UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata

Dipartimento di Fisica

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata

Ciclo XXXV

From the Dark Matter to the Dark Sector: search for new physics in final states with a photon and missing transverse momentum with the ATLAS detector

Settore Scientifico Disciplinare FIS/01, FIS/04

Supervisori: Dott.ssa Silvia RESCONI
Prof. Marcello FANTI

Coordinatore: Prof. Matteo PARIS

Tesi di Dottorato di:

Federica Piazza

Anno Accademico 2021/2022

Referees:

Prof. Philip Harris

Dr. Valerio Ippolito

Commission of the final examination:

Prof. Daniele Del Re

Prof. Emidio Gabrielli

Prof. Marie-Hélène Genest

Final examination:

17 April 2023

Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy

Settore scientifico disciplinare:

FIS/04 - FIS/01

PACS:

14.80.-j - 14.80.Cp

Cover illustration:

Artwork by Sandbox Studio, Chicago with Ana Kova: https://www.symmetrymagazine.org/article/voyage-into-the-dark-sector
Event display: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HDBS-2019-13/figaux_07.png
ATLAS art mural at CERN by Josef Kristofoletti

Al nonno Edo

Contents

Introduction

According to the latest measurement of Cosmic Microwave Background (CMB) anisotropies, provided by the Planck and WMAP experiments, the ordinary baryonic matter accounts for only 5% of the Universe mass-energy, while the largest part of the Universe is composed by Dark Energy (68%) and Dark Matter (27%). Despite several astrophysical and cosmological observations have provided compelling evidences of the existence of Dark Matter (DM) as an invisible and gravitationally interacting component of matter, its nature and properties are still largely unknown. The DM problem is therefore one of the main open questions in physics, which solution would give a relevant contribution to our understanding not only in astrophysics and cosmology, but also in particle physics. As a matter of fact, the existence of an unknown component of matter constitutes an important hint of the incompleteness of the Standard Model (SM) of elementary particles. Among the several proposed DM models, one of the most credited and experimentally explored ones is the WIMP (Weakly Interactive Massive Particles) hypothesis, which predicts the existence of non-baryonic and non-relativistic particles, with masses in the range GeV-TeV, weakly interacting with ordinary matter and stable over cosmological scales, while an appealing scenario which is gaining growing interest in the last years, predicts the existence of a full "hidden" or Dark Sector, with its own internal gauge structures and particle content. The DM would be only part of this sector, potentially interacting also with the SM through so-called "portal" interactions.
Assuming that non-gravitational interactions between DM and ordinary matter are possible, multiple approaches to DM detection are fundamental to explore the variety of possible DM scenarios. This thesis focuses on WIMP and Dark Photon searches in proton proton collisions at $\sqrt{s}$=13 TeV at the Large Hadron Collider (LHC), using the full Run-2 data collected by the ATLAS experiment between 2015 and 2018, for a total integrated luminosity of 139 fb^{-1}. Final states involving a photon and missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$), an imbalance in the total transverse momentum of the final state due to undetected particles, are considered, the latter being a potential signature of DM or Dark Sector particles.
The first search targets mono-photon final states, with an high energy photon and $E_{\mathrm{T}}^{\mathrm{miss}}$, and interprets the results in terms of simplified DM models, where weakly interacting DM candidates are pair-produced via an s-channel axial-vector or vector mediator, with a photon from Initial State Radiation. In the presence of a signal, an excess of events in the $E_{\mathrm{T}}^{\mathrm{miss}}$ tails with respect to SM expectations would be observed. The SM background is estimated by normalizing MC simulations to data in appropriate Control Regions (CRs), enriched with a specific background process, for contributions involving true photons, while data-driven techniques are employed for background processes entering the SR due

to electrons or jets mistakenly reconstructed as photons. No excess is observed, therefore exclusion limits in the $m_\chi - M_{med}$ plane are set, with competitive results with respect to Direct Detection experiments.
The second part of this work explores the Dark Sector, by looking for signals of Dark Photon (γ_D) production. The γ_D is predicted as the gauge boson of a new U(1) symmetry group, mediating the interactions in the Dark Sector. An interesting production channel consists in the decay of a (SM or BSM) Higgs boson into a photon and a γ_D, through a loop interactions featuring an additional BSM messenger field, coupled both to the SM and the Dark Sector. A search is performed, for the first time in ATLAS, targeting Higgs bosons produced in association with a Z boson, leptonically decaying into two electrons or two muons, giving rise to a $\ell\ell + E_{\mathrm{T}}^{\mathrm{miss}} + \gamma$ final state. The background estimation is strongly based on data-driven techniques, with a new method developed to estimate the dominant background arising from "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ mainly due to mismeasured jets. No excess is observed, and upper limits on the branching fraction BR($H \to \gamma\gamma_D$) are set. Finally, a reinterpretation of the mono-photon analysis is performed, to provide exclusion limits on the same decay for BSM Higgs with masses between 400 GeV and 3 TeV, in the gluon-gluon fusion and Vector Boson Fusion production modes. These analyses provide competitive results with respect to other LHC searches, and a statistical combination with other ATLAS results is foreseen to gain further improvement.
The searches described in this thesis are part of a huge effort carried on within the ATLAS collaboration, towards the understanding of DM nature. Collider searches have a key role in this field, ensuring good complementariety with Direct Detection and Indirect Detection strategies, which is of fundamental importance to extend the possible reach of research in probing the extremely wide range of possible DM models and masses.

Structure of the thesis This thesis is structured as follows: in Chapters 1 and 2, the general theoretical framework is presented, starting with the presentation of the Standard Model of elementary particles, its successes and its open issues, before moving to an overview of the Dark Matter problem and possible candidates and detection strategies. After describing the Lrge Hadron Collider and ATLAS experiment in Chapter 3, a review of the complex reconstruction and identification of the physics objects in ATLAS is given in Chapter 4. In Chapter 5, the main results of my Qualification Task are presented: it consists in the implementation of an updated $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithm towards a Global Particle Flow strategy, its validation and preliminary performance studies. The remaining part of the thesis, focuses on the Dark Matter related searches which were the main focus of my work. The phenomenology and models related to the analyses described later are summarized, including a brief status of the art, in Chapter 6. In Chapter 7, common methods and tools employed in the search for new physics are described. Chapter 8, 9 and 10 describe in details the three analyses mentioned above, focusing in particular on my personal contributions.

Main personal contributions

- Mono-photon analysis: analysis Framework and N-tuple production; selection criteria optimization; $E_{\mathrm{T}}^{\mathrm{miss}}$ and jet related performance studies; Control regions and γ+jets background studies, probe-e CRs rescaling for $e \to \gamma$ data-driven estimation; experimental systematic uncertainties, theoretical uncertainties on the signal; statistical analysis with HistFitter: set-up of the HF macro, background-only fit,

model-independent and model-dependent exclusion limits, reinterpretations in different simplified models, translation into limits on the DM-nucleon scattering cross-section and comparison with Direct Detection experiments.

- Dark-photon analysis: contribution to the maintenance of the analysis framework and N-tuple production; $E_{\mathrm{T}}^{\mathrm{miss}}$ related performance studies; background estimation and modelling: optimization of ABCD method for the estimation of the dominant background from "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$, contribution to $e \to \gamma$ estimate, definition of $VV\gamma$ CR and top VR; statistical analysis with HistFitter.

- Mono-photon reinterpretation: integration of the new signal in the RECAST workflow; theoretical uncertainties; independent reinterpretation to cross-check the results of the RECAST.

Chapter 1

The Standard Model of elementary particles

The Standard Model (SM) of particle physics is the quantum field theory that describes the elementary particles and their fundamental interactions, except for gravitational one [1]. The model was developed during the XX century, and has been extensively tested for decades, with oustanding success, culminated with the most important recent confirmation of the validity of the model: the discovery of the Higgs boson, responsible for the mass term, in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (CERN) [2, 3, 4].
The Standard Model consists in a non-abelian gauge theory, invariant under the gauge symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$: $SU(3)_C$ is the symmetry group related to colour charge C, describing strong interactions in the Quantum Chromodynamics (QCD) model, while $SU(2)_L \times U(1)_Y$ is the symmetry group associated to the electroweak interactions, with the L sub-script indicating that the weak charged currents interact only with left-handed fermions, and Y being the hypercharge.

1.1 Particles

The fundamental particles of the model are summarized in Figure 1.1. The gauge bosons are spin-1 fields, mediators of the interactions between *quarks* and *leptons*. They arise from the 12 generators of the theory: 8 gluons are associated to the SU(3) group and mediators of strong interactions, W^1, W^2, W^3 are the 3 generators of the SU(2) group and B is the gauge boson associated to the U(1) group. By means of a mixing described in the following section, the generators of the SU(2) and U(1) groups give rise to the physical mediators of the weak interactions, $W^{\pm}$ and Z, and the electromagnetic one, the photon A.
The matter fields are spin 1/2 particles, the fermions, which include three generations of *quarks* and *leptons*. The left-handed (ψ_L) and right-handed (ψ_R) states behave differently under weak interactions, since weak charged currents couple only to ψ_L: left-handed fields are isospin doublets, while right-handed ones are isospin singlets. Considering only the first generation of leptons and quarks, the quantum numbers are summarized in Table 1.1. While leptons are charged only under the electroweak symmetry group, quarks have an additional quantum number, the *colour* charge, thus being subject also to strong interactions. The last constituent of the model is the Higgs boson, a neutral scalar field, responsible for the gauge bosons, leptons and quarks masses, through the Higgs mechanism described in Section 1.3.1.

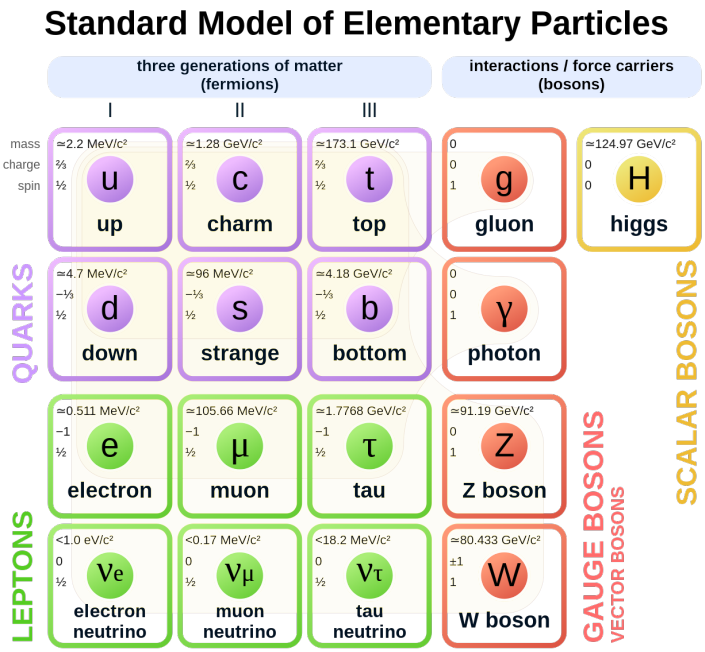


Figure 1.1: Standard Model elementary particles [5]

	T^3	Q	Y
$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$+1/2$ $-1/2$	$+2/3$ $-1/3$	$1/3$
u_R	0	$+2/3$	$4/3$
d_R	0	$-1/3$	$2/3$

	T^3	Q	Y
$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$	$+1/2$ $-1/2$	0 -1	-1
$(\nu_e)_R$	0	0	0
e^-_R	0	-1	-2

Table 1.1: Summary of the Isospin (T^3) electric charge (Q) and hypercharge (Y) for the first generations of leptons and quarks. Similar quantum numbers, but different masses, characterize the other two generations

1.2 Interactions

1.2.1 Quantum Electrodynamics QED

A simple example of a gauge theory is Quantum Electrodynamics (QED) [1] describing electromagnetic interactions mediated by a vector field, A^μ.
As mentioned before, electromagnetic interactions occur between leptons, described by spinor fields ψ with spin 1/2 and mass m. The Lagrangian of QED can be therefore built starting from a Dirac Lagrangian.

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \tag{1.1}$$

This Lagrangian is required to be invariant under local $U(1)$ gauge transformations, i.e.:

$$\psi \to e^{iq\theta(x)}\psi \tag{1.2}$$

where x denotes time and space coordinates. Local invariance can be obtained only by the introduction of an extra vector field, the photon A_μ, with the following transformation:

$$A_\mu \to A_\mu + \partial_\mu \theta(x) \tag{1.3}$$

The derivative in the Dirac Lagrangian can be rewritten as:

$$\partial_\mu \to D_\mu = \partial_\mu - iqA_\mu \tag{1.4}$$

And the locally gauge invariant lagrangian of QED is

$$\mathcal{L}_{QED} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi + (q\bar{\psi}\gamma^\mu \psi)A_\mu \tag{1.5}$$

with $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$.
The fine-structure constant α, quantifying the strength of electromagnetic interactions, is given, at low energy scales, by:

$$\alpha = \frac{e^2}{4\pi\epsilon_0 \hbar c} \simeq 1/137 \tag{1.6}$$

where $e = 1.602176 \times 10^{-19} C$, $\epsilon_0 = 8.854187 \times 10^{-12} \mathrm{Fm}^{-1}$ the vacuum dielectric constant, $\hbar = h/2\pi = 1.054571 \times 10^{-34}$ Js with h the Planck constant and $c = 299792458$ m/s the speed of light in the vacuum.

1.2.2 The electroweak interactions

Electroweak interactions are described by the Glashow-Weinberg-Salam model [6, 7, 8], according to which the weak and electromagnetic interactions are unified at high energy scales ($\sim$ 100 GeV), but manifest themseles in the form of two separate interactions at lower scales as the ones reached at the LHC.
The SM is based on the local gauge invariance of the Lagrangian. The invariance of the electroweak sector of the lagrangian under the $SU(2)_L \times U(1)_Y$ symmetry group is ensured by a redefinition of the derivative in the Dirac lagrangian through the addition of the three gauge bosons W^i of the $SU(2)_L$ group, and the B gauge boson of $U(1)_Y$:

$$\partial_\mu \to D_\mu = \partial_\mu + igT^i W_\mu^i + ig'\frac{1}{2}YB_\mu \tag{1.7}$$

where g and g' are the coupling constants, T^i is the i-th component of the weak isospin and Y is the weak hypercharge quantum number defined by the Gell-Mann-Nishijima relation $Q = T^3 + \frac{1}{2}Y$.
The Dirac Lagrangian therefore assumes the form:

$$\mathcal{L}_{ewk} = i\bar{\psi}\gamma^\mu D_\mu \psi - \frac{1}{4}W^k_{\mu\nu}W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{1.8}$$

where ψ are the fermion spinors and the "reduced fields" can be written as:

$$W^k_{\mu\nu} = \partial_\mu W^k_\nu - \partial_\nu W^k_\mu + g\epsilon^{klm}W_{\mu,l}W_{\nu,m} \tag{1.9}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.10}$$

Therefore, the interaction term of the electroweak Lagrangian is given by:

$$\mathcal{L}^{int}_{ewk} = -\frac{g'}{2} Y(\bar{\psi}\gamma^\mu\psi)B_\mu - g(\bar{\psi}_L\gamma^\mu T_i\psi_L)W^i_\mu \tag{1.11}$$

$$= -\frac{g'}{2}\sum_{i=u,d} Y^i_R(\bar{\psi}^i_R \not{B}\psi^i_R) - \frac{1}{2}(\bar{\psi}^u_L\ \bar{\psi}^d_L)\begin{pmatrix} g'Y_L\not{B} + g\not{W}^3 & g(\not{W}^1 - i\not{W}^2) \\ g(\not{W}^1 + i\not{W}^2) & g'Y_L\not{B} - g\not{W}^3 \end{pmatrix}\begin{pmatrix}\psi^u_L \\ \psi^d_L\end{pmatrix} \tag{1.12}$$

where the *slashed* notation indicates $\not{B} = \gamma_\mu B^\mu$ and $\not{W} = \gamma_\mu W^\mu$, and the coupling of the W gauge fields of the $SU(2)_L$ group to only the *left*–handed component of the spinor fields is made explicit.
The physical fields mediating the weak and the electromagnetic interactions can be obtained as a combination of the above mentioned generator fields, highlighting the fact that these two interactions are actually a manifestation of an unified force. In particular, the charged gauge bosons of the weak interactions arise from a linear combination of W^μ_1 and W^μ_2

$$W^\pm = \frac{1}{\sqrt{2}}(W^\mu_1 \mp W^\mu_2) \tag{1.13}$$

while the neutral boson of the weak interactions (Z^μ) and the photon (A^μ) result from the mixing of W^μ_3 and B^μ:

$$\begin{pmatrix} Z^\mu \\ A^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} W^\mu_3 \\ B^\mu \end{pmatrix} \tag{1.14}$$

with θ_W the mixing Weinberg angle. It can be shown that the electric charge e and the weak neutral charge g_Z can be expressed as a function of the coupling constants g and g':

$$e = g\sin\theta_W = g'\sin\theta_W \tag{1.15}$$

$$g_Z = \frac{e}{2\sin\theta_W\cos\theta_W}(T_3 - 2Q\sin^2\theta_W) \tag{1.16}$$

The interaction term of the electroweak lagrangian can indeed be written in the form:

$$\mathcal{L}^{int}_{ewk} = -eA^\mu J^{em}_\mu - \frac{g}{\sqrt{2}}(W^{+\mu}J^-_\mu + W^{-\mu}J^+_\mu) \tag{1.17}$$

$$+ \frac{e}{2\sin\theta_W\cos\theta_W}\bar{\psi}\not{Z}\left[(T_3 - 2Q\sin^2\theta_W) - T_3\gamma^5\right]\psi \tag{1.18}$$

where the electromagnetic current and the charged weak current are defined as $J^{em}_\mu = \bar{\psi}\gamma_\mu Q\psi$ and $J^\pm_\mu = \bar{\psi}^{u/d}\gamma_\mu\psi^{d/u}$ respectively, while the last term is the weak neutral current.

1.2.3 The strong interactions in QCD

The strong force is described by Quantum Chromodynamics, QCD [9, 10], a Yang-Mills non Abelian field theory [11] based on the $SU(3)_C$ symmetry group, where C indicates the colour charge. The quarks are triplets under colour charge, which can carry *red*, *green* or *blue* charge. The strong interactions between quarks are mediated by the 8 massless

gluons of the theory (G_a^μ with a the colour quantum number), which carry a colour charge, due to the theory being non abelian, allowing QCD gluons self-couplings.
Labeling the quark flavour with f and indicating with q_{fa} the quark field of flavour f and colour charge a, the Lagrangian is

$$\mathcal{L}_{QCD} = -\frac{1}{4} G_a^{\mu\nu} G^a_{\mu\nu} + \sum_f \bar{q}_{fa} i\gamma^\mu D^a_{\mu b} q^b_f \tag{1.19}$$

where a and b run over the colour quantum numbers.
The field strength tensor $G^a_{\mu\nu}$ for the gluon field G^a_μ is given by

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu \tag{1.20}$$

with f_{abc} the structure constants, and the last factor being the gluon self-interaction, with strong coupling g_s.
The covariant derivative is:

$$D^a_{\mu b} = \partial_\mu \delta^a_b + i g_s \lambda^{ia}_b G^i_\mu \tag{1.21}$$

where i is the colour index and λ^i are the $SU(3)$ Gell-Mann matrices. This derivative ensures the required local invariance under the $SU(3)$ transformation:

$$q_a(x) \to e^{i g_s \sum_i \theta_i(a) \lambda^b_a} q_b \tag{1.22}$$

The presence of gluon-gluon couplings modifies the dependency of the coupling constant α_s on the energy scale, resulting from renormalization. The coupling constant can be expressed as a function of the renormalization scale μ_R

$$\alpha_s(\mu_R) = \frac{g_s^2}{4\pi} = \frac{12\pi}{(11N_C - 2N_f)\ln(\frac{\mu_R^2}{\Lambda^2_{QCD}})} \tag{1.23}$$

where, Λ_{QCD} is a characteristic energy scale with experimental value ~ 300 MeV, N_C is the number of colours and N_f the number of flavours. This value gives an estimate of the effective strength of a given process, for $\mu_R^2 \sim Q^2$, with Q the transferred momentum in the process.
While in abelian theories like QED the coupling constant increases for increasing energy scales, in QCD the coupling constant value at low energies increases, while at high energies it becomes smaller, resulting in the so-called *asymptotic freedom*. Due to this phenomenon, perturbative QCD is only appliable at high energy scales.
Another characteristic phenomenon of QCD, observed empirically, is the *colour confinement*: quarks and gluons (the *partons*) cannot be observed as coloured elementary particles, but only in form of composite colourless particles (the *hadrons*). Moreover, when trying to extract a quark from a hadron, due to the increase of potential energy, the production of a new quark-antiquark pair and gluons becomes more energetically convenient, giving rise to the processes of *fragmentation* and subsequent *hadronization*, i.e. creation of colourless hadrons from quark-antiquark pairs and gluons.

1.3 The Higgs boson

Since a mass term for the gauge bosons of the form $M^2 V_\mu V^\mu$ would break the local gauge invariance of the Lagrangian, gauge bosons in the Standard Model are required to be massless. The same requirement is true for fermions, because the mass term

$$m_f \bar{\psi}\psi = m_f(\bar{\psi}\psi_L + \bar{\psi}_L\psi_R) \tag{1.24}$$

includes the terms $\bar{\psi}_R\psi_L$ and $\bar{\psi}_L\psi_R$ which are not gauge invariant since they transform differently under $SU(2)_L$.
On the other side, both weak gauge bosons and fermions have experimentally been observed to have a mass, thus suggesting the incompleteness of the SM as described until now.

1.3.1 The Brout-Englert-Higgs mechanism

A possible mechanism to provide mass terms without breaking the $SU(2) \times U(1)$ local gauge symmetry was proposed in 1964 by Higgs [12, 13, 14], and independently Brout and Englert [15], based on the mechanism of *spontaneous symmetry breaking*. This mechanism applies only to the $SU(2) \times U(1)$ group, because the gluons of $SU(3)$ are massless. This leads to the definition of the Higgs field as an isospin doublet of complex scalar fields:

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.25}$$

Since after the symmetry breaking the $U(1)$ group of QED must be unbroken, due to the fact that the photon is massless, the Higgs electric charge must be null, thus leading to an hypercharge $Y = \pm 1$. Selecting $Y = 1$, we obtain the field:

$$\Phi(x) = \begin{pmatrix} 0 \\ \phi^0(x) \end{pmatrix} \tag{1.26}$$

The Higgs potential is:

$$V(\phi) = \lambda(\Phi^\dagger\Phi)^2 + \mu^2\Phi^\dagger\Phi \tag{1.27}$$

where λ is taken positive, and the full Lagrangian of the Higgs sector is:

$$\mathcal{L}_\phi = (D^\mu\Phi)^\dagger(D^\mu\Phi) - [\lambda(\Phi^\dagger\Phi)^2 + \mu^2\Phi^\dagger\Phi] \tag{1.28}$$

where the covariant derivative is the one of the electroweak theory.
For $\mu^2 > 0$ the potential includes a mass term for the field Φ and has a minimum at $\Phi = 0$. The case of interest is instead $\mu^2 < 0$: in this situation, the potential assumes a continuum of minima, as shown in Figure 1.2, for

$$|\phi_0|^2 = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2} \tag{1.29}$$

This infinite number of minima results in the spontaneous symmetry breaking, when a particular choice of the the minimum (or vacuum) is selected. Choosing the minimum

$$< \Phi >= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.30}$$

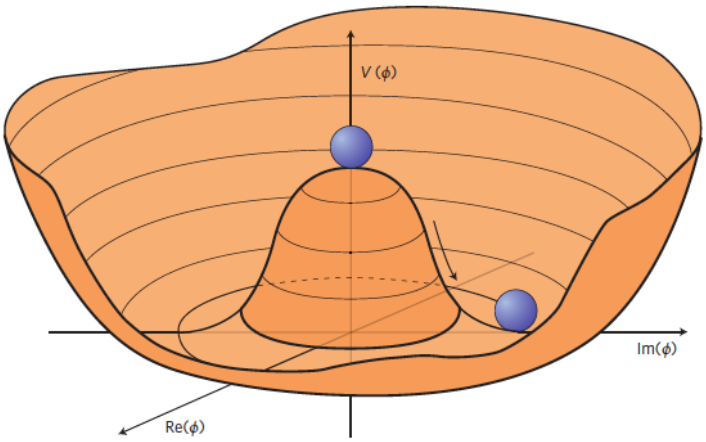


Figure 1.2: Higgs potential

where v is the vacuum expectation value (vev), the particle field can be expressed in terms of perturbation around the minimum, such that:

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} \phi_1 + i\phi_2 \\ v + h + i\eta \end{pmatrix} \tag{1.31}$$

where $v = \sqrt{-\mu^2/(2\lambda)}$, h is the scalar Higgs field and ϕ_1, ϕ_2 and η are the Goldstone bosons arising from the breaking of 3 of the 4 generators of the $SU(2) \times U(1)$ group, which can be set to 0 in the unitary gauge.
Including the covariant derivatives of the electroweak theory in Eq. 1.28, and using the fact that physical gauge bosons are a combination of the electroweak generators, the Higgs sector is described by the lagrangian:

$$\mathcal{L} = \frac{1}{2}\partial_\mu h \partial^\mu h + \mu^2 h^2 - \lambda\nu h^3 - \frac{\lambda}{4}h^4 + \tag{1.32}$$

$$+ \left(\frac{gv}{2}\right)^2 W_\mu^+ W^{-\mu} + \frac{1}{2}\left(\frac{g^2 + g'^2}{4}\right) v^2 Z_\mu Z^\mu + \tag{1.33}$$

$$+ \frac{1}{2}g^2 v W_\mu^+ W^{-\mu} h + \frac{g^2 v}{2}\frac{1}{2\cos\theta_w^2} Z_\mu Z^\mu h + \frac{g^2}{4} W_\mu^+ W^{-\mu} h^2 + \frac{g^2}{8\cos\theta} Z_\mu Z^\mu h^2 \tag{1.34}$$

where the terms in the first row are the kinetic and Higgs self-coupling terms including the mass term leading to the definition of the Higgs mass $m_H = \sqrt{2}\mu = \sqrt{2\lambda}v$. The second row contains the gauge bosons mass terms, and the last one the couplings of the Higgs boson to the gauge bosons.
Therefore, we obtain the following relations for the W and Z masses:

$$m_W = \frac{gv}{2} = 80.4 \text{ GeV} \tag{1.35}$$

$$m_Z = \frac{v\sqrt{g^2 g'^2}}{2} = \frac{gv}{2\cos\theta_W} = \frac{m_W}{\cos\theta_W} = 91.2 \text{ GeV} \tag{1.36}$$

where $v = \sqrt{1/(\sqrt{2}G_F)} \sim 246$ GeV, given the Fermi coupling constant $G_F = 1.1663787 \times 10^{-5}$ GeV^{-2}.
In addition to gauge bosons, the Higgs mechanism is also responsible for the fermion masses. The Higgs doublet allows to build fermionic mass terms that are invariant under $SU(2)_L \times U(1)_Y$, through the Higgs field, for the down component of the isosping doublet, or its charge conjugate for the up component:

$$\Phi_C = -i\sigma_2\Phi^* = \frac{1}{\sqrt{2}}\begin{pmatrix} v+h \\ 0 \end{pmatrix} \tag{1.37}$$

Indicating with $(\nu\ \ell)$ the lepton doublets and with $(u\ d)$ the quark doublets, the mass terms are obtained through couplings of the left-handed doublets and right-handed fermionic singlets to the Φ_C and Φ fields, as described in the Yukawa lagrangian

$$\mathcal{L}_f = -y_u(\bar{u}_L\ \bar{d}_L)\Phi_C u_R - y_d(\bar{u}_L\ \bar{d}_L)\Phi_C d_R - y_\ell(\bar{\nu}_L\ \bar{\ell}_L)\Phi\ell_R + h.c. \tag{1.38}$$

$$= -\frac{1}{\sqrt{2}}\left[y_u(\bar{u}_L\ \bar{d}_L)\begin{pmatrix} v+h \\ 0 \end{pmatrix}u_R + y_d(\bar{u}_L\ \bar{d}_L)\begin{pmatrix} v+h \\ 0 \end{pmatrix}d_R + (\bar{\nu},\bar{\ell})_L\begin{pmatrix} 0 \\ v+h \end{pmatrix}\ell_R\right] + h.c. \tag{1.39}$$

$$= -\frac{v}{\sqrt{2}}\left[y_u\bar{u}u + y_d\bar{d}d + y_\ell\bar{\ell}\ell\right] + -\frac{h}{\sqrt{2}}\left[y_u\bar{u}u + y_d\bar{d}d + y_\ell\bar{\ell}\ell\right] \tag{1.40}$$

where $y_{u,d,\ell}$ are the Yukawa couplings and the first parenthesis contains the mass terms, thus leading to:

$$m_f = \frac{\lambda_f v}{\sqrt{2}} \tag{1.41}$$

1.3.2 Higgs boson phenomenology

The Higgs boson is characterized by a very rich phenomenology in proton-proton collisions, with multiple possible production modes, and several decay channels leading to different possible signature of its production. Despite the Higgs production cross-section at the LHC is much smaller than QCD background and several other SM processes such ad W or Z production, as summarized in Figure 1.3, a relatively high number of Higgs bosons has been collected during Run-1 and Run-2 by the ATLAS and CMS experiments, allowing important precision measurements to be performed.

Higgs boson production modes

A summary of all possible production modes of the Higgs boson is shown in Figure 1.4, where the Feynman diagrams of the different processes are displayed. At the LHC energies of $\sqrt{s} = 13$ TeV, the dominant production process is the gluon-gluon fusion (ggF) (a), followed by Vector Boson Fusion (VBF) (b), associated production with a W or Z boson (ZH, WH), also named Higgstrahlung (c) and gluon-gluon initiated ZH (d). Other processes including quarks t or b in the final states are subdominant. The cross-sections of different production modes are shown in Figure 1.5, as a function of the Higgs mass, with a zoom in the observed Higgs mass in the plot on the right.
Despite the higher cross-section of the ggF process, the VBF mode is often more suitable for analyses, thanks to the clean final state including 2 jets and the Higgs boson. VH processes are also characterized by clean signatures, where the decay products of the Z or W bosons can be used to tag the event. Finally, the $t\bar{t}$H, tH and $b\bar{b}$H modes, although

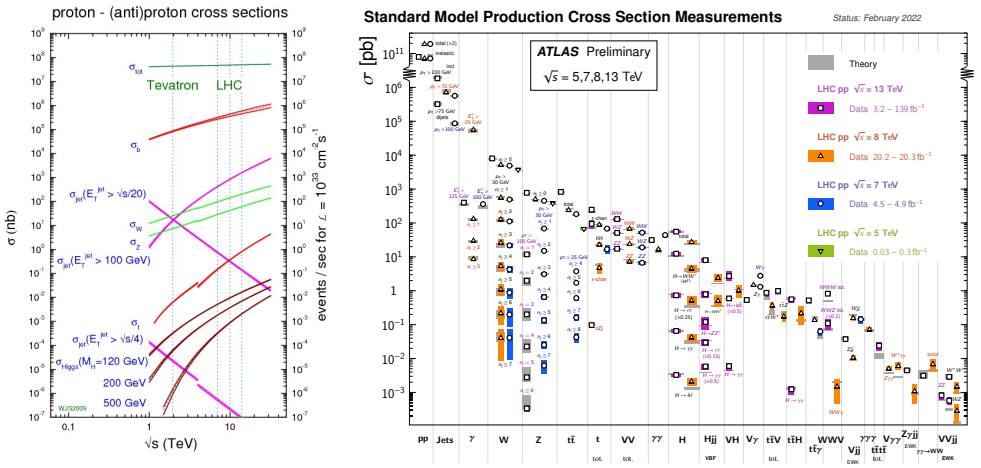


Figure 1.3: The theoric cross-sections for several processes as a function of the $\sqrt{s}$ (left), and the measured cross-sections from ATLAS Run-1 and partial Run-2 dataset [16].

challenging due to the low cross-section, are particularly interesting, as they give direct access to Yukawa couplings.

Higgs boson decay channels

The Higgs boson can decay in many possible channels, with a behaviour that is strongly dependent on the Higgs mass, as shown in Figure 1.6. The Branching Ratio (BR) of the Higgs in a given final state i is defined as

$$BR(H \to X_i) = \frac{\Gamma(H \to X_i)}{\sum_j \Gamma(H \to X_j)} \tag{1.42}$$

where Γ is the decay width, $\Gamma_H = 4.07 \times 10^{-3}(^{+4.0\%}_{-3.9\%})$ GeV [18].
The dominant decay channel, for the measured mass of the Higgs, is $H \to b\bar{b}$. This channel is typically not characterized by high sensitivity though, due to the high background from QCD jets. Its sensitivity can be enhanced in VH or $t\bar{t}H$ production modes, where the additional particles in the final state can help tagging the event. The best channel for Higgs measurements is $H \to ZZ^* \to 4\ell$, thanks to its extremely clean signature. Another important channel is the $H \to \gamma\gamma$ one, despite having a relatively small BR, thanks to its excellent mass resolution, as well as being sensitive to scales far beyond the Higgs boson mass.

1.3.3 Higgs boson discovery and measurements

The Higgs boson discovery was announced in July 2012, by the ATLAS and CMS collaborations. A new particle, compatible with the Higgs boson and with mass of about 125 GeV was observed, using data collected during Run 1, with ~ 5 fb^{-1} luminosity at $\sqrt{s} = 7$ TeV and ~ 6 fb^{-1} luminosity at $\sqrt{s} = 8$ TeV. The different decay channels were combined, with dominant contribution from $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ as shown in figure 1.7, showing the local significance for the background only hypothesis as a function

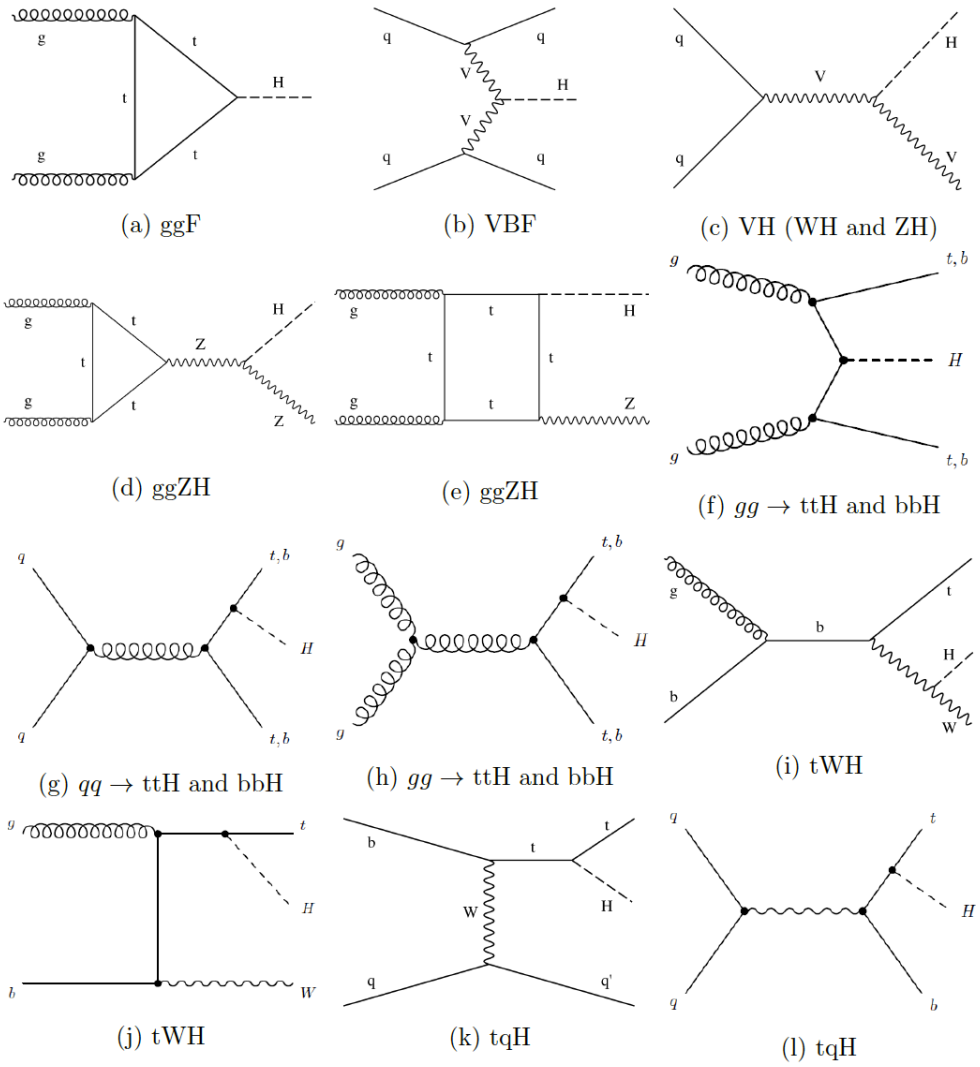


Figure 1.4: List of all possible Leading Order Feynman diagrams for Higgs production

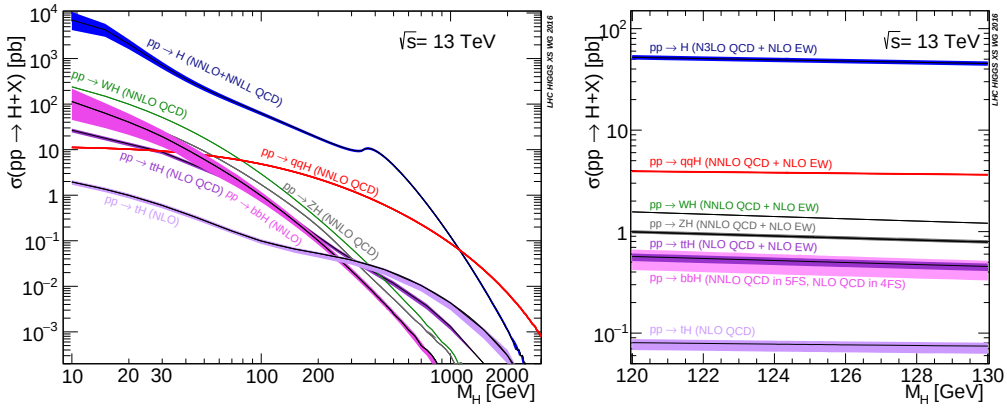


Figure 1.5: Theoretical Higgs production cross-sections for different production modes, as a function of the Higgs mass. The right plot shows a zoom in a range of masses close to the measured one [17].

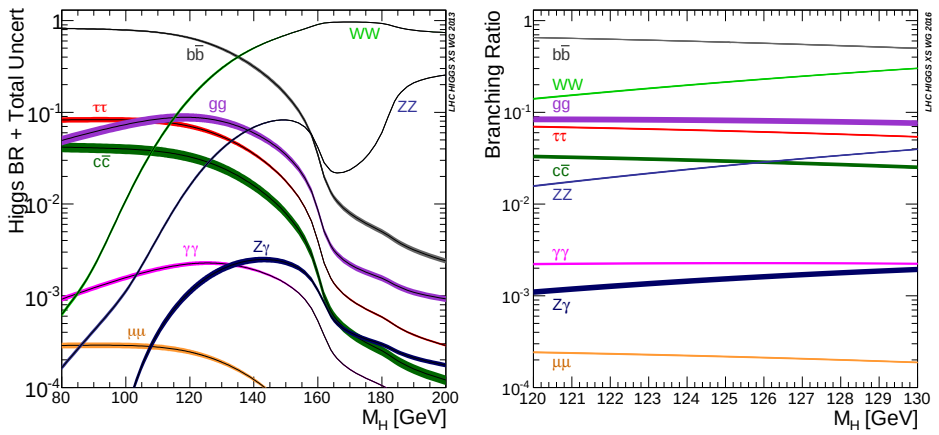


Figure 1.6: Theoretical BR for different Higgs decay channels, as a function of the Higgs mass. The right plot shows a zoom in a range of masses close to the measured one [17].

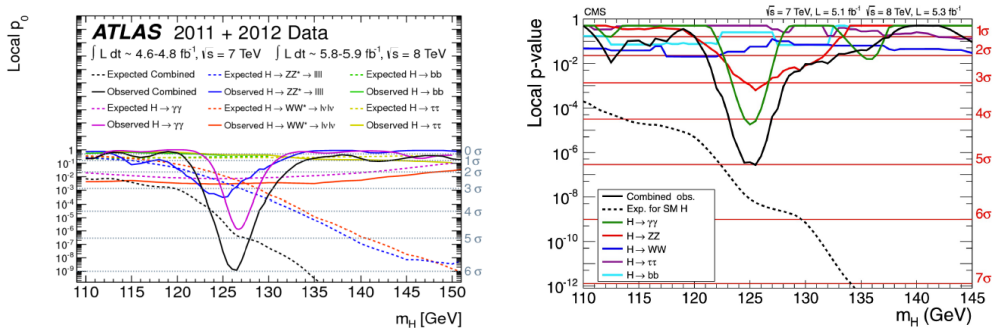


Figure 1.7: The expected and observed local p_0 measured by ATLAS [2] (upper plot) and CMS [3] (lower plot) experiments for the background only hypothesis, for each individual channels present in the combination. The solid black lines represent the combined result. Horizontal lines link p_0 values to the corresponding significance

of the hypothesized Higgs mass. Since the discovery, the focus of Higgs physics has been on precision measurements of the Higgs boson properties, such as its mass and decay rates, but also measurement of the spin, parity and charge conjugation, with particular interest towards possible deviations from SM expectation.
In particular, the Higgs boson mass measurement is of fundamental importance for the SM, being it a free parameter on which several properties of the Higgs boson and its interactions with the other particles of the SM depend. The Run 1 measurement was based on the $H \to \gamma\gamma$ and $H \to 4\ell$ channels. The Higgs boson manifests itself as a peak in the invariant mass of the final state system, thus allowing the measurement of its mass by means of a fit in the invariant mass distribution. The distributions for the two considered channels are shown in figure 1.8: the signal is evident on top of continuous non-resonant background dominated by QCD $\gamma\gamma$ production and non-resonant ZZ respectively. All Run 1 results are summarized in Figure 1.10, showing both ATLAS and CMS measured masses either in the single channels or in the combined one. The final combination of all channels and ATLAS+CMS analyses is also shown, resulting in the final mass value [4]:

$$m_H = 125.09 \pm 0.21(stat) \pm 0.11(syst) \text{ GeV} \tag{1.43}$$

1.4 Open issues of the Standard Model

In the last decades, the Standard Model has been experimentally tested successfully and with excellent accuracy, for instance with the high precision measurements of the Z and W boson masses, which results consistent with Standard Model prediction within 0.3 σ and 1.6 σ respectively (and with most of the main parameters consistent with SM predictions within less than 1σ) [20]. The observation of the predicted Higgs boson, and the absence of any evidence of deviation with respect to expected properties up to now, has been a fundamental step towards the confirmation of the theory. Despite this, some questions remain open, suggesting that some physics Beyond the Standard Model (BSM) might be needed and the model can not be considered complete yet.
The main open issues are the impossibility to describe gravitational interactions within the Standard Model framework, the hierarchy problem, the absence of an explanation for Dark Matter, or of a mechanism to provide masses to neutrinos, and the matter-

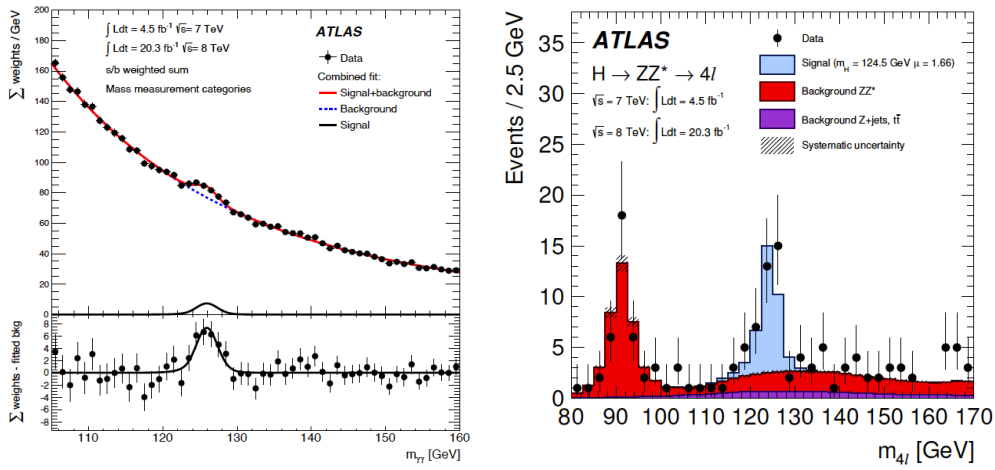


Figure 1.8: Invariant mass spectra for the diphoton (left) and four leptons (right) channels for the full Run 1 dataset collected by the ATLAS experiment, with the result of the simultaneous fit over analysis categories [19].

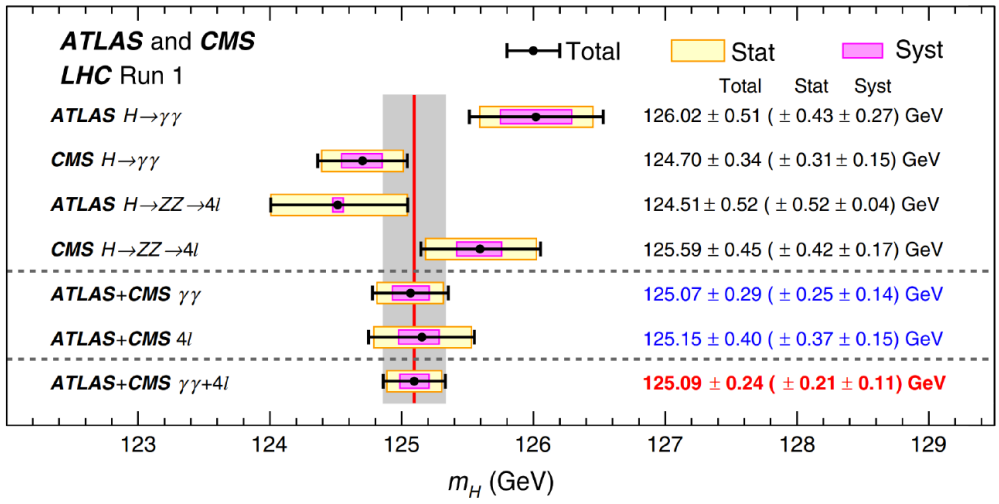


Figure 1.9: Summary of Higgs boson mass measurements from the individual analyses of ATLAS and CMS and from the combined analysis. The systematic (narrower, magenta shaded bands), statistical (wider, yellow-shaded bands), and total (black error bars) uncertainties are indicated. The (red) vertical line and corresponding (gray) shaded column indicate the central value and the total uncertainty of the combined measurement, respectively [4].

antimatter asymmetry. The Dark Matter problem, being the main focus of the searches that will be described in this thesis, will be detailed in Chapter 2. In this section, an overview of the other mentioned issues is presented.

1.4.1 Gravity and hierarchy problem

A way to implement the theory of General Relativity, describing the gravitational force, within the framework of a quantum field theory has not been found yet. For this reason gravity can not be described by the Standard Model. The impossibility to include gravitation in the general description of particle interactions is not critical at the TeV scales, due to gravity being extremely weak with respect to strong and electroweak interactions at the microscopic level. Nevertheless, at higher energies, and in particular at the Planck scale $M_p \sim 10^{19}$ GeV [1], gravitation quantum effects become relevant.
This large scale difference between weak ($M_W \sim 100$ GeV) and Planck scale questions the naturalness of the theory and also results in the so-called *hierarchy problem*, consisting in the need of a fine-tuning of some SM parameters in order to provide the cancellation of quantum effects which would be expected to result in a much higher Higgs boson mass than observed, namely close to the Planck scale.
In particular, quadratically-divergent radiative corrections to the squared Higgs mass are expected from virtual effects of each SM particle (with one-loop diagrams shown in Figure 1.11). The squared mass would be:

$$m_H^2 = m_{H,bare}^2 - O(\lambda, g^2, h^2)\Lambda^2 + O(\ln(\Lambda^2)) \tag{1.44}$$

where Λ is the Planck scale, up to which the SM is assumed to be valid. As an example, the corrections arising from the fermion terms would be of the form

$$\Delta m_H^2 = -\frac{|y_f|^2}{8\pi^2}\Lambda^2 + ... \tag{1.45}$$

In order to obtain the observed mass Higgs of about 125 GeV, a fine-tuned cancellation is needed, of the order of $\left(\frac{m_{H,bare}}{m_P}\right)^2 = \left(\frac{10^2}{10^{19}}\right)^2$.
Among possible solutions to this problem, one is provided by Supersymmetry [21], assuming the existence of a new scalar field for each fermionic one, to cancel the correction with a term of the form $+\frac{y_s}{16\pi^2}\Lambda^2 + ...$ with $y_S = 2|y_f|^2$. The existence of a new $U(1)$ symmetry in an hidden sector, associated to a new gauge boson traditionally named dark-photon, could also help explaining the hierarchy problem. More details about the *dark-photon* will be given in the following chapter.

1.4.2 Neutrino masses

Important proofs of the non-zero masses of neutrinos come from atmospheric and solar neutrino experiments, with a key contribution from Super-Kamiokande (1998) [22] and SNO (2001, 2002) [23, 24], through the observation of oscillations between neutrino flavours, more specifically a discrepancy between the observed ratio of the muonic and electronic atmospheric neutrino fluxes and the expected one (Super-Kamiokande), and the observation of muonic and tauonic neutrinos in the solar neutrino fluxes (SNO).
The explanation of these oscillations requires the existence of different neutrino mass

[1] The scale at which quantum gravitational effects are expected to become relevant and the quantum field theory and general relativity laws are broken

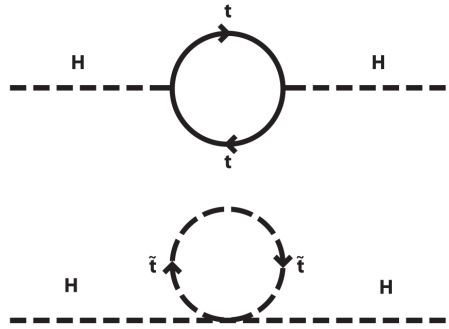


Figure 1.10: Feynman diagrams of one-loop quantum corrections to the Higgs squared mass due to a Dirac fermion (top) or a scalar (bottom) particle [21].

eigenstates, which do not coincide with the flavour eigenstates (e, μ and τ), so that each neutrino flavour is a superposition of mass eigenstates [25]:

$$|\nu_\alpha >= \sum_i U^*_{\alpha i} \nu_i \tag{1.46}$$

where α labels different flavour eigenstates, i labels the mass eigenstates, and $U_{\alpha i}$ is a 3×3 rotation matrix analogous to the CKM matrix of quarks and known as the Pontecorvo–Maki– Nakagawa–Sakata (PMNS) matrix. The PMNS matrix can be expressed in terms of 4 parameters, 3 mixing angles ($\theta_{12}, \theta_{13}, \theta_{23}$) and one CP-violating phase (δ). Two additional phases (α_1 and α_2) arise if neutrinos are Majorana particles (i.e. the particle coincides with the anti-particle), but give no contribution to oscillations. In conclusion, the PMNS matrix is given by:

$$U_{\alpha i} = \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ -s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & sc_{23}c_{13} \end{bmatrix} \begin{bmatrix} e^{i\ \alpha_1/2} & 0 & 0 \\ 0 & e^{i\alpha_2/2} & \\ 0 & 0 & 1 \end{bmatrix} \tag{1.47}$$

The measurement of the PMNS matrix parameters and of the differences between the three mass eigenstates (Δm_{ij}) is based on the measurement of the oscillation probability among different flavour states, $|\langle \nu_\alpha | \nu_\beta \rangle|^2$, which depend on these parameters. A rich experimental program has been carried out and is still ongoing to investigate these properties, with different experiments being sensitive to different subsets of parameters [26], depending on the targeted oscillations and the experimental parameter L/E, with L the distance between the neutrino source and the detector, and E the neutrino energy. In addition, the observation of neutrino-less double β decays could be a signature of Majorana neutrinos, as a Majorana neutrino could be emitted by one nucleus and absorbed by the other one, thus resulting in the absence of neutrinos in the final state. The experimental sensitivity to this process depends on the mass hierarchy, as shown in Figure 1.11. Despite important progresses in the measurements of the relative masses of different neutrino flavours, the measurement of absolute masses has not been possible yet, and an ambiguity remains between normal hierarchy ($m_3 > m_2 > m_1$) and inverted hierarchy ($m_2 > m_1 > m_3$).

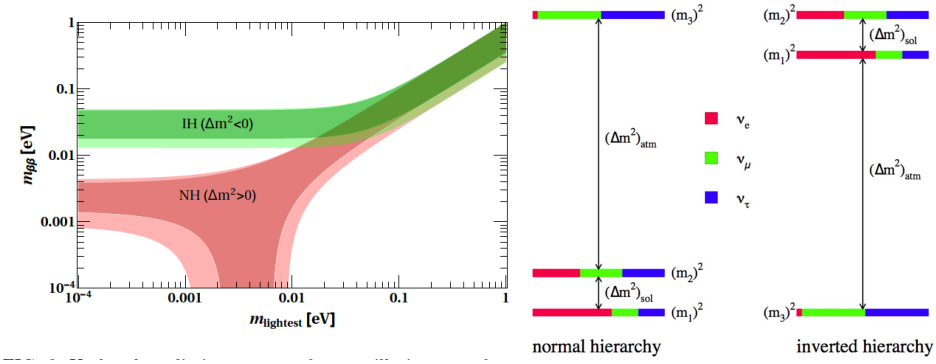


Figure 1.11: Predictions on the effective Majorana mass $m_{\beta\beta}$ from oscillations as a function of the lightest neutrino mass (left [27]) and sketch of mass hierarchies (right [28])

In contrast with the experimental observation of neutrino oscillations, SM neutrinos are massless: the neutrino fields are present in the SM only with *left-handed* (LH) chirality, thus implying the impossibility to build a Dirac mass term in the Lagrangian.
The mechanism through which neutrinos acquire mass is therefore unknown yet. Some physics beyond the SM is needed in order to explain neutrino masses, and several possibles models have been proposed, including Grand Unification Theories or other extensions of the SM predicting the existence of *right-handed* RH neutrinos, as well as models considering neutrinos as Majorana particles (in which case the lepton number conservation or the $U(1)_Y$ symmetry would be violated) [27]. The hypothetical additional RH neutrinos predicted in some theories would mix with the ordinary neutrinos via the Dirac mass term [29]. Moreover, being not subject to SM interactions, they can acquire a Majorana mass without breaking the SM gauge symmetry, resulting in an extension of the mass term of the Lagrangian by [29]:

$$\mathcal{L}_\nu \supset \frac{i}{2}\overline{\nu}_R \not{\partial} \nu_R - Y_\nu \overline{\ell}_L i\sigma_2 H^* \nu_R - \frac{1}{2}\nu_R^T C M_R \nu_R + h.c. \tag{1.48}$$

where ν_R are RH neutrinos, which are singlets under SM gauge symmetries, Y_ν are the Yukava couplings, ℓ_L represents SM lepton doublet, $\sigma_2 H^*$ the Higgs doublet , and M_R a symmetric matrix of L-violating Majorana masses, with L the lepton number. Several models proposed to explain neutrino masses by predicting the existence of additional neutrinos, are collectively known as *see-saw* models. As an example, in the Type I *see-saw* model, RH neutrinos are assumed to be heavy, with a mass much higher than the scale of electroweak symmetry breaking, and the SM neutrinos can acquire a mass term by integrating heavy neutrinos away, and the SM neutrino mass matrix is given by:

$$m_\nu = m_D M_R^{-1} m_D^T \tag{1.49}$$

where $(m_D)_{ij} = Y_\nu^{ij} v/\sqrt{2}$ is the $3 \times m$ Dirac mass matrix with m the number of additional RH neutrinos and v the SM Higgs vacuum expectation value. Other realizations of the Seesaw mechanism predict only additional Dirac RH neutrinos, with Yukawa couplings constrained to be $< 10^{-12}$, the neutrino masses < 0.1 eV and Majorana mass term forbidden [29].

1.4.3 Matter-antimatter asymmetry

Within the framework of the SM the matter-antimatter asymmetry problem [30] is generally considered to be related to the baryon asymmetry problem, i.e. the imbalance of baryonic matter and antibaryonic matter in the observable universe. Starting from the generally accepted absumption that matter and anti-matter are predicted to be produced in equal amount ad the origin of the Universe, the observed imbalance must be originated from some baryon number violating physical process, during the Universe expansion and cooling.
Sakharov proposed three necessary conditions [31], to produce baryons and antibaryons at different rates within the SM: violation of baryon number; violation of both charge conjugation symmetry, C, and charge conjugation-parity symmetry, CP; and the process must not be in thermal equilibrium. The research in this field is therefore focused on probing these three conditions.
Up to now, no evidence of baryon number "direct" violation has been observed. Concerning CP violation, this phenomenon is indeed observed in weak interactions, where it originates from charge-changing (CC) weak interactions that change the charge and flavor of quarks. These CC interactions introduce a mixing between up-like and down-like quarks, with transition amplitudes described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix elements. Nevertheless, this effect has been found to be too tiny to explain the observed asymmetry. Finally, departure from thermal equilibrium is generally assumed to also occur within the electroweak sector of the SM during the so-called electroweak phase transition, a first order transition between massless W and Z gauge bosons state to the massive one.
In conclusion, despite some physical processes satisfying the Sakharow conditions do exist in the SM, they are not enough to explain matter-antimatter asymmetry, thus leading to the need for new physics. The search for CP violating processes gains particular interest at LHC, where this hypothesis is probed, for instance, in the Higgs sector.

Chapter 2

The Dark Matter problem

In the last decades, a variety of unrelated astrophysical measurements have provided compelling evidence of the need for Dark Matter (DM) in order to explain the observation of missing mass at various scales [32, 33].

Velocity dispersion in Coma Cluster The first hint of an additional "dark" contribution to mass, dates back to 1933 when Zwicky observed an anomaly in the velocity dispersions of galaxies in the Coma Cluster [34]. By inferring the mass of the Coma Cluster from luminosity measurements and applying the virial theorem, he estimated an expected velocity dispersion of about 80 km/s, as opposed with the measured one of about 1000 km/s, too high to be explained by visible matter alone.

Rotational curves of galaxies Since 1970s, crucial observations supporting the DM hypotesis have taken place, related to the rotation curves of spiral galaxies, i.e. the circular velocity profile of the stars and gas in galaxies as a function of their distance from the galactic center. According to the Newtonian theory, the orbit velocity of stars and gas in the galaxies would be expected to decrease with increasing distance from the galactic center, following the equation

$$v(r) = \sqrt{\frac{G_N M(r)}{r}} \tag{2.1}$$

where G_N is the gravitational constant and $M(r)$ the mass of the galaxy within the distance r from the centre. Nevertheless, the observed rotational curves are actually observed to be flat up to large radii outside the disk, suggesting the presence of an halo of DM, able to speed up the orbit, with a density going as

$$\rho_{DM}(r) \sim 1/r^2 \tag{2.2}$$

As an example, in Figure 2.1 the velocity profile of galaxy NGC 6503 is shown, as a function of radial distance from the galactic center. The line fitting data is obtained by adding a DM component, while it's clear that the data can't be explained by baryonic matter (gas and stars in the disk) alone.
Since the first observation of flat rotation curve for the Andromeda galaxy by Rubin and Ford in 1970s' [35], similar results have been obtained for almost all the studied galaxies.

Gravitational lensing Another hint of the existence of DM comes from gravitational lensing. This phenomenon is predicted in the context of general relativity and consists

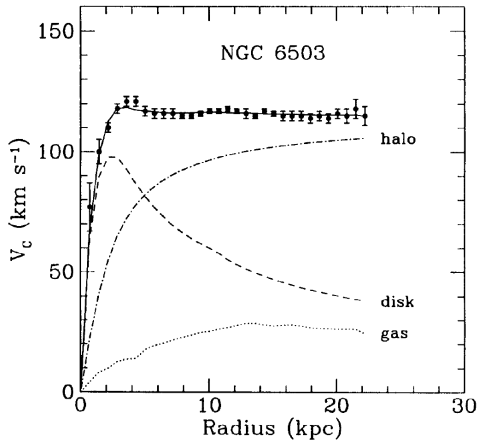
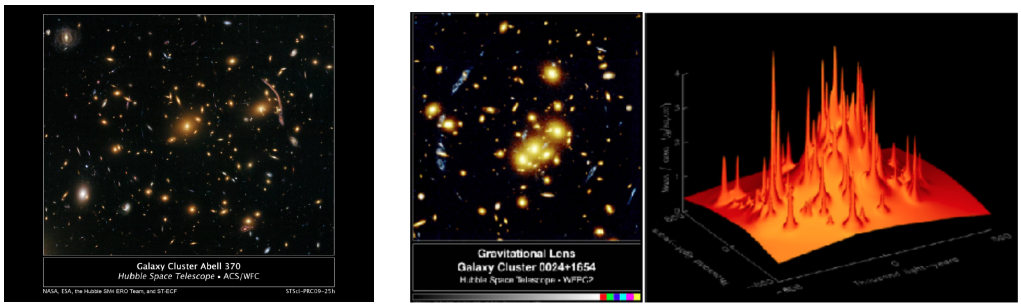


Figure 2.1: Galactic rotation curve for NGC 6503 showing disk and gas contribution plus the dark matter halo contribution needed to match the data.

in the deviation of the light due to space-time deformation produced by the presence of a mass distribution. The mass of the object producing the gravitatonal lensing effect, named "lens", can be inferred through the analysis of the lensed object.
Strong lensing effects happen when the lens mass is high enough to split the lensed astronomical source into multiple images or to distort it in giant arcs or rings. In these cases, it is possible to measure the mass of the foreground object through the observation of its effect on a single distant source. For instance, the observation of giant arcs around the Abell 370 cluster, Figure 2.2a, provided an observed mass about 300 times larger than the one derived from luminosity measurements. In Figure 2.2b [36] another example of strong lensing is shown, in which the background blue galaxy is split into multiple images. In the right panel of Figure 2.2b, a simulation of the lens is displayed: the peaks represent the galaxies in the cluster that constitute the lens, while the smooth component can be related to the DM in the cluster itself. An alternative approach consists of a statistical



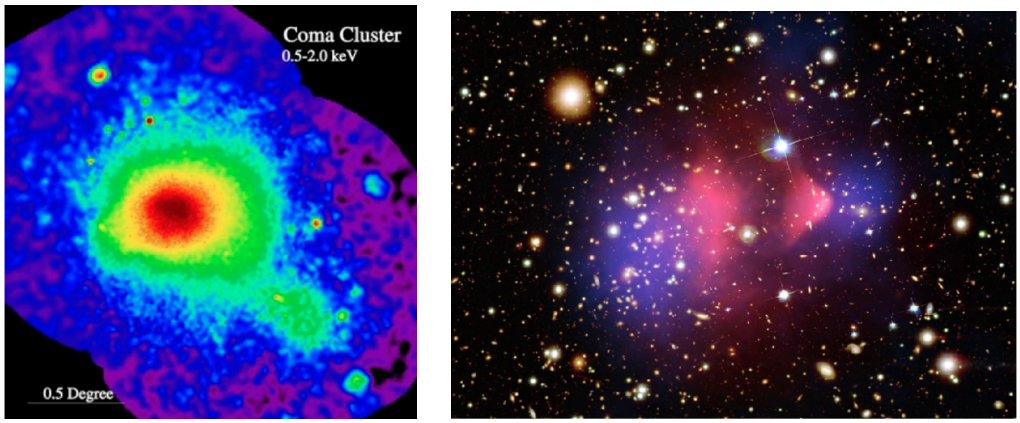
(a) **(b)**

Figure 2.2: (a) Image of the massive galaxy cluster Abell 370 taken by the NASA/ESA Hubble Space Telescope, where giant arcs due to gravitational lensing can be clearly seen. (b) On the left: The foreground cluster of galaxies gravitationally lenses the blue background galaxy into multiple images. On the right: A computer reconstruction of the lens shows a smooth background component not accounted for by the mass of the luminous objects [36].

analysis (performed, for instance, by the Sloan Digital Sky Survey [37]) of *weak lensing* phenomena, resulting in a magnification and shear of multiple distant sources.
The gravitational lensing allows therefore to point out the presence of non-luminous matter, by analyzing the distortion of the images of distant objects, and makes it possible to indirectly observe DM at distances up to 200kPc [1] from the center of the galaxies. Moreover, gravitational lensing observations show the presence of DM distributed in filaments (cosmic web) on Mpc scales.

Hot gas in clusters X-ray emissions observed by ROSAT in the Coma Cluster [38](Figure 2.3a), have shown the presence of hot gas which could only be explained by assuming the presence of a large DM fraction, able to avoid the gas evaporation.

Bullet cluster A further evidence of DM existence is given by combined studies of X-ray emissions and lensing effects, allowing to show the DM interplay with the stellar and gas component of a bullet cluster (cluster originated by the collision of two smaller ones). The X-ray image (ordinary matter) has been collected by the Chandra observatory and is shown in pink in Figure 2.3b. The DM (inferred by lensing measurements) is instead represented in blue. The clearly inhomogeneous distribution of ordinary matter and DM suggests a different behaviour between the two components during collision: while the former has passed through the collision point without interacting, the gas and stars have been slowed due to collisions, and coalesced in the center of the merged cluster.



(a) **(b)**

Figure 2.3: (a) X-ray image of COMA cluster from ROSAT satellite [39]. (b) A collision of galactic clusters shows baryonic matter (pink) as separate from dark matter (blue), whose distribution is deduced from gravitational lensing [40].

Cosmological evidences In addition to the quoted astrophysical measurements, cosmological observations also support the DM hypothesis.
High-precision cosmological simulations underlined that the large-scale structures of the Universe can hardly be explained if ordinary matter alone is taken into account, while a DM component would allow to clump together ordinary matter before recombination,

[1] 1Pc$\sim$ 3.3 light-years$\sim 3.09 \times 10^{16}$m

and would provide the potential wells for structure formation at a later time.
But one of the most important observations is related to the CMB anisotropy measurements (performed by Planck [41] and WMAP [42]), which, in association with the baryonic abundance derived from studies of primordial nucleosynthesis, allows to establish the Universe composition.
The CMB is the remnant radiation from early Universe at $T = 2.73$ K. The observed temperature anisotropy spectrum is the result of photon temperature oscillations, frozen in at redshift x=1100, when the photons decoupled from the baryonic matter.
In Figure 2.4 the temperature fluctuations are shown as a function of angular scales. The anisotropy spectrum is expanded in series of spherical harmonics

$$\delta T(\theta,\phi) = \sum_{\ell=2}^{+\infty} \sum_{m=-\ell}^{+\ell} a_{\ell m} Y_{\ell m}(\theta,\phi) \tag{2.3}$$

and the quantity

$$D_\ell = \frac{\ell(\ell+1)C_\ell}{2\pi} \propto \sum_{m=-\ell}^{+\ell} |a_{\ell m}|^2 \tag{2.4}$$

is plotted, where $C_\ell = \langle |a_{\ell m}|^2 \rangle$ is the angular power spectrum. The first peak is compatible with a flat universe, the second one suggests a 5% abundance of ordinary matter (coherent with predictions from primordial nucleosynthesis), and the third one can be explained by a 27% contribution from DM.

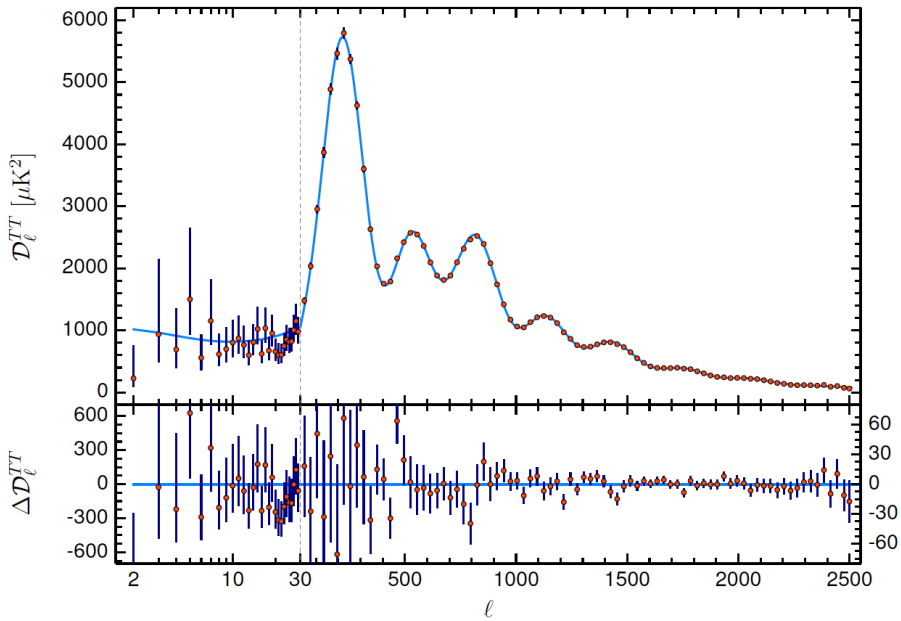


Figure 2.4: Planck's power spectrum of temperature fluctuations in the CMB. The fluctuations are shown at different angular scales on the sky. Red dots with error bars are the Planck data. The blue curve represents the standard model of cosmology, ΛCDM. The peak at 1 degree is consistent with a flat geometry of the universe, the height of the second peak with 5% ordinary matter, and the second and third peaks with 27% DM [41].

2.1 Dark Matter candidates

Based on little information that can be deduced by astrophysical and cosmological observations, a huge number of DM hypotheses and scenarios is possible, predicting DM candidates with masses that range from 10^{30} eV to 10^{30} kg, as summarized in Figure 2.5.

Among this plethora of possibilities, the most credited and experimentally explored

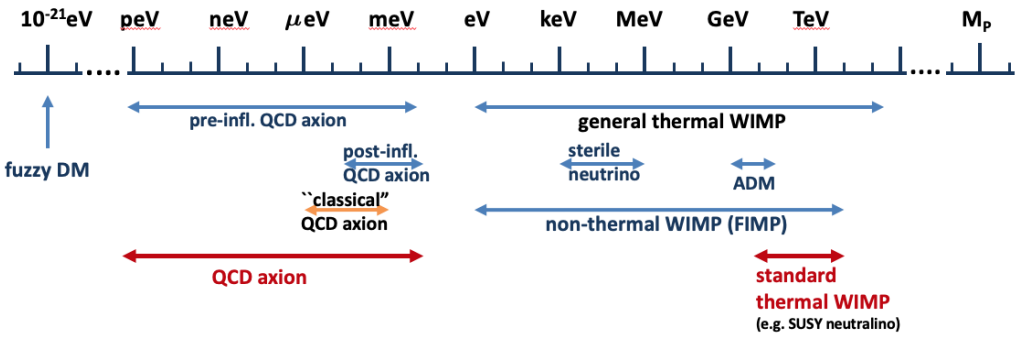


Figure 2.5: A summary of possible DM candidates and their typical predicted masses [43].

candidates are non-baryonic and cold DM particles such as WIMPs (Weakly Interacting Massive Particles). However, in the following part of this chapter, also other hypoteses will be briefly described.

2.1.1 Baryonic Dark Matter

One of the first hypotheses that have been considered, interprets the observed missing mass in terms of a baryonic DM, constituted by Massive Compact Halo Objects (MACHOs) such as faint stars, substellar objects, stellar remnants like brown dwarf or neutron stars, primordial black holes. At present, a baryonic nature of DM has been widely ruled out by strong constraints from several experiments, such as the microlensing experiments MACHO [44] and EROS [45, 46], which pointed out that MACHOs could account for no more than 8% of the missing mass in our galaxy. An overview of the limits set over the percent contribution of MACHOs to DM halo, as a function of the DM mass, is shown in Figure 2.6 . Primordial Black Holes [47], i.e. black holes that formed before the epoch of Big Bang nucleosynthesis and with masses below the sensitivity range of microlensing surveys, were recently explored as a still viable DM candidate, but the latest measurements provided by the Subaru Hyper Suprime-Cam (HSC) [48] strongly constrain this hypothesis. Moreover, a rather strong confutation of the baryonic hypotesis comes from the already mentioned predictions from primordial nucleosynthesis, leading to the 5% estimate of the baryonic abundance.

2.1.2 Non-baryonic Dark Matter

Since the possibility of a baryonic DM is largely excluded, nowadays the most supported theories interpret DM as composed of new, non Standard Model, particles. Given the crucial role of DM in structure formation, these particles are expected to be a relic from the Big Bang and therefore are requested to be neutral, very weakly interacting with ordinary matter and stable over cosmological scales in order to survive with the observed abundance up to the present. Moreover, in order to be coherent with an Universe evolution

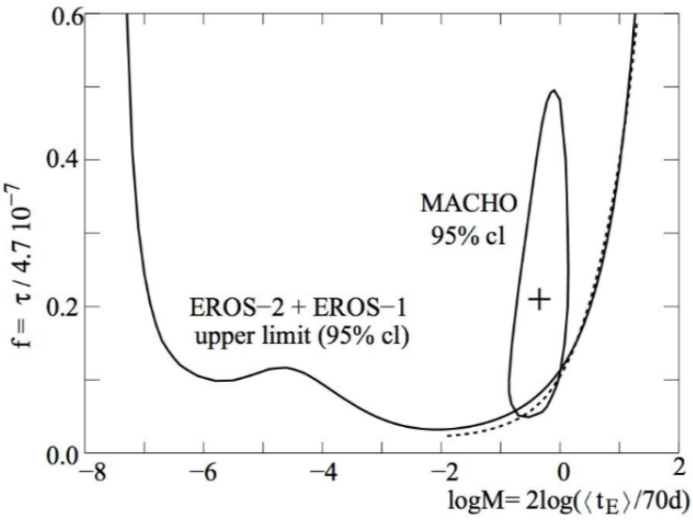


Figure 2.6: Excluded fraction at 95% CL of DM halo constituted by machos, f, as a function of the macho mass, M, for the combined analysis of the EROS surveys. The likelihood contour at 95% CL observed by MACHO is also reported [46].

from small-scale to larger-scale structures, they should be non-relativistic (Cold DM) at *freeze-out*, i.e. the time when the annihilation of DM particles is inhibited by the Universe expansion.
On the other hand, Hot DM candidates (such as ordinary or sterile neutrinos which will be described later), consists in particles that are relativistic at *freeze-out* time, thus favouring an evolution of the Universe from large scale structures to smaller scale ones.

WIMPs Among the non baryonic cold DM candidates, one of the most credited are the WIMPs [49], i.e. weakly interacting massive particles, neutral, stable, and produced as the thermal relic of a *freeze-out* process.
The WIMPs satisfy all the requirements for a possible DM candidate, and are automatically predicted in the context of various models proposed to solve open questions not directly related to DM. For instance, possible WIMP particles could be the lightest neutralino in the SUSY models, or the lightest Kaluza-Klein particle in Extra Dimension theories [50].
Moreover, the search for WIMP particles has been strongly encouraged by the so-called "WIMP miracle", consisting in the observation that this kind of particles, with a mass in the GeV-TeV range, can be naturally produced with a relic density consistent with the observed DM abundance. In the early universe, WIMP are considered to be in thermal equilibrium with baryonic matter, meaning that the reaction

$$\chi\chi \leftrightarrow XX \tag{2.5}$$

(where χ is the DM and X the ordinary matter) is balanced and the DM annihilation rate equals its production rate. With the expansion and consequent cooling down of the Universe, at $T < m_\chi$, the thermal energy becomes insufficient for DM production, thus suppressing this reaction. DM annihilation still happens, resulting in a decrease of DM

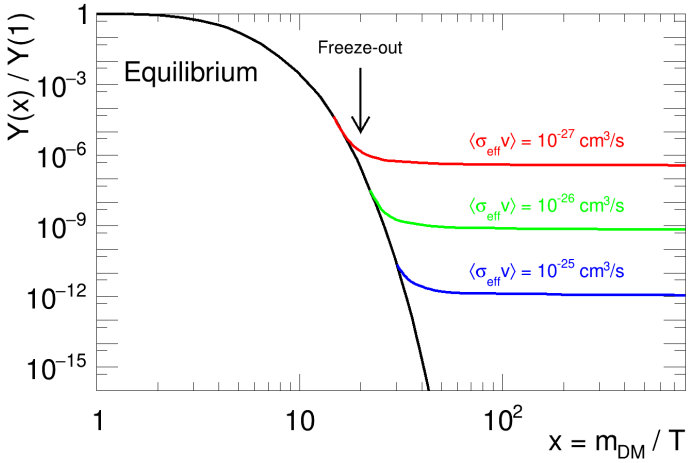


Figure 2.7: Dark matter abundance Y normalized to its value at $T = m_\chi$, as a function of $x = m_\chi/T$, for different choices of the averaged annihilation cross sections. The black line corresponds to thermal equilibrium and the coloured lines to the values of Y after freeze-out. [51].

density, until *freeze-out*. From this moment on, the DM density remains constant up to the present, as shown in Figure 2.7.

More quantitatively, assuming kinetic equilibrium with SM particles (i.e. $\chi X \leftrightarrow \chi X$) and thermal equilibrium with photons (the dominant component in the radiation-dominated early Universe, for red-shift $z > 3000$), the density distribution in time is described by the Boltzmann equation:

$$\frac{dn}{dt} + 3Hn = -\langle\sigma v\rangle(n^2 - n_{eq}^2) \tag{2.6}$$

where H is the Hubble constant, n the DM density at time t, $\langle\sigma v\rangle$ the thermally averaged annihilation cross section and n_{eq} the DM density at thermal equilibrium, following the Boltzmann distribution:

$$n_{eq}(T) = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T} \tag{2.7}$$

In Eq 2.6 the term $3HN$ is related to the decrease in density due to universe expansion, while the term on the right accounts for equilibrium, "forcing" n to approach the equilibrium value. This equation remains valid until *freeze-out*, which happens when the condition:

$$H = n\langle\sigma v\rangle \tag{2.8}$$

is satisfied. Numerical solutions of the Boltzman equation lead to the estimated relic density of the DM:

$$\Omega_\chi h^2 \simeq \frac{3\times10^{-27}\ \mathrm{cm^2 s^{-1}}}{\langle\sigma v\rangle} \simeq 0.1\left(\frac{0.01}{\alpha}\right)^2\left(\frac{m}{100\ \mathrm{GeV}}\right) \tag{2.9}$$

with $h = H_0/100$ km s^{-1} Mpc^{-1} (H_0 the present Hubble constant) and α the coupling strength between DM and ordinary matter. A coupling strength $\alpha \simeq 0.01$ (of the order of

SM weak interaction) would reproduce the observed DM abundance for $m_\chi \simeq 100$ GeV. Although this mass and coupling have actually already been disfavoured by observations, weaker interactions or different WIMP masses are still not excluded and compatible with the observed relic density.

Non-WIMP Cold Dark Matter Other cold DM candidates are WIMP-like particles with similar properties as WIMPs, but a different production mechanism.
For instance, the DM relic density could be explained by a *freeze-in* process, consisting in the production of DM particles through SM particles annihilation or decay, starting from a negligible DM abundance in the early Universe. This production continues until the expansion rate of the Universe becomes high enough and the temperature low enough to suppress the production mechanisms, resulting in a fixed number of DM particles.
Another possibility consists in Asymmetric DM [52], a model that interprets the present DM abundance as the result of an asymmetry in the early universe, similar to the baryonic one:

$$\eta_\chi = \frac{n_\chi - n_{\bar{\chi}}}{n_\gamma} \tag{2.10}$$

where $n_{\chi(\bar{\chi})}$ and n_γ are the number of particle (antiparticle) DM and photons respectively. This interpretation would also give a possible explanation to the observed ratio between baryonic and dark matter abundances ($\Omega_\chi/\Omega_B \sim 5$) as due to a common production mechanism.

Axions The last cold DM candidates described in this incomplete list, are axions, i.e. neutral, pseudo-scalar and stable particles that were originally theorized as the (pseudo-) Goldstone bosons of an anomalous $U(1)$ broken symmetry introduced in order to explain the strong CP problem, i.e. the non observation of CP violation in strong interactions, despite the fact that the conservation of CP symmetry is not theoretically expected in QCD.
Strong constraints rule out the original formulation, the so-called PQWW axion (Peccei-Quinn-Weinberg-Wilzcek) solution [53], which relates the scale of the interactions to the scale of electroweak symmetry breaking.
Despite this, other possible models are still viable solutions to the strong CP problem. For instance the DFSZ (Dine- Fischler-Srednicki-Zhitnitsky) [54, 55] and the KSVZ (Kim-Shifman-Vainshtein-Zakharov) [56] models, related to invisible axions (axions with very weak couplings to ordinary matter) or new formulations of Peccei-Quinn solution theorizing lighter axions.
Axion-like particles (ALPs) are also predicted in the context of other extensions (non related to the strong CP problem) of the SM, which introduce spontaneously broken global $U(1)$ symmetries.

Neutrinos Concerning the Hot DM hypothesis, ordinary neutrinos would be the main candidate, but this hypothesis is disfavoured both because of its inconsistency with the large scale structures of the Universe, and because of the the low masses of these particles that would hardly give a sufficient contribution to the measured DM abundance. On the other side, a viable warm DM candidate would be a sterile neutrino [57], predicted as additional right-handed neutrino, potentially coupling with ordinary neutrinos through kinetic mixing.

2.1.3 MOND

Although the existence of DM is a generally accepted hypothesis by the largest part of the physics community, for completeness the MOND theory (Modified Gravity) has to be mentioned as an alternative solution to the DM problem.
These theories are based on the assumption that no additional matter component exists and that the apparent "missing mass" can be explained by means of a modification of the gravity laws.
More precisely, according to the originally proposed MOND hypothesis [58], for very low accelerations ($a \ll a_0 \sim 1.2 \times 10^{-10}\,\mathrm{ms}^{-2}$), the Newtonian force $F = ma$ has to be scaled:

$$F = \begin{cases} ma & a > a_0 \\ ma^2/a_0 & a \sim a_0 \end{cases} \tag{2.11}$$

In the past years, more complex modified gravity theories have been developed and the leading one, at present, is the TeVeS theory (Tensor-Vector-Scalar gravity) [59], proposed by Bekenstein in 2004 and predicting two additional fields beyond those of general relativity.
The Modified Gravity allows to explain the gravitational lensing observations and anomalies in rotational curves of the galaxies. Moreover, it can reproduce the large scale structures of the Universe without the need for DM.
Despite its ability to explain these observations with an appropriate choice of the parameters, it fails to reproduce the third peak in CMB anisotropies, thus being strongly disfavoured by cosmological evidences. Moreover, the recent observation of two galaxies almost devoid of DM (NGC 1052-DF2 and NGC 1052-DF4) [60, 61], seems to potentially further rule out this hypothesis, being these observations considered consistent with tidal disruption phenomena due to interaction with a neighbour galaxy [62], while it might be difficult to understand such loss in generality of a physics law in modified gravity models. Some recent publications, nevertheless, claim the possibility to explain these two galaxies also within a modified gravity framework [63].

2.2 The Dark Sector scenario

An hypothesis which is gaining growing interest in the last years, also due to increasingly wide regions of the phase-space being excluded by DM searches, predicts the existence of a full "hidden" or Dark Sector, with its own internal gauge structure and particles content [64, 65, 66]. The DM would be only part of this sector, mainly interacting with other dark particles, but potentially coupled also to the Standard Model through so-called "portal" interactions.
The Dark Sector hypothesis opens to an extremely rich phenomenology [67, 68], and does affect the modelling of DM production and evolution, for instance through self-interactions potentially enhancing the annihilation rate, or allowing additional possible decay channels into either SM or other Dark Sector particles.

2.2.1 Portal interactions

The Dark Sector particles are predicted to be neutral under the SM interactions. The assumed "portal" interactions [66, 68] are therefore an exception to the SM gauge structure of interactions, since these kind of forces could be mediated by direct couplings between

Dark Sector and Standard Model particles rather than by gauge bosons of an associated symmetry group. This interactions are required to conserve the SM gauge invariance, thus leading to a relatively limited number of viable interactions, including vector, neutrino, Higgs and axion portals. In addition, gauge interactions are still a possibility, as long as a new symmetry group is introduced.
A general form of the "portal" interaction lagrangian is

$$\mathcal{L}_{portal} = \sum O_{SM} \times O_{DS} \tag{2.12}$$

where O_{SM} and O_{DS} represent any generic operator related to SM and Dark Sector fields respectively.

The vector portal A simple realization of the vector portal is a dimension 4 operator, consisting in a coupling between the SM photon and an additional vector field A^{μ}_{d}, with strength tensor $F_{d,\mu\nu}$, the so-called Dark Photon, gauge boson of a new $U(1)_d$ symmetry group of the Dark Sector:

$$\mathcal{L}_{vector} = \mathcal{L}_{SM} + \mathcal{L}_{DS} - \frac{\epsilon}{2} F_{d,\mu\nu} F^{\mu\nu} \tag{2.13}$$

where ϵ is the kinetic mixing parameter [69], which scales with the Dark Photon coupling to the electromagnetic current, and $\mathcal{L}_{DS}$ the Dark Sector lagrangian, which might include a DM field χ:

$$\mathcal{L}_{DS} = -\frac{1}{4} F^2_{d,\mu\nu} + \frac{1}{2} m^2_{A_d} A^2_{d,\mu} + |(\partial_\mu + i g_d A_{d,\mu})\chi|^2 \tag{2.14}$$

The coupling via kinetic mixing to the SM photon is the simplest and most general one, but other cases are considered in the literature, such as $B-L$ and $L_\mu - L_\tau$ models, where the Dark Photon is assumed to be coupled to either total lepton current, or total baryon currents [70, 71].

The Higgs portal The Higgs portal is based on the existence of a new scalar field (S), originally predicted in the context of next-to-minimal SUSY models [72]. This scalar field couples with the Higgs boson (H) with dimension 3 or 4 operators, with coupling constants μ and λ respectively. Typically, the interactions scale with the Yukawa couplings (resulting in a suppression of the portal interaction for high-mass loops). The corresponding lagrangian is:

$$\mathcal{L}_{scalar} = \mathcal{L}_{SM} + \mathcal{L}_{DS} - (\mu S + \lambda S^2) H^\dagger H \tag{2.15}$$

where an interaction between the scalar S and DM might emerge in the DS term in the form $\mathcal{L}_{DS} = S\bar{\chi}\chi + ...$

Neutrino portal In this scenario, potentially related to neutrino mass generation, an additional heavy neutrino N couples with a fermion L and Higgs boson H in a 4-dimension operator. The Lagrangian is

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_{DS} + \sum F_{\alpha I} (\bar{L}_\alpha H) N_I \tag{2.16}$$

where the summation runs over the flavours α of the lepton doublets L_α and the additional heavy neutrinos N_I, and the $F_{\alpha I}$ indicates the corresponding Yukawa couplings. Here, the DS Lagrangian will include the mass terms, either Majorana or Dirac, of the new neutrinos N_I.

Axion portal This interaction consists in a dimension 5 interaction, where a new pseudoscalar particle (the axion or an axion-like-particle ALP described in Section 2.1.2) interacts with electromagnetic fields, gluons or fermions through the term:

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_{DS} + \frac{a}{4f_\gamma} F_{\mu\nu}\tilde{F}_{\mu\nu} + \frac{a}{4f_G}\mathrm{Tr}G_{\mu\nu}\tilde{G}_{\mu\nu} + \frac{\partial_\mu a}{f_l}\sum_\alpha \bar{l}_\alpha\gamma_\mu\gamma_5 l_\alpha + \frac{\partial_\mu a}{f_q}\sum_\beta \bar{q}_\beta\gamma_\mu\gamma_5 q_\beta \tag{2.17}$$

where a is the new pseudoscalar particle, F indicates the electromagnetic field, G the gluons, l the leptons and q the quarks, with the corresponding interaction strengths which depend on the scales of the theory, f_X.

SM charge gauges The most generic scenario predicts a new vector particle V_μ with interaction strength ϵ_V, with the requirement of an additional globally-conserved current of the SM (e.g. $B-L$, $L_\mu - L_\tau$).

$$\mathcal{L} = \epsilon_V \bar{f}\gamma^\mu q_f f V_\mu \tag{2.18}$$

2.2.2 Minimal Dark Sector and the Dark Photon

A minimal extension of the Lagrangian to accomodate the Dark Sector would consists in an additional $U(1)_d$ symmetry group, already mentioned in the vector "portal" scenario, with its associated gauge boson, the Dark Photon (γ_d) [66]. This gauge boson can be either massless or massive, with significant differences in the phenomenologies between the two cases.
It is useful to start from a generic Lagrangian involving two $U(1)$ symmetry groups:

$$\mathcal{L} = -\frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} - \frac{1}{4}F'_{d,\mu\nu}F'^{\mu\nu}_d - \frac{\epsilon}{2}F'_{\mu\nu}F'^{\mu\nu}_d + eJ_\mu A'^\mu + e_d J_{d,\mu}A'^\mu_d + \frac{1}{2}m^2_{\gamma_d}A'_{d,\mu}A'^\mu_d \tag{2.19}$$

where the label d identifies the Dark Sector and $F'^{\mu\nu}_{(d)}$ is the $A'^\mu_{(d)}$ strength tensor. The A'_μ field is only coupled to the SM J_μ current, while the dark field $A'_{d,\mu}$ to the dark current $J_{d,\mu}$. Nevertheless, in the massive Dark Photon case, a coupling between the dark photon and the SM current arises from kinetic mixing between the Dark Photon and the photon, while in the massless case an interaction can happen through operator with higher than 4 dimension. In the following, the massless and massive case will be considered separately.

Massless Dark Photon If the Dark Photon mass is null, $m_{\gamma_d} = 0$, it is possible to cancel the kinetic coupling term between the photon and the Dark Photon by means of a redefinition of the Dark Photon field. In particular, it is always possible to diagonalize the kinetic terms through rotations with an angle θ:

$$\begin{pmatrix} A'^\mu_d \\ A'^\mu \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1-\epsilon^2}} & 0 \\ -\frac{\epsilon}{\sqrt{1-\epsilon^2}} & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} A^\mu_d \\ A^\mu \end{pmatrix}$$

where the A^μ and A_d^μ are the physics fields respectively of the SM photon and of the Dark Photon. The mixing term in the Lagrangian can be rewritten :

$$\mathcal{L} = \left[\frac{e_d \cos\theta}{\sqrt{1-\epsilon^2}} J_{d,\mu} + e\left(\sin\theta - \frac{\epsilon\cos\theta}{\sqrt{1-\epsilon^2}}\right) J_\mu\right] A_d^\mu + \left[-\frac{e_d \sin\theta}{\sqrt{1-\epsilon^2}} J_{d,\mu} + e\left(\cos\theta + \frac{\epsilon\sin\theta}{\sqrt{1-\epsilon^2}}\right) J_\mu\right] A^\mu \tag{2.20}$$

Given that θ can assume an arbitrary value in the absence of a Dark Photon mass, two possible choices can be considered, to simplify the couplings structure such that either the SM or the dark current are coupled to only one gauge boson. In particular, if $\sin\theta = 0$, the SM photon is coupled only to the SM current, while the Dark Photon couples to both SM and dark sector currents. Oppositely, if $\sin\theta = \epsilon$, the Dark Photon couples only to the dark current, while the SM photon to both. More explicitly:

$$\sin\theta = 0 \rightarrow \mathcal{L} = \left[\frac{e_d}{\sqrt{1-\epsilon^2}} J_{d,\mu} - \frac{e\epsilon}{\sqrt{1-\epsilon^2}} J_\mu\right] A_d^\mu + e J_\mu A^\mu \tag{2.21}$$

$$\sin\theta = \epsilon \rightarrow \mathcal{L} = e_d J_{d,\mu} A_d^\mu + \left[\frac{e}{\sqrt{1-\epsilon^2}} J_\mu - \frac{e_d\epsilon}{\sqrt{1-\epsilon^2}} J_{d,\mu}\right] A^\mu \tag{2.22}$$

In the second case, the additional interaction between the photon and the dark current introduces a new charge commonly referred to as *milli-charge*, which is strongly constrained experimentally [73], as shown in the top left plot of Figure 6.4. a symmetry left over after the spontaneous breaking of a larger nonAbelian group.
Interactions between the Dark Photon and the SM would occur only through higher-order operators in this case. The simplest realization consists in a dimension-5 operator:

$$\frac{e_d}{2\Lambda^2} \bar{\psi}^i \sigma_{\mu\nu} \left(\mathbb{D}_M^{ij} + i\gamma_5 \mathbb{D}_E^{ij}\right) \psi^j F_d^{\mu\nu} \tag{2.23}$$

where $\sigma_{\mu\nu} = i[\gamma_\mu, \gamma_\nu]$, $\mathbb{D}_M$ and $\mathbb{D}_R$ are coefficients associated to the magnetic and electric dipole operators respectively and the i and j indexes label the fermion flavour.

Massive dark-photon In the massive case, the θ parameter is no more arbitrary. From a gauge theory point of view, this can be explained as the result of the symmetry breaking giving rise to the Dark Photon mass term, while breaking the degeneracy of the possible θ values. Alternatively, the dark photon might acquire mass by means of the Stueckelberg mechanism [74].
The θ value that originates the Dark Photon mass eigenstate is such that

$$\sin\theta = \frac{\delta\sqrt{1-\epsilon^2}}{\sqrt{1-2\delta\epsilon+\delta^2}} \tag{2.24}$$

where $\delta = M_{A'_d}/M_{A'}$ is the ratio of the A' and A'_d masses. In the Lagrangian, both the photon and the Dark Photon couples to the SM and the Dark Sector currents, meaning that the Dark Photon has now direct interactions, at tree-level, with the SM, and an additional coupling arises also for the photon.

$$\mathcal{L} = \frac{1}{\sqrt{1-2\delta\epsilon+\delta^2}} \left[\left(\frac{e_d(1-\delta\epsilon)}{\sqrt{1-\epsilon^2}} J_{d,\mu} + \frac{e(\delta-\epsilon)}{\sqrt{1-\epsilon^2}} J_\mu\right) A_d^\mu + \left(e J_\mu - \delta e_d J_{d,\mu}\right) A^\mu\right] \tag{2.25}$$

What described until now can be extended to a coupling to the SM hyper-charge rather

than the electromagnetic one, with the $U(1)_d$ gauge boson taking part, if massive, to the diagonalization of the B and W^3 gauge bosons to produce the physical Z, A and A_d bosons. In this case, also the Z boson will have additional coupling to the dark current, and the mixing terms of the Lagrangian, assuming $\delta = 0$ and therefore no coupling between the photon and the dark-sector, will be of the form

$$\mathcal{L} = +e_d \epsilon \tan\theta_W J_d^\mu Z_\mu + (e_d J_d^\mu - e\epsilon J^\mu) A_{d,\mu} \tag{2.26}$$

In the massless case, the only effect of a coupling to the $U(1)_Y$ would be that also the Z boson would be coupled to the dark-sector current.

2.3 Detection strategies

In order to understand the nature of DM, it is essential to observe possible non gravitational interactions with ordinary matter.
The search for DM is performed by means of three different approaches, based on different processes and interactions between DM and SM particles, as illustrated in Figure 2.8:

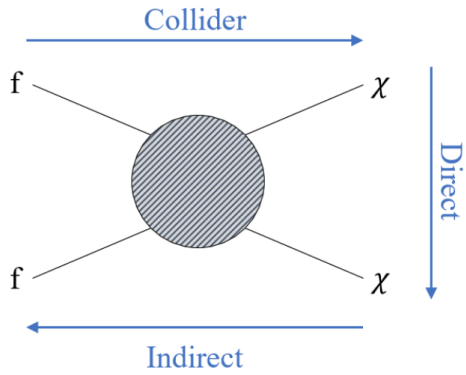


Figure 2.8: A scheme of the different approaches to DM search.

- Direct Detection (DD), looking for scattering of DM off ordinary matter;
- Indirect Detection (ID), looking for products of annihilation of pairs of DM particles;
- Production at colliders, aiming to produce Dark Matter (or generally Dark Sector) particles through high-energy collisions of Standard Model Particles.

The importance of multiple approaches to DM search lies in their complementarity, which provides sensitivities to different DM candidates and interactions and allows to explore the several open scenarios.

2.3.1 Direct Detection and exclusion limits on WIMPs

Direct Detection experiments are especially performant in the context of WIMP searches. They are based on the assumption that DM particles interact with ordinary matter through elastic scattering off nuclei (and more recently electrons).

The observed signal of DM interaction with the detector is the nuclear recoil energy which, for WIMP masses in the range 20 GeV-1 TeV, is of about 1 keV-50 keV. Therefore, it's essential to obtain an extremely low energy treshold.
Moreover, the event rate comes out to be extremely low (about $10^{-1}-10^{-6}$events/kg/day) due to very weak interactions, thus requiring high mass detectors, in order to enhance the probability to observe a signal, and a suppression as effective as possible of the various background sources (cosmogenic background, environmental radioactivity, detector radioactivity). The background suppression is pursued by shielding techniques, signal background discrimination (for instance by means of Pulse Shape Analysis, hybrid detectors or by using a neutron detector to isolate neutron recoils off nuclei), the choice of radio-pure materials and of underground sites for the detector.
These experiments are based on some theoretical assumption, for instance related to the DM Halo shape or to the velocity distribution of the DM 'wind", which is commonly assumed to follow a Maxwellian distribution. Another theoretical assumption relies on the type of DM-nucleus interaction: in Spin Independent (SI) interactions, the scattering is coherent and scales as the atomic mass squared, A^2, thus allowing to enhance the experiment sensitivity by using heavier nuclei; on the other hand, the Spin Dependent (SD) scattering is due to the interaction of a WIMP with the spin of the nucleus and takes place only in those detector isotopes with an unpaired proton and/or unpaired neutron. At present, the stronger limit are set for the SI scenario.
Besides looking for a scattering of DM particles off nuclei in background-free environments, another strategy can be applied, which doesn't require a strong background suppression: it takes advantage of the expected annual modulation of the DM signal, due to the relative motion of the Earth with respect to the WIMP "wind". As a matter of fact, the random motion of WIMPs in the Galaxy, combined with the Sun's motion, creates (on the average) a relative velocity between Earth and the WIMPs. On top of that, because of the Earth revolution, the relative velocity of the Earth with the WIMP wind varies with the time of year, leading to a sinusoidal modulation peaking in June and reaching its minimum in December.
Various types of detecting techniques are employed, based on ionization, scintillation or heat detection: for instance, high-purity scintillation detectors (such as DAMA/LIBRA, Dark Side, Picasso), high-purity semi-conductors (such as CoGENT, IGEX), liquid noble gases detectors (XMASS, XENON100, LUX, DEAP-3600, DarkSide-50), bolometric detectors (COUPP, Cuoricino), cryogenic crystal detectors (CDMS II), hybrid detectors (CREEST-II, CDMS, EDELWEISS).
Among the several DD experiments, DAMA/LIBRA [75, 76, 77] claimed a 13.7σ detection of " anomalous" modulation. Nevertheless, other experiments do not observe a signal compatible with DAMA/LIBRA or even exclude that region of the phase space, and no confirmation of this result has been found until now. Some experiments, such as SABRE, COSINE-100, ANAIS and a collaboration between DM-ICE and KIMS, using the same detector material (NaI) as DAMA, aim at confirming or rejecting the DAMA/LIBRA observation. SABRE and DM-ICE will also be able to exclude the hypothesis of a seasonal origin of the observed modulation, thanks to the fact that they're sited in the opposite emisphere with respect to DAMA.

2.3.2 Indirect Detection

Indirect detection experiments look for the product of WIMP annihilation, including gamma rays, positrons, antiprotons, antinuclei or neutrinos. These experiments are based

on the assumption that WIMPs can be gravitationally trapped in the Earth, the Sun or the galaxy halo, therefore enhancing the annihilation probability.
WIMPs annihilating in the Sun could be a source of muon neutrinos which can interact in the Earth. This would give rise to the observation of upward going muons, which can be detected in large neutrino telescopes such as MACRO, BAKSAN, SuperKamiokande, Baikal, AMANDA, ANTARES, NESTOR, and IceCube.
On the other hand, WIMP annihilation in the halo can give a continuous spectrum of gamma rays. The HESS Cherenkov detector and the Fermi-LAT Telescope have observed an excess respectively of 1 TeV and 1 GeV γ radiation near the galactic center. At present, an interpretation in terms of DM is disfavoured, although DM could give a contribution of at most 5% to the Fermi-LAT excess [78, 79].
Antiparticles produced in DM annihilation in the halo can be detected by the PAMELA satellite and the AMS-02 experiment on the International Space Station. An excess in the positron flux with energies grater than 10 GeV has been observed by PAMELA, despite not associated to an antiproton excess. The same excess is also confirmed by the AMS-02 experiment, but it is likely explainable without recurring to DM hypotheses.
Several other experiments are conducted, focusing on specific models or candidates, such as axion searches by means of polarization experiment, haloscopes or helioscopes, typically exploiting the expected mixing between axions and photons. Similar technologies can be used also for Dark Photon searches (see for instance [80]).

2.3.3 Production at Colliders

The search for DM at the LHC looks for an hint of DM production in high-energy pp collisions. Being weakly interacting and long-lived, DM or Dark Sector particles produced at LHC could not be directly observed by the ATLAS detector, but need to be produced in association to a visible SM particle. In this case, the invisible DM particles, similarly to neutrinos, can be detected as an imbalance in the total transverse momentum of the final state, called "missing transverse momentum" ($E_{\mathrm{T}}^{\mathrm{miss}}$), recoiling against the visible particles.
Different final states are exploited by the ATLAS Collaboration, and interpreted in terms of several different models, with searches focusing not only on the observation of a signature of DM production, but also on BSM resonances providing an hint of the existence of some new mediator field which can be coupled both with the SM and an hypothetical dark sector.
In Chapter 6, some more details will be provided about potentialities of these kind of searches, focusing on specific models of interest for the analyses described later in this thesis.

Chapter 3

LHC and ATLAS

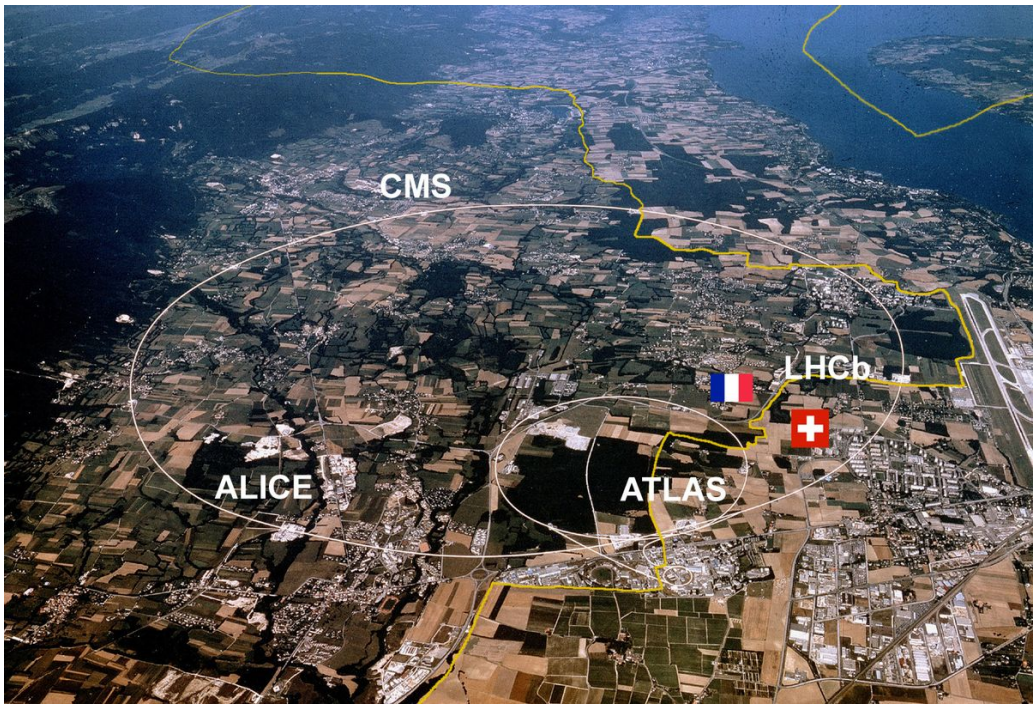


Figure 3.1: Aereal view of the Large Hadron Collider, with the locations of the main experiments along the ring highlighted [81].

The Large Hadron Collider (LHC) [82] at CERN , is the largest existing circular protons and heavy ion accelerator and collider. It consists of a 27 km ring, equipped with superconducting magnets which provide the electromagnetic field to guide the proton beams throughout their trajectory, and located in an underground tunnel, which develops between France and Switzerland (Figure 3.1) at a depth ranging between 50 m and 175 m. It was built in the same tunnel previously hosting the Large Electron Positron collider (LEP) which was dismantled in 2000 after 11 year of operation and after having provided important precision measurements of the Z and W bosons.
Four main detectors are placed along the LHC ring, in four collision points: the two multipurpose experiments ATLAS (A Toroidal LHC ApparatuS) [83] and CMS (Compact Muon Spectometer) [84], ALICE (A Large Ion Collider Experiment) [85], dedicated to heavy ion collisions to study the strong interactions of quarks–gluon plasma, and LHCb

(Large Hadron Collider beauty experiment) [86], focused on B hadron. In the four mentioned collision points, proton beams, composed of 2808 bunches of 10^{11} protons, collide every 25 ns, corresponding to a beam collision frequency of 40 MHz.

3.1 The Large Hadron Collider

The LHC was designed to reach record energies in the centre-of-mass and luminosities, overcoming the limits arising from large radiative energy loss of an electron-positron collider and pushing towards the discovery of new physics thanks to the large range of accessible collisions energies in composite objects collisions. The initial target centre-of-mass energy and luminosity were 14 TeV and 10^{34} cm^{-2}s^{-1} respectively. After the first Run-1 data-taking at 7 TeV in 2010-2011 and 8 TeV in 2012, the centre-of-mass was increased to 13 TeV during Run-2 between 2015 and 2018, and 13.6 TeV at the start of Run-3 in July 2022.

3.1.1 Proton-proton collisions

As previously mentioned, synchroton radiation would make e^+e^- collisions extremely challenging at the energies of interest, being the energy loss proportional to E^4/m^4 [87], E and m being the accelerated particle's energy and mass, resulting in an energy loss about 10^{13} higher for accelerated e^+/e^- than for protons, for a given target energy.
The advantage of accelerating protons rather then electrons comes with the drawback of proton's compositeness. The interactions of interest for physics at the LHC are the *hard collisions*, i.e. short-range interactions among the partons (quarks and gluons) in the proton, with high transferred momentum. These are the rarest events though, while the largest fraction of interactions happening in each bunch crossing are long-distance interactions with low transferred momentum, the *soft collisions*. The latter constitute an important background for physics searches at the LHC: a large number of *pile-up* soft collisions between other protons in the same bunch crossing (*in-time pile-up*) or neighbour bunch crossings (*out-of-time pile-up*), as well as interactions between other partons in the same protons (*Underlying Events*, UE) happen simultaneously to the events of interest, thus producing an extremely dense environment and degrading the reconstruction performance and resolution for hard-scatter interactions.
In the *hard collisions*, each parton carries a fraction x, unknown a priori, of the proton energy, thus reducing the available centre-of-mass energy $\sqrt{\hat{s}}$ in the parton-parton collision with respect to the proton-proton one ($\sqrt{s}$) according to the equation:

$$\sqrt{\hat{s}} = \sqrt{x_a x_b s} \tag{3.1}$$

The cross-section for the pp hard scattering interactions is given by

$$\sigma_{pp\to X} = \sum_{a,b} \int dx_a dx_b f_a(x_a, Q^2) f_b(x_b, Q^2) \hat{\sigma}_{ab\to X}(x_a, x_b) \tag{3.2}$$

where $\hat{\sigma}_{ab\to X}$ is the cross-section of the parton-parton interaction and $f_i(x_i, Q^2)$ are the Parton Distribution Functions (PDFs) of the two colliding partons at energy scale Q energy scale, Q being the transferred momentum in the collision. The PDFs for gluons and different flavour quarks are shown in Figure 3.2, with $Q^2 = 10$ GeV^2 on the left, and $Q^2 = 10^4$ GeV^2 on the right, in the MMHT 2014 PDFs computation [88].

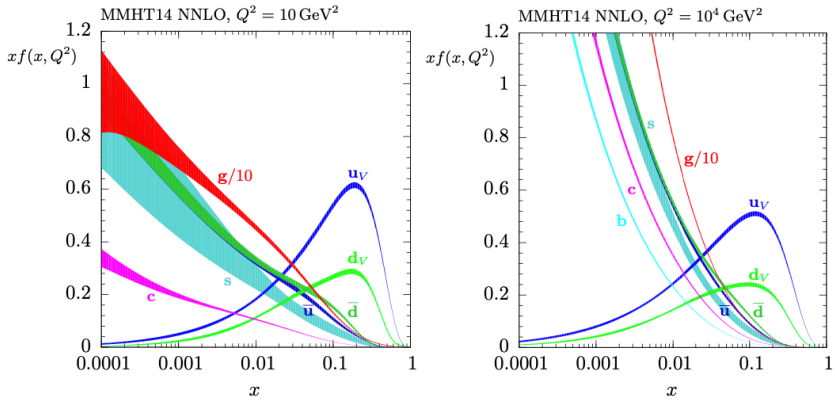


Figure 3.2: Parton Distribution Functions, times the parton energy fraction, at $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 from MMHT2014 NNLO computation [88]

Theoretical computations of the PDFs are therefore a key ingredient in the determination of the cross-section of a given process, and give rise to one of the main theoretical systematic uncertainties. On the other side, as anticipated, the proton compositeness gives access to a wide range of parton scattering energies in hadronic colliders, which is an interesting features towards the discovery of new physics.

3.1.2 The accelerator complex

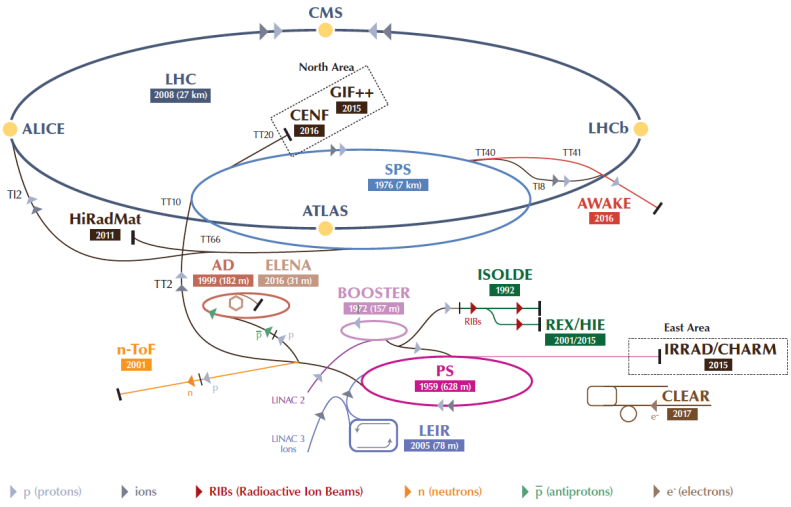


Figure 3.3: The LHC accelerator chain [89].

