 1σ. For p_{T} distributions, the choice was also made to scale the bin width logarithmically to ensure good statistical precision in each bin. Figures 7.2 and 7.3 show the migration matrices for the p_{T} distribution of Z + jets and γ + jets channels. It can be observed how the chosen binning ensure the matrices to be almost diagonal, with small contributions to off-diagonal bins.

7.2. Systematic uncertainties

This section is dedicated to describing the different sources of systematic errors considered in the analysis. Two types of uncertainties are discussed, affecting the data measurements and theoretical predictions.

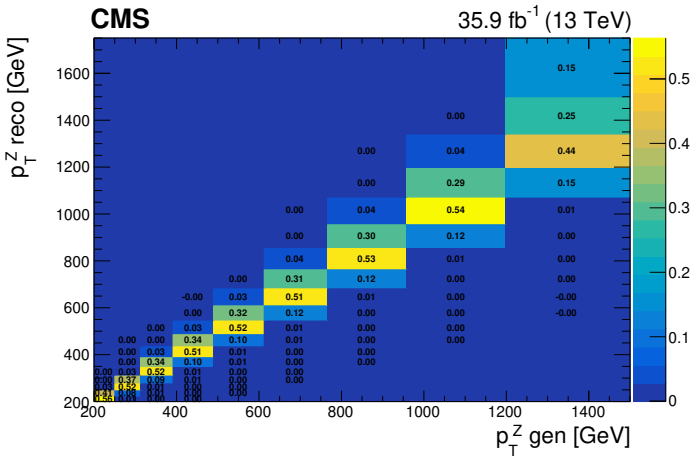


Figure 7.2.: The correlation matrix showing the bin migration in the p_T distribution of the Z boson. The matrix is almost diagonal, with a small fraction of events populating the off-diagonal bins.

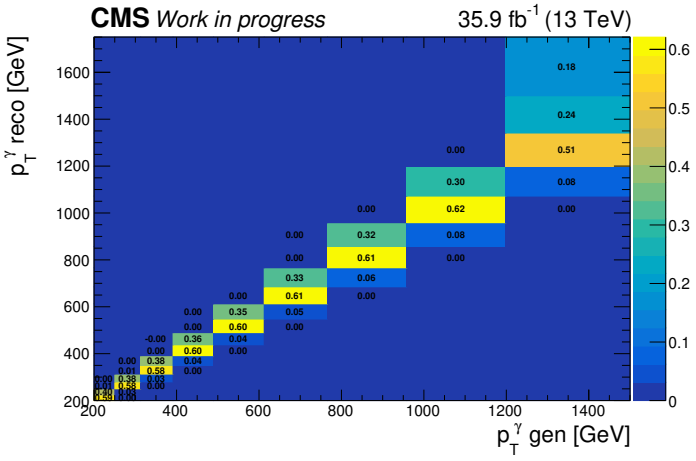


Figure 7.3.: The correlation matrix showing the bin migration in the photon p_T distribution. The matrix is almost diagonal, with a small fraction of events migrating between different bins.

7.2.1. Measurement uncertainties

- **Muon efficiencies**
  The corrections for muon efficiencies, determined with the tag-and-probe method as described in Sec. 5.2.2. The efficiencies for tracking, isolation and identification depend on η, p_T, as well as the data-taking period. Uncertainties on the corrections arise from event statistics in the considered bin and from the overall tag-and-probe and fitting strategy, the latter being considered as a systematic uncertainty. While statistical errors are not correlated across the bins, systematic uncertainties do correlate across different bins. Given that Z p_T is reconstructed from muons that can be reconstructed in different p_T and η bins, particular care is needed when propagating uncertainty effects to the Z p_T

distribution. In general, the scale factors are applied per event and propagated through the unfolding procedure by using the alternative MC distributions as input, obtained by a $\pm 1\sigma$ variation on the efficiencies. To properly take into consideration the correlation between the muon and the Z boson p_{T}, the SF have been varied in each of the p_{T} bins independently, deriving a set of systematic variations. The total systematic uncertainty on the unfolded distribution, coming from the muon SF, is derived by summing in quadrature these variations. The overall uncertainty is found by summing in quadrature the statistical and systematic variations. The muon selection efficiency uncertainty, which includes the identification, isolation, and tracking efficiency, contributes at the 0.8-1.0% level on the Z + jets cross-section.

- **Muon trigger**
  The systematic uncertainty on the muon trigger efficiency is derived in a similar manner as the isolation and identification uncertainty. Also in this case, a set of corrections from varying each p_{T} bin independently is obtained, with the final systematic uncertainty obtained summing in quadrature the different variations. The impact of the muon trigger efficiency uncertainty is given varying the single muon trigger SF in the binomial formula, reported in Eq. 5.6. Given that the overall event weight is very close to 1, the impact on the final distribution of the single muon trigger efficiency uncertainty is found to be negligible.

- **Photon efficiencies**
  The systematic uncertainty on the photon ID efficiency is determined using the tag-and-probe method and is provided as a function of the photon η and p_{T}. The efficiencies are applied per event. To estimate the impact of the scale factor systematics on the differential cross-section measurement, the alternative distributions obtained with a $\pm$ 1 σ variation on the efficiency central value are implemented in the unfolding procedure and fully propagated to the final measurement. The photon selection efficiency uncertainty, which includes the identification and isolation efficiency, contributes at the 2.5-2.6% level on the γ + jets cross-section across the full p_{T} range.

- **Photon trigger**
  The strategy to measure the single photon trigger efficiency is described in Sec. 5.3.6, with the statistical component of the uncertainty being derived from the JetHT absolute efficiency method, while a further systematic uncertainty is defined by considering the efficiency variations obtained in each p_{T} bin, using the relative efficiency methodology. The overall uncertainty on the single photon trigger efficiency is obtained by summing the statistical and systematic uncertainties in quadrature. The systematic uncertainty on the γ + jets p_{T} distribution is derived repeating the unfolding with alternative migration matrices obtained taking a $\pm 1\sigma$ variation on the single photon trigger efficiency. The

photon trigger efficiency uncertainty contributes at the 0.2-2.2% level on the γ + jets cross-section growing with p_{T}.

- **Muon energy scale**
  The systematic uncertainty on the muon energy scale and resolution is obtained by combining muon scale calibration and resolution error on the muon p_{T}, derived from the best reconstructed track associated with the muon. The analysis is repeated substituting the nominal muon p_{T} with a $\pm 1\sigma$ variation. As a result, a different number of events is selected, and the difference in the unfolding spectrum is taken as an uncertainty. The systematic uncertainty in the muon momentum scale is the dominant systematic uncertainty in the Z + jets cross-section. The uncertainty in the scale of up to 1 (6)% at $p_{\mathrm{T}} < 100\,\mathrm{GeV}$ (1 TeV) for each muon, results in an uncertainty in the cross-section ranging from 1.7% at low p_{T} up to 21.9% in the highest p_{T} bin.

- **Photon scale**
  The systematic uncertainty on the photon energy scale and resolution is taken considering resolution and scale errors and combining them in quadrature on a per-event basis. As for the muon scale uncertainty, the analysis is repeated substituting the nominal photon p_{T} with the $\pm 1\sigma$ variation. The difference in the unfolding spectrum, due to the different number of events fulfilling analysis selections, is taken as the uncertainty. The uncertainty in the photon energy scale and resolution of 1-2% results in an uncertainty in the cross-section of ($\leq$1%) at low p_{T} and up to 8.6% at high p_{T}, becoming the dominant systematic uncertainty on the γ + jets p_{T} distribution.

- **Pile up reweighting and luminosity**
  The "Pile-up reweighting" is applied to correct the MC distributions to match pileup conditions as measured in data. Corrections are derived by comparing simulated samples with minimum bias datasets (see Sec. 5.3.1). To derive an uncertainty on this procedure, the recommended 5% uncertainty is applied to the minimum bias cross-section. This affects the shape of the number of interaction vertices per bunch crossing variable and, therefore, different weights for simulated events. Alternative migration matrices are provided in the unfolding step containing such reweighting variations, however a negligible (less than 1%) effect on the Z + jets and γ + jets cross-sections is found.

  A 2.5% uncertainty on the total integrated luminosity is applied, as per standard CMS prescriptions [113]. The uncertainty on the luminosity differs from the ones described so far, as it directly affects the number of events in data. This uncertainty is therefore applied directly to the unfolded data before extracting the cross-section.

- **Jet energy corrections**
  The effect on the measurement from the JES and JER uncertainty is evaluated by varying the jet four-momenta using the uncertainties in the correction factors that depend on

the jet p_T and η for the JES and jet η for the JER. Details on jet correction uncertainties determination have been provided in Sec. 5.3.4. As for other objects, scale and resolution variations are substituted for nominal jet p_T event content, and the analysis is repeated with the alternative samples. The JES and JER uncertainties are subdominant (below or at the % level) for all three event categories due to the high threshold on the p_T of the leading jet.

- **Background subtraction**
  Given the high purity of the implemented muon selection, the $Z+$ jets background contribution is taken directly from simulations. To derive an uncertainty on this subtraction, all the uncertainties on the $Z+$ jets selection described in this section are applied to the considered background sample. The total uncertainty on the background prediction is obtained by scaling up and down coherently all the systematic variations and summing them in quadrature with the statistical uncertainty of the background samples. The background subtraction is performed in the unfolding procedure, and the error from the subtraction is propagated to the unfolded distribution automatically.

- **Purity estimation**
  The method described in Section 6.2.2 implements a template fit, using shapes derived from both data and simulation. A discussion of the systematic uncertainty on the signal and background templates is provided in Section 6.2.2 with more details in Appendix A. In summary, alternative signal templates obtained from different MC predictions and alternative background templates obtained varying the sideband region in the data are provided to the fit. Changing the number of bin shows a negligible impact on the purity results, provided a good statistical population of the bins and a meaningful low $\sigma_{i\eta i\eta}$ and high $\sigma_{i\eta i\eta}$ separation. Systematic uncertainties from each of the effects described so far for γ + jets selection are included in the purity fit via alternative shapes obtained from up and down variation from each of these effects. It should be noted that the change in Monte Carlo shape and normalization will also affect the background template as the predicted fraction of real γ + jets events in the sideband is varying. Another systematic source is associated with the $\sigma_{i\eta i\eta}$ remapping due to the E/Gamma POG Tag and Probe sample's finite size. MC driven background subtraction is also performed using Vγ+jets and t$\bar{\text{t}}\gamma$ predictions, but as such backgrounds are subdominant, the effect of this is negligible. The overall impact of photon background uncertainty is up to 1.1% at low photon p_T and down to 0.2% at high p_T.

- **Unfolding**
  Additional uncertainties are assigned to the unfolding procedure from the following sources: uncertainty due to the limited statistics of the response matrix, uncertainty due to the Monte Carlo spectrum inside the unfolded bins and bias introduced by the regularisation process. Since the response matrices of the distributions of interest are

fairly diagonal and present a rather small condition number, the optimal regularisation strength, determined with the L-curve scan method is found to be negligible. For this reason, no regularisation has been applied in the unfolding, and therefore it is not necessary to include a systematic uncertainty covering this. The error associated with the statistics of the response matrix is defined considering two alternative matrices obtained with up and down variations on the bin contents. To take into account the uncertainty coming from the Monte Carlo model, a comparison of the unfolding distributions has been implemented. The alternative migration matrices are obtained with the nominal NLO MADGRAPH5_aMC@NLO v2.2.2 ($\leq$ 2 jets) and with the LO MADGRAPH5_aMC@NLO v2.2.2 ($\leq$ 4 jets) sample for the $Z+$ jets selection. Similarly, for the γ + jets selection, a comparison is performed on the unfolded distribution obtained with the nominal NLO MADGRAPH5_aMC@NLO v2.2.2 γ + 1 jet and with the LO SHERPA v2.2.2 ($\leq$ 2 jets) sample. The overall unfolding uncertainty is obtained by summing in quadrature the contribution from the uncertainty associated with the simulation model and the migration matrix's statistical uncertainty. The uncertainty in the unfolding is the dominant uncertainty at high boson p_{T} for the $Z+$ jets cross-section, contributing up to 18.6% in the highest Z boson p_{T} bin and the subdominant uncertainty in the γ + jets cross-section, contributing up to 6.4% at high photon p_{T}. The unfolding uncertainty is driven by the statistical uncertainty of the migration matrix, growing with the p_{T} of the vector bosons. The uncertainty in the unfolding for the $\Delta R^{\mathrm{min}}_{Z,j}$ distribution is among the dominant uncertainties.

To provide a general overview of the impacts of the different systematic uncertainties on the measured quantities, a summary of the contributions from each source to the differential cross-section measurements of $Z+$ jets, γ + jets, Z/γ ratio, and collinear Z boson emission (with leading jet p_{T} threshold set to 300 GeV) is shown in Table 7.1. More detailed systematic contributions, split per analysis bin, are reported in the Tab. 7.2 Tab. 7.4 and Tab. 7.6 for $Z+$ jets, γ + jets and Z/γ p_{T} distributions respectively. Similarly, Tab. 7.7 and Tab. 7.8 reports per bin contribution on the $\Delta R^{\mathrm{min}}_{Z,j}$ selection for lead jet p_{T} thresholds of 300 GeV and 500 GeV.

Common sources of systematic uncertainties such as those from JES, JER, and integrated luminosity are treated as correlated between $Z+$ jets and γ + jets and therefore mostly cancel in the $Z+$ jets ratio, while sources of uncertainty such as the lepton efficiency, trigger and photon purity are treated as uncorrelated.

7.2.2. Theory uncertainties

Besides uncertainty coming from the statistical significance of the different theoretical predictions used as comparison to the presented measurement, for NLO QCD MC predictions,

Table 7.1.: The contributions to the uncertainty in the differential cross-section measurements for the Z + jets, γ + jets, the Z/γ ratio, and collinear Z boson regions. The uncertainties are expressed in per cent, and a range represents the minimum and maximum effect observed.

Systematic source	Z + jets [%]	γ + jets [%]	(Z + jets)/(γ + jets) [%]	Collinear region [%]
Trigger	0.0	0.2 – 2.2	0.2 – 2.2	0.0 – 0.2
Muon reconstruction and selection	0.8 – 1.0	—	0.8 – 1.0	0.9 – 1.1
Photon reconstruction and selection	—	2.5 – 2.6	2.5 – 2.6	—
Photon energy scale	—	0.5 – 8.6	0.5 – 8.6	—
Muon momentum scale	1.7 – 21.9	—	1.7 – 21.9	0.1 – 12.2
Photon purity	—	0.2 – 1.1	0.2 – 1.1	—
Background yields	0.7 – 1.5	—	0.5 – 1.6	0.9 – 10.9
Pileup	0.0 – 0.7	0.0 – 0.3	0.0 – 0.4	0.2 – 0.9
Integrated luminosity	2.5	2.5	0.0	2.5
Unfolding	0.3 – 18.6	1.1 – 6.4	1.1 – 19.7	1.2 – 10.8
JES/JER	0.0 – 0.2	0.0 – 0.2	$\leq$ 0.04	0.3 – 1.5
Total	3.3 – 28.9	4.0 – 11.5	4.4 – 30.9	3.5 – 16.8

theoretical uncertainties are also considered. Such uncertainties are determined by varying the QCD factorisation and renormalisation scale, α_s and considering alternative pdf variations.

- **Scale variations**

  The scale uncertainties in the NLO predictions from MADGRAPH5_aMC@NLO v2.2.2 and JETPHOX are estimated by varying the values of the renormalization, factorization and, for JETPHOX only, fragmentation scales up or down by a factor of 2. The variation is applied on the scales either individually or in a correlated fashion. However, the cases where one scale is multiplied by 2 and the other by 0.5 are considered as non-physical. The final scale uncertainty is estimated by taking the largest deviation from the nominal value among all the possible variations.

- **PDF variations**

  The PDF uncertainty on the predictions from JETPHOX is estimated using the latest PDF4LHC recommendations where the 100 replicas from the NNPDF30nlo analysis are used [114]:

$$\delta^{pdf}\sigma = \pm\sqrt{\frac{1}{N_{var}-1}\sum_{k=1}^{N_{var}}(\sigma^{(k)} - \langle\sigma\rangle)^2}\,, \tag{7.6}$$

where N_{var} is the number of variations and $< \sigma >$ is the mean value of the cross-section and it is calculated as:

$$\langle \sigma \rangle = \frac{1}{N_{var}} \sum_{k=1}^{N_{var}} \sigma^{(k)} \, . \tag{7.7}$$

- **α_s variations**

  The α_s uncertainty is estimated using two sets of proton PDFs from the NNPDF30 analysis with different values of α_s assumed in the fits, $\alpha_s = 0.119$ and $\alpha_s = 0.117$ (with nominal α_s being 0.118) were used here:

$$\delta^{\alpha_s}\sigma = \frac{\sigma(\alpha_s = 0.119) - \sigma(\alpha_s = 0.117)}{2} \, . \tag{7.8}$$

7.3. Z+jets p_{T} measurement

The differential production cross section for Z + jets events is calculated using the following equation:

$$\frac{d\sigma}{dp_T^Z} = \frac{u(N^Z)}{\Delta p_T^Z} \frac{1}{L_{int}} \, , \tag{7.9}$$

where $u(N^Z)$ corresponds to the unfolded Z yields in each Z boson p_T bin of width Δp_T^Z and L_{int} is the integrated luminosity.

