






Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



Search for new phenomena in events with jets and large missing transverse momentum at the high-energy LHC Run II using the ATLAS detector

José L. Muñoz Martínez

PH.D. THESIS

Institut de Física d'Altes Energies
Universitat Autònoma de Barcelona
Departament de Física
Facultat de Ciències

Programa de Doctorat en Física

September, 2021

Directed by

Mario Martínez Pérez

Institució Catalana de Recerca i Estudis Avançat
Institut de Física d'Altes Energies
Universitat Autònoma de Barcelona

“First principle: never to let one’s self be beaten down by persons or by events”

Maria Salomea Skłodowska-Curie

Contents

INTRODUCTION	I
1 THE STANDARD MODEL AND BEYOND	3
1.1 The Standard Model	3
1.1.1 Quantum Electrodynamics	5
1.1.2 Quantum Chromodynamics	6
1.1.3 Electroweak interaction and the Higgs mechanism	8
1.2 Hadron collisions	12
1.2.1 Proton structure	12
1.2.2 Event simulation	14
1.3 Beyond the Standard Model	17
1.3.1 Large Extra Dimensions	19
1.3.2 Dark Matter production	21
1.3.3 Supersymmetric quark-pair production	25
1.3.4 Invisible-decaying Higgs	30
1.3.5 Other models	31
2 MC SIMULATION	33
2.1 Simulation Software	33
2.1.1 MC generators	33
2.1.2 Detector simulation	34
2.2 Simulated Samples for the Monojet Analysis	35
2.2.1 Background samples	35
2.2.2 Signal samples	36
2.2.3 MC correction weights	38
2.3 V+jets reweighting	39
3 THE ATLAS EXPERIMENT AT THE LHC	45
3.1 The Large Hadron Collider	45
3.2 The ATLAS Detector	47
3.2.1 Inner Detector	48

	3.2.2	Calorimeters	50
	3.2.3	Muon Spectrometers	53
	3.2.4	Trigger system	54
	3.2.5	Luminosity definition and Good Run List	56
4		EVENT RECONSTRUCTION	59
	4.1	Tracks and Primary Vertex	59
	4.1.1	Track reconstruction	59
	4.1.2	Vertex reconstruction	60
	4.2	Electrons	61
	4.2.1	Electron Reconstruction	61
	4.2.2	Electron Identification	62
	4.2.3	Electron Isolation	63
	4.3	Photons	64
	4.3.1	Photon Reconstruction	64
	4.3.2	Photon Identification	65
	4.4	Muons	66
	4.4.1	Muon Reconstruction	66
	4.4.2	Muon Identification	68
	4.5	Jets	69
	4.5.1	Jet Reconstruction	69
	4.5.2	Jet Calibration	70
	4.5.3	Jet identification	75
	4.5.4	Jet cleaning	77
	4.5.5	b -tagging	78
	4.6	Taus	79
	4.6.1	Tau Reconstruction	79
	4.6.2	Tau Identification	80
	4.7	Missing Transverse Energy	81
	4.7.1	E_T^{miss} reconstruction	82
5		THE STATISTICAL MODEL	85
	5.1	Introduction	85
	5.2	Statistical test	86
	5.2.1	Statistical significance	87
	5.2.2	The CL_s method	88
	5.3	The Maximum Likelihood method	89
	5.3.1	The Likelihood function	89
	5.3.2	The Profile Likelihood Ratio	90
	5.4	Fit configurations	91

6	THE MONOJET ANALYSIS	93
6.1	Data sample	94
6.1.1	Trigger	94
6.2	Event selection	95
6.2.1	Object definitions and Overlap Removal	95
6.2.2	Pre-selection	98
6.2.3	Signal Region definition	98
6.2.4	Control Regions definitions	99
6.3	Background estimation	101
6.3.1	Analysis strategy	101
6.3.2	Control Regions fit	103
6.3.3	Non-Collision Background	104
6.3.4	Multi-jet background	105
6.4	Pre-fit distributions	107
7	SYSTEMATIC UNCERTAINTIES	115
7.1	Implementation in the fit	115
7.1.1	Systematics smoothing	116
7.1.2	Systematics pruning	116
7.2	Experimental uncertainties	116
7.2.1	Luminosity	117
7.2.2	Pile-up re-weighting	117
7.2.3	Jets	117
7.2.4	E_T^{miss} track soft term	121
7.2.5	Leptons and photons	122
7.3	Theoretical uncertainties	124
7.3.1	Background processes	124
7.3.2	Signal processes	131
8	RESULTS AND INTERPRETATIONS	137
8.1	Background-only Fit	137
8.1.1	Fit parameters	137
8.1.2	Event yields and post-fit distributions in the CRs	142
8.2	Results in the SR	153
8.2.1	Background estimation	153
8.2.2	Goodness of the fit	156
8.2.3	Impact of systematic uncertainties	156
8.2.4	Kinematic distributions	158
8.3	Model-Independent Limits	160
8.4	Interpretations	161
8.4.1	Large Extra Dimensions	161

8.4.2	Dark Matter production	163
8.4.3	SUSY Quark-Pair production	166
8.4.4	Invisible-decaying Higgs	168
8.4.5	Other models	168
CONCLUSION		170
APPENDIX A IBL 3D PIXEL SENSORS PERFORMANCE STUDIES		173
A.1	I-V curves studies	175
A.2	HV scans	177
A.2.1	Efficiency vs. HV	177
A.2.2	Cluster size, ToT and charge collection vs HV	178
APPENDIX B ANTI-SFs CALCULATION METHOD		181
APPENDIX C NCB STUDIES		185
APPENDIX D ADDITIONAL MATERIAL FROM THE BACKGROUND-ONLY FIT		189
APPENDIX E SYSTEMATICS IMPACT		197
APPENDIX F SIGNALS CUTFLOW		199
APPENDIX G V +JETS REWEIGHTING IMPACT		203
REFERENCES		218
ACKNOWLEDGMENTS		219

Introduction

Understanding the nature of the Universe has been always a main concern across History. Our description of matter has evolved significantly over the centuries: from Democritus and Leucippus with their concept of ‘atoms’ in the Ancient Greek times, passing through Dalton and his modern atomic theory, Rutherford’s experiments, Bohr’s atomic model, the formulation of quantum mechanics in the mid 1920’s, the discoveries of the neutron and the positron by Chadwick and Anderson, the measurement of the β -decay by Fermi, the idea of *quarks* introduced by Gell-Mann and Zweig... ending up with the formulation of the Standard Model (SM) of particle physics, a complete theory which was systematically confirmed by the observations of the W and Z bosons in 1983 and 1984 at the Super Proton Synchrotron at CERN, the observation of the top-quark in 1995 at Fermilab and, finally, the discovery of the Higgs boson in 2012 at the Large Hadron Collider.

In parallel, Astronomy tells us that we live in a Universe which was originated in a Big Bang, has been expanding and cooling down ever since, and now is formed by clusters and galaxies bounded by the action of gravity, which we are able to describe thanks to the contributions from Newton and Einstein. Moreover, Cosmology has revealed that such Universe is governed by an unknown substance called *dark energy*, which is accelerating the expansion, but also that there is another kind of matter, called *dark matter* as it barely interacts with regular matter. The SM does not explain any of these observations, among many others of similar relevance, so it is commonly accepted that the theory must be extended somehow. For this reason, many theoretical models of physics beyond the Standard Model have been proposed in the last decades, most of them predicting the existence of new particles, and an effort has been deployed by the experiments looking for evidences of such new phenomena in many different scenarios. However, the SM has proved to be quite hard to break and no significant evidences of new particles have been found, but rather exclusion limits have been set to the proposed alternative models.

The work described in this thesis represents another attempt to explore the limits of the SM. Data from proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV, collected by the ATLAS detector during 2015–2018, are used to test the validity of the predictions given by the SM.

Via a complex statistical analysis, the data is also used to set limits on the parameters of a number of different models of new physics. The analysis looks for events with a jet recoiling a large amount of missing transverse momentum, this is why it receives the name of *monojet analysis*, and it is known for being a very powerful tool in searches for new phenomena. The assumption is that if a new particle is created in the collision it would escape undetected, hence the analysis looks for excesses of events with high missing transverse momentum over the SM predictions.

This thesis is organized as follows. In Chapter 1, a short introduction to the underlying theoretical framework is given. Chapter 2 summarizes the preparation of the simulated data samples that are used in the analysis. A brief description of the experimental setup is provided in Chapters 3 and 4, the former describing both the accelerator facilities of the Large Hadron Collider at CERN and the most important technical aspects of the ATLAS detector, and the later discussing how the collision products measured by ATLAS are reconstructed and identified. The monojet analysis is then introduced in Chapter 6, detailing the selection applied to the data along with the strategy followed. After that, Chapter 5 gives a brief description of the tools that are employed in the statistical analysis and Chapter 7 details the treatment of the systematical uncertainties. The obtained results are finally presented and discussed in Chapter 8, and they are also interpreted in the context of some Beyond Standard Model theories, including the ADD model of extra dimensions, Supersymmetry in compressed scenarios, Dark Matter pair production and Higgs decaying into invisible. In addition, a number of appendices are included with complementing material detector and analysis related.

This work has contributed to the following publications:

- *Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector* (Phys. Rev. D, 103:112006, Jun 2021).
- *Search for new phenomena in events with jets and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector* (August 2020, ATLAS-CONF-2020-048).
- *Dark matter summary plots for s-channel mediators*. (Mar 2021, ATL-PHYS-PUB-2021-006).
- *SUSY June 2021 Summary Plot Update*. June 2021, ATL-PHYS-PUB-2021-019).

1

The Standard Model and Beyond

The work presented in this thesis is essentially a test to the Standard Model of particle physics, whose predictions are contrasted against a large amount of data collected from proton–proton collisions at the Large Hadron Collider via a statistical analysis. The basics of the theoretical background of the analysis are outlined in this chapter, starting from a brief introduction to the Standard Model, then describing the phenomenology of the hadron collisions and, finally, introducing some theoretical models of physics beyond the Standard Model that are going to be tested in the analysis.

1.1 THE STANDARD MODEL

The SM is a renormalizable quantum field theory (QFT) that describes the dynamics of the elementary (point-like, with no further substructure) particles that conform the regular matter in the Universe. It was developed during the 1960's and 1970's decades by the combined contribution from several physicists [1–3] and is, up to the date, the most precise scientific theory in terms of the level of accuracy of its predictions. Such predictions have been confirmed systematically by the experiments, being the most recent example the discovery of the Higgs boson, whose existence was proposed simultaneously by Brout, Englert, Higgs and other authors [4–6] in 1964 and empirically observed by the ATLAS

Fermions						Bosons		
Quarks	Q	m [MeV/ c^2]	Leptons	Q	m [MeV/ c^2]	Gauge bosons	Q	m [GeV/ c^2]
Up (u/\bar{u})	$\pm\frac{2}{3}$	$2.16^{+0.49}_{-0.26}$	Electron/positron (e^\pm)	± 1	0.511	Photon (γ)	0	0
Down (d/\bar{d})	$\mp\frac{1}{3}$	$4.67^{+0.48}_{-0.17}$	Electron neutrino ($\nu_e/\bar{\nu}_e$)	0	$< 1.1 \cdot 10^{-3}$	W^\pm	± 1	80.379 ± 0.012
Charm (c/\bar{c})	$\pm\frac{2}{3}$	$(1.27 \pm 0.02) \cdot 10^3$	Muon (μ^\pm)	± 1	105.7	Z	0	91.1876 ± 0.0021
Strange (s/\bar{s})	$\mp\frac{1}{3}$	93^{+11}_{-5}	Muon neutrino ($\nu_\mu/\bar{\nu}_\mu$)	0	< 0.19	Gluon (g)	0	0
Top (t/\bar{t})	$\pm\frac{2}{3}$	$(172.76 \pm 0.30) \cdot 10^3$	Tau (τ^\pm)	± 1	776.86 ± 0.12	Scalar bosons	Q	m [GeV]
Bottom (b/\bar{b})	$\mp\frac{1}{3}$	$(4.18^{+0.03}_{-0.02}) \cdot 10^3$	Tau neutrino ($\nu_\tau/\bar{\nu}_\tau$)	0	< 18.2	Higgs (H)	0	125.25 ± 0.17

Table 1.1: Elementary particles in the SM with the corresponding mass and electric charge, the later given in units of the charge of the electron [9]. Quarks and leptons are shown together with their respective anti-particles.

and CMS experiments in 2012 [7, 8].

Schematically, SM elementary particles can be summarized as shown in Table 1.1: divided into fermions and bosons*, the former split again into quarks and leptons and the later into vector and scalar bosons. Quarks and leptons have both spin $J = 1/2$ and are grouped in three families, or generations, ranked by increasing mass: the first one formed by the up (u) and down (d) quarks plus the electron (e) and the electron neutrino (ν_e); the second family with the charm (c) and strange (s) quarks and the muon (μ) and the muon neutrino (ν_μ); and, finally, the third generation with the top (t) and bottom (b) quarks, together with the tau-lepton (τ) and the corresponding tau neutrino (ν_τ). They are the building blocks of everyday matter. Due to the *color confinement*, one of the main characteristics of the strong interaction (see Section 1.1.2), quarks are only found forming composite particles named baryons, which can be either hadrons (combinations of three quarks, e.g the proton) or mesons (quark-anti-quark pairs, e.g. pions). Electrons are typically found being part of the atoms and bounding molecules, while muons are normally generated in high-energetic events in the atmosphere, travelling free until they decay into electrons, and taus are harder to find in nature due to their short mean lifetime.

On the other hand, the SM is formulated as a *gauge* theory, which means that its Lagrangian is invariant under local (gauge) transformations. In this context, the interactions between the different particles arise from a gauge symmetry and are explained in terms of the exchange of a mediator particle, a role played by vector bosons (with $J = 1$). The SM accommodates three of the four known fundamental interactions, namely the electromagnetic, weak and strong interactions, whereas the gravita-

*Fermions are defined as those particles whose quantum state is described by an anti-symmetric wave function, i.e. it will change sign under permutations. These type of particles are characterized for having half-odd integer spin values and follow Fermi-Dirac statistics, being one of its main consequences the so-called Pauli exclusion principle. On the other hand, bosons are described by symmetric wave functions, have typically integer spin values and behave obeying Bose-Einstein statistics.

tional one has not been possible to fit within a QFT framework yet. The electromagnetic interaction originates from a $U_{\text{EM}}(1)$ symmetry and is mediated by the massless photon (γ). The weak interaction comes from a symmetry described by the group $SU(2)_L \otimes U(1)_Y$, where L denotes that only left-handed particles participate in the interaction and the label Y refers to the weak hypercharge (related to the electric charge as $Q = T_3 + \frac{1}{2}Y$, with T_3 as the third component of the weak isospin). This means that the unification with the electromagnetic interaction is already included in the gauge group. The weak interaction is governed by three mediators, the W^- , W^+ and Z bosons and it has a strength (relative to the electromagnetism) of $\sim 10^{-3}$. The strong interaction arises from the $SU(3)_C$ symmetry group, where C stands for “color”, and is ruled by the gluons, which are massless particles that can be of 8 types. It is of about one order of magnitude stronger than the electromagnetic force. A more detailed description of the fundamental interactions is given in the next sections. Finally, there is another boson with spin $J = 0$ which is the Higgs (H), responsible for giving their corresponding masses to the rest of SM particles through the Higgs mechanism, as discussed later.

1.1.1 QUANTUM ELECTRODYNAMICS

Quantum Electrodynamics (QED) was historically the first successful attempt to describe a fundamental interaction (electromagnetism, in this case) within a quantum-relativistic frame. It was developed between the late 40’s and the 50’s by Feynman, Schwinger and Tomonaga [10–12]. It is constructed as invariant under phase transformations,

$$\psi \rightarrow e^{iQ\theta} \psi, \quad (1.1)$$

where ψ is the Dirac spinor, the plane-wave solution of the free Dirac equation[†]:

$$(i\gamma^\mu \partial_\mu - m) \psi = 0. \quad (1.2)$$

In Eq. (1.2), γ^μ are the Dirac matrices and ∂_μ is the partial derivative, expressed both in covariant notation. In order to make the Lagrangian invariant under gauge transformations, a modification must be made in order to promote the global symmetry given by Eq. (1.1) to a local one, with $\theta = \theta(x)$. This is done by replacing ∂_μ by the *covariant derivative*,

$$\partial_\mu \rightarrow D_\mu \equiv \partial_\mu - ieQA_\mu, \quad (1.3)$$

where the four-potential $A_\mu = (\Phi, -\mathbf{A})$ describes the interaction with an external electromagnetic potential and is defined such that remains invariant under gauge transformations.

[†]Natural units are used in the following, i.e. $c = 1$ and $\hbar = 1$.

The complete QED Lagrangian has then the form

$$\mathcal{L}_{\text{QED}} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi - eQ\bar{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \quad (1.4)$$

where $\bar{\psi}$ denotes the conjugate of ψ . The last term in Eq. (1.4) is a kinematic factor that has to be added as a consequence of the inclusion of the electromagnetic potential, with $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. The second term of the equation represents the interaction with the field A_μ , which is identified with the photon. A mass term of the form $m^2 A_\mu A^\mu$ would violate the gauge invariance of the Lagrangian, hence the photon has to be massless. The quantity e represents the coupling with the photon, measured experimentally, which increases with the energy scale. This is usually expressed in terms of α , the *fine structure* constant, which at low energies is

$$\alpha = \frac{e^2}{4\pi} \simeq \frac{1}{137}. \quad (1.5)$$

1.1.2 QUANTUM CHROMODYNAMICS

The strong interaction is responsible for bounding protons and neutrons in the nucleus, due to the interactions between the quarks and gluons that compose them, as well as for the clustering of the quarks into hadrons and mesons. Quantum chromodynamics (QCD) is the QFT that describes such interaction. The theory, which was developed by Gell-Mann and Fritzsche in the 70's [13], is built with the inclusion of a new quantum number, the *color*, which can take three different values (red, green and blue) and only quarks and gluons can have it. If q_i is the field of a quark with color i , then the QCD Lagrangian is constructed such that is invariant under the following gauge transformation:

$$q_i \rightarrow e^{i\left(\frac{\lambda^a}{2}\theta(x)\right)_{ij}} q_j, \quad (1.6)$$

where λ^a are the so-called *Gell-Mann matrices*, the generators of the $SU(3)_C$ group previously mentioned, with $a = 1, \dots, 8$. Due to the non-abelian character of the $SU(3)_C$ group, the Gell-Mann matrices do not commute, i.e,

$$[\lambda^a, \lambda^b] = if^{abc}\lambda^c, \quad (1.7)$$

and f^{abc} are the substructure constants of QCD. Like in QED, the symmetry is introduced by including the covariant derivative into the Lagrangian, which in this case takes the form

$$D_\mu = \partial_\mu + ig_s \frac{\lambda^a}{2} A_\mu^a. \quad (1.8)$$

In Eq. (1.8), g_S is the coupling constant of the QCD interaction and A_μ^a are the eight fields that represent the gluons. The QCD Lagrangian then looks like

$$\mathcal{L}_{\text{QCD}} = \bar{q}_i \left[i (\gamma^\mu D_\mu)_{ij} - m \delta_{ij} \right] q_j - \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu}, \quad (1.9)$$

with $F_{\mu\nu}^a$ as the gluon strength tensor, analogous to the QED case and defined as

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g_S f^{abc} A_\mu^b A_\nu^c, \quad (1.10)$$

where the last term represents the gluon-gluon interaction, and is a consequence of the non-commutative character of the Gell-Mann matrices, the later caused by the non-abelian nature of the symmetry group, as mentioned. Gluons coupling themselves are the cause of the the color confinement, previously mentioned. When separating two quarks from each other, the gluon self-coupling terms in the Lagrangian would lead to a cloud of virtual gluons carrying color charge around the individual quark, and increasing the potential between the two quarks as the distance increases. At some point, the energy will be large enough to create a new quark-anti-quark pair. This is the reason of the quark confinement.

The coupling constant is usually expressed in terms of the strong coupling constant α_S as

$$\alpha_S = \frac{g_S^2}{4\pi}, \quad (1.11)$$

similarly as the QED fine structure constant. This quantity also depends on the energy scale of the interaction such that the coupling strength increases as the scale of energy raises (or, equivalently, when the distance between quarks decreases). The Renormalization Group Equation (RGE) allows to determine the running of α_S with the scale μ^2 . At one-loop, the solution of the RGE yields [14]

$$\alpha_S(\mu^2) = \frac{12\pi}{(11n - 2f) \ln \left(\frac{\mu^2}{\Lambda_{\text{QCD}}^2} \right)}, \quad (1.12)$$

where n is the number of colors (3 in the SM), f is the number of quark flavours (6 in the SM) and Λ_{QCD} is a constant that determines the scale at which the denominator of Eq. (1.12) blows up and α_S diverges. Λ_{QCD} can be interpreted then as a limit for the validity of perturbation theory in α_S . This behaviour of the QCD coupling constant at high energies is called *asymptotic freedom* and is a major characteristic of the strong interaction, because it implies that quarks and gluons behave almost as free particles at very small distances, and therefore they can be described by using perturbative approaches at very hard regimes. Figure 1.1 shows a recent measurement of α_S as a function of the transferred mo-

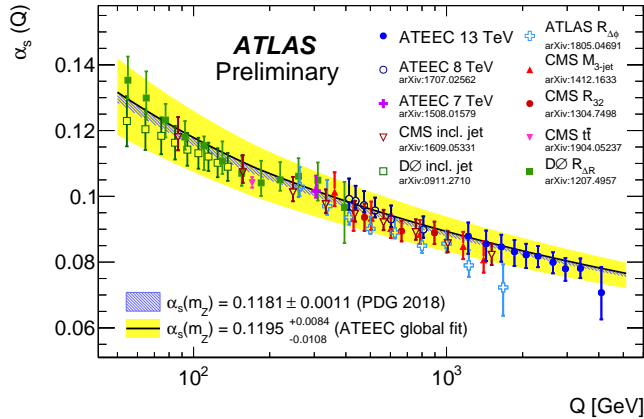


Figure 1.1: Measurement of the strong coupling constant as a function of the transferred momentum Q by the ATLAS experiment at $\sqrt{s} = 13$ TeV and compared with previous observations [15].

mentum Q by the ATLAS experiment, where both the asymptotic freedom and the color confinement features can be inferred from the behaviour of the coupling constant at high and low scales.

1.1.3 ELECTROWEAK INTERACTION AND THE HIGGS MECHANISM

The first attempt to describe the weak interactions came by the hand of E. Fermi in 1933 [16], when trying to explain the β -radiation of the nuclei. According to his theory, the neutron decayed into a proton plus an electron, emitting also an electron anti-neutrino in the process. The strength of Fermi's weak interaction was given by the coupling constant G_F , actually known as the Fermi constant, and its predictions were in good agreement with the observations although it was not renormalizable.

The actual formulation of the weak interaction as a QFT was done by Glashow, Weinberg and Salam later in the 60's [2, 3, 17] and it came already unified with QED, this is why it is known as the Electroweak (EW) theory. As mentioned before, the EW theory is built under the symmetry group $SU(2)_L \otimes U(1)_Y$, where $U(1)_Y$ is a one-dimensional symmetry group generated by the weak hypercharge \hat{Y} and refers to the electromagnetic part of the interaction, governed by the photon. The group $SU(2)_L$, the pure weak part of the interaction, is a three-dimensional group, so it has three generators given by the weak isospin components $T_i = \sigma_i/2$ (with $i = 1, 2, 3$), where σ_i are the Pauli matrices. The new quantum number for this part of the interaction is then the weak isospin, which is 1/2 for left-handed particles (right-handed anti-particles) and 0 for right-handed (left-handed anti-particles).

The left- and right-handed chiralities (denoted by ψ_L and ψ_R) are defined as

$$\psi_{L/R} = \frac{1 \mp \gamma^5}{2} \psi, \quad (1.13)$$

where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. As a consequence of their different weak isospin, left-handed particles will behave as doublets, $\psi_L^i = \begin{pmatrix} \nu_L^i \\ \ell_L^i \end{pmatrix}$, $\begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix}$, and right-handed particles will do as singlets, i.e. $\psi_R^i = \ell_R^i, u_R^i, d_R^i$, in a similar way as bosons and fermions behave differently in relation to the spin. As a consequence, only left-handed fermions will participate in the weak interaction when they transform under the $SU(2)_L \otimes U(1)_Y$ group:

$$\begin{aligned} \psi_L &\rightarrow e^{i[\frac{1}{2}g \boldsymbol{\alpha}(x) \cdot \boldsymbol{\sigma} + g' \beta(x) Y]} \psi_L \\ \psi_R &\rightarrow e^{ig' \beta(x) Y} \psi_R \end{aligned} \quad (1.14)$$

In Eq. (1.14), $\boldsymbol{\alpha}(x)$ and $\beta(x)$ are arbitrary local phases, the vector $\boldsymbol{\sigma}$ contains the Pauli matrices σ_i and g and g' are the coupling constants of the $SU(2)_L$ and $U(1)_Y$ groups, respectively.

Again, the invariance of the Lagrangian under Eq. (1.14) is ensured by introducing the covariant derivative, which in this case is defined as

$$D_\mu = \partial_\mu - ig\mathbf{T} \cdot \mathbf{W}_\mu - ig' \frac{1}{2} B_\mu, \quad (1.15)$$

where the isotriplet field $\mathbf{W}_\mu = (W_{1\mu}, W_{2\mu}, W_{3\mu})$ and the singlet field B_μ are introduced as the mediators for the $SU(2)_L$ and $U(1)_Y$ parts of the interaction, respectively. The EW Lagrangian takes then the form

$$\mathcal{L}_{\text{EW}} = \bar{\psi}_L (i\gamma^\mu D_\mu) \psi_L + i\bar{\psi}_R \gamma^\mu \left(\partial_\mu - ig' \frac{1}{2} B_\mu \right) \psi_R - \frac{1}{4} \mathbf{W}^{\mu\nu} \cdot \mathbf{W}_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu}, \quad (1.16)$$

where the last two terms are the kinetic terms corresponding to the bosons of the two interactions,

$$\begin{aligned} W_i^{\mu\nu} &= \partial^\mu W_i^\nu - \partial^\nu W_i^\mu + g\epsilon_{ijk} W_j^\mu W_k^\nu \\ B^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu \end{aligned} \quad (1.17)$$

The actual observed bosons (γ , W^{\pm} and Z) are linear combinations of $W_{i\mu}$ and B_{μ} :

$$\begin{pmatrix} A_{\mu} \\ Z_{\mu} \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_{\mu} \\ W_{3\mu} \end{pmatrix} \quad (1.18)$$

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

The parameter θ_W added in Eq. (1.18), also known as Weinberg angle, relates the couplings g and g' as $\tan \theta_W = g/g'$. The couplings g and g' can be related to the electromagnetic coupling, e in Eq. (1.4), via $e = g \sin \theta_W = g' \cos \theta_W$. Recent experimental results [9] indicate values of $\sin^2 \theta_W = 0.23121 \pm 0.00002$. On the other hand, the weak coupling g relates to the Fermi constant as

$$G_F = \frac{\sqrt{2}g^2}{8m_W^2}, \quad (1.19)$$

where the mass of the W boson, m_W , enters in the definition.

In the EW Lagrangian given by Eq. (1.16) there are no mass terms for the gauge bosons since that would violate the gauge invariance, but this is in conflict with the experimental observations which suggest otherwise ($m_W = 80.379 \pm 0.012$ GeV and $m_Z = 91.1876 \pm 0.0021$ GeV from recent measurements [9]). Moreover, the fermionic mass terms are not gauge-invariant either due to the dependence on the chirality. This controversial is fixed by the Higgs mechanism, explained as follows.

A solution for the inclusion of the mass terms in the Lagrangian was found simultaneously by several authors [4–6], and it is based on a Spontaneous Symmetry Breaking (SSB) mechanism. In this scenario, a new doublet of the $SU(2)_L \otimes U(1)_Y$ group is introduced,

$$\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (1.20)$$

where the ‘+’ and ‘0’ labels indicate the electric charge of the field. The Lagrangian corresponding to the field ϕ will take the form

$$\mathcal{L}_{\phi} = (D^{\mu}\phi)^{\dagger} (D_{\mu}\phi) - V(\phi), \quad (1.21)$$

where $V(\phi)$ is a potential associated to ϕ which depends on two parameters, μ and λ :

$$V(\phi) = \mu^2 \phi^{\dagger} \phi + \lambda (\phi^{\dagger} \phi)^2. \quad (1.22)$$

If the parameters are chosen such as $\mu^2 < 0$ and $\lambda > 0$, the potential $V(\phi)$ has a minimum in

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}, \quad (1.23)$$

i.e., this potential would have a minimum on a circle of radius $v = \sqrt{-\mu^2/2\lambda}$ around $\phi = 0$. This result implies a vacuum expectation value different from zero: $\langle 0|\phi|0\rangle = v/\sqrt{2}$. According to the Goldstone theorem [18], a non-zero expectation value of a field implies the appearance of a massless boson (usually called *Goldstone bosons*). Given that the ground state of ϕ is degenerated, a choice of λ and μ can be made such that only the neutral component of ϕ has a non-zero expectation value[‡],

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.24)$$

Applying now a perturbative expansion around the vacuum state,

$$\phi(x) = e^{i\frac{\sigma\cdot\theta(x)}{v}} \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}, \quad (1.25)$$

four independent fields ($\theta(x)$ and $H(x)$) are introduced to parametrize the fluctuations around ϕ_0 . The field $H(x)$ is a real scalar field (referred to as the Higgs field) that represents small perturbations around the vacuum expectation value. The three other fields $\theta_i(x)$ are massless Goldstone bosons that will be reabsorbed in the redefinition of the gauge bosons, which will get then get their masses and an extra polarization. Therefore they can be removed from the Lagrangian and the resulting field can be simplified as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \quad (1.26)$$

Substituting Eq. (1.24) in Eq. (1.21) and writing the fields in terms of A_μ, Z_μ and W_μ^\pm via Eq. (1.18), the boson masses can be expressed as a function of the couplings and the vacuum expectation value:

$$\begin{aligned} m_W &= \frac{1}{2}vg \\ m_Z &= \frac{1}{2}v\sqrt{g^2 + g'^2}, \end{aligned} \quad (1.27)$$

[‡]This particular choice is conveniently done as it ensures the invariance under the $U(1)_Y$ group, keeping the photon as a massless boson.

and the excitation of the field $H(x)$ gives the scalar Higgs boson with its mass,

$$m_H = \sqrt{2\lambda}v. \quad (1.28)$$

The vacuum expectation value can be estimated by measuring m_W , and it is $v \simeq 246$ GeV, value that sets the scale for the EW symmetry breaking. The mass of the Higgs boson, however, can not be predicted because λ is a free parameter.

Finally, the scalar Higgs field is used to generate the quark and lepton masses. This is achieved by adding the so-called Yukawa term to the EW Lagrangian,

$$\mathcal{L}_Y = \lambda_f^{ij} \left(\bar{\psi}_L^i \phi \psi_R^j + \bar{\psi}_R^i \phi \psi_L^j \right), \quad (1.29)$$

where $i, j = 1, 2, 3$ and the matrices λ_f contain the Yukawa couplings between the Higgs doublet and the corresponding fermion. Substituting the expansion of ϕ given by Eq. (1.25) in Eq. (1.29), the fermion masses can be determined at leading order as

$$m_f = \lambda_f \frac{v}{2}. \quad (1.30)$$

1.2 HADRON COLLISIONS

Predictions from the SM must be cross-checked with empirical observations, and this is usually done with particle collision experiments. For this reason, it is necessary to understand all the physics involved in such collisions, essentially because the way to compare the SM predictions with the data is by generating simulated collision events, and it needs to be done accurately. In particular, the analysis of this thesis is done with data from proton–proton (pp) collisions at the Large Hadron Collider. Since protons are composite particles (formed by three quarks: uud), the physics involved in the collisions is governed by the interactions between the quarks and gluons that compose them (mainly QCD), so it is important to discuss first the internal structure of the proton.

1.2.1 PROTON STRUCTURE

In order to study the initial state of the proton collision, protons are described as a combination of *partons*, a name coined by Feynman in the 60's to refer to the inner constituents of the proton when trying to understand Bjorken's scaling [19], nowadays identified with quarks and gluons.

In the parton model, each parton i carries a fraction x of the proton momentum P , with a proba-

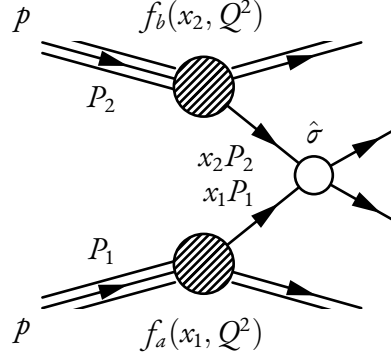


Figure 1.2: Feynmann diagram of a generic proton–proton collision, separating the hard and soft parts of the interaction. The cross-section of the hard process is denoted by $\hat{\sigma}$ and the soft processes are included in the PDFs f_1 and f_2 .

bility described by the so-called *parton distribution function* (PDF), $f_i(x, Q^2)$, where Q^2 is the squared transferred momentum between the two protons involved in the collision and represents the energy scale of the process. According to the factorization theorem [20], the interaction can be factorized into a *hard* part, which can be described with perturbative QCD and depends on the renormalization scale μ_R , and a *soft* interaction that includes non-perturbative contributions and is separated by a factorization scale μ_F . This separation is illustrated in the diagram of Figure 1.2. Then, the cross-section for a typical scattering process between two protons with momenta P_1 and P_2 , at a center-of-mass energy of $s = (P_1 + P_2)^2$, is expressed as

$$\sigma_{pp \rightarrow X}(x_1, x_2, Q^2, \sqrt{s}) = \sum_{a,b} \int dx_1 \int dx_2 f_a(x_1, \mu_F^2) f_b(x_2, \mu_F^2) \hat{\sigma}_{ab \rightarrow X}(x_1 P_1, x_2 P_2, \mu_F^2, \mu_R^2, Q^2, \sqrt{s}), \quad (1.31)$$

where the sum runs over all partons a and b of the incoming protons. The cross-section $\hat{\sigma}_{ab \rightarrow X}$ corresponds to the hard part of the interaction, so it depends on μ_R and can be calculated perturbatively. Such calculation is usually known as *matrix element* (ME) calculation. On the other hand, the non-perturbative contributions are parametrized by the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Paris) equations [21]:

$$\frac{d}{dQ^2} f_i(x, Q^2) = \sum_j \frac{\alpha_S(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{j \rightarrow i}(z) f_j(x/z, Q^2), \quad (1.32)$$

where $P_{i \rightarrow j}(z)$ is called the *splitting kernel*, and represents the probability of a parton i becoming a parton j carrying a fraction z of its original momentum.

The parameters of the PDF are typically measured at a certain Q^2 in deep inelastic scattering (DIS) experiments, in which lepton-proton collisions are carried out to probe the internal structure of the proton. Figure 1.3 shows the proton PDFs for values of Q^2 (equivalently, μ) of 10 and 6500 GeV, as

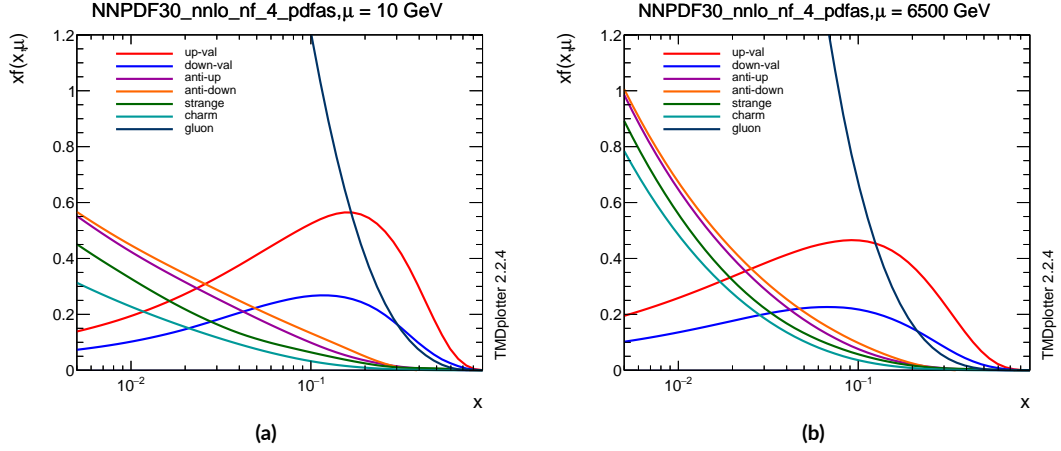


Figure 1.3: Proton PDFs for $\mu = 10 \text{ GeV}^2$ (a) and $\mu = 6.5 \text{ TeV}$ (b), with the NNLO NNPDF3.0 global analysis [22]. Valence u and d quarks are shown together with gluons and the sea quarks s , c , \bar{u} and \bar{d} . Produced with TMDplotter [23].

calculated by the NNPDF group [22]. In the figure is shown that the valence quarks u and d are the ones carrying the largest fraction of momentum (about half of it), while at low- x regimes virtual gluons dominate together with the sea quarks, where higher-flavour ones such as s and c also contribute.

1.2.2 EVENT SIMULATION

Simulated pp collisions are generated starting from the hard process ME calculation as described by Eq. (1.31). This calculation is performed at a fixed level of accuracy in perturbative QCD. Leading order (LO) calculations are those for which only tree-level Feynman diagrams are included, next-to-leading order (NLO) refers to calculations including one-loop diagram, etc. Computationally, this is addressed with Monte Carlo (MC) techniques, which rely on the use of pseudo-random numbers to resolve integrals numerically.

Apart from the description of the proton structure with the PDFs, discussed in the previous section, the collision itself is a very complex process, as illustrated in Fig. 1.4. In the figure, the hard scattering process is represented by red blobs. Initial- or final-state quarks or gluons can undergo QCD Bremsstrahlung, developing the so-called *parton shower* (PS), drawn in pink. Eventually, the showered particles will collapse forming hadrons, a process known as *hadronization* indicated in green in the figure, which can even decay further. Moreover, proton remnants can also interact, which is called the *underlying event*, shown in purple. All these steps, discussed with more detail as follows, are described independently with the MC generators.

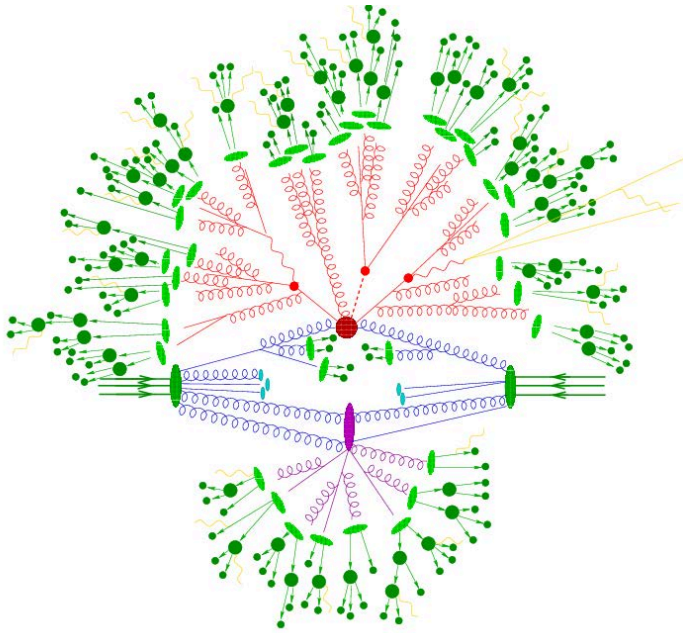


Figure 1.4: Schematic view of the physics implemented in standard multi-purpose event generators for pp collisions [24].

PARTON SHOWER

PS is included in the MC generation as an approximation for higher-order processes that were not included in the QCD calculation, in order to emulate a more complete final state. As mentioned, it consists on the emission of soft initial- or final-state partons, and is the origin of the jets that are later observed in the experiments.

Processes such as $q \rightarrow qg$, $g \rightarrow gg$ and $g \rightarrow q\bar{q}$ are simulated by using the DGLAP equations and included iteratively until the Λ_{QCD} scale is reached (see Eq. (1.12)), which represents the limit for the validity of the perturbative calculations. This simulation is approximate since it assumes that parton emissions are completely independent and they do not interact with each other. The parton emitted from the hard process can then split into a collimated shower of partons, with a probability given by the splitting kernels $P_{j \rightarrow i}$ of Eq. (1.32). On the other hand, the probability of the parton to evolve without branching is given by the so-called *Sudakov factors* [25]. The evolution of the cascade until the lower cut-off scale is reached can be described by using different variables, such as Q^2 , the emission angle of the parton, or its transverse momentum.

The PS implementation can be applied to initial-state partons, known as initial-state radiation (ISR) and final-state partons as well, referred to analogously as final-state radiation (FSR). In both

cases the simulation is equivalent, although it is applied backwards in time for the ISR case, i.e. the parton momentum is increased until it matches that of the value given by the PDF.

Once the PS simulation is done it has to be matched with the ME calculation, performed in the previous step of the event generation, in order to avoid potential double counting in some regions of the phase space (which can happen if the ME calculation includes the emission of one or more additional partons). This combination is not straightforward in general. There are two main ME–PS matching approaches: the CKKW (Catani-Krauss-Kuhn-Webber) procedure [26] and the MLM (Michelangelo L. Mangano) algorithm [27].

The CKKW algorithm relies on the jet resolution as defined in the jet k_t algorithm[§], which decides whether to cluster two objects together depending on a distance parameter that depends on the energy of the objects. If the resolution is larger than a certain threshold, the cross-sections and distributions are given by the ME and weighted by the Sudakov factors; otherwise, they are determined by PS, which provides a better description of the internal jet structure. The MLM procedure, on the other hand, starts by separating the events in exclusive samples with a fixed number of partons in the final state, and then the PS is performed. After that, the parton configuration is processed with a cone jet-clustering algorithm and, if the number of reconstructed jets is equal to the number of partons from the ME calculation, the event is accepted and otherwise is rejected.

HADRONIZATION

The hadronization is the next step in the generation and concerns the collapsing of the outgoing showers of gluons and quarks into hadrons and mesons due to the QCD color confinement effect. Two different theoretical approaches are commonly adopted to describe this process:

- The *string fragmentation* model [28] is based on the assumption of a linear confinement. This means that the confinement potential is seen as a string between quarks, whose tension raises as the separation distance increases until the energy of the string reaches the mass of a $q\bar{q}$ pair, then the string breaks into smaller strings along its length and a color singlet pair is created. In this model, gluon emissions represent kinks on the strings. The string fragmentation is repeated until all the energy is converted into quark pairs.
- The *cluster fragmentation* model [29] forces the gluons from the PS to branch into $q\bar{q}$ pairs, and clusters are formed with the neighbouring pairs. The clusters decay eventually into pairs of baryons depending on the available energy.

[§]Jet clustering algorithms will be discussed in Chapter 4.

The produced hadrons, which may decay further, are finally combined into jets, reproducing approximately the kinematics of the original partons.

UNDERLYING EVENT

The underlying event (UE) refers to the interaction between the residual proton remnants, as mentioned. These kind of processes can not be calculated in QCD, so data-based phenomenological models [30, 31] are used instead, which must be tuned later to match with real observed events. In addition, it might happen that a residual but non-negligible contribution from hard interactions is present in the UE. This is referred to as multiple parton interactions (MPI), and are typically described with a combination of perturbative calculations, also tuned with experimental data.

1.3 BEYOND THE STANDARD MODEL

The SM provides a very successful description of all the known phenomena in high energy physics. However, there is a number of open questions emerging from the theory, together with a wide range of empirical observations that are suggesting that a more general theoretical framework could be necessary.

First of all, the SM does not explain the gravitational interaction, as pointed out at the beginning of the chapter. So far, gravity has been well described by Einstein's General Relativity [32], including successful predictions such as the Gravitational Waves (GW), finally observed by the LIGO and Virgo collaborations [33] one hundred years after predicted. Quantum gravitational effects would be expected at the Planck scale ($M_P \sim (8\pi G)^{-1/2} = 2.4 \cdot 10^{18}$ GeV), and therefore a new model would be needed at such regimes, but the description of gravity as a renormalizable QFT is a mathematical problem that theorist have no been able to solve yet.

The so-called *hierarchy problem* arises from the huge difference between the Planck and EW scales ($M_{EW} \sim 10^2$ GeV), which makes the Higgs boson sensitive to big corrections to its mass via loop contributions from any particle up to 17 orders of magnitude more massive that couples to it. This happens because there is no symmetry protecting the mass of the Higgs, unlike fermions and gauge bosons. Although this does not make the theory inconsistent, it is an uncomfortable feature as it requires a fine tuning, which is not considered natural. In fact, the masses of all particles in the SM can not be predicted since the model has many free parameters (19), so they have to be measured experimentally, and there is no explanation for the origin of the generations of fermions and gauge symmetries.

The existence of dark matter (DM) has been proved empirically by astronomical observations, such as the measurement of rotational velocities of galaxies [34, 35], gravitational lensing, or the multipolar analysis of the fluctuations in the temperature of the cosmic microwave background (CMB), provided by the WMAP and Planck missions (the later shown in Fig. 1.5). However, the SM does not provide any explanation for this kind of matter. In fact, the latest cosmological measurements [36, 37] reveal that ordinary matter, the one described by the SM, represents only less than 5% of the universe, while DM makes up about 27% of it. The 68% left is the so-called *dark energy* (DE), predicted by the Standard Model of Cosmology (usually referred to as the Λ CDM model) as the cause behind the accelerated expansion of the universe, observed in 1998 by looking at in high-redshift[‡] supernovae of type Ia [38, 39].

Neutrino flavour oscillations have been extensively observed [40], but this enters in contradiction with the SM since it would require neutrinos to have non-zero mass, as Pontecorvo pointed out first in 1957 [41, 42]. Therefore, the SM would need to be extended in order to account for this. However, this is rather an open question in the SM than an indication of new physics, as it is related to whether neutrinos are Dirac or Majorana particles.

The asymmetry between matter and anti-matter that is observed in the universe can not be explained with the SM. It includes *CP* violation but it is not enough to account for such level of imbalance, especially because QCD seems to preserve this symmetry for no particular reason. This is known as the *strong CP problem*. Experiments like Belle and LHCb have been reporting results compatible with *CP* violation beyond the SM predictions in the last years, hence there is an effort to look for additional sources of *CP*-violation, but not successfully yet.

Additionally, two different experiments have released recently very exciting results: first, LHCb reported evidences of lepton universality violation [43] (shown in Fig. 1.5), and the Muon $g-2$ experiment published a new measurement of the anomalous momentum of the muon, which combined with previous results turns out to differ from the SM predicted value by more than 4σ [44]. In both cases, evidence of new physics was claimed, although they are still under discussion as theorists say such discrepancies are caused by the lack of higher-order corrections in the theoretical predictions.

Altogether, there is a fairly decent amount of arguments pointing towards the necessity of new models that can complete the SM and explain all these observations that SM can not. All these scenarios are usually called *Beyond Standard Model* (BSM) theories. In the analysis of this thesis, a number of BSM models are tested with data from *pp* collisions. Such models are briefly described below.

[‡]Redshift refers to the shift between the emitted wavelength of an object, λ_e , and the one measured by the observer, λ_o , caused by the Doppler effect. It is defined as $1 + z \equiv \lambda_o/\lambda_e$.

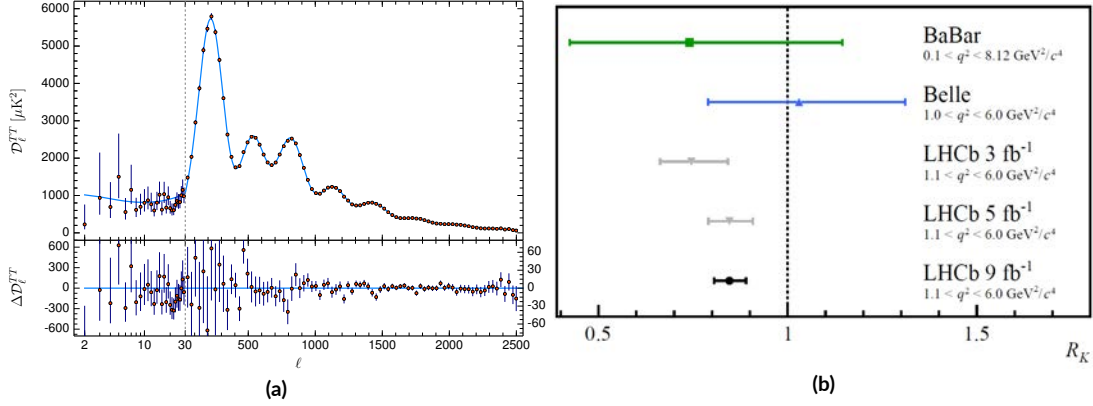


Figure 1.5: (a) Power spectrum of from the CMB temperature fluctuations measured by the Planck mission (ESA) [36]. (b) Lepton universality conservation violation as measured by LHCb compared with previous measurements [43].

1.3.1 LARGE EXTRA DIMENSIONS

In 1919, a description of both electromagnetism and gravitation (the two known interactions at that time) was proposed by T. Kaluza [45], based on the inclusion of an extra spacial dimension. In his theory, General Relativity works in a 5-dimensional spacetime, R_5 , parametrized by the metric tensor

$$\tilde{g}_{ab} = \begin{pmatrix} g_{\mu\nu} + \phi^2 A_\mu A_\nu & \phi^2 A_\nu \\ \phi^2 A_\nu & \phi^2 \end{pmatrix}, \quad (1.33)$$

where $g_{\mu\nu}$ is the usual 4-dimensional metric, ϕ an scalar field and A_μ a vector field identified with the electromagnetic potential. Everyday world is then just a four-dimensional part of R_5 , conveniently constrained by the so-called *cylindrical conditions*, i.e. the derivatives of all physical observables with respect to the new parameter vanish ($\partial\tilde{g}_{ab}/\partial x^5 = 0$), or at least they are considered to be small as they are of higher order. This explains the weakness of gravity in relation to electromagnetism and solves the hierarchy problem as well.

A quantum interpretation of Kaluza's theory was given by O. Klein in 1926 [46, 47], who considered the fifth dimension as compactified within a circle of radius R . Being that dimension closed, a massless scalar field $\Phi(x^\mu, x^5)$ will satisfy the periodic condition $\Phi(x^\mu, x^5) = \Phi(x^\mu, x^5 + 2\pi R)$, and the Fourier expansion of the field will yield

$$\Phi(x^\mu, x^5) = \sum_{n=-\infty}^{+\infty} \Phi_n(x^\mu) e^{inx^5/R}. \quad (1.34)$$

Then, an electric charge moving in the 5-th dimension can be described as a combination of stand-

ing waves, one for each mode of Eq. (1.34), and therefore its momentum's 5-th component will be quantized as $P_5 = n/R$. This implies that for any field one would find infinite wave functions, the so-called *Kaluza-Klein towers*, which applies not only for scalar fields but also for fields with non-zero spin like the graviton (spin $J = 2$). Using a 'de Broglie' relation of the type $P_5 = h/R$, Klein obtained the following expression for the 0-th mode of such waves:

$$R = \frac{hc\sqrt{2G}}{q} \simeq 0.8 \cdot 10^{-30} \text{ cm}, \quad (1.35)$$

where h and G are the Planck and Newton constants, respectively, c is the speed of light and q is the electron charge. The small value of the radius obtained is taken as a support of the cylindrical conditions in Kaluza's theory.

In 1998, an extra-dimensions model was proposed by Arkani-Hamed, Dimopoulos and Dvali [48]. The so-called **ADD model**, named after the name of its authors, is an effective field theory (EFT)^{||} in which the cut-off is set at the EW scale, and therefore yields to measurable effects at TeV energies. In this framework, n extra dimensions are proposed, all of them confined within a circle of radius R . A higher-dimensional space named as *bulk* where the gravitational interaction can propagate via its mediator, the graviton, while all the SM fermions and bosons are bound to the known 4 dimensions, denoted as *branes*.

In this context, the gravitational potential between two masses m_1 and m_2 takes the form

$$V(r) \sim \frac{m_1 m_2}{M_D^{n+2}} \times \begin{cases} 1/r^{n+1} & (r < R) \\ 1/(R^n r) & (r \gg R) \end{cases}, \quad (1.36)$$

hence the effect of the additional dimensions is only noticeable when the distances between the masses are small enough in comparison with R , while the potential recovers the usual $\sim 1/r$ behaviour at distances larger than R . In Eq. (1.36), M_D represents the effective Planck scale in the $4+n$ dimensional space, related to the 4D Planck scale as

$$M_P^2 \sim M_D^{n+2} R^n. \quad (1.37)$$

The ADD model becomes especially interesting when R is small enough such that M_D is close to M_{EW} , since it would solve naturally the hierarchy problem and also provides predictions that can be

^{||}An EFT is a tool to describe physics related to a more complete theory at energy regimes much lower than a certain cut-off scale.

tested in particle collider experiments. At these regimes, the typical radius can be estimated as

$$R \sim 10^{30/n-17} \times \left(\frac{1 \text{ TeV}}{m_{\text{EW}}} \right)^{1+2/n} \text{ cm} . \quad (1.38)$$

The $n = 1$ case is automatically excluded as it would imply a radius of about 10^{13} cm, with consequences in the gravitational potential at solar system scales, which would have been observed already. This is why the name *large extra dimensions* (LED), is usually utilized in the context of the ADD scenarios, as only models with $n \geq 2$ extra dimensions are typically considered.

In this framework, the weakness of gravity is explained in terms of a Kaluza-Klein graviton propagating through n extra dimensions. The interaction term in Lagrangian of the graviton field, $G_{\mu\nu}$, is determined by Einstein's equations in $4 + n$ dimensions, and the interaction term takes the form

$$\mathcal{L}_{\text{grav}} = -\frac{\sqrt{8\pi}}{M_P} G^{\mu\nu} T_{\mu\nu} , \quad (1.39)$$

where $T_{\mu\nu}$ is the energy-momentum tensor. The decay rate of the graviton is then suppressed by a factor $\sim M_P^2$, which points to a stable –or, at least, long-lived– particle. Therefore, a graviton produced in a proton–proton collision would escape through the detector without interacting with it, leaving a signature of *missing transverse momentum*. Scenarios where the graviton is produced in association with ISR or FSR jets are highly compatible with the typical signature of the monojet analysis (discussed later in Chapter 6), and therefore it can be used to set exclusion limits on the parameter space of the ADD model. In previous versions of this analysis [49], values of M_D up to 7.7 and 4.8 TeV have been excluded for $n = 2$ and $n = 6$ extra dimensions by the ATLAS collaboration.

1.3.2 DARK MATTER PRODUCTION

As mentioned, there is a non-baryonic type of matter about five times more prevalent than ordinary matter in the universe, as confirmed by different astronomical and cosmological observations over the last century. There are only a few known things about the nature of dark matter [50]:

- DM does not interact via electromagnetism or QCD. Hence the name ‘dark’.
- It interacts gravitationally. It must be at most weakly interacting with SM particles.
- It has to be stable, or at least long-lived in cosmological time scales, since it has survived from the *freeze-out* era (i.e. the moment when dark matter decoupled from the thermal bath). Otherwise it would exhibit a smaller abundance today.

- It must have the right mass and abundance in order to match with the measured relic density, as estimated by Planck [36]:

$$\Omega_c h^2 = 0.1198 \pm 0.0012 \quad (1.40)$$

The only DM candidate SM can provide is the neutrino, as it actually has very similar characteristics. However it has been ruled out since its abundance in the universe is not large enough. Many different candidates have been proposed in different BSM theories, such as sterile neutrinos, axions, light gravitinos, primordial black holes, etc [51], but the most popular ones are the so-called Weakly Interacting Massive Particles (WIMPs).

WIMPs have masses m_χ allowed in the range between 10 MeV and 1 TeV and interact with SM particles with cross-sections similar to that of weak processes. There is an interesting feature of WIMPs, usually referred to as the *WIMP miracle*, which is that if a WIMP exists and is stable, it is naturally produced with a relic density consistent with that required of DM. This also implies that WIMPs must annihilate to other particles. Assuming these are SM particles, there are three ways to detect DM experimentally:

- Direct detection experiments, based on the idea of measuring the nuclear recoil that follows an elastic scattering between a WIMP and the nuclei that compose the active material of a certain detector. Recoil energies depend on the mass of the WIMP and typically range between 1 and 100 keV. Depending on the kind of coupling with the nuclei, two types of interactions are studied: spin-independent, with scalar or vector coupling; and spin-dependent, with pseudo-scalar or axial-vector coupling, sensitive to the spin of the corresponding nucleon. There are many DM direct detection experiments, such as XENON [52], CDMS [53], LUX [54] and CRESST [55].
- Indirect detection experiments, looking for decay products from WIMP-pair annihilation processes. The assumption is that WIMPs can be captured gravitationally by heavy objects like the Sun or the galactic center, where the high local density can make possible the annihilation. Most of the products are immediately absorbed except for the neutrinos, which can be later detected by neutrino telescopes such as MAGIC [56] and Fermi-LAT [57], working on space, or Earth-based ones like IceCube [58], Super-Kamiokande [59] and ANTARES [60].
- Pair-production in colliders. Finally, pairs of WIMP could be created in particle collisions with sufficient center-of-mass energy. If the rate of collisions is high enough, rare processes with very low cross-sections can be produced with enough statistics to be discriminated against other

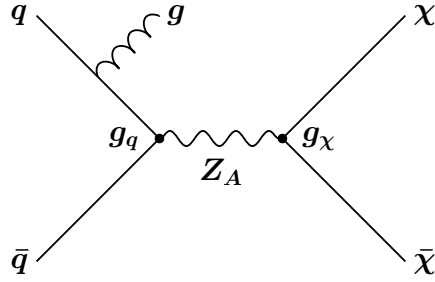


Figure 1.6: Diagram of a DM-pair production in association with one jet, assuming an axial-vector mediator exchange in the s -channel within a simplified model.

processes more usual. In the case of a DM-pair creation, they would escape through the detector leaving a signature in the form of missing momentum, just like neutrinos.

The later is the method studied in the analysis of this thesis. The idea is to exploit the similarity of the monojet signature with respect to events where a pair of WIMPs is produced in association with an ISR/FSR jet. In fact, the monojet analysis is known as the *golden channel* for DM searches in colliders due to its high sensitivity to these kind of processes.

In the analysis, the WIMP-pair production is modelled by using *simplified models*, which rely on the introduction of a new mediator particle to be exchanged between the WIMPs and the SM particles [61, 62]. The considered models assume mediators with minimal decay width, as described in Ref. [63], so only decays into SM particles or WIMPs are allowed. Only Dirac DM particles are considered, since the choice of both Majorana fermions or scalars would produce changes in the kinematic distributions of the visible particle. The models depend then on the following parameters: the masses the DM particles, m_χ , and the mediator, M_{med} , the coupling between the mediator and the partons, g_g and the coupling between the mediator and the dark sector, g_χ . Moreover, the simplified models considered for the DM pair production are the s -channel processes, in which the propagator can be written in the Breit-Wigner form

$$\frac{1}{Q^2 - M_{\text{med}}^2 + iM_{\text{med}}\Gamma}, \quad (1.41)$$

where Q is the transferred momentum and Γ is the width of the mediator. A typical choice is to assume a Z' -like mediator, i.e. a heavier relative of the EW Z boson. This boson arises from an additional $U(1)$ gauge symmetry that is introduced, and its nature (in particular, the spin) affects directly to the kinematic distributions of the final states. Therefore, two different scenarios are considered in the analysis: an axial-vector mediator model and a pseudo-scalar mediator model.

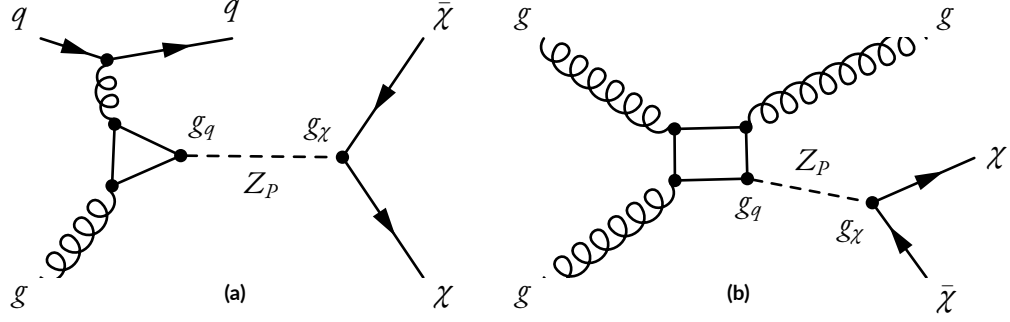


Figure 1.7: One-loop diagrams of a DM-pair production in association with one jet, assuming a pseudo-scalar mediator exchange via s -channel in a simplified model.

The axial-vector mediator simplified model (in the following, the **DMA model**) assumes a spin-1 mediator Z_A exchanged in the production of a DM-pair, as illustrated in Figure 1.6. The Lagrangian for this kind of process takes the form

$$\mathcal{L}_A = g_q \sum_q Z_A^\mu \bar{q} \gamma_\mu \gamma_5 q + g_\chi Z_A^\mu \bar{\chi} \gamma_\mu \gamma_5 \chi \quad (1.42)$$

where the sum runs over all quark flavours ($q = u, d, s, c, t, b$) and m_{Z_A} is the mass of the corresponding mediator ($\mathcal{M}_{\text{med}} \equiv m_{Z_A}$). The coupling g_q is assumed to be universal, i.e. the same for all quarks. As mentioned, the width of the mediator is assumed to be the minimal one, taking the form

$$\Gamma_{\text{min}}^A = \frac{g_\chi^2 m_{Z_A}}{12\pi} \beta_\chi^3 \Theta(m_{Z_A} - 2m_\chi) + \sum_q \frac{3g_q^2 m_{Z_A}}{12\pi} \beta_q^3 \Theta(m_{Z_A} - 2m_\chi), \quad (1.43)$$

where $\Theta(x)$ is the Heaviside step function and $\beta_f = \sqrt{1 - 4m_f^2/m_{Z_A}^2}$ is the velocity of the fermion f with mass m_f in the rest frame relative to the mediator Z_A . If there are no additional visible or invisible decays contributing to the width of the mediator, the minimal width is fixed by the choices of g_q and g_χ . A scan over the couplings was performed at the beginning of Run 2 with simulated DMA samples generated at NLO [61], and it was found that the choice $g_q = 0.25$ and $g_\chi = 1$ is optimal to cover a wide range of the m_χ - m_{Z_A} plane, since the kinematic distributions do not change significantly with the coupling values for fixed masses and setting $g = 0.25$ helps suppressing constrains from di-jets events.

Similarly, a model with a pseudo-scalar spin-0 mediator is also considered (the **DMP model**). Assuming minimal flavour violation (MFV), such particle would behave in a similar manner as the Higgs boson, and therefore one could include mixing terms with the Higgs sector. But this would lead to a much more complex phenomenology with respect to what concerns to the analysis of this

thesis. Then, for simplicity the SM extension to include the mediator will not consider any mixing with the Higgs, and therefore the processes that derive from the pp collisions are loop-suppressed as illustrated in the diagrams of Figure 1.7. In this case, the Lagrangian looks like

$$\mathcal{L}_P = \frac{ig_q}{\sqrt{2}} Z_P \sum_i \left(\lambda_i^u \bar{u}_i \gamma_5 u_i + \lambda_i^d \bar{d}_i \gamma_5 d_i + \lambda_i^\ell \bar{\ell}_i \gamma_5 \ell_i \right) + ig_\chi Z_P \bar{\chi} \gamma_5 \chi, \quad (1.44)$$

where i runs over fermion generations, Z_P is the pseudo-scalar mediator and the Yukawa couplings λ_i^f (being f the fermion) are normalized to the Higgs vacuum expectation value, i.e. $\lambda_i^f = \sqrt{2} m_i^f / v$. Again, universality is assumed for the coupling with the fermions g_q . The minimal mediator width, neglecting the small contributions from other quarks than t , is expressed as

$$\Gamma_{\min}^P = \sum_f N_C \frac{y_f^2 g_q^2 m_{Z_P}}{12\pi} \left(1 - \frac{4m_f^2}{m_{Z_P}^2} \right)^{1/2} + \frac{g_\chi^2 m_{Z_P}}{8\pi} \left(1 - \frac{4m_f^2}{m_{Z_P}^2} \right)^{1/2} + \frac{\alpha_s^2 y_t^2 g_q^2 m_{Z_P}^3}{32\pi^3 v^2} \left| f_P \left(\frac{4m_t^2}{m_{Z_P}^2} \right) \right|^2, \quad (1.45)$$

where the loop integral, with f_P as a complex function, is defined as

$$f_P(x) = x \arctan^2 \left(\frac{1}{\sqrt{x-1}} \right), \quad (1.46)$$

being $x \equiv 4m_t^2 / m_{Z_P}^2$. A similar parameter scan as done for the axial-vector model was performed with DMP simulated samples, revealing the optimal choice of $g_q = g_\chi = 1$ in this case.

1.3.3 SUPERSYMMETRIC QUARK-PAIR PRODUCTION

Supersymmetry (SUSY) [64–67] is a concept born in the 70's which essentially proposes a new symmetry between fermions and bosons. A supersymmetric theory is one that is invariant under such symmetry, and the SM is not. Therefore SUSY models predict the existence of new particles complementing the SM in such a way that this symmetry is conserved. These kind of theories not only provide an elegant solution to the hierarchy problem, but also provide dark matter candidates and even allow for the inclusion of the missing gravitational interaction. For these reasons, Supersymmetry has become quite popular over the years and many experiments have been dedicated to probe predictions from many different SUSY models. However, no evidence of SUSY has been found up to the date.

The generator of the proposed symmetry, \hat{Q} , is an anti-commuting spinor with spin-1/2 that transforms fermions into bosons and vice-versa like $\hat{Q}|\text{fermion}\rangle \propto |\text{boson}\rangle$ and $\hat{Q}|\text{boson}\rangle \propto |\text{fermion}\rangle$. The operator \hat{Q} commutes with all the generators of the SM gauge symmetries, which implies that, for each SM particle, the corresponding super-partner (sparticle) will have the same quantum numbers under the $SU(2)_L \otimes U(1)_Y \otimes SU(3)_C$ group.

The complete set of particles and sparticles of a SUSY theory can be grouped into *supermultiplets* as follows. Chiral supermultiplets are formed by a complex scalar field, composed by two scalars, and a Weyl fermion** with two helicity states. SM fermions plus their scalar super-partners enter in this kind of groups. The gauge supermultiplets are formed by a massless vector field with two chirality states and a Weyl fermion with two helicity states ('gauginos'). These groups include the SM gauge bosons and their fermionic super-partners. Finally, a gravitational supermultiplet can be defined as well, composed by the spin-2 graviton and the spin-3/2 gravitino. This later group is only included in some SUSY models.

The simplest extension of the SM to make it a SUSY theory is usually called Minimal Supersymmetric Standard Model (MSSM). The number of extra particles introduced then is just the minimum necessary, with no additional interactions. Table 1.2 summarizes the particle content of the MSSM. Usually, the super-partners of the SM fermions are named by adding an 's-' prefix at the beginning of the name of the corresponding SM particle (e.g. top \rightarrow stop). On the other hand, super-partners of the SM gauge bosons are named by their SM name plus an '-ino' suffix at the end (e.g. gluon \rightarrow gluino). In the table, the Weyl notation has been chosen, this is why conjugates like \tilde{u}_R^* appear. As mentioned, the quantum numbers corresponding to the SM gauge symmetry are preserved under supersymmetric transformations, and therefore the super-partners of SM right-handed fermions also behave as singlets under the $SU(2)_L$ group. The Higgs sector is extended in MSSM models, with two complex scalar Higgs fields to generate the masses of u - and d -type fermions, respectively. This is done because is it not possible to conjugate the Higgs field in order to generate down-fermion masses as in the SM. This leads to a total of four Higgs bosons, two of them charged, and four higgsinos. The SM-like Higgs is recovered by combining H_u^0 and H_d^0 . As in the SM, the neutral-charged fields \tilde{W}^0 and \tilde{B}^0 can be mixed to form the so-called *photino* and *zino*. In a similar way, higgsinos, wino and bino mix with each other resulting in six mass eigenstates: two charginos, $\tilde{\chi}_{1,2}^\pm$, and four neutralinos, $\tilde{\chi}_{1,2,3,4}^0$.

In principle, the new symmetry introduced when building the MSSM implies that each new particle would have the same mass as their respective counterpart. But it is clear from the experimental evidence that this is not the case. In order to resolve this, an spontaneous symmetry breaking mechanism is introduced in a similar way as EW, only that, instead of adding a new scalar field, a soft breaking

**Weyl fermion refers to the solution of a two-dimensional version of the Dirac equation for massless fields, usually known as the Weyl equation.