The LHC is the final stage of an accelerating chain depicted in Figure 3.3, which progressively increases proton beams energies through four pre-accelerators. In the first stage, protons are accelerated to 50 MeV in the linear LINAC accelerator. The Syn-

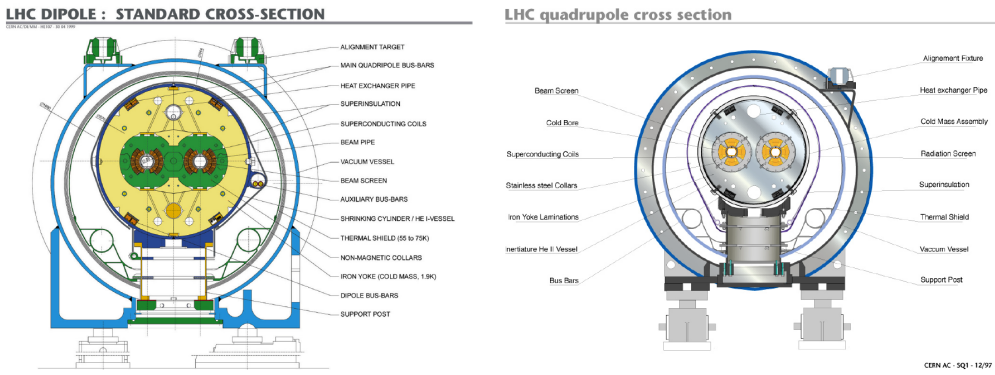


Figure 3.4: Cross-section of the LHC dipole [91] (left) and quadrupole [92] (right)

chrotron Booster provides a further acceleration up to 1.4 GeV, followed by the Proton Synchrotron (PS) where an energy of 25 GeV is reached. Finally, proton beams pass through the Super Proton Synchrotron (SPS), where they are accelerated to 450 GeV. With this energy, they are injected, in the two opposite directions, in the LHC ring.
In the LHC, the acceleration is provided by eight single-cell superconducting Radio Frequency cavities (RF) [90], each delivering 2 MV (5.3 MV/m) at 400 MHz and operating at 4.5 K. The system relies on dipole and quadrupole magnets (Figure 3.4) to keep the trajectory along the ring and focus the beams. The magnets are grouped in cells composed by six dipole magnets and two quadrupole magnets with opposite polarity. A total of 1232 Niobium-Titanium superconducting dipole magnets, operating at a temperature of 1.9 K, produce a 8.4 T magnetic field at a current of 11,700 A to deflect the protons, while the beam focus in the horizontal or vertical plane, depending on the magnet polarity, is obtained by means of 392 quadrupole magnets. Higher order magnets, i.e. sextupoles, octupoles and decapoles are employed to correct for non-linear and chromatic effects, and to counteract external interactions - such as gravitational interactions over protons, electromagnetic interactions among bunches, electron clouds from the pipe wall among others - to ensure the required beam stability over time. Finally, eight sets of "inner triplet" magnets focus the particle beams into the four colliding areas, squeezing the bunch transverse size from 0.2 mm down to 16μm.

3.1.3 Luminosity

An important quantity at the LHC is the instantaneous luminosity $\mathcal{L}$ [93], a proportionality factor between the event rate and the cross section σ for a given process, defined as:

$$\frac{dN}{dt} = \mathcal{L}\sigma \tag{3.3}$$

This quantity depends on the beam parameters, according to:

$$\mathcal{L} = \frac{N_b^2 n_b f_{ref} \gamma_r}{4\pi\epsilon_n \beta^*} F \ [\mathrm{cm}^{-2}\mathrm{s}^{-1}] \tag{3.4}$$

where N_b and n_b indicate respectively the number of particles per bunch and the number of bunches, f_{rev} is the revolution frequency, γ_r is the relativistic gamma factor, ϵ_n the transverse normalized beam emittance, β^* the beta focusing function at the collision

point, and F the geometric luminosity reduction factor due to the crossing angle at the interaction point.
A summary of the beams parameters employed during different periods of the Run-2 data-taking and the design ones is reported in Table 3.1, while Figure 3.5 shows the integrated luminosities delivered by LHC and collected by the ATLAS experiment, together with the ones "good for physics" in the same period. The peak luminosity per fill are also shown, for separate years of the Run-2 data-taking, in Figure 3.6, highlighting the gradually increasing reached instantaneous luminosities.

Parameter	**Design**	**2015**	**2016**	**2017**	**2018**
Energy	7.0	6.5	6.5	6.5	6.5
f_{rev} [kHz]	11.2	11.2	11.2	11.2	11.2
Number of bunches	2808	2244	2220	2556-1868	2556
N_b [10^{11} particles]	1.15	1.2	1.25	1.25	1.1
β^* [cm]	55	80	40	40→30	30→27→25
ϵ_n [μm rad]	3.75	2.6 - 3.5	1.8-2	1.8-2.2	1.8-2.2
Crossing angle θ_c [μrad]	285	370	370-280	300-240	320-260
γ_r	7462	6929	6929	6929	6929
F [%]	84	84	65	72	61
Peak luminosity [10^{34}cm^{-2}s^{-1}]	1.0	< 0.6	1.5	2.0	2.1

Table 3.1: Summary of beam and machine parameters during the four years of Run 2, compared to the design values [94]

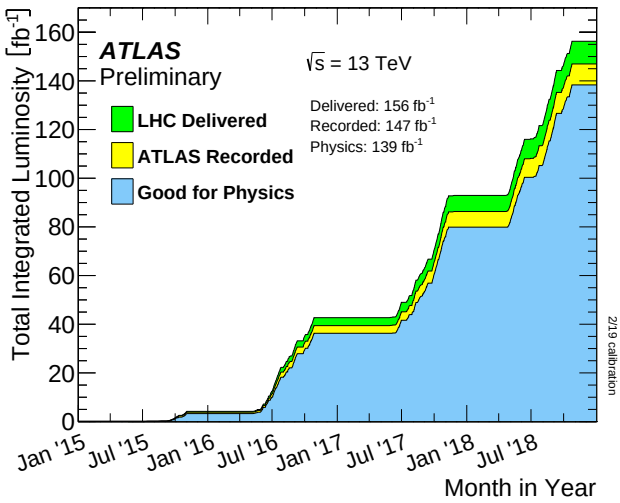


Figure 3.5: Integrated luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow) and good for physics (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in LHC Run-2 [95].

An important correlation exists between the instantaneous luminosity and the average number of inelastic interactions per bunch crossing, μ, which quantifies the *pile-up*

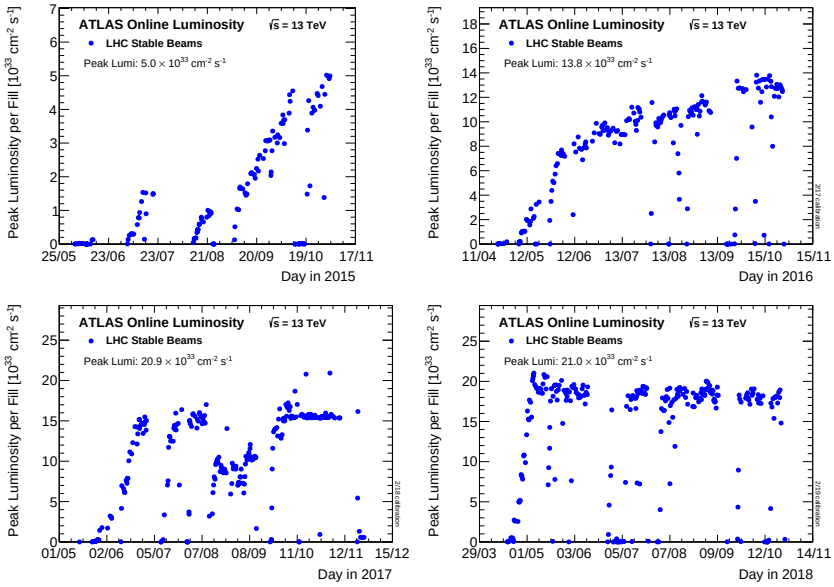


Figure 3.6: Peak instantaneous luminosity delivered to ATLAS during stable beams for pp collisions at $\sqrt{s}$=13 TeV for each LHC fill as a function of time in 2015, 2016, 2017 and 2018 [95].

conditions. In particular, higher luminosities correspond to higher μ:

$$\mathcal{L} = \frac{\mu n_b f_{rev}}{\sigma_{inelastic}} \tag{3.5}$$

It is therefore important to find a good compromise between the necessity of pushing the luminosity to maximize statistics and the consequent worsening of pile-up conditions. The distribution of the average μ for the full Run-2 *pp* collision data are shown in Figure 3.7, showing the increasing pile-up year-by-year, consistently with the enhancement of the delivered luminosity.

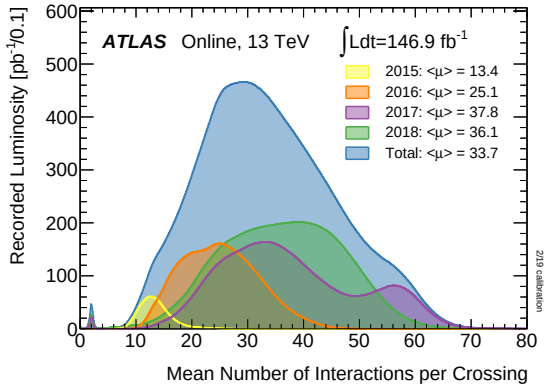


Figure 3.7: Luminosity-weighted distribution of the mean number of interactions per crossing for the full Run-2 *pp* collision data at $\sqrt{s} = 13$ TeV [95].

3.2 The ATLAS detector

The ATLAS experiment [83] is one of the multi-purpose experiments at the LHC, together with CMS.

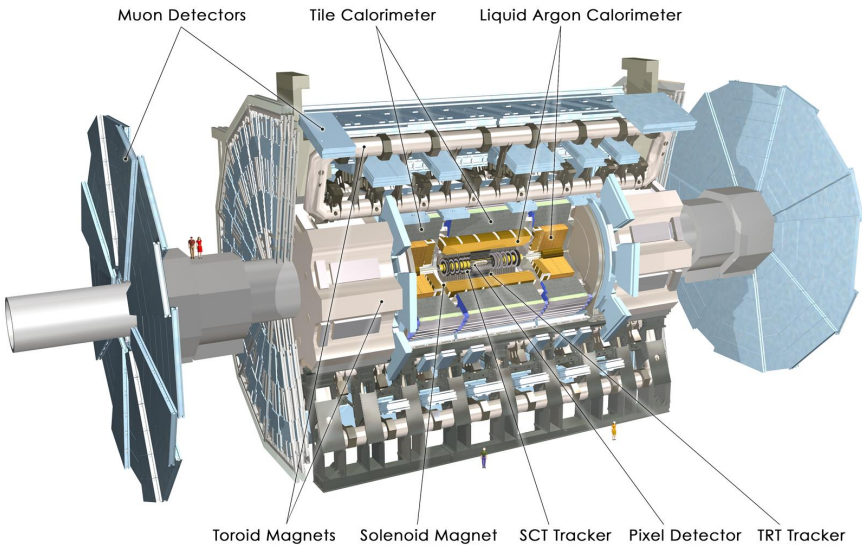


Figure 3.8: Schematic view of the ATLAS detector [96]

It is the largest detector at the LHC, with a forward–backward symmetric cylindrical geometry, covering nearly 4π in solid angle, with a radius of 12 m, a length of 44 m and approximately 7000 tons of weight. The detector is composed of 4 concentric sub-detectors, as shown in Figure 3.8. From the most central to the outward:

- The Inner Detector (ID) for the tracking of charged particles and reconstruction of interaction vertex;
- The Electromagnetic Calorimeter (EM) for the measurement of electromagnetic showers of electrons and photons;
- The Hadronic Calorimeter (HEC) for hadronic showers produced by charged and neutral hadrons;
- The Muon Spectrometer (MS) for muon identification and measurement.

3.2.1 Reference frame

A right-handed coordinate system, as shown in Figure 3.9 is used, with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe, defining the transverse plane as the x-y one. The x-axis points from the IP to the centre of the LHC ring, while the y-axis points upward. A cylindrical system is also used, with ϕ the azimuthal angle measured around the beam direction and θ the polar angle with respect to the z-axis. Since interactions are typically boosted along the z-axis due to different energy fractions carried by the two colliding partons, a z-boost Lorentz invariant

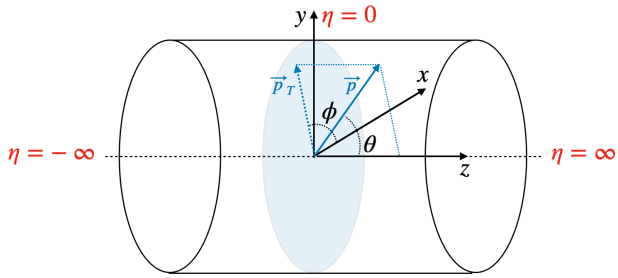


Figure 3.9: ATLAS coordinate system

quantity is more suitable. This invariance is ensured in the (η, ϕ) coordinate system by the $\Delta\eta$ variable, with η the *pseudo-rapidity*:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{3.6}$$

derived as the massless limit of the rapidity $Y = \frac{1}{2}\ln\left(\frac{E+p_{||}}{E-p_{||}}\right)$. Another variable of interest is the angular distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$

3.2.2 The magnets system

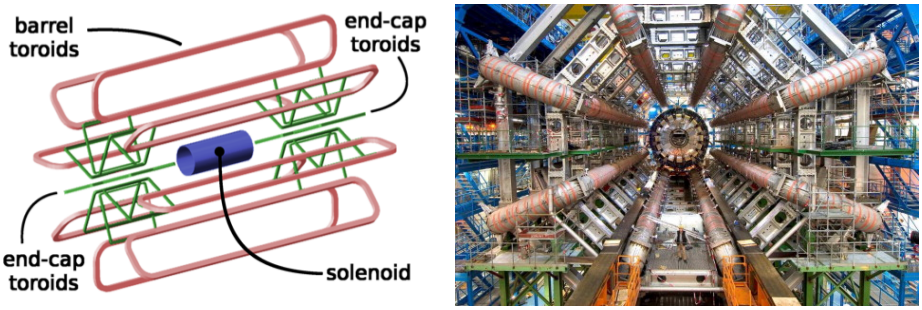


Figure 3.10: A schematic representation of the ATLAS magnets system layout (left) and a view of the barrel toroids (right).

The superconducting magnet system, schematically represented in Figure 3.10, provides the electromagnetic field needed for charged particle momenta measurement in the Inner Detector and in the Muon Spectrometer. The ID is served by a solenoidal central system, while the MS by two end-caps and one barrel toroids. The *solenoids* are coils of superconducting material located inside the calorimeter cryostats and providing an axial magnetic field of 2 T to the ID and operating at a temperature of 4.5 K. The full system covers a 5.8 m long region, with an inner radius of 1.23 m, and an outer one of 1.28 m, corresponding to 0.66 radiation lengths.
On the other hand, the *toroidal* system is splitted into a barrel toroid, composed of 8 separate coils providing 0.5 T magnetic field, and two end-caps toroids consisting each of

eight superconducting coils inside an insulating vacuum vessel, providing 1 T field. These magnets are located outside the calorimeters. The central toroid extends for a length of 25.3 m, with internal diameter of 9.4 m and external one of 20.1 m, while the end-caps vessel has a 10.7 m diameters and a 5 m wideness.

3.2.3 The Inner Detector

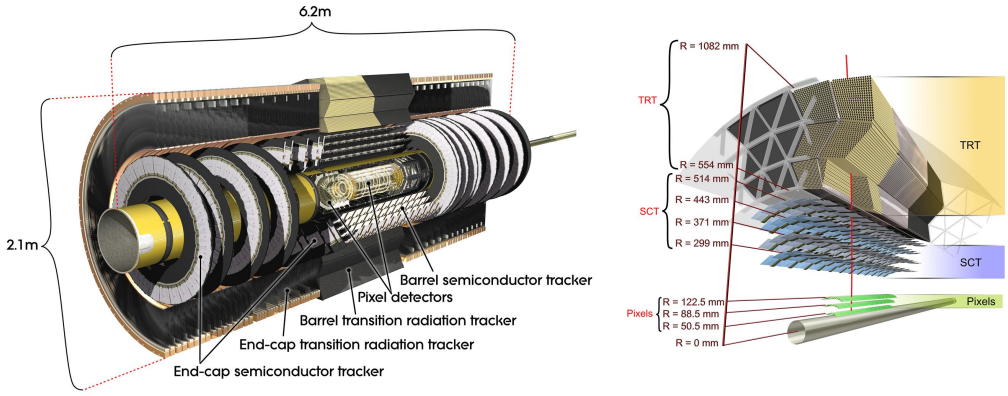


Figure 3.11: The ATLAS Inner Detector (ID) [97]

The inner tracking detector (ID), covering the pseudorapidity range $|\eta| < 2.5$, has a diameter of 2.1 m and is 6.2 m long and is splitted into a barrel and two end-cap regions. As shown in Figure 3.11, it consists of a silicon pixel detector including the insertable B-layer [98], which was added around a new smaller-radius beam-pipe before the start of Run 2; a silicon microstrip detector; and, for $|\eta| < 2.0$, a straw-tube transition radiation tracker (TRT). The 2 T axial magnetic field provided by the solenoid allows to bend charged particles with a radius and directions dependent respectively on the intensity and direction of their momenta, thus allowing for momentum and charge measurements. The structure of this sub-detector was optimized to concurrently satisfy multiple requirements: a limited amount of material in order to avoid multiple scattering and minimize photon conversion and bremsstrahlung emissions; a good radiation resistance to cope with high exposure to radiation damages due to the small distance from the interaction point; a high granularity of tracking, especially at smallest radii, in order to correctly resolve multiple overlapping interactions.
The overall resolution on the reconstruction of a charged particle transverse momentum was $0.05\% \cdot p_T \bigoplus 1\%$ during Run-1, with a further improvement arising from the insertion of the Insertable B-Layer (IBL) during Run-2. The first term is related to the resolution of the curvature measurement, while the constant term of 1% accounts for the impact of multiple Coulomb scattering.

Pixel detector

The first component is the Pixel Detector, the closest one to the beam pipe, with a minimum distance of 31 mm and extending up to a radium of 123 mm, covering the $|\eta| < 2.5$ region. It is characterized by the highest granularity and constituted of pixels silicon sensors segmented in $R - \phi$ and z and arranged in four concentric barrel layers and two end-caps composed of three disks each. The typical pixel size is of $50 \times 400 \mu m^2$,

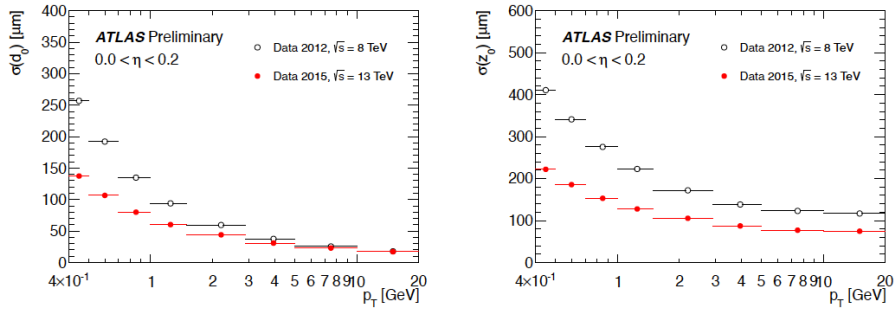


Figure 3.12: Transverse (left) and longitudinal (right) impact parameter resolution as a function of the transverse momentum p_T in data collected in 2015 (Run-2) with the IBL installed, compared with 2012 (Run-1).

resulting in a single hit resolution of about $13\mu m$ in the (R, ϕ) plane and $72 - 115\mu m$ in the z direction. The best resolution in the z direction is obtained in the Insertable B-Layer, installed as a first barrel layer before Run-2, at a distance of 3.3 cm from a new smaller-radius beam pipe, to improve the resolution on the transverse and longitudinal impact parameters (Figure 3.12) and recover the tracking and b-tagging efficiency of the Pixel detector that was partially degraded due to radiation damage.

Semi Conductor Tracker

The Semi-Conductor Tracker extends between 30 cm and 51 cm in radius, arranged in eight concentric barrel layers of silicon micro-strips with typical pitch size of 80 μm, covering the $|\eta| < 1.4$ range, and two end-caps of nine disks each, with pitch size ranging from 56.9 μm to 90.4 μm, reaching up to $|\eta| = 2.5$. The barrel layers are organized in 4 cylinders, providing 4 additional hits per track and made of two layers of sensors glued back-to-back with a 40 mrad stereo angle. The intrinsic hit resolution of the strips is of about 16 μm along the $R - \phi$ plane and about 580 μm along the z-axis.

Transition Radiation Tracker

The most external component of the ID is the Transition Radiation Tracker (TRT), located at radii between 55 cm and 108 cm and serving as a drift chamber measuring the charge drift time and as transition radiation detector for electron identification. It consists of 370 000 drift tubes (called straw), filled with Xenon or Argon gas mixture, among which 50000 tubes are located in the barrel, parallel to the beam, while the remaining ones are placed orthogonally in the end-caps. The charged particles drift is produced through a potential difference of -1.5 kV between a gold-plated tungsten wire kept at ground potential at the centre of the tube, and the walls. In addition, the spaces between the straws are filled with polymer fibers in the barrel and foils in the endcap, in order to create transition radiation emitted by charged particles passing through the material boundaries, which provides an additional energy release in the gas with consequent enhancement of the readout signals with an amplitude that can exceed the 6 keV high threshold (HT). The dependency of the emitted photons spectrum on the γ factor of the incident particles ensure a good discrimination power between electrons and poins up to 200 GeV, thanks to the fact that only the most energetic pions have large enough γ factor to produce TR.

The TRT provides up to 36 additional hits, with an $R-\phi$ resolution of 130 μm for charged tracks with $|\eta| < 2$ and $p_T > 0.5$ GeV,

3.2.4 Calorimeters

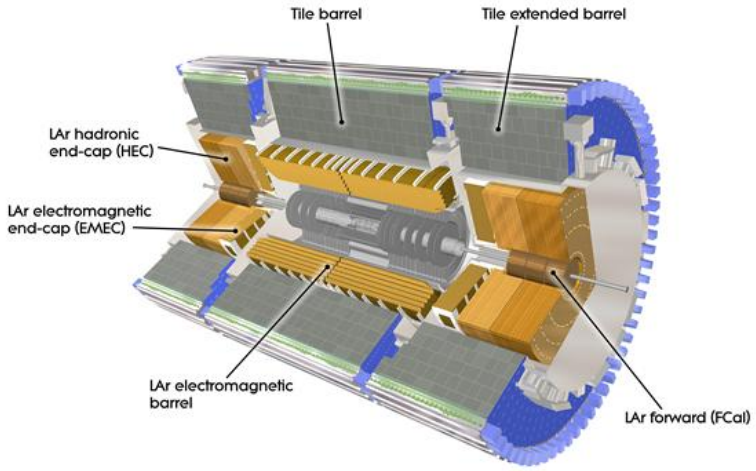


Figure 3.13: The ATLAS calorimetric system layout [99]

The calorimetric system (Figure 3.13) aims at measuring the energy deposits released by both neutral and charged particles (electrons, photons or hadrons), and their directions, having a key role also in the reconstruction of the missing transverse momentum. Three different calorimeters compose this part of the detector, covering different $|\eta|$ regions and targeting different signals: the Electromagnetic Calorimeter (EM) covers the pseudorapidity region $|\eta|$ <3.2 and is designed to collect the energy deposits from electrons and photons, the Hadronic Calorimeter (Had) covers $|\eta| < 3.9$ and is addressed to hadronic measurements, while the Forward Calorimeter (FCal) targets noth electrons/photons and hadrons 3.1< $|\eta|$ <4.9.
When particles interact with the detector material, secondary particles are created, which further interact with the detector thus starting a cascade process which goes under the name of "shower". Depending on the nature of the initial particle, the shower will be entirely electromagnetic (starting from electrons or photon) or electromagnetic and hadronic (starting from hadrons). Electromagnetic and hadronic "showers" develope differently:

- The electromagnetic "shower" is fed by e^+e^- pair production from photon interaction with the detector material, and the emission of photons from high energy electrons via bremsstrahlung. The first process stops as soon as the photon energy goes below the pair production threshold, while the latter when the critical energy $E_c = (610 \text{ MeV})/(Z + 1.24)$ (with Z the atomic number of the detector material) is reached, meaning that the energy loss by bremmstrahlung equals the one by ionization. The electromagnetic shower is mainly characterized by its depth and its width. The firs quantifies the amount of material which is able to contain 95% of

the shower energy in the longitudinal plane, and is parametrized as:

$$L(95\%) \sim \left[\ln\left(\frac{E_0}{E_c}\right) \pm 0.5 \right] + 0.08Z + 9.6 \; [X_0] \tag{3.7}$$

where E_0 is the energy of the incident particle, the $+0.5$ and -0.5 values are taken respectively for photon and electron, and X_0 is the radiation length quantifying the amount of material where an electron energy would be reduced by a factor e:

$$X_0 = \frac{716.4 \; [\text{g cm}^{-2}] \times A}{Z(Z+1)\ln(287/\sqrt{Z})} \tag{3.8}$$

The "shower" width, on the other side, is quantified by the *Molière radius* R_M, defined such that 95% of the shower develops within $2 \times R_M$:

$$R_M = \frac{21 \text{ MeV}}{E_c} X_0 \tag{3.9}$$

- The hadronic "shower" includes an electromagnetic component constituting about $30\% - 60\%$ of the total energy and arising from π^0 and η decays into photons by absorption, together with an hadronic component from remaining processes. The latter is characterized by about a 40% of *invisible energy* mainly due to the nucleon binding energy, which doesn't produce any signal. Hadronic 'showers" are described by the nuclear interaction length

$$\lambda \sim 35A^{1/3} \text{ g cm}^{-2} \tag{3.10}$$

which can quantify the containment of the longitudinal energy profile: the 95% of the longitudinal energy deposits are within $t_{95\%} = t_{max} + 2\lambda E^{0.13}$, with $t_{max} = 0.2\ln(E/1\text{GeV}) + 0.7$ $[\lambda]$. This translates in a much higher longitudinal length for hadronic than for electromagnetic shower, thus making it possible to discriminate between e/γ or hadronic incident particles by studying the fraction of energy deposits in the electromagnetic calorimeters, located closer to the interaction point, over the deposits in the hadronic calorimeters, surrounding the electromagnetic ones.

The ATLAS detector includes exclusively *sampling calorimeters*, with Liquid Argon (LAr) or polystyrene scintillator as an active medium, where the "showers" develope releasing measurable signals, and lead (Pb), copper or iron as passive material (or *absorber*), with shorter radiation and interaction lengths, inserted to accelerate the energy degradation of the "shower". The advantage of sampling calorimeters lies not only in economic reasons, but also in a higher radiation hardness and in the segmentation which provides direction information. The drawback is a lower energy resolution, due to the energy loss in the passive materials, with respect to an homogeneous calorimeter. In particular, the resolution is reduced by a factor $\sqrt{t}$, t being the thickness of the absorber.

The Electromagnetic Calorimeter

The Electromagnetic Calorimeter (EM) is divided into a barrel detector covering the $|\eta| < 1.475$ region and two end-caps (EMEC) covering $1.375 < |\eta| < 3.2$, for a global extension of 6.65 m in length and outer radius of 2.25 m. It consists in a sampling

calorimeter arranged in an accordion geometry (Figure 3.14) to provide a full coverage in ϕ and avoid cracks due to the outgoing readout system. The full containment of the EM shower is ensured by the global thickness of the calorimetric system, $> 22X_0$ in the barrel and $> 24X_0$ in the end-caps. These conditions are reached employing the LAr as active medium, arranged in 2 mm layers and interleaved with copper electrodes collecting the ionization charges produced by the EM showers, and lead plates as absorbers, with a thickness of 1.5 mm for $|\eta| < 0.8$ and 1.13 mm for $|\eta| > 0.8$.
The system is longitudinally segmented into 4 layers, each one further segmented in $\eta - \phi$ with a granularity ranging from 0.003×0.025 to 0.1×0.1 depending on the η region and layer (Figure 3.14). The four longitudinal layers include a *presampler* and 3 further layers.

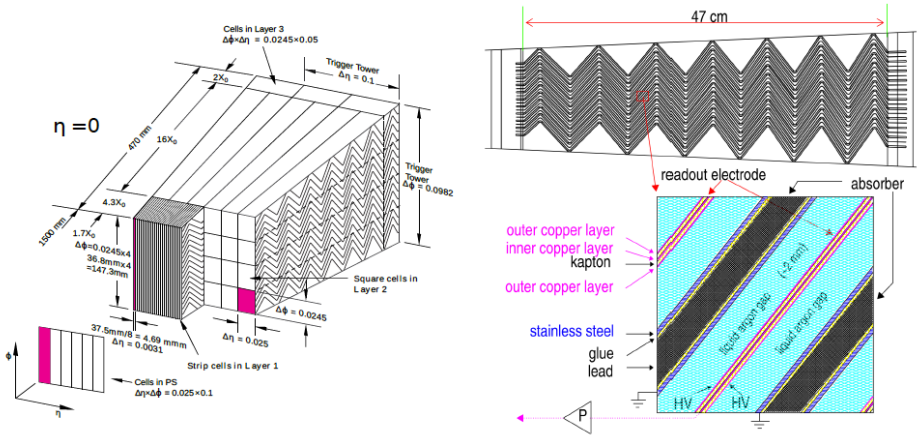


Figure 3.14: Electromagnetic Calorimeter accordion structure and barrel granularity, with a detailed description of the active and passive material layer composition

- The *presampler* (PS) is the closest layer to the beam pipe, covering the $|\eta| < 1.8$ region with a thickness of 1.1 cm. It is located inside the solenoid and aims at measuring and correcting for the particle energy loss in the upstream material;
- The *strips* layer (Layer 1) is the first in the accordion shape, extending up to $6X_0$. It is characterized by the finest granularity in η, with $\Delta\eta = 0.003$, aiming at the discrimination between prompt photonss and $\pi^0 \to \gamma\gamma$ with almost collinear photons;
- The *middle* layer (Layer 2) reaches $22X_0$, with finer segmentation $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ in the barrel, and up to $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the end-caps. It's the layer where the largest part of the shower energy is deposited;
- The *back* layer (Layer 3) covers the last 2 radiation lengths with $\Delta\eta \times \Delta\phi = 0.050 \times 0.025$, and is included with the purpose of estimating the possible energy leakage of EM shower into the hadronic calorimeter;

The nominal resolution of the EM calorimeter is:

$$\frac{\sigma(E)}{E} \sim \frac{(10\% - 17\%)}{\sqrt{E}} \oplus 0.7\% \tag{3.11}$$

where the stochastic term depends on the η region.

The Hadronic Calorimeter

The Hadronic calorimeter (HAD) surrounds the EM Calorimeter, to collect energy deposits from the hadronic showers, characterized by a larger longitudinal development than the electromagnetic ones.
With its 4.2 m external radius and 11.5 m length, it contains up to 11λ and it covers the $|\eta| < 3.9$ pseudo-rapidity range, divided into two sub-system, the *Barrel* and the *End-cap* (HEC).
The *Barrel* calorimeter is further divided into *Tile Barrel* (Figure 3.15), for central $|\eta|$ and *Extended Barrel* for the most forward region. Is develops concentrically to the beam pipe and EM calorimeter, and is composed of steel absorber and scintillating plates (tiles) as active material. The tiles are displaced perpendicularly to the colliding beams, with segmentation $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$.
The *End-Cap* calorimeter covers the $1.5 < |\eta| < 3.2$ region and is located after the EM end-cap (Figure 3.15), in the longitudinal direction, in two wheels with copper absorber (arranged either in 25 mm plates or 50 mm) and LAr as active medium.

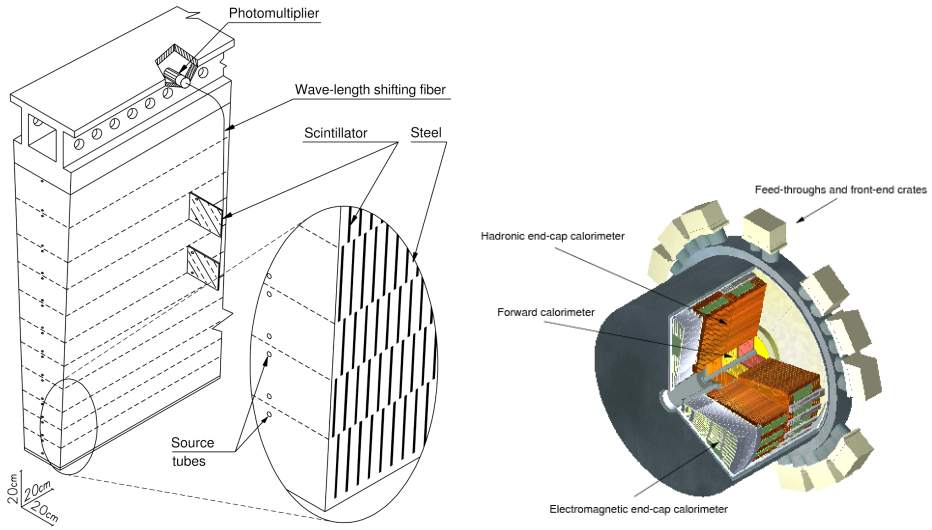


Figure 3.15: The ATLAS Tile Calorimeter layout (left) and a zoom on the end-cap region where the hadronic end-cap is located (right)

The global resolution is:

$$\frac{\sigma(E)}{E} \sim \frac{(50\%)}{\sqrt{E}} \oplus 3\% \tag{3.12}$$

The Forward Calorimeter

The Forward Calorimeter (FCal) closes the coverage at high pseudo-rapidity, $3.1 < |\eta| < 4.9$. The measurements of the showers in this region has a key role in forward jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. This component of the calorimetric system is located inside the HEC wheels, at a distance of about 5 m from the interaction point, and longitudinally

segmented into three sections (Figure 3.16). The employed active medium is still LAr, while the absorbers are copper, in the first longitudinal layer, and tungsten in the remaining two.
This calorimetric sub-system is the one with worse resolution:

$$\frac{\sigma(E)}{E} \sim \frac{(100\%)}{\sqrt{E}} \oplus 10\% \tag{3.13}$$

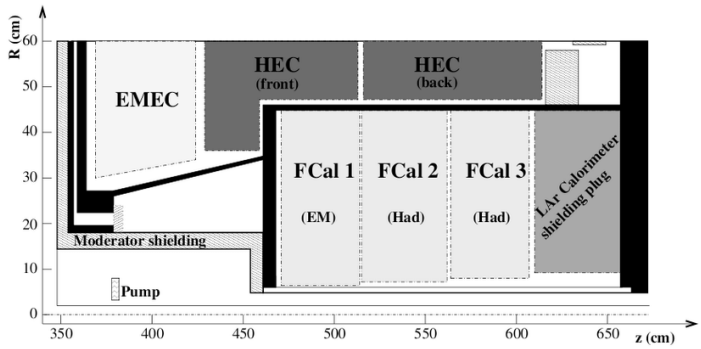


Figure 3.16: Layout of the ATLAS Forward Calorimeter.

3.2.5 The Muon Spectrometer

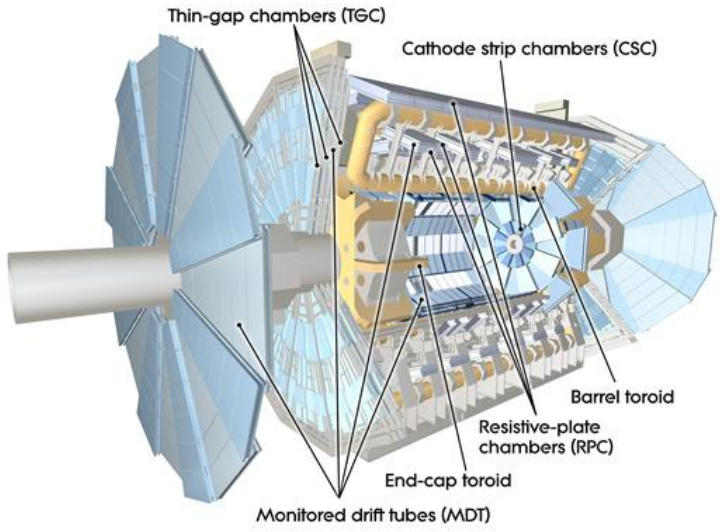


Figure 3.17: Sketch of the ATLAS Muon Spectrometer system [100].

The Muon Spectrometer (MS), represented in Figure 3.17, surrounds the other subdetectors, constituting the most outward section of the ATLAS detector, and is designed to trigger muons, which are expected to not be contained in the calorimeters, and measure

their momenta exploiting their deflection in the magnetic field provided by the toroidal magnets system. The system is divided into a barrel and two end-caps (Figure 3.18), with a total acceptance in pseudo-rapidity reaching up to $|\eta| < 2.7$, with full coverage, except for the $|\eta| < 0.1$.

The momentum measurement is performed by means of Monitored Drift Tubes (MTD)

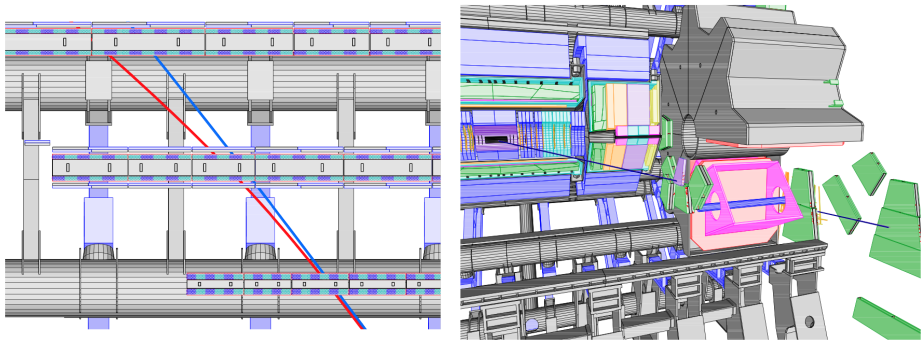


Figure 3.18: Sketch of the barrel (left) and end-cap (right) ATLAS Muon Spectrometer system.

in the barrel ($|\eta| < 2.7$) and Cathode Strip Chambers (CSC) in the forward region ($2 < |\eta| < 2.7$), arranged in three cylindrical layers around the beam axis in the barrel, and in three orthogonal layers to the beam in the end-caps. A resolution ranging between 2-3% and 10% for transverse momenta between 10 GeV and 1 TeV can be reached. The trigger system, with coverage limited to $|\eta| < 2.4$, includes Resistive Plate Chambers (RPC's) in the barrel and Thin Gap Chambers (TGC's) in the end-cap regions and is used to provide a fast muon trigger, recording and delivering track information within 15-25 ns. The main requirements are a good discrimination on muon transverse momentum, bunch-crossing identification, fast and coarse tracking information to be used in the high-level trigger stages, second coordinate measurement in the ϕ-projection to complement the MDT measurement, robustness towards random hits due to n/γ-background in the experimental hall.

The Monitored Drift Tubes

The Monitored Drift Chambers consist of 3-8 layers of drift tubes, with a diameter of about 30 mm, operating with a gas mixture of argon (Ar) and carbon dioxide (CO2) at 3 bar, and equipped with a central tungsten-rhenium wire with a diameter of 50 μm, at a potential of 3080 V.

This system can achieve a resolution of about 80 μm per tube and a total resolution 35 μm per chamber.

Cathode-strip chambers

Cathode-strip chambers (CSCs), with higher granularity, are used in the inner-most tracking layer of the forward region, where the limit of 150 Hz/cm^2 for safe operation of the MDTs is exceeded. The CSCs combine high spatial, time and double track resolution with high-rate capability and low neutron sensitivity.

They consist of multiwire proportional chambers with the wires oriented radially and the 2 cathodes segmented into strips arranged perpendicularly to the wires in one cathode, and parallel in the other, in order to reconstruct the track by an interpolation between

the charges induced on neighbouring cathode strips. The full system is composed of two disks with eight chambers each, each camber containing four CSC planes, thus providing four independent measurements in η and ϕ for each track.
This design provides a resolution of 60 μm in the bending plane and 5 mm in the transverse plane.

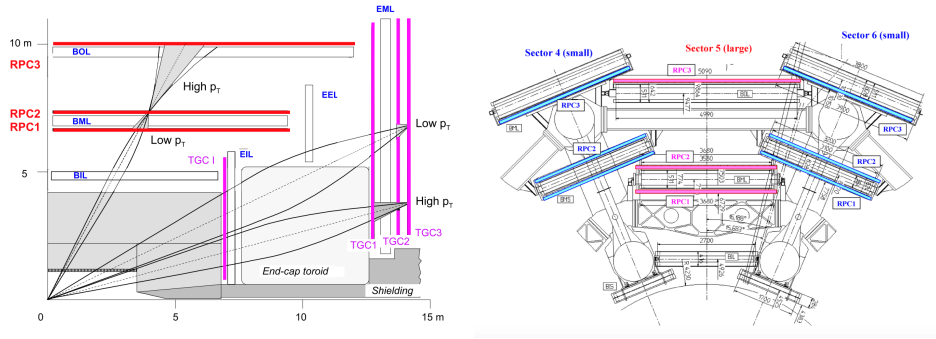


Figure 3.19: Sketch of the trigger system of ATLAS Muon Spectrometer (left) and of zoom on the arrangement of the Resistive Plate Chambers (right).

Resistive Plate Chambers

The Resistive Plate Chambers (RPCs), highlighted in red in Figure 3.19, cover the $|\eta| < 1.05$ region, to ensure good spatial and time resolution and an adequate rate capability. The system is divided into three "trigger stations", cylindrical layers concentrically located around the beam axis. Each station is further composed of two independent detector layers, each measuring both η and ϕ coordinate, thus resulting in a maximum of 6 track measurements.
Each RPC consists of a gaseous parallel electrode-plate detector, using a mixture of C2H2F4/Iso-C4H10/SF6. Two parallel resistive plates, of phenolic-melaminic plastic laminate, at a distance of 2 mm, provide an electric field of about 4.9 kV/mm, which allows avalanches to form along the ionising tracks towards the anode. The choice of operating the system in avalanche mode ensures an higher rate capacity than the streamer mode and a rate-independent time resolution.

Thin Gap Chambers

Finally, the Thin Gap Chambers (TGC) complete the trigger system in the $1.05 < |\eta| < 2.4$ pseudo-rapidity region, providing good time resolution and high rate capability. This system allows the determination of the second, azimuthal coordinate to complement the measurement of the MDTs in the radial direction, and it consists of multi-wire proportional chambers with the wire-to-cathode distance (1.4mm) smaller than the wire-to-wire distance (1.8 mm), using a highly quenching gas mixture of CO2 and n-C5H12 (n-pentane) Seven layers of TGCs complement the MDT middle layer in the end-cap, while only two TGC layers are associated to the inner MDT one. The inner layer is segmented radially into two non-overlapping regions: end-cap and forward.

3.3 Trigger

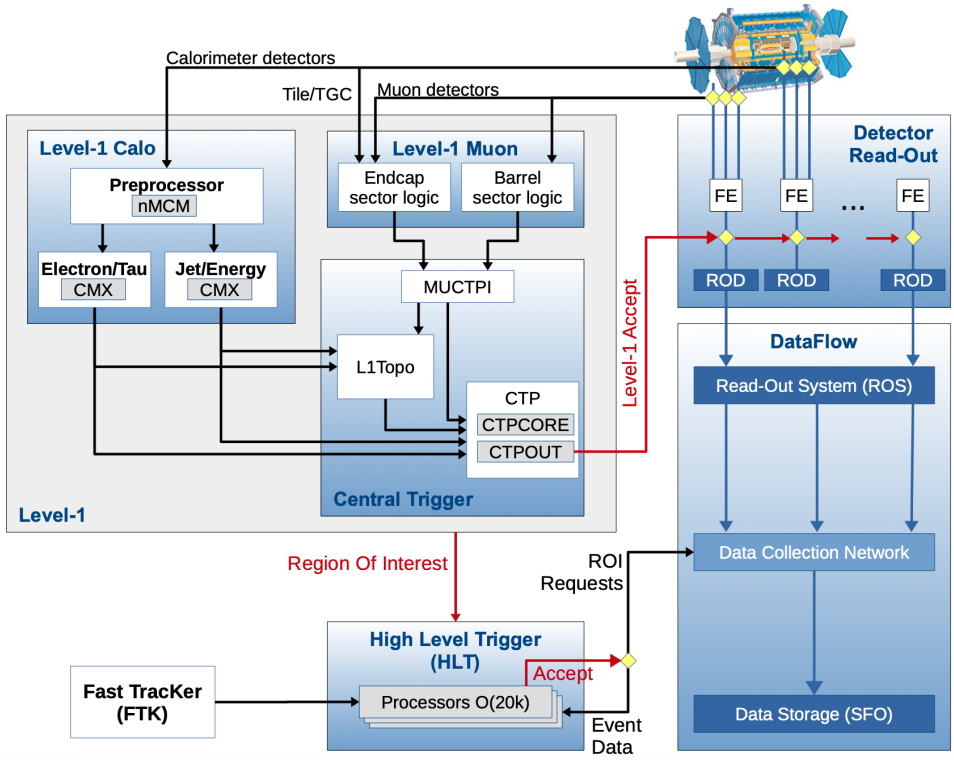


Figure 3.20: Sketch of the ATLAS Trigger and Data Acquisition system.

The ATLAS Trigger and Data Acquisition (TDAQ) system plays the critical role of handling the huge amount of data delivered by the LHC, which can not be fully acquired and processed, reducing the event rate from 40 MHz down to 1.2 KHz by means of fast online selections of events considered interesting for physics.
The trigger chain, schematically summarized in Figure 3.20, develops in two sub-systems, the Level 1 (L1) and High Level Trigger (HLT).

Level 1 Trigger

At the first stage, an hardware-based trigger system select the events using reduced-granularity information only from the calorimeters and the Muon Spectrometer, aiming at reducing the rate down to about 100 KHz, with a short response time of 2.5μs. The selection criteria are based on event-level quantities (e.g. the total energy in the calorimeter), the multiplicity of objects above thresholds (e.g. the transverse momentum of a muon, etc.), or topological requirements (such as invariant masses or angular distances). The L1 trigger also identifies, for each selected event, a Region Of Interest (ROI) in $\eta - \phi$ which will be investigated by the HLT.
The read-out chain starts with the transfer of data off-detector, to the ReadOut Drivers (RODs), where the initial processing and formatting is performed. Afterwards, data

are sent to the ReadOut System (ROS) where they are temporarily stored before being forwarded to the HLT, if requested.

High Level Trigger

The HLT trigger is a software-based system, which refines the selections performed at the previous stage by exploiting full-granularity information collected by all sub-detectors, in the ROI defined by the L1 trigger. Dedicated trigger algorithms are employed, with a decision time of about 200 ms, to further reduce the output rate to about 1.2 KHz.

Chapter 4

Event Reconstruction

The final state produced in the pp collisions can be reconstructed by combining the information provided by the signals released by the particles passing through the different subdetectors, and converted into electrical signals recorded by the data acquisition system (DAQ). Charged particles manifest themselves at first through *hits* in the layers of the inner detector, and subsequently release most of their energy in the calorimeters. The hits information can be exploited to derive precise directional information and track the particle down to the event interaction vertex. On the other side, neutral particles do not interact with the Inner Detector and release most of their energy in the calorimeters. The photons and electrons showers are typically contained within the electromagnetic one, while hadronic jets reach the hadronic calorimeters. Finally, muons are the only particles reaching and interacting with the Muon Spectrometer, while neutrinos are completely undetected. A sketch of the interactions and signals associated to different particles is shown in Figure 4.1.

4.1 Tracks and vertexes

The first foundamental ingredients for the reconstruction of hard scattering events are the reconstruction of the tracks from charged particles and the identification of the interaction vertex, which is mainly based on track information itself.
The main elements for the **track reconstruction** [101, 102] are the hits produced by charged particles in the ID. Primary tracks, produced in the hard-scattering vertex, are reconstructed with an inside-out algorithm, starting from a seed constituted by three hits in the silicon detector, primarily SCT. Compatible hits in the other layers of the ID are subsequently added with a Kalman filter algorithm [103]. Secondary tracks, arising from photon conversion or secondary vertices, are built starting from TRT segments not associated with primary tracks, with a backward extrapolation to the silicon detectors. Finally an ambiguity solving procedure is applied among tracks which share hits.
The **vertex reconstruction** [104] is based on a vertex finding algorithm which is able to identify the vertex position starting from track information, being robust enough against mis-associated or mis-measured tracks. Among the multiple vertices reconstructed for each collision, the primary vertex associated to the hard scattering interaction is identified as the one with highest sum of the squared transverse momenta of its associated tracks.

4.2 Electrons and photons

Electrons and photons reconstruction [105] is based on the combination of signals in the EM calorimeter and in the ID.

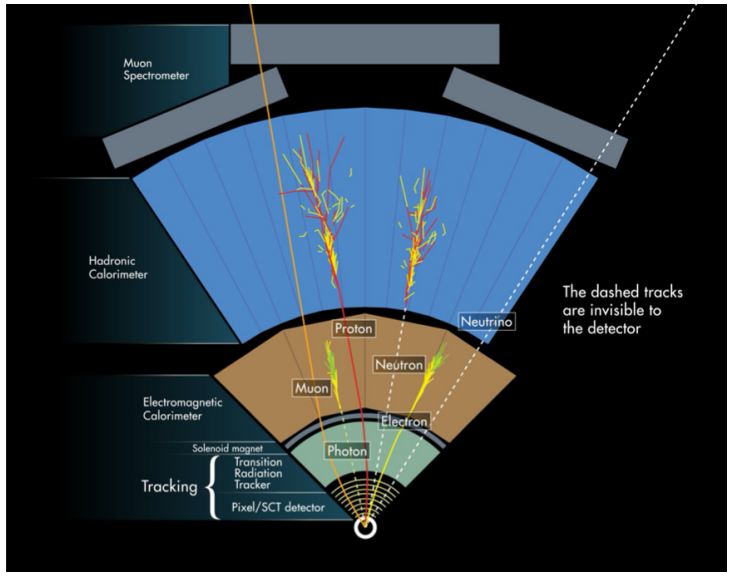


Figure 4.1: Sketch of a wedge of the ATLAS detector, with different subdetector and physics objects interactions highlighted. Dashed lines represent undetected particles in the considered subdetector.

Both these two particles produce an electromagnetic shower in the EM calorimeter, thus leading to an important interplay between the two reconstructions. Their quite similar shower profiles result in the need for additional information, in order to discriminate between the two. For this reason a matching between the energy deposit in the calorimeter and one or more tracks (if present) is performed: an electron produces one track in the ID, pointing to the primary vertex; on the other hand, a photon can either be associated to no track, or to a single or double track in the ID, pointing to a secondary vertex. The latter cases happen when the photon goes through a conversion into an e^+e^- pair before reaching the EM calorimeter. About 20% of photons at low $|\eta|$ convert in the ID, and up to about 65% convert at $|\eta| \geq 2.3$.
In order to discriminate electrons and photons from background, some Identification and Isolation Working Points (WP) are defined by applying appropriate selection criteria over variables related to geometrical and energetic properties of the shower. In this section, the reconstruction procedure, summarized in Figure 4.2, is presented, as well as the definition of the different Identification and Isolation WPs for both electrons and photons.

4.2.1 Electrons and photons reconstruction

Photon and electron candidates in $|\eta| < 2.5$ are at first identified through the construction of topological clusters (topo-clusters) of energy deposits in the EM calorimeter. The matchig of these clusters to ID tracks and conversion vertices, if present, is then performed. Afterwards, dynamic and variable-size clusters, named superclusters, are separately built for electrons and photons taking as input the defined topo-cluster. After initial position corrections and energy calibration, electron superclusters are associated to tracks, and photon superclusters to conversion vertices.
At the end of the reconstruction procedure, an electron will manifest itself as a cluster matched to an ID track, a converted photon as a cluster matched to a conversion vertex,

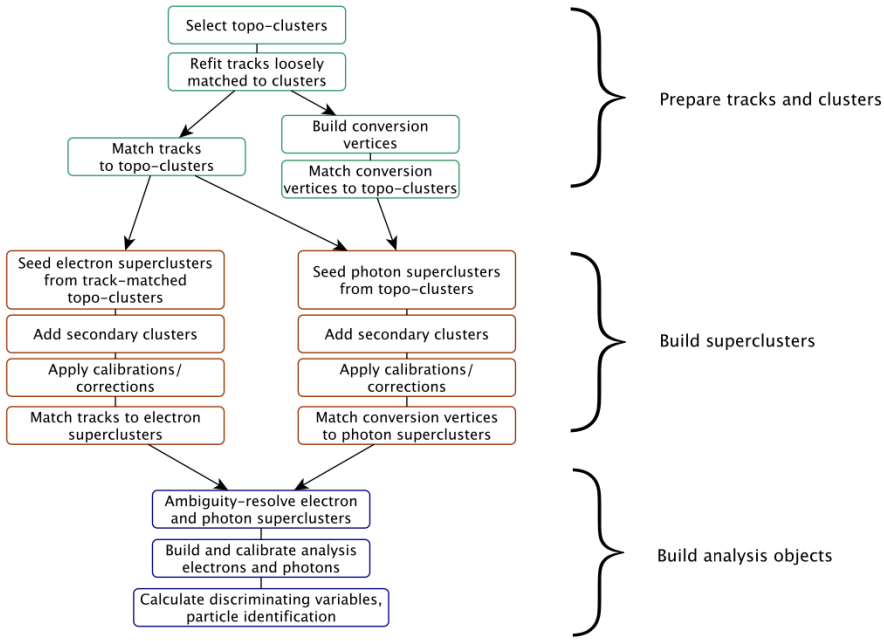


Figure 4.2: Diagram showing the algorithm flow for electrons and photons reconstruction.

while an unconverted photon as a cluster matched to neither an electron track nor a conversion vertex.
In the following, the various steps of electron/photon reconstruction and discrimination will be described in more details.

Topo-cluster building Topo-clusters [106, 107] consist of clusters of energy deposits measured in topologically connected EM and hadronic calorimeter cells. In order to avoid pile-up contribution from hadronic cells, these are removed from the topo-cluster if it is identified as originated from an electromagnetic shower. The identification criterion is $f_{EM} > 0.5$, where f_{EM} is the EM fraction, which allows a $\sim 60\%$ rejection of pile-up, without affecting electron and photon reconstruction efficiency. After hadronic cells removal, only topo-cluster with EM energy greater than 400MeV are retained.
The topo-cluster building consists of subsequent seed-and-collect steps which iteratively associate neighbour cells to a seed cluster following a signal-significance pattern [1]. The algorithm is based on the signal-significance observable, defined as:

$$\varsigma_{\text{cell}}^{\text{EM}} = \left| \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise,cell}}^{\text{EM}}} \right|$$

where $E_{\text{cell}}^{\text{EM}}$ is the cell energy at the EM scale and $\sigma_{\text{noise,cell}}^{\text{EM}}$ is the expected cell noise, including the known electronic noise and an estimate of the pile-up noise corresponding to the average instantaneous luminosity expected for Run-2. The absolute value is taken to cope with negative-energy cells due to calorimeter noise.

[1] Neighbour cells are the ones directly adjacent in a given sampling layer, or, if in adjacent layers, having at least partial overlap in the (η, ϕ) plane.

The first step is the identification of a seed cell with high signal significance, which defines the proto-cluster around which the topo-cluster will be built by adding neighbouring cells. In this initial stage, cells from the presampler and the first LAr EM calorimeter layer are excluded from initiating proto-clusters, to suppress the formation of noise clusters. The seed cell must satisfy the signal-significance condition $\varsigma^{\mathrm{EM}}_{\mathrm{cell}} > 4$. Neighbour cells are then added if $\varsigma^{\mathrm{EM}}_{\mathrm{cell}} > 2$, merging two topo-cluster if they share a cell. The topo-cluster is completed by the association of the nearest-neighbour cells to the ones added in the previous step. One proto-cluster can be merged with a neighbour one if they share a neighbour cell, or it can be split into separate clusters if it has two or more local maxima (cell with $E^{\mathrm{EM}}_{\mathrm{cell}} > 500$ MeV).

Matching to tracks and conversion vertices The following step consists in matching the built topo-clusters to tracks reconstructed in the ID tracker or to conversion vertices, if present.

A refined **track reconstruction** is performed [108]. Sets of three space-points in the silicon-detector layers form the track seeds, starting from which the tracks are built through a path-recognition algorithm that is performed in a Region-Of-Interest (ROI) defined by using fixed-size clusters in the calorimeter with a longitudinal and lateral shower profile compatible with that of an EM shower. If the standard algorithm for track recognition [101] fails to deduce the pattern within the ROI, a different algorithm, still based on Kalman filter formalism [103], is used, which allows for up to 30% of energy loss due to Brehmsstrahlung. A χ^2 fit [109] is performed for track candidates with $p_T > 400$MeV, and possible ambiguities from track candidates sharing hits are resolved. Finally, for tracks with at least 4 Silicon hits and loosely matched [2] to the fixed-size cluster, a Gaussian Sum Filter (GSF) algorithm [110] consisting in a non-linear generalization of the Kalman filter [3] is used in order to re-fit the tracks.

These re-fitted tracks are the ones given as input to the **matching** algorithm. One track is considered matched to a topo-cluster if it satisfies the requirements $\mid \eta_{\mathrm{cluster}} - \eta_{\mathrm{track}} \mid < 0.05$ and $-0.10 < q \times (\phi_{\mathrm{track}} - \phi_{\mathrm{cluster}}) < 0.05$ (where q is the particle charge). The $(\eta_{\mathrm{track}}, \phi_{\mathrm{track}})$ identifies the intersection of the extrapolated track with the second layer of the EM calorimeter and the extrapolation is performed using either the measured track momentum or, in order to take into account high bremsstrahlung energy loss, the momentum rescaled to match the cluster energy. If multiple tracks are matched to a cluster, tracks with hits in the pixel detector are preferred over tracks with hits only in the SCT. Moreover, preference is given to tracks with better ΔR match.

The **conversion vertex reconstruction** [111] takes as input both Silicon tracks and TRT tracks which are loosely matched to fixed-size clusters. Two-track conversion vertices are identified by two opposite-charge tracks associated to a vertex compatible with a massless particle, while single-track conversion vertices are given by tracks without hits in the innermost layers.

The **conversion vertex-matching** to the EM topo-cluster is then performed requiring $|\Delta\eta| < 0.05$ and $|\Delta\phi| < 0.05$ after extrapolation in case of Silicon tracks, while $|\Delta\eta| < 0.35(0.2)$ and $|\Delta\phi| < 0.02$ for TRT tracks with the first track in the barrel (end-

[2] tracks satisfying the criteria $|\eta_{\mathrm{cluster}} - \eta_{\mathrm{track}}| < 0.05$ and either $-0.20 < \Delta\phi < 0.05$ or $-0.10 < \Delta\phi_{\mathrm{res}} < 0.05$ where $\Delta\phi_{(res)}$ is the azimuthal separation between the track and the cluster, multiplied by the opposite of the charge sign and rescaled to the energy of the cluster for $\Delta\phi_{\mathrm{res}}$

[3] Within the GSF, experimental noise is modelled by a sum of Gaussian functions. The GSF therefore consists of a number of Kalman filters running in parallel, the result of which is that each track parameter is approximated by a weighted sum of Gaussian functions

cap) TRT. If multiple conversion vertices are matched to a single cluster, preference is given to the vertex with the smallest conversion radius. Moreover, two-track conversions are preferred over single-track ones and, among the double-track conversions, the ones with two Silicon tracks have priority.

Superclusters reconstruction The superclusters [112] are reconstructed separately for electrons and photons. Each EM topo-cluster becomes a supercluster seed if it satisfies the following requirements:

- Electron supercluster seed: the EM topo-cluster has $E_{\mathrm{T}} > 1\mathrm{GeV}$ and is matched to a track with at least four hits in the Silicon tracking detectors;
- Photon supercluster seed: the EM topo-cluster has $E_{\mathrm{T}} > 1.5\mathrm{GeV}$. No track or conversion vertex requirement is needed.

Once the supercluster seeds are identified, other clusters, likely associated to secondary shower from the same electron/photon, are added to them as satellite clusters according to the following criteria:

- Electron supercluster: if they fall within window $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ around the seed cluster barycentre or if they fall in the window $\Delta\eta \times \Delta\phi = 0.125 \times 0.300$ and share their "best-matched" track with the seed cluster;
- Photon supercluster: if they fall within a window $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ around the seed cluster barycentre or they are associated to conversion vertices made up only of tracks containing silicon hits and their best-matched (electron) track belongs to the conversion vertex matched to the seed cluster.

Figure 4.3 shows a scheme of the superclustering algorithm. Afterwards, the association of the calorimeter cells to the superclusters is performed, taking into account only cells from the presampler and the first three LAr calorimeter layers, and in addition, in the transition region $1.4 < |\eta| < 1.6$, the scintillator between the calorimeter cryostats.

Ambiguity resolution An initial energy calibration and position correction is applied to the supercluster before the matching to tracks and conversion vertex, which proceeds in a similar way than the matching to the EM topo-clusters. At this point, in cases in which the same seed cluster is associated to both a photon and an electron, the ambiguity is resolved through an algorithm described in Figure 4.4 and based on appropriate criteria over the tracks associated to the objects (such as the number of tracks, the number of hits in the different component of the Inner Detector, ...). At the end of this procedure, if it has been possible to effectively discriminate between photon or electron, only the selected one is retained. Otherwise, both are kept, labeled as "ambiguous", and their overlap is handled at analysis-level.

4.2.2 Energy calibration

The energy of the final electrons and photons must be accurately calibrated [113, 114]. The calibration proceeds through multiple steps, summarized in Figure 4.5.

- The energy of the electron or photon is derived from the energy deposits in the different layers of the calorimeter, accounting for upstream energy loss and lateral and longitudinal leakage. A single correction, including all the mentioned effects, is

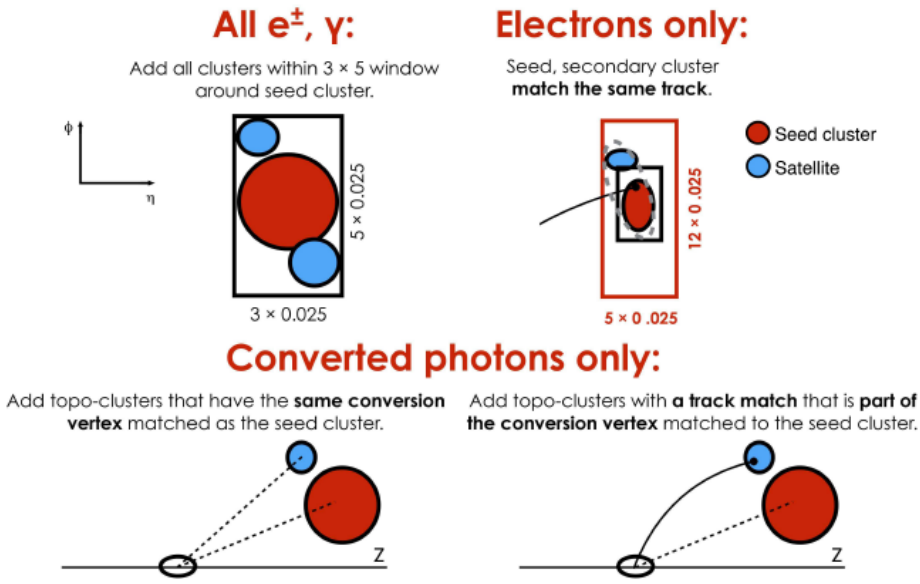


Figure 4.3: Illustration of the superclustering algorithm for electrons and photons. Seed clusters are shown in red, satellite clusters in blue.

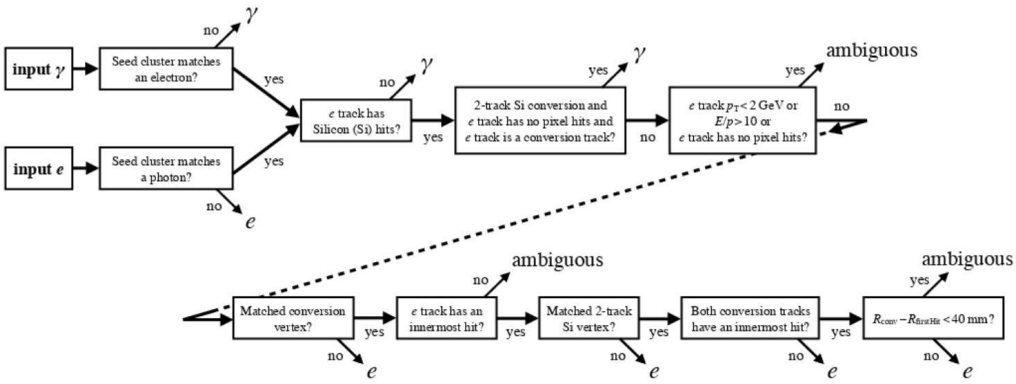


Figure 4.4: Electron/photon ambiguity resolution logic. An "innermost hit" is a hit in the pixel nearest to the beam-line along the track trajectory, E/p is the ratio of the energy in the supercluster to the matched track momentum, R_{conv} is the radial position of the conversion vertex, and R_{firstHit} is the smallest radial position of a hit in the track or tracks that make a conversion vertex [105].

provided by a multivariate regression algorithm trained through a Boosted Decision Tree strategy on samples of simulated single particles without pile-up, in intervals of $|\eta|$ and of transverse energy.

- Further corrections are applied to the data before the estimation of the electron or photon energy, to account for residual differences between data and simulation such as the relative energy scales of the different layers, presampler energy scale, correc-

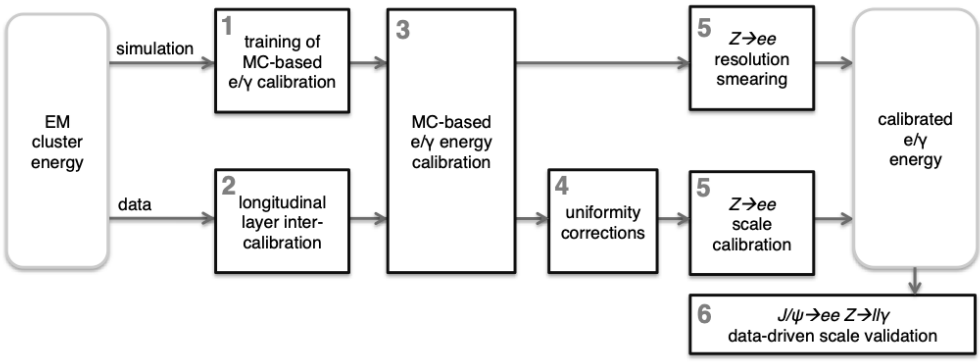


Figure 4.5: Electrons and photons calibration workflow applied to MC simulations and data

tions for energy shifts induced by pile-up and corrections of residual nonuniformities in the calorimeter response.

- The residual differences between data and simulations in energy scale and resolution are corrected through data-driven techniques using $Z \to ee$ samples and comparing the shapes of the dielectron invariant mass distributions in data and simulations in different (η, ϕ) regions. The corrections are provided in form of two factors, named α_i and c_i respectively for energy scale and resolution, such that:

$$E^{\text{data}} = E^{MC}(1 + \alpha_i)$$
$$\left(\frac{\sigma_E}{E}\right)^{\text{data}} = \left(\frac{\sigma_E}{E}\right)^{MC} \oplus c_i$$

  where $\oplus$ represents the sum in quadrature and i labels the η region. Pseudo-data samples generated from simulations are used to validate the corrections and estimate the possible residual bias, by comparing the α_i and c_i coefficients obtained from the comparison between data and pseudo-data, with the values used to generate the pseudo-data sample. The bias, which is assigned as an uncertainty, is typically (0.001–0.01)% for α_i and (0.01–0.03)% for c_i, depending on $|\eta$.

- The energy scale universality is validated using $J/\Psi \to ee$ and radiative Z decays. The first samples are used for a cross-check of the energy scale of low-energy electrons, by appliying the full energy scale calibration procedure to the MC samples and then comparing to data. The Z radiative decays are instead used as a cross-check for the energy scale of photons, especially in the low-energy region, with a separate study for converted and unconverted photons.

4.2.3 Energy scale and resolution systematic uncertainties

A set of 64 independent variations, some of which in multiple $|\eta|$ regions, describes the systematic error related to the electron and photon energy scale calibration, accounting for different sources of uncertainty arising from the $Z \to ee$ calibration procedure, Cell energy non-linearity, detector response, upstream energy loss or lateral/longitudinal leakage and conversion reconstruction.
In addition, systematic uncertainties also come from electron and photon energy resolution. A 10% uncertainty is assumed for the instrinsic energy resolution, i.e. the expected

resolution in the absence of upstream material and with uniform response, derived from the energy resolution in the simulation of genuine unconverted photons. Other sources of uncertainty arise from uncertainties in the impact of the detector material upstream of the calorimeter or from electronics and pile-up noise modelling. A systematic uncertainty on the constant term derived from data/MC comparison of the energy resolution for electrons from $Z \to ee$ decays is also added. The uncertainty ranges from about 5% to about 10% for electrons and photons with a transverse energy in the range $30 - 60$GeV, while at higher energy the relative uncertainty goes up to $20 - 50\%$.

4.2.4 Performance

A set of *Identification* and *Isolation* criteria [105] are applied to both photons and electrons, aiming at further discriminating between these particles and hadronic jets. Different Working Points (WP) are optimized, maximizing either the physics object acceptance or it's "quality". The choice of the WP to adopt depend on the needs of each analysis. For each identification or isolation WP, the efficiency is measured both for data and MC samples, and the discrepancies between the two measurements define the data to MC Scale Factors (SFs), which are applied at analysis-level as correction factors for MC simulations, with their uncertainty included as a systematic uncertainty.

Photon Identification

In order to discriminate between prompt, isolated photons and hadronic jets, *identification* quality criteria are applied through a cut-based selection over the shower-shape variables summarized in Table 4.1 and represented in Figure 4.6: narrower signal in the EM calorimeter with small energy fraction in the HCAL are more likely to be prompt photons. This additional selection goes under the name of Identification [36].
Three Working Points (WP) are defined, selecting objects with increasing quality: Loose, Medium and Tight. These WP differ not only in the cut value chosen, but also in the shower shape variable considered. For instance, Loose and Medium WP are mainly based on variables related to hadronic leakage (the fraction of energy released in the HCAL) or to the second layer of EM calorimeter [4], while Tight selections also take into account the first layer of the EM calorimeter, thus allowing to better reject possible background from π_0 decay by resolving the double energy maximum produced by this process in the L1. Moreover, while Loose and Medium WP are only η dependent, the Tight one also include a dependence on E_{T} and on the converted or unconverted nature of the photon.
The Loose photon identification reaches a 99% efficiency[111] for a photon with $p_T > 40$ GeV, with a background rejection factor of about 1000. The Tight selection ensures a background rejection factor of about 5000, with a reduced efficiency of 85% for a photon $p_T > 40$ GeV, increasing up to about 92% for unconverted photons and 95% for converted ones, with $p_T > 100$ GeV.
The identification efficiency, measured with 3 different methods [111], are shown in Figures 4.7 and for unconverted and converted photons respectively.

[4] The Medium WP also includes loose criteria on E_{ratio} .

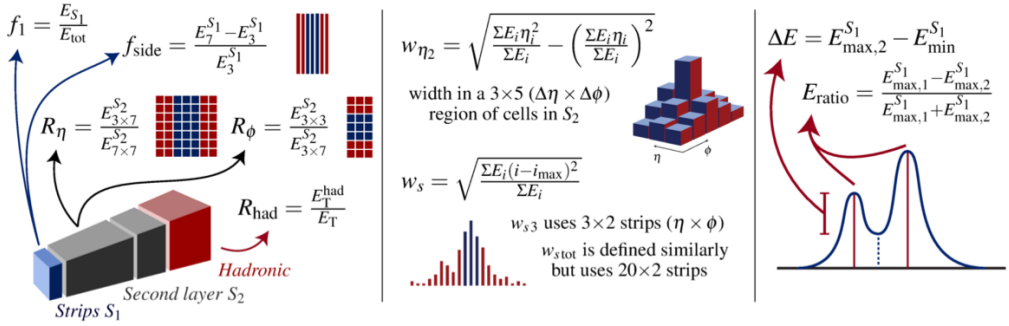


Figure 4.6: Photon identification discriminating variables, described in Table 4.1.

Category	Description	Name	Loose	Medium	Tight
Hadronic Leakage	Ratio of E_{T} in the 1st sampling layer of the hadronic calorimeter to E_{T} of the EM cluster (over the range $\lvert\eta\rvert < 0.8$ in $\lvert\eta > 1.37$)	R_{had_1}	✓	✓	✓
	Ratio of E_{T} in all the hadronic calorimeter samplings to E_{T} of the EM cluster (over the range $0.8 < \lvert\eta\rvert < 1.37$)	R_{had}	✓	✓	✓
EM 2^{th} layer	Ratio in η of cell energies in 3×3 cells over 3×7	R_η	✓	✓	✓
	Lateral width of the shower	w_{η_2}	✓	✓	✓
	Ratio in ϕ of cell energies in 3×3 over 3×7	R_ϕ			✓
EM 1^{th} layer	Shower width calculated from three strips around the strip with maximum energy deposit	w_{s_3}			✓
	Total lateral shower width	$w_{s_{tot}}$			✓
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips	F_{side}			✓
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima	ΔE			✓
	Ratio of the energy difference associated with the largest and second largest energy deposits to the sum of these energies	E_{ratio}			✓

Table 4.1: Photon shower shape variables and identification criteria definitions

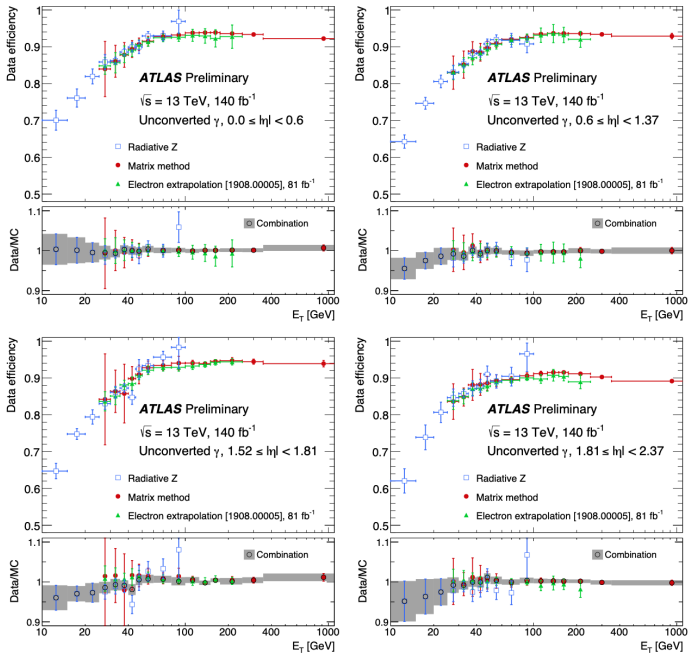


Figure 4.7: The photon Tight identification efficiency, and the ratio of data to MC efficiencies, for unconverted photons (top two rows) and converted photons (bottom two rows) with a Loose isolation requirement as preselection, as a function of E_T in different $|\eta|$ regions. The different measurement strategies are described in [105]. The combined scale factor, obtained using a weighted average of scale factors from the individual measurements, is also presented and the band represents the total uncertainty [115].

Photon Isolation

Requiring the photon to be highly isolated from surrounding activity, allows to further reject background from hadronic decays.
Two variables are used for photon Isolation [105], related to calorimetric measurements or to reconstructed tracks:

- E_T^{cone} : is the calorimeter isolation variable, defined as

$$E_T^{\text{cone}} = E_T - E_T^{\text{core}} - E_T^{\text{leakage}}\ (E_T, \eta, \Delta R) - E_T^{\text{pile-up}}\ (\eta, \Delta R)$$

  the sum of the transverse energies of the topoclusters in a cone ΔR (usually set to 0.4 or 0.2) around the direction of the photon candidate (E_T), removing the contribution from the photon candidate (E_T^{core}) by subtracting the energy deposit in a 5×7 window in the EM calorimeter around the particle barycentre. The last two terms of the equation account for leakage (i.e. the photon EM energy which is not contained in the subtracted window) corrections and pile-up and underlying-events ones.

- p_T^{cone} : is the track isolation variable, defined as the sum of the transverse momentum of selected tracks not associated to the photon, within a cone centred around the photon cluster direction. The tracks entering the calculation must satisfy some

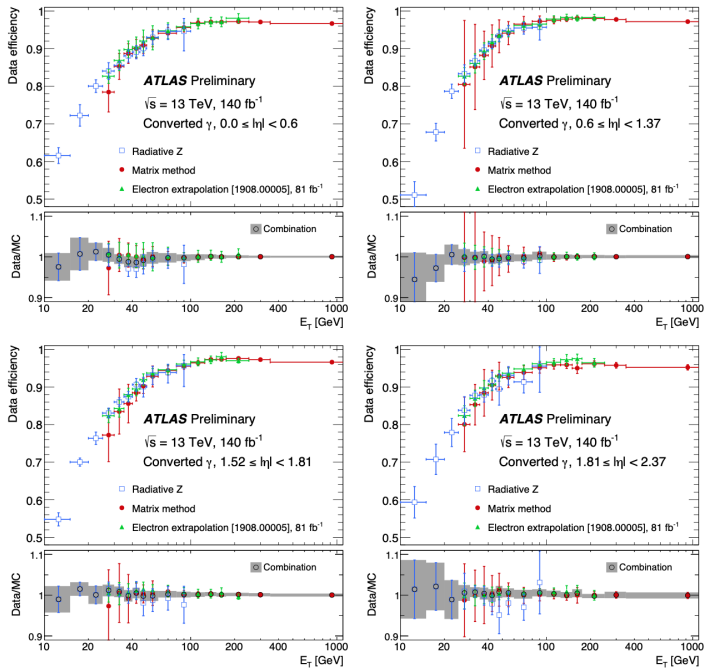


Figure 4.8: The photon Tight identification efficiencies, and the ratio of data to MC efficiencies, for unconverted photons (top two rows) and converted photons (bottom two rows) with a Loose isolation requirement as preselection, as a function of E_T in different $|\eta|$ regions. The different measurement strategies are described in [105]. The combined scale factor, obtained using a weighted average of scale factors from the individual measurements, is also presented and the band represents the total uncertainty [115].

requirements: they must have $p_T > 1$GeV and $|\eta| < 2.5$, they must have at least seven silicon (Pixel + SCT) hits, at most one shared hit, two silicon holes (i.e. missing hits in the pixel and SCT detectors) and one pixel hole. Finally it's required that $|\Delta z_0| \sin\theta < 3$ mm where $|\Delta z_0|$ is the longitudinal impact parameter with respect to the primary vertex.

The recommended working points are summarized in table 4.2, and the measured photon isolation efficiencies in Figure 4.9, together with data to MC SFs.

Working point	**Calorimeter**	**Track**
Loose	$E_{\mathrm{T}}^{cone20} < 0.65 \times E_{\mathrm{T}}$	$p_{\mathrm{T}}^{cone20}/E_{\mathrm{T}} < 0.05$
Tight	$E_{\mathrm{T}}^{cone40} < 0.022 \times E_{\mathrm{T}} + 2.54$ GeV	$p_{\mathrm{T}}^{cone20}/E_{\mathrm{T}} < 0.05$
TightCaloOnly	$E_{\mathrm{T}}^{cone40} < 0.022 \times E_{\mathrm{T}} + 2.54$ GeV	-

Table 4.2: Photon isolation WPs.

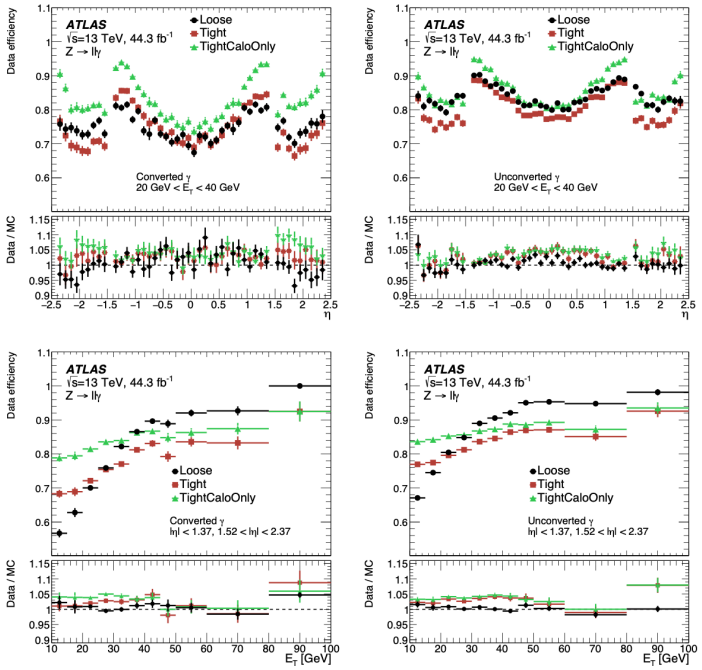


Figure 4.9: Photon isolation efficiency for different WPs, using $Z \to \ell\ell\gamma$ events, for converted (left) and unconverted (right) photons as a function of photon $|\eta|$ (top) and E_{T} (bottom). The lower panel shows the ratio of the efficiencies measured in data and in simulation. The total uncertainty is shown, including the statistical and systematic components [105]

4.2.5 Electron Identification

Electron Identification [108] allows to discriminate between prompt, isolated electrons and energy deposits from hadronic jets, converted photons or electrons produced in the decays of heavy-flavour hadrons.
Differently from the photon Identification, electron Identification is based on selection criteria applied to a likelihood discriminant, which collects information from several discriminant variables, summarized in Table 4.3 related to the primary electron track, the lateral and longitudinal development of the electromagnetic shower in the EM calorimeter, and the spatial compatibility of the primary electron track with the reconstructed cluster.
The WPs Loose, Medium and Tight are all $|\eta|$ and E_{T} dependent, and optimized in order to get a predefined Identification selection.
Some further requirements are applied in addition to the likelihood discriminant selection, in order to better reject background from converted photons: electrons with a two-track silicon conversion vertex reconstructed with a momentum closer to the cluster energy than that of the primary electron track are rejected, and $E/p < 10$ and $p_{\mathrm{T}} > 2$GeV is required for the Tight WP.
The identification efficiencies are reported in Figure 4.10, as well as the comparison between data and MC from which the SFs are derived.

Category	Description	Name
Hadronic Leakage	Ratio of E_{T} in the 1st sampling layer of the hadronic calorimeter to E_{T} of the EM cluster (over the range $\lvert\eta\rvert < 0.8$ ir $\lvert\eta > 1.37$)	R_{had_1}
	Ratio of E_{T} in all the hadronic calorimeter samplings to E_{T} of the EM cluster (over the range $0.8 < \lvert\eta < 1.37$)	R_{had}
EM 3^{th} layer	Ratio of the energy in the back layer to the total energy in the EM accordion. This variable is only used below 100 GeV because it is known to be inefficient at high energies	f_3
EM 2^{th} layer	Ratio in η of cell energies in 3×7 cells over 7×7	R_η
	Lateral width of the shower	w_{η_2}
	Ratio in ϕ of cell energies in 3×3 over 3×7	R_ϕ
EM 1^{th} layer	Total lateral shower width	$w_{s_{tot}}$
	Ratio of the energy difference associated with the largest and second largest energy deposits to the sum of these energies	E_{ratio}
	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter	f_1
Track conditions	Number of hits in the innermost pixel layer; discriminates against photon conversions	n_{IBL}
	Number of hits in the pixel detector	n_{Pixel}
	Number of total hits in the pixel and SCT detectors	n_{Si}
	Transverse impact parameter with respect to the beam-line	d_0
	Significance of transverse impact parameter defined as the ratio of d_0 and its uncertainty	d_0/σ_{d_0}
	Momentum lost by the track between the perigee and the last measurement point divided by the original momentum	$\Delta p/p$
TRT	Likelihood probability based on transition radiation in the TRT	eProbabilityHT
Track-cluster matching	$\Delta\eta$ between the cluster position in the middle layer and the track extrapolated from the perigee	$\Delta\eta_1$
	$\Delta\phi$ between the cluster position in the strip layer and the extrapolated track	$\Delta\phi_2$
	Defined as $\Delta\phi_2$ but the track momentum is rescaled to the cluster energy before extrapolating the track from the perigee to the middle layer of the calorimeter	$\Delta\phi_r ed$
	Ratio of the cluster energy to the track momentum	E/p

Table 4.3: Electron shower shape variables entering the likelihood that defines the identification WPs

4.2.6 Electron Isolation

Electron Isolation [105] proceeds in a similar way as for photons, with some differences in the track isolation variable definition.
A variable cone size ($p_{\mathrm{T}}^{\mathrm{varcone}}$) is used for the electron, in order to take into account the possible presence of other decay products in case of electrons coming from the decay of a

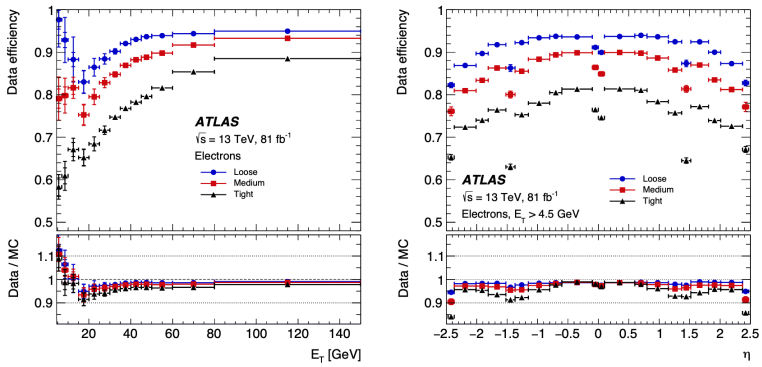


Figure 4.10: The electron identification efficiency for different WP, in $Z \to ee$ events in data as a function of E_{T}(left) and as a function of η (right). The efficiencies are obtained by applying data-to-simulation efficiency ratios measured in $J/\Psi \to ee$ and $Z \to ee$ events to $Z \to ee$ simulation. The inner uncertainties are statistical and the total uncertainties are the statistical and systematic uncertainties in the data-to-simulation efficiency ratio added in quadrature. For both plots, the bottom panel shows the data-to-simulation ratios.

high-momentum heavy particle. The difference lies in the ΔR definition:

$$\Delta R = \min\left(\frac{10}{p_{\mathrm{T}}[\mathrm{GeV}]}, \Delta R_{\mathrm{max}}\right)$$

where ΔR_{max} is the chosen maximum size of the cone (usually set to 0.2 for electrons). Moreover the tracks are required to have a loose vertex association, i.e. the track was used in the primary vertex fit, or it was not used in any vertex fit but satisfies $|\Delta z_0| \sin\theta < 3$ mm.

The used Working Points for electron isolation are reported in Table 4.4, while the measured efficiencies and data to MC SFs are shown in Figure 4.11

Working point	**Calorimeter**	**Track**
Gradient	$\epsilon = 0.1143 \times p_{\mathrm{T}} + 92.14\%$ (with E_{T}^{cone20})	$\epsilon = 0.1143 \times p_{\mathrm{T}} + 92.14\%$ (with $p_{\mathrm{T}}^{varcone20}$)
Loose	$E_{\mathrm{T}}^{cone20}/p_{\mathrm{T}} < 0.20$	$p_{\mathrm{T}}^{varcone20}/p_{\mathrm{T}} < 0.15$
Tight	$E_{\mathrm{T}}^{cone20}/p_{\mathrm{T}} < 0.06$	$p_{\mathrm{T}}^{varcone20}/p_{\mathrm{T}} < 0.06$
HighPtCaloOnly	$E_{\mathrm{T}}^{cone20} < \max(0.015 \times p_{\mathrm{T}}, 3.5\ \mathrm{GeV})$	-

Table 4.4: Electron isolation WPs

4.3 Muons

Muon reconstruction [116] is strongly based on signals released in the most external part of the detector, the MS, due to the fact that these are the most penetrating particles, after neutrinos, and are not contained within the electromagnetic or the hadronic calorimeters. Tracks in the ID, associated to the MS signals are also used to improve directional measurements and association to the primary vertex, as well as calorimetric measurements to account for muon energy loss in the calorimeters.

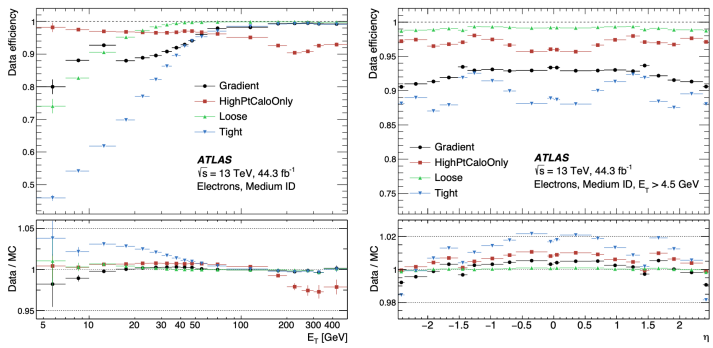


Figure 4.11: Efficiency of the different isolation WPs for electrons from inclusive $Z \to ee$ events as a function of the electron E_{T} (left) and η (right). The electrons are required to fulfil the Medium selection from the likelihood-based electron identification. The lower panel shows the ratio of the efficiencies measured in data and in MC simulations. The total uncertainties are shown, including the statistical and systematic components [105].

The first step of reconstruction consists in the search for hit patterns in the different muon chambers, to build track segments. The track segments are identified by means of the Hough transform pattern recognition algorithm, which searches for hits aligned on a trajectory in the bending plane. In addition, the coordinate orthogonal to the bending plane is measured by the trigger chambers (RPC and TGC). Afterwards, all these informations are combined to create a 3D track candidate, starting from a seed in the middle layers and merging track segments inwards and outwards. As a final step, a global χ^2 fit is performed, taking into account possible interactions with the detector material and misalignments between the different chambers. Ambiguities are resolved by removing tracks that share a large fraction of hits with higher-quality tracks, with the exception of tracks that are identical in two stations but share no hits in a third station, to ensure a high efficiency for boosted low-mass dimuon systems. After ambiguities removal, a new fit is done with a loose IP constraint and taking into account the energy loss in the calorimeters.
Once the tracks in the Muon Spectrometer are reconstructed, they can be combined with ID tracks or calorimetric signals to provide the final muon candidate. Depending on which and how sub-detector information are employed in the reconstruction, 5 reconstruction strategies can be identified:

Combined (CB) Muons Informations from all detectors are exploited for this reconstruction. MS and ID tracks are combined with a track fit and energy losses in the calorimeters are taken into account. After the fit, an update of the MS hits might be done, possibly adding or removing hits from the track, before performing a new combined track fit. In the $|\eta| > 2.5$ region, MS tracks may be combined with short track segments reconstructed from hits in the pixel and SCT detectors, forming the so-called silicon-associated forward (SiF) muons.

Inside-Out (IO) Muons A complementary inside-out algorithm is exploited, which extrapolates ID tracks to the MS and searches for at least three MS hits with loose alignment. A combined track fit is then performed using the ID track, the energy loss in the calorimeters and the MS hits. This reconstruction allows an efficiency improvement

for example in regions with limited MS coverage, as it does not rely on independently-reconstructed MS tracks.

Extrapolated and Standalone (ME) muons This strategy is employed in case the matching between the MS track and ID tracks fails. In this case, the MS track parameters are extrapolated to the beamline to define a ME muon. This reconstruction is used to cover regions outside the ID acceptance, thus reaching the $|\eta| < 2.7$ coverage of the MS.

Segment-Tagged (ST) muons In this case, the muon parameters are taken directly from the ID track fit, but a tight matching between ID tracks extrapolated to the MS and at least one reconstructed MS segment is required.

Calorimeter-Tagged (CT) muons This strategy is based on ID tracks and energy losses in the calorimeters. ID tracks are extrapolated to the calorimeters, to search for energy deposits consistent with a minimum-ionising particle. If such energy deposits are found, the ID track is tagged as a muon, and muon parameters are taken from the ID track fit. In this case, a larger p_T threshold of 5 GeV (instead of 2 GeV) is used for tracks to avoid large background contamination.

4.3.1 Identification

The muon identification is performed by means of set of selection requirements on the number of hits in the different ID subdetectors and different MS stations, on the track fit properties, and on variables that test the compatibility of the individual measurements in the two detector systems. Tighter or looser criteria are applied depending on the adopted Working Point (WP). Similarly to electrons and photons, efficiency measurements are performed, providing data to MC SFs and relative systematic uncertainties.

Medium WP This is the default WP, selecting CB and IO muons within the ID acceptance of $|\eta| < 2.5$, with at least two precision stations, defined as the number of MS stations in which the muon has at least three hits in the MDT or CSC detectors. For $|\eta| < 0.1$ muons, muons with only one precision station are accepted, given that they have at most one precision hole station, defined as a station where the muon has less than three hits and is missing at least three hits that are expected given its trajectory and the detector layout and operational status. Another selection criterion is based on the q/p compatibility, a variable which quantifies the compatibility between the ratio of muon charge and momentum as measured from ID tracks and from MS ones. This variable is required to be less than seven. Moreover, the acceptance is extended in the range $2.5 < |\eta| < 2.7$ by including ME and SiF muons with hits in at least three precision stations.

Loose WP The Loose WP aims at increasing the efficiency in the $|\eta| < 0.1$ range, where the CB muon efficiency is affected by the gap in the MS coverage, and for low-p_T muons. In addition to muons passing the Medium one, this WP accepts CT and ST muons in the $|\eta| < 0.1$ range. Moreover, IO muons with $p_T < 7$ GeV and only one precision station are accepted in the $|\eta| < 1.3$ range, if they are also reconstructed as ST muons. The efficiency increase of the Loose WP compared to Medium is around 20% for 3 GeV $< p_T < 5$ GeV and approximately 1–2% for higher p_T.

Tight WP This is the tightest WP, accepting CB and IO muons with at least two precision stations among the ones passing Medium WP. The combined track fit is required to have normalised χ^2 less than 8, to reject pathological tracks due to hadron decays in flight. In order to provide better background rejection for lower-p_T muons, further requirements are optimized, depending on p_T and η, on q/p compatibility and ρ', the latter being defined as the absolute difference between the ID and MS p_T measurements divided by the p_T of the combined track. For the region 6 GeV $< p_T <$20 GeV, the Tight WP achieves a background reduction of more than 50% compared to Medium, with a corresponding efficiency loss for prompt muons of approximately 6%.

High-p_T WP This WP is designed to improve the resolution of selected high-p_T muons, taking into account that, for these muons, the limiting factors are the intrinsic detector resolution of the individual measurements along the track and the knowledge of the relative alignment between the corresponding detector element. The same requirements as for Medium WP are applied. In addition, at least three precision stations are required. As an exception, muons traversing the B-field inversion zones instrumented with additional chambers, must have at least four precision stations, while muons with 2 precision stations can be accepted in the $|\eta| < 1.3$ region, if the missing hits are in the inner station. In addition, muons in regions where the relative alignment is not known with sufficient precision, i.e. in the barrel–endcap overlap region $1.0 < |\eta| < 1.1$, and partially in $1.1 < |\eta| < 1.3$ and $|\eta| < 1$ regions corresponding to the detector support structures, are rejected.

Low-p_T WP In this WP, addressing muons with $p_T < 18$ GeV, CB and IO muons are included, with a largest fraction from IO due to the fact that very low p_T muons will not reach the middle station of the MS, or even the MS itself. The IO muons are required to be also reconstructed as ST ones. At least one precision station is required in general, but 2 precision stations are required in the $|\eta| > 1.3$ region, where muons with $p_T > 1.3$ can typically reach the second station. Further selection requirements are imposed to reject light-hadron decays, exploiting the distinctive *kink* observed for light-hadron decays along the trajectory in the ID, due to undetected neutrino. Both a cut-based and a Boosted Decision Tree (BDT) approaches are used, defining two variants of the low-p_T WP.

A summary of the efficiencies for different identification WPs, and the data/MC comparison, is shown in Figure 4.12

4.3.2 Isolation

The muon isolation makes use of similar variable as the ones defined for electrons and photons, with the addition of a Particle Flow [5] based variable, the $E_T^{nflow20}$, defined as the sum of track-based isolation ($p_T^{varcone30}$ for $p_T < 50$ GeV and p_T^{cone20} for $p_T > 50$ GeV) and the transverse energy of neutral particle-flow objects in a cone of size $\Delta R = 0.2$ around the muon. The different adopted WPs are summarized in Table 4.5, and their efficiencies and SF, with associated systematic uncertainties, are shown in Figure 4.13 for the Loose and Tight WPs, taken as example.

[5] See Section 4.4 for an overview of the Particle Flow strategy

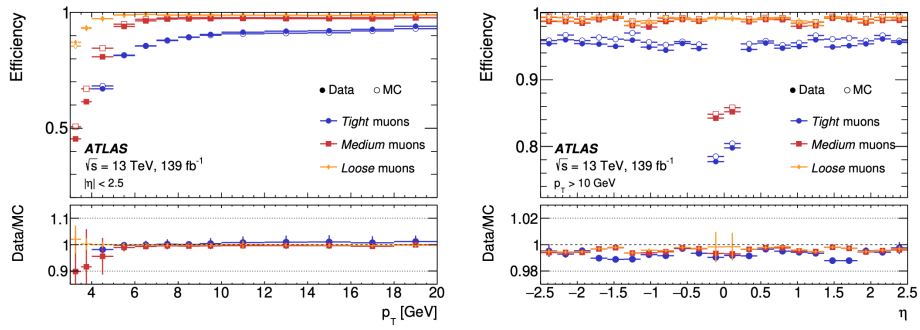


Figure 4.12: Muon reconstruction and identification efficiencies for the Loose, Medium, and Tight criteria. The left plot shows the efficiencies measured in $J/\Psi \to \mu\mu$ events as function of p_{T}. The right plot displays the efficiencies measured in $z \to \mu\mu$ events as a function of η, for muons with $p_{\mathrm{T}} > 10$ GeV. The predicted efficiencies are depicted as open markers, while filled markers illustrate the result of the measurement in collision data. The statistical uncertainty in the efficiency measurement is smaller than the size of the markers, and thus not displayed. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties [116].

Isolation WP	**Definition**	**Track p_T requirement**
PFlowLoose	$p_T^{varcone30} + 0.4 \cdot E_T^{neflow20} < 0.16 \cdot p_T^{\mu}$	$p_T > 500$ MeV
PFlowTight	$p_T^{varcone30} + 0.4 \cdot E_T^{neflow20} < 0.045 \cdot p_T^{\mu}$	
Loose	$p_T^{varcone30} < 0.15 \cdot p_T^{\mu}, E_T^{topoetcone20} < 0.3 \cdot p_T^{\mu}$	$p_T > 1$ GeV
Tight	$p_T^{varcone30} < 0.04 \cdot p_T^{\mu}, E_T^{topoetcone20} < 0.15 \cdot p_T^{\mu}$	
HighPtTrackOnly	$p_T^{cone20} < 1.25$ GeV	$p_T > 1$ GeV
TightTrackOnly	$p_T^{varcone30} < 0.06 \cdot p_T^{\mu}$	
PLBDTLoose (Tight)	$p_T^{varcone30} < max(1.8\ \mathrm{GeV}, 0.15 \cdot p_T^{\mu})$ BDT cut to mimic *TightTrackOnly(Tight)* efficiency	$p_T > 1$ GeV

Table 4.5: Definition of different muon isolation WPs

4.4 Jets

Quarks and gluons produced in final states in high energy pp collisions can't be directly detected because of the colour confinement of the strong interaction resulting in the formation of colorless hadrons, through the *fragmentation* and *hadronization* processes.
Fragmentation gives rise to the developement of a *parton-shower*, represented in Figure 2.10 as simulated by a Monte Carlo generator [117]. The quarks and gluons resulting from the fragmentation process recombine to form a collimated jet of hadrons (hadronization) with a total momentum pointing approximately in the same direction as the initial parton.
The signature of jets in the ATLAS detector comprises tracks in the ID and energy deposits in the calorimeters produced by the hadrons resulting from hadronization. The hadronic shower is typically wider than the electromagnetic one produced by electrons and photons and, despite including an EM fraction, it develops mainly in the Hadronic

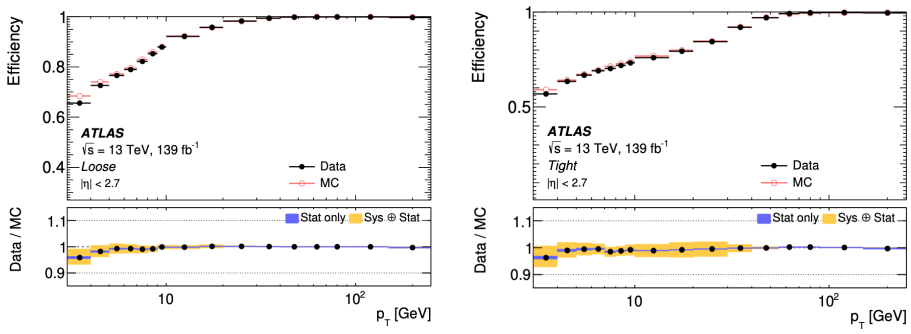


Figure 4.13: Muon isolation efficiency measured in $Z \to \mu\mu$ events for the Loose (left plot) and Tight (right plot) WP, as a function of p_{T}. The statistical uncertainty in the efficiency measurement is smaller than the size of the markers, and thus not displayed. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties [116].

calorimeter (HCAL). It is therefore possible to discriminate between jets and electrons or photons through the already mentioned Identification and Isolation criteria.

During Run-1 and the start of Run-2, the jets were reconstructed using the so-called EMTopo reconstruction algorithm, based exclusively on energy deposits in the electromagnetic and hadronic calorimeters, with track information used only for origin corrections. For full Run-2 analyses, the reconstruction strategy has been improved through the implementation of the Particle Flow (PFlow) strategy, which combines both calorimetric and track informations. In the analyses described in this thesis, both the two reconstruction have been employed: EMTopo jets are used in the mono-photon analysis (chapter 8 and 10), while the PFlow ones in the Dark Photon analysis (chapter 9).
The reconstruction workflows for the two methods are summarized in Figure 4.15: the main difference, other than the inputs given to the *jet-finding* algorithm, consists in the *origin correction* step, needed only in the EMTopo algorithm. In the following the Particle Flow Objects used in the PFlow jet reconstruction will be described, followed by the jet-finding algorithm for EMTopo and the remaining pile-up correction and calibration steps, common to both EMTopo and PFlow reconstructions.

4.4.1 Particle Flow Objects

The Particle Flow Objects (PFOs) which are given as input to PFlow jet-finding are built from a combination of calorimetric and track informations. The algorithm [118], illustrated in Figure 4.16, develops in different steps, to match each track to one or more topo-clusters and perform a cell-based energy subtraction to avoid double counting of energy contributions.

Track selection Stringent quality criteria are applied to select tracks that will feed PFlow jets. At least 9 hits are required in the silicon detectors, and no missing Pixel hit. Only tracks with $p_{\mathrm{T}} > 0.5$ GeV and $|\eta| < 2.5$ are included, without any requirement of association to the hard scatter vertex. Tracks with high transverse momentum ($p_{\mathrm{T}} > 40$ GeV) are typically not well isolated from nearby activity, with negative impact on the accuracy of the subsequent energy subtraction step. To overcome this issue, no subtraction is performed for $p^{trk} > 100$ GeV, and for all tracks up to $p^{trk} = 100$ GeV where the energy

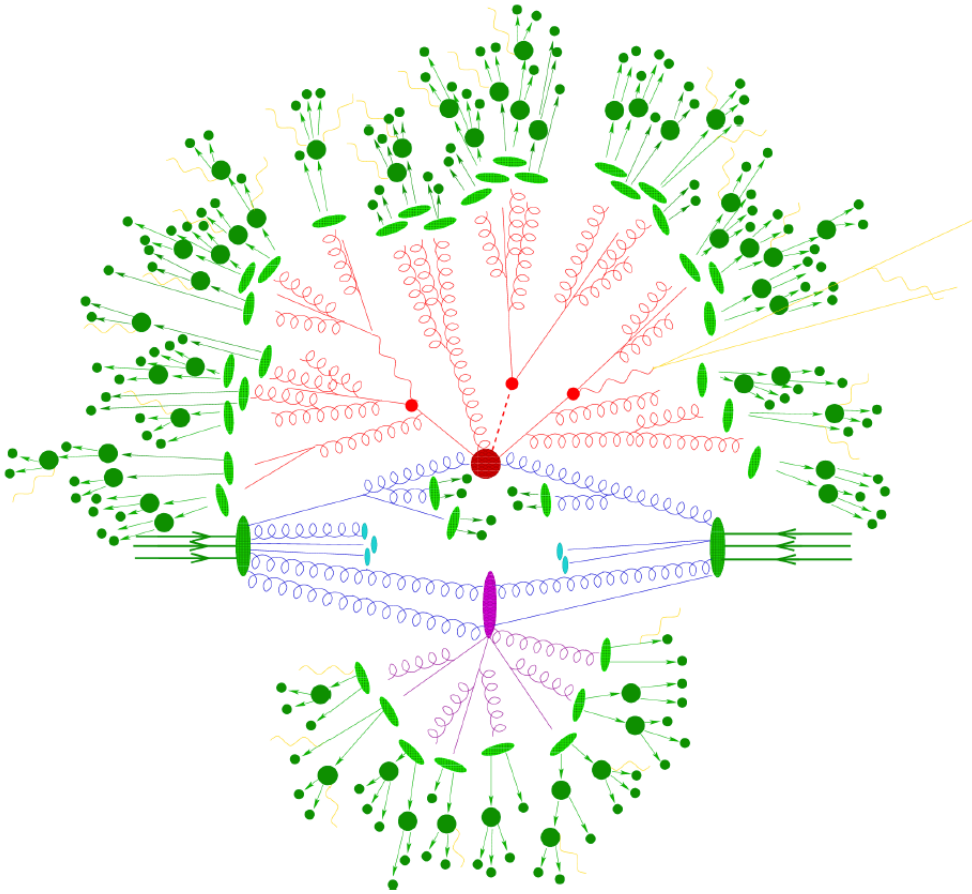


Figure 4.14: Sketch of a hadron-hadron collision as simulated by a Monte-Carlo event generator. The red blob in the center represents the hard collision, surrounded by a tree-like structure representing QCD radiation as simulated by parton showers. The purple blob indicates a secondary hard scattering event. Parton-to-hadron transitions are represented by light green blobs, dark green blobs indicate hadron decays, while yellow lines signal soft photon radiation [117].

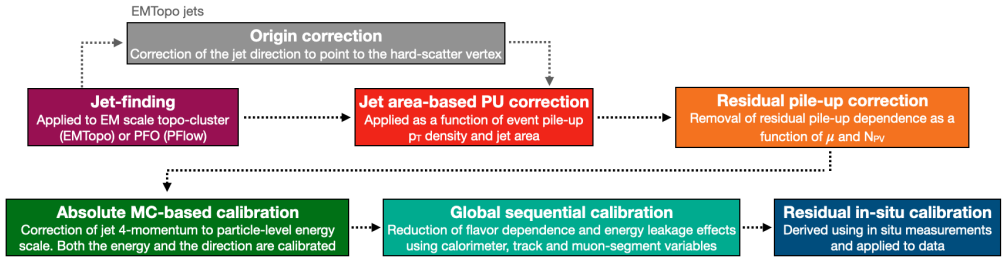


Figure 4.15: Overview of the reconstruction stages for PFlow and EMTopo jets. Each step of the calibration corrects the full four-momentum of the jet, with the exception of the origin corrections for EMTopo jet.

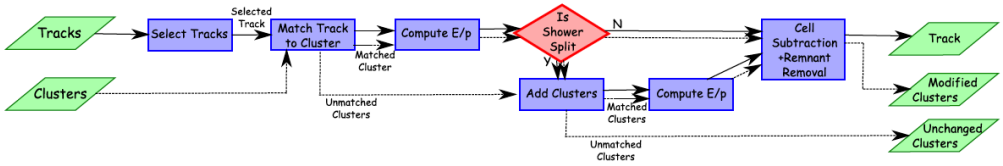


Figure 4.16: PFlow algorithm workflow[118].

E^{clus} in a $\Delta R = 0.14$ cone around the extrapolated particle satisfies

$$\frac{E^{clus} - \langle E_{dep} \rangle}{\sigma \langle E_{dep} \rangle} > 33.2 \times \log_{10}(40 \text{ GeV}/p_T^{trk}) \tag{4.1}$$

In addition, being the algorithm optimized for hadronic showers, tracks matched to medium quality electrons or muons, with no isolation requirement, are masked. These selection criteria were optimized on $\sqrt{s} = 8$ TeV collisions, with a pile-up condition similar to 2012 data taking, and were shown to ensure a track reconstruction efficiency of about 90% (increasing with the trak p_{T}) in the most central region ($|\eta^{true}| < 1.0$) and an efficiency ranging from $> 70\%$ to $\sim 80\%$ for the $2.0 < |\eta^{true}| < 2.5$ region.

Track-topocluster matching The matching between the selected tracks and the topo-clusters is based on the angular diastances between the barycentre of the topo-cluster and the track, extrapolated to the second layer of the EM calorimeter. The ranking variable is

$$\Delta R' = \sqrt{\left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta\eta}{\sigma_\eta}\right)^2} \tag{4.2}$$

where $\sigma_{\eta/\phi}$ denote the angular topo-cluster widths defined as the the standard deviation of the displacements of the topo-cluster constituent $\eta - \phi$ cells with respect to the barycentre. This variable has an excellent discrimination power between the correctly matched topo-cluster and neighbour ones, as shown in the top plots of Figure 4.17 showing a comparison of $\Delta R'$ distributions for the topo-cluster containing 90% of the particle true energy and the other ones. The topo-cluster to be matched to the considered track are preselected by requiring an high ratio between the topo-cluster energy and the track momentum, $E^{clus}/p^{trk} > 0.1$, which allows a good discrimination between "correct" matches and neighbour topo-clusters for particles with $p_{\mathrm{T}} > 5$ GeV, reaching a $40 - 50\%$ rejection of incorrect clusters, while for softer particles in the $1 < p_{\mathrm{T}} < 2$ GeV range the rejection is reduced to about 10%, as shown in the central plots of Figure 4.17. The matching is therefore performed taking the lowest $\Delta R'$ (up to 1.64) preselected topo-cluster.