Systematic uncertainties are propagated through the unfolding procedure, and their impact on the final distributions are summarised in Tab. 7.2. The overall uncertainty contribution for each analysis bin, including statistical data uncertainty, is reported in Fig. 7.4.

In the first bins, the largest systematic uncertainty comes from the luminosity measurement. Above 300 GeV the uncertainties on the unfolding and the muon scale rises, becoming the dominant ones. The differential measurement uncertainty is dominated by systematic effects up to 800 GeV and by statistical uncertainty above that, reaching $\sim$ 55% in the highest bin. The large impact of muon scale uncertainty is due to the variation of bin contents induced by the variation in the p_{T} of the muons. The unfolding uncertainty evolution across the p_{T} range results from the MC's statistical uncertainty, which grows similarly to the data statistical uncertainty, and the difference in shape obtained when using alternative MC for building the migration matrix.

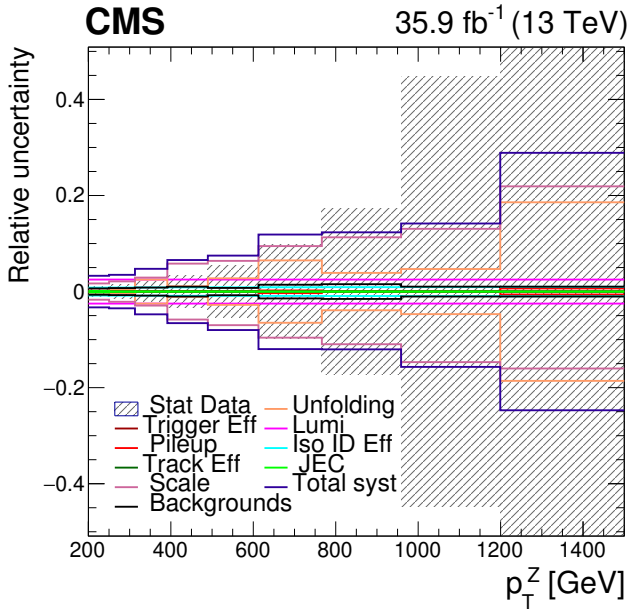


Figure 7.4.: Summary of the statistical and systematic uncertainties contributing to the Z + jets p_T measurement.

Table 7.2.: Breakdown of the sources of systematic uncertainty for the Z + jets channel as the relative error per p_T bin of the analysis.

p_T bin [GeV]	Pile-up [%]		ID-iso [%]	tracking [%]	JEC [%]		Scale [%]		Lumi [%]	Trigger [%]	Backgrounds [%]	Unfolding [%]	**Total [%]**	
200-250	-0.18	+0.19	±0.8	±0.02	-0.18	+0.14	-1.69	+1.69	±2.50	±0.01	±0.66	±0.71	-3.29	+3.29
250-312	-0.15	+0.15	±0.8	±0.02	-0.05	+0.04	-2.12	+2.12	±2.50	±0.01	±0.73	±0.34	-3.49	+3.49
312-391	-0.01	+-0.00	±0.9	±0.02	-0.01	+0.01	-2.92	+2.92	±2.50	±0.01	±0.84	±2.50	-4.74	+4.74
391-489	-0.07	+0.10	±0.9	±0.02	-0.00	+0.00	-5.82	+5.82	±2.50	±0.01	±1.02	±1.10	-6.57	+6.57
489-612	-0.01	+-0.00	±0.9	±0.02	-0.00	+-0.00	-6.99	+6.39	±2.50	±0.01	±0.76	±2.76	-8.01	+7.49
612-766	-0.32	+0.24	±0.9	±0.02	-0.00	+0.00	-9.59	+9.47	±2.50	±0.01	±1.43	±6.49	-11.97	+11.87
766-958	-0.02	+0.00	±0.9	±0.02	-0.00	+0.00	-10.96	+11.30	±2.50	±0.01	±1.53	±3.89	-12.03	+12.34
958-1199	-0.01	+0.01	±0.9	±0.02	-0.00	+0.00	-14.69	+13.06	±2.50	±0.01	±1.03	±4.69	-15.68	+14.17
1199-1500	-0.61	+0.65	±1.0	±0.02	-0.00	+-0.00	-16.01	+21.91	±2.50	±0.01	±1.02	±18.59	-24.71	+28.88

The Z p_T distribution is compared with the particle level prediction from the MADGRAPH5_aMC@NLO v2.2.2 both at NLO and LO precision and with SHERPA v2.2.8 +OPENLOOPS v2.0 at NLO QCDxEW. SHERPA v2.2.8 and MADGRAPH5_aMC@NLO v2.2.2 LO predictions display statistical errors only, while MADGRAPH5_aMC@NLO v2.2.2 NLO errors include theoretical uncertainties from variation to renormalization and factorization scales, pdf variations and α_s variation.

The measured differential cross-section is shown in Table 7.3. A comparison of the unfolded cross-section of Z + jets events as a function of the Z p_T to several theoretical predictions is shown in Figure 7.5. The unfolded data are compared to the MADGRAPH5_aMC@NLO v2.2.2 prediction at NLO and LO and the NLO QCD+EW prediction from SHERPA v2.2.8 + OPENLOOPS v2.0. The data's statistical uncertainty is shown by the error bars, while the shaded region gives the total statistical and systematic uncertainty. Systematic uncertainties from the muon selection efficiency dominate the low p_T region, and the key source of systematic uncertainty in the high p_T region is the unfolding uncertainty. The high p_T region is limited by the statistical uncertainty on the data and the statistical uncertainty arising from the simulation samples' limited size. The data shows agreement within uncertainties with all predictions across almost the full p_T range. A deviation of 2.8σ (local) is observed in the data between 950 GeV and 1200 GeV and the effect is also visible in the out-of-the-box Data/MC comparison in Fig. 6.4.

p_T bin [GeV]	cross-section [pb]
200 - 250	6.13e-01 $\pm$ 1.32e-02
250 - 313	2.93e-01 $\pm$ 1.89e-02
313 - 391	1.35e-01 $\pm$ 6.62e-03
391 - 490	5.87e-02 $\pm$ 3.04e-03
490 - 613	2.34e-02 $\pm$ 1.94e-03
613 - 766	7.35e-03 $\pm$ 1.02e-03
766 - 959	2.40e-03 $\pm$ 4.22e-04
959 - 1199	3.37e-04 $\pm$ 1.70e-04
1199 - Inf	1.19e-04 $\pm$ 1.18e-04

Table 7.3.: Z + jets cross-section differential in p_T. The error includes both systematic and statistical uncertainties.

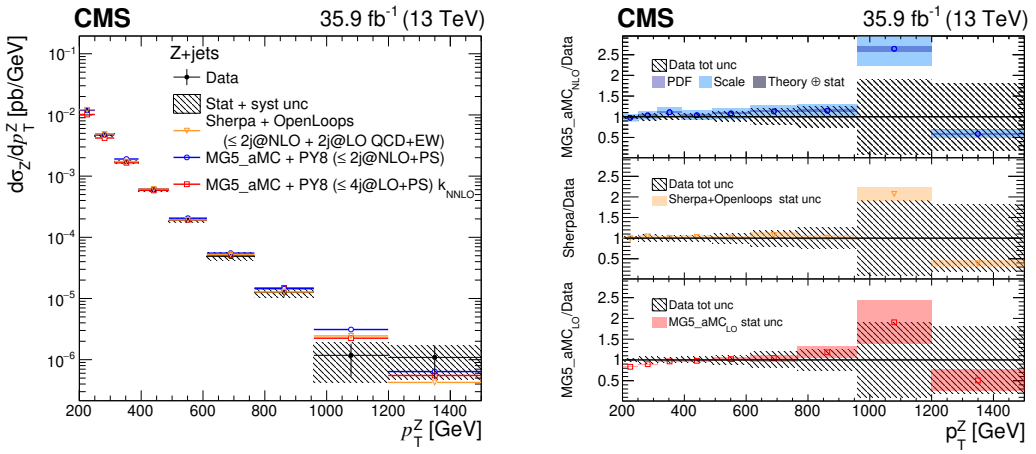


Figure 7.5.: Measured differential cross sections as a function of the boson p_T for Z + jets and the comparison with several theoretical predictions. The error bars in the upper panels represent the statistical uncertainty on the measurement, and the hatched band in the lower and upper panels is the sum in quadrature of the statistical and systematic uncertainty on the measurement. The lower panel shows the ratio of the theory predictions to the unfolded data. The shaded error band on the LO MADGRAPH5_aMC@NLO v2.2.2 and SHERPA v2.2.8 + OPENLOOPS v2.0 calculations is the statistical uncertainty. The dark (light) shaded error band on the NLO prediction from MADGRAPH5_aMC@NLO v2.2.2 represents the PDF (scale) uncertainties.

7.4. γ + jets p_T measurement

The differential production cross section for photon + jet events is calculated similarly to the Z + jets p_T distribution, following the equation:

$$\frac{d\sigma}{dp_T^\gamma} = \frac{u(N^\gamma)}{\Delta p_T^\gamma} \frac{1}{L_{int}} \,. \tag{7.10}$$

While LO and SHERPA v2.2.8 +OPENLOOPS v2.0 predictions errors include statistical uncertainty only, the uncertainty of theoretical predictions for γ + 1 jet MADGRAPH5_aMC@NLO v2.2.2 and JETPHOX samples is obtained by combining MC statistical uncertainty with theoretical uncertainties due to renormalization and factorization scales, pdf variations and α_s variation.

The relative uncertainties from each of these sources of uncertainty are shown in Fig. 7.6. The dominant uncertainty is from variations in the scale for most of the p_T range, with the PDF uncertainty dominating at very high p_T.

The statistical uncertainty and the different systematic uncertainty sources are shown in Fig. 7.7 and in Tab. 7.4. The integrated luminosity uncertainty and the photon isolation and identification efficiencies dominate the low p_T region, each of them leading to a 2.5% impact.

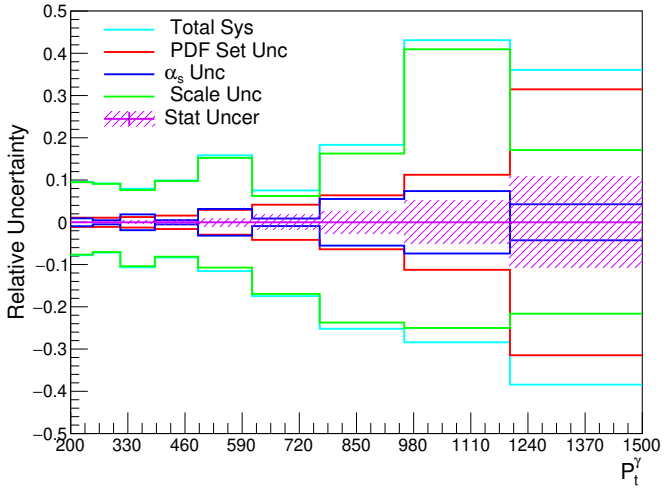


Figure 7.6.: The theoretical uncertainties on the JETPHOX predictions as a function of the photon p_T.

The unfolding uncertainty also contributes significantly, especially in the second bin, driven by the variation in the p_T spectrum obtained when using alternative MC samples for building the migration matrix. The key source of the systematic uncertainty in the high p_T region is from the unfolding and the photon scale uncertainties. Systematic uncertainties overall dominate the measurement of the γ + jets p_T distribution up to 1200 GeV and above that systematic and statistical errors have a similar impact. The precision in the high p_T region is therefore limited by the statistical uncertainties in the data and the limited size of the MC samples, which results in large unfolding uncertainty.

Table 7.4.: Breakdown of the sources of systematic uncertainty for the γ + jets channel as the relative error per p_T bin of the analysis.

p_T bin [GeV]	Purity [%]	Pile-up [%]		photon ID-iso [%]	JEC [%]		Scale [%]		Lumi [%]	Trigger [%]	Unfolding [%]	**Total [%]**
200-250	±1.1	-0.04	+0.02	±2.55	-0.18	+0.17	-1.55	+2.16	±2.50	±0.23	±1.43	-4.31 +4.56
250-312	±0.8	-0.03	+0.03	±2.55	-0.07	+0.06	-0.49	+0.54	±2.50	±0.40	±2.27	-4.35 +4.36
312-391	±0.6	-0.07	+0.09	±2.55	-0.02	+0.02	-0.58	+1.17	±2.50	±0.44	±1.43	-3.96 +4.09
391-489	±0.5	-0.03	+0.02	±2.55	-0.00	+0.00	-2.13	+2.15	±2.50	±0.37	±1.46	-4.45 +4.46
489-612	±0.3	-0.11	+0.09	±2.55	-0.00	+-0.00	-3.55	+3.69	±2.50	±0.47	±1.69	-5.35 +5.44
612-766	±0.4	-0.05	+0.02	±2.54	-0.00	+0.00	-4.67	+4.79	±2.50	±0.56	±1.11	-6.02 +6.11
766-958	±0.3	-0.05	+0.03	±2.53	-0.00	+0.00	-5.54	+6.02	±2.50	±0.91	±1.90	-6.92 +7.31
958-1199	±0.2	-0.11	+0.10	±2.52	-0.00	+0.00	-6.24	+7.24	±2.50	±0.84	±4.61	-8.58 +9.33
1199-1500	±0.2	-0.31	+0.29	±2.50	-0.00	+-0.00	-6.96	+8.63	±2.50	±2.21	±6.39	-10.33 +11.53

The unfolded photon cross-section, differential in p_T is reported in Table 7.5. A comparison of the unfolded distribution to theoretical predictions from three different generators; JETPHOX, SHERPA v2.2.8 + OPENLOOPS v2.0, and MADGRAPH5_aMC@NLO v2.2.2 at two orders of the perturbative expansion, LO and NLO, is shown in Figure 7.8. The LO prediction from MADGRAPH5_aMC@NLO v2.2.2 shows significant disagreement in the shape of the photon

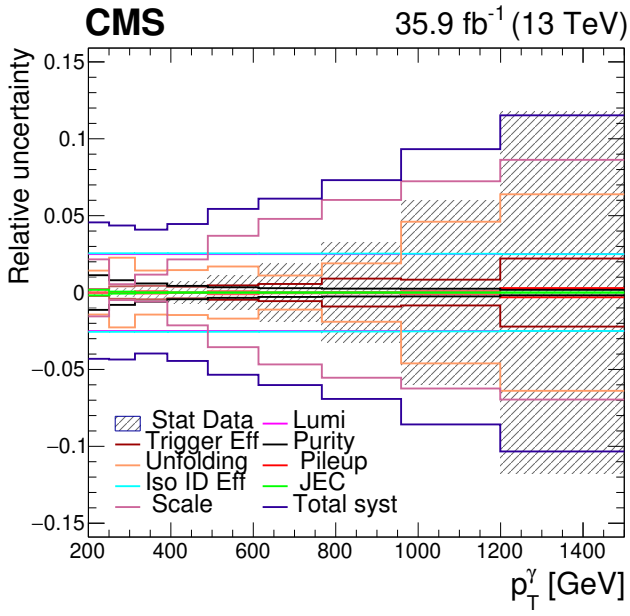


Figure 7.7.: Summary of the statistical and systematic uncertainties contributing to the γ + jets p_T measurement.

p_T distribution, particularly at boson p_T below around 600 GeV. The corresponding NLO calculation shows much better agreement across the full range of p_T. The SHERPA v2.2.8 + OPENLOOPS v2.0 calculation overpredicts the data by $20-30\%$ in the p_T region below 500 GeV and then is consistent within uncertainties for the rest of the p_T range. The NLO prediction from JETPHOX is shown with uncertainties from the variation in PDFs and the renormalization, factorization, and fragmentation scales. The scale uncertainty is estimated by varying the values of the three scales up or down by a factor of two. This variation is applied on the scales either simultaneously, individually or by fixing one and varying the other two. The final scale uncertainty is estimated by taking the largest deviation from the nominal value among all the possible variations. The theoretical uncertainties are at the level of 10% at low p_T and increase to 40% at high p_T. The scale uncertainty is dominant at both low and high p_T, while the PDF uncertainty increases from the per cent level at low p_T to 10% at high p_T and 30% in the last bin. The prediction is mostly consistent with data within uncertainties with a general overprediction at the level of around 20% below p_T of 500 GeV.