Names		$J = 0$	$J = 1/2$	$J = 1$
Chiral supermultiplets				
Quarks & squarks	Q_i	$(\tilde{u}_L, \tilde{d}_L)_i$	$(u_L, d_L)_i$	–
	$(i = 1, 2, 3)$	\bar{u}_i	$(\tilde{u}_R^*)_i$	$(\tilde{u}_R^\dagger)_i$
		\bar{d}_i	$(\tilde{d}_R^*)_i$	$(\tilde{d}_R^\dagger)_i$
Leptons & sleptons	L_i	$(\tilde{\nu}_L, \tilde{e}_L)_i$	$(\nu_L, e_L)_i$	–
	$(i = 1, 2, 3)$	\bar{e}_i	$(\tilde{e}_R^*)_i$	$(\tilde{e}_R^\dagger)_i$
Higgs & higgsinos	H_u	(H_u^+, H_u^0)	$(\tilde{H}_u^+, \tilde{H}_u^0)$	–
	H_d	(H_d^+, H_d^0)	$(\tilde{H}_d^+, \tilde{H}_d^0)$	–
Gauge supermultiplets				
Gluon & gluino		–	\tilde{g}	g
W -bosons & winos		–	$\tilde{W}^\pm, \tilde{W}^0$	W^\pm, W^0
B -boson & bino		–	\tilde{B}^0	B^0

Table 1.2: Elementary particles composing the chiral and gauge supermultiplets in a generic MSSM theory.

term is simply added to the Lagrangian,

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}}, \quad (1.47)$$

where the last term solves the mass degeneracy between SM and SUSY particles by making the later heavier. The breaking of the symmetry is then regarded as a perturbation of the SUSY Lagrangian without introducing unwanted quadratic divergences. As mentioned previously, the hierarchy problem originates from the fact that corrections to the mass of the Higgs could extend up to very high values as there is a huge gap until reaching the Planck scale. The addition of the new set of particles makes the contributions from the fermionic sector to cancel with the contributions from their bosonic super-partners, and same applies for the SM bosons. The remaining contributions arising from the soft term are assumed to be not larger than the TeV scale. Therefore, at higher energies new particles be produced and thus MSSM models can be probed in collider experiments.

In the construction of the Lagrangian, all possible interaction terms that satisfy the SM symme-

tries are in principle included, but some of these terms can lead to violations of the baryon and lepton numbers conservation, which is forbidden in the SM and supported empirically by evidences such as the proton decay. In order to avoid this, a new symmetry called R -parity is introduced, whose conserved quantity is defined as

$$R = (-1)^{3(B-L)+2S}, \quad (1.48)$$

where B and L are the baryon and lepton numbers, respectively, and S denotes the spin of the particle. Eq. (1.48) implies that $R = 1$ for SM particles and $R = -1$ for sparticles. The conservation of R preserves the conservation of the baryon and lepton quantum numbers, as required, and has major consequences in SUSY phenomenology:

- sparticles must be produced in pairs,
- no mixing is allowed between SM and SUSY particles,
- the lightest supersymmetric particle (LSP) can not decay,
- all the other sparticles decay into final states with an odd number of LSPs.

The conservation of R -parity is not always required depending on the SUSY model. In those models in which it is conserved, as is the case to what respect to the analysis of this thesis, the LSP is usually identified with the lightest neutralino ($\tilde{\chi}_1^0$) and is considered a candidate for dark matter.

SUSY models can be tested in particle collider experiments. In particular this analysis searches for events where a squark-pair is produced from a pp collision. The production of sparticles is typically dominated by QCD, with processes like $pp \rightarrow \tilde{g}\tilde{g}$, $pp \rightarrow \tilde{g}\tilde{q}$ or $pp \rightarrow \tilde{q}\tilde{q}$, depending on the masses of the squark and the gluino. The squark-pair would then decay into final states with SM quarks and light neutralinos. The later would escape undetected as it is assumed to be stable, therefore leaving a signature in form of missing transverse momentum that the monojet analysis is sensitive to. In particular, the analysis focuses in those scenarios in which the difference between the masses of the squark and the neutralino, $\Delta m = m_{\tilde{q}} - m_{\tilde{\chi}_1^0}$, is small (the so-called *compressed scenarios*), of the order of a few GeV, since in such cases the SM decay products will be low-energetic and might not be reconstructed by the detector, so these kind of events can only be identified if an ISR jet recoils against the squark system. In this way, the monojet analysis provides a unique access to this particular region of the parameter space.

Four different scenarios are considered for the squark-pair production, all of them conserving R -parity as mentioned. Figure 1.8 shows typical diagrams for each of the models considered. The production of pairs of light squarks which decay directly into their respective SM quarks is treated individually. This is the case known as the **SS model**, the process $\tilde{q} \rightarrow q\tilde{\chi}_1^0$ with $q = u, d, s, c$. The production of heavier-flavour squarks is studied separately. First, a sbottom-pair decaying into b -quarks

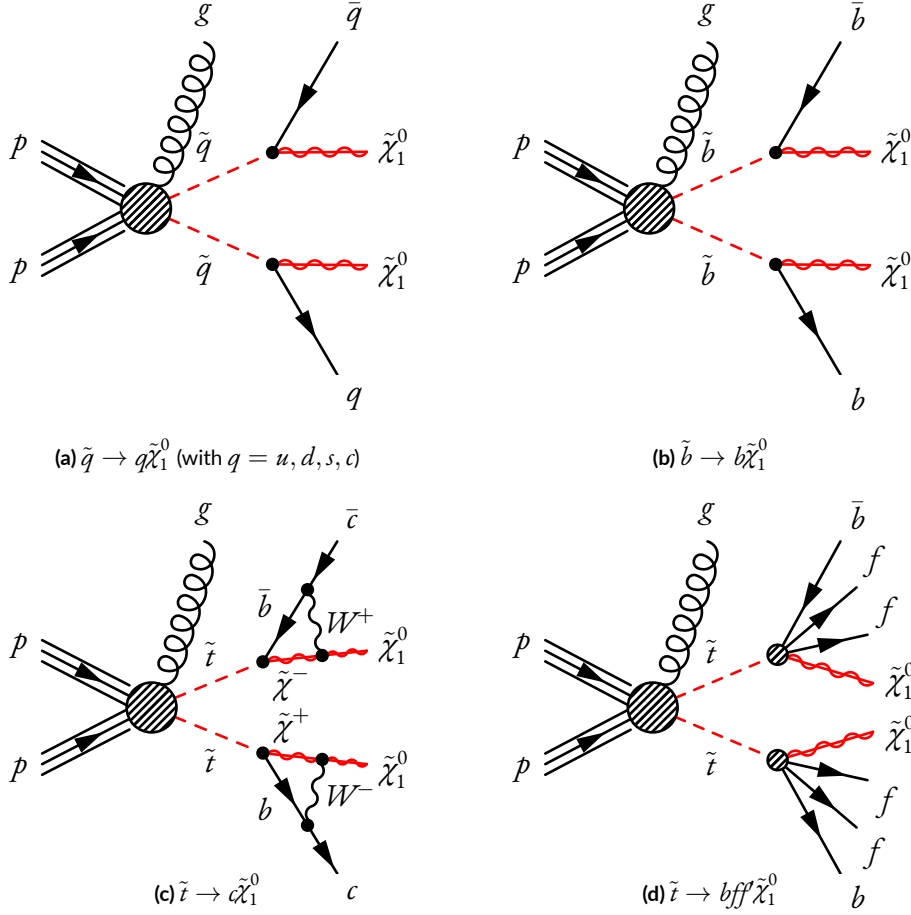


Figure 1.8: Squark-pair production diagrams in association with a jet plus their subsequent decays according to the four compressed SUSY models considered in the analysis, namely (a) SS, (b) BB, (c) TT and (d) T4body.

plus neutralinos: $\tilde{b} \rightarrow b\tilde{\chi}_1^0$ (the **BB model**), a similar scenario to the SS model. Then, two different cases are considered for the stop production: a stop decaying into a c -quark and a neutralino, $\tilde{t} \rightarrow c\tilde{\chi}_1^0$ (the **TT model**), which happens only if $\Delta m < m_b + m_W$ so decaying into a b -quark is not allowed; and the so-called stop four-body decay, $\tilde{t} \rightarrow bff\tilde{\chi}_1^0$ (named as the **T4body model**). In all cases, the branching fraction of the process is assumed to be of 100% and the parameter space is explored up to the most compressed scenarios, with the exception of the $\Delta m < 5$ GeV region for the third generation cases, as the phase space would not allow for b -quark production. Since in the monojet analysis events are tagged by the presence of a jet, which is often initiated by a gluon, the flavour of the produced squarks is not really important.

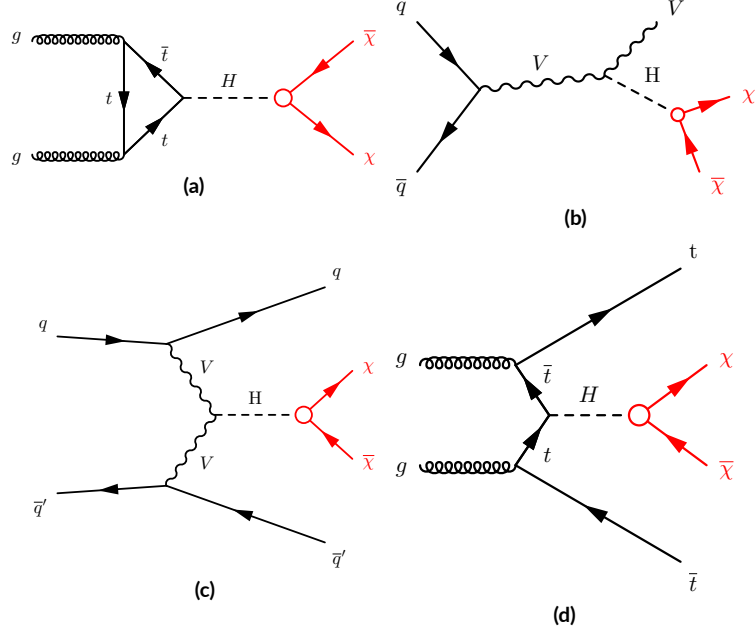


Figure 1.9: Diagrams of invisible-decaying Higgs for different production channels, namely (a) ggF, (b) VH, (c) VBF and (d) ttH

1.3.4 INVISIBLE-DECAYING HIGGS

The Higgs boson is unique within the SM framework, as it is the only scalar boson, and provides a critical element in the theory which is the generation of fermions and gauge bosons masses via spontaneous EW symmetry breaking (described in Section 1.1.3). Studying its properties is fundamental to get a better understanding of the dynamics related to the SM particles.

But the Higgs can be also studied as a portal to hidden sectors where BSM physics might lie [68]. In particular, some WIMP models predict a relation between the SM and the dark sector through Higgs, for example via Yukawa couplings to fermionic DM candidates. This can be explored in particle collision experiments by searching for whenever a Higgs boson is produced and decays into a pair of WIMPs. These kind of events would lead to signatures of missing transverse momentum, since the DM-pair is stable and does not interact with the detector, so the process $H \rightarrow \chi\chi$ can be expressed as $H \rightarrow \text{invisible}$. There is actually a process of this type allowed in the SM, which is a Higgs decaying into two Z -bosons, which decay subsequently into neutrinos, $H \rightarrow ZZ^* \rightarrow \nu\nu\nu$, with a branching ratio (BR) of $\Gamma_{H \rightarrow \text{inv.}} \simeq 1.12 \cdot 10^{-3}$ [69]. The addition of $H \rightarrow \chi\chi$ would increase the total BR above the SM predicted value, therefore measuring $\Gamma_{H \rightarrow \text{inv.}}$ is a powerful way to search for evidences of BSM physics.

In pp collisions, a Higgs boson can be produced via different processes, as shown in the diagrams of Fig. 1.9. The dominant production channel is the so-called gluon–gluon fusion (ggF) mechanism, with $gg \rightarrow H$ and $gg \rightarrow ZH$, followed by the vector boson fusion (VBF) process, with $ZZ/W^+W^- \rightarrow H$. Another processes are the Higgs production in association with a W/Z boson (called VH process), or with a $t\bar{t}$ (ttH) in the final state.

Previous searches for invisible Higgs decays using the monojet signature have been performed by the ATLAS and CMS experiments, setting upper limits on the BR of 0.26 [70] and 0.19 [71], respectively.

1.3.5 OTHER MODELS

In addition to the mentioned models above, another two signatures of BSM have been explored, although they will not be discussed in detail in this thesis.

- A Dark Energy inspired model, based on the inclusion of a new scalar field and implemented as an EFT [72]. In this context, the new scalar particle is stable and produced in pairs, therefore escaping through the detector and leaving a monojet-like signature, when produced in association with a gluon.
- An model including Axion-Like Particles (ALPs), which arise from the breaking of a new $U(1)$ symmetry that is introduced as an attempt to solve the lack of CP -violation in QCD. These kind of particles are also seen as solid DM candidates. An EFT implementation of the ALP model, with an effective scale f_a , is then considered in which ALPs are produced in association with a gluon, a process governed by the ALP-gluon coupling, $c_{\tilde{G}}$. Again, the produced ALP is considered stable and escapes without interaction with the detector material, therefore leaving a signature that can be studied via the monojet channel.

2

MC Simulation

In order to test the theoretical predictions against the data, simulated samples are generated with MC techniques for both SM processes that can lead to final states similar to the selected events, and for each of the signal models considered. In this chapter, the preparation of such set of samples is detailed.

2.1 SIMULATION SOFTWARE

Different algorithms are used for the different stages in the generation of events, as well as the simulation of the interactions with the detector.

2.1.1 MC GENERATORS

A number of software options are available for the different stages of the event generation.

- SHERPA [73] is a general-purpose MC event generator widely used by LHC experiments for Run 1 and 2 analyses. It is specialized on simulating $2 \rightarrow n$ processes in inclusive samples. Provides ME calculations up to NLO or NNLO in pQCD precision, depending on the process,

and up to NLO accuracy in EW, and provides also generic interface to external generators. The PS is modelled by using Q^2 to order the cascade evolution, with typical inter-parton separation scales down to 1 GeV^2 . The UE is also modelled within SHERPA. The cluster fragmentation model is used for the hadronization, and the CKKW algorithm for the ME-PS matching.

- PYTHIA [74–76] is general purpose event generator, mostly focused on $2 \rightarrow 1$ and $2 \rightarrow 2$ processes, with some $2 \rightarrow 3$ processes available in the latest releases. The ME is calculated at LO. The PS modelling is based on the dipole-style p_{\perp} -ordered evolution. Underlying events and minimum bias events can be simulated with PYTHIA as well. The string fragmentation model is used for the hadronization modelling.
- HERWIG++ [77] is also a multi-purpose generator. It calculates the ME at NLO accuracy. HERWIG++ simulates the PS by the angular ordering of successive emissions and uses the cluster fragmentation model for the hadronization. The UE is simulated using an eikonal multiple parton-parton scattering model.
- POWHEG-BOX [78–80] is a generator that provides ME calculations at NLO in QCD. It is often interfaced with other frameworks for the PS and UE simulations.
- MADGRAPH [81–83] is an event generator specialized on $2 \rightarrow n$ processes, with ME calculations at up to NLO in QCD. MadGraph is usually interfaced with Pythia8 for the PS, hadronization and UE simulation.
- EVTGEN [84] is framework widely used in B -physics experiments because it provides precise predictions for B -meson decays, although it also includes other resonances. It uses decay amplitudes instead of probabilities for the simulation of decays.

2.1.2 DETECTOR SIMULATION

Once the event is generated at truth level, the interaction with the different components of the detector has to be simulated as well, including effects such as Bremsstrahlung radiation, multiple scattering, pair creation, etc. In this way, the event can be reconstructed and processed with the same code chain as done for the data. Such simulation of the detector is carried out with the software GEANT4 [85], widely used in nuclear and particle physics for emulating radiation-material interaction processes.

This is the step that consumes most of the computing time and resources of the whole sample generation chain. For this reason, a balance must be found between an accurate modelling and a realistic CPU time. For this reason, and depending on the necessities for the specific samples, two different configurations are provided for the detector simulation in ATLAS: *full simulation* [86] and a *fast*

simulation option, provided by the ATLEFAST package [87]. Both options include a detailed modelling of the geometry of the ATLAS detector components and the interactions of the particles with the material, the main difference is the simplified parametrization of the calorimeter energy deposits implemented in the fast simulation, as this is the most CPU consuming part.

2.2 SIMULATED SAMPLES FOR THE MONOJET ANALYSIS

The MC samples production is split into three campaigns in order to account for the different pile-up conditions in the description of the corresponding data collected in 2015 and 2016, 2017 and 2018.

2.2.1 BACKGROUND SAMPLES

There are several SM processes that contribute in a non-negligible way to the monojet signature and for which dedicated samples are produced, listed below. The full detector simulation is used for all the background samples.

- *V*+jets. Samples for *W*+jets and *Z*+jets processes are simulated using SHERPA 2.2.1 interfaced with COMIX [88] and OPENLOOPS [89] for the ME calculation for up to 2 partons at NLO and 4 partons at LO, and merged with the SHERPA PS, following the ME+PS@NLO prescriptions detailed in Ref. [90]. The NNLO PDF set NNPDF3.0 [22] is used. The total cross-sections are computed at NNLO. In the case of dilepton processes, a cut on $m(\ell^+\ell^-) > 40$ GeV is applied, thus an additional set of samples with $10 \text{ GeV} < m(\ell^+\ell^-) < 40$ GeV is also generated. These samples have a special $p_T(V)$ -based reweighting as it is described in Section 2.3.

In addition, a set of samples where the *Z/W* boson comes from VBF are generated using HERWIG 7.1.3. These samples are produced at NLO in pQCD using VBFNLO v3.0.0 [91]. The NNPDF3.0 PDF set was used together with the default set of tuned parameters for the PS, hadronization and UE. For these samples the muon channels are regenerated in HERWIG 7.2 due to a displaced vertex issue. The samples are produced requiring a *t*-channel colour singlet exchange to avoid overlap with the semi-leptonic diboson topologies, as it has been done in other Run2 analyses (e.g. ATLAS VBF+MET [92]).

- In the generation of samples of $t\bar{t}$ events, the ME are calculated using the POWHEG-Box v2 generator with the NNPDF3.0NLO PDF set. The generator uses the 5-flavour scheme for the

NLO ME calculations. The top-quarks are made to decay using MADSPIN [93] to preserve all spin correlations, while the PS, fragmentation, and the UE were simulated using PYTHIA (version 8.2.30) with the NNPDF2.3LO PDF set and the ATLAS A14 tune [94].

- **single- t .** Samples with single top-quark events produced in the Wt and s -channel are generated with POWHEG-BOX (v2) and the ME are calculated using the CT10 PDF sets. EW t -channel single- t events are generated using POWHEG-BOX (v1). This generator uses the 4-flavour scheme for the NLO ME calculations together with the fixed 4-flavour PDF set CT10f4. Like for $t\bar{t}$, spin correlations are preserved (for t -channel, top-quarks are decayed using MADSPIN). The PS, fragmentation, and the UE are simulated using PYTHIA 6.428 with the CTEQ6L1 PDF sets and the Perugia 2012 tune (P2012) [95]. The top mass is set to 172.5 GeV. The EVTGEN v1.2.0 program is used for properties of the bottom and charm hadron decays.
- **Diboson** samples (WW , WZ and ZZ production) with the VV system decaying into 4ℓ , $3\ell + \nu$ or $2\ell + 2\nu$ are simulated with SHERPA (versions 2.2.1 and 2.2.2). Likewise the V +jets case, the ME are calculated for up to 1 partons at NLO and up to 3 partons at LO using COMIX and OPENLOOPS, and merged with the SHERPA PS, according to the ME+PS@NLO prescription. The NNPDF3.0NNLO PDF set is used together with the default PS tuning. The event generator cross-sections are calculated at NLO.

2.2.2 SIGNAL SAMPLES

There are several signal models considered in the analysis of this thesis, for which dedicated samples are produced. With the exception of the invisible Higgs decay model (for which full simulation is used), the signal samples are produced with detector fast simulation. In addition to the samples listed below, events are also generated by following similar procedures for an ALPs EFT model and a Dark Energy inspired model as well.

- **ADD LED** samples are generated using PYTHIA 8.205 with the NNPDF2.3 PDF set [96] and using the ATLAS14 tune for the PS. The cross-section is computed at NLO in α_S . The renormalization scale is set to $\sqrt{(p_{T,G}^2 + m_G^2)(p_{T,p}^2 + m_p^2)}$, where $p_{T,G}$ and m_G are the transverse momentum and the mass of the KK graviton, and $p_{T,p}$ and m_p are the p_T and the mass of the parton. The factorization scale is set to the smallest of the transverse masses of the graviton

and the parton, $\sqrt{p_T^2 + m^2}$. The samples are generated for $n = 2 - 6$ extra dimensions and setting the fundamental scale M_D to values close to exclusion limits set in Run 1 [97], for fair comparison.

- **H** \rightarrow **inv.** Samples with a Higgs boson produced with a mass of 125 GeV are generated with POWHEG-Box (version 2), with NLO accuracy. Particularities in the simulation of the different production channels are listed below.
 - ggF: the generation of events is interfaced with PYTHIA 8.212 for PS, hadronization and UE, using the AZNLO tune [98] and the NNPDF3.0+CTEQ6L1 PDF set [22]. The total cross-section is calculated at NNNLO order in QCD and NLO EW corrections are applied for the $gg \rightarrow H$ sample, while the $gg \rightarrow H$ sample is normalized to a cross-section calculated at NLO in QCD.
 - VH: like the gluon–gluon fusion samples, the generation is interfaced with PYTHIA 8.212 for PS, hadronization and UE using the AZNLO tune. The CT10 PDF set is used instead, and the MiNLO (*Multiscale Improved NLO*) procedure [99] is applied. The cross-sections for these samples are calculated with NNLO accuracy in QCD with NLO EW corrections.
 - VBF: these samples are also generated interfacing POWHEG-Box with PYTHIA 8.212 for the PS, hadronization and UE, using the AZNLO tune, but the NNPDF3.0 PDF set is used in this case. The samples are normalized to cross-sections calculated at NNLO in QCD with NLO EW corrections.
 - ttH: samples are generated at NLO accuracy in QCD using the POWHEG-Box generator alone, with the PDF4LHC15 [100] set of PDFs. The cross-sections are determined at NLO in QCD, with NLO EW corrections.

In all processes the Higgs boson is forced to decay into two Z bosons, which decay into neutrinos ($H \rightarrow Z^* Z \rightarrow 4\nu$), because such final state is consistent with those from models with invisibly decaying Higgs bosons.

- **DM_s**-channel samples of the type $\chi\bar{\chi}$ +jet are generated at NLO precision with the POWHEG-Box framework, considering two simplified models:
 - The DMA model with spin-1 axial-vector mediator exchange at NLO.
 - The DMP model used for spin-0 pseudo-scalar mediator exchange with the full quark-loop calculation at LO.

Renormalization and factorization scales are set to $H_T/2$, where $H_T = \sqrt{m_{\chi\chi}^2 + p_{T,j1}^2} + p_{T,j1}$ is defined by the invariant mass of the WIMP pair ($m_{\chi\chi}$). A Breit-Wigner expression is chosen for the mediator propagator. The generation of samples is interfaced with PYTHIA 8.205 with the ATLAS14 tune for the PS, hadronization and UE simulation, and the NNPDF3.0 PDF set is used. As mentioned, the couplings of the mediator to DM particles and SM quarks are set to $g_\chi=1$ and $g_q=1/4$ for the DMA model, while for the DMP model both are set to $g_\chi=g_q=1$. A set of samples is produced for DM masses ranging between 1 – 1000 GeV and mediator masses varying within 10 – 10000 GeV.

- **SUSY.** As explained in the previous chapter, four squark-pair production scenarios are considered in the analysis of this thesis, namely SS, BB, TT and T₄body. All signal samples are generated with MadGraph5_aMC@NLO v2.2.3, interfaced with PYTHIA 8.186 with the ATLAS 14 tune for modeling of the squark decay, PS, hadronization, and the UE. The PDF set NNPDF2.3LO is used. The ME calculation is performed at LO, and includes the emission of up to two additional partons. The ME–PS matching is done using the CKKW-L prescription, with a matching scale set to 1/4 of the pair-produced superpartner mass. The renormalization and factorization scales are set to $\mu = \sum_i \sqrt{m_i^2 + p_{T,i}^2}$, where the sum runs over all final-state particles from the hard-scatter process. The samples are normalized to cross sections calculated to approximate NNLO in α_S , adding the resummation of soft gluon emission at next-to-next-to-leading-logarithm accuracy (approximate NNLO+NNLL) [101–104]. In all models, the compressed scenario is considered. A grid of samples is then produced with values of $m_{\tilde{q}}$ varying from 250 GeV to 1.2 TeV and differences in mass, $\Delta m = m_{\tilde{q}} - m_{\tilde{\chi}_1^0}$, between 5–50 GeV.

2.2.3 MC CORRECTION WEIGHTS

Every event generated is corrected by a set of weights which are specific of the algorithm used for the simulation and the performance in the identification and reconstruction of the physical objects. The final weight is defined by the product of all the individual contributions, listed below.

- MC weights: event-by-event generator weights, specific of the simulation algorithm.
- Pile-up weights: correction factors accounting for differences between simulations and data in pile-up. These weights are extracted by overlaying the MC sample with simulated minimum-bias events, which are generated with PYTHIA 8.186 using the A3 set of parameters [105] and the NNPDF2.3LO PDF set, and are distributed according to the data frequency.

- Lepton efficiency scale factors (SFs): quantities derived from the ratio of efficiencies measured in data and simulated events, $\varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$, accounting for mismodeling of the detector efficiency in the reconstruction, identification, isolation or triggering of leptons and photons. The SFs act as event-level weights to be used when the presence of such object is requested.
- Anti scale factors (anti-SFs): similarly to the SFs, the anti-SFs correct for detector inefficiencies when vetoing objects. The calculation method is detailed in the Appendix B. The anti-SFs are applied every time the absence of an specific object is requested in the event.

Finally, the MC samples are normalized to be compared to the data, taking into account the process cross-section and the luminosity of the corresponding data-taking period and also accounting for generator-level cuts applied at the time of the generation of the samples.

2.3 V +JETS REWEIGHTING

In order to achieve a better description of the W/Z + jets backgrounds, which are largely dominant in the monojet analysis, higher-order perturbative corrections in QCD and EW processes are applied to the V +jets MC samples*. The reweighting method is based on Ref. [106], which includes QCD corrections at NNLO [107–110] and EW corrections at NLO [111–114], supplemented by Sudakov logarithms at two loops [115–118]. The correction described below is one of the major improvements with respect to the previous iteration of the analysis [49], where the calculation provided was performed only at NLO in QCD (version–1 of [106]). Uncertainties related to these corrections and their correlations across processes are described in Section 7.3.

Eq. (2.1) describes the procedure for the one-dimensional reweighting of the V +jets MC samples in a generic variable x (the momentum of the boson in this case, $x = p_T(V)$), expressed in terms of the cross-section as computed from MC simulation ($\sigma_{\text{MC}}^{(V)}$) and from the theory calculation ($\sigma_{\text{TH}}^{(V)}$),

$$\frac{d}{dx} \frac{d}{d\mathbf{y}} \sigma^{(V)}(\varepsilon_{\text{MC}}, \varepsilon_{\text{TH}}) = \frac{d}{dx} \frac{d}{d\mathbf{y}} \sigma_{\text{MC}}^{(V)}(\varepsilon_{\text{MC}}) \left[\frac{\frac{d}{dx} \sigma_{\text{TH}}^{(V)}(\varepsilon_{\text{TH}})}{\frac{d}{dx} \sigma_{\text{MC}}^{(V)}(\varepsilon_{\text{MC}})} \right], \quad (2.1)$$

where \mathbf{y} refers to the remaining kinematic variables included in the simulation, ε_{MC} represents the set of experimental and modelling uncertainties from independent sources on the variable x , and ε_{TH} denotes the set of theoretical uncertainties associated with the corrections. The term in brackets depends only on the variable x and represents the reweighting factor applied to the differential cross-section,

*This procedure is prescribed only for V +jets events not initiated by VBF.

therefore this procedure is inclusive in all observables except of x . The same event selection cuts must be used for the numerator and the denominator of such term in brackets. One of the advantages of this method is that the three terms on the right hand side of Eq. (2.1) do not need to be computed with the same numerical setup, as long as the definition and the binning of x is the same. On the other hand, the two terms of the ratio given by

$$R_{\text{MC}}(x, \mathbf{y}) = \frac{\frac{d}{dx} \frac{d}{dy} \sigma_{\text{MC}}^{(V)}}{\frac{d}{dx} \sigma_{\text{MC}}^{(V)}} \quad (2.2)$$

has to be fully consistent in terms of PDFs, scale choices and input parameters. While the numerator in Eq. (2.2) is the nominal yield given by the MC sample, the denominator needs to be adapted in order to meet the aforementioned condition. In addition, in order to make an optimal use of the reweighting procedure two conditions should be passed: the theoretical predictions should be at least equal or more precise than the MC ones; and the correlations across x and the other variables \mathbf{y} should be at least as precise (or more) in the MC predictions than in the theory calculations. For these reasons, although $p_{\text{T}}(V)$ is a natural choice given its minimal sensitivity to multiple jet emissions, the reweighting procedure is only valid for inclusive samples and the region $p_{\text{T}}(V) \ll m_V$ must be excluded from it, since in that regime the analysis becomes sensitive to soft QCD interactions. Moreover, any non-perturbative aspect of the MC simulation (hadronization, UE, hadron decays) should be excluded from the definition of x , and therefore uncertainties related to those features will remain the same after the reweighting.

The theoretical corrections provided can be split into several components as

$$\frac{d}{dx} \sigma_{\text{TH}}^{(V)} = \frac{d}{dx} \sigma_{\text{QCD}}^{(V)} + \frac{d}{dx} \Delta \sigma_{\text{EW}}^{(V)} + \frac{d}{dx} \sigma_{\text{mix}}^{(V)} + \frac{d}{dx} \sigma_{\gamma\text{-ind}}^{(V)}, \quad (2.3)$$

where the labels refer to the higher-order QCD and EW corrections, mixed QCD and EW calculations and photon-induced contributions. Each term of Eq. (2.3) will be discussed in the following.

- QCD predictions are provided at NNLO (i.e. $\mathcal{O}(\alpha_s^3)$). The higher-order corrections are applied as normalization factors $K_{\text{NNLO}}^{(V)}(x, \boldsymbol{\mu})$ to the LO prediction as

$$\frac{d}{dx} \sigma_{\text{NNLO QCD}}^{(V)}(\boldsymbol{\mu}) = K_{\text{NNLO}}^{(V)}(x, \boldsymbol{\mu}) \frac{d}{dx} \sigma_{\text{LO QCD}}^{(V)}(\boldsymbol{\mu}_0), \quad (2.4)$$

where the dependence on the renormalization and factorization scales, $\boldsymbol{\mu} = (\mu_R, \mu_F)$ is absorbed by $K_{\text{NNLO}}^{(V)}(x, \boldsymbol{\mu})$. The LO predicted differential cross-section in Eq. (2.4) is taken at the central scale, which is adopted by convention as $\mu_{R,0} = \mu_{F,0} = \hat{H}_{\text{T}}'/2$ with \hat{H}_{T}' being the

scalar sum of the transverse energy of all the parton-level objects in the final state.

- For the EW corrections, the following notation is used:

$$\begin{aligned}\frac{d}{dx}\sigma_{\text{NLO EW}}^{(V)} &= \frac{d}{dx}\sigma_{\text{LO QCD}}^{(V)} + \frac{d}{dx}\Delta\sigma_{\text{NLO EW}}^{(V)}, \\ \frac{d}{dx}\sigma_{\text{nNLO EW}}^{(V)} &= \frac{d}{dx}\sigma_{\text{NLO EW}}^{(V)} + \frac{d}{dx}\Delta\sigma_{\text{NNLO Sud}}^{(V)}\end{aligned}\quad (2.5)$$

where $\Delta\sigma_{\text{NLO EW}}^{(V)}$ denotes exact $\mathcal{O}(\alpha^2\alpha_S)$ contributions and ‘NNLO Sud’ stands for $\mathcal{O}(\alpha^3\alpha_S)$ EW Sudakov logarithms in NLL approximation. Their combination is labeled nNLO EW as it accounts for the dominant EW effects at NNLO. Similarly as done for QCD, nNLO EW corrections are expressed as correction factors, $\kappa_{\text{nNLO EW}}^{(V)}(x, \vec{\mu})$, applied to the LO QCD predictions as

$$\frac{d}{dx}\sigma_{\text{nNLO EW}}^{(V)}(\boldsymbol{\mu}) = \left[1 + \kappa_{\text{nNLO EW}}^{(V)}(x, \boldsymbol{\mu})\right] \frac{d}{dx}\sigma_{\text{LO QCD}}^{(V)}(\boldsymbol{\mu}), \quad (2.6)$$

where $\kappa_{\text{nNLO EW}}^{(V)}(x)$ is split into a pure EW component, evaluated at NLO, and NNLO Sudakov contributions:

$$\kappa_{\text{nNLO EW}}^{(V)}(x) = \kappa_{\text{NLO EW}}^{(V)}(x) + \kappa_{\text{NNLO Sud}}^{(V)}(x). \quad (2.7)$$

The former includes virtual EW corrections to $q\bar{q} \rightarrow Vg$, photon Bremsstrahlung corrections, virtual QCD corrections to $q\bar{q} \rightarrow V\gamma$ (needed to cancel soft-gluon singularities from photon Bremsstrahlung) and $q\bar{q} \rightarrow Vq'\bar{q}'$ Bremsstrahlung corrections as well. The later accounts for the inclusion of higher-order Sudakov logarithms that need to be implemented at high momentum transfer regimes, where all energy scales are way above m_V [25], and are evaluated at next-to-leading-logarithm (NLL) level of accuracy. The LO QCD cross-section is taken at the same scale as the EW correction, since it has no dependency on the scale.

- Mixed QCD–EW corrections are yet not known. However, in order to include partially such effects, higher-order QCD and EW corrections are factorized as

$$K_{\text{TH}}^{(V)} = K_{\text{NNLO}}^{(V)}(x, \boldsymbol{\mu}) \left[1 + \kappa_{\text{nNLO EW}}^{(V)}(x, \boldsymbol{\mu})\right]. \quad (2.8)$$

This approach allows to include mixed correction terms of the order $\mathcal{O}(\alpha\alpha_S)$.

- Photon-induced contributions arise from $\gamma q \rightarrow Vq'$ processes, where the photon comes from the photon PDF inside the proton. The impact on W +jets processes is evaluated by compar-

ing different sets of PDFs (namely, LUXqed [119] and CT14qed_inc [120]). The impact on Z +jets is, on the other hand, negligible.

The final correction has then the form

$$\frac{d}{dx}\sigma_{\text{TH}}^{(V)} = K_{\text{TH}}^{(V)} \frac{d}{dx}\sigma_{\text{QCD}}^{(V)} + \frac{d}{dx}\sigma_{\gamma\text{-ind}}^{(V)}, \quad (2.9)$$

Figure 2.1 shows the size of the QCD and EW corrections when applied to different V +jets processes as a function of the boson p_{T} , as well as their individual uncertainties. NNLO QCD corrections seem to have a mostly 7% flat effect on $Z \rightarrow \ell\ell$ +jets processes while raise from 5% to 15% with $p_{\text{T}}(V)$ for $W \rightarrow \ell\nu$ +jets, scaling up the number of events in both cases. nNLO EW corrections, on the other hand, appear to have a more relevant impact, with a similar reduction of events growing with $p_{\text{T}}(V)$ from less than 2% up to 30 – 45%. In the analysis, the V +jets corrections combined translate into an increasing reduction of the number of events as a function of $E_{\text{T}}^{\text{miss}}$, ranging between $\sim 2\%$ and $\sim 20\%$ depending on the process and the specific selection region of the analysis. For more details about the size of the V +jets corrections on the samples of the analysis, see Appendix G.

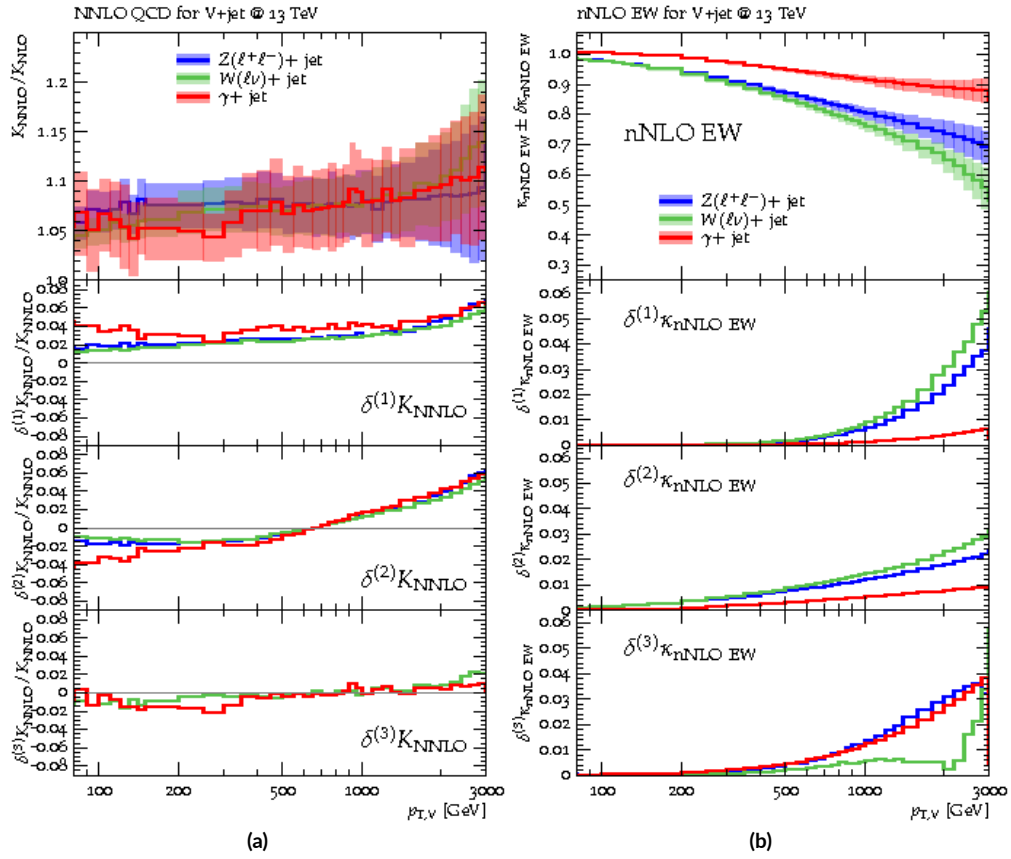


Figure 2.1: (a) NNLO QCD K -factors and (b) nNLO EW κ -factors for different V +jets processes at 13 TeV with their individual uncertainties. Figures taken from Ref. 2.1.

3

The ATLAS Experiment at the LHC

In this chapter the main aspects of the ATLAS (*A Toroidal LHC Apparatus*) Experiment are introduced, starting from a brief description of the Large Hadron Collider at CERN, followed by an overview of the multiple components of the ATLAS detector and the mechanisms that are used for the data acquisition.

3.1 THE LARGE HADRON COLLIDER

The Large Hadron Collider (LHC) [121] is the most ambitious physics experiment up to date. It consists of a ~ 27 km-long underground ring located about 100 m below the border between France and Switzerland, inside of which two beams of charged particles (mainly protons, but also Pb and Xe ions have been used) travel in opposite directions in separate pipes kept at ultra-high vacuum, and are made to collide in four interaction points placed along the ring, where the four main LHC experiments (namely, ATLAS, CMS, LHCb and ALICE) operate, measuring and analyzing the outcome of the collisions. ATLAS [122] and CMS [123] are multipurpose experiments, LHCb [124] studies heavy-hadron physics and ALICE [125] focuses on heavy ion collisions.

CERN's Accelerator Complex

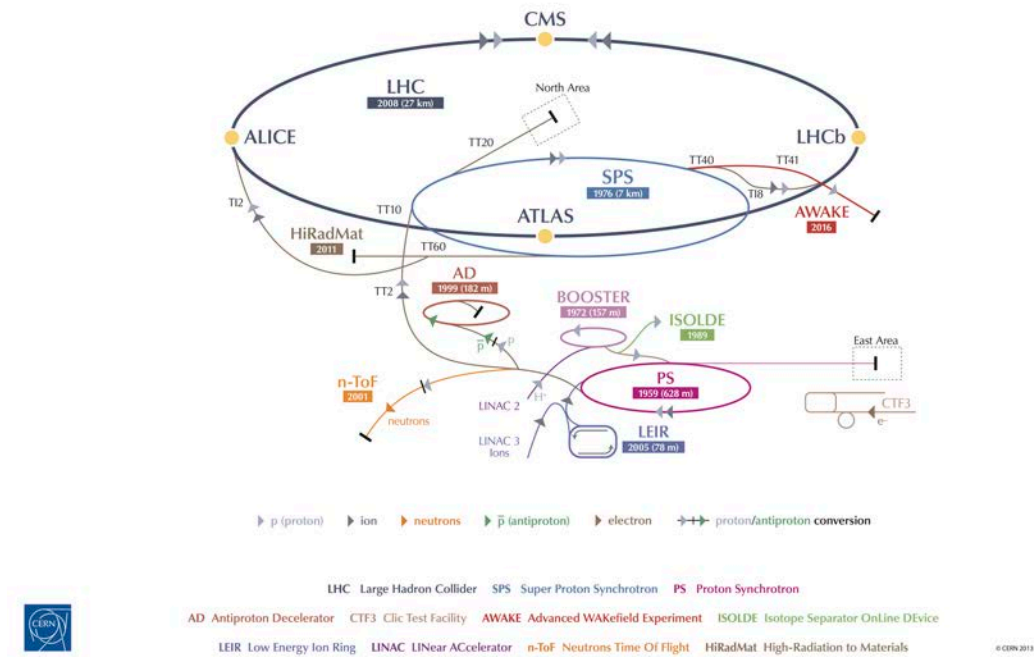


Figure 3.1: Schematic view of the CERN acceleration complex. Source: CERN.

Originally, the LHC was designed to collide protons at $\sqrt{s} = 14$ TeV. During the first period of operations, known as the Run 1 (years 2010–2012), the proton–proton collisions were produced at $\sqrt{s} = 7$ and 8 TeV. After the Run 1 the collider went off for three years, a period known as the Long Shutdown 1 where many upgrades were implemented in both collider and detectors. Between 2015 and 2018 (Run 2), the LHC was working at $\sqrt{s} = 13.7$ TeV, after of which the Long Shutdown 2 started.

The collider chain, as shown in figure 3.1, is governed by LHC magnets operations and initiates such that protons are obtained from a single bottle of hydrogen gas, replaced twice per year. Once the hydrogen atoms are stripped off their electrons by using an electric field, the protons start being accelerated up to 50 MeV with a linear accelerator (LINAC 2), followed by a circular accelerator (Booster) where an energy of 1.4 GeV is reached. Then, the protons pass consecutively through the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), raising their energies up to 25 and 450 GeV,

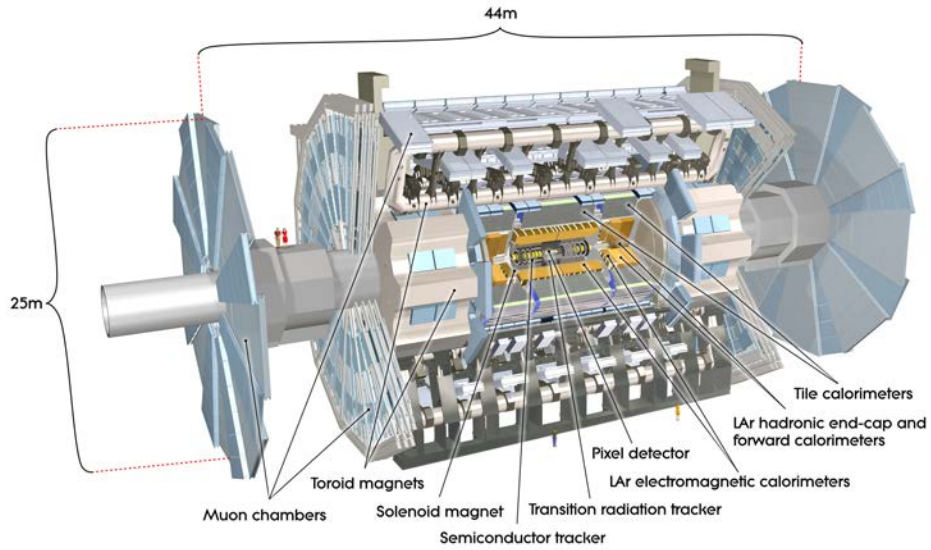


Figure 3.2: Schematic view of the ATLAS detector [122].

respectively. Finally, bunches of $\sim 1.1 \times 10^{11}$ protons are injected into the LHC (each beam contains a total of 2808 bunches, separated from each other by 25 ns), where the acceleration keeps up to final energies of 7–8 TeV.

3.2 THE ATLAS DETECTOR

The ATLAS detector is a cylindrical-shaped particle detector that covers a very wide range of physical phenomena studies, although it was originally optimized for the search of the Higgs boson. It is the largest LHC detector, measuring 46 m long \times 25 m diameter.

The detector is designed to identify and reconstruct the main products of the pp collisions, i.e. leptons (electrons, muons and taus), photons, jets and missing transverse energy. For this purpose the apparatus is composed of several concentric layers around the beam pipe, as shown in Figure 3.2, each one divided in multiple sub-detectors that are specialized in the detection of different types of particles. They can be grouped in three major parts, from inside out: the Inner Detector (ID), designed for the tracking of charged particles; the Calorimeters system, which contains and measures the energies of electrons, photons and jets; and the Muon Spectrometers (MS), optimized for measuring the

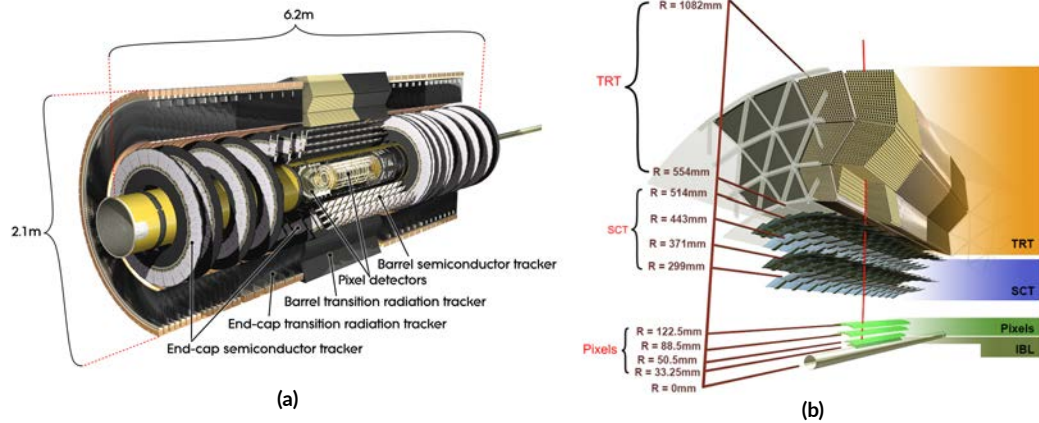


Figure 3.3: Schematic view of (a) the Inner Detector [122] and (b) its multiple layers in the barrel [126].

muon properties. In addition, a solenoid magnet is built around the ID to bend the trajectories of the charged particles, and a toroidal magnetic field surrounds the MS.

Before giving a more detailed description of the several sub-detectors, it is convenient to define the system of coordinates that will be used in the following. Setting the origin at the interaction point and the z -axis along the beam direction, a Cartesian right-handed system is chosen such that the x -axis points towards the center of the LHC ring and the y -axis points upwards. From there, two quantities are derived to define the trajectories of the particles: the azimuthal angle $\varphi = \arctan(y/x)$ and the pseudo-rapidity $\eta = -\ln[\tan(\theta/2)]$, being θ the polar angle measured from the z -axis. The transverse momentum (p_T) and the missing transverse energy (E_T^{miss}) are defined in the x - y plane.

3.2.1 INNER DETECTOR

The ID is the most internal component of ATLAS. It measures 6.2 m long \times 2.1 m diameter, and has a coverage $|\eta| < 2.5$. Immersed in a solenoidal magnetic field of 2 T, the ID provides information for reconstructing the tracks of the charged particles and the interaction vertices with very high precision. It is composed of three different (sub)sub-detectors, as shown in figure 3.3, detailed below.

- The **Pixel Detector** is composed of silicon sensors (pixels) and has about 80.4 million readout channels, being the instrument with the highest granularity in the ATLAS detector. Provides a resolution of 10 (115) μm in the R - φ plane (in z), which makes of it an essential instrument in

Item	Radial Extension [mm]	Length [mm]	Staves/sectors	Modules	Pixels ($\times 10^6$)
IBL	$\langle R \rangle = 25.7$	$ Z < 332$	14	224	6.02
B-layer	$\langle R \rangle = 50.5$	$ Z < 400.5$	22	286	13.2
Layer 1	$\langle R \rangle = 88.5$	$ Z < 400.5$	38	494	22.8
Layer 2	$\langle R \rangle = 122.5$	$ Z < 400.5$	52	676	31.2
Disk 1	$88.8 < R < 149.6 = 88.5$	$\langle Z \rangle = 495$	8×2	48×2	4.4
Disk 2	$88.8 < R < 149.6 = 88.5$	$\langle Z \rangle = 580$	8×2	48×2	4.4
Disk 3	$88.8 < R < 149.6 = 88.5$	$\langle Z \rangle = 650$	8×2	48×2	4.4
<i>Pixel Total</i>					80.4

Table 3.1: Main parameters of the Pixel Detector system [126].

the reconstruction of the primary vertex. The Pixel Detector is structured in three barrel layers plus the Insertable B-Layer (IBL), which was assembled in 2016 during the Long Shutdown 1 [126], and three disks in each endcap. The main technical aspects of each component are listed in Table 3.1. Most of the pixel sensors are of the type known as 'planar', with a size of $50 \times 400 \mu\text{m}^2$, except for a few ones, called '3D', that are much smaller ($50 \times 250 \mu\text{m}^2$), and are installed only at the two extremes of the IBL (more details about the 3D pixel sensors are given in Appendix A).

- The **Semi-Conductor Tracker (SCT)** is a silicon microstrip detector composed of four concentric layers of stereo strips (barrel, see figure 3.2) and nine disks at the endcaps. The strips are glued in pairs back-to-back with an angle of 40 mrad to provide as much points as possible for the reconstruction of the position of a crossing charged particle. Making use of 6.3 million readout channels, the SCT gives a resolution of $17 \mu\text{m}$ in the R - ϕ plane and $580 \mu\text{m}$ in z (R) in the barrel (endcap).
- The **Transition Radiation Tracker (TRT)**. The outermost part of the ID, made out of 4 mm diameter straw tubes supplied with Xe gas and a $30 \mu\text{m}$ -diameter gold-plated tungsten sense wire inserted in the center. The tubes are 144 cm long in the barrel, where they sit parallel to the beam pipe, and 37 cm long at the endcaps, where they are radially arranged. A plastic transition radiation material (polyethylene) is filling the space between the tubes. With about 351000 readout channels, the TRT has a resolution of $130 \mu\text{m}$, only in the R - ϕ plane.

In combination, the three ID sub-components give a total resolution in p_T of

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\% \quad (3.1)$$

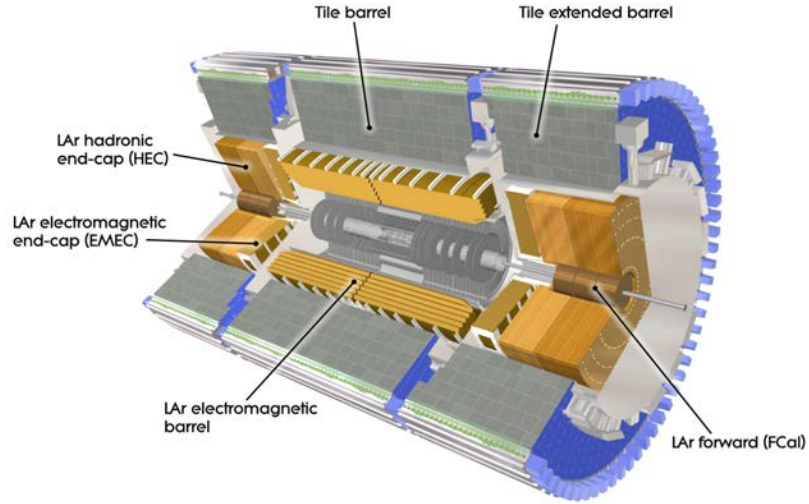


Figure 3.4: Schematic view of the ATLAS calorimeters system [122].

3.2.2 CALORIMETERS

The calorimeters are mounted around the ID, with a coverage of $|\eta| < 4.9$ and a radial extension of 4.25 m. They are divided into the electromagnetic calorimeter (EMCal), designed for the reconstruction of electrons and photons, and the hadronic calorimeter (HCal), devoted to the precise measurement of the energy of jets and taus. In total, the calorimeters system has 187648 cells and about 375000 readout channels. A view of the calorimeters is shown in Fig. 3.4 and their main technical parameters are listed in Table 3.2.

- The **EMCal** is a Pb–LAr (liquid Argon) detector with accordion-shaped electrodes and Pb absorber plates. It extends along the same η range as the ID and is divided into a barrel part, with three layers, and two wheel sections at the endcaps. The accordion geometry provides full φ -coverage without azimuthal cracks. When charged particles cross the active material (LAr), ion and electron pairs are created and drifted in opposite directions by an electric field, and finally collected by kapton electrodes. In the central region, a presampler detector is used to correct for the energy lost by electrons and photons in the dead parts upstream of the calorimeter. A sketch of a LAr module is shown in Fig. 3.5. The EMCal provides a resolution in energy of

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \quad (3.2)$$

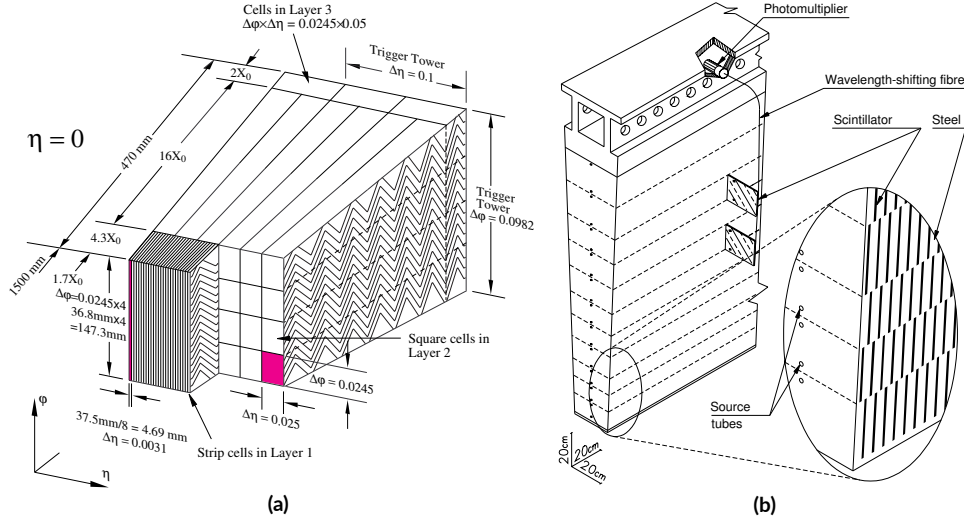


Figure 3.5: Schematic view of (a) an EMCal LAr and (b) a TileCal modules [122].

- The **HCal** comprises three independent sections: the Tile calorimeter (TileCal), the LAr hadronic endcap calorimeter (HEC) and the LAr forward calorimeter (FCal).
 - The **TileCal**, placed right over the EMCal, consists of a barrel and two extended barrels. It is composed by tiles of scintillating plastic (active material) and low-carbon steel (absorber). The tiles emit light when a charged particle crosses, and this light is collected and converted in pulses by photomultipliers. A sketch of a TileCal module is shown on the right-side of figure 3.5.
 - The **HEC** consists of two independent wheels made of LAr (active medium) and Cu plates (absorber), placed after the EM calorimeter ones at the endcaps.
 - The **FCal** is formed by three modules located 1.2 m away from the EMCal at the endcaps. All of them use LAr as active material, but the first one has copper as absorber and is optimized for electromagnetic measurements, while the other two have tungsten instead and take care of the hadronic interactions.

The HCal has a resolution of

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \text{ (TileCal and HEC)} \quad (3.3)$$

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \text{ (FCal)}. \quad (3.4)$$

		Barrel	Endcap	
EM calorimeter				
Number of layers and $ \eta $ coverage				
Presampler	1	$ \eta < 1.52$	1	$1.5 < \eta < 1.8$
Calorimeter	3	$ \eta < 1.35$	2	$1.375 < \eta < 1.5$
	2	$1.35 < \eta < 1.475$	3	$1.5 < \eta < 2.5$
			2	$2.5 < \eta < 3.2$
Granularity $\Delta\eta \times \Delta\phi$ versus $ \eta $				
Presampler	0.025×0.1	$ \eta < 1.52$	0.025×0.1	$1.5 < \eta < 1.8$
Calorimeter 1st layer	$0.025/8 \times 0.1$	$ \eta < 1.40$	0.050×0.1	$1.375 < \eta < 1.425$
	0.025×0.025	$1.40 < \eta < 1.475$	0.025×0.1	$1.425 < \eta < 1.5$
			$0.025/8 \times 0.1$	$1.5 < \eta < 1.8$
			$0.025/6 \times 0.1$	$1.8 < \eta < 2.0$
			$0.025/4 \times 0.1$	$2.0 < \eta < 2.4$
			0.025×0.1	$2.4 < \eta < 2.5$
Calorimeter 2nd layer	0.025×0.025	$ \eta < 1.40$	0.050×0.025	$1.375 < \eta < 1.425$
	0.075×0.025	$1.40 < \eta < 1.475$	0.025×0.025	$1.425 < \eta < 2.5$
			0.1×0.1	$2.5 < \eta < 3.2$
Calorimeter 3rd layer	0.050×0.025	$ \eta < 1.35$	0.050×0.025	$1.5 < \eta < 2.5$
Number of readout channels				
Presampler	7808		1536 (both sides)	
Calorimeter	101760		62208 (both sides)	
LAr hadronic endcap				
$ \eta $ coverage			$1.5 < \eta < 3.2$	
Granularity $\Delta\eta \times \Delta\phi$			0.1×0.1	$1.5 < \eta < 2.5$
			0.2×0.2	$2.5 < \eta < 3.2$
Readout channels			5632 (both sides)	
LAr forward calorimeter				
$ \eta $ coverage			$3.1 < \eta < 4.9$	
Granularity $\Delta x \times \Delta y$ (cm)			FCal1: 3.0×2.6	$3.15 < \eta < 4.30$
			FCal1: \sim four times finer	$3.10 < \eta < 3.15,$ $4.30 < \eta < 4.83$
			FCal2: 3.3×4.2	$3.24 < \eta < 4.50$
			FCal2: \sim four times finer	$3.20 < \eta < 3.24,$ $4.50 < \eta < 4.81$
			FCal3: 5.4×4.7	$3.32 < \eta < 4.60$
			FCal3: \sim four times finer	$3.29 < \eta < 3.32,$ $4.60 < \eta < 4.75$
	Readout channels		3524 (both sides)	
	Tile calorimeter			
	Barrel	Extended barrel		
$ \eta $ coverage	$ \eta < 1.0$	$0.8 < \eta < 1.7$		
Granularity $\Delta\eta \times \Delta\phi$	0.1×0.1	0.1×0.1		
Last layer	0.2×0.1	0.2×0.1		
Readout channels	5760	4092 (both sides)		

Table 3.2: Main parameters of the ATLAS calorimeters system [122].

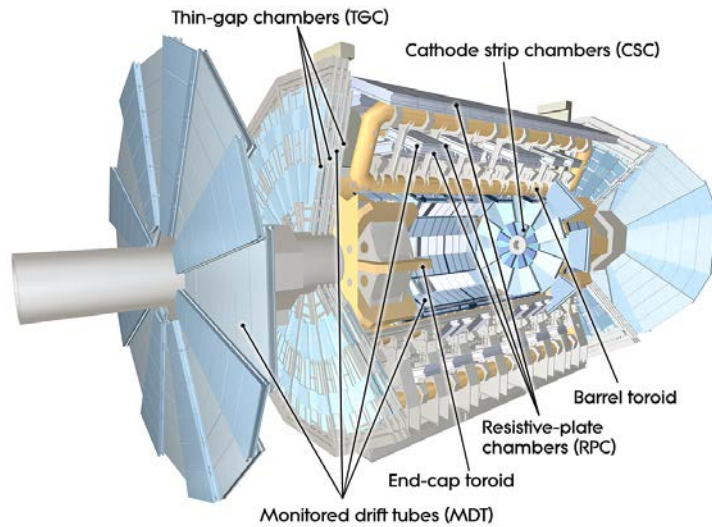


Figure 3.6: Schematic view of the ATLAS muon spectrometers and the toroidal magnets [122].

3.2.3 MUON SPECTROMETERS

The MS are the most external part of the ATLAS detector and have the role of identifying and measuring high- p_T muons. A global view is shown in Figure 3.6, where its several components can be identified: the Monitored Drift Tubes (MDT), the Cathode Strip Chambers (CSC), the Resistive Plate Chambers (RPC) and the Thin Gap Chambers (TGC), all detailed below with their main technical aspects listed in Table 3.3. Moreover, the MS is immersed in a toroidal magnetic field which deflects the trajectories of the muons. The magnetic system is divided into three toroids: one in the central part ($|\eta| < 1.4$), providing a 0.5 T field and another two in both sides of the barrel that create a field of 1 T.

- The **MDT** are aluminium tubes filled with a mixture of Ar (93%) and CO₂ (7%) and a tungsten-rhenium wire inserted in the center. The passing muons ionize the gas, producing electrons that are collected by the wire. The MDT provide precision coordinate measurement in the bending direction of the toroidal magnet, therefore providing the muon momentum measurement in the central region (see Table 3.3).
- The **CSC** are multi-wire proportional chambers with two cathodes segmented into strips and filled with the same gas mixture as the MDT. They are located in the closest layer to the beam pipe at the endcaps, and given the more demanding conditions of the region they are designed to provide high resolution and fast time response.

	MDT	CSC	RPC	TGC
Function	Precision tracking	Precision tracking	Trigger	Trigger
$ \eta $ coverage	< 2.7 (< 2.0 innermost layer)	2.0–2.7	< 1.05	1.05–2.7
Resolution (z/R)	$35 \mu\text{m}$ (z)	$40 \mu\text{m}$ (R)	10 mm (z)	2.6 m (R)
Resolution (φ)	–	5 mm	10 mm	3–7 mm
Resolution (time)	–	7 ns	1.5 ns	4 ns

Table 3.3: Technical parameters of the ATLAS Muon Spectrometers [122]

- The **RPC** and the **TGC** are trigger chambers located in the barrel and the endcaps, respectively. They are used for the trigger and provide measurement of the muon φ coordinate, as well as bunch-crossing identification and well defined p_T thresholds.

3.2.4 TRIGGER SYSTEM

The ratio of the pp collisions produced at the LHC is of the order of bunches of $\sim 1.5 \times 10^{11}$ protons crossing each other every 25 ns. This means that ATLAS has to handle up to ~ 1.7 billion collisions per second, translating into more than 60 Mb/s. This rate needs to be reduced, firstly because the storage of such amount of data is not manageable, and secondly because most of the processes that occur in the collisions are rather not interesting for physics analyses. The Trigger and Data Acquisition (TDAQ) systems help to reduce the flow of data by deciding whether or not to save a collision event from a given bunch-crossing (BC) interaction. The event that meets any of the multiple conditions considered (trigger menu) will pass the trigger and be stored for later study. The trigger system has two levels: the **L1**, fully hardware based, and the High Level Trigger (HLT), a software-based system that uses similar algorithms to the ones used for the offline reconstruction.

- The **L1** trigger uses information provided by RPC and TGC components of the MS (**L1**Muon), the calorimeters system (**L1**Calo) and topological clusters (**L1**Topo), from where it identifies Regions of Interest (RoI's). The final decision is made by the Central Trigger Processor (CTP), which is also responsible for limiting the minimum time between two consecutive **L1** accepts. The CTP provides 512 possible trigger selections, each one with a *prescale* of $n \geq 1^d$ such that $1/n$ random events passing the selection are accepted by the CTP. Most of the analyses are interested in events with very high- p_T objects in the final state and make use of unprescaled triggers, hence using all the events selected by the trigger ($n = 1$). If the event passes the filter,

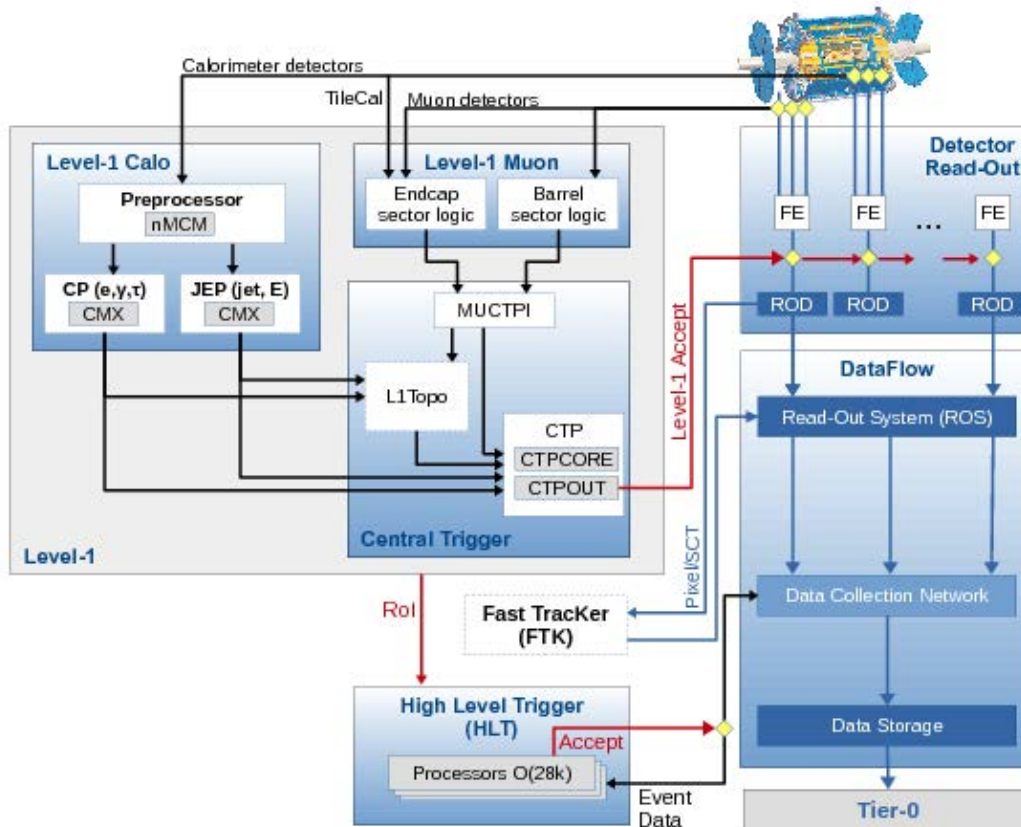


Figure 3.7: Scheme of the ATLAS TDAQ systems in Run 2 [127].

the L1 will buffer the RoIs in the Readout System (ROS). Globally, the L1 reduces the event rate from 30 MHz to 100 kHz, within a decision time of $2.5 \mu\text{s}$.

- The **HLT** is a large farm of CPU's that takes the RoI's from L1 as input. Using fully reconstructed data, the HLT applies a chain of selection steps (the trigger menu). If the event passes the selection it will be stored on disk at the CERN Tier-0 centre. The HLT reduces the event rate up to 1 kHz within approximately 200 ms.

3.2.5 LUMINOSITY DEFINITION AND GOOD RUN LIST

The luminosity \mathcal{L} is defined as the ability of a particle collider to produce a required number of interactions of a given physical process, and it is the proportionality factor that relates the observed event rate dN/dt and the cross-section σ for such process,

$$\frac{dN}{dt} = \sigma \mathcal{L}. \quad (3.5)$$

A precise measurement of the luminosity is instrumental for any experiment in particle physics, especially in cross-section measurements but also in new physics searches, where an accurate determination of the luminosity is necessary in order to provide a good description of the background predictions and the sensitivity to new phenomena signatures.

The *instantaneous luminosity*, defined by Eq. (3.5), is expressed in $\text{cm}^{-2}\text{s}^{-1}$ and reflects the performance of the collider. It depends on the number of collisions, so it will decrease over time for each fill due to fewer protons from one BC to another*. The *integrated luminosity*, $\int \mathcal{L} dt$, concerns the instantaneous luminosity accumulated over a certain period of time and is usually quoted in units of inverse barns (inverse picobarns pb^{-1} , femtobarns fb^{-1} , etc). A distinction must be made between *absolute* and *relative* luminosity: the first one is expressed on a given absolute scale through a calibration procedure, while the later refers to the monitoring of the relative variations of the instantaneous luminosity over time.