Expected deposited particle energy In order to correctly subtract the track energy from associated topo-clusters, an accurate estimation of the expected energy deposited on average in calorimeters by a particle with track momentum p^{trk} is foundamental. This is quantified as

$$\langle E_{dep} \rangle = p^{trk} \langle E_{ref}^{clus}/p_{ref}^{trk} \rangle \tag{4.3}$$

where the ref subscript indicates the values determined using single-pion samples without pile-up, by summing the energies of topo-clusters in a ΔR cone of size 0.4 around the track position, extrapolated to the second layer of the EM calorimeter.

Recover split showers It is possible that a shower is not fully contained into one, but splitted into more topo-clusters. In this case, the track should be associated to all of them. In order to identify splitted showers, the significance of the difference between the expected energy and that of the matched topo-cluster is used as discriminant variable:

$$S(E^{clus}) = \frac{E^{clus} - \langle E_{dep} \rangle}{\sigma(E_{dep})} \tag{4.4}$$

By comparing the distribution of this variables for topo-clusters containing more than 90% or less than 70% of the total shower energy (bottom plots of Figure 4.17), a threshold at $S(E^{clus}) = -1$ has been optimized: if $S(E^{clus}) < -1$, topo-cluster within a $\Delta R = 0.2$ cone around the track position extrapolated to the second EM calorimeter layer are considered matched to the track.

Energy subtraction The final step consists in the subtraction of the track energy from the matched topo-cluster(s). If the $\langle E_{dep} \rangle$ (eq. 4.3) exceeds the total energy of matched topo-clusters, these are completely removed. Otherwise, a cell-by-cell subtraction algorithm is employed, based on a step-by-step removal of cells within increasing radius rings, proceeding until the energy contained in the ring exceeds the expected deposited energy, thus being only partially subtracted. An illustration of this process is given in Figure 4.18, together with a visualization of the effect of the subtraction in the second layer of the EM calorimeter.

This procedure provides a list of neutral PFOs, given by topo-cluster after track energy subtraction, and charged PFOs consisting in the tracks.

4.4.2 Jet-finding

The jet finding is performed by means of the so-called anti-k_t algorithm [119], which ensures that the final jets are "infra-red and collinear safe", as required to obtain a finite prediction at all perturbative orders. The algorithm is similar for EMTopo and PFlow jets, but fed by different inputs. The topological clusters at the LC (Local Cell) scale are used for the first strategy, with track information providing an a posteriori correction, while the latter uses the charged and neutral PFOs, with an additional preselection on charged PFOs impact parameter $|z_0 \sin\theta| < 2$ mm (where z_0 is the distance of closest approach of the track to the hard-scatter primary vertex along the z-axis), aiming at reducing pile-up contributions.
The anti-k_t algorithm is based on the two distance parameters:

$$d_{i,j} = \min\left(\frac{1}{p^2_{\mathrm{T,i}}}, \frac{1}{p^2_{\mathrm{T,j}}}\right)\frac{\Delta^2_{ij}}{R^2}$$
$$d_{iB} = \frac{1}{p^2_{\mathrm{T,i}}}$$

where $\Delta^2_{ij} = \left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2$ and R is the distance parameter that defines the maximum radius of the cone, commonly set to $R = 0.4$. For each pair of input particles, if $d_{ij} < d_{iB}$ the objects i and j are combined, their combination is added as input for the next iteration while the single objects are removed. Otherwise, the object i is considered as a jet and removed from the input list. The algorithm proceeds iteratively and ends when all the objects have been combined in jets.

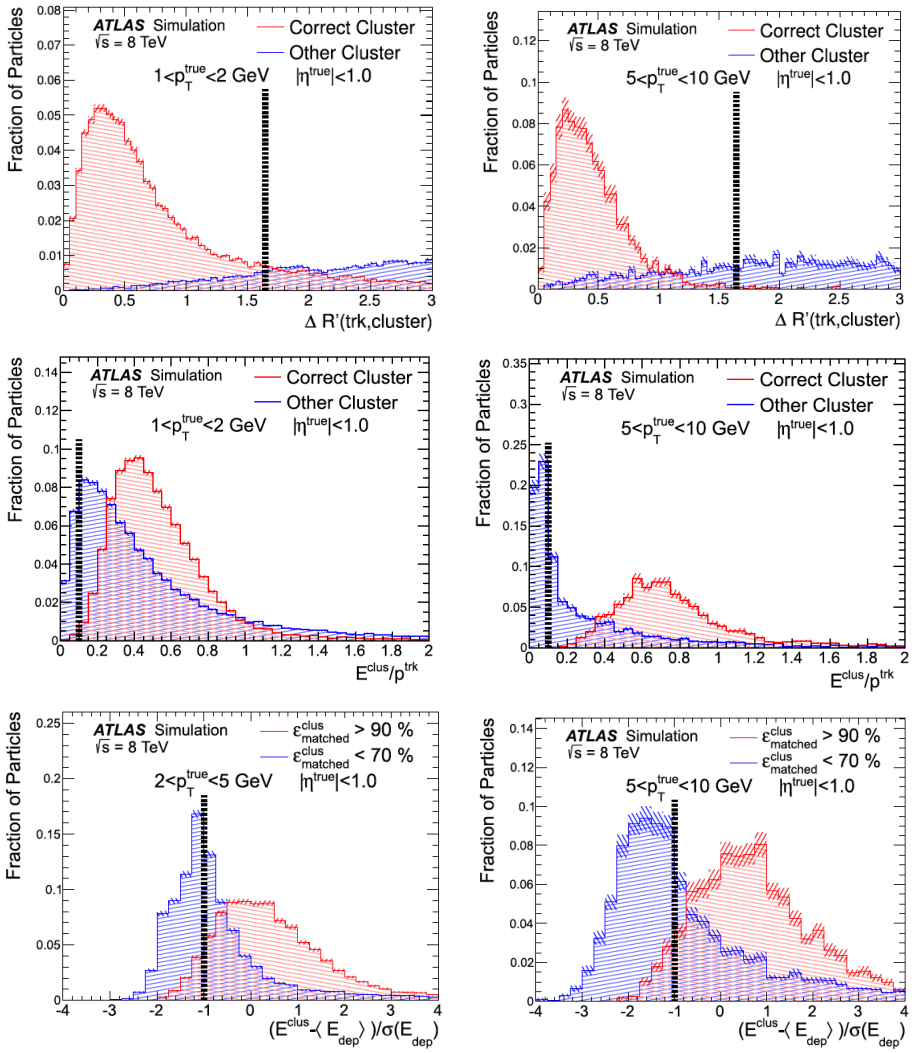


Figure 4.17: Comparison of relevant distributions for the topo-cluster containing $> 90\%$ of the true energy of the particle (correct cluster) and the closest other topo-cluster in $\Delta R'$ (other cluster), for two p_{T} and $|\eta|$ bins taken as an example. The variables shown are the $\Delta R'$ (top), E^{clus}/p^{trk} (middle) and $\frac{E^{clus}-\langle E_{dep}\rangle}{\sigma(E_{dep})}$ (bottom).

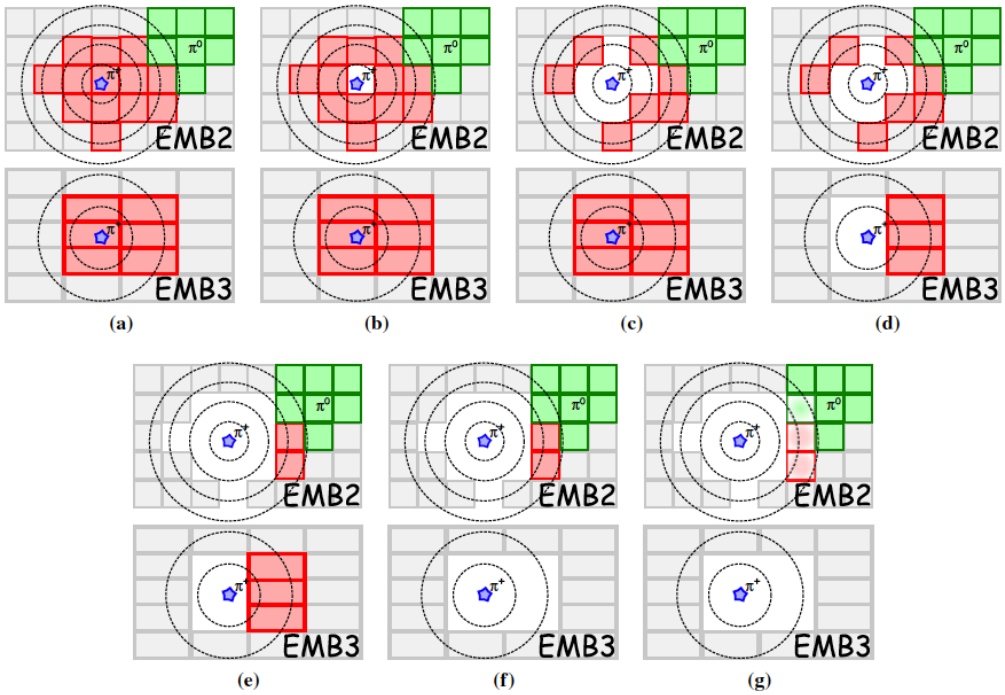


Figure 4.18: An idealised example of how the cell-by-cell subtraction works. Cells in two adjacent calorimeter layers (EMB2 and EMB3) are shown in grey if they are not in clusters, red and green if they belong to a π^+ cluster and a pi^0 meson respectively. The stars represent the extrapolated tracks, and the steps of the ring-based removal procedure are illustrated in the different subfigures. The final ring contains more energy than the expected energy, hence this is only partially subtracted (g), indicated by a lighter shading

In the EMTopo reconstruction, a further step, referred to as the *origin correction* is needed to correct the jet direction, through the recalculation of the four-momentum such that it points to the hard-scatter primary vertex rather than the center of the detector.

4.4.3 Jet Energy Calibration

At this point, the jets are calibrated at the EM energy scale and need to be recalibrated in order to restore the jet energy scale to that of *truth* jets, i.e. jets reconstructed at the particle-level, using stable final-state particles from MC generators, without accounting for detector effects. The calibration develops similarly for EMTopo [120] jets (applied in the $20 < p_{\mathrm{T}} < 2000$ GeV range) and PFlow jets [118] (applied only in the $20 < p_{\mathrm{T}} < 1500$ GeV range), in the following steps:

- The Pile-up corrections [121] allow the removal of the energy contribution coming from pile-up events. It is performed in two steps.
  The first correction is the jet-area pile-up subtraction [122] , which subtracts the pileup contribution to the p_{T} of each jet according to its area. The jets used for the pile-up subtraction are reconstructed by applying the k_t-algorithm [6] (sensitive to soft radiation) with R = 0.4 to positive-energy topo-clusters, which are required to have $|\eta| < 2$ because of the higher calorimeter occupancy in forward regions. The pile-up contribution is calculated from the median p_{T} density of jets in the $\eta - \phi$ plane, defined as $\rho = p_{\mathrm{T}}/A$, where A is the area of the jet estimated through ghost association [7]. The correction factor, applied to the four-momentum of the jet, is derived as the ratio of the pile-up-subtracted jet p_{T} to the uncorrected one. In the PFlow jet case, given that in-time pile-up is largely removed by the matching of charged PFOs to the Primary Vertex, this reduction mainly addresses charged Underlying Event hadrons and particles from out-of-time interactions, as well as neutral pile-up contributions.
  This correction has the limit of a non accurate description of the pile-up in the forward or higher-occupancy regions of the calorimeter. An additional MC-based correction [121] is therefore required in order to correct the residual dependence of the reconstructed jet p_{T} on the number of primary vertex (related to in-time pileup) and the average number of interactions per crossing μ (related to out-of-time pile-up). These effects are quantified by the difference between the truth-jets p_{T} and the p_{T} of the reconstructed jets with $p_{\mathrm{T}} > 10$GeV, matched to the truth jets within $\Delta R = 0.3$.
  The final pile-up-corrected p_{T} of the jet is given by:

$$p_{\mathrm{T}}^{\mathrm{corrected}} = p_{\mathrm{T}}^{\mathrm{reco}} - \rho \times A - \alpha \times (N_{PV} - 1) - \beta \times \mu$$

  where α and β are the coefficients parametrizing the almost linear dependence of the jet p_{T} respectively on the N_{PV} and on μ.
- The MC-based absolute jet energy scale (JES) and η calibration [120, 123] corrects the reconstructed jet four-momentum to the particle-level energy scale and accounts for biases in the jet η reconstruction, mainly due to transitions between different calorimeter technologies and sudden changes in calorimeter granularity.

[6] The kt algorithm is similar to the anti-kt, but defining $d_{i,j} = \min\left(p_{\mathrm{T,i}}^2, p_{\mathrm{T,j}}^2\right)\frac{\Delta_{ij}^2}{R^2}$ and $d_{iB} = p_{\mathrm{T,i}}^2$

[7] Simulated ghost particles of infinitesimal momentum are added uniformly in solid angle to the event before jet reconstruction. The area of a jet is then measured from the relative number of ghost particles associated with a jet after clustering.

For the JES calibration, isolated reconstructed jets [8] are matched to truth jets within $\Delta R = 0.3$. The calibration factor is derived, through a numerical inversion procedure, as the inverse of the energy response, defined as the mean of the Gaussian fit to the core of the $E^{\text{reco}}/E^{\text{truth}}$ distribution. This distribution is binned in E^{truth} and η_{det}, i.e. the jet η pointing to the geometric center of the detector. The η_{det} binning allows to remove possible ambiguities related to the region of the detector which is measuring the jet.
The mentioned biases in η produce an increase in the energy of one side of the jet with respect to the other, altering the reconstructed four-momentum. For this reason an η correction is needed and is derived as the difference between the reconstructed η^{reco} and the truth η^{truth}, parametrized as a function of E^{truth} and η^{det}. Afterwards, corrections to E^{reco} are derived again by repeating the numerical inversion procedure. Jets calibrated with the full jet energy scale and the η calibration are considered to be at the EM+JES scale.

- Global Sequential energy scale calibration (GSC) [124] allows to correct residual dependencies on longitudinal and transverse properties of the jets, due to fluctuations in the jet particle composition and in the energy distribution within the jet. In fact, the average particle composition and shower shape of a jet depend on whether the initiating particle is a quark or a gluon. The GSC is based on the following observables:
  - f_{charged} the fraction of the jet energy measured from constituent tracks (charged fraction), i.e. tracks associated to the jet, applied only to PFlow jets;
  - $f_{\text{Tile 0}}$ the fraction of jet energy measured in the first layer of the hadronic Tile calorimeter ($|\eta_{\text{det}}| < 1.7$);
  - f_{LAr3} the fraction of jet energy measured in the third layer of the electromagnetic LAr calorimeter ($|\eta_{\text{det}}| < 3.5$);
  - n_{trk} the number of tracks with $p_{\text{T}} > 1$GeV ghost-associated with the jet ($|\eta_{\text{det}}| < 2.5$);
  - W_{trk} the average p_{T}-weighted transverse distance in the $\eta - \phi$ plane between the jet axis and all tracks of $p_{\text{T}} > 1$GeV ghost-associated to the jet ($|\eta_{\text{det}}| < 2.5$);
  - n_{segments} the number of muon track segments ghost-associated with the jet ($|\eta_{\text{det}}| < 32.7$)

  For each observable, an independent correction of the jet four-momentum is evaluated and applied sequentially as a function of $p_{\text{T}}^{\text{truth}}$ and $|\eta_{\text{det}}|$ through numerical inversion of the jet energy response. The correlations between the observables are neglected.
- Finally η-intercalibration and in-situ calibration [120] are performed, accounting for differences in the jet response between data and MC simulation due to imperfect description of the detector response and of the physics of the event in MC simulation. The differences are quantified by looking at jets balanced against other well-measured reference objects, in different stages:
  - the η-intercalibration corrects the average response of forward jets ($0.8 < |\eta| < 4.5$) to that of well-measured central jets ($|\eta| < 0.8$) using dijet events with $p_{\text{T}}^{jet} > 25$ GeV;

[8] No other calorimeter jet of $p_{\text{T}} > 7$GeV within $\Delta R = 0.6$, and only one truth jet of $p_{\text{T}}^{\text{truth}} > 7$GeV within $\Delta R = 1.0$.

- the Z+jet and γ+jet analyses [125] balance the hadronic recoil in an event against the p_{T} of a calibrated Z boson or photon. The Z+jet calibration is limited to the statistically significant p_{T} range of Z boson production of $20 < p_{\mathrm{T}} < 500$ GeV, while the γ+jet calibration is limited by the small number of events at high p_{T} and by both dijet contamination and the artificial reduction of the number of events due to the prescaled triggers at low p_{T}, limiting the calibration to $36 < p_{\mathrm{T}} < 950$ GeV. The $E_{\mathrm{T}}^{\mathrm{miss}}$ Projection Fraction (MPF) method is used, based on the full hadronic recoil to compute the balance to help mitigate effects of pile-up and jet reconstruction thresholds which otherwise make low-p_{T} measurements challenging. The corrections are provided in bins of p_{T};
- the multi-jet balance (MJB) analysis [126] uses a system of well-calibrated low-p_{T} jets to calibrate a single high-p_{T} jet, covering the $300 < p_{\mathrm{T}} < 2000$ GeV region.

For each in-situ estimate, the correction is derived from numerical inversion of the ratio $R_{in-situ}^{data}/R_{in-situ}^{MC}$ where $R_{in-situ}$ is the response, defined as the average ratio of jet p_{T} to reference object p_{T} (or η), binned in regions of the reference object p_{T} (η).
The data-to-MC ratio and the associated systematic uncertainties derived from the orthogonal Z+jet, γ+ jet, and multi-jet calibrations are combined across overlapping regions of jet p_{T}.

Jet Energy Scale systematic uncertainties

The calibration procedure results in a set of 125 Jet Energy Scale (JES) systematic uncertainty terms [120] from the individual calibrations.
The largest contribution comes from Z+ jet, γ+ jet, and multi-jet in-situ calibrations, with systematic sources related to the assumptions in the event topology, MC simulation, sample statistics, and propagated uncertainties of the electron, muon, and photon energy scales.

In Figure 4.19 combination of all uncertainties is shown as a function of jet p_{T} at $\eta = 0$ and η at $p_{\mathrm{T}} = 60$GeV, assuming a flavor composition taken from the inclusive dijet selection in Pythia, for both EMTopo and PFlow jets as well as a direct comparison of the two. The different contributions are considered fully correlated in p_{T} and η, but uncorrelated with one another, with the exceptions of the electron and photon energy scale measurements, which are also treated as fully correlated. In the left plot, the sharp increase in the uncertainties for $p_{\mathrm{T}} > 2$ TeV is due to the fact that the in-situ methods can't reach these momenta, and the uncertainty is evaluated from single-jet response studies.

4.4.4 Jet Energy Resolution and systematic uncertainty

Precise knowledge of the jet energy resolution (JER) [118] is important for detailed measurements of SM jet production, measurements and studies of the properties of the SM particles that decay to jets (e.g. W/Z bosons, top quarks), as well as searches for physics beyond the SM involving jets. The JER also affects the missing transverse momentum, which plays an indispensable role in many searches for new physics and measurements

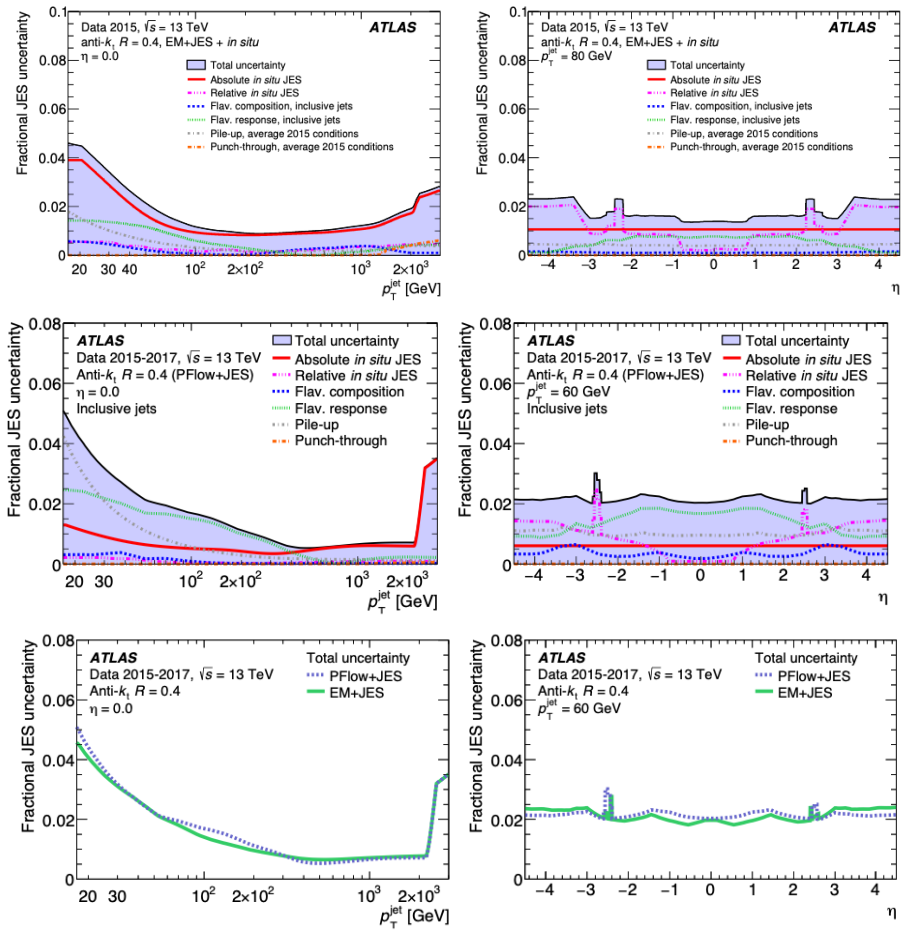


Figure 4.19: Combined uncertainty in the JES of fully calibrated jets as a function of jet p_{T} at $\eta = 0$ (left) and η at $p_{\mathrm{T}} = 80$ GeV (right), for EMTopo jets [120](top), PFlow jets (center) and a comparison of the two [118] (bottom). Absolute in-situ JES indicates Z/γ+ jet and multijet calibrations, while relative in situ JES the η-intercalibrations. The flavor composition and response uncertainties assume a quark and gluon composition taken from Pythia dijet MC simulation (inclusive jets)

involving particles that decay into neutrinos, and thus rely on well-reconstructed missing transverse momentum.

The JER can be parametrized with three terms, each characterized by different dependence on the jet p_{T}:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C \tag{4.5}$$

where the first term is the electronic and pile-up *noise*, expected to be dominant at $p_{\mathrm{T}} < 30$ GeV, the S term is the *stochastic* contribution, dominant for higher p_{T} jets and up to hundreds of GeV, and the last term is the *constant* term, originated from p_{T} independent fluctuations such as energy deposited in passive material, the starting point of the hadronic shower or non-uniform responses across the calorimeter. The last term is expected to be dominant for jet $p_{\mathrm{T}} > 400$ GeV.

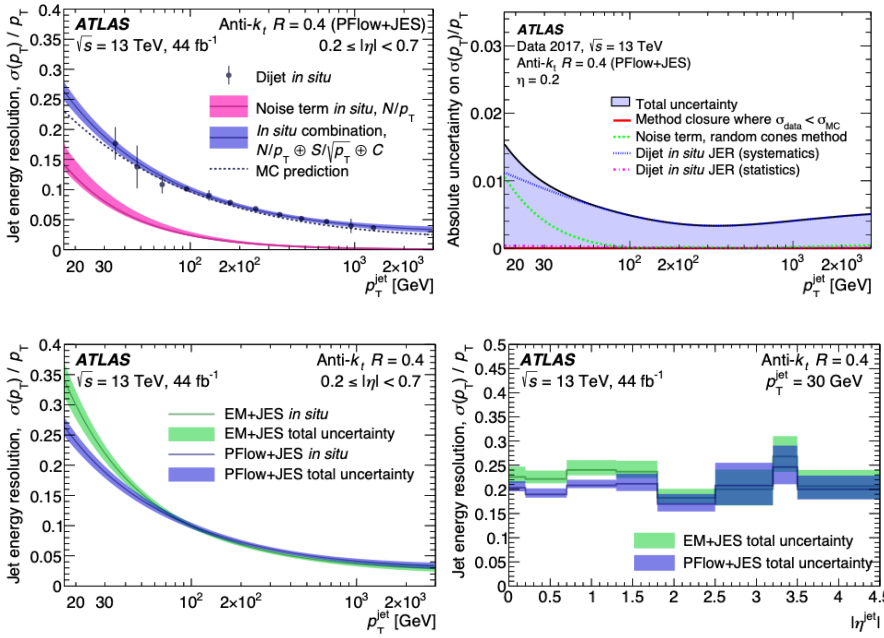


Figure 4.20: Jet Energy Resolution (top left) and its uncertainty (top right) in the $0.2 \leq |\eta| \leq 0.7$ region for PFlow jets in data and MC. The different terms are highlighted. A direct comparison between EMTopo jets and PFlow jets is also shown in the same $|\eta$ region (bottom left) and for jet $p_{\mathrm{T}} = 30$ GeV as a function of the jet $|\eta|$ (bottom right). [118].

Different *in-situ* methods can be employed to measure jets resolution, including recoiling jets against a precisely measured object or a well balanced dijet system, and a noise-related approach:

- the *dijet in-situ measurement* relies on the definition of an *asymmetry* variable quantifying the deviation from exact balance, due to either experimental resolution, additional radiation or biases introduced by event selections:

$$\mathcal{A} = \frac{p_{\mathrm{T}}^{\mathrm{probe}} - p_{\mathrm{T}}^{\mathrm{ref}}}{p_{\mathrm{T}}^{\mathrm{avr}}} \tag{4.6}$$

  where the *reference* jet is chosen to be located in a well calibrated region of the detector and the *probe* jet is the measured one, with no preferential location;

- the *noise in-situ measurement* are obtained by measuring the fluctuations in the energy deposits due to pile-up using data samples that are collected by random unbiased triggers. The measurement is based on a *random cones* method in which energy deposits in the calorimeter are summed at the constituent energy scale in circular areas analogous to the jet area for anti-kt with $R = 0.4$ jets.

The two measured are combined to provide final results, shown in Figure 4.20 for PFlow jets only and the two contributions highlighted, as well as compared with EMTopo jets.

4.4.5 Jet-Vertex-Tagger (JVT) for pile-up suppression

The reconstruction of the jets coming from the hard-scattering is strongly affected by both in-time pile-up, arising from additional pp interactions in the current bunch-crossing, and out-of-time pile-up from previous and following bunch crossings relative to the triggered event. Despite the already mentioned pile-up subtraction strategies, local fluctuations in the pile-up activity, may result in spurious pile-up jets. For this reason, a Jet-VertexTagger (JVT) method [127] has been implemented in order to further discriminate between pile-up and hard-scatter jets.
The JVT consists of a multivariate combination of two track-based variables. The first one is similar to the previously used Jet-Vertex-Fraction (JVF), defined as the ratio between the scalar p_{T} sum of the tracks that are associated with the jet and originate from the hard-scatter vertex, and the scalar p_{T} sum of all associated tracks (therefore including pileup). Compared to the JVF, the new variable is corrected in order to take into account the linear increase of the average scalar sum of p_{T} from pile-up tracks with the number of primary vertices:

$$\mathrm{corrJVF} = \frac{\sum_k p_{\mathrm{T},k}^{\mathrm{track}}\ (\mathrm{PV}_0)}{\sum_l p_{\mathrm{T},l}^{\mathrm{track}}\ (\mathrm{PV}_0) + \frac{\sum_{n\geq 1}\sum_{l'} p_{\mathrm{T},l'}^{\mathrm{track}\ (\mathrm{PV}_n)}}{(k\cdot n_{\mathrm{track}}^{\mathrm{PU}})}}$$

where PV_0 denotes the hard-scatter primary vertex, while PV_n the pile-up vertices. The $k \cdot n_{\mathrm{track}}^{\mathrm{PU}}$ in the second term of the denominator is the mentioned correction.
The second variable of the JVT is the scalar p_{T} sum of the tracks that are associated with the jet and originate from the hard-scatter vertex divided by the fully calibrated jet p_{T}, which includes pile-up subtraction:

$$R_{p\ \mathrm{T}} = \frac{\sum_k p_{\mathrm{T},k}^{\mathrm{track}}\ (\mathrm{PV}_0)}{p_{\mathrm{T}}^{\mathrm{jet}}}$$

Low values of these two variables are likely associated to pile-up jets.
The discriminant variable JVT, shown in Figure 4.21 (left), is derived by combining corrJVF and $R_{p\ \mathrm{T}}$ in a 2-dimensional likelihood.
The performance of JVT requirements have been studied on dijet simulated samples: signal jet efficiencies of 80%, 90% and 95% are achieved for pile-up fake rates of respectively 0.4%, 1.0% and 3%, as shown in Figure 4.21 (right) which also presents a comparison with the discrimination power reached using as discriminant the JVF variables or one between corrJVF and $R_{p\ \mathrm{T}}$.
The JVT cut is not efficient against forward jets, in the $2.5 < |\eta| < 4.5$ region, outside the coverage of the Inner Detector. An alternative tagger has therefore been defined, referred to as the *forward JVT* (fJVT) [128]. This tagger discriminates PU jets based on the fact that a balance is expected between a PU forward jet and the transverse momentum associated to the corresponding PU vertex. The first step is therefore providing a reliable estimate of the transverse momentum associated to each PU vertex, in the central region. Central jets are reconstructed and calibrated for all PU vertices and QCD PU jets, arising from a single vertex, are discriminated from stochastic PU contribution based on a modified version of the $R_{p_{\mathrm{T}}}$ variable centered in PU vertices rather than the HS one. A value of $R_{p_{\mathrm{T}}} > 0.1$ identifies QCD PU contributions. The HS jets are rejected by requiring $JVT < 0.2$, ensuring a rejection power of 98.8%.

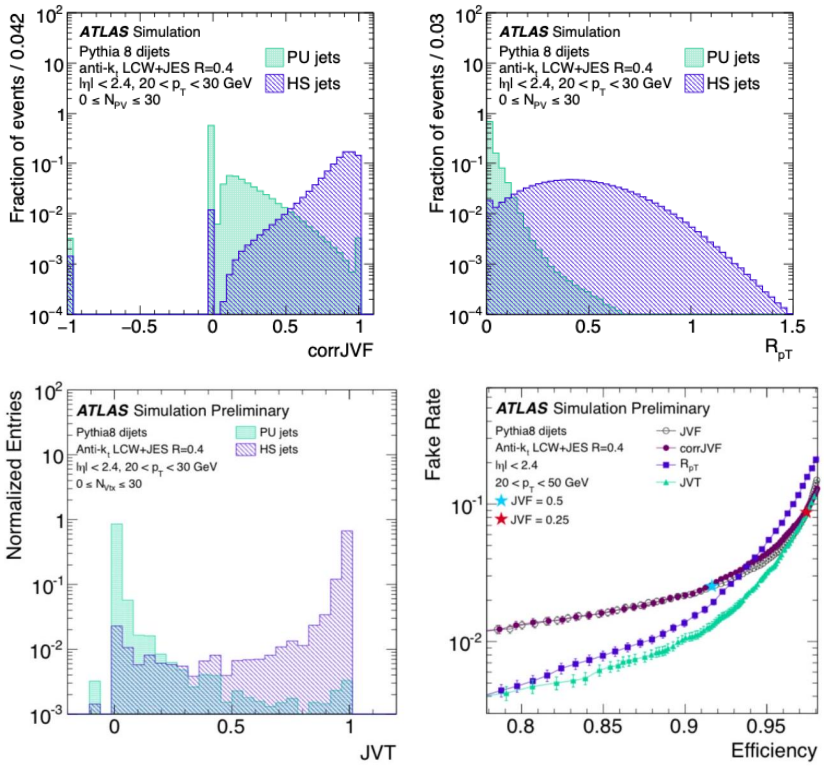


Figure 4.21: Distribution of the corrJVF (top left), R_{p_T} (top right) and JVT variables (bottom left) for pile-up (PU) and hard-scatter (HS) jets at the LC scale, with $20 < p_{\mathrm{T}} < 30$GeV. In the bottom right plot, fake rate, at LC scale, from pile-up jets versus hard-scatter jet efficiency curves for JVF, corrJVF, $\mathrm{R}_{p\ \mathrm{T}}$, and JVT [129].

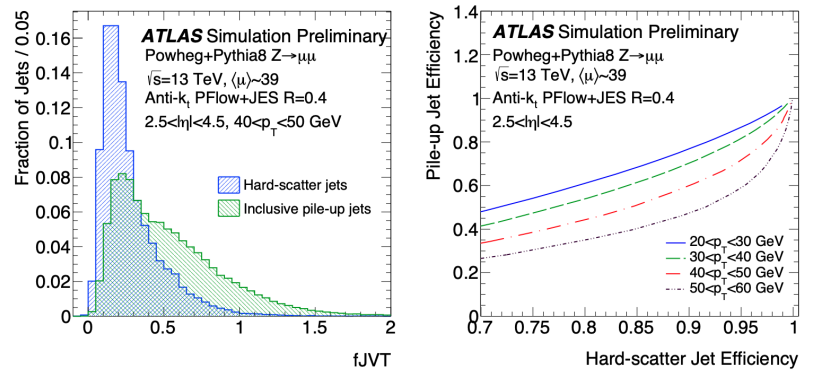


Figure 4.22: Distribution of the fJVT variable for hard-scatter and pile-up forward PFlow jets in the $40 < p_{\mathrm{T}} < 50$ GeV range (left), and efficiency for PU forward jets in simulated Z+jets events as a function of the efficiency for hard-scatter jets for different jet p_{T} ranges [128].

EMTopo		
Loose	$20 < p_T < 60$ GeV and $\|\eta\| < 2.4$	$JVT \geq 0.11$
Medium (Default)	$20 < p_T < 60/120$ GeV and $\|\eta\| < 2.4$ $20 < p_T < 60/120$ GeV and $2.4 < \|\eta\| < 2.5$	$JVT \geq 0.59$ $JVT \geq 0.11$
Tight	$20 < p_T < 60$ GeV and $\|\eta\| < 2.4$	$JVT \geq 0.91$
PFlow		
Medium	$20 < p_T < 60$ GeV and $\|\eta\| < 2.4$	$JVT \geq 0.2$
Tight (Default)	$20 < p_T < 60$ GeV and $\|\eta\| < 2.4$	$JVT \geq 0.5$
Forward		
Loose	$20 < p_T < 120(60)$ GeV and $2.5 < \|\eta\| < 4.5$	$fJVT < 0.5$
Tight	$20 < p_T < 120(60)$ GeV and $2.5 < \|\eta\| < 4.5$	$fJVT < 0.4$

Table 4.6: List of recommended Working Points for EMTopo and PFlow jets.

The missing transverse momentum associated to a given PU vertex i is therefore estimated as:

$$\mathbf{p}_{\mathrm{T},i}^{\mathrm{miss}} = -\left(\sum_{p_{\mathrm{T}}^{\mathrm{jet}} > 20 \text{ GeV}} \mathbf{p}_{\mathrm{T}}^{\mathrm{jet}} + \sum_{\mathbf{p}_{\mathrm{T}}^{\mathrm{jet}} < \mathbf{20} \text{ GeV}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}} + \sum_{\text{jets fail } \mathrm{R}_{\mathrm{p_T}}^{\mathrm{i}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}} \right) \tag{4.7}$$

where the terms are central jets and tracks associated to them. The fJVT score is finally defined as the maximum of the normalized projection of $\mathbf{p}_{\mathrm{T},i}^{\mathrm{miss}}$ on the forward jet direction:

$$\mathrm{fJVT} = \max_i \left(\frac{\mathbf{p}_{\mathrm{T},i}^{\mathrm{miss}} \cdot \mathbf{p}_{\mathrm{T}}^{\mathrm{fwd}}}{\left|\mathbf{p}_{\mathrm{T}}^{\mathrm{fwd}}\right|^2} \right) \tag{4.8}$$

This variable is expected to be closer to 1 for forward PU jets, which should be well balanced with the $\mathbf{p}_{\mathrm{T}}^{\mathrm{fwd}}$, and closer to 0 for HS forward jets, as shown in the left plot of Figure 4.22, for PFlow jets. It is therefore used as an upper threshold to define the forward jet PU rejection, with efficiencies shown in the right plot of Figure 4.22.
Different Working Points are defined, corresponding to different JVT (and possibly fJVT) cut thresholds, as described in table 4.6.

4.4.6 Jet cleaning

The jet cleaning procedure [130] allows to discriminate between jets coming from pp collisions and jets of non-collision origin. The latter comprise the cosmic ray background and the Beam induced background (BIB), composed of muons arising from beam losses which can be reconstructed as fake jets with energy as high as the beam energy. Another residual background which can be suppressed by means of the jet cleaning is the one given by phenomena of coherent noise or isolated pathological cells in the calorimeters. These are mainly subtracted before the object reconstruction, by masking the problematic cells, but a fraction can remain undetected.
The jet cleaning selection reject all the jets which don't satisfy all the required Quality

criteria, involving variables based on signal pulse shape in the LAr calorimeters to account for calorimeter noise, energy ratio variables such as the electromagnetic fraction or the HEC energy fraction, and track-based variables, more specifically the charged fraction and the ratio of the charged fraction over the energy fraction in the layer with maximum energy deposit.
Two Working points are defined: the Loose WP reject all the jets tagged as LooseBad, ensuring a 99.9% selection efficiency on good jets, while the Tight WP rejects jets tagged as TightBad, with a 99.5% selection efficiency on good jets.

4.5 Hadronically decaying τ leptons

The τ-lepton decays with a lifetime of about 2.9×10^{-13} s, thus being detected by the ATLAS experiment through its decay products. Hadronic decays constitute about 65% of the total decays, and display characteristic displacement, multiplicity and kinematic properties which can be exploited for the τ-lepton reconstruction, identification and calibration [131]. In this thesis, hadronically decaying τ-leptons are only used in the τ veto of the mono-photon analysis, therefore only a brief overview of their reconstruction in ATLAS is given in this section.

Reconstruction Hadronically decaying τ-leptons ($\tau_{had-vis}$) are reconstructed by combining information from the calorimeters and the ID. The τ-lepton reconstruction algorithm is seeded by jets with $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$, reconstructed with the *anti-kt* algorithm with $R = 0.4$ and calibrated using a local hadronic calibration (LC) as inputs. The p_{T} of the $\tau_{had-vis}$ at LC scale is set to the energy of the TopoClusters within $R = 0.2$. A specific τ *vertex association* algorithm is employed, to improve the identification of the correct vertex, which can be challenging especially in high *pile-up* conditions. All tracks in the $\Delta R < 0.2$ region aroung the seed jet direction are taken as input, and the τ vertex (TV) is identified as the primary vertex candidate to which the largest fraction of the track p_{T} sum is associated.
Since $\tau_{had-vis}$ are required to have 1 or 3 associated charged-particle tracks (to identify the 1- or 3-prongs final states), *track association* is an important step of $\tau_{had-vis}$ reconstruction. Only tracks in the $\Delta R < 0.2$ region around the $\tau_{had-vis}$ direction are selected. In addition, they are required to have $p_{\mathrm{T}} > 1$ GeV, at least two associated hits in the pixel detector (including the IBL), and at least seven hits in total in the pixel and the SCT detectors. Finally, requirements are imposed on the distance of closest approach of the track to the τ vertex (TV) in the transverse ($|d_0| < 1.0$ mm) and longitudinal ($|\Delta z_0 sin\theta| < 1.5$ mm) plane.
The reconstruction efficiencies for 1- and 3-prong decays as a function of the truth p_{T} of the τ-leptons, defined as the fraction of 1-prong (3-prong) hadronic tau decays which are reconstructed as 1-track (3-track) $\tau_{had-vis}$, are shown in Figure 4.23: an approximately constant efficiency of about 70% is observed for the 1-prong case, while the 3-prong $\tau_{had-vis}$ shows a reduction in the low-p_{T} bins, due to the minimum transverse momentum requirement on the charged decay products, and at high-p_{T} due to the increased collimation of the decay products that results in an increased probability to miss a track because of overlapping trajectories.

Identification The discrimination between $\tau_{had-vis}$ and hadronic jets is performed in the identification step, based on Boosted Decision Tree (BDT) algorithms trained

separately for 1-track and 3-track candidates, on the $Z/\gamma^* \rightarrow \tau\tau$ signal sample and di-jets background sample. The input variables, listed in Table 4.7 exploits tracks and TopoClusters information, in a core or isolation region around the direction of the $\tau_{had-vis}$ candidate. Three working points are defined, Loose, Medium and Tight, with global reconstruction and identification efficiencies ranging from 40% (30%) to 80% for 1-prong (2-prong) candidates, as shown in Figure 4.24

Variable	Description	1-prong	3-prong
f_{cent}	Fraction of the calorimeter transverse energy deposited in the region $\Delta R < 0.1$ with respect to the total energy in $\Delta R < 0.2$ around the $\tau_{had-vis}$ candidate	✓	✓
$f^{-1}_{leadtrack}$	Transverse energy sum, at the EM energy scale, deposited in all cells belonging to TopoClusters in the core region of the $\tau_{had-vis}$ candidate, divided by the transverse momentum of the highest-p_{T} charged particle in the core region	✓	✓
$R^{0.2}_{track}$	p_{T}-weighted ΔR distance of the associated tracks to the τ direction, using $\tau_{had-vis}$ only tracks in the core region	✓	✓
$\lvert S_{leadtrack}\rvert$	Absolute value of transverse impact parameter of the highest-p_{T} track in the core region, calculated with respect to the TV, divided by its estimated uncertainty	✓	
f^{track}_{iso}	Fraction of tracks p_{T} in the isolation region $0.2 < \Delta R < 0.4$		
ΔR_{max}	Maximum ΔR between a track associated with the $\tau_{had-vis}$ candidate and its direction		✓
S_T^{flight}	Decay length of the secondary vertex (vertex reconstructed from the tracks associated with the core region of the $\tau_{had-vis}$ candidate) in the transverse plane, calculated with respect to the TV, divided by its estimated uncertainty. It is defined only for multi-track $\tau_{had-vis}$ candidates		✓
m_{track}	Invariant mass calculated from the sum of the four-momentum of all tracks in the core and isolation regions, assuming a pion mass for each track	✓	✓
$f^{EM}_{track-HAD}$	Fraction of EM energy from charged pions	✓	✓
f^{EM}_{track}	Ratio of EM energy to track momentum	✓	✓
$m_{EM+track}$	Invariant mass of the system composed of the tracks and up to two most energetic EM clusters in the core region	✓	✓
$p_{\mathrm{T}}^{EM+track}/p_{\mathrm{T}}$	Ratio of the $\tau_{had-vis}$ p_{T}, estimated using the vector sum of track momenta and up to two most energetic EM clusters in the core region to the calorimeter-only measurement of $\tau_{had-vis}$ p_{T}	✓	✓

Table 4.7: Input variables for the $\tau_{had-vis}$ identification BDT algorithm

Calibration The LC scale mostly corrects for the calorimeter non-compensation and for the energy deposited in dead material or outside TopoClusters. Subsequently, the calibration to the τ energy scale consists of two additional corrections to recover the true visible energy. First, the energy contribution originating from pile-up interactions is subtracted. A response correction is then applied to account for possible additional effects: decay products not reaching the calorimeter, not depositing enough energy to create TopoClusters, or not detected within $\Delta R < 0.2$ of the reconstructed $\tau_{had-vis}$

candidate. The $\tau_{had-vis}$ momentum is computed from the energy assuming zero mass. The τ energy resolution for 1-prong and 3-prong candidates is reported in Figure 4.25 as a function of the $\tau_{had-vis}$ energy and $|\eta|$ region: it ranges from about 25-30% at low energies, down to 5-10% at higher energies, reaching a plateau for calibrated energies close to 300 GeV.

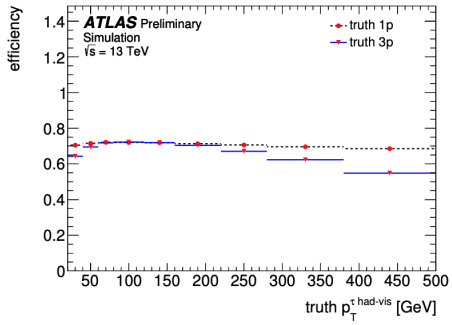


Figure 4.23: Reconstruction efficiencies for 1- and 3-prong decays as a function of the truth p_{T} of the τ-lepton.

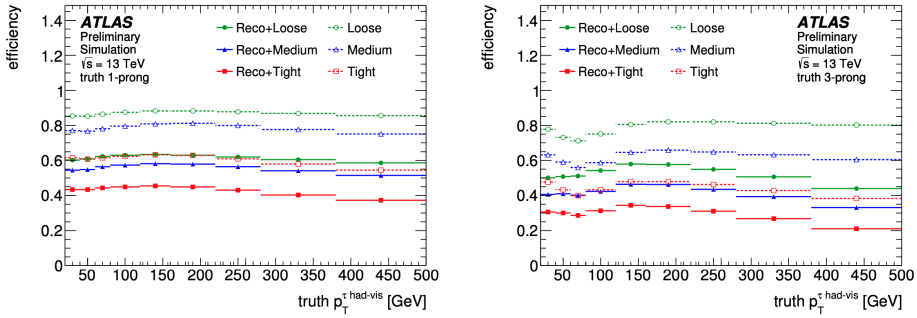


Figure 4.24: Efficiency for $\tau_{had-vis}$ identification (open symbols) and combined reconstruction and identification efficiency (full symbols) as a function of the $\tau_{had-vis}$ p_{T}, for 1-prong (left) and 3-pronf (right) candidates.

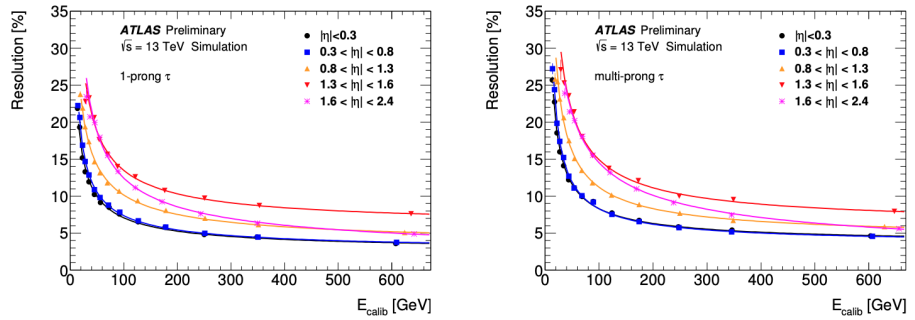


Figure 4.25: Resolution on the $\tau_{had-vis}$ calibrated energy for reconstructed 1-prong (left) and multi-prong (right) candidates, as a function of the energy and $|\eta|$ region.

4.6 The Missing Transverse Momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$)

In final states of pp collisions at LHC, the momenta in the plane transverse to the beam axis of all particles in the final state are expected to sum to zero, due to momentum conservation. This provides a possible signature for neutrinos or other Beyond Standard Model particles that pass through the ATLAS detector without interacting. The transverse momentum carried by these undetected particles produces an imbalance in the visible total transverse momentum of the final state that goes under the name of Missing Transverse Momentum $\left(E_{\mathrm{T}}^{\mathrm{miss}}\right)$.
The reconstruction of this variable [132] is based on information collected from all the detector subsystems and requires the most complete and unambiguous representation of the hard interaction of interest by calorimeter and tracking signals. Therefore, it is deeply affected by detector-related limitations such as acceptance or defects, as well as by objects misreconstructions or miscalibrations or contributions from particles coming from pile-up or underlying events and mistakenly associated to the hard scattering event.
The missing transverse momentum is defined as the negative vectorial sum of the transverse momenta of all the reconstructed particles (hard objects) and of the clusters or tracks not associated to any physics particle (soft-term).
The component of the $E_{\mathrm{T}}^{\mathrm{miss}}$ are therefore defined as:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum \vec{p}_{\mathrm{T}}^{\,\mu} - \sum \vec{p}_{\mathrm{T}}^{\,e} - \sum \vec{p}_{\mathrm{T}}^{\,\gamma} - \sum \vec{p}_{\mathrm{T}}^{\,\tau} - \sum \vec{p}_{\mathrm{T}}^{\,jets} - \sum \vec{p}_{\mathrm{T}}^{\,soft}$$

where the order of the terms reflects the most commonly chosen priority order following which the energy deposits in the calorimeters and tracks are matched to reconstructed objects in order to minimise double-counting of detector signals due to overlap of close-by objects.
Another important observable is the total transverse energy in the detector $\sum E_{\mathrm{T}}$, which quantifies the total event activity and can give information about $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. It's defined as the scalar sum of the transverse momenta of reconstructed objects and soft-term signals that contribute to the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction:

$$\sum E_{\mathrm{T}} = \sum p_{\mathrm{T}}^{\mu} + \sum p_{\mathrm{T}}^{e} + \sum p_{\mathrm{T}}^{\gamma} + \sum p_{\mathrm{T}}^{\tau} + \sum p_{\mathrm{T}}^{jets} + \sum p_{\mathrm{T}}^{\mathrm{soft}}$$

Hard-objects definition The hard-objects entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation are defined as the "preselected" or "baseline" and calibrated objects of the analysis, passing an Overlap Removal (OR) procedure described in the following paragraphs. Depending on the used jet collection, whether PFlow or EMTopo jets, some differences are expected in the Jet Term, while the other terms are expected to be unvaried.
At analysis level, appropriate selections are applied on the reconstructed objects in each event, in order to reject fake or otherwise problematic signatures. The same selections must therefore be applied to the hard-objects entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation to ensure a consistent interpretation of the event. For this reason, $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction software provides flexibility for the choice of the selection criteria for all the particles, with exception of the jets which are selected by choosing one among the predefined operating points, summarized in Table 4.8 for $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructed either from PFlow or EMTopo jets.

Overlap Removal An Overlap Removal procedure is performed on the preselected and calibrated objects entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, in order to resolve possible overlapping energy deposits matched to more than one hard-object and avoid double-countings. The

WP	Jet kinematics	JVT cut
	EMTopo	
Loose	$20 < p_T < 60$ GeV and $\lvert\eta\rvert < 2.4$	$JVT \geq 0.59$
Tight	$20 < p_T < 60$ GeV and $\lvert\eta\rvert < 2.4$ $p_T > 30$ GeV and $\lvert\eta\rvert \geq 2.4$	$JVT \geq 0.59$ Tight fJVT (optional)
Tighter	$20 < p_T < 60$ GeV and $\lvert\eta\rvert < 2.4$ $p_T > 35$ GeV and $\lvert\eta\rvert \geq 2.4$	$JVT \geq 0.59$ Tight fJVT (optional)
Tenacious	$20 < p_T < 40$ GeV and $\lvert\eta\rvert < 2.4$ $40 < p_T < 60$ GeV and $\lvert\eta\rvert < 2.4$ $60 < p_T < 120$ GeV and $\lvert\eta\rvert < 2.4$ $p_T > 35$ GeV and $\lvert\eta\rvert \geq 2.4$	$JVT \geq 0.91$ $JVT \geq 0.59$ $JVT \geq 0.11$ Medium fJVT
	PFlow	
Loose	$20 < p_T < 60$ GeV and $\lvert\eta\rvert < 2.4$	$JVT \geq 0.5$
Tight	$20 < p_T < 60$ GeV and $\lvert\eta\rvert < 2.4$ $p_T > 30$ GeV and $\lvert\eta\rvert \geq 2.4$	$JVT \geq 0.5$ Tight fJVT (optional)
Tighter	$20 < p_T < 60$ GeV and $\lvert\eta\rvert < 2.4$ $p_T > 35$ GeV and $\lvert\eta\rvert \geq 2.4$	$JVT \geq 0.59$ Tight fJVT (optional)

Table 4.8: Summary of the jet selections applied in $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, depending on the $E_{\mathrm{T}}^{\mathrm{miss}}$ WP.

priority order shown in eq 2.13 is followed. Therefore all electrons passing the selection enter the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation while the lower-priority reconstructed particles are fully rejected if they overlap with a higher-priority object that has already been added to the calculation. Two objects are considered to be overlapped if they share their calorimetric signal (for EMTopo $E_{\mathrm{T}}^{\mathrm{miss}}$) or a PFO (for PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$), i.e. if there are topoclusters/PFOs which are geometrically matched to both the hard objects in a ΔR cone.
For the jets, additional criteria have to be applied in cases of electrons, photons or hadronically-decaying τ-leptons which are close to a real jet with a partial overlap, in order to decide the amount of a jet's energy signature to be associated to the jet. In the past, the main discriminating variable used to be the electron/photon/hadronically decaying τ-lepton energy fraction in the given jet:

$$f_{\mathrm{overlap}} = \frac{E_{e(\gamma,\tau)}^{\mathrm{EM}}}{e_{\mathrm{jet}}^{\mathrm{EM}}}$$

If $f_{\mathrm{overlap}} > 0.5$ the jet was assigned to the soft term, and therefore only the tracks associated with the jet were included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. Otherwise, the jet was included with its calibrated transverse momentum scaled by f_{overlap} . Nevertheless, this procedure suffers from a relatively high rejection of real jets, and in these cases the neutral component of the jet is completely lost because of the track-based definition of the soft-term (described in the next section), which includes only the charged component. This results in fake missing transverse momentum. To overcome this, a new discriminating variable has been studied:

$$\Delta p_{\mathrm{T}}^{\mathrm{EM},e(\gamma,\tau),jet} = p_{\mathrm{T}}^{\mathrm{EM},jet} - p_{\mathrm{T}}^{\mathrm{EM},e(\gamma,\tau)}$$

Choosing $f_{\text{overlap}} < 1.0$ and $\Delta p_{\text{T}}^{\text{EM},e(\gamma,\tau),\text{ jet}} > 20\text{GeV}$, the rejection of real jets is reduced. Finally, a specific μ-jet overlap is implemented, in order to solve ambiguities between jets and muons, due to the fact that muons can also experience energy loss in the calorimeters:

- a muon overlapping with a pile-up jet can lead to a JVT mis-tag of the jet, because the ID track from the muon represents a significant amount of p_{T} from the hard-scatter vertex and thus increases the JVT value;
- in cases in which the muon goes through a significant energy loss in the calorimeter, it can be reconstructed as a hard-scatter jet from the primary vertex, which is included in the $E_{\text{T}}^{\text{miss}}$ calculation, resulting in a double-counting of this p_{T} component.
- Muons can radiate hard photons at small angles, which are wrongly reconstructed as a jet with an associated muon ID track. As the transverse momentum carried by the FSR photon is not recovered in muon reconstruction, jets representing this photon need to be included in the $E_{\text{T}}^{\text{miss}}$ reconstruction.

The strategies for the resolution of these ambiguities is described in [132], section 6.6.2.

Soft Term As already mentioned, the soft term includes all the detector signals not matched to the reconstructed objects and may contain contributions from the hard scattering as well as the underlying event and pile-up interactions. In Run-1 analyses, a Calorimeter Soft Term (CST) reconstructed from calorimeter-only measurements was used, but due to the high pile-up environment of the LHC from 2012 onwards, the soft component suffers from a high degree of contamination from interactions besides the hard scattering vertex of interest. For this reason, a Track Soft Term (TST) calculated from the tracks is used for Run-2 analysis, which provides a more robust measurement against pile-up, thanks to a more accurate track-to-vertex association.
In this track-based reconstruction of the Soft Term, tracks not associated to hard objects (in EMTopo $E_{\text{T}}^{\text{miss}}$) or tracks associated to charged PFO not associated to hard objects (for the PFlow $E_{\text{T}}^{\text{miss}}$) are associated to the hard scatter vertex by considering the transverse and the longitudinal impact parameters (d_0 and z_0 respectively), with respect to the hard scatter vertex position. The exact requirements are summarized, for EMTopo and PFlow $E_{\text{T}}^{\text{miss}}$, in Table 4.9. To avoid the double counting of tracks already included

EMTopo $E_{\text{T}}^{\text{miss}}$	**PFlow $E_{\text{T}}^{\text{miss}}$**
$p_{\text{T}} > 0.4$ GeV	$p_{\text{T}} > 0.5$ GeV
$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.5$
$\lvert\frac{d_0}{\sigma(d_0)}\rvert < 2.0$ mm	-
$\lvert z_0 \sin\theta\rvert < 3.0$ mm	$\lvert z_0 \sin\theta\rvert < 3.0$ mm

Table 4.9: Track selections for the TST

in other physics objects, the tracks matched to the hard-term objects are excluded by applying the following selection criteria:

- tracks within $\Delta\eta < 0.2$ and $\Delta\phi < 0.05$ around electrons and photons;
- tracks within a cone of size $\Delta R = 0.2$ around hadronically decaying τ-leptons;

- ID tracks associated with identified muons;
- tracks matched to jets using the ghost association technique;
- isolated tracks with $p_{\mathrm{T}}^{\mathrm{track}} > 120$GeV, or $p_{\mathrm{T}}^{\mathrm{track}} > 200$ GeV for $|\eta_{\mathrm{track}}| < 1.5$, having estimated relative resolution on their p_{T} larger than 40% or having no associated calorimeter energy deposit with p_{T} larger than 65% of the track p_{T}. This is a cleaning cut to remove mismeasured tracks.

4.6.1 $E_{\mathrm{T}}^{\mathrm{miss}}$ Systematics

The $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainty includes the propagation of all the systematics associated to each object entering the hard-term, and additional systematics related to the reconstruction of the track-based soft-term and depending on the precision of the soft-term modeling in simulations.
In order to take into account the fact that the balance expected between the soft term momentum $p_{\mathrm{T}}^{\mathrm{soft}}$ and the hard term $p_{\mathrm{T}}^{\mathrm{hard}}$ in events with no true $E_{\mathrm{T}}^{\mathrm{miss}}$ is spoilt by detector resolution effects, the modelling of the soft term is studied using different projections of $p_{\mathrm{T}}^{\mathrm{soft}}$ along $p_{\mathrm{T}}^{\mathrm{hard}}$ in data and MC (Figure 4.26):

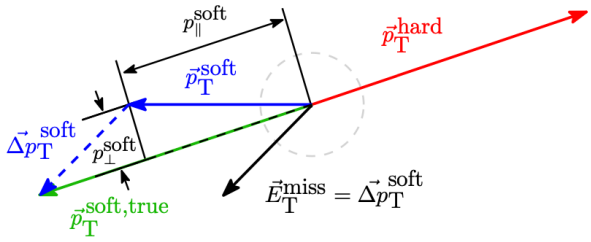


Figure 4.26: Representation of the track-based soft term projections with respect to $p_{\mathrm{T}}^{\mathrm{hard}}$ for the calculation of the TST systematic uncertainties.

- The parallel scale (Δ_L), representing the mean value of the parallel projection $\left(p_{//}^{\mathrm{soft}}\right)$ of $p_{\mathrm{T}}^{\mathrm{soft}}$ along $p_{\mathrm{T}}^{\mathrm{hard}}$
- The parallel resolution $\left(\sigma_{\parallel}\right)$, defined as the root-mean-square (RMS) of $p_{\parallel}^{\mathrm{soft}}$.
- The transverse resolution ($\sigma_{\perp}$), defined as the root-mean-square (RMS) of the perpendicular component of $p_{\mathrm{T}}^{\mathrm{soft}}$ with respect to $p_{\mathrm{T}}^{\mathrm{hard}}$.

The systematic uncertainty is defined as the maximal disagreement between data and different Monte Carlo generators for a certain set of $p_{\mathrm{T}}^{\mathrm{hard}}$ bins. In addition, the total systematic is splitted into jet-inclusive and jet-veto selections and taking the maximal variation of these two cases in order to account for possible differences between event topologies with or without jets.
The three uncertainties contributions, for EMTopo and PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$, are shown in Figure 4.27 for the inclusive jet case, taken as an example, while Figure 4.28 summarizes the three components of the final uncertainty.

4.6.2 $E_{\mathrm{T}}^{\mathrm{miss}}$ significance

The $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is defined as the ratio between $E_{\mathrm{T}}^{\mathrm{miss}}$ and its resolution, and it is an useful variable to discriminate between events with real $E_{\mathrm{T}}^{\mathrm{miss}}$ and events with fake

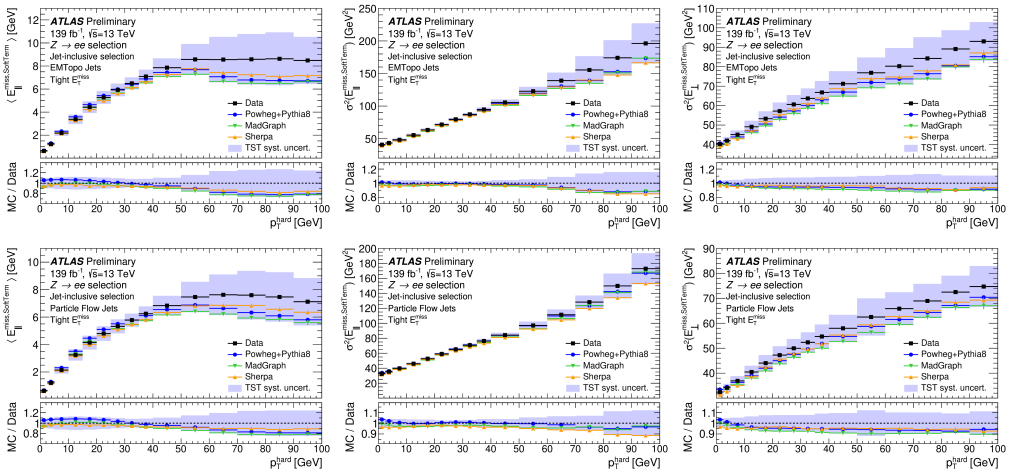


Figure 4.27: Parallel scale (left), squared parallel resolution (center) and squared transverse resolution (right) in bins of the vector sum of the transverse momenta of hard objects, for a $Z \to ee$ data (black points) and MC sample simulated with different generators (other coloured points) using full Run-2. Tight EMTopo $E_{\mathrm{T}}^{\mathrm{miss}}$ (top) and PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ (bottom) are both shown. The shaded band shows the TST systematic uncertainty, symmetrized around data, derived as the maximal discrepancy between data and the different MC generators in both the jet-inclusive and 0-jet selections (not shown here) for each bin [133].

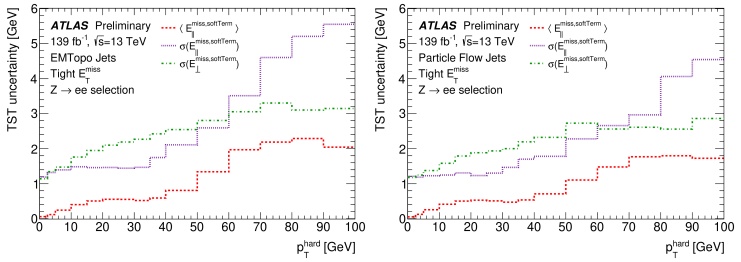


Figure 4.28: Components of the TST uncertainties in bins of the vector sum of hard objects transverse momentum, for a $Z \to ee$ dataset using full Run-2. Tight EMTopo $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ (right) are both shown. [133].

$E_{\mathrm{T}}^{\mathrm{miss}}$ arising for example from object misreconstructions or miscalibrations and detector defects. In fact, a high value of significance is an hint for real $E_{\mathrm{T}}^{\mathrm{miss}}$, given by undetected particles in the event, since it indicates that the observed $E_{\mathrm{T}}^{\mathrm{miss}}$ is not well explained by resolution smearing alone. On the other hand, lower significance values are likely associated to events with fake $E_{\mathrm{T}}^{\mathrm{miss}}$.

An event-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is defined as $\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sum_{\mathrm{T}} E_{\mathrm{T}}}$, according to the proportionality between the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and the $\sum E_{\mathrm{T}}$ variable. Despite this approximation is fully valid only under the assumption that the $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed from purely calorimetric measurements, it provides an acceptable rejection of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ also when applied to TST $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation.

In order to get a better evaluation of this variable, an object-based definition has been developed [134], which is calculated event by event taking into account the expected reso-

lutions and likelihood of mismeasurement of all the objects entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation (a plot showing the resolutions for different objects as a function of the transverse momentum and for different values of pseudorapidity is shown in Figure 4.29), as well as directional correlations between measurements.

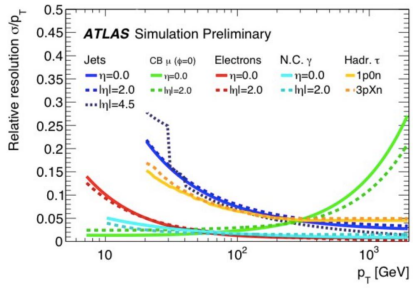


Figure 4.29: Resolution of the objects entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation as a function of p_{T} and for different values of $|\eta|$. The label N.C. γ indicates unconverted photons and Hadr. τ hadronically decaying τ-leptons. 1p0n (3pXn) corresponds to the decay mode with one (three) charged hadron, zero (one or more) neutral hadrons and a non detectable τ-neutrino.

The oject-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is defined as the square root $\mathcal{S}$ of:

$$\mathcal{S}^2 = 2\ln\left(\frac{\max_{p_{\mathrm{T}}^{\mathrm{inv}}\neq 0}\mathcal{L}\left(E_{\mathrm{T}}^{\mathrm{miss}} \mid p_{\mathrm{T}}^{\mathrm{inv}}\right)}{\max_{p_{\mathrm{T}}^{\mathrm{inv}}=0}\mathcal{L}\left(E_{\mathrm{T}}^{\mathrm{miss}} \mid p_{\mathrm{T}}^{\mathrm{inv}}\right)}\right)$$

which is basically the log-likelihod to test the hypothesis that the total transverse momentum carried by invisible particles $\left(p_{\mathrm{T}}^{\mathrm{inv}}\right)$ is equal to zero against the hypothesis that $p_{\mathrm{T}}^{\mathrm{inv}}$ is different from zero. The numerator and denominator are calculated maximising on the parameters of the likelihood with the constraints $p_{\mathrm{T}}^{\mathrm{inv}} \neq 0$ and $p_{\mathrm{T}}^{\mathrm{inv}} = 0$ respectively. The likelihoods are defined as:

$$\mathcal{L}\left(E_{\mathrm{T}}^{\mathrm{miss}} \mid p_{\mathrm{T}}^{\mathrm{inv}}\right) \propto \exp\left[-1/2\left(E_{\mathrm{T}}^{\mathrm{miss}} - p_{\mathrm{T}}^{\mathrm{inv}}\right)^{\mathrm{T}}\left(\sum_i V^i\right)^{-1}\left(E_{\mathrm{T}}^{\mathrm{miss}} - p_{\mathrm{T}}^{\mathrm{inv}}\right)\right]$$

where it has been assumed that the measurement of each reconstructed object i is independent from others, that the vectorial sum of all the true transverse momentum associated to the hard-objects is $\sum \pi_{\mathrm{T}}^i = -p_{\mathrm{T}}^{\mathrm{inv}}$ and that the probability distribution of measuring p_{T}^i given a true transverse momentum π_{T}^i is a gaussian with covariance matrix V^i. The $\mathcal{S}^2$ can be now rewritten as:

$$\mathcal{S}^2 = 2\ln\left(\frac{\mathcal{L}\left(E_{\mathrm{T}}^{\mathrm{miss}} \mid E_{\mathrm{T}}^{\mathrm{miss}}\right)}{\mathcal{L}\left(E_{\mathrm{T}}^{\mathrm{miss}} \mid 0\right)}\right) = \left(E_{\mathrm{T}}^{\mathrm{miss}}\right)^{\mathrm{T}}\left(\sum_i V^i\right)^{-1}\left(E_{\mathrm{T}}^{\mathrm{miss}}\right)$$

where the index i indicates each object entering the calculation and the covariance matrixes include the effects of the different hard-objects resolutions, plus one covariant matrix associated to the soft term. Through appropriate rotations, a final synthetic expression can be written:

$$\mathcal{S}^2 = \left(E_{\mathrm{T}}^{\mathrm{miss}} 0\right) \begin{pmatrix} \sigma_{\mathrm{L}}^2 & \rho_{\mathrm{LT}}\sigma_{\mathrm{L}}\sigma_{\mathrm{T}} \\ \rho_{\mathrm{LT}}\sigma_{\mathrm{L}}\sigma_{\mathrm{T}} & \sigma_{\mathrm{T}}^2 \end{pmatrix}^{-1} \begin{pmatrix} E_{\mathrm{T}}^{\mathrm{miss}} \\ 0 \end{pmatrix} = \frac{\left|E_{\mathrm{T}}^{\mathrm{miss}}\right|^2}{\sigma_{\mathrm{L}}^2\left(1-\rho_{\mathrm{LT}}^2\right)}$$

where σ_{L}^2 and σ_{T}^2 are the total variances in the longitudinal and transverse direction respectively with respect to the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction and ρ_{LT} is the correlation factor of the longitudinal and transverse measurements.

4.6.3 $E_{\mathrm{T}}^{\mathrm{miss}}$ performance in Run-2

A comparison between EMTopo and PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ with Loose WP, in a $Z \to ee$ sample, is shown for 2015-2016 Run-2 data and MC16a simulations in Figure 4.30, together with $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of $\langle\mu\rangle$ and of the number of reconstructed Primary Vertices (N_{PV}), in order to test its stability against pile-up. In addition, Figure 4.31 shows the latest $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance distributions, in the same topology, using the full Run-2 dataset and a Tight PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$.

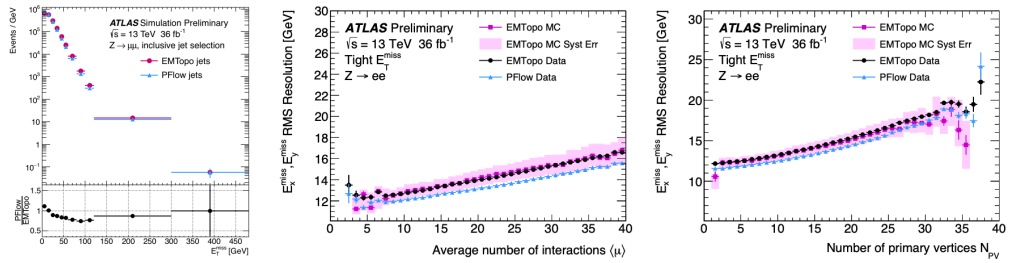


Figure 4.30: Comparison of the EMTopo and PFlow Loose $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions in MC16a simulations for the $Z \to \mu\mu$ topology (left), and of EMTopo and PFlow Tight $E_{\mathrm{T}}^{\mathrm{miss}}$ resolutions in 2015-2016 data and MC as a function of $\langle\mu\rangle$ (center) and N_{PV} (right). The resolution is defined as the RMS from the combined E_x^{miss} and E_y^{miss} distributions [132].

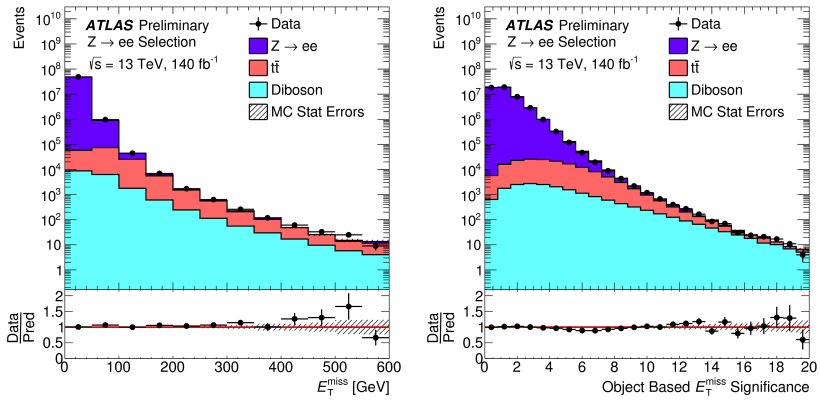


Figure 4.31: Distribution of Tight PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance (right) in data and MC using full Run-2, in the $Z \to ee$ topology [135].

Chapter 5

The $E_{\mathrm{T}}^{\mathrm{miss}}$: a Global Particle Flow reconstruction strategy for Run-3

For Run-3, an updated $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction strategy has been implemented, as part of the developments towards a Global Particle Flow (GPF) strategy. The aim of Global Particle Flow reconstruction is to uniquely associate neutral and charged Particle Flow Objects (PFOs, see Section 4.4.1) to all reconstructed stable particles, and provide unambiguous signals representing the event Energy Flow, avoiding energy deposits omissions or double countings. This can be obtained by extending to photons and leptons the Particle Flow reconstruction, illustrated in Figure 5.1 and already adopted for jets, as described in section 4.4.1. As a first step towards this, direct links between photons/leptons and their constituent PFOs have been set in Release 22 of the ATLAS reconstruction framework, `Athena`. This direct association can be included in $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction as well, replacing the association strategy used during Run-2 and potentially increasing the accuracy of ambiguities resolution, with possible improvements in the handling of overlap removal within $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, towards better $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and performance.

5.1 The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction software

The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, described in section 4.6, is based on the $E_{\mathrm{T}}^{\mathrm{miss}}$ Event Data Model (EDM), a C++ package within the `Athena` framework which provides a scheme for dynamic book-keeping of all the information needed for the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, including the kinematics of different physics objects in the event, a flag indicating if that specific object entered the calculation after analysis selections and internal overlap removal, and any modifications to the kinematics of these objects applied in the process.
The $E_{\mathrm{T}}^{\mathrm{miss}}$ EDM is based on two `xAOD` [1] classes:

- the `MissingET` collects the results of the reconstruction, in form of 2-vectors (in the trasverse plane) representing the global $E_{\mathrm{T}}^{\mathrm{miss}}$ and the different $E_{\mathrm{T}}^{\mathrm{miss}}$ terms. One `MissingET` object for each term of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined, and they are globally held in the `MissingETContainer`.;
- the `MissingEtAssociationMap` is a compact representation of the possible combinations of distinct objects whose transverse momenta should be summed in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation.

[1] The xAOD structures data in a tree, using the `TTree` ROOT class, while maintaining an "AOD-like" organization into objects (`AuxElement`), which may be grouped into containers (`AuxVectorData`).

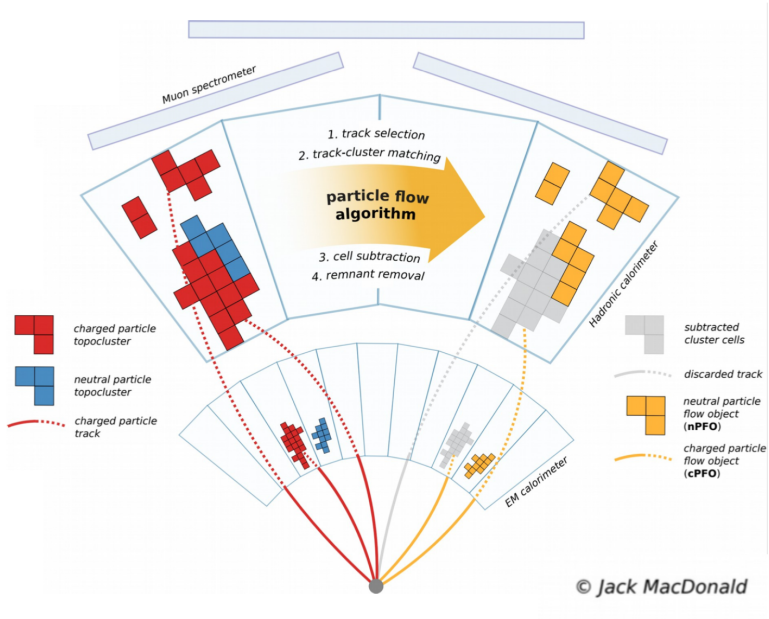


Figure 5.1: Illustration of the differences between standard handling of topo-cluster and tracks information, and the Particle Flow algorithm.

To each jet is associated a `MissingEtAssociation` object which contains the `ElementLink` objects, template classes mapping each reconstructed jet to photons or leptons that share constituents with that jet (most jets will be associated to at least one lepton or photon, since the latter are usualy reconstructed also as jets). In addition, for each jet `MissingEtAssociation` object, the constituents shared between two or more leptons/photons are stored in form of index pairs, as well as their four-momenta. This information is then used for analysis-level $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, to apply the overlap removal according to the analysis-specific priorities and physics objects selections. Finally, `ElementLink` objects and overlap informations are stored also for all physics objects not associated to any jets, collected in a separate `MissingEtAssociation` object which goes under the name of "miscellaneous association". A schematic representation of the `MissingEtAssociation` and `MissingETContainer` logic is shown in Figure 5.2 and 5.3 respectively.
The final $E_{\mathrm{T}}^{\mathrm{miss}}$ then needs to be computed at analysis-level, based on the analysis physics objects selection. The calculation is done by the `METMaker` tool: the `rebuildMET(...)` function performs the photons and leptons overlap removal and returns the corresponding $E_{\mathrm{T}}^{\mathrm{miss}}$ terms, while the `rebuildJetMET(...)` specifically addresses the Jet Term and Soft Term reconstruction. The latter include the "core" Soft Term, including PFOs not associated to any (analysis based) hard object entering $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, as well as a "miscellaneous" term built from PFOs belonging to the "miscellaneous association". The global $E_{\mathrm{T}}^{\mathrm{miss}}$ value is finally provided by the `buildMETSum(...)` function.

5.2 Global Particle Flow E_T^{miss} reconstruction

As already mentioned, the Global PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction has been implemented in the Release 22 of `Athena` framework, replacing the Run-2 PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ strategy for Run-3.

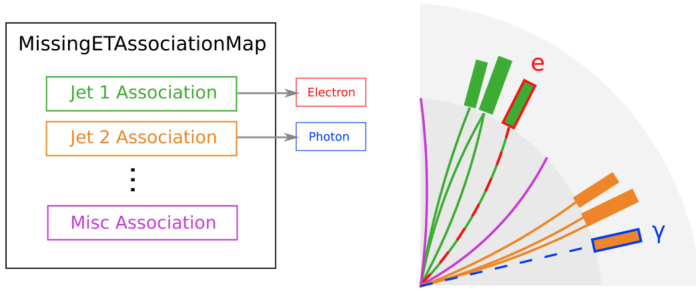


Figure 5.2: Illustration of the structure of the `MissingETAssociationMap` EDM and the typical detector signals (tracks and calorimeter clusters) corresponding to each element [136]

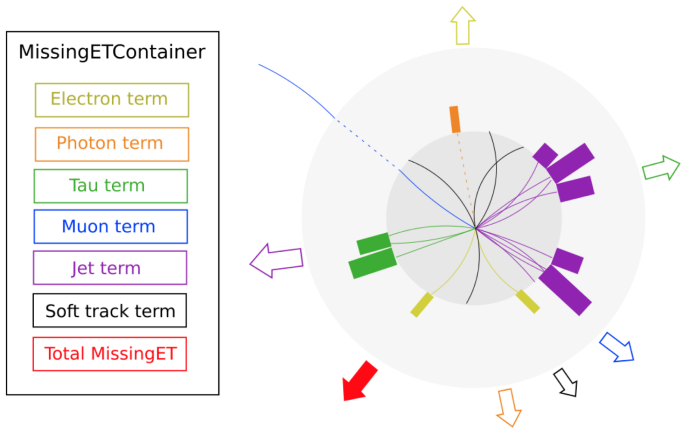


Figure 5.3: Structure of the `MissingETContainer` EDM and the typical detector signals (tracks and calorimeter clusters) corresponding to each element [136]

It relies on a prior implementation of new `ElementLink` objects between photons/leptons and PFOs. In addition, an update is required at the level of PFOs reconstruction itself: whilst the Run-2 reconstruction removed medium muons and electrons from the list of charged particle tracks, the updated PFOs are built from all tracks. Since each PFO is directly linked to the physics objects it feeds, the overlap removal can be easily performed in a subsequent step. In the next section, the Run-2 PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ strategy is briefly described, before moving to the implementation of the new reconstruction.

5.2.1 Run-2 Particle Flow $E_{\mathrm{T}}^{\mathrm{miss}}$

The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction based on PFOs in Run-2 relies on ΔR cones and track association to create links between the physics objects and the PFOs in the `MissingEtAssociationMap`, according to the following matching strategy:

- **Electrons**: match charged PFOs corresponding to the tracks associated to the electron by the electrons and photons reconstruction. Match neutral PFOs with

$\Delta R(PFO, e) < 0.1$, starting from the highest neutral PFO in this cone that has less than 150% of the electron cluster energy, and continuing to add neutral PFOs as long as adding another PFO would reduce the absolute difference between the PFO energy sum and the electron cluster. If only PFOs above the 150% cutoff exist, take the best energy match amongst these.

- **Photons** Match any charged PFOs corresponding to the tracks associated to the photon by the electrons and photons reconstruction procedures, adding tracks without an IBL hit and within $\Delta R(PFO, \gamma) < 0.05$. Match neutral PFOs according to the same rules as electrons.
- **Muons** Match the charged PFO corresponding to the muon's ID track, if it has one. A neutral PFO, associated to energy deposits arising from muon energy loss in the calorimeter, is matched if the fraction of energy it shares with the muon cluster (build only from the cells crossed by the muon track) exceeds a given threshold fraction of the neutral PFOs energy.
- **Taus** Match any PFOs within $\Delta R(PFO, \tau) < 0.2$ of the seed jet axis.

These associations are performed by specific `METAssociator` tools.
Once the `MissingEtAssociationMap` is built, the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation can be performed to provide the `MissingETContainer`, collecting the different `MissingET` objects. The "core" soft term is built in the association step, using the `METSoftAssociator`, from all PFOs not associated to any hard object (jets, leptons or photons).

5.2.2 Software updates for Global Particle Flow $E_{\mathrm{T}}^{\mathrm{miss}}$

The bulk of the implementation is contained in the `MissingETAssociationMap`, and it will be described, in this section, only for electrons and links to PFOs, as an example. The implementation for other physics objects is similar. In addition, for Run-3, the PFO container (`<xAOD::PFOContainer>`) is being replaced by an improved version, the FlowElement (FE) container `<xAOD::FlowElementContainer>`. Direct links have been set also to this new container, with an identical implementation, only changing the name of the container itself.
The main difference with respect to the Run-2 strategy is that the association is simply based on the previously set `ElementLinks` between electrons and PFOs, rather than track and ΔR association. The `ElementLinks` can be dereferenced through `ElementLink<xAOD::ElectronContainer>` and `ElementLink<xAOD::PFOContainer>` to retrieve respectively the PFOs associated to the electron or the electrons associated to the PFO. The main function declaration takes as input the `xAOD::Egamma` object (i.e. electrons and photons), a pointer to the vector which will be filled with the constituents associated to it (`std::vector<const xAOD::IParticle*>`) and the container of the constituents given as input to the
`MissingETAssociationMap` (`const met::METAssociator::ConstitHolder`).

```
METEgammaAssociator::extractPFOsFromLinks(const xAOD::Egamma* eg,
        std::vector<const xAOD::IParticle*>& pfolist,
        const met::METAssociator::ConstitHolder& constits) const
```

The links information is extracted through the `ReadDecorHandle` function. The last step is the actual association of constituents to the electrons. The
`METAssociator::constitHolder` contains the `CHSParticleFlowObjects` constituents collection, wich is derived from the `JetETmissParticleFlowObjects` one by applying four-vector corrections and adding a decorator for the matching to Primary Vertex (PV) of

charged constituents. On the other side, the `ElementLinks` are included only to the `JetETmissParticleFlowObjects` collection. The first step towards using the `ElementLinks` is therefore an index-based matching between the two collections. Further selections on the association to PV ($|z_0 \sin\theta| < 2$ mm, with z_0 the distance of closest approach of the track to the hard-scatter primary vertex along the z-axis), as well as on the quality of the PFO and the E^{clus}/p^{trk} ratio (with E^{clus} the energy of the cluster before track energy subtraction and p^{trk} the track momentum) are applied for charged PFOs, while neutral PFOs are entirely included in the Association Map. Figure 5.4 shows the impact of including the previously mentioned selection criteria on charged PFOs, for a sample of Z bosons decaying into two electrons ($Z \to ee$): the mean value of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is slightly higher when links are used with no additional selection criterion (blue) with respect to standard reconstruction (red), suggesting a degradation in $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. The PV matching is the main selection in order to avoid this small loss in performance, while the E/p criterion gives a subdominant contribution to reach standard reconstruction performance.

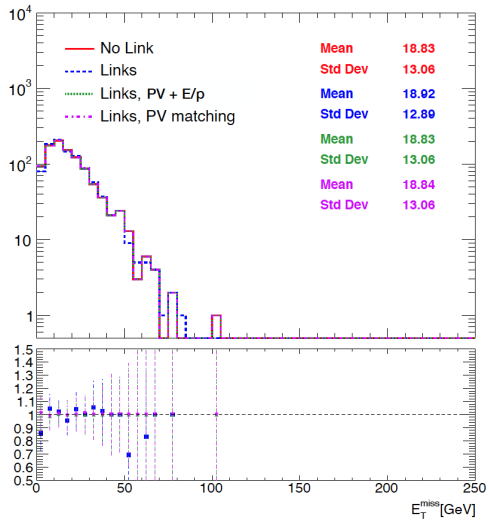


Figure 5.4: Comparison of reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for a $Z \to ee$ MC sample with the Run-2 PFlow strategy (in red) and with direct-links association, with or without the additional selection criteria on charged PFOs matching to PV and E/p.

5.2.3 Preliminary validation of the PFO reconstruction

The implementation described above is validated against the Run-2 PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ in multiple final states, with two different physics object selections and testing both the CST and the TST soft term (4.6).
The first performance studies were carried out on final states including electrons or photons, being these the first physics objects for which direct links to PFOs were implemented in the Athena software, while signatures with muons were not considered at first. In a following step, a general validation with increased statistics was performed, including also muons, the outcome of which is shown later.

In order to have a general overview of the performance in different conditions, the following topologies have been tested:

- $Z \to ee$ events: no true $E_{\mathrm{T}}^{\mathrm{miss}}$ is expected in this final state, making it the main scenario to easily investigate the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution;
- $W \to e\nu$ events: a true $E_{\mathrm{T}}^{\mathrm{miss}}$ is expected, thus allowing good linearity studies;
- $\gamma + jets$ events: this final state allows to investigate the performance for photons, with particular stress on the overlap removal with jets.

The isolation and identification Working Points (WP), as well as the kinematic selections applied to reconstructed particles in the final states, for the Tighter and Looser criteria are summarized in table 5.1.

In the plots, the results obtained with the Looser criteria are shown with empty square

Physics object	**Loose criteria**	**Tight criteria**
Electron	Loose ID WP $p_T > 5$ GeV	Medium ID WP $p_T > 10$ GeV
Photon	Loose ID WP $p_T > 5$ GeV	Tight ID WP $p_T > 25$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$	"PFlow" WP ($JVT > 0.2$)	

Table 5.1: Selection criteria for electrons, photons and $E_{\mathrm{T}}^{\mathrm{miss}}$ WP in the two configurations considered

markers, while the ones with tighter selection with fill triangle markers. The mean and standard deviation are reported for the Tighter and Looser selections respectively in the first two and last two boxes in the plots. Finally, the standard $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is shown in red, while the one based on direct links between physics objects and PFO in blue. All the $E_{\mathrm{T}}^{\mathrm{miss}}$ terms were analyzed, together with linearity and resolution as a function of the $\sum E_{\mathrm{T}}^{\mathrm{miss}}$ and of $< \mu >$, and 2D plots were built, in addition to distribution comparison, to provide an event-by-event comparison. For simplicity, only the global $E_{\mathrm{T}}^{\mathrm{miss}}$ plots are reported in the following. In addition, the plots with CST $E_{\mathrm{T}}^{\mathrm{miss}}$ are shown only for the first sample.

$Z \to ee$ final state

For $Z \to ee$ events, the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution comparison between links-based and Run-2 PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ is shown in Figure 5.5, the 2D plots are shown in Figure 5.6, while resolution studies are summarized in Figure 5.7. A minimal selection is applied, requiring exactly two electrons in the final state, satisfying the Tight or Loose selection criteria. An excellent agreement between the two $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions is observed, providing a good validation of the implementation, but also suggesting no evident performance improvement from the new $E_{\mathrm{T}}^{\mathrm{miss}}$ implementation.

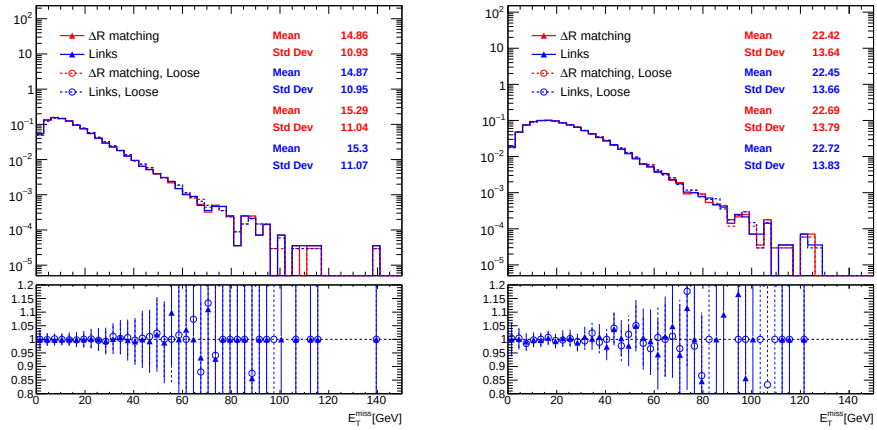


Figure 5.5: Distribution comparison, for $Z \rightarrow ee$ events, of $E_{\mathrm{T}}^{\mathrm{miss}}$ with the Run-2 PFlow reconstruction strategy (red) and the links-based one (blue) using the TST (left plots) and the CST (right plots) $E_{\mathrm{T}}^{\mathrm{miss}}$, for Looser and Tighter selection criteria. The first two mean and standard deviation values in top right box are associated to Tighter selections, the last two to Looser ones. The error bars include statistical uncertainties only.

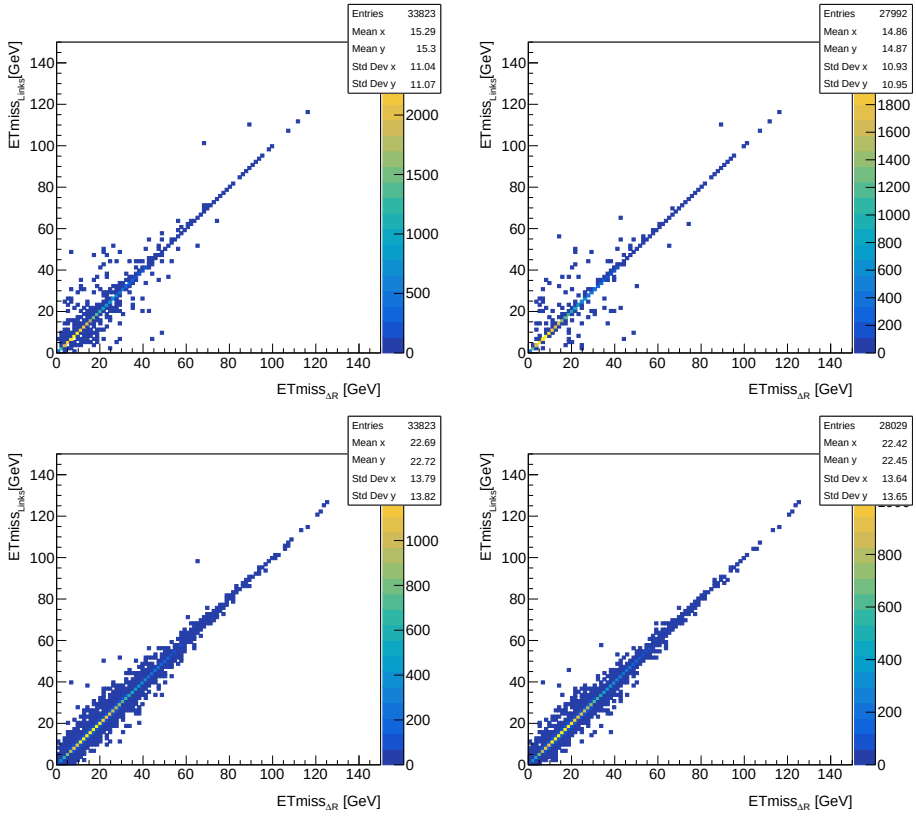


Figure 5.6: Event-by-Event comparison, for $Z \rightarrow ee$ events, of $E_{\mathrm{T}}^{\mathrm{miss}}$ with the Run-2 PFlow strategy (x axis) and the links-based one (y axis) using the TST Soft Term (top plots) and the CST one (bottom plots), for Looser (left plots) and Tighter (right plots) selection criteria.

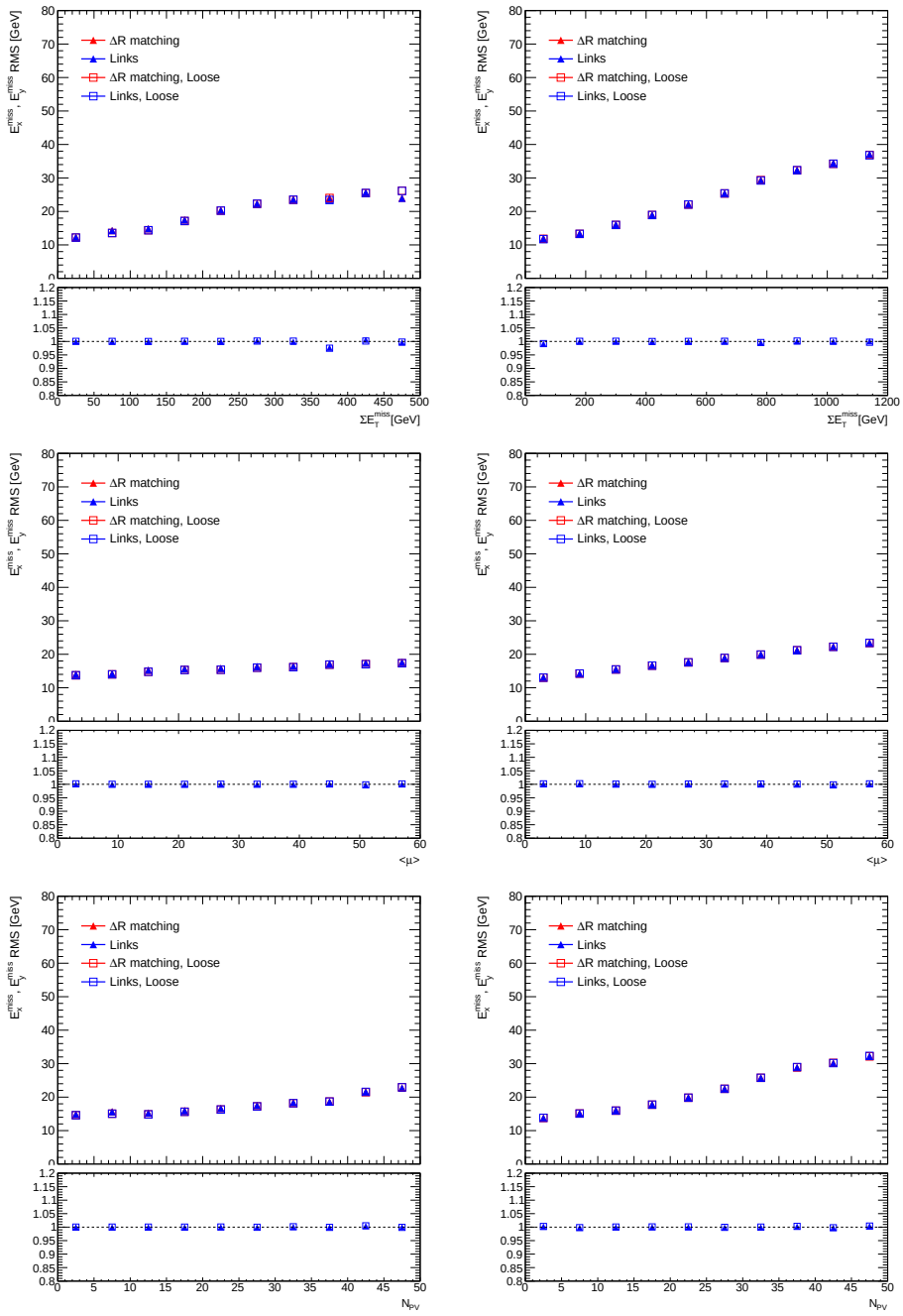


Figure 5.7: Resolution comparisons, for $Z \rightarrow ee$ events, of $E_{\mathrm{T}}^{\mathrm{miss}}$ with the Run-2 PFlow strategy (x axis) and the links-based one (y axis) using the TST (left plots) and the CST (right plots) $E_{\mathrm{T}}^{\mathrm{miss}}$, for Looser and Tighter selection criteria. The resolution is shown as a function of $\sum E_{\mathrm{T}}^{\mathrm{miss}}$ (top), $\langle\mu\rangle$ (middle) and N_{PV} (bottom).

$W \to e\nu$ final state

The MC simulations for the $W \to e\nu$ process come in slices of p_{T} of the W boson. For the performance studies, two slices have been used: one at low p_{T}, in the range 0-70 GeV, and one at very high $p_{\mathrm{T}} > 1$ TeV. Similar plots as for the $Z \to ee$ sample are shown to compare the $E_{\mathrm{T}}^{\mathrm{miss}}$ values with the two reconstruction strategies, in 5.8, 5.9, selecting events with exactly 1 electron satisfying the Tight or Loose selection criteria. The 2D plots show the event-by-event comparison of the linearity, rather than the $E_{\mathrm{T}}^{\mathrm{miss}}$. The linearity as a function of $E_{\mathrm{T}}^{\mathrm{miss},true}$ is shown in figure 5.10. Also in this case, an excellent agreement between the two is observed, with no relevant improvement from the new $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

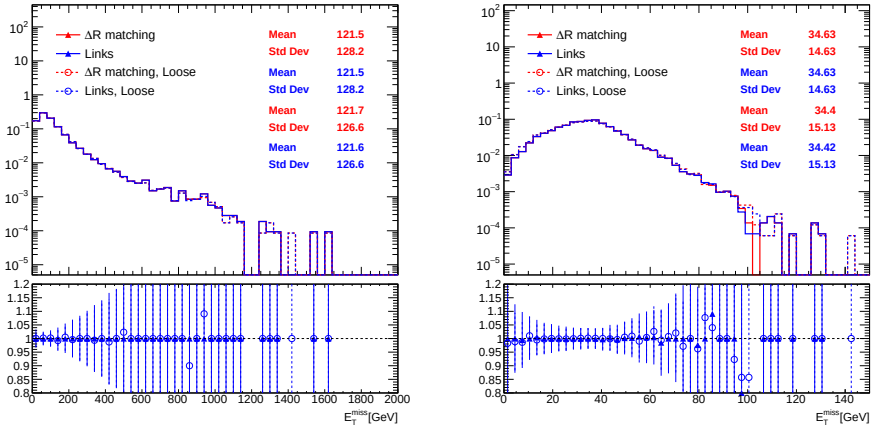


Figure 5.8: Comparison of TST $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions with the Run-2 PFlow and link-based $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, in $W \to e\nu$ final states with high W boson p_{T} (left), and low W boson p_{T} (right), for Looser and Tighter selection criteria. The first two mean and standard deviation values in top right box are associated to Tighter selections, the last two to Looser ones. The error bars include statistical uncertainties only.

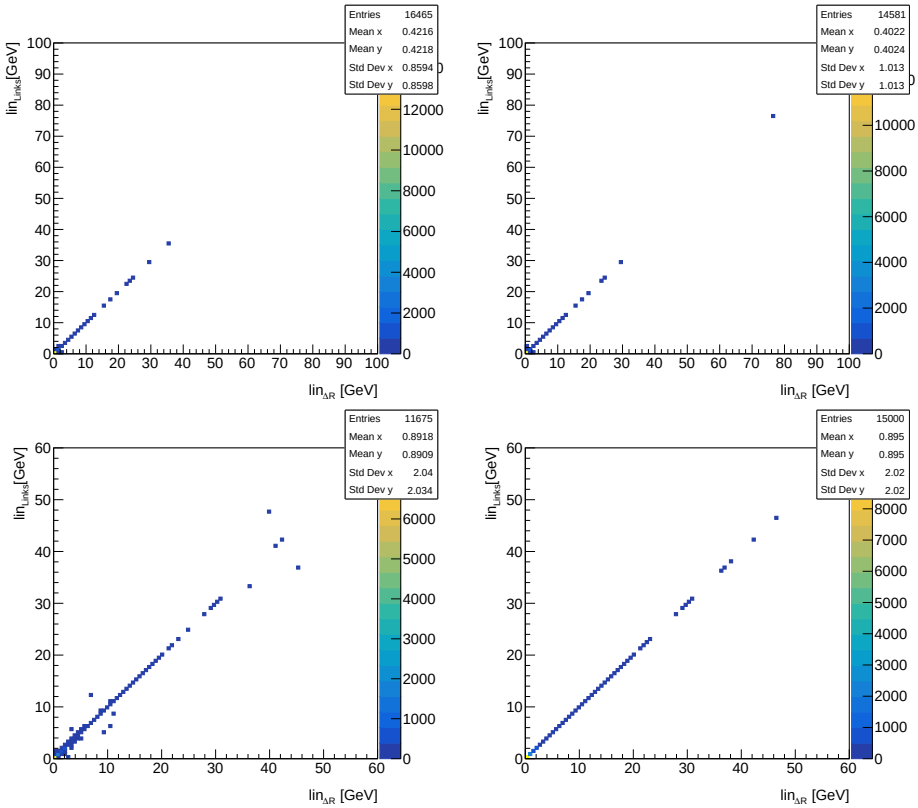


Figure 5.9: Event-by-Event comparison of TST $E_{\mathrm{T}}^{\mathrm{miss}}$ with the Run-2 PFlow strategy (x axis) and the links-based one (y axis) in $W \to e\nu$ final state with low W boson p_{T} (top plots), and high W boson p_{T} (bottom plots), for Looser (left plots) and Tighter (right plots) selection criteria.

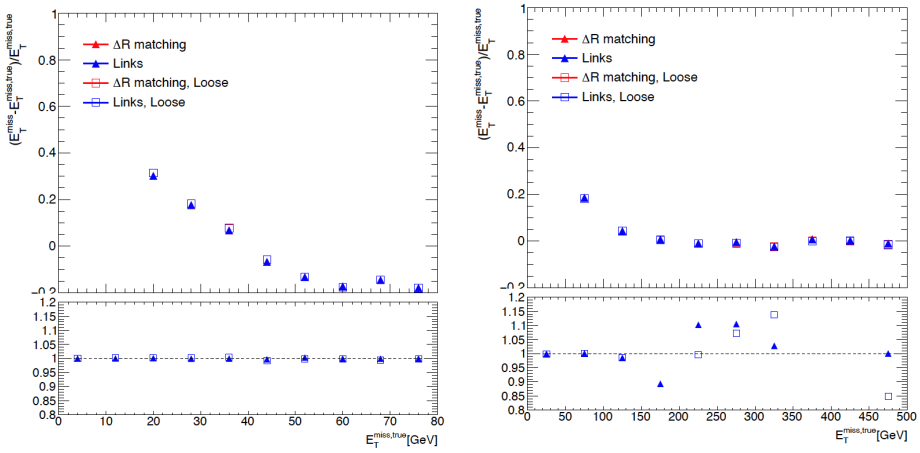


Figure 5.10: Comparison of the linearities for the Run-2 PFlow and links-based TST $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions, in $W \to e\nu$ final states with low W boson p_{T} (left plot) and high W boson p_{T} with (right plot), for Looser (left plots) and Tighter (right plots) selection criteria.

γ+jets final state

Finally, γ+jets events are studied. Similarly to $W \to e\nu$ MC simulations, also γ+jets ones are splitted into slices of photon p_{T}. In this case, the considered ranges are 35-50 GeV (only TST $E_{\mathrm{T}}^{\mathrm{miss}}$ tested) and 140-280 GeV, and exactly one selected photon is required, which results in no events passing Tight selections in the low p_{T} slice. The $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions and event-by-event comparisons between standard and links-based $E_{\mathrm{T}}^{\mathrm{miss}}$ recostructions are shown in Figure 5.11 and Figure 5.12 respectively, while the resolution plots are shown in figures 5.13 for low p_T and high p_T cases, with TST soft term only. Similar conclusions as for the two previous cases can be drawn, despite larger differences observed between the two reconstruction strategies, due to the higher jet multiplicity in this topology, enhancing the probability of inaccurate ΔR-based associations of PFOs to the physics objects (more specifically photons, in this case). The large discrepancy observed in one of the bins of the resolution plot as a function of $\sum E_{\mathrm{T}}^{\mathrm{miss}}$ (top left of Figure 5.13) arises from one single event, as highlighted in Figure 5.14 showing the E_x^{miss} variable distribution in that $\sum E_{\mathrm{T}}^{\mathrm{miss}}$ bin, for Tight physics objects selections with the "standard" association (in blue) and with the links-based one, in red.

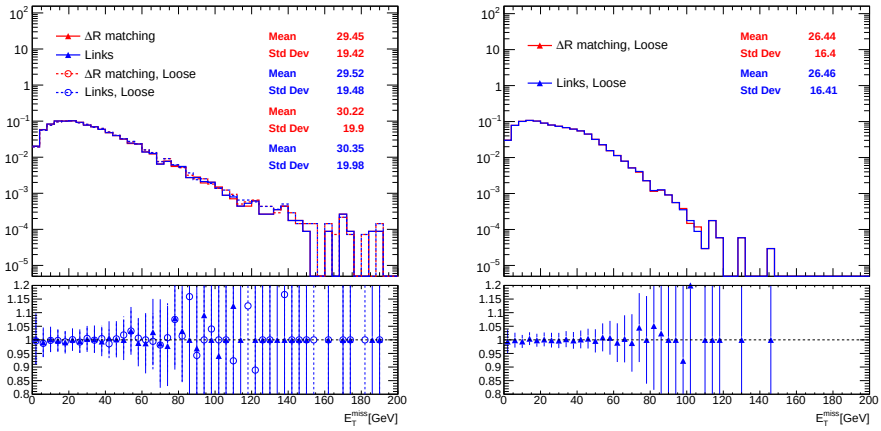


Figure 5.11: Comparison of TST $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions with the Run-2 PFlow and links-based $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, in γ+jets final states with high photon p_{T} (left) and low photon p_{T} with (right), for Looser and Tighter selection criteria. The first two mean and standard deviation values in top right box are associated to Tighter selections, the last two to Looser ones. The error bars include statistical uncertainties only.

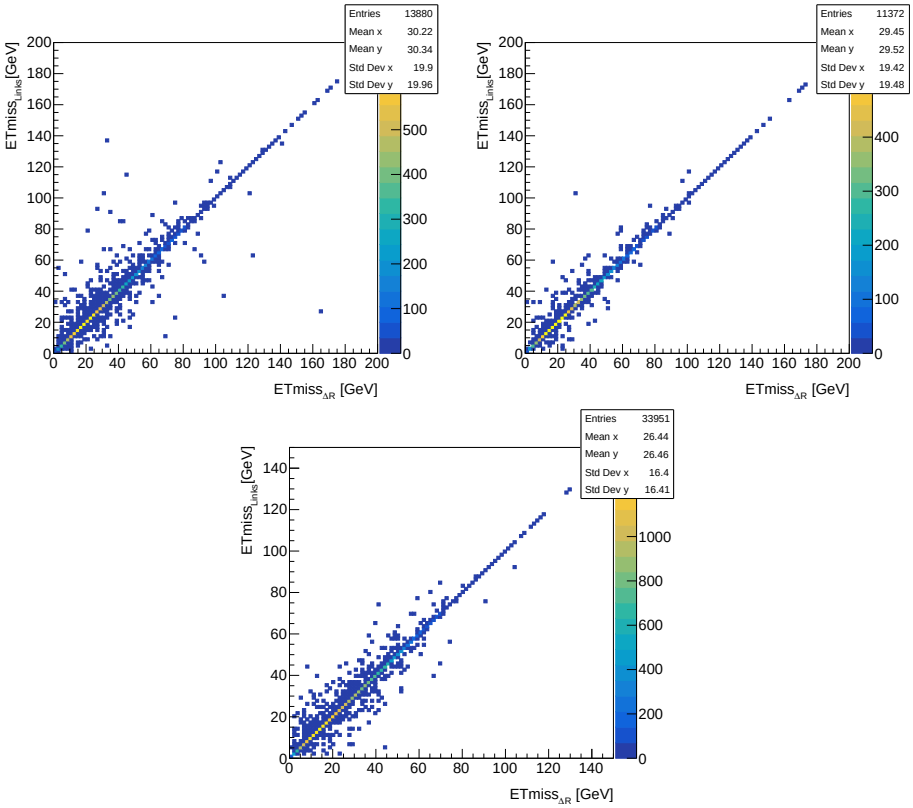


Figure 5.12: Event-by-Event comparison of TST $E_{\mathrm{T}}^{\mathrm{miss}}$ with the Run-2 PFlow reconstruction strategy (x axis) and the links-based (y axis) in γ+jets final state with high photon p_{T} with (top plots) and low photon p_{T} (top plots), for Looser and Tighter selection criteria.