7.5. Z/γ ratio

Since the Z + jets and γ + jets cross-sections have been extracted in the same fiducial region, as described in Section 5.4, a direct comparison between the two processes can be made. The

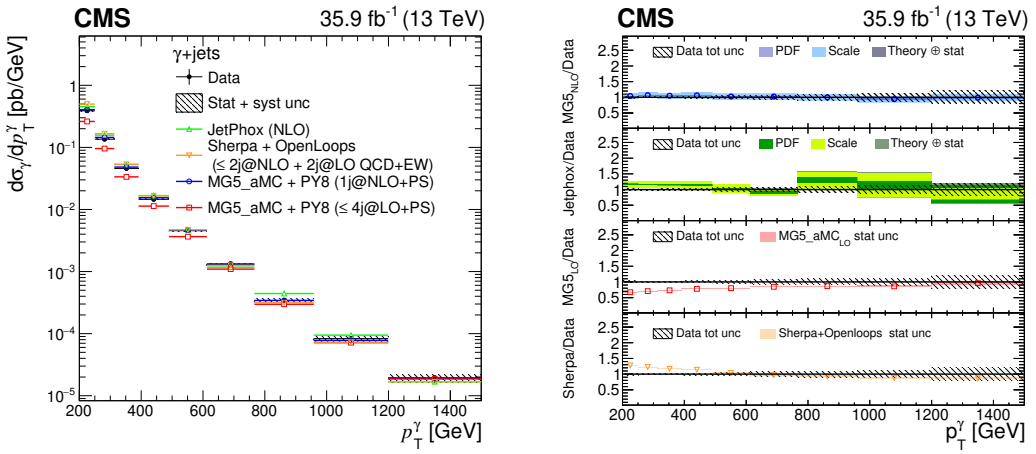


Figure 7.8.: Measured differential cross sections as a function of the boson p_T for γ + jets and the comparisons with several theoretical predictions. The error bars in the upper panels represent the statistical uncertainty on the measurement, and the hatched band in the lower and upper panels is the sum in quadrature of the statistical and systematic uncertainty on the measurement. The lower panel shows the ratio of the theory predictions to the unfolded data. The shaded error band on the LO MADGRAPH5_aMC@NLO v2.2.2 and SHERPA v2.2.8 + OPENLOOPS v2.0 calculations is the statistical uncertainty. The dark (light) shaded error band on the NLO prediction from MADGRAPH5_aMC@NLO v2.2.2 and the JETPHOX prediction represents the PDF (scale) uncertainties.

p_T bin [GeV]	cross-section [pb]
200 - 250	1.94e+01 $\pm$ 4.48e-01
250 - 313	8.42e+00 $\pm$ 7.67e-02
313 - 391	3.56e+00 $\pm$ 4.51e-02
391 - 490	1.42e+00 $\pm$ 3.19e-02
490 - 613	5.54e-01 $\pm$ 2.01e-02
613 - 766	1.96e-01 $\pm$ 9.84e-03
766 - 959	6.44e-02 $\pm$ 4.32e-03
959 - 1199	1.96e-02 $\pm$ 1.68e-03
1199 - Inf	5.62e-03 $\pm$ 7.91e-04

Table 7.5.: γ + jets cross-section differential in p_T. The error includes both systematic and statistical uncertainties.

cross-section ratio is measured differentially in p_T, restricting to $p_T^V > 200$ GeV and $|y^V|<1.4$ and requiring the presence of at least one jet with $p_T > 100$ GeV.

The statistical and systematic uncertainties on the ratio are propagated to the ratio considering the systematics affecting both channels to be 100% correlated. The errors are propagated in each p_T bin separately. This allows the correct cancellation between systematics uncertainties correlated across the two processes, such as PU reweighting, JEC and luminosity uncertainty.

The remaining uncertainties are from object selection efficiency, photon purity estimation and the unfolding procedure. A detailed breakdown of the uncertainties on the ratio between the $Z+$ jets and γ + jets channels is presented in Table 7.6.

The differential cross-section ratio of the two processes $Z+$ jets and γ + jets is shown in Fig. 7.9 and compared to the theoretical prediction at NLO from MADGRAPH5_aMC@NLO v2.2.2 and NLO QCD+EW from SHERPA v2.2.8 + OPENLOOPS v2.0. The comparison with MADGRAPH5_aMC@NLO v2.2.2 shows consistency within the uncertainties across the entire p_{T} range. In contrast, SHERPA v2.2.8 + OPENLOOPS v2.0 underpredicts the data by 10-20% at low p_{T} but then is consistent with data within uncertainties for $p_{\mathrm{T}} > 300\,\mathrm{GeV}$.

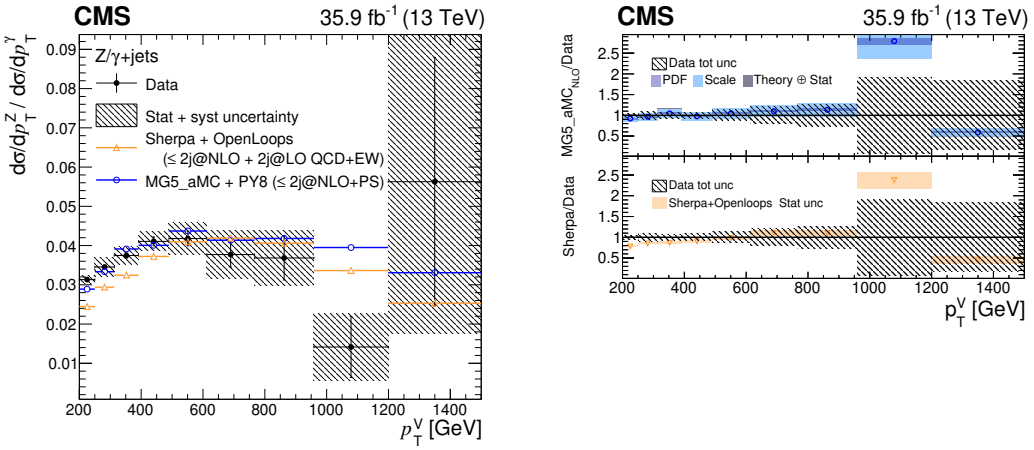


Figure 7.9.: Differential cross-section ratio of $Z+$ jets over γ + jets as a function of the vector boson transverse momentum compared to the theoretical prediction from MADGRAPH5_aMC@NLO v2.2.2 and SHERPA v2.2.8 + OPENLOOPS v2.0. Only vector bosons produced centrally, with $|y^V|<1.4$, in association with one or more jets are considered. The error bars in the upper panel represent the statistical uncertainty on the measurement, and the hatched band in the lower and upper panels is the sum in quadrature of the statistical and systematic uncertainty on the measurement. The lower panel shows the ratio of the theoretical prediction to the unfolded data. The shaded error band on the SHERPA v2.2.8 + OPENLOOPS v2.0 calculation is the statistical uncertainty. The dark (light) shaded error band on the NLO prediction from MADGRAPH5_aMC@NLO v2.2.2 represents the PDF (scale) uncertainties.

Table 7.6.: Breakdown of the uncertainty on the $Z+$ jets/γ + jets ratio per p_{T} bin of the analysis.

p_{T} bin [GeV]	Pile-up [%]	JEC [%]	Lumi [%]	μ tracking [%]	μ ID-iso [%]	γ ID-iso [%]	μ trigger [%]	γ trigger [%]	Z background [%]	γ purity [%]	μ scale [%]	γ scale [%]	Z unfolding [%]	γ unfolding [%]	stat [%]	**Total [%]**
200-250	0.03	0.02	0.00	0.00	0.85	2.55	0.01	0.23	0.49	1.07	1.69	1.98	0.71	1.43	0.79	3.55
250-312	0.07	0.04	0.00	0.00	0.89	2.55	0.01	0.40	0.62	0.82	2.12	0.25	0.34	2.27	1.18	6.93
312-391	0.08	0.01	0.00	0.00	0.99	2.55	0.01	0.44	0.74	0.63	2.92	0.91	2.50	1.43	1.78	5.88
391-489	0.04	0.00	0.00	0.00	0.90	2.55	0.01	0.37	0.86	0.49	5.82	2.00	1.10	1.46	2.75	5.96
489-612	0.05	0.00	0.00	0.00	0.91	2.55	0.01	0.47	0.64	0.30	6.39	3.41	2.76	1.69	4.47	9.61
612-766	0.17	0.00	0.00	0.00	0.98	2.54	0.01	0.56	1.21	0.35	9.47	4.68	6.49	1.11	8.49	16.11
766-958	0.16	0.00	0.00	0.00	0.92	2.53	0.01	0.91	1.25	0.32	11.30	5.87	3.89	1.90	15.22	19.11
958-1199	0.08	0.00	0.00	0.00	0.96	2.52	0.01	0.84	1.57	0.20	13.06	6.43	4.69	4.61	54.92	60.41
1199-1500	0.27	0.00	0.00	0.01	1.02	2.50	0.01	2.21	1.07	0.25	21.91	9.81	18.59	6.39	56.29	68.68

7.6. Collinear Z boson emission

The angular separation between the Z boson and the closest jet ($\Delta R^{\min}_{Z,j}$) is an interesting observable that, at high p_{T} of the jets, is sensitive to the real emission of Z bosons in dijet events. For Z + 1-jet like production at LO, the Z and jets are produced back-to-back, the hadronic recoil in the event is balanced by the boson p_{T} and the angular separation between the Z and the jet is π. At NLO, both QCD and EWK corrections can modify this observable with both real and virtual contributions. In particular, a real Z boson can be emitted from one of the jets such that the angular separation between the two is small, so $\Delta R^{\min}_{Z,j}$ tends to small values. The relative Feynman diagrams contributing to the angular distribution are shown in Figure 7.10. This region is also populated by Z + multi-jets events at LO, but the contribution from collinear Z emission can be enhanced by increasing the p_{T} threshold on the leading jet. This is done to select highly energetic di-jet systems, where one of the two jets irradiates a real Z boson. For this study of the angular Z distribution, the p_{T} and η cuts on the Z boson are removed and instead, the threshold on the leading jet is progressively increased up to 500 GeV. Figures 7.11 and 7.12 shows the ΔR distribution between the Z boson and the leading and subleading jets, with the first set of plots obtained considering lead jets with $p_{\mathrm{T}} > 100$ GeV, and the second one implements a 500 GeV threshold on the leading jet p_{T}. It can be observed how the higher threshold requirement, does not affect the ΔR distribution for the leading jet, while the collinear region is enhanced in the subleading jet distribution. This is expected as the lead jet threshold requirement increases the probability for the subleading jet to irradiate the Z boson. Figure 7.13 shows the $\Delta R^{\min}_{Z,j}$ for progressively higher thresholds on the lead jet: nominal lead jet selection, 300 and 500 GeV respectively. The enhancement of collinear Z events is shown in the low $\Delta R^{\min}_{Z,j}$ region.

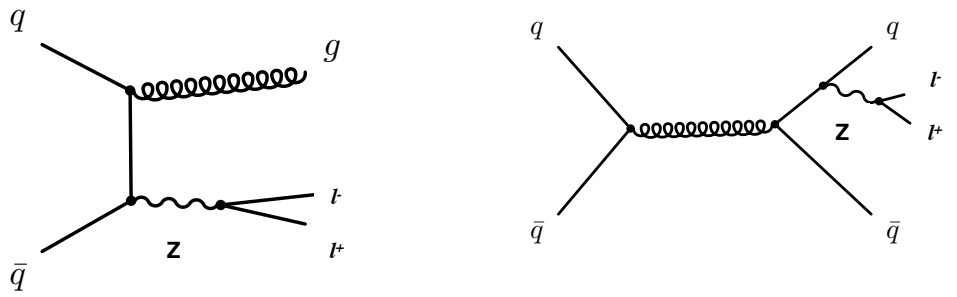


Figure 7.10.: Diagrams showing the configuration where (a) the jet and Z boson are back to back (b) where the Z boson is emitted collinear to one of the jets in the dijet final state.

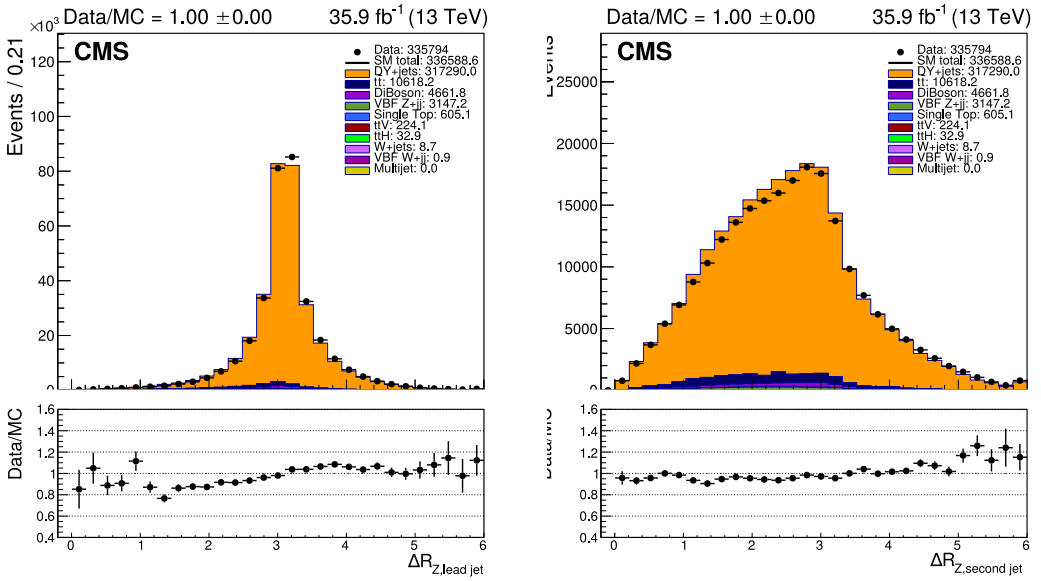


Figure 7.11.: Distribution of the angular separation between the Z boson and the lead jet (a) or the subleading jet (b). Distributions are obtained selecting events which satisfy the nominal analysis jet requirement jet$_1$ $p_T > 100$ GeV.

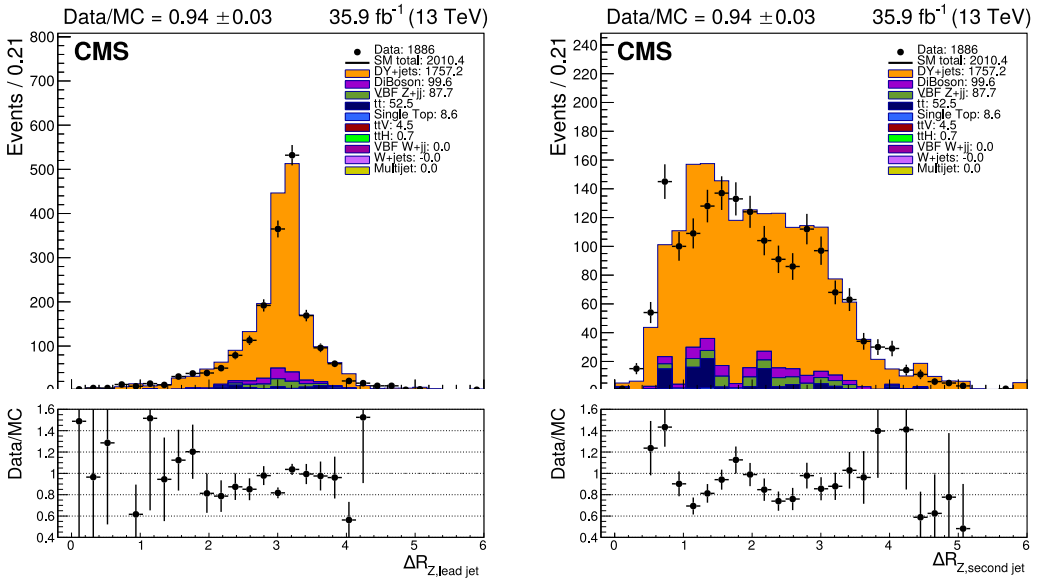
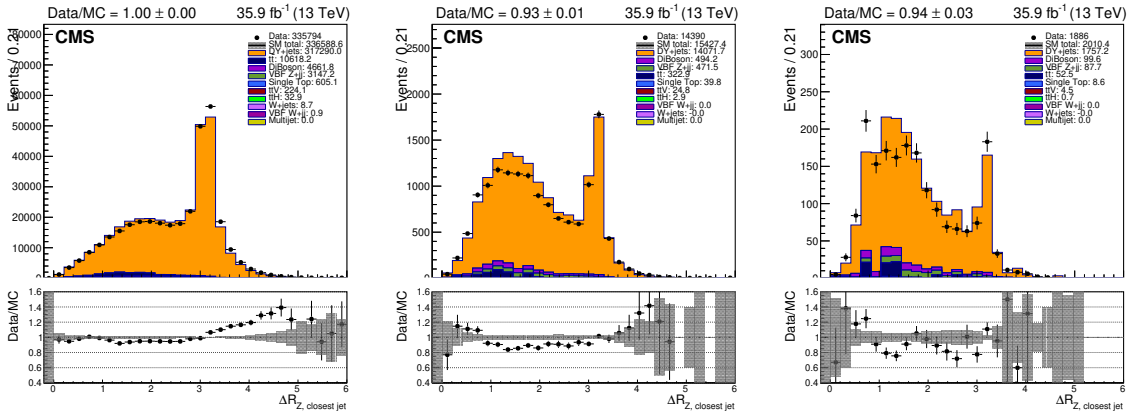


Figure 7.12.: Distribution of the angular separation between the the Z boson and the lead jet (a) or the subleading jet (b), for a threshold on the p_T of the lead jet of jet$_1$ $p_T > 500$ GeV.