At the LHC, the beams are squeezed by a system of quadrupole magnets near to the collision points, increasing the proton density with respect to the collisional cross-section, β^* , and thus the probability of collision. The peak luminosity at the start of the fills ($\mathcal{L}_{\text{peak}}$) raised significantly during the Run 2 as the beam conditions were evolving, increasing the number of bunches, n_b , and reducing the spatial sizes of the beams, as shown in Table 3.4. In ATLAS, the luminosity calibration relies on multiple and redundant sub-detectors and algorithms [128], being the Van der Meer (vdM) method the main approach that is used. First, the instantaneous luminosity is given in terms of the beam parameters as

$$\mathcal{L} = \frac{n_b f n_1 n_2}{2\pi \Sigma_x \Sigma_y}, \quad (3.6)$$

*In ATLAS, a run is defined as the period of data-taking before a new fill. Runs are typically divided into time intervals where the instantaneous luminosity is about constant, called *lumiblocks*

	2015	2016	2017	2018	Comb.
n_b	2232	2208	2544	2544	–
β^* [m]	0.8	0.4	0.3	0.3 – 0.25	–
$\mathcal{L}_{\text{peak}}$ [$10^{33} \text{ cm}^{-2}\text{s}^{-1}$]	5	13	16	19	–
Total delivered luminosity $\int \mathcal{L} dt$ [fb^{-1}]	4.0	38.5	50.2	63.4	139.0
Total uncertainty $\delta\mathcal{L}/\mathcal{L}$ [%]		2.1	2.4	2.0	1.7

Table 3.4: Selected LHC parameters, total integrated luminosity and total uncertainties in Run 2 [128].

where f_r is the LHC revolution frequency and n_1 (n_2) the number of protons in the bunch 1 (2). Σ_x and Σ_y are the convoluted widths of the beam in the horizontal and vertical directions, respectively, which can be determined by vdM scans [129], using dedicated low-intensity runs where the beams are moved in controlled steps. Secondly, Eq. (3.5) can be re-written taking the inelastic pp cross-section σ_{inel} as reference[†], so the luminosity can be computed in a beam-independent way, after applying some event selection, as

$$\mathcal{L} = \frac{(dN/dt)_{\text{inel}}}{\sigma_{\text{inel}}} = \frac{\langle\mu\rangle n_b f_r}{\sigma_{\text{inel}}} = \frac{\langle\mu\rangle_{\text{vis}} n_b f_r}{\sigma_{\text{vis}}}, \quad (3.7)$$

where $(dN/dt)_{\text{inel}}$ is the rate of inelastic collisions and $\langle\mu\rangle$ the mean number of inelastic interactions per BC, averaged over all colliding bunch-pairs (μ is known also as the *pile-up* parameter). If ε is the efficiency for one inelastic collision to pass the selection criteria, $\langle\mu\rangle_{\text{vis}} = \varepsilon \langle\mu\rangle$ is the visible number of interactions per BC and $\sigma_{\text{vis}} = \varepsilon \sigma_{\text{inel}}$ is the visible cross-section, which can be obtained by combining Eqs. (3.6) and (3.7). The visible cross-section σ_{vis} is the absolute calibration constant specific for every sub-detector and algorithm, and is determined using vdM scans once a year. The luminosity can be characterized as a function of the pile-up via Eq. (3.7), being this the method of monitoring the luminosity bunch-by-bunch in ATLAS. Figure 3.8 shows the relative luminosity recorded by ATLAS as a function of μ in Run 2.

The final integrated luminosity used in physics analyses is determined taking only into account what is called ‘good data’, so the status of the detector is permanently monitored and recorded by

[†]The luminosity depends on the total cross-section, which is not accessible since the elastic cross-section σ_{el} is not measurable at $t=0$, and so the diffractive term. Hence the luminosity calibration depends mostly on the efficiency of the detector to measure σ_{inel} . There is an alternative approach, used in ATLAS as a complementary measurement, using the Roman Pots [130] to measure the proton distribution at the forward region, determining σ_{el} at very small angles so the total cross-section can be estimated via the Optical Theorem.

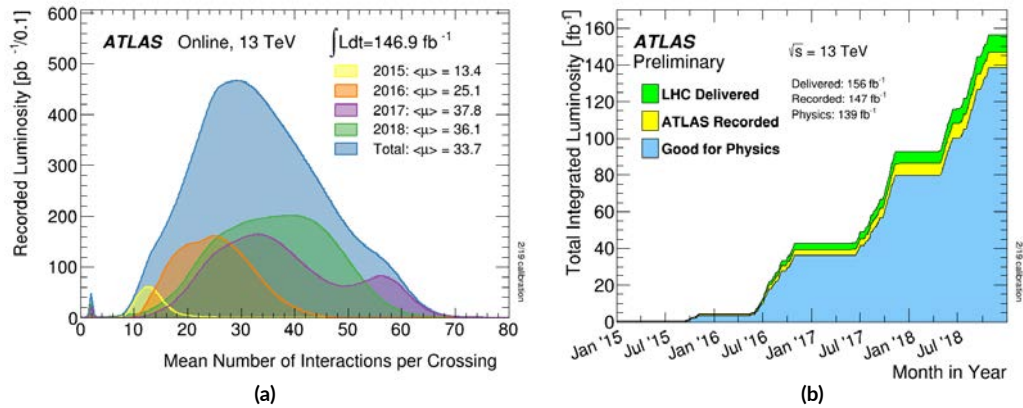


Figure 3.8: (a) Luminosity-weighted distribution of the mean number of interactions per BC for the 2015–2018 pp collision data at $\sqrt{s} = 13 \text{ TeV}$. (b) Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified as good-quality data (blue) during stable beams for pp collisions at $\sqrt{s} = 13 \text{ TeV}$ in 2015–2018.

detector experts, that provide feedback on the quality of the data by delivering the so-called **Good Run List (GRL)**, containing the recommended runs for analysis use. The total integrated luminosity recorded by ATLAS in Run 2 is shown in Fig. 3.8 and Table 3.4, with a result of 139 fb^{-1} of good data and a total uncertainty of 1.7%.

4

Event Reconstruction

The collision events recorded by ATLAS are reconstructed using the information provided by the different sub-detectors, in order to be available for further analysis. This Chapter is focused on the reconstruction and identification of the physical objects that are most relevant for the analysis of this thesis, i.e. electrons, muons, taus, photons, jets, b -jets and E_T^{miss} . However, it is convenient to discuss first the tracks and vertices, since they are essential in the reconstruction of all the rest of the objects.

4.1 TRACKS AND PRIMARY VERTEX

4.1.1 TRACK RECONSTRUCTION

A track is defined as the trajectory of an electrically charged particle through the detector. Assuming no multiple scattering and negligible Bremsstrahlung losses, such trajectory of a charged particle in a solenoidal magnetic field can be described as a helix. Therefore the track can be parametrized by a set of 5 parameters $P = P(d_0, z_0, \varphi, \theta, q/p)$, where d_0 and z_0 are the transverse and longitudinal impact parameters (minimum distance between the track and the centre of the detector), respectively, φ and

θ the corresponding longitudinal and azimuthal angles and q/p is the ratio between the electric charge and the momentum of the particle. By convention, all these quantities are measured at the distance of closest approach to the z -axis (known as the track perigee).

The reconstruction of the tracks relies on the use of hits (i.e., local ionized signals released by the particle in the detector on its way through) left in the ID, and it is performed by several algorithms [131, 132], being the most common the ones so-called *inside-out* and *outside-in* finding methods. First, the hits in the Pixel Detector and the SCT are clustered into *space-points* (defined as three-dimensional representations of individual particle-detector interactions). Groups of three space-points are used to set the seed for the inside-out finding algorithm, which estimates a preliminary trajectory P by extrapolating from the IBL towards the outermost layers of the ID, adding hits and recomputing the goodness χ^2 of the track fit. Hits with large χ^2 will be tagged as *outliers* and removed from the track candidate. Once the inside-out sequence has finished, the complementary outside-in method is ran, starting from hits in the TRT that were not assigned to any track candidate during the previous stage, and extrapolating all the way down to the centre of the ID to find associated hits that were not matched to inside-out tracks. Space-points can not be constructed in the TRT due to its inability to provide coordinate information, so TRT segments are used to seed tracks instead.

This whole process results in a list of track candidates that will be sorted according to a *track score*, based on the energy of the track, the number of shared hits with other tracks and the total number of hits and holes (sensors where a hit is expected given the trajectory extrapolated, but none is found). Finally, candidates with the lowest scores are removed and further requirements are applied in order to keep only the best quality tracks.

4.1.2 VERTEX RECONSTRUCTION

Vertices are points where a physical interaction occurs. In particular, the point where the pp interaction took place in a collision event is called Primary Vertex (PV). The number of PVs gives an estimation of the pile-up, which is essential for the analysis.

Once the tracks reconstruction step is completed, the vertices are reconstructed as follows [133]. First, the seed for the vertex finding algorithm is set such that its transverse position is taken as the center of the beam spot (defined as the spatial region around where the beam profiles overlap), and the z -coordinate is chosen as the global maximum in the z -coordinate distribution among all tracks.

Tracks within 7σ of the seed, $p_T > 400$ MeV, 9 SCT hits with no holes and $|d_0| < 4$ mm are considered as candidates for inclusion in the reconstructed vertex. Then, the Adaptive Vertex Fitting algorithm [134], runs over the ID track candidates an iterative fit in order to estimate the vertex position. As the fit progresses, tracks more compatible with the seed position are assigned higher weights, and vice-versa. Once no significant change is noted, the iterative procedure stops and the PV position is defined, while tracks that are not compatible are removed and used in the determination of another vertex. This method is repeated until all tracks in the event are assigned to a PV.

The vertex with the largest sum of the squared momenta of the associated tracks, $\sum p_T^2$, is defined as the main vertex, the source of the event, while the rest are considered as pile-up vertices. Once all PVs are defined, the impact parameters of their associated tracks are re-calculated with respect to the position of the vertex. Finally, vertices originated in secondary interactions, such as particle decays, are called Secondary Vertices and can be reconstructed by looking at the displacement of tracks with respect to the PV. The identification of the SV is of especial importance in analysis and searches B -physics related, but in the analysis of this thesis they are only used in the identification of b -hadron initiated jets.

4.2 ELECTRONS

4.2.1 ELECTRON RECONSTRUCTION

Electrons are reconstructed by using combined information from energy deposits within the calorimeter system and tracks in the ID. The reconstruction process works as follows.

First, topo-clusters are formed by collecting groups of cells in the EMCal and HCal according to the cell energy and certain noise thresholds criteria [135]. Topo-clusters with an EM energy (defined as the sum of energies only from EMCal cells) higher than 400 MeV and a fraction of EM energy over the total cluster energy larger than 0.5 are selected. The process continues by re-fitting and extrapolating tracks from the ID up to the second layer of the EMCal, since it is where most of the energy of the electrons is deposited, given the small depth of the first layer and because low-energy electrons will not have enough energy left for outer layers. The tracks are matched to topo-clusters such that they must be at most $\Delta\eta = 0.05$ away and satisfy $-0.1 < q\Delta\phi < 0.05$, where q is the reconstructed charge of the track. In the case of multiple tracks compatible with the matching criteria, those with hits in the Pixel Detector are preferred, and the one with the smallest ΔR is chosen.

After the track-matching stage is completed, EM topo-clusters with a transverse energy E_T higher

than 1 GeV and matched to a track with at least four SCT hits are targeted as seed clusters. Clusters within a $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ rectangle around the seed are considered as satellite clusters. The seed clusters with their associated satellites are called ‘*superclusters*’. After applying initial position corrections, the energy is initially estimated by the cells in the first three layers of the EMCal associated to the supercluster. Then, the track-matching procedure is repeated, this time using the superclusters instead. Since superclusters are built independently, a given seed can produce both an electron or a photon supercluster (photons reconstruction will be discussed in the next section). In cases where the object can be easily identified as an electron (good-quality track, no conversion vertex) or a photon (no good track associated), only an electron or a photon will be created for analysis, but in those cases where the ambiguity is especially explicit both will be saved, allowing the final classification for each analysis.

Finally, once all electron candidates are defined, their energies are calibrated via multivariate techniques based on the properties of the shower development in the EMCal [136]. Simulated $Z \rightarrow ee$ events are used to adjust the absolute energy scale, validated with $J/\Psi \rightarrow ee$ simulated samples. The efficiency of the electron reconstruction is estimated with a tag-and-probe method using MC $Z \rightarrow ee$ samples, and provided in bins of E_T and η . Figure 4.1 illustrates the efficiency of the electron reconstruction at the different stages as a function of E_T , showing how the total efficiency is dominated by the effect of the cluster reconstruction, which starts being efficient only for electrons with $E_T > 5$ GeV.

4.2.2 ELECTRON IDENTIFICATION

Not all reconstructed electrons are prompt electrons, i.e., not all of them come from the main pp interaction or from the decay of heavy resonances such as Higgs or W/Z bosons. Some might come from light-flavour jets, photon conversions, or from semi-leptonic decays from heavy-flavour hadrons. Electron identification algorithms are employed to select such signal electrons among the full set of candidates. This method uses a likelihood function which takes inputs from the tracking and calorimeter systems, but also quantities that combine both tracking and calorimeter information [135]. The likelihood function is composed with the signal and background pdfs, which are constructed based on $Z \rightarrow ee$ (for $E_T > 15$ GeV) and $J/\Psi \rightarrow ee$ ($E_T < 15$ GeV) events of data recorded in 2015 and 2016, via a tag-and-probe method described in Ref. [137].

A multivariate analysis is then performed, and several identification criteria are defined, which are usually known as *working points* (WPs). The WPs are provided for analysis use and are named, in increasing order of threshold for the likelihood discriminant (and therefore decreasing efficiency), LOOSE, MEDIUM and TIGHT. The LOOSE WP is optimized for light-flavour jets and photon conversion rejection.

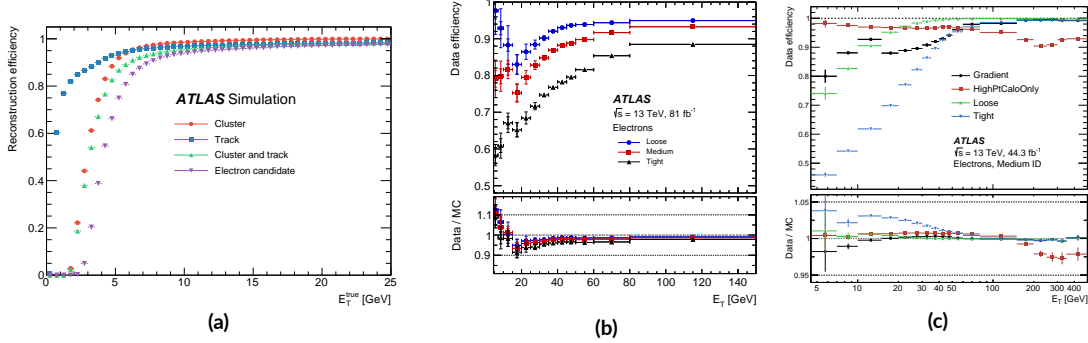


Figure 4.1: (a) Cluster, track, cluster and track, and electron reconstruction efficiencies as a function of the generated electron E_T . (b) Electron identification efficiency in $Z \rightarrow ee$ events in data as a function of E_T for the Loose, Medium and Tight WPs. The efficiencies are obtained by applying data-to-simulation efficiency ratios measured in $J/\Psi \rightarrow ee$ and $Z \rightarrow ee$ events to $Z \rightarrow ee$ simulations. (c) Efficiency of the different isolation WPs for electrons from inclusive $Z \rightarrow ee$ events as a function of the electron E_T . The electrons are required to fulfil the Medium identification selection. For the later two plots, the bottom panel shows the Data/MC ratios [135].

tion, while MEDIUM and TIGHT include additional variables for better suppression of heavy-flavour jets. The efficiency of the electron identification is calculated in bins of E_T and η via a tag-and-probe method using $Z \rightarrow ee$ and J/Ψ MC samples [138]. As shown in Fig. 4.1, the efficiencies for identifying a prompt electron for the LOOSE, MEDIUM and TIGHT WPs with E_T are, on average, 93%, 88% and 80%, respectively, and gradually increase with E_T .

4.2.3 ELECTRON ISOLATION

Even electrons that pass the identification can still be contaminated by background processes such as the ones mentioned in the previous subsection. Such signals leave a characteristic signature represented by some small activity in the surrounding area of the candidate, in both calorimeter and inner detector systems. Isolation variables are constructed to quantify this amount of activity, in order to further purify the electron sample:

- **Calorimeter-based isolation:** the variable $E_T^{\text{cone}\Delta R}$ is used, defined as the sum of the transverse energy of topo-clusters within a cone of a typical size $\Delta R = 0.2$ around the supercluster barycentre. The core energy is subtracted by removing cells in the central rectangle $\Delta\eta \times \Delta\phi = 5 \times 7$ (in EMCal middle layer units). Additional fine-tune techniques are employed to suppress pile-up and underlying event contributions (more details can be found in Ref. [139]), as well as core leakage corrections.

- Track-based isolation: similarly, the variable $p_T^{\text{cone}\Delta R}$ is defined as the sum of the p_T of a subset of tracks inside a cone around the electron. Tracks with $p_T > 1$ GeV, $|\eta| < 2.5$, and fulfilling basic good-quality track criteria are used. An additional requirement on $|z_0| \sin \theta < 3$ mm is applied to minimize contributions from pile-up. The track- p_T contribution of the candidate electron is subtracted from the cone.

Like at the identification stage, several isolation WPs are established, allowing to an analysis-specific implementation. The WPs are defined either targeting a fixed value of efficiency, or with fixed cuts on the isolation variables. The `GRADIENT` WP is defined to give an efficiency of 90% and 99% at $E_T = 25$ and 60 GeV, respectively, uniform in η . The `HIGHPTCALOONLY`, `LOOSE` and `TIGHT` WPs are designed with fixed requirements on the track and/or calorimeter isolation variables. The electron isolation efficiency is evaluated by using simulated samples enriched in $Z \rightarrow ee$ events, complemented with a J/Ψ simulated sample for the $E_T < 15$ GeV regime, and provided in bins of electron E_T and η for analysis use. Figure 4.1 shows the isolation efficiency provided by different isolation WPs, for electrons satisfying the `MEDIUM` identification selection, as a function of E_T .

4.3 PHOTONS

Two types of photons are distinguished at reconstruction level: converted and unconverted. The difference is simply that converted photons are those that decay into a e^+e^- pair when interacting with the material in the EMCal.

4.3.1 PHOTON RECONSTRUCTION

Photons are reconstructed following an almost identical procedure as electrons, with small differences. The topo-clusters selection and track-matching steps are the same as described in previous section.

After that, the reconstruction of the conversion vertex is carried out. Converted photons would have two tracks pointing to clusters in the calorimeter, but often it might happen that the e^+e^- pair created is asymmetric and one of the tracks has very low p_T , so only the other one is reconstructed. These tracks usually do not have hits in the ID, hence matched tracks with hits in the SCT or the TRT are used as input for the reconstruction of the conversion vertex. Double-track vertices are reconstructed from two opposite-charge tracks converging into a vertex consistent with that of a massless

particle, while single-track vertices are basically tracks without hits in the innermost layers. Further requirements and cross-checks are implemented in order to reduce the fraction of unconverted photons mistakenly reconstructed as converted photons, due to tracks wrongly assigned to the photon cluster (more details can be found in Ref. [135]). Once reconstructed, the conversion vertices are matched to the topo-clusters. In case of multiple vertices matching, double-track conversions with SCT tracks are preferred, and the vertex with the smallest radius is chosen.

Clusters with $E_T > 1.5$ GeV are selected to set the seed for photon superclusters, with no requirements on tracks or conversion vertex matching. The superclusters are built then by adding satellite clusters around the seed, as described for electrons. Initial position corrections are applied to the photon supercluster, and its energy is estimated from cells in the EMCal. Then, the vertex-matching procedure is repeated, this time with the supercluster. As mentioned in the previous section, at this stage the reconstruction algorithm decides whether to save the reconstructed object as a photon or an electron, or even both, since the superclusters reconstruction is ran independently.

The energy of the reconstructed photons is calibrated following the same procedure prescribed for electrons [136], and validated by using $Z \rightarrow \ell\ell\gamma$ events. The reconstruction efficiency is evaluated using a clean photon sample of $Z \rightarrow \mu\mu\gamma$ selected events. Figure 4.2 shows the reconstruction efficiency for converted photons as a function of the E_T , where the combined effect of the reconstruction of the several types of tracks results in total reconstruction efficiency values of about 70% for converted photons with $E_T > 25$ GeV.

4.3.2 PHOTON IDENTIFICATION

Photon fake signatures might come from neutral hadron jets decay, or QCD jets depositing a large fraction of energy in the EMCal. Typically, prompt photons would leave narrower energy deposits in the EMCal and have smaller leakage to the HCal. The discrimination between prompt and non-prompt photons is therefore addressed via a ‘*cut-based selection*’, which uses a number of shower shape and energy variables described in Ref. [135]. Variables using the first EMCal layer are of special importance in rejecting the $\pi^0 \rightarrow \gamma\gamma$ background process. Three WPs are defined, namely LOOSE, MEDIUM and TIGHT, in increasing order of background rejection. The LOOSE and MEDIUM operation points, less restrictive, are used for trigger algorithms and are the same for converted and unconverted photons, since the ATLAS trigger system does not make such distinction. The TIGHT identification WP is optimized using a series of simulated samples, and performed separately for converted and unconverted

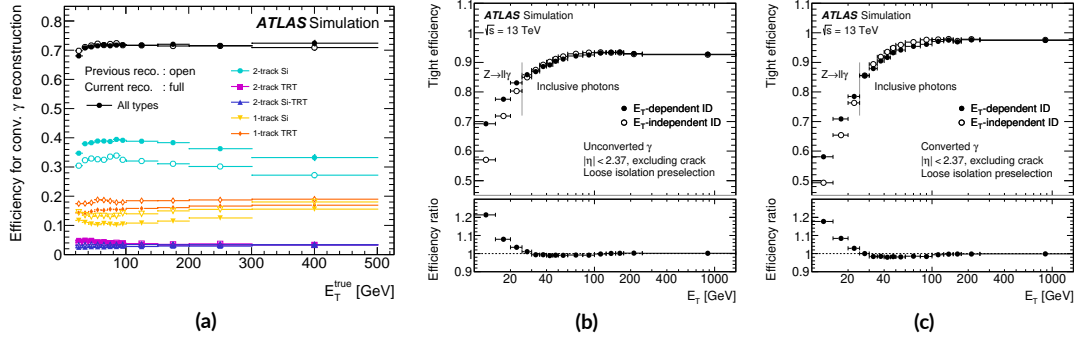


Figure 4.2: (a) Converted photon reconstruction efficiency and contributions of the different conversion types as a function of E_T^{true} . Efficiency of Tight photon identification for unconverted (b) and converted (c) photons as a function of photon E_T . The signal events are taken from a $Z \rightarrow \ell\ell\gamma$ sample with $E_T < 25$ GeV, and from inclusive-photon production simulated events with $E_T > 25$ GeV. In each case, the E_T -independent and E_T -dependent selections are compared and the bottom panels show the ratios between the E_T -dependent and independent efficiencies [135].

photons. The cut-based selection criteria for the three operation points are optimized separately in bins of $|\eta|$. Figure 4.2 shows the unconverted and converted photon identification efficiencies for the TIGHT WP as a function of E_T , as calculated with $Z \rightarrow \ell\ell\gamma$ simulated events for $10 < E_T < 25$ GeV and an inclusive-photon MC sample for $E_T > 25$ GeV. In the figure, the selection is also optimized in bins of E_T , and compared with a reference selection that does not change with E_T .

4.4 MUONS

Muon candidates are reconstructed and identified using mainly track information from the ID and the MS, and energy deposits in the calorimeter up to a lesser extent.

4.4.1 MUON RECONSTRUCTION

First, the muon tracks are reconstructed independently in the ID and the MS. ID tracks are reconstructed as described in Section 4.1. The reconstruction in the MS starts by looking for a hit pattern in the muon chambers in order to form segments [140]. Hits in the MDT are fitted to a straight line to create such segments. The segments in the CSC are formed by fitting hits in both η and ϕ planes. The RPC and TGC hits are used to measure the orthogonal coordinate to the bending plane. Then, the MS muon tracks are built by fitting segments from different layers altogether. The seed for the track

reconstruction is set in segments generated in the middle layers, where more trigger hits are available, and extending the search to the inner and outer layers.

Secondly, with the track information provided by the ID and the MS, four muon types are considered in the reconstruction process:

- Combined (CB) muons: a track is formed by a global fit combining hits from the ID and the MS. The fit follows an outside-in pattern, extrapolating from MS hits towards the ID. MS hits may be added or removed during the process in order to improve the fit quality. A complementary inside-out sequence is ran afterwards.
- Segment-tagged (ST) muons: a track from the ID is extrapolated up to the MS and matched to at least one segment in the MDT or the CSC. ST muons are used for low- p_T muons or just muons simply falling in MS regions with reduced acceptance.
- Calorimeter-tagged (CT) muons: ID extrapolated tracks are matched to an energy deposit in the calorimeters system if it is compatible with a minimum-ionizing particle. This is the lowest-purity type of muon but is useful for regions where the MS is only partially instrumented.
- Extrapolated (ME) muons: only tracks from the MS are used for the reconstruction. Loose requirements on compatibility with the PV are applied. This type of muons are used to extend the acceptance range in η up to the $2.5 < |\eta| < 2.7$ region, which is out of the ID coverage.

Overlaps between the different types of muons are resolved such that the CB muons type has the highest priority, followed by ST muons, CT muons and, finally, ME muons. Once the corresponding sets of reconstructed muons are defined, CB muons are used to set the calibration of the muon momentum. The calibration procedure, described in detail in Ref. [140], uses $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$ MC samples to identify corrections in the p_T distribution. Figure 4.3 shows the di-muon transverse mass distribution for a selection of $Z \rightarrow \mu\mu$ events, comparing the data to the MC predictions before (dashed grey line) and after (solid red line) the muon- p_T corrections are applied. The figure shows how the agreement between data and simulations becomes almost perfect except for small fluctuations at the tails of the distribution, far away from the mass of the Z .

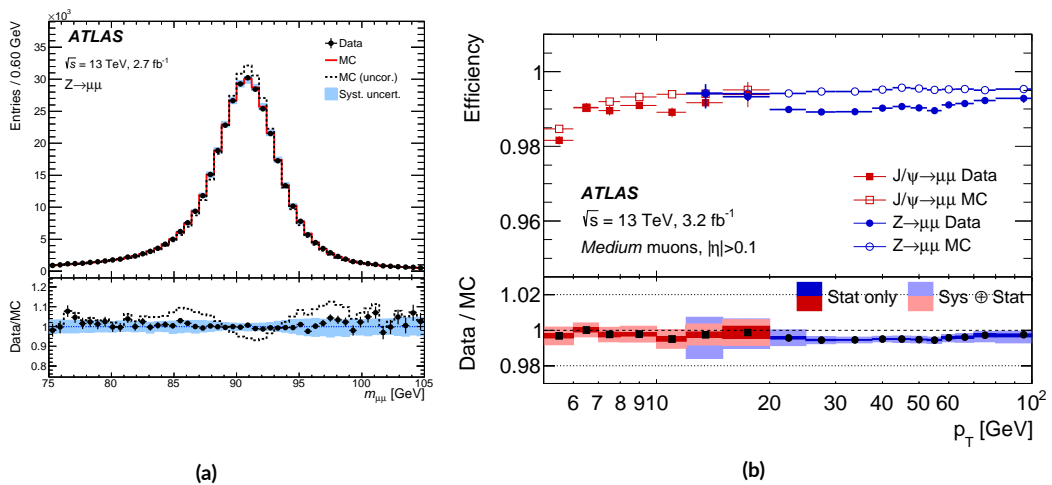


Figure 4.3: (a) Di-muon invariant mass distribution of $Z \rightarrow \mu\mu$ candidate events reconstructed with CB muons. The upper panel shows the distribution for data (black points) and for the signal simulation plus the background estimate. The continuous (dashed) line denotes the simulation with (without) the MC p_T corrections applied. The lower panels show the data to MC ratios. (b) Reconstruction efficiency for the Medium muon identification WP as a function of the muon p_T , in the region $0.1 < |\eta| < 2.5$ as obtained with $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$ events. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties [140].

4.4.2 MUON IDENTIFICATION

Prompt muons are identified mostly against pion and kaon decays backgrounds. Several discrimination variables are derived from MC $t\bar{t}$ events. Muons from W decays are considered as signal muons while those from light-hadron decays would be tagged as background. Such background muons normally leave a characteristic signature of a poorly fitted combined track, hence variables based on the number of hits and differences between ID and MS reconstructed tracks are employed for the discrimination. Four identification WPs are defined, provided for specific analysis use: **LOOSE** (optimized for Higgs four-lepton decay reconstruction, uses all types of muons), **MEDIUM** (default selection in ATLAS, uses only CM and ME muons), **TIGHT** (designed to maximize the purity of the reconstructed sample sacrificing some efficiency, only a subset of CB muons passing the **MEDIUM** criteria is used) and **HIGH- p_T** (aiming to maximize the momentum resolution for $p_T > 100$ GeV, CB muons passing the **MEDIUM** selection and with at least three MS hits are used).

The muon reconstruction efficiency is measured via a tag-and-probe method [140] using simulated samples of $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$ events. Efficiencies above 98% are obtained for the **LOOSE** and **MEDIUM** muons, as shown in Figure 4.3, where the reconstruction efficiency for **MEDIUM** muons

is drawn as a function of the muon p_T . Typical efficiencies between 90–98% are obtained for the TIGHT WP.

4.5 JETS

Jets are flows of particles fired from the fragmentation of a high-energetic quark or a gluon via strong interaction. They are the most common objects produced in pp collisions and one of the major components of the monojet final state. Typical jets signatures in the detector are collimated showers in the calorimeters with associated tracks in the ID.

4.5.1 JET RECONSTRUCTION

Jets considered in the analysis of this thesis are reconstructed using the *anti- k_t* jet finding algorithm [141], as implemented in the FASTJET software tool [142]. The algorithm takes inputs in the form of four-vector objects called ‘*PseudoJets*’, which are constructed from energy deposits in the calorimeter.

Cells in the calorimeter are grouped into topo-clusters using a nearest-neighbour algorithm described in Ref. [143], selecting cells according to their signal-over-noise ratio around a seed cell with a highly significant signal value over the expected noise. Contributions from pile-up and electronic noise are taken into account when defining the expected noise thresholds. The energy of the topo-cluster is given by the sum of the cell energies and its position, initially given by the cells, is corrected to be consistent with a jet originated at the PV. Only topo-clusters with positive energy are used as PseudoJets.

The anti- k_t recombination sequence runs over the full list of PseudoJets and decides whether to cluster the elements i and j of the list into a jet based on two parameters: the distance d_{ij} between the elements i and j and the distance d_{iB} between the i -th element and the beam, defined both as

$$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2}) \frac{\Delta_{ij}^2}{R^2} \quad (4.1)$$

and

$$d_{iB} = k_{ti}^{-2}, \quad (4.2)$$

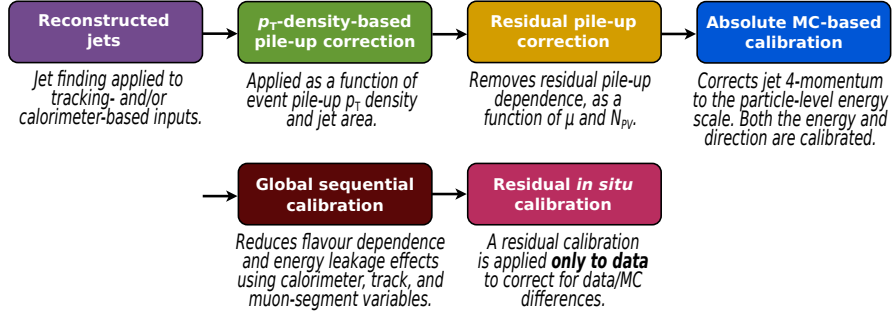


Figure 4.4: Schematic view of the jet calibration procedure [145].

where k_{T_i} (k_{T_j}) is the transverse momentum of the element i (j), $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\varphi_i - \varphi_j)^2$ is the squared distance between the elements i and j in the η - φ plane and R is the parameter that approximates the size of the jet. In the analysis of this thesis, jets are reconstructed with this parameter tuned to a value of $R = 0.4$. For a given element i , the algorithm computes the two quantities defined by Eqs. (4.1) and (4.2) with respect to all the rest of elements in the list, and identifies the smallest number obtained. If the minimum is d_{ij} (for any j), the elements i and j are recombined, while otherwise the i -th element is called a jet and removed from the list. The distances are recalculated and the procedure is repeated until the list is empty.

Only jets identified by the anti- k_t algorithm that pass a p_T threshold of 7 GeV are reconstructed. Finally, good-quality tracks with $p_T > 400$ GeV and associated to the PV are matched to the reconstructed jets via ghost association, a procedure described in Ref. [144].

4.5.2 JET CALIBRATION

The energy of the reconstructed jets is initially estimated at the EM scale. A calibration procedure is carried out in order to restore the energy of the jets to that of jets reconstructed in the MC simulations (usually referred to as particle, or truth, level) [145]. Simulation-based calibrations are first applied to correct for pile-up effects and set the absolute energy scale, followed by ‘in-situ’ corrections to account for differences between simulation and data. Figure 4.4 illustrates the full chain of corrections applied.

PILE-UP CORRECTIONS

The measured energy of the reconstructed jets may be affected by contributions that are not coming from the collision event of interest but from additional pp collisions. These contributions is what is usually called as pile-up, and they can be of two types: in-time pile-up (extra collisions within the same bunch crossing) or out-of-time pile-up (interactions from surrounding bunch-crossings). Since pile-up can alter the measurements in different ways, corrections are applied in two steps.

First, a jet area-based method [139] is employed to subtract pile-up contributions per event to the p_T of the jets. A number of simulated ghost particles are added homogeneously before jet reconstruction, which is performed with the k_t algorithm [146] with $R = 0.4$ ^{*}, and the area of a jet is then estimated from the relative number of particles associated to the jet after clustering. The jet area A is a measure of the susceptibility of the jet to pile-up. The pile-up contribution is then calculated from the median of the p_T density in the y - ϕ plane,

$$\rho = \text{median} \left\{ \frac{p_T}{A} \right\}, \quad (4.3)$$

The ratio of the jet p_T with ρ subtracted, as defined in Eq. (4.3), to the uncorrected jet p_T provides an offset correction for pile-up. This correction is applied as a global scale factor to the jet energy and does not affect to the y and ϕ coordinates.

The estimation of ρ given by Eq. (4.3) is derived from central jets ($|\eta| < 2$), which gives a rather meaningful measure of the pile-up activity than when using the entire η range, but it does not fully describe the pile-up conditions in the forward region, where a residual contribution still remains left. Consequently, a residual pile-up correction is derived. There is an observed dependency of the p_T of the jets after the area-based correction on the number of PVs, N_{PV} , and μ , the average number of interactions per bunch-crossing defined by Eq. (3.7). The correction is derived from the difference in p_T between the reconstructed jet and the jet at truth level. Such correction is estimated as a function of N_{PV} and μ , with a linear dependence observed for both variables, independent from each other. The corrected jet p_T after both area-based and residual pile-up corrections is then given by

$$p_T^{\text{corr}} = p_T^{\text{reco}} - A\rho - \alpha(N_{PV} - 1) - \beta\mu, \quad (4.4)$$

where p_T^{reco} denotes the p_T of the jets before any pile-up correction and α and β represent the residual

^{*}Here the k_t algorithm is used because it typically reconstructs jets including an uniform soft background [141].

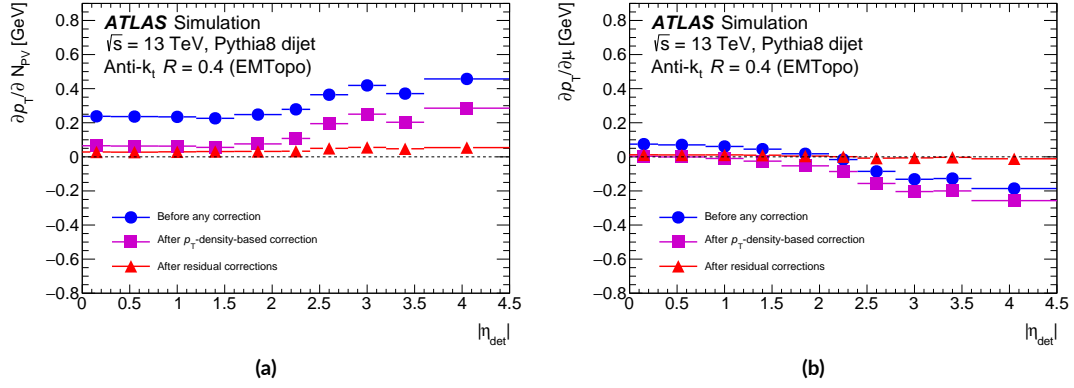


Figure 4.5: Dependence of jet p_T on in-time (a) and out-of-time (b) pile-up, as a function of $|\eta|$ for $p_T^{\text{truth}} = 25$ GeV [145].

p_T dependences with N_{pV} and μ , respectively, which are calculated from linear fits in simulations. Figure 4.5 shows the dependence of the corrections applied on N_{pV} and μ as a function of η for simulated jets with $p_T = 25$ GeV, separately for in-time and out-of-time jets, where the offset effect of the jet area-based correction is clearly visible, and also how the residual corrections become relevant only for $|\eta|$ values above 1.5, translating into an overall flat η distribution after both corrections.

ENERGY SCALE CALIBRATION

The energy of the reconstructed jets is corrected to the scale of truth jets, and mismeasurements in the η distribution caused by inhomogeneities in the calorimeter are fixed as well. This is again addressed with simulations. $R = 0.4$ anti- k_t truth jets from MC di-jet generated samples, and with the pile-up corrections applied, are used. Reconstructed jets are geometrically matched to the truth jets within $\Delta R = 0.3$, selecting only isolated ones [147].

The average jet energy response, \mathcal{R} , is estimated as the mean of a Gaussian fit to the $E^{\text{reco}}/E^{\text{truth}}$ distribution, in bins of E^{truth} and jet η . The corrections are parametrized as a function of E^{reco} using a numerical inversion method, described in Ref [148]. Such method works in two steps: first, computes the jet energy response as a function of E^{truth} like $\mathcal{R}(E^{\text{truth}}) = E^{\text{reco}}/E^{\text{truth}}$; then, estimates $\mathcal{R}(E^{\text{reco}})$ by using $E^{\text{reco}} = \mathcal{R}(E^{\text{truth}}) \times E^{\text{truth}}$. The resulting corrections range between 10 and about 50%, being larger for lower values of E^{reco} and depending of the η range of the jet, as shown in Figure 4.6a.

After correcting the jet energy scale, a remaining bias in the η distribution is observed, as shown in Figure 4.6b. This is caused by changes in the calorimeter geometry or technology (e.g. the barrel-

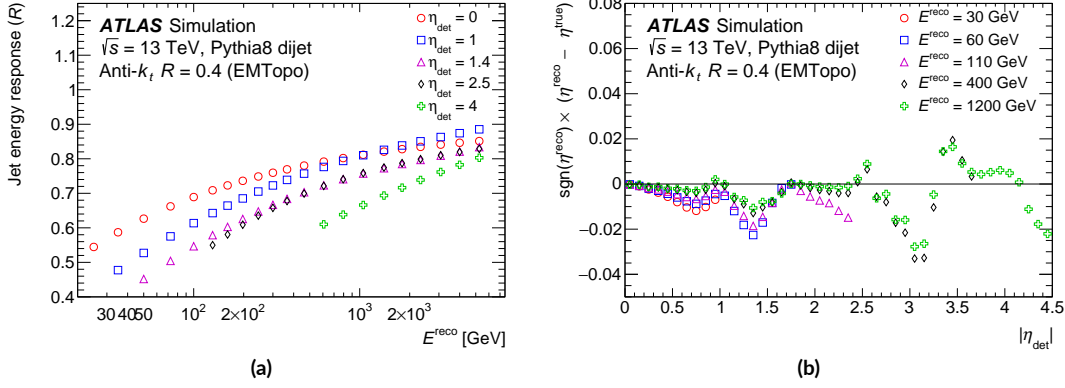


Figure 4.6: (a) Average energy response as a function of reconstructed jet p_T . (b) Signed difference between the reconstructed and truth jet η , denoted by η^{reco} and η^{true} [145].

endcap and the endcap–forward transition regions at $|\eta| \sim 1.4$ and $|\eta| \sim 3.1$, respectively), altering the reconstructed four-momentum of the jet. A second correction is then derived as the difference between the η distributions of reconstructed and truth jets, binned in E^{truth} and η . The aforementioned numerical inversion method is employed to parametrize such corrections in terms of E^{reco} . This later calibration affects the jet p_T and η distributions (corrections of about 1–4% in η), but not the full four-momentum.

GLOBAL SEQUENTIAL CALIBRATION

After the previous calibrations are applied, the response can vary from jet to jet depending on their composition and fluctuations of the jet development in the calorimeter. For instance, jets initiated by quarks often include high- p_T hadrons that reach deeper into the calorimeter, while gluon-initiated jets usually include more soft- p_T particles, since gluons typically emit more particles with lower energies. The global sequential calibration procedure (GSC in the following) is a series of multiplicative corrections applied to reduce the effect from such fluctuations and to improve the jet resolution without affecting the average jet energy response [149]. A sequence of correction factors is derived as a function of different shower shape variables [145]:

1. f_{Tile0} , the fraction of the jet energy measured in the first TileCal layer ($|\eta| < 1.7$);
2. f_{LAR3} , the fraction of jet energy measured in the third EMCal layer ($|\eta| < 3.5$);
3. n_{trk} , the number of tracks with $p_T > 1$ GeV associated to the jet ($|\eta| < 2.5$);

4. w_{trk} (called track width), the average p_T -weighted transverse distance in the η - ϕ plane between the jet axis and all tracks of $p_T > 1$ GeV associated to the jet ($|\eta| < 2.5$);
5. n_{segments} , the number of muon track segments associated with the jet ($|\eta| < 2.7$).

All corrections are derived as a function of the jet p_T , except for the n_{segments} correction, which is derived as a function of the jet energy. For each of the variables described above, an independent correction is derived in a similar way as proceeded for the energy scale calibration: jets are matched to simulated truth jets, a response function is derived in bins of p_T^{truth} and the numerical inversion method is used to extract the correction in terms of p_T^{reco} .

IN-SITU CALIBRATIONS

A final calibration is needed to account for differences between the jet response in the simulation and data. Such differences are caused by a non optimal modelation of detector response, hard scatter interaction and underlying event, jet formation, pile-up and interactions with the detector. The in-situ calibration measures these differences by balancing the jet p_T against another well calibrated reference object, like another jet, a photon, a Z boson or a multi-jet system. The response $\mathcal{R}_{in\ situ}$ is defined as the average of a Gaussian fit of ratio in p_T of the jet to the reference object, in bins of the reference object p_T [145]. The double ratio of the jet response in data and MC is then defined as

$$c = \frac{\mathcal{R}_{in\ situ}^{\text{data}}}{\mathcal{R}_{in\ situ}^{\text{MC}}}. \quad (4.5)$$

The double ratio defined by Eq. (4.5) is parametrized as a function of the jet p_T via numerical inversion. The final in-situ corrections are derived sequentially in three stages:

- First, the η inter-calibration analysis calibrates the forward jets ($0.8 \leq |\eta| < 4.5$) to the energy scale of those in the central region ($|\eta| < 0.8$) by using the p_T balance in di-jet events.
- Second, the jet recoil against the p_T of a Z boson or a photon is balanced by the Z +jet and γ +jet analyses, using the full hadronic recoil instead of a jet to reduce effects of pile-up and jet reconstruction threshold that would compromise low- p_T measurements [150].
- Finally, the multi-jet balance analysis uses a well calibrated low- p_T jets system to calibrate a single high- p_T jet [151]. This is done in order to access very high- p_T jets.

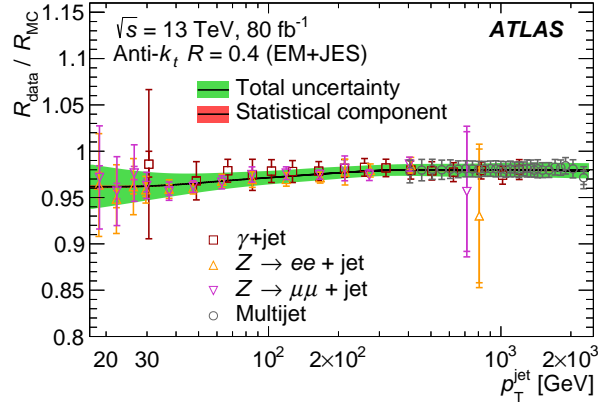


Figure 4.7: Ratio of the jet response in data to that in the nominal MC event generators as a function of jet p_T for Z +jet, γ +jet, and multi-jet in-situ calibrations. The inner horizontal ticks in the error bars give the size of the statistical uncertainty while the outer horizontal ticks indicate the total uncertainty (statistical and systematic uncertainties added in quadrature). The final correction and its statistical and total uncertainty bands are also shown [145].

Figure 4.7 shows the final in-situ corrections as a function of the jet p_T for anti- k_t , $R = 0.4$ reconstructed jets, decreasing from 4 to 2% as the jet p_T increases.

4.5.3 JET IDENTIFICATION

Reconstructed hard-scatter jets are discriminated against jets coming from pile-up interactions. A combination of track-based variables is used to construct the so-called *jet vertex tagger* (JVT) [152]. The JVT is developed such that the resulting jet efficiency is stable as a function of N_{PV} , and it is based on the use of two variables named corrJVF and R_{p_T} .

In Run 1, a cut on the *jet vertex fraction* (JVF) was employed to suppress pile-up jets. Such variable is defined as the fraction of tracks associated to a jet that come from the main PV. The performance of the JVF cut was noted to be dependent on N_{PV} . The variable corrJVF is a similar quantity to JVF, but corrected for such dependence. It is defined as

$$\text{corrJVF} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV}_0)}{\sum_l p_T^{\text{trk}_l}(\text{PV}_0) + \sum_{n \geq 1} \sum_l p_T^{\text{trk}_l}(\text{PV}_n)} / (k - n_{\text{trk}}^{\text{PU}}) \quad (4.6)$$

where $\sum_k p_T^{\text{trk}_k}(\text{PV}_0)$ is the scalar sum in p_T over tracks associated with the jet and originated from the hard-scatter vertex. The term $p_T^{\text{PU}} = \sum_{n \geq 1} \sum_l p_T^{\text{trk}_l}(\text{PV}_n)$ is the corresponding scalar p_T sum

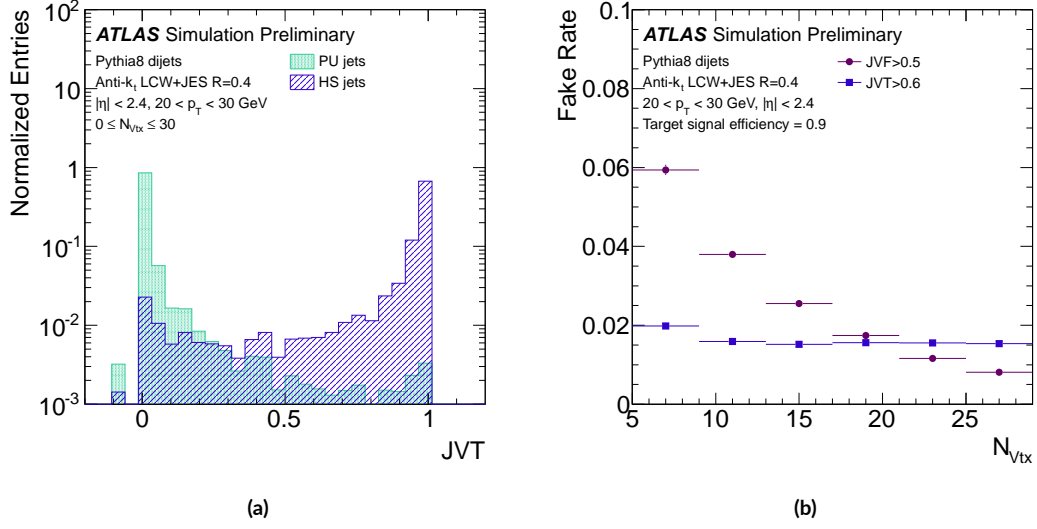


Figure 4.8: (a) JVT distribution for pile-up and hard-scatter jets with $20 < p_T < 30$ GeV. (b) Dependence on the number of PVs of the pile-up jet fake rate when imposing cuts on JVT (blue) and JVF (violet) such that the inclusive hard-scatter jet efficiency is 90% [152].

of tracks from pile-up interactions. The factor $(k - n_{\text{trk}}^{\text{PU}})$ is added to correct for the dependence on N_{PV} , where n^{PU} is the number of pile-up tracks per event and k is a scaling factor, taken as the slope of $\langle p_T^{\text{PU}} \rangle$ (resulting $k = 0.01$). The variable R_{p_T} is defined as the scalar sum in p_T over all tracks assigned to the jet and originated in the main interaction vertex, divided by the p_T of the jet after calibrations and pile-up corrections:

$$R_{p_T} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV}_0)}{p_T^{\text{jet}}}. \quad (4.7)$$

The variable corrJVF gives a measure of how likely the jet is a hard-scatter one. R_{p_T} , on the other hand, peaks at 0 and falls steeply for pile-up jets. Both variables are used to construct the 2-dimensional likelihood discriminant JVT. The discriminant is derived by using MC samples where the pile-up is emulated with minimum bias events overlaid. Figure 4.8 shows the resulting JVT values for hard-scatter and pile-up jets with $20 < p_T < 30$ GeV. Using JVT, efficiency values of 80%, 90% and 95% are obtained for pile-up fake rates of 0.4%, 1% and 3%, respectively. The dependence on N_{PV} present in the formerly used JVF method is almost negligible for JVT, as shown on the right side of Fig. 4.8.

4.5.4 JET CLEANING

Even after pile-up rejection, reconstructed jets must be still distinguished from misidentified jets of non-collision origin [153]. The main backgrounds of non-collision jets are beam-induced background (BIB), cosmic muons and calorimeter noise. The first two sources are usually referred to as non-collision background (NCB), whose contribution to the monojet final state is discussed in detail in Appendix C. A jet cleaning selection criteria is applied to the candidates sample to reject jets from such background processes (fake jets in the following) while keeping those initiated in pp collisions (good jets) with the highest efficiency.

A number of variables, based on signal pulse shapes in the calorimeters, energy ratios and track information, are built in order to discriminate good and fake jets. A detailed description of these variables can be found in Ref. [153]. Two levels of jet cleaning selection are then provided, called the BADLOOSE and BADTIGHT jet selections. A set of loose conditions is defined to identify jets originated either from sporadic noise bursts in the hadronic endcap calorimeter, from large coherent noise or isolated pathological cells in the EMCal, from hardware issues, or from NCB interactions. A jet is then tagged as a BADLOOSE jet if it meets any of such loose criteria. The BADTIGHT selection gives a much higher fake jets rejection, with a very small inefficiency for good jets. Jets are identified as BADTIGHT jets if they are already tagged as BADLOOSE or they pass a single additional requirement:

$$\frac{f_{\text{ch}}}{f_{\text{max}}} < 0.1 \text{ and } |\eta| < 2.4, \quad (4.8)$$

where

- $f_{\text{ch}} = \sum p_{\text{T}}^{\text{track}} / p_{\text{T}}^{\text{jet}}$ is the fraction of the momentum of the jet associated to tracks reconstructed in the ID;
- f_{max} is the maximum fraction of jet energy deposited in any layer of the calorimeter.

Such ratio in Eq. (4.8) gives a very high efficiency at discriminating fake jets (with typical f_{ch} values close to 0 and f_{max} close to 1) and good jets, usually with $f_{\text{ch}} > 0$ and $f_{\text{max}} < 1$. This particular criteria was initially implemented in ATLAS by the monojet analysis, since it needs a high restrictive jet selection due to the similar topologies between NCB signatures and the monojet final state. For analysis use, the LOOSE and TIGHT selections for jet cleaning are provided, such that a jet is identified as LOOSE if is not identified as a BADLOOSE jet and, correspondingly, a TIGHT jet is that which is not tagged as BADTIGHT. In Figure 4.9 some kinematic distributions are shown for a fake jets enriched sample. The TIGHT selection provides a much higher level of fake jets suppression, especially

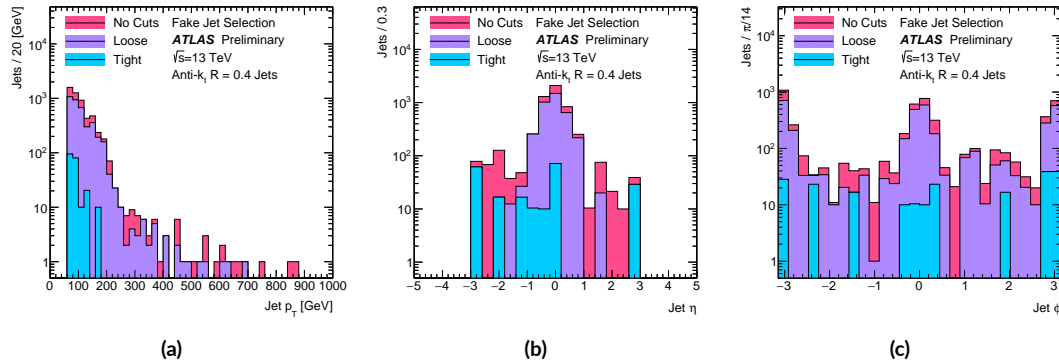


Figure 4.9: Jet p_T (a), η (b) and ϕ (c) distributions for fake jets enriched samples before and after the Loose and Tight jet cleaning selections [153].

at high p_T . The peaks at $\phi = 0$ and π , characteristic BIB signatures [154], are rejected by the TIGHT selection while they are not with the LOOSE criteria.

4.5.5 b -TAGGING

The identification of jets initiated by b hadrons (known as b -jets) is an important tool for many analyses in ATLAS, since allows to select events enriched in certain processes. For instance, top-quarks decay into b -quarks plus a W boson in most of the cases, hence having a way to select b -jets allows to construct a region enriched in top production processes.

Different algorithms have been developed in ATLAS to identify b -jets. Some of them by exploiting the long lifetime of the b -hadrons via looking for displaced secondary vertices, others are based on the impact parameters information from tracks associated to the jet, or try to exploit the topological structure of the b -hadron decay chain inside the jet [155]. The most discriminating observables from these algorithms are combined to train a Boosted Decisions Tree (BDT) algorithm [156], using the ROOT Toolkit for Multivariate Data Analysis (TMVA) tool [157]. This results in the multivariate MV2 algorithm [158].

The training of the MV2 classifier is carried out with $t\bar{t}$ MC samples with b -jets considered as signal, while c -hadron initiated jets (c -jets) and light-flavour jets are considered as background. Depending on the needs of the different analyses, several MV2 variants are provided, where the fraction of the c -jets background is varied in the training. In the analysis of this thesis, the MV2c10 variant is

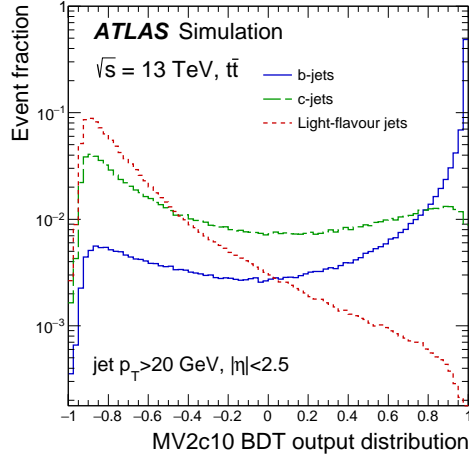


Figure 4.10: MV2c10 output for b - (solid line), c - (dashed line) and light-flavour jets (dotted line) in MC $t\bar{t}$ events [159].

used. The c -jet fraction of the training for MV2c10 is set at 7% of the total background composition, being the 93% left light-flavour jets. Figure 4.10 shows the output of the MV2c10 BDT for the signal and background components. Operation points are defined by applying a single cut on the MV2 output, such that they provide specific efficiency values on b -jets identification. In the analysis of this thesis a BDT cut value of 0.9349 is chosen, resulting into a b -tagging efficiency of 60%.

4.6 TAUS

Tau-leptons have a mean life of $\sim 290 \times 10^{15}$ s, which translates into a decay length of $\sim 87 \mu\text{m}$ [9], much smaller than the detector resolution. For this reason, no secondary vertex can be reconstructed, and therefore they can only be identified by their reconstructed decay products. Taus can decay either leptonically, with $\tau \rightarrow \ell \nu_\ell \nu_\tau$ ($\ell = e, \mu$), or hadronically, like $\tau \rightarrow \text{hadrons} + \nu_\tau$. In the analysis of this thesis, only hadronically-decaying taus (τ_{had} in the following) are considered. The visible decay products of τ_{had} are referred to as $\tau_{\text{had-vis}}$.

4.6.1 TAU RECONSTRUCTION

The seed for the $\tau_{\text{had-vis}}$ reconstruction process is set by using energy deposits in the calorimeter [160]. Such deposits must have been reconstructed as jets with the anti- k_t algorithm with $R = 0.4$ [141].

Jets with $p_T > 10$ GeV and $|\eta| < 2.5$ are selected for the seed.

In order to reduce pile-up effects, the tau-lepton production vertex (TV) is identified among the set of PVs previously reconstructed (the reconstruction of the primary vertex is described in Sec. 4.1). Using good-quality tracks with $p_T > 1$ GeV within $\Delta R < 0.2$ around the seed, the TV is defined as the PV to which the largest fraction of the p_T sum over these tracks is matched to.

The momentum of the $\tau_{\text{had-vis}}$ candidate is determined by the sum of the four-momenta of all topo-clusters in the core region (i.e. $\Delta R < 0.2$ around the barycentre), which are calibrated at the local hadronic scale [161] and recalculated in the TV coordinate-system. The mass of the $\tau_{\text{had-vis}}$ is defined as zero. Then, tracks are assigned to the candidate if they meet the following criteria: be in the $\Delta R < 0.2$ region around the $\tau_{\text{had-vis}}$ direction, have $p_T > 1$ GeV, at least two Pixel hits and at least seven hits in the Pixel + SCT detectors; in addition, they are required to have $|d_0| < 1$ mm and $|z_0 \sin \theta| < 1.5$ mm.

Finally, the energy of the tau candidate is re-calibrated to the tau-lepton energy scale with a response curve correction method using $Z \rightarrow \tau\tau$, $W\tau\nu$ and $Z' \rightarrow \tau\tau$ MC samples [162].

4.6.2 TAU IDENTIFICATION

Reconstructed $\tau_{\text{had-vis}}$ are discriminated against the jet background, which is composed of jets in which a quark or a gluon is the dominated particle, called quark-like and gluon-like jets, respectively. In general, gluon-like jets are easier to reject than quark-like jets since the later are typically more collimated and have fewer tracks.

A number of discriminating variables is constructed by using information from tracks within the so-called isolation region, i.e. tracks in the region $0.2 < \Delta R < 0.4$ around the $\tau_{\text{had-vis}}$ direction, and shower shape information provided by the associated topo-clusters in both core and isolation regions. The full list of variables employed is described in detail in Ref. [163]. Then, separate BDT algorithms [156] are trained for one-track and one-track $\tau_{\text{had-vis}}$ decays using $Z \rightarrow \tau\tau$ and di-jets simulated samples.

Three working points, corresponding to different tau identification efficiencies, are provided:

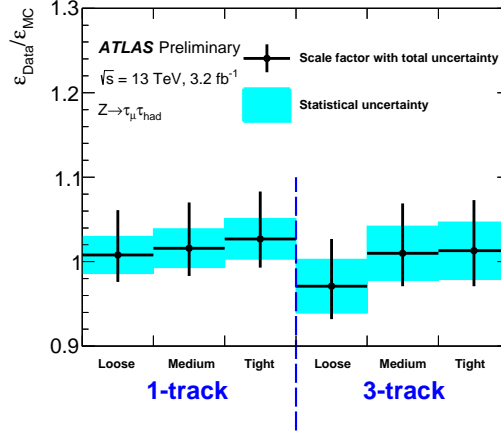


Figure 4.11: Ratio of identification efficiencies as obtained for data ($\varepsilon_{\text{Data}}$) over MC (ε_{MC}), for one-track and three-track $\tau_{\text{had-vis}}$ candidates with $p_T > 20$ GeV [162].

LOOSE, MEDIUM and TIGHT. For each WP, requirements on the BDT score are determined as a function of the $\tau_{\text{had-vis}}$ p_T , in order to ensure constant efficiency values for the reconstruction and the identification. Figure 4.11 shows the ratio of the $\tau_{\text{had-vis}}$ identification efficiencies measured for data over simulations using $Z \rightarrow \tau\tau$ events for the different identification WPs. This number is called *scale-factor* and is used in the analysis as a weight for the MC events in order to bring the tau identification efficiency in simulations to the level observed in the data.

In addition, a further discrimination between one-track $\tau_{\text{had-vis}}$ and electrons is carried out. This is done by using the Likelihood function described in Section 4.2. Reconstructed $\tau_{\text{had-vis}}$ candidates are matched to electrons with $p_T > 5$ GeV within a distance of $\Delta R < 0.4$, and rejected if the associated electron passes the VERYLOOSE electron identification criteria. This cut is tuned to return a 95% efficiency for hadronically decaying taus [162].

4.7 MISSING TRANSVERSE ENERGY

Momentum conservation requires the total sum of the transverse momenta of all the collision products to be at least close to zero (detector imperfections can lead to small imbalances in the total transverse momentum measurement). When this does not happen, is because an undetected particle has been produced in the collision event. In the Standard Model, the only particles that would escape without leaving any signal in the detector are neutrinos. Moreover, many Beyond the Standard Model

theories predict the existence of different types of particle that would escape undetected as well. Therefore, the precise measurement of this imbalance is very important for physics analysis, but more especially for new physics searches. Such imbalance is measured in the transverse plane and called missing transverse momentum, $\mathbf{p}_T^{\text{miss}}$, and its magnitude is the so-called missing transverse energy, E_T^{miss} . The E_T^{miss} is the most important component of the monojet final state.

4.7.1 E_T^{MISS} RECONSTRUCTION

Since in ATLAS the reconstruction of each kind of particle is carried out independently, it is possible that some ambiguities can emerge between objects, e.g. the same calorimeter energy deposit assigned to reconstruct an electron might be used as well to reconstruct a jet. For this reason, it is important to resolve explicitly any potential signal ambiguity prior addressing the E_T^{miss} calculation, in order to avoid double-counting effects. Regarding the analysis of this thesis, the solution adopted for this issue is called *overlap removal*, and is discussed in Chapter 6.

Once all the physics objects are reconstructed and properly identified and selected, $\mathbf{p}_T^{\text{miss}}$ is calculated as the negative vectorial sum of momenta of all the objects in the event:

$$\mathbf{p}_T^{\text{miss}} = - \underbrace{\sum_{\text{electrons}} \mathbf{p}_T^e - \sum_{\text{photons}} \mathbf{p}_T^\gamma - \sum_{\text{muons}} \mathbf{p}_T^\mu - \sum_{\text{taus}} \mathbf{p}_T^{\tau_{\text{had}}} - \sum_{\text{jets}} \mathbf{p}_T^{\text{jet}}}_{\text{hard term}} - \overbrace{\sum_{\text{unused tracks}} \mathbf{p}_T^{\text{track}}}^{\text{soft term}}. \quad (4.9)$$

Unused tracks in Eq. (4.9) refers to tracks associated with the hard-scatter vertex but not with any hard object [164]. Such contributions are taken into account via the so-called *soft term*. On the contrary, terms relative to the contributions from reconstructed objects are grouped and denoted as the *hard term*. The particular choice of using only tracks for the soft term helps to suppress pile-up contributions, improving the E_T^{miss} resolution as well. Only high-quality tracks passing certain signal-overlap conditions are included in the soft term calculation [164].

The performance of the E_T^{miss} reconstruction and resolution is studied by comparing the distributions in data and MC simulations for some well known SM processes, such as $Z \rightarrow \ell\ell$ and $W \rightarrow \ell\nu$. Figure 4.12 shows the reconstructed E_T^{miss} distributions for $Z \rightarrow \mu\mu$ and $W \rightarrow e\nu$, where good level of agreement between simulations and data is observed. The $Z \rightarrow \mu\mu$ process, under a certain event selec-

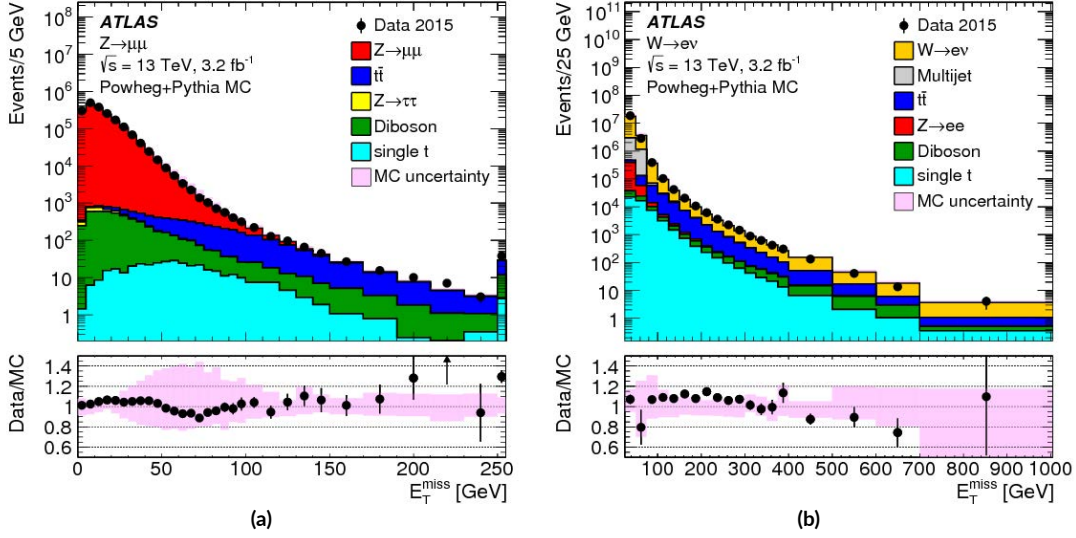


Figure 4.12: Distributions of E_T^{miss} for inclusive samples of $Z \rightarrow \mu\mu$ (a) and $W \rightarrow e\nu$ (b) events extracted from data and compared to MC simulations including all relevant backgrounds. The shaded areas indicate the total uncertainty for MC simulations, including the overall statistical uncertainty combined with systematic uncertainties. The respective ratios between data and MC simulations are shown below the distributions, with the shaded areas showing the total uncertainties for MC simulations [164].

tion around the Z boson peak, provides a very high purity sample where no genuine E_T^{miss} is expected. The $W \rightarrow e\nu$ process, on the other hand, provide a source of genuine E_T^{miss} due to the neutrino.

The E_T^{miss} resolution is studied comparing the measurements in data with MC simulations using $Z \rightarrow \mu\mu$ and $W \rightarrow e\nu$ samples. The resolution is measured in bins of $\sum E_T$, a variable used to measure activity of the event and defined as the scalar sum of all transverse momenta from the objects contributing to the E_T^{miss} ,

$$\sum E_T = \sum_{\text{electrons}} p_T^e + \sum_{\text{photons}} p_T^\gamma + \sum_{\text{muons}} p_T^\mu + \sum_{\text{taus}} p_T^{\tau_{\text{had}}} + \sum_{\text{jets}} p_T^{\text{jet}} + \sum_{\text{unused tracks}} p_T^{\text{track}}. \quad (4.10)$$

In each bin, the E_T^{miss} resolution is calculated as the root mean squared (RMS) of the combined X and Y E_T^{miss} components. The measured resolution as a function of $\sum E_T$ is shown in Figure 4.13, where good agreement is noted between data and predictions, and the measured resolution scales quickly from 4 to 20 GeV within the range $50 < \sum E_T < 250$ GeV, and then slowly up to 25 GeV for $\sum E_T = 700$ GeV.

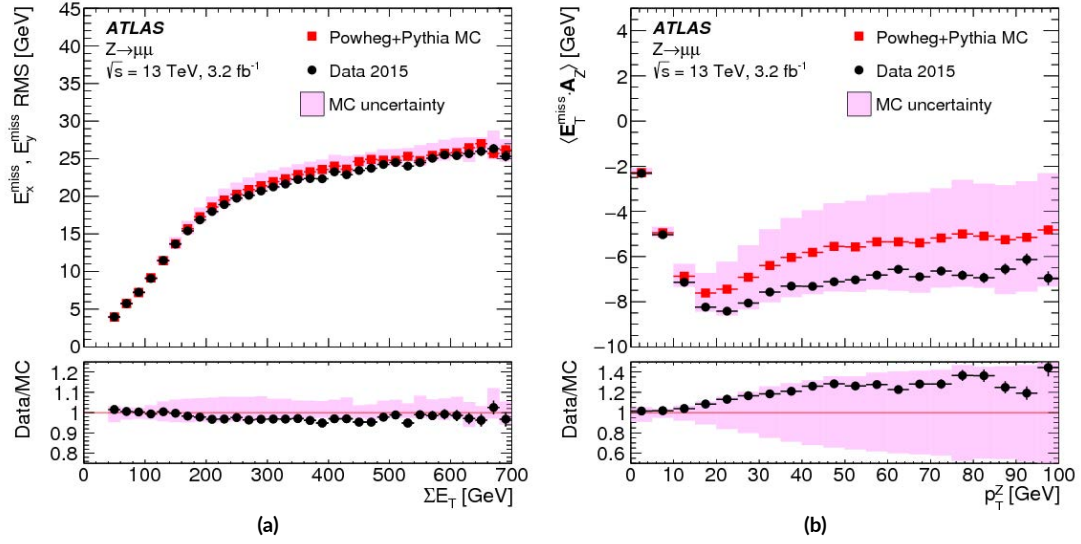


Figure 4.13: (a) the RMS width of the $E_{x(y)}^{\text{miss}}$ distributions in bins of ΣE_T in an inclusive sample of $Z \rightarrow \mu\mu$ events. (b) The average projection of $\mathbf{p}_T^{\text{miss}}$ onto the direction \mathbf{A}_Z of the Z boson's transverse momentum vector \mathbf{p}_T^Z as a function of $p_T^Z = |\mathbf{p}_T^Z|$ in inclusive $Z \rightarrow \mu\mu$ events. Predictions from MC simulations are overlaid on the data points, and the ratios are shown below the respective plot. The shaded bands indicate the combined statistical and systematic uncertainties of the resolution measurements. [164].