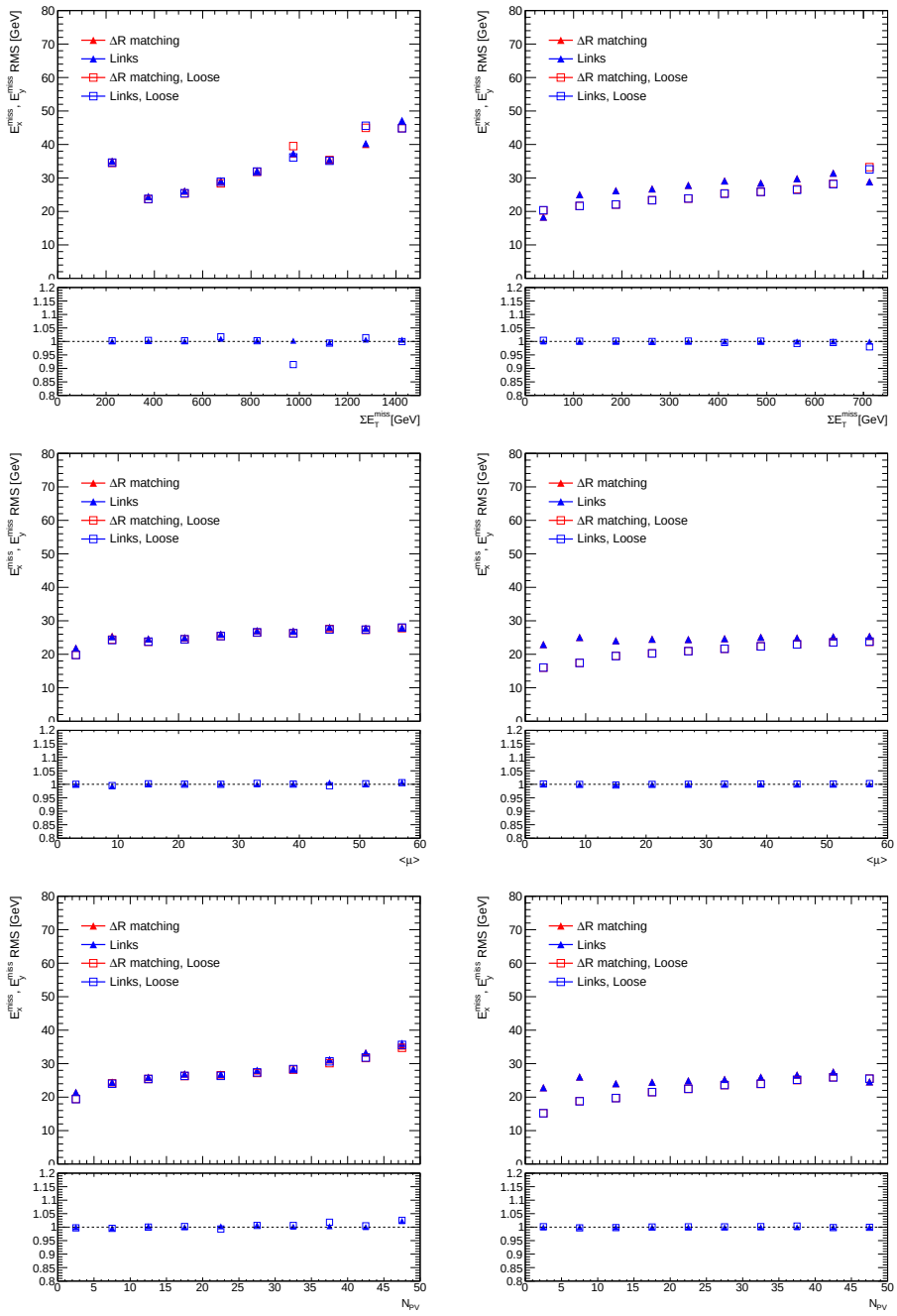


Figure 5.13: Resolution comparisons, in γ+jets sample, of TST $E_{\mathrm{T}}^{\mathrm{miss}}$ with the Run-2 PFlow reconstruction strategy (red) and the links-based (blue), with high photon p_{T} (left), and low photon p_{T} (right), for Looser and Tigher selection criteria. The resolution is shown as a function of $\sum E_{\mathrm{T}}^{\mathrm{miss}}$ (top), $\langle\mu\rangle$ (middle) and N_{PV} (bottom).

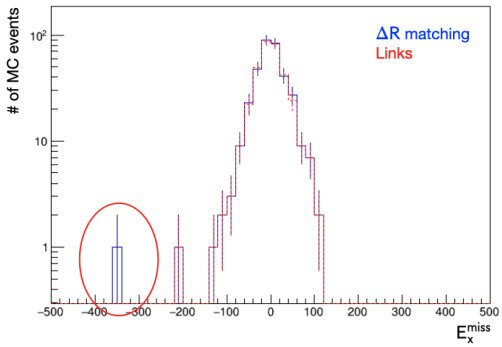


Figure 5.14: Comparison of E_x^{miss} distributions for $900 < \sum E_{\mathrm{T}}^{\mathrm{miss}} < 1050$ GeV, in γ+jets sample, with the Run-2 PFlow reconstruction strategy (red) and the links-based (blue), with high photon p_{T}

5.2.4 Final validation with higher statistics

A further validation is performed with a higher number of MC events, and also including direct links between muons/taus and PFOs, and links to FEs (Flow Elements) as well. An official derivation production is used, with both the Run-2 PFlow and the links-based $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions stored. Samples of $H \to \gamma\gamma$, $Z \to \tau\tau$ and $t\bar{t}$ are tested, to validate electron, photon, τ lepton and μ/e lepton links respectively. In the plots in the following figures 5.15–5.17, a comparison between 4 cases is shown:

- PFO: $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction based on the standard association between physics objects and PFOs;
- FE: $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction based on the standard association between physics objects and FlowElements;
- PFO+GPF links: new $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction based on the direct links between physics objects and PFOs;
- FE+FE links: new $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction based on the direct links between physics objects and FlowElements.

The ratios between Run-2 PFLow and links-based $E_{\mathrm{T}}^{\mathrm{miss}}$ using PFO or FE are shown (Links/NoLinks), as well as the ratios between PFO and FE with standard and links-based $E_{\mathrm{T}}^{\mathrm{miss}}$ (FE/PFO). Both TST and CST $E_{\mathrm{T}}^{\mathrm{miss}}$ are shown in the following plots. Links to FE or to PFO are expected to be identical. The observed differences are due to independent jet calibrations, affecting the Jet and Soft Term, while all other terms are perfectly consistent, as shown in Figure 5.18 in some example cases, for TST $E_{\mathrm{T}}^{\mathrm{miss}}$. The Jet, Soft and Photon terms are shown for $H \to \gamma\gamma$, while Electron/Muon and Tau Terms are shown for $t\bar{t}$ and $Z \to \tau\tau$ terms respectively.

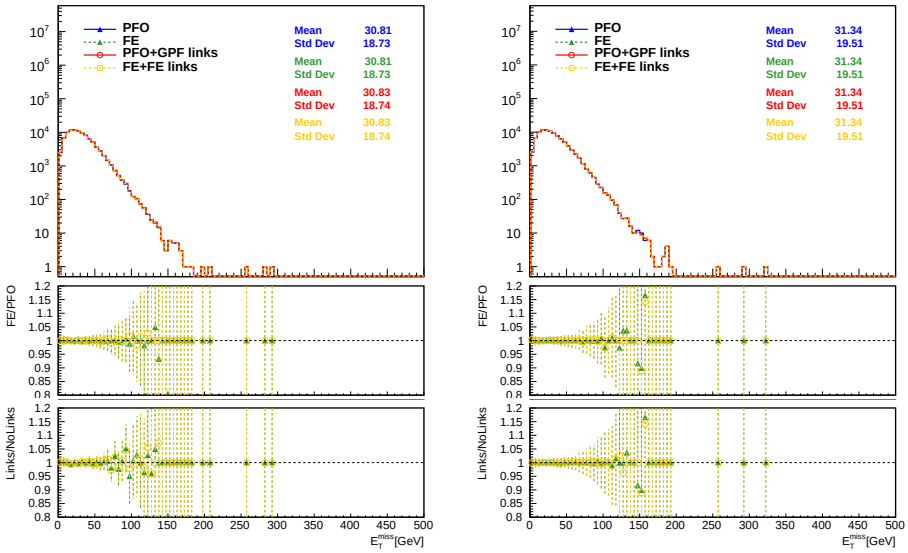


Figure 5.15: Comparison of $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions with Run-2 PFLow and links-based reconstruction, both using PFO or FE collections, for a $\gamma\gamma$ sample. The error bars include statistical uncertainties only.

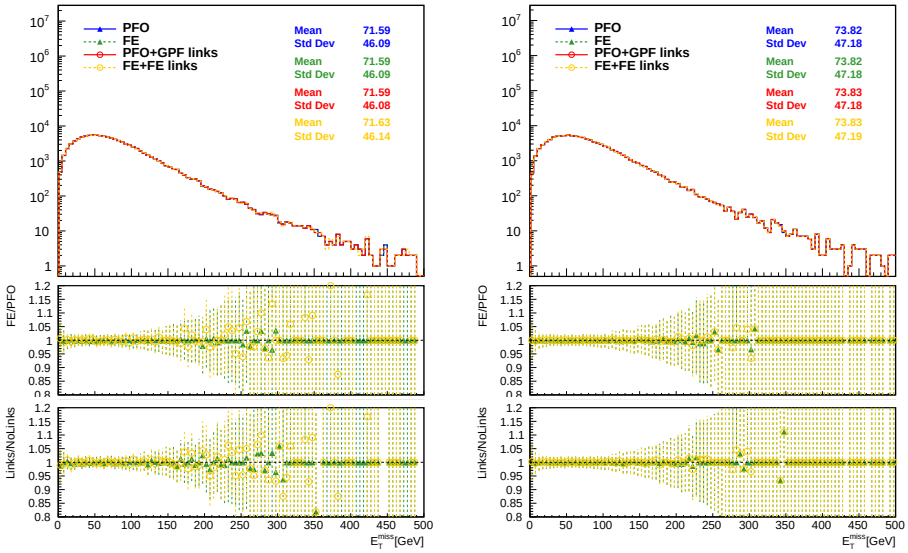


Figure 5.16: Comparison of $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions with Run-2 PFLow and links-based reconstruction, both using PFO or FE collections, for a $t\bar{t}$ sample. The error bars include statistical uncertainties only.

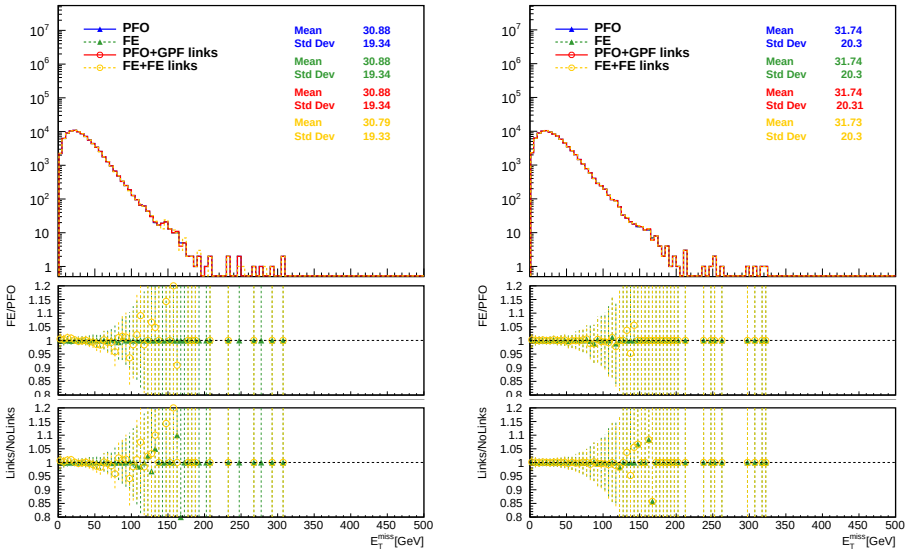


Figure 5.17: Comparison of $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions with Run-2 PFLow and links-based reconstruction, both using PFO or FE collections, for a $Z \to \tau\tau$ sample. The error bars include statistical uncertainties only.

The conclusions drawn from the preliminary studies described in the previous section are confirmed: similar performance are observed between the two reconstruction strategies. In addition, the implementation for `FlowElement` is validated against the PFO one.

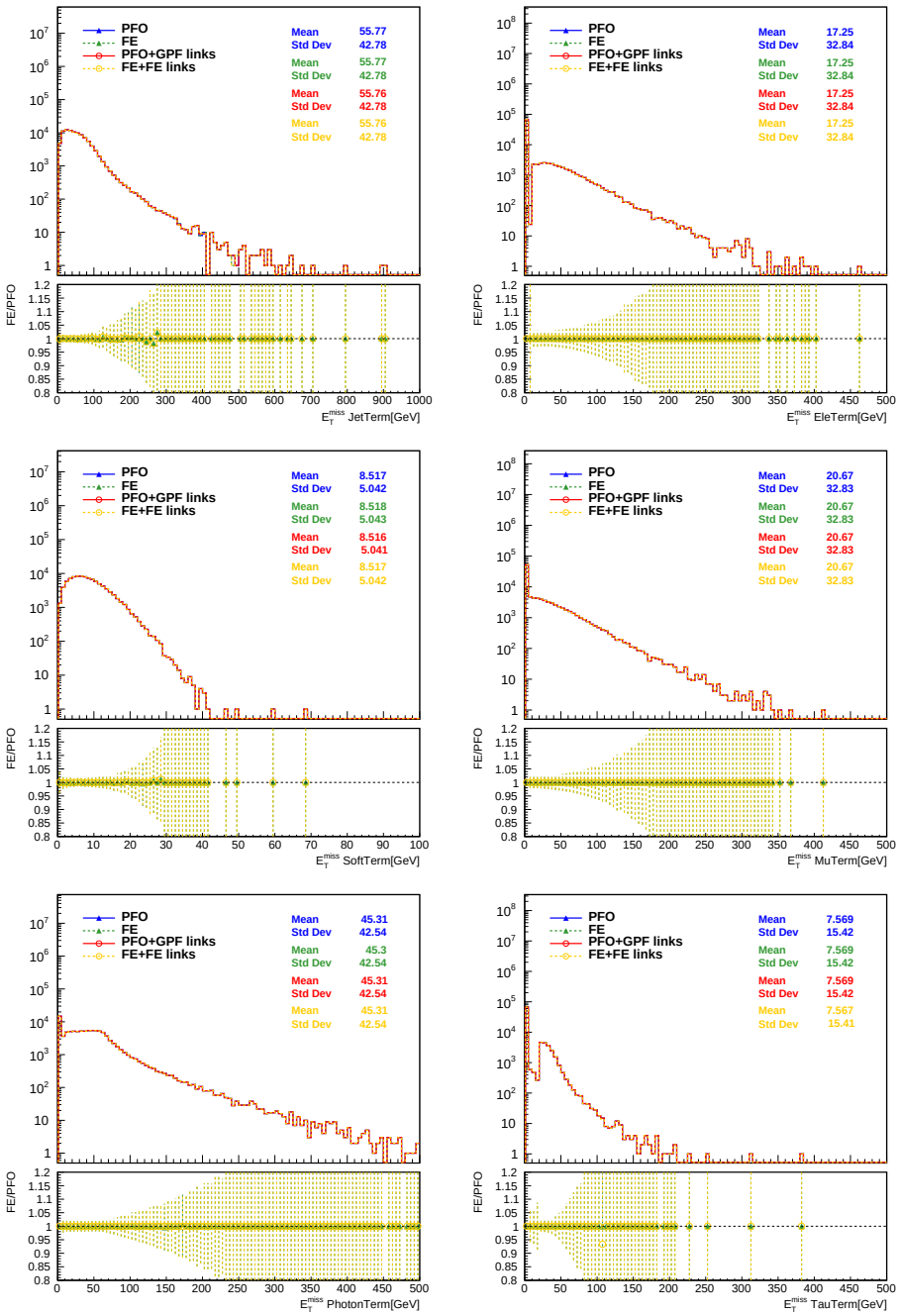


Figure 5.18: Jet, Soft and Photon Term for $H \to \gamma\gamma$ (left), Electron and Muon Terms for $t\bar{t}$ (right top and central) and Tau Term for $Z \to \tau\tau$ (right bottom). The error bars include statistical uncertainties only.

5.3 Improving the overlap removal in E_T^{miss} reconstruction

The Global Particle Flow strategy gives also an opportunity to potentially improve the overlap removal between jets and other reconstructed physics objects. In the standard PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, the $E_{\mathrm{T}}^{\mathrm{miss}}$ Jet Term is built from the reconstructed jets in the `AntiKt4EMPFlowJet` container, and the overlap removal with other physics objects is handled within the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithm according to the following procedure:

- If a reconstructed physics object (with selection criteria and WP defined at analysis-level) overlaps with a jet, the energy of the overlapping object is subtracted from the jet at constituent level, i.e. through PFO/FE, or track/cluster depending on whether the standard or the link-based $E_{\mathrm{T}}^{\mathrm{miss}}$ is employed.

- If the residual energy of the jet is higher than 20 GeV, the energy is rescaled to recover calibration, and the jet is included into the Jet Term. This is the case for events with "partial" overlap between a jet and another particle.

- If the residual energy of the jet is smaller than 20 GeV, its contribution enters the Soft Term, with energy computed at constituent level (i.e. without calibration). This is the case for events with an "exact" overlap, likely meaning that the jet is not an hadronic jet, but rather it is associated to the overlapping particle, since each physics object is also reconstructed as a jet.

This procedure is represented in the left part of Figure 5.19, in the case of "exact" overlap, in which the jet is completely removed, and "partial" overlap. In this section, an alternative approach is investigated, illustrated on the right of the same figure:

- The Jet Term is built starting from a new collection of jets, in the following called "Overlap Removed (OR)" jets, reconstructed only from constituents (neutral and charged PFO, with easy generalization to FE) that have been not previously associated to other selected physics objects entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation (with selection criteria for the physics objects defined at analysis level).

- In this case, constituents associated to selected physics objects are not given as input to jet reconstruction at all, thus allowing to naturally handle both exact and partial overlap removal. In case of partial overlap, the jet will be correctly calibrated, without need for any energy subtraction and subsequent correction to recover jet calibration.

- Any other "soft" constituent still enters the Soft Term (either in the Soft Core or in the Miscellaneous term), or potentially feeds a neighbour jet.

5.3.1 Technical implementation

The workflow for this alternative Jet Term reconstruction is represented in figure 5.20. The bulk of the implementation is included in the `METUtilities` package in the Athena framework.

- In `METMaker`: added `retrieveOverlapRemovedSignals(...)` function relying on `MissingETAssociation::GetOverlapRemovedSignals(...)` to extract PFOs not associated to selected objects

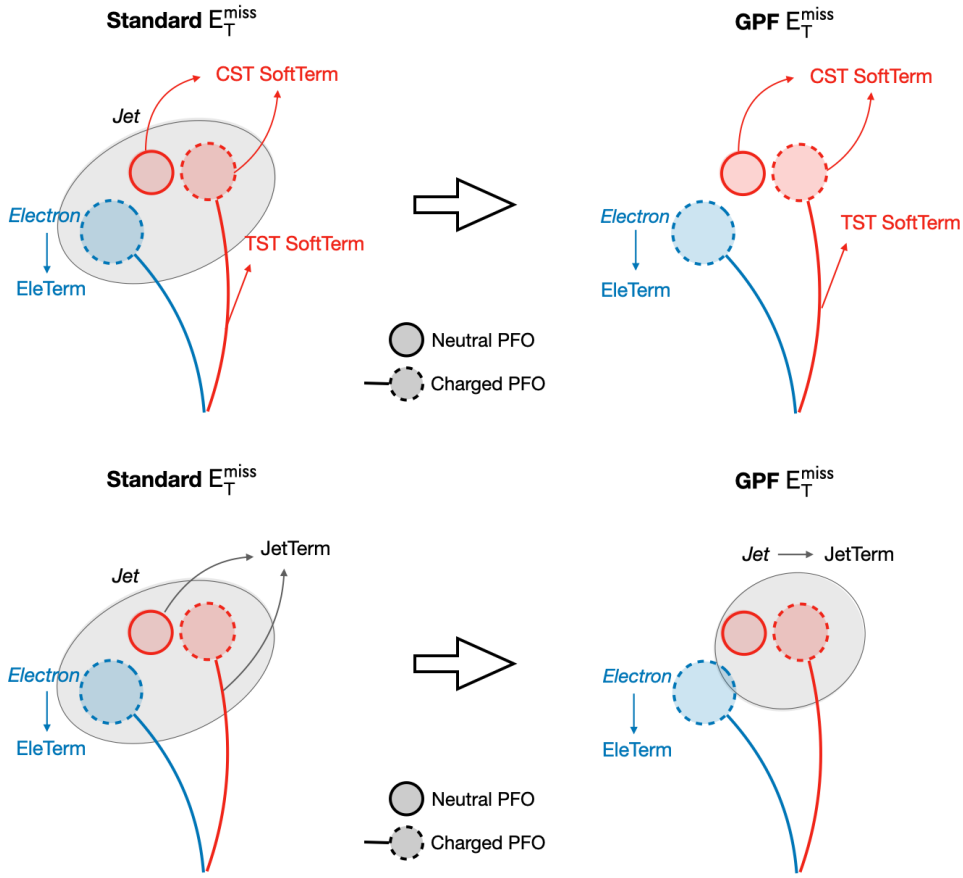


Figure 5.19: Comparison between standard Jet Term and “”Overlap Removed one for exact overlap (top) and partial overlap (bottom) cases.

- New MET maker algoritm (`ORMETMakerAlg`): to be run twice. The first run (with option `doRetrieveORsignals = true`) is needed to extract the “OR” PFO collection after building the other terms (except for muon term for which the overlap removal strategy present some additional complexity). The second run (`doORMET = true`) performs the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction using the new jet collection built, in an intermediate step, from the “OR” PFO collection.
- The full procedure is implemented in a python script (`run_ORMETReco.py`)

This implementation was tested for overlap removal against electrons, τ leptons and photon. The μ-jet overlap, is more critical to handle (see Section 4.6), as shown later, therefore the standard procedure is used in this first validation and performance studies, by retaining constituents associated to muons when building the “OR” PFO collection.

5.3.2 Validation and performance studies

The implementation is validated, and its performance are investigated, in $Z \to ee$, $t\bar{t}$ and γ+jets topologies, for both the TST and the CST Soft Term. The $t\bar{t}$ signature is

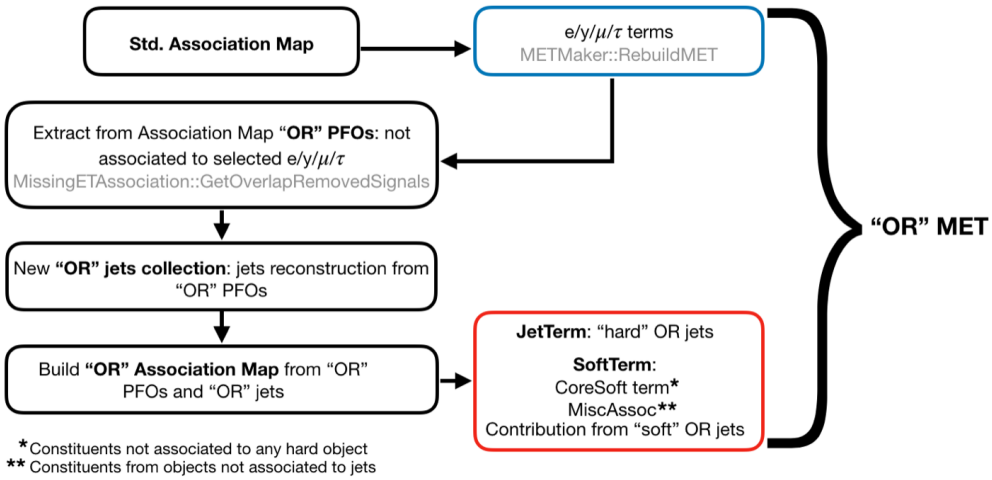


Figure 5.20: Workflow for $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction with 'Overlap Removed" Jet Term.

considered in order to study the jet-electron overlap removal in a high jet multiplicity environment.
The comparison has been performed inclusively and for three different categories of events:

No overlap between jets and other selected objects: no difference is expected between standard $E_{\mathrm{T}}^{\mathrm{miss}}$ and "OR" one, thus being a good data subset for validation.

Exact overlap between jets and other selected objects: similar performance are expected, with potentially some small differences in the Jet Term due to soft constituents from overlapping (and therefore removed) jets which, in the "OR" $E_{\mathrm{T}}^{\mathrm{miss}}$ can be included in a nearby jet. If the CST Soft Term is used, a test performed without applying jet calibration is expected to show that the Soft Term would compensate for the Jet Term differences between the two $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions, since the soft constituents will still enter the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation in the Soft Term. On the other side, the TST SoftTerm can happen not to compensate Jet Term differences, due to neutral PFO or not PV matched charged PFO not feeding the Soft Term. This behaviour is demonstrated in figure 5.21, for a γ+jets sample, with uncalibrated jets.

Partial overlap between jets and other selected objects: this is the case were most of the potential performance improvement is expected, due to a better handling of calibrated signals. As a matter of fact, instead of rescaling the jet energy, after subtraction of the overlapping contribution, to recover the jet calibration scale, the calibration is performed directly on the jet reconstructed without including the overlapping constituents, thus allowing a more accurate calibration.

The physics objects selection criteria employed for the following studies are summarized in table 5.2.

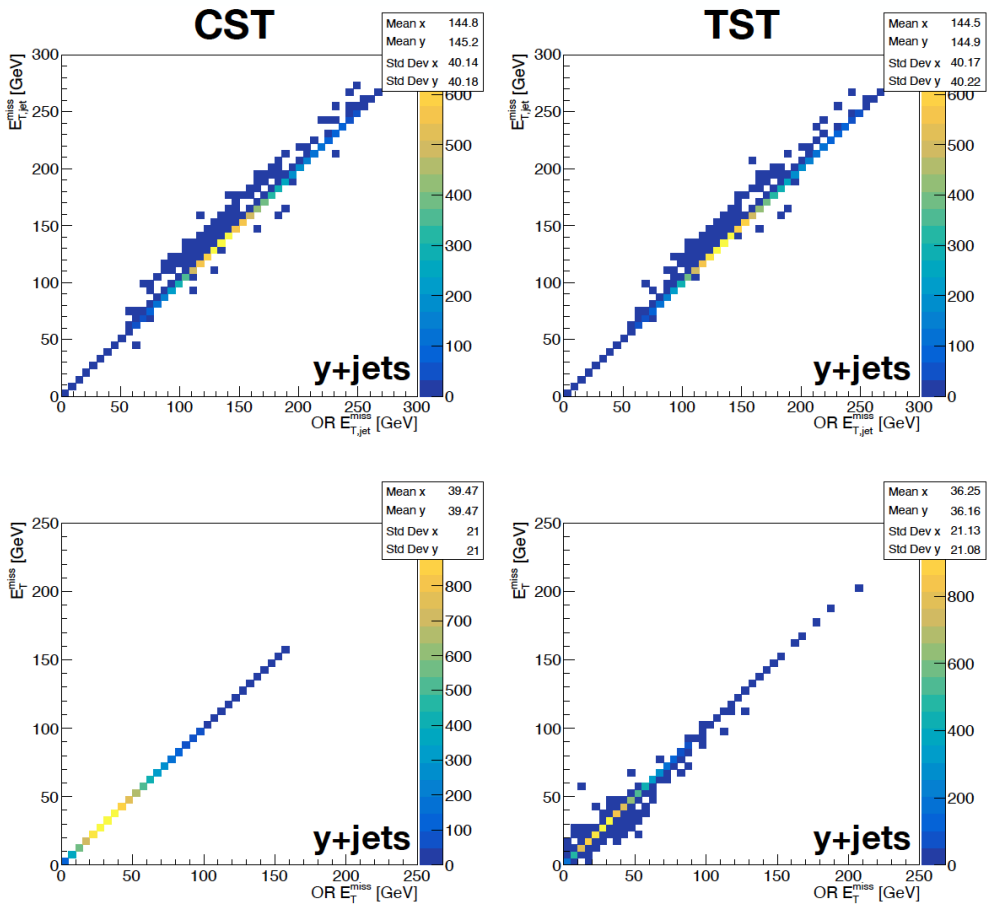


Figure 5.21: Event-by-event comparison between the "OR" and standard Jet Terms (top plots) and total $E_{\mathrm{T}}^{\mathrm{miss}}$ (bottom plots) in γ+jets events. The CST $E_{\mathrm{T}}^{\mathrm{miss}}$ is shown in the left plots, while TST one on the right.

Physics object	Criteria
Electron	Medium ID WP
	$p_{\mathrm{T}} > 10$ GeV
Muon	Medium ID WP
	$p_{\mathrm{T}} > 10$ GeV
Photon	Tight ID WP
	$p_{\mathrm{T}} > 10$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$	"PFlow" WP ($JVT > 0.2$)

Table 5.2: selection criteria for electrons, photons, muons and $E_{\mathrm{T}}^{\mathrm{miss}}$ WP

Comparison of standard and "Overlap Removed" jets containers

As a first step, the jet reconstruction procedure is validated with a γ+jets sample, with photon p_{T} in the 140-280 GeV range, and $Z \to ee$ events. The left plot of figure 5.22

demonstrates that, as expected, the jets p_{T} from the original (before overlap removal) and "overlap removed" containers are exactly the same when no overlap exists (here the results are shown before calibration, to rule out small discrepancies arising from the random seed used in the calibration procedure). On the other side, the middle and right plot in figure 5.22 show the impact on the p_{T} distribution of calibrated jets, for γ+jets and $Z \to ee$ final states respectively, of jets reconstruction only from constituents not associated to other selected physics objects: as expected, in the γ+jets events the peak at about 140 GeV (the p_T threshold of the photons in the MC sample used), given by jets actually associated to photons, is correctly removed, and similar behaviour is observed in $Z \to ee$ events, with the removal of the feature due to jet associated to electrons.

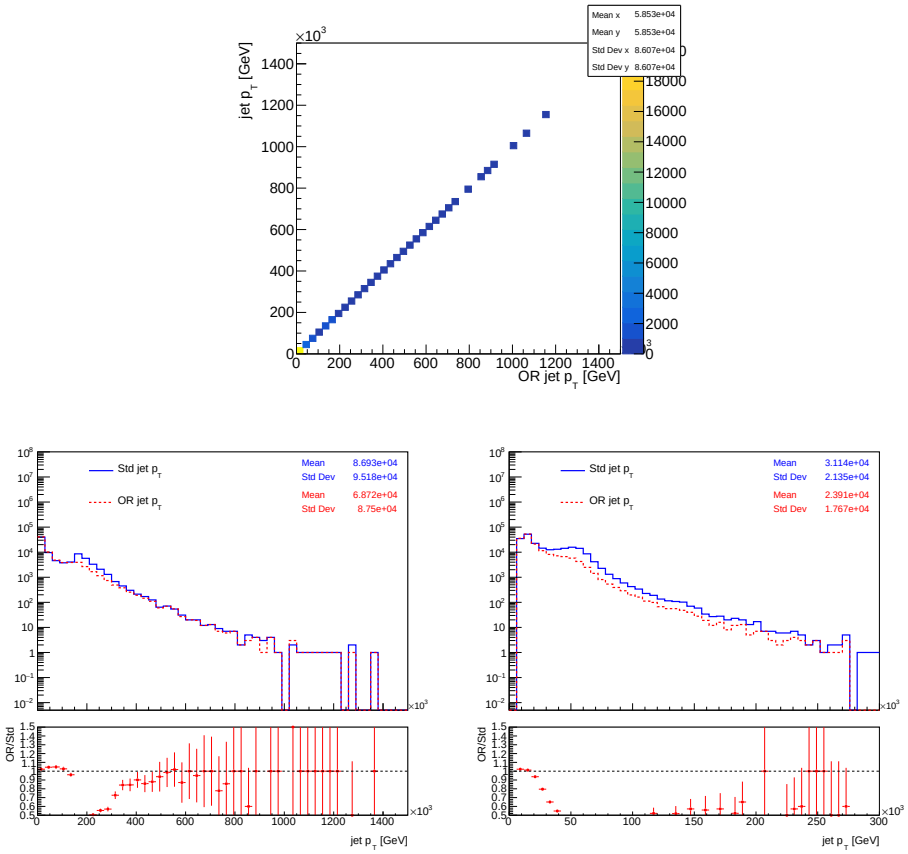


Figure 5.22: Event-by-event comparison of the standard-jets and "OR" jets p_{T} in γ+jets (top) and comparison of the corresponding distributions in γ+jets (bottom left) and $Z \to ee$ (bottom right) events.

$Z \to ee$ final state

The $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions are compared for events with no overlap, events with exact overlap and events with partial overlap, in figure 5.23, using TST Soft Terms. There is an overall optimum consistency between the two $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions in all cases, suggesting no relevant improvement from the modified treatement of the overlap removal in this final

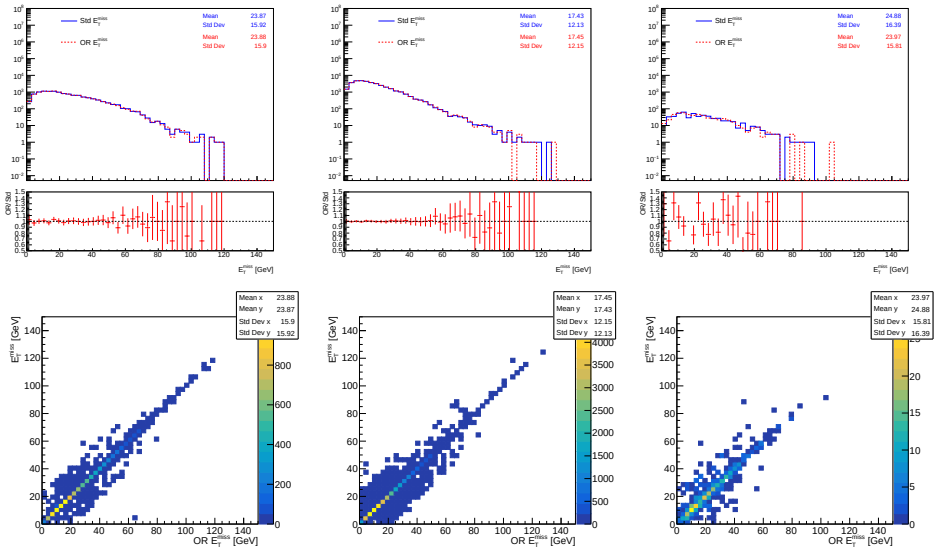


Figure 5.23: Comparison of $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction with "OR" Jet Term and standard one, in $Z \to ee$ events. The global $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions are compared, as well as the event-by-event values, for events with no overlap (left), with exact overlap (center), and with partial overlap (right) between jets and other physics objects, both considering the TST $E_{\mathrm{T}}^{\mathrm{miss}}$.

state. A slight decrease of the mean value and the removal of few events in tails can be observed, though, in the partial overlap case. The differences observed in the case without any overlap arise from calibration and disappear if uncalibrated jets are considered. An example is given in Figure 5.24, for the TST case. To give a complete picture of constituent level differences, results are shown, using uncalibrated jets, also for events with exact and partial overlap .

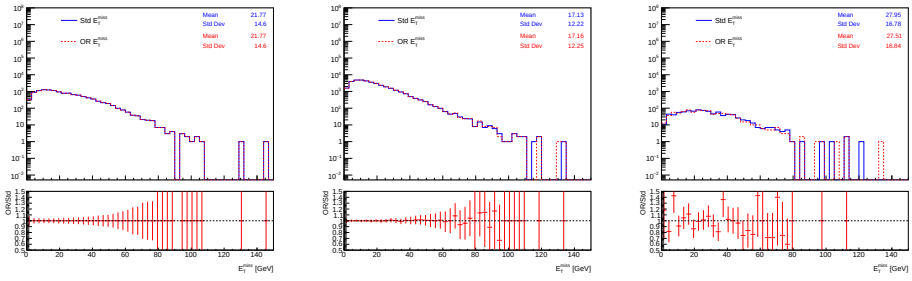


Figure 5.24: Comparison of the standard and "OR" $E_{\mathrm{T}}^{\mathrm{miss}}$ in $Z \to ee$ events, using non-calibrated jets to rule out differences arising from calibration random seed.

$t\bar{t}$ final state

The same results as in the previous section are shown for $t\bar{t}$ events in Figure 5.25. Since the standard jet-muon overlap removal is retained, as mentioned before, events with muons in the final state are discarded, to focus on the impact of the new overlap removal between

jets and electrons in an environment with higher jets multiplicity than the $Z \to ee$ final state. Also in this case, very similar performance are observed between the two $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions.

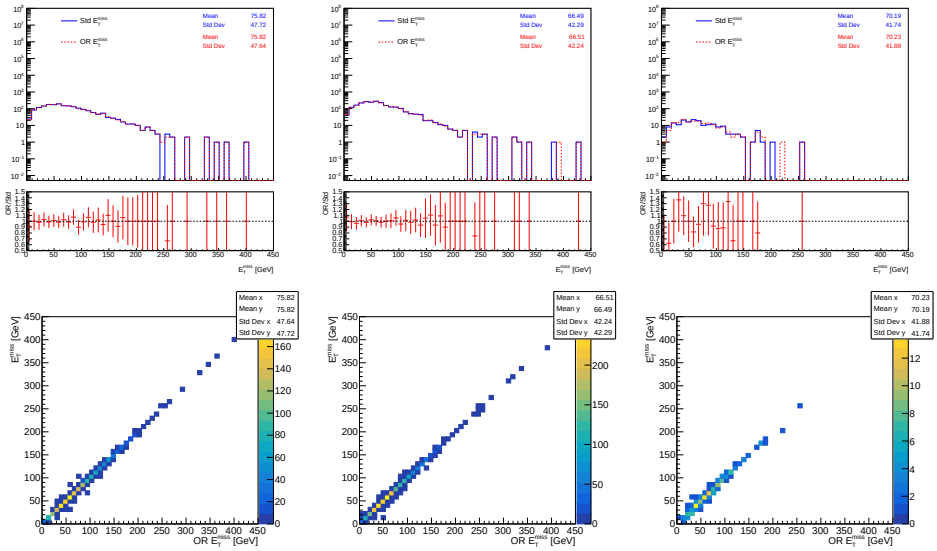


Figure 5.25: Comparison of $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction with "OR" Jet Term and standard one, in $t\bar{t}$ events. The global $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions are compared, as well as the event-by-event values, for events with no overlap (left), with exact overlap (center), and with partial overlap (right) between jets and other physics objects, with TST $E_{\mathrm{T}}^{\mathrm{miss}}$.

γ+jets final state

For γ+jets final states, similar considerations as for the previous processes are valid (Figure 5.26), but some hints of more relevant performance improvements can be observed in the partial overlap case. In this case, the results with non-calibrated jets are shown in Figure 5.27 for both TST and CST $E_{\mathrm{T}}^{\mathrm{miss}}$, to allow a comparison with fully calibrated events and provide a qualitative idea of the impact of calibration which is expected to enhance the "OR" $E_{\mathrm{T}}^{\mathrm{miss}}$ improvement. Whilst the relative difference between the mean values with the two reconstructions is of about 3% for TST and 1.7% for CST $E_{\mathrm{T}}^{\mathrm{miss}}$ with non-calibrated jets, a slightly higher difference (towards improved $E_{\mathrm{T}}^{\mathrm{miss}}$) is observed when the calibration is applied, reaching about 4% for TST $E_{\mathrm{T}}^{\mathrm{miss}}$ and 2.5% for the CST one. The statistics is too poor to draw conclusions, but observations at least go in the expected direction.

The muon-jet overlap removal

The muon-jet overlap is particularly tricky, due to the variety of possible cases, as briefly described in the paragraph "Overlap Removal" of section 4.6. In figure 5.28, the impact of applying the new overlap removal strategy in $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction also for the jet-muon overlap is shown in γ+jets events, with CST (top) and TST (bottom) Soft Term. The performance are clearly negatively affected with respect to the standard procedure,

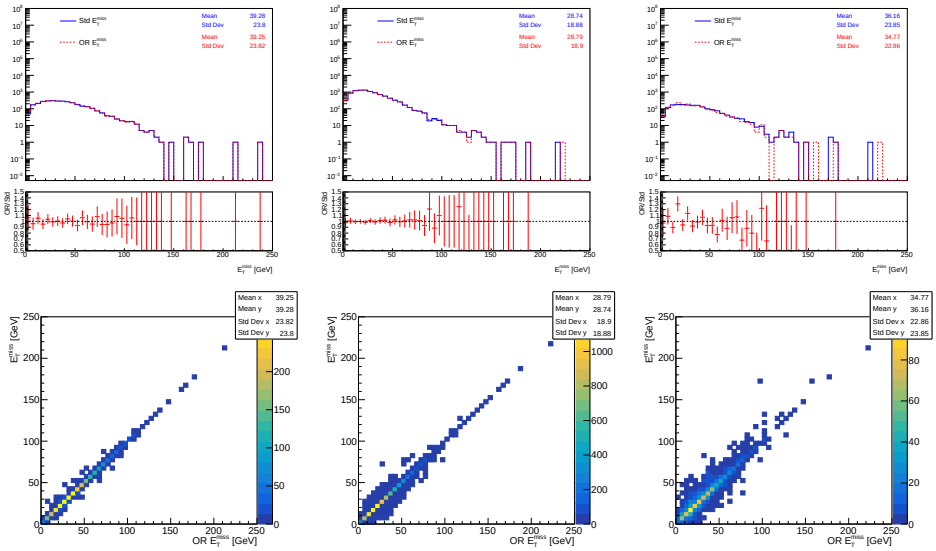


Figure 5.26: Comparison of $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction with "OR" Jet Term and standard one, in γ+jets events. The global $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions are compared, as well as the event-by-event values, for events with no overlap (left), with exact overlap (center), and with partial overlap (right) between jets and other physics objects, both considering the TST $E_{\mathrm{T}}^{\mathrm{miss}}$.

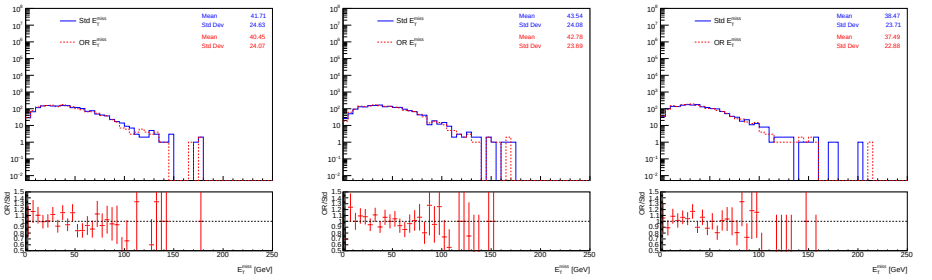


Figure 5.27: Comparison of the standard and "OR" TST (left) and CST (center) $E_{\mathrm{T}}^{\mathrm{miss}}$ in $\gamma+jets$ events with partial overlap, using non-calibrated jets to rule out the effect of calibration. In the right plot, the CST $E_{\mathrm{T}}^{\mathrm{miss}}$ with calibrated jet is shown, for reference, while the TST $E_{\mathrm{T}}^{\mathrm{miss}}$ case with calinìbrated jet is shown in the top right plot of Figure 5.26

probably due to the incorrect handling of hard photons radiation off muons. As a matter of fact, in these cases the jet should be retained, as the energy carried by the radiated photon is not included in the total energy measurement of the muon and would get lost. On the other hand, potential improvements in the muon-jet performance could be expected in events in which the jets should indeed be removed, i.e. in case of significant muon energy losses (already accounted for as part of the muon transverse momentum) or events where a pile-up jet is mistakenly tagged as originating from a hard-scatter due to overlapping muon tracks. Further studies should be carried out in order to better identify the categories of muon-jet overlaps which can be correctly handled by the "OR" $E_{\mathrm{T}}^{\mathrm{miss}}$ strategy.

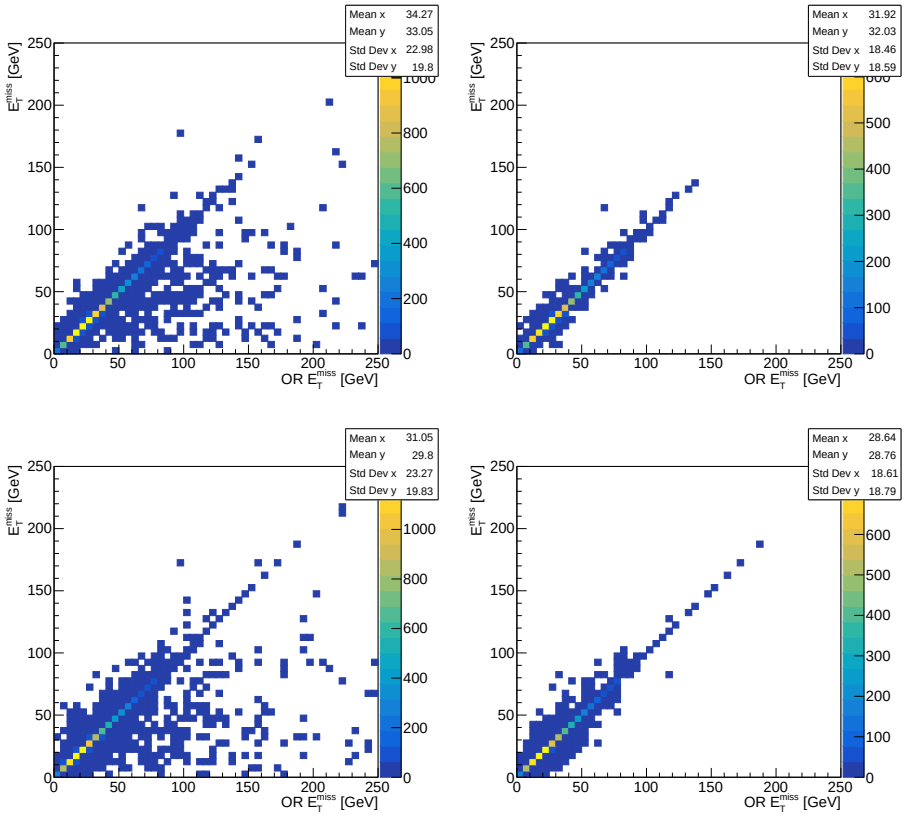


Figure 5.28: Event-by-event comparison of the standard and "OR" $E_{\mathrm{T}}^{\mathrm{miss}}$ in γ+jets sample, including (left) or removing (right) events with some muon-jet overlap. The CST $E_{\mathrm{T}}^{\mathrm{miss}}$ is shown in the top plots, while TST one in the bottom.

Chapter 6

Phenomenology of simplified DM model and Dark Photon production in Higgs boson decay at LHC

Profiting of the high energies in the centre-of-mass and high reached luminosities, proton-proton collisions at the LHC have good potentialities to produce DM or other BSM particles in the Dark Sector up to the TeV scale. As previously mentioned, the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ would be a clear signature of such particles, since they are predicted to be weakly coupled to the SM sector, thus being invisible to the detector.
Several searches are conducted, within the ATLAS and CMS Collaborations, in final states involving a high $E_{\mathrm{T}}^{\mathrm{miss}}$. Interpretations in terms of different DM signal models are possible, including complete models such as SUSY (predicting for example neutralinos as a possible DM candidate), Effective Field Theories (EFT) or simplified models predicting some minimal extension of the SM Lagrangian to introduce DM candidates and mediators of their interaction with the SM. Moreover, the Dark Sector is explored considering different viable "portal" interactions.
The $E_{\mathrm{T}}^{\mathrm{miss}}$ signature is not the only instrument to investigate Dark Matter at LHC: an additional source of important information lies in new resonance searches. These can be exploited to look for BSM mediators decaying into SM particles, but potentially coupling also to DM or Dark Sector particles. Together with model independent searches, specific models are studied, for example predicting the existence of new symmetry groups to provide a common explanation to DM and other open issues such as neutrino masses or the hierarchy problem.
In this chapter, a brief overview, certainly with no claims to completeness, of the LHC and ATLAS potentialities in the search for DM or Dark Sector particles is given, with a focus on the two scenarios of main interest for this thesis: simplified DM models, and Dark Photon production from (SM or BSM) Higgs boson decay.

6.1 Mono-X searches and simplified DM models

As already mentioned in Chapter 2, the WIMP model is one of the most credited DM scenarios, and largely explored. To give an (incomplete) idea of the status of the art on this field, Figure 6.1 is the most up-to-date summary plot of exclusion limits on the DM-nucleon Spin Independent scattering cross-section, as a function of DM mass, provided by Direct Detection experiments (Section 2.3.1). Spin Dependent results are also available from DD detection, and several constraints have been additionally set by Indirect Detection (ID) searches (Section 2.3.2), excluding a large area of the phase space.
A category of final states which can be interpreted in terms of DM production at LHC, goes under the collective name of mono-X channels, including processes in which the final

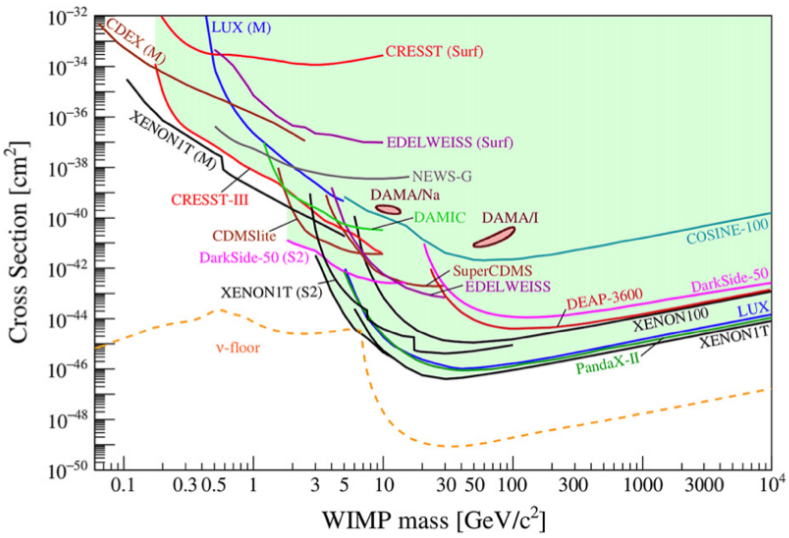


Figure 6.1: WIMP cross sections (normalized to a single nucleon) for spin-independent coupling versus mass. The DAMA/LIBRA enclosed area is the only region of interest for possible signal events, not confirmed by other experiments. The neutrino-floor [137], shown here for a Ge target, is a discovery limit arising from neutrinos from astrophysical sources which perfectly mimic a WIMP signal. It is defined as the cross-section σ_d at which a given experiment has a 90% probability to detect a WIMP with a scattering cross-section $\sigma > \sigma_d$ at $\leq 3\sigma$ [43].

state is composed of a single visible particle X (either a photon, a jet, or a Z or W boson), needed to trigger the event, and high $E_{\mathrm{T}}^{\mathrm{miss}}$. The mono-photon analysis described in the next chapter is part of this effort. Additional channels for the DM search include the di-jets, di-leptons or di-bosons final states, which can be interpreted in terms of the decay of the DM-SM mediator into two SM particles, rather than DM. Here, DM particle are not directly produced, but the observation of a new mediator is an implicit hint of new physics, potentially related to DM.

6.1.1 Simplified Dark Matter Models

These final states are typically interpreted in terms of simplified DM models. As briefly explained in Chapter 2, several theoretical models include a DM candidate as a natural consequence of the theory itself, but a **complete theory** typically involves a large number of parameters, making it challenging to test. Moreover, searching for a particular candidate of DM, defined in the context of a specific model, lacks of generality and is potentially insensitive to other possible scenarios.
These limits can be overcome by employing a more general strategy, with reduced dependence on theoretical absumptions. In this context, some benchmark DM models have been proposed by the ATLAS-CMS Dark Matter Forum [138], thanks to a collaboration among ATLAS, CMS and theoretical physicists, in order to cover a variety of possible kinematically distinct signatures of DM production.
Effective Field Theories (EFT) were extensively used for Run-1 data collected at a centre-of-mass energy $\sqrt{s} = 7 - 8$ TeV. These theories depend only on the the DM mass and on the couplings, since they approximate the DM-SM interactions to contact

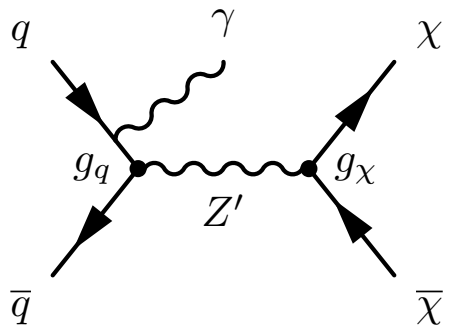


Figure 6.2: Example of a feynman diagram for an s-channel decay of a vector or axial-vector mediator into DM particles, with a photon radiated from the initial state.

interactions, by integrating out the propagator. Their limit comes from the fact that this approximation is only valid when the mass of the mediator of the interaction is lower than or comparable to the transferred momentum.
Therefore, a set of **Simplified DM Models** has been proposed to cope with the higher center-of-mass energy $\sqrt{s} = 13$ TeV of Run-2. These models consist of a minimal extension of the SM, including a new Dirac particle playing the role of the WIMP candidate, and one mediator of the interactions between SM and WIMP particles. They describe the full DM kinematics better than an EFT, without adding the complexity of a complete theory. Thanks to this, they keep a certain degree of generality and depend on a relatively small number of free parameters: the mass of DM particle (m_χ), the mass of the mediator (M_{med}), the mediator couplings to DM and SM particles (g_χ, g_{SM}) and the mediator width Γ. The proposed models have been chosen in order to explore various interactions, depending on whether the interaction proceeds through an s-channel or a t-channel, as well as on the nature of the mediator (scalar, pseudo-scalar, vector or axial-vector), and on different assumptions for its couplings to SM and DM particles. As an example, the feynman diagram for an s-channel case, which is the main focus of this thesis, is shown in Figure 6.2. More details about the specific model considered in the search described in this thesis are given in Section 8.1.1.

Grounding assumptions

Some grounding assumptions are introduced as part of the theoretical framework of Simplified DM Models adopted at LHC:

- The Dark Matter is assumed to be a Dirac fermion WIMP, stable on collider timescales and non-interacting with the detector. A different spin for the DM particle would lead to similar results. Moreover, each Simplified Model assumes the existence of a single DM particle and a single mediator of the SM-DM interaction. Of course this is not necessarily true, but this minimal hypothesis can be sufficient in the context of discovery searches, as long as it's taken into account that the observation of a signal in a specific Simplified Model can't be considered a priori as fulfilling the possible complexity of the Dark Matter particles and interactions;
- The Minimal Flavour Violation hypothesis is considered. This means that the coupling between Dark Matter and SM particles follows a similar flavour structure

as the SM interactions, so that it is possible to automatically satisfy the observed flavour simmetry, thus avoiding violation of flavour constraints;

- The mediator width value is set to the minimal width, assuming that the mediator only decays into DM and quarks, which alone can ensure the self-consistency of the theory. As long as $\Gamma < M_{med}$ and the narrow mediators are sufficiently light (Narrow Width Approximation, NWA), a generalization of this assumption is still possible by means of a rescaling procedure, the width variation in this regime would only affect cross-section with negligible impact on the kinematic distributions.

Kinematic studies and mass and couplings scan

Detailed studies, reported in [138], have been performed on the kinematics related to different choices of the parameters and of the model features. The main outcomes can be summarized as following:

- models with different mediator parity differ only by their cross sections;
- the choice of the couplings only affects the production cross-section, not the kinematics, in the range of values which ensure the validity of perturbation theory ($g < \sqrt{4\pi}$);
- for a given DM mass, the kinematics is almost independent on the mediator mass, in the on-shell region and in the NWA.

These studies converged in a set of recommendations for the parameter choices in these analyses [139], including a proposal for "standard" couplings to be explored in the provided interpretations, in order to satisfy the Narrow Width Approximation, $\Gamma_{med}/m_{med} < 10\%$, and elude stringent constraints from di-jets/leptons searches.

Overview of the results

The results of the analyses carried out at LHC, in the s-channel simplified DM models, are presented in form of model-dependent limits in the M_{med}–$m\chi$ plane, where M_{med} is the mediator mass, and m_χ the DM mass. The top plots of Figure 6.3 [140] show a comparison between different DM searches by the ATLAS Collaboration, interpreted in terms of different mediator models. Both searches in the mono-X and di-jets/leptons final states are shown, the latter giving no information about DM mass, as they aim at the observation of mediator decay into SM particles, as previously explained. In the bottom plots, the results are extrapolated to a m_χ–$\sigma_{\chi-nucleon}$ plane, where $\sigma_{\chi-nucleon}$ is the scattering cross-section between DM and SM nucleons, in order to compare the LHC sensitivity to Spin Dependent (left plot) and Spin Independent (right plot) DD results, highlighting a good complementarity between the two detection strategies. In particular, the mono-photon analysis described in Chapter 8 is part of the mono-X searches, which are shown to be competitive, with respect to DD experiments, for low DM masses.

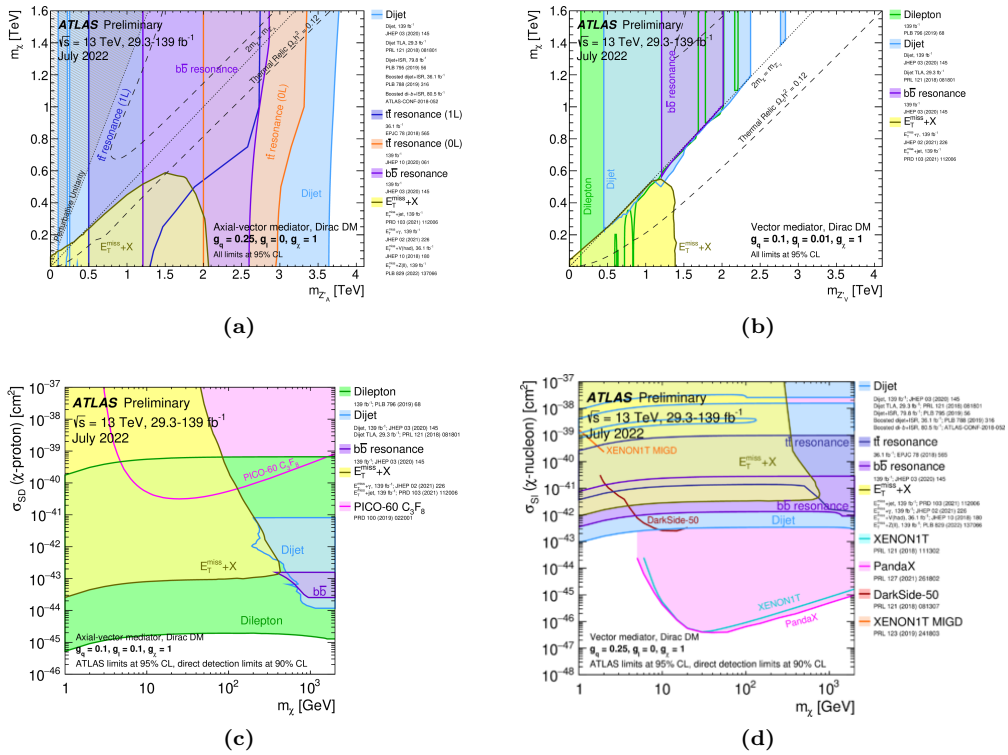


Figure 6.3: In (a) and (b): comparison of the limits in $M_{med}-m_{\chi}$ plane set by analyses carried out in different channels on data collected in 2015-2016 by the ATLAS detector, considering two different signal models. In (c) and (d): comparison between the results obtained at LHC, with limits from DD experiment, considering Spin Dependent (c) and Independent (d) DD experiment and different signal models for the LHC result interpretations [140].

6.2 Exploring the dark sector through dark photon searches

The Dark Sector scenario, described in Section 2.2, gives origin to an extremely rich phenomenology potentially observable at the LHC [141, 142, 143, 144].
In general, a strong effort has been done particularly in the search for massive Dark Photons, with several constraints set by independent experiments on the kinetic-mixing parameter ϵ. Sub-MeV Dark Photons have also been explored, as well as *milli-charge* scenarios (Section 2.2.2) with massless Dark Photon. An incomplete set of results is summarized in Figure 6.4, to provide a minimum context, while a complete overview of the status of the art in the Dark Photon search is provided in reference [66].
At colliders, searches for possible new fields, either dark fermions or dark bosons, can be conducted exploring the different possible "portal" interactions with the SM. Typical signatures which can be interpreted in terms of new fields in the Dark Sector are new resonances arising from the decay of a dark particle into SM particles, final states involving $E_{\mathrm{T}}^{\mathrm{miss}}$ from invisible dark particles or displaced vertex from the decay of long lived dark particles.
In particular, this work focuses on the search for the decay of a Higgs boson - either SM or BSM - into a photon and a Dark Photon ($H \to \gamma\gamma_d$).

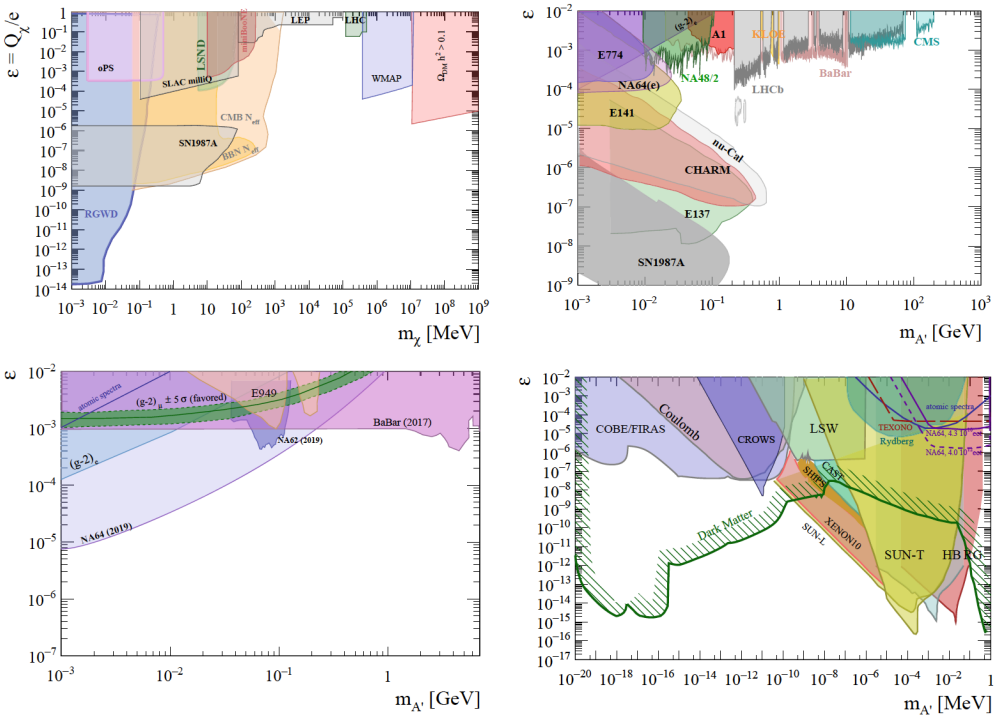


Figure 6.4: Overview of some of the main results from Dark Photon searches, provided by experiments on milli-charged Dark Sector matter (top left), massive Dark Photon with $m_{A'} > 1$ MeV, including results from di-leptons searches at colliders (top right), massive Dark Photon decaying into invisible final states (bottom left) and massive, but "light" Dark Photon with $m_{A'} < 1$ MeV (bottom right) [66]. The exclusion limits are set on the kinetic-mixing parameter ϵ.

6.2.1 Dark photon from Higgs boson decay

The BSM decay of the Higgs boson into a photon and a Dark Photon is particularly interesting in the massless γ_d case, where kinetic mixing between the γ_d and the electromagnetic or hypercharge currents can be canceled out thus leading to its complete decoupling, at tree level, from the SM [66]. This allows to avoid the tight existing constraints on the kinetic mixing coupling from other experiments, opening to a wide, not yet excluded phase-space. In this case, as already mentioned in Section 2.2.2, the interaction between the γ_d and the SM sector might happen through higher dimensional interactions with effective scale $\lambda_{\gamma\gamma_d}$, featuring also an additional BSM field which is coupled both with the SM and the Dark Sector. The effective scale is expected to be proportional to the mass of this messenger sector, resulting in a suppression of the decay width by a factor $1/\Lambda$ for large messenger masses.
In the case of interest in this thesis the γ_d is produced through the decay of an Higgs boson into a photon and a γ_d, whith a "dark" messenger field coupled both to the Higgs and the γ_d [145, 143]. An interesting feature of this kind of process is that, similarly to the SM $H \to \gamma\gamma$ decay, the non-decoupling nature of the Higgs boson removes the loop suppression mentioned above.

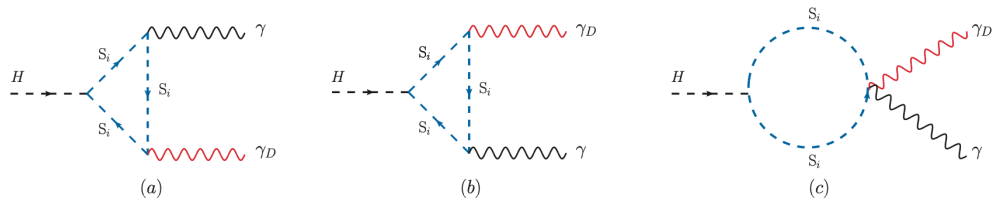


Figure 6.5: Feynman diagrams through which the Higgs boson decay into a photon and a dark photon can be realized, under the assumption of the existence of a new field S_i coupled both to the Higgs (Higgs portal) and the dark photon (vector portal).

A benchmark model

Different models can include this decay. As a benchmark model, we consider here a scenario which connects the DM and the mass hierarchy problems by assuming the Yukawa couplings to be radiatively generated from the non-perturbative breaking of a new chiral symmetry in the dark-sector [146, 145, 143].
In addition to the γ_d, the model requires a new messenger sector. This sector is composed of 3 generations of coloured and electroweak scalar messengers: the coloured messengers include the left-doublets $(\hat{S}_L^{Ui}\ \hat{S}_L^{Di})$ and two right-singlets S_R^{Ui}, S_R^{Di} of the $SU(2)_L$ group, while the electroweak messengers consist in the corresponding $(\hat{S}_L^{Ei}\ \hat{S}_L^{Ni})$ and S_R^{Ei}, S_R^{Ni}, with an universal mass term for each generation. These messengers are assumed to have universal Yukawa couplings to dark fermions and SM quarks/leptons, and to be also charged under the $U(1)_D$ group.
Focusing on the coloured sector, the interaction Lagrangian assumes the form:

$$\mathcal{L} = \lambda_S S_0 \big(\tilde{H}^\dagger S_L^U S_R^U + H^\dagger S_L^D S_R^D\big) + h.c. \tag{6.1}$$

where S_0 is an additional singlet scalars which can get a vev $\langle S\rangle$, giving rise to a trilinear Higgs coupling which can induce $H \to \gamma\gamma_d$ or $H \to \gamma\gamma$ decays in 1-loop diagrams as presented in Figure 6.5, with decay rate proportional to $\mu_S = \lambda_S\langle S\rangle$. After Electroweak symmetry breaking, a mixing mass term in the left and right messenger sectors arises:

$$\mathcal{L}_0 = \partial_\mu \hat{S}^\dagger \partial \hat{S} - \hat{S}^\dagger M_S^2 \hat{S} \tag{6.2}$$

where $\hat{S} = (S_L, S_R)$ and the mass term is expressed as:

$$M_S = \begin{pmatrix} m_L^2 & \mu_S v \\ \mu_S v & m_R^2 \end{pmatrix} \tag{6.3}$$

where v represents the Higgs field *vev*.
The interaction between the γ_d and the messenger sector can be introduced by replacing the partial derivative in the kinetic Lagrangian term with the covariant derivative $D^\mu = \partial^\mu + ie_D q A_D^\mu$ where A_D^μ is the Dark Photon field, e_D the unit charge of $U(1)_D$ and q the charge eigenvalue of the field to which the covariant derivative applies.

The decay width is derived to be:

$$\Gamma(H \to \gamma\gamma_D) = \frac{m_H^3}{32\pi\Lambda^2_{\gamma\gamma_D}} \tag{6.4}$$

If the Higgs boson mass can be neglected with respect to the messenger mass, and in a minimal scenario assuming mass universality in the messenger sectors, with $m_L \sim m_R$, the effective scale results to be equal to $\Lambda_{\gamma\gamma_D} = \frac{6\pi v}{R\sqrt{\alpha\alpha_D}} \frac{1-\xi^2}{\xi^2}$, where $\xi = (\mu_S v)/\bar{m}^2$ is the mixing parameter and $R = N_C \sum_{i=1}^{3} (\frac{2}{3} q_{U_i} - \frac{1}{3} q_{D_i})$, with q_{U_i/D_i} the $U(1)_D$ charges in the up and down sectors. This value of $\Lambda_{\gamma\gamma_D}$ is indepedent on the messenger mass and finite for $\xi < 1$, thus the non-decoupling nature of the Higgs boson is realized. It has been observed that the validity of this description of the BSM Higgs decay can be actually extended to "light" Dark Photon masses: despite the switch on of the kinetic mixing, given the present constraints, its contribution to the global decay width is expected to be overall negligible with respect to the loop diagrams.

Present results of Dark Photon searches from Higgs boson decay at LHC

In Chapter 9, the search for γ_d from the Higgs decay in the ZH production mode will be detailed. A similar analysis, in the VBF production mode (Figure 6.6), has been performed in the ATLAS Collaboration [147], setting exclusion limits on the brancing fraction $BR(H \to \gamma\gamma_d)$ as a function of the Higgs mass, considering also BSM Higgs bosons. The results are shown in Figure 6.7, together with similar constraints from the CMS searches in the ZH and VBF [148, 149] production mode. Concerning the ggF production mode, neither the ATLAS Collaboration nor the CMS one have explored it yet, for a SM Higgs boson, due to sensitivity limitations arising from high trigger $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds used during Run-2. Nevertheless, in Chapter 10, the ggF production mode will be explored, for heavy BSM Higgs bosons with masses above 400 GeV, and a brief discussion of prospects for future studies will be presented.

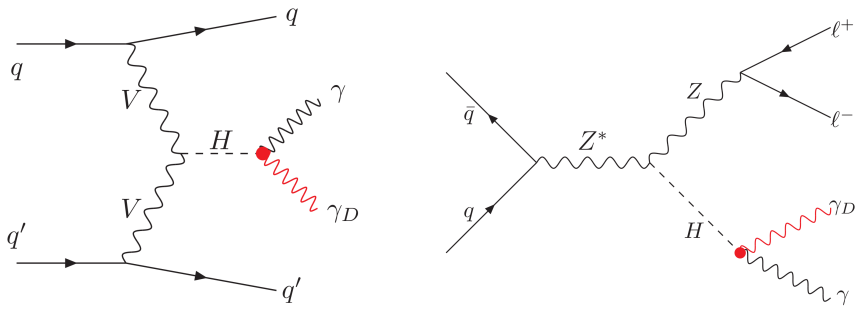


Figure 6.6: Feynman diagrams for the Higgs production via VBF (left), and ZH (right) channels at the LHC.

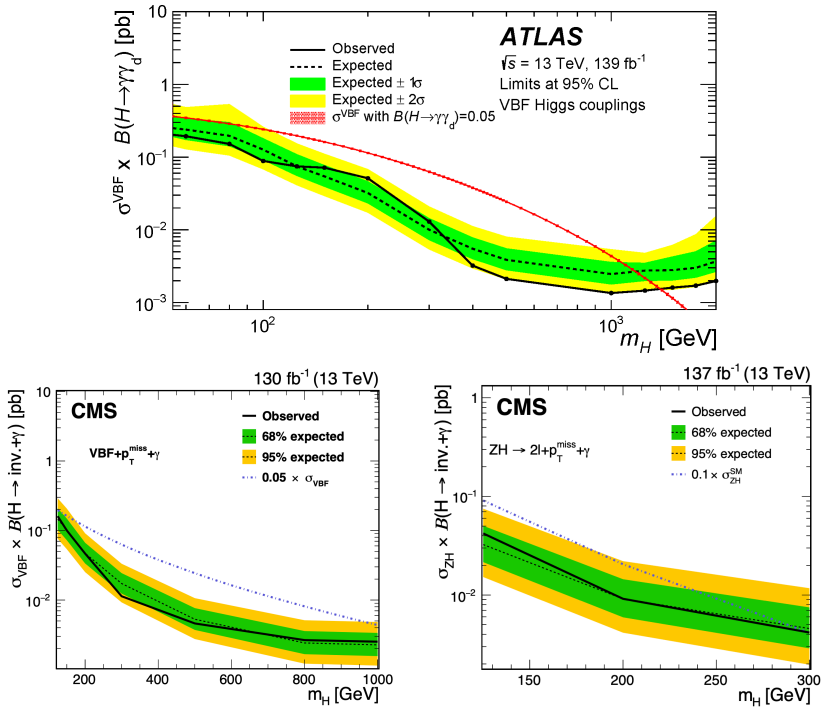


Figure 6.7: Exclusion limit at 95% CL on $\sigma \times BR(H \to \gamma\gamma_d)$ as a function of Higgs mass, from ATLAS search in the VBF production mode [147] (top plot) and from CMS searches in the VBF [149] (bottom left) and ZH [148] (bottom right).

Chapter 7

General tools and methods in searches for new physics with the ATLAS detector

Within the ATLAS Collaboration, a large set of methods, frameworks and tools are commonly developed to harmonize as much as possible the results from the plethora of analyses that are performed by the different groups. In this chapter, some general information is given about how a typical BSM search is performed. A brief description of the data pre-selections, the MC samples production and the reconstruction chain is given, followed by a description of common methods for the evaluation of systematic uncertainties and statistical analysis.

7.1 Data and MC simulations

7.1.1 Data format

The searches described in this thesis are based on the full Run 2 dataset of proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the ATLAS experiment between 2015 and 2018. The dataset corresponds to a total integratel luminosity of 139 fb^{-1} with an uncertainty of $\pm 1.7\%$ [93, 150], with peak instantaneous luminosity up to $L = 2.1 \times 10^{34}$ cm^{-2}s^{-1} and an average number of interactions in the same or neighbouring bunch crossings of $\langle \mu \rangle = 33.7$. The integrated luminosity includes only data fulfilling quality requirements on the stability of the beams, the operational status of all ATLAS detector components, and the quality of the recorded data (i.e. data entering the so-called Good Run List, GRL). In addition, trigger selection requirements are applied to reduce the event rate from the about 40 MHz delivered by LHC down to the kHz scale which can be handled and stored.
The signals released in each subdetector by the particles produced in the collision events passing the OR of all employed triggers, are stored as Raw Data in RDO (Raw Data Object) files. Afterwards, at analysis level, a set of "cleaning" pre-selections is applied:

- Data quality: the event must be in the Good Run List (GRL);
- Good vertex: a primary vertex must be reconstructed with at least two associated good-quality tracks see [151] with $p_{\mathrm{T}} > 400$ MeV and $|\eta| < 2.5$;
- Jet cleaning: it is introduced in order to discriminate between jets originating from hard scatter processes and jets from non-collision background. Events with any *LooseBad* jet (Section 4.4.6), overlapping with neither leptons nor photons, with calibrated $p_{\mathrm{T}} > 20$ GeV are rejected.

The last two pre-selections are applied also on MC simulations.

7.1.2 MC Production chain

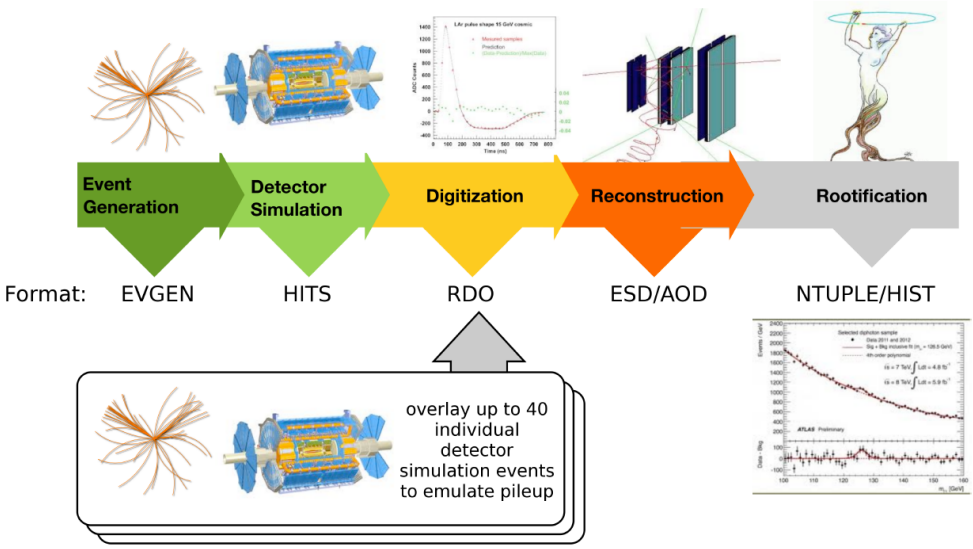


Figure 7.1: Sketch of the ATLAS MC production chain [152].

The signals and the physics processes which constitute a possible analysis background are modelled using Monte Carlo (MC) simulations. The MC samples production is performed centrally in the ATLAS Collaboration within the ATLAS Simulation Infrastructure, based on the Athena Framework [153, 154], with a standard production chain represented in Figure 7.1 and including the following steps:

Event generation At this stage, the truth-level simulation of the parton-parton scattering and of the subsequent parton-shower, hadronization and decay into the final state stable particles are performed, and the results are stored in the EVGEN format. This step is strongly dependent on theoretical QCD computations, and all the degrees of "arbitrariety" in the theoretical computation needs to be taken into account in form of systematic uncertainties at analysis-level. Matrix Element (ME) calculations are used for the parton-parton hard-scattering simulations, with different possible correction orders, ranging from the Leading Order (LO) up to the Next-to-Next-to Leading Order (NNLO). The results depend on the choice of the PDF set, as well as QCD factorization (μ_F) and renormalization (μ_R) scales. The Parton Shower (PS) due to parton QCD radiation is simulated as a cascade process fed by a single parton split into two partons, with flavour, 4-momentum and unitarity conservation [117, 155]. Different PS algorithms, based on collinear and soft approximation, are emploied by different MC generators, but full ME simulations are also needed for parton emission which do not satisfy neither of the two approximations. A matching between PS and ME contributions is therefore needed to avoid double-counting, with a *matching scale* defining the threshold above which PS algorithms are not appliable anymore. The ME calculation can be performed at tree-level only (ME+PS matching) or at NLO (NLO+PS matching) [156]. The final step of the *hadronization* is based on phenomenological models, either the Lund string [157] or the

cluster fragmentation [158] one. These models are based on experimentally determined parameters, defining the MC PS Tuning;

Detector simulation The particles interaction with the detector is simulated either in full or in fast simulation, the first being performed through `Geant4` generator [159], the latter based on a simplified geometry description and parametrization of the physics of interaction between the particles and the detector material to reduce the required CPU time. The Fast simulation can be realized with multiple possible strategies: the Fast G4 Simulation introduce only a simplification of low energy electromagnetic particles simulations by using pre-simulated "frozen" showers; The ATLAS FAST 2 simulation [160] (AF2) and the most recent ATLAS FAST 3 [161] (AF3) are based on parametrizing the longitudinal and lateral energy profiles of the showers in the calorimeters using the Fast Calorimeter Simulation (FastCaloSim); the ATLAS Fast 2 simulation Fatras [162] (AF2F), finally, exploits also the Fast ATLAS Tracking Simulation (Fatras), providing a simplified description of the ID and MS geometry. The simulated energy deposits in the detector are computed and stored in HITS files;

Digitization The simulated signals are converted into detector responses which mimic real data, in Raw Data Objects (RDO) format. In this step, low momentum pile-up events are also simulated with Pythia 8 [163], and injected into the production chain. Three separate MC production campaigns (MC16a, MC16d, MC16e) allow to account for the specific pile-up conditions of each period of data taking, respectively 2015-2016, 2017 and 2018.

7.1.3 Reconstruction

After MC digitization, data and MC simulations undergo the same event reconstruction chain, as detailed in Chapter 4. The event reconstruction is performed using the Athena framework [150], based on the Gaudi framework used within the ATLAS, LHCb and FCC collaborations. The information associated to each reconstructed event is stored in Event Summary Data (ESD) and Analysis Object Data (AOD) formats. The final step provides an intermediate data format (the derivation, DAOD) optimized for specific categories of analyses. An appropriate trigger and skimming selection, as well as the reduction of the stored variables and information allows a reduction of the size from the ∼PB scale of the DAOD files to the ∼TB of the DAOD files.

7.1.4 Analysis framework

The derivations are processed through analysis specific frameworks, based on the Athena projects, which extract and process the needed information and store them into ROOT trees [164]. In this step, some skimming selections can be applied to preselect the events and reduce the global size of the final trees. In addition, the p_{T} and η thresholds for the different reconstructed physics objects, as well as the WP for jet, $E_{\mathrm{T}}^{\mathrm{miss}}$ and leptons/photons isolation and identification, are applied. Two categories of reconstructed physics objects are defined, with looser ("baseline") and tighter ("selected") criteria, the latter being a subset of the first. An Overlap Removal (OR) procedure is performed, to avoid ambiguities in the identification of the particles in the final state: if two reconstructed baseline physics objects are found overlapping, only one of them is retained according to the priority decided at analysis level, following general recommendations [165] if no specific analysis need must be satisfied. In particular, in the analyses described in this

thesis, the Overlap Removal procedure described in Table 7.1 is adopted. Baseline physics objects are also given as input to $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, which is performed at analysis level to ensure consistency with the physics objects selections. A set of event-based corrections are needed to correct MC simulations: the scale factors described in Chapter 4, accounting for residual discrepancies between MC simulations and data for different physics objects properties, are stored in forms of event-by-event weights providing the needed correction to each event, together with the pile-up reweighting corrections provided for the three MC campaignes, needed to reproduce the luminosity profile ($\langle\mu\rangle$) observed in data (Figure 3.7), and the MC weights arising from NLO generations.
The frameworks used in the analyses described in this work, are based on a package initially developed for SUSY searches, the SUSYTools [166]. This allows to "automatically" integrate in the framework the several recommendations provided by the different Combined Performance (CP) groups.

Remove	**Keep**	**Criteria**	**Reason**
e	μ	Shared inner detector track	Remove electron candidates coming from muon bremstrahlung followed by photon conversion
γ	μ	$\Delta R < 0.4$	Photons from muon Bremstrahlung
γ	e	$\Delta R < 0.4$	Reduce $e \to \gamma$
jet	μ	$\Delta R < 0.4$ and number of tracks with $p_{\mathrm{T}} > 0.5$ GeV associated with the jet <3	Prompt muon emitting a photon
μ	jet	$\Delta R < 0.4$ and number of tracks with $p_{\mathrm{T}} > 0.5$ GeV associated with the jet ≥ 3	Muon from hadron decay
jet	e	$\Delta R < 0.2$	Remove duplication of electron as jet
e	jet	$0.2 < \Delta R < 0.4$	Region enriched with real hadronic jets close to electrons
jet	γ	$\Delta R < 0.4$	Remove duplication of photon as jet

Table 7.1: Summary of the applied overlap removal procedure in order of priority

7.2 Systematic uncertainties

Experimental and theoretical systematic uncertainties need to be estimated for all MC simulated processes. While some of these uncertainties are specific of each analysis, most contributions, mainly related to the event reconstruction and theoretical assumptions adopted in the events simulation, are common and their evaluation is harmonized among the different analyses.

Experimental systematics A set of experimental systematic variations associated to the event and physics objects reconstruction is centrally computed by the relevant CP groups as detailed in Chapter 4. Two categories of uncertainty can be identified: the first one is related to the efficiency of reconstruction, identification, isolation and trigger for

each physics object, and it doesn't directly affect the kinematics of each MC event, but only the Scale Factors (SFs) applied as event weights. This results in an event-by-event reweighting which can lead to a variation of the global kinematic distributions. For each event and physics object, the up and down variations ($\pm 1\sigma$) of the SFs associated to each systematic source are provided by the CP groups and stored in the analysis trees. The second category collects all the systematic effects related to energy measurement and calibration, which affect the value of the relevant kinematic variables. In this case, a complete event reconstruction for each systematic variation is required, in order to evaluate their impact. For a given source of uncertainty, the corresponding variation on the final yield for each background process is obtained by varying the relevant quantity and by propagating its impact through the full analysis chain.
Table 7.2 summarizes the common experimental uncertainty sources.

Theoretical systematics Theoretical uncertainties account for the uncertainty in the description of the PDF and the simulation of the Parton Shower (PS), by evaluating how different choices of the PDF set and of the parameters used in the parton-level and PS simulation at generation level (see the "Event generation" paragraph in Section 7.1.2) affect the background and signal expectations.
Two categories can be identified. The first one, computed following the PDF4LHC recipes [167], includes three main systematic sources:

- Uncertainties from missing higher orders are evaluated [167] as variations of the QCD factorisation and renormalisation scales in the matrix elements. The scales are varied by factors of 0.5 and 2 avoiding variations in opposite directions, and the scale uncertainty, $\sigma(scales)$, is then given by the maximum shift of the envelope with respect to the nominal.
- The NNPDF set provides an ensemble of 101 PDF replicas. The nominal value corresponds to the mean of all the replicas and the standard deviation is the uncertainty. The results are also cross-checked using the central values of the `CT14nnlo` [168] and `MMHT2014 NLO` [88] PDF sets. The envelope is used as the combined uncertainty, $\sigma(PDF)$.
- The uncertainty on the strong coupling constant α_S is assessed by variations of ± 0.001. The combined uncertainty is calculated as $\sigma(\alpha_s) = \frac{\sigma_{down} - \sigma_{up}}{2}$.

The second category includes the uncertainties associated to the FSR and ISR, as well as the Multi Parton Interaction (MPI) and underlying events (UE) effects which are parametrized in the MC PS tuning. These uncertainties are typically estimated by comparing truth-level MC simulations produced with different parameters.

7.3 Analysis strategy and statistical methods

In this section, the main concepts of the fit strategies for background estimation and results interpretation is described. Several tools and frameworks are available to implement the statistical data analysis detailed in the following: the analyses reported in this work use the HistFitter package [169], which is based on HistFactory [170] to build the PDFs through a combination of ROOT histograms and store them in a `RooWorkspace`, and on RooFit [171] and RooStats for the fit to data and statistical tests.

Event related		
	LUMI_syst	computation of the integrated luminosity
	PRW_DATASF	variation of data scale factor for pile-up reweighting
Efficiencies		
Photons	PH_EFF	identification of leading photon
	PH_EFF_(TRK_)ISO	(track) calorimetric isolation efficiency
	PH_EFF_TRIG_EFF	trigger efficiency on leading photon
Electrons	EL_EFF_RECO	reconstruction efficiency of electrons
	EL_EFF_ID	identification efficiency of electrons
	EL_EFF_ISO	isolation efficiency of electrons
Muons	MU_EFF_SYST(_LOWPT)	reconstruction efficiency of muons
	MU_EFF_STAT	
	MU_ISO_SYST(/STAT)	Systematic(/statistical) uncertainty on isolation efficiency
	MU_EFF_TTVA_SYST(/STAT)	Systematic(/statistical) uncertainty on TTVA SF
Jets	JVT_EFF	Jet Vertex Tagger efficiency
Photons and leptons energy		
Photon and electrons	EG_SCALE	energy scale uncertainty from calibration
	EG_RESO	energy resolution uncertainty from caliobration
Muons	MUON_SCALE	energy scale uncertainty
	MUON_ID	energy resolution uncertainty from inner detector
	MUON_MS	energy resolution uncertainty from muon system
	MUON_SAGITTA_RESBIAS(/RHO)	Muon sagitta-related uncertainties
Jets related		
	JET_EffectiveNP	energy scale uncertainty split into 15 components
	JET_EtaIntercalibration_Modelling	inter-calibration uncertainty
	JET_EtaIntercalibration_NonClosure	
	JET_EtaIntercalibration_NonClosure	inter-calibration uncertainty with 4 components:
	JET_EtaIntercalibration_NonClosure	2018data, highE, negEta, posEta
	JET_EtaIntercalibration_NonClosure	
	JET_EtaIntercalibration_TotalStat	inter-calibration uncertainty
	JET_Flavor_Composition(/Response)	Uncertainties related to jet flavour reconstruction
	JET_JER_DataVsMC_MC16	Energy resolution uncertainty from data/MC comparison
	JET_JER_EffectiveNP	Energy resolution uncertainty split into 11 parameters
	JET_Pileup_OffsetMu(/NPV)	
	JET_Pileup_PtTerm	Pileup related uncertainties
	JET_Pileup_RhoTopology	
	JET_PunchThrough_MC16	jet-related uncertainty
	JET_SingleParticle_HighPt	jet-related uncertainty
Favour tagging		
	FT_EFF_B_systematics	b-jet tagging efficiency
	FT_EFF_C_systematics	c-jet tagging efficiency
	FT_EFF_Light_systematics	light-jet tagging efficiency
	FT_EFF_extrapolation	b-jet tagging efficiency with high pt extrapolation
	FT_EFF_extrapolation_from_charm	c-jet tagging efficiency with high pt extrapolation
$E_{\mathrm{T}}^{\mathrm{miss}}$ related		
	MET_SoftTrk_ResoPerp(/Para)	Transverse (/longitudinal) resolution uncertainty of TST
	MET_SoftTrk_ScaleUp(/Down)	Longitudinal scale uncertainties of TST

Table 7.2: List and description of the systematic contributions of uncertainty.

7.3.1 Analysis regions

The first step of an analysis is the definition of selection criteria identifying a Signal Region (SR), where the background contribution are suppressed while maximizing the signal acceptance. In order to improve the background estimation, specific Control Regions (CRs) are defined by inverting or loosening some of the SR selection criteria, in order to enrich each CR with a given background, while ensuring a negligible signal contamination. Finally, Validation Regions (VRs) can be used to test the statistical method: they should lay in-between the SR and the CRs, being characterized by a background composition as close as possible to the SR, but with negligible signal contamination.

The CRs are used to normalize the background MC simulations to data by means of a fitting procedure based on statistically independent CRs, which ensures that they can be modeled by separate Probability Distribution Functions (PDFs) and combined into a simultaneous fit. A simplified description, in a 2D plane, of the logic and usage of CRs, VRs and SRs is shown in Figure 7.2, where the arrows represent the extrapolation of the normalization factors from the CRs.

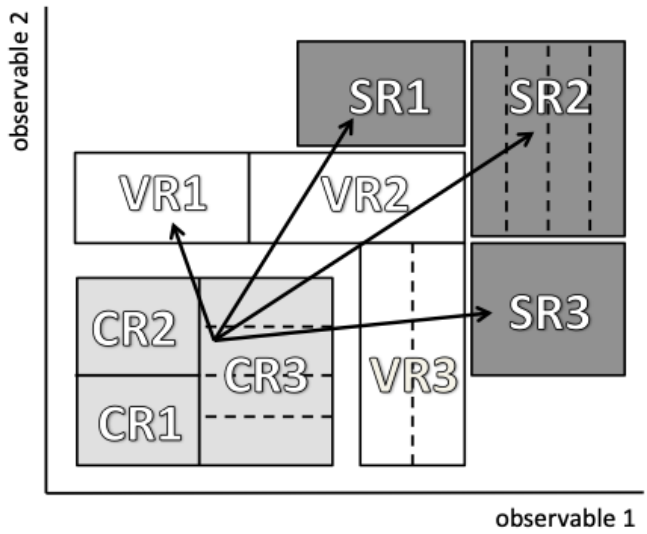


Figure 7.2: Sketch of an analysis strategy based on multiple CRs, VRs and SRs [169].

7.3.2 Likelihood function

The fitting strategy is based on the definition of a likelihood function describing the background and signal models. The parton distribution function (PDF) is built as a product of Poisson distributions of event counts in the different bins of the SR(s) or VR(s) and CR(s) $P(n_b; \boldsymbol{\theta}, \mathbf{k}, \mu_{sig})$, with n_b the number of events in the bin b, $\mathbf{k}$ and μ_{sig} the background normalization factors and the signal strength respectively, both included as free parameters of the fit, and $\boldsymbol{\theta}$ the systematic uncertainties treated as nuisance parameters. The NPs associated to the experimental and theoretical systematic uncertainties (typically identified with α) are included with a gaussian constraint $G(\hat{\alpha}|\alpha, \hat{\Delta}_\alpha)$, where $\hat{\alpha}$ is the central value of the auxiliary measurement and the Gaussian width, $\hat{\Delta}_\alpha$, is the value of the considered systematic uncertainty. Another systematic contribution is related to the statistical uncertainty of MC simulations (the γ NPs): these NPs are treated as uncorrelated among different regions, but correlated among the background processes (in order to provide a global uncertainty for each region) and modelled with a poissonian constraint. This is a simplification of the Barlow and Beeston method (in which to each sample is associated a different NP). The parameter of the Poisson distribution is taken as $\lambda_b = n_b/\delta_b$ where n_b is the MC yield in the considered bin, and δ_b the MC statistical uncertainty. The constraint term in the likelihood can be expressed as $P(\hat{\lambda}_b|\gamma_b\tau_b)$, where λ_b is treated as an auxiliary variable, γ_b is the nuisance parameter and $\tau_b = (n_b/\delta_b)^2$

is a fixed value such that the γ_b parameter is centered in 1. The likelihood function is therefore defined as

$$\mathcal{L}(\mathbf{n}, \hat{\boldsymbol{\theta}}|\boldsymbol{\theta}, \mathbf{k}, \mu_{sig}) = \prod_{b \in CRs+SR\ bins} P(n_b^{obs}|n_b^{exp}(\boldsymbol{\theta}, \mathbf{k}, \mu_{sig})) \times C_b(\hat{\boldsymbol{\theta}}) \tag{7.1}$$

where $C_b(\hat{\boldsymbol{\theta}})$ is the NPs constraint for the bin b

$$C_b(\hat{\boldsymbol{\theta}}) = \left[\prod_s G(\hat{\alpha}_s|\alpha_s, \hat{\Delta}_{s,b})\right] \times P(\hat{\lambda}_b|\gamma_b\tau_b) \tag{7.2}$$

and $n_b^{exp}(\boldsymbol{\theta}, \mathbf{k}, \mu_{sig})$ are the expected yields as a function of the nuisance and free parameters:

$$n_b^{exp}(\boldsymbol{\theta}, \mathbf{k}, \mu_{sig}) = n_{b,sig}^{exp}(\boldsymbol{\theta}) \times \mu_{sig} + \sum_{B \in bkgs} n_{b,B}^{exp}(\boldsymbol{\theta}) \times k_B \tag{7.3}$$

In HistFitter, the dependence of expected yields on the NP is parametrized such that the $\hat{\alpha}$ auxiliary measurements are set to 0 with gaussian width equal to 1 ($G(\hat{0}|\alpha, \hat{1})$, therefore:

$$n_b^{exp}(\boldsymbol{\theta}) = \gamma_b \times n_b^{MC} \times \left(1 + \sum_s (\alpha_{s,b}\Delta_{s,b})\right) \tag{7.4}$$

7.3.3 Fit strategies

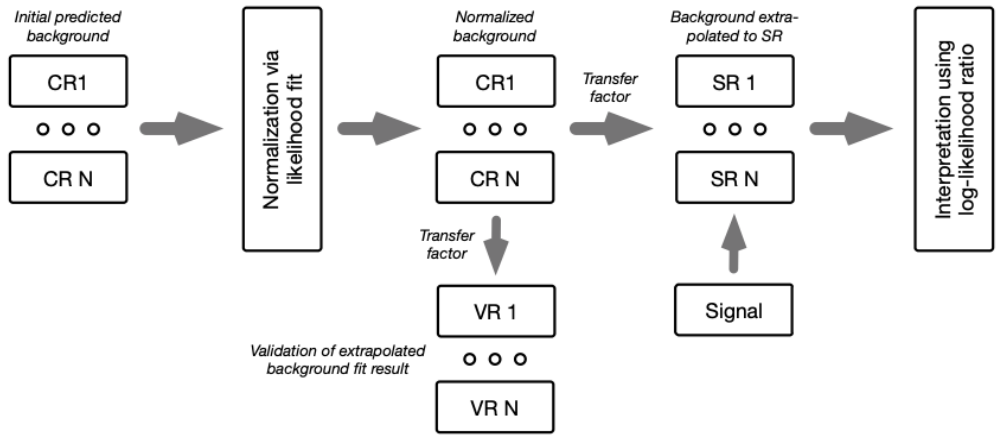


Figure 7.3: A sketch of the typical background and signal fit workflow in new physics searches in ATLAS, using the HistFitter package [169].

The discovery analyses are mainly based on two fit strategies, as summarized in Figure 7.3: the *background-only* fit, performed in the CRs to correct the background estimation in the SR and look for excesses of data with respect to SM expectations, and the *exclusion* fit in SRs+CRs, used to test the signal hypothesis and provide upper limits on the cross-section of BSM processes, if no excess is highlighted by the background-only fit. Both exploit a simultaneous fitting technique, which allows a coherent normalization of the different background sources to data in multiple regions with "shared" parameters, and ensures a coherent treatment of the correlation of the systematic uncertainties across

the different regions. This leads to their partial cancelation in the fit. The behaviour of the NPs in the fit can be cross-checked by looking at the plots of correlations among different NPs, and the pull-plots. The latter show the NP central value after the fit, as well as its uncertainty, pointing out possible unexpected "pulls" or under/over constraints, the former consisting in a NP central value after the fit "pulled" away from the central value, the latter in uncertainties much larger(smaller) than the initial one. While some pulls and constraint might be expected, anomalous behaviours can sometimes be an hint of an instability of the fit, resulting in some NPs having a larger impact than expected, for instance due to their 'attempt" to artificially cover possible discrepancies between data and background expectations. Typical cases when this might happen can be high statistical fluctuations affecting the bin-by-bin systematic uncertainty, or underestimated systematic sources, leading to the other uncertainty contributions taking over.

Background-only fit

The background-only fit typically consists in a maximum likelihood fit to data in the CRs only, to validate the background estimation avoiding possible bias due to the hypothetical presence of a signal. The signal contribution is assumed to be null, and the free parameters are the background normalization factors, shared among the different CRs and the SRs (or VRs) as shown in Eq. 7.1, in order to extrapolate the results of the fit in the CRs to the background predictions in the SRs (VRs). The simplest fit configuration is a single-bin fit, including a single SR and its associated CRs, to provide a global normalization factor on the background yields (pure counting experiment). More refined results can be obtained by means of multiple SRs defined in bins of a given discriminating variable. In this case, a "shape" fit can be performed, in which the background expectations are normalized to data bin-by-bin, thus exploiting also the shape information.

Exclusion fit

The exclusion fit is based on the profiled log-likelihood ratio test statistics, including the SR in the fit and potentially accounting for signal contamination also in the CRs. In addition to the background normalization factors, the signal strength ($\mu_{sig} = \sigma_{obs}/\sigma_{theo}$) is introduced as the POI. The fit can be performed either with a model-independent ("discovery" fit in HistFitter) or a model-dependent approach ("exclusion" fit in HistFitter), depending on whether or not a specific signal model is considered. The first case can be implemented only in form of a single-bin fit, due to the absence of any assumption on the signal kinematic, while the latter allows the implementation of the "shape" fit strategy. In order to derive the upper limits on the signal strength μ_{sig}, a scan is performed testing different μ_{sig} and the upper limit is taken as the highest value such that the p-value is $p_\mu > 1 - CL$. In high energy physics, a p-value of 0.05 is commonly used to set exclusion limits. Expected and observed limits are provided: the former are obtained by fitting on a bagkround-only asimov dataset, built from the best fit in SRs+CRs in the background-only hypothesis, while the latter is the result obtained from the fit to real data.

Log-likelihood ratio As previously mentioned, the fit strategy emploied for the exclusion limit is based on the profiled log-likelihood ratio test statistics [172]. Considering μ

the POI and $\boldsymbol{\theta}$ the nuisance parameters, the likelihood-ratio test is defined as

$$\lambda(\mu) = \frac{\mathcal{L}(\hat{\hat{\boldsymbol{\theta}}}, \mu)}{\mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mu})} \tag{7.5}$$

The denominator is the "global" maximum likelihood, obtained with both $\boldsymbol{\theta}$ and μ included as free parameters and defines the background+signal model that best fit observed data. The numerator is the conditional maximum likelihood obtained for a fixed value of the μ parameter. The more data are consistent with background predictions, the more the $\hat{\mu}$ value will be close to 0 and the best fit model close to the background-only scenario. For computational simplicity, the fit is peformed by minimizing the negative log-likelihood ratio rather then maximizing $\lambda(\mu)$. The used test statistics is therefore the one-sided profile log-likelihood ratio, defined to avoid non physical negative values of the fitted μ_{sig} and ignore negative tails :

$$q_\mu = \begin{cases} -2 \ln \left(\frac{\mathcal{L}(\hat{\hat{\boldsymbol{\theta}}}, \mu)}{\mathcal{L}(\hat{\boldsymbol{\theta}}(0), 0)} \right) & \text{if } 0 \geq \hat{\mu}; \\ -2 \ln \left(\frac{\mathcal{L}(\hat{\hat{\boldsymbol{\theta}}}, \mu)}{\mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mu})} \right) & \text{if } \mu \geq \hat{\mu} \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Frequentist hypothesis testing and asymptotic approach In order to test how well a given hypothesis describes data, the p-value is evaluated, i.e. the probability, for the given hypothesis, of obtaining a new measurement with equal or lower compatibility to the tested hypothesis than observed. Defining $f(q_\mu|\mu)$ the probability distribution of the q_μ value, and taking into account that we are performing a one-side test, the p-value is defined as:

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|\mu) \, dq_\mu \tag{7.6}$$

The probability distribution function is not known a priori and should be in principle derived numerically through simulated "toys" experiments. Nevertheless, according to the Wilk's theorem [173], in the asymptotic approximation, valid for a large enough dataset, the $f(q_\mu)$ follow a χ^2 distribution with one degree of freedom and q_μ follows the Wald approximation [174]:

$$q_\mu \sim \begin{cases} \left(\frac{\mu - \hat{\mu}}{\sigma} \right)^2 & \text{if } \mu \geq \hat{\mu} \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

The p-value is given by:

$$p_\mu = 1 - \Phi(\sqrt{q_\mu}) \tag{7.7}$$

with Φ the cumulative function of the standard Gaussian distribution, corresponding to a significance $Z_\mu = \Phi^{-1}(1 - p_\mu) = \sqrt{q_\mu}$

The CLs method In ATLAS, a modified frequentist test is usually emploied in new physics searches, aiming at avoiding false positives for the "new physics" hypothesis. This method is based on the definition of CL_s value [175]:

$$CL_s = \frac{p_1}{1 - p_0} \tag{7.8}$$

where p_1 is the p-value of the signal hypothesis ($\mu_{sig} = 1$) and p_0 the one associated to the background only hypothesis. In conclusion, signal strength corresponging to $CL_s \leq 0.05$ are excluded at 95% CL.

Chapter 8

Search for Dark Matter in mono-photon signature

The mono-photon analysis [176] searches for an excess of events, with respect to the SM expectation, in final states with a high p_{T} photon and high $E_{\mathrm{T}}^{\mathrm{miss}}$, the latter being potentially a striking signature of new physics.
A mono-photon final state is predicted in several models, including supersymmetric and extra-dimentions models. In particular, the analysis presented here provides exclusion limits on DM related interpretations, as part of an extensive program of Dark Matter searches at LHC, carried out by the ATLAS and CMS Collaborations, and aiming at exploring different signatures to detect possible DM particles produced in high energy pp collisions. The results are interpreted in terms of a simplified DM model (Section 6.1) and of Axion-Like-Particles (ALPs) production, but this work will focus on the first interpretation. The model predicts a minimal extension of the Lagrangian with an additional $U(1)$ symmetry group, opening to possible interactions between DM and SM particles. This interaction happens by means of a vector or axial-vector mediator in an s-channel process. The photon characterizing the mono-photon signature arises from Initial State Radiation (ISR), as shown in the Feynman diagram of the process, in Fig. 8.1.

The analysis described in this chapter is performed using the full Run-2 data collected by ATLAS at $\sqrt{s} = 13$ TeV from 2015 to 2018, with an enhanced sensitivity with respect to the previously published results, obtained from 2015-2016 data at $\sqrt{s} = 13$ TeV [177] thanks to the increase of the integrated luminosity from 36.1 fb^{-1} to 139 fb^{-1} and the improvements in the physics objects reconstruction and calibration performance.

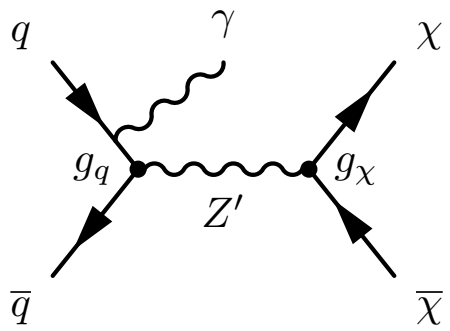


Figure 8.1: Feynman diagram representing the mono-photon signature: an s-channel pair production of DM particles, in association with one photon from ISR.

8.1 Monte Carlo samples

8.1.1 The signal MC samples

The signal model

The considered simplified DM model, in case of a vector ($\mathcal{L}_V$) or an axial-vector ($\mathcal{L}_{AV}$) mediator, is described by the Lagrangian [138]:

$$\mathcal{L}_V = g_q \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q}\gamma^\mu q + g_\chi Z'_\mu \bar{\chi}\gamma^\mu \chi \tag{8.1}$$

$$\mathcal{L}_{AV} = g_q \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q}\gamma^\mu\gamma^5 q + g_\chi Z'_\mu \bar{\chi}\gamma^\mu\gamma^5 \chi \tag{8.2}$$

where $q/\bar{q}$ and $\chi, \bar{\chi}$ are the Dirac spinor fields respectively of the SM quarks and of the DM particles and the coupling g_q with the BSM mediator Z'_μ is assumed to be universal to all quarks. Moreover, here the couplings to leptons are considered null in order to ensure avoiding constraints already set by searches for Drell-Yan processes, but a different scenario, including small coupling with leptons, is also considered.
The decay width is taken as the minimal width (Section 6.1.1), fixed by the choice of the couplings and of the mediator and DM masses by:

$$\Gamma^V_{min} = \frac{g_\chi^2 M_{\text{med}}}{12\pi}\left(1+\frac{2m_\chi^2}{M_{\text{med}}^2}\right)\beta_{DM}\theta(M_{\text{med}}-2m_\chi)+\sum_q \frac{3g_q^2 M_{\text{med}}}{12\pi}\left(1+\frac{2m_q^2}{M_{\text{med}}^2}\right)\beta_q\theta(M_{\text{med}}-2m_q) \tag{8.3}$$

$$\Gamma^{AV}_{min} = \frac{g_\chi^2 M_{\text{med}}}{12\pi}\beta_{DM}^3\theta(M_{\text{med}}-2m_\chi) + \sum_q \frac{3g_q^2 M_{\text{med}}}{12\pi}\beta_q^3\theta(M_{\text{med}}-2m_q) \tag{8.4}$$

where $\theta(x)$ is the Heaviside step function, while $\beta_X = \sqrt{1-4m_X^2/M_{\text{med}}^2}$ is the speed of the fermion $X = q, \chi$ with mass m_f in the mediator rest frame. For the considered signals, it varies from about $\Gamma_{med}/M_{\text{med}} \approx 2\%$ in the off-shell regime to $\Gamma_{med}/M_{\text{med}} \approx 6\%$ in the on-shell regime for heavy mediators.

MC production

The Monte Carlo simulations for the considered signal processes have been generated following the recommendations from the joint ATLAS/CMS Dark Matter Forum [138, 139] and anticipated in Section 6.1.1. The loose dependence of the kinematics on the spin structure of the interaction and on the couplings allows to generate only one model between the vector and the axial-vector with one choice of the couplings, and reinterpret the results in terms of other scenarios by means of cross-section rescaling [178]. The axial-vector mediator case is simulated, with the coupling to quarks set to $g_q = 0.25$ (universal for all quark flavours), and DM coupling $g_\chi = 1$. This choice allows to avoid strong constraints from di-jet searches, while satisfying the Narrow Width Approximation (NWA), with a width varying in the range $\Gamma_{med}/m_{med} \sim 2\%$ to $\Gamma_{med}/m_{med} \sim 6\%$.
MC samples corresponding to different masses of the DM and of the mediator are produced, covering a grid in the m_χ/M_{med} plane as shown in Figure 8.2, with an extension with respect to the analysis published on 2015+2016 data [177] to take into account the expected increase in sensitivity. The mass points density is higher in the transition region

from on-shell to off-shell mediator $M_{\text{med}} = 2m_\chi$, where the cross-section changes quickly, being enhanced by resonant production when reaching $M_{\text{med}} \simeq 2m_\chi$ and suppressed in the off-shell region at $M_{\text{med}} < 2m_\chi$.
The kinematics in the on-shell region has been verified to be overall independent also on the DM mass, as shown in the next section, thus leading to similar acceptance for a given mediator mass and allowing to employ the cross section rescaling procedure to extrapolate the results from the lowest to higher DM masses. Only the mass points close to the off-shell regions and the ones with lowest DM masses (the red ones in Figure 8.2) are genereated at reconstruction-level, including the detector effect simulation, while the remaining ones are generated only at truth-level for cross-section computation.
In total, 163 signal samples are generated at NLO, with a filter on the photon transverse momentum $p_{\text{T}}^{\gamma} > 130$ GeV, using the official ATLAS production chain, with MadGraph5_aMC@NLO v2.6.2 [156] in association with Pythia 8.235 [163] for the Parton Shower simulation, and the NNPDF3.0NLO set of Parton Distribution Functions [179]. The model used for the ME generation in MadGraph5_aMC@NLO is the DMsimp model with spin 1 [180, 181]. Among these signal samples, the 48 generated at reconstruction-level include the detector simulation in Fast Simulation (AF2 [160]); the differences with respect to Full Simulation (using Geant4 [159]) have been cross-checked to be negligible.

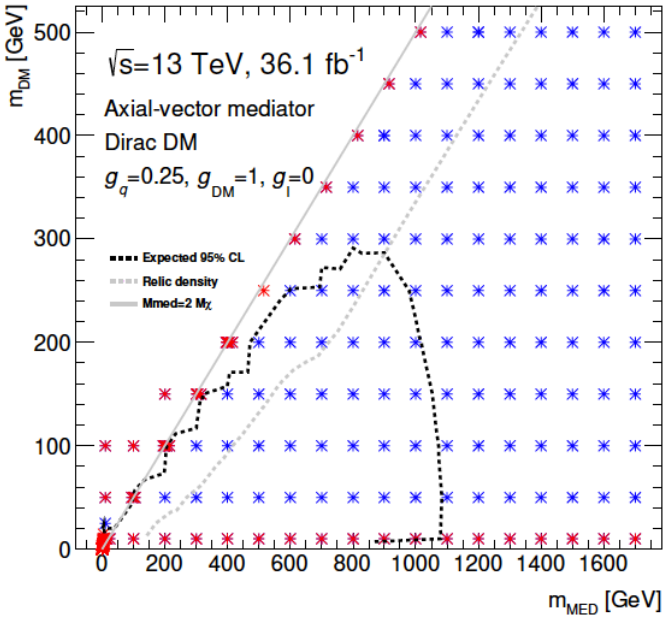


Figure 8.2: Grid of mono-photon signals generated in the $M_{\text{med}} - m_{DM}$ plane for the Full Run-2 analysis. The red points correspond to the signals generated ad reconstruction level, while the blue ones to the Truth-level simulations. The dotted line represents the limit obtained from the analysis of 2015+2016 data [177], corresponding to an integrated luminosity of 36.1 fb^{-1}.