(a) (b) (c)

Figure 7.13.: Distribution of the angular separation between the Z boson and the closest jet for progressively higher thresholds on the p_{T} of the leading jet of (a) 100 GeV, (b) 300 GeV, and (c) 500 GeV.

7.6.1. Unfolding

The details of the unfolding are very similar to those discussed in Sec. 7.1. However, in the $\Delta R^{\mathrm{min}}_{Z,j}$ distribution, it is crucial to have a matching not only between the generator level Z boson and the reconstructed candidate but also with the closest jet. When building the migration matrix, the first step consists of selecting generator level events in the fiducial region, then the Z boson is matched to a reconstructed candidate, and once the closest jet is found, a second matching procedure is performed to identify the generator level jet. The ΔR between generator-level objects and between the matched detector-level ones is used to populate the migration matrix. The unfolding of the $\Delta R^{\mathrm{min}}_{Z,j}$ between the reconstructed Z boson and the closest jet has also been performed using the NLO MADGRAPH Z + jets. The binning for the unfolding has been chosen to have at least 68% of events in the diagonal. The resulting response matrix is shown in Fig.7.14.

7.6.2. Systematics

The sources of systematic uncertainties considered for the collinear emission study are very similar to the Z p_{T}, and details can be found in Section 7.2.

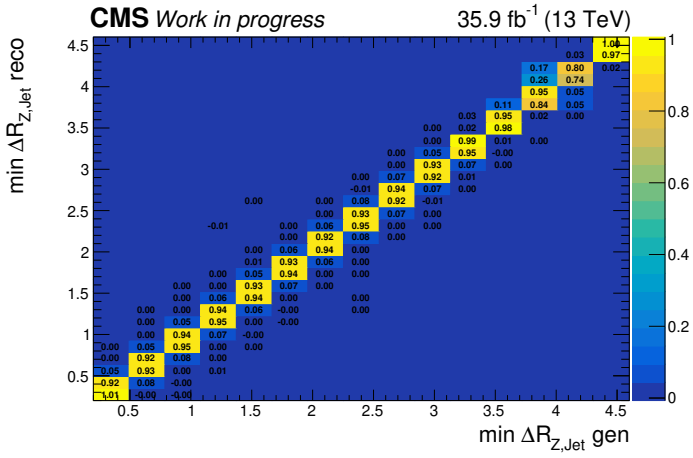
An overview of the systematic uncertainties is shown in Fig. 7.15. The breakdown of the different systematic sources in each analysis bin is shown in Tab. 7.7 and Tab. 7.8 for the 300 and 500 GeV threshold, respectively.

Table 7.7.: Breakdown of the sources of systematic uncertainty for the collinear Z study, as the relative error per ΔR bin of the analysis for the 300 GeV threshold on the leading jet.

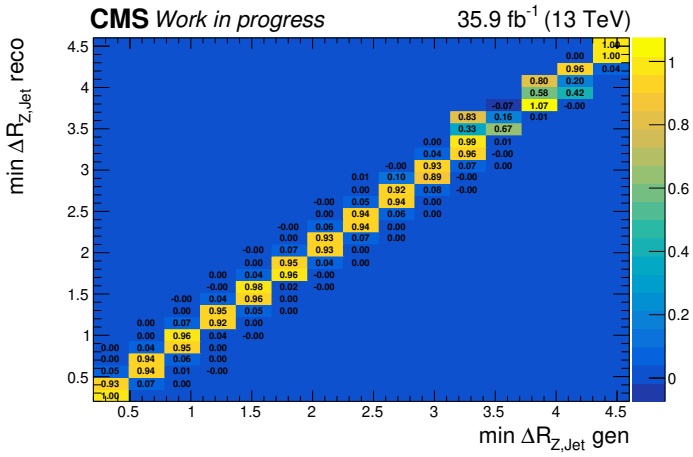
min $\Delta R_{Z,j}$ bin	Pile-up [%]		ID-iso [%]	tracking [%]	JEC [%]		Scale [%]		Lumi [%]	Trigger [%]	Backgrounds [%]	Unfolding [%]	**Total [%]**
0.2-0.5	-0.55	+0.40	±0.9	±0.02	-10.59	+9.98	-8.04	+10.79	±2.50	±0.02	±9.32	±8.35	-18.46 +19.49
0.5-0.8	-0.61	+0.51	±1.0	±0.02	-0.23	+2.78	-0.17	+1.99	±2.50	±0.04	±10.89	±2.28	-11.47 +11.96
0.8-1.1	-0.30	+0.30	±1.1	±0.02	-3.19	+1.67	-3.56	+2.74	±2.50	±0.05	±10.31	±1.59	-11.80 +11.25
1.1-1.4	-0.36	+0.32	±1.1	±0.02	-5.49	+3.35	-3.85	+3.78	±2.50	±0.05	±7.57	±1.42	-10.57 +9.61
1.4-1.7	-0.15	+0.15	±1.1	±0.02	-1.07	+0.06	-0.21	+0.14	±2.50	±0.05	±7.90	±1.43	-8.54 +8.47
1.7-2.0	-0.34	+0.26	±1.1	±0.02	-0.93	+2.53	-1.38	+2.06	±2.50	±0.05	±5.55	±1.54	-6.59 +7.16
2.0-2.3	-0.48	+0.40	±1.1	±0.03	-0.08	+2.28	-1.42	+1.71	±2.50	±0.07	±4.81	±1.71	-5.98 +6.46
2.3-2.5	-0.43	+0.50	±1.1	±0.03	-1.89	+3.99	-3.01	+3.94	±2.50	±0.08	±4.58	±2.17	-6.78 +8.05
2.5-2.8	-0.81	+0.91	±1.1	±0.03	-0.82	+1.79	-1.20	+0.22	±2.50	±0.09	±4.75	±2.36	-6.19 +6.30
2.8-3.1	-0.41	+0.42	±1.1	±0.02	-0.12	+1.86	-0.59	+0.95	±2.50	±0.08	±2.83	±1.29	-4.21 +4.66
3.1-3.4	-0.39	+0.41	±1.1	±0.02	-1.70	+1.20	-0.24	+0.07	±2.50	±0.11	±0.88	±1.17	-3.58 +3.37
3.4-3.7	-0.81	+0.81	±1.1	±0.03	-4.21	+6.79	-6.73	+4.34	±2.50	±0.15	±1.91	±3.75	-9.43 +9.53
3.7-4.0	-0.38	+0.31	±1.0	±0.03	-4.38	+6.60	-12.21	+7.64	±2.50	±0.18	±2.12	±10.81	-17.24 +15.19

Table 7.8.: Breakdown of the sources of systematic uncertainty for the collinear Z study, as the relative error per ΔR bin of the analysis for the 500 GeV threshold on the leading jet.

min $\Delta R_{Z,j}$ bin	Pile-up [%]		ID-iso [%]	tracking [%]	JEC [%]		Scale [%]		Lumi [%]	Trigger [%]	Backgrounds [%]	Unfolding [%]	**Total [%]**
0.2-0.5	-1.15	+1.01	±0.9	±0.02	-1.09	+0.15	-31.77	+20.74	±2.50	±0.01	±14.62	±18.14	-39.51 +31.32
0.5-0.8	-0.86	+0.75	±1.0	±0.02	-0.79	+0.17	-3.49	+0.98	±2.50	±0.05	±9.56	±4.51	-11.52 +10.98
0.8-1.1	-0.13	+0.13	±1.1	±0.02	-0.97	+0.66	-16.63	+15.45	±2.50	±0.05	±10.87	±3.69	-20.42 +19.45
1.1-1.4	-0.16	+0.26	±1.1	±0.02	-0.91	+0.73	-7.66	+8.34	±2.50	±0.05	±11.57	±3.73	-14.65 +15.01
1.4-1.7	-0.20	+0.15	±1.1	±0.02	-0.83	+0.46	-2.98	+2.84	±2.50	±0.05	±6.99	±3.57	-8.87 +8.80
1.7-2.0	-0.14	+0.01	±1.1	±0.02	-0.81	+1.23	-2.91	+0.42	±2.50	±0.06	±5.59	±4.46	-8.22 +7.75
2.0-2.3	-1.93	+1.71	±1.1	±0.02	-1.46	+1.00	-0.02	+3.95	±2.50	±0.08	±7.75	±5.90	-10.40 +11.04
2.3-2.5	-1.53	+1.28	±1.1	±0.03	-0.56	+0.70	-1.18	+6.84	±2.50	±0.07	±5.32	±7.09	-9.49 +11.61
2.5-2.8	-2.00	+1.71	±1.1	±0.03	-1.22	+0.86	-4.69	+5.09	±2.50	±0.08	±5.30	±7.52	-10.93 +11.03
2.8-3.1	-0.27	+0.25	±1.1	±0.02	-0.73	+0.92	-2.22	+0.42	±2.50	±0.07	±7.35	±7.01	-10.78 +10.57
3.1-3.4	-0.38	+0.26	±1.2	±0.02	-1.66	+1.68	-1.24	+1.60	±2.50	±0.11	±1.17	±5.05	-6.24 +6.32
3.4-3.7	-5.22	+6.40	±1.2	±0.01	-2.39	+0.68	-3.09	+3.98	±2.50	±0.25	±8.68	±80.33	-81.11 +81.20
3.7-4.0	-3.55	+2.74	±0.7	±0.03	-1.21	+1.88	-28.15	+3.34	±2.50	±0.13	±1.61	±45.56	-53.78 +45.91



(a)



(b)

Figure 7.14.: The correlation matrices showing the bin migration in the unfolded $\Delta R^{\min}_{Z,j}$ distribution of the Z boson for a leading jet threshold of 300 (a) and 500 (b) GeV. The matrices are almost diagonal, with at least 68% of events populating the same bin both in reconstructed and generator level distributions.

7.6.3. $\Delta R^{\min}_{Z,j}$ measurement

The unfolded distribution for the angular separation between the Z and closest jet is shown in Figure 7.16 compared to the LO and NLO predictions from MADGRAPH5_aMC@NLO v2.2.2. The theoretical predictions are generally consistent with the data within uncertainties for the case where the leading jet p_{T} is above 500 GeV. The LO prediction from MADGRAPH5_aMC@NLO v2.2.2 underpredicts the data for $\Delta R^{\min}_{Z,j}$ above 1.8, while the NLO prediction is consistent within uncertainties for the bulk of the distribution, with the largest discrepancies at $\Delta R^{\min}_{Z,j}$ of 0.5-0.8 and around 3.2.

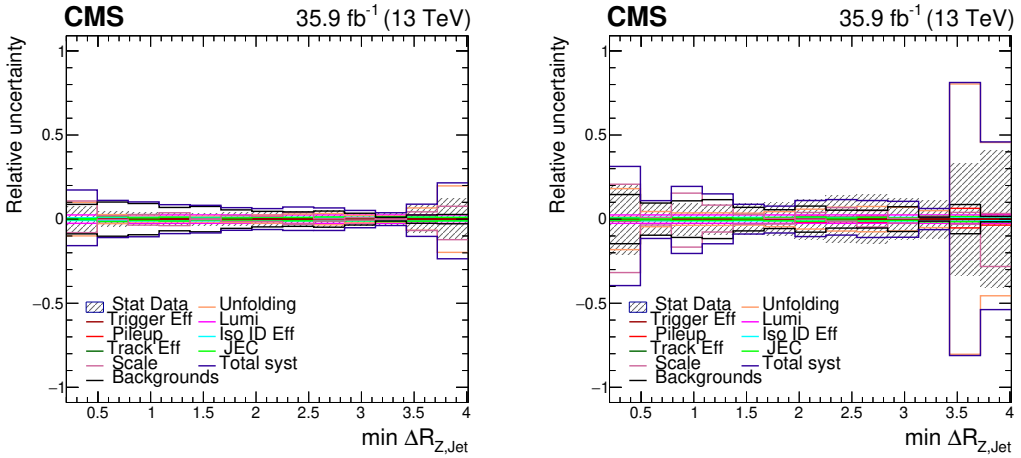


Figure 7.15.: Summary of the statistical and systematic uncertainty contributions to the $\Delta R_{Z,j}^{\min}$ distribution where the leading jet p_{T} is (a)$>$ 300 GeV and (b) $>$ 500 GeV.

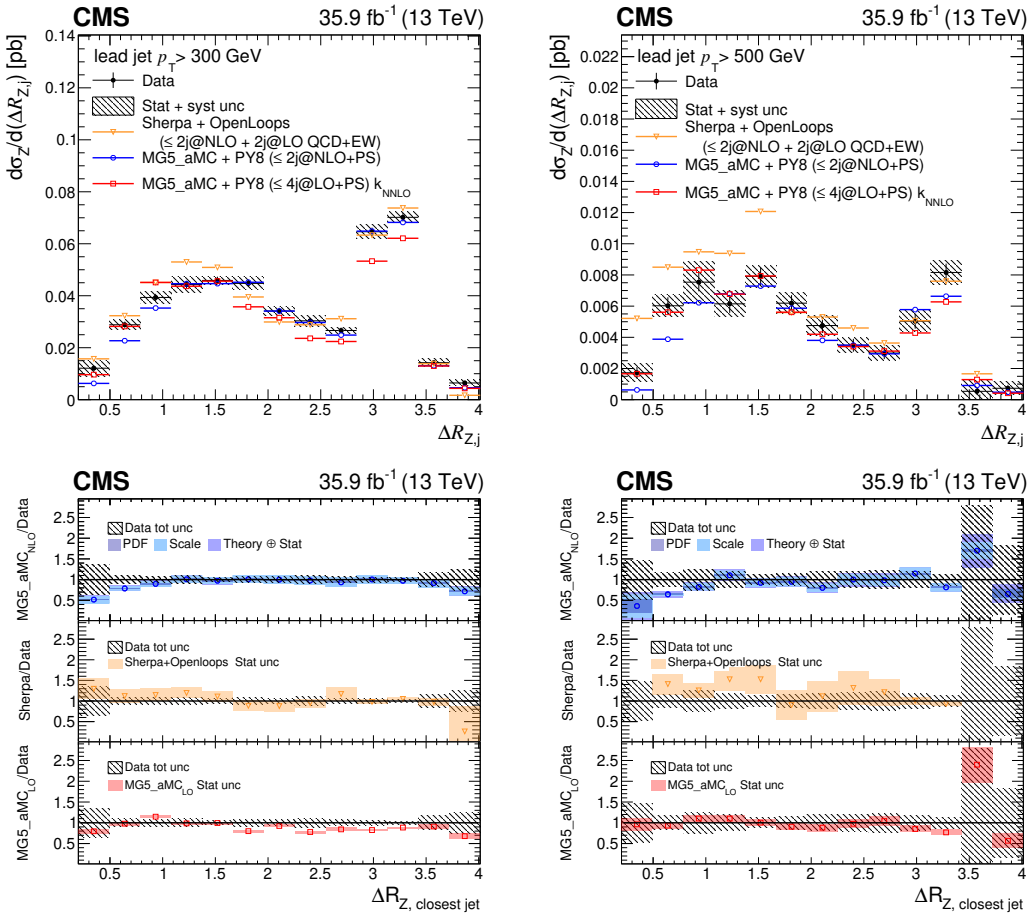


Figure 7.16.: Measured differential cross-section of $Z+$jets as a function of the angular separation between the Z boson and the closest jet and the comparison with theory predictions from MADGRAPH5_aMC@NLO v2.2.2 and SHERPA v2.2.8 + OPENLOOPS v2.0 where the leading jet p_T is > 300 GeV (left) and 500 GeV (right). The error bars in the upper panel represent the statistical uncertainty on the measurement, and the hatched band in the lower and upper panels is the sum in quadrature of the statistical and systematic uncertainty on the measurement. The lower panel shows the ratio of the theory predictions to the unfolded data. The shaded error band on the theory calculation is the statistical uncertainty.

Chapter 8.

Conclusions

Measurements of the cross-section of $Z+$ jets, γ + jets and their ratio differential in p_T, as well as the measurement of collinear Z boson production in the proximity of a jet, have been performed using the 35.9 fb^{-1} dataset collected by the CMS experiment during the 2016 proton-proton collisions at an energy of $\sqrt{s} = 13$ TeV.

The data have been selected requiring muon and photon triggers for the $Z+$ jets and γ + jets channels, respectively. In addition, muon and photon identification and isolation requirements have been applied.

Simulated samples, both for the signal and background processes, have been selected similarly to the data, and dedicated corrections have been derived and applied to improve the simulated description of experimental data.

Different approaches have been adopted to estimate the background contribution to the different analysis channels. A simulation-based approach has been implemented to estimate diboson and VBF Z+jets backgrounds in the $Z+$ jets channel, while a data-driven technique is used to estimate the level of contamination in the γ + jets selection coming from misidentified hadrons.