The E_T^{miss} scale is determined in $Z \rightarrow \mu\mu$ events, and is defined as the average of the E_T^{miss} projection in the direction of the Z boson p_T , $\langle \mathbf{A}_Z \cdot \mathbf{p}_T^{\text{miss}} \rangle$, where \mathbf{A}_Z is the unity vector in the direction of the Z boson,

$$\mathbf{A}_Z = \frac{\mathbf{p}_T^{\mu^+} + \mathbf{p}_T^{\mu^-}}{|\mathbf{p}_T^{\mu^+} + \mathbf{p}_T^{\mu^-}|}. \quad (4.11)$$

Figure 4.13 shows the E_T^{miss} scale as a function of the p_T of the Z boson. The predictions exhibit overall agreement with the data within uncertainties, and the estimated E_T^{miss} scale starts from -2 in the ideal case of $p_T(Z) = 0$, when perfect balance between the hard objects and the hadronic recoil would be reached, drops down to -8 as $p_T(Z)$ grows up to ~ 25 GeV, and then starts raising smoothly as the momentum of the Z increases, and so do the associated uncertainties.

5

The Statistical Model

The work presented in this thesis relies in a series of statistical tools, whose theoretical bases are introduced briefly in this chapter together with the definitions of the different techniques that are employed to analyze the data.

5.1 INTRODUCTION

A system is said to be random if the outcome of a measurement over that system varies unpredictably when repeating the experiment. Such randomness, which can be caused by the measuring device or by the intrinsic unpredictable nature of the system (e.g. a system governed by quantum mechanics), is quantified by the concept of probability [165]. There are two main “schools” on how to define the probability: the Bayesian approach, widely used in astrophysics, which sees the probability as a degree of belief in a certain event based on a prior knowledge about the event; and the frequentist (also known as classical) one, most commonly adopted in particle physics, which interprets the probability as the relative frequency of an event in n measurements when $n \rightarrow \infty$.

In a given experiment, the probability to measure the variable X within the infinitesimal interval

$[x, x + dx]$ is given by the probability density function (PDF), $f(X)$, as

$$\mathcal{P}(X \in [x, x + dx]) = f(x)dx. \quad (5.1)$$

The PDF is defined such that is normalized to the unity. If X is a random variable, the expectation value $E[X]$ (or called simply mean) represents the value where X is most likely to be measured, and the variance $V[X]$ gives a measure of how X spreads around its mean. Depending on the characteristics of the variable and the type of measurement, the associated PDF can be very different. For instance, if X is a discrete variable and the number of repetitions of the experiment, n , is large, the PDF will follow a Poisson distribution,

$$\mathcal{P}(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad (5.2)$$

where λ is a real positive number that is equal to the expected value of X and its variance, $E[X] = V[X] = \lambda$. Poisson distributions are typically used in decay measurements of radioactive materials (usually introducing $\lambda = t/\tau$, where t is the time and τ is the mean lifetime of the material), and in particle collider experiments, where the mean number of events of a certain process is given by the product of its cross-section, the efficiency of the detector and the luminosity of the collider, $\lambda = \sigma\mathcal{L}$.

On the other hand, random variables with continuous real values will exhibit PDFs in the form of a Gaussian distribution,

$$G(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (5.3)$$

where $\mu = E[X]$ is the mean value of the distribution and σ is the standard deviation, given by the square root of $V[X]$. If $\mu = 0$ and $\sigma = 1$, the PDF given by Eq. (5.3) is called a normal distribution. Gaussian distributions play an important role in physics experiments thanks to the Central Limit Theorem [165], which holds that the sum of independent random variables becomes a Gaussian random variable, regardless of the form of the individual PDFs. This justifies treating systematic uncertainties as Gaussian variables, as the total error is the sum of a large number of small contributions.

5.2 STATISTICAL TEST

A data analysis is usually formulated in terms of a hypothesis test. The idea is to give a quantitative measure of the level of agreement between the sample of data collected by the experiment and a certain prediction, which will be called *null hypothesis* (H_0), against an *alternative hypothesis* (H_1). In new physics searches, as the analysis of this thesis, H_0 usually represents the case when the data can be described only by SM processes (background-only), while H_1 corresponds to the scenario in which the background plus signal prediction matches with the observations. Quantitatively, this can be sim-

plified by introducing a *signal strength* parameter (μ), which scales the predicted signal cross-section such that $\mu = 0$ is equivalent to H_0 and consequently $\mu = 1$ corresponds to H_1^* .

Mathematically, this is addressed by constructing a new variable as a function of μ called *test-statistic*, q_μ , for which each of the hypotheses will imply a different PDF: $f(q_\mu|H_0)$ and $f(q_\mu|H_1)$. The point is to define q_μ with the same or less degrees of freedom than the measured variable, to reduce the amount of data without losing the ability to discriminate between hypotheses.

5.2.1 STATISTICAL SIGNIFICANCE

Once a test-statistic is defined and evaluated with the data, the obtained value q_{obs} allows to calculate the so-called *p*-value. This quantity is given by the upper-tail area under the curve $f(q_\mu|H_0)$ from q_{obs} ,

$$p = \int_{q_{\text{obs}}}^{\infty} f(q_\mu|H_0) d\mu, \quad (5.4)$$

and represents the probability, assuming the hypothesis H_0 is true, of obtaining a result as compatible or less with the H_0 than actually observed. Figure 5.1 illustrates the concept of the *p*-values as computed by using Gaussian distributions for simplicity. In the figure, the obtained values p_0 and p_1 are, respectively, 0.09366 and 0.02385, meaning that there is a probability of 9.3% (2.4%) of measuring a result incompatible with H_0 (H_1) by accident.

In order to decide whether to accept or reject H_0 , a certain threshold q_{cut} is set prior the measurement (so the experiment is not biased). This threshold is chosen such as the probability of measuring q_μ beyond q_{cut} , under the assumption of H_0 , is some value α , known as the *significance level* of the test, $\alpha = \int_{q_{\text{cut}}}^{\infty} f(q_\mu|H_0)$. If the obtained *p*-value is lower than α (or, equivalently, $|q_{\text{obs}}| > |q_{\text{cut}}|$), the hypothesis H_0 is accepted, and otherwise is rejected.

Usually, the *p*-values are translated into an equivalent significance, Z , defined such that a Gaussian distributed variable found Z standard deviations above its mean has an upper-tail probability equal to p [166],

$$z = \Phi^{-1}(1 - p), \quad (5.5)$$

where Φ^{-1} is the inverse of the cumulative normal distribution. Typically, $Z = 5$ is required to claim a ‘discovery’ (corresponding to $p = 2.9 \cdot 10^{-7}$), and ‘evidence’ of new physics is reported with $Z = 3$ ($p = 1.3 \cdot 10^{-3}$). For the purpose of excluding a signal hypothesis, a threshold of $\alpha = 0.05$ is often

*It is important to remark that this is the approach that is typically adopted in searches targeting a discovery (e.g. the Higgs boson discovery in 2012 [7, 8]), while in analyses where the goal is to set exclusion limits into signal models the roles of H_0 and H_1 are usually inverted.

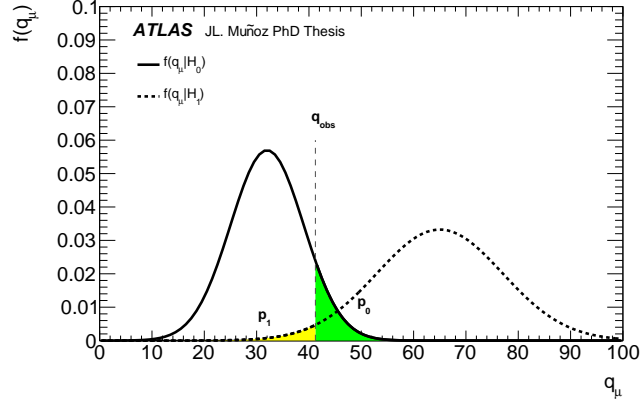


Figure 5.1: Example of PDFs for a test-statistic taken as a Gaussian variable under the assumption of H_0 and H_1 . The green (yellow) area denotes the p -value obtained for H_0 (H_1) for a given q_{obs} .

used, i.e. models are excluded at 95% confidence level (CL).

5.2.2 THE CL_s METHOD

In analyses where the number of events registered does not exceed significantly the background predictions, the goal is no longer to discover a new particle but rather to put limits to the validity of some signal models. As mentioned before, in these cases the signal plus background hypothesis ($s + b$) is set as H_0 , being the alternative hypothesis the one considering background only (b). For this kind of analyses the use of the p -values can lead to nonphysical conclusions in cases where the sensitivity of the experiment to the signal model is low, since the associated PDFs to the test-statistic would look very similar under the two hypotheses and the effect of a statistical fluctuation could lead to the exclusion of both $s + b$ and b hypotheses. In order to avoid this issue, the CL_s method [167] is introduced. This method relies on the use of the confidence levels CLs determined for the two hypotheses,

$$\begin{aligned} CL_{s+b} &= p_{s+b} \\ CL_b &= 1 - p_b \end{aligned}, \quad (5.6)$$

where p_{s+b} and p_b are the p -values computed for the $s + b$ and b hypotheses, respectively. The quantity CL_s is then defined as the ratio of the CLs defined by Eq. (5.6):

$$CL_s = \frac{CL_{s+b}}{CL_b}. \quad (5.7)$$

Finally, the CL_s , defined by Eq. (5.7) is used to determine whether a model is excluded or not. In the analysis of this thesis, models for which the computed CL_s is lower than 0.05 are excluded. Therefore the exclusion limits will be reported at 95% CL.

5.3 THE MAXIMUM LIKELIHOOD METHOD

In ATLAS, a procedure widely used in search analyses is based on a frequentist significance test using a profile likelihood ratio as a test-statistic [166]. In addition to parameters of interest such as the signal strength, the signal and background models will contain in general *nuisance parameters* (NPs). The Maximum Likelihood (ML) method [168] is a technique for estimating such parameters from a given sample of data.

5.3.1 THE LIKELIHOOD FUNCTION

The likelihood function is constructed as follows. Consider an experiment that measures a given variable X collecting events in a histogram of N bins. The expected number of events in the i -th bin can be expressed as

$$E[n_i] = \mu s_i + b_i, \quad (5.8)$$

where s_i and b_i denote the expected number of signal and background events in the bin i , respectively, and are given by the expectation value within the boundaries of the bin of their corresponding PDFs, $f_s(X; \theta)$ and $f_b(X; \theta)$, whose shape is determined by a set of NPs denoted by θ .

Assuming that the data follows a Poisson distribution (see Eq. (5.2)), the probability of measuring n^{obs} events in the data is then given by the likelihood function $L(\mu, \theta)$, defined as the product of the Poisson probabilities in each bin:

$$L(\mu, \theta) = \prod_i^N \mathcal{P}(n_i^{\text{obs}} | \mu s_i + b_i) \quad (5.9)$$

In many cases, a set of parameters κ is introduced to account for the normalization of the number of background events. These factors are usually estimated from a fit to the data in R control regions enriched in background events. In these cases, the likelihood function is the total product of the

likelihoods in each region r ,

$$L(\mu, \boldsymbol{\kappa}, \boldsymbol{\theta}) = \prod_r^R \prod_i^N \mathcal{P} \left(n_{ir}^{\text{obs}} | \mu s_{ir} + b_{ir}(\boldsymbol{\kappa}) \right), \quad (5.10)$$

In addition, some complementary measurements are often performed in order to help to constrain the nuisance parameters. The variations are parametrized as normal distributions $G(\theta_p | 0, 1)$, where θ_p represents the relative size of the NP p in units of RMS. A constraining function is then constructed as the product of the Gaussian distributions of each parameter: $f_{\text{constr}}(\boldsymbol{\theta}) = \prod_p G(\theta_p | 0, 1)$. The complete likelihood function is then

$$L(\mu, \boldsymbol{\kappa}, \boldsymbol{\theta}) = \prod_r^R \prod_i^N \mathcal{P} \left(n_{ir}^{\text{obs}} | \mu s_{ir} + b_{ir}(\boldsymbol{\kappa}) \right) f_{\text{constr}}(\boldsymbol{\theta}), \quad (5.11)$$

Finally, the likelihood function given by Eq. (5.11) is maximized to extract the optimal values for the parameters[†]. This is actually done by taking the negative logarithm of the likelihood function (NLL), and finding the minimum of this function instead.

5.3.2 THE PROFILE LIKELIHOOD RATIO

In order to perform the hypothesis test, the profile likelihood ratio is defined as

$$\lambda(\mu) = \frac{L(\mu, \hat{\boldsymbol{\theta}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \quad (5.12)$$

where the dependence of the likelihood function on $\boldsymbol{\kappa}$ given by Eq. (5.11) is omitted for simplicity. In Eq. (5.12), the numerator is what is called the *profile likelihood*, where $\hat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$ that maximizes L for a given μ , i.e. is the conditional ML estimator of $\boldsymbol{\theta}$, and therefore a function of μ . The denominator is the maximized (unconditional) likelihood function, i.e. $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the values of μ and $\boldsymbol{\theta}$ that maximize L . Based on Eq. (5.12), the test-statistic q_μ is then defined as

$$q_\mu = -2 \ln \lambda(\mu) \quad (5.13)$$

[†]Note that the dependence of s_i and b_i on the nuisance parameters $\boldsymbol{\theta}$ is omitted in Eqs. (5.9), (5.10) and (5.11) for simplicity.

From the definition of $\lambda(\mu)$, it can be inferred that values of λ close to 1 imply good agreement between the data and the hypothesized value of μ . Therefore, higher values of q_μ will correspond to increasing incompatibility between the data and μ .

From q_μ the p -value be computed for a given μ' by using Eq. (5.4). However, the estimation associated PDF, $f(q_\mu|\mu')$ is not straightforward. In general, this is solved by generating a large number of random pseudo-experiments (MC toys), but these methods are computationally heavy, so what is done in some cases is to use the asymptotic approximation [166], which holds that, assuming the data are distributed according to μ' , the function given by Eq. (5.13) can be approximated by

$$q_\mu = -2 \ln \lambda(\mu) \simeq \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}), \quad (5.14)$$

where $\hat{\mu}$ follows a Gaussian distribution with mean μ' and standard deviation σ , and N denotes the size of the data sample. In the large sample limit the asymptotic approximation becomes exact and the term $\mathcal{O}(1/\sqrt{N})$ in Eq. (5.14) can be neglected. In this case, the associated PDF takes the form of a χ^2 distribution and can be then expressed as

$$f(q_\mu|\mu') = \frac{1}{2\sqrt{q_\mu}} \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{1}{2}\left(\sqrt{q_\mu} + \frac{\mu - \mu'}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{\mu - \mu'}{\sigma}\right)^2} \right]. \quad (5.15)$$

5.4 FIT CONFIGURATIONS

In the analysis of this thesis, a number of requirements is applied to the data in order to select a region that gives an optimal sensitivity to the signal models that are considered. This region is called *signal region*. Together with this region, a number of control regions is also defined similarly in order to constrain the expected number of background events. The definitions of the signal and control regions are detailed in Chapter 6. Three different configurations of the fit are used for different purposes, detailed below, and are performed using the HistFitter software [169].

- **Background-only fit.** Only the control regions are considered to constrain the parameters of the fit. Potential contributions from signal processes are neglected in all regions. This configuration is used to estimate the background contributions to the signal region.
- **Model-independent fit.** Both signal and control regions are included in the fit. Any signal

contribution in the control regions is neglected. This configuration is used to obtain 95% CL limits on the visible cross-section of any potential signal process.

- **Model-dependent fit.** Again, signal and control regions are used and signal contributions in the control regions are neglected. Signal predictions for specific models are injected in the signal region. This configuration is used to exclude signal models when the computed CL_s is lower than 0.05.

6

The Monojet Analysis

This chapter describes the monojet analysis of the full Run 2 dataset of pp collisions collected by ATLAS at $\sqrt{s} = 13$ TeV. The monojet signature is known for being a very sensitive channel for a wide variety of BSM searches, especially for DM production models, but also for extra dimensions or compressed SUSY scenarios, among many others. Such processes will predominantly leave signatures in the detector similar to that of the ATLAS event display in Figure 6.1: one energetic jet recoiling against a large amount of E_T^{miss} . The monojet analysis strategy is designed to maximize the sensitivity to new phenomena, providing a precise estimation of SM backgrounds. Compared to previous versions of this analysis, a superior precision is achieved by an extended E_T^{miss} range, a revision of the lepton definitions and an enhanced treatment of the W/Z +jets SM predictions.

The chapter is organized as follows. The data sample used in the analysis is described in Section 6.1. The object definitions and the event selection applied are detailed in Section 6.2. The background estimation procedure is discussed in Section 6.3. Finally, some distributions are shown in Section 6.4.

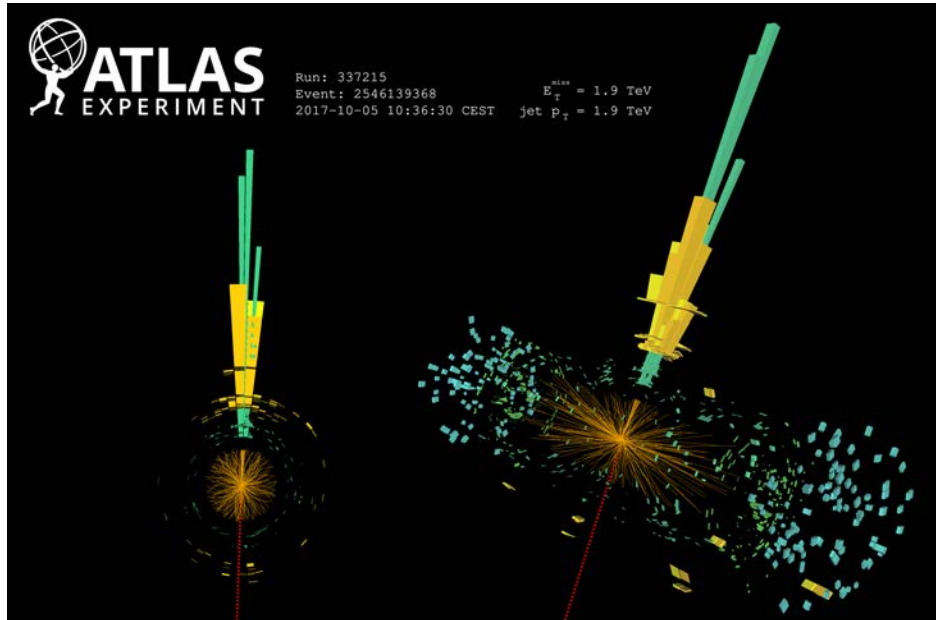


Figure 6.1: ATLAS event display of a monojet event recorded in 2017 with $\sqrt{s} = 13$ TeV. One jet is measured with $p_T = 1.9$ TeV, indicated by the green and yellow bars corresponding to the energy deposition in the calorimeters, and a missing transverse momentum of $E_T^{\text{miss}} = 1.9$ TeV is measured in the opposite φ -direction (red dashed line).

6.1 DATA SAMPLE

The dataset used in this analysis was recorded by ATLAS between 2015 and 2018 at a center-of-mass energy of 13 TeV, corresponding to a total integrated luminosity of 139 fb^{-1} . The events are accumulated after applying data-quality requirements such as tracking detectors, calorimeters, muon chambers and magnets were fully functional during the data-taking operations, i.e. satisfying the GRL criteria. Both real and simulated events are reconstructed and analyzed with the same analysis code chain.

6.1.1 TRIGGER

Different trigger strategies are implemented to analyze data from the 2015, 2016, 2017 and 2018 data-taking periods, in order to reflect the different trigger menus. Also, different triggers are employed depending on different event selection choices.

The triggers used in most of the selections are calorimeter-based E_T^{miss} triggers, for which muons

Data period		Trigger
2015	All runs	HLT_xe70_mht
	$296939 \leq \text{run} \leq 302872$	HLT_xe90_mht_L1XE50
2016	$302919 \leq \text{run} \leq 303892$	HLT_xe100_mht_L1XE50
	$303943 \leq \text{run} \leq 311481$	HLT_xe110_mht_L1XE50
2017	$325713 \leq \text{run} \leq 331975$	HLT_xe110_pufit_L1XE55
	$332303 \leq \text{run} \leq 341649$	HLT_xe110_pufit_L1XE50
2018	$348197 \leq \text{run} \leq 350066$	HLT_xe110_pufit_xe70_L1XE50
	$350067 \leq \text{run} \leq 363400$	HLT_xe110_pufit_xe65_L1XE50

Table 6.1: Lowest unprescaled E_T^{miss} triggers in the 2015–2018 data periods. The names indicate the calorimeter energy thresholds in GeV at HLT level, followed by the E_T^{miss} calculation algorithm used and by the initial L1 trigger threshold.

are considered as invisible particles in the evaluation of the E_T^{miss} at trigger level. In these regions, the lowest unprescaled E_T^{miss} triggers are used. As Table 6.1 shows, in these regions the lowest unprescaled E_T^{miss} trigger thresholds change during the data-taking periods due to increasing instantaneous luminosity conditions. Single-electron triggers are used in selections that do not include muons but electrons are selected. In such regions, the lowest unprescaled single-electron triggers are used combined in an OR logical (see Table 6.2).

6.2 EVENT SELECTION

In this section, the list of requirements that are used in the definition of the **Signal Region (SR)** and the **Control Regions (CRs)** are detailed.

6.2.1 OBJECT DEFINITIONS AND OVERLAP REMOVAL

Once reconstructed as discussed in Chapter 4, electrons, muons, photons, taus and jets are required a few more conditions before being used in the analysis. In particular, two categories of increasing level of requirements are defined for electrons and muons: *baseline* objects, which are used in the overlap removal and for vetoes; and *signal* objects, used for actual event selection.

Data period	Trigger
2015	HLT_e24_lhmedium_L1EM20VH or HLT_e60_lhmedium or HLT_e120_lhloose
2016,2017 and 2018	HLT_e26_lhtight_nod0_ivarloose or HLT_e60_lhmedium_nod0 or HLT_e140_lhloose_nod0 or HLT_e300_etcut

Table 6.2: Lowest unrescaled single-electron triggers in 2015–2018 data, combined in a logical OR. The names indicate the offline electron transverse energy threshold in GeV at the HLT level, followed by an electron ID and (possible) isolation requirements applied, and by the initial L1 threshold.

- **Baseline electrons** are electrons with $p_T > 7$ GeV and $|\eta| < 2.47$. The baseline electron must satisfy the **LOOSE** electron identification criteria. A requirement on the longitudinal impact parameter, $|z_0| \sin \theta < 0.5$ mm, is also applied. **Signal electrons** are baseline electrons that pass the **TIGHT** electron likelihood criteria. Their p_T threshold is raised up to 30 GeV and a cut on the transverse longitudinal parameter of the electron track, $d_0/\sigma_0 < 5$, is applied.
- **Baseline muons** are muons selected with p_T above 7 GeV and $|\eta| < 2.5$. They also have to meet the **MEDIUM** identification criteria. **Signal muons** are a subset of baseline muons with $p_T > 10$ GeV and passing additional cuts on the track impact parameters: $|z_0| \sin \theta < 0.5$ mm and $d_0/\sigma_0 < 3$.
- **Photons** are required to satisfy the **TIGHT** identification requirement and to have $p_T > 10$ GeV and $|\eta| < 2.37$. Photons are used at pre-selection level, only for vetoes.
- Hadronically-decaying **tau-leptons** are selected with $p_T > 20$ GeV, $|\eta| < 2.5$ (excluding the crack region $1.37 < |\eta| < 1.52$ between the EMCal barrel and the endcap calorimeters). The reconstructed taus are required to pass the **LOOSE** identification criteria. As photons, taus are only used in the veto selection.
- All **jets** are required to have $p_T > 30$ GeV and $|\eta| < 2.8$. The jet must meet the **MEDIUM JVT** working point. In this way, jets with $p_T > 120$ GeV and $|\eta| < 2.4$ must satisfy $JVT > 0.59$, while for jets with $2.4 < |\eta| < 2.5$ the requirement lowered to $JVT > 0.11$. In addition, the **LOOSE** jet cleaning criteria is required to all jets in the event, to remove anomalous energy depositions due to coherent noise and electronic noise bursts in the calorimeter.

- ***b*-jets.** The subset of jets with $|\eta| < 2.5$ which are identified by the MV2C10 *b*-tagger discriminant described in the previous chapter are treated as independent objects.

In order to resolve ambiguities between the different physical objects defined above (caused by the fact that their reconstructions are developed independently), an **overlap removal** technique is employed. The goal is not only to remove duplicated objects but also to wipe out close-by objects which can bias each other's position or energy reconstruction. A number of discrimination rules, extensively used within the ATLAS collaboration, is applied to baseline electrons, baseline muons, taus and jets, as defined above, in the following order:

1. If an electron and a muon are sharing the same track, the muon is prioritized as the electron is mainly coming from a converted photon radiated by the muon.
2. If a *b*-jet is found within a distance of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$ from a baseline electron or muon, the lepton is removed since it is most likely originated from a semi-leptonic *b* decay.
3. Jets closer than $\Delta R = 0.2$ from a well identified electron are typically electrons misreconstructed as jets, therefore they are removed from the event. However, the area right outside such duplication region ($0.2 < \Delta R < 0.4$) is typically populated by real hadronic jets with a fraction of shared clusters with the electron, biasing its reconstruction, hence in this case the jet is kept and the electron removed.
4. Jets and muons in a region of $\Delta R < 0.4$ are resolved in the following way:
 - if $\Delta R < 0.2$, the decision to keep the muon or the jet is based on the number of tracks associated to the jet. If they are 3 or more, the muon is most likely coming from a hadron decay, so the jet is kept. Otherwise the jet is likely a misreconstructed photon emitted by a prompt muon, or simply a low- p_T muon reconstructed as a jet, thus the muon is kept;
 - if $\Delta R > 0.2$ the jet is always kept.
5. Photons overlapping with electrons, muons and jets are resolved (in that order) within a cone of $\Delta R < 0.4$ simply by prioritizing leptons over photons, and the later over jets.
6. Similarly, overlaps between tau-leptons and electrons, muons or jets are resolved in a $\Delta R < 0.2$ cone by keeping the electron/muon in the leptonic case and removing the jet otherwise. The overlap between taus and jets is a tau identification step rather than an actual overlap removal.

Finally, once the ambiguities are resolved, the **missing transverse momentum**, E_T^{miss} , is calculated as discussed in the previous chapter, i.e. as the negative vectorial sum of the p_T associated to baseline

electrons and muons, taus, photons and jets, as well as the track-based soft term. In addition, the variable p_T^{recoil} is constructed by adding the transverse momentum of the leptons in the event to the E_T^{miss} , thus acting as a proxy for the transverse momentum of the system recoiling against the jets.

6.2.2 PRE-SELECTION

Once the physical objects are properly defined, a preliminary set of cuts is applied in all regions. The full pre-selection criteria is listed in Table 6.3.

Events accepted by the trigger and fulfilling the GRL requirements are first required to have at least one primary vertex with at least two associated tracks with $p_T > 500$ MeV.

At least one jet with $|\eta| < 2.4$ and $p_T > 150$ GeV is required, and up to 3 more jets are allowed. Non-collision background (NCB) contributions, such as energy deposits in the calorimeter from cosmic muons or beam-induced interactions, are suppressed by requiring the leading jet (the one with the largest p_T) to pass the TIGHT jet cleaning criteria.

QCD multi-jet events, mainly coming from jet energy mismeasurements, are rejected by requiring a minimum separation on the azimuthal angle between the jets and p_T^{recoil} :

$$\begin{cases} \min[\Delta\varphi(\text{jets}, p_T^{\text{recoil}})] > 0.6 & (p_T^{\text{recoil}} \leq 250 \text{ GeV}) \\ \min[\Delta\varphi(\text{jets}, p_T^{\text{recoil}})] > 0.4 & (p_T^{\text{recoil}} > 250 \text{ GeV}) \end{cases} \quad (6.1)$$

Note that the requirement in Eq. (6.1) is tightened for the low- p_T^{recoil} region, since in this regime the multi-jet background becomes more relevant.

Finally, no photons and no τ -leptons are allowed in the final state. The inclusion of the tau veto is one of the most important improvements with respect to the previous versions of the monojet analysis, allowing to reduce the total background in the SR by $\sim 10\%$.

6.2.3 SIGNAL REGION DEFINITION

In addition to the pre-selection requirements, the following cuts define the SR:

- $E_T^{\text{miss}} > 200$ GeV. All E_T^{miss} triggers are fully efficient at this regime.
- Lepton veto. No baseline electrons or muons are allowed in the final state.

Category	Selection Criteria
Trigger	Logic defined for the specific region
GRL	PHYS_StandardGRL_all_Good_25ns
Vertex	$N_{\text{vtx}} \geq 1$ (with $N_{\text{trk}} \geq 2$ and $p_{\text{T}} > 500$ MeV)
Leading jet	$ \eta < 2.4$
	$p_{\text{T}} > 150$ GeV
	$f_{\text{ch}}/f_{\text{max}} > 0.1$ (NCB rejection)
Jet multiplicity	$0 < N_{\text{jets}} \leq 4$
Multijet suppression	$\min[\Delta\phi(\text{jets}, p_{\text{T}}^{\text{recoil}})] > (0.6)0.4$
Photon veto	$N_{\gamma} = 0$
Tau veto	$N_{\tau} = 0$

Table 6.3: List of pre-selection cuts.

Note that the quantities $E_{\text{T}}^{\text{miss}}$ and $p_{\text{T}}^{\text{recoil}}$ become equivalent after the lepton veto. For this reason, they will be treated indistinguishably in the following.

In comparison to the previous iteration of this analysis, the $E_{\text{T}}^{\text{miss}}$ threshold has been lowered from 250 GeV to 200 GeV as a way to enhance the sensitivity to some signal models that exhibit a soft $E_{\text{T}}^{\text{miss}}$ spectrum, in particular to the invisible-decaying Higgs scenario. The p_{T} thresholds in the definition of baseline electrons, baseline muons and photons have been lowered as well, resulting in an increased rejection of background events in the SR when vetoing such objects.

6.2.4 CONTROL REGIONS DEFINITIONS

Five CRs are used to constrain the most important background contributions in the analysis. They are defined to be orthogonal to the SR and to each other, i.e. there is no overlap of events between regions. Events in the CRs are required to pass the pre-selection cuts. In addition, a common cut on $p_{\text{T}}^{\text{recoil}} > 200$ GeV is applied. The description of every CR is glossed below and the specific cuts are outlined in Table 6.4.

- The muon control region, $W \rightarrow \mu\nu$, is enriched in the background process $W(\mu\nu) + \text{jets}$. This is achieved by requiring exactly one signal muon in the event, no baseline electrons and no b -jets in the event (to ensure orthogonality with respect to the Top CR). In addition, a requirement

$W \rightarrow \mu\nu$	$Z \rightarrow \mu\mu$	$W \rightarrow e\nu$	$Z \rightarrow ee$	Top
1 signal μ	2 signal μ	No baseline μ	No baseline μ	Same as $W \rightarrow \mu\nu$ or same as $W \rightarrow e\nu$
No baseline e	No baseline e	1 signal e , $ \eta_e \notin (1.37, 1.52)$, TIGHT isolation	2 signal e	
$30 < m_T < 100$	$66 < m_{\mu\mu} < 116$	$30 < m_T < 100$	$66 < m_{ee} < 116$	
-	-	$E_T^{\text{miss}} > 70 \text{ GeV}$, $E_T^{\text{miss}}/\sqrt{H_T} > 5 \text{ GeV}^{1/2}$	-	
No b -jets	-	No b -jets	-	$\geq 1b$ -jet

Table 6.4: Control region specific cuts. The masses are expressed in GeV.

on the transverse mass of the muon–neutrino system is applied: $30 \text{ GeV} < m_T < 100 \text{ GeV}$,

$$\text{where } m_T = \sqrt{2 p_{T\mu} p_{T\nu} [1 - \cos(\varphi_\mu - \varphi_\nu)]}.$$

- The 2-muon CR, $Z \rightarrow \mu\mu$, is enriched in the $Z(\mu\mu) + \text{jets}$ background process. Events are selected if there are not baseline electrons and exactly two signal muons are found. Also, a very high purity is reached by applying a cut on the invariant mass of the muon–muon system around the mass of the Z , $66 \text{ GeV} < m_{\mu\mu} < 116 \text{ GeV}$.
- The electron control region, $W \rightarrow e\nu$, is defined such that is dominated by the background process $W(e\nu) + \text{jets}$. One signal electron is required and no baseline muons are allowed. In order to ensure high purity of the sample, electrons reconstructed within the crack region between the EM and endcap calorimeters are rejected (via a cut in the electron pseudo-rapidity as shown in Table 6.4), and selected electrons must pass the TIGHT isolation criteria. Likewise the $W \rightarrow \mu\nu$ CR, b -jets are vetoed and cut on the transverse mass is applied, $30 \text{ GeV} < m_T < 100 \text{ GeV}$, where this time $m_T = \sqrt{2 p_{Te} p_{T\nu} [1 - \cos(\varphi_e - \varphi_\nu)]}$. In addition, cuts on $E_T^{\text{miss}} > 70 \text{ GeV}$ and $E_T^{\text{miss}}/\sqrt{H_T} > 5 \text{ GeV}^{1/2}$ are applied to further suppress contamination from the multi-jet background, where H_T is defined as the sum of all jets momentum. This is also the reason for requiring isolation on the electron.
- The 2-electron region, $Z \rightarrow ee$, is enriched in the $Z(ee) + \text{jets}$ process. This region is constructed by inverting the roles of muons and electrons in the $Z \rightarrow \mu\mu$ selection. Therefore, two signal electrons are required, baseline muons are vetoed and the invariant mass is constrained around m_Z , $66 \text{ GeV} < m_{ee} < 116 \text{ GeV}$.
- The Top CR is dominated by single- t and $t\bar{t}$ processes. It is defined by events that pass the

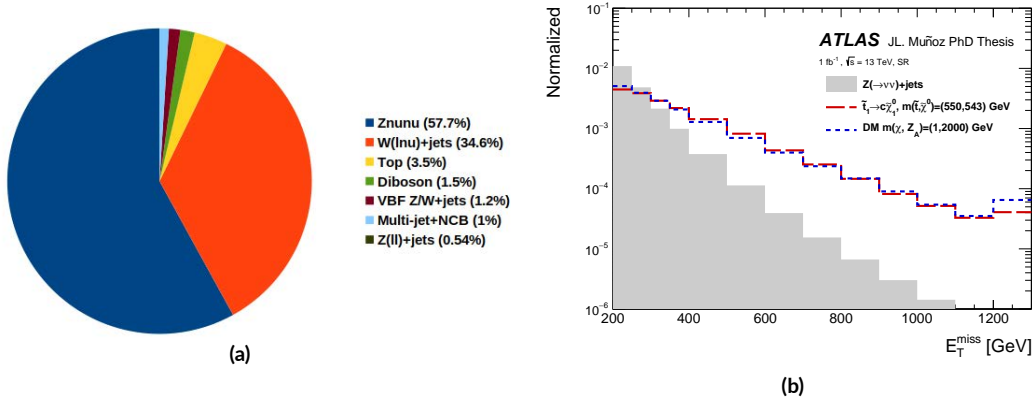


Figure 6.2: (a) Scheme of the background composition of the monojet SR. (b) E_T^{miss} shape comparison of the dominant background $Z(\nu\nu) + \text{jets}$ with respect to three benchmark signals, namely TT with $m(\tilde{t}, \tilde{\chi}_1^0) = (550, 543) \text{ GeV}$ and DMA with $m(\chi, Z_A) = (1, 2000) \text{ GeV}$.

selection of the $W \rightarrow \mu\nu$ or the $W \rightarrow e\nu$ but reverting the b -jet veto.

6.3 BACKGROUND ESTIMATION

After applying the SR selection described in Section 6.2, the monojet signature is compatible with a wide variety of SM processes. Figure 6.2a shows a scheme of the full background composition of the SR. The most relevant background sources are given by EW processes, where a vector boson W or Z is produced in association with jets, decaying into leptons and/or neutrinos. In particular, the $Z(\nu\nu) + \text{jets}$ process dominates largely ($\sim 58\%$) the scene, being also an irreducible background, i.e. there is no possible selection of events that can help to discriminate this kind of background against potential signal phenomena. There are also significant contributions from $W(\ell\nu) + \text{jets}$ processes ($\sim 35\%$), followed by small contributions from top-production ($t\bar{t}$ and single- t), diboson (WW, ZZ and WZ), vector-boson-fusion (VBF) Z/W production processes, and finally some minor contributions from di-lepton ($Z(\rightarrow \ell\ell)$), multi-jet and NCB processes.

6.3.1 ANALYSIS STRATEGY

The idea of the analysis is to exploit the differences in the shape of the E_T^{miss} distribution to discriminate background versus signal events. This is illustrated in Figure 6.2b, where two of the signal models considered in the analysis exhibit an enhanced tail in E_T^{miss} , while the background spectrum (represented by $Z(\nu\nu) + \text{jets}$) falls faster. For this reason, the SR is divided into 13 bins of E_T^{miss} , listed in Table 6.5,

Exclusive bins		Inclusive bins	
Name	E_T^{miss} [GeV]	Name	E_T^{miss} [GeV]
EM ₀	(200, 250]	IM ₀	(200, +∞)
EM ₁	(250, 300]	IM ₁	(250, +∞)
EM ₂	(300, 350]	IM ₂	(300, +∞)
EM ₃	(350, 400]	IM ₃	(350, +∞)
EM ₄	(400, 500]	IM ₄	(400, +∞)
EM ₅	(500, 600]	IM ₅	(500, +∞)
EM ₆	(600, 700]	IM ₆	(600, +∞)
EM ₇	(700, 800]	IM ₇	(700, +∞)
EM ₈	(800, 900]	IM ₈	(800, +∞)
EM ₉	(900, 1000]	IM ₉	(900, +∞)
EM ₁₀	(1000, 1100]	IM ₁₀	(1000, +∞)
EM ₁₁	(1100, 1200]	IM ₁₁	(1100, +∞)
EM ₁₂	(1200, +∞)	IM ₁₂	(1200, +∞)

Table 6.5: Bins of E_T^{miss} used in the analysis.

which are optimized for giving good resolution on the E_T^{miss} shape without losing statistical power. Inclusive and exclusive bins are defined for different purposes. Regarding the previous version of the analysis, 3 more bins have been included, one for the lowered E_T^{miss} threshold (250 \rightarrow 200 GeV) as mentioned, and adding two more bins between 1.0 and 1.3 TeV to profit from the increased statistics.

In order to optimize such discrimination, it is crucial to measure the total background contributions with the highest level of precision as possible. The contributions from the different background processes are estimated via a number of different methods that are summarized in Table 6.6, and explained as follows.

MC samples of simulated events are generated for most of the background processes described above, with the exception of the multi-jet and non-collision backgrounds which need a different treatment based on data-driven techniques (described in Sub-ses. 6.3.4 and 6.3.3, respectively). The generation of the MC samples is fully described in Chapter 2. These samples provide already a preliminary estimation of the background yields, once the events are properly normalized to the luminosity of the data sample. Diboson and VBF W/Z +jets contributions are directly estimated from such MC predictions.

Process	Strategy
$Z(\nu\nu) + \text{jets}$	MC shape + theory corrections + CR fit
$W(\mu\nu) + \text{jets}$	MC shape + theory corrections + CR fit
$W(e\nu) + \text{jets}$	MC shape + theory corrections + CR fit
$W(\tau\nu) + \text{jets}$	MC shape + theory corrections + CR fit
$Z(\mu\mu) + \text{jets}$	MC shape + theory corrections + CR fit
$Z(\tau\tau) + \text{jets}$	MC shape + theory corrections + CR fit
$Z(ee) + \text{jets}$	MC shape + theory corrections + CR fit
single- t	MC shape + CR fit
$t\bar{t}$	MC shape + CR fit
diboson	From MC
VBF $W(\ell\nu)+\text{jets}$	From MC
VBF $Z(\ell\ell/\nu\nu)+\text{jets}$	From MC
Multi-jet	Jet Smearing
NCB	Tagger

Table 6.6: Background estimation methods used for the different processes contributing to the SR.

An additional higher-order $p_T(V)$ -based reweighting is performed on the V +jets samples, achieving a much preciser description of the data for these processes, which are especially relevant in the analysis given their large contributions in the SR. Such corrections not only give a better precision, but also provide a way to constrain the $Z(\nu\nu) + \text{jets}$ background. This is because the correlations of the theoretical uncertainties across the different W/Z +jets processes are taken into account in the corrections, therefore the residual differences bin-by-bin are covered by the resulting systematic uncertainties. Taking advantage of this feature, the strategy then is to perform a simultaneous fit to the data in the CRs, and use such statistical power to extract a unique normalization factor for all the V +jets processes. Similarly, another two floating normalization factors are used for the single- t and $t\bar{t}$ processes in the same fit. The CRs fit is discussed in detail in the next section.

6.3.2 CONTROL REGIONS FIT

A simultaneous, binned likelihood fit is carried out using the p_T^{recoil} distribution in the five CRs defined in Sub-section 6.2.4. Every region is split into exclusive bins of p_T^{recoil} as listed in Table 6.5, which are fitted simultaneously so the p_T^{recoil} -shape information is used to extract the normalization for the V +jets, single- t and $t\bar{t}$ processes.

The number of events in each bin is treated as a random variable with a Poisson distribution

function \mathcal{P} . The expectation value is given by the sum of the background predictions in the bin, corrected by a global κ -factor, which is specific for different background processes and is treated as a free parameter in the fit. The normalization factors assigned to V +jets, single- t and $t\bar{t}$ are named κ^V , κ^t and $\kappa^{t\bar{t}}$, respectively*. The likelihood function is then defined as

$$\mathcal{L}(\mu, \boldsymbol{\kappa}, \boldsymbol{\theta}) = \prod_i \prod_j \mathcal{P} \left(N_{ij}^{\text{obs}} \mid \mu N_{ij}^{\text{sig}}(\boldsymbol{\theta}) + N_{ij}^{\text{bkg}}(\boldsymbol{\kappa}, \boldsymbol{\theta}) \right) f_{\text{constr}}(\boldsymbol{\theta}), \quad (6.2)$$

where the indexes i and j run over p_T^{recoil} bins and regions, respectively, and N_{ij}^X denotes the number of observed ($X = \text{obs}$), signal ($X = \text{sig}$) or background ($X = \text{bkg}$) events in the i -th bin and region j . The number of signal events is weighted by the signal strength, μ , and is expressed as a function of the nuisance parameters, represented by the vector $\boldsymbol{\theta}$. The final background prediction, whose dependence on the κ -factors is denoted by $\boldsymbol{\kappa} = (\kappa^V, \kappa^t, \kappa^{t\bar{t}})$, is expressed in terms of the individual background contributions as

$$\begin{aligned} N_{ij}^{\text{bkg}}(\boldsymbol{\kappa}, \boldsymbol{\theta}) = & \kappa^V \left[N_{ij}^{Z(\nu\nu)+\text{jets}} + N_{ij}^{Z(\ell\ell)+\text{jets}} + N_{ij}^{W(\ell\nu)+\text{jets}} \right] + \kappa^t N_{ij}^{\text{single-}t} + \kappa^{t\bar{t}} N_{ij}^{t\bar{t}} \\ & + N_{ij}^{\text{VBF } W(\rightarrow\ell\nu)+\text{jets}} + N_{ij}^{\text{VBF } Z(\rightarrow\ell\ell)+\text{jets}} + N_{ij}^{\text{diboson}} + N_{ij}^{\text{multi-jet}} + N_{ij}^{\text{NCB}}, \end{aligned} \quad (6.3)$$

where the dependence on $\boldsymbol{\theta}$ is omitted for simplicity. The term $f_{\text{constr}}(\boldsymbol{\theta})$ in Eq. (6.2) represents the product of the Gaussian constrains applied to each nuisance parameter. The inclusion and treatment of the systematic uncertainties is discussed in the next chapter.

6.3.3 NON-COLLISION BACKGROUND

Non-Collision Background refers to physical processes that are not initiated by pp collisions but leave a signal in the detector with very similar topologies to the monojet final state. Typically, it consists of cosmic muons originated in the atmosphere and BIB, i.e. muons travelling parallel to the beam pipe produced in inelastic interactions between protons in the beam halo and residual gases in the vacuum chambers. This kind of events are efficiently rejected by a factor $\mathcal{O}(10^3)$ with the jet TIGHT cleaning criteria, but a small remnant can still contribute to the SR. Such residual contamination can not be simulated, therefore a data-driven method is employed instead. This method, described in detail in Appendix C, relies on the use of the jet-timing variable to discriminate jets consistent with an origin

*In the previous version of the analysis, a single κ -factor was used for both single- t and $t\bar{t}$ processes instead. The usage of two independent κ -factors improves the description of the data since it allows to exploit the shape differences in p_T^{recoil} of the different top-production channels.

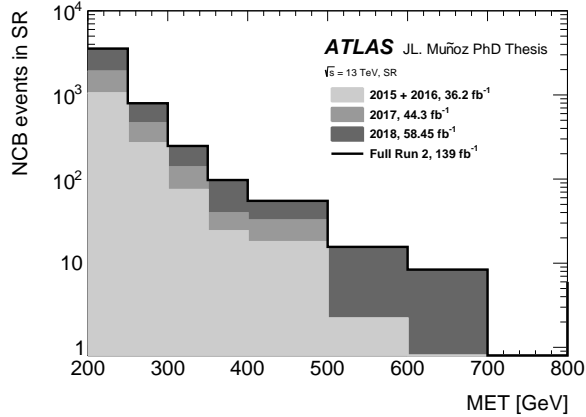


Figure 6.3: NCB contribution to the SR split in years of data taking.

in the main hard collision against those that do not, which typically leave a signal in the calorimeter shifted in time with respect to that of the bunch-crossing.

The results indicate a residual contribution of about 0.2% and 0.1% of NCB events in the exclusive E_T^{miss} bins EM₀ and EM_I, respectively, and below 0.1% for $E_T^{\text{miss}} > 300$ GeV. Figure 6.3 shows the breakdown of the NCB residual estimations in the different years of data-taking. NCB contributions in the CRs are neglected and a conservative uncertainty of 100% is assigned to this background in the fit.

6.3.4 MULTI-JET BACKGROUND

The multi-jet background refers to QCD multi-jet events for which there is no precise QCD prediction. They come mainly from misreconstructions of the jet energy in the calorimeter and, up to a lesser extent, from the presence of neutrinos in the jet cone from heavy-flavour hadron decays. These events are efficiently rejected by applying a cut on the azimuthal separation between E_T^{miss} and the jets, as defined by Eq. (6.1), but still some contribution to the SR remains left.

Such signatures are dominated by fluctuations of the jet response in the calorimeter. Given the difficulty to reproduce such non-Gaussian effects on the jet measurement in the MC simulations, a data-driven technique called *Jet Smearing Method* is employed. This procedure, described in detail in Ref. [170], relies on the use of a multi-jet enriched control region, where the jet response is measured from simulated di-jet events and matched to *smear*ed seed events, selected from data. This region is constructed by applying a SR-like selection, where the cut on $\Delta\phi$ is modified such that events with

E_T^{miss} interval [GeV]	Events in SR
(200, 250]	18604.7
(250, 300]	5421.8
(300, 350]	1130.6
(350, 400]	387.3
(400, 500]	57.2
(500, 600]	4.5
(600, 700]	0.6
(700, 800]	0.1
(800, $+\infty$)	0.1

Table 6.7: Multi-jet estimate in the exclusive signal region bins used in the analysis.

$\min[\Delta\phi(\text{jets}, E_T^{\text{miss}})] < 0.3$ are accepted instead. The seed events are selected by using the E_T^{miss} -significance variable, which is defined as

$$S = \frac{E_T^{\text{miss}} - 8}{\sqrt{\sum E_T}}, \quad (6.4)$$

where $\sum E_T$ is the scalar sum of the event transverse energy. Seed events are those with $S < 0.05$ GeV. The normalization of the multi-jet background, $k_{\text{multi-jet}}$, is then extracted via

$$k_{\text{multi-jet}} = \frac{N_{\text{multi-jet}}^{\text{data}} - N_{\text{multi-jet}}^{\text{MC bkg}}}{N_{\text{multi-jet}}^{\text{smeared}}}, \quad (6.5)$$

where $N_{\text{multi-jet}}^{\text{data}}$ denotes the total number of data events in the multi-jet region, $N_{\text{multi-jet}}^{\text{MC bkg}}$ refers to other backgrounds in the region (namely, W/Z -jets, single- t , $t\bar{t}$ and diboson), which are subtracted by using their respective MC yields, and, finally, $N_{\text{multi-jet}}^{\text{smeared}}$ denotes the number of smeared events (defined above). In addition, a similar multi-jet region is constructed in the interval $0.3 < \min[\Delta\phi(\text{jets}, E_T^{\text{miss}})] \leq 0.4$, where the procedure described above is repeated in order to validate the normalization of the predicted QCD background.

The resulting estimations are of about 1.2, 0.8, 0.4 and 0.3% of multi-jet contribution in the exclusive bins EM₀, EM₁, EM₂ and EM₃, respectively, and below 0.1% for $E_T^{\text{miss}} > 400$ GeV. The full multi-jet estimations are shown in Table 6.7. As for the NCB, no multi-jet contributions are considered in the CRs and a 100% uncertainty is adopted.

6.4 PRE-FIT DISTRIBUTIONS

The p_T^{recoil} distributions in the CRs are shown in Figure 6.4 as they are before performing the fit to the data. In the following Figures 6.5–6.11 some other kinematic variables are also shown, namely the leading jet p_T and $|\eta|$ distributions, the jet multiplicity, the p_T distributions of the second (sub-leading) and third (sub-sub-leading) jets and, finally, the lepton-neutrino system transverse mass and the di-lepton invariant mass distributions for the W and Z control regions, respectively. Systematical uncertainties are not included in all these plots, therefore only statistical uncertainties are taken into account. Only those background processes contributing in a non-negligible way to the corresponding region are shown in the figures, although all samples are included in the calculation of the total background (denoted by a solid black line) and the total uncertainties (dashed grey shadow around the total background line). All W +jets and Z +jets processes are shown merged as two individual background sources in the figures. Top-production processes are also included together as a unique process, with the exception of the Top CR, where single- t and $t\bar{t}$ are shown separately to remark their shape differences, which are especially relevant for the fit in this particular region.

In the $W \rightarrow \mu\nu$, $W \rightarrow e\nu$, $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ CRs, which are dominated by V +jets processes, a flat discrepancy of about 15% between predictions and data is observed, where the MC estimations seem to underestimate the data. This effect is caused by the V +jets re-weighting discussed in Chapter 2. The offset is caused by the fact that such theoretical corrections are prescribed for inclusive processes. In fact, in Fig. 6.7 it can be seen how the discrepancy is coming mainly from events with one or two jets. Besides the offset, the ratios data/MC are mostly flat. The correct normalization will be extracted by the simultaneous fit to the data.

In the Top CR, no offset between data and background predictions is observed, since this region is dominated by single- t and $t\bar{t}$ processes, and the data are compatible with the MC estimations within statistical uncertainties. There is a deficit of about 15% of predicted events with exactly one jet, as it can be seen in Figure 6.7. In this particular sub-region, the presence of V +jets dominates over top-production processes, thus the disagreement is likely stemming from the aforementioned normalization offset among V +jets samples.

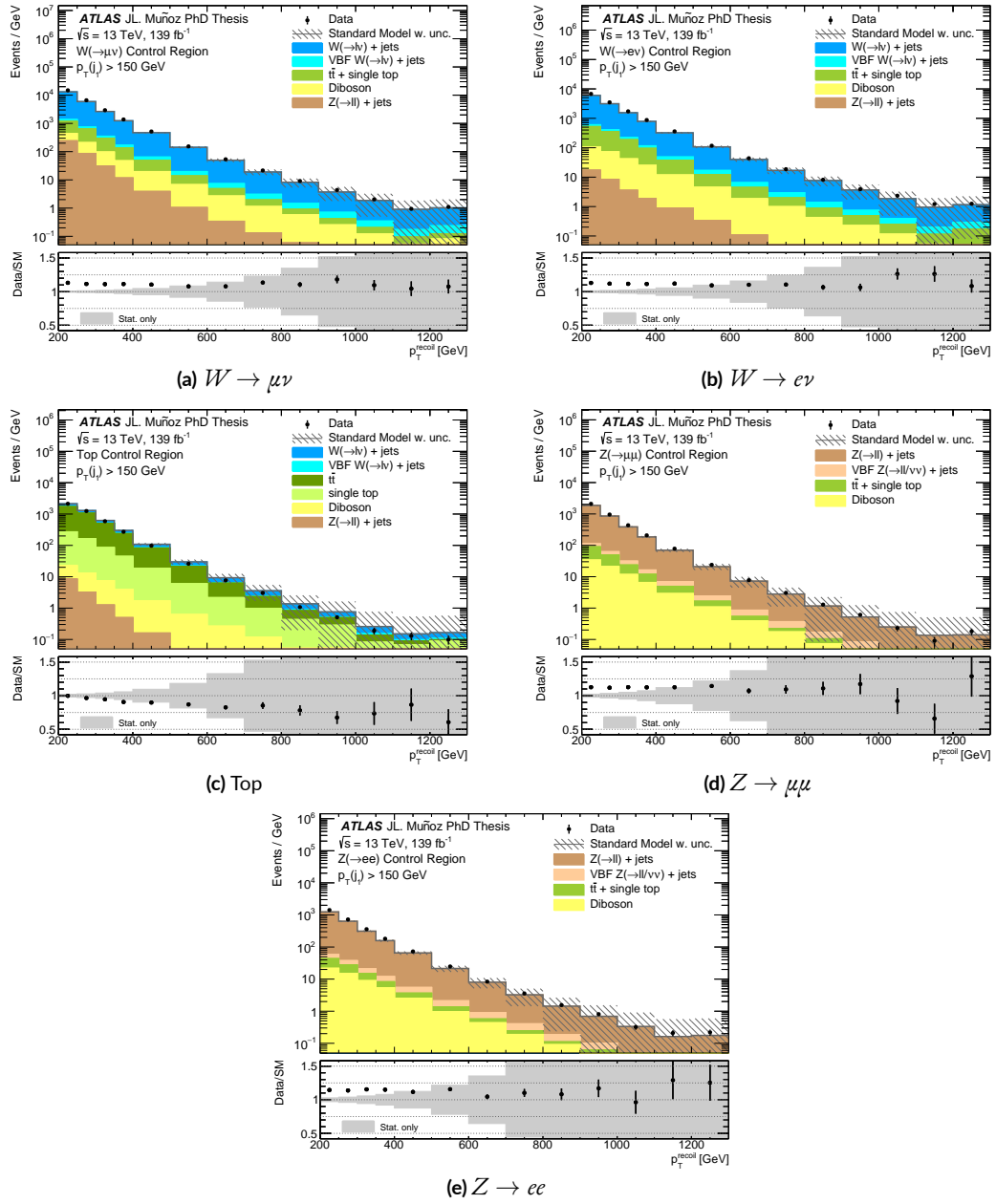


Figure 6.4: p_T^{recoil} pre-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes only statistical uncertainties. The lower panels show the ratio of data over background prediction.

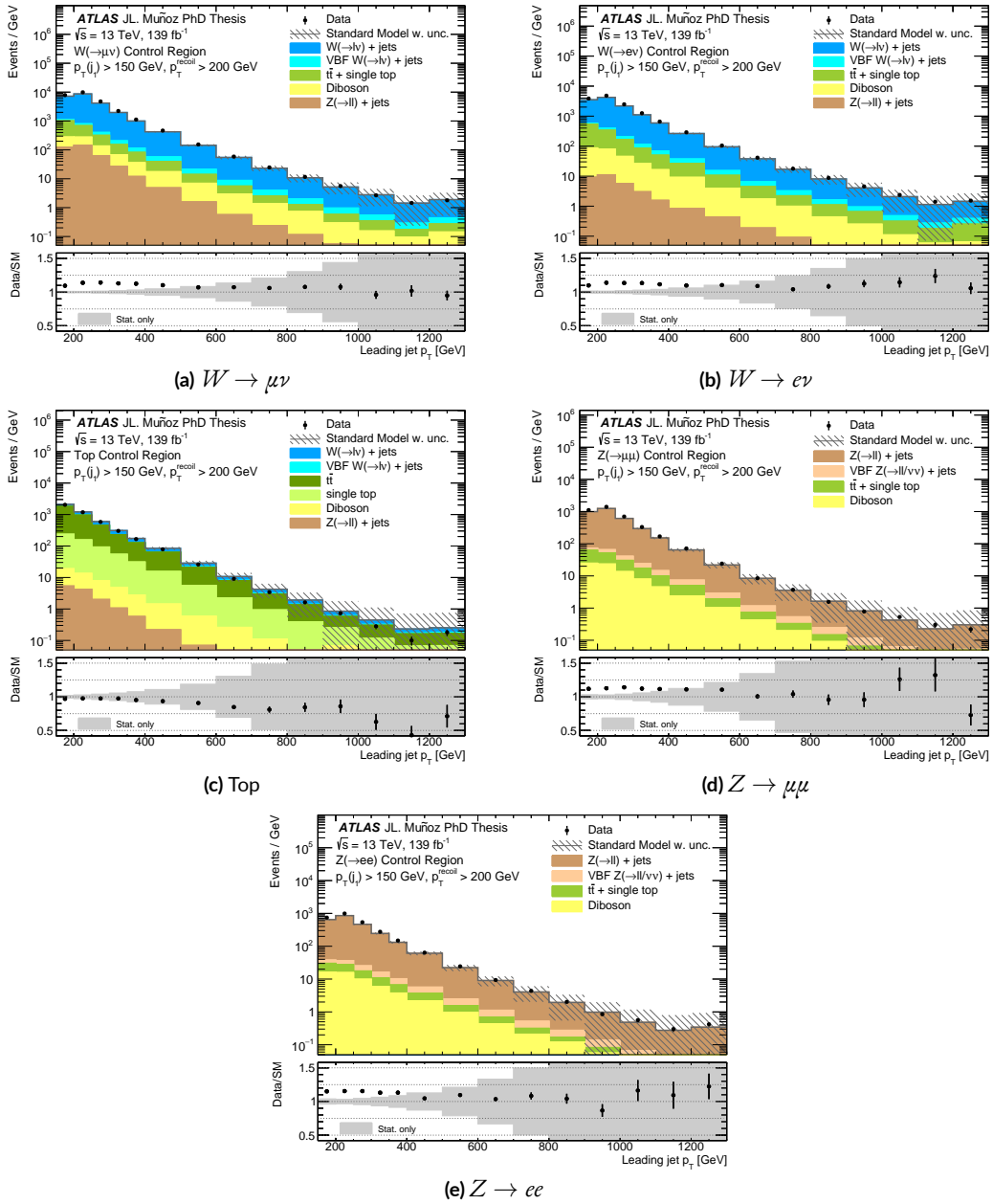


Figure 6.5: Leading jet p_T pre-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes only statistical uncertainties. The lower panels show the ratio of data over background prediction.

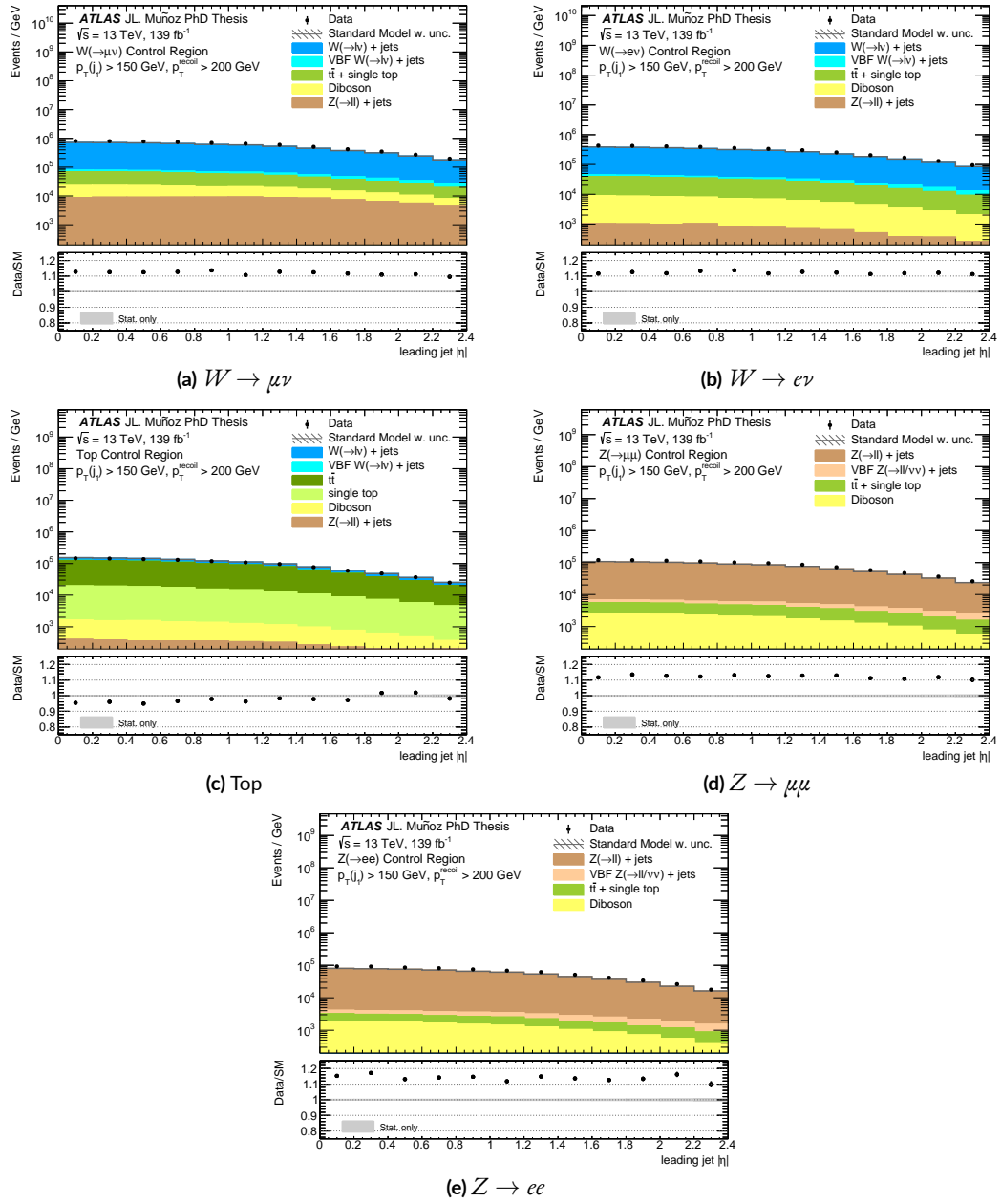


Figure 6.6: Leading jet p_T pre-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes only statistical uncertainties. The lower panels show the ratio of data over background prediction.

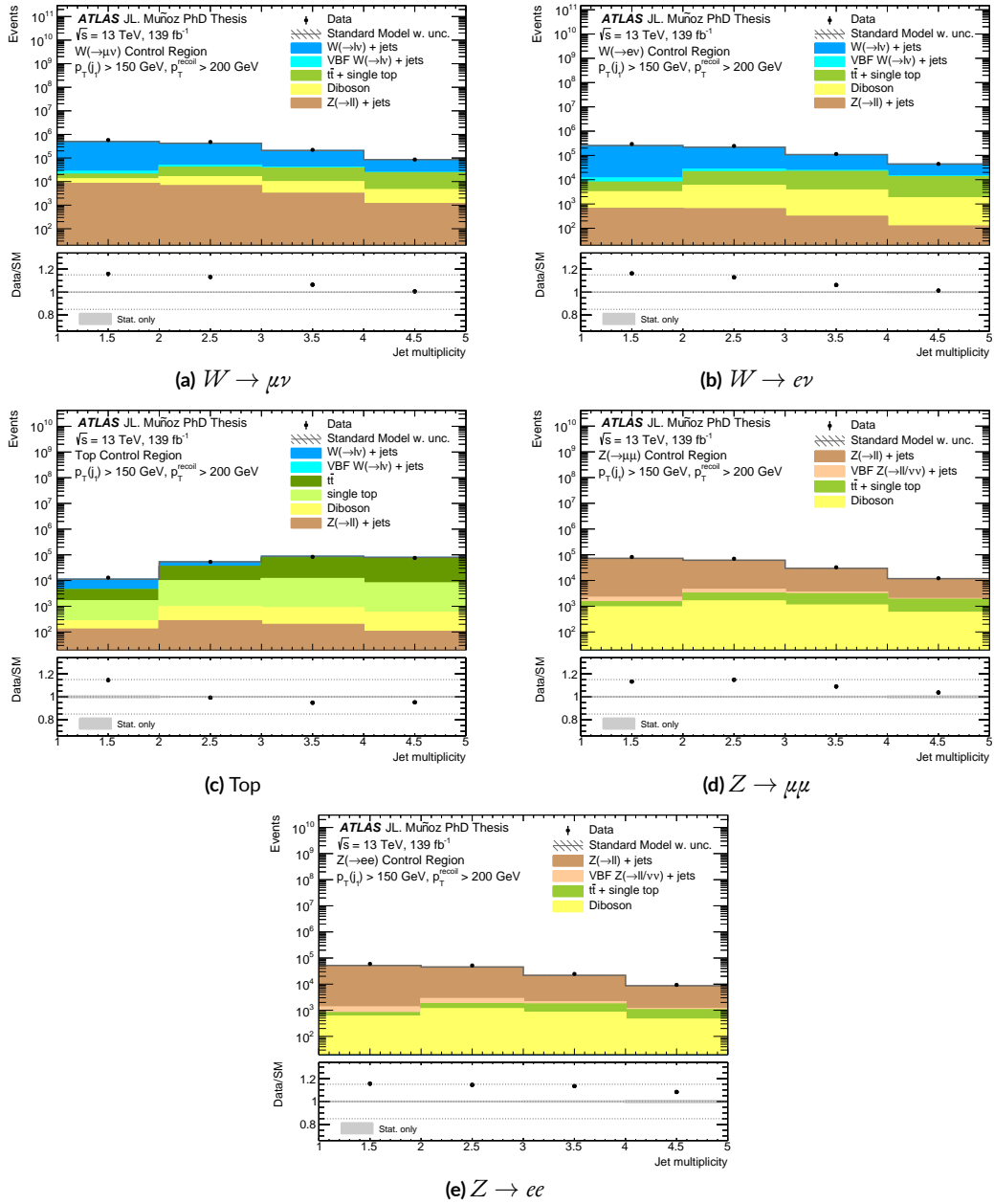


Figure 6.7: Distribution pre-fit of number of jets in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes only statistical uncertainties. The lower panels show the ratio of data over background prediction.

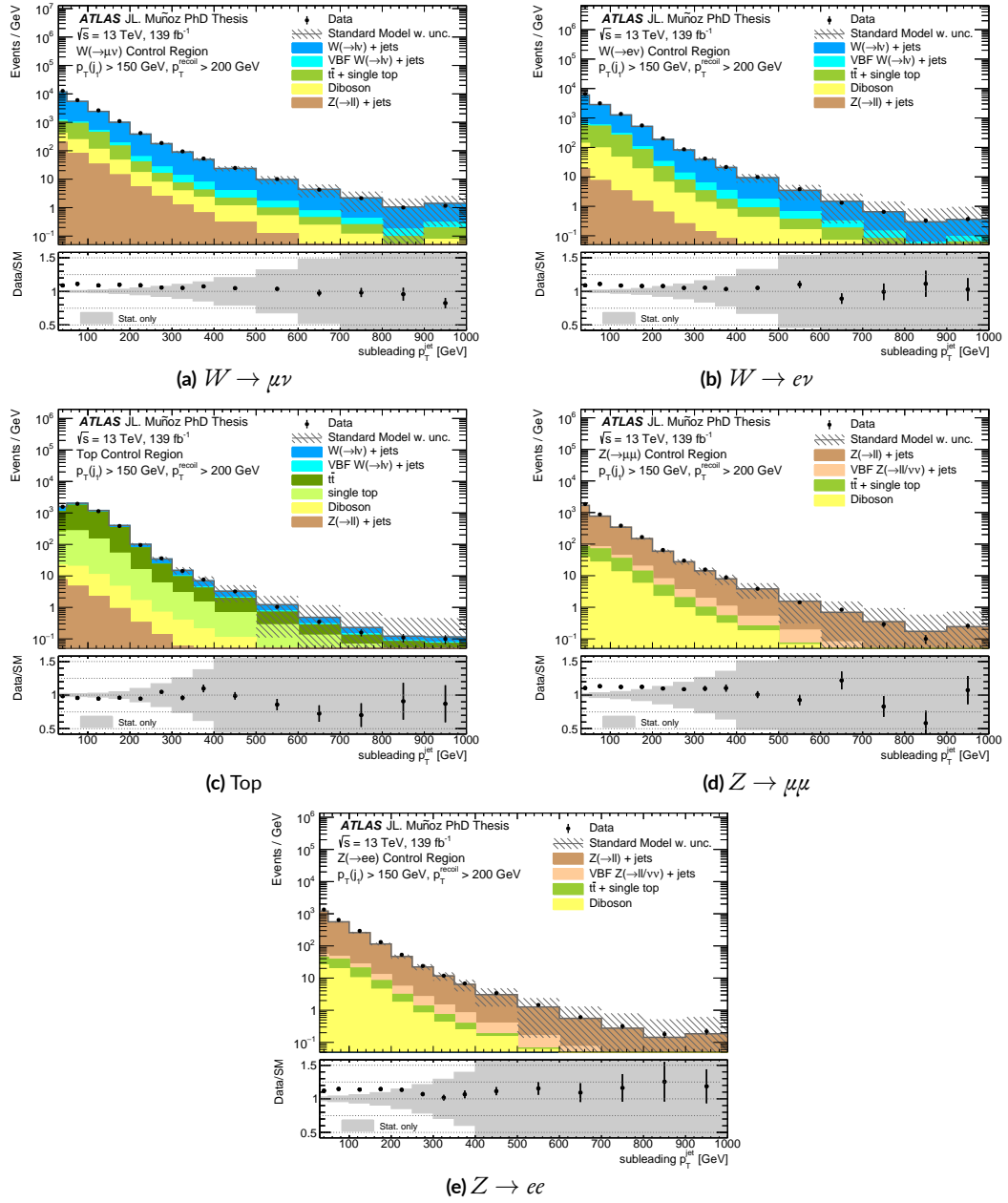


Figure 6.8: Second jet p_T pre-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes only statistical uncertainties. The lower panels show the ratio of data over background prediction.