Kinematic distributions In Figure 8.3, the distributions of $E_{\text{T}}^{\text{miss}}$ and of the photon p_{T} and η, normalized to the signal cross section, are shown for some of the generated MC16a signal samples with axial-vector mediator, scanning different choices of the mediator and

DM masses. The ratio is evaluated on the distributions normalized to unit area, to highlight the shape dependence on the DM and mediator masses. In the plots on the

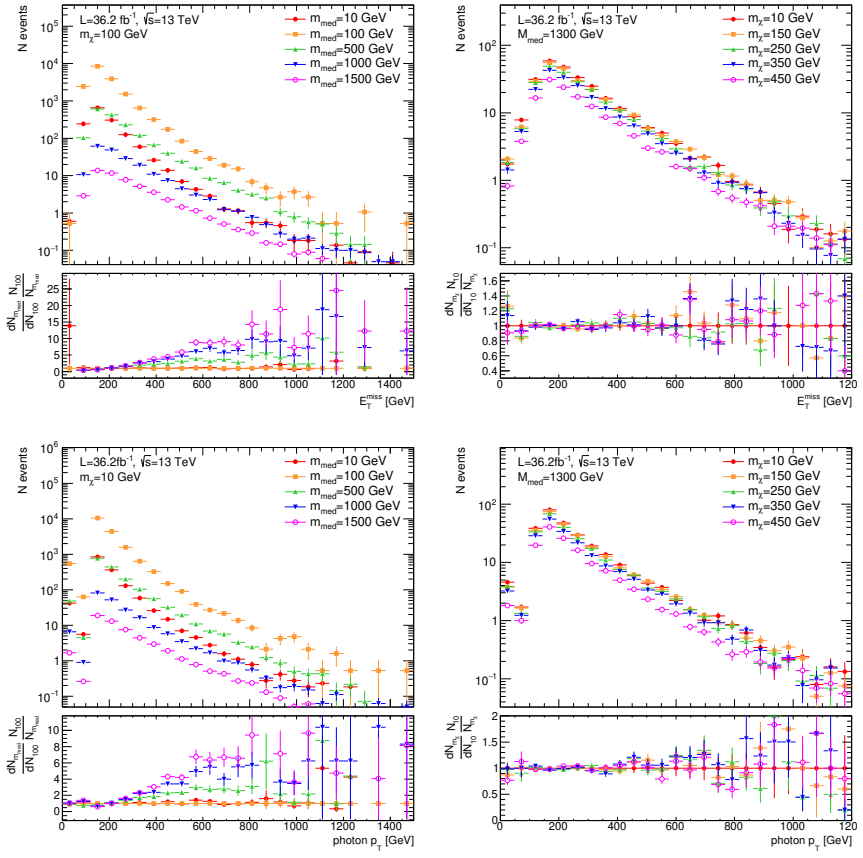


Figure 8.3: Distribution of the $E_{\mathrm{T}}^{\mathrm{miss}}$, and photon p_{T} and η for signal samples with an axial-vector mediator, $g_\chi = 0.25$ and $g_q = 1$. Different combinations of the mediator and DM masses are considered. In (a) the dependence on the M_{med} is shown, for a fixed $m_\chi = 10$ GeV. In (b) the dependence on the m_χ, for a fixed $M_{\mathrm{med}} = 1300$ GeV. The distributions are normalized to the cross-sections of the respective signal samples in the top panels, and to unity in the ratio plots. The ratios are evaluated with respect to the model with $M_{\mathrm{med}} = 100$ GeV and $m_\chi = 10$ GeV respectively in (a) and in (b).

left (Figure 8.3a), the mass of the DM is fixed ($m_\chi = 10$ GeV) and different values of the mediator mass are considered. The profiles of the kinematic variables show a dependence on the mass of the mediator. Moreover, as expected, the production cross section increases with decreasing mass of the mediator, and drops when reaching the off-shell limit $2m_\chi > M_{\mathrm{med}}$ (red points).
The plots on the right (Figure 8.3b) show the kinematic distributions for a fixed mass of the mediator ($M_{\mathrm{med}} = 1300$ GeV) and variable DM mass: no significant difference in shape is observed in this case when varying the mass of the DM particle within the on-shell regime, resulting in an almost constant acceptance of the analysis selections as a function of the DM mass. More specifically, the acceptance ranges from 0.567 ± 0.004 to 0.574 ± 0.004 for a DM mass in the range 10-300 GeV.

8.1.2 Background Monte Carlo samples

The following Standard Model processes constitute possible background sources for the analysis, mimicking the signal final state:

- $\gamma + Z(\to \nu\nu)$ ($Z(\nu\nu)\gamma$): irreducible background, with invisible neutrinos in place of DM particles;
- $\gamma + W(\to e\nu)$ ($W\gamma$): the electron is not reconstructed or it is reconstructed as a photon;
- $\gamma + W(\to \mu\nu)$ ($W\gamma$): the muon is not reconstructed;
- $\gamma + W(\to \tau\nu)$ ($W\gamma$): either a leptonically decaying τ-lepton is lost because the leptons are not reconstructed, or an hadronically decaying τ-lepton is reconstructed as a jet;
- $\gamma + Z(\to \ell\ell)$ ($Z(\ell\ell)\gamma$): both leptons are not reconstructed;
- jet+$Z(\to \nu\nu)$ (Z+jets): the jet is mistakenly reconstructed as a photon (jet faking photon, jet$\to \gamma$);
- jet+$W(\to e\nu)$ (W+jets): the electron (electron faking photon, $e \to \gamma$) or the jet is mistakenly reconstructed as a photon;
- jet+$W(\to \mu\nu)$ and jet+$W(\to \tau\nu)$ (W+jets): the jet fakes a photon and the lepton is not reconstructed, or the τ-lepton is reconstructed as a jet;
- γ+jets: a high fake $E_{\mathrm{T}}^{\mathrm{miss}}$ is produced by a miscalibration or misreconstruction of a jet or a photon;
- single t and diboson: similar to W/Z+jets.

The single t or diboson contributions are considered negligible. For the other ones, MC simulations are needed for the background estimation. The processes are simulated with Sherpa 2.2 MC event generator [182, 183], with Comix [184] and OpenLoops [185] matrix-element generators, matched to the Sherpa Parton Shower using the ME+PS@NLO prescription [186]. The PDF set is NNPDF3.0NNLO [179], in conjunction with a dedicated parton shower tuning developed by the Sherpa authors. For the Z/γ^* leptonic decays, a cut on the dilepton invariant mass $m_{ll} > 10$ GeV is applied at generator level.

Table 8.1 summarizes the details of the MC generation for the signal and SM background processes considered in the analysis.

Process	**Generators**	**PDF sets**	**Order**
DMsimp	MG5_aMC@NLO v2.6.2 + Pythia 8.235	NNPDF3.0NLO	NLO
$W/Z + \gamma$	Sherpa 2.2.2	NNPDF3.0NNLO	0,1j@NLO + 2,3j@LO
γ+jets	Sherpa 2.2.2	NNPDF3.0NNLO	1,2j@NLO + 3,4j@LO
W/Z+jets	Sherpa 2.2.1	NNPDF3.0NNLO	0,1,2j@NLO + 3,4j@LO

Table 8.1: Details of the generation of events for the signal samples and SM backgrounds considered in the analysis.

8.2 Physics Object definition and Overlap Removal

The definition of the physics objects selection criteria, as well as their isolation and identification Working Points (WPs), defined in Chapter 4, are summarized in Table 8.2, together with the $E_{\mathrm{T}}^{\mathrm{miss}}$ and jets WPs. Two categories of objects are introduced: the "baseline" objects enter the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation (Section 4.6), the Overlap Removal (OR) (Section 7.1.4), and are used to define the lepton vetoes in the analysis Control Regions; the "selected" ones are the objects that pass the analysis selections. In addition to p_{T} and $|\eta|$ thresholds, isolation and identification WPs, in order to suppress contribution from non-promp electrons and muons, selection cuts on the impact parameters with respect to the primary vertex are applied: more specifically, these selection criteria are defined on the significance of the transverse impact parameter ($|d_0/\sigma(d_0)|$) and on the longitudinal impact parameter z_0. Concerning hadronic jets, the EMTopo reconstruction is adopted, since using PFlow jets (see Section 4.4 for details about both the reconstruction strategies) has been observed to strongly affect the γ+jets yield in SR, as will be detailed later, in section 8.4.4. The OR procedure mentioned in Section 7.1.4 is applied as described in Table 7.1, following the recommendations by the Harmonization effort [165], to resolve ambiguities which can happen in the physics object reconstruction. In case of ambiguity between an electron and a photon, the electron is retained to improve the rejection of the $e \to \gamma$ background.

8.3 Event Selection

The final ROOT tree production from DAODs is performed through a SUSYTools [166] based analysis framework, in the AnalysisBase project, Release 21.2.99. The selection criteria defining the Signal Regions (SRs) and the Control Regions (CRs) of the analysis, mainly based on the 2015+2016 Run-2 analysis [177], are described in the following.
The data and MC events stored in the trees are pre-selected as described in Section 7.1.1. In addition, the event must pass the `HLT_g140_loose` trigger, which selects events with a photon satisfying the "loose" quality criteria and $p_{\mathrm{T}} > 140$ GeV. At DAOD level, also events passing a larger set of photon and $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers are stored. For the analysis at 36 fb^{-1}, the trigger efficiency was estimated with respect to an OR of the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers (`HLT_MET`):

$$\epsilon = \frac{Events\ passing\ \texttt{HLT\_g140\_loose}\ AND\ \texttt{HLT\_MET}}{Events\ passing\ \texttt{HLT\_MET}} \tag{8.5}$$

The resulting turn-on curve, as a function of the photon p_{T} is shown in Figure 8.4, showing an efficiency of more than 98.5% for photons with $p_{\mathrm{T}} > 150$ GeV, which are the ones selected in this analysis.

A signal event is characterized by a high $E_{\mathrm{T}}^{\mathrm{miss}}$, recoiling against a high p_{T} photon. The lowest $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold employed in the analysis is $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV. In total, four inclusive SRs with increasing $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds are defined, as well as three exclusive ones corresponding to different $E_{\mathrm{T}}^{\mathrm{miss}}$ bins, introduced to allow the "shape" fit described in Section 7.3.3. A summary of the $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds and bins is given in Table 8.3. In addition, a cut on the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance variable defined in Section 4.6.2 ($\mathcal{S} > 8.5$) is applied in order to reduce the contribution from γ+jets events, thanks to its discrimination power between events with real $E_{\mathrm{T}}^{\mathrm{miss}}$ and events where the $E_{\mathrm{T}}^{\mathrm{miss}}$ arises from mismeasurements or detector defects ("fake" $E_{\mathrm{T}}^{\mathrm{miss}}$). Figure 8.5 shows a comparison of the $\mathcal{S}$ distributions in γ+jets events and in two signal samples, for both the event-based and the object-based $\mathcal{S}$, defined in Section 4.6.2, highlighting higher values for the signals, as expected.

Object	Feature	"Baseline"	"Selected"
Photons	p_{T}	> 10 GeV	$p_{\mathrm{T}} > 150$ GeV
	$\lvert\eta\rvert$	< 2.37	< 2.37
	Identification WP (sec. 4.2.4)	*Loose*	*Tight*
	Isolation WP (sec. 4.2.4)	-	*FixedCutTight*
Electrons	p_{T}	> 7 GeV	> 7 GeV
	$\lvert\eta\rvert$	< 2.47	$\lvert\eta\rvert < 2.47$
	Track-to-vertex association	$\lvert z_0\rvert \sin\theta < 0.5$ mm $\lvert d_0\rvert/\sigma_{d_0} < 5.0$	$\lvert z_0\rvert \sin\theta < 0.5$ mm $\lvert d_0\rvert/\sigma_{d_0} < 5.0$
	Identification WP (sec. 4.2.5)	*MediumLH*	*MediumLH*
	Isolation WP (sec. 4.2.6)	-	*FixedCutLoose*
Muons	p_{T}	> 7 GeV	> 7 GeV
	$\lvert\eta\rvert$	< 2.7	< 2.7
	Track-to-vertex association	$\lvert z_0\rvert \sin\theta < 0.5$ mm $\lvert d_0\rvert/\sigma_{d_0} < 3.0$	$\lvert z_0\rvert \sin\theta < 0.5$ mm $\lvert d_0\rvert/\sigma_{d_0} < 3.0$
	Identification WP (sec. 4.3.1)	*Medium*	*Medium*
	Isolation WP (sec. 4.3.2)	-	*FCLoose*
Taus	p_{T}	> 20 GeV	-
	$\lvert\eta\rvert$	< 2.5	-
	Track selection	1 or 3 tracks	-
	Jet rejection	BDT-based (Loose)	
	Electron rejection	BDT-based	
	Muon rejection	OR ($\Delta R < 0.2$) Muons must not be CT	
Jets	Collection	AntiKt4EMTopoJets (EMTopo)	
	p_{T}	> 20 GeV	> 30 GeV
	$\lvert\eta\rvert$	< 4.5	< 4.5
	JVT WP (tab. 4.6)	-	Medium
	Jet Cleaning (sec. 4.4.6)	-	LooseBad
$E_{\mathrm{T}}^{\mathrm{miss}}$	WP (tab. 4.8)	Default	
	Soft Term	Track Soft Term (TST)	

Table 8.2: Selection criteria and isolation/identification Working Points for different physics objects

At least one "selected" photon is required, with $p_{\mathrm{T}} > 150$ GeV and $|\eta| < 2.37$, excluding the calorimeter crack region $1.37 < |\eta| < 1.52$. In addition it must not overlap with $E_{\mathrm{T}}^{\mathrm{miss}}$ ($\Delta\phi(\gamma, E_{\mathrm{T}}^{\mathrm{miss}}) > 0.4$) and the z coordinate, pointed by the photon with respect to the identified primary vertex, must be smaller than 250 mm ($|\Delta z_\gamma| < 250$ mm). This last criterion allows to suppress the beam-induced background [1] in CRs used for $jet \to \gamma$ background. In the other regions, this background is already strongly reduced by removing all events with at least one jet not satisfying loose jet cleaning criteria (Section 7.1.1).
The jet veto selects only events with at most one "selected" jet; if there is a jet, it must satisfy $\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, jet) > 0.4$, in order to suppress the contribution from events in which the $E_{\mathrm{T}}^{\mathrm{miss}}$ comes from a mismeasurement of the jet itself.
Finally, the lepton veto rejects all events with any "baseline" electron, muon or tau lepton.

[1] See section 4.4.6

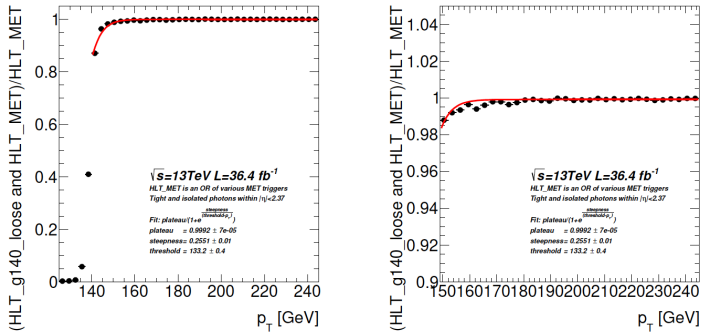


Figure 8.4: Trigger efficiency in data for the `HLT_g140_loose` trigger shown as a function of the leading photon p_{T} (left). A zoom into the region of interest is shown on the right.

	I1	I2	I3	I4	E1	E2	E3
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	> 200	> 250	> 300	> 375	200–250	250–300	300–375

Table 8.3: List of the $E_{\mathrm{T}}^{\mathrm{miss}}$ bins defining the analysis SRs.

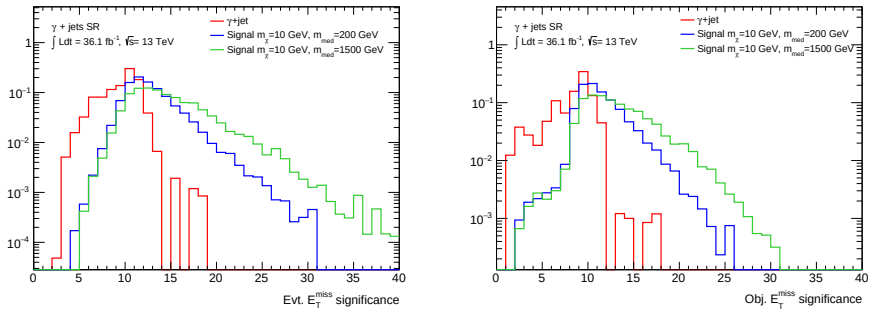


Figure 8.5: Comparison between the normalized $E_{\mathrm{T}}^{\mathrm{miss}}$ significance distribution in SR of mc16a $\gamma+$ jets and two signals with $m_{\chi} = 10$ GeV and $M_{med} = 200$ GeV or $M_{med} = 1500$ GeV. The event-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is shown on the left, and the object-based one on the right.

Main update with respect to 36 fb^{-1} analysis With respect to the previous version of the analysis [177], which used only 3 inclusive and 2 exclusive SRs, the $E_{\mathrm{T}}^{\mathrm{miss}}$ bins were reoptimized to profit from the increased statistics, and the lowest $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold had to be increased from 150 GeV to 200 GeV, to further suppress the contribution from the γ+jets background. A reoptimization was needed also for the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance ($\mathcal{S}$) cut due to non negligible differences in the variable shape, with respect to the previous analysis, arising from several updates introduced in the `Athena` framework at reconstruction level. The new object-based definition of this variable was also studied against the event-based one. Figure 8.6 shows the ROC curves obtained in the previous version of the analysis (red curve), with event-based definition, compared to the event-based and object-based ones from the new simulations (MC16a only), suggesting an overall similar performance between event- and object-based $\mathcal{S}$, but a slight degradation with respect to

the 2016 analysis. The choice of the final cut value prioritizes an high signal acceptance, required to be above 95%. The $\mathcal{S} > 8.5$ criterion is applied, ensuring a signal acceptance between 96% and 97% (depending on the mediator mass), with a corresponding rejection of $\sim$ 40% on $\gamma+$ jets events. Moving down to $\mathcal{S} > 8$, is observed to increase the signal effieicncy

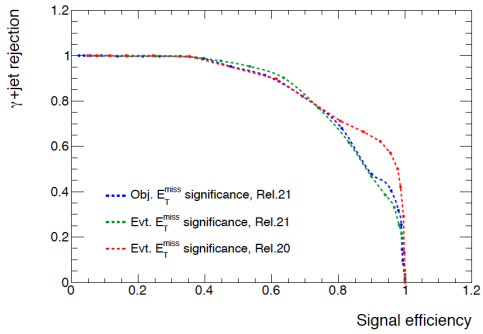


Figure 8.6: The $\gamma+$ jets rejection as a function of the signal efficiency, varying the $\mathcal{S}$ selection value. The results obtained with the previous Rel. 20 and with the event-based definition are compared with the ones from Rel. 21 with both the event-based and the object-based $\mathcal{S}$. The signal corresponds to a model with $m_{\chi} = 10$ GeV and $M_{med} = 200$ Gev

To conclude, Table 8.4 shows the number of events and the acceptance times efficiency ($A \times \epsilon$) for an example signal, after each selection applied in the SRI1, sorresponding to $E_{\mathrm{T}}^{\mathrm{miss}}$ >200 GeV.

Selection	**Yield**	$\boldsymbol{A \times \epsilon}$ **(%)**
Total	2409.7	100.00
Preselection	2119.2 ± 9.1	88.0
Photon trigger	1899.4 ± 8.6	78.8
$E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV	1170.2 ± 6.7	48.6
$E_{\mathrm{T}}^{\mathrm{miss}}$ significance > 8.5	1129.6 ± 6.6	46.9
Photon $E_{\mathrm{T}} > 150$ GeV	1105.5 ± 6.5	45.9
Photon ID	1040.7 ± 6.3	43.2
Photon isolation	979.8 ± 6.1	40.7
$\Delta\phi(\gamma,\ \vec{E}_{\mathrm{T}}^{\mathrm{miss}}) > 0.4$	976.9 ± 6.1	40.5
$\lvert\Delta z_{\gamma}\rvert$	964.7 ± 6.1	40.0
Jet veto	717.8 ± 5.2	29.8
Lepton veto	715.2 ± 5.2	29.7

Table 8.4: Yields and $A \times \epsilon$ after each cut of SRI1 for 139 fb^{-1} for a simplified model of dark matter production involving an axial-vector mediator and couplings $g_{\chi} = 1$, $g_q = 0.25$ and $g_{\ell} = 0$ with m_{χ}= 10 GeV and m_{med} =1 TeV. Generator-level filters are taken into account in the determination of the total number of events. Only the statistical uncertainty is shown.

8.4 Background estimation and modelling

The contribution of the background processes to the total observed events in the SR is estimated following different strategies:

- the dominant and irreducible background $Z(\nu\nu)\gamma$, as well as the secondary contributions from $W\gamma$, $Z(\ell\ell)\gamma$ and γ+jets, are estimated by constraining the MC expectation to data in the Control Regions (CR) described in Section 8.4.1 and built inverting one or more cuts of the SR, in order to enhance the contribution of one particular process and suppress the signal, as anticipated in Section 7.3.1.

- Data-driven (or in-situ) techniques are used for W/Z+jets , top and di-boson contributions and are based on electron/jet photon "fake" studies described in 8.4.2 and section 8.4.3.

8.4.1 Definition of the Control Regions

Four control regions (CR) are defined to constrain the normalization of the dominant $Z(\nu\nu)\gamma$ and secondary $W\gamma$, $Z(\ell\ell)\gamma$ and γ+jet backgrounds:

- **Single-Muon (1μ-CR or CR1mu)**: used to extract the normalization of the $W\gamma$ background in the signal region. The same selection of the SR is applied, but the muon veto is inverted, requiring exactly one "selected" muon in the event. The muons are treated as invisible particles in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, subtracting the muon term;
- **Two-Muon/Electron ($2\mu/e$-CR or CR2el/mu)**: used to constrain the normalization of both the $Z(\nu\nu)\gamma$ and the $Z(\ell\ell)\gamma$ in the SR. The same selection of the SR is applied, but the lepton veto which is inverted by requiring exactly two "selected" muons (electrons) in the event and no "baseline" electron (muon). As in the Single-Muon CR the muon (electron) term is subtracted from the $E_{\mathrm{T}}^{\mathrm{miss}}$ computation, such that muons (electrons) are treated as invisible particles. The two-muon (two-electron) invariant mass is required to be greater than 10 GeV to be coherent with the generator level cut of the $W\gamma$, $Z\gamma$ background samples. Both two-muon and two-electron CRs are considered, in order to improve the statistics;
- **Photon-Jet (γj-CRphjet)**: used to constrain the normalization of γ+jets background. The same selection of the SR is applied except for a lower $E_{\mathrm{T}}^{\mathrm{miss}}$ range: 85 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 110 GeV to enrich this region of γ+jets background. The requirement $\Delta\phi(\gamma, E_{\mathrm{T}}^{\mathrm{miss}}) < 3.0$ is applied to reduce possible signal contamination.

The modified $E_{\mathrm{T}}^{\mathrm{miss}}$ definition with "invisible" leptons is introduced in the lepton CRs in order to ensure a similar $E_{\mathrm{T}}^{\mathrm{miss}}$ spectrum as the in the SR, as shown in Figure 8.7, where the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution of $W\gamma$ and $Z\gamma$ processes in SR is compared with the $E_{\mathrm{T}}^{\mathrm{miss}}$ with invisible leptons in the 1μ-CR and the $2\mu/e$-CRs respectively. Moreover, in all the CRs, the criterion on the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is not applied, in order to enhance the background acceptance.

All the analysis regions are summarized in Table 8.5, where the full list of CRs selections is reported, compared with SR ones.

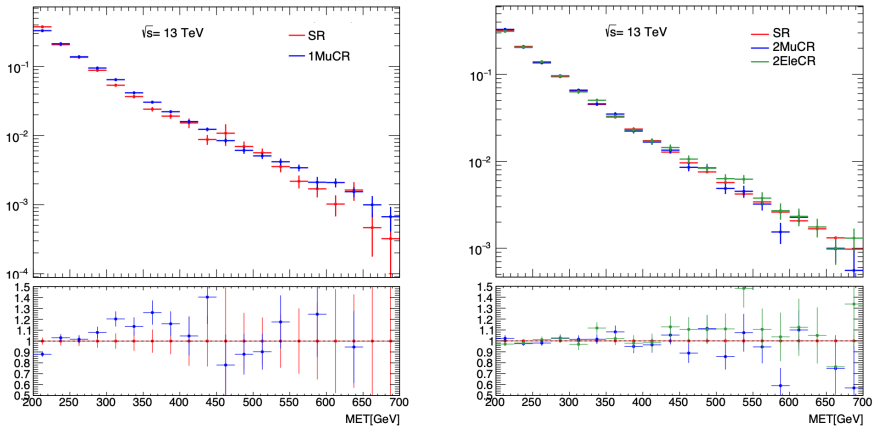


Figure 8.7: Comparison of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions, normalized to unit area, of $W\gamma$ (left) and $Z\gamma$ (right) events in the SR and in the lepton-CRs. In the latter, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is redefined by treating the leptons as invisible particles, in order to reproduce the kinematics in the SR.

Selection	**SR**	**1μ-CR**	**2μ-CR**	**2e-CR**	**γ+jet-CR**
Leptons	0 ℓ	1 μ	2 μ	2 e	0 ℓ
Jets	0 or 1 with p_{T} 30 GeV, $\lvert\eta\rvert < 4.5$ and $\Delta\phi$(jet, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$) > 0.4				
Leading γ	$p_{\mathrm{T}}^{\gamma} > 150$ GeV $\lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.37$ tight, isolated, $\lvert\Delta z_{\gamma}\rvert < 250$ mm				
	$\Delta\phi(\gamma,\ \vec{E}_{\mathrm{T}}^{\mathrm{miss}}) > 0.4$				$0.4 < \Delta\phi(\gamma,\ \vec{E}_{\mathrm{T}}^{\mathrm{miss}}) < 3.0$
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 200 GeV				85–110
$\mathcal{S}$	> 8.5	-			

Table 8.5: Summary of the selection criteria for the SRs and CRs of the mono-photon analysis.

8.4.2 Electrons faking photons

The "electons faking photons" ($e \to \gamma$) background consists of events, mainly from $W(\to e\nu)$+jets processes, with an electron mistakenly reconstructed and identified as a photon due to the partial ambiguity between the two, arising from their similar signatures in the EM calorimeter (Section 4.2.1).
The fraction of misidentified electrons is not easily reproducible by MC, thus making the employment of data-driven techniques extremely important to correctly estimate this background. The method used consists in 2 steps:

- evaluation of the *electron-to-photon fake rate*, which gives an estimate of the probability of electrons to be mistakenly reconstructed and identified as photons;
- rescaling of data in appropriate CRs (called the *probe-electron* CRs) by the previously measured *fake-rates*.

The electron-to-photon fake rate

The fake-rate is defined as:

$$\rho = \frac{N_{e^{true} \to \gamma^{reco}}}{N_{e^{true}}}. \tag{8.6}$$

In practice, this value can be measured from $Z \to ee$ data, selected requiring two "selected" electrons or a photon and an electron, with invariant mass compatible with the Z mass. Since a $Z \to e\gamma$ decay is not possible, the latter case can be associated to a $Z \to ee$ process, where one of the electrons has been misidentified as a photon. Given $N_{e\gamma}$ and N_{ee} the number of events with an $e\gamma$ or an ee pair respectively, the fake-rate can therefore be expressed in terms of these two measurable quantities, with $N_{e\gamma} << N_{ee}$:

$$F_{e\to\gamma} = \frac{N_{e\gamma}}{N_{ee}}, \tag{8.7}$$

More specifically, the two numbers are obtained by fitting the invariant mass distributions separately for the ee and $e\gamma$ cases, with a signal+background model: the signal is modelled with a double-sided crystal ball (DSCB) function, i.e. a gaussian distribution core with two power-law tails, while the background is given by an exponential of second-degree polynomial e^{-ax-bx^2}. The invariant mass fits on the inclusive m_{ee} and $m_{e\gamma}$ are shown in figure 8.8 top and bottom respectively, as an example.

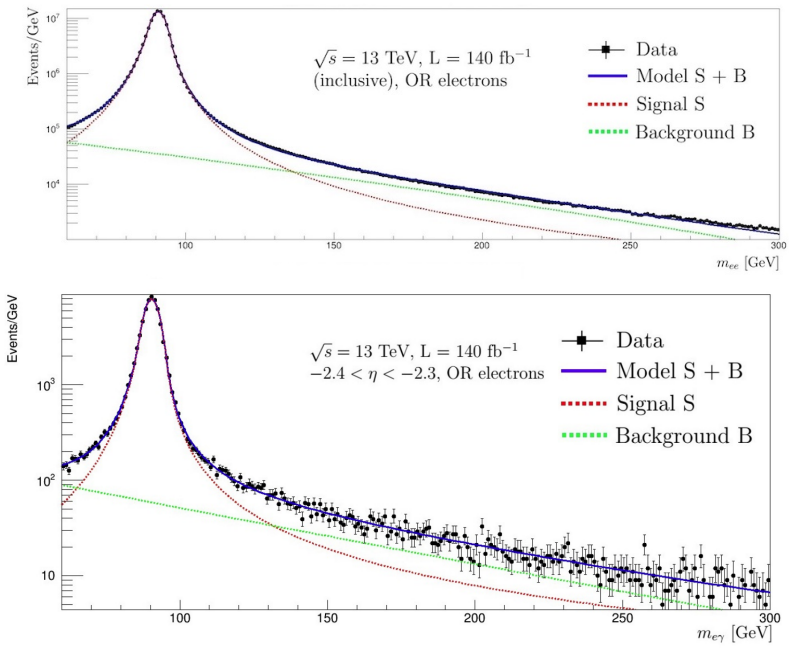


Figure 8.8: The two signal + background model fit on the inclusive invariant mass distribution m_{ee}. The signal and background models are shown by the red and green dotted line respectively, while the blue line shows the global model fit.

In the derivation of N_{ee}, each event is counted twice, if both electrons satisfy the analysis photon selection criteria, to account for the fact that either of the two might be

misidentified as a photon. The final number of ee and $e\gamma$ events is defined as the integral of the fitted signal curve in a range of $[m_Z^{peak} - 3\sigma, m_Z^{peak} + 3\sigma]$, where σ is the width of the gaussian core of the DSCB distribution.
Due to the possible dependence on p_{T} and η, the fake-rates are derived as a function of these two variables: the $|\eta|$ range is divided in 6 bins and in every $|\eta|$ bin the p_T range is divided in 8 bins for a total of 48 bins:

- 6 bins in $|\eta|$: 0-0.8, 0.8-1.15, 1.15-1.37, 1.52-1.81, 1.81-2.01, 2.01-2.37;
- 8 bins in p_T (in GeV): 25-35, 35-45, 45-55, 55-65, 65-75, 75-100, 100-150, 150-250.

Systematic uncertainties Three main sources of systematic uncertainties have been identified and estimated: the invariant mass range w_3 used to calculate the integrals of the signal distributions S_{ee} and $S_{e\gamma}$, the effect of the background subtraction and the bias in the reconstructed photons energy due to the fact that the photon energy calibration is applied for an object that should actually be an electron.
The first contribution is derived as the impact, on the fake-rate value, of varying the mass window from the nominal value $w_3 = [\mu_0 - 3\sigma_{ee}, \mu_0 + 3\sigma_{ee}]$, to $w_2 = [\mu_0 - 2\sigma_{ee}, \mu_0 + 2\sigma_{ee}]$ and $w_4 = [\mu_0 - 4\sigma_{ee}, \mu_0 + 4\sigma_{ee}]$. The uncertainty, taken as the maximum absolute variation with respect to nominal, is $\sim (1-4)\%$.
The systematic on the background subtraction is evaluated by integrating the m_{ee} and $m_{e\gamma}$ distributions in the mass range w_3, without performing any signal + background fit. The impact of this systematic uncertainty is about $\sim (3-8)\%$.
Finally, in order to estimate the last systematic source, the energy of the fake photons is increased and decreased by 1.5% (value obtained from a comparison between the m_{ee} and $m_{e\gamma}$ peak positions) before applying the selection criteria, and the maximum absolute difference between the nominal fake-rate and the up/down variated ones is quoted as uncertainty. It ranges from 8% to 10% depending on the η and p_{T} bin.

The results of the described procedure in bins of η and p_{T} are summarize in figure 8.9.

Probe-electron CRs

The number of fake photons in the SR and in the CRs can be obtained through the scaling, by the *fake-rates*, of the data yields in the so called probe-electron (probe-e) CRs, defined as the analysis region with the replacement of the selected photon with a probe-e (i.e. a "selected" electron satisfying the "selected" photon kinematic criteria). The trigger used for these CRs is more inclusive than for the other analysis regions, being an OR of the photon and $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers, in order to maximize the acceptance. The statistical uncertainty on the probe-e CR yields is propagated to the final estimate.
Table 8.6 summarizes the data yields in the probe-e CRs associated to the inclusive and exclusive SRs, as well as the final estimate of the $e \to \gamma$ contributions in the corresponding analysis regions. Same results have been derived for all the CRs and are reported, for inclusive regions only, in Appendix A.1. In general, the major contribution to the total uncertainty is the systematic uncertainty related to the electron-to-photon fake-rate.

8.4.3 Jets faking photons

Similarly to electrons, the jets can be mistakenly reconstructed as photons. The requirement of Tight Identification and Isolation for the photon largely suppresses this background, but the residual contribution is not well described by MC simulations and

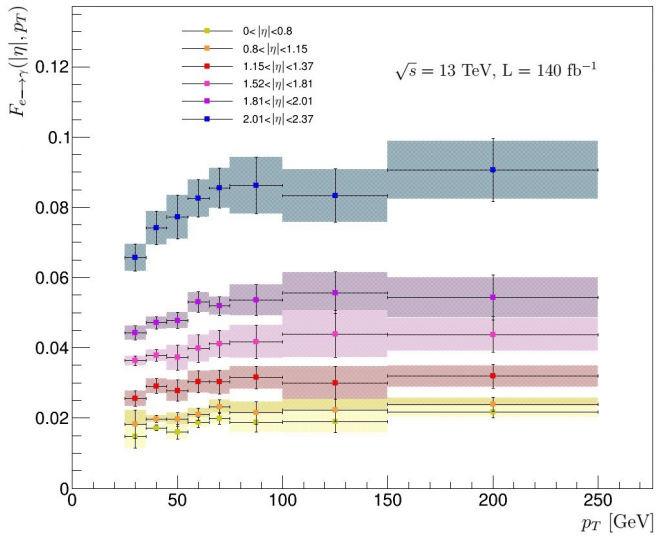


Figure 8.9: Fake rate as a function of p_T in different η bins. The error bars represent the total uncertainty, while the colored bands show the systematic contribution for each bin.

Region	Yields in Probe-e CR	$e \to \gamma$	Probe-e CR stat.	Fake-rate stat.	Fake-rate syst.	Total unc.
SRI1	26650	859.6	6.1	31.2	73.9	80.44
SRI2	11445	348.5	3.7	12.4	29.7	32.44
SRI3	5568	160.6	2.4	5.6	13.5	14.8
SRI4	2162	59.7	1.4	2.0	5.0	5.6
SRE1	15205	511.1	4.8	18.8	44.2	48.3
SRE2	5877	187.9	2.8	6.8	16.2	17.8
SRE3	3406	100.9	2.0	3.5	8.6	9.5

Table 8.6: Probe-e CR yields and final estimate of $e \to \gamma$s in each SR. The uncertainty is expressed in three terms: statistical uncertainty of data in the probe-e CR; statistical and systematic uncertainties related to the electron fake rate. The total uncertainty is the quadratic some of the three.

needs to be estimated in-situ.
Its estimation is performed using an ABCD method. The generalities of this method will be described in the next section, before going into the details of the specific realization for jets-faking-photons background estimation.

The ABCD method

The ABCD method is based on the definition of a 2-dimensional side-band matrix, obtained by inverting or loosening the selection criteria on two variables, to define 4 regions, as shown in Figure 8.10: the signal region (A), designed to be enriched with the signal, and three CRs (B, C and D) enriched with the background of interest. The leakage into the CRs of the signal and of different background processes then the one under study

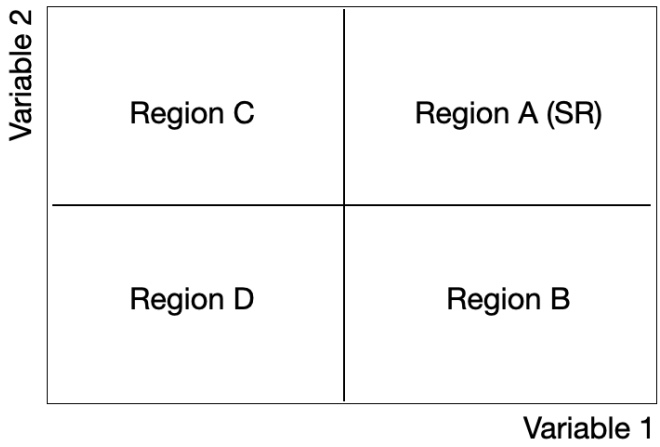


Figure 8.10: Definition of the four regions of the ABCD method.

should be minimized, and correctly taken into account for a reliable estimation.
Under the assumption that the CRs are devoid of signal and other background contribution, and that the two variables defining the ABCD plane are completely un-correlated, meaning that $N_A/N_B = N_C/N_D$, the number of background events in the region A can be derived from the data population in the three CRs:

$$N_A^b = \frac{N_B^{data} N_C^{data}}{N_D^{data}} \tag{8.8}$$

where the b super-script denotes the background under study. Since the two assumptions are typically not realized in practice, a subtraction of possible contamination in the CRs must be performed, and a correction factor R, estimated from MC simulations, can be introduced to account for possible correlation

$$N_A^b = R \times \frac{N_B^{b,data} N_C^{b,data}}{N_D^{b,data}}, \; R = \frac{N_A^{b,MC} N_D^{b,MC}}{N_B^{b,MC} N_C^{b,MC}} \tag{8.9}$$

In this last equation, the super-script $b, data$ denotes the data events after subtraction of unrelated background contributions.

The ABCD for jets-faking-photons background estimation

For the background under consideration, the two variables chosen to define the ABCD plane are the isolation and identification requirements on the photon. In the following, we indicate the number of events in the signal region, selecting tight and isolated photons, as N_A, and the yields in the other three regions as N_B (tight and non isolated photons), M_A (non tight and isolated photons), and M_B (non tight and non isolated photons), as sketched in figure 8.11. An isolation gap region is introduced between 0 GeV and 3 GeV, to minimize the leakage of signal photons in the non isolated CRs.

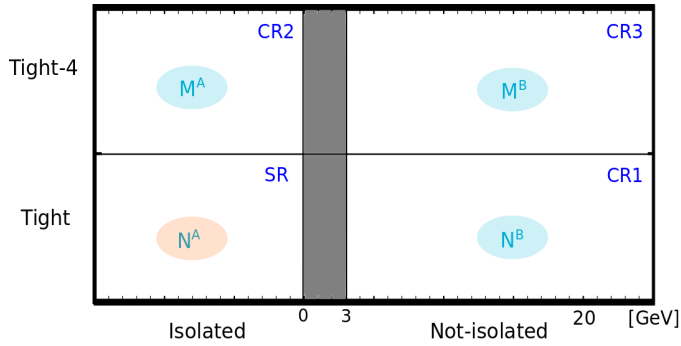


Figure 8.11: The definition of the four regions defining the ABCD method for the estimation of $jet \to \gamma$.

More specifically, isolated and non isolated photons are defined as (Table 4.2):

$$\text{Isolated photons} \to \text{TopoEtCone40} - 0.022\, p_{\mathrm{T}}^{\gamma} - 2.45\ \text{GeV} < 0 \tag{8.10}$$

$$\text{and ptcone20}/p_{\mathrm{T}}^{\gamma} < 0.05 \tag{8.11}$$

$$\text{Non isolated photons} \to \text{TopoEtCone40} - 0.022\, p_{\mathrm{T}}^{\gamma} - 2.45\ \text{GeV} > 3\,\text{GeV} \tag{8.12}$$

where the first component of the isolation is calorimetric, the second one is related to track isolation and no requirement on it is applied to define non isolated photons, in order to enhance the statistics. On the other side, different possible choices of the *non-tight* definition are possible, depending on which tightness requirements (Table 4.1) are removed:

$$\text{Tight} - 3 : F_{\text{side}},\ \Delta E,\ w_{s_3} \tag{8.13}$$

$$\text{Tight} - 4 : F_{\text{side}},\ \Delta E,\ w_{s_3},\ E_{ratio} \tag{8.14}$$

$$\text{Tight} - 5 : F_{\text{side}},\ \Delta E,\ w_{s_3},\ E_{ratio},\ w_{s_{tot}} \tag{8.15}$$

For the nominal results, the Tight-4 option was observed to be the one ensuring the smallest correlation among isolation and tightness variable, as required for the method. The other two options are used to derive systematic uncertainties on the final estimate. In the ideal case in which the two assumption previously described are satisfied, the number of background photons in the analysis regions can be estimated as:

$$N_{bkg}^{A} = \frac{N^{B} M^{A}}{M_{B}} \tag{8.16}$$

and consequently the number of signal photons in each region is:

$$N_{sig}^{A} = N_{A} - \frac{N^{B} M^{A}}{M_{B}} \tag{8.17}$$

from which we can define the purity as:

$$P = \frac{N_{sign}^{A}}{N^{A}} \tag{8.18}$$

In order to correct for deviations from the assumptions, the correlation factor R_{MC} is estimated from MC, as well as signal leakage coefficients c_1, c_2, c_3, accounting for possible contamination of the B,C and D regions with true photons

$$R_{MC} = \frac{N^A_{bkgMC}}{N^B_{bkgMC}} \frac{M^B_{bkgMC}}{M^A_{bkgMC}}, \quad c_1 = \frac{N^B_{sign}}{N^A_{sign}}, \quad c_2 = \frac{M^A_{sign}}{N^A_{sign}}, \quad c_3 = \frac{M^B_{sign}}{N^A_{sign}} \tag{8.19}$$

Including these two terms in the formula,

$$N^A_{sig} = N^A - \left((N^B - N^A_{sig} c_1) \frac{M^A - N^A_{sig} c_2}{M^B - N^A_{sig} c_3} \right) R_{MC} \tag{8.20}$$

from where the N^A_{sig} and the corrected purity can be derived:

$$\begin{aligned} P = & \frac{\left(M^B + N^A c_3 - N^B c_2 R_{MC} - M^A c_1 R_{MC}\right)}{2N^A (c_1 c_2 R_{MC} - c_3)} \\ & \cdot \left(-1 + \sqrt{1 + \frac{4(c_1 c_2 R_{MC} - c_3)(N^A M^B - N^B M^A R_{MC})}{\left(M^B + N^A c_3 - N^B c_2 R_{MC} - M^A c_1 R_{MC}\right)^2}} \right) \end{aligned} \tag{8.21}$$

The number of fake photons in the region A is given by:

$$N_{jet \to \gamma} = (1 - P) N^A \tag{8.22}$$

Estimation of the MC coefficients

- The correlation factor is estimated separatedly on background photons from Z+jets and W+jets MC samples, as a cross-check of the consistency of the results. The two values are then combined for the final R_{MC} estimate.
- Concerning the signal leakage coefficients, three estimates are performed for each of them, using $Z(\ell\ell)\gamma$, $Z(\nu\nu)\gamma$ and $W\gamma$ samples. While good consistency is observed between the 3 values for c_2 and c_3, a systematic difference is observed between the c_1 value from $W\gamma$ and the two values from the $Z\gamma$ samples. This difference is tracked down to be due to different tails in the isolation profile, and it is accounted for as a source of systematic uncertainty not only on c_1, but also on the other coefficient related to isolation, c_3.

Systematic uncertainties on isolation and tightness

In addition to the systematic uncertainties related to the signal leakage coefficients, two other systematic sources are accounted for: the uncertainty related to the identification is evaluated as the maximum difference between the nominal purities (based on Tight-4 non-tight selection) and the ones obtained with Tight-3 and Tight-5 definition of the non-Tight CR; similarly, the uncertainty on the isolation is derived by varying the isolation gap from 3 GeV to 2 GeV and 4 GeV and taking the maximum impact. The total systematic uncertainty is the square sum of all systematic uncertainties.

Final estimate of the jet$\to \gamma$ background

Table 8.7 summarizes the results for the 7 SRs adopted in the analysis. The same estimation was performed also in the corresponding CRs and are reported in Appendix A.1. The

results for the jets-faking-photons in each region are dominated by the systematic error coming from the tightness control regions and by the statistical uncertainty on R_{MC}.

Region	$N_{jet\to\gamma}$	Stat.	ID syst	Iso. syst	c_1 stat	c_2 stat	c_3 stat	R stat	c_1 syst	c_3 syst	Total syst	Total error
SRI2	114	16	31	2	2	7	0	15	10	0	37	40
SRI3	54	11	10	0	1	5	0	11	5	0	17	20
SRI4	20	7	5	1	1	3	0	7	2	0	9	11
SRE1	136	16	10	4	2	8	0	16	10	0	23	28
SRE2	56	11	23	1	1	5	0	10	5	0	26	29
SRE3	33	9	8	1	1	4	0	8	3	0	13	16

Table 8.7: Jets-faking-photons contribution in the analysis regions, as estimated using the ABCD method. All the statistical and systematic uncertainty contributions are reported, as well as the total uncertainty on the estimate.

8.4.4 The γ+jets background

The modelling of γ+jets background arising from "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ is particulary challenging and dependent on the pile-up conditions. An important contribution to the yield in the SR was found to be given by particularly anomalous events in which a signal jet is not reconstructed, for example due to an incorrect identification of the primary vertex or energy mismeasurement and imprecise pile-up discrimination, thus perfectly reproducing the signal signature and even providing high $E_{\mathrm{T}}^{\mathrm{miss}}$ significance $\mathcal{S}$. In order to reduce this contribution, which was observed to be not precisely modelled, the lowest $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold was increased from 150 to 200 GeV with respect to the analysis at 36 fb^{-1}, as previously mentioned. Indeed, the combination of photon $p_{\mathrm{T}} > 150$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV results particularly effective in selecting events with a lost signal jet.
This background was investigated in detail. Among the several studies performed, a cross-check of the γjet-CR reliability in a VR with $110 < E_{\mathrm{T}}^{\mathrm{miss}} < 150$ GeV is reported, as well as a performance comparison between EMTopo jets and PFlow jets.

Validation Region

In order to cross-check the validity of the low $E_{\mathrm{T}}^{\mathrm{miss}}$ CR for the γ+jets background, a validation region (VR) is defined with the same criteria used for the SR, except for a lower $E_{\mathrm{T}}^{\mathrm{miss}}$ range covering the $E_{\mathrm{T}}^{\mathrm{miss}}$ region between 110 and 150 GeV, in-between the SR and the CR. In addition, in order to suppress a possible signal contamination, the requirement $\Delta\phi(\gamma, E_{\mathrm{T}}^{\mathrm{miss}}) < 3.0$ is applied. The $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions of data in this region and in the CR are compared with MC background expectations, with V+jets MC rescaled to the $(e \to \gamma) + (j \to \gamma)$ in-situ estimates, in Figure 8.12. Except for the global normalization, no evident trend is observed, thus suggesting a good reliability of the CR.

Comparison between EMTopo and PFlow jets

As already mentioned, the EMTopo jets were used in this analysis due to the loss in performance of PFlow jets in the γ+jets topology. These studies were performed with a $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV threshold, before taking the decision to move to 200 GeV. As reported in Table 8.8, when using the PFlow jets, the number of γ+jets events entering the inclusive

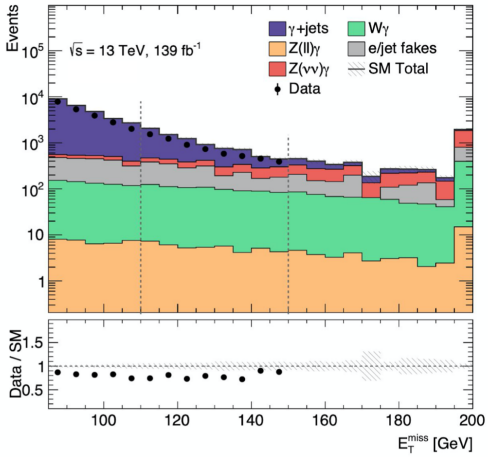


Figure 8.12: Data and background $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions in γjet-CR ($85 < E_{\mathrm{T}}^{\mathrm{miss}} < 110$ GeV) and γjet-VR ($110 < E_{\mathrm{T}}^{\mathrm{miss}} < 150$ GeV).

SR with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV is increased by more than a factor three with respect to EMTopo ones, mainly as a result of different $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance cut acceptances.

	PFlow	**EMTopo**
SR w/o $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\mathcal{S}$ cuts	23007	25479
SR	985	266

Table 8.8: The γ+jets yields with EMTopo and PFlow jets, before applying $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance selection (w/o $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\mathcal{S}$ cuts) and after (SR).

Figure 8.13 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance distributions: the Medium WP for EMTopo jets and both Medium and Tight for PFlow jets are shown (see Table 4.6 for the WP definition), while the Tight WP is adopted for the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution (Table 4.8). With PFlow jets, a larger amount of events with $E_{\mathrm{T}}^{\mathrm{miss}}$ close to the photon p_{T} threshold of 150 GeV is observed, as well as higher values of $E_{\mathrm{T}}^{\mathrm{miss}}$ significance. These are hints of an higher probability of loosing a signal jet, thus leaving the photon unbalanced. The higher $E_{\mathrm{T}}^{\mathrm{miss}}$ significance values, and the drop in events at $E_{\mathrm{T}}^{\mathrm{miss}}$ significance ~ 2 can be interpreted as a consequence of the lost jes: as a matter of fact, the denominator of the object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is related to the resolution of all hard objects entering $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, thus being expected to be higher for events with a larger jet multiplicity, given the worse resolution of jets with respect to photons.

In figures 8.14 an event-by-event comparison between the PFlow and EMTopo jets is shown, for events passing SR requirements using PFlow jet reconstruction, while being rejected with the EMTopo one. The p_{T} of the leading reconstructed PFlow and EMTopo jets, matched to the same truth jet based on ΔR, is shown (top left), as well as their $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$, i.e. the ratio between their reconstructed transverse momentum and the truth-level one (top right), and their JVT value (bottom) (section 4.4.5). What is observed from the top plot is that the PFlow jets in the events of interest are affected by an underestimation of the p_{T}, such that, differently from the corresponding EMTopo jets, they enter the p_{T} range ($20 < p_{\mathrm{T}} < 60$ GeV) in which the JVT cut is applied and are

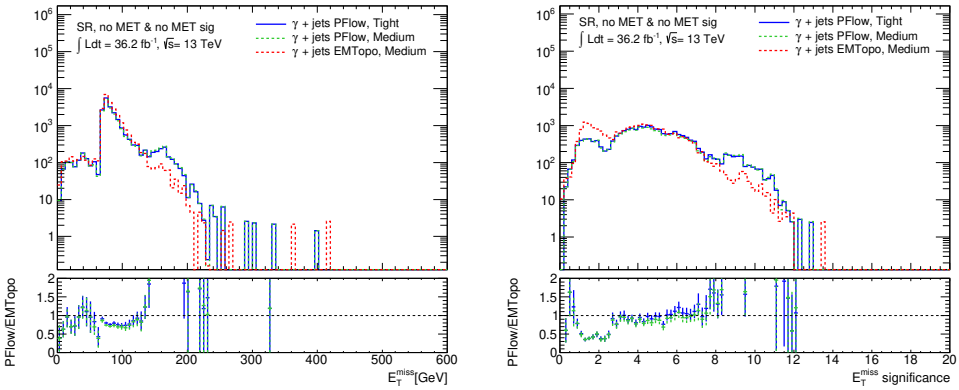


Figure 8.13: Distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance (right) with EMTopo jets with Medium JVT working point (red line) and PFlow jets with Tight (blue line) and Medium (green line) JVT working point. All the SR seletion criteria are applied, except for $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance. The step behaviour in $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution at 70 GeV is due to preselections applied on an OR of the different $E_{\mathrm{T}}^{\mathrm{miss}}$ definition employed in the analysis (standard $E_{\mathrm{T}}^{\mathrm{miss}}$ and the ones with muons or electrons treated as invisible particles).

therefore rejected due to their extremely low JVT value.
In order to better characterize these jets, a study of the properties of the leading PFlow jet has been performed over a larger sample, including all preselected events. In the left plot of figure 8.15 the charge fraction f_{ch} versus $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ of the leading PFlow jet, matched to a truth jet, is shown. Two populations can be distinguished, with a bulk of events centered at about $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth} \sim 1$ and $f_{ch} \sim 0.6$, and a thin band of events centered at $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth} < 1$ and with $f_{ch} < 0.05$, the latter being the population of interest. The plot on the right, shows the JVT versus $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ for this population, selected with $f_{ch} < 0.1$. The mismeasured jets have very low JVT (meaning no or few tracks associated to the PV), while events with negative JVT (i.e. without tracks at all) are well reconstructed in transverse momentum. The underestimation seems therefore due to possible errors either in the track association to the PV or in the PV identification. In these cases, the energy of the particle associated to the track is included in EMTopo jet reconstruction, by including its calorimetric component. On the other hand, in PFlow jet reconstruction, the track is completely lost since it's not included as charged component of the PFlow object (being mistakenly not associated to the PV and therefore not selected), but its energy is subtracted from the cluster at cell-subtraction stage, due to the fact that no PV matching is performed as part of the track selection for the subtraction step (as explained in section 4.4.1).

8.4.5 Background composition in the analysis regions

An overview of the background as resulting from data-driven estimates for $e/jet \rightarrow \gamma$ backgrounds and pure MC for the ones with genuine photon, is given here. Figure 8.16 shows the background composition in all the analysis SRs. In addition, Table 8.9 summarizes all background yields in each separate bin of $E_{\mathrm{T}}^{\mathrm{miss}}$, for SRs and CRs.

In order to validate the CRs, detailed comparisons of the main kinematic variable distributions in data and background expectation were performed. For simplicity, only the $E_{\mathrm{T}}^{\mathrm{miss}}$, p_{T}^{γ} and $\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma)$ distributions in the CRs, are shown in Figure 8.17. An

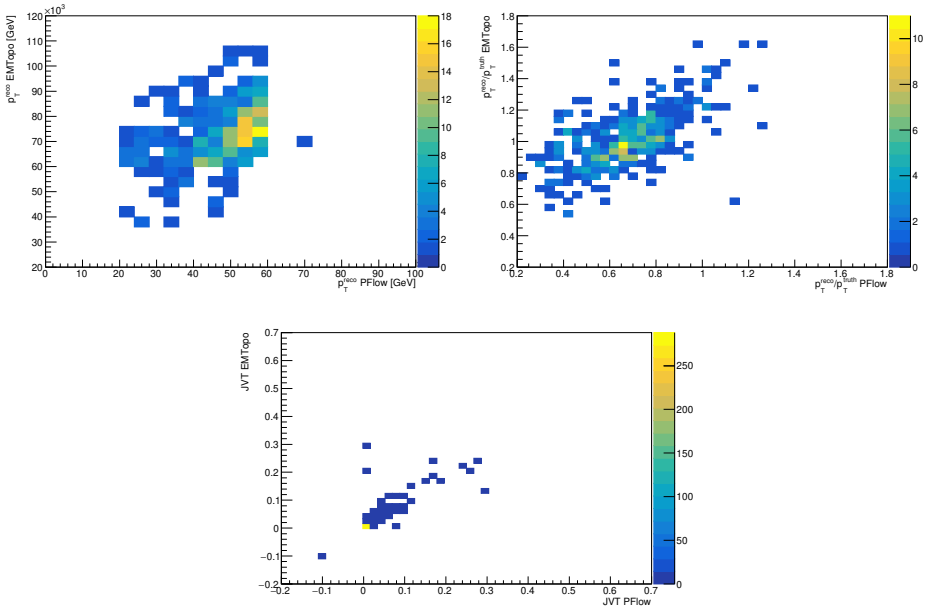


Figure 8.14: Event-by-event comparison between the leading truth-matched PFlow jet (x-axis) and EMTopo jet (y-axis). Only events satisfying the SR selection criteria with PFlow jets, but rejected with the EMTopo ones are included. The reco p_{T} (top left), $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ (top right) and JVT value (bottom) are shown.

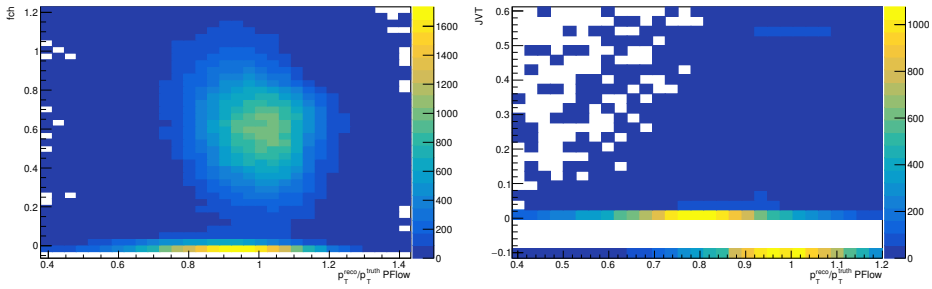


Figure 8.15: Left: f_{ch} versus $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ of the leading truth-matched PFlow jet, for preselected events. Right: JVT versus $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ of the leading truth-matched PFlow jet, for preselected events with $fch < 0.1$

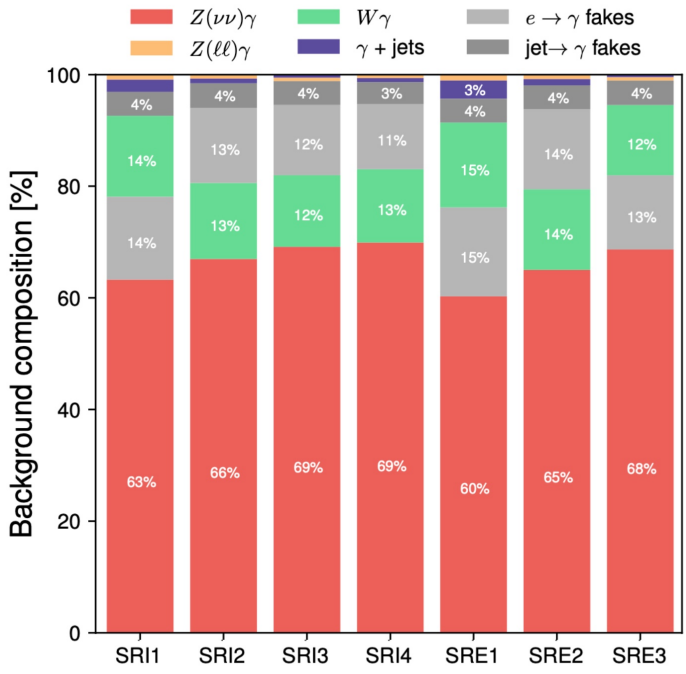


Figure 8.16: The background composition in the SRs. The $jet \to \gamma$ and $e \to \gamma$ backgrounds are obtained from data-driven estimates, while the other backgrounds are estimated from MC.

offset in the global normalization is expected, and will be corrected through the fit to data in the CRs, as explained in the next sections, but the ratios between data and background expectations show in general a relatively flat behaviour, especially for the $E_{\mathrm{T}}^{\mathrm{miss}}$ variable, highlighting an overall good modelling of MC simulation and providing a confirmation of the CRs reliability.

SR	**[200,250] GeV**	**[250,300] GeV**	**[300,375] GeV**	**>375 GeV**
$Z(\nu\nu)\gamma$	1940 ± 1300	850 ± 58	522 ± 36	358 ± 26
$Z(\ell\ell)\gamma$	33.7 ± 3.2	10.5 ± 0.98	4.75 ± 0.45	3.15 ± 0.30
$W\gamma$	490 ± 50	189 ± 19	96 ± 9.8	67.2 ± 7
γ + jets	105 ± 55	14.4 ± 7.6	3.26 ± 1.7	3.57 ± 1.9
$e \to \gamma$ fakes	511 ± 48	188 ± 18	101 ± 9.4	59.7 ± 5.5
$j \to \gamma$ fakes	136 ± 28	56.1 ± 29	33.3 ± 16	20.4 ± 11
1μ-CR	**[200,250] GeV**	**[250,300] GeV**	**[300,375] GeV**	**>375 GeV**
SM events	1290 ± 110	549 ± 52	328 ± 32	210 ± 21
$Z(\nu\nu)\gamma$	0.95 ± 0.06	0.5 ± 0.04	0.31 ± 0.05	0.09 ± 0.02
$Z(\ell\ell)\gamma$	82.3 ± 7.7	32.2 ± 3.8	17.2 ± 1.7	9.69 ± 1.3
$W\gamma$	1070 ± 100	457 ± 43	268 ± 25	174 ± 18
γ + jets	13.6 ± 11	0 ± 0	$2.6^{+6.6}_{-2.6}$	$0.6^{+3.9}_{-0.6}$
$e \to \gamma$ fakes	38.9 ± 3.8	13.8 ± 1.4	7.32 ± 0.83	3.84 ± 0.51
$j \to \gamma$ fakes	91.8 ± 19	45.5 ± 26	33 ± 15	20.9 ± 11
2μ-CR	**[200,250] GeV**	**[250,300] GeV**	**[300,375] GeV**	**>375 GeV**
SM events	271 ± 21	119 ± 9.5	74.5 ± 6.4	44.8 ± 4
$Z(\nu\nu)\gamma$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
$Z(\ell\ell)\gamma$	262 ± 19	114 ± 8.6	71.8 ± 5.9	44.1 ± 3.9
$W\gamma$	$0.02^{+0.16}_{-0.02}$	$0.16^{+0.21}_{-0.16}$	0.14 ± 0.11	0.17 ± 0.01
γ + jets	0 ± 0	0 ± 0	0 ± 0	0 ± 0
$e \to \gamma$ fakes	1.03 ± 0.21	0.59 ± 0.16	0.18 ± 0.07	0.11 ± 0.05
$j \to \gamma$ fakes	$8.2^{+9.3}_{-8.2}$	4.6 ± 3.8	$2.4^{+2.5}_{-2.4}$	$0.41^{+0.70}_{-0.41}$
2e-CR	**[200,250] GeV**	**[250,300] GeV**	**[300,375] GeV**	**>375 GeV**
$Z(\nu\nu)\gamma$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
$Z(\ell\ell)\gamma$	200 ± 11	91.1 ± 7.2	56.5 ± 4.3	39.4 ± 3.1
$W\gamma$	0.4 ± 0.17	0.14 ± 0.07	0.25 ± 0.2	0.21 ± 0.06
γ + jets	0 ± 0	0 ± 0	0 ± 0	$0.40^{+0.40}_{-0.40}$
$e \to \gamma$ fakes	0.1 ± 0.06	0.14 ± 0.06	0.07 ± 0.04	0.23 ± 0.14
$j \to \gamma$ fakes	$2.9^{+5.3}_{-2.9}$	5.1 ± 4.5	3.5 ± 2.9	$1.1^{+2.3}_{-1.1}$
γ+j-CR	**$E_{\mathrm{T}}^{\mathrm{miss}} \in [85, 110]$ GeV**			
SM events	26500 ± 6200			
$Z(\nu\nu)\gamma$	400 ± 61			
$Z(\ell\ell)\gamma$	38.5 ± 5.4			
$W\gamma$	656 ± 92			
γ + jets	24000 ± 6100			
$e \to \gamma$ fakes	694 ± 66			
$j \to \gamma$ fakes	750 ± 230			

Table 8.9: Background yields in each bin of $E_{\mathrm{T}}^{\mathrm{miss}}$, for SRs and CRs. The uncertainties include both statistical and systematic sources. The $j \to \gamma$ and $e \to \gamma$ contributions are estimated through data-driven techniques, while the other background are estimated from MC.

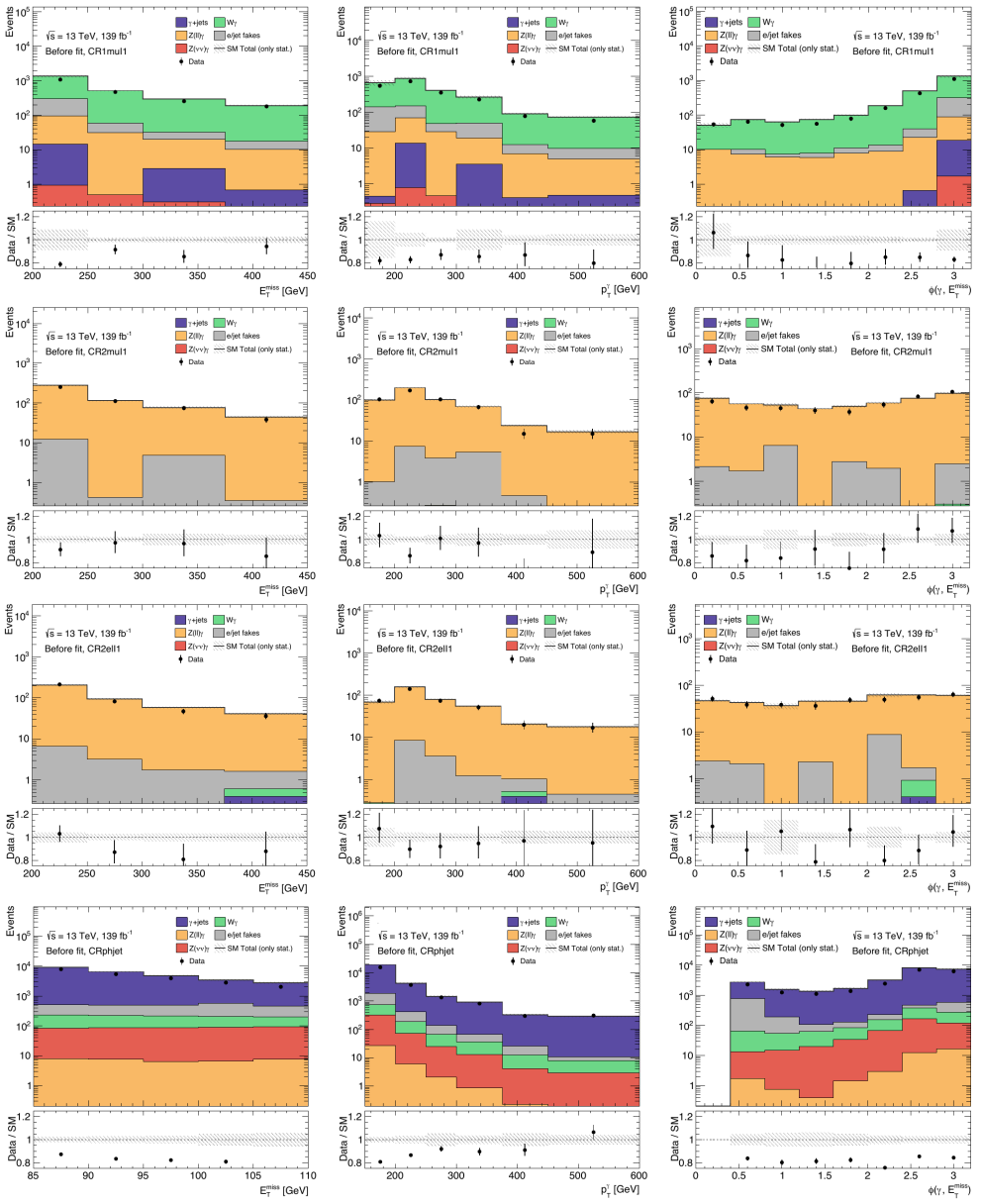


Figure 8.17: Distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$, p_{T}^{γ} and $\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma)$ in the 1μ-CR, 2μ-CR, $2e$-CR and γjet-CR for data and background. The background predictions are based on pure MC for backgrounds with a genuine photon, and data-driven methods for $e/jet \rightarrow \gamma$. The dashed band includes only statistical uncertainties. The bottom part of each plot shows the ratios of data to expected-background event yields.

8.5 Systematic uncertainties

Several sources of experimental and theoretical systematic uncertainties are taken into account in the analysis for the background estimation and signal modelling.

8.5.1 Experimental systematics

A list of all general systematic sources is provided in Table 7.2, including all uncertainties provided by the Combined Performance (CP) groups (see Section 7.2), i.e. photon and lepton reconstruction/identification/isolation efficiency, jet JVT efficiency, a contribution from pile-up reweighting, energy scale and resolutions for all physics objects and $E_{\mathrm{T}}^{\mathrm{miss}}$ related uncertaities. The luminosity uncertainty of 1.7% is also included, as well as the few analysis-specific uncertainties related to the data-driven background estimates, listed in Table 8.10.

Uncertainties from data-driven methods		
$jet \to \gamma$	JetFake_syst	Variation of isolation and identification requirements
	JetFake_stat_REGION	Statistics of the ABCD control regions
$e \to \gamma$	EleFake_syst	Variation of mass window and difference from MC
	EleFake_stat	Statistics of the data for fake rate
	EleFake_statCR_REGION	Statistics of probe-e control regions

Table 8.10: List and description of the systematic contributions of uncertainty.

Smoothing of the experimental systematics In case of statistical fluctuations, the systematic uncertainty contribution in some $E_{\mathrm{T}}^{\mathrm{miss}}$ bins might be over-estimated. To overcome this limit, which could potentially result in unwanted fit instabilities, a smoothing procedure has been applied to each systematic contribution and MC sample in the SRs and CRs, using a standard tool of common usage within ATLAS, the *CommonSystSmoothingTool*. Few examples of the impact of the smoothing on some relevant systematic sources are shown in Figure 8.18.

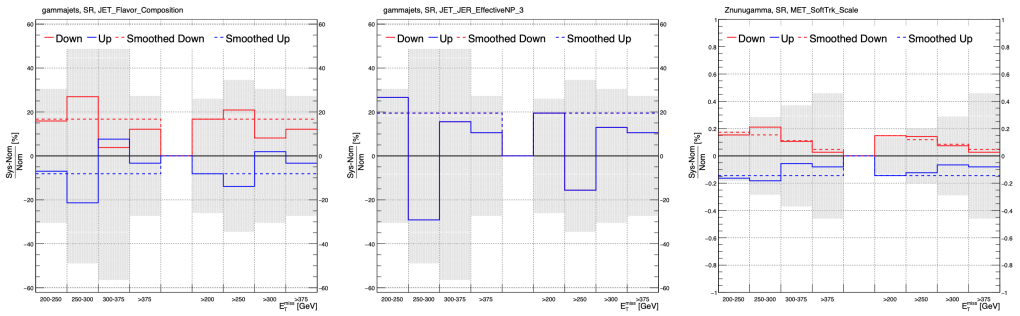


Figure 8.18: Some examples of the effect of the smoothing algorithm on some systematic uncertainties: jet flavor composition (left) and JER EffectiveNP_3 (middle) for the γ+jets background in the SR; $E_{\mathrm{T}}^{\mathrm{miss}}$ SoftTrk Scale for the $Z(\to \nu\nu)+\gamma$ background in the SR (right). The smoothed distributions are shown with dashed lines.

Anti-SF The uncertainties in the efficiency may also affect the lepton veto in the SR and in the CRs. For this reason, an additional systematic on the reconstruction, identification and isolation inefficiency of the vetoed leptons has been implemented, by means of the definition of muon, electron and tau anti-SFs, as an overall value for each MC sample, defined as[2]:

$$\text{anti} - \text{SF} = 1 + (1 - < \text{SF}_{\text{tot}} >) \frac{N_{>0\ell}}{N_{0\ell}} \tag{8.23}$$

where SF_{tot} is the mean of the "preselected" leptons SFs in each event; $< \text{SF}_{\text{tot}} >$ is the mean over all the events passing a selection region (antiSF-CR or antiSF-SR) defined from the most inclusive SR and CRs, without applying the lepton veto/selections (the dependence of the antiSF on the $E_{\text{T}}^{\text{miss}}$ bin was verified to be negligible); $N_{>0\ell}$ and $N_{0\ell}$ are respectively the number of events in the antiSF-SR/CR with at least one lepton or with no leptons. The anti-SF values in the SR are listed, for different background processes, in Table 8.11, showing very small deviations from 1. Similar results are obtained also in the CRs and reported in Appendix A.5.

SR	e-antiSF	μ-antiSF	τ-antiSF	Total antiSF
$\text{Z}(\nu\nu) + \gamma$	0.9946	0.9946	0.9945	0.9945
$\text{Z}\gamma$	1.0039	1.0048	0.9990	1.0077
$\text{W}\gamma$	1.0081	1.0049	0.9994	1.0124
γ+jets	1.0004	1.0016	0.9997	1.0016

Table 8.11: e-antiSF, μ-antiSF, τ-antiSF and their product in the inclusive SR.

8.5.2 Theoretical uncertainties on the signals and backgrounds

Theoretical uncertainties associated to background and signals are evaluated following the PDF4LHC recipes [167].The total PDF+α_S uncertainty is the quadratic sum of the contributions detailed in 7.2. All the needed variations are present in the NLO Sherpa Monte Carlo samples and in MadGraph ones as *on-the-fly* weights to be applied event-by-event. This allows a full statistical correlation between the varied samples and the nominal ones. In addition, PS uncertainties are derived for the signal samples, as described in Section 8.5.2.

Theoretical uncertainties on the backgrounds

The relative uncertainties for each variation are shown in Table 8.12. Due to low statistics in high-$E_{\text{T}}^{\text{miss}}$ bins, the uncertainty evaluated in the first inclusive bin is used for all the bins. The single uncertainty contributions are combined in quadrature for each $E_{\text{T}}^{\text{miss}}$ bin and for each MC sample, and included as a single nuisance parameter in the likelihood fit. Full correlation is assumed between SRs and CRs, as well as among the $E_{\text{T}}^{\text{miss}}$ bins, for a given process: $Z\gamma$ (with $Z(\nu\nu) + \gamma$ and $Z(\ell\ell) + \gamma$ considered correlated), $W\gamma$ or γ+jets. The impact of uncorrelating these systematic uncertainties among $E_{\text{T}}^{\text{miss}}$ bins has been cross-checked to be negligible, while the correlation between SR and CRs in each individual bin introduces a large cancellation of the theory uncertainties.

[2] From the assumption: $N_{\text{tot}} =< \text{SF}_{\text{tot}} > N_{>0\ell} + \text{anti} - \text{SF} N_{0\ell} = N_{>0\ell} + N_{0\ell}$

γ + jets	Nominal (NNPDF30)	altPDF (CT14nnlo)	altPDF (MMHT)	alphaS	scale_ME		PDF_ME	Total Uncertainty	
CR1muI	15.83	+1.42%	+1.03%	± 3.42%	-25.83%	+22.42%	± 2.47%	-26.17%	+22.81%
CR2muI	0.00	-	-	-	-	-	-	-	-
CR2elI	0.00	-	-	-	-	-	-	-	-
CRphjet	23977.90	+0.23%	-0.16%	±2.13%	-13.94%	+16.35%	±1.00%	-14.13%	+16.52%
SRI	126.22	+0.96%	+0.97%	± 2.43%	-15.42%	+21.01%	±1.46%	-15.67%	+21.20%
$W\gamma$									
CR1muI	1965.49	-0.02%	+2.09%	± 1.11%	-6.59%	+6.43%	±1.37%	-6.83%	+6.67%
CR2muI	0.49	+0.03%	+1.77%	± 0.79%	-9.11%	+3.19%	±2.46%	-9.47%	+4.11%
CR2elI	1.00	-0.87%	+2.30%	± 1.28%	-6.84%	+6.97%	±2.32%	-7.34%	+7.46%
CRphjet	650.24	+0.49%	+1.70%	±1.31%	-9.76%	+12.25%	±1.20%	-9.92%	+12.38%
SRI	841.81	+0.50%	+2.36%	±0.90%	-6.39%	+6.66%	±1.54%	-6.63%	+6.89%
$Z\gamma$									
CR1muI	141.32	+0.35%	+1.71%	±1.08%	-6.37%	+4.84%	±1.76%	-6.69%	+5.26%
CR2muI	485.77	+0.28%	+1.36%	±1.16%	-5.08%	+1.93%	±1.61%	-5.45%	+2.76%
CR2elI	386.97	+0.51%	+1.97%	±1.06%	-3.87%	+1.38%	±1.74%	-4.37%	+2.46%
CRphjet	36.95	+0.97%	+1.67%	±1.05%	-6.77%	+6.93%	±1.56%	-7.02%	+7.18%
SRI	52.03	-0.46%	+1.12%	±0.95%	-5.81%	+6.39%	±1.83%	-6.17%	+6.71%
$Z(\nu\nu)\gamma$									
CR1muI	1.85	+0.19%	+1.32%	±1.32%	-5.90%	+2.17%	±1.47%	-6.23%	+2.94%
CR2muI	0.00	-	-	-	-	-	-	-	-
CR2elI	0.00	-	-	-	-	-	-	-	-
CRphjet	399.99	+0.33%	+0.89%	±1.64%	-12.02%	+15.17%	±1.10%	-12.19%	+15.30%
SRI	3664.87	+0.25%	+1.62%	±1.15%	-4.53%	+2.59%	±1.65%	-4.96%	+3.28%

Table 8.12: Impact of theoretical uncertainties, in percentage, on the event yields of $V\gamma$ and γ+jets samples in SRI1 ($E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV) and in the corresponding CRs.

Impact of NLO electroweak corrections to $Z + \gamma$ samples

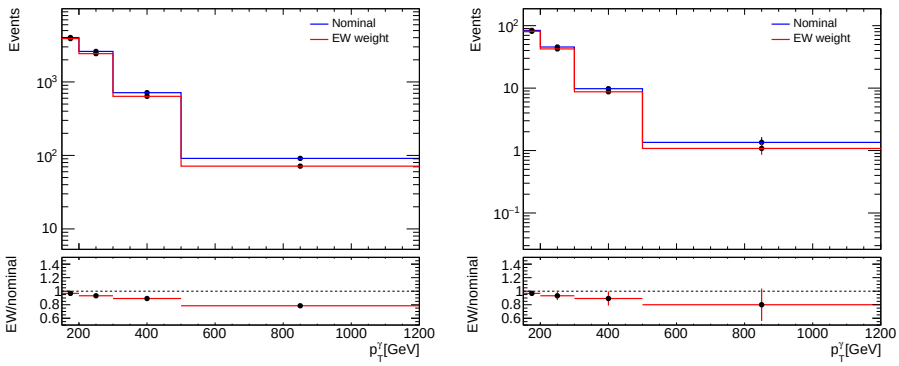


Figure 8.19: E_T^γ distribution reweighted with the NLO electro-weak corrections in an inclusive signal region SRI0 with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV and $E_{T^\gamma} > 150$ GeV, for the $Z(\nu\nu) + \gamma$ samples (left) and the $Z(\ell\ell) + \gamma$ samples (right). The bottom panels show the ratio of the reweighted distributions to the nominal ones. Only statistical uncertainties are shown.

Electroweak (EW) corrections at NLO for the $Z(\ell\ell) + \gamma$ and $Z(\nu\nu) + \gamma$ samples were made available, in bins of E_T^γ, when the analysis was in an already advanced status. This corrections has been applied to estimate their impact on the analysis results. The reweighted E_T^γ distributions are shown in Figure 8.19 for an inclusive signal region SRI0 with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV and $E_T^\gamma > 150$ GeV. As detailed in the Appendix A.3, these corrections were tested to not affect significantly the results, and are not applied nor considered as a source of uncertainty in the final fit.

Theoretical uncertainties on DM simplified model

The total $PDF + \alpha_S$ systematic uncertainties on the acceptances for a subset of signal points (the complete table can be found in Appendix A.4) and in the exclusive and last inclusive regions are shown in Table 8.13, along with the uncertainties on the cross-section. The latter includes the same sources of systematic uncertainties, but with an estimation of their impact on the global cross-section rather than on the number of events passing the analysis selections. They are not included in the fit but only shown as an error band on the observed limit in the exclusion plots.
Final and Initial State Radiation (FSR/ISR) and Multi Parton Interaction (MPI) uncertainties are also evaluated through the generation of additional samples with variations of the A14 MC tune[3]. One pair mainly accounts for underlying event effects (VAR1), one pair mainly for jet structure effects (VAR2), and three pairs for different aspects of extra jet production (VAR3a, VAR3b, VAR3c). For each MC sample, all the variation pairs are combined to get full coverage of the tune uncertainties: the total uncertainty, reported in the last column of Table 8.13, is derived in the first inclusive $E_{\mathrm{T}}^{\mathrm{miss}}$ bin, as the sum in quadrature of the highest between up and down impacts of each variation pair, and added as a single nuisance parameter in the likelihood fit.

m_χ (GeV)	m_{med} (GeV)	Acceptance (% uncertainty) SRE1		SRE2		SRE3		SRI4		Cross-section (% uncertainty)		MC Tuning (% uncertainty)
10	10	-2.18	+2.39	-2.36	+2.68	-3.13	+3.60	-3.52	+3.85	-4.27	+5.18	±9.15
10	100	-2.17	+2.49	-2.60	+3.07	-3.21	+3.79	-3.80	+4.26	-4.27	+5.18	±13.59
10	500	-2.09	+2.16	-2.39	+2.48	-2.61	+2.71	-3.41	+3.52	-2.93	+3.33	±5.19
10	1000	-2.59	+2.47	-2.77	+2.63	-3.51	+3.38	-4.10	+3.36	-3.27	+3.30	±5.97
10	1500	-3.81	+2.90	-3.71	+3.04	-4.11	+3.09	-5.89	+4.04	-4.42	+3.60	±4.32
100	10	-2.05	+2.13	-2.43	+2.52	-2.84	+2.98	-3.72	+3.85	-2.96	+3.40	±8.22
100	100	-2.07	+2.15	-2.44	+2.56	-2.56	+2.63	-3.57	+3.69	-2.96	+3.40	±6.47
100	215	-1.99	+2.16	-2.47	+2.81	-2.71	+3.02	-3.24	+3.44	-3.14	+3.80	±5.81
500	1015	-2.92	+2.75	-3.11	+2.93	-3.60	+3.06	-4.93	+4.09	-3.67	+3.50	±8.31

Table 8.13: PDF and scale uncertainties, in %, on the acceptance in bins of $E_{\mathrm{T}}^{\mathrm{miss}}$ and on the cross section of simplified dark matter models. In the last column, the MC tuning uncertainties are reported for inclusive $E_{\mathrm{T}}^{\mathrm{miss}}$.

8.6 Statistical analysis

The likelihood function describing the signal and background model is given by eq. 7.1, with

$$N_R^{exp} \propto \mu_{sig} \times n_R^{sig}(\boldsymbol{\theta}) \tag{8.24}$$
$$+ \; k_{Z\gamma}(n_R^{Z(\nu\nu)+\gamma}(\boldsymbol{\theta}) + n_R^{Z(\ell\ell)+\gamma}(\boldsymbol{\theta})) \tag{8.25}$$
$$+ \; k_{W\gamma} n_R^{W(\ell\nu)+\gamma}(\boldsymbol{\theta}) \tag{8.26}$$
$$+ \; k_{\gamma\mathrm{jets}} n_R^{\gamma+\mathrm{jets}}(\boldsymbol{\theta}) \tag{8.27}$$
$$+ \sum_{B \neq W\gamma, Z\gamma, \gamma\mathrm{jets}} n_R^{(B)}(\boldsymbol{\theta}) \tag{8.28}$$

[3] https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCTuningRecommendations

and the θ dependence as defined in Eq 7.4. The final background estimation is obtained by means of the background-only maximum-likelihood fit performed simultaneously in all CRs as described in Section 7.3.3. A sketch of the CRs and the normalization factors ($k_{Z\gamma}$, $k_{W\gamma}$ and $k_{\gamma\text{jet}}$) they constrain is shown in Figure 8.20.

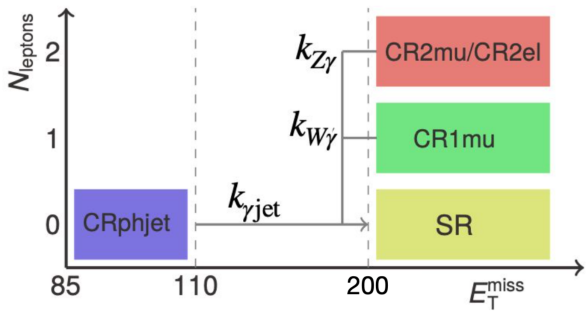


Figure 8.20: Definition of the CRs used in the analysis and corresponding normalization factors.

8.6.1 Results in Validation Region

In order to test the fit machinery and the background estimation, a single bin background-only fit to data in CRs with $E_{\text{T}}^{\text{miss}} > 150$ GeV is performed, and the normalization factors obtained from the fit are extrapolated to the VR described in Section 8.6.1, and compared to observed data. Pre- and post-fit results are reported in Table 8.14 and Figure 8.21, showing good agreement, within 1σ, between data and expected background yields after the fit.

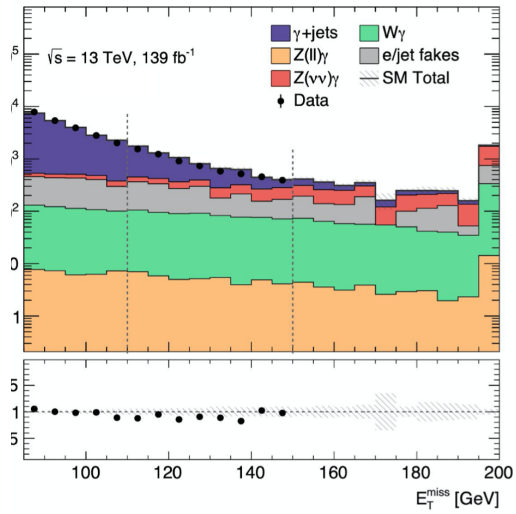


Figure 8.21: Data and background $E_{\text{T}}^{\text{miss}}$ distributions in γjet-CR ($85 < E_{\text{T}}^{\text{miss}} < 110$ GeV) and γjet-VR ($110 < E_{\text{T}}^{\text{miss}} < 150$ GeV) after the background-only fit, to be compared with pre-fit distributions in Figure 8.12.

	VR	**1μ-CR**	**2μ-CR**	**2e-CR**	**γj-CR**
Observed events	1017	4425	1005	801	21991
Post-fit SM events	1102 ± 95	4425 ± 67	1000 ± 27	805 ± 23	21990 ± 1.50
Post-fit $Z(\nu\nu)\gamma$	310 ± 19	3.87 ± 0.18	0 ± 0	0 ± 0	379 ± 18
Post-fit $Z(\ell\ell)\gamma$	8.41 ± 0.82	317 ± 11	965 ± 28	761 ± 23	36.6 ± 4.7
Post-fit $W\gamma$	219 ± 15	3520 ± 100	0.81 ± 0.13	1.98 ± 0.62	555 ± 24
Post-fit $\gamma + jets$	311 ± 74	41.7 ± 8.5	0 ± 0	10.9 ± 2.2	19580 ± 290
Post-fit $e \to \gamma$ fakes	229 ± 22	152 ± 14	4.45 ± 0.58	0.72 ± 0.26	694 ± 65
Post-fit $j \to \gamma$ fakes	26 ± 12	394 ± 74	30 ± 11	31 ± 10	750 ± 230
Pre-fit SM events	1230 ± 150	5110 ± 300	1050 ± 54	855 ± 43	26500 ± 4600
Pre-fit $Z(\nu\nu)\gamma$	327 ± 28	4.08 ± 0.29	0 ± 0	0 ± 0	400 ± 31
Pre-fit $Z(\ell\ell)\gamma$	8.88 ± 0.87	335 ± 20	1019 ± 53	807 ± 41	38.5 ± 4.7
Pre-fit $W\gamma$	259 ± 30	4170 ± 260	0.96 ± 0.16	2.37 ± 0.69	656 ± 56
Pre-fit $\gamma + jets$	379 ± 100	51.1 ± 10	0 ± 0	13.6 ± 1.1	24000 ± 4600
Pre-fit $e \to \gamma$ fakes	229 ± 22	152 ± 14	4.45 ± 0.58	0.72 ± 0.26	694 ± 66
Pre-fit $j \to \gamma$ fakes	26 ± 12	394 ± 74	22 ± 11	31 ± 11	750 ± 230

Table 8.14: Results of the background estimation in the VR defined with $110 < E_{\mathrm{T}}^{\mathrm{miss}} < 150$ GeV using full Run-2 data. The CRs used in the estimation are those corresponding to the signal region defined with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV. Results are obtained with a background-only fit in the CRs. Each component is reported before (bottom) and after the fit (top). The errors shown include both the statistical and systematics uncertainties.

8.6.2 Results in Signal Region

The results of the background-only simplified-shape fit in the SR are presented in this section. The estimated factors after the fit are reported in Table 8.15, while table 8.16 summarizes the expected background yields with their uncertainties after the fit and the number of observed events in data in the different analysis regions. The same results are summarized in Figure 8.22 and Figure 8.23, showing respectively post fit $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions in the CRs and both pre- and post-fit ones in the SR. Finally, Figure 8.24 summarizes the post-fit data and background yields in all the analysis regions. The bottom panel in the Figure show the significance ($(n_{obs} - n_{exp})/\sigma$) in each region, highlighting no excess of data with respect to the SM expectation.

	SRE1	**SRE2**	**SRE3**	**SRI4**
$\mu_{W\gamma}$	0.81 ± 0.09	0.84 ± 0.11	0.74 ± 0.11	0.85 ± 0.13
$\mu_{Z\gamma}$	0.99 ± 0.08	0.89 ± 0.09	0.90 ± 0.11	0.86 ± 0.12
$\mu_{\gamma+\mathrm{jets}}$		0.82 ± 0.21		

Table 8.15: Normalization factors obtained from a background-only simplified shape fit performed on the exclusive $E_{\mathrm{T}}^{\mathrm{miss}}$ bins defined in Table 8.8. The errors shown include both the statistical and systematic uncertainties.

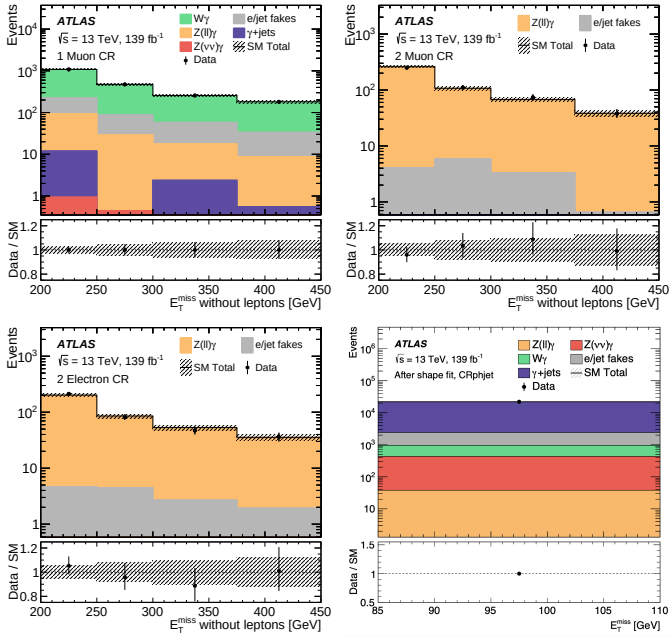


Figure 8.22: Post-fit distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the 1muCR, 2muCR, 2eCR and PhJetCR for background predicted with a shape fit to the luminosity of the full Run-2 dataset. The dashed band includes statistical and systematic uncertainties. The lower part of the figure shows the ratios of data to expected-background event yields.

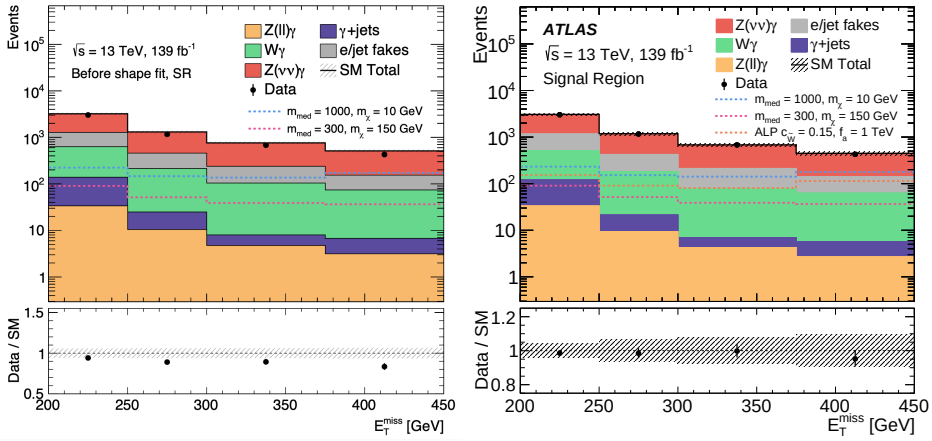


Figure 8.23: Pre-fit (left) and post-fit (right) distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the SR for background predicted with a shape fit to the luminosity of the full Run-2 dataset. The dashed band includes statistical and systematic uncertainties. The lower part of the figure shows the ratios of data to expected-background event yields.

SR	**[200,250] GeV**	**[250,300] GeV**	**[300,375] GeV**	**>375 GeV**
Observed events	3023	1164	679	427
SM events	3070 ± 130	1182 ± 75	680 ± 53	448 ± 42
$Z(\nu\nu)\gamma$	1910 ± 110	758 ± 65	468 ± 49	306 ± 40
$W\gamma$	394 ± 22	159 ± 15	71.0 ± 8.2	56.7 ± 7.1
$Z(\ell\ell)\gamma$	33.2 ± 2.4	9.32 ± 0.89	4.26 ± 0.48	2.69 ± 0.37
γ + jets	87 ± 35	11.9 ± 4.8	2.7 ± 1.1	3.0 ± 1.2
$e \to \gamma$ fakes	511 ± 48	188 ± 18	100.9 ± 9.5	59.7 ± 5.6
$j \to \gamma$ fakes	136 ± 28	56 ± 29	33 ± 16	20 ± 11
1 μ CR	**[200,250] GeV**	**[250,300] GeV**	**[300,375] GeV**	**>375 GeV**
Observed events	1081	473	256	181
SM events	1081 ± 33	473 ± 22	256 ± 16	181 ± 13
$Z(\nu\nu)\gamma$	0.938 ± 0.062	0.444 ± 0.041	0.278 ± 0.031	0.075 ± 0.010
$Wgamma$	857 ± 40	385 ± 35	198 ± 22	147 ± 17
$Z(\ell\ell)\gamma$	81.0 ± 5.9	28.7 ± 2.8	15.4 ± 1.8	8.3 ± 1.1
γ + jets	11.0 ± 6.3	–	2.1 ± 1.2	0.48 ± 0.27
$e \to \gamma$ fakes	38.9 ± 3.7	13.8 ± 1.4	7.32 ± 0.82	3.84 ± 0.51
$j \to \gamma$ fakes	92 ± 19	46 ± 26	33 ± 15	21 ± 11
2 μ CR	**[200,250] GeV**	**[250,300] GeV**	**[300,375] GeV**	**>375 GeV**
Observed events	250	111	74	38
SM events	261 ± 13	107.4 ± 8.4	68.0 ± 6.6	38.4 ± 4.8
$Z(\nu\nu)\gamma$	–	–	–	–
$W\gamma$	0.0187 ± 0.0050	0.134 ± 0.037	0.101 ± 0.029	0.143 ± 0.041
$Z(\ell\ell)\gamma$	257 ± 14	101.6 ± 8.6	64.7 ± 6.7	37.7 ± 4.8
γ + jets	–	–	–	–
$e \to \gamma$ fakes	1.03 ± 0.21	0.60 ± 0.16	0.177 ± 0.071	0.111 ± 0.050
$j \to \gamma$ fakes	3.0 ± 5.4	5.1 ± 4.0	3.1 ± 2.7	0.40 ± 0.53
2 e CR	**[200,250] GeV**	**[250,300] GeV**	**[300,375] GeV**	**>375 GeV**
Observed events	213	82	47	36
SM events	202 ± 11	85.6 ± 6.8	53.0 ± 5.2	35.6 ± 4.3
$Z(\nu\nu)\gamma$	–	–	–	–
$W\gamma$	0.324 ± 0.075	0.113 ± 0.027	0.183 ± 0.046	0.178 ± 0.045
$Z(\ell\ell)\gamma$	198 ± 11	81.1 ± 7.0	50.3 ± 5.3	33.7 ± 4.4
γ + jets	–	–	–	0.329 ± 0.091
$e \to \gamma$ fakes	0.098 ± 0.063	0.143 ± 0.060	0.065 ± 0.038	0.23 ± 0.14
$j \to \gamma$ fakes	4.2 ± 5.1	4.2 ± 3.6	2.4 ± 2.1	1.2 ± 1.6
γ+jets CR	**$E_{\mathrm{T}}^{\mathrm{miss}} \in [85, 110]$ GeV**			
Observed events	21991			
SM events	21990 ± 150			
$Z(\nu\nu)\gamma$	393 ± 47			
$W\gamma$	527 ± 39			
$Z(\ell\ell)\gamma$	37.9 ± 5.2			
γ + jets	19588 ± 290			
$e \to \gamma$ fakes	694 ± 65			
$j \to \gamma$ fakes	750 ± 230			

Table 8.16: Post-fit background yields and observed data in all $E_{\mathrm{T}}^{\mathrm{miss}}$ bins of SR and CRs.

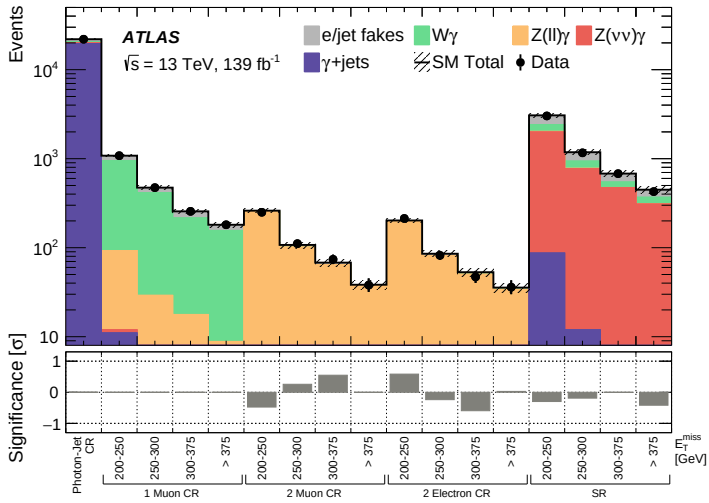


Figure 8.24: Summary of the post-fit event yields in the different regions of the analysis for background predicted with a shape fit to the luminosity of the full Run-2 dataset. The dashed band includes statistical and systematic uncertainties. The lower panel shows the significance of the difference between data and the background prediction.

A single-bin fit is also performed in each inclusive SR. The main results are quickly summarized here for completeness: the k-factors resulting from this fit are reported in Table 8.17, while the data and post-fit background expectations in all CRs and SRs (defined with inclusive $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds) are shown in Figure 8.25.

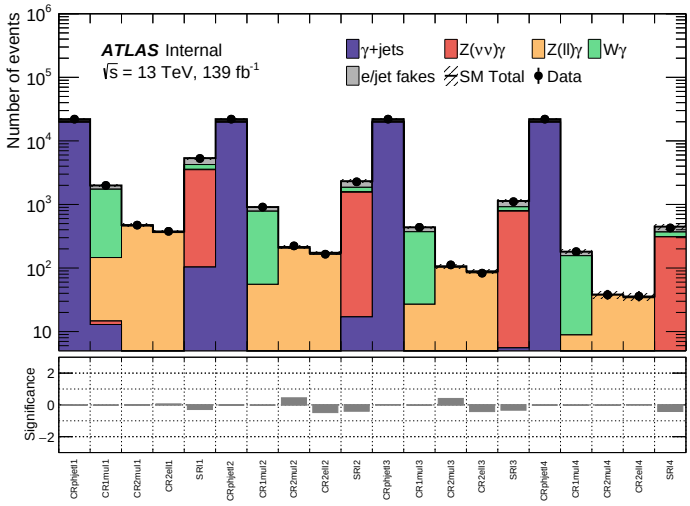


Figure 8.25: Summary of the post-fit event yields in the different regions of the analysis for background predicted with a single-bin fit to the luminosity of the full Run-2 dataset. The dashed band includes statistical and systematic uncertainties. The lower panel shows the significance of the difference between data and the background prediction.

	SRI1	SRI2	SRI3	SRI4
$\mu_{W\gamma}$	0.81 ± 0.08	0.80 ± 0.09	0.78 ± 0.10	0.85 ± 0.13
$\mu_{Z\gamma}$	0.93 ± 0.07	0.88 ± 0.07	0.88 ± 0.09	0.86 ± 0.12
$\mu_{\gamma+\text{jets}}$	0.82 ± 0.21	0.82 ± 0.20	0.81 ± 0.19	0.82 ± 0.21

Table 8.17: Normalization factors obtained from a background-only single-bin fit performed on the inclusive $E_{\text{T}}^{\text{miss}}$ bins defined in Table 8.8. The errors shown include both the statistical and systematic uncertainties.

Fit with a single normalization for $Z\gamma$ and $W\gamma$

Table 8.15 shows that the normalization factors for the $W\gamma$ and for the $Z\gamma$, extracted respectively from the 1μ-CR and the $2\mu/e$-CRs, are different, especially in the low $E_{\text{T}}^{\text{miss}}$ regions, supporting the choice to normalize the two background processes separately, despite the drawback of the potential loss in sensitivity due to higher uncertainties on the normalization factors arising from the separate treatement of the CRs. The effect of applying a single normalizion to both $W\gamma$ and $Z\gamma$, based on the combination of the 1μ-CR and the $2\mu/e$-CR was cross-checked to be relatively small. Table 8.18 summarizes the normalization factors obtained in this new configuration, with their uncertainty, while the post-fit expected backgrounds are shown in Table 8.19, compared with the results obtained in the nominal configuration: the uncertainty on the normalization factor for $W/Z\gamma$ processes is reduced from $\sim 8-15\%$ to $\sim 6-10\%$, but the maximum impact on the global uncertainty is a slight increase from 8% to 10% in the last $E_{\text{T}}^{\text{miss}}$ bin. On the other side, the background modelling results to be worse in the alternative fit configuration, with the expected yields being slightly less consistent with observed data than in the nominal fit with $W\gamma$ and $Z\gamma$ separately fitted.

	SRE1	SRE2	SRE3	SRI4
$\mu_{Z/W\gamma}$	0.93 ± 0.06	0.91 ± 0.07	0.86 ± 0.08	0.90 ± 0.09
$\mu_{\gamma+\text{jets}}$		0.83 ± 0.21		

Table 8.18: Normalization factors obtained from a background-only "shape" fit, a single normalizion to both $W\gamma$ and $Z\gamma$, based on the combination of the 1μ-CR and the $2\mu/e$-CR.

	SRE1	SRE2	SRE3	SRI4
Observed events	3023	1164	679	427
Fit with μ_Z and μ_W	$3067 \pm 153(5\%)$	$1183 \pm 75(6\%)$	$679 \pm 53(8\%)$	$447 \pm 44(10\%)$
Fit with μ_{ZW}	$2975 \pm 119(4\%)$	$1189 \pm 68(6\%)$	$661 \pm 44(7\%)$	$458 \pm 36(8\%)$

Table 8.19: Comparison between the background expectation, after a "shape" fit, in the standard fit configuration with separate normalization for the $W\gamma$ and the $z\gamma$ background processes (first row) and in the alternative configuration with one single normalization factor shared among the two backgrounds, using full Run2 data. The parentheses show the relative error in percent.

8.6.3 Overview of systematic uncertainties

The NPs pull-plots in Figure 8.26 do not show any anomalous pull or constraint: the only slightly pulled NPs are the ones associated to $jet \to \gamma$ background estimation. The correlation matrix among the NPs has been produced as well, and is shown in Figure 8.27.
A summary of the relative uncertainties, grouped in categories of contributions, on the expected SM background, is shown in Table 8.20, for the inclusive regions as resulting from the single-bin background-only fit, and for the exclusive regions as resulting from the "shape" fit. They are determined as the post-fit impact of the nuisance parameters associated to a given group of systematic uncertainties, obtained by repeating the background-only fit setting the considered NPs to $\pm 1\sigma$ (with the other NPs free to float) and evaluating the corresponding variation on the total background yields. The relative uncertainty ranges from approximately 3.5% to 9.8%. The purely statistical uncertainty is dominant, varying from 2.4% to 8.5% with a relevant contribution from the statistical precision of the $2\mu/e$-CRs adopted to constrain the normalisation of the leading $Z(\to \nu\nu)\gamma$ background. Concerning the systematic uncertainties, the main contribution arises from the data-driven estimates of the $jet \to \gamma$ and $e \to \gamma$ background processes, as well as the jet energy scale and resolutions. The breakdown of the single systematics are shown, for each SR bin, in the ranking plots in Appendix A.6.

	SRI1 [%]	**SRI2** [%]	**SRI3** [%]	**SRI4** [%]	**SRE1** [%]	**SRE2** [%]	**SRE3** [%]
Total (stat+syst) uncertainty	3.5	4.8	6.2	9.5	4.3	6.3	7.8
Statistical uncertainty	2.4	3.6	5.3	8.5	3.3	5.0	6.7
Fake photons from jets	1.4	2.5	2.8	4.1	1.4	3.6	3.7
Jet Energy scale/resol	1.6	2.2	2.5	2.7	2.2	2.2	2.3
Fake photons from electrons	2.1	2.0	2.0	2.0	2.3	2.3	2.1
Electrons reco/id/isolation eff.	1.0	1.2	1.3	1.4	1.0	1.0	1.2
Electron/Photon Energy scale/resol.	0.8	0.6	0.7	0.9	0.9	0.9	0.6
Muon reco/isolation eff.	0.7	0.8	0.9	1.0	0.6	0.7	0.9
$E_{\mathrm{T}}^{\mathrm{miss}}$ soft term scale/resolution	0.1	0.4	0.7	0.9	0.5	0.2	0.5
Theoretical $W/Z\gamma$, γ+jets	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$\langle\mu\rangle$ reweighting in MC	0.3	0.2	0.2	0.2	0.3	0.2	0.2

Table 8.20: Summary of the uncertainties (%) in the background estimate for the inclusive SRs after the background-only fit and for the exclusive SRs after the "shape" fit. Due to correlations among the different uncertainty contributions, the uncertainties do not add in quadrature to provide the total uncertainty.

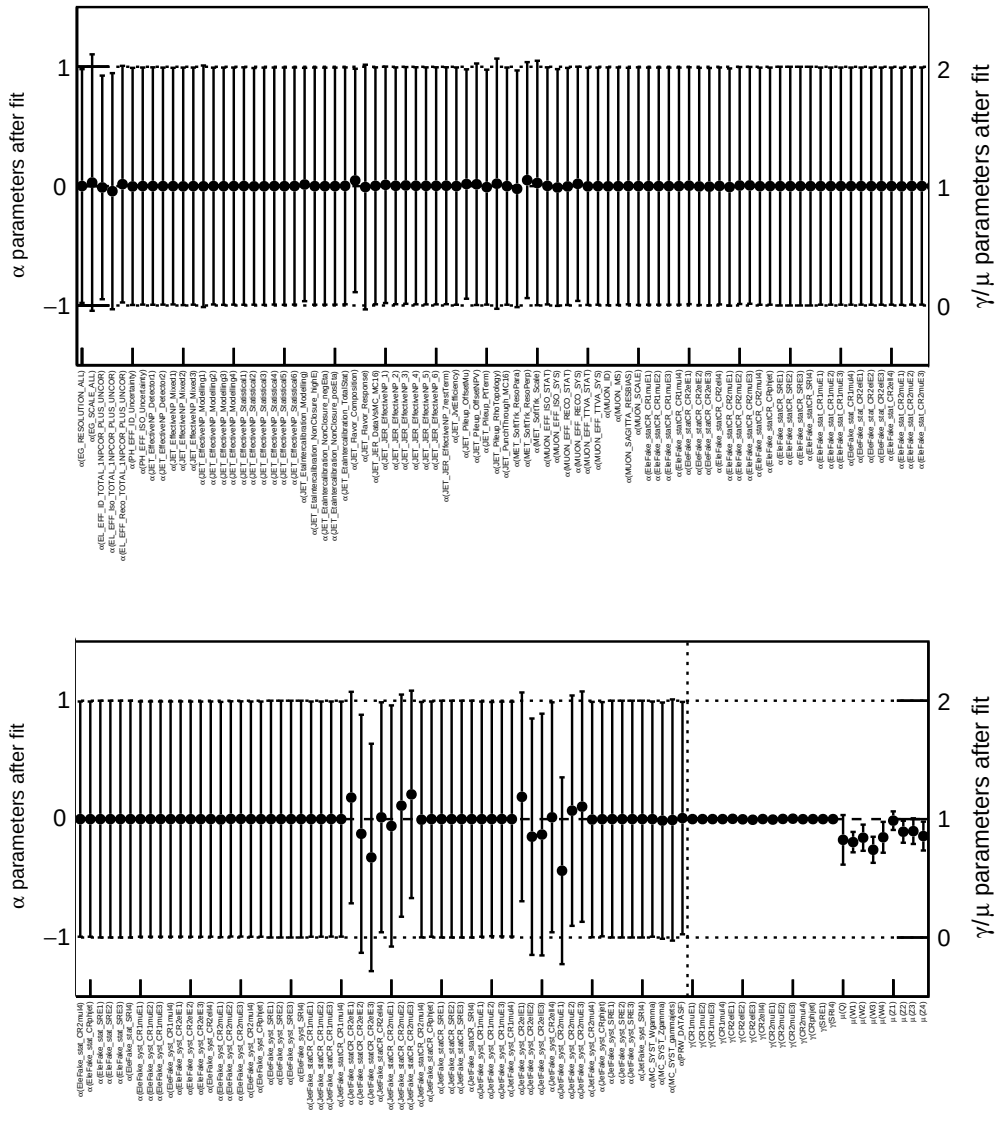


Figure 8.26: Post-fit nuisance parameter values and errors in the shape fit on full Run2 data. The plot is splitted into two rows to make the NPs more readable.