After background contributions are subtracted, the data distributions are unfolded to the particle level and compared with different theoretical predictions in different fiducial regions, targeting specifically the high p_T region for the Z/γ ratio, and real Z boson collinear emission. The Z/γ p_T measurement is performed in a final state characterized by $p_T^V > 200$ GeV and $|y^V| < 1.4$ and at least one jet with $p_T > 100$ GeV and $|\eta| < 2.4$. For the collinear Z boson production, the fiducial region is defined by having a Z boson reconstructed from muons with $p_T > 30$ GeV and $|\eta| < 2.4$ together with at least one jet with $|\eta| < 2.4$ and $p_T > 300$ or 500 GeV.

In the unfolding stage, systematic uncertainties are included considering various sources that can affect the physics objects considered and the overall event reconstruction. Among them, the dominant ones are found to be the energy calibration of the leptons and photons,

due to the large effects on bin contents in the high p_T range, as well as the unfolding procedure, due to the limited statistics of simulation used to build the migration matrices.

Overall, good agreement is observed between the measurement and the most precise theoretical predictions. The $Z+$ jets p_T distribution is modelled well by MADGRAPH5_aMC@NLO v2.2.2 Z+0,1,2 jets NLO prediction, with a small overshoot of the data across the p_T spectrum. The comparison with MADGRAPH5_aMC@NLO v2.2.2 Z+0,1,2,3,4 jets generated at LO and normalized to NNLO QCD cross-section shows a trend in the p_T shape, underestimating the data in the low p_T region while predicting more events in the high p_T bins. The prediction that performs better is the SHERPA v2.2.8 + OPENLOOPS v2.0 which has NLO QCDxEW precision. The γ + jets p_T distribution instead is predicted better by the JETPHOX and MADGRAPH5_aMC@NLO v2.2.2 γ+0,1 jets at NLO in QCD, with the overall normalization of JETPHOX overshooting data by 10-15%, while the MADGRAPH NLO prediction provides a precise description of the p_T spectrum in every bin. The SHERPA prediction exhibits a trend in the p_T spectrum overestimating the data in the 200-250 GeV bin by 30% and decreasing with p_T with a $\sim$10% undershoot in the last bin. The LO MADGRAPH prediction shows an opposite trend with respect to SHERPA, underestimating by 40% the data in the low end of the spectrum and improving data description up to the high p_T region. The unfolded Z/γ ratio is compared with MADGRAPH5_aMC@NLO v2.2.2 NLO and SHERPA v2.2.8 +OPENLOOPS v2.0 predictions and the comparison follows the described behaviour in both $Z+$ jets and γ + jets channels with MADGRAPH performing better in the low p_T range due to the SHERPA γ + jets overestimate, and SHERPA following the evolution of the ratio better in the high p_T region thanks to a better description of the $Z+$ jets spectrum. As expected, looking at the Z/γ ratio, the NLO EWxQCD prediction exhibits a reduction in the cross-section growing with p_T. This is due to the EW corrections reducing $Z+$ jets production with respect to γ + jets processes.

For the $\Delta R^{\min}_{Z,j}$ distributions, the same predictions of the $Z+$ jets p_T analysis have been compared to the unfolded distributions. In the $p_T^{j_1} > 300$ GeV selection, the different MC are in agreement with the data. In particular, the MADGRAPH NLO sample describes well the data from $1 < \Delta R^{\min}_{Z,j} < 3.4$, while it underestimates the data in the high ΔR region and especially in the collinear region. SHERPA v2.2.8 + OPENLOOPS v2.0 predictions present an opposite behaviour, overshooting the measurement but providing a better description in the small ΔR range, while MADGRAPH LO provides a stable agreement across the ΔR range. It has to be noted that statistical fluctuations have a considerable impact in the MADGRAPH and especially the SHERPA samples, however in the back-to-back peak, the better agreement is provided by SHERPA and MADGRAPH NLO predictions, while LO calculation shows a normalization deficit. The effect is more pronounced in the $p_T^{j_1} > 500$ GeV selection, where both LO and NLO MADGRAPH undershoot the back-to-back peak, while the inclusion of NLO EW corrections provides a 10-15% cross-section enhancement and better predicts the data. In this selection, MC statistics are limited and all three predictions are affected by fluctuations. Overall, the

better description of the data distribution is provided by LO MADGRAPH predictions normalized to the NNLO (in QCD) cross-section. This can be related to the higher jet multiplicity considered in the matrix element calculation for MADGRAPH LO with respect to MADGRAPH NLO predictions. As mentioned, SHERPA provides a better description of the back-to-back region, while overestimates data in the small ΔR range, however, statistical fluctuations are a limiting factor in the interpretation.

Further improvements that can be implemented in the analysis would be including the full Run 2 dataset and adding the electron channel to improve data statistics in the high p_{T} region of the $\mathrm{Z}+$ jets channel, as well as improving the statistics of the simulated samples to constrain the systematic uncertainty related to the unfolding procedure. Nevertheless, the results presented in this thesis extend previous CMS 8 TeV results of the Z/γ [115] ratio well beyond the 1 TeV mark, being the first measurement of this quantity performed at a centre of mass energy of 13 TeV. The collinear region study constitutes the first measurement at colliders of real Z boson radiation from a jet, providing crucial landmarks for theoretical calculation development. The comparison with MADGRAPH5_aMC@NLO v2.2.2 Z+0,1,2,3,4 jets LO normalized at NNLO cross-section shows a trend in the p_{T} shape.

Appendix A.

Photon Purity Studies

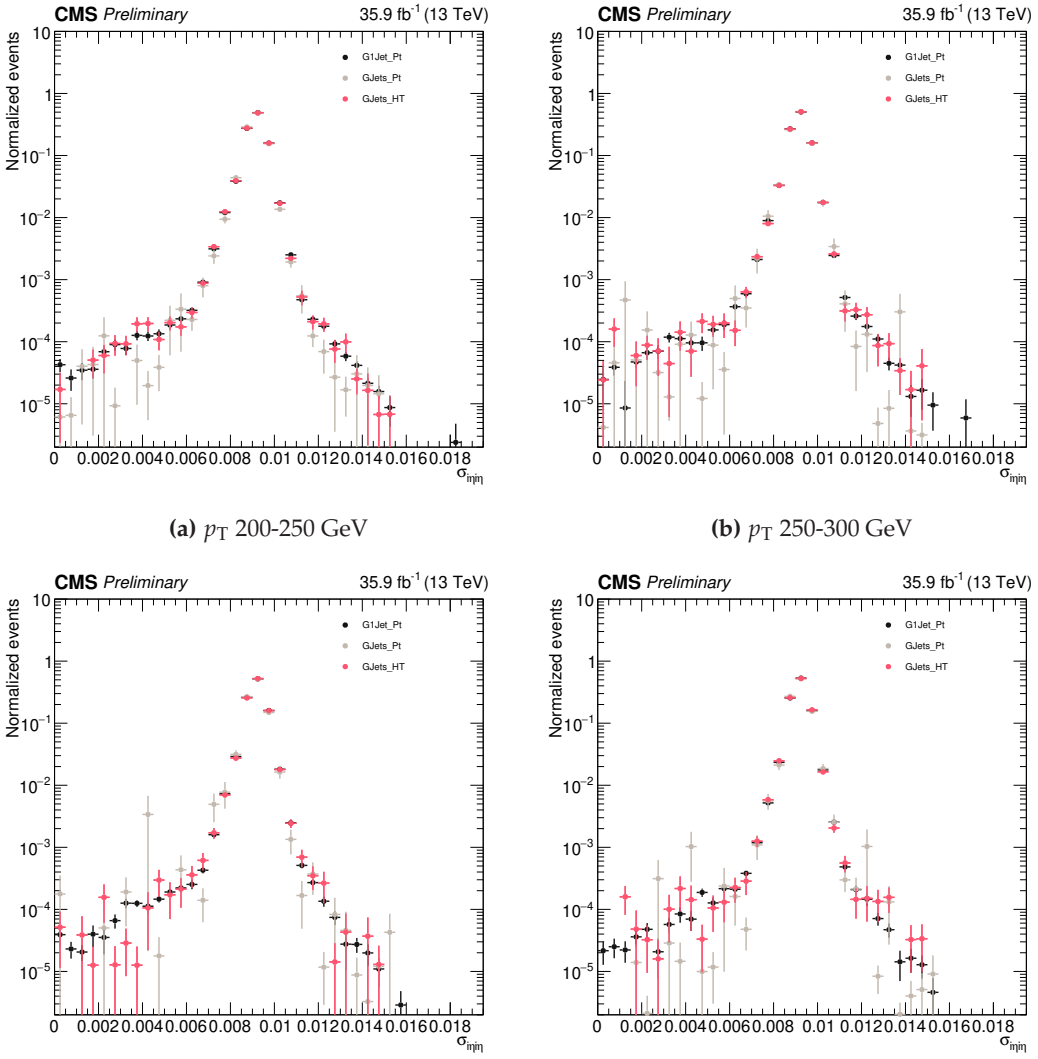
As explained in Sec.6.2.2 the purity of the selected γ + jets events (as described in Sec. 5.4) is determined with a *Template fit* method and a purity value is extracted in different p_{T} bins. The fitting is performed on an extended range (up to 0.02) of the $\sigma_{i\eta i\eta}$ variable, to include a region which is largely dominated by hadronic objects misidentified as photons, and therefore allowing for a stronger constraint on the background normalization. The data distribution is fitted to a signal and a background templates. In the following sections a detailed description on how the templates are built and which uncertainties affect the template shapes is provided.

A.1. Signal template

The signal template is obtained from MADGRAPH γ + 1 jet MC sample, selecting all the photon candidates matched to a particle level isolated photon coming from the hard scattering. Alternatively, Sherpa and MADGRAPH LO γ + jets samples have been analysed, without providing significant discrepancies (see Figure A.1 and Figure A.2). Given the larger statistics of NLO γ + 1 jet MC sample, the extracted signal template provides a smoother shape in the tails of the $\sigma_{i\eta i\eta}$ distribution, the maximum variation from the nominal template in each bin is considered as a source of systematic uncertainty on the signal shape in the fit.

$\sigma_{i\eta i\eta}$ reweight

The reweighting procedure, briefly described in Sec.6.2.2, consists of a histogram remapping based on the percentile distribution, correcting the MC $\sigma_{i\eta i\eta}$ shape to be closer to the one observed in data. The correction is derived from Z$\rightarrow$ *ee* tag-and-probe samples, as electron and photon showers develop in a similar way inside the ECAL crystals. Events from data and simulations are required to present two isolated electrons, opposite in charge, with an invariant mass within 20 GeV of the Z boson mass. Tight identification requirements are applied to the



(a) p_T 200-250 GeV

(b) p_T 250-300 GeV

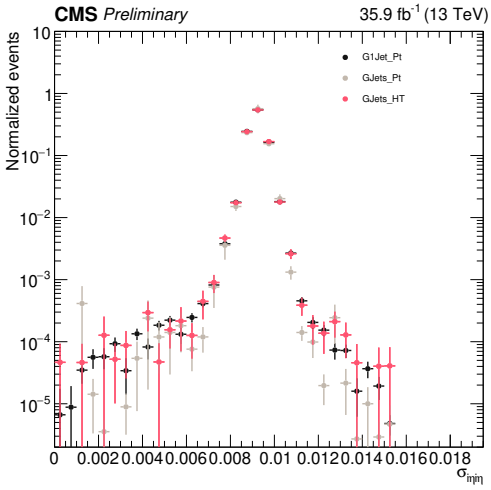
(c) p_T 300-350 GeV

(d) p_T 350-450 GeV

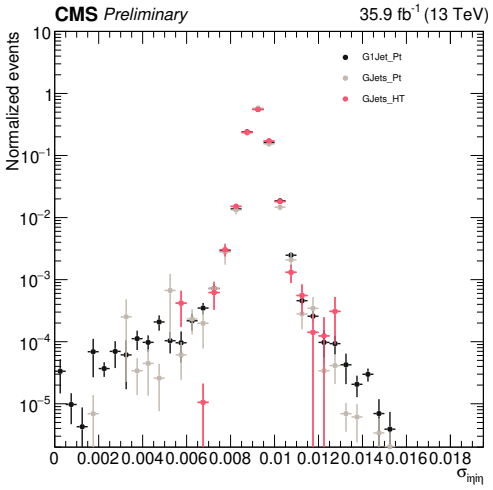
Figure A.1.: Distribution of the photon shower shape variable $\sigma_{i\eta i\eta}$ in the photon transverse momentum bins, ranging from 200 to 450 GeV. The distributions are obtained with different MC γ + jets samples. In black is reported theMADGRAPH5_aMC@NLO v2.2.2 NLO γ + 1 jet, in red MADGRAPH5_aMC@NLO v2.2.2 LO γ + jets sample and the grey points represent the SHERPA v2.2.8 LO sample. To compare the differences in shape the distribution has been normalized to the same area

tag objects and the invariant mass distribution of the probes is fitted to extract the fraction of signal electrons.

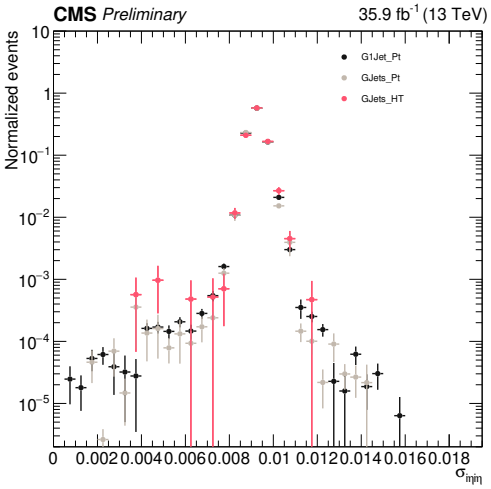
Starting from the $\sigma_{i\eta i\eta}$ distributions in data and simulation, the normalized cumulative distribution is derived (percentile distribution) and used as input for determining the correction map. An association between MC and data $\sigma_{i\eta i\eta}$ value is implemented, based on the percentile value.



(a) p_T 450-600 GeV



(b) p_T 600-800 GeV



(c) p_T 800-Inf GeV

Figure A.2.: Distribution of the photon shower shape variable $\sigma_{i\eta i\eta}$ in the photon transverse momentum bins, starting from 450 GeV. The distributions are obtained with different MC γ + jets samples. In black is reported theMADGRAPH5_aMC@NLO v2.2.2 NLO γ + 1 jet, in red MADGRAPH5_aMC@NLO v2.2.2 LO γ + jets sample and in gray the SHERPA v2.2.8 LO sample. To compare the difference in shape the distribution has been normalized to the same area

In Figure A.3 the percentile distribution for data and MC is reported. To derive a mapping granular enough to be used in a per-event correction a very fine binning (10000) has been chosen.

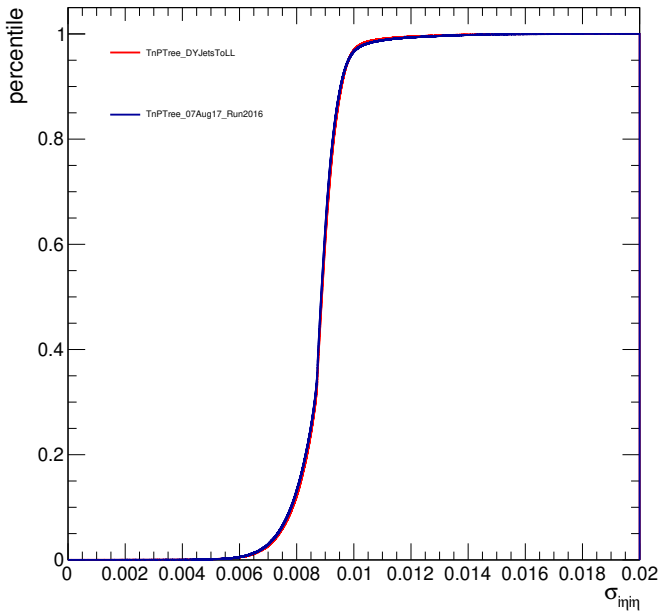


Figure A.3.: Percentile distribution of $\sigma_{i\eta i\eta}$ variable

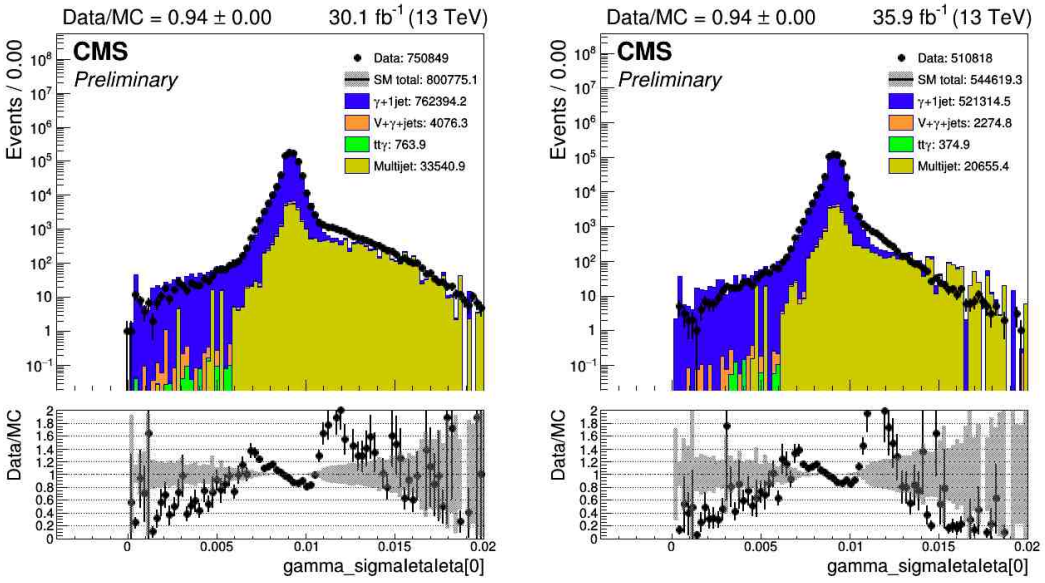
A.2. Background template

The data/MC comparison displayed in Figure A.4 and Figure A.5 shows how QCD MC fails to reproduce the $\sigma_{i\eta i\eta}$ distribution, resulting in large discrepancies in the high $\sigma_{i\eta i\eta}$ region.