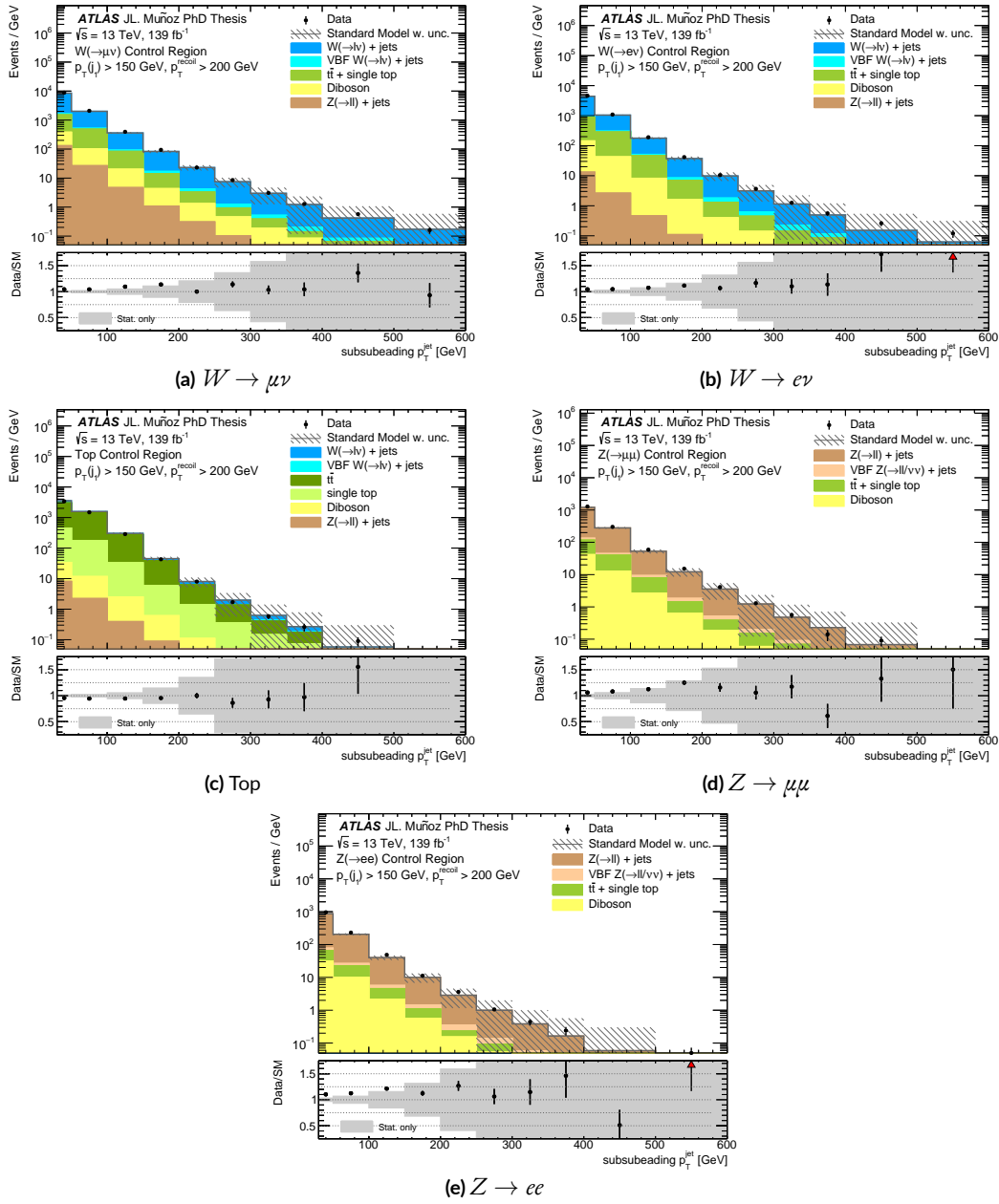


Figure 6.9: Third jet p_T pre-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes only statistical uncertainties. The lower panels show the ratio of data over background prediction.

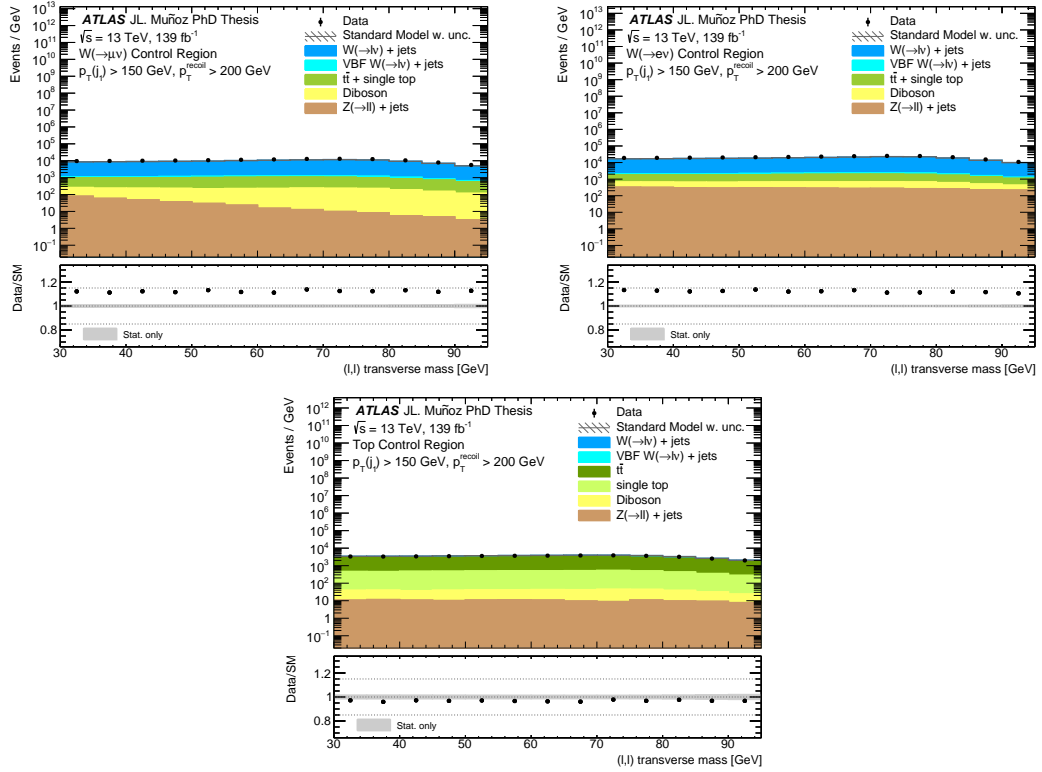


Figure 6.10: Lepton-neutrino transverse mass pre-fit distribution in the one-lepton CRs, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes only statistical uncertainties. The lower panels show the ratio of data over background prediction.

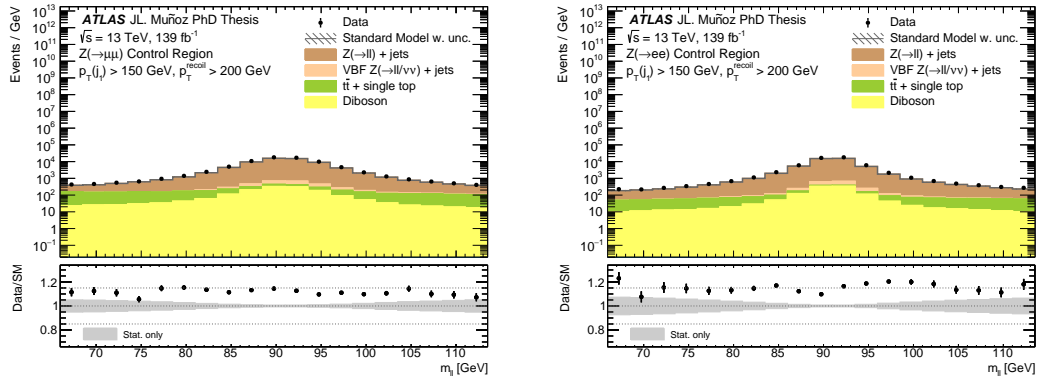


Figure 6.11: Di-lepton invariant mass pre-fit distribution in the two-lepton CRs (d and e), based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes only statistical uncertainties. The lower panels show the ratio of data over background prediction.

7

Systematic Uncertainties

A large number of uncertainties have to be considered in the analysis in order to account for potential mismeasurements and imperfections in the experimental setup, as well as for the limited precision of the theoretical predictions, which are propagated through the event reconstruction and the analysis chain and might affect the normalization and/or the p_T^{recoil} shape of any signal or background process. Many different techniques are then employed to evaluate such set of systematic uncertainties in bins of p_T^{recoil} . Unless otherwise is said, the uncertainties are correlated across bins. The description of the systematics evaluation is discussed in this Chapter separately for experimental and theoretical uncertainties, dividing the later into background and signal uncertainties in regard of the kind of process affected.

7.1 IMPLEMENTATION IN THE FIT

Systematic uncertainties are included in the fit via Gaussian distributed nuisance parameters with an initial value of 0 for the mean and a standard deviation of 1, represented by the term $f(\theta)$ in the likelihood function described in Eq. (6.2). An additional Gaussian constraint is included to reflect the data statistical uncertainties in a given region and bin only if $\sqrt{N_{\text{data}}}/\sigma_{\text{MC}}^{\text{stat}} < 10$, where $\sigma_{\text{MC}}^{\text{stat}}$ is the sum in

quadrature of the MC statistical uncertainties of background processes in the bin.

7.1.1 SYSTEMATICS SMOOTHING

The monojet Run 2 analysis is largely limited by systematic uncertainties, given the large integrated luminosity available. This makes the fit highly sensitive to the size and the shape of the systematic variations, especially in the first p_T^{recoil} bins which contains the core of the statistical power. For this reason, any non-physical fluctuation in the systematics can alter significantly not only the outcome of the fit but also its very convergence and numerical stability. In order to reduce such effects, mostly caused by the limited statistics of the MC samples used, a smoothing algorithm is applied to all the systematic variations before being implemented in the fit.

The algorithm runs in two steps. First, the p_T^{recoil} variation distribution is re-binned until no local extrema remain. Then, starting from the right edge of the distribution, bin contents are merged until the statistical uncertainty of the resulted bin is lower than 5%. A more careful approach is adopted for some specific cases to preserve large systematic variations at high p_T^{recoil} .

7.1.2 SYSTEMATICS PRUNING

In order to further ensure the numerical stability of the fit, a minimum set of requirements is used to select the systematic variations that are included in the fit. This procedure is called ‘pruning’ and is applied only for experimental uncertainties. The pruning is performed in three steps as follows. If σ_X^r is a systematic variation applied to the process X in the region r ,

1. The integrated p_T^{recoil} distribution of σ_X^r must be $> 0.2\%$ of $\sum_N X$, where N is the total number of processes in the region r ;
2. $\sigma_X^r > 0.2\%$ in any p_T^{recoil} bin;
3. $\sigma_X^r \cdot X > 0.002\%$ of X in any p_T^{recoil} bin.

If the three conditions are passed, σ_X^r is accepted by the fitting algorithm.

7.2 EXPERIMENTAL UNCERTAINTIES

Experimental uncertainties are those that stem from the reconstruction, calibration, identification and selection efficiencies of the several physical objects used in the analysis, as well as from the measure-

ment of various quantities that are included in the overall event selection efficiency (e.g. the luminosity). In the following, the different sources of experimental systematics are listed and described briefly, together with their contribution to the total background uncertainty in the SR before performing the fit. A list with of all the experimental uncertainties considered is shown in Table 7.1.

7.2.1 LUMINOSITY

An uncertainty of 1.7% is assigned to the luminosity measurement corresponding to the 2015–2018 period. This is a combination of the various sources of uncertainty on the luminosity calibration, calculated, as explained in Ref. [128], taking into account the correlations between the different uncertainties and across the years. The largest individual source of uncertainty stems from the calibration transfer procedure, used to export the calibration obtained with vdM scans from low- μ runs to the actual physics data-taking conditions. One NP is included in the fit for all MC background and signal samples to account for this uncertainty, which translates into a contribution to the total background uncertainty in the SR of about 1.68% for $p_{\text{T}}^{\text{recoil}} = 200$ GeV and of 1.7% for $p_{\text{T}}^{\text{recoil}}$ above 400 GeV.

7.2.2 PILE-UP RE-WEIGHTING

A single NP is included in the fit to account for the uncertainty associated to the pile-up correction procedure, which is applied to all MC samples. The impact of this systematic in the SR is of less than 0.1% for $p_{\text{T}}^{\text{recoil}} = 200$ GeV and about 0.2% for $p_{\text{T}}^{\text{recoil}} = 1.2$ TeV.

7.2.3 JETS

Different sources of jet-related uncertainties are considered in the analysis. Such uncertainties are propagated through the whole analysis chain for the signal and background MC samples that are used in the analysis, and the final up and down variations are extracted from the difference in yields between the nominal and varied samples in bins of $p_{\text{T}}^{\text{recoil}}$.

JET ENERGY SCALE

The jet energy scale (JES) is one of the dominant sources of uncertainty in the analysis. It comprises a large number of systematic uncertainties, derived from the in-situ calibrations, pile-up effects and flavour dependence, among other uncertainties related to the jet calibration procedure [145]. Such uncertainties are evaluated as a function of the jet p_{T} and η and, as shown in Figures 7.1a and 7.1b,

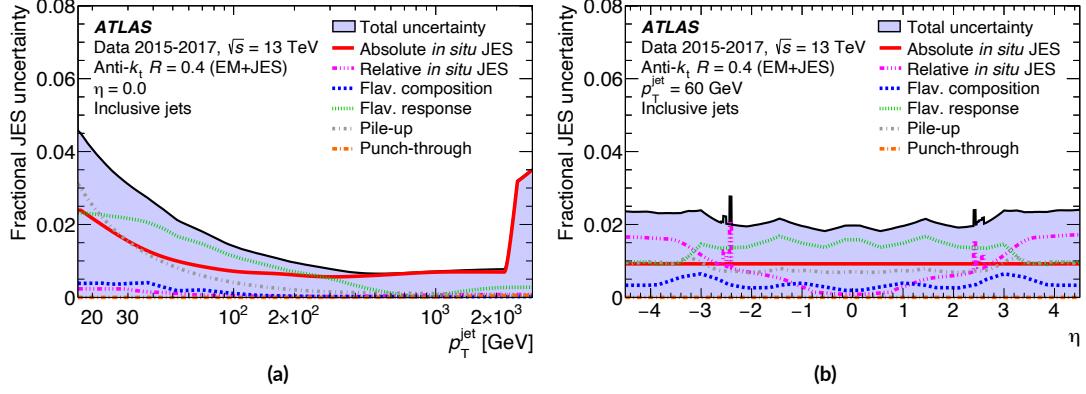


Figure 7.1: Jet energy scale systematic uncertainty components as a function of p_T at $\eta = 0$ (a) and as a function of η at $p_T = 60$ GeV (b). The total uncertainty, determined as the quadrature sum of all components, is shown as a filled region topped by a solid black line. Flavour-dependent components shown here assume a di-jet flavour composition [145].

they are relatively flat in η while range between 1% for a wide range of high- p_T jets ($250 < p_T < 2000$ GeV), 5% at very low p_T (20 GeV) and 3.5% at very high p_T (> 2.5 TeV).

The JES uncertainty is implemented via a total number of 30 nuisance parameters. Most of them (15) are related to in-situ measurements, covering effects such as selection cuts, MC mismodellings or statistical limitations involved in the jet in-situ analyses. One of the largest of these uncertainties is shown in Figure 7.2a for the $Z(\nu\nu) + \text{jets}$ sample in the SR, where varies between 2%–2.5% at low- p_T^{recoil} and $\sim 1.5\%$ at high- p_T^{recoil} . Five NPs are included to account for uncertainties related to the jet η inter-calibration analysis: one for systematic effects, another one for statistical uncertainty, and three more to parameterize the non-closure. Four NPs describe pile-up effects, accounting for offsets and p_T dependence in $\langle \mu \rangle$ and N_{PV} , as well as event topology dependence of the density metric ρ . Two NPs are used to represent the flavour dependence uncertainties, derived from simulation and accounting for differing responses to quark and gluon jets. An additional flavour-related uncertainty is applied only to b -jets to cover the difference in response between light- and heavy-flavour jets. These are some of the largest jet-related uncertainties, as shown in Fig. 7.2b, where the jet flavour response uncertainty for the $t\bar{t}$ sample in the Top region varies between 2.5%, less than 0.5% and 1.5% for values of p_T^{recoil} of 200, 600 and 1200 GeV, respectively. The punch-through uncertainty accounts for mismodelling of the GSC correction to jets which pass through the calorimeter into the MS, taking the difference in jet response between data and MC simulation as the uncertainty. Finally, the high- p_T ‘single particle’ uncertainty is derived from studies of the response to individual hadrons and reflects the lack of statistical power of the multi-jet balance analysis in the region above 2.4 TeV. In addition, for samples generated using fast simulation (most of the signal samples, excepting the ones dedicated to the invisible-decaying Higgs interpretation), an additional non-closure uncertainty is included to

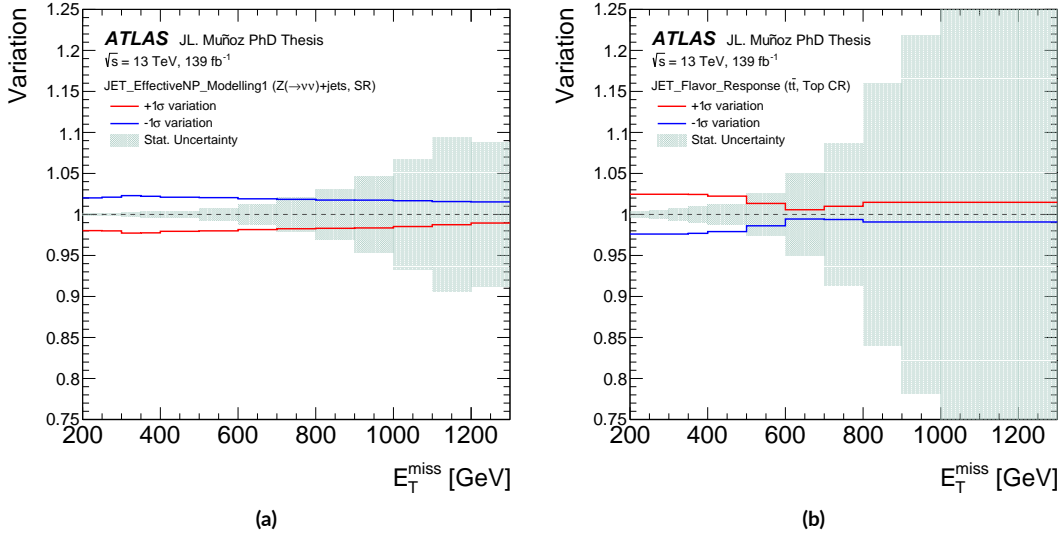


Figure 7.2: Impact of two of the largest JES uncertainties of the analysis in different regions: (a) one of the NPs related to the uncertainties in the in-situ calibrations in the SR for the $Z(\nu\nu) + \text{jets}$ sample; and (b) the jet flavour response uncertainty in the Top CR for the $t\bar{t}$ sample. The shaded area denotes the statistical uncertainty.

account for the difference in jet response between such samples and those which used full detector simulation.

JET ENERGY RESOLUTION

Jet energy resolution (JER) uncertainties are propagated through the analysis chain by smearing jets as described in Ref. [145]. Such smearing is performed to ensure that the resolution of the jet energy scale in MC events matches to that of the data wherever is possible and is carried out in different regions of jet p_T . Anti-correlations between the different JER components are taken into account in the smearing as follows. If σ_{NP} denotes the 1σ variation of a given uncertainty component, the smearing is applied to simulation events in those jet- p_T regions where $\sigma_{\text{NP}} > 0$, and applied to the data otherwise. Regarding the $V + \text{jets}$ processes, since they all have very similar topologies, the JER systematics are evaluated using only the $Z(\nu\nu) + \text{jets}$ sample and the resulting variations are applied then to the rest of $V + \text{jets}$ samples, in order to reduce statistical fluctuations.

Figure 7.3a shows the absolute JER uncertainty as a function of jet p_T , as measured with simulated di-jet events. This uncertainty is fairly flat for the region that concerns to the monojet analysis (i.e. $p_T > 150$ GeV) except for a small bump around 900 GeV that stems from the single-particle uncertainty (which is part of the JES uncertainty, discussed in the previous section), but always below

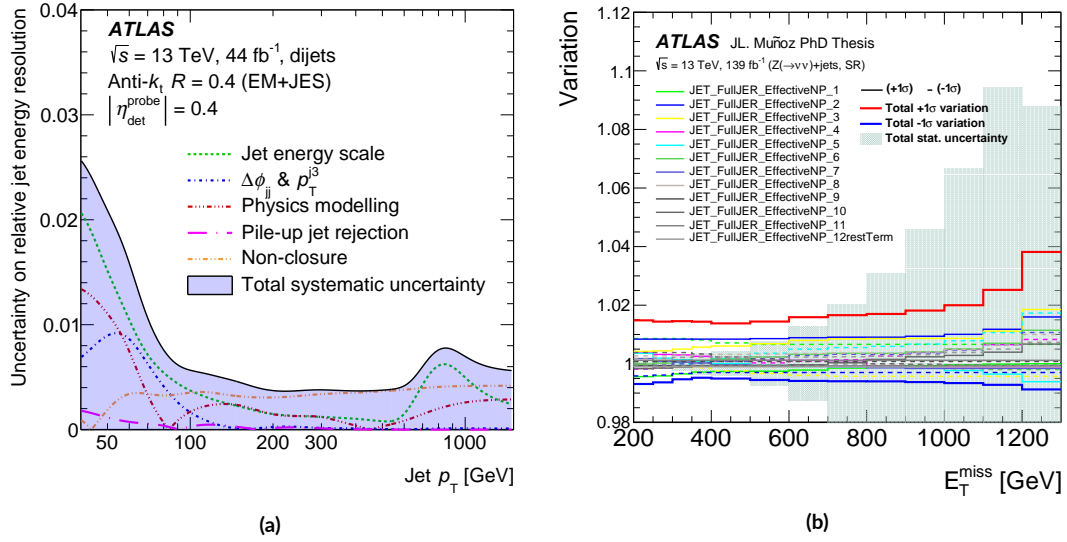


Figure 7.3: (a) Absolute uncertainty on the relative JER as a function of p_T for central jets [145]. (b) The JER uncertainties in the SR for the $Z(\nu\nu) + \text{jets}$ sample, split into its 12 components. The solid red (blue) line represents the sum in quadrature of all contributions in the up (down) side direction and the shaded area denotes the statistical uncertainty.

1%. Figure 7.3b shows the impact of such uncertainty, split into its 12 components, on the $Z(\nu\nu) + \text{jets}$ sample in the SR. In the figure it is shown that most of the JES uncertainties are one-sided, and their combination grows smoothly from $\sim 1.5\%$ to $\sim 2\%$ as p_T^{recoil} increases up to 1000 GeV, becoming more relevant at the tail where it raises up to almost 4%.

JVT EFFICIENCY

The uncertainty on the efficiency of the JVT requirement, used to reject pile-up jets, is estimated from simulated $Z(\mu\mu) + \text{jets}$ events and ranges from 1% to 2% [152]. This uncertainty has only a relevant impact in the Top CR, where its contribution to the total uncertainty ranges between values of about 0.23% and 0.1% across the p_T^{recoil} spectrum.

FLAVOUR TAGGING

Uncertainties on the flavour-tagging efficiency are included in those regions where b -jets are used in the selection, i.e. the $W \rightarrow \mu\nu$, $W \rightarrow e\nu$ and Top CRs. Five NPs are included: the uncertainty on the b -jets tagging efficiency, the c -tag efficiency, the mistag rate of light-flavour jets and the uncertainties on the extrapolation of the b - and c -jets tagging efficiency to high- p_T regimes. Figures 7.4a and 7.4b

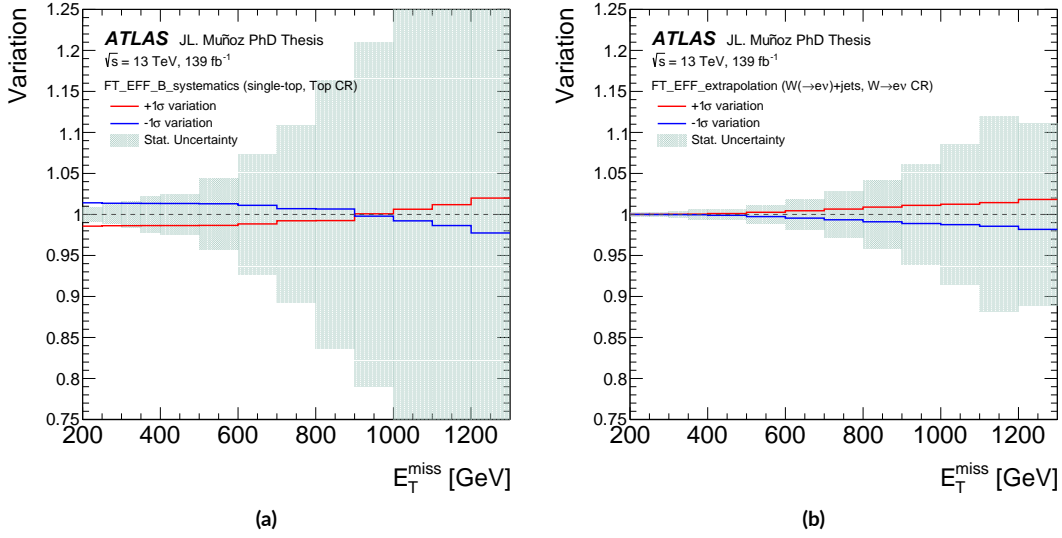


Figure 7.4: (a) b -tagging efficiency uncertainty for the single- t sample in the Top CR; (b) uncertainty on the extrapolation of the b -tagging efficiency to high- p_T regimes for the $W(e\nu) + \text{jets}$ sample in the $W \rightarrow e\nu$ CR. The shaded area denotes the statistical uncertainty of the samples.

show the impact of the b -tagging uncertainty on the $W(e\nu) + \text{jets}$ sample in the $W \rightarrow e\nu$ region and the uncertainty on the high- p_T extrapolation of the b jets tagging efficiency on the single- t sample in the Top CR, respectively. Both are relatively small and exhibit a smooth shape across the p_T^{recoil} bins, varying between values close to 0% and 2%. These systematics combined have a total impact in the Top region between 1.8% and 2.8%.

7.2.4 E_T^{MISS} TRACK SOFT TERM

Uncertainties related to the E_T^{miss} scale and resolution due to the track-based soft contributions to the E_T^{miss} calculation are derived by comparing data and simulated $Z \rightarrow \mu\mu$ events with exactly 0 jets [164]. In these kind of topologies, the momentum of the Z boson approximates the *hard* component of the E_T^{miss} , $\mathbf{p}_T^{\text{hard}}$, defined as the scalar sum of the p_T of the leptons, photons and jets participating in the E_T^{miss} calculation. Then, the parallel and perpendicular projections of the soft term, $\mathbf{p}_T^{\text{soft}}$ (defined by $E_T^{\text{miss}} = -(\mathbf{p}_T^{\text{hard}} + \mathbf{p}_T^{\text{soft}})$), onto $\mathbf{p}_T^{\text{hard}}$ are used to evaluate such uncertainties. The uncertainty on the energy scale is derived from the parallel projection, while the uncertainty on the energy resolution is evaluated separately in the longitudinal direction, for which the parallel projection is used, and the transverse direction, using the perpendicular projection instead. Therefore, three NPs are included in the fit to account for these uncertainties. The E_T^{miss} track soft term uncertainties range between

$\sim 0.52\%$ and less than 0.1% in the SR.

7.2.5 LEPTONS AND PHOTONS

Different sources of uncertainties related to electrons, muons, taus and photons are considered in the analysis, concerning different experimental aspects such as uncertainties on the energy scale and resolution measurements, or uncertainties on the efficiency on the identification, reconstruction and definition of the physical objects.

ENERGY/MOMENTUM SCALE AND RESOLUTION

Two nuisance parameters are included to account for the uncertainties on the energy scale and resolution of electrons and photons, which are measured by comparing the invariant mass distributions of real and simulated $Z \rightarrow ee$ samples [136]. These uncertainties are smaller than 1% in all regions, giving the largest contribution in the $W \rightarrow e\nu$ CR where the uncertainty on the e/γ energy scale (EG_SCALE_ALL) raises from $\sim 0.5\%$ at $p_T^{\text{recoil}} = 200$ GeV up to $\sim 0.8\%$ for $p_T^{\text{recoil}} \geq 700$ GeV.

Regarding muons, one NP covers the uncertainty on the momentum scale calibration and two independent NPs account for the different track resolution uncertainties from the ID and MS systems. In addition, uncertainties on the *sagitta bias* charge-dependent correction to the momentum scale are included. This is correction applied offline to account for displacements of the reconstructed hits in the bending plane orthogonal to the muon track, resulting in a charge-asymmetric alteration of the track curvature. Such uncertainties are covered by including two NPs: one to account for variations in the momentum scale based on a combination of corrections on combined ID and MS measurements; and a second one to cover variations based on the residual charge-dependent bias before the corrections are applied. These uncertainties are only relevant in the $W \rightarrow \mu\nu$ CR, being the MUON_SAGITTA_RESBIAS the largest one, with contributions ranging from less than 1% at low p_T^{recoil} to 1.8% at $p_T^{\text{recoil}} = 1200$ GeV.

EFFICIENCY SCALE FACTORS

As explained in Chapter 2, efficiency SFs are applied at event-level to correct for mismodelling in the reconstruction, identification, isolation or triggering of leptons and photons. Uncertainties related to these kind of experimental aspects are propagated via such SFs in the regions where the specific particle is selected. For those regions where the presence of leptons or photons is vetoed the uncertainties are propagated via the anti-SFs, which are re-calculated (following the procedure described in App. B)

for every systematic variation.

Uncertainties on the electron reconstruction, isolation and triggering efficiencies are included via individual NPs. Since different electron identification WPs are used for selected and vetoed electrons, the treatment of the uncertainty on the electron identification varies between selections. For regions where at least one electron is required (i.e. the $W \rightarrow e\nu$, $Z \rightarrow ee$ and Top CRs), the electron identification uncertainties are treated as uncorrelated with respect to the uncertainties applied in the other regions. Moreover, in these regions the ATLAS SIMPLIFIED systematics scheme is adopted, consisting in the utilization of 18 uncorrelated uncertainties in bins of $p_T \times \eta$ plus 16 correlated uncertainties across $p_T \times \eta$ bins. In regions where electrons are vetoed, the TOTAL scheme is adopted instead, resulting in a single NP. The electron efficiency uncertainties have a small impact ($< 0.1\%$) in the SR but reach up to 5.2% in the $Z \rightarrow ee$ CR, and are largely dominated by the uncertainties on the identification efficiency. Figure 7.5a shows one of the correlated uncertainties in the electron identification efficiency for the $Z(ee) + \text{jets}$ sample in the $Z \rightarrow ee$ region, where it grows from values close to 0% at low p_T^{recoil} up to almost 2% at the tail of the distribution.

The uncertainty on the photon identification efficiency is included via a single NP in the fit, with negligible effect in all regions.

Uncertainties on the muon reconstruction efficiency are evaluated separately for muons with p_T below and above 15 GeV, and they are included splitting the statistical and systematical component of each one, resulting in 4 NPs in total to cover the uncertainty on the reconstruction efficiency. Similarly, uncertainties on the track-to-vertex association are included separately for the statistical and systematical components, with 2 NPs. Additionally, a set of 15 uncorrelated NPs defined in bins of muon p_T and η are included to account for MC mismodelling on high- p_T muons with the Medium identification WP. Combined, the muon efficiency scale factor uncertainties have an impact between 0.5% and 3% in the $W \rightarrow \mu\nu$ CR, and between 1% and 15% in the $Z \rightarrow \mu\mu$ CR, with the uncertainties on the modelling of high- p_T Medium muons dominating at high p_T^{recoil} while the low and medium p_T^{recoil} regimes are governed by the uncertainty on the muon reconstruction. The later is shown in Figure 7.5b for the $W(\mu\nu) + \text{jets}$ sample in the $W \rightarrow \mu\nu$ region, ranging between values of ~ 1 and $\sim 2\%$ across the p_T^{recoil} spectrum.

Uncertainties on the reconstruction efficiency of hadronically-decaying taus are taken into account with a single NP, plus an additional uncertainty added for the uncertainty when reconstructing taus with $p_T > 100$ GeV. Uncertainties on the $\tau_{\text{had-vis}}$ identification are treated as follows. To cover

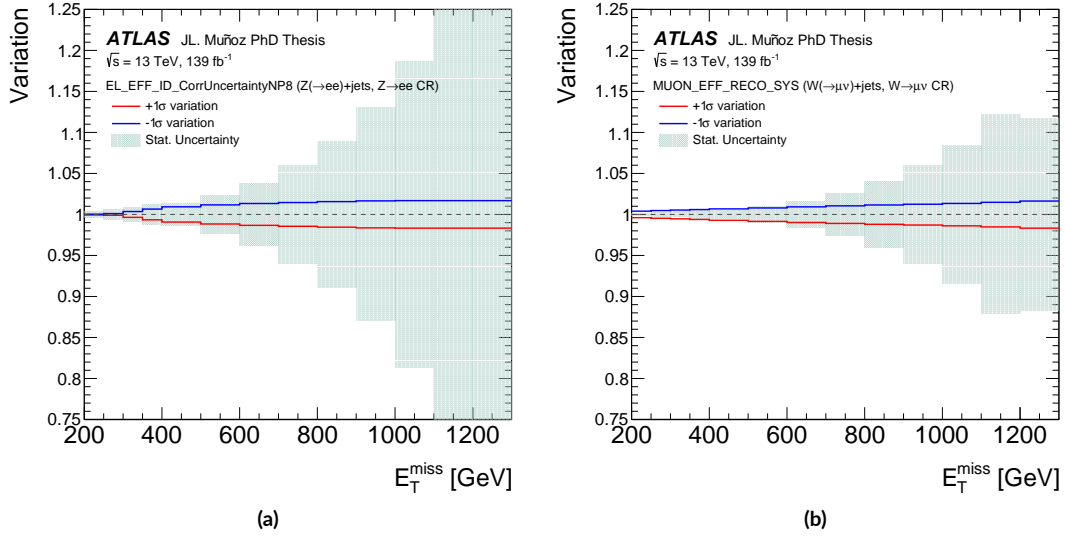


Figure 7.5: (a) Electron identification efficiency uncertainty for the $Z(ee) + \text{jets}$ sample in the $Z \rightarrow ee$ CR; (b) Muon reconstruction efficiency uncertainty for the $W(\mu\nu) + \text{jets}$ sample in the $W \rightarrow \mu\nu$ CR. The shaded area denotes the statistical uncertainty of the samples.

the uncertainty on the discrimination against electrons one NP is used. Uncertainties on the discrimination against jets are included separately for 1-prong taus, for which 4 NPs defined in bins of tau- p_T are used, 3-prong taus, using two NPs for different ranges in p_T , taus with $p_T > 100$, for which an additional NP is included, and, finally, systematical effects, which are covered by a single NP. Since tau-leptons are vetoed in all regions in the analysis, these systematic uncertainties are only propagated via the anti-SFs, which have a globally negligible effect (see App. B) and so do the tau uncertainties.

7.3 THEORETICAL UNCERTAINTIES

Theoretical uncertainties are those related to the modelling of the physical processes involved in the analysis.

7.3.1 BACKGROUND PROCESSES

Uncertainties on the simulation of the varied background processes are detailed below. A summary of the nuisance parameters used in the fit can be found in Table 7.3.

V +JETS CORRECTIONS

The $p_T(V)$ -based higher-order corrections applied to the V +jets samples ensure that a consistent set of theoretical uncertainties can be used to cover simultaneously any potential mismodelling in W +jets and Z +jets processes. Uncertainties related to such corrections, discussed in Chapter 2, are derived by introducing the following set of nuisance parameters:

- `vjets_d1K_NNLO`, assigned to the uncertainty associated to the truncation of the perturbative expansion in α_S . This uncertainty is estimated by varying, separately and simultaneously, the QCD renormalization and factorization scales by factors of 0.5 and 2, extracting the changes in the differential cross-section in bins of $p_T(V)$. Then, the center of the resulting band is taken as the nominal value and the half of its width as the systematic uncertainty.
- `vjets_d2K_NNLO` covers the uncertainties in the shape of the $p_T(V)$ distribution, which are relevant for the extrapolation of low- p_T measurements to high- p_T regimes. This uncertainty is estimated by a shape distortion of the scale uncertainty, parametrized as a function of $p_T(V)$ between 200–2000 GeV as $(p_T^2 - p_{T,0}^2) / (p_T^2 + p_{T,0}^2)$, with $p_{T,0} = 650$ GeV.
- `vjets_d3K_NNLO` is introduced to account for the uncertainty on the correlation among the different processes, i.e. it covers the residual differences in QCD corrections between W +jets and Z +jets processes. This uncertainty is estimated from the difference in QCD NNLO κ -factors with respect to Z +jets production.
- `vjets_d1kappa_EW` accounts for unknown higher-order Sudakov logarithms at high- p_T .
- `vjets_d2kappa_EW` covers the uncertainty of possible NNLO EW effects not included in the corrections. Since the universality is not demonstrated for EW corrections, this uncertainty is treated as uncorrelated across processes and therefore three NPs are included, corresponding to the W +jets, $Z(\rightarrow \ell\ell)$ +jets and $Z(\nu\nu)$ + jets processes. These uncertainties are estimated via a conservative approach by taking the 5% of the full NLO EW correction.
- `vjets_d3kappa_EW` accounts for the uncertainty on the limited precision of the Sudakov approximation at two loops and is treated as uncorrelated between W + and Z +jets processes, resulting into two independent NPs. These uncertainties are estimated as the difference between the NLL Sudakov approximation and the exponentiation of the full NLO EW correction.
- `vjets_dK_NNLO_mix` represents the uncertainty on the mixed QCD-EW corrections, which is assumed to be proportional to the difference between the additive and multiplicative combination of QCD and EW corrections.

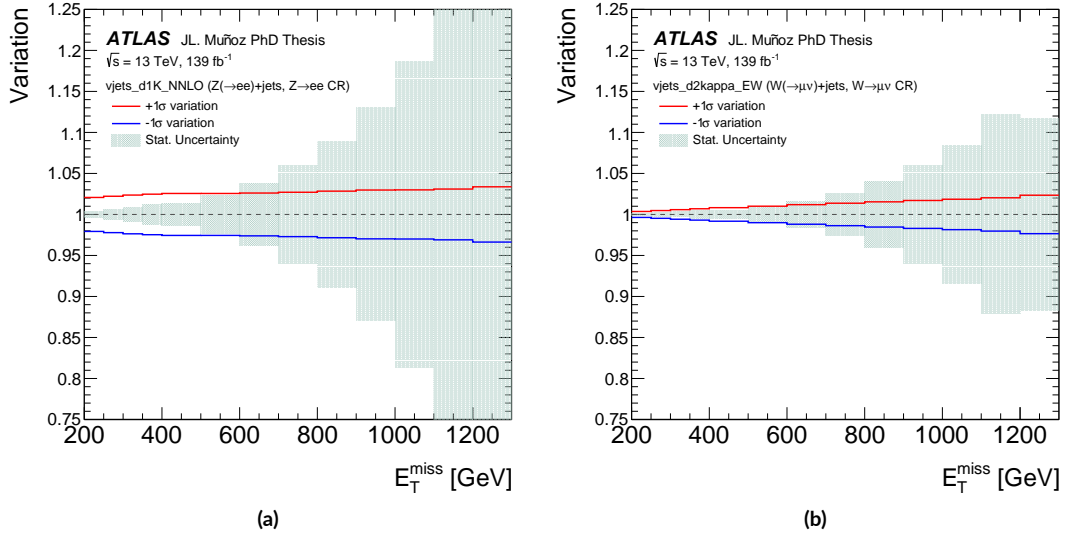


Figure 7.6: (a) QCD normalization correction uncertainty for the $Z(ee) + \text{jets}$ sample in the $Z \rightarrow ee$ CR; (b) Uncertainty on EW additional NNLO corrections not included for the $W(\mu\nu) + \text{jets}$ sample in the $W \rightarrow \mu\nu$ CR. The shaded area denotes the statistical uncertainty of the samples.

- `vjets_dK_PDF` accounts for the uncertainty related to the different PDF choice. It is estimated by the sum in quadrature of the 107 independent PDFs provided by the PDF set `LUXqed_plus_PDF4LHC15_nnlo` [17].
- An additional uncertainty, represented by the NP `taus_PTV_definition`, is introduced to account for possible caused by the different definition of τ -leptons used in the MC sample at truth level with respect to the one used in the theoretical calculations from Ref. [106].

All these uncertainties are treated as independent and correlated across p_T^{recoil} bins and V +jets processes, with the exceptions of the EW components `vjets_d2kappa_EW` and `vjets_d3kappa_EW`, already mentioned.

Table 7.2 summarizes the overall impact pre-fit of the systematic uncertainties related to the V +jets correction in the SR and in the CRs dominated by W/Z +jets processes. The QCD-related uncertainties `vjets_d1K_NNLO` and `vjets_d1K_NNLO` dominate largely across the p_T^{recoil} spectrum in all those regions (with impacts of 1.8% and 1.5%, respectively), followed by the uncertainty on the PDF choice (0.7%), which plays an important role at high- p_T^{recoil} regimes. Figures 7.6a and 7.6b show two examples of the size of such uncertainties: `vjets_d1K_NNLO` in the $Z \rightarrow ee$ region on the $W(e\nu) + \text{jets}$ sample, ranging between values of 2% and $\sim 3.5\%$; and `vjets_d2kappa_EW_W` on the $W(\mu\nu) + \text{jets}$ sample in the $W \rightarrow \mu\nu$ CR, growing from less than 0.1% up to about 2% with p_T^{recoil} .

VBF V +JETS

Theoretical uncertainties related to the simulation of the VBF initiated V +jets processes are evaluated using an alternative set of MC samples generated with Sherpa. Uncertainties on the scale and PDF are taken into account, together with an additional uncertainty for residual differences caused by the different generators. Combined, these uncertainties have an impact in the SR that varies from $\sim 0.1\%$ at $p_T^{\text{recoil}} = 200$ GeV to $\sim 6\%$ for p_T^{recoil} above 1.2 TeV.

DIBOSON

Uncertainties in the modelling of the diboson background include: variations in the QCD scale, which are estimated similarly as for the V +jets samples, i.e. varying the factorization and renormalization scales in seven combinations and extracting the systematic uncertainty from the resulting envelope as half of its width; uncertainties on the PDF choice, extracted from a combination of 100 NNPDF replicas; and uncertainties related to the modelling of the parton shower. The later is estimated as follows. Variations of the resummation (QSF) and the matching (CKKW) scales are evaluated by generating MC samples with 3 leptons in the final state. A region enriched in WZ is then constructed by applying the pre-selection criteria and requiring $2e$ and 1μ or $1e$ and $2mu$ in the final state. The systematic impacts are extracted as the difference in yields with respect to the nominal sample.

The quadratic sum of the diboson uncertainties described above is implemented in the fit as a single NP, giving an impact in the SR before the fit between $\sim 0.1\%$ in the bin EMO and $\sim 2.5\%$ in EM12.

TOP-QUARK PRODUCTION

Systematic uncertainties related to the simulation of top-production events are treated separately for the single- t and $t\bar{t}$ processes. Uncertainties on the parton shower modelling are estimated by using two MC samples generated both with Powheg but using Pythia8 in one of the samples for the PS modelling and Herwig7 in the other one, extracting the systematic variation from the differences between the two samples. Similarly, uncertainties related to the initial- and final-state soft gluon radiation are evaluated by using dedicated MC samples, varying the corresponding parameters in the simulation and extracting the final variation by comparing predictions. This method is described in detail in Ref. [172]. The degree of interference between single- t in the Wt -channel and $t\bar{t}$ when using the DR and DS schemes, described in Ref. [173], is also taken into account.

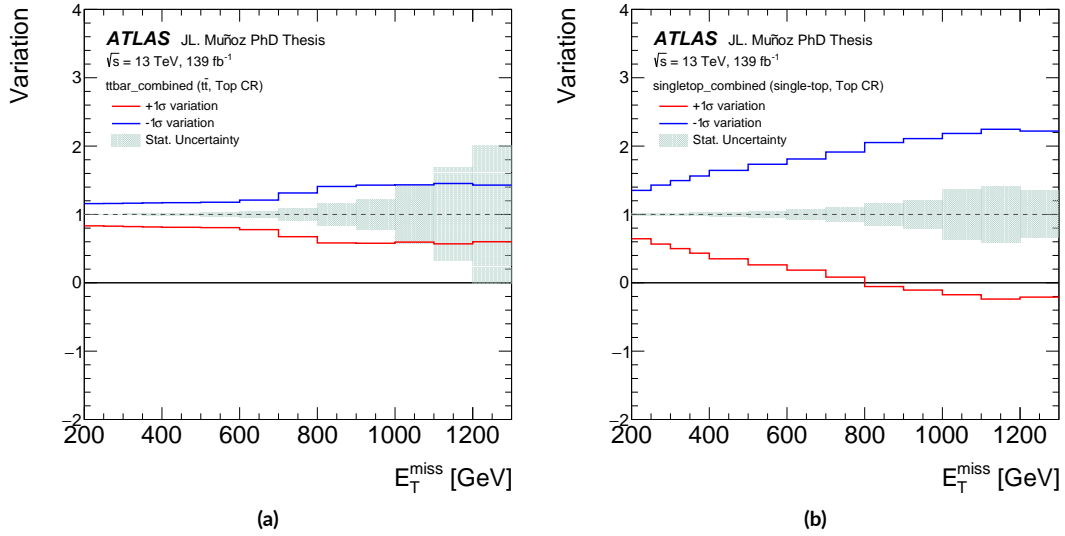


Figure 7.7: Top-production theoretical uncertainties, split into $t\bar{t}$ (a) and single- t (b) processes, in the Top CR. The shaded area denotes the statistical uncertainty of the samples.

All uncertainties described above are treated as uncorrelated. The several systematic variations are summed in quadrature and, finally, included in the fit as two independent NPs, one for each top-process. These uncertainties are shown in the Top CR for the $t\bar{t}$ and single- t samples in Figures 7.7a and 7.7b, respectively. Both are relatively large compared to the other systematic uncertainties previously discussed: the $t\bar{t}$ uncertainty varies between $\sim 16\%$ and $\sim 45\%$ and the single- t one ranges between $\sim 35\%$ – $\sim 124\%$ in the Top region, being the later the largest systematic uncertainty in the analysis. The overall contribution to the total background uncertainty in the SR varies between 1% and 12% across the p_T^{recoil} spectrum in the case of $t\bar{t}$, and rises from 4% at $p_T^{\text{recoil}} = 200$ GeV up to $\sim 60\%$ for $p_T^{\text{recoil}} > 1200$ GeV. The contribution of top-production events to the SR is however small compared to other SM processes, and therefore the impact of these uncertainties in the fit is expected to be reduced after the fit.

MULTI-JET AND NON-COLLISION BACKGROUNDS

The QCD multi-jet and NCB estimations are not derived from simulations, as explained in the previous Chapter. Therefore there is no uncertainty associated to the modelling of such contributions. Instead, a conservative uncertainty of 100% is assigned to both backgrounds, translating into a combined uncertainty of 1.2% at $p_T^{\text{recoil}} = 200$ GeV and less than 0.1% for $p_T^{\text{recoil}} > 400$ GeV.

NP name	Short description	
lumiSys	Uncertainty on the total integrated luminosity (1.7% [128])	
PRW_DATASF	Uncertainty on the pile-up reweighting scale factor computation	
JET_EffectiveNP_Detector (2 NPs)	Jet energy scale uncertainties, split into 15 components	
JET_EffectiveNP_Mixed (3 NPs)		
JET_EffectiveNP_Modelling (4 NPs)		
JET_EffectiveNP_Statistical (6 NPs)		
JET_EtaIntercalibration		JES of forward jets wrt. central jets, split into 5 components
JET_Flavor_Response		Jet flavour-related uncertainties
JET_Flavor_Composition		
JET_BJES		
JET_Pileup		Pile-up uncertainties (5 NPs for p_T , η , N_{PV} , and $\langle\mu\rangle$ dependence)
JET_PunchThrough_MC16		Jet punch-through uncertainty
JET_SingleParticle_HighPt	Jet absolute in-situ propagation of single-particle uncertainty	
JET_FullJER	Jet energy resolution uncertainty, split into 13 components	
JET_JvtEfficiency	JVT efficiency uncertainty	
FT_EFF_B_systematics	Jet b -tagging uncertainty	
FT_EFF_C_systematics	Jet c -tagging uncertainty	
FT_EFF_Light_systematics	Jet light-flavour tagging uncertainty	
FT_EFF_extrapolation	Uncertainty on high- p_T extrapolation of flavour-tagging efficiency	
MET_SoftTrk_ResoPerp	E_T^{miss} track soft term transverse resolution uncertainty	
MET_SoftTrk_ResoPara	E_T^{miss} track soft term longitudinal resolution uncertainty	
MET_SoftTrk_Scale	E_T^{miss} track soft term scale uncertainty	
EL_EFF_Trigger_TOTAL_1NPCOR_PLUS_UNCOR	Electron trigger efficiency uncertainty	
EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	Electron reconstruction efficiency uncertainty	
EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR	Electron isolation efficiency uncertainty	
EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR	Electron identification efficiency uncertainty (no- e CRs)	
EL_EFF_ID_SIMPLIFIED_CorrUncertainty	16 correlated electron identification efficiency uncertainties (e CRs)	
EL_EFF_ID_SIMPLIFIED_UncorrUncertainty	18 uncorrelated electron identification efficiency uncertainties (e CRs)	
EG_SCALE_ALL	Electron/photon energy scale uncertainty	
EG_RESOLUTION_ALL	Electron/photon energy resolution uncertainty	
PH_EFF_ID_Uncertainty	Photon identification efficiency uncertainty	
MUON_EFF_RECO_STAT	Muon reconstruction and identification uncertainties ($p_T > 15$ GeV)	
MUON_EFF_RECO_SYS		
MUON_EFF_RECO_STAT_LOWPT	Muon reconstruction and identification uncertainties ($p_T < 15$ GeV)	
MUON_EFF_RECO_SYS_LOWPT		
MUON_EFF_TTVA_STAT	Muon track-to-vertex association efficiency uncertainties	
MUON_EFF_TTVA_SYS		
MUON_SCALE	Muon energy scale uncertainty	
MUON_SAGITTA_RHO	Muon momentum scale variations (charge dependent)	
MUON_SAGITTA_RESBIAS	Muon momentum scale variations (charge dependent) – symmetrized	
MUON_ID	Muon energy resolution uncertainty from inner detector	
MUON_MS	Muon energy resolution uncertainty from muon system	
MUON_EFF_ID_HighPT	Uncertainty for high- p_T muons mismodelling (15 NPs)	
TAUS_TRUEHADTAU_EFF_RECO_TOTAL	Tau reconstruction efficiency uncertainty	
TAUS_TRUEHADTAU_EFF_RECO_HIGHPT	Tau reconstruction efficiency uncertainty for $p_T > 100$ GeV	
TAUS_TRUEHADTAU_EFF_ELEOLR_TOTAL	Uncertainty on tau identification against electrons	
TAUS_TRUEHADTAU_EFF_JETID_1PRONGSTATSYSTUNCORR2025	1-prong tau identification uncertainty for $p_T \in [20, 25]$ GeV	
TAUS_TRUEHADTAU_EFF_JETID_1PRONGSTATSYSTUNCORR2530	1-prong tau identification uncertainty for $p_T \in [25, 30]$ GeV	
TAUS_TRUEHADTAU_EFF_JETID_1PRONGSTATSYSTUNCORR3040	1-prong tau identification uncertainty for $p_T \in [30, 40]$ GeV	
TAUS_TRUEHADTAU_EFF_JETID_1PRONGSTATSYSTUNCORRGE40	1-prong tau identification uncertainty for $p_T > 40$ GeV	
TAUS_TRUEHADTAU_EFF_JETID_3PRONGSTATSYSTUNCORR2030	3-prong tau identification uncertainty for $p_T \in [20, 30]$ GeV	
TAUS_TRUEHADTAU_EFF_JETID_3PRONGSTATSYSTUNCORRGE30	3-prong tau identification uncertainty for $p_T > 30$ GeV	
TAUS_TRUEHADTAU_EFF_JETID_HIGHPT	Tau identification uncertainty for $p_T > 100$ GeV	
TAUS_TRUEHADTAU_EFF_JETID_SYST	Tau identification uncertainty	

Table 7.1: Summary of the experimental systematic uncertainties considered in the analysis.

	SR	$W \rightarrow \mu\nu$	$Z \rightarrow \mu\mu$	$W \rightarrow e\nu$	$Z \rightarrow ee$
vjets_d1K_NNLO [%]	1.83–2.77	1.72–2.54	1.94–2.82	1.69–2.44	1.97–2.82
vjets_d2K_NNLO [%]	1.53–1.54	1.44–1.49	1.63–1.66	1.41–1.42	1.65–1.64
vjets_d3K_NNLO [%]	0.15–0.20	0.37–1.14	0.00–0.00	0.37–1.10	0.00–0.00
vjets_d1kappa_EW [%]	0.01–1.26	0.01–1.56	0.01–1.19	0.01–1.49	0.01–1.18
vjets_d2kappa_EW_W [%]	0.13–0.31	0.32–1.81	0.00–0.00	0.33–1.75	0.00–0.00
vjets_d2kappa_EW_Z11 [%]	0.00–0.01	0.01–0.01	0.31–1.49	0.00–0.00	0.32–1.49
vjets_d2kappa_EW_Zvv [%]	0.11–1.23	0.00–0.00	0.00–0.00	0.00–0.00	0.00–0.00
vjets_d3kappa_EW_W [%]	0.01–0.10	0.01–0.59	0.00–0.00	0.01–0.57	0.00–0.00
vjets_d3kappa_EW_Z [%]	0.09–1.58	0.00–0.02	0.03–2.08	0.00–0.00	0.04–2.07
vjets_dK_NNLO_mix [%]	0.28–1.25	0.48–2.51	0.23–0.98	0.48–2.42	0.23–0.98
vjets_dK_PDF [%]	0.72–2.87	0.75–2.96	0.73–3.09	0.74–2.83	0.74–3.10
taus_PTV_definition [%]	0.05–0.21	0.01–0.16	0.00–0.00	0.01–0.11	0.00–0.00

Table 7.2: Impact pre-fit of the V +jets correction-related uncertainties, expressed in %, in the p_T^{recoil} exclusive bins EMO and EM12 in the regions SR, $W \rightarrow \mu\nu$, $Z \rightarrow \mu\mu$, $W \rightarrow e\nu$ and $Z \rightarrow ee$.

NP name	Short description
vjets_d1K_NNLO	Uncertainty on truncation of the expansion in α_S
vjets_d2K_NNLO	Uncertainty on $p_T(V)$ shape and high- p_T extrapolation
vjets_d3K_NNLO	Differences in QCD corrections between W +jets and Z +jets
vjets_d1kappa_EW	Uncertainty on Sudakov logarithms beyond NNLO
vjets_d2kappa_EW_W	Additional possible EW NNLO effects (split into 3 NPs)
vjets_d2kappa_EW_Z11	
vjets_d2kappa_EW_Zvv	
vjets_d3kappa_EW_W	Uncertainty due to limitations in the Sudakov approximation (2 NPs)
vjets_d3kappa_EW_Z	
vjets_dK_NNLO_mix	Interference term between QCD and EW corrections
vjets_dK_PDF	PDF-related uncertainties, combined in a single NP
taus_PTV_definition	Uncertainty covering different τ definitions effects
VBF_theo_sys	Uncertainty on the modelling of the VBF V +jets processes
diboson_Sys	Uncertainty on the modelling of the diboson process
singletop_singletop_combined_Sys	Uncertainty on the modelling of the single- t process
ttbar_ttbar_combined_Sys	Uncertainty on the modelling of the $t\bar{t}$ process
NCB_Sys	Uncertainty on the non-collision background
multijet_Sys	Uncertainty on the multi-jet background

Table 7.3: Summary of the theoretical uncertainties considered for the background samples in the analysis.

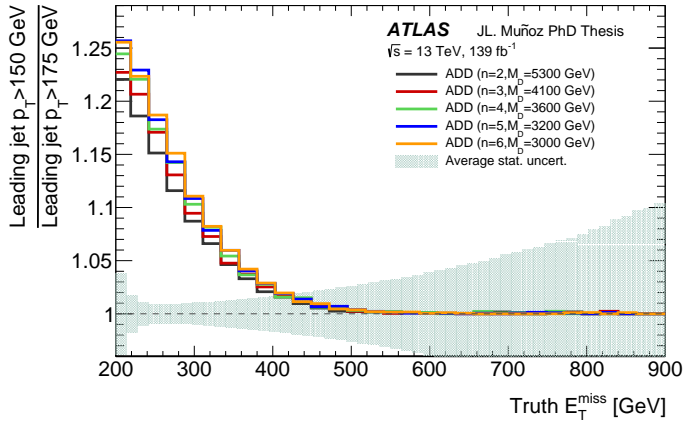


Figure 7.8: Ratio of events with leading jet $p_T > 150 \text{ GeV}$ over events with leading jet $p_T > 175 \text{ GeV}$, as a function of truth-level E_T^{miss} , for ADD samples simulated with $n = 2, 3, 4, 5$ and 6 extra dimensions. The shaded band denotes the average statistical uncertainty of the MC samples

7.3.2 SIGNAL PROCESSES

Several sources of uncertainty in the predicted signal events are considered separately for every model of new physics. For these uncertainties, there is a differentiation made between the variation caused in the total cross-section ($\Delta\sigma$), given by the change on the total number of signal events of the specific model, and the effect in the signal acceptance (ΔA), defined by the variation of the ratio of events determined at truth-level in a particular SR E_T^{miss} bin divided by the total number of events. The estimation of the systematical uncertainties corresponding to the ADD, DM production, compressed SUSY and invisible-decaying Higgs models are detailed below. Theoretical uncertainties are estimated for the ALPs and DE models following similar procedures.

ADD LARGE EXTRA DIMENSIONS

Uncertainties for the ADD model include variations in the PDF, renormalization and factorization scales, parton shower tuning and initial- and final-state radiation. This set of systematic uncertainties is the same that what was used in the previous version of the analysis [49], which was already a re-calculation of the uncertainties used in the previous iteration [174]. Since the ADD samples are generated at LO, no parton matching scale uncertainty is considered. The samples are produced with a truth-level cut on the transverse momentum of the KK graviton at 150 GeV , for this reason the E_T^{miss} distribution is biased for values below 400 GeV , as shown in Figure 7.8. Therefore, only the region with E_T^{miss} above 400 GeV is considered in the analysis for this model.

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
$\Delta\sigma$ [%]	11	18	27	35	43
ΔA (EM ₄) [%]	8	11	12	13	13
ΔA (EM ₅) [%]	13	13	15	11	14
ΔA (EM ₆) [%]	16	16	18	20	8
ΔA (EM ₇) [%]	18	21	19	15	20
ΔA (EM ₈) [%]	21	19	20	12	9
ΔA (EM ₉) [%]	20	22	21	15	19
ΔA (EM ₁₀) [%]	32	26	28	30	26
ΔA (EM ₁₁) [%]	32	26	28	30	26
ΔA (EM ₁₂) [%]	32	26	28	30	26

Table 7.4: Systematic uncertainties on PDFs for the ADD model, expressed in %, separated in the acceptance (A) and cross-section (σ) components. The uncertainties are computed as the envelop that contains the signal yields from the three PDF families plus their error bands.

The PDF uncertainties affect both the ADD cross-section (normalization of the sample) and the signal acceptance. For the evaluation of these uncertainties, two sources of uncertainties are then considered: intra-PDF uncertainty, which is the corresponding uncertainty within a specific PDF set; and the inter-PDF uncertainty, covering the relative variations when replacing one PDF set by another. Three different sets of PDFs are considered: the NNPDF2.3 set [96] (which is the one used for the generation of the ADD samples), the CT₁₀ set at LO [175] and the MMHT₂₀₁₄ [176] set. Instead of generating new samples for each PDF family, the events are re-weighted in the original sample as if they would have had generated with an alternative PDF. This, and the estimation of the intra-PDF uncertainties is done by using the LHAPDF method [177]. The inter-PDF uncertainties are evaluated by following the recommendations from the PDF₄LHC group [100]. The final PDF uncertainty is the envelope that contains the error bands of the three PDF families. These uncertainties, which are shown in Table 7.4 separated into their cross-section and signal acceptance components, range between 8%–13% in the E_T^{miss} bin EM₄ and 26%–32% in EM₁₂, depending on the model.

The uncertainty on the renormalization and factorization scales is estimated by varying the corresponding parameters by factors of 2 and 0.5 at truth level in different MC samples. The final uncertainty is the average of the up and down variations. This uncertainty affects only to the total cross-section and the results, collected in Table 7.5, vary from 23% to 36% as the number of extra dimensions considered increases from 2 to 6.

Uncertainties on the initial and final state radiation are estimated as follows. Five tune parameters

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
$\Delta\sigma$ [%]	23	27	30	33	36

Table 7.5: Systematic uncertainty on the factorization and renormalization scales for the ADD model, expressed in %. The scales are varied up and down simultaneously, the final uncertainty is the average of these variations.

	EM ₄	EM ₅	EM ₆	EM ₇	EM ₈	EM ₉	EM ₁₀	EM ₁₁	EM ₁₂
$\Delta\mathcal{A}$ [%]	7	7	10	13	18	13	9	9	9

Table 7.6: Initial and final state radiation uncertainties of the ADD model, expressed in % and given in exclusive bins of E_T^{miss} . The final value is a common envelope valid for all the ADD models between $n = 2$ –6 dimensions.

are varied in order to account for uncertainties from underlying event effects, jet structure effects and those aspects of the MC generation that might provide extra-jet production. For each ADD model, ten systematic samples are produced and analyzed at truth-level, and $\Delta\mathcal{A}$ is then evaluated in the different E_T^{miss} exclusive bins. The final uncertainty in each bin, as shown in Table 7.6, is a common envelope valid for the different extra dimensions models ($n = 2$ to 6), and ranges between 7% and 18% across the E_T^{miss} spectrum.

DM PRODUCTION

Scale-related uncertainties are estimated for the DM production models by calculating the acceptance bins of E_T^{miss} , using the usual set of 7 variations of the renormalization and factorization parameters, $[\mu_R, \mu_F] = [0.5-2, 0.5-2]$. The uncertainty for each variation is then calculated as $100 \times (A - A_0)/A_0$, where A_0 is $\mu_R = \mu_F = 1.0$. The high-variation is then taken as the largest variation in the positive direction, and the low-variation as the most negative one. Regarding the axial-vector mediator model, the estimated variations range between 0.1% and 6.5%, increasing as a function of E_T^{miss} with the exception of two samples: $(m_\chi, m_A) = (1, 10 \text{ and } 50) \text{ GeV}$, for which larger uncertainties are obtained (8–20% and 4–12%, respectively). Globally, higher masses result in lower scale uncertainty for all E_T^{miss} ranges, although mediator mass appears to have a larger effect overall. In the pseudo-scalar mediator scenario, these uncertainties vary from 0.4% to 21% across the E_T^{miss} spectrum, with higher mediator masses translating into lower variations.

The PDF uncertainty is estimated via measuring the variation of the acceptance in each bin of E_T^{miss} for 100 different PDF sets, via internal weights in the nominal sample. The PDF uncertainty is then evaluated as the standard deviation of the resulting acceptances (ΔA^{PDF}), divided by the acceptance given by the nominal PDF choice, $\Delta A^{\text{PDF}}/A_0$. Statistical fluctuations due to the limited

	Merging	Renorm/Factor	ISR/FSR	Intra-PDF	Inter-PDF
$\Delta\mathcal{A}$ [%]	7.0	2.8	7.6	1.0	5.0

Table 7.7: Uncertainties on the merging scale, renormalization and factorization scales, initial- and final-state radiation modelling and PDF choice for the SUSY signal samples, expressed in %, as used in the analysis.

statistical power of the MC samples are reduced by performing a linear fit. For the axial-vector mediator scenario this uncertainty ranges between 0.1% and about 10%, increasing with E_T^{miss} . There is no significant dependence on either mediator or WIMP mass. For the pseudo-scalar mediator model, the PDF-related uncertainties vary from 0.2% to 20%, increasing as a function of E_T^{miss} . In this case, higher DM masses result in lower PDF uncertainties.

The ISR/FSR modeling uncertainty is estimated via varying the corresponding parameters in dedicated generated samples, and summing in quadrature the relative variations obtained in order to extract a conservative full coverage of these uncertainties. In the DMA scenario, the estimated uncertainty sizes about 3%, while in the DMP scenario it is of about 6%. The estimated impact of the varied scale and PDF systematic uncertainties on the DMA (DMP) predicted cross-section is of 10% ($^{+50\%}_{-30\%}$) and 5% (20%).

SUSY QUARK-PAIR PRODUCTION

Theoretical uncertainties on the SUSY signal models considered (SS, BB, TT and T₄body) include variations in the merging scale, the renormalization/factorization scales, initial/final state radiation (ISR/FSR) modeling and PDF. These uncertainties are included as a single value for each source for all E_T^{miss} bins, given a general lack of trend in E_T^{miss} observed during the calculation, and only the largest of the up or down variations is kept in each case in order to symmetrize the results. These uncertainties are evaluated with the same methods as described for the ADD model. During the calculation process, it was noted that the variations obtained were mostly independent with respect to the model or the mass of the squark of the sample used, and therefore it was decided to use one single sample (TT sample with masses (700, 693) GeV) and then apply the resulting variations to the rest of the samples. The obtained uncertainties are shown in Table 7.7, ranging between 1% and 8%, depending on the source.

Uncertainties on the predicted cross-section are provided by calculations at approximate NNLO+NNLL accuracy [103, 104], including uncertainties on the PDF and α_S , as a function of the mass of the squark as shown in Figures 7.9a and 7.9b. These uncertainties range between 7% for a mass of 100 GeV and about 11% for a mass of about 1 TeV.

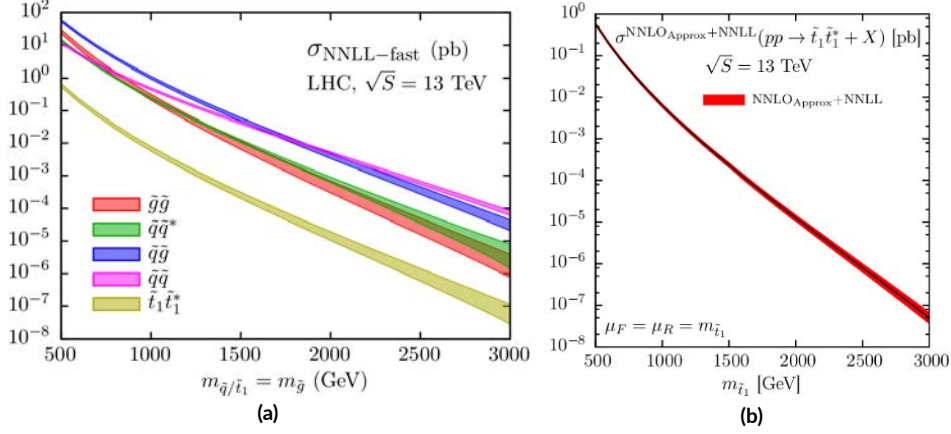


Figure 7.9: Cross-section predictions for squark and gluino (a) and stop-antistop (b) production at the LHC with $\sqrt{s} = 13$ TeV at NNLO_{Approx}+NNLL accuracy. The error bands denote the theoretical uncertainty due to scale variation and the pdf+ α_S error, as described in Refs. [103, 104].

INVISIBLE-DECAYING HIGGS

Theoretical uncertainties in the $b \rightarrow \text{inv}$ model include variations on the PDF, renormalization/factorization scales, parton shower modelling and higher-order EW corrections.

PDF-related uncertainties are evaluated using 32 event weights generated from different PDF sets. The standard deviation of the resulting envelope is taken as the uncertainty. These uncertainties, which affect both the acceptance and the cross-section, translate into variations between 0.4% and 0.8% in the final yields as E_T^{miss} increases, as shown in Table 7.8.

Parton shower uncertainties are estimated by varying the corresponding tune parameters in the Pythia8 simulator, which are stored as internal weights. The obtained uncertainties, shown in Table 7.8 vary between 3% and 9% as E_T^{miss} increases.