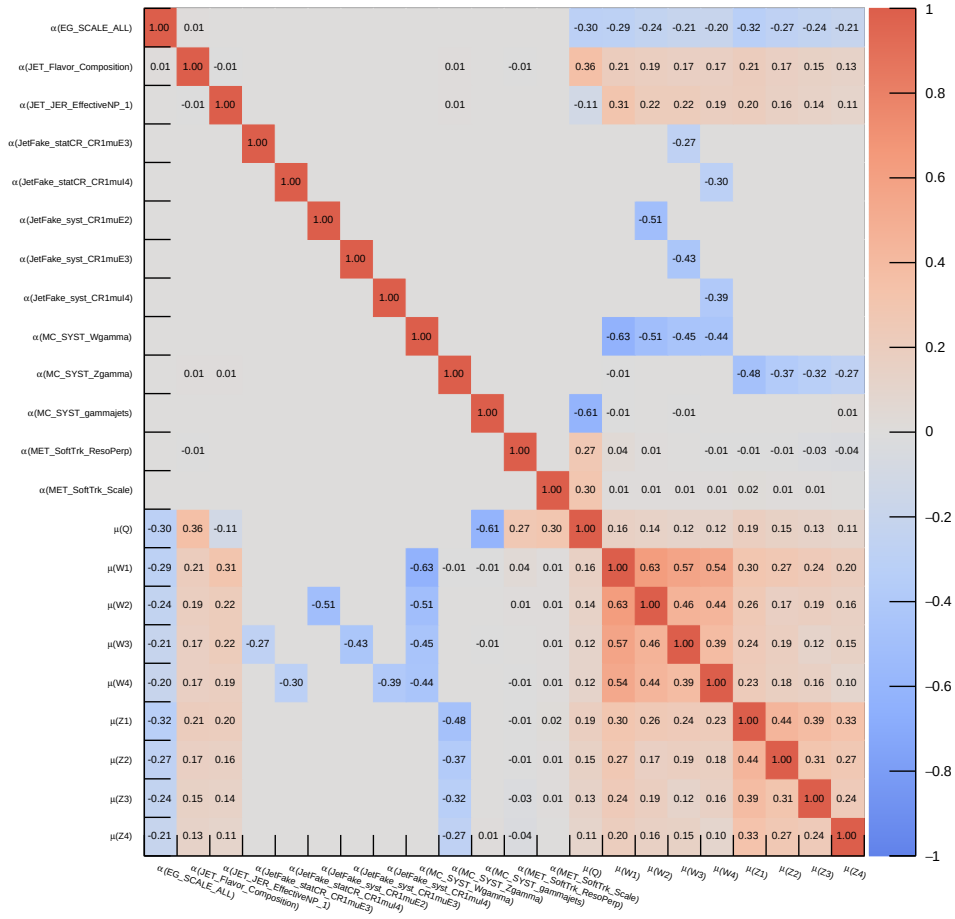


Figure 8.27: Correlation matrix between the NPs of background-only simplified-shape fit. Only NPs with at least one correlation exceeding 10% are shown.

8.7 Interpretations

Since no excess of data with respect to SM expectations is observed in the SR, exclusion limits on new physics can be set, following the methods described in Section 7.3.3. Both *model-independent* and *model-dependent* interpretations are provided.

8.7.1 Model-independent limit

Model-independent exclusion limits on the visible cross-section of new physics, without any assumption on the specific model, are provided in all the SRs, either inclusive or exclusive.

Exclusion limits on the visible cross-section ($\sigma \times A \times \epsilon$)

The first step consists in setting exclusion limits on the number of observable events of new physics. The expected selected events coming from a potential new physics process with a certain cross-section σ is $N_{\text{new}} = L \times \sigma \times A \times \epsilon$, where L is the integrated luminosity, A is the acceptance of the analysis selection criteria and ϵ the reconstruction efficiency, accounting for detector effects. The *visible* cross-section, i.e. the cross-section observable in the analysis SR at ATLAS, is defined as $\sigma \times A \times \epsilon$. In order to set these upper limits, a set of single-bin fits is performed in SR+CRs, enabling a "dummy" signal component included as a single event with a normalization factor included as parameter of interest (POI). A scan of the CLs values for increasing value of the POI is performed, providing upper limits at 95% CL on the number of events from new physics, S^{95}. The expected and observed exclusion limits (in the absence of new physics) are shown in Table 8.21, both in terms of S^{95}, and in terms of the visible cross-section $\sigma \times A \times \epsilon = S^{95}/L$ (identified with $\langle \epsilon\sigma \rangle^{95}$): the expected limits are the ones obtained based on an Asimov dataset [172] built from post-fit background expectations (as resulting from the SR+CRs simultaneous fit), while the observed ones are based on a fit to actual data. The limits in the exclusive bins improve the ones in the inclusive regions and can therefore be advantageous for reinterpretations in terms of any model predicting a signal largely contained within the bin $E_{\text{T}}^{\text{miss}}$ range.

SR	$\langle \epsilon\sigma \rangle^{95}_{obs}$[fb]	$\langle \epsilon\sigma \rangle^{95}_{exp}$[fb]	S^{95}_{obs}	S^{95}_{exp}
$E_{\text{T}}^{\text{miss}} > 200$ GeV	2.45	$2.82^{+1.08}_{-0.78}$	340.0	$391.7^{+149.7}_{-107.9}$
$E_{\text{T}}^{\text{miss}} > 250$ GeV	1.42	$1.68^{+0.63}_{-0.46}$	197.8	$233.7^{+88.1}_{-64.2}$
$E_{\text{T}}^{\text{miss}} > 300$ GeV	0.93	$1.07^{+0.40}_{-0.29}$	128.8	$148.3^{+55.5}_{-40.5}$
$E_{\text{T}}^{\text{miss}} > 375$ GeV	0.53	$0.63^{+0.23}_{-0.17}$	74.20	$87.30^{+32.1}_{-23.5}$
$200 < E_{\text{T}}^{\text{miss}} < 250$ GeV	1.80	$2.03^{+0.77}_{-0.56}$	249.6	$282.4^{+106.6}_{-77.4}$
$250 < E_{\text{T}}^{\text{miss}} < 300$ GeV	1.04	$1.15^{+0.43}_{-0.31}$	144.6	$159.4^{+59.8}_{-43.4}$
$300 < E_{\text{T}}^{\text{miss}} < 375$ GeV	0.79	$0.82^{+0.31}_{-0.22}$	109.4	$113.5^{+42.6}_{-31.0}$

Table 8.21: The observed and expected limit at 95% confidence level on the visible cross-section $\sigma \times A \times \epsilon$ ($\langle \epsilon\sigma \rangle^{95}_{obs}$) and on the number of events (S^{95}) in the four inclusive and three exclusive $E_{\text{T}}^{\text{miss}}$ bins.

Fiducial limit ($\sigma \times A$)

In order to provide useful constraints on the $\sigma \times A$ of new physics, which rule out the dependence on the detector and are therefore more suitable for reinterpretations outside ATLAS, a fiducial signal region is defined at truth-level:

- photon selection: $E_{T,\gamma} > 10$ GeV, $|\eta_\gamma| < 2.37$, not in $1.37 < |\eta_\gamma| < 1.52$
- electron selection: $p_{T,e} > 7$ GeV, $|\eta_e| < 2.47$
- muon selection: $p_{T,\mu} > 6$ GeV, $|\eta_\mu| < 2.5$
- tau selection: $p_{T,\tau} > 20$ GeV, $|\eta_\tau| < 2.5$
- jet selection: $p_{T,jet} > 30$ GeV, $|\eta_{jet}| < 4.5$, not overlapping with electrons or photons within $\Delta R < 0.4$
- $E_{\text{T}}^{\text{miss}} > 200$ GeV, $E_{\text{T}}^{\text{miss}} > 250$ GeV, $E_{\text{T}}^{\text{miss}} > 300$ GeV or $E_{\text{T}}^{\text{miss}} > 375$ GeV, depending on the considered inclusive signal region

- leading photon with $E_T > 150$ GeV, $\Delta\phi(\gamma, E_{\mathrm{T}}^{\mathrm{miss}}) > 0.4$
- No electron, no muon, no hadronic taus
- At most 1 jet with $p_T > 30$ GeV and $\Delta\phi(jet, E_{\mathrm{T}}^{\mathrm{miss}}) > 0.4$.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ significance cut is not applied, to avoid the inconsistency between the object-based definition used at reconstruction-level, and the event-based one ($E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{\sum E_{\mathrm{T}}^{\mathrm{miss}}}$) that would be used at truth-level. The impact of this selection on signal events passing all the other criteria is less than 5% and can therefore be considered negligible.
As already mentioned, the visible cross-section is given by $\sigma \times A \times \epsilon$ where $A \times \epsilon$ is the total signal efficiency ϵ_{tot}, decomposed into fiducial acceptance A and fiducial reconstruction efficiency ϵ.

$$\epsilon_{tot} = \frac{N_{reco}}{N_{total}} = A \times \epsilon, \tag{8.29}$$

$$A = \frac{N_{fiducial}}{N_{total}}, \tag{8.30}$$

$$\epsilon = \frac{N_{reco}}{N_{fiducial}}, \tag{8.31}$$

where N_{total} is the total number of generated events in the truth-level MC sample, $N_{fiducial}$ is the number of events passing truth-level fiducial selection and N_{reco} is the number of events passing signal region selection at reconstruction level. The fiducial efficiency ϵ therefore includes the reconstruction efficiency, thus the difference between truth-level and reconstruction-level selections. Given the ϵ_{tot}, known from reconstruction-level selections, and the fiducial acceptance A obtained from truth-level selections, the fiducial efficiency ϵ can be easily derived. It is therefore possible to easily convert the visible cross section limits ($\sigma \times A \times \epsilon$) into fiducial cross-section limits ($\sigma \times A$), which may be used to set constraints on other signals not considered in this analysis. The fiducial acceptances for some example simplified DM models can be found in Tables A.12. The full list of acceptances is reported in Appendix A.7, together with the efficiencies. For $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 375$ GeV, the efficiency ranges from 76% to 96%, from 74% to 94%, from 72% to 86% and from 67% to 88%, respectively. For the corresponding exclusive bins, it ranges from 75% to 97%, from 75% to 100% and from 71% to 91%. The lowest efficiency for each signal region is used in a conservative way to set the fiducial cross section limit, reported in Table 8.23.

8.7.2 Simplified Dark Matter model

The results of the simplified shape fit are interpreted in the context a simplified DM model described in Section 6.1.

Rescaling procedure

The signal+background fit is performed only for the signal points generated at reconstruction-level in the axial-vector model, providing upper limits on the signal strength μ_S. Since the signal acceptance has been shown to be independent on the DM mass for a fixed value of the mediator mass, a cross-section rescaling can be applied to derive, from these results, also the upper limits for the other signals in the grid, generated at truth-level.
A similar rescaling procedure, following recommendations in Ref. [178], can be applied

m_χ (GeV)	m_{med} (GeV)	A_{SRI1} (%)	A_{SRI2} (%)	A_{SRI3} (%)	A_{SRI4} (%)	A_{SRE1} (%)	A_{SRE2} (%)	A_{SRE3} (%)
10	10	13.90	7.00	3.70	1.60	6.90	3.30	2.10
10	100	16.00	7.50	3.80	1.70	8.40	3.70	2.10
10	500	28.50	17.10	10.50	5.60	11.50	6.60	4.90
10	1000	35.60	23.80	15.80	9.10	11.90	8.00	6.70
10	1500	37.60	26.40	19.40	12.20	11.20	7.00	7.20
50	10	22.90	12.90	7.40	3.40	10.00	5.50	4.00
50	115	16.00	8.10	4.20	1.90	8.00	3.80	2.30
100	10	26.90	16.20	9.80	5.20	10.70	6.40	4.60
100	215	21.50	11.10	6.20	2.70	10.40	4.90	3.60
150	200	29.50	17.90	11.20	5.70	11.60	6.70	5.50
150	315	25.80	14.20	8.50	4.10	11.70	5.70	4.40
200	395	29.20	17.80	11.30	5.90	11.40	6.40	5.40
200	415	28.70	16.80	10.50	5.10	11.90	6.30	5.40
450	915	36.00	24.70	16.70	9.70	11.30	8.00	7.00

Table 8.22: Fiducial acceptances for a subset of signal mass points in the simplified DM model, in the four inclusive and three exclusive signal regions.

Signal channel	$\langle\sigma\rangle^{95}_{obs}$[fb]	$\langle\sigma\rangle^{95}_{exp}$[fb]
$E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV	3.24	$3.73^{+1.42}_{-1.03}$
$E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV	1.93	$2.28^{+0.86}_{-0.63}$
$E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV	1.28	$1.47^{+0.55}_{-0.40}$
$E_{\mathrm{T}}^{\mathrm{miss}} > 375$ GeV	0.80	$0.94^{+0.35}_{-0.25}$
$200 < E_{\mathrm{T}}^{\mathrm{miss}} < 250$ GeV	2.39	$2.71^{+1.02}_{-0.74}$
$250 < E_{\mathrm{T}}^{\mathrm{miss}} < 300$ GeV	1.39	$1.53^{+0.57}_{-0.42}$
$300 < E_{\mathrm{T}}^{\mathrm{miss}} < 375$ GeV	1.10	$1.14^{+0.43}_{-0.31}$

Table 8.23: The observed and expected limit at 95% confidence level on the fiducial cross-section $\sigma \times A$ ($\langle\sigma\rangle^{95}_{obs}$) in the four inclusive and three exclusive $E_{\mathrm{T}}^{\mathrm{miss}}$ bins. The expected limits at $\pm 1\sigma$ are also reported ($\langle\sigma\rangle^{95}_{exp}$).

also for reinterpretations in terms of a vector mediator model or with different couplings, because the choice of mediator and couplings only affects the cross section and not the signal acceptance. In the on-shell region an analytical approximation of the cross-section can be applied:

$$\sigma_{AV/V} \propto \frac{\Gamma^{q\bar{q}}_{AV/V}\Gamma^{\chi\bar{\chi}}_{AV/V}}{\Gamma^{\chi\bar{\chi}}_{AV/V} + \Gamma^{q\bar{q}}_{AV/V} + \Gamma^{l\bar{l}}_{AV/V}} \tag{8.32}$$

where the widths for the axial-vector and vector models are given by:

$$\Gamma^{XX}_{AV} = \frac{N_C^X g_X^2 M_{med}}{12\pi}\beta_\chi^3 \theta(M_{med} - 2m_X) \tag{8.33}$$

$$\Gamma^{XX}_{V} = \frac{N_C^X g_X^2 M_{med}}{12\pi}\left(1 + \frac{2m_X^2}{M_{med}^2}\right)\beta_X \theta(M_{med} - 2m_X) \tag{8.34}$$

with X labelling either the χ particle or a quark or lepton, N_C^X the number of color charges for the given particle (i.e. 3 for the quarks and 1 otherwise), $\beta = \sqrt{1 - 4\frac{m_\chi^2}{M_{med}^2}}$ and θ the Heaviside function.

When reaching the off-shell region, this approximation is no more valid. Therefore, the cross-section for the off-shell and transition region is taken from the truth-level generation of the diagonal signal points in the grid for each choice of the mediator and couplings.

Results

A simplified shape fit corresponding to the 4-bin configuration: 200-250-300-375 GeV is performed for the simplified model with an axial-vector mediator and couplings $g_q = 0.25$, g_χ =1 and g_ℓ=0.0. Figure 8.28 shows the observed and expected contours corresponding to a 95% CL exclusion as a function of m_{med} and m_χ for this initial scenario and for other scenarios obtained through the rescaling procedure. The limit curve is given by the interpolation of mass points corresponding to an upper limit $\mu_{sig}^{limit} = 1$. The region of the plane under the limit curves is excluded, while the region not allowed due to perturbative unitarity violation is the one at the left of the line defined by $m_\chi = \sqrt{\pi/2} m_{\text{med}}$. The line corresponding to the DM thermal relic abundance measured by the Planck collaboration [41] is also indicated and calculated following prescriptions in [178]. In Figure 8.29, the grid of the upper limits on the signal strength and on the cross-section of new physics (obtained by multiplying the limit on the signal strength for each signal by its theoretical cross-section) are shown, for the different signals, and for the nominal scenario with leptophobic Axial-Vector mediator. $\sigma_{theo} = \mu_{sig}^{limit} \times \sigma_{vis}$ (obtained by multiplying the upper limits on signal strength).

Comparison with Direct Detection experiments

To highlight the complementarity with DM Direct Detection (DD) searches (briefly described in Section 2.3.1), the contours in the m_χ-m_{med} plane obtained for a specific choice of mediator and couplings can be directly translated into bounds on the χ-nucleon scattering cross-sections. Three different results are presented by DD experiments, depending on the scattering spin dependence and on the nucleon against which the hypothetical DM recoils: for Spin Independent (SI) interactions, the scattering cross-section with protons and neutrons are assumed to be equal ($\sigma_{SI}^p = \sigma_{SI}^n$), thus providing a single result, while Spin Dependent (SD) results are separately shown for protons (σ_{SD}^p) and neutrons (σ_{SD}^n) as recoiling nucleons.
Concerning simplified DM models, different scattering interactions can happen depending on the nature of the mediator. For axial-vector mediators (and pseudo-scalar ones), SD scattering are induced, while vector (or scalar) mediators give origin to SI interactions.
The translation is performed following the procedure described in Ref. [178]. For SI and vector mediators, the DM-nucleon scattering cross-section can be expressed in the form:

$$\sigma_{SI} = \frac{f^2(g_q) g_\chi^2 \mu_{n\chi}^2}{\pi M_{med}^4} \tag{8.35}$$

where $\mu_{n\chi}^2 = m_n m_\chi / (m_n + m_\chi)$ is the DM-nucleon reduced mass with $m_n \simeq 0.939$ GeV, g_χ is the mediator-DM coupling, and $f(g_q)$ is the mediator coupling to the nucleon. The dependence of the latter on the mediator-quark coupling is the main ingredient of the

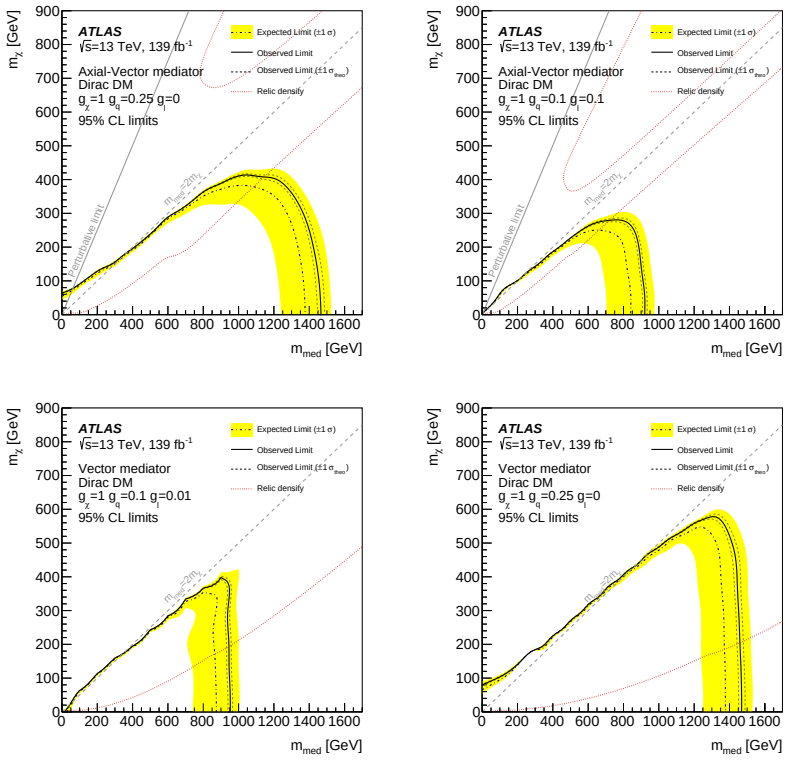


Figure 8.28: The observed (solid line) and expected (dot-dashed line) 95% CL exclusion contours on the m_χ–m_{med} plane for a simplified DM model involving an axial-vector mediator with couplings $g_q = 0.25$, $g_\chi = 1$ and $g_\ell = 0$. The area under the limit curve is excluded. The region on the left of the line defined by $m_\chi = \sqrt{\pi/2} m_{\mathrm{med}}$ is excluded by the perturbative limit which is relevant for axial-vector mediators. The relic density curve is also shown.

translation procedure itself: for the vector mediator case, this term assumes the simple form $f(g_q) = 3g_q$. It is therefore possible to express σ_{SI} in terms of the simplified DM model parameters:

$$\sigma_{SI} \simeq 6.9 \times 10^{-41}\ \mathrm{cm}^2 \cdot \left(\frac{g_q g_\chi}{0.25}\right)^2 \left(\frac{1\ \mathrm{TeV}}{M_{med}}\right)^4 \left(\frac{\mu_{n\chi}}{1\ \mathrm{GeV}}\right)^2 \tag{8.36}$$

Similarly, for the SD and axial-vector mediator case, the scattering cross-section is

$$\sigma_{SI} = \frac{3 f^2(g_q) g_\chi^2 \mu_{n\chi}^2}{\pi M_{med}^4} \tag{8.37}$$

with $f(g_q) = 0.32 g_q$ under the assumption the the g_q coupling is the same for all quarks, thus leading to:

$$\sigma_{SD} \simeq 2.4 \times 10^{-42}\ \mathrm{cm}^2 \left(\frac{g_q g_\chi}{0.25}\right)^2 \left(\frac{1\ \mathrm{TeV}}{M_{med}}\right)^4 \left(\frac{\mu_{n\chi}}{1\ \mathrm{GeV}}\right)^2 \tag{8.38}$$

Since DD experiments provide their results at 90% CL, exclusion limits at the same CL are provided also in the mono-photon analysis, for a coherent comparison. The top plots of Figure 8.30 show the 90% CL exclusion limits on the χ–proton and χ–neutron spin-dependent (SD) scattering cross-sections versus m_χ in the axial-vector model with couplings $g_q = 0.25$, $g_\chi = 1$ and $g_\ell = 0$. The bottom plot of the same figure, shows the 90% CL exclusion limits on the χ–nucleon spin-independent (SI) scattering cross-section versus m_χ for the vector model with couplings $g_q = 0.25$, $g_\chi = 1$ and $g_\ell = 0$. In this case, direct detection searches provide generally stronger limits, but the mono-photon results are still competitive for $m_\chi < 2$ GeV.

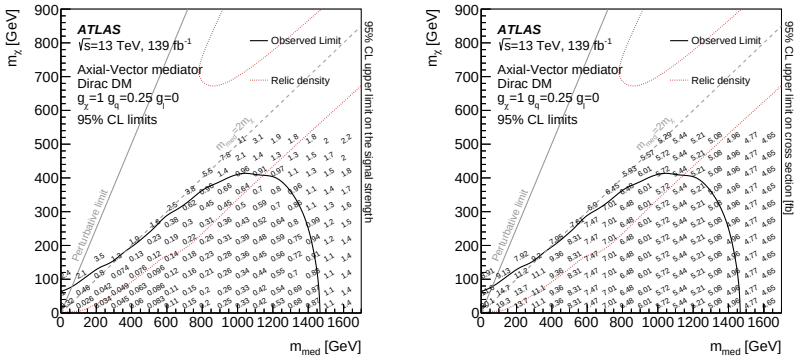


Figure 8.29: Upper limits on the signal strength (left) and on the cross-section (right) for the different signal points in the m_χ-m_{med} grid. The observed exclusion limit at 95% is also shown, as well as the perturbative limit and the relic density curve.

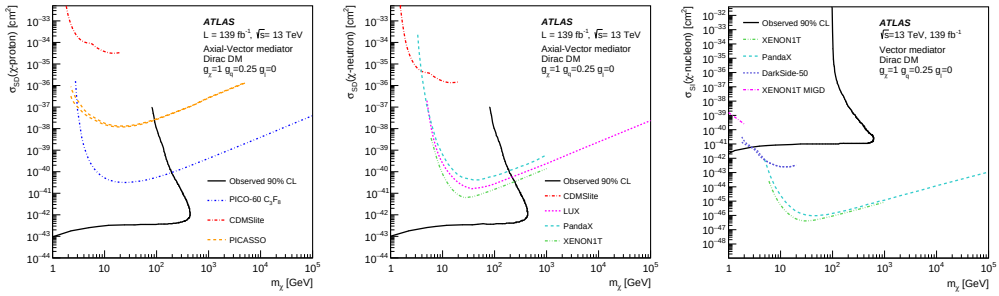


Figure 8.30: The 90% CL exclusion limit, as a function of the dark-matter mass m_χ on: (top left) χ–proton spin-dependent scattering cross section in an axial-vector model with couplings $g_\chi = 1$, $g_q = 0.25$ and $g_\ell = 0$, compared to direct DM searches [187, 188, 189]; (top right) χ–neutron spin-dependent scattering cross section in an axial-vector model with couplings $g_\chi = 1$, $g_q = 0.25$ and $g_\ell = 0$, compared to direct DM searches [190, 188, 191, 192]; (bottom) χ–nucleon spin-independent scattering cross section in a vector model with couplings $g_\chi = 1$, $g_q = 0.25$ and $g_\ell = 0$, compared to direct DM searches [193, 194, 195, 196]

Chapter 9

Search for Dark Photon from Higgs boson decay, in ZH production model

The observation of a Dark Photon (γ_d) would be an hint of the existence of the Dark Sector, thus opening to several scenarios towards the explanation of the nature of DM. In addition to being a possible way to explore the Dark Sector, strong motivations for the search of this particle arises from its potential contribution in solving the small-scale structure formation problems [197, 198], as well as being a possible source for the Sommerfeld enhancement of DM annihilation required to explain the PAMELA-Fermi-MS2 positron anomaly in terms of DM [199], or making asymmetric DM scenarios viable [200]. Going beyond DM related possible consequences of the existence of a Dark Photon, this new particle would also be part of a new paradigm to solve the hyerarchy problem by generating exponentially spread SM Yukawa coupling [146] from unbroken $\mathrm{U}(1)_F$ quantum numbers in the Dark Sector, as mentioned in section 6.2.1.
The analysis described in this Chapter [201] is a search for a Dark Photon produced from the Higgs boson decay into a photon and a Dark Photon, in the associated Higgs boson production mode (ZH), as shown in the Feynman diagram in Figure 9.1. The final state is therefore composed of two opposite sign and same flavour leptons from the Z boson decay, one photon and missing transverse momentum (signature of the "invisible" γ_d).
The analysis has been performed on the full Run-2 dataset, collected by the ATLAS experiment between 2015 and 2018.

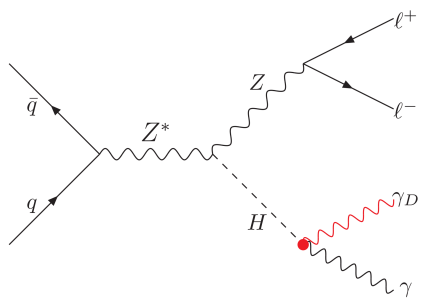


Figure 9.1: Feynman diagram for the $H \to \gamma\gamma_d$ decay via ZH production mode.

9.1 Data and MC simulations

9.1.1 Data

Similarly to the mono-photon analysis, the pre-selections described in section 7.1.1 are applied to both data and MC simulations. A high efficiency unprescaled trigger is employed, selecting single- or di-lepton events, i.e. events with at least one o two electrons or muons [202, 203] with p_{T} thresholds defined based on the data-taking conditions during different periods [204]. In particular, single-lepton triggers with low p_{T} threshold and lepton isolation requirements are combined in a logical OR with higher-threshold triggers without isolation requirements. In addition, di-lepton triggers complement the single lepton trigger, requiring two leptons satisfying loose identification criteria, with low p_{T} thresholds for electrons and muons (including asymmetric thresholds in the muon case). The addition of the di-lepton triggers allows to enhance the signal efficiency by about 3% - 3.6%, depending on the Dark Photon mass points for combined $Z \to ee$ and $Z \to \mu\mu$ final states. The trigger definitions are summarized in Table 9.1.

Period	Electron	Muon
Single-lepton		
2015	$p_{\mathrm{T}} > 24$, 60, 120 GeV	$p_{\mathrm{T}} > 20$, 50 GeV
2016-2018	$p_{\mathrm{T}} > 26$, 60, 140 GeV	$p_{\mathrm{T}} > 26$, 50 GeV
Di-lepton		
2015	$p_T^1, p_T^2 > 12$ GeV	$p_T^1 > 18$ GeV, $p_T^2 > 8$ GeV $p_T^1, p_T^2 > 10$ GeV
2016, 2018	$p_T^1, p_T^2 > 17$ GeV	$p_T^1 > 22$ GeV, $p_T^2 > 8$ GeV $p_T^1, p_T^2 > 14$ GeV
2017, 2018	$p_T^1, p_T^2 > 24$ GeV	$p_T^1 > 22$ GeV, $p_T^2 > 8$ GeV $p_T^1, p_T^2 > 14$ GeV

Table 9.1: List of single-lepton and di-lepton triggers.

9.1.2 Monte Carlo simulations

The configuration for the Monte Carlo (MC) simulated samples for signals and backgrounds are summarized in Table 10.1.
The simulation of the particle interactions with the ATLAS detector is performed mostly using the GEANT [159] generator, except for the signals with massive Dark Photon and the $tW\gamma$ background process, for which fast parameterized simulation of the ATLAS calorimeters [161] was used. The pile-up was modelled from minimum-bias interactions generated with PYTHIA8.186 [163] and with the A3 parameters tuning [205] and overlaid onto the simulated hard-scatter events according to the luminosity profile of the recorded data. All simulated samples, except those produced with the SHERPA 2.2.1 [182] event generator, used `EvtGen` 1.2.0 [206] to model the decays of heavy-flavour hadrons.

9.1.3 Signal

The signal Model

A "Hidden Valley" scenario is explored [207]: it consists in a model featuring a new light hidden sector, decoupled from the SM sector at relatively low energy scales (of the order

Process	Generator	ME Order	PDF	Parton Shower	Tune
Signal samples					
$ZH, H \to \gamma\gamma_d$	Powheg-Box [v2]	NLO	NNPDF[3.0nlo]	Pythia [8.245]	AZNLO
SM background samples					
$V\gamma^{QCD}$	Sherpa v2.2.8	NLO (up to 2 jets), LO (up to 3 jets)	NNPDF[3.0nnlo]	Sherpa MEPS@NLO	Sherpa
$V\gamma^{EWK}$	MadGraph5_aMC@NLO [v2.6.5]	LO	NNPDF[2.3lo]	Pythia [8.240]	A14
Z^{QCD}	Sherpa v2.2.1	NLO (up to 2 jets), LO (up to 4 jets)	NNPDF[3.0nnlo]	Sherpa MEPS@NLO	Sherpa
Z^{EWK}	Sherpa v2.2.1	NLO (up to 2 jets), LO (up to 4 jets)	NNPDF[3.0nnlo]	Sherpa MEPS@NLO	Sherpa
Single t-quark/$t\bar{t}$	Powheg-Box [v2]	NLO	NNPDF[2.3lo]	Pythia [8.230]	A14
$t\bar{t}$, (V, VV), $Wt\gamma$	MadGraph5_aMC@NLO [v2.2.3]	NLO	NNPDF[2.3lo]	Pythia [8.210]	A14
SM Higgs boson	Powheg-Box [v2]	NNLO	PDF4LHC15	Pythia [8.230]	AZNLO
$VV\gamma$	Sherpa v2.2.11	NLO (0 jets), LO (up to 3 jets)	NNPDF[3.0nnlo]	Sherpa MEPS@NLO	Sherpa
VV/VVV	Sherpa v2.2.2	NLO (0 jets), LO (up to 3 jets)	NNPDF[3.0nnlo]	Sherpa MEPS@NLO	Sherpa

Table 9.2: The configurations used for event generation of signal and background processes. V refers to an electroweak boson (W or Z/γ^*). The matrix element (ME) order refers to the order in the strong coupling constant of the perturbative calculation. The tune refers to the underlying-event tune of the parton shower.

of the TeV scale) to evade current experimental limits. The valley particles are assumed to be charged under a given valley group (an $U(1)$ group in the Dark Photon case), but neutral under SM gauge groups. On the other hand, SM particles are required to be neutral under the valley group. A messenger field is therefore required, coupling both to the SM and the hidden sector. This kind of models represent a good generalization of more specific UV complete models, thus providing a certain degree of model-independence to the search.

MC production

For the signal process $H \to \gamma\gamma_d$, the ZH Higgs boson production mode is considered, including both the contributions from qq and gg initial states, for a global production cross-section of 0.884 pb [208]. The Higgs boson decay into a photon and a Dark Photon, $H \to \gamma\gamma_d$ is simulated in the Hidden Valley scenario [207], as implemented in Pythia8. The matrix elements are estimated using Powhegv2 [209, 210] with the NNPDF3.0 PDF libraries [179]. The $q\bar{q} \to ZH$ and $gg \to ZH$ events are generated at NLO and LO respectively, while the Higgs boson decay and the hadronization and showering are performed with Pythia8.245 [163] and CTEQ6L1.AZNLO tuning [211]. The SM Higgs boson is considered, with mass of 125 GeV, width set to the SM one equal to 4 MeV [208] and the complex pole scheme [212] turned off.
Different Dark Photon masses are considered, for a total of six Monte Carlo samples with masses equal to 0, 1, 10, 20, 30, 40 GeV. The mass range is chosen to provide similar kinematics among different signals, thus ensuring the analysis optimality in the full range.

9.1.4 Background

The analysis is dominated by reducible background, while the irreducible one from $VV\gamma$ final states (V being any of W, Z) with both V bosons decaying leptonically, is overall subdominant. Monte Carlo simulations for all processes potentially entering the analysis selection were performed, including electrons-faking-photons backgrounds for preliminary studies, despite them being ultimately estimated through a purely data-driven method.

The analysis background

The main background processes are listed here, according to their categorization adopted in the statistical analysis:

- $VV\gamma$: mainly constituted by the irreducible background from $Z(\to \ell\ell)Z(\to \nu\nu)\gamma$ events, with some contribution from $W^+(\to \ell^+\nu)W^-(\to \ell^-\bar{\nu})\gamma$ when the W bosons decay into same flavour and opposite sign leptons, casually falling in the Z mass window;
- Electrons-faking-photons ($e \to \gamma$): dominated by $Z(\to \ell\ell)W(\to e\nu)$. Other events can give contribution to this background due to undetected leptons, $Z(\to \ell\ell)Z(\to ee)$ and VVV with $V = Z, W$, or due to failed b-tagging in $\ell\ell Vt$ and $VVt\bar{t}$;
- Fake $E_{\mathrm{T}}^{\mathrm{miss}}$: arising mostly from jet energy mismeasurement and dominated by $Z\gamma$ and $Z+$jets production ($Z\gamma+$jets and $Z+$jets. In the latter, the reconstructed photon in the final state is not a real photon, but rather a jet mistakenly identified as a photon ($j \to \gamma$). Both electroweak and strong components are considered, differing by the electroweak coupling constant order, equal to α_{EWK}^3 for the strong component, and α_{EWK}^5 for the electroweak one; Additionaly, $Z\gamma\gamma$ and $ZH(\to \gamma\gamma)$ with an undetected photon, and ggF or VBF Higgs boson with $H \to Z\gamma$ give a further negligible contribution.
- Background from top quark production: including additional events with top quarks in the final states and with a genuine photon, $tW\gamma$ and $t\bar{t}\gamma$, or a photon from jet misidentification, $t\bar{t}$ and single top-quark production;
- $W\gamma$: subdominant background characterized by a jet faking a lepton;
- Other background from Higgs boson production: category collecting other extremely subdominant (less than 0.1% contribution) background processes from Higgs boson decays into $Z\gamma$: $t\bar{t}H(\to Z\gamma)$ events failing b-veto, the irreducible $Z(\to \nu\nu)H(\to Z\gamma)$ and $W(\to \ell\nu)H(\to Z\gamma)$ with an undetected lepton.

MC simulation

The $Z\gamma+$jets and $Z+$jets processes, as well as the Z electroweak, $VV\gamma$, VV and VVV ones, were generated using the Sherpa generator, at NLO, with NNPDF30_nnlo_as_0118 libraries and tuning developed by the Sherpa authors [179] for the Parton Shower. The MadGraph generator [156] was used for the $Z\gamma$ electroweak process, simulated at LO, and for the NLO simulation of the $t\bar{t}\gamma$ and the $Wt\gamma$ backgrounds, while single-top, $t\bar{t}$ and the Higgs boson backgrounds were produced using Powheg-Box [209], the first ones at NLO, the latter at NNLO. Both the MadGraph and Powheg-Box generators were coupled to Pythia [163] for the parton shower simulation and the hadronization, mostly with NNPDF2.3LO libraries [213] and the A14 tune [205], except for the SM

Higgs boson backgrounds for which the PDF4LHC15 libraries with AZNLO tune were employed. Additional details about the background simulation is given in Table 10.1.

9.2 Event selection

9.2.1 Physics objects selection and overlap removal

The physics objects selection is summarized in Table 9.3, dividing into "baseline" and "selected" objects as described also for the mono-photon analysis in Section 8.2. To solve ambiguities, the same Overlap Removal procedure as in Table 7.1 is applied.

Object	**Feature**	**"Baseline"**	**"Selected"**
Photons	p_{T}	> 10 GeV	$p_{\mathrm{T}} > 20$ GeV
	$\lvert\eta\rvert$	< 2.37	< 2.37
	Identification WP (sec. 4.2.4)	*Loose*	*Tight*
	Isolation WP (sec. 4.2.4)	-	*FixedCutTight*
Electrons	p_{T}	> 10 GeV	> 20 GeV
	$\lvert\eta\rvert$	< 2.47	$\lvert\eta\rvert < 2.47$
	Track-to-vertex association	$\lvert z_0\rvert \sin\theta < 0.5$ mm	$\lvert z_0\rvert \sin\theta < 0.5$ mm $\lvert d_0\rvert/\sigma_{d_0} < 5.0$
	Identification WP (sec. 4.2.5)	*LooseBLayer*	*MediumLH*
	Isolation WP (sec. 4.2.6)	-	*FixedCutLoose*
Muons	p_{T}	> 7 GeV	> 7 GeV
	$\lvert\eta\rvert$	< 2.5	< 2.5
	Track-to-vertex association	$\lvert z_0\rvert \sin\theta < 0.5$ mm	$\lvert z_0\rvert \sin\theta < 0.5$ mm $\lvert d_0\rvert/\sigma_{d_0} < 3.0$
	Identification WP (sec. 4.3.1)	*Medium*	*Medium*
	Isolation WP (sec. 4.3.2)	-	*FCLoose*
Jets	Collection	AntiKt4EMPFlowJets (PFlow)	
	p_{T}	> 30 GeV	
	$\lvert\eta\rvert$	< 4.5	
	JVT WP (tab. 4.6)	Default	
	Jet Cleaning (sec. 4.4.6)	LooseBad	
$E_{\mathrm{T}}^{\mathrm{miss}}$	WP (tab. 4.8)	Tight (PFlow $E_{\mathrm{T}}^{\mathrm{miss}}$)	
	Soft Term	Track Soft Term (TST)	

Table 9.3: Selections criteria and isolation/identification WPs for different physics objects

9.2.2 Signal Region definition

The signal signature is expected to be characterized by sizeable $E_{\mathrm{T}}^{\mathrm{miss}}$ recoiling against the $\gamma\ell\ell$ system, with the two leptons being same flavour and opposite sign, and having invariant mass consistent with the Z boson one. The Signal Region (SR) selection criteria are optimized by means of the AHOI (A Horrible Optimisation Instrument) tool [214], which performs a scan of rectangular cuts testing all possible combination of cuts on a given set of variables, based on a chosen metric. In particular, the AMS metric was adopted, $AMS = \sqrt{2[(s+b)\ln(1+s/b) - s]}$. A first optimization step was based on 9 input variables and 5 cut values each. The optimization was then refined progressively reducing in two further steps the number of variables and increasing the cut values: 7

variables and 10 cut values, and finally 4 variables and 100 cut values.
The final SR of the analysis, detailed in Table 9.4, is the results of this optimization and other additional selections introduced based on the expected final state: one photon and two opposite sign and same flavour leptons are required, with a di-lepton invariant mass consistent with the Z boson mass (in an AHOI optimized window). The lepton p_{T} thresholds are chosen to ensure that the trigger is fully efficient. At most 2 jets are accepted, and b-tagged jets are vetoed, in order to suppress background contributions from processes involving top quarks. In addition, a selection criterion on the invariant mass of the leptons and photon system, $m_{\ell\ell\gamma} > 100$ GeV is optimized through the AHOI procedure and is particularly effective in the rejection of the background from Drell-Yann processes in $Z^*\gamma$ events. The threshold on $E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV is defined by AHOI scan as well. Finally, a high separation between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the leptons and photon system is ensured by the $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}, \vec{p}_{\mathrm{T}}^{\,\ell\ell\gamma})$, originally included as part of the ABCD method definition detailed in section 9.3.1. The background yields, according to pure MC simulations, and

Two same flavour, opposite sign, medium ID and loose isolated leptons, with leading $p_{\mathrm{T}} > 27$ GeV, sub-leading $p_{\mathrm{T}} > 20$ GeV
Veto events with additional lepton(s) with loose ID and $p_{\mathrm{T}} > 10$ GeV
76 GeV $< m_{\ell\ell} <$ 116 GeV
Only one tight ID, tight isolated photon with $E_{\mathrm{T}}^{\gamma} > 25$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV with $\Delta\phi(\vec{p}_T^{\,miss}, \vec{p}_T^{\,\ell\ell\gamma}) > 2.4$ rad
$m_{\ell\ell\gamma} > 100$ GeV
$\mathrm{N}_{\mathrm{jet}} \leq 2$, with $p_{\mathrm{T}}^{\mathrm{jet}} > 30$ GeV, $\lvert\eta\rvert < 4.5$
Veto events with b-jet(s)

Table 9.4: Optimised kinematic selections defining the signal region for $\ell\ell{+}\gamma{+}E_{\mathrm{T}}^{\mathrm{miss}}$.

the signal yields for the massless Dark Photon case after SR selections are reported in Table 9.5. The event yields and acceptance after each cut of the SR, for the signal with massless dark photon, are reported in Table 9.6.

Channel	Signal	$Z\gamma$	Z+jets	Top	$t\bar{t}$	Topγ	$VV\gamma$	VV	SM Higgs boson	$W\gamma$	Total background
ee	19.3 ± 0.2	155 ± 15	274 ± 55	3.5 ± 0.7	25 ± 1	1.9 ± 0.1	26 ± 1	27 ± 1	0.41 ± 0.01	3.5 ± 1.5	517 ± 57
$\mu\mu$	22.4 ± 0.2	283 ± 18	380 ± 63	4.6 ± 0.8	26 ± 1	2.4 ± 0.1	35 ± 1	24 ± 1	0.54 ± 0.01	1.6 ± 1.1	758 ± 66

Table 9.5: Expected event yields for signal and background in the SR corresponding to $\mathcal{L} = 139$ fb^{-1}. Signal events are for massless γ_d, assuming BR($H \to \gamma\gamma_d$) = 5%. Events for background processes are categorised as $Z\gamma$ (QCD+EWK $Z\gamma$), Z+jets (QCD+EWK Z+jets), Top (single top-quark, Wt), $t\bar{t}$ ($t\bar{t}, t\bar{t}V, t\bar{t}VV$), Top$\gamma$ ($Wt\gamma$), $VV\gamma$ ($WW\gamma, WZ\gamma, ZZ\gamma$), VV (WW, WZ, ZZ), SM Higgs boson (ggH, VH, VBF H) and $W\gamma$ (QCD+EWK $W\gamma$). Only statistical uncertainties on the simulated samples are shown.

Boosted Decision Tree As highlighted in Table 9.5, the SR is still strongly dominated by backgrounds over the signal process. The sensitivity of this analysis is indeed driven by the employment of Machine Learning techniques, rather than the simple cut and count method itself. In particular, the Boosted Decision Tree (BDT) algorithm is used, with the XGBoost classifier [215], to gain an optimal discrimination between the signal and the different background processes.

Channel	***ee***		***μμ***	
Selection	**Yields**	**$\epsilon \times A$ [%]**	**Yields**	**$\epsilon \times A$ [%]**
Filter efficiency	396	100	396	100
Preselections	101.1 ± 0.5	25.5	101.1 ± 0.5	25.5
Trigger	100.7 ± 0.5	25.4	100.7 ± 0.5	25.4
ee or $\mu\mu$ channel	38.0 ± 0.3	9.6	45.3 ± 0.3	11.4
Veto extra photons	37.9 ± 0.3	9.6	45.2 ± 0.3	11.4
Veto extra leptons	37.8 ± 0.3	9.5	45.2 ± 0.3	11.4
Leading lepton p_T	37.8 ± 0.3	9.5	45.2 ± 0.3	11.4
Photon p_T	36.8 ± 0.3	9.3	44.1 ± 0.3	11.1
$E_{\mathrm{T}}^{\mathrm{miss}}$	22.6 ± 0.2	5.7	26.5 ± 0.3	6.7
m_{ll}	21.7 ± 0.2	5.5	25.2 ± 0.2	6.3
$m_{ll\gamma}$	21.7 ± 0.2	5.5	25.1 ± 0.2	6.3
$\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \vec{p}_{\mathrm{T}}^{\ell\ell\gamma})$	19.3 ± 0.2	4.9	22.4 ± 0.2	5.7

Table 9.6: Yields and $A \times \epsilon$ after each cut of the SR for 139 fb^{-1} for a $ZH, H \to \gamma\gamma_d$ signal with massless dar photon. Generator-level filters are taken into account in the determination of the total number of events. Only the statistical uncertainty is shown.

The training of the BDT is performed using MC events satisfying the SR selection criteria, with a 5-fold cross-validation to avoid over-fitting, consisting in splitting all samples in 5 subsets and associating the BDT score to each of the 5 sets, based on a training performed on the remaining 4. The same subdivision is then propagated also to data. In addition, the two-samples Kolmogorov-Smirnov (KS) test [216] is implemented, against over-training. Several input variables were tested, observing a good stability of the BDT with respect to the inputs. The final input variables, listed in order of importance (as shown in top right plot of Figure 9.2), are: $E_{\mathrm{T}}^{\mathrm{miss}}$ significance, $m_{\mathrm{T}}(\gamma, E_{\mathrm{T}}^{\mathrm{miss}})$, $m_{\ell\ell}$, p_{T}^{γ}, $m_{\ell\ell\gamma}$ and $Abs_{p_{\mathrm{T}}} = \left| \frac{|\vec{E}_{\mathrm{T}}^{\mathrm{miss}} + \vec{p}_{\mathrm{T}}^{\gamma}| - \vec{p}_{\mathrm{T}}^{\ell\ell}}{p_{\mathrm{T}}^{\ell\ell}} \right|$, the latter exploiting the expected balance in signal events.
The hyperparameter tuning for the BDT training is based on a combination of scikit-learn's RandomizedSearchCV, which randomly scans the hyper-parameter space by sampling a fixed number of parameter settings from given distributions, and GridSearchCV, which tests all possible combinations of hyper-parameters. The 5-fold cross-validation is used also for this test, and the area under the ROC curve is taken as a metric. The decision score and the ROC curve obtained for each fold, are presented in Figure 9.2, together with a comparison of the ROC curves obtained using a different set of signals events (i.e. all Dark Photon masses or one single mass) or using looser selections (noAHOI) for the training. These two comparisons underline an overall good stability of the BDT. In particular, the similar performance when performing the training with different Dark Photon masses are a further hint of the similarity in the kinematics. Similar tests were also performed varying the input variables, showing no relevant improvement from the inclusion of additional variables in the training.

9.3 Background estimation

As already anticipated, the background estimation relies on data-driven techniques for the dominant background arising from "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ and for electrons-faking-photons ($e \to \gamma$), with the first background source including also events with jets-faking-photons. The irreducible background is normalized to data in a specific $VV\gamma$ CR, while the remaining

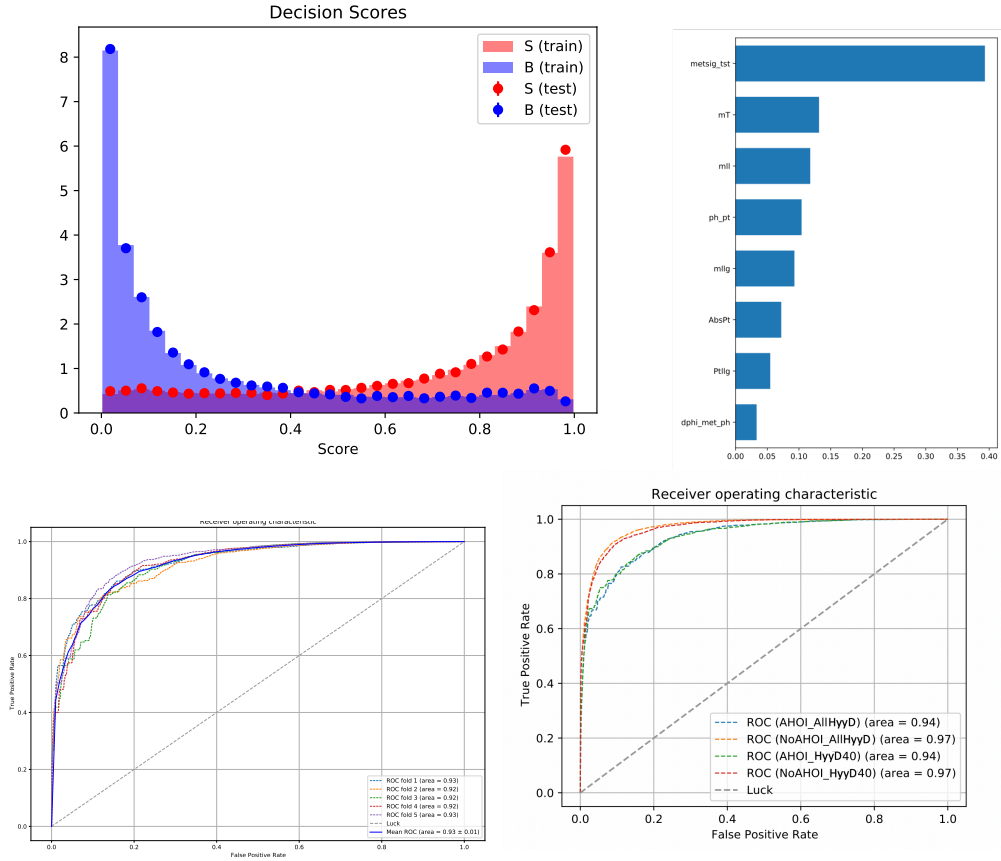


Figure 9.2: The BDT decision score (top left) and ROC curves compared among different folds (bottom left) and combination of γ_d masses (bottom right) with a feature set of 6 variables, and impurity-based feature importance as resulting from a training with 8 input variables.

background sources are extracted from pure MC simulations.

9.3.1 Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background

The "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ background is derived by means of an ABCD method (the generalities of which are described in paragraph "ABCD method" in section 8.4.3), specifically optimized for this analysis. Being all regions dominated by "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ processes, differently than the ABCD method described in section 8.4.3, the leakage of true $E_{\mathrm{T}}^{\mathrm{miss}}$ events in the CR is accounted for by a simple subtraction, in each region, based on MC and on the $e \to \gamma$ in-situ estimate described in the following section, rather then on leakage coefficients analogous to the ones employed in the mono-photon analysis, which were observed to be extremely unstable due to small statistics.
Defining A the SR and B, C and D the CRs, the "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ background in region A is obtained as:

$$N_A^b = R \times \frac{N_B^{b,data} N_C^{b,data}}{N_D^{b,data}}, \; R = \frac{N_A^{b,MC} N_D^{b,MC}}{N_B^{b,MC} N_C^{b,MC}} \tag{9.1}$$

where $N_X^{b,data}$ and $N_X^{b,MC}$ are respectively the number of data after the subtraction of true $E_{\mathrm{T}}^{\mathrm{miss}}$ contribution based on MC, and the MC yields for the "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ background in region N_X^{data}.

Method optimization

Several variables and selection criteria were tested, in order to identify the best choice for the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ ABCD, looking for a pair of weakly correlated variables, providing good discrimination between the signal and the background in region A, while maintaining a statistically relevant population in regions B, C and D.
More specifically, the following optimisation criteria were used:

- R stability: weak dependence of the R coefficient on the choice of the cut values, to ensure good solidity of the method and reduce potential sources of systematic uncertainties;
- R value close to 1, to ensure low correlation among the two variables;
- High signal efficiency of the selection criteria applied to the two ABCD variables ($\epsilon_{sig} = \frac{N_A^{sig}}{N_{A+B+C+D}^{sig}}$), to enhance the analysis sensitivity and minimise the leakage of the signal in the B,C and D regions.
- Good statistics in all the CRs to minimize uncertainties of statistical origin.

An optimal handling of the R factor, and detailed cross-checks of its reliability are a key point of the method optimization, being it the only ingredient which strongly depend on the MC simulation of exactly those processes with "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ for which we are developing a data-driven estimate assuming them to be poorely modelled.
The optimization focused on the $Z\gamma$ + jets process alone, entering the SR only due to "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$, while temporarily not considering the Z+jets process which is characterised also by jets misidentified as photons. The same optimization procedure, applied to a composition of $Z\gamma$ + jets and Z + jets, was afterwards proved to provide similar results (Appendix B.1).

Choice of the variables defining the ABCD plane The first step was the choice of the variables, based on the R stability and signal efficiency criteria. For each tested pair of variables, a scan over different cut values was performed, with an optimization metric based on a Ranking Value (RV), defined to evaluate a single combination of cuts for a given pair of variables, preferring high signal efficiency, and R values closer to 1 and with lower statistical uncertainty ($\frac{\sigma_R}{R}$):

$$RV = \frac{\epsilon_{sig}}{(R + \frac{1}{R})\frac{\sigma_R}{R}} \tag{9.2}$$

and a Figure of Merit (FoM), defined as:

$$FoM = \frac{RV_{best}}{\chi^2_R} \tag{9.3}$$

where RV_{best} is the highest RV achieved in the cuts scan and χ^2_R is the result of the reduced χ^2 test performed, for each pair of variables, over all R values obtained with the different tested cuts.
The $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance variables were tested as one of the two variables of the ABCD, the latter requiring a further cut on top of the SR while the first is already part of SR as previously defined (an attempt was also made to combine the two, as detailed in Appendix B.2, showing no particular improvement). As a second variable, several quantities were tested, exploiting the discrimination power arising from the anti-alignement between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the leptons and photons in signal due to momentum conservation, as opposed to some degree of randomness expected in the studied background:

- $\Delta\phi(\vec{E}_T^{miss} + \vec{p}_T^{\gamma}, \vec{p}_T^{ll})$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$, defined as the azimuthal angle difference between $\vec{E}_T^{miss} + \vec{p}_T^{\gamma}$ and $\vec{p}_T^{ll}$ or $\vec{E}_T^{miss}$ and $\vec{p}_T^{ll\gamma}$ respectively, where $\vec{p}_T^{ll}$ ($\vec{p}_T^{ll\gamma}$) are the vector sum of the leptons (leptons and photon) p_{T}. Since the photon and the leptons are expected to balance the missing transverse momentum produced by the Dark Photon in the signal, $\Delta\phi(\vec{E}_T^{miss} + \vec{p}_T^{\gamma}, \vec{p}_T^{ll})$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ are expected to peak at π, while the background will have a more flat distribution, as the "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ which characterizes these events will have no preferential direction with respect to the di-lepton system;
- $|\vec{E}_T^{miss} + \vec{p}_T^{lly}|/E_T^{miss}$ and $|\vec{E}_T^{miss} + \vec{p}_T^{lly}|/|E_T^{miss} + p_T^{lly}|$, expected to be close to zero for signal events;
- $\Delta\phi(\vec{E}_T^{miss} + \vec{p}_T^{lly}, nearest(\vec{E}_T^{miss}, \vec{p}_T^{lly}))$, defined as the difference between the sum $\vec{E}_T^{miss} + \vec{p}_T^{lly}$ and the nearest object among the two addends. As a consequence of the definition, the mathematical upper boundary is $\frac{\pi}{2}$;
- $min_{obj}\{\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{obj.})\}$ defined as the $\Delta\phi$ between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the closest reconstructed object. The background is characterised by an higher contribution from events with $min_{obj}\{\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{obj.})\}$ close to zero, typically due to a mis-reconstructed object producing most of the fake $E_{\mathrm{T}}^{\mathrm{miss}}$;
- $|\vec{p}_T^{lly}|$, defined as the absolute value of the vector sum of the two leptons and the photon. In the signal events, due to the momentum conservation, $\vec{E}_T^{miss} = -\vec{p}_T^{lly}$.

Using $E_{\mathrm{T}}^{\mathrm{miss}}$ significance as a first variable provided worse performance than $E_{\mathrm{T}}^{\mathrm{miss}}$, also due to the lower statistics. The performance, focusing on R stability and value, is shown in Figures 9.3 and 9.4, which display the R values for different cut values, for the best combination including $E_{\mathrm{T}}^{\mathrm{miss}}$ significance and $E_{\mathrm{T}}^{\mathrm{miss}}$ respectively. Each line shows the

value of R for different cut values on the first variable of the pair (x axis), while different colours represent different cut values on the second variable of the pair. An optimal behaviour would manifest with a flat distribution and almost overlying different colours. The best performance was obtained for $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$. More quantitative results of the scan are shown in Table 9.7, showing the R, FoM and χ^2 values for all tested configurations with $E_{\mathrm{T}}^{\mathrm{miss}}$.

Var	R (ee)	R ($\mu\mu$)	FoM (ee)	FoM ($\mu\mu$)	$\chi^2_R(ee)$	$\chi^2_R(\mu\mu)$
$\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$	0.9412 ± 0.0180	1.1208 ± 0.0205	20.8627	8.4688	0.2721	0.6920
$\Delta\phi(\vec{E}_T^{miss} + \vec{p}_T^{\gamma}, \vec{p}_T^{ll})$	0.6677 ± 0.0167	0.7252 ± 0.0176	5.5884	6.1424	0.8653	0.8505
$\lvert\vec{E}_T^{miss} + \vec{p}_T^{lly}\rvert / E_T^{miss}$	1.1656 ± 0.0299	1.2967 ± 0.0297	0.4854	0.8103	11.0121	6.6320
$\lvert\vec{E}_T^{miss} + \vec{p}_T^{lly}\rvert / \lvert E_T^{miss} + p_T^{lly}\rvert$	0.5177 ± 0.0127	0.6052 ± 0.0136	0.6314	1.8626	7.7196	2.8236
$\Delta\phi(\vec{E}_T^{miss} + \vec{p}_T^{lly}, nearest(\vec{E}_T^{miss}, \vec{p}_T^{lly}))$	0.7515 ± 0.0166	0.7780 ± 0.0160	2.2223	3.6549	2.5473	1.6708
$min_{obj}\{\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{obj.})\}$	1.0282 ± 0.0218	0.8636 ± 0.0182	4.7656	7.3926	1.0640	0.6588
$\lvert\vec{p}_T^{lly}\rvert$	1.2715 ± 0.0225	1.3739 ± 0.0223	1.9635	1.1373	3.4322	6.3268

Table 9.7: Results of the ABCD tests based on $E_{\mathrm{T}}^{\mathrm{miss}}$ variable. The values shown are the ones corresponding to the highest RV among a set of cuts tested. The error shown corresponds to the statistical uncertainty.

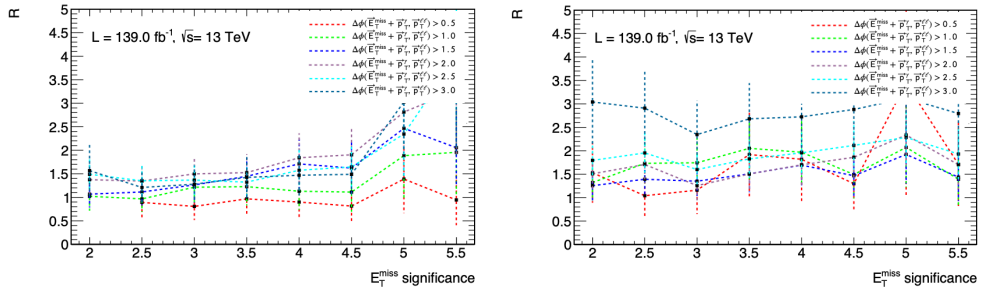


Figure 9.3: Stability of R varying the cut values, for the best found combination using $E_{\mathrm{T}}^{\mathrm{miss}}$ significance ($E_{\mathrm{T}}^{\mathrm{miss}}$ significance and $\Delta\phi(\vec{E}_T^{miss} + \vec{p}_T^{\gamma}, \vec{p}_T^{ll})$), in the ee channel (left) and $\mu\mu$ channel (right)

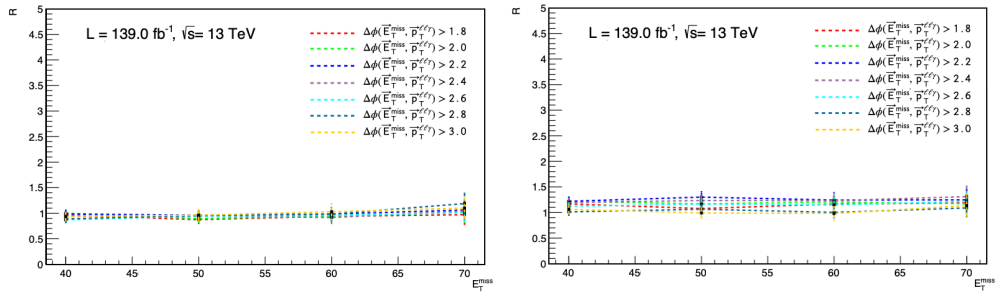


Figure 9.4: Stability of R varying the cut values, for the best found combination using $E_{\mathrm{T}}^{\mathrm{miss}}$ ($E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$), in the ee channel (left) and $\mu\mu$ channel (right).

To help visualizing the low correlation between these two variables for fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background, as well as the discrimination with respect to the signal, the 2D plots of $E_{\mathrm{T}}^{\mathrm{miss}}$

versus $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ versus $E_{\mathrm{T}}^{\mathrm{miss}}$ variables for the signal and for $Z\gamma$ + jets events are shown in figure 9.5, for both ee and $\mu\mu$ channels. The histograms drawn in black on top of the 2D plots correspond to the mean of the y-axis variable in bins of the x-axis variable: the flat behaviour for $Z\gamma$ is a further hint of the fact that the $E_{\mathrm{T}}^{\mathrm{miss}}$ shape does not strongly depend on the $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ region considered.

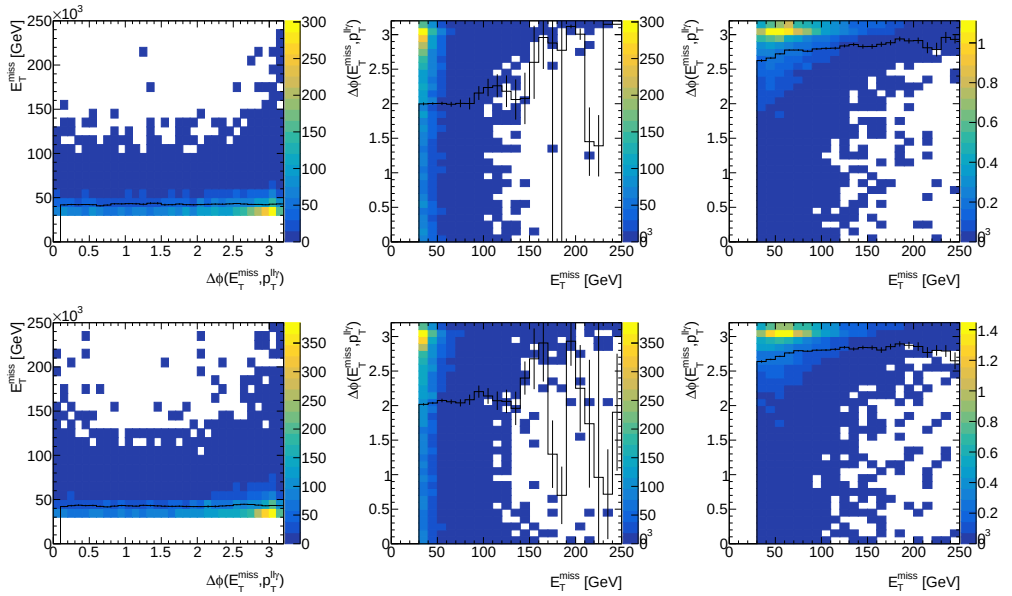


Figure 9.5: 2D plots, in the ee (top) and $\mu\mu$ (bottom) channeld, of $E_{\mathrm{T}}^{\mathrm{miss}}$ VS $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ (left) and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ VS $E_{\mathrm{T}}^{\mathrm{miss}}$ (center) for the $Z\gamma$ + jets process. As a comparison, $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ VS $E_{\mathrm{T}}^{\mathrm{miss}}$ is also shown for the signal (right). The histograms in black represent the mean value of $E_{\mathrm{T}}^{\mathrm{miss}}$ in each $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ bin (and viceversa).

Impact of Z + jets The impact, on the R value, of considering both $Z\gamma$ + jets and Z + jets instead of $Z\gamma$ + jets only was also tested, and is shown in 9.6: the overall R stability is conserved, and the value remains close to 1.

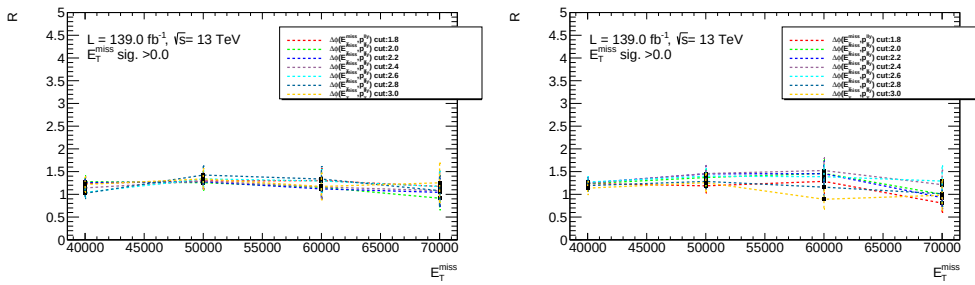


Figure 9.6: Stability of the R coefficient estimated from $Z\gamma$ + jets and Z + jets MC samples, , in the ee channel (left) and $\mu\mu$ channel (right).

Cuts optimization The selections defining the 4 regions of the method have been optimized looking for a compromise between the maximization of signal efficiency in the SR (A) and ensuring a reasonable statistics in all regions. A gap region in $E_{\mathrm{T}}^{\mathrm{miss}}$ have been added, defining two additional A' and C' regions (Figure 9.7), both to further suppress the leakage of Dark Photon signal and true $E_{\mathrm{T}}^{\mathrm{miss}}$ events in the B and D regions and to build a Validation Region where to test the method and the accuracy of the MC description of the correlation factor R.

The final ABCD method

Summarizing, the $E_{\mathrm{T}}^{\mathrm{miss}}$ regions are defined as:

- 30 GeV$<E_{\mathrm{T}}^{\mathrm{miss}}<$40 GeV for the B and D regions
- 40 GeV$<E_{\mathrm{T}}^{\mathrm{miss}}<$60 GeV for the A' and C' regions
- $E_{\mathrm{T}}^{\mathrm{miss}} >$ 60 GeV for the A and C regions

while the $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{ll\gamma})$ cut has been set to $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{ll\gamma}) > 2.4$ for the SR. A sketch of the different regions is shown in figure 9.7.

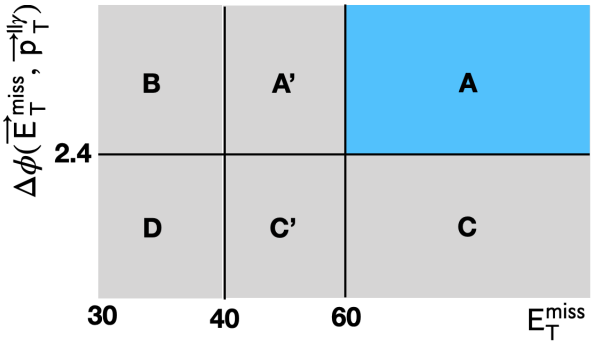


Figure 9.7: Scheme of the ABCD regions with the A' and C' regions. The $x-$ and $y-$axis are inverted w-r-t Figure 9.5.

Figures 9.8 and 9.9 show the data/MC agreement for the two chosen variables, $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{ll\gamma})$, in the ee and $\mu\mu$ channels, in all the CRs of the method. The overall agreement is good, but some mismodelling can be observed in the $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{ll\gamma})$ variable (resulting in the offset between data and MC in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ region). Nevertheless, when combining MC information in the computation of the R value, this mismodelling is observed to have no negative impact, as the R value itself is validated to data in the VR (see Section 9.3.1). This highlights one of the advantages of this method which has the potentialities to somehow "rule out" the mismodelling taking advantage of the non-correlation feature.

The R coefficient is estimated from a composition of all the MC samples associated to processes with fake $E_{\mathrm{T}}^{\mathrm{miss}}$, with dominant contribution from $Z\gamma$ + jets and Z + jets and its uncertainty is obtained through error propagation of the statistical uncertainties in the different regions.
Taking advantage of the R stability, it is possible to estimate an "inclusive" value of R, using A+A' and C+C' regions in place of A and C. This choice allows to overcome the

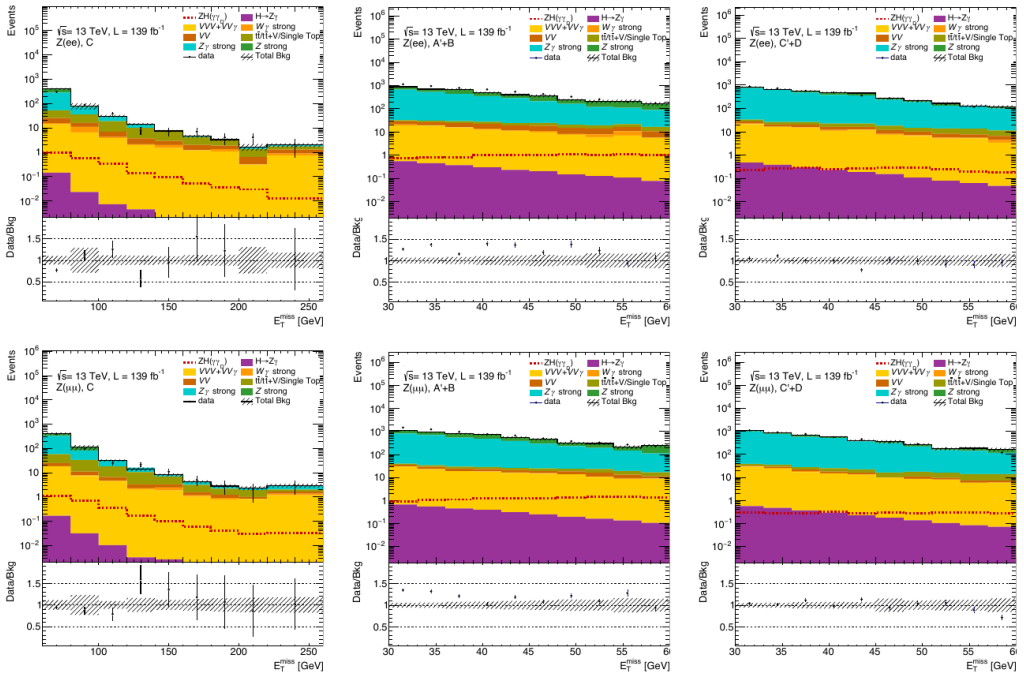


Figure 9.8: Distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the ee channel (top) and $\mu\mu$ channel (bottom), for background processes (mc16a+mc16d+mc16e) and full Run 2 data, in the $E_{\mathrm{T}}^{\mathrm{miss}}$ ¿60 GeV region with $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly}) < 2.4$ (left), and in the 30 $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 60 GeV region with $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly}) > 2.4$ (center) and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly}) < 2.4$ (right). The dashed band and error bars show the statistical uncertainty on the background and data respectively.

low statistics in the C region, observed in particular in the $\mu\mu$ channel (which is indeed the one where the higest difference between the "inclusive" R and the "standard" one is observed). The comparison between the two R values is shown in Table 9.8.

Channel	**R**	$\mathbf{R}_{incl}$
ee	1.201 ± 0.236	1.115 ± 0.109
$\mu\mu$	1.627 ± 0.297	1.244 ± 0.113

Table 9.8: Comparison between the "inclusive" R obtained from (A+A'), B, (C+C') and D, and the "standard" value obtained from A, B, C and D regions.

Concerning the subtraction of background with true $E_{\mathrm{T}}^{\mathrm{miss}}$ from data in the B, C and D regions, it is based on MC for all the processes including a true photon, and for subdominant contribution from $t\bar{t}$ and single top. For the other processes, characterised by the presence of an electron misidentified as a photon, a data-driven estimate is performed as described in section 8.4.2.

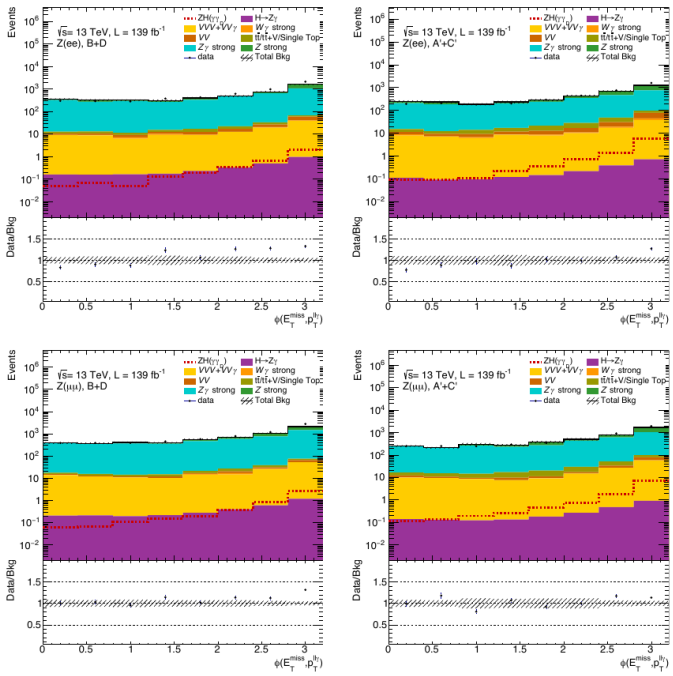


Figure 9.9: Distribution of $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ in the *ee* channel (top) and $\mu\mu$ channel (bottom), for background processes (mc16a+mc16d+mc16e) and full Run 2 data, in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ regions with 30 $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 40 GeV (left) and 40 $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 60 GeV (right). The dashed band and error bars show the statistical uncertainty on the background and data respectively.

ABCD method validation and investigation of systematic uncertainties

Possible sources of systematic uncertainty could arise from the impact of a variation of the cut values, or from a mis-modelling in the relative contribution of different processes. The first one is covered by statistical uncertainties, as demonstrated in Figure 9.10, showing the impact of varying the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut values for $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ cut fixed at 2.4, and the $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ cut for $E_{\mathrm{T}}^{\mathrm{miss}}$ cut value fixed at 40 GeV. The coloured bands show the uncertainty on the R values obtained for the "nominal" configuration ($E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly}) < 2.4$): all the other points are consistent with the nominal value, within uncertainties.

Concerning the second one, the Z + jets background have been re-scaled by 0.5 and 1.5 to test the impact of a high variation of the background composition on the R values: the effect is still covered by statistical uncertainties, as reported in the left part of Table 9.9. Finally, being the R value the only part of the method which is highly relying on MC, the estimate obtained from MC has been validated in the VR, by evaluating an R' ($R' = \frac{N_b^{A'} N_b^{D}}{N_b^{B} N_b^{C'}}$) both using data and MC: the results, reported on the right in Table 9.9, show compatibility within the uncertainties between the two values, supporting the reliability of MC simulations for R estimation.

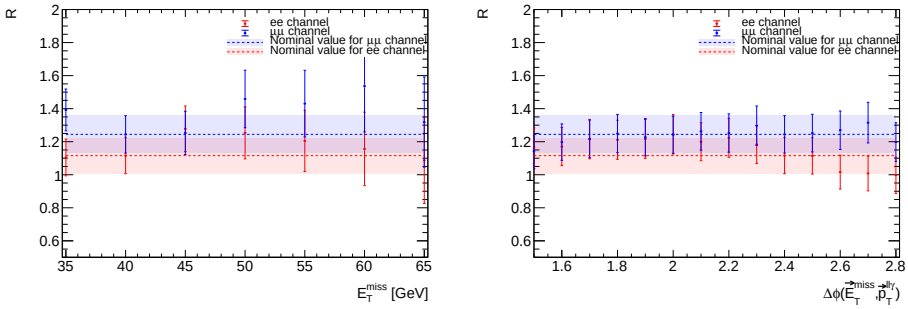


Figure 9.10: R values in the ee (red) and $\mu\mu$ (blue) channel, obtained for different $E_{\mathrm{T}}^{\mathrm{miss}}$ cuts and fixed cut on $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ at 2.4 (left plot) and for different $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ cuts and a fixed cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ at 2.4. The coloured bands correspond to the error bands on the chosen configuration, with $E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly}) < 2.4$. Being the events used to extract the different R values not statistically independent, uncertainties are correlated among different points in the same channel.

Channel	R_{nom}	$R_{Z\ strong\times 1.5}$	$R_{Z\ strong\times 0.5}$	R'_{MC}	R'_{data}
ee	1.12 ± 0.11	1.15 ± 0.13	1.06 ± 0.081	1.094 ± 0.111	1.159 ± 0.056
$\mu\mu$	1.24 ± 0.11	1.25 ± 0.14	1.23 ± 0.089	1.151 ± 0.111	1.181 ± 0.051

Table 9.9: Impact of the variation of Z + jets contribution on the R value, and comparison between R' from data and from $Z\gamma$ strong + Z strong or all background MC simulations in the A'BC'D regions. The error shown corresponds to the statistical uncertainty.

Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background in SR

The final fake $E_{\mathrm{T}}^{\mathrm{miss}}$ estimates, with "inclusive" R from MC associated to all processes characterised by fake $E_{\mathrm{T}}^{\mathrm{miss}}$ ($Z\gamma$ + jets, Z + jets, V$\gamma\gamma$ and $ggH, H \to Z\gamma$), is reported in Table 9.10. The uncertainty on the data-driven estimate is the propagation of uncertainty on R and of statistical uncertainty of data in the ABCD CRs.

Channel	MC	Data-driven
ee	433.3 ± 56.9	413.1 ± 50.2
$\mu\mu$	670.8 ± 66.1	580.8 ± 64.1

Table 9.10: Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background in SR, from pure MC simulations and using the ABCD method. The error shown is the statistical uncertainty (including the uncertainty of R for the data-driven estimate).

9.3.2 Electrons-faking-photons

The electrons-faking-photon ($e \to \gamma$) estimation is performed by means of the same data-driven technique already explained in section 8.4.2 and adopted in the mono-photon analysis. The fake-rate were re-evaluated specifically for the dark-photon analysis, applying the the same e/γ overlap removal procedure and requiring, in place of the $\mu\mu\gamma$ or $ee\gamma$ final states, either ee or $e\gamma$ pairs from Z decay, with $p_{\mathrm{T}} > 20$ GeV thresholds

and identification/isolation WPs consistent with the "selected" photon and electron ones. The probe-e CRs are defined starting from the SR, but selecting $\mu\mu e$ and eee final states.

The electron-to-photon fake rate

The fake rate is measured as a function of $|\eta|$ and p_{T}, $F_{e\to\gamma}(|\eta|, p_T)$, with a total of 48 bins, same as for the mono-photon analysis:
- 6 bins in $|\eta|$: $0-0.8$, $0.8-1.15$, $1.15-1.37$, $1.52-1.81$, $1.81-2.01$, $2.01-2.37$;
- 8 bins in p_{T} (in GeV): 25–35, 35–45, 45–55, 55–65, 65–75, 75–100, 100–150, 150–250.

As an example, in Figures 9.11 and 9.12 the invariant mass distributions are shown for the bin $(\eta, p_{\mathrm{T}}) = ([0, 0.8], [25, 35])$, of Monte Carlo and data respectively.

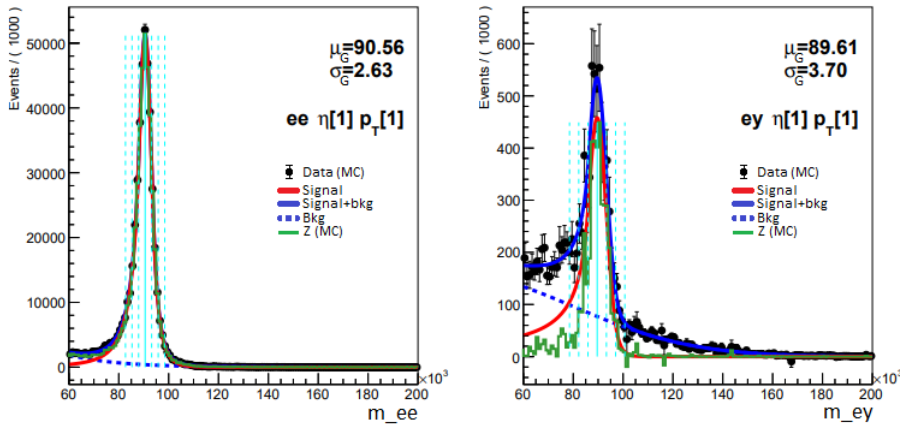


Figure 9.11: Fits for the distributions m_{ee} (left) and $m_{e\gamma}$ (right) in MeVusing MC simulations, for $\eta \in [0, 0.8]$ and $p_T \in [25, 35]$. The green line represents the number of events from the simulated Z_{strong} and Z_{ewk} backgrounds, to be compared with the red line, which models the signal. The dotted blue line models the background events, which results in the continuous blue line when added to the signal. The vertical dotted lines show the intervals in term of σ around the peak.

The background subtraction is performed by means of a background+signal fit, similar to what already described in section 8.4.2:

- Z decays are modelled by a Double-Sided Crystal Ball (DSCB), a probability distribution function (PDF) with a Gaussian core and two power-law tails. The parameters are the mean μ_G and the standard deviation σ_G of the Gaussian, the normalization factor N_Z, the powers, n_{Low} and n_{High}, of the left and right tails, and two parameters describing the transition points between the Gaussian core and the tails, α_{Low} and α_{High};

- for the non-resonant processes an exponential of a 3rd-degree polynomial $N_B e^{c_1 x + c_2 x^2 + c_3 x^3}$ is used, with $x = m_{inv}/(100\ \mathrm{GeV})$ and N_B the normalization factor. This choice was found to provide more accurate fit results in the different p_{T} and $|\eta|$ bins, with respect to the 2nd-degree polynomial used in the mono-photon search.

Two examples of such fits, for an ee pair and an $e\gamma$ one are shown in figure 9.12. The number of ee and $e\gamma$ pairs defining the fake-rate are given by the integral of the distribution

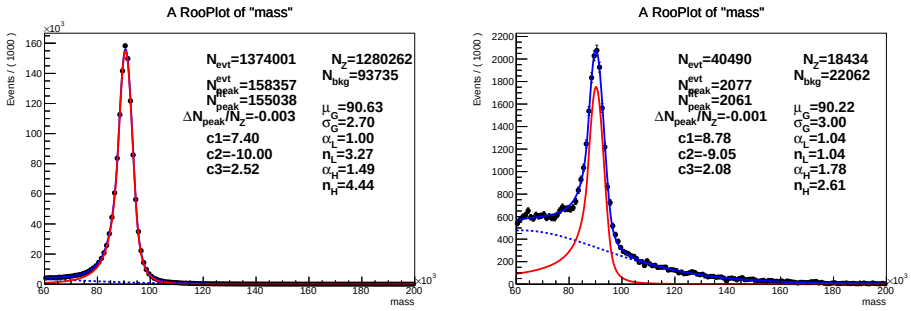


Figure 9.12: Fits for the distributions m_{ee} (left) and $m_{e\gamma}$ (right) in MeV, using data, for $\eta \in [0, 0.8]$ and $p_T \in [25, 35]$ GeV. The signal and the background are represented respectively by the red line and the dotted blue line. The global trend of the fit is represented by the blue line.

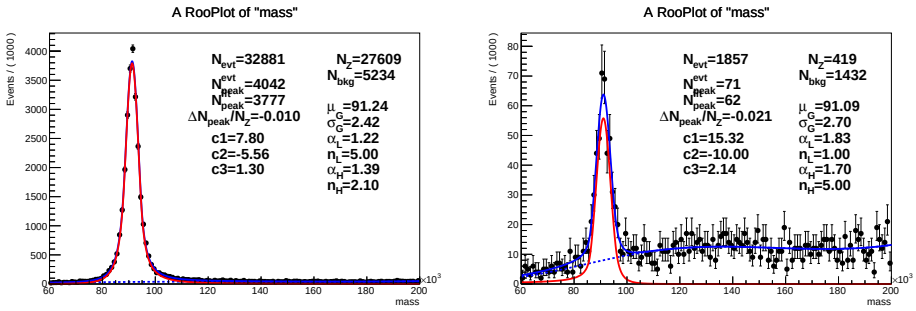


Figure 9.13: Examples of "bad" fits, giving rise to a potential source of systematic uncertainty from mismodelling. The plots correspond to the fits for the distributions m_{ee} (left) and $m_{e\gamma}$ (right) in MeV, using data, for $\eta \in [0.8, 1.15]$ and $p_T \in [100, 150]$ GeV. The signal and the background are represented respectively by the red line and the dotted blue line. The global trend of the fit is represented by the blue line.

of the fitted signal on an interval around the peak, thus excluding the part of the tails dominated by the background events, where the signal may be not well modelled. The uncertainties on the fake-rate are evaluated in a different way than for the mono-photon analysis.

A first contribution to systematic uncertainty is given by inaccurate fits (such as the ones presented in Figure 9.13): their impact is evaluated as the difference $\Delta N_{peak} = |N^{fit}_{peak} - N^{hist}_{peak}|$ between the fitted peak value and the highest bin content in the histogram, reduced by the uncertainty on the highest bin content ($\Delta N_{peak} - \sigma^{hist}_{peak}$). If this difference is negative, it is set to 0, therefore leading to the definition: $\sigma_{fit} = \max\{0, \Delta N_{peak} - \sigma^{hist}_{peak}\}$ where σ^{hist}_{peak} is the uncertainty on the fitted N_{peak} parameter. In addition, the effect of data statistical fluctuations is accounted for by summing in quadrature a poissonian term, resulting in a global uncertainty $\sigma_{N_Z} = \sqrt{\sigma^2_{fit} + N_Z}$ arising from the fit itself.

Another source of systematic uncertainty is the choice of the fitting interval and of the final integral. In order to evaluate it, the fit procedure has applied to the Monte Carlo events (see e.g. Figure 9.11), in order to derive the deviation of the fitted fake rate values ($F^{fit}_{e\to\gamma}$) from the true ones ($F^{true}_{e\to\gamma}$) and introduce a MC-based correction factor for each

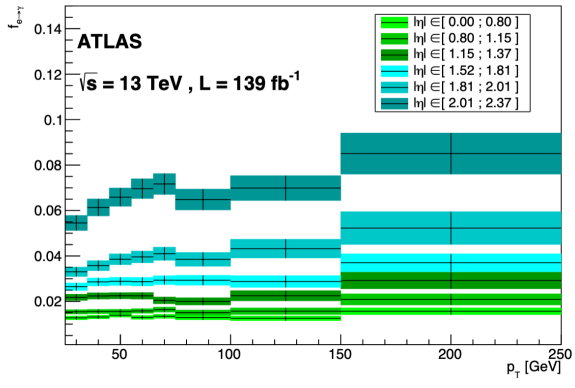


Figure 9.14: Values of fake rates as a function of p_{T} for all η regions.

region in η, p_T:

$$C^{F_{e\to\gamma}} = \frac{F^{true}_{e\to\gamma}}{F^{fit}_{e\to\gamma}} \tag{9.4}$$

The systematic error for each bin is defined as:

$$\epsilon_{syst} = 1 - C^{F_{e\to\gamma}} \tag{9.5}$$

The choice of the fitting and integration ranges was based on a minimization of the quadratic sum, S, of the $(\epsilon_{syst})_{ij}$ for the i-th η and the j-th p_{T} bin. The best found ranges are $[60; 200]$ GeV for the fit and $[-3\sigma_G; +3\sigma_G]$ around μ_G for the integration. Since the ϵ_{syst} values for different bins are distributed symmetrically around a mean value close to 0, suggesting this contribution to be dominated by random errors, a global relative uncertainty for all bins is defined as $\sigma^{intervals}_{rel} = \sqrt{S/48}$, added in quadrature, for each bin, to the previously described systematic sources. The 48 e-to-γ *fake-rates*, with their uncertainties, are displayed in figure 9.14, as function of p_{T}, for different η regions.

Probe-e CRs

The probe-e CRs are defined for the SR and for all the CRs of the "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ ABCD method, by replacing the leading photon requirements with a selected electron. For the eee final states, events characterized by two possible combinations of the ee from Z decays (i.e. opposite sign, same flavour, in the SR invariant mass window) and the probe-electron, are counted twice, as both the identified probe-electrons can indeed mimic a photon. Figure 9.15 shows the good data/MC agreement in the probe-e CR associated to the SR for $E_{\mathrm{T}}^{\mathrm{miss}}$ and $m_T(\gamma, E_{\mathrm{T}}^{\mathrm{miss}})$ variables.

In addition, the possible contamination from $j \to e$, mainly from Z + jets and $t\bar{t}$ events, in the probe-e CR needs to be taken into account and subtracted. This subtraction is performed based on truth information in MC simulations, specifically defining a "scale factor" S which represents the fraction of genuine electrons over all events that enter each

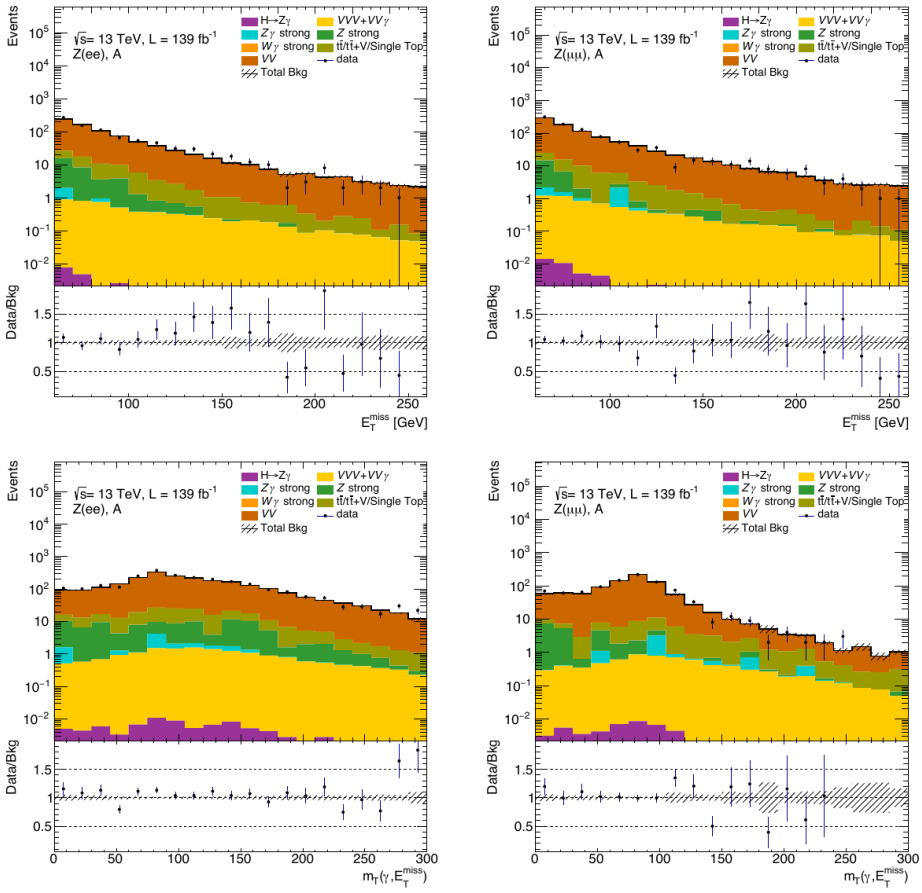


Figure 9.15: Distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ and $m_T(\gamma, E_{\mathrm{T}}^{\mathrm{miss}})$ for the probe-e CR associated to the SR in the ee (left) and $\mu\mu$ channel (right).

probe-e CR:

$$S^{fact} = \frac{N_{\mathrm{true}-e}}{N_{\mathrm{true}-e} + N_{\mathrm{fake}-e}} = \frac{N_{\mathrm{true}-e}}{N_{\mathrm{tot}}} \tag{9.6}$$

where $N_{\mathrm{true}-e}$ is the number of genuine electrons and $N_{\mathrm{fake}-e}$ is the number of jets mis-reconstructed as electrons. The calculated scale factors for all the analysis regions, ranging from $\sim$ 80% up to $\sim$ 95%, are reported in Table 9.11. A 100% uncertainty is then included on this correction factor, evaluated as σ_{SF} = 1 - S^{fact}.

Results in SR

The final $e \to \gamma$ background yields in all analysis regions, obtained by rescaling the yield in the corresponding probe-e CR by the fake-rate in the appropriate $|\eta|$ and p_{T} bin, are summarized in the Table 9.12, together with the number of events in the associated probe-e CR (already rescaled by the previously descriped scale factor).

Probe-e	***ee* channel**	**$\mu\mu$ channel**
SR (Region A)	0.9264 ± 0.0736	0.9403 ± 0.0597
Region B	0.8528 ± 0.1472	0.9068 ± 0.0932
Region C	0.9229 ± 0.0771	0.9242 ± 0.0758
Region D	0.7808 ± 0.2192	0.7887 ± 0.2113
Region A'	0.8969 ± 0.1031	0.9501 ± 0.0499
Region C'	0.8953 ± 0.1047	0.8397 ± 0.1603

Table 9.11: Scale factors with their respective uncertainties for all the probe-e CRs of the analysis.

Region	**Channel**	**Events in probe-e CR**	$e \rightarrow \gamma$
SR (A)	ee	1119 ± 39	$20.975 \pm 0.846 \pm 0.407 \pm 1.466 \pm 1.666$
	$\mu\mu$	1043 ± 32	$20.402 \pm 0.841 \pm 0.343 \pm 1.405 \pm 1.295$
B	$\mu\mu$	988 ± 31	$18.646 \pm 0.854 \pm 0.293 \pm 1.332 \pm 1.916$
	ee	1029.0 ± 38.223	$17 \pm 0.8 \pm 0.313 \pm 1.354 \pm 3.077$
C	$\mu\mu$	583 ± 24	$11.057 \pm 0.632 \pm 0.244 \pm 0.775 \pm 0.907$
	ee	662.0 ± 30.496	$12 \pm 0.7 \pm 0.289 \pm 0.869 \pm 1.034$
D	$\mu\mu$	569 ± 24	$8.332 \pm 0.580 \pm 0.201 \pm 0.684 \pm 2.232$
	ee	565 ± 28	$8.755 \pm 0.621 \pm 0.188 \pm 0.726 \pm 2.458$
A'	$\mu\mu$	1364 ± 37	$26.691 \pm 0.982 \pm 0.385 \pm 1.820 \pm 1.402$
	ee	1448 ± 46	$25.303 \pm 0.944 \pm 0.380 \pm 1.827 \pm 2.909$
C'	$\mu\mu$	656 ± 26	$10.782 \pm 0.653 \pm 0.247 \pm 0.832 \pm 2.058$
	ee	670 ± 31	$11.903 \pm 0.674 \pm 0.215 \pm 0.861 \pm 1.392$

Table 9.12: Number of observed events in the probe-e CR and estimated $e \rightarrow \gamma$ yields in the corresponding analysis region, for full Run-2 dataset, ee and $\mu\mu$ channel. The uncertainty on the $e \rightarrow \gamma$ yields is expressed in four terms: statistical error of the histogram used to count the events, the second and third come respectively from the statistical and systematic error of the fake rate, while the fourth is evaluated as the difference in the estimates of electron faking photons before and after applying the scale factor of probe-e Control Region.

9.3.3 VVγ and top backgrounds

VVγ CR For the irreducible background from $VV\gamma$ events, a CR is defined in order to correct possible mismodelling in the normalization and to cross-check the modelling of the input variables of the BDT. The identification of a suitable CR was challenging: the general idea is to build a $WZ\gamma$ or $ZZ\gamma$ CR in which the neutrinos that characterize the events entering the SR are replaced by leptons, i.e. the Z boson decays into leptons. In order to recover a similar kinematics than in the SR, these additional leptons would be treated as invisible particles in $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, with a similar strategy as the one adopted in the mono-photon analysis. All possible final states were investigated, including either three or four leptons, where at least two opposite sign and same flavour leptons with invariant mass in the Z mass window are present.
From preliminary studies, the global statistics was found to be extremely small, thus leading to the choice of removing the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ selection criteria, to enhance it. Due to the fact that the population in four leptons final states is too limited to allow building a CR specifically for $ZZ\gamma$ processes (as shown in Table 9.13), the focus moves to

the 3 leptons CR option: profiting from the similarity between $ZZ\gamma$ and $WZ\gamma$ processes (see Appendix B.4 for some kinematic comparison between the two samples), a good normalisation for WZγ can be expected to imply similarly good normalisation for $ZZ\gamma$. Finally, final states including electrons are found to provide a less clean CR, with high contamination from $e \to \gamma$ events and potentially $j \to e$, as shown in Table 9.13: in the $3e$ and 2μ+1e final states, the dominant background contribution arises from $e \to \gamma$ (VV process) which approximately double the $VV\gamma$ events. The 2 e+1 μ region looks promising as for the background composition according to MC simulations, but 25 data events are observed in this region, with a high discrepancy with respect to MC suggesting the presence of a not well understood background contribution, possibly from $j \to e$ events. Data/MC comparison in the 2 e+1 μ region for the BDT input variables, highlighting the mismodelling, are shown in Appendix B.3. In order to select the cleanest region, the adopted CR is therefore the 3 muons one (including also 4 muons gives negligible impact on the global statistics, while including a further complexity in the identification of "signal" muons against the "additional" ones to be treated as invisible). Different $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds were investigated, highlighting no particular dependence of the data/MC ratio on the $E_{\mathrm{T}}^{\mathrm{miss}}$ region (Table 9.14), to prove the reliability of releasing the $E_{\mathrm{T}}^{\mathrm{miss}}$ selection.

Lepton sel.	**Tot. bkg**	**VVγ**	**Zγ**	**Z+jets**	**VV**
Inclusive μ	92 ± 3.9	41 ± 0.46	7.6 ± 2.7	2.8 ± 2.7	37 ± 0.61
2 μ+1 e	64 ± 3.6	16 ± 0.3	6 ± 2.4	3.1 ± 2.7	36 ± 0.58
3 μ	24 ± 1.4	20 ± 0.34	1.6 ± 1.3	-0.25 ± 0.2	0.89 ± 0.17
4 μ	2.3 ± 0.051	2.2 ± 0.047	0 ± 0	0 ± 0	0.0099 ± 0.018
Inclusive e	67.41 ± 2.97	27.92 ± 0.37	6.06 ± 2.58	1.27 ± 1.26	28.48 ± 0.58
3 e	52.95 ± 2.95	14.92 ± 0.27	5.96 ± 2.58	1.27 ± 1.26	28.12 ± 0.57
2 e+1 μ	13.04 ± 0.32	11.65 ± 0.25	0.10 ± 0.10	0.00 ± 0.00	11.74 ± 0.25
2 e+2 μ	-	-	-	-	-
4 e	1.42 ± 0.04	1.36 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.03 ± 0.01

Table 9.13: Yields of relevant backgrounds in different leptons categories of $VV\gamma$ CR. The $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ selections are not applied, in order to maximize the statistics

$E_{\mathrm{T}}^{\mathrm{miss}}$ region	**Lepton sel.**	**Data**	**Tot. Bkg**	**VVγ**	**Data/Bkg**
$E_{\mathrm{T}}^{\mathrm{miss}} > 0$ GeV	3 μ	32 ± 5.7	24 ± 1.4	20 ± 0.34	1.3 ± 0.25
$E_{\mathrm{T}}^{\mathrm{miss}} > 10$ GeV	3 μ	31 ± 5.6	23 ± 1.4	20 ± 0.34	1.3 ± 0.25
$E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV	3 μ	29 ± 5.4	21 ± 0.52	19 ± 0.33	1.4 ± 0.26
$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV	3 μ	26 ± 5.1	19 ± 0.51	17 ± 0.32	1.4 ± 0.27
$E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV	3 μ	22 ± 4.7	17 ± 0.41	15 ± 0.3	1.3 ± 0.28
$E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV	3 μ	20 ± 4.5	15 ± 0.39	13 ± 0.28	1.4 ± 0.31
$E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV	3 μ	20 ± 4.5	13 ± 0.37	11 ± 0.26	1.6 ± 0.36

Table 9.14: Yields of data and background (total and contribution from $VV\gamma$) for different $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds in the 3μ category of $VV\gamma$ CR. The $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ selection is not applied, in order to increase the statistics.

The plots in figure 9.16 show a comparison between data and MC for all the input

variables, suggesting, despite the limits of low statistics, no evident mismodelling. Same conclusions can be drawn for BDT modelling, shown in Figure 9.17.

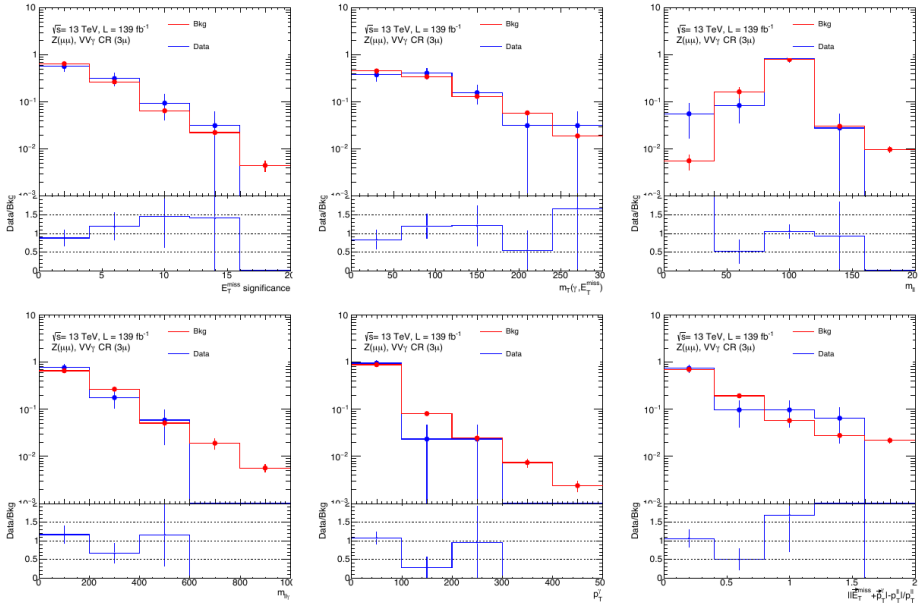


Figure 9.16: Shape comparison of the BDT input variables distributions between data and MC, in the $\mu\mu\mu$ channel of the VVγ CR. The error includes only the statistical uncertainty.

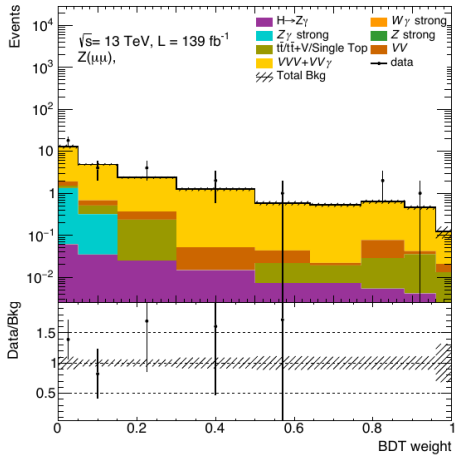


Figure 9.17: BDT weight distribution for data and background in the VVγ CR with 3 muons and no $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ selections. The error shown corresponds to the statistical uncertainty.

Top validation A VR is defined to cross-check the accuracy of the simulations of processes including top quarks in their final states. This top-VR is defined like the SR but inverting b-tagged jet veto and requiring at least one b-tagged jet (with the

same identification criteria as applied for the veto), to select top backgrounds. The $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ and the invariant mass cut are removed in order to maximize the statistics, while the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut is retained in order to suppress the contamination from fake $E_{\mathrm{T}}^{\mathrm{miss}}$ processes. In addition to the inclusive CR ($N_{bjet} > 0$), two separate categories has been studied, requiring exactly 1 b-tagged jet (topCR1b) or 2 bjets (topCR2b). The number of events in data and the data/background ratio, for $E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV, are summarised in table 9.15 for the ee and $\mu\mu$ channel. A discrepancy of less than 20% is observed between data and MC total yields, while the shapes are similar. As a matter of fact, figure 9.18 shows the BDT weight distribution for data and background in the inclusive topCR with $E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV showing overall good agreement on the full range. The shape comparison of the BDT input variables distributions between MC and data, in figures 9.19 and 9.20, show a good agreement as well. The $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ distributions are also shown in figure 9.21. Being these background contribution subdominant in the full BDT range, this region is not used as a CR, but a conservative systematic uncertainty of 20% is associated to these background processes, to account for the observed normalization discrepancy.

Chan	Top sel.	Data	Tot. Bkg	Single top	$t\bar{t}$	$t\gamma$	Data/Bkg
ee	inclusive	513 ± 23	435 ± 7	27 ± 1.9	353 ± 3.7	21 ± 0.18	1.18 ± 0.055
ee	topVR1b	378 ± 19	339 ± 6.7	24 ± 1.8	265 ± 3.2	19 ± 0.17	1.11 ± 0.06
ee	topVR2b	135 ± 12	96 ± 2.3	2.7 ± 0.6	88 ± 1.8	1.8 ± 0.053	1.41 ± 0.13
$\mu\mu$	inclusive	591 ± 24	508 ± 9.5	35 ± 2.2	409 ± 3.9	23 ± 0.19	1.16 ± 0.05
$\mu\mu$	topVR1b	462 ± 21	393 ± 9.2	30 ± 2	303 ± 3.4	21 ± 0.18	1.18 ± 0.06
$\mu\mu$	topVR2b	129 ± 11	116 ± 2.4	5.5 ± 0.86	106 ± 2	2.1 ± 0.056	1.11 ± 0.10

Table 9.15: Yields of data and relevant backgrounds for $E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV and different top categories of top VR, ee and $\mu\mu$ channels. The $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ and invariant mass selections are not applied, in order to increase the statistics.

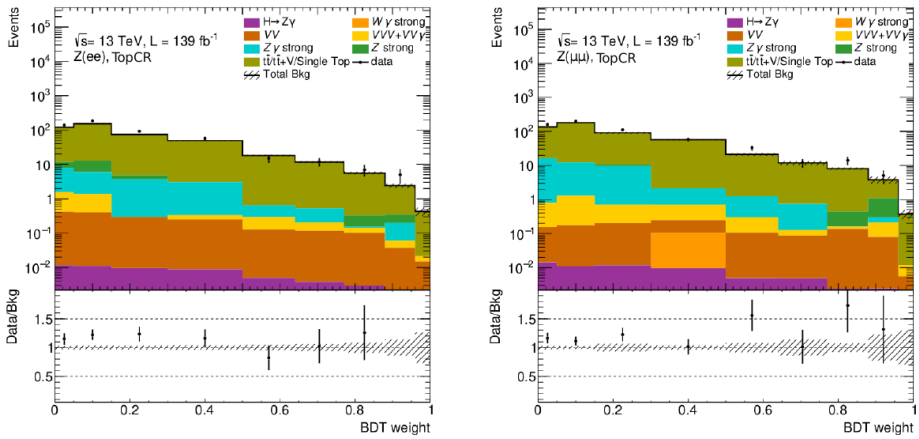


Figure 9.18: BDT weight distribution for data and background in the inclusive TopCR ($N_{bjet} > 0$) for the ee (left) and $\mu\mu$ (right) channel, with no $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ and invariant mass cuts applied and $E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV. The error shown corresponds to the statistical uncertainty.

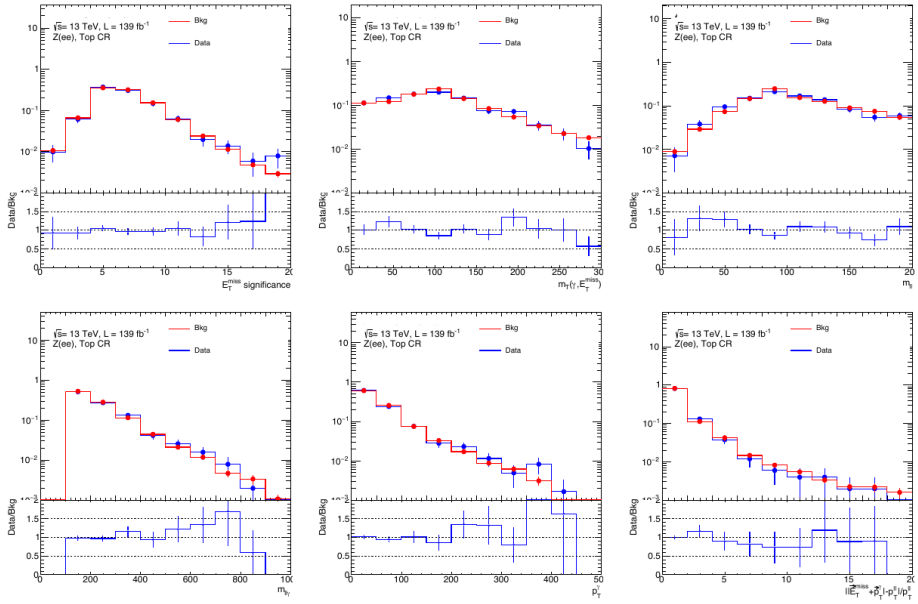


Figure 9.19: Shape comparison of the BDT input variables distributions between data and MC, in the *ee* channel of the top CR. The error includes only the statistical uncertainty.

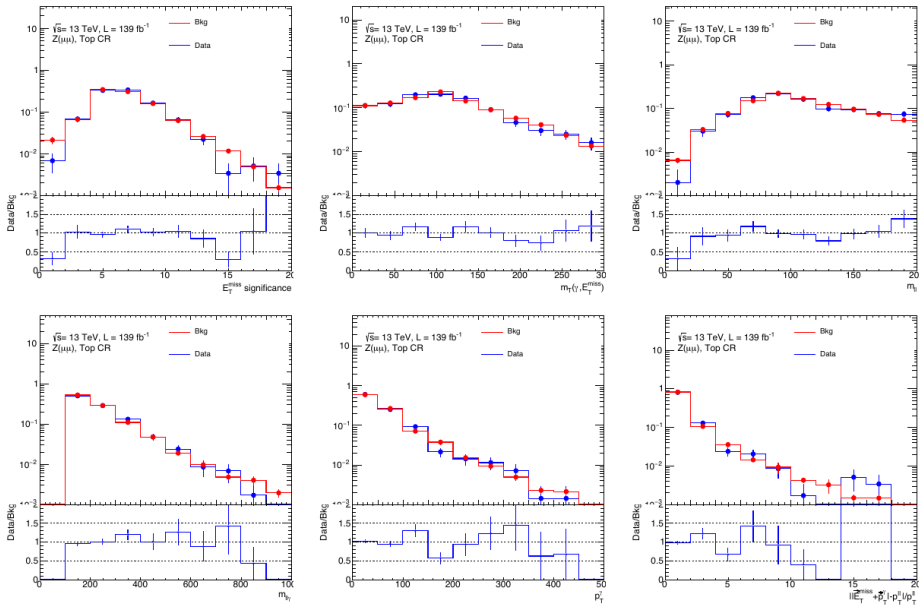


Figure 9.20: Shape comparison of the BDT input variables distributions between data and MC, in the $\mu\mu$ channel of the top CR. The error includes only the statistical uncertainty.