The background template is therefore taken from data. To define a fake enriched region from which taking the background template we choose a sideband in the charged hadron isolation variable (I_{ph}^{chHad}). The fiducial I_{ph}^{chHad} interval for "tight" photons is $I_{ph}^{chHad} < 0.202$ GeV. To produce the background template, the $\sigma_{i\eta i\eta}$ distribution from photon candidates with $I_{ph}^{chHad} \in [10.0 - 15.0]$ GeV has been choosen.

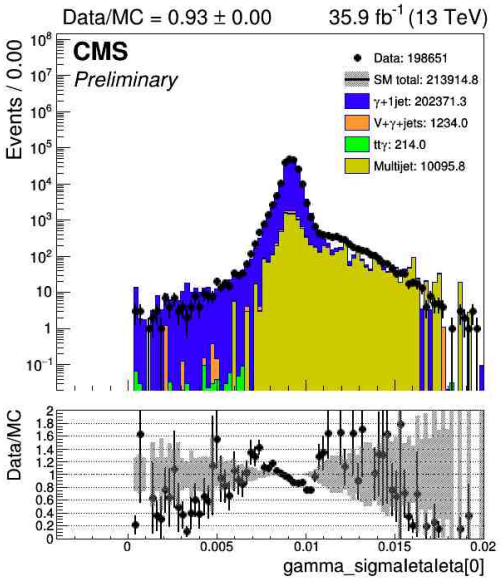
Care has to be taken in defining a suitable sideband, as the presence of true photons in the chosen region can lead to a background template that looks like signal and therefore an unphysical low purity estimation. To correct the template shape from residual contribution from true photons, we subtract the fraction of leaked signal events, estimating it with the MC. However, this procedure relies on the correct normalization of the MC prediction in a region of the I_{ph}^{chHad} which is not subject to precise calibration. A test of sanity for the procedure of signal leakage subtraction is to check if the normalized $\sigma_{i\eta i\eta}$ shape varies across different sidebands. To evaluate the shape difference we define the quantity $R_{\sigma_{i\eta i\eta}}$:

$$R_{\sigma_{i\eta i\eta}} = \frac{N_{\sigma_{i\eta i\eta}<0.0094}}{N_{\sigma_{i\eta i\eta}>0.014}}$$

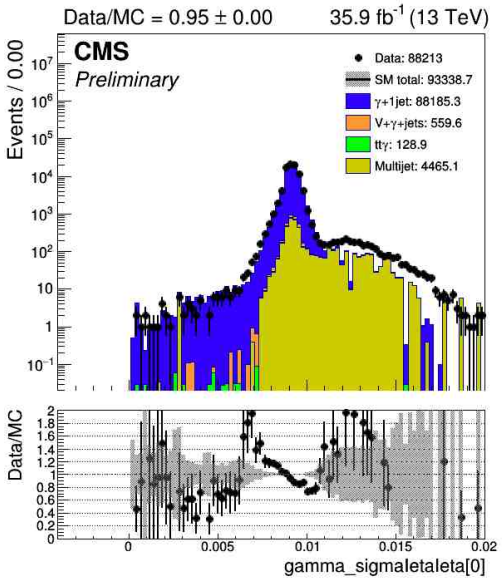


(a) Inclusive

(b) p_{T} 200-250 GeV



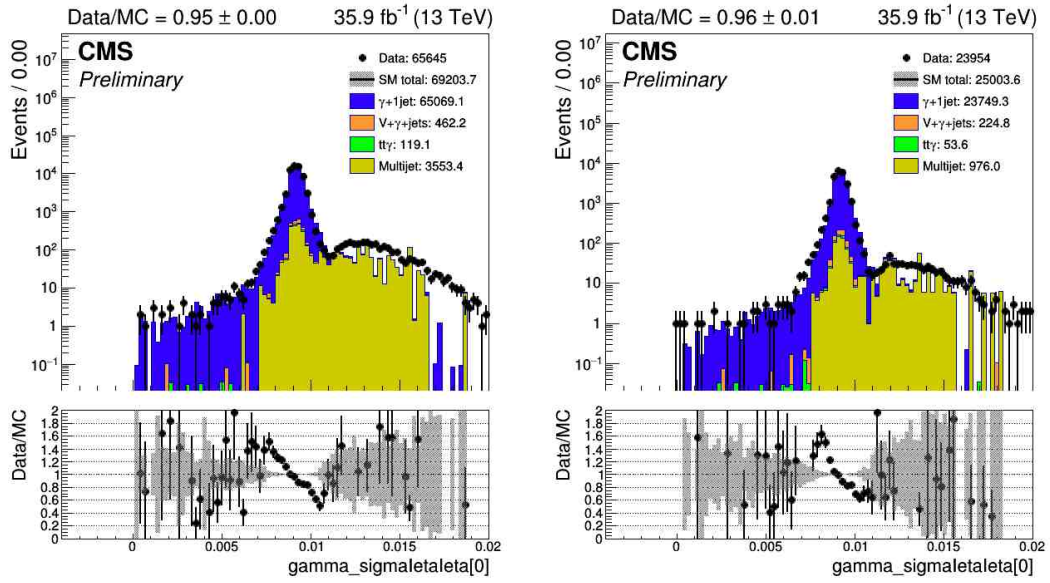
(c) p_{T} 250-300 GeV



(d) p_{T} 300-350 GeV

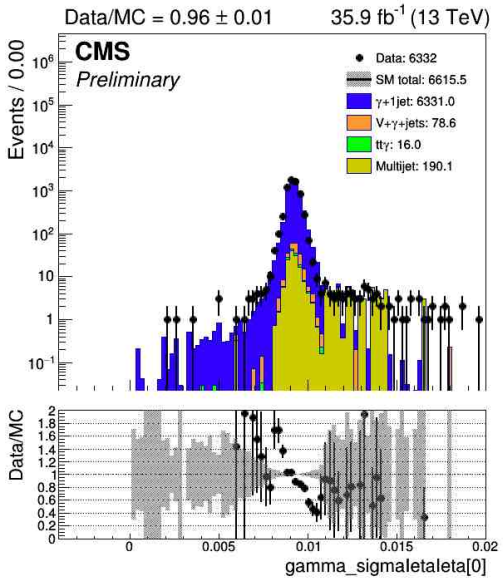
Figure A.4.: Distribution of the photon shower shape variable $\sigma_{i\eta i\eta}$ in the inclusive case (a) and in the photon transverse momentum bins (b-d), covering the range from 200 to 450 GeV. Especially in the most statistically significant bins the Data/MC comparison in the high $\sigma_{i\eta i\eta}$ region shows a large QCD MC mismodelling.

as the ratio of the number of events in the tight region, where true photons are expected to contribute, and the number of events in the high $\sigma_{i\eta i\eta}$ tail, which is dominated by fake photons ($N_{\sigma_{i\eta i\eta}} > 0.014$).

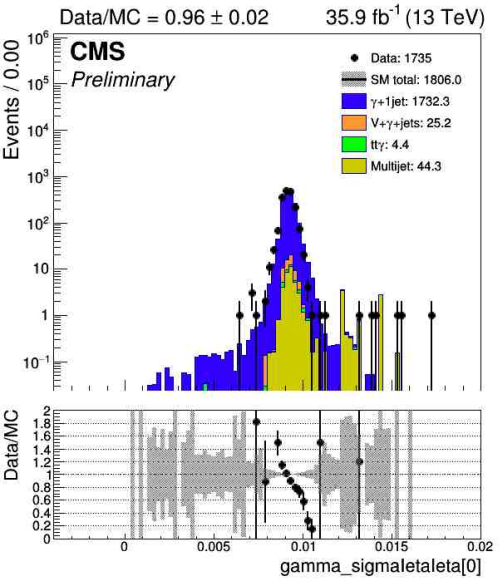


(a) p_T 350-450 GeV

(b) p_T 450-600 GeV



(c) p_T 600-800 GeV



(d) p_T 800-Inf GeV

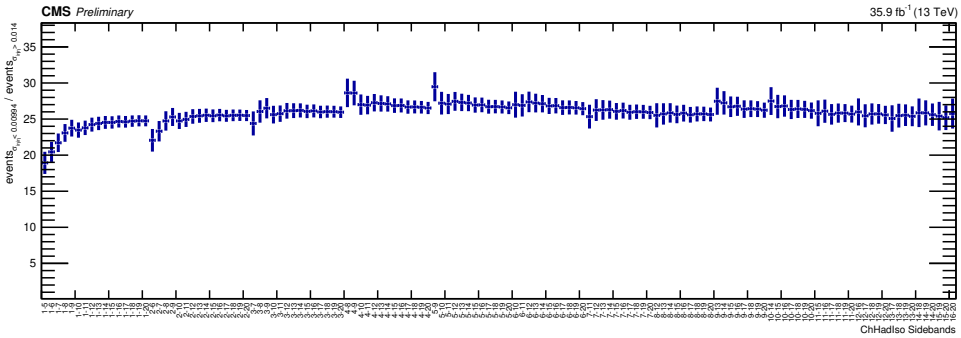
Figure A.5.: Distribution of the photon shower shape variable $\sigma_{i\eta i\eta}$ in the photon transverse momentum bins, starting from 350 GeV. Especially in the most statistically significant bins the Data/MC comparison in the high $\sigma_{i\eta i\eta}$ region shows a large QCD MC mismodelling.

In Figure A.6 and A.7 the behaviour of the ratio $R_{\sigma_{i\eta i\eta}}$ can be observed across different I_{ph}^{chHad} sidebands. 190 sidebands have been investigated, considering all possible combinations of integer boundaries between 1 and 20 GeV. It can be observed how $R_{\sigma_{i\eta i\eta}}$ is varying across sidebands with a lower boundary close to the signal region, whereas it stabilizes when the lower boundary is far enough, at around 8 GeV for all the p_T bins. This is due to an imperfect signal leakage subtraction as can be seen in Figures A.8 and A.9, where the ratio is plotted

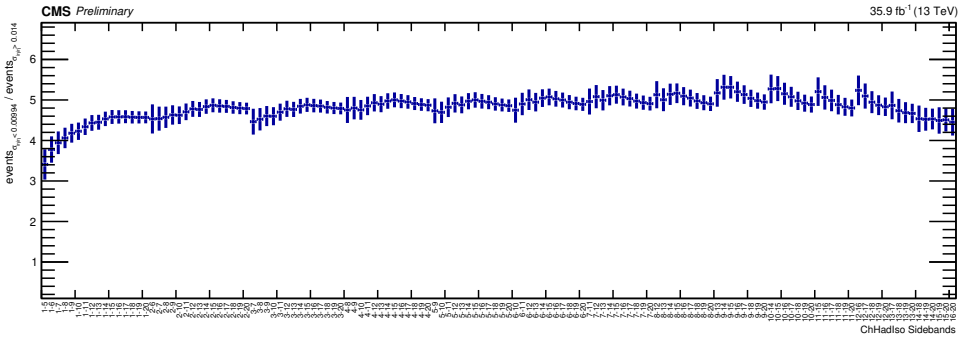
against the fraction of γ + jets events estimated with the MC; the $R_{\sigma_{i\eta i\eta}}$ values are very similar for sidebands with low signal leakage, whereas when the leakage increases, a clear trend is observed.

The sideband $I_{ph}^{chHad} \in [10.0 - 15.0]$ has been chosen optimizing the low signal leakage expected across the different p_T bins while maintaining good statistics.

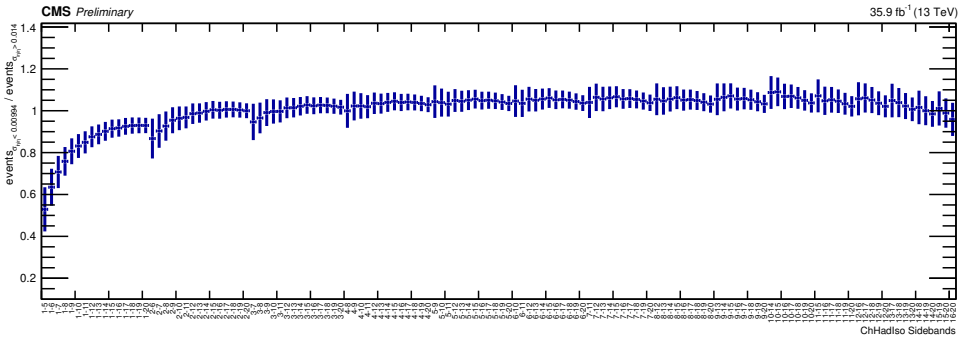
In the fit to determine the purity of the selected γ + jets events, we included as a systematic effect on the sideband choice, the maximum variation from the nominal background template, across all the sidebands that presents a $R_{\sigma_{i\eta i\eta}}$ value independent of the signal leakage. In FiguresA.10 A.11 all the templates from the selected sidebands are plotted normalized to the unit area and despite the large number of sidebands considered, the shape difference is found to be small, compatible within statistical error.



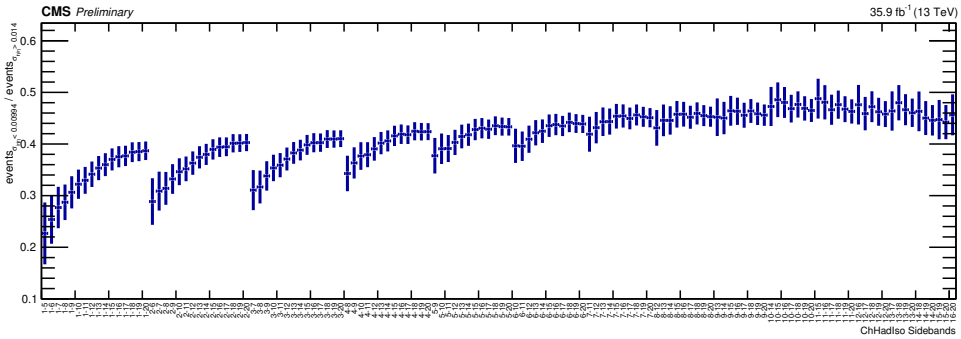
(a) p_{T} 200-250 GeV



(b) p_{T} 250-300 GeV

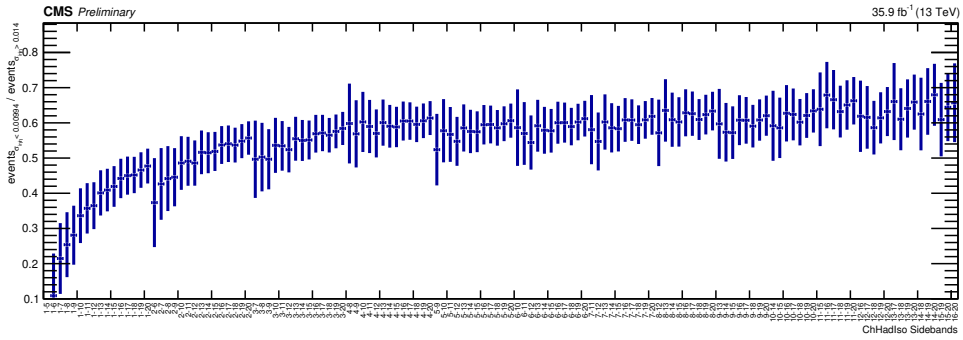


(c) p_{T} 300-350 GeV

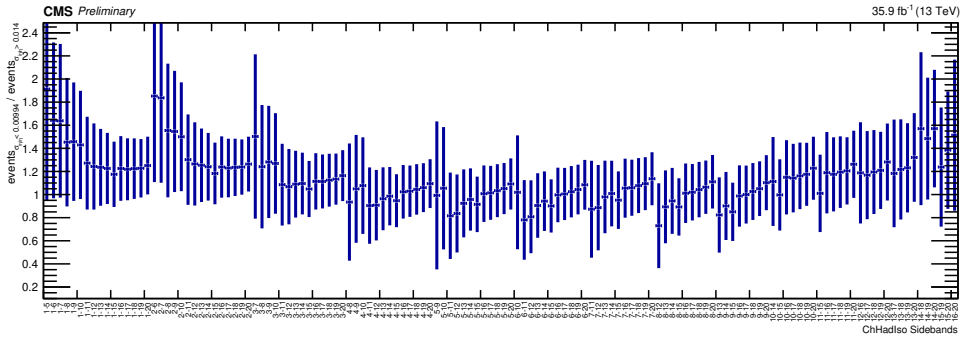


(d) p_{T} 350-450 GeV

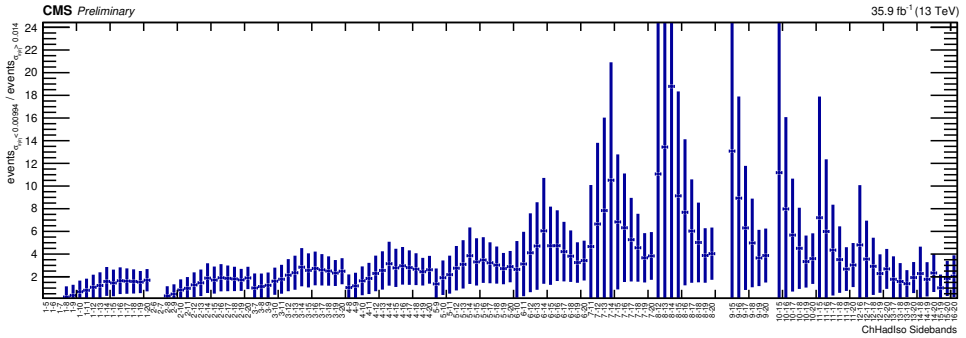
Figure A.6.: Distribution of the $R_{\sigma_{i\eta i\eta}}$ for each charged hadron isolation sideband in the photon transverse momentum bins. The bins range from 200 to 450 GeV. It can be observed how the ratio stabilizes as the lower edge of the sidebands is increasingly far from the tight region.