Scale-variations uncertainties are evaluated by varying the corresponding Pythia8 parameters in 7 steps, as done for the other models, and systematic uncertainties are estimated from the difference in the resulting acceptance with respect to the nominal one in bins of E_T^{miss} for each variation of the parameters. In addition, a linear fit is performed to smooth statistical fluctuations and then the uncertainties are symmetrized. The final uncertainty is extracted as the sum in quadrature of all the variations. The obtained uncertainties are of about 10% in the E_T^{miss} spectrum.

Uncertainties in the EW corrections are particularly important for VBF and VH processes. Since the contribution from VH is small ($\sim 8\%$), only corrections for the VBF process are considered. The uncertainty related to this process is obtained using the HAWK program [178], and it is parametrized

	PDF [%]	PS [%]	EW [%]
EM ₀	0.38	3.00	1.40
EM ₁	0.39	3.30	1.70
EM ₂	0.39	3.60	2.10
EM ₃	0.41	3.90	2.50
EM ₄	0.45	4.20	3.00
EM ₅	0.47	4.70	3.80
EM ₆	0.50	5.30	4.60
EM ₇	0.55	6.00	5.40
EM ₈	0.60	6.60	6.20
EM ₉	0.67	7.60	6.90
EM ₁₀	0.68	7.90	7.90
EM ₁₁	0.70	8.30	8.50
EM ₁₂	0.76	8.80	10.00

Table 7.8: PDF, parton shower modelling and EWK correction uncertainties for the invisible-decaying Higgs model.

as a function of $p_T(H)$. The results translate into uncertainties in the signal yield that vary between 1.4% and 10% with increasing E_T^{miss} , as shown in Table 7.8.

8

Results and Interpretations

The results of the Run 2 monojet analysis are presented in this Chapter. Details of the background determination via the background-only fit are shown first, followed by a discussion of the obtained results in the SR plus several interpretations in terms of limits to the existence of new phenomena.

8.1 BACKGROUND-ONLY FIT

As discussed in Section 6.3, five CRs are used to constrain the V +jets and top backgrounds in the SR via three floating normalization factors that are extracted from a simultaneous fit to the data. In the fit, a total 97 nuisance parameters are included to constrain the systematic uncertainties.

8.1.1 FIT PARAMETERS

The fitted κ -factors associated to the V +jets, single- t and $t\bar{t}$ processes are shown in Table 8.1. The fitted NPs are shown in Figure 8.1, and tables with their exact values are included in Appendix D. A

κ -factor	Fitted value
κ^V	1.13 ± 0.01
κ^t	1.6 ± 0.4
$\kappa^{\bar{t}\bar{t}}$	0.9 ± 0.1

Table 8.1: Normalization factors for the V +jets and top production processes obtained from the CRs simultaneous fit.

visualization of the correlations among the different NPs is shown in Fig. 8.2, where only uncertainties that exhibit any correlation coefficient above $|0.2|$ are included.

As expected, given the original $\sim 15\%$ offset between data and predictions in regions dominated by V +jets events (see Section 6.4), a value of 1.13 ± 0.01 is obtained for κ^V , resolving the normalization discrepancy. The fitted values for κ^t and $\kappa^{\bar{t}\bar{t}}$ are 1.6 ± 0.4 and 0.9 ± 0.1 , respectively. The larger value obtained for the single- t normalization might seem a bit unexpected given the good agreement pre-fit, but this is due to an interplay between κ^t and the NP associated to the single-top theoretical uncertainty (named ‘singletop_single-top_combined_Sys’), which is fitted to a value of $-1.57^{+0.29}_{-0.46}$. This means that the total yield is reduced by a factor of about 2.2σ of this uncertainty, which ranges between 35%–124% in the Top CR. This compensates the effect of the large value of κ^t and the good agreement MC/data is recovered.

Most of the fitted NPs are compatible with 0 and their uncertainties with 1 (in units of RMS), therefore most of them have not a significant influence in the fit. There are, however, some parameters whose fitted values are pulled away from 0, but always within the $\pm 1\sigma$ band (excepting the case of the top theory systematic, already discussed), and their uncertainties are below 1σ , so they are constrained in the fit. Regarding the experimental uncertainties, the NPs mostly constrained are the ones corresponding to the uncertainty on the muon reconstruction efficiency (slightly pulled but still consistent with 0 within uncertainties), the e/γ scale uncertainty and one of the JES uncertainty components. The JET Flavor Response and Composition NPs are slightly constrained as well, since the quark and gluon composition are not completely optimized in the analysis. The uncertainty on the scale of the momentum of the muons is also constrained.

In the theoretical uncertainties, both of the two top-production systematics are constrained because of their very large pre-fit variations in the high-boosted regime. The single- t NP is the one pulled the most among all parameters. The VBF V +jets theoretical uncertainty is constrained and largely pulled (it is fitted to a value of ~ 0.95) because it is estimated from SHERPA samples gener-

ated at LO and then applied to the nominal HERWIG samples, generated at NLO. Regarding the NPs related to the V +jets higher-order corrections, the parameter `vjets_d3K_NL0`, associated to the uncertainty on the non-universality of QCD corrections across Z +jets, W +jets and γ +jets processes, is fitted to a value of -0.9 and its uncertainty is constrained below $|0.7|$.

A correlation between the electron and muon systematic uncertainties is observed, expected because of the use of CRs to estimate all the V +jets processes simultaneously with a unique κ -factor. The normalization factors κ^t and $\kappa^{\bar{t}\bar{t}}$ are anti-correlated, because they both are extracted from the same CR. The κ^t factor is correlated with the single- t theory uncertainty, because of the interplay between these two factors before discussed, and with the $\bar{t}\bar{t}$ uncertainty, while $\kappa^{\bar{t}\bar{t}}$ is found to be significantly anti-correlated with both uncertainties. There is a correlation between the three κ -factors and the uncertainty on the luminosity, expected since they all play a part in the total background normalization. The Jet Flavor Response and Composition NPs are correlated as well. The parameters `vjets_d1K_NNL0` and `vjets_d2K_NNL0`, related to the uncertainties on the V +jets re-weighting, are anti-correlated and correlated, respectively, with the V +jets global normalization factor κ^V , as expected since these components account for the QCD scale normalization and shape uncertainties. This means that an up variation of κ^V requires a compensation of `vjets_d1K_NNL0` in the opposite direction. On the other hand, such variation on `vjets_d2K_NNL0` would induce a reduction of events below $p_T^{\text{recoil}} = 600$ GeV, and since that is the region that dominates the fit this implies the correlation observed.

Globally the fit behaves as expected, the simultaneous shape fit constrains some of the uncertainties that were dominating before the fit by using the information from the exclusive bins in p_T^{recoil} .

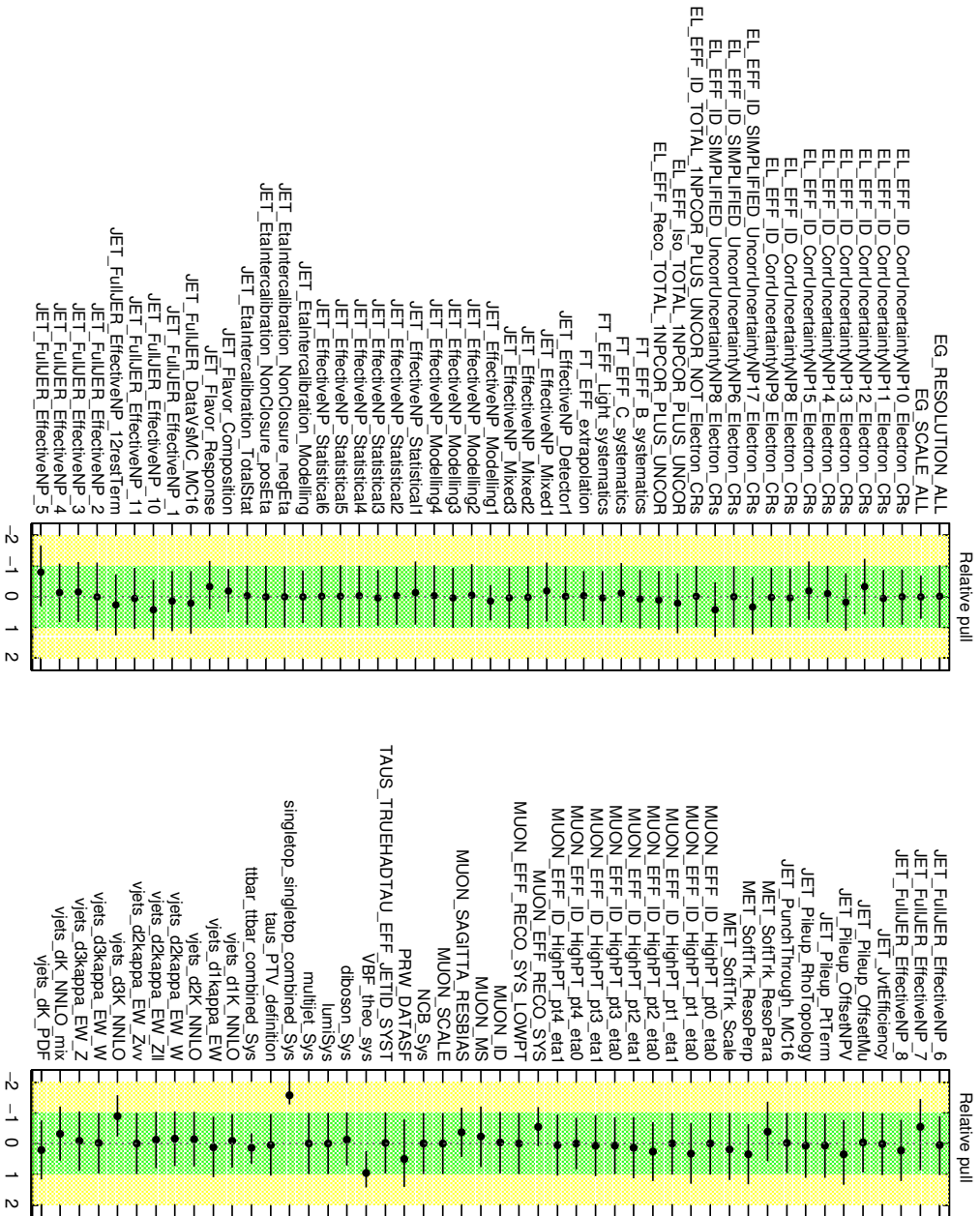


Figure 8.1: Fitted nuisance parameters in the background-only fit.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	1364958	699674	225606	196800	145531
Background post-fit prediction	$1\,364\,800 \pm 3\,300$	$699\,700 \pm 1\,800$	$225\,590 \pm 1\,100$	$196\,960 \pm 600$	$145\,500 \pm 500$
$W(\mu\nu) + \text{jets}$	$1\,148\,700 \pm 6\,000$	–	$31\,600 \pm 1\,700$	–	–
$W(e\nu) + \text{jets}$	–	$578\,800 \pm 3\,400$	$16\,070 \pm 900$	–	–
$W(\tau\nu) + \text{jets}$	$71\,480 \pm 800$	$45\,170 \pm 500$	$3\,380 \pm 180$	–	–
$Z(\mu\mu) + \text{jets}$	$21\,520 \pm 500$	–	778 ± 20	$185\,200 \pm 900$	–
$Z(ee) + \text{jets}$	–	–	–	–	$138\,140 \pm 700$
$Z(\tau\tau) + \text{jets}$	–	$1\,896 \pm 50$	–	–	–
Diboson	$23\,000 \pm 4\,000$	$12\,400 \pm 1\,900$	$1\,880 \pm 340$	$4\,000 \pm 700$	$2\,900 \pm 500$
single- t	$22\,100 \pm 6\,000$	$13\,200 \pm 3\,500$	$33\,000 \pm 10\,000$	350 ± 170	109 ± 60
$t\bar{t}$	$51\,800 \pm 4\,000$	$34\,000 \pm 2\,800$	$136\,700 \pm 9\,000$	$4\,050 \pm 400$	$1\,790 \pm 180$
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	$3\,320 \pm 400$	$2\,530 \pm 320$
VBF $W(\ell\nu) + \text{jets}$	$26\,200 \pm 3\,100$	$14\,300 \pm 1\,700$	$2\,020 \pm 340$	–	–
Background pre-fit prediction	$1\,216\,000 \pm 60\,000$	$623\,000 \pm 32\,000$	$233\,000 \pm 31\,000$	$175\,000 \pm 9\,000$	$127\,400 \pm 7\,000$
$W(\mu\nu) + \text{jets}$	$1\,014\,000 \pm 50\,000$	–	$28\,000 \pm 2\,300$	–	–
$W(e\nu) + \text{jets}$	–	$509\,000 \pm 27\,000$	$14\,200 \pm 1\,200$	–	–
$W(\tau\nu) + \text{jets}$	$63\,000 \pm 3\,400$	$39\,800 \pm 2\,100$	$2\,970 \pm 250$	–	–
$Z(\mu\mu) + \text{jets}$	$18\,900 \pm 1\,100$	–	689 ± 23	$162\,900 \pm 9\,000$	–
$Z(ee) + \text{jets}$	–	–	–	–	$120\,100 \pm 7\,000$
$Z(\tau\tau) + \text{jets}$	–	$1\,680 \pm 60$	–	–	–
Diboson	$23\,400 \pm 4\,000$	$12\,600 \pm 2\,300$	$1\,930 \pm 400$	$4\,100 \pm 800$	$2\,960 \pm 600$
single- t	$16\,100 \pm 6\,000$	$9\,800 \pm 4\,000$	$28\,000 \pm 13\,000$	680 ± 500	280 ± 210
$t\bar{t}$	$59\,600 \pm 7\,000$	$39\,000 \pm 5\,000$	$155\,000 \pm 27\,000$	$4\,600 \pm 1\,300$	$2\,000 \pm 700$
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	$2\,730 \pm 500$	$2\,040 \pm 400$
VBF $W(\ell\nu) + \text{jets}$	$21\,600 \pm 4\,000$	$11\,700 \pm 2\,300$	$1\,500 \pm 500$	–	–

Table 8.2: Predicted (pre- and post-fit) and observed background events in the control regions in the inclusive p_T^{recoil} bin IMO.

8.1.2 EVENT YIELDS AND POST-FIT DISTRIBUTIONS IN THE CRs

The total event yields obtained for every background in each CR are listed for the inclusive selection $p_T^{\text{recoil}} > 200$ GeV, before and after the fit, in Table 8.2. The p_T^{recoil} fitted distributions in the CRs are presented in Figure 8.3, and several post-fit distributions for some other kinematic variables (namely, leading jet p_T and $|\eta|$, jet multiplicity, sub-leading jet p_T , sub-sub-leading jet p_T , transverse mass and invariant mass) are shown in Figs. 8.4–8.9. A closer look to the outcome of the fit is given in in Tables 8.3, 8.4, 8.5 and 8.6, where the exact values of the fitted yields are shown in the exclusive p_T^{recoil} regions EM₀, EM₄, EM₈ and EM₁₂. The tables corresponding with the regions EM₁, EM₂, EM₃, EM₅, EM₆, EM₇, EM₉, EM₁₀ and EM₁₁ are included in Appendix D. Contributions from $Z(\nu\nu) + \text{jets}$ are omitted in the tables and figures as they are negligible in the CRs. The quoted uncertainties include all systematic uncertainties and the statistical uncertainty after the fit.

Overall, good agreement between data and predictions is observed in all regions. The total uncertainty is dominated by systematics in most of the $p_{\text{T}}^{\text{recoil}}$ range with the exception of the higher regime, where the statistical component becomes more relevant. Moreover, a strong reduction of the background uncertainty in all regions is noted, as shown in Table 8.2, where the relative background uncertainties are reduced up to less than 1% after the fit.

Quantitatively, the goodness of the fit is tested by computing the corresponding p -values for the background-only hypothesis using the profile likelihood described by Eq. (6.2). Values of 0.49, 0.73, 0.96, 0.75 and 0.53 are obtained correspondingly for the regions $W \rightarrow \mu\nu$, $W \rightarrow e\nu$, Top, $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$, thus reinforcing the conclusion of good agreement.

The post-fit jet multiplicity distributions (Fig. 8.6) show that the description of the data is still not optimal. The MC predictions seem to overestimate the number of jets per event than actually observed. This might be a remaining effect of the V +jets re-weighting, which was already discussed at the pre-fit stage in Chapter 6, but it could also be pointing to a potential mismodelling in the jet fragmentation and the parton shower, which is actually a known feature of SHERPA that has been observed in other analyses. There is also some discrepancy observed in the p_{T} distributions of the second and third jets (Figs. 8.7 and 8.8), likely connected to the jet multiplicity mismodelling. Apart from this feature, the leading jet p_{T} and $|\eta|$ distributions (most sensitive to this monojet analysis) are well described. The lepton-neutrino transverse mass is well modelled as well in the one-lepton CRs. There are however small remaining discrepancies ($< 10\%$) at the tails of the di-lepton invariant mass distributions for both $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ CRs. These features were already present before the fit, and might be pointing to a non-perfect modelling of the electron and muon kinematics in these regions.

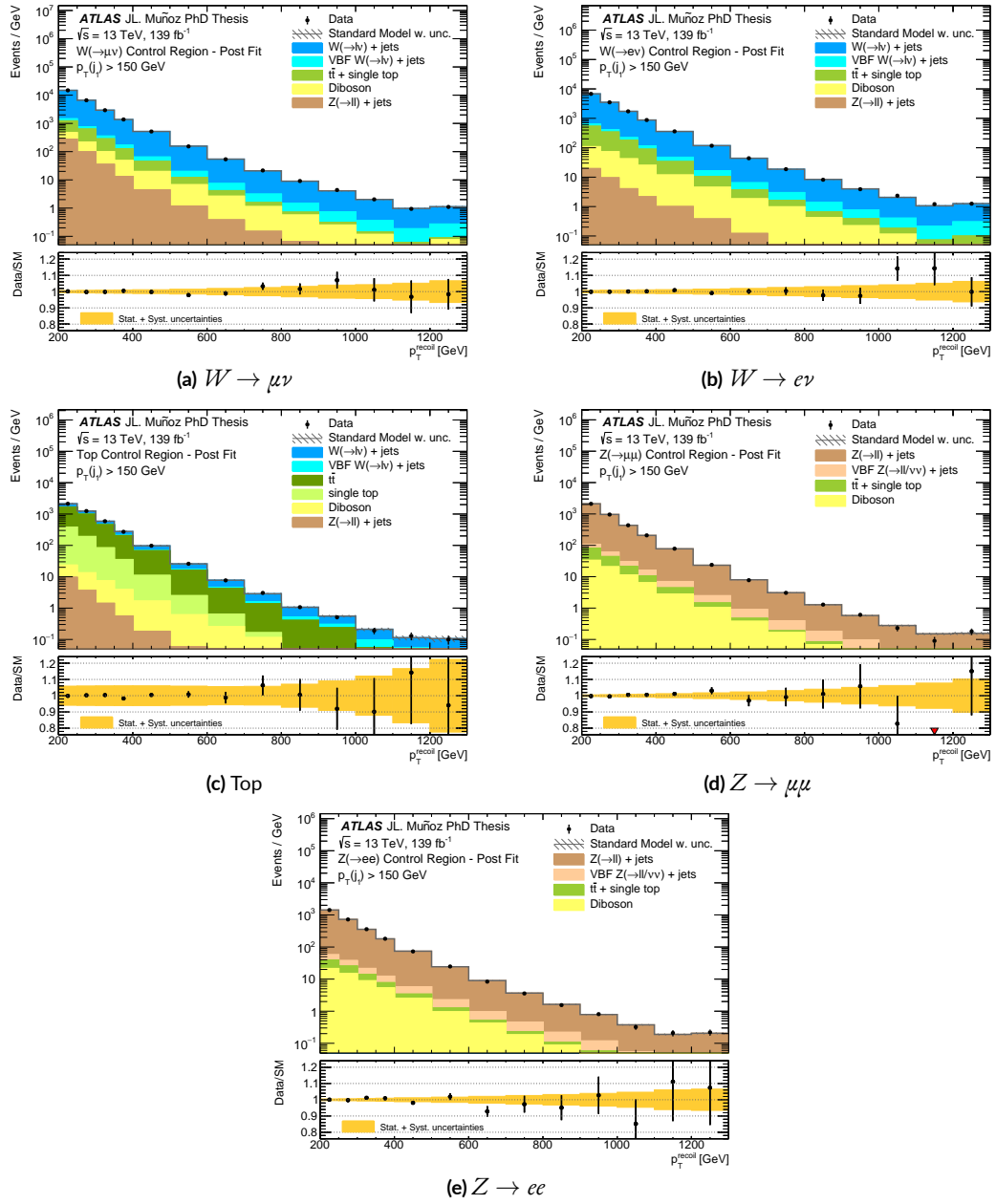


Figure 8.3: p_T^{recoil} post-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panels show the ratio of data over background prediction.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	742740	338998	106425	105155	71075
Background post-fit prediction	741 600±1 900	339 330±920	106 520±610	105 400±330	71 050±270
$W(\mu\nu) + \text{jets}$	631 300±3 000	–	15 260±870	–	–
$W(e\nu) + \text{jets}$	–	285 900±1 500	6 720±380	–	–
$W(\tau\nu) + \text{jets}$	38 520±430	22 050±220	1 544±89	–	–
$Z(\mu\mu) + \text{jets}$	13 640±370	–	473±12	100 100±420	–
$Z(ee) + \text{jets}$	–	–	–	–	68 220±310
$Z(\tau\tau) + \text{jets}$	–	966±25	–	–	–
Diboson	9 600±1 400	4 410±650	690±110	1 640±270	1 060±170
single- t	12 500±3 000	6 800±1 600	17 900±4 500	200±86	39±23
$t\bar{t}$	25 000±2 100	14 300±1 200	63 300±4 200	2 150±170	846±71
VBF $Z(\ell\ell/\nu\nu)+\text{jets}$	–	–	–	1 310±140	887±99
VBF $W(\ell\nu)+\text{jets}$	11 000±1 200	4 900±510	712±98	–	–

Table 8.3: Predicted (post-fit) and observed background events in the control regions, in the exclusive $p_{\text{T}}^{\text{recoil}}$ bin EM0.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	51748	36025	9749	7822	7278
Background post-fit prediction	51 840±180	35 720±130	9 706±76	7 736±42	7 415±50
$W(\mu\nu) + \text{jets}$	42 680±390	–	1 591±83	–	–
$W(e\nu) + \text{jets}$	–	28 750±280	1 065±54	–	–
$W(\tau\nu) + \text{jets}$	2 812±39	2 267±31	198.3±9.2	–	–
$Z(\mu\mu) + \text{jets}$	434.6±6.4	–	17.92±0.50	7 040±68	–
$Z(ee) + \text{jets}$	–	–	–	–	6 843±69
$Z(\tau\tau) + \text{jets}$	–	98.5±2.8	–	–	–
Diboson	1 520±270	1 070±180	147±31	278±53	249±45
single- t	600±220	420±160	920±430	11.7±7.6	4.2±2.9
$t\bar{t}$	1 870±160	1 750±150	5 580±390	145±20	83±14
VBF $Z(\ell\ell/\nu\nu)+\text{jets}$	–	–	–	261±39	235±34
VBF $W(\ell\nu)+\text{jets}$	1 910±270	1 370±200	181±38	–	–

Table 8.4: Predicted (post-fit) and observed background events in the control regions, in the exclusive $p_{\text{T}}^{\text{recoil}}$ bin EM4.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	911	817	107	129	155
Background post-fit prediction	896 ± 16	836 ± 14	106.4 ± 5.8	127.7 ± 3.9	163.0 ± 3.4
$W(\mu\nu) + \text{jets}$	696 ± 20	–	26.7 ± 2.0	–	–
$W(e\nu) + \text{jets}$	–	646 ± 18	24.3 ± 1.8	–	–
$W(\tau\nu) + \text{jets}$	52.5 ± 1.7	53.3 ± 1.8	4.17 ± 0.28	–	–
$Z(\mu\mu) + \text{jets}$	6.51 ± 0.20	–	–	109.8 ± 4.3	–
$Z(ee) + \text{jets}$	–	–	–	–	141.3 ± 4.0
$Z(\tau\tau) + \text{jets}$	–	1.690 ± 0.070	–	–	–
Diboson	49 ± 12	39.6 ± 8.7	4.8 ± 1.7	6.8 ± 1.8	8.7 ± 2.1
single- t	1.0 ± 1.0	2.2 ± 2	–	–	–
$t\bar{t}$	14.1 ± 1.7	21.9 ± 2.5	35.1 ± 4.9	1.57 ± 0.34	1.87 ± 0.41
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	9.5 ± 2.2	11.1 ± 2.6
VBF $W(\ell\nu) + \text{jets}$	78 ± 17	72 ± 16	11.2 ± 4.8	–	–

Table 8.5: Predicted (post-fit) and observed background events in the control regions, in the exclusive $p_{\text{T}}^{\text{recoil}}$ bin EM8.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	110	125	10	18	22
Background post-fit prediction	111.8 ± 4.5	125.2 ± 5.1	10.6 ± 2.1	15.6 ± 1.6	20.49 ± 0.93
$W(\mu\nu) + \text{jets}$	78.9 ± 4.3	–	2.99 ± 0.41	–	–
$W(e\nu) + \text{jets}$	–	88.9 ± 4.8	2.87 ± 0.40	–	–
$W(\tau\nu) + \text{jets}$	6 ± 0.37	6.53 ± 0.39	–	–	–
$Z(\mu\mu) + \text{jets}$	0.760 ± 0.040	–	–	12.5 ± 1.4	–
$Z(ee) + \text{jets}$	–	–	–	–	16.61 ± 0.92
$Z(\tau\tau) + \text{jets}$	–	–	–	–	–
Diboson	6.8 ± 2.2	3.6 ± 1.1	–	1.21 ± 0.44	1.47 ± 0.53
single- t	–	–	–	–	–
$t\bar{t}$	1.13 ± 0.23	6.1 ± 1.0	0.87 ± 0.17	–	–
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	1.89 ± 0.64	2.41 ± 0.82
VBF $W(\ell\nu) + \text{jets}$	18.2 ± 5.7	20.0 ± 6.1	3.3 ± 2.3	–	–

Table 8.6: Predicted (post-fit) and observed background events in the control regions, in the exclusive $p_{\text{T}}^{\text{recoil}}$ bin EM12.

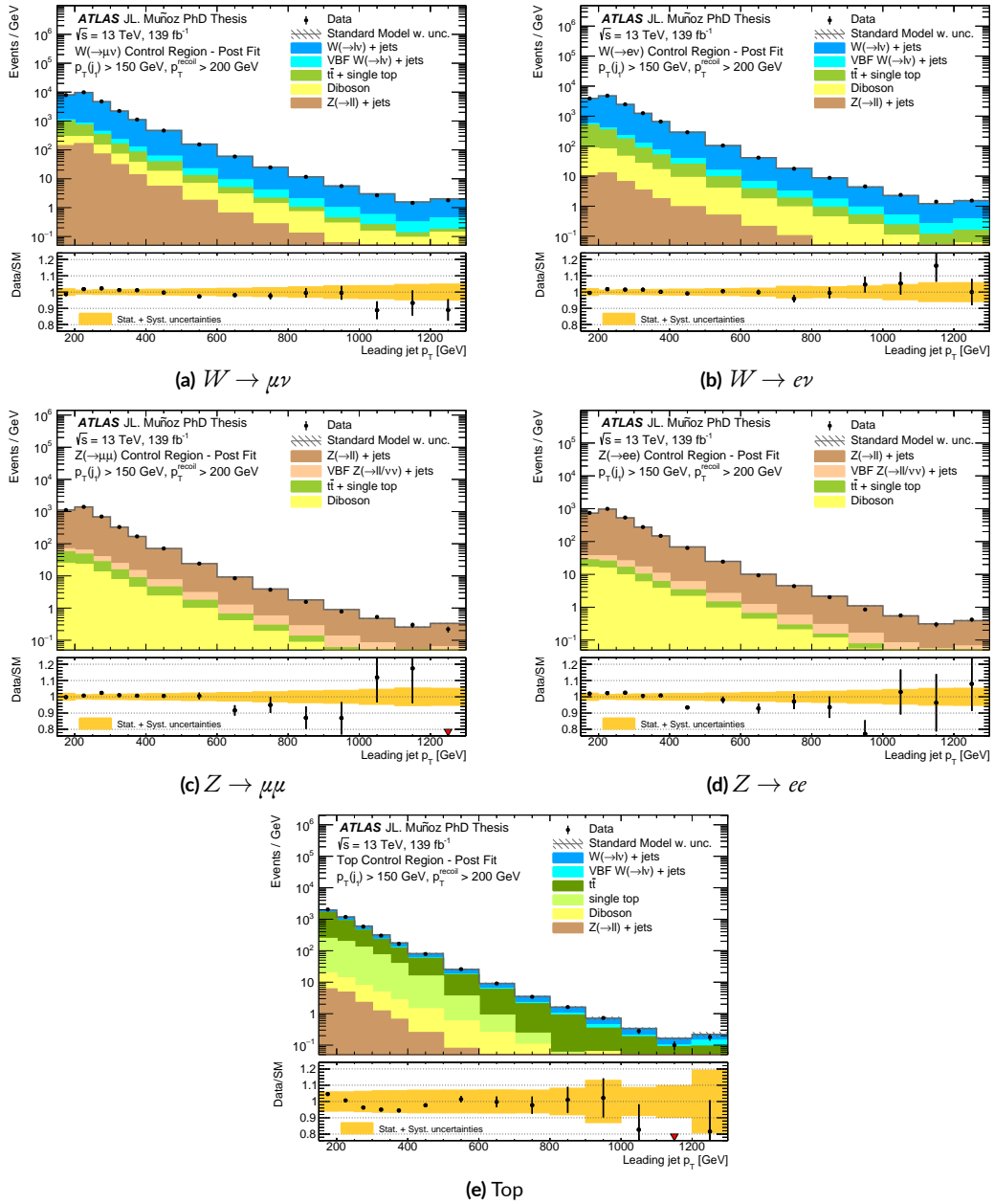


Figure 8.4: Leading jet p_T post-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panels show the ratio of data over background prediction.

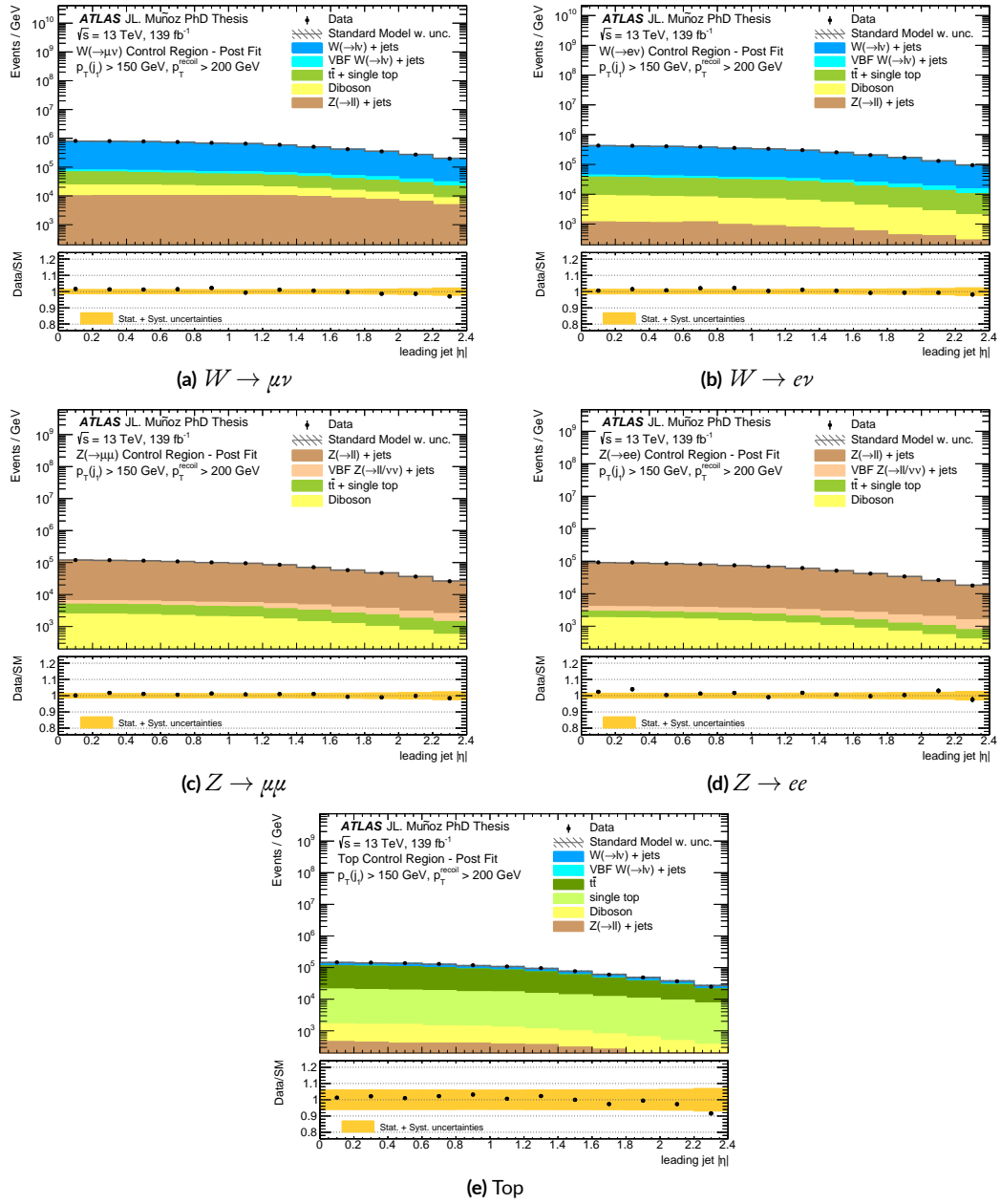


Figure 8.5: Leading jet $|\eta|$ post-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panels show the ratio of data over background prediction.

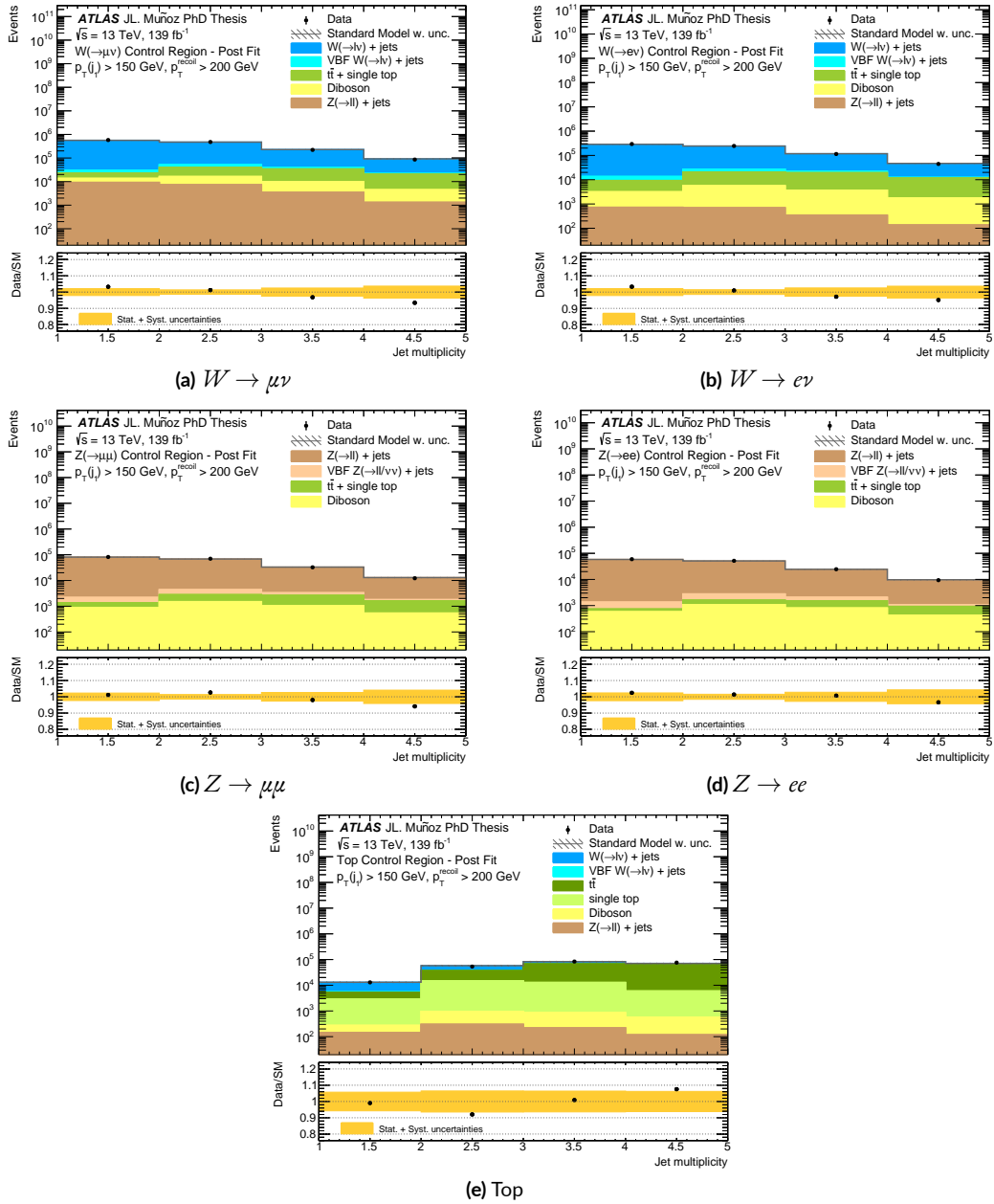


Figure 8.6: Jets multiplicity post-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panels show the ratio of data over background prediction.

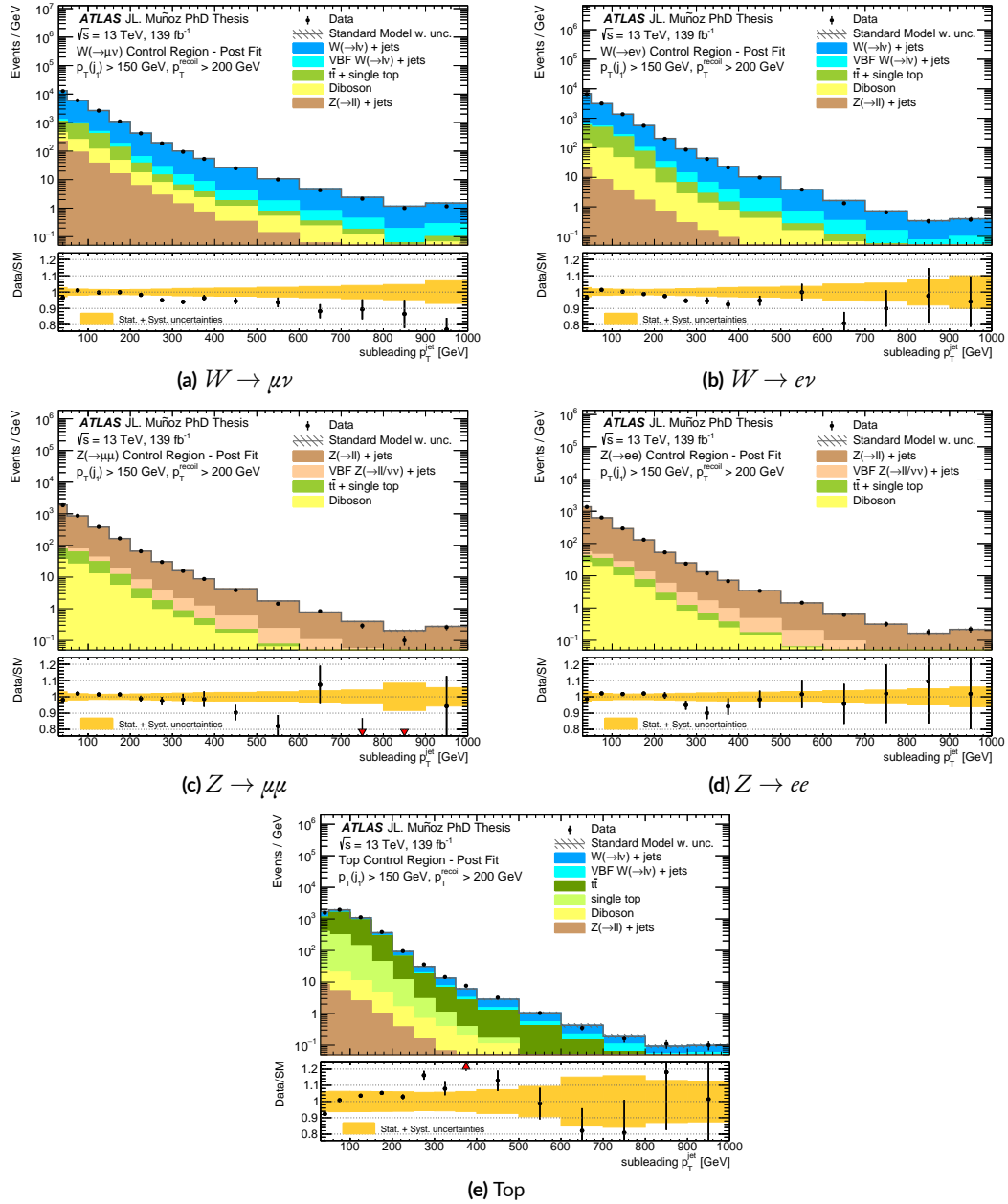


Figure 8.7: Sub-leading jet p_T post-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panels show the ratio of data over background prediction.

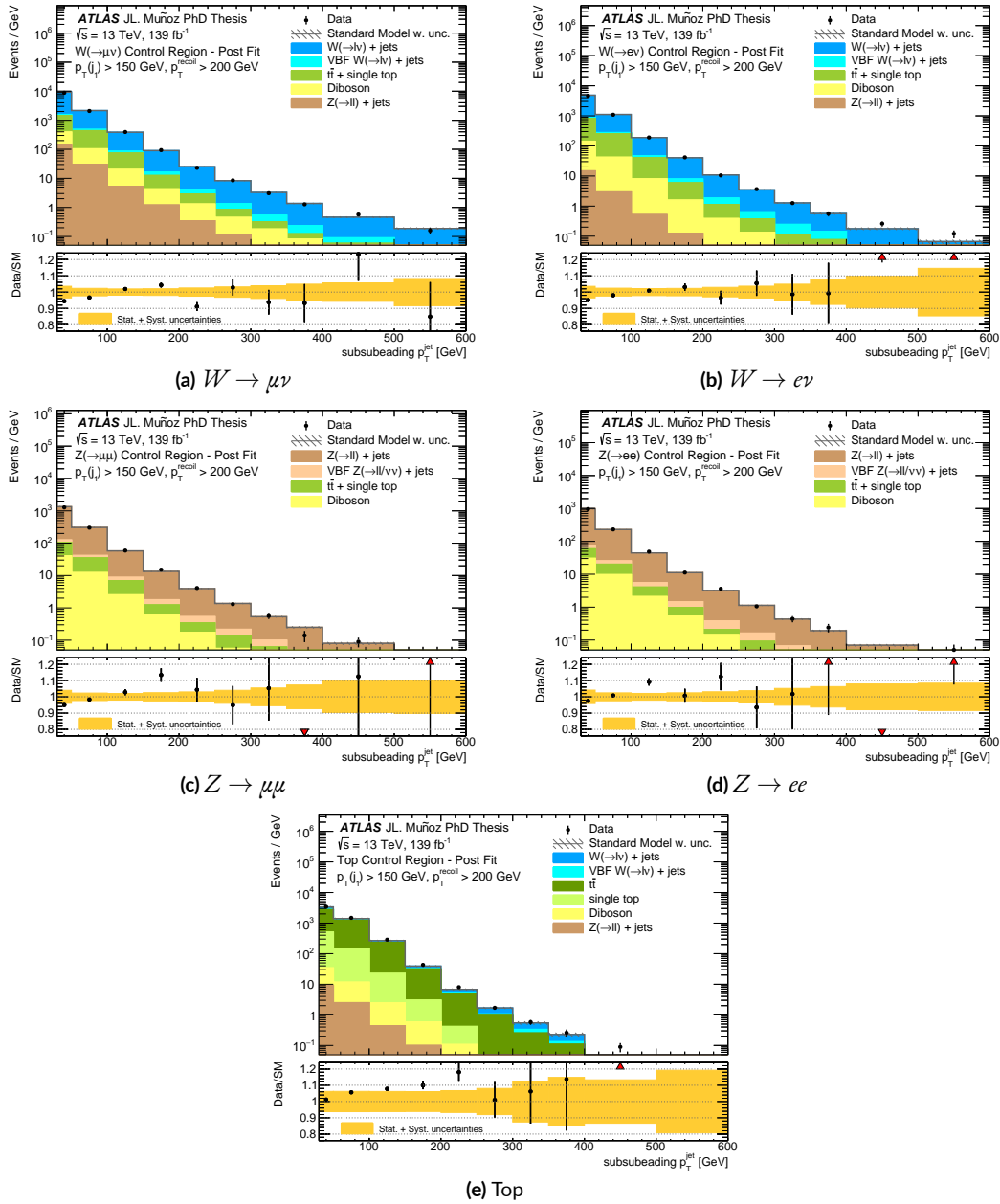


Figure 8.8: Sub-sub-leading jet p_T post-fit distribution in the control regions, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panels show the ratio of data over background prediction.

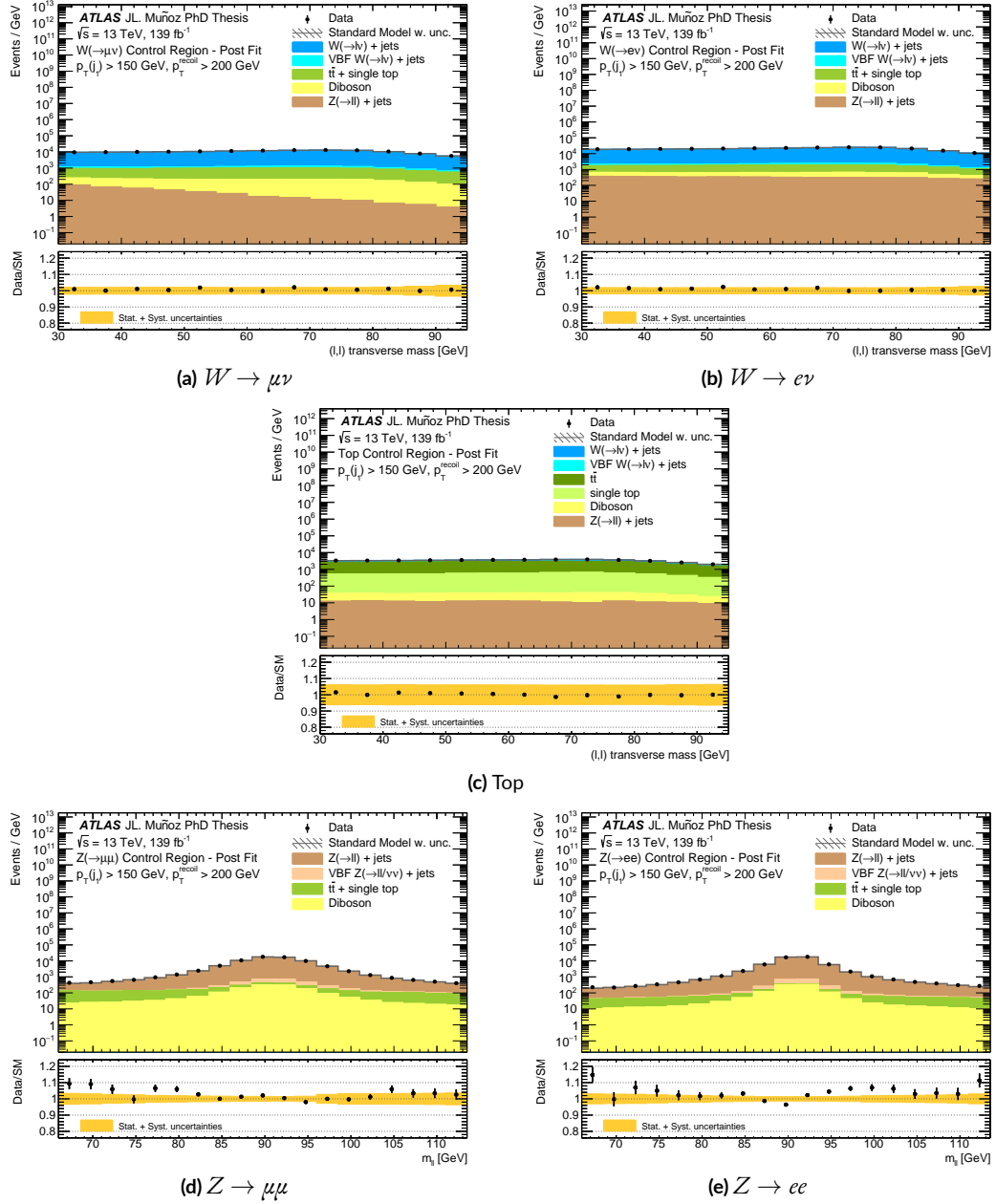


Figure 8.9: Lepton-neutrino transverse mass post-fit distribution in the one-lepton CRs (a, b and c) and di-lepton invariant mass post-fit distribution in the two-lepton CRs (d and e), based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panels show the ratio of data over background prediction.

8.2 RESULTS IN THE SR

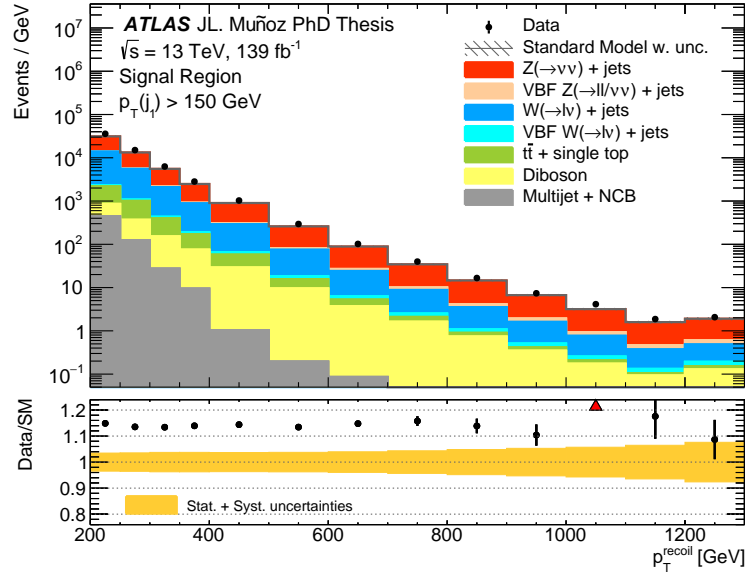
The normalization factors obtained from the background-only fit are propagated into the SR. In this way, the background normalization and the nuisance parameters are adjusted as discussed in the previous section.

8.2.1 BACKGROUND ESTIMATION

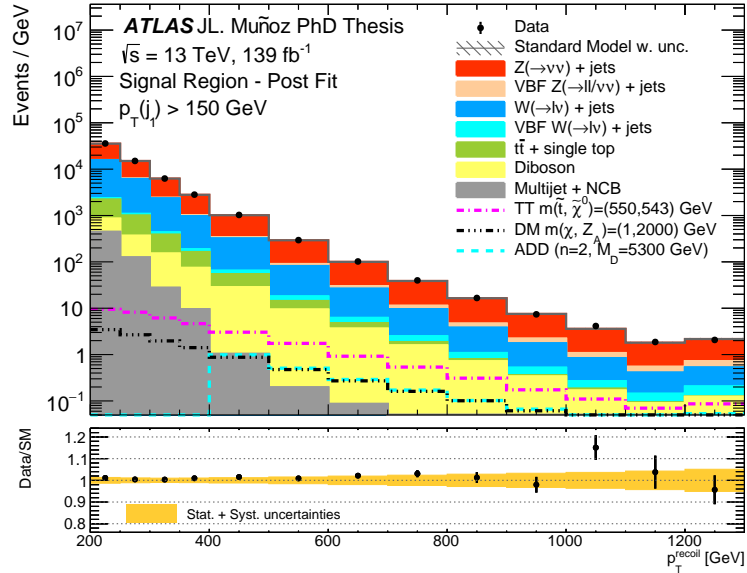
The final background prediction in the SR is shown in Figure 8.10 and is compared to the data. In the figure, the estimations before and after the fit are compared. The actual event yields for the several background processes are outlined for every exclusive bin of p_T^{recoil} in Table 8.7. In addition, the ratio between the data and the total MC prediction in each bin is included in the table.

As expected, the background process $Z(\nu\nu) + \text{jets}$ dominates largely the full p_T^{recoil} spectrum, with contributions varying from $\sim 55\%$ in EM0 to $\sim 66\%$ in EM12. It is followed by $W(\tau\nu) + \text{jets}$, with an impact of $\sim 20\%$ in EM0 and $\sim 7\%$ in EM12. The third largest background source is $W(\mu\nu) + \text{jets}$, whose contribution ranges between $\sim 20\%$ and $\sim 6\%$ in EM0 and EM12, respectively. The contributions from $W(e\nu) + \text{jets}$ falls from $\sim 6.5\%$ in EM0 to $\sim 1.4\%$ in EM12, while the diboson background raises from $\sim 1.1\%$ to $\sim 5.5\%$ in the same bins. The contribution from top-production processes is dominated by $t\bar{t}$ and decreases rapidly from $\sim 3.7\%$ in EM0 to $\sim 0.49\%$ in the region EM8, becoming negligible for $p_T^{\text{recoil}} > 900$ GeV. VBF-initiated $W/Z + \text{jets}$ processes, dominated by the VBF $Z(\nu\nu) + \text{jets}$ channel, play a sub-dominant role in the fit but exhibit a soft decreasing behaviour which makes them become especially relevant at the tail, giving a total contribution of $\sim 1.1\%$ in EM0 and $\sim 12.8\%$ in EM12. The multi-jet and non-collision backgrounds are only relevant at low p_T^{recoil} , with a contribution of $\sim 1.2\%$ in the first bin, but become negligible for p_T^{recoil} above 600 GeV. All other background sources contribute below 1%.

As anticipated by the fitted κ -factors, the total background is scaled up by a factor of about 13% after the fit. The total uncertainty gets reduced significantly, resulting in a precise SM prediction in almost the full p_T^{recoil} spectrum. As an example, the total background uncertainty in the region EM0 amounts to $\sim 1.5\%$, it goes slightly down to 1.2% in EM4 and for the region EM12 it raises up to $\sim 4.2\%$.



(a)



(b)

Figure 8.10: p_T^{recoil} distribution in SR before (a) and after (b) the fit, based on full Run 2 dataset. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panel shows the ratio of data over background prediction. Additionally, three signal models are overlaid with the backgrounds in (b): DM production with $m(Z_A, \chi) = (1, 2000) \text{ GeV}$ denoted by a black dash-dotted line, SUSY TT with $m(\tilde{t}, \tilde{\chi}^0) = (550, 543) \text{ GeV}$ (magenta dashed line) and ADD with $n = 3$ and $M_D = 5300 \text{ GeV}$ as a light blue dashed line.

Exclusive p_T^{recoil} bin	EM ₀	EM ₁	EM ₂	EM ₃	EM ₄
Observed events	1791624	752328	313912	141036	102888
Background post-fit prediction	1 783 000±26 000	753 000±9 400	314 000±3 500	140 100±1 600	101 600±1 200
$Z(\nu\nu)$ + jets	987 000±7 800	436 600±4 000	192 800±2 100	89 100±1 100	67 400±1 000
$W(\mu\nu)$ + jets	174 000±4 400	64 400±1 500	23 570±480	9 310±190	5 940±120
$W(e\nu)$ + jets	116 600±7 200	44 100±2 700	16 200±1 000	6 460±400	3 980±250
$W(\tau\nu)$ + jets	365 100±4 700	143 400±2 000	54 920±800	23 270±360	15 430±260
$Z(\mu\mu)$ + jets	6 380±160	2 032±53	597±16	188.0±5.1	97.4±2.7
$Z(ee)$ + jets	–	–	–	–	–
$Z(\tau\tau)$ + jets	4 990±130	1 642±43	530±14	194.3±5.3	115.6±3.3
Diboson	20 400±3 100	12 000±1 800	6 130±970	3 220±520	2 750±460
single- t	19 400±4 400	7 800±1 900	2 750±730	1 010±290	550±180
$t\bar{t}$	47 300±3 600	23 400±1 800	8 880±660	3 430±250	2 000±150
VBF $Z(\ell\ell/\nu\nu)$ +jets	11 000±1 200	6 820±790	3 880±480	2 270±300	2 170±310
VBF $W(\ell\nu)$ +jets	9 100±1 000	4 620±550	2 340±300	1 190±160	1 010±150
Multijet + NCB	22 000±19 000	6 200±5 500	1 400±1 200	470±400	103±73
Ratio Data/MC	0.995±0.015	1.00±0.013	1.00±0.011	0.993±0.011	0.987±0.012

Exclusive p_T^{recoil} bin	EM ₅₀₀	EM ₆₀₀	EM ₇₀₀	EM ₈₀₀	EM ₉₀₀
Observed events	29458	10203	3986	1663	738
Background post-fit prediction	29 240±420	10 000±180	3 873±79	1 645±40	754±20
$Z(\nu\nu)$ + jets	20 230±380	7 000±170	2 744±77	1 179±39	534±20
$W(\mu\nu)$ + jets	1 463±33	481±13	158.5±4.8	66.8±2.3	31.2±1.2
$W(e\nu)$ + jets	974±64	280±19	98.2±6.9	35.8±2.6	13.4±1
$W(\tau\nu)$ + jets	3 894±77	1 243±29	445±12	167.2±5.5	77.4±2.9
$Z(\mu\mu)$ + jets	14.99±0.44	4.51±0.15	1.650±0.060	1.490±0.050	0.670±0.020
$Z(ee)$ + jets	–	–	–	–	–
$Z(\tau\tau)$ + jets	27.17±0.82	8.31±0.28	3.15±0.11	0.960±0.040	–
Diboson	910±160	352±67	157±32	71±16	33.5±7.9
single- t	83±36	15.2±8.3	2.9±2.3	–	–
$t\bar{t}$	415±33	100.1±8.3	26.1±2.4	8.2±1.0	2.41±0.36
VBF $Z(\ell\ell/\nu\nu)$ +jets	850±130	370±64	174±32	86±17	45.4±9.8
VBF $W(\ell\nu)$ +jets	358±61	140±27	61±14	28.6±7.4	16.0±4.7
Multijet + NCB	20±16	8.6±8.0	–	–	–
Ratio Data/MC	0.993±0.014	0.980±0.018	0.972±0.020	0.989±0.024	1.02±0.027

Exclusive p_T^{recoil} bin	EM ₁₀₀₀	EM ₁₁₀₀	EM ₁₂₀₀
Observed events	413	187	207
Background post-fit prediction	359±11	182.4±6.4	218.1±9.2
$Z(\nu\nu)$ + jets	251±11	125.9±6.2	144.4±8.5
$W(\mu\nu)$ + jets	16.22±0.68	7.82±0.37	13.41±0.67
$W(e\nu)$ + jets	6.74±0.51	3.01±0.24	2.59±0.22
$W(\tau\nu)$ + jets	33.7±1.5	15.54±0.80	15.9±1
$Z(\mu\mu)$ + jets	–	–	1.190±0.040
$Z(ee)$ + jets	–	–	–
$Z(\tau\tau)$ + jets	–	2.100±0.080	0.550±0.040
Diboson	16.7±4.3	8.9±2.4	12.2±3.6
single- t	–	–	–
$t\bar{t}$	1.56±0.23	–	–
VBF $Z(\ell\ell/\nu\nu)$ +jets	24.4±5.6	13.7±3.3	19.7±5.1
VBF $W(\ell\nu)$ +jets	8.4±2.8	5±1.9	8.1±3.3
Multijet + NCB	–	–	–
Ratio Data/MC	0.870±0.030	0.976±0.035	1.05±0.042

Table 8.7: Predicted (post-fit) and observed background events in the signal region, in exclusive bins of p_T^{recoil} .

8.2.2 GOODNESS OF THE FIT

The agreement between predictions and data is overall good. A χ^2 -statistical test is performed to probe potential shape discrepancies in the p_T^{recoil} distribution. The χ^2 expression is defined as

$$\chi^2 = \sum_{i=1}^{N_{\text{bins}}} \frac{[d_i - b_i(\alpha)]^2}{\Delta d_i^2 + \Delta b_i(\alpha)^2} + \sum_j^{N_{\text{NPs}}} \alpha_j^2, \quad (8.1)$$

where d_i and t_i denote the measured data and the background prediction post-fit in the bin i , respectively ($N_{\text{bins}} = 13$), after minimizing the NLL including uncertainties as α NPs. Δd_i and $\Delta b_i(\alpha)$ represent the uncertainties on the data and the background predictions, respectively, and the second term in Eq. 8.1 sums over all NPs with their fitted values α_j ($N_{\text{NPs}} = 97$). This approximate formula assumes that the NPs are independent and correlated across all bins in p_T^{recoil} .

With the inputs from the background-only fit and given the number of degrees of freedom (13 bins), the obtained χ^2 is 10.05, which translates into a p -value of 0.69. This, which can be considered as a simplified probability for the background-only hypothesis, transforms, via Eq. (5.5), into a statistical significance of $z = 0.5\sigma$. This indicates a good agreement of the data with SM background-only hypothesis.

8.2.3 IMPACT OF SYSTEMATIC UNCERTAINTIES

In order to illustrate the impact of each systematic uncertainty in the SR, the background-only fit is repeated by shifting and fixing each NP at a time by $+1\sigma$ (or -1σ) and then evaluating the relative impact in terms of the variation of the total yields in each SR bin.

The obtained variations are outlined in Figure 8.11, where experimental and theoretical uncertainties are shown separately on the left and right sides of the figure. The impacts are combined as the sum in quadrature of the different components for each category of uncertainty. More detailed numbers are given in Appendix E.

The most relevant uncertainties from experimental sources are related to the electron and muon reconstruction and identification uncertainties, which dominate over the middle range of the E_T^{miss} spectrum and have an overall leading contribution, ranging between 0.4%–1.9%, in the case of the muons, and between 0.6% and 1% for the electrons. Jet-related uncertainties become relevant at high-boosted regimes, with the JES uncertainties dominating (varying between 0.17%–1%) and the JER ones giving the largest impact (1.3%) in the last bin. The E_T^{miss} track soft term resolution and scale uncertainties are relevant at the lowest regions but globally subdominant, ranging from 0.63% to 0.3%.

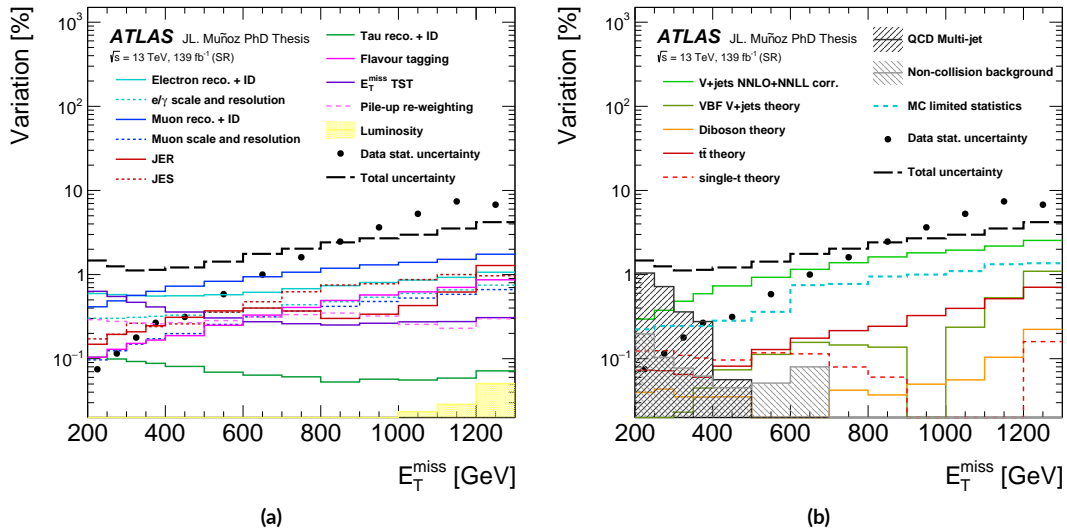


Figure 8.11: Impact of experimental (a) and theoretical (b) uncertainties in the SR, expressed in % and compared to the total background uncertainty and the data statistical uncertainty.

Regarding the theoretical uncertainties, the multi-jet uncertainty has the largest impact in the first two bins (1% and 0.7% in EM_0 and EM_1 , respectively), essentially being relevant only as its relative contribution to the total background is non-negligible. The uncertainties related to the V +jets higher-order corrections, largely governed by the EW components, are one of the most important sources of uncertainty, dominating the E_T^{miss} spectrum above 500 GeV and ranging between 0.3%–2.5%. The MC statistical uncertainty is non-negligible at high- E_T^{miss} , where reaches up to 1.0%, but it is still much less relevant than the data statistical uncertainty, which is about 7% in the last bin.

Comparing to their relative size before the fit, most of the systematic uncertainties are reduced. Table 8.8 includes the signed difference in % of the impact in the bins EM_0 , EM_6 and EM_{12} for those uncertainties that are most significantly reduced (differences larger than 0.1%). The averaged difference over all bins is also included. The uncertainty on the muon reconstruction efficiency, which is one of the dominant systematics, is the one that experiences the largest reduction (almost 0.9% in difference), followed by the electron/photon energy scale uncertainty and one of the components of electron identification uncertainty (about 0.4% of difference in both cases).

Nuisance Parameter	EM ₀	EM ₆	EM ₁₂	Average
vjets_d2kappa_EW_W	+0.08	-0.18	-0.73	-0.21
PRW_DATASF	-0.21	-0.28	-0.15	-0.22
EL_EFF_ID_CorrUncertaintyNP14_Electron_CRs	-0.03	-0.13	-0.25	-0.13
EL_EFF_ID_CorrUncertaintyNP13_Electron_CRs	-0.01	-0.11	-0.25	-0.11
MUON_EFF_ID_HighPT_pt2_eta0	-0.03	-0.12	-0.34	-0.14
EG_SCALE_ALL	-0.26	-0.38	-0.73	-0.42
MUON_EFF_ID_HighPT_pt3_eta0	-0.03	-0.11	-0.52	-0.17
FT_EFF_extrapolation	-0.04	-0.23	-0.78	-0.29
MUON_SAGITTA_RESBIAS	-0.09	-0.31	-0.65	-0.33
MUON_EFF_RECO_SYS	-0.35	-0.89	-1.45	-0.89
MET_SoftTrk_Scale	-0.12	-0.15	-0.24	-0.16
EL_EFF_ID_CorrUncertaintyNP9_Electron_CRs	-0.03	-0.10	-0.19	-0.10
vjets_d2kappa_EW_Z11	-0.08	-0.18	-0.28	-0.19
MUON_EFF_ID_HighPT_pt4_eta0	-0.04	-0.10	-0.59	-0.18
EL_EFF_ID_CorrUncertaintyNP11_Electron_CRs	-0.09	-0.20	-0.36	-0.21
FT_EFF_Light_systematics	-0.04	-0.20	-0.16	-0.14
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP8_Electron_CRs	-0.12	-0.35	-0.72	-0.38
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP17_Electron_CRs	-0.11	-0.24	-0.37	-0.24

Table 8.8: Difference, expressed in %, for each uncertainty in the SR with respect to their values before the fit. The differences are shown in the exclusive bins of p_T^{recoil} EM₀, EM₆ and EM₁₂, and the average difference is shown as well. Numbers in green (red) denote a reduction (increase) of the systematic uncertainty. Only uncertainties for which the average difference is larger than 0.1% are shown.

8.2.4 KINEMATIC DISTRIBUTIONS

Figure 8.12 shows the distributions of the leading jet p_T , the leading jet $|\eta|$, the jet multiplicity, the second jet p_T and the third jet p_T in the SR obtained with the fit to the full Run 2 dataset. Likewise in the CRs, there is a mismodelling in the prediction of the number of jets, which leads to discrepancies across the other jet kinematic variables. The leading jet $|\eta|$ and the third jet p_T predictions, however, seem to be in good agreement with the data within the background uncertainties.