9.3.4 Validation of the background estimate

The data-driven estimates for $e \to \gamma$ and fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background processes are validated in the Validation Region (A') with 40 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 60 GeV. For the other background con-

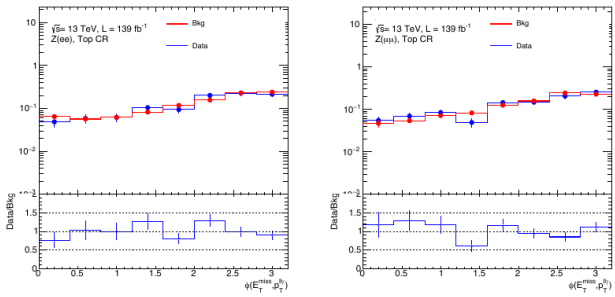


Figure 9.21: Shape comparison of the $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ variable distribution between data and MC, in the ee (left) and $\mu\mu$ (right) channel of the top CR. The error includes only the statistical uncertainty.

tributions, pure MC estimates are used. The total expected background shows outstanding compatibility with data, as highlighted in figures 9.22 showing $E_{\mathrm{T}}^{\mathrm{miss}}$ and $m_T(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma)$ distributions, with a non negligible improvement gained from the employment of data-driven estimate, as demonstrated by the comparison between data and total background using data-driven estimates or pure MC for all processes in Table 9.16.

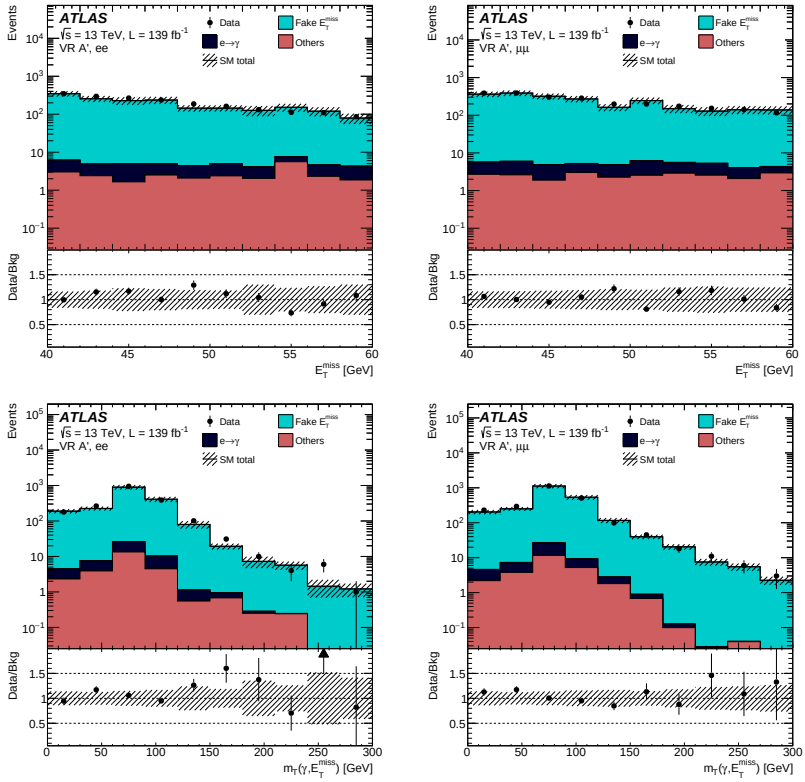


Figure 9.22: Comparison between data and total expected background of the $E_{\mathrm{T}}^{\mathrm{miss}}$ (top) and $m_T(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma)$ (bottom) in the VR. The results are shown in ee (left) and $\mu\mu$ (right) channel.

Channel	***data***	***MC***	***Data − driven***
ee	1936 ± 44	1580 ± 93.5	1838 ± 225
$\mu\mu$	2347 ± 48.4	1995 ± 109	2309 ± 266

Table 9.16: Comparison between the observed data in A', the total observed background obtained using the data-driven method for fake $E_{\mathrm{T}}^{\mathrm{miss}}$ and $e \to \gamma$, and the pure MC expectations. The error shown is the statistical uncertainty and uncertainties associated to the data-driven methods.

Global yields in the SR

The final yields in the ee+$\mu\mu$ SR are summarized in table 9.17 and in the pie-chart in figure 9.23. The table shows the yields in the bins of BDT which will be used in the fit (section 9.5).

BDT bin	**0 - 0.50**	**0.50 - 0.64**	**0.64 - 0.77**	**0.77 - 0.88**	**0.88 - 0.96**	**0.96 - 1**	**CR 0 - 1**
Tot. SM	900 ± 120	90 ± 35	65 ± 27	53 ± 24	35 ± 22	7.8 ± 4.4	24 ± 4.7
Fake E_T^{miss}	800 ± 110	77 ± 33	51 ± 23	41 ± 21	23 ± 19	$3.2^{+4.0}_{-3.2}$	$1.8^{+3.2}_{-1.8}$
$e \to \gamma$	21.4 ± 2.3	3.62 ± 0.86	4.1 ± 1.3	5.3 ± 1.7	5.4 ± 1.9	1.52 ± 0.25	1.24 ± 0.07
$VV\gamma$	32.7 ± 6.8	4.3 ± 1.2	4.8 ± 1.7	3.9 ± 1.5	4.1 ± 1.6	2.56 ± 0.65	20.21 ± 3.29
$t\bar{t}$, $t\bar{t}\gamma$, single t	43 ± 17	4.7 ± 1.9	3.7 ± 1.7	3.0 ± 1.4	2.0 ± 1.0	0.52 ± 0.22	0.65 ± 0.23
$W\gamma$	3.3 ± 1.5	0.42 ± 0.21	1.28 ± 0.69	–	0.04 ± 0.02	–	–
$t\bar{t}H$, VH	0.15 ± 0.03	0.03 ± 0.01	0.04 ± 0.01	0.05 ± 0.02	0.08 ± 0.03	0.03 ± 0.01	$0.15^{+0.18}_{-0.15}$

Table 9.17: Expected yields from SM backgrounds before the background-only fit for the ee+$\mu\mu$ channel in SR and in $VV\gamma$ CR. The uncertainty includes both the statistical and systematic uncertainties.

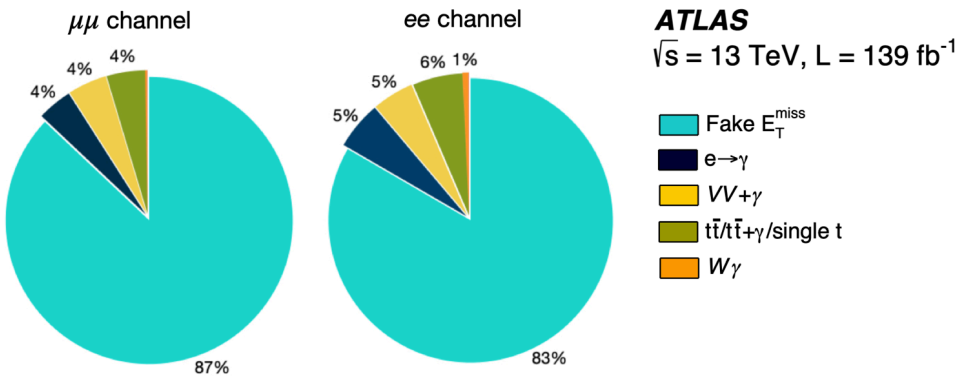


Figure 9.23: Expected relative proportions of different background categories in the signal region, for the $\mu\mu$ (left) and ee (right) channels. Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ and $e \to \gamma$ yields are evaluated data-driven, as explained in the text, while other backgrounds are from MC simulation.

9.3.5 Modelling of the BDT distribution for the "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ background

While the global "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ yields are provided by the data-driven ABCD method, the shape modelling of this background does still relies on MC simulations and needs to be cross-checked. Since the discriminant variable of the analysis is the BDT score, its distribution was studied in the VR and in a low BDT SR, i.e. the BDT<0.5 region with $E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV. Important discrepancies between data and background expectations were observed, as shown in Figure 9.25, which were tracked down to be mainly due to high statistical fluctuations in the $Z+$jets MC simulations (see Appendix B.5). As a matter of fact, despite the high statistics of centrally generated events in these samples, only an extremely low fraction of events enter the $ZH, H \to \gamma\gamma_d$ analysis SR, with less than 10 MC events per bin being selected in the lowest p_{T} slices, the ones with highest cross-section and therefore entering the SR with a highest weight. This effect is not covered by the MC statistical uncertainty alone.
To overcome this, several studies were performed, aiming at finding some strategy to avoid relying on $Z+$jets MC simulations at all, either by extrapolating the BDT shape in the SR from the data shape in the VR ($40 < E_{\mathrm{T}}^{\mathrm{miss}} < 60GeV$), of by using only $Z\gamma+$jets events to model the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ shape. Neither of the two methods resulted promising enough, as detailed in Appendix B.5. Eventually, it was decided to apply a gaussian smoothing to the $Z+$jets BDT shape, trying to "heal" the dips due to statistical issues. The smoothing was applied starting from a 50 bins configuration, using different width gaussian kernels for each bin, depending on its uncertainty. The basic idea is to collect information from neighbour bins, in order to "deduce" the expected bin content based on the rest of the distribution, with a larger window from were to collect the information for more "depleted" bins. The impact of this procedure is shown in Figure 9.24, for the ee and $\mu\mu$ channels in both the VR and the SR. In Figure 9.25, the global expected background with smoothed $Z+$jets, is superimposed in blue to the one pre-smoothing, highlighting the improvement in the modelling which can be gained with this method. Due to a certain degree of arbitrariety in the proposed method (and potential "extreme" effects arising from boundary handling), the smoothed distribution will not replace the one from MC, but rather be used to define an additional systematic uncertainty source, in a conservative approach.

9.4 Systematic uncertainties

9.4.1 Experimental uncertainties

The experimental systematic uncertainties for both the background and signal processes are the same as reported in 7.2, with an additional 20% uncertainty on the top process backgrounds derived from the studies in the top VR and the ones associated to $e \to \gamma$ and "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ data-driven estimates. All these additional contributions are listed in Table 9.18. The "smooth" systematic is the one derived from the $Z+$jets gaussian smoothing described in the previous section, and it is defined as the difference between the BDT shapes (normalized by data-driven estimates) from pure MC and the ones varied by applying the smoothing to the $Z+$jets distribution.

9.4.2 Theoretical uncertainties

The theoretical uncertainties are evaluated similarly for signal and background MC samples, with an additional contribution from electroweak corrections on the signal, and they

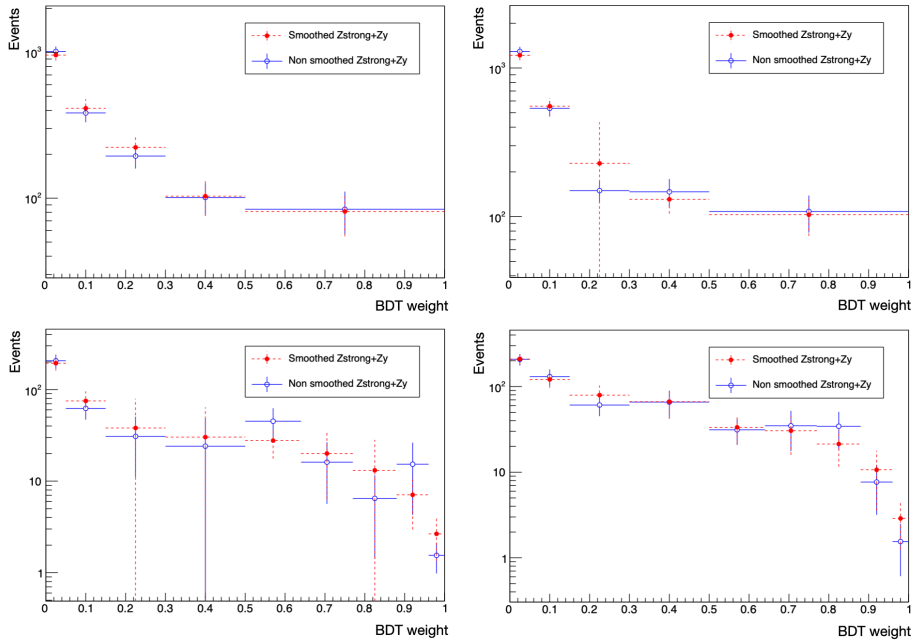


Figure 9.24: Comparison between fake $E_{\mathrm{T}}^{\mathrm{miss}}$ from pure MC (blue) and with smoothed $Z+$jets (red) in the ee (left) and $\mu\mu$ (right) channels of the VR (top) and SR (bottom).

Uncertainties from data-driven methods	
FAKEMET_SYST	uncertainty on the data-driven fake $E_{\mathrm{T}}^{\mathrm{miss}}$ estimate
ELEFAKE_SYST	uncertainty on the data-driven $e \to \gamma$ estimate
Smooth	uncertainty on the shape of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ due to $Z+$jets low statistics
TOP_SYST	uncertainty on top backgrounds normalisation from top CR

Table 9.18: List and description of the additional systematic uncertainty related to data-driven estimates and top VR.

are provided in bins of BDT score for both the SR and the VR.
The scale and PDF variations are derived using the *on-the-fly* weights present in the MC samples, with the procedure described in section 7.2. The main contribution arises from renormalization and factorization scale variations, leading to an uncertainty ranging from 2% up to about 25%, with a strong dependence on the MC sample considered, but overall stable values in different bins of BDT. On the other side, the PDF set and α_S uncertainties reach a maximum value of about 1.5% for signal samples, and range from 0.5% to about 5% in most background samples, reaching higher values (up to about 10%) in subdominant background contributions.
The Parton shower (PS) uncertainties are evaluated by comparing the nominal samples showered with Pythia8, with alternative samples based on Herwig7. *On-the-fly* weights are available and used for the $V\gamma_{QCD}$ and $VV\gamma$ process: in this case, the PS contribution is included in the scale one mentioned above.

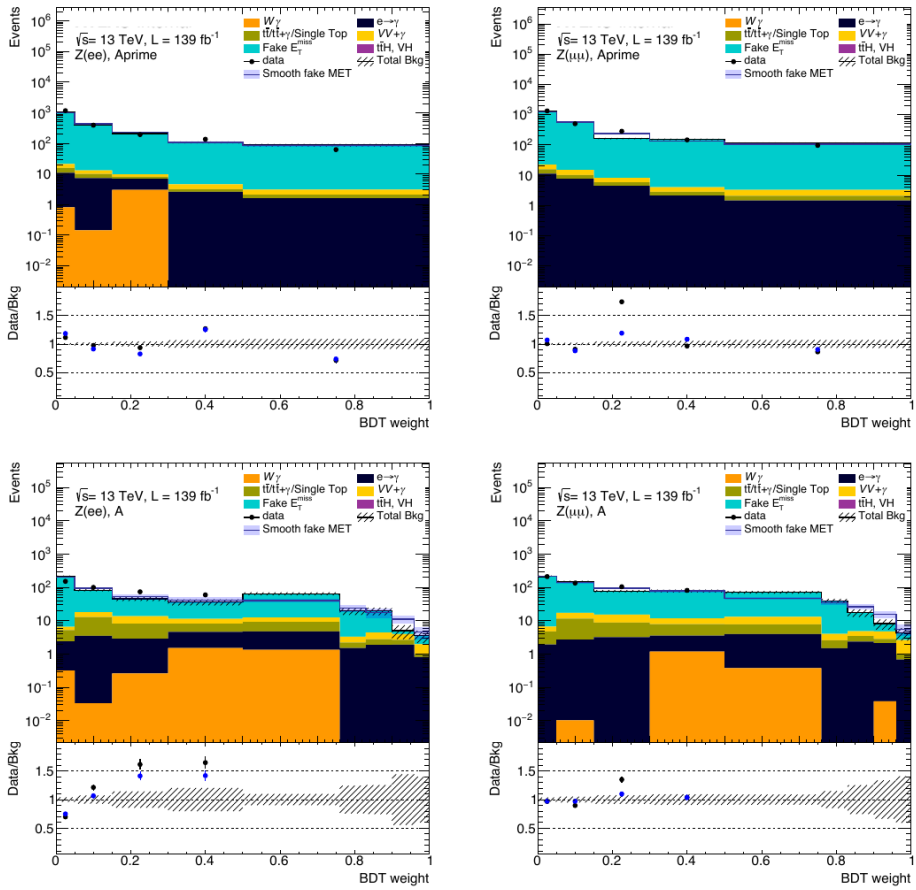


Figure 9.25: Comparison between BDT distributions with and without $Z+$jets smoothing, in the VR (top plots) and SR (bottom plots). The blue lines and blue point correspond to the total background with $Z+$jets events smoothed, and the error bands include only the statistical uncertainty.

Electroweak correction uncertainties

In the MC simulation based on Powheg+Pythia8 generators, NLO electroweak (EW) corrections are not accounted for. The impact of such corrections on the Higgs boson momentum distribution is evaluated through the computation of the ZH cross-section in bins of p_{T}, by means of the HAWK MC software [217]. The corrections are provided by the LHC cross-section working group and amount to a maximum value of $\delta_{EW} = -5.3\%$ for $ZH, Z \to \ell\ell$ process at 13 TeV.
The impact of these corrections on the signal cross-section has been propagated to the BDT score distribution, both in the SR and in the VR. The overall effect is an increase of the total uncertainty of about 0.2-0.3%.

9.4.3 Smoothing of the systematics

A smoothing is applied to both experimental and theoretical systematic, to avoid an overestimation due to limited statistics (in particular for the $Z+$jets process, as detailed in

Appendix B.5): the poor statistics results in huge fluctuations in the systematic variations, potentially leading to a double counting of statistical errors and fit instabilities. All systematic, for all samples, have been smoothed using the CommonSystSmoothingTool [1], with the TRExDefault method (other algorithms were also tested, but shown to be less effective) which consists of two main steps:

- bins are merged until a MC statistical uncertainty below a given threshold is reached. After testing different thresholds, the tolerance has been set to a conservative value of 50% to cope with the large statistical uncertainties in Z + jets;
- an additional smoothing is applied, based on the "353QH twice" algorithm (Friedman in Proc. of the 1974 CERN School of Computing, [218]) to avoid artificially flat uncertainties.

Figure 9.26 shows the impact of the smoothing for the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background, for some of the dominant and most problematic systematic uncertainties.

9.5 The statistical analysis

The fit strategy for this analysis is strongly based on Nuisance Parameters, due to the fact that most of the background is derived in-situ, and only the irreducible $VV\gamma$ background source is normalized to data in a Control Region.
The statistical treatment of the results is performed by means of the HistFitter package ([169]), and the BDT output classifier is used as discriminant variable, with a shape-fit performed in 6 bins. The binning optimization was based on merging bins from the highest down to the lowest BDT bin, starting from a 100 bins distribution and with the lowest bin set to 0-0.5 (used for validation purposes), until a minimum in the expected limit is reached.
The background processes are grouped in six templates:

- Fake $E_{\mathrm{T}}^{\mathrm{miss}}$: $Z\gamma$ + jets, Z + jets, $V\gamma\gamma$, Higgs boson related processes with fake $E_{\mathrm{T}}^{\mathrm{miss}}$
- Electrons faking photons: VV, VVV, top processes with fake photons
- $VV\gamma$
- $W\gamma$
- $t\bar{t}\gamma$, single top, $t\bar{t}$, $Wt\gamma$
- other Higgs boson related processes

The BDT shape for all the templates, except $e \to \gamma$, is taken from MC. For the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background, the normalisation is corrected by the data-driven estimates, while the $e \to \gamma$ BDT distribution is obtained directly from data in the probe-e CR, re-scaled by the fake-rate.

9.5.1 The likelihood function

The statistical model is described by a binned likelihood defined as in Eq. 7.1, with the signal strength as parameter of interest, the $VV\gamma$ normalization factor as an additional

[1] https://gitlab.cern.ch/atlas-phys/exot/CommonSystSmoothingTool

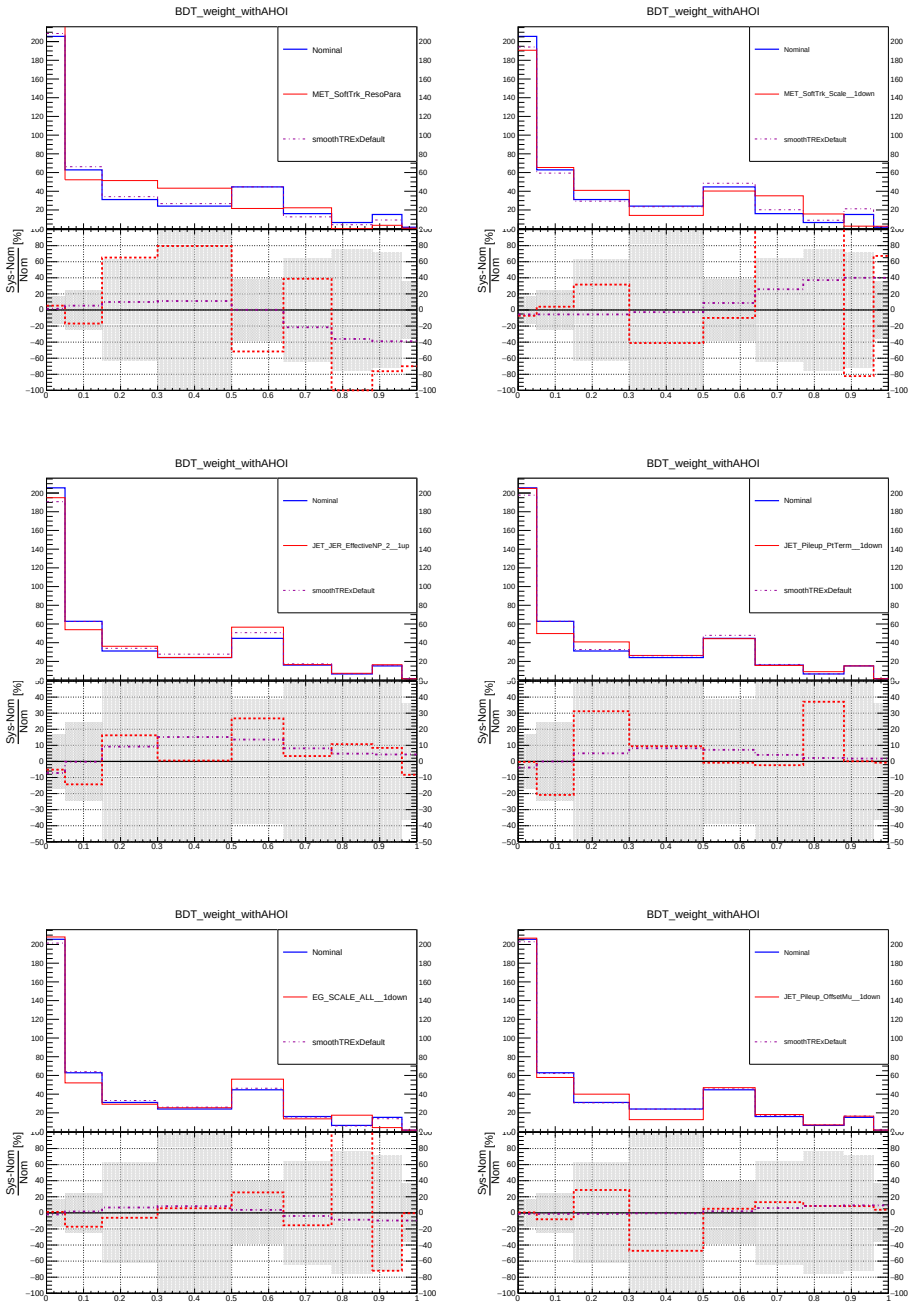


Figure 9.26: Impact of the smoothing on some of the most relevant systematic uncertainties for the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background in the *ee* channel of SR. The systematic variations before smoothing are shown in Red, while the smoothed ones in purple.

free parameter, and the systematic uncertainties and MC statistical ones included as

nuisance parameters with a Gaussian and Poissonian constraint respectively:

$$\mathcal{L}(\boldsymbol{n}|\mu_{sig}, k_{VV\gamma}, \boldsymbol{\theta}) = \prod_{i\in\{\text{BDT bins}\}} \text{Pois}(n_i^{data}|n_i(\mu_{sig}, k_{VV\gamma}\boldsymbol{\theta})) \times C_i(\boldsymbol{\theta}) \tag{9.7}$$

where the $C(\boldsymbol{\theta})$ constraint is defined as in Eq. 7.2 and

$$n_i(\mu_{sig}, k_{VV\gamma}\boldsymbol{\theta}) = n_{i,sig}(\boldsymbol{\theta}) \times \mu_{sig} + n_{i,VV\gamma}(\boldsymbol{\theta}) \times k_{VV\gamma} + \sum_{bkg\neq VV\gamma} n_{i,bkg}(\boldsymbol{\theta}) \tag{9.8}$$

with the θ dependence defined in Eq 7.4. The systematic uncertainties from data-driven methods are included as correlated among different bins, as well as the other experimental systematic variations. In addition, as already mentioned in section 9.3.3, an uncertainty on the BDT shape of fake $E_{\text{T}}^{\text{miss}}$ background is included, and a conservative 20% systematic uncertainty is associated to the subdominant top background to take into account the data/background ratio observed in the top CR.

9.5.2 Validation of the method

The fit strategy went through different steps of validation before fitting to data in the SR. A first fit was performed in the already defined VR (i.e. the A' of the fake $E_{\text{T}}^{\text{miss}}$ ABCD method), in the ee and $\mu\mu$ channel separately. Afterwards, the low BDT region (with BDT < 0.5) was explored, after verifying the negligible contribution from signal in this region. For all the fits, the nuisance parameter pull-plots and correlations were studied. The results of these studies will be summarized in this section, before moving to the fit on data in the SR.

Validation region

For the first validation step in the low $E_{\text{T}}^{\text{miss}}$ VR ($40 < E_{\text{T}}^{\text{miss}} < 60$ GeV), the low BDT region (BDT < 0.5) is splitted into 4 bins, while BDT> 0.5 is merged in a single bin to ensure a signal contribution of less than 5%. The $VV\gamma$ CR is not included in this fit, therefore results are completely based on the NPs described in the previous section and provides a good test of the stability of such fit. The background expectations before and after the fit are compared to data in figure 9.27, showing a good agreement within uncertainties for the ee and $\mu\mu$ channel separately, as well as for the combination of the two. The impact of the fit on the NP is shown in figure 9.28 for the combined channel only: it can be observed that most of the main pulls and constraints are related, as expected, to the dominant "fake" $E_{\text{T}}^{\text{miss}}$ data-driven estimate. The only other constrained NP is the one associated to $E_{\text{T}}^{\text{miss}}$ Soft Term uncertainty, which shows indeed some non negligible correlations with the NPs associated to statistical uncertainties ($\gamma(SRi)$), as shown in the correlation plot in Figure 9.29.

Low BDT region

The BDT < 0.5 bin is splitted in 4 bins, similarly to the VR, for the second validation step, while the highest region is blinded. This test aims at verifying that the fit strategy is stable enough also in the SR $E_{\text{T}}^{\text{miss}}$ range. In this fit, the $VV\gamma$ CR is still not included, being the $VV\gamma$ background completely negligible at low BDT.

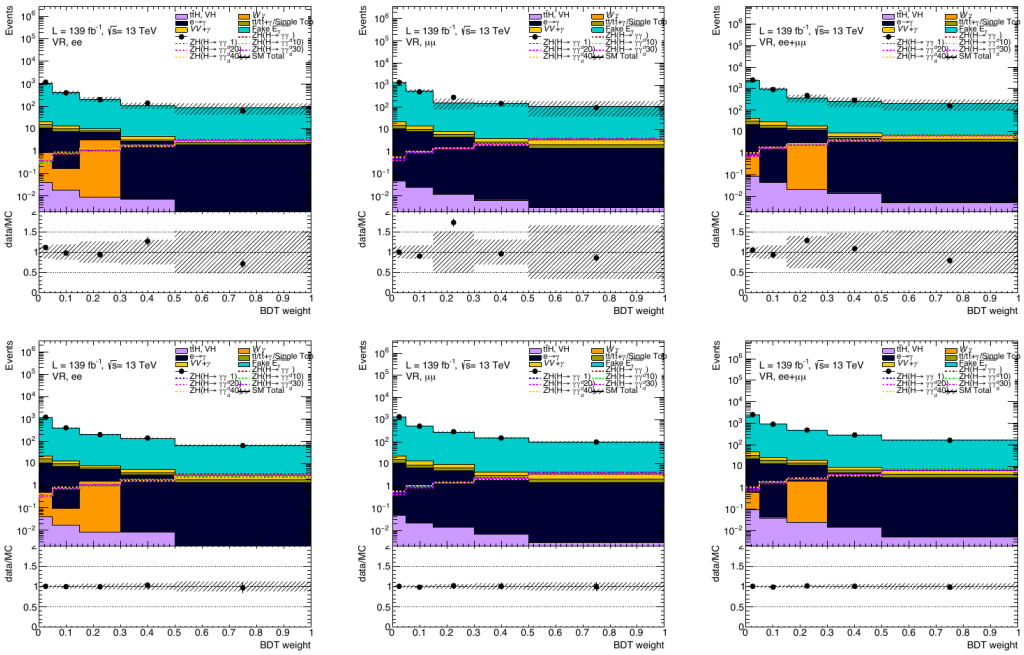


Figure 9.27: Comparison between observed data and expected background in the VR ($40 < E_{\mathrm{T}}^{\mathrm{miss}} < 60$ GeV), ee (left), $\mu\mu$ (center) and combined (right) channels, before (top) and after (bottom) the fit. The error bars include all statistical and systematic uncertainties.

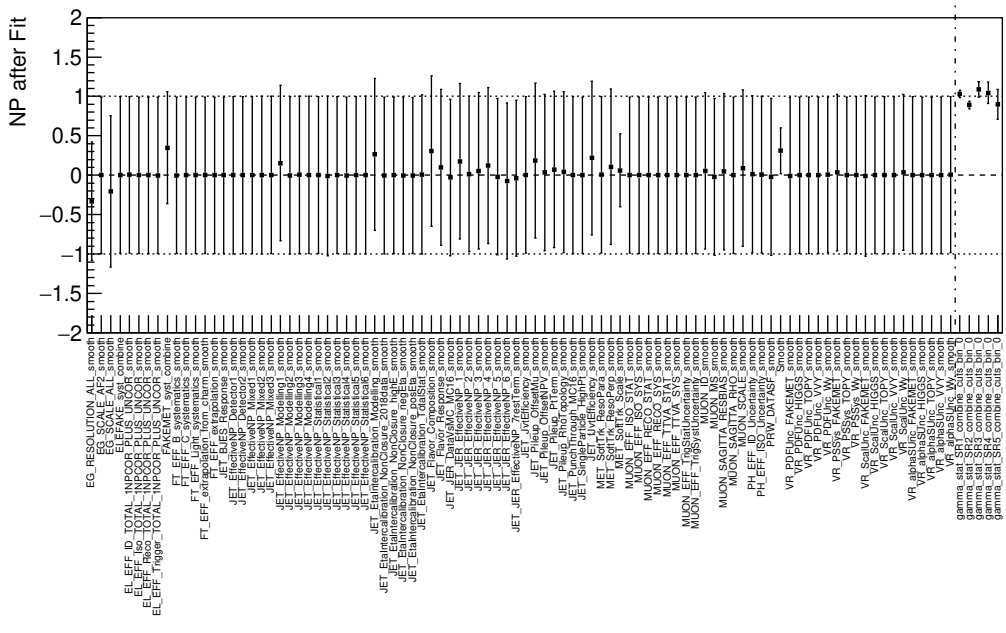


Figure 9.28: Nuisance parameters pull-plots in the VR ($40 < E_{\mathrm{T}}^{\mathrm{miss}} < 60$ GeV) combined channel.

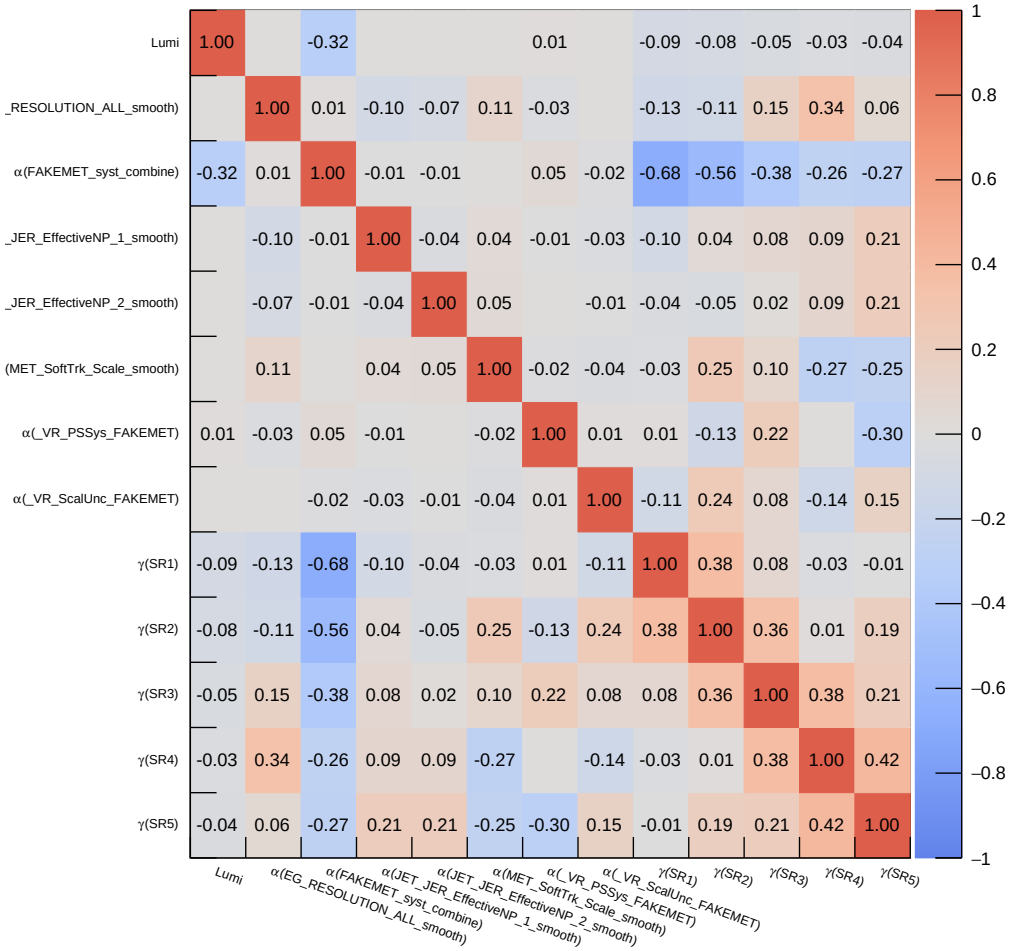


Figure 9.29: Correlation matrix of the nuisance parameters for the combined channel in the VR.

The signal contribution is preliminarily verified to be less than 5% in all bins. The same results as for the VR test are shown in Figures 9.30, 9.31 and 9.32, allowing to draw similar conclusions: a relatively good agreement between data and background expectations can be observed after the fit, and overall stable NPs after the fit, with no relevant over/under constraint other than the somehow expected ones for the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ "smooth" systematic (which is also strongly correlated with the statistical ones), and few pulls only for some JER uncertainties. In this case, the uncertainties before fit are quite large, especially in the single channel case, with an important contribution from the "smooth" systematic itself.

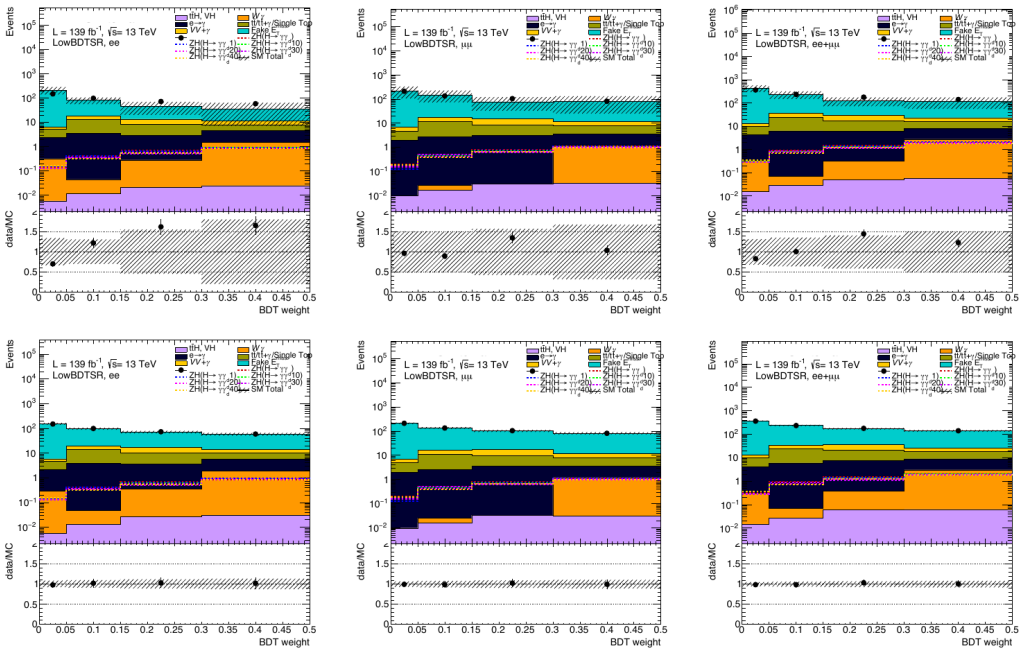


Figure 9.30: Comparison between observed data and expected background in the low BDT (< 0.5) region, *ee* (left), $\mu\mu$ (center) and combined (right) channels, before (top) and after (bottom) the fit. The error bars include all statistical and systematic uncertainties.

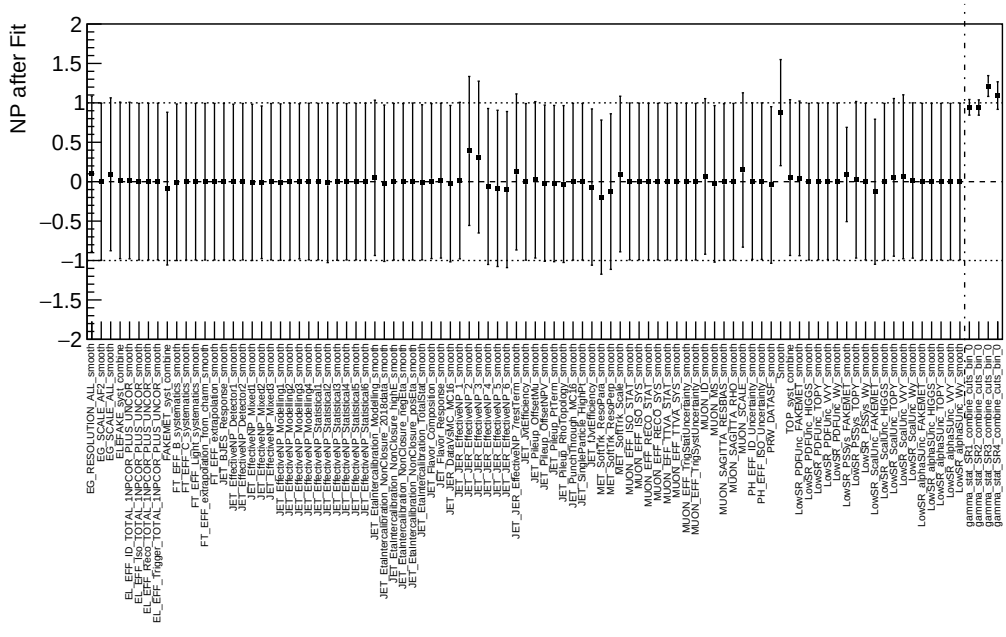


Figure 9.31: Nuisance parameters pull-plots in the low BDT (BDT<0.5) region, combined channel.

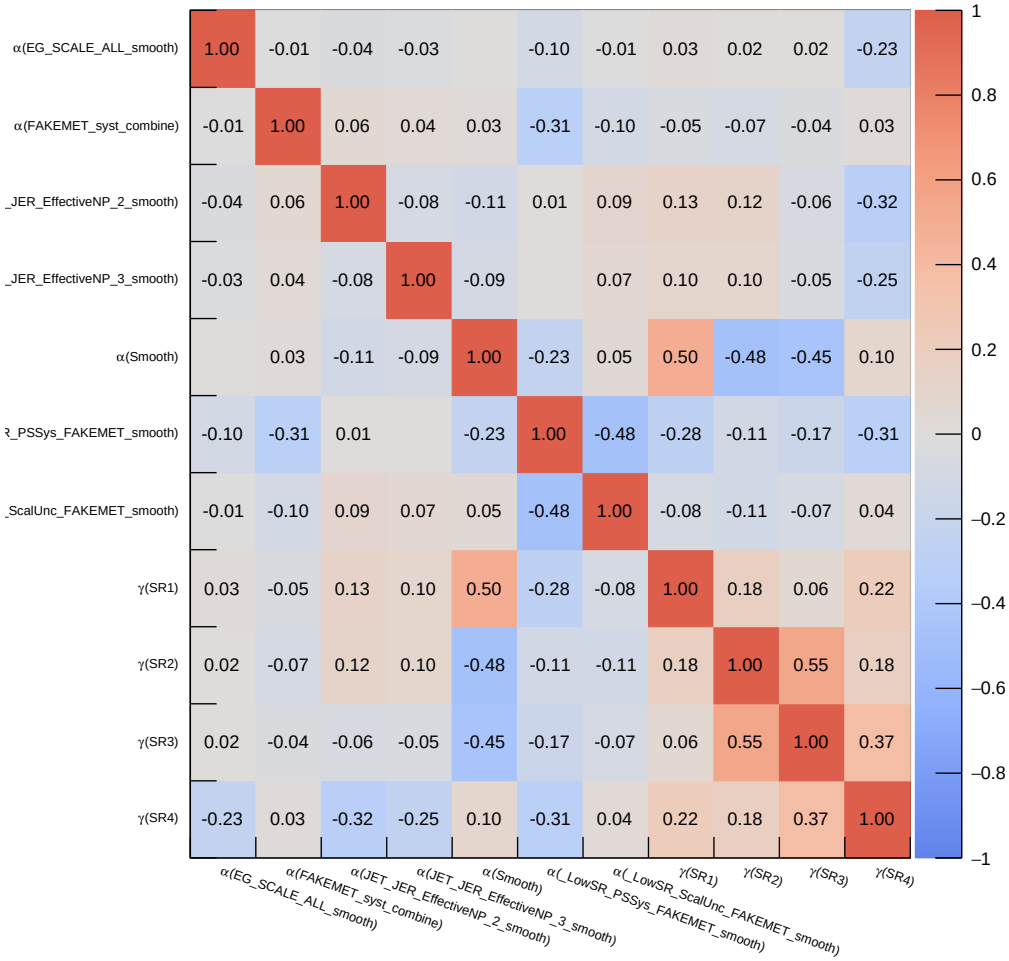


Figure 9.32: Correlation matrix of the nuisance parameters for the combined channel in the low BDT SR.

Impact of the "smooth" systematic uncertainty

From the first fit performed during the validation step, without including the "smooth" uncertainty, some NPs were anomalously pulled, trying to heal the observed discrepancies. in In Figures 9.33 and 9.34, as an example of the impact of thys systematic uncertainty, the pull-plots before and after including it, in the $\mu\mu$ channel of the VR (the one with the highest observed discrepancy), are shown.

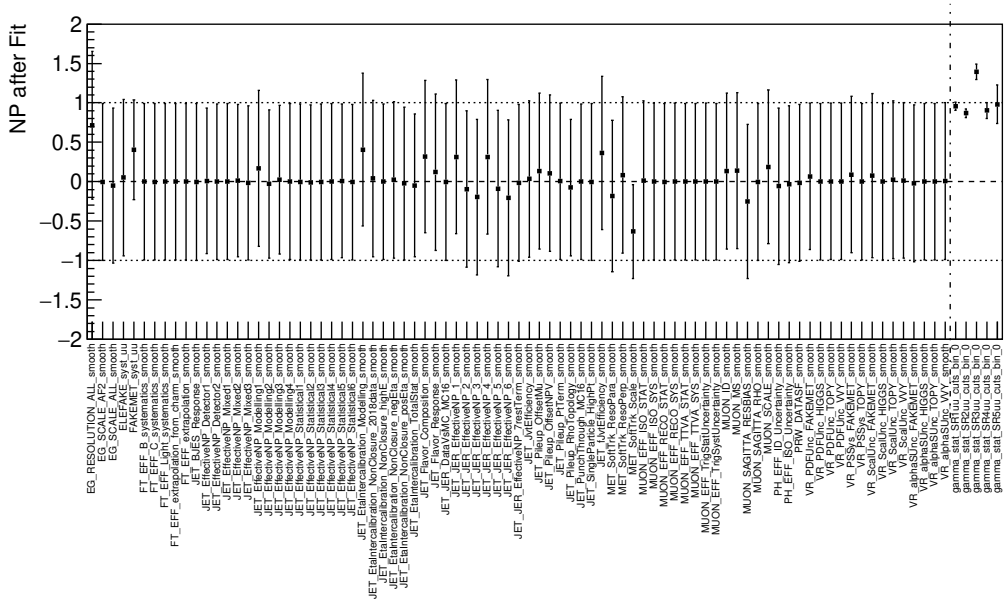


Figure 9.33: Nuisance parameters pull-plots in the VR, $\mu\mu$ channel, without including the "Smooth" systematic.

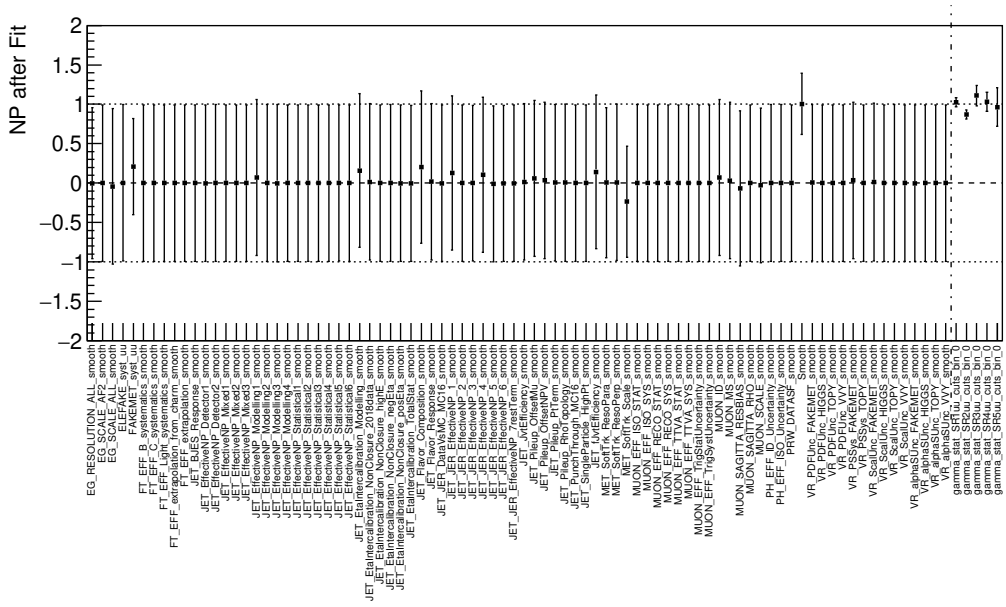


Figure 9.34: Nuisance parameters pull-plots in the VR, $\mu\mu$ channel, including the "Smooth" systematic.

Signal injection test

For a further cross-check of the stability of the fit and its sensitivity to possible signal, in this section some tests performed by injecting a $ZH, H \to \gamma\gamma_D$ signal with BR=5% and different μ_{sig} are reported. Three fitting configurations are tested: a fit without VVγ CR, one with the 1-bin $VV\gamma$ CR used for the final results, and one with a 2-bin VVγ CR (with BDT bing 0-0.5-1) which was considered in the beginning and finally not adopted due to lack of statistics in the second bin (only 4 data events). In addition, two different signal+background Asimov datasets have been built:

- VVγ background expectation from pure MC;
- VVγ background expectation from MC rescaled by the k-factors obtained from a background only fit in the 2-bin VVγ CR configuration.

This is done in order to assess the fit stability in the two most "extreme" scenarios considered, where the VVγ is assumed to be well described either by pure MC or by the most conservative VVγ CR approach.
The results of a signal+background fit to the Asimov dataset in the $ee + \mu\mu$ SR, with an injected signal strength $\mu_{sig} = 1$, are reported in table 9.19, for the MC VVγ and normalized VVγ scenarios. The fitted μ_{sig} are all consistent with 1, but with some deviation (despite within uncertainties) in the two "extreme" configurations: the 2-bin VVγ CR configuration when the VVγ background is assumed to be well described by pure MC, or the configuration with no VVγ CR when the VVγ background is assumed to be well described by MC rescaled by the 2-bin k-factors. The 1-bin VVγ CR, on the other side, seems a good compromise, being able to fit with good accuracy the signal in both cases. The correlation matrix is also shown, in figure 9.35, for the 1-bin VVγ CR fit, on the Asimov dataset build with rescaled VVγ. The signal strength parameter is strongly anti-correlated to the NP associated to the systematic uncertainty from Z strong smoothing, being the latter one of the main uncertainty sources for the dominant background.

$VV\gamma$ from ...	Fit configuration	μ_{sig}	$k_{1,VV\gamma}$	$k_{2,VV\gamma}$
MC	No VVγ CR	1.004 ± 0.400	-	-
	1-bin VVγ CR	0.946 ± 0.399	1.387 ± 0.328	-
	2-bin VVγ CR	0.888 ± 0.418	1.340 ± 0.347	1.774 ± 1.66
2-bin CR	No VVγ CR	1.105 ± 0.420	-	-
	1-bin VVγ CR	1.052 ± 0.421	1.397 ± 0.328	-
	2-bin VVγ CR	0.993 ± 0.436	1.350 ± 0.346	1.772 ± 1.64

Table 9.19: Post fit parameters, for different fitting configuration, as resulting from a signal+background fit to Asimov dataset in SR, combined channel. The Asimov dataset is built including a $ZH, H \to \gamma\gamma_D$ signal with BR=5% and $\mu_{sig} = 1$, and background with VVγ contribution from pure MC and from a background only fit in the 2-bin VVγ CR configuration.

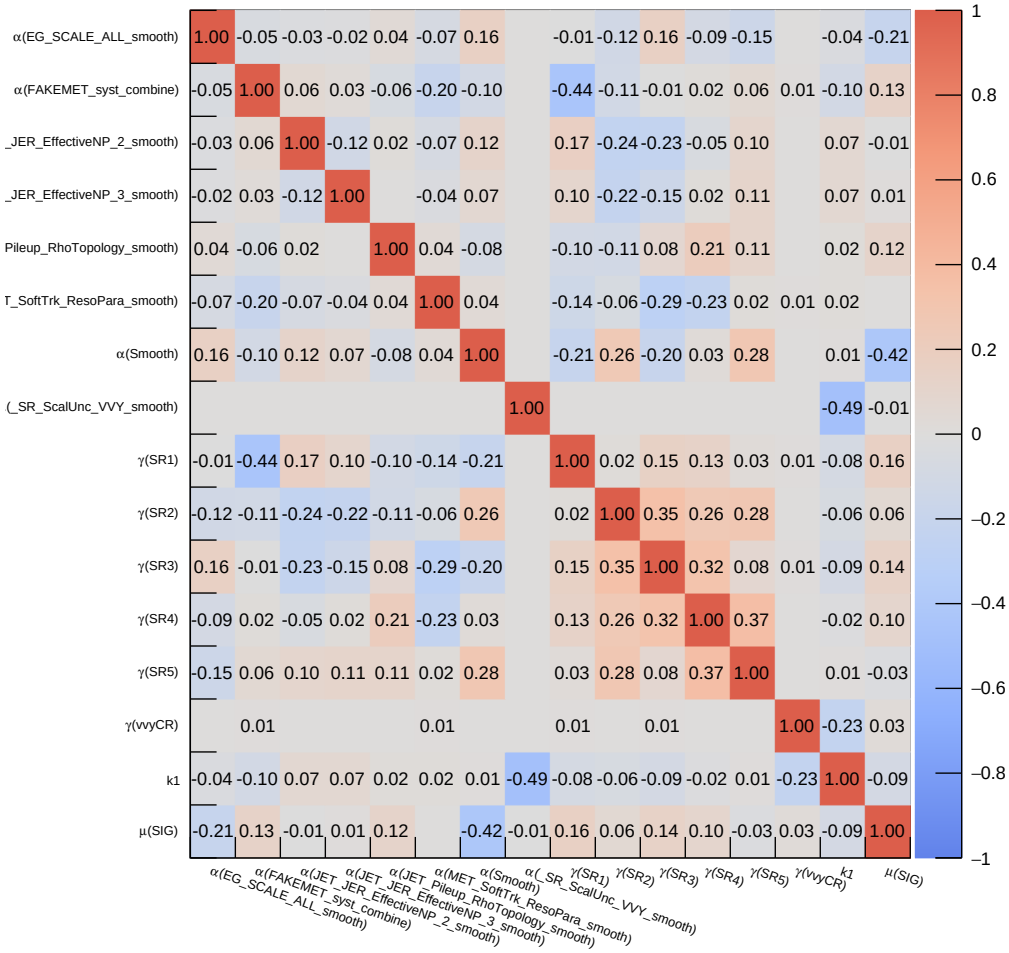


Figure 9.35: Correlation matrix of the nuisance parameters for the combined channel in the SR region, from the simultaneous background+signal fit to Asimov background+signal dataset, with VVγ contribution from the MC rescaled by the 2-bin k-factors. The single-bin VVγ CR is employed in the fit.

In addition, the fit with 1-bin $VV\gamma$ CR configuration, is redone with different injected signal strengths, in order to test the fit linearity. The Asimov dataset is built from the signal rescaled by the considered signal strength and the background taking the $VV\gamma$ contribution from MC normalized to data in 1-bin CR (consistently with the adopted fit configuration). Results are summarized in Table 9.20 and in the plot in figure 9.36.

Injected μ_{sig}	Fitted μ_{sig}	$k_{1,VV\gamma}$
0.5	0.498 ± 0.339	1.395 ± 0.329
1	0.997 ± 0.410	1.394 ± 0.328
1.5	1.501 ± 0.478	1.392 ± 0.328
2	1.998 ± 0.543	1.395 ± 0.329
3	2.998 ± 0.668	1.394 ± 0.329

Table 9.20: Post fit parameters as resulting from a signal+background fit to Asimov dataset in SR, combined channel. The Asimov dataset is built including a $ZH, H \to \gamma\gamma_D$ signal with BR=5% and different injected μ_{sig}, and background with VVγ contribution from MC normalized to data in the 1-bin CR. The 1-bin$VV\gamma$ CR is included in the fit.

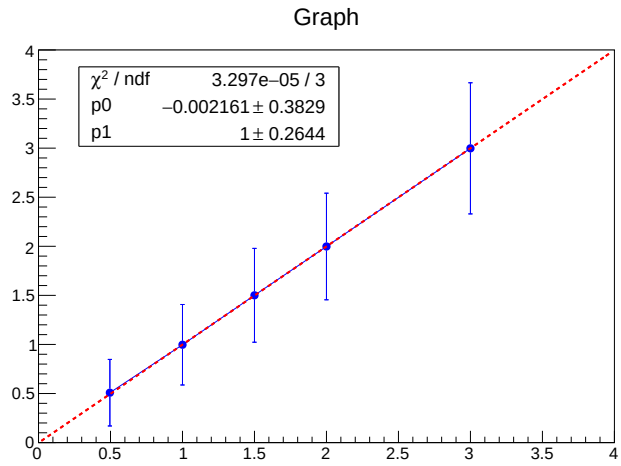


Figure 9.36: The fitted μ_{sig} as a function of the injected ones, as resulting from a signal+background fit to an Asimov dataset built from signal+background, where the VVγ contribution is given by MC normalized to data in the 1-bin CR.

9.5.3 BDT input variables in the SR

Before moving to the background-only fit in the SR, the background estimation and modelling in the SR is tested for the BDT input variables, including also the systematic uncertainties. The data/background comparison is shown in Figure 9.37: the modelling is overall good enough, with discrepancies observed mainly in lowest statistics bins.

9.5.4 Background-only fit in SR+VVγ CR

The final results are based on a background only simplified-shape fit to data, performed simultaneously in the SR BDT bins and in the single-bin $VV\gamma$ CR, using the optimized BDT binning. The SR is included in the fit due to the fact that the analysis is not based on CRs, therefore a fit in the 1-bin $VV\gamma$ CR only would not add particular information with respect to pre-fit results. This configuration basically provides the background-only hypothesis tested against the background+signal one to extract the exclusion limits. Pre- and post-fit BDT distributions in SR and the yields in the single-bin VVγ CR are shown

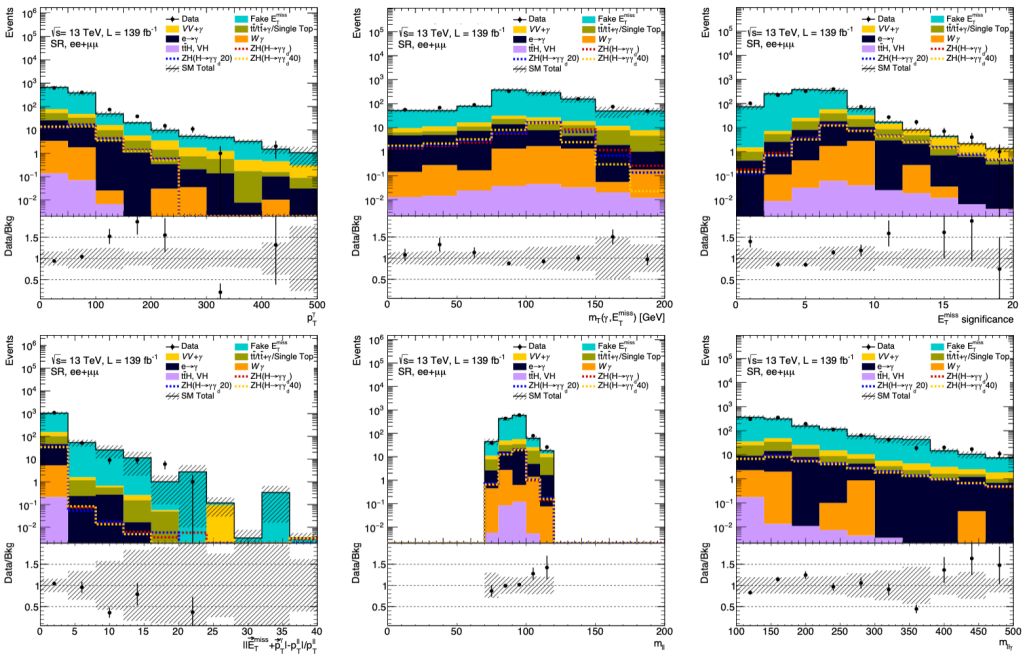


Figure 9.37: Comparison between data and expected background distribution for the BDT input variables in SR. The data-driven estimates are used for the $e \to \gamma$ and "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ background contributions, and the error bands include both statistical and systematic uncertainties.

in figures 9.38, highlighting a good data/background agreement already before the fit, while the yields for each background process, after the fit, are reported in Table 9.21. The VVγ k-factor is 1.350 ± 0.378.

9.5.5 Overview of the systematic uncertainties

The NPs pull plot and correlation matrix in the combined channel, as resulting from the previously described background-only fit, are shown in Figures 9.39 and 9.40 respectively. As expected, the main correlations are observed among the NPs associated to MC statistical uncertainties in the different bins and the "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ related systematics (with correlation level up to about 50%). In addition, a summary of the post-fit systematic impacts, computed as described in Section 8.6.3, is shown in Table 9.22, bin-by-bin and grouped in categories of NPs. The different contributions are reported in order of importance in the last BDT bin: the statistical uncertainty is dominant in all BDT bins except for the last one, where non negligible systematic contributions arise mainly from "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ related uncertainties and from jet energy scale and resolution.

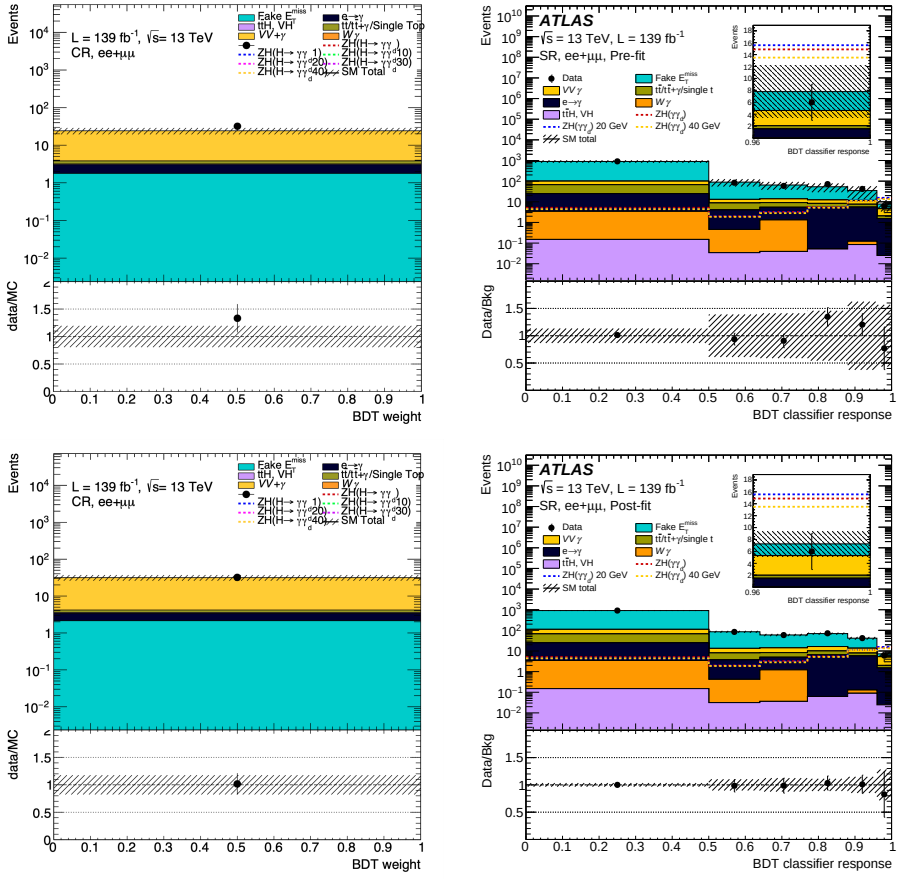


Figure 9.38: BDT score distribution in the single-bin VVγ CR and $ee + \mu\mu$ SR, before (top) and after (bottom) the simultaneous background only fit to data in SR. The error bands include statistical uncertainties, as well as experimental and theoretical systematic uncertainties for the SR.

$ee + \mu\mu$	CR 0 - 1	0 - 0.50	0.50 - 0.64	0.64 - 0.77	0.77 - 0.88	0.88 - 0.96	0.96 - 1
Observed	32	910	84	59	72	42	6
Fitted	31.48 ± 5.39	909.65 ± 29.04	85.44 ± 8.74	59.86 ± 7.33	69.61 ± 7.76	41.59 ± 6.10	7.27 ± 2.04
Fake E_T^{miss}	$2.14^{+3.51}_{-2.14}$	798.17 ± 33.85	72.08 ± 8.30	45.71 ± 6.45	53.17 ± 7.13	27.97 ± 6.13	1.97 ± 1.93
$e \to \gamma$	1.24 ± 0.07	21.46 ± 2.38	3.33 ± 0.65	3.75 ± 0.77	6.40 ± 1.18	5.66 ± 1.47	1.47 ± 0.26
$VV\gamma$	27.30 ± 6.41	44.13 ± 12.48	5.29 ± 1.55	5.84 ± 1.72	6.42 ± 1.84	5.70 ± 1.92	3.30 ± 0.97
top	0.63 ± 0.22	42.44 ± 14.64	4.31 ± 1.51	3.35 ± 1.18	3.56 ± 1.23	2.13 ± 0.80	0.50 ± 0.18
$W\gamma$	0.00 ± 0.00	3.31 ± 1.47	0.39 ± 0.18	1.18 ± 0.55	0.00 ± 0.00	0.04 ± 0.02	0.00 ± 0.00
Higgs boson	$0.17^{+0.18}_{-0.17}$	0.15 ± 0.02	0.03 ± 0.01	0.04 ± 0.01	0.06 ± 0.01	0.09 ± 0.03	0.02 ± 0.01

Table 9.21: Post-fit yields in SR, in the Combined channel, from a background-only fit to data, including a single-bin VVγ CR. The errors shown include both the statistical and systematic uncertainties.

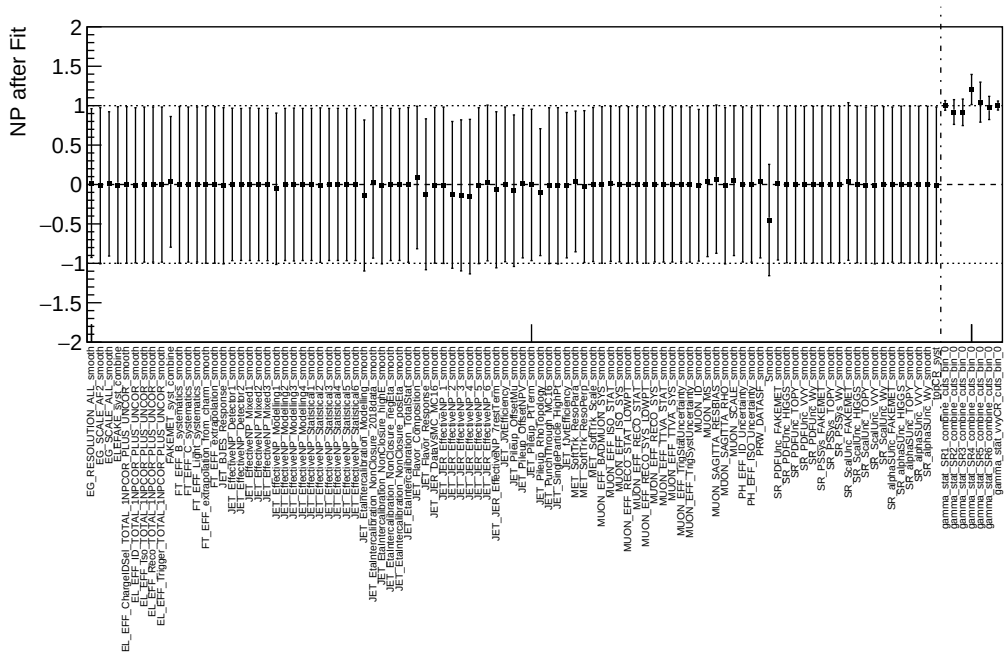


Figure 9.39: Nuisance parameter pull-plot from the simultaneous background only fit, including a single-bin VVγ CR, in the SR for the $ee + \mu\mu$ channel.

BDT bin	**0 - 0.50**	**0.50 - 0.64**	**0.64 - 0.77**	**0.77 - 0.88**	**0.88 - 0.96**	**0.96 - 1**
Total (stat+syst) uncertainty	3.1%	10%	12%	11%	15%	28%
Statistical uncertainty	3.1%	9.9%	12%	11%	14%	16%
Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ shape	0.17%	1.2%	0.15%	0.76%	2.4%	21%
Jet energy scale and resolution	0.09%	4.3%	2.6%	1.1%	0.65%	16%
e/γ energy scale and resolution	0.04%	0.60%	1.0%	0.21%	1.7%	7.4%
μ energy scale and resolution	0.08%	0.10%	0.28%	0.67%	1.3%	4.2%
Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ data-driven	0.52%	0.29%	0.04%	0.04%	0.29%	3.3%
$E_{\mathrm{T}}^{\mathrm{miss}}$ soft term scale/resolution	0.27%	0.10%	0.50%	0.35%	0.33%	0.85%
Top normalization	0.07%	0.09%	0.14%	0.13%	0.07%	0.42%
$e \to \gamma$ data-driven	0.03%	0.10%	0.12%	0.12%	0.06%	0.34%
$\langle\mu\rangle$ reweighting in MC	0.07%	0.09%	0.28%	0.33%	0.20%	0.23%
Flavour tagging eff.	0.02%	0.15%	0.13%	0.11%	0.07%	0.20%
Photon ID/iso/reco eff.	0.00%	0.10%	0.12%	0.11%	0.08%	0.17%
Muon trigger/ID/iso/reco eff.	0.01%	0.15%	0.10%	0.06%	0.07%	0.16%
Electron trigger/ID/iso/reco eff.	0.01%	0.18%	0.09%	0.13%	0.11%	0.15%
Theoretical top	0.09%	0.04%	0.03%	0.17%	0.17%	0.46%
Theoretical $W\gamma$	0.03%	0.13%	0.10%	0.15%	0.09%	0.44%
Theoretical $VV\gamma$	0.03%	0.10%	0.12%	0.08%	0.08%	0.35%
Theoretical Higgs boson	0.01%	0.10%	0.14%	0.08%	0.07%	0.22%
Theoretical fake $E_{\mathrm{T}}^{\mathrm{miss}}$	0.06%	0.13%	0.15%	0.28%	0.36%	0.31%

Table 9.22: Summary of the uncertainties in the background estimate for the BDT bins after the background-only fit. The individual uncertainties can be correlated and do not necessarily add in quadrature to equal the total background uncertainty.

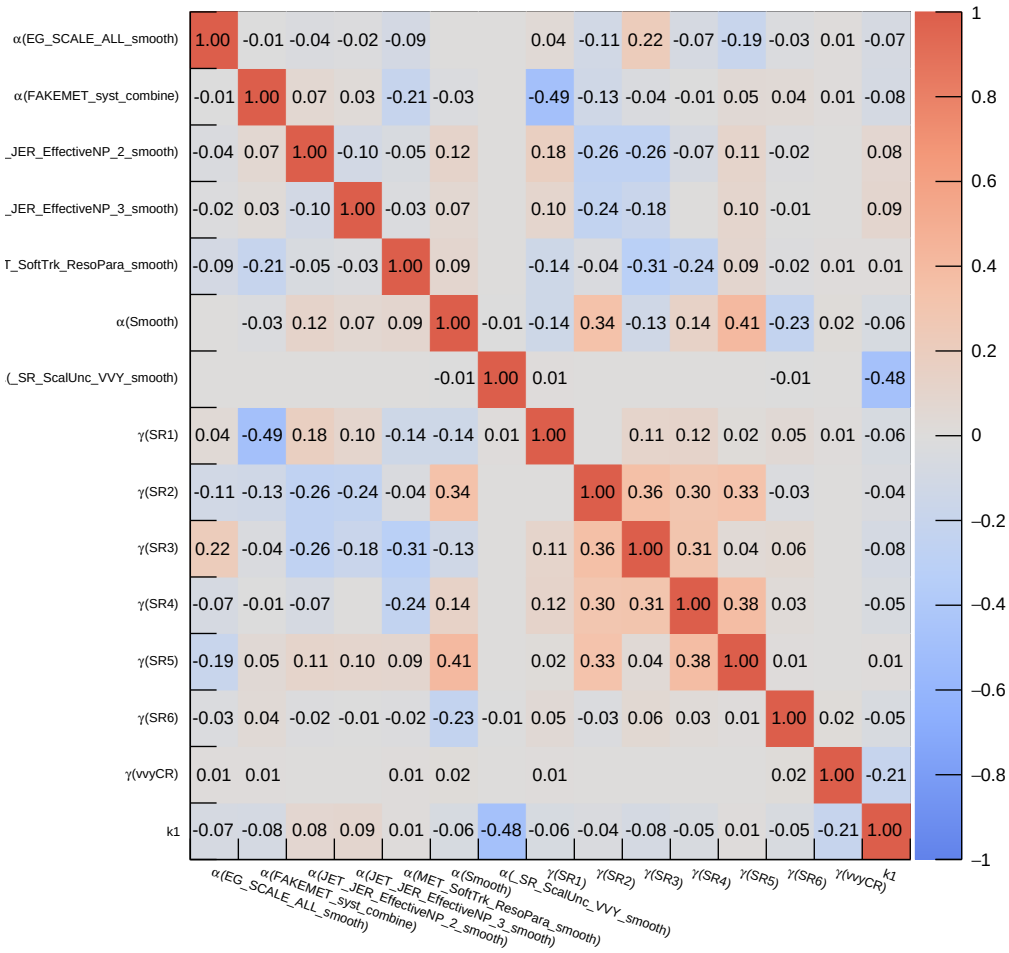


Figure 9.40: Correlation matrix of the nuisance parameters for the combined channel in the SR region, from the simultaneous background only fit to data, including a single-bin VVγ CR.

9.6 Interpretations

The results are interpreted in terms of the $ZH, H \to \gamma\gamma_d$ signal. Since no excess is observed, an exclusion fit is performed in the SR+CR.The signal is taken as the Higgs boson production in ZH mode with Higgs boson decaying into γ and γ_d with a benchmark branching ratio of $BR = 5\%$. The exclusion limit have been performed for all the considered γ_d masses and the results are reported in tables 9.23 (limits with respect to a signal with BR($H \to \gamma\gamma_D$)=5%) and 9.24 (limits in terms of BR($H \to \gamma\gamma_D$)), as well as in Figure 9.41.

$ee + \mu\mu$ channel - limits on μ_{sig}						
m_{γ_D} [GeV]	Obs	Exp	Exp -1σ	Exp $+1\sigma$	Exp -2σ	Exp $+2\sigma$
0	0.456	0.564	0.395	0.830	0.292	1.212
1	0.437	0.541	0.379	0.798	0.280	1.168
10	0.441	0.547	0.383	0.805	0.282	1.174
20	0.434	0.538	0.376	0.794	0.278	1.162
30	0.464	0.574	0.402	0.845	0.296	1.235
40	0.503	0.622	0.435	0.917	0.321	1.344

Table 9.23: Expected and observed exclusion limits on the signal strength assuming BR($H \to \gamma\gamma_D$)=5%, as a function of the γ_D mass, including a 1-bin VVγ CR in the fit

$ee + \mu\mu$ channel - limit on BR [%]						
m_{γ_D} [GeV]	Obs	Exp	Exp -1σ	Exp $+1\sigma$	Exp -2σ	Exp $+2\sigma$
0	2.278	2.820	1.976	4.151	1.458	6.060
1	2.185	2.707	1.895	3.991	1.398	5.842
10	2.207	2.733	1.913	4.024	1.412	5.872
20	2.168	2.689	1.881	3.968	1.388	5.810
30	2.318	2.869	2.009	4.225	1.482	6.174
40	2.517	3.108	2.174	4.587	1.604	6.720

Table 9.24: Expected and observed exclusion limits on the $H \to \gamma\gamma_D$ BR, as a function of the γ_D mass, including a 1-bin VVγ CR in the fit

Table 9.25: Observed and expected limits at 95% CL on BR($H \to \gamma\gamma_d$) for different values of the γ_d mass for the $ee + \mu\mu$ channel. The asymmetric error corresponds to the $\pm 1\sigma$

Smooth syst.	$\text{BR}(H \to \gamma\gamma_d)^{95\%\text{ CL}}_{\text{obs}}$	$\text{BR}(H \to \gamma\gamma_d)^{95\%\text{ CL}}_{\text{exp}}$
	[%]	[%]
Symmetrized	2.28	$2.82^{+1.33}_{-0.84}$
Not Symmetrized	2.11	$2.89^{+1.44}_{-0.91}$
Not included	2.11	$2.89^{+1.44}_{-0.91}$

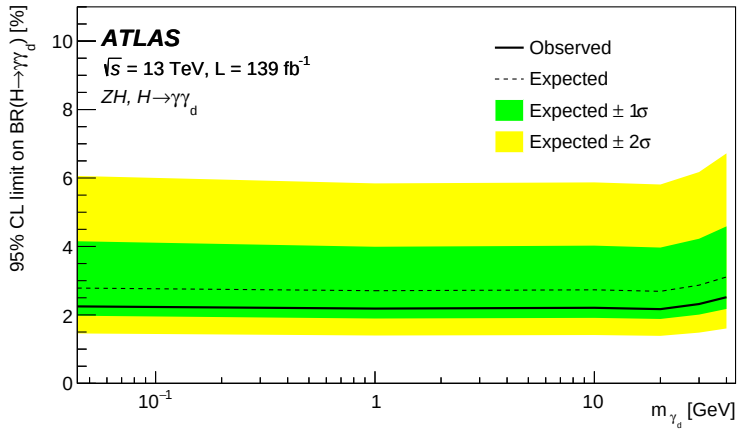


Figure 9.41: Expected and observed exclusion limits as function of the γ_D mass, including a single-bin VVγ CR in the fit. The green and yellow bands show respectively the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties.

Chapter 10

A reinterpretation of the mono-photon analysis in the context of the Dark Photon search

The photon+$E_{\mathrm{T}}^{\mathrm{miss}}$ final state explored in the mono-photon analysis described in Chapter 8, can be also a signature of an Higgs boson decay into a photon and a Dark Photon, with the Higgs produced in the gluon gluon fusion mode. A reinterpretation of the mono-photon analysis in terms of an Hidden Valley model, as described in Section 9.1.3, is therefore an interesting effort, which can in principle provide complementary results with respect to the search in the ZH (Chapter 9) and VBF production modes [147].
The signature of such process would present a photon with energy $E_\gamma = m_H/2$ in the heavy Higgs boson centre-of-mass frame and a similar amount of missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) which originates from the escaping γ_d [143]. A photon+$E_{\mathrm{T}}^{\mathrm{miss}}$ resonance at about the heavy Higgs boson mass is expected, with enhanced γ_d production. Unfortunately, the $E_{\mathrm{T}}^{\mathrm{miss}}$ and photon p_{T} selections defining the mono-photon SRs are tight enough to strongly suppress this kind of signature for the SM Higgs. Nevertheless, the analysis is proven sensitive to BSM Higgs bososn with masses higher than 400 GeV. While the ZH search focused on SM Higgs boson only, BSM Higgs bosons have already been considered in the VBF search and the ggF production mode was partially taken into account in the latter. Similarly, in the analysis here described, both ggF and VBF signals will be considered, since the contribution from the VBF production mode is observed to be non-negligible in the mono-photon SR, due to the acceptance of events with 1 jet. The preliminary studies performed on the analysis acceptance showed good sensitivity to this production mode, especially at higher Higgs boson masses, thus providing an hint of potential complementarity with respect to the VBF analysis.
To conclude, this reinterpretation focuses on massless dark photon, and high Higgs boson masses in the range 400 GeV - 3 TeV, both in the gluon-gluon Fusion (ggF) production mode and in the VBF one, as in the Feynman diagrams shown in Figure 10.1.

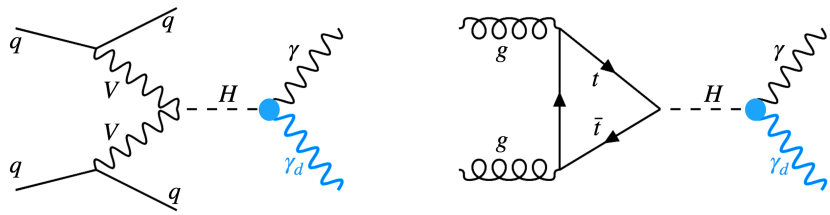


Figure 10.1: Leading order Feynman diagrams for H→ $\gamma\gamma_d$ in ggF (right) and VBF (left) production modes.

The RECAST strategy (Request Efficiency Computation for Alternative Signal Theories) [219] is employed, which consists in a framework designed to ensure preservation

of the original analysis strategy, by reusing the exact same background input histograms, systematic uncertainties and statistical analysis. It is therefore possible to provide reinterpreted results by simply generating MC simulations for the new signals, computing the relative experimental and systematic uncertainties, and providing it as an input to the statistical analysis framework.

10.1 MC samples

Both the ggF and VBF signals predictions are based on MC simulations. The Higgs boson production is simulated using the POWHEG generator, interfaced with PYTHIA for the simulation of the hadronization and parton shower, as well as the Higgs boson decay. The latter is simulated under the Hidden-Valley dark photon scenario [207] implemented in PYTHIA since 8.150 version.
More specifically, the ggF signals are generated with POWHEG [v2 NNLOPS] [220, 221, 222, 223, 224] with the CT10 set of parton distribution functions (PDFs) at NNLO accuracy in α_S [225]. On the other hand, VBF signals are generated using POWHEG [v2] [209, 210] at NLO accuracy in α_S, using the PDF4LHC15_nlo_30_pdfas set of PDFs [167]. Both are interfaced with PYTHIA [8.306] [226] using the CTEQ[6L1] PDF [211] set and the AZNLO underlying-event tune and generated with the width set to the SM value of 4 MeV [208] for all masses, and the complex pole scheme [212] turned off. Finally, the detector geometry and response is simulated with GEANT [159].
A total of eight MC samples per production process were generated with heavy Higgs boson masses equal to 400, 600, 800, 1000, 1500, 2000, 2500 and 3000 GeV. A summary of the mentioned configurations is shown in Table 10.1.

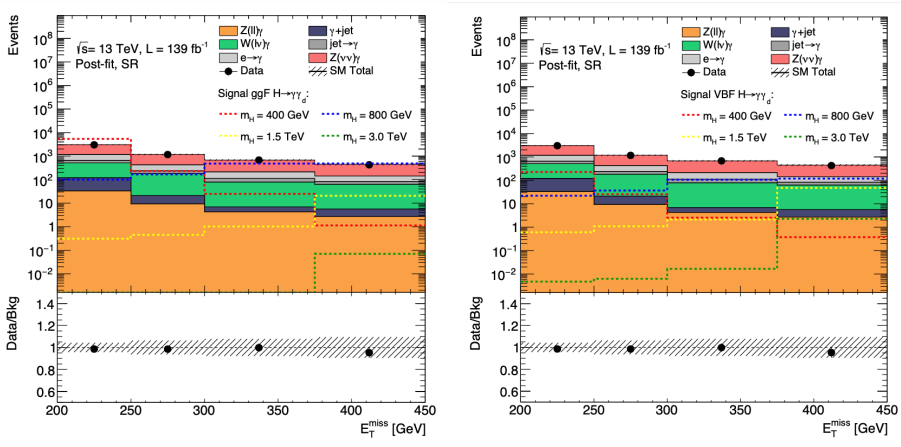
Process	**Generator**	**ME Order**	**Parton Shower**	**PDF**	**Tune**
$ggF, H \to \gamma\gamma_d$	POWHEG [v2 NNLOPS]	NNLO	PYTHIA [8.306]	CT[10], CTEQ[6L1]	AZNLO
$VBF, H \to \gamma\gamma_d$	POWHEG [v2]	NLO	PYTHIA [8.306]	PDF4LHC15, CTEQ[6L1]	AZNLO

Table 10.1: The configurations used for event generation of the signal processes. The matrix element (ME) order refers to the order in the strong coupling constant of the perturbative calculation, while the tune refers to the underlying-event tune of the parton shower.

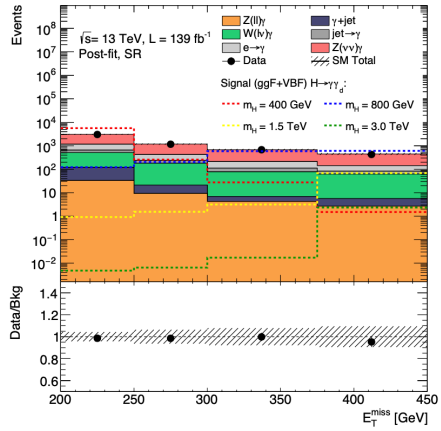
The signal acceptance in the exclusive and last inclusive SR (Table 10.2) is reported in Table 10.2, while the shape of the different $H \to \gamma\ \gamma_d$ signals is shown in Figure 10.2, superimposed with the data and background distribution after a background only fit as described in section 8.6.2. The signal are normalized for the theoretical production cross-sections from [17], reported in Table 10.3, and assuming $BR(H \to \gamma\gamma_d)$=5% as a benchmark value.

10.2 Theoretical uncertainties

The theoretical uncertainties for each ggF and VBF signal are computed in all $E_{\mathrm{T}}^{\mathrm{miss}}$ bins following the same procedure as reported in Section 8.5.2. The sources of uncertainty



(a) **(b)**



(c)

Figure 10.2: Distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ for (a) ggF, (b) VBF (c) ggF+VBF production modes of $H \to \gamma\,\gamma_d$ assuming the BR($H \to \gamma\,\gamma_d$) = 0.05. The signals are superimposed to the expected SM background and data. after performing the 'simplified shape fit'. The error bars are statistical, and the dashed band includes statistical and systematic uncertainties determined by the fit. The lower panel shows the ratio of data to expected background event yields.

	SRE1		SRE2		SRE3		SRI4	
m_H (GeV)	ggF (%)	VBF (%)	ggF (%)	VBF (%)	ggF (%)	VBF (%)	ggF (%)	VBF (%)
400	8.15	4.30	0.35	0.49	0.04	0.05	0.00	0.01
600	9.05	4.95	18.9	9.10	7.74	5.44	0.35	0.53
800	3.21	1.96	5.33	3.27	15.4	9.39	15.6	10.5
1000	1.63	1.24	2.50	1.72	5.92	4.01	29.4	21.2
1500	0.50	0.38	0.73	0.69	1.65	1.33	33.3	30.0
2000	0.22	0.21	0.35	0.33	0.67	0.69	32.7	34.3
2500	0.10	0.09	0.16	0.18	0.35	0.41	29.6	38.0
3000	0.04	0.08	0.08	0.11	0.19	0.29	28.9	39.6

Table 10.2: Signal acceptance (%) for ggF and VBF processes in all the exclusive and the last inclusive SRs used in the "simplified shape fit".

m_H (GeV)	400	600	800	1000	1500	2000	2500	3000
σ_{ggF} (pb)	9.516	2.006	0.449	0.123	8.91×10^{-3}	1.08×10^{-3}	1.78×10^{-4}	3.50×10^{-5}
σ_{VBF} (pb)	0.758	0.327	0.162	0.087	2.28×10^{-2}	7.05×10^{-3}	2.36×10^{-3}	8.25×10^{-4}

Table 10.3: Cross-section σ values of heavy Higgs boson productions in ggF and VBF processes [17].

include the variation of the QCD factorisation and renormalisation scales [227], the α_S scale, the choice of the PDF set and its intrinsic uncertainty, and the uncertainty related to the choice of the tuning parameters for the Parton Shower simulation (splitted into four up and down variations named `VAR1`, `VAR2`, `Ren` and `MPI`), affecting the description of initial- and final-state radiation, underlying events and multiparton interactions.
The global $PDF + \alpha_S$ and PS tuning uncertainties as shown in Tables 10.4 and 10.5 respectively. In the first table, the signal acceptance in each SR is reported to give an estimation of the available statistics. The PS Tuning statistics, on the other hand, is computed using truth level samples with 10000 events each: the number of events passing each SR is directly shown in this case. Only one symmetrized value is provided, conservatively determined as the quadratic sum of the maximum between the up and down variations for each of the uncertainty contributions.
No clear dependence on the $E_{\mathrm{T}}^{\mathrm{miss}}$ bin is observed neither for $PDF + \alpha_S$, nor for the PS uncertainties. The tuning uncertainties, in particular, show particularly unstable results in different bins of exclusive $E_{\mathrm{T}}^{\mathrm{miss}}$, and typically lowee uncertainties for the most inclusive region with $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV than in any of the exclusive bins. To better visualize these systematics the total up and down variations are shown in Figure 10.3 in bins of $E_{\mathrm{T}}^{\mathrm{miss}}$, for the lowest masses (for $m_H > 1$ TeV most of events fall in the last inclusive bin leaving too small statistics at lower $E_{\mathrm{T}}^{\mathrm{miss}}$), together with the four separate contributions: the presence of high statistical fluctuations is clear for $m_H > 400$ GeV and $m_H > 800$ GeV, while some sort of trend in the separate contributions seems present for $m_H > 600$ GeV, despite being "lost" when merging the 4 contribution in a global uncertainty.

ggF				VBF			
m_H = 400 GeV	Acc. [%]	Total		m_H = 400 GeV	Acc. [%]	Total	
SRE1	8.15	-14.54	+16.79	SRE1	4.30	-1.68	+1.45
SRE2	0.35	-15.17	+17.99	SRE2	0.49	-1.04	+1.02
SRE3	0.04	-19.35	+25.37	SRE3	0.05	-3.24	+2.34
SRI4	0.00	-18.72	+24.34	SRI4	0.01	-7.58	+8.93
SRI1	0.00	-14.59	+16.88	SRI1	4.84	-1.54	+1.29
m_H = 600 GeV	Acc. [%]	Total		m_H = 600 GeV	Acc. [%]	Total	
SRE1	9.05	-14.64	+16.05	SRE1	4.95	-2.58	+2.00
SRE2	18.94	-14.67	+16.08	SRE2	9.10	-2.81	+2.44
SRE3	7.74	-14.77	+16.19	SRE3	5.44	-2.27	+1.63
SRI4	0.35	-16.25	+19.23	SRI4	0.53	-1.72	+1.33
SRI1	36.08	-14.70	+16.13	SRI1	20.02	-2.57	+2.04
m_H = 800 GeV	Acc. [%]	Total		m_H = 800 GeV	Acc. [%]	Total	
SRE1	3.21	-14.96	+16.09	SRE1	1.96	-3.02	+2.08
SRE2	5.33	-14.95	+16.17	SRE2	3.27	-3.09	+2.38
SRE3	15.41	-14.94	+16.12	SRE3	0.39	-2.82	+2.05
SRI4	15.63	-15.03	+16.27	SRI4	10.52	-2.32	+1.49
SRI1	39.58	-14.98	+16.18	SRI1	25.14	-2.63	+1.82
m_H = 1 TeV	Acc. [%]	Total		m_H = 1 TeV	Acc. [%]	Total	
SRE1	1.63	-15.26	+16.28	SRE1	1.24	-5.03	+4.36
SRE2	2.50	-15.24	+16.19	SRE2	1.72	-5.13	+4.72
SRE3	5.92	-15.35	+16.31	SRE3	4.01	-4.04	+3.24
SRI4	29.39	-15.37	+16.40	SRI4	21.21	-4.19	+3.28
SRI1	39.44	-15.36	+16.37	SRI1	28.19	-4.21	+3.41
m_H = 1.5 TeV	Acc. [%]	Total		m_H = 1.5 TeV	Acc. [%]	Total	
SRE1	0.50	-15.87	+16.51	SRE1	0.38	-13.90	+6.08
SRE2	0.73	-15.61	+16.42	SRE2	0.69	-6.81	+6.45
SRE3	1.65	-16.21	+17.20	SRE3	1.33	-2.97	+0.77
SRI4	33.29	-16.27	+17.25	SRI4	30.02	-5.52	+4.46
SRI1	36.16	-16.25	+17.22	SRI1	32.43	-5.28	+4.01
m_H = 2 TeV	Acc. [%]	Total		m_H = 2 TeV	Acc. [%]	Total	
SRE1	0.22	-14.46	+13.76	SRE1	0.21	-10.71	+12.49
SRE2	0.35	-15.12	+13.64	SRE2	0.33	-5.58	+4.68
SRE3	0.67	-12.34	+12.73	SRE3	0.69	-4.88	+1.11
SRI4	32.71	-13.34	+13.34	SRI4	34.28	-6.55	+5.20
SRI1	33.94	-13.35	+13.33	SRI1	35.51	-6.46	+5.05
m_H = 2.5 TeV	Acc. [%]	Total		m_H = 2.5 TeV	Acc. [%]	Total	
SRE1	-	-	-	SRE1	0.09	-13.16	+15.41
SRE2	-	-	-	SRE2	0.18	-17.41	+22.87
SRE3	-	-	-	SRE3	0.41	-4.47	+2.26
SRI4	-	-	-	SRI4	37.99	-8.14	+7.35
SRI1	-	-	-	SRI1	38.68	-8.15	+7.38
m_H = 3 TeV	Acc. [%]	Total		m_H = 3 TeV	Acc. [%]	Total	
SRE1	0.04	-22.59	+23.50	SRE1	0.09	-7.81	+4.17
SRE2	0.08	-21.35	+22.79	SRE2	0.12	-20.96	+28.92
SRE3	0.19	-20.42	+21.25	SRE3	0.29	-16.28	+20.58
SRI4	28.92	-22.12	+23.14	SRI4	39.63	-8.96	+7.38
SRI1	29.22	-22.10	+23.13	SRI1	40.12	-9.03	+7.52

Table 10.4: Theoretical uncertainties on PDF+α_S, for the ggF and VBF $H \to \gamma\gamma_d$ signals in different $E_{\mathrm{T}}^{\mathrm{miss}}$ ranges.

ggF			VBF		
m_H = 400 GeV	Nominal	Tuning [%]	m_H = 400 GeV	Nominal	Tuning [%]
SRE1	1479	x	SRE1	804.0	9.378
SRE2	62	x	SRE2	100.0	53.889
SRE3	4	x	SRE3	14.0	99.745
SRI4	0	x	SRI4	11.0	109.469
SRI1	1545	x	SRI1	929.0	10.036
m_H = 600 GeV	Nominal	Tuning [%]	m_H = 600 GeV	Nominal	Tuning [%]
SRE1	1318	7.664	SRE1	847.0	15.656
SRE2	2681	3.875	SRE2	1486.0	5.319
SRE3	1166	5.238	SRE3	880.0	11.305
SRI4	81	24.129	SRI4	114.0	36.905
SRI1	5246	2.197	SRI1	3327.0	2.818
m_H = 800 GeV	Nominal	Tuning [%]	m_H = 800 GeV	Nominal	Tuning [%]
SRE1	487	12.085	SRE1	330.0	10.330
SRE2	753	14.162	SRE2	475.0	31.580
SRE3	2075	11.153	SRE3	1434.0	6.007
SRI4	2275	4.184	SRI4	1745.0	11.024
SRI1	5590	6.493	SRI1	3984.0	2.944
m_H = 1 TeV	Nominal	Tuning [%]	m_H = 1 TeV	Nominal	Tuning [%]
SRE1	250	24.508	SRE1	196.0	19.957
SRE2	367	21.576	SRE2	294.0	15.498
SRE3	800	8.045	SRE3	612.0	11.889
SRI4	4041	4.975	SRI4	3180.0	4.215
SRI1	5458	4.543	SRI1	4282.0	3.429
m_H = 1.5 TeV	Nominal	Tuning [%]	m_H = 1.5 TeV	Nominal	Tuning [%]
SRE1	86	49.402	SRE1	75.0	24.404
SRE2	123	34.175	SRE2	111.0	20.265
SRE3	238	14.909	SRE3	213.0	18.394
SRI4	4465	4.497	SRI4	4451.0	2.055
SRI1	4912	3.006	SRI1	4850.0	2.505
m_H = 2 TeV	Nominal	Tuning [%]	m_H = 2 TeV	Nominal	Tuning [%]
SRE1	34	25.641	SRE1	40.0	46.704
SRE2	38	64.245	SRE2	53.0	25.173
SRE3	112	36.585	SRE3	123.0	30.376
SRI4	4419	1.863	SRI4	4914.0	1.604
SRI1	4603	2.174	SRI1	5130.0	1.463
m_H = 2.5 TeV	Nominal	Tuning [%]	m_H = 2.5 TeV	Nominal	Tuning [%]
SRE1	22	103.553	SRE1	16.0	98.027
SRE2	24	37.500	SRE2	31.0	33.212
SRE3	65	57.089	SRE3	65.0	24.997
SRI4	4242	2.972	SRI4	5303.0	1.101
SRI1	4353	2.364	SRI1	5415.0	1.086
m_H = 3 TeV	Nominal	Tuning [%]	m_H = 3 TeV	Nominal	Tuning [%]
SRE1	10	61.644	SRE1	19.0	132.732
SRE2	13	76.537	SRE2	16.0	81.009
SRE3	32	51.349	SRE3	45.0	43.829
SRI4	4185	3.656	SRI4	5572.0	1.995
SRI1	4240	3.627	SRI1	5652.0	2.287

Table 10.5: Theoretical uncertainties on the PS tuning, for the ggF and VBF $H \to \gamma\gamma_d$ signals in different $E_{\mathrm{T}}^{\mathrm{miss}}$ ranges.

In order to minimize the impact of statistical fluctuations, it was eventually chosen to

m_H (GeV)	ggF, $H \to \gamma\gamma_d$			VBF, $H \to \gamma\gamma_d$		
	PDF+scale (%)		PS (%)	PDF+scale (%)		PS (%)
400	(-14.6	+16.9)	2.20	(-1.54	+1.29)	10.0
600	(-14.7	+16.1)	2.20	(-2.57	+2.04)	2.82
800	(-15.0	+16.2)	6.49	(-2.63	+1.82)	2.94
1000	(-15.4	+16.4)	4.54	(-4.21	+3.41)	3.43
1500	(-16.3	+17.2)	3.01	(-5.28	+4.01)	2.51
2000	(-13.3	+13.3)	2.17	(-6.46	+5.05)	1.46
2500	(-22.1	+23.1)	2.36	(-8.15	+7.38)	1.09
3000	(-22.1	+23.1)	3.60	(-9.03	+7.52)	2.28

Table 10.6: Theoretical uncertainties (%) for ggF and VBF production modes and for different heavy Higgs boson masses computed in the most inclusive SRI1, with $E_T^{miss} > 200$ GeV.

use the uncertainty computed in the most inclusive SR (SRI1) rather than distinguish in $E_{\mathrm{T}}^{\mathrm{miss}}$ bins. The final values implemented in the fit for both the production processes and for the different mass points are summarised in Table 10.6.

10.3 Results

The reinterpretation of the Run-2 mono-photon search is performed including the BSM Higgs boson signals decaying to a photon and a massless dark photon into the simultaneous likelihood fit, with its relative experimental and theoretical uncertainties added as NPs. For the background expectation, the same inputs used in the original mono-photon analysis described in chapter 8 are employed. The simplified-shape fit is performed for each signal using the HistFitter package, to extract exclusion limits at 95% CL on $\sigma\times$ BR($H \to \gamma\ \gamma_d$) by means of a CLs scan. Three sets of results are provided: the limits on the ggF and VBF signal separately, assuming the other production mode to give negligible contribution, and the combined limits accounting simultaneously for both the two contributions. In the latter case, the combination is based on the assumption that the relative contribution from each of the two processes is correctly described by the ratio of the theoretical cross-sections in Table 10.3. The two production modes are therefore merged proportionally to their cross-section and globally normalized to unitary cross-section (in pb) to get the exclusion limits on $\sigma\times$ BR($H \to \gamma\ \gamma_d$):

$$N_{ggF+VBF}^{\sigma=1pb} = \frac{\sigma_{ggF} \times A_{ggF} + \sigma_{VBF} \times A_{VBF}}{\sigma_{ggF} + \sigma_{VBF}} \times L \tag{10.1}$$

An overview of the impact of experimental and theoretical systematics of the signal on the computed observed limit, for the ggF and VBF separate fits, is presented in Table 10.7.

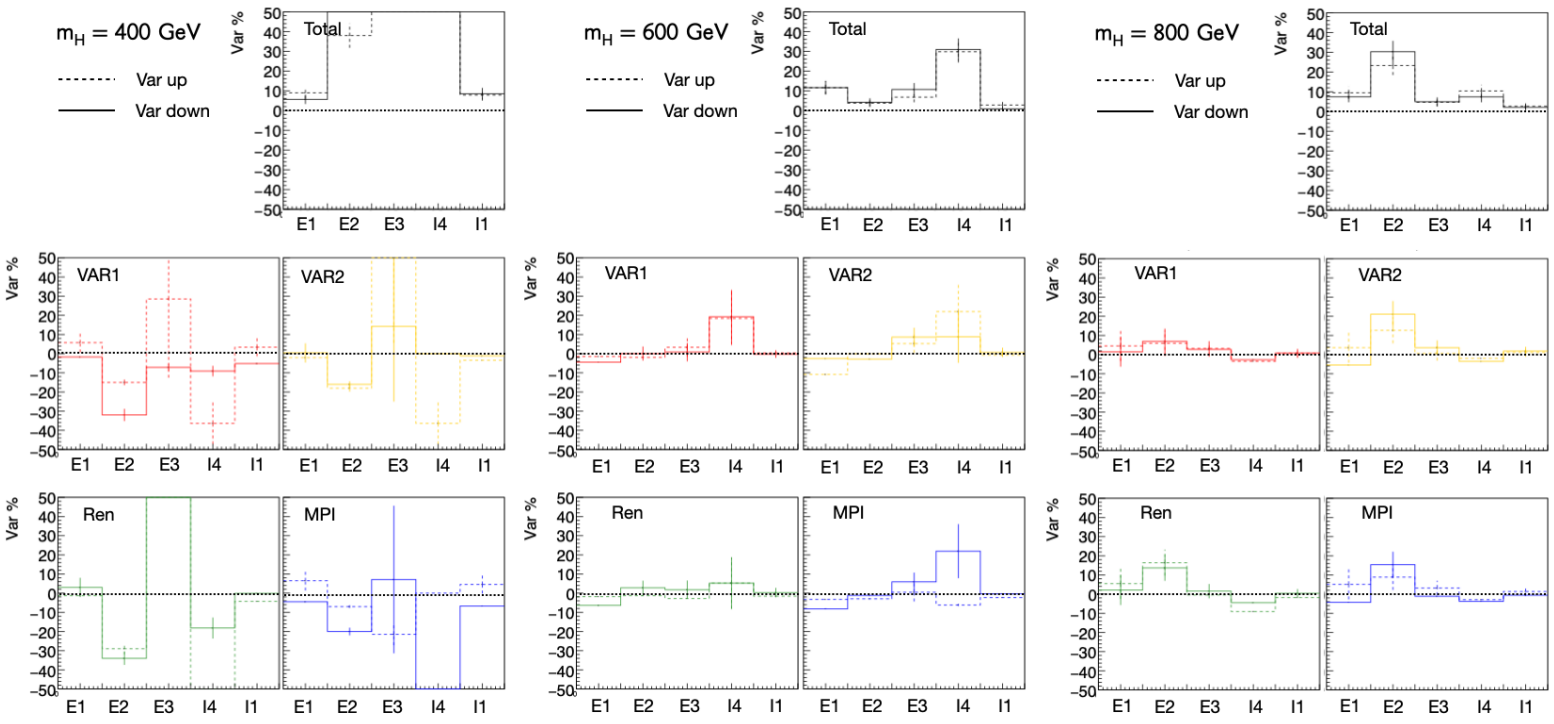


Figure 10.3: Uncertainties on the PS tuning for three signal samples with $m_H = 400$ GeV, $m_H = 600$ GeV and $m_H = 800$ GeV (chosen in order to have a reasonable statistics in all exclusive bins).

95% CL limits	**ggF**		**VBF**	
m_H	**Impact of exp.**	**Impact of theo.**	**Impact of exp.**	**Impact of theo.**
400 GeV	6.6%	5.7%	6.9%	2.1%
600 GeV	2.7%	3.7%	-	-
800 GeV	3.0%	5.0%	4.2%	0.7%
1 TeV	2.0%	4.9%	2.6%	1.2%
1.5 TeV	2.2%	5.2%	2.2%	1.1%
2 TeV	3.0%	3.0%	2.5%	1.2%
2.5 TeV	2.8%	10.8%	2.1%	2.1%
3 TeV	2.5%	11.3%	1.5%	2.2%

Table 10.7: Impact, on the observed limits, of adding experimental uncertainties and theoretical uncertainties of the signal, for different heavy Higgs boson masses and both the production modes.

Results for the Higgs boson signal produced via gluon-gluon Fusion (ggF), vector-boson Fusion (VBF) as well as ggF+VBF are shown in Figure 10.4 for the considered Higgs mass values, and summarized in Table 10.8. The theoretically predicted cross-section of a BSM Higgs boson produced via gluon-gluon Fusion is superimposed assuming the BR($H \to \gamma\ \gamma_d$) = 5%. While the limit on the ggF signal is more stringent than the VBF one up to 1.5 TeV, for higher masses the limits on the VBF production mode start to be competitive due to a combination of higher cross-section and higher acceptance, as shown in Tables 10.3 and 10.2

95% CL Limits	$\sigma_{ggF} \times BR$ [fb]		$\sigma_{VBF} \times BR$ [fb]		$\sigma_{(ggF+VBF)} \times BR$ [fb]	
m_H	**Obs.**	**Exp.** $^{+1\sigma}_{-1\sigma}$	**Obs.**	**Exp.** $^{+1\sigma}_{-1\sigma}$	**Obs.**	**Exp.** $^{+1\sigma}_{-1\sigma}$
400 GeV	23.9	$27.3\ ^{+12.7}_{-7.80}$	44.4	$51.4\ ^{+23.1}_{-15.2}$	24.2	$28.1\ ^{+12.7}_{-8.30}$
600 GeV	5.08	$5.72\ ^{+2.60}_{-1.65}$	9.30	$10.43\ ^{+4.35}_{-2.94}$	5.45	$6.10\ ^{+2.66}_{-1.80}$
800 GeV	3.04	$3.56\ ^{+1.63}_{-1.05}$	4.54	$5.28\ ^{+2.17}_{-1.46}$	3.31	$3.86\ ^{+1.69}_{-1.13}$
1 TeV	1.93	$2.30\ ^{+1.02}_{-0.67}$	2.62	$3.08\ ^{+1.21}_{-0.85}$	2.14	$2.52\ ^{+1.04}_{-0.71}$
1.5 TeV	1.73	$2.05\ ^{+0.92}_{-0.60}$	1.86	$2.17\ ^{+0.87}_{-0.59}$	1.79	$2.09\ ^{+0.84}_{-0.57}$
2 TeV	1.74	$2.05\ ^{+0.88}_{-0.59}$	1.64	$1.92\ ^{+0.78}_{-0.52}$	1.64	$1.93\ ^{+0.76}_{-0.54}$
2.5 TeV	2.06	$2.44\ ^{+1.24}_{-0.77}$	1.48	$1.73\ ^{+0.70}_{-0.49}$	1.50	$1.75\ ^{+0.71}_{-0.48}$
3 TeV	2.11	$2.50\ ^{+1.27}_{-0.79}$	1.42	$1.66\ ^{+0.69}_{-0.46}$	1.44	$1.68\ ^{+0.69}_{-0.46}$

Table 10.8: Observed and expected exclusion limits at 95% CL on $\sigma \times BR(H \to \gamma\gamma_d)$ obtained with the simplified shape-fit for ggF, VBF and combined channels.

In Table 10.9 and Figure 10.5, the same results are also reported in form of exclusion limits on the $BR(H \to \gamma\gamma_d)$, assuming the theoretical cross-sections (Table 10.3). When focusing on the upper limits on BR only, factorizing out the cross-section, the combined limits is observed, as expected, to improve the ggF and VBF only ones. In this interpretation, an exclusion limit exceeding 100% indicates that a BSM Higgs boson signal with the corresponding mass cannot be excluded.

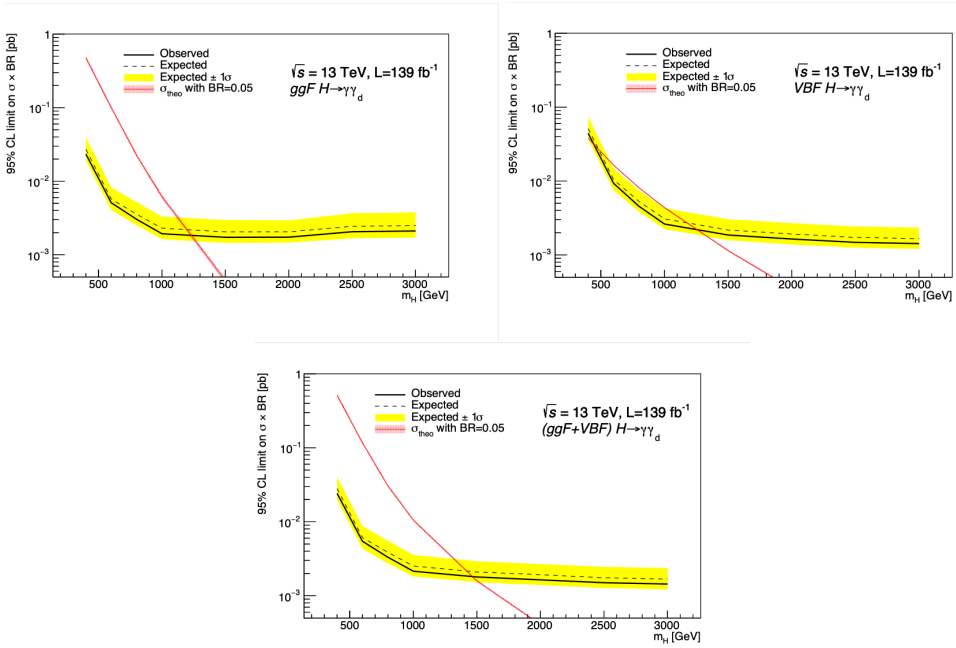


Figure 10.4: Observed and expected exclusion limits at 95% CL on $\sigma\times$ BR($H \to \gamma\, \gamma_d$) as function of the BSM Higgs mass, in gluon-gluon Fusion (top left), VBF (top right) and $ggF{+}VBF$ (bottom) production mode. The yellow band shows the $\pm 1\sigma$ uncertainties. The theoretically predicted cross-section is superimposed assuming the BR($H \to \gamma\, \gamma_d$) = 0.05 and the theoretical production cross-section reported in [17].

95% CL Limits	***BR*** **[%], ggF**		***BR*** **[%], VBF**		***BR*** **[%], ggF+VBF**	
m_H	**Obs.**	**Exp.** $^{+1\sigma}_{-1\sigma}$	**Obs.**	**Exp.** $^{+1\sigma}_{-1\sigma}$	**Obs.**	**Exp.** $^{+1\sigma}_{-1\sigma}$
400 GeV	0.25	0.29	5.86	6.78	0.24	0.27
600 GeV	0.25	0.29	2.84	3.19	0.23	0.26
800 GeV	0.68	0.79	2.80	3.26	0.54	0.63
1 TeV	1.57	1.86	3.00	3.52	1.02	1.20
1.5 TeV	19.39	22.95	8.13	9.49	5.64	6.58
2 TeV	160	188	8.13	9.49	20.1	23.7
2.5 TeV	1157	1372	62.7	73.2	59.1	68.9
3 TeV	6018	7131	173	201	167	195

Table 10.9: Observed and expected exclusion limits at 95% CL on $BR(H \to \gamma\gamma_d)$ obtained with the simplified shape-fit for ggF, VBF and combined channels.