(a) p_{T} 450-600 GeV

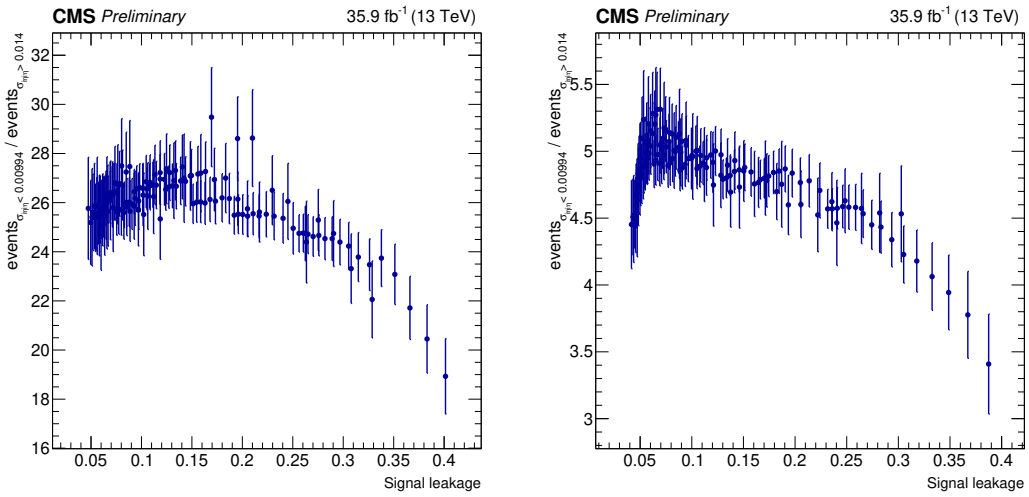


(b) p_{T} 600-800 GeV



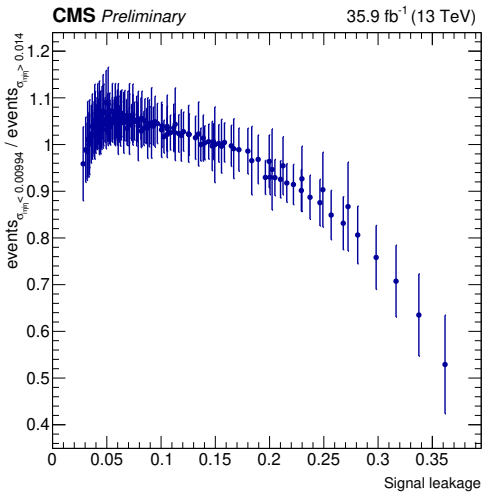
(c) p_{T} 800-Inf GeV

Figure A.7.: Distribution of the $R_{\sigma_{i\eta i\eta}}$ for each charged hadron isolation sideband in the photon transverse momentum bins, starting form 450 GeV. It can be observed how the ratio stabilizes as the lower edge of the sidebands is increasingly far from the tight region.

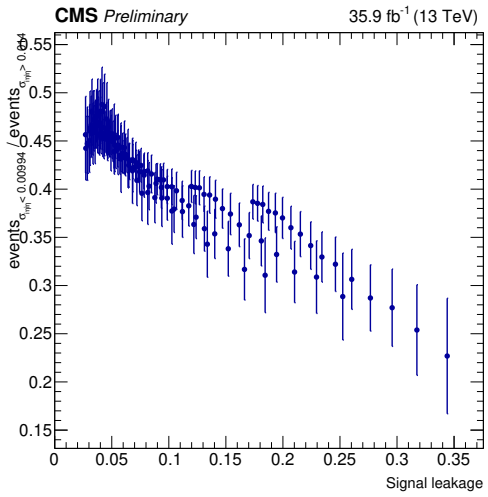


(a) p_{T} 200-250 GeV

(b) p_{T} 250-300 GeV

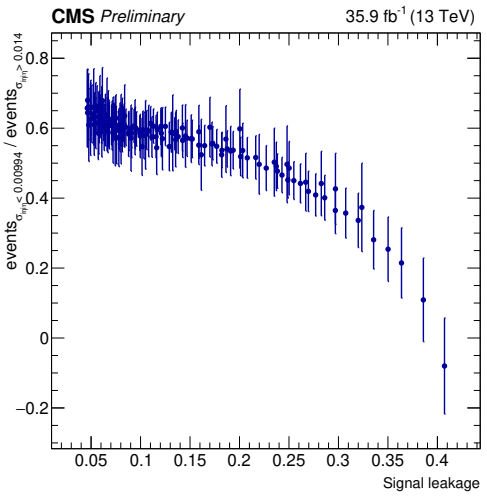


(c) p_{T} 300-350 GeV

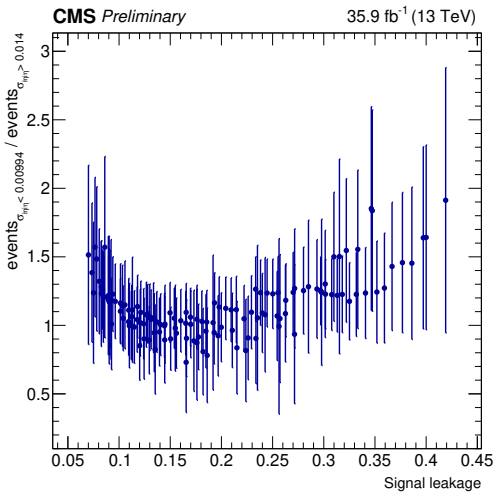


(d) p_{T} 350-450 GeV

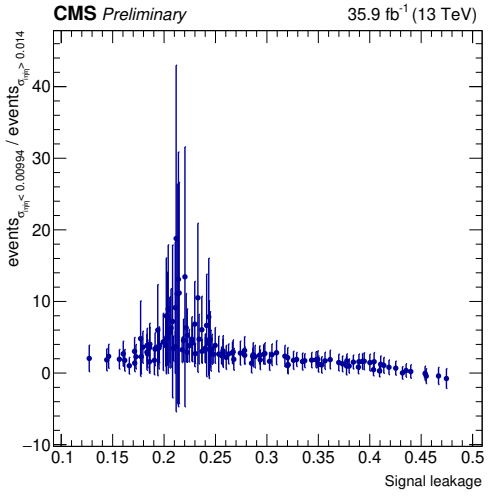
Figure A.8.: Distribution of the $R_{\sigma_{i\eta i\eta}}$ with respect to the signal leakage expected in each sideband. The plots are obtained in the photon transverse momentum bins, from 200 to 450 GeV. It can be observed how the ratio is mostly stable for small values of leakage, while shows a strong dependence when the fraction of real photons in the sidebands increases.



(a) p_{T} 450-600 GeV

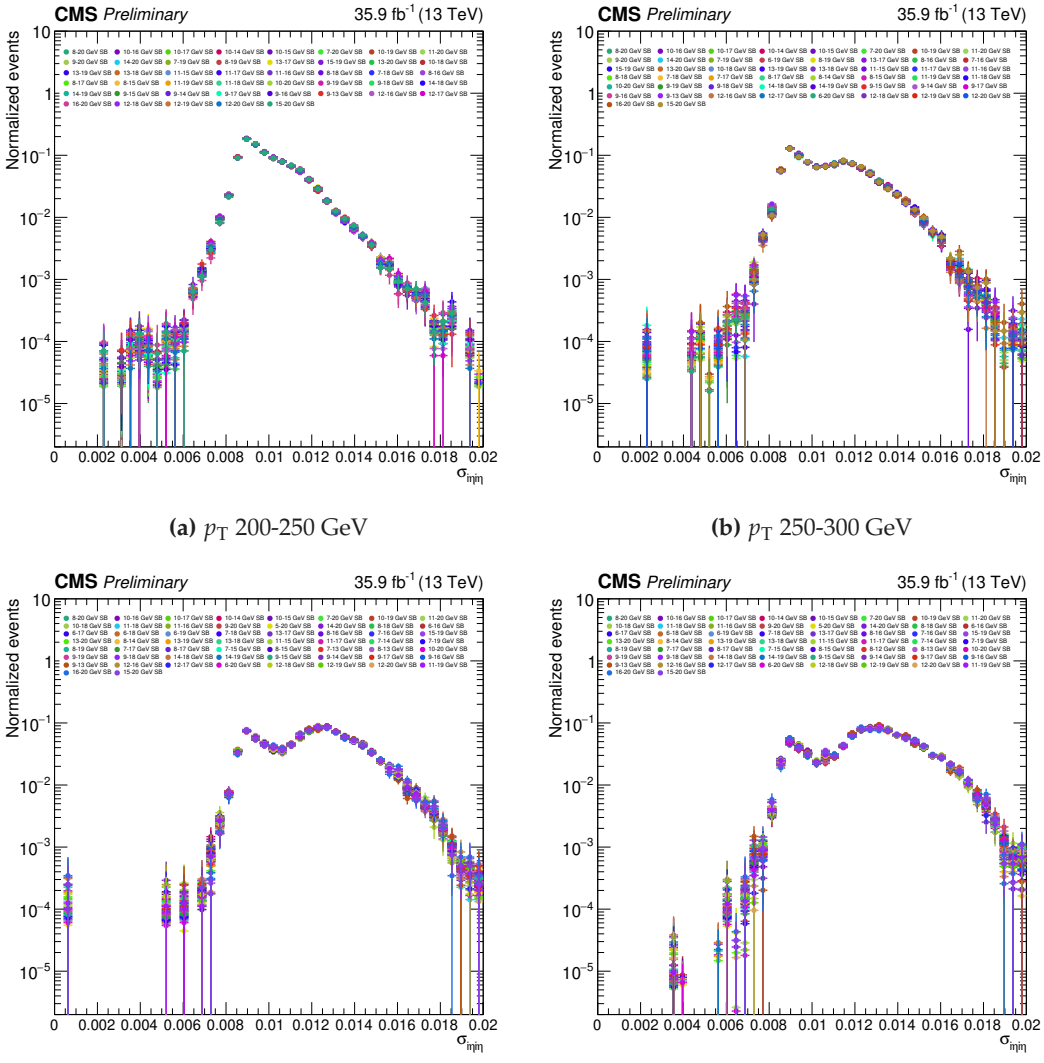


(b) p_{T} 600-800 GeV



(c) p_{T} 800-Inf GeV

Figure A.9.: Distribution of the $R_{\sigma_{i\eta i\eta}}$ with respect to the signal leakage expected in each sideband. The plots are obtained in the photon transverse momentum bins, starting from 450 GeV. It can be observed how the ratio is mostly stable for small values of leakage, while shows a strong dependence when the fraction of real photons in the sidebands increases.



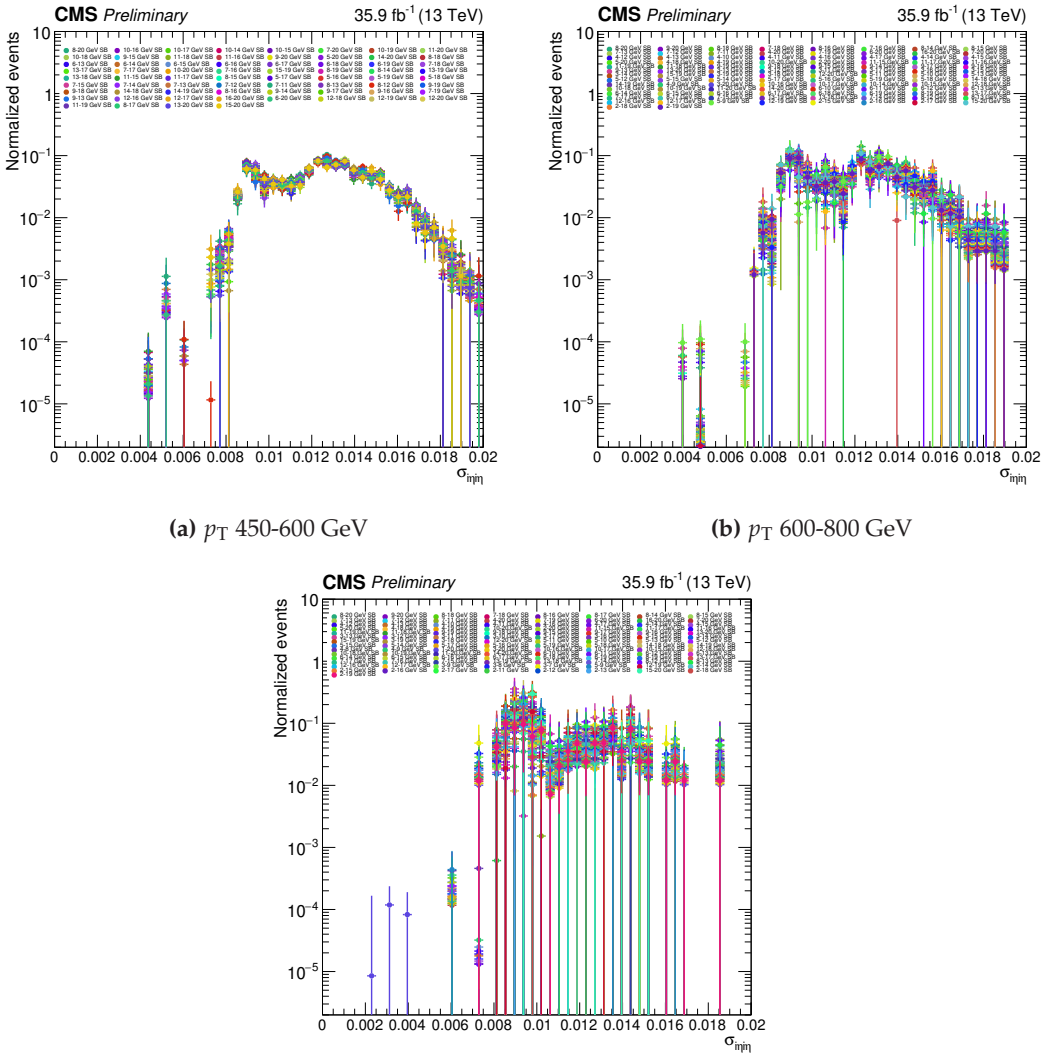
(a) p_{T} 200-250 GeV

(b) p_{T} 250-300 GeV

(c) p_{T} 300-350 GeV

(d) p_{T} 350-450 GeV

Figure A.10.: Distribution of the $\sigma_{i\eta i\eta}$ variable for the different charged hadron isolation sidebands. Different photon transverse momentum bins are shown, ranging from 200 to 450 GeV. The $\sigma_{i\eta i\eta}$ shape is consistent across the many $\mathrm{I}_{\mathrm{ph}}^{\mathrm{chHad}}$ regions considered.



(a) p_{T} 450-600 GeV

(b) p_{T} 600-800 GeV

(c) p_{T} 800-Inf GeV

Figure A.11.: Distribution of the $\sigma_{i\eta i\eta}$ variable for the different charged hadron isolation sidebands. Different photon transverse momentum bins are shown, starting from 450 GeV. The $\sigma_{i\eta i\eta}$ shape is consistent across the many $\mathrm{I}_{\mathrm{ph}}^{\mathrm{chHad}}$ regions considered.