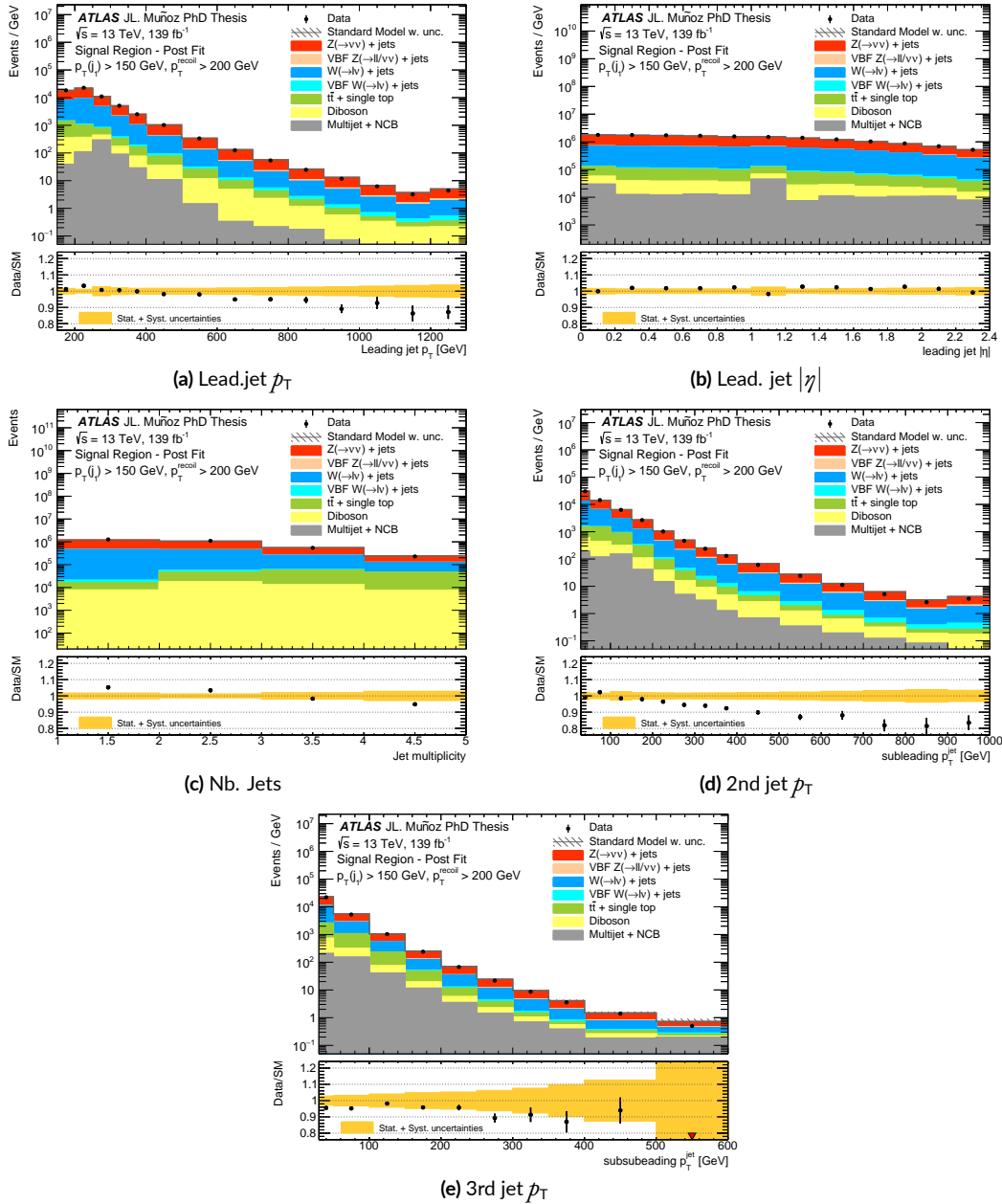


Figure 8.12: Kinematic post-fit distribution in the SR, based on full Run 2 dataset. The leading jet p_T and $|\eta|$, jet multiplicity, sub-leading jet p_T and sub-sub-leading jet p_T distributions are included. The data is shown as black dots, the total background prediction as the grey histogram. The hashed uncertainty band includes statistical and all systematic uncertainties. The lower panels show the ratio of data over background prediction.

SR region	$\langle\sigma\rangle_{\text{obs}}^{95}$ [fb]	S_{obs}^{95}	S_{exp}^{95}
IM ₀	736	102 274	83 000 ^{+22 000} _{-23 000}
IM ₁	296	41 158	33 800 ^{+11 000} _{-9 400}
IM ₂	150	20 893	15 400 ^{+5 900} _{-4 300}
IM ₃	86	11 937	8 300 ^{+3 100} _{-2 300}
IM ₄	52	7 214	4 700 ^{+1 800} _{-1 300}
IM ₅	21	2 918	1 930 ⁺⁷³⁰ ₋₅₄₀
IM ₆	10	1 391	940 ⁺³⁶⁰ ₋₂₆₀
IM ₇	4.1	574	490 ⁺¹⁹⁰ ₋₁₄₀
IM ₈	2.1	298	277 ⁺¹¹⁰ ₋₇₇
IM ₉	1.2	164	168 ⁺⁶⁵ ₋₄₇
IM _{IO}	1.3	186	119 ⁺⁴⁵ ₋₃₃
IM _{II}	0.52	73	75 ⁺²⁸ ₋₂₁
IM _{I2}	0.29	40	49 ⁺¹⁹ ₋₁₄

Table 8.9: Expected and observed 95% CL upper limits on the number of signal events, S_{obs}^{95} and S_{exp}^{95} , and on the visible cross-section, defined as the product of cross-section, acceptance and efficiency, $\langle\sigma\rangle_{\text{obs}}^{95}$, for the inclusive selections.

8.3 MODEL-INDEPENDENT LIMITS

The agreement between the data and the SM predictions translates into limits on the existence of new physics phenomena. In particular, one can set limits to the visible cross-section of any new physics process in a model-independent way. The model-independent fit, as described in Chapter 5, is performed. Since an inclusive approach is adopted in this case, the $p_{\text{T}}^{\text{recoil}}$ shape information is lost and therefore there is no longer benefit in splitting the top-production processes into $t\bar{t}$ and single- t , so they are treated as a single background process with a unique normalization factor. On the other hand, individual theoretical uncertainties for $t\bar{t}$ and single- t are kept separated, as they were in the background-only fit. In this way, two normalization factors are fitted for the backgrounds: κ^V for all the V +jets processes and κ^t for the top-production processes.

In each of the inclusive regions, 95% CL limits are obtained on the signal strength μ_{sig} , which is interpreted as the maximum amount of signal events allowed in addition to the background prediction. This quantity is denoted as S_{exp}^{95} for the expected limit and S_{obs}^{95} for the observed limit. These

numbers can be translated into a 95% CL limit on the visible cross-section, defined as

$$\langle \sigma \rangle_{\text{obs}}^{95} = \sigma \times A \times \varepsilon, \quad (8.2)$$

where σ is the cross-section of the unknown process, A is the acceptance of the detector and ε is the efficiency of the selection criteria applied in the monojet SR. The limits $\langle \sigma \rangle_{\text{obs}}^{95}$ are obtained by dividing S_{obs}^{95} by the integrated luminosity of the corresponding dataset, taking into consideration the systematic uncertainties in the SM predictions and the uncertainty on the luminosity. The final results are listed in Table 8.9. Values of $\sigma \times A \times \varepsilon$ above 736 fb for IMo and above 0.3 fb for IM12 are excluded at 95% CL.

8.4 INTERPRETATIONS

Limits into specific models of new physics can be also set from the obtained agreement between the data and the background predictions. In Chapter 1, the phenomenology of a number of BSM theories was introduced. A study of the acceptance times efficiency of the different models in the monojet signature ($A \times \varepsilon$) is included in Appendix F.

To obtain the limits, an exclusion fit is carried out using the profile likelihood defined in Chapter 6 to construct a profile likelihood which is used as test-statistic. The CL_s method is then used to set limits at 95% CL on μ_{sig} . The observed limits are calculated using the observed event yield in data in the SR, while for the expected limits the nominal event yield is set to the nominal background expectation. The $\pm 1\sigma$ uncertainty on the observed limit is calculated from the theoretical uncertainty on the specific cross-section of the model, while for the expected limits the $\pm 1\sigma$ and $\pm 2\sigma$ bands are calculated from the background uncertainties and the theoretical uncertainties on the signal acceptance. The computed exclusion limits obtained for a number of BSM scenarios are reported below.

8.4.1 LARGE EXTRA DIMENSIONS

The results can be interpreted as limits on the fundamental scale, M_D , in $4 + n$ spatial dimensions, in the context of the ADD LED signal model. For the generated samples, values of $A \times \varepsilon$, as computed from simulated events with $E_T^{\text{miss}} > 350$ GeV, are of the order of 23% for EM4 and about 2% for EM12.

As mentioned previously and like in previous iterations of the analysis [49, 174, 179], only the

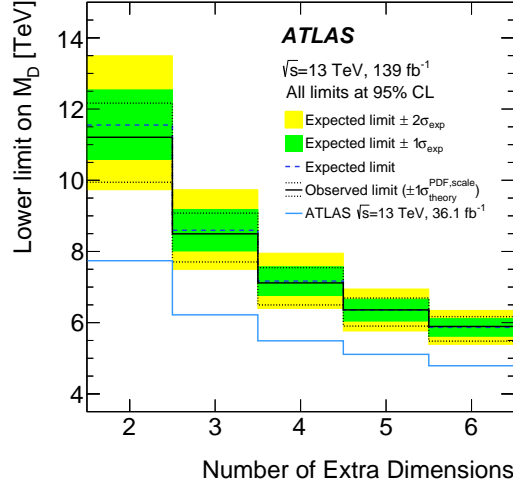


Figure 8.13: Observed and expected 95% CL lower limits on the fundamental Planck scale in $4 + n$ dimensions, M_D , as a function of the number of extra dimensions. The bands indicate the $\pm 1\sigma$ theory uncertainties in the observed limit, and the $\pm 1\sigma$ and $\pm 2\sigma$ ranges of the expected limit in the absence of a signal. Previous results from ATLAS using 36.1 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data [49] are included for comparison.

exclusive SR regions with $p_T^{\text{recoil}} > 400 \text{ GeV}$, with sufficient sensitivity to the signal, are included in the fit. The limits on M_D are calculated as follows. By definition, the excluded signal strength μ_{excl} is equal to the ratio

$$\mu_{\text{excl}} = \frac{\sigma_{\text{excl}}}{\sigma_{\text{nom}}}, \quad (8.3)$$

where σ_{excl} is the production cross-section for any signal model and σ_{nom} is the nominal value of the cross-section used in the MC generation of the signal sample. In the ADD scenario, the following relation between the Kaluza-Klein graviton production cross-section and M_D has been proved [180]:

$$\sigma \propto M_D^{-(n+2)}. \quad (8.4)$$

So one can combine Eqs.(8.3) and (8.4), and finally obtain the following relation:

$$M_D^{\text{excl}} = \frac{M_D^{\text{nom}}}{\mu_{\text{excl}}^{1/(n+2)}}, \quad (8.5)$$

where M_D^{nom} is the nominal value of the fundamental scale in the $4 + n$ dimensional space that is used at the MC generation stage for the signal sample.

95% CL Limits on M_D [TeV]		
ADD model	Expected	Observed
$n = 2$	$11.6^{+1.0}_{-1.0}$	$11.2^{+1.0}_{-1.3}$
$n = 3$	$8.6^{+0.6}_{-0.6}$	$8.5^{+0.6}_{-0.8}$
$n = 4$	$7.2^{+0.4}_{-0.4}$	$7.1^{+0.4}_{-0.6}$
$n = 5$	$6.4^{+0.3}_{-0.3}$	$6.4^{+0.3}_{-0.5}$
$n = 6$	$5.9^{+0.2}_{-0.2}$	$5.9^{+0.3}_{-0.4}$

Table 8.10: Expected and observed 95% CL lower limits on M_D as a function of the number of extra dimensions n in the context of the ADD signal model. The impact of the $\pm 1\sigma$ uncertainty from the theory on the observed limits and the expected $\pm 1\sigma$ range of limits in absence of a signal is reported.

The results from the exclusion fit are translated via Eq.(8.5) into 95% CL limits on M_D , and reported in Figure 8.13 and Table 8.10. In the figure, the dashed blue line shows the expected limits with the $\pm 1\sigma$ and $\pm 2\sigma$ error bands (green and yellow bands, respectively), and the black solid is the observed limit, with the $\pm 1\sigma$ uncertainty on the signal cross-section as dashed lines. The limits on M_D decrease with increasing n due to the cross-section scaling with $\sim 1/M_D^{n+2}$ (Eq. 8.4). The obtained results allow to exclude values of M_D up to 11.2 TeV and 5.9 TeV for $n = 2$ and $n = 6$ extra dimensions, respectively, at 95% CL, improving previous results using 36.1 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data [49]. The observed limits are in good agreement with the expected ones, within their 1σ uncertainty bands.

In previous versions of the analysis, the validity of the EFT was evaluated by performing a truncation of the obtained limits on M_D . This was essentially done via event-level damping weights, applied when the transferred momentum \hat{s} of the event becomes comparable to the scale of the theory, i.e. $\hat{s} \geq M_D^2$. However, in Ref. [49] it was already noticed that the suppression of this kinematic region had negligible effect on the results, therefore no truncation is considered this time.

8.4.2 DARK MATTER PRODUCTION

As discussed in Chapter 2, two DM simplified models are considered: the axial-vector mediator model and the pseudo-scalar mediator model. In the case of the DMA model for DM-pair production with $m_{Z_A} > 2m_\chi$ and a mediator mass of 2 TeV, $\mathcal{A} \times \varepsilon$ values range from 13% to less than 1% for the EMO and EM12 selections, respectively, while in the case of the DMP model values between 13% and less than 1% are typical for masses $m_{Z_p} = 350 \text{ GeV}$ and $m_\chi = 1 \text{ GeV}$. These values refer to simulated

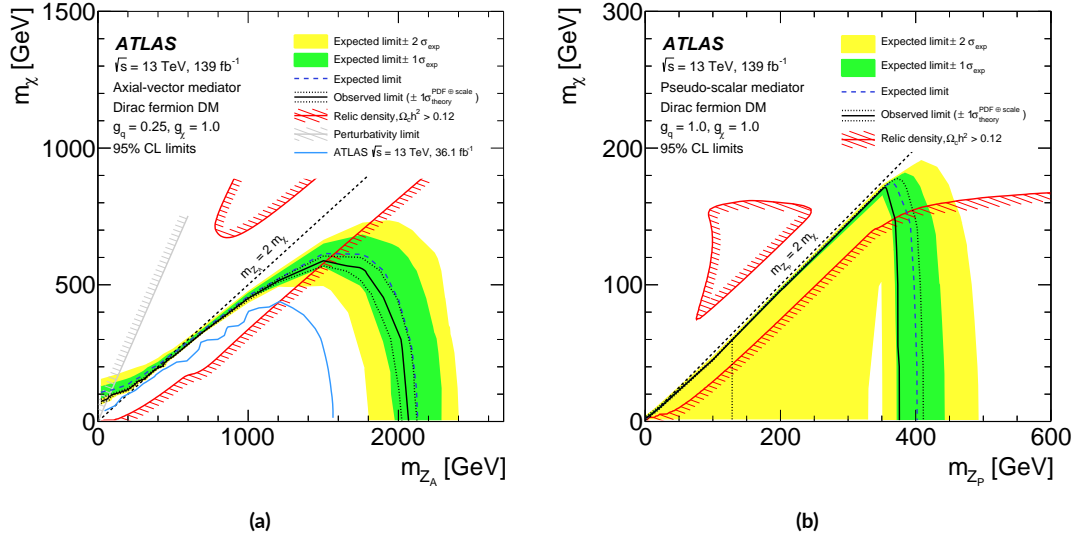


Figure 8.14: 95% CL exclusion contours in the (a) $m_{Z_A} - m_{\chi}$ plane for the DMA model and (b) in the $m_{Z_P} - m_{\chi}$ plane for the DMP model. The solid (dashed) curves show the observed (expected) limits, while the bands indicate the $\pm 1\sigma$ theoretical uncertainties on the observed limit and the $\pm 1\sigma$ and $\pm 2\sigma$ ranges of the expected limit. The red curves represents the region where the expected relic DM density is consistent with the WMAP measurements ($\Omega_{\chi} = 0.12$), as computed with MADDM [181], while the area on the hashed side of the curves is the region inconsistent with such measurements. The gray hatched area indicates the region excluded due to perturbativity, defined by $m_{\chi} > \sqrt{\pi}/2m_{Z_{A,P}}$. The dotted lines indicate the kinematic limit for on-shell production $m_{Z_{A,P}} = 2 \times m_{\chi}$. In (a), the results are compared with previous results from ATLAS at $\sqrt{s} = 13$ TeV using 36.1 fb $^{-1}$ [49].

samples generated with a minimum E_T^{miss} of 150 GeV.

In both scenarios, the fit results are translated into exclusion contours in the $m_{Z_{A,P}} - m_{\chi}$ plane as follows. The exclusion fit is performed with the full E_T^{miss} spectrum at 95% CL for a two sets of signal samples: one with $m_{\chi} = 1$ GeV and different $m_{Z_{A,P}}$ values within the on-shell regime (i.e. $m_{Z_{A,P}} > 2m_{\chi}$), and another one for different $(m_{Z_{A,P}}, m_{\chi})$ points off-shell. For a given mediator mass, the shape of the E_T^{miss} distribution does not change as a function of m_{χ} in the on-shell regime, therefore the limits on the signal strength obtained for $m_{\chi} = 1$ GeV in this region of the parameter-space are re-scaled by the cross-section via

$$\mu'_{\text{excl}} = \mu_{\text{excl}} \times \sigma/\sigma', \quad (8.6)$$

where μ_{excl} and σ are the limit and the cross-section of the reference sample, and μ'_{excl} and σ' the corresponding limit and cross-section of the point in the parameter space to where to extrapolate. The limits are then extrapolated all the way up to the diagonal defined by $m_{Z_{A,P}} = 2m_{\chi}$. Figure 8.14 shows the observed and expected exclusion contours at 95% CL on the $m_{Z_A} - m_{\chi}$ ($m_{Z_P} - m_{\chi}$) plane on the left

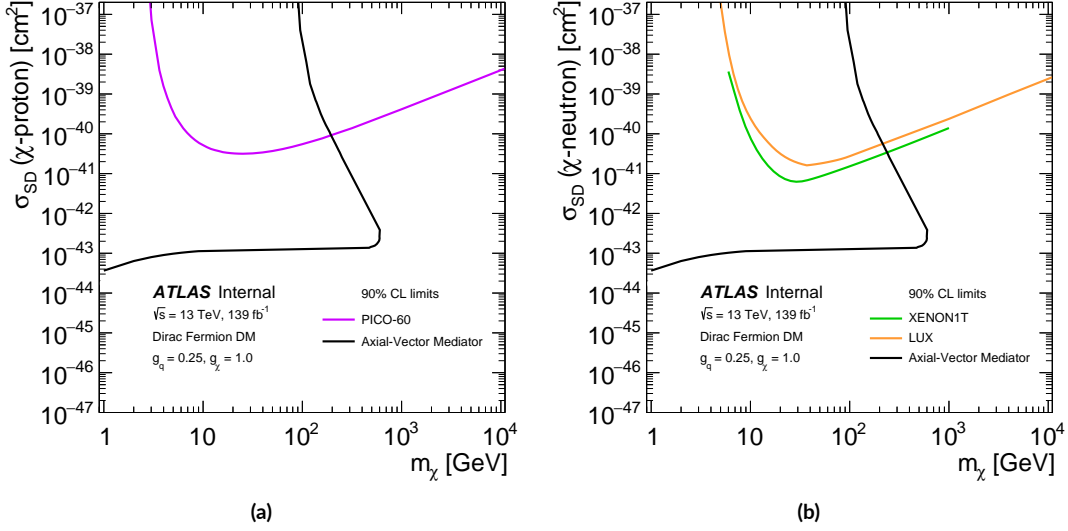


Figure 8.15: Comparison of the obtained 90% CL limits for the DMA model (black line) with respect to results from direct-detection experiments on the spin-dependent WIMP–nucleon scattering cross-section (left: proton, right: neutron) as a function of the DM mass. The region to the left of the contour is excluded. Limits from the PICO [182] (purple line), LUX [183] (orange line), and XENON1T [184] (green line) experiments are included for comparison, which is only valid in the context of this model, assuming minimal mediator width and coupling values of $g_q = 1/4$ and $g_\chi = 1$.

(right) side for an axial-vector (pseudo-scalar) mediator with $g_q = 0.25$ and $g_\chi = 1.0$ ($g_q = g_\chi = 1$). The region below the curve is excluded. The region for which the models predict a DM relic density higher than measured by Planck [36] and WMAP [37], namely $\Omega b^2 > 0.12$, is computed using MadDM [181]. In the case of the DMA model, the observed (expected) limits extend up to about $m_A = 2067$ (2128) TeV for DM mass candidates of $M_\chi = 1$ GeV, extending the previous results [49] by about 450 (400) GeV. For the DMP scenario, the observed (expected) limits extend up to about $m_{Z_p} = 376$ (402) GeV for DM mass candidates of $m_\chi = 1$ GeV. For the first time the ATLAS monojet analysis have the required sensitivity to exclude part of the parameter space of this model.

The axial-vector limits are also converted into 90% CL limits on the spin-dependent WIMP-proton and WIMP-neutron scattering cross-section as a function of m_χ , in order to compare the results with the ones obtained by DM direct detection experiments (left and right sides on Figure 8.15, respectively). This is done first by re-computing the exclusion fit at 90% CL in order to be comparable to the direct-detection measurements, and then by using the following relation, discussed in Ref. [62]:

$$\sigma_{SD} = 2.4 \times 10^{-41} \text{ cm}^2 \times \left(\frac{g_\chi g_q}{0.25} \right)^2 \times \left(\frac{1 \text{ TeV}}{m_{Z_A}} \right)^4 \times \left(\frac{\mu_{n\chi}}{1 \text{ GeV}} \right)^2, \quad (8.7)$$

where σ_{SD} denotes the spin-dependent WIMP-nucleon scattering cross-section and μ_{m_χ} the reduced mass of the WIMP-nucleon system, $\mu_{m_\chi} = m_\chi m_n / (m_\chi + m_n)$. The obtained limits, translated via Eq. (8.7) into limits on the scattering cross-section, are of the order of $1.4 \times 10^{43} \text{ cm}^2$ for DM masses of about 100 GeV, and $3 \times 10^{44} \text{ cm}^2$ for $m_\chi < 10 \text{ GeV}$, complementing the results from direct-detection experiments.

8.4.3 SUSY QUARK-PAIR PRODUCTION

As mentioned previously, different models for squark-pair production are considered in the analysis. Namely, light squark-pair production with $\tilde{q} \rightarrow q\tilde{\chi}_1^0$ and $q = (u, d, c, s)$ (denoted as SS), sbottom-pair production with $\tilde{b} \rightarrow b\tilde{\chi}_1^0$ (BB), stop-pair production with $\tilde{t} \rightarrow c\tilde{\chi}_1^0$ (TT) and, finally, stop-pair production with $\tilde{t} \rightarrow bff'\tilde{\chi}_1^0$ (T4body). A 100% branching ratio is assumed separately in each of the four scenarios. For all cases, the results are translated into 95% CL exclusion contours in the $m_{\tilde{q}}-m_{\tilde{\chi}_1^0}$ parameter space. The analysis is mainly sensitive to compressed scenarios, i.e. those where the difference in mass between the squark and the neutralino, Δm , is small. The region with $\Delta m < 5 \text{ GeV}$ is not considered for the sbottom and stop cases, since in this regime the squarks could become long-lived. Typical $\mathcal{A} \times \varepsilon$ values, computed for samples generated with $E_{\text{T}}^{\text{miss}} > 150 \text{ GeV}$, in the EMO and EM12 regions are:

- 10% and less than 1% for an SS sample with $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0}) = (900, 895) \text{ GeV}$;
- 13% and less than 1% for a BB sample with $(m_{\tilde{b}}, m_{\tilde{\chi}_1^0}) = (500, 300) \text{ GeV}$;
- 11% and less than 1% for a TT sample with $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0}) = (600, 593) \text{ GeV}$;
- 11% and less than 1% for a T4body sample with $(m_{\tilde{t}}, m_{\tilde{\chi}_1^0}) = (450, 443) \text{ GeV}$.

The exclusion fit is performed using the full $p_{\text{T}}^{\text{recoil}}$ spectrum for all the generated SUSY MC samples. The fitted signal strengths are interpolated between the grid of mass points and the exclusion contours are drawn where $\mu_{\text{sig}} = 1$. The $\pm 1\sigma$ uncertainty band on the observed limit is obtained by re-scaling the observed limits by the $\pm 1\sigma$ theoretical uncertainty on the NNLO+NNLL cross-section.

The expected and observed 95% exclusion contours for all the models are displayed in Figure 8.16. The SS limits are presented in the Δm vs. $m_{\tilde{q}}$ plane, while in the BB, TT and T4body cases the limits are displayed as a function of $m_{\tilde{\chi}_1^0}$ vs $m_{\tilde{b}}/m_{\tilde{t}}$ instead. For the third generation squark contours, the grey dashed lines indicated the range in which the decays are allowed. In the most compressed scenarios, masses below 925 GeV are excluded for the SS model, and sbottom and stop masses up to about

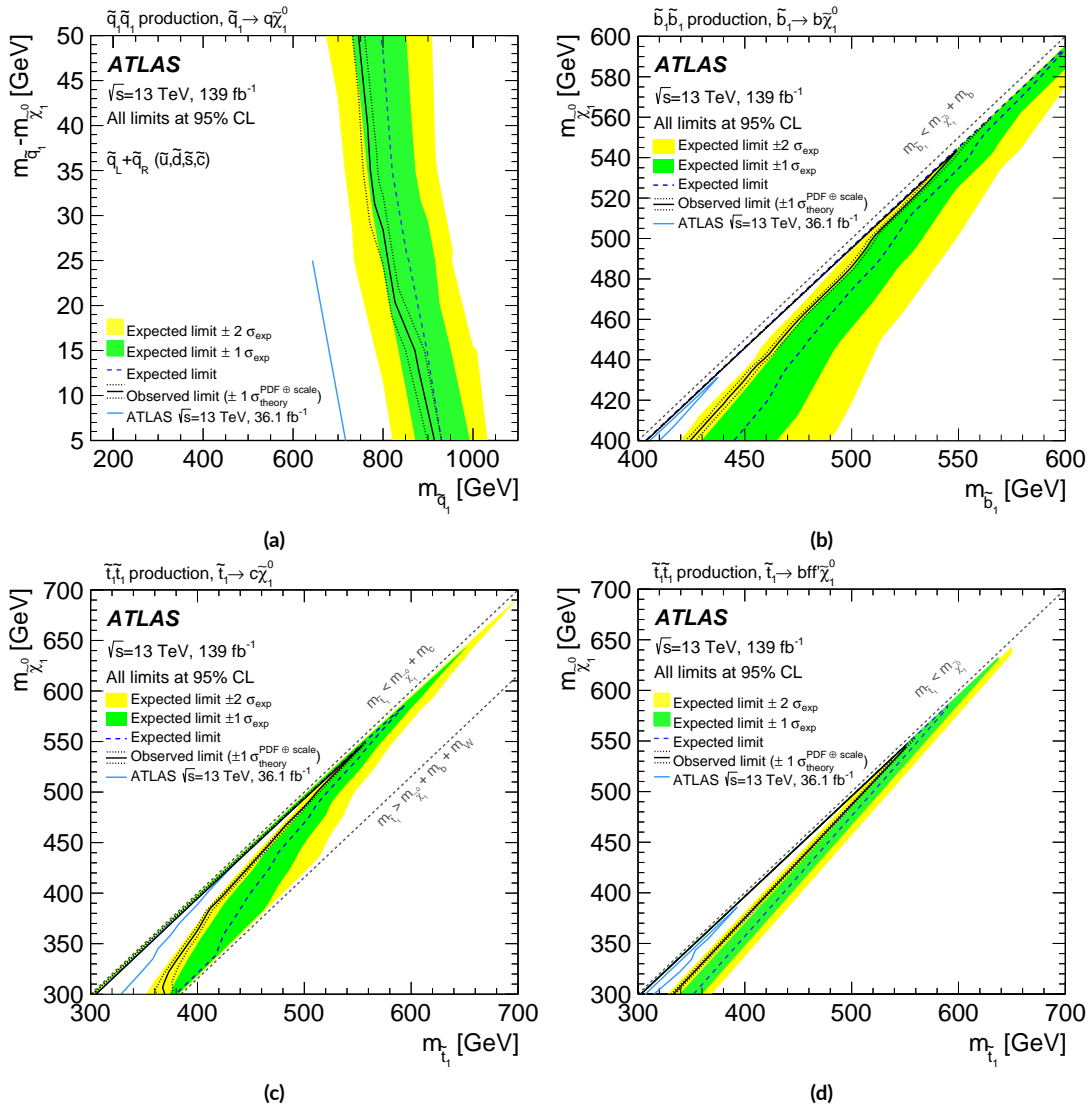


Figure 8.16: Exclusion region at 95% CL as a function of (a) the mass of the squark and the difference in mass with respect to that of the neutralino for the SS model, and as a function of the mass of the squark versus the mass of the neutralino for the (b) BB, (c) TT and (d) T4body models. The dotted lines around the observed limits indicate the $\pm 1\sigma$ variations of the NNLO+NNLL SUSY cross-section predictions. The bands around the expected limits denote the expected $\pm 1\sigma$ and $\pm 2\sigma$ ranges of limits in the absence of a signal. In all cases, the limits are compared with previous results from ATLAS at $\sqrt{s} = 13$ TeV using 36.1 fb $^{-1}$ [49]. In the case of the BB, TT and T4body models, the grey dashed lines indicate the allowed region for the decays.

545 GeV are excluded correspondingly for the BB, TT and T4body signatures, all of them at 95%

95% CL Limits on $\Gamma(H \rightarrow \text{inv})$	
Expected	Observed
$0.39^{+0.16(0.19)}_{-0.11(0.38)}$	0.34

Table 8.11: Expected and observed 95% CL upper limits on the branching ratio for an invisible-decaying Higgs boson. The $\pm 1\sigma$ uncertainties on the expected limit are quoted, and the $\pm 2\sigma$ errors are added within parenthesis.

CL. These results improve significantly the limits set in the previous iteration of the analysis [49], extending the exclusion limit on the mass of the squark by more than 100 GeV in the very compressed scenario.

8.4.4 INVISIBLE-DECAYING HIGGS

The results are also translated into a 95% CL upper limit on the branching ratio for a Higgs boson decaying into invisible products. The limits are calculated assuming SM production cross-sections and combining the contributions from ggF (73%), VBF (18%), VH (8%) and ttH (1%) processes. Typical $\mathcal{A} \times \varepsilon$ values, calculated for samples generated with $E_T^{\text{miss}} > 150$ GeV, are of about 11% in EMo and decrease rapidly, becoming already lower than 0.1% in EM7. The low E_T^{miss} region plays therefore an important role in enhancing the sensitivity of the data to this particular signal, so the full set of exclusive bins is employed in the exclusion fit.

Table 8.11 shows an obtained expected limit for the $H \rightarrow \text{inv}$ process of $0.39^{+0.16}_{-0.11}$, in good agreement with the observed limit calculated of 0.34. This limit is weaker in comparison with respect to the one obtained by the ATLAS VBF analysis, which is actually the channel with the best sensitivity to this interpretation, with an upper limit on the BR of 0.13 [92]. Moreover, the strongest limit reported up to date is $\Gamma(H \rightarrow \text{inv}) < 0.11$, produced by the Run 1 + Run 2 combination of the ATLAS VBF and ttH channels [185]. The results provided by the monojet analysis must be interpreted then as complementary sensitivity to be used in potential combinations with other channels such as mono- V , as it has been done by CMS [186].

8.4.5 OTHER MODELS

In addition, exclusion limits are set on the ALPs and DE models.

- For the ALPs interpretation, the results are translated into 95% CL exclusion contours in the

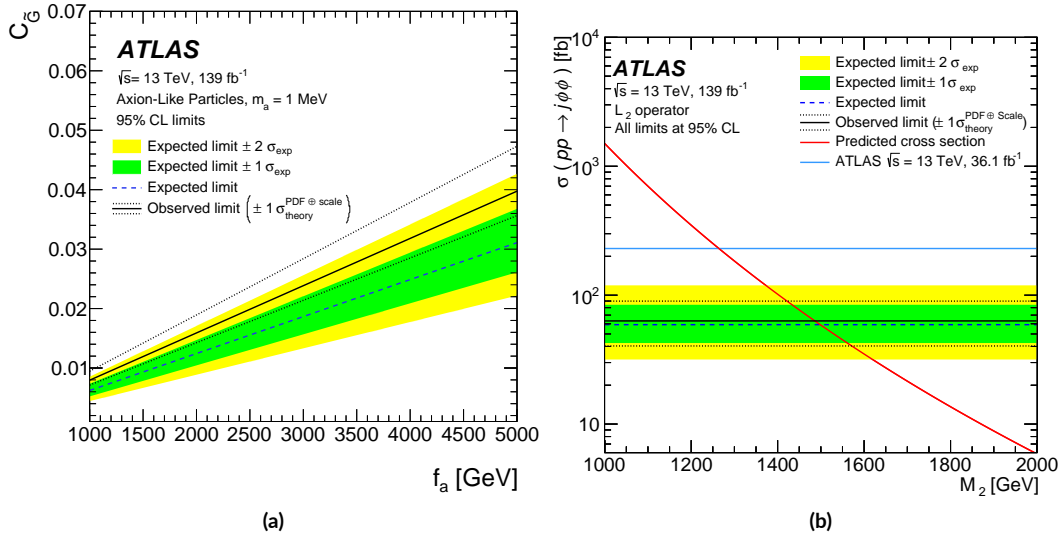


Figure 8.17: (a) Observed and expected 95% CL upper-limits on the coupling $c_{\tilde{G}}$ as a function of the effective scale f_a for an ALP mass of 1 MeV. The 95% CL limits are computed with no suppression of the events with $\hat{s} > f_a^2$. (b) Expected and observed and expected exclusion limits (dashed and solid line, respectively) at 95% CL on the Horndeski Dark Energy model for $m_\phi = 0.1$ GeV and $c_{i \neq 2} = 0$, $c_2 = 1$, expressed in terms of the visible cross-section as a function of the suppression scale M_2 and compared with theoretical predictions (red solid line). The results from this analysis are compared with previous results from ATLAS at $\sqrt{s} = 13$ TeV using 36.1 fb^{-1} . In both cases, the bands indicate the $\pm 1\sigma$ theoretical uncertainties in the observed limit and the $\pm 1\sigma$ and $\pm 2\sigma$ ranges of the expected limit.

$c_{\tilde{G}}-f_a$ plane, for a given mass of the ALP of 1 MeV. As Figure 8.17a shows, the obtained limits on the coupling $c_{\tilde{G}}$ increase linearly with the scale of the theory, f_a , and couplings with values above 0.008 are excluded for $f_a = 1$ TeV. Moreover, the limits computed with the fit to the data are in good agreement with respect to the expected curve given by the predictions within the uncertainty bands.

- Similarly, 95% CL exclusion limits are computed for the Horndeski Dark Energy inspired model with $m_\phi = 0.1$ GeV and $c_i = \delta_{i,2}$. Figure 8.17b shows the expected and observed exclusion contours on the $\sigma-M_2$ plane obtained for such model. The final limit set on the mass scale is then where the predicted cross-section (red solid line in Fig. 8.17b) meets the observed limit contour (solid black line). Values of M_2 below 1486 GeV are excluded, improving previous results [187].

Conclusions

The monojet analysis, as performed with a sample of proton-proton collision events at $\sqrt{s} = 13$ TeV collected by ATLAS between 2015 and 2018, corresponding to a total integrated luminosity of 139 fb^{-1} , is presented in this thesis. Events passing a E_T^{miss} -based trigger requirement are selected with at least one high-energetic jet recoiling a large amount of E_T^{miss} and with no leptons and photons in the final state. Simulated samples are generated for each background and signal process expected to be compatible with the monojet signature. An extra reweighting is applied to the boson p_T distribution of all V +jets samples, based in theoretical QCD corrections at NNLO precision and EW corrections at NLO precision supplemented by Sudakov logarithms at two loops. The analysis strategy relies on a simultaneous binned likelihood fit to the E_T^{miss} distribution, which is divided in thirteen bins starting at 200 GeV and being the highest one at $E_T^{\text{miss}} > 1.2$ TeV. Five orthogonal control regions are defined to constrain the dominant backgrounds, composed by V +jets and top-quark production processes. Systematic uncertainties are included in the fit as nuisance parameters and treated as random Gaussian variables, and correlations across different processes are taken into account as well. Such strategy allows a good understanding of the $Z(\nu\nu) + \text{jets}$ background process that can be estimated with high precision from the fit of the other V +jets backgrounds in the control regions.

In comparison with respect to its previous version, the analysis has definitely profited from a factor-4 increase of data. But there are also some other major improvements that have contributed to a much better sensitivity. The inclusion of the τ -lepton veto plus the lowered p_T -threshold in the definition of baseline objects have increased the background rejection power. The implementation of the higher-order V +jets corrections is probably the most important improvement of the analysis, as it made possible the estimation of the dominant $Z(\nu\nu) + \text{jets}$ contributions with an unprecedented precision. In addition, the lowered cuts on E_T^{miss} and jet- p_T have allowed the analysis to be sensitive to softer E_T^{miss} spectrum signatures such as the invisible-decaying Higgs interpretation.

Good agreement is found between the observed data and the SM background predictions, within a total uncertainty that ranges between 1.5% and 4.7% across the E_T^{miss} spectrum. Model-independent limits are reported for thirteen inclusive E_T^{miss} regions, excluding visible cross-section values at 95% CL

above 736 fb for $E_T^{\text{miss}} > 200$ GeV and above 0.3 fb for $E_T^{\text{miss}} > 1200$ GeV. The obtained agreement between data and predictions is translated into exclusion limits on the parameter space of a number of BSM models using the CL_s method. An ADD model of large extra dimensions is considered, and lower limits on the fundamental Planck scale in $4 + n$ dimensions are reported at 95% CL, varying from 11.2 TeV for $n = 2$ to 5.9 TeV for $n = 6$. Two types of simplified models for DM WIMP-pair production in the s -channel are explored, one involving the exchange of an axial-vector mediator and the other one with a pseudo-scalar one. Axial-vector mediators with masses below 2.1 TeV for very light WIMPs and coupling values of $g_q = 0.25$ and $g_\chi = 1$ are excluded, and pseudoscalar mediators with masses below 376 GeV, for very low WIMP masses and coupling values $g_q = g_\chi = 1$ are excluded as well, both at 95% CL. The limits on the axial-vector mediator model improve the previous results by about 450 GeV while the ones reported for the pseudo-scalar model are genuinely new, as ATLAS did not have enough sensitivity for this model before. In addition, the obtained results for the axial-vector model are translated into 90% CL limits on the WIMP-nucleon cross-section, providing complementary sensitivity for direct detection searches. Similarly, the results are interpreted in terms of a search for squark-pair production in a compressed mass scenario. In the case of light-flavour squark-pair production with $\tilde{q} \rightarrow q\tilde{\chi}_1^0$ ($q = u, d, c, s$), masses of the squark below 925 GeV are excluded at 95% CL. Third generation squark cases are studied separately for stop- and sbottom-pair production with $\tilde{b} \rightarrow b\tilde{\chi}_1^0$, $\tilde{t} \rightarrow c\tilde{\chi}_1^0$ or $\tilde{t} \rightarrow bff\tilde{\chi}_1^0$, respectively. In all of them, squark masses below about 550 GeV are excluded at 95% CL, improving previous boundaries by almost 100 GeV. The monojet analysis provides then unique access to region close to the diagonal in the $m_{\tilde{q}}-m_{\tilde{\chi}}$ plane, which complements other ATLAS SUSY searches with better sensitivity in the rest of the parameter space. Limits are also obtained for the scenario of a Higgs boson decaying into invisible particles, excluding branching fraction values above 0.39 at 95% CL. This result contributes to future combined results using different channels such as V +MET.

Altogether, the reported limits are much stronger and prove the monojet analysis to keep being a very powerful tool for BSM searches for a large variety of models. Given the level of precision already achieved by the analysis, an additional order or magnitude in data statistics will be needed before the sensitivity can be improved at the tail of the E_T^{miss} distribution. The monojet channel will remain the golden channel for dark matter searches at hadron colliders with increasing center-of-mass energies.



IBL 3D pixel sensors performance studies

This appendix summarizes some studies that were carried out on the performance of the 3D pixel sensors in the ATLAS Insertable B-Layer.

The IBL is the most internal layer of the ATLAS Pixel detector. It was added to the previous three existing layers before the start of Run 2 (see Figure A.1a) in order to improve the quality of the impact parameter reconstruction for tracks thanks to smaller pixel occupancy, hence improving vertexing and b -tagging performance, as well as to add complementary robustness to the overall tracking procedure by additional redundancy in track measurements, necessary to deal with the factor-2 of increased luminosity that was expected with respect to Run 1.

The IBL has a radial extension of $31 \text{ mm} < R < 40 \text{ mm}$ and a length of 33.2 cm [188]. Its original design was projected to handle a total integrated luminosity of 550 fb^{-1} and a peak luminosity of $3 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, and it was built to withstand 250 MRad of ionizing dose and $5 \cdot 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ non-ionizing dose. It consists of 14 staves arranged around the beam pipe with a small azimuthal overlap between each other of $\Delta\varphi = 0.18^\circ$ to ensure full coverage in φ , as shown in Fig. A.1b. The average distance from the staves to the beam pipe is 32.25 mm . Each staff is 64 cm long and covers up to $|\eta| \leq 2.9$.

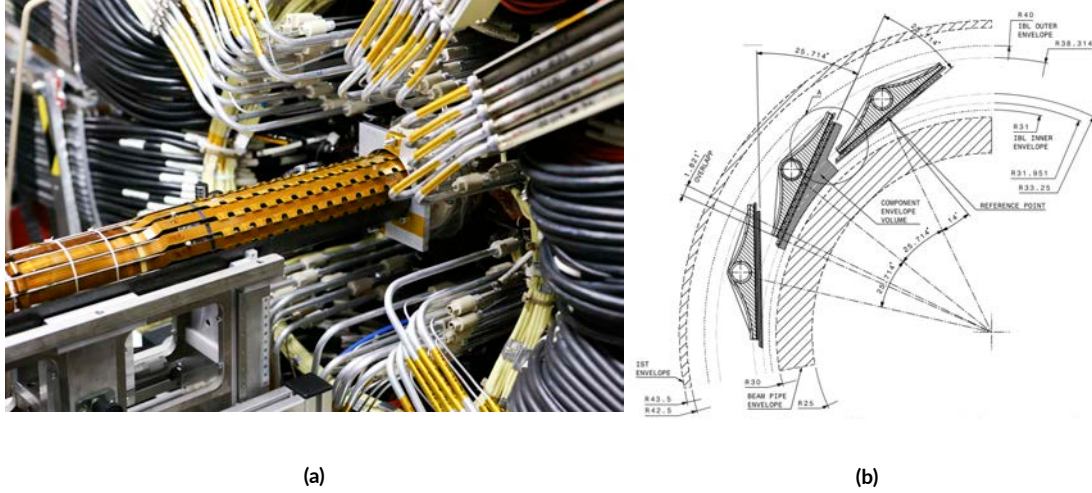
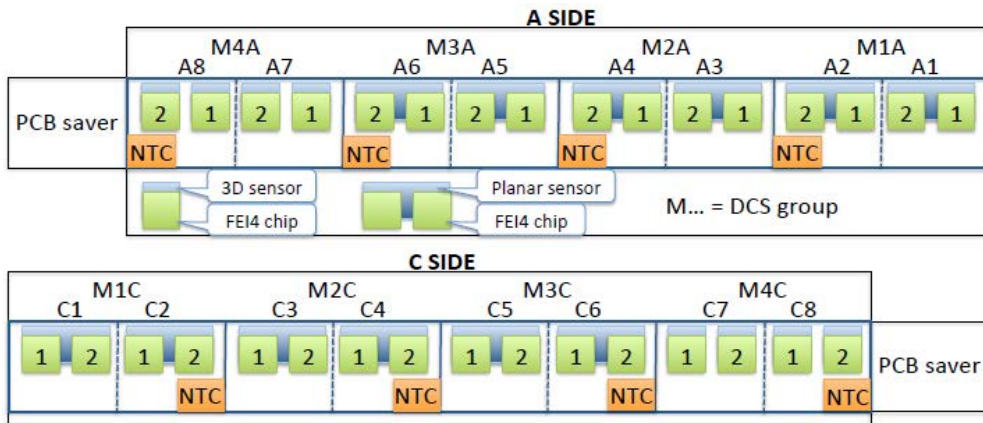


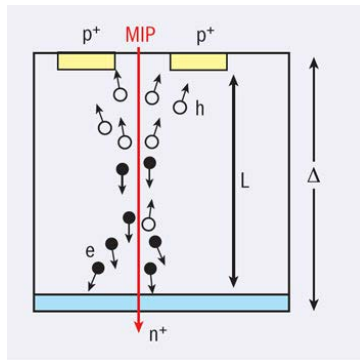
Figure A.1: (a) Assembly of the IBL into ATLAS in 2015 [189]. (b) IBL structure in the $r\phi$ - plane [188].

As Figure A.2a illustrates, each stave is divided symmetrically into its A- and C-sides*, on which silicon pixel sensors are grouped into modules. Each sensor is then connected to a FE-I4 front-end chip [190]. Two sensor technologies are used in the IBL: a conservative planar [191] type of sensors, and a novel 3D technology [192]. Planar sensors are $41.3 \times 19.2 \text{ mm}^2$ and have 160×336 pixels, while 3D sensors size $20.5 \times 18.5 \text{ mm}^2$ and have 80×336 pixels. In both cases pixels are $250 \times 50 \mu\text{m}^2$, with the longer side along the z -axis, but while the planar pixels are $200 \mu\text{m}$ thick the 3D ones are $230 \mu\text{m}$. Planar sensors are the ones used in most of the IBL, covering 75% of the active area, while 3D sensors are only placed at the extremes of the stave ($|\eta| > 2.6$), covering the remaining 25%. In the case of the planar sensors, the module will consist of two front-end chips and one sensor (2-chip module), while in the case of the 3D sensors, it will be a single chip and one sensor (1-chip module). The main difference between planar and 3D sensors is on their design. Planars are very similar to the other sensors used in the ATLAS Pixel Detector, with a slim n-in-n design and a $\sim 200 \mu\text{m}$ inactive region. On the other hand, 3D sensors are built with a n-in-p design, using the innovative concept of electrodes passing through the bulk, not only on the surface of the pixel as in the planar case (see Figures A.2b and A.2c). The advantage of this type of design is that, when a charged particle crosses the sensor, the pair electron-hole created would have to travel a much smaller distance than in the planar case, implying that these sensors need lower bias voltage to operate (planar sensors use to work at $\sim 300 - 400 \text{ V}$ while 3Ds do at around $30 - 40 \text{ V}$), resulting on lesser thermal losses and

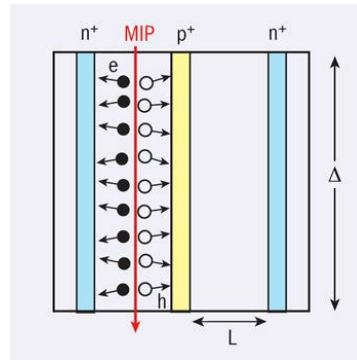
*This is a standard nomenclature in ATLAS, the A-side of the detector is along the positive z -axis and faces towards the direction of Geneva's airport and the Salève mountain, while the C-side is along the negative z -axis and is in the direction of the Jura. The letter 'B' corresponds to the central barrel.



(a)



(b)



(c)

Figure A.2: (a) Scheme of the modules arrangement on an IBL stave. Scheme of a (b) planar and a (c) 3D pixel sensor.

better performance after radiation. Such feature is shown in Figure A.2, where it can be seen that the depletion length, L , is proportional to the thickness Δ of the sensor in the planar case, while in the 3D case it is proportional to the distance between electrodes. The counter-side of the 3D technology is essentially the higher cost for large scale production, against the mature planar technology, more standardized within ATLAS and with lower production costs.

A.1 I-V CURVES STUDIES

An I-V curve (short for ‘current-voltage characteristic curve’), is a graphical representation of the relationship between the voltage applied across an electrical device and the current flowing through it. It is one of the most common methods of determining how an electrical device functions in a circuit. 3D pixel sensors, as semiconducting devices, only allow current to flow through in one direction. At

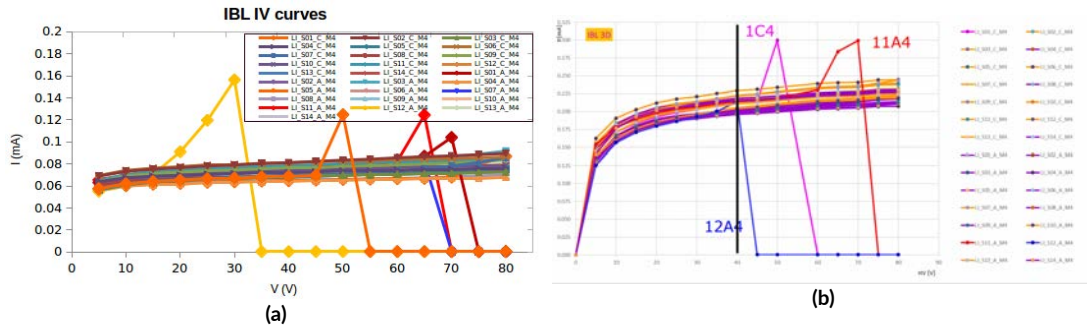


Figure A.3: 3D sensors I-V curves studied: (a) measured in 2017 and (b) measured in 2018.

positive voltages, the curve rises exponentially, indicating that current is free to flow through the device. At negative voltages, the current remains nearly at zero. However, a sufficiently large negative voltage (known as the *breakdown voltage*, V_{bd}) will cause the sensor to become conductive to negative current. The idea is to keep the sensors at sufficient negative voltages to be fully depleted but below the V_{bd} threshold, after of which the current would flow free and therefore no particle could be detected. Hence I-V scans can be used to measure V_{bd} and to decide at which voltage the sensors will operate for data-taking.

Two I-V scans carried out with IBL 3D sensors were studied: one performed on April 2017 (shown in Figure A.3a), and the other one on April 2018 (Fig. A.3b). In the figures, each line corresponds with a different module on a different IBL stave. The naming is not the same in both figures but it essentially denotes the number of the stave and the module number on the A or C sides.

- 2017 I-V curves showed three failing modules, with a measured V_{bd} significantly lower than the other modules. Comparing with respect to I-V scans performed before the IBL assembling, it was found that such modules were already showing early breakdowns at that stage. The comparison also indicated an overall increase of V_{bd} after irradiation for most of the modules. In addition, two modules were found to exhibit strange wiggles on their curves, indicating a non-stable behaviour. Finally, a recommendation to operate sensors at 40 – 50 V was given, with the exception of one module which should be operated at 20 – 25 V instead.
- 2018 I-V scans were studied in comparison with respect to the 2017 ones. No modules were showing unstable behaviour this time, and the three early-breakdown modules identified in 2017 were still showing low V_{bd} values, although higher than back then. Again, an overall increment of V_{bd} was noted for all modules, indicating a potential trend as consequence of the accumulated radiation damage.

A.2 HV SCANS

Some low-luminosity runs, not interesting for physics analysis use, are used from time to time to perform the so-called HV scans, which consist in operations where the sensors bias voltage is increased in controlled steps for fixed ranges of lumiblocks during the collisions are recorded. In this way, the performance of the sensors can be studied as a function of the voltage supplied with real data.

A.2.1 EFFICIENCY VS. HV

The efficiency of the pixel 3D sensors as a function of the voltage had never been studied before. However, there are a few issues that make the efficiency of the IBL 3D sensors particularly hard to measure:

- The standard efficiency measurement approach, as performed with other sections of the Pixel Detector, is not reliable when measuring for the IBL since it is based on track reconstruction (consists in looking for hits assigned to tracks, so the track can be extrapolated and count the number of times when a hit is expected on a given layer but is not found). Given that the seed for the track reconstruction algorithm (see Chapter 4) is set on the IBL, this method would be biased.
- The location of the 3D sensors at the high- $|\eta|$ region implies very low statistics (few thousands of tracks are measured per lumiblock). But such location is also highly sensitive to bad-quality tracks, such as those coming from pile-up events, hence a reasonably tight track selection is needed, which leads to even lower statistics.

An alternative approach was proposed for measuring the efficiency of the 3D sensors. The method is based on the use of the overlaps between staves (Fig. A.1b), so tracks with at least one hit in the IBL are selected and the relative efficiency is then calculated as

$$\varepsilon = \frac{\text{tracks with } \geq 2 \text{ hits}}{\text{tracks with } \geq 1 \text{ hits}}. \quad (\text{A.1})$$

A HV scan performed in May 2017 (namely Run 324340) was used to test the efficiency calculation defined above. Tracks with $p_T > 1$ GeV, $d_0 < 2$ mm and with at least 8 hits on the SCT were selected. Figure A.4a shows the efficiencies, inclusive in V , calculated for the different modules. In the figure, a fast drop of the efficiency is noted for the 3D modules the higher- $|\eta|$ (modules are numbered from -10 to 9 corresponding to their location on the staffe from the A to the C side: the 3D modules are the first and last four ones, denoted by the grey area), caused by the lack of statistics at such external regimes. This was tried to fix by adding the outlier hits (those discarded during the track

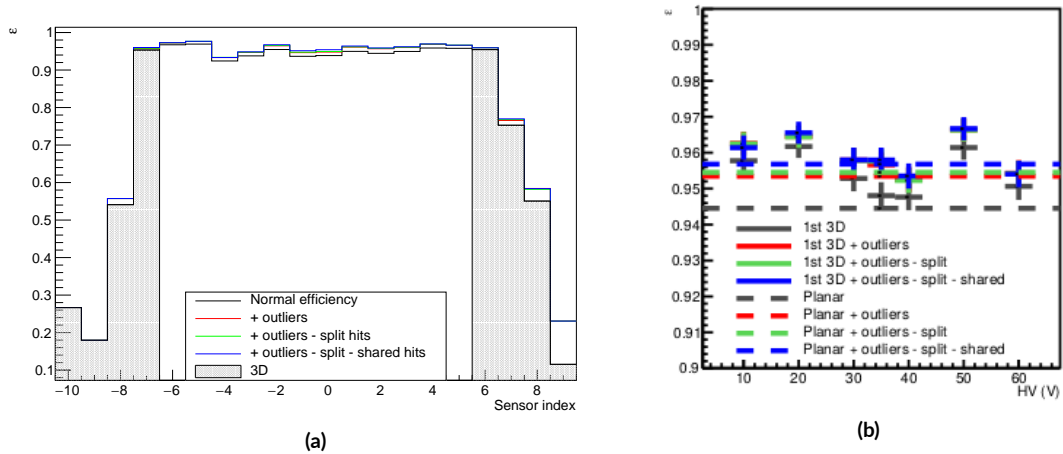


Figure A.4: (a) Sensor efficiency as calculated via the stave-overlap approach for each IBL module. The 3D sensors are indicated by the grey shaded area. (b) Efficiency as a function of HV measured for the most central 3D module on each side and compared to planar.

reconstruction due to fit-quality criteria) to the reconstructed track, and removing split and shared hits as well. However, the results did not change much. The conclusion is that the only 3D modules for which the efficiency calculation is reliable, and can be compared to that of the planars, are the most central ones. Figure A.4b shows the efficiency as a function of the voltage as calculated only such 3D sensors, and compared to the planar efficiency (planar sensors were kept operating at nominal voltages during the HV scan). The results, although seem to indicate slightly higher efficiencies of 3D sensors compared to planars, do not reveal any clear trend with the bias voltage.

The conclusion of the study is that, with the available statistics at that time another method must be found to measure the efficiency of the 3D sensors in a way that is less dependent on the track reconstruction efficiency. Otherwise, longer HV scans with larger periods of lumiblocks per HV step would allow to apply a tighter track selection (tracks with $p_T > 5$ GeV would be needed, according to previous efficiency measurements on the IBL [193]), which is crucial to ensure the validity of the proposed method.

A.2.2 CLUSTER SIZE, ToT AND CHARGE COLLECTION VS HV

Clusters are formed by grouping individual nearby pixels from the same sensor that get triggered by a crossing charged particle. The average size of the clusters in both longitudinal and transverse directions gives a hint of the performance of a sensor. On the other hand, the deposited charge is estimated in terms of the time-over-threshold (ToT), which is essentially the duration time of the electric pulse

emitted by the pixel, and is measured in units of the LHC bunch-crossing rate 25 ns. The conversion between ToT and collected charge is tuned specifically for each type of sensor.

A study of the average cluster size, ToT and collected charge as a function of the voltage supplied was performed using a HV scan performed in November 2017, namely Run 339957, in addition to the previously mentioned Run 324340 from May 2017. In the study, two different track selection criteria were tested: first, tracks with $p_T > 1$ GeV, $d_0 < 2$ mm and with at least 2 Pixel hits; and, secondly, tracks fulfilling the previous requirements plus having at least 8 hits on the SCT.

The results are shown in Figures A.5 and A.6. In all cases the additional cut on the number of SCT hits seems not to improve the results. Therefore one of the conclusions of the study is to stop using this kind of cuts in 3D sensors performance studies as it would only reduce the available statistics. A clear direct dependence is observed in both cluster size and charge collection with the voltage. Finally, comparing the two runs shows that such dependence with the voltage is enhanced with time, and therefore with the accumulated exposure of the pixel sensors to irradiation. Both cluster size and deposited charge show smaller values, especially the later. Altogether, the results seem to indicate a potential radiation damage effect on the 3D sensors.

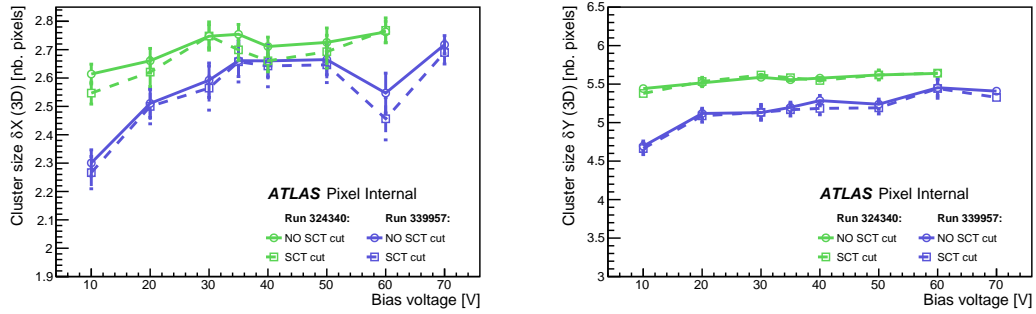


Figure A.5: Average cluster size of IBL 3D sensors as a function of the bias voltage in the (left) transverse and (right) longitudinal directions, as measured with two different HV scan runs.

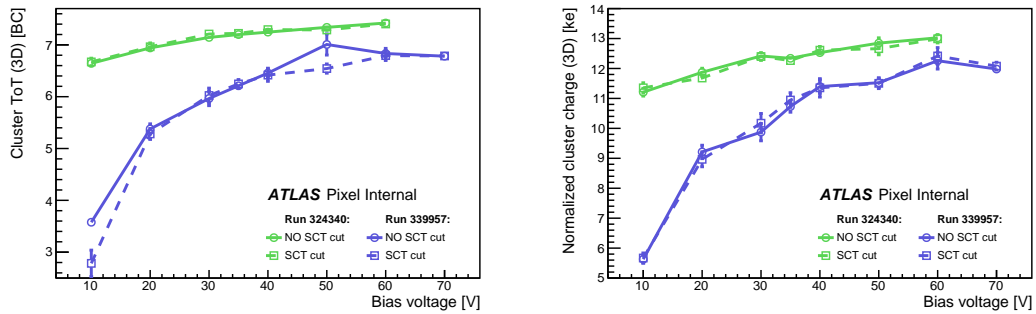


Figure A.6: (Left) ToT and (b) collected charge of IBL 3D sensors as a function of the bias voltage, as measured with two different HV scan runs.

B

Anti-SFs calculation method

As explained in Chapter 2, potential mismodelings on the lepton and photon reconstruction or identification are corrected by SFs that are calculated as the ratio of the reconstruction or identification efficiency in MC compared to data. The overall effect on an MC sample when the SFs are applied is to reduce or increase the number of events where the corresponding objects are reconstructed and identified. This means, however, that the number of events that can be expected to be found in the lepton- or photon-veto regions should be varied consequently. In order to account for this effect, the analysis applies sample-level *anti-scale factors* (anti-SFs) to correct the number of events falling into any veto region.

Anti-SFs are calculated for each background and signal process considered in the analysis. A separate anti-SF is computed for each vetoed object (electrons, muons, photons, and taus) using the procedure described below:

$$\text{anti-SF} = 1 + (1 - \langle \text{SF}_{\text{Tot}} \rangle) \frac{N_{>0\ell}}{N_{0\ell}} \quad (\text{B.1})$$

1. A SR-like region is constructed using the SR selection except for the lepton veto. For each event with *at least one baseline lepton* the mean value of the baseline lepton SFs, $\sum_i \text{SF}_i^n / n$,

where n is the total number of leptons in the event, is recorded. The weighted total number of events with at least one lepton ($N_{>0\ell}$) and the weighted number with zero leptons ($N_{0\ell}$) are also recorded.

2. The mean of the SFs is calculated as $\langle \text{SF}_{\text{Tot}} \rangle$.
3. The anti-SF are applied to each channel according to Eq. (B.1).

Note that this procedure* uses the baseline definitions of leptons and photons, since those are the vetoed objects in the analysis. Three sets of anti-SF are evaluated for each particle, accounting for the different $E_{\text{T}}^{\text{miss}}$ definition in each region (electrons and muons are treated as invisible particles in the electron and muon CRs, respectively, i.e. the lepton- p_{T} is added to the $E_{\text{T}}^{\text{miss}}$ in such regions). Anti-SFs binned in $E_{\text{T}}^{\text{miss}}$ were found to be compatible, within statistical uncertainties, with a unique value across bins, hence a single anti-SFs is then calculated for each background and signal process†.

There are several minor changes in this procedure compared to the previous iteration of the analysis. Regarding the use of the baseline SFs in the anti-SF calculation, the product of all SFs in the event $\prod_i \text{SF}_i$ was used in the past. This implies that only those cases when *all leptons* are identified were taken into account, not considering other possible combinations. However, taking the mean value of the SFs gives a better approximation to the actual probability of having *at least one* of the leptons identified, which is already enough for vetoing the event.

The anti-SF procedure is sensitive to cases where most of the events in a MC sample contain a reconstructed and identified baseline lepton, i.e. when the efficiencies are naturally high. To account for a slight mismodeling in the large number of events with the identified leptons, a large anti-SF must correct the relatively tiny number of events in the veto-region. In other words, the calculation is statistically sensitive due to the ratio $N_{>0\ell}/N_{0\ell}$ in Eq. (B.1). This issue happens for the di-lepton background samples, for which the limited MC statistics lead to nonphysical large values. In order to solve this, the conservative strategy chosen is to take as the anti-SF the value from the equivalent W flavour sample (e.g. the anti-SF computed for the $W(e\nu) + \text{jets}$ sample is applied to the $Z(ee) + \text{jets}$ one).

Table B.1 shows some examples of the total anti-SF values applied for some representative processes. In the table, the total product of all the anti-SFs utilized in the corresponding region is reported. As it can be seen, the corrections values are almost negligible, with an overall impact on the total SM

*Eq. B.1 is derived from the assumptions $N_{\text{Tot}} = N_{>0\ell} + N_{0\ell}$ and $N_{\text{Tot}} = \langle \text{SF}_{\text{Tot}} \rangle N_{>0\ell} + \text{anti-SF} \cdot N_{0\ell}$, where N_{Tot} is the total number of events of the sample.

†In addition, the impact of a $E_{\text{T}}^{\text{miss}}$ -dependent anti-SFs implementation was found to be of less than 0.05% on the SM predictions in the SR.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$	SR
$Z(\nu\nu) + \text{jets}$	1	1	1	1	1
$W(\mu\nu) + \text{jets}$	1.00041	1.01421	1.00041	1.01421	1.01464
$W(e\nu) + \text{jets}$	1.10824	1.00001	1.10824	1.00001	1.10824
$W(\tau\nu) + \text{jets}$	1.00411	0.998237	1.00411	0.998237	1.00571
Diboson	1.00800	1.00073	1.00800	1.00073	1.01333
ADD ($d = 2, M_D = 5300$)	1.00017	1.00001	1.00017	1.00001	1.00017
DM $m(\chi, Z_A) = (1, 2000)$ GeV	1.00012	0.999997	1.00012	0.999997	1.00012
TT $m(\tilde{t}, \tilde{\chi}^0) = (550, 543)$ GeV	1.00018	1.00002	1.00018	1.00002	1.00017
H(\rightarrow inv)	1.00072	0.999981	1.00072	0.999981	1.00083

Table B.1: Global anti-SFs applied to some background and signal samples in the control regions. The reported numbers correspond to the samples generated for the emulation of the 2018 data-taking period

prediction of less than 1%. The only apparent exception are those of the $W(e\nu) + \text{jets}$ sample reported for the muon and signal regions, where this process is absolutely negligible.

C

NCB studies

Non-collision background jets can mimic the signature of a high-pt jet recoiling against E_T^{miss} . There are two sources of NCB to be considered: beam-induced background (BIB), where muons are created via inelastic beam-gas interactions or losses on the LHC collimators, and cosmic ray backgrounds. In both cases, the muons traverse the ATLAS detector and may deposit significant amount of energy in the calorimeters. TIGHT jet cleaning criteria (see Chapter 4) have been deployed to efficiently reject NCB in the monojet SR, suppressing non-collision rates by $O(10^3)$. Figure C.1 shows the distributions of the leading jet p_T , φ and timing before and after the TIGHT jet cleaning criteria is applied. In the figure it can be seen that NCB jets have a mostly homogeneous p_T distribution, while the φ distribution exhibit peaks at $\varphi = 0$ and $\varphi = \pi$, characteristic of cosmic muons.

The residual NCB in the SR is estimated using a data-driven method which relies on the characteristic difference in jet timing t_{jet} for non-collision jets and jets originating from pp collisions. The timing t_{jet} is calculated from the energy-weighted average of the individual times of the jet energy deposits, taking as reference the time recorded by the trigger. The zero-time for each cell is set to when a light-speed particle from the interaction point would reach that cell. Jets with $|t_{\text{jet}}| > 5$ ns are predominantly non-collision jets (as can be seen in Fig. C.1). The NCB estimation method is performed in each bin of E_T^{miss} as follows:

Tagging selection	ε_{NCB} [%]	$N^{\text{NCB,SR}}$
2015 + 2016		
$t_{\text{jet}} < -5$ ns	36.40 ± 0.05	1434 ± 63
$t_{\text{jet}} > 5$ ns	7.84 ± 0.02	1977 ± 158
2017		
$t_{\text{jet}} < -5$ ns	34.70 ± 0.07	1153 ± 58
$t_{\text{jet}} > 5$ ns	5.68 ± 0.02	3098 ± 234
2018		
$t_{\text{jet}} < -5$ ns	30.41 ± 0.05	2263 ± 86
$t_{\text{jet}} > 5$ ns	5.67 ± 0.02	4952 ± 296

Table C.1: NCB tagging efficiency and total amount of NCB in the SR in the different years of data-taking. Errors shown are statistical uncertainties.

1. A region enriched in NCB is constructed using the SR selection but inverting the NCB-rejection cut (given by Eq. (4.8)).
2. The number of events is counted in the NCB region: N^{NCB} .
3. The number of events with $t_{\text{jet}} < -5$ ns, also known as “out-of-time” (OOT) jets, is counted in the NCB region: $N_{\text{OOT}}^{\text{NCB}}$.
4. The ratio between the number of OOT jet events and the total number of events in the NCB region is taken as the NCB tagging efficiency: $\varepsilon_{\text{NCB}} = N_{\text{OOT}}^{\text{NCB}}/N^{\text{NCB}}$.
5. The number of OOT events is counted in the signal region: $N_{\text{OOT}}^{\text{SR}}$.
6. Finally, the NCB estimate is obtained by scaling the number of tagged events in the signal region by the inverse of the tagging efficiency: $N^{\text{NCB,SR}} = N_{\text{OOT}}^{\text{SR}}/\varepsilon_{\text{NCB}}$.

As the kinematics for positively and negatively OOT jets are different [154], a validation check is performed using $t_{\text{jet}} > 5$ ns. The obtained results are shown in Table C.1 for the inclusive SR and in bins of $E_{\text{T}}^{\text{miss}}$ in Table C.2. The estimated NCB contributions are consistent over the years and negligible for $E_{\text{T}}^{\text{miss}} > 400$ GeV in the SR.