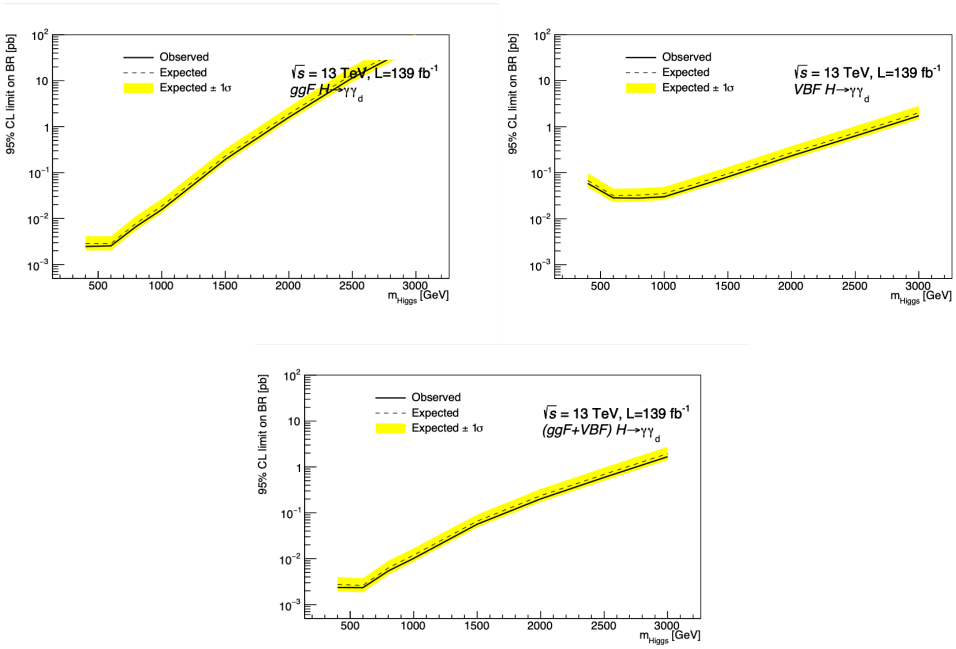


Figure 10.5: Observed and expected exclusion limits at 95% CL on the BR($H \to \gamma\ \gamma_d$), assuming the theoretical cross-section, as a function of the BSM Higgs mass. The results are obtained considering ggF (top left) and VBF (top right) only production modes, or a combination of the two (bottom). The yellow band shows the $\pm 1\sigma$ uncertainties.

Chapter 11

Conclusions and future prospects

In this thesis the final state involving $E_{\mathrm{T}}^{\mathrm{miss}}$ and a photon from ATLAS Run-2 pp collisions data delivered by the LHC has been exploited to address the Dark Matter problem. This study provides a complete picture for this signature, starting from the search for WIMP-like particles and proceeding to the search for a Dark Photon to explore the Dark Sector. To draw the conclusions of this effort, the results described in this thesis are here summarized and framed into the more general context of ATLAS and CMS available results, before mentioning the future prospects in this important field.

Dark Matter search: the Mono-photon analysis The results provided by this analysis in the $E_{\mathrm{T}}^{\mathrm{miss}} + \gamma$ final state (chapter refchap:monophoton)permits to set stringent exclusion limits on the production of DM from the s-channel decay of a Vector or Axial-Vector mediator, in the context of simplified DM models (Section 6.1)). It provides the second best limits among the mono-X searches in ATLAS [16], being second only to the the mono-jet search, as shown in Figure 11.1, but taking advantage of a cleaner final state. These are also the only results in this final state with full Run-2 dataset, as CMS published the results only using the partial data collected in 2015-2016 [228] (Figure 11.1). The search for DM in simplified DM models, either in the mono-X or in the di-jets/di-leptons signature, the latter investigating the mediator decay into SM particles, is an important joint effort, in which all contributions play an important role in building a complete picture and exploiting at best the potentialities of research at LHC, providing competitive limits at low DM masses with respect to Direct Detection searches (Figures 8.30 and 6.3). Moreover, the information about the interaction mediator, which these results can give, have a key role in the understanding of DM properties and its relation with the SM sector.

Dark Photon searches The $H \to \gamma\gamma_d$ decay has been probed in different production modes, starting with a new analysis specifically designed for this process in the ZH Higgs production mode, and concluding with a reinterpretation of the mono-photon analysis in terms of BSM heavy Higgs decays in the ggF and VBF production modes.
The first analysis (Chapter 9) set competitive upper limits on the production of a massless Dark Photon from SM Higgs boson decay, and it's the only one to further extend the reach to higher Dark Photon masses. A summary of the available results at LHC already mentioned in Section 6.2.1 [201, 147, 149, 148], both from the ATLAS and CMS Collaborations, is reported in Table 11.1, while the exclusion limits plots are shown in Figure 11.2 Concerning the reinterpretation of the mono-photon analysis, described in chapter 10, it is a good (though partial) completion of the set of analyses exploring the $H \to \gamma\gamma_d$ decay. It provides a first insight into the potentialities of the search in the ggF production mode, with the limit of being sensitive only to heavy BSM Higgs bosons. Moreover,

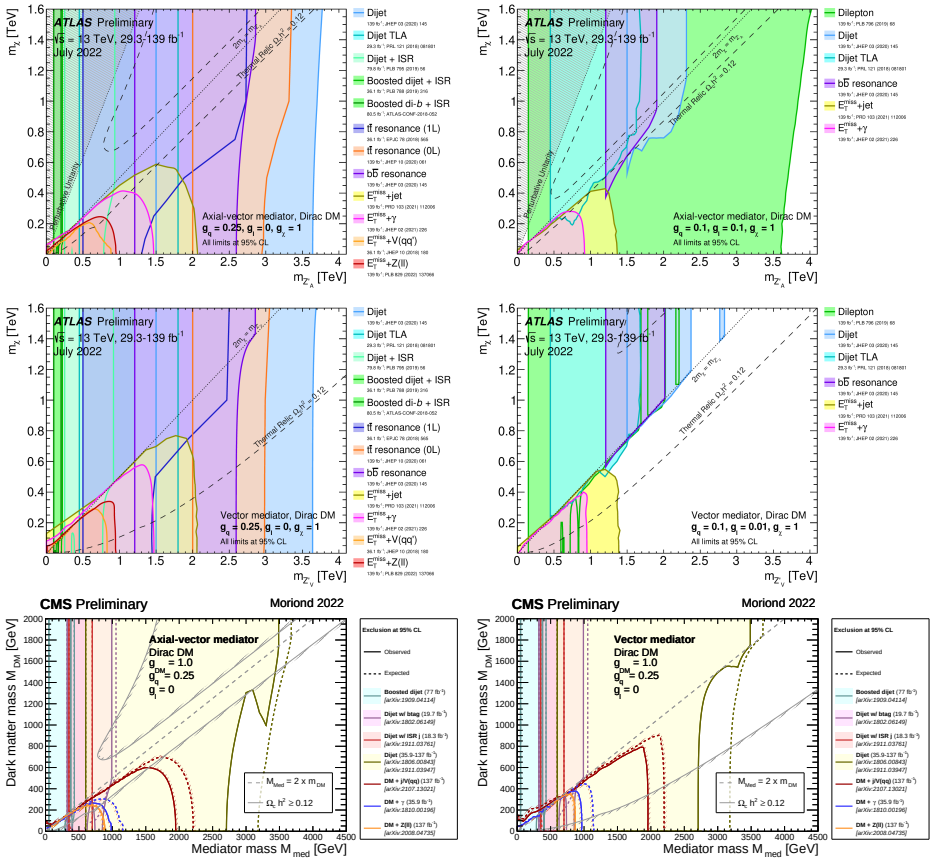


Figure 11.1: Comparison of the limits in $M_{med} - m_{\chi}$ plane set by Run-2 data analyses carried out, in different channels, by the ATLAS [140] (first two rows) and CMS [229] (last row) collaborations, considering four different signal models.

the comparison between VBF analysis results and results provided by the mono-photon reinterpretation in the same production mode (both reproposed in the central plots of Figure 11.2), highlight an improvement of the limits from the latter at $m_H > 1$ TeV. In this region, indeed, the VBF analysis starts loosing in sensitivity, while the mono-photon reinterpretation becomes competitive.

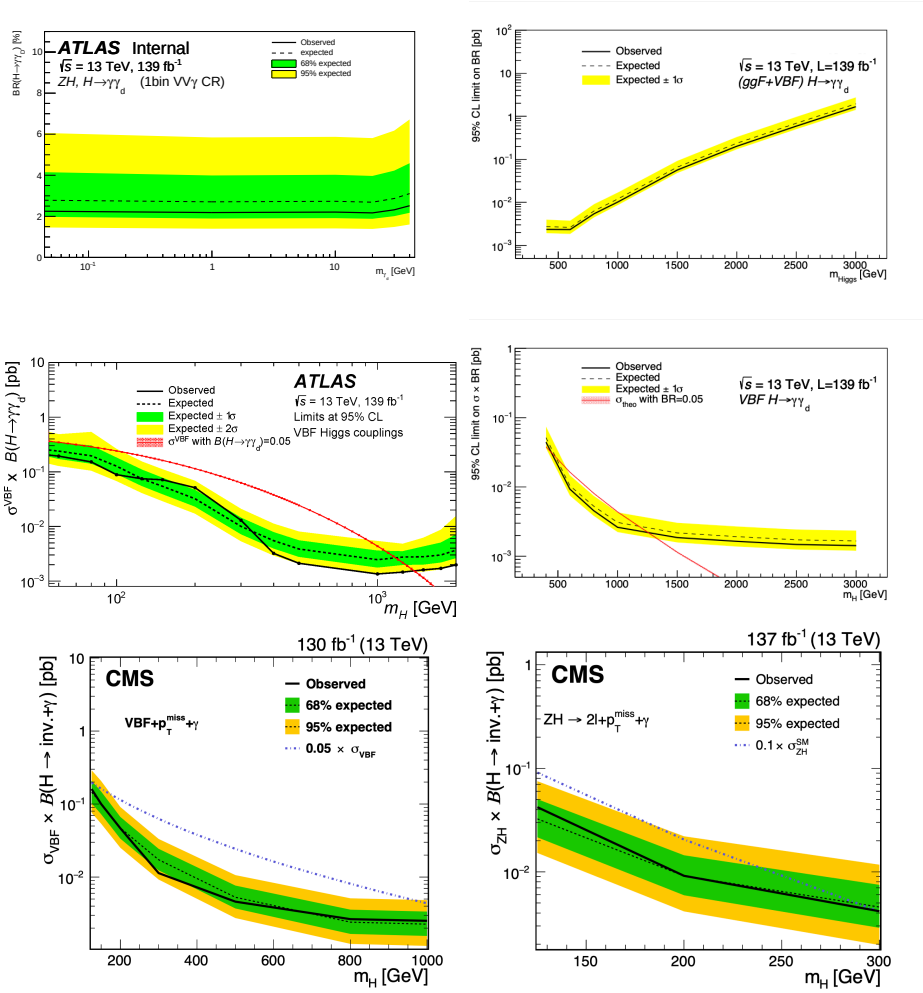


Figure 11.2: In the top plots, exclusion limits at 95% CL on $BR(H \to \gamma\gamma_d)$ as a function of the Dark Photon mass in the ZH production mode [201](left), and as a function of BSM Higgs mass in the ggF+VBF production mode (right). In the central plots, exclusion limits at 95% CL on $\sigma \times BR(H \to \gamma\gamma_d)$ as a function of the Higgs mass, in the VBF Higgs production mode, provided by the VBF analysis [147] (left) and mono-photon reinterpretation (right). The red line shows the expected values under the assumption of BR=5% and theoretical production cross-section reported in [17]. Finally, in the bottom plots, the CMS limits on the $\sigma \times BR(H \to \gamma\gamma_d)$, as a function of the Higgs mass in the VBF [149] and ZH [148] production mode. The theoretical expectations are shown in blue dashed line, assuming a 5% and 1% BR

Experiment	Production mode	Observed	Expected
CMS	VBF	3.5%	2.8%
CMS	ZH	4.6%	3.6%
ATLAS	VBF	1.8%	1.7%
ATLAS	ZH	2.3%	2.8%

Table 11.1: Exclusion limits at 95% CL set by the ATLAS and CMS Collaborations on the branching fraction $BR(H \to \gamma\gamma_d)$ in the VBF [147, 149] and ZH [201, 148] production mode. A massless Dark Photon and the SM Higgs boson are considered.

Future prospects towards Run-3 For the future, further developements in this research field are foreseen, both to conclude Run-2 analyses exploiting at maximum the strengths of the different analyses, and to start analyzing Run-3 data.
A statistical combination of the Run-2 analysis is ongoing, aiming at merging the results in the ZH and VBF production modes for massless Dark Photon and SM Higgs, potentially extending also the ZH analysis to cover BSM Higgs and the VBF one to massive Dark Photons, as well as combining the mono-photon reinterpretation with VBF limits at high BSM Higgs masses.
In addition, to complete the picture, a search in the ggF mode, for a SM Higgs would be desirable. This effort was not pursued using full Run 2 dataset, neither by the ATLAS nor by the CMS collaboration: despite its higher cross-section (49.85 pb) with respect to the other production modes, the too high $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds in the triggers used during Run-2 strongly suppress the signal acceptance, reducing the sensitivity to this process. However, feasibility studies are ongoing, trying to push the sensitivity to ggF production mode in Run-2, despite the challenges of suppressed signal acceptance and extremely high background from γ+jets events mimicking the signal due to sizeable fake $E_{\mathrm{T}}^{\mathrm{miss}}$. In parallel, new triggers are under study and should be made available during Run-3, with lowered $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds combined with low p_T photons and a cut on the transverse mass $m_T(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma) = \sqrt{2p_T^{\gamma} E_T^{miss}(1 - \cos\Delta\phi(\vec{p}_T^{\,miss}, \vec{p}_T^{\,\gamma}))}$, the latter being expected to be peaked around the Higgs mass for the signal. These new triggers could give access to the phase space of interest for the $ggH, H \to \gamma\gamma_d$ signal, thus making this channel potentially competitive in the search for Dark Photon in Run-3.

Appendix A

Additional material for mono-photon analysis CRs

A.1 Data-driven background estimates in the mono-photon analysis CRs

Electrons faking photons

Electron fakes	Stat. for number	Stat. for of events in probe-e CR	Syst. for fake rate	Total unc. fake rate	
		$E_T^{miss} \geq 200$ GeV			
1μ CR	63.85	1.56	2.23	5.36	6.01
2μ CR	1.91	0.26	0.06	0.16	0.31
2e CR	0.54	0.15	0.02	0.05	0.16
PhJetCR	694.45	5.66	25.65	60.23	65.71
		$E_T^{miss} \geq 250$ GeV			
1μ CR	24.91	0.95	0.85	2.08	2.44
2μ CR	0.88	0.17	0.03	0.07	0.19
2e CR	0.44	0.14	0.02	0.04	0.15
PhJetCR	694.45	5.66	25.65	60.23	65.71
		$E_T^{miss} \geq 300$ GeV			
1μ CR	11.16	0.63	0.38	0.93	1.19
2μ CR	0.29	0.09	0.01	0.02	0.09
2e CR	0.30	0.12	0.01	0.03	0.12
PhJetCR	694.45	5.66	25.65	60.23	65.71
		$E_T^{miss} \geq 375$ GeV			
1μ CR	3.84	0.38	0.13	0.32	0.51
2μ CR	0.11	0.05	0.00	0.01	0.05
2e CR	0.23	0.12	0.01	0.02	0.12
PhJetCR	694.45	5.66	25.65	60.23	65.71

Table A.1: Number of electrons faking photons estimated in the CRs in the inclusive E_T^{miss} regions, for full Run-2 data. The uncertainty is expressed in three terms: the first term is the statistical uncertainty related to the number of events found in the probe-e CR; the second and third terms are the statistical and systematic uncertainties related to the electron fake rate, respectively. The total uncertainty, calculated as the combination of the three uncertainties listed above, is also shown.

Jets faking photons

Regions	Mean	Stat.	Tightness syst.	Isolation syst.	c1 stat.	c2 stat.	c3 stat.	R stat.	c1 syst.	c3 syst.	Total syst.	Total error	Total (%) error
1muCR - ISR1	192.0	18.0	42.0	3.0	1.0	4.0	0.0	17.0	7.0	0.0	46.0	49.0	25.5
1muCR - ISR2	102.0	14.0	38.0	2.0	1.0	3.0	0.0	13.0	4.0	0.0	41.0	43.0	42.2
1muCR - ISR3	54.0	10.0	14.0	1.0	1.0	2.0	0.0	11.0	2.0	0.0	18.0	21.0	38.9
1muCR - ISR4	21.0	7.0	5.0	0.0	0.0	1.0	0.0	7.0	1.0	0.0	9.0	11.0	52.4
1muCR - ESR1	92.0	12.0	10.0	2.0	1.0	2.0	0.0	11.0	3.0	0.0	15.0	19.0	20.7
1muCR - ESR2	46.0	9.0	23.0	1.0	1.0	2.0	0.0	8.0	2.0	0.0	25.0	26.0	56.5
1muCR - ESR3	33.0	8.0	9.0	1.0	0.0	1.0	0.0	8.0	1.0	0.0	13.0	15.0	45.5
2muCR - ISR1	16.0	5.0	7.0	0.0	0.0	1.0	0.0	1.0	2.0	0.0	7.0	9.0	56.2
2muCR - ISR2	7.0	4.0	2.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	2.0	5.0	71.4
2muCR - ISR3	3.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	2.0	66.7
2muCR - ISR4	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
2muCR - ESR1	8.0	4.0	9.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	9.0	10.0	125.0
2muCR - ESR2	5.0	3.0	2.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	2.0	4.0	80.0
2muCR - ESR3	2.0	2.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	3.0	150.0
2eCR - ISR1	13.0	5.0	7.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	7.0	9.0	69.2
2eCR - ISR2	11.0	5.0	4.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	4.0	6.0	54.5
2eCR - ISR3	5.0	3.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	3.0	60.0
2eCR - ISR4	1.0	2.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	3.0	300.0
2eCR - ESR1	3.0	3.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	7.0	233.3
2eCR - ESR2	5.0	3.0	3.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	3.0	4.0	80.0
2eCR - ESR3	4.0	2.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	3.0	75.0
gammajetCR	750.0	25.0	168.0	13.0	14.0	52.0	2.0	148.0	1.0	1.0	231.0	232.0	30.9

Table A.2: Final estimates of $jet \to \gamma$ background and their uncertainties.

A.2 Migration uncertainties

The PDF and scale variations can affect the relative acceptances in SR and CRs, potentially leading to a migration of events from the CRs to the SR.
To account for this further theoretical systematic contribution, migration uncertainties have been estimated at reconstruction level for each background process and CR.
For each considered variation, the double ratio acceptance between the SR and the CRs is evaluated:

$$\mathrm{Acc}^{i}_{CR/SR} = \frac{N(CR, i^{th}\ variation)/N(CR, nominal)}{N(SR, i^{th}\ variation)/N(SR, nominal)} \tag{A.1}$$

where the yield variations are obtained by reweighing, using the PMG weights of the MC samples (the same used for theoretical uncertainties estimation, as reported in subsection 8.5.2).
Two contributions are considered:

- Inter-PDF variations: considering CT14nnlo and MMHT2014 NLO as alternative PDF sets;
- QCD factorisation and renormalisation scale variations.

The global uncertainties for each of these contributions are given by the maximum deviation of the double ratio acceptance from the unity:

$$\max(|1 - \mathrm{Acc}^{i}_{CR/SR}|)_{i \in \mathrm{inter-PDF/scale\ var.}} \tag{A.2}$$

		$E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV			$E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV			$E_{\mathrm{T}}^{\mathrm{miss}} > 375$ GeV		
Sample	Region	Scale (%)	PDF (%)	Total (%)	Scale (%)	PDF (%)	Total (%)	Scale (%)	PDF (%)	Total (%)
γ+jets	1μ CR	10.819	1.131	10.88	-	-	-	-	-	-
	2μ CR	-	-	-	-	-	-	-	-	-
	$2e$ CR	-	-	-	-	-	-	-	-	-
	γj CR	5.145	0.914	5.23	6.34	1.61	6.54	6.50	0.63	6.51
Zγ	1μ CR	0.780	0.317	0.84	3.51	0.58	3.56	0.75	1.66	1.82
	2μ CR	1.012	0.448	1.11	2.04	0.84	2.21	3.17	0.84	3.28
	$2e$ CR	0.658	0.594	0.89	1.76	0.52	1.84	4.26	0.45	4.28
	γj CR	13.428	0.298	13.43	14.22	0.70	14.24	16.44	1.32	16.49
Wγ	1μ CR	0.288	0.100	0.30	0.88	0.26	0.92	2.58	1.87	3.19
	2μ CR	7.851	0.199	7.85	-	-	-	-	-	-
	$2e$ CR	7.558	0.378	7.56	-	-	-	-	-	-
	γj CR	7.321	0.252	7.32	5.89	0.73	5.94	2.51	1.39	2.87

Table A.3: Migration uncertainties estimated in the $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV bin. $Z\gamma$ includes both $Z(\to \nu\nu) + \gamma$ and $Z(\to \ell\ell) + \gamma$

The results in an inclusive SR with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV (SRI0) and in SRI2 ($E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV) and SRI5 ($E_{\mathrm{T}}^{\mathrm{miss}} > 375$ GeV) are reported in Table A.3. The total uncertainty is obtained by summing in quadrature the PDF and scale uncertainties. Both $Z(\to \nu\nu) + \gamma$ and $Z(\to \ell\ell) + \gamma$ background processes are included in $Z\gamma$. For each CR the uncertainty related to the dominant background process is lower or close to 1%, except for γ+jets in the γ+jets CR which reaches about 5%. On the other hand, the highest uncertainties are associated to subdominant backgrounds, thus giving a minor contribution.

These uncertainties have been included in the fit as nuisance parameters in the CRs, not correlated among different CRs, and their impact on the total post-fit uncertainty is described in Table A.4 for different $E_{\mathrm{T}}^{\mathrm{miss}}$ thresholds. The total uncertainty increase is only 0.05% in SRI1 and up to 0.2% in the last bin. In addition, the results for the higher bins are likely affected by low statistics and less reliable: applying the uncertainties estimated in SRI1 also to the fit in SRI5, the increase in the total post-fit uncertainty reduces to only 0.02%.

Since the impact has been proved to be negligible, we chose not to include this systematic

Region	Nominal syst. (%)	With SR/CR syst. (%)	Increase (%)
$E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV	3.27	3.32	0.05
$E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV	5.78	5.85	0.07
$E_{\mathrm{T}}^{\mathrm{miss}} > 375$ GeV	9.64	9.81	0.17

Table A.4: Comparison between the post-fit uncertainty withouth including migration systematics (Nominal syst) and including them (with SR/CR syst.)

contribution in our results.

For completeness, in Tables A.5 the migration uncertainties in SRI3 and SRI5, used to derive the results in Table A.4, are also reported.

Sample	Region	Scale	PDF set	Total
γ+jets	1μ CR	-	-	-
	2μ CR	-	-	-
	$2e$ CR	-	-	-
	γ+jets CR	6.34%	1.61%	6.54%
$Z\gamma$	1μ CR	3.51%	0.58%	3.56%
	2μ CR	2.04%	0.84%	2.21%
	$2e$ CR	1.76%	0.52%	1.84%
	γ+jets CR	14.22%	0.70%	14.24%
$W\gamma$	1μ CR	0.88%	0.26%	0.92%
	2μ CR	-	-	-
	$2e$ CR	-	-	-
	γ+jets CR	5.89%	0.73%	5.94%

Sample	Region	Scale	PDF set	Total
γ+jets	1μ CR	-	-	-
	2μ CR	-	-	-
	$2e$ CR	-	-	-
	γ+jets CR	6.50%	0.63%	6.51%
$Z\gamma$	1μ CR	0.75%	1.66%	1.82%
	2μ CR	3.17%	0.84%	3.28%
	$2e$ CR	4.26%	0.45%	4.28%
	γ+jets CR	16.44%	1.32%	16.49%
$W\gamma$	1μ CR	2.58%	1.87%	3.19%
	2μ CR	-	-	-
	$2e$ CR	-	-	-
	γ+jets CR	2.51%	1.39%	2.87%

Table A.5: Migration uncertainties estimated in the $E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV bin (left) and $E_{\mathrm{T}}^{\mathrm{miss}} > 375$ GeV bin (right).

A.3 Impact of NLO electroweak corrections to $Z + \gamma$ samples

In this appendix, some studies showing the impact of the $Z\gamma$ electro-weak corrections (Figure 8.19) on the final results are shown.
As a first test, the $Z(\nu\nu)\gamma$ and the $Z(\ell\ell)\gamma$ samples were reweighted applying these corrections, and the impact of the reweighting on the results of the background-only fit was studied in the SRI0 (inclusive SR with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV). Table A.6 shows the impact of the EW corrections on the normalisation factors. As expected, the reweighted $Z + \gamma$ samples bring μ_Z closer to 1, but the change is within the uncertainties of the nominal value.

k-factor	EW corr.	no EW corr.
$\mu_{\gamma+jets}$	0.816 ± 0.204	0.816 ± 0.206
μ_W	0.843 ± 0.076	0.843 ± 0.076
μ_Z	1.006 ± 0.070	0.952 ± 0.066

Table A.6: Normalisation factors obtained from a background-only fit performed in the SR defined with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV. The errors shown include both the statistical and systematic uncertainties.

An additional test is performed taking the difference between the nominal and reweighted $Z + \gamma$ samples as a systematic uncertainty, which is included in the likelihood as an additional nuisance parameter. When adding this uncertainty, the total background expectation in the SRI0 after the background-only fit remaing almos unvaried, moving from 12776.48 ± 381.93 to 12778.09 ± 381.86. Indeed, the post-fit impact of this uncertainty is absolutely negligible when compared with the other theoretical uncertainties, as reported in Table A.7. Similar conclusion can be deduced by repeating the same test in a "shape" fit: the comparison between post-fit background yields in the different $E_{\mathrm{T}}^{\mathrm{miss}}$ bins with and without the additional systematic uncertainty from EW corrections (correlated among different bins) is shown in Table A.8

Syst. source	Down (%)	Up (%)
PDF	+0.11	-0.35
EW	-0.02	0.01

Table A.7: Impact of the EW and the PDF theoretical uncertainties on the background estimation in the SR defined with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV using full Run2 data.

	SRE0	SRE1	SRE2	SRE3	SRI4
No EW syst.	7419.99 ± 251.66	3067.17 ± 152.61	1182.60 ± 75.02	678.53 ± 52.96	446.68 ± 43.61
With EW syst.	7420.23 ± 252.67	3068.32 ± 153.36	1182.61 ± 75.05	678.53 ± 52.97	446.66 ± 43.63

Table A.8: Impact of the EW systematic uncertainty on the background yields after a "shape" fit in a SR with $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV using full Run2 data.

A.4 Theoretical uncertainties on all simplified DM signal samples

PDF+α_S

m_χ (GeV)	m_{med} (GeV)	Acceptance (% uncertainty) SRE1		SRE2		SRE3		SRI4		Cross-section (% uncertainty)	
10	10	-2.18	+2.39	-2.36	+2.68	-3.13	+3.60	-3.52	+3.85	-4.27	+5.18
10	15	-2.21	+2.48	-2.44	+2.82	-2.52	+2.80	-3.35	+3.61	-4.43	+3.60
10	25	-2.57	+3.14	-3.02	+3.60	-3.73	+4.44	-3.95	+4.50	-6.79	+7.94
10	100	-2.17	+2.49	-2.60	+3.07	-3.21	+3.79	-3.80	+4.26	-4.27	+5.18
10	200	-1.99	+2.13	-2.44	+2.80	-2.84	+3.23	-3.60	+3.97	-3.30	+4.03
10	300	-1.98	+2.17	-2.39	+2.71	-2.83	+3.16	-3.44	+3.71	-3.07	+3.70
10	400	-1.97	+2.03	-2.32	+2.48	-2.65	+2.83	-3.40	+3.59	-2.92	+3.41
10	500	-2.09	+2.16	-2.39	+2.48	-2.61	+2.71	-3.41	+3.52	-2.93	+3.33
10	600	-2.26	+2.31	-2.49	+2.53	-2.58	+2.57	-3.64	+3.37	-3.02	+3.34
10	700	-2.22	+2.20	-2.42	+2.36	-2.72	+2.69	-3.42	+3.10	-3.06	+3.29
10	800	-2.38	+2.32	-2.66	+2.62	-2.88	+2.82	-3.90	+3.80	-3.15	+3.31
10	900	-2.54	+2.45	-2.87	+2.81	-2.98	+2.90	-3.97	+3.55	-3.20	+3.29
10	1000	-2.59	+2.47	-2.77	+2.63	-3.51	+3.38	-4.10	+3.36	-3.27	+3.30
10	1100	-2.63	+2.47	-2.83	+2.65	-2.91	+2.69	-4.27	+3.73	-3.41	+3.38
10	1200	-2.80	+2.62	-3.08	+2.65	-3.58	+3.10	-4.89	+3.61	-3.76	+3.49
10	1300	-3.43	+2.74	-3.28	+2.84	-3.87	+3.20	-5.23	+3.77	-4.02	+3.51
10	1400	-3.37	+2.69	-3.93	+3.02	-3.94	+3.32	-5.27	+3.92	-4.25	+3.54
10	1500	-3.81	+2.90	-3.71	+3.04	-4.11	+3.09	-5.89	+4.04	-4.42	+3.60
10	1600	-3.52	+2.72	-4.23	+3.11	-4.66	+3.43	-5.87	+4.13	-4.70	+3.63
10	1700	-3.81	+3.01	-4.52	+3.15	-4.86	+3.46	-6.45	+4.33	-4.94	+3.64
25	10	-2.04	+2.18	-2.32	+2.61	-2.85	+3.18	-3.53	+3.76	-3.70	+4.47
25	45	-2.12	+2.39	-2.58	+2.98	-3.51	+4.12	-3.51	+3.79	-4.19	+5.07
50	10	-1.96	+2.15	-2.37	+2.62	-2.82	+3.10	-3.52	+3.74	-3.91	+4.76
50	95	-2.02	+2.15	-2.34	+2.60	-2.90	+3.24	-3.41	+3.65	-3.42	+4.14
50	100	-2.09	+2.25	-2.33	+2.65	-2.89	+3.26	-3.55	+3.87	-3.58	+4.34
50	105	-2.14	+2.33	-2.26	+2.58	-2.95	+3.40	-3.58	+3.95	-3.91	+4.76
50	115	-2.15	+2.35	-2.46	+2.87	-3.07	+3.55	-3.40	+3.74	-3.96	+4.81
100	10	-2.05	+2.13	-2.43	+2.52	-2.84	+2.98	-3.72	+3.85	-2.96	+3.40
100	100	-2.07	+2.15	-2.44	+2.56	-2.56	+2.63	-3.57	+3.69	-2.96	+3.40
100	195	-1.97	+2.05	-2.40	+2.59	-2.77	+2.98	-3.40	+3.56	-3.03	+3.58
100	200	-1.98	+2.16	-2.22	+2.41	-2.60	+2.81	-3.47	+3.69	-3.03	+3.59
100	205	-1.96	+2.06	-2.31	+2.57	-2.80	+3.08	-3.54	+3.81	-3.10	+3.72
100	215	-1.99	+2.16	-2.47	+2.81	-2.71	+3.02	-3.24	+3.44	-3.14	+3.80
150	200	-2.15	+2.18	-2.40	+2.42	-2.77	+2.81	-3.61	+3.56	-2.97	+3.28
150	295	-2.08	+2.17	-2.42	+2.55	-2.70	+2.81	-3.64	+3.72	-2.96	+3.39
150	300	-1.99	+2.08	-2.40	+2.56	-2.88	+3.08	-3.55	+3.76	-2.93	+3.38
150	305	-2.02	+2.13	-2.39	+2.58	-2.62	+2.76	-3.55	+3.69	-2.96	+3.44
150	315	-2.04	+2.20	-2.30	+2.45	-2.76	+2.99	-3.57	+3.84	-2.92	+3.42
200	395	-2.08	+2.12	-2.60	+2.68	-2.62	+2.63	-3.48	+3.46	-3.00	+3.31
200	400	-2.23	+2.29	-2.37	+2.42	-2.69	+2.74	-3.60	+3.59	-2.95	+3.29
200	405	-2.14	+2.19	-2.58	+2.72	-2.83	+2.95	-3.57	+3.68	-2.98	+3.35
200	415	-2.12	+2.19	-2.42	+2.51	-2.77	+2.90	-3.35	+3.44	-2.94	+3.34
250	515	-2.22	+2.25	-2.52	+2.57	-2.81	+2.87	-3.79	+3.80	-3.01	+3.30
300	615	-2.43	+2.42	-2.64	+2.63	-2.88	+2.87	-3.84	+3.66	-3.08	+3.28
350	715	-2.59	+2.55	-2.79	+2.75	-3.06	+3.03	-4.21	+3.66	-3.21	+3.35
400	815	-2.69	+2.60	-2.90	+2.81	-3.28	+3.22	-4.10	+3.72	-3.31	+3.43
450	915	-2.92	+2.82	-2.98	+2.81	-3.36	+3.24	-4.73	+3.92	-3.46	+3.48
500	1015	-2.92	+2.75	-3.11	+2.93	-3.60	+3.06	-4.93	+4.09	-3.67	+3.50

Table A.9: PDF and scale uncertainties, in %, on the acceptance and the cross section of simplified dark matter models.

Tuning uncertainties

m_χ	m_{med}	Nominal yield	Uncertainty (%)
10	10	1350	9.152
10	15	1271	5.522
10	25	716	21.027
10	100	1499	13.593
10	200	1855	8.929
10	300	2276	7.543
10	400	2586	7.462
10	500	2713	5.190
10	600	2960	3.966
10	700	3126	8.723
10	800	3240	7.757
10	900	3275	6.085
10	1000	3411	5.970
10	1100	3440	5.732
10	1200	3544	5.509
10	1300	3546	4.852
10	1400	3649	5.978
10	1500	3646	4.320
10	1600	3702	5.111
10	1700	3680	5.922
25	10	1820	9.631
25	45	1601	11.654
50	10	2175	7.308
50	95	1837	11.193
50	100	1736	9.357
50	105	1631	9.804
50	115	1538	6.378
100	10	2553	8.224
100	100	2574	6.465
100	195	2310	7.147
100	200	2204	9.490
100	205	2139	6.179
100	215	2038	5.818
150	200	2833	6.977
150	295	2609	4.876
150	300	2570	2.988
150	305	2444	9.336
150	315	2498	8.388
200	395	2793	7.820
200	400	2872	6.370
200	405	2808	6.148
200	415	2744	5.761
250	515	3040	9.584
300	615	3101	6.573
350	715	3145	11.219
400	815	3287	8.314
450	915	3445	3.761
500	1015	3492	8.306

Table A.10: MC tuning uncertainties, in %, for simplified DM models.

A.5 Anti-SF in CRs

CR1mu	e-antiSF	μ-antiSF	τ-antiSF	Total antiSF
$Z(\nu\nu)+\gamma$	1.0000	-	0.9999	0.9999
$Z\gamma$	1.0021	-	0.9989	1.0010
$W\gamma$	1.0003	-	0.9996	0.9999
γ+jets	1.0007	-	0.9998	1.0007

CR2mu	e-antiSF	μ-antiSF	τ-antiSF	Total antiSF
$Z\gamma$	1.0001	-	0.9993	0.9994
$W\gamma$	1.0001	-	0.9996	0.9997

CR2el	e-antiSF	μ-antiSF	τ-antiSF	Total antiSF
$Z\gamma$	-	1.0001	0.9999	1.0
$W\gamma$	-	1.0003	0.9997	1.0
γ+jets	-	1.0002	0.9998	1.0

CRphjet	e-antiSF	μ-antiSF	τ-antiSF	Total antiSF
$Z(\nu\nu)+\gamma$	1.0000	1.0000	0.9999	0.9999
$Z\gamma$	1.0209	1.0437	0.9978	1.0624
$W\gamma$	1.0216	1.0087	0.9992	1.0295
γ+jets	1.0007	1.0003	0.9998	1.0007

Table A.11: e-antiSF, μ-antiSF, τ-antiSF and their product in the CRs, evaluated at $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV. The $\mu(e)$-antiSF are not applied in CR1mu, CR2mu (CR2el)

A.6 Ranking plots of the systematics

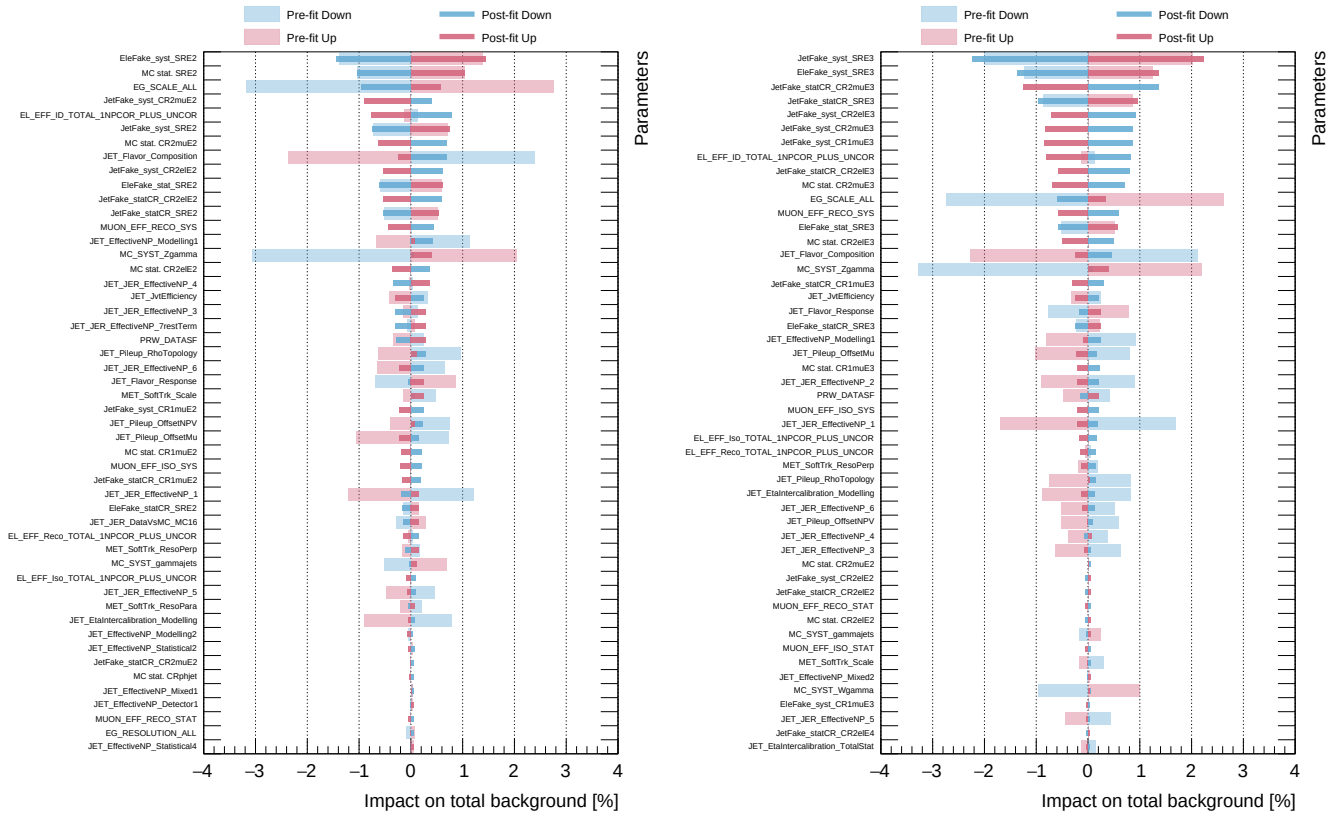


Figure A.1: Breakdown of dominant systematics uncertainties before and after the background-only shape-fit in SRE1 and SRE2.

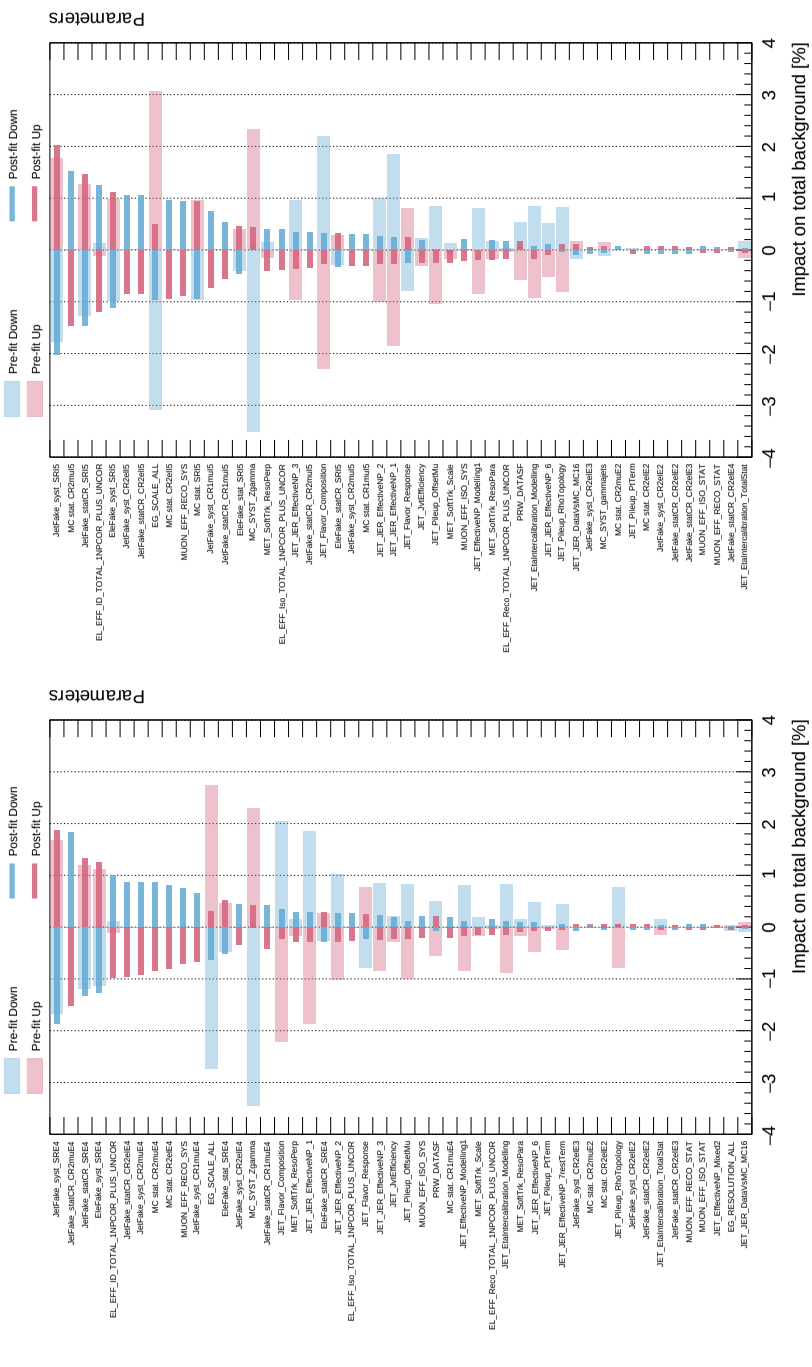


Figure A.2: Breakdown of dominant systematics uncertainties before and after the background-only shape-fit in SRE3 and SRI4.

A.7 Fiducial acceptance and efficiency for all signal samples

m_χ [GeV]	m_{med} [GeV]	Acc [%] $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV	Acc [%] $E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV	Acc [%] $E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV	Acc [%] $E_{\mathrm{T}}^{\mathrm{miss}} > 375$ GeV
10	10	13.90	7.00	3.70	1.60
10	15	13.40	6.80	3.50	1.50
10	25	7.30	3.40	1.90	0.80
10	100	16.00	7.50	3.80	1.70
10	200	19.90	9.90	5.40	2.20
10	300	23.80	12.80	7.10	3.40
10	400	27.10	15.40	9.20	4.60
10	500	28.50	17.10	10.50	5.60
10	600	31.10	19.60	12.30	6.10
10	700	32.30	20.30	13.00	7.10
10	800	34.00	22.20	14.50	8.20
10	900	34.40	22.80	15.20	8.50
10	1000	35.60	23.80	15.80	9.10
10	1100	36.00	24.10	16.10	9.50
10	1200	36.80	24.90	17.50	10.10
10	1300	37.40	26.30	18.50	10.50
10	1400	38.40	26.50	18.60	11.60
10	1500	37.60	26.40	19.40	12.20
10	1600	38.70	28.20	20.30	12.60
10	1700	38.70	27.50	19.90	12.00
25	10	19.00	10.60	5.80	2.70
25	45	16.80	8.70	4.50	2.10
50	10	22.90	12.90	7.40	3.40
50	95	19.40	10.00	5.50	2.50
50	100	18.60	9.60	5.00	2.20
50	105	17.60	8.90	4.80	2.00
50	115	16.00	8.10	4.20	1.90
100	10	26.90	16.20	9.80	5.20
100	100	27.30	16.40	10.30	5.20
100	195	24.30	13.60	8.00	3.60
100	200	23.40	13.10	7.60	3.30
100	205	22.70	12.70	7.20	3.50
100	215	21.50	11.10	6.20	2.70
150	200	29.50	17.90	11.20	5.70
150	295	27.10	16.30	10.00	4.70
150	300	26.80	15.30	9.10	4.20
150	305	25.60	14.90	8.90	4.20
150	315	25.80	14.20	8.50	4.10
200	395	29.20	17.80	11.30	5.90
200	400	30.20	18.00	10.90	5.60
200	405	29.30	17.40	10.70	5.20
200	415	28.70	16.80	10.50	5.10
250	515	31.80	19.50	12.40	6.60
300	615	32.20	20.20	12.90	7.00
350	715	33.00	21.70	14.10	8.00
400	815	34.30	23.30	15.60	9.00
450	915	36.00	24.70	16.70	9.70
500	1015	36.80	25.20	17.20	10.00

Table A.12: Fiducial acceptances for the dark matter simplified models in the four inclusive signal regions.

m_χ [GeV]	m_{med} [GeV]	Acc [%] $200 < E_{\mathrm{T}}^{\mathrm{miss}} < 250$ GeV	Acc [%] $250 < E_{\mathrm{T}}^{\mathrm{miss}} < 300$ GeV	Acc [%] $300 < E_{\mathrm{T}}^{\mathrm{miss}} < 375$ GeV
10	10	6.90	3.30	2.10
10	15	6.60	3.30	2.00
10	25	3.90	1.50	1.10
10	100	8.40	3.70	2.10
10	200	10.00	4.50	3.20
10	300	11.00	5.70	3.70
10	400	11.70	6.20	4.60
10	500	11.50	6.60	4.90
10	600	11.50	7.30	6.20
10	700	12.10	7.30	5.90
10	800	11.80	7.70	6.20
10	900	11.60	7.60	6.60
10	1000	11.90	8.00	6.70
10	1100	12.00	7.90	6.70
10	1200	11.90	7.40	7.30
10	1300	11.10	7.70	8.00
10	1400	11.90	7.90	7.00
10	1500	11.20	7.00	7.20
10	1600	10.50	7.80	7.70
10	1700	11.20	7.50	7.90
25	10	8.40	4.80	3.00
25	45	8.00	4.20	2.40
50	10	10.00	5.50	4.00
50	95	9.40	4.50	3.00
50	100	9.00	4.60	2.80
50	105	8.80	4.10	2.80
50	115	8.00	3.80	2.30
100	10	10.70	6.40	4.60
100	100	11.00	6.10	5.10
100	195	10.70	5.60	4.40
100	200	10.30	5.50	4.30
100	205	10.00	5.50	3.70
100	215	10.40	4.90	3.60
150	200	11.60	6.70	5.50
150	295	10.80	6.30	5.30
150	300	11.50	6.10	5.00
150	305	10.70	6.00	4.70
150	315	11.70	5.70	4.40
200	395	11.40	6.40	5.40
200	400	12.30	7.10	5.30
200	405	11.90	6.70	5.50
200	415	11.90	6.30	5.40
250	515	12.30	7.10	5.80
300	615	12.00	7.30	5.90
350	715	11.30	7.60	6.10
400	815	11.10	7.70	6.60
450	915	11.30	8.00	7.00
500	1015	11.60	8.00	7.20

Table A.13: Fiducial acceptances for the dark matter simplified models in the exclusive signal regions.

m_χ [GeV]	m_{med} [GeV]	$\epsilon_{E_T^{miss}>200\text{GeV}}$	$\epsilon_{E_T^{miss}>250\text{GeV}}$	$\epsilon_{E_T^{miss}>300\text{GeV}}$	$\epsilon_{E_T^{miss}>375\text{GeV}}$
10	10	0.813	0.814	0.811	0.812
10	15	0.806	0.779	0.800	0.800
10	25	0.959	0.941	0.842	0.750
10	100	0.756	0.773	0.789	0.706
10	200	0.819	0.848	0.833	0.864
10	300	0.815	0.812	0.831	0.794
10	400	0.815	0.825	0.804	0.739
10	500	0.839	0.830	0.800	0.750
10	600	0.823	0.801	0.797	0.803
10	700	0.830	0.828	0.823	0.775
10	800	0.824	0.806	0.800	0.756
10	900	0.846	0.833	0.822	0.812
10	1000	0.834	0.832	0.848	0.813
10	1100	0.842	0.851	0.863	0.853
10	1200	0.840	0.843	0.829	0.851
10	1300	0.834	0.829	0.816	0.829
10	1400	0.820	0.826	0.828	0.793
10	1500	0.859	0.860	0.825	0.787
10	1600	0.829	0.809	0.803	0.802
10	1700	0.840	0.847	0.829	0.850
25	10	0.795	0.736	0.724	0.667
25	45	0.804	0.793	0.822	0.762
50	10	0.808	0.775	0.784	0.794
50	95	0.845	0.860	0.855	0.840
50	100	0.823	0.823	0.840	0.864
50	105	0.784	0.764	0.750	0.750
50	115	0.831	0.802	0.810	0.737
100	10	0.851	0.827	0.816	0.750
100	100	0.813	0.793	0.748	0.731
100	195	0.823	0.816	0.800	0.833
100	200	0.825	0.817	0.803	0.848
100	205	0.815	0.780	0.778	0.714
100	215	0.823	0.847	0.823	0.852
150	200	0.837	0.832	0.830	0.807
150	295	0.841	0.822	0.810	0.830
150	300	0.825	0.837	0.835	0.881
150	305	0.848	0.826	0.798	0.810
150	315	0.818	0.824	0.788	0.756
200	395	0.846	0.837	0.814	0.780
200	400	0.815	0.822	0.835	0.821
200	405	0.823	0.822	0.822	0.846
200	415	0.812	0.810	0.781	0.765
250	515	0.805	0.795	0.782	0.758
300	615	0.832	0.842	0.837	0.814
350	715	0.852	0.843	0.844	0.812
400	815	0.854	0.833	0.827	0.800
450	915	0.836	0.818	0.826	0.804
500	1015	0.840	0.837	0.843	0.840

Table A.14: Fiducial efficiencies for the dark matter simplified models in the four inclusive signal regions.

m_χ [GeV]	m_{med} [GeV]	$\epsilon_{200<E_{\mathrm{T}}^{\mathrm{miss}}<250\ \mathrm{GeV}}$	$\epsilon_{250<E_{\mathrm{T}}^{\mathrm{miss}}<300\ \mathrm{GeV}}$	$\epsilon_{300<E_{\mathrm{T}}^{\mathrm{miss}}<375\ \mathrm{GeV}}$
10	10	0.812	0.818	0.810
10	15	0.833	0.758	0.800
10	25	0.974	1.000	0.909
10	100	0.750	0.784	0.810
10	200	0.790	0.867	0.812
10	300	0.809	0.807	0.865
10	400	0.803	0.855	0.848
10	500	0.843	0.879	0.857
10	600	0.861	0.808	0.774
10	700	0.826	0.836	0.864
10	800	0.847	0.818	0.871
10	900	0.871	0.842	0.848
10	1000	0.832	0.800	0.896
10	1100	0.817	0.835	0.881
10	1200	0.832	0.892	0.808
10	1300	0.847	0.870	0.787
10	1400	0.798	0.835	0.886
10	1500	0.848	0.957	0.889
10	1600	0.886	0.833	0.805
10	1700	0.821	0.893	0.797
25	10	0.869	0.750	0.767
25	45	0.825	0.762	0.875
50	10	0.840	0.782	0.775
50	95	0.830	0.867	0.867
50	100	0.822	0.804	0.857
50	105	0.784	0.805	0.714
50	115	0.850	0.816	0.870
100	10	0.888	0.844	0.891
100	100	0.836	0.869	0.784
100	195	0.832	0.857	0.773
100	200	0.835	0.836	0.767
100	205	0.850	0.800	0.838
100	215	0.808	0.857	0.806
150	200	0.845	0.851	0.836
150	295	0.880	0.841	0.792
150	300	0.809	0.852	0.780
150	305	0.879	0.867	0.787
150	315	0.803	0.877	0.818
200	395	0.860	0.891	0.852
200	400	0.797	0.803	0.830
200	405	0.815	0.836	0.800
200	415	0.815	0.873	0.778
250	515	0.821	0.831	0.810
300	615	0.817	0.849	0.864
350	715	0.867	0.842	0.885
400	815	0.892	0.857	0.864
450	915	0.876	0.800	0.857
500	1015	0.845	0.825	0.833

Table A.15: Fiducial efficiencies for the dark matter simplified models in the exclusive signal regions.

A.8 Background only fit with Asimov dataset

The behaviour of the complete fit in CR+SR, performed to provide the exclusion limits, has been tested with a simplified-shape fit on Asimov data, before the fit to actual data, as an additional cross-check. Results are shown in the following figures: the pull-plot of nuisance parameters in figure A.3, and the correlation matrix in figure A.4.

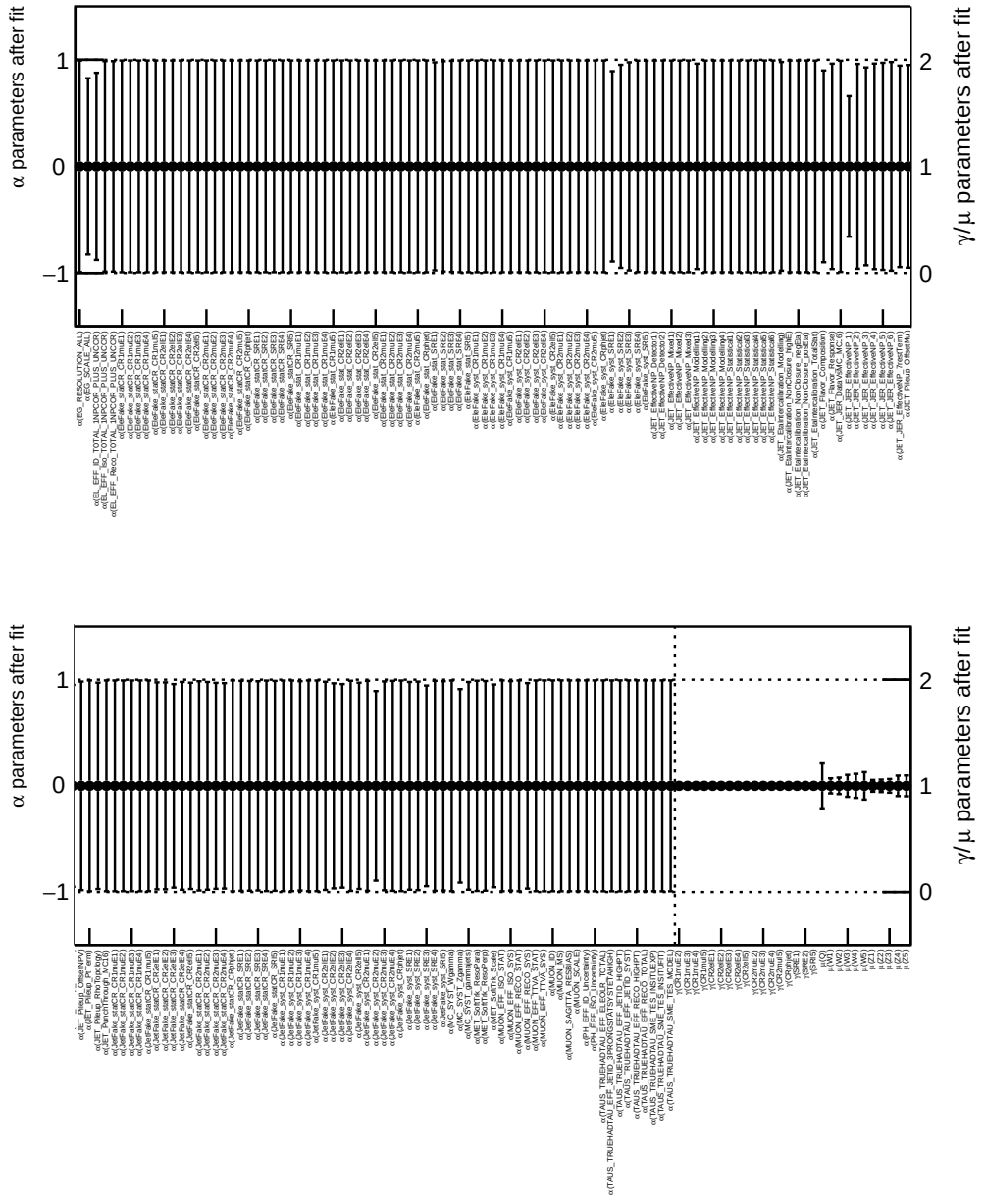


Figure A.3: Post-fit nuisance parameter values and errors in the shape fit on Asimov data. The plot is splitted into two rows to make the NPs more readable.

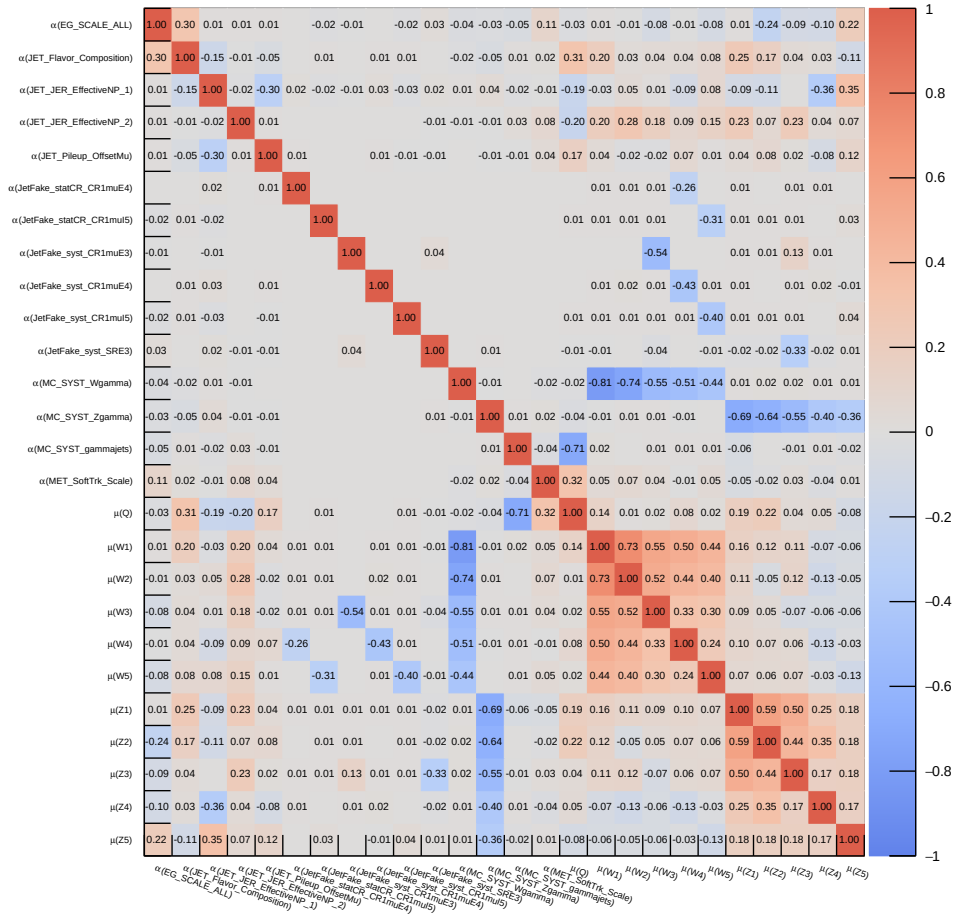


Figure A.4: Correlation matrix between the various sources of systematic uncertainty that are taken into account in the shape fit on Asimov data. Only systematics with a correlation higher that 0.25 are shown.

Appendix B

Additional material for the Dark Photon analysis in ZH production mode

B.1 Fake $E_{\mathrm{T}}^{\mathrm{miss}}$ ABCD optimisation based on Zgamma strong + Z strong

This appendix reports, in Table B.1, the results of the optimisation process described in section 9.3.1, performed including both $Z\gamma$ strong and Z strong MC samples. The best pair is the same obtained with $Z\gamma$ strong only.

Var	R (ee)	R ($\mu\mu$)	FoM (ee)	FoM ($\mu\mu$)	$\chi^2_R(ee)$	$\chi^2_R(\mu\mu)$
$\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$	1.1970 ± 0.0334	1.2347 ± 0.0323	9.4582	5.9460	0.4271	0.7215
$\Delta\phi(\vec{E}_T^{miss} + \vec{p}_T^{ll}, \vec{p}_T^{ll})$	0.8210 ± 0.0305	0.7368 ± 0.0251	10.7970	2.9066	0.3308	1.3446
$\lvert\vec{E}_T^{miss} + \vec{p}_T^{lly}\rvert / E_T^{miss}$	1.7367 ± 0.0547	1.6724 ± 0.0472	0.7342	1.1412	5.6291	3.7866
$\lvert\vec{E}_T^{miss} + \vec{p}_T^{lly}\rvert / \lvert E_T^{miss} + p_T^{lly}\rvert$	0.7741 ± 0.0243	0.6850 ± 0.0207	1.7423	1.6150	2.1652	2.5194
$\Delta\phi(\vec{E}_T^{miss} + \vec{p}_T^{lly}, nearest(\vec{E}_T^{miss}, \vec{p}_T^{lly}))$	0.9373 ± 0.0290	0.8675 ± 0.0243	4.2592	4.3970	0.9262	0.9774
$min_{obj}\{\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{obj.})\}$	1.4434 ± 0.0475	0.9178 ± 0.0269	7.2472	20.7160	0.4626	0.1767
$\lvert\vec{p}_T^{lly}\rvert$	1.3443 ± 0.0325	1.4779 ± 0.0321	3.1301	2.7168	1.3553	1.7760

Table B.1: Results of the ABCD tests for the best pair of variables ($E_{\mathrm{T}}^{\mathrm{miss}}$ variable as a first variable), including both $Z\gamma$ strong and Z strong. The error shown corresponds to the propagation of statistical uncertainties.

B.2 Including $E_{\mathrm{T}}^{\mathrm{miss}}$ significance in "fake" $E_{\mathrm{T}}^{\mathrm{miss}}$ ABCD method

In order to combine the discrimination power of $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance, the possibility to include a $E_{\mathrm{T}}^{\mathrm{miss}}$ significance selection in the high $E_{\mathrm{T}}^{\mathrm{miss}}$ regions, as represented in Figure B.1, have been explored. Figure B.2 shows the evolution of the $E_{\mathrm{T}}^{\mathrm{miss}}$-$\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{lly})$ correlations by changing the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance cut. The correlation and its stability gets worse with the raise of the cut above 3, while it maintains a decent performance for lower selections. Since no improvement in significance is observed with such small cut values on $E_{\mathrm{T}}^{\mathrm{miss}}$ significance, this hypothesis was eventually abandoned.

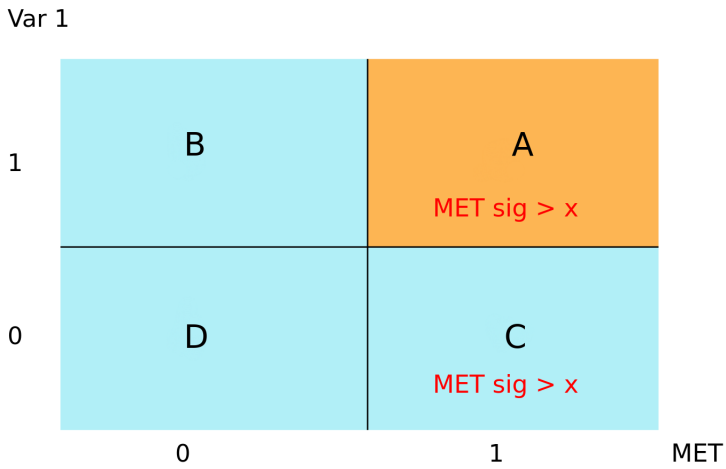


Figure B.1: Scheme of the ABCD regions with the cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ significance.

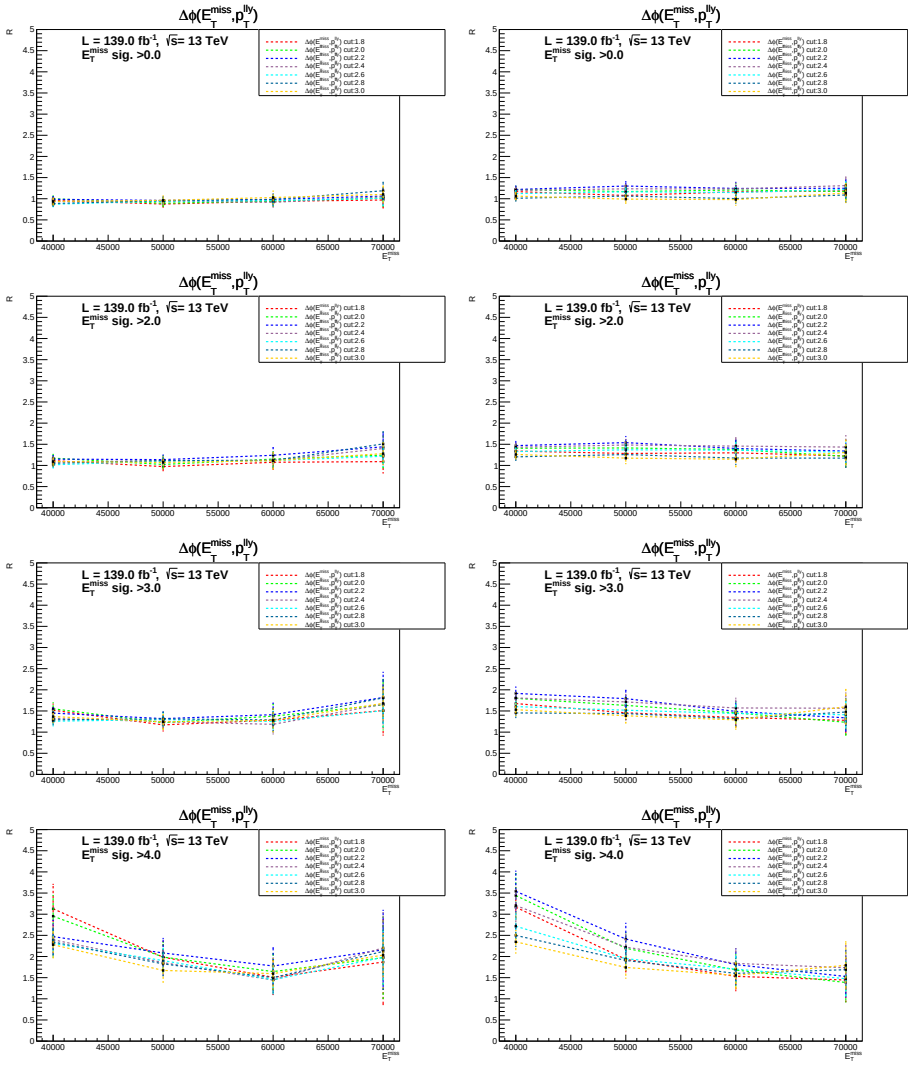


Figure B.2: R stability in the ee (left) and $\mu\mu$ (right) channel, for increasing values (0,2,3,4) of $E_{\mathrm{T}}^{\mathrm{miss}}$ significance cut included in A and C.

B.3 Kinematic distributions in a $2e{+}1\mu$ $VV\gamma$ CR

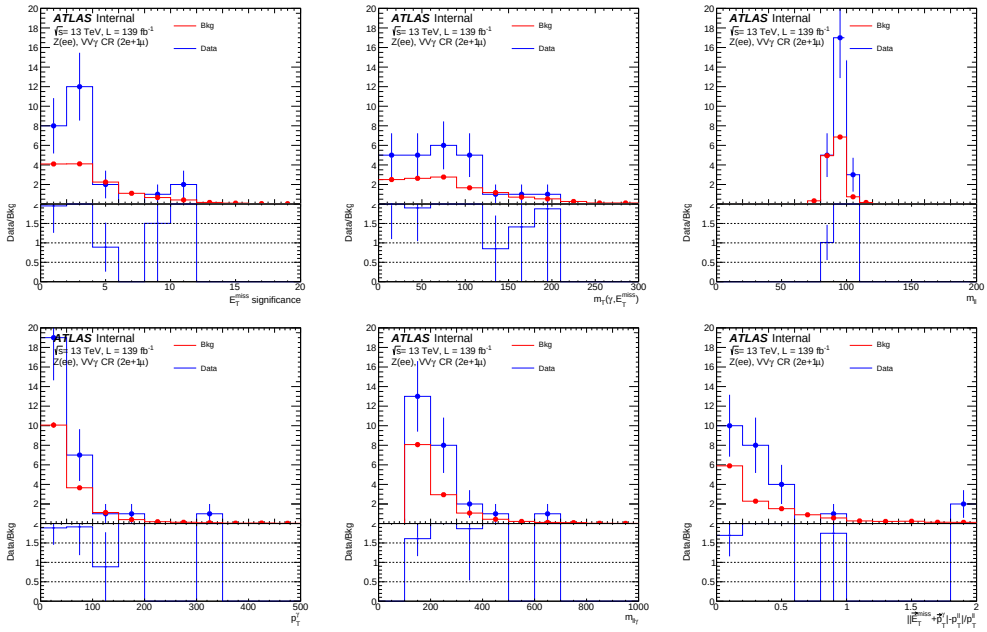


Figure B.3: Shape comparison of relevant variables distributions between data and MC, in the $ee\mu$ channel of the $VV\gamma$ CR.

B.4 Comparison between $ZZ\gamma$ and $WZ\gamma$ processes

The main kinematic distribution show, as expected, good similarities between the two samples. The highest discrepancies are observed in the photon p_{T} distribution. Nevertheless, the analysis cut of $p_{\mathrm{T}}^{\gamma} > 30$ GeV provide very similar acceptances for the two processes, equal to $\sim 80\%$ for the $ZZ\gamma$ and $\sim 77\%$ for the $WZ\gamma$, thus reducing the impact of these differences on the extrapolation of the normalization extracted from $WZ\gamma$ samples to the $ZZ\gamma$ ones.

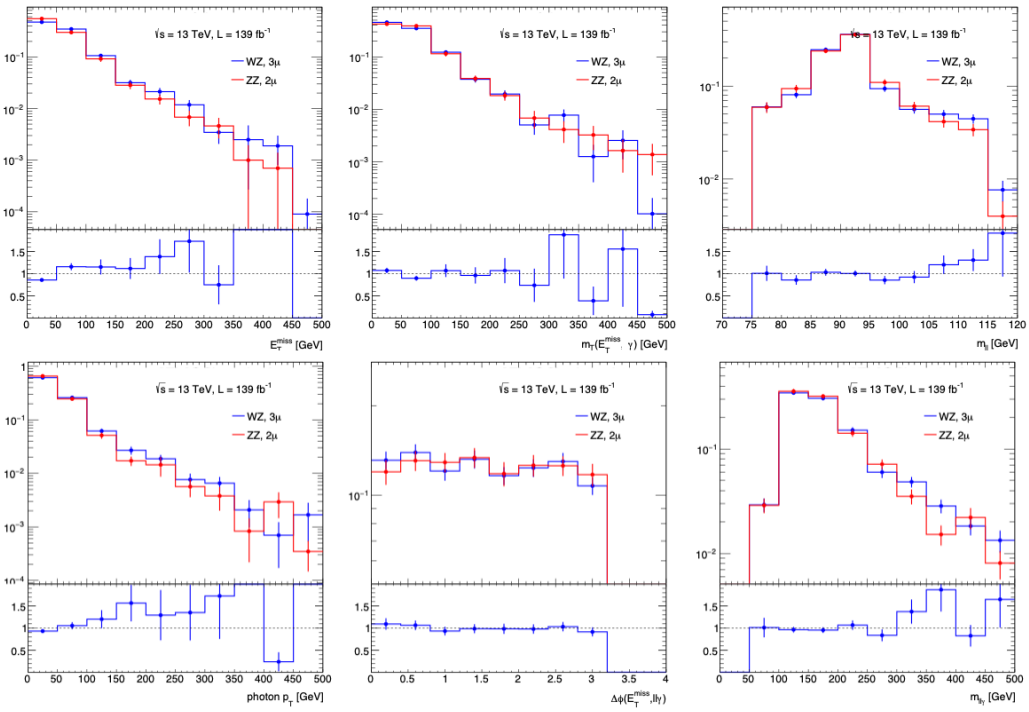


Figure B.4: Comparison of some relevant kinematic distributions in $ZZ\gamma$ and $WZ\gamma$ events.

B.5 Studies about BDT shape for fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background

This appendix summarises some studies aiming at defining an alternative to MC for the BDT shape of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ background.
The main issue of using MC derives from the low statistics of Z strong events, giving rise to non-smooth shape as shown in Figure B.5: the yields per bin width are plotted for the two channels and in both SR and VR, and dips and large uncertainties are observed in the Z strong background for all bins where a disagreement between data and background is observed.

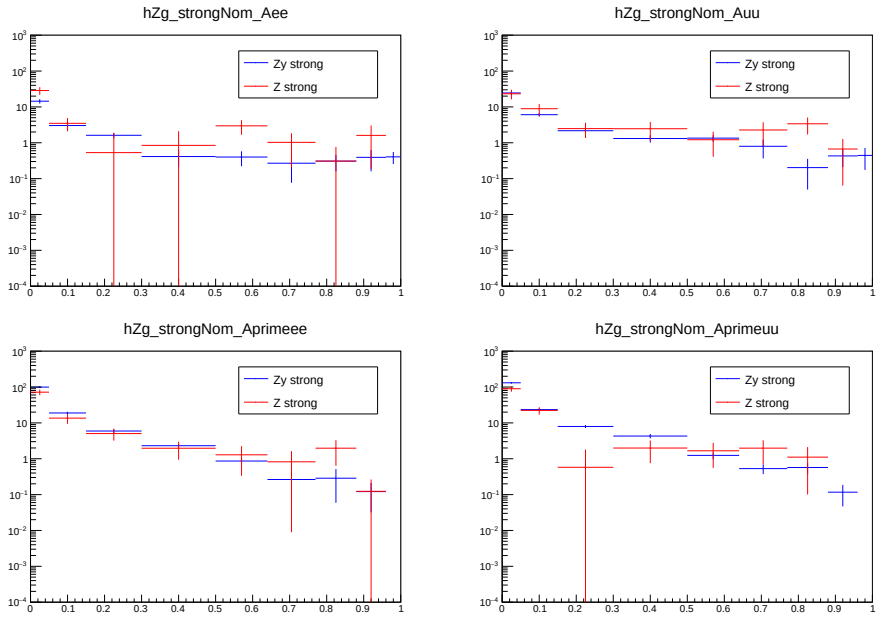


Figure B.5: BDT score distributions in the ee (left) and $\mu\mu$ (right) channel, of $Z\gamma$ strong (blue) and Z strong (red) processes in SR (top) and in the VR (bottom). The yields per bin width are shown, and the uncertainties are statistical from MC.

In tables B.2,B.3,B.4,B.5, the number of MC events (without any reweighting) passing SR or VR selections in each bin and for each MC sample is reported, underlying the difference in the number of MC events between $Z\gamma$ strong and Z strong processes, as well as the very small statistics in the problematic bins, especially from the lowest p_T slices, contributing with the largest weights due to cross-section normalisation.

Sample	Run	0,0.05	0.15	0.30	0.50	0.64	0.77	0.88	0.96	1
$Z\gamma$ strong	700011	465	377	100	53	26	19	10	5	1
$Z\gamma$ strong	700013	1	41	29	5	2	4	1	2	0
$Z\gamma$ strong	700018	104	104	55	44	20	12	8	7	8
$Z\gamma$ strong	700020	0	2	2	1	0	0	1	0	0
Z strong	364114	14	2	6	5	4	1	1	1	0
Z strong	364115	13	3	1	3	1	0	3	1	0
Z strong	364116	11	8	2	7	1	2	0	3	0
Z strong	364117	6	6	3	1	2	1	0	0	0
Z strong	364118	6	1	3	0	1	0	0	0	0
Z strong	364119	5	5	8	0	1	1	5	0	0
Z strong	364120	12	3	1	3	2	0	0	1	0
Z strong	364121	6	2	2	0	1	0	0	0	0
Z strong	364122	11	8	8	3	1	4	2	1	0
Z strong	364123	1	3	1	0	0	0	0	0	0
Z strong	364125	4	2	1	2	0	0	1	0	0
Z strong	364126	5	2	1	0	0	0	0	1	0
Z strong	364127	5	2	0	0	0	0	0	0	0

Table B.2: Number of pure MC events in the ee SR, for each $Z\gamma$ strong and Z strong MC sample, in each BDT bin.

Sample	Run	0,0.05	0.15	0.30	0.50	0.64	0.77	0.88	0.96	1
$Z\gamma$ strong	700012	628	551	200	100	58	43	24	15	1
$Z\gamma$ strong	700013	1	17	10	6	4	2	1	1	1
$Z\gamma$ strong	700019	176	216	110	62	30	27	15	17	3
$Z\gamma$ strong	700020	0	1	1	0	0	0	0	0	0
Z strong	364100	11	9	3	6	2	4	3	0	0
Z strong	364101	9	7	3	2	1	2	2	4	0
Z strong	364102	22	6	8	4	3	8	6	0	0
Z strong	364103	11	7	7	3	0	0	1	0	0
Z strong	364104	1	2	1	1	1	0	2	0	0
Z strong	364105	4	10	3	3	1	3	3	5	0
Z strong	364106	10	9	2	1	1	0	0	2	0
Z strong	364107	9	1	2	2	2	0	0	0	0
Z strong	364108	18	17	3	4	5	2	2	1	0
Z strong	364109	5	2	4	3	0	1	0	0	0
Z strong	364110	2	0	0	1	0	0	0	0	0
Z strong	364111	4	3	1	0	0	0	0	0	0
Z strong	364112	12	8	1	0	0	0	0	0	0
Z strong	364113	2	1	1	0	0	0	0	0	0
Z strong	364131	0	0	1	0	0	0	0	0	0
Z strong	364134	1	0	0	0	0	0	0	0	0

Table B.3: Number of pure MC events in the $\mu\mu$ SR, for each $Z\gamma$ strong and Z strong MC sample, in each BDT bin.

Sample	Run	0,0.05	0.15	0.30	0.50	0.64	0.77	0.88	0.96	1
$Z\gamma$ strong	700011	2100	1650	661	427	104	27	17	4	0
$Z\gamma$ strong	700013	1	2	11	9	2	4	1	0	0
$Z\gamma$ strong	700018	568	464	312	168	46	16	14	5	0
$Z\gamma$ strong	700020	0	1	2	0	0	0	1	0	0
Z strong	364114	44	14	8	5	2	1	2	0	0
Z strong	364115	40	9	8	1	0	0	0	0	0
Z strong	364116	56	12	7	3	0	0	0	1	0
Z strong	364117	48	8	1	5	1	0	0	0	0
Z strong	364118	14	5	1	0	0	0	1	0	0
Z strong	364119	37	9	4	1	1	0	0	0	0
Z strong	364120	31	5	5	12	3	0	0	0	0
Z strong	364121	19	2	3	3	2	2	0	1	0
Z strong	364122	52	11	7	3	3	1	0	0	0
Z strong	364123	12	1	1	2	0	0	0	0	0
Z strong	364124	6	2	2	0	0	1	0	0	0
Z strong	364125	6	7	4	3	1	1	0	0	0
Z strong	364126	12	4	1	0	0	0	0	0	0
Z strong	364127	3	0	0	0	0	0	0	0	0

Table B.4: Number of pure MC events in the ee VR, for each $Z\gamma$ strong and Z strong MC sample, in each BDT bin.

Sample	Run	0,0.05	0.15	0.30	0.50	0.64	0.77	0.88	0.96	1
$Z\gamma$ strong	700012	2620	2260	1060	844	171	92	37	7	0
$Z\gamma$ strong	700013	3	0	2	8	3	2	2	1	0
$Z\gamma$ strong	700019	697	604	391	231	52	22	12	8	0
$Z\gamma$ strong	700020	0	0	1	0	1	0	1	0	0
Z strong	364100	53	23	2	5	2	2	1	0	0
Z strong	364101	53	12	6	2	0	0	0	0	0
Z strong	364102	78	22	9	4	4	0	1	0	0
Z strong	364103	57	17	5	4	0	2	0	0	0
Z strong	364104	18	5	1	0	0	0	0	0	0
Z strong	364105	46	16	4	1	1	0	0	0	0
Z strong	364106	33	8	5	3	3	0	0	0	0
Z strong	364107	20	8	4	1	1	0	0	0	0
Z strong	364108	74	15	8	12	0	1	1	0	0
Z strong	364109	13	3	2	0	0	0	0	0	0
Z strong	364110	5	2	0	1	0	0	0	0	0
Z strong	364111	11	6	1	4	0	0	0	0	0
Z strong	364112	27	6	0	0	0	0	0	0	0
Z strong	364113	5	1	0	0	0	0	0	0	0
Z strong	364129	0	1	0	0	0	0	0	0	0

Table B.5: Number of pure MC events in the $\mu\mu$ VR, for each $Z\gamma$ strong and Z strong MC sample, in each BDT bin.

Bibliography

[1] M. E. Peskin and D. V. Schroeder. *QED: The Strange Theory of Light and Matter*. Addison-Wesley, Reading, USA, 1995.

[2] 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.

[3] 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.

[4] ATLAS and CMS Collaborations. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[5] Wikipedia. Fundamental interaction. https://en.wikipedia.org/wiki/Fundamental_interaction.

[6] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961.

[7] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[8] Abdus Salam. Gauge unification of fundamental forces. *Rev. Mod. Phys.*, 52:525–538, Jul 1980.

[9] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215, 1964.

[10] G. Zweig. An SU(3) model for strong interaction symmetry and its breaking. Version 1. 1 1964.

[11] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954.

[12] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[13] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[14] Peter W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966.

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

[16] ATLAS Collaboration. Standard model summary plots february 2022. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-009.

[17] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 10 2016.

[18] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[19] ATLAS Collaboration. Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data. *Phys. Rev. D*, 90(5):052004, 2014.

[20] R. L. Workman et al. Review of Particle Physics - Electroweak Model and Constraints on New Physics. *PTEP*, 2022:083C01, 2022.

[21] Stephen P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 18:1–98, 1998.

[22] Super-Kamiokande Collaboration. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998.

[23] Q. R. Ahmad et al. Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, 2001.

[24] Q. R. Ahmad et al. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, 2002.

[25] G. L. Fogli, Eligio Lisi, A. Marrone, A. Melchiorri, P. Serra, Joseph I. Silk, A. Slosar, and A. Palazzo. Neutrino mass and mixing: 2006 status. *Nucl. Phys. B Proc. Suppl.*, 168:341–343, 2007.

[26] R. L. Workman et al. Review of Particle Physics - Neutrino Masses, Mixing and Oscillations. *PTEP*, 2022:083C01, 2022.

[27] Stefano Dell'Oro, Simone Marcocci, Matteo Viel, and Francesco Vissani. Neutrinoless double beta decay: 2015 review. *Advances in High Energy Physics*, 2016:1–37, 2016.

[28] A. de Gouvea et al. Working Group Report: Neutrinos. In *Community Summer Study 2013: Snowmass on the Mississippi*, 10 2013.

[29] Pavel Fileviez Pérez, Clara Murgui, and Alexis D. Plascencia. Neutrino-Dark Matter Connections in Gauge Theories. *Phys. Rev. D*, 100(3):035041, 2019.

[30] Brian Albert Robson. The matter-antimatter asymmetry problem. *Understanding Gravity*, 2018.

[31] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967.

[32] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.*, 405:279–390, 2005.

[33] Katherine Freese. Status of Dark Matter in the Universe. *Int. J. Mod. Phys.*, 1(06):325–355, 2017.

[34] F. Zwicky. Republication of: The redshift of extragalactic nebulae. *General Relativity and Gravitation*, 41(1):207–224, January 2009.

[35] V C Rubin and W K Ford, Jr. Rotation of the andromeda nebula from a spectroscopic survey of emission regions. *Astrophys. J. 159: 379-403(Feb 1970).*, 1 1970.

[36] J. Anthony Tyson, Greg P. Kochanski, and Ian P. Dell'Antonio. "detailed mass map of CL 0024+1654 from strong lensing". *The Astrophysical Journal*, 498(2):L107–L110, may 1998.

[37] Jennifer K. Adelman-McCarthy et al. The Fourth Data Release of the Sloan Digital Sky Survey. *Astrophys. J. Suppl.*, 162:38–48, 2006.

[38] Ulrich G. Briel and J. Patrick Henry. An x-ray temperature map of coma. 11 1997.

[39] The coma cluster. https://heasarc.gsfc.nasa.gov/docs/rosat/gallery/clus_coma.html.

[40] The bullet cluter. https://www.esa.int/ESA_Multimedia/Images/2007/07/The_Bullet_Cluster2.

[41] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[42] Hinshaw et al. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. 208(2):19, October 2013.

[43] Julien Billard et al. Direct detection of dark matter—APPEC committee report*. *Rept. Prog. Phys.*, 85(5):056201, 2022.

[44] C. Alcock et al. The MACHO project: Microlensing results from 5.7 years of large magellanic cloud observations. *The Astrophysical Journal*, 542(1):281–307, oct 2000.

[45] R. Ansari. EROS: A Galactic microlensing odyssey. In *International Conference on Cosmic Rays and Dark Matter*, 7 2004.

[46] P. Tisserand et al. Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds. *Astron. Astrophys.*, 469:387–404, 2007.

[47] Bernard Carr, Florian Kuhnel, and Marit Sandstad. Primordial Black Holes as Dark Matter. *Phys. Rev. D*, 94(8):083504, 2016.

[48] Hiroko Niikura et al. Microlensing constraints on primordial black holes with Subaru/HSC Andromeda observations. *Nature Astron.*, 3(6):524–534, 2019.

[49] Leszek Roszkowski, Enrico Maria Sessolo, and Sebastian Trojanowski. WIMP dark matter candidates and searches—current status and future prospects. *Rept. Prog. Phys.*, 81(6):066201, 2018.

[50] Hsin-Chia Cheng, Jonathan L. Feng, and Konstantin T. Matchev. Kaluza-klein dark matter. *Phys. Rev. Lett.*, 89:211301, Oct 2002.

[51] A. Arbey and F. Mahmoudi. Dark matter and the early Universe: a review. *Prog. Part. Nucl. Phys.*, 119:103865, 2021.

[52] Kalliopi Petraki and Raymond R. Volkas. Review of asymmetric dark matter. *Int. J. Mod. Phys. A*, 28:1330028, 2013.

[53] R. D. Peccei and Helen R. Quinn. CP conservation in the presence of pseudoparticles. *Phys. Rev. Lett.*, 38:1440–1443, Jun 1977.

[54] Michael Dine, Willy Fischler, and Mark Srednicki. A Simple Solution to the Strong CP Problem with a Harmless Axion. *Phys. Lett. B*, 104:199–202, 1981.

[55] A. R. Zhitnitsky. On Possible Suppression of the Axion Hadron Interactions. (In Russian). *Sov. J. Nucl. Phys.*, 31:260, 1980.

[56] Jihn E. Kim. Weak Interaction Singlet and Strong CP Invariance. *Phys. Rev. Lett.*, 43:103, 1979.

[57] Scott Dodelson and Lawrence M. Widrow. Sterile neutrinos as dark matter. *Phys. Rev. Lett.*, 72:17–20, Jan 1994.

[58] Mordehai Milgrom. A modification of the newtonian dynamics as a possible alternative to the hidden mass hypothesis. *The Astrophysical Journal*, 270:365–370, 1983.

[59] Jacob D. Bekenstein. Relativistic gravitation theory for the MOND paradigm. *Phys. Rev. D*, 70:083509, 2004. [Erratum: Phys.Rev.D 71, 069901 (2005)].

[60] Pieter van Dokkum, Shany Danieli, Yotam Cohen, Allison Merritt, Aaron J. Romanowsky, Roberto Abraham, Jean Brodie, Charlie Conroy, Deborah Lokhorst, Lamiya Mowla, Ewan O'Sullivan, and Jielai Zhang. A galaxy lacking dark matter. *Nature*, 555(7698):629–632, mar 2018.

[61] Mireia Montes, Raúl Infante-Sainz, Alberto Madrigal-Aguado, Javier Román, Matteo Monelli, Alejandro S. Borlaff, and Ignacio Trujillo. The Galaxy "Missing Dark Matter" NGC 1052-DF4 is Undergoing Tidal Disruption. *Astrophys. J.*, 904(2):114, 2020.

[62] Mireia Montes, Raúl Infante-Sainz, Alberto Madrigal-Aguado, Javier Román, Matteo Monelli, Alejandro S. Borlaff, and Ignacio Trujillo. The galaxy "missing dark matter" ngc 1052-df4 is undergoing tidal disruption. *The Astrophysical Journal*, 904(2):114, nov 2020.

[63] John Moffat and V. Toth. Ngc 1052-df2 and modified gravity (mog) without dark matter. *Monthly Notices of the Royal Astronomical Society: Letters*, 482:L1–L3, 01 2019.

[64] Rouven Essig et al. Working Group Report: New Light Weakly Coupled Particles. In *Community Summer Study 2013: Snowmass on the Mississippi*, 10 2013.

[65] Jim Alexander et al. Dark Sectors 2016 Workshop: Community Report. 8 2016.

[66] Marco Fabbrichesi, Emidio Gabrielli, and Gaia Lanfranchi. The physics of the dark photon. *SpringerBriefs in Physics*, 2021.

[67] M. A. Deliyergiyev. Recent Progress in Search for Dark Sector Signatures. *Open Phys.*, 14(1):281–303, 2016.

[68] J. Beacham et al. Physics Beyond Colliders at CERN: Beyond the Standard Model Working Group Report. *J. Phys. G*, 47(1):010501, 2020.

[69] Bob Holdom. Two U(1)'s and Epsilon Charge Shifts. *Phys. Lett. B*, 166:196–198, 1986.

[70] Chengcheng Han, M.L. Lopez-Ibanez, Bo Peng, and Jin Min Yang. Dirac dark matter in u(1)b-l with the stueckelberg mechanism. *Nuclear Physics B*, 959:115154, 2020.

[71] Julian Heeck and Werner Rodejohann. Neutrino Phenomenology of gauged $L_\mu - L_\tau$: MINOS and beyond. *AIP Conf. Proc.*, 1382(1):144–146, 2011.

[72] R. L. Workman et al. Review of Particle Physics - Supersymmetry, Part I (Theory). *PTEP*, 2022:083C01, 2022.

[73] Sacha Davidson, Steen Hannestad, and Georg Raffelt. Updated bounds on millicharged particles. *JHEP*, 05:003, 2000.

[74] E. C. G. Stueckelberg. Interaction energy in electrodynamics and in the field theory of nuclear forces. *Helv. Phys. Acta*, 11:225–244, 1938.

[75] R. Bernabei et al. Final model independent result of DAMA/LIBRA-phase1. *Eur. Phys. J. C*, 73:2648, 2013.

[76] R. Bernabei et al. First model independent results from DAMA/LIBRA-phase2. *Nucl. Phys. Atom. Energy*, 19(4):307–325, 2018.

[77] R. Bernabei et al. Further results from DAMA/LIBRA–phase2 and perspectives. *Nucl. Phys. At. Energy*, 22(329), 2021.

[78] M. Ackermann et al. The Fermi Galactic Center GeV Excess and Implications for Dark Matter. *Astrophys. J.*, 840(1):43, 2017.

[79] Simona Murgia. The Fermi–LAT Galactic Center Excess: Evidence of Annihilating Dark Matter? *Ann. Rev. Nucl. Part. Sci.*, 70:455–483, 2020.

[80] Sumita Ghosh, E. P. Ruddy, Michael J. Jewell, Alexander F. Leder, and Reina H. Maruyama. Searching for dark photons with existing haloscope data. *Phys. Rev. D*, 104(9):092016, 2021.

[81] https://spectrum.ieee.org/analyzing-the-lhc-magnet-quenches.

[82] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[83] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[84] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[85] ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[86] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[87] John David Jackson. *Classical electrodynamics; 2nd ed.* Wiley, New York, NY, 1975.

[88] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J. C*, 75(5):204, 2015.

[89] Ewa Lopienska. The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022. 2022. https://cds.cern.ch/record/2800984.

[90] Daniel Boussard and Trevor Paul R Linnecar. The LHC Superconducting RF System. Technical report, CERN, Geneva, 1999.

[91] AC Team. Diagram of an LHC dipole magnet. Schéma d'un aimant dipôle du LHC. 1999.

[92] Jean-Luc Caron. LHC quadrupole cross section. AC Collection. Legacy of AC. Pictures from 1992 to 2002., 1998.

[93] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ using the ATLAS detector at the LHC. 2022.

[94] R Bruce, N Fuster-Martínez, A Mereghetti, D Mirarchi, and S Redaelli. Review of LHC Run 2 Machine Configurations. pages 187–197, 2019.

[95] ATLAS Collaboration. Public atlas luminosity results for run-2 of the lhc. [Online; accessed 15-March-2019].

[96] Joao Pequenao. Computer generated image of the whole ATLAS detector. https://cds.cern.ch/record/1095924, 2008.

[97] Joao Pequenao. Computer generated image of the ATLAS inner detector. https://cds.cern.ch/record/1095926, 2008.

[98] M Capeans, G Darbo, K Einsweiller, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical report, 2010.

[99] Joao Pequenao. Computer Generated image of the ATLAS calorimeter. https://cds.cern.ch/record/1095927, 2008.

[100] Joao Pequenao. Computer generated image of the ATLAS Muons subsystem. https://cds.cern.ch/record/1095929, 2008.

[101] T Cornelissen, M Elsing, S Fleischmann, W Liebig, E Moyse, and A Salzburger. Concepts, Design and Implementation of the ATLAS New Tracking (NEWT). Technical report, CERN, Geneva, 2007. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2007-007.

[102] ATLAS Collaboration. Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2. *Eur. Phys. J. C*, 77:673, 2017.

[103] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth. A*, 262:444–450, 1987.

[104] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. Adaptive vertex fitting. *J. Phys. G*, 34:N343, 2007.

[105] ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data. *JINST*, 14:P12006, 2019.

[106] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal. Calorimeter clustering algorithms: Description and performance. 5 2008.

[107] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77:490, 2017.

[108] ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 79:639, 2019.

[109] Thijs G. Cornelissen, M. Elsing, I. Gavrilenko, J. F. Laporte, W. Liebig, M. Limper, K. Nikolopoulos, A. Poppleton, and A. Salzburger. The global chi**2 track fitter in ATLAS. *J. Phys. Conf. Ser.*, 119:032013, 2008.

[110] Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung. Technical report, CERN, Geneva, 2012. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2012-047.

[111] Morad Aaboud et al. Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016. *Eur. Phys. J. C*, 79(3):205, 2019.

[112] Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach. Technical report, CERN, Geneva, 2017. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-022.

[113] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *Eur. Phys. J. C*, 74:3071, 2014.

[114] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data. *JINST*, 14:P03017, 2019.

[115] ATLAS Collaboration. Atlas egamma public plots: Photon identification efficiencies and scale factors in run 2. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/EGAM-2021-01.

[116] ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 81:578, 2021.

[117] Stefan Höche. Introduction to parton-shower event generators. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pages 235–295, 2015.

[118] ATLAS Collaboration. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. *Eur. Phys. J. C*, 81:689, 2020.

[119] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[120] ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. *Phys. Rev. D*, 96:072002, 2017.

[121] ATLAS Collaboration. Pile-up subtraction and suppression for jets in ATLAS. 8 2013.

[122] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119–126, 2008.

[123] ATLAS Collaboration. Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS. 8 2015.

[124] ATLAS Collaboration. Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV. Technical report, CERN, Geneva, 2015. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-002.

[125] ATLAS Collaboration. Determination of the jet energy scale and resolution at ATLAS using Z/γ–jet events in data at $\sqrt{s} = 8$ TeV. Technical report, CERN, Geneva, 2015. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-057.

[126] ATLAS Collaboration. TeV-scale jet energy calibration using multijet events including close-by jet effects at the ATLAS experiment. 1 2013.

[127] ATLAS Collaboration. Tagging and suppression of pileup jets. 5 2014.

[128] ATLAS Collaboration. Forward jet vertex tagging using the particle flow algorithm. Technical report, CERN, Geneva, 2019. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-026.

[129] ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in *pp* collisions at $\sqrt{s} = 8\,\text{TeV}$ using the ATLAS detector. *Eur. Phys. J. C*, 76:581, 2016.

[130] ATLAS Collaboration. Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector. Technical report, CERN, Geneva, 2015. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-029.

[131] ATLAS Collaboration. Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. Technical report, CERN, Geneva, 2015. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-045.

[132] ATLAS Collaboration. $E_{\mathrm{T}}^{\mathrm{miss}}$ performance in the ATLAS detector using 2015-2016 LHC pp collisions. 6 2018.

[133] ATLAS Collaboration. Tst systematics using 2015-2018 data. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2020-02/.

[134] Object-based missing transverse momentum significance in the ATLAS detector. Technical report, CERN, Geneva, 2018. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2018-038.

[135] ATLAS Collaboration. Missing transverse momentum performance using the full run 2 pp data at 13 tev. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2019-03/.

[136] William Keaton Balunas, Donatella Cantore-Cavalli, Teng Jian Khoo, Matthew Henry Klein, Peter Loch, Federica Piazza, Caterina Pizio, Silvia Resconi, Douglas Michael Schaefer, Russell Woods Smith, and Sarah Louise Williams. A flexible and efficient approach to reconstruction of missing transverse momentum. Technical report, CERN, Geneva, 2022. This is the working draft of the 2022 MET software paper.

[137] J. Billard, L. Strigari, and E. Figueroa-Feliciano. Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments. *Phys. Rev. D*, 89(2):023524, 2014.

[138] Daniel Abercrombie et al. Dark Matter benchmark models for early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum. *Phys. Dark Univ.*, 27:100371, 2020.

[139] Antonio Boveia et al. Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter. *Phys. Dark Univ.*, 27:100365, 2020.

[140] ATLAS Collaboration. Dark matter summary plots for s-channel, 2HDM+a and Dark Higgs models. Technical report, CERN, Geneva, 2022. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-036.

[141] Arpit Gupta, Reinard Primulando, and Prashant Saraswat. A New Probe of Dark Sector Dynamics at the LHC. *JHEP*, 09:079, 2015.

[142] Timothy Cohen, Mariangela Lisanti, Hou Keong Lou, and Siddharth Mishra-Sharma. LHC Searches for Dark Sector Showers. *JHEP*, 11:196, 2017.

[143] Sanjoy Biswas, Emidio Gabrielli, and Barbara Mele. Dark Photon Searches via Higgs Boson Production at the LHC and Beyond. *Symmetry*, 14(8):1522, 2022.

[144] Theodota Lagouri. Review on Higgs Hidden–Dark Sector Physics at High-Energy Colliders. *Symmetry*, 14(7):1299, 2022.

[145] Emidio Gabrielli, Matti Heikinheimo, Barbara Mele, and Martti Raidal. Dark photons and resonant monophoton signatures in Higgs boson decays at the LHC. *Phys. Rev. D*, 90(5):055032, 2014.

[146] Emidio Gabrielli and Raidal Marti. Exponentially spread dynamical yukawa couplings from non-perturbative chiral symmetry breaking in the dark sector. *Phys. Rev. D*, 89:015008, 2014.

[147] ATLAS Collaboration. Observation of electroweak production of two jets in association with an isolated photon and missing transverse momentum, and search for a Higgs boson decaying into invisible particles at 13 TeV with the ATLAS detector. *Eur. Phys. J. C*, 82:105, 2021.

[148] CMS Collaboration. Search for dark photons in decays of Higgs bosons produced in association with Z bosons in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JHEP*, 10:139, 2019.

[149] CMS Collaboration. Search for dark photons in Higgs boson production via vector boson fusion in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JHEP*, 03:011, 2021.

[150] ATLAS Collaboration. ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity: Technical Design Report, 2008.

[151] ATLAS Collaboration. Vertex Reconstruction Performance of the ATLAS Detector at "\sqrts = 13˜\textrmTeV". 2015.

[152] Chiara Debenedetti. Concepts for fast large scale Monte Carlo production for the ATLAS experiment. *J. Phys. Conf. Ser.*, 513:022006, 2014.

[153] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823, 2010.

[154] ATLAS Collaboration. ATLAS Computing: Technical Design Report, 2005.

[155] G. Marchesini and B. R. Webber. Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation. *Nucl. Phys. B*, 310:461–526, 1988.

[156] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[157] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.

[158] B. R. Webber. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys. B*, 238:492–528, 1984.

[159] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[160] W Lukas. Fast Simulation for ATLAS: Atlfast-II and ISF. Technical report, CERN, Geneva, 2012.

[161] ATLAS Collaboration. AtlFast3: The Next Generation Of Fast Simulation in ATLAS. *Comput. Softw. Big Sci.*, 6:7, 2021.

[162] K Edmonds, S Fleischmann, T Lenz, C Magass, J Mechnich, and A Salzburger. The Fast ATLAS Track Simulation (FATRAS). Technical report, CERN, Geneva, 2008. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2008-001.

[163] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852, 2008.

[164] Rene Brun and Fons Rademakers. Root – an object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389(1):81 – 86, 1997.

[165] D Adams, C Anastopoulos, A Andreazza, M Aoki, L Asquith, M Begel, F Bernlochner, U Blumenschein, A Bocci, S Cheatham, W Davey, P-A Delsart, P-O DeViveiros, A Dewhurst, D Duschinger, F Filthaut, P Francavilla, F Garberson, S Head, A Henrichs, A Hoecker, M Kagan, B Kersevan, TJ Khoo, B Lenzi, D Lopez Mateos, B Malaescu, Z Marshall, T Martin, C Meyer, A Morley, W Murray, M zur Nedden, R Nicolaidou, S Pagan Griso, G Pasztor, P Petroff, C Pizio, R Polifka, X Poveda, R Reece, F Ruehr, F Salvatore, R Sandstroem, T Scanlon, D Scheirich, S Schramm, A Schwartzman, K Suruliz, M Sutton, E Thompson, M Tripiana, A Tuna, S Viel, M Vincter, I Vivarelli, M Wielers, A Wildauer, and Z Zinonos. Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014. Technical report, CERN, Geneva, 2014.

[166] ATLAS Collaboration. Susy object definitions (rel 21). https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SusyObjectDefinitionsr2113TeV#SUSYTools.

[167] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.

[168] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93(3):033006, 2016.

[169] G. J. Besjes, M. Baak, D. Côté, A. Koutsman, J. M. Lorenz, and D. Short. HistFitter: a flexible framework for statistical data analysis. *J. Phys. Conf. Ser.*, 664(7):072004, 2015.

[170] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. 2012.

[171] Wouter Verkerke and David Kirkby. The RooFit toolkit for data modeling, 2003.

[172] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011.

[173] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals Math. Statist.*, 9(1):60–62, 1938.

[174] Abraham Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54:426–482, 1943.

[175] Alexander L. Read. Presentation of search results: the CL_S technique. *J. Phys. G*, 28:2693, 2002.

[176] ATLAS Collaboration. Search for dark matter in association with an energetic photon in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. *JHEP*, 02:226, 2021.

[177] ATLAS Collaboration. Search for dark matter at $\sqrt{s} = 13\,\text{TeV}$ in final states containing an energetic photon and large missing transverse momentum with the ATLAS detector. *Eur. Phys. J. C*, 77:393, 2017.

[178] Andreas Albert et al. Recommendations of the LHC Dark Matter Working Group: Comparing LHC searches for dark matter mediators in visible and invisible decay channels and calculations of the thermal relic density. *Phys. Dark Univ.*, 26:100377, 2019.

[179] Richard D. Ball et al. Parton distributions for the LHC run II. *JHEP*, 04:040, 2015.

[180] Mihailo Backović, Michael Krämer, Fabio Maltoni, Antony Martini, Kentarou Mawatari, and Mathieu Pellen. Higher-order QCD predictions for dark matter production at the LHC in simplified models with s-channel mediators. *Eur. Phys. J. C*, 75(10):482, 2015.

[181] Dmsimp: Simplified dark matter models. https://feynrules.irmp.ucl.ac.be/wiki/DMsimp.

[182] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and Winter J. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[183] Enrico Bothmann, Gurpreet Singh Chahal, Stefan Höche, Johannes Krause, Frank Krauss, Silvan Kuttimalai, Sebastian Liebschner, Davide Napoletano, Marek Schönherr, Holger Schulz, Steffen Schumann, and Frank Siegert. Event generation with sherpa 2.2. *SciPost Physics*, 7(3), sep 2019.

[184] Tanju Gleisberg and Stefan Höche. Comix, a new matrix element generator. *JHEP*, 12:039, 2008.

[185] Fabio Cascioli, Philipp Maierhöfer, and Stefano Pozzorini. Scattering Amplitudes with Open Loops. *Phys. Rev. Lett.*, 108:111601, 2012.

[186] Stefan Höche, Frank Krauss, Steffen Schumann, and Frank Siegert. QCD matrix elements and truncated showers. *JHEP*, 05:053, 2009.

[187] PICASSO Collaboration. Final results of the picasso dark matter search experiment. *Astropart. Phys.*, 90:85–92, 2017.

[188] SuperCDMS Collaboration. Low-mass dark matter search with cdmslite. *Phys. Rev. D*, 97:022002, 2018.

[189] PICO Collaboration. Dark matter search results from the complete exposure of the pico-60 c_3f_8 bubble chamber. *Phys. Rev. D*, 100(2):022001, 2019.

[190] XENON Collaboration. Constraining the spin-dependent wimp-nucleon cross sections with xenon1t. *Phys. Rev. Lett.*, 122:141301, 2019.

[191] LUX Collaboration. Limits on spin-dependent wimp-nucleon cross section obtained from the complete lux exposure. *Phys. Rev. Lett.*, 118(25):251302, 2017.

[192] PandaX-II Collaboration. Spin-dependent weakly-interacting-massive-particle–nucleon cross section limits from first data of pandax-ii experiment. *Phys. Rev. Lett.*, 118:071301, 2017.

[193] XENON Collaboration. Dark matter search results from a ton-year exposure of xenon1t. *Phys. Rev. Lett.*, 121:111302, 2018.

[194] PandaX-II Collaboration. Dark matter results from 54-ton-day exposure of pandax-ii experiment. *Phys. Rev. Lett.*, 119:181302, 2017.

[195] DarkSide Collaboration. Low-mass dark matter search with the darkside-50 experiment. *Phys. Rev. Lett.*, 121:081307, 2018.

[196] XENON Collaboration. Search for light dark matter interactions enhanced by the migdal effect or bremsstrahlung in xenon1t. *Phys. Rev. Lett.*, 123:241803, 2019.

[197] JiJi Fan, Andrey Katz, Lisa Randall, and Matthew Reece. Dark-Disk Universe. *Phys. Rev. Lett.*, 110(21):211302, 2013.

[198] Matti Heikinheimo, Martti Raidal, Christian Spethmann, and Hardi Veermäe. Dark matter self-interactions via collisionless shocks in cluster mergers. *Phys. Lett. B*, 749:236–241, 2015.

[199] Nima Arkani-Hamed, Douglas P. Finkbeiner, Tracy R. Slatyer, and Neal Weiner. A Theory of Dark Matter. *Phys. Rev. D*, 79:015014, 2009.

[200] Kathryn M. Zurek. Asymmetric Dark Matter: Theories, Signatures, and Constraints. *Phys. Rept.*, 537:91–121, 2014.

[201] ATLAS Collaboration. Search for dark photons from Higgs boson decays via ZH production with a photon plus missing transverse momentum signature from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 12 2022.

[202] ATLAS Collaboration. Performance of the ATLAS muon triggers in Run 2. *JINST*, 15:P09015, 2020.

[203] ATLAS Collaboration. Performance of electron and photon triggers in ATLAS during LHC Run 2. *Eur. Phys. J. C*, 80:47, 2020.

[204] ATLAS Collaboration. Operation of the ATLAS trigger system in Run 2. *JINST*, 15:P10004, 2020.

[205] The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model. Technical report, CERN, Geneva, 2016. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-017.

[206] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152, 2001.

[207] Lisa Carloni and Torbjörn Sjöstrand. Visible effects of invisible hidden valley radiation. *Journal of High Energy Physics*, 2010(9), sep 2010.

[208] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016.

[209] John M. Campbell, R. Keith Ellis, Rikkert Frederix, Paolo Nason, Carlo Oleari, and Ciaran Williams. NLO Higgs boson production plus one and two jets using the POWHEG BOX, MadGraph4 and MCFM. *JHEP*, 07:092, 2012.

[210] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. $HW^{\pm}/HZ$ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013.

[211] J. Pumplin et al. New Generation of Parton Distributions with Uncertainties from Global QCD Analysis. *JHEP*, 07:012, 2002.

[212] Stefano Goria, Giampiero Passarino, and Dario Rosco. The Higgs Boson Lineshape. *Nucl. Phys. B*, 864:530–579, 2012.

[213] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244, 2013.

[214] Nikolai Hartmann. ahoi (a horrible optimization instrument), 2020.

[215] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[216] Richard Simard and Pierre L'Ecuyer. Computing the two-sided kolmogorov-smirnov distribution. *Journal of Statistical Software*, 39(11):1–18, 2011.

[217] Ansgar Denner, Stefan Dittmaier, Stefan Kallweit, and Alexander Mück. Hawk 2.0: A monte carlo program for higgs production in vector-boson fusion and higgs strahlung at hadron colliders. *Computer Physics Communications*, 195:161–171, Oct 2015.

[218] J H Friedman. Data analysis techniques for high energy particle physics. page 96 p, Oct 1974.

[219] I. Yavin K. Cranmer. RECAST: Extending the Impact of Existing Analyses. *JHEP*, 3:S08001, 2011.

[220] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[221] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[222] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[223] Keith Hamilton, Paolo Nason, Carlo Oleari, and Giulia Zanderighi. Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching. *JHEP*, 05:082, 2013.

[224] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. MINLO: multi-scale improved NLO. *JHEP*, 10:155, 2012.

[225] J. Gao et al. CT10 next-to-next-to-leading order global analysis of QCD. *Phys. Rev. D*, 89:033009, 2014.

[226] Christian Bierlich et al. A comprehensive guide to the physics and usage of PYTHIA 8.3. 3 2022.

[227] Enrico Bothmann, Marek Schönherr, and Steffen Schumann. Reweighting QCD matrix-element and parton-shower calculations. *Eur. Phys. J. C*, 76(11):590, 2016.

[228] CMS Collaboration. Search for new phenomena in monophoton final states in proton–proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *Phys. Lett. B*, 755:102, 2016.

[229] CMS Collaboration. Dm summary plots. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryPlotsEXO13TeV#DM_summary_plots.