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***, *1* (2012).

[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***, *30* (2012).

[3] Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters", *Astron. Astrophys.* **594** (2016) A13, doi:10.1051/0004-6361/201525830, arXiv:1502.01589.

[4] E. Corbelli and P. Salucci, "The Extended Rotation Curve and the Dark Matter Halo of M33", *Mon. Not. Roy. Astron. Soc.* **311** (2000) 441–447, doi:10.1046/j.1365-8711.2000.03075.x, arXiv:astro-ph/9909252.

[5] V. Rubin, "Spectral displacement of extra galactic nebulae", *Scientific American* **248** (1983) 96–108.

[6] W. J. Stirling and E. Vryonidou, "Electroweak corrections and Bloch-Nordsieck violations in 2-to-2 processes at the LHC", *JHEP* **04** (2013) 155, doi:10.1007/JHEP04(2013)155, arXiv:1212.6537.

[7] A. Denner, S. Dittmaier, T. Kasprzik et al., "Electroweak corrections to W + jet hadroproduction including leptonic W-boson decays", *JHEP* **08** (2009) 075, doi:10.1088/1126-6708/2009/08/075, arXiv:0906.1656.

[8] A. Denner, S. Dittmaier, T. Kasprzik et al., "Electroweak corrections to dilepton + jet production at hadron colliders", *JHEP* **06** (2011) 069, doi:10.1007/JHEP06(2011)069, arXiv:1103.0914.

[9] A. Denner, S. Dittmaier, T. Kasprzik et al., "Electroweak corrections to monojet production at the LHC", *Eur. Phys. J. C* **73** (2013) 2297, doi:10.1140/epjc/s10052-013-2297-x, arXiv:1211.5078.

[10] A. Denner, L. Hofer, A. Scharf et al., "Electroweak corrections to lepton pair production in association with two hard jets at the LHC", *JHEP* **01** (2015) 094, doi:10.1007/JHEP01(2015)094, arXiv:1411.0916.

[11] S. Kallweit, J. M. Lindert, P. Maierhofer et al., "NLO QCD+EW predictions for V+jets including off-shell vector-boson decays and multijet merging", *JHEP* **04** (2016) 021, doi:10.1007/JHEP04(2016)021, arXiv:1511.08692.

[12] CMS Collaboration, "Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **97** (2018), no. 9, 092005, doi:10.1103/PhysRevD.97.092005, arXiv:1712.02345.

[13] U. Baur, "Weak boson emission in hadron collider processes", *Phys. Rev. D* **75** (2007) 013005, doi:10.1103/PhysRevD.75.013005, arXiv:hep-ph/0611241.

[14] J. R. Christiansen and T. Sjöstrand, "Weak gauge boson radiation in parton showers", *JHEP* **04** (2014) 115, doi:10.1007/JHEP04(2014)115, arXiv:1401.5238.

[15] F. Krauss, P. Petrov, M. Schönherr et al., "Measuring collinear W emissions inside jets", *Phys. Rev. D* **89** (2014) 114006, doi:10.1103/PhysRevD.89.114006, arXiv:1403.4788.

[16] ATLAS Collaboration, "Measurement of W boson angular distributions in events with high transverse momentum jets at $\sqrt{s} = 8$ TeV using the ATLAS detector", *Phys. Lett. B* **765** (2017) 132, doi:10.1016/j.physletb.2016.12.005, arXiv:1609.07045.

[17] CMS Collaboration, "Measurement of the differential cross sections for the associated production of a W boson and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **96** (2017) 072005, doi:10.1103/PhysRevD.96.072005, arXiv:1707.05979.

[18] M. Cush, "Standard Model of Elementary Particles". licensed under the Creative Commons Attribution , Wikimedia Commons, 2020, https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[19] E. Fermi, "Tentativo di una Teoria Dei Raggi β", *Il Nuovo Cimento (1924-1942)* **11** (Sep, 2008) 1, doi:10.1007/BF02959820.

[20] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[21] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[22] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C* **680519** (1968) 367–377, doi:10.1142/9789812795915_0034.

[23] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **98** (Aug, 2018) 030001, doi:10.1103/PhysRevD.98.030001. https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[24] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys.Lett. 12 (1964) 321–323* (1964).

[25] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys.Lett. 13 (1964) 132–133* (1964).

[26] C. Guralnik, C. Hagen, and T. Kibble, "Global Conservation Laws and Massless Particles", *Phys.Lett. 12 (1964) 585–587* (1964).

[27] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory". Addison-Wesley, Reading, USA, 1995.

[28] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (Jun, 1963) 531–533, doi:10.1103/PhysRevLett.10.531. https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[29] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.

[30] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge", *Sov. Phys. JETP* **7** (1958) 172–173.

[31] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles", *Prog. Theor. Phys.* **28** (1962) 870–880, doi:10.1143/PTP.28.870.

[32] CMS collaboration, "Summaries of CMS cross-section measurements". https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined.

[33] S. Kallweit, J. M. Lindert, P. Maierhöfer et al., "NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging", *Journal of High Energy Physics* **2016** (Apr, 2016) 1–51, doi:10.1007/jhep04(2016)021. http://dx.doi.org/10.1007/JHEP04(2016)021.

[34] J. M. Lindert et al., "Precise predictions for $V+$ jets dark matter backgrounds", *Eur. Phys. J.* **C77** (2017), no. 12, 829, doi:10.1140/epjc/s10052-017-5389-1, arXiv:1705.04664.

[35] M. Schönherr, S. Kallweit, J. M. Lindert et al., "NLO QCD+EW for V+jets", in *Proceedings, 4th Large Hadron Collider Physics Conference (LHCP 2016): Lund, Sweden, June 13-18, 2016*, volume LHCP2016, p. 58. 2016. arXiv:1609.01445. doi:10.22323/1.276.0058.

[36] F. Buccioni, J.-N. Lang, J. M. Lindert et al., "OpenLoops 2", *Eur. Phys. J. C* **79** (2019) 866, doi:10.1140/epjc/s10052-019-7306-2, arXiv:1907.13071.

[37] R. Frederix, S. Frixione, V. Hirschi et al., "The automation of next-to-leading order

electroweak calculations", *JHEP* **07** (2018) 185, doi:10.1007/JHEP07(2018)185, arXiv:1804.10017.

[38] CMS Collaboration, "Comparison of the Z/γ^* + jets to γ + jets cross sections in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **10** (2015) 128, doi:10.1007/JHEP04(2016)010, arXiv:1505.06520. [Erratum: JHEP04,010(2016)].

[39] J.-L. Caron, "Overall view of LHC experiments. Vue d'ensemble des experiences du LHC.", (May, 1998). http://cds.cern.ch/record/841555.

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

[41] P. Lebrun, "Interim Summary Report on the Analysis of the 19 September 2008 Incident at the LHC". 2008.

[42] CERN, "LEP Design Report". CERN, Geneva, 1984. http://cds.cern.ch/record/102083.

[43] R. D. Ball, V. Bertone, S. Carrazza et al., "Parton distributions from high-precision collider data", *The European Physical Journal C* **77** (Oct, 2017) doi:10.1140/epjc/s10052-017-5199-5.

[44] R. Bruce, C. Bracco, R. De Maria et al., "Reaching record-low β* at the CERN Large Hadron Collider using a novel scheme of collimator settings and optics", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **848** (2017) 19 – 30, doi:https://doi.org/10.1016/j.nima.2016.12.039. http://www.sciencedirect.com/science/article/pii/S0168900216313092.

[45] CMS Collaboration, "CMS Luminosity - Public Result". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[46] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08004.

[47] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[48] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[49] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08002.

[50] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08007.

[51] S. A. Khalek, B. Allongue, F. Anghinolfi et al., "The ALFA Roman Pot detectors of ATLAS", *Journal of Instrumentation* **11** (Nov, 2016) P11013–P11013, doi:10.1088/1748-0221/11/11/p11013.

[52] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08006.

[53] J. L. Pinfold, "The MoEDAL Experiment at the LHC – a New Light on the Terascale Frontier", *Journal of Physics: Conference Series* **631** (2015), no. 1, 012014.

[54] CMS Collaboration, G. L. Bayatian, S. Chatrchyan, G. Hmayakyan et al., "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006. https://cds.cern.ch/record/922757.

[55] CMS Collaboration, "Cutaway diagrams of CMS detector", Technical Report CMS-OUTREACH-2019-001, (May, 2019). https://cds.cern.ch/record/2665537.

[56] S. D. Breeze, "Precision measurement of the Z invisible width with the CMS experiment". PhD thesis, Imperial College London, 2019.

[57] L. Borrello, A. Messineo, E. Focardi et al., "Sensor Design for the CMS Silicon Strip Tracker", Technical Report CMS-NOTE-2003-020, CERN, Geneva, (Aug, 2003). http://cds.cern.ch/record/687861.

[58] CMS Collaboration, "The CMS ECAL performance with examples", Technical Report CMS-CR-2013-430, CERN, Geneva, (Nov, 2013). https://cds.cern.ch/record/1632384.

[59] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[60] J. Mans, J. Anderson, B. Dahmes et al., "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", Technical Report CERN-LHCC-2012-015. CMS-TDR-10, (Sep, 2012). https://cds.cern.ch/record/1481837.

[61] G. Bayatian et al., "Design, performance and calibration of the CMS forward calorimeter wedges", *Eur. Phys. J.* **C53** (2008) 139–166, doi:10.1140/epjc/s10052-007-0459-4.

[62] CMS Collaboration, "Performance of the CMS Muon Detectors in 2016 collision runs", Technical Report CMS-DP-2016-046, (Jul, 2016). http://cds.cern.ch/record/2202964.

[63] CMS Collaboration, G. L. Bayatyan, N. Grigorian, V. G. Khachatrian et al., "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems". Technical Design Report CMS. https://cds.cern.ch/record/706847.

[64] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade",

Technical Report CERN-LHCC-2013-011. CMS-TDR-12, (Jun, 2013). https://cds.cern.ch/record/1556311.

[65] CMS Collaboration, P. Sphicas, "CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger". Technical Design Report CMS.

[66] M. de Gruttola, S. di Guida, D. Futyan et al., "Persistent storage of non-event data in the CMS databases", *JINST* **5** (Jan, 2010) P04003. 20 p, doi:10.1088/1748-0221/5/04/P04003. https://cds.cern.ch/record/1232872.

[67] F. Cavallari, M. de Gruttola, S. Di Guida et al., "Real time monitoring system for applications performing the population of condition databases for CMS non-event data", in *17th IEEE NPSS Real Time Conference: RT 2010*, IEEE Nucl.Sci.Symp.Conf.Rec., p. 5750460. 2010. doi:10.1109/RTC.2010.5750460.

[68] I. Bird, P. Buncic, F. Carminati et al., "Update of the Computing Models of the WLCG and the LHC Experiments", Technical Report CERN-LHCC-2014-014. LCG-TDR-002, (Apr, 2014). http://cds.cern.ch/record/1695401.

[69] CMS Collaboration, "CMS computing: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 2005. http://cds.cern.ch/record/838359.

[70] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (May, 2014) P10009. 80 p, doi:10.1088/1748-0221/9/10/P10009. http://cds.cern.ch/record/1704291.

[71] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME–Journal of Basic Engineering* **82** (1960), no. Series D, 35–45.

[72] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation* **9** (Oct, 2014) P10009–P10009, doi:10.1088/1748-0221/9/10/p10009.

[73] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting", Technical Report CMS-NOTE-2007-008, CERN, Geneva, (Mar, 2007). https://cds.cern.ch/record/1027031.

[74] T. Speer, K. Prokofiev, R. Frühwirth et al., "Vertex Fitting in the CMS Tracker", Technical Report CMS-NOTE-2006-032, CERN, Geneva, (Feb, 2006). https://cds.cern.ch/record/927395.

[75] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[76] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[77] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[78] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *Journal of Instrumentation* **6** (Nov, 2011) P11002–P11002, doi:10.1088/1748-0221/6/11/p11002, arXiv:1107.4277.

[79] A. Sirunyan, A. Tumasyan, W. Adam et al., "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *Journal of Instrumentation* **14** (Jul, 2019) P07004–P07004, doi:10.1088/1748-0221/14/07/p07004.

[80] CMS Collaboration, "CMS Luminosity Measurement for the 2016 Data Taking Period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, (2017). https://cds.cern.ch/record/2257069.

[81] J. Alwall, R. Frederix, S. Frixione et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[82] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5: going beyond", *Journal of High Energy Physics* **2011** (Jun, 2011) doi:10.1007/jhep06(2011)128.

[83] T. Sjöstrand, S. Ask, J. R. Christiansen et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[84] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[85] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[86] S. Alioli, P. Nason, C. Oleari et al., "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111, doi:10.1007/JHEP02(2010)011,10.1088/1126-6708/2009/09/111, arXiv:0907.4076.

[87] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J.* **C71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[88] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[89] GEANT4 Collaboration, "GEANT4: A simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[90] Sherpa Collaboration, "Event Generation with Sherpa 2.2", *SciPost Phys.* **7** (2019) 034, doi:10.21468/SciPostPhys.7.3.034, arXiv:1905.09127.

[91] T. Gleisberg and S. Hoeche, "Comix, a new matrix element generator", *JHEP* **12** (2008) 039, doi:10.1088/1126-6708/2008/12/039, arXiv:0808.3674.

[92] S. Schumann and F. Krauss, "A Parton shower algorithm based on Catani-Seymour dipole factorisation", *JHEP* **03** (2008) 038, doi:10.1088/1126-6708/2008/03/038, arXiv:0709.1027.

[93] S. Hoeche, F. Krauss, M. Schonherr et al., "A critical appraisal of NLO+PS matching methods", *JHEP* **09** (2012) 049, doi:10.1007/JHEP09(2012)049, arXiv:1111.1220.

[94] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.

[95] S. Catani, M. Fontannaz, J. P. Guillet et al., "Cross-section of isolated prompt photons in hadron hadron collisions", *JHEP* **05** (2002) 028, doi:10.1088/1126-6708/2002/05/028, arXiv:hep-ph/0204023.

[96] M. L. Bourhis, J. Fontannaz, and Guillet, "Quark and gluon fragmentation functions into photons", *European Physical Journal C* (1998) 529–537, doi:doi:10.1007/s100529800708.

[97] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through O($\alpha(s)^2$)", *Phys. Rev. D* **74** (2006) 114017, doi:10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.

[98] NNPDF Collaboration, "Parton distributions for the LHC Run 2", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[99] CMS Collaboration, "Muon Identification and Isolation efficiency on full 2016 dataset". https://cds.cern.ch/record/2257968, Mar, 2017.

[100] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Journal of Instrumentation* **10** (Aug, 2015) P08010–P08010, doi:10.1088/1748-0221/10/08/p08010.

[101] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2018) 161, doi:10.1007/JHEP07(2018)161, arXiv:1802.02613.

[102] CMS Collaboration, "Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **10** (2011) 132, doi:10.1007/JHEP10(2011)132, arXiv:1107.4789.

[103] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[104] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[105] CMS Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016". http://cds.cern.ch/record/2622157, Jun, 2018.

[106] CMS Collaboration, "Muon HLT performance on 2016 data". https://cds.cern.ch/record/2297529, Dec, 2017.

[107] CMS Collaboration, "Measurement of the differential Z+jets and γ+jets cross sections, their ratio, and collinear Z boson emission in pp collisions at $\sqrt{s} = 13$ TeV ", Technical Report CMS-PAS-SMP-19-010, CERN, Geneva, (2020).

[108] S. Hageboeck, "What the new RooFit can do for your analysis", *Proceedings of 40th International Conference on High Energy physics — PoS(ICHEP2020)* (Feb, 2021) doi:10.22323/1.390.0910.

[109] G. Schott, "RooStats for Searches", *PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding* (2012) arXiv:1203.1547.

[110] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Computer Physics Communications* **77** (1993), no. 2, 219 – 228, doi:https://doi.org/10.1016/0010-4655(93)90005-W. http://www.sciencedirect.com/science/article/pii/001046559390005W.

[111] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics", *JINST* **7** (2012) T10003, doi:10.1088/1748-0221/7/10/T10003, arXiv:1205.6201.

[112] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method", *Soviet Math. Dokl.* **4** (1963) 1035.

[113] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, (2017). https://cds.cern.ch/record/2257069.

[114] J. Butterworth, S. Carrazza, A. Cooper-Sarkar et al., "PDF4LHC recommendations for LHC Run II", *Journal of Physics G: Nuclear and Particle Physics* **43** (Jan, 2016) 023001, doi:10.1088/0954-3899/43/2/023001.

[115] CMS Collaboration, "Comparison of the Z/γ* + jets to gamma + jets cross sections in pp collisions at $\sqrt{(s)} = 8$ TeV", *Journal of High Energy Physics* **2015** (Oct, 2015) doi:10.1007/jhep10(2015)128.