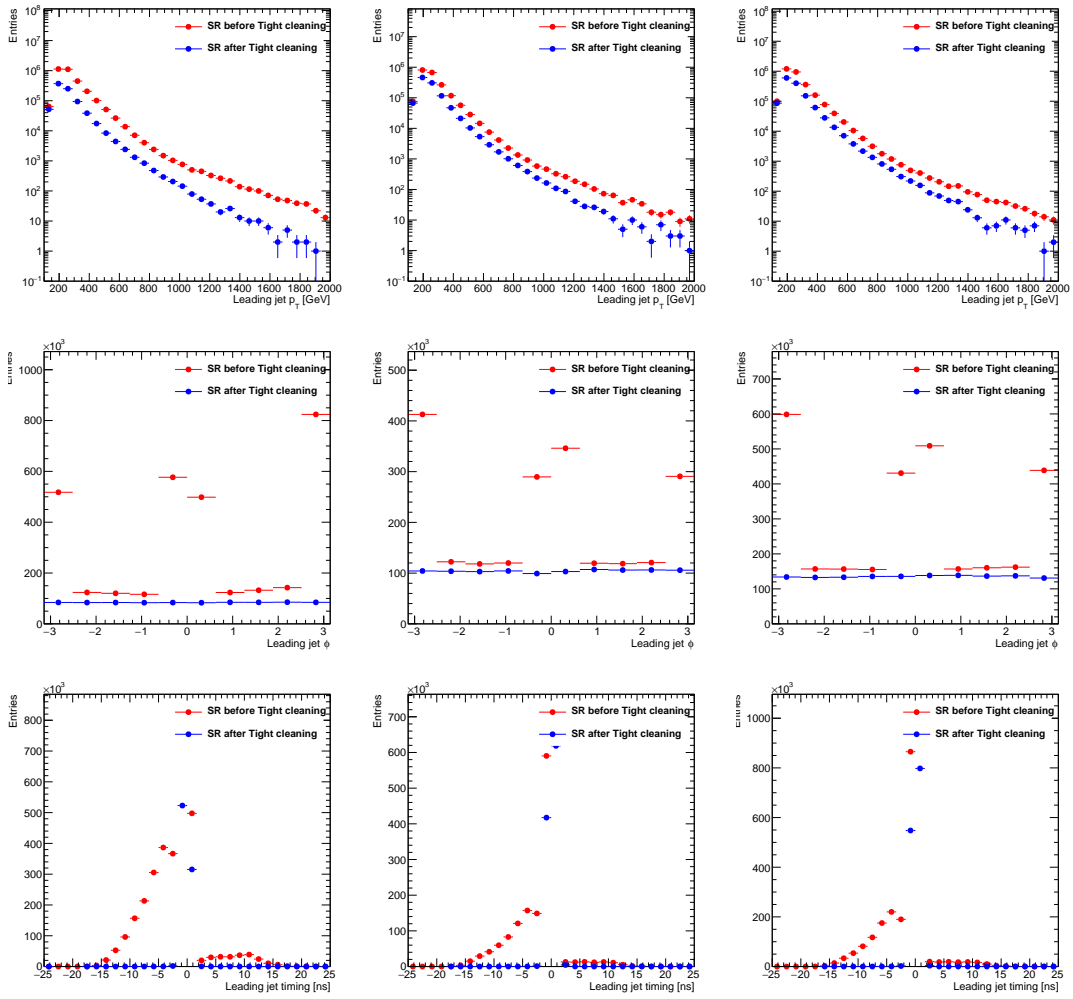


Figure C.1: Leading jet p_T , ϕ and timing in the signal region before and after Tight jet cleaning. Left: 2015+2016 data; center: 2017 data; right: 2018 data.

E_T^{miss} bin in SR	$N^{\text{NCB,SR}}$ using $t_{\text{jet}} < -5$ ns	$N^{\text{NCB,SR}}$ using $t_{\text{jet}} > 5$ ns
2015 + 2016		
$200 \text{ GeV} < E_T^{\text{miss}} \leq 250 \text{ GeV}$	1039 ± 19	1420 ± 10
$250 \text{ GeV} < E_T^{\text{miss}} \leq 300 \text{ GeV}$	266 ± 9	297 ± 5
$300 \text{ GeV} < E_T^{\text{miss}} \leq 350 \text{ GeV}$	74 ± 5	105 ± 3
$350 \text{ GeV} < E_T^{\text{miss}} \leq 400 \text{ GeV}$	24 ± 3	41 ± 1
$400 \text{ GeV} < E_T^{\text{miss}} \leq 500 \text{ GeV}$	13 ± 2	0 ± 0
$500 \text{ GeV} < E_T^{\text{miss}} \leq 600 \text{ GeV}$	2 ± 1	0 ± 0
$600 \text{ GeV} < E_T^{\text{miss}} \leq 700 \text{ GeV}$	0 ± 0	0 ± 0
$E_T^{\text{miss}} > 700 \text{ GeV}$	0 ± 0	0 ± 0
2017		
$200 \text{ GeV} < E_T^{\text{miss}} \leq 250 \text{ GeV}$	845 ± 16	2112 ± 11
$250 \text{ GeV} < E_T^{\text{miss}} \leq 300 \text{ GeV}$	191 ± 8	445 ± 5
$300 \text{ GeV} < E_T^{\text{miss}} \leq 350 \text{ GeV}$	61 ± 4	128 ± 2
$350 \text{ GeV} < E_T^{\text{miss}} \leq 400 \text{ GeV}$	12 ± 2	20 ± 1
$400 \text{ GeV} < E_T^{\text{miss}} \leq 500 \text{ GeV}$	14 ± 2	66 ± 1
$500 \text{ GeV} < E_T^{\text{miss}} \leq 600 \text{ GeV}$	0 ± 0	0 ± 0
$600 \text{ GeV} < E_T^{\text{miss}} \leq 700 \text{ GeV}$	0 ± 0	0 ± 0
$E_T^{\text{miss}} > 700 \text{ GeV}$	0 ± 0	0 ± 0
2018		
$200 \text{ GeV} < E_T^{\text{miss}} \leq 250 \text{ GeV}$	1629 ± 22	3231 ± 13
$250 \text{ GeV} < E_T^{\text{miss}} \leq 300 \text{ GeV}$	324 ± 9	853 ± 7
$300 \text{ GeV} < E_T^{\text{miss}} \leq 350 \text{ GeV}$	108 ± 5	148 ± 3
$350 \text{ GeV} < E_T^{\text{miss}} \leq 400 \text{ GeV}$	51 ± 3	172 ± 2
$400 \text{ GeV} < E_T^{\text{miss}} \leq 500 \text{ GeV}$	19 ± 2	0 ± 0
$500 \text{ GeV} < E_T^{\text{miss}} \leq 600 \text{ GeV}$	13 ± 2	59 ± 1
$600 \text{ GeV} < E_T^{\text{miss}} \leq 700 \text{ GeV}$	8 ± 1	0 ± 0
$E_T^{\text{miss}} > 700 \text{ GeV}$	0 ± 0	0 ± 0

Table C.2: Total amount of NCB in the specified regions. Errors shown are statistical uncertainties.

D

Additional Material from the Background-Only Fit

This appendix is devoted to the material that was not shown in Chapter 8 in relation to the results obtained from the background-only fit. The exact values of the fitted NPs are shown in Tables D.1 and D.2, and the CRs event yields in exclusive bins of E_T^{miss} are shown in Tables D.3–D.11.

Nuisance Parameter	Value	Upper unc.	Lower unc.
EG_RESOLUTION_ALL	-0.00892	1.01865	-1.01753
EG_SCALE_ALL	-0.00014	0.72780	-0.69614
EL_EFF_ID_CorrUncertaintyNP10_Electron_CRs	0.00974	0.89842	-0.90009
EL_EFF_ID_CorrUncertaintyNP11_Electron_CRs	0.05919	0.93220	-0.93258
EL_EFF_ID_CorrUncertaintyNP12_Electron_CRs	-0.32996	0.91585	-0.90189
EL_EFF_ID_CorrUncertaintyNP13_Electron_CRs	0.17298	0.92886	-0.93591
EL_EFF_ID_CorrUncertaintyNP14_Electron_CRs	-0.09435	0.93139	-0.91816
EL_EFF_ID_CorrUncertaintyNP15_Electron_CRs	-0.19701	0.95313	-0.95533
EL_EFF_ID_CorrUncertaintyNP8_Electron_CRs	0.03002	0.95609	-0.95643
EL_EFF_ID_CorrUncertaintyNP9_Electron_CRs	0.01825	0.96949	-0.96666
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP17_Electron_CRs	0.33869	0.89285	-0.97876
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP6_Electron_CRs	0.00130	1.00105	-1.00107
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP8_Electron_CRs	0.42944	0.88514	-0.89036
EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR_NOT_Electron_CRs	-0.01955	0.99581	-0.99536
EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR	0.22024	0.98111	-0.98273
EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	0.11213	0.95430	-0.95060
FT_EFF_B_systematics	0.07411	0.95759	-0.95279
FT_EFF_C_systematics	-0.12617	0.97210	-0.96712
FT_EFF_Light_systematics	0.03695	0.87202	-0.86901
FT_EFF_extrapolation	-0.02515	0.80914	-0.80969
JET_EffectiveNP_Detector1	-0.02183	0.96378	-0.96786
JET_EffectiveNP_Mixed1	-0.18873	0.99511	-0.92767
JET_EffectiveNP_Mixed2	0.02291	1.02602	-1.02493
JET_EffectiveNP_Mixed3	0.03710	0.99003	-0.98816
JET_EffectiveNP_Modelling1	0.14316	0.63701	-0.52147
JET_EffectiveNP_Modelling2	-0.04237	1.03259	-1.01462
JET_EffectiveNP_Modelling3	0.03475	0.98227	-0.97754
JET_EffectiveNP_Modelling4	-0.02571	1.00122	-1.00132
JET_EffectiveNP_Statistical1	-0.14050	1.05304	-1.00581
JET_EffectiveNP_Statistical2	-0.02508	0.94635	-0.95231
JET_EffectiveNP_Statistical3	0.03238	0.91213	-0.90310
JET_EffectiveNP_Statistical4	-0.02452	0.98617	-0.98710
JET_EffectiveNP_Statistical5	-0.01260	1.01091	-1.00919
JET_EffectiveNP_Statistical6	-0.01242	0.99697	-0.99658
JET_EtaIntercalibration_Modelling	0.00356	0.86274	-0.86412
JET_EtaIntercalibration_NonClosure_negEta	0.00006	1.00144	-1.00146
JET_EtaIntercalibration_NonClosure_posEta	0.00712	1.00979	-1.00995
JET_EtaIntercalibration_TotalStat	-0.03930	0.95614	-0.96345
JET_Flavor_Composition	-0.19389	0.69705	-0.72080
JET_Flavor_Response	-0.32216	0.72642	-0.85027
JET_FullJER_DataVsMC_MC16	0.21022	1.00580	-1.04412
JET_FullJER_EffectiveNP_1	0.14166	0.98211	-0.98361
JET_FullJER_EffectiveNP_10	0.42063	0.97410	-0.97832
JET_FullJER_EffectiveNP_11	0.05694	0.99825	-0.99901
JET_FullJER_EffectiveNP_12restTerm	0.27266	0.99089	-0.99620
JET_FullJER_EffectiveNP_2	-0.00368	1.10034	-1.11258
JET_FullJER_EffectiveNP_3	-0.14851	0.98006	-0.97864
JET_FullJER_EffectiveNP_4	-0.12813	0.95251	-0.96054
JET_FullJER_EffectiveNP_5	-0.80282	1.12955	-0.87745

Table D.1: Values and uncertainties of the fitted nuisance parameters from the background-only fit.

Nuisance Parameter	Value	Upper unc.	Lower unc.
JET_FullJER_EffectiveNP_6	0.05038	0.97058	-0.95324
JET_FullJER_EffectiveNP_7	-0.54164	1.41710	-0.92270
JET_FullJER_EffectiveNP_8	0.22190	0.99012	-1.00109
JET_JvtEfficiency	0.01686	1.00326	-1.00630
JET_Pileup_OffsetMu	-0.04286	0.98447	-0.99211
JET_Pileup_OffsetNPV	0.35196	0.98917	-1.13251
JET_Pileup_PtTerm	0.07071	1.03810	-1.06337
JET_Pileup_RhoTopology	0.06609	1.04691	-1.08014
JET_PunchThrough_MC16	-0.02457	0.98023	-0.98292
MET_SoftTrk_ResoPara	-0.39287	0.96993	-0.97102
MET_SoftTrk_ResoPerp	0.34285	0.98513	-0.98242
MET_SoftTrk_Scale	0.17950	1.00498	-1.18444
MUON_EFF_ID_HighPT_pt0_etao	0.00014	1.00105	-1.00105
MUON_EFF_ID_HighPT_pt1_etao	0.32682	0.98818	-0.98849
MUON_EFF_ID_HighPT_pt1_eta1	0.00004	1.00106	-1.00106
MUON_EFF_ID_HighPT_pt2_etao	0.25262	0.96103	-0.95985
MUON_EFF_ID_HighPT_pt2_eta1	0.13441	0.99390	-0.99408
MUON_EFF_ID_HighPT_pt3_etao	0.06139	0.92904	-0.92774
MUON_EFF_ID_HighPT_pt3_eta1	0.06865	0.99415	-0.99397
MUON_EFF_ID_HighPT_pt4_etao	-0.01155	0.84576	-0.82730
MUON_EFF_ID_HighPT_pt4_eta1	0.04926	0.99379	-0.99370
MUON_EFF_RECO_SYS	-0.54792	0.62635	-0.64758
MUON_EFF_RECO_SYS_LOWPT	0.00000	1.00000	-1.00000
MUON_ID	-0.03248	1.01293	-1.00917
MUON_MS	-0.22973	0.99224	-0.99093
MUON_SAGITTA_RESBIAS	-0.37268	0.80054	-0.80351
MUON_SCALE	-0.00764	0.99894	-0.99835
NCB_Sys	0.00000	1.00000	-1.00000
PRW_DATASF	0.50066	0.91437	-1.30127
TAUS_TRUEHADTAU_EFF_JETID_SYST	-0.02360	1.00042	-0.99996
VBF_theo_sys	0.95499	0.47348	-0.70830
diboson_Sys	-0.13171	0.83654	-0.89400
lumiSys	0.00195	0.99492	-1.00033
multijet_Sys	0.00000	1.00000	-1.00000
singletop_singletop_combined_Sys	-1.57696	0.29096	-0.46215
taus_PTV_definition	0.04031	1.00077	-1.00027
ttbar_ttbar_combined_Sys	0.13033	0.52670	-0.46118
vjets_d1K_NNLO	-0.09847	0.87725	-0.87581
vjets_d1kappa_EW	0.11350	0.99026	-0.98783
vjets_d2K_NNLO	-0.14087	0.89407	-0.89512
vjets_d2kappa_EW_W	-0.16003	0.89416	-0.89739
vjets_d2kappa_EW_Zll	-0.12216	0.92249	-0.91963
vjets_d2kappa_EW_Zvv	0.00000	1.00000	-1.00000
vjets_d3K_NNLO	-0.90058	0.66801	-0.67910
vjets_d3kappa_EW_W	-0.01722	0.99155	-0.99153
vjets_d3kappa_EW_Z	-0.08450	0.96491	-0.96444
vjets_dK_NNLO_mix	-0.32425	0.87578	-0.88754
vjets_dK_PDF	0.19789	0.96931	-0.95813

Table D.2: Values and uncertainties of the fitted nuisance parameters from the background-only fit (part 2).

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	330096	174890	62421	47931	36162
Background post-fit prediction	330 800±830	175 010±470	62 250±300	48 140±160	36 260±130
$W(\mu\nu) + \text{jets}$	276 300±1 600	–	8 030±430	–	–
$W(e\nu) + \text{jets}$	–	143 510±860	4 200±230	–	–
$W(\tau\nu) + \text{jets}$	17 220±210	11 260±120	843±48	–	–
$Z(\mu\mu) + \text{jets}$	4 860±120	–	180.2±4.7	45 160±220	–
$Z(ee) + \text{jets}$	–	–	–	–	34 400±170
$Z(\tau\tau) + \text{jets}$	–	468±12	–	–	–
Diboson	5 930±920	3 150±470	481±84	1 020±170	740±120
single- t	5 500±1 500	3 520±950	8 800±2 600	84±43	35±18
$t\bar{t}$	14 700±1 200	9 710±790	39 200±2 600	1 060±100	482±46
VBF $Z(\ell\ell/\nu\nu)+\text{jets}$	–	–	–	814±95	609±72
VBF $W(\ell\nu)+\text{jets}$	6 340±720	3 390±380	475±71	–	–

Table D.3: Predicted (post-fit) and observed background events in the control regions, in the exclusive p_T^{recoil} bin EM1.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	146813	86339	29551	21794	17973
Background post-fit prediction	147 030±430	86 260±260	29 440±150	21 677±85	17 767±75
$W(\mu\nu) + \text{jets}$	121 750±820	–	3 920±200	–	–
$W(e\nu) + \text{jets}$	–	69 890±490	2 220±110	–	–
$W(\tau\nu) + \text{jets}$	7 727±93	5 480±63	457±24	–	–
$Z(\mu\mu) + \text{jets}$	1 744±37	–	70.2±1.9	20 160±120	–
$Z(ee) + \text{jets}$	–	–	–	–	16 700±100
$Z(\tau\tau) + \text{jets}$	–	200.0±5.3	–	–	–
Diboson	3 200±510	1 870±290	301±55	568±99	438±75
single- t	2 320±690	1 620±480	3 800±1 300	34±19	23±11
$t\bar{t}$	6 750±560	5 100±420	18 400±1 200	442±49	226±27
VBF $Z(\ell\ell/\nu\nu)+\text{jets}$	–	–	–	473±59	378±48
VBF $W(\ell\nu)+\text{jets}$	3 550±430	2 100±260	305±50	–	–

Table D.4: Predicted (post-fit) and observed background events in the control regions, in the exclusive p_T^{recoil} bin EM2.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	68951	43781	13592	10390	9065
Background post-fit prediction	$68\,650 \pm 200$	$43\,710 \pm 130$	$13\,829 \pm 82$	$10\,336 \pm 44$	$8\,987 \pm 48$
$W(\mu\nu) + \text{jets}$	$56\,560 \pm 440$	–	$1\,990 \pm 100$	–	–
$W(e\nu) + \text{jets}$	–	$35\,130 \pm 290$	$1\,281 \pm 64$	–	–
$W(\tau\nu) + \text{jets}$	$3\,797 \pm 48$	$2\,851 \pm 35$	229 ± 11	–	–
$Z(\mu\mu) + \text{jets}$	657 ± 11	–	27.78 ± 0.75	$9\,532 \pm 70$	–
$Z(ee) + \text{jets}$	–	–	–	–	$8\,374 \pm 65$
$Z(\tau\tau) + \text{jets}$	–	106.2 ± 2.9	–	–	–
Diboson	$1\,730 \pm 290$	$1\,130 \pm 180$	164 ± 32	316 ± 57	263 ± 46
single- t	970 ± 320	750 ± 240	$1\,560 \pm 610$	15.9 ± 9.2	7.8 ± 4.4
$t\bar{t}$	$2\,920 \pm 240$	$2\,410 \pm 200$	$8\,400 \pm 560$	197 ± 24	111 ± 16
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	275 ± 37	231 ± 31
VBF $W(\ell\nu) + \text{jets}$	$2\,020 \pm 260$	$1\,330 \pm 170$	172 ± 31	–	–

Table D.5: Predicted (post-fit) and observed background events in the control regions, in the exclusive $p_{\text{T}}^{\text{recoil}}$ bin EM3.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	15339	11695	2596	2388	2476
Background post-fit prediction	$15\,667 \pm 65$	$11\,792 \pm 69$	$2\,576 \pm 31$	$2\,317 \pm 22$	$2\,430 \pm 23$
$W(\mu\nu) + \text{jets}$	$12\,810 \pm 160$	–	490 ± 26	–	–
$W(e\nu) + \text{jets}$	–	$9\,400 \pm 120$	346 ± 19	–	–
$W(\tau\nu) + \text{jets}$	887 ± 15	761 ± 14	62.2 ± 3.0	–	–
$Z(\mu\mu) + \text{jets}$	116.5 ± 1.7	–	5.77 ± 0.18	$2\,072 \pm 30$	–
$Z(ee) + \text{jets}$	–	–	–	–	$2\,207 \pm 30$
$Z(\tau\tau) + \text{jets}$	–	37.5 ± 1.1	–	–	–
Diboson	550 ± 100	408 ± 72	55 ± 13	103 ± 21	95 ± 18
single- t	124 ± 56	95 ± 48	190 ± 110	1.9 ± 1.6	0.71 ± 0.66
$t\bar{t}$	429 ± 38	485 ± 44	$1\,348 \pm 99$	38.6 ± 6.4	27.9 ± 5.4
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	101 ± 17	99 ± 16
VBF $W(\ell\nu) + \text{jets}$	750 ± 120	609 ± 100	84 ± 22	–	–

Table D.6: Predicted (post-fit) and observed background events in the control regions, in the exclusive $p_{\text{T}}^{\text{recoil}}$ bin EM5.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	5358	4392	767	774	834
Background post-fit prediction	$5\,415 \pm 43$	$4\,384 \pm 36$	776 ± 15	798 ± 11	899 ± 11
$W(\mu\nu) + \text{jets}$	$4\,396 \pm 74$	–	161.3 ± 8.7	–	–
$W(e\nu) + \text{jets}$	–	$3\,462 \pm 59$	133.2 ± 7.9	–	–
$W(\tau\nu) + \text{jets}$	289.6 ± 6.1	288.8 ± 6.8	23.2 ± 1.2	–	–
$Z(\mu\mu) + \text{jets}$	37.85 ± 0.72	–	1.930 ± 0.070	709 ± 15	–
$Z(ee) + \text{jets}$	–	–	–	–	803 ± 14
$Z(\tau\tau) + \text{jets}$	–	12.02 ± 0.40	–	–	–
Diboson	222 ± 45	168 ± 32	23.4 ± 6.4	37.3 ± 8.2	41.9 ± 8.7
single- t	32 ± 18	19 ± 12	38 ± 28	–	–
$t\bar{t}$	119 ± 11	164 ± 16	355 ± 28	9.1 ± 1.8	8.8 ± 1.9
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	42.6 ± 8.1	44.4 ± 8.3
VBF $W(\ell\nu) + \text{jets}$	320 ± 58	270 ± 50	41 ± 13	–	–

Table D.7: Predicted (post-fit) and observed background events in the control regions, in the exclusive p_T^{recoil} bin EM6.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	2156	1868	305	306	357
Background post-fit prediction	$2\,087 \pm 25$	$1\,862 \pm 25$	286.9 ± 10.0	308.6 ± 6.5	367.0 ± 5.8
$W(\mu\nu) + \text{jets}$	$1\,654 \pm 36$	–	65.7 ± 4.0	–	–
$W(e\nu) + \text{jets}$	–	$1\,447 \pm 33$	55.5 ± 3.7	–	–
$W(\tau\nu) + \text{jets}$	124.7 ± 3.2	118.5 ± 3.5	10.40 ± 0.59	–	–
$Z(\mu\mu) + \text{jets}$	14.81 ± 0.36	–	0.850 ± 0.040	269.3 ± 7.6	–
$Z(ee) + \text{jets}$	–	–	–	–	321.8 ± 7.3
$Z(\tau\tau) + \text{jets}$	–	5.03 ± 0.19	–	–	–
Diboson	96 ± 21	92 ± 19	10.5 ± 3.3	16.4 ± 3.9	18.1 ± 4.1
single- t	4.7 ± 4.1	5.4 ± 4.3	4.9 ± 5.2	–	–
$t\bar{t}$	38.7 ± 3.9	57.3 ± 5.9	119 ± 12	2.24 ± 0.48	4.8 ± 1.0
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	20.6 ± 4.4	22.2 ± 4.6
VBF $W(\ell\nu) + \text{jets}$	154 ± 31	136 ± 28	19.7 ± 7.3	–	–

Table D.8: Predicted (post-fit) and observed background events in the control regions, in the exclusive p_T^{recoil} bin EM7.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	438	389	51	61	81
Background post-fit prediction	409±11	399.4±9.1	55.5±4.2	57.7±2.3	78.8±2.1
$W(\mu\nu)$ + jets	312±11	–	12.3±1.2	–	–
$W(e\nu)$ + jets	–	303±10	10.5±1	–	–
$W(\tau\nu)$ + jets	24.5±1.0	20.80±0.84	2.66±0.23	–	–
$Z(\mu\mu)$ + jets	2.76±0.11	–	–	48.2±2.4	–
$Z(ee)$ + jets	–	–	–	–	67.2±2.4
$Z(\tau\tau)$ + jets	–	0.830±0.040	–	–	–
Diboson	22.1±5.7	21.1±5.0	4.9±2.0	4.1±1.2	4.5±1.2
single- t	–	0.51±0.54	–	–	–
$t\bar{t}$	5.86±0.77	15.6±2.0	18.3±2.9	–	1.15±0.25
VBF $Z(\ell\ell/\nu\nu)$ +jets	–	–	–	5.1±1.3	6.0±1.6
VBF $W(\ell\nu)$ +jets	42±10	37.3±9.1	6.7±3.3	–	–

Table D.9: Predicted (post-fit) and observed background events in the control regions, in the exclusive p_T^{recoil} bin EM9.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	204	234	19	23	32
Background post-fit prediction	201.8±5.0	204.9±5.1	21.1±2.3	27.8±1.5	37.6±1.3
$W(\mu\nu)$ + jets	154.3±5.8	–	5.30±0.57	–	–
$W(e\nu)$ + jets	–	153.4±5.7	5.43±0.59	–	–
$W(\tau\nu)$ + jets	11.65±0.51	11.76±0.52	0.900±0.090	–	–
$Z(\mu\mu)$ + jets	1.200±0.050	–	–	23.3±1.5	–
$Z(ee)$ + jets	–	–	–	–	32.2±1.4
$Z(\tau\tau)$ + jets	–	–	–	–	–
Diboson	10.3±2.8	11.0±2.8	0.54±0.26	1.78±0.55	1.86±0.56
single- t	–	–	–	–	–
$t\bar{t}$	2.77±0.40	6.60±0.93	4.81±0.81	–	–
VBF $Z(\ell\ell/\nu\nu)$ +jets	–	–	–	2.74±0.78	3.38±0.96
VBF $W(\ell\nu)$ +jets	21.6±5.7	21.8±5.8	4.0±2.3	–	–

Table D.10: Predicted (post-fit) and observed background events in the control regions, in the exclusive p_T^{recoil} bin EM10.

Control Region	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$	Top	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$
Observed events	94	121	13	9	21
Background post-fit prediction	97.1 ± 3.1	105.8 ± 3.3	11.4 ± 1.8	15.1 ± 1.1	18.90 ± 0.92
$W(\mu\nu) + \text{jets}$	73.4 ± 3.3	–	2.78 ± 0.38	–	–
$W(e\nu) + \text{jets}$	–	78.9 ± 3.5	2.74 ± 0.38	–	–
$W(\tau\nu) + \text{jets}$	5.43 ± 0.28	5.61 ± 0.28	–	–	–
$Z(\mu\mu) + \text{jets}$	0.650 ± 0.030	–	–	12.3 ± 1	–
$Z(ee) + \text{jets}$	–	–	–	–	15.73 ± 0.88
$Z(\tau\tau) + \text{jets}$	–	–	–	–	–
Diboson	4.4 ± 1.3	5.0 ± 1.4	0.73 ± 0.40	1.14 ± 0.38	1.23 ± 0.41
single- t	–	–	–	–	–
$t\bar{t}$	0.94 ± 0.15	2.03 ± 0.30	1.97 ± 0.39	–	–
VBF $Z(\ell\ell/\nu\nu) + \text{jets}$	–	–	–	1.52 ± 0.47	1.94 ± 0.61
VBF $W(\ell\nu) + \text{jets}$	12.3 ± 3.6	14.1 ± 4.0	2.7 ± 1.7	–	–

Table D.11: Predicted (post-fit) and observed background events in the control regions, in the exclusive p_T^{recoil} bin EM11.

E

Systematics Impact

The exact values of the impact of each systematic uncertainty in every SR exclusive bin of E_T^{miss} are listed in Table 8.8.

Source [%]	EM ₀	EM ₁	EM ₂	EM ₃	EM ₄	EM ₅	EM ₆	EM ₇	EM ₈	EM ₉	EM ₁₀	EM ₁₁	EM ₁₂
Flavor tagging	0.10	0.13	0.15	0.17	0.19	0.25	0.33	0.41	0.49	0.57	0.62	0.70	0.87
Jet energy scale	0.17	0.20	0.27	0.25	0.28	0.39	0.50	0.63	0.80	0.89	0.90	1.00	0.97
Jet energy resolution	0.15	0.20	0.21	0.25	0.32	0.37	0.40	0.37	0.31	0.34	0.43	0.62	1.28
Jet JVT efficiency	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.0	0.01	0.01	0.02	0.03
Pile-up re-weighting	0.37	0.35	0.34	0.33	0.32	0.30	0.32	0.34	0.36	0.33	0.26	0.24	0.30
L_T^{miss} resolution	0.34	0.31	0.25	0.20	0.18	0.14	0.08	0.05	0.04	0.05	0.05	0.06	0.07
E_T^{miss} scale	0.53	0.46	0.40	0.36	0.31	0.27	0.26	0.26	0.25	0.26	0.27	0.27	0.30
e/γ energy resolution	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.08
e/γ energy scale	0.30	0.30	0.31	0.32	0.33	0.36	0.4	0.44	0.48	0.54	0.6	0.66	0.74
e identification efficiency	0.58	0.56	0.54	0.54	0.54	0.56	0.60	0.67	0.73	0.80	0.85	0.92	1.06
e reconstruction efficiency	0.16	0.16	0.16	0.16	0.16	0.15	0.15	0.16	0.16	0.17	0.18	0.18	0.2
μ identification efficiency	0.06	0.07	0.07	0.04	0.03	0.11	0.20	0.29	0.38	0.48	0.58	0.70	0.9
μ reconstruction efficiency	0.41	0.48	0.56	0.63	0.73	0.82	0.92	1.03	1.13	1.21	1.27	1.35	1.49
μ momentum scale	0.10	0.12	0.15	0.17	0.20	0.26	0.31	0.37	0.42	0.48	0.53	0.58	0.66
τ identification efficiency	0.10	0.10	0.09	0.09	0.08	0.07	0.06	0.06	0.05	0.06	0.06	0.06	0.07
Luminosity	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.03	0.05
Diboson theory	0.04	0.04	0.04	0.04	0.04	0.02	0.01	0.04	0.04	0.05	0.06	0.10	0.22
NCB	0.20	0.10	0.08	0.06	0.05	0.05	0.08	0.00	0.00	0.00	0.00	0.00	0.00
Multi-jet background	1.04	0.72	0.36	0.28	0.06	0.02	0.01	0.0	0.01	0.00	0.00	0.00	0.00
single- t theory	0.15	0.15	0.14	0.15	0.17	0.22	0.24	0.22	0.23	0.17	0.13	0.13	0.28
$t\bar{t}$ theory	0.07	0.07	0.06	0.06	0.08	0.13	0.18	0.22	0.24	0.33	0.4	0.52	0.71
$V+$ jets τ -lepton definition	0.05	0.04	0.04	0.04	0.04	0.04	0.06	0.07	0.08	0.10	0.12	0.16	0.13
$V+$ jets pure QCD corrections	0.24	0.28	0.33	0.39	0.44	0.50	0.61	0.73	0.83	0.90	0.89	0.96	1.07
$V+$ jets pure EW corrections	0.17	0.25	0.35	0.45	0.58	0.78	0.96	1.15	1.36	1.53	1.69	1.89	2.20
$V+$ jets mixed corrections	0.02	0.04	0.06	0.08	0.12	0.18	0.23	0.28	0.33	0.40	0.45	0.54	0.70
$V+$ jets PDF	0.02	0.03	0.01	0.03	0.04	0.08	0.16	0.23	0.30	0.35	0.39	0.48	0.69
VBF EW $V+$ jets bkgs.	0.02	0.03	0.04	0.06	0.09	0.13	0.17	0.16	0.16	0.03	0.24	0.53	1.10
Limited MC statistics	0.05	0.06	0.06	0.06	0.08	0.13	0.56	0.60	0.90	1.00	1.21	1.76	1.86
Total Uncertainty	1.48	1.26	1.13	1.14	1.22	1.42	1.76	2.03	2.41	2.69	2.98	3.53	4.21

Table E.1: Summary of the impact post-fit of systematic uncertainties on the total background in exclusive bins of p_T^{recoil} in the SR, as obtained from the background-only fit. The impact of each source of systematic is shown as the sum in quadrature of the individual contributions represented by the corresponding NPs.

F

Signals Cutflow

A study on the acceptance \times efficiency of the signal models considered in the analysis is performed. One representative sample is chosen for each interpretation and the ratio of remaining events is counted as the monojet SR cuts are applied one by one. A Dark Energy inspired model and an ALPs model, not discussed in detail in the core of this thesis, are also included in the study. The obtained results are shown in Tables [F.1](#), [F.2](#) and [F.3](#).

A truth-level cut is applied in all samples in order to avoid any effect from generation-level filters. This cut is set at $E_T^{\text{miss}} > 150$ GeV, with the exception of the ADD and ALPs models, in which case the cut is raised up to 350 GeV because of the filter that was applied in the generation of the samples.

SR cut	DMA (1, 2000)		DMP (1, 350)		SS (900, 895)		T4body (450, 443)	
Total events (truth- $E_T^{\text{miss}} > 150$)	10282	100.00%	199254	100.00%	5750	100.00%	39598	100.00%
Trigger	10101	98.23%	193342	97.03%	5651	98.27%	38851	98.11%
Event cleaning	10091	98.14%	193094	96.91%	5642	98.11%	38783	97.94%
Lepton veto	9788	95.19%	187094	93.90%	5435	94.51%	37547	94.82%
$1 \leq N_{\text{jets}} \leq 4$	9455	91.95%	176978	88.82%	5142	89.43%	35412	89.43%
$\min[\Delta\varphi(\text{jets}, E_T^{\text{miss}})]$ cut	9104	88.54%	168962	84.80%	4838	84.14%	33319	84.14%
Lead. Jet quality	8963	87.17%	160714	80.66%	4687	81.50%	31870	80.48%
Lead. Jet $p_T > 150$ & $ \eta < 2.4$	6642	64.60%	90366	45.35%	3508	61.00%	23134	58.42%
$E_T^{\text{miss}} > 200$	5317	51.71%	60133	30.18%	3018	52.48%	18801	47.48%
EM ₀	1346	13.09%	25162	12.63%	562	9.77%	4488	11.34%
EM ₁	1045	10.17%	15549	7.80%	536	9.32%	3789	9.57%
EM ₂	771	7.49%	8648	4.34%	416	7.23%	2857	7.21%
EM ₃	552	5.36%	4717	2.37%	316	5.50%	2111	5.33%
EM ₄	684	6.65%	4034	2.02%	439	7.63%	2618	6.61%
EM ₅	371	3.61%	1303	0.65%	267	4.65%	1352	3.41%
EM ₆	212	2.06%	444	0.22%	177	3.08%	712	1.80%
EM ₇	126	1.22%	156	0.08%	110	1.92%	393	0.99%
EM ₈	79	0.77%	67	0.03%	71	1.23%	204	0.52%
EM ₉	48	0.47%	28	0.01%	48	0.84%	122	0.31%
EM ₁₀	29	0.28%	12	0.01%	28	0.50%	58	0.15%
EM ₁₁	19	0.18%	7	0.00%	17	0.30%	42	0.11%
EM ₁₂	35	0.34%	6	0.00%	29	0.51%	55	0.14%

Table F.1: $A \times \varepsilon$ values for several signal benchmarks, namely DMA with $m(\chi, Z_A) = (1, 2000)$ GeV, DMP with $m(\chi, Z_P) = (1, 350)$ GeV, SS with $m(\tilde{q}, \tilde{\chi}_1^0) = (900, 895)$ GeV and T4body with $m(\tilde{t}, \tilde{\chi}_1^0) = (450, 443)$ GeV. A generator level cut on the truth E_T^{miss} at 150 GeV is applied for all the signatures. Missing transverse energies and jet p_T are expressed in GeV.

SR cut	BB (500, 300)		TT 600, 593)		DE ($C_2 = 1, M_2 = 1000$)		H(\rightarrow inv)	
Total events (truth- $E_T^{\text{miss}} > 150$)	95200	100.00%	8857	100.00%	102275	100.00%	289224	100.00%
Trigger	93802	98.53%	8694	98.15%	100300	98.07%	277197	95.84%
Event cleaning	93679	98.40%	8677	97.97%	100135	97.91%	276356	95.55%
Lepton veto	89103	93.60%	8352	94.29%	95799	93.67%	255785	88.44%
$1 \leq N_{\text{jets}} \leq 4$	74701	78.47%	7924	89.46%	86034	84.12%	235545	81.44%
$\min[\Delta\varphi(\text{jets}, E_T^{\text{miss}})]$ cut	66128	69.46%	7463	84.26%	78632	76.88%	225333	77.91%
Lead. Jet quality	64964	68.24%	7197	81.26%	76516	74.81%	211656	73.18%
Lead. Jet $p_T > 150$ & $ \eta < 2.4$	48148	50.58%	5379	60.73%	56942	55.68%	110272	38.13%
$E_T^{\text{miss}} > 200$	37203	39.08%	4444	50.17%	49799	48.69%	69456	24.01%
EM ₀	11972	12.58%	968	10.93%	8394	8.21%	31426	10.87%
EM ₁	11167	11.73%	804	9.08%	8282	8.10%	18081	6.25%
EM ₂	6670	7.01%	662	7.48%	6801	6.65%	9377	3.24%
EM ₃	3266	3.43%	493	5.57%	5424	5.30%	4816	1.67%
EM ₄	2670	2.80%	640	7.23%	7604	7.44%	3780	1.31%
EM ₅	870	0.91%	379	4.28%	4711	4.61%	1237	0.43%
EM ₆	332	0.35%	222	2.50%	2981	2.91%	426	0.15%
EM ₇	132	0.14%	112	1.27%	1950	1.91%	173	0.06%
EM ₈	61	0.06%	64	0.72%	1236	1.21%	69	0.02%
EM ₉	35	0.04%	40	0.45%	801	0.78%	36	0.01%
EM ₁₀	14	0.02%	26	0.29%	542	0.53%	17	0.01%
EM ₁₁	6	0.01%	12	0.13%	348	0.34%	8	0.00%
EM ₁₂	7	0.01%	21	0.24%	725	0.71%	10	0.00%

Table F.2: $\mathcal{A} \times \varepsilon$ values for several signal benchmarks, namely BB with $m(\tilde{b}, \tilde{\chi}_1^0) = (500, 300)$ GeV, TT with $m(\tilde{t}, \tilde{\chi}_1^0) = (600, 593)$ GeV, a Dark energy-inspired model with $C_2 = 1$ and $M_2 = 1000$ GeV and the invisible-decaying Higgs interpretation. A generator level cut on the truth E_T^{miss} at 150 GeV is applied for all the signatures. Missing transverse energies and jet p_T are expressed in GeV.

SR cut	ADD ($n = 4, M_D = 3600$)		ALPs ($C_{\tilde{G}} = 0.02, f_a = 1 \text{ TeV}$)	
Total events (truth- $E_T^{\text{miss}} > 350 \text{ GeV}$)	91657	100.00%	420880	100.00%
Trigger	91657	100.00%	420880	100.00%
Event cleaning	91494	99.82%	420455	99.90%
Lepton veto	87792	95.78%	412108	97.92%
$1 \leq N_{\text{jets}} \leq 4$	80072	87.36%	410466	97.53%
$\min[\Delta\phi(\text{jets}, E_T^{\text{miss}})]$ cut	74714	81.52%	408341	97.02%
Lead. Jet quality	74216	80.97%	402395	95.61%
Lead. Jet $p_T > 150 \text{ GeV}$ & $ \eta < 2.4$	73314	79.99%	397258	94.39%
$E_T^{\text{miss}} > 400 \text{ GeV}$	54059	58.98%	277923	66.03%
EM ₄	21140	23.06%	153724	36.52%
EM ₅	12155	13.26%	62404	14.83%
EM ₆	7145	7.79%	26916	6.40%
EM ₇	4560	4.98%	14227	3.38%
EM ₈	2960	3.23%	7054	1.68%
EM ₉	1927	2.10%	4145	0.98%
EM ₁₀	1319	1.44%	3305	0.79%
EM ₁₁	924	1.01%	2119	0.50%
EM ₁₂	1930	2.11%	4028	0.96%

Table F.3: $A \times \varepsilon$ values for an ADD model with $n = 4$ and $M_D = 3600$, and ALPs model with $C_{\tilde{G}} = 0.02$ and $f_a = 1 \text{ TeV}$. A generator level cut on the truth E_T^{miss} at 350 GeV is applied. Missing transverse energies and jet p_T are expressed in GeV.

G

V +jets reweighting impact

The resulting variation in the nominal yield as a consequence of the V +jets reweighting procedure, described in Chapter 2, is shown as a function of E_T^{miss} for the different V +jets processes in each region of the analysis in Figure G.1. An overall reduction of the number of predicted events is observed for all samples in all regions, excepting those cases where the process is negligible. The variation becomes bigger as E_T^{miss} increases, ranging between 2% and about 20% depending on the sample and the region.

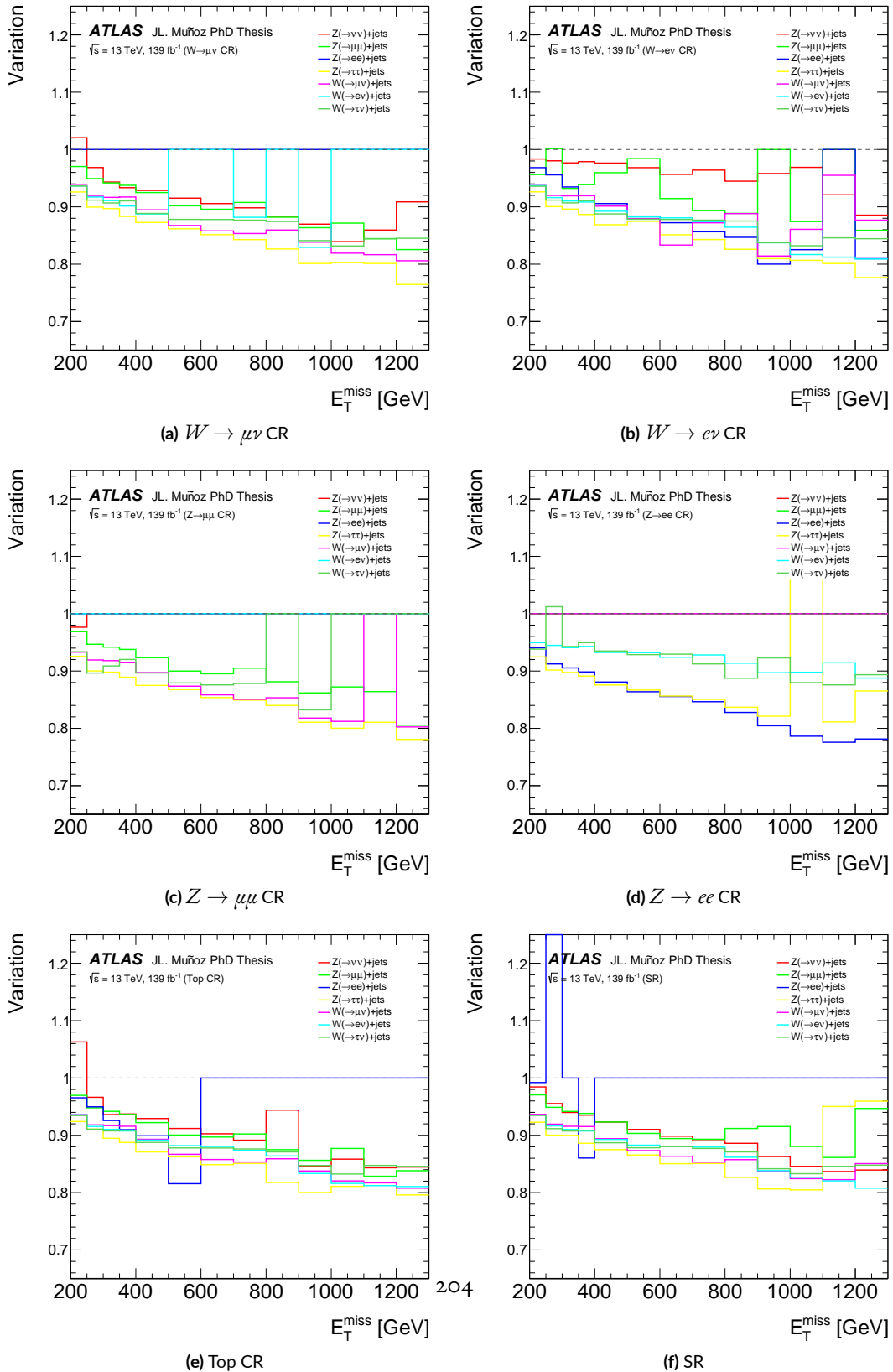


Figure G.1: Total variation in bins of E_T^{miss} after the V +jets corrections are applied

References

- [1] David J. Gross and Frank Wilczek. Asymptotically free gauge theories. ii. *Phys. Rev. D*, 9:980–993, Feb 1974.
- [2] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968.
- [3] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.
- [4] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [5] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.
- [6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964.
- [7] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.
- [8] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.
- [9] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.
- [10] R. P. Feynman. Mathematical formulation of the quantum theory of electromagnetic interaction. *Phys. Rev.*, 80:440–457, 1950.
- [11] Schwinger, Julian. Quantum Electrodynamics. I. A Covariant Formulation. *Phys. Rev.*, 74:1439–1461, Nov 1948.
- [12] S. Tomonaga. On a relativistically invariant formulation of the quantum theory of wave fields. *Prog. Theor. Phys.*, 1:27–42, 08 1946.
- [13] H. Fritzsch, Murray Gell-Mann, and H. Leutwyler. Advantages of the Color Octet Gluon Picture. *Phys. Lett. B*, 47:365–368, 1973.

- [14] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008.
- [15] Determination of the strong coupling constant and test of asymptotic freedom from Transverse Energy-Energy Correlations in multijet events at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical report, CERN, Geneva, Jul 2020. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2020-025>.
- [16] Fermi, Enrico. Tentativo di una Teoria Dei Raggi β . *Il Nuovo Cimento*, 11(1):1–19, jan 1934.
- [17] Howard Georgi and Sheldon L. Glashow. Unified weak and electromagnetic interactions without neutral currents. *Phys. Rev. Lett.*, 28:1494, 1972.
- [18] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken Symmetries. *Phys. Rev.*, 127:965–970, 1962.
- [19] Tung-Mow Yan and Sidney D. Drell. The Parton Model and its Applications. *Int. J. Mod. Phys. A*, 29:0071, 2014.
- [20] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989.
- [21] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.
- [22] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.
- [23] F. Hautmann, H. Jung, M. Krämer, P. J. Mulders, E. R. Nocera, T. C. Rogers, and A. Signori. TMDlib and TMDplotter: library and plotting tools for transverse-momentum-dependent parton distributions. *Eur. Phys. J. C*, 74:3220, 2014.
- [24] S. Hoche. Introduction to parton-shower event generators. 2014.
- [25] V. V. Sudakov. Vertex parts at very high-energies in quantum electrodynamics. *Sov. Phys. JETP*, 3:65–71, 1956.
- [26] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber. QCD matrix elements + parton showers. *JHEP*, 11:063, 2001.
- [27] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Treccani. Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP*, 01:013, 2007.
- [28] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2):31–145, 1983.

- [29] B.R. Webber. A qcd model for jet fragmentation including soft gluon interference. *Nuclear Physics B*, 238(3):492–528, 1984.
- [30] T. Sjostrand and Peter Z. Skands. Multiple interactions and the structure of beam remnants. *JHEP*, 03:053, 2004.
- [31] Georges Aad et al. Measurement of the underlying event in jet events from 7 TeV proton-proton collisions with the ATLAS detector. *Eur. Phys. J. C*, 74(8):2965, 2014.
- [32] Albert Einstein. The Foundation of the General Theory of Relativity. *Annalen Phys.*, 49(7):769–822, 1916.
- [33] B. P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, 2016.
- [34] F. Zwicky. On the Masses of Nebulae and of Clusters of Nebulae. *Astrophys. J.*, 86:217–246, 1937.
- [35] Vera C. Rubin and Jr. Ford, W. Kent. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *Astrophys. J.*, 159:379, February 1970. Provided by the SAO/NASA Astrophysics Data System.
- [36] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020.
- [37] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. NINE-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: COSMOLOGICAL PARAMETER RESULTS. *The Astrophysical Journal Supplement Series*, 208(2):19, sep 2013.
- [38] S. Perlmutter et al. Discovery of a supernova explosion at half the age of the Universe and its cosmological implications. *Nature*, 391:51–54, 1998.
- [39] Peter M. Garnavich et al. Constraints on cosmological models from Hubble Space Telescope observations of high z supernovae. *Astrophys. J. Lett.*, 493:L53–57, 1998.
- [40] M.C. Gonzalez-Garcia, Michele Maltoni, and Thomas Schwetz. Global analyses of neutrino oscillation experiments. *Nuclear Physics B*, 908:199–217, 2016. Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015.
- [41] B. Pontecorvo. Mesonium and anti-mesonium. *Sov. Phys. JETP*, 6:429, 1957.
- [42] B. Pontecorvo. Neutrino Experiments and the Problem of Conservation of Leptonic Charge. *Zh. Eksp. Teor. Fiz.*, 53:1717–1725, 1967.

- [43] Roel Aaij et al. Test of lepton universality in beauty-quark decays. 3 2021.
- [44] B. Abi et al. Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm. *Phys. Rev. Lett.*, 126(14):141801, 2021.
- [45] Th. Kaluza. Zum Unitätsproblem der Physik. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1921:966–972, 1921.
- [46] Oskar Klein. Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English). *Z. Phys.*, 37:895–906, 1926.
- [47] O. Klein. The Atomicity of Electricity as a Quantum Theory Law. *Nature*, 118:516, 1926.
- [48] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. The Hierarchy problem and new dimensions at a millimeter. *Phys. Lett. B*, 429:263–272, 1998.
- [49] Morad Aaboud et al. Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector. *JHEP*, 01:126, 2018.
- [50] A. Arbey and F. Mahmoudi. Dark matter and the early Universe: a review. *Prog. Part. Nucl. Phys.*, 119:103865, 2021.
- [51] Jonathan L. Feng. Dark Matter Candidates from Particle Physics and Methods of Detection. *Ann. Rev. Astron. Astrophys.*, 48:495–545, 2010.
- [52] E. Aprile et al. The XENON100 Dark Matter Experiment. *Astropart. Phys.*, 35:573–590, 2012.
- [53] Z. Ahmed et al. Dark Matter Search Results from the CDMS II Experiment. *Science*, 327:1619–1621, 2010.
- [54] D. S. Akerib et al. Improved Limits on Scattering of Weakly Interacting Massive Particles from Reanalysis of 2013 LUX Data. *Phys. Rev. Lett.*, 116(16):161301, 2016.
- [55] Rafael F Lang and Wolfgang Seidel. Search for dark matter with CRESST. *New Journal of Physics*, 11(10):105017, oct 2009.
- [56] J. Cortina, F. Goebel, Thomas Schweizer Institut de Fisica d’Altes Energies, and M. F. Physik. Technical performance of the magic telescopes. *arXiv: Instrumentation and Methods for Astrophysics*, 2009.
- [57] W. B. Atwood, A. A. Abdo, M. Ackermann, W. Althouse, B. Anderson, M. Axelsson, L. Baldini, J. Ballet, D. L. Band, G. Barbiellini, and et al. The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission. *The Astrophysical Journal*, 697(2):1071–1102, June 2009.

- [58] A. Achterberg et al. First Year Performance of The IceCube Neutrino Telescope. *Astropart. Phys.*, 26:155–173, 2006.
- [59] Christopher W. Walter. The Super-Kamiokande Experiment. 2 2008.
- [60] E. Aslanides et al. A deep sea telescope for high-energy neutrinos. 5 1999.
- [61] 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.
- [62] Giorgio Busoni 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.
- [63] Oliver Buchmueller, Matthew J. Dolan, Sarah A. Malik, and Christopher McCabe. Characterising dark matter searches at colliders and direct detection experiments: Vector mediators. *JHEP*, 01:037, 2015.
- [64] Pierre Fayet. Supersymmetry and Weak, Electromagnetic and Strong Interactions. *Phys. Lett. B*, 64:159, 1976.
- [65] Pierre Fayet. Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions. *Phys. Lett. B*, 69:489, 1977.
- [66] Adel Bilal. Introduction to supersymmetry. 1 2001.
- [67] J.-L. Gervais and B. Sakita. Field theory interpretation of supergauges in dual models. *Nuclear Physics B*, 34(2):632–639, 1971.
- [68] Brian Patt and Frank Wilczek. Higgs-field portal into hidden sectors. 5 2006.
- [69] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 7 2013.
- [70] Morad Aaboud et al. Combination of searches for invisible Higgs boson decays with the ATLAS experiment. *Phys. Rev. Lett.*, 122(23):231801, 2019.
- [71] Albert M Sirunyan et al. Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 793:520–551, 2019.
- [72] Philippe Brax, Clare Burrage, Christoph Englert, and Michael Spannowsky. LHC Signatures Of Scalar Dark Energy. *Phys. Rev. D*, 94(8):084054, 2016.
- [73] Enrico Bothmann et al. Event Generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019.
- [74] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

- [75] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [76] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.
- [77] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J. C*, 58:639–707, 2008.
- [78] Carlo Oleari. The POWHEG-BOX. *Nucl. Phys. B Proc. Suppl.*, 205-206:36–41, 2010.
- [79] 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.
- [80] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.
- [81] 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.
- [82] Fabio Maltoni and Tim Stelzer. MadEvent: Automatic event generation with MadGraph. *JHEP*, 02:027, 2003.
- [83] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.
- [84] David J. Lange. The evtgen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152 – 155, 2001. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.
- [85] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.
- [86] G. Aad et al. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823–874, 2010.
- [87] Elzbieta Richter-Was, D. Froidevaux, and Luc Poggioli. ATLFAST 2.0 a fast simulation package for ATLAS. 11 1998.
- [88] Tanju Gleisberg and Stefan Höche. Comix, a new matrix element generator. *JHEP*, 0812:039, 2008.
- [89] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini. Scattering Amplitudes with Open Loops. *Phys. Rev. Lett.*, 108:111601, 2012.

- [90] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. QCD matrix elements + parton showers: The NLO case. *JHEP*, 04:027, 2013.
- [91] K. Arnold, M. Bähr, G. Bozzi, F. Campanario, C. Englert, T. Figy, N. Greiner, C. Hackstein, V. Hankele, B. Jäger, and et al. Vbfno: A parton level monte carlo for processes with electroweak bosons. *Computer Physics Communications*, 180(9):1661–1670, Sep 2009.
- [92] Search for invisible Higgs boson decays with vector boson fusion signatures with the ATLAS detector using an integrated luminosity of 139 fb^{-1} . Technical report, CERN, Geneva, Apr 2020. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2020-008>.
- [93] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013.
- [94] ATLAS Pythia 8 tunes to 7 TeV data. 11 2014.
- [95] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev. D*, 82:074018, 2010.
- [96] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244–289, 2013.
- [97] Georges Aad et al. Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ with the ATLAS detector. *Eur. Phys. J. C*, 75(7):299, 2015. [Erratum: *Eur.Phys.J.C* 75, 408 (2015)].
- [98] Georges Aad et al. Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ with the ATLAS detector. *JHEP*, 09:145, 2014.
- [99] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. MINLO: Multi-Scale Improved NLO. *JHEP*, 10:155, 2012.
- [100] Michiel Botje et al. The PDF4LHC Working Group Interim Recommendations. 1 2011.
- [101] W. Beenakker, M. Kramer, T. Plehn, M. Spira, and P. M. Zerwas. Stop production at hadron colliders. *Nucl. Phys. B*, 515:3–14, 1998.
- [102] Wim Beenakker, Silja Breusung, Michael Kramer, Anna Kulesza, Eric Laenen, and Irene Niessen. Supersymmetric top and bottom squark production at hadron colliders. *JHEP*, 08:098, 2010.
- [103] Wim Beenakker, Christoph Borschensky, Michael Krämer, Anna Kulesza, and Eric Laenen. NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation. *JHEP*, 12:133, 2016.

- [104] Wim Beenakker, Christoph Borschensky, Raphael Heger, Michael Krämer, Anna Kulesza, and Eric Laenen. NNLL resummation for stop pair-production at the LHC. *JHEP*, 05:153, 2016.
- [105] The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model. 8 2016.
- [106] J. M. Lindert et al. Precise predictions for $V+$ jets dark matter backgrounds. *Eur. Phys. J. C*, 77(12):829, 2017.
- [107] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, and T. A. Morgan. Precise QCD predictions for the production of a Z boson in association with a hadronic jet. *Phys. Rev. Lett.*, 117(2):022001, 2016.
- [108] Aude Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, and T. A. Morgan. The NNLO QCD corrections to Z boson production at large transverse momentum. *JHEP*, 07:133, 2016.
- [109] Radja Boughezal, Xiaohui Liu, and Frank Petriello. Phenomenology of the Z-boson plus jet process at NNLO. *Phys. Rev. D*, 94(7):074015, 2016.
- [110] Radja Boughezal, Xiaohui Liu, and Frank Petriello. W-boson plus jet differential distributions at NNLO in QCD. *Phys. Rev. D*, 94(11):113009, 2016.
- [111] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Muck. Electroweak corrections to dilepton + jet production at hadron colliders. *JHEP*, 06:069, 2011.
- [112] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Mück. Electroweak corrections to monojet production at the LHC. *Eur. Phys. J. C*, 73(2):2297, 2013.
- [113] Stefan Kallweit, Jonas M. Lindert, Philipp Maierhofer, Stefano Pozzorini, and Marek Schönherr. NLO QCD+EW predictions for $V +$ jets including off-shell vector-boson decays and multijet merging. *JHEP*, 04:021, 2016.
- [114] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Muck. Electroweak corrections to $W +$ jet hadroproduction including leptonic W-boson decays. *JHEP*, 08:075, 2009.
- [115] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. Logarithmic electroweak corrections to hadronic $Z+1$ jet production at large transverse momentum. *Phys. Lett. B*, 609:277–285, 2005.
- [116] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. Electroweak corrections to large transverse momentum production of W bosons at the LHC. *Phys. Lett. B*, 651:160–165, 2007.
- [117] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. One-loop weak corrections to hadronic production of Z bosons at large transverse momenta. *Nucl. Phys. B*, 727:368–394, 2005.

- [118] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. Electroweak corrections to hadronic production of W bosons at large transverse momenta. *Nucl. Phys. B*, 797:27–77, 2008.
- [119] Aneesh Manohar, Paolo Nason, Gavin P. Salam, and Giulia Zanderighi. How bright is the proton? A precise determination of the photon parton distribution function. *Phys. Rev. Lett.*, 117(24):242002, 2016.
- [120] Carl Schmidt, Jon Pumplin, Daniel Stump, and C. P. Yuan. CT14QED parton distribution functions from isolated photon production in deep inelastic scattering. *Phys. Rev. D*, 93(11):114015, 2016.
- [121] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001–S08001, 2008.
- [122] G Aad et al. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation*, 3(08):S08003–S08003, 2008.
- [123] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004–S08004, aug 2008.
- [124] Jr. Alves, A. Augusto et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.
- [125] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.
- [126] M Capeans, G Darbo, K Einsweiler, 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 CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010.
- [127] G Aad et al. Performance of the ATLAS Trigger System in 2015. Performance of the ATLAS Trigger System in 2015. *Eur. Phys. J. C*, 77(CERN-EP-2016-241. 5):317. 76 p, Nov 2016.
- [128] Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. Technical Report ATLAS-CONF-2019-021, CERN, Geneva, Jun 2019.
- [129] P. Grafström and W. Kozanecki. Luminosity determination at proton colliders. *Progress in Particle and Nuclear Physics*, 81:97 – 148, 2015.
- [130] Peter Jenni, Markus Nordberg, Marzio Nessi, and Kerstin Jon-And. *ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity*. Technical Design Report ATLAS. CERN, Geneva, 2008.
- [131] M. Aaboud et al. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2. *Eur. Phys. J. C*, 77(10):673, 2017.

- [132] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyses, and A. Salzburger. The new ATLAS track reconstruction (NEWT). *J. Phys. Conf. Ser.*, 119:032014, 2008.
- [133] S. Boutle et al. Primary vertex reconstruction at the ATLAS experiment. *J. Phys. Conf. Ser.*, 898(4):042056, 2017.
- [134] Wolfgang Waltenberger. Adaptive vertex reconstruction. 7 2008.
- [135] Georges Aad et al. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *JINST*, 14(12):P12006, 2019.
- [136] Morad Aaboud et al. Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data. *JINST*, 14(03):P03017, 2019.
- [137] Morad Aaboud et al. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data. *Eur. Phys. J. C*, 77(3):195, 2017.
- [138] Morad Aaboud et al. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(8):639, 2019.
- [139] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119–126, 2008.
- [140] Georges Aad et al. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 76(5):292, 2016.
- [141] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008.
- [142] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012.
- [143] Georges Aad et al. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77:490, 2017.
- [144] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119–126, 2008.
- [145] Georges Aad et al. Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 7 2020.
- [146] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, 1993.
- [147] Morad Aaboud et al. In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector. *Eur. Phys. J. C*, 79(2):135, 2019.

- [148] Georges Aad et al. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 73(3):2304, 2013.
- [149] M. Aaboud et al. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 96(7):072002, 2017.
- [150] Determination of the jet energy scale and resolution at ATLAS using Z/γ -jet events in data at $\sqrt{s} = 8$ TeV. 10 2015.
- [151] TeV-scale jet energy calibration using multijet events including close-by jet effects at the ATLAS experiment. 1 2013.
- [152] Tagging and suppression of pileup jets. 5 2014.
- [153] Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector. Technical Report ATLAS-CONF-2015-029, CERN, Geneva, Jul 2015.
- [154] Georges Aad et al. Characterisation and mitigation of beam-induced backgrounds observed in the ATLAS detector during the 2011 proton-proton run. *JINST*, 8:P07004, 2013.
- [155] Georges Aad et al. Performance of b -Jet Identification in the ATLAS Experiment. *JINST*, 11(04):P04008, 2016.
- [156] Yann Coadou. Boosted Decision Trees and Applications. *EPJ Web Conf.*, 55:02004, 2013.
- [157] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. 3 2007.
- [158] Optimisation of the ATLAS b -tagging performance for the 2016 LHC Run. 2016.
- [159] Morad Aaboud et al. Measurements of b -jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *JHEP*, 08:089, 2018.
- [160] Georges Aad et al. Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s}=8$ TeV. *Eur. Phys. J. C*, 75(7):303, 2015.
- [161] T. Barillari et al. Local hadronic calibration. 2009.
- [162] Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV. 5 2017.
- [163] Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. 2015.
- [164] Morad Aaboud et al. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 78(11):903, 2018.

- [165] G. Cowan. *Statistical data analysis*. Oxford University Press, USA, 1998.
- [166] 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. [Erratum: *Eur. Phys. J. C* 73, 2501 (2013)].
- [167] Alexander L. Read. Presentation of search results: The CL(s) technique. *J. Phys. G*, 28:2693–2704, 2002.
- [168] John Aldrich. R.A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3):162 – 176, 1997.
- [169] M. Baak, G. J. Besjes, D. Côte, A. Koutsman, J. Lorenz, and D. Short. HistFitter software framework for statistical data analysis. *Eur. Phys. J. C*, 75:153, 2015.
- [170] Georges Aad et al. Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using 4.7 fb⁻¹ of $\sqrt{s} = 7$ TeV proton-proton collision data. *Phys. Rev. D*, 87(1):012008, 2013.
- [171] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.
- [172] Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO. Technical report, CERN, Geneva, May 2017. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-007>.
- [173] Studies on top-quark Monte Carlo modelling for Top2016. Technical report, CERN, Geneva, Sep 2016. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-020>.
- [174] Morad Aaboud et al. Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Phys. Rev. D*, 94(3):032005, 2016.
- [175] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C. P. Yuan. CT10 next-to-next-to-leading order global analysis of QCD. *Phys. Rev. D*, 89(3):033009, 2014.
- [176] 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.
- [177] Andy Buckley, James Ferrando, Stephen Lloyd, Karl Nordström, Ben Page, Martin Rufenacht, Marek Schönherr, and Graeme Watt. LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J. C*, 75:132, 2015.

- [178] 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. *Comput. Phys. Commun.*, 195:161–171, 2015.
- [179] Vardan Khachatryan et al. Search for dark matter, extra dimensions, and unparticles in monojet events in proton–proton collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, 75(5):235, 2015.
- [180] Gian F. Giudice, Riccardo Rattazzi, and James D. Wells. Quantum gravity and extra dimensions at high-energy colliders. *Nucl. Phys. B*, 544:3–38, 1999.
- [181] Mihailo Backović, Antony Martini, Olivier Mattelaer, Kyoungchul Kong, and Gopolang Mohlabeng. Direct Detection of Dark Matter with MadDM v.2.0. *Phys. Dark Univ.*, 9-10:37–50, 5 2015.
- [182] C. Amole et al. Dark Matter Search Results from the Complete Exposure of the PICO-60 C₃F₈ Bubble Chamber. *Phys. Rev. D*, 100(2):022001, 2019.
- [183] D. S. Akerib et al. Limits on spin-dependent WIMP-nucleon cross section obtained from the complete LUX exposure. *Phys. Rev. Lett.*, 118(25):251302, 2017.
- [184] E. Aprile et al. Constraining the spin-dependent WIMP-nucleon cross sections with XENON1T. *Phys. Rev. Lett.*, 122(14):141301, 2019.
- [185] Combination of searches for invisible Higgs boson decays with the ATLAS experiment. Technical report, CERN, Geneva, Oct 2020. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2020-052>.
- [186] Armen Tumasyan et al. Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. 7 2021.
- [187] Search for scalar dark energy in $t\bar{t} + E_T^{\text{miss}}$ and mono-jet final states with the ATLAS detector. Technical report, CERN, Geneva, Jun 2018. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-008>.
- [188] M. Capeans, G. Darbo, K. Einsweiler, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet. ATLAS Insertable B-Layer Technical Design Report. 9 2010.
- [189] Claudia Marcelloni De Oliveira. IBL installation into the inner detector of the ATLAS Experiment side C. General Photo, May 2014.
- [190] V Zivkovic, J-D Schipper, M Garcia-Sciveres, A Mekkaoui, M Barbero, G Darbo, D Gnani, T Hemperek, M Menouni, D Fougeron, F Gensolen, F Jensen, L Caminada, V Gromov,

R Kluit, J Fleury, H Krüger, M Backhaus, X Fang, L Gonella, A Rozanov, and D Arutinov. The FE-i4 pixel readout system-on-chip resubmission for the insertable b-layer project. *Journal of Instrumentation*, 7(02):C02050–C02050, feb 2012.

- [191] C. Goessling, R. Klingenberg, D. Muenstermann, A. Rummler, G. Troska, and T. Wittig. Planar n+-in-n silicon pixel sensors for the atlas ibl upgrade. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 650(1):198–201, 2011. International Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010.
- [192] S.I. Parker, C.J. Kenney, and J. Segal. 3d — a proposed new architecture for solid-state radiation detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 395(3):328–343, 1997. Proceedings of the Third International Workshop on Semiconductor Pixel Detectors for Particles and X-rays.
- [193] IBL Efficiency and Single Point Resolution in Collision Events. Technical report, CERN, Geneva, Aug 2016. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-INDET-PUB-2016-001>.

Acknowledgments

Behind the work described in these pages there are years of learning, suffering and traveling. A complete journey full of ups and downs in which I had the chance to interact with many people who ended up contributing significantly to bring this journey to a happy end.

I would like to start thanking my supervisor, Mario Martínez, for his infinite patience. Without his clear mind and straight feedback this thesis would definitely not exist at all. I would also like to extend my gratitude to Martine Bosman, Aurelio Juste and Imma Riu for their assistance at every stage of the project. In general, a big thanks to IFAE for this great opportunity. In addition, I would like to thank Sebastian Grinstein, from IFAE, and Marco Battaglia, from UCSC, for supervising my work during my ATLAS Qualification Tasks.

First steps are never easy and mine in ATLAS were not an exception. For this reason, I am deeply thankful for having worked with Rachel Rosten because without her supervision my life would have been much harder, thanks for being available at any time even after changing jobs. I would also like to thank Giuliano Gustavino for his kindly leadership of the analysis team, extremely useful suggestions and interesting conversations. Also, thanks to Remi Zaidan, Vagelis Gkougkousis and Danijela Bogavac for helping me with the work at different stages.

Either in Barcelona or Geneva, I had a great time sharing the office and occasional meals with different colleagues. In this regard, thanks to Sergio, Bernat, Marco, Daniel, Marc, Eli, Stergios, Chiara and Paula for making the experience much more enjoyable. And, especially, a big thanks to Andrea for showing me around when I first arrived to CERN and becoming almost a big sister for me.

Work apart, there is a large number of people who had a direct influence on this academic journey, forgive me if I forget someone to mention. Thanks to Ceci, Roger, Nacho and Fabio for being a small family during my first months, to my incredible flatmates Irene, Will and Orestis, to Aleks for the never ending conversations about life in OB (that was probably how I got decently fluent in English), to Owen, Danielle, Alex, Hans and the rest of the ‘gang’ for all those R1 Fridays and awesome trips, to Ana for showing me there is life out of CERN in Geneva and her very special friendship, to

the ‘foxes’ Pedro, Riccardo and Óscar for the fun in our rehearsals and performing at the Hardronic festival (which is one of my favourite memories ever), to Mar for making me laugh every 5 minutes, and, to complete the list: thanks to Adrianna, Pablo, Carlito, Emma, Barbosa, Oksana, Alessandro, Esther, Michael, Pascal, Prajita, Kinga, Tjebbe, Alexia, Micol, Amanda, Giannis and Flory.

To my parents and my sister for their unconditional support, trust and not very hopeful attempts to understand what I was doing. A very special thanks also to Marina and Simon, my ‘adoptive’ Catalan parents, for their enormous generosity and kindness.

Thanks to my older friends Irene, Lidia, Jose Antonio, Fran, Álvaro, Práxedes, Fátima, Amaia, Sergio and Pablo, among many others, for their support from the distance and eventual visits. Same gratitude shall be extended to my cousins and my uncles. In addition, a special thanks to Andreu Buenafuente and Berto Romero for the priceless company in so many nights of work and/or insomnia.

Finally, to Ceci (again). Because you were there at the beginning but you showed up again at the end of the journey, and never stopped making my life easier. You were there in some of the darkest moments and I do not know how to thank you enough, but I will never forget it. This thesis is as yours as mine.

The last stages of my PhD –including the writing of this thesis– took place during the outbreak of the pandemic caused by the COVID-19 disease. For this reason, I could not finish these lines without a deep thank to all those healthcare and essential workers that held the line through the toughest moments. Long life to the public health system